Complementary Bayesian Method of Moments Strategies

A. Ronald Gallant
Penn State University

Paper: http://www.aronaldg.org/papers/cb.pdf
Slides: http://www.aronaldg.org/papers/cbclr.pdf

Background

- The GMM (Hansen, 1982) minimum chi-square function (Neyman and Pearson, 1928) can be used to define a likelihood for Bayesian inference.

- Examples are Romeo (2007), Gallant and Hong (2007), Duan and Mela (2009), Guosheng (2009), Gallant (2016a), Gallant, Giacomini, and Ragusa (2017), and Gallant and Tauchen (2017, 2018).

- In the econometric literature the best known instance of the GMM representation of the likelihood is Chernozhukov and Hong (2003), although there it is used for frequentist inference.

Bayesian GMM

- For data $x = [x_1, \ldots, x_n]$ and parameter θ define

$$Z(x, \theta) = \sqrt{n}\, [W(x, \theta)]^{-\frac{1}{2}} [\bar{m}(x, \theta)]$$

- Moment conditions

$$\bar{m}(x, \theta) = \frac{1}{n} \sum_{t=1}^{n} m(x_t, \theta)$$

- Weighting matrix (use HAC for time series)

$$W(x, \theta) = \frac{1}{n} \sum_{t=1}^{n} [m(x_t, \theta) - \bar{m}(x, \theta)] [m(x_t, \theta) - \bar{m}(x, \theta)]'$$

- GMM representation of the likelihood

$$p^*(x|\theta) \propto \exp[-\frac{1}{2} Z'(x, \theta) Z(x, \theta)]$$

- Given prior $p^o(\theta)$, apply MCMC to $p^*(x|\theta) p^o(\theta)$

Concerns

- The GMM representation of the likelihood, as commonly used and as above, is missing a Jacobian term.

- The standard approach of assuming normality of the underlying random variable Z used to form the likelihood is suspect.

Strategy to Address Concerns

- Consider two similar examples:
  - A normality assumption is acceptable for the first
  - It is not acceptable for the second.

- For both:
  - Check if the adjustment term matters.
    (It doesn't).
  - Check if a proposed non-normality detection scheme works.
    (It does.)

Derivation of the Adjustment Term

- Bayesian analysis proceeds as if the data $x \in \mathcal{X}$ and parameter $\theta \in \Theta$ were random variables on $\mathcal{X} \times \Theta$ with joint density $p^o(x, \theta) = p^o(x \,|\, \theta) p^o(x \,|\, \theta) =$ likelihood x prior

- Suppose $z = Z(x, \theta)$ has density $\psi(z)$, then, under a semi-pivotal condition

$$p^*(z \,|\, \theta) = \psi(z)\Big|_{z=Z(x,\theta)}$$

where x is random and θ is fixed.

- If x has the same dimension as z, this implies

$$p^*(x \,|\, \theta) = \left|\text{det}[(\partial/\partial x')Z(x, \theta)]\right| \psi[Z(x, \theta)].$$

If the Dimension of x is Larger than θ

- Suppose one can find mappings $u = U(x, \theta)$ and $x = X(u, \theta)$ such that u has the same dimension as z and $z = Z[X(u, \theta), \theta]$.

- Then

$$p^*(x \,|\, \theta) = \left|\det \left\{(\partial/\partial u')Z[X(u, \theta), \theta]\right\}\right| \psi[Z(x, \theta)].$$

- Methods for constructing $u = U(x, \theta)$ and $x = X(u, \theta)$ are proposed in the paper.

- $p^*(x \,|\, \theta)$ is invariant to the choice of $u = U(x, \theta)$.

On Asymptotic Normality

- Define $F_n(z|\theta) = \int I[Z(x,\theta) \le z]\, p^o(x|\theta) dx$.

- The finite sample distribution of Z is
$$\Psi_n(z) = \int F_n(z|\theta)\, p^o(\theta)\, d\theta$$

- Dominated convergence theorem:
$$\lim_{n\to\infty} \Psi_n(z) = \int \lim_{n\to\infty} F_n(z|\theta)\, p^o(\theta)\, d\theta = \int \Phi(z)\, p^o(\theta)\, d\theta = \Phi(z).$$

- The density $\psi_n(z)$ converges to the normal if it is bounded and asymptotically equicontinuous (Sweeting, 1986)

Examples use CRRA Utility:

- Parameter: $\theta = (\beta, \gamma) =$ (discount factor, risk aversion)

- Data: $x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} \mathsf{lsr}_t \\ \mathsf{lcg}_t \end{pmatrix} = \begin{pmatrix} \text{log stock returns} \\ \text{log endowment growth} \end{pmatrix}$

- Moments: $m(x_t, \theta) = \begin{pmatrix} 1 \\ \mathsf{lsr}_{t-1} \\ \mathsf{lcg}_{t-1} \end{pmatrix} [1 - \exp(\log \beta - \gamma \mathsf{lcg}_t + \mathsf{lsr}_t)]$

- Adjustment: $\text{adj}(x, \theta) = 4(1-e)^2 \left| \frac{1-\tanh\left(\frac{1}{4} z_1\right)}{1-\left[\tanh\left(\frac{1}{4} z_1\right)\right]^2} \right|$

Data Generating Processes

- A data generating process has the form $p^o(x_t \,|\, \rho, \theta)p^o(\rho)p^o(\theta)$ where ρ are general equilibrium nuisance parameters and $\theta = (\beta, \gamma)$.

- First example is a Lucus exchange economy where endowment growth follows an autoregression with parameters $\rho = (\mu, \alpha, \sigma)$. Mean values of the prior are $\bar{\rho} = (0, 0.95, 0.2)$ and $\bar{\theta} = (0.95, 12.5)$. Standard deviations are (0.0, 0.01, 0.01) and (0.01, 2.0).

- The second example uses a non-parametrically extracted stochastic discount factor, data on corporate profits, and data on GDP from Gallant and Tauchen (2018). A trivariate VAR for $[\log(SDF), \log(GDP) - \log(CP), \Delta \log(CP)]$ is fit to these data. The estimated parameters from this fit determines $p^o(\rho)$; $p^o(\theta)$ has mean $\bar{\theta} = (0.9532, 24.5030)$ and standard deviations (0.547, 28.2425). Given a (ρ, θ) draw from the prior, the return to corporate profits is computed analytically from the VAR and the SDF is transformed to endowment growth by inverting the formula for the SDF implied by CRRA utility.

Assessment of Adjustment Term

- Draw from the prior $N = 1000$ times and retain that (ρ, θ) for which the max minus min adjustment is the largest over an exhaustive θ grid.

- For that (ρ, θ) evaluate the log likelihood and the log adjustment over the θ grid.

- Determine by inspection if the adjustment is large enough to affect the accept/reject decision of an MCMC trial.

- For the worst case (ρ, θ) data, check with MCMC.

- Worst case for Example 1 with $n = 100$ (next slide) occurs for $(\rho, \theta) = (0, 0.9410, 0.01000, 0.9542, 15.00)$ at $\theta = (0.97, 0.50)$

Table 3. Exchange Economy Likelihood and Adjustment, $n = 100$

Log Likelihood

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	-211.99	-16640.04	-170.40	-39.47	-15.89	-8.78	-6.06	-4.92	-4.50	-4.45	-4.61
10.45	-155.11	-13398.32	-126.74	-29.62	-12.51	-7.39	-5.46	-4.70	-4.47	-4.53	-4.75
20.40	-109.32	-10306.79	-91.12	-21.81	-9.87	-6.33	-5.03	-4.58	-4.51	-4.66	-4.92
30.35	-73.48	-7474.01	-62.69	-15.78	-7.88	-5.56	-4.76	-4.55	-4.61	-4.82	-5.12
40.30	-46.57	-5002.77	-40.72	-11.29	-6.43	-5.05	-4.63	-4.61	-4.76	-5.02	-5.33
50.25	-27.68	-2978.95	-24.59	-8.12	-5.46	-4.76	-4.64	-4.74	-4.97	-5.25	-5.56
60.20	-15.73	-1462.56	-13.67	-6.07	-4.91	-4.69	-4.76	-4.95	-5.21	-5.50	-5.80
70.15	-9.34	-484.22	-7.26	-4.97	-4.72	-4.80	-4.98	-5.22	-5.49	-5.78	-6.06
80.10	-7.08	-58.66	-4.61	-4.70	-4.88	-5.08	-5.30	-5.55	-5.81	-6.07	-6.33
90.05	-8.10	-51.30	-5.04	-5.16	-5.33	-5.52	-5.72	-5.93	-6.16	-6.39	-6.60
100.00	-12.42	-508.94	-8.15	-6.29	-6.07	-6.10	-6.21	-6.36	-6.54	-6.72	-6.89

Log Adjustment

γ/β	0.80	0.82	0.84	0.86	0.88	0.90	0.91	0.93	0.95	0.97	0.99
0.50	1.78	1.78	1.78	1.79	1.86	2.01	2.21	2.42	2.63	2.81	2.95
10.45	1.78	1.78	1.78	1.80	1.90	2.08	2.29	2.50	2.70	2.87	3.00
20.40	1.78	1.78	1.78	1.82	1.96	2.15	2.37	2.59	2.78	2.93	3.06
30.35	1.78	1.78	1.78	1.86	2.03	2.24	2.47	2.67	2.85	2.99	3.11
40.30	1.79	1.78	1.79	1.92	2.12	2.35	2.56	2.76	2.92	3.06	3.16
50.25	1.81	1.78	1.81	2.00	2.23	2.46	2.66	2.85	3.00	3.12	3.20
60.20	1.87	1.78	1.87	2.13	2.37	2.58	2.77	2.93	3.07	3.17	3.25
70.15	2.05	1.78	2.01	2.30	2.52	2.71	2.88	3.02	3.14	3.23	3.30
80.10	2.43	1.78	2.30	2.51	2.68	2.84	2.98	3.11	3.21	3.29	3.34
90.05	3.04	6.38	2.75	2.77	2.87	2.98	3.09	3.20	3.28	3.34	3.38
100.00	3.84	15.45	3.37	3.06	3.06	3.12	3.20	3.28	3.35	3.39	3.42

Table 5. Exchange Economy Estimates

	No Adjustment			Adjustment		
Parameter	Mean	Mode	Sdev	Mean	Mode	Sdev
			$n = 50$, tight prior			
β	0.9530	0.9550	0.00513	0.9540	0.9559	0.00510
γ	13.650	12.365	1.4849	13.603	12.233	1.4800
			$n = 100$, tight prior			
β	0.9595	0.9603	0.00422	0.9604	0.9611	0.00428
γ	13.603	12.233	1.4800	13.994	12.994	1.3948
			$n = 1000$, tight prior			
β	0.9312	0.9315	0.00242	0.9442	0.9445	0.00169
γ	14.995	14.673	0.8594	14.900	14.539	0.7894
			$n = 50$, loose prior			
β	0.9177	0.9561	0.03884	0.9228	0.9569	0.03904
γ	31.145	12.025	12.893	31.725	11.939	14.016
			$n = 100$, loose prior			
β	0.9416	0.9614	0.03017	0.9449	0.9623	0.03022
γ	33.272	13.301	13.402	34.613	13.189	14.213
			$n = 1000$, loose prior			
β	0.9280	0.9296	0.00493	0.9433	0.9441	0.00282
γ	16.539	15.414	2.0841	15.870	14.912	1.8355

Corporate Profits Example

- Adjustment term results for the discounted corporate profits example are analogous.

- We next consider the effect of non-normality and how to detect non-normality.

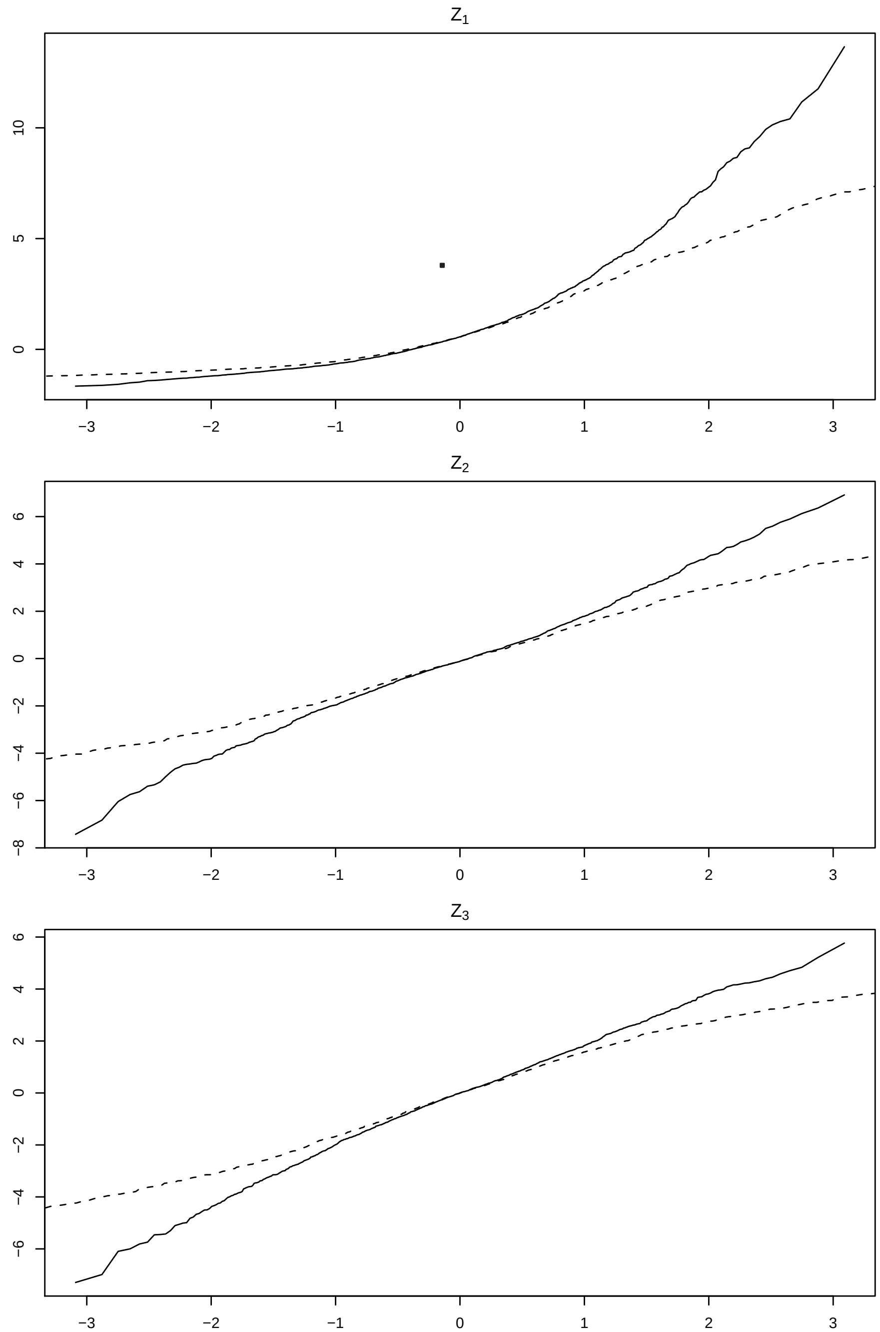

Figure 2. Corporate Profits Q-Q Plots. Solid line normal; dashed t(3).

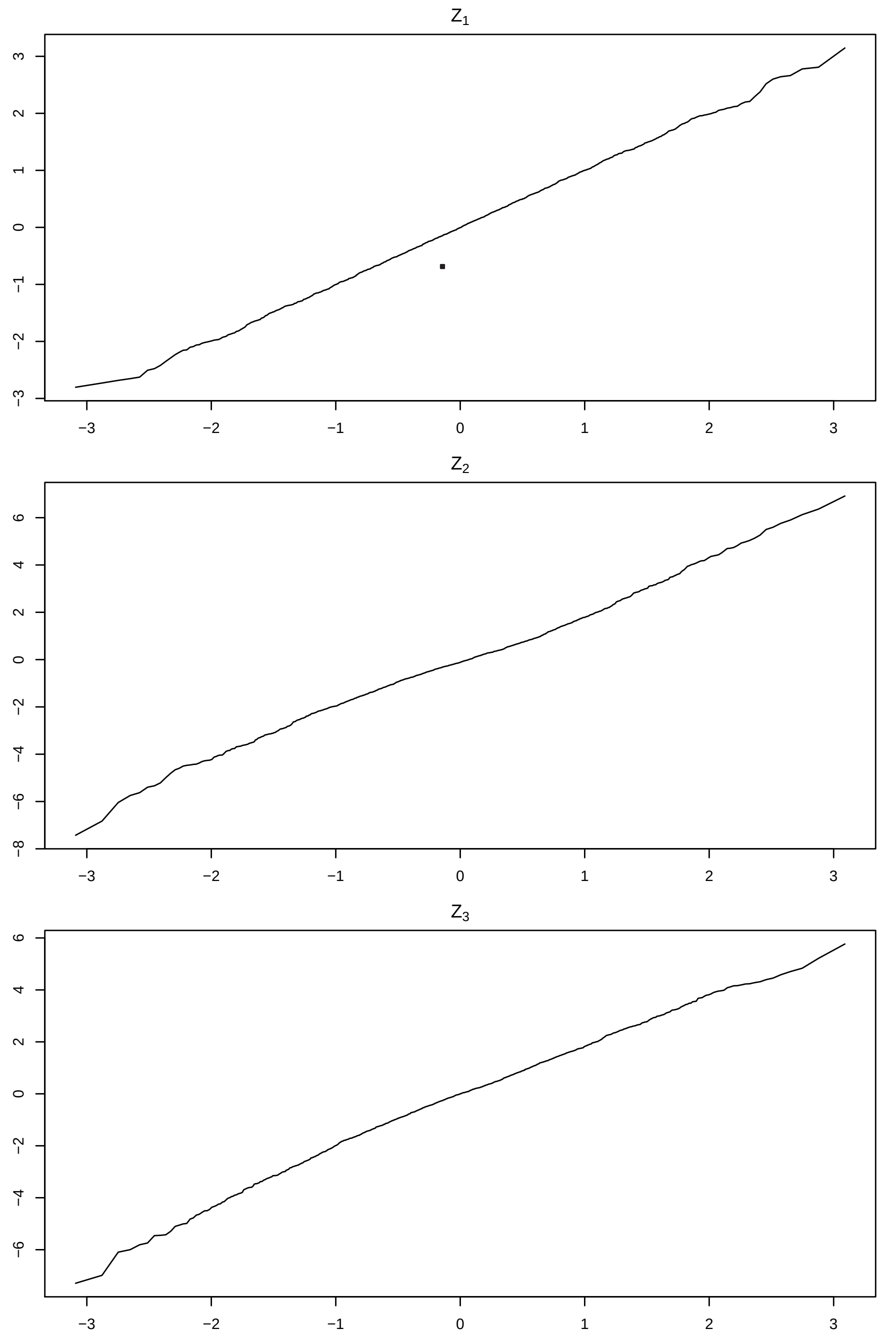

Figure 3. Transformed Corporate Profits Q-Q Plots. Solid line normal.

Table 11. Discounted Cash Flow Estimates

	No Adjustment			Adjustment		
Parameter	Mean	Mode	IQR	Mean	Mode	IQR
			$n = 50$			
β	0.9328	0.9556	0.05590	0.9328	0.9559	0.05582
γ	39.595	29.902	17.422	42.082	31.859	18.922
			$n = 100$			
β	0.9311	0.9541	0.05653	0.9311	0.9538	0.05666
γ	30.151	25.574	8.6111	31.737	26.849	9.3107
			$n = 1000$			
β	0.9310	0.9556	0.05631	0.9305	0.9528	0.05669
γ	40.288	28.940	4.3800	37.489	25.680	5.7771
			$n = 50$, transformed Z			
β	0.9327	0.9558	0.05562	0.9327	0.9574	0.05606
γ	35.500	29.562	18.321	39.249	38.540	18.720
			$n = 100$, transformed Z			
β	0.9315	0.9539	0.05652	0.9308	0.9533	0.05677
γ	26.797	25.657	8.4857	28.833	29.564	8.3368
			$n = 1000$, transformed Z			
β	0.9327	0.9604	0.05570	0.9322	0.9567	0.05606
γ	32.624	27.330	3.0234	35.546	28.259	5.0233

Complementary Methods

- Use an explicit likelihood $f(x \,|\, \rho)$ that has a non-parametric interpretation such as a sieve.

- Estimate subject to

$$0 = \bar{g}(\rho, \theta) = \int \bar{m}(x, \theta)\, f(x|\rho)\, dx$$

- The difficulty is that the parameter space

$$\{(\rho, \theta) \in \mathcal{R} \times \Theta \,|\, 0 = \bar{g}(\rho, \theta)\}$$

  has Lebesgue measure zero when $\bar{m}(x, \theta)$ is over identified.

- On this see Bornn, Luke, Neil Shephard, and Reza Solgi (2018), "Moment Conditions and Bayesian Nonparametrics," Journal of the Royal Statistical Society, Series B 81(1), 5–43.

Approximate Method: λ-prior

- Use a sieve, e.g. SNP, for $f(x \mid \rho)$.

- Use λ-prior

$$p_\lambda(\rho,\theta) \propto p^o(\theta) \times p(\rho) \times \exp\left[-\lambda \frac{n}{2} \bar{g}'(\rho,\theta)\bar{g}(\rho,\theta)\right]$$

- By trial and error find the largest λ such that MCMC draws mix.

- Reference: Gallant, Hong, Leung, and Lee (2019).

Non-normality Detection

- Discretize the prior $p^o(\theta)$ using a quadrature rule to obtain support points θ_i^o and weights p_i

- For each support point θ_i^o of the quadrature rule apply the λ-prior method with prior

$$p_\lambda^i(\rho, \theta) \propto p^o(\theta) \times \exp\left[-\lambda \frac{n}{2} \bar{g}'(\rho, \theta_i^o)\bar{g}(\rho, \theta_i^o)\right]$$

- Let ρ_i be the corresponding posterior modes.

- Draw z by drawing i with probability p_i, draw x from $f(x \mid \rho_i)$, and put $z = Z(x, \theta_i)$.

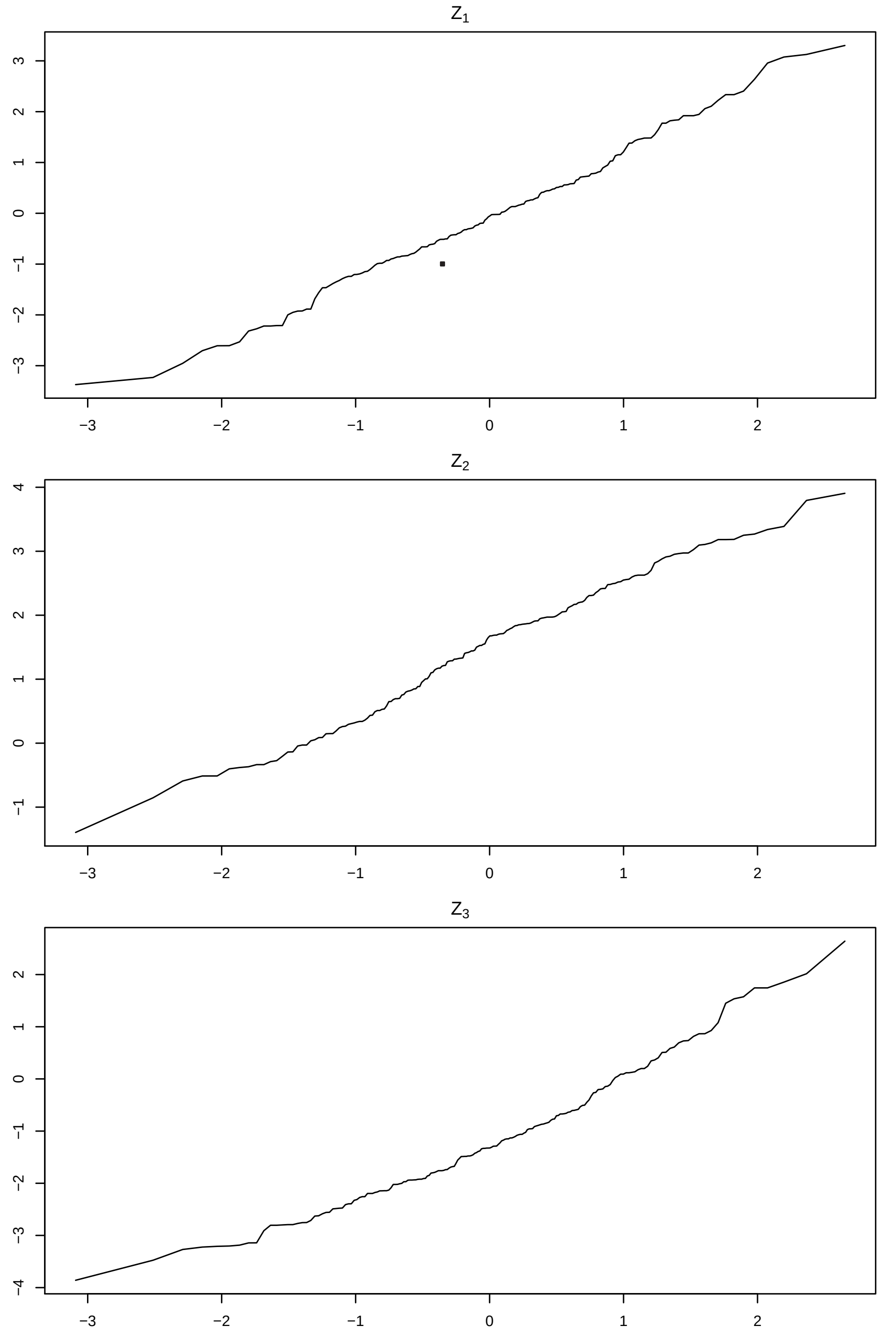

Figure 4. Complementary Q-Q Plots for Exchange Economy

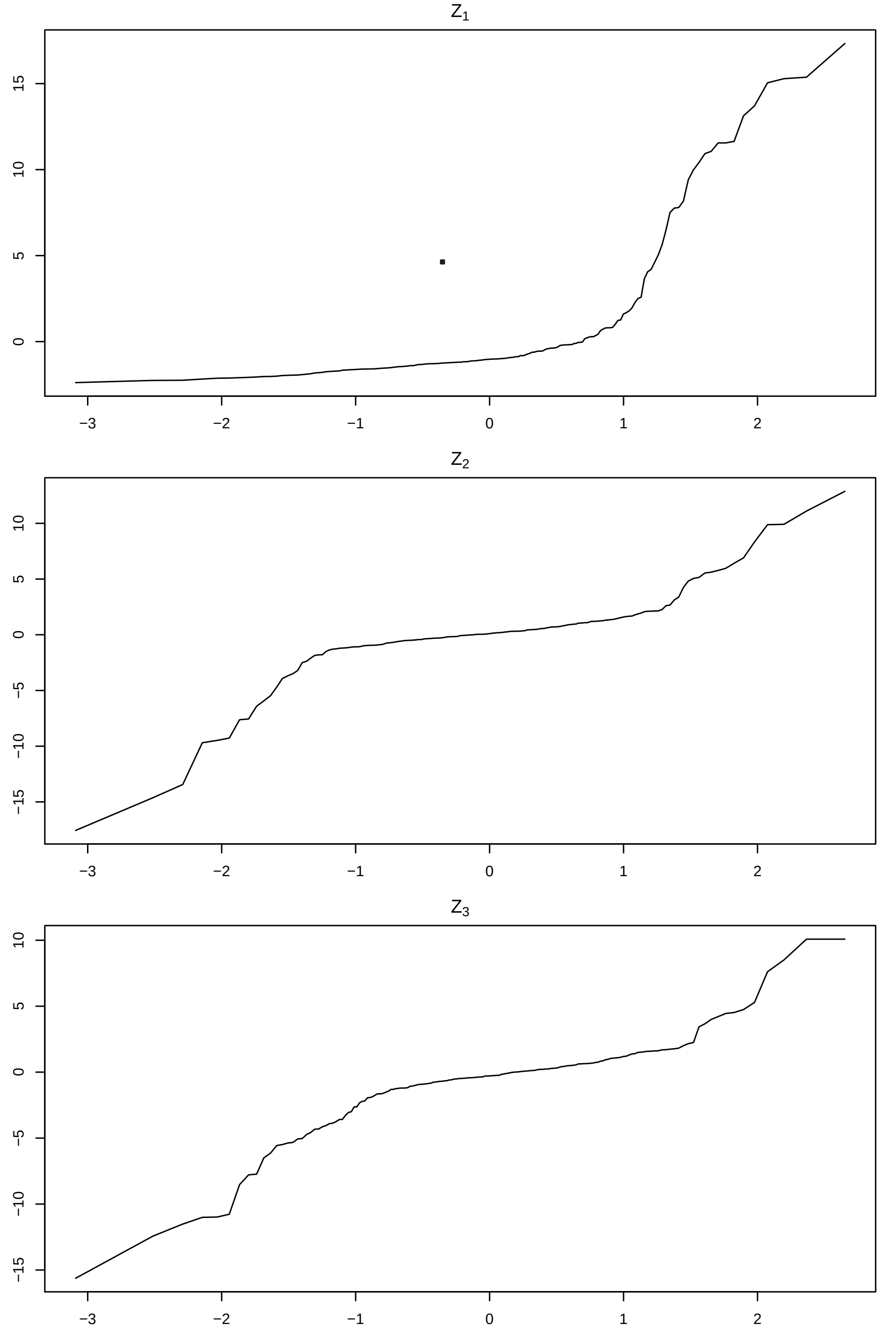

Figure 5. Complementary Q-Q Plots for Discounted Corporate Profits

Conclusion

- Addressed are two issues that arise in application of the GMM representation of the likelihood in Bayesian inference: (1) a missing Jacobian term and (2) a normality assumption.

- Practicable methods for addressing these two problems in an application are proposed and illustrated by application to an endowment economy whose representative agent has CRRA utility.

- While the main contribution is the proposed methodology, the illustration does provide some interesting anecdotal evidence that is consistent with the anecdotal evidence in Gallant (2016b): violation of (1) or (2) has less serious consequences in applications than one might expect for moderately large sample sizes.