An Analysis of the Superior Performance of GEL over GMM*

Ian Dew-Becker

Harvard University

April 5, 2010

Abstract

I examine the performance of the Generalized Method of Moments (GMM) and Empirical Likelihood (GEL) in an empirical example. I find GMM provides estimates that vary widely depending on the set of moment conditions used and cannot be reconciled with one another, while GEL estimates are statistically indistinguishable. I calibrate simulations to match the sample and show that GMM is substantially biased in this setting and GEL is not. GMM's performance degrades under weak identification more quickly than EL. Expansions giving the asymptotic bias of GEL and GMM are helpful for understanding the patterns of estimates from the estimators.

Keywords: Generalized Empirical Likelihood, Generalized Method of Moments, Weak Identification, Consumption, Household wealth, Dynamic panel models

JEL Codes: E21, C13, C14

*I thank Guido Imbens, Nick Bloom, John Campbell, Naomi Hausman, David Laibson, and Jim Stock for helpful comments and discussions, and John Zedlewski for code samples. I also acknowledge support from an NSF graduate research fellowship. All errors are my own.

1 Introduction

It is well known that in a variety of settings, Generalized Method of Moments (GMM) has poor small sample properties. Even in large samples, if the number of instruments is too large or weak, GMM can be inconsistent; two prime examples of this come from Angrist and Krueger's (1991) study of the returns to schooling, and estimates of the CAPM in finance. Further, in nonlinear models in particular, GMM can have substantial small sample biases.[1] Many solutions have been proposed, but almost all involve throwing information away from the data. Empirical Likelihood has been proposed as an alternative method that uses all available information and estimates exactly the same models that GMM does, but does not share some of its flaws.[2] It achieves the bias of optimally weighted, infeasible, GMM. That is, EL has small sample properties similar to what we would obtain with GMM if we knew the true optimal weighting matrix.

EL, however, has not caught on in empirical analysis. One goal of this article is to show that EL can provide a substantial improvement to inference in a realistic setting. I estimate a simple dynamic model for household wealth. The particular model exploits the two prime weaknesses of GMM: weak instruments and a high degree of overidentification. After estimating the model with both GMM and EL, I calibrate a simulation to match the data as closely as possible. Histograms of the estimates from this simulation are plotted in figure 1. GMM is biased down by more than one standard deviation and the GMM 95 percent confidence interval rejects the true value of the parameter as much as 30 percent of the time. On the other hand, EL is unbiased and the simulated distribution closely matches the asymptotic distribution.

Past studies of models similar to the one I work with, and studies comparing EL and GMM, have tended to use Monte Carlo evidence to support one estimator or another. Recently, though, Newey and Smith (2004) have provided analytic formulas for the small sample

[1] See Newey and Smith, 2004.

[2] See, for example, the work of Imbens, 1997; Kitamura and Stutzer, 1997; Owen, 1988; and Qin and Lawless, 1994.

bias of GMM and EL. If these formulas are accurate, then there is no need to run simulations; rather, we can directly assess the bias to help understand differences between various estimators. In section 4, I expand the simulations from figure 1 and section 3.3 and test how well the Newey–Smith predictions match the simulated bias. I find them to be accurate. Moreover, across a broad range of parameter sets and degrees of overidentification, the bias of GMM remains far larger than that of EL. When scaled by the standard error, the smallest bias I observe for GMM is larger than the largest bias I observe for EL.

A number of alternatives generalizations to EL have been proposed (generalized empirical likelihood, GEL). I discuss one in particular, Hansen, Heaton, and Yaron's (1996) continuously updated GMM estimator (CUE, as distinct from 1 and 2-step GMM). While CUE does not achieve the higher order efficiency of EL, the two estimators (along with GMM) are first order equivalent. Moreover, CUE captures much of the higher order efficiency gain of EL over GMM and has the advantage of being easier to implement than EL. In small samples, as long as the moments are not very skewed (which is simple to test), CUE and EL have similar higher order properties. In simulations with no skewness in the moments by construction, I find EL and CUE estimates are over 98 percent correlated. While EL is not at this point too computationally intensive for the average researcher to carry out, CUE may be more convenient for quick analyses.

The findings presented in this article provide further support for using GEL rather than GMM. The empirical analysis I report in section 3 would be simply unreportable if I used GMM – the estimates from different moments are wildly inconsistent. While I focus on one specific model, the problems with GMM are found in a variety of areas of research. Two good examples are finance, where variants of the CAPM are routinely estimated with 25 moments for 4 parameters, and models with many weak instruments, e.g. Angrist and Krueger (1991).[3] The superiority of EL and CUE is by no means restricted to this model. General theoretical results show that these estimators are preferable to GMM.

[3]While LIML and JIVE of Angrist, Krueger, and Imbens, 1999, both also work in the linear IV setting with many instruments, EL has the advantage of applying to a more general class of problems.

The remainder of the paper is organized as follows. In section 2 I describe the basic model used in the rest of the paper, review the standard methods of identifying the parameter of interest, and define the estimators. In section 3 I apply the dynamic panel model to household wealth data from the Health and Retirement Survey. The analysis returns an estimate that indicates wealth shocks are highly persistent, and that wealth and consumption can be modeled as a random walk for this sample. Next, section 4 examines a large set of Monte Carlo simulations to see whether the Newey–Smith predictions for the bias are accurate in a sample calibrated to match the data used in section 3. I find that the Newey–Smith predictions are effective for both EL and GMM. This means that the Newey–Smith analytic results on the dominance of EL are relevant in real world samples, and that rather than running endless Monte Carlo simulations, researchers can rely on these results in the future.

2 Identification and Estimators

2.1 The model and identification

The basic model I analyze is the widely studied panel AR model,[4]

$$y_{it} = \theta y_{it-1} + \alpha_i + \varepsilon_{it} \tag{1}$$

where $t \in \{1, 2, ..., T\}$, ε_{it} is assumed to have zero mean and be serially uncorrelated, and α_i is an unobserved individual-specific fixed effect. I use the model here as a compact framework for describing wealth dynamics, with y_{it} representing log asset wealth. θ measures how persistent wealth shocks are. When a household's wealth rises above its long run trend, θ determines how quickly it comes back. θ is also related to the difference between the average and marginal propensities to consume from wealth; an MPC higher than the APC leads to $\theta < 1$. Models in which wealth is the only state variable and which are scale-independent

[4]See, for example, Anderson and Hsiao, 1981, Holtz-Eakin et al., 1988, Arellano and Bond, 1991, and Blundell and Bond, 1998.

have $\theta = 1$. The buffer stock model (e.g. Carroll, 1997) adds income as a state variable and has $\theta < 1$. The results below will show that to a first approximation, asset wealth of older households can be modeled as following a random walk (meaning that consumption also follows a random walk and is a constant share of asset wealth).

In the empirical analysis, it is important to account for measurement error. If we measure y imperfectly as $w_{it} = y_{it} + \nu_{it}$, where ν_{it} is measurement error mean independent of wealth (though not necessarily homoskedastic), then the model becomes

$$w_{it} = \theta w_{it-1} + \alpha_i + \varepsilon_{it} + \nu_{it} - \theta \nu_{it-1}. \tag{2}$$

When the number of available time periods observed is finite, OLS is inconsistent for this model. Anderson and Hsiao (1981), Holtz-Eakin et al. (1988), and Arellano and Bond (1991) suggest differencing (2) and using lagged observations of w_{it} as instruments for Δw_{it}. This leads to the following $(T-3)(T-2)/2$ "difference" moments,

$$\mathbb{E}\left[(\Delta w_{it} - \theta \Delta w_{it-1}) \cdot w_{it-j}\right] = 0 \text{ for } j \geq 3. \tag{3}$$

The validity of these moments relies only on the assumption that ε_{it} is not serially correlated; it need not be homoskedastic or uncorrelated with α_i. I will refer to models that use only these moments as "DIF", following Blundell and Bond (1998).

Blundell and Bond (1998) show that in the case where ε_{it} and α_i are uncorrelated, we can take advantage of equation (2) without differencing, and use Δw_{it-2} as an instrument for w_{it-1}.[5] This gives us $T-3$ more "levels" moments,

$$\mathbb{E}\left[(w_{it} - \theta w_{it-1}) \Delta w_{it-2}\right] = 0. \tag{4}$$

When I use the moments from both (3) and (4) I will refer to the models as "ALL".

[5] Note that by the assumption of no serial correlation in ε_{it}, all that this moment actually requires is that ε_{i0}, the first observation of ε_i, is uncorrelated with α_i.

Blundell and Bond (1998) show that the moments in (4) are relatively more powerful when θ approaches 1. One way to see this is to rewrite the two moments as

$$\begin{aligned} \text{Differences:} \qquad & \mathbb{E}\left[\Delta w_{it} \cdot w_{it-j}\right] - \theta \mathbb{E}\left[\Delta w_{it-1} \cdot w_{it-j}\right] = 0 \\ \text{Levels:} \qquad & \mathbb{E}\left[w_{it} \cdot \Delta w_{it-2}\right] - \theta \mathbb{E}\left[w_{it-1} \cdot \Delta w_{it-2}\right] = 0. \end{aligned} \tag{5}$$

In line (5), when the true value of θ approaches 1, both $\mathbb{E}\left[\Delta w_{it} \cdot w_{it-j}\right]$ and $\mathbb{E}\left[\Delta w_{it-1} \cdot w_{it-j}\right]$ approach zero (since Δw becomes unpredictable), which means that for any value of θ, the condition will hold. In the levels moment, the two expectation terms are different from zero, so even with $\theta = 1$ it is still identified.

2.2 Estimators

In general, this type of model is estimated with the Generalized Method of Moments (GMM).[6] Denote the moment function $g(z_i, \theta)$, where z_i is an observation of data and g maps z and θ to an $m \times 1$ vector of moments, and $\mathbb{E}g(z, \theta) = 0$. 1-step GMM is

$$\hat{\theta}_{1GMM} = \arg\min_{\theta} \bar{g}(z, \theta)' I_m \bar{g}(z, \theta),$$

where I_m is the m-dimensional identity matrix and $\bar{g}(z, \theta) = n^{-1} \sum_i g(z_i, \theta)$. Two-step GMM is

$$\begin{aligned} \hat{\theta}_{2GMM} &= \arg\min_{\theta} \bar{g}(z, \theta)' \hat{\Omega}(\theta_{1GMM})^{-1} \bar{g}(z, \theta) \\ \hat{\Omega}(\theta) &= n^{-1} \sum_i g(z_i, \theta) g(z_i, \theta)'. \end{aligned}$$

A variety of Monte Carlo and empirical analyses have shown GMM is likely to be biased in samples encountered in the real world.[7] An alternative estimator is Empirical Likelihood

[6] Though see Binder, Hsiao and Pesaran (2005) for a VAR with fixed effects estimated with Quasi-Maximum Likelihood.

[7] Among others, see Blundell and Bond (1998) and Imbens (2002).

(EL). EL begins by treating the observed data as the support for the true distribution.[8] We estimate that distribution: to each point, we assign some probability π_i such that $\sum_i \pi_i = 1$. EL maximizes the log likelihood, $\sum_i \log \pi_i$ such that the moment conditions hold, with the expectations taken with respect to the π_i's. That is,

$$\begin{aligned} \hat{\theta}_{EL}, \{\hat{\pi}_i\} &= \arg\max_{\theta, \{\pi_i\}} \sum_i \log \pi_i \\ \textit{such that } \sum_i \pi_i g(z_i, \theta) &= 0 \textit{ and } \sum_i \pi_i = 1 \end{aligned}$$

In the exactly identified case, we can find a θ such that the moment conditions hold for any set $\{\pi_i\}$, so $\pi_i = 1/n$, where n is the sample size. When the model is overidentified, the π_i's will not necessarily all be equal. Intuitively, we are using information from the overidentifying restrictions to tell us how to reweight the data and obtain more efficient estimates. This is analogous to using two-step GMM to select a weighting matrix that reweights the data more efficiently.

In fact, a simple modification to standard GMM captures most of the benefits of reweighting in EL. Hansen, Heaton, and Yaron (1996) propose the continuously updated estimator (CUE):

$$\theta_{CUE} = \arg\min_{\theta} \bar{g}(z, \theta)' \left[\sum_i n^{-1} g(z_i, \theta) g(z_i, \theta)' \right]^{-1} \bar{g}(z, \theta)'.$$

The only change from standard GMM is simply to simultaneously update the weighting matrix, rather than updating it iteratively after obtaining an estimate of θ with a fixed weighting matrix.

Newey and Smith (2004) show that this estimator fits into the family of Generalized Empirical Likelihood (GEL) estimators. EL maximizes the log likelihood, $\sum_i \log \pi_i$, which is the same as minimizing a particular distance between $\{\pi_i\}$ and a flat likelihood, $1/n$. In particular, $\sum_i \log \pi_i$ and $-\sum_i \log \left(\frac{1/n}{\pi_i} \right)$ differ only by a constant, so the distance metric for

[8] Unobserved points could be included in the support, but any estimator will give them zero probability, so we can restrict our attention to the observed data.

EL is the log. Following Corcoran (1998), GEL uses the Cressie-Read discrepancy statistic to measure this same distance. GEL still involves choosing a set $\{\pi_i\}$ such that the moment conditions hold, but the objective function becomes,

$$\arg\min_{\{\pi_i\}} \frac{1}{\lambda(1+\lambda)} \sum_i \frac{1}{n} \left[\left(\frac{1/n}{\pi_i} \right)^{\lambda} - 1 \right]$$

As $\lambda \to 0$, this function approaches the EL objective function. As $\lambda \to -1$, it approaches the exponential tilting (ET) estimator of Kitamura and Stutzer (1997) and Imbens, Spady, and Johnson (1998). Newey and Smith (2004) show $\lambda = -2$ corresponds to CUE. ET and CUE are the two main GEL estimators analyzed in the literature because ET has convenient information theoretic interpretations and may be computationally stable, and CUE has a simple interpretation as an extension of GMM. In principle, however, any value of λ will produce an estimator that is first order equivalent to GMM and EL. In section 4 I examine further the differences between EL, CUE, and GMM.

There are two types of papers that have argued that EL is superior to GMM. The first uses Monte Carlo simulations and shows that GMM tends to be biased and have a sampling distribution that is not well approximated by the asymptotic distribution.[9] The second type of argument derives analytic results about the distribution of EL and GMM. The prime theoretical result about EL is that when the data are discrete, EL is exactly equal to maximum likelihood estimation. Therefore, with large samples drawn from a continuous distribution, EL comes arbitrarily close to MLE, and hence inherits its higher order efficiency properties.

Newey and Smith (2004) provide the most extensive results in the second area. They provide formulas for the bias of GEL and GMM. I discuss these expansions below. Furthermore, I attempt to unify the two types of analysis, Monte Carlo and theoretical. In section 4 I report simulation results that match the Newey–Smith predictions closely. In other words, to analyze GMM and GEL, researchers can rely exclusively on formulas rather

[9] See, e.g. Imbens, 2002.

than expending time and computer processing power on simulations.

3 Empirical Analysis

3.1 The data

For the empirical example, I draw wealth data from the RAND extract of the Health and Retirement Survey (HRS). I measure wealth as total wealth excluding IRAs, which includes home equity (excluding second homes), financial wealth, and the value of any private businesses, vehicles, and other savings (including, e.g., jewelry and collectibles) net of nonhousing debt. Observations of wealth occur at two-year intervals, running from 1994 to 2006 – seven total surveys.[10] There are 1,315 households with 7 consecutive observations in the dataset. Households are eliminated from the sample for three reasons. First, I only keep households which begin the sample with at least $25,000 in wealth in order to limit the effects of outside sources of wealth, such as social security, on behavior.[11] This eliminates 205 households. Second, households can never have negative net wealth because the model is estimated in logs, which eliminates 43 households. Last, I drop households with a recorded wealth change greater than a factor of 10 to eliminate outliers. This leaves a total of 991 usable households in the preferred sample, with $7 \times 991 = 6937$ observations.

Table 1 reports summary statistics for the sample. The median respondent is 63 years old, with $209,000 of wealth, $47,000 in income, and a high school education. Almost exactly half of the sample is currently employed, and nearly the entire sample is white. The people in the sample have above average socioeconomic status, but the 75th percentile still has only a college education and under a half million dollars in nonhousing assets.

[10] I drop the 1992 sample due to changes in the survey.

[11] If a household has substantial assets that we do not observe and that cannot be adjusted (e.g. social security or a defined benefit contribution), then the estimate of θ will be biased down. Intuitively, when we only observe the most responsive margin of adjustment, we will overestimate responsiveness. However, the minimum level of initial assets does not have a substantial effect on point estimates of θ. IRAs are excluded so that rather than capturing some pension wealth but not all of it, we just exclude pensions entirely. The results are not sensitive to this decision. They are also not sensitive to the required level of initial wealth.

3.2 Estimates of θ

To make the model suitable for estimation, I add six time dummies, μ_t, to equation (2), leading to the full model,

$$w_{it} = \theta w_{it-1} + \alpha_i + \mu_t + \varepsilon_{it} + \nu_{it} - \theta \nu_{it-1}. \tag{6}$$

We then also have a set of moments that identifies μ_t,

$$\mathbb{E}\left[w_{it} - \theta w_{it-1} - \mu_t\right] = 0$$

Table 2 reports estimates of θ using EL and GMM.[12] The first row reports results using only the differences moments from line (3), while the second row also includes the levels moments from line (4). Unsurprisingly, the estimates for DIF have much larger standard errors than those for ALL, due to the relative weakness of the differences moments when θ approaches 1. Notably, however, the point estimate of θ from EL in both rows is roughly the same, while the GMM estimate is vastly different.

If we only had GMM, these results would be unreportable. One set of moments gives a 99 percent confidence interval with an upper bound of 0.77, while the other moment set returns the same confidence interval with a lower bound of 0.85. For GMM, a Hausman test rejects the hypothesis that the true value of θ is the same in the two models with a p-value of 7×10^{-4}. One can imagine two responses a researcher might have to these results. The first is that the levels moments are invalid and are causing bias. The researcher would then drop the levels moments and end up with a point estimate of 0.18. A more cautious researcher might go further and note that the p-value for the overidentifying test for GMM with DIF is low, and therefore conclude that the model must simply be misspecified.[13]

Both of these outcomes would be disappointing. With EL, the results do not conflict.

[12] Estimation was carried out with John Zedlewski's MatELike package for Matlab (http://people.fas.harvard.edu/~jzedlewski).

[13] This is in fact exactly what I did when I first analyzed this dataset.

The 95 percent confidence interval for DIF easily contains the point estimate from ALL, and the Hausman test gives a p-value of 0.54. Moreover, a likelihood ratio test of either model does not reject the null hypothesis of correct specification. A good estimator leads a researcher to the correct estimate. GMM does not do that here.

But how do we know EL is giving the correct estimates and GMM is misleading? While table 2 is certainly suggestive that the EL estimates with the full moment set is correct, it is premature to conclude that GMM is biased based on the circumstantial evidence in table 2. In the next section I calibrate simulated data sets to match the empirical sample. I find that GMM is in fact substantially biased and leads to poor inference with both DIF and ALL.

3.3 Simulations

To calibrate the model in equation (6), we need to know the variances of α, ε, and ν. These can be calculated from the residuals, u_{it},

$$u_{it} = \alpha_i + \varepsilon_{it} + \nu_{it} - \theta\nu_{it-1} = w_{it} - \theta w_{it-1} - \mu_t.$$

The variance of u_{it} along with its autocovariances identify the variances of α, ε, and ν, assuming independence and homoskedasticity in the shocks.[14] I estimate these parameters using the same sample as above. I then generate data sets with random draws from normal distributions. The parameters I use for the simulation are,

$\theta = 0.96$	$\sigma_\alpha/(1-\theta) = 0.74$
$\sigma_\varepsilon = 0.30$	$\sigma_\nu = 0.34$
$N = 991$	

[14] In reality, there is heteroskedasticity, which means that there are multiple ways to estimate the variances which need not all give the same answer. I estimate the variances using just the first two autocovariances, but the results are similar if alternative lags are used.

The time dummies μ_t all set to zero, but I allow them to be estimated freely in the simulations. These parameters have empirical value on their own. The magnitude of the innovations to wealth is substantial: their standard deviation is 0.3 – a 30 percent shock to wealth in two years (since the data is biennial). Moreover, measurement error is as important as ε, meaning that year-to-year comparisons of household wealth, as are sometimes done with data sets like the Survey of Consumer Finances, are probably not informative. Since shocks are so persistent due to the high value of θ, the majority of the variation in wealth across households is driven by past values of ε rather than by the fixed effect α_i. That is, households generally do not have low wealth simply due to preferences, i.e. a low target level of wealth.

The results from the simulations are summarized in table 3 and figure 1. Figure 1 plots the distribution of the estimates of θ from two-step GMM and EL; estimates using DIF are in the top panel, with ALL in the bottom. The difference between EL and GMM is readily apparent: whether we use DIF or ALL, GMM is clearly biased downwards while EL is median unbiased. Moreover, for ALL the GMM distribution is noticeably less peaked than that of EL. In these simulations, EL and CUE give estimates that are over 98 percent correlated. I do not replicate the result of Hansen, Heaton, and Yaron (1996) that CUE has heavy tails.[15]

Table 3 reports a number of statistics that confirm the problems noted in figure 1. As in the figure, the top panel reports results for DIF, and the bottom panel ALL. Results are reported for GMM, EL, and CUE, and the expected results using the asymptotic distribution are in the far right column.

GMM is biased down by 0.14 with DIF, and 0.06 with ALL. EL and CUE, on the other hand, have no bias in their medians, and only minimal bias in their means. Looking at the higher moments, the standard deviations of all three estimators are inflated slightly, but more for GMM. For ALL, the GMM standard deviation is twice that of the analytic distribution.

[15] They find that even the 10th and 90th percentiles are far from the asymptotic distribution, while I find no problems at the 5th and 95th percentiles, and excess kurtosis of between 4 and 6.

To gauge the quality of inference, the key statistics are the confidence interval coverage rates. The EL and CUE 90 and 95 percent confidence intervals have nearly exactly correct coverage rates. The GMM coverage rates are far below their nominal levels. For ALL, the true value of θ is contained in the 95 percent confidence interval only 69 percent of the time. The coverage rates are actually worse for ALL than for DIF. Moreover, even though the absolute size of GMM's bias is larger for DIF than ALL, it is larger relative to the standard error for ALL. This is a troubling finding. Normally we think that by adding informative moments we improve inference, but in this case inference actually becomes worse.

Figure 2 shows what happens as θ is allowed to vary. I estimate DIF while keeping the parameters fixed at their levels in the above simulation (including scaling σ_α by $(1-\theta)$), and allowing θ to approach 1 geometrically. As noted above, when θ approaches 1, the instruments in DIF become irrelevant, so each estimator should fail. The top panel plots 90 percent confidence interval coverage rates for 1 and 2-step GMM, while the bottom panel shows the median bias of each estimator divided by the standard error.[16] The GMM coverage rates are never close to their nominal levels – 1-step GMM performs particularly poorly, even with $\theta = 0.68$. The EL coverage rate, though, never falls below 0.75 even as θ rises above 0.9999.[17] Looking at the bottom panel, it is clear that much of the poor inference with GMM is due to bias – both 1 and 2-step GMM are biased substantially, whereas the bias of EL only becomes problematic as rises past 0.9975. The results for CUE for both metrics in figure 2 are identical to those of EL.

The findings in figure 2 help interpret the literature on weak identification in GMM and GEL. Stock and Wright (2000) first showed that GMM is inconsistent under weak identification. Guggenberger and Smith (2005) extend Stock and Wright's (2000) results to GEL and find that EL and CUE are asymptotically first order equivalent under weak identification and inconsistent. The results reported in figure 2 show, however, that the

[16] I report results for 1-step GMM because it has been proposed as a more robust estimator than 2-step GMM, by Altonji and Segal, 1996, among others.

[17] For the highest values of θ used in figure 2, I find estimated standard errors of approximately 0.6, which is large considering the qualitative differences between $|\theta| \geq 1$, $0 < |\theta| < 1$, and $\theta = 0$.

inconsistency of GEL is substantially less problematic than that of GMM. 2-step GMM delivers poor inference even with $\theta = 0.64$, but then as θ rises above 0.96, the results are disastrous. EL, on the other hand, does not decay until θ is above 0.9975, and even then the 90 percent confidence coverage rate is at worst 80 percent.

The results of these simulations can be summarized as follows. When we create a data sample similar to the empirical sample studied in section 3, GMM is biased down by about one standard error for both DIF and ALL. Furthermore, the shape of its distribution is far from the asymptotic distribution and its confidence interval coverage rates are well below the nominal levels. EL and CUE are median unbiased and have confidence intervals with nearly perfect coverage rates. When θ approaches 1 and the model becomes unidentified, the performance of GMM degrades far sooner than that of EL. The distributions CUE and EL are essentially identical, meaning that one can choose between them based on ease of use.

4 Predicting the bias

Newey and Smith (2004) provide second-order expansions that give the asymptotic bias for GMM and GEL. As the analysis of this paper is about understanding the small sample performance of various estimators, we wish to check whether the Newey–Smith predictions for the bias are accurate in a small sample. I find that the bias terms in fact explain a substantial amount of the variation in the simulations.

Newey and Smith divide the bias into four unobservable terms, which they call B_I, B_G, B_W, and B_Ω. It is not easy to calculate these terms for arbitrary models. One can think of them as four essentially random terms, which may be correlated, but for which we do not know the correlation. In general, then, an estimator with more random terms defining its bias will have a more uncertainty. This leads to the result that EL is higher order efficient compared to the other estimators considered here – its bias terms create less uncertainty.

Because definition of the terms involves a good deal of notation, I define the complete formulas in the appendix. For all estimators considered here, the bias can be written in terms of four components. Newey and Smith (2004) interpret the terms as follows: B_I is the asymptotic bias for GMM with the optimal weighting matrix. It is also the only term in the bias of EL. B_G is driven by error in estimating the derivatives of the moment conditions; B_W is an error due to the preliminary estimator in 2-step and iterated GMM, and is related to the weighting matrix; and B_Ω comes from estimation of the covariance matrix of the moments. B_I and B_G are both driven by correlation between the moments and their derivatives, while B_Ω is driven by a cubic function of the moments.[18]

The biases of the estimators are,

$$\begin{aligned} bias_{EL} &= B_I N^{-1} \\ bias_{GEL} &= \left[B_I - \frac{\lambda}{2} B_\Omega \right] N^{-1} \\ bias_{CUE} &= (B_I + B_\Omega)\, N^{-1} \\ bias_{GMM} &= (B_I + B_\Omega + B_G + B_W)\, N^{-1} \end{aligned}$$

where λ is the Cressie-Read parameter (noted in section 2.2) defining the GEL estimator. $\lambda = 0$ corresponds to EL and $\lambda = -2$ corresponds to CUE. The asymptotic bias of (infeasible) optimal GMM, B_I, is shared by all GEL estimators. GEL estimators with $\lambda \neq 0$, i.e. estimators other than EL, also have a bias term due to estimation of the covariance matrix of the moments, which determines the weights placed on the moments. In situations where the moments have little skewness (which is true, for example, of a moment which is the product of a normally distributed instrument and a normally distributed mean zero error term), B_Ω will tend to be small, meaning that there should be little difference between EL and CUE. The reason that EL is in general more efficient than other GEL estimators and GMM is that it has the fewest of these random terms. Similarly, the bias of GMM includes $B_G + B_W$, which does not affect GEL.

[18] Since the moments in the simulations all have zero skewness, many of the terms of the cubic function will be zero, but the cross products need not be zero. Neverthelsss, B_Ω tends to be very small in the simulations.

To actually estimate these terms, I take the formulas from the appendix and estimate them on a large sample ($N = 1.6 \times 10^6$) of simulated data. Figure 3 shows the predicted bias for ALL as simulated above, except with σ_μ^2 multiplied by 2/3 in order to make the graphs easier to read.[19] Figure 3 has four panels, plotting the four terms from Newey–Smith. In each panel there are five lines, depending on the degree of overidentification. As θ approaches 1, four of the moments in ALL become irrelevant.[20] So the five lines include zero to four of those relatively uninformative moments. The y-axis gives the value of the bias term, and the x-axis is the value of θ.

The top left panel plots B_I, the bias of both EL and optimally weighted (infeasible) GMM. B_I is unaffected by the degree of overidentification, showing that EL is robust to the addition of irrelevant moments. As θ rises, the bias of EL increases, tripling as θ moves from 0.900 to 0.975. EL seems unaffected by the number of instruments, but it is negatively affected by their weakness.

The top right panel of figure 3 plots B_Ω, the term that drives the differences among GEL estimators. The magnitude is comparable to the bias of EL – between -0.001 and -0.002. The magnitude of B_Ω is unrelated to the value of θ, which reflects the fact that θ has no effect on the skewness of the moments. B_Ω is affected somewhat by the number of irrelevant moments, which could simply be due to the fact that they have a different distribution from the other moments. For CUE, the values of B_Ω imply a bias roughly double that of EL. An estimator with $\lambda = 2$ would eliminate the bias of EL. Globally, EL that is not bias corrected

[19] When I use the estimated value of σ_μ^2, the predicted bias of GMM is nearly zero – in the simulations below, I show that the Newey–Smith formulas tend to underpredict the bias of GMM for this model.

[20] In preceding sections, I wrote the moments as,

$$\begin{aligned} \mathbb{E}\left[(\Delta w_{it} - \theta \Delta w_{it-1}) \cdot w_{it-j}\right] &= 0, \; for \; j \geq 3 \\ \mathbb{E}\left[(w_{it} - \theta w_{it-1}) \, \Delta w_{it-2}\right] &= 0, \end{aligned}$$

following Blundell and Bond, 1998. But one can also write them as

$$\begin{aligned} \mathbb{E}\left[(\Delta w_{ik} - \theta \Delta w_{ik-1}) \cdot w_{i0}\right] &= 0, \; for \; k \geq 4 \\ \mathbb{E}\left[(w_{it} - \theta w_{it-1}) \, \Delta w_{it-j}\right] &= 0, \; for \; j \geq 2. \end{aligned}$$

In that case, the moments in the first line become irrelevant as θ approaches 1. The second line remains relevant. The moments from the first line are the ones I drop, starting with the highest values of k.

need not be optimal in every situation. That is clear from this case, where an estimator with $\lambda = 2$ would have a smaller bias. However, in general cases where we do not know B_I and B_Ω, EL has less higher order error associated with it.

The bottom left panel plots B_G. This term is, like B_I, driven by correlation between the moments and their derivatives, but it is weighted differently. This is the term that drives the bias of GMM, both in this setting and also in the simulations I report below. It is an order of magnitude larger than B_I and B_Ω in figure 3. As expected, B_G rises with the degree of overidentification. Somewhat surprisingly, it actually shrinks as θ approaches 1. This result is in conflict with the results form figure 2, which is explained by the fact that the Newey-Smith expansions are not designed to model the effects of weak identification in the sense of Stock and Wright (2000), which occurs as θ approaches 1.

The last panel of figure 3 plots B_W, the term driven by the initial estimate for the weighting matrix. Since 2-step GMM uses an efficient estimate of the weighting matrix, we in general would not expect this term to be important, and it clearly is not – it is smaller than $5 * 10^{-5}$ except for a few higher values of θ.

These plots are related to the literature on GMM with large numbers of weak instruments, e.g. Bekker (1994), Han and Phillips (2005), and Newey and Windmeijer (2005). I find that GMM performs more poorly as more moments are added, but that EL is unaffected by the number of moments. The studies of GMM with asymptotically large numbers of instruments use fundamentally different methods from Newey and Smith (2004), studying first order distributions, but the results are clearly informative for small samples

4.1 Simulations

To test the relevance of the asymptotic bias predictions in a small sample, I expand the simulations above. If we have sets of estimates for a given model with a variety of Cressie-Read parameters, then we can identify B_I and B_Ω from a simple regression of the estimated bias of an estimator on its Cressie-Read parameter. This section reports results from a large set of simulations varying the models simulated by the number of time periods in the data, the variances of the shocks, and the moments used for estimation. This allows us to see how well variation in the bias is predicted by the asymptotic predictions.

I run three sets of simulations, with approximately 1000 replications in each. There are numerous differences across the simulations in terms of parameters used, the number of time periods in the data sets, and the sets of moments. This is by design – I want to see how well the Newey–Smith predictions work in a wide variety of models.

Simulations *1a* and *1b* are differentiated by whether I include the levels moments or not. I set θ to a lower value than in section 3.3 which will make the differences moments more powerful. Within each simulation, I estimate 15 different models, varying the variances of the shocks and the number of lags used as instruments in the differences moments, which controls the degree of overidentification.[21] The parameters are as follows,

$\theta = 0.81$	$\sigma_\varepsilon \in \{0.27, 0.54\}$
$\sigma_\alpha / (1-\theta) \in \{0.51, 1.02, 2.04\}$	$\sigma_\nu \in \{0.17, 0.34\}$
$N = 991$	

I estimate each model using GMM and GEL, with the Cressie-Read parameter varying from -3 to 1 in increments of 0.5. In both the simulations, μ_t is set to zero for all t, but I still freely estimate the μ_t's.

In the third set of simulations, rather than varying the parameters, I vary the number of time periods in the model in order to allow for greater degrees of overidentification. The set

[21] There are 5 parameter sets: the original, ones with σ_α and σ_ν individually divided by 2, and ones with σ_α and σ_ε individually multiplied by 2. For each of these five models, I use a maximum of between 2 and 4 lags of wealth as instruments in the differences moments.

of parameters used is slightly different from above,

$\theta = 0.60$	$\sigma_\alpha/(1-\theta) = 1.53$
$\sigma_\varepsilon = 0.30$	$\sigma_\nu = 0.10$
$N = 761$	

As above, μ_t is set to zero, but here I do not estimate it freely – the estimators all assume $\mu_t = 0$. Since θ is smaller, the levels moments are not as crucial to identification, so I restrict attention in these simulations to just the differences moments. This allows me to expend processing time on varying the Cressie-Read parameter; I use 19 different estimators with λ varying from -4 to 2, in addition to GMM. As in the first two sets of simulations, I vary the maximum number of lags used as instruments from 2 to $T-3$. The number of time periods, T, varies from 7 to 11, meaning the number of available moments varies from 10 to 36. To summarize, the simulations vary the sample along every available dimension: θ, the variances of all the shocks, the number of time periods, and the moments.[22]

Figure 4 gives a first look at the results of the simulations. The top panel plots the estimates of B_I, the EL bias. The x-axis is the predicted value and the y-axis is the actual value. Nearly all of the estimates lie very close to the 45-degree line. For simulation *1a*, the R^2 is 0.10, but it rises to 0.70 when the simulations with $\sigma_\alpha/(1-\theta) = 2.32$ are dropped. Similarly, for simulation *1b*, the R^2 is 0.82 when the one outlier is excluded. For simulation 2, the R^2 is 0.98. Across a very broad range of parameters, the Newey–Smith predictions perform well for EL. This has two implications: first, large scale Monte Carlo simulations may be unnecessary for EL since analytic results describe the small sample behavior well; second, estimates of B_I may be informative in telling whether EL is biased.

The second panel of figure 4 plots results for B_Ω. As noted above, we expect B_Ω to be small, and it clearly is. The scale in this plot is an order of magnitude smaller than the scale

[22] The results reported below are means of the biases. It is possible that the estimators do not have moments in small samples. I therefore also looked at medians and trimmed means, and the results were similar.

for B_I. For simulations *1a* and *1b*, there is little connection between the predictions and the estimated values. The R^2 's are both near zero. Since the estimates necessarily involve simulation error, this is not surprising. What is surprising is that the predictions of B_Ω for simulation *2* are actually reasonably successful, with an R^2 of 0.55.

The bottom panel of figure 4 plots results for the bias of GMM, $B_I + B_G + B_\Omega + B_W$. The predictions apparently work well for simulations *1a* and *1b*, with a strong positive relationship between the actual and predicted values. However, there is no connection between the actual and predicted bias for simulation *2.* In all three simulations, the actual bias is larger than the prediction. While the simulations do not allow us to identify B_G and B_W individually, if we trust the Newey–Smith predictions for them, then we can see what drives the bias of GMM. for these simulations, the bias is driven almost entirely by B_G, which comes from correlation between the moments and their derivatives.

The results of the simulations are summarized in table 4. The parameters are summarized, and the R^2's from the plots are reported. I also report the median and range for the bias terms divided by standard errors. These are the bias estimates from the simulations divided by analytic standard errors. For all three sets of simulations, the maximum bias for EL is smaller than the minimum for GMM. Even the median bias for GMM is worrisomely large. It is 0.7 for simulations *1a* and *1b*, and 1.91 for simulation *2*.

4.2 EL or CUE?

In all of the simulations I have reported here, EL and CUE have been essentially identical. In the models I study, CUE does not have heavy tails, as in Hansen, Heaton, and Yaron (1996). That does not necessarily mean that CUE is a perfect replacement for EL. Newey and Smith's (2004) formulas help give us an idea when it is safe to use CUE and when to use EL. First, it is instructive to write down the first order conditions for EL and CUE,[23]

$$\begin{aligned} EL &: \sum_i \pi_i G(z_i,\theta)\left(\sum_i \pi_i g(z_i,\theta)\, g(z_i,\theta)'\right)^{-1} \sum_i n^{-1} g(z_i,\theta) \\ CUE &: \sum_i \pi_i G(z_i,\theta)\left(\sum_i n^{-1} g(z_i,\theta)\, g(z_i,\theta)'\right)^{-1} \sum_i n^{-1} g(z_i,\theta). \end{aligned}$$

Where $G(z_i,\theta) = \partial g(z_i,\theta)/\partial\theta$. The difference between EL and CUE is that EL uses an efficient estimate of the weighting matrix, whereas CUE, like GMM, uses an inefficient estimate that places equal weight on each observation. In situations where the concern about bias in GMM does not come from poor estimation of the weighting matrix, CUE will be appropriate. On the other hand, if the weighting matrix is being poorly estimated, then EL will be preferable to CUE.

To obtain a more concrete indicator of whether CUE is acceptable, we need only look at Newey and Smith's (2004) formulas for the bias of GEL (reproduced in the appendix of this paper). The difference between EL and CUE is driven by skewness in the moments. The larger the skewness (and the fewer the number of observations one has), the larger will be the expected difference between CUE and EL. Cases where the skewness is relatively small are precisely the situations when mismeasurement of the weighting matrix will not be an issue. As the moments in most empirical applications tend not to be highly skewed (since researchers often take the log of highly skewed datasets), CUE will often be an acceptable alternative. In an analysis of, for example, raw income data, which will be highly skewed, CUE might not be appropriate.

5 Conclusion

This article reports an empirical case in which GMM provides misleading inference and Empirical Likelihood solves the problem. Furthermore, the problems with GMM are confirmed

[23] Newey and Smith (2004) report these first order conditions along with those for GEL and GMM.

in simulations. When data is generated to match the empirical sample, GMM is biased both with the full set of available moments and with a smaller set that requires fewer assumptions. While the full moment set provides better identification in the sense that standard errors are smaller, inference from GMM becomes poorer as the confidence intervals have coverage rates that are 18 to 30 percentage points too low. EL, on the other hand, suffers from none of these drawbacks. It is median unbiased, and its confidence intervals have the correct coverage rates. In large scale simulations, I find that analytic predictions of the bias of GMM and GEL are effective, especially for EL. this indicates that researchers need not rely on Monte Carlo analyses of the small sample bias of these estimators. As the number of weak instruments becomes large, GMM performs more poorly, but EL still provides good inference.

For practical work, CUE can be implemented with nearly identical computer code to that of GMM. A simple switch from GMM to CUE would improve inference in countless economic applications, including the dynamic panel setting studied above, linear instrumental variables estimation, estimation of financial models, and any other overidentified moment-based estimation.

A Appendix

A.1 Complete Characterization of the Bias

Following directly from Newey and Smith (2004), I here reproduce the second order bias terms.

As above, $g_i = g(z_i, \theta)$ is the $m \times 1$ moment function, z_i is one observation of data, and θ is the $p \times 1$ parameter. The identifying assumption is $E[g(z, \theta_0)] = 0$. The following terms will be used,

$$\begin{array}{cc}
G_i = \dfrac{\partial g(z_i, \theta)}{\partial \theta} & \Omega = E\left[g(\theta_0)\, g(\theta_0)'\right] \\
\Sigma = \left(G'\Omega^{-1}G\right)^{-1} & H = \Sigma G' \Omega^{-1} \\
P = \Omega^{-1} - \Omega^{-1} G \Sigma G' \Omega^{-1} & H_W = \left(G'WG\right)^{-1} G'W^{-1} \\
a_j = tr\left(\Sigma \mathbb{E} \partial^2 g_j(\theta_0) / \partial \theta^2\right) & \bar{\Omega}_{\theta_j} = E\left[\partial\left(g(\theta_0)\, g(\theta_0)'\right) / \partial \theta_j\right]
\end{array}$$

Where, in a_j, $g_j(\theta_0)$ is the jth element of $g(\theta_0)$.

The bias terms are then defined as follows,

$$\begin{aligned}
B_I &= H\left(-a + \mathbb{E}\left[G_i H g_i\right]\right) \\
B_G &= -\Sigma E\left[G_i' P g_i\right] \\
B_\Omega &= H E\left[g_i g_i' P g_i\right] \\
B_w &= -H \textstyle\sum_{j=1}^{p} \bar{\Omega}_{\theta_j} \left(H_W - H\right)' e_j
\end{aligned}$$

where e_j is the jth unit vector.

A.2 Computation

Computation of GEL and GMM was carried out using John Zedlewski's MatELike, available at

http://people.fas.harvard.edu/~jzedlews. MatELike also includes Siegfried M. Rump's Intlab optimization package, available at http://www.ti3.tu-harburg.de/rump/intlab. Replication files for the main results are available at http://people.fas.harvard.edu/~idew.

References

[1] Altonji, J.G. and L.M. Segal. 1996. "Small-Sample Bias in GMM Estimation of Covariance Structures." *Journal of Business & Economic Statistics*, 14(3), pp. 353–366.

[2] Angrist, J., G.W. Imbens and A. Krueger. 1999. "Jacknife Instrumental Variables Estimation." *Journal of Applied Econometrics*, 14(1), pp. 57–67.

[3] Angrist, J., and A. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics*, 106(4), pp. 979–1014.

[4] Anderson, T.W. and Hsiao, C. 1981. "Estimation of Dynamic Models with Error Components." *Journal of the American Statistical Association*, 76, pp. 598–606.

[5] Arellano, M. and S. Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *The Review of Economic Studies*, 58(2), pp. 277–297.

[6] Bekker, P.A. 1994. "Alternative Approximations to the Distributions of Instrumental Variable Estimators." *Econometrica*, 62(3), pp. 657–681.

[7] Binder, M., C. Hsiao and M.H. Pesaran. 2005. "Estimation and Inference in Short Panel Vector Autoregressions With Unit Roots and Cointegration." *Econometric Theory*, 21(4), pp. 795–837.

[8] Blundell, R. and S. Bond. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics*, 87(1), pp. 115–143.

[9] Brown, B.W. and W.K. Newey. 2002. "Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference." *Journal of Business & Economic Statistics,* 20(4), pp. 507–517.

[10] Carroll, C. D. 1997. "Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis." *The Quarterly Journal of Economics*, 112(1), pp. 1–55.

[11] Corcoran, S.A. 1998. "Bartlett Adjustment of Empirical Discrepancy Statistics." *Biometrika*, 85(4), pp. 967–972.

[12] Guggenberger, P. and R.J. Smith. 2005. "Generalized Empirical Likelihood Estimators and Tests Under Partial, Weak, and Strong Identification." *Econometric Theory*, 21(4), pp. 667–709.

[13] Han, C. and P.C.B. Phillips. 2005. "GMM with Many Moment Conditions." *Econometrica*, 74(1), pp. 147–192.

[14] Hansen, L. P., J. Heaton, and A. Yaron. 1996. "Finite-Sample Properties of Some Alternative GMM Estimators." *Journal of Business & Economic Statistics*, 14(3), pp. 262–280.

[15] Holtz-Eakin, D., W. K. Newey and H.S. Rosen. 1988. "Estimating Vector Autoregressions with Panel Data." *Econometrica,* 56(6), pp. 1371–1395.

[16] Horowitz, J.L. 1998. "Bootstrap Methods for Covariance Structures." *Journal of Human Resources*, 33, pp. 38–61

[17] Imbens, G.W. 1997. "One-Step Estimators for Overidentified Generalized Method of Moments Models." *Review of Economic Studies* 64(3), pp. 359–383.

[18] ______. 2002. "Generalized method of Moments and Empirical Likelihood." *Journal of Business & Economic Statistics*, 20(4), pp. 493–506.

[19] ______, R.H. Spady, and P. Johnson. 1998. "Information Theoretic Approaches to Inference in Moment Condition Models." *Econometrica*, 66(2), pp. 333–357.

[20] Kitamura, Y. and M. Stutzer. 1997. "An Information-Theoretic Alternative to Generalized Method of Moments Estimation." *Econometrica,* 65(4), pp. 861–874.

[21] Newey, W.K. and R.J. Smith. 2004. "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators." *Econometrica*, 72(1), pp. 219–255.

[22] Newey, W.K. and F. Windmeijer. 2005. "GMM with Many Weak Moment Conditions." CEMMAP Workig Paper No. CWP18/05.

[23] Owen, A. 1988. "Empirical Likelihood Ratio Confidence Intervals for a Single functional." *Biometrika*, 75(1), pp. 237–249.

[24] Qin, J. and J. Lawless. 1994. "Empirical Likelihood and General Estimating Equations." *The Annals of Statistics*, 22(1), pp. 300–325.

Table 1: Summary Statistics

Percentile	Age	Wealth ($k)	Income ($k)	Years Education
5	53	43	12	9
25	56	106	28	12
50	59	209	47	12
75	61	425	74	16
95	63	1,285	177	17
Mean	58	428	69	13

Working?		Retired?		Race	
Yes	49.6	No	37.0	White	91.2
No	50.4	Partially	14.0	Black	6.9
		Yes	48.9	Other	1.9

Summary statistics for the sample used in the empirical exercise. The data is drawn from the health and retirement survey, and households are kept only if they have observations for all seven biennial surveys from 1994 to 2006.

Table 2: Estimates of θ

Moments	EL	GMM
DIF	0.79	0.18
	(0.28)	(0.23)
	[0.66]	[0.16]
ALL	0.96	0.95
	(0.03)	(0.03)
	[0.42]	[0.48]
N	991	

Standard errors are in parentheses. EL standard errors are calculated with the standard asymptotic GMM estimator. Likelihood ratio test p-values are in brackets for EL, and J-statistic p-values for GMM. The Arellano-Bond and Blundell-Bond estimators are used for an AR(1) model for household wealth, allowing for the existence of fixed effects and measurement error

Table 3. Simulation Summary Statistics

Moments:	DIF				ALL			
Estimator:	GMM	EL	CUE	Analytic	GMM	EL	CUE	Analytic
Mean bias	-0.14	0.02	0.02	0	-0.06	0.00	0.00	0
Median bias	-0.12	0.00	0.00	0	-0.04	0.00	0.00	0
Std. Dev.	0.20	0.18	0.18	0.16	0.09	0.05	0.05	0.04
90% CI coverage	0.70	0.90	0.90	0.90	0.60	0.88	0.88	0.90
95% CI coverage	0.77	0.95	0.95	0.95	0.69	0.94	0.94	0.95

Summary statics from 20,000 simulations of the benchmark model. Each column gives results from a different estimator. GMM is two-step GMM using the identity matrix for weights in the first step. The confidence intervals for all estimators are estimated based on the standard GMM standard errors. The column labeled "Analytic" reports values that would be obtained using the standard first order asymptotic distributions of the estimators.

Table 4: Simulation results

Simulation:	*1a*	*1b*	*2*
Moments	Differences	All	Differences
Year Dummies?	Yes	Yes	No
σ_α	{0.11, 0.22, 0.44}	{0.11, 0.22, 0.44}	0.61
σ_ε	{0.27, 0.54}	{0.27, 0.54}	0.30
σ_ν	{0.17, 0.34}	{0.17, 0.34}	0.10
θ	0.81	0.81	0.6
N	991	991	762
R^2 :			
EL	0.717	-0.085	0.982
GMM	0.861	-0.360	-60.694
$\lvert B_I/SE\rvert$			
Min	0.0174	0.0398	0.0022
Median	0.0423	0.0768	0.0261
Max	0.2352	0.2110	0.1401
$\lvert GMM\ Bias/SE\rvert$			
Min	0.3225	0.2918	0.6510
Median	0.7283	0.7366	1.9104
Max	1.7348	2.6940	4.1851

Results of simulations varying the parameters in the underlying model. 1000 simulations were run for each parameter set. The rows with greek lettes give the values of the parameters used in the simulations for each column. N is the number of observations in each simulation. R^2 is the fraction of the variance of the Newey–Smith bias terms explained by the analytic formulas. $|B_I/SE|$ is the absolute value of the simulated bias for EL divided by the standard error.

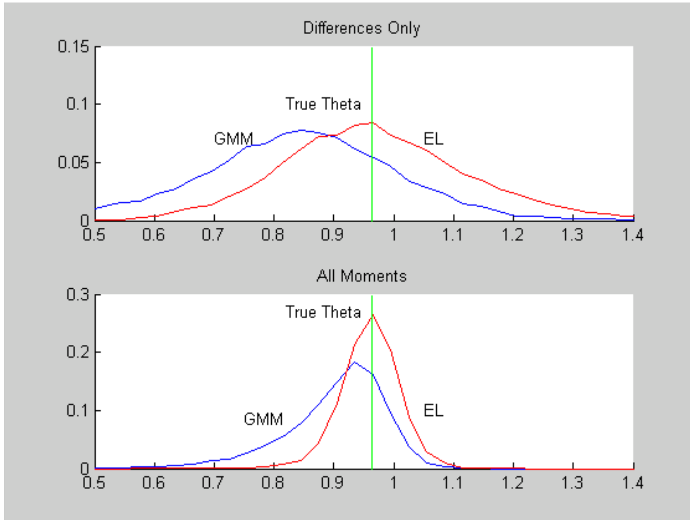

Figure 1: Histograms of estimates from simulated data. The top frame is for DIF, the bottom ALL. The true value of θ is 0.96.

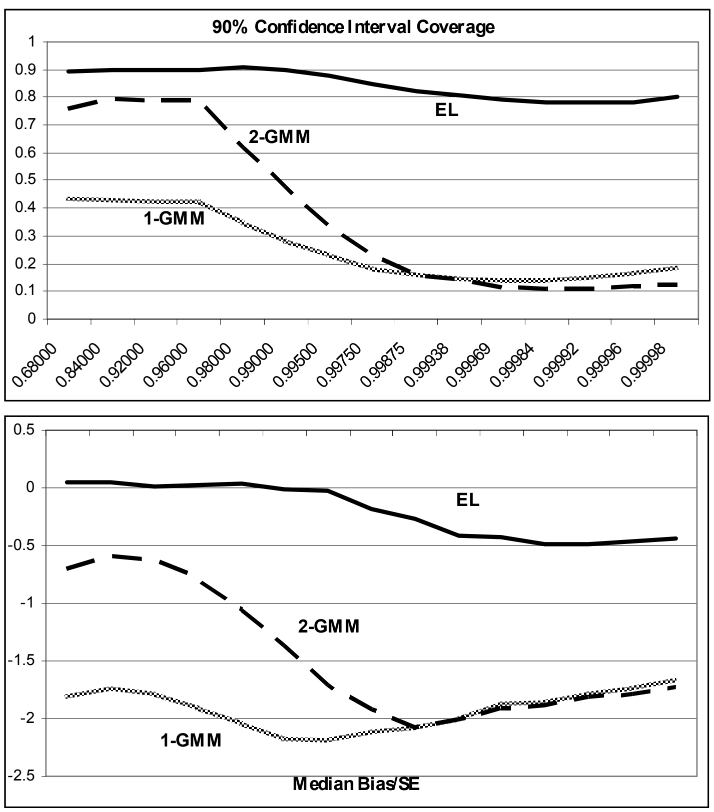

Figure 2: 90 percent confidence interval coverage rates and median bias scaled by the standard error for 1 and 2-step GMM and EL. The x-axis is θ, where each successive point on the scale represents reducing $(1 - \theta)$ by a factor of 2. Similar results are obtained with the 95 percent confidence interval or using CUE instead of EL.

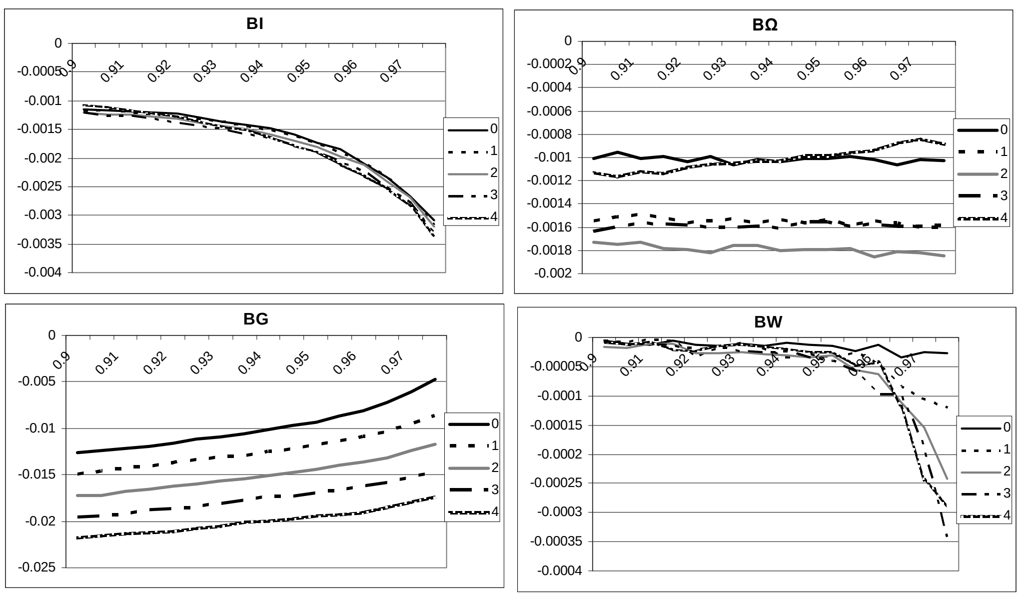

Figure 3: Plots of the four bias terms. Legends denote the number of extraneous (uninformative as $\theta \to 1$) moments included. The x-axis is the value of θ, the autoregressive parameter, and the y-axis is the magnitude of the bias term for $N = 991$. Each term is estimated from 1.6×10^6 simulations.

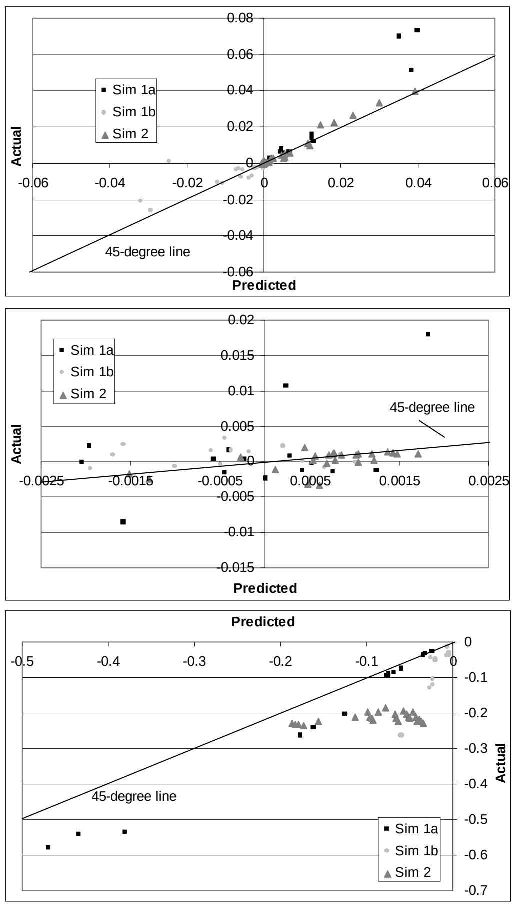

Figure 4: Scatter plots of actual vs. predicted biases. The top frame is B_I (the bias of EL), the middle B_Ω (the difference between EL and CUE), and the bottom the bias of GMM ($B_I + B_G + B_\Omega + B_W$). The solid line is the 45-degree line.