Incorporating Higher Moments into Financial Data Analysis

By

James M. Sfiridis
Associate Professor
Department of Finance
University of Connecticut

June 9, 2006

The Problem

Ample empirical evidence that short-term security returns are not normally distributed.

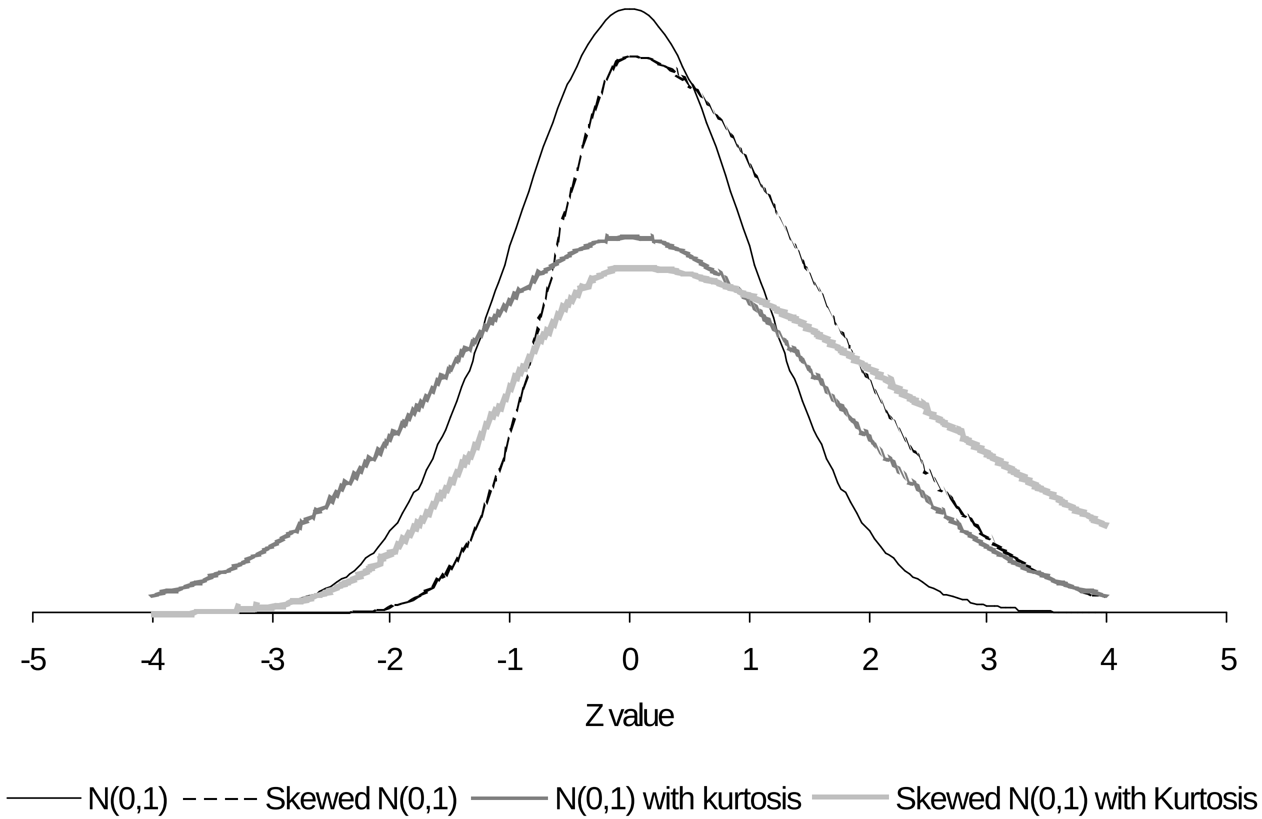
- *Fat tails* or *leptokurtosis* in the fourth moment - return *outliers* occur with greater frequency than expected assuming normality
  - Mandelbrot (1963)
  - Fama (1976)
  - Campbell, Lo and MacKinlay (1997)

- *Skewness* occurs in the third moment – return distribution is asymmetric because of the tendency for either positive or negative returns to persist or *cluster*.
  - Campbell, Lo and MacKinlay (1997)
  - Conine and Tamarkin (1981)
  - Harvey and Siddique (2000)
  - Hwang and Satchell (1999)

Ignoring these higher statistical moments can lead to

- **biased and/or inefficient parameter estimates**
- **the loss of data-based information**

Figure 1. Illustrating the effect of skewness and kurtosis. Simulations shown below are all from an initial Normal (0,1) distribution with various levels of skewness and kurtosis.



Past Solutions

A. Distribution-free non-parametric tests
   - Corrado (1989)
   - Cowan (1992)

B. The bootstrap
   - Efron and Gong (1983)
   - Patro and Wu (2004)

C. The jackknife
   - Quenouille (1956)
   - Tukey (1958)

D. A better model?
   - Additional cross-sectional factors
     - Fama and French (1996)
   - Time-series models
     - Stochastic coefficients for the first moment
       - Jostova and Philipov (2004), Ghysels (1998)
     - Time-varying variance for the second moment
       - ARCH, GARCH, EGARCH, etc.

E. Ignore the higher moments by assuming normality
   - OLS regressions that assume $\varepsilon \sim N(0, \sigma^2)$

F. VaR research has just started to incorporate higher moments
   - Allen and Bali (2005)

Goal of this Paper

Incorporate four statistical moments in a likelihood-based framework to

(1) More accurately model the return data-generating process

(2) Clarify the affect of the higher moments on statistical testing and point estimation.

Bayesian Methodology

Bayes Theorem states

$$p(q|data)\,f\,(data) = h(data|q)p\,(q)$$

where p(.) is the joint posterior distribution of the parameter vector, θ; f(data) is a constant of proportionality; h(.) is the likelihood function and specifies the data distribution and π(.) is the prior distribution of the parameter vector. We can also say that

$$p(q|data) \propto h(data|q)p\,(q)$$

The joint posterior distribution is of interest here. Sampling-based Bayesian methods, such as Markov chain Monte Carlo (MCMC), allow determination of the marginal posterior distribution of individual components of the parameter vector, or

$$p(q_i|data) \propto h(data|q)p\,(q_i)$$

by drawing large samples from the marginal posterior distribution via the right-hand side of the above equation. Here θ_i represents just one component of the θ parameter vector.

If the *likelihood x prior* specification provides a known distribution to sample from, the GIBBS SAMPLER can be used. If the *likelihood x prior* specification gives an unknown distribution to sample from, the METROPOLIS-HASTINGS algorithm can be used to get samples from the desired marginal posterior distribution.

For our purposes a portion of the parameter vector θ includes m, σ^2, ν and γ.

Bayesian Methodology

A likelihood function is used based on Fernandez and Steel (1998) as

$$h(r_i|\mu,\sigma^2,\lambda,\gamma)=\left(\frac{2}{\pi}\right)^{1/2}\left(\frac{1}{\gamma+\frac{1}{\gamma}}\right)\frac{\lambda_i^{1/2}}{\sigma}\exp\left[-\frac{\lambda_i(r_i-\mu)^2}{2\sigma^2}\left\{\frac{1}{\gamma^2}I_{[0,\infty)}(r_i-\mu)+\gamma^2 I_{(-\infty,0)}(r_i-\mu)\right\}\right]$$

where in addition to the first two moments, μ and σ^2, the parameter vector is supplemented by λ and γ to introduce kurtosis and skewness respectively.

- λ and γ are unobserved and must be elicited from the Bayesian MCMC analysis.
- Data augmentation is used to replicate kurtosis by a scale mixture of normals, or

$$h(r_i|\mu,\sigma^2,\lambda,v,\gamma)=\left(\frac{2}{\pi}\right)^{1/2}\left(\frac{1}{\gamma+\frac{1}{\gamma}}\right)\int_0^{\infty}\frac{\lambda_i^{1/2}}{\sigma}\exp\left[-\frac{\lambda_i(r_i-\mu)^2}{2\sigma^2}\left\{\frac{1}{\gamma^2}I_{[0,\infty)}(r_i-\mu)+\gamma^2 I_{(-\infty,0)}(r_i-\mu)\right\}\right]$$

$$\text{x } f_G\left(\lambda_i\left|\frac{\nu}{2},\frac{\nu}{2}\right.\right)d\lambda_i$$

so that the λ mixing parameter for each return observation is defined by degrees of freedom, ν. Thus, the parameter vector, θ, is defined as

$$[\mu,\sigma^2,\gamma,\lambda_1\ldots\lambda_N|\nu]$$

having N+4 dimensions.

How well does the MCMC methodology replicate a known non-normal distribution?

PANEL A. With kurtosis. Parameters from data generated from target distribution are mean = -0.03, standard deviation = 1.31, skewness = -0.20, kurtosis = 2.93

Parameters	*mode*	*mean*	*median*	*scale*	*degrees of freedom*	*gamma*
Target distribution	**0**	0.0	0	1.0	4.0	1.0
MCMC output	**0.03**	0.0	0	1.0	5.18	0.99
Data-based output	**-0.03**	-0.03	-0.03	1.3	30.0	1.0

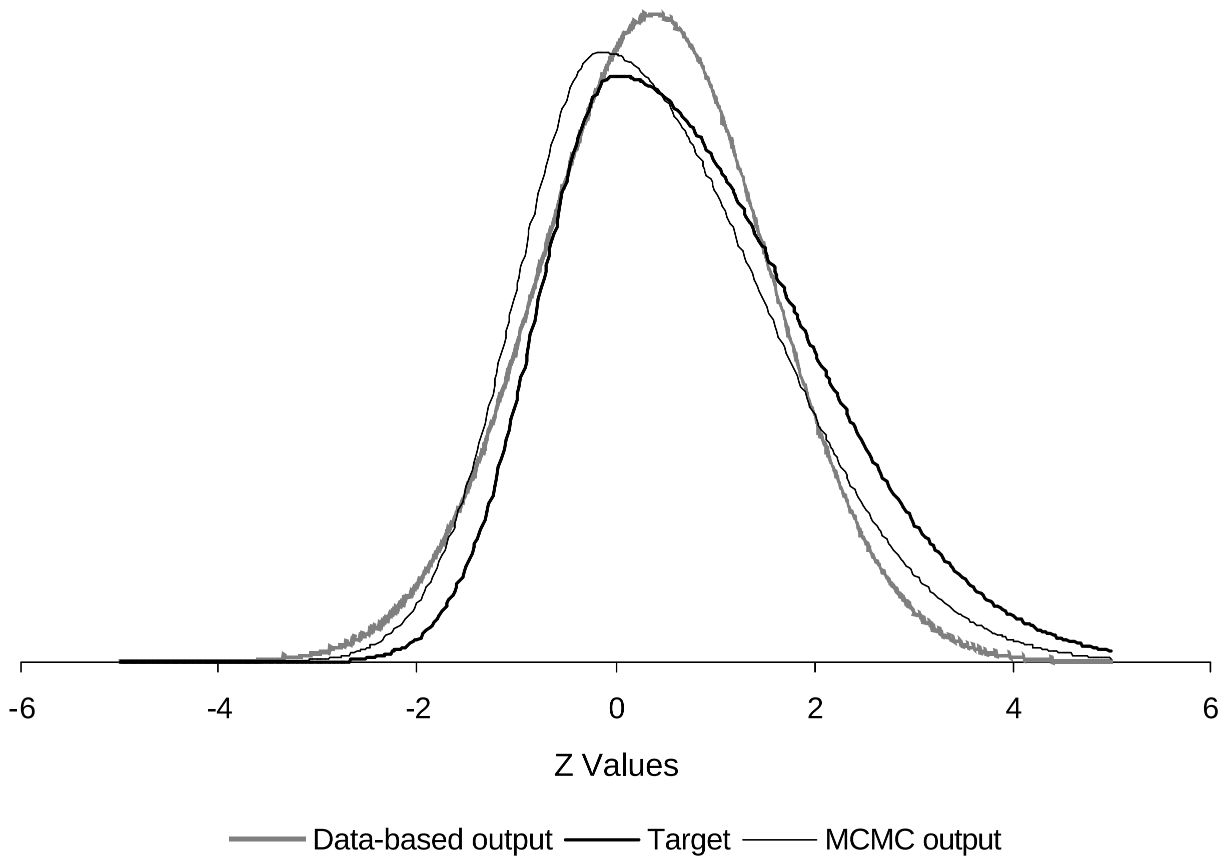
PANEL B. With skewness. Parameters from data generated from target distribution are mean = -0.32, standard deviation = 1.04, skewness = -0.54, kurtosis = 0.06

Parameters	*mode*	*mean*	*median*	*scale*	*degrees of freedom*	*gamma*
Target distribution	**0**	-0.49	-0.40	1.0	30.0	0.75
MCMC output	**0.122**	-0.29	-0.23	0.94	23.65	0.77
Data-based output	**-0.32**	-0.32	-0.32	1.04	30.0	1.0

PANEL C. With skewness and kurtosis. Parameters from data generated from target distribution are mean= 0.39, standard deviation= 1.34, skewness= 0.95, kurtosis= 1.85

Parameters	*mode*	*mean*	*median*	*scale*	*degrees of freedom*	*gamma*
Target distribution	**0**	0.77	0.62	1.0	7.0	1.50
MCMC output	**-0.15**	0.44	0.30	1.07	9.3	1.37
Data-based output	**0.39**	0.39	0.39	1.34	30.0	1.0

Figure 4. Simulation from an initial N (0,1) distribution having positive skewness and kurtosis.

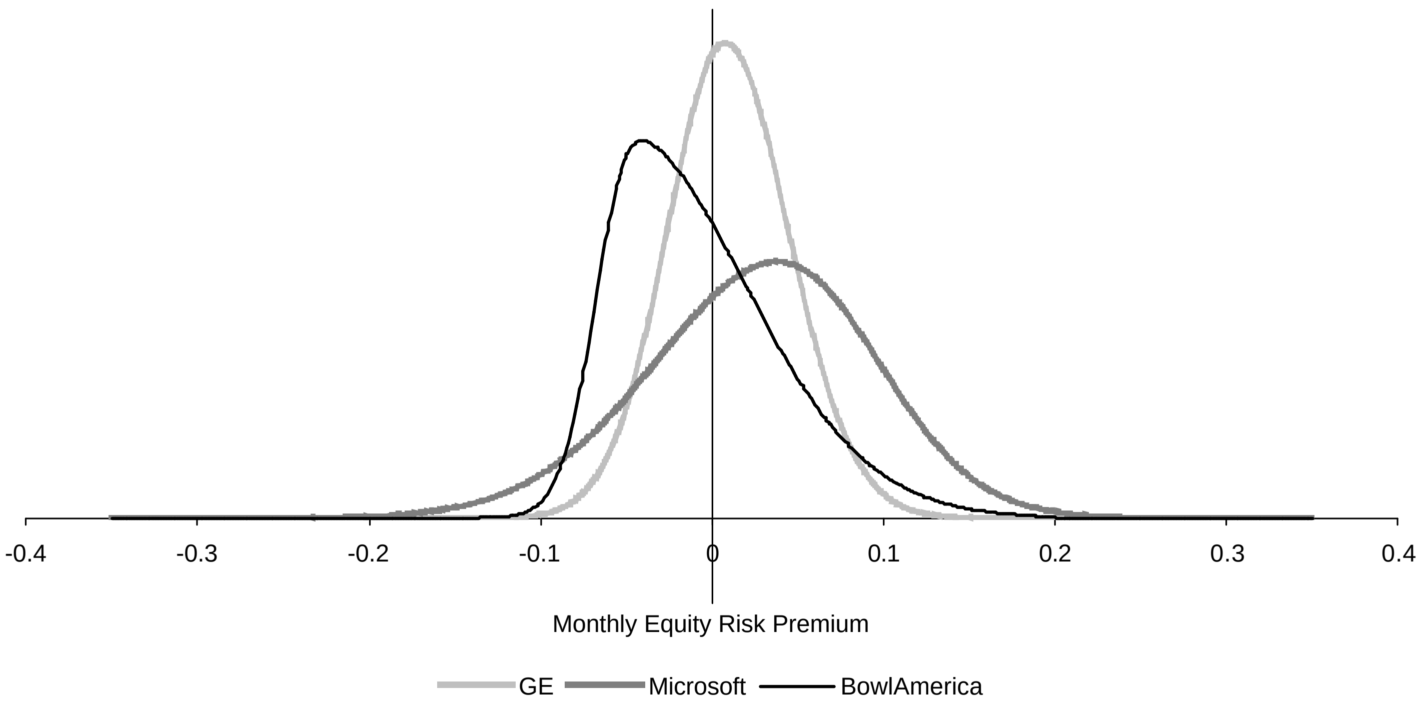


Do higher moments make a difference in tests of financial asset pricing models, e.g., the S-L CAPM?

Higher moments and testing the S-L CAPM

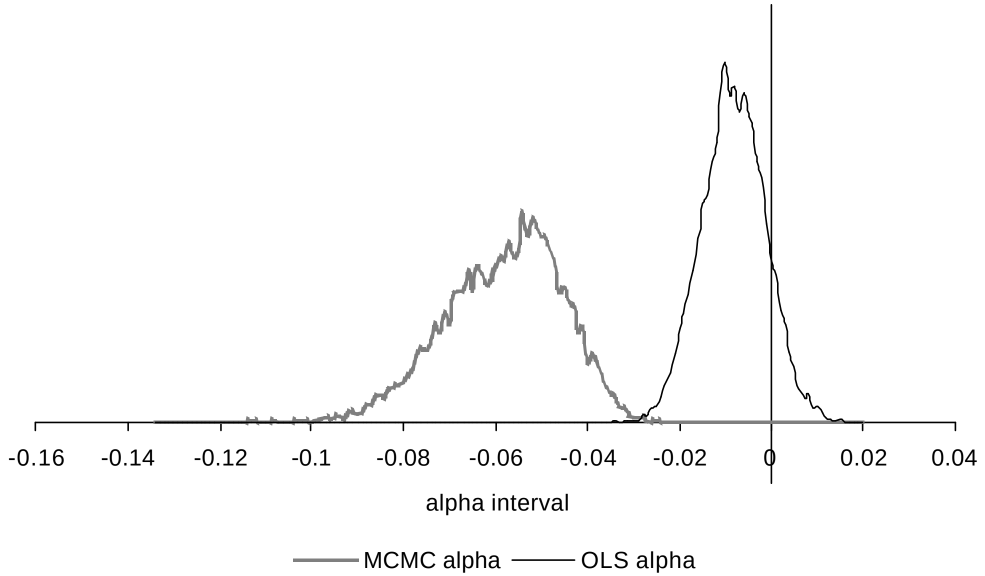
Panel A. Descriptive Statistics and Ordinary Least Squares (OLS) Analysis Results					
	G.E.	*Microsoft*	*Genentech*	*BowlAmerica*	*Osteotech*
Market Cap	$377.6B	$282.9B	$49.8B	$51.7M	$67.5M
Exchange	NYSE	Nasdaq	NYSE	AMEX	Nasdaq
Time Period	01/87-12/91	01/92-12/96	01/94-12/98	01/94-12/98	01/99-12/03
OLS Market Model Analysis					
Adj. R Square	0.75	0.11	0.09	0.07	0.18
Standard Error (σ_{OLS})	0.037	0.067	0.04	0.052	0.204
No. of Observations	60	60	60	60	60
F-test statistic	173	7.13	5.53	5.37	12.33
p-value	<0.001	0.01	0.022	0.024	<0.001
OLS alpha (α_{OLS})	**0.001**	**0.016**	**0.001**	**-0.009**	**0.002**
p-value	**0.413**	**0.038**	**0.417**	**0.106**	**0.472**
OLS beta (β_{OLS})	1.20	0.95	0.303	0.385	1.791
p-value	<0.001	0.009	0.022	0.024	<0.001
Skewness	**0.26**	**-0.50**	**0.66**	**1.58**	**0.60**
Kurtosis	**-0.42**	**0.13**	**1.26**	**4.60**	**0.35**
Panel B. Parameter Median Values from the Markov chain Monte Carlo (MCMC) Analysis					
alpha (α)	**-0.001**	**0.034**	**-0.010**	**-0.057**	**-0.100**
Sampling-based p-value	**0.413**	**0.016**	**0.19**	**<0.001**	**0.054**
beta (β)	1.20	0.92	0.266	0.31	1.746
Standard deviation (σ)	0.034	0.062	0.031	0.030	0.186
Degrees of freedom (ν)	**22.46**	**18.80**	**5.35**	**7.52**	**18.29**
Gamma (γ)	**1.06**	**0.87**	**1.21**	**2.26**	**1.33**
Panel C. Monthly Equity Risk Premium					
OLS mean	0.0095	0.0242	0.0050	-0.0038	0.0003
MCMC mean	0.0100	0.0269	0.0032	-0.0071	-0.0213
MCMC median	0.0093	0.0293	0.0010	-0.0150	-0.0350
MCMC mode	0.0080	0.0380	-0.0040	-0.0410	-0.075

Figure 5. Monthly equity risk premium predictive distributions for three of the five companies in the sample. Only GE's distribution approaches normality. Microsoft has negative skewness and BowlAmerica has high levels of both positive skewness and leptokurtosis.



BowlAmerica alpha simulation comparison

	OLS	MCMC
Mean	= -0.009	-0.059
Median	= -0.009 (0.106)	-0.057 (<0.001)
Mode	= -0.009	-0.054
Skewness	= 0.001	-0.61
Kurtosis	= -0.002	0.946



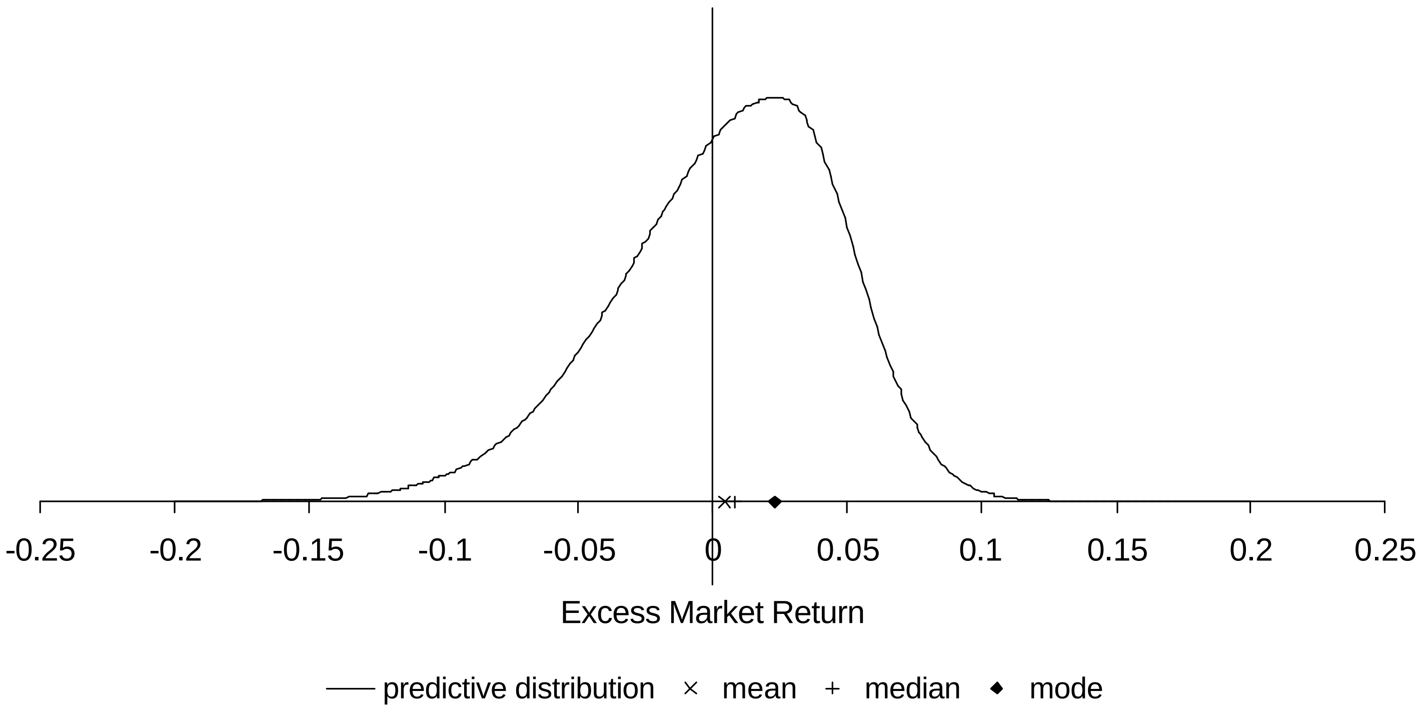
CONCLUSION: A statistical test that wrongly assumes normality biases the result towards acceptance of H_0:$\alpha_i = 0$ when it may be false, i.e., type II error.

NOTE: What happens if α_i had been shifted to the right? Now the wrong assumption of normality biases the result towards rejection of H_0 when it may be true, i.e., type I error.

Table 4. Descriptive statistics for the monthly market risk premium for various time periods.

Number of Years	*Time Period*	*Mean*	*Standard Deviation*	*Skewness*	*Kurtosis*
5	199401-199812	0.0128	0.040	-1.4707	4.605
5	199901-200312	-0.00081	0.052	-0.266	-0.833
10	199401-200312	0.0060	0.0469	-0.743	0.561
20	198401-200312	0.0064	0.0457	-0.93	3.019
30	197401-200312	0.0046	0.0473	-0.500	1.980
40	**196401-200312**	**0.0048**	**0.0448**	**-0.500**	**1.987**
50	195401-200312	0.0056	0.0433	-0.500	1.925

Figure 6. Market Risk premium predictive distribution using monthly returns, 1964-2003. Moments under an assumption of normality are given in table 4. Annualized arithmetic mean return is 5.64%. Annualized mean, median and mode from the MCMC non-normal specification is 5.76%, 10.2% and 27.6% respectively.



Do higher moments affect point estimation of important financial parameters, e.g., the market risk premium?

- Pastor and Stambaugh (2001) find the market risk premium is 3.9%-6% using data from 1834-1997.
- Using data from 1926-1997 Mayfield (2004) estimates an unconditional forward market risk premium of 8.5% with a 95% confidence interval of 4.4%-12.2%. This compares to a simple arithmetic average of 8.3%.
- Kim, Morley and Nelson (2005) estimate the market risk premium of 9% with a posterior probability of 6%-12% using monthly data since World War II. This compares to an arithmetic mean of 6.5%.
- Using data from 1964-2003 I estimate the market risk premium to be 10.2% with a 50% posterior probability of 6.72% to 16.8%. This compares to a simple mean of 5.64%.

CONCLUSION - A Bayesian model that explicitly incorporates skewness and kurtosis provides a satisfactory point estimator of the variable of interest, but precision does suffer compared to a model that may accurately account for higher moments through reformulation of the first two moments.