L'Actualité économique

A Shrinkage Instrumental Variable Estimator for Large Datasets

Andrea Carriero, George Kapetanios et Massilimiano Marcellino

Volume 91, numéro 1-2, mars–juin 2015

Identification, Simulation and Finite-Sample Inference

URI : https://id.erudit.org/iderudit/1036914ar
DOI : https://doi.org/10.7202/1036914ar

Aller au sommaire du numéro

Éditeur(s)

HEC Montréal

ISSN

0001-771X (imprimé)
1710-3991 (numérique)

Découvrir la revue

Citer cet article

Carriero, A., Kapetanios, G. & Marcellino, M. (2015). A Shrinkage Instrumental Variable Estimator for Large Datasets. *L'Actualité économique*, *91*(1-2), 67–87. https://doi.org/10.7202/1036914ar

Résumé de l'article

This paper proposes and discusses an instrumental variable estimator that can be of particular relevance when many instruments are available and/or the number of instruments is large relative to the total number of observations. Intuition and recent work (see, e.g., Hahn, 2002) suggest that parsimonious devices used in the construction of the final instruments may provide effective estimation strategies. Shrinkage is a well known approach that promotes parsimony. We consider a new shrinkage 2SLS estimator. We derive a consistency result for this estimator under general conditions, and via Monte Carlo simulation show that this estimator has good potential for inference in small samples.

L'Actualité économique, Revue d'analyse économique, vol. 91, nos 1-2, mars-juin 2015

A SHRINKAGE INSTRUMENTAL VARIABLE ESTIMATOR FOR LARGE DATASETS*

Andrea CARRIERO
Queen Mary
University of London

George KAPETANIOS
Queen Mary
University of London

Massilimiano MARCELLINO
Bocconi University
IGIER
CEPR

Abstract–This paper proposes and discusses an instrumental variable estimator that can be of particular relevance when many instruments are available and/or the number of instruments is large relative to the total number of observations. Intuition and recent work (see, e.g., Hahn, 2002) suggest that parsimonious devices used in the construction of the final instruments may provide effective estimation strategies. Shrinkage is a well known approach that promotes parsimony. We consider a new shrinkage 2SLS estimator. We derive a consistency result for this estimator under general conditions, and via Monte Carlo simulation show that this estimator has good potential for inference in small samples.

Introduction

Recent theoretical work in instrumental variable estimation has focused on the consequences of having "weak instruments", "many instruments", or a combination of these two cases. Instrument weakness is, unfortunately, rather likely in economic applications, and the availability of larger and larger information sets makes the many instruments case also relevant for empirical analyses. Hence, the theoretical contributions on IV estimation have a vast range of practical applicability.

By "weak instruments" we label the case where instrumental variables are only weakly correlated with the endogenous explanatory variables of an instrumental

* We would like to thank an anonymous referee for helpful comments on a previous draft.

variables (IV) regression. A natural measure of instrument weakness (or strength) in a linear IV framework is the so-called concentration parameter, see, e.g. Phillips (1983), Rothenberg (1984), Stock and Yogo (2003b) and Chao and Swanson (2005). In standard analysis the concentration parameter is taken to grow at the rate of the sample size whereas in the case of weak instruments this parameter grows more slowly or, in the extreme case introduced and considered by Staiger and Stock (1997), it remains finite asymptotically. Weak instruments affect the properties of IV estimators such as the two stage least squares (2SLS) and the limited information maximum likelihood (LIML) estimators, in particular they can become inconsistent.

The "many instruments" case was first analysed by Morimune (1983) and later generalized by Bekker (1994). Other relevant papers include Donald and Newey (2001), Hahn, Hausman, and Kuersteiner (2001), Hahn (2002), and Chao and Swanson (2004). In general, the larger the available information set the more efficient the resulting estimator. However, when the number of instruments becomes too large, standard IV estimators can become inconsistent, as in the weak instrument case though for different reasons.

These two developments in the IV literature have later been combined to provide a comprehensive framework for the analysis of the properties of IV estimators in the case of many weak instruments. Work in this area includes Stock and Yogo (2003a), Newey (2004), Chao and Swanson (2005), and Hansen, Hausman, and Newey (2006). The Chao and Swanson paper is closest to the spirit of the analysis of the current paper. A clear conclusion from this work suggests that inconsistency of IV estimators is a probable outcome when many weak instruments are used.

Given this problem, recent research focuses on considering parsimonious modeling methods for the large set of potentially weak instruments to avoid IV estimator inconsistency. In particular, Kapetanios and Marcellino (2010) and Bai and Ng (2010) suggest that imposing a factor structure on the set of instruments, extracting estimates of these factors and using them as instruments can be very useful. Of course, an issue with this approach is the need to assume a factor structure, albeit a possibly weak one, as discussed in detail in Kapetanios and Marcellino (2010). Simulation evidence suggests that if no factor structure exists then assuming one is problematic for IV estimation as one would expect. Another approach similar but designed to parsimoniously summarize large sets of instruments in the complete absence of a factor structure is proposed by Kapetanios and Marcellino (2007). The basic idea is that a finite number of cross-sectional weighted averages of the available instruments can, under certain conditions, be valid instruments themselves.

The current paper provides a new approach to deal with the IV inconsistency issue, which shares the search for parsimony with the papers mentioned in the previous paragraph but can be applied under more general conditions. In particular, we suggest that a shrinkage estimator be considered in the first stage of IV regression to construct appropriate instruments which can then be used in a standard way in the second stage to estimate the parameters of the structural equation. Shrinkage promotes parsimony in the first stage of estimation. In addition to the reasonably strong case for parsimony for IV estimation made in the cited literature, Hahn (2002) provides grounds for parsimony also in terms of optimal inference when many instruments are available.

After introducing our new estimator, that we label Two Stages Least Squares Shrinkage (2SLSS), we provide a formal proof of its consistency under general conditions on the instrument set. Further, we carry out a Monte Carlo study which provides clear evidence in favor of the new estimator compared with existing estimators such as 2SLS or LIML (also when the number of instruments is large relative to the total number of observations). Finally, we apply the new estimator to the Angrist and Krueger (1991) dataset which has been repeatedly used in the literature, in the context of analyzing new methodological IV-related advances. We propose an innovative way of using this dataset in order to evaluate the performance of the new estimator relative to existing ones.

The paper is structured as follows: Section 1 presents the theoretical results. Section 2 reports results of the Monte Carlo study. Section 3 presents the results of our empirical application. Finally, we conclude. Proofs are relegated to an Appendix.

1. Theoretical Results

The model is given by

$$y_{1n} = Y_{2n}\beta + u_n , \tag{1}$$

$$Y_{2n} = Z_n\Pi_n + V_n , \tag{2}$$

where y_{1n} and Y_{2n} are respectively an n x 1 vector and an n x G matrix of observations on the $G + 1$ endogenous variables of the system, Z_n is an n x K_n matrix of observations on the K_n instrumental variables, and $u_n = (u_1,...,u_i,...,u_n)'$ and $V_n = (v_1,...,v_i,...,v_n)'$ are, respectively, an n x 1 vector and an n x G matrix of random disturbances.

We propose a two stage shrinkage estimator for β obtained as follows: in the first stage, we obtain instruments by using a standard shrinkage estimator to estimate Π_n in (2). Then, we use these instruments in a standard fashion to obtain a second stage estimator for β. For simplicity we use the following shrinkage estimator:

$$\hat{\Pi}_n = (Z'_n Z_n + s_n I)^{-1} Z'_n Y_{2n}.$$

Then, straightforwardly, the two stage estimator is given by

$$\hat{\beta}_{2SLSS} = \left(Y'_{2n} Z_n (Z'_n Z_n + s_n I)^{-1} Z'_n Y_{2n}\right)^{-1} Y'_{2n} Z_n (Z'_n Z_n + s_n I)^{-1} Z'_n y_{1n}. \tag{3}$$

We refer to this estimator as the 2SLS Shrinkage (2SLSS) estimator. This estimator becomes of interest if the shrinkage parameter s_n becomes large enough to promote parsimony asymptotically. As we will see, for this it is required that $n/s_n = o(1)$. We make the following assumptions.

Assumption 1 *(i)* $K_n \to \infty$ *as* $n \to \infty$ *such that* $K_n / n \to \tau, 0 \leq \tau \leq C < \infty$. *(ii)* $\forall n$, $Z'_n Z_n + s_n I$ *has full rank. (iii) There exist two nondecreasing sequences of real numbers,* r_n *and* s_n*, such that as* $n \to \infty$ $r_n / n \to \mu$ *for some nonnegative constant* μ, $n / s_n = o(1)$ *and* $s_n / nK_n = o(1)$*, and such that*

$$\frac{q_n \Pi'_n Z'_n Z_n (Z'_n Z_n + s_n I)^{-1} Z'_n Z_n \Pi_n}{r_n} \to \Psi , \tag{4}$$

where $q_n = s_n / n$*, almost surely for some positive definite matrix* Ψ *and*

$$\frac{q_n \Pi'_n Z'_n Z_n (Z'_n Z_n + s_n I)^{-1} Z'_n Z_n (Z'_n Z_n + s_n I)^{-1} Z_n Z_n \Pi_n}{r_n} \to 0$$

almost surely. (iv) The eigenvalues of $Z'_n Z_n / n$ *are bounded away from zero and infinity for all n.*

Assumption 2 *(i)* Z_n *and* $\eta_i = (u_i, v'_i)'$ *are independent for all i,n, (ii)*

$\eta_i \sim i.i.d.(0, \Sigma)$*, where* $\Sigma = \begin{pmatrix} \sigma_{uu} & \sigma'_{Vu} \\ \sigma_{Vu} & \Sigma_{VV} \end{pmatrix}$*, (iii)* η_i *has finite fourth moments.*

Given the above, we have the following theorem:

Theorem 1 *Let* $P_{Z_n}^{s_n} = Z_n (Z'_n Z_n + s_n I)^{-1} Z'_n$. *Let* $q_n = s_n / n$ *such that* $q_n \to \infty$ *and* $\frac{K_n}{q_n} \to \infty$ *and* $\frac{r_n}{q_n} \to \infty$. *Let the shrinkage estimator be given by*

$$\hat{\beta}_{2SLSS} = \left(Y'_{2n} P_{Z_n}^{s_n} Y_{2n} \right)^{-1} \left(Y'_{2n} P_{Z_n}^{s_n} y_{1n} \right).$$

Assume that $\frac{K_n}{r_n} \to 0$. *Then,* $\hat{\beta}_{2SLSS}$ *is consistent for* β_0.

Some comments on the assumptions are in order. In particular, assumption 1(iii) is worthy of comment. The first part of assumption 1(iii) is the counterpart of the assumption relating to the concentration parameter made usually in the literature concerning the 2SLS and other IV estimators. Note that there is no need for the sequence r_n satisfying assumption 1(iii) for the 2SLSS estimator to be the same or have the same order of magnitude as that required for the 2SLS estimator.

The importance of parsimony for IV estimation has been pointed out by? Hahn (2002) who conjectured that a 2SLS estimator using a small subset of available instruments, when the number of available instruments is large, may be optimal. We view our shrinkage estimator in the same spirit as the estimator suggested by Hahn (2002). It is important to note condition 1 of Hahn (2002) which requires that the fit of a parsimonious estimator be comparable to that of the 2SLS estimator using all instruments. In this sense it is reasonable to expect that the fit of the shrinkage estimator may, under certain conditions relating to the structure of Π_n, be close to that of the 2SLS estimator using all instruments, thereby implying that the r_n sequence for the 2SLSS estimator be of a larger order of magnitude than the analogous sequence for the 2SLS estimator. However, it is difficult to envisage specific conditions for Π_n that ensure this is the case.

We have chosen to focus on the simplest shrinkage estimator on the grounds of simplifying the asymptotic analysis. This estimator shrinks, in a uniform way, the parameter estimates towards zero. It may in fact be more appropriate to shrink towards a nonzero constant or vary the degree of shrinkage depending on the instrument. For such shrinkage estimators the condition (4) would have a different form and therefore it is entirely possible that such estimator will have different and possibly superior consistency properties, depending of course on the data generating process for z_t. We leave theoretical investigation of this possibility to future work mainly because there are many possibilities for the shrinkage setup. However, in the Monte Carlo section we consider uniform shrinkage to a nonzero constant and obtain interesting results.

2. Monte Carlo Evidence

In this section we provide a Monte Carlo study of the 2SLS Shrinkage (*2SLSS*) estimator and its relative performance compared to the traditional *2SLS* estimator, the *LIML* estimator, and the bias corrected Nagar's B*2SLS* estimator. These estimators fall in the class of *k*-estimators, and can be written as:

$$\hat{\beta}_K = \left(Y_{2n}'Z_n(Z_n'Z_n)^{-1}Z_n'Y_{2n} - \lambda Y_{2n}'Y_{2n}\right)^{-1}\left(Y_{2n}'Z_n(Z_n'Z_n)^{-1}Z_n'y_{1n} - \kappa Y_{2n}'y_{1n}\right), \tag{5}$$

where the *2SLS* estimator corresponds to $\lambda = 0$, the *LIML* corresponds to setting λ to the minimum of $\dfrac{(Y - Y_{2n}\beta)'Z_n(Z_n'Z_n)^{-1}Z_n'(Y - Y_{2n}\beta)}{(Y - Y_{2n}\beta)'(Y - Y_{2n}\beta)}$, and the B*2SLS* corresponds to $\lambda = \dfrac{K_n - 2}{n}$.

The basic setup of the Monte Carlo experiments is:

$$y_i = x_i + \varepsilon_i, i = 1,\dots,n \tag{6}$$

$$z_{ij} = e_{ij},\ j = 1,\dots,K_n, i = 1,\dots,n \tag{7}$$

$$x_i = \sum_{j=1}^{K_n} K_n^{-1/2}(1+\alpha_j)z_{ij} + u_i, \tag{8}$$

where $e_{ij} \sim i.i.d.N(0,1)$, $cov(e_{il},e_{sj}) = 0$ for $i \neq s$ or $l \neq j$, $\alpha_j \sim N(0,c^2)$ with c = 0.1,0.5,1. Let $\kappa_i = (\varepsilon_i,u_i)'$. Then, $\kappa_i = P\eta_i$, where $\eta_i = (\eta_{1,i},\eta_{2,i})'$ $\eta_{j,i} \sim i.i.d.N(0,1)$ and $P = [p_{ij}]$, $p_{ij} \sim i.i.d.N(0,1)$. The errors e_{ij} and u_s are related as follows:

$$\varepsilon_i = \rho u_i + \sqrt{1-\rho^2}\, v_i, \tag{9}$$

where u and v are both $i.i.d.N(0,1)$. We run experiments with p = 0.25, 0.5, 0.75.

The *2SLSS* estimator is computed for a grid of values of the tightness parameter s_n. In particular we use the grid $s_n = 0, 10, 10^3, 10^5$. For $s_n = 0$ the *2SLSS* and *2SLS* are equivalent, therefore we do not report results for this case. Higher values of s_n correspond to stronger shrinkage. We consider two different shrinkage setups: one where we shrink towards $q = 0$ and one where we shrink towards $q = 1/\sqrt{K_n}$. The latter corresponds to the true value of the coefficients in the setup of the Monte Carlo. We have also considered shrinking towards $1/K_n$ with very similar results[1] to those for $1/\sqrt{K_n}$ giving us some comfort that the actual choice of the non-zero constant may not be crucial.

Results are reported in Tables 1-3. The tables display the relative mean squared error (*RMSE*) of each estimator with respect to the *2SLS* estimator, i.e. the ratio between the mean quared error *(MSE)* of a given estimator and the *MSE* of the *2SLS* estimator. For the *2SLS* estimator we do not report the ratio (as it will be equal to 1) but the *MSE*. The tables are organized so that on the rows are reported results for different numbers of observations n while on the columns are displayed results for different proportions of the number of instruments to the number of observations, i.e. K_n/n. The tables are vertically divided in three subpanels providing results for the three cases $c = 0.1, 0.5, 1$.

In Tables 1-3 a figure smaller than 1 signals that the considered estimator is more efficient than *2SLS*. As is clear, both the *LIML* and the *B2SLS* estimators substantially improve on the traditional *2SLS* in all the cases in which $K_n < n$ (with large n), while in the case $n = K_n$ the *LIML* performs very poorly, and the *B2SLS* is by construction equivalent to *2SLS*[2].

Turning our attention to the *2SLSS* estimator, two main results are apparent. First, the *2SLSS* features a systematically smaller *MSE* than both *2SLS* and, to a smaller extent, *LIML* and *B2SLS*. Second, when both n and K_n are large the *MSEs* of *2SLSS* with prior mean $q = 1/\sqrt{K_n}$ become remarkably small.

Finally we focus on the case $K_n > n$. As for this case the competitor estimators are not implementable, we provide results only for the *2SLSS*. Table 4 displays the *MSEs* of the *2SLSS* estimator in the cases $K_n/n = 1$ and $K_n/n = 1.1$, as well as their ratio. The ratios are systematically close to 1, showing that the *2SLSS* estimator can handle the $K_n > n$ case almost as efficiently as the case $n = K_n$.

These results confirm our theoretical findings and, further, show that using shrinkage in the first stage may significantly improve the small sample efficiency of the estimator. Our results for the case $q = 1/\sqrt{K_n}$ suggest that shrinking the coefficients towards an appropriate direction might improve the results even further, possibly indicating that the consistency properties of this shrinkage estimator are

1. These results are not reported but are available upon request.

2. *B2SLS* and *2SLS* are also equivalent to *OLS* when $n = K_n$. The two equivalences are obvious once one notes that for $n = K_n$, $Y_{2n}'Z_n(Z_n'Z_n)^{-1}Z_n'Y_{2n} = Y_{2n}'Y_{2n}$ and $Y_{2n}'Z_n(Z_n'Z_n)^{-1}Z_n'y_{1n} = Y_{2n}'y_{1n}$.

TABLE 1

Mean Squared Errors, $p = 0.25$

	c=0.1			c=0.5			c=1		
$Kn/n\rightarrow$ $n\downarrow$	0.60	0.80	1.00	0.60	0.80	1.00	0.60	0.80	1.00
				2SLS (level)					
50	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01
100	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
200	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.01
400	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.00	0.01
				B2SLS					
50	4.58	12.2	1.00	2.78	7.13	1.00	1.63	4.82	1.00
100	2.01	6.71	1.00	1.57	3.28	1.00	1.24	1.72	1.00
200	1.09	1.84	1.00	1.01	1.53	1.00	0.82	1.18	1.00
400	0.62	0.96	1.00	0.58	0.81	1.00	0.53	0.67	1.00
				LIML					
50	11.5	25.5	114.3	9.88	27.8	112.6	2.63	18.4	166.1
100	3.00	13.4	142.5	1.87	7.10	164.4	1.35	4.30	217.2
200	1.09	6.26	159.8	1.01	2.24	191.1	0.81	1.32	310.7
400	0.59	1.00	166.9	0.57	0.82	238.2	0.52	0.66	315.1
				2SLSS					
				q=1/n; s_n= 10					
50	0.90	0.84	0.80	0.91	0.86	0.80	0.94	0.88	0.84
100	0.89	0.86	0.80	0.90	0.85	0.79	0.93	0.87	0.80
200	0.91	0.87	0.83	0.92	0.88	0.81	0.93	0.88	0.81
400	0.94	0.91	0.86	0.95	0.91	0.85	0.95	0.91	0.83
				q=1/n; s_n= 10^3					
50	0.98	0.87	0.81	1.12	1.03	0.85	1.65	1.39	1.27
100	0.66	0.57	0.45	0.79	0.62	0.53	1.08	0.83	0.70
200	0.43	0.34	0.29	0.50	0.40	0.32	0.67	0.51	0.41
400	0.31	0.27	0.25	0.36	0.30	0.27	0.46	0.37	0.32
				q=1/n; s_n= 10^5					
50	1.04	0.94	0.90	1.22	1.17	0.97	2.11	2.02	1.69
100	0.71	0.62	0.50	0.90	0.73	0.64	1.46	1.16	1.00
200	0.46	0.36	0.29	0.59	0.45	0.35	0.96	0.74	0.59
400	0.26	0.20	0.16	0.33	0.26	0.20	0.57	0.42	0.31
				q=0; s_n= 10					
50	0.96	0.90	0.87	0.96	0.92	0.85	0.97	0.91	0.87
100	0.94	0.91	0.84	0.95	0.89	0.83	0.95	0.90	0.83
200	0.95	0.91	0.86	0.95	0.91	0.84	0.95	0.90	0.82
400	0.97	0.93	0.87	0.96	0.93	0.86	0.96	0.92	0.84
				q=0; s_n= 10^3					
50	1.05	0.96	0.88	1.06	0.96	0.87	1.08	0.95	0.90
100	0.87	0.83	0.72	0.91	0.80	0.72	0.96	0.84	0.73
200	0.77	0.69	0.62	0.77	0.69	0.60	0.81	0.69	0.59
400	0.70	0.63	0.57	0.69	0.62	0.55	0.70	0.60	0.53
				q=0; s_n= 10^5					
50	1.06	0.97	0.90	1.08	0.98	0.88	1.10	0.97	0.92
100	0.89	0.84	0.72	0.93	0.81	0.73	0.98	0.85	0.74
200	0.77	0.67	0.60	0.77	0.68	0.59	0.81	0.69	0.59
400	0.66	0.58	0.51	0.65	0.57	0.51	0.68	0.56	0.48

Note : The Table displays the Relative Mean Squared Error (*RMSE*) of each estimator with respect to the 2*SLS* estimator, i.e. the ratio between the Mean Squared Error (*MSE*) of a given estimator and the *MSE* of the 2*SLS* estimator. For the 2*SLS* estimator we do not report the ratio (as it will be equal to 1) but the *MSE*. On the rows are reported results for different numbers of observations n while on the columns are displayed results for different proportions of the number of instruments to the number of observations, i.e. K_n/n. Results are computed with $p = 0.25$

TABLE 2

MEAN SQUARED ERRORS, $p = 0.50$

	c=0.1			c=0.5			c=1		
$Kn/n\rightarrow$ $n\downarrow$	0.60	0.80	1.00	0.60	0.80	1.00	0.60	0.80	1.00
2SLS (level)									
50	0.03	0.04	0.04	0.02	0.03	0.04	0.01	0.02	0.02
100	0.02	0.03	0.04	0.02	0.03	0.03	0.01	0.02	0.02
200	0.02	0.03	0.04	0.02	0.02	0.03	0.01	0.01	0.02
400	0.02	0.03	0.04	0.02	0.02	0.03	0.01	0.01	0.02
B2SLS									
50	3.04	5.01	1.00	1.80	4.82	1.00	1.40	1.90	1.00
100	0.88	3.21	1.00	0.81	2.09	1.00	0.64	0.78	1.00
200	0.41	1.32	1.00	0.36	0.62	1.00	0.34	0.41	1.00
400	0.19	0.29	1.00	0.19	0.24	1.00	0.18	0.21	1.00
LIML									
50	3.53	8.39	40.88	1.93	6.95	42.32	2.05	3.04	56.24
100	0.73	3.07	43.58	0.65	1.64	51.43	0.62	1.00	65.69
200	0.31	0.68	46.65	0.30	0.44	52.19	0.29	0.34	84.41
400	0.15	0.22	43.83	0.15	0.19	56.17	0.15	0.16	81.71
2SLSS									
$q=1/n$; $s_n= 10$									
50	0.77	0.71	0.66	0.78	0.72	0.67	0.81	0.75	0.68
100	0.82	0.78	0.73	0.83	0.78	0.72	0.85	0.79	0.72
200	0.89	0.85	0.79	0.89	0.85	0.78	0.90	0.84	0.77
400	0.93	0.90	0.84	0.93	0.90	0.83	0.94	0.90	0.82
$q=1/n$; $s_n= 10^3$									
50	0.49	0.35	0.30	0.61	0.43	0.35	0.96	0.71	0.54
100	0.27	0.20	0.16	0.32	0.22	0.19	0.49	0.34	0.27
200	0.17	0.14	0.13	0.21	0.16	0.14	0.29	0.23	0.19
400	0.18	0.17	0.17	0.21	0.19	0.18	0.27	0.24	0.21
$q=1/n$; $s_n= 10^5$									
50	0.52	0.39	0.33	0.68	0.49	0.42	1.25	1.81	0.73
100	0.29	0.21	0.18	0.37	0.25	0.20	0.66	0.46	0.37
200	0.15	0.11	0.09	0.20	0.14	0.10	0.35	0.24	0.18
400	0.07	0.05	0.04	0.10	0.07	0.05	0.18	0.12	0.09
$q=0$; $s_n= 10$									
50	0.89	0.81	0.76	0.88	0.81	0.75	0.88	0.81	0.73
100	0.90	0.85	0.79	0.90	0.84	0.77	0.90	0.83	0.75
200	0.93	0.89	0.82	0.93	0.88	0.81	0.92	0.87	0.79
400	0.96	0.93	0.86	0.96	0.92	0.85	0.95	0.91	0.83
$q=0$; $s_n= 10^3$									
50	0.79	0.66	0.60	0.79	0.67	0.59	0.80	0.69	0.58
100	0.68	0.60	0.53	0.67	0.58	0.52	0.69	0.58	0.49
200	0.63	0.55	0.51	0.62	0.55	0.48	0.61	0.52	0.45
400	0.63	0.56	0.52	0.61	0.54	0.50	0.59	0.51	0.45
$q=0$; $s_n= 10^5$									
50	0.79	0.66	0.59	0.79	0.67	0.59	0.81	0.69	0.58
100	0.67	0.59	0.52	0.67	0.57	0.50	0.68	0.57	0.48
200	0.60	0.52	0.47	0.59	0.51	0.45	0.58	0.49	0.42
400	0.56	0.49	0.45	0.54	0.47	0.43	0.52	0.44	0.38

NOTE : The Table displays the Relative Mean Squared Error (*RMSE*) of each estimator with respect to the 2*SLS* estimator, i.e. the ratio between the Mean Squared Error (*MSE*) of a given estimator and the *MSE* of the 2*SLS* estimator. For the 2*SLS* estimator we do not report the ratio (as it will be equal to 1) but the *MSE*. On the rows are reported results for different numbers of observations *n* while on the columns are displayed results for different proportions of the number of instruments to the number of observations, i.e. K_n/n. Results are computed with $p = 0.50$

TABLE 3

MEAN SQUARED ERRORS, $p = 0.75$

	c=0.1			c=0.5			c=1		
$Kn/n\rightarrow$ $n\downarrow$	0.60	0.80	1.00	0.60	0.80	1.00	0.60	0.80	1.00
				2SLS (level)					
50	0.05	0.07	0.09	0.04	0.05	0.07	0.02	0.03	0.05
100	0.05	0.07	0.08	0.04	0.05	0.07	0.02	0.03	0.04
200	0.05	0.07	0.08	0.04	0.05	0.07	0.02	0.03	0.04
400	0.04	0.06	0.08	0.03	0.05	0.07	0.02	0.03	0.04
				B2SLS					
50	1.97	3.68	1.00	1.25	2.46	1.00	0.77	1.83	1.00
100	0.55	1.34	1.00	0.53	1.93	1.00	0.37	0.47	1.00
200	0.22	0.61	1.00	0.20	0.33	1.00	0.18	0.21	1.00
400	0.10	0.17	1.00	0.09	0.14	1.00	0.09	0.11	1.00
				LIML					
50	0.63	3.65	18.17	0.91	1.82	20.90	0.50	0.80	29.56
100	0.31	0.47	19.28	0.22	0.36	23.71	0.25	0.28	28.97
200	0.12	0.15	19.78	0.11	0.13	22.01	0.12	0.11	38.13
400	0.06	0.06	18.70	0.05	0.06	22.77	0.06	0.05	32.62
				2SLSS					
				q=1/n; s_n= 10					
50	0.70	0.66	0.62	0.72	0.67	0.62	0.75	0.68	0.62
100	0.80	0.76	0.71	0.81	0.76	0.70	0.82	0.76	0.69
200	0.88	0.84	0.78	0.88	0.84	0.77	0.89	0.83	0.75
400	0.93	0.90	0.84	0.93	0.90	0.83	0.93	0.89	0.81
				q=1/n; s_n= 10^3					
50	0.28	0.19	0.16	0.32	0.25	0.18	0.90	0.39	0.28
100	0.14	0.10	0.09	0.16	0.12	0.09	0.26	0.18	0.14
200	0.10	0.09	0.09	0.12	0.11	0.10	0.18	0.15	0.13
400	0.15	0.15	0.15	0.17	0.17	0.17	0.22	0.20	0.19
				q=1/n; s_n= 10^5					
50	0.31	0.22	0.18	0.37	0.29	0.22	0.97	0.78	0.45
100	0.15	0.10	0.08	0.19	0.13	0.10	0.35	0.23	0.17
200	0.07	0.05	0.04	0.09	0.06	0.05	0.17	0.11	0.08
400	0.04	0.02	0.02	0.05	0.03	0.02	0.08	0.05	0.04
				q=0; s_n= 10					
50	0.85	0.79	0.72	0.85	0.78	0.71	0.84	0.76	0.68
100	0.89	0.83	0.77	0.88	0.82	0.75	0.88	0.80	0.72
200	0.93	0.88	0.82	0.93	0.87	0.80	0.92	0.86	0.77
400	0.96	0.92	0.86	0.96	0.92	0.85	0.95	0.91	0.82
				q=0; s_n= 10^3					
50	0.65	0.58	0.51	0.66	0.56	0.49	0.66	0.55	0.47
100	0.60	0.54	0.49	0.59	0.52	0.46	0.59	0.49	0.43
200	0.59	0.53	0.49	0.58	0.51	0.46	0.56	0.48	0.42
400	0.61	0.55	0.51	0.59	0.53	0.48	0.57	0.49	0.44
				q=0; s_n= 10^5					
50	0.65	0.57	0.50	0.65	0.55	0.48	0.66	0.54	0.46
100	0.59	0.52	0.47	0.58	0.50	0.44	0.58	0.47	0.41
200	0.55	0.49	0.45	0.54	0.47	0.42	0.52	0.44	0.38
400	0.54	0.48	0.43	0.52	0.46	0.41	0.49	0.41	0.37

NOTE : The Table displays the Relative Mean Squared Error (*RMSE*) of each estimator with respect to the 2*SLS* estimator, i.e. the ratio between the Mean Squared Error (*MSE*) of a given estimator and the *MSE* of the 2*SLS* estimator. For the 2*SLS* estimator we do not report the ratio (as it will be equal to 1) but the *MSE*. On the rows are reported results for different numbers of observations n while on the columns are displayed results for different proportions of the number of instruments to the number of observations, i.e. K_n/n. Results are computed with $p = 0.75$

TABLE 4

MEAN SQUARED ERRORS OF THE 2SLSS. CASE $K_n > n$

	$p = 0.25$			$p = 0.50$			$p = 0.75$		
$Kn/n\rightarrow$ $n\downarrow$	1	1.1	Ratio	1	1.1	Ratio	1	1.1	Ratio
				$q=1/n$ $s_n = 10$					
50	0.011	0.011	1.018	0.024	0.025	0.960	0.044	0.048	0.913
100	0.008	0.009	0.953	0.024	0.024	0.979	0.048	0.051	0.955
200	0.007	0.008	0.923	0.024	0.025	0.941	0.051	0.054	0.946
400	0.007	0.007	0.972	0.025	0.026	0.984	0.056	0.058	0.959
				$q=1/n$ $s_n = 10^3$					
50	0.012	0.012	1.008	0.013	0.012	1.066	0.012	0.013	0.932
100	0.006	0.006	0.982	0.006	0.006	1.034	0.007	0.007	0.985
200	0.003	0.003	0.935	0.004	0.005	0.936	0.007	0.008	0.893
400	0.002	0.002	0.957	0.006	0.006	0.932	0.011	0.012	0.911
				$q=1/n$ $s_n = 10^5$					
50	0.014	0.014	0.986	0.015	0.014	1.063	0.015	0.016	0.936
100	0.007	0.007	0.971	0.007	0.007	1.030	0.007	0.007	1.015
200	0.003	0.003	0.970	0.003	0.003	0.970	0.003	0.003	1.031
400	0.002	0.002	1.000	0.002	0.002	1.000	0.002	0.002	1.133
				$q=0$; $s_n = 10$					
50	0.012	0.012	1.017	0.027	0.028	0.961	0.050	0.054	0.921
100	0.009	0.009	0.956	0.025	0.026	0.984	0.052	0.054	0.959
200	0.007	0.008	0.925	0.025	0.026	0.943	0.053	0.056	0.950
400	0.007	0.007	0.972	0.026	0.026	0.985	0.057	0.059	0.960
				$q=0$; $s_n = 10^3$					
50	0.012	0.012	1.042	0.022	0.022	0.977	0.035	0.038	0.913
100	0.007	0.008	0.949	0.017	0.017	1.000	0.032	0.033	0.967
200	0.005	0.006	0.914	0.015	0.016	0.937	0.031	0.033	0.939
400	0.005	0.005	0.978	0.015	0.016	0.968	0.033	0.034	0.948
				$q=0$; $s_n = 10^5$					
50	0.013	0.012	1.041	0.021	0.022	0.977	0.034	0.037	0.914
100	0.008	0.008	0.949	0.017	0.016	1.012	0.031	0.032	0.968
200	0.005	0.006	0.929	0.014	0.015	0.939	0.028	0.030	0.943
400	0.004	0.004	0.976	0.013	0.013	0.977	0.028	0.029	0.955

NOTE : The table displays the *MSE* of the 2*SLSS* estimator in the cases $K_n/n = 1$ and $K_n/n = 1.1$, and their ratio (i.e. *MSE* in the case $K_n/n = 1$ divided by *MSE* in the case $K_n/n = 1.1$). The three verstical subpanels display results for $p = 0.25,5,0.75$. The parameter c is set to $c = 0.5$.

different to those of the simple one analysed theoretically in the previous section. As we noted in remark 3 this is entirely possible since the relevant consistency condition (4) will be different for the two shrinkage estimators. This is a topic of interest for future work.

3. APPLICATION TO ANGRIST-KRUEGER (1991) DATASET

In this Section we evaluate the properties of the 2*SLSS* estimator by using the Krueger (1991) dataset. This dataset has been repeatedly used when evaluating new IV related methods (see, e.g., Donald and Newey, 2001). The dataset is composed of 329,509 observations on men born between 1930-1939 and is taken from the US Census. Angrist and Krueger (1991) estimate an equation where the dependent variable is the log of the weekly wage, and the explanatory variable of interest, featuring endogeneity, is the number of years of schooling. They consider several models, differing in the set of exogenous explanatory variables which are included in the equation of interest. The particular version of the model we consider is the same selected also by Donald and Newey (2001), and is the one in which the exogenous explanatory variables include an intercept, 9 year-of-birth dummies, and 50 state-of-birth dummies (for a total of 60 variables). All the 60 exogenous explanatory variables are used as instruments. Additionally, the instrument set includes 3 quarter-of-birth dummies, 27 interactions of the 3 quarter-of-birth dummies with the 9 year-of-birth dummies, and 150 interactions of the 3 quarter-of-birth dummies with the 50 state-of-birth dummies. This gives a total of 240 instruments. This particular model and dataset correspond to column 2 of Table VII in Angrist and Krueger (1991), and row 4 in Table VIII of Donald and Newey (2001). Using the whole sample, the *2SLS* estimate of the coefficient on years of schooling is 0.0928, with a standard deviation of 0.0093. One interesting feature of this dataset is that the number of observations greatly exceeds the number of instruments. This implies that we are not faced with a relatively large number of instruments and therefore that standard methods such as 2SLS can be expected to perform reasonably well.

The above feature of this widely analysed dataset can be used to evaluate the performance of the *2SLSS* against the traditional *2SLS* estimator, in, what we view as, an innovative way. We propose the following approach. We draw from the complete dataset of 329,509 observations, a total of 600 subsamples of 500 observations each. Each subsample is composed of equally spaced observations of the randomly ordered complete sample[3]. If in some samples, some dummies are not active causing perfect multicollinearity they are removed. These 600 subsamples can be then used to estimate the coefficient of interest, and this will provide a distribution of estimates. If we knew the true value of the coefficient, this distribution would provide us with an estimate of the bias, variance, and mean squared error (*MSE*) of the estimator used. As the dataset is composed of 329,509 observations, we can reasonably argue that the estimated coefficient of 0.0928 obtained, via 2SLS, using the whole sample is close enough to the truth to be considered a proxy for the actual coefficient. Therefore, we can compute an estimate of the MSE for all the estimators at hand by using the estimated distributions and assuming that the true value of the coefficient is 0.0928.

Results for this experiment are displayed in Table 5. The first row of the table reports the *2SLS* estimate obtained using the whole sample, which is identical to the result reported in Angrist and Krueger (1991) and Donald and Newey (2001). The remaining rows report the estimated *MSE*, bias, variance, and standard deviation

3. In particular we take observations spaced by 658 places, e.g. subsample 1 is composed by observations 1, 659, 1,317,..., 328,343, subsample 2 is composed by observations 2, 660, 1318,..., 328,344, and so on until the last subsample considered which is composed by observations 600, 1258, 1,916,..., 328,942.

TABLE 5

APPLICATION TO ANGRIST KRUEGER (1991) DATA

	Coefficient	Standard Error			
2SLS all sample	0.0928	0.0093			
	Average Coefficient	**Mean Squared Error**	**Bias**	**Variance**	**Standard Deviation**
2SLS	0.0360	0.0045	–0.0568	0.0012	0.0350
LIML	0.0007	0.0085	–0.0922	0.0000	0.0052
B2SLS	0.0000	0.0086	–0.0928	0.0000	0.0003
S2SLS					
$s_n = 0$	0.0360	0.0045	–0.0568	0.0012	0.0350
$s_n = 0.1$	0.0736	0.0008	–0.0192	0.0004	0.0202
$s_n = 0.5$	0.0859	0.0007	–0.0069	0.0006	0.0246
$s_n = 1$	0.0956	0.0008	0.0027	0.0008	0.0285
$s_n = 2$	0.1095	0.0015	0.0167	0.0012	0.0346
$s_n = 3$	0.1204	0.0023	0.0276	0.0016	0.0398
$s_n = 5$	0.1379	0.0044	0.0451	0.0024	0.0486
$s_n = 10$	0.1707	0.0106	0.0779	0.0045	0.0672
$s_n = 10^3$	2.0872	7.8401	1.9943	3.8627	1.9654
$s_n = s^*$	0.1450	0.0056	0.0522	0.0029	0.0537

NOTE : Average coef and the other statistics are computed by splitting the sample in 600 subsamples of 500 observations each. The used instruments are 60 exogenous variables plus 3 quarter of birth dummy, plus 27 interactions of the 3 quarter of birth dummies with 9 year of birth dummies, plus 150 interactions of the 3 quarter of birth dummies with 50 state of birth dummies, for a total of 240 instruments. If in some samples some dummies are not active causing perfect multicollinearity they are removed.

of the alternative estimators, computed by using the 600 subsamples of 500 observations. The estimators we consider are *2SLS, LIML, B2SLS,* and *2SLSS* (with several different values of the shrinkage parameter). As is clear from the table, *2SLS* seems superior to *B2SLS* and *LIML* in terms of *MSE*, and this is due to a much smaller bias[4].

On the other hand, there are values of the shrinkage parameter s_n (e.g. 0.1, 0.5 and 1) such that the *MSE* associated with *2SLSS* is more than six times smaller that that of *2SLS*, with gains coming both in the form of reduced bias and in reduced variance. Note that with this dataset a value of S_n between 0.1

4. Note that this result is by no means driven by the fact that we are using the 2*SLS* estimates on the whole sample rather than the *LIML* or the *B*2*SLS* as a proxy for the true value of the coefficient. Indeed, although the *LIML* and *B*2*SLS* estimates using the whole sample are different from those of 2*SLS*, the difference is both relatively small (*LIML* = 0.1064 and *B*2*SLS* = 0.1086) and implies even larger bias in the subsample estimates.

and 5 is reasonably large if one compares it with the scale of the data. To give a rough idea for this, we note that, for the present dataset, the trace of the matrix Z'Z is about 1,670,499, so that 1,670,499/(nK_n) is roughly equal to 0.02. We conclude that for a reasonable set of values for the shrinkage parameter 2SLSS can perform as well and in some case much better than existing estimators. Only when the shrinkage parameter deviates considerably from reasonable values, as discussed above, is 2SLSS performing badly. We consider this result as suggestive of a considerable amount of robustness for 2SLSS with respect to this tuning parameter.

Conclusion

Estimation of structural equations using instrumental variable techniques, in the presence of a large number of, possibly weak, instruments, is a topic that has received substantial attention in the literature. Most work has focused on the properties of existing estimators in the case of many, possibly weak, instruments. These estimators include the 2SLS estimator and the LIML estimator.

This paper is part of a small literature that discusses estimators that can be of particular relevance when many instruments are available. Intuition and recent work (see, e.g., Hahn,2002) suggests that parsimonious devices used in the construction of the final instruments may provide effective estimation strategies. Shrinkage is a well known approach that promotes parsimony. We consider a new shrinkage 2SLS estimator. We derive a consistency result for this estimator under general conditions, and, via both Monte Carlo simulations and an empirical application, show that it has also good potential for inference in small samples.

An open and interesting question for future research relates to the choice of the shrinkage parameter, s_n. It is of interest to develop a data-dependent way of determining this. An interesting possibility is to derive approximations of the MSE of the 2SLS shrinkage estimator and optimise the choice of s_n with respect to this measure, in the spirit of Donald and Newey (2001) We consider such an investigation to be the next step in our research agenda on this topic.

APPENDIX

Lemmas

Lemma 1–*Let assumptions 1-2 hold. Let* $P_{Z_n}^{s_n} = Z_n(Z_n'Z_n + s_nI)^{-1}Z_n'$. *Define* $q_n = s_n / n$ *such that* $q_n \to \infty$, $\frac{K_n}{q_n} \to \infty$ *and* $\frac{r_n}{q_n} \to \infty$. *Then, (i) for some constant C,*

$$V_n' P_{Z_n}^{s_n} V_n \Big/ \left(\frac{CK_n}{q_n}\right) = \Sigma_{VV} + O_p\left(\sqrt{\frac{q_n}{K_n}}\right),$$

(ii) for some constant C,

$$V_n' P_{Z_n}^{s_n} u_n / \left(\frac{CK_n}{q_n} \right) = \sigma_{Vu} + O_p \left(\sqrt{\frac{q_n}{K_n}} \right),$$

(iii)

$$V_n' P_{Z_n}^{s_n} Z_n \Pi_n / \left(\frac{r_n}{q_n} \right) = O_p \left(\sqrt{\frac{q_n}{r_n}} \right)$$

and (iv)

$$u_n' P_{Z_n}^{s_n} Z_n \Pi_n / \left(\frac{r_n}{q_n} \right) = O_p \left(\sqrt{\frac{q_n}{r_n}} \right).$$

Proof–C denotes constants which may be differ across contexts. To prove (i) it is sufficient to prove the statement for the *g, h*-th element of $V_n' P_{Z_n}^{s_n} V_n$ denoted by $V_{gn}' P_{Z_n}^{s_n} V_{hn}$ where V_{gn} denotes the *g*-th column of V_n. It is sufficient to show that

$$E\left(V_{gn}' P_{Z_n}^{s_n} V_{hn} / \left(\frac{CK_n}{q_n} \right) - \Sigma_{VV}^{gh} \right)^2 = O_p \left(\frac{q_n}{K_n} \right),$$

where Σ_{VV}^{gh} denotes the *g, h*-th element of Σ_{vv}. So, denoting the (*i,j*)-th element of $P_{Z_n}^{s_n}$ by $p_{ij,n}^{s_n}$, we have

$$E\left(V_{gn}' P_{Z_n}^{s_n} V_{hn} / \left(\frac{CK_n}{q_n} \right) - \Sigma_{VV}^{gh} \right)^2 =$$

$$\left(\frac{q_n}{CK_n} \right)^2 \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} E\left(p_{ij,n}^{s_n} p_{kl,n}^{s_n} \right) E\left(v_{ig} v_{jh} v_{kg} v_{lh} \right) -$$

$$\frac{2q_n \Sigma_{VV}^{gh}}{CK_n} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left(p_{ij,n}^{s_n} \right) E\left(v_{ig} v_{jh} \right) + \left(\Sigma_{VV}^{gh} \right)^2 =$$

$$\left(\frac{q_n}{CK_n} \right)^2 E\left(v_{ig}^2 v_{jh}^2 \right) \left[\sum_{i=1}^{n} E\left(\left(p_{ii,n}^{s_n} \right)^2 \right) \right] + \left(\frac{\sqrt{2} q_n}{CK_n} \right)^2 \Sigma_{VV}^{gg} \Sigma_{VV}^{hh} \left[\sum_{i=2}^{n} \sum_{j=1}^{i-1} E\left(\left(p_{ij,n}^{s_n} \right)^2 \right) \right] +$$

$$\left\{ \left(\frac{\sqrt{2} \Sigma_{VV}^{gh} q_n}{CK_n} \right)^2 \left[\sum_{i=2}^{n} \sum_{j=1}^{i-1} E\left(p_{ii,n}^{s_n} p_{jj,n}^{s_n} + \left(p_{ii,n}^{s_n} \right)^2 \right) \right] - \right.$$

$$\left. \frac{2\left(\Sigma_{VV}^{gh} \right)^2 q_n}{CK_n} \sum_{i=1}^{n} E\left(p_{ii,n}^{s_n} \right) + \left(\Sigma_{VV}^{gh} \right)^2 \right\} = A_{1n} + A_{2n} + A_{3n} .$$

We examine each of A_{1n}, A_{2n} and A_{3n} in turn. Starting with A_{1n} we have that

$$\left(\frac{q_n}{CK_n}\right)^2 E\left(v_{ig}^2 v_{jh}^2\right)\left[\sum_{i=1}^{n} E\left(\left(p_{ii,n}^{s_n}\right)^2\right)\right] \leq$$

$$\left(\frac{q_n}{CK_n}\right)^2 \sqrt{E\left(v_{ig}^4\right)}\sqrt{E\left(v_{jh}^4\right)}\left[\sum_{i=1}^{n} E\left(\left(p_{ii,n}^{s_n}\right)^2\right)\right] \leq$$

$$\left(\frac{q_n}{CK_n}\right)\sqrt{E\left(v_{ig}^4\right)}\sqrt{E\left(v_{jh}^4\right)} = O\left(\frac{q_n}{K_n}\right).$$

The second inequality follows from the fact that

$$\sum_{i=1}^{n} E\left(\left(p_{ii,n}^{s_n}\right)^2\right) \leq \sum_{i=1}^{n} E\left(p_{ii,n}^{s_n}\right)$$

which follows from the fact that $0 \leq p_{ii,n}^{s_n} \leq 1$. This in turn follows from the fact that $0 \leq p_{ii,n} \leq 1$ where $p_{ii,n}$ is the (i,i)-th element of $Z_n(Z_n'Z_n)^{-1}Z_n'$. The result then follows from Lemma 2. Next, focusing on A_{2n} we have that

$$\left(\frac{\sqrt{2}q_n}{CK_n}\right)^2 \Sigma_{VV}^{gg}\Sigma_{VV}^{hh}\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1} E\left(\left(p_{ij,n}^{s_n}\right)^2\right)\right] \leq$$

$$\left(\frac{q_n}{CK_n}\right)^2 \Sigma_{VV}^{gg}\Sigma_{VV}^{hh}\left[\sum_{i=1}^{n} E\left(\left(p_{ii,n}^{s_n}\right)^2\right) + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1} E\left(\left(p_{ij,n}^{s_n}\right)^2\right)\right].$$

But

$$\left[\sum_{i=1}^{n} E\left(\left(p_{ii,n}^{s_n}\right)^2\right) + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1} E\left(\left(p_{ij,n}^{s_n}\right)^2\right)\right] = tr\left(E\left(P_{Z_n}^{s_n}P_{Z_n}^{s_n}\right)\right) \leq \frac{CK_n}{q_n^2}$$

by Lemma 2. So

$$\left(\frac{\sqrt{2}q_n}{CK_n}\right)^2 \Sigma_{VV}^{gg}\Sigma_{VV}^{hh}\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1} E\left(\left(p_{ij,n}^{s_n}\right)^2\right)\right] \leq \frac{Cq_n\Sigma_{VV}^{gg}\Sigma_{VV}^{hh}}{K_n} = O\left(\frac{q_n}{K_n}\right).$$

Finally, we consider A_{3n}. We have that

$$|A_{3n}| = \left|\left(\frac{\sqrt{2}\Sigma_{VV}^{gh}q_n}{CK_n}\right)^2\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1} E\left(p_{ii,n}^{s_n}p_{jj,n}^{s_n} + \left(p_{ii,n}^{s_n}\right)^2\right)\right] - \right.$$

$$\left.\frac{2\left(\Sigma_{VV}^{gh}\right)^2 q_n}{CK_n}\sum_{i=1}^{n} E\left(p_{ii,n}^{s_n}\right) + \left(\Sigma_{VV}^{gh}\right)^2\right| =$$

$$\left|\left(\frac{\Sigma_{VV}^{gh} q_n}{CK_n}\right)^2 \left\{\left(tr\left(E\left(P_{Z_n}^{s_n}\right)\right)\right)^2 + tr\left(E\left(P_{Z_n}^{s_n} P_{Z_n}^{s_n}\right)\right) - 2\sum_{i=1}^{n} E\left(\left(p_{ii,n}^{s_n}\right)^2\right)\right\} - \right.$$

$$\left. \frac{2\left(\Sigma_{VV}^{gh}\right)^2 q_n}{CK_n} \sum_{i=1}^{n} E\left(p_{ii,n}^{s_n}\right) + \left(\Sigma_{VV}^{gh}\right)^2 \right| \leq$$

$$\left|\left(\frac{\Sigma_{VV}^{gh} q_n}{CK_n}\right)^2 \left\{\left(tr\left(E\left(P_{Z_n}^{s_n}\right)\right)\right)^2 - 2\sum_{i=1}^{n} E\left(\left(p_{ii,n}^{s_n}\right)^2\right)\right\} - \right. \tag{10}$$

$$\left. \frac{2\left(\Sigma_{VV}^{gh}\right)^2 q_n}{CK_n} \sum_{i=1}^{n} E\left(p_{ii,n}^{s_n}\right) + \left(\Sigma_{VV}^{gh}\right)^2 \right| + \left|\left(\frac{\Sigma_{VV}^{gh} q_n}{CK_n}\right)^2 tr\left(E\left(P_{Z_n}^{s_n} P_{Z_n}^{s_n}\right)\right)\right|.$$

The second term of (10) is $O\left(\frac{1}{K_n}\right) = o\left(\frac{q_n}{K_n}\right)$ by Lemma 2. Focusing on the first term we note that by Lemma 2 there exists a constant such that $tr\left(E\left(P_{Z_n}^{s_n}\right)\right) = \frac{CK_n}{q_n}$. Thus, we have

$$\left|\left(\frac{\Sigma_{VV}^{gh} q_n}{CK_n}\right)^2 \left\{\left(tr\left(E\left(P_{Z_n}^{s_n}\right)\right)\right)^2 - 2\sum_{i=1}^{n} E\left(\left(p_{ii,n}^{s_n}\right)^2\right)\right\} - \right.$$

$$\left. \frac{2\left(\Sigma_{VV}^{gh}\right)^2 q_n}{CK_n} \sum_{i=1}^{n} E\left(p_{ii,n}^{s_n}\right) + \left(\Sigma_{VV}^{gh}\right)^2 \right| \leq$$

$$\frac{\left(\Sigma_{VV}^{gh}\right)^2}{K_n} + \frac{2\left(\Sigma_{VV}^{gh}\right)^2 q_n^2 \sum_{i=1}^{n} E\left(\left(p_{ii,n}^{s_n}\right)\right)}{K_n^2} \leq \frac{C\left(\Sigma_{VV}^{gh}\right)^2 q_n}{K_n} = O\left(\frac{q_n}{K_n}\right),$$

where the first inequality follows from Lemma 2. This concludes the proof of part (i) of Lemma 1. Part (ii) is proven similarly. Next, we move on to part (iii). We have

$$E\left(\left\|\frac{q_n V_n' P_{Z_n}^{s_n} \Pi_n}{r_n}\right\|\right)^2 =$$

$$E\left(\left\|\frac{q_n^2 \Pi_n' Z_n' Z_n (Z_n' Z_n + s_n I)^{-1} Z_n' V_n V_n' Z_n (Z_n' Z_n + s_n I)^{-1} Z_n' Z_n \Pi_n}{r_n^2}\right\|\right)^2 =$$

$$tr\left(\Sigma_{VV}\right)E\left(tr\left[\frac{q_n^2\Pi_n'Z_n'Z_n(Z_n'Z_n+s_nI)^{-1}Z_n'Z_n(Z_n'Z_n+s_nI)^{-1}Z_n'Z_n\Pi_n}{r_n^2}\right]\right)\leq$$

$$\frac{q_n}{r_n}tr\left(\Sigma_{VV}\right)E\left(tr\left[\frac{q_n\Pi_n'Z_n'Z_n(Z_n'Z_n+s_nI)^{-1}Z_n'Z_n(Z_n'Z_n+s_nI)^{-1}Z_n'Z_n\Pi_n}{r_n}\right]\right)\leq$$

$$\frac{q_n}{r_n}tr\left(\Sigma_{VV}\right)E\left(tr\left[\frac{q_n\Pi_n'Z_n'Z_n(Z_n'Z_n+s_nI)^{-1}Z_n'Z_n\Pi_n}{r_n}\right]\right)\leq$$

$$\frac{q_nC}{r_n}=O\left(\frac{q_n}{r_n}\right),$$

where the last but one inequality follows by the second part of assumption 1(iii) and the last by the first part of assumption 1(iii). Part (iv) can be proved similarly.

Lemma 2–*Let* $P_{Z_n}^{s_n}=Z_n(Z_n'Z_n+s_nI)^{-1}Z_n'$ *and let assumptions 1-2 hold. Then, for all n,* $tr\left(P_{Z_n}^{s_n}\right)$*, and therefore* $E\left(tr\left(P_{Z_n}^{s_n}\right)\right)$ *are bounded from above by* $\frac{CnK_n}{s_n}$*. Further, for all n,* $tr\left(\left(P_{Z_n}^{s_n}\right)^2\right)$*, and therefore,* $E\left(tr\left(\left(P_{Z_n}^{s_n}\right)^2\right)\right)$*, are bounded from above by* $\frac{Cn^2K_n}{s_n^2}$.

Proof–We have that

$$tr\left(Z_n(Z_n'Z_n+s_nI)^{-1}Z_n'\right)=tr\left(Z_n'Z_n(Z_n'Z_n+s_nI)^{-1}\right)=$$

$$tr\left(\frac{Z_n'Z_n}{n}\left(\frac{Z_n'Z_n}{n}+\frac{s_n}{n}I\right)^{-1}\right)=tr\left(\left(I+\frac{s_n}{n}\left(\frac{Z_n'Z_n}{n}\right)^{-1}\right)^{-1}\right).\qquad(11)$$

We next use standard results (see, e.g., Bai and Golub, 1997) on upper bounds of the trace of the inverse of a matrix. We first note the following: By assumption 1(iv), for all *n*, all eigenvalues of $\frac{Z_n'Z_n}{n}$, and therefore those of $\left(\frac{Z_n'Z_n}{n}\right)^{-1}$, are bounded and bounded away from zero. As a result all eigenvalues of $\frac{s_n}{n}\left(\frac{Z_n'Z_n}{n}\right)^{-1}$, and therefore of $I+\frac{s_n}{n}\left(\frac{Z_n'Z_n}{n}\right)^{-1}$, are $O\left(\frac{s_n}{n}\right)$. Then, we use Kantorovich's inequality for a square *m* x *m* matrix *A*, which states that for the *i,i*-th element of A^{-1} the following holds:

$$(A^{-1})_{ii}\leq\frac{1}{4a_{ii}}\left(\frac{\alpha}{\beta}+\frac{\beta}{\alpha}+2\right),\qquad(12)$$

where a_{ii} is the i,i-th element of A, $\alpha = \min_i \lambda_i(A)$, $\beta = \max_i \lambda_i(A)$ and λ_i denotes the i-th eigenvalue of A. Applying this result to our case gives

$$tr\left(\left(I + \frac{s_n}{n}\left(\frac{Z_n'Z_n}{n}\right)^{-1}\right)^{-1}\right) \leq C\frac{nK_n}{s_n}$$

noting firstly that both $\frac{\alpha}{\beta}$ and $\frac{\beta}{\alpha}$, in (12), in our case are $O(1)$ and secondly that $a_{ii} = O\left(\frac{s_n}{n}\right)$. Note that since we assume that $n = o(s_n)$, it follows that

$$\lim_{n\to\infty} E\left(tr\left(Z_n(Z_n'Z_n + s_nI)^{-1}Z_n'\right)\right) = o(K_n).$$

This proves the first part of the Lemma. Given the preceding analysis, in order to prove the second part of the Lemma, it is simply sufficient to analyse the behaviour of $tr\left(\left(I + \frac{s_n}{n}\left(\frac{Z_n'Z_n}{n}\right)^{-1}\right)^{-2}\right)$. Then, we have

$$tr\left(\left(I + \frac{s_n}{n}\left(\frac{Z_n'Z_n}{n}\right)^{-1}\right)^{-2}\right) =$$

$$tr\left(\left(\left(I + \frac{s_n}{n}\left(\frac{Z_n'Z_n}{n}\right)^{-1}\right)\left(I + \frac{s_n}{n}\left(\frac{Z_n'Z_n}{n}\right)^{-1}\right)\right)^{-1}\right) =$$

$$tr\left(\left(I + 2\frac{s_n}{n}\left(\frac{Z_n'Z_n}{n}\right)^{-1} + \left(\frac{s_n}{n}\right)^2\left(\frac{Z_n'Z_n}{n}\right)^{-2}\right)^{-1}\right) = tr\left(\left(I + \tilde{\Sigma}_n\right)^{-1}\right),$$

where

$$\tilde{\Sigma}_n = 2\frac{s_n}{n}\left(\frac{Z_n'Z_n}{n}\right)^{-1} + \left(\frac{s_n}{n}\right)^2\left(\frac{Z_n'Z_n}{n}\right)^{-2}.$$

Since all eigenvalues of $\left(\frac{Z_n'Z_n}{n}\right)^{-1}$ are bounded, and bounded away from zero, the same follows for the eigenvalues of $\left(\frac{Z_n'Z_n}{n}\right)^{-2}$. Then, it is easy to see that all eigenvalues of $I + \tilde{\Sigma}_n$ are $O\left(\left(\frac{s_n}{n}\right)^2\right)$. Then, by a similar analysis as that used for $tr\left(\left(I + \frac{s_n}{n}\left(\frac{Z_n'Z_n}{n}\right)^{-1}\right)^{-1}\right)$ we have that

$$tr\left(\left(I + \frac{s_n}{n}\Sigma^{-1}\right)^{-2}\right) \leq C\frac{n^2K_n}{s_n^2}.$$

Proof of Theorem 1

We have that

$$\hat{\beta}_{2SLSS} - \beta_0 = \left(Y_{2n}' P_{Z_n}^{s_n} Y_{2n}\right)^{-1} \left(Y_{2n}' P_{Z_n}^{s_n} u_n\right) = \left(\frac{q_n Y_{2n}' P_{Z_n}^{s_n} Y_{2n}}{r_n}\right)^{-1} \left(\frac{q_n Y_{2n}' P_{Z_n}^{s_n} u_n}{r_n}\right). \quad (13)$$

We analyse each term of the product of the RHS of (13) in turn. We have,

$$\frac{q_n Y_{2n}' P_{Z_n}^{s_n} Y_{2n}}{r_n} = \frac{q_n \Pi_n' Z_n' P_{Z_n}^{s_n} Z_n \Pi_n}{r_n} + \frac{q_n V_n' P_{Z_n}^{s_n} Z_n \Pi_n}{r_n} +$$

$$\frac{q_n \Pi_n' Z_n' P_{Z_n}^{s_n} V_n}{r_n} + \left(\frac{K_n}{r_n}\right) \frac{q_n V_n' P_{Z_n}^{s_n} V_n}{K_n} \xrightarrow{p} \Psi$$

by the fact that $\frac{K_n}{r_n} \to 0$, Assumption 1(iii) and Lemma 1(i),(iii). Next, we have that

$$\frac{q_n Y_{2n}' P_{Z_n}^{s_n} u_n}{r_n} = \frac{q_n \Pi_n' Z_n' P_{Z_n}^{s_n} u_n}{r_n} + \left(\frac{K_n}{r_n}\right) \frac{q_n V_n' P_{Z_n}^{s_n} u_n}{K_n}.$$

By Lemma 1 (ii), $\frac{q_n V_n' P_{Z_n}^{s_n} u_n}{K_n} = O_p(1)$ and so

$$\left(\frac{K_n}{r_n}\right) \frac{q_n V_n' P_{Z_n}^{s_n} u_n}{K_n} = O_p\left(\frac{K_n}{r_n}\right).$$

Further, by lemma 1 (iv)

$$\frac{q_n \Pi_n' Z_n' \tilde{P}_{Z_n}^{s_n} u_n}{r_n} = O_p\left(\sqrt{\frac{q_n}{r_n}}\right).$$

Overall,

$$\frac{q_n Y_{2n}' \tilde{P}_{Z_n}^{s_n} u_n}{r_n} = O_p\left(\frac{K_n}{r_n}\right) + O_p\left(\sqrt{\frac{q_n}{r_n}}\right) = o_p(1)$$

under the assumptions of Theorem 1.

REFERENCES

ANGRIST, J., and A. KRUEGER (1991). "Bootstrap and Randomization Tests of Some Nonparametric Hypotheses." *Quarterly Journal of Economics*, 106: 979-1014.

BAI, J., and S. NG (2010). "Instrumental Variable Estimation in a Data Rich Environment." *Econometric Theory*, 26: 1577-1606.

BAI, Z., and G. H. GOLUB (1997). "Bounds for the Trace of the Inverse and the Determinant of Symmetric Positive Definite Matrices." *Annals of Numerical Mathematics*, 4: 29-38.

BEKKER, P. A. (1994). "Alternative Approximations to the Distributions of Instrumental Variable Estimators." *Econometrica*, 62: 657- 681.

CHAO, J. C., and N. R. SWANSON (2004). "Asymptotic Normality of Single-Equation Estimators for the Case with a Large Number of Weak Instruments." in CORBAE, D., S. DURLAUF and B. HANSEN (Eds), *Frontiers of Analysis and Applied Research: Essays in Honor of Peter C. B. Phillips*, Cambridge, U.K.: Cambridge University Press.

CHAO, J. C., and N. R. SWANSON (2005), "Consistent Estimation with a Large Number of Weak Instruments." *Econometrica*, 73: 1673-1692.

DONALD, S. G., and W. K. NEWEY (2001), "Choosing the Number of Instruments." *Econometrica*, 69: 1161-1191.

HAHN, J. (2002), "Optimal Inference with Many Instruments." *Econometric Theory*, 18: 140-168.

HAHN, J., J. HAUSMAN, and G. KUERSTEINER (2001). "Accuracy of Higher Order Bias and MSE Approximations." *Econometrics Journal*, 7: 272-306.

HANSEN, C., J. HAUSMAN, and W. K. NEWEY (2006). "Many Instruments, Weak Instruments and Microeconometric Practice." Working Paper, MIT.

KAPETANIOS, G., and M. MARCELLINO (2007). "Cross-sectional Averaging and Instrumental Variable Estimation with Many Weak Instruments." *Economics Letters*, 108: 36-39.

KAPETANIOS, G., and M. MARCELLINO (2010). "Factor-GMM Estimation with Large Sets of PossiblyWeak Instruments." *Computational Statistics and Data Analysis*, 54: 2655-2675.

MORIMUNE, K. (1983). "Approximate Distributions of k-class Estimators when the Degree of Overidentification is Large Compared with the Sample Size." *Econometrica*, 51: 821-842.

NEWEY, W. K. (2004). "Many Weak Moment Asymptotics for the Continuously Updated GMM Estimator." Working Paper, MIT.

PHILLIPS, P. C. B. (1983). "Exact Small Sample Theory in the Simultaneous Equations Model." in Z. GRILICHES and M. D. INTRILIGATOR (Eds), *Handbook of Econometrics*, Vol. I, Amsterdam: North-Holland, pp. 449-516.

ROTHENBERG, T. (1984). "Approximating the Distributions of Econometric Estimators and Test Statistics." in Z. GRILICHES and M. D. INTRILIGATOR (Eds), *Handbook of Econometrics*, Vol. I, Amsterdam: North-Holland, pp. 881.935.

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

Stock, J. H., and M. Yogo (2003a). “Asymptotic Distributions of Instrumental Variables Statistics with Many Weak Instruments’.” in Andrews, D. W. K. and J. H. Stock (Eds), *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, Cambridge, U.K.: Cambridge University Press.

Stock, J. H., and M. Yogo (2003b). “Testing for Weak Instruments in Linear IV Regression.” in Andrews, D. W. K. and J. H. Stock (Eds), *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, Cambridge, U.K.: Cambridge University Press.