DSSR

Discussion Paper No. 99

Jensen's alpha measured and decomposed under skew symmetric semi-parametric model for error terms in the market model

Navruzbek Karamatov
Ryozo Miura

July, 2019

Data Science and Service Research Discussion Paper

Center for Data Science and Service Research
Graduate School of Economic and Management
Tohoku University
27-1 Kawauchi, Aobaku
Sendai 980-8576, JAPAN

Jensen's alpha measured and decomposed under skew symmetric semi-parametric model for error terms in the market model

Navruzbek Karamatov[1], Ryozo Miura[2]

[1] Graduate School of Economics and Management, Tohoku University
[2] Graduate School of Business, Hitotsubashi University

Abstract

A simple estimation method, namely the Ordinary Least Squares (LS) is applied for nearly all empirical analysis to estimate β. However, Jensen (1968) made clear that CAPM is not able to explain abnormal returns and α is used to account for this unobserved factors. More importantly $Jensen's\ Alpha$ is obtained as a mean value of residuals from a simple regression. Nonetheless, LS is sensitive to outliers and this could make estimators to be vulnerable. As empirical studies states, observed residuals are not symmetrically distributed.

Can asymmetry in error term distribution explain *Jensen's Alpha*? This research tries to find the answer by applying robust Rank statistics, in comparison with Least Squares, to fit a simple linear regression into Nikkei 225, FTSE 100 and S&P 500 stocks. Furthermore, the Generalized Lehmann's Alternative Model (GLAM) is applied to observed residuals to analyze the location and asymmetry of the residuals distribution.

We found that residuals are, indeed, noticeably skewed. GLAM model shows that majority of stocks in all three markets experience asymmetry, especially during the financially stressful periods in 2008. In addition, our asymmetry parameter θ possesses a statistically significant relation to α and to the *skew effect* which is defined as a difference between α and location (μ). Furthermore, in order to obtain the underlying F distribution we fitted t distribution with varying degrees of freedom. Our results show that most of the stocks experience smaller degrees of freedom meaning that R estimate is more efficient than its counterpart LS. Moreover we found that R approach is suitable even in the case of high degrees of freedom (close to normal) but large θ values. Next, we also found that LS underestimates α and β for majority of stocks with smaller degrees of freedom.

Keywords: *Jensen's Alpha, CAPM, ordinary least squares, LSE, rank statistics, R-estimate, Generalized Lehmann's Alternative Model*

***JEL Classification**: G11, G12*

[1] *E-mail*: karamatov.navruzbek.p5@dc.tohoku.ac.jp

[2] Professor Emeritus, Graduate School of Business Administration, Hitotsubashi University and Specially Appointed (Visiting) Professor, Graduate School of Economics and Management, Tohoku University, *E-mail*: ryozomiura@gmail.com

1 Introduction

Ordinary Least Squares (OLS) is primarily used for estimation of beta in a simple regression model which is called a market model in the field of finance. It is because OLS is the best linear unbiased estimator (BLUE) and the best estimate among all the linear estimates (linear functions of observations), thus, it has been used in almost every empirical study of the market model.

In this study we apply an estimation method based on rank statistics (R estimate in short). It provides us a nonlinear estimate (a nonlinear function of observations) of β and it has been known to be robust against outliers. As it is well known that outliers can be often observed when the distribution of error terms in a linear regression has a heavy tail.

Asymptotic accuracy of estimate of β can be measured by the asymptotic variance of these estimators and the relative accuracy of OLS and R-estimate depends on the heaviness or lightness of the tail of distribution. For instance, Chapter 5 in Lehmann (1983), page 384, shows that R-estimator is better than of OLS when the distribution function of error term has a heavy tail, but it reverses when the distribution is light, i.e. it is Normal or close to Normal. We will show in our study that more than half of individual stocks in the major markets: Tokyo, London and New York, have rather heavy tails. Our view is that more accurate estimate of β will provide us more accurate residuals, so that the important parameter α can be more accurately estimated.

Study looks at the symmetry and asymmetry of the distribution of error terms in the market model by applying GLAM which is a semi-parametric model. GLAM can describe a family of distributions including an underlying symmetric distribution F which centers around a location parameter μ and GLAM represent how much F is skewed (or asymmetric) along with a parameter θ. Besides, paper shows theoretically and mathematically that the residuals can be used to estimate θ (asymmetry parameter) and μ (location parameter) as well as F based on Z_i in Miura and Tsukahara (1993).

α is estimated simultaneously with β under OLS method. But method based on rank statistics estimates β without having to estimate α. Then, α in this approach estimated by the sample mean of residuals. This is concordant with OLS of α as it can be defined by expectation of [α + *error term*] in a simple linear regression model. This approach makes us able to decompose α as a sum of location (μ) and asymmetry effect (θ).

We found that depending on the period a large part of α is contributed by a skew effect whose degree is indicated by θ especially in US stocks.

Grouping stocks based on *df* clearly illustrated that in 5-15 *df* subgroup α estimated by LS is often underestimated when the error term distribution has a heavy tail, compared with α based on residuals brought by R approach.

Following the empirical study of relations among those parameters, we propose a certain recommendation on when to use LS or R estimate so that the empirical work may have more accuracy both in academics and practice.

This paper is organized as follows. Section 2 reviews previous studies related to this study and section 3 reviews statistical properties of our methodologies. Section 4 presents data and its descriptive summary. In addition, section meticulously introduces to LS and R methods as well as to models that is employed by this study. Estimated β based on two approaches and residual analysis are in section 5. Besides, this section also includes cross sectional study of GLAM and skew-t distribution parameters. Our main results, *Jensen's Alpha* decomposition and its relation to asymmetry parameter also presented in section 5. Next, section 6 presents empirical findings for estimation of underlying distribution of observed residuals. Lastly, section 7, sums up main findings and concludes with possible directions for future research.

2 Review of previous studies

Among others, study by Jensen (1968) made clear that CAPM is not able to explain abnormal stock or portfolio returns and α, intercept of the linear regression, is added as an additional variable to account for extra variability that is left unexplained by market return. Empirical researches proved α has a non-constant nature and fluctuates during the time period (B. Arnott., *et al,* (2018)). It is known as a *Jensen's Alpha* and applied as one of the portfolio strategies that exist out in the market today.

Nonetheless, LS alternatives and modifications of it are based on a number of assumptions and sensitive to outliers clustering found in Onder and Zaman (2003, 2005). Moreover, Hettmansperger and Sheather (1992) showed that the Least Median Squares is instable when centrally located data changes. Recently, Denhere and Bindele (2015) compared Rank based estimation with LS and LAD estimators, and found that R estimators are robust compared to parametric methods when data has outlying observations and fat-tailed error distribution. Besides, we found that finance literature also lacks of study for an application of robust estimation technique for CAPM β and *Jensen's Alpha* estimation, such as a distribution free Rank based methods.

Nonparametric methods gained popularity due to several advantages than traditional approaches and rank statistics is one of the widely used approach. Rank method has been developed extensively by a number of studies such as Jureckova (1971) and Jaeckel (1972). In specific, Jureckova (1971) mathematically establishes the asymptotic linearity of rank statistics and infers its asymptotic normality for a multiple linear regression case. Besides, Jackel (1972) introduces dispersion measures and minimization procedure in order to derive regression parameters. Asymptotic normality is also shown to be the same as in Jureckova (1971) case. Especially, in the case of a simple linear regression, estimator is a weighted mean of pairwise slopes $(Y_j - Y_i)/(c^j - c^i)$ $\{j \neq i\}$.

Rank method does not require the underlying observations to follow any specific distribution such as normal distributions and it provides distribution free estimation - which is the main reason for its popularity. Moreover, being insensitive to outliers and efficiency properties are the key reasons for applying these methods in the analysis rather than LS (Hettmansperger and McKean (1977)).

Miura(1985a,b) computed estimates of beta based on monthly data for the period from 1952 January to 1981 December and showed the difference of the two estimates of beta based on LS and nonparametric estimate based on R statistics. Also he fitted Log-Normal distribution to the residuals and showed the relations between the estimated scale parameter of Log-Normal distribution and the estimate of asymptotic variance of the two estimators. However, the model was not adequate because the choice of the location was ad-hoc and it did not cover the case of asymmetric distribution. Zhou(2001) followed the same scheme as Miura(1985a,b) to compute beta based on daily data. In this paper we use Generalized Lehmanns Alternative model which can take good care of location and asymmetry. This corrects an ad-hoc treatment of location in the Log-Normal fitting in Miura(1985a,b) and Zhou(2001).

3 Review of statistical properties

Study employs a simple linear regression in Eq. (1) where $i = 1, ..., n$. Error terms (ϵ_i) are expected to be i.i.d and have a distribution $G(x)$.

$$Y_i = \alpha + \beta x_i + \epsilon_i \quad (1) \qquad \eta_i = Y_i - \beta x_i = \alpha + \epsilon_i \quad (2)$$

$$\eta \sim G(x - \mu) \equiv h(F(x - \mu) : \theta) \quad (3)$$

as defined later in Eq. (12)

3.1 Optimality of Least Squares (β estimation)

Least Squares estimate β is considered to be the best linear unbiased estimate (BLUE) based on Gauss-Markov theorem. It states that β is the minimum variance and linear unbiased estimator of true β, as long as the assumptions of classical linear regression model are hold (Greene, 2012).

However, R-estimate of β is a non-linear function of Y_i. It as been known that asymptotic variance of R-estimate is smaller then LSE β when the distribution of ϵ_i (or η_i) has a heavy tail.

3.2 Asymptotic normality of estimates

Asymptotic normality of LS and R estimates are presented below in Eq. (4) and (5), respectively. When n is large enough, both estimates will reach to the true parameter β. In addition, variances of both estimates are presented in Eq. (6) and (7), respectively (Lehman (1983, Chapter 5)).

$$\sqrt{n}(\hat{\beta}_{LS} - \beta) \to N(0, \sigma^2_\beta) \quad (4) \qquad \sqrt{n}(\hat{\beta}_R - \beta) \to N(0, \sigma^2_\beta) \quad (5)$$

$$\sigma^2_{\beta,LS} = \frac{1}{c^2}\int_{-\infty}^{\infty} x^2 g(x)dx \quad (6) \qquad \sigma^2_{\beta,R} = \frac{1}{12c^2\{\int_{-\infty}^{\infty} g^2(x)dx\}^2} \quad (7)$$

$$g(x) = G'(x) = h'(F(x) : \theta)f(x) \quad (8)$$

$$c^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad (9) \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (10)$$

We further focus on error terms by applying Generalized Lehamnn's Alternative Model.

3.3 The Generalized Lehmann's Alternative Model

The GLAM method is semi-parametric and based on rank statistics. The following definitions and assumptions of GLAM is from Miura and Tsukahara (1993) and we keep notations unchanged for simplicity.

Let Θ be interval in real line. A function $h(t;\theta)$ for $t \in (0,1)$ and $\theta \in \Theta$ which satisfies the following (1) and (2) is called the Generalized Lehmann's Alternative model:

(1) $h = (0;\theta) = 0$ and $h(1;\theta) = 1$ for any $\theta \in \Theta$. $h(t;\theta)$ is strictly monotone function of t.

(2) There exists $\theta^* \in \Theta$ such that $h(t;\theta^*) = t$ for $t \in (0,1)$. And for $\theta < \theta'$, $h(t;\theta) < h(t;\theta')$ for all t.

X observations are assumed to be i.i.d and have an empirical distribution function given by $G(x : \mu, \theta)$. Deformation in $G(x : \mu, \theta)$ is captured by the parameter θ.

$$h(t;\theta) = 1 - (1 - t)^\theta \quad (11)$$

$$G(x:\mu,\theta) = h(F(x-\mu);\theta) = 1-(1-F(x-\mu))^{\theta} \tag{12}$$

3.4 Estimation of θ and μ based on η

To obtain μ and θ parameters we followed the estimation procedure presented in Miura and Tsukahara (1993).

Regard the following X_i as our η_i. Our estimation of θ and μ based on the residuals after estimating β. It will be described in the subsection 4.2.2 where estimation of β based on Rank statistics is also described. Regard there $e_i(\beta_0) \equiv \eta_i$. The section 3.5 provides a mathematical statement with a proof which makes a bridge between $e_i(\hat{\beta})$ and $e_i(\beta_0)$, in other words makes the estimation procedure in Miura and Tsukahara (1993) usable being based on the residuals $e_i(\hat{\beta})$ rather than $\eta_i \equiv X_i$.

Let $X_1, ...X_n$ are $i.i.d$ random variables following $G(x:\mu,\theta)$. First, the empirical distribution function for observation X_i is defined as follows.

$$G_n(x) = n^{-1}\sum_{i=1}^{n} I_{[X_i<x]} \tag{13}$$

Next, the empirical distribution function is linearized.

$X_{(1)} < X_{(2)}... < X_{(n)}$ are ordered values of X_i's for $i=1,...,n$. $X_{(0)} = X_{(1)} - 1/n$ and $X_{(n+1)} = X_{(n)} + 1/n$ are set, respectively.

$$\tilde{G}_n(x) = \frac{x + iX_{(i+1)} - (i+1)X_{(i)}}{(n+1)(X_{(i+1)} - X_{(i)})} \tag{14}$$

where, $x \in (X_{(i)}, X_{(i+1)}]$.

Following the linearization, Z_i values are defined by the inverse of $\tilde{G}_n(x)$.

$$Z_i(r) = \tilde{G}_n^{-1}(h(\frac{i}{n+1};r)) \tag{15}$$

for $i=1,...,n$ and r is a tentative parameter for θ.

Then, $R_i^+(r,q)$ are estimated for a given tentative location parameter q.

$$R_i^+(r,q) = (the \quad number \quad of \quad \{j : |Z_j(r)-q| \leq |Z_i(r)-q|\}) \tag{16}$$

The rank statistics used for (θ, μ) inference are defined as follows.

$$S_{\theta,n}(r,q) = \frac{1}{n}\sum_{i:Z_i(r)>q} J_\theta((1-\frac{R_i^+(r,q)}{n+1})/2) + \frac{1}{n}\sum_{i:Z_i(r)\leq q} J_\theta((1+\frac{R_i^+(r,q)}{n+1})/2) \tag{17}$$

$$S_{\mu,n}(r,q) = \frac{1}{n}\sum_{i:Z_i(r)>q} J_\mu((1-\frac{R_i^+(r,q)}{n+1})/2) + \frac{1}{n}\sum_{i:Z_i(r)\leq q} J_\mu((1+\frac{R_i^+(r,q)}{n+1})/2) \tag{18}$$

Score functions given by Eq. (43) and (44) are used for Eq. (17) and (18) to estimate θ and μ parameters simultaneously. Statistics are simultaneously minimized as in Eq. (19) to obtain optimal parameters of $\hat{\mu}$ and $\hat{\theta}$.

$$
\begin{aligned}
S_{\theta,n} &\approx 0 \\
S_{\mu,n} &\approx 0 \\
D_n \triangleq \{(r,q) : \sum_{k=1}^{2} &|S_{k,n}(r,q)| = min\}
\end{aligned} \tag{19}
$$

Asymptotic normality of $\hat{\theta}$ and $\hat{\mu}$ are given in great detail in Theorem 3.2 in Miura and Tsukahara (1993).

$$
\sqrt{n}\begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \hat{\mu}_n - \mu_0 \end{pmatrix} \to N(0, D^{-1}\Sigma(D^{-1})') \tag{20}
$$

Here, $D = [d_{k,l}]$ and Σ are covariance matrix of T as given in Miura and Tsukahara (1993).

$$
d_{k,1} = \int_0^1 \{\frac{h_2(t;\theta_0)}{h_1(t;\theta_0)} + \frac{h_2(1-t;\theta_0)}{h_1(1-t;\theta_0)}\} dJ_k(t), \tag{21}
$$

$$
d_{k,2} = -2\int_0^1 f(F^{-1}(t))dJ_k(t) \tag{22}
$$

$$
T_k = \int_0^1 \{\frac{U(h(t;\theta_0))}{h(t;\theta_0)} + \frac{U(h(1-t;\theta_0))}{h_1(1-t;\theta_0)}\} dJ_k(t), \tag{23}
$$

$k = 1, 2$.

We employ η_i from Eq. (3) for GLAM instead of X_i. The statistical properties of applicability of η_i are meticulously presented in the next section.

3.5 Estimation based on residuals

Assume that we have an estimate $\hat{\beta}$ of β which has $\sqrt{n}$ - asymptotic normality. For instance, $\hat{\beta}$ can be either $\hat{\beta}_{LSE}$ based on LSE or $\hat{\beta}_R$ based on rank statistics.

Now we can write the rank statistics for θ and μ.

$$
\begin{aligned}
S_{n,\theta}((r,q):\hat{\beta}_n) &= \frac{1}{n}\sum_{i:Z_i^*>q} J_\theta((1+\frac{R_i^+(r,q:\hat{\beta}_n)}{n+1})/2) + \frac{1}{n}\sum_{i:Z_i^*\le q} J_\theta((1-\frac{R_i^+(r,q:\hat{\beta}_n)}{n+1})/2) \\
S_{n,\mu}((r,q):\hat{\beta}_n) &= \frac{1}{n}\sum_{i:Z_i^*>q} J_\mu((1+\frac{R_i^+((r,q):\hat{\beta}_n)}{n+1})/2) + \frac{1}{n}\sum_{i:Z_i^*\le q} J_\mu((1-\frac{R_i^+(r,q:\hat{\beta}_n)}{n+1})/2)
\end{aligned} \tag{24}
$$

where J_θ and J_μ are the score functions for θ and μ.

Proposition A-1

Let β_0 be the true value of β.

For

$$
\begin{aligned}
r &= \theta_0 + \frac{b_1}{\sqrt{n}} \\
q &= \mu_0 + \frac{b_2}{\sqrt{n}} \\
|b_1| &\leq B \\
|b_2| &\leq B
\end{aligned} \tag{25}
$$

$$
\begin{aligned}
&\sqrt{n}\big\{S_{n,\theta}((r,q):\hat{\beta}_n) - S_{n,\theta}((r,q):\beta_0)\big\} \\
&\quad \to T(\bar{x}) \int_{-\infty}^{\infty} f(x) dJ_\theta(F(x))
\end{aligned} \tag{26}
$$

as $n \to \infty$ and

$$
\begin{aligned}
&\sqrt{n}\big\{S_{n,\mu}((r,q):\hat{\beta}_n) - S_{n,\mu}((r,q):\beta_0)\big\} \\
&\quad \to T(\bar{x}) \int_{-\infty}^{\infty} f(x) dJ_\mu(F(x))
\end{aligned} \tag{27}
$$

Proof for A-1 is given in Appendix.

4 Data and estimation procedure

4.1 Data

Paper relies on three stock market index constituents, Nikkei 225 (N225), FTSE 100 and S&P500 for this study. N225 data is obtained from Quick Financial Data Provider and it is a set of stock prices of all Nikkei 225 stocks in a daily frequency. Similarly FTSE 100 and S&P 500 are in daily frequency as well and obtained through Thomson Reuters Database. Time period coverage by datasets varied depending on the market. N225 data time span is from Q1 1998 until Q3 2017, FTSE100 data time span is from Q1 1986 to Q3 2017 and S&P500 data time span ranged from Q1 1994 to Q3 2018. Rate of returns are estimated as the difference of prices (P_t - P_{t-1}) over price at $t-1$. As a risk free rate - overnight call money rate of the Bank of Japan is employed[3] for N225 stocks, London Interbank Offered Rate (LIBOR) for FTSE 100 stocks and 1-month US Treasury Bill rate for S&P500 stocks. The chosen risk free rate is in line with previous researches for Japanese market (Kubota and Takehara (2010)). Descriptive statistics for index and risk free rates are presented in Table (1).

Table 1

Statistic	Quarters	Mean	St. Dev.	Min	Max
N225	79	0.0001	0.015	−0.114	0.142
Call money rate	79	0.001	0.002	−0.001	0.007
FTSE 100	131	0.0001	0.012	−0.088	0.098
1 month LIBOR	131	0.031	0.024	0.002	0.078
S&P 500	99	0.0003	0.012	−0.090	0.116
1 month Treasury bill	99	0.020	0.021	0.000	0.064

[3]https://www.boj.or.jp/en/statistics/market/short/mutan/index.htm/

4.2 β estimation

4.2.1 LS Method

A simple linear regression model is given in Eq. (1). LS method relies on minimizing the sum of squared residuals (29). Estimation window consisted of moving and non overlapping 3 month. For each stock all available rate of returns are divided into quarters with a given month and year information. Number of returns are not the same for each quarter due to trading and non trading day differences for every month. However, the available number of observations for stock returns per quarter are found to be in the range of 59 and 63. This approach of analysis ensures our estimates to be conducted for every single quarter of the year and makes it possible to gain extra insight of a given stock behavior during the period. Hence, more than 100 β values are estimated for each stock names depending on the availability of stock returns for all sample period.

$$R_{i,q,t} - R_{f,q,t} = \alpha_{i,q} + \beta^{ls}_{i,q}(R_{m,q,t} - R_{f,q,t}) + \epsilon_{i,q,t} \tag{28}$$

$$SSR_{i,q}(\epsilon_{i,q}) = \sum_{t=1}^{T}((R_{i,q,t} - R_{f,q,t}) - \alpha_{i,q} - (R_{m,q,t} - R_{f,q,t})\beta^{ls}_{i,q})^2 \tag{29}$$

$$u_{i,q,t} = (R_{i,q,t} - R_{f,q,t}) - \alpha_{i,q} - \hat{\beta}^{LS}_{i}(R_{m,q,t} - R_{f,q,t}) \tag{30}$$

$$i = \{1, ..., 225\}, q = \{1, ..., N\}, t = \{1, ..., T\} \tag{31}$$

Here, R_i - stock rate of return, R_f - risk free rate, R_m - market rate of return, ϵ_i - LS error term, u_i - LS residual. i is the available stocks in our data set and varies depending on a stock market, N is the a maximum number of quarters available for a given stock and T is the maximum number of stock returns available for a given quarter.

4.2.2 R Method

Eq. (32) presents R approach. Similar to LS method, estimation window consisted of moving and non overlapping 3 month. For each stock all available rate of returns are divided into quarters with a given month and year information. However, in the case of rank statistics not sum of squared residuals but the sum of dispersions in Eq. (34) are minimized. We employed the simplest and commonly applied score function - Wilcoxon scores (Jaeckel (1972)) as in (33).

$$R_{i,q,t} - R_{f,q,t} = \beta^{R}_{i,q}(R_{m,q,t} - R_{f,q,t}) + \eta_{i,q,t} \tag{32}$$

$$W_T(R_\eta) = \frac{R_\eta}{T+1} - \frac{1}{2}(\Leftrightarrow J_\beta(t) = t - \frac{1}{2}) \tag{33}$$

$$D_{i,q}(\eta_{i,q}) = \sum_{t=1}^{T}(\frac{R_{\eta_{i,q,t}}}{T+1} - \frac{1}{2})((R_{i,q,t} - R_{f,q,t}) - (R_{m,q,t} - R_{f,q,t})\beta^{R}_{i,q}) \tag{34}$$

Here, $D_i(\eta_i)$ - sum of dispersion, R_{η_i} - rank of η_i, $W_T(R_\eta)$ - Wilcoxon scores.

On the basis of observed values, the following $v_{i,q,t}$ are the estimates of $\eta_{i,q,t}$. v_i is residual obtained by R approach.

$$v_{i,q,t} = (R_{i,q,t} - R_{f,q,t}) - \hat{\beta}^R_{i,q}(R_{m,q,t} - R_{f,q,t}) \tag{35}$$

$$i = \{1, ..., 225\}, q = \{1, ..., N\}, t = \{1, ..., T\} \tag{36}$$

Here, i is the available stocks in our data set, N is the a maximum number of quarters available for a given stock and T is the maximum number of stock returns available for a given quarter.

Rank statistics for β is described below. Here, b is estimate.

$R_i(b)$ = rank of $e_i(b)$ among $\{e_j(b), j = 1, 2, ...n\} = \sum_{i=1}^{n} I\{e_j(b) \leq e_i(b)\}$

$R_i(b)$ does not change even some constant value is subtracted from e_i and makes it possible to estimate β without estimating α.

$$\begin{gathered} S_{n,\beta}(b) = \frac{1}{n}\sum_{i=1}^{n} J_\beta(\frac{R_i(b)}{n+1})(x_i - \bar{x}) = \frac{1}{n}\sum_{i=1}^{n} J_\beta(\frac{R_i(b)}{n+1})c_i \\ c_i = (x_i - \bar{x}) \\ J_\beta(t, g) = -\frac{g\prime(G^{-1}(t))}{g(G^{-1}(t))} \end{gathered} \tag{37}$$

$\hat{\beta}$ is the value of b which makes $|S_{n,\beta}(b)|$ closest to zero.

4.3 GLAM

Following the estimation of $\hat{\beta}$, residuals (u_i, v_i) are observed for every stock and quarterly period. Here, we present the procedure to obtain $\hat{\theta}$ and $\hat{\mu}$.

Here J_1 and J_2 are score functions for θ and μ respectively. The optimal score functions can be derived as following:

$$g(x : \mu, \theta) = \frac{dG(x : \mu, \theta)}{dx} \tag{38}$$

$$g_\theta(x : \mu, \theta) = \frac{dg(x : \mu, \theta)}{d\theta} \tag{39}$$

$$g_\mu(x : \mu, \theta) = \frac{dg(x : \mu, \theta)}{d\mu} \tag{40}$$

$$J_\theta(t) = \frac{g_\theta(G^{-1}_{\mu.\theta}(t) : \mu, \theta)}{g(G^{-1}_{\mu.\theta}(t) : \mu, \theta)} \tag{41}$$

$$J_\mu(t) = -\frac{g_\mu(G^{-1}_{\mu.\theta}(t) : \mu, \theta)}{g(G^{-1}_{\mu.\theta}(t) : \mu, \theta)} \tag{42}$$

However, these optimal scores are not available since the fundamental form of F is unknown. Here, the *logistic distribution* is employed to derive score functions.

$$J_\theta(t) = \frac{1}{\theta} + ln(1 - [1 - (1-t)^{1/\theta}]) = \frac{1}{\theta} + ln(1-t)^{1/\theta} \tag{43}$$

$$J_\mu(t) = -\frac{1}{s}\left[(\theta - 1)(-1)\left[1 - (1-t)^{1/\theta}\right] + 1 - 2(1-t)^{1/\theta}\right] \tag{44}$$

This is because we used $J_\beta(t) = t - \frac{1}{2}$ in Eq. (33) for estimating β which is an optimal score function for the case of $G_{\mu,\theta} \equiv F(x-\mu)$ with $\theta = 1$ and F is logistic. This makes us keep a consistency of our view on F.

Score functions given by Eq. (43) and (44) are used for Eq. (17) and (18) to estimate θ and μ parameters simultaneously. Statistics are simultaneously minimized as in Eq. (45) to obtain estimates of μ and θ.

$$\begin{aligned} S_{\theta,n} &\approx 0 \\ S_{\mu,n} &\approx 0 \\ D_n \triangleq \{(r,q) &: \sum_{k=1}^{2} |S_{k,n}(r,q)| = min\} \end{aligned} \tag{45}$$

$\hat{\theta}$ and $\hat{\mu}$ are obtained by minimizing Eq. (45), as explained in Eq. (19).

4.4 Skew-t distribution

Random values from a normal distribution have no skewness on either side of the distribution and displays a bell-shape form. However, this behavior is not observed in residuals (ϵ) from a simple linear regression Eq. (2) fitted into stock return. Hence, we applied a semi-parametric approach - GLAM to capture deformation by asymmetry parameter θ.

To estimate a skewness a widely used skew-t distribution (Azzalini, A., 1985) is used as well which is a parametric approach in order to compare with our semi-parametric approach by GLAM. To make a ground for fair comparison we choose degrees of freedom 8 which makes t distribution close to logistic distribution.

In Eq. (46) is presented a linear transformation of random variable Y which follows skew-t distribution[4]. Here, ξ is location, w scale parameters and γ skew parameters. And again we keep notations unchanged as in the original study.

$$Y \sim St(\xi, w^2, \gamma) \tag{46}$$

$$Y = \xi + wX \tag{47}$$

Probability distribution function of X is shown in Eq. (48) where υ is degrees of freedom, Γ is a gamma function and Φ is a cumulative t-distribution function.

$$f(x) = 2\phi(x)\Phi(x) \tag{48}$$

$$\phi(x) = \frac{\Gamma(\frac{\upsilon+1}{2})}{\sqrt{\upsilon\pi}\Gamma(\frac{\upsilon}{2})}(1+\frac{t^2}{\upsilon})^{-\frac{\upsilon+1}{2}} \tag{49}$$

We fitted skew-t distribution into observed residuals ($v_{i,t}$, $u_{i,t}$) from a simple linear regression and estimated all three parameters by Maximum Likelihood method. Our objective is to use γ and ξ to compare with θ and μ from GLAM.

However, due to a singularity problem (Azzalini, A., 2013) of information matrix, we used centralized parameters rather than direct parameters and estimated location ξ and skewness γ. Comparison of different parameters is beyond the scope of this research.

[4] http://azzalini.stat.unipd.it/SN/Intro/intro.html

5 Empirical results

5.1 β

β in a simple linear regression (1) is estimated by LS and R methods for N225, FTSE100 and S&P500 stock returns. β is estimated for non-overlapping 79 quarterly windows. Average $\hat{\beta}$ across Nikkei 225 stocks presented in Table (2). Thus, R and LS produce distinctive $\hat{\beta}$ as well as standard deviations, minimum and maximum values. Similarly, Tables (3) and (4) present descriptive statistics of β for FTSE100 and S&P500 stocks. This overall averages do not provide much insight. But, Figures (1) - (3) illustrate quarterly average β over time period of each sample.

Tables (5) and (6) present descriptive statistics of estimated β by R and LS methods for a sample of 6 Japanese stocks from various industries. Two approaches estimated comparable β, nonetheless, discrepancy is clear and supports previous result in Table (2). Especially standard deviation of β from R approach are smaller than its counterpart for most of the cases. Depending on terms, estimated β is low as -0.001 or high as 2.2. However, this behavior is different depending on stocks. A possible explanation for this variation in β is the nature of industry where companies belong.

Table 2: Average β of N225, Q1 1998 - Q3 2017

Statistic	N	Mean	St. Dev.	Min	Max
R	79	0.939	0.094	0.663	1.150
LS	79	0.946	0.092	0.700	1.138

Table 3: FTSE100, Q1 1998 - Q3 2017

Statistic	N	Mean	St. Dev.	Min	Max
R	79	0.860	0.166	0.417	1.149
LS	79	0.875	0.166	0.433	1.183

Table 4: S&P500, Q1 1998 - Q3 2017

Statistic	N	Mean	St. Dev.	Min	Max
R	79	0.993	0.137	0.597	1.269
LS	79	0.999	0.135	0.579	1.271

We can observe this nature of β by looking at the stocks one by one for each time period, but lack of a statistical method to capture an overall image will not allow us except conditioning or restricting analysis by industry-wise. So we randomly choose a widely known company stock and present results. Results for other stocks are available upon request.

Quarterly estimated β for Canon stocks illustrated in Figure (4). In 1999, Canon stock behaved quite distinctly than the rest of the market as it is clear from a very low β. Especially, during the end of 2000 Canon β was fluctuating and hit the highest peak for the last 20 years period of time. From 2005 until 2009, β has increasing trend in a small range, nonetheless, European Sovereign Debt crisis in 2011 possibly caused stocks to plummet sharply in 2011 - 2012. Afterwards, starting from 2013 Canon experienced less volatile and smaller β until the end of data period. This non constant behavior of β is in line with previous studies (Jagannnathan and Wang (1996), Lewellen and Nagel (2006)) in contrary to the static CAPM. LS and R estimates are comparable and the divergence is minimal. Notably, for 2002 and 2011 LS estimate β are quite different than its counterpart R estimate β.

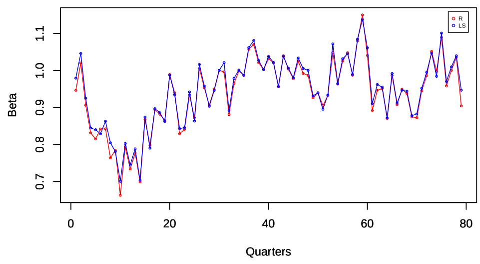

Figure 1: Quarterly average beta, N225

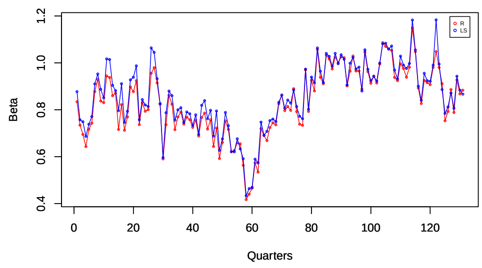

Figure 2: Quarterly average beta, FTSE100

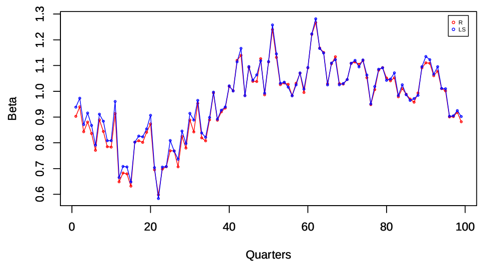

Figure 3: Quarterly average beta, S&P500

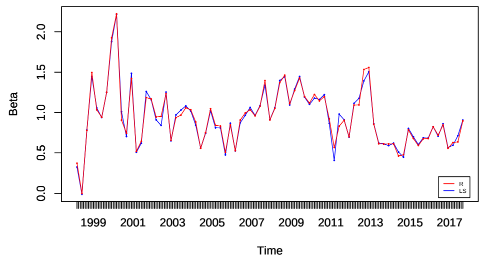

Figure 4: Beta Canon Inc

Table 5: Descriptive statistics of β, R method

Statistic	N	Mean	St. Dev.	Min	Max
Toyota Motor Corp	79	0.913	0.234	0.303	1.519
Taisei Corp	79	0.894	0.324	0.127	2.006
Takashimaya Co	79	0.911	0.295	0.238	1.726
Nippon Express Co Ltd	79	0.814	0.239	0.093	1.267
Canon Inc	79	0.937	0.353	−0.001	2.222
Mitsubishi Corp	79	1.182	0.243	0.399	1.713

Table 6: Descriptive statistics of β, LS method

Statistic	N	Mean	St. Dev.	Min	Max
Toyota Motor Corp	79	0.917	0.232	0.366	1.489
Taisei Corp	79	0.907	0.345	0.126	2.070
Takashimaya Co	79	0.909	0.296	0.258	1.808
Nippon Express Co Ltd	79	0.822	0.247	0.141	1.334
Canon Inc	79	0.934	0.353	−0.012	2.215
Mitsubishi Corp	79	1.192	0.242	0.420	1.667

Cross sectional analysis of β

In order to get a deeper intuition regarding our estimates we look into cross sectional distribution of β for a chosen quarter. Fig. (5) - (6) present histograms for 2008 Q2 and 2017 Q3, for R and LS cases, respectively. Obviously, estimates are different during crisis and relatively peaceful periods in market. LS histograms have fat tails and more width. In contrast, R β histograms display slightly centralized distribution and it is stronger for Q3 in 2017.

Fig (9) - (10) illustrate the cross sectional scatter plots of two distinct β for the same quarter as shown in previous histograms. Clearly, β form stronger similarity in Q2 of 2008 than Q3 in 2017.

Fig. (7) displays β difference between LS and R estimates. Histogram clearly illustrates the persistent discrepancy between β across all N225 stocks in Q2 2008. Some of the stocks have a significantly distinct β estimates. This is more obvious in Q3 2017 in Fig. (8). Maximum and minimum of β difference is significantly bigger than estimates in crisis period. A possible explanation for this lies in the fundamental variety of LS and R methods. In brief, during the volatile market, LS and R β are at similar level, contrary to less volatile period estimates.

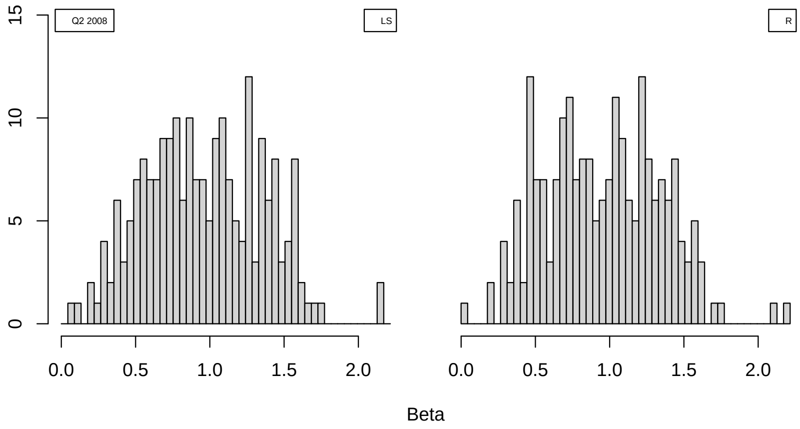

Figure 5: Cross sectional beta, N225, Q2 2008

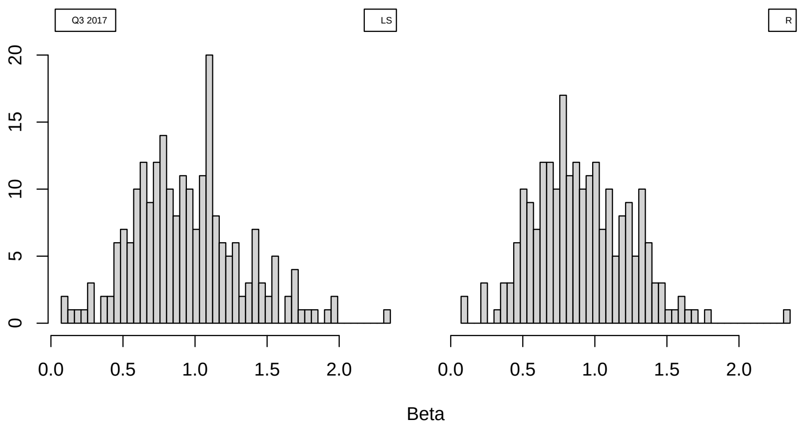

Figure 6: Cross sectional beta, N225, Q3 2017

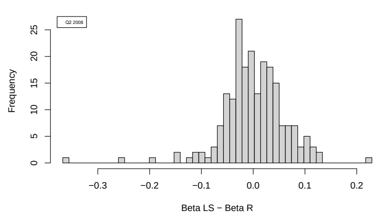

Figure 7: Beta difference, N225, Q2 2008

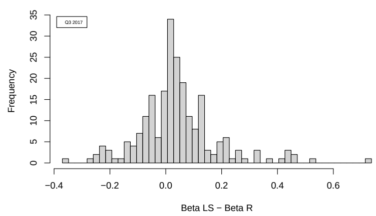

Figure 8: Beta difference, N225, Q3 2017

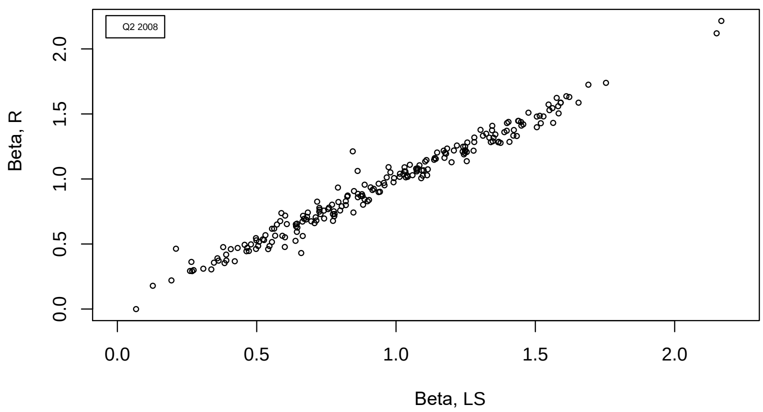

Figure 9: Cross sectional beta scatter plot, Q2 2008

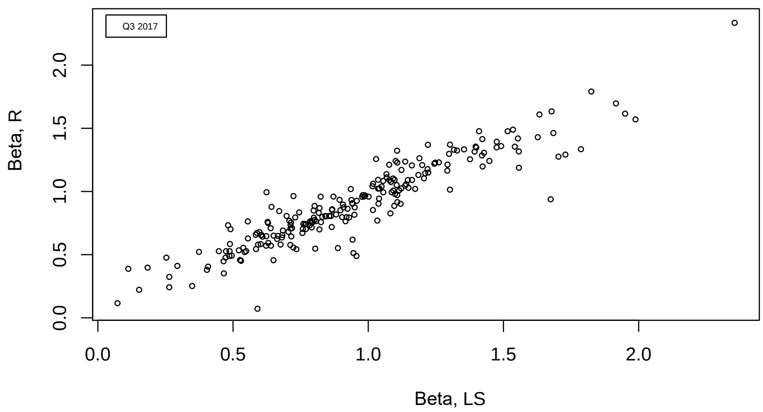

Figure 10: Cross sectional beta scatter plot, Q3 2017

5.2 Shape of residual distribution

In this subsection, we start to investigate the residuals from a simple linear regression. Fig. (83), (84), (85) and (86) in Appendix illustrate histogram of residuals from LS and R methods for Canon stocks as an example. Similar behavior of residuals could be observed for other company stocks as well. As it is illustrated, histograms have a noticeable skewness and have heavy tails.

In addition, Table (7) presents average skewness and kurtosis of residual distribution for N225 stocks. Normal distribution has 0.03 skewness and 2.96 kurtosis which verifies a symmetry of distribution. However, average skewness and kurtosis among N225 stocks are far from being close to normal distribution. Tables (8) - (9) in Appendix 8.6 present average skewness and kurtosis of residual distribution for stocks in USA and UK market, respectively.

Table 7: Descriptive statistics of average skewness and kurtosis, N225

Statistic	N	Mean	St. Dev.	Min	Max
Skewness LS	79	0.262	0.157	−0.054	0.626
Skewness R	79	0.267	0.161	−0.058	0.646
Kurtosis LS	79	1.693	0.804	0.310	4.221
Kurtosis R	79	1.819	0.880	0.353	4.791

5.3 μ and θ

Relying on derived score functions for logistic distribution and statistics for parameter inference, GLAM parameters θ and μ are estimated. The following Fig. (11) and (12) illustrate estimates for θ and μ for Canon stocks.

θ for Canon stocks has a significant fluctuation during the sample period. Values are higher than one for most of the observation and fluctuation becomes wider from 2006 until 2009. This exceptional variation could be a possible reaction of Canon stock prices to financial market distress around 2008. Interestingly, θ behavior changed after 2011, however, from 2017 it revives noticeable fluctuations.

μ shows similar pattern. Fluctuations in a narrow corridor is followed by movements in wide range during 2006 and 2009. Especially, in 2009 μ plummets to the lowest points twice in a year and decline is obviously the effect of stagnation and downfall in financial markets occurred in 2009.

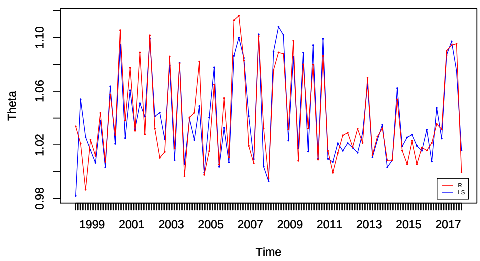

Figure 11: θ Canon Inc

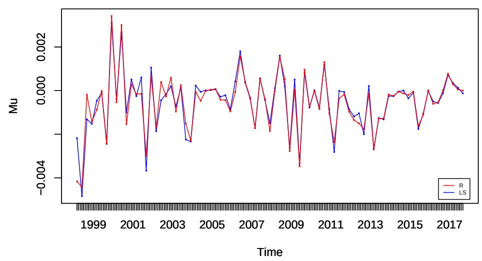

Figure 12: μ Canon Inc

Following Fig. (13) and (14) depict estimated θ and μ for Mitsubishi Corp., respectively.

θ for Mitsubishi stocks also has a significant fluctuation throughout the sample period. However, before 2007 variation usually happens in a smaller range and some quarters have quite low θ estimates. Extreme fluctuation is persistent and periodic, especially after 2007 and a similar behavior is observed until the end of observation period.

μ shows similar pattern with the case of Canon. High variation is observed only from 1998 until 2003. On the contrary, estimated parameter exhibits a clear increasing trend prior to the crisis in 2008 - 2009. Afterwards, μ only has a fluctuation in a narrow range.

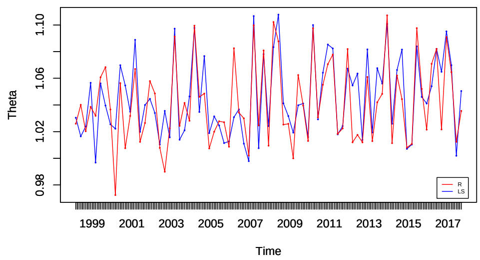

Figure 13: θ Mitsubishi Corp.

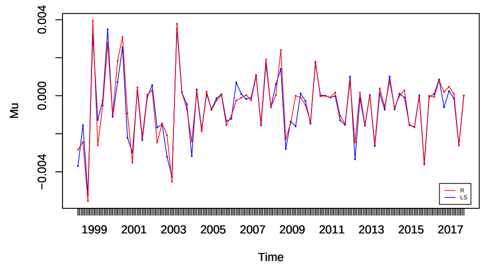

Figure 14: μ Mitsubishi Corp

As such each individual stock has different behavior of their estimated parameters, but the time period at which the behavior changes seem common to all the stocks.

Tables (10) and (11) in Appendix 8.6 present descriptive statistics of estimated θ for 6 Japanese stocks based on R and LS approaches, respectively. Mean values of θ clearly indicate that on average residuals have unsymmetrical distribution and θ is higher than 1 throughout our sample period.

In addition, Tables (12) and (13) in Appendix 8.6 present descriptive statistics for estimated μ parameters for 6 Japanese stocks based on R and LS, respectively. Clearly, mean values of μ are small, around -0.001.

Results support our expectations meaning that a simple linear regression residuals are unsymmetrical. Tables (14) and (19) in Appendix 8.6 present descriptive statistics of μ and θ across N225 stocks. Obviously, results are not different from the case of 6 stocks, such as μ parameter is -0.001 and θ is 1.039.

Cross sectional analysis of μ and θ

Distribution of estimated μ across 225 stocks are presented below in Fig. (15) - (16) for LS and R residuals, respectively. Starting with Fig. (15), in 2008 μ has a noticeable left skewed shape for both approaches, nonetheless, this nature is weak in 2015. In addition, the range of estimated μ is slightly larger for R case.

Similar skewed distribution is observed for θ across 225 stocks as well, but to the right side as illustrated in Fig. (17) - (18). In addition, θ values noticeably form two clusterings around 1 and 1.05 in Q2 2008. This behavior is still weakly persistent in Q1 2015, especially in θ from R residuals.

Scatter plots of θ estimated from LS and R residuals are illustrated in Fig. (21) - (22). Clearly, residuals from both approaches yield similar θ values. In comparison, Fig. (19) - (20) display estimated μ from LS and R residuals. Relation is stronger than θ case and plots are similar for both quarters from 2008 and 2015.

In markets during financially stressful periods abnormal behavior in stock prices could be observed. As our findings for θ and μ depicted, this nature of stocks is persists in residuals and it is not explained by market excess return in a simple linear regression. Thus, asymmetry in residual distribution caused by irregularity in stock return could leave traditional results in doubt. Moreover, company specific and industry related factors are possible drivers of unsymmetrical and non-normal shape of error term distribution, and GLAM accurately captures those factors in stock returns.

Fig. (23) and (24) display differences of θs and μs estimated from LS and R residuals. Histogram clearly supports the notion that both approaches deliver distinct residuals and this discrepancy is consistent across 225 stocks in Q2 2008. Some of the stocks have a significantly diverse θ and μ estimates, e.g., -0.08 (far left side of Fig. (23)).

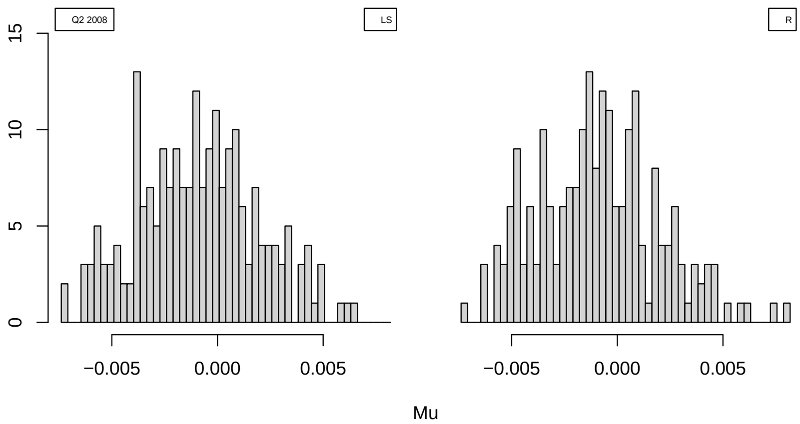

Figure 15: Cross sectional μ, N225, Q2 2008

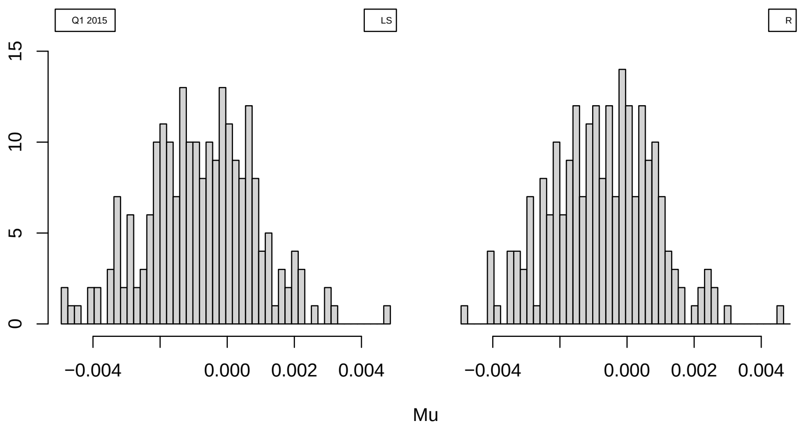

Figure 16: Cross sectional μ, N225, Q1 2015

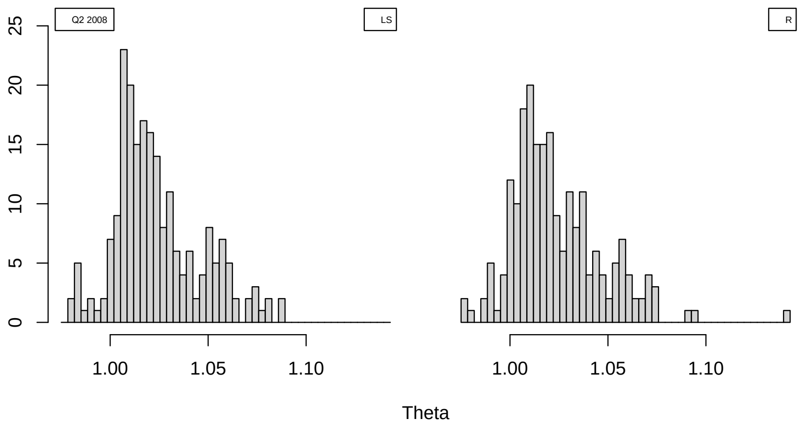

Figure 17: Cross sectional θ, N225, Q2 2008

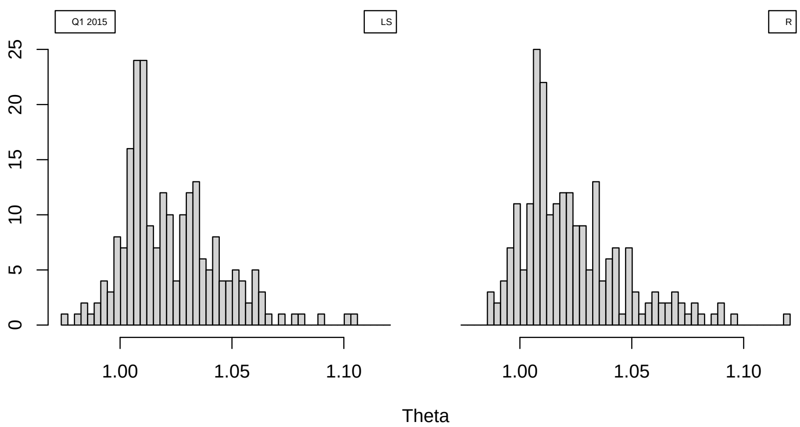

Figure 18: Cross sectional θ, N225, Q1 2015

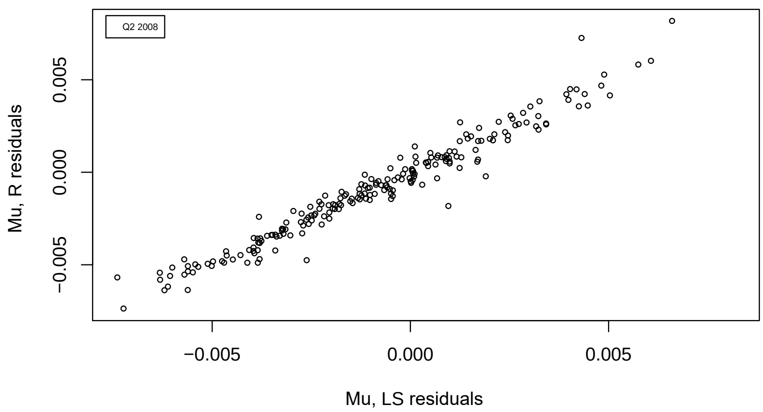

Figure 19: Cross sectional μ scatter plot, Q2 2008

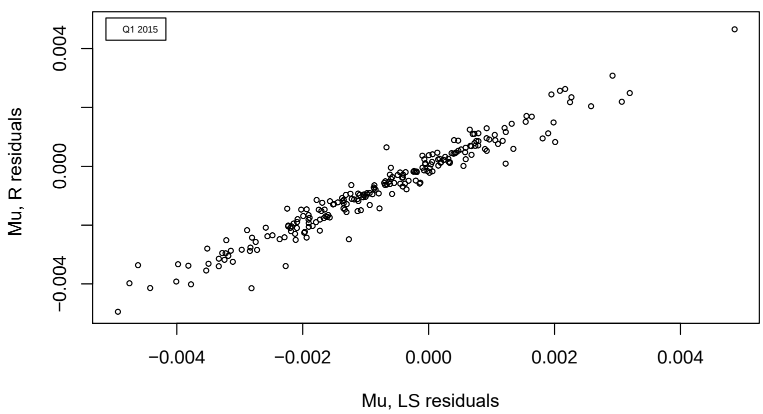

Figure 20: Cross sectional μ scatter plot, Q1 2015

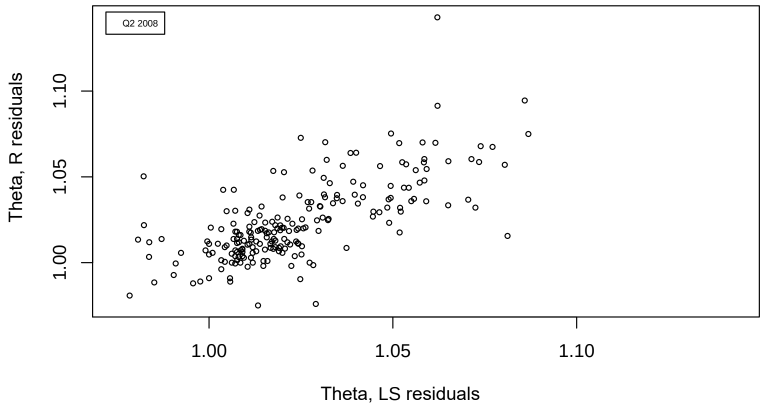

Figure 21: Cross sectional θ scatter plot, Q2 2008

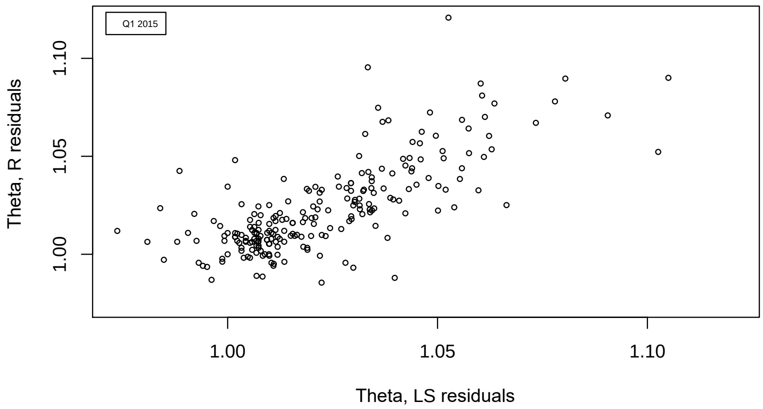

Figure 22: Cross sectional θ scatter plot, Q1 2015

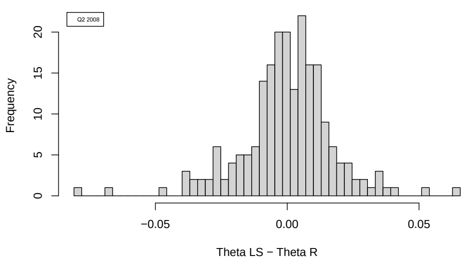

Figure 23: θ difference, N225, Q2 2008

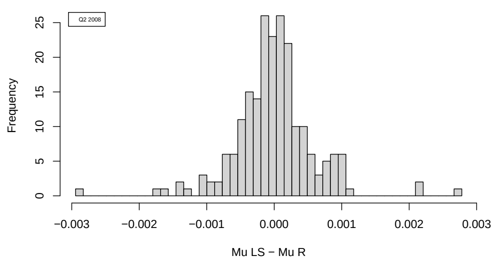

Figure 24: μ difference, N225, Q2 2008

β and θ

Fig. (25) illustrate difference of estimated β for Canon stocks. Plot has no pronounced time trend, however, the magnitude of contrast is quite significant. Especially, in 2011 and 2013 the divergence of two β is noticeable. Similar plot for θ parameter is presented in Fig. (26). θ difference also has obscure trend by time but the range of fluctuation decays gradually.

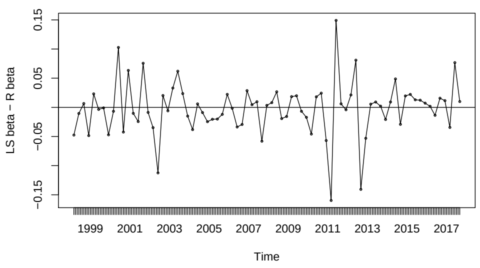

Figure 25: β difference, Canon Inc

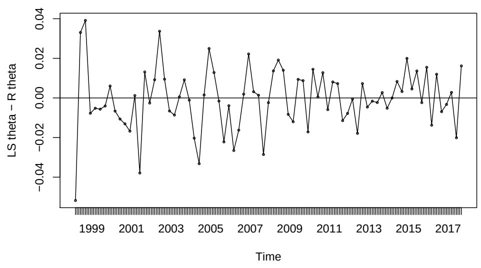

Figure 26: θ difference, Canon Inc

Moreover, β and θ do not show any sign of correlation as illustrated by Fig. (27) and (28). Observed θ values form two distinct clusters with mean being lower and higher than one. However, this behavior of θ is not related to β.

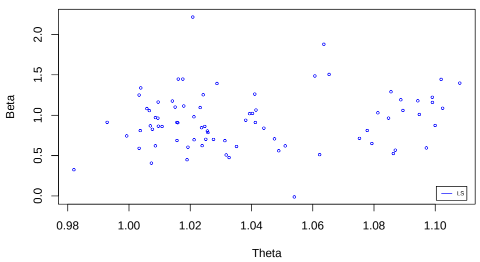

Figure 27: β and θ, Canon Inc

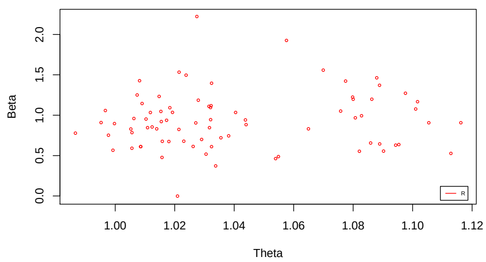

Figure 28: β and θ, Canon Inc

5.4 Skew-t distribution results

Following the estimation of θ and μ, we fitted skew-t distribution into observed residuals ($v_{i,t}$,$u_{i,t}$) and estimated w, γ and ξ. As as example Fig. (29) and (30) illustrate estimated parameters after fitting into residuals for Canon stocks.

Skew-t distribution's ξ parameter represents the location of the residual distribution and ξ should be comparable with μ from GLAM. μ and ξ share a similar path in the beginning of the period with high fluctuations. However, ξ plummets significantly in 2009, while μ shows only high fluctuations (Fig. (12)). In addition, μ varies in the range of -0.002 and 0.002, but ξ has a range of -0.004 and 0.004 which is almost two times wider. This obviously shows the fundamental difference of both approaches to model error terms from a simple linear regression.

Shape parameter in Fig. (30) has a distinct behavior. Initially, γ fluctuates in a small range but later reaches the highest point in 2009 and the lowest in 2016. Interestingly, for some periods γ is zero which means that residual distribution has skewness on neither side and has a symmetrical form. However, θ from GLAM in Fig. (11) fluctuates quite noticeably during the time period with no sign of symmetricalness. Once again this could be due to a fundamental difference inherited into two approaches.

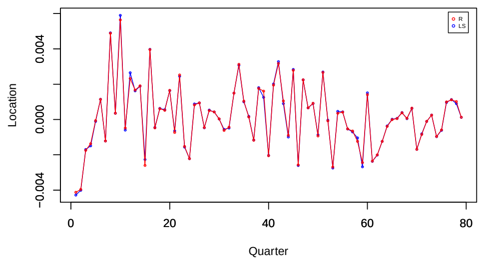

Figure 29: Location estimate ξ, Canon Inc

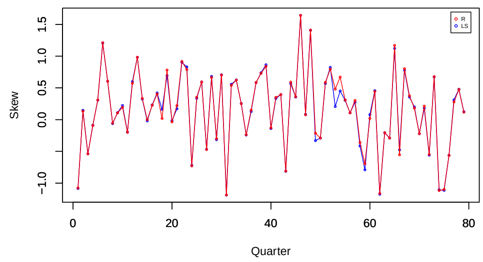

Figure 30: Skew estimate γ, Canon Inc

Cross sectional distribution of estimated 225 γs for Q2 in 2008 are displayed in Fig. (31) and (32), for R and LS residuals, respectively. Both histograms illustrate a similar distribution of skewness parameters among 225 stocks. Moreover, γ forms two clusterings, one is more negative and the other on a positive side, and it is stronger in case of LS residuals.

This is a possible indication that for some stocks residuals are left skewed and for others residuals are right skewed, and it in concordance with our previous θ results. Similar behavior is observed for other periods as well, such as in the Q1 of 2015 in Fig. (33) and (34) which show the histogram of residuals. We choose to present findings for γ only for crisis and relatively peaceful periods, nonetheless, result for the rest of the time period is available upon request.

Moreover, scatter plots of estimated γs for the same quarter as in histogram are illustrated in Fig. (35) and (36). Clearly, γ estimated from both methods (R and LS) are very similar as depicted in Figures.

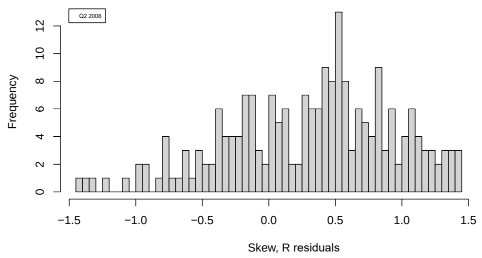

Figure 31: Cross sectional γ, Q2 2008

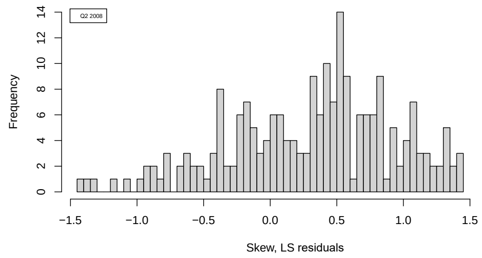

Figure 32: Cross sectional γ, Q2 2008

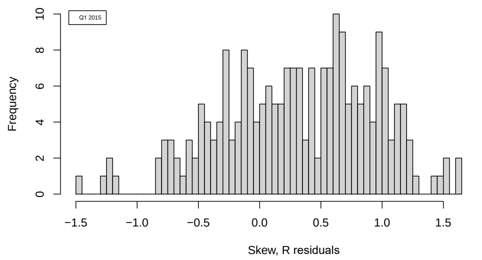

Figure 33: Cross sectional γ, Q1 2015

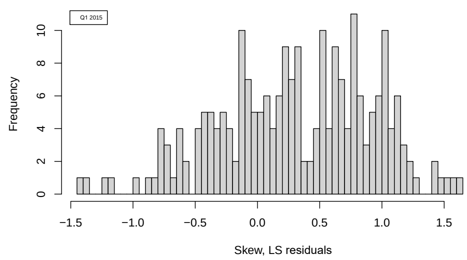

Figure 34: Cross sectional γ, Q1 2015

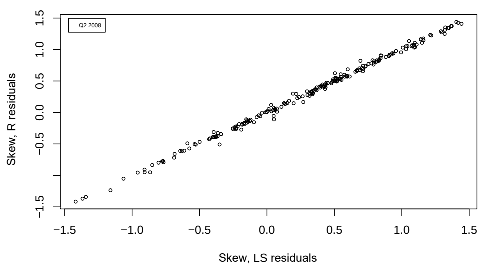

Figure 35: γ scatter, Q2 2008

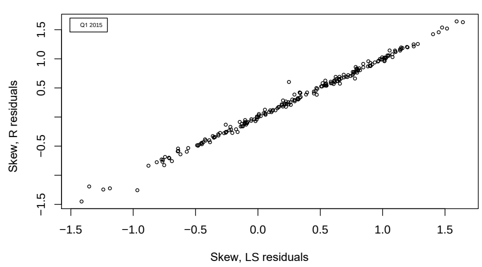

Figure 36: γ scatter, Q1 2015

5.5 Jensen's Alpha decomposition

Skew effect

β in a linear regression captures the sensitiveness of excess return to excess market return and in line with CAPM. Nonetheless, a growing number of papers analyze the rate of return with inclusion of intercept term in the regression known as "Jensen's Alpha" and introduced by M. Jensen (1968).

$$\alpha_i = E[R_i - R_f - \beta_i(R_m - R_f)] \tag{50}$$

Difference of α and location parameter μ gives a skew effect as shown below.

$$\alpha_i = E[\eta] = \mu_i + E[\epsilon] = \mu_i + \int_{-\infty}^{\infty} xdh(F(x) : \theta) \tag{51}$$

$$\alpha_i - \mu_i = \int_{-\infty}^{\infty} xdh(F(x) : \theta) \tag{52}$$

Fig. (37) - (40) illustrate the cross sectional distribution of skew-effect for different quarters. In 2005 Q2, histograms are centered between 0 and 0.001, and has a fat tails on the right side. Skew effect from R and LS do not differ significantly and has a very similar shape of distribution. However, in 2008 skew effects are quite distinct and R case has a noticeable right tail.

Besides, Fig. (41) - (42) illustrate scatter plots of skew effect based on μ from GLAM. Clearly, skew effect derived based on μ from LS and R residuals, are close to each other as shown in plots

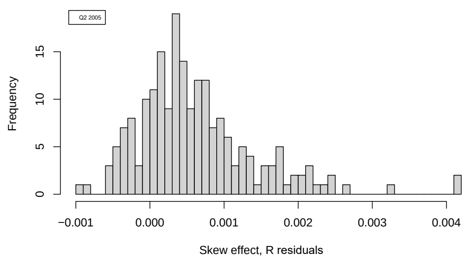

Figure 37: skew effect ($\alpha_i - \mu_i$), Q2 2005

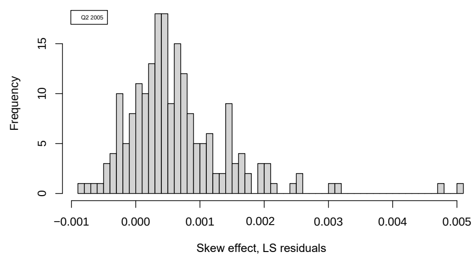

Figure 38: skew effect ($\alpha_i - \mu_i$), Q2 2005

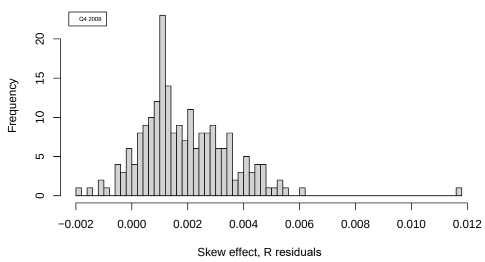

Figure 39: skew effect ($\alpha_i - \mu_i$), Q4 2008

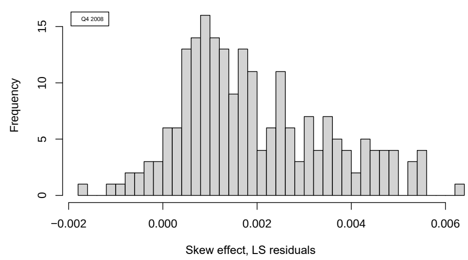

Figure 40: skew effect ($\alpha_i - \mu_i$), Q4 2008

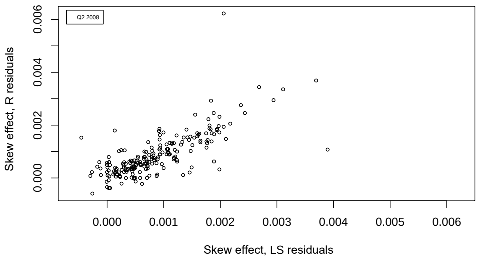

Figure 41: skew effect scatter, Q2 2008

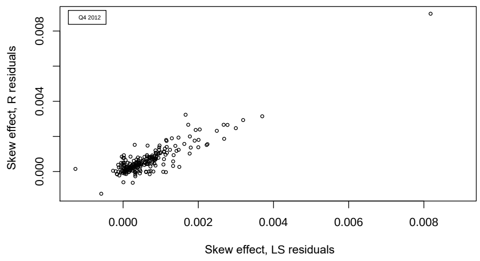

Figure 42: skew effect scatter, Q4 2012

In addition, skew effect based on skew-t distribution's location parameter as well in Eq. (55). Fig. (43) - (46) illustrate cross sectional skew effect obtained by subtracting skew-t location parameter μ from α. In Q2 2005, skew effects from LS and R are centered around 0 as well as have a similar shape. In comparison, skew effect based on skew-t location parameter has a smaller magnitude than GLAM counterpart but still it has a fat right tail. In 2008 Q4 skew effect has more balanced distribution than Fig. (39) and (40). Possible explanation is the intrinsic difference of GLAM and skew-t to capture the location parameter. GLAM seems to capture the location more accurately and has asymmetrical skew effect during the crisis time.

Moreover, Fig. (47) - (48) display scatter plots of skew effect derived based on ξ from skew-t distribution. In comparison with skew effect based on μ from GLAM, scatter plots show strong relation of skew effect obtained based on LS and R residuals.

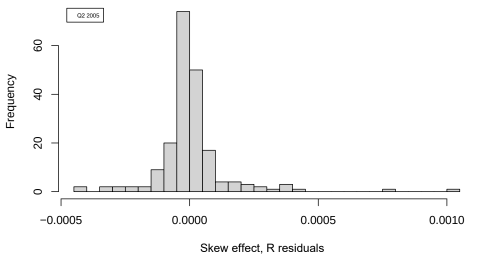

Figure 43: skew effect $(\alpha_i - \xi_i)$, Q2 2005

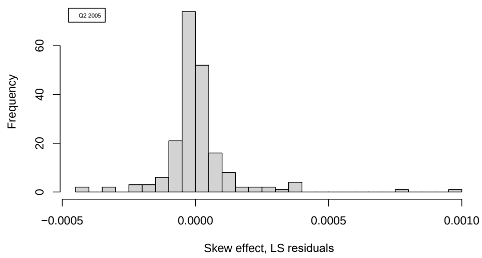

Figure 44: skew effect $(\alpha_i - \xi_i)$, Q2 2005

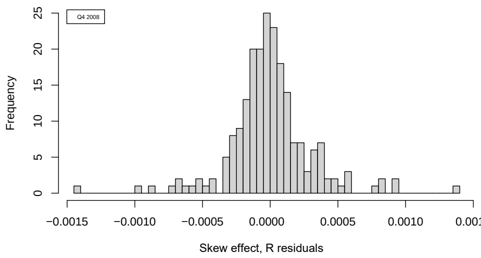

Figure 45: skew effect $(\alpha_i - \xi_i)$, Q4 2008

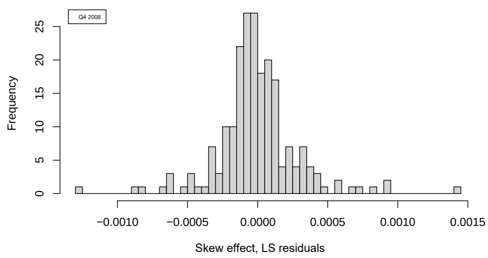

Figure 46: skew effect $(\alpha_i - \xi_i)$, Q4 2008

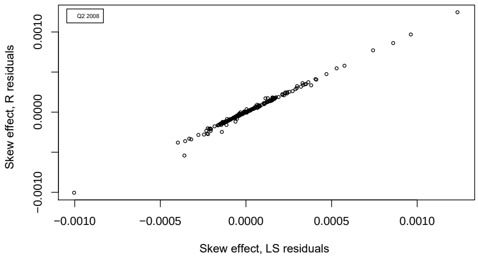

Figure 47: Skew effect scatter, Q2 2008

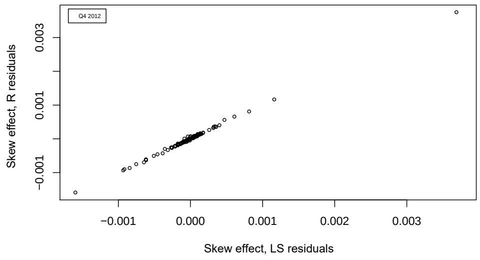

Figure 48: Skew effect scatter, Q4 2012

Skew effect regression on asymmetry

In order to have a comparison between GLAM and skew-t, skew effects regressed onto asymmetry parameters as presented in Eq. (54) and (56), respectively.

Here, $i = \{1...225\}$ stocks and $q = \{1..79\}$ quarters.

$$skew\ effect^{glam}_{i,q} = \alpha_{i,q} - \mu^{glam}_{i,q} \tag{53}$$

$$skew\ effect^{glam}_{i,q} = \kappa_0 + \kappa_1\theta_{i,q} + \epsilon_{i,q} \tag{54}$$

$$skew\ effect^{skew-t}_{i,q} = \alpha_{i,q} - \xi^{skew-t}_{i,q} \tag{55}$$

$$skew\ effect^{skew-t}_{i,q} = \kappa^*_0 + \kappa^*_1\gamma_{i,q} + \epsilon_{i,q} \tag{56}$$

Here, κ_1 and κ^*_1 is the sensitiveness of skew-effect on asymmetry parameter θ and γ. Fig. (52) and (54) illustrate estimated κ_1 and κ^*_1 for both R and LS cases. Time period is given in quarters from Q1 1998 until Q3 2017. Noticeably, both κs have a completely different path and magnitude, due to the fact skew-effects are different in Eq. (53) and (55).

For instance, regression result for Q1 2009 is presented below in Eq. (57) (t - stats are given in parenthesis) for Japanese stocks. For this regression only considered asymmetry and location parameters that are obtained from R residuals. Clearly, when θ is equal to 1 which means symmetry, skew effect is almost zero for GLAM case (κ_0 and κ_1 cancel each other). Similar relation between skew effect and θ could be observed for other quarters as well.

Moreover, regression result for Q1 2009 presented below in Eq. (58) for USA stocks as well. In comparison with N225 stocks, S&P500 stocks' skew effect are less sensitive to asymmetry (0.0219). One possible explanation is high liquidity levels of S&P500 stocks. However, similarly to N225 stocks, S&P500 stocks do not experience skew effect when θ is equal to 1 which means the intercept and coefficient sum up to zero.

$$\begin{aligned} skew\ effect^{glam,jp}_{i} &= -0.0335 + 0.0333 * \theta_i + \epsilon_i \\ & \quad\ (-13.90) \quad (14.47) \end{aligned} \tag{57}$$

$$\begin{aligned} skew\ effect^{glam,usa}_{i} &= -0.0197 + 0.0219 * \theta_i + \epsilon_i \\ & \quad\ (-1.97) \quad (2.26) \end{aligned} \tag{58}$$

In comparison, below in Eq. (59) (t - stats are given in parenthesis) is presented regression result for Q1 2009 for skew-t case for Japanese stocks. For this regression only considered asymmetry and location parameters that are obtained from R residuals. Assuming symmetrical error term distribution, skew effect is equal to the sum of κ^*_0 and $\kappa^*_1 * 0$. More importantly, skew-t's γ parameter does not explain skew effect as shown in Eq. (59), κ_1^* is insignificant, in comparison to Eq. (57), κ_1 is statistically significant.

$$\begin{aligned} skew\ effect_i^{skew-t} = -0.0000123 - 0.0000139 * \gamma_i + \epsilon_i \\ (-0.76) \qquad (-0.54) \end{aligned} \tag{59}$$

Besides, comparison of both approaches (GLAM and skew-t parameters) based on p -values of κ_1 and κ_1^* from quarterly regressions' results reveals that θ explains skew-effect in all quarters across our data time span (49). Skewness parameter of skew-t fails to explain skew-effect in most of the quarters and could not reject the null hypothesis that κ_1^* is equal to 0 (50).

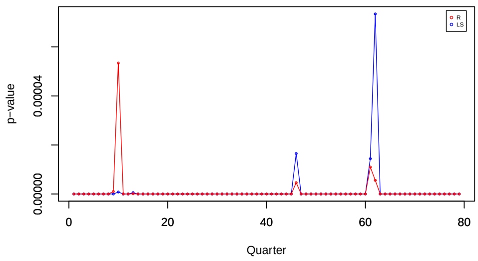

Figure 49: P-values of κ_1, GLAM

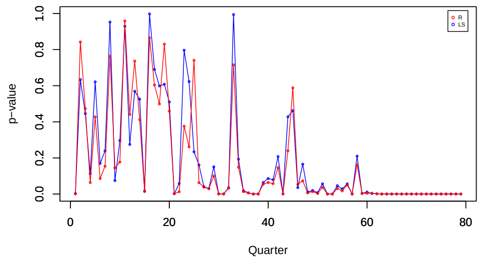

Figure 50: P-values of κ_1, Skew-t

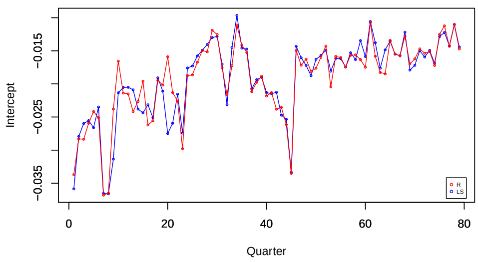

Figure 51: κ_0, GLAM

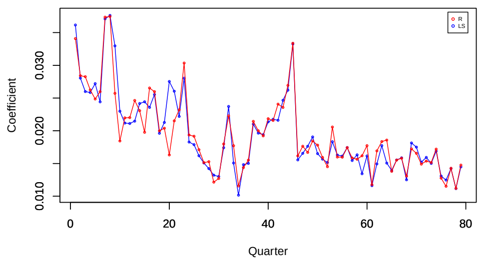

Figure 52: κ_1, GLAM

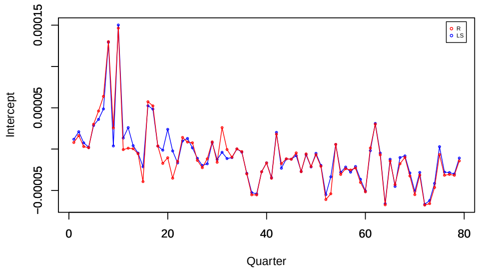

Figure 53: κ_0, Skew-t

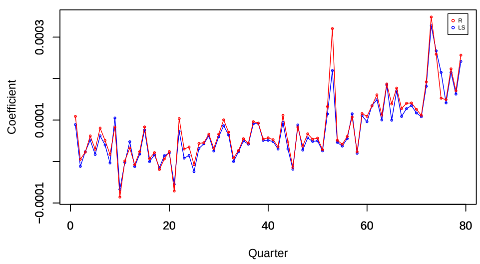

Figure 54: κ_1, Skew-t

α decomposition

Following the GLAM and skew-t distribution fit, we estimated location and asymmetry parameters for each quarters and markets, respectively.

Jensen's Alpha obtained as in Eq.s (60) and (61) (Jensen, 1968).

$$\alpha_i^{LS} = E[R_i - R_f - \beta_i^{LS}(R_m - R_f)] \tag{60}$$

$$\alpha_i^{R} = E[R_i - R_f - \beta_i^{R}(R_m - R_f)] \tag{61}$$

Next, we regressed α from a simple linear regression on μ, θ, ξ and γ for every quarter separately to analyze if α is explained by location and asymmetry of residual distribution as in Eq. (62) and (63).

$$\alpha_{i,q} = \kappa_0 + \kappa_1\mu_{i,q} + \kappa_2\theta_{i,q} + \epsilon_{i,q} \tag{62}$$

$$\alpha_{i,q} = \kappa_0^* + \kappa_1^*\xi_{i,q} + \kappa_2^*\gamma_{i,q} + \epsilon_{i,q}^* \tag{63}$$

Here, $i = \{1, ..., 225\}$ stocks and $q = \{1, ..., 79\}$ quarters.

Fig. (55), (56) and (57) illustrate κ_0, κ_1 and κ_2, respectively, for Japanese stocks. Starting with κ_1 in Fig. (56), estimated coefficient for location parameter μ fluctuates noticeably around one and this result is in line with the study of Jensen (1968). *Jensen's Alpha* is equal to expected value of error terms from a simple linear regression in Eq. (2) ($\alpha_i = E[\eta_i]$).

In Fig. (55) and (57), κ_0 and κ_2 have upward and downward sloping path, respectively, and obviously coefficients have a negative correlation. As figures illustrate, during the crisis period in 2008, α was quite sensitive to θ than other periods.

As an example, regression results for Q1 2009 presented in Eq. (64) for N225 stocks. For this regression only considered asymmetry and location parameters that are obtained from R residuals. Assuming no asymmetry ($\theta = 1$) in error term distribution from a simple linear regression, κ_0 and κ_2 sum up to zero and *Jensen's Alpha* is only equal to $1.0229 * \mu$. Similar relation between α and θ could be observed for other quarters as well.

Regression results for Q1 2009 presented in Eq. (65) for S&P500 stocks as well. κ_0 and κ_2 sum up to zero in case of symmetry and α is only equal to $0.6484 * \mu$. Clearly, α is less sensitive to μ for SP500 than N225 case. This is possible due to the difference in nature of US stocks and trading behavior of market participants.

$$\begin{aligned} \alpha_i^{jp} = {} & -0.0325 + 1.0229 * \mu_i + 0.0325 * \theta_i + \epsilon_i \\ & (-11.91) \quad (32.27) \quad (12.49) \end{aligned} \tag{64}$$

$$\begin{aligned} \alpha_i^{usa} = {} & -0.0169 + 0.6484 * \mu_i + 0.0185 * \theta_i + \epsilon_i \\ & (-1.986) \quad (17.06) \quad (2.24) \end{aligned} \tag{65}$$

Fig. (58), (59) and (60) illustrate κ_0^*, κ_1^* and κ_2^*, respectively, for the case of skew-t and for Japanese stocks. Similarly, Eq. (66) presents regression result for Q1 2009 for N225 stocks and this regression only considered asymmetry and location parameters that are obtained from R residuals. Skew-t's γ parameter does not explain α as shown in Eq. (66), κ_2^* is insignificant.

$$\begin{aligned} \alpha_i = {} & -0.000012 + 1.0093 * \xi_i - 0.000036 * \gamma_i + \epsilon_i^* \\ & (-0.73) \quad (215.49) \quad (-1.30) \end{aligned} \tag{66}$$

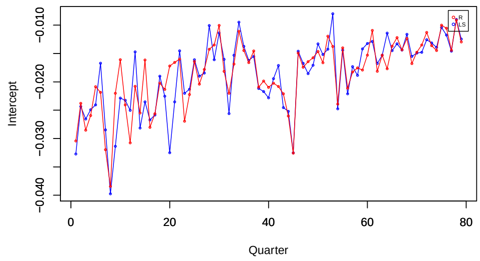

Figure 55: κ_0, GLAM

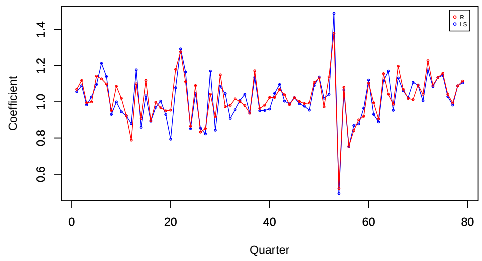

Figure 56: κ_1, GLAM

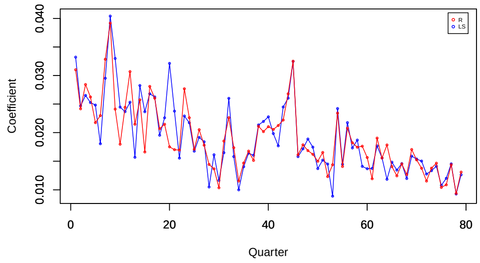

Figure 57: κ_2, GLAM

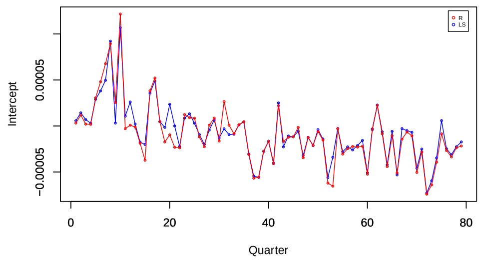

Figure 58: κ_0, Skew-t

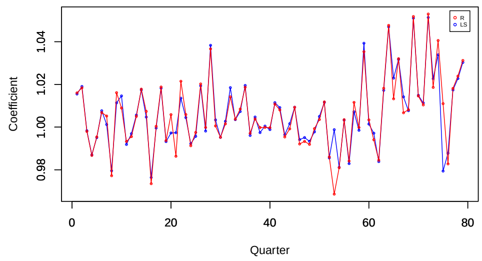

Figure 59: κ_1, Skew-t

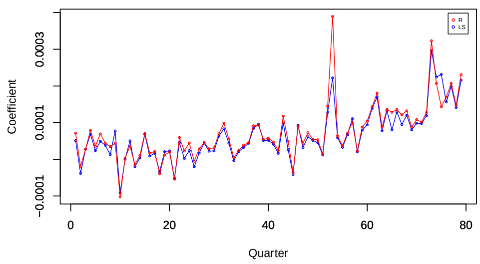

Figure 60: κ_2, Skew-t

Following the regression result we can conclude that GLAM parameters μ and θ are superior to skew-t distribution parameters to decompose the total exposure to α into location and asymmetry.

6 F distribution and relative efficiency

6.1 Estimation of F distribution

In this section we estimate F which is assumed to be a symmetric around zero but to have unknown form of distribution. Z_i values are obtained based on optimal $\hat{\theta}$ estimated in section 4. The proof applies the convergence arguments for estimated empirical distribution functions in section 5 of chapter 5 of Shorack and Wellner (1986).

The advantage of GLAM is that we can estimate μ and θ without knowing the functional form of F. However, in order to see the accuracy of estimators of β and a comparison of asymptotic variance of estimation error of these two methodology: LS and R-estimator, the functional form of distribution F and the density f are required. The role of F in GLAM model includes, as well as its symmetry, a representation of dispersion of the underlying distribution in the tails of F, while the transformation function $h(F;\theta)$ only represents how and how much the underlying distribution F is skewed/asymmetric to fit to the distribution of observed residuals. Thus we choose t-distribution which degree of freedom parametrize the tail-heaviness and unimodal symmetric shape of distribution ranging from Normal to almost close to Cauchy.

As indicated in M&T (1993) the empirical distribution function of $Z_i(r:\beta_0)$ approximates asymptotically the empirical distribution function of $Z_i(\theta_0:\beta_0)$ for $i = 1, 2, ..., n$.

That is, $Z_i(r:\hat{\beta}_n)$ also asymptotically approximate the $i-th$ order statistics of

$$e_1(\beta_0), e_2(\beta_0), ..., e_n(\beta_0) \quad i.i.d \sim G(x-\mu) \equiv h(F(x-\mu):\theta_0) \tag{67}$$

Note that $e_i(\beta) = \eta_i$

We will prove here that the empirical distribution function of $Z_i(\hat{\theta}_n : \hat{\beta}_n)$ estimates the underlying unknown distribution function F asymptotically.

Now denote the empirical distribution function of $Z_i(r:\beta)$.

$$L_{n,r}(x:\beta) = \frac{1}{n}\sum_{i=1}^{n} I\{Z_i(r:\beta) \leq x\} \tag{68}$$

Proposition A-2

$$\sqrt{n}\{L_{n,\hat{\theta}_n}(x:\hat{\beta}_n) - F(x-\mu))\} \tag{69}$$

converges in distribution to a limit random variable LEF with $N(0,\sigma_F^2)$, as $n \to \infty$.

Proof for A-2 is given in Appendix 8.5.

For each quarter and each stock, we fitted t-distribution to a set of Z_i (with estimated θ). Then we found that the estimated degree of freedom varies cross-sectionally from as small as 3 or 4 to as large as 40 and to 100-120 in every quarter during the year 1998-2017.

After obtaining Z_i values we fitted t distribution and obtained degrees of freedom (*df*). Fig. 61 illustrates the cross sectional distribution of *df* for N225 stocks in the last quarter of 2002.

Undoubtedly, residuals do not have a specific distribution but it varies depending on the stocks. Nonetheless, two distinct clusterings are emerged, one is centered around 20 *df* and the other one around 100 *df*. Besides, based on Fig. 61 it is clear that majority of stocks experience heavy tail (*df* is small than 20) and it is in line with our previous expectations regarding the heavy tailed residual distribution.

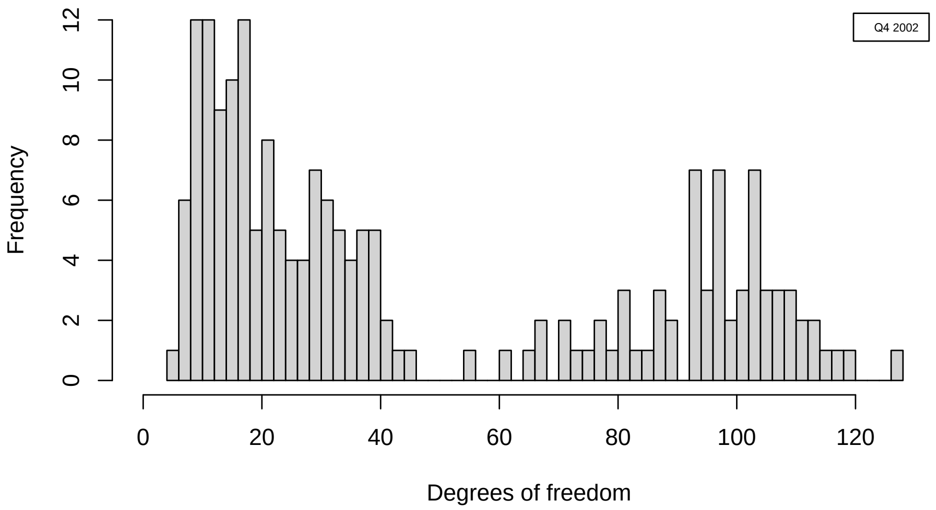

Figure 61: DF of t dist. fitted into R residuals, N225, Q4 2002

6.2 Asymptotic variance ratio

We obtain a parametric form of asymptotic variance for LS and R estimates of β by substituting the functional form of the density of t-distribution into the Eq. (6) and (7) in Section 2. It shows that the asymptotic variances are functions of the degrees of freedom. Specifically, for each quarter we estimated asymptotic variance for LS and R cases, respectively. Since, we used the t - distribution for estimation, its degrees of freedom changed depending on the quarter and stock.

Following the estimation of F we took the ratio of asymptotic variances $\hat{\sigma}^2(\hat{\beta}_{LS})/\hat{\sigma}^2(\hat{\beta}_R)$ by substituting into F, the estimate of t distribution with the degree of freedom (see Eq. (6) and (7)). Especially, variance ratio is higher when the degrees of freedom is smaller than approximately 20 (the tail is heavier than the Normal distribution and close to the Logistic). This behavior reverses when the degrees of freedom is bigger than 20 (it is close to Normal distribution).

Figures 62 - 63 illustrate cross sectional scatter plots of degrees of freedom and variance ratio. Besides, plots are colored based on the estimated $\hat{\theta}$. Thus, each dots in the plot represents statistics for individual stock for a given quarter, respectively. As we mentioned earlier, we fitted t distribution into residuals to estimate the suitable degrees of freedom, obtained variance ratio and also estimated $\hat{\theta}$. Hence, plots jointly represents all these three estimates for comparison purpose.

Obviously, LS is more efficient than R estimates (variance ratio is smaller than 0.95) when underlying t distribution's degrees of freedom is bigger than approximately 20. This is in line with our expectation since t with high df (close to normal) LS is more efficient. However, in case of strong asymmetry, R estimate is still more efficient (for instance, red dots) even df is larger than 20.

For cases of df smaller than 20, R estimate is clearly efficient with high variance ratio. Small

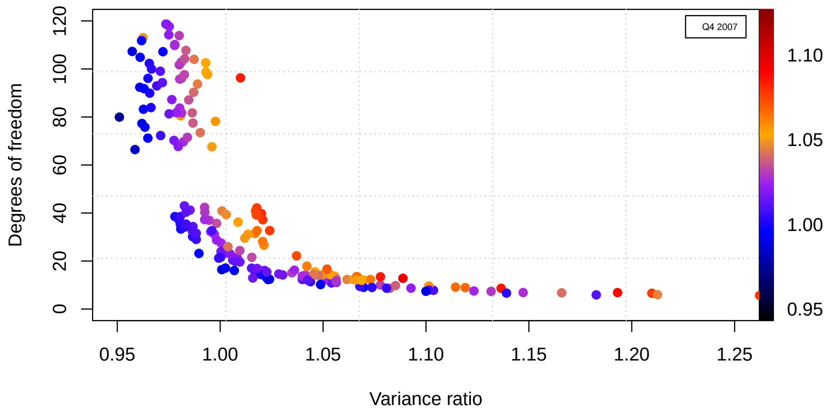

Figure 62: DF, variance ratio and θ relation, Q4 2007

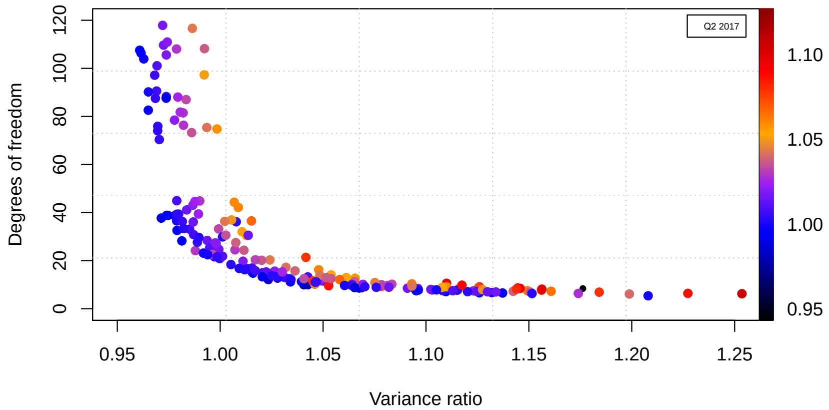

Figure 63: DF, variance ratio and θ relation, Q2 2007

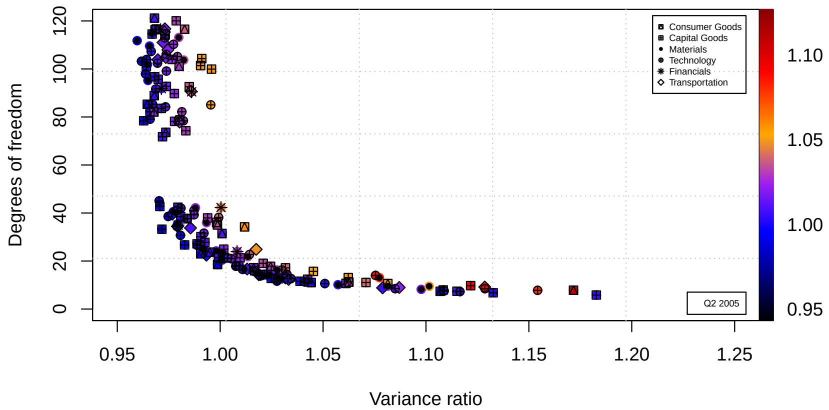

Figure 64: DF, variance ratio and θ relation, industry wise, Q2 2005

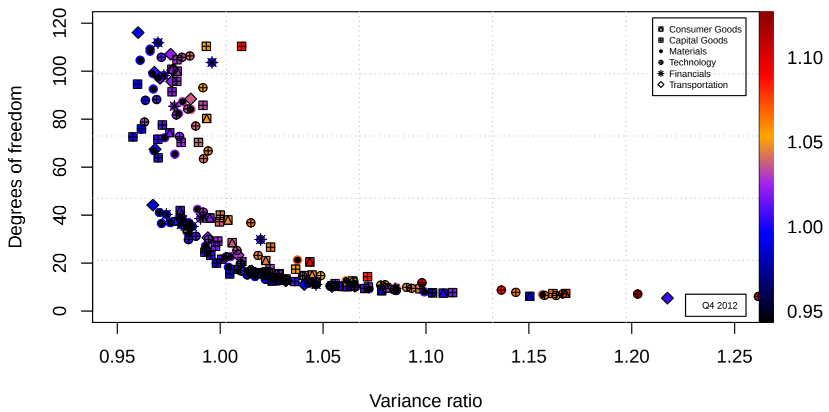

Figure 65: DF, variance ratio and θ relation, industry wise, Q4 2012

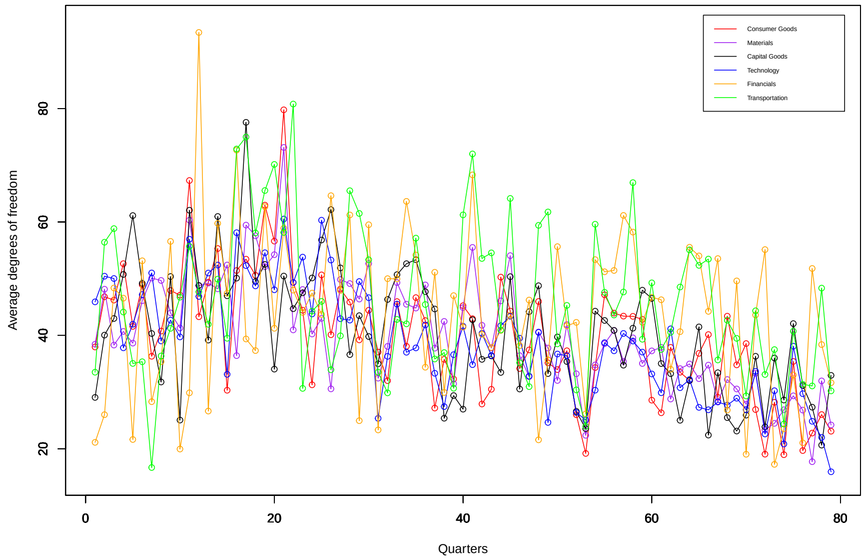

Figure 66: Average DF industry wise

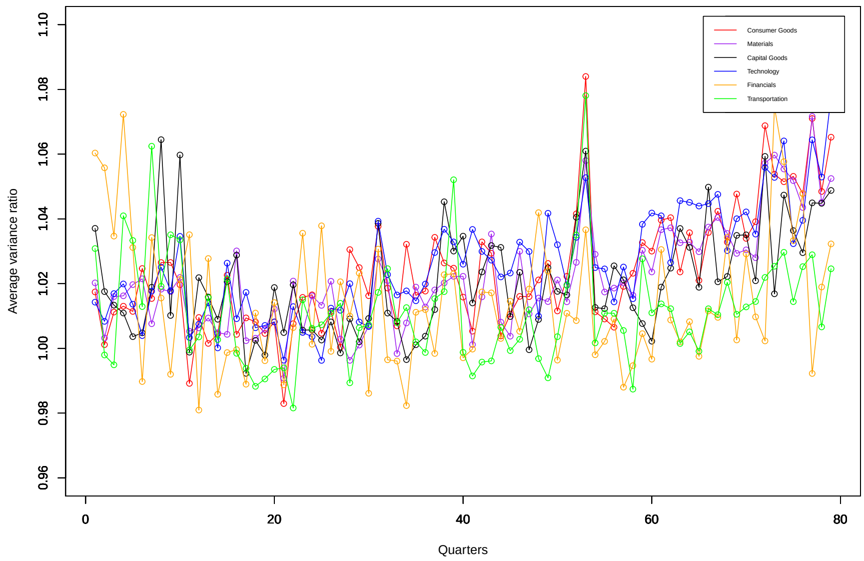

Figure 67: Average variance ratio industry wise

degrees of freedom indicates a heavy tailed distribution than Normal. Thus, our results indicate the efficiency of R estimates when residuals have heavier tails. Moreover, findings are in line with Lehmann's (1983) results given in Table 6.1 in Chapter 5. For instance, Table 6.1 shows the relative efficiency ratio to be 1.24 (for t-distribution with $df = 5$) which is in line with our findings.

Similarly, figures 64 - 65 illustrate the same relationship between 3 parameters (θ, variance ratio and degrees of freedom) and also displays the industry of companies, respectively. Interestingly, we can see a close patterns of stocks in one industry. For instance, "Capital Goods" stocks have high θ and high variance ratio but "Materials" stocks have high θ and lower variance ratio in Figure (64).

To investigate further by industry-wise, we estimated the average degrees of freedom and variance ratio for industries, respectively. As Figures 66 - 67 illustrate, "Financial" and "Transportation" company stocks have distinct patterns than the rest. More importantly, figures reveal that for most of the industries, the observed residuals are asymmetric on average and it has been increasing significantly in the last 5 years.

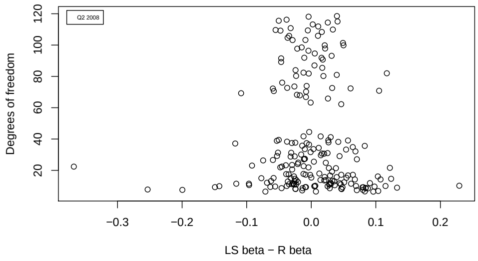

Figure 68: DF and β diff., N225, Q2 2008

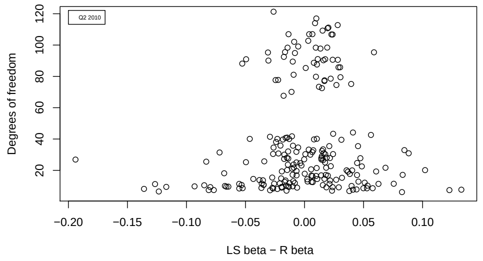

Figure 69: DF and β diff., N225, Q2 2010

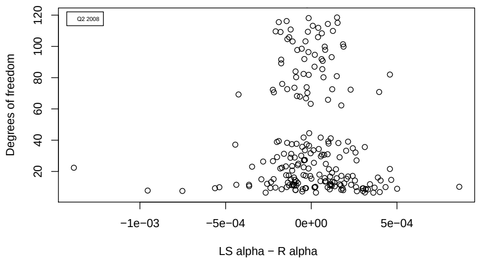

Figure 70: DF and α diff., N225, Q2 2008

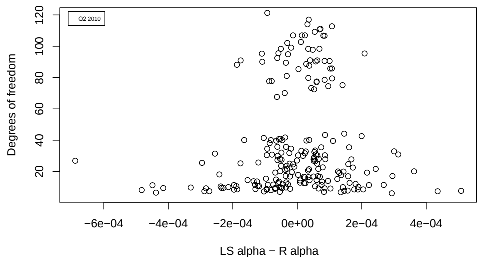

Figure 71: DF and α diff., N225, Q2 2010

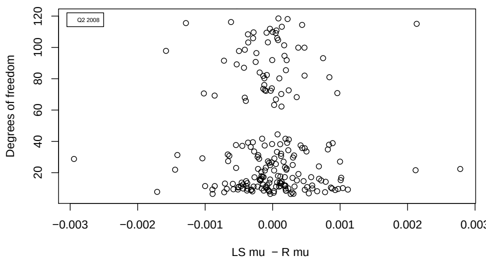

Figure 72: DF and μ diff., N225, Q2 2008

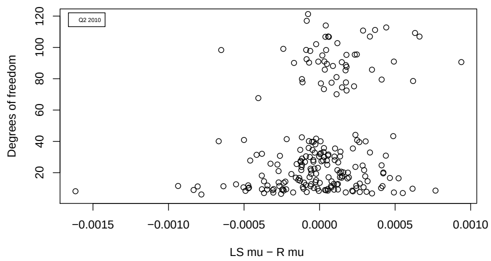

Figure 73: DF and μ diff., N225, Q2 2010

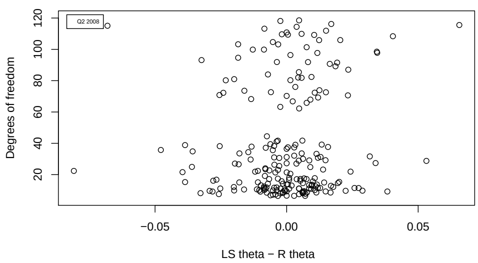

Figure 74: DF and θ diff., N225, Q2 2008

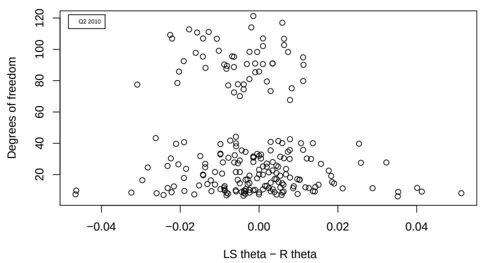

Figure 75: DF and θ diff., N225, Q2 2010

Figures 68 - 75 illustrate scatter plots of degrees of freedom and differences of estimates (β, α, μ, θ) for N225 stocks. The relation is quite distinct depending on the degrees of freedom. High group (60 - 120 *df*) of degrees of freedom have quite smaller differences of estimates while low group (4 - 40) have larger difference of estimates. Actually, t distribution with high degrees of freedom closer to normal and smaller degrees of freedom have heavy tails. Thus, figures clearly illustrate that LS and R estimates are noticeably distinctive when the underlying distribution has heavy tails.

In order to further investigate, we looked at the average θ (from R residuals). Based on *df* stocks are grouped into 5 - 15, 15 - 40 and 40 - 120 groups. Next, each groups' stocks are further divided into two subgroups based on positive and negative α difference (LS α - R α). Fig. (76) illustrates the average θ for each 6 groups, respectively. Clearly, following the end of 2008 (Q60) the average θs has a noticeable variation. Interestingly, the average θ increases for groups of stocks with negative α difference and for the case of heavy tail group (*df* between 5-15) the average θ is the highest.

In addition, Fig. (79) - (82) illustrate plots of θ and α difference for pre-crisis and post-crisis quarters for groups of stocks with *df* 5-15 and 15-40, respectively. Obviously, θ and α difference have a negative relation in post-crisis period.

Moreover, Fig. (77) illustrates the average θ for each 6 subsamples of stocks, respectively, when 3 *df* groups mentioned above are divided based on positive and negative β (LS β - R β) difference. This figure supports the previous result that in post-crisis period average θ for each groups are quite different and average θ increases in post crisis period for groups with negative β differences.

If we look at the number of stocks for each of 6 groups based on *df* and α difference (positive or negative) it is clear from Fig. (78) that stocks with smaller *df* has been increasing during the post crisis period.

Thus, when residuals have a noticeable asymmetry (average θ in Fig. (76) - (77)) LS underestimates α and β (negative α and β differences). This effect is even significant for cases when residuals have heavy tailed distribution (smaller *df* groups).

R estimate of β is more suitable and accurate when residuals have heavy tailed distribution or close to normal distribution but with noticeable asymmetry. R estimate is asymptotically more efficient than its counterpart LS and leads to precise estimation of $Jensen's\ \alpha$. As our results showed, majority of stocks experience heavy tailed distribution and rank statistics should be employed in order to estimate precise β and α.

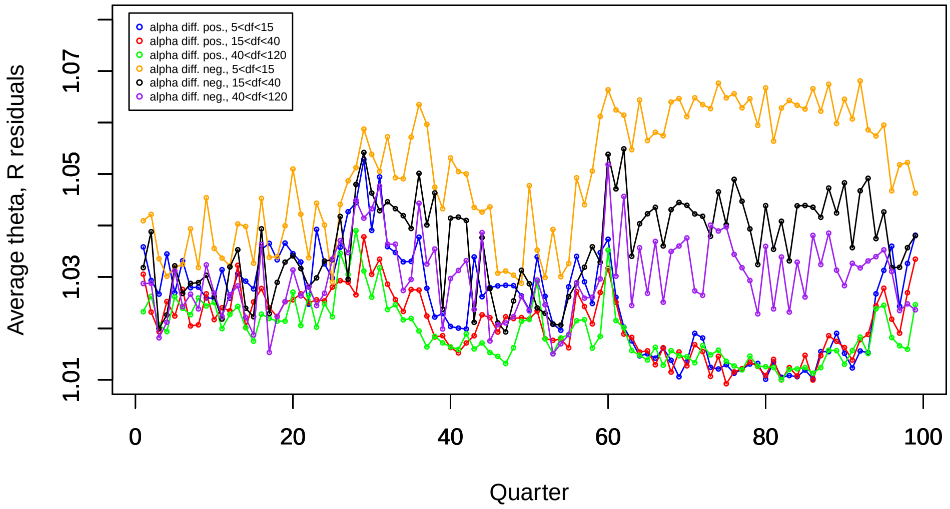

Figure 76: Average theta for 6 subsample of stocks

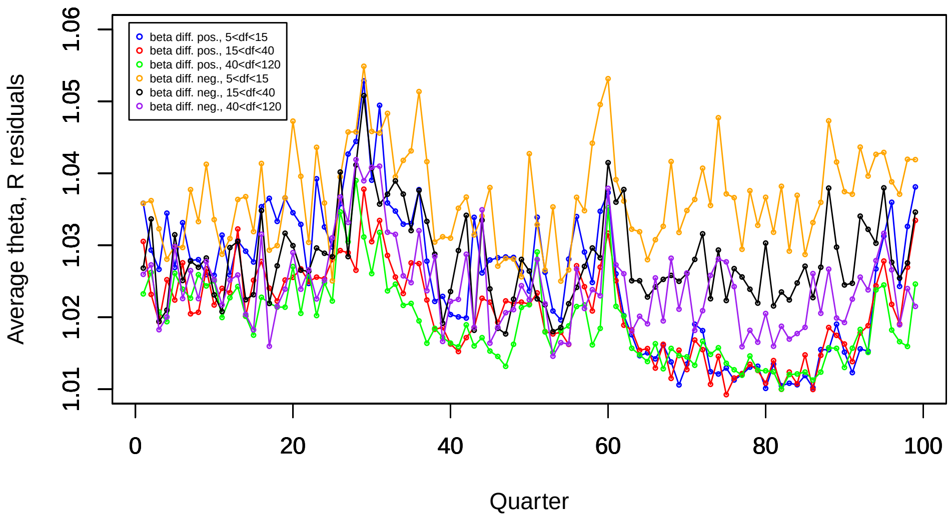

Figure 77: Average theta for 6 subsample of stocks

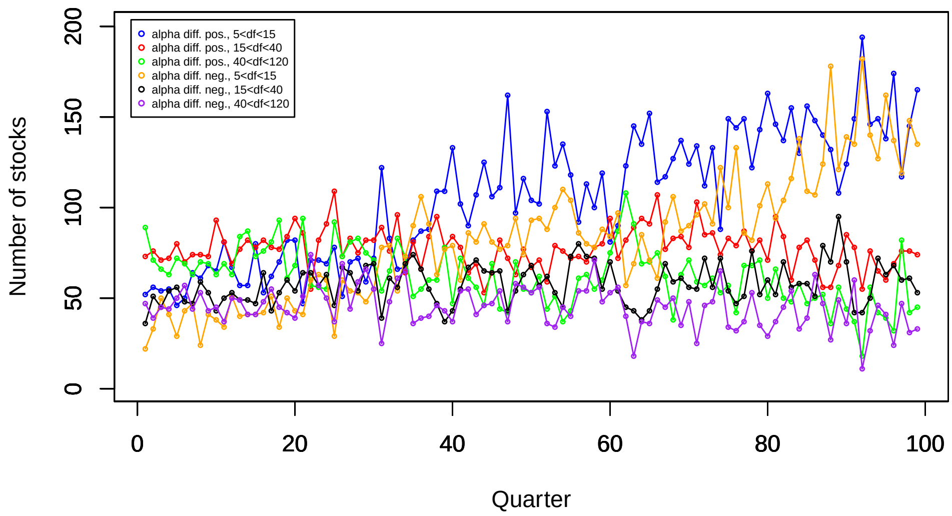

Figure 78: Number of stocks

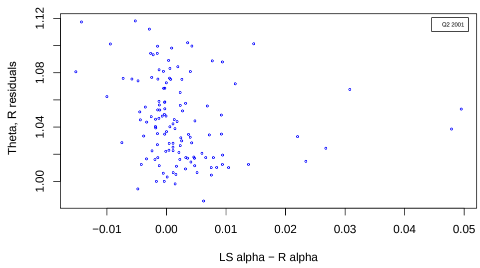

Figure 79: θ and α diff. (5-df-15), Q2 2001

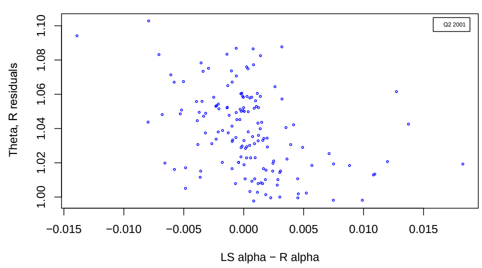

Figure 80: θ and α diff. (15-df-40), Q2 2001

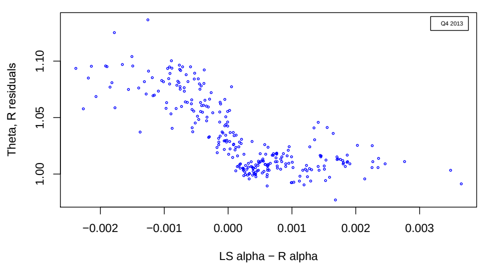

Figure 81: θ and α diff. (5-df-15), Q4 2003

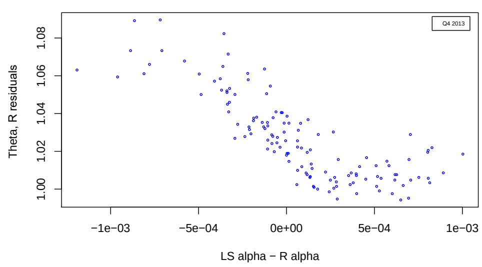

Figure 82: θ and α diff. (15-df-40), Q4 2003

7 Conclusion

This paper focused on a simple market model and applied two distinct approaches (LS and R) to estimate β for the purpose of accuracy comparison. LS is common and simple method in social sciences, on the contrary, rank is a distribution free, robust and widely applied approach in statistics field. Thus, CAPM β is estimated by both methods for each stock in 3 markets.

Moreover, study showed mathematically that the residuals can be used to estimate θ and μ. Thus, following the regression, the GLAM method is employed onto observed residuals to estimate the asymmetry (θ) and location (μ) parameters. Estimated θ showed that majority of stocks have a strong asymmetry over the sample period across stocks in Nikkei 225, FTSE 100 and S&P 500.

Study also found that asymmetry parameter θ is statistically significant to explain α from a simple linear regression. Especially, during the crisis periods (2007-2009) sensitiveness of α to θ more than tripled in Japanese stocks. Hence, decomposition of *Jensen's Alpha* showed that variation in α can be explained by the magnitude of asymmetry in error terms. In addition, α also explained by location (ξ) and skew (γ) from skew-t distribution, nonetheless, as the results showed skew parameter from skew-t distribution unable to explain α. Thus, obviously GLAM is suitable to measure asymmetry for α decomposition than parametric approach - skew-t.

Paper introduced "Estimation procedure of F" based on Z_i that enabled us to precisely estimate the underlying distribution F by fitting t distribution and obtaining degrees of freedom (*df*) for each stocks, respectively.

Furthermore, paper showed that *df*, asymmetry (θ) and variance ratio ($\sigma^2_{\beta,LS}/\sigma^2_{\beta,R}$) have a strong relation in common. Specifically, high *df* related with low variance ratio but variance ratio is larger as well in case of high θ, meaning that R estimate is more efficient. Also, grouping based on *df* and parameter (β, α) differences made clear that LS underestimates β and α when residual distribution is heavy tailed.

Thus, R is more accurate than LS in such cases when the error term is heavy tailed (which is the case for most of the stocks in our study) or have a significant asymmetry in its distribution (high θ values).

Our research sheds light on analyzing *Jensen's Alpha* from prospective of asymmetry in error term distribution and applying robust non-parametric approaches to estimate stock β. Application of θ indicator for portfolio construction could be a possible innovative approach and this is a topic for future research.

Acknowledgements
This work was supported by JSPS KAKENHI Grant Number 17H02508

References

[1] Arnott, R., Kalesnik, V. and Schuesler, T., Is Your Alpha Big Enough to Cover Its Taxes? A Quarter-Century Retrospective. The Journal of Portfolio Management, 44(5), pp.78-102. (2018)

[2] Azzalini, A., A class of distributions which includes the normal ones. Scandinavian Journal of Statistics, 12, pp.171178. (1985)

[3] Azzalini, A., The skew-normal and related families (Vol. 3). Cambridge University Press. (2013)

[4] Black, F., Capital market equilibrium with restricted borrowing. The Journal of Business, 45(3), pp.444-455. (1972)

[5] Da, Z., Guo, R.J. and Jagannathan, R., CAPM for estimating the cost of equity capital: Interpreting the empirical evidence. Journal of Financial Economics, 103(1), pp.204-220. (2012)

[6] Denhere, M. and Bindele, H.F., Rank estimation for the functional linear model. Journal of Applied Statistics, 43(10), pp.1928-1944. (2016)

[7] Edgeworth, F.Y., On observations relating to several quantities. Hermathena, 6(13), pp.279-285. (1887)

[8] Fama, E.F. and French, K.R., Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), pp.3-56. (1993)

[9] Fama, E.F. and French, K.R., The capital asset pricing model: Theory and evidence. Journal of Economic Perspectives, 18(3), pp.25-46. (2004)

[10] Gibbons, J.D., Nonparametric Methods for Quantitative Analysis, 3rd ed. American Sciences Press, Syracuse, New York. (1997)

[11] Graham, J.R. and Harvey, C.R., The theory and practice of corporate finance: evidence from the field. Journal of Financial Economics, 60(2-3), pp.187-243. (2001)

[12] Greene, W,H., Econometric Analysis. Pearson Education. (2012)

[13] Hettmansperger, T.P. and McKean, J.W., A robust alternative based on ranks to least squares in analyzing linear models. Technometrics, 19(3), pp.275-284. (1977)

[14] Hettmansperger, T.P. and Sheather, S.J., A cautionary note on the method of least median squares. The American Statistician, 46(2), pp.79-83. (1992)

[15] Hollander, M. and Sethuraman, J., Testing for agreement between two groups of judges. Biometrika, 65(2), pp.403-410. (1978)

[16] Hou, K., Xue, C. and Zhang, L., Digesting anomalies: An investment approach. The Review of Financial Studies, 28(3), pp.650-705. (2015)

[17] Jaeckel, L.A., Estimating regression coefficients by minimizing the dispersion of the residuals. The Annals of Mathematical Statistics, pp.1449-1458. (1972)

[18] Jagannathan, R. and Wang, Z., The conditional CAPM and the crosssection of expected returns. The Journal of Finance, 51(1), pp.3-53. (1996)

[19] Jensen, M.C., The performance of mutual funds in the period 1945-1964. The Journal of Finance, 23(2), pp.389-416. (1968)

[20] Jureckova, J., Nonparametric estimate of regression coefficients. The Annals of Mathematical Statistics, pp.1328-1338. (1971)

[21] Krasker, W.S. and Welsch, R.E., Efficient bounded-influence regression estimation. Journal of the American Statistical Association, 77(379), pp.595-604. (1982)

[22] Kubota, K. and Takehara, H., Expected return, liquidity risk, and contrarian strategy: evidence from the Tokyo Stock Exchange. Managerial Finance, 36(8), pp.655-679. (2010)

[23] Lehmann, E.L., Nonparametrics: Statistical Methods Based on Ranks. Holden Day, San Francisco, CA. (1975)

[24] Lehmann, E.L, Theory of point estimation. Springer, New York. (1983)

[25] Lewellen, J. and Nagel, S., The conditional CAPM does not explain asset-pricing anomalies. Journal of Financial Economics, 82(2), pp.289-314. (2006)

[26] Lintner, J., Security prices, risk, and maximal gains from diversification. The Journal of Finance, 20(4), pp.587-615. (1965)

[27] Markowitz, H., Portfolio Selection, Efficient Diversification of Investments. J. Wiley. (1959)

[28] Miura, R., Estimation of β based on Rank statistics: Japanese stock market. Management Research, Osaka City University Management Society 35 (6), pp.43-71. (In Japanese). (1985a)

[29] Miura, R., Estimation of β based on Rank statistics: supplement. Management Research, Osaka City University Management Society 36 (3), pp.51-68. (In Japanese). (1985b)

[30] Miura, R., and Tsukahara, H., One-sample estimation for generalized Lehmann's alternative models. Statistica Sinica, pp.83-101. (1993)

[31] Miura, R., A simple linear regression model with Generalized Lehmann's Alternative Model and R-estimators. Shogaku Kenkyu, 61, 3. (In Japanese). (2014)

[32] Mossin, J., Equilibrium in a capital asset market. Econometrica: Journal of the Econometric Society, pp.768-783. (1966)

[33] Onder, A.O. and Zaman, A., A test for normality based on robust regression residuals. In Developments in Robust Statistics pp.296-306. Physica, Heidelberg. (2003)

[34] Onder, A.O. and Zaman, A., Robust tests for normality of errors in regression models. Economics Letters , 86, pp.63-68. (2005)

[35] Perold, A.F., The capital asset pricing model. Journal of Economic Perspectives, 18(3), pp.3-24. (2004)

[36] Ross, S.A., The arbitrage theory of capital asset pricing. Journal of Economic Theory, 13(3), pp.341-360. (1976)

[37] Rousseeuw, P.J., Least Median of Squares Regression. Journal of American Statistical Association., 79, pp.871-880. (1984)

[38] Sharpe, W.F., Capital asset prices: A theory of market equilibrium under conditions of risk. The Journal of Finance, 19(3), pp.425-442. (1964)

[39] Shorack, G.R. and Wellner, J.A., Empirical Processes with Applications to Statistics. 1986. John Wiley&Sons. (1986)

[40] Welch, I., The consensus estimate for the equity premium by academic financial economists in December 2007. Available at SSRN 1084918. (2008)

[41] Zhou Tian Yun, Empirical study of beta in Japanese Stock Market : A Comparison Of Least Square Estimation and Estimator based on Rank Statistics. (in Japanese) Master Thesis. Graduate School of Commerce Hitotsubashi University. (2001)

8 Appendix

Estimation of (θ, μ) and F based on residuals

Assume that we have an estimate of β which has asymptotic normality.

$$\sqrt{n}(\hat{\beta} - \beta) \Rightarrow N(0, \sigma^2_{\hat{\beta}}) \tag{70}$$

as $n \to \infty$.

Converges in distribution to a normal distribution with mean 0 and variance $\sigma^2_{\hat{\beta}}$.

In the special construction, the convergence in distribution in the original probability space can be represented as a convergence almost every where (p.93 in Shorack and Wellner (1986), and p.757 in Pyke and Shorack (1968)). The advantage of this representation is that we can write a limit of convergence as an explicit random variable. It is very convenient for mathematical discussions of empirical processes.

We denote the limit of (70) by T in this Appendix.

8.1 Model

A simple linear regression model is

$$\begin{gathered} Y_i = \beta x_i + \alpha + \epsilon_i \\ i = 1, 2, ..., n \end{gathered} \tag{71}$$

where ϵ_i are *i.i.d.*

Least square method estimate α and β simultaneously, however, the estimate based on rank statistics (R-estimate) can estimate β without concerning α. Hence, for R-estimate the simple linear regression model can as well be written.

$$\begin{gathered} Y_i = \beta x_i + \eta_i \\ i = 1, 2, ..., n \end{gathered} \tag{72}$$

where

$$\begin{gathered} \eta_i = \alpha + \epsilon_i \\ E[\epsilon_i] = 0 \end{gathered} \tag{73}$$

and

$$\begin{gathered} \eta_i \sim G_{\mu,\theta}(x) = h(F(x - \mu) : \theta) \\ \theta \in (0, \infty) \end{gathered} \tag{74}$$

We call the right hand side a Generalized Lehmann's Alternative Model (in short GLAM). The relation of residual variables in the expectation is as follows.

$$
\begin{aligned}
E[\eta_i] = \int_{-\infty}^{\infty} x dG_{\mu,\theta}(x) &= \mu + \int_{-\infty}^{\infty} x dG_{0,\theta}(x) \\
&= \mu + \int_0^1 t h'(t:\theta) f(F^{-1}(t)) dt \\
&= \alpha + E[\epsilon_i] \\
&= \alpha
\end{aligned}
\tag{75}
$$

Thus,

$$
\alpha = \mu + \int_0^1 t f(F^{-1}(t)) d(h(t:\theta)) \tag{76}
$$

Second term is zero when θ is equal to θ^* for which $h(t:\theta^*) = t$. Assume that F is symmetric around zero.

Assume that functional form of F is unknown. We take the advantage of R-estimate of μ and θ which works well for this semi-nonparametric model.

8.2 Empirical distribution functions

Denote,

$$
\begin{aligned}
e_i(\beta) &= Y_i - \beta x_i \\
i &= 1, 2, ..., n
\end{aligned}
\tag{77}
$$

Then,

$$
\begin{aligned}
e_i(\hat{\beta}) &= Y_i - \hat{\beta} x_i \\
i &= 1, 2, ..., n
\end{aligned}
\tag{78}
$$

$e_i(\hat{\beta})$ are the residuals. Denote β_0 is a true parameter value. Then,

$$
\begin{aligned}
e_i(\beta_0) = Y_i - \beta_0 x_i &= \eta_i = \alpha + \epsilon i \\
i &= 1, 2, ..., n
\end{aligned}
\tag{79}
$$

are $i.i.d$ (independent and identically distributed).

Let

$$
G_n(x:\beta) = \frac{1}{n+1} \sum_{i=1}^{n} I\{e_i(\beta) \leq x\} \tag{80}
$$

and let $\tilde{G}_{n,\beta}$ be its linearized version.

Since the proofs are very similar to those in Miura and Tsukahara (1993) (M&T (1993), here after), we try to use the same notation so that it makes easier to see the reference and to simplify the proof-writing.

Denote

$$Z_i(r:\beta) = \tilde{G}_n^{-1}(h(\frac{i}{n+1}:r):\beta)$$
$$i = 1, 2, ..., n \tag{81}$$

Note that $Z_i(r:\beta_0)$ is $Z_i(r)$ in M&T (1993)

Now denote the empirical distribution function of $Z_i(r:\beta)$.

$$L_{n,r}(x:\beta) = \frac{1}{n}\sum_{i=1}^{n} I\{Z_i(r:\beta) \leq x\} \tag{82}$$

Then,

$$L_{n,r}(x:\beta) = u_n(h^{-1}(\tilde{G}_n(x:\beta):r)) \tag{83}$$

where,

$$u_n(t) = \frac{1}{n}\sum_{i=1}^{n} I\{\frac{i}{n+1} \leq t\}$$
$$t \in [0,1] \tag{84}$$

Note that $G_n(x:\beta_0)$ and $L_{n,r}(x:\beta_0)$ is the same process as $G_n(x)$ and $L_{n,r}(x)$, respectively in M&T (1993).

Lemma A-1, A-2 and A-3 in appendix tell us that the empirical distribution function of residuals with estimated $\hat{\beta}$ converges asymptotically to the empirical distribution function of residuals with true value of β which are $i.i.d$ random variables. Thus we can go for Proposition A-1 and A-2 to use $Z_i(\theta)$ in M&T (1993) for estimation of parameters (θ, μ) and F in GLAM, where $Z_i(\theta)$ in M&T (1993) is constructed based on $i.i.d$ random variables which correspond to residuals with true value of β.

Denote the rank of $|Z_i(r:\beta) - q|$

$$R_i^+(r,q:\beta) = the \quad number \quad of \quad \{j : |Z_j(r:\beta) - q| \leq |Z_i(r:\beta) - q|\} \tag{85}$$

Note that $R_i^+(r,q:\beta_0)$ is the same as $R_i^+(r,q)$ in M & T (1993).

Define, for $x > 0$, the empirical distribution of $|Z_i(r:\hat{\beta}_n) - q|$ be

$$H_{n,r,q}(x:\hat{\beta}_n) \triangleq \frac{1}{n+1}(the \quad number \quad of \quad \{i : |Z_i(r:\hat{\beta}_n) - q| \leq x\}) \tag{86}$$

, for $x > 0$.

Then, we can write

$$R_i^+(r,q:\hat{\beta}_n) = (n+1)H_{n,r,q}(|Z_i(r:\hat{\beta}_n) - q|) \tag{87}$$

$\beta = \beta_0$ relates to $i.i.d$ case in M&T (1993), while $\beta = \hat{\beta}_n$ is for residuals. If we put $\hat{\beta}_n$ in the notation, it means that the empirical process is based on residuals.

8.3 Lemmas

Lemma A-1

$$\begin{gathered}\sqrt{n}\{\tilde{G}_n(x:\hat{\beta}_n) - \tilde{G}_n(x:\beta_0)\} \\ \Rightarrow \text{converges to} \\ T\bar{x}h_1(F(x-\mu):\theta)f(x-\mu)\end{gathered} \tag{88}$$

Proof.

Let $\epsilon_i' = \eta_i - \mu$ and $\hat{\beta}_n$ is an estimate of β_0.

$$\begin{gathered}I\{e_i(\hat{\beta}_n) \le x\} = I\{Y_i - \hat{\beta}_n x_i \le x\} = \\ I\{\mu + \epsilon_i' \le x + (\hat{\beta}_n - \beta_0)x_i\} = I\{\epsilon_i' \le x - \mu + (\hat{\beta}_n - \beta_0)x_i\}\end{gathered} \tag{89}$$

and

$$\begin{gathered}I\{e_i(\beta_0) \le x\} = I\{\mu + \epsilon_i' \le x\} = \\ I\{\epsilon_i' \le x - \mu\}\end{gathered} \tag{90}$$

where,

$$\epsilon_i' \sim G_{0,\theta}(x) = h(F(x):\theta) \tag{91}$$

So,

$$\begin{gathered}\sqrt{n}\{\tilde{G}_n(x:\hat{\beta}_n) - \tilde{G}_n(x:\beta_0)\} \\ = \sqrt{n}\left[\frac{1}{n+1}\sum_{i=1}^{n} I\{\epsilon_i' \le x - \mu + (\hat{\beta}_n - \beta_0)x_i\} - \frac{1}{n+1}\sum_{i=1}^{n} I\{\epsilon_i' \le x - \mu\}\right]\end{gathered} \tag{92}$$

We know that,

$$\sqrt{n}\Big\{\frac{1}{n+1}\sum_{i=1}^{n} I\{\epsilon_i' \le x - \mu\} - h(F(x-\mu):\theta)\Big\} \tag{93}$$

converges in distribution to $U(G_{\mu,\theta}(x))$ where $U(\cdot)$ is a Brownian Bridge (see Shorack and Wellner (1986), p120).

Now

$$
\begin{aligned}
&\sqrt{n}\{\tilde{G}_n(x:\hat{\beta}_n) - \tilde{G}_n(x:\beta_0)\} \\
&= \sqrt{n}\Big\{\tilde{G}_n(x:\hat{\beta}_n) - \frac{1}{n+1}\sum_{i=1}^{n} h(F(x-\mu+(\hat{\beta}_n-\beta_0)x_i):\theta)\Big\} \\
&+\sqrt{n}\Big\{\frac{1}{n+1}\sum_{i=1}^{n} h(F(x-\mu+(\hat{\beta}_n-\beta_0)x_i):\theta) - h(F(x-\mu):\theta)\Big\} \\
&-\sqrt{n}\Big\{G_n(x:\beta_0) - G_{\mu,\theta}(x)\Big\} \\
&= ① + ② - ③
\end{aligned}
\tag{94}
$$

$$
\begin{aligned}
① &= \sqrt{n}\frac{1}{n+1}\sum_{i=1}^{n}\Big[I\{\epsilon_i' \le x-\mu+(\hat{\beta}_n-\beta_0)x_i\} - h(F(x-\mu+(\hat{\beta}_n-\beta_0)x_i):\theta_0)\Big] \\
② &= \sqrt{n}\frac{1}{n+1}\sum_{i=1}^{n}\Big[h(F(x-\mu+(\hat{\beta}_n-\beta_0)x_i):\theta_0) - \frac{n+1}{n}h(F(x-\mu):\theta_0)\Big] \\
③ &= \sqrt{n}\Big\{G_n(x:\beta_0) - G_{\mu,\theta}(x)\Big\}
\end{aligned}
\tag{95}
$$

① and ③ converges to the same Brownian Bridge $U(G_{\mu,\theta}(x))$, so that ① - ③ converges to 0.

$$
\begin{aligned}
② &= \frac{1}{n+1}\sum_{i=1}^{n}\frac{d}{dt}h(t:\theta)\Big|_{t=F(x-\mu)} f(x-\mu)\sqrt{n}(\hat{\beta}-\beta)x_i \\
&= \frac{1}{n+1}\Big(\sum_{i=1}^{n}x_i\Big)\frac{d}{dt}h(t:\theta)\Big|_{t=F(x-\mu)} f(x-\mu)\sqrt{n}(\hat{\beta}-\beta) \\
&\to T\bar{x}h_1(F(x-\mu):\theta)f(x-\mu)
\end{aligned}
\tag{96}
$$

where T and $\bar{x}$ are a limits of $\sqrt{n}(\hat{\beta}-\beta)$ and $\frac{1}{n}\sum_{i=1}^{n}x_i$, respectively, and h_1 is $\frac{d}{dt}h(t:\theta)$

Lemma A-2

(1) For

$$
\begin{aligned}
r &= \theta + \frac{s}{\sqrt{n}} \\
0 &< s < C < \infty
\end{aligned}
\tag{97}
$$

the following hold uniformly in s, as $n \to \infty$.

$$
\sqrt{n}\{L_{n,r}(x:\hat{\beta}_n) - L_{n,r}(x:\beta_0)\} \to T(J_\beta,h,g)\bar{x}f(x-\mu), -\infty < x < \infty \tag{98}
$$

(2) Further

$$
\sqrt{n}\{H_{n,r,q}(x:\hat{\beta}_n) - H_{n,r,q}(x:\beta_0)\} \to T\bar{x}\{f(x+q-\mu) + f(-x+q-\mu)\} \tag{99}
$$

Proof.
Noting that as $n \to \infty$

$$\begin{aligned} h(\tilde{G}_n(x:\cdot):r) &\approx h(G_n(x:\cdot):r) \\ r &= \theta + \frac{s}{\sqrt{n}}, \\ 0 < |s| &< C < \infty \end{aligned} \tag{100}$$

and that by definitions, of $u_n(\cdot)$ and $Z_i(r:\hat{\beta}_n)$

$$L_{n,r}(x:\hat{\beta}_n) = \frac{1}{n}\sum_{i=1}^{n} I\{Z_i(r:\hat{\beta}_n) \le x)\} = u_n(h^{-1}\tilde{G}_n(x:\hat{\beta}_n):r)) \tag{101}$$

Then, we have

$$\begin{aligned} &\sqrt{n}\{L_{n,r}(x:\hat{\beta}_n) - L_{n,r}(x:\hat{\beta}_0)\} \\ &= \sqrt{n}[\frac{1}{n}\sum_{i=1}^{n} I\{Z_i(r:\hat{\beta}_n) \le x\} - \frac{1}{n}\sum_{i=1}^{n} I\{Z_i(r:\beta_0) \le x\}] \\ &= \sqrt{n}[\frac{1}{n}\sum_{i=1}^{n} I\{\frac{i}{n+1} \le h^{-1}(\tilde{G}_n(x:\hat{\beta}_n):r)\} - \frac{1}{n}\sum_{i=1}^{n} I\{\frac{i}{n+1} \le h^{-1}(\tilde{G}_n(x:\beta_0):r)\}] \\ &\approx \sqrt{n}[\frac{1}{h_1(h^{-1}(G(x)^*:r))}\{\tilde{G}_n(x:\hat{\beta}_n):r) - \tilde{G}_n(x:\beta_0):r)\}] \\ &\to \frac{1}{h_1(h^{-1}(G(x)^*:\theta))} T(J_\beta,h,g)\bar{x}h_1(F(x-\mu):\theta)f(x-\mu) \\ &= T(J_\beta,h,g)\bar{x}f(x-\mu) \end{aligned} \tag{102}$$

where $G(x)^*$ is some value between $\tilde{G}_n(x:\hat{\beta}_n)$ and $\tilde{G}_n(x:\beta_0)$
Proof for the second statement

$$
\begin{aligned}
&\sqrt{n}\{H_{n,r,q}(x:\hat{\beta}_n)-H_{n,r,q}(x:\beta_0)\}\\
&=\sqrt{n}(\frac{n+1}{n})\Big[\{L_{n,r}(x+q:\hat{\beta}_n)-L_{n,r}(-x+q:\hat{\beta}_n)\}\\
&\quad-\{L_{n,r}(x+q:\beta_0)-L_{n,r}(-x+q:\beta_0)\}\Big]\\
&=\frac{n+1}{n}\Big[\sqrt{n}\{L_{n,r}(x+q:\hat{\beta}_n)-L_{n,r}(x+q:\beta_0)\}\\
&\quad-\sqrt{n}\{L_{n,r}(-x+q:\hat{\beta}_n)-L_{n,r}(-x+q:\beta_0)\}\Big]\\
&=\frac{n+1}{n}\Big[\sqrt{n}\{u_n(h^{-1}(\tilde{G}_n(x+q:\hat{\beta}_n):r))-u_n(h^{-1}(\tilde{G}_n(x+q:\beta_0):r))\}\\
&\quad-\sqrt{n}\{u_n(h^{-1}(\tilde{G}_n(-x+q:\hat{\beta}_n):r))-u_n(h^{-1}(\tilde{G}_n(-x+q:\beta_0):r))\}\Big]\\
&\approx\sqrt{n}\{h^{-1}(\tilde{G}_n(x+q:\hat{\beta}_n):r)-h^{-1}(\tilde{G}_n(x+q:\beta_0):r)\}\\
&\quad-\sqrt{n}\{h^{-1}(\tilde{G}_n(-x+q:\hat{\beta}_n):r)-h^{-1}(\tilde{G}_n(-x+q:\beta_0):r)\}\\
&\approx\frac{1}{h_1(h^{-1}(G_n^*:r):r)}\sqrt{n}\{\tilde{G}_n(x+q:\hat{\beta}_n)-\tilde{G}_n(x+q:\beta_0)\}\\
&\quad-\frac{1}{h_1(h^{-1}(G_n^{**}:r):r)}(-1)\sqrt{n}\{\tilde{G}_n(-x+q:\hat{\beta}_n)-\tilde{G}_n(-x+q:\beta_0)\}
\end{aligned}
\tag{103}
$$

$$
\begin{aligned}
&\to\frac{1}{h_1(F(x+q-\mu):\theta)}T\bar{x}h_1(F(x+q-\mu):\theta)f(x+q-\mu)\\
&\quad-\frac{1}{h_1(F(-x+q-\mu):\theta)}T\bar{x}h_1(F(-x+q-\mu):\theta)(-1)f(-x+q-\mu)\\
&=T\bar{x}\{f(x+q-\mu)+f(-x+q-\mu)\}
\end{aligned}
\tag{104}
$$

where G_n^* is some value between $\tilde{G}_n(x+q:\hat{\beta}_n)$ and $\tilde{G}_n(x+q:\beta_0)$ and G_n^{**} is similarly some value between $\tilde{G}_n(-x+q:\hat{\beta}_n)$ and $\tilde{G}_n(-x+q:\beta_0)$.

Lemma A-3

(1) For

$$
\begin{aligned}
r&=\theta+\frac{s}{\sqrt{n}}\\
0&<s<C<\infty
\end{aligned}
\tag{105}
$$

we have the following uniformly in s, as $n\to\infty$.

$$
\begin{aligned}
&\sqrt{n}\{L_{n,r}(x:\hat{\beta}_n)-F(x-\mu)\}\\
&\to T(J_\beta,h,g)\bar{x}f(x-\mu)+\frac{U(h(F(x-\mu):\theta))}{h_1(F(x-\mu):\theta)}-s\frac{h_2(F(x-\mu):\theta)}{h_1(F(x-\mu):\theta)}
\end{aligned}
\tag{106}
$$

where $U(t)$, $0<t<1$ is the same as in Lemma A-1 which is the limit of $\sqrt{n}[G_n(G^{-1}_{\mu,\theta}(t:\beta_0):\beta_0)-t]$, $0<t<1$ and is a Brownian Bridge.

Further, we have the following

$$\begin{aligned}&\sqrt{n}[H_{n,r,q}(x:\hat{\beta}_n)-\{F(x-\mu)-F(-(x-\mu))\}]\\ &\to T\bar{x}\{f(x+q-\mu)+f(-x+q-\mu)\}+[\frac{U(h(F(x-\mu):\theta))}{h_1(F(x-\mu):\theta)}-\frac{U(h(1-F(x-\mu):\theta))}{h_1(1-F(x-\mu):\theta)}]\\ &-s[\frac{h_2(F(x-\mu):\theta)}{h_1(F(x-\mu):\theta)}-\frac{h_2(1-F(x-\mu):\theta)}{h_1(1-F(x-\mu):\theta)}]\end{aligned} \tag{107}$$

where $h_1=\frac{d}{dt}h(t:\theta)$ and $h_2=\frac{d}{d\theta}h(t:\theta)$.

Proof.

We have

$$\begin{aligned}&\sqrt{n}\{L_{n,r}(x:\hat{\beta}_n)-F(x-\mu)\}\\ &=\sqrt{n}\{L_{n,r}(x:\hat{\beta}_n)-L_{n,r}(x:\beta_0)\}+\sqrt{n}\{L_{n,r}(x:\beta_0)-F(x-\mu)\}\\ &=①+②\end{aligned} \tag{108}$$

The limit of ① is provided by Lemma A-2, and the limit of ② is given in M&T (1993).

Further in a similar way, we have

$$\begin{aligned}&\sqrt{n}[H_{n,r,q}(x:\hat{\beta}_n)-\{F(x-\mu)-F(-(x-\mu))\}]\\ &=\sqrt{n}[H_{n,r,q}(x:\hat{\beta}_n)-H_{n,r,q}(x:\beta_0)]+\sqrt{n}[H_{n,r,q}(x:\beta_0)-\{F(x-\mu)-F(-(x-\mu))\}]\\ &=①+②\end{aligned} \tag{109}$$

As shown in Lemma A-2, ① converge to $T\bar{x}\{f(x+q-\mu)+f(-x+q-\mu)\}$ and the limit of ② is shown in M&T (1993),

8.4 Estimation of (θ, μ)

We will prove that $S_{n,\theta}((r,q):\hat{\beta}_n)$ approximates $S_{n,\theta}((r,q):\beta_0)$. Then, by remarking that $S_{n,\theta}((r,q):\beta_0)$ is the same rank statistic as in M&T (1993). The same comment applies to the other statistic $S_{n,\mu}((r,q):\hat{\beta}_n)$.

Thus, in order to prove that asymptotic linearity of $(S_{n,\theta}((r,q):\hat{\beta}_n), S_{n,\mu}((r,q):\hat{\beta}_n))$, it is enough to work on $(S_{n,\theta}((r,q):\hat{\beta}_n) - S_{n,\theta}((r,q):\beta_0), S_{n,\mu}((r,q):\hat{\beta}_n) - S_{n,\mu}((r,q):\beta_0))$ since M&T (1993) proved asymptotic linearity of $(S_{n,\theta}((r,q):\beta_0), S_{n,\mu}((r,q):\beta_0))$.

We define, as in M&T (1993) our estimate $(\hat{\theta}, \hat{\mu})$ of θ and μ based on the regression residuals be the values in

$$D^*_{n,\hat{\beta}_n} \triangleq \left\{(r,q) : |S_{n,\theta}(r,q:\hat{\beta}_n)| + |S_{n,\mu}(r,q:\hat{\beta}_n)| = min\right\} \tag{110}$$

Now, we go on to prove asymptotic linearity of $S_{n,\theta}(r,q:\hat{\beta}_n), S_{n,\mu}(r,q:\hat{\beta}_n)$.

Our rank statistic can be written as

$$\begin{aligned}
S_{n,\theta}((r,q):\hat{\beta}_n) &= \int_q^\infty J_\theta(\frac{1}{2} + \frac{1}{2}H_{n,r,q}(x-q:\hat{\beta}_n))dL_{n,r}(x:\hat{\beta}_n)\\
&+ \int_{-\infty}^q J_\theta(\frac{1}{2} - \frac{1}{2}H_{n,r,q}(-(x-q):\hat{\beta}_n)))dL_{n,r}(x:\hat{\beta}_n)\\
S_{n,\mu}((r,q):\hat{\beta}_n) &= \int_q^\infty J_\mu(\frac{1}{2} + \frac{1}{2}H_{n,r,q}(x-q:\hat{\beta}_n))dL_{n,r}(x:\hat{\beta}_n)\\
&+ \int_{-\infty}^q J_\mu(\frac{1}{2} - \frac{1}{2}H_{n,r,q}(-(x-q):\hat{\beta}_n))dL_{n,r}(x:\hat{\beta}_n)
\end{aligned} \tag{111}$$

Proof of A-1.

$$\begin{aligned}
&S_{n,\theta}((r,q):\hat{\beta}_n) - S_{n,\theta}((r,q):\beta_0)\\
&= \int_q^\infty J_\theta(\frac{1}{2} + \frac{1}{2}H_{n,r,q}(x-q:\hat{\beta}_n))dL_{n,r}(x:\hat{\beta}_n)\\
&- \int_q^\infty J_\theta(\frac{1}{2} + \frac{1}{2}H_{n,r,q}(x-q:\beta_0))dL_{n,r}(x:\beta_0)\\
&+ \int_{-\infty}^q J_\theta(\frac{1}{2} - \frac{1}{2}H_{n,r,q}(-(x-q):\hat{\beta}_n))dL_{n,r}(x:\hat{\beta}_n)\\
&- \int_{-\infty}^q J_\theta(\frac{1}{2} - \frac{1}{2}H_{n,r,q}(-(x-q):\beta_0))dL_{n,r}(x:\beta_0)\\
&= [A] + [B]
\end{aligned} \tag{112}$$

We will prove for $[A]$ only. The proof for $[B]$ goes in a very similar way as for $[A]$.

$$
\begin{aligned}
[A] = \Big[\int_q^\infty & J_\theta(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\hat\beta_n))dL_{n,r}(x:\hat\beta_n) \\
& -\int_q^\infty J_\theta(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\hat\beta_n))dL_{n,r}(x:\beta_0)\Big] \\
& +\Big[\int_q^\infty J_\theta(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\hat\beta_n))dL_{n,r}(x:\beta_0) \\
& -\int_q^\infty J_\theta(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\beta_0))dL_{n,r}(x:\beta_0)\Big] \\
& = [A_1]+[A_2]
\end{aligned}
\tag{113}
$$

Thanks to the mean - value theorem for a continuous function J_θ, we have

$$
\begin{aligned}
& \sqrt{n}\Big\{J_\theta(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\hat\beta_n))-J_\theta(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\beta_0))\Big\} \\
& =\sqrt{n}J'_\theta(\frac{1}{2}+\frac{1}{2}x^*)\frac{1}{2}\{H_{n,r,q}(x-q:\hat\beta_n)-H_{n,r,q}(x-q:\beta_0)\} \\
& \to \frac{1}{2}J'_\theta(F(x-\mu))T\bar{x}\{f(x-\mu)+f(-(x-\mu))\}
\end{aligned}
\tag{114}
$$

for some x^* between $H_{n,r,q}(x-q:\hat\beta_n)$ and $H_{n,r,q}(x-q:\beta_0)$.
Thus, for the first term $[A_1]$ in $[A]$, we have

$$
\begin{aligned}
& \sqrt{n}\Big\{\int_q^\infty J_\theta(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\hat\beta_n))dL_{n,r}(x:\hat\beta_n) \\
& -\int_q^\infty J_\theta(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\hat\beta_n))dL_{n,r}(x:\beta_0)\Big\} \\
= & \sqrt{n}\int_q^\infty J_\theta(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\hat\beta_n))\big\{dL_{n,r}(x:\hat\beta_n)-dL_{n,r}(x:\beta_0)\big\} \\
& \to \int_\mu^\infty J_\theta(F(x-\mu))d\big\{T(\lim_{n\to\infty}\bar{x})f(x-\mu)\big\}
\end{aligned}
\tag{115}
$$

The second term $[A_2]$ in $[A]$ can be written as

$$
\int_q^\infty \big\{J_\theta(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\hat\beta_n))-J_\theta(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\beta_0))\big\}dL_{n,r}(x:\beta_0) \tag{116}
$$

As discussed in the above the integrand converges to $\frac{1}{2}J'_\theta(F(x-\mu))T\bar{x}\{f(x-\mu)+f(-(x-\mu))\}$ in $\sqrt{n}$ order as $n\to\infty$. Also we know that $L_{n,r}(x-q:\beta_0)$ converges to $F(x-\mu)$.

Thus, we have for the second term in $[A]$, as $n\to\infty$

$$\sqrt{n}[A_2] = \sqrt{n}\int_q^\infty \{J_\theta(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\hat{\beta}_n)) - J_\theta(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\beta_0))\}dL_{n,r}(x:\beta_0)$$
$$\to \int_0^\infty \frac{1}{2}J'_\theta(F(x-\mu))T\bar{x}\{f(x-\mu)+f(-(x-\mu))\}dF(x-\mu) \tag{117}$$

Thus, as $n \to \infty$

$$\begin{gathered}\sqrt{n}[A] \to T\bar{x}\int_0^\infty J_\theta(F(x))df(x) + T\bar{x}\int_0^\infty J'_\theta(F(x))f(x)dF(x) \\ (\equiv 2T(\bar{x})\int_{\frac{1}{2}}^1 f(F^{-1}(t))dJ_\theta(t)\end{gathered} \tag{118}$$

Assuming $\lim_{x\to I\infty} f(x)J_\theta(F(x)=0)$.

For $[B]$, it can be proved in a very similar way that

$$\sqrt{n}[B] \to T(\bar{x})\int_{-\infty}^0 J_\theta(F(x))df(x) + T\bar{x}\int_{-\infty}^0 J'_\theta(F(x))f(x)dF(x) \tag{119}$$

Thus, we have

$$\begin{gathered}\sqrt{n}([A]+[B]) \to T\bar{x}\int_{-\infty}^\infty J_\theta(F(x))df(x) + T\bar{x}\int_{-\infty}^\infty J'_\theta(F(x))f(x)dF(x) \\ (\equiv 2T\bar{x}\int_0^1 f(F^{-1}(t))dJ_\theta(t))\end{gathered} \tag{120}$$

Thus, we have, uniformly in b_1 and b_2, where

$$\begin{aligned} r &= \theta_0 + \frac{b_1}{\sqrt{n}} \\ q &= \mu_0 + \frac{b_2}{\sqrt{n}} \end{aligned} \tag{121}$$

for $|b_1| \le B, |b_2| \le B$

as $n \to \infty$

Now for

$$\begin{gathered} r = \theta_0 + \frac{b_1}{\sqrt{n}} \\ q = \mu_0 + \frac{b_2}{\sqrt{n}} \\ |b_1| \le B_1 \\ |b_2| \le B_2 \end{gathered} \tag{122}$$

$$
\begin{aligned}
&S_{n,\mu}((r,q):\hat{\beta}_n) - S_{n,\mu}((r,q):\beta_0) \\
&= \int_{-\infty}^{\infty} J_\mu(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\hat{\beta}_n))dL_{n,r}(x:\hat{\beta}_n) \\
&- \int_{-\infty}^{\infty} J_\mu(\frac{1}{2}-\frac{1}{2}H_{n,r,q}(x-q:\hat{\beta}_0))dL_{n,r}(x:\beta_0) \\
&= \int_{q}^{\infty} J_\mu(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\hat{\beta}_n))dL_{n,r}(x:\hat{\beta}_n) \\
&- \int_{q}^{\infty} J_\mu(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\beta_0))dL_{n,r}(x:\beta_0) \\
&+ \int_{-\infty}^{q} J_\mu(\frac{1}{2}-\frac{1}{2}H_{n,r,q}(-(x-q):\hat{\beta}_n))dL_{n,r}(x:\hat{\beta}_n) \\
&- \int_{-\infty}^{q} J_\mu(\frac{1}{2}-\frac{1}{2}H_{n,r,q}(-(x-q):\beta_0))dL_{n,r}(x:\beta_0) \\
&= [C]+[D]
\end{aligned} \tag{123}
$$

(This can be worked out in a very similar way as for $S_{n,\theta}$)

$$
\begin{aligned}
[C] &= \int_{q}^{\infty} J_\mu(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\hat{\beta}_n))dL_{n,r}(x:\hat{\beta}_n) \\
&- \int_{q}^{\infty} J_\mu(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\hat{\beta}_n))dL_{n,r}(x:\beta_0) \\
&+ \int_{q}^{\infty} J_\mu(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\hat{\beta}_n))dL_{n,r}(x:\beta_0) \\
&- \int_{q}^{\infty} J_\mu(\frac{1}{2}+\frac{1}{2}H_{n,r,q}(x-q:\beta_0))dL_{n,r}(x:\beta_0) \\
&= [C1]+[C2]
\end{aligned} \tag{124}
$$

As discussed for $[A]$, it can be proved that as $n \to \infty$

$$
\sqrt{n}[C1] \to T(\bar{x}) \int_{0}^{\infty} J_\mu(F(x))df(x) \tag{125}
$$

Also, we have that as discussed for $[A]$, as $n \to \infty$.

$$
\sqrt{n}[C2] \to T\bar{x} \int_{0}^{\infty} J'_\mu(F(x))f(x)dF(x) \tag{126}
$$

Thus, we have that as $n \to \infty$

$$
\begin{aligned}
&\sqrt{n}[C] \to \\
&T\bar{x}\{\int_{0}^{\infty} J_\mu(F(x))df(x) + \int_{0}^{\infty} J'_\mu(F(x))f(x)dF(x)\}
\end{aligned} \tag{127}
$$

uniformly in (b_1, b_2) for $|b_1| \le B$ and $|b_2| \le B$.
Very similarly as for $[B]$, we have uniformly in b_1 and b_2,

$$
\sqrt{n}[D] \to T\bar{x}\{\int_{-\infty}^{0} J_\mu(F(x))df(x) + \int_{-\infty}^{0} J'_\mu(F(x))f(x)dF(x)\} \tag{128}
$$

as $n \to \infty$.

Thus, we have, uniformly in b_1, b_2.

$$\begin{aligned} &\sqrt{n}\{S_{n,\mu}((r,q):\hat{\beta}_n) - S_{n,\mu}((r,q):\beta_0)\} \\ \to T\bar{x}\{&\int_{-\infty}^{\infty} J_\mu(F(x))df(x) + \int_{-\infty}^{\infty} J'_\mu(F(x))f(x)dF(x)\} \end{aligned} \tag{129}$$

Now, noting that M&T (1993) on page 97 proved that $S_{n,\theta}((r,q):\beta_0)$ and $S_{n,\mu}((r,q):\beta_0)$ converges to T_1^* and T_2^*, respectively, we have $S_{n,\theta}((r,q):\hat{\beta}_n)$ and $S_{n,\mu}((r,q):\hat{\beta}_n)$ converges to

$$T\bar{x}\{\int_{-\infty}^{\infty} J_\theta(F(x))df(x) + \int_{-\infty}^{\infty} J'_\theta(F(x))f(x)dF(x)\} + T_1^* \tag{130}$$

and

$$T\bar{x}\{\int_{-\infty}^{\infty} J_\mu(F(x))df(x) + \int_{-\infty}^{\infty} J'_\mu(F(x))f(x)dF(x)\} + T_2^* \tag{131}$$

, respectively.

The asymptotic linearity of $S_{n,\theta}((r,q):\beta_0)$ and $S_{n,\mu}((r,q):\beta_0)$ can be derived just as M & T (1993) does without having affected by the asymptotic behavior of $\{S_{n,\mu}((r,q):\hat{\beta}_n) - S_{n,\mu}((r,q):\beta_0)\}$ and $\{S_{n,\theta}((r,q):\hat{\beta}_n) - S_{n,\theta}((r,q):\beta_0)\}$ as seen in the above, of course, since the estimation error $\hat{\beta}_n - \beta_0$ bring out T, the asymptotic variance of $\hat{\theta}_n$ and $\hat{\mu}_n$ are enlarged by the corresponding terms which include T, although their expectations (means) are zero.

This means that the estimation for the parameters (θ, μ) can be well done, being based on the residuals provided by any suitable estimate of β which has $\sqrt{n}$ - order asymptotic normality such as the usual least square estimate and R-estimate.

In order to see the limit of $\sqrt{n}\{\hat{\theta}_n - \theta_0\}$ where $\hat{\theta}_n$ is constructed using residuals, we look at the limit of $\sqrt{n}S_{n,\theta}((r,q):\hat{\beta}_n)$ with $r = \theta_0 + \frac{b_1}{\sqrt{n}}$, $|b_1| \le B_1$, $q = \mu_0 + \frac{b_2}{\sqrt{n}}$, $|b_2| \le B_2$, which is

$$\begin{gathered} T\bar{x}\{\int_{-\infty}^{\infty} J_\theta(F(x))df(x) + \int_{-\infty}^{\infty} J'_\theta(F(x))f(x)dF(x)\} + T_1^* \\ T\bar{x}\{\int_{-\infty}^{\infty} J_\theta(F(x))df(x) + \int_{-\infty}^{\infty} J'_\theta(F(x))f(x)dF(x)\}\frac{1}{2}T_1 \\ -\frac{1}{2}d_{1,1}b_1 + d_{1,2}b_2 \end{gathered} \tag{132}$$

Under this asymptotic linearity the limit of $\sqrt{n}(\theta_n - \theta_0)$ can be obtained by solving the following linear equation for b_1

$$\begin{bmatrix} \sqrt{n}S_{n,\theta}((r,q):\hat{\beta}_n) \\ \sqrt{n}S_{n,\mu}((r,q):\hat{\beta}_n) \end{bmatrix} = \begin{bmatrix} T\bar{x}\{\int_{-\infty}^{\infty} J_\theta(F(x))df(x) + \int_{-\infty}^{\infty} J'_\theta(F(x))f(x)dF(x)\} \\ T\bar{x}\{\int_{-\infty}^{\infty} J_\mu(F(x))df(x) + \int_{-\infty}^{\infty} J'_\mu(F(x))f(x)dF(x)\} \end{bmatrix} \tag{133}$$

$$+ \begin{bmatrix} \int_{-\infty}^{\infty}\{\frac{U(h(F(x):\theta_0))}{h_1(F(x):\theta_0)} + \frac{U(h(1-F(x):\theta_0))}{h_1(1-F(x):\theta_0)}\}dJ_\theta(F(x)) \\ \int_{-\infty}^{\infty}\{\frac{U(h(F(x):\theta_0))}{h_1(F(x):\theta_0)} + \frac{U(h(1-F(x):\theta_0))}{h_1(1-F(x):\theta_0)}\}dJ_\mu(F(x)) \end{bmatrix} \tag{134}$$

$$+ \begin{bmatrix} d_{11}, & d_{12} \\ d_{21}, & d_{22} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \equiv M_\beta + M_{(\theta,\mu)} + D_{(\theta,\mu)}\underset{\sim}{b} \tag{135}$$

Thus,

$$\underset{\sim}{b}^* = \begin{bmatrix} b_1^* \\ b_2^* \end{bmatrix} = -D_{(\theta,\mu)}^{-1}(M_\beta + M_{(\theta,\mu)}) \tag{136}$$

by setting the right side equal to zero.

Proposition A-1 makes a bridge between estimation for θ and μ based on residuals e($\hat{\beta}$) and that based on e(β_0). M&T (1993) defined estimators for θ and μ based on ϵ (β_0) and proved asymptotic Normality of estimators. So what we have added here is the difference in the case where we estimate θ and μ based on residuals e($\hat{\beta}_n$) instead of e(β_0). Proposition A-1 reveals that the limit of the rank statistics based on residuals differ from β_0 known case, only by a single term whose randomness comes from estimation of β. This means that the estimation error based on residuals just adds a term caused by β estimation part to the original estimation error based on the case of β known (as in M&T (1993)).

By looking at the form of limit which is a simple sum of a few integrals of Brownian Bridge, it is clear that their expectation is zero and Normality of distribution follows. However, we will not provide their specific form of their asymptotic variances. This is because we do not go further into the investigation variances of these estimators.

Remark

Denote the limit variable of $\sqrt{n}(\hat{\beta} - \beta_0)$ be T. In case of R-estimate of β with score function J_β, T can be written as

$$T = T(J_\beta, h, g) = \frac{\int_0^1 J_\beta(t)dW(t)}{\int_0^1 J_\beta(t)J(t,g)dt(\lim_{n\to\infty}\sum_{i=1}^n \frac{x_i^2}{n})} \tag{137}$$

where

$$J_\beta(t:g)dt = \frac{g'(G^{-1}(t))}{g(G^{-1}(t))} \tag{138}$$

and $G(t)$ is the distribution function of $(\eta_i - \mu)$.

Since $\hat{\beta}_n$ is defined as the value of β that minimizes

$$|S_{n,\beta} = \sum_{i=1}^{n} J_\beta(\frac{R_i(\beta)}{n+1})x_i| \tag{139}$$

where, $R_i(\beta)$ is the rank of $Y_i - \beta x_i$ among $\{Y_j - \beta x_j : j = 1, 2, ...n\}$.
Note that $W(t)$ is a limit of the weighted empirical process $W_n(t)$.

$$W_{n,}(t) = \frac{1}{\sqrt{x'x}} \sum_{i=1}^{n} x_i \Big[I\{G(\eta_i) \leq t\} - t\Big] \tag{140}$$

8.5 Estimation of F

Lemma A-3 proved convergence of $L_{n,r}(x:\hat{\beta}_n) - L_{n,r}(x:\beta_0)$, uniformly in r for $|r| \le \beta$. Here, we use $\hat{\theta}_n$ in place of r in order to estimate F. This means that we estimate the unknown F after obtaining the estimates of β and θ. (Note that we do not need estimate of μ here).

We will prove that $L_{n,\hat{\theta}_n}(x:\hat{\beta}_n)$ is an estimator of F and has $\sqrt{n}$ - order asymptotic normality

Proposition A-2.

Let $\hat{\theta}_n$ be the estimate of θ defined in Appendix 8.4 Eq. (110), just as in M&T (1993).

$$\sqrt{n}\big\{L_{n,\hat{\theta}_n}(x:\hat{\beta}_n) - F(x-\mu)\big\} \to N(0,\sigma_F^2) \tag{141}$$

in distribution as $n \to \infty$

Proof of A-2

Note that $L_{n,\hat{\theta}}(x:\hat{\beta}_n)$ is using the estimate of θ discussed in Appendix 8.4,

Noting that $F(x-\mu) = h^{-1}(G_{\mu,\theta}(x))$, and also that $L_{n,\hat{\theta}}(x:\hat{\beta}_n) = u_n h^{-1}(\tilde{G}_n(x:\hat{\beta}_n):\hat{\theta})$

$$\begin{aligned} L_{n,\hat{\theta}_n}(x:\hat{\beta}_n) - F(x-\mu) &= \\ +\Big[\frac{1}{n+1}\sum_{i=1}^{n} I\{Z_i(\hat{\theta}_n:\hat{\beta}_n)\le x\} &- \frac{i}{n+1}\sum_{i=1}^{n} I\{Z_i(\hat{\theta}_n:\beta_0)\le x\}\Big] \\ +\Big[\frac{1}{n+1}\sum_{i=1}^{n} I\{Z_i(\hat{\theta}_n:\beta_0)\le x\} &- \frac{i}{n+1}\sum_{i=1}^{n} I\{Z_i(\theta_0:\beta_0)\le x\}\Big] \\ +\Big[\frac{1}{n+1}\sum_{i=1}^{n} I\{Z_i(\theta_0:\beta_0)\le x\} &- F(x-\mu)\Big] \\ &= ① + ② + ③ \end{aligned} \tag{142}$$

For ①,

$$\begin{aligned} ① &= \frac{1}{n+1}\sum_{i=1}^{n}\Big[I\{\tilde{G}^{-1}_{n,\hat{\beta}_n} h(\frac{i}{n+1}:\hat{\theta}_n)\le x\} - I\{\tilde{G}^{-1}_{n,\beta_0} h(\frac{i}{n+1}:\hat{\theta}_n)\le x\}\Big] \\ &= \frac{1}{n+1}\sum_{i=1}^{n}\Big[I\{\frac{i}{n+1}\le h^{-1}(\tilde{G}_{n,\hat{\beta}_n}(x):\hat{\theta}_n)\} - I\{\frac{i}{n+1}\le h^{-1}(\tilde{G}_{n,\beta_0}(x):\hat{\theta})\}\Big] \\ &= \frac{n}{n+1}\Big[u_n(h^{-1}(\tilde{G}_{n,\hat{\beta}_n}(x):\hat{\theta}_n)) - u_n(h^{-1}(\tilde{G}_{n,\beta_0}(x):\hat{\theta}_n))\Big] \\ &= \frac{n}{n+1}\Big[u_n(h^{-1}(\tilde{G}_{n,\beta_0}(x)^*:\hat{\theta}_n))\Big\{h^{-1}(\tilde{G}_{n,\hat{\beta}_n}(x):\hat{\theta}_n) - h^{-1}(\tilde{G}_{n,\beta_0}(x):\hat{\theta}_n)\Big\}\Big] \end{aligned} \tag{143}$$

where $\tilde{G}_{n,\beta_0}(x)^*$ is some value between $\tilde{G}_{n,\hat{\beta}_n}(x)$ and $\tilde{G}_{n,\beta_0}(x)$

Then, as $n \to \infty$ $\sqrt{n}①$ converges in distribution to

$$
\begin{aligned}
&h^{-1}(G(x):\theta_0)\Big[\frac{d}{dr}h^{-1}(t:\theta_0)\Big|_{t\,=\,G(x)}\Big]\lim_{n\to\infty}\Big\{\sqrt{n}\{\tilde{G}_{n,\hat{\beta}_n}(x)-\tilde{G}_{n,\beta_0}(x)\}\Big\} \\
&= F(x-\mu)\frac{1}{h_1(h^{-1}(G(x):\theta_0):\theta_0)}T(\lim_{n\to\infty}\bar{x})h_1(F(x-\mu):\theta_0)f(x-\mu) \\
&= F(x-\mu)\frac{1}{h_1(F(x-\mu):\theta_0)}T(\lim_{n\to\infty}\bar{x})h_1(F(x-\mu):\theta_0)f(x-\mu) \\
&= T(\lim_{n\to\infty}\bar{x})F(x-\mu)f(x-\mu)
\end{aligned}
\tag{144}
$$

In the same way, we have

$$
\begin{aligned}
② &= \frac{1}{n+1}\sum_{i=1}^{n}\Big[I\{Z_i(\hat{\theta}_n:\beta_0)\le x\}-I\{Z_i(\theta_0:\beta_0)\le x\}\Big] \\
&= \frac{1}{n+1}\sum_{i=1}^{n}\Big[I\{\tilde{G}^{-1}_{n,\beta_0}(h(\frac{i}{n+1}:\hat{\theta}_n))\le x\}-I\{\tilde{G}^{-1}_{n,\beta_0}(h(\frac{i}{n+1}:\theta_0))\le x\}\Big] \\
&= \frac{n}{n+1}\Big[u_n(h^{-1}(\tilde{G}_{n,\beta_0}(x):\hat{\theta}_n))-u_n(h^{-1}(\tilde{G}_{n,\beta_0}(x):\theta_0))\Big] \\
&= \frac{n}{n+1}\Big[u_n(h^{-1}(\tilde{G}_{n,\beta_0}(x):\hat{\theta}_n)^*)\Big\{h^{-1}(\tilde{G}_{n,\beta_0}(x):\hat{\theta}_n)-h^{-1}(\tilde{G}_{n,\beta_0}(x):\theta_0)\Big\}\Big]
\end{aligned}
\tag{145}
$$

where $h^{-1}(\tilde{G}_{n,\beta_0}(x):\hat{\theta}_n)^*$ is some value between $h^{-1}(\tilde{G}_{n,\beta_0}(x):\hat{\theta}_n)$ and $h^{-1}(\tilde{G}_{n,\beta_0}(x):\theta_0)$

Then, as $n\to\infty$ the limit of $\sqrt{n}②$ is

$$
\begin{aligned}
&\lim_{n\to\infty}u_n(h^{-1}(G(x):\theta_0))\Big[\frac{d}{d\theta}h^{-1}(t:\theta)\Big|_{\substack{t\,=\,G(x)\\ \theta\,=\,\theta_0}}\Big]\sqrt{n}\Big\{\hat{\theta}_n-\theta_0\Big\} \\
&= h^{-1}(G(x):\theta_0)\frac{h_2(h^{-1}(t:\theta_0):\theta_0)}{h_1(h^{-1}(t:\theta_0):\theta_0)}\Big|_{t\,=\,G(x)}\lim_{n\to\infty}\sqrt{n}\Big\{\hat{\theta}_n-\theta_0\Big\} \\
&= h^{-1}(h(F(x-\mu):\theta_0):\theta_0)\frac{h_2(h^{-1}(G(x):\theta_0):\theta_0)}{h_1(h^{-1}(G(x):\theta_0):\theta_0)}\lim_{n\to\infty}(\sqrt{n}\Big\{\hat{\theta}_n-\theta_0\Big\}) \\
&= F(x-\mu)\frac{h_2(F(x-\mu):\theta_0)}{h_1(F(x-\mu):\theta_0)}\lim_{n\to\infty}(\sqrt{n}\Big\{\hat{\theta}_n-\theta_0\Big\})
\end{aligned}
\tag{146}
$$

$$
\begin{aligned}
③ &= \frac{1}{n+1}\sum_{i=1}^{n}I\{\tilde{G}^{-1}_{n,\beta_0}(h(\frac{i}{n+1}:\theta_0))\le x\}-F(x-\mu) \\
&= \frac{1}{n+1}\sum_{i=1}^{n}I\{\frac{i}{n+1}\le h^{-1}(\tilde{G}_{n,\beta_0}(x):\theta_0)\}-F(x-\mu) \\
&= \frac{n}{n+1}\sum_{i=1}^{n}u_n(h^{-1}(\tilde{G}_{n,\beta_0}(x):\theta_0))-F(x-\mu)
\end{aligned}
\tag{147}
$$

Then, $\sqrt{n}③$ converges, as $n\to\infty$, to

$$
\begin{aligned}
&\lim_{n\to\infty}\sqrt{n}\Big\{u_n(h^{-1}(\tilde{G}_{n,\beta_0}(x):\theta_0)) - u_n(F(x-\mu)) + u_n(F(x-\mu)) - F(x-\mu)\sqrt{n}\Big\} \\
&\lim_{n\to\infty}\sqrt{n}\Big\{h^{-1}(\tilde{G}_{n,\beta_0}(x):\theta_0) - h^{-1}(G(x)) + \{u_n(F(x-\mu)) - F(x-\mu)\}\sqrt{n}\Big\} \\
&= \frac{d}{d\theta}h^{-1}(t:\theta)\Bigg|_{\substack{t=G(x)\\ \theta=\theta_0}}\sqrt{n}\{\tilde{G}_{n,\beta_0}(x) - G(x)\} \\
&= \frac{1}{h_1(h^{-1}(G(x):\theta_0):\theta_0))}\lim_{n\to\infty}\sqrt{n}\{\tilde{G}_{n,\beta_0}(x) - G(x)\} \\
&= \frac{\lim_{n\to\infty}\sqrt{n}\{\tilde{G}_{n,\beta_0}(x) - G(x)\}}{h_1(F(x-\mu):\theta_0)}
\end{aligned} \tag{148}
$$

As shown in Lemma A-1 the limit of $\sqrt{n}\{\tilde{G}_{n,\beta_0}(x) - G(x)\}$ is $T(\lim_{n\to\infty}\bar{x})h_1(F(x-\mu):\theta_0)f(x-\mu)$. Hence $h_1(F(x-\mu):\theta_0)$ cancels out and we have the limit of $\sqrt{n}③$ is $T(\lim_{n\to\infty}\bar{x})f(x-\mu)$.

Let LEF (limit variable of $\sqrt{n}$ order error of estimation of F) denote the limit of $\sqrt{n}(① + ② + ③)$. We know it is normally distributed with mean 0 and variance σ_F^2.

8.6 Tables

Table 8: Descriptive statistics of average skewness and kurtosis, S&P500

Statistic	N	Mean	St. Dev.	Min	Max
Skewness LS	99	0.156	0.147	−0.285	0.573
Skewness R	99	0.159	0.151	−0.288	0.580
Kurtosis LS	99	2.909	1.144	1.119	5.286
Kurtosis R	99	3.063	1.175	1.188	5.562

Table 9: Descriptive statistics of average skewness and kurtosis, FTSE100

Statistic	N	Mean	St. Dev.	Min	Max
Skewness LS	131	0.205	0.216	−0.486	0.888
Skewness R	131	0.210	0.228	−0.616	0.929
Kurtosis LS	131	2.353	0.743	1.103	5.436
Kurtosis R	131	2.527	0.816	1.168	5.711

Table 10: Descriptive statistics of θ, R

Statistic	N	Mean	St. Dev.	Min	Max
Toyota Motor Corp	79	1.034	0.035	0.989	1.107
Taisei Corp	79	1.041	0.036	0.987	1.115
Takashimaya Co	79	1.036	0.031	0.984	1.104
Nippon Express Co Ltd	79	1.034	0.033	0.984	1.118
Canon Inc	79	1.041	0.035	0.987	1.116
Mitsubishi Corp	79	1.041	0.031	0.972	1.107

Table 11: Descriptive statistics of θ, LS

Statistic	N	Mean	St. Dev.	Min	Max
Toyota Motor Corp	79	1.035	0.035	0.989	1.107
Taisei Corp	79	1.042	0.036	0.965	1.107
Takashimaya Co	79	1.036	0.033	0.982	1.108
Nippon Express Co Ltd	79	1.034	0.032	0.984	1.112
Canon Inc	79	1.040	0.033	0.982	1.108
Mitsubishi Corp	79	1.045	0.030	0.997	1.108

Table 12: Descriptive statistics of μ, R

Statistic	N	Mean	St. Dev.	Min	Max
Toyota Motor Corp	79	−0.001	0.001	−0.004	0.002
Taisei Corp	79	−0.001	0.002	−0.006	0.004
Takashimaya Co	79	−0.001	0.002	−0.011	0.003
Nippon Express Co Ltd	79	−0.001	0.002	−0.005	0.002
Canon Inc	79	−0.001	0.001	−0.004	0.003
Mitsubishi Corp	79	−0.001	0.002	−0.006	0.004

Table 13: Descriptive statistics of μ, LS

Statistic	N	Mean	St. Dev.	Min	Max
Toyota Motor Corp	79	−0.001	0.001	−0.004	0.002
Taisei Corp	79	−0.001	0.002	−0.006	0.007
Takashimaya Co	79	−0.001	0.002	−0.010	0.003
Nippon Express Co Ltd	79	−0.001	0.002	−0.005	0.004
Canon Inc	79	−0.001	0.001	−0.005	0.003
Mitsubishi Corp	79	−0.001	0.002	−0.005	0.004

Table 14: Average μ of N225 stocks

Statistic	N	Mean	St. Dev.	Min	Max
R	79	−0.001	0.001	−0.004	0.001
LS	79	−0.001	0.001	−0.004	0.001

Table 15: FTSE100 stocks

Statistic	N	Mean	St. Dev.	Min	Max
R	79	−0.007	0.010	−0.035	0.002
LS	79	−0.006	0.010	−0.034	0.002

Table 16: S&P500 stocks

Statistic	N	Mean	St. Dev.	Min	Max
R	79	−0.002	0.006	−0.019	0.011
LS	79	−0.002	0.006	−0.020	0.012

Table 17: Average θ of N225 stocks

Statistic	N	Mean	St. Dev.	Min	Max
R	79	1.039	0.006	1.026	1.058
LS	79	1.039	0.006	1.026	1.060

Table 18: FTSE100 stocks

Statistic	N	Mean	St. Dev.	Min	Max
R	79	1.026	0.005	1.016	1.040
LS	79	1.025	0.004	1.017	1.039

Table 19: S&P500 stocks

Statistic	N	Mean	St. Dev.	Min	Max
R	79	1.031	0.005	1.022	1.045
LS	79	1.031	0.005	1.022	1.046

8.7 Figures

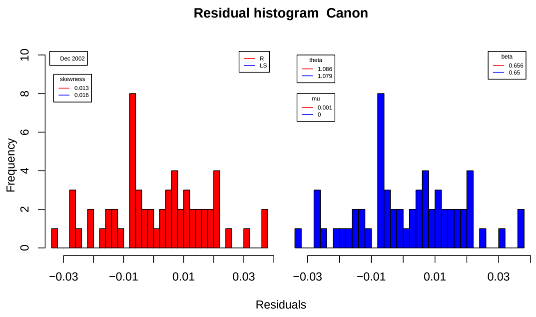

Figure 83: Residual hist. Q4 2002

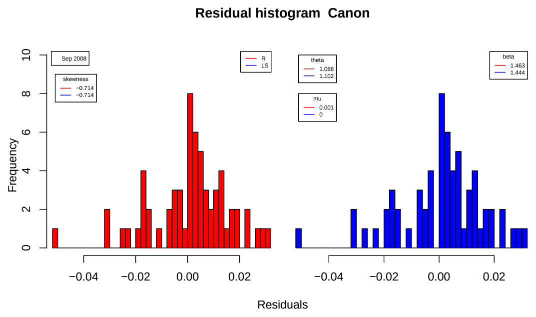

Figure 84: Residual hist. Q3 2008

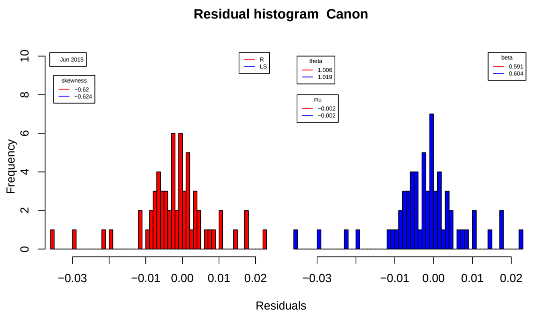

Figure 85: Residual hist. Q2 2015

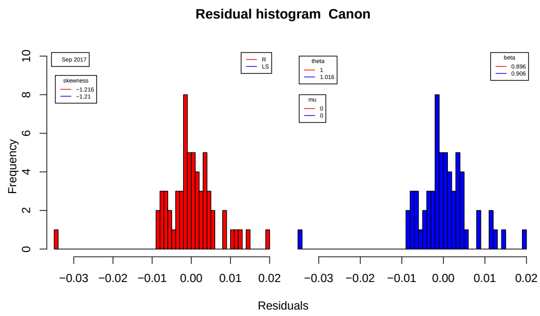

Figure 86: Residual hist. Q3 2017