Non-Normality Issue and Hypothesis-Inferring: Testing the Monte Carlo Process

Syed Kashif Saeed*
Farooq Rasheed**

Abstract

Ordinary Least Square (OLS) estimator is assumed to be an unbiased estimator and the errors are normally distributed. However often is the case that stock returns characteristically have non-symmetric distribution which leads to problems related to inferential part by using the estimates of regression analysis. Markov-chain Monte Carlo simulation approach offers advantage in better estimates of the model and has become an important tool in risk management. In this article we compare the critical t-statistics estimated by Monte Carlo Simulation process with the standard asymptotic t-distribution which subsist under the assumption that the error terms are normally distributed. Sample of 6 stock companies from the Karachi Stock Exchange (KSE) 100 index was taken. Daily data of 406 closing prices and KSE 100 index from January 2010 to June 2011 is taken from Daily "Business Recorder". Jarque Bera Test shows that regression error terms in all these six estimated models were not normally distributed. Following Monte Carlo Simulation procedure, the critical t-values were simulated at 5% level of significance. These values were found to be almost closer to the asymptotic standards of t-distribution. Thus it can be concluded that Monte Carlo based simulation approach is a preferred one for assessing statistical significance due to its property to transform unsymmetrical distribution into symmetrical distribution.

Key Words: Monte Carlo simulation process, Asymptotic t-values, Normality assumption.

* Syed Kashif Saeed, Assistant Professor, PMAS-Arid Agriculture University, Rawalpindi, Pakistan.
** Farooq Rasheed, Assistant Professor, Air University, Islamabad, Pakistan.

1. Introduction

Monte Carlo methods were first introduced to finance by David B. Hertz in 1964. Hertz is known for his scholarly work in operations research. Monte Carlo Simulations (MCS) is a technique that studies regular routes through conducting trials. Researches using MCS techniques are appearing more frequently in the literature. Monte Carlo experiments with econometric applications are now being used extensively e.g., the analysis of portfolio holdings and financial balance sheets incorporates MCS.

Ciuiu (2009) used Jarque-Bera normality test to verify whether the residuals of the regression model are normally distributed or not. The Author presented Monte Carlo method to obtain normal residuals for the cases where the Jarque-Bera test rejects the hypothesis that residuals are normally distributed. After performing data simulation via Monte Carlo process, the author re-obtained regression estimates of the model and found that the new statistic of Jarque-Bera test do not reject normality in the error terms.

A study by Brooks (2002) scrutinized whether or not stock returns characteristically have non-symmetric distribution. If so then this would lead to the problem related to inferential part of the study by using the estimates of regression analysis. Frequent usage of daily data with trendy least square regression thus results in insignificant studies. Simpson (2001) analyzed that in various studies, the statistical significance of returns is measured frequently using t-tests and F-tests i.e., based on the estimated standard errors of the regression but the normality assumption is ignored.

In a linear regression equation the parameter of interest is the slope. OLS regression assumptions include zero mean of the errors, estimators are unbiased and the errors are normally distributed. Thus it is conclusive that the estimators estimated under OLS regression have a normal distribution. However an estimated slope parameter could be smaller or larger than its true parameter. This issue can be tackled if the required estimates are computed from large random data. It will estimate parameters equal to the true parameters. To perform this regression by means of data diffusion simulation process, samples of data can be simulated that are consistent with the regression assumptions.

Hammond and Sun (2001) findings show that in large economies a standard stochastic framework presumes a random macroeconomic shock combined with micro shocks that can be represented by a random process consisting of a range of unsystematic nature of variables. Authors were of the view that this process satisfies a joint measurability condition only if there is essentially no unusual risk patterns prevail at all by validating stochastic framework under MCS.

Christoffersen and Jacobs (2004) compare a range of financial models along a different dimension, using option prices and returns under the risk-neutral as well as the physical probability measure. By evaluating an objective function based on option prices they judge the relative performance of various models. They find that option-based objective function favors a relatively parsimonious model. Their analysis favors a model that, besides volatility clustering, only allows for a standard leverage effect, when evaluated out-of-sample through simulation process.

Gabriel (2003) compares the relative performance of several tests for the null hypothesis of co-integration, in terms of power and size in restricted samples, generally carried out by using MSC for a range of likely data-generating processes. Authors also examine the shock on size and power of selecting various procedures to estimate the variance of the error terms. Their study found that the parametrically tuned tests are the most well-balanced one characteristically as they display relatively smaller distortions in the long run situation.

Chou (1988) investigated the issues of volatility persistence and the changing risk premium in the stock market. Data cannot reject the non-stationary volatility process specification because the persistence of shocks to the stock return was so high. The parameter estimates and the non- stationary test are both robust to changes in the frequency of data measurements.

Francq and Zakoian (2000) detected a sequential correlation in the squared regression errors. This can be awkward because such autocorrelation structures are companionable with severe misspecifications. Standard (quasi-) maximum likelihood procedures can be inconsistent if the conditional first two moments are unspecified. To assuage these troubles of potential misspecification, they deem weak representations characterized by the squared error terms. The weak representation eliminates the need for correct requirement of the first two unconfirmed moments. However, using confidence intervals based on strong assumptions can be ambiguous and the need of models with simulation like Monte Carlo's arrives.

Hayakawa and Kurozumi (2006) signify the first difference of the variable to be integrated in the dynamic least square estimation of co-integrating models. Authors demonstrated that the role of leads is related to Granger causality test but in some situations such leads are needless in the dynamic OLS co-integrating regression models. Under MCS they found that the dynamic least square estimator, without leads significantly do better than with leads and lags and thus recommend the test of Granger non-causality.

Finance studies show that the parameters are usually found to be weak if using discrete time data rather than equations formulated in continuous time data. If the available discrete data is not transformable into continuous data by its original values then some technique is required to address this issue. Pedersen (1995) pioneered the idea of augmenting the actual low-frequency values with simulated extra frequencies, thus offers advantage in better estimates of the model. This is also known as discrete diffusion process through simulations and has become an important tool in finance particularly for financial derivatives and risk management purposes.

Eraker (2001) proposed an advanced parameter-estimation approach in diffusion models based on Markov-chain Monte Carlo (MCMC) procedure. Markov-chain Monte Carlo methodology specially applies to a system with unobservable state variables. This indicates that Eraker's MCMC approach initiate simulated auxiliary data points in between each pair of discrete observations for non-violation of the assumption of normality.

Dunkel & Weber (2005) discuss efficient Monte Carlo methods for the estimation of risk measures in Portfolio credit risk models. They are of the view that such an analysis of large financial losses in realistic portfolio models requires extensive numerical simulations. Thus they demonstrated that sampling with exponential twists can be used to construct numerically efficient estimators within the framework of the credit risk models. Their study found that the numerical simulations of test portfolios demonstrate good performance of the proposed estimators.

Zhu et al. (2011) evaluated the robust regression method when de-trending the crop yield records. They used a Monte Carlo simulation approach and the performance was compared with previous estimators found in the studies performed under OLS and method of moments for the models of crop yield. The study found that the outcome of acquiring more accurate de-trending method offers an improvement in the accuracy of models used in rating crop insurance contracts.

In this article the objective is to compare the critical t-statistics estimated by Monte Carlo Simulation process with the standard asymptotic t-distribution which subsists under the assumption that the error terms are normally distributed.

2. Data and Methodology

Authors have taken the sample of 6 companies listed in Karachi Stock Exchange 100 index that includes Pakistan State Oil (PSO), Engro Corporation Ltd. (ENGRO), Oil & Gas Development Corporation Limited (OGDCL), Muslim Commercial Bank (MCB), LUCKY CEMENT (LUC) and HUB Power Company Limited (HUBCO). Daily data of 406 closing prices and KSE 100 index from January 2010 to June 2011 is taken from daily "Business Recorder" to get stock returns (R_t) and market returns (MR_t). Following Ford (2003) the model for returns for each stock is as follows:

$$R_{it} = a_i + b_1MR_t + b_2MR_{t-1} + \varepsilon_{it} \quad (1)$$

Where R_{it} is the return of stock at time t for i^{th} firm calculated by the formula $(P_t - P_{t-1})/P_{t-1}$, MR_t is the market return on the KSE 100 index at time t, one time period lag value of index return is also incorporated in the model and ε is the error term. The equations' slope parameters are b_1 and b_2.

3. Results

Test of Normality

First of all the normality issue of the error terms obtained via equation (1) is checked by applying Jarque-Bera test for all the six companies selected.

Table 3.1
Jarque-Bera Test of Normality

	PSO	ENGRO	OGDCL	MCB	LUC	HUBCO
Observations	406	406	406	406	406	406
Jarque-Bera	19.05263	6.828940	43.56013	88.23147	29.00048	61.87618
Probability	0.000073	0.042233	0.000000	0.000000	0.000019	0.000000

It is found that error terms in all the cases were not normally distributed (see table 3.1). Following Monte Carlo Simulation (MacKinnon 2002), empirical distributions of the test statistics are estimated by using equation (2).

$$\check{R}_{it} = \alpha + \beta_1MR_t + \beta_2MR_{t-1} + \mu_{it} \quad (2)$$

Where $\check{R}_{it}$ is the forecasted return of stock at time t observed simply by adding Monte Carlo based residuals (ε^{MC}) into R_{it}. Note that Monte Carlo based residuals (ε^{MC}) were estimated by multiplying ε_{it} with $[n/(n - k)]^{0.5}$ where n is the number of observation and k is the number of parameters estimated.

T-Statistics under Monte Carlo Simulation Procedure

Using altered returns variable $\check{R}_t$, equation (2) was iterated for each stock 500 times such that a distribution of t-statistics[1] was thus generated. Finally 5 percent unique critical value based on average for each company were estimated using the formula $(\alpha/2)*(n)$. In our case (0.05/2)*(500) and (1 - 0.05/2)*(500) gives approximately 12th and 488th values of the sorted simulated t-statistics for β_1 and β_2 each for all 6 firms selected.

Table 3.2
Confidence Interval for Simulated t-Statistics

	Simulated Average "t-value" for β_1	Standard Deviation	Simulated Average "t-value" for β_2	Standard Deviation
PSO	1.98	0.008	1.97	0.007
ENGRO	1.97	0.009	1.98	0.008
OGDCL	1.98	0.007	1.89	0.009
MCB	1.97	0.009	1.92	0.007
LUC	1.99	0.008	1.95	0.008
HUBCO	1.99	0.009	1.81	0.009
Average	1.98	0.008	1.92	0.008
St Dev	0.0008	0.0082	0.0632	0.0009

Table 3.2 summarizes simulation based estimations. These values have returned the mean t-value of 1.98 and the mean t-value of 1.92 for each stock through simulation process for β_1 and β_2 respectively. Monte Carlo approach is a preferred one for assessing statistical significance due to its transforming property that critical statistics are generally symmetric. The standard deviations in all the cases were found to be very low suggesting the existence of consistency in these iterated estimations for β_1 and β_2. The standard asymptotic t-values are ±1.962 at 5 percent level of significance and

[1] Simulated t-statistics is calculated by $(\beta_j - b_j) / SE_{\beta j}$ }; for j = 1 to 2 as mentioned in equations (1) and (2).

our estimates of these simulated critical t-values are almost closer to the asymptotic standards of t-distribution.

4. Conclusion

Generally financial analysis is performed using regression tools and the statistical significance is examined by using the asymptotic critical values. However the stock returns are known to be abnormally distributed often in large data and thus asymptotic values cannot be considered as the appropriate one. Comparing Monte Carlo based simulated critical values with the standard asymptotic values depict that simulated values are more reliable measure of confidence intervals. The model was thus simulated to transform a series of non-normally distributed error terms in a normally distributed series.

References

Brooks, C. (2002). *Introductory Econometrics for Finance.* Cambridge, England*:* Cambridge University Press.

Chou, R. Y. (1988). Volatility Persistence and Stock Valuation. *Journal of Applied Econometrics, 3*(4).

Christoffersen, P., and Jacobs, K. (2004). Which GARCH Model for Option Valuation. *Management Sciences, 50*(9).

Ciuiu, D. (2009). Numerical and Monte Carlo Methods to Make Normal Residues in Regression. *Romanian Journal of Economic Forecasting, 1*(9).

Dunkel, J., & Weber, S. (2005). *Efficient Monte Carlo Methods for Convex Risk Measures in Portfolio Credit Risk Models.* Proceedings of the 2007 Winter Simulation Conference.

Eraker, B. (2001). MCMC Analysis of Diffusion Models with Application to Finance. *Journal of Business & Economic Statistics, 19*(2).

Ford, G. S. (2003), *Daily Stock Returns, Non-Normality and Hypothesis Testing*, Telepolicy.com Working Paper, July 2003.

Francq C., and Zakoian, J. M. (2000). Multivariate ARMA Models with Generalized Autoregressive Linear Innovation. *Stochastic Analysis and Applications, 18*(2).

Gabriel, V. J. (2003). Cointegration and the Joint Confirmation Hypothesis. *Economics Letters, 78*(1).

Hammond, P., and Sun, Y. (2001). *Monte Carlo Simulation of Macroeconomic Risk with a Continuum of Agents: The Symmetric Case* (Working Papers 01015). California, USA: Department of Economics, Stanford University.

Hayakawa, K., and Kurozumi, E. (2006). *The Role of "Leads" in the Dynamic OLS Estimation of Cointegrating Regression Models* (Hi-Stat Discussion Paper Series d06-194). Tokyo, Japan: Institute of Economic Research, Hitotsubashi University.

MacKinnon, J. G. (2002). Bootstrap Inference in Econometrics. *Canadian Journal of Economics, 35*(4).

Pedersen, A. (1995). A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations. *Scandinavian Journal of Statistics, 22*(1).

Zhu, Y., Goodwin, B. K., and Ghosh, S. K. (2011). Modeling Yield Risk Under Technological Change: Dynamic Yield Distributions and the U.S. Crop Insurance Program. *Journal of Agriculture and Resource Economics, 36*(1).