Robust Estimation of Multiple Regression Model with asymmetric innovations and Its Applicability on Asset Pricing Model

Wing-Keung Wong
Department of Economics
National University of Singapore

Guorui Bian
Department of Statistics
East China Normal University, China

May 1, 2005

Abstract: In this paper, we first develop the modified maximum likelihood (MML) estimators for the multiple regression coefficients in linear model with the underlying distribution assumed to be symmetric, one of Student's t family. We obtain the closed form of the estimators and derive their asymptotic properties. In addition, we demonstrate that the MML estimators are more appropriate to estimate the parameters in the Capital Asset Pricing Model by comparing its performance with that of least squares estimators (LSE) on the monthly returns of US portfolios. Our empirical study reveals that the MML estimators are more efficient than the LSE in terms of relative efficiency of one-step-ahead forecast mean square error for small samples.

1.   Introduction

The estimation of coefficients in a simple linear model is one of the oldest and most important problems and has received tremendous attention in the literature in Statistics and Econometrics. Most of the work reported is, however, 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 coefficients in multiple regressive model when the underlying distribution is non-normal. Naturally, one would prefer closed form estimators which are fully efficient (or nearly so). Preferably, these estimators should also be robust to plausible deviations from an assumed model. That is exactly what has been achieved in the series of our papers including the present one. The underlying distribution is assumed to be symmetric and to be Student's t family for illustration. The method of modified maximum likelihood (MML) estimation (Tiku 1968; Tiku et al. 1999, 2000, 2001) is invoked.

This paper first extends the results given in Bian and Tiku (1997), Tiku et al. (1999, 2000, 2001) and Wong and Bian (2005). Tiku et al. (1999) develop the MML estimators for simple linear regression with symmetric innovation; Tiku et al. (2000) come up with the MML estimators for the first order autoregressive model with symmetric Innovation; Tiku et al. (2001) refine the MML estimator for the simple linear regression model with innovation from Student's t family while Bian and Tiku (1997) adopt the Bayesian approach to study a standard multiple regression model with identical and independent distributed (iid) error term. This paper extends their work to derive the MML estimators for the multiple regression model with the underlying distribution assumed to be symmetric, one of Student's t family. The likelihood equations have no explicit solutions and have to be solved by iterative method which is a formidable task. Thus, the maximum likelihood (ML) estimators are not readily available. Following Tiku et al. (1999, 2000, 2001), we derive the MML estimators. These estimators are explicit functions of sample observations and hence easy to compute. Moreover, they are essentially as efficient as the ML estimators (see for example, Tiku et al.,1999, 2000, 2001; and Wong and Bian 2005). We further derive the asymptotic properties for the MML estimators.

We note that the MML estimators have been extensively demonstrated by simulation study to be robust and remarkably efficient and clearly superior to the traditional normal-theory estimators in all the models being studied, including the autoregressive model (Tiku et al. 2000), simple linear regression model (Tiku et al. 2001) and simple linear regressive model with autoregressive innovation (Tiku et al. 1999; Wong and Bian 2005). As the multiple linear regression model is a simple extension of the above models, the properties of the robustness and efficiency for its estimators will be similar to that of the simple linear regression. As such, the resulting estimators in this paper are explicit functions of sample observations and are asymptotically fully efficient. Since they are almost fully efficient for small sample sizes and are remarkably robust, we skip reporting the simulation results in this paper.

We then study the applicability of the MML estimators to finance and economics by demonstrating that the MML estimators are more appropriate to estimate 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 concerned by both financial economists and econometricians. Fama (1963; 1965a, b) and many others analyze the empirical data. They conclude that the normality assumption in the distribution of a security or portfolio return is violated such that the distribution is 'flat-tailed' and suggest the family of stable Paretian distributions between normal and Cauchy distributions for the stock returns. On the other hand, Blattberg and Gonedes (1974) examine the return to security and suggested student-t as an alternative 'flat-tail' distribution for the return. Clark (1973), Kon (1984) and Tse (1991) suggest a mixture of normal distributions for the stock return. However, Fielitz and Rozelle (1983) suggest that a mixture of non-normal stable distributions would be a better representation of the distribution of the return.

Harvey and Zhou (1993) show that the distributional structure of the return 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 disturbance in the CAPM. As the MML estimators for the simple linear regression with t-distributed innovation 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 academic or practitioners to apply the MML estimators developed in our paper for the estimation of the parameters of the CAPM for the stock returns and we hypothesize that the MML estimators developed in our paper is more appropriate in the estimation of the CAPM in the sense that it is more efficient than the LSE.

To illustrate the superiority of the proposed MML estimators and to test the above hypothesis, we apply the one-step ahead forecasting technique to compare the MML estimators with the traditional least squares estimators, LSE, in the estimation of the parameters in the CAPM for the US monthly stock returns. The one-step ahead forecasting technique is commonly used to compare the performance of different models (Clements and Hendry 1997). Our empirical study reveals that the MML estimators are more efficient than the LSE in terms of the relative efficiency of one-step-ahead forecast mean square error for small samples. Hence, we recommend the MML estimators for the estimation of the CAPM.

This paper is organized as follows. We first derive the MML estimators in the next section and reveal the asymptotic properties of the MML estimators in Section 3. Section 4 reviews the theory of the standard CAPM and the 'flat-tail' distribution of the security return and demonstrates the superiority of the MML estimators in CAPM. Section 5 is the conclusion.

2.   Modified Maximum Likelihood Estimators

Consider the multiple regression model

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

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, $\beta=(\beta_1, ..., \beta_q)'$ is a qx1 vector of regression coefficients, and $\mathbf{e}$ is a nx1 vector of random errors $(e_i , e_i , ..., e_i)'$.

It is assumed that the innovations e_i are iid errors. The linear model (1) has many applications, for example, in the estimation of the CAPM as illustrated in this paper and in the prediction of the 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).

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

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

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 $T = \sqrt{v}e/\sigma\sqrt{k}$ has Student's 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 $N(0, \sigma^2)$.

The likelihood function for the model in (1) is

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

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

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

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 in terms of the function $g(z_i)$ and hence intractable. Solving them by iterative methods is a formidable task and can be very problematic especially for small values of p in which one may encounter multiple roots, slow convergence, or convergence to wrong values or even

---

(1) g(z) is the nonlinear part of the derivative of lnf(z), where f(z) is the standard distribution of the error term with f(z) = c(1+z²/k)⁻ᵖ.

4

divergence (Barnett 1966a; Lee et al. 1980; Tiku and Suresh 1992). See also Pearson and Hartley (1972, p89) who give examples where the iterations involved in determining ML estimates do not converge rapidly enough. 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 by now 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 ith observation corresponding to the order statistic $z_{([i])}$. Clearly,

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

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

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

where

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

Since g(z) is almost linear in any small interval (Tiku 1968; Tiku and Suresh 1992), under some very general regularity conditions, $z_{(i)}$ converges to $t_{(i)}$ as 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). Utilizing (6), the following modified likelihood equations are obtained:

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

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

---

2 $z_{(i)}$ and $t_{(i)}$ are the percentiles of the empirical distribution $F_n(x)$ and theoretical distribution F(x).

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

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

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

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

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

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

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

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

It is clear that all the above MML estimators have closed-formed algebraic expressions and are, therefore, easy to compute. From (8), the MML estimator β is found to consist of two components with the main component $\hat{\beta}_w$ being a weighted least squares estimator of β and unbiased for β. We remark that for $p = \infty$ (normally distributed errors), $a_i = 0$, $b_i = 1$ and $2p/k = 1$. Consequently, in this situation the MML estimators (8) are reduced to the usual LS (least squares) estimators. For computations, 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 estimates of β and σ from (8). Replacing the LS estimates by these MML estimates, we repeat the computation for a few more iterations till the estimates stabilize. In all our computations, some of which are presented in this paper, no more than three iterations were needed for the estimates to stabilize. Any further iteration hardly changes the values of the estimates and are, therefore, not necessary. This is shown in our extensive computations in the present paper and in the computations of our past papers.

6

3.   Asymptotic Properties of MML Estimators

The asymptotic properties of MML estimators can be 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 the covariance for β and σ are given by*

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

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

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

The proofs are in the Appendix.

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

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

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

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

Since 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 below:

Lemma 3.

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

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

(iii) the estimators $\hat{\beta}$ 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

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

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

4.   An Application in Finance

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

$$R_i = a_i + b_i R_m + e_i , \qquad (14)$$

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

We choose CAPM for the illustration of our MMLE approach as CAPM is one of the simplest models in Finance, yet complicated enough that the usual LSE cannot handle well. If MMLE outperforms LSE in this simple model, MMLE is expected to outperform all other more complicated models in Finance and Economics like Arbitrary Pricing Theory. Though the CAPM looks simple as shown in (14), it is complicated as the measure of beta is empirically

nonstationary over time (Leavy 1971; Blume 1975). Besides, the distributions of both the security or portfolio return and the disturbance are 'flat-tail' and hence violate the normality assumption (Fama 1963, 1965a,b; Pettit and Westerfield 1974).

To handle the non-stationarity of beta, one may estimate the model from a reasonably short period in order to capture the stationary Beta parameter (Wong and Bian 2000). To handle the 'flat-tail' distribution for both the return and the disturbance, Fama (1963; 1965a,b) suggest the family of stable Paretian distributions between normal and Cauchy distributions for the stock returns while Blattberg and Gonedes (1974) examine the security returns and suggest student-t as an alternative 'flat-tail' distribution. Clark (1973), Kon (1984), and Tse (1991) recommend a mixture of normal distributions for the stock return while Fielitz and Rozelle (1983) believe that a mixture of non-normal stable distributions would be a better representation of the distribution of security or portfolio return. In this paper, we will demonstrate that the MMLE with t-distributed innovations will be a good approach in handling the non-normality situation since MMLE has been extensively studied (see for example Tiku et al.1999, 2000, 2001) to be robust enough 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 easy 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 employed in the study. The industry classifications conform to Sharpe (1982), Breeden et al. (1989) and Gibbons et al. (1989). The portfolios are value-weighted and the market return is the weighted NYSE return and the monthly returns from the period 1926-1987 are in excess of 30-day Treasury-bill rate. The portfolios 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 the MML estimators are more appropriate in the estimation of the parameters of the CAPM because it is more efficient than the LSE. To illustrate, we use the MMLE model with Student's t distribution, plus 7 degrees of freedom to treat the CAPM for the US monthly stock returns as the MMLE estimators are one of the most

9

robust estimators (Tiku et al. 1999, 2000, 2001; Wong and Bian 2005). Twelve industrial portfolios of US data are employed in the study. We adopt the one-step-ahead forecast bias and MSE as the basis to evaluate the performance of LSE and MMLE over the CAPM. In the computation, we choose a small sample size n of 12 (i.e., a year period) in order to capture a stationary *b* parameter.

We first compute the skewness and kurtosis coefficients and 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, like the modified Shapiro-Wilk statistic, Anderson-Darling test and Kolmogorov-Smirnov test. However, as Jarque-Bera statistic is one of the best test statistics for normality and the results for other normality statistics are similar, we only report the results of Jarque-Bera statistic and its corresponding skewness and kurtosis coefficients in this paper. The 0.01 level of significance shown in the table lead us to reject the null hypothesis of normality for the monthly excess returns R 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 return departs from normal, which may be attributed to the ARCH or GARCH effects. However, temporal aggregation will reduce this ARCH or GARCH effects, for examples, see Drost and Nijman (1995).

[PLACE TABLE 1 ABOUT HERE]

Table 1. Tests for departure from normality for monthly excess portfolio returns and the corresponding residuals in CAPM by industrial classifications.

We adopt the one-step-ahead forecast MSE (see Clements and Hendry (1997) and Wong and Bian (2000) for more detail), as a basis for comparison between LSE and MMLE for the US monthly data. For the given sample size n=12, the estimates of both LSE and MMLE are first computed for each of the 12 industrial portfolios for t = n, ... , T-1. We then 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 the average one-step-ahead of forecast

MSE for both LSE and MMLE, is then computed and displaced in Table 2. We note that the values of the bias and MSE in the table are 1000 times the original values.

[PLACE TABLE 2 ABOUT HERE]

Table 2. The one-step ahead forecast bias and MSE of MML and LS approaches for US stock monthly returns (n=12)

From Table 2, we find from the tabulated values that MMLE has both smaller one-step-ahead forecast bias and smaller MSE and is more efficient than LSE in all industries except the Finance and Real Estate and Transportation. The average values depicted in the table also show that MMLE attains a smaller average one-step-ahead forecast bias (-0.0007175) and smaller one-step-ahead forecast MSE (0.00070874) than those of LSE (-0.0009641 and 0.0007140 respectively) with average relative efficiency to be 1.0082. This implies that the MML estimators are remarkably more efficient and robust than the LSE. Hence, MMLE is clearly superior to the traditional normal-theory estimators.

5.   **Concluding Remarks**

It is generally recognized that nonnormal samples occur very frequently in practice. In this paper, we extend 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 innovation to be asymmetric and from a Student-t family. The likelihood equations are intractable. Solving them by the iterative methods is tedious and time consuming and the results obtained might even be unreliable. We, therefore, use the methodology of 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 the 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 have 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 have demonstrated their very high efficiencies not shared by the Gaussian estimators. In fact, we do not recommend the use of the Gaussian estimators for nonnormal innovations.

For further study, one may consider incorporating the Bayesian approach (see Matsumura et al. 1990; Bian and Tiku 1997) into the MMLE estimation. We note that the distribution of the stock returns in our illustration is not only heavy-tailed but also strongly skewed (refer to Table 1). Hence, it is possible to improve the forecasting by using skewed error distributions. The MMLE model with asymmetric innovations will be an interesting issue for the extension in this situation. Further extension includes studying the applicability of the MMLE linear model to other prominent Economics or Finance models in, for example, Wong and Chan (2004) and Fong et al. (2005). Another possible area for further research is to compare the beta in this study with the equity cost of capital for each portfolio (Thompson and Wong 1991, 1996).

There are many other approaches in the study of linear models, for example, no distributional assumptions on the measurement errors (Wong and Miller 1990; Li 2002; Li and Hsiao 2004), the nonlinear regression models (Amemiya, 1985; Hsiao, 1989; Hausman et al., 1998, Honore and Hu, 2004), multinomial models (Hsiao and Sun 1999), and count models (Li et al. 2003). Nevertheless, it is well-known that if the distribution of the disturbances is known to be from Student's t family, parametric approaches like ours will yield estimators which outperform the estimators without distributional assumptions (Li and Hsiao 2004). As such, our approach performs better when the distribution is known. In additional, our approach could be incorporated to improve the estimation in other models, like the nonlinear regression models, multinomial models and count models.

Table 1. Tests for departure from normality for monthly excess portfolio returns and the corresponding residuals in CAPM by industrial classifications.

Portfolio	Returns			Residuals		
	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**

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

Table 2. The one-step ahead forecast bias and MSE of MML and LS approaches for US stock monthly 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 values of 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 Student's t distribution, it immediately follows that

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

and

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

where $\mathbf{e}$ = Y - Xβ. 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 variance-covariance matrix of the MML estimators is given by the inverse of

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

Since

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

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 $E(a_{[i]}) = (1/n)\,3a_i = 0$. Also

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

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

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

References

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

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 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 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.

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.

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 American Statistical Association 78, 28-36.

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.

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.

Hsiao, C., Sun, B., 1999. Modeling survey response bias—with an analysis of the demand for an advanced electronic device. Journal of Econometrics 89, 15–39.

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

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.

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.

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.

Li, T., 2002. Robust and consistent estimation of nonlinear errors-in-variables models. Journal of Econometrics 110, 1–26.

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

Li, T., Trivedi, P.K., Guo, J., 2003. Modeling response bias in count: a structural approach with an application to the national crime victimization survey data. Sociological Methods and Research 31, 514–544.

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.

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.

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.

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 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., (ed). 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., 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., Miller, R.B., 1990. Analysis of ARIMA-Noise Models with Repeated Time Series. Journal of Business and Economic Statistics 8(2), 243-250.