Robust Estimation and Forecasting of the Capital Asset Pricing Model*

Guorui Bian
Department of Statistics
East China Normal University

Michael McAleer
Econometric Institute
Erasmus School of Economics
Erasmus University Rotterdam
and
Tinbergen Institute
The Netherlands
and
Institute of Economic Research
Kyoto University

Wing-Keung Wong
Department of Economics
Hong Kong Baptist University

EI 2010-65

Revised: October 2010

* The third author would like to thank Robert B. Miller and Howard E. Thompson for their continuous guidance and encouragement. For financial support, the first author is grateful to East China Normal University, the second author acknowledges the Australian Research Council, National Science Council, Taiwan, and the Japan Society for the Promotion of Science, and the third author wishes to acknowledge Hong Kong Baptist University.

Abstract

In this paper, we develop a modified maximum likelihood (MML) estimator for the multiple linear regression model with underlying student t distribution. We obtain the closed form of the estimators, derive the asymptotic properties, and demonstrate that the MML estimator is more appropriate for estimating the parameters of the Capital Asset Pricing Model by comparing its performance with least squares estimators (LSE) on the monthly returns of US portfolios. The empirical results reveal that the MML estimators are more efficient than LSE in terms of the relative efficiency of one-step-ahead forecast mean square error in small samples.

JEL classifications: C1, C2, G1.

Keywords: Maximum likelihood estimators, Modified maximum likelihood estimators, Student t family, Capital asset pricing model, Robustness.

1. Introduction

The estimation of parameters in a linear regression model has received significant attention in the statistics and econometrics literature. Much of the work reported is based on the assumption of normality (Lawrence and Arthur, 1990). In recent years, however, it has been recognized that the underlying distribution is, in most situations, basically not normal, especially in economics and finance (Huber, 1981; Tiku et al., 1986). The solution, therefore, is to develop efficient estimators of the parameters in the multiple regression model when the underlying distribution is non-normal. It would be preferred to have closed-form estimators that are fully efficient, and would also be robust to plausible deviations from the assumed model. In this paper, the underlying distribution is assumed to be symmetric and student t, and the method of modified maximum likelihood (MML) estimation (Tiku, 1968; Tiku et al., 1999, 2000, 2001) is used.

This paper extends the results given in Bian and Tiku (1997), Tiku et al. (1999, 2000, 2001), Wong and Bian (2005), and Islam and Tiku (2005, 2010). Tiku et al. (1999) develop MML estimators for the simple linear regression model with symmetric innovations, Tiku et al. (2000) derive MML estimators for the first-order autoregressive model with symmetric innovations, Tiku et al. (2001) refine the MML estimator for the simple linear regression model with t distribution innovations, and Bian and Tiku (1997) adopt the Bayesian approach to examine a standard multiple regression model with independently and identically distributed (iid) errors.

This paper extends previous research to derive MML estimators for the multiple regression model with t distribution errors. The likelihood equations have no explicit solutions and have to be solved by an iterative method, which can be a formidable task. As maximum likelihood (ML) estimators are not readily available, following Tiku et al. (1999, 2000, 2001), we derive MML estimators. These estimators are explicit functions of sample observations, and hence are straightforward to compute. Moreover, they are as efficient as ML estimators (Tiku et al., 1999, 2000, 2001; Wong and Bian, 2005). We also derive the asymptotic properties of the MML estimators.

We note that MML estimators have been extensively demonstrated in simulation studies to be robust, remarkably efficient, and clearly superior to the traditional estimators under normality in all the models examined, including the autoregressive model (Tiku et al., 2000), simple linear regression model (Tiku et al., 2001), and simple linear regression model with autoregressive innovations (Tiku et al., 1999; Wong and Bian, 2005). As the multiple linear regression model is a straightforward extension of the above models, the properties of robustness and efficiency for the estimators will be similar to that of the simple linear regression model. As such, the resulting estimators developed here are explicit functions of the sample observations, and are asymptotically fully efficient. We note that Tiku et al. (1999, 2000, 2001) and Islam and Tiku (2005, 2010) have conducted extensive simulation studies for MML estimators of both the simple and multiple linear regression models. They observed that MML estimators are robust and remarkably efficient, and are clearly superior to the traditional estimators under normality.

We examine the applicability of MML estimators in finance and economics by demonstrating that MML estimators are more appropriate for estimating the parameters in the Capital Asset Pricing Model (CAPM), one of the most prominent models in finance, by comparing its performance with that of least squares estimators (LSE) on the monthly returns of US portfolios. The distributions of stock market returns have been widely analyzed by financial economists and econometricians. Fama (1963, 1965a, b) and many others have analyzed the empirical data. They have concluded that the normality assumption of a security or portfolio return is violated, such that the distribution is 'flat-tailed', and have suggested the family of stable Paretian distributions between the normal and Cauchy distributions for stock returns. On the other hand, Blattberg and Gonedes (1974) have examined security returns and suggested the student t as an alternative 'flat-tail' distribution for returns. Clark (1973), Kon (1984) and Tse (1991) suggested a mixture of normal distributions for stock returns. However, Fielitz and Rozelle (1983) suggested that a mixture of non-normal stable distributions would be a better representation of the returns distribution.

Harvey and Zhou (1993) showed that the distributional structure of returns may carry over into the structure of the disturbance in the Capital Asset Pricing Model (CAPM). In this

situation, the mixture of normal distributions, or mixture of normal and Cauchy distributions or t distributions, may give a better description of the distribution of the disturbances in CAPM. As MML estimators for simple linear regression with t innovations has been demonstrated to be robust, and based on the 'flat-tail' characteristic on the distributions of the security or portfolio returns and their corresponding disturbances in the CAPM, we recommend using the MML estimators developed in this paper for estimating the parameters of CAPM for stock returns.

In order to illustrate the superiority of the proposed MML estimators, we use one-step ahead forecasting to compare MML estimators with the traditional least squares estimators (LSE) in estimating the parameters of CAPM for US monthly stock returns. One-step ahead forecasting is commonly used to compare the performance of different models (Clements and Hendry, 1997; Chiang et al., 2009). The empirical analysis given below reveals that MML estimators are more efficient than LSE in terms of the relative efficiency of one-step-ahead forecast mean square errors in small samples.

This paper is organized as follows. We derive MML estimators in Section 2, and show their asymptotic properties in Section 3. Section 4 reviews the theory of the standard CAPM and the 'flat-tail' distribution of security returns, and demonstrates the superiority of MML estimators in CAPM. Section 5 gives some concluding remarks.

2. Modified Maximum Likelihood Estimators

Consider the multiple linear regression model:

$$y = X\beta + \boldsymbol{e} \tag{1}$$

where y is an $nx1$ vector of the observations of the endogenous variable regressed on the exogenous variables, X, an nxq ($n>q$) matrix of rank q, $\beta=(\beta_1, \ldots, \beta_q)'$ is a $qx1$ vector of regression coefficients, and $\boldsymbol{e}$ is a $nx1$ vector of random errors.

It is assumed that the innovations, e_i, are iid errors. The linear model (1) has many applications, for example, in the estimation of CAPM in the prediction of future stock prices. Numerous other applications of the above model, besides business and economics, are in agricultural, biological, and biomedical problems (see, for example, Lawrence and Arthur (1990) for further details).

Assume that the common distribution of e_i is symmetric and is, for illustrative purposes, given by

$$f(e) = \frac{1}{\sigma\sqrt{k}B(1/2, p-1/2)}\left(1+\frac{e^2}{k\sigma^2}\right)^{-p}, -\infty < e < +\infty \qquad (2)$$

where $k = 2p\text{-}3$, $p \geq 2$, and **B**(.,.) is the beta function. We note that E(e_i) = 0, V(e_i) = σ^2, and $T = \sqrt{v}e/\sigma\sqrt{k}$ has the student t distribution, with v=2p-1 degrees of freedom. For $1 \leq p < 2$, the constant k in (2) is equal to 1, in which σ is simply the scale parameter. For $p = \infty$, (2) is reduced to a normal distribution, *N(0, σ^2).*

The likelihood function for the model in (1) is given by:

$$L(\beta,\sigma) \propto \sigma^{-n}\prod_{i=1}^{n}\left[1+\left(\frac{1}{k\sigma^2}\right)\left(y_i-\sum_{j=1}^{q}x_{ij}\beta_j\right)^2\right]^{-p} \qquad (3)$$

which gives

$$\ln L(\beta,\sigma) = c - n\ln(\sigma) - p\sum_{i=1}^{n}\ln\left[1+(k\sigma^2)^{-1}\left(y_i-\sum_{j=1}^{q}x_{ij}\beta_j\right)^2\right].$$

Setting

$$z_i = \frac{1}{\sigma}\left(y_i-\sum_{j=1}^{n}x_{ij}\beta_j\right) \qquad (4)$$

and

$$g(z) = \frac{z}{1 + z^2 / k},$$

where *g(z)* is the nonlinear part of the derivative of *lnf(z), f(z)* is the standard distribution of the error term with $f(z) = c(1+z^2/k)^{-p}$, we obtain the likelihood equations, $\partial \ln L(\beta, \sigma) / \partial \beta_j = 0$ and $\partial \ln L(\beta, \sigma) / \partial \sigma = 0$, which are given in terms of the function $g(z_i)$, and hence are intractable. Solving them by iterative methods is a formidable task, and can be problematic, especially for small values of p, in which one may encounter multiple roots, slow convergence, convergence to wrong values, or even divergence (Barnett, 1966a; Lee et al., 1980; Tiku and Suresh, 1992). Pearson and Hartley (1972, p. 89) give examples where the iterations involved in determining ML estimates do not converge sufficiently rapidly. In addition, the solutions provided by different iterative methods are not necessarily identical (Barnett, 1966a).

In order to obtain efficient closed-formed estimators, we invoke Tiku's modified likelihood estimation approach, which is well established (Smith et al., 1973; Lee et al., 1980; Tan, 1985; Tiku et al., 1986, 1999, 2000, 2001; Schneider, 1986; Vaughan, 1992; Wong and Bian, 2005). Let $z_{(1)} \le z_{(2)} \le \cdots \le z_{(n)}$ (arranged in ascending order) be the order statistics of $z_i (1 \le i \le n)$, and denote *[i]* as the concomitant index of the i^{th} observation corresponding to the order statistic, $z_{([I])}$. Clearly,

$$[i] = j \quad if \quad z_i = z_{(j)} \tag{5}$$

In order to linearize the intractable term $g(z_{(i)})$, we use the first two terms of a Taylor series expansion, such that:

$$g(z_{(i)}) \approx a_i + b_i z_{(i)} \qquad (i=1,2,\ldots,n) \tag{6}$$

where

$$a_i = \frac{(2/k)t_{(i)}^3}{\left[1 + (1/k)t_{(i)}^2\right]^2}, \quad b_i = \frac{1 - (1/k)t_{(i)}^2}{\left[1 + (1/k)t_{(i)}^2\right]^2} \text{ and } t_{(i)} = E[z_{(i)}]$$

where $z_{(i)}$ and $t_{(i)}$ are the percentiles of the empirical distribution, $F_n(x)$, and theoretical distribution $F(x)$, respectively. As $g(z)$ is almost linear in any small interval (Tiku, 1968; Tiku and Suresh, 1992), under some general regularity conditions, $z_{(i)}$ converges to $t_{(i)}$ as the sample size becomes large. If $p>3$, then $b_i>0$ for $i = 1,2, ...,n$. On the other hand, if $p=\infty$ (normal distribution), then $a_i =0$ and $b_i =1$. The expected values, variances, and covariances of standardized order statistics are available (Barnett, 1966b; Vaughan, 1992, 1994; Tiku et al., 1999, 2000, 2001; Wong and Bian, 2005). Using (6), the following modified likelihood equations are obtained:

$$\frac{\partial \ln L(\beta,\sigma)}{\partial \beta_j} \cong \frac{\partial \ln L^*(\beta,\sigma)}{\partial \beta_j} = \frac{2p}{k\sigma}\sum_{i=1}^{n} x_{ij}(a_{[i]}+b_{[i]}z_i) = 0 \qquad (j=1,2,\ldots,q),$$

$$\frac{\partial \ln L(\beta,\sigma)}{\partial \sigma} \cong \frac{\partial \ln L^*(\beta,\sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p}{k\sigma}\sum_{i=1}^{n} z_i(a_{[i]}+b_{[i]}z_i) = 0\ . \qquad (7)$$

Solving the estimating equations (7), we obtain the MML estimators:

$$\hat{\beta} = \hat{\beta}_w + D\hat{\sigma}$$

and

$$\hat{\sigma} = \frac{1}{2n}(B + \sqrt{B^2+4nC}) \qquad (8)$$

where

$$\hat{\beta}_w = (X'WX)^{-1}X'Wy\ ,$$

$$D = (X'WX)^{-1}X'a\ ,$$

$$B = (2p/k)y'[I - WX(X'WX)^{-1}X']a\ ,$$

$$C = (2p/k)\left\|y - X\hat{\beta}_w\right\|_W^2\ ,$$

$$W = Diagnol\left(b_{[1]}, b_{[2]}, \cdots b_{[n]}\right),$$

$$a' = \left(a_{[1]}, a_{[2]}, \cdots a_{[n]}\right)\ \text{, and}$$

$$\|x\|_W = x'Wx\ .$$

It is clear that all of the above MML estimators have closed-formed algebraic expressions and are, therefore, easy to compute. From (8), the MML estimator of β is found to consist of two components, with the main component $\hat{\beta}_w$ being an LSE of β, and is unbiased for β. For $p = \infty$ (normally distributed errors), $a_i = 0$, $b_i = 1$ and $2p/k = 1$. Consequently, the MML estimators (8) are reduced to the LSE. For computation, we first calculate the usual LS estimates of β and σ, which are used as initial estimates to compute z_i . We then order $z_i (1 \le i \le n)$, and compute the MML estimators of β and σ from (8). Replacing the LSE by their MML counterparts, we repeat the computation for further iterations until convergence. In all our computations, some of which are presented in this paper, no more than three iterations were needed for the estimates to converge.

3. Asymptotic Properties of MML Estimators

The asymptotic properties of the MML estimators are summarized in the following two lemmas:

Lemma 1. *The MML estimators $\hat{\beta}$ and $\hat{\sigma}$ are asymptotically unbiased for β and σ, respectively.*

Lemma 2. *The asymptotic variances and covariance of β and σ are given by:*

(I) $Cov(\hat{\beta}, \hat{\sigma}) = 0$ *;*

(ii) $$V(\hat{\beta}) = \frac{p - 3/2}{p} E^{-1}(b_{[1]})(X'X)^{-1}\sigma^2 \rightarrow \frac{(p+1)(p-3/2)}{p(p-1/2)}(X'X)^{-1}\sigma^2 ; \qquad (9)$$

(iii) $$V(\hat{\sigma}) \cong \frac{p+1}{p-1/2}\frac{\sigma^2}{2n} .$$

The proofs of Lemmas 1 and 2 are given in the Appendix.

Solving the differential equations (7), we obtain the modified likelihood function:

$$L^*(\hat{\beta},\hat{\sigma}) \propto \sigma^{-n} \exp\left[\frac{-1}{2\sigma^2}\left\{h(\hat{\beta}-\beta)'(X'X)(\hat{\beta}-\beta)+(n-q)\hat{\sigma}^2\right\}\right]G(y) \qquad (10)$$

where $G(y)$ is an analytical function free of β and σ; and

$$h = p/(p\text{-}3/2)\mathrm{E}(b_{[1]}) \rightarrow p(p\text{-}1/2)/[(p\text{+}1)(p\text{-}3/2)]. \qquad (11)$$

As the likelihood function-like L^* in (10) is asymptotically equivalent to the corresponding likelihood function L in (3) (Tan, 1985), the asymptotic properties of $\hat{\beta}$ and $\hat{\sigma}$ follow immediately, as shown in the following lemma:

Lemma 3.

(i) The vector $\hat{\beta}$ has a q-variate normal distribution, with mean vector β and covariance matrix given in Lemma 2;

(ii) $(n-q)\hat{\sigma}^2/\sigma^2$ is distributed as chi-square, with n-q degrees of freedom; and

(iii) $\hat{\beta}$ and $\hat{\sigma}$ are independent.

In addition, following the argument of Vaughan (1992), a close approximation of the joint distribution of $\hat{\beta}$ and $\hat{\sigma}$ is given by

$$\sigma^{-n} \exp\left[\frac{-1}{2\sigma^2}\left\{h_1(\hat{\beta}-\beta)'(X'X)(\hat{\beta}-\beta)+(n-q)\hat{\sigma}^2\right\}\right] \qquad (12)$$

where h_1 is an adjusted value of h in (11) given by

$$h_1^{-1} = \frac{2}{n}\left[\frac{\Gamma(n/2)}{\Gamma[(n-1)/2]}\right]^2 h^{-1} \qquad (13)$$

(see Bian and Tiku (1997) for further details). We note that, for large n, $h_1 = h$.

4. Application in Finance

In this section, we examine the perfoemance of applying MML estimators in finance, with an illustration of the Capital Asset Pricing Model (CAPM) on monthly returns of US portfolios. We hypothesize that MML estimators are more appropriate for CAPM, a parsimonious general equilibrium model (Sharpe, 1963; Lintner, 1965) whose excess returns, R, on a security from the risk-free rate, R_f, is formulated by:

$$R_i = a_i + b_i R_m + e_i , \tag{14}$$

where R_i (R_m) is the excess returns of portfolio i (market portfolio) from the risk-free rate, R_f , a_i measures the abnormal performance of the portfolio i, b_i measures the portfolio's level of systematic risk in relation to the market portfolio, and e_i is the random error term, with zero expectation.

We choose CAPM for MML as it is one of the simplest models in finance, yet sufficiently complicated that LSE does not perform well. If MML outperforms LSE for this simple model, MML would be expected to outperform more complicated models in finance and economics, such as Arbitrary Pricing Theory. Although CAPM may appear simple, as shown in (14), it is complicated as the measure of beta is empirically nonstationary over time (Leavy, 1971; Blume, 1975). Moreover, the distributions of both the security or portfolio returns and the disturbance are 'flat-tailed', and hence violate the normality assumption (Fama, 1963, 1965a, b; Pettit and Westerfield, 1974).

In order to handle the non-stationarity of beta, one may estimate the model for a reasonably short period in order to capture stationary beta (Wong and Bian, 2000). Thus, to handle the 'flat-tailed' distribution for both returns and disturbance, Fama (1963, 1965a, b) suggested the family of stable Paretian distributions between normal and Cauchy distributions for stock returns, while Blattberg and Gonedes (1974) examined security returns and suggested the student t as an alternative 'flat-tailed' distribution. Clark (1973), Kon (1984), and Tse (1991) recommended a mixture of normal distributions for stock returns, while Fielitz and Rozelle (1983) argued that a mixture of non-normal stable distributions would be a better

representation of the distribution of security or portfolio returns. In this paper, we demonstrate that MML with t distribution innovations will be a good approach to handle non-normality as MML has been studied extensively (see, for example, Tiku et al., 1999, 2000, 2001) to be sufficiently robust to represent many different distributions, including a family of t distributions, a mixture of normal distributions, and a mixture of non-normal stable distributions.

For comparison, we use the same dataset as in Harvey and Zhou (1993) and Wong and Bian (2000), in which twelve industrial portfolios of US monthly data are examined. The industry classifications conform to Sharpe (1982), Breeden et al. (1989) and Gibbons et al. (1989). The portfolios are value weighted, the market return is the weighted NYSE return, and monthly returns for the period 1926-1987 are in excess of the 30-day Treasury Bill rate. The portfolio returns are available from the Center for Research in Security Prices (CRSP) at the University of Chicago, while the 30-day Treasury Bill rate is available from Ibbotson Associates.

In this paper, we hypothesize that MML estimators are more appropriate to estimate the parameters of CAPM and to forecast using CAPM as it is more efficient than LSE. As an illustration, we use the robust MML with the student t distribution and 7 degrees of freedom to analyze CAPM for US monthly stock returns. Twelve industrial portfolios of US data are used. We adopt the one-step-ahead forecast bias and MSE as the basis to evaluate the performance of LSE and MML for CAPM. In the computations, we choose a small sample size n of 12, namely for a one-year period) to capture a stationary b parameter.

We compute the skewness and kurtosis coefficients and the Jarque-Bera statistic for the returns and the corresponding residuals in the CAPM to test the normality hypothesis for both excess returns, R, and their corresponding disturbances, e, in (14). The results are shown in Table 1. Several other statistics can be used to test normality, such as the modified Shapiro-Wilk statistic, Anderson-Darling test and Kolmogorov-Smirnov test. However, as the Jarque-Bera statistic is one of the most powerful tests of normality, and the results of the other statistics are similar, we report only the results of the Jarque-Bera statistic and its corresponding skewness and kurtosis coefficients. The 1% level of significance shown in the

table leads to rejection of the null hypothesis of normality for the monthly excess returns, as well as their corresponding disturbances. These findings support the hypothesis that the non-normality in the returns will carry over into the non-normality of the disturbances in the CAPM (Harvey and Zhou, 1993). We note that the returns depart from normality, which may be attributed to ARCH or GARCH effects (Qiao et al., 2008a, b, c). However, temporal aggregation will reduce any ARCH or GARCH effects (see, for example, Drost and Nijman, 1995).

[Table 1 goes here]

We use the one-step-ahead forecast MSE (see, for example, Clements and Hendry, 1997; Wong and Bian, 2000; Chiang et al., 2009) for further details) as a basis for comparison between LSE and MML for the US monthly data. For the given sample size n=12, the estimates of both LSE and MML are computed for each of the 12 industrial portfolios for $t = n,..., T-1$. Then we compute their one-step-ahead forecasts by applying both LSE and MMLE to each portfolio for $t = n+1,... ,T$. After that, the one-step-ahead forecast bias and MSE for each portfolio are calculated. Their relative efficiency (REF), the ratio of average one-step-ahead of forecast MSE for both LSE and MML, is computed and displayed in Table 2. We note that the values of the bias and MSE in the table are 1000 times the original values.

[Table 2 goes here]

From Table 2, we find that MML has both smaller one-step-ahead forecast bias and smaller MSE, and is more efficient than LSE in all industries, except Finance and Real Estate and Transportation. The average values also show that MML attains a smaller average one-step-ahead forecast bias (-0.000717) and smaller one-step-ahead forecast MSE (0.0007087) than those of LSE, -0.0009641 and 0.0007140, respectively, with average relative efficiency of 1.0082. This implies that the MML estimators are far more efficient and robust than their LSE counterparts.

5. Concluding Remarks

It is generally recognized that nonnormal samples occur frequently in practice. In this paper, we extended the results of Bian and Tiku (1997), Tiku et al. (1999, 2000, 2001), and Wong and Bian (2005) to the linear model by assuming the innovations to be asymmetric and from a student t distribution. The likelihood equations are intractable. Solving them by iterative methods is tedious and time consuming, and the results obtained may be unreliable. Therefore, we used modified likelihood estimation.

In the context of iid random sampling and survey sampling, this method is known to yield asymptotically fully efficient estimators (Tiku, 1970; Bhattacharyya, 1985), and almost as efficient as maximum likelihood estimators for small n (Smith et al., 1973; Tan, 1985; Schneider, 1986; Tiku and Suresh, 1992; Vaughan, 1992). An attractive feature of the method is that it yields MML estimators which can be expressed explicitly as functions of sample observations and are, therefore, easy to compute and can be studied analytically. We derived the MML estimators here in the context of linear models. These estimators are as attractive as in the classical framework of iid random observations. We demonstrated their very high efficiency that is not shared by their Gaussian counterparts.

One might consider incorporating the Bayesian approach (Matsumura, et al., 1990; Bian and Tiku, 1997; Wong and Bian, 2000) into MML estimation. We note that the distribution of the stock returns in our illustration is not only heavy-tailed, but is also strongly skewed. Hence, it is possible to improve forecasting by using skewed error distributions or other distributions (see, for example, Fong and Wong, 2006). The MML with asymmetric innovations would be an interesting issue as an extension. Further extensions include the applicability of the MML for linear models to other prominent economics or finance models (as in, for example, Fong et al., 2005, 2008). Another possible area for further research is to compare the beta coefficients estimated above with the equity cost of capital for each portfolio (Thompson and Wong, 1991, 1996; Wong and Chan, 2004).

There are many other approaches in the analysis of linear models, for example, no distributional assumptions on the measurement errors (Wong and Miller 1990), and other models such as nonlinear regression models (Amemiya, 1985; Hsiao, 1989; Hausman et al., 1998, Honore and Hu, 2004). Nevertheless, it is well known that if the distribution of the

disturbances is known to be from a student t distribution, parametric approaches yield estimators which outperform estimators without distributional assumptions (Li and Hsiao, 2004). As such, our approach performs better when the distribution is known. In addition, the MML approach could be incorporated to improve estimation in other models, such as nonlinear regression models.

Applying our approach to estimate the parameters in the Capital Asset Pricing Model or other financial models could enable investors to obtain better estimates, thereby leading to better decision in investments. In order to provide more helpful information for decision making, investors could incorporate our approach with other theories, as in behaviourial finance (Lam et al., 2010a, b), technical analysis (Wong et al., 2001, 2003; Kung and Wong, 2009; Wong and McAleer, 2009), stochastic dominance theory (Wong and Li, 1999; Wong, 2007; Wong and Ma, 2008; Sriboonchita et al., 2009), portfolio optimization (Bai et al, 2009a, b; Egozcue and Wong, 2010). They may obtain more information from markets, as in information company performance (Thompson and Wong, 1991, 1996; Wong and Chan, 2004), other measurement techniques (Leung and Wong, 2008; Qiao et al, 2009; Ma and Wong, 2010; Bai et al., 2010), information of some economic/financial phenomena (Broll et al., 2006; Gasbarro et al., 2007; Wong et al., 2008), and incorporate the behaviour of other investors (Li and Wong, 1999; Wong and Chan, 2008; Broll et al., 2010) to lead to better decision making.

Table 1. Tests for Departure from Normality of Monthly Excess Portfolio Returns and the Corresponding Residuals in CAPM by Industrial Classifications

	Returns			Residuals		
Portfolio	**Skewness**	**kurtosis**	**Jarque-Bera statistic**	**Skewness**	**kurtosis**	**Jarque-Bera statistic**
NYSE value-weighted	0.3059**	10.6030**	1803.58**	---	---	
Petroleum	0.3103**	7.4277**	619.68**	0.2477**	4.1315**	47.30**
Finance & Real Estate	0.2257**	10.6255**	1808.91**	0.006	4.7600**	96.03**
Consumer Durables	1.0134**	15.3646**	4866.73**	0.6193**	10.7926**	1930.02**
BasicIndustries	0.8691**	13.6209**	3590.57**	0.6333**	9.6177**	1407.35**
Food &Tobacco	0.0178	10.1611**	1589.76**	-0.1866*	4.9496**	122.15**
Construction	0.8995**	11.5376**	2359.94**	0.5306**	6.6211**	441.39**
Capital Goods	0.2375**	9.0959**	1158.95**	0.1785*	4.7571**	99.66**
Transportation	1.1614**	15.2275**	4802.12**	1.1199**	8.7320**	1174.05**
Utilities	0.1446	10.7665**	1872.47**	-0.0405	5.0824**	134.63**
Textile & Trade	0.1218	8.6145**	979.04**	-0.094	4.8637**	108.77**
Services	0.0349	7.0560**	510.14**	0.3336**	11.8533**	2443.61**

Note: * $p < 5\%$, ** $p < 1\%$

Table 2. One-step Ahead Forecast Bias and MSE of MML and LS for US Monthly Stock Returns

Portfolio	LS Method		MML method		REF
	Bias	MSE	Bias	MSE	
Petroleum	-1.614	1.1228	-1.588	1.1173	1.005
Finance & Real Estate	-0.296	0.3525	-0.275	0.3555	0.992
Consumer Durables	-0.980	0.4021	-0.611	0.3939	1.021
Basic Industries	-0.827	0.2482	-0.644	0.2476	1.002
Food & Tobacco	-0.313	0.5971	-0.121	0.5917	1.009
Construction	-1.139	0.6678	-1.034	0.6542	1.021
Capital Goods	-0.463	0.3329	-0.376	0.3318	1.003
Transportation	-0.709	0.8608	-0.853	0.8511	1.011
Utilities	-0.667	0.8865	-0.380	0.8858	1.001
Textile & Trade	-1.365	0.7954	-0.570	0.7767	1.024
Services	-1.774	1.5085	-1.604	1.5187	0.993
Recreation	-1.423	0.7940	-0.554	0.7806	1.017
Average	-0.9641	0.7140	-0.7175	0.70874	1.0082

Note: The bias and MSE in the table are 1000 times the original values.

Appendix

Proof of Lemma 1:

The result follows immediately from the first two terms of the Taylor series expansions of $\partial \ln L^* / \partial \beta_j$ (j=1,2,...,q) and $\partial \ln L^* / \partial \sigma$, and the fact that $1/n \mid \partial \ln L / \partial \beta_j - \partial \ln L^* / \partial \beta_j \mid$ and $1/n \mid \partial \ln L / \partial \sigma - \partial \ln L^* / \partial \sigma \mid$ tend to zero as n tends to infinity (Kendall and Stuart, 1979, Chapter 18).

Proof of Lemma 2:

From the symmetry of the student t distribution, it follows immediately that

$$E(a_i * b_{[1]}, b_{[2]}, ..., b_{[n]}) = 0, \text{ for all I = 1, 2, ..., n}$$

and

$$E(\mathbf{e} * b_{[1]}, b_{[2]}, ..., b_{[n]}) = \mathbf{0}, \tag{15}$$

where $\mathbf{e} = y - X\beta$. Thus,

$$E(\hat{\beta}_w) = \beta + E\left[(X'WX)^{-1} X'WE(e \mid b_{[1]}, b_{[2]}, \cdots, b_{[n]})\right] = \beta$$

and

$$E(X'WX)^{-1} X'a) = E\left[(X'WX)^{-1} X'E(a \mid b_{[1]}, b_{[2]}, \cdots, b_{[n]})\right] = \beta \ .$$

Therefore, the MML estimator $\hat{\beta}$ is unbiased for β as σ is known.

The asymptotic covariance matrix of the MML estimators is given by the inverse of

$$I(\beta, \sigma) \cong \begin{bmatrix} -E\left(\frac{\partial^2 L^*(\beta,\sigma)}{\partial\beta\partial\beta}\right) & -E\left(\frac{\partial^2 L^*(\beta,\sigma)}{\partial\beta\partial\sigma}\right) \\ -E\left(\frac{\partial^2 L^*(\beta,\sigma)}{\partial\sigma\partial\beta}\right) & -E\left(\frac{\partial^2 L^*(\beta,\sigma)}{\partial\sigma\partial\sigma}\right) \end{bmatrix}.$$

As

$$\frac{\partial^2 L^*(\beta,\sigma)}{\partial\beta_j\partial\beta_l} = -\frac{2p}{k\sigma^2}\sum_{i=1}^{n} b_{[i]}x_{ij}x_{il} \ ,$$

we have

$$-E\left(\frac{\partial^2 L^*(\beta,\sigma)}{\partial\beta\partial\beta}\right) = \frac{2p}{k\sigma^2}X'E(W)X = \frac{2p}{k\sigma^2}E(b_{[1]})X'X$$

where

$$E(b_{[1]}) = \frac{1}{n}\sum_{i=1}^{n} b_i \to \frac{p-1/2}{p+1} \quad \text{as } n\to\infty \ .$$

Similarly,

$$\frac{\partial^2 L^*(\beta,\sigma)}{\partial\beta_j\partial\sigma} = -\frac{2p}{k\sigma^2}\left(\sum_{i=1}^{n}\alpha_{[i]}x_{ij} + 2\sum_{i=1}^{n}x_{ij}b_{[i]}z_i\right)$$

and

$$-E\left(\frac{\partial^2 L^*(\beta,\sigma)}{\partial\beta\partial\sigma}\right) = \frac{2p}{k\sigma^2}\left[X'E(a) + 2X'E(Wz)\right] = 0 \ .$$

This follows immediately from (15) and the fact that E($a_{[i]}$) = (1/n) 3a_i = 0. Moreover,

$$\frac{\partial^2 L^*(\beta,\sigma)}{\partial^2\sigma} = \frac{n}{\sigma^2} - \frac{2p}{k\sigma^2}\sum_{i=1}^{n}(2\alpha_{[i]}z_i + 3b_{[i]}z_i^2)$$

which gives (see, for example, Bian and Tiku, 1997):

$$-\frac{1}{n}E\left(\frac{\partial^2 L^*(\beta,\sigma)}{\partial^2\sigma}\right) \to \frac{p-1/2}{p+1}\frac{2}{\sigma^2} \quad \text{as} \quad n\to\infty .$$

References

Amemiya, Y., 1985. Instrumental variable estimator for the nonlinear errors-in-variables model. Journal of Econometrics, 28, 273–290.

Bai, Z.D., H.X. Liu and W.K. Wong (2009a). Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory. Mathematical Finance, 19(4), 639-667.

Bai, Z.D., H.X. Liu and W.K. Wong (2009b). On the Markowitz mean-variance analysis of self-financing portfolios. Risk and Decision Analysis, 1(1), 35-42.

Bai, Z.D., W.K. Wong, B. Zhang, (2010). Multivariate linear and non-linear causality tests, Mathematics and Computers in Simulation, 81, 5–17.

Barnett, V.D., 1966a. Evaluation of the maximum likelihood estimator where the likelihood equation has multiple roots. Biometrika, 53, 151-165.

Barnett, V.D. 1966b. Order statistics estimators of the location of the Cauchy distribution. Journal of the American Statistical Association, 61, 1205-1218.

Bhattacharyya, G.K., 1985. The asymptotics of maximum likelihood and related estimators based on Type II censored data. Journal of the American Statistical Association, 80, 398-404.

Bian, G., Tiku, L.M., 1997. Bayesian inference based on robust priors and MML estimators: Part I, symmetric location-scale distributions. Statistics, 29, 317-345.

Bian, G., Wong, W.K., 1997. An alternative approach to estimate regression coefficients. Journal of Applied Statistical Science, 6, 21-44.

Blattberg, R.C., Gonedes, N.J.,1974. A comparison of stable and student distribution as statistical models for stock prices, Journal of Business, 47, 244-280.

Blume, M.E., 1975. Betas and their regressions tendencies. Journal of Finance, 30, 785-795.

Breeden, D.T., Gibbons M., Litzenberger, R.H., 1989. Empirical tests of the consumption based on CAPM, Journal of Finance, 44, 231-262.

Broll, U., M. Egozcue, W.K. Wong, and R. Zitikis, (2010), Prospect theory, indifference curves, and hedging risks, Applied Mathematics Research Express, 2010(2), 142–153.

Broll, U., J.E. Wahl, and W.K. Wong, 2006, Elasticity of risk aversion and international trade, Economics Letters, 91(1), 126-130.

Chiang, T.C., Qiao, Z., Wong, W.K., 2009, New evidence on the relation between return volatility and trading volume, Journal of Forecasting, 29(5), 502 – 515.

Clark, P.K., 1973. A subordinated stochastic process model with finite variance for speculative prices. Econometrica, 37, 135-155.

Clements, M.P., Hendry, D.F., 1997. An empirical study of seasonal unit roots in forecasting.

International Journal of Forecasting, 13, 341-355.

Drost, F.C., Nijman, T.E., 1995. Temporal aggregation of GARCH processes, in Engle, R.F., ed. ARCH: Selected Readings, Oxford University Press.

Egozcue, M., Wong, W.K., 2010. Gains from diversification: A majorization and stochastic dominance approach. European Journal of Operational Research, 200, 893–900.

Fama, E.F., 1963. Mandelbrot and the stable Paretian hypothesis. Journal of Business, 36, 420-429.

Fama, E.F., 1965a. The behaviour of stock market prices, Journal of Business, 38, 34-105.

Fama, E.F., 1965b, Portfolio analysis in a stable Paretian market, Management Science, 11, 401-419.

Fielitz, B.D., Rozelle, J.P., 1983. Stable distributions and mixtures of distributions hypotheses for common stock return, Journal of the American Statistical Association, 78, 28-36.

Fong, W.M., Lean, H.H., Wong, W.K., 2008. Stochastic dominance and behavior towards risk: the market for internet stocks. Journal of Economic Behavior and Organization, 68(1), 194-208.

Fong, W.M., Wong, W.K., 2006. The modified mixture of distributions model: A revisit, Annals of Finance, 2(2), 167 – 178.

Fong, W.M., Wong, W.K., Lean, H.H., 2005. Stochastic dominance and the rationality of the momentum effect across markets, Journal of Financial Markets, 8, 89-109.

Gasbarro, D., Wong, W.K., Zumwalt, J.K., 2007. Stochastic dominance analysis of iShares, European Journal of Finance, 13, 89–101.

Gibbons, M.R., Ross, S.A., Shanken, J., 1989. A test of efficiency of a given portfolio, Econometrica, 57, 1121-1152.

Harvey, M.C., Zhou, G., 1993. International asset pricing with alternative distributional specifications, Journal of Empirical Finance, 1, 107-131.

Hausman, J.A., Abrevaya, J., Scott-Morton, F.M., 1998. Misclassification of the dependent variable in a discrete-response setting. Journal of Econometrics, 87, 237–269.

Honor, B.E., Hu, L.J., 2004. Estimation of cross sectional and panel data censored regression models with endogeneity. Journal of Econometrics, 122, 293– 316.

Hsiao, C., 1989. Consistent estimation for some nonlinear errors-in-variables models. Journal of Econometrics, 41, 159–185.

Huber, P.J., 1981. Robust Statistics. Wiley, New York.

Islam, M.Q., Tiku, M.L. 2005. Multiple linear regression model under nonnormality, Communications in Statistics - Theory and Methods, 33(10), 2443-2467.

Islam, M.Q., Tiku, M.L. 2010. Multiple linear regression model with stochastic design variables, Journal of Applied Statistics, 37(6), 923 - 943.

Kendall, M.G. Stuart, A., 1979. The Advanced Theory of Statistics. Charles Gri.n, London.

Kon, S.J., 1984. Models of stock returns - a comparison. Journal of Finance, 39, 147-165.

Kung, J.J., Wong, W.K. 2009. Efficiency of Taiwan stock market, Japanese Economic Review, 60(3), 389-394.

Lam, K., Liu, T., Wong, W.K. 2010a. A pseudo-Bayesian model in financial decision making with implications to market volatility, under- and overreaction, European Journal of Operational Research, 203(1), 166-175.

Lam, K., Liu, T., Wong, W.K. 2010b, A new pseudo Bayesian model with implications to financial anomalies and investors' behaviors, Journal of Behavioral Finance, to appear.

Lawrence, K.D., Arthur, J.L., 1990. Robust Regression. Marcel Dekker, New York.

Leavy, R.A., 1971. On the short term stationarity of Beta coefficient. Financial Analysis Journal, 27, 55-62.

Lean, H.H., McAleer, M., Wong, W.K., (2010), Market efficiency of oil spot and futures: A mean-variance and stochastic dominance approach, Energy Economics, 32, 979–986.

Lee, K.R., Kapadia, C.H., Dwight, B.B., 1980. On estimating the scale parameters of the Rayleigh distribution from doubly censored samples. Statist. Hefte, 21, 14-29.

Leung, P.L., Wong, W.K., 2008, On testing the equality of the multiple Sharpe ratios, with application on the evaluation of IShares, Journal of Risk, 10(3), 1-16.

Li, C.K., Wong, W.K., 1999. Extension of stochastic dominance theory to random variables. RAIRO Recherche Opérationnelle, 33(4), 509-524.

Li, T., Hsiao, C., 2004. Robust estimation of generalized linear models with measurement errors. Journal of Econometrics, 118, 51–65.

Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics, 47, 13-37.

Ma, C. and W.K. Wong, (2010) Stochastic dominance and risk measure: A decision-theoretic foundation for VaR and C-VaR, European Journal of Operational Research, 207, 927-935.

Matsumura, E.M., Tsui, K.W., Wong, W.K., 1990. An extended multinomial-Dirichlet model for error bounds for dollar-unit sampling, Contemporary Accounting Research, 6(2-I), 485-500.

Pearson, E.S., and Hartley, H.O, 1972. Biometrika Tables for Statisticians, Vol II: University Press, Cambridge.

Pettit, R.R. and Westerfield, R., 1974. Using the capital asset pricing model and returns. Journal of Financial and Quantitative Analysis, 9, 579-605.

Qiao, Z., Chiang, T.C. and Wong, W.K., 2008a. Long-run equilibrium, short-term adjustment, and spillover effects across Chinese segmented stock markets, Journal of International Financial Markets, Institutions & Money, 18, 425–437.

Qiao, Z., Li, Y. and Wong, W.K., 2008b. Policy change and lead-lag relations among China’s segmented stock markets, Journal of Multinational Financial Management, 18, 276–289.

Qiao, Z., McAleer, M. and Wong, W.K., 2009. Linear and nonlinear causality between changes in consumption and consumer attitudes, Economics Letters, 102(3), 161–164.

Qiao, Z., Smyth, R. and Wong, W.K., 2008c. Volatility switching and regime interdependence between information technology stocks 1995-2005, Global Finance Journal, 19, 139-156.

Schneider, H., 1986. Truncated and censored samples from normal populations. New York Marcel Dekker.

Sharpe, W., 1963. A simplified model for portfolio analysis. Management Science, 9, 277-293.

Sharpe, W., 1982. Factors in New York Stock Exchange security returns, 1931-1979. Journal of Portfolio Management, 8, 5-19.

Smith, W.B., Zeis, C.D., Syler, G.W., 1973. Three parameter lognormal estimation from censored data. Journal of Indian Statistical Association, 11, 15-31.

Sriboonchita, S., Wong, WK. S. Dhompongsa, Nguyen, H.T., 2009. Stochastic Dominance and Applications to Finance, Risk and Economics, Chapman and Hall/CRC, Taylor and Francis, Boca Raton, Florida, USA.

Tan, W.Y., 1985. On Tiku's robust procedure - a Bayesian insight. Journal of Statistical Planning and Inference, 11, 329-340.

Thompson, H E, Wong, W.K., 1991. On the unavoidability of 'unscientific' judgment in estimating the cost of capital, Managerial and Decision Economics, 12, 27-42.

Thompson, H E, Wong, W.K., 1996. Revisiting 'dividend yield plus growth' and its applicability, Engineering Economist, 41(2), 123-147.

Tiku, M.L., 1968. Estimating the parameters of log-normal distribution from censored samples. Journal of the American Statistical Association, 63, 134-140.

Tiku, M.L., 1970. Monte Carlo study of some simple estimators in censored normal samples. Biometrika, 57, 207-211.

Tiku, M.L., Islam, M.Q., Selcuk, A.S., 2001. Non-normal regression, II: Symmetric distributions. Communications in Statistics: Theory and Methods, 30, 1021-1045.

Tiku, M.L., Suresh, R.D., 1992. A new method of estimation for location and scale parameters. Journal of Statistical Planning and Inference, 30, 281-292.

Tiku, M.L., Tan, W.Y., Balakrishnan, N., 1986. Robust Inference. Marcel Dekker, New York.

Tiku, M.L., Wong, W.K., Bian, G., 1999. Estimating parameters in autoregressive models in non-normal situations: Symmetric innovations. Communications in Statistics: Theory and Methods, 28(2), 315-341.

Tiku, M.L., Wong, W.K., Vaughan, D.C., Bian, G., 2000. Time series models with nonnormal innovations: Symmetric location-scale distributions. Journal of Time Series Analysis, 21(5), 571-596.

Tse, Y.K., 1991. Price and volume in the Tokyo stock exchange: An exploratory study, in Ziemba, W., Bailey, T., Hamao, W., (eds.). Japanese Financial Market Research, 91-119.

Vaughan, D.C., 1992. On the Tiku-Suresh method of estimation. Communications in Statistics: Theory and Methods, 21, 451-469.

Vaughan, D.C., 1994. The exact values of the expected values, variances, and covariances of order statistics from the Cauchy distribution. Journal of Statistical Computation and Simulation, 49, 21-32.

Wong, W.K., 2007, Stochastic dominance and mean-variance measures of profit and loss for business planning and investment. European Journal of Operational Research, 182, 829-843.

Wong, W.K., Bian, G., 2000. Robust Bayesian inference in asset pricing estimation, Journal of Applied Mathematics and Decision Sciences, 4(1), 65-82.

Wong, W.K., Bian, G., 2005. Estimating parameters in autoregressive models with asymmetric innovations. Statistics and Probability Letters, 71, 61-70.

Wong, W.K., Chan, R., 2004. The estimation of the cost of capital and its reliability. Quantitative Finance, 4(3), 365-372.

Wong, W.K., Chan, R., 2008. Markowitz and prospect stochastic dominances. Annals of Finance, 4(1), 105-129.

Wong, W.K., Chew, B.K., Sikorski, D., 2001. Can P/E ratio and bond yield be used to beat stock markets?, Multinational Finance Journal, 5(1), 59-86.

Wong, W.K., Li, C.K., 1999. A note on convex stochastic dominance theory. Economics Letters, 62, 293-300.

Wong, W.K., Ma, C., 2008. Preferences over Meyer's location-scale family. Economic Theory, 37(1), 119-146.

Wong W.K., Manzur, M., Chew, B.K., (2003), How rewarding is technical analysis? Evidence from Singapore stock market, Applied Financial Economics, 13(7), 543-551.

Wong, W.K., McAleer, M., 2009, Mapping the Presidential Election Cycle in US stock markets, Mathematics and Computers in Simulation, 79(11), 3267-3277.

Wong, W.K., Miller, R.B., 1990. Analysis of ARIMA-noise models with repeated time series. Journal of Business and Economic Statistics, 8(2), 243-250.

Wong, W.K., Phoon, K.F., Lean, H.H., 2008. Stochastic dominance analysis of Asian hedge funds. Pacific-Basin Finance Journal, 16(3), 204-223.