GMM-based Model Averaging*

Luis F. Martins
Department of Quantitative Methods, ISCTE-LUI, Portugal
Centre for International Macroeconomic Studies (CIMS), UK
(luis.martins@iscte.pt)

Vasco J. Gabriel
CIMS, University of Surrey, UK and NIPE-UM
(v.gabriel@surrey.ac.uk)

This version: March 2012
(very preliminary, comments welcome)

Abstract

This paper considers moment conditions model averaging estimators in the Generalized Method of Moments (GMM) framework. We propose using moment selection criteria to select weights for averaging across GMM estimates. This can be achieved by direct smoothing of selection criteria arising from the estimation stage, or by numerical minimization of a specific criterion, as in Hansen (2007). We derive some asymptotic properties assuming correctly specified models. Monte Carlo experiments show that our procedure compares favourably with existing alternatives in many relevant setups.

Keywords: Generalized Method of Moments; Model Selection; Model Averaging.

JEL Classification: C22; C52; E43; E52

*This study was initiated while the first author was visiting the Economics Research Department (DEE) at the Bank of Portugal (BP) for a short-term fellowship. The author acknowledges financial support from the BP and thanks its members for useful comments. The analysis, opinions and findings of this paper represent the views of the authors, they are not necessarily those of the BP or the Eurosystem.

1 Introduction

In this paper, we consider model averaging (MA) estimation methods for unconditional moment-based models. The model averaging estimator is a weighted average of estimates obtained using different sets of moment conditions. We propose selecting empirical weights based on GMM moment selection criteria. This can be achieved by direct smoothing of information criteria arising from the estimation stage, or by numerical minimization of a specific criterion, as in Hansen (2007). We show that, under correct model specification, the MA estimator is consistent and normally distributed.

In many applications of instrumental variables and generalized method of moments (GMM) estimation, there is often a large set of candidate variables that can be used as instruments. However, the properties of moment conditions and instrumental variables estimators are very sensitive to the choice (and the characteristics) of the instrument set. Indeed, instruments might be poorly correlated with the endogenous variables, which invalidates conventional inference procedures (Staiger and Stock, 1997 and Stock and Wright, 2000). On the other hand, using many (potentially weak) instruments, while desirable (see Hansen, Hausman and Newey, 2008), may lead to biases and substantial deviations from the usual Gaussian asymptotic approximation (see Chao and Swanson, 2005, Han and Phillips, 2006 and Newey and Windmeijer, 2009).

Thus, much of the literature has focused on procedures for the selection of the appropriate moments/instruments.[1] Given that the rejection of the J-statistic is an indicator that some moment conditions are invalid, Andrews (1999) developed GMM analogues of model selection criteria based on this statistic in order to consistently select the largest set of valid moment conditions. This was extended to the case of jointly picking the moments and the parameter vector by Andrews and Lu (2001) and Hong, Preston and Shum (2003) for GEL estimators. As an alternative, Hall, Inoue, Jana and Shin (2007) suggest selecting moment conditions according to the relevant moment selection criterion (RMSC), based on the entropy of the limiting distribution of the GMM estimator. On the other hand, Donald and Newey (2001) propose a selection procedure such that an approximate mean-square error is minimized over all existing instruments deemed to be valid, while Hall and Peixe (2003) propose a canonical correlations information criteria (CCIC) for instrument selection.

Model selection entails choosing one of the estimated competing models under consideration, possibly by setting some of the moment conditions equal to zero. Testing competing, non-nested formulations, in which the outcome may not be the selection of one particular model, can be carried out in a moment conditions framework using the tests of Smith (1992) and Smith and

[1] Though our paper focuses on unconditional moment models, we should note that there is an extensive literature on the choice of optimal instruments in a conditional moments setting, see Newey (1993) and, more recently, Donald, Imbens and Newey (2009), for example.

Ramalho (2002).

Here, we pursue the alternative approach of model averaging, in which parameter estimates are constructed based on a weighted average of estimates from a number of possible specifications. By making use of the information conveyed by otherwise discarded alternative specifications, model averaging as an estimation strategy may yield significant gains in terms of bias and efficiency when compared to procedures that make use of a single set of moment conditions.

Our work is a natural extension of the literature, in which model averaging usually involve weights obtained from functions of model selection criteria, such as the BIC, AIC, etc. Indeed, there is a large literature on model averaging, both in the Bayesian tradition and in a frequentist framework (see Claeskens and Hjort, 2008 for a review). Recently, Hansen (2007) proposed a Mallows criterion for the selection of weights for averaging across least squares estimates obtained from a set of approximating models, in which regressors (or groups of regressors) are added sequentially. Kuersteiner and Okui (2010) suggest using Hansen's (2007) method as a first step to construct optimal instruments IV estimation with 2SLS, LIML and Fuller estimators. The weights are chosen to minimize the approximate mean squared error (MSE), as in Donald and Newey (2001).

Shrinkage methods for moment conditions estimators are an alternative to model selection. Caner (2009) proposes a LASSO-type GMM estimator, while Canay (2010) and Okui (2011) propose shrinkage-type estimators for linear models with many instruments, in which kernel weights are used to shrink the first-stage coefficient estimators toward zero. Note that shrinkage estimators can be viewed as a special case of model averaging, in which some moment conditions/instruments receive weights approaching zero. In general, two-stage shrinking reduces bias in the first-stage, at the cost of reduced efficiency in the second. Furthermore, kernel methods make the weighting scheme somewhat inflexible once a particular kernel is chosen, as pointed out by Kuersteiner and Okui (2010).

Although model averaging in the linear IV context has seen some recent developments, model averaging in general GMM estimation is still an open issue. We intend to fill this gap in the literature. Contrary to the above-mentioned methods, we do not confine the analysis to linear IV models, but consider a general GMM setup. Moreover, our approach is not two-stepped and, in our case, the list of candidate models does not depend on ordered instruments from the full-instrument matrix. Indeed, with m instruments we can consider m models (each model including an extra instrument as in Hansen, 2007), but also any possible combination of these. This makes our approach more general and not restricted by how the instruments are ordered. Also, unlike 2SLS kernel-based weighting, our procedure does not depend on the choice of kernels or arbitrarily user-chosen smoothing parameters.

Thus, we are able to combine the estimation of general moment conditions models with one-step and information criteria-based model averaging estimation. Furthermore, Monte Carlo experiments show that, in several setups, our model averaging estimation procedure outperforms the methods of Donald and Newey (2001) and Kuersteiner and Okui (2010) in terms of median bias and absolute deviation, namely in models with possibly weak instruments.

Next, section 2 introduces assumptions and definitions. In section 3, we introduce our moment conditions model averaging approach. In section 4, we derive the asymptotic properties of the GMM model averaging estimator. Section 5 presents a Monte Carlo simulation study providing evidence in support of our MA procedures and Section 6 concludes.

2 Definitions

Assume that the data $\{y_t\}$ is an infinite sequence of stationary and ergodic (possibly weakly dependent) variables such that the standard assumptions in the GMM context hold. The estimation of the unique p-dimensional parameter vector $\theta_0 = (\theta_{0,1}, ..., \theta_{0,p}) \in \Theta \subset \Re^p$ is based on $m \geq p$ moment conditions of the form $E[g(y_t, \theta_0)] \equiv E[g_t(\theta_0)] = 0$, for all t, with corresponding empirical moments $\widehat{g}_T(\theta) = (1/T)\sum_{t=1}^{T} g(y_t, \theta)$. The GMM estimator is defined as

$$\widehat{\theta}_T(W) = \arg\min_{\theta\in\Theta} \widehat{g}_T(\theta)' W_T \widehat{g}_T(\theta), \tag{1}$$

where W_T is a weighting matrix such that $\operatorname{plim} W_T = S^{-1}$, where

$$S = \lim_{T\to\infty} Var\left[T^{-1/2}\sum_{t=1}^{T} g(y_t, \theta_0)\right]. \tag{2}$$

is the $m \times m$ long-run variance matrix of the process $\widehat{g}_T(\theta)$. Under some regularity conditions, $\widehat{\theta}_T(W)$ is consistent and asymptotically normal, with asymptotic variance

$$V = \left(G' S^{-1} G\right)^{-1}, \tag{3}$$

where G is a $m \times p$, full-column ranked, Jacobian matrix defined as

$$G = E\left(\left.\frac{\partial g(y_t, \theta)}{\partial \theta'}\right|_{\theta=\theta_0}\right). \tag{4}$$

Also, the J test statistic for over-identifying restrictions is

$$J_T = T \inf_{\theta\in\Theta} \widehat{g}_T(\theta)' W_T \widehat{g}_T(\theta). \tag{5}$$

Given that the rejection of the J-statistic is an indicator that some moment conditions are invalid, this can then be used to consistently select the correct moment conditions. Let $\mathcal{M}$ be the collection of candidate moment conditions models. Here, $\mathcal{M}$ is a countable/finite or an uncountable set, such that model M_i belongs to the family of models $\mathcal{M} : M_i \in \mathcal{M}$. The "true" model may or may not be an element of $\mathcal{M}$. In our model averaging procedure, we specify a subset of $\mathcal{M}$ from which we define the MA estimator. For now, take any particular moment conditions model, M_i, which is characterized by a particular set of instruments.

Following Andrews (1999), the GMM moment selection criteria for a given model is defined as

$$MSC_T(c) = J_T(c) - \kappa_T(|c| - p), \tag{6}$$

where $c \in \Re^m$ is a moment selection vector that represents a list of "selected" moment conditions (a subset of g), $|c|$ denotes the number of the selected moments c (here $|c| \leq m$), $J_T(c)$ is the J-statistic computed using the selected moments $c, |c| - p$ is the number of over-identifying restrictions and $\kappa_T = o(T)$ is a sequence that defines the selection criterion ($\kappa_T = 2$ for the AIC; $\kappa_T = \log T$ for the BIC; and $\kappa_T = Q \log\log T$ for some $Q > 2$ for the HQ-type criterion). Defining the unit-simplex set

$$C = \left\{c \in \Re^m \backslash \{0\} : c_j = 0 \text{ or } 1, \forall 1 \leq j \leq m, \text{ where } c = (c_1, ..., c_m)'\right\}, \tag{7}$$

c is a vector of zeros (excluded conditions) and ones (included conditions) and $|c| = \sum_j^m c_j$ for $c \in C$. The moment selection criteria estimator is defined as

$$\widehat{c}_{msc} = \arg\min_{c\in\mathcal{C}} MSC_T(c) = \arg\min_{c\in\mathcal{C}} \left(J_T(c) - \kappa_T(|c| - p)\right), \tag{8}$$

where $\mathcal{C} \subset C$, with $\{0\} \in \mathcal{C}$, is some parameter space for the moment selection vector. The estimator $\widehat{c}_{msc}$ picks the moment conditions c over $\mathcal{C}$ such that the increase in $J_T(c)$ that typically occurs when moment conditions are added (even if correct) is offset by the "bonus term" $\kappa_T(|c| - p)$ that rewards selection vectors that utilize more moment conditions.

Under relatively standard assumptions, Andrews (1999) shows that $\widehat{c}$ is a consistent estimator of c_0, assumed to be the single "correct" selection vector.[2] If, additionally, one assumes that $E(g_{c_0}(\theta)) = 0$ has a unique solution $\theta_0 \in \Theta$ (the "true" value of θ, set at c_0), then $\widehat{c}$ consistently estimates both c_0 and θ_0.

While the criteria above stress the satisfaction of *orthogonality* conditions, other procedures have been proposed in which the focus is on the *relevance* of moment conditions. Under somewhat more restrictive assumptions, Hall et al. (2007) suggest selecting a model according to the relevant moment selection criterion

$$RMSC_T(c) = \ln\left(\left|\widehat{V}_c\right|\right) + \kappa_T(|c| - p), \tag{9}$$

where the efficient GMM variance-covariance matrix $\widehat{V}_c$ is evaluated at $\widehat{\theta}_{Tc}$. On the other hand, Hall and Peixe (2003) consider the problem of instrument selection based on a combination of the efficiency and non-redundancy conditions

$$CCIC_T(c) = T\sum_{i=1}^{p} \ln\left[1 - r_{i,T}^2(c)\right] + \kappa_T(|c| - p), \tag{10}$$

[2]The GMM-AIC is not consistent and it has positive probability (even asymptotically) of selecting too few moments.

where $r_{i,T}(c)$ is the i^{th} sample canonical correlation between $d_t\left(\widetilde{\theta}_T\right)$ and $z_t(c)$, with $d_t(\theta) = \frac{\partial u_t(\theta)}{\partial \theta}$ and $\widetilde{\theta}_T$ is a $\sqrt{T}-$ consistent preliminary estimator. Here, $u_t(\theta)$ is scalar and, if the model is linear, $d_t(\theta) = -x_t$.

3 Moment Conditions Model Averaging Estimators

In this section, we present a methodology whereby we average across candidate specifications to obtain an averaged estimator. Note that this differs from previous literature (namely Kuersteiner and Okui, 2010, Kapetanios and Marcellno, 2010 and Okui, 2011) in that we are not averaging across instruments to obtain an optimal set of instruments. Instead, we propose averaging different estimates of θ_0 obtained from distinct sets of moment conditions. The weights associated with each estimate are chosen according to GMM information criteria.

3.1 The Procedure

Consider m and c as defined in (7). Quantities such as $\widehat{g}_{Tc}(\theta), W_{Tc}$ and $\widehat{\theta}_{Tc}$ are obtained after deleting the moments j corresponding to $c_j = 0$. Here, $\widehat{g}_{Tc}(\theta)$ is a $|c| \times 1$ vector. Now, let $\omega = \left(\omega_1, ..., \omega_{|C|}\right)'$ be a weight vector in the unit-simplex in $\Re^{|C|}$, where $|C| = 2^m - \sum_{j=0}^{p-1}\binom{m}{j} = \sum_{j=p}^{m}\binom{m}{j}$ with the binomial coefficients $\binom{m}{j} = \frac{m!}{j!(m-j)!}$ representing the number of different elements[3] in C :

$$H_m = \{\omega \in [0,1]^{|C|} : \sum_{c \in C} \omega_c = 1\}. \tag{11}$$

Thus, a model averaging estimator of the unknown $p \times 1$ vector θ_0 is

$$\widehat{\theta}_T(\omega) = \sum_{c \in C} \omega_c \widehat{\theta}_{Tc}. \tag{12}$$

Clearly, the standard GMM estimation is a special case for which no model averaging occurs: $\omega_{c^*} = 1$ for some c^* and $\omega_{c'} = 0$ for $c' \neq c^*$ and $\widehat{\theta}_T(\omega) = \widehat{\theta}_{Tc^*}$.

In general, the optimal vector ω will be unknown. As in much of the literature on model averaging, a data-dependent procedure will have to be used to determine the weights in order to implement estimation according to (12). Thus, we suggest linking the problem of selecting weights with model selection criteria obtained in the estimation stage. This can be achieved either by direct 'smoothing' or by minimization of a consistent MSC.

3.2 Smooth Moment Selection Criteria Weights

As suggested by Buckland, Burnham and Augustin (1997) (see also Burnham and Anderson, 2002), a simple averaging scheme can be obtained by using weights proportional to the expo-

[3] We need to exclude $\sum_{j=0}^{p-1}\binom{m}{j}$ from the total of combinations 2^m, those for which $m < p$.

nential form of a given MSC. Thus, a smooth AIC, BIC, etc. scheme is based on weights for candidate model M,

$$\widehat{\omega}_M(MSC) = \frac{\exp(\frac{1}{2}MSC_M)}{\sum_{M'\in\mathcal{M}}\exp(\frac{1}{2}MSC_{M'\in\mathcal{M}})} \tag{13}$$

where the sum term encompasses all, not necessarily nested, $M' \in \mathcal{M}$ models of interest.[4] Note that in our case, given that the AIC is not consistent, we will only consider smooth BIC (denoted as $\widehat{\omega}_{S\text{-}BIC}$) and smooth Hannan-Quinn ($\widehat{\omega}_{S\text{-}HQ}$) weigths. Other simplified weighting schemes have been explored in the literature and can potentially be employed, see Claeskens and Hjort (2008).

3.3 Selecting Weights by Minimizing GMM Moment Selection Criteria

In the spirit of Hansen (2007), we propose obtaining the weight vector ω by numerical minimization of consistent GMM moment selection criteria evaluated at $\widehat{\theta}_T(\omega)$. Using Andrews's (1999) MSC, the empirical selected weight vector is defined as

$$\widehat{\omega}_{MSC} = \arg\min_{\omega\in H_m} MSC_{T\overline{c}}(\omega) = \arg\min_{\omega\in H_m}\left(J_{T\overline{c}}(\omega) - \kappa_T(|\overline{c}| - p)\right), \tag{14}$$

where $J_{T\overline{c}}(\omega) = T\widehat{g}_{T\overline{c}}\left(\widehat{\theta}_T(\omega)\right)' W_{T\overline{c}}\widehat{g}_{T\overline{c}}\left(\widehat{\theta}_T(\omega)\right)$, for a given set of moment conditions $\overline{c}$ and given $W_{T\overline{c}}$.

Alternatively, weight selection could be pursued through

$$\begin{aligned}\widehat{\omega}_{RMSC} &= \arg\min_{\omega\in H_m} RMSC_{T\overline{c}}(\omega) = \arg\min_{\omega\in H_m}\left[\ln\left(\omega' diag\left(\left|\widehat{V}_1\right|, ..., \left|\widehat{V}_{|C|}\right|\right)\omega\right)\right] \qquad (15)\\ &= \arg\min_{\omega\in H_m}\left[\omega' diag\left(\left|\widehat{V}_1\right|, ..., \left|\widehat{V}_{|C|}\right|\right)\omega\right]\end{aligned}$$

or

$$\widehat{\omega}_{CCIC} = \arg\min_{\omega\in H_m} CCIC_{Tc}(\omega) \tag{16}$$

$$= \arg\min_{\omega\in H_m}\left[\omega' diag\left(\sum_{i=1}^{p}\ln\left[1 - r^2_{i,T,1}\right], ..., \sum_{i=1}^{p}\ln\left[1 - r^2_{i,T,|C|}\right]\right)\omega\right], \tag{17}$$

where $diag(\cdot)$ refers to a $|C| \times |C|$ diagonal matrix.

Remark 1. As in Hansen (2007), the solution $\widehat{\omega}$ is found by numerical algorithms. It solves a constrained optimization problem with nonnegativity and summation constraints ($\omega_c \in [0, 1]$, for all c and $\sum_{c\in C}\omega_c = 1$, respectively). The (asymptotic) distribution of $\widehat{\omega}$ is beyond the scope of this paper (indeed, this is not even know for the case of least squares MA estimators, see Hansen, 2007).

[4] For numerical stability, it is sometimes recommended that the maximum MSC_T value is subtracted to the each $MSC(M)$.

Remark 2. Note that, although averaging occurs over specifications using different combinations of moment conditions, the minimization of GMM selection criteria in (14) depends on the J-statistic. This, in turn, requires the weight matrix to be chosen and, therefore, a set of moment conditions $\bar{c}$ to be fixed. Moreover, and unlike the least squares MA estimator of Hansen (2007) and the two-step MA instruments estimators of Kuersteiner and Okui (2010), which have distinct number of parameters to estimate for each individual model, in our case $p_c = p$ for all c. Hence,

$$\min_{\omega \in H_m} MSC_{T\bar{c}}(\omega) = \min_{\omega \in H_m} J_{T\bar{c}}(\omega) \tag{18}$$

for any penalty term κ_T. Thus, an MA estimator that minimizes a GMM selection criterion will be solely based on the $J_T(\omega)$-statistic. For sake of efficiency, one can pick $\bar{c} = \iota_m$, a vector of ones, which implies using the whole set of moment conditions (in this case, $|\bar{c}| = m$ and, in terms of notation, "c" is dropped):

$$J_T(\omega) = T\widehat{g}_T\left(\widehat{\theta}_T(\omega)\right)' W_T\widehat{g}_T\left(\widehat{\theta}_T(\omega)\right). \tag{19}$$

For the linear IV/2SLS case, for a set of variables x_t and instruments z_t, such that $y_t = (x_t', z_t')'$, then

$$J_{T\bar{c}}(\omega) = T\left(\frac{1}{T}\sum_{t=1}^{T} z_{\bar{c},t}\left(y_t - x_t'\sum_{c\in C}\omega_c\widehat{\theta}_{Tc}\right)\right)' W_{T\bar{c}}\left(\frac{1}{T}\sum_{t=1}^{T} z_{\bar{c},t}\left(y_t - x_t'\sum_{c\in C}\omega_c\widehat{\theta}_{Tc}\right)\right). \tag{20}$$

Remark 3. Given the consistency of the GMM MSC, our averaging procedure can be applied even if incorrect moment conditions are used, as discussed below. Andrews (1999) proposes "downward" and "upward" testing procedures that can consistently pick the "correct" moment conditions amongst a set that may contain incorrect ones. Thus, if the researcher is unsure whether or not the moment conditions are valid, these testing procedures could be used to select the (sub)set of correct moment conditions and then averaging would take place over estimates obtained from valid specifications utilizing different combinations of the selected moment conditions.

4 Properties of the MA GMM Estimator

Given our initial assumptions in section 2 and the consistency properties of the moment selection criteria as defined in Andrews (1999), we can assume our procedure is averaging over valid specifications. Indeed, for a given ω, the limit statistical properties of $\widehat{\theta}_T(\omega) = \sum_{c\in C}\omega_c\widehat{\theta}_{Tc}$ depend on a linear combination of the random processes $\widehat{\theta}_{Tc}, c \in C$. Thus, under correct model specification, plim $\widehat{\theta}_{Tc} = \theta_0$ for all $c \in C$ and $\widehat{\theta}_{Tc}$ is $\sqrt{T}$-gaussian with asymptotic variance

$$V_c = \left(G_c'W_cG_c\right)^{-1}\left(G_c'W_cS_cW_cG_c\right)\left(G_c'W_cG_c\right)^{-1} \tag{21}$$

where G_c, W_c, S_c and V_c denote limit quantities corresponding the particular specification using selected moments c. The asymptotic variance of the efficient GMM estimator is given by

$$V_c = \left(G_c' S_c^{-1} G_c\right)^{-1}. \tag{22}$$

However, we need to take into account the fact that, in our MA estimator, the moment functions $\widehat{g}_{Tc}(\theta_0)$ are different across model specifications c, which could complicate the derivation of their limiting behaviour. We circumvent this problem by defining a selection matrix that contains certain rows with zeros, operating on the full moment functions, as in Domowitz and White (1982), see also Newey (1985). Consider the GMM estimator obtained using the whole set of moment conditions, $c = \iota_m$, where $|c| = m$ (see Remark 2). Now, define the matrix Λ_c of dimension $|c|$ by m, such that each row $j = 1, ..., |c|$ contains zeros, except a single "1" at position i that corresponds to the moment condition as defined in model $c = \iota_m$. [5] Then,

$$\widehat{g}_{Tc}(\theta_0) = \Lambda_c \widehat{g}_T(\theta_0), \tag{23}$$

that is, we write the moment functions as a linear function of the 'full' specification, which will allow us to obtain the limiting distribution of our MA estimator, as shown in the following theorem.

Theorem 1 *(Distribution of the MA estimator): Assume that the model is correctly specified. As $T \to \infty$, for any $\omega \in H_m$,*

$$\widehat{\theta}_T(\omega) = \sum_{c \in C} \omega_c \widehat{\theta}_{Tc} \xrightarrow{p} \theta_0, \tag{24}$$

where $\widehat{\theta}_{Tc}, c \in C$, is the GMM estimator. Moreover,

$$\sqrt{T}\left(\widehat{\theta}_T(\omega) - \theta_0\right) \xrightarrow{d} N(0, V_\omega), \tag{25}$$

where

$$V_\omega = \left(\sum_{c \in C} \omega_c \left(G_c' W_c G_c\right)^{-1} G_c' W_c \Lambda_c\right) V \left(\sum_{c \in C} \omega_c \Lambda_c' W_c G_c \left(G_c' W_c G_c\right)^{-1}\right). \tag{26}$$

In the case of efficient GMM estimation, then $V_c = (22)$, so

$$V_\omega = \left(\sum_{c \in C} \omega_c V_c G_c' S_c^{-1} \Lambda_c\right) V \left(\sum_{c \in C} \omega_c \Lambda_c' S_c^{-1} G_c V_c\right). \tag{27}$$

Proof: Consistency follows from

$$\widehat{\theta}_T(\omega) = \sum_{c \in C} \omega_c \widehat{\theta}_{Tc} \xrightarrow{p} \sum_{c \in C} \omega_c \theta_0 = \theta_0. \tag{28}$$

[5] Taking, for example, $m = 3$ (three moment conditions) and the particular specification c using conditions one and three, Λ_c is 2 by 3 with rows $(1, 0, 0)$ and $(0, 0, 1)$.

The asymptotic distribution follows from the limiting law for $\sqrt{T}\left(\widehat{\theta}_{Tc}-\theta_0\right)$, noting that $\sqrt{T}\left(\widehat{\theta}_T(\omega)-\theta_0\right)$ equals

$$\sqrt{T}\left(\sum_{c\in C}\omega_c\widehat{\theta}_{Tc}-\sum_{c\in C}\omega_c\theta_0\right)=\sum_{c\in C}\omega_c\sqrt{T}\left(\widehat{\theta}_{Tc}-\theta_0\right) \tag{29}$$

because $\sum_{c\in C}\omega_c=1$. For a given $c\in C$, $\widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)'W_{Tc}\widehat{g}_{Tc}\left(\widehat{\theta}_{Tc}\right)=0$, and expanding $\widehat{g}_{Tc}\left(\widehat{\theta}_{Tc}\right)$ around $\widehat{g}_{Tc}(\theta_0)$,

$$\widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)'W_{Tc}\widehat{g}_{Tc}(\theta_0)+\widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)'W_{Tc}\widehat{G}_{Tc}\left(\overline{\theta}_{Tc}\right)\left(\widehat{\theta}_{Tc}-\theta_0\right)=0, \tag{30}$$

where $\overline{\theta}_{Tc}$ is some value "between" $\widehat{\theta}_{Tc}$ and θ_0, by the Mean Value Theorem. Rearranging terms,

$$\begin{aligned}\sqrt{T}\left(\widehat{\theta}_{Tc}-\theta_0\right) &= -\left[\widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)'W_{Tc}\widehat{G}_{Tc}\left(\overline{\theta}_{Tc}\right)\right]^{-1}\widehat{G}_{Tc}\left(\widehat{\theta}_{Tc}\right)'W_{Tc}\sqrt{T}\widehat{g}_{Tc}(\theta_0) && (31)\\ &= -\left(G_c'W_cG_c\right)^{-1}G_c'W_c\sqrt{T}\widehat{g}_{Tc}(\theta_0)+o_p(1) && (32)\end{aligned}$$

(see Newey, 1985).

Given (23),

$$\sqrt{T}\left(\widehat{\theta}_T(\omega)-\theta_0\right)=-\sum_{c\in C}\omega_c\left(G_c'W_cG_c\right)^{-1}G_c'W_c\Lambda_c\sqrt{T}\widehat{g}_T(\theta_0)+o_p(1), \tag{33}$$

with $\sqrt{T}\widehat{g}_T(\theta_0)$ not indexed by c, and the asymptotic variance-covariance matrix of $\sqrt{T}\left(\widehat{\theta}_T(\omega)-\theta_0\right)$ is given by

$$V_\omega=\left(\sum_{c\in C}\omega_c\left(G_c'W_cG_c\right)^{-1}G_c'W_c\Lambda_c\right)V\left(\sum_{c\in C}\omega_c\Lambda_c'W_cG_c\left(G_c'W_cG_c\right)^{-1}\right) \tag{34}$$

QED.

Remark 5. The results in the previous Theorem, namely the closed form expression of the asymptotic covariance of the MA estimator is very general in the context of GMM-type of estimation procedures. First, it covers the cases of linear IV and maximum likelihood estimators. Second, since model selection is indeed a special case of MA whenever $\omega_{\widetilde{c}}=1$ and $\omega_{c'}=0$, for all $c'\neq\widetilde{c}$, for some model $c=\widetilde{c}$, we have $\widehat{\theta}_T(\omega)=\widehat{\theta}_{T\widetilde{c}}$ and, more importantly, $V_\omega=V_{\widetilde{c}}$.

In general, V_ω does not equal the (squared) weighted sum of variances V_c, that is,

$$V_\omega=\left(\sum_{c\in C}\omega_cV_cG_c'S_c^{-1}\Lambda_c\right)V\left(\sum_{c\in C}\omega_c\Lambda_c'S_c^{-1}G_cV_c\right)\neq\sum_{c\in C}\omega_c^2V_c, \tag{35}$$

since there are pairs $c_1,c_2\in C, c_1\neq c_2$, such that c_1 and c_2 share commom moment conditions. Therefore, finding, as a new weighting selection criteria, the argument $\widehat{\omega}$ that minimizes the trace of V_ω,

$$tr\left(V_\omega\right)=tr\left(\left(\sum_{c\in C}\omega_cV_cG_c'S_c^{-1}\Lambda_c\right)V\left(\sum_{c\in C}\omega_c\Lambda_c'S_c^{-1}G_cV_c\right)\right), \tag{36}$$

may be hard to accomplish in practice and, on the other hand, finding the argument $\widehat{\omega}$ that minimizes

$$tr\left(\sum_{c\in C}\omega_c^2 V_c\right) = \sum_{c\in C}\omega_c^2.tr\left(V_c\right) \tag{37}$$

would not be reasonable as it does not equal $tr\left(V_\omega\right)$.

Thus, for a given ω, and noting that Λ_c is known for all $c \in C$, a consistent estimator of V_ω can be obtained using consistent estimators for G_c and W_c, for all $c \in C$, and for V as well, and inference can be carried out in the usual way. Any value for ω can be chosen, in fact. In the previous section, we provide and recommend the use of any of the several criteria for selecting ω. Clearly, the asymptotic covariance matrix V_ω will differ across methods for obtaining $\widehat{\omega}$. Still, notice that the expression for V_ω is derived independently on the criteria in section 3 that we pick for replacing ω by $\widehat{\omega}$.

5 Monte Carlo Study

In this section, we report some results of an extensive Monte Carlo study assessing the finite sample properties of the proposed model averaging estimators.[6] To facilitate comparisons, we base our experiments on the design used in Donald and Newey (2001), and subsequently employed in Donald, Imbens and Newey (2009), Eryuruk et al. (2009), Kuersteiner and Okui (2010) and Okui (2011), for example. As in Kuersteiner and Okui (2010) and Okui (2011), we use the selection method of Donald and Newey (2001) as the benchmark against which our procedures are compared.

The data generating process is

$$y_i = \beta_0 Y_i + \varepsilon_i, \qquad Y_i = \pi' Z_i + u_i, \qquad i = 1, ..., N, \tag{38}$$

where the true parameter of interest is the scalar β_0, which is fixed at 0.1, Y_i is a scalar, while

$$\left(\varepsilon_i, u_i, Z_i'\right)' \sim i.i.d.N\left(0, \Sigma\right), \text{ where } \Sigma = \begin{pmatrix} 1 & \sigma_{\varepsilon u} & 0_{1\times M} \\ \sigma_{\varepsilon u} & 1 & 0_{1\times M} \\ 0_{M\times 1} & 0_{M\times 1} & I_M \end{pmatrix}. \tag{39}$$

The degree of endogeneity is defined by $\sigma_{\varepsilon u}$, which we set as $\sigma_{\varepsilon u} \in \{0.1, 0.5, 0.9\}$. The number of observations is $N \in \{100, 1000\}$ and the number of instruments is

$$M = \begin{cases} 10, 20 \text{ if } N = 100 \\ 30, \text{ if } N = 1000 \end{cases}$$

[6]The full set of results, not reported here, is available upon request - more on this below.

and the number of replications is 5000. In terms of specifications for π, we have, for $m = 1, ..., M$,

$$\text{Model A (equal coefficients)} \quad : \quad \pi_m = \sqrt{\frac{R_f^2}{M\left(1 - R_f^2\right)}} \text{ and} \tag{40}$$

$$\text{Model B (declining coefficients)} \quad : \quad \pi_m = c(M)\left(1 - \frac{m}{M+1}\right)^4, \tag{41}$$

$$\text{Model C (irrelevant instruments)} \quad : \quad \pi_m = \begin{cases} 0, \\ c(M)\left(1 - \frac{m - M/2}{M/2+1}\right)^4, \end{cases} \tag{42}$$

where $c(M)$ is set so that π satisfies $\pi'\pi = R_f^2/\left(1 - R_f^2\right)$, where $R_f^2 \in \{0.1, 0.01\}$. The value of $R_f^2 = 0.01$ can be interpreted as the "weak instruments" case, quite common in empirical applications.

In model A, all instruments are equally important (and relatively weak), which means that instrument selection methods may not be very effective. In model B, the strength of the instruments declines gradually, but the ordering matters. Finally, in model C we allow for the first $M/2$ instruments to be irrelevant, while the others remain informative. The overall purpose is to examine how well each MA procedure estimates β along a number of distinct dimensions: model specification, sample size (N), degree of endogeneity ($\sigma_{\varepsilon u}$), number of instruments (M) and strength of instruments (R_f^2).

We consider two sets of experiments. First, estimation is conducted for the unconditional moments $E[Z_{c,i}(y_i - \beta Y_i)] = 0$, where $Z_{c,i}$ is $c \times 1$ for all models $c = m = 1, ..., M$, i.e. instruments are added sequentially and then we compute the MA estimators as a linear combination of M weights, where each Z_m is the matrix of the first m elements of Z.[7] By not taking all possible combinations of instruments, one avoids estimating a large number of weights, assuming implicitly that β_0 is identified by all the subsets of the candidate set considered. Notice, however, that for Models B and C the instruments are included in the order of their explanatory power, which corresponds to the case in which the practitioner knows the relative importance of the instruments. Thus, these conditions may favour selection methods such as the benchmark case of Donald and Newey (2001) - the experiments in Okui (2011) and Eryuruk et al. (2009) suggest that sequentially adding the instruments in the 'wrong' order causes the performance of selection methods to deteriorate.

Then, in a second set of experiments, we allow for different combinations of instruments. For some settings, the number of combinations is inevitably quite large, which may lead to poor estimation of the weight vector and additional computational difficulties (for example, there are 1023 possible combinations when $M = 10$). As recommended by Andrews (1999), we circumvent this issue by fixing a block of moment conditions (M_{fixed}), assumed to be valid. Thus, for each replication, we select the instruments for the fixed block that maximise the correlation with the

[7]For example, when $M = 10$, then we average over 10 different outcomes, i.e., estimates with $m = 1, 2, ..., 10$.

endogenous regressor Y_i, while using all possible combinations of the remaining instruments, up to a maximum of 255 combinations (i.e., all possible combinations of 8 'free' instruments).

Following Donald and Newey (2001), we compute their estimator (2SLS-DN) using a cross-validation criterion for the first-stage reduced-form model. For each estimator, we compute the median bias and the median absolute deviation relative to that of 2SLS-DN, as in Kuersteiner and Okui (2010). The MA-MSC estimator is computed using $W_{T\bar{c}} = \left(\frac{Z'Z}{T}\right)^{-1}$ and with $\bar{c} = \iota_m$ (all instruments).[8]

The results can be found in Tables 1-9 and a few general conclusions should be highlighted. First, there are several instances in which a model averaging approach, regardless of the specific type of estimation procedure, performs better than the selection procedure of Donald and Newey (2001). There can be substantial gains when $R_f^2 = 0.1$, but results then depend on the particular degree of endogeneity, whilst for $R_f^2 = 0.01$ the advantages are less striking, but are much more consistent. One estimator in particular, the MA-CCIC, emerges as a remarkably robust procedure, steadily delivering a good balance between lower median bias and median absolute deviations, irrespective of the experimental setting. The advantages of this method are particularly noticeable in terms of RMAD and when instruments are weaker ($R_f^2 = 0.01$).

Second, the use of smooth estimators often leads to considerable reductions in bias, although these gains depend on the particular parameterization of the simulations, i.e. the behaviour of this type of estimators is somewhat less congruent. Third, the performance of the MA-MSC appears to be particularly sensitive to the degree of endogeneity, e.g, this procedure tends to perform worse for $\sigma_{\varepsilon u} = 0.1$, but it is often the best method when $\sigma_{\varepsilon u} = 0.9$.

As expected, outcomes differ according to the model under study and to whether instruments are added sequentially or different combinations are considered, so we first analyse the experiments based on the sequential strategy outline above. In this case, the MA estimators largely dominate in terms of median bias, delivering in general less biased estimates than 2SLS-DN (in some cases quite dramatically, for example for the smoothing MA estimators in model B). Moreover, MA estimators are more precise than 2SLS-DN - as measured by RMAD - for medium and high degrees of endogeneity ($\sigma_{\varepsilon u} = 0.5$ and $\sigma_{\varepsilon u} = 0.9$). In particular, the MA-CICC (and the MA-RMSC as well) seems to be quite robust and in most cases performs better than 2SLS-DN in terms of MAD.

The results are similar across different models, but the advantage of MA estimators is more marked for models B and C, in particular for smoothing procedures. This is remarkable, as we would expect these models to favour selection models, as mentioned before. Interestingly, the relative improvements when $\sigma_{\varepsilon u}$ increases are more noticeable in terms of RMAD than RMB, although MA estimators remain better. The exception seems to be when $M = 30$ (also

[8] We also experimented with using $W_{T\bar{c}} = \widehat{S}_{T\bar{c}}^{-1}$ (HAC formula) and $\bar{c} = (1, 0_{m-1})'$ (only the first instrument), but the results changed very little.

corresponding to $N = 1000$), in which case the performance of non-smooth estimators worsens when $\sigma_{\varepsilon u} = 0.9$.

Turning to the experimental setting in which averaging is done over all possible combinations of 'free' instruments, in Tables 10 to 11, we found that the number of instruments M, the size of the fixed instruments M_{fixed} and the sample size N were of secondary importance in explaining the performance of the estimators. Rather, differences across models, namely between models A and B/C, were more relevant, so to save space, we only report results for models A and B, with $M = 10$ and $M_{fixed} = 2$. In the case of model A, results are similar to the sequential case.[9] For model B (which should, in principle, favour selection methods), the behaviour of MA estimators deteriorates, relative to 2SLS-DN and the sequential experiment, when $R_f^2 = 0.1$. However, in the case of 'weaker' identification ($R_f^2 = 0.01$), the MA estimators regain some of their competitiveness and are, at worst, indistinguishable from 2SLS-DN.

Given that our focus in on the MA estimators, we say little about the estimation of the weights. Nonetheless, we also analyze how the behaviour of empirical weights changes with the experiments and how it influences the properties of the MA estimators. In particular, we look at the Relative Mean Difference (RMD) of weights, which gives us a measure of dispersion of the empirical weights relative to the simple weight $\frac{1}{M}$ (the RMD is half of the Gini coefficient). We also consider the measure KW used in Kuersteiner and Okui (2010), which is the average value of $\sum_{m=1}^{M} m \max(\omega_m, 0)$, thus conveying information on the relative importance of the 'dimension' of the specifications used in model averaging (e.g., a higher KW means that, on average, specifications with more instruments are given larger weights). We then analyse the correlation between these measures and absolute bias ($\rho_{RMD,|bias|}$ and $\rho_{KW,|bias|}$) for each estimator. Note that we are essentially analysing features of the unconditional distribution of the empirical weights, that is, irrespective of which specification is attached to each weight, which is what actually determines the bias properties of the estimators

First, we note that results are similar in both the sequential and combined settings. Regarding the RMD, we note that it drops as $\sigma_{\varepsilon u}$ increases, perhaps suggesting that the procedures are less able to distinguish between the different specifications and, therefore, weights tend to be more equal. This is also the case when M increases, except for for smooth estimators, as these have a less flexible weighting scheme (also note that in general, RMD is larger for smooth estimators, while its the smallest for MSC, indicating that the empirical weights are often close to $1/M$).

In terms of the KW measure, it tends to be higher for non-smooth estimators, in particular the RMSC method. This is expected, as the smooth weights are directly based on actual information criteria for each specification, thus favouring more parsimonious models, whereas

[9] Results tend to deteriorate slightly with a lower M_{fixed}, while MA-MSC performs poorly when $\sigma_{\varepsilon u} = 0.1$, but otherwise similar for higher endogeneity.

the empirical weights for non-smooth estimators are obtained from minimizing the corresponding criterion. Changes in the KW measure are not very significant across different $\sigma_{\varepsilon u}$'s, models and R_f^2, with the exception of the sequential setting, in which case KW measure tends to approximate $M/2$ when $\sigma_{\varepsilon u}$ increases for smaller M's.

However, we found a less clear cut pattern in terms of the relation of these measures with estimation bias. For smooth estimators, more concentrated weighting schemes induce larger biases, but this effect is attenuated for higher endogeneity and weaker instruments in the combined setting. In the case of non-smooth estimators, correlations are usually negative (mainly for CCIC), but results vary substantially. Thus, it seems that while the distribution of the empirical weights plays a role in the performance of the estimators, a more detailed analysis of the specific pairs weight-specification would have to be carried out.[10]

As mentioned earlier, the full Monte Carlo experiment took into account other details, which, nonetheless, did not significantly affect the overall results. For example, we considered 'trimming' the combinatorial set when this was large, averaging only over the 50 or 100 best combinations of instruments, where the latter are selected based on the minimization of the different model selection criteria. Results improve slightly, but they are qualitatively similar to the case where no such trimming occurs. As a robustness check, we also analysed the 'degree' of averaging, given by the number of combined instruments as a proportion of M, but we found that the outcomes are qualitatively the same when we increase the proportion of free instruments for possible combinations for the same number of instruments (from 4 out of 10 to 6 out of 10 and 8 out of 10), but also when the total number of instruments increase (to 20 and 30), in which case the proportion of fixed instruments increase. Thus, we feel that the above results provide a concise description of the finite sample properties of the MA GMM estimators studied in this paper.

6 Conclusion

This paper develops GMM-based model averaging estimators. We use different moment selection criteria to select weights for averaging across GMM estimates. This can be achieved by direct smoothing of information criteria arising from the estimation stage, or by numerical minimization of a specific criterion, as in Hansen (2007). We study the asymptotic properties of the resulting estimators for correctly specified models. Monte Carlo experiments show that our MA estimation procedure outperforms the optimal instrument selection method of Donald and Newey (2001)

[10]We also considered other measures (not reported here), such as the range of the weights (which tends to be larger for smooth estimators, as expected from the results discussed before) and the correlation of bias with the number of instruments associated with the specification with the largest weight, but it is difficult to pin down an underlying behaviour.

in many relevant setups, including models with weak instruments.

There are a few issues that should deserve further attention. Indeed, it would be interesting to study the behaviour of the estimator under misspecification of the moment conditions. In addition, one should also look at the relative efficiency of the MA estimator when compared, for example, with the GMM estimator using the full moment conditions. Also, it would be important to further investigate the statistical properties of the MA estimators for models with many moments, weak identification and with conditional moment restrictions.

Note that, following the literature on model averaging, we consider weighting schemes based on the GMM counterparts of typical model selection criteria (Andrews, 1999), to which we add criteria that emphasize the validity and relevance of orthogonality conditions (Hall and Peixe, 2003 and Hall et al., 2007). Another possibility would be to study different criteria, such as a Mallows-type criterion, which focuses on (approximate) mean square error minimization, as in Donald and Newey (2001) and Hansen (2007).

A further extension would include selection criteria for estimators that belong to the class of Generalized Empirical Likelihood (GEL) estimators, see Newey and Smith (2004) for a general framework. With respect to GEL estimation method, it would be interesting to develop MA estimators with weights based on the GEL model selection criteria of Hong, Preston and Shum (2003). Contrary to our GMM approach, this alternative does not depend on any weight matrix in the J-statistic, but has the disadvantage of needing to estimate a Lagrange multiplier which, in model averaging, is likely complicate the analysis. Still, this is an important topic in the literature on MA for moment conditions models that is left for future research.

References

[1] Andrews, D. W. K. (1999), "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica,* 67, 543-564.

[2] Andrews, D. W. K. and Lu, B. (2001), "Consistent Model and Moment Selection Criteria for GMM Estimation with Applications to Dynamic Panel Data Models," *Journal of Econometrics,* 101, 123-164.

[3] Buckland, S. T., Burnham, K. P. and Augustin, N. H. (1997), "Model Selection: an Integral Part of Inference", *Biometrics,* 53, 603-618.

[4] Burnham, K. P and Anderson, D. R. (2002), *Model Selection and Multimodal Inference: a Practical Information-Theoretic Approach*, Berlin, Springer-Verlag.

[5] Canay, I. A. (2010), "Simultaneous Selection and Weighting of Moments in GMM Using a Trapezoidal Kernel," *Journal of Econometrics,* 156, 284-303.

[6] Caner, M. (2009), "LASSO Type GMM Estimator," *Econometric Theory*, 25, 1-23.

[7] Chao, J. C. and Swanson, N. R. (2005), "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica*, 73, 1673-1692.

[8] Claeskens, G. and Hjort, N. L. (2008), *Model Selection and Model Averaging*, Cambridge University Press.

[9] Domowitz, I. and White, H. (1982), "Misspecified Models with Dependent Observations," *Journal of Econometrics*, 20, 35-58.

[10] Donald, S. G., Imbens, G. W. and Newey, W. K. (2009), "Choosing Instrumental Variables in Conditional Moment Restriction Models," *Journal of Econometrics,* 152, 28-36.

[11] Donald, S. G., Newey, W. K. (2001), "Choosing the Number of Instruments," *Econometrica,* 69, 1161-1192.

[12] Eryuruk, G., Hall, A. R., and Janas, K. (2009), "A Comparative Study of Three Data-Based Methods of Instrument Selection," *Economics Letters,* 105, 280-283.

[13] 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.

[14] Hall, A. R. and Peixe, F. P .M. (2003), "A Consistent Method for the Selection of Relevant Instruments," *Econometric Reviews,* 22, 269-287.

[15] Han, C. and Phillips, P. C. B. (2006), "GMM with Many Moment Conditions," *Econometrica* 74, 147-182.

[16] Hansen, B. E. (2007), "Least Squares Model Averaging," *Econometrica*, 75, 1175-1189.

[17] Hansen, C., Hausman, J. and Newey, W. K. (2008), "Estimation with Many Instrumental Variables," *Journal of Business and Economics Statistics*, 26, 398-422.

[18] Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

[19] Hong, H., Preston, B. and Shum, M. (2003), "Generalized Empirical Likelihood-Based Model Selection Criteria for Moment Condition Models," *Econometric Theory*, 19, 923-943.

[20] Kapetanios, G., Marcellino, M. (2010), "Cross-sectional Averaging and Instrumental Variable Estimation with Many Weak Instruments," *Economics Letters*, 108, 36-39.

[21] Kuersteiner, G. and Okui, R. (2010), “Constructing Optimal Instruments by First-Stage Prediction Averaging,” *Econometrica*, 78, 697-718.

[22] Newey, W. K. (1985), “Generalized Method of Moments Specification Testing,” *Journal of Econometrics,* 29, 229-256.

[23] Newey, W. K. (1993), ”Efficient Estimation of Models with Conditional Moment Restrictions,” in *Handbook of Statistics*, Volume 11: Econometrics, ed. by G. S. Maddala, C. R. Rao, and H. D. Vinod. Amsterdam: North-Holland.

[24] Newey, W. K. and Smith, R. J. (2004), “Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators,” *Econometrica*, 72, 219-255.

[25] Newey, W. K. and Windmeijer, F. (2009), “Generalized Method of Moments with Many Weak Moment Conditions,” *Econometrica*, 77, 687-719.

[26] Okui, R. (2011), “Instrumental Variable Estimation in the Presence of Many Moment Conditions,” *Journal of Econometrics* , 165, 70-86.

[27] Smith, R. J. (1992), “Non-nested Tests for Competing Models Estimated by Generalized Method of Moments,” *Econometrica*, 60, 973-980.

[28] Smith, R. J. and Ramalho, J. S. (2002), “Generalized Empirical Likelihood Non-Nested Tests,” *Journal of Econometrics*, 107, 99-125.

[29] Staiger, D. and Stock, J. H. (1997), “Instrumental Variables Regression with Weak Instruments, *Econometrica*, 65, 55786.

[30] Stock, J. H. and Wright, J. H. (2000), “GMM With Weak Identification,” *Econometrica*, 68, 1055-1096.

A Tables

Table 1: Instruments added sequentially, Model A, $M = 10$, $N = 100$

		DN2SLS	S-BIC	S-AIC	S-HQ	MA-MSC	MA-$RMSC$	MA-$CCIC$
$\sigma_{\varepsilon u} = 0.1$, $R_f^2 = 0.1$	MB	0.0600	0.0524	0.0529	0.0529	0.0785	0.0562	0.0509
	$[RMB]$		[0.8728]	[0.8809]	[0.8817]	[1.3077]	[0.9372]	[0.8481]
	MAD	0.1587	0.3081	0.3111	0.3152	0.3008	0.1675	0.1201
	$[RMAD]$		[1.9411]	[1.9601]	[1.9856]	[1.8951]	[1.0552]	[0.7565]
	RMD		1.0029	1.0273	1.0540	0.4492	0.6688	0.7482
			(0.4259)	(0.4197)	(0.4118)	(0.4502)	(0.2178)	(0.4132)
	KW		3.3096	3.2372	3.1621	4.7716	7.0034	6.1788
	$\rho_{RMD,\lvert bias\rvert}$		0.1098	0.1078	0.1057	0.0070	0.0685	-0.0693
			(0.0140)	(0.0140)	(0.0140)	(0.0141)	(0.0141)	(0.0141)
	$\rho_{KW,\lvert bias\rvert}$		-0.1080	-0.1064	-0.1048	-0.0468	0.0657	0.0653
			(0.0140)	(0.0140)	(0.0140)	(0.0141)	(0.0141)	(0.0141)
$R_f^2 = 0.01$	MB	0.0870	0.0872	0.0860	0.0852	0.1146	0.0853	0.0750
	$[RMB]$		[1.0015]	[0.9877]	[0.9787]	[1.3163]	[0.9801]	[0.8623]
	MAD	0.2043	0.3964	0.3993	0.4035	0.3847	0.2192	0.1503
	$[RMAD]$		[1.9398]	[1.9543]	[1.9746]	[1.8828]	[1.0729]	[0.7354]
	RMD		1.0064	1.0308	1.0573	0.4582	0.6534	0.8144
			(0.4226)	(0.4163)	(0.4084)	(0.4533)	(0.2454)	(0.4232)
	KW		3.3028	3.2306	3.1557	4.7498	6.9660	6.1792
	$\rho_{RMD,\lvert bias\rvert}$		0.2298	0.2284	0.2271	-0.0062	0.0588	-0.1510
			(0.0134)	(0.0134)	(0.0134)	(0.0141)	(0.0141)	(0.0138)
	$\rho_{KW,\lvert bias\rvert}$		-0.2274	-0.2267	-0.2260	-0.1058	0.0516	0.0572
			(0.0134)	(0.0134)	(0.0134)	(0.0140)	(0.0141)	(0.0141)
$\sigma_{\varepsilon u} = 0.5$, $R_f^2 = 0.1$	MB	0.3067	0.2904	0.2910	0.2920	0.2885	0.2755	0.2948
	$[RMB]$		[0.9469]	[0.9487]	[0.9519]	[0.9406]	[0.8983]	[0.9610]
	MAD	0.3202	0.3691	0.3711	0.3740	0.3645	0.2832	0.2992
	$[RMAD]$		[1.1526]	[1.1591]	[1.1682]	[1.1384]	[0.8843]	[0.9345]
	RMD		0.9104	0.9365	0.9652	0.3005	0.6591	0.2825
			(0.4152)	(0.4106)	(0.4041)	(0.4068)	(0.2231)	(0.2175)
	KW		3.5326	3.4527	3.3696	4.9963	6.9802	5.8215
	$\rho_{RMD,\lvert bias\rvert}$		0.2441	0.2424	0.2407	0.0011	-0.0220	-0.3272
			(0.0133)	(0.0133)	(0.0133)	(0.0141)	(0.0141)	(0.0126)
	$\rho_{KW,\lvert bias\rvert}$		-0.2415	-0.2406	-0.2396	-0.0854	-0.0197	-0.0228
			(0.0133)	(0.0133)	(0.0133)	(0.0140)	(0.0141)	(0.0141)
$R_f^2 = 0.01$	MB	0.4658	0.4571	0.4579	0.4590	0.4594	0.4574	0.4550
	$[RMB]$		[0.9815]	[0.9831]	[0.9856]	[0.9865]	[0.9821]	[0.9768]
	MAD	0.4803	0.5158	0.5171	0.5194	0.5168	0.4598	0.4570
	$[RMAD]$		[1.0739]	[1.0766]	[1.0813]	[1.0760]	[0.9574]	[0.9515]
	RMD		0.8255	0.8525	0.8826	0.2554	0.6571	0.2909
			(0.4055)	(0.4032)	(0.3990)	(0.3843)	(0.2460)	(0.2195)
	KW		3.7391	3.6524	3.5621	5.0592	6.9743	5.8967
	$\rho_{RMD,\lvert bias\rvert}$		0.2193	0.2183	0.2173	-0.0799	0.0071	-0.0250
			(0.0135)	(0.0135)	(0.0135)	(0.0141)	(0.0141)	(0.0141)
	$\rho_{KW,\lvert bias\rvert}$		-0.2152	-0.2152	-0.2152	0.0006	0.0061	0.0461
			(0.0135)	(0.0135)	(0.0135)	(0.0141)	(0.0141)	(0.0141)
$\sigma_{\varepsilon u} = 0.9$, $R_f^2 = 0.1$	MB	0.5506	0.4908	0.4915	0.4914	0.4892	0.5013	0.5035
	$[RMB]$		[0.8913]	[0.8925]	[0.8924]	[0.8884]	[0.9104]	[0.9144]
	MAD	0.5608	0.5000	0.5013	0.5039	0.5041	0.5013	0.5035
	$[RMAD]$		[0.8917]	[0.8939]	[0.8986]	[0.8990]	[0.8940]	[0.8979]
	RMD		0.6190	0.6485	0.6825	0.1183	0.6433	0.1611
			(0.3444)	(0.3477)	(0.3488)	(0.2585)	(0.2568)	(0.0984)
	KW		4.2370	4.1345	4.0272	5.2865	6.9252	5.5988
	$\rho_{RMD,\lvert bias\rvert}$		0.2738	0.2737	0.2733	-0.0394	-0.1069	-0.4917
			(0.0131)	(0.0131)	(0.0131)	(0.0141)	(0.0140)	(0.0107)
	$\rho_{KW,\lvert bias\rvert}$		-0.2728	-0.2728	-0.2728	-0.0260	-0.0944	-0.2360
			(0.0131)	(0.0131)	(0.0131)	(0.0141)	(0.0140)	(0.0134)
$R_f^2 = 0.01$	MB	0.8472	0.8270	0.8279	0.8282	0.8304	0.8253	0.8322
	$[RMB]$		[0.9762]	[0.9772]	[0.9776]	[0.9801]	[0.9741]	[0.9823]
	MAD	0.8480	0.8271	0.8281	0.8286	0.8318	0.8253	0.8322
	$[RMAD]$		[0.9754]	[0.9765]	[0.9771]	[0.9809]	[0.9732]	[0.9814]
	RMD		0.2401	0.2425	0.2546	0.0404	0.6614	0.1040
			(0.1356)	(0.1518)	(0.1698)	(0.1141)	(0.2482)	(0.0809)
	KW		5.3773	5.2551	5.1249	5.4241	6.9811	5.6519
	$\rho_{RMD,\lvert bias\rvert}$		0.0432	0.0466	0.0490	-0.1028	-0.0027	-0.1145
			(0.0141)	(0.0141)	(0.0141)	(0.0140)	(0.0141)	(0.0140)
	$\rho_{KW,\lvert bias\rvert}$		-0.0494	-0.0506	-0.0519	0.0742	0.0028	-0.0101
			(0.0141)	(0.0141)	(0.0141)	(0.0141)	(0.0141)	(0.0141)

Notes: numbers in square brackets are measures relative to the Donald-Newey estimator; standard deviations in round brackets.

Table 2: Instruments added sequentially, Model A, $M = 20$, $N = 100$

		DN2SLS	S-BIC	S-AIC	S-HQ	MA-MSC	MA-$RMSC$	MA-$CCIC$
$\sigma_{\varepsilon u} = 0.1$, $R_f^2 = 0.1$	MB	0.0743	0.0664	0.0661	0.0671	0.0872	0.0692	0.0658
	$[RMB]$		[0.8940]	[0.8905]	[0.9037]	[1.1744]	[0.9314]	[0.8859]
	MAD	0.1377	0.3284	0.3327	0.3394	0.2640	0.1461	0.1151
	$[RMAD]$		[2.3844]	[2.4155]	[2.4637]	[1.9163]	[1.0603]	[0.8357]
	RMD		1.4036	1.4286	1.4542	0.4158	0.6583	0.6446
			(0.3089)	(0.2948)	(0.2797)	(0.4333)	(0.1590)	(0.3774)
	KW		3.9111	3.7708	3.6307	9.0947	13.6260	11.8533
	$\rho_{RMD,\|bias\|}$		0.2850	0.2835	0.2823	0.1883	0.0621	−0.0587
			(0.0130)	(0.0130)	(0.0130)	(0.0136)	(0.0141)	(0.0141)
	$\rho_{KW,\|bias\|}$		−0.2828	−0.2821	−0.2815	−0.3057	0.0612	0.0597
			(0.0130)	(0.0130)	(0.0130)	(0.0128)	(0.0141)	(0.0141)
$R_f^2 = 0.01$	MB	0.0944	0.0817	0.0817	0.0804	0.1139	0.0967	0.0812
	$[RMB]$		[0.8659]	[0.8662]	[0.8517]	[1.2069]	[1.0250]	[0.8604]
	MAD	0.1617	0.3717	0.3763	0.3810	0.3042	0.1725	0.1306
	$[RMAD]$		[2.2987]	[2.3275]	[2.3564]	[1.8814]	[1.0668]	[0.8077]
	RMD		1.3945	1.4201	1.4465	0.4296	0.6649	0.6888
			(0.3127)	(0.2982)	(0.2828)	(0.4453)	(0.1701)	(0.3802)
	KW		3.9561	3.8118	3.6677	9.0556	13.6562	11.8495
	$\rho_{RMD,\|bias\|}$		0.1964	0.1953	0.1945	0.1220	0.0823	−0.0828
			(0.0136)	(0.0136)	(0.0136)	(0.0139)	(0.0140)	(0.0140)
	$\rho_{KW,\|bias\|}$		−0.1945	−0.1940	−0.1937	−0.2291	0.0804	0.0507
			(0.0136)	(0.0136)	(0.0136)	(0.0134)	(0.0141)	(0.0141)
$\sigma_{\varepsilon u} = 0.5$, $R_f^2 = 0.1$	MB	0.3721	0.3634	0.3633	0.3636	0.3503	0.3498	0.3539
	$[RMB]$		[0.9768]	[0.9765]	[0.9774]	[0.9417]	[0.9403]	[0.9513]
	MAD	0.3792	0.4256	0.4287	0.4313	0.3825	0.3500	0.3541
	$[RMAD]$		[1.1223]	[1.1304]	[1.1375]	[1.0086]	[0.9229]	[0.9337]
	RMD		1.2805	1.3131	1.3469	0.1961	0.6574	0.2038
			(0.3462)	(0.3310)	(0.3142)	(0.3350)	(0.1641)	(0.1419)
	KW		4.5238	4.3374	4.1516	9.7958	13.6214	11.1052
	$\rho_{RMD,\|bias\|}$		0.2384	0.2378	0.2377	0.0140	−0.0343	−0.3451
			(0.0133)	(0.0133)	(0.0133)	(0.0141)	(0.0141)	(0.0125)
	$\rho_{KW,\|bias\|}$		−0.2361	−0.2365	−0.2371	−0.1025	−0.0333	−0.0959
			(0.0134)	(0.0134)	(0.0133)	(0.0140)	(0.0141)	(0.0140)
$R_f^2 = 0.01$	MB	0.4842	0.4940	0.4944	0.4965	0.4823	0.4798	0.4792
	$[RMB]$		[1.0202]	[1.0210]	[1.0254]	[0.9962]	[0.9909]	[0.9898]
	MAD	0.4872	0.5396	0.5413	0.5448	0.5040	0.4798	0.4792
	$[RMAD]$		[1.1075]	[1.1110]	[1.1182]	[1.0345]	[0.9847]	[0.9836]
	RMD		1.2042	1.2410	1.2793	0.1626	0.6674	0.2033
			(0.3506)	(0.3361)	(0.3195)	(0.3056)	(0.1682)	(0.1357)
	KW		4.9052	4.6927	4.4806	9.8938	13.6685	11.1562
	$\rho_{RMD,\|bias\|}$		0.1954	0.1953	0.1952	−0.0740	−0.0130	−0.1384
			(0.0136)	(0.0136)	(0.0136)	(0.0141)	(0.0141)	(0.0139)
	$\rho_{KW,\|bias\|}$		−0.1932	−0.1939	−0.1946	0.0055	−0.0128	−0.0491
			(0.0136)	(0.0136)	(0.0136)	(0.0141)	(0.0141)	(0.0141)
$\sigma_{\varepsilon u} = 0.9$, $R_f^2 = 0.1$	MB	0.6577	0.6174	0.6187	0.6181	0.6195	0.6240	0.6269
	$[RMB]$		[0.9387]	[0.9406]	[0.9397]	[0.9419]	[0.9487]	[0.9532]
	MAD	0.6629	0.6195	0.6206	0.6208	0.6216	0.6240	0.6269
	$[RMAD]$		[0.9346]	[0.9362]	[0.9365]	[0.9378]	[0.9413]	[0.9458]
	RMD		0.7009	0.7584	0.8236	0.0382	0.6572	0.1037
			(0.3318)	(0.3333)	(0.3282)	(0.0992)	(0.1774)	(0.0630)
	KW		7.4944	7.1063	6.7122	10.3538	13.6141	10.7355
	$\rho_{RMD,\|bias\|}$		0.1626	0.1631	0.1634	−0.1967	−0.0637	−0.4452
			(0.0138)	(0.0138)	(0.0138)	(0.0136)	(0.0141)	(0.0113)
	$\rho_{KW,\|bias\|}$		−0.1615	−0.1621	−0.1629	0.1592	−0.0589	−0.2048
			(0.0138)	(0.0138)	(0.0138)	(0.0138)	(0.0141)	(0.0136)
$R_f^2 = 0.01$	MB	0.8705	0.8569	0.8568	0.8567	0.8581	0.8581	0.8593
	$[RMB]$		[0.9844]	[0.9843]	[0.9841]	[0.9858]	[0.9858]	[0.9871]
	MAD	0.8706	0.8569	0.8568	0.8568	0.8586	0.8581	0.8593
	$[RMAD]$		[0.9844]	[0.9842]	[0.9842]	[0.9862]	[0.9857]	[0.9871]
	RMD		0.3043	0.2987	0.3193	0.0180	0.6684	0.0732
			(0.1338)	(0.1555)	(0.1858)	(0.0533)	(0.1718)	(0.0558)
	KW		10.5695	10.0816	9.5636	10.4295	13.6724	10.7579
	$\rho_{RMD,\|bias\|}$		−0.0002	0.0070	0.0127	−0.1326	0.0057	−0.0924
			(0.0141)	(0.0141)	(0.0141)	(0.0139)	(0.0141)	(0.0140)
	$\rho_{KW,\|bias\|}$		−0.0131	−0.0152	−0.0173	0.1182	0.0065	−0.0221
			(0.0141)	(0.0141)	(0.0141)	(0.0139)	(0.0141)	(0.0141)

Table 3: Instruments added sequentially, Model A, $M = 30$,N=1000

		DN2SLS	S-BIC	S-AIC	S-HQ	MA-MSC	MA-$RMSC$	MA-$CCIC$
$\sigma_{\varepsilon u} = 0.1$, $R_f^2 = 0.1$	MB	0.0250	0.0063	0.0047	0.0048	0.0286	0.0260	0.0348
	$[RMB]$		[0.2523]	[0.1896]	[0.1938]	[1.1441]	[1.0393]	[1.3921]
	MAD	0.0647	0.2306	0.2354	0.2338	0.1296	0.0732	0.0697
	$[RMAD]$		[3.5627]	[3.6368]	[3.6119]	[2.0017]	[1.1307]	[1.0768]
	RMD		1.6735	1.6950	1.6880	0.3480	0.6695	0.3635
			(0.1674)	(0.1535)	(0.1580)	(0.4062)	(0.0909)	(0.2051)
	KW		3.3822	3.2167	3.2704	13.6964	20.3542	16.8734
	$\rho_{RMD,\|bias\|}$		0.1738	0.1735	0.1736	0.3481	0.0296	0.1715
			(0.0217)	(0.0217)	(0.0217)	(0.0197)	(0.0223)	(0.0217)
	$\rho_{KW,\|bias\|}$		−0.1738	−0.1736	−0.1737	−0.4097	0.0295	0.1780
			(0.0217)	(0.0217)	(0.0217)	(0.0186)	(0.0223)	(0.0217)
$R_f^2 = 0.01$	MB	0.0782	0.0740	0.0754	0.0744	0.0901	0.0715	0.0776
	$[RMB]$		[0.9464]	[0.9644]	[0.9520]	[1.1526]	[0.9144]	[0.9923]
	MAD	0.1151	0.3734	0.3851	0.3801	0.2386	0.1278	0.1124
	$[RMAD]$		[3.2436]	[3.3457]	[3.3021]	[2.0725]	[1.1105]	[0.9763]
	RMD		1.6686	1.6909	1.6836	0.3783	0.6719	0.5380
			(0.1804)	(0.1646)	(0.1698)	(0.4133)	(0.1295)	(0.2688)
	KW		3.4182	3.2465	3.3023	13.5785	20.3704	17.3602
	$\rho_{RMD,\|bias\|}$		0.1888	0.1905	0.1900	0.2205	0.0766	0.3044
			(0.0216)	(0.0216)	(0.0216)	(0.0213)	(0.0222)	(0.0203)
	$\rho_{KW,\|bias\|}$		−0.1890	−0.1906	−0.1901	−0.3170	0.0762	0.1129
			(0.0216)	(0.0216)	(0.0216)	(0.0201)	(0.0222)	(0.0221)
$\sigma_{\varepsilon u} = 0.5$, $R_f^2 = 0.1$	MB	0.1632	0.1438	0.1445	0.1447	0.1252	0.1310	0.1440
	$[RMB]$		[0.8812]	[0.8859]	[0.8867]	[0.7673]	[0.8031]	[0.8825]
	MAD	0.1701	0.2387	0.2442	0.2421	0.1580	0.1325	0.1450
	$[RMAD]$		[1.4034]	[1.4358]	[1.4234]	[0.9292]	[0.7793]	[0.8525]
	RMD		1.6457	1.6708	1.6626	0.2046	0.6625	0.1871
			(0.1797)	(0.1632)	(0.1686)	(0.3414)	(0.1016)	(0.1099)
	KW		3.5849	3.3923	3.4547	14.4103	20.3026	15.8000
	$\rho_{RMD,\|bias\|}$		0.2734	0.2775	0.2761	0.2753	−0.1680	−0.5011
			(0.0207)	(0.0206)	(0.0207)	(0.0207)	(0.0217)	(0.0168)
	$\rho_{KW,\|bias\|}$		−0.2735	−0.2775	−0.2762	−0.3606	−0.1675	−0.3421
			(0.0207)	(0.0206)	(0.0207)	(0.0195)	(0.0217)	(0.0197)
$R_f^2 = 0.01$	MB	0.4119	0.3915	0.3944	0.3944	0.3790	0.3822	0.3897
	$[RMB]$		[0.9506]	[0.9576]	[0.9575]	[0.9202]	[0.9280]	[0.9462]
	MAD	0.4145	0.4532	0.4575	0.4554	0.3915	0.3822	0.3897
	$[RMAD]$		[1.0933]	[1.1038]	[1.0987]	[0.9445]	[0.9221]	[0.9402]
	RMD		1.5796	1.6126	1.6019	0.1286	0.6651	0.1598
			(0.2082)	(0.1866)	(0.1936)	(0.2393)	(0.1345)	(0.0974)
	KW		4.0684	3.8151	3.8969	14.7986	20.3210	16.2045
	$\rho_{RMD,\|bias\|}$		0.2106	0.2131	0.2122	−0.1117	−0.0252	−0.3487
			(0.0214)	(0.0214)	(0.0214)	(0.0221)	(0.0224)	(0.0196)
	$\rho_{KW,\|bias\|}$		−0.2104	−0.2131	−0.2122	0.0360	−0.0245	−0.0553
			(0.0214)	(0.0214)	(0.0214)	(0.0223)	(0.0224)	(0.0223)
$\sigma_{\varepsilon u} = 0.9$, $R_f^2 = 0.1$	MB	0.2980	0.2262	0.2267	0.2265	0.2207	0.2321	0.2323
	$[RMB]$		[0.7590]	[0.7607]	[0.7602]	[0.7406]	[0.7789]	[0.7796]
	MAD	0.3128	0.2644	0.2683	0.2673	0.2260	0.2321	0.2323
	$[RMAD]$		[0.8452]	[0.8576]	[0.8545]	[0.7226]	[0.7420]	[0.7427]
	RMD		1.5865	1.6195	1.6088	0.0894	0.6477	0.1432
			(0.1974)	(0.1771)	(0.1836)	(0.2096)	(0.1358)	(0.0636)
	KW		4.0170	3.7655	3.8465	14.9881	20.1826	15.5971
	$\rho_{RMD,\|bias\|}$		0.2303	0.2361	0.2342	0.0111	−0.3379	−0.5726
			(0.0212)	(0.0211)	(0.0211)	(0.0224)	(0.0198)	(0.0150)
	$\rho_{KW,\|bias\|}$		−0.2305	−0.2362	−0.2343	−0.0724	−0.3377	−0.5292
			(0.0212)	(0.0211)	(0.0211)	(0.0222)	(0.0198)	(0.0161)
$R_f^2 = 0.01$	MB	0.7289	0.6824	0.6854	0.6847	0.6775	0.6827	0.6836
	$[RMB]$		[0.9361]	[0.9403]	[0.9393]	[0.9295]	[0.9367]	[0.9378]
	MAD	0.7312	0.6827	0.6858	0.6851	0.6775	0.6827	0.6836
	$[RMAD]$		[0.9337]	[0.9379]	[0.9370]	[0.9266]	[0.9338]	[0.9349]
	RMD		1.0996	1.2070	1.1723	0.0175	0.6621	0.0858
			(0.2785)	(0.2425)	(0.2547)	(0.0510)	(0.1467)	(0.0470)
	KW		7.5779	6.7624	7.0228	15.4000	20.2914	15.7350
	$\rho_{RMD,\|bias\|}$		0.1373	0.1411	0.1399	−0.1786	−0.0660	−0.4252
			(0.0219)	(0.0219)	(0.0219)	(0.0217)	(0.0223)	(0.0183)
	$\rho_{KW,\|bias\|}$		−0.1389	−0.1416	−0.1407	0.1608	−0.0610	−0.1799
			(0.0219)	(0.0219)	(0.0219)	(0.0218)	(0.0223)	(0.0216)

Table 4: Instruments added sequentially, Model B, $M = 10$, $N = 100$

		DN2SLS	S-BIC	S-AIC	S-HQ	MA-MSC	MA-$RMSC$	MA-$CCIC$
$\sigma_{\varepsilon u} = 0.1$, $R_f^2 = 0.1$	MB	0.0497	0.0169	0.0171	0.0164	0.0342	0.0381	0.0383
	$[RMB]$		[0.3408]	[0.3444]	[0.3297]	[0.6890]	[0.7674]	[0.7717]
	MAD	0.1462	0.1939	0.1950	0.1965	0.1945	0.1526	0.1186
	$[RMAD]$		[1.3265]	[1.3340]	[1.3444]	[1.3305]	[1.0441]	[0.8115]
	RMD		0.9993	1.0237	1.0504	0.3498	0.3346	0.5402
			(0.4170)	(0.4110)	(0.4034)	(0.4448)	(0.1555)	(0.4051)
	KW		3.3203	3.2476	3.1720	4.9659	6.2518	5.8267
	$\rho_{RMD,\|bias\|}$		0.2160	0.2146	0.2134	−0.0461	0.1416	−0.0582
			(0.0135)	(0.0135)	(0.0135)	(0.0141)	(0.0139)	(0.0141)
	$\rho_{KW,\|bias\|}$		−0.2129	−0.2123	−0.2118	−0.0758	0.1381	0.0372
			(0.0135)	(0.0135)	(0.0135)	(0.0141)	(0.0139)	(0.0141)
$R_f^2 = 0.01$	MB	0.0980	0.0728	0.0738	0.0747	0.1117	0.0879	0.0771
	$[RMB]$		[0.7428]	[0.7533]	[0.7623]	[1.1398]	[0.8965]	[0.7862]
	MAD	0.2097	0.3667	0.3710	0.3753	0.3661	0.2235	0.1532
	$[RMAD]$		[1.7484]	[1.7691]	[1.7893]	[1.7455]	[1.0657]	[0.7303]
	RMD		1.0078	1.0321	1.0585	0.4562	0.6041	0.7896
			(0.4199)	(0.4139)	(0.4061)	(0.4672)	(0.2424)	(0.4258)
	KW		3.3000	3.2277	3.1527	4.7704	6.8550	6.1008
	$\rho_{RMD,\|bias\|}$		0.0930	0.0919	0.0908	−0.0151	0.0898	−0.1366
			(0.0140)	(0.0140)	(0.0140)	(0.0141)	(0.0140)	(0.0139)
	$\rho_{KW,\|bias\|}$		−0.0918	−0.0910	−0.0902	−0.0409	0.0825	0.0639
			(0.0140)	(0.0140)	(0.0140)	(0.0141)	(0.0140)	(0.0141)
$\sigma_{\varepsilon u} = 0.5$, $R_f^2 = 0.1$	MB	0.2284	0.1203	0.1193	0.1181	0.1530	0.1904	0.1849
	$[RMB]$		[0.5269]	[0.5226]	[0.5171]	[0.6699]	[0.8339]	[0.8097]
	MAD	0.2631	0.2109	0.2109	0.2117	0.2294	0.2120	0.1983
	$[RMAD]$		[0.8019]	[0.8016]	[0.8047]	[0.8722]	[0.8059]	[0.7537]
	RMD		0.9089	0.9353	0.9643	0.2894	0.3661	0.2470
			(0.4222)	(0.4179)	(0.4118)	(0.4231)	(0.1772)	(0.2386)
	KW		3.5371	3.4560	3.3718	4.9877	6.3215	5.8638
	$\rho_{RMD,\|bias\|}$		0.1470	0.1469	0.1469	0.0344	−0.0055	−0.4041
			(0.0138)	(0.0138)	(0.0138)	(0.0141)	(0.0141)	(0.0118)
	$\rho_{KW,\|bias\|}$		−0.1427	−0.1438	−0.1448	−0.1097	−0.0040	−0.2784
			(0.0139)	(0.0139)	(0.0138)	(0.0140)	(0.0141)	(0.0130)
$R_f^2 = 0.01$	MB	0.4612	0.4149	0.4143	0.4131	0.4242	0.4425	0.4359
	$[RMB]$		[0.8995]	[0.8983]	[0.8956]	[0.9196]	[0.9594]	[0.9450]
	MAD	0.4742	0.4757	0.4766	0.4786	0.4907	0.4443	0.4380
	$[RMAD]$		[1.0031]	[1.0050]	[1.0093]	[1.0347]	[0.9370]	[0.9237]
	RMD		0.8234	0.8509	0.8817	0.2661	0.6032	0.2751
			(0.4117)	(0.4091)	(0.4044)	(0.4074)	(0.2489)	(0.2194)
	KW		3.7410	3.6537	3.5627	5.0300	6.8532	5.8662
	$\rho_{RMD,\|bias\|}$		0.2167	0.2157	0.2148	−0.0567	0.0579	−0.0690
			(0.0135)	(0.0135)	(0.0135)	(0.0141)	(0.0141)	(0.0141)
	$\rho_{KW,\|bias\|}$		−0.2122	−0.2124	−0.2125	−0.0202	0.0571	−0.0399
			(0.0135)	(0.0135)	(0.0135)	(0.0141)	(0.0141)	(0.0141)
$\sigma_{\varepsilon u} = 0.9$, $R_f^2 = 0.1$	MB	0.2511	0.2197	0.2160	0.2131	0.2726	0.3531	0.3120
	$[RMB]$		[0.8750]	[0.8600]	[0.8487]	[1.0853]	[1.4059]	[1.2424]
	MAD	0.2794	0.2601	0.2582	0.2562	0.3053	0.3533	0.3125
	$[RMAD]$		[0.9310]	[0.9240]	[0.9169]	[1.0927]	[1.2643]	[1.1186]
	RMD		0.6621	0.6922	0.7268	0.2565	0.5121	0.2147
			(0.3825)	(0.3853)	(0.3857)	(0.3918)	(0.2334)	(0.1304)
	KW		4.1436	4.0383	3.9284	4.9525	6.6480	5.9496
	$\rho_{RMD,\|bias\|}$		0.0903	0.0904	0.0913	0.0188	0.0490	−0.5672
			(0.0140)	(0.0140)	(0.0140)	(0.0141)	(0.0141)	(0.0096)
	$\rho_{KW,\|bias\|}$		−0.0749	−0.0796	−0.0844	−0.0323	0.0493	−0.5376
			(0.0141)	(0.0141)	(0.0140)	(0.0141)	(0.0141)	(0.0101)
$R_f^2 = 0.01$	MB	0.8249	0.7521	0.7501	0.7470	0.7533	0.7944	0.7691
	$[RMB]$		[0.9117]	[0.9093]	[0.9056]	[0.9132]	[0.9629]	[0.9323]
	MAD	0.8266	0.7529	0.7505	0.7483	0.7567	0.7944	0.7691
	$[RMAD]$		[0.9108]	[0.9079]	[0.9053]	[0.9154]	[0.9610]	[0.9305]
	RMD		0.2300	0.2248	0.2296	0.0652	0.6441	0.1193
			(0.1232)	(0.1383)	(0.1567)	(0.1738)	(0.2531)	(0.0915)
	KW		5.4851	5.3594	5.2253	5.3667	6.9444	5.7158
	$\rho_{RMD,\|bias\|}$		−0.0666	−0.0839	−0.0931	−0.2387	0.0510	−0.2582
			(0.0141)	(0.0140)	(0.0140)	(0.0133)	(0.0141)	(0.0132)
	$\rho_{KW,\|bias\|}$		0.1252	0.1198	0.1136	0.2255	0.0538	−0.2425
			(0.0139)	(0.0139)	(0.0140)	(0.0134)	(0.0141)	(0.0133)

Table 5: Instruments added sequentially, Model B, $M = 20$, $N = 100$

		DN2SLS	S-BIC	S-AIC	S-HQ	MA-MSC	MA-$RMSC$	MA-$CCIC$
$\sigma_{\varepsilon u} = 0.1$, $R_f^2 = 0.1$	MB	0.0632	0.0248	0.0247	0.0250	0.0411	0.0456	0.0447
	$[RMB]$		[0.3928]	[0.3905]	[0.3958]	[0.6495]	[0.7208]	[0.7069]
	MAD	0.1308	0.2163	0.2186	0.2203	0.1889	0.1348	0.1107
	$[RMAD]$		[1.6542]	[1.6714]	[1.6848]	[1.4443]	[1.0309]	[0.8463]
	RMD		1.4014 (0.3168)	1.4268 (0.3022)	1.4528 (0.2867)	0.3575 (0.4326)	0.4303 (0.1358)	0.5154 (0.3747)
	KW		3.9228	3.7804	3.6382	9.3309	12.5435	11.3680
	$\rho_{RMD,\lvert bias\rvert}$		0.2459 (0.0133)	0.2455 (0.0133)	0.2455 (0.0133)	0.0803 (0.0141)	0.1037 (0.0140)	−0.0499 (0.0141)
	$\rho_{KW,\lvert bias\rvert}$		−0.2433 (0.0133)	−0.2437 (0.0133)	−0.2443 (0.0133)	−0.1879 (0.0136)	0.1030 (0.0140)	0.0274 (0.0141)
$R_f^2 = 0.01$	MB	0.0968	0.0873	0.0881	0.0890	0.1151	0.0953	0.0829
	$[RMB]$		[0.9022]	[0.9097]	[0.9192]	[1.1887]	[0.9846]	[0.8564]
	MAD	0.1610	0.3570	0.3614	0.3661	0.2939	0.1668	0.1267
	$[RMAD]$		[2.2176]	[2.2452]	[2.2742]	[1.8256]	[1.0364]	[0.7870]
	RMD		1.4064 (0.3116)	1.4315 (0.2972)	1.4573 (0.2820)	0.4192 (0.4412)	0.6329 (0.1702)	0.6657 (0.3807)
	KW		3.8952	3.7548	3.6146	9.1041	13.5043	11.7642
	$\rho_{RMD,\lvert bias\rvert}$		0.2744 (0.0131)	0.2738 (0.0131)	0.2735 (0.0131)	0.1500 (0.0138)	0.0752 (0.0141)	−0.0692 (0.0141)
	$\rho_{KW,\lvert bias\rvert}$		−0.2724 (0.0131)	−0.2724 (0.0131)	−0.2726 (0.0131)	−0.2848 (0.0130)	0.0731 (0.0141)	0.0331 (0.0141)
$\sigma_{\varepsilon u} = 0.5$, $R_f^2 = 0.1$	MB	0.3092	0.1568	0.1549	0.1523	0.2234	0.2696	0.2568
	$[RMB]$		[0.5070]	[0.5010]	[0.4924]	[0.7223]	[0.8720]	[0.8305]
	MAD	0.3220	0.2420	0.2430	0.2443	0.2668	0.2711	0.2578
	$[RMAD]$		[0.7513]	[0.7547]	[0.7587]	[0.8285]	[0.8420]	[0.8005]
	RMD		1.3026 (0.3498)	1.3352 (0.3337)	1.3687 (0.3161)	0.2561 (0.4029)	0.4547 (0.1450)	0.1917 (0.1618)
	KW		4.4105	4.2268	4.0442	9.4525	12.6587	11.1865
	$\rho_{RMD,\lvert bias\rvert}$		0.1456 (0.0138)	0.1483 (0.0138)	0.1514 (0.0138)	−0.0136 (0.0141)	−0.0476 (0.0141)	−0.5408 (0.0100)
	$\rho_{KW,\lvert bias\rvert}$		−0.1435 (0.0139)	−0.1469 (0.0138)	−0.1506 (0.0138)	−0.0309 (0.0141)	−0.0468 (0.0141)	−0.4538 (0.0112)
$R_f^2 = 0.01$	MB	0.4772	0.4345	0.4335	0.4324	0.4536	0.4691	0.4623
	$[RMB]$		[0.9105]	[0.9086]	[0.9063]	[0.9506]	[0.9830]	[0.9688]
	MAD	0.4803	0.4879	0.4866	0.4880	0.4782	0.4691	0.4623
	$[RMAD]$		[1.0158]	[1.0131]	[1.0160]	[0.9957]	[0.9766]	[0.9625]
	RMD		1.1990 (0.3568)	1.2367 (0.3418)	1.2758 (0.3246)	0.1713 (0.3165)	0.6344 (0.1721)	0.1978 (0.1356)
	KW		4.9304	4.7135	4.4972	9.8563	13.5115	11.1315
	$\rho_{RMD,\lvert bias\rvert}$		0.1637 (0.0138)	0.1647 (0.0138)	0.1661 (0.0138)	−0.0768 (0.0141)	0.0308 (0.0141)	−0.1666 (0.0138)
	$\rho_{KW,\lvert bias\rvert}$		−0.1617 (0.0138)	−0.1635 (0.0138)	−0.1655 (0.0138)	0.0129 (0.0141)	0.0311 (0.0141)	−0.1095 (0.0140)
$\sigma_{\varepsilon u} = 0.9$, $R_f^2 = 0.1$	MB	0.3572	0.3223	0.3152	0.3079	0.4034	0.4934	0.4438
	$[RMB]$		[0.9023]	[0.8822]	[0.8620]	[1.1293]	[1.3812]	[1.2424]
	MAD	0.3643	0.3420	0.3346	0.3292	0.4104	0.4934	0.4438
	$[RMAD]$		[0.9386]	[0.9183]	[0.9036]	[1.1265]	[1.3543]	[1.2181]
	RMD		0.8210 (0.3977)	0.8798 (0.3965)	0.9459 (0.3884)	0.1510 (0.2566)	0.5910 (0.1804)	0.1757 (0.0790)
	KW		6.9432	6.5387	6.1332	9.8078	13.3037	11.2875
	$\rho_{RMD,\lvert bias\rvert}$		−0.0727 (0.0141)	−0.0617 (0.0141)	−0.0482 (0.0141)	−0.2401 (0.0133)	0.0377 (0.0141)	−0.6100 (0.0089)
	$\rho_{KW,\lvert bias\rvert}$		0.0938 (0.0140)	0.0747 (0.0141)	0.0552 (0.0141)	0.2369 (0.0133)	0.0379 (0.0141)	−0.5864 (0.0093)
$R_f^2 = 0.01$	MB	0.8597	0.8129	0.8090	0.8048	0.8136	0.8427	0.8212
	$[RMB]$		[0.9456]	[0.9411]	[0.9362]	[0.9464]	[0.9803]	[0.9552]
	MAD	0.8600	0.8133	0.8096	0.8052	0.8140	0.8427	0.8212
	$[RMAD]$		[0.9457]	[0.9415]	[0.9363]	[0.9465]	[0.9799]	[0.9549]
	RMD		0.3070 (0.1261)	0.2921 (0.1462)	0.3045 (0.1784)	0.0250 (0.0670)	0.6473 (0.1746)	0.0824 (0.0601)
	KW		10.7615	10.2557	9.7167	10.3960	13.5725	10.8241
	$\rho_{RMD,\lvert bias\rvert}$		−0.0559 (0.0141)	−0.0963 (0.0140)	−0.1110 (0.0140)	−0.2494 (0.0133)	−0.0111 (0.0141)	−0.3077 (0.0128)
	$\rho_{KW,\lvert bias\rvert}$		0.1336 (0.0139)	0.1269 (0.0139)	0.1190 (0.0139)	0.2377 (0.0133)	−0.0103 (0.0141)	−0.2903 (0.0130)

Table 6: Instruments added sequentially, Model B, $M = 30$, $N = 1000$

		DN2SLS	S-BIC	S-AIC	S-HQ	MA-MSC	MA-$RMSC$	MA-$CCIC$
$\sigma_{\varepsilon u} = 0.1$, $R_f^2 = 0.1$	MB	0.0214	0.0033	0.0033	0.0034	0.0112	0.0143	0.0163
	$[RMB]$		[0.1545]	[0.1546]	[0.1604]	[0.5230]	[0.6674]	[0.7593]
	MAD	0.0596	0.0862	0.0864	0.0865	0.0604	0.0595	0.0561
	$[RMAD]$		[1.4468]	[1.4496]	[1.4520]	[1.0131]	[0.9976]	[0.9415]
	RMD		1.6683	1.6906	1.6833	0.1727	0.2280	0.1739
			(0.1734)	(0.1584)	(0.1633)	(0.3341)	(0.0375)	(0.1735)
	KW		3.4205	3.2487	3.3045	14.5864	17.1526	16.0212
	$\rho_{RMD,\lvert bias\rvert}$		0.2904	0.2968	0.2947	0.3734	0.0258	0.1895
			(0.0205)	(0.0204)	(0.0204)	(0.0192)	(0.0224)	(0.0216)
	$\rho_{KW,\lvert bias\rvert}$		−0.2901	−0.2968	−0.2946	−0.3814	0.0258	0.0733
			(0.0205)	(0.0204)	(0.0204)	(0.0191)	(0.0224)	(0.0222)
$R_f^2 = 0.01$	MB	0.0791	0.0368	0.0368	0.0369	0.0574	0.0622	0.0670
	$[RMB]$		[0.4650]	[0.4656]	[0.4665]	[0.7262]	[0.7861]	[0.8477]
	MAD	0.1203	0.2368	0.2401	0.2385	0.1876	0.1211	0.1054
	$[RMAD]$		[1.9696]	[1.9966]	[1.9835]	[1.5597]	[1.0074]	[0.8768]
	RMD		1.6651	1.6876	1.6803	0.3550	0.4894	0.4347
			(0.1741)	(0.1584)	(0.1635)	(0.4269)	(0.1192)	(0.2572)
	KW		3.4439	3.2703	3.3267	13.7308	19.0477	16.8289
	$\rho_{RMD,\lvert bias\rvert}$		0.2742	0.2759	0.2753	0.1964	0.1351	0.2662
			(0.0207)	(0.0207)	(0.0207)	(0.0215)	(0.0220)	(0.0208)
	$\rho_{KW,\lvert bias\rvert}$		−0.2743	−0.2760	−0.2754	−0.3047	0.1345	0.0343
			(0.0207)	(0.0207)	(0.0207)	(0.0203)	(0.0220)	(0.0223)
$\sigma_{\varepsilon u} = 0.5$, $R_f^2 = 0.1$	MB	0.0511	0.0204	0.0206	0.0203	0.0598	0.0698	0.0694
	$[RMB]$		[0.3995]	[0.4025]	[0.3979]	[1.1707]	[1.3661]	[1.3580]
	MAD	0.0734	0.0882	0.0889	0.0883	0.0816	0.0778	0.0756
	$[RMAD]$		[1.2015]	[1.2098]	[1.2021]	[1.1104]	[1.0596]	[1.0287]
	RMD		1.6569	1.6808	1.6729	0.1914	0.2519	0.1628
			(0.1776)	(0.1611)	(0.1664)	(0.3389)	(0.0487)	(0.1188)
	KW		3.5042	3.3203	3.3799	14.3037	17.3259	16.3900
	$\rho_{RMD,\lvert bias\rvert}$		0.2600	0.2687	0.2658	0.2624	−0.1296	−0.3257
			(0.0209)	(0.0208)	(0.0208)	(0.0208)	(0.0220)	(0.0200)
	$\rho_{KW,\lvert bias\rvert}$		−0.2594	−0.2685	−0.2655	−0.2837	−0.1293	−0.2582
			(0.0209)	(0.0208)	(0.0208)	(0.0206)	(0.0220)	(0.0209)
$R_f^2 = 0.01$	MB	0.3801	0.1829	0.1804	0.1810	0.2738	0.3271	0.3092
	$[RMB]$		[0.4812]	[0.4745]	[0.4762]	[0.7202]	[0.8606]	[0.8135]
	MAD	0.3837	0.2781	0.2780	0.2777	0.2948	0.3274	0.3094
	$[RMAD]$		[0.7247]	[0.7245]	[0.7239]	[0.7684]	[0.8532]	[0.8063]
	RMD		1.6050	1.6366	1.6263	0.2006	0.5152	0.1594
			(0.2012)	(0.1807)	(0.1873)	(0.3493)	(0.1294)	(0.1130)
	KW		3.8830	3.6416	3.7193	14.2345	19.2340	16.3722
	$\rho_{RMD,\lvert bias\rvert}$		0.2140	0.2233	0.2201	−0.0932	−0.0672	−0.6127
			(0.0213)	(0.0213)	(0.0213)	(0.0222)	(0.0223)	(0.0140)
	$\rho_{KW,\lvert bias\rvert}$		−0.2140	−0.2233	−0.2202	0.0525	−0.0664	−0.4971
			(0.0213)	(0.0213)	(0.0213)	(0.0223)	(0.0223)	(0.0168)
$\sigma_{\varepsilon u} = 0.9$, $R_f^2 = 0.1$	MB	0.0706	0.0392	0.0378	0.0381	0.1021	0.1262	0.1184
	$[RMB]$		[0.5556]	[0.5355]	[0.5397]	[1.4460]	[1.7876]	[1.6763]
	MAD	0.0866	0.0959	0.0961	0.0957	0.1131	0.1263	0.1185
	$[RMAD]$		[1.1066]	[1.1099]	[1.1050]	[1.3051]	[1.4579]	[1.3677]
	RMD		1.6335	1.6616	1.6525	0.1989	0.3218	0.1806
			(0.2001)	(0.1800)	(0.1865)	(0.2968)	(0.0728)	(0.0783)
	KW		3.6745	3.4592	3.5285	14.1448	17.8318	16.6923
	$\rho_{RMD,\lvert bias\rvert}$		0.2466	0.2589	0.2549	−0.1111	−0.1386	−0.5699
			(0.0210)	(0.0209)	(0.0209)	(0.0221)	(0.0219)	(0.0151)
	$\rho_{KW,\lvert bias\rvert}$		−0.2464	−0.2589	−0.2549	0.1004	−0.1386	−0.5057
			(0.0210)	(0.0209)	(0.0209)	(0.0221)	(0.0219)	(0.0166)
$R_f^2 = 0.01$	MB	0.4231	0.3737	0.3586	0.3622	0.5063	0.5961	0.5398
	$[RMB]$		[0.8831]	[0.8475]	[0.8561]	[1.1967]	[1.4088]	[1.2758]
	MAD	0.4277	0.3893	0.3802	0.3829	0.5076	0.5961	0.5398
	$[RMAD]$		[0.9104]	[0.8890]	[0.8954]	[1.1868]	[1.3937]	[1.2622]
	RMD		1.1822	1.2993	1.2619	0.0877	0.6511	0.1575
			(0.3297)	(0.2860)	(0.3007)	(0.1690)	(0.1585)	(0.0636)
	KW		6.9807	6.0929	6.3725	14.8899	20.2167	16.5780
	$\rho_{RMD,\lvert bias\rvert}$		0.0124	0.0431	0.0329	−0.3644	0.0085	−0.5959
			(0.0224)	(0.0223)	(0.0223)	(0.0194)	(0.0224)	(0.0144)
	$\rho_{KW,\lvert bias\rvert}$		−0.0110	−0.0430	−0.0326	0.3620	0.0091	−0.5693
			(0.0224)	(0.0223)	(0.0223)	(0.0194)	(0.0224)	(0.0151)

Table 7: Instruments added sequentially, Model C, $M = 10$, $N = 100$

		DN2SLS	S-BIC	S-AIC	S-HQ	MA-MSC	MA-$RMSC$	MA-$CCIC$
$\sigma_{\varepsilon u} = 0.1$, $R_f^2 = 0.1$	MB	0.0468	0.0127	0.0126	0.0126	0.0347	0.0351	0.0375
	$[RMB]$		[0.2709]	[0.2682]	[0.2690]	[0.7400]	[0.7494]	[0.8003]
	MAD	0.1515	0.1958	0.1967	0.1980	0.1985	0.1566	0.1234
	$[RMAD]$		[1.2924]	[1.2985]	[1.3068]	[1.3102]	[1.0339]	[0.8143]
	RMD		1.0037	1.0279	1.0542	0.3367	0.3339	0.5441
			(0.4254)	(0.4193)	(0.4115)	(0.4347)	(0.1540)	(0.4029)
	KW		3.3099	3.2377	3.1628	4.9934	6.2503	5.8221
	$\rho_{RMD,\lvert bias\rvert}$		0.2290	0.2278	0.2267	−0.0620	0.1181	−0.0501
			(0.0134)	(0.0134)	(0.0134)	(0.0141)	(0.0139)	(0.0141)
	$\rho_{KW,\lvert bias\rvert}$		−0.2253	−0.2250	−0.2247	−0.0614	0.1142	0.0456
			(0.0134)	(0.0134)	(0.0134)	(0.0141)	(0.0140)	(0.0141)
$R_f^2 = 0.01$	MB	0.0791	0.0616	0.0602	0.0600	0.0990	0.0841	0.0689
	$[RMB]$		[0.7786]	[0.7602]	[0.7584]	[1.2509]	[1.0630]	[0.8706]
	MAD	0.2018	0.3587	0.3623	0.3662	0.3595	0.2177	0.1504
	$[RMAD]$		[1.7775]	[1.7955]	[1.8147]	[1.7815]	[1.0791]	[0.7453]
	RMD		1.0034	1.0274	1.0537	0.4551	0.6037	0.7840
			(0.4299)	(0.4238)	(0.4159)	(0.4652)	(0.2500)	(0.4294)
	KW		3.3121	3.2398	3.1647	4.7705	6.8541	6.1073
	$\rho_{RMD,\lvert bias\rvert}$		0.2737	0.2712	0.2688	−0.0034	0.0870	−0.1401
			(0.0131)	(0.0131)	(0.0131)	(0.0141)	(0.0140)	(0.0139)
	$\rho_{KW,\lvert bias\rvert}$		−0.2711	−0.2694	−0.2677	−0.1255	0.0795	0.0365
			(0.0131)	(0.0131)	(0.0131)	(0.0139)	(0.0141)	(0.0141)
$\sigma_{\varepsilon u} = 0.5$, $R_f^2 = 0.1$	MB	0.2278	0.1121	0.1110	0.1102	0.1406	0.1889	0.1819
	$[RMB]$		[0.4922]	[0.4874]	[0.4835]	[0.6174]	[0.8293]	[0.7987]
	MAD	0.2592	0.2138	0.2151	0.2154	0.2319	0.2104	0.1965
	$[RMAD]$		[0.8249]	[0.8300]	[0.8311]	[0.8947]	[0.8118]	[0.7583]
	RMD		0.9138	0.9400	0.9690	0.3083	0.3685	0.2490
			(0.4331)	(0.4284)	(0.4218)	(0.4402)	(0.1819)	(0.2356)
	KW		3.5267	3.4456	3.3614	4.9349	6.3269	5.8753
	$\rho_{RMD,\lvert bias\rvert}$		0.1806	0.1805	0.1806	0.0741	−0.0055	−0.4178
			(0.0137)	(0.0137)	(0.0137)	(0.0141)	(0.0141)	(0.0117)
	$\rho_{KW,\lvert bias\rvert}$		−0.1741	−0.1757	−0.1774	−0.1420	−0.0038	−0.2875
			(0.0137)	(0.0137)	(0.0137)	(0.0139)	(0.0141)	(0.0130)
$R_f^2 = 0.01$	MB	0.4634	0.3913	0.3903	0.3891	0.4090	0.4403	0.4293
	$[RMB]$		[0.8444]	[0.8422]	[0.8397]	[0.8824]	[0.9501]	[0.9264]
	MAD	0.4795	0.4573	0.4584	0.4583	0.4705	0.4426	0.4314
	$[RMAD]$		[0.9537]	[0.9560]	[0.9559]	[0.9813]	[0.9230]	[0.8997]
	RMD		0.8127	0.8403	0.8711	0.2614	0.5981	0.2728
			(0.4057)	(0.4031)	(0.3984)	(0.3940)	(0.2473)	(0.2180)
	KW		3.7677	3.6798	3.5881	5.0403	6.8416	5.8612
	$\rho_{RMD,\lvert bias\rvert}$		0.2059	0.2048	0.2038	−0.0813	0.0592	−0.0657
			(0.0135)	(0.0136)	(0.0136)	(0.0141)	(0.0141)	(0.0141)
	$\rho_{KW,\lvert bias\rvert}$		−0.2019	−0.2019	−0.2018	0.0046	0.0582	−0.0386
			(0.0136)	(0.0136)	(0.0136)	(0.0141)	(0.0141)	(0.0141)
$\sigma_{\varepsilon u} = 0.9$, $R_f^2 = 0.1$	MB	0.2539	0.2287	0.2260	0.2223	0.2789	0.3533	0.3146
	$[RMB]$		[0.9009]	[0.8900]	[0.8756]	[1.0987]	[1.3916]	[1.2389]
	MAD	0.2785	0.2671	0.2636	0.2628	0.3079	0.3535	0.3151
	$[RMAD]$		[0.9590]	[0.9465]	[0.9435]	[1.1055]	[1.2692]	[1.1313]
	RMD		0.6520	0.6822	0.7172	0.2494	0.5016	0.2107
			(0.3799)	(0.3827)	(0.3829)	(0.3939)	(0.2314)	(0.1289)
	KW		4.1665	4.0606	3.9500	4.9686	6.6245	5.9406
	$\rho_{RMD,\lvert bias\rvert}$		0.0617	0.0628	0.0652	0.0409	0.0122	−0.5832
			(0.0141)	(0.0141)	(0.0141)	(0.0141)	(0.0141)	(0.0093)
	$\rho_{KW,\lvert bias\rvert}$		−0.0484	−0.0542	−0.0600	−0.0573	0.0127	−0.5501
			(0.0141)	(0.0141)	(0.0141)	(0.0141)	(0.0141)	(0.0099)
$R_f^2 = 0.01$	MB	0.8233	0.7451	0.7422	0.7394	0.7455	0.7924	0.7608
	$[RMB]$		[0.9050]	[0.9015]	[0.8981]	[0.9055]	[0.9625]	[0.9242]
	MAD	0.8248	0.7456	0.7426	0.7402	0.7510	0.7924	0.7608
	$[RMAD]$		[0.9039]	[0.9003]	[0.8974]	[0.9105]	[0.9607]	[0.9224]
	RMD		0.2315	0.2277	0.2341	0.0705	0.6473	0.1223
			(0.1249)	(0.1400)	(0.1578)	(0.1826)	(0.2601)	(0.0934)
	KW		5.4690	5.3431	5.2089	5.3560	6.9514	5.7231
	$\rho_{RMD,\lvert bias\rvert}$		0.0030	−0.0153	−0.0283	−0.1835	0.0268	−0.2619
			(0.0141)	(0.0141)	(0.0141)	(0.0137)	(0.0141)	(0.0132)
	$\rho_{KW,\lvert bias\rvert}$		0.0666	0.0615	0.0558	0.1685	0.0290	−0.2465
			(0.0141)	(0.0141)	(0.0141)	(0.0137)	(0.0141)	(0.0133)

Table 8: Instruments added sequentially, Model C, $M = 20$, $N = 100$

		DN2SLS	S-BIC	S-AIC	S-HQ	MA-MSC	MA-$RMSC$	MA-$CCIC$
$\sigma_{\varepsilon u} = 0.1$, $R_f^2 = 0.1$	MB	0.0654	0.0257	0.0262	0.0263	0.0490	0.0537	0.0498
	$[RMB]$		[0.3924]	[0.4010]	[0.4020]	[0.7496]	[0.8208]	[0.7612]
	MAD	0.1286	0.2210	0.2227	0.2245	0.1890	0.1375	0.1105
	$[RMAD]$		[1.7187]	[1.7325]	[1.7461]	[1.4700]	[1.0697]	[0.8598]
	RMD		1.4044 (0.3215)	1.4294 (0.3072)	1.4553 (0.2918)	0.3478 (0.4274)	0.4331 (0.1380)	0.5131 (0.3714)
	KW		3.9102	3.7684	3.6268	9.3615	12.5566	11.3549
	$\rho_{RMD,\lvert bias\rvert}$		0.2281 (0.0134)	0.2281 (0.0134)	0.2285 (0.0134)	0.1381 (0.0139)	0.1166 (0.0140)	−0.0197 (0.0141)
	$\rho_{KW,\lvert bias\rvert}$		−0.2264 (0.0134)	−0.2271 (0.0134)	−0.2280 (0.0134)	−0.2461 (0.0133)	0.1158 (0.0140)	0.0346 (0.0141)
$R_f^2 = 0.01$	MB	0.0939	0.0678	0.0669	0.0674	0.1038	0.0889	0.0762
	$[RMB]$		[0.7224]	[0.7126]	[0.7178]	[1.1048]	[0.9468]	[0.8116]
	MAD	0.1639	0.3528	0.3564	0.3619	0.2995	0.1718	0.1300
	$[RMAD]$		[2.1522]	[2.1739]	[2.2077]	[1.8271]	[1.0481]	[0.7929]
	RMD		1.4046 (0.3136)	1.4298 (0.2996)	1.4557 (0.2845)	0.4177 (0.4472)	0.6341 (0.1700)	0.6760 (0.3855)
	KW		3.9060	3.7646	3.6234	9.1011	13.5101	11.7701
	$\rho_{RMD,\lvert bias\rvert}$		0.2575 (0.0132)	0.2557 (0.0132)	0.2544 (0.0132)	0.1294 (0.0139)	0.0778 (0.0141)	−0.1013 (0.0140)
	$\rho_{KW,\lvert bias\rvert}$		−0.2555 (0.0132)	−0.2546 (0.0132)	−0.2538 (0.0132)	−0.2507 (0.0133)	0.0762 (0.0141)	0.0262 (0.0141)
$\sigma_{\varepsilon u} = 0.5$, $R_f^2 = 0.1$	MB	0.3070	0.1441	0.1415	0.1386	0.2119	0.2650	0.2513
	$[RMB]$		[0.4694]	[0.4608]	[0.4514]	[0.6901]	[0.8630]	[0.8186]
	MAD	0.3190	0.2367	0.2365	0.2372	0.2564	0.2666	0.2524
	$[RMAD]$		[0.7423]	[0.7415]	[0.7437]	[0.8038]	[0.8359]	[0.7913]
	RMD		1.2994 (0.3484)	1.3323 (0.3328)	1.3662 (0.3155)	0.2573 (0.4012)	0.4528 (0.1440)	0.1918 (0.1565)
	KW		4.4289	4.2427	4.0575	9.4656	12.6494	11.1706
	$\rho_{RMD,\lvert bias\rvert}$		0.1610 (0.0138)	0.1647 (0.0138)	0.1686 (0.0137)	0.0041 (0.0141)	−0.0204 (0.0141)	−0.5410 (0.0100)
	$\rho_{KW,\lvert bias\rvert}$		−0.1583 (0.0138)	−0.1630 (0.0138)	−0.1678 (0.0137)	−0.0588 (0.0141)	−0.0194 (0.0141)	−0.4198 (0.0117)
$R_f^2 = 0.01$	MB	0.4767	0.4320	0.4280	0.4290	0.4501	0.4677	0.4589
	$[RMB]$		[0.9062]	[0.8979]	[0.9000]	[0.9443]	[0.9812]	[0.9627]
	MAD	0.4811	0.4839	0.4859	0.4877	0.4724	0.4677	0.4589
	$[RMAD]$		[1.0058]	[1.0099]	[1.0136]	[0.9817]	[0.9721]	[0.9538]
	RMD		1.2086 (0.3567)	1.2459 (0.3417)	1.2845 (0.3244)	0.1733 (0.3239)	0.6299 (0.1731)	0.1960 (0.1358)
	KW		4.8806	4.6668	4.4537	9.8434	13.4903	11.1317
	$\rho_{RMD,\lvert bias\rvert}$		0.1578 (0.0138)	0.1585 (0.0138)	0.1595 (0.0138)	−0.0758 (0.0141)	0.0250 (0.0141)	−0.1911 (0.0136)
	$\rho_{KW,\lvert bias\rvert}$		−0.1552 (0.0138)	−0.1567 (0.0138)	−0.1585 (0.0138)	0.0124 (0.0141)	0.0252 (0.0141)	−0.1478 (0.0138)
$\sigma_{\varepsilon u} = 0.9$, $R_f^2 = 0.1$	MB	0.3594	0.3220	0.3139	0.3060	0.4000	0.4962	0.4430
	$[RMB]$		[0.8959]	[0.8733]	[0.8513]	[1.1128]	[1.3806]	[1.2326]
	MAD	0.3674	0.3407	0.3342	0.3277	0.4077	0.4962	0.4430
	$[RMAD]$		[0.9274]	[0.9098]	[0.8921]	[1.1097]	[1.3506]	[1.2058]
	RMD		0.8133 (0.3889)	0.8730 (0.3876)	0.9400 (0.3797)	0.1522 (0.2555)	0.5946 (0.1804)	0.1767 (0.0783)
	KW		6.9769	6.5689	6.1599	9.8044	13.3210	11.2919
	$\rho_{RMD,\lvert bias\rvert}$		−0.0380 (0.0141)	−0.0297 (0.0141)	−0.0198 (0.0141)	−0.2277 (0.0134)	0.0710 (0.0141)	−0.5924 (0.0092)
	$\rho_{KW,\lvert bias\rvert}$		0.0533 (0.0141)	0.0390 (0.0141)	0.0249 (0.0141)	0.2237 (0.0134)	0.0713 (0.0141)	−0.5671 (0.0096)
$R_f^2 = 0.01$	MB	0.8576	0.8105	0.8069	0.8033	0.8108	0.8372	0.8176
	$[RMB]$		[0.9450]	[0.9409]	[0.9367]	[0.9454]	[0.9761]	[0.9533]
	MAD	0.8578	0.8106	0.8071	0.8036	0.8111	0.8372	0.8176
	$[RMAD]$		[0.9450]	[0.9410]	[0.9368]	[0.9456]	[0.9760]	[0.9532]
	RMD		0.3103 (0.1278)	0.2965 (0.1490)	0.3098 (0.1816)	0.0267 (0.0719)	0.6555 (0.1762)	0.0845 (0.0620)
	KW		10.7308	10.2241	9.6845	10.3882	13.6115	10.8325
	$\rho_{RMD,\lvert bias\rvert}$		−0.0232 (0.0141)	−0.0695 (0.0141)	−0.0945 (0.0140)	−0.3554 (0.0124)	0.0131 (0.0141)	−0.3300 (0.0126)
	$\rho_{KW,\lvert bias\rvert}$		0.1253 (0.0139)	0.1200 (0.0139)	0.1138 (0.0140)	0.3449 (0.0125)	0.0140 (0.0141)	−0.3166 (0.0127)

Table 9: Instruments added sequentially, Model C, $M = 30$, $N = 1000$

		DN2SLS	S-BIC	S-AIC	S-HQ	MA-MSC	MA-$RMSC$	MA-$CCIC$
$\sigma_{\varepsilon u} = 0.1$, $R_f^2 = 0.1$	MB	0.0240	0.0055	0.0050	0.0053	0.0134	0.0140	0.0173
	$[RMB]$		[0.2297]	[0.2094]	[0.2221]	[0.5611]	[0.5830]	[0.7218]
	MAD	0.0603	0.0885	0.0889	0.0889	0.0634	0.0599	0.0571
	$[RMAD]$		[1.4675]	[1.4744]	[1.4741]	[1.0518]	[0.9930]	[0.9461]
	RMD		1.6711	1.6930	1.6859	0.1689	0.2274	0.1736
			(0.1778)	(0.1620)	(0.1671)	(0.3215)	(0.0374)	(0.1808)
	KW		3.4002	3.2310	3.2859	14.6383	17.1489	15.9526
	$\rho_{RMD,\lvert bias\rvert}$		0.2672	0.2742	0.2718	0.3317	0.0744	0.1865
			(0.0208)	(0.0207)	(0.0207)	(0.0199)	(0.0222)	(0.0216)
	$\rho_{KW,\lvert bias\rvert}$		−0.2670	−0.2741	−0.2717	−0.3402	0.0744	0.0200
			(0.0208)	(0.0207)	(0.0207)	(0.0198)	(0.0222)	(0.0224)
$R_f^2 = 0.01$	MB	0.0783	0.0260	0.0256	0.0250	0.0567	0.0645	0.0674
	$[RMB]$		[0.3327]	[0.3275]	[0.3191]	[0.7249]	[0.8236]	[0.8609]
	MAD	0.1206	0.2505	0.2555	0.2539	0.1884	0.1231	0.1055
	$[RMAD]$		[2.0774]	[2.1192]	[2.1058]	[1.5628]	[1.0213]	[0.8749]
	RMD		1.6673	1.6896	1.6823	0.3724	0.4974	0.4366
			(0.1775)	(0.1621)	(0.1671)	(0.4436)	(0.1229)	(0.2498)
	KW		3.4281	3.2560	3.3119	13.5382	19.1059	16.8190
	$\rho_{RMD,\lvert bias\rvert}$		0.2794	0.2789	0.2791	0.2419	0.0333	0.2261
			(0.0206)	(0.0206)	(0.0206)	(0.0211)	(0.0223)	(0.0212)
	$\rho_{KW,\lvert bias\rvert}$		−0.2789	−0.2788	−0.2788	−0.3340	0.0327	−0.0271
			(0.0206)	(0.0206)	(0.0206)	(0.0199)	(0.0223)	(0.0223)
$\sigma_{\varepsilon u} = 0.5$, $R_f^2 = 0.1$	MB	0.0493	0.0208	0.0199	0.0203	0.0551	0.0682	0.0673
	$[RMB]$		[0.4220]	[0.4040]	[0.4112]	[1.1185]	[1.3843]	[1.3666]
	MAD	0.0727	0.0899	0.0911	0.0906	0.0816	0.0781	0.0759
	$[RMAD]$		[1.2371]	[1.2538]	[1.2465]	[1.1230]	[1.0743]	[1.0438]
	RMD		1.6640	1.6874	1.6797	0.1986	0.2515	0.1661
			(0.1777)	(0.1616)	(0.1668)	(0.3429)	(0.0498)	(0.1217)
	KW		3.4526	3.2725	3.3308	14.2560	17.3228	16.4009
	$\rho_{RMD,\lvert bias\rvert}$		0.3045	0.3118	0.3094	0.2477	−0.0914	−0.2692
			(0.0203)	(0.0202)	(0.0202)	(0.0210)	(0.0222)	(0.0207)
	$\rho_{KW,\lvert bias\rvert}$		−0.3044	−0.3118	−0.3094	−0.2690	−0.0911	−0.2017
			(0.0203)	(0.0202)	(0.0202)	(0.0207)	(0.0222)	(0.0215)
$R_f^2 = 0.01$	MB	0.3800	0.1960	0.1927	0.1943	0.2808	0.3277	0.3122
	$[RMB]$		[0.5159]	[0.5071]	[0.5112]	[0.7390]	[0.8624]	[0.8216]
	MAD	0.3830	0.2813	0.2831	0.2811	0.3009	0.3277	0.3122
	$[RMAD]$		[0.7344]	[0.7391]	[0.7339]	[0.7855]	[0.8555]	[0.8151]
	RMD		1.6080	1.6390	1.6290	0.2002	0.5079	0.1573
			(0.1963)	(0.1765)	(0.1829)	(0.3572)	(0.1264)	(0.1091)
	KW		3.8604	3.6232	3.6996	14.2343	19.1815	16.3625
	$\rho_{RMD,\lvert bias\rvert}$		0.1645	0.1682	0.1670	−0.0999	−0.0927	−0.6579
			(0.0218)	(0.0217)	(0.0217)	(0.0221)	(0.0222)	(0.0127)
	$\rho_{KW,\lvert bias\rvert}$		−0.1642	−0.1682	−0.1669	0.0694	−0.0919	−0.5452
			(0.0218)	(0.0217)	(0.0217)	(0.0223)	(0.0222)	(0.0157)
$\sigma_{\varepsilon u} = 0.9$, $R_f^2 = 0.1$	MB	0.0689	0.0368	0.0364	0.0366	0.0992	0.1256	0.1167
	$[RMB]$		[0.5349]	[0.5290]	[0.5314]	[1.4412]	[1.8238]	[1.6955]
	MAD	0.0838	0.0926	0.0927	0.0925	0.1099	0.1260	0.1172
	$[RMAD]$		[1.1050]	[1.1053]	[1.1030]	[1.3112]	[1.5037]	[1.3977]
	RMD		1.6338	1.6619	1.6528	0.1932	0.3224	0.1820
			(0.1956)	(0.1760)	(0.1823)	(0.2822)	(0.0726)	(0.0724)
	KW		3.6717	3.4573	3.5263	14.1835	17.8363	16.7046
	$\rho_{RMD,\lvert bias\rvert}$		0.2805	0.2961	0.2910	−0.0780	−0.1300	−0.6134
			(0.0206)	(0.0204)	(0.0205)	(0.0222)	(0.0220)	(0.0140)
	$\rho_{KW,\lvert bias\rvert}$		−0.2805	−0.2962	−0.2911	0.0634	−0.1299	−0.5334
			(0.0206)	(0.0204)	(0.0205)	(0.0223)	(0.0220)	(0.0160)
$R_f^2 = 0.01$	MB	0.4141	0.3623	0.3472	0.3510	0.4967	0.5950	0.5332
	$[RMB]$		[0.8749]	[0.8384]	[0.8476]	[1.1994]	[1.4370]	[1.2876]
	MAD	0.4169	0.3783	0.3687	0.3719	0.4986	0.5950	0.5332
	$[RMAD]$		[0.9075]	[0.8843]	[0.8921]	[1.1960]	[1.4273]	[1.2789]
	RMD		1.1872	1.3038	1.2666	0.0918	0.6631	0.1613
			(0.3272)	(0.2831)	(0.2979)	(0.1474)	(0.1594)	(0.0640)
	KW		6.9432	6.0599	6.3380	14.8582	20.3039	16.6092
	$\rho_{RMD,\lvert bias\rvert}$		0.0023	0.0291	0.0203	−0.6135	−0.0180	−0.6334
			(0.0224)	(0.0223)	(0.0224)	(0.0139)	(0.0224)	(0.0134)
	$\rho_{KW,\lvert bias\rvert}$		−0.0015	−0.0291	−0.0201	0.6108	−0.0176	−0.6086
			(0.0224)	(0.0223)	(0.0224)	(0.0140)	(0.0224)	(0.0141)

Table 10: Instruments combined, $M = 10$, $M_{fixed} = 2$, Model A, $N = 100$

		DN2SLS	S-BIC	S-AIC	S-HQ	MA-MSC	MA-$RMSC$	MA-$CCIC$
$\sigma_{\varepsilon u} = 0.1$, $R_f^2 = 0.1$	MB	0.0600	0.0615	0.0614	0.0605	0.0877	0.0584	0.0551
	$[RMB]$		[1.0242]	[1.0225]	[1.0086]	[1.4618]	[0.9734]	[0.9179]
	MAD	0.1587	0.2106	0.2107	0.2115	0.2336	0.1589	0.1058
	$[RMAD]$		[1.3266]	[1.3275]	[1.3321]	[1.4717]	[1.0010]	[0.6663]
	RMD		0.8142 (0.3398)	0.8232 (0.3394)	0.8334 (0.3385)	0.3437 (0.3751)	0.5086 (0.1060)	0.7051 (0.3400)
	KW		97.8813	96.5682	95.1723	121.2911	150.6818	134.6008
	$\rho_{RMD,\lvert bias\rvert}$		0.2750 (0.0131)	0.2762 (0.0131)	0.2776 (0.0131)	0.1071 (0.0140)	0.0533 (0.0141)	−0.1991 (0.0136)
	$\rho_{KW,\lvert bias\rvert}$		−0.2837 (0.0130)	−0.2857 (0.0130)	−0.2878 (0.0130)	−0.2018 (0.0136)	−0.0129 (0.0141)	−0.1850 (0.0137)
$R_f^2 = 0.01$	MB	0.0856	0.0876	0.0874	0.0873	0.1267	0.0880	0.0734
	$[RMB]$		[1.0230]	[1.0214]	[1.0198]	[1.4799]	[1.0276]	[0.8574]
	MAD	0.2044	0.2785	0.2791	0.2802	0.3183	0.1997	0.1230
	$[RMAD]$		[1.3624]	[1.3654]	[1.3705]	[1.5568]	[0.9769]	[0.6017]
	RMD		0.8206 (0.3421)	0.8295 (0.3417)	0.8396 (0.3407)	0.3868 (0.3982)	0.5301 (0.1319)	0.7905 (0.3407)
	KW		97.7340	96.4232	95.0298	120.0696	150.4726	134.6758
	$\rho_{RMD,\lvert bias\rvert}$		0.2720 (0.0131)	0.2727 (0.0131)	0.2736 (0.0131)	0.0839 (0.0140)	0.0253 (0.0141)	−0.2242 (0.0134)
	$\rho_{KW,\lvert bias\rvert}$		−0.2904 (0.0130)	−0.2919 (0.0129)	−0.2934 (0.0129)	−0.1750 (0.0137)	−0.0666 (0.0141)	−0.2037 (0.0136)
$\sigma_{\varepsilon u} = 0.5$, $R_f^2 = 0.1$	MB	0.3075	0.2710	0.2713	0.2714	0.2752	0.2858	0.2972
	$[RMB]$		[0.8816]	[0.8824]	[0.8827]	[0.8950]	[0.9296]	[0.9665]
	MAD	0.3220	0.2928	0.2935	0.2936	0.3045	0.2899	0.2985
	$[RMAD]$		[0.9094]	[0.9114]	[0.9117]	[0.9455]	[0.9004]	[0.9271]
	RMD		0.7179 (0.3227)	0.7269 (0.3230)	0.7373 (0.3228)	0.1926 (0.2954)	0.5130 (0.1082)	0.2415 (0.1148)
	KW		102.7437	101.3882	99.9462	124.2605	150.4146	130.9121
	$\rho_{RMD,\lvert bias\rvert}$		0.1987 (0.0136)	0.2000 (0.0136)	0.2015 (0.0136)	−0.0848 (0.0140)	0.0222 (0.0141)	−0.4662 (0.0111)
	$\rho_{KW,\lvert bias\rvert}$		−0.1990 (0.0136)	−0.2008 (0.0136)	−0.2027 (0.0136)	−0.0427 (0.0141)	0.0611 (0.0141)	0.2847 (0.0130)
$R_f^2 = 0.01$	MB	0.4716	0.4592	0.4594	0.4591	0.4669	0.4629	0.4630
	$[RMB]$		[0.9739]	[0.9742]	[0.9737]	[0.9901]	[0.9816]	[0.9818]
	MAD	0.4854	0.4689	0.4691	0.4687	0.4826	0.4636	0.4637
	$[RMAD]$		[0.9660]	[0.9665]	[0.9657]	[0.9943]	[0.9551]	[0.9553]
	RMD		0.6636 (0.3091)	0.6723 (0.3098)	0.6825 (0.3100)	0.1659 (0.2756)	0.5311 (0.1318)	0.2576 (0.1246)
	KW		105.5362	104.1611	102.6977	124.5752	150.4621	132.2204
	$\rho_{RMD,\lvert bias\rvert}$		0.1293 (0.0139)	0.1302 (0.0139)	0.1312 (0.0139)	−0.2295 (0.0134)	0.0225 (0.0141)	−0.1264 (0.0139)
	$\rho_{KW,\lvert bias\rvert}$		−0.1311 (0.0139)	−0.1326 (0.0139)	−0.1343 (0.0139)	0.0989 (0.0140)	0.0168 (0.0141)	0.0435 (0.0141)
$\sigma_{\varepsilon u} = 0.9$, $R_f^2 = 0.1$	MB	0.5528	0.4673	0.4672	0.4677	0.4782	0.5130	0.4977
	$[RMB]$		[0.8453]	[0.8452]	[0.8461]	[0.8651]	[0.9281]	[0.9004]
	MAD	0.5635	0.4680	0.4681	0.4681	0.4797	0.5130	0.4977
	$[RMAD]$		[0.8305]	[0.8306]	[0.8307]	[0.8514]	[0.9104]	[0.8833]
	RMD		0.4634 (0.2407)	0.4703 (0.2440)	0.4791 (0.2467)	0.0497 (0.1236)	0.5403 (0.1307)	0.1396 (0.0524)
	KW		115.2822	113.8379	112.2983	126.9998	148.9728	128.6098
	$\rho_{RMD,\lvert bias\rvert}$		0.1463 (0.0138)	0.1495 (0.0138)	0.1524 (0.0138)	−0.2474 (0.0133)	−0.0375 (0.0141)	−0.4461 (0.0113)
	$\rho_{KW,\lvert bias\rvert}$		−0.2015 (0.0136)	−0.2024 (0.0136)	−0.2033 (0.0136)	0.1414 (0.0139)	0.0925 (0.0140)	0.1080 (0.0140)
$R_f^2 = 0.01$	MB	0.8462	0.8290	0.8293	0.8293	0.8286	0.8318	0.8323
	$[RMB]$		[0.9796]	[0.9800]	[0.9800]	[0.9792]	[0.9830]	[0.9836]
	MAD	0.8472	0.8290	0.8293	0.8293	0.8286	0.8318	0.8323
	$[RMAD]$		[0.9785]	[0.9789]	[0.9789]	[0.9781]	[0.9818]	[0.9824]
	RMD		0.2349 (0.1193)	0.2339 (0.1254)	0.2355 (0.1313)	0.0134 (0.0236)	0.5370 (0.1347)	0.0953 (0.0488)
	KW		126.8451	125.3597	123.7729	127.7533	150.4115	129.7454
	$\rho_{RMD,\lvert bias\rvert}$		0.0359 (0.0141)	0.0378 (0.0141)	0.0394 (0.0141)	−0.2939 (0.0129)	0.0275 (0.0141)	−0.1814 (0.0137)
	$\rho_{KW,\lvert bias\rvert}$		−0.0444 (0.0141)	−0.0445 (0.0141)	−0.0446 (0.0141)	0.2199 (0.0135)	0.0750 (0.0141)	0.1786 (0.0137)

Table 11: Instruments combined, $M = 10$, $M_{fixed} = 2$, Model B, $N = 100$

		DN2SLS	S-BIC	S-AIC	S-HQ	MA-MSC	MA-$RMSC$	MA-$CCIC$
$\sigma_{\varepsilon u} = 0.1$, $R_f^2 = 0.1$	MB [RMB]	0.0483	0.0547 [1.1329]	0.0544 [1.1265]	0.0552 [1.1432]	0.0895 [1.8548]	0.0509 [1.0548]	0.0526 [1.0908]
	MAD [$RMAD$]	0.1468	0.2045 [1.3934]	0.2056 [1.4009]	0.2063 [1.4055]	0.2348 [1.5997]	0.1466 [0.9989]	0.0998 [0.6799]
	RMD		0.8138 (0.3426)	0.8227 (0.3422)	0.8329 (0.3413)	0.3568 (0.3809)	0.5589 (0.1195)	0.6952 (0.3334)
	KW		97.9231	96.6113	95.2167	122.0180	147.2977	133.9939
	$\rho_{RMD,\lvert bias\rvert}$		0.2930 (0.0129)	0.2943 (0.0129)	0.2957 (0.0129)	0.2115 (0.0135)	−0.0086 (0.0141)	−0.1659 (0.0138)
	$\rho_{KW,\lvert bias\rvert}$		−0.2841 (0.0130)	−0.2857 (0.0130)	−0.2874 (0.0130)	−0.2807 (0.0130)	0.0202 (0.0141)	−0.2154 (0.0135)
$R_f^2 = 0.01$	MB [RMB]	0.0857	0.0907 [1.0574]	0.0907 [1.0584]	0.0912 [1.0637]	0.1273 [1.4842]	0.0842 [0.9820]	0.0720 [0.8395]
	MAD [$RMAD$]	0.2070	0.2745 [1.3259]	0.2760 [1.3331]	0.2774 [1.3402]	0.3196 [1.5438]	0.1984 [0.9586]	0.1223 [0.5907]
	RMD		0.8179 (0.3434)	0.8268 (0.3431)	0.8370 (0.3422)	0.3928 (0.4056)	0.5310 (0.1338)	0.7911 (0.3423)
	KW		97.8123	96.5012	95.1075	120.0593	149.9291	134.4852
	$\rho_{RMD,\lvert bias\rvert}$		0.2893 (0.0130)	0.2903 (0.0130)	0.2914 (0.0129)	0.1360 (0.0139)	0.0152 (0.0141)	−0.2216 (0.0134)
	$\rho_{KW,\lvert bias\rvert}$		−0.2953 (0.0129)	−0.2970 (0.0129)	−0.2988 (0.0129)	−0.2308 (0.0134)	−0.0649 (0.0141)	−0.2255 (0.0134)
$\sigma_{\varepsilon u} = 0.5$, $R_f^2 = 0.1$	MB [RMB]	0.2257	0.2832 [1.2546]	0.2834 [1.2557]	0.2839 [1.2576]	0.2866 [1.2696]	0.2583 [1.1443]	0.2891 [1.2806]
	MAD [$RMAD$]	0.2612	0.3012 [1.1528]	0.3027 [1.1585]	0.3033 [1.1610]	0.3075 [1.1770]	0.2628 [1.0061]	0.2910 [1.1139]
	RMD		0.7455 (0.3250)	0.7544 (0.3251)	0.7646 (0.3248)	0.1941 (0.2812)	0.5554 (0.1206)	0.2513 (0.1154)
	KW		102.1783	100.8328	99.4015	124.9574	147.5912	131.3710
	$\rho_{RMD,\lvert bias\rvert}$		0.2333 (0.0134)	0.2347 (0.0134)	0.2361 (0.0134)	−0.0262 (0.0141)	−0.0118 (0.0141)	−0.4353 (0.0115)
	$\rho_{KW,\lvert bias\rvert}$		−0.1729 (0.0137)	−0.1749 (0.0137)	−0.1770 (0.0137)	−0.0928 (0.0140)	0.0510 (0.0141)	0.1683 (0.0137)
$R_f^2 = 0.01$	MB [RMB]	0.4555	0.4519 [0.9921]	0.4518 [0.9919]	0.4516 [0.9914]	0.4586 [1.0067]	0.4580 [1.0055]	0.4584 [1.0063]
	MAD [$RMAD$]	0.4733	0.4605 [0.9730]	0.4615 [0.9751]	0.4623 [0.9768]	0.4754 [1.0044]	0.4584 [0.9685]	0.4587 [0.9692]
	RMD		0.6696 (0.3103)	0.6783 (0.3111)	0.6885 (0.3113)	0.1654 (0.2795)	0.5315 (0.1330)	0.2588 (0.1240)
	KW		105.5174	104.1431	102.6804	124.5427	149.9424	132.1730
	$\rho_{RMD,\lvert bias\rvert}$		0.1341 (0.0139)	0.1355 (0.0139)	0.1370 (0.0139)	−0.1962 (0.0136)	0.0183 (0.0141)	−0.1285 (0.0139)
	$\rho_{KW,\lvert bias\rvert}$		−0.1284 (0.0139)	−0.1303 (0.0139)	−0.1323 (0.0139)	0.0703 (0.0141)	0.0232 (0.0141)	−0.0018 (0.0141)
$\sigma_{\varepsilon u} = 0.9$, $R_f^2 = 0.1$	MB [RMB]	0.2518	0.5219 [2.0726]	0.5227 [2.0755]	0.5236 [2.0791]	0.4883 [1.9389]	0.4898 [1.9449]	0.5082 [2.0182]
	MAD [$RMAD$]	0.2756	0.5219 [1.8938]	0.5227 [1.8964]	0.5236 [1.8998]	0.4886 [1.7727]	0.4898 [1.7771]	0.5082 [1.8440]
	RMD		0.6061 (0.2710)	0.6142 (0.2717)	0.6239 (0.2720)	0.0626 (0.1086)	0.5448 (0.1298)	0.1747 (0.0534)
	KW		112.5528	111.1462	109.6476	127.3946	148.5435	129.2696
	$\rho_{RMD,\lvert bias\rvert}$		0.3812 (0.0121)	0.3795 (0.0121)	0.3777 (0.0121)	−0.1534 (0.0138)	−0.0078 (0.0141)	−0.4482 (0.0113)
	$\rho_{KW,\lvert bias\rvert}$		−0.1897 (0.0136)	−0.1890 (0.0136)	−0.1881 (0.0136)	0.0326 (0.0141)	0.2123 (0.0135)	0.2974 (0.0129)
$R_f^2 = 0.01$	MB [RMB]	0.8267	0.8305 [1.0046]	0.8305 [1.0046]	0.8307 [1.0048]	0.8291 [1.0029]	0.8344 [1.0094]	0.8343 [1.0092]
	MAD [$RMAD$]	0.8288	0.8305 [1.0021]	0.8305 [1.0021]	0.8307 [1.0023]	0.8291 [1.0003]	0.8344 [1.0068]	0.8343 [1.0066]
	RMD		0.2452 (0.1265)	0.2442 (0.1324)	0.2459 (0.1383)	0.0147 (0.0355)	0.5385 (0.1350)	0.0976 (0.0502)
	KW		127.4893	126.0038	124.4168	127.7378	150.3331	129.9351
	$\rho_{RMD,\lvert bias\rvert}$		0.0428 (0.0141)	0.0430 (0.0141)	0.0430 (0.0141)	−0.2618 (0.0132)	0.0245 (0.0141)	−0.1938 (0.0136)
	$\rho_{KW,\lvert bias\rvert}$		−0.0141 (0.0141)	−0.0141 (0.0141)	−0.0142 (0.0141)	0.1983 (0.0136)	0.0802 (0.0141)	0.0584 (0.0141)