*Econometric Theory*, **26**, 2010, 1820–1837.
doi:10.1017/S0266466609990806

# A NEW PROJECTION-TYPE SPLIT-SAMPLE SCORE TEST IN LINEAR INSTRUMENTAL VARIABLES REGRESSION

SARASWATA CHAUDHURI
*University of North Carolina—Chapel Hill*

THOMAS RICHARDSON
*University of Washington*

JAMES ROBINS
*Harvard University*

ERIC ZIVOT
*University of Washington*

In this paper we introduce a new method of projection-type inference and describe it in the context of two stage least squares–based split-sample inference on subsets of structural coefficients in a linear instrumental variables regression model. The use of the new method not only guards against the uncontrolled overrejection of the true value of the parameters of interest but also reduces the conservativeness of the usual method of projection proposed by Dufour and his coauthors (Dufour, 1997, *Econometrica* 65, 1365–1388; Dufour and Jasiak, 2001, *International Economic Review* 41, 815–843; Dufour and Taamouti, 2005, discussion paper; Dufour and Taamouti, 2005, *Econometrica* 73, 1351–1365; Dufour and Taamouti, 2007, *Journal of Econometrics* 139, 133–153).

## 1. INTRODUCTION

Dufour (1997), Dufour and Jasiak (2001), and Dufour and Taamouti (2005b, 2007) show that projection based on a test for a set of parameters can be used to test for subsets of parameters. If the former test has correct size, then such a projection-type test for subsets of parameters cannot be oversized. The projection-type test can subsequently be inverted to obtain confidence regions having (at least) the correct coverage probability. We refer to this method of inference as the “usual” (method of) projection-type inference for subsets of parameters.

We thank the seminar participants at the economics department of the University of Washington and the session participants at the summer meeting (2007) of the North American Econometric Society for their comments. Comments and suggestions from the coeditor and two anonymous referees improved the paper substantially. Research was supported by the Gary Waterman Distinguished Scholar Fund, a seed grant from the Center for Statistics and the Social Sciences at the University of Washington (PI: Zivot), NSF grant DMS 0505865 (PI: Richardson), NIH grant R01 AI032475 (PI: Robins). Address correspondence to Eric Zivot, Department of Economics, Box 353330, University of Washington, Seattle, WA 98195-3330, USA; e-mail: ezivot@u.washington.edu.

  0266-4666/10 $15.00

However, in spite of the attractiveness in terms of size and coverage, the usual method of projection-type inference can often be too conservative (see, e.g., Moreira, 2003; Zivot, Startz, and Nelson, 2006). In this paper we address the problem of conservativeness. We propose a new method of projection-type asymptotic inference that is quite generally less conservative than the usual method of projection, and at the same time retains the desirable characteristics in terms of size of the resulting tests and coverage probability of the derived confidence regions. The idea behind this new method of projection-type asymptotic inference is derived from Robins (2004).

We introduce this new method in the context of split-sample-based inference on structural coefficients in a linear instrumental variables (IV) regression. In particular, we restrict the focus of this paper to projection-type inference based on the "split-sample statistic" for structural coefficients corresponding to the endogenous regressors.

Linear IV regression with "weak instruments" has received considerable attention recently. In the presence of weak instruments, the standard techniques of asymptotic inference on structural coefficients are a poor guide to finite-sample inference (see, among others, Dufour, 1997; Zivot, Startz, and Nelson, 1998). However, the split-sample statistic can be used to (asymptotically) test for structural coefficients without overrejecting their true value (even in finite samples) (see Staiger and Stock, 1997; Dufour and Jasiak, 2001; Kleibergen, 2002).

The split-sample statistic can be also interpreted as a split-sample version of the two-stage-least-squares (TSLS) score statistic considered by Wang and Zivot (1998) and Dufour and Taamouti (2005b). This split-sample (TSLS) score statistic provides a simple and interesting framework for the exposition of our method of projection-type inference.[1]

In Section 2, we state the linear IV model with multiple structural coefficients. We describe the usual method of projection-type split-sample score test for subsets of structural coefficients in Section 3, and we describe our method of projection-type split-sample score test for subsets of structural coefficients and discuss the asymptotic properties of the test in Section 4. Monte Carlo experiments in Section 5 indicate that the asymptotic results from Section 4 provide a good approximation to the finite-sample behavior of our projection-type split-sample score test.

We use the following notation throughout. For any $n \times m$ matrix $A$, let $P(A) = A(A'A)^{-1}A'$ and $N(A) = I_n - P(A)$.

## 2. LINEAR IV MODEL AND ASSUMPTIONS

Consider the following model:

$$\left.\begin{aligned} y &= X\beta + W\gamma + u \\ X &= Z\Pi_x + V_x \\ W &= Z\Pi_w + V_w \end{aligned}\right\}, \tag{2.1}$$

where $y$ is the dependent variable, $X$ and $W$ are the endogenous regressors, $u$, $V_x$, and $V_w$ are the unobserved correlated structural errors, and $Z$ is the matrix of instruments.[2] Let the dimensions of $\beta$, $\gamma$, $\Pi_x$, and $\Pi_w$ be, respectively, $m_x \times 1$, $m_w \times 1$, $k \times m_x$, and $k \times m_w$. Let $m = m_x + m_w$ and $m_x, m_w$, and $k$ be fixed and finite integers. We assume that the order condition $k \geq m$ is satisfied. We do not, however, impose the restriction of full column rank on $\Pi = [\Pi_x, \Pi_w]$.

Suppose that there are $n$ observations on $y$, $X$, $W$, and $Z$ and we randomly split the sample into two subsamples—the first one containing $n_1$ observations and the second one containing $n_2 = n - n_1$ observations such that $\min\{n_1, n_2\} > k$ and $\lim_{n\to\infty} n_1/n = \zeta \in (0,1)$ is a fixed number. Let $y_i$, $X_i$, $W_i$, and $Z_i$ represent the matrices containing the $n_i$ observations in the $i$th subsample ($i = 1, 2$) where the observations are stacked in rows.

Without loss of generality, let $\beta$ be the parameters of interest. We are concerned with the projection-type split-sample methods for testing hypotheses of the form $H : \beta = \beta_0$ and subsequently inverting the tests to obtain confidence regions for arbitrary functions of $\beta$.

The sole purpose of this paper is to propose a modification to the usual method of projection-type inference to reduce its conservativeness. The split-sample TSLS score test in a linear IV regression provides a relatively simple framework for the exposition of our method. The simplicity of the framework, however, comes at a cost—loss of information. As a result, our method, in this context, does not necessarily lead to a test as powerful as, say, the subset-K test of Kleibergen (2004) or the tests described in Chaudhuri and Zivot (2008) and Chaudhuri (2009).[3] Therefore, the reference to local optimality, while we discuss the asymptotic properties of our method, is confined to the somewhat restrictive framework of split-sample TSLS score tests for $H : \beta = \beta_0$ (treating $\gamma$ as unknown).

The discussion on the asymptotic properties is facilitated by the following set of high-level assumptions on the joint asymptotic behavior of the structural errors and the instruments. We summarize them under Assumption M.

**Assumption M** (Structural errors and instruments). The following convergence results hold jointly as $n \to \infty$ for $i = 1, 2$:

M1. $n_i^{-1}(u_i, V_{xi}, V_{wi})'(u_i, V_{xi}, V_{wi}) \xrightarrow{P} \Sigma = \begin{pmatrix} \sigma_{uu} & \sigma_{ux} & \sigma_{uw} \\ \sigma_{xu} & \sigma_{xx} & \sigma_{xw} \\ \sigma_{wu} & \sigma_{wx} & \sigma_{ww} \end{pmatrix}$ where $\Sigma$ is a symmetric, positive definite matrix.

M2. $n_i^{-1} Z_i' Z_i \xrightarrow{P} Q$ where $Q$ is a symmetric, positive definite matrix.

M3. $n_i^{-1/2} Z_i'(u_i, V_{xi}, V_{wi}) \xrightarrow{d} Q^{1/2}(\Psi_{Zui}, \Psi_{Zxi}, \Psi_{Zwi})$ where

$\mathrm{vec}(\Psi_{Zui}, \Psi_{Zxi}, \Psi_{Zwi}) \sim N(0, \Sigma \otimes I_k)$.

M4. $\Psi_{Zu1}, \Psi_{Zx1}, \Psi_{Zw1}$ are uncorrelated with $\Psi_{Zx2}$ and $\Psi_{Zw2}$.

See Staiger and Stock (1997) and Kleibergen (2002) for discussion of Assumptions M1–M3. Assumption M4 ensures that the relevant random functions based on subsample 1 are asymptotically uncorrelated with those based on subsample 2.

It is well known that under Assumption M, the standard techniques of asymptotic inference on the structural coefficients degenerate when $\Pi$ is rank deficient (see Phillips, 1989; Choi and Phillips, 1992).

Of more practical interest is the case where $\Pi$ is close to being rank deficient (i.e., near rank deficient). Because not all convergences are uniform in $\Pi$ the standard first-order asymptotic theory, which treats $\Pi$ as fixed, provides poor approximation to the finite-sample behavior of the estimators and the tests for the structural coefficients. In particular, the asymptotic size of the standard Wald, likelihood ratio (LR), and score tests for the structural coefficients can hugely underestimate the size in finite samples (see, e.g., Staiger and Stock, 1997; Zivot et al., 1998).

To understand the properties of the projection-type tests, it is useful to characterize the near rank deficiency of $\Pi_x$ and $\Pi_w$ following the weak-instrument framework of Staiger and Stock (1997). This is summarized under Assumption WI.

**Assumption WI** (Partial identification of $\beta$ and $\gamma$). $\Pi_x = 0_{k\times m_x}1_{[\delta_x=0]} + (\mathbb{C}_x/\sqrt{n})1_{[\delta_x=1/2]} + \mathbb{C}_x 1_{[\delta_x=1]}$ and $\Pi_w = 0_{k\times m_w}1_{[\delta_w=0]} + (\mathbb{C}_w/\sqrt{n})1_{[\delta_w=1/2]} + \mathbb{C}_w 1_{[\delta_w=1]}$ where $\mathbb{C}_x$ and $\mathbb{C}_w$ are $k\times m_x$ and $k\times m_w$ matrices of fixed and bounded elements such that $\mathbb{C} = [\mathbb{C}_x, \mathbb{C}_w]$ is full column rank. The terms $\delta_x$ and $\delta_w$ are constants such that $1_{[\delta_x=0]} + 1_{[\delta_x=1/2]} + 1_{[\delta_x=1]} = 1$ and $1_{[\delta_w=0]} + 1_{[\delta_w=1/2]} + 1_{[\delta_w=1]} = 1$.

The nonrandom indicator functions involving the $\delta$'s delineate the nine cases of partial identification of the structural coefficients $\beta$ and $\gamma$: the (asymptotic) rank deficiency of $\Pi_x$ leads to the (asymptotic) nonidentification of $\beta$, and similarly the (asymptotic) rank deficiency of $\Pi_w$ leads to the (asymptotic) nonidentification of $\gamma$.[4] Under Assumption M, $\delta_x = \delta_w = 0$ (i.e., $\Pi_x = 0$ and $\Pi_w = 0$) corresponds to the case of complete unidentification and is referred to as the "leading case" by Phillips (1989). The case with $\delta_x = \delta_w = 1$ (i.e., $\Pi_x = \mathbb{C}_x$ and $\Pi_w = \mathbb{C}_w$) corresponds to the standard linear IV regression, and the standard techniques of asymptotic inference can be reliably employed only under this case.

Although the preceding canonical representation of the different cases of (weak) partial identification in Assumption WI is by no means exhaustive, it is sufficiently rich to produce the nondegenerate asymptotic results in this paper. We maintain Assumptions M and WI throughout the paper.

## 3. THE USUAL METHOD OF PROJECTION

The split-sample statistic for $\beta$ and $\gamma$, considered by Staiger and Stock (1997) and Dufour and Jasiak (2001), is defined as

$$\mathcal{SSLM}(\beta,\gamma) = \frac{(y_1 - X_1\beta - W_1\gamma)' P\left([\widehat{X}_{12}, \widehat{W}_{12}]\right)(y_1 - X_1\beta - W_1\gamma)}{\frac{1}{n_1-k}(y_1 - X_1\beta - W_1\gamma)' N(Z_1)(y_1 - X_1\beta - W_1\gamma)}, \quad \textbf{(3.1)}$$

where $\widehat{X}_{1i} = Z_1\widehat{\Pi}_{xi}$, $\widehat{W}_{1i} = Z_1\widehat{\Pi}_{wi}$, $\widehat{\Pi}_{xi} = (Z_i'Z_i)^{-1}Z_i'X_i$, and $\widehat{\Pi}_{wi} = (Z_i'Z_i)^{-1}$ $Z_i'W_i$ for $i = 1,2$. The asymptotic test "based on generated regressors," proposed by Dufour and Jasiak (2001), rejects the null hypothesis $(\beta,\gamma) = (\beta_0,\gamma_0)$ against the alternative $(\beta,\gamma) \neq (\beta_0,\gamma_0)$ if $\mathcal{SSLM}(\beta_0,\gamma_0) > \chi^2_m(1-\alpha)$. Although Staiger and Stock (1997) call (3.1) the split-sample Anderson–Rubin statistic, it is more naturally interpreted as a split-sample version of the score statistic considered by Wang and Zivot (1998) and Dufour and Taamouti (2005b). This interpretation follows once we note that (3.1) is the score statistic for $(\beta,\gamma)$ when inference on these parameters is based on the objective function

$$\min_{\beta,\gamma} \frac{1}{2}(y_1 - X_1\beta - W_1\gamma)'[\widehat{X}_{12},\widehat{W}_{12}]([X_1,W_1]'[\widehat{X}_{12},\widehat{W}_{12}])^{-1} \times [\widehat{X}_{12},\widehat{W}_{12}]'(y_1 - X_1\beta - W_1\gamma).$$

Sample splitting ensures asymptotic independence between the normalized gradient of the objective function and the estimator of its variance even under rank deficiency of $\Pi$, and thus, under Assumptions M and WI, $\mathcal{SSLM}(\beta,\gamma) \xrightarrow{d} \chi^2_m$.

Following Dufour and Taamouti (2005b, 2007), the usual projection-type split-sample score test for the null hypothesis $H:\beta = \beta_0$ against the alternative $K:\beta \neq \beta_0$ can be defined as

$$\text{reject } H:\beta = \beta_0 \quad \text{against } K:\beta \neq \beta_0 \quad \text{if } \inf_{\gamma_0 \in \mathbb{R}^{m_w}} \mathcal{SSLM}(\beta_0,\gamma_0) > \chi^2_m(1-\alpha), \tag{3.2}$$

and a confidence region for any arbitrary function $g(\beta)$ of $\beta$ can be obtained as

$$\mathcal{C}(g(\beta),1-\alpha) = \left\{ g(\beta_0) : \inf_{\gamma_0 \in \mathbb{R}^{m_w}} \mathcal{SSLM}(\beta_0,\gamma_0) \leq \chi^2_m(1-\alpha) \right\}. \tag{3.3}$$

Analytic methods for computing (3.3) are discussed in Dufour and Taamouti (2005b). Under Assumptions M and WI, the usual projection-type split-sample score test in (3.2) has asymptotic size of at most $\alpha$, and hence the confidence region in (3.3) has asymptotic coverage probability of at least $1-\alpha$.

However, our simulations in Section 5 reveal that the test can be very conservative. The conservativeness of the usual method of projection can be attributed to two factors: (i) the degrees of freedom implicitly used for the test is $m$, which can be much larger than the number of restrictions being tested, i.e., $m_x$; and (ii) the split-sample statistic $\mathcal{SSLM}(\beta_0,\gamma)$, which asymptotically converges to $\chi^2_m$ under the null, can be much larger than the unrestricted infimum $\inf_{\gamma_0 \in \mathbb{R}^{m_w}} \mathcal{SSLM}$ $(\beta_0,\gamma_0)$.[5]

## 4. AN ALTERNATIVE METHOD OF PROJECTION

Consider testing the null hypothesis $H:\beta = \beta_0$ against the alternative $K:\beta \neq \beta_0$. The size-$\alpha$ TSLS score test, based on subsample 1, rejects the null hypothesis

$H:\beta=\beta_0$ against the alternative $K:\beta\neq\beta_0$ if $\mathcal{LM}_\beta(\beta_0,\widehat{\gamma}_{11}(\beta_0))>\chi^2_{m_x}(1-\alpha)$ where

$$\mathcal{LM}_\beta(\beta,\gamma)=\frac{(y_1-X_1\beta-W_1\gamma)'P\left(N(\widehat{W}_{11})\widehat{X}_{11}\right)(y_1-X_1\beta-W_1\gamma)}{\frac{1}{n_1-k}(y_1-X_1\beta-W_1\gamma)'N(Z_1)(y_1-X_1\beta-W_1\gamma)}\quad\text{and}$$

$$\widehat{\gamma}_{11}(\beta)=\left(\widehat{W}_{11}'W_1\right)^{-1}\widehat{W}_{11}'(y_1-X_1\beta).$$

We note that $\mathcal{LM}_\beta(\beta,\gamma)$ is the efficient score statistic for $\beta$ and $\mathcal{LM}_\beta(\beta_0,\gamma_0)=\mathcal{LM}_\beta(\beta_0,\gamma)+o_p(1)$ for any $\gamma_0$ in a $\sqrt{n}$-neighborhood of $\gamma$, where $\mathcal{LM}_\beta(\beta_0,\gamma)$ is the infeasible efficient score statistic for $\beta$ that uses the unknown true value of $\gamma$. Under Assumptions M and WI, the statistic $\mathcal{LM}_\beta(\beta,\widehat{\gamma}_{11}(\beta))$ does not necessarily converge to a $\chi^2_{m_x}$ distribution unless $\Pi=\mathbb{C}$.

One way to (partly) avoid the problem of (near) rank deficiency of $\Pi_x$ and $\Pi_w$ is to replace $\widehat{X}_{11}$ and $\widehat{W}_{11}$ by $\widehat{X}_{12}$ and $\widehat{W}_{12}$, respectively in $\mathcal{LM}_\beta(\beta,\gamma)$ and $\widehat{\gamma}_{11}(\beta)$ (see Angrist and Krueger, 1995; Dufour and Jasiak, 2001). The resulting test rejects $H:\beta=\beta_0$ against $K:\beta\neq\beta_0$ if $\mathcal{SSLM}_\beta(\beta_0,\widehat{\gamma}_{12}(\beta_0))>\chi^2_{m_x}(1-\alpha)$ where

$$\mathcal{SSLM}_\beta(\beta,\gamma)=\frac{(y_1-X_1\beta-W_1\gamma)'P\left(N(\widehat{W}_{12})\widehat{X}_{12}\right)(y_1-X_1\beta-W_1\gamma)}{\frac{1}{n_1-k}(y_1-X_1\beta-W_1\gamma)'N(Z_1)(y_1-X_1\beta-W_1\gamma)}\quad\text{and}\tag{4.1}$$

$$\widehat{\gamma}_{12}(\beta_0)=\left(\widehat{W}_{12}'W_1\right)^{-1}\widehat{W}_{12}'(y_1-X_1\beta_0).$$

Although this leads to a size-$\alpha$ test irrespective of the (near) rank deficiency of $\Pi_x$, the problem with the (near) rank deficiency of $\Pi_w$ persists. This is because $\widehat{\gamma}_{12}(\beta)$ is inconsistent unless $\Pi_w=\mathbb{C}_w$. Chaudhuri, Richardson, Robins, and Zivot (2007) call this the unbiased split-sample instrumental variables (USSIV) score test and point out that the inconsistency of $\widehat{\gamma}_{12}(\beta)$ can cause severe upward size distortion, especially if the regressors $W$ are highly endogenous.

Under Assumption M, when $\Pi_w=\mathbb{C}_w$ and $\beta_0=\beta+O(n^{-1/2})$, we show in the Appendix that

$$\mathcal{SSLM}_\beta(\beta_0,\widehat{\gamma}_{12}(\beta_0))=\mathcal{SSLM}_\beta(\beta_0,\gamma)+o_p(1),\tag{4.2}$$

and hence the USSIV score test is (locally) asymptotically equivalent to the size-$\alpha$ "infeasible" split-sample score test that rejects $H:\beta=\beta_0$ against $K:\beta\neq\beta_0$ if $\mathcal{SSLM}_\beta(\beta_0,\gamma)>\chi^2_{m_x}(1-\alpha)$. The latter test is infeasible because it uses the unknown true value of $\gamma$. The (local) asymptotic equivalence in (4.2) follows from standard contiguity arguments once we note that $\widehat{\gamma}_{12}(\beta_0)$ is $\sqrt{n}$-consistent for $\gamma$ whenever Assumption M holds, $\Pi_w=\mathbb{C}_w$, and $\beta_0=\beta+O(n^{-1/2})$. Further, when $\Pi_w=\mathbb{C}_w$, the diameter of the confidence region based on inverting the USSIV score test corresponds to the semiparametric variance bound for $\beta$ based on $n_1$ observations in model (2.1).

Hence, although the USSIV score test for $H : \beta = \beta_0$ should not be used unless $\Pi_w = \mathbb{C}_w$, it provides a valuable insight: if the unknown $\gamma$ is replaced by a $\sqrt{n}$-consistent estimator in $\mathcal{SSLM}_\beta(\beta_0, \gamma)$ then (local) asymptotic equivalence with the infeasible split-sample score test can be achieved. This motivates our new projection-type split-sample score test that achieves (local) asymptotic equivalence with the infeasible split-sample score test when $\Pi_w = \mathbb{C}_w$. Because it is not possible to find a ($\sqrt{n}$-)consistent estimator of $\gamma$ unless $\Pi_w = \mathbb{C}_w$, the use of the projection technique in our new test at least guards against the uncontrolled overrejection of the true value of the parameters of interest $\beta$ that occurs in the USSIV test.

### 4.1. The New Projection-Type Split-Sample Score Test

The new projection-type split-sample score test rejects the null hypothesis $H : \beta = \beta_0$ against the alternative $K : \beta \neq \beta_0$ if

$$\inf_{\gamma_0 \in \mathcal{C}(\gamma, 1-\tau, \beta_0)} \mathcal{SSLM}_\beta(\beta_0, \gamma_0) > \chi^2_{m_x}(1-\epsilon), \tag{4.3}$$

where $\mathcal{C}(\gamma, 1-\tau, \beta_0)$ is a confidence region for $\gamma$ (restricted by $H : \beta = \beta_0$) such that

$$\mathcal{C}(\gamma, 1-\tau, \beta_0) = \left\{ \gamma_0 : \mathcal{SSLM}^*_\gamma(\beta_0, \gamma_0) \leq \chi^2_{m_w}(1-\tau) \right\} \quad \text{and} \tag{4.4}$$

$$\mathcal{SSLM}^*_\gamma(\beta, \gamma) = \frac{(y_1 - X_1\beta - W_1\gamma)' P(\widehat{W}_{12})(y_1 - X_1\beta - W_1\gamma)}{\frac{1}{n_1 - k}(y_1 - X_1\beta - W_1\gamma)' N(Z_1)(y_1 - X_1\beta - W_1\gamma)}. \tag{4.5}$$

This can be seen as a two-step procedure: in the first stage we construct a restricted confidence region for $\gamma$ such that the region has correct asymptotic coverage probability $1-\tau$ under the null hypothesis $H : \beta = \beta_0$, and in the second step we reject the null hypothesis if the infimum (with respect to $\gamma_0$ in the confidence region) of the statistic $\mathcal{SSLM}_\beta(\beta_0, \gamma_0)$ is larger than the $\chi^2_{m_x}(1-\epsilon)$ critical value. This method is motivated by Theorem 5.1 in Robins (2004).

The new projection-type split-sample score test relies on projection based on the statistic $\mathcal{SSLM}_\beta(\beta_0, \gamma)$, whose reference distribution is $\chi^2_{m_x}$, whereas the usual projection-type test uses the statistic $\mathcal{SSLM}(\beta_0, \gamma_0)$, whose reference distribution is $\chi^2_m$. Furthermore, unlike the usual methods of projection, here we project from a restricted space, a confidence region for $\gamma$, and not from the entire parameter space for $\gamma$. Hence, this new projection-type split-sample score test is, by construction, at least as powerful as the projection-type test that rejects $H : \beta = \beta_0$ against $K : \beta \neq \beta_0$ if $\inf_{\gamma_0 \in \mathbb{R}^{m_w}} \mathcal{SSLM}_\beta(\beta_0, \gamma_0) > \chi^2_{m_x}(1-\epsilon)$.

Theorem 4.1 is helpful to understand the asymptotic properties of our projection-type split-sample score test described in (4.3)–(4.4) under Assumptions M and WI.

THEOREM 4.1. *Let* $0 < \epsilon, \tau < 1$, *and* $\beta_0 = \beta - n(\delta_x)^{-1} d_\beta$ *where* $n(\delta_x) = n^{1/2} 1_{[\delta_x = 1]} + \left(1 - 1_{[\delta_x=1]}\right)$ *and* $d_\beta \in \mathbb{R}^{m_x}$. *Under Assumptions M and WI,*

*(i)* $\lim_{n\to\infty} \Pr_{\beta,\gamma}\left[\mathcal{SSLM}^*_\gamma(\beta,\gamma) > \chi^2_{m_w}(1-\tau)\right] = \tau$,

*(ii)* $\lim_{n\to\infty} \Pr_{\beta,\gamma}\left[\mathcal{SSLM}_\beta(\beta,\gamma) > \chi^2_{m_x}(1-\epsilon)\right] = \epsilon$, *and*

*(iii) if* $\Pi_w = \mathbb{C}_w$, *then* $\inf_{\gamma_0 \in \mathcal{C}(\gamma, 1-\tau, \beta_0)} \mathcal{SSLM}_\beta(\beta_0, \gamma_0) = \mathcal{SSLM}_\beta (\beta_0, \gamma) + o_p(1)$.

Theorem 4.1 holds for arbitrary "true values" $\beta \in \mathbb{R}^{m_x}$ and $\gamma \in \mathbb{R}^{m_w}$. Hence using Bonferroni's inequality, it follows from (i) and (ii) that the asymptotic size of our projection-type split-sample score test (described in (4.3)–(4.4)) cannot exceed $\tau + \epsilon$. Furthermore, (iii) implies that when $\Pi_w = \mathbb{C}_w$, then our projection-type split-sample score test is asymptotically equivalent to the size-$\epsilon$ "infeasible" split-sample score test.

A confidence region for any arbitrary function $g(\beta)$ of $\beta$ can be obtained by inverting the new projection-type split-sample score test as

$$\mathcal{C}(g(\beta), \tau+\epsilon) = \left\{ g(\beta_0) : \inf_{\gamma_0 \in \mathcal{C}(\gamma, 1-\tau, \beta_0)} \mathcal{SSLM}_\beta(\beta_0, \gamma_0) \leq \chi^2_{m_x}(1-\epsilon) \right\}. \tag{4.6}$$

This is a conservative $1 - (\tau + \epsilon)$ confidence region for $g(\beta)$. Moreover, when $\Pi_w = \mathbb{C}_w$, it follows from Theorem 4.1(iii) that the asymptotic length and coverage of this region are same as those of the infeasible region $\left\{g(\beta_0) : \mathcal{SSLM}_\beta (\beta_0, \gamma) \leq \chi^2_{m_x}(1-\epsilon)\right\}$ obtained by inverting the size-$\epsilon$ infeasible split-sample score test.

**Remarks.**

1. If $\alpha$ is the maximum allowable asymptotic size for testing $H : \beta = \beta_0$, then one should choose $\tau$ and $\epsilon$ such that $\tau + \epsilon = \alpha$. Although an analytical discussion on the choice of $\tau$ and $\epsilon$ is beyond the scope of this paper, we can at least conclude that when $\Pi_w = \mathbb{C}_w$, the choice of $\tau$ does not matter asymptotically.
2. All the split-sample tests mentioned in this paper treat subsample 1 as the working sample; information from subsample 2 is used only to deal with the (asymptotic) rank deficiency of $\Pi$. Hence the power of our test increases with the proportion of observations in subsample 1. In fact, when $\Pi_x = \mathbb{C}_x$ and $\Pi_w = \mathbb{C}_w$, i.e., in a standard linear IV regression, the noncentrality

parameter of the asymptotic $\chi^2_{m_x}$ distribution of $\inf_{\gamma_0 \in \mathcal{C}(\gamma,1-\tau,\beta_0)} \mathcal{SSLM}_\beta(\beta_0,\gamma_0)$ is $\zeta$ $(= \lim_{n\to\infty} n_1/n)$ times the noncentrality parameter of the asymptotic $\chi^2_{m_x}$ distribution of the standard score statistic that treats $\gamma$ as unknown (see eqn. (A.4) in the Appendix).

3. The rejection rule for the new projection-type split-sample score test can alternatively be expressed as follows: reject $H : \beta = \beta_0$ against $K : \beta \neq \beta_0$ if

$$\left\{\gamma_0 | \gamma_0' A_1 \gamma_0 - 2B_1\gamma_0 + C_1 \leq 0\right\} \cap \left\{\gamma_0 | \gamma_0' A_2 \gamma_0 - 2B_2\gamma_0 + C_2 \leq 0\right\} = \varnothing, \tag{4.7}$$

where $A_1 = W_1' H_1 W_1$, $B_1 = W_1' H_1 (y_1 - X_1\beta_0)$, $C_1 = (y_1 - X_1\beta_0)' H_1 (y_1 - X_1\beta_0)$, $H_1 = P(\widehat{W}_{12}) - (n_1 - k)^{-1} \chi^2_{m_w}(1-\tau) N(Z_1)$, $A_2 = W_1' H_2 W_1$, $B_2 = W_1' H_2 (y_1 - X_1\beta_0)$, $C_2 = (y_1 - X_1\beta_0)' H_2 (y_1 - X_1\beta_0)$, $H_2 = P(N(\widehat{W}_{12})\widehat{X}_{12}) - (n_1 - k)^{-1} \chi^2_{m_x}(1-\epsilon) N(Z_1)$, and $\varnothing$ stands for an empty set. This alternative representation does not require us to find the restricted infimum $\inf_{\gamma_0 \in \mathcal{C}(\gamma,1-\tau,\beta_0)} \mathcal{SSLM}_\beta(\beta_0,\gamma_0)$ explicitly as discussed by Dufour and Taamouti (2005b). As in the case of the usual projection-type split-sample score test, this may significantly reduce the computational cost of the new test.

## 4.2. Comparison with the Usual Method of Projection

The statistics on which the usual and the new projection-type split-sample score tests have been designed are different, and these two statistics converge to different distributions. As a result, a direct analytical comparison between the asymptotic conservativeness of the two tests is not possible without further assumptions. Hence we take recourse to Monte Carlo experiments in the next section to compare the performance of these two methods in finite samples.

However, when $\Pi_w = \mathbb{C}_w$ and $\epsilon = \alpha$, we can appeal to (4.2) and Theorem 4.1(iii) to conclude that for any fixed $\tau \in (0,1)$, the new projection-type split-sample score test is (locally) asymptotically at least as powerful as the usual projection-type split-sample score test. This is evident once we note that, when $\Pi_w = \mathbb{C}_w$ and $\beta_0 = \beta + O(n^{-1/2})$,

$$\inf_{\gamma_0 \in \mathcal{C}(\gamma,1-\tau,\beta_0)} \mathcal{SSLM}_\beta(\beta_0,\gamma_0) = \mathcal{SSLM}_\beta(\beta_0,\gamma) + o_p(1)$$

$$= \mathcal{SSLM}_\beta(\beta_0,\widehat{\gamma}_{12}(\beta_0)) + o_p(1) \quad \text{and}$$

$$\Pr_{\beta,\gamma}\left[\inf_{\gamma_0 \in \mathbb{R}^{m_w}} \mathcal{SSLM}(\beta_0,\gamma_0) \leq \chi^2_m(1-\alpha)\right]$$

$$\geq \Pr_{\beta,\gamma}\left[\mathcal{SSLM}(\beta_0,\widehat{\gamma}_{12}(\beta_0)) \leq \chi^2_m(1-\alpha)\right]$$

$$= \Pr_{\beta,\gamma}\left[\mathcal{SSLM}_{\beta}(\beta_0,\widehat{\gamma}_{12}(\beta_0)) \le \chi^2_m(1-\alpha)\right]$$

$$\ge \Pr_{\beta,\gamma}\left[\mathcal{SSLM}_{\beta}(\beta_0,\widehat{\gamma}_{12}(\beta_0)) \le \chi^2_{m_x}(1-\alpha)\right].$$

Simulations in the next section suggest that the reduction in conservativeness due to the new method can be significant in finite samples even for a small value of $\tau$.

The new method of projection-type split-sample score test hinges on two important considerations: (i) the projection is restricted to a $\sqrt{n}$-neighborhood of the true value of $\gamma$ whenever $\Pi_w = \mathbb{C}_w$, and (ii) in the second step we use the split-sample efficient score statistic, and not the usual score statistic, for the parameters of interest $\beta$. Whereas the first consideration helps in the reduction of conservativeness, the second one ensures asymptotic equivalence with the size-$\epsilon$ "infeasible" split-sample score test when $\Pi_w = \mathbb{C}_w$.

A natural question is, What happens if such a restricted projection is applied to the statistic (3.2) used in the usual method of projection? Of course this will reduce its conservativeness, but not as much. To see this, note that the confidence region $\mathcal{C}(\gamma, 1-\tau, \beta_0)$ always contains $\widehat{\gamma}_{12}(\beta_0)$ and hence when $\Pi_w = \mathbb{C}_w$ and $\beta_0 = \beta + O(n^{-1/2})$,

$$\inf_{\gamma_0\in\mathcal{C}(\gamma,1-\tau,\beta_0)} \mathcal{SSLM}(\beta_0,\gamma_0) \le \mathcal{SSLM}_{\beta}(\beta_0,\widehat{\gamma}_{12}(\beta_0))$$

$$= \inf_{\gamma_0\in\mathcal{C}(\gamma,1-\tau,\beta_0)} \mathcal{SSLM}_{\beta}(\beta_0,\gamma_0) + o_p(1).$$

Therefore, not only will our test statistic asymptotically (stochastically) dominate the statistic $\inf_{\gamma_0\in\mathcal{C}(\gamma,1-\tau,\beta_0)} \mathcal{SSLM}(\beta_0,\gamma_0)$, but also because of a less conservative critical value, our method is as powerful as when the restricted projection is applied to the usual method.

## 5. FINITE-SAMPLE PROPERTIES: SIMULATION STUDY

In this section, we study the finite-sample properties of the usual and the new projection-type split-sample score tests using Monte Carlo methods. The simulations show that (1) the new test is not as conservative as the usual test and (2) in a standard linear IV model in which there is no rank deficiency of $\Pi_x$ and $\Pi_w$, the finite-sample power of the new projection-type split-sample score test "almost" attains the "infeasible power envelope" provided by the finite-sample power of the infeasible split-sample score test.

We consider a data generating process similar to that in Dufour and Taamouti (2005a) in which $m_x = m_w = 1$. This special case is the standard setup in simulation studies in the weak instrument literature and is common in empirical applications. The results in this section are also supported by a more extensive simulation study conducted by Chaudhuri et al. (2007).

We generate data from the model in (2.1) such that we have the following results.

1. The structural errors, $(u_t, V_{xt}, V_{wt}) \overset{\text{i.i.d.}}{\sim} N(0, \Sigma)$ for $t = 1, \ldots, n$ where $\sigma_{uu} = \sigma_{xx} = \sigma_{ww} = 1$, $\sigma_{ux} = \sigma_{uw} = 0.8$, and $\sigma_{xw} = 0.3$.
2. The first column of $Z$ is an $n \times 1$ column of ones, and the elements in the other $k - 1$ columns are generated as independent and identically distributed (i.i.d.) $N(0, 1)$ variables but are kept fixed over simulations. We report the results for $k = 4$ and $k = 10$ (for size comparison only). The results are similar for other choices of $k$ (not reported) that are not too large as compared to $n_1$ and $n_2$.
3. The matrix $\Pi$ is constructed such that $\Pi = \mathbb{C}/\sqrt{n}$ where $\mathbb{C} = [\mathbb{C}_x, \mathbb{C}_w]$, and the elements of $\mathbb{C}_j$ are set at 0, 1.1547, and 20 when $\delta_j = 0$, 1/2, and 1, respectively, for $j = X, W$. This satisfies the classification of "unidentification," "weak identification," and "strong identification" by Dufour and Taamouti (2005a).[6]
4. The structural coefficients are set at $\beta = 0.5$ and $\gamma = 1$.
5. We consider sample size $n = 100$ and randomly split the sample into two subsamples, the first one containing $n_1 = 75$ and the second one containing $n_2 = 25$ observations.
6. We report the simulation results based on 10,000 replications.

The results are summarized in Figure 1 and Table 1. The usual and new projection-type split-sample score tests never overreject the true value of $\beta$ even in finite samples. In fact, if the allowable rate of type-I error (ARTIE) is 5% (say), a gain in power can be achieved for the usual test by choosing less conservative critical values. The results are similar even if we consider sample sizes as large as 10,000 ($n_1 = 7{,}500$ and $n_2 = 2{,}500$). It is evident that the new method of projection is considerably less conservative than the usual method; e.g., the rejection rate of the new test with $(1\% + 5\% = )$ 6% ARTIE uniformly dominates the rejection rate of the usual projection-type split-sample score test with 10% ARTIE when $\gamma$ is strongly identified. Regarding the choice of $\tau$ and $\epsilon$: the conservativeness of the new test decreases more rapidly when $\epsilon$ increases. Moreover, when $\gamma$ is strongly identified, the effect of the choice of $\tau$ on the overall conservativeness of the new test seems to be negligible. The simulations provide more pronounced support to the (local) asymptotic equivalence between the new test and the infeasible split-sample score test when $\beta$ is weakly identified or strongly identified. Based on the preceding observations, better performance is achieved by fixing $\epsilon =$ ARTIE and setting a less conservative value (say, equal to ARTIE itself) for $\tau$ while testing the hypothesis $H : \beta = \beta_0$ using the new projection-type split-sample score test.

The simulations, without any exception, provide compelling evidence in support of the results discussed in this paper and show the usefulness of the new method of projection. The new method reduces the conservativeness of the usual method of projection by trimming the set of $\gamma$ over which the projection is made

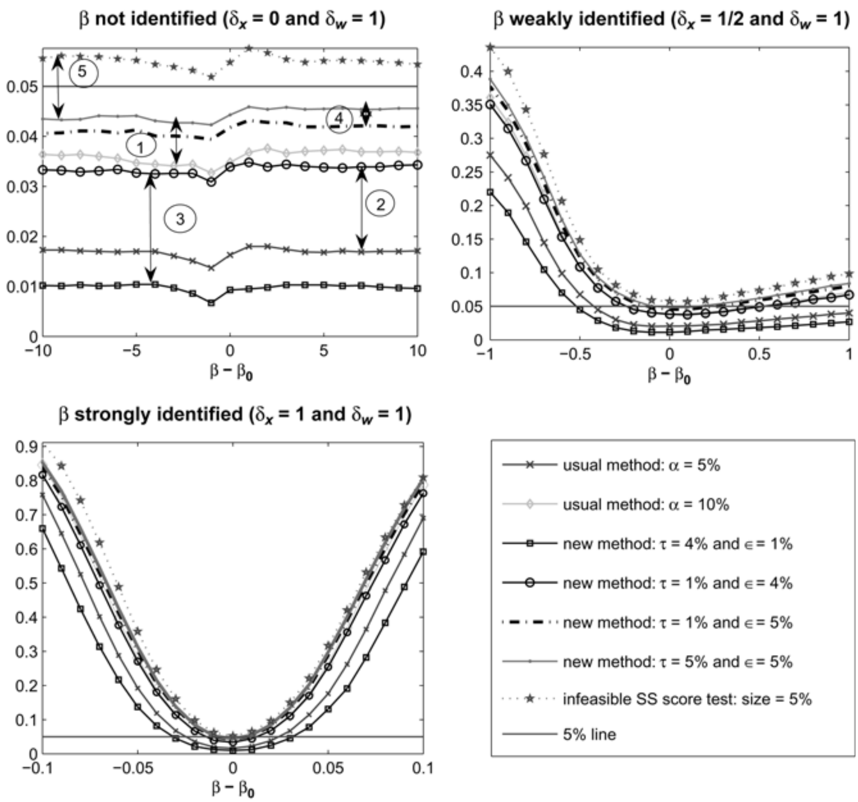


**FIGURE 1.** Rejection rates for $H: \beta = \beta_0$ when $\gamma$ is strongly identified ($n_1 = 75, n_2 = 25$, and $k = 4$). The pointers on the upper left plot respectively, point to the differences in rejection rates of (1) the usual and the new method with 10% upper bound on size, (2) the usual and the new method with 5% upper bound on size, (3) the new method with different $\tau$ and $\epsilon$ such that $\tau + \epsilon = 5\%$, (4) the new method with same $\epsilon$ but different $\tau$, (5) the new method and the power envelope (i.e., the infeasible split-sample score test).

and by decreasing the nominal critical value from the $1-\alpha$ quantile of a $\chi^2_m$ to that of a $\chi^2_{m_x}$. Such a reduction is likely to be even more significant when the dimension of $\gamma$ is large. In that sense the simulation results based on a scalar probably reflect the lower bound of the reduction in conservativeness due to our method.

## NOTES

1. The framework of this paper is based on TSLS, and we do not mention “TSLS” explicitly hereafter unless there is a possibility of confusion.

**TABLE 1.** Rejection rates for $H : \beta = \beta_0$ in finite samples ($n_1$ observations in subsample 1 and $n_2$ observations in subsample 2)

| Sample size | Identification status for $\beta$ | No. of instruments | Identification status for $\gamma$: Unidentified (i.e., $\delta_w = 0$) $k = 4$ | Unidentified $k = 10$ | Weakly identified (i.e., $\delta_w = 1/2$) $k = 4$ | Weakly identified $k = 10$ | Strongly identified (i.e., $\delta_w = 1$) $k = 4$ | Strongly identified $k = 10$ |
|---|---|---|---|---|---|---|---|---|
| $n = 100$ ($n_1 = 75$, $n_2 = 25$) | | | | | | | | |
| | unidentified (i.e., $\delta_x = 0$) | usual: 5% | 0.0 | 0.0 | 0.7 | 0.7 | 1.6 | 1.4 |
| | | usual: 10% | 0.0 | 0.0 | 2.1 | 1.7 | 3.5 | 3.5 |
| | | new (4% + 1%) | 0.0 | 0.0 | 0.0 | 0.0 | 0.9 | 0.8 |
| | | new (1% + 4%) | 0.0 | 0.0 | 0.0 | 0.0 | 3.4 | 3.3 |
| | | new (1% + 5%) | 0.0 | 0.0 | 0.2 | 0.1 | 4.2 | 4.2 |
| | | new (5% + 5%) | 0.0 | 0.0 | 0.5 | 0.4 | 4.4 | 4.5 |
| | | infeasible (5%) | 5.4 | 5.5 | 5.3 | 5.4 | 5.5 | 5.4 |
| | weakly identified (i.e., $\delta_x = 1/2$) | usual: 5% | 0.0 | 0.0 | 0.8 | 0.7 | 2.0 | 1.5 |
| | | usual: 10% | 0.2 | 0.1 | 2.1 | 1.8 | 3.9 | 3.4 |
| | | new (4% + 1%) | 0.0 | 0.0 | 0.2 | 0.0 | 1.1 | 0.8 |
| | | new (1% + 4%) | 0.0 | 0.0 | 0.2 | 0.0 | 3.8 | 3.2 |
| | | new (1% + 5%) | 0.0 | 0.0 | 0.2 | 0.1 | 4.6 | 4.1 |
| | | new (5% + 5%) | 0.0 | 0.0 | 1.0 | 0.4 | 4.8 | 4.3 |
| | | infeasible (5%) | 5.4 | 5.5 | 5.4 | 5.4 | 5.7 | 5.2 |
| | strongly identified (i.e., $\delta_x = 1$) | usual: 5% | 0.0 | 0.1 | 1.3 | 0.6 | 1.7 | 1.8 |
| | | usual: 10% | 0.2 | 0.2 | 3.0 | 1.6 | 3.4 | 3.5 |
| | | new (4% + 1%) | 0.0 | 0.0 | 0.2 | 0.1 | 1.0 | 1.1 |
| | | new (1% + 4%) | 0.0 | 0.0 | 0.2 | 0.1 | 3.4 | 3.7 |
| | | new (1% + 5%) | 0.0 | 0.0 | 0.2 | 0.1 | 4.2 | 4.5 |
| | | new (5% + 5%) | 0.0 | 0.0 | 0.8 | 0.4 | 4.4 | 4.7 |
| | | infeasible (5%) | 5.3 | 5.2 | 5.3 | 5.4 | 5.2 | 5.3 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $n = 10^4$ ($n_1 = 7500$, $n_2 = 2500$) | | | | | | | |
| | usual: 5% | 0.1 | 0.1 | 0.7 | 0.9 | 1.6 | 1.5 |
| | usual: 10% | 0.3 | 0.2 | 1.6 | 2.2 | 3.2 | 3.3 |
| | new (4% + 1%) | 0.0 | 0.0 | 0.0 | 0.1 | 0.9 | 0.8 |
| unidentified | new (1% + 4%) | 0.0 | 0.0 | 0.1 | 0.1 | 3.3 | 3.5 |
| (i.e., $\delta_x = 0$) | new (1% + 5%) | 0.0 | 0.0 | 0.1 | 0.1 | 4.0 | 4.5 |
| | new (5% + 5%) | 0.0 | 0.0 | 0.4 | 0.5 | 4.3 | 4.6 |
| | infeasible (5%) | 5.0 | 5.6 | 4.8 | 5.1 | 5.1 | 5.2 |
| | usual: 5% | 0.1 | 0.0 | 0.7 | 0.9 | 1.4 | 1.4 |
| | usual: 10% | 0.2 | 0.2 | 1.5 | 2.1 | 3.2 | 3.2 |
| | new (4% + 1%) | 0.0 | 0.0 | 0.0 | 0.0 | 0.9 | 0.8 |
| weakly | new (1% + 4%) | 0.0 | 0.0 | 0.0 | 0.1 | 3.0 | 3.4 |
| identified | new (1% + 5%) | 0.0 | 0.0 | 0.0 | 0.1 | 3.8 | 4.2 |
| (i.e., $\delta_x = 1/2$) | new (5% + 5%) | 0.1 | 0.0 | 0.4 | 0.6 | 4.2 | 4.4 |
| | infeasible (5%) | 5.4 | 5.1 | 4.8 | 5.2 | 5.2 | 5.0 |
| | usual: 5% | 0.1 | 0.1 | 0.6 | 1.1 | 1.6 | 1.5 |
| | usual: 10% | 0.2 | 0.1 | 1.5 | 2.8 | 3.3 | 3.2 |
| | new (4% + 1%) | 0.0 | 0.0 | 0.0 | 0.2 | 1.0 | 1.0 |
| strongly | new (1% + 4%) | 0.0 | 0.0 | 0.1 | 0.1 | 3.3 | 3.4 |
| identified | new (1% + 5%) | 0.0 | 0.0 | 0.1 | 0.1 | 4.0 | 4.3 |
| (i.e., $\delta_x = 1$) | new (5% + 5%) | 0.0 | 0.0 | 0.3 | 0.6 | 4.2 | 4.5 |
| | infeasible (5%) | 5.1 | 4.8 | 4.7 | 4.9 | 5.0 | 5.0 |

2. Adding included-exogenous variables to the model does not entail any fundamental change in our results because it is possible to find a $\sqrt{n}$-consistent, asymptotically unbiased estimator for the corresponding coefficients, even when the true values of $\beta$ and $\gamma$ are unknown.

3. Likewise, our focus on TSLS also excludes LR-based tests from consideration. These are the subject of active ongoing research on our part.

4. Under Assumptions M and WI, the parameters $\beta$ ($\gamma$) are identified as long as $\Pi_x = \mathbb{C}_x$ ($\Pi_w = \mathbb{C}_w$). However, the asymptotic normality of the unrestricted split-sample estimator (see Angrist and Krueger, 1995) of $\beta$ ($\gamma$) does not hold unless $\Pi_w = \mathbb{C}_w$ ($\Pi_x = \mathbb{C}_x$). A general result along this line can be found in Stock and Wright (2000).

5. The unrestricted infimum is the minimum eigenvalue of the matrix $A^{-1}B$ where $A = (n_1 - k)^{-1}(y_1 - X_1\beta_0, W_1)'N(Z_1)(y_1 - X_1\beta_0, W_1)$ and $B = (y_1 - X_1\beta_0, W_1)'P([\widehat{X}_{12}, \widehat{W}_{12}])(y_1 - X_1\beta_0, W_1)$. Nothing guarantees that its difference from $\mathcal{SSLM}(\beta_0, \gamma)$ is small unless $\Pi = \mathbb{C}$.

6. Although the magnitude of the minimum eigenvalue of the (population) concentration matrix, $\frac{\zeta}{k}[\mathrm{Var}(V_{xt}, V_{wt})]^{-1/2'}\mathbb{C}'Q\mathbb{C}[\mathrm{Var}(V_{xt}, V_{wt})]^{-1/2}$, is not meant to measure weak identification of individual structural coefficients, we note that our choice of $\Pi$ results in this minimum eigenvalue being (i) zero if any coefficient is unidentified and (ii) at most 3.35 if one coefficient is weakly identified and the other strongly identified (see Stock and Yogo, 2005). The (population) concentration parameter corresponding to any structural coefficient (ignoring the others) is modeled as 0 and 1, respectively, when that coefficient is unidentified and weakly identified.

## *REFERENCES*


Angrist, J. & A.B. Krueger (1995) Split-sample instrumental variables estimates of the return to schooling. *Journal of Business & Economics Statistics* 13, 225–235.

Chaudhuri, S. (2009) Projection-Based GEL Score Test for Subsets of Parameters with Possible Weak Identification. Manuscript, University of North Carolina.

Chaudhuri, S., T. Richardson, J. Robins, & E. Zivot (2007) Split-Sample Score Tests in Linear Instrumental Variables Regression. Technical report 73, CSSS, University of Washington.

Chaudhuri, S. & E. Zivot (2008) A New Method of Projection-Based Inference in GMM with Weakly Identified Nuisance Parameters. Manuscript, University of North Carolina.

Choi, I. & P.C.B. Phillips (1992) Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations. *Journal of Econometrics* 51, 113–150.

Dufour, J.M. (1997) Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica* 65, 1365–1388.

Dufour, J.M. & 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. & M. Taamouti (2005a) Further Results on Projection-Based Inference in IV Regressions with Weak, Collinear or Missing Instruments. Discussion paper, Université de Montréal.

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

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

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

Kleibergen, F. (2004) Testing subsets of parameters in the instrumental variables regression model. *Review of Economics and Statistics* 86, 418–423.

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

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

Robins, J.M. (2004) Optimal structural nested models for optimal sequential decisions. In D.Y. Lin & P. Heagerty (eds.), *Proceedings of the Second Seattle Symposium on Biostatistics*, pp. 189–326. Springer-Verlag.

Staiger, D. & J.H. Stock (1997) Instrumental variables regression with weak instruments. *Econometrica* 65, 557–586.
Stock, J.H. & J.H. Wright (2000) GMM with weak identification. *Econometrica* 68, 1055–1096.
Stock, J.H. & M. Yogo (2005) Testing for weak instruments in linear IV regression. In D.W.K. Andrews & J.H. Stock (eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pp. 80–108. Cambridge University Press.
Wang, J. & E. Zivot (1998) Inference on a structural parameter in instrumental variables regression with weak instruments. *Econometrica* 66, 1389–1404.
Zivot, E., R. Startz, & C. Nelson (1998) Valid confidence intervals and inference in the presence of weak instruments. *International Economic Review* 39, 1119–1144.
Zivot, E., R. Startz, & C. Nelson (2006) Inference in weakly identified instrumental variables regression. In D. Corbae, S.N. Durlauf, & B.E. Hansen (eds.), *Frontiers in Analysis and Applied Research: Essays in Honor of Peter C. B. Phillips*, pp. 125–166. Cambridge University Press.

# APPENDIX: Proofs

**Proof of Theorem 4.1.** We use the following definitions. Recall that $\lim_{n\to\infty} n_1/n = \zeta \in (0,1)$ is a fixed number. For $a = x, w$, we define $\lambda_a = Q^{1/2}\mathbb{C}_a$ and

$$n_1(\delta_a) = n_1^{1/2} 1_{[\delta_a=1]} + \left(1 - 1_{[\delta_a=1]}\right),$$

$$\nu_{a1} = \left(1 - 1_{[\delta_a=1]}\right)\Psi_{Za1} + \zeta^{1/2}\lambda_a\left((1 - 1_{[\delta_a=1]}) + \zeta^{-1/2}1_{[\delta_a=1]}\right),$$

$$\nu_{a2} = \left(1 - 1_{[\delta_a=1]}\right)\Psi_{Za2} + (1-\zeta)^{1/2}\lambda_a\left((1 - 1_{[\delta_a=1]}) + \zeta^{-1/2}1_{[\delta_a=1]}\right).$$

Using Assumptions M and WI, it follows that for $a, b = x, w$ and $A, B = X, W$,

(a) $n_1^{-1/2} n_2^{1/2} n_1(\delta_a)^{-1}\widehat{A}'_{12}u_1 \xrightarrow{d} \nu'_{a2}\Psi_{Zu1}$
(b) $n_1^{-1/2} n_2^{1/2} n_1(\delta_a)^{-1} n_1(\delta_b)^{-1}\widehat{A}'_{12}B_1 \xrightarrow{d} \nu'_{a2}\nu_{b1}$
(c) $n_1^{-1} n_2 n_1(\delta_a)^{-1} n_1(\delta_b)^{-1}\widehat{A}'_{12}\widehat{B}_{12} \xrightarrow{d} \nu'_{a2}\nu_{b2}$.

Further define $\Delta_n(\beta_0,\gamma_0) = [1,(\beta-\beta_0)',(\gamma-\gamma_0)']\Sigma[1,(\beta-\beta_0)',(\gamma-\gamma_0)']'$. Then Assumption M implies that $1/(n_1-k)(y_1 - X_1\beta_0 - W_1\gamma_0)'N(Z_1)(y_1 - X_1\beta_0 - W_1\gamma_0) - \Delta_n(\beta_0,\gamma_0) = o_p(1)$.

Part (i). By Slutsky's theorem we get

$$\mathcal{SSLM}^*_\gamma(\beta_0,\gamma_0) \xrightarrow{d} \lim_{n\to\infty} \frac{\phi_1(\beta_0,\gamma_0)'P(\nu_{w2})\phi_1(\beta_0,\gamma_0)}{\Delta_n(\beta_0,\gamma_0)}, \tag{A.1}$$

where $\phi_1(\beta_0,\gamma_0) = \Psi_{Zu1} + \nu_{x1}(\beta-\beta_0)n_1(\delta_x) + \nu_{w1}(\gamma-\gamma_0)n_1(\delta_w)$. Using Assumption M4, it follows that $\mathcal{SSLM}^*_\gamma(\beta,\gamma) \xrightarrow{d} \sigma_{uu}^{-1}\Psi'_{Zu1}P(\nu_{w2})\Psi_{Zu1} \sim \chi^2_{m_w}$ conditional on $\Psi_{Zw2}$ (and hence unconditionally). This implies

$$\begin{aligned} 1 - \lim_{n\to\infty}\Pr_{\beta,\gamma}\left[\mathcal{SSLM}^*_\gamma(\beta,\gamma) \le \chi^2_{m_w}(1-\tau)\right] \\ = 1 - \Pr\left[\chi^2_{m_w} \le \chi^2_{m_w}(1-\tau)\right] = \tau. \end{aligned}$$

Part (ii). By Slutsky's theorem we get

$$\mathcal{SSLM}_{\beta}(\beta_0,\gamma_0)\xrightarrow{d}\lim_{n\to\infty}\frac{\phi_1(\beta_0,\gamma_0)'P\left(N(\nu_{w2})\nu_{x2}\right)\phi_1(\beta_0,\gamma_0)}{\Delta_n(\beta_0,\gamma_0)}. \tag{A.2}$$

Using Assumption M4, it follows that

$$\mathcal{SSLM}_{\beta}(\beta,\gamma)\xrightarrow{d}\sigma_{uu}^{-1}\Psi_{Zu1}'P\left(N(\nu_{w2})\nu_{x2}\right)\Psi_{Zu1}\sim\chi^2_{m_x}$$

conditional on $\Psi_{Zx2}$ and $\Psi_{Zw2}$ (and hence unconditionally). Again, as before, this implies

$$1-\lim_{n\to\infty}\Pr_{\beta,\gamma}\left[\mathcal{SSLM}_{\beta}(\beta,\gamma)\le\chi^2_{m_x}(1-\epsilon)\right]=1-\Pr\left[\chi^2_{m_x}\le\chi^2_{m_x}(1-\epsilon)\right]=\epsilon.$$

Part (iii). For any $\beta_0$, $\mathcal{SSLM}_{\gamma}^{*}\left(\beta_0,\widehat{\gamma}_{12}(\beta_0)\right)\equiv 0$ where $\widehat{\gamma}_{12}(\beta_0)=(\widehat{W}_{12}'W_1)^{-1}\widehat{W}_{12}'(y_1-X_1\beta_0)$. Therefore, $\mathcal{C}(\gamma,1-\tau,\beta_0)$ cannot be empty, and hence $\inf_{\gamma_0\in\mathcal{C}(\gamma,1-\tau,\beta_0)}\mathcal{SSLM}_{\beta}(\beta_0,\gamma_0)$ exists. Furthermore, if $\Pi_w=\mathbb{C}_w$, then by definition of $\phi_1(\beta_0,\gamma_0)$, we get that

$$\phi_1(\beta_0,\gamma_0)\xrightarrow{d}\Psi_{Zu1}+\lim_{n\to\infty}\nu_{x1}(\beta-\beta_0)n_1(\delta_x)+\lim_{n\to\infty}\lambda_w(\gamma-\gamma_0)n_1^{1/2}.$$

Hence noting the order of magnitude (as a function of $\gamma-\gamma_0$) of the numerator and the denominator of (A.1), it is evident that $\lim_{n\to\infty}\Pr_{\beta,\gamma}\left[\mathcal{SSLM}_{\gamma}^{*}(\beta_0,\gamma_0)<\infty\right]>0$ only if $\gamma-\gamma_0=O_p(n^{-1/2})$. So, if $\gamma_0$ is outside a $\sqrt{n}$-neighborhood of $\gamma$, then the probability with which it is contained in $\mathcal{C}(\gamma,1-\tau,\beta_0)$ is asymptotically equal to zero. Therefore, if $\Pi_w=\mathbb{C}_w$ and $n\to\infty$ then, by construction, $\inf_{\gamma_0\in\mathcal{C}(\gamma,1-\zeta,\beta_0)}\mathcal{SSLM}_{\beta}(\beta_0,\gamma_0)$ is attained at some $\widehat{\gamma}^{\inf}(\beta_0)$ in a $\sqrt{n}$-neighborhood of $\gamma$ with probability approaching one.

Now consider any $\gamma_0$ in a $\sqrt{n}$-neighborhood of $\gamma$ and model it as $\gamma_0=\gamma-d_{\gamma}n^{-1/2}$ for some bounded $d_{\gamma}$. If $\Pi_w=\mathbb{C}_w$, then from (A.2) we get

$$\mathcal{SSLM}_{\beta}(\beta_0,\gamma_0)$$
$$\xrightarrow{d}\lim_{n\to\infty}\frac{\left(\phi_2(\beta_0)+\sqrt{\zeta}\lambda_w d_{\gamma}\right)'P\left(N(\lambda_w)\nu_{x2}\right)\left(\phi_2(\beta_0)+\sqrt{\zeta}\lambda_w d_{\gamma}\right)}{\left[\sigma_{uu}+2\sigma_{ux}(\beta-\beta_0)+(\beta-\beta_0)'\sigma_{xx}(\beta-\beta_0)\right]}, \tag{A.3}$$

$$\mathcal{SSLM}_{\beta}(\beta_0,\gamma)\xrightarrow{d}\lim_{n\to\infty}\frac{\phi_2(\beta_0)'P\left(N(\lambda_w)\nu_{x2}\right)\phi_2(\beta_0)}{\left[\sigma_{uu}+2\sigma_{ux}(\beta-\beta_0)+(\beta-\beta_0)'\sigma_{xx}(\beta-\beta_0)\right]},$$

where $\phi_2(\beta_0)=\Psi_{Zu1}+\nu_{x1}(\beta-\beta_0)n_1(\delta_x)$. Noting that by definition $N(\lambda_w)\lambda_w=0$, it follows that $\mathcal{SSLM}_{\beta}(\beta_0,\gamma_0)$ and $\mathcal{SSLM}_{\beta}(\beta_0,\gamma)$ converge in distribution to the same random variable. Furthermore, to see that they even have the same probability, first note that $\Delta_n(\beta_0,\gamma_0)=\Delta_n(\beta_0,\gamma)+o_p(1)\neq 0$. Second, when $\Pi_w=\mathcal{C}_w$, from (A.2) and (A.3) we have $n_2^{1/2}n_1^{-3/2}N(\widehat{W}_{12})W_1\xrightarrow{P}0$, and hence the numerators of $\mathcal{SSLM}_{\beta}(\beta_0,\gamma_0)$ and $\mathcal{SSLM}_{\beta}(\beta_0,\gamma)$ also have the same probability limit. Therefore, the statistics have the same probability limit. Hence, when $\Pi_w=\mathbb{C}_w$ it follows from our previous arguments that

$$\inf_{\gamma_0\in\mathcal{C}(\gamma,1-\tau,\beta_0)}\mathcal{SSLM}_{\beta}(\beta_0,\gamma_0)=\mathcal{SSLM}_{\beta}(\beta_0,\gamma)+o_p(1)$$

and the asymptotic distribution is given by the right-hand side of (A.4). ∎

**Proof of Equation (4.2).** For $\beta_0 = \beta - n(\delta_x)^{-1} d_\beta$, as defined in the statement of Theorem 4.1, note that

$$n_1(\delta_w)(\widehat{\gamma}_{12}(\beta_0) - \gamma) = (\nu'_{w2}\nu_{w1})^{-1}\nu'_{w2}\left[\Psi_{Zu1} + \nu_{x1}d_\beta\left(1_{[\delta_x \neq 1]} + \sqrt{\zeta}1_{[\delta_x = 1]}\right)\right]$$

and hence when $\Pi_w = \mathbb{C}_w$, the restricted USSIV estimator $\widehat{\gamma}_{12}(\beta_0)$ is $\sqrt{n}$-consistent for $\gamma$. Therefore, it follows using the same strategy as in the proof of Theorem 4.1(iii) that $\mathcal{SSLM}_\beta(\beta_0, \widehat{\gamma}_{12}(\beta_0)) = \mathcal{SSLM}_\beta(\beta_0, \gamma) + o_p(1)$. ■

In this context it is also interesting to note that in a standard linear IV regression (where $\Pi_x = \mathbb{C}_x$ and $\Pi_w = \mathbb{C}_w$), if $\beta_0 = \beta - d_\beta n^{-1/2}$ then

$$\inf_{\gamma_0 \in \mathcal{C}(\gamma, 1-\tau, \beta_0)} \mathcal{SSLM}_\beta(\beta_0, \gamma_0) \xrightarrow{d} \chi^2_{m_x} \tag{A.4}$$

with noncentrality parameter $\zeta \sigma_{uu}^{-1} d'_\beta \lambda'_x N(\lambda_w)\lambda_x d_\beta$. The limiting distribution is the same as that of the usual score test for $\beta = \beta_0$ based on subsample 1 (see Wang and Zivot, 1998).