Method-of-moments estimation and choice of instruments: numerical computations

by

Stanislav Anatolyev*
New Economic School

Abstract

We numerically evaluate asymptotic variances and biases of various method-of-moments estimators in the Hansen–Singleton model calibrated to real data. Inspection of resulting figures leads a conclusion that applied researchers do not always form instrument sets judiciously.

JEL classification codes: C13, C22

Key words: instruments, generalized method of moments, empirical likelihood, asymptotic variance, asymptotic bias.

*Address: Stanislav Anatolyev, New Economic School, Nakhimovsky Prospekt, 47, Moscow, 117418 Russia. E-mail: sanatoly@nes.ru. Phone: +7-495-129-1700. Fax: +7-495-129-3722. I benefited from discussions with Grigory Kosenok, Geert Dhaene and Dick van Dijk. I also thank seminar audiences at the New Economic School and Erasmus University Rotterdam.

1 Introduction

For decades applied researchers have been using the method of instrumental variables when estimating and testing nonlinear models with rational expectations, like a consumption-based capital asset pricing model (C–CAPM) introduced in Hansen and Singleton (1982). In such situations a researcher faces a need to form an instrument set, and carelessness in implementation of this important step can lead to unreliable results. However, practitioners often complain of lack of guides in selecting instruments in such models.

The present study uses numerical computations as a tool to shed some light on how instrument sets should be formed in nonlinear models with rational expectations. We take the one- and two-period variations of the one-return Hansen–Singleton model estimated by a number of method-of-moments estimators. These estimators are first order asymptotically equivalent but differ in their finite sample properties. They are: the generalized method of moments (GMM) estimator (Hansen, 1982), and two versions of the empirical likelihood (EL) estimator (Imbens, 1997; Smith, 1997) that differ by how they take account of serial dependence. The GMM is still most popular method for this class of models, while the EL has been recruited over years to improve finite sample properties of GMM estimators.

For the estimators of interest, we derive exact formulas for asymptotic variances and second order asymptotic biases while varying a composition of a typical instrument set. To this end, we use a simple but plausible dynamic lognormal specification so that it is possible to find closed form expressions for various moments and calculate *exact* numerical values for asymptotic variances and biases. The resulting numerical comparisons of these across various estimators can provide valuable information on their finite sample properties, in addition to existing Monte Carlo studies. A technical Appendix containing details of computations is available from `http://www.nes.ru/~sanatoly/papers/CCAPMapp.pdf`. Even though the computations are performed for concrete simple models, the tendencies we discover are likely to prevail in more complex situations.

2 Models and estimators

Let $x_{1,t}$ be the one period rate of return, and $x_{2,t}$ be one period consumption growth. The basic $(q+1)$-period one-return C–CAPM of Hansen and Singleton (1982) with CRRA utility implies the Euler equation

$$E\left[\beta^{q+1}x_{1,t+1}x_{1,t+2}\cdots x_{1,t+q+1}x_{2,t+1}^{\alpha}x_{2,t+2}^{\alpha}\cdots x_{2,t+q+1}^{\alpha}-1|I_t\right]=0,$$

where I_t is time t information, β is a discount factor, and α indexes risk aversion. The vector of parameters is $\theta=(\beta,\alpha)'$. The following vector of instruments is used:

$$z_t=\left(x_{1,t},x_{1,t-1},\cdots,x_{1,t-nl_1+1},x_{2,t},x_{2,t-1},\cdots,x_{2,t-nl_2+1},1\right)'.$$

Thus, along with a constant, we employ nl_1 current and most recent lagged values of $x_{1,t}$ and nl_2 current and most recent lagged values of $x_{2,t}$, totaling to $\ell=1+nl_1+nl_2$ instruments. Denote by m_t the moment indicator $z_t\left(\beta^{q+1}x_{1,t+1}x_{1,t+2}\cdots x_{1,t+q+1}x_{2,t+1}^{\alpha}x_{2,t+2}^{\alpha}\cdots x_{2,t+q+1}^{\alpha}-1\right)$, and by $m_{\theta t}$ its derivative with respect to θ. Denote for future use $\Sigma=\left(Q'V^{-1}Q\right)^{-1}$, $\Xi=\Sigma Q'V^{-1}$, $\Omega=V^{-1}-V^{-1}Q\Xi$, where $Q=E\left[m_{\theta t}\right]$ and $V=\sum_{s=-q}^{+q}E\left[m_t m_{t-s}'\right]$.

To derive expressions for asymptotic variances and biases, we impose that the vector $(x_{1,t},x_{2,t})'$ is lognormally distributed, and the law of motion for $(\log x_{1,t},\log x_{2,t})'$ is stationary VAR(1) with normal innovations. Note that such lognormal specification is often adopted when maximum likelihood is applied or is presumed when log-linearization is performed. To quantify the asymptotic variances and biases, we calibrate the model using the data from the Hansen/Heaton/Ogaki GMM package. In what follows, we consider the one-period problem (M_1) corresponding to $q = 0$, and the two-period problem (M_2) corresponding to $q = 1$. Interestingly, the serial correlation in M_2 is pretty severe.

The GMM estimator uses a HAC weight matrix in the Hansen–Hodrick form suitable for problems with serial correlation of finite order; for M_1 it reduces to a non-HAC one. By default we presume that the weight matrix is based on an asymptotically efficient preliminary parameter estimate; this is possible if a GMM estimator is iterated once or to convergence. The EL estimators are adapted to autocorrelation in one of two ways. One named CEL (corrected EL) is derived from the estimating equations for the baseline EL estimator by temporally adding up the moment indicator in the tilting function (Imbens, 1997). Another estimator named SEL (smoothed EL) uses the moment indicator that is smoothed from the outset with a kernel function (Smith, 1997). We assume that the truncated kernel is used; this leads to simplification of some bias expressions (Anatolyev, 2005). Note that for M_1 CEL reduces to baseline EL, while smoothing in SEL is not necessary.

3 Asymptotic variance

The optimal instrument is the one that attains the efficiency bound, the greatest lower bound for the asymptotic variance of GMM/EL estimators (Hansen, 1985). For M_1, the optimal instrument yields the following minimal asymptotic variances for estimates of β and α: 6.878×10^{-3} and 6.202, respectively; efficiency can be attained by using the vector of instruments $(\log x_{1,t}, \log x_{2,t}, 1)'$. For M_2 the optimal instrument yields the following minimal asymptotic variances for estimates of β and α: 7.121×10^{-3} and 6.487, respectively; the optimal instrument is a linear function of all lags of $\log x_{1,t}$ and $\log x_{2,t}$, hence the efficiency bound cannot be attained using finite instrument sets.

All estimators of interest are asymptotically normal with asymptotic variance equal to Σ. Tables 1A and 2A present the asymptotic variance as a function of the composition of the instrument vector. They hide the fact that not including $x_{1,t}$, the rate of return, as an instrument is disastrous for the asymptotic variance for either parameter. That is, the variable that enters the moment function linearly should always be used as an instrument. Also, exclusion of a constant from an instrument set brings sharp efficiency losses, much higher than are attainable by exclusion of several additional lags of regular instruments.

In M_1 the asymptotic variance is stable over instrument sets and quickly reaches an asymptote when the instrument set is expanded. Further, provided that $x_{1,t}$ is included, inclusion of only $x_{2,t}$ allows to nearly reach the variance asymptote. Note that the asymptote with variances of 6.880×10^{-3} for β and 6.222 for α is not that far from the asymptotic variance bounds 6.878×10^{-3} and 6.202, respectively, differing at most by meager 0.3%. Hence, it is not worthwhile to use complex nonlinear functions (other than logs) or many lags for the sake of attaining more asymptotic efficiency.

In contrast, in M_2, the asymptotic variance for both parameters can be decreased by adding more and more lags of either variable, although with sharply decreasing returns. For instance, adding two more lags (which researchers sometimes do in order to check for robustness) of $x_{1,t}$ and $x_{2,t}$ to the instrument set $(x_{1,t}, x_{2,t}, 1)'$ reduces the asymptotic variances for β and α by 1.24% and 2.99%, respectively. Adding twenty more lags reduces the asymptotic variances further by $0.05 \div 0.20\%$. However, again in contrast to M_1, all these gains fall short of what the efficiency bound provides: attaining it could deliver about 12% of efficiency gains for β and about 50% – for α. Switching from levels to logs is able to reduce asymptotic variance more significantly than expanding the instrument set.

4 Asymptotic bias

The results of Anatolyev (2005, Theorem 1), and second-order asymptotic expansions analogous to those in Anatolyev (2005, proof of Theorem 2) lead to the following expressions for the asymptotic biases of order T^{-1} for the GMM, CEL and SEL estimators:

$$\begin{aligned} B_{GMM} &= B_0 + \sum_{s=-\infty}^{+\infty} B_1(s) + \sum_{u=-q}^{q} \sum_{v=-\infty}^{+\infty} B_2(u,v), \\ B_{CEL} &= B_0 + \sum_{|s|>q} B_1(s) + \sum_{u=-q}^{q} \sum_{|v|>q} B_2(u,v), \\ B_{SEL} &= B_0, \end{aligned}$$

where

$$\begin{aligned} B_0 &= \Xi \left(\sum_{s=-\infty}^{+\infty} E\left[m_{\theta t} \Xi m_{t-s}\right] - E\left[\sum_{j=1}^{k} \frac{\partial m_{\theta t}}{\partial \theta_j} \frac{\Sigma}{2} e_j \right] \right), \\ B_1(s) &= -\Sigma E\left[m_{\theta t}' \Omega m_{t-s}\right], \\ B_2(u,v) &= \Xi E\left[m_t m_{t-u}' \Omega m_{t-v}\right], \end{aligned}$$

k is the dimensionality of the parameter vector, and e_j is the j^{th} column of the identity matrix. In the above expressions, some summands are identically zero by the conditional moment restriction.

Tables 1B–1D and 2B–2D show quantified dependence of second order biases on the instrument set, for M_1 and M_2, respectively. The biases of GMM estimators for estimation of β are big and quickly rise when more instruments are exploited. In contrast, the biases of EL estimators for estimation of β are small and stable over instrument combinations; in addition, they are of the opposite sign. The growth of bias for GMM is provided exclusively by the B_2-terms that are not present in the bias expression for CEL in M_1, and by similar B_1- and B_2-terms in M_2. Note that because of cancellations among different bias components CEL sometimes exhibits a smaller bias than SEL. As far as estimation of α is concerned, the bias of GMM in most cases turns out to be smaller in absolute value than those for the EL-type estimators because of heavy cancellations. Cancellations also lead to an interesting phenomenon that extending an instrument set sometimes results in a much smaller bias.

5 Some practical implications

Applied researchers are guided by asymptotic variances when choosing estimators. Often they discover that the actual stochastic properties of these estimators in practice substantially deviate from predictions of asymptotic theory. Most often researchers complain that the estimators are biased, and the impact of bias is comparable with uncertainty implied by the asymptotic variance. Thus, it makes sense to analyze the trade-off between the *first order* asymptotic variance and the *second order* asymptotic bias. To this end, we define an efficiency measure of interest as

$$MSE = \Sigma + \frac{1}{T} B^2,$$

where Σ is the asymptotic variance common to all estimators, and B is the asymptotic bias specific for each estimator. Minimization of MSE over instrument sets yields an optimal instrument combination $nl^* = (nl_1^*, nl_2^*)$. In the experiments below we allow the number of lags of either variable to be no larger than 6 and set $T = 100$.

For M_1, GMM yield $nl^* = (1,1)$ for β and $nl^* = (4,1)$ for α, while both EL estimators show $nl^* = (3,0)$ for both β and α. Switching from GMM to EL leads to less that 0.01% lower MSE for β, but about 0.2% *higher* MSE for α. This is a consequence of fortunate cancellations among the summands of bias components for GMM. In contrast, for M_2 much more efficiency can be attained by switching from GMM to EL: it results in about 0.3% lower MSE for β, and about 0.2% lower for CEL, and about 0.4% *higher* MSE for SEL as far as α is concerned. The optimal strategies exploit many more lags than in the case of M_1: when GMM is employed, $nl^* = (3,2)$ for β and $nl^* = (4,4)$ for α; when EL is employed, $nl^* = (4,3)$ for β and again $nl^* = (4,4)$ for α.

Interestingly, a much sharper contrast between estimators in favor of SEL results if we sum absolute values of bias components thus forbidding numerous cancellations among various components. Hence, in practice the number of components in a formula for asymptotic bias may not be a good measure of biasedness because of cancellations. At the same time, small amounts of reported discrepancies attest that the issue of how many instruments to use in this sort of models is not a very serious issue provided that minimal requirements (like the presence of a constant) are satisfied.

In addition to the reported experiments, we found that using the textbook two-step GMM (with an identity weight matrix at the first step) in place of the iterative GMM does not significantly affect the conclusions.

References

Anatolyev, S. (2005) GMM, GEL, serial correlation, and asymptotic bias. *Econometrica* 73, 983–1002.

Hansen, L.P. (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.

Hansen, L.P. (1985) A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *Journal of Econometrics* 30, 203–238.

Hansen, L.P. and K.J. Singleton (1982) Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* 50, 1269–1286.

Imbens, G.W. (1997) One-step estimators for over-identified generalized method of moments models. *Review of Economic Studies* 64, 359–383.

Smith, R.J. (1997) Alternative semi-parametric likelihood approaches to generalized method of moments estimation. *Economic Journal* 107, 503–519.

Table 1. Asymptotic variances and biases, one-period problem

1A. Asymptotic variance

	$nl_1 \downarrow nl_2 \rightarrow$	0	1	2	3	4
β, $\times 10^{-3}$	1	6.892	6.880	6.880	6.880	6.880
	2	6.881	6.880	6.880	6.880	6.880
	3	6.881	6.880	6.880	6.880	6.880
	4	6.881	6.880	6.880	6.880	6.880
α	1	6.318	6.222	6.222	6.222	6.222
	2	6.226	6.222	6.222	6.222	6.222
	3	6.225	6.222	6.222	6.222	6.222
	4	6.225	6.222	6.222	6.222	6.222

1B. Asymptotic bias for GMM

	$nl_1 \downarrow nl_2 \rightarrow$	0	1	2	3	4
β, $\times 10^{-2}$	1	−0.574	1.160	3.232	5.299	7.363
	2	1.920	3.502	5.859	7.925	9.990
	3	4.096	5.814	7.983	10.098	12.163
	4	6.173	7.897	10.086	12.172	14.245
α	1	0.819	0.947	0.768	0.593	0.419
	2	0.262	0.524	0.087	−0.087	−0.261
	3	−0.011	0.127	−0.139	−0.358	−0.532
	4	−0.196	−0.063	−0.347	−0.540	−0.721

1C. Asymptotic bias for CEL

	$nl_1 \downarrow nl_2 \rightarrow$	0	1	2	3	4
β, $\times 10^{-2}$	1	−0.574	−0.894	−0.894	−0.894	−0.894
	2	−0.576	−0.902	−0.884	−0.8844	−0.884
	3	−0.538	−0.863	−0.878	−0.8764	−0.876
	4	−0.538	−0.861	−0.878	−0.877	−0.877
α	1	0.819	1.111	1.111	1.111	1.111
	2	0.825	1.118	1.102	1.102	1.102
	3	0.790	1.083	1.096	1.095	1.095
	4	0.791	1.081	1.096	1.095	1.095

1D. Asymptotic bias for SEL

	$nl_1 \downarrow nl_2 \rightarrow$	0	1	2	3	4
β, $\times 10^{-2}$	1	−0.574	−0.894	−0.894	−0.894	−0.894
	2	−0.570	−0.906	−0.916	−0.916	−0.916
	3	−0.543	−0.913	−0.916	−0.916	−0.916
	4	−0.538	−0.913	−0.916	−0.916	−0.916
α	1	0.819	1.111	1.111	1.111	1.111
	2	0.819	1.122	1.132	1.132	1.132
	3	0.795	1.128	1.131	1.131	1.131
	4	0.791	1.129	1.131	1.131	1.131

Table 2. Asymptotic variances and biases, two-period problem

2A. Asymptotic variance

	$nl_1 \downarrow nl_2 \rightarrow$	0	1	2	3	4
β, $\times 10^{-3}$	1	8.273	8.185	8.160	8.158	8.158
	2	8.185	8.167	8.136	8.132	8.132
	3	8.168	8.137	8.130	8.123	8.122
	4	8.166	8.135	8.123	8.122	8.120
α	1	14.137	13.559	13.399	13.383	13.382
	2	13.558	13.445	13.241	13.212	13.210
	3	13.447	13.245	13.197	13.154	13.147
	4	13.434	13.229	13.155	13.144	13.134

2B. Asymptotic bias for GMM

	$nl_1 \downarrow nl_2 \rightarrow$	0	1	2	3	4
β, $\times 10^{-2}$	1	−3.489	0.5472	3.571	5.978	8.823
	2	−0.076	0.766	3.787	5.943	8.758
	3	3.035	3.649	4.441	6.716	9.459
	4	5.524	6.196	6.618	7.361	10.184
α	1	5.218	2.836	1.792	1.500	0.879
	2	4.334	3.543	2.623	2.544	1.946
	3	3.487	2.550	2.442	2.334	1.784
	4	3.116	2.125	2.217	2.212	1.677

2C. Asymptotic bias for CEL

	$nl_1 \downarrow nl_2 \rightarrow$	0	1	2	3	4
β, $\times 10^{-2}$	1	−3.489	−2.625	0.839	0.512	1.430
	2	−4.615	−3.755	0.006	−0.614	0.437
	3	−3.456	−1.859	−0.156	−1.031	0.446
	4	−3.366	−1.661	0.748	−0.204	1.859
α	1	5.218	3.411	0.260	0.461	−0.321
	2	6.189	4.808	1.466	1.903	1.014
	3	5.224	2.965	1.516	2.143	0.898
	4	5.113	2.757	0.689	1.446	−0.291

2D. Asymptotic bias for SEL

	$nl_1 \downarrow nl_2 \rightarrow$	0	1	2	3	4
β, $\times 10^{-2}$	1	−3.489	−2.625	−1.091	−1.562	−1.435
	2	−3.706	−3.136	−1.485	−2.150	−1.979
	3	−3.220	−2.198	−1.538	−2.359	−2.097
	4	−3.398	−2.392	−1.658	−2.128	−1.765
α	1	5.218	3.411	1.854	2.204	2.098
	2	5.463	4.313	2.708	3.229	3.089
	3	5.078	3.294	2.693	3.329	3.111
	4	5.224	3.444	2.761	3.134	2.833