Household Wealth Dynamics and the Performance of Alternative Estimators

Or: How I learned to stop worrying and love empirical likelihood

Ian Dew-Becker

Harvard University

August, 2008

Introduction

- There is a growing literature arguing that EL is superior to GMM
- I present a simple empirical case in which EL outperforms GMM
- Next, I run simulations with data similar to the empirical sample
  - The problems with GMM also occur in the simulations
  - Analytic formulas for the bias of EL and GMM work fairly well
  - GEL estimators are all roughly identical in this case

- I work with the dynamic panel data model, $w_{it} = \alpha_i + \theta w_{it-1} + \varepsilon_{it}$
- A series of papers has noted GMM works poorly with this model
  - Highly overidentified, with weak instruments
  - E.g. Altonji and Segal (1998) and Blundell and Bond (1998)
- There are numerous suggestions for how to improve GMM
  - 1-step, principal components of weighting matrix, bootstrap, etc.
- Authors also argue for EL – Imbens (2002)

Empirical Framework

- The goal is to estimate a simple dynamic model for individual wealth:

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

    - $\alpha_i / (1 - \theta)$ is the equilibrium level of wealth, $E[\varepsilon] = E\left[\varepsilon_{it} \cdot w_{it-j}\right] = 0$

- I haven't seen anybody estimate a similar model – we don't know much about the dynamics of individual wealth

- If wealth follows a random walk with or without drift, $\theta = 1$; if people have some target wealth level, $\theta < 1$
- The model also allows to decompose the variance of measured wealth
    - Variance from steady states (α), annual shocks (ε), and measurement error (μ)

A simple consumption model

- W_t is wealth, $w_t = \log W_t$
- Suppose consumption is a function only of wealth and is scale independent $(C(\mu W) = \mu C(W))$, then $C = \mu W$
- With $C_t = \mu W_t$,

$$
\begin{aligned}
W_{t+1} &= (1 + R_{t+1})(1 - \mu) W_t \\
w_{t+1} &= w_t + \log(1 - \mu) + \log(1 + R_{t+1})
\end{aligned}
$$

- Buffer stock models say $\theta < 1$ with income also a state variable
- θ is the difference between the marginal and average propensities to consume from wealth

Identification and Inference

- OLS is inconsistent for dynamic panel models with fixed effects

$$\Delta w_{it} = \theta \Delta w_{it-1} + \Delta \varepsilon_{it}$$

  - Δw_{it-1} is correlated with $\Delta \varepsilon_{it}$ through ε_{it-1}

- Standard technique (Arellano and Bond, etc.) is to instrument for Δw with past values of w; this gives the "differences" moments,

$$E\left[w_{it-j}\left(\Delta w_{it} - \theta \Delta w_{it-1}\right)\right] = 0$$

  for all $j \geq 2$

  - These moments only require that ε is not autocorrelated
  - ε need not be homoskedastic or independent of α

- If α_i is also independent of ε_{it} for all i, t, then we have the "levels" moments

$$E\left[\Delta w_{it-1}\left(w_{it} - \theta w_{it-1}\right)\right] = 0$$

- In most economic applications, measurement error will be an issue; easy to correct here
- We just lag the instruments
- If we observe $z_{it} = w_{it} + v_{it}$, with v_{it} random measurement error, we have

$$E\left[z_{it-j}\left(\Delta z_{it} - \theta \Delta z_{it-1}\right)\right] = 0$$

for $j \geq 3$, and with stationarity,

$$E\left[\Delta z_{it-2}\left(z_{it} - \theta z_{it-1}\right)\right]$$

Estimation

Estimation

- Normally, estimation is performed with GMM
  - GMM has poor small sample properties, especially when highly overidentified
  - Bias is widely studied with the present model; especially bad with θ near 1
- Empirical likelihood (EL) is thought to have better bias properties
  - EL and GMM are both consistent and identical to the first order
  - EL is higher order efficient – dominates GMM (wrt terms that get small at rate N)
  - Inherits the characteristics of MLE
- The bias of GMM rises as more moments are added, but for EL it is unchanged
  - Important here since we have lots of available moments

- There are lots of proposed solutions to the GMM bias problem
  - Bootstrapping the bias (Horowitz, 1998)
  - Using 1-step GMM
  - Alternate estimators for the weighting matrix
- These methods involve throwing out information from the data
  - They also only fix problems associated with the weighting matrix
- EL just uses the data efficiently

Review of empirical likelihood

- Suppose the observed data forms the support for the true distributioin
- We assign probabilities π_i to each point
- We have a moment function g such that, $\mathbb{E}\left[g\left(z, \theta_0\right)\right] = 0$
- Then we maximize the empirical likelihood, $\sum_i \ln \pi_i$ such that $\sum_i \pi_i = 1$ and the moment conditions hold

$$\sum_i \pi_i g\left(z_i, \theta\right) = 0$$

  - With exact identification, $\pi_i = 1/n \; \forall i$, and EL replicates GMM
  - Overidentifying restrictions tell us what the probabilites must be, leads to greater efficiency

Generalized empirical likelihood

- $\sum_i \ln \pi_i$ is maximized when $\pi_i = 1/n$, so EL just minimizes the difference $\sum_i \ln n^{-1} - \sum_i \ln \pi_i = -\sum_i \ln\left(\pi_i / n^{-1}\right)$
- We could use a function other than the log to measure that difference
- Leads to GEL – using Cressie-Read discrepancy statistic
- Objective is

$$\frac{1}{\rho(\rho-1)} \sum_i n^{-1} \left[\left(\pi_i / n^{-1}\right)^{1-\rho} - 1 \right]$$

- EL is $\rho \to 0$, ET if $\rho = 1$, CUE is $\rho = 2$

The root of the problem

- First order conditions that characterize the estimators:
- GMM:

$$\left[\sum_i n^{-1} G_i\left(\hat{\theta}_{GMM}\right)\right]' \left[\sum_i n^{-1} g\left(z_i, \tilde{\theta}\right)' g\left(z_i, \tilde{\theta}\right)\right]^{-1} \bar{g}\left(\hat{\theta}_{GMM}\right) = 0$$

EL:

$$\left[\sum_i \hat{\pi}_i G_i\left(\hat{\theta}_{GMM}\right)\right]' \left[\sum_i \hat{\pi}_i g\left(z_i, \hat{\theta}\right)' g\left(z_i, \hat{\theta}\right)\right]^{-1} \bar{g}\left(\hat{\theta}_{GMM}\right) = 0$$

where $G_i\left(\theta\right) = \partial g\left(z_i, \theta\right)/\partial\theta$, $\bar{g}\left(\hat{\theta}_{EL}\right) = \left[\sum_i \hat{\pi}_i g\left(z_i, \hat{\theta}_{EL}\right)\right]'$ for EL and $\bar{g}\left(\hat{\theta}_{GMM}\right) = \left[\sum_i n^{-1} G_i\left(\hat{\theta}_{GMM}\right)\right]'$ for GMM

- GMM uses sample averages, while EL uses efficient estimators
- The biases arise from nonzero correlations between the terms in the FOC's – $E\left[X \cdot Y\right] = EX \cdot EY + Cov\left(X, Y\right)$
  - The efficiency of the EL estimates of the jacobian and second moments means they are asymptotically uncorrelated with $\bar{g}$.

Estimation

- I'll estimate the model with both EL and GMM
- EL is computationally much more difficult
    - Nothing a modern computer can't handle
    - See John Zedlewski's website: http://people.fas.harvard.edu/~jzedlews

Data and Estimates

Data

- I draw data from the RAND version of the Health and Retirement Survey (HRS)
  - Biennial survey from 1994 to 2006 (1992 dropped due to changes in the survey)
  - Samples households at or near retirement
- The HRS has extensive wealth data
  - Covers basically everything except for social security and employer-based pensions
  - Housing (but not 2^{nd} house), real estate, business, vehicles, financial wealth
  - Missing some wealth will bias θ down (if the other wealth is fixed, e.g. social security)
    - Intuition: if we only observe the most responsive margin of adjustment, we'll think adjustment is fast

Sample Selection

- In the preferred sample I drop households for 4 reasons
  - Change in number of members
  - Negative wealth
  - Change in wealth greater than a factor of 10
  - Initial wealth below $25k
- I vary minimum initial wealth; makes little difference

Summary Statistics

- Summary stats for nonhousing wealth (financial + businesses + real estate + vehicles), \$'000

Year	Mean	Median	Std. Dev.	Log Std. Dev.
1 (1994)	352	170	720	1.07
2 (1996)	376	180	700	1.14
3 (1998)	439	216	781	1.19
4 (2000)	552	235	1,990	1.24
5 (2002)	484	231	1,076	1.25
6 (2004)	534	240	1,431	1.26
7 (2006)	618	248	1,512	1.35

- These people are wealthy; consistent with being near retirement

Sample Characteristics

Percentile	Age	Wealth ($k)	Income ($k)	Years Education
5	54	30	15	10
25	59	94	32	12
50	63	215	52	13
75	67	500	82	16
95	72	1,551	201	17
Mean	63.28	479	79	13.63

Working?		Retired?		Race?	
Yes	50.13%	No	39.07%	White	94.88%
No	49.87%	Partially	15.99%	Black	3.67%
		Yes	44.94%	Other	1.44%

Results

Levels?	EL	GMM
No	0.89	-0.09
	(0.73)	(0.29)
Yes	0.96	0.95
	(0.03)	(0.03)

- Estimates of θ for nonhousing wealth (financial + businesses + real estate + vehicles), standard errors in parentheses
- The overidentifying restrictions do not reject the model; p-values are around 0.5 or higher.
    - The small standard errors along with no rejection by the OID statistic indicate this isn't badly misspecified
- Note that GMM would provide horrible inference in the first row

Alternative samples

	Diffs. Only		All	
Wealth Measure	EL	GMM	EL	GMM
Total	1.04	0.20	0.96	0.96
	(0.56)	(0.31)	(0.03)	(0.03)
Total. excl. IRA	0.79	0.18	0.96	0.95
	(0.28)	(0.23)	(0.04)	(0.04)
Nonhousing	0.89	-0.09	0.96	0.95
	(0.73)	(0.29)	(0.03)	(0.03)
Financial	2.31	0.96	0.95	0.95
	(1.93)	(0.49)	(0.04)	(0.04)

Simulations

Simulations

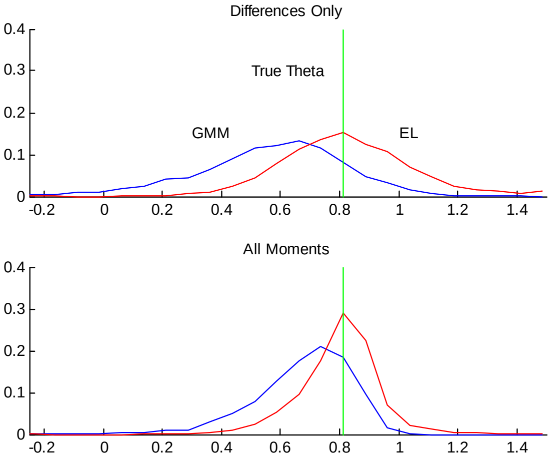
- One way to check the performance of the estimators is to simulate the model
- I work from the non-IRA sample
  - There may be rollovers from 401k's to IRA's
  - Using the EL estimates, I calculate the variances of the shocks from the residuals (estimates aren't unique)
- Basic results:
  - GMM is biased in the simulation, especially with just the differences moments
  - GMM 95% confidence interval coverage rate 70% or below
  - EL is nearly unbiased for a broad range of parameters
  - Mixed success predicting the bias

Simulation parameters

θ	0.81
σ_α	0.22
σ_ε	0.27
σ_μ	0.34
N	991

- All shocks are drawn from mean zero normal distributions
  - [θ is an old, inefficient estimate; variances of shocks are unchanged with new estimates of θ]

Simulation Results



Simulation Results

	Diffs. Only		All	
Percentile:	EL	GMM	EL	GMM
5	0.47	0.05	0.56	0.38
25	0.68	0.42	0.73	0.59
50	0.81	0.59	0.81	0.70
75	0.96	0.73	0.88	0.79
95	1.23	0.94	1.01	0.89
90% CI coverage	0.90	0.58	0.87	0.64
95% CI coverage	0.95	0.66	0.92	0.72

- True value of $\theta = 0.81$
- EL is median unbiased in both cases
- 95% CI is thinner for EL in both cases, CI's have correct coverage rate

Analytic Bias Predictions

Understanding the bias

- Newey and Smith (2004) give analytic formulas for the bias of GMM and EL
  - They solve for the terms in the estimator that get small at rate N
  - We don't know much about whether these are relevant in real-world samples
- I run simulations allowing the parameters, number of moments, and number of time periods to vary to test the predictions

$$\text{GMM:} \quad \left[\sum_i n^{-1} G_i \left(\hat{\theta}_{GMM}\right)\right]' \Omega^{-1} \bar{g}\left(\hat{\theta}_{GMM}\right) = 0$$

$$\text{EL:} \quad \left[\sum_i \hat{\pi}_i G_i \left(\hat{\theta}_{GMM}\right)\right]' \left[\sum \hat{\pi}_i g\left(z_i, \hat{\theta}\right)' g\left(z_i, \hat{\theta}\right)\right]^{-1} \bar{g}\left(\hat{\theta}_{GMM}\right) = 0$$

- Newey and Smith divide the small sample bias into 4 terms:
  - B_I – asymptotic bias of optimally weighted (infeasible) GMM
  - B_G – term due to error in estimating the derivatives of the moments
  - B_W – error from initial estimate for the weighting matrix
  - B_Ω – error from estimating the weighting matrix
- Bias:

$$\begin{array}{cc} \text{GMM:} & B_I + B_\Omega + B_G + B_W \\ \text{EL:} & B_I \\ \text{GEL:} & B_I + \left(1 + \frac{\rho}{2}\right) B_\Omega \end{array}$$

More simulations

- Simulations **1a** and **1b**:
  - Same θ as above, except error variances change:

θ	0.81
σ_α	{0.11, 0.22, 0.44}
σ_ε	{0.27, 0.54}
σ_μ	{0.17, 0.34}
N	991

- Simulations can include or exclude the levels moments
  - **1a** differences moments only
  - **1b** differences and levels moments

More simulations

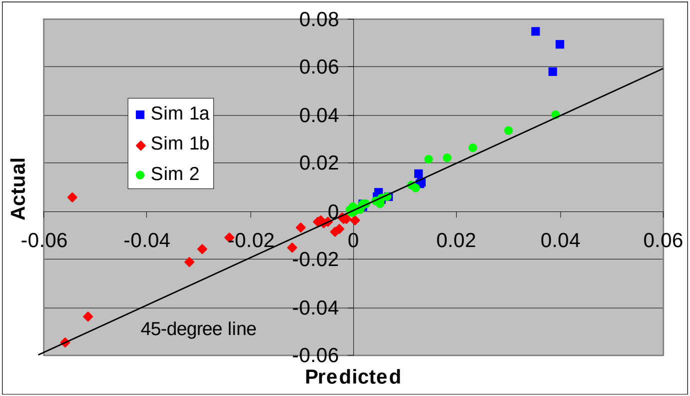
- Simulation **2**:
  - 7–11 periods of data
  - Differences moments only
  - Slightly different parameters:

θ	0.6
σ_α	0.61
σ_ε	0.30
σ_μ	0.10
N	762

More simulations

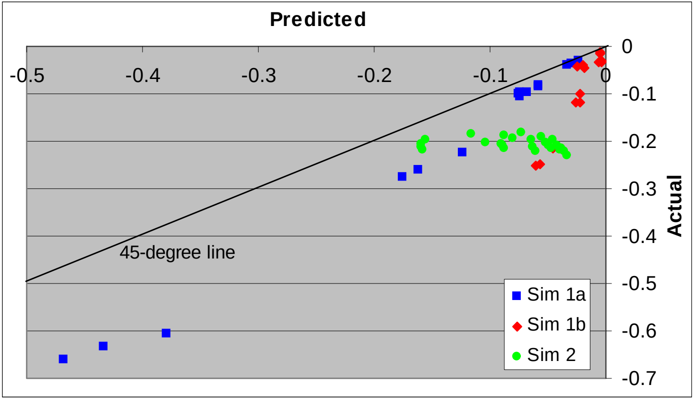
- Each set of simulations gives estimates for B_I (the bias of EL), B_Ω (difference between GEL estimators), and the bias of GMM
- B_Ω is driven by the skewness of the moments, which is small by construction
  - ε is normal, instruments are normal, product of independent normals has zero skewness
  - This might be a direction to expand the simulations

EL Bias



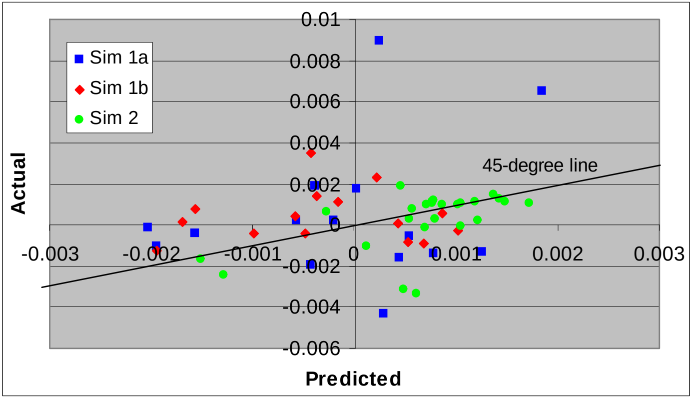
Sim 1a
Sim 1b
Sim 2
Actual
Predicted
45-degree line
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
0.02
0.04
0.06

GMM Bias



- Works ok for **1a** and **1b**, not well for **2;** bias is always larger than predicted

Differences between GEL estimators



- Note how small the scale is – simulation error dominates

Simulations

- The bias of GMM is substantial across a variety of parameter sets
  - The predictions don't work very well
- We have a good handle on the bias of EL

Summary

- We have an empirical case where GMM fails
- Simulations reproduce the result
  - GMM CI coverage far below nominal rate
- Newey–Smith expansions work well for EL