Serdica J. Computing **8** (2014), No 3, 255–268

Serdica
Journal of Computing

Bulgarian Academy of Sciences
Institute of Mathematics and Informatics

ON THE L_p-NORM REGRESSION MODELS FOR ESTIMATING VALUE-AT-RISK*

Pranesh Kumar, Faramarz Kashanchi

Abstract. Analysis of risk measures associated with price series data movements and its predictions are of strategic importance in the financial markets as well as to policy makers in particular for short- and long-term planning for setting up economic growth targets. For example, oil-price risk-management focuses primarily on when and how an organization can best prevent the costly exposure to price risk. Value-at-Risk (VaR) is the commonly practised instrument to measure risk and is evaluated by analysing the negative/positive tail of the probability distributions of the returns (profit or loss). In modelling applications, least-squares estimation (LSE)-based linear regression models are often employed for modeling and analyzing correlated data. These linear models are optimal and perform relatively well under conditions such as errors following normal or approximately normal distributions, being free of large size outliers and satisfying the Gauss-Markov assumptions. However, often in practical situations, the LSE-based linear regression models fail to provide optimal results, for instance, in non-Gaussian situations especially when the errors follow distributions with fat tails and error terms possess a finite variance. This is the

ACM Computing Classification System (1998): B.1.2, F.1.3, F.2.3, G.3, J.2.

Key words: Value-at-Risk (VaR), Quantile distributions, Least-squares estimation, L_p-norms.

*This study was supported by Dr. Pranesh Kumar's NSERC Discovery Grant.

situation in case of risk analysis which involves analyzing tail distributions. Thus, applications of the LSE-based regression models may be questioned for appropriateness and may have limited applicability. We have carried out the risk analysis of Iranian crude oil price data based on the L_p-norm regression models and have noted that the LSE-based models do not always perform the best. We discuss results from the L_1, L_2 and L_∞-norm based linear regression models.

1. Introduction. The most popular measure of risk in financial markets is volatility, which is calculated in terms of the probability. Risk measurement for regulating the financial markets has been a key issue since the systematic developments in present financial history. Portfolio theory ([7], [8], [9]) placed financial risk measurement in practice which assumes that all risk could be partitioned into systematic, market risk and the residual, company-specific risk. The Capital Asset Pricing (CAP) model which emerged practically as an easy risk measure model hypothesized that since market risk is relevant for securities pricing, only the market risk measurement β (beta) is necessary. The CAP model provided a readily measurable risk estimate that could be applied practically in real time market conditions. However, the problem was that β metrics have only a tenuous connection to the actual security returns, thus raising questions on the appropriateness of βs as the true risk measure ([2], [7], [14]).

With doubts on the appropriateness of βs as the correct risk measure, practitioners searched for alternative risk measures which were both accurate and relatively inexpensive to estimate. While many other risk measures and models have been considered, Value-at-Risk (VaR) has been widely adopted. Since accurate risk measurement is essential to the financial institutions for proper risk management, many such internal models were developed in-house by financial institutions.

Value-at-Risk is a statistical measure of possible portfolio losses and is a measure of losses due to normal market movements. Losses greater than the value at risk occur with only a specified small probability. Subject to the simplifying assumptions, value at risk aggregates all risks in a portfolio into a single number suitable for practical uses like reporting to regulators or disclosure in annual reports. The concept of value-at-Risk is to understand and describe the magnitude of the likely losses on the portfolio.

The VaR model is complementary to many other internal risk measures such as RAROC developed by Bankers Trust in the 1970s [18]. However, market forces during the late 1990s led to the evolution of VaR as a dominant risk measurement tool for financial firms. The primary reason for the widespread adoption of VaR was JP Morgan's decision to develop an open architecture methodology known as RiskMetrics^{TM} ([15], [19]). RiskMetrics^{TM} was supported by a publicly available database containing the critical inputs required to estimate the model. Another reason for the adoption and popularity of VaR was the introduction in 1998, by the Bank for International Settlements (BIS), of international bank capital requirements that allowed relatively sophisticated banks to calculate their capital requirements based on their own internal modes such as VaR. RiskMetrics^{TM} quickly became the industry benchmark in risk measurement. Bank regulators worldwide allowed commercial banks to measure their market risk exposures using internal models that were often VaR-based. The market risk amendments to the Basel accord made in-house risk measurement models a mainstay in the financial sector ([20], [21]).

In data analysis, the least-squares estimation (LSE)-based linear regression models are often employed especially for modeling and analyzing time-series and cross-sectional data. These linear models are optimal and perform relatively well under conditions such as errors following normal or approximately normal distributions, being free of large size outliers and satisfying the Gauss-Markov assumptions. However, often in practical situations, the LSE-based linear regression models fail to provide optimal results, for instance, in non-Gaussian situations especially when the errors follow distributions with fat tails and error terms possess a finite variance. This is the situation in case of risk analysis which involves analyzing tail distributions. Thus, applications of the LSE-based regression models may be questioned for appropriateness and may have limited applicability.

In section 2 we describe the Value-at-Risk as a measure of risk. In Section 3, we define L_p-norm linear regression models. In section 4, measures for checking the adequacy of the model are given. The exploratory data analysis of Iran's crude oil prices and export is described in section 5. In section 6, we present the L_p-norm based results of the risk analysis from the data on Iran's crude oil prices. We offer concluding remarks in Section 7.

2. Value-at-Risk (VaR). Value-at-Risk (VaR) is defined as a statistical risk measure to answer a basic question in financial markets: "How much can we lose with α% probability on a trading portfolio over a pre-set horizon?" VaR measures the worst expected loss under normal market conditions over a specific time period at a given confidence level (quantile). In statistical terms, a bad day is defined so that there is only an α% probability that daily losses will exceed this amount for a given distribution of all possible daily returns over some recent past period. Thus, a bad day is defined so that there is only an α% probability of an even worse day. The 1% VaR denoted as VaR1% indicates a daily loss that will be equalled or exceeded only 1 percent of the time. Alternatively, VaR1% means that there is a 99% chance that tomorrow's daily portfolio value will exceed today's value less VaR1% [19].

Setting X as the portfolio value over the d-period horizon and $f(x)$ its probability density function, then for a confidence level $\alpha\%$,

$$\alpha = \Pr\left(Value > 100 - VaR\right) = \int\limits_{100-VaR}^{\infty} f\left(x\right) dx. \tag{2.1}$$

or, if $g(x)$ is the probability density function of losses, then for a confidence level $\alpha\%$,

$$\alpha = \Pr\left(Loss > -VaR\right) = \int\limits_{-VaR}^{\infty} g\left(x\right) dx. \tag{2.2}$$

The time period (d) and confidence level (α) are two main parameters to be selected appropriately for the risk measure.

In mathematical terms, Value-at-Risk (VaR) is the $(1-\alpha)^{\text{th}}$ quantile of the distribution of the d-period return value of a given portfolio P. Thus,

$$VaR_{\alpha,d}\left(P\right) = -F_{P^d}^{-1}\left(1-\alpha\right) \times PV\left(P\right), \tag{2.3}$$

where α is the confidence level, P^d is the change in value of the portfolio P over time period (d), i.e., the d-period return, F_{P^d} is the cumulative probability distribution of P^d and $PV(P)$ is the present worth of the portfolio.

3. L_p-norm Linear Regression Models. The least-squares estimation (*LSE*) technique was first published by Legendre [12]. LSE is used for estimating linear regression models (LRMs). LRMs based on *LSE* perform well provided the errors follow a normal or approximately normal distribution, do not possess large size outliers and follow Gauss-Markov assumptions. Under these conditions, the *LSEs* are optimal and provide the best linear unbiased estimators of the model parameters. Alternatives to the *LSE* which are more robust to departures from the usual least squares assumptions have been investigated by several researchers; cf. interesting work by Boscovich [1], Legendre [13], Gauss [5], [6], Laplace [11], Stigler [17], Farebrother [4].

We consider a linear regression model for a single response variable y given k independent variables:

$$y = X\beta + \varepsilon, \tag{3.1}$$

where y is the vector of n responses, X is the $n \times k$ matrix of values of k independent variables, β is the vector of $k+1$ model parameters and ε is the vector of n residual values. Errors ε's are assumed to follow a multivariate normal distribution.

Definition. *The L_p-norm of the residual vector ε*

$$\|\varepsilon\|_p = \begin{cases} (\Sigma|\varepsilon|^p)^{1/p}, & \text{for } p \in [1;\infty), \\ \max|\varepsilon|, & \text{for } p \to \infty. \end{cases} \tag{3.2}$$

An estimator minimizing a L_p-norm of the residual vector ε is called an L_p-norm estimator. Measuring the size of ε using the L_p-norm, we arrive at the L_p-regression problem.

In regression analysis, the goal is to find β that attains the minimum L_p-norm for the difference between y and $X\beta$. Thus, the L_p-regression problem is to determine β such that

$$\min \|X\beta - y\|_p. \tag{3.3}$$

Setting $p = 1$ in (3.3), the L_1-norm regression problem becomes $\min\|X\beta - y\|_1$, which is written as the linear programming (LP) problem

$$\min \Sigma t_i : -t_i \le x_i^T\beta - y_i \le t_i, \quad i = 1, \ldots, n. \tag{3.4}$$

A methodology to estimate unknown parameters in L_1-norm regression model was first introduced by Boscovich [1].

Setting $p = 2$ in (3.3), the L_2-norm regression problem becomes $\min\|X\beta - y\|_2$. This is equivalent to minimizing with respect to β:

$$\Sigma(y_i - \Sigma x_{ij}\beta_j)^2. \tag{3.5}$$

The solution of β for L_2-norm regression problem is commonly known as the least square estimators (LSE). It may be noted from the works of Legendre [13] and Gauss [5] that they proposed to minimize the sum of the squares of the measurement errors and, thereafter, the method of least squares became the most popular estimating technique. The main reasons for the LSE's popularity are presumably easy computations and due the fact that when the residuals are independent and identically normally distributed, the least squares estimators regression model is also the best linear unbiased estimator as well as equivalent to the maximum likelihood estimator, implying that the inference is easily performed (Nyquist [16]). However, it has been noted that the least squares estimates are sensitive to departures from the assumptions, for example, normally distributed errors.

Setting $p = \infty$ in (3.3), the L_∞-norm regression problem becomes $\min \|X\beta - y\|_\infty$ which can be written as the linear programming (LP) problem

$$\min t : -t \leq x_i^T\beta - y_i \leq t, \quad i = 1, \dots, n. \tag{3.6}$$

This minimization problem is often referred to as the Chebyshev approximation. Laplace [11] and Edgeworth [3] have shown that the L_p-norm estimator is preferable to least squares when estimating a simple linear regression model with fat-tailed distributed residuals. Nyquist [16] has investigated the L_p-norm estimators of linear regression models. In particular, he discussed results on the existence, uniqueness and asymptotic distributions of L_p-norm estimators and gave geometrical interpretations of L_p-norm estimation.

4. Checking Model Adequacy. For assessing model adequacy, one commonly used measure based on estimated residuals is the coefficient of determination R^2 which is defined as the proportion of the total response variance that is explained by the model

$$R^2 = 100\left[1 - \frac{\Sigma\varepsilon^2}{\Sigma(y-\bar{y})^2}\right] \tag{4.1}$$

Another measure [Kumar and Kashanchi] denoted by $\|R^2\|_1$ based on the estimated residuals for checking model accuracy is

$$\|R^2\|_1 = 100\left[1 - \frac{\Sigma|\varepsilon|}{\Sigma|y-\bar{y}|}\right] \tag{4.2}$$

Either measure compares how well model fits. A higher value of $\|R^2|_1$ or R^2 indicates a better fit.

5. Descriptive Data Analysis of Iran's Crude Oil Prices and Export. Iran is one of the largest crude oil exporting counties and a risk analysis of this product may affect major economic policy decisions. Iran is a member of the Organization of the Petroleum Exporting Countries (OPEC) which sets crude oil prices in the world and is a reliable data source. The Central Bank of Iran is another reliable data source which provides economical and statistical data and reports about Iran. We referred to OPEC and the Central Bank of Iran and compiled the following data: (i) Monthly prices for a barrel of crude oil from OPEC and Iran's light and heavy crude oil (in US dollar) from January 1997 to December 2008 [22] and Number of barrels of crude oil exported by Iran per month from January 1997 to December 2008 [23]. Note that these data on export were presented in the Iranian calendar and we have converted it to the Western calendar.

In Figure 5.1 we can see the time series plots of the data (there are 144 observations in all). The left-hand side of Figure 5.1 shows the oil prices of OPEC and Iran light and heavy crude oil prices which are strongly positively correlated in the period of twelve years (January 1997–December 2008). The right-hand side of Figure 5.1 shows the number of barrels of crude oil (in 1000) exported by Iran and we can see significant fluctuations in the years 2001 and 2002.

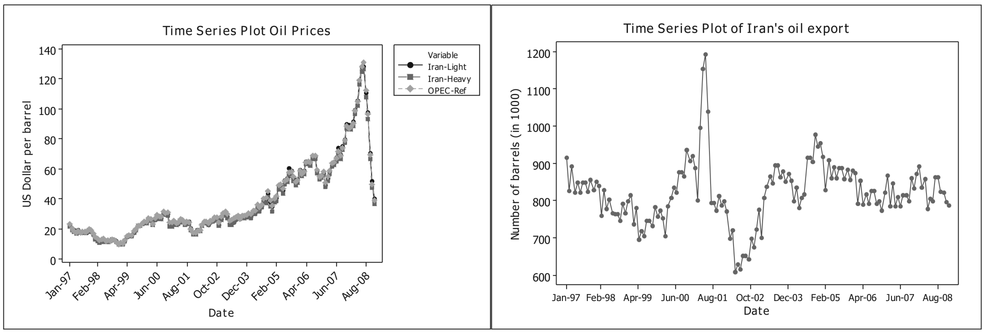
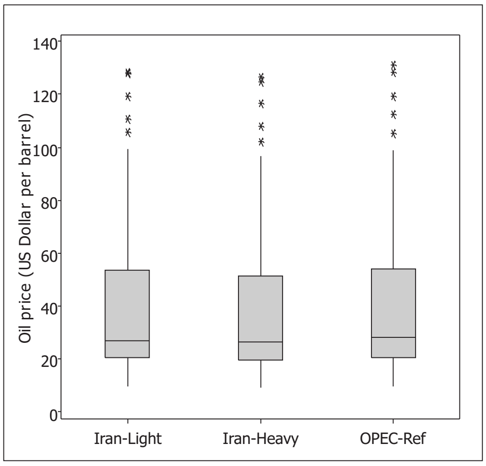


Fig. 5.1. Time series plots of the crude oil prices and export per month data, January 1997–December 2008

In Figure 5.2, the graphical summaries of the above data are presented. The distribution of crude oil prices are skewed to the right. During April 2008 and August 2008 the prices are above 97 US Dollars per barrel. We also noted suspected outliers in Figure 5.2 left box plots. The right side graph of Figure 5.2 shows the distribution of Iran's crude oil export that is almost symmetric with few suspected outliers at both ends. May 2001 and April 2002 with 1192 and 608 thousand barrels per day respectively were the maximum and minimum of Iran's crude oil export.

The numerical description of the data is provided in Table 5.1. Some



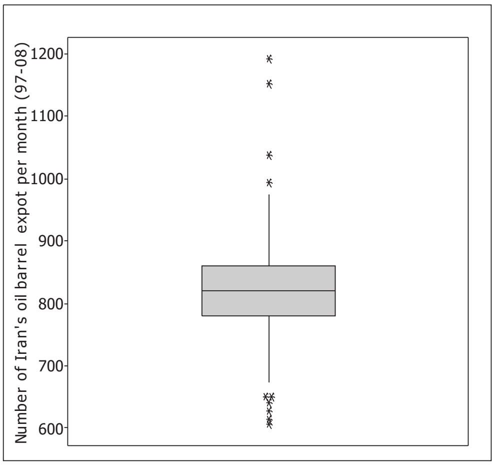


Fig. 5.2. Left is the box plots of the crude oil prices and right is the boxplot of the crude oil export per month, January 1997–December 2008

interesting observations:

1. Crude oil prices: skewed to the right, suspected outliers, Righ variability.
2. Crude oil export: almost symmetric, few suspected outliers at both ends, very high variability.
3. Linear correlation between crude oil prices and export: 0.149.

Table 5.1. Descriptive statistics

$N = 144$	Minimum	Maximum	Mean		Standard Deviation
	Statistic	Statistic	Statistic	Std. Error	Statistic
Iran-Light ($)	9.80	128.19	38.14	2.167	26.008
Iran-Heavy ($)	9.51	126.75	36.83	2.097	25.163
OPEC-Ref ($)	9.69	131.22	38.54	2.146	25.756
Export (000)	608	1192	820.25	6.999	83.985

6. VaR Calculations. In this section we will discuss the results of the risk analysis for the data and copula simulated data of the Iran's heavy and light crude oil prices. Here the Value-at-Risk (VaR) is calculated in terms of the relative change of the monthly crude oil prices. We are defining risk as the unexpected rare event of observing either a very low relative change (drop in oil price) or a very high relative change (high oil price).

We carry out the calculation steps:

(i) Fit L_1, L_2 and L_∞-norm regression models.

$$y_1 = 0.4020X + 797.9541 \tag{6.1}$$

$$y_2 = 0.4715X + 802.8855 \tag{6.2}$$

$$y_\infty = 1.8726X + 789.0337 \tag{6.3}$$

(ii) Check model adequacy

Table 6.1. Model accuracy measures

Model adequacy measure	L_1	L_2	L_∞
R^2	1.2	2.0	0.8
$\Vert R^2\Vert_1$	92.9	92.8	90.9

(iii) Predict export and predicted relative errors for lower and upper percentiles of crude oil prices.

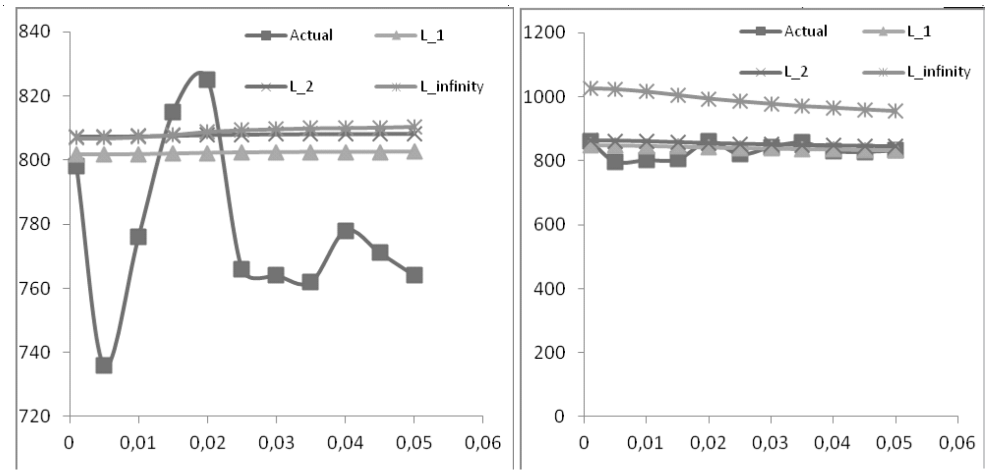


Fig. 6.1. Predicted crude oil export (000 Barrels) from L_1, L_2 and L_∞-norms

Table 6.2. Predicted crude oil export (000 Barrels) from L_1, L_2 and L_∞-norms

Probability	Price ($/Barrel)	Export (000)	Export L_1 (000)	Export L_2 (000)	Export L_∞ (000)
0.001	9.5143	798	801.7788	807.3715	806.8502
0.005	9.5315	736	801.7857	807.3796	806.8823
0.010	9.7249	776	801.8635	807.4708	807.2445
0.150	10.0773	815	802.0052	807.6369	807.9045
0.200	10.6064	825	802.2179	807.8864	808.8952
0.250	10.8595	766	802.3196	808.0058	809.3692
0.300	11.0280	764	802.3874	808.0852	809.6847
0.350	11.1703	762	802.4446	808.1523	809.9512
0.400	11.2132	778	802.4618	808.1725	810.0315
0.450	11.2909	771	802.4930	808.2092	810.1770
0.050	11.4135	764	802.5423	808.2670	810.4066
0.050	88.3090	833	833.4543	844.5232	954.4011
0.450	91.0695	828	834.5640	845.8247	959.5704
0.400	94.0592	831	835.7659	847.2344	965.1690
0.350	96.6618	857	836.8121	848.4615	970.0426
0.300	100.6205	843	838.4035	850.3281	977.4556
0.250	104.7248	820	840.0534	852.2632	985.1413
0.200	109.2718	861	841.8814	854.4072	993.6561
0.150	115.2564	804	844.2872	857.2289	1004.8627
0.010	121.1383	802	846.6517	860.0022	1015.8773
0.005	125.2557	797	848.3069	861.9435	1023.5874
0.001	126.4511	861	848.7875	862.5072	1025.8261

From Table 6.2 (Figure 6.1) and Table 6.3 (Figure 6.2), we have the following indicative conclusions:

- Models in order of goodness of fit (First is the best): L_1, L_2, L_∞-norm regression models.
- Model adequacy: R^2 values are very small for all models because of lower degree of linear association and assumptions of optimality of linear models not met.

  Values of $\|R^2\|_1$ measure are high ($> 90\%$) for all three models.
- Predicted export for lower and upper percentiles of crude oil prices: Predicted values are close enough. However, absolute errors are minimum for the L_1-regression model. The L_2 (or usual least square) regression model is not optimal, particularly for predicting end-tail values.

Table 6.3. Predicted relative errors (%) from L_1, L_2 and L_∞-norms

Probability	|Relative Error|	|Relative Error|	|Relative Error|
0.001	0.4735	1.1744	1.1090
0.005	0.4744	1.1754	1.1131
0.010	0.4841	1.1868	1.1585
0.150	0.5019	1.2076	1.2412
0.200	0.5286	1.2389	1.3653
0.250	0.5413	1.2539	1.4247
0.300	0.5498	1.2638	1.4643
0.350	0.5570	1.2722	1.4976
0.400	0.5591	1.2748	1.5077
0.450	0.5630	1.2793	1.5259
0.050	0.5692	1.2866	1.5547
0.050	4.4429	5.8300	19.5991
0.450	4.5820	5.9931	20.2469
0.400	4.7326	6.1697	20.9485
0.350	4.8637	6.3235	21.5592
0.300	5.0631	6.5574	22.4882
0.250	5.2699	6.7999	23.4513
0.200	5.4989	7.0686	24.5183
0.150	5.8004	7.4222	25.9226
0.010	6.0967	7.7697	27.3029
0.005	6.3041	8.0130	28.2691
0.001	6.3643	8.0836	28.5496
Sum (Lower)	5.8020	13.6137	14.9620
Sum (Upper)	59.0185	76.0306	262.8558
Average (Lower)	0.5275	1.2376	1.3602
Average (Upper)	5.3653	6.9119	23.8960

- For considered crude oil price and export data series, the L_1-regression model is recommended over the L_2 (or least square) regression model and the L_∞-regression model.

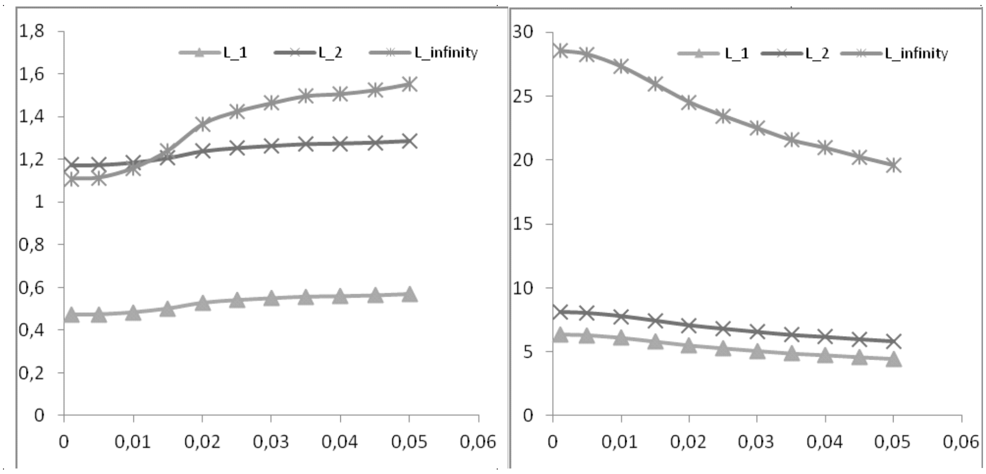


Fig. 6.2. Predicted relative errors (%) from L_1, L_2 and L_∞-norms

7. Concluding Remarks. Although the least squares estimation (LSE) is simple and algebraically highly developed, studies have shown that LSE-based linear regression may not be the optimal model when one or more of its assumptions fail. In our analysis, we have estimated L_1, L_2, L_∞-norm regression models to calculate VaR for Iranian crude oil price and export data. Our findings are in agreement with those in some earlier studies about the L_p-norm based linear regression models.

Our results also indicate a need for further investigations like: (i) distributional properties of the L_p-norm based linear regression estimated models, (ii) effects of deviating from the assumptions of LSE on the L_p-norm linear regression models, (iii) statistical inference issues such as interval estimation, hypothesis testing and prediction bands, etc., for the L_p-norm models, and (iv) for given applications, how to determine optimal choice of the L_p-norm.

REFERENCES

[1] Boscovich R. J. De literaria expeditione per ponticiam ditionem et synopsis amplioris operis. *Bononiensi Scientiarum et Artum Instituto atque Academia Commentarii*, **4** (1757), 353–396. (Reprinted with a Croatian translation by N. Cubranic, Institute of Higher Geodesy, Zagreb, 1961).

[2] Duffie D., J. Pan. An overview of Value at Risk. *J. of Derivatives*, **4** (1997), 7–49.

[3] Edgeworth F. Y. On observations relating to several quantities. *Phil. Mag.* (5th Series), **6** (1887), 279–285.

[4] Farebrother R. W. Fitting Linear Relationships: A History of the Calculus of Observations 1750-1900. Springler-Verlag, New York Inc, 1999.

[5] Gauss C. F. Theoria motus corporum coellestium in sectionibus conicis solem ambientium. Hamburg, 1809.

[6] Gauss C. F. Theoria combinations observationum erroribus minimus obnoxiae pars prior. Printed in Werke, Gottingen, 1820.

[7] Grundy B. D., Z. Wiener. The analysis of VAR, Deltas and State prices: A New Approach. Working paper of the Wharton School, University of Pennsylvania, 1996.

[8] Hull J., A. White. Incorporating volatility updating into the Historical Simulation method for Value at Risk. *J. of Risk*, **1** (1998), 5–19.

[9] Joe H. Multivariate Models and Dependent Concepts. Chapman & Hall, New York, 1997.

[10] Jorion P. How informative are Value-at-Risk disclosures? *Accounting Review*, **77** (2002), 911–932.

[11] Laplace P. S. Theorie analytique des probabilities. Paris, 1812.

[12] Legendre A. M. Nouvelles methodes pour la determination des orbites des cometes. Paris, 1805.

[13] Legendre A. M. Nouvelles methodes pour la determination des orbites des cometes. Encyclopdia Britannica, 2014.
`http://www.britannica.com/EBchecked/topic/420949/`
`Nouvelles-methods-pourla-determination-des-orbites-des-cometes`, 1806

[14] Linsmeier T., N. Pearson. Risk Measurement: An Introduction to Value at Risk. Working paper of the University of Illinois, 1996.

[15] Morgan J. P. RiskMetrics – Technical Document. Morgan Guaranty Trust Company, New York, 1996.

[16] NYQUIST H. Recent Studies on Lp-norm Estimation. Doctoral thesis, University of Umeå, 1980.

[17] STIGLER S. M. The History of Statistics: The Measurement of Uncertainty Before 1900. Harvard University Press, 1990.

[18] For Bankers Trust and Modern Risk Management. (WEB- Resources) http://fic.wharton.upenn.edu/fic/case%20studies/Birth%20of%20Modern%20Risk%20Management%2003-06-2008.pdf

[19] For RiskMetrics and CreditMetrics. (WEB- Resources) http://www.riskmetrics.com

[20] Basel Committee on Banking Supervision, International Convergence of Capital Measurement and Capital Standards: A Revised Framework, Bank for International Settlement. (WEB- Resources) http://www.bis.org/publ/bcbs107.htm

[21] Basel Committee on Banking Supervision, Supervisory Framework for the Use of "Backtesting" in Conjunction with the Internal Models Approach to Market Risk Capital Requirements, Bank for International Settlement, January 1996. (WEB- Resources) http://www.bis.org/publ/bcbs22.htm

[22] Data Sources. http://www.opec.org/opec_web/en/publications/202.htm

[23] Data Sources. http://tsd.cbi.ir/DisplayEn/MostReq.aspx

Pranesh Kumar
Department of Mathematics and Statistics
University of Northern British Columbia
Prince George, BC V2N 4Z9, Canada
e-mail: Pranesh.kumar@unbc.ca

Faramarz Kashanchi
Planning and Performance Improvement
Northern Health
Prince George, BC V2L 5B8, Canada
e-mail: Faramarz.Kashanchi@northernhealth.ca

Received December 4, 2014
Final Accepted February 23, 2015