A Comparative Study of Three Data-Based Methods of Instrument Selection

Gunce Eryuruk

ITAM[1]

Alastair R. Hall

University of Manchester[2]

and

Kalidas Jana

University of Texas at Brownsville[3]

August 14, 2009

[1]CIE, Instituto Tecnológico Autónomo de México, Avenida Camino a Santa Teresa 930, 10700 México D.F., Mexico. E-mail: gunce.eryuruk@itam.mx

[2]Corresponding author: Economics, School of Social Sciences, University of Manchester, Manchester M13 9PL, UK. E-mail: alastair.hall@manchester.ac.uk

[3]Department of Business Administration, University of Texas at Brownsville, 80 Fort Brown, Brownsville, TX 78520, USA. E-mail: kalidas.jana@utb.edu

Abstract

We assess relative performance of three recently proposed instrument selection methods via a Monte Carlo study that investigates the finite sample behavior of the post-selection estimator of a simple linear IV model. Our results suggest that no one method dominates.

Key Words: Instrument Selection; Canonical Correlations Information Criterion; Relevant Moments Selection Criterion; Approximate Mean Square Error Criterion; Two-Stage Least Squares.

JEL Classification: C13, C30.

1   Introduction

This paper compares the properties of three recently proposed methods for instrument selection,namely the approximate Mean Square Error Criterion (AMSE) of Donald & Newey (2001), the Canonical Correlations Information Criterion (CCIC) of Hall & Peixe (2003), and the Relevant Moments Selection Criterion (RMSC) of Hall, Inoue, Jana & Shin (2007). While the three methods under study are tied by the common goal of instrument selection, they are different in terms of their underlying objectives. Donald and Newey's (2001) objective is to achieve an improved finite sample risk property of the estimators. They attain this goal by minimizing the approximate Mean Square Error Criterion based on higher-order asymptotics. The objective of Hall & Peixe (2003) and Hall, Inoue, Jana & Shin (2007), on the other hand, is to achieve an improved quality of asymptotic approximation to the finite sample behavior of the estimators. They gain this objective by eliminating the redundant moment conditions based on certain canonical correlations: CCIC exploits explicitly the canonical correlations (CCs) between the regressors and instruments; RMSC exploits implicitly the long run canonical correlations (LRCCs) between the unknown true score vector and the product of the instrument vector and error. While the properties of each of the methods have been explored by their proponents, there have been no comparative studies of these methods to date. Our paper intends to fill that gap.

2   Three methods of instrument selection

In this section, we describe the three methods of instrument selection mentioned in the Introduction in the context of the model used in our simulation study. Accordingly, consider the model

$$y_t = x_t \theta_0 + u_t,$$
$$\underset{(1\times1)}{y_t} = \underset{(1\times1)(1\times1)}{x_t \; \theta_0} + \underset{(1\times1)}{u_t},$$

$$\underset{(1\times q)(q\times1)}{x_t = z_t' \; \pi_0} + \underset{(1\times1)}{e_t}, \quad t = 1, 2, \dots, T$$

$$\text{(1)}$$

where

$$\left(\begin{array}{c} u_t \\ e_t \end{array} \middle| z_t \right) \overset{iid}{\sim} N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{ue} \\ \sigma_{ue} & \sigma_e^2 \end{bmatrix} \right). \tag{2}$$

We assume that the candidate set of moment conditions is given by

$$E[z_t u_t(\theta_0)] = 0 \tag{3}$$

where $u_t(\theta) = y_t - x_t\theta$ and z_t is the $q \times 1$ vector of instruments in (1). In this case, the only difference between various choices of moments lies in the chosen instrument vector and so we refer to z_t as the candidate set of instruments. We use a $q \times 1$ selection vector c in the notation of Andrews (1999) to denote which elements of the instrument vector z_t are included in a particular moment condition: if $c_j = 1$ then the jth element of z_t is included; if $c_j = 0$ then the jth element of z_t is excluded. The case in which all instruments are used is denoted by $c = \iota_q$ where ι_q is $q \times 1$ vector of ones. The moments associated with c are written as

$$E[z_t(c) u_t(\theta_0)] = 0 \tag{4}$$

where $z_t(c) = S(c) z_t$ and $S(c)$ is a selection matrix that picks out the elements of z_t indicated by c. Note that $|c| = c'c$ equals the number of elements in $z_t(c)$. The set of all possible selection vectors is denoted by $\mathcal{C}$, that is

$$\mathcal{C} = \left\{ c \in \Re^q; \; c_j = 0, 1, \text{ for } j = 1, 2, \dots q, \text{ and } c = (c_1, \dots c_q)', \; |c| \geq 1 \right\}.$$

For brevity, statistics of interest are now indexed by c and so, for example, $\hat{\theta}_T(c)$ denotes the 2SLS estimator based on (4).

Specializing the definitions to the model of this section, the three instrument selection methods are as follows.

CCIC: The selected instrument vector is $\hat{c}_{CCIC}$ where[1]

$$\hat{c}_{CCIC} \;=\; argmin_{c \in \mathcal{C}} \left\{ T \, ln[1 - R_T^2(c)] + (|c| - 1) \, ln[T] \right\} \tag{5}$$

where $R_T(c)$ is the sample multiple correlation coefficient between x_t and $z_t(c)$, and T is the sample size.

RMSC: The selected instrument vector is $\hat{c}_{RMSC}$ where[2]

$$\hat{c}_{RMSC} \;=\; argmin_{c \in \mathcal{C}} \, RMSC(c) \tag{6}$$

and

$$
\begin{aligned}
RMSC(c) \;=\; & ln \left[\hat{\sigma}_u^2(c) \left\{ T^{-1} \sum_{t=1}^{T} x_t z_t(c)' \{ T^{-1} \sum_{t=1}^{T} z_t(c) z_t(c)' \}^{-1} \right. \right. \\
& \left. \left. \times \; T^{-1} \sum_{t=1}^{T} z_t(c) x_t \right\}^{-1} \right] + \frac{(|c| - 1)}{\sqrt{T}} \, ln\sqrt{T}
\end{aligned} \tag{7}
$$

with $\hat{\sigma}_u^2(c) = T^{-1} \sum_{t=1}^{T} \{ u_t(\hat{\theta}_T(c)) \}^2$.

Approximate MSE criterion: The selected instrument vector is $\hat{c}_{AMSE}$ where

$$\hat{c}_{AMSE} \;=\; argmin_{c \in \mathcal{C}} \, AMSE(c) \tag{8}$$

and

$$AMSE(c) = \{ \hat{\sigma}_{au} \}^2 \frac{|c|^2}{T} + \tilde{\sigma}_u^2 \left(\hat{R}(c) - \hat{\sigma}^2 \frac{|c|}{T} \right) \tag{9}$$

---

[1]Note that we use the penalty term associated with BIC. Other choices are possible but Hall & Peixe (2003) found this choice to work best of those they considered; see Hall & Peixe (2003) for further discussion.

[2]Note that we again use the penalty term associated with BIC (Schwarz, (1978)). As with CCIC, other choices are possible but this choice appears to work best; see Hall, Inoue, Jana & Shin (2007) for further discussion.

where $\hat{\sigma}_{au} = T^{-1}\tilde{a}'\tilde{u}$, $\tilde{a} = \tilde{e}[\tilde{\pi}'Z(\iota_q)'Z(\iota_q)\tilde{\pi}/T]^{-1}$, $\tilde{e} = [I-P(\iota_q)]x$, $P(c) = Z(c)[Z(c)'Z(c)]^{-1}Z(c)'$, $Z(c)$ is a $T \times |c|$ matrix with t^{th} row $z_t(c)'$, $\tilde{\pi}$ is a preliminary estimator of π_0, $\tilde{u} = y - x\tilde{\theta}$ is a $T \times 1$ residual vector with t^{th} element $u_t(\tilde{\theta})$, $\tilde{\theta}$ is a preliminary estimator of θ_0, $\tilde{\sigma}_u^2 = T^{-1}\tilde{u}'\tilde{u}$, $\hat{\sigma}^2 = T^{-1}\tilde{a}'\tilde{a}$, and $\hat{R}(c)$ is a measure of the goodness of fit of the first stage reduced form model, $x_t = z_t(c)'\pi(c) + e_t(c)$. Donald & Newey (2001) consider two choices for $\hat{R}(c)$; for brevity, we focus on just one of these, the cross-validation criterion,

$$\hat{R}^{cv}(c) = \frac{1}{T}\sum_{i=1}^{T}\frac{[\hat{a}_i(c)]^2}{[1-P_{ii}(c)]^2}$$

where $\hat{a}_i(c)$ is the ith element of $\hat{a}(c) = \hat{e}(c)[\tilde{\pi}'Z(\iota_q)'Z(\iota_q)\tilde{\pi}/T]^{-1}$ and $\hat{e}(c) = [I_T - P(c)]x$.

3 Simulation design and results

We now describe the simulation design. Data are generated from (1)-(3) with $p = 1$, $\theta_0 = 0.1$, different specifications of π_0 described below, $[u_t, e_t, z_t']' \sim N(0, \Sigma)$, Σ is a matrix whose diagonal elements are all one and whose only non-zero off diagonal elements are $cov(u_t, e_t) = \sigma_{ue} \in \{0.1, 0.5, 0.9\}$, and sample size $T \in \{100, 500\}$. We consider three specifications for π_0 in the equation for x_t:

- *Model 1: Declining coefficients*

$$\pi_{0,i}^{(1)} = k(q)\left(1 - \frac{i}{q+1}\right)^4 \quad \text{for } i = 1, \dots, q, \tag{10}$$

  where the constant $k(q)$ is chosen so that the squared population multiple correlation between x_t and z_t, R_1^2 say, is 0.5.

- *Model 2: Equal coefficients*

$$\pi_{0,i}^{(2)} = \pi_0^{(2)} = \sqrt{\frac{0.5^2}{q(1-0.5^2)}} \quad \text{for } i = 1, \dots, q, \tag{11}$$

  which implies $R_1^2 = 0.5$.

4

- *Model 3: Two non-zero coefficients and the rest zero*

$$\pi_{0,i} = k(q)\left(1 - \frac{i}{q+1}\right)^4, \ i = 1, 2;$$
$$= 0 \quad \text{for} \ i = 3, \ldots, q,$$

where the constant $k(q)$ is chosen so that (12) holds for $R_1^2 = 0.5$. Note that $(z_{t,3}, z_{t,4}, \ldots, z_{t,q})$ are redundant given $(z_{t,1}, z_{t,2})$ and so $(z_{t,1}, z_{t,2})$ are referred to as the "relevant" instruments in this model.

There are two natural approaches to instrument selection: a sequential strategy in which only q possible choices are considered in a sequence such as $(z1)$, $(z1, z2)$, $(z1, z2, z3)$ *etc.*, and a strategy in which all possible combinations are considered.

In their investigation of $AMSE(c)$, Donald & Newey (2001) use Models 1 and 2 with $q \in \{20, 30\}$ and employ a sequential strategy in which the sequence of instruments considered is given by $c_1, c_2, \ldots, c_q$ where $c_i = [\iota_i', 0_{q-i}']'$ and ι_i is a $i \times 1$ vector of ones and 0_{q-i} is a $(q-i) \times 1$ vector of zeros, $i = 1, 2, \ldots, q$. In our simulations, we generate data from both Models 1 and 2, and select instruments using this sequential strategy based on all three instrument selection criteria. However, for brevity we only report results for $q = 20$; the results for $q = 30$ are qualitatively the same as for $q = 20$ and are available from the authors upon request. In their evaluation of $CCIC(c)$, Hall & Peixe (2003) consider Model 3 with $q = 8$ and selection is based on a comparison of all possible combinations of all instruments, that is $\{c \in \mathcal{C}\}$; note that there are 255 possible combinations within Model 3.[3] We compare the performance of all three criteria within this design.

To assess the relative performance of the CCIC, the RMSC and the AMSE, we focus on the finite sample performance of the post-selection 2SLS estimator. To assess the adequacy of standard asymptotic theory, we report the median bias and the coverage probability of 90% confidence intervals based on conventional asymptotic theory. However, due to potential concerns about non-existence of moments, we also report the median absolute deviation

---

[3]Hall, Inoue, Jana & Shin (2007) also use this set up in their evaluation of $RMSC(c)$.

and the interdecile range. To compute AMSE, in Models 1 and 2, we obtain the preliminary estimators $\tilde{\pi}$ and $\tilde{\theta}$ using the number of instruments that minimized the first stage cross-validation criteria. In Model 3 we obtain $\tilde{\pi}$ and $\tilde{\theta}$ using all eight instruments. The results are presented in Tables 1 through 3. All results correspond to 10,000 replications.

We begin with Model 1. Following Donald & Newey (2001), we consider the sequential strategy for instrument selection described above for this model. Notice that the selection strategy involves including the instruments in the order of their explanatory power for x_t, and so the order of selection is from "best to worst". This scenario corresponds to the case in which a practitioner has *a priori* information about the relative importance of the instruments. The results for this case are shown in Table 1. It can be seen from the table that for both sample sizes, $T = 100$ and $T = 500$, and for all three values of endogeneity considered, the median bias and coverage probabilities under AMSE, CCIC and RMSC are similar and thus no unique ranking among the three criteria emerges. In terms of median absolute deviation and interdecile range, AMSE and CCIC are similar but both dominate RMSC. It can also be seen from Table 1 that the median bias increases and the coverage probability distortions worsen with the increase in endogeneity under all three criteria.

Table 2 contains the results for Model 2. Here again, we use the sequential selection strategy. However, this time, all the instruments have equal explanatory power for x_t and so there are no "best" or "worst" instruments *per se*. It can be seen that for all configurations considered the median biases under AMSE and CCIC are similar and smaller than the median bias under RMSC. Interestingly, the biases under each criterion tend to be larger under Model 2 than under Model 1. The coverage probabilities also exhibit greater distortions in this model than under Model 1. No one criterion yields coverage probabilities that are closer to the nominal level than the others over all configurations. As in Model 1, the performance of the post-selection estimator deteriorates as the degree of endogeneity increases. In terms of both median absolute deviation and interdecile range, AMSE dominates for $\sigma_{ue} \in \{0.1, 0.5\}$ but no criterion dominates for $\sigma_{ue} = 0.9$.

Table 3 gives the results for Model 3. Recall that in this case only two of the eight instruments can explain x_t and selection is over all possible combinations. Following Hall and Peixe (2003) and Hall, Inoue, Jana & Shin (2007), we consider selection over all possible combinations of instruments for this case. The results reveal that under each parameter configuration the median biases and coverage probabilities are similar for each criterion, and thus no unique ranking emerges. In fact, in all configurations, all three methods perform very well and the empirical coverage rate under all three methods is almost equal to the nominal rate of 90%. All three methods also lead to post-selection estimators with comparable median absolute deviations and interdecile ranges.

Taking all our simulation results together, it seems to us that no one instrument selection criterion clearly dominates the others in terms of median bias, the closeness of the coverage probability to the nominal level, and dispersion. In some designs, we find that *all* three criteria lead to post-selection estimators with small bias and coverage probabilities close to the nominal level; but in other designs, we find that *all* three criteria lead to post-selection estimators whose finite sample behaviour is poorly approximated by conventional asymptotic theory.

References

Andrews, D. W. K. (1999). Consistent moment selection procedures for generalized method of moments estimation. *Econometrica, 67*, 543–564.

Donald, S. G. & Newey, W. K. (2001). Choosing the number of instruments. *Econometrica, 69*, 1161–1192.

Hall, A. R., Inoue, A., Jana, K., & Shin, C. (2007). Information in generalized method of moments estimation and entropy based moment selection. *Journal of Econometrics, 138*, 488–512.

Hall, A. R. & Peixe, F. P. M. (2003). A consistent method for the selection of relevant instruments. *Econometric Reviews*, *22*(3), 269–287.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464.

Table 1: Properties of *AMSE(c)*, *CCIC(c)* & *RMSC(c)*

Model 1: sequential selection strategy

$q_{max} = 20$	T=100			T=500		

$R_1^2 = 0.5$; $\sigma_{ue} = 0.1$.

	AMSE	CCIC	RMSC	AMSE	CCIC	RMSC
Med Bias	0.010	0.006	0.008	0.003	0.002	0.002
Cov Rate	0.886	0.906	0.914	0.898	0.904	0.905
Interdecile Range	0.251	0.257	0.291	0.114	0.114	0.127
Med AD	0.066	0.068	0.077	0.030	0.030	0.033

$R_1^2 = 0.5$; $\sigma_{ue} = 0.5$.

	AMSE	CCIC	RMSC	AMSE	CCIC	RMSC
Med Bias	0.036	0.035	0.037	0.009	0.008	0.010
Cov Rate	0.850	0.861	0.875	0.892	0.894	0.894
Interdecile Range	0.261	0.256	0.288	0.113	0.114	0.124
Med AD	0.068	0.067	0.075	0.030	0.031	0.033

$R_1^2 = 0.5$; $\sigma_{ue} = 0.9$.

	AMSE	CCIC	RMSC	AMSE	CCIC	RMSC
Med Bias	0.053	0.060	0.059	0.014	0.016	0.016
Cov Rate	0.818	0.800	0.832	0.873	0.871	0.873
Interdecile Range	0.252	0.244	0.273	0.114	0.112	0.125
Med AD	0.066	0.063	0.072	0.030	0.029	0.032

Notes: Med Bias = median bias, *Cov Rate* = the empirical coverage of a nominal 90% confidence interval, *Med AD* = median absolute deviation.

Table 2: Properties of *AMSE(c)*, *CCIC(c)* & *RMSC(c)*

Model 2: sequential selection strategy

$q_{max} = 20$	T=100			T=500		
$R_1^2 = 0.5;\ \sigma_{ue} = 0.1.$						
	AMSE	CCIC	RMSC	AMSE	CCIC	RMSC
Med Bias	0.013	0.022	0.021	0.004	0.003	0.008
Cov Rate	0.891	0.918	0.939	0.897	0.895	0.909
Interdecile Range	0.240	0.418	0.403	0.113	0.113	0.178
Med AD	0.062	0.088	0.102	0.029	0.030	0.046
$R_1^2 = 0.5;\ \sigma_{ue} = 0.5.$						
	AMSE	CCIC	RMSC	AMSE	CCIC	RMSC
Med Bias	0.094	0.096	0.108	0.018	0.018	0.038
Cov Rate	0.728	0.771	0.841	0.865	0.863	0.858
Interdecile Range	0.253	0.409	0.392	0.112	0.111	0.172
Med AD	0.063	0.082	0.100	0.029	0.029	0.044
$R_1^2 = 0.5;\ \sigma_{ue} = 0.9.$						
	AMSE	CCIC	RMSC	AMSE	CCIC	RMSC
Med Bias	0.166	0.172	0.185	0.046	0.033	0.064
Cov Rate	0.634	0.479	0.637	0.726	0.782	0.749
Interdecile Range	0.417	0.376	0.358	0.110	0.108	0.170
Med AD	0.094	0.069	0.092	0.028	0.029	0.044

Notes: See Table 1.

Table 3: Properties of $AMSE(c)$, $CCIC(c)$ & $RMSC(c)$

Model 3: selection over all possible combinations

$q_{max} = 8$	T=100			T=500		

$R_1^2 = 0.5;\ \sigma_{ue} = 0.1.$

	AMSE	CCIC	RMSC	AMSE	CCIC	RMSC
Med Bias	0.008	0.004	0.004	0.002	0.000	0.001
Cov Rate	0.892	0.897	0.900	0.899	0.895	0.897
Interdecile Range	0.257	0.259	0.262	0.114	0.116	0.116
Med AD	0.066	0.068	0.068	0.030	0.031	0.030

$R_1^2 = 0.5;\ \sigma_{ue} = 0.5.$

	AMSE	CCIC	RMSC	AMSE	CCIC	RMSC
Med Bias	0.020	0.012	0.017	0.004	0.002	0.001
Cov Rate	0.874	0.895	0.890	0.895	0.904	0.902
Interdecile Range	0.256	0.255	0.261	0.116	0.114	0.113
Med AD	0.067	0.067	0.067	0.030	0.030	0.030

$R_1^2 = 0.5;\ \sigma_{ue} = 0.9.$

	AMSE	CCIC	RMSC	AMSE	CCIC	RMSC
Med Bias	0.022	0.020	0.028	0.003	0.003	0.003
Cov Rate	0.868	0.867	0.869	0.894	0.896	0.901
Interdecile Range	0.261	0.256	0.265	0.117	0.115	0.115
Med AD	0.068	0.067	0.068	0.030	0.031	0.031

Notes: See Table 1.