The Econometrics Journal

*Econometrics Journal* (2014), volume **17**, pp. 165–187.
doi: 10.1111/ectj.12021

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

FIRMIN DOKO TCHATOKA[†] AND JEAN-MARIE DUFOUR[‡]

[†]*School of Economics and Finance, University of Tasmania, Private Bag 85, Hobart, TAS 7001, Australia*
E-mail: Firmin.dokotchatoka@utas.edu.au

[‡]*Department of Economics, McGill University, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada*
E-mail: jean-marie.dufour@mcgill.ca

First version received: August 2012; final version accepted: September 2013

**Summary** We provide a generalization of the Anderson–Rubin (AR) procedure for inference on parameters that represent the dependence between possibly endogenous explanatory variables and disturbances in a linear structural equation (endogeneity parameters). We stress the distinction between regression and covariance endogeneity parameters. Such parameters have intrinsic interest (because they measure the effect of latent variables, which induce simultaneity) and play a central role in selecting an estimation method (such as ordinary least-squares or instrumental variable methods). We observe that endogeneity parameters might not be identifiable and we give the relevant identification conditions. These conditions entail a simple identification correspondence between regression endogeneity parameters and the usual structural parameters, while the identification of covariance endogeneity parameters typically fails as soon as global identification fails. 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 Fisher critical values. For a wide class of parametric non-Gaussian errors (possibly heavy-tailed), we 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 special cases. The proposed tests are applied to two empirical examples: the relation between trade and economic growth, and the widely studied problem of returns to education.

**Keywords:** *AR-type statistic, Endogeneity, Identification-robust confidence sets, Partial exogeneity test, Projection-based techniques.*

## 1. INTRODUCTION

Instrumental variable (IV) regressions are typically motivated by the fact that explanatory variables can be correlated with the error term, so least-squares methods yield biased inconsistent estimators of model coefficients. As is well known, IV estimates are obtained by isolating the variation in endogenous explanatory variables due to exogenous regressors excluded from the explanatory variables, and correlating this variation with that of the dependent variable of interest. Even though coefficients estimated in this way might have interesting interpretations from the point of view of economic theory, inference on such structural parameters faces identification difficulties. Furthermore, it is well known that IV estimators can be very imprecise, while 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 body of literature on reliable inference in the presence of weak instruments; see the reviews of Stock et al. (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 that 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 latent variables, which are typically unobserved and can simultaneously affect a number of observable endogenous variables. These latent variables 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 can represent unobserved latent variables, such as surprise variables, which play a role in models with expectations; see Barro (1977), and 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) might be important for selecting statistical procedures. Even if instruments are strong, it is well known that IV estimators can be considerably less efficient than least-squares estimators; see Kiviet and Niemczyk (2007, 2012), Doko Tchatoka and Dufour (2011a), Kiviet and Pleus (2012), and Kiviet (2013). Indeed, this might 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.

We stress here the view that linear structural models (IV regressions) can be interpreted as regressions with missing regressors. If the latter 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 to a linear regression where all the explanatory variables are uncorrelated with the error term. We call this equation the orthogonalized structural equation, and we use it extensively. Interestingly, the latter is not a reduced-form equation. Instead, it involves the structural parameters of interest, but also includes endogeneity parameters that are hidden in the original structural equation.

The problem stems from the fact that the missing regressors are unobserved. Despite this difficulty, we show that procedures similar to the Anderson–Rubin (AR) procedure, as proposed by Anderson and Rubin (1949), 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 characterized as identification-robust. For more general inference on transformations of the parameters in 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 can 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 for $a$ itself is then derived by exploiting the possibility of making identification-robust inference on $\beta$. Then, inference on covariances (i.e., $\sigma_{Vu}$) between $u$ and $Y$ (covariance endogeneity parameters) can be derived by considering 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 non-singular), but the relationship is more complex in other cases.

In this paper, we emphasize cases where $a \neq 0$. We first study formally the identification of endogeneity parameters. We establish a simple identification correspondence between the components of $\beta$ and $a$: each component of $a$ is identifiable if and only if the corresponding component of $\beta$ is identifiable. In contrast, this does not hold in general for the covariances $\sigma_{Vu}$: as soon as one element of $\beta$ is not identifiable, all components of $\sigma_{Vu}$ typically fail to be identifiable. In this sense, $\sigma_{Vu}$ is more difficult to interpret than $a$. Because of the failure of the exogeneity hypothesis, the distributions of the test statistics are much more complex. It is relatively easy to produce finite-sample inference for $a$, but not for $\sigma_{Vu}$. So, for $\sigma_{Vu}$, we propose asymptotic tests and confidence sets. It is important to note that stronger assumptions are needed for making inference on $\sigma_{Vu}$ (as opposed to $a$). Indeed, we describe general distributional set-ups where $\sigma_{Vu}$ might not be well defined (because of heterogeneity in the model for $Y$, or the non-existence of moments), while $a$ remains well defined and statistically meaningful. In such cases, inference on $a$ is feasible, while inference on $\sigma_{Vu}$ might not be (even when all parameters in the structural equation of interest are identifiable).

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, 1983), Revankar and Hartley (1973), 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), Smith (1983, 1984, 1985), Staiger and Stock (1997), and Doko Tchatoka and Dufour (2011a, b). By contrast, we consider here the problem of testing any value of $a$ (or $\sigma_{Vu}$) and build confidence sets for these parameters. By allowing for weak instruments, we extend the results in Dufour (1979, 1987) where Wald-type tests and confidence sets are proposed for inference on $a$ and $\sigma_{Vu}$, under assumptions that exclude weak instruments. Finally, by considering inference on $a$ and $\sigma_{Vu}$, we extend a procedure proposed by Dufour and Jasiak (2001) for inference on the aggregate parameter $\theta = \beta + a$ (but not $a$ or $\sigma_{Vu}$) in the context of a 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 errors follow a completely specified distribution (up to an unknown scale – possibly random – parameter), which might be non-Gaussian. Because of this invariance property, exact Monte Carlo tests can be performed without a Gaussian assumption, as described by Dufour (2006). In particular, we show this is feasible under general assumptions, which allow considerable heterogeneity in the reduced-form model for $Y$, even a completely unspecified model for $Y$. 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: the relationship between trade and economic growth (Frankel and Romer, 1999, and Dufour and Taamouti, 2007), and the model of returns to education studied by Card (1995) and Kleibergen (2004, Table 2, p. 421).

The paper is organized as follows. In Section 2, we describe the model and the identification conditions for endogeneity parameters. In Section 3, we present the finite-sample theory for inference on regression endogeneity parameters. In Section 4, we discuss asymptotic theory and inference for covariance endogeneity parameters. In Section 5, we present empirical applications.

## 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 $y = [y_1, \ldots, y_T]'$ is a $T \times 1$ vector of observations on a dependent variable, $Y = [Y_1, \ldots, Y_T]'$ 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, and $\beta$ and $\gamma$ 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), $X = [X_1, X_2] = [X_{\bullet 1}, \ldots, X_{\bullet T}]'$ has full-column rank $k = k_1 + k_2$, $\Pi_1$ and $\Pi_2$ 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), we obtain 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 means (although this assumption could easily be replaced by another location assumption, such as zero medians), the usual necessary and sufficient condition for identification of $\beta$ and $\gamma$ (from the first moments of $y$ and $Y$) in (2.1) and (2.2) is

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

If $\Pi_2 = 0$, the instruments $X_2$ are irrelevant, and $\beta$ is completely unidentified. If $1 \leq \text{rank}(\Pi_2) < G$, $\beta$ is not identifiable, but some linear combinations of the elements of $\beta$ are identifiable (see Dufour and Hsiao, 2008). If $\Pi_2$ 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 $\beta$ are ill determined by the data, a situation often called weak identification (see Dufour, 2003).

Throughout this paper, $I_m$ is the identity matrix of order $m$. For any full-column-rank $T \times m$ matrix $A$, $P(A) = A(A'A)^{-1}A'$, $M(A) = I_T - P(A)$, vec$(A)$ is the $(Tm) \times 1$ column vectorization of $A$ and $\|A\| = [\text{tr}(A'A)]^{1/2}$ is the matrix Euclidian norm. For $A$ square, $A > 0$ means $A$ is positive definite (p.d.), and $A \geq 0$ means $A$ is positive semi-definite (p.s.d.). We use $\overset{p}{\to}$ to denote convergence in probability, and $\overset{d}{\to}$ to denote convergence in distribution.

### *2.1. Identification of endogeneity parameters*

We now wish to represent the fact that $u$ and $V$ can be correlated, allowing for the possibility of identification failure. It is important to note that the structural error $u_t$ might not be uniquely determined by the data when $\beta$ and $\gamma$ are not identified. For this, it will be useful to consider two alternative set-ups for the disturbance distribution. In the first set-up, the disturbance vectors $(u_t, V_t')'$ have common finite second moments (structural homoscedasticity). In the second set-up, we allow for a large amount of heterogeneity in the distributions of reduced-form errors (reduced-form heterogeneity). The second set-up is more appropriate for practical work, and we wish to go as far as possible in that direction. However, it will be illuminating to consider the first set-up.

ASSUMPTION 2.1 (STRUCTURAL HOMOSCEDASTICITY). *The vectors $U_t = (u_t, V_t')'$, $t = 1, \ldots, T$, all have finite second moments with mean zero and the same covariance matrix*

$$\Sigma_U = E[U_t U_t'] = \begin{bmatrix} \sigma_u^2 & \sigma_{Vu}' \\ \sigma_{Vu} & \Sigma_V \end{bmatrix}, \tag{2.5}$$

*where $\Sigma_V = E[V_t V_t']$ is non-singular.*

Under the above assumption, we have

$$\sigma_{Vv} = E[V_t v_t] = E[V_t(V_t'\beta + u_t] = \Sigma_V \beta + \sigma_{Vu}, \quad \sigma_v^2 = \sigma_u^2 + \beta'\Sigma_V\beta + 2\beta'\sigma_{Vu}. \tag{2.6}$$

The covariance vector $\sigma_{Vu}$ indicates which variables in $Y$ are correlated with $u_t$, so it is a basic determinant of the level of endogeneity of these variables. Note, however, that $\sigma_{Vu}$ is not identifiable when $\beta$ is not (for then 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 $V_t$,

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

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

$$v_t = V_t'\theta + \eta_t, \quad t = 1, \ldots, T, \tag{2.8}$$

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

$$\sigma_{Vu} = \Sigma_V a, \qquad \sigma_u^2 = \sigma_e^2 + a'\Sigma_V a = \sigma_e^2 + \sigma_{Vu}'\Sigma_V^{-1}\sigma_{Vu}, \tag{2.9}$$

where $E[e_t^2] = \sigma_e^2$ 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 $a = 0$. However, there is no simple match between the components of $a$ and $\sigma_{Vu}$ (unless $\Sigma_V$ is a diagonal matrix). For example, if $a = (a_1', a_2')'$ and $\sigma_{Vu} = (\sigma_{Vu1}', \sigma_{Vu2}')'$ where $a_1$ and $\sigma_{Vu1}$ have dimension $G_1 < G$, $a_1 = 0$ is not equivalent to $\sigma_{Vu1} = 0$. We call $a$ the regression endogeneity parameter, and $\sigma_{Vu}$ the covariance endogeneity parameter.

As long as the identification condition (2.4) holds, both $\sigma_{Vu}$ and $a$ are identifiable. This is not the case, however, when (2.4) does not hold. By contrast, the regression coefficient $\theta$ 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:

$$\theta = \Sigma_V^{-1}\sigma_{Vv} = \Sigma_V^{-1}(\Sigma_V\beta + \sigma_{Vu}) = \beta + a. \tag{2.10}$$

In other words, the sum $\beta + a$ is equal to the regression coefficient of $v_t$ on $V_t$. Even though $\beta$ and $a$ might not be identifiable, the sum $\beta + a$ is identifiable (from the first and second moments of $v$ and $V$). Further, for any fixed $G \times 1$ vector $w$, $w'\theta$ is identifiable, so the identities $w'a = w'\theta - w'\beta$ and $\sigma_{Vu} = \Sigma_V a$ along with the invertibility of $\Sigma_V$ entail the following equivalences:

$$\beta \text{ is identifiable } \Leftrightarrow a \text{ is identifiable} \Leftrightarrow \sigma_{Vu} \text{ is identifiable;} \tag{2.11}$$

$$w'\beta \text{ is identifiable} \Leftrightarrow w'a \text{ is identifiable} \Leftrightarrow w'\Sigma_V^{-1}\sigma_{Vu} \text{ is identifiable.} \tag{2.12}$$

In particular, (2.12) entails a simple identification correspondence between the components of $\beta$ and $a$: for each $1 \le i \le G$, $a_i$ is identifiable $\Leftrightarrow$ $\beta_i$ is identifiable. In other words, the identification conditions for $\beta$ and $a$ are identical. In contrast, the equivalences [$w'\sigma_{Vu}$ is identifiable $\Leftrightarrow$ $w'\beta$ is identifiable] and [$\sigma_{Vui}$ is identifiable $\Leftrightarrow$ $\beta_i$ is identifiable] do not hold, in general: as soon as one element of $\beta$ is not identifiable, all components of $\sigma_{Vu}$ typically fail to be identifiable. In this sense, $\sigma_{Vu}$ is more difficult to interpret than $a$.

The first set-up (Assumption 2.1) requires that the reduced-form disturbances $V_t$, $t = 1, \ldots, T$, have identical second moments. In many practical situations, this might not be appropriate, especially in a limited-information analysis, which 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 consider the following alternative assumptions (where $X_{\bullet t}$ is the $t$th row of $X$).

ASSUMPTION 2.2 (SECOND-ORDER REDUCED-FORM HETEROGENEITY). *For some fixed vector $a$ in $\mathbb{R}^G$, we have*

$$u = Va + e \tag{2.13}$$

*where $e$, $V$, and $X$ have finite second moments, $E[e] = 0$, and $e$ is uncorrelated with $V$ and $X$.*

ASSUMPTION 2.3 (REDUCED-FORM HETEROGENEITY). *Equation (2.13) holds with $E[e_t|V_t, X_{\bullet t}] = 0,\ t = 1, \ldots, T$.*

Assumptions 2.2 and 2.3 allow substantial heterogeneity in the distribution of the disturbances $V_t$, $t = 1, \ldots, T$. The latter need not be identically distributed or independent.

Assumption 2.2 maintains the existence of second moments (even though the covariance matrices $E[V_t V_t']$ can vary with $t$) and defines $e$ through a zero mean and orthogonality with $V$ and $X$. Assumption 2.3 replaces this condition by a zero conditional mean; no further restriction on $V$ is imposed. The existence of moments for $V_t$ and $X_{\bullet t}$ is not required. An important case where Assumption 2.2 holds is the one where $V$ and $e$ are independent (strong linear structural decomposition). Given equation (2.1), the three conditions $E[e_t|V_t, X_{\bullet t}] = 0$, $E[e_t|Y_t, X_{\bullet t}] = 0$, and $E[e_t|Y_t, V_t, X_{\bullet t}] = 0$ are equivalent. In such cases, $\sigma_{Vu}$ might not be well defined (because of heterogeneity in the model for $Y$, or the non-existence of moments), but $a$ remains statistically meaningful.

In view of the decomposition (2.13), equation (2.1) can be viewed as a regression model with missing regressors. On substituting (2.13) into (2.1), we obtain

$$y = Y\beta + X_1\gamma + Va + e, \tag{2.14}$$

where $e$ is uncorrelated with all the regressors. Because of this property, we call (2.14) the orthogonalized structural equation associated with (2.2), and we call $e$ the orthogonalized structural disturbance vector.[1] In this equation, the original structural parameters ($\beta$ and $\gamma$) can be interpreted as regression coefficients, along with the regression endogeneity parameter $a$. We see that $a$ represents the effect of the latent variable $V$. Even though (2.14) is a regression equation (i.e., $(Y, X_1, V)$ is orthogonal to the disturbance $e$), it is quite distinct from the reduced-form equation (2.3) for $y$.

The orthogonalized structural equation is quite helpful for interpreting model coefficients. A structural model of the form (2.1) and (2.2) often represents a causal structure to explain $y$. The endogenous variables ($y$ and $Y$) are determined by two types of inputs: observable exogenous variables ($X_1$ and $X_2$) and unobserved variables ($V$ and $e$). Here, $X_1$ has both a direct effect ($X_1\gamma$) on $y$ and an indirect effect ($X_1\Pi_1\beta$ through $Y$), while $X_2$ only has an indirect effect ($X_2\Pi_2\beta$). Similarly, $V$ represents unobserved variables (e.g., shocks, latent variables, expectation errors), which have both a direct effect ($Va$) and an indirect effect ($V\beta$), while $e$ represents idiosyncratic shocks to $y$, which are orthogonal to $Y$. Finally, we can interpret the sum $V\beta + Va = V(\beta + a)$ as the net final effect (both direct and indirect) of $V$ on $y$. In the context of a causal interpretation, the coefficient vectors $\beta$, $a$, and $\beta + a$ have useful distinct interpretations: $\beta$ represents the impact of $Y$ (in particular, its systematic component $E[Y] = X_1\Pi_1 + X_2\Pi_2$) on $y$, $a$ the direct effect of the latent variable $V$ on $y$, and $\beta + a$ the total effect of $V$ on $y$. Statistical inference on each of the coefficients has its own independent interest.

The identification of $a$ can be studied through the orthogonalized structural equation. By equation (2.2),

$$y = Y\theta + X_1\pi_1^* + X_2\pi_2^* + e, \tag{2.15}$$

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 (2.15) is 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 call (2.15) 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 (almost surely (a.s.)), the parameters of equation (2.15)

[1] The form (2.14) was originally proposed by Revankar and Hartley (1973) for the purpose of testing complete exogeneity ($a = 0$). As pointed out by Dufour (1979, 1987), the distributional theory is substantially simpler in that case and does not allow one to test more general restrictions on $a$ (because the covariance matrix is modified).

are identifiable (a.s.), because they are uniquely determined by the linear projections of $y_t$ on $Y_t$ and $X_{\bullet t}$ for $t = 1, \ldots, T$ (under Assumption 2.2) or by the corresponding conditional means (under assumption 2.3). 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 (a.s., conditional on $X$), for example, if its distribution is absolutely continuous. This entails again that $a$ is identifiable if and only if $\beta$ is identifiable, and similarly between $w'a$ and $w'\beta$ for any $w \in \mathbb{R}^G$. This establishes the following identification result for $a$, where identification refers to the conditional distributions of $y_t$ given $Y_t$ and $X_{\bullet t}$, $t = 1, \ldots, T$.

PROPOSITION 2.1 (IDENTIFICATION OF REGRESSION ENDOGENEITY PARAMETERS). *Under the model given by (2.2), (2.3), and Assumption 2.2 or 2.3, suppose the matrix* $[Y, X_1, X_2]$ *has full-column rank. Then,* $a + \beta$ *is identifiable, and the following two equivalences hold:*

$$a \text{ is identifiable} \Leftrightarrow \beta \text{ is identifiable}; \tag{2.16}$$

$$\text{for any } w \in \mathbb{R}^G, \; w'a \text{ is identifiable} \Leftrightarrow w'\beta \text{ is identifiable}. \tag{2.17}$$

Under Assumption 2.2, covariance endogeneity parameters can depend on $t$. Indeed, it is easy to see that $E[V_t u_t] = E[V_t V_t^{'}]a \equiv \sigma_{Vut}$, which might depend on $t$ if $E[V_t V_t^{'}]$ does. However, the identification of the parameters $\sigma_{Vut}$ remains determined by the identification of $a$, whenever the reduced-form covariances (which are parameters of reduced forms) are identifiable. Inference on covariance endogeneity parameters requires additional assumptions. In Sections 3 and 4, we see that finite-sample inference methods can be derived for regression endogeneity parameters under the relatively weak Assumption 2.2, while only asymptotically justified methods are proposed for covariance endogeneity parameters. For covariances, we focus on the case where $\sigma_{Vut}$ is constant.

### 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 ($\sigma_{Vu}$), allowing for the possibility of identification failure (or weak identification). We develop inference procedures for the full vectors $a$ and $\sigma_{Vu}$, 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 of the identification of endogeneity parameters, we observe that inference on $a$ can be tackled more easily than inference on $\sigma_{Vu}$, so we study this problem first. The problem of testing hypotheses of the form $H_a(a_0) : a = a_0$ can be viewed as an extension of the classical AR problem on testing $H_\beta(\beta_0) : \beta = \beta_0$. However, there is 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, 1987, 1990, Abdelkhalek and Dufour, 1998, Dufour and Jasiak, 2001, and 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 $\alpha_1$ and $\alpha_2$ such that $0<\alpha=\alpha_1+\alpha_2<1$, $0<\alpha_1<1$, and $0<\alpha_2<1$. We can then proceed as follows.

STEP 1. We build an identification-robust confidence set with level $1-\alpha_1$ for $\beta$. There are several methods available to do this. In view of the existence of a finite-sample distributional theory (as well as computational simplicity), we focus on the AR approach, but alternative procedures could be exploited for this purpose.[2]

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

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

STEP 4. Confidence sets for $a=\theta-\beta$ and any linear transformation $w'a$ can then be derived by projection. These confidence sets have level $1-\alpha$.

STEP 5. Confidence sets for $\sigma_{Vu}$ and $w'\sigma_{Vu}$ can finally be built using 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 $\beta$ (the AR-type statistic) and $\theta$ enjoy invariance properties that 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 $\beta$ and $\theta$ through a small number of Monte Carlo simulations (see Dufour, 2006). For inference on both regression and covariance endogeneity parameters ($a$ and $\sigma_{Vu}$), we also provide a large-sample distributional theory based on standard asymptotic assumptions, which relax various restrictions used in the finite-sample theory. None of the proposed methods makes identification assumptions on $\beta$, 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 point of view. Along with (2.1) and (2.2), we suppose that Assumption 2.2 holds with the following condition on $u$.

ASSUMPTION 3.1 (CONDITIONAL SCALE MODEL FOR STRUCTURAL ERRORS). $u=\sigma(X)\upsilon$, *where* $\sigma(X)$ *is a (possibly random) function of* $X$ *such that* $\mathbb{P}[\sigma(X)\neq 0|X]=1$, *and the conditional distribution of* $\upsilon$ *given* $X$ *is completely specified.*

ASSUMPTION 3.2 (CONDITIONAL SCALE MODEL FOR ORTHOGONALIZED STRUCTURAL ERRORS). $e=\sigma_1(X)\varepsilon$, *where* $\sigma_1(X)$ *is a (possibly random) function of* $X$ *such that* $\mathbb{P}[\sigma_1(X)\neq 0|X]=1$, *and the conditional distribution of* $\varepsilon$ *given* $X$ *is completely specified.*

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

Assumption 3.1 means the distribution of $u$ given $X$ only depends on $X$ and a (typically unknown) scale factor $\sigma(X)$. The scale factor can also be random, so we can have $\sigma(X) = \bar{\sigma}(X, \upsilon)$. Of course, this holds whenever $u = \sigma\upsilon$, where $\sigma$ is an unknown positive constant and $\upsilon$ is independent of $X$ with a completely specified distribution. In this context, the standard Gaussian assumption is obtained by taking $\upsilon \sim \mathrm{N}[0, I_T]$. However, non-Gaussian distributions are covered, including heavy-tailed distributions, which can lack moments (such as the Cauchy distribution). Similarly, Assumption 3.2 means the distribution of $e$ given $X$ only depends on $X$ and a (typically unknown, possibly random) scale factor $\sigma_1(X)$. So, again, a standard Gaussian model is obtained by assuming that $\sigma_1(X)$ is fixed (given $X$) and $\varepsilon \sim \mathrm{N}[0, I_T]$. 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. (given $X$) with finite second moments and the decomposition specified by Assumption 2.2 holds. This will be the case *a fortiori* if the vectors $(u_t, V_t')'$, $t, \ldots, T$, are i.i.d. multinormal (given $X$).

We study in turn the following problems: (1) we test and build confidence sets for $\beta$; (2) we test and build confidence sets for $\theta = \beta + a$; (3) we test and build confidence sets for $a$; (4) we test and build confidence sets for scalar linear transformations $w'a$.

### *3.1. AR-type tests for $\beta$ with possibly non-Gaussian errors*

Because this is a basic building block for inference on endogeneity parameters, we consider first the problem of testing the hypothesis

$$H_\beta(\beta_0) : \beta = \beta_0, \tag{3.1}$$

where $\beta_0$ is any given possible value of $\beta$. Several methods have been proposed for this purpose. However, because we wish to use an identification-robust procedure for which a finite-sample theory can easily be obtained and that does not require assumptions on the distribution of $Y$, we focus on the AR procedure. So we consider the transformed equation

$$y - Y\beta_0 = X_1\pi_1^0 + X_2\pi_2^0 + \upsilon^0, \tag{3.2}$$

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

$$AR(\beta_0) = \frac{(y - Y\beta_0)'(M_1 - M)(y - Y\beta_0)/k_2}{(y - Y\beta_0)'M(y - Y\beta_0)/(T - k)}, \tag{3.3}$$

where $M_1 \equiv M(X_1)$ and $M \equiv M(X)$. 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 by invariance the distribution of $AR(\beta_0)$ under the general Assumption 3.1.

PROPOSITION 3.1 (NULL DISTRIBUTION OF AR STATISTICS UNDER SCALE STRUCTURAL ERROR MODEL). *Suppose equations (2.1) and (2.2) and Assumption 3.1 hold. If $\beta = \beta_0$, we have*

$$AR(\beta_0) = \frac{\upsilon'(M_1 - M)\upsilon/k_2}{\upsilon'M\upsilon/(T - k)}, \tag{3.4}$$

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

The proof is given in the Appendix. This proposition means that the conditional null distribution of $AR(\beta_0)$, given $X$, only depends on the distribution of $\upsilon$. The distribution of $V$ plays no role here, so no decomposition assumption (such as Assumption 2.1 or 2.2) is needed. If the distribution of $\upsilon|X$ can be simulated, we can obtain 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 (e.g., the Cauchy distribution).[3] This might be useful in financial models with fat-tailed error distributions, such as the Student-$t$ distribution.

When the normality assumption holds $\big(\upsilon \sim \mathrm{N}[0, I_T]\big)$ 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 $F(k_2, T-k)$ distribution. A confidence set with level $1-\alpha$ for $\beta$ is then given by

$$\mathcal{C}_\beta(\alpha) = \{\beta_0 : AR(\beta_0) \leq F_\alpha(k_2, T-k)\} = \{\beta : Q(\beta) \leq 0\}, \tag{3.5}$$

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

### 3.2. *Inference on $\theta$*

Let us now consider the problem of testing the hypothesis $H_\theta(\theta_0) : \theta = \theta_0$, where $\theta_0$ is a given vector of dimension $G$, and Assumption 3.2 holds. This can be done by considering the extended reduced form in (2.15). By Assumption 3.2, $e$ is independent of $Y$, $X_1$, and $X_2$, and (2.15) is a linear regression model. As soon as the matrix $[Y, X_1, X_2]$ has full-column rank, the parameters of equation (2.15) can be tested through standard $F$-tests.

We 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

$$F_\theta(\theta_0) = \frac{(\hat{\theta} - \theta_0)'(Y'MY)(\hat{\theta} - \theta_0)/G}{y'M(Z)y/(T-G-k)}, \tag{3.6}$$

where $\hat{\theta} = (Y'MY)^{-1}Y'My$ is the ordinary least-squares (OLS) estimate of $\theta$ in (2.15), $M = M(X)$, $X = [X_1, X_2]$, and $Z = [Y, X_1, X_2]$. When $\upsilon \sim \mathrm{N}[0, I_T]$, we have $F_\theta(\theta_0) \sim F(G, T-k-G)$ under $H_\theta(\theta_0)$. Under the more general Assumption 3.2, it is easy to see that

$$F_\theta(\theta_0) = \frac{\varepsilon' MY(Y'MY)^{-1}Y'M\varepsilon/G}{\varepsilon' M(Z)\varepsilon/(T-G-k)}, \tag{3.7}$$

[3] By "robustness to weak instruments", we mean the fact that the null distribution of the test statistic remains valid even if $\mathrm{rank}[\Pi_2] < G$, so $\beta$ might not be identifiable from the available data. By "robustness to excluded instruments", we mean that the test remains valid even if $Y$ depends on additional explanatory variables ($X_3$), which are not taken in IV-based inference; for further discussion of this issue, see Dufour and Taamouti (2007). Of course, identification failure (or weak identification) typically affects test power and confidence set precision. For example, if identification fails completely ($\mathrm{rank}[\Pi_2] = 0$), it is impossible to distinguish between alternative values of $\beta$, and a valid test of $H_\beta(\beta_0)$ should have power no larger than its level. Further, confidence sets of unidentified parameters should be uninformative (e.g., unbounded) with high probability (see Dufour, 1997).

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 $\theta_0$, so the same critical value can be applied irrespective of $\theta_0$. The main difference with the Gaussian case is that the critical value can depend on $Y$ and $X$. Irrespective of the case, we denote by $c(\alpha_2)$ the critical value for $F_\theta(\theta_0)$.

From (3.6), a confidence set with level $1-\alpha$ for $\theta$ can be obtained by inverting $F_\theta(\theta_0)$

$$\mathcal{C}_\theta(\alpha) = \left\{\theta_0 : F_\theta(\theta_0) \le \bar{f}(\alpha)\right\} = \left\{\theta_0 : \bar{Q}(\theta_0) \le 0\right\}, \tag{3.8}$$

where $\bar{Q}(\theta) = (\hat{\theta}-\theta)'(Y'MY)(\hat{\theta}-\theta) - \bar{c}_0 = \theta'\bar{A}\theta + \bar{b}'\theta + \bar{c}$, $\bar{c}_0 = \bar{f}(\alpha)Gs^2$, $s^2 = y'M(Z)y/(T-G-k)$, $\bar{A} = Y'MY$, $\bar{b} = -2\bar{A}\hat{\theta} = -2Y'My$, $\bar{c} = \hat{\theta}'\bar{A}\hat{\theta} - \bar{c}_0 = \hat{\theta}'(Y'MY)\hat{\theta} - \bar{c}_0 = y'\tilde{H}y$, and $\bar{H} = P(MY) - \bar{f}(\alpha)(G/(T-G-k))M_1$. Because the matrix $\bar{A}$ is positive definite (with probability one), the quadric set $\mathcal{C}_\theta(\alpha)$ is an ellipsoid (hence bounded); see Dufour and Taamouti (2005, 2007). This reflects the fact that $\theta$ 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 (2.15) as a linear regression, we can test in the same way linear restrictions of the form $H_{w'\theta}(\gamma_0) : w'\theta = \gamma_0$, where $w$ is a $G \times 1$ vector and $\gamma_0$ is known constant. We can then use the corresponding $t$-statistic

$$t_{w'\theta}(\gamma_0) = \frac{w'\hat{\theta} - \gamma_0}{s\left(w'(Y'MY)^{-1}w\right)^{1/2}}, \tag{3.9}$$

and reject $H_{w'\theta}(\gamma_0)$ when $|t_{w'\theta}(\gamma_0)| > c_w(\alpha)$, where $c_w(\alpha)$ is the critical value for a test with level $\alpha$. 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 $\varepsilon$ follows a non-Gaussian distribution, we have

$$t_{w'\theta}(\gamma_0) = \frac{(T-G-k)^{1/2} w'(Y'MY)^{-1}Y'M\varepsilon}{\left(\varepsilon'M(Z)\varepsilon\right)^{1/2}\left(w'(Y'MY)^{-1}w\right)^{1/2}} \tag{3.10}$$

under $H_{w'\theta}(\gamma_0)$, so that the distribution of $t_{w'\theta}(\gamma_0)$ can be simulated like $F_\theta(\theta_0)$ in (3.7).

### 3.3. Joint inference on $\beta$ and regression endogeneity parameters

We can now derive confidence sets for the vectors $(\beta', a')'$ and $(\beta', \theta')'$. Consider the set

$$\begin{aligned}\mathcal{C}_{(\beta,\theta)}(\alpha_1,\alpha_2) &= \{(\theta_0', \beta_0')' : \beta_0 \in \mathcal{C}_\beta(\alpha_1), \theta_0 \in \mathcal{C}_\theta(\alpha_2)\} \\ &= \{(\theta_0', \beta_0')' : Q(\beta_0) \le 0, \bar{Q}(\theta_0) \le 0\}.\end{aligned}$$

Using the Boole–Bonferroni inequality, we have

$$\mathbb{P}[\beta \in \mathcal{C}_\beta(\alpha_1) \quad \text{and} \quad \theta \in \mathcal{C}_\theta(\alpha_2)] \ge 1 - \mathbb{P}[\beta \notin \mathcal{C}_\beta(\alpha_1)] - \mathbb{P}[\theta \notin \mathcal{C}_\theta(\alpha_2)] \ge 1 - \alpha_1 - \alpha_2. \tag{3.11}$$

So, $\mathcal{C}_{(\beta,\theta)}(\alpha_1,\alpha_2)$ is a confidence set for $(\beta', \theta')'$ 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 (3.8) as a function of $\beta$ and $a$:

$$\bar{Q}(\theta) = \bar{Q}(\beta + a) = a'\bar{A}a + (\bar{b} + 2\bar{A}\beta)'a + \left(\bar{c} + \bar{b}'\beta + \beta'\bar{A}\beta\right).$$

Thus, we obtain a confidence set with level $1 - \alpha$ for $\beta$ and $a$ by taking

$$\bar{\mathcal{C}}_{(\beta,a)}(\alpha) = \{(\beta_0', a_0')' : Q(\beta_0) \leq 0 \text{ and } \bar{Q}(\beta_0 + a_0) \leq 0\}. \tag{3.12}$$

Thus, finite-sample inference on the structural (possibly unidentifiable) parameter $a$ can be achieved. 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 $\beta$ and $a$. Because the event $(\beta, a) \in \bar{\mathcal{C}}_{(\beta,a)}(\alpha)$ entails $g(\beta, a) \in g[\bar{\mathcal{C}}_{(\beta,a)}(\alpha)]$, where $g[\bar{\mathcal{C}}_{(\beta,a)}(\alpha)] = \{g(\beta, a) : (\beta, a) \in \bar{\mathcal{C}}_{(\beta,a)}(\alpha)\}$, we have

$$\mathbb{P}\left[g(\beta, a) \in g[\bar{\mathcal{C}}_{(\beta,a)}(\alpha)\right] \geq \mathbb{P}[(\beta, a) \in \bar{\mathcal{C}}_{(\beta,a)}(\alpha)] \geq 1 - \alpha. \tag{3.13}$$

On taking $g(\beta, a) = a \in \mathbb{R}^G$, we see that

$$\begin{aligned} \mathcal{C}_a(\alpha) &= \{a : (\beta, a) \in \bar{\mathcal{C}}_{(\beta,a)}(\alpha) \text{ for some } \beta\} \\ &= \{a : \bar{Q}(\beta + a) \leq 0 \text{ and } Q(\beta) \leq 0 \text{ for some } \beta\} \end{aligned}$$

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 (3.8) reduce to scalars, and the different confidence sets take the following simple forms:

$$\mathcal{C}_\beta(\alpha_1) = \left\{\beta : A\beta^2 + b\beta + c \leq 0\right\}, \quad \mathcal{C}_\theta(\alpha_2) = \{\theta : \bar{A}\theta^2 + \bar{b}\theta + \bar{c} \leq 0\}, \tag{3.14}$$

$$\mathcal{C}_a(\alpha) = \{a : A\beta^2 + b\beta + c \leq 0, \bar{A}a^2 + (\bar{b} + 2\bar{A}\beta)a + [\bar{c} + \bar{b}\beta + \bar{A}\beta^2] \leq 0\}. \tag{3.15}$$

Closed forms for the sets $\mathcal{C}_\beta(\alpha_1)$ and $\mathcal{C}_\theta(\alpha_2)$ 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 $\mathcal{C}_a(\alpha)$ can be obtained by finding the roots of $\bar{A}a^2 + \bar{b}(\beta)a + \bar{c}(\beta) = 0$, where $\bar{b}(\beta) = \bar{b} + 2\bar{A}\beta$ and $\bar{c}(\beta) = \bar{c} + \bar{b}\beta + \bar{A}\beta^2$ for each $\beta \in \mathcal{C}_\beta(\alpha_1)$.

We 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 to $\mathcal{C}_a(\alpha)$, which yields the confidence set

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

However, it will be more efficient to exploit the linear structure of model (3.15), which allows us 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 $\mathcal{C}_\beta(\alpha_1)$ and $\mathcal{C}_\theta(\alpha_2)$ as

$$\begin{aligned} \mathcal{C}_{w'\beta}(\alpha_1) \equiv g_w[\mathcal{C}_\beta(\alpha_1)] &= \{x_1 : x_1 = w'\beta, \ Q(\beta) \leq 0\} \\ &= \{x_1 : x_1 = w'\beta, \ \beta' A\beta + b'\beta + c \leq 0\}, \end{aligned}$$

where $A$, $b$, and $c$ are defined as in (3.5). For $w'\theta$, we can use a $t$-type confidence interval based on $t(\gamma_0)$:

$$\bar{\mathcal{C}}_{w'\theta}(\alpha_2) \equiv \bar{g}_w[\mathcal{C}_\theta(\alpha_2)] = \{\gamma_0 : |t_{w'\theta}(\gamma_0)| < c_w(\alpha_2)\} = \{\gamma_0 : |w'\hat{\theta} - \gamma_0| < \bar{D}(\alpha_2)\}. \quad (3.16)$$

Here, $\bar{D}(\alpha_2) = c_w(\alpha_2)\,\hat{\sigma}(w'\hat{\theta})$, $\hat{\sigma}(w'\hat{\theta}) = s[w'(Y'MY)^{-1}w]^{1/2}$ with $s = (y'M(Z)y)^{1/2}/(T - G - k)^{1/2}$, and $c_w(\alpha_2)$ is the critical value for a test with level $\alpha_2$ based on $t_{w'\theta}(\gamma_0)$ (in (3.9)). Setting

$$\mathcal{C}_{(w'\beta, w'\theta)}(\alpha_1, \alpha_2) = \{(x, y)' : x \in \mathcal{C}_{w'\beta}(\alpha_1) \text{ and } y \in \bar{\mathcal{C}}_{w'\theta}(\alpha_2)\}, \quad (3.17)$$

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

$$\mathbb{P}[(w'\beta, w'\theta) \in \mathcal{C}_{(w'\beta, w'\theta)}(\alpha_1, \alpha_2)] = \mathbb{P}[w'\beta \in \mathcal{C}_{w'\beta}(\alpha_1) \text{ and } w'\theta \in \bar{\mathcal{C}}_{w'\theta}(\alpha_2)] \geq 1 - \alpha, \quad (3.18)$$

where $\alpha = \alpha_1 + \alpha_2$. For any point $x \in \mathbb{R}$ and any subset $A \subseteq \mathbb{R}$, set $x - A = \{z \in \mathbb{R} : z = x - y \text{ and } y \in A\}$. Because $w'a = w'\theta - w'\beta$, it is clear that

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

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

Now, consider the set

$$\mathcal{C}_{w'a}(\alpha_1, \alpha_2) = \{z \in \mathbb{R} : z \in y - \mathcal{C}_{w'\beta}(\alpha_1) \text{ for some } y \in \bar{\mathcal{C}}_{w'\theta}(\alpha_2)\}. \quad (3.19)$$

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

$$\mathbb{P}[w'a \in \mathcal{C}_{w'a}(\alpha_1, \alpha_2)] \geq \mathbb{P}[w'\beta \in \mathcal{C}_{w'\beta}(\alpha_1) \text{ and } w'\theta \in \bar{\mathcal{C}}_{w'\theta}(\alpha_2)] \geq 1 - \alpha_1 - \alpha_2, \quad (3.20)$$

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

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

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

where $\tilde{\beta} = -(1/2)A^{-1}b$, $d = (1/4)b'A^{-1}b - c$, and $D(\alpha_1) = \sqrt{d\,(w'A^{-1}w)}$. If $A$ has exactly one negative eigenvalue and $d < 0$,

$$\begin{aligned}\mathcal{C}_{w'\beta}(\alpha_1) &= ]-\infty, w'\tilde{\beta} - D(\alpha_1)] \cup [w'\tilde{\beta} + D(\alpha_1), +\infty[, &&\text{if } w'A^{-1}w < 0, \\ &= \mathbb{R}\backslash\{w'\tilde{\beta}\}, &&\text{if } w'A^{-1}w = 0;\end{aligned} \quad (3.21)$$

otherwise, $\mathcal{C}_{w'\beta}(\alpha_1) = \mathbb{R}$. Here, $\mathcal{C}_{w'\beta}(\alpha_1) = \emptyset$ corresponds to a case where the model is not consistent with the data (so that $\mathcal{C}_{w'a}(\alpha_1, \alpha_2) = \emptyset$ as well), while $\mathcal{C}_{w'\beta}(\alpha_1) = \mathbb{R}$ and $\mathcal{C}_{w'\beta}(\alpha_1) = \mathbb{R}\backslash\{w'\tilde{\beta}\}$ indicate that $w'\beta$ is not identifiable and, similarly, for $w'a$ (so that

$\mathcal{C}_{w'a}(\alpha_1, \alpha_2) = \mathbb{R}$). This yields the following confidence sets for $w'a$. If $A$ is positive definite,

$$\begin{aligned}\mathcal{C}_{w'a}(\alpha_1, \alpha_2) &= [w'(\hat{\theta} - \tilde{\beta}) - D_U(\alpha_1, \alpha_2), w'(\hat{\theta} - \tilde{\beta}) + D_U(\alpha_1, \alpha_2)], \quad \text{if } d \geq 0,\\ &= \emptyset, \quad \text{if } d < 0,\end{aligned} \tag{3.22}$$

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 $d < 0$,

$$\mathcal{C}_{w'a}(\alpha_1, \alpha_2) = ]-\infty, w'(\hat{\theta} - \tilde{\beta}) - D_L(\alpha_1, \alpha_2)] \cup [w'(\hat{\theta} - \tilde{\beta}) + D_L(\alpha_1, \alpha_2), +\infty[, \tag{3.23}$$

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

### *3.5. Exact Monte Carlo identification-robust tests with non-Gaussian errors*

Suppose now that the conditional distribution of $\upsilon$ (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 $\alpha$ $(0 < \alpha < 1)$.

STEP 1. Choose $\alpha^*$ and $N$ so that $\alpha = \big(I[\alpha^* N] + 1\big)/(N + 1)$.
STEP 2. For given $\beta_0$, compute the test statistic $AR^{(0)}(\beta_0)$ based on the observed data.
STEP 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 (3.4). Note that the distribution of $AR(\beta_0)$ does not depend on the specific value $\beta_0$ tested, so there is no need to make it depend on $\beta_0$.
STEP 4. Compute the simulated $p$-value function: $\hat{p}_N[x] = (1 + \sum_{j=1}^{N} 1[AR^{(j)} \geq x])/(N + 1)$, where $1[C] = 1$ if condition $C$ holds, and $1[C] = 0$ otherwise.
STEP 5. Reject the null hypothesis $H_\beta(\beta_0)$ at level $\alpha$ when $\hat{p}_N[AR^{(0)}(\beta_0)] \leq \alpha$.

Under the null hypothesis $H_\beta(\beta_0)$, $\mathbb{P}\big[\hat{p}_N[AR^{(0)}(\beta_0)] \leq \alpha\big] = \alpha$, so that we have a test with level $\alpha$. If the distribution of the test statistic is not continuous, the Monte Carlo test procedure can easily be adapted by using the tie-breaking method described by Dufour (2006).[4] Correspondingly, a confidence set with level $1 - \alpha$ for $\beta$ is given by the set of all values $\beta_0$, which are not rejected by the above Monte Carlo test. More precisely, the set

$$\mathcal{C}_\beta(\alpha) = \{\beta_0 : \hat{p}_N[AR^{(0)}(\beta_0)] > \alpha\} \tag{3.24}$$

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

$$\mathcal{C}_\beta(\alpha; N) = \{\beta_0 : AR^{(0)} < \hat{f}_N(\alpha^*)\} \tag{3.25}$$

is equivalent to $\mathcal{C}_\beta(\alpha)$ – with probability one – and so has level $1 - \alpha$. On replacing $>$ and $<$ by $\geq$ and $\leq$ in (3.24) and (3.25), it is also clear that the sets $\big\{\beta_0 : \hat{p}_N[AR^{(0)}(\beta_0)] \geq \alpha\big\}$ and

$$\bar{\mathcal{C}}_\beta(\alpha; N) = \{\beta_0 : AR^{(0)}(\beta_0) \leq \hat{f}_N(\alpha^*)\} \tag{3.26}$$

[4] Without the 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 than the nominal level $(\alpha_1)$).

constitute confidence sets for $\beta$ with level $1 - \alpha$ (though possibly a little larger than $1 - \alpha$). The quadric form given in (3.5) also remains valid with $f(\alpha) = \hat{f}_N(\alpha^*)$.

## 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 (3.2) and (2.15) 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 assumption.

ASSUMPTION 4.1. *When the sample size $T$ converges to infinity, the following convergence results hold jointly: (a) $(1/T)X'u \xrightarrow{p} 0$; (b) $(1/T)u'u \xrightarrow{p} \sigma_u^2 > 0$, $(1/T)X'X \xrightarrow{p} \Sigma_X$ with $\det(X'X) \neq 0$; (c) $(1/\sqrt{T})X'u \xrightarrow{d} \psi_{Xu}$, $\psi_{Xu} \sim N[0,\ \sigma_u^2 \Sigma_X]$, where $X = [X_1, X_2]$.*

The above conditions are easy to interpret: (a) represents the asymptotic orthogonality between $u$ and the instruments in $X$, (b) can be viewed as the laws of large numbers for $u$ and $X$, while (c) is a central limit property. Then, it is a simple exercise to see that

$$AR(\beta_0) \xrightarrow{d} \chi^2(k_2)/k_2, \quad \text{when} \quad \beta = \beta_0. \tag{4.1}$$

Similarly, for $F_\theta(\theta_0)$, we can suppose the following.

ASSUMPTION 4.2. *When the sample size $T$ converges to infinity, the following convergence results hold jointly: (a) $(1/T)Z'e \xrightarrow{p} 0$; (b) $(1/T)e'e \xrightarrow{p} \sigma_e^2 > 0$, $(1/T)Z'Z \xrightarrow{p} \Sigma_Z$ with $\det(Z'Z) \neq 0$; (c) $(1/\sqrt{T})Z'e \xrightarrow{d} \psi_{Ze}$, $\psi_{Ze} \sim N[0, \sigma_e^2 \Sigma_Z]$, where $Z = [Y, X_1, X_2]$.*

Then

$$F_\theta(\theta_0) \xrightarrow{d} \chi^2(G)/G, \quad \text{when} \quad \theta = \theta_0. \tag{4.2}$$

The asymptotic distributions in equations (4.1) and (4.2) hold irrespective whether the instruments $X$ are weak or strong. Furthermore, as soon as Assumptions 4.1 and 4.2 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 the $1 - \alpha_1$ and $1 - \alpha_2$ quantiles, respectively, of the corresponding $\chi^2$ distributions. Of course, the Gaussian-based Fisher critical values can also be used (because they converge to the $\chi^2$ critical values as $T \to \infty$).

We can now consider inference for covariance endogeneity parameters $\sigma_{Vu}$. The problem of building confidence sets for $\sigma_{Vu}$ is especially important for assessing partial exogeneity hypotheses. Because $a_j = 0$, $j = 1, \ldots, G$ does not entail $\sigma_{uVj} = 0$ (where $1 \leq j \leq G$), confidence sets on the components of $a$ cannot directly be used to assess, for example, the exogeneity of each regressor $Y_j$, $j = 1, \ldots, G$. Confidence sets and tests for $\sigma_{uV}$ can be deduced from those on $a$ through the relationship $\sigma_{Vu} = \Sigma_V a$ given in (2.9). On replacing $a$ by $\Sigma_V^{-1}\sigma_{Vu}$

in $\mathcal{C}_a(\alpha)$, we see that the set

$$\mathcal{C}_{\sigma_{Vu}}(\alpha; \Sigma_V) = \{\sigma_{Vu} \in \mathbb{R}^G : \sigma_{Vu} = \Sigma_V a \quad \text{and} \quad a \in \mathcal{C}_a(\alpha)\}$$
$$= \{\sigma_{Vu} \in \mathbb{R}^G : \bar{Q}(\beta + \Sigma_V^{-1}\sigma_{Vu}) \leq 0 \quad \text{and} \quad Q(\beta) \leq 0 \text{ for some } \beta\} \quad (4.3)$$

is a confidence set with level $1 - \alpha$ for $\sigma_{Vu}$. This set is simply the image of $\mathcal{C}_a(\alpha)$ by the linear transformation $g(x) = \Sigma_V x$. The difficulty here comes from the fact that $\Sigma_V$ is unknown. Let $\hat{\Sigma}_V = \hat{V}'\hat{V}/(T - k)$, 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

$$\hat{\Sigma}_V \xrightarrow{p} \Sigma_V, \quad (4.4)$$

where $\det(\Sigma_V) > 0$. If $\beta_0$ and $a_0$ are the true values of $\beta$ 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

$$F_\theta(\beta_0 + \Sigma_V^{-1}\sigma_{Vu0}) = \frac{(\hat{\theta} - \beta_0 - \Sigma_V^{-1}\sigma_{Vu0})'(Y'MY)(\hat{\theta} - \beta_0 - \Sigma_V^{-1}\sigma_{Vu0})/G}{y'M(Z)y/(T - G - k)}. \quad (4.5)$$

Replacing $\Sigma_V$ by $\hat{\Sigma}_V$, we obtain the approximate pivotal function $F_\theta(\beta_0 + \hat{\Sigma}_V^{-1}\sigma_{Vu0})$. If (4.4) 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 non-degenerate distribution, when $\beta_0$ and $\sigma_{Vu0}$ are the true parameter values. Consequently, the confidence set of type $\mathcal{C}_{\sigma_{Vu}}(\alpha)$ 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 $\mathcal{C}_a(\alpha)$ by the linear transformation $\hat{g}(x) = \hat{\Sigma}_V x$, that is

$$\mathcal{C}_{\sigma_{Vu}}(\alpha; \hat{\Sigma}_V) = \{\sigma_{Vu} \in \mathbb{R}^G : \bar{Q}(\beta + \hat{\Sigma}_V^{-1}\sigma_{Vu}) \leq 0 \quad \text{and} \quad Q(\beta) \leq 0 \text{ for some } \beta\}. \quad (4.6)$$

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

## 5. EMPIRICAL APPLICATIONS

We now apply the methods proposed to two empirical examples: a model of the relation between trade and economic growth, previously studied by Frankel and Romer (1999) and Dufour and Taamouti (2007), and the model of returns to education studied by Card (1995) and Kleibergen (2004, Table 2, p. 421).

### *5.1. Trade and growth*

The trade and growth model studies the relationship between standards of living and openness. Frankel and Romer (1999) have argued that trade share (the ratio of imports or exports to GDP), which is the commonly used indicator of openness, might be endogenous. The equation studied is given by

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

where Income is the income per capita, Trade is measured as a ratio of imports and exports to GDP, Pop is the logarithm of the country population, and Area is the logarithm of the country area. The instrument suggested is constructed on the basis of geographic characteristics. The first stage equation is then given by

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

where $Z_i$ is a constructed instrument. We use the sample of 150 countries and the data are for 1985. Dufour and Taamouti (2005) have shown that the fitted instrument in this sample is not very weak.[5]

The identification-robust confidence intervals with level 97.5% for $\beta$ and $\theta = \beta_1 + a$, which result on inverting $AR(\beta_0)$ and $t_\theta(\gamma_0)$, are given by $\mathcal{C}_\beta(\alpha) = \{\beta_0 : 0.23\beta_0^2 - 4.76\beta_0 + 0.04 \leq 0\} = [0.01, 20.62]$ and $\mathcal{C}_\theta(\alpha) = [-0.05, 0.47]$. The results reported are based on the critical values of the $F$-distributions of Section 3. The Monte Carlo method, as described in Section 3.5, gives similar results even with 1000 replications. We see that $\mathcal{C}_\beta(\alpha)$ is a bounded interval, thus confirming that identification is not weak in this model. The estimates of regression and covariance endogeneity parameters are given by $\hat{a} = -1.82$ and $\hat{\sigma}_{uV} = -0.38$, respectively. The confidence intervals with level 95% for $a$ and $\sigma_{Vu}$ are given by[6]

$$\mathcal{C}_a(\alpha) = [-20.67, 0.46] \quad \text{and} \quad \mathcal{C}_{\sigma_{Vu}}(\alpha) = [-4.33, 0.09].$$

Both confidence intervals are bounded and contain the estimates of $a$ and $\sigma_{Vu}$ from observed data. Both confidence intervals, although including zero, are left skewed at zero. In particular, the upper bound for $\mathcal{C}_{\sigma_{Vu}}(\alpha)$ is very close to zero. So, the true covariance and regression endogeneity parameters can actually be large, thus indicating the importance of omitting variables bias (for $a$) and trade share endogeneity (for $\sigma_{Vu}$). The latter is likely plausible because the discrepancy between the OLS estimate of $\beta$ ($\hat{\beta}_{\text{OLS}} = 0.28$) and the two-stage least-squares (2SLS) estimate ($\hat{\beta}_{\text{2SLS}} = 2.03$) is relatively large.

### 5.2. *Card model of education and earnings*

We also apply the methods proposed to the following alternative model studied by Card (1995) for the return of education to earnings:

$$y_i = Y_{1i}\beta_1 + Y_{2i}\beta_2 + Y_{3i}\beta_3 + X_{1i}'\gamma + u_i; \tag{5.3}$$

$$(Y_{1i},\ Y_{2i},\ Y_{3i}) = X_{1i}'\Pi_1 + X_{2i}'\Pi_2 + V_i. \tag{5.4}$$

Here, $Y_{1i}$ is the length of education of individual $i$, $(Y_{2i}, Y_{3i}) = (exper_i, exper_i^2)$ contains the experience ($exper$) and experience squared of individual $i$, where $exper_i = age_i - 6 - Y_{1i}; X_{1i} = (1, race_i, smsa_i, south_i)'$ consists of a constant and indicator variables for race, residence in a metropolitan area, and residence in the south of the United States, and $y_i$ is the logarithm of the wage of individual $i$. All variables in $X_1$ are assumed exogenous. $X_{2i}$ is the vector of instruments that contains $age$, $age^2$ of individual $i$, and proximity-to-college indicators for educational attainment; these are proximity to two- and four-year college. Kleibergen (2004,

[5] The $F$-statistic in the first stage (5.2) is about 13; see also Frankel and Romer (1999, Table 2, p. 385).

[6] Note that the confidence intervals with level 95% for $a$ and $\sigma_{Vu}$, obtained on inverting $AR(\beta_0)$ and $F_\theta(\theta_0)$, are similar to those reported here.

Table 2, p. 421) shows that the proximity-to-college indicator instruments are not very strong. Hence, it is important to be careful when interpreting the 2SLS estimates of this model. We follow the methodology developed in this paper for building projection-based confidence intervals of the components of the regression and covariance endogeneity parameters $a = (a_1, a_2, a_3)'$ and $\sigma_{Vu} = (\sigma_{Vu1}, \sigma_{Vu2}, \sigma_{Vu3})'$.

The data analysed are from the National Longitudinal Survey of Young Men (from 1966 to 1981). We use the cross-sectional 1976 subsample, which contains 3010 observations. After accounting for missing data, the final sample has 2061 observations. The variables contained in the data set are: two variables indicating the proximity to college, the length of education, log wages, experience, IQ score, age, racial, metropolitan, family, and regional indicators.

To build confidence sets with level 95% for $a$ and $\sigma_{Vu}$, we take $\alpha_1 = \alpha_2 = 0.025$. The identification-robust confidence sets with level 97.5% for $\beta = (\beta_1, \beta_2, \beta_3)'$ and $\theta = \beta + a$, based on inverting $AR(\beta_0)$ and $F_\theta(\theta_0)$, are given by $\mathcal{C}_\beta(\alpha) = \{\beta_0 : \beta_0' A \beta_0 - b'\beta_0 + 0.37 \le 0\}$ and $\mathcal{C}_\theta(\alpha) = \{\theta_0 : \theta_0' \bar{A} \theta_0 + \bar{b}'\theta_0 + 0.63 \le 0\}$, where

$$A = \begin{pmatrix} 0.7 & 6.17 & 87.34 \\ 6.14 & 170.88 & 3210.82 \\ 87.34 & 3210.82 & 61730.62 \end{pmatrix}, \quad \bar{A} = \begin{pmatrix} 770.72 & -770.70 & -13287.73 \\ -770.70 & 770.72 & 13287.70 \\ -13287.73 & 13287.70 & 270277.74 \end{pmatrix}, \tag{5.5}$$

$b = (-0.8, -15.62, -285.9)'$ and $\bar{b} = (-33.59, 33.59, 838.17)'$. The matrix $A$ has exactly one negative eigenvalue, while all eigenvalues of $\bar{A}$ are positive. Hence, $\mathcal{C}_\beta(\alpha)$ is an unbounded ellipsoid, while $\mathcal{C}_\theta(\alpha)$ is a bounded ellipsoid, thus confirming that $\theta$ is identified while $\beta$ is not. Then, for any scalar linear transformations $w'\theta$, a confidence set with level $1 - \alpha_2$ is given by (3.16) with $\hat{\theta} = (0.279, 0.312, -0.003)$ and $\bar{D}(\alpha_2) = 0.72[w'(Y'MY)^{-1}w]^{1/2}$. For $w'\beta$, we can obtain a projection-based confidence set with level $1 - \alpha_1$ by using (3.21) with $\tilde{\beta} = (-0.361, 0.218, -0.010)$, $d = -1.55 < 0$ and $D(\alpha_1) = [-1.55 w'A^{-1}w]^{1/2}$ when $w'A^{-1}w < 0$. For inference on $a$, we also use the following estimates:[7]

$$\hat{a} = \begin{pmatrix} -0.102 \\ 0.102 \\ -0.004 \end{pmatrix}, \quad \hat{\sigma}_{Vu} = \begin{pmatrix} -0.492 \\ 0.492 \\ 7.634 \end{pmatrix}, \quad \hat{\Sigma}_V = \begin{pmatrix} 3.76 & -3.75 & -64.75 \\ -3.75 & 3.74 & 64.76 \\ -64.75 & 64.76 & 1317.14 \end{pmatrix}.$$

The 2SLS estimate of $\beta$ is $\hat{\beta}_{2SLS} = (0.190, 0.019, 0.001)'$, and the eigenvalues of $\hat{\Pi}_2'\hat{\Pi}_2$, where $\hat{\Pi}_2$ is the OLS estimate of $\Pi_2$ from (5.4), are (0.0003, 0.095, 3858.326). The value 0.0003 is quite close to zero, which suggests that the instruments are weak.

Table 1 presents the projection-based confidence intervals with level 95% for individual components of endogeneity parameters ($a$ and $\sigma_{Vu}$). In the first part of the table, the $IQ$ variable is omitted from the model, but it is included in the second part. The results are similar with and without this variable: the confidence intervals for all components of $a$ and $\sigma_{Vu}$ are unbounded. So, all components of both endogeneity parameters are weakly identified. While the estimate of $a_3$ ($\hat{a}_3 = -0.004$) seems very close to zero, the corresponding covariance estimate $\hat{\sigma}_{Vu3} = 7.634$ is relatively large, which confirms the fact that $a_i = 0$ does not necessarily imply that $\sigma_{Vui} = 0$, as argued in Section 2.1. All confidence intervals, though unbounded, contain zero, suggesting that there is not enough information from the data to support the presence of bias as a result of omitted

[7] The results reported are based on the critical values of the $F$-distributions of Section 3. The Monte Carlo method as described in Section 3.5 gives similar results even with 1000 replications, for both (1) Gaussian errors, and (2) Student type errors with three degrees of freedom.

**Table 1.** Card model of education and earnings.

| Regression endogeneity | | Covariance endogeneity | |
|---|---|---|---|
| Without $IQ$ variable | | | |
| $\mathscr{C}_{a_1}$ | $]-\infty, 0.47] \cup [1.45, +\infty[$ | $\mathscr{C}_{\sigma_{Vu1}}$ | $]-\infty, 0.41] \cup [9.08, +\infty[$ |
| $\mathscr{C}_{a_2}$ | $]-\infty, -0.12] \cup [-0.03, +\infty[$ | $\mathscr{C}_{\sigma_{Vu3}}$ | $]-\infty, -9.08] \cup [-0.41, +\infty[$ |
| $\mathscr{C}_{a_3}$ | $]-\infty, 0.002] \cup [0.03, +\infty[$ | $\mathscr{C}_{\sigma_{Vu3}}$ | $]-\infty, -165.35] \cup [-7.65, +\infty[$ |
| With $IQ$ variable | | | |
| $\mathscr{C}_{a_1}$ | $]-\infty, 0.55] \cup [0.73, +\infty[$ | $\mathscr{C}_{\sigma_{Vu1}}$ | $]-\infty, 0.24] \cup [3.19, +\infty[$ |
| $\mathscr{C}_{a_2}$ | $\mathbb{R}$ | $\mathscr{C}_{\sigma_{Vu3}}$ | $]-\infty, -3.19] \cup [-0.24, +\infty[$ |
| $\mathscr{C}_{a_3}$ | $]-\infty, 0.001] \cup [0.013, +\infty[$ | $\mathscr{C}_{\sigma_{Vu3}}$ | $]-\infty, -52.05] \cup [-4.37, +\infty[$ |

variables (regression endogeneity parameters $a_i$, $i = 1, 2, 3$, measure the importance of omitted variables) or to reject the partial exogeneity of the *schooling* and *experience* variables (covariance endogeneity parameters $\sigma_{Vui}$, $i = 1, 2, 3$, measure the endogeneity of the corresponding variable $Y_i$). Meanwhile, though zero belongs to the 95% confidence intervals of all these parameters, it might be the case that the true values of these parameters are actually large, because the 95% corresponding confidence intervals are unbounded. So, the use of the standard $t$-type statistics based on the estimates of $a$ and $\sigma_{Vu}$ in the extended regression (2.14), where $V$ is replaced by $\hat{V} = MY$, to build confidence intervals for scalar linear transformations $w'a$ and $w'\sigma_{Vu}$ can be misleading when identification is weak. The Monte Carlo simulations indicate that such $t$-type confidence intervals have poor coverage probabilities (which might even be equal to zero) when identification is weak, while the coverage probabilities of the projection method developed in this paper are always above $1 - \alpha$, irrespective of whether identification is strong or weak, where $\alpha$ is the nominal level.

## ACKNOWLEDGEMENTS

The authors are grateful to Professor Elie Tamer, Co-Editor of the *Econometrics Journal*, and to two anonymous referees for constructive comments and suggestions. The authors also thank Marine Carrasco, Jan Kiviet, Benoit Perron, and Purevdorj Tuvaandorj for several useful comments. This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), the Toulouse School of Economics (Pierre-de-Fermat Chair of Excellence), the Universitad Carlos III de Madrid (Banco Santander de Madrid Chair of Excellence), 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).

## REFERENCES

Abdelkhalek, T. and J.-M. Dufour (1998). Statistical inference for computable general equilibrium models with applications to a model of the Moroccan economy. *Review of Economics and Statistics 80*, 520–34.

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

Barro, R. (1977). Unanticipated money growth and unemployment in the United States. *American Economic Review 67*, 101–15.

Card, D. (1995). Using geographic variation in college proximity to estimate the return to schooling. In L. N. Christofides, E. K. Grant, and R. Swidinsky (Eds.), *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, 201–22. Toronto: University of Toronto Press.

Doko Tchatoka, F. and J.-M. Dufour (2011a). Exogeneity tests and estimation in IV regressions. Working paper, Department of Economics, McGill University, Montréal, Canada.

Doko Tchatoka, F. and J.-M. Dufour (2011b). On the finite-sample theory of exogeneity tests with possibly non-Gaussian errors and weak identification. Working paper, Department of Economics, McGill University, Montréal, Canada.

Dufour, J.-M. (1979). *Methods for Specification Errors Analysis with Macroeconomic Applications*. Unpublished Ph.D. thesis, University of Chicago.

Dufour, J.-M. (1987). Linear Wald methods for inference on covariances and weak exogeneity tests in structural equations. In I. B. MacNeill and G. J. Umphrey (Eds.), *Advances in the Statistical Sciences: Festschrift in Honour of Professor V.M. Joshi's 70th Birthday. Time Series and Econometric Modelling, Volume III*, 317–38. Dordrecht: Reidel.

Dufour, J.-M. (1990). Exact tests and confidence sets in linear regressions with autocorrelated errors. *Econometrica 58*, 475–94.

Dufour, J.-M. (1997). Some impossibility theorems in econometrics, with applications to structural and dynamic models. *Econometrica 65*, 1365–89.

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

Dufour, J.-M. (2006). Monte Carlo tests with nuisance parameters: a general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics 133*, 443–77.

Dufour, J.-M. and C. Hsiao (2008). Identification. In L. E. Blume and S. N. Durlauf (Eds.), *The New Palgrave Dictionary of Economics*, 2nd edn. Basingstoke: Palgrave Macmillan.

Dufour, J.-M. and J. Jasiak (2001). Finite sample limited information inference methods for structural equations and models with generated regressors. *International Economic Review 42*, 815–43.

Dufour, J.-M. and M. Taamouti (2005). Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica 73*, 1351–65.

Dufour, J.-M. and M. Taamouti (2007). Further results on projection-based inference in IV regressions with weak, collinear or missing instruments. *Journal of Econometrics 139*, 133–53.

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.

Frankel, J. A. and D. Romer (1999). Does trade cause growth? *American Economic Review 89*, 379–99.

Hausman, J. (1978). Specification tests in econometrics. *Econometrica 46*, 1251–72.

Hausman, J. and W. E. Taylor (1981). A generalized specification test. *Economics Letters 8*, 239–45.

Hwang, H-S. (1980). Test of independence between a subset of stochastic regressors and disturbances. *International Economic Review 21*, 749–60.

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

Kiviet, J. F. (2013). Identification and inference in a simultaneous equation under alternative information sets and sampling schemes. *Econometrics Journal 16*, S24–S59.

Kiviet, J. F. and J. Niemczyk (2007). The asymptotic and finite-sample distributions of OLS and simple IV in simultaneous equations. *Computational Statistics and Data Analysis 51*, 3296–318.

Kiviet, J. F. and J. Niemczyk (2012). Comparing the asymptotic and empirical (un)conditional distributions of OLS and IV in a linear static simultaneous equation. *Computational Statistics and Data Analysis 56*, 3567–86.

Kiviet, J. F. and M. Pleus (2012). The performance of tests on endogeneity of subsets of explanatory variables scanned by simulation. Working paper, Amsterdam School of Economics, Amsterdam, The Netherlands.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica 70*, 1781–803.

Kleibergen, F. (2004). Testing subsets of structural coefficients in the IV regression model. *Review of Economics and Statistics 86*, 418–23.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica 71*, 1027–48.

Nakamura, A. and M. Nakamura (1981). On the relationships among several specification error tests presented by Durbin, Wu and Hausman. *Econometrica 49*, 1583–88.

Revankar, N. S. (1978). Asymptotic relative efficiency analysis of certain tests in structural sysytems. *International Economic Review 19*, 165–79.

Revankar, N. S. and M. J. Hartley (1973). An independence test and conditional unbiased predictions in the context of simultaneous equation systems. *International Economic Review 14*, 625–31.

Smith, R. J. (1983). On the classical nature of the Wu–Hausman statistics for independence of stochastic regressors and disturbance. *Economics Letters 11*, 357–64.

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*, 263–69.

Smith, R. J. (1985). Wald tests for the independence of stochastic variables and disturbance of a single linear stochastic simultaneous equation. *Economics Letters 17*, 87–90.

Spencer, D. E. and K. N. Berk (1981). A limited-information specification test. *Econometrica 49*, 1079–85. Erratum (1982). *Econometrica 50*, 1087.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*, 557–86.

Stock, J. H., J. H. Wright and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics 20*, 518–29.

Wu, D-M. (1973). Alternative tests of independence between stochastic regressors and disturbances. *Econometrica 41*, 733–50.

Wu, D-M. (1974). Alternative tests of independence between stochastic regressors and disturbances: finite sample results. *Econometrica 42*, 529–46.

Wu, D-M. (1983). A remark on a generalized specification test. *Economics Letters 11*, 365–70.

## APPENDIX A: PROOF OF RESULTS

**Proof of Lemma 3.1:** On multiplying the two sides of equation (3.2) by $M$ and $M_1 - M$, we see that

$$M(y - Y\beta_0) = Mu + MV(\beta - \beta_0), \tag{A.1}$$

$$(M_1 - M)(y - Y\beta_0) = M_1 X_2 \Pi_2(\beta - \beta_0) + (M_1 - M)u + (M_1 - M)V(\beta - \beta_0).$$

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

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

Thus, the $AR$-statistic in equation (3.3) 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 $\upsilon$ and $X$. If normality holds conditional on $X$ (i.e., $\upsilon|X \sim \mathrm{N}[0, , I_T]$), we have $\upsilon' M\upsilon \sim \chi^2(T - k)$ and $\upsilon'(M_1 - M)\upsilon \sim \chi^2(k_2)$. Because $M(M_1 - M) = 0$, hence $\upsilon' M\upsilon$ and $\upsilon'(M_1 - M)\upsilon$ are independent conditional on $X$. Consequently, $AR(\beta_0) \sim F(k_2, T - k)$. □