Bayesian Inference Using the EMM Objective Function: With Application to the Dominant Generalized Blumenthal-Getoor Index of an Itô Semimartingale

A. Ronald Gallant

Penn State University

George Tauchen

Duke University

Paper: http://www.aronaldg.org/papers/hfemm.pdf

Slides: http://www.aronaldg.org/papers/hfemmclr.pdf

Bayesian Inference Using the EMM Objective Function

Changes to frequentist EMM required:

- Need a proper prior

- Use either a continuously updated weighting matrix or a static weighting matrix obtained by averaging over the prior.

- Nead to determine the distribution of the mean score either by simulation or a CLT. Due to averaging over the prior, issues of uniform convergence do not arise.

- Need a Jacobian term.

Application to the Dominant Index

	Posterior		95% Credibility	
Moneyness	Mean	Mode	Interval	
IV1	1.0134	1.0073	1.0026	1.0257
IV2	1.1266	1.1231	1.1125	1.1556
IV3	1.3041	1.3090	1.2915	1.3097
IV4	1.5586	1.5617	1.5398	1.5737
IV5	1.7672	1.7706	1.7438	1.7902
IV6	1.8240	1.8210	1.8023	1.8664
IV7	1.8808	1.8808	1.8807	1.8810
Futures	1.8808	1.8808	1.8799	1.8812

The Setting

- The standard jump-diffusion model used for modeling many stochastic processes is an Itô semimartingale given by the following differential equation

$$dX_t = \sigma_{t-} dW_t + dY_t, \tag{1}$$

- W_t is a Brownian motion

- Y_t is an Itô semimartingale process of pure jump type

Dominant (Brownian) Piece

- At high-frequencies, provided σ_t does not vanish, the dominant component of X_t is its continuous martingale component

$$\sigma_{t-} dW_t$$

- At these frequencies the increments of X_t in (1) behave like scaled and independent Gaussian random variables.

$$\frac{1}{\sqrt{h}}(X_{t+h} - X_t) \xrightarrow{\mathcal{L}} \sigma_t\,(B_{t+1} - B_t), \tag{2}$$

as $h \to 0$

Implications

There are two distinctive features of the convergence in (2):

1. The scaling factor of the increments on the left side of (2) is the square-root of the length of the high-frequency interval.

2. The limiting distribution of the (scaled) increments on the right side of (2) is mixed Gaussian (the mixing given by σ_t).

Only Jumps, Small and Large

- The pure jump diffusion model is an Itô semimartingale given by the following differential equation

$$dX_t = \sigma_{t-} dS_t + dY_t \tag{3}$$

- σ_t and Y_t are processes with càdlàg paths

- Y_t is a pure jump process

- S_t is a symmetric stable process with a characteristic function given by

$$\log \left[\mathbb{E} \left(e^{iuS_t} \right) \right] = -t|cu|^{\beta}, \quad \beta \in [0, 2] \tag{4}$$

- If $\beta = 2$, then X_t becomes the jump diffusion model (1)

Role of the Big Jump Process Y_t

- When $\beta < 2$, X is of pure-jump type, Y_t in (3) plays the role of a "residual" jump component at high frequencies

- Y_t can have dependence with S_t (and σ_t), therefore X_t does not inherit the tail properties of the stable process S_t.

$$Y \stackrel{\mathcal{D}}{=} X - S$$

- Under reasonable assumptions the local result (2) extends:

$$h^{-1/\beta}(X_{t+h} - X_t) \xrightarrow{\mathcal{L}} \sigma_t \left(S_{t+1} - S_t\right) \tag{5}$$

as $h \to 0$

Inference Strategy

- View the data as doubly indexed:

$$X_{i,t} \in [t-1, t] = \text{one day}$$

$i = 1, 2, \ldots, n$ within day index, $t = 1, 2, \ldots, T$ indexes days

- Use fill-in theory to indicate what to expect for

$$\Delta_i X = X_{i,t} - X_{i-1,t}$$

as $n \to \infty$ for $X_{i,t} \in [t-1, t]$

- Use multiple spans $[t-1, t]$ to improve estimation accuracy, $t = 1, 2, \ldots, T$

De-Volatizing Returns

- Proceed as in the locally Gaussian case (1) by using a bi-power variation estimate of local scale.

- Estimate σ_t locally by a leave-one-out block estimate $\widehat{V}_j^n(i)$ that has k_n increments within each block j and then divide the high-frequency increments by this estimate.

- The procedure is both automatically self-scaling and valid in the locally β–stable situation of (3).

Estimating Local Scale

- One divides the interval $[0, 1]$ into blocks, each of which contains k_n increments, for some deterministic sequence $k_n \to \infty$ with $k_n/n \to 0$.

- On each of the blocks the local estimator of σ_t^2 is given by

$$\widehat{V}_j^n = \frac{\pi}{2} \frac{n}{k_n - 1} \sum_{i=(j-1)k_n+2}^{jk_n} |\Delta_{i-1}^n X||\Delta_i^n X|, \tag{6}$$

$j = 1, \ldots, \lfloor n/k_n \rfloor$

- $\widehat{V}_j^n$ is the bi-power variation estimator proposed by Barndorff-Nielson and Shephard for measuring the quadratic variation of the diffusion component of X.

Re-Scaling

The scaling of every high-frequency increment is done after adjusting $\widehat{V}_j^n$ to exclude the contribution of that increment in its formation

$$\widehat{V}_j^n(i) = \begin{cases} \frac{k_n-1}{k_n-3}\widehat{V}_j^n - \frac{\pi}{2}\frac{n}{k_n-3}|\Delta_i^n X||\Delta_{i+1}^n X|, \\ \quad \text{for } i = (j-1)k_n + 1, \\ \frac{k_n-1}{k_n-3}\widehat{V}_j^n - \frac{\pi}{2}\frac{n}{k_n-3}\left(|\Delta_{i-1}^n X||\Delta_i^n X| + |\Delta_i^n X||\Delta_{i+1}^n X|\right), \\ \quad \text{for } i = (j-1)k_n + 2, ..., jk_n - 1, \\ \frac{k_n-1}{k_n-3}\widehat{V}_j^n - \frac{\pi}{2}\frac{n}{k_n-3}|\Delta_{i-1}^n X||\Delta_i^n X|, \\ \quad \text{for } i = jk_n. \end{cases} \tag{7}$$

Jump Removal

- De-volatilized increment:

$$\frac{\Delta_i^n X}{\sqrt{\widehat{V}_j^n(i)/n}}, \tag{8}$$

- Retain only the increments that do not contain big jumps:

$$\frac{\Delta_i^n X}{\sqrt{\widehat{V}_j^n(i)/n}} \mathbf{1}\left(|\Delta_i^n X| \leq \alpha \sqrt{\widehat{V}_j^n} n^{-\varpi} \right), \tag{9}$$

  $\alpha > 0$, $\varpi \in (0, 1/2)$.

- The time-varying jump threshold accounts for the time-varying σ_t.

CDF of Adjusted Data

- The empirical distribution of these increments is

$$\widehat{F}_n(\tau) = \frac{1}{N^n(\alpha,\varpi)} \sum_{j=1}^{\lfloor n/k_n \rfloor} \sum_{i=(j-1)k_n+1}^{(j-1)k_n+m_n} 1\left\{ \frac{\Delta_i^n X}{\sqrt{\widehat{V}_j^n(i)/n}} \leq \tau \right\} 1_{\left\{|\Delta_i^n X| \leq \alpha\sqrt{\widehat{V}_j^n} n^{-\varpi}\right\}} \quad (10)$$

- The divisor is the total number of retained increments

$$N^n(\alpha,\varpi) = \sum_{j=1}^{\lfloor n/k_n \rfloor} \sum_{i=(j-1)k_n+1}^{(j-1)k_n+m_n} 1\left(|\Delta_i^n X| \leq \alpha\sqrt{\widehat{V}_j^n} n^{-\varpi} \right) \quad (11)$$

- $\widehat{F}_n(\tau)$ is simply the empirical CDF of the de-volatilized increments that do not contain large jumps.

- In the jump-diffusion case of (1), $\widehat{F}_n(\tau)$ should be approximately the CDF of a standard normal random variable.

Limit Theorem

THEOREM 1 For either the processes (1) or (3), assume the block size grows at the rate

$$k_n \;\sim\; n^q, \quad \text{for some } q \in (0,1), \tag{12}$$

and $m_n \to \infty$ as $n \to \infty$. Then if $\beta \in (1,2]$,

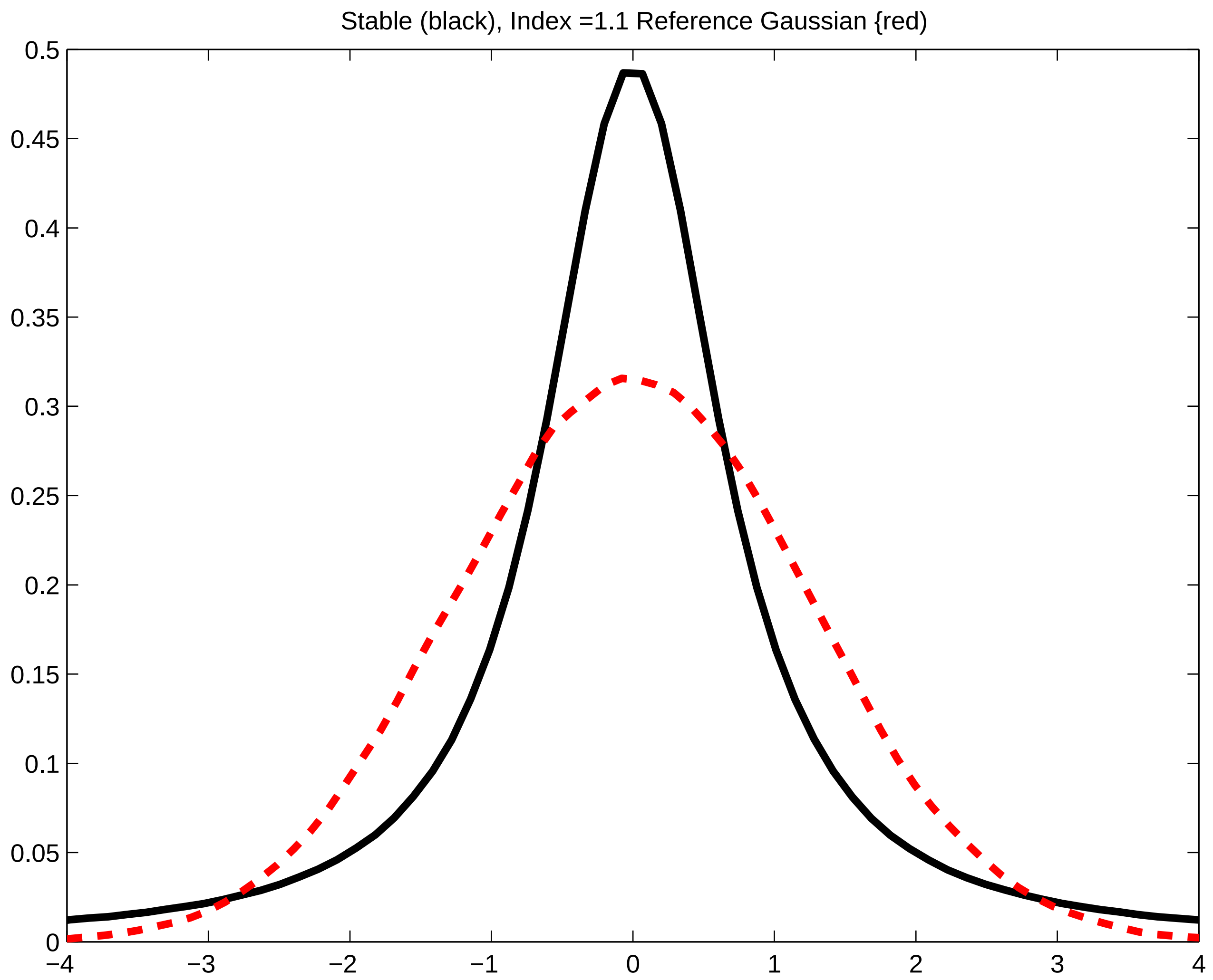
$$\boxed{\widehat{F}_n(\tau) \;\xrightarrow{\mathbb{P}}\; F_\beta(\tau), \quad \text{as } n \to \infty} \tag{13}$$

uniformly in τ over compact subsets of $\mathbb{R}$

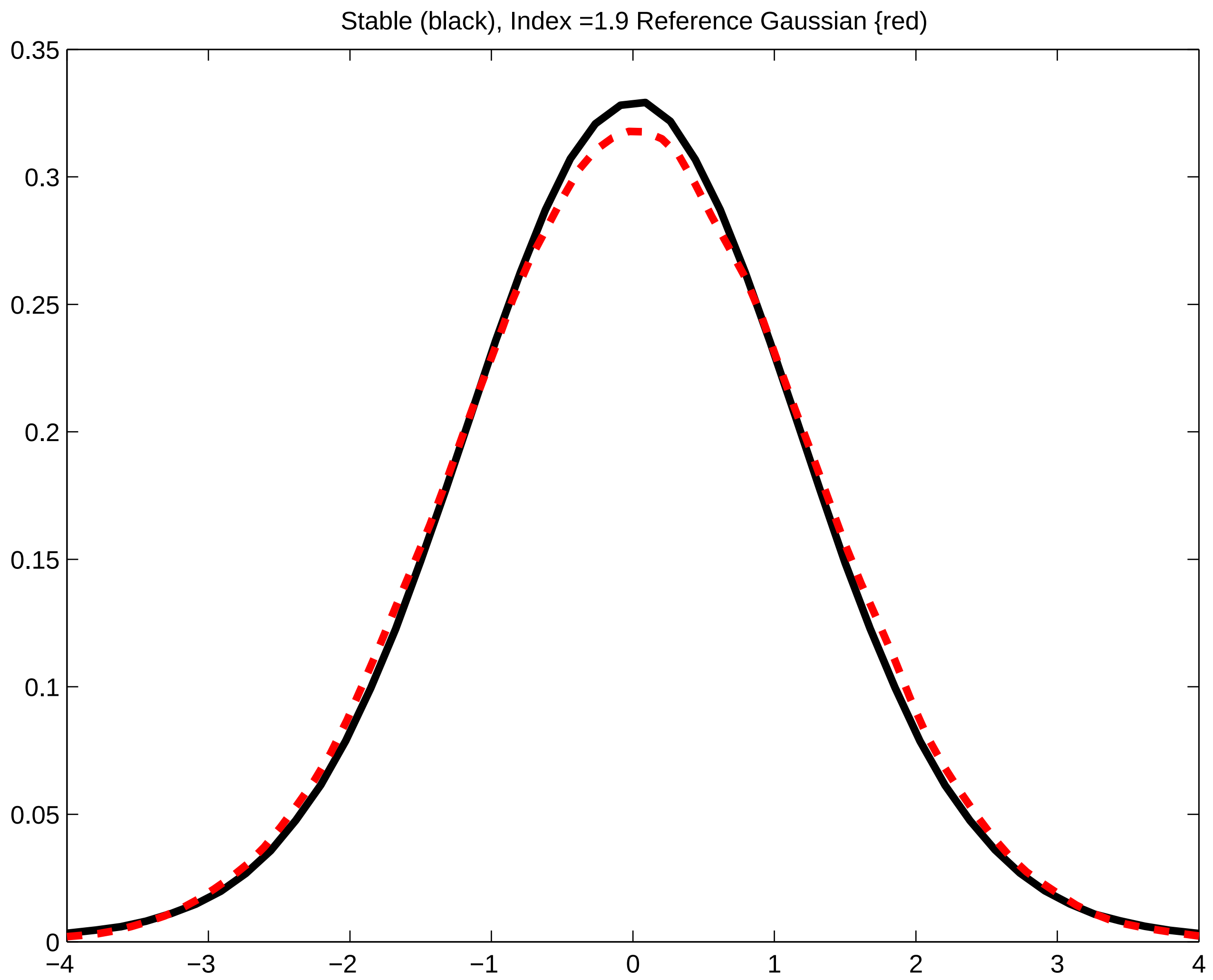
$F_\beta(\tau)$ is the CDF of $\sqrt{\frac{2}{\pi}}\frac{S_1}{\mathbb{E}|S_1|}$ (S_1 is the value of the β-stable process S_t at time 1) and $F_2(\tau)$ is the CDF of a standard normal.

The limit result in (13) shows that when S_t is stable with $\beta < 2$, $\widehat{F}_n(\tau)$ estimates the cdf of a β-stable random variable.

Predicted Density: $\beta = 1.10$



Predicted Density: $\beta = 1.90$



Data: S&P E-Mini Options and Futures

- We use the very skillfully assembled options data set developed and in Andersen, Bondarenko, Gonzalez-Perez (2013) for the purpose of decomposing and analyzing the CBOE volatility index (VIX).

- Hereafter, we refer to this as the ABG data.

- The data set was years in construction, and it provides a unique set of very clean high frequency observations of index options classified by moneyness.

Moneyness

- At expiration put options finish in-the-money (ITM), out-of-the-money (OTM), or at-the-money (ATM), presuming European style options.

- Before expiration a measure of how far an option is in or out of the money is useful because one might posit that options will load on different risk factors depending on moneyness.

- ABG moneyness metric is

$$m = \frac{\log(K/F)}{\sigma_{BS}\sqrt{\tau}}$$

  where

$$\begin{aligned} K &= \text{strike price} \\ F &= \text{forward price} \\ \sigma_{BS} &= \text{ATM Black-Scholes implied volatility} \\ \tau &= \text{time to maturity, 30 days for our data} \end{aligned}$$

- When m is negative a put option is out of the money

Moneyness Categories and S&P 500 Futures

Groups		
IV1	m=-4	Way out of the money
IV2	m=-3	Way out of the money
IV3	m=-2	Out of the money
IV4	m=-1	Out of the money
IV5	m=-.5	Nearly at the money
IV6	m=0	At the money
IV7	m=1	In the money
Futures		E-mini S&P futures

Data are 15 sec. pre-averaged to 5 min., de-volatilized, and de-jumped as per the foregoing.
January 3, 2007, to March 22, 2011, 1062 trading days.

EMM – Structural Model

- Structural model: The observed data are from a sampled, transformed, Levy process.

- On compacts, the transformed data are distributed as the stable distribution with parameter β, $1 < \beta \leq 2$
  - We use $[-4.5, 4.5]$

- The likelihood is not known in closed form; standard approximations such as inversion of the characteristic function are not numerically stable.

- The data can be simulated: $\hat{y}_t$, $t = 1, \ldots, N$
  - We use $N = 55000$

EMM – Auxiliary Model

- SNP density:

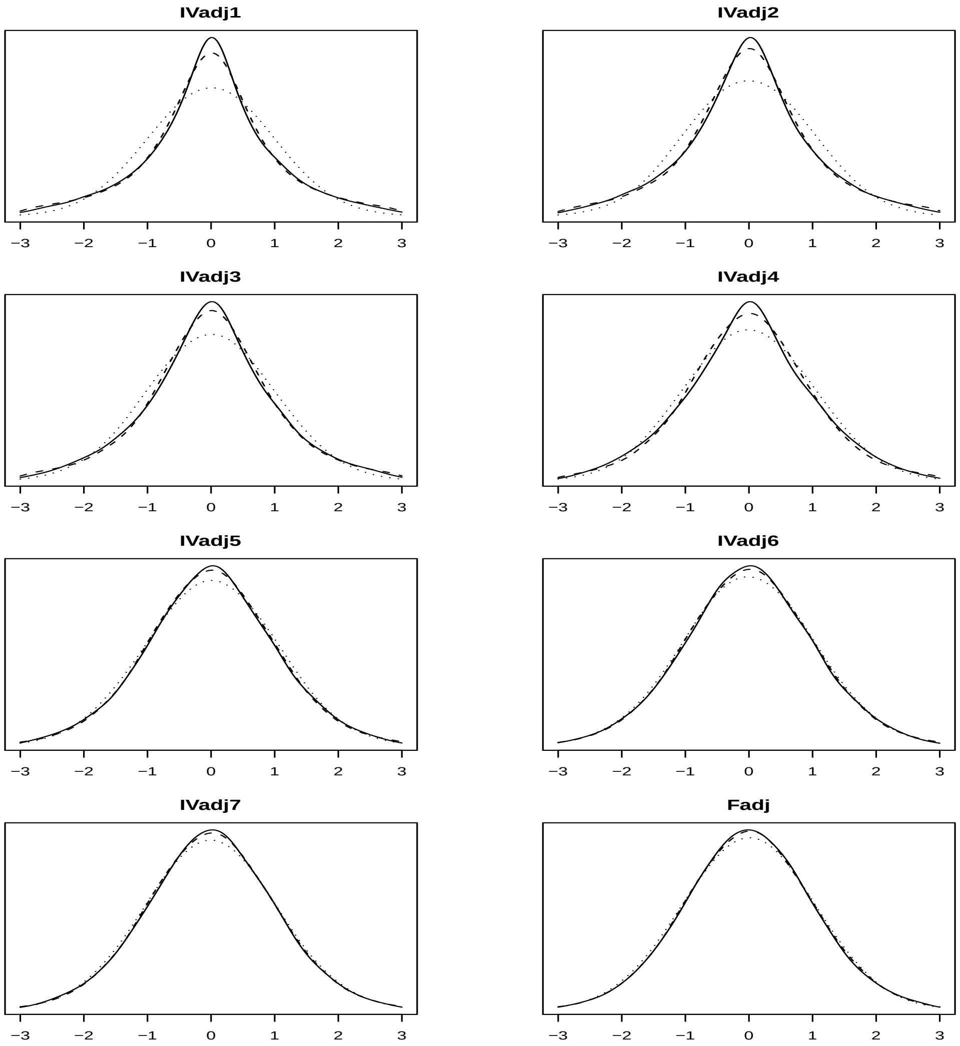
$$
\begin{aligned}
y &= Rv + \mu \\
h_K(v) &= \frac{[\mathcal{P}_K(v)]^2 \phi(v)}{\int [\mathcal{P}_K(u)]^2 \phi(u)\, du} \\
\mathcal{P}_K(v) &= \sum_{|\lambda|=0}^{K} a_\lambda v^\lambda
\end{aligned}
$$

  - We use $K = 8$ with even powers only
  - $\phi(v)$ is the standard normal

- MLE:

$$\tilde{\theta} = \underset{\theta \in \Theta}{\text{argmax}} \frac{1}{n} \sum_{t=1}^{n} \log[f_K(y_t \,|\, \theta)]$$

Auxiliary Model



Solid line is data, dashed line is SNP fit, dotted line is normal

EMM – Moment conditions

- Moment conditions

$$m(\beta, \tilde{\theta}) = \mathcal{E}_\beta \left\{ \frac{\partial}{\partial \theta} \log[f_K(y \,|\, \tilde{\theta})] \right\}$$

- Identification

$$m(\beta, \tilde{\theta}) = 0 \iff \beta = \text{ true value}$$

- Moment conditions can be computed by simulation

$$m(\beta, \tilde{\theta}) \doteq \frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \theta} \log[f_K(\hat{y}_i \,|\, \tilde{\theta})]$$

EMM – Variance Estimate

- Either, continuously updated

$$\mathcal{I}_\beta = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\partial}{\partial\theta}\log f(\widehat{y}_i \,|\, \tilde{\theta})\right]\left[\frac{\partial}{\partial\theta}\log f(\widehat{y}_i \,|\, \tilde{\theta})\right]',$$

  where, recall, $\{\widehat{y}_i\}_{i=1}^{N}$ are draws from the stable with index β

- or static

$$\mathcal{I} = \int \mathcal{I}_\beta \, \pi(\beta)\, d\beta,$$

  where $\mathcal{I}$ is computed by averaging $\mathcal{I}_\beta$ over draws from the prior $\pi(\beta)$

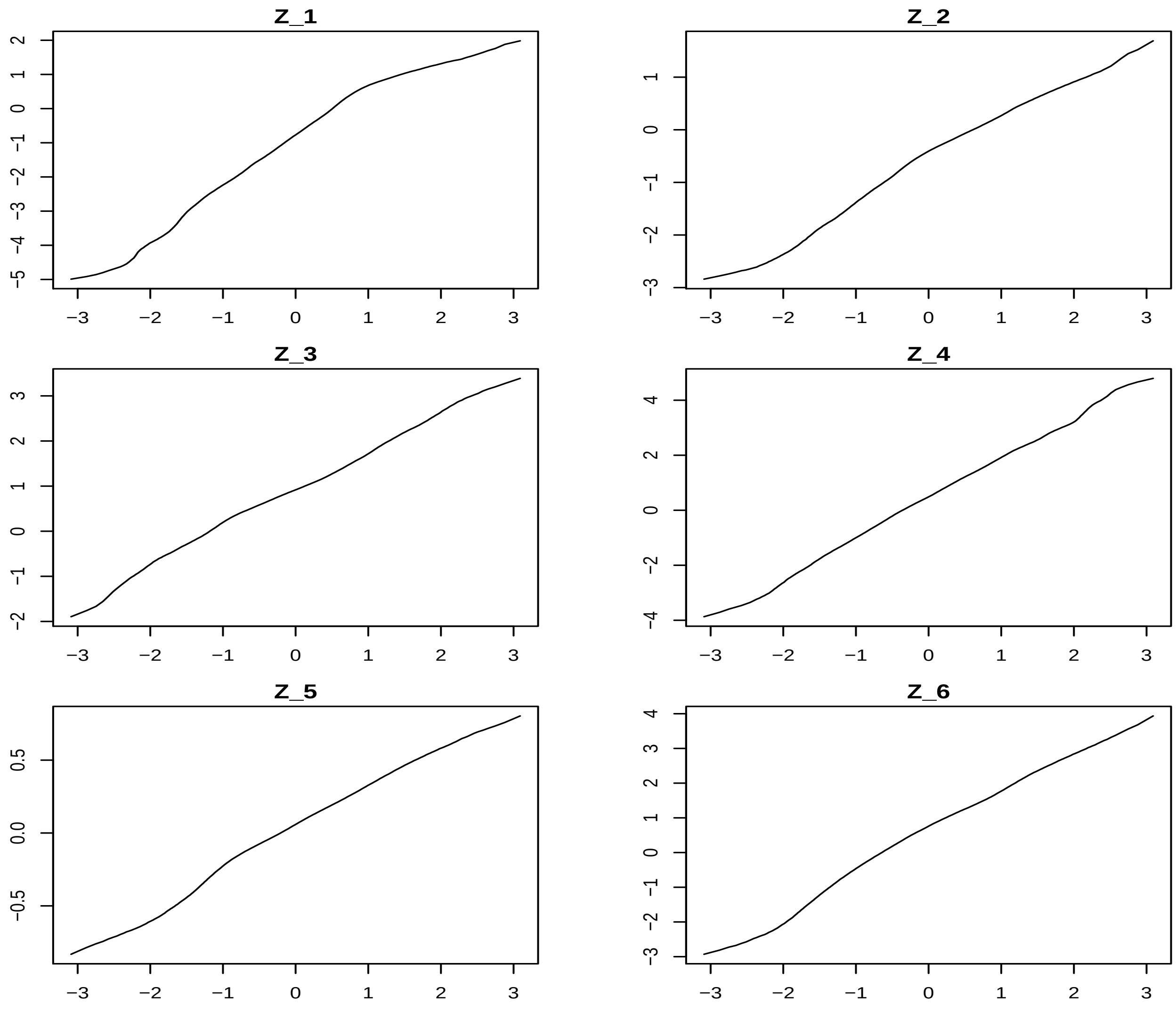
Distribution of Moment Conditions

In Bayesian inference the unknown parameter is manipulated as if it were random even though one might actually regard it as fixed.

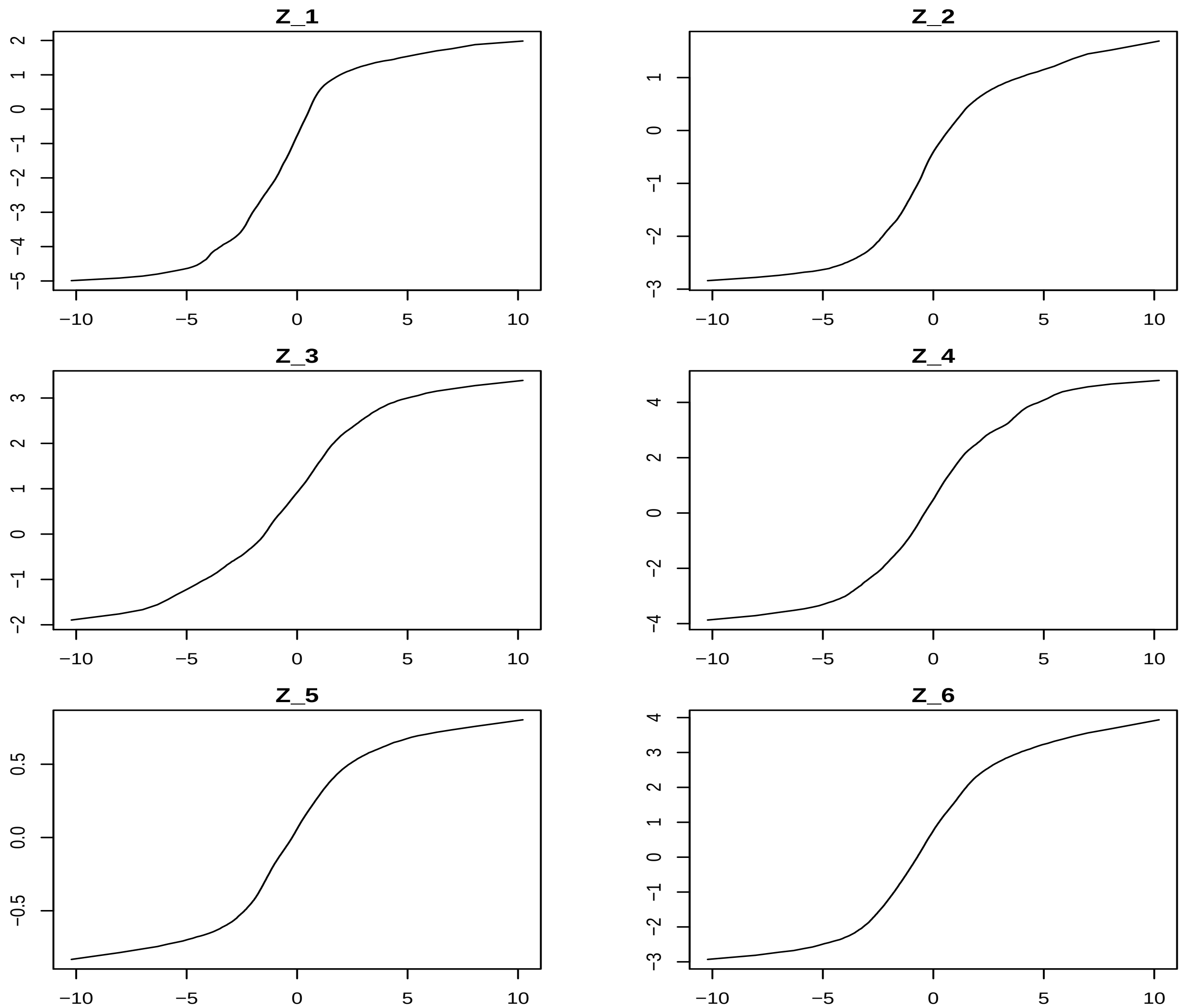
Consider the following exercise:

- Draw β from the uniform over $1 < \beta \leq 2$
- Simulate a sample of size n from Levy process at that β
- Compute the SNP estimate $\tilde{\theta}$
- Simulate a sample of size N from Levy process at that β
- Compute $z = \sqrt{n}\,(\mathcal{I}_{\beta})^{-1/2} m(\beta, \tilde{\theta})$ at that β
- Plot z quantiles vs. Normal quantiles and vs. t quantiles

Q-Q Plot of z Draws vs. Normal



Q-Q Plot of z Draws vs. 3 d.f. t



Conclusion Regarding Distribution of Z

- Under a uniform (1,2] prior on β and a Levy likelihood for the data, it is plausible to assume that the distribution of

$$Z_\beta = Z(\tilde{\theta}, \beta) = \sqrt{n}\,(\mathcal{I}_\beta)^{-1/2} m(\beta, \tilde{\theta})$$

  is normal.

- A standard CLT for iid data applies here so one would expect normality.
  - The issue of uniform convergence does not arise.
  - Differentiates our procedure from other ABC methods.

- Normal distribution denoted Φ normal density denoted ϕ

Exact Bayesian Inference

The central idea is that moment equations can be used to construct a "method of moments representation" of the likelihood of a structural model that can be used in otherwise standard Bayesian inference.

Exact Bayesian Inference - Notation

Temporary change of notation to agree with

Gallant, A. Ronald (2016), "Reflections on the Probability Space Induced by Moment Conditions with Implications for Bayesian Inference," *Journal of Financial Econometrics* 14, 284–294.

Gallant, A. Ronald (2016), "Reply to Comment on Reflections," *Journal of Financial Econometrics* 14, 284–294.

$$
\begin{aligned}
\tilde{\theta} &\rightarrow y \in \mathcal{Y}, \text{ which is viewed as the data} \\
\beta &\rightarrow \theta \in \Theta \\
m(\beta, \tilde{\theta}) &\rightarrow m(y, \theta) \\
Z(\beta, \tilde{\theta}) &\rightarrow Z(y, \theta)
\end{aligned}
$$

Exact Bayesian Inference - Structural P-Space

- The structural model and prior imply a joint distribution $P^o(y, \theta)$ defined over $\mathcal{Y} \times \Theta$, whose density is the product of a likelihood $p(y \,|\, \theta)$ times a prior $\pi(\theta)$.

- The joint probability space is, therefore,

$$(\mathcal{Y} \times \Theta, \mathcal{C}^o, P^o),$$

where $\mathcal{C}^o$ denotes the Borel subsets of $\mathcal{Y} \times \Theta$.

Exact Bayesian Inference - Moment Induced P-Space

- The random variable $z = Z(y, \theta)$ over $(\mathcal{Y} \times \Theta, \mathcal{C}^o, P^o)$ has some distribution Ψ with a support $\mathcal{Z}$. $\Psi = \Phi$ in our application.

- Let $\mathcal{C}$ be the smallest σ-algebra containing the preimages $C = Z^{-1}(B)$ where B ranges over the Borel subsets of $\mathcal{Z}$.

- Because the distribution $\Psi(z)$ of $z = Z(y, \theta)$ is determined by P^o the probability measure $P[C = Z^{-1}(B)] = \int_B d\Psi(z)$ over $(\mathcal{Y} \times \Theta, \mathcal{C})$ will satisfy $P(C) = P^o(C)$ for every $C \in \mathcal{C}$.

Exact Bayesian Inference - Extension of Moment Induced P-Space

- Define $\mathcal{C}^*$ to be the smallest σ-algebra that contains all sets in $\mathcal{C}$ plus all sets of the form $R_B = (\mathcal{Y} \times B)$, where B is a Borel subset of Θ.

- If $Z(y, \theta)$ is a semi-pivotal, there is an extension to a space

$$(\mathcal{Y} \times \Theta, \mathcal{C}^*, P^*)$$

  such that $P^o(C) = P^*(C)$ for all $C \in \mathcal{C}^*$. (Gallant, 2016)

  - Sufficient to be semi-pivotal is that Z is continuous and unbounded in at least one element of y

- The σ-algebras involved satisfy $\mathcal{C} \subset \mathcal{C}^* \subset \mathcal{C}^o$.

Exact Bayesian Inference - Method of Moments Likelihood

- The "method of moments" likelihood on the extended space $(\mathcal{Y} \times \Theta, \mathcal{C}^*, P^*)$ is

$$\text{adj}(y, \theta)\psi[Z(y, \theta)]$$

  where $\text{adj}(y, \theta)$ is analogous to a Jacobian term.

  - Actually is a Jacobian in our application

- The key insight that allows substitution of the "method of moments representation" of the likelihood for the likelihood under the structural model in a Bayesian analysis is the fact that both probability measures P^o and P^* assign the same probability to sets in $\mathcal{C}^*$.

Exact Bayesian Inference - Information Loss

- Because $\mathcal{C}^*$ is a subset of $\mathcal{C}^o$, some information is lost

- Intuitively this is similar to the information loss that occurs when one divides the range of a continuous variable into intervals and uses a discrete distribution to assign probability to each interval. Both the continuous and discrete distributions assign the same probability to each interval but the discrete distribution cannot assign probability to subintervals.

- How much information is lost depends on how well one chooses moment conditions.

EMM – Objective Function

- Semi-pivotal:

$$Z_\beta = Z(\tilde{\theta}, \beta) = \sqrt{n}\,(\mathcal{I}_\beta)^{-1/2} m(\beta, \tilde{\theta})$$

- EMM objective function

$$\text{adj}(\tilde{\theta}, \beta)\, \phi[Z(\tilde{\theta}, \beta)]\, \pi(\beta)$$

  – $\text{adj}(\tilde{\theta}, \beta) = \left|\det\left[\frac{\partial}{\partial\theta'} Z(\tilde{\theta}, \beta)\right]\right|$

  – $\text{adj}(\tilde{\theta}, \beta)\phi[Z(\tilde{\theta}, \beta)]$ is the likelihood

  – $\pi(\beta)$ is the prior

  – We use $p(\beta)$ uniform on $(1, 2]$

EMM – MCMC

Given a current β^o obtain the next β' by:

1. Draw β^* according to $q(\beta^o, \beta^*) = n[\beta^* \,|\, \beta^o, (\sigma^*)^2]$.

2. Simulate $\{\widehat{y}_t\}_{t=1}^N$ from the structural model with parameter set to β^*.

3. Compute $m(\beta^*, \tilde{\theta})$, $Z_{\beta^*} = \sqrt{n}\,(\mathcal{I}_\beta)^{-1/2} m(\beta^*, \tilde{\theta})$, and $a_{\beta^*} = \mathrm{adj}(\tilde{\theta}, \beta^*)$

4. Compute $\alpha = \min\left(1, \frac{a_{\beta^*}\,\phi(Z_{\beta^*})\,\pi(\beta^*)\,q(\beta^*, \beta^o)}{a_{\beta^o}\,\phi(Z_{\beta^o})\,\pi(\beta^o)\,q(\beta^o, \beta^*)}\right)$.

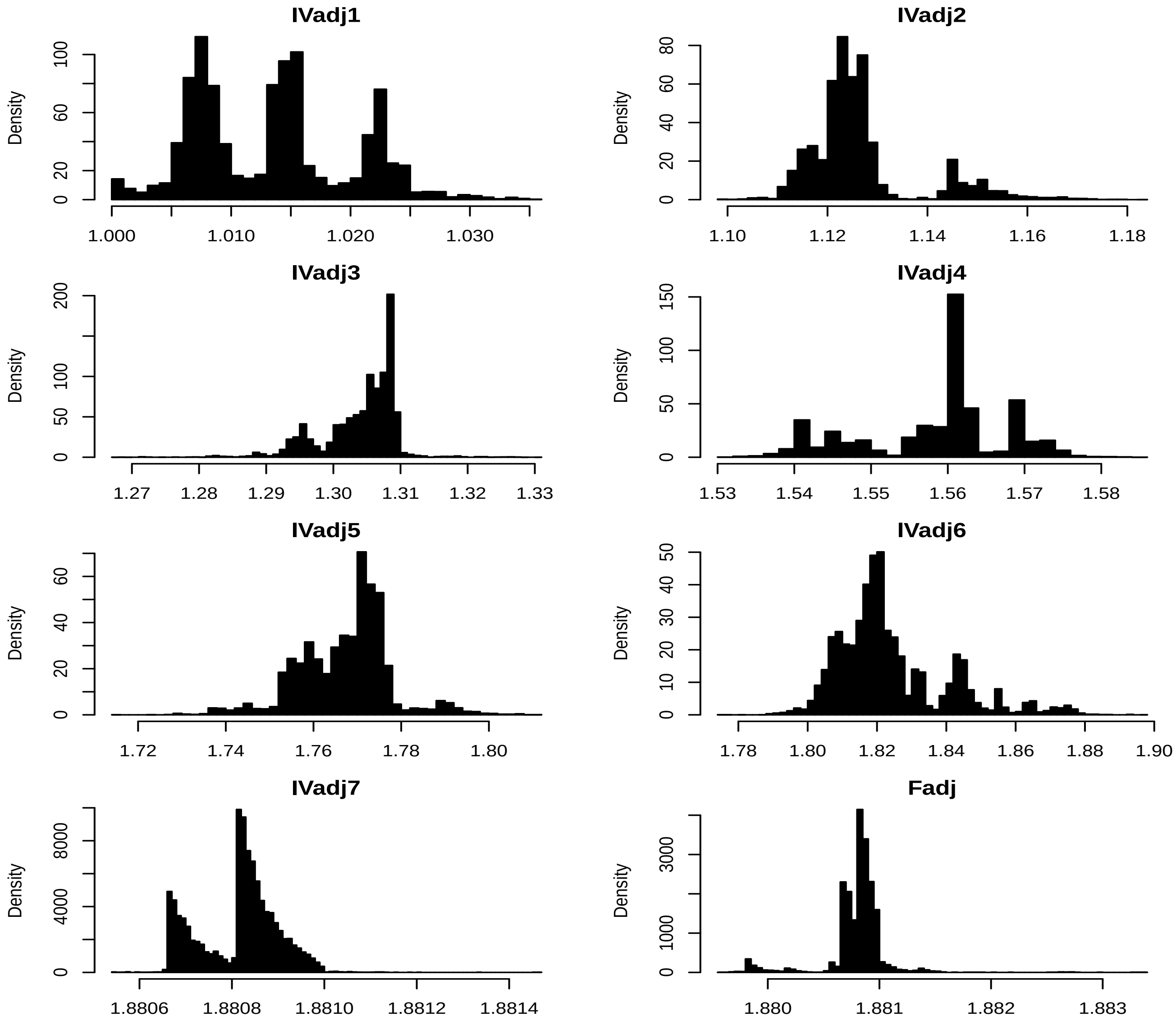
5. With probability α, set $\beta' = \beta^*$, otherwise set $\beta' = \beta^o$. Return to step 1 with the new β^o set to β'.

Result is an MCMC chain β_t $t = 1, \ldots, R$; We used $R = 50000$.

Estimate is either $\beta_{mode} = \underset{\beta_t}{\mathrm{argmax}}\{\phi(Z_{\beta_t})\,\pi(\beta_t)\}$ or $\beta_{mean} = \frac{1}{R}\sum_{t=1}^R \beta_t$

Credibility interval is the 2.5% and 97.5% quantiles of $\{\beta_t \,|\, t = 1, \ldots, R\}$

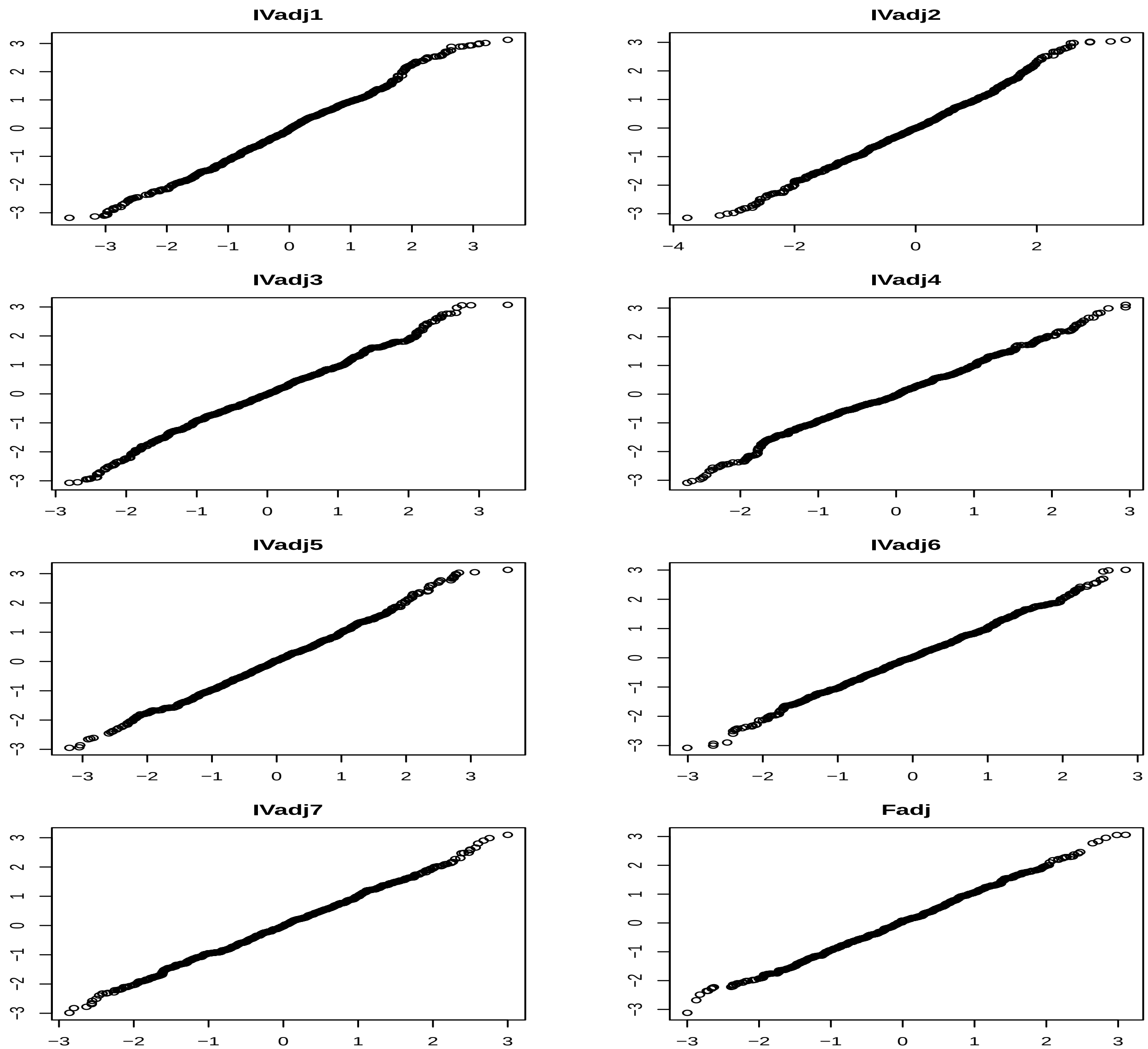
Histograms of Beta Draws



EMM – Estimates of β

Moneyness	Posterior Mean	Posterior Mode	95% Credibility Interval	
IV1	1.0134	1.0073	1.0026	1.0257
IV2	1.1266	1.1231	1.1125	1.1556
IV3	1.3041	1.3090	1.2915	1.3097
IV4	1.5586	1.5617	1.5398	1.5737
IV5	1.7672	1.7706	1.7438	1.7902
IV6	1.8240	1.8210	1.8023	1.8664
IV7	1.8808	1.8808	1.8807	1.8810
Futures	1.8808	1.8808	1.8799	1.8812

Cross Check – Q-Q Plot of Draws from Model at Est. β vs. Data



Cross Check – Downward Bias

- Our estimates of the index β are significantly lower than those found elsewhere, raising the possibility that the procedure itself might be downward biased.

- To investigate, we did a Monte Carlo (10,000 reps) by applying the entire procedure to simulated data sets of length equal to our sample size under the model distribution implied by Theorem 1 for two separate parameter settings.

Estimator	Mean	Std.Dev.
$\beta = 1.10$		
posterior mean	1.1003	0.0162
posterior mode	1.1006	0.0170
$\beta = 1.90$		
posterior mean	1.8698	0.0194
posterior mode	1.8679	0.0221

- Evidently, the Monte Carlo values are nearly spot-on adding credence to our claimed empirical findings.

Conclusion: Exact Bayes Implies

	Posterior		95% Credibility	
Moneyness	Mean	Mode	Interval	
IV1	1.0134	1.0073	1.0026	1.0257
IV2	1.1266	1.1231	1.1125	1.1556
IV3	1.3041	1.3090	1.2915	1.3097
IV4	1.5586	1.5617	1.5398	1.5737
IV5	1.7672	1.7706	1.7438	1.7902
IV6	1.8240	1.8210	1.8023	1.8664
IV7	1.8808	1.8808	1.8807	1.8810
Futures	1.8808	1.8808	1.8799	1.8812