King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 404

Apr 2009

ESTIMATION OF KURTOSIS FOR MULTIVARIATE DATA

S. E. AHMED, M. H. OMAR, AND A. H. JOARDER

DHAHRAN 31261 ● SAUDI ARABIA ● www.kfupm.edu.sa/math/ ● E-mail: c-math@kfupm.edu.sa

ESTIMATION OF KURTOSIS FOR MULTIVARIATE DATA

S. E. AHMED, M. H. OMAR, AND A. H. JOARDER

Abstract. In this paper, we suggest some improved estimation strategies for the kurtosis parameter based on shrinkage and pretest methodologies in the presence of non-sample information (NSI) regarding the kurtosis parameter. Indeed, the shrinkage and pretest methods use the NSI in some optimal sense. In practice, NSI is readily available in the form of a realistic conjecture based on the experimenter's knowledge and experience with model and data. It is advantageous to utilize NSI in the estimation process to construct improved estimation for the kurtosis parameter. A large sample theory of the suggested estimators are developed where the properties of these estimators are examined both analytically. Further research directions are also discussed.

1. Introduction and Preliminaries

Skewness and kurtosis have been used in tests of normality, robustness, outliers, modified tests and estimation, large sample inferences, and other situations. The kurtosis parameter is embedded in many inference problems. For example, the asymptotic variance of process capability indices, coefficient of variation and effect size index depend on kurtosis parameter, among other parameters. In its own right, kurtosis measure the "peakedness" of a distribution. Generally speaking, the kurtosis parameter and its estimation is not "stable", specially in the presence of outliers. In this communication, our parameter of interest is kurtosis and we consider some improved alternative estimation strategies. Our objective is to combine sample and non-sample information in the estimation process for the kurtosis parameter of a multivariate normal distribution. The kurtosis parameter estimation is embedded in many statistical estimation problems and applications, see Douglas (2006) and An and Ahmed (2008). The asymptotic variance of many important indices are a function of kurtosis parameter and hence an accurate and precise estimation of kurtosis is needed. Kim and White (2004) argued that the role of higher moments has become increasingly important in the literature mainly because the traditional measure of risk, variance (or standard deviation), has failed to capture fully the "true risk" of the distribution of stock market returns; see also Harvey and Siddique (2000).

This research is motivated by diverse applications and involvement. Some motivating examples are given below.

Date: April 10, 2009.

2000 *Mathematics Subject Classification.* Primary 62E20,65C05; Secondary 62F12, 62J07, 65C60,62F03,62F05.

Key words and phrases. Parameter stability, shrinkage and pretest estimation, local alternatives, asymptotic properties, relative precision, Monte-Carlo simulation.

This paper is in final form and no version of it will be submitted for publication elsewhere.

Let X be a p-dimensional random variable with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then the kurtosis parameter is defined as

$$\beta = E[\{(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\}^2].$$

For multivariate normal distribution, we have $\beta = p(p+2)$, and for the univariate normal distribution, since $p = 1$, the kurtosis parameter will have a value 3. For the bivariate case, with $p = 2$, β has a simplified version in terms of centered product moments (Joarder and Abujiya, 2008).

The estimate of the kurtosis measure based on a sample $(\mathbf{X}_1, , \ldots , \mathbf{X}_n)$ is

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\{(\mathbf{X}_i - \bar{\mathbf{X}})'\mathbf{S}^{-1}(\mathbf{X}_i - \bar{\mathbf{X}})\}^2$$

where $\bar{\mathbf{X}}$ and $\mathbf{S}$ are sample mean and sample covariance matrix given by

$$\bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X_i} \text{ and } \mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X_i} - \bar{\mathbf{X}})(\mathbf{X_i} - \bar{\mathbf{X}})'$$

respectively. Further, the asymptotic normality of $\hat{\beta}$ is presented in the following lemma.

Lemma 1. *(Mardia,1970). Let the p-dimensional random vector* $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *then as* $n \to \infty$

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow[D]{} \mathcal{N}(0, 8\beta),$$

where the notation $\xrightarrow[D]{}$ *means convergence in distribution and* $\mathcal{N}_p$ *is the multivariate normal density function.*

Now, we state another related definition (Srivastava, 1984). Let $\lambda_1, \cdots, \lambda_p$ be the eigenvalues of $\boldsymbol{\Sigma}$ and let $\gamma_1, \cdots, \gamma_p$ be a columns of a matrix $\boldsymbol{\Gamma}$ such that $\boldsymbol{\Gamma}'\boldsymbol{\Sigma}\boldsymbol{\Gamma}$ $= diag(\lambda_1, \cdots, \lambda_p)$. Let $Y_i = \gamma_i'\mathbf{X}$ and $\theta_i = \gamma_i'\boldsymbol{\mu}$ for $i = 1, \cdots p$. Then, the kurtosis parameter is defined as

$$\beta_* = \frac{1}{p}\sum_{i=i}^{p}\frac{E[(Y_i - \theta_i)^4]}{\lambda_i^2}.$$

Let $\hat{\beta}_*$ be the corresponding sample measurements, then under multivariate normality $\frac{np}{24}(\hat{\beta}_* - 3)$ is asymptotically distributed as standard normal.

However, both estimators are very sensitive to outliers or unusual observation. A few contaminated observations may have diverse effect on its sample estimates. For this reason we try to stabilize the kurtosis parameter estimation by incorporating the available parameter information in the estimation process. In this article, although we focus on the estimation of β, the obtained results can also be implemented for the estimation of β_*.

A plan of this paper is as follows. The improved estimation procedures based on shrinkage and the preliminary test method are considered in Section 2 along with some asymptotic results. In Section 3, we compare our estimators with the sample estimate and show that our methods are asymptotically superior to sample estimate when the non-sample information (*NSI)* regarding the kurtosis parameter is credible or even nearly credible. The results of the simulation experiment are given in Section 4 to illustrate our methods. Further, examples are given in section 5. Finally, we provide concluding remarks in Section 6.

2. Improved Estimation Strategies

Our main focus here is to improve the estimation of β when it is assumed that the data come from a multivariate normal distribution. However, the data may be contaminated by a few observation, which will have a very negative impact on the sample estimate, $\overset{\wedge}{\beta}$. Hence, in an effort to stabilize the parameter estimation of β, we investigate alternative strategies for parameter estimation. In a number of real problems, the practitioner has available both an approximation of β that provides a constant β_o and a sample based estimator that provides a point estimator $\hat{\beta}$. The quality of β_o is unknown; the analyst, however, respect its ability to approximate β. Our problem is to is to combine the approximation or opinion on β_o and the sample result $\hat{\beta}$. Consequentially, we consider estimators based on shrinkage and pretest estimation.

For both shrinkage and preliminary test estimators we discuss performance as measured by mean squared error (MSE), the sum of squared bias and variance. The approximation β_o is deterministic, which can be interpreted as zero variance; its error, $\beta - \beta_o$, can be interpreted as bias. Suppose the analyst wishes to report the point estimator defined by the linear combination

$$\overset{\wedge}{\beta}^S = c\beta_o + (1-c)\overset{\wedge}{\beta}, \tag{2.1}$$

in which we would choose, in ideal circumstances, the coefficient c so as to minimize the mean squared error (*MSE*). Further, c may also be defined as the degree of confidence in the prior information β_o. The value of $c \in [0, 1]$ may be assigned by the experimenter according to confidence in the prior value of β_o. If $c = 0$, then we use the sample data only. We may choose an estimator of optimal c that minimizes the variance. However, the optimal value of c depends on the unknown parameter θ and thus it is not accessible (oracle). Estimators constructed as linear (or, more precisely, convex) combinations of other estimators or guessed values as in (2.1), are called *composite* estimators. The composite estimator $\overset{\wedge}{\beta}^S$ can be interpreted as shrinkage estimator (*SE*), as it shrinks the sample estimator $\overset{\wedge}{\beta}$ towards β_o. Ledoit and Wolf (2003) applied this strategy to estimate the covariance matrix. They suggested to shrink the MLE of the covariance matrix towards structured covariance matrices that can have relatively small estimation error in comparison with the MLE.

Ahmed and Krzanowski (2004) and others pointed out that such an estimator yields smaller mean squared error (*MSE*) when a priori information β_o is correct or nearly correct, however at the expense of poorer performance in the rest of the parameter space induced by the prior information. We will demonstrate that $\overset{\wedge}{\beta}^S$ will have a smaller *MSE* than $\overset{\wedge}{\beta}$ near the restriction, that is, β_o. However, $\overset{\wedge}{\beta}^S$ becomes considerably biased and inefficient when the restriction may not be judiciously justified. Thus, the performance of this shrinkage procedure depends upon the correctness of the uncertain prior information. As such, when the prior information is rather not trustworthy, it may be desirable to formulate a shrinkage pretest estimator (*SPE)* denoted by $\hat{\beta}^{SP}$ which incorporates a pretest on β_o. Thus, the estimator yields either $\hat{\beta}$ or $\hat{\beta}^S$ depending upon the outcome of the pretest. If the prior information is tenable, one may use $\hat{\beta}^S$, while $\hat{\beta}$ may be chosen otherwise.

Thus, we consider the shrinkage pretest estimator which is defined as

$$\hat{\beta}^{SP} = \hat{\beta} I(\mathcal{L}_n \geq c_\alpha) + [(1-c)\hat{\beta} + c\beta_o] I(\mathcal{L}_n < c_\alpha), \tag{2.2}$$

where $I(A)$ is the indicator function of a set A and $\mathcal{L}_n$ is the test statistic for the null hypothesis H_o: $\beta = \beta_o$ defined below. Based on the result of lemma 1, we consider the following test statistics for H_o: $\beta = \beta_o$ against H_a: $\beta \neq \beta_o$ (or $\beta < \beta_o$ or $\beta > \beta_o$). A natural choice of β_o will be $\beta_o = p(p+2)$. In other words, we are testing the normality of the parent population using the kurtosis measure. However, the purpose here is to improve the estimation of the kurtosis parameter β. Hence, the statistics is given by

$$\mathcal{L}_n = \frac{\{\sqrt{n}(\hat{\beta} - \beta_o)\}^2}{8p(p+2)}.$$

For large $n(\geq 50)$ and under the null hypothesis, the test statistics $\mathcal{L}_n$ follows a χ^2-distribution with one degree of freedom, which provides the asymptotic critical values. It is important to note that for a fixed alternative that is different from the null hypothesis, the power of the test statistics will converge to one as $n \to \infty$. Hence, to explore the asymptotic power properties of $\mathcal{L}_n$, we confine ourselves to a sequence of local alternatives $\{K_n\}$. In the present work, such a sequence is specified by

$$K_n : \beta_n = \beta_o + \frac{\delta}{\sqrt{n}}, \tag{2.3}$$

where δ is a fixed real number. Stochastic convergence of $\hat{\beta}$ to the parameter β ensures that $\hat{\beta} \xrightarrow[p]{} \beta$ under local alternatives as well, where the notation $\xrightarrow[p]{}$ means *convergence in probability.*

The following theorem, which we present without proof, characterizes the asymptotic powers of the three test statistics under local alternatives.

Theorem 1. *Under local alternatives in (2.3) the following results hold:*

1. $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow[D]{} \mathcal{N}(\delta, 8p(p+2))$,

2. $\mathcal{L}_n$ *has asymptotically a noncentral* χ^2*-distribution with 1 degree of freedom and non-centrality parameter* $\Delta = \dfrac{\delta^2}{(8p(p+2))^2}$.

Hence, the power calculations of the proposed test statistic can be accomplished by using noncentral χ^2-distribution. Thus, the critical value c_α of $\mathcal{L}_n$ may be approximated by $\chi^2_{1,\alpha}$, the upper $100\alpha\%$ critical value of the χ^2-distribution with 1 degree of freedom.

Further, shrinkage pretest estimator (SPE) can be written in a more computationally attractive form as follows:

$$\hat{\beta}^{SP} = \hat{\beta} - c(\hat{\beta} - \beta_o)I(\mathcal{L}_n < c_\alpha). \tag{2.4}$$

Thus, the classical pretest estimator (PE) is readily obtained, by substituting $c = 1$ in above relation,

$$\hat{\beta}^{P} = \hat{\beta} - (\hat{\beta} - \beta_o)I(\mathcal{L}_n < c_\alpha). \tag{2.5}$$

The above PE is due to Bancroft (1944). The proposed SPE (Ahmed, 1992) may be viewed as an improved PE which represents both $\hat{\beta}$ and PE for $c = 0$ and $c = 1$ respectively. In the literature, a discussion about pretesting can be found in Giles and Giles (1993), Magnus (1999), Ohanti (1999), Reif and Vlček (2002), Khan and Ahmed (2003), among many others.

3. Asymptotic Bias and Mean Squared Error

We will assess the performance of all these listed estimator using the mean squared error (MSE) criterion. The MSE of an estimator $\tilde{\beta}$ aimed at the target β is defined as

$$MSE(\tilde{\beta}; \beta) = E\{(\tilde{\beta} - \beta)^2\},$$

where the notation E is the expectation operator with reference to hypothetical replications of the sampling process. The bias of an estimator $\tilde{\beta}$ of β is denoted by $B(\tilde{\beta}; \beta)$, so

$$MSE(\tilde{\beta}; \beta) = var(\tilde{\beta}) + B(\tilde{\beta}; \beta)^2.$$

We regard MSE as a measure of efficiency. However, MSE is usually not known and its value may depend on one or several parameters, sometimes on the target parameter itself. In addition, an estimator may be efficient for some values of the parameters but not for others. All this makes the search for the most efficient estimator a challenging problem. To meet some of these challenges, we will express the MSE for the listed estimators as a function of non-centrality parameter Δ for a smooth reading of the findings of the article.

Further, note that our results are based on the asymptotic normality of $\hat{\beta}$, so our results will be of asymptotic nature. The asymptotic bias of an estimator $\tilde{\beta}$ of β is defined as

$$AB(\tilde{\beta};\beta) = \lim_{n\to\infty} E\{\sqrt{n}(\tilde{\beta}-\beta)\} \tag{3.1}$$

Under the local alternatives $AB(\hat{\beta}^S) = -c\,\delta$, an unbounded function of δ. The expression of $AB(\hat{\beta}^{SP})$ is obtained with the aid of the following lemma.

Lemma 2. *Let $Z \sim \mathcal{N}(\mu, 1)$. Then*

$$E\{ZI(0 < Z^2 < x)\} = \mu P(\chi^2_{3,\frac{\mu^2}{2}} < x)$$

where $\chi^2_{3,\frac{\mu^2}{2}}$ is distributed as a chi-square with 3 degrees of freedom and non-centrality parameter $\frac{\mu^2}{2}$.

For proof of the lemma, readers are referred to Judge and Bock (1978).

Using Lemmas 1 and 2, the following relation is established.

$$AB(\hat{\beta}^{SP};\beta) = -c\delta G_3(\chi^2_{1,\alpha};\Delta),$$

where $G_q(.;\Delta)$ is the cumulative distribution of a noncentral χ^2-distribution with q degrees of freedom and non-centrality parameter Δ. Since $lim_{\delta\to\infty}\delta G_3(\chi^2_{1,\alpha};\Delta) = 0$, we safely conclude that $\hat{\beta}^{SP}$ is asymptotically unbiased, with respect to δ. For $c = 1$, $AB(\hat{\beta}^{P};\beta) = -\delta G_3(\chi^2_{1,\alpha};\Delta)$. The $AB(\hat{\beta}^{SP};\beta)$ and $AB(\hat{\beta}^{P};\beta)$ are 0 at $\Delta = 0$. The bias functions of both pretest estimators increases to maximum as *delta* increases, then decreases towards 0 as Δ further increases. Further, it is seen from the AMSE expression that the larger the value of c, the greater is the variation in the bias values.

Under the local alternatives in (2.3) we present the expressions for the $AMSE$ for the estimators under consideration.

$$AMSE(\hat{\beta}^S;\beta) = AMSE(\hat{\beta};\beta) - AMSE(\hat{\beta};\beta)c(2-c) + AMSE(\hat{\beta};\beta)c^2\Delta,$$

where $AMSE(\hat{\beta};\beta) = 8p(p+2)$.

$$\begin{aligned} AMSE(\hat{\beta}^{SP};\beta) &= AMSE(\hat{\beta};\beta) - AMSE(\hat{\beta};\beta)c(2-c)G_3(\chi^2_{1,\alpha};\Delta) \\ &\quad + AMSE(\hat{\beta};\beta)c^2\Delta\{2G_3(\chi^2_{1,\alpha};\Delta) - (2-c)G_5(\chi^2_{1,\alpha};\Delta)\}. \end{aligned}$$

The expression of $AMSE(\hat{\beta}^{SP};\beta)$ is readily obtained with the use of the following lemma.

Lemma 3. *Let $Z \sim \mathcal{N}(\mu, 1)$. Then*

$$E\{Z^2I(0 < Z^2 < x)\} = P\left(\chi^2_{3,\frac{\mu^2}{2}} < x\right) + \mu^2 P\left(\chi^2_{5,\frac{\mu^2}{2}} < x\right)$$

The proof of this lemma can be found in Judge and Bock (1978).

The $AMSE(\hat{\beta}^S;\beta)$ is a straight line function of Δ, which intersects the $AMSE(\hat{\beta})$ at $\Delta = (2-c)/c$. Under the null hypothesis, the $AMSE$ of $\hat{\beta}^S$ is less than the $AMSE$ of $\hat{\beta}$. Specifically,

$$AMSE(\hat{\beta}^S;\beta) \leq AMSE(\hat{\beta};\beta) \text{ whenever } \Delta \in [0,(2-c)/c].$$

On the other hand, $AMSE(\hat{\beta}^{SP};\beta) \geq AMSE(\hat{\beta};\beta)$ if

$$\Delta \geq (2-c)G_3(\chi^2_{1,\alpha};\Delta)\{2G_3(\chi^2_{1,\alpha};\Delta) - (2-c)G_5(\chi^2_{1,\alpha};\Delta)\}^{-1}. \tag{3.2}$$

Alternatively, $\hat{\beta}^{SP}$ performs better than $\hat{\beta}$ if

$$\Delta < (2-c)G_3(\chi^2_{1,\alpha};\Delta)\{2G_3(\chi^2_{1,\alpha};\Delta) - (2-c)G_5(\chi^2_{1,\alpha};\Delta)\}^{-1}.$$

The $AMSE$ of pretest estimators are a function of α, the level of the statistical significance. As α approaches one, $AMSE(\hat{\beta}^{SP};\beta)$ tends to $AMSE(\hat{\beta};\beta)$. Also, when Δ increases and tends to infinity, the $AMSE(\hat{\beta}^{SP};\beta)$ approaches the $AMSE(\hat{\beta};\beta)$. Indeed, for larger values of Δ, the value of the $AMSE(\hat{\beta}^{SP};\beta)$ increases, reaches its maximum after crossing the $AMSE(\hat{\beta};\beta)$ and then monotonically decreases and approaches the $AMSE(\hat{\beta};\beta)$. It appears from the AMSE expression that the smaller the value of α, the greater is the variation in the maximum and minimum of $AMSE(\hat{\beta}^{SP};\beta)$.

For $c = 1$ we get the $AMSE$ of $\hat{\beta}^P$ as follows

$$\begin{aligned} AMSE(\hat{\beta}^P;\beta) =& 8p(p+2) + 8p(p+2)\Delta\{2G_3(\chi^2_{1,\alpha};\Delta) - G_5(\chi^2_{1,\alpha};\Delta)\} \\ &- 8p(p+2)\Delta G_3(\chi^2_{1,\alpha};\Delta) \end{aligned}$$

and $AMSE(\hat{\beta}^P;\beta) \geq AMSE(\hat{\beta};\beta)$. Accordingly,

$$\Delta \geq G_3(\chi^2_{1,\alpha};\Delta)\{2G_3(\chi^2_{1,\alpha};\Delta) - G_5(\chi^2_{1,\alpha};\Delta)\}^{-1}. \tag{3.3}$$

Thus, we notice that the range of the parameter space in (2.4) is smaller than that in (3.2).

The risk difference

$$\begin{aligned} AMSE(\hat{\beta}^P;\beta) - AMSE(\hat{\beta}^{SP};\beta) =& \Delta 8p(p+2)\{2(1-c)G_3(\chi^2_{1,\alpha};\Delta) \\ & - (1-c)^2 G_5(\chi^2_{1,\alpha};\Delta)\} \\ & - 8p(p+2)(1-c)^2 G_3(\chi^2_{1,\alpha};\Delta), \end{aligned}$$

suggests that $AMSE(\hat{\beta}^P;\beta) \leq AMSE(\hat{\beta}^{SP};\beta)$ since

$$\Delta \leq (1-c)G_3(\chi^2_{1,\alpha};\Delta)\{2G_3(\chi^2_{1,\alpha};\Delta) - (1-c)G_5(\chi^2_{1,\alpha};\Delta)\}^{-1}.$$

Thus, $\hat{\beta}^{SP}$ outshines $\hat{\beta}^P$ when

$$\Delta > (1-c)G_3(\chi^2_{1,\alpha};\Delta)\{2G_3(\chi^2_{1,\alpha};\Delta) - (1-c)G_5(\chi^2_{1,\alpha};\Delta)\}^{-1}.$$

Hence, it can be safely concluded that none of the estimators perform better than the other three. However, at Δ=0, the shrinkage estimator will be the best choice. Also, both pretest estimators have smaller AMSE than that of $\hat{\beta}$ when the null hypothesis is tenable.

4. Direction for Further Research

To examine the behavior of the relative precisions of $\hat{\beta}^S, \hat{\beta}^P$, and $\hat{\beta}^{SP}$ to $\hat{\beta}$, we can consider $H_o : \beta = \beta_o$ against $H_a : \beta = \beta_o + \delta$ where δ is a shift real number in the neighborhood domain of β from various data distributions. Using Monte Carlo simulations, the various kurtosis estimators discussed earlier can be calculated . The performance of these kurtosis parameters can be studied by comparing their simulated relative precisions (SRP) for various values of δ, where generally for an estimator $\tilde{\beta}$

$$SRP(\hat{\beta};\tilde{\beta}) = \frac{SMSE(\hat{\beta})}{SMSE(\tilde{\beta})}$$

where $SMSE(\tilde{\beta})$ and $SMSE(\hat{\beta})$ are the empirical mean square errors of $\tilde{\beta}$ and $\hat{\beta}$ respectively.

5. Acknowledgements

The research work of Professor Ahmed is supported by a grant from the Natural Sciences and Engineering Council of Canada and a part of this investigation has been conducted while he was visiting the King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, KSA. Professors Joarder and Omar would also like to express gratitude to KFUPM for providing facilities for this research.

6. References

Ahmed, S.E. (1992). Shrinkage preliminary test estimation in multivariate normal distributions. Journal of Statistical Computation and Simulation, vol. 43, 177-195.

Ahmed, S.E. and Krzanowski, W.J. (2004): Biased estimation in a simple multivariate regression model, Computational Statistics and Data Analysis, 45, pp 689-696.

An, L. and Ahmed, S. E. (2008). Improving the performance of kurtosis estimator. *Comput. Statist. Data Anal.*, 52, 2669 – 2681.

Douglas, G.B., 2006. Confidence interval for a coefficient of quartile variation. *Comput. Statist. Data Anal.* 50, 2953 - 2957.

Giles, Judith A.; Giles, David E.A. (1993). Preliminary-test estimation of the regression scale parameter when the loss function is asymmetric. *Commun. Stat., Theory Methods*, 22, no.6, 1709-1733 .

Harvey, C.R., Siddique, A., (2000). Conditional skewness in asset pricing tests. *Journal of Finance* 55, 1263–1295.

Joarder, A.H., Abujiya, M.R. (2008). Standardized Moments of Bivariate Chi-square Distribution. *Journal of Applied Statistical Science*, 4, 1-9.

Judge, G.G., Bock, M.E. (1978). *The Statistical Implications of Pre-test and Stein-rule Estimators in Econometrics,* Amsterdam: North-Holland Publishing Co.

Khan, B.U.; Ahmed, S.E. (2003). Improved estimation of coefficient vector in a regression model. *Commun. Stat., Simulation Comput.,* 32, no. 3, 747-769.

Kim, T-H and White, H. (2004). On more robust estimation of skewness and kurtosis. *Finance Research Letters*, 1, 56–73.

Ledoit, P. and Wolf, M. (2003). Improve estimation of of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*,10, 603-621.

Magnus, J. R. (1999).The traditional pretest estimator. *Theory Probab. Appl.,* 44, No. 2, 293-308

Mardia, K.V.(1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika,* 57, 519-530.

Ohanti (1999)

Perez-Quiros, G., Timmermann, A., (2001). Business cycle asymmetries in stock returns: Evidence from higher order moments and conditional densities. *Journal of Econometrics* 103, 259–306.

Reif, J., Vlček, K. (2002). Optimal pre-test estimators in regression. *J. Econometrics,* 110, 91-102.

Timmermann, A., (2000). Moments of Markov switching models. *Journal of Econometrics,* 96, 75–111.

Srivastava, M.S. (1984). A measure of skewness and kurtosis and a graphical method of assessing multivariate normality. *Statistics & Probability Letters*, 5, 15-18.

Dept of Mathematics & Statistics, University of Windsor, Canada.
E-mail address: seahmed@uwindsor.edu.ca

Current address: Dept of Mathematics and Statistics, King Fahd University of Petroleum & Minerals, Saudi Arabia
E-mail address: omarmh@kfupm.edu.sa
URL: http://faculty.kfupm.edu.sa/omarmh.htm

Dept of Mathematics and Statistics
Current address: King Fahd University of Petroleum & Minerals, Saudi Arabia
E-mail address: anwarj@kfupm.edu.sa
URL: http://faculty.kfupm.edu.sa/anwarj.htm