Identification-robust inference for endogeneity parameters in linear structural models

Firmin Doko Tchatoka and Jean-Marie Dufour

McGill University, University of Tasmania, Department of Economics, Hobart, Australia

16 August 2012

Identification-robust inference for endogeneity parameters in linear structural models *

Firmin Doko Tchatoka[†] Jean-Marie Dufour [‡]
University of Tasmania McGill University

August 2012

* The authors thank Marine Carrasco, Jan Kiviet and Benoit Perron for several useful comments. This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), 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, and the Fonds de recherche sur la société et la culture (Québec).

† School of Economics and Finance, University of Tasmania, Private Bag 85, Hobart TAS 7001, Tel: +613 6226 7226, Fax:+61 3 6226 7587; e-mail: Firmin.dokotchatoka@utas.edu.au. Homepage: http://www.fdokotchatoka.com

‡ William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 4400 ext. 09156; FAX: (1) 514 398 4800; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

//

ABSTRACT

We provide a generalization of the Anderson-Rubin (AR) procedure for inference on parameters which represent the dependence between possibly endogenous explanatory variables and disturbances in a linear structural equation (*endogeneity parameters*). We focus on second-order dependence and stress the distinction between *regression* and *covariance endogeneity parameters*. Such parameters have intrinsic interest (because they measure the effect of "common factors" which induce simultaneity) and play a central role in selecting an estimation method (because they determine "simultaneity biases" associated with least-squares methods). We observe that endogeneity parameters may not identifiable and we give the relevant identification conditions. We develop identification-robust finite-sample tests for joint hypotheses involving structural and regression endogeneity parameters, as well as marginal hypotheses on regression endogeneity parameters. For Gaussian errors, we provide tests and confidence sets based on standard-type Fisher critical values. For a wide class of parametric non-Gaussian errors (possibly heavy-tailed), we also show that exact Monte Carlo procedures can be applied using the statistics considered. As a special case, this result also holds for usual AR-type tests on structural coefficients. For covariance endogeneity parameters, we supply an asymptotic (identification-robust) distributional theory. Tests for partial exogeneity hypotheses (for individual potentially endogenous explanatory variables) are covered as instances of the class of proposed procedures. The proposed procedures are applied to two empirical examples: the relation between trade and economic growth, and the widely studied problem of returns to education.

Key words: Identification-robust confidence sets; endogeneity; AR-type statistic; projection-based techniques; partial exogeneity test.

Journal of Economic Literature classification: C3; C12; C15; C52.

Contents

List of Tables

List of Figures

List of Definitions, Assumptions, Propositions and Theorems

1. Introduction

Instrumental variable (IV) regressions are typically motivated by the fact that "explanatory variables" may be correlated with the error term, so least-squares methods yield biased inconsistent estimators of model coefficients. Such IV parameter estimates can be interpreted as measures of the relationship between variables, once the "effect" of common "driving" (or "exogenous") variables has been eliminated. Even though coefficients estimated in this way may have interesting interpretations from the viewpoint of economic theory, inference on such "structural parameters" raises identification difficulties. Further, it is well known that IV estimators may be very imprecise, and inference procedures (such as tests and confidence sets) can be highly unreliable, especially when instruments are weakly associated with model variables (weak instruments). This has led to a large literature aimed at producing reliable inference in the presence of weak instruments; see the reviews of Stock, Wright and Yogo (2002) and Dufour (2003).

Research on weak instruments has focused on inference for the coefficients of endogenous variables in so-called "IV regressions". This leaves out the parameters which specifically determine simultaneity features, such as the covariances between endogenous explanatory variables and disturbances. These parameters can be of interest for several reasons. *First*, they provide direct measures of the importance of "common factors" which induce simultaneity. Such factors are in a sense "left out" from "structural equations", but they remain "hidden" in "structural disturbances". For example, in a wide set of economic models, they may represent unobserved latent variables, such as "surprise variables" which play a role in models with expectations [see Barro (1977), Dufour and Jasiak (2001)]. *Second*, the simultaneity covariance (or regression) coefficients determine the estimation bias of least-squares methods. Information on the size of such biases can be useful in interpreting least-squares estimates and related statistics. *Third*, information on the parameters of hidden variables (which induce simultaneity) may be important for selecting statistical procedures. Even if instruments are "strong", it is well known that IV estimators may be considerably less efficient than least-squares estimators; see Kiviet and Niemczyk (2007) and Doko Tchatoka and Dufour (2011). Indeed, this may be the case even when endogeneity is present. If a variable is not correlated (or only weakly correlated) with the error term, instrumenting it can lead to sizable efficiency losses in estimation. Assessing when and which variables should be instrumented is an important issue for the estimation of structural models.

In this paper, we stress the view that linear structural models (IV regressions) can be interpreted as regressions with missing regressors. If the missing regressors were included, there would be no simultaneity bias, so no correction for simultaneity – such as IV methods – would be needed. This feature allows one to define a model transformation that maps a linear structural equation (with simultaneity) to a linear regression where all the explanatory variables are uncorrelated with the error term. We call this transformed equation the *orthogonalized structural equation*. Interestingly, the latter is not a reduced-form equation. Rather, the orthogonalized structural equation still involves the structural parameters of interest, but also includes *endogeneity parameters* which are "hidden" in the original structural equation. We focus here on this orthogonalized structural equation.

The problem stems from the fact that the missing regressors are unobserved. Despite this difficulty, we show that procedures similar to the one proposed by Anderson and Rubin (1949, AR) can be applied to the orthogonalized equation. This allows one to make inference jointly on both the parameters of the original structural equation and endogeneity parameters. Two types of endogeneity parameters are considered: *regression endogeneity parameters* and *covariance endogeneity parameters*. Under standard conditions, where instruments are strictly exogenous and errors are Gaussian, the tests and confidence sets derived in this way are exact. The proposed methods do not require identification assumptions, so they can be described as *identification-robust*. For more general inference on transformations of the parameters of the orthogonalized structural equation, we propose projection methods, for such techniques allow for a simple finite-sample distributional theory and preserve robustness to identification assumptions.

To be more specific, we consider a model of the form

$$y = Y\beta + X_1\gamma + u$$

where y is an observed dependent variable, Y is a matrix of observed (possibly) endogenous regressors, and X_1 is a matrix of exogenous variables. We observe that AR-type procedures may be applied to test hypotheses on the transformed parameter $\theta = \beta + a$, where a represents regression coefficients of u on the reduced-form errors of Y (*regression endogeneity parameters*). Identification-robust inference on a itself is then derived by exploiting the possibility of making identification-robust inference on β. Then, inference on covariances (say σ_{Vu}) between u and Y (*covariance endogeneity parameters*) can be derived by considering appropriate linear transformations of a.

We stress that regression and covariance endogeneity parameters – though theoretically related – play distinct but complementary roles: regression endogeneity parameters represent the effect of reduced-form innovations on y, while covariance endogeneity parameters determine the need to instrument different variables in Y. When $\sigma_{Vu} = 0$, Y can be treated as exogenous (so IV estimation is not warranted). So-called exogeneity tests typically test the hypothesis $\sigma_{Vu} = 0$. It is easy to see that $\sigma_{Vu} = 0$ if and only if $a = 0$ (provided the covariance matrix between reduced-form errors is nonsingular), but the relationship is more complex in other cases. In this paper, we emphasize cases where $a \neq 0$. Due to the failure of the exogeneity hypothesis, the distributions of various test statistics are much more complex. Interestingly, it is relatively easy to produce finite-sample inference on a, but not on σ_{Vu}. So, for σ_{Vu}, we only propose asymptotically valid tests and confidence sets.

By allowing $a \neq 0$ (or $\sigma_{Vu} \neq 0$), we extend earlier results on exogeneity tests, which focus on the null hypothesis $H_a : a = 0$. The literature on this topic, is considerable; see, for example, Durbin (1954), Wu (1973, 1974, 1983a, 1983b), Revankar and Hartley (1973), Farebrother (1976), Hausman (1978), Revankar (1978), Dufour (1979, 1987), Hwang (1980), Kariya and Hodoshima (1980), Hausman and Taylor (1981), Spencer and Berk (1981), Nakamura and Nakamura (1981), Engle (1982), Holly (1982), Reynolds (1982), Smith (1984), Staiger and Stock (1997), Doko Tchatoka and Dufour (2010, 2011). By contrast, we consider here the problem of testing any value of a (or σ_{Vu}) and build confidence sets for these parameters. By allowing weak instruments, we extend the results in Dufour (1987) where Wald-type tests and confidence sets are proposed for inference on a and σ_{Vu}, under assumptions which exclude weak instruments. Finally, by considering inference on a and σ_{Vu}, we extend a procedure proposed in Dufour and Jasiak (2001) for inference on the aggregate parameter $\theta = \beta + a$ (but not a or σ_{Vu}) in the context of a somewhat different model.

On exploiting results from Dufour and Taamouti (2005, 2007), we supply analytical forms for the proposed confidence sets, and we give the necessary and sufficient conditions under which they are bounded. These results can be used to assess partial exogeneity hypotheses even when identification is deficient or weak.

In order to allow for alternative assumptions on error distributions, we show that the proposed AR-type statistics are pivotal as long as the errors follow a completely specified distribution (up to an unknown scale parameter), which may be non-Gaussian. Under such conditions, exact Monte Carlo tests can be performed without a Gaussian assumptions [as described in Dufour (2006)]. On allowing for more general error distributions and weakly exogenous instruments (along with standard high-level asymptotic assumptions), we also show that the proposed procedures remain asymptotically valid and identification-robust.

Finally, we apply the proposed methods to two empirical examples, previously considered in Dufour and Taamouti (2007): a study of the relationship between trade and economic growth [Frankel and Romer (1999)], and the widely considered example of returns to education [Bound, Jaeger and Baker (1995)].

The paper is organized as follows. Section 2 formulates the model considered. Section 3 presents the finite-sample theory for inference on regression endogeneity parameters. Section 4 discusses asymptotic theory and inference for covariance endogeneity parameters. Section 5 illustrates the theoretical results through two empirical applications: a model of the relationship between trade and growth model, and returns to schooling. We conclude in Section 6. Proofs are presented in appendix.

Throughout the paper, I_m stands for the identity matrix of order m. For any full rank $T \times m$ matrix A, $P(A) = A(A'A)^{-1}A'$ is the projection matrix on the space spanned by the columns of A, $M(A) = I_T - P(A)$, and $vec(A)$ is the $(Tm) \times 1$ dimensional column vectorization of A. For any squared matrix B, the notation $B > 0$ means that B is positive definite (p.d.), while $B \geq 0$ means it is positive semidefinite (p.s.d.). Finally, " $\xrightarrow{p}$ " stands for convergence in probability while " $\xrightarrow{L}$ " is for convergence in distribution. Finally, $\|A\|$ is the Euclidian norm of a vector or matrix, *i.e.*, $\|A\| = [tr(A'A)]^{\frac{1}{2}}$.

2. Framework: endogeneity parameters and their identification

We consider a standard linear structural equation of the form:

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

where y is a $T \times 1$ vector of observations on a dependent variable, Y is a $T \times G$ matrix of observations on (possibly) endogenous explanatory variables $(G \geq 1)$, X_1 is a $T \times k_1$ full-column-rank matrix of strictly exogenous variables, $u = [u_1, \ldots, u_T]'$ is a vector of structural disturbances, β and γ are $G \times 1$ and $k_1 \times 1$ unknown coefficient vectors. Further, Y satisfies the model:

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

where X_2 is a $T \times k_2$ matrix of observations on exogenous variables (instruments), $X = [X_1, X_2]$ has full-column rank $k = k_1 + k_2$, Π_1 and Π_2 are $k_1 \times G$ and $k_2 \times G$ coefficient matrices, $\Pi = [\Pi_1, \Pi_2]$, and $V = [V_1, \ldots, V_T]'$ is a $T \times G$ matrix of reduced-form disturbances. Equation (2.1) is the "structural equation" of interest, while (2.2) represents the "reduced form" for Y. On substituting (2.2) into (2.1) and reexpressing y in terms of exogenous variables, we get the reduced form for y:

$$y = X_1\pi_1 + X_2\pi_2 + v \tag{2.3}$$

where $\pi_1 = \gamma + \Pi_1\beta$, $\pi_2 = \Pi_2\beta$, and $v = V\beta + u = [v_1, \ldots, v_T]'$.

When the errors u and V have finite zero means (although this assumption could easily be replaced by another "centering assumption", such as zero medians), the usual necessary and sufficient condition for identification of β in (2.1)-(2.2) is:

$$\operatorname{rank}(\Pi_2) = G\,. \tag{2.4}$$

If $\Pi_2 = 0$, the instruments X_2 are irrelevant, and β is completely unidentified. If $1 \leq \operatorname{rank}(\Pi_2) < G$, β is not identifiable, but some linear combinations of the elements of β are identifiable [see Dufour and Hsiao (2008)]. If Π_2 is close not to have full rank [*e.g.*, if some eigenvalues of $\Pi_2'\Pi_2$ are close to zero], some linear combinations of β are ill-determined by the data, a situation often called "weak identification" in this type of setup [see Dufour (2003)].

2.1. Identification of endogeneity parameters

We now wish to represent the fact that u and V are not independent and may be correlated, taking into account the fact that structural parameters (such as β and γ) may not be identifiable. In this context, it is important to note that the "structural error" u_t is not uniquely determined by the data when identification conditions for β and γ do not hold. For that, it will be useful to consider two alternative setups for the disturbance distribution: (A) in the first one, the disturbance vectors $(u_t, V_t')'$ have common finite second moments (*structural homoskedasticity*); (A) in the second one, we allow for a large amount of heterogeneity in the distributions of reduced-form errors (*reduced-form heterogeneity*). The second setup is clearly more appropriate for practical work, and we wish to go as far as possible in that direction. But it will be illuminating to consider first the more restrictive assumption.

In setup A, we suppose that:

the vectors $U_t = (u_t, V_t')'$, $t = 1, \ldots, T$, all have mean zero and finite covariance matrix (2.5)

$$\Sigma_U = \mathsf{E}\big[U_t U_t'\big] = \begin{bmatrix} \sigma_u^2 & \sigma_{Vu}' \\ \sigma_{Vu} & \Sigma_V \end{bmatrix} \tag{2.6}$$

where $\Sigma_V = \mathsf{E}\big[V_t V_t'\big]$ is nonsingular. In this case, the reduced-form errors $W_t = (v_t, V_t')'$, $t = 1, \ldots, T$, also have mean zero and covariance matrix

$$\Sigma_W = \mathsf{E}\big[W_t W_t'\big] = \begin{bmatrix} \sigma_v^2 & \sigma_{Vv}' \\ \sigma_{Vv} & \Sigma_V \end{bmatrix} \tag{2.7}$$

where

$$\sigma_{Vv} = \mathsf{E}[V_t v_t] = \mathsf{E}[V_t(V_t'\beta + u_t] = \Sigma_V\beta + \sigma_{Vu}\,,\ \sigma_v^2 = \sigma_u^2 + \beta'\Sigma_V\beta + 2\beta'\sigma_{Vu}\,. \tag{2.8}$$

The covariance vector σ_{Vu} indicates which variables in Y are "correlated" with u_t, so it provides a natural measure of the "endogeneity" of these variables. Note, however, that σ_{Vu} is not identifiable when β is not (because, in this case, the "structural error" u_t is not uniquely determined by the data).

In this context, it will be illuminating to look at the following two regressions: (1) the linear regression of u_t on V_t ,

$$u_t = V_t'a + e_t,\ t = 1, \ldots, T, \tag{2.9}$$

where $a = \Sigma_V^{-1}\sigma_{Vu}$ and $\mathsf{E}[V_t e_t] = 0$ for all t; and (2) the linear regression of v_t on V_t ,

$$v_t = V_t'b + \eta_t,\ t = 1, \ldots, T, \tag{2.10}$$

where $b = \Sigma_V^{-1}\sigma_{Vv}$ and $\mathsf{E}[V_t\eta_t] = 0$ for all t. It is easy to see that

$$\sigma_{Vu} = \Sigma_V a, \quad \sigma_u^2 = \sigma_e^2 + a'\Sigma_V a = \sigma_e^2 + \sigma_{Vu}'\Sigma_V^{-1}\sigma_{Vu} \tag{2.11}$$

where $\mathsf{E}[e_t^2] = \sigma_e^2$ for all t. This entails that: $a = 0$ if and only if $\sigma_{Vu} = 0$, so the exogeneity of Y can be assessed by testing whether $a = 0$. There is however no simple match between the components of a and σ_{Vu} (unless Σ_V is a diagonal matrix). For example, if $a = (a_1', a_2')'$ and $\sigma_{Vu} = (\sigma_{Vu1}', \sigma_{Vu2}')'$ where a_1 and σ_{Vu1} have dimension $G_1 < G$, $a_1 = 0$ is not equivalent to $\sigma_{Vu1} = 0$. In such a setup, we call a the "regression endogeneity parameter" and σ_{Vu} the "covariance endogeneity parameter".

As long as the identification condition (2.4) holds, both σ_{Vu} and a are identifiable. This is not the case, however, when (2.4) does not hold. By contrast, the regression coefficient b is always identifiable, because it is uniquely determined by the second moments of reduced-form errors. It is then useful to observe the following identity:

$$b = \Sigma_V^{-1}\sigma_{Vv} = \Sigma_V^{-1}(\Sigma_V\beta + \sigma_{Vu}) = \beta + a\,. \tag{2.12}$$

In other words, the sum $\beta + a$ is equal to the regression coefficient of v_t on V_t. Even though β and a may not be identifiable, the sum $\beta + a$ is identifiable. Further, for any fixed $G \times 1$ vector w, $w'b$ is identifiable, and the identities

$$w'a = w'b - w'\beta\,,\ \sigma_{Vu} = \Sigma_V a \tag{2.13}$$

along with the invertibility of Σ_V entail the following equivalences:

$$\beta \text{ is identifiable} \quad \Leftrightarrow \quad a \text{ is identifiable} \Leftrightarrow \sigma_{Vu} \text{ is identifiable}\,; \tag{2.14}$$

$$w'\beta \text{ is identifiable} \quad \Leftrightarrow \quad w'a \text{ is identifiable} \Leftrightarrow w'\Sigma_V^{-1}\sigma_{Vu} \text{ is identifiable}\,. \tag{2.15}$$

In particular, it is interesting to observe a simple identification correspondence between the components of β and a:

$$a_i \text{ is identifiable} \Leftrightarrow \beta_i \text{ is identifiable} \tag{2.16}$$

for $i = 1, \ldots, G$. In other words, the identification conditions for β and a are identical. In contrast, the equivalences [$w'\sigma_{Vu}$ is identifiable $\Leftrightarrow$ $w'\beta$ is identifiable] and [σ_{Vui} is identifiable $\Leftrightarrow$ β_i is identifiable] *do not hold* in general. Below, we will see that inference on b can be obtained through standard linear regression methods, so that this can be combined with identification-robust inference on β in order to obtain identification-robust inference on endogeneity parameters.

The setup (2.5) - (2.6) requires that the reduced-form disturbances V_t, $t = 1, \ldots, T$, have identical second moments. In many practical situations, this may not be appropriate, especially in a limited-information analysis that focuses on the structural equation of interest (2.1), rather than the marginal distribution of the explanatory variables Y. To allow for more heterogeneity among the observations in Y, we can however directly assume that:

$$u = Va + e\,, \tag{2.17}$$

$$e \text{ has mean zero and is uncorrelated with } V \text{ and } X\,, \tag{2.18}$$

for some fixed vector a in $\mathbb{R}^G$ (setup B). Later on, however, we shall consider setups where this assumption is modified, for example in order to allow for cases where e does not have finite first or second moments. There is no further restriction on the distribution of V, such as identical covariance matrices $[\mathsf{E}(V_t V_t') = \Sigma_V$ for all $t]$. An attractive feature of this assumption is that it remains "agnostic" concerning the distribution of V. In particular, the rows of V need not be identically distributed (for example, arbitrary heteroskedasticity is allowed) or independent. In fact, the assumption of finite second moments for e, V and X – entailed by the orthogonality condition (2.18) – can be relaxed if it is replaced by a similar assumption that does not require the existence of moments [such as independence between e and $(V,\, X)$]. Clearly, (2.5) - (2.6) is a special case of (2.17). We will see below that finite-sample inference on model parameters remains possible under the assumptions (2.17) - (2.18).

In view of (2.17), equation (2.1) can be viewed as a regression model with missing regressors. On substituting (2.17) into (2.1), we get:

$$y = Y\beta + X_1\gamma + Va + e \tag{2.19}$$

where e is uncorrelated with all the regressors. Because of the latter property, we call (2.19) the *orthogonalized structural equation* associated with (2.2), and e the *orthogonalized structural disturbance* vector. This equation contains the parameters of the original structural equation as regression coefficients, plus the regression endogeneity parameter a. We see that a represents the effects of the latent variable V. Even though (2.19) is a regression equation [in the sense that all regressors (Y, X_1, V) are orthogonal to the disturbance vector e], it is quite distinct from the reduced-form equation (2.3) for y.

The identification of a can be studied through the orthogonalized structural equation. Using the reduced form (2.2), we see that

$$\begin{aligned} y &= Y\beta + X_1\gamma + (Y - X_1\Pi_1 - X_2\Pi_2)a + e \\ &= Y\theta + X_1\pi_1^* + X_2\pi_2^* + e \end{aligned} \tag{2.20}$$

where $\theta = \beta + a$, $\pi_1^* = \gamma - \Pi_1 a$, $\pi_2^* = -\Pi_2 a$, and e is uncorrelated with all the regressors $(Y, X_1$ and $X_2)$. Equation (2.20) is thus a regression equation obtained by adding X_2 to the original structural equation or, equivalently, by adding Y to the reduced form (2.3) for y. We will call equation (2.20) the *extended reduced form* associated with (2.2). As soon as the matrix $Z = [Y, X_1, X_2]$ has full-column rank with probability one, the parameters of equation (2.20) are identifiable. This is the case in particular for $\theta = \beta + a$ (with probability one) when Z has full-column rank with probability one. This rank condition holds in particular when the matrix V has full column rank with probability one (conditional on X), *e.g.* if its distribution is absolutely continuous. This entails again that a is identifiable if and only β is identifiable, and similarly between $w'a$ and $w'\beta$ for any $w \in \mathbb{R}^G$.

This establishes the following identification lemma for a.

Lemma 2.1 IDENTIFICATION OF REGRESSION ENDOGENEITY PARAMETERS. *Under the assumptions* (2.2), (2.3) *and* (2.17), *suppose the matrix* $[Y, X_1, X_2]$ *has full column rank with probability one. Then* $a + \beta$ *is identifiable, and the following two equivalences hold:*

$$a \text{ is identifiable} \Leftrightarrow \beta \text{ is identifiable}\,; \tag{2.21}$$

$$\text{for any } w \in \mathbb{R}^G,\ w'a \text{ is identifiable} \Leftrightarrow w'\beta \text{ is identifiable}\,. \tag{2.22}$$

The decomposition assumption (2.17) can also be formulated in terms of the reduced-form disturbance v [as in (2.10)] rather than the structural disturbance u:

$$v = Vb + \eta \tag{2.23}$$

for some fixed vector b in $\mathbb{R}^G$, where each element of η has mean zero and is uncorrelated with V and X, again without any other assumption on the distribution of V. This means that the linear regressions $v_t = V_t'b + \eta_t,\ t, , \ldots, T$, can all be written in terms of the same coefficient vector b. The latter is uniquely determined (identifiable) as soon as the matrix V has full column rank (with probability one), so the identification of β is irrelevant. Even though conditions (2.17) and (2.23) look quite different (because the dependent variable is not the same), they are in fact equivalent in the context of the model we study here. This can be seen by rewriting the reduced form (2.3) as follows:

$$\begin{aligned} y &= X_1\pi_1 + X_2\pi_2 + v = X_1(\gamma + \Pi_1\beta) + X_2(\Pi_2\beta) + Vb + \eta \\ &= (X_1\Pi_1 + X_2\Pi_2)\beta + X_1\gamma + Vb + \eta \\ &= Y\beta + X_1\gamma + V(b - \beta) + \eta\,. \end{aligned} \tag{2.24}$$

Through matching the latter equation with the structural form (2.1), we get

$$u = V(b - \beta) + \eta \tag{2.25}$$

provided $[Y, X_1]$ has full-column rank. Since η and V are uncorrelated, this entails that (2.17) holds with $a = b - \beta$ and $e = \eta$. Conversely, under the assumption (2.17), we have from the reduced form (2.3):

$$v = V\beta + u = V(\beta + a) + e \tag{2.26}$$

which is equivalent to (2.23) with $b = \beta + a$ and $\eta = e$. We can thus state the following lemma.

Lemma 2.2 EQUIVALENCE BETWEEN STRUCTURAL AND REDUCED-FORM ERROR DECOMPOSITIONS. *Under the assumptions* (2.2) *and* (2.3), *suppose the matrix* $[Y, X_1, X_2]$ *has full column rank with probability one. Then the assumptions* (2.17) *and* (2.23) *are equivalent with* $b = \beta + a$ *and* $\eta = e$.

The identity $\eta = e$ entails that the residual vector from the regression of u on V is uniquely determined (identifiable) even if u itself may not be. The orthogonalized structural equation (2.19) may thus be rewritten as

$$\begin{aligned} y &= Y\beta + X_1\gamma + V(b-\beta) + \eta \\ &= (X\Pi)\beta + X_1\gamma + Vb + \eta \end{aligned} \tag{2.27}$$

where b is a regression vector between two reduced-form disturbances $(v$ on $V)$ and η the corresponding error. This shows clearly that different regression endogeneity parameters $a = b - \beta$ are obtained by "sweeping" β over its identification set.

Under the general assumption (2.17), covariance endogeneity parameters may depend on t. Indeed, it is easy to see that

$$\mathsf{E}\left[V_t u_t\right] = \mathsf{E}\left[V_t V_t^{\prime}\right] a \equiv \sigma_{Vut} \tag{2.28}$$

which may depend on t if $\mathsf{E}\left[V_t V_t^{\prime}\right]$ does. However, identification of the parameters σ_{Vut} remains determined by the identification of a, whenever the reduced-form covariance (which are parameters of reduced forms) are identifiable. Of course, inference on covariance endogeneity parameters requires additional assumptions. Indeed, we will see below that finite-sample inference methods can be derived for regression endogeneity parameters under the "weak assumptions" (2.17) - (2.18), while only asymptotically justified methods will be proposed for covariance endogeneity parameters. In particular for covariances we will focus on the case where σ_{Vut} does not depend on t $(\sigma_{Vut} = \sigma_{Vu}$ for all $t)$.

2.2. Statistical problems

In this paper, we consider the problem of testing hypotheses and building confidence sets for regression endogeneity parameters (a) and covariance endogeneity parameters (σ_{Vu}), allowing for the possibility of identification failure (or weak identification). We develop inference procedures for the full vectors a and σ_{Vu}, as well as linear transformations of these parameters $w'a$ and $w'\sigma_{Vu}$. In view of the identification difficulties present here, we emphasize methods for which a finite-sample distributional theory is possible [see Dufour (1997, 2003)], at least partially.

In line with the above discussion on identification of endogeneity parameters, we observe that inference on a can be tackled more easily than inference on σ_{Vu}, so we study this problem first. The problem of testing hypotheses of the form

$$H_a(a_0) : a = a_0 \tag{2.29}$$

can be viewed as an extension of the classical Anderson and Rubin (1949, AR) problem on testing $H_\beta(\beta_0) : \beta = \beta_0$. There is, however, an additional complication: the variable V is not observable.

For this reason, substantial adjustments are required. To achieve our purpose, we propose a strategy that builds on two-stage confidence procedures [Dufour (1990)], projection methods [Dufour (1990, 1987), Abdelkhalek and Dufour (1998), Dufour and Jasiak (2001), Dufour and Taamouti (2005)], and Monte Carlo tests [Dufour (2006)].

Specifically, in order to build a confidence set with level $1-\alpha$ for a, choose α_1 and α_2 such that $0 < \alpha = \alpha_1 + \alpha_2 < 1$, $0 < \alpha_1 < 1$ and $0 < \alpha_2 < 1$. We can then proceed as follows:

1. we build an identification-robust confidence set with level $1-\alpha_1$ for β; various procedures are already available for that purpose; in view of the existence of a finite-sample distributional theory (as well as computational simplicity), we focus on the Anderson and Rubin (1949, AR) approach; but alternative procedures could be exploited for that purpose;[1]

2. we build an identification-robust confidence set for the sum $\theta = \beta + a$, which happens to be an identifiable parameter; we show this can be done easily though simple regression methods;

3. the confidence sets for β and θ are combined to obtain a simultaneous confidence set for the stacked parameter vector $\varphi = (\beta', \theta')'$; by the Boole-Bonferroni inequality, this yields a confidence set for φ with level $1-\alpha$ (at least), as in Dufour (1990);

4. confidence sets for $a = \theta - \beta$ and any linear transformation $w'a$ may then be derived by projection; these confidence sets have level $1-\alpha$;

5. confidence sets for σ_{Vu} and $w'\sigma_{Vu}$ can finally be built on exploiting the relationship $\sigma_{Vu} = \Sigma_V a$.

For inference on a, we develop a finite-sample approach which remains valid irrespective of assumptions on the distribution of V. In addition, we observe that the test statistics used for inference on β [the AR-type statistic] and θ enjoy invariance properties which allow the application of Monte Carlo test methods: as long as the distribution of the errors u is specified up to an unknown scale parameter, exact tests can be performed on β and θ through a small number of Monte Carlo simulations [see Dufour (2006)]. For inference on both regression and covariance endogeneity parameters (a and σ_{Vu}), we also provide a large-sample distributional theory based on standard asymptotic assumptions which relax various restrictions used in the finite-sample theory. All proposed methods do not make identification assumptions on β, either in finite samples or asymptotically.

3. Finite-sample inference for regression endogeneity parameters

In this section, we study the problem of building identification-robust tests and confidence sets for the regression endogeneity parameter a from a finite-sample viewpoint. Along with the basic model assumptions (2.2) - (2.3), we suppose that (2.17) and the following assumption on the error distribution hold.

[1] Such procedures include, for example, the methods proposed by Kleibergen (2002) or Moreira (2003). No finite-sample distributional theory is, however, available for these methods. Further, these are not robust to missing instruments; see Dufour (2003) and Dufour and Taamouti (2007).

Assumption 3.1 CONDITIONAL SCALE MODEL FOR THE STRUCTURAL ERROR DISTRIBUTION. *The conditional distribution of u given $X = [X_1, X_2]$ is completely specified up to an unknown scalar factor,* i.e.

$$u\,|\,X \sim \sigma(X)\,\upsilon \tag{3.1}$$

where $\sigma(X)$ is a fixed function of X, and υ has a completely specified distribution $($which may depend on $X)$.

Assumption 3.2 CONDITIONAL SCALE MODEL FOR STRUCTURAL ERROR DISTRIBUTION. *The conditional distribution of $e = u - Va$ given $X = [X_1, X_2]$ is completely specified up to unknown scalar factor,* i.e.

$$e\,|\,X \sim \sigma_1(X)\,\varepsilon \tag{3.2}$$

where $\sigma_1(X)$ is a fixed function of X, and υ has a completely specified distribution $($which may depend on $X)$.

Assumption **3.1** means that the distribution of u given X only depends on X and a (typically unknown) scale factor $\sigma(X)$. Of course, this holds whenever u is independent of X with a distribution of the form $u \sim \sigma\,\upsilon$, where υ has a specified distribution and σ is an unknown positive constant. In this context, the standard Gaussian assumption is obtained by taking

$$\upsilon \sim \mathrm{N}[0, I_T]\,. \tag{3.3}$$

But non-Gaussian distributions are covered, including heavy-tailed distributions which may lack moments (such as the Cauchy distribution). Similarly, Assumption **3.2** means that the distribution of e given X only depends on X and a (typically unknown) scale factor $\sigma_1(X)$, so again a standard Gaussian model is obtained by assuming

$$\varepsilon \sim \mathrm{N}[0, I_T]\,. \tag{3.4}$$

In general, assumptions **3.1** and **3.2** do not entail each other. However, it is easy to see that both hold when the vectors $[u_t, V_t^{'}]^{'}$, $t, , \ldots, T$, are i.i.d. (conditional on X) with finite second moments and the decomposition assumption (2.17) - (2.18) holds. This will be the case *a fortiori* if the vectors $[u_t, V_t^{'}]^{'}$, $t, , \ldots, T$, are i.i.d. multinormal (conditional on X).

We will study in turn the following problems:

1. test and build confidence sets for β;

2. test and build confidence sets for $\theta = \beta + a$;

3. test and build confidence sets for a;

4. test and build confidence sets for scalar linear transformations $w'a$.

3.1. AR-type tests for β with possibly non-Gaussian errors

Since this will be a basic building block for inference on endogeneity parameters, we consider first the problem of testing the hypothesis

$$H_\beta(\beta_0) : \beta = \beta_0 \tag{3.5}$$

where β_0 is any given possible value of β. Several procedures have been proposed for that purpose. However, since we wish to use an identification-robust procedure for which a finite-sample theory can easily be easily obtained and does not require assumptions on the distribution of Y, we focus on the Anderson and Rubin (1949, AR) procedure. So we consider the transformed equation:

$$y - Y\beta_0 = X_1\pi_1^0 + X_2\pi_2^0 + v^0 \tag{3.6}$$

where $\pi_1^0 = \gamma + \Pi_1(\beta - \beta_0)$, $\pi_2^0 = \Pi_2(\beta - \beta_0)$ and $v^0 = u + V(\beta - \beta_0)$. Since $\pi_2^0 = 0$ under $H_\beta(\beta_0)$, it is natural to consider the corresponding F-statistic in order to test $H_\beta(\beta_0)$:

$$AR(\beta_0) = \frac{(y - Y\beta_0)'(M_1 - M)(y - Y\beta_0)/k_2}{(y - Y\beta_0)'M(y - Y\beta_0)/(T - k)} \tag{3.7}$$

where $M_1 \equiv M(X_1)$ and $M \equiv M(X)$; for any full-column rank matrix A, we set $P(A) = A(A'A)^{-1}A'$ and $M(A) = I - P(A)$. Under the usual assumption where $u \sim \mathrm{N}[0, \sigma^2 I_T]$ independently of X, the conditional distribution of $AR(\beta_0)$ under $H_\beta(\beta_0)$ is $F(k_2, T-k)$. In the following proposition, we characterize the null distribution of $AR(\beta_0)$ under the more general Assumption **3.1**.

Proposition 3.3 NULL DISTRIBUTION OF AR STATISTICS UNDER SCALE STRUCTURAL ERROR MODEL. *Suppose the assumptions* (2.1), (2.2) *and* **3.1** *hold. If $\beta = \beta_0$, we have:*

$$AR(\beta_0) = \frac{\upsilon'(M_1 - M)\upsilon/k_2}{\upsilon'M\upsilon/(T - k)} \tag{3.8}$$

and the conditional distribution of $AR(\beta_0)$ given X only depends on X and the distribution of υ.

The latter proposition means that the conditional null distribution of $AR(\beta_0)$, given X, only depends on the distribution of υ. Note the distribution of V plays no role here, so no decomposition assumption [such as (2.17) - (2.18) or (2.23)] is needed. If the distribution of $\upsilon\,|X$ can be simulated, one can get exact tests based on $AR(\beta_0)$ through the Monte Carlo test method [see Dufour (2006)], even if this conditional distribution is non-Gaussian. Furthermore, the exact test obtained in this way is robust to weak instruments as well as instrument exclusion even if the distribution of $u\,|\,X$ does not have moments (the Cauchy distribution, for example). This may be useful for example in financial models with fat-tailed error distributions, such as the Student t distribution.

When the normality assumption (3.3) holds and X is exogenous, we have $AR(\beta_0) \sim F(k_2, T-k)$, so that $H_\beta(\beta_0)$ can be assessed by using a critical region of the form $\{AR(\beta_0) > f(\alpha)\}$, where $f(\alpha) = F_\alpha(k_2, T-k)$ is the $1-\alpha$ quantile of the F-distribution with $(k_2, T-k)$ degrees of freedom.

A confidence set with level $1-\alpha$ for β is then given by

$$\mathscr{C}_{\beta}(\alpha) = \{\beta_0 : AR(\beta_0) \leq F_{\alpha}(k_2, T-k)\} = \{\beta : Q(\beta) \leq 0\} \tag{3.9}$$

where $Q(\beta) = \beta' A\beta + b'\beta + c$, $A = Y'HY$, $b = -2Y'Hy$, $c = y'Hy$, $H = M_1 - [1 + f(\alpha)(\frac{k_2}{T-k})]M$, and $f(\alpha) = F_{\alpha}(k_2, T-k)$; see Dufour and Taamouti (2005).

Suppose now that the conditional distribution of υ (given X) is continuous, so that the conditional distribution of $AR(\beta_0)$ under the null hypothesis $H_{\beta}(\beta_0)$ is also continuous. We can then proceed as follows to obtain an exact Monte Carlo test of $H_{\beta}(\beta_0)$ with level α $(0 < \alpha < 1)$:

1. choose α^* and N so that
$$\alpha = \frac{I[\alpha^* N] + 1}{N+1}\,; \tag{3.10}$$
2. for given β_0, compute the test statistic $AR^{(0)}(\beta_0)$ based on the observed data;
3. generate N *i.i.d.* error vectors $\upsilon^{(j)} = [\upsilon_1^{(j)}, \ldots, \upsilon_T^{(j)}]'$, $j = 1, \ldots, N$, according to the specified distribution of $\upsilon|X$, and compute the corresponding statistic $AR^{(j)}$, $j = 1, \ldots, N$, following (3.8); note the distribution of $AR(\beta_0)$ does not depend on the specific value β_0 tested, so there is no need to make it depend on β_0;
4. compute the empirical distribution function based on $AR^{(j)}$, $j = 1, \ldots, N$,
$$\hat{F}_N(x) = \frac{\sum_{j=1}^{N} \mathbb{1}[AR^{(j)} \leq x]}{N+1}\,, \tag{3.11}$$
or, equivalently, the simulated p-value function
$$\hat{p}_N[x] = \frac{1 + \sum_{j=1}^{N} \mathbb{1}[AR^{(j)} \geq x]}{N+1} \tag{3.12}$$
where $\mathbb{1}[C] = 1$ if condition C holds, and $\mathbb{1}[C] = 0$ otherwise;
5. reject the null hypothesis $H_{\beta}(\beta_0)$ at level α when $AR^{(0)}(\beta_0) \geq \hat{F}_N^{-1}(1-\alpha^*)$, where $\hat{F}_N^{-1}(q) = \inf\{x : \hat{F}_N(x) \geq q\}$ is the generalized inverse of $\hat{F}_N(\cdot)$, or (equivalently) when $\hat{p}_N[AR^{(0)}(\beta_0)] \leq \alpha$.

Under the null hypothesis $H_{\beta}(\beta_0)$,

$$\mathbb{P}\big[AR^{(0)}(\beta_0) \geq \hat{F}_N^{-1}(1-\alpha^*)\big] = \mathbb{P}\big[\hat{p}_N[AR^{(0)}(\beta_0)] \leq \alpha\big] = \alpha \tag{3.13}$$

so that we have a test with level α. If the distribution of the test statistic is not continuous, the MC test procedure can easily be adapted by using "tie-breaking" method described in Dufour (2006).[2]

[2]It is also useful to note that, without correction for continuity, the algorithm proposed for statistics with continuous distributions yields a conservative test, *i.e.* the probability of rejection under the null hypothesis is not larger the nominal level (α_1).

Correspondingly, a confidence set with level $1-\alpha$ for β is given by the set of all values β_0 which are not rejected by the above MC test. More precisely, the set

$$\mathscr{C}_\beta(\alpha) = \left\{\beta_0 : \hat{p}_N[AR^{(0)}(\beta_0)] > \alpha\right\} \tag{3.14}$$

is a confidence set with level $1-\alpha$ for β. On noting that the distribution of $AR(\beta_0)$ does not depend on β_0, we can use a single simulation for all values β_0: setting $\hat{f}_N(\alpha^*) = \hat{F}_N^{-1}(1-\alpha^*)$, the set

$$\mathscr{C}_\beta(\alpha; N) = \left\{\beta_0 : AR^{(0)} < \hat{f}_N(\alpha^*)\right\} \tag{3.15}$$

is equivalent to $\mathscr{C}_\beta(\alpha)$ – with probability one – and so has level $1-\alpha$. On replacing $>$ and $<$ by $\geq$ and $\leq$ in (3.14) - (3.15), it is also clear that the sets $\left\{\beta_0 : \hat{p}_N[AR^{(0)}(\beta_0)] \geq \alpha\right\}$ and

$$\bar{\mathscr{C}}_\beta(\alpha; N) = \{\beta_0 : AR^{(0)}(\beta_0) \leq \hat{f}_N(\alpha^*)\} \tag{3.16}$$

constitute confidence sets for β with level $1-\alpha$ (though possibly a little larger than $1-\alpha$). The quadric form given in (3.9) also remains valid with $f(\alpha) = \hat{f}_N(\alpha^*)$.

3.2. Inference on θ

Let us now consider the problem of testing the hypothesis

$$H_\theta(\theta_0) : \theta = \theta_0 \tag{3.17}$$

where θ_0 is a given vector of dimension G, and Assumption **3.2** holds. This can be done by considering the extended reduced form in (2.20):

$$y = Y\theta + X_1\pi_1^* + X_2\pi_2^* + e \tag{3.18}$$

where $\theta = \beta + a$, $\pi_1^* = \gamma - \Pi_1 a$, $\pi_2^* = -\Pi_2 a$, and e is independent of Y, X_1 and X_2. Thus the extended reduced form is a linear regression model. As soon as the matrix $[Y, X_1, X_2]$ has full-column rank, the parameters of equation (3.18) can be tested through standard F-tests.

We will now assume that $[Y, X_1, X_2]$ has full-column rank with probability one. This property holds as soon as $X = [X_1, X_2]$ has full column rank and Y has a continuous distribution (conditional on X). The F-statistic for testing $H_\theta(\theta_0)$ is

$$F_\theta(\theta_0) = \frac{(\hat{\theta}-\theta_0)'(Y'MY)(\hat{\theta}-\theta_0)/G}{y'M(Z)y/(T-G-k)} \tag{3.19}$$

where $\hat{\theta} = (Y'MY)^{-1}Y'My$ is the OLS estimate of θ in (3.18), $M = M(X)$, $X = [X_1, X_2]$, and $Z = [Y, X_1, X_2]$. Under the normality assumption (3.4), we have:

$$F_\theta(\theta_0) \sim F(G, T-k-G). \tag{3.20}$$

Under the more general assumption **3.2**, it is easy to see that

$$F_\theta(\theta_0) = \frac{\varepsilon' MY(Y'MY)^{-1}Y'M\varepsilon/G}{\varepsilon' M(Z)\varepsilon/(T-G-k)} \tag{3.21}$$

under $H_\theta(\theta_0)$. On observing that the conditional distribution of $F_\theta(\theta_0)$, given Y and X, does not involve any nuisance parameter, the critical value can be obtained by simulation. It is also important to note that this distribution does not depend on θ_0, so the same critical value can be applied irrespective of θ_0. The main difference with the Gaussian case is that the critical value may depend on Y and X. Irrespective of the case considered [(3.20) or (3.21)], we shall denote by $c(\alpha_2)$ the critical value used for $F_\theta(\theta_0)$.

From (3.19), a confidence set with level $1-\alpha$ for θ can be obtained by inverting $F_\theta(\theta_0)$:

$$\mathscr{C}_\theta(\alpha) = \{\theta_0 : F_\theta(\theta_0) \le \bar{f}(\alpha)\} = \{\theta_0 : \bar{Q}(\theta_0) \le 0\} \tag{3.22}$$

where

$$a\bar{Q}(\theta) = (\hat{\theta}-\theta)'(Y'MY)(\hat{\theta}-\theta) - \bar{c}_0 = \theta'\bar{A}\theta + \bar{b}'\theta + \bar{c}, \tag{3.23}$$

where $\bar{c}_0 = \bar{f}(\alpha)Gs^2$, $s^2 = y'M(Z)y/(T-G-k)$,

$$\bar{A} = Y'MY,\ \bar{b} = -2\bar{A}\hat{\theta} = -2Y'My,\ \bar{c} = \hat{\theta}'\bar{A}\hat{\theta} - \bar{c}_0 = \hat{\theta}'(Y'MY)\hat{\theta} - \bar{c}_0 = y'\tilde{H}y, \tag{3.24}$$

and $\bar{H} = P(MY) - \bar{f}(\alpha)[G/(T-G-k)]M_1$. Since the matrix $\bar{A}$ is positive definite (with probability one), the quadric set $\mathscr{C}_\theta(\alpha)$ is an ellipsoid (hence bounded); see Dufour and Taamouti (2005, 2007). This reflects the fact that θ is an identifiable parameter. As a result, the corresponding projection-based confidence sets for scalar transformations $w'\theta$ are also bounded intervals.

In view of the form of model (3.18) as a linear regression, we can test in the same way linear restrictions of the form

$$H_{w'\theta}(\gamma_0) : w'\theta = \gamma_0 \tag{3.25}$$

where w is a $G\times 1$ vector and γ_0 is known constant. We can then use the corresponding t statistic

$$t_{w'\theta}(\gamma_0) = \frac{w'\hat{\theta}-\gamma_0}{s[w'(Z'Z)^{-1}w]^{1/2}} \tag{3.26}$$

and reject $H_{w'\theta}(\gamma_0)$ when

$$|t_{w'\theta}(\gamma_0)| > c_w(\alpha) \tag{3.27}$$

where $c_w(\alpha)$ is the critical value for a test with level α. In the Gaussian case, $t_{w'\theta}(\gamma_0)$ follows a Student distribution with $T-G-k$ degrees of freedom, so we can take $c_w(\alpha) = t(\alpha_2; T-G-k)$. When ε follows a non-Gaussian distribution, we have

$$t_{w'\theta}(\gamma_0) = \frac{(T-G-k)^{1/2}(Y'MY)^{-1}Y'M\varepsilon}{\left(\varepsilon' M(Z)\varepsilon\right)^{1/2}[w'(Z'Z)^{-1}w]^{1/2}} \tag{3.28}$$

under $H_w(\gamma_0)$, so that the distribution of $t(\gamma_0)$ can be simulated like $F_\theta(\theta_0)$ in (3.21).

3.3. Joint inference on β and regression endogeneity parameters

We can now derive confidence sets for the vectors $(\beta', a')'$ and $(\beta', \theta')'$. By the Boole-Bonferroni inequality, we have:

$$\mathbb{P}[\beta \in \mathscr{C}_\beta(\alpha_1) \text{ and } \theta \in \mathscr{C}_\theta(\alpha_2)] \geq 1 - \mathbb{P}[\beta \notin \mathscr{C}_\beta(\alpha_1)] - \mathbb{P}[\theta \notin \mathscr{C}_\theta(\alpha_2)] \geq 1 - \alpha_1 - \alpha_2 \quad (3.29)$$

The set

$$\begin{aligned} \mathscr{C}_{(\beta,\,\theta)}(\alpha_1, \alpha_2) &= \{(\theta_0', \beta_0')' : \beta_0 \in \mathscr{C}_\beta(\alpha_1)\,,\ \theta_0 \in \mathscr{C}_\theta(\alpha_2)\} \\ &= \{(\theta_0', \beta_0')' : Q(\beta_0) \leq 0\,,\ \bar{Q}(\theta_0) \leq 0\} \end{aligned} \quad (3.30)$$

is thus a confidence set with level $1-\alpha$ where $\alpha = \alpha_1 + \alpha_2$.

In view of the identity $\theta = \beta + a$, we can write $\bar{Q}(\theta)$ in (3.23) as a function of β and a:

$$\bar{Q}(\theta) = \bar{Q}(\beta + a) = a'\bar{A}a + (\bar{b} + 2\bar{A}\beta)'a + [\bar{c} + \bar{b}'\beta + \beta'\bar{A}\beta]\,.$$

Then the set

$$\bar{\mathscr{C}}_{(\beta,\,a)}(\alpha) = \{(\beta_0', a_0')' : Q(\beta_0) \leq 0 \text{ and } \bar{Q}(\beta_0 + a_0) \leq 0\} \quad (3.31)$$

is in turn a joint confidence set with level $1-\alpha$ for β and a. Thus, finite-sample inference on the structural (possibly unidentifiable) parameter a is possible. Of course, if a is not identified, a valid confidence set will cover the set of all possible values (or be unbounded) with probability $1-\alpha$ [see Dufour (1997)].

3.4. Confidence sets for regression endogeneity parameters

We can now build "marginal" confidence sets for the endogeneity coefficient vector a. In view of the possibility of identification failure, this is most easily done by projection techniques. Let $g(\beta, a)$ be any function of β and a. Since the event $(\beta, a) \in \bar{\mathscr{C}}_{(\beta,\,a)}(\alpha)$ entails $g(\beta, a) \in g[\bar{\mathscr{C}}_{(\beta,\,a)}(\alpha)]$, where $g[\bar{\mathscr{C}}_{(\beta,\,a)}(\alpha)] = \{g(\beta, a) : (\beta, a) \in \bar{\mathscr{C}}_{(\beta,\,a)}(\alpha)\}$, we have:

$$\mathbb{P}\left[g(\beta, a) \in g[\bar{\mathscr{C}}_{(\beta,\,a)}(\alpha)\right] \geq \mathbb{P}[(\beta, a) \in \bar{\mathscr{C}}_{(\beta,\,a)}(\alpha)] \geq 1 - \alpha\,. \quad (3.32)$$

On taking $g(\beta, a) = a,$ we see that

$$\begin{aligned} \mathscr{C}_a(\alpha) &= \{a \in \mathbb{R}^G : (\beta, a) \in \bar{\mathscr{C}}_{(\beta,\,a)}(\alpha) \text{ for some } \beta\} \\ &= \{a \in \mathbb{R}^G : \bar{Q}(\beta + a) \leq 0 \text{ and } Q(\beta) \leq 0 \text{ for some } \beta\} \end{aligned} \quad (3.33)$$

is a confidence set with level $1-\alpha$ for a.

When $G = 1,$ the matrices A, $\bar{A}$, b, $\bar{b}$, c and $\bar{c}$ in (3.23) reduce to scalars, and the different confidence sets take the following simple forms:

$$\mathscr{C}_{\beta}(\alpha_1) = \left\{\beta : A\beta^2 + b\beta + c \le 0\right\}, \quad \mathscr{C}_{\theta}(\alpha_2) = \{\theta : \bar{A}\theta^2 + \bar{b}\theta + \bar{c} \le 0\}, \tag{3.34}$$

$$\mathscr{C}_{a}(\alpha) = \{a : A\beta^2 + b\beta + c \le 0, \bar{A}a^2 + (\bar{b} + 2\bar{A}\beta)a + [\bar{c} + \bar{b}\beta + \bar{A}\beta^2] \le 0\}. \tag{3.35}$$

Closed-form for the sets $\mathscr{C}_{\beta}(\alpha_1)$ and $\mathscr{C}_{\theta}(\alpha_2)$ are easily derived by finding the roots of the second-order polynomial equations $A\beta^2 + b\beta + c = 0$ and $\bar{A}\theta^2 + \bar{b}\theta + \bar{c} = 0$ [as in Dufour and Jasiak (2001)], while the set $\mathscr{C}_{a}(\alpha)$ can be obtained by finding the roots of the equation

$$\bar{A}a^2 + \bar{b}(\beta)a + \bar{c}(\beta) = 0 \text{ where } \bar{b}(\beta) = \bar{b} + 2\bar{A}\beta \text{ and } \bar{c}(\beta) = \bar{c} + \bar{b}\beta + \bar{A}\beta^2 \tag{3.36}$$

for each $\beta \in \mathscr{C}_{\beta}(\alpha_1)$.

We shall now focus on building confidence sets for scalar linear transformations $g(a) = w'a = w'\theta - w'\beta$, where w is a $G \times 1$ vector. Conceptually, the simplest approach consists in applying the projection method from $\mathscr{C}_{a}(\alpha)$, which yields the confidence set:

$$\begin{aligned} \mathscr{C}_{w'a}(\alpha) &= g_w[\mathscr{C}_{a}(\alpha)] = \{d : d = w'a \text{ for some } a \in \mathscr{C}_{a}(\alpha)\} \\ &= \{d : d = w'a\,,\ \bar{Q}(\beta + a) \le 0 \text{ and } Q(\beta) \le 0 \text{ for some } \beta\}. \end{aligned}$$

But it will more efficient to exploit the linear structure of model (3.18), which allows one to build a confidence interval for $w'\theta$.

Following Dufour and Taamouti (2005, 2007), confidence sets for $g_w(\beta) = w'\beta$ and $g_w(\theta) = g_w = w'\theta$ can be derived from $\mathscr{C}_{\beta}(\alpha_1)$ and $\mathscr{C}_{\theta}(\alpha_2)$ as follows:

$$\begin{aligned} \mathscr{C}_{w'\beta}(\alpha_1) &\equiv g_w[\mathscr{C}_{\beta}(\alpha_1)] = \{x_1 : x_1 = w'\beta \text{ where } Q(\beta) \le 0\} \\ &= \{x_1 : x_1 = w'\beta \text{ where } \beta'A\beta + b'\beta + c \le 0\} \end{aligned} \tag{3.37}$$

where A, b and c are defined as in (3.9). For $w'\theta$, we can use a $t-$type confidence interval based on $t(\gamma_0)$:

$$\begin{aligned} \bar{\mathscr{C}}_{w'\theta}(\alpha_2) &\equiv \bar{g}_w[\mathscr{C}_{\theta}(\alpha_2)] = \{\gamma_0 : |t_{w'\theta}(\gamma_0)| < c_w(\alpha_2)\} \\ &= \{\gamma_0 : |w'\hat{\theta} - \gamma_0| < \bar{D}(\alpha_2)\} \end{aligned} \tag{3.38}$$

where $\bar{D}(\alpha_2) = c_w(\alpha_2)\,\hat{\sigma}(w'\hat{\theta})$, $\hat{\sigma}(w'\hat{\theta}) = s[w'(Z'Z)^{-1}w]^{1/2}$ and $c_w(\alpha_2)$ is the critical value for a test with level α_2 [determined as in (3.27)]. Setting

$$\mathscr{C}_{(w'\beta,\, w'\theta)}(\alpha_1, \alpha_2) = \{(x, y)' : x \in \mathscr{C}_{w'\beta}(\alpha_1) \text{ and } y \in \bar{\mathscr{C}}_{w'\theta}(\alpha_2)\}, \tag{3.39}$$

we see that $\mathscr{C}_{(w'\beta,\, w'\theta)}(\alpha_1, \alpha_2)$ is a confidence set for $(w'\beta,\, w'\theta)$ with level $1 - \alpha_1 - \alpha_2$:

$$\mathbb{P}[(w'\beta,\, w'\theta) \in \mathscr{C}_{(w'\beta,\, w'\theta)}(\alpha_1, \alpha_2)] = \mathbb{P}[w'\beta \in \mathscr{C}_{w'\beta}(\alpha_1) \text{ and } w'\theta \in \bar{\mathscr{C}}_{w'\theta}(\alpha_2)] \ge 1 - \alpha \tag{3.40}$$

where $\alpha = \alpha_1 + \alpha_2$. For any point $x \in \mathbb{R}$ and any subset $A \subseteq \mathbb{R}$, set $x - A = \{z \in \mathbb{R} : z = x - y \text{ and } y \in$

$A\}$. Since $w'a = w'\theta - w'\beta$, it is clear that

$$(w'\beta,\, w'\theta) \in \mathscr{C}_{(w'\beta,\, w'\theta)}(\alpha_1,\, \alpha_2) \;\; \begin{aligned} &\Leftrightarrow w'\theta - w'a \in \mathscr{C}_{w'\beta}(\alpha_1) \text{ and } w'\theta \in \bar{\mathscr{C}}_{\underline{w}'\theta}(\alpha_2) \\ &\Leftrightarrow w'a \in w'\theta - \mathscr{C}_{w'\beta}(\alpha_1) \text{ and } w'\theta \in \bar{\mathscr{C}}_{w'\theta}(\alpha_2) \end{aligned} \tag{3.41}$$

so that

$$\begin{aligned} \mathbb{P}[w'a \in w'\theta - \mathscr{C}_{w'\beta}(\alpha_1) \text{ and } w'\theta \in \bar{\mathscr{C}}_{w'\theta}(\alpha_2)] &= \mathbb{P}[w'\beta \in \mathscr{C}_{w'\beta}(\alpha_1) \text{ and } w'\theta \in \bar{\mathscr{C}}_{w'\theta}(\alpha_2)] \\ &\geq 1 - \alpha_1 - \alpha_2\,. \end{aligned} \tag{3.42}$$

Now, consider the set

$$\mathscr{C}_{w'a}(\alpha_1,\, \alpha_2) = \{z \in \mathbb{R} : z \in y - \mathscr{C}_{w'\beta}(\alpha_1) \text{ for some } y \in \bar{\mathscr{C}}_{w'\theta}(\alpha_2)\}. \tag{3.43}$$

Since the event $\{w'a \in w'\theta - \mathscr{C}_{w'\beta}(\alpha_1)$ and $w'\theta \in \bar{\mathscr{C}}_{w'\theta}(\alpha_2)\}$ entails $w'a \in \mathscr{C}_{w'a}(\alpha_1,\, \alpha_2)$, we have:

$$\mathbb{P}[w'a \in \mathscr{C}_{w'a}(\alpha_1,\, \alpha_2)] \geq \mathbb{P}[w'\beta \in \mathscr{C}_{w'\beta}(\alpha_1) \text{ and } w'\theta \in \bar{\mathscr{C}}_{w'\theta}(\alpha_2)] \geq 1 - \alpha_1 - \alpha_2 \tag{3.44}$$

and $\mathscr{C}_{w'a}(\alpha_{\underline{1}},\, \alpha_2)$ is a confidence set with level $1 - \alpha_1 - \alpha_2$ for $w'a$.

Since $\bar{\mathscr{C}}_{w'\theta}(\alpha_2)$ is a bounded interval, the shape of $\mathscr{C}_{w'a}(\alpha_1,\, \alpha_2)$ can be deduced easily by using the results given in Dufour and Taamouti (2005, 2007). We focus on the case where A is nonsingular [an event with probability one as soon as the distribution of $AR(\beta_0)$ is continuous] and $w \neq 0$. Then the set $\mathscr{C}_{w'\beta}(\alpha_1)$ may then rewritten as follows: if A is positive definite,

$$\begin{aligned} \mathscr{C}_{w'\beta}(\alpha_1) &= \left[w'\tilde{\beta} - D(\alpha_1),\, w'\tilde{\beta} + D(\alpha_1)\right], && \text{if } d \geq 0\,, \\ &= \emptyset\,, && \text{if } d < 0\,, \end{aligned} \tag{3.45}$$

where $\tilde{\beta} = -\frac{1}{2}A^{-1}b$, $d = \frac{1}{4}b'A^{-1}b - c$ and $D(\alpha_1) = \sqrt{d\,(w'A^{-1}w)}$; if A has exactly one negative eigenvalue,

$$\begin{aligned} \mathscr{C}_{w'\beta}(\alpha_1) &= \left]-\infty,\, w'\tilde{\beta} - D(\alpha_1)\right] \cup \left[w'\tilde{\beta} + D(\alpha_1)\,,+\infty\right[\,, && \text{if } w'A^{-1}w < 0 \text{ and } d < 0, \\ &= \mathbb{R}\backslash\{w'\tilde{\beta}\}\,, && \text{if } w'A^{-1}w = 0 \text{ and } d < 0 \\ &= \mathbb{R}\,, && \text{otherwise;} \end{aligned} \tag{3.46}$$

otherwise, $\mathscr{C}_{w'\beta}(\alpha_1) = \mathbb{R}$. $\mathscr{C}_{w'\beta}(\alpha_1) = \emptyset$ corresponds to a case where the model is not consistent with the data [so that $\mathscr{C}_{w'a}(\alpha_1,\, \alpha_2) = \emptyset$ as well], while $\mathscr{C}_{w'\beta}(\alpha_1) = \mathbb{R}$ and $\mathscr{C}_{w'\beta}(\alpha_1) = \mathbb{R}\backslash\{w'\tilde{\beta}\}$ indicate that $w'\beta$ is not identifiable and similarly for $w'a$ [so that $\mathscr{C}_{w'a}(\alpha_1,\, \alpha_2) = \mathbb{R}$]. This yields the following confidence sets for $w'a$: if A is positive definite,

$$\begin{aligned} \mathscr{C}_{w'a}(\alpha_1,\, \alpha_2) &= \left[w'(\hat{\theta} - \tilde{\beta}) - D_U(\alpha_1,\, \alpha_2)\,,\, w'(\hat{\theta} - \tilde{\beta}) + D_U(\alpha_1,\, \alpha_2)\right], && \text{if } d \geq 0\,, \\ &= \emptyset\,, && \text{if } d < 0\,, \end{aligned} \tag{3.47}$$

where $D_U(\alpha_1, \alpha_2) = D(\alpha_1) + \bar{D}(\alpha_2)$; if A has exactly one negative eigenvalue, $w'A^{-1}w < 0$ and $d < 0$,

$$\mathscr{C}_{w'a}(\alpha_1, \alpha_2) = \left]-\infty,\ w'(\hat{\theta} - \tilde{\beta}) - D_L(\alpha_1, \alpha_2)\right] \cup \left[w'(\hat{\theta} - \tilde{\beta}) + D_L(\alpha_1, \alpha_2),\, +\infty\right[\tag{3.48}$$

where $D_L(\alpha_1, \alpha_2) = D(\alpha_1) - \bar{D}(\alpha_2)$; otherwise, $\mathscr{C}_{w'a}(\alpha_1, \alpha_2) = \mathbb{R}$. These results may be extended to cases where A is singular, as done by Dufour and Taamouti (2007).

4. Asymptotic theory for inference on endogeneity parameters

In this section, we examine the validity of the procedures developed in Section 3 under weaker distributional assumptions, and we show how inference on covariance endogeneity parameters can be made. On noting that equations (3.6) and (3.18) constitute standard linear regression models (at least under the null hypothesis $\beta = \beta_0$), it is straightforward to find high-level regularity conditions under which the tests based on $AR(\beta_0)$ and $F_\theta(\theta_0)$ are asymptotically valid.

For $AR(\beta_0)$, we can consider the following general assumptions:

$$\frac{1}{T}X'u \xrightarrow{p} 0\,, \tag{4.1}$$

$$\frac{1}{T}u'u \xrightarrow{p} \sigma_u^2 > 0\,, \tag{4.2}$$

$$\frac{1}{T}X'X \xrightarrow{p} \Sigma_X \text{ with } \det(X) \neq 0\,, \tag{4.3}$$

$$\frac{1}{\sqrt{T}}X'u \xrightarrow{L} \psi_{Xu}\,,\ \psi_{Xu} \sim N\left[0,\ \sigma_u^2\Sigma_X\right], \tag{4.4}$$

where $X = [X_1,\, X_2]$. The above conditions are easy to interpret: (4.1) represents the asymptotic orthogonality between u and the instruments in X, (4.2) and (4.3) may be viewed as laws of large numbers for u and X, while (4.4) is a central limit property. Then, it is simple exercise to see that

$$AR(\beta_0) \xrightarrow{L} \frac{1}{k_2}\chi^2(k_2) \text{ when } \beta = \beta_0. \tag{4.5}$$

Similarly, for $F_\theta(\theta_0)$, suppose:

$$\frac{1}{T}Z'e \xrightarrow{p} 0\,, \tag{4.6}$$

$$\frac{1}{T}e'e \xrightarrow{p} \sigma_e^2\,, \tag{4.7}$$

$$\frac{1}{T}Z'Z \xrightarrow{p} \Sigma_Z \text{ with } \det(Z) \neq 0\,, \tag{4.8}$$

$$\frac{1}{\sqrt{T}}Z'e \xrightarrow{L} \psi_{Xe}\,,\ \psi_{Xe} \sim N\left[0,\ \sigma_e^2\Sigma_Z\right], \tag{4.9}$$

where $Z = [Y, X_1, X_2]$. Then

$$F_\theta(\theta_0) \xrightarrow{L} \frac{1}{G}\chi^2(G) \text{ when } \theta = \theta_0 \,. \tag{4.10}$$

The asymptotic distributions in (4.5) and (4.10) hold irrespective whether the instruments X are weak or strong. Further, as soon as (4.1) - (4.1) and (4.6) - (4.9) hold, the confidence procedures described in Section 3 remain "asymptotically valid" with $f(\alpha_1) = \chi^2(\alpha_1; k_2)/k_2$ and $\bar{f}(\alpha_2) = \chi^2(\alpha_2; G)/G$, where $\chi^2(\alpha_1; k_2)$ and $\chi^2(\alpha_2; G)$ are respectively the $1-\alpha_1$ and $1-\alpha_2$ quantiles of the corresponding χ^2 distributions. Of course, the Gaussian-based Fisher critical values may also be used (for they converge to the chi-square critical values as $T \to \infty$).

We can now consider inference for covariance endogeneity parameters σ_{Vu}. The problem of building confidence sets for σ_{Vu} is especially important for assessing partial exogeneity hypotheses. Since $a_j = 0,\ j = 1, \ldots, G$ does not entail $\sigma_{uVj} = 0$(where $1 \le j \le G$), confidence sets on the components of a cannot directly be used to assess for example, the exogeneity of each regressor $Y_j = 0,\ j = 1, \ldots, G$.

Confidence sets and tests for σ_{uV} can be deduced from those on a through the relationship $\sigma_{Vu} = \Sigma_V a$ given in (2.11). On replacing a by $\Sigma_V^{-1}\sigma_{Vu}$ in $\mathscr{C}_a(\alpha)$, we see that the set

$$\begin{aligned} \mathscr{C}_{\sigma_{Vu}}(\alpha; \Sigma_V) &= \{\sigma_{Vu} \in \mathbb{R}^G : \sigma_{Vu} = \Sigma_V a \text{ and } a \in \mathscr{C}_a(\alpha)\} \\ &= \{\sigma_{Vu} \in \mathbb{R}^G : \bar{Q}(\beta + \Sigma_V^{-1}\sigma_{Vu}) \le 0 \text{ and } Q(\beta) \le 0 \text{ for some } \beta\} \end{aligned} \tag{4.11}$$

is a confidence set with level a for σ_{Vu}. This set is simply the image of $\mathscr{C}_a(\alpha)$ by the linear transformation $g(x) = \Sigma_V x$. The difficulty here comes from the fact that Σ_V is unknown. Let

$$\hat{\Sigma}_V = \hat{V}'\hat{V}/(T-k) \tag{4.12}$$

where $\hat{V} = M(X)Y$ is the matrix of least-squares residuals from the first-step regression (2.2). Under standard regularity conditions, we have:

$$\hat{\Sigma}_V \xrightarrow{p} \Sigma_V \tag{4.13}$$

where $\det(\Sigma_V) > 0$. If β_0 and a_0 are the true values of β and a, the relations $\theta_0 = \beta_0 + a_0$ and $\sigma_{Vu0} = \Sigma_V a_0$ entail that $F_\theta(\theta_0)$ can be rewritten as follows:

$$F_\theta(\beta_0 + \Sigma_V^{-1}\sigma_{Vu0}) = \frac{(\hat{\theta} - \beta_0 - \Sigma_V^{-1}\sigma_{Vu0})'(Y'MY)(\hat{\theta} - \beta_0 - \Sigma_V^{-1}\sigma_{Vu0})/G}{y'M(Z)y/(T-G-k)} \,. \tag{4.14}$$

Replacing Σ_V by $\hat{\Sigma}_V$, we get the approximate pivotal function:

$$F_\theta(\beta_0 + \hat{\Sigma}_V^{-1}\sigma_{Vu0}) = \frac{(\hat{\theta} - \beta_0 - \hat{\Sigma}_V^{-1}\sigma_{Vu0})'(Y'MY)(\hat{\theta} - \beta_0 - \hat{\Sigma}_V^{-1}\sigma_{Vu0})/G}{y'M(Z)y/(T-G-k)} \tag{4.15}$$

If (4.13) holds, it is easy to see (by continuity) that $F_\theta(\beta_0 + \hat{\Sigma}_V^{-1}\sigma_{Vu0})$ and $F_\theta(\beta_0 + \Sigma_V^{-1}\sigma_{Vu0})$

are asymptotically equivalent with a nondegenerate distribution, when β_0 and σ_{Vu0} are the true parameter values. Consequently, the confidence set of type $\mathscr{C}_{\sigma_{Vu}}(\alpha)$ based on $F_\theta(\beta_0 + \hat{\Sigma}_V^{-1}\sigma_{Vu0})$ as opposed to $F_\theta(\beta_0 + \Sigma_V^{-1}\sigma_{Vu0})$ has level $1-\alpha$ asymptotically. This set is simply the image of $\mathscr{C}_a(\alpha)$ by the linear transformation $\hat{g}(x) = \hat{\Sigma}_V x$, *i.e.*

$$\begin{aligned} \mathscr{C}_{\sigma_{Vu}}(\alpha; \hat{\Sigma}_V) &= \{\sigma_{Vu} \in \mathbb{R}^G : \sigma_{Vu} = \hat{\Sigma}_V a \text{ and } a \in \mathscr{C}_a(\alpha)\} \\ &= \{\sigma_{Vu} \in \mathbb{R}^G : \bar{Q}(\beta + \hat{\Sigma}_V^{-1}\sigma_{Vu}) \leq 0 \text{ and } Q(\beta) \leq 0 \text{ for some } \beta\}. \end{aligned} \tag{4.16}$$

Finally, confidence sets for the components of σ_{Vu}, and more generally for linear combinations $w'\sigma_{Vu}$, can be derived from those on $w'a$ as described in Section 3.4. For Σ_V given, the relation $\sigma_{Vu} = \Sigma_V a$ entails that a confidence set for $w'\sigma_{Vu}$ (with level $1-\alpha$) can be obtained by computing a confidence set (at level $1-\alpha$) for $w_1'a$ with $w_1 = \Sigma_V w$. When Σ_V is estimated by $\hat{\Sigma}_V$, taking $w_1 = \hat{\Sigma}_V w$ yields a confidence set for σ_{Vu} with level $1-\alpha$ asymptotically.

5. Empirical applications

We will now apply the methods proposed above to two empirical examples: the relation between trade and growth [Dufour and Taamouti (2007), Irwin and Tervio (2002), Frankel and Romer (1999), Harrison (1996), Mankiw, Romer and Weil (1992)] and the well known problem of returns to schooling [Doko Tchatoka and Dufour (2009), Dufour and Taamouti (2007), Angrist and Krueger (1991), Angrist and Krueger (1995), Mankiw et al. (1992)].

5.1. Trade and growth

The trade and growth model studies the relationship between standards of living and openness. Frankel and Romer (1999) argued that trade share (ratio of imports or exports to GDP) which is the commonly used indicator of openness should be viewed as endogenous. The authors then suggest to estimate the income-trade relationship using an IV method. The equation studied is given by:

$$\ln(\text{Income}_i) = \beta_0 + \beta_1 \text{Trade}_i + \gamma_1 \ln(\text{Pop}_i) + \gamma_2 \ln(\text{Area}_i) + u_i,\ i = 1, \ldots, N \tag{5.1}$$

where Income_i is the income per capita in country $i,$ Trade_i is the trade share (measured as a ratio of imports and exports to GDP), Pop_i is the logarithm of population of country $i,$ and Area_i is the logarithm of country $i,$ area. The instrument suggested by Frankel and Romer (1999) is constructed on the basis of geographic characteristics. The first stage equation is given by:

$$\text{Trade}_i = b_0 + b_1 Z_i + c_1 \text{Pop}_i + c_2 \text{Area}_i + V_i,\ i = 1, \ldots, N \tag{5.2}$$

where Z_i is a constructed instrument. We use the sample of 150 countries and the data include for each country the trade share in 1985, the area and population (1985), per capita income (1985), and the fitted trade share (instrument). As showed in Dufour and Taamouti (2005), it is not clear how“weak ”the instruments are for this sample.[3]

We follow the methodology developed in this paper to build projection-based confidence sets for regression endogeneity “a” and covariance endogeneity “σ_{Vu}” . We have also reported IV-based confidence intervals for the identified parameter “$\theta = \beta + a$” .

The estimate of the regression endogeneity parameter “a” in the transformed equation

$$\ln(\text{Income}_i) = \beta_0 + \beta_1 \text{Trade}_i + \gamma_1 \ln(\text{Pop}_i) + \gamma_2 \ln(\text{Area}_i) + \hat{V}_i a + e_i \tag{5.3}$$

is around $\hat{a} = -1.817,$ while the estimate of Σ_V from the first-step regression is $\hat{\Sigma}_V = 0.209.$ Hence, the estimate of of the covariance endogeneity parameter σ_{Vu} is about $\hat{\sigma_{uV}} = \hat{\Sigma}_V \hat{a} = -0.3805$. Table 1 presents the confidence sets at levels 97.5% and 95% for β_1 and $\theta = \beta_1 + a,$ and at levels 95% and 90% for a and σ_{Vu}. The results show clearly that both $\mathscr{C}_a(\alpha)$ and $\mathscr{C}_{\sigma_{Vu}}(\alpha)$ are bounded in all cases. However, the confidence interval that result from projection are large compare with alternative IV-based confidence intervals. This suggests that the instruments may not be very strong in this model. Moreover, we observe that both $\mathscr{C}_a(\alpha)$ and $\mathscr{C}_{\sigma_{Vu}}(\alpha)$ contain 0, so the exogeneity of the trade share variable cannot be rejected at levels $\alpha = 5\%$ or $\alpha = 10\%$.

5.2. Education and earnings

We now consider the problem of estimating the returns to schooling. The model studies a relationship between log weekly earning and the number of years of education and several other covariates (age, squared age, year of birth, ...). Several authors including Angrist and Krueger (1991) argued that schooling may be endogenous in this model and proposed to use the birth quarter as an instrument to estimate the returns to schooling consistently. The reason is individuals born in the first quarter of the year start school at an older age, and can therefore drop out after completing less schooling than individuals born near the end of the year. Hence, individuals born at the beginning of the year are likely to earn less than those born during the rest of the year. Bound et al. (1995) however, showed that the quarter of birth instruments are very weak. Doko Tchatoka and Dufour (2010, 2011) showed that DWH-tests cannot detect the endogeneity of schooling in this model, since the instruments have poor quality [see Dufour and Taamouti (2007)].

Here, we assess whether schooling is exogenous by using the projection method developed in this paper. The model is specified by:

$$y = \beta_0 + \beta_1 E + \sum_{i=1}^{k_1} \gamma_i X_i + u\,, \tag{5.4}$$

$$E = \pi_0 + \sum_{i=1}^{k_2} \pi_i Z_i + \sum_{i=1}^{k_1} \phi_i X_i + V \tag{5.5}$$

where y is log-weekly earnings, E is the number of years of education (possibly endogenous), X contains the exogenous covariates (age, age squared, 10 dummies for birth of year). Z contains

[3] The F-statistic in the first stage is about 13 as indicated in Frankel and Romer (1999, Table 2, p.385), which is not to high compared to the rule of thumb of 10 suggested by Staiger and Stock (1997) in the weak instruments case.

Table 1. Projection-based confidence sets for different parameters in growth model

AR-type CS's	97.5%	95%
$C_{\beta_1}(\alpha)$	$\{\beta_1 : 0.2306\beta_1^2 - 4.757\beta_1 + 0.043 \leq 0\}$ $= [0.009, 20.623]$	$\{\beta_1 : 0.478\beta_1^2 - 4.86\beta_1 + 1.271 \leq 0\}$ $= [0.2685, 9.896]$
$C_\theta(\alpha)$	$\{\theta : 0.305\theta^2 - 0.127\theta - 0.039 \leq 0\}$ $= [-0.205, 0.621]$	$\{\theta : 0.306\theta^2 - 0.127\theta - 0.027 \leq 0\}$ $= [-0.153, 0.569]$
$C_\theta(\alpha)$ based on $t_{w'\theta}(\gamma_0)$	$[-0.051, 0.466]$	$[-0.018, 0.433]$
Scheffé-type CS's	95%	90%
$\mathscr{C}_a(\alpha)$	$[-20.828, 0.612]$	$[-10.049, 0.300]$
$\mathscr{C}_a(\alpha)$ based on $t_{w'\theta}(\gamma_0)$	$[-20.674, 0.457]$	$[-9.9140.165]$
$C_{\sigma_{Vu}}(\alpha)$	$[-4.361, 0.128]$	$[-2.104, 0.063]$
$C_{\sigma_{Vu}}(\alpha)$ based on $t_{w'\theta}(\gamma_0)$	$[-4.329, 0.096]$	$[-2.076, 0.035]$

40 dummies obtained by interacting the quarter of birth with the year of birth. In this model, β_1 measures the return to education. The data set consists of the 5% public-use sample of the 1980 US census for men born between 1930 and 1939. The sample size is 329 509 observations.

Table 2 presents the results. We observe that $\mathscr{C}_{\beta}(\alpha)$ is unbounded indicating that β is not identified. However, $\mathscr{C}_{\theta}(\alpha)$ is bounded The latter result confirms the fact θ is always identified even if identification is weak (weak instrument). As a result, $\mathscr{C}_{a}(\alpha)$ and $\mathscr{C}_{\sigma_{Vu}}(\alpha)$ are unbounded in all cases. That indicates clearly that identification is an issue in this model.

6. Conclusion

In this paper, we have studied the problem of testing hypotheses and building confidence sets on endogeneity parameters. Such parameters have both intrinsic and statistical interest, because they represent the effect of "common factors" which induce simultaneity and determine simultaneity biases (along with other features of the data). We stressed the usefulness of distinguishing between *regression endogeneity parameters* (a) and *covariance endogeneity parameters* (σ_{Vu}): regression endogeneity parameters measure the effect of "missing variables" in linear structural equations, while covariance endogeneity parameters directly indicate which variables may be treated as "exogenous" in statistical inference. Further, regression endogeneity parameters may be tested relatively easily, and we proposed finite-sample inference methods for these. Inference on covariance endogeneity parameters involves additional nuisance parameters (e.g., the unknown covariance matrix Σ_V), so only asymptotically justified methods were given for σ_{Vu}.

The identification of endogeneity parameters was also discussed. After formulating necessary and sufficient conditions for the identification of such parameters, we observed a simple equivalence between the identification of individual regression endogeneity parameters (a_i) and the identification of the corresponding structural parameters (β_i), while this feature does not hold for covariance endogeneity parameters. In view of the possibility of identification failure, identification-robust inference procedures were proposed for endogeneity parameters. For joint hypotheses involving structural and regression endogeneity parameters, as well as marginal hypotheses on regression endogeneity parameters, finite-sample procedures were proposed. Under Gaussian errors, the tests and confidence sets are based on standard Fisher critical values. For a wide class of parametric non-Gaussian errors (possibly heavy-tailed), exact Monte Carlo procedures can be applied using the statistics considered. As a special case, this result also holds for usual AR-type tests and confidence sets on structural coefficients.

We showed that the proposed finite-sample procedures (*e.g.*, those based on a Gaussian assumption on the errors) remain asymptotically valid under weaker distributional assumptions. Tests of partial exogeneity hypotheses (for individual potentially endogenous explanatory variables) are covered as instances of the class of proposed procedures. The asymptotic theory also yields inference for covariance endogeneity. Even though the asymptotic theory is only approximate in finite samples, it is robust to identification assumptions. Finally, the proposed procedures were applied to two empirical examples: the relation between trade and economic growth, and the widely studied problem of returns to education.

Table 2. Projection-based confidence sets for different parameters in earning equation

AR-type CS's	97.5%	95%
$C_{\beta_1}(\alpha)$	$\{\beta_1 : -2.382\beta_1^2 + 0.332\beta_1 - 0.107 \leq 0\}$ $= \mathbb{R}$	$\{\beta_1 : -2.229\beta_1^2 + 0.31\beta_1 - 0.1 \leq 0\}$ $= \mathbb{R}$
$C_\theta(\alpha)$	$\{\theta : 3.527\theta^2 - 0.5\theta + 0.018 \leq 0\}$ $= [0.0701, 0.0716]$	$\{\theta : 3.527\theta^2 - .5\theta + 0.018 \leq 0\}$ $= [.0702, .0715]$
$C_\theta(\alpha)$ based on $t_{w'\theta}(\gamma_0)$	$[.0707, .0710]$	$[.0707, .0710]$
Scheffé-type CS's	95%	90%
$\mathscr{C}_a(\alpha)$	$\mathbb{R}$	$\mathbb{R}$
$C_{\sigma_{Vu}}(\alpha)$	$\mathbb{R}$	$\mathbb{R}$

APPENDIX

A. Proofs

PROOF OF LEMMA **3.3** On multiplying the two sides of (3.6) by M and $M_1 - M$, we see that:

$$\begin{aligned} M(y - Y\beta_0) &= Mu + MV(\beta - \beta_0), \\ (M_1 - M)(y - Y\beta_0) &= M_1 X_2 \Pi_2(\beta - \beta_0) + (M_1 - M)u + (M_1 - M)V(\beta - \beta_0). \end{aligned} \tag{A.1}$$

When Assumption **3.1** holds and $\beta = \beta_0$, this entails:

$$M(y - Y\beta_0) = \sigma(X)M\upsilon,\ (M_1 - M)(y - Y\beta_0) = \sigma(X)(M_1 - M)\upsilon.$$

Thus, the AR-statistic in (3.7) can be rewritten as:

$$AR(\beta_0) = \frac{\sigma(X)^2\upsilon'(M_1 - M)\upsilon/k_2}{\sigma(X)^2\upsilon'M\upsilon/(T-k)} = \frac{\upsilon'(M_1 - M)\upsilon/k_2}{\upsilon'M\upsilon/(T-k)}.$$

Hence, the null conditional distribution of $AR(\beta_0)$, given X, only depends on υ and X. If the normality assumption (3.3) also holds and υ is independent of X, then

$$\upsilon'M\upsilon \sim \chi^2(T-k),\ \upsilon'(M_1 - M)\upsilon \sim \chi^2(k_2);$$

further, since $M(M_1 - M) = 0$, $\upsilon'M\upsilon$ and $\upsilon'(M_1 - M)\upsilon$ are independent conditional on X. Consequently, $AR(\beta_0) \sim F(k_2, T-k)$. □

References

Abdelkhalek, T., Dufour, J.-M., 1998. Statistical inference for computable general equilibrium models with applications to a model of the moroccan economy. Review of Economics and Statistics, pp. 520–534.

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

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

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

Barro, R., 1977. Unanticipated money growth and unemployment in united states. American Economic Review 67, 101–115.

Bound, J., Jaeger, D. A., Baker, R. M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American Statistical Association 90, 443–450.

Doko Tchatoka, F., Dufour, J.-M., 2009. Exogeneity tests, non-Gaussian distributions and weak identification : Finite-sample and asymptotic distributional theory. Technical report, C.R.D.E., Université de Montréal.

Doko Tchatoka, F., Dufour, J.-M., 2010. On the finite-sample theory of exogeneity tests with possibly non-Gaussian errors and weak identification. Technical report, Department of Economics, McGill University, Canada Montréal, Canada.

Doko Tchatoka, F., Dufour, J.-M., 2011. Exogeneity tests and estimation in IV regressions. Technical report, Department of Economics, McGill University, Canada Montréal, Canada.

Dufour, J.-M., 1979. Methods for Specification Errors Analysis with Macroeconomic Applications PhD thesis University of Chicago. 257 + XIV pages. Thesis committee: Arnold Zellner (Chairman), Robert E. Lucas and Nicholas Kiefer.

Dufour, J.-M., 1987. Linear Wald methods for inference on covariances and weak exogeneity tests in structural equations. In: I. B. MacNeill, G. J. Umphrey, eds, Advances in the Statistical Sciences: Festschrift in Honour of Professor V.M. Joshi's 70th Birthday. Volume III, Time Series and Econometric Modelling. D. Reidel, Dordrecht, The Netherlands, pp. 317–338.

Dufour, J.-M., 1990. Exact tests and confidence sets in linear regressions with autocorrelated errors. Econometrica 58, 475–494.

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

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

Dufour, J.-M. , 2006. Monte carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics. Journal of Econometrics 138, 2649–2661.

Dufour, J.-M. , Hsiao, C. , 2008. Identification. In: L. E. Blume , S. N. Durlauf, eds, The New Palgrave Dictionary of Economics 2nd edn . Palgrave Macmillan, Basingstoke, Hampshire, England. forthcoming.

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

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

Dufour, J.-M., Taamouti, M., 2007. Further results on projection-based inference in IV regressions with weak, collinear or missing instruments. Journal of Econometrics 139(1), 133–153.

Durbin, J., 1954. Errors in variables. Review of the International Statistical Institute 22, 23–32.

Engle, R. F., 1982. A general approach to Lagrange multiplier diagnostics. Journal of Econometrics 20, 83–104.

Farebrother, R. W., 1976. A remark on the Wu test. Econometrica 44, 475–477.

Frankel, J. A., Romer, D., 1999. Does trade cause growth?. American Economic Review 89(3), 379–399.

Harrison, A. , 1996. Oponness and growth: a time-series, cross-country analysis for developing countries. Journal of Development Economics 48, 419–447.

Hausman, J., 1978. Specification tests in econometrics. Econometrica 46, 1251–1272.

Hausman, J., Taylor, W. E., 1981. A generalized specification test. Economics Letters 8, 239–245.

Holly, A., 1982. A remark on Hausman's test. Econometrica 50, 749–759.

Hwang, H.-S. , 1980. Test of independence between a subset of stochastic regressors and disturbances. International Economic Review 21, 749–760.

Irwin, A.-D., Tervio, M., 2002. Does trade raise income? evidence from twentieth century. Journal of International Economics 58, 1–18.

Kariya, T., Hodoshima, H., 1980. Finite-sample properties of the tests for independence in structural systems and LRT. The Quarterly Journal of Economics 31, 45–56.

Kiviet, J. F., Niemczyk, J., 2007. The asymptotic and finite-sample distributions of OLS and simple IV in simultaneous equations. Computational Statistics and Data Analysis 51, 3296–3318.

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

Mankiw, N. G., Romer, D., Weil, D. N., 1992. A contribution to the empirics of economic growth. The Quarterly Journal of Economics 107(2), 407–437.

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

Nakamura, A., Nakamura, M., 1981. On the relationships among several specification error tests presented by Durbin, Wu and Hausman. Econometrica 49, 1583–1588.

Revankar, N. S., 1978. Asymptotic relative efficiency analysis of certain tests in structural sysytems. International Economic Review 19, 165–179.

Revankar, N. S., Hartley, M. J., 1973. An independence test and conditional unbiased predictions in the context of simultaneous equation systems. International Economic Review 14, 625–631.

Reynolds, R. A., 1982. Posterior odds for the hypothesis of independence between stochastic regressors and disturbances. International Economic Review 23(2), 479–490.

Smith, R. J., 1984. A note on likelihood ratio tests for the independence between a subset of stochastic regressors and disturbances. International Economic Review 25, 263–269.

Spencer, D. E., Berk, K. N., 1981. A limited-information specification test. Econometrica 49, 1079–1085.

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. Journal of Business and Economic Statistics 20(4), 518–529.

Wu, D.-M., 1973. Alternative tests of independence between stochastic regressors and disturbances. Econometrica 41, 733–750.

Wu, D.-M., 1974. Alternative tests of independence between stochastic regressors and disturbances: Finite sample results. Econometrica 42, 529–546.

Wu, D.-M., 1983a. A remark on a generalized specification test. Economics Letters 11, 365–370.

Wu, D.-M., 1983b. Tests of causality, predeterminedness and exogeneity. International Economic Review 24(3), 547–558.