Identification-robust inference for endogeneity parameters in models with an incomplete reduced form [a]

Jean-Marie Dufour [b]
McGill University

Vinh Nguyen [c]
McGill University

December 2020

[a] The authors thank two anonymous refereees and the Editor Cheng Hsiao 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), 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).

[b] 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 414, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 6071; FAX: (1) 514 398 4800; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

[c] Department of Economics, McGill University. Email: vinh.nguyen3@mail.mcgill.ca

ABSTRACT

We propose inference methods for endogeneity parameters in linear simultaneous equation models allowing for weak identification and missing instruments. Endogeneity parameters measure the impact of unobserved variables which may be correlated with observed explanatory variables, and play a central role in determining the "bias" associated with endogeneity and measurement errors in structural equations. These results expand, in several ways, the finite-sample theory in Doko and Dufour [*Econometrics J.*, 2014] for this problem. The latter theory relies on relatively restrictive assumptions, in particular the hypothesis that the reduced form is complete (*e.g.*, contains all the relevant instruments), which is questionable in many practical situations. While the new proposed inference methods retain identification robustness, they also allow the reduced form to be incomplete, *e.g.* due to missing instruments. We propose easily applicable inference methods for endogeneity parameters – in particular, two-stage procedures [similar to those in Dufour [*Econometrica*, 1990]. An application to a model of returns to schooling is presented.

Keywords: endogeneity; exogeneity; instrumental variables; simultaneous equations; IV regression; missing instruments; identification; identification robust; projection; hypothesis testing; confidence set.

Journal of Economic Literature classification: C01, C12, C26, C3, C36, C52, D1, E2.

MSC2020 classification: 62, 62P20, 62P25, 62F03, 62F05, 62F25.

1 Introduction

In a linear model where some explanatory variables are correlated with the error term, least squares yield inconsistent estimators of model coefficients. One popular solution consists in using instrumental variables (IV), which are uncorrelated with the error term, but correlated with explanatory endogenous variables. However, when this correlation is weak, IV estimation can be very imprecise and yield unreliable tests and confidence sets. This is the weak-instrument problem which has attracted considerable attention [for example, see Bound et al. (1995), Dufour (1997), and Staiger and Stock (1997); for reviews, see Dufour (2003) and Mikusheva (2013)].

An important objective of the weak-instrument literature is the development of identification-robust inference procedures, which remain valid even when the instruments are weak. This literature, however, has focused on the coefficients of the explanatory variables. In contrast, we are interested in the parameters that give rise to the source of endogeneity in the first place: the covariances between the (possibly) endogenous variables and the error term.

As argued in Doko Tchatoka and Dufour (2014), there are several reasons to study *endogeneity parameters*. *First*, these provide a measure of the importance of latent variables, which are unobserved but can influence the (observed) endogenous variables. *Second*, they give information on the estimation bias of least-squares estimation. Understanding this bias is helpful in interpreting least-squares estimated and related statistics. *Third*, knowledge of the degree of endogeneity can help the investigator select the appropriate estimation method: it is known that IV can be less efficient than least-squares when endogeneity is small, and this is true even when the instruments are strong [see Kiviet and Niemczyk (2012), Doko Tchatoka and Dufour (2014), and Kiviet and Pleus (2017)].

In this paper, we develop *three new* inference approaches for the covariances between the endogenous variables and the error term. All three approaches are robust to missing or unobserved instruments. This constitutes an important extension of the previous work Doko Tchatoka and Dufour (2014) which also takes interest in endogeneity parameters, but assumes that the reduced-form equation is complete (*i.e.* includes all relevant instruments). This latter assumption is suspect in many practical situations.

(1) The first approach (**Section 4**) is asymptotically robust to weak instruments. It is based on the following insight: *if we knew the regression coefficients of the endogenous variables, there would be a natural plug-in estimator for the covariances of the endogenous variables and the error term.* Since we do *not* know the endogenous-variable coefficients, we construct a two-stage inference procedure, using the Anderson and Rubin (1949) test for the first stage and Dufour (1990) for the second stage. As we allow for weak instruments, this approach is an extension of the results of Dufour (1987), whose assumptions exclude weak instruments.

(2) The second approach (**Section 5**) modifies the inference from Doko Tchatoka and Dufour (2014) for the *total effect*, *i.e.* the sum of the direct and the indirect effects of the endogenous variables on the dependent variable. Our modification yields another two-stage procedure which is robust to both unobserved (or missing) and weak instruments. By considering inference from the perspective of the total effect, we extend a procedure in Dufour and Jasiak (2001), who study the total effect in the context of a different model.

(3) The third approach (**Section 6**) is inspired by the literature on exogeneity testing, which focuses on the difference between least-squares and IV estimators; see Durbin (1954), Wu (1973), Revankar and Hartley (1973), Hausman (1978), Doko Tchatoka and Dufour (2017), and Doko Tchatoka and Dufour (2020)]. On observing that this difference can be viewed as a measure of the OLS bias, which is directly related to the endogeneity covariances, we develop an alternative approach to the problem.

The paper is organized as follows. In **Sections 2** and **3**, we state and discuss the assumptions of our framework. The assumptions in Section 2 focus on specification, algebra, and notation, and those in Section 3 are statistical in nature. In **Sections 4 – 6**, we develop the three inference approaches for the covariances between the endogenous variables and the structural error. In **Section 7**, we demonstrate how to implement our theory to returns to education. Finally, in **Section 8**, we summarize the results and discuss some ideas for future work. Proofs are available in an (online) Appendix.[1]

[1] See also the discussion paper version of the article [Dufour and Nguyen (2020)].

2 Framework and notation

In this section, we describe our general framework, and we introduce relevant notation. Asymptotic assumptions are discussed in Section 3.

2.1 Basic structural framework

Assumption 1 (Structural equation)**.**

$$y = Y\beta + X_1\gamma + u\,, \quad \mathbb{E}(u) = 0 \tag{2.1}$$

where y is a $T \times 1$ vector of dependent variables, Y is a $T \times G$ matrix of (possibly) endogenous variables, and X_1 is a $T \times k_1$ matrix of exogenous variables.

Assumption 1 is maintained throughout the paper. Here, the word "exogenous" may mean "strictly exogenous", "weakly exogenous" or "predetermined". In particular, X_1 does not have to be nonstochastic or strictly exogenous. Explicit asymptotic assumptions are stated in Section 3.

We next consider several alternatives for the reduced-form equations for Y in (2.1). Any new notation is immediately clarified after these alternatives.

Assumption 2 (Reduced-form equation: general form I)**.**

$$Y = g(\bar{X},\, \bar{\Pi},\, V). \tag{2.2}$$

Assumption 3 (Reduced-form equation: general form II)**.**

$$Y = g(\bar{X},\, \bar{\Pi}) + V\,. \tag{2.3}$$

Assumption 4 (Linear reduced-form equation I)**.**

$$Y = X_1\Pi_1 + X_2\Pi_2 + X_3\Pi_3 + V\,. \tag{2.4}$$

Assumption 5 (Linear reduced-form equation II)**.**

$$Y = X_1\Pi_1 + X_2\Pi_2 + V\,. \tag{2.5}$$

Above, V is a $T \times G$ matrix of "reduced-form errors", while X_i and Π_i are $T \times k_i$ and $k_i \times G$ matrices of exogenous variables and fixed coefficients, respectively, where $i = 1, 2, 3$. The matrix $\bar{\Pi} = [\Pi_1', \Pi_2', \Pi_3']'$ has dimension $\bar{k} \times G$ and contains unknown coefficients. For convenience, we set:

$$\bar{Y} := [Y,\, X_1]\,, \quad X := [X_1,\, X_2]\,, \quad k := k_1 + k_2\,, \tag{2.6}$$

$$\bar{X} := [X_1, X_2, X_3]\,, \quad \bar{k} := k_1 + k_2 + k_3\,, \tag{2.7}$$

$$\text{for } g \text{ in (2.3)}: \quad \bar{g} := g(\bar{X}, \bar{\Pi}). \tag{2.8}$$

Here, := means "equal by definition".

In Assumptions 2–5, X_2 represents instruments excluded from the structural equation of interest but used for IV estimation, and X_3 represents missing or unobserved instruments. It is clear that Assumptions 2 - 5 go from most general to most specific. The combination of Assumptions 1 and 5 is the most popular one in studies of simultaneous equation models. If (2.5) is assumed but the actual reduced-form equation follows (2.4), this constitutes a misspecification in which the variables in X_3 are omitted. Similarly, if (2.5) holds but the actual reduced-form equation follows (2.2) or (2.3), the misspecification entails *both* the omission of X_3 and arbitrary deviation from the correct form.

To ensure the validity of many algebraic manipulations below, we also make the following assumption. Below, the matrices $P(\cdot)$ and $M(\cdot) = I - P(\cdot)$ yield the orthogonal projections onto and off the column space of the input matrix.

Assumption 6 (Column rank)**.** $\bar{X}$, $[Y, X]$ and $[P(X)Y, X_1]$ have full-column rank with probability one (conditional on X).

In this paper, we focus on inference for endogeneity parameters which capture the covariances between Y and u when Assumption 3 holds. In our framework, the covariances between Y and

u (for each row) are the same as the corresponding covariances between V and u. By allowing missing instruments (X_3), our framework also accommodates various forms of heteroskedasticity.

In this context, identification-robust (IR) inference on β can be achieved by using the Anderson and Rubin (1949) (AR) statistic for testing $\beta = \beta_0$:

$$\mathrm{AR}(\beta_0) := \frac{(y - Y\beta_0)'[M_1 - M](y - Y\beta_0)/k_2}{(y - Y\beta_0)'M(y - Y\beta_0)/(T - k)} \tag{2.9}$$

where $M_1 := M(X_1)$, $M := M(X)$ and $\beta_0 \in \mathbb{R}^G$. However, we focus here on the link between u and Y. For this purpose, we consider the decomposition

$$u = Va + e \tag{2.10}$$

where a (the "regression endogeneity parameter") is a fixed vector, and e is, in some sense, uncorrelated with Y and $\bar{X}$. The coefficient a can represent the effect of V on u which then gets transmitted to y. The unobserved variable V also affects y through Y, so that the "total effect" of V on y is given by

$$\theta := \beta + a\,. \tag{2.11}$$

Under Assumption 5, θ may be estimated by considering the "orthogonalized structural equation"

$$y = Y\beta + X_1\gamma + Va + e = Y\theta + X_1(\gamma - \Pi_1 a) + X_2(-\Pi_2 a) + e \tag{2.12}$$

and values of θ can be tested by the corresponding AR-type F statistics for $\theta = \theta_0$:

$$F(\theta_0) := \frac{(y - Y\theta_0)'[M - M(Z)](y - Y\theta_0)/G}{y'M(Z)y/(T - G - k)} \tag{2.13}$$

where $Z := [Y, X_1, X_2]$. and $\theta_0 \in \mathbb{R}^G$; see Doko Tchatoka and Dufour (2014).

In this paper, we do not wish to be restricted by the assumption of a complete reduced form (Assumption 5). When Assumption 4 or Assumption 3 holds, the strategy above for θ does *not* work. In Section 5, we show how to use F in the presence of missing or unobserved instruments to obtain inference on the covariance between Y and u.

2.2 Notation

We collect here several definitions used throughout the paper. The motivation for some of these will be become clearer when the asymptotic framework will be introduced (Section 3). We start with those that have already appeared but are reproduced here for completeness:

$$
\begin{gathered}
\text{endogenous variables}: \quad y,\, Y \quad (\text{dimensions } T\times 1,\, T\times G)\,;\\
\text{exogenous variables}: \quad X_1,\, X_2,\, X_3 \quad (\text{dimensions } T\times k_1,\, T\times k_2,\, T\times k_3)\,;\\
\bar{Y} = [Y,\ X_1], \quad X = [X_1,\, X_2], \quad \bar{X} = [X_1,\, X_2,\, X_3], \quad Z = (Y,\, X_1,\, X_2)\,;\\
k = k_1 + k_2, \quad \bar{k} = k_1 + k_2 + k_3, \quad \bar{\Pi} = [\Pi_1',\, \Pi_2',\, \Pi_3']'\,. \qquad (2.14)
\end{gathered}
$$

From these, we consider the following projection-based matrices:

$$P_1 := P(X_1), \quad M_1 := M(X_1), \quad P := P(X), \quad M := M(X), \qquad (2.15)$$

$$N_1 := PM_1 = M_1P = M_1 - M = P - P_1; \qquad (2.16)$$

$$B_1 := (Y'M_1Y)^{-1}Y'M_1, \quad B_2 := (Y'N_1Y)^{-1}Y'N_1\,. \qquad (2.17)$$

P_1, M_1, P and M are useful abbreviations, while N_1 and B_2 are used in the second-stage regression in estimating β (as opposed to γ) using IV. B_1 is the OLS counterpart of B_2 when one estimates β using least squares. It is easy to verify

$$B_1Y = I_G = B_2Y, \quad B_1X_1 = 0 = B_2X_1\,. \qquad (2.18)$$

In this context, the OLS estimator ($\hat{\beta}$) and the IV estimator ($\tilde{\beta}$) of β are given by:

$$\hat{\beta} = B_1y, \quad \tilde{\beta} = B_2y\,. \qquad (2.19)$$

We also consider matrices related to the asymptotic covariance matrices of $\hat{\beta}$ and $\tilde{\beta}$:

$$\hat{\Omega}_{\text{OLS}} := \frac{1}{T}Y'M_1Y, \quad \hat{\Omega}_{\text{IV}} := \frac{1}{T}Y'N_1Y \quad \hat{\Omega}_{\text{OLS},\,2} := \frac{1}{T}Y'MY, \tag{2.20}$$

where $\hat{\Omega}_{\text{OLS},\,2}$ corresponds to estimating the (asymptotic) covariance of the OLS estimator of θ in (2.12). Further,

$$C_1 := B_2 - B_1, \quad N_2 := I_T - M_1YB_2, \quad \hat{\Delta} := \hat{\Omega}_{\text{IV}}^{-1} - \hat{\Omega}_{\text{OLS}}^{-1}, \tag{2.21}$$

$$\Lambda_1 := \frac{1}{T}M_1N_2'N_2M_1, \quad \Lambda_2 := \frac{1}{T}M[\bar{Y}] = \frac{1}{T}M_1M(M_1Y)M_1\,. \tag{2.22}$$

The equality of the two equivalent expressions for Λ_2 follows from the Frisch-Waugh-Lovell (FWL) theorem. The matrices C_1 and N_2 are associated with the difference $\tilde{\beta} - \hat{\beta}$:

$$\tilde{\beta} - \hat{\beta} = C_1y = C_1u = -\hat{\Omega}_{\text{OLS}}^{-1}\big[\frac{1}{T}Y'M_1N_2u\big]\,. \tag{2.23}$$

$\hat{\Delta}$ appears in the (asymptotic) covariance matrix of $\tilde{\beta} - \hat{\beta}$, while Λ_1 and Λ_2 are associated with the estimation of the variance of u:

$$\tilde{\sigma}^2 := y'\Lambda_1y, \quad \hat{\sigma}^2 := y'\Lambda_2y, \quad \tilde{\Sigma} := \tilde{\sigma}^2\hat{\Delta}, \quad \kappa := T - k_1 - G\,. \tag{2.24}$$

$\hat{\sigma}^2$ is the OLS estimator of the (structural) error variance. We note also that the IV residuals $y - Y\tilde{\beta} - X_1\tilde{\gamma}$ belong to the column space of M_1 [where $\tilde{\gamma} = (X_1'X_1)^{-1}X_1'(y - PY\tilde{\beta})$ is the IV estimator of γ], hence

$$y - Y\tilde{\beta} - X_1\tilde{\gamma} = M_1(y - Y\tilde{\beta} - X_1\tilde{\gamma}) = M_1(y - Y\tilde{\beta}) = N_2M_1y\,. \tag{2.25}$$

Consequently, $\tilde{\sigma}^2$ gives the IV estimator for the variance of the structural error. We use (2.24) to define

$$\mathscr{T} = \kappa(\tilde{\beta} - \hat{\beta})'\tilde{\Sigma}^{-1}(\tilde{\beta} - \hat{\beta}), \quad \mathscr{H} := (T/\kappa)\mathscr{T}\,. \tag{2.26}$$

The statistic $\mathscr{T}$ was originally proposed by Wu (1973), while $\mathscr{H}$ is a variant of the Hausman (1978) statistic considered by Hahn et al. (2011).

Finally, let $\overset{a}{\sim}$ be the binary relation indicating both terms having the same limiting variable. For any symmetric matrices $\bar{A}$ and $\bar{B}$, define $\bar{A} \succeq \bar{B}$ ($\bar{A} \preceq \bar{B}$) to mean that $\bar{A} - \bar{B}$ is positive (negative) semidefinite. Lastly, $\succ$ and $\prec$ are defined analogously.

3 Asymptotic assumptions

In this section, we describe general asymptotic assumptions for studying the behavior of inference procedures when instruments are missing. We also discuss their meaning and derive a useful dominance lemma.

Assumption 7 (Convergence, I)**.** For $\bar{g} := g(\bar{X}, \bar{\Pi})$ in (2.8), we have:

$$\frac{1}{T}\bar{g}'[u,\, V] \xrightarrow[T\to\infty]{\text{p}} 0, \quad \frac{1}{T}X'[u,\, V] \xrightarrow[T\to\infty]{\text{p}} 0, \quad \frac{1}{T}[u,\, V]'[u,\, V] \xrightarrow[T\to\infty]{\text{p}} \Sigma := \begin{pmatrix} \sigma_u^2 & \sigma_{uV} \\ \sigma_{Vu} & \Sigma_V \end{pmatrix}, \tag{3.1}$$

$$\frac{1}{T}[X,\, \bar{g}]'[X,\, \bar{g}] \xrightarrow[T\to\infty]{\text{p}} \bar{\Sigma} := \begin{pmatrix} \Sigma_1 & \Sigma_{12} & \Sigma_{1g} \\ \Sigma_{21} & \Sigma_2 & \Sigma_{2g} \\ \Sigma_{g1} & \Sigma_{g2} & \Sigma_g \end{pmatrix} = \begin{pmatrix} \Sigma_X & \Sigma_{Xg} \\ \Sigma_{gX} & \Sigma_g \end{pmatrix}, \tag{3.2}$$

where Σ, Σ_X and $\Sigma_{g1}\Sigma_1^{-1}\Sigma_{1g}$ are positive definite.

Assumption 7 entails we can use laws of large numbers in our context. It encompasses cases with a set of "missing" instruments in $\bar{g}$. It is easy to see that Assumption 7 entails the following convergence properties:

$$\Omega_{\text{OLS}} := \text{plim}\, \hat{\Omega}_{\text{OLS}}\,, \quad \Omega_{\text{OLS},\,2} := \text{plim}\, \hat{\Omega}_{\text{OLS},\,2}, \tag{3.3}$$

$$\Omega_{\text{IV}} := \text{plim}\, \frac{1}{T}Y'N_1Y\,, \quad \Sigma_Y := \text{plim}\, \frac{1}{T}Y'Y\,. \tag{3.4}$$

We have the following identities:

$$\Sigma_Y = \Sigma_V + \Sigma_g\,, \quad \Omega_{\text{OLS}} = \Sigma_V + \Sigma_g - \Sigma_{g1}\Sigma_1^{-1}\Sigma_{1g}\,, \tag{3.5}$$

$$\Omega_{\text{OLS},\,2} = \Sigma_V + \Sigma_g - \Sigma_{gX}\Sigma_X^{-1}\Sigma_{Xg}\,, \quad \Omega_{\text{IV}} = \Sigma_{gX}\Sigma_X^{-1}\Sigma_{Xg} - \Sigma_{g1}\Sigma_1^{-1}\Sigma_{1g}\,. \tag{3.6}$$

Further, the following lemma gives a number of useful related inequalities.

Lemma 1 (Upper and lower bounds for Σ_V)**.** *Under Assumptions 1, 3, 6, and 7, we have:*

$$\sigma_{Vu}(\sigma_u^2)^{-1}\sigma_{uV} \preceq \Sigma_V \preceq \Omega_{OLS,\,2} \preceq \Omega_{OLS} \preceq \Sigma_Y\,, \tag{3.7}$$

where the parameters σ_u^2, $\Omega_{OLS,\,2}$, Ω_{OLS}, and Σ_Y can be estimated consistently.

The consistent estimation of σ_u^2, $\Omega_{\text{OLS},\,2}$, Ω_{OLS}, and Σ_Y is obvious. Lemma 1 represents the price we have to pay to have the flexibility of Assumption 3 or Assumption 4. The (matrix) bounds on Σ_V can be consistently estimated, but not Σ_V itself. As a result, we design our inference approaches so that they do *not* rely on a consistent estimator of Σ_V. Even when a consistent estimator of Σ_V is available (*e.g.* under Assumption 5), we need to use it judiciously (this will be discussed below.)

Under Assumption 7, we can **define** a and e as follows:

$$a := \Sigma_V^{-1}\sigma_{Vu}, \quad e := u - Va\,. \tag{3.8}$$

Then, it is not difficult to show that

$$\frac{1}{T}e'e = \frac{1}{T}(u - Va)'(u - Va) \xrightarrow[T\to\infty]{\text{p}} \sigma_u^2 - \sigma_{uV}\Sigma_V^{-1}\sigma_{Vu} := \sigma_e^2\,. \tag{3.9}$$

To state results regarding convergence in distribution, we need an additional assumption. We consider here the following one.

Assumption 8 (Convergence, II)**.**

$$\frac{1}{\sqrt{T}}\,\text{vec}\left[\begin{pmatrix} X'u & X'V \\ \bar{g}'u & \bar{g}'V \end{pmatrix}\right] \xrightarrow[T\to\infty]{\text{d}} \text{vec}\left[\begin{pmatrix} \Psi_{Xu} & \Psi_{XV} \\ \Psi_{gu} & \Psi_{gV} \end{pmatrix}\right] \sim N\left(0,\, \Sigma \otimes \bar{\Sigma}\right), \tag{3.10}$$

$$\sqrt{T}\big[\,\frac{1}{T}V'u - \sigma_{Vu}\big] \xrightarrow[T\to\infty]{\text{d}} \Psi_{Vu} \sim N[0,\, \Omega_{Vu}] \tag{3.11}$$

where $\Omega_{Vu} := \sigma_u^2 \Sigma_V + \sigma_{Vu}\sigma_{uV}$. Moreover, $\{\Psi_{Xu}, \Psi_{XV}, \Psi_{\bar{g}u}, \Psi_{\bar{g}V}\}$ and Ψ_{Vu} are independent.

As with Assumption 7, Assumption 8 can readily accommodate unobserved instruments, as well as both strong and weak-instrument asymptotics. The multiplicative form of the covariance matrix in (3.10) follows from the asymptotic orthogonality between $\{X, \bar{g}\}$ and $\{u, V\}$ (Assumption 7). When

$$\bar{g} = X_1\Pi_1 + X_2\Pi_2 + X_3\Pi_3\,, \tag{3.12}$$

(3.10) is implied by Assumption 4.1 in Doko Tchatoka and Dufour (2020). The same asymptotic orthogonality between $\{X, \bar{g}\}$ and $\{u, V\}$ motivates the independence specification on the last line of Assumption 8. The additive form of the asymptotic covariance matrix in (3.11) is suggested by Isserlis's Theorem [Isserlis (1918)], also known as Wick's Theorem [Wick (1950)].[2] The intuition here can be seen by considering the case where the vectors $(u_t, V_t')'$ are i.i.d. with covariance matrix Σ [defined in Assumption 7]. Then, (3.11) holds with $\Omega_{Vu} = \mathbb{E}[u_t^2 V_t V_t']$. Note that $(u_t, V_t')'$ do not have to be multivariate normal, but if they are so asymptotically, then we can use Isserlis' Theorem to compute $\mathbb{E}(u_t^2 V_t V_t')$ *as if* $(u_t, V_t')'$ is multivariate normal.

Assumptions 7 and 8 hold in cases where the vectors $(u_t,\, V_t')'$ are i.i.d. (with appropriate finite moments), but they also allow for some error heteroskedasticity and autocorrelation in finite samples. However, the influence of these features disappear in the asymptotic distribution of the statistics we consider below.

[2] The variant sufficient for our purpose can be stated as follows.

Theorem [Isserlis]. If $(\mathcal{Z}_1, \mathcal{Z}_2, \mathcal{Z}_3, \mathcal{Z}_4)$ is a *zero-mean* multivariate normal random vector, then

$$E(\mathcal{Z}_{i_1}\mathcal{Z}_{i_2}\mathcal{Z}_{i_3}\mathcal{Z}_{i_4}) = E(\mathcal{Z}_{i_1}\mathcal{Z}_{i_2})E(\mathcal{Z}_{i_3}\mathcal{Z}_{i_4}) + E(\mathcal{Z}_{i_1}\mathcal{Z}_{i_3})E(\mathcal{Z}_{i_2}\mathcal{Z}_{i_4}) + E(\mathcal{Z}_{i_1}\mathcal{Z}_{i_4})E(\mathcal{Z}_{i_2}\mathcal{Z}_{i_3})$$

where the indices $1 \leqslant i_1, i_2, i_3, i_4 \leqslant 4$ can be equal to one another.

4 Two-stage inference for endogeneity covariances

In this section, we discuss inference methods based on a two-stage process which makes joint inference on the structural parameter β and the endogeneity covariances σ_{Vu}. Due to the possibility of identification failure (or weak identification), this turns out to be important for building identification-robust inference methods. We rely on the fact that identification-robust inference on β is already known to be feasible. This feature is then ported to joint inference on β and σ_{Vu}, and finally to inference on σ_{Vu} alone.

4.1 Inference for structural parameter β

We discuss inference for the structural parameter β, which can be viewed as a first step in the process considered here.

Proposition 1. *Suppose that Assumptions 1, 3, 6, 7 and 8 hold. Then,*

$$\mathrm{AR}(\beta) \xrightarrow[T\to\infty]{\mathrm{d}} \chi^2_{k_2}/k_2\,. \tag{4.1}$$

This asymptotic convergence holds regardless of whether the instruments are strong or weak. The null distribution of the AR statistic does not depend on the form of the reduced-form equation. Using (4.1), we can construct an asymptotic confidence set for β:

$$\mathbb{I}_\beta^{\mathrm{AR}}(\alpha) := \{\beta_0 \in \mathbb{R}^G : \mathrm{AR}(\beta_0) \leqslant Q_{1-\alpha}(\chi^2_{k_2})/k_2\}\,. \tag{4.2}$$

This confidence set can be unbounded, which indicates identification is weak. In this case, the unbounded confidence set remains useful as it should lead to searching for other instruments or changing the model.[3]

[3] As pointed out at the end of Section 3, Assumptions 7 and 8 do not preclude a certain level of heteroskadasticity (or autocorrelation) in the error vectors, but these complications are asymptotically negligible. Allowing for stronger forms of heteroskedasticity or autocorrelation would require modifications of the test statistics (along with the corresponding confidence set procedures). But this would go beyond the scope of the current paper.

4.2 Conditional point estimation of endogeneity covariances

We now develop the *first* inference approach to σ_{Vu}. On setting

$$\sigma_{Yu} := \text{plim}_{T\to\infty} \frac{1}{T} Y'u \tag{4.3}$$

and using Assumption 7, it is easy to see that

$$\sigma_{Yu} = \sigma_{Vu}. \tag{4.4}$$

An advantage of viewing σ_{Vu} as σ_{Yu} is that Y is observed. If $\grave{u}$ is an estimator of u, then

$$\hat{\sigma}_{Yu} = \frac{1}{T} Y'\grave{u} \tag{4.5}$$

is an estimator of σ_{Yu}. In particular, for any candidate value β_0 for the true β, we can form

$$\hat{\sigma}_{Vu}(\beta_0) := \hat{\sigma}_{Yu}(\beta_0) := \frac{1}{T} Y'\grave{u}(\beta_0)\,, \quad \grave{u}(\beta_0) := M_1(y - Y\beta_0). \tag{4.6}$$

One can verify that

$$\hat{\sigma}_{Vu}(\beta_0) - \sigma_{Vu} = \hat{\Omega}_{\text{OLS}}\big(\hat{\beta} - \beta_0 - \hat{\Omega}_{\text{OLS}}^{-1}\sigma_{Vu}\big)\,. \tag{4.7}$$

This closely links the conditional estimator of σ_{Vu} ("conditional" in the sense that $\hat{\sigma}_{Vu}$ entails having a candidate for β) to a bias-corrected estimator $\hat{\beta} - \hat{\Omega}_{\text{OLS}}^{-1}\sigma_{Vu}$ for β. In the next subsection, we look at this bias-corrected estimator.

4.3 Joint inference for structural parameters and endogeneity covariances

The bias-corrected estimator $\hat{\beta} - \hat{\Omega}_{\text{OLS}}^{-1}\sigma_{Vu}$ satisfies the following convergence result.

Lemma 2. *Suppose that Assumptions 1, 3, 6, 7 and 8 hold. Then,*

$$\sqrt{T}(\hat{\beta} - \beta - \hat{\Omega}_{OLS}^{-1}\sigma_{Vu}) \xrightarrow[T\to\infty]{\text{d}} N\big[0,\, \sigma_u^2\Omega_{OLS}^{-1} + \Omega_{OLS}^{-1}\sigma_{Vu}\sigma_{uV}\Omega_{OLS}^{-1}\big]\,. \tag{4.8}$$

In our context, a candidate for σ_{Vu} will be supplied by a hypothesis test. To use (4.8), we need a consistent estimator of σ_u^2. We know that $\hat{\sigma}^2$ underestimates σ_u^2 asymptotically, and while $\tilde{\sigma}^2$ is consistent for σ_u^2, though convergence may be slow when the instruments in X_2 are weak. Given for β_0 and σ_0 (both with dimension $G \times 1$), we consider:

$$\acute{\sigma}_u^2(\beta_0) := \frac{1}{T}(y - Y\beta_0)' M_1 (y - Y\beta_0)\,, \tag{4.9}$$

$$V(\beta_0,\, \sigma_0) := \acute{\sigma}_u^2(\beta_0)\hat{\Omega}_{\text{OLS}}^{-1} + \hat{\Omega}_{\text{OLS}}^{-1}\sigma_0\sigma_0'\hat{\Omega}_{\text{OLS}}^{-1}\,, \tag{4.10}$$

$$W(\beta_0,\, \sigma_0) := T(\hat{\beta} - \beta_0 - \hat{\Omega}_{\text{OLS}}^{-1}\sigma_0)'[V(\beta_0,\, \sigma_0)]^{-1}(\hat{\beta} - \beta_0 - \hat{\Omega}_{\text{OLS}}^{-1}\sigma_0)\,. \tag{4.11}$$

We can then prove the following convergence properties.

Theorem 1. *Suppose that Assumptions 1, 3, 6, 7 and 8 hold, and let $\acute{\sigma}_u^2$, V, and W be defined as in (4.9) and (4.10). Then,*

$$\acute{\sigma}_u^2(\beta) \xrightarrow[T\to\infty]{\text{p}} \sigma_u^2, \quad V(\beta,\, \sigma_{Vu}) \xrightarrow[T\to\infty]{\text{p}} \sigma_u^2\Omega_{OLS}^{-1} + \Omega_{OLS}^{-1}\sigma_{Vu}\sigma_{uV}\Omega_{OLS}^{-1}, \tag{4.12}$$

$$W(\beta,\, \sigma_{Vu}) \xrightarrow[T\to\infty]{\text{d}} \chi_G^2\,. \tag{4.13}$$

Theorem 1 can be exploited in two ways. First, (4.13) allows one to perform an asymptotic test for the joint hypothesis $\beta = \beta_0$, $\sigma_{Vu} = \sigma_0$ by comparing $W(\beta_0, \sigma_0)$ with the appropriate quantile of χ_G^2. Second, (4.13) allows one to perform an asymptotic test for σ_{Vu} *given* β. Under the assumptions of Theorem 1, we have:

$$\sqrt{T}\left[\hat{\sigma}_{Vu}(\beta) - \sigma_{Vu}\right] \xrightarrow[T\to\infty]{\text{d}} N\left[0,\, \sigma_u^2\Omega_{\text{OLS}} + \sigma_{Vu}\sigma_{uV}\right]. \tag{4.14}$$

Then, the function W – which is originally defined as a function of $\hat{\beta} - \beta_0$ – has the interpretation:

$$W(\beta_0,\, \sigma_0) = T(\hat{\sigma}_{Vu}(\beta_0) - \sigma_0)'\left[\acute{\sigma}_u^2(\beta_0)\hat{\Omega}_{\text{OLS}} + \sigma_0\sigma_0'\right]^{-1}(\hat{\sigma}_{Vu}(\beta_0) - \sigma_0)\,. \tag{4.15}$$

This perspective of inference for σ_{Vu} conditional on β induces the application of the results of Dufour (1990) in the next subsection.

4.4 Projection-based inference for endogeneity covariances

Let $1-\alpha$ be the nominal level for the desired confidence set for σ_{Vu}. Let α_1 and α_2 be any positive real numbers such that $\alpha_1 + \alpha_2 = \alpha$. Define: $\forall \beta_0 \in \mathbb{R}^G$,

$$\mathbb{I}^1_{\sigma_{Vu}}(\alpha_2; \beta_0) := \{\sigma_0 \in \mathbb{R}^G : W(\beta_0, \sigma_0) \leqslant Q_{1-\alpha_2}(\chi^2_G)\}, \tag{4.16}$$

$$\mathbb{G}^1_{(\beta,\,\sigma_{Vu})}(\alpha_1, \alpha_2) := \{(\beta_0, \sigma_0) : \beta_0 \in \mathbb{I}^{\text{AR}}_{\beta}(\alpha_1) \text{ and } \sigma_0 \in \mathbb{I}^1_{\sigma_{Vu}}(\alpha_2; \beta_0)\}, \tag{4.17}$$

$$\mathbb{I}^1_{\sigma_{Vu}}(\alpha_1, \alpha_2) := \{\sigma_0 \in \mathbb{R}^G : (\beta_0, \sigma_0) \in \mathbb{G}^1_{(\beta,\,\sigma_{Vu})}(\alpha_1, \alpha_2) \text{ for some } \beta_0 \in \mathbb{I}^{\text{AR}}_{\beta}(\alpha_1)\}. \tag{4.18}$$

These can be interpreted as follows: (1) $\mathbb{I}^1_{\sigma_{Vu}}(\alpha_2; \beta_0)$ is an asymptotic confidence set for σ_{Vu} *given* $\beta = \beta_0$; (2) $\mathbb{G}^1_{(\beta,\,\sigma_{Vu})}(\alpha_1, \alpha_2)$ is a joint confidence set for (β, σ) obtained by taking the graph of the set-valued map $\beta_0 \mapsto \mathbb{I}^1_{\sigma_{Vu}}(\alpha_2; \beta_0)$ for $\beta_0 \in \mathbb{I}^{\text{AR}}_{\beta}(\alpha_1)$; (3) $\mathbb{I}^1_{\sigma_{Vu}}(\alpha_1, \alpha_2)$ is corresponding the projection-based confidence set for σ. The validity of these confidence sets is established by the following theorem.

Theorem 2. *Suppose that Assumptions 1, 3, 6, 7, and 8 hold, and let* $1-\alpha$ *be any nominal coverage probability and let* $\mathbb{I}^{\text{AR}}_{\beta}(\alpha_1)$ *be as constructed in* (4.2)*. Then, for all* $\alpha_1 > 0$ *and* $\alpha_2 > 0$ *such that* $\alpha_1 + \alpha_2 = \alpha$, $\mathbb{G}^1_{(\beta,\,\sigma_{Vu})}(\alpha_1, \alpha_2)$ *and* $\mathbb{I}^1_{\sigma_{Vu}}(\alpha_1, \alpha_2)$ *are asymptotic confidence sets of nominal confidence level* $1-\alpha$ *for* (β, σ_{Vu}) *and* σ_{Vu}*, respectively,* i.e.

$$\lim_{T\to\infty} \Pr[(\beta, \sigma_{Vu}) \in \mathbb{G}^1_{(\beta,\,\sigma_{Vu})}(\alpha_1, \alpha_2)] \geqslant 1-\alpha, \tag{4.19}$$

$$\lim_{T\to\infty} \Pr[\sigma_{Vu} \in \mathbb{I}^1_{\sigma_{Vu}}(\alpha_1, \alpha_2)] \geqslant 1-\alpha. \tag{4.20}$$

This confidence set for σ_{Vu} does not rely on any specification of g other than the general regularity assumptions stated in Assumptions 7 and 8. In particular, we allow for nonlinearity, missing X_3, and no obvious candidate as a consistent estimator for Σ_V.

5 Inference for total effect

In this section, we provide a *second* approach for inference on σ_{Vu}. We do this by extending Doko Tchatoka and Dufour (2014) to allow for missing or unobserved instruments. Like with $\mathbb{I}^1_{\sigma_{Vu}}$, the confidence set built in this section ($\mathbb{I}^3_{\sigma_{Vu}}$ below) is also robust to weak instruments.

These authors observe that, under Assumption 5, (2.12) is correct, and, under regularity assumptions similar to those in Assumptions 7 and 8, it follows that

$$F(\beta + a) = \frac{[y - Y(\beta + a)]'[M - M(Z)][y - Y(\beta + a)]/G}{y'M(Z)y/(T - G - k)} \xrightarrow[T\to\infty]{\mathrm{d}} \chi^2_G/G\,. \tag{5.1}$$

Since $a = \Sigma_V^{-1}\sigma_{Vu}$, we have: $F(\beta + \Sigma_V^{-1}\sigma_{Vu}) \xrightarrow[T\to\infty]{\mathrm{d}} \chi^2_G/G$. This suggests the following definitions: for any $G \times G$ positive definite matrix Σ_0 and any $G \times 1$ vector β_0,

$$\begin{aligned}
\mathbb{I}^2_{\sigma_{Vu}}(\alpha_2; \beta_0, \Sigma_0) &:= \{\sigma_0 \in \mathbb{R}^G : F(\beta_0 + \Sigma_0^{-1}\sigma_0) \leqslant Q_{1-\alpha_2}(\chi^2_G)/G\},\\
\mathbb{G}^2_{(\beta,\,\sigma_{Vu})}(\alpha_1,\, \alpha_2; \Sigma_0) &:= \{(\beta_0,\, \sigma_0) : \beta_0 \in \mathbb{I}^{\mathrm{AR}}_\beta(\alpha_1) \text{ and } \sigma_0 \in \mathbb{I}^2_{\sigma_{Vu}}(\alpha_2; \beta_0,\, \Sigma_0)\},\\
\mathbb{I}^2_{\sigma_{Vu}}(\alpha_1,\, \alpha_2; \Sigma_0) &:= \{\sigma_0 \in \mathbb{R}^G : (\beta_0,\, \sigma_0) \in \mathbb{G}^2_{(\beta,\,\sigma_{Vu})}(\alpha_1,\, \alpha_2; \Sigma_0) \text{ for some } \beta_0 \in \mathbb{I}^{\mathrm{AR}}_\beta(\alpha_1)\}.
\end{aligned}$$

If Assumption 5 holds, then

$$\lim_{T\to\infty} \Pr[\sigma_{Vu} \in \mathbb{I}^2_{\sigma_{Vu}}(\alpha_1,\, \alpha_2; \Sigma_V)] \geqslant 1 - \alpha \text{ whenever } \alpha_1 + \alpha_2 = \alpha\,.$$

The performance of the confidence set $\mathbb{I}^2_{\sigma_{Vu}}$ depends on the on the strength of X_2 as instrument. *However, it is clearly vulnerable to misspecification of* (2.5)*. Another issue is that* Σ_V *is unavailable.*

When (2.5) holds, a natural candidate for estimating Σ_V is

$$\hat{\Sigma}_V := \frac{1}{T - k} Y'MY = \frac{T}{T - k}\hat{\Omega}_{\mathrm{OLS},\,2}\,. \tag{5.2}$$

However, $F(\beta + \Sigma_V^{-1}\sigma_{Vu})$ does not have the same asymptotic distribution as $F(\beta + \hat{\Sigma}_V^{-1}\sigma_{Vu})$. For

$\hat{\theta} := (Y'MY)^{-1}Y'My$, we have:

$$\sqrt{T}(\hat{\theta} - \beta - \hat{\Sigma}_V^{-1}\sigma_{Vu}) = \sqrt{T}(\hat{\theta} - \beta - a) + \hat{\Sigma}_V^{-1}\sqrt{T}(\hat{\Sigma}_V - \Sigma_V)a\,. \tag{5.3}$$

In the latter equation, the second term contributes to the asymptotic distribution of $\sqrt{T}(\hat{\theta} - \beta - \hat{\Sigma}_V^{-1}\sigma_{Vu})$, even if $\hat{\Sigma}_V \xrightarrow[T\to\infty]{\text{p}} \Sigma_V$. The problem is more unpredictable when (2.5) is misspecified because $\hat{\Sigma}_V \xrightarrow[T\to\infty]{\text{p}} \Sigma_V$ no longer holds. In the present framework, it is easy to rectify the situation using the following lemma.

Lemma 3. *Suppose that Assumptions 1, 3, 6, 7, and 8 hold. Then,*

$$\sqrt{T}(\hat{\theta} - \beta - \hat{\Sigma}_V^{-1}\sigma_{Vu}) \xrightarrow[T\to\infty]{\text{d}} N\big[0,\, \sigma_u^2\Omega_{OLS,\,2}^{-1} + \Omega_{OLS,\,2}^{-1}\sigma_{Vu}\sigma_{uV}\Omega_{OLS,\,2}^{-1}\big]\,. \tag{5.4}$$

Under Assumption 5, $\mathbb{I}^2_{\sigma_{Vu}}(\cdot;\beta,\hat{\Sigma})$ would treat the left-hand side (LHS) of (5.4) *as if* its asymptotic covariance matrix were

$$\sigma_e^2\Omega_{\text{OLS},\,2}^{-1} = \sigma_u^2\Omega_{\text{OLS},\,2}^{-1} - \sigma_{uV}\Sigma_V^{-1}\sigma_{Vu}\Omega_{\text{OLS},\,2}^{-1} \tag{5.5}$$

where σ_e^2 is defined in (3.9). It is clear that the quantity in (5.5) is smaller than the *actual* asymptotic covariance matrix in (5.4). Thus, even when Assumption 5 is true, the joint test for $\beta = \beta_0$ and $\sigma_{Vu} = \sigma_0$ as using $F(\beta_0 + \hat{\Sigma}_V^{-1}\sigma_0)$ and a critical value of χ^2_G/G will *overreject* asymptotically, thus failing to achieve asymptotic level control.

On the other hand, for the same joint hypothesis, we can use (5.4) to achieve asymptotic level control with:

$$\bar{W}(\beta_0,\, \sigma_0) := T\,(\hat{\theta} - \beta_0 - \hat{\Sigma}_V^{-1}\sigma_0)'\hat{\Sigma}(\beta_0,\, \sigma_0)^{-1}(\hat{\theta} - \beta_0 - \hat{\Sigma}_V^{-1}\sigma_0) \tag{5.6}$$

where

$$\hat{\Sigma}(\beta_0,\, \sigma_0) := \acute{\sigma}_u(\beta_0)^2\hat{\Omega}_{\text{OLS},\,2}^{-1} + \hat{\Omega}_{\text{OLS},\,2}^{-1}\sigma_0\sigma_0'\hat{\Omega}_{\text{OLS},\,2}^{-1}\,. \tag{5.7}$$

Even if Assumption 5 is not satisfied, we have (under the null hypothesis):

$$\bar{W}(\beta,\, \sigma_{Vu}) \xrightarrow[T\to\infty]{\text{d}} \chi^2_G \tag{5.8}$$

and valid confidence sets for σ_{Vu} [denoted $\mathbb{I}^3_{\sigma_{Vu}}(\alpha_2; \beta_0)$ and $\mathbb{I}^3_{\sigma_{Vu}}(\alpha_1,\, \alpha_2)$] can be built on replacing W by $\bar{W}$ in (4.16) - (4.18). This is formally stated in the following theorem.

Theorem 3. *Suppose that Assumptions 1, 3, 6, 7, and 8 hold, and let $\alpha \in (0,1)$, $\alpha_1 > 0$ and $\alpha_2 > 0$ with $\alpha_1 + \alpha_2 = \alpha$. Then,*

$$\lim_{T\to\infty} \Pr\left[\bar{W}(\beta,\, \sigma_{Vu}) > Q_{1-\alpha}(\chi^2_G)\right] = \alpha, \quad \lim_{T\to\infty} \Pr\left[\sigma_{Vu} \in \mathbb{I}^3_{\sigma_{Vu}}(\alpha_1,\, \alpha_2)\right] \geqslant 1 - \alpha \tag{5.9}$$

where $\mathbb{I}^3_{\sigma_{Vu}}(\alpha_1,\, \alpha_2)$ is defined by (4.16) - (4.18) with W replaced by $\bar{W}$.

6 Exogeneity tests

In this section, we present a *third* approach for inference on σ_{Vu}. Its rationale derives from the observation that $\tilde{\beta} - \hat{\beta}$ can be interpreted as a measure of the OLS bias, which in turn depends on the endogeneity covariances σ_{Vu}. Under Assumption 7, we have:

$$\frac{1}{T} Y' N_1 u = \frac{1}{T} Y' P u - \frac{1}{T} Y' P_1 u \xrightarrow[T\to\infty]{\text{p}} 0 - 0 = 0\,, \tag{6.1}$$

$$\frac{1}{T} Y' M_1 u = \frac{1}{T} (\bar{g} + V)' u - \frac{1}{T} Y' P_1 u \xrightarrow[T\to\infty]{\text{p}} \sigma_{Vu}\,. \tag{6.2}$$

We can then take the probability limit of the difference in (2.23):

$$\hat{\beta} - \tilde{\beta} = -C_1 u = B_1 u - B_2 u \xrightarrow[T\to\infty]{\text{p}} \Omega^{-1}_{\text{OLS}} \sigma_{Vu} - \Omega^{-1}_{\text{IV}} 0 = \Omega^{-1}_{\text{OLS}} \sigma_{Vu}\,. \tag{6.3}$$

This suggests to consider $\hat{\beta} - \tilde{\beta} - \hat{\Omega}^{-1}_{\text{OLS}} \sigma_{Vu}$ as for inference on σ_{Vu}.

Theorem 4. *Suppose that Assumptions 3, 6, 7, and 8 hold, and set*

$$\underline{W}(\sigma_0) := T\big[(\hat{\beta}-\tilde{\beta}) - \hat{\Omega}_{OLS}^{-1}\sigma_0\big]'\big[\hat{\Omega}_{OLS}^{-1}\sigma_0\sigma_0'\hat{\Omega}_{OLS}^{-1} + \tilde{\sigma}^2\hat{\Delta}\big]^{-1}\big[(\hat{\beta}-\tilde{\beta}) - \hat{\Omega}_{OLS}^{-1}\sigma_0\big] \tag{6.4}$$

where Ω_{OLS} and Ω_{IV} are as defined in (3.4). Then,

$$\sqrt{T}\big[\hat{\Omega}_{OLS}(\hat{\beta}-\tilde{\beta}) - \sigma_{Vu}\big] \xrightarrow[T\to\infty]{\mathrm{d}} N[0,\, \sigma_{Vu}\sigma_{uV} + \sigma_u^2(\Omega_{OLS}\Omega_{IV}^{-1}\Omega_{OLS} - \Omega_{OLS})]\,, \tag{6.5}$$

$$\sqrt{T}\big[(\hat{\beta}-\tilde{\beta}) - \hat{\Omega}_{OLS}^{-1}\sigma_{Vu}\big] \xrightarrow[T\to\infty]{\mathrm{d}} N[0,\, \Omega_{OLS}^{-1}\sigma_{Vu}\sigma_{uV}\Omega_{OLS}^{-1} + \sigma_u^2(\Omega_{IV}^{-1} - \Omega_{OLS}^{-1})]\,. \tag{6.6}$$

Further, for any $G \times 1$ fixed vector σ_0, $0 < \alpha < 1$, and

$$\mathbb{I}_{\sigma_{Vu}}^4(\alpha) := \{\sigma_0 \in \mathbb{R}^G : \underline{W}(\sigma_0) \leqslant Q_{1-\alpha}(\chi_G^2)\}\,, \tag{6.7}$$

we have:

$$\underline{W}(\sigma_{Vu}) \xrightarrow[T\to\infty]{\mathrm{d}} \chi_G^2\,, \tag{6.8}$$

$$\lim_{T\to\infty} \Pr[\sigma_{Vu} \in \mathbb{I}_{\sigma_{Vu}}^4(\alpha)] = 1-\alpha\,. \tag{6.9}$$

We note that $\underline{W}(0)$ coincides with the Hausman statistic $\mathscr{H} = (T/\kappa)\mathscr{T}$. Moreover, the confidence set $\mathbb{I}_{\sigma_{Vu}}^4$ does not rely on a consistent estimator for Σ_V and is robust to general specifications of $\bar{g}$. On the other hand, as $\tilde{\beta}$ and $\tilde{\sigma}^2$ are involved, $\mathbb{I}_{\sigma_{Vu}}^4$ is vulnerable if X_2 contains weak instruments. However, when X_2 is strong, then $\mathbb{I}_{\sigma_{Vu}}^4(\alpha)$ should exhibit better size control in finite samples as it is a direct confidence set for σ_{Vu} that does not go through any two-stage construction.

7 Application to return-to-schooling model

In this section, we illustrate the theoretical results with an empirical example using return to schooling data provided in Hayashi (2000). In turn, this data set is a subset of that from Blackburn and Neumark (1992). A motivation for our selection of this application is that it has been studied in the related works of Kiviet (2020) and Kiviet and Pleus (2017). This should give us comparability.

We begin by describing the data, the variables, and performing some preliminary calculations. We will compute $\mathbb{I}_{\beta}^{\mathrm{AR}}$ and $\mathbb{I}_{\sigma_{Vu}}^{1}$. To simplify the presentation, we do not compute $\mathbb{I}_{\sigma_{Vu}}^{3}$ because, as we have seen in our simulation experiments, $\mathbb{I}_{\sigma_{Vu}}^{3}$ performs very similarly to $\mathbb{I}_{\sigma_{Vu}}^{1}$. On the other hand, while $\mathbb{I}_{\sigma_{Vu}}^{4}$ is generally not as reliable as $\mathbb{I}_{\sigma_{Vu}}^{1}$, the exogeneity-test based confidence set is easier to compute, so it is included for comparison. Along the way, we also provide remarks on computation implementation.

A deep dive into this data set is within neither the intention nor the scope of this paper. Indeed, the objective here is to demonstrate the practical applicability of the tools developed in this paper. We will, however, provide a short discussion at the end of the section.

7.1 Data

Hayashi's data contain a panel on nonblack men, each of whom appears in the data two times, one before 1980 and one in 1980. We focus on the earlier appearance. The dependent variable is log wage (`lw`) and the explanatory variables are: constant, dummy of residence in the southern states (`rns`), dummy for residency in metropolitan areas (`smsa`), tenure in years (`tenure`), experience in years (`expr`), iq (`iq`), age (`age`), schooling in years (`s`), and year dummies for the years 1967–1971 and 1973 (there is no observation for 1972 and the year 1966 is excluded as a constant is already included). There is also `kww` (score on the test "Knowledge of the World of Work"), which is used as an excluded instrument, along with its square and the squares of `age` and `exp`. Overall, there are $T = 758$ observations.

In our notation, the above translates to:

$$T = 758, \quad y = \texttt{lw}, \quad X_2 = [\texttt{kww}, \texttt{kww}^2, \texttt{age}^2, \texttt{expr}^2], \quad k_2 = 4\,. \tag{7.1}$$

As for Y and X_1, Kiviet and Pleus (2017) consider:

$$\text{IV} \quad : \quad \begin{aligned} & G = 2, \quad k_1 = 12, \quad Y = [\mathtt{s},\ \mathtt{iq}], \\ & X_1 = [\iota,\ \mathtt{expr},\ \mathtt{rns},\ \mathtt{tenure},\ \mathtt{smsa},\ \mathtt{age},\ \text{year dummies}]\,; \end{aligned} \tag{7.2}$$

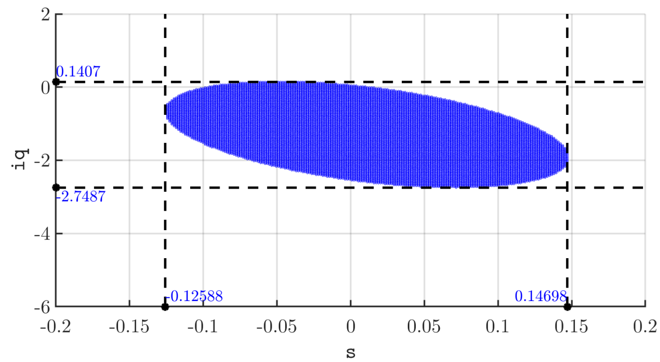
$$\text{IV}_1 \quad : \quad \begin{aligned} & G = 1, \quad k_1 = 13, \quad Y = \mathtt{iq}, \\ & X_1 = [\iota,\ \mathtt{s},\ \mathtt{expr},\ \mathtt{rns},\ \mathtt{tenure},\ \mathtt{smsa},\ \mathtt{age},\ \text{year dummies}]; \end{aligned} \tag{7.3}$$

$$\text{IV}_2 \quad : \quad \begin{aligned} & G = 1, \quad k_1 = 13, \quad Y = \mathtt{s}, \\ & X_1 = [\iota,\ \mathtt{iq},\ \mathtt{expr},\ \mathtt{rns},\ \mathtt{tenure},\ \mathtt{smsa},\ \mathtt{age},\ \text{year dummies}]\,. \end{aligned} \tag{7.4}$$

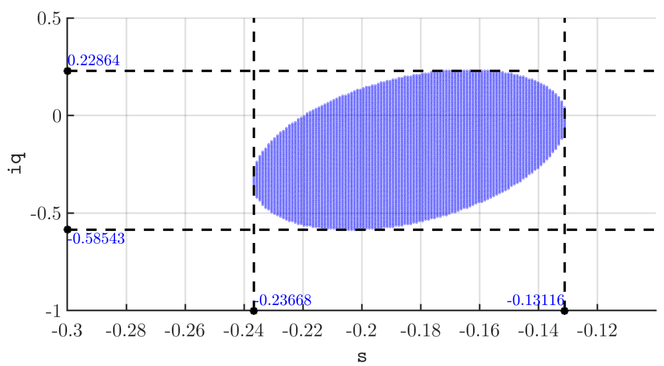
For brevity here, we focus on `IV` configuration. The working paper version of this work provides additional results for $\mathtt{IV}_1$ and $\mathtt{IV}_2$.

7.2 Results

When $\mathbb{I}^1_{\sigma_{Vu}}$ has to be computed in two stages, there are two complications: $\mathbb{I}^1_{\sigma_{Vu}}$ is conservative and may be unbounded. In our example here, $Y = [\mathtt{s}, \mathtt{iq}]$, the eigenvalues of $Y'M(\alpha_1)Y$ are both positive, so the AR confidence set for β will be bounded [Dufour and Taamouti (2005)], so constructing $\mathbb{I}^1_{\sigma_{Vu}}$ is almost as easy as constructing $\mathbb{I}^4_{\sigma_{Vu}}$.



(a) $\mathbb{I}^4_{\sigma_{Vu}}(\alpha)$ with $\alpha = 0.05$



(b) $\mathbb{I}^1_{\sigma_{Vu}}(\alpha_1, \alpha_2)$ with $\alpha_1 = \alpha_2 = 0.025$

Figure 1: Confidence regions for σ_{Vu}

In Figure 1, we show both the $\mathbb{I}^1_{\sigma_{Vu}}$ and $\mathbb{I}^4_{\sigma_{Vu}}$ confidence sets for σ_{Vu} in (6.7). By comparing panels (a) and (b), we see that, despite its inherent conservativeness, $\mathbb{I}^1_{\sigma_{Vu}}$ is not always larger than $\mathbb{I}^4_{\sigma_{Vu}}$. Kiviet and Pleus (2017) warn of "slightly weak" instruments when both `iq` and `s` are treated as endogenous, but this is not an issue for us as $\mathbb{I}^1_{\sigma_{Vu}}$ is identification-robust.

8 Conclusion

In this paper, we take an interest in endogeneity parameters: the covariances between the endogenous variables and the error term. We have developed three new inference procedures for these, all of which are robust to missing instruments. Two of the procedures are robust to weak instruments, whereas the third procedure can have better size control when the excluded instruments are strong.

A venue for future work is the bootstrap implications of our results. The plug-in estimator of Section 4 naturally suggests bootstrap inference. And, as we have computed various asymptotic covariance matrices, it is easy to enjoy bootstrap refinement by working only with pre-pivoted statistics.

References

Abadir, K. M. and J. R. Magnus (2005): *Matrix Algebra*, vol. 1 of *Econometric Exericses*, New York, NY 10013-2473, USA: Cambridge University Press.

Anderson, T. 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.

Blackburn, M. and D. Neumark (1992): "Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials," *Quarterly Journal of Economics*, 107, 1421–1436.

Bound, J., D. Jaeger, and R. Baker (1995): "Problems with Instrument Variables Estimnation 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. and J.-M. Dufour (2014): "Identifiction-robust inference for endogeneity parameters in linear structural models," *The Econometrics Journal*, 17, 165–187.

——— (2017): "Exogeneity tests and weak identification in IV regressions: asymptotic theory and point estimation," Working paper.

——— (2020): "Exogeneity tests, weak identification, incomplete models and non-Gaussian distributions: Invariance and finite-sample distributional theory," *Journal of Econometrics*, 219, 390–418.

Dufour, J.-M. (1987): "Linear Wald methods for inference on covariances and weak exogeneity tests in structural equations," in *Advances in the Statistical Sciences: Festschrift in Honour of*

Professor V.M. Joshi's 70th Birthday. Time Series and Econometric Modelling, Volume III, ed. by I. B. MacNeill and G. J. Umphrey, Reidel: Dordrecht, 317–338.

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

——— (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1387.

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

Dufour, J.-M. and J. Jasiak (2001): "Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors," *International Economic Review*, 42, 815–843.

Dufour, J.-M. and V. Nguyen (2020): "Identification-robust Inference for Endogeneity Parameters in Models with an Incomplete Reduced Form," Tech. rep., Department of Economics, McGill University, Montréal, Québec, Canada.

Dufour, J.-M. and M. Taamouti (2005): "Projection-based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica*, 73, 1351–1365.

Durbin, J. (1954): "Errors in variables," *Revue de l'institut International de Statistique*, 23–32.

Gourieroux, C. and A. Monfort (1995): *Statistics and Econometric Models, Volume Two*, New York, NY: Cambridge University Press.

Hahn, J., J. C. Ham, and H. R. Moon (2011): "The Hausman test and weak instruments," *Journal of Econometrics*, 160, 289–299.

Hausman, J. A. (1978): "Specification tests in econometrics," *Econometrica: Journal of the econometric society*, 1251–1271.

Hayashi, F. (2000): *Econometrics*, Princeton, NJ: Princeton University Press.

Isserlis, L. (1918): "On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables," *Biometrika*, 12, 134–139.

Kiviet, J. F. (2020): "Testing the impossible: identifying exclusion restrictions," *Journal of Econometrics*, 218, 294–316.

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," *Journal Computational Statistics & Data Analysis*, 56, 3567–3586.

Kiviet, J. F. and M. Pleus (2017): "The performance of tests on endogeneity of subsets of explanatory variables scanned by simulation," *Econometrics and Statistics*, 2, 1–27.

Mikusheva, A. (2013): "Survey on Statistical Inferences in Weakly-identified Instrumental Variable Models," *Applied Econometrics*, 29, 117–131.

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*, 625–631.

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

Tripathi, G. (1999): "A matrix extension of the Cauchy-Schwarz inequality," *Economics Letters*, 63, 1–3.

Wick, G.-C. (1950): "The evaluation of the collision matrix," *Physical review*, 80, 268.

Wu, D.-M. (1973): "Alternative tests of independence between stochastic regressors and disturbances," *Econometrica: journal of the Econometric Society*, 733–750.

Identification-robust inference for endogeneity parameters in models with an incomplete reduced form

Jean-Marie Dufour and Vinh Nguyen

Technical appendix

December 2020

A Proofs

In this section, we provide proofs for the results in the main text of the paper.

Proof of Lemma 1. By (3.5), we have $\Sigma_Y \geq \Sigma_V$, and by the matrix Cauchy-Schwarz inequality [for the relevant variant, see Gourieroux and Monfort (1995), p. 469, or Tripathi (1999)], $\Omega_{\mathrm{OLS}} \geq \Sigma_V$ and $\Omega_{\mathrm{OLS},\,2} \geq \Sigma_V$. Thus, Σ_Y, Ω_{OLS} and $\Omega_{\mathrm{OLS},\,2}$ provide upper bounds for Σ_V. Moreover, we have the ordering

$$\Sigma_Y \geq \Omega_{\mathrm{OLS}} \geq \Omega_{\mathrm{OLS},\,2} \tag{A.1}$$

where the second inequality uses $\Omega_{\mathrm{OLS}} - \Omega_{\mathrm{OLS},\,2} = \Omega_{\mathrm{IV}} \geq 0$. By the Cauchy-Schwarz inequality again, we get

$$\Sigma_V \geq \sigma_{Vu}(\sigma_u^2)^{-1}\sigma_{uV}\,. \tag{A.2}$$

□

Proof of Proposition 1. Because $y = Y\beta + X_1\gamma + u$, we can write

$$\mathrm{AR}(\beta) = \frac{u'[M_1 - M]u/k_2}{u'Mu/(T-k)}. \tag{A.3}$$

For the numerator, we use $M_1 - M = P(M_1 X_2)$ to write:

$$(X_2' M_1 X_2)^{-1/2} X_2' M_1 u \overset{a}{\sim} \left(\Sigma_2 - \Sigma_{21}\Sigma_1^{-1}\Sigma_{12}\right)^{-1/2}\left[-\Sigma_{21}\Sigma_1^{-1} \,\vdots\, I_{k_2}\right] \frac{1}{\sqrt{T}} X' u \tag{A.4}$$

which converges in distribution to $N(0, \sigma_u^2 I_{k_2})$. Therefore, the numerator of AR(β) converges in distribution to $\sigma_u^2 \chi^2_{k_2}/k_2$. Now the claim follows because, for the denominator, we have:

$$\frac{1}{T-k} u' M u \overset{a}{\sim} \frac{1}{T} u'u - \frac{1}{T} u' X \left(\frac{1}{T} X' X\right)^{-1} \frac{1}{T} X' u \xrightarrow{\text{P}} \sigma_u^2. \tag{A.5}$$

□

Proof of Lemma 2. We have

$$\begin{aligned} \sqrt{T}(\hat\beta - \beta - \hat\Omega_{\text{OLS}}^{-1}\sigma_{Vu}) &= \hat\Omega_{\text{OLS}}^{-1}\sqrt{T}\left[\frac{1}{T} Y' M_1 u - \sigma_{Vu}\right] \\ &\overset{a}{\sim} \Omega_{\text{OLS}}^{-1}\left[\sqrt{T}\left(\frac{1}{T}V'u - \sigma_{Vu}\right) + \left[-\Sigma_{g1}\Sigma_1^{-1} \,\vdots\, I_G\right]\frac{1}{\sqrt{T}}\begin{pmatrix} X_1' u \\ \bar g' u\end{pmatrix}\right] \end{aligned} \tag{A.6}$$

which converges in distribution to a normal distribution with mean 0 and covariance matrix

$$\begin{aligned} &\Omega_{\text{OLS}}^{-1}\left[\sigma_u^2 \Sigma_V + \sigma_{Vu}\sigma_{uV} + \sigma_u^2 \begin{pmatrix} -\Sigma_{g1}\Sigma_1^{-1} & I_G \end{pmatrix}\begin{pmatrix} \Sigma_1 & \Sigma_{1g} \\ \Sigma_{g1} & \Sigma_g \end{pmatrix}\begin{pmatrix} -\Sigma_1^{-1}\Sigma_{1g} \\ I_G\end{pmatrix}\right]\Omega_{\text{OLS}}^{-1} \\ &\quad = \Omega_{\text{OLS}}^{-1}\left[\sigma_u^2\Sigma_V + \sigma_{Vu}\sigma_{uV} + \sigma_u^2\left(\Sigma_g - \Sigma_{g1}\Sigma_1^{-1}\Sigma_{1g}\right)\right]\Omega_{\text{OLS}}^{-1} \\ &\quad = \sigma_u^2 \Omega_{\text{OLS}}^{-1} + \Omega_{\text{OLS}}^{-1}\sigma_{Vu}\sigma_{uV}\Omega_{\text{OLS}}^{-1}. \end{aligned} \tag{A.7}$$

□

Proof of Theorem 2. Boole's inequality implies

$$\begin{aligned} \Pr[(\beta, \sigma_{Vu}) \in \mathbb{G}^1_{(\beta,\sigma_{Vu})}(\alpha_1, \alpha_2)] &= \Pr[\beta \in \mathbb{I}^{\text{AR}}_\beta(\alpha_1) \text{ and } \sigma_{Vu} \in \mathbb{I}^1_{\sigma_{Vu}}(\alpha_2; \beta_0)] \\ &\geqslant \Pr[\beta \in \mathbb{I}^{\text{AR}}_\beta(\alpha_1)] + \Pr[\sigma_{Vu} \in \mathbb{I}^1_{\sigma_{Vu}}(\alpha_2; \beta_0)] - 1, \end{aligned} \tag{A.8}$$

which converges to $(1-\alpha_1)+(1-\alpha_2)-1 = 1-\alpha$, thanks to (4.1) and Theorem 1. This proves the first inequality in (4.19). Next, because $(\beta, \sigma_{Vu}) \in \mathbb{G}^1_{(\beta, \sigma_{Vu})}(\alpha_1, \alpha_2)$ implies $\sigma_{Vu} \in \mathbb{I}^1_{\sigma_{Vu}}(\alpha_1, \alpha_2)$, we have

$$\Pr[\sigma_{Vu} \in \mathbb{I}^1_{\sigma_{Vu}}(\alpha_1,\, \alpha_2)] \geqslant \Pr[(\beta,\, \sigma_{Vu}) \in \mathbb{G}^1_{(\beta,\, \sigma_{Vu})}(\alpha_1,\, \alpha_2)]\,. \tag{A.9}$$

Taking the limits of both sides, we have the second inequality in (4.19).

□

Proof of Lemma 3. We have:

$$\begin{aligned}
\sqrt{T}\left[\hat{\theta} - \beta - \hat{\Sigma}_V^{-1}\sigma_{Vu}\right] &= \sqrt{T}\left[\hat{\Omega}^{-1}_{\mathrm{OLS},\,2}\Big(\frac{1}{T}Y'Mu\Big) - \frac{T-k}{T}\hat{\Omega}^{-1}_{\mathrm{OLS},\,2}\sigma_{Vu}\right] \\
&\overset{a}{\sim} \hat{\Omega}^{-1}_{\mathrm{OLS},\,2}\sqrt{T}\left[\frac{1}{T}Y'Mu - \sigma_{Vu}\right] \\
&\xrightarrow[T\to\infty]{\mathrm{d}} N\left(0,\, \sigma_u^2\Omega^{-1}_{\mathrm{OLS},\,2} + \Omega^{-1}_{\mathrm{OLS},\,2}\sigma_{Vu}\sigma_{uV}\Omega^{-1}_{\mathrm{OLS},\,2}\right)\,.
\end{aligned} \tag{A.10}$$

The last convergence above follows in a similar manner to (4.8).

□

Proof of Theorem 4. To start, we use (6.3) to write:

$$\begin{aligned}
\sqrt{T}\Big[\frac{1}{T}(Y'M_1Y)(\hat{\beta}-\tilde{\beta}) - \sigma_{Vu}\Big] &= \frac{1}{\sqrt{T}}(Y'M_1Y)(B_1-B_2)u - \sqrt{T}\sigma_{Vu} \\
&= \frac{1}{\sqrt{T}}V'u - \sqrt{T}\sigma_{Vu} - \frac{1}{\sqrt{T}}V'P_1u + \frac{1}{\sqrt{T}}\bar{g}'M_1u - \frac{1}{\sqrt{T}}(Y'M_1Y)(Y'N_1Y)^{-1}Y'(P-P_1)u \\
&\overset{a}{\sim} \sqrt{T}\left(\frac{1}{T}V'u - \sigma_{Vu}\right) + \frac{1}{\sqrt{T}}\bar{g}'M_1u - \frac{1}{\sqrt{T}}(Y'M_1Y)(Y'N_1Y)^{-1}\bar{g}'(P-P_1)u \\
&\overset{a}{\sim} \sqrt{T}\left(\frac{1}{T}V'u - \sigma_{Vu}\right) + \frac{1}{\sqrt{T}}\bar{g}'u - \Sigma_{g1}\Sigma_1^{-1}\frac{1}{\sqrt{T}}X_1'u - \Omega_{\mathrm{OLS}}\Omega_{\mathrm{IV}}^{-1}\Sigma_{gX}\Sigma_X^{-1}\frac{1}{\sqrt{T}}X'u \\
&\quad + \Omega_{\mathrm{OLS}}\Omega_{\mathrm{IV}}^{-1}\Sigma_{g1}\Sigma_1^{-1}\frac{1}{\sqrt{T}}X_1'u.
\end{aligned} \tag{A.11}$$

Here, we have used the following convergence results:

$$V'X_1 \xrightarrow[T\to\infty]{\mathrm{p}} 0\,, \quad V'X \xrightarrow[T\to\infty]{\mathrm{p}} 0\,, \tag{A.12}$$

$$\frac{1}{\sqrt{T}} V' P_1 u \xrightarrow[T\to\infty]{\mathrm{d}} 0\,, \quad \frac{1}{\sqrt{T}} V'(P - P_1) u \xrightarrow[T\to\infty]{\mathrm{d}} 0\,. \tag{A.13}$$

The term $\sqrt{T}(\frac{1}{T}V'u - \sigma_{Vu}) \xrightarrow[T\to\infty]{\mathrm{d}} \Psi_{Vu}$ is directly given by Assumption 8. To deal with the remaining terms in the last two lines above, we write:

$$\Sigma_X^{-1} = \begin{bmatrix} \Sigma^1 & \Sigma^{12} \\ \Sigma^{21} & \Sigma^2 \end{bmatrix} = \begin{bmatrix} \Sigma^{\cdot 1} & \Sigma^{\cdot 2} \end{bmatrix}, \quad \Sigma^{\cdot 1} := \begin{bmatrix} \Sigma^1 \\ \Sigma^{21} \end{bmatrix}, \quad \Sigma^{\cdot 2} := \begin{bmatrix} \Sigma^{12} \\ \Sigma^2 \end{bmatrix}, \tag{A.14}$$

$$\sqrt{T}\big[\frac{1}{T}(Y' M_1 Y)(\hat{\beta} - \tilde{\beta}) - \sigma_{Vu}\big] \overset{a}{\sim} \sqrt{T}\big(\frac{1}{T} V'u - \sigma_{Vu}\big) + \delta\, \varphi_T \tag{A.15}$$

where

$$\varphi_T := \frac{1}{\sqrt{T}} \begin{pmatrix} X_1' u \\ X_2' u \\ \bar{g}' u \end{pmatrix} \xrightarrow[T\to\infty]{\mathrm{d}} N[0,\, \sigma_u^2 \bar{\Sigma}]\,, \tag{A.16}$$

$$\bar{\Sigma} := \begin{bmatrix} \Sigma_X & \Sigma_{Xg} \\ \Sigma_{gX} & \Sigma_g \end{bmatrix} = \begin{bmatrix} \Sigma_1 & \Sigma_{12} & \Sigma_{1g} \\ \Sigma_{21} & \Sigma_2 & \Sigma_{2g} \\ \Sigma_{g1} & \Sigma_{g2} & \Sigma_g \end{bmatrix}, \tag{A.17}$$

$$\delta := \big[-\Sigma_{g1}\Sigma_1^{-1} + \Omega_{\mathrm{OLS}}\Omega_{\mathrm{IV}}^{-1}\Sigma_{g1}\Sigma_1^{-1} - \Omega_{\mathrm{OLS}}\Omega_{\mathrm{IV}}^{-1}\Sigma_{gX}\,\Sigma^{\cdot 1} \,\vdots\, -\Omega_{\mathrm{OLS}}\Omega_{\mathrm{IV}}^{-1}\Sigma_{gX}\,\Sigma^{\cdot 2} \,\vdots\, I_G \big]\,. \tag{A.18}$$

Since $\varphi_T \sim N(0, \sigma_u^2 \bar{\Sigma})$, we compute $\delta\,\bar{\Sigma}\,\delta'$. We can do this step-by-step as follows:

$$\begin{aligned} \delta \begin{bmatrix} \Sigma_1 \\ \Sigma_{21} \\ \Sigma_{g1} \end{bmatrix} &= -\Sigma_{g1} + \Omega_{\mathrm{OLS}}\Omega_{\mathrm{IV}}^{-1}\Sigma_{g1} - \Omega_{\mathrm{OLS}}\Omega_{\mathrm{IV}}^{-1}\Sigma_{gX} \begin{bmatrix} \Sigma^1\Sigma_1 + \Sigma^{12}\Sigma_{21} \\ \Sigma^{21}\Sigma_1 + \Sigma^2\Sigma_{21} \end{bmatrix} + \Sigma_{g1} \\ &= \Omega_{\mathrm{OLS}}\Omega_{\mathrm{IV}}^{-1}\Sigma_{g1} - \Omega_{\mathrm{OLS}}\Omega_{\mathrm{IV}}^{-1}\big[\Sigma_{g1} \,\vdots\, \Sigma_{g2}\big] \begin{bmatrix} I_{k_1} \\ 0_{k_2 \times k_1} \end{bmatrix} \\ &= \Omega_{\mathrm{OLS}}\Omega_{\mathrm{IV}}^{-1}\Sigma_{g1} - \Omega_{\mathrm{OLS}}\Omega_{\mathrm{IV}}^{-1}\Sigma_{g1} = 0\,, \end{aligned} \tag{A.19}$$

$$
\delta\begin{pmatrix}\Sigma_{12}\\ \Sigma_2\\ \Sigma_{g2}\end{pmatrix} = -\Sigma_{g1}\Sigma_1^{-1}\Sigma_{12} + \Omega_{\text{OLS}}\Omega_{\text{IV}}^{-1}\Sigma_{g1}\Sigma_1^{-1}\Sigma_{12} - \Omega_{\text{OLS}}\Omega_{\text{IV}}^{-1}\Sigma_{gX}\begin{bmatrix}\Sigma^1\Sigma_{12}+\Sigma^{12}\Sigma_2\\ \Sigma^{21}\Sigma_{12}+\Sigma^2\Sigma_2\end{bmatrix} + \Sigma_{g2}
$$

$$
= [\Sigma_{g2} - \Sigma_{g1}\Sigma_1^{-1}\Sigma_{12}] + \Omega_{\text{OLS}}\Omega_{\text{IV}}^{-1}\Sigma_{g1}\Sigma_1^{-1}\Sigma_{12} - \Omega_{\text{OLS}}\Omega_{\text{IV}}^{-1}[\Sigma_{g1} \vdots \Sigma_{g2}]\begin{bmatrix}0_{k_1\times k_2}\\ I_{k_2}\end{bmatrix}
$$

$$
= [\Sigma_{g2} - \Sigma_{g1}\Sigma_1^{-1}\Sigma_{12}] - \Omega_{\text{OLS}}\Omega_{\text{IV}}^{-1}[\Sigma_{g2} - \Sigma_{g1}\Sigma_1^{-1}\Sigma_{12}]
$$

$$
= (I_G - \Omega_{\text{OLS}}\Omega_{\text{IV}}^{-1})[\Sigma_{g2} - \Sigma_{g1}\Sigma_1^{-1}\Sigma_{12}]\,, \tag{A.20}
$$

$$
\delta\begin{pmatrix}\Sigma_{1g}\\ \Sigma_{2g}\\ \Sigma_{g}\end{pmatrix} = -\Sigma_{g1}\Sigma_1^{-1}\Sigma_{1g} + \Omega_{\text{OLS}}\Omega_{\text{IV}}^{-1}\Sigma_{g1}\Sigma_1^{-1}\Sigma_{1g} - \Omega_{\text{OLS}}\Omega_{\text{IV}}^{-1}\Sigma_{gX}\Sigma_X^{-1}\Sigma_{Xg} + \Sigma_g
$$

$$
= -\Sigma_{g1}\Sigma_1^{-1}\Sigma_{1g} - \Omega_{\text{OLS}}\Omega_{\text{IV}}^{-1}(\Sigma_{gX}\Sigma_X^{-1}\Sigma_{Xg} - \Sigma_{g1}\Sigma_1^{-1}\Sigma_{1g}) + \Sigma_g
$$

$$
= -\Sigma_{g1}\Sigma_1^{-1}\Sigma_{1g} - \Omega_{\text{OLS}}\Omega_{\text{IV}}^{-1}\Omega_{\text{IV}} + \Sigma_g = -\Sigma_{g1}\Sigma_1^{-1}\Sigma_{1g} - \Omega_{\text{OLS}} + \Sigma_g = -\Sigma_V \tag{A.21}
$$

and so

$$
\delta\begin{pmatrix}\Sigma_1 & \Sigma_{12} & \Sigma_{1g}\\ \Sigma_{21} & \Sigma_2 & \Sigma_{2g}\\ \Sigma_{g1} & \Sigma_{g2} & \Sigma_g\end{pmatrix}\delta' = \left[0 \vdots \left(I_G - \Omega_{\text{OLS}}\Omega_{\text{IV}}^{-1}\right)\left(\Sigma_{g2} - \Sigma_{g1}\Sigma_1^{-1}\Sigma_{12}\right) \vdots\ -\Sigma_V\right]\delta'
$$

$$
= -(I_G - \Omega_{\text{OLS}}\Omega_{\text{IV}}^{-1})[\Sigma_{g2} - \Sigma_{g1}\Sigma_1^{-1}\Sigma_{12}]\left[\Sigma^{21} \vdots \Sigma^2\right]\begin{bmatrix}\Sigma_{1g}\\ \Sigma_{2g}\end{bmatrix}\Omega_{\text{IV}}^{-1}\Omega_{\text{OLS}} - \Sigma_V
$$

$$
= -(I_G - \Omega_{\text{OLS}}\Omega_{\text{IV}}^{-1})\Big[\Sigma_{g2}\Sigma^{21}\Sigma_{1g} + \Sigma_{g2}\Sigma^2\Sigma_{2g}
$$

$$
- \Sigma_{g1}\Sigma_1^{-1}\Sigma_{12}\Sigma^{21}\Sigma_{1g} - \Sigma_{g1}\Sigma_1^{-1}\Sigma_{12}\Sigma^2\Sigma_{2g}\Big]\Omega_{\text{IV}}^{-1}\Omega_{\text{OLS}} - \Sigma_V\,. \tag{A.22}
$$

Since

$$
\Sigma_{g2}\Sigma^{21}\Sigma_{1g} + \Sigma_{g2}\Sigma^2\Sigma_{2g} = \Omega_{\text{IV}} - \Sigma_{g1}\Sigma^1\Sigma_{1g} - \Sigma_{g1}\Sigma^{12}\Sigma_{2g} + \Sigma_{g1}\Sigma_1^{-1}\Sigma_{1g}\,, \tag{A.23}
$$

we have

$$
\begin{aligned}
&\Sigma_{g2}\Sigma^{21}\Sigma_{1g} + \Sigma_{g2}\Sigma^{2}\Sigma_{2g} - \Sigma_{g1}\Sigma_1^{-1}\Sigma_{12}\Sigma^{21}\Sigma_{1g} - \Sigma_{g1}\Sigma_1^{-1}\Sigma_{12}\Sigma^{2}\Sigma_{2g} \\
&= \Omega_{\text{IV}} - \Sigma_{g1}(\Sigma^{1} + \Sigma_1^{-1}\Sigma_{12}\Sigma^{21})\Sigma_{1g} - \Sigma_{g1}(\Sigma^{12} + \Sigma_1^{-1}\Sigma_{12}\Sigma^{2})\Sigma_{2g} + \Sigma_{g1}\Sigma_1^{-1}\Sigma_{1g}\,. \qquad \text{(A.24)}
\end{aligned}
$$

Using block-inversion formulas [*e.g.*, Abadir and Magnus (2005), Exercise 5.16, p. 106], we can check that

$$
\Sigma^{1} + \Sigma_1^{-1}\Sigma_{12}\Sigma^{21} = \Sigma_1^{-1} \quad \text{and} \quad \Sigma^{12} + \Sigma_1^{-1}\Sigma_{12}\Sigma^{2} = 0_{k_1\times k_2}\,, \qquad \text{(A.25)}
$$

which in turn implies

$$
\Sigma_{g2}\Sigma^{21}\Sigma_{1g} + \Sigma_{g2}\Sigma^{2}\Sigma_{2g} - \Sigma_{g1}\Sigma_1^{-1}\Sigma_{12}\Sigma^{21}\Sigma_{1g} - \Sigma_{g1}\Sigma_1^{-1}\Sigma_{12}\Sigma^{2}\Sigma_{2g} = \Omega_{\text{IV}}\,. \qquad \text{(A.26)}
$$

We finally arrive at

$$
\begin{aligned}
\delta\,\bar{\Sigma}\,\delta' &= (\Omega_{\text{OLS}}\Omega_{\text{IV}}^{-1} - I_G)\Omega_{\text{IV}}\Omega_{\text{IV}}^{-1}\Omega_{\text{OLS}} - \Sigma_V \\
&= \Omega_{\text{OLS}}\Omega_{\text{IV}}^{-1}\Omega_{\text{OLS}} - \Omega_{\text{OLS}} - \Sigma_V\,. \qquad \text{(A.27)}
\end{aligned}
$$

Putting everything together, we have the following asymptotic covariance for the LHS of (A.15):

$$
\sigma_u^2\Sigma_V + \sigma_{Vu}\sigma_{uV} + \sigma_u^2(\Omega_{\text{OLS}}\Omega_{\text{IV}}^{-1}\Omega_{\text{OLS}} - \Omega_{\text{OLS}} - \Sigma_V) = \sigma_{Vu}\sigma_{uV} + \sigma_u^2(\Omega_{\text{OLS}}\Omega_{\text{IV}}^{-1}\Omega_{\text{OLS}} - \Omega_{\text{OLS}})\,. \quad \text{(A.28)}
$$

□