Federal Reserve Bank of New York
Staff Reports

Parsimonious Estimation with Many Instruments

Jan J. J. Groen
George Kapetanios

Staff Report no. 386
August 2009

Parsimonious Estimation with Many Instruments
Jan J. J. Groen and George Kapetanios
Federal Reserve Bank of New York Staff Reports, no. 386
August 2009
JEL classification: C30, C59, C13

Abstract

We suggest a way to perform parsimonious instrumental variables estimation in the presence of many, and potentially weak, instruments. In contrast to standard methods, our approach yields consistent estimates when the set of instrumental variables complies with a factor structure. In this sense, our method is equivalent to instrumental variables estimation that is based on principal components. However, even if the factor structure is weak or nonexistent, our method, unlike the principal components approach, still yields consistent estimates. Indeed, simulations indicate that our approach always dominates standard instrumental variables estimation, regardless of whether the factor relationship underlying the set of instruments is strong, weak, or absent.

Key words: instrumental variables estimation, many instruments, factor models

---

Groen: Federal Reserve Bank of New York (e-mail: jan.groen@ny.frb.org). Kapetanios: Queen Mary University of London (e-mail: g.kapetanios@qmul.ac.uk). The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

1 Introduction

Recent work in instrumental variable estimation has considered two distinct routes. The first is one where instrumental variables are only weakly correlated with the endogenous explanatory variables of an instrumental variables (IV) regression. Work by, e.g., Phillips (1983), Rothenberg (1984), Stock and Yogo (2003b) and Chao and Swanson (2005) consider a natural measure of instrument weakness (or strength) in a linear IV framework to be the so-called concentration parameter. 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. In the case of weak instruments, the properties of IV estimators such as two stage least squares (2SLS) and limited information maximum likelihood (LIML) are affected relative to the case of strong instruments and the estimators may, in fact, be inconsistent.

Another direction in IV research involves the case where the number of available instruments is large. This approach was first taken by Morimune (1983) and later generalized by Bekker (1994). Other relevant papers include Donald and Newey (2001), Hahn *et al.* (2001), Hahn and Kuersteiner (2002), and Chao and Swanson (2004). More recently, the two different stands have been combined to provide a comprehensive framework for the analysis of the properties of IV estimators in the case of many weak instruments. Work on this includes Hansen *et al.* (2006), Stock and Yogo (2003a), Newey (2004) and Chao and Swanson (2005). A clear conclusion from this work suggests that inconsistency of IV estimators is a probable outcome when many weak instruments are used.

With this in mind, further recent developments focus on considering parsimonious modeling assumptions for the set of instruments to avoid IV estimator inconsistency. In particular, Kapetanios and Marcellino (2006) 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. This factor-based IV estimation is powerful when factor structures exist but loses its desirable properties in the absence of a factor structure. In this paper we suggest an alternative parsimonious approach for carrying out IV estimation in the presence of many (and therefore potentially partially weak) instruments. This approach is based on using partial least squares (PLS) to obtain a fit for each of the endogenous explanatory variables using the instruments. The advantages of using PLS based IV estimation are many. We can show that, in the presence of a factor structure, PLS is equivalent to principal component based Factor IV estimation. However, if the factor structure is weak or even non-existent PLS IV remains consistent.

The paper is structured as follows: Section 2 develops the theoretical properties of the PLS instrumental variable estimator. Section 3 studies the finite sample properties of PLS-IV estimation using Monte Carlo experiments. Finally, Section 4 summarizes and concludes. All proofs are contained in the Appendix.

2 Theory

Let the equation of interest be

$$y_t = x_t'\beta + \epsilon_t, \quad t = 1, \ldots, T, \tag{1}$$

where the k regressors in $x_t = (x_{1,t}, ..., x_{k,t})'$ are possibly correlated with the error term ϵ_t. A usual source of endogeneity is, of course, simultaneity, which is widespread in applied macroeconomic applications based on single equation estimation. Let us assume that there exist N instrumental variables, $z_t = (z_{1,t}, ..., z_{N,t})'$, generated by a factor model with $r \geq k$ unobservable factors, $f_t = (f_{1,t}, ..., f_{r,t})'$:

$$z_t = \Lambda^{0'} f_t + v_t, \tag{2}$$

where r is much smaller than N. Therefore, each instrumental variable can be decomposed into a common component (an element of $\Lambda^{0'} f_t$) that is driven by a few common forces, the factors, and an idiosyncratic component (an element of v_t). When the latter is small compared to the former, the information in the large set of N instrumental variables z_t can be efficiently summarized by the r factors f_t. We will not solely focus our analysis to the above factor model since we will consider the case $\Lambda^0 = 0$ which implies the lack of a factor structure.

As the data generation mechanisms for x_t we consider

$$x_t = A_Z^{0'} z_t + u_t, \tag{3}$$

with $E(u_t'\epsilon_t) \neq 0$ to introduce simultaneity in (1).

We will consider estimation of β based on a variant of two stage least squares where each of the first stage regressions ($x_{j,t}$ on z_t) is carried out using partial least squares (PLS) and for this we now review PLS. Introduced by Herman Wold and co-workers between 1975 and 1982,[1] partial least squares (PLS) is a relatively new method for estimating regression equations, that has received much attention in a variety of disciplines and especially in chemometrics. The basic idea is similar to principal component analysis in that factors or components, which are linear combinations of the original regression variables, are used, instead of the original variables, as regressors. A major difference between PC and PLS is that, whereas in PC regressions the factors are constructed taking into account only the values of the z_t variables, in PLS, the relationship between x_t and z_t is considered as well in constructing the factors.

There are a variety of definitions for PLS and accompanying specific PLS algorithms that inevitably have much in common. We provide a brief description of these definitions and algorithms we feel are most appropriate for conveying the essence of PLS. A conceptually powerful way of defining PLS is to note that the PLS factors are those linear combinations of z_t, denoted by Λz_t, that give maximum covariance between x_t and Λz_t while being orthogonal to each other. Of course, in analogy to PC factors, an identification assumption is needed, to construct PLS factors, in the usual form of a normalization.

A simple algorithm to construct k_1 PLS factors is discussed among others, in detail, in Helland (1990). Assuming for simplicity that both x_t and z_t have been normalised to have zero mean, a simplified version of the algorithm for $x_{j,t}$ is given below

Algorithm 1

1. *Set* $u_{j,t} = x_{j,t}$ *and* $v_t = (v_{1,t}, , ..., v_{N,t})'$, $v_{i,t} = z_{i,t}$, $i = 1, ...N$. *Set* $s = 1$.

2. *Determine the* $N \times 1$ *vector of indicator variable weights or loadings* $w_{js} = (w_{1js} \cdots w_{Njs})'$ *by computing individual covariances:* $w_{ijs} = Cov(u_{j,t},\ v_{i,t})$, $i = 1, ..., N$. *Construct the s-th PLS factor by taking the linear combination given by* $w_{js}' v_t$ *and denote this factor by* $f_{j,s,t}$.

[1] See, e.g., Wold (1982).

3. *Regress $u_{j,t}$ and $v_{i,t}$, $i = 1, ..., N$ on $f_{j,s,t}$. Denote the residuals of these regressions by $\tilde{u}_{j,t}$ and $\tilde{v}_{i,t}$ respectively.*

4. *If $s = k_1$ stop, else set $u_{j,t} = \tilde{u}_{j,t}$, $v_{i,t} = \tilde{v}_{i,t}$ $i = 1, .., N$ and $s = s + 1$ and go to step 2.*

This algorithm makes clear that PLS is computationally tractable for very large data sets. Once PLS factors are constructed x_t can be modeled or forecast by regressing $x_{j,t}$ on $f_{j,s,t}$ $s = 1, ..., k_1$. Helland (1988, 1990) provide a general description of the partial least squares (PLS) regression problem. Helland (1988) shows that the PLS estimates of the regression coefficients, α_j, of the regression of $x_{j,t}$ on z_t obtained implicitly via PLS Algorithm 1 and a regression of $x_{j,t}$ on $f_{j,s,t}$ $s = 1, ..., m$, can be equivalently obtained by the following formula

$$\hat{\alpha}_{j,PLS} = V_{k_1}(V'_{k_1} Z' Z V_{k_1})^{-1} V'_{k_1} Z' x_j \tag{4}$$

where $V_{k_1} = (Z'x_j, Z'ZZ'x_j, ..., (Z'^{k_1-1} Z'x_j)$, $Z = (z_1, ..., z_T)'$ and $x_j = (x_{j,1}, ..., x_{j,T})'$.

Having briefly reviewed PLS we refer to Groen and Kapetanios (2008) for a more detailed discussion. Next, we introduce our estimator. Stacking observations across time for our model presented above gives:

$$y = X\beta + \epsilon \tag{5}$$

$$Z = F\Lambda^0 + v \tag{6}$$

$$X = ZA^0_Z + u \tag{7}$$

where $y = (y_1, ..., y_T)'$, $F = (f_1, ..., f_T)'$, $u = (u_1, ..., u_T)'$, $v = (v_1, ..., v_T)'$ and $\epsilon = (\epsilon_1, ..., \epsilon_T)'$. We define the PLS-IV estimator as:

$$\hat{\beta}_{2SPLS} = \left(X'_{PLS} X\right)^{-1} X'_{PLS} y, \tag{8}$$

where $X_{PLS} = (x_{1,PLS}, \ldots, x_{T,PLS})'$ denote the matrix of fitted values for X using a PLS regression.

We make the following assumptions which are standard in the principal component (PC) based factor estimation literature.

Assumption 1 *1. $E||f_t||^4 \leq M < \infty$, $T^{-1} \sum_{t=1}^{T} f_t f_t' \xrightarrow{p} \Sigma_f$ for some $r \times r$ positive definite matrix Σ_f. Λ^0 has bounded elements. Further $||\Lambda^0 \Lambda^0 / N - D|| \to 0$ where D is a positive definite matrix.*

2. *$E(v_{i,t})=0$, $E|v_{i,t}|^8 \leq M$ where $v_t=(v_{1,t},...,v_{N,t})'$ The variance of v_t is denoted by Σ_v. f_s and v_t are independent for all s,t.*

3. *For $\tau_{i,j,t,s} \equiv E(v_{i,t}v_{j,s})$ the following hold*

   - $(NT)^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}|\sum_{i=1}^{N}\tau_{i,i,t,s}| \leq M$
   - $|1/N\sum_{i=1}^{N}\tau_{i,i,s,s}| \leq M$ *for all* s
   - $N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}|\tau_{i,j,s,s}| \leq M$
   - $(NT)^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}|\tau_{i,j,t,s}| \leq M$
   - *For every (t,s), $E|(N)^{-1/2}\sum_{i=1}^{N}(v_{i,s}v_{i,t}-\tau_{i,i,s,t})|^4 \leq M$*

Assumption 2 *ϵ_t is a martingale difference sequence with finite fourth moment and $E(\epsilon_t^2|\mathcal{F}_t)=\sigma^2<\infty$ where $\mathcal{F}_t$ is the σ-field generated by (f_s,z_s), $s \leq t$.*

Assumption 3 *(x_t',z_t') are jointly stationary. z_t is predetermined, so that $E(z_{it}\epsilon_t)=0$, $i=1,...,N$. The probability limit of $\frac{z_tz_t'}{T}$ is finite and nonsingular. $E(z_tx_t')$ has full column rank k. x_t and z_t have finite fourth moments.*

Assumption 1 is standard in the factor literature. In particular, it is used in, e.g., Stock and Watson (2002) and Bai (2003) to prove consistency and asymptotic normality (at certain rates) of the principal component based estimator of the factors, and by Bai and Ng (2006) to show consistency of the parameter estimators in factor augmented regressions. Assumption 3 guarantees that standard IV estimation using z_t as instruments is feasible, and Assumption 2 that it is efficient.

We make the following additional assumptions:

Assumption 4 *Let $\Sigma=\Sigma_N=[\sigma_{ij}]$ denote the $N \times N$ second moment matrix of Z. Σ can be factorised as follows:*

$$\Sigma=\tilde{S}\tilde{\Psi}\tilde{S}'+R$$

where $\tilde{S}\tilde{S}'=I$, $\tilde{\Psi}=diag(\tilde{\psi}_{N1},...,\tilde{\psi}_{Nr})$, $r<N$ and $||R||=o(N)$.

Assumption 5 *Uniformly over $i,j=1,...,N$*

$$\frac{1}{T}\sum_{t=1}^{T}x_{j,t}z_{i,t}-\sigma_{ji,xz}=O_p\left(T^{-1/2}\right)$$

where $\sigma_{ji,xz}=E\left(x_{j,t}z_{i,t}\right)$.

These assumptions deserve some comment. Assumption 4 states that the variables in Z are asymptotically with respect to N collinear. It is instructive to compare this assumption with the standard factor assumption. In one sense this assumption is stronger than the standard factor assumption because in standard factor models the sum of all bounded eigenvalues of the covariance matrix is $O(N)$ whereas under assumption 4 it is $o(N)$. On the other hand this assumption is weaker. Under a standard factor assumption there can be only a finite number of unbounded eigenvalues, and hence unbounded singular values, for the covariance matrix of the data. In our case we can have an infinity of unbounded eigenvalues as long as the sum of all but the first r eigenvalues is $o(N)$. In particular the remainder term, R, can, in fact, be parameterized as a neglected 'weak' factor model whose eigenvalue characterisation allows for unbounded eigenvalues which, however, have to grow at a rate slower than N. Assumption 2 is a mild, high level, assumption. It is sufficient to have a central limit theorem for $(z_{i,t}z_{j,t} - \sigma_{i,j})$ for this assumption to hold. The above assumptions are needed to derive an equivalence between PC and PLS which in turn will lead to desirable properties for PLS-IV following the analysis of Factor-IV of Kapetanios and Marcellino (2003).

We therefore need to impose both sets of assumptions to obtain our first result. Note that the combination of both sets of assumptions is basically restrictive only to the extent that it implies that the eigenvalues of the population covariance matrix of the idiosyncratic term are both bounded and their sum does not grow as fast as N. We now present our first result

Theorem 1 *Let assumptions 4-5 hold. Let $\hat{\beta}_{2SPLS}$ and $\hat{\beta}_{2SPC}$ denote the PLS IV estimator and Factor-IV estimator of Kapetanios and Marcellino (2003) respectively. Then, there exists a finite number of PLS factors such that for that number of PLS factors*

$$\sqrt{T}(\hat{\beta}_{2SPLS} - \hat{\beta}_{2SPC}) = o_p(1).$$

If further, assumptions 1-3 hold then, for the same number of PLS factors

$$\sqrt{T}(\hat{\beta}_{2SPLS} - \beta) \xrightarrow{d} N\left(0, \left(\left(A_Z^{0\prime}\Lambda_Z^{0\prime}\right)\Sigma_f\left(\Lambda_Z^0 A_Z^0\right)\right)^{-1}\right)$$

where Σ_f denotes the covariance matrix of f_t.

As we said in the Introduction, the above result is of secondary importance given the existence of Factor-IV. However, the next result provides a powerful justification for considering PLS-IV. For this, we need the following assumption

Assumption 6 *$E(z_t x_t')$ is full column rank. $E(z_t z_t')$ is nonsingular. Uniformly on j, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(z_{j,t} - E(z_{j,t})) \xrightarrow{d} N(0, \sigma_j^2)$ where σ_j^2 denotes the variance of $z_{j,t}$.*

Then, we have the following result

Theorem 2 *Let assumptions 5 and 6 hold. Let $\hat{\beta}_{2SPLS}$ denote the PLS IV estimator where the PLS factors have been obtained from steps 1-3 of Algorithm 1. Then,*

$$\left\|\hat{\beta}_{2SPLS} - \beta\right\| = O_p\left(\left(\frac{N}{T}\right)^{-1/2}\right)$$

This theorem has a number of implications. Firstly, a cursory examination of the proof immediately suggests that, for finite N, PLS-IV is $\sqrt{T}$-consistent and asymptotically normal. Secondly, we note that the result is not just a consistency result as is common in the literature on many instruments but also a rate result under reasonably mild assumptions. Finally, we note that the setup of (3) is one where not all instruments can be strong, in the sense that some elements of $A_Z^{0'}$ are either zero or implicitly become smaller as N increases. Two leading cases can be mentioned: One is where a finite subset of z_t are strong instruments, in the sense that their coefficients in $A_Z^{0'}$, remain bounded away from zero as N and the rest are weak. The other leading case is one where all instruments are weak. In both cases PLS-IV will be consistent as long as $\lim_{N\to\infty} E(z_t x_t')$ has full column rank. We do not explore formally such issues preferring to leave our formal analysis to the level of generality posited in Theorem 2.

3 Monte Carlo Study

The setup of the Monte Carlo experiments is:

$$y_t = \sum_{i=1}^{k} x_{it} + \epsilon_t \tag{9}$$

$$z_{it} = c\sum_{j=1}^{r} N^{-p} f_{jt} + c_1 e_{it}, \quad i = 1, ..., N \tag{10}$$

and either

$$x_{it} = \sum_{j=1}^{r} c_1^{-1} f_{jt} + u_{it}, \quad i = 1, ..., k \tag{11}$$

or

$$x_{it} = \sum_{j=1}^{N} N^{-q} d_j z_{jt} + u_{it}, \quad i = 1, ..., k \tag{12}$$

or

where $e_{it} \sim i.i.d.N(0,1)$, $f_{it} \sim i.i.d.N(0,1)$ and $cov(e_{it}, e_{sj}) = 0$ for $i \neq s$. Let $\kappa_t = (\epsilon_t, u_{1t}, ..., u_{kt})'$. Then, $\kappa_t = P\eta_t$, where $\eta_t = (\eta_{1,t}, ..., \eta_{k+1,t})'$, $\eta_{i,t} \sim i.i.d.N(0,1)$ and $P = [p_{ij}]$, $p_{ij} \sim i.i.d.N(0,1)$. The errors e_{it} and u_{is} are independent for each i and s. We only report results for $k = 1$, $r = 1$, since there are no qualitative changes by increasing the number of endogenous variables or of factors.[2]

The instrumental variables z_{it} are generated by the model in (10). The parameter p controls the "strength" of the factor structure. We consider the values $p = 0, 0.1, 0.25, 0.33, 0.45, 0.5$. When $p = 0$ we are in the standard case analyzed by, e.g., Stock and Watson (2002) and Bai (2003). When $p > 0$ we are in the weak factor structure case, as discussed in Kapetanios and Marcellino (2003). The parameter c_1 controls the relative size of the idiosyncratic component, so that a larger value of c_1 makes factor estimation harder, at least for small values of N. In the base case we set $c_1 = 1$, but we also consider experiments with $c_1 = 0.5$. Note also that c_1 appears in (10) so that high values of c_1 will reduce the influence of f_t on x_t. So, overall increasing the value of c_1 should have a negative influence on the performance of a factor based IV estimator although not necessarily PLS-IV in view of Theorem 2. We use two different generating mechanisms for x_{it}: (11) and (12). We also consider two cases that relate to the presence of factors. In the first case, $c = 1$, and in the second case, $c = 0$. If $c = 1$ we set $q = 1$ since the cross-sectional sum of z_{jt} needs to be normalised by N^{-1}, in this case, to remain bounded in probability. whereas the normalisation should be $N^{-1/2}$ when $c = 0$. Finally, if $c = 1$ we set $d_j = 1$ whereas if $c = 0$, we consider two cases: In the first case, we set $d_j \sim N(1,1)$. This implies that all the variables, z_{it}, get the same weight on average in determining x_{it}. In the second case we wish to have different weights on different z_{jt}. This is important as it is likely in practical applications for different instruments to be of different importance for the first-stage regression. Therefore, we adopt the scheme

[2] Results for k and r larger than one are available upon request.

used in Model 1 of the simulation study of Donald and Newey (2001), which is given by $d_j = c(N)\left(1 - \frac{j}{N+1}\right)^4$ where $c(N)$ is chosen so that the R^2 of (12) is 0.5. Results on the MSE of the PLS-IV and Standard IV estimators are reported in Tables 1-3.

Results make interesting reading. PLS-IV dominates standard IV in all cases considered. This dominance is clearest when there is a strong factor structure in the data but it extends to cases where the factor structure is weak or completely nonexistent. Interestingly and in accordance with the results of Theorem 2, the performance of PLS-IV does not deteriorate sharply as the factor structure weakens and, unlike Factor-IV, remains dominant over standard IV, irrespective of the coefficient pattern for the first-stage regression.

4 Conclusion

In this paper we suggest a parsimonious approach for carrying out IV estimation in the presence of many (and therefore potentially partially weak) instruments. This approach is based on using partial least squares (PLS) to obtain a fit for each of the endogenous explanatory variables using the instruments. The advantages of using PLS based IV estimation are many. We have shown that, in the presence of a factor structure, PLS is equivalent to principal component based Factor IV estimation. However, if the factor structure is weak or even non-existent PLS IV remains consistent and can even outperform IV estimation based on cross-sectional averages. Further PLS IV does not depend on a tuning parameter, at least to the extent that shrinkage IV estimation does. Our Monte Carlo study suggests that PLS-IV is as good as PC based factor IV when there exists a factor structure but remains better than standard IV when there is no factor structure. As a result this method combines the desirable properties of Factor-IV in the special case of a factor structure while still providing very good performance in the general case where no factor structure appears in the data.

References

Bai, J., 2003, Inferential Theory for Factor Models of Large Dimensions, *Econometrica* **71**, 135–173.

Bai, J. and S. Ng, 2006, Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions, *Econometrica* **74**, 1133–1150.

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* D. Corbae, S. Durlauf and B. Hansen (editors), *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.

Groen, J. J. J. and G. Kapetanios, 2008, Revisiting Useful Approaches to Data-Rich Macroeconomic Forecasting, *Working Paper 624*, Queen Mary, University of London.

Hahn, J. and G. Kuersteiner, 2002, Discontinuities of Weak Instrument Limiting Distributions, *Economics Letters* **75**, 325–331.

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* .

Helland, I. S., 1988, On the Structure of Partial Least Squares Regression, *Communications in Statistics - Simulation and Computation* **17**, 581–607.

Helland, I. S., 1990, Partial Least Squares Regression and Statistical Models, *Scandinavian Journal of Statistics* **17**, 97–114.

Kapetanios, G. and M. Marcellino, 2003, A Comparison of Estimation Methods for Dynamic Factor Models of Large Dimensions, *Queen Mary, University of London Working Paper No. 489* .

Kapetanios, G. and M. Marcellino, 2006, Factor-GMM Estimation with Large Sets of Possibly Weak Instruments, *Queen Mary University of London Working Paper No. 577* .

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 (editors), *Handbook of Econometrics, Vol. I*, Amsterdam: North-Holland, p. 449516.

Rothenberg, T., 1984, Approximating the Distributions of Econometric Estimators and Test Statistics, *in* Z. Griliches and M. D. Intriligator (editors), *Handbook of Econometrics, Vol. II*, 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. W. Watson, 2002, Macroeconomic Forecasting Using Diffusion Indices, *Journal of Business and Economic Statistics* **20**, 147–162.

Stock, J. H. and M. Yogo, 2003a, Asymptotic Distributions of Instrumental Variables Statistics with Many Weak Instruments, *in* D. W. K. Andrews and J. H. Stock (editors), *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* D. W. K. Andrews and J. H. Stock (editors), *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, Cambridge, U.K.: Cambridge University Press, pp. 80–108.

Wold, H., 1982, Soft Modeling. The Basic Design and Some Extensions, *in* K.-G. Jöreskog and H. Wold (editors), *Systems Under Indirect Observation, Vol. 2*, North-Holland, Amsterdam.

Appendix

Proof of Theorem 1

We have that

$$\hat{\beta}^{PC} = \left(X'\hat{F}_{PC}(\hat{F}'_{PC}\hat{F}_{PC})^{-1}\hat{F}'_{PC}X\right)^{-1} X'\hat{F}_{PC}(\hat{F}'_{PC}\hat{F}_{PC})^{-1}\hat{F}'_{PC}y,$$

and

$$\hat{\beta}^{PLS} = \left(X'\hat{F}_{PLS}(\hat{F}'_{PLS}\hat{F}_{PLS})^{-1}\hat{F}'_{PLS}X\right)^{-1} X'\hat{F}_{PLS}(\hat{F}'_{PLS}\hat{F}_{PLS})^{-1}\hat{F}'_{PLS}y.$$

Note that these estimators can be written as

$$\hat{\beta}^{PC} = \left(X'_{PC}X\right)^{-1} X'_{PC}y$$

and

$$\hat{\beta}^{PLS} = \left(X'_{PLS}X\right)^{-1} X'_{PLS}y,$$

respectively, where $X_{PC} = (x_{1,PC}, \ldots, x_{T,PC})'$ and $X_{PLS} = (x_{1,PLS}, \ldots, x_{T,PLS})'$ denote the matrix of fitted values for X using a PC or PLS regression respectively. For future use define $x_t = (\ x_{1,t}, \ldots, x_{k,t})'$ $x_{t,PC} = (\ x_{1,t,PC}, \ldots, x_{k,t,PC})'$ and $x_{t,PC} = (\ x_{1,t,PLS}, \ldots, x_{k,t,PLS})'$. We wish to prove that

$$\left\|\hat{\beta}^{PC} - \hat{\beta}^{PLS}\right\| = o_p(T^{-1/2})$$

We need to prove that

$$\left\|\frac{X'_{PLS}X}{T} - \frac{X'_{PC}X}{T}\right\| = o_p(T^{-1/2}) \tag{13}$$

and

$$\left\|\frac{X'_{PLS}y}{T} - \frac{X'_{PC}y}{T}\right\| = o_p(T^{-1/2}) \tag{14}$$

It is sufficient to show that

$$\left\|\frac{1}{T}\sum_{t=1}^{T} x_{j,t}x_{j,t,PC} - \frac{1}{T}\sum_{t=1}^{T} x_{j,t}x_{j,t,PLS}\right\| = o_p(T^{-1/2}) \tag{15}$$

for $j = 1, \ldots k$. Then, (13) immediately follows and (14) can be proven similarly. Let $\hat{\alpha}_{j,PC}$ and $\hat{\alpha}_{j,PLS}$ denote the implied estimated regression coefficients from the PC or PLS regression respectively. Then,

$$\left\|\frac{1}{T}\sum_{t=1}^{T} x_{j,t}x_{j,t,PC} - \frac{1}{T}\sum_{t=1}^{T} x_{j,t}x_{j,t,PLS}\right\| = \left\|\frac{1}{T}\sum_{t=1}^{T} x_{j,t}\left(\hat{\alpha}'_{j,PC} - \hat{\alpha}'_{j,PLS}\right)z_t\right\| \leq$$

$$\|\hat{\alpha}_{j,PC} - \hat{\alpha}_{j,PLS}\| \left\| \frac{1}{T} \sum_{t=1}^{T} x_{j,t} z_t \right\|$$

Bu, by assumption 5 the LLN for stationary, ergodic processes implies that

$$\left\| \frac{1}{T} \sum_{t=1}^{T} x_{j,t} z_t \right\| \xrightarrow{p} \|\sigma_{j,xz}\|$$

where $\sigma_{j,xz} = E\left(x_{j,t} z_t\right)$. But $\|\sigma_{j,xz}\| = O\left(N^{1/2}\right)$. Noting that, by theorem 1 of Groen and Kapetanios (2008),

$$\|\hat{\alpha}_{j,PC} - \hat{\alpha}_{j,PLS}\| = o_p\left((NT)^{-1/2}\right),$$

(15) follows immediately, proving the result of the Theorem.

Proof of Theorem 2

To prove this Theorem we need the following Lemma

Lemma 1 *Let assumptions 5 and 6 hold. Then,*

$$\frac{1}{T} \sum_{t=1}^{T} x_t' \tilde{x}_{t,PLS} \xrightarrow{p} \Psi$$

where $\tilde{x}_{t,PLS} = (\ \tilde{x}_{1,t,PLS}, \ldots, \tilde{x}_{k,t,PLS})'$, $x_{j,t,PLS} = z_t' \hat{a}_{j,PLS}$, $\hat{a}_{j,PLS} = \frac{1}{T} \sum_{t=1}^{T} x_{j,t} z_t$, *and* Ψ *is a positive definite matrix.*

To prove the required result, we assume without loss of generality that $E(z_t) = 0$. We need to show that

$$\frac{1}{T} \sum_{t=1}^{T} x_t' \tilde{x}_{t,PLS} \xrightarrow{p} \Psi$$

We first show that

$$\frac{1}{T} \sum_{t=1}^{T} x_{j,t} \tilde{x}_{j,t,PLS} \xrightarrow{p} c > 0$$

We note that

$$x_{j,t} = z_t' \pi_j + v_t$$

Further, note that

$$\pi_j = \Sigma_{zz}^{-1} \sigma_{j,xz},$$

where $\sigma_{j,xz} = E\left(x_{j,t} z_t\right)$. Then, it follows that

$$x_{j,t} \tilde{x}_{j,t,PLS} = \hat{a}_{PLS}' z_t z_t' \Sigma_{zz}^{-1} \sigma_{j,xz}$$

It suffices to show that

$$\hat{a}'_{PLS}\left(\frac{1}{T}\sum_{t=1}^{T} z_t z_t'\right)\Sigma_{zz}^{-1}\sigma_{j,xz} \xrightarrow{p} c > 0 \tag{16}$$

But, $\left(\frac{1}{T}\sum_{t=1}^{T} z_t z_t'\right)\Sigma_{zz}^{-1}$ is positive definite for all $T > N_0$ for some $N_0 > N$, and therefore (16) follows if

$$\left\|\frac{1}{T}\sum_{t=1}^{T} x_{j,t}z_t - \sigma_{j,xz}\right\| = o_p(1) \tag{17}$$

But, by assumption 5

$$\frac{1}{T}\sum_{t=1}^{T} x_{j,t}z_{i,t} - \sigma_{j,xz} = O_p\left(T^{-1/2}\right), \text{ uniformly over } i. \tag{18}$$

Hence, (16) follows if $N/T \to 0$. The result is proven if we note that, by assumption 6 $\Sigma_{xz} = E\left(z_t x_t\right)$ is full column rank and $p\lim_{T,N\to\infty} Z'Z = \Sigma_{zz}$ is nonsingular. This proves Lemma 1.

To prove Theorem 2, we have that

$$\hat{\beta}_{2SPLS} - \beta = \left(\frac{X'_{PLS}X}{T}\right)^{-1}\frac{X'_{PLS}u}{T}$$

We first examine $\frac{X'_{PLS}X}{T}$. We have

$$X'_{PLS}X = \hat{A}'_{PLS}Z'Z\Pi + A'_{PLS}Z'V$$

By Lemma 1,

$$\hat{A}'_{PLS}Z'Z\Pi/T \xrightarrow{p} \Psi.$$

We next examine $\hat{A}'_{PLS}Z'V/T$. By arguments similar to those used in the proof of Lemma 1, we have that

$$\left\|Z\hat{A}_{PLS} - Z\Sigma_{xz}\right\| = o_p(1)$$

But given that $x_{j,t}$ and v_t are independent and have finite variance, it follows that $z_t'\pi_j$ has finite variance. Since, further, $z_{i,t}$ follow uniformly a CLT, it follows that

$$\frac{\Sigma'_{xz}Z'V}{T} = o_p(1)$$

Thus, overall

$$\frac{X'_{PLS}X}{T} \xrightarrow{p} \Psi.$$

Next, we examine

$$\frac{X'_{PLS}u}{T} = \frac{\hat{A}'_{PLS}Z'u}{T}$$

Then, using (18) and similarly to above we have that

$$\left\|\frac{\hat{A}'_{PLS}Z'u}{T}-\frac{\Sigma'_{xz}Z'u}{T}\right\|=O_p\left(\left(\frac{N}{T}\right)^{-1/2}\right)$$

and

$$\left\|\frac{\Sigma'_{xz}Z'u}{T}\right\|=O_p\left(T^{-1/2}\right)$$

Overall, it then follows that

$$\left\|\hat{\beta}_{2SPLS}-\beta\right\|=O_p\left(\left(\frac{N}{T}\right)^{-1/2}\right)$$

proving the Theorem.

Table 1. Results for IV estimators in the case of Equation (11)

MSE Results

	c_2	0.5				1				0.5				1			
p	T/N	30	50	100	200	30	50	100	200	30	50	100	200	30	50	100	200
		PLS IV								Standard IV							
0	30	0.139	0.135	0.141	0.143	0.198	0.188	0.192	0.194	0.253	0.256	0.255	0.251	0.372	0.369	0.373	0.369
0	50	0.106	0.107	0.107	0.103	0.153	0.147	0.146	0.146	0.183	0.250	0.244	0.247	0.284	0.360	0.368	0.371
0	100	0.071	0.072	0.072	0.074	0.103	0.100	0.104	0.099	0.106	0.148	0.244	0.241	0.181	0.242	0.358	0.359
0	200	0.052	0.049	0.051	0.052	0.072	0.071	0.071	0.069	0.069	0.088	0.143	0.235	0.108	0.150	0.240	0.352
0.1	30	0.142	0.139	0.139	0.135	0.186	0.184	0.181	0.174	0.258	0.258	0.248	0.258	0.368	0.373	0.375	0.368
0.1	50	0.102	0.103	0.102	0.104	0.152	0.144	0.134	0.138	0.182	0.248	0.250	0.246	0.282	0.362	0.364	0.364
0.1	100	0.074	0.070	0.071	0.072	0.104	0.102	0.100	0.097	0.111	0.153	0.244	0.245	0.182	0.248	0.360	0.359
0.1	200	0.053	0.051	0.051	0.053	0.071	0.069	0.072	0.072	0.068	0.090	0.150	0.238	0.111	0.152	0.241	0.355
0.25	30	0.133	0.132	0.134	0.137	0.202	0.203	0.215	0.233	0.252	0.255	0.251	0.249	0.372	0.366	0.372	0.370
0.25	50	0.103	0.105	0.104	0.104	0.153	0.147	0.163	0.176	0.180	0.246	0.251	0.254	0.293	0.356	0.363	0.362
0.25	100	0.074	0.074	0.069	0.074	0.109	0.104	0.104	0.118	0.116	0.154	0.240	0.245	0.192	0.254	0.356	0.358
0.25	200	0.052	0.052	0.051	0.052	0.079	0.077	0.077	0.078	0.068	0.091	0.148	0.238	0.119	0.165	0.241	0.354
0.33	30	0.139	0.139	0.150	0.162	0.222	0.231	0.252	0.287	0.254	0.251	0.255	0.255	0.372	0.370	0.361	0.372
0.33	50	0.109	0.107	0.112	0.122	0.166	0.184	0.199	0.233	0.190	0.247	0.247	0.251	0.301	0.372	0.361	0.367
0.33	100	0.080	0.074	0.077	0.081	0.119	0.122	0.137	0.169	0.118	0.156	0.242	0.238	0.207	0.264	0.364	0.360
0.33	200	0.053	0.053	0.053	0.055	0.080	0.082	0.091	0.106	0.071	0.095	0.152	0.237	0.124	0.171	0.254	0.356
0.45	30	0.156	0.165	0.184	0.211	0.266	0.291	0.318	0.349	0.256	0.255	0.251	0.253	0.366	0.364	0.369	0.381
0.45	50	0.121	0.129	0.149	0.179	0.219	0.248	0.285	0.315	0.196	0.244	0.249	0.248	0.315	0.369	0.365	0.365
0.45	100	0.081	0.091	0.102	0.131	0.147	0.168	0.220	0.264	0.125	0.170	0.238	0.244	0.222	0.278	0.356	0.359
0.45	200	0.060	0.060	0.068	0.085	0.106	0.115	0.157	0.202	0.080	0.105	0.165	0.236	0.154	0.197	0.279	0.355
0.5	30	0.171	0.188	0.203	0.227	0.297	0.321	0.337	0.343	0.257	0.257	0.252	0.255	0.375	0.381	0.370	0.359
0.5	50	0.127	0.146	0.176	0.196	0.244	0.265	0.308	0.328	0.190	0.247	0.251	0.246	0.325	0.359	0.366	0.361
0.5	100	0.090	0.098	0.126	0.154	0.171	0.202	0.255	0.302	0.133	0.172	0.243	0.239	0.241	0.290	0.361	0.366
0.5	200	0.060	0.065	0.083	0.114	0.118	0.137	0.186	0.241	0.081	0.111	0.167	0.239	0.162	0.209	0.283	0.350

The table reports the Mean Squared Error for the standard and PLS-IV estimators. The Monte Carlo design is as in (9)-(11), with r=1, k=1.

Table 2. Results for IV estimators in the case of Equation (12)																	
MSE Results																	
	c_2	0.5				1				0.5				1			
p	T/N	30	50	100	200	30	50	100	200	30	50	100	200	30	50	100	200
		PLS IV								Standard IV							
0	30	0.205	0.214	0.210	0.207	0.186	0.191	0.195	0.202	0.377	0.369	0.368	0.378	0.369	0.368	0.369	0.372
0	50	0.155	0.142	0.153	0.152	0.147	0.145	0.139	0.151	0.283	0.365	0.360	0.362	0.273	0.360	0.359	0.363
0	100	0.095	0.105	0.101	0.106	0.101	0.098	0.097	0.105	0.176	0.240	0.361	0.359	0.175	0.241	0.357	0.360
0	200	0.069	0.071	0.069	0.072	0.070	0.072	0.071	0.072	0.108	0.149	0.237	0.353	0.104	0.147	0.236	0.357
0.1	30	0.258	0.261	0.268	0.272	0.241	0.251	0.254	0.271	0.492	0.504	0.516	0.551	0.479	0.489	0.527	0.539
0.1	50	0.215	0.209	0.213	0.230	0.175	0.199	0.199	0.207	0.398	0.498	0.509	0.540	0.390	0.488	0.510	0.534
0.1	100	0.143	0.148	0.156	0.167	0.137	0.145	0.154	0.156	0.268	0.368	0.509	0.529	0.269	0.366	0.506	0.536
0.1	200	0.102	0.103	0.115	0.120	0.099	0.098	0.109	0.116	0.167	0.249	0.397	0.534	0.164	0.250	0.391	0.523
0.25	30	0.364	0.428	0.502	0.579	0.389	0.455	0.553	0.612	0.611	0.633	0.644	0.667	0.596	0.620	0.655	0.672
0.25	50	0.271	0.318	0.403	0.520	0.283	0.367	0.475	0.559	0.544	0.630	0.644	0.670	0.533	0.625	0.636	0.663
0.25	100	0.203	0.234	0.281	0.394	0.203	0.259	0.349	0.473	0.433	0.548	0.641	0.661	0.427	0.552	0.643	0.670
0.25	200	0.160	0.172	0.198	0.278	0.149	0.176	0.242	0.350	0.317	0.454	0.581	0.670	0.304	0.440	0.584	0.658
0.33	30	0.477	0.558	0.632	0.664	0.493	0.579	0.650	0.683	0.647	0.678	0.687	0.687	0.626	0.667	0.688	0.696
0.33	50	0.376	0.466	0.586	0.629	0.413	0.514	0.613	0.657	0.606	0.647	0.682	0.676	0.593	0.653	0.684	0.686
0.33	100	0.255	0.348	0.477	0.586	0.288	0.406	0.535	0.631	0.508	0.621	0.657	0.677	0.495	0.604	0.666	0.694
0.33	200	0.180	0.233	0.367	0.507	0.198	0.279	0.446	0.567	0.394	0.537	0.636	0.677	0.376	0.519	0.644	0.686
0.45	30	0.605	0.663	0.705	0.696	0.590	0.664	0.686	0.710	0.678	0.692	0.716	0.698	0.651	0.688	0.691	0.713
0.45	50	0.560	0.625	0.683	0.712	0.552	0.629	0.671	0.689	0.661	0.681	0.702	0.715	0.636	0.680	0.682	0.691
0.45	100	0.444	0.556	0.653	0.703	0.459	0.580	0.658	0.701	0.601	0.654	0.697	0.717	0.580	0.662	0.694	0.710
0.45	200	0.319	0.458	0.602	0.676	0.335	0.486	0.619	0.674	0.523	0.615	0.673	0.702	0.474	0.601	0.674	0.693
0.5	30	0.642	0.677	0.703	0.704	0.646	0.671	0.708	0.718	0.687	0.695	0.707	0.704	0.691	0.685	0.710	0.719
0.5	50	0.592	0.658	0.697	0.702	0.584	0.658	0.701	0.690	0.663	0.694	0.704	0.703	0.641	0.689	0.709	0.692
0.5	100	0.503	0.607	0.676	0.705	0.498	0.620	0.686	0.685	0.621	0.670	0.698	0.710	0.591	0.673	0.706	0.690
0.5	200	0.393	0.545	0.639	0.690	0.389	0.532	0.642	0.688	0.552	0.649	0.681	0.701	0.496	0.608	0.673	0.698

The table reports the Mean Squared Error for the standard and PLS-IV estimators. The Monte Carlo design is as in (9)-(11), with r=1, k=1.

Table 3. Results for IV estimators using (12) with no factors.								
MSE Results								
Equal average weights for each z_{jt}								
T/N	30	50	100	200	30	50	100	200
	PLS IV				Standard IV			
30	0.275	0.287	0.273	0.264	0.348	0.350	0.345	0.346
50	0.212	0.230	0.224	0.237	0.265	0.336	0.324	0.333
100	0.154	0.169	0.199	0.209	0.173	0.226	0.325	0.326
200	0.098	0.121	0.154	0.177	0.104	0.146	0.225	0.313
Unequal weights for each z_{jt}								
30	0.297	0.271	0.290	0.270	0.375	0.377	0.382	0.372
50	0.220	0.225	0.240	0.247	0.277	0.359	0.359	0.366
100	0.155	0.181	0.196	0.217	0.183	0.244	0.351	0.357
200	0.099	0.122	0.156	0.188	0.106	0.147	0.241	0.358

The table reports the Mean Squared Error for the standard and PLS-IV estimators. The Monte Carlo design is as in (9)-(11), with k=1 and no factor structure ($c = 0$).