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 2010-62

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 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:

\[ y = X\beta + e \quad (1) \]

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

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

(2)

where \( k = 2p-3, \ p \geq 2, \) and \( B(., .) \) is the beta function. We note that \( E(e_i) = 0, \ V(e_i) = \sigma^2, \) and \( T = \sqrt{ve / \sigma \sqrt{k}} \) 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 \( \sigma \) is simply the scale parameter. For \( p = \infty, \ (2) \) is reduced to a normal distribution, \( N(0, \sigma^2). \)

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

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

(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

\[
z_i = \frac{1}{\sigma} \left(y_i - \sum_{j=1}^{q} x_{ij} \beta_j\right)
\]  

(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, \(\frac{\partial \ln L(\beta, \sigma)}{\partial \beta_j} = 0\) and \(\frac{\partial \ln L(\beta, \sigma)}{\partial \sigma} = 0\), 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 \(z_{(1)} \leq z_{(2)} \leq \cdots \leq z_{(n)}\) (arranged in ascending order) be the order statistics of \(z_i (1 \leq i \leq n)\), and denote \(\lfloor i \rfloor\) as the concomitant index of the \(i^{th}\) observation corresponding to the order statistic, \(z_{\lfloor i \rfloor}\). Clearly,
\[
\lfloor i \rfloor = j \quad \text{if} \quad z_i = z_{(j)} \quad (5)
\]

In order to linearize the intractable term \(g(z_{\lfloor i \rfloor})\), we use the first two terms of a Taylor series expansion, such that:
\[
g(z_{\lfloor i \rfloor}) \approx a_i + b_i z_{\lfloor i \rfloor} \quad (i=1,2,\ldots,n) \quad (6)
\]

where
\[
a_i = \frac{(2/k)t_{\lfloor i \rfloor}^2}{[1 + (1/k)t_{\lfloor i \rfloor}^2]^2} , \quad b_i = \frac{1-(1/k)t_{\lfloor i \rfloor}^2}{[1 + (1/k)t_{\lfloor i \rfloor}^2]^2} \quad \text{and} \quad t_{\lfloor i \rfloor} = \text{E}[z_{\lfloor i \rfloor}]\]
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 \( i = 1, 2, \ldots, n \). 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:

\[
\frac{\partial \ln L(\beta, \sigma)}{\partial \beta_j} \approx \frac{\partial \ln L^*(\beta, \sigma)}{\partial \beta_j} = \frac{2p}{k\sigma} \sum_{i=1}^{n} x_i (a_i + b_i z_i) = 0 \quad (j=1,2,\ldots,q),
\]

\[
\frac{\partial \ln L(\beta, \sigma)}{\partial \sigma} \approx \frac{\partial \ln L^*(\beta, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p}{k\sigma} \sum_{i=1}^{n} z_i (a_i + b_i z_i) = 0 .
\]

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

\[
\hat{\beta} = \hat{\beta}_w + D\hat{\sigma}
\]

and

\[
\hat{\sigma} = \frac{1}{2n} \left( B + \sqrt{B^2 + 4nC} \right)
\]

where

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

\[
D = (X'^{\prime}WX)^{-1}X'^{\prime}a,
\]

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

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

\[
W = \text{Diag}(b_{(1)}, b_{(2)}, \ldots, b_{(n)}) ,
\]

\[
a' = (a_{(1)}, a_{(2)}, \ldots, a_{(n)}) ,
\]

\[
\| x \|_w = x'^{\prime}Wx .
\]
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 $\beta$ is found to consist of two components, with the main component $\hat{\beta}_w$ being an LSE of $\beta$, and is unbiased for $\beta$. 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 $\beta$ and $\sigma$, which are used as initial estimates to compute $z_i$. We then order $z_i (1 \leq i \leq n)$, and compute the MML estimators of $\beta$ and $\sigma$ 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 $\beta$ and $\sigma$, respectively.

**Lemma 2.** The asymptotic variances and covariance of $\beta$ and $\sigma$ are given by:

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

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

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

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

Solving the differential equations (7), we obtain the modified likelihood function:
\[ L' (\hat{\beta}, \hat{\sigma}) \propto \sigma^{-n} \exp \left[ \frac{-1}{2\sigma^2} \left\{ h(\hat{\beta} - \beta)'(X'X)(\hat{\beta} - \beta) + (n - q)\hat{\sigma}^2 \right\} \right] G(y) \]  

(10)

where \( G(y) \) is an analytical function free of \( \beta \) and \( \sigma \); and

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

(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 \( \beta \) 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

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

(12)

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

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

(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 performance 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 Paretoian 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 $t = n, \ldots, T-1$. Then we compute their one-step-ahead forecasts by applying both LSE and MMLE to each portfolio for $t = n+1, \ldots, T$. 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 (-0.0007717) 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 behavioural 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

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

Note: * p < 5%, ** p < 1%
Table 2. One-step Ahead Forecast Bias and MSE of MML and LS for US Monthly Stock Returns

<table>
<thead>
<tr>
<th>Portfolio</th>
<th>LS Method</th>
<th>MML method</th>
<th>REF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>MSE</td>
<td>Bias</td>
</tr>
<tr>
<td>Petroleum</td>
<td>-1.614</td>
<td>1.1228</td>
<td>-1.588</td>
</tr>
<tr>
<td>Finance &amp; Real Estate</td>
<td>-0.296</td>
<td>0.3525</td>
<td>-0.275</td>
</tr>
<tr>
<td>Consumer Durables</td>
<td>-0.980</td>
<td>0.4021</td>
<td>-0.611</td>
</tr>
<tr>
<td>Basic Industries</td>
<td>-0.827</td>
<td>0.2482</td>
<td>-0.644</td>
</tr>
<tr>
<td>Food &amp; Tobacco</td>
<td>-0.313</td>
<td>0.5971</td>
<td>-0.121</td>
</tr>
<tr>
<td>Construction</td>
<td>-1.139</td>
<td>0.6678</td>
<td>-1.034</td>
</tr>
<tr>
<td>Capital Goods</td>
<td>-0.463</td>
<td>0.3329</td>
<td>-0.376</td>
</tr>
<tr>
<td>Transportation</td>
<td>-0.709</td>
<td>0.8608</td>
<td>-0.853</td>
</tr>
<tr>
<td>Utilities</td>
<td>-0.667</td>
<td>0.8865</td>
<td>-0.380</td>
</tr>
<tr>
<td>Textile &amp; Trade</td>
<td>-1.365</td>
<td>0.7954</td>
<td>-0.570</td>
</tr>
<tr>
<td>Services</td>
<td>-1.774</td>
<td>1.5085</td>
<td>-1.604</td>
</tr>
<tr>
<td>Recreation</td>
<td>-1.423</td>
<td>0.7940</td>
<td>-0.554</td>
</tr>
<tr>
<td>Average</td>
<td>-0.9641</td>
<td>0.7140</td>
<td>-0.7175</td>
</tr>
</tbody>
</table>

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 \( \frac{\partial \ln L^*}{\partial \beta_j} \) (j=1,2,...,q) and \( \frac{\partial \ln L^*}{\partial \sigma} \), and the fact that \( 1/n |\frac{\partial \ln L}{\partial \beta_j} - \frac{\partial \ln L^*}{\partial \beta_j} | \) and \( 1/n |\frac{\partial \ln L}{\partial \sigma} - \frac{\partial \ln L^*}{\partial \sigma} | \) 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

\[
E(a_i \ast b_{[1]}, b_{[2]}, ..., b_{[n]}) = 0, \quad \text{for all } i = 1, 2, ..., n
\]

and

\[
E(e \ast b_{[1]}, b_{[2]}, ..., b_{[n]}) = 0, \quad (15)
\]

where \( e = y - X\beta \). Thus,

\[
E(\hat{\beta}_n) = \beta + E[(X'WX)^{-1} XWE(e | b_{[1]}, b_{[2]}, ..., b_{[n]})] = \beta
\]

and

\[
E(X'WX)^{-1} X'a = E[(X'WX)^{-1} XE(a | b_{[1]}, b_{[2]}, ..., b_{[n]})] = \beta.
\]

Therefore, the MML estimator \( \hat{\beta} \) is unbiased for \( \beta \) as \( \sigma \) is known.

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

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

As
\[
\frac{\partial^3 L'(\beta, \sigma)}{\partial \beta_j \partial \beta_i} = -\frac{2p}{k\sigma^2} \sum_{i=1}^{n} b_{(i)} x_j x_{il},
\]
we have
\[
-E\left(\frac{\partial^3 L'(\beta, \sigma)}{\partial \beta \partial \beta \partial \beta}\right) = \frac{2p}{k\sigma^2} X' E(W) X = \frac{2p}{k\sigma^2} E(b_{(i)}) X' X
\]
where
\[
E(b_{(i)}) = \frac{1}{n} \sum_{i=1}^{n} b_{i} \rightarrow \frac{p-1/2}{p+1} \quad \text{as } n \rightarrow \infty.
\]
Similarly,
\[
\frac{\partial^3 L'(\beta, \sigma)}{\partial \beta_j \partial \sigma} = -\frac{2p}{k\sigma^2} \left( \sum_{i=1}^{n} \alpha_{(i)} x_j + 2 \sum_{i=1}^{n} x_j b_{(i)} z_i \right)
\]
and
\[
-E\left(\frac{\partial^3 L'(\beta, \sigma)}{\partial \beta \partial \sigma}\right) = \frac{2p}{k\sigma^2} \left[ X' E(a) + 2X' E(Wz) \right] = 0.
\]
This follows immediately from (15) and the fact that \(E(a_{(i)}) = (1/n) 3a_i = 0\). Moreover,
\[
\frac{\partial^3 L'(\beta, \sigma)}{\partial^2 \sigma} = \frac{n}{\sigma^3} - \frac{2p}{k\sigma^4} \sum_{i=1}^{n} (2\alpha_{(i)} z_i + 3b_{(i)} z_i^2)
\]
which gives (see, for example, Bian and Tiku, 1997):
\[
-\frac{1}{n} E\left(\frac{\partial^3 L'(\beta, \sigma)}{\partial^2 \sigma}\right) \rightarrow \frac{p-1/2}{p+1} \frac{2}{\sigma^3} \quad \text{as } n \rightarrow \infty.
\]
References


