Estimating mean-variance ratios of financial data.

H.E.T. Holgersson[*], Rashid Mansoor[*] and Peter S. Karlsson[*]

[*]Jönköping International, Business School, Box 1026, 55 111 Sweden

Abstract.

Keywords. Asymptotics, return-risk ratio, high dimensions.

1. Introduction

Investigating the relation of the second central moment of a distribution to its first raw moment is of great interest in many situations it. There are many other applications in different areas such as in economics and medical applications. For example, according to the theory of finance, an investor is compensated by increased returns for taking higher risks (i.e. variance of the investments returns), which is represented by the Sharpe ratio μ/σ. One way to quantify the relation between expected excess return (i.e. return on an asset minus the risk free return) and risk is to calculate the Sharpe ratio (Sharpe, 1966). In order to make inference of the Sharpe ratio, it is necessary to define a simple model which explains that the relation between mean excess return and standard deviation of excess return for stocks is linear and allows unsystematic deviations from the linear relation. In investigating potential estimators for Sharpe ratio, two widely observed characteristics of financial data need to be taken into consideration. Firstly, the number of the variables (number of assets in the cross-section), n, and the number of observations over time, T, are usually of comparable sizes, so the asymptotic property of any estimator of Sharpe ratio should allow n and T to be proportional. Secondly, the data follow a factor structure based on pricing theory of financial assets. The impact of the later characteristic (the assumed underlying factor structure) implies that the covariance matrix may be expressed with a reduced dimension. This fact has an important impact on the convergence rate of the variance of each estimator of Sharpe ratio. In this paper we suggest three different estimators of the Sharpe ratio, then those estimators are applied on a real data set.

The paper develops as follows: in section 2 we present the assumed underlying model of the mean-variance ratio and present three potential estimators of the Sharpe ratio. In this section we also discuss the asymptotics of the three suggested estimators in an context where both a parameter space and a observations space limit infinity. In section 3 an empirical application of the estimators to a dataset from the Stockholm Stock exchange is presented.

2. Modeling mean-standard deviation in high dimensions

The relation between signal (mean value) and noise (standard deviation) has long been considered in statistics. The most commonly used expression is probably that of the coefficient of variation suggested by Pearson (1895), defined by the ratio between the standard deviation to the mean value, i.e. σ/μ, and has played a central role in fields such as quality control, economics and biology. Its reciprocal value, μ/σ has also been frequently used in imaging, finance (Sharpe ratio), etc. In most cases these applications has involved the problem of making inference of the relation between μ and σ of a single variable, say X. In some contexts, however, there may be a systematic relation between the standard deviations and mean values of a large number of heterogeneous variables. This may happen for example when analyzing the blood pressure of patients in a longitudinal study or the relation between excess return and risk of a financial assets. In order to make inference in such cases it is convenient to set up a simple model for the data. For example, we may set

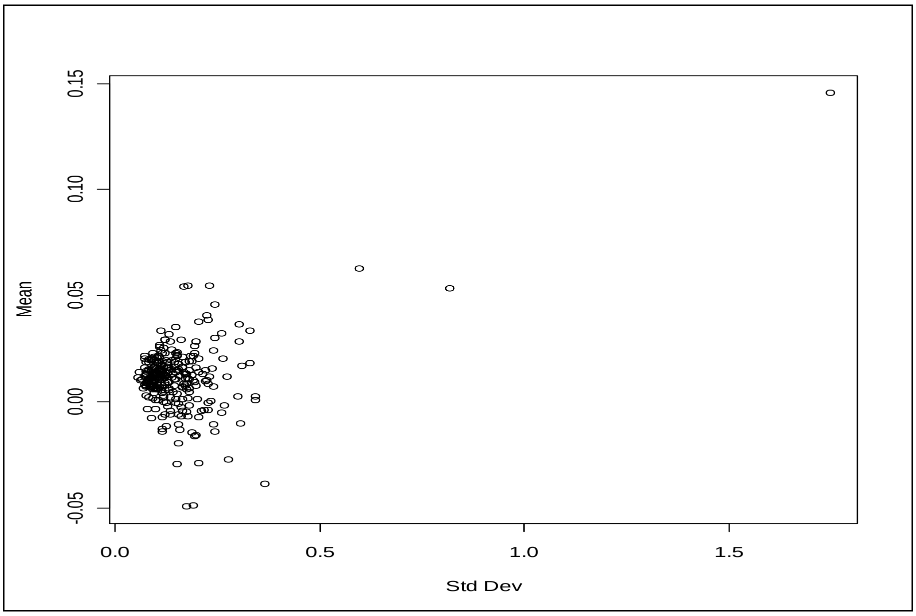
$$\mu_i = \beta\sigma_i \tag{1.1}$$

where $\mu_i = E\left[X_{it}\right]$ and $\sigma_i = \sqrt{E\left[X_{it} - \mu_i\right]^2}$, $i = 1,...,n$ $t = 1,...,T$. So all random variables will have a common β coefficient, which is the parameter of interest in this paper. Such a setting may be justified by Figure 1 presenting the relation between the sample mean and standard deviation of 285 stocks of the Stockholm exchange market from 1995-2010. In this paper we will consider estimating β in the context of financial application to Stockholm stock exchange. There are two characteristics of this type of data. Firstly, the number of variable n and the number of observations T are usually of comparable size. hence the asymptotic property of any estimator of β should allow for n and T to be proportional, say $n/T = c$, even asymptotically. Secondly, from pricing theory (Roll, 1976; Roll and Ross, 1980) such data are frequently assumed to follow a factor structure (Mardia et al, 1979) in the sense that

$$\mathbf{x}_t - \boldsymbol{\mu} = \mathbf{L}_{n\times k}\mathbf{F}_{k\times 1} + \boldsymbol{\varepsilon}_t \tag{1.2}$$

where $Cov\left[\mathbf{x}_t\right] = \boldsymbol{\Sigma}_{(n\times n)} = \mathbf{LL}' + \boldsymbol{\psi}$, $E\left[\mathbf{x}_t\right] = \boldsymbol{\mu}$, $E\left[\left(\mathbf{x}_t - \boldsymbol{\mu}\right)\left(\mathbf{x}_s - \boldsymbol{\mu}\right)\right] = \mathbf{0}$ $t \neq s$, $\boldsymbol{\psi}$ is a diagonal matrix. The dimension of the factor model, *k*, is usually taken to be fixed but we will further on relax this assumption to let *k* grow asymptotically with *n*. Hence the sum of all elements of $\boldsymbol{\Sigma}$ is given by $\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1} = \mathbf{1}'\left(\mathbf{LL}+\boldsymbol{\psi}\right)\mathbf{1} = \mathbf{1}'\mathbf{LL}'\mathbf{1} + \mathbf{1}'\boldsymbol{\psi}\mathbf{1} = \mathrm{O}\left(kn\right) + \mathrm{O}\left(n\right) = \mathrm{O}\left(kn\right)$. This should be compared to the $\mathrm{O}\left(p^2\right)$ that would result without the factor structure. Hence, the assumed underlying factor structure implies that the covariance matrix may be expressed with a reduced dimension. Thereby the number of elements in the variance of any statistic that is a function of all the random variables in the cross-section, will be reduced from n^2 elements to less or equal to n elements. This fact will have an important impact on the convergence rate of the variance of each estimators of β. Hence the variance of the estimators will limit zero much faster as n limits infinity than for the case of a full rank covariance matrix.

Figure 1: Mean values and std. deviations of all stocks from the Stockholm stock exchange.



In order to identify a family of estimators of β let $\bar{X}_i$ and S_i denote the sample mean and standard deviation respectively, expand both sides of (1.1) with $\bar{X}_i + \beta S_i$ and then solve the lhs for sample mean. This gives

$$\bar{X}_i = \beta S_i + \nu_i + \eta_i \tag{1.3}$$

where $v_i = \bar{X}_i - \mu_i$ and $\eta_i = \beta\left(\sigma_i - S_i\right)$. It should be evident that there are a number of possible estimators of β from this setting. In the following three alternative estimators will be considered with respect to their large sample properties. By using the fact that $p\lim\left[v_i\right] = p\lim\left[\eta_i\right] = 0$ as $T \to \infty$ it is obvious that $\bar{X}_i / S_i \xrightarrow{p} \mu_i / \sigma_i = \beta$. Hence an intuitive estimator may be defined by

$$\hat{\beta}_I = n^{-1}\sum_{i=1}^{n}\left(\bar{X}_i / S_i\right). \tag{1.4}$$

By using the same argument, an alternative estimator may be defined by

$$\hat{\beta}_{II} = \frac{n^{-1}\sum_{i=1}^{n}\bar{X}_i}{n^{-1}\sum_{i=1}^{n}S_i}. \tag{1.5}$$

A third possibility may be obtained by considering (1.1) as a regression model and define a pseudo-least square solution by

$$\hat{\beta}_{III} = \frac{\sum_{i=1}^{n}\bar{X}_i S_i}{\sum_{i=1}^{n}S_i^2}. \tag{1.6}$$

Other possible estimators of β involves weighted least square estimates (as (1.3) is a heteroscedastic model) or GMM estimators. In this paper, however, we will restrict ourselves to (1.4) - (1.6) and their large sample properties in the context discussed above.

In order to investigate the asymptotic properties of (1.4) – (1.6) it should be stressed that random vector $\mathbf{x}_t$ is observed over time and the dimension of the random vector is not a fixed quantity. Hence, the asymptotic situation at hand is depending of two dimensions, observations in the time dimension and the number of random variables in the cross-sectional dimension, in contrast to the traditional context where the cross-sectional dimension is fixed. We will mainly look at the convergence of the suggested estimators in L_2-norm, which also imply convergence in probability.

Initially the random vector at hand, $\mathbf{x}_t$, will be assumed to follow a multivariate normal distribution, but this will be relaxed further on. In a normal distribution setting the statistic based on the ratio between the sample mean and standard deviation is well known to have a non-central student's *t*-distribution. First, we note that the L_2-norm of an estimator of β can be written as follows:

$$E\left|\hat{\beta}_T - \beta\right|^2 = Var\left(\hat{\beta}_T\right) + Bias^2\left(\hat{\beta}_T\right).$$

By the properties of the *t*-distribution the asymptotic variance and bias of (1.4) as n/T approach a constant $c \in (0,1)$, are derived in Appendix 2. Hence $\hat{\beta}_I$ is asymptotically unbiased. The variance of estimator (1.4) converge to zero as we increase both time dimension and the cross-sectional dimension, so that the ratio belonging to the open interval between zero and one.

$$V\left(\hat{\beta}_I\right) \le o\left(Tn^{3/2}\right) Max\left\{\lambda_i \mathbf{1}_n' \boldsymbol{\gamma}_{(i)} \boldsymbol{\gamma}_{(i)}' \mathbf{1}_n\right\}_{i=1}^{\sqrt{n}} + o\left(Tn^2\right) \to 0 \text{ as } n \to \infty, T \to \infty.$$

we have that $E\left|\hat{\beta}_I - \beta\right|^2 \to 0$ as $n \to \infty, T \to \infty$ and $n \le T$, which implies that $\hat{\beta}_I - \beta = o_p(1)$.

Moreover, by appendix A3 it is shown that $n^{-1}\sum_{i=1}^n \bar{X}_i \xrightarrow{L_2} \beta n^{-1}\sum_{i=1}^n \sigma_i$ and $n^{-1}\sum_{i=1}^n S_i \xrightarrow{L_2} n^{-1}\sum_{i=1}^n \sigma_i$. Hence $\hat{\beta}_{II} - \beta = o_p(1)$.

Finally, by appendix A4 it is shown that:

$$n^{-1}\sum_{i=1}^n \bar{X}_i S_i \xrightarrow{L_2} \beta n^{-1}\sum_{i=1}^n \sigma_i^2 \text{ and } n^{-1}\sum_{i=1}^n S_i^2 \xrightarrow{L_2} n^{-1}\sum_{i=1}^n \sigma_i^2 \text{, so that } \hat{\beta}_{III} - \beta = o_p(1).$$

Hence all three estimators converge in probability to β as T limits infinity though some terms depends also on the size in the cross-section. The assumed model here, however, is unrealistic since it claims that all stocks share the exact same relation between the mean and standard deviation. As a consequence, it would be sufficient with one stock to estimate β , which in turn appears unrealistic.

Therefore an alternative model which allows the relations/points to fall away from the line with slope β, should be considered. A reasonable model is given by

$$X_{i,t} = \beta\sigma_i + \delta_i + \varepsilon_{i,t}, \tag{1.7}$$

where $\delta_i \sim N\left(0,\sigma_\delta^2\right)$, $\varepsilon_{i,t} \sim N\left(0,\sigma_i^2\right)$ and $\varepsilon_{i,t}$ and δ_i are individually and mutually independent. This implies that the conditional marginal distribution of each observation is:

$$X_{i,t}\left|\tilde{\delta}_i \sim N\left(\beta\sigma_i + \tilde{\delta}_i,\, \sigma_i^2\right),\right. \tag{1.8}$$

where the factor structure is now defined by

$$Cov\left[\mathbf{x}_t\middle|\boldsymbol{\delta}\right] = \boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\psi}\ ,\ E\left[\mathbf{x}_t\right] = \boldsymbol{\mu}\ \forall t,\ E\left[\left(\mathbf{x}_t - \boldsymbol{\mu}\right)\left(\mathbf{x}_s - \boldsymbol{\mu}\right)\right] = \mathbf{0}\,\forall\ t \neq s.$$

Thus the conditional expectation of each observation consists of two parts where one is the functional form $\beta\sigma_i$ seen in the former section and the second part is a random term $\tilde{\delta}_i$ that allows observations to deviate from the common beta relation. Hence the conditioned expectation is a random variable.

The bias of **first** estimator (1.4) is (for proof see appendix A5):

$$Bias\left(\hat{\beta}_I\right) = \beta\frac{\Gamma\left[(T-2)/2\right]}{\Gamma\left[(T-1)/2\right]}\sqrt{\frac{T-1}{2}} + \frac{\Gamma\left[(T-2)/2\right]}{\Gamma\left[(T-1)/2\right]}\sqrt{\frac{T-1}{2}}n^{-1}\sum_{i=1}^{n}\frac{\delta_i}{\sigma_i} - \beta =$$

$$\beta o(T) + \left(1 + o(T)\right)o\left(n^{1/2}\right) \to 0 \text{ as } n \to \infty, T \to \infty, \tag{1.9}$$

hence the first estimator is asymptotic unbiased as *n* and *T* limits infinity and the variance of (1.4) is:

$$Var\left(\hat{\beta}_I\right) = V\left(n^{-1}\sum_{i=1}^{n}\frac{\bar{X}_i}{S_i}\middle|\boldsymbol{\delta}\right) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov\left(\frac{\bar{X}_i}{S_i},\frac{\bar{X}_j}{S_j}\middle|\delta_i,\delta_j\right) =$$

$$\frac{1}{Tn^2}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov\left(\frac{\bar{X}_i}{S_i/\sqrt{T}},\frac{\bar{X}_j}{S_j/\sqrt{T}}\middle|\delta_i,\delta_j\right) = \frac{1}{Tn^2}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov\left(t_i',t_j'\middle|\delta_i,\delta_j\right) =$$

$$\frac{1}{Tn^2}\mathbf{1}'\boldsymbol{\Sigma}_{\mathbf{t}'}\mathbf{1} = \frac{1}{Tn^2}\mathbf{1}'\left(\boldsymbol{\Gamma}\tilde{\boldsymbol{\Lambda}}\boldsymbol{\Gamma}' + \boldsymbol{\Gamma}\boldsymbol{\Xi}\boldsymbol{\Gamma}'\right)\mathbf{1} = \frac{1}{Tn^2}\left(\sum_{i=1}^{\sqrt{n}}\lambda_i\mathbf{1}'\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}_{(i)}'\mathbf{1} + O(1)\right) \le$$

$$o(nT)\,\mathrm{Max}\left\{\lambda_i\mathbf{1}'\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}_{(i)}'\mathbf{1}\right\}_{i=1}^{n} + o(nT) \to 0 \text{ as } n\to\infty, T\to\infty . \tag{1.10}$$

Hence, from (1.9) and (1.10) we have that $E\left|\hat{\beta}_I-\beta\right|^2 \to 0$ as $n\to\infty, T\to\infty$, which implies that $\hat{\beta}_I-\beta = o_p(1)$.

For the case of the second suggested estimator, (1.5), the bias of numerator is (for proof see appendix A6):

$$Bias\left(n^{-1}\sum_{i=1}^{n}\bar{X}_iS_i\right) = E\left[n^{-1}\sum_{i=1}^{n}\bar{X}_iS_i\middle|\boldsymbol{\delta}\right] - \beta\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 = \frac{1}{n}\sum_{i=1}^{n}(\beta\sigma_i+\delta_i)\sigma_i\sqrt{\frac{2}{T-1}}\frac{\Gamma[T/2]}{\Gamma[(T-1)/2]} -$$

$$\beta\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 = o(T)\beta\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 + \left[1+o(T)\right]o\left(n^{1/2}\right) \to 0 \text{ as } n\to\infty,\ T\to\infty, \tag{1.11}$$

and the variance of the numerator is:

$$V\left(n^{-1}\sum_{i=1}^{n}\bar{X}_iS_i\middle|\boldsymbol{\delta}\right) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{i,j}\sqrt{V\left(\bar{X}_iS_i\left|\delta_i\right.\right)V\left(\bar{X}_jS_j\left|\delta_j\right.\right)} \le$$

$$n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sqrt{V\left(\bar{X}_iS_i\left|\delta_i\right.\right)}\sqrt{V\left(\bar{X}_jS_j\left|\delta_j\right.\right)} = o\left(n^{3/2}\right)Max\left\{\lambda_i\mathbf{1}'\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}_{(i)}'\right\}_{i=1}^{n} + o\left(n^2\right) =$$

$$o\left(n^{3/2}\right) \to 0 \text{ as } n\to\infty,\ T\to\infty . \tag{1.12}$$

So that the numerator converge in probability to $\beta\sum_{i=1}^{n}\sigma_i^2$. Further the bias of the denominator is zero and the variance of the denominator is:

$$V\left(n^{-1}\sum_{i=1}^{n}S_i^2\right)=\frac{1}{n^2T^2}V\left(vec\left(\mathbf{X}\right)'\,\tilde{\mathbf{H}}\,vec\left(\mathbf{X}\right)\right)=\frac{2}{n^2T^2}\left[tr\left(\tilde{\mathbf{H}}\boldsymbol{\Omega}\tilde{\mathbf{H}}\boldsymbol{\Omega}\right)+2\tilde{\boldsymbol{\mu}}'\tilde{\mathbf{H}}\boldsymbol{\Omega}\tilde{\mathbf{H}}\tilde{\boldsymbol{\mu}}\right]=$$

$$\frac{4}{n^2T^2}\left(\boldsymbol{\mu}'\otimes\mathbf{1}'\right)\left(\boldsymbol{\Sigma}\otimes\mathbf{H}\right)\left(\boldsymbol{\mu}\otimes\mathbf{1}\right)=\frac{2}{n^2T^2}tr\left(\boldsymbol{\Gamma\Lambda\Gamma}'\boldsymbol{\Gamma\Lambda\Gamma}'\right)tr\left(\mathbf{H}\right)+\frac{4}{n^2T^2}\left(\boldsymbol{\mu}'\boldsymbol{\Sigma\mu}\otimes\mathbf{1}'\mathbf{H1}\right)=$$

$$\frac{2}{n^2T^2}tr\left(\boldsymbol{\Gamma\Lambda}^2\boldsymbol{\Gamma}'\right)\left(T-1\right)=\frac{2}{n^2T^2}\left(\sum_{i=1}^{\sqrt{n}}\lambda_i^{\,2}+O\left(1\right)\right)\left(T-1\right)\le o\left(n^{3/2}T\right)2Max\left\{\lambda_i^{\,2}\right\}_{i=1}^{n}\to 0 \text{ as}$$

$$n\to\infty,\, T\to\infty\,, \tag{1.13}$$

where $\tilde{\mathbf{H}}$ is a block matrix consisting of the centering matrix $\mathbf{H}_{(T\times T)}$ in each diagonal block.

Hence the dominator converge in probability to $\sum_{i=1}^{n}\sigma_i^2$, and so $\hat{\beta}_{II}-\beta=o_p\left(1\right)$.

Finally, for the case of the third suggested estimator, (1.6), the bias of the numerator is (for proof see appendix A7):

$$Bias\left(n^{-1}\sum_{i=1}^{n}\bar{X}_i\right)=E\left[n^{-1}\sum_{i=1}^{n}\bar{X}_i\middle|\boldsymbol{\delta}\right]-\beta n^{-1}\sum_{i=1}^{n}\sigma_i=\left(Tn\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}E\left[X_{i,t}\left|\delta_i\right.\right]-\beta n^{-1}\sum_{i=1}^{n}\sigma_i=$$

$$\left(Tn\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\beta\sigma_i+\delta_i\right)-\beta n^{-1}\sum_{i=1}^{n}\sigma_i=\bar{\delta}\to 0 \text{ as } n\to\infty, T\to\infty\,, \tag{1.14}$$

and the variance of the numerator is:

$$Var\left(n^{-1}\sum_{i=1}^{n}\bar{X}_i\middle|\boldsymbol{\delta}\right)=n^{-2}Var\left(\mathbf{1}_n'\bar{\mathbf{x}}\right)=n^{-2}\mathbf{1}_n'Var\left(\bar{\mathbf{x}}\right)\mathbf{1}_n=n^{-2}\mathbf{1}_n'\left(T^{-1}\boldsymbol{\Sigma}_{\mathbf{x}}\right)\mathbf{1}_n=$$

$$\left(Tn^2\right)^{-1}\left(\sum_{i=1}^{\sqrt{n}}\lambda_i\mathbf{1}_n'\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}_{(i)}'\mathbf{1}_n+O\left(1\right)\right)\le\left(Tn^{3/2}\right)^{-1}\lambda_{\max}\mathbf{1}_n'\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}_{(i)}'\mathbf{1}_n+\left(Tn^2\right)^{-1}O\left(1\right)\to 0 \text{ as}$$

$$n\to\infty,\; T\to\infty. \tag{1.15}$$

Hence the numerator converge in probability to $\beta\sum_{i=1}^{n}\sigma_i$. Further, the bias of the denominator is:

$$Bias\left(n^{-1}\sum_{i=1}^{n}S_i\right)=n^{-1}E\left[tr\left\{\mathbf{S}^{1/2}\right\}\right]-n^{-1}tr\left\{\boldsymbol{\Sigma}^{1/2}\right\}=\left[1+o\left(T\right)\right]n^{-1}tr\left\{\boldsymbol{\Sigma}^{1/2}\right\}-n^{-1}tr\left\{\boldsymbol{\Sigma}^{1/2}\right\}=$$

$$o\left(T\right)n^{-1}tr\left\{\sum_{i=1}^{\sqrt{n}}\lambda_i^{1/2}\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}'_{(i)}+O\left(1\right)\right\}=o\left(Tn^{\delta}\right)\to 0 \quad \text{when} \quad n\to\infty, \quad T\to\infty \quad \text{and}$$

$$\delta\in\left]0,1/2\right[, \tag{1.16}$$

and the variance of the denominator is:

$$Var\left(n^{-1}\sum_{i=1}^{n}S_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}Var\left(S_i\right)+\frac{1}{n^2}\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n}Cov\left(S_i,S_j\right)\leq o\left(Tn^2\right)\sum_{i=1}^{n}\sigma_i^2+$$

$$o\left(Tn^2\right)\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n}\sigma_i\sigma_j\to 0 \text{ when } n\to\infty, T\to\infty. \tag{1.17}$$

Thus the dominator converge in probability to $\sum_{i=1}^{n}\sigma_i$, and consequently $\hat{\beta}_{III}-\beta=o_p\left(1\right)$.

Thus, all three estimators converge in probability to β as both *n* and *T* limits infinity, and all three estimators are asymptotically unbiased. In order to demonstrate the potential of the three estimators, we will in the following section apply the estimators on a real dataset and also investigate if the relation (1.3) is constant between different groups of stocks.

3. An empirical application – Sharpe ratio for the Stockholm stock exchange.

This empirical analysis is performed on the whole population of stocks listed on the Stockholm stock exchange over the period of Jun 1995-Jun 2010, which are retrieved from the *Data Stream*. The data set consist of monthly excess returns which is calculated as follows:

$$\tilde{r}_j = r_j - r_f \qquad (2.1)$$

where $\tilde{r}_j :=$ excess return, $r_j :=$ monthly return on the j:th asset and $r_f :=$ risk-free return (monthly return on 3-months treasure bills).

The population is divided into three subpopulation based on the market capitalization of each company by following Stockholm stock exchange segmentation of stocks into the three different segments: Large Cap, Mid Cap and Small Cap. The subpopulations/segments of Large Cap and the Mid Cap consist of 77 stocks, respectively, and the subpopulation/segment Small Cap consists of 131 stocks. Of these stocks only 44 Large Cap stocks has a complete history of observations, whereas 21 Mid Cap stocks and 30 Small Cap stocks have complete history of observations. A non-complete history of the observations over the sample period can be, for example, due to a company which might have been acquired by another company (that may be either listed or not on the stock exchange), resulting in a sudden end to the series of the observation. The same pattern can prevail if a company goes bankrupt due to some financial distress. It is also the case of having non-complete history of observation when new companies enter into the stock exchange, producing series of observation which cannot initially be observed.

Now, the question is how well the three suggested estimators, (1.4) – (1.6), fit the data set at hand and if there is homogeneity for β among different subpopulations/segments in the market, i.e. if the β_i differ among the different segments $i = 1, 2, 3$.

By first studying the scatter plots, 2a to 4b, we find that there is an upward trend for all three subpopulations, indicating a positive relationship between mean excess returns and standard deviation of excess returns. Hence, as the risk (standard deviation of excess return) increases so does the mean excess return. Another interesting observation is that the two subpopulations Mid Cap (Figure 3a) and Small Cap

(Figure 4a) have both two extreme observations, where for Small Cap the most extreme observation has nearly five times higher standard deviation than the rest of the non extreme observations.

Figure 2a: Scatter plot of Large Cap.

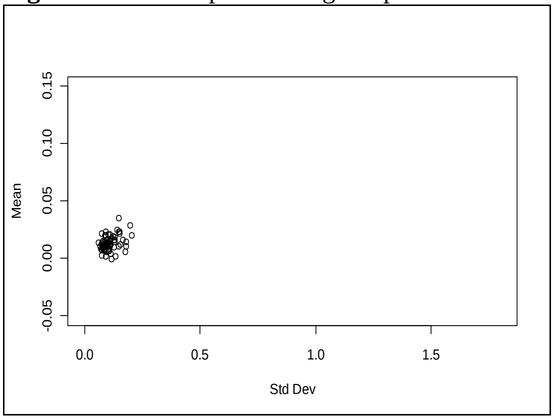


Figure 2b: Scatter plot of Large Cap.

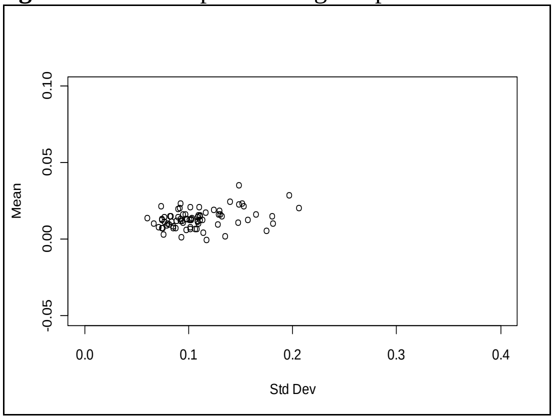


Figure 3a: scatter plot of Mid Cap.

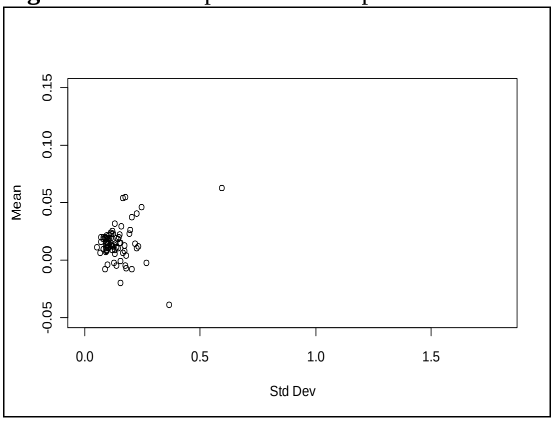


Figure 3a: scatter plot of Mid Cap.

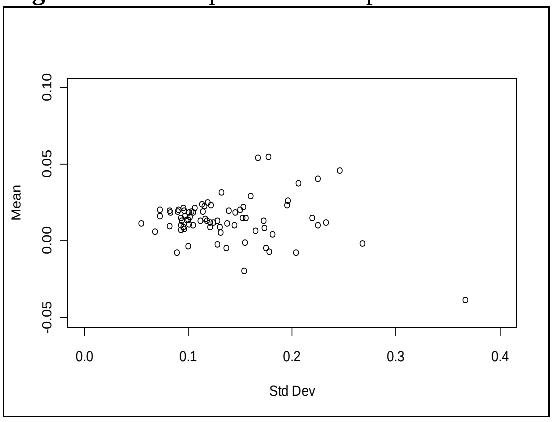


Figure 4a: scatter plot of Small Cap.

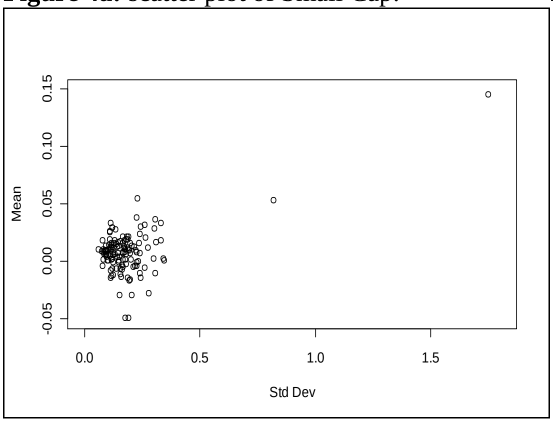
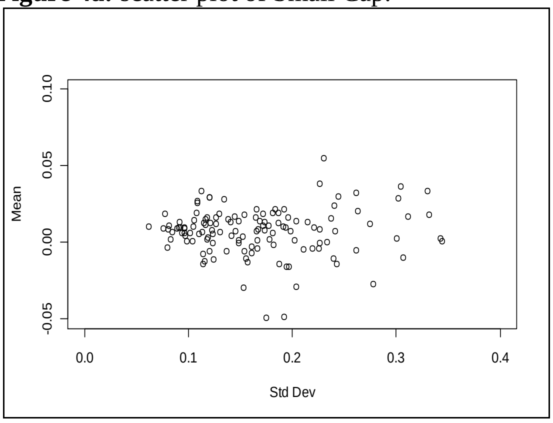


Figure 4a: scatter plot of Small Cap.



The positive linear relation between mean excess returns and standard deviation of excess returns is also supported by the result of the three estimators for the three different subpopulations as seen in table 1. In general, the estimated β diminishes for all three estimators as we shift market segment from Large Cap to Mid Cap and Mid Cap to Small Cap. Clearly, the β is not homogenous between the three market segments as seen in table 1, which violates the assumption of a common beta for all stocks. An additional result is that for all segments the two first estimators gives approximately the same value, while the third estimator is smaller, except for the segment Small Cap of which the third estimator is larger than the other two estimators. It turns out that if the two extreme observations for the Small Cap segment is excluded, then the third estimator is smaller than the other two estimators, as seen in table 2. Hence, it is evident that the third estimator is more sensitive to extreme values. In general, the third estimator tends to estimate a lower β than the two other estimators.

The reason why the three estimators differ in all segments is due to the fact that all three estimators are asymptotically unbiased for an infinite population, but the population used here is finite.

Table 1: Estimates of the Sharpe ratio.

Segment	β_I	β_{II}	β_{III}
Large Cap	0.127972	0.1236918	0.1195114
Mid Cap	0.1180131	0.1042785	0.0898506
Small Cap	0.0516006	0.04705	0.0583177

Table 2: Estimates of the Sharpe ratio when extreme values are excluded.

Segment	β_I	β_{II}	β_{III}
Mid Cap	0.1211573	0.1118937	0.1037336
Small Cap	0.0512375	0.0434029	0.0387099

4. Summary.

In this paper the relation of the second central moment of a distribution to its first raw moment is investigated in an financial context. According to the theory of finance, an investor is compensated by increased returns for taking higher risks (i.e. variance of the investments returns), which is represented by the Sharpe ratio μ/σ. In order to make inference of Sharpe ratio we define two models. The first model assumes that the relation between risks and returns is exactly linear, making all the observations fall on an straight line, while the later model allows the observations to deviate unsystematically from the straight line. Three estimators in this case for estimating the Sharpe ratio are suggested.

It is been widely observed that there are two characteristics of the financial data. Firstly, the number of the variables (number of assets in the cross-section), *n*, and the number of observations over time, *T*, are usually of comparable sizes, so the asymptotic property of any estimator of β should allow *n* and *T* to be proportional. Secondly, the data follow a factor structure based on pricing theory of financial assets. The impact of the later characteristic (the assumed underlying factor structure) implies that the covariance matrix may be expressed with a reduced dimension, therefore, the number of elements in the variance of any statistic that is a function of all the random variables in the cross-section is reduced. This fact has an important impact on the convergence rate of the variance of each estimators of β. Under these settings all three estimators are proven to be consistent as both the number of observations and number of stocks limit infinity.

A case study is also conducted in this study. The whole population of stocks listed on the Stockholm stock exchange during Jun 1995-Jun 2010 is included in the case study. The population was divided into three subpopulations depending on the market capitalization value of each company listed on the Stockholm stock exchange. It is empirically shown that the Sharpe ratio for each subpopulation are heterogeneous. Moreover, it is also empirically shown that the third estimator suffers more from extreme values than the two other estimators.

Appendix

A1:

The expected value of the first estimator, (1.4), with only one observation is:

$$E\left[\hat{\beta}_i\right]=E\left[\bar{X}_iS_i^{-1}\right]=E\left[\bar{X}_iS_i^{-1}-\mu_iS_i^{-1}+\beta\sigma_iS_i^{-1}\right]=E\left[\bar{X}_iS_i^{-1}\right]-\mu_iE\left[S_i^{-1}\right]+\beta\sigma_iE\left[S_i^{-1}\right] \tag{3.1}$$

From Basu et al (1975), we have that sample mean is independent of any translation invariant statistic, hence the expected value of (3.1) becomes:

$$E\left[\hat{\beta}_i\right]=E\left[\bar{X}_i\right]E\left[S_i^{-1}\right]-\mu_iE\left[S_i^{-1}\right]+\beta\sigma_iE\left[S_i^{-1}\right]=\beta\sigma_iE\left[S_i^{-1}\right] \tag{3.2}$$

Since $X_i \sim N\left(\mu_i,\sigma_i^2\right)$, the sample variance is $S^2 \sim \frac{\sigma^2}{n-1}\chi^2_{(n-1)}$ and $S^2\frac{n-1}{\sigma^2} \sim \chi^2_{(n-1)}$.

So that the expected value of S_i^{-1} is: $E\left[S^{-1}\right]=E\left[\left(S^2\right)^{-\frac{1}{2}}\right]=$

$$E\left[\left(S^2\frac{T-1}{\sigma^2}\frac{\sigma^2}{(T-1)}\right)^{-\frac{1}{2}}\right]=E\left[\left(\frac{T-1}{\sigma^2}\right)^{\frac{1}{2}}\left(S^2\frac{T-1}{\sigma^2}\right)^{-\frac{1}{2}}\right]=\sqrt{\frac{T-1}{\sigma^2}}E\left[Y^{-\frac{1}{2}}\right] \tag{3.3}$$

Due to Y is chi-square distributed, we have the following:

$$E\left[Y^{-\frac{1}{2}}\right]=\int_0^\infty Y^{-\frac{1}{2}}\frac{(1/2)^{\left(\frac{T-1}{2}\right)}}{\Gamma\left[(T-1)/2\right]}y^{\frac{T-1}{2}-1}e^{-\lambda y}dy=\int_0^\infty\frac{(1/2)^{\left(\frac{T-1}{2}\right)}}{\Gamma\left[(T-1)/2\right]}y^{\frac{T-2}{2}-1}e^{-\lambda y}dy=$$

$$\frac{\Gamma\left[(T-2)/2\right](1/2)^{\left(\frac{1}{2}\right)}}{\Gamma\left[(T-1)/2\right]}\int_0^\infty\frac{(1/2)^{\left(\frac{T-2}{2}\right)}}{\Gamma\left[(T-2)/2\right]}y^{\frac{T-2}{2}-1}e^{-\lambda y}dy=\frac{\Gamma\left[(T-2)/2\right]}{\sqrt{2}\,\Gamma\left[(T-1)/2\right]} \tag{3.4}$$

By substituting (3.4) into (3.3), this gives:

$$E\left[S^{-1}\right]=\sqrt{\frac{T-1}{\sigma^2}}E\left[Y^{-\frac{1}{2}}\right]=\sqrt{\frac{T-1}{\sigma^2}}\frac{\Gamma\left[(T-2)/2\right]}{\sqrt{2}\,\Gamma\left[(T-1)/2\right]}=\sigma^{-1}\sqrt{\frac{T-1}{2}}\frac{\Gamma\left[(T-2)/2\right]}{\Gamma\left[(T-1)/2\right]}, \tag{3.5}$$

so that: $E\left[S_i^{-1}\right]=\sigma^{-1}\sqrt{\frac{T-1}{2}}\frac{\Gamma\left[(T-2)/2\right]}{\Gamma\left[(T-1)/2\right]}$. (3.6)

Finally, by substituting (3.6) into (3.2) we then have following:

$$E\left[\hat{\beta}_i\right]=\beta\sigma_iE\left[S_i^{-1}\right]=\beta\sigma_i\frac{\sqrt{T-1}}{\sigma_i\sqrt{2}}\frac{\Gamma\left[(T-2)/2\right]}{\Gamma\left[(T-1)/2\right]}=\beta+o(T). \tag{3.7}$$

The variance of the of the first estimator, (1.4), with only one observation is:

$$V\left(\hat{\beta}_i\right) = E\left[\left(\hat{\beta}_i - E\left[\hat{\beta}_i\right]\right)^2\right] = E\left[\hat{\beta}_i^2\right] - E^2\left[\hat{\beta}_i\right], \tag{3.8}$$

where: $E\left[\hat{\beta}_i^2\right] = E\left[\left(\bar{X}_i\left(S_i\right)^{-1} - \mu_i\left(S_i\right)^{-1} + \beta\sigma_i\left(S_i\right)^{-1}\right)^2\right] =$

$$E\left[\left\{\bar{X}_i^2 - 2\mu_i\bar{X} + 2\beta\sigma_i\bar{X}_i - 2\beta\sigma_i\mu + \mu_i^2 + \beta^2\sigma_i^2\right\}S_i^{-2}\right]. \tag{3.9}$$

Further, by independence between $\bar{X}_i$, S_i^{-1}, (3.9) reduces to:

$$E\left[\hat{\beta}_i^2\right] = E\left[\bar{X}_i^2 - 2\mu_i\bar{X} + 2\beta\sigma_i\bar{X}_i - 2\beta\sigma_i\mu + \mu_i^2 + \beta^2\sigma_i^2\right]E\left[S_i^{-2}\right]$$

$$\left\{E\left[\bar{X}_i^2\right] - 2\mu_i^2 + 2\beta\sigma_i\mu - 2\beta\sigma_i\mu + \mu_i^2 + \beta^2\sigma_i^2\right\}E\left[S_i^{-2}\right] =$$

$$\left\{\left(\frac{\sigma_i^2}{T} + \mu_i^2\right) - \mu_i^2 + \beta^2\sigma_i^2\right\}E\left[S_i^{-2}\right] = \left\{\frac{\sigma_i^2}{T} + \beta^2\sigma_i^2\right\}E\left[S_i^{-2}\right] = \sigma_i^2\left\{\frac{1}{T} + \beta^2\right\}E\left[S_i^{-2}\right]. \tag{3.10}$$

To simplify (3.10), it is first needed to calculate $E\left[S_i^{-2}\right]$. Let $Y = S^2\left(T-1\right)/\sigma^2$ then

$Y \sim \chi^2_{(T-1)}$ so that: $S^{-2} = \sigma^2\left(T-1\right)^{-1}Y^{-1}$. (3.11)

The expected value of (3.11) is: $E\left[S^{-2}\right] = E\left[\left(T-1\right)/Y\sigma^2\right] = \left[\left(T-1\right)/\sigma^2\right]E\left[Y^{-1}\right]$. (3.12)

Due to Y is chi-square distributed, the expected value of the inverse of *Y* is equal to:

$$E\left[Y^{-1}\right] = \int_0^\infty y^{-1}\frac{(1/2)^{\left(\frac{T-1}{2}\right)}}{\Gamma\left[\left(T-1\right)/2\right]}y^{\frac{n-1}{2}-1}e^{-\frac{1}{2}y}dy =$$

$$\frac{(1/2)}{\left[\left(T-1\right)/2-1\right]}\int_0^\infty \frac{(1/2)^{\left(\frac{T-1}{2}-1\right)}}{\Gamma\left[\left(T-1\right)/2-1\right]}y^{\frac{T-1}{2}-1-1}e^{-(1/2)y}dy = \frac{1}{T-3}. \tag{3.13}$$

Thus, by substituting (3.13) into (3.12) we then have following:

$$E\left[S^{-2}\right] = \sigma^{-2}\left[\left(T-1\right)/\left(T-3\right)\right]. \tag{3.14}$$

Hence (3.10) is then by (3.14) equal to: $E\left[\hat{\beta}_i^2\right] = \sigma_i^2\left\{1/T + \beta^2\right\}\sigma_i^{-2}\left(T-1\right)/\left(T-3\right) =$

$$(T-1)/T(T-3)+\beta^2(T-1)/(T-3)=o(T)+(1+o(T))\beta^2 \qquad (3.15)$$

Finally, from (3.15) and(3.7) the variance of $\hat{\beta}_i$ is:

$$V\left(\hat{\beta}_i\right)=E\left[\hat{\beta}_i^2\right]-E^2\left[\hat{\beta}_i\right]=\left(\frac{(T-1)}{T(T-3)}+\frac{(T-1)}{(T-3)}\beta^2\right)-\left(\beta\sqrt{\frac{T-1}{2}}\frac{\Gamma\left[(1/2)(T-2)\right]}{\Gamma\left[(1/2)(T-1)\right]}\right)^2=$$

$$\beta^2\left(\frac{(T-1)}{T(T-3)}+\frac{(T-1)}{(T-3)}-\frac{(T-1)}{2}\frac{\Gamma^2\left[(1/2)(T-2)\right]}{\Gamma^2\left[(1/2)(T-1)\right]}\right)=$$

$$\beta^2\left(\frac{T-1}{T(T-3)}+\frac{T-1}{T-3}-\frac{T-1}{2}\frac{\Gamma^2\left[(1/2)(T-2)\right]}{\frac{T-3}{2}\Gamma\left[(1/2)(T-1)\right]\Gamma\left[(1/2)(T-3)\right]}\right)=$$

$$\beta^2\frac{T-1}{T-3}\left(\frac{1}{T}+1-\frac{\Gamma^2\left[(1/2)(T-2)\right]}{\Gamma\left[(1/2)(T-1)\right]\Gamma\left[(1/2)(T-3)\right]}\right)=o(T)$$

A2:

The first estimator, (1.4), can be rewritten into the following form:

$$\hat{\beta}_I=n^{-1}\sum_{i=1}^{n}\frac{\bar{X}_i}{S_i}=n^{-1}\sum_{i=1}^{n}\frac{1}{\sqrt{T}}\frac{\bar{X}_i}{S_i/\sqrt{T}}=\frac{1}{n\sqrt{T}}\sum_{i=1}^{n}t_i. \qquad (4.1)$$

From (3.7) it follows that the bias of (4.1) is:

$$Bias\left(\hat{\beta}_I\right)=E\left[\hat{\beta}_I\right]-\beta=n^{-1}\sum_{i=1}^{n}E\left[\frac{\bar{X}_i}{S_i}\right]-\beta=$$

$$\beta\left(n^{-1}\sum_{i=1}^{n}\frac{\Gamma\left[(T-2)/2\right]}{\Gamma\left[(T-1)/2\right]}\sqrt{\frac{T-1}{2}}-1\right)\leq o(T). \qquad (4.2)$$

Assuming that $\mathbf{x}_t$ is multivariate normal distributed, this implies that $\bar{X}_i, S_i$ are independent and then t_i' in (4.1) has a non-central student's t distribution with T degree of freedom and non-centrality parameter μ_i. From Rohatgi (1976) follows that:

$$E\left[t_i'\right]=(T/2)^{1/2}\frac{\Gamma\left[(T-1)/2\right]}{\Gamma\left[T/2\right]}\mu_i,\ \ V\left(t_i'\right)=\frac{T\left(1+\mu_i^2\right)}{T-2}-\frac{T}{2}\frac{\Gamma^2\left[(T-1)/2\right]}{\Gamma^2\left[T/2\right]}\mu_i^2, \qquad (4.3)$$

and the variance of (4.1) is by the assumed underlying factor structure equal to:

$$V\left(\hat{\beta}_I\right)=\frac{1}{n^2}V\left(\sum_{i=1}^{n}T^{-1/2}t_i'\right)=\frac{1}{Tn^2}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov\left(t_i',t_j'\right)=\frac{1}{Tn^2}\mathbf{1}'\Sigma_{\mathbf{t}}\mathbf{1}=$$

$$\frac{1}{Tn^2}\mathbf{1}'\left(\sum_{i=1}^{\sqrt{n}}\lambda_i\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}_{(i)}'+\sum_{i=1+\sqrt{n}}^{n}\lambda_i\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}_{(i)}'\right)\mathbf{1}=\frac{1}{Tn^2}\left(\sum_{i=1}^{\sqrt{n}}\lambda_i\mathbf{1}_n'\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}_{(i)}'\mathbf{1}_n+o(1)\right)\leq$$

$$o\left(Tn^{3/2}\right)Max\left\{\lambda_i\mathbf{1}_n'\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}_{(i)}'\mathbf{1}_n\right\}_{i=1}^{\sqrt{n}}+o\left(Tn^2\right). \tag{4.4}$$

Hence, the variance of the mean naïve estimator limits zero as $n\to\infty, T\to\infty$. Thus, from (4.2) and (4.4) we have that $\text{plim}\left(\hat{\beta}_I\right)=\beta$.

A3:

The third estimator, (1.6), can be rewritten as follows:

$$\hat{\beta}_{III}=\sum_{i=1}^{n}\bar{X}_iS_i\Big/\sum_{i=1}^{n}S_i^2=n^{-1}\sum_{i=1}^{n}\bar{X}_iS_i\Big/n^{-1}\sum_{i=1}^{n}S_i^2. \tag{5.1}$$

To derive the expected value and variance of (5.1), we first establish following:

Let $S=\sqrt{S^2}$ and $Y=S^2\dfrac{T-1}{\sigma^2}$ so that $Y\in\chi^2_{(T-1)}$, then we have: $S=\sqrt{\dfrac{\sigma^2}{T-1}}\sqrt{Y}$.

(5.2)

The expected value of (5.2) is:

$$E[S]=E\left[\sqrt{\frac{\sigma^2}{T-1}}\sqrt{S^2\frac{T-1}{\sigma^2}}\right]=\sqrt{\frac{\sigma^2}{T-1}}E\left[Y^{1/2}\right]=$$

$$\sqrt{\frac{\sigma^2}{T-1}}\int_0^\infty y^{1/2}\frac{(1/2)^{\left(\frac{T-1}{2}\right)}}{\Gamma\left((T-1)/2\right)}y^{\frac{T-1}{2}-1}e^{-(1/2)y}dy=$$

$$\sqrt{\frac{\sigma^2}{T-1}}\frac{(1/2)^{(-0.5)}\Gamma[T/2]}{\Gamma[(T-1)/2]}\int_0^\infty y^{1/2}\frac{(1/2)^{(T/2)}}{\Gamma[T/2]}y^{\frac{T}{2}-1}e^{-(1/2)y}dy=\sigma\sqrt{\frac{2}{(T-1)}}\frac{\Gamma(T/2)}{\Gamma\left((T-1)/2\right)}. \tag{5.3}$$

The expected value of the numerator in (5.1) is, by (5.3) and independence between $\bar{X}_i$ and S_i, equal to:

$$E\left[\frac{1}{n}\sum_{i=1}^{n}\bar{X}_iS_i\right]=\frac{1}{n}\sum_{i=1}^{n}E\left[\bar{X}_iS_i\right]=\frac{1}{n}\sum_{i=1}^{n}E\left[\bar{X}_i\right]E\left[S_i\right]=\frac{1}{n}\sum_{i=1}^{n}\mu_i\sigma_i\sqrt{\frac{2}{T-1}}\frac{\Gamma[T/2]}{\Gamma[(T-1)/2]}. \tag{5.4}$$

By substituting (1.1) into (5.4) the expected value of the numerator becomes:

$$E\left[\frac{1}{n}\sum_{i=1}^{n}\bar{X}_i S_i\right]=\beta\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2\sqrt{\frac{2}{(T-1)}}\frac{\Gamma(T/2)}{\Gamma((T-1)/2)}, \tag{5.5}$$

so the asymptotic properties of the bias of the numerator in (5.1) is:

$$E\left[\frac{1}{n}\sum_{i=1}^{n}\bar{X}_i S_i\right]-\beta\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2=\left(\sqrt{\frac{2}{(T-1)}}\frac{\Gamma(T/2)}{\Gamma((T-1)/2)}-1\right)\beta\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2\le o(T)\beta Max\left\{\sigma_i^2\right\}_{i=1}^{n}\to 0$$

as $n\to\infty, T\to\infty$. (5.6)

The variance of the numerator in (5.1) is:

$$V\left(\frac{1}{n}\sum_{i=1}^{n}\bar{X}_i S_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}V\left(\bar{X}_i S_i\right)+\frac{1}{n^2}\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n}Cov\left(\bar{X}_i S_i,\bar{X}_j S_j\right), \tag{5.7}$$

where each of the variances in the rhs of (5.7) are by equal to:

$$V\left(\bar{X}_i S_i\right)=E\left[\left(\bar{X}_i S_i-E\left[\left(\bar{X}_i S_i\right)\right]\right)^2\right]=E\left[\bar{X}_i^2 S_i^2\right]-E^2\left[\bar{X}_i S_i\right]=$$

$$E\left[\bar{X}_i^2\right]E\left[S_i^2\right]-E^2\left[\bar{X}_i\right]E^2\left[S_i\right]=\left\{E\left[\bar{X}_{i,t}^2\right]=\frac{\sigma_i^2}{T}+\mu_i^2, E\left[X_{i,t}X_{i,\tau}\right]=\mu_i^2\right\}=$$

$$\left(\frac{\sigma_i^2}{T}+\mu_i^2\right)\sigma_i^2-\mu_i^2\sigma_i^2\frac{2}{T-1}\frac{\Gamma^2(T/2)}{\Gamma^2((T-1)/2)}=\frac{\sigma_i^4}{T}+\mu_i^2\sigma_i^2\left(1-\frac{2}{T-1}\frac{\Gamma^2(T/2)}{\Gamma^2((T-1)/2)}\right)=$$

$$=o(T)+\mu_i^2\sigma_i^2 o(T). \tag{5.8}$$

The first sum in the rhs of (5.7) is equal to:

$$\frac{1}{n^2}\sum_{i=1}^{n}V\left(\bar{X}_i S_i\right)=o(T)\frac{1}{n^2}\sum_{i=1}^{n}\mu_i^2\sigma_i^2+o(Tn)\le o(T)Max\left\{\mu_i^2\sigma_i^2\right\}_{i,j}+o(Tn)\to 0 \text{ as}$$

$n\to\infty, T\to\infty$. (5.9)

Due to the fact that $|Cov(X,Y)|=|\rho_{X,Y}|\sigma_X\sigma_Y\le\sigma_X\sigma_Y$ by the Cauchy-Schwartz inequality (Casella et al,1990), we have that each of the covariances in the rhs of (5.7) is bounded above by:

$$Cov\left(\bar{X}_i S_i,\bar{X}_j S_j\right)\le\left(Var\left(\bar{X}_i S_i\right)\right)^{1/2}\left(Var\left(\bar{X}_j S_j\right)\right)^{1/2}, \tag{5.10}$$

Then by substituting (5.8) into (5.10) we get:

$$Cov\left(\bar{X}_iS_i,\bar{X}_jS_j\right)\le\left(o(T)+\mu_i^2\sigma_i^2o(T)\right)^{1/2}\left(o(T)+\mu_j^2\sigma_j^2o(T)\right)^{1/2}=\Big[o\left(T^2\right)+o\left(T^2\right)\mu_i^2\sigma_i^2+ o\left(T^2\right)\mu_j^2\sigma_j^2+o\left(T^2\right)\mu_i^2\sigma_i^2\mu_j^2\sigma_j^2\Big]^{1/2}=o(T)\left[1+\mu_i^2\sigma_i^2+\mu_j^2\sigma_j^2+\mu_i^2\sigma_i^2\mu_j^2\sigma_j^2\right]^{1/2}, \quad (5.11)$$

So that the second summand in (5.7) is bounded by:

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}Cov\left(\bar{X}_iS_i,\bar{X}_jS_j\right)\le o\left(n^2T\right)\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\left[1+\mu_i^2\sigma_i^2+\mu_j^2\sigma_j^2+\mu_i^2\sigma_i^2\mu_j^2\sigma_j^2\right]^{1/2}\le o(T)\left(1-o(n)\right)\left[1+Max\left\{\mu_i^2\sigma_i^2\right\}_{i,j}\right]\to 0 \text{ as } n\to\infty, T\to\infty. \quad (5.12)$$

Thus from (5.9) and (5.12) we have that the variance of the numerator limits:

$$V\left(\frac{1}{n}\sum_{i=1}^{n}\bar{X}_iS_i\right)\le o(T)Max\left\{\mu_i^2\sigma_i^2\right\}_{i,j}+o(Tn)+o(T)\left(1-o(n)\right)\left[1+Max\left\{\mu_i^2\sigma_i^2\right\}_{i,j}\right]= o(T)\to 0 \text{ as } n\to\infty, T\to\infty. \quad (5.13)$$

Thus, from (5.6) and (5.13) we have that $E\left|\frac{1}{n}\sum_{i=1}^{n}\bar{X}_iS_i-\frac{1}{n}\sum_{i=1}^{n}\mu_i\sigma_i\right|^2\to 0$ as $n\to\infty$, $T\to\infty$, which by substation of μ_i by (1.1) implies that $\text{plim}\left(\sum_{i=1}^{n}\bar{X}_iS_i\right)=\beta\sum_{i=1}^{n}\sigma_i^2.$

(5.14)

The expected value of the denominator in (5.1) is:

$$E\left[\frac{1}{n}\sum_{i=1}^{n}S_i^2\right]=\frac{1}{n}\sum_{i=1}^{n}E\left[S_i^2\right]=\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2, \quad (5.15)$$

so that the bias of the denominator is:

$$Bias\left(\frac{1}{n}\sum_{i=1}^{n}S_i^2\right)=\frac{1}{n}\sum_{i=1}^{n}E\left[S_i^2\right]-\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2=0. \quad (5.16)$$

In matrix form the denominator in (5.1) is equal to:

$$n^{-1}\sum_{i=1}^{n}S_i^2=\frac{1}{nT}vec(\mathbf{X})'\,\tilde{\mathbf{H}}\,vec(\mathbf{X}), \quad (5.17)$$

where $\tilde{\mathbf{H}}$ is a block matrix consisting of the centering matrix $\mathbf{H}_{(T\times T)}$ in each diagonal block and $\mathbf{X}$ is the $(T\times n)$ data matrix .

The variance of the denominator in (5.1) is then by theorem 10.22 (Schott, 2005, p. 418) and by letting $\mathbf{y}_{(nT\times 1)} = vec\left(\mathbf{X}_{(T\times n)}\right)$, equal to:

$$V\left(n^{-1}\sum_{i=1}^{n}S_i^2\right) = V\left(\frac{1}{nT}\mathbf{y}'\tilde{\mathbf{H}}\,\mathbf{y}\right) = \frac{1}{n^2T^2}V\left(\mathbf{y}'\tilde{\mathbf{H}}\,\mathbf{y}\right) = \frac{2}{n^2T^2}\left[tr\left(\tilde{\mathbf{H}}\boldsymbol{\Omega}\tilde{\mathbf{H}}\boldsymbol{\Omega}\right) + 2\tilde{\boldsymbol{\mu}}'\tilde{\mathbf{H}}\boldsymbol{\Omega}\tilde{\mathbf{H}}\tilde{\boldsymbol{\mu}}\right] =$$
$$\frac{2}{n^2T^2}\left[tr\left((\mathbf{I}\otimes\mathbf{H})(\boldsymbol{\Sigma}\otimes\mathbf{I})(\mathbf{I}\otimes\mathbf{H})(\boldsymbol{\Sigma}\otimes\mathbf{I})\right)\right] + \frac{4}{n^2T^2}(\boldsymbol{\mu}'\otimes\mathbf{1}')(\mathbf{I}\otimes\mathbf{H})(\boldsymbol{\Sigma}\otimes\mathbf{I})(\mathbf{I}\otimes\mathbf{H})(\boldsymbol{\mu}\otimes\mathbf{1}) =$$
$$\frac{2}{n^2T^2}tr\left((\boldsymbol{\Sigma}\boldsymbol{\Sigma})\otimes(\mathbf{H}\mathbf{H})\right) + \frac{4}{n^2T^2}(\boldsymbol{\mu}'\otimes\mathbf{1}')(\boldsymbol{\Sigma}\otimes\mathbf{H})(\boldsymbol{\mu}\otimes\mathbf{1}) = \frac{2}{n^2T^2}tr\left(\boldsymbol{\Gamma}\boldsymbol{\Lambda}^2\boldsymbol{\Gamma}'\right)tr(\mathbf{H}) +$$
$$\frac{4}{n^2T^2}(\boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu}\otimes\mathbf{1}'\mathbf{H}\mathbf{1}) = \frac{2}{n^2T^2}tr\left(\boldsymbol{\Gamma}\boldsymbol{\Lambda}^2\boldsymbol{\Gamma}'\right)(T-1) \qquad (5.18)$$

By the assumption of an underlying factor structure, (5.18) can be written as follows:

$$V\left(n^{-1}\sum_{i=1}^{n}S_i^2\right) = \frac{2}{n^2T^2}\left(\sum_{i=1}^{\sqrt{n}}\lambda_i^2 + \sum_{i=1+\sqrt{n}}^{n}\lambda_i^2\right)(T-1) = \frac{2}{n^2T^2}\left(\sum_{i=1}^{\sqrt{n}}\lambda_i^2 + O(1)\right)(T-1) \le$$
$$\frac{2}{n^{3/2}T}Max\left\{\lambda_i^2\right\}_{i=1}^{n} \to 0 \text{ as } n\to\infty, T\to\infty. \qquad (5.19)$$

Thus, from (5.15) and (5.19) we have that $E\left|\frac{1}{n}\sum_{i=1}^{n}S_i^2 - \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2\right|^2 \to 0$ as $n\to\infty$, $T\to\infty$, which implies that $\text{plim}\left(\frac{1}{n}\sum_{i=1}^{n}S_i^2\right) = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2$. (5.20)

Finally, since both the denominator and the nominator in (5.1) converge in probability to (5.14) and (5.20) respectively when $n\to\infty, T\to\infty$, we have by Slutsky's theorems (Ferguson, 1996) that the ratio converge in probability to β, hence:

$$\text{plim}\left(\sum_{i=1}^{n}\bar{X}_iS_i \Big/ \sum_{i=1}^{n}S_i^2\right) = \beta\sum_{i=1}^{n}\sigma_i^2 \Big/ \sum_{i=1}^{n}\sigma_i^2 = \beta. \qquad (5.21)$$

A4:

The second estimator, (1.5), is: $\hat{\beta}_{II} = \frac{1}{n}\sum_{i=1}^{n}\bar{X}_i \Big/ \frac{1}{n}\sum_{i=1}^{n}S_i$. (6.1)

The expected value of the numerator in (6.1) is by substituting (1.3) into the numerator, equal to:

$$E\left[\frac{1}{n}\sum_{i=1}^{n}\bar{X}_i\right]=\frac{1}{n}\sum_{i=1}^{n}E\left[\left(\bar{X}_i-\mu_i\right)+\beta\left(\sigma_i-S_i\right)+\beta S_i\right]=\beta\frac{1}{n}\sum_{i=1}^{n}\sigma_i. \tag{6.2}$$

Hence, the bias of the numerator is zero. The variance of the numerator in (6.1) is:

$$Var\left(n^{-1}\sum_{i=1}^{n}\bar{X}_i\right)=n^{-2}Var\left(\sum_{i=1}^{n}\bar{X}_i\right)=n^{-2}Var\left(\mathbf{1}'\bar{\mathbf{x}}\right)=n^{-2}\mathbf{1}'Var\left(\bar{\mathbf{x}}\right)\mathbf{1}=\left(n^2T\right)^{-1}\mathbf{1}'\mathbf{\Sigma}_{\mathbf{x}}\mathbf{1}=$$
$$\left(Tn^2\right)^{-1}\mathbf{1}'\left(\left(\mathbf{\Gamma}\tilde{\mathbf{\Lambda}}\mathbf{\Gamma}'+\mathbf{\Gamma}\mathbf{\Xi}\mathbf{\Gamma}'\right)\right)\mathbf{1}=\left(Tn^2\right)^{-1}\left(\sum_{i=1}^{\sqrt{n}}\lambda_i\mathbf{1}'\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}'_{(i)}\mathbf{1}+O(1)\right)\leq$$
$$o\left(Tn^{3/2}\right)\left(Max\left\{\lambda_i\mathbf{1}'\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}'_{(i)}\mathbf{1}\right\}_{i=1}^{n}\right)+o\left(Tn^2\right)\to 0 \text{ as } n\to\infty,\ T\to\infty. \tag{6.3}$$

So from (6.2), (6.3) and (1.1) we have $E\left|\frac{1}{n}\sum_{i=1}^{n}\bar{X}_i-\beta\frac{1}{n}\sum_{i=1}^{n}\sigma_i\right|^2\to 0$ as $n\to\infty, T\to\infty$

, which implies that $\operatorname{plim}\left(\frac{1}{n}\sum_{i=1}^{n}\bar{X}_i\right)=\beta\frac{1}{n}\sum_{i=1}^{n}\sigma_i.$ (6.4)

The expected value of the denominator in (6.1) is by (5.3) equal to:

$$E\left[\frac{1}{n}\sum_{i=1}^{n}S_i\right]=\frac{1}{n}\sum_{i=1}^{n}E\left[S_i\right]=\sqrt{\frac{2}{(T-1)}}\frac{\Gamma(T/2)}{\Gamma\left((T-1)/2\right)}\frac{1}{n}\sum_{i=1}^{n}\sigma_i. \tag{6.5}$$

Hence the asymptotic bias of the denominator in (6.1) is:

$$Bias\left(\frac{1}{n}\sum_{i=1}^{n}S_i\right)=\sqrt{\frac{2}{(T-1)}}\frac{\Gamma(T/2)}{\Gamma\left((T-1)/2\right)}\frac{1}{n}\sum_{i=1}^{n}\sigma_i-\frac{1}{n}\sum_{i=1}^{n}\sigma_i=o(T)\frac{1}{n}\sum_{i=1}^{n}\sigma_i\leq$$
$$o(T)Max\left\{\sigma_i\right\}_{i=1}^{n}\to 0 \text{ when } n\to\infty, T\to\infty. \tag{6.6}$$

Further, the variance of the denominator of (6.1) is:

$$Var\left(\frac{1}{n}\sum_{i=1}^{n}S_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}Var\left(S_i\right)+\frac{1}{n^2}\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n}Cov\left(S_i,S_j\right), \tag{6.7}$$

where the first summand in rhs of (6.7) is by (5.3) equal to:

$$\frac{1}{n^2}\sum_{i=1}^{n}Var\left(S_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}\left(E\left[S_i^2\right]-E^2\left[S_i\right]\right)=\left(1-\frac{2}{(T-1)}\frac{\Gamma^2(T/2)}{\Gamma^2\left((T-1)/2\right)}\right)\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2=$$
$$o(nT)\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2\leq o(nT)\mathrm{Max}\left\{\sigma_i^2\right\}_{i=1}^{n}, \tag{6.8}$$

and the second summand in rhs of (6.7) is by (6.8) bounded above by:

$$\frac{1}{n^2}\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n}\left|Cov\left(S_i,S_j\right)\right| = \frac{1}{n^2}\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n}\left|\rho_{S_i,S_j}\sigma_{S_i}\sigma_{S_j}\right| \leq \frac{1}{n^2}\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n}\sigma_{S_i}\sigma_{S_j} =$$

$$\frac{1}{n^2}\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n}\sqrt{o(T)\sigma_i^2}\sqrt{o(T)\sigma_j^2} = \leq o(T)\left(1-o(n)\right)Max\left\{\sigma_i\sigma_j\right\}_{i,j}^{n,n} \to 0 \quad \text{when} \quad n\to\infty,$$

$$T\to\infty . \tag{6.9}$$

Thus we have from (6.8) and (6.9) that (6.7) limits:

$$Var\left(\frac{1}{n}\sum_{i=1}^{n}S_i\right) \leq o(nT)\text{Max}\left\{\sigma_i^2\right\}_{i=1}^{n} + o(T)\left(1-o(n)\right)Max\left\{\sigma_i\sigma_j\right\}_{i,j}^{n,n} \to 0 \text{ when } n\to\infty,$$

$$T\to\infty . \tag{6.10}$$

Hence, by (6.6) and (6.10) we have $E\left|\frac{1}{n}\sum_{i=1}^{n}S_i - \frac{1}{n}\sum_{i=1}^{n}\sigma_i\right|^2 \to 0$ as $n\to\infty, T\to\infty$,

which implies that $\text{plim}\left(\sum_{i=1}^{n}S_i\right) = \sum_{i=1}^{n}\sigma_i .$ (6.11)

Finally, since both the denominator and the nominator in (6.1) converge in probability to (6.4) and (6.11) respectively, we have by Slutsky’s theorems that the ratio converge in probability to β, hence: $\text{plim}\left(\sum_{i=1}^{n}\bar{X}_i \Big/ \sum_{i=1}^{n}S_i\right) = \beta\sum_{i=1}^{n}\sigma_i \Big/ \sum_{i=1}^{n}\sigma_i = \beta .$ (6.12)

A5:

From now on the setting is changed so that each observation is equal to: $X_{i,t} = \beta\sigma_i + \delta_i + \varepsilon_{i,t}$, where $\delta_i \sim N\left(0,\sigma_\delta^2\right)$, $\varepsilon_{i,t} \sim N\left(0,\sigma_i^2\right)$, so that $X_{i,t}\left|\delta_i\right. \sim N\left(\beta\sigma_i + \delta_i, \sigma_i^2\right)$.

The bias of the first estimator, $\hat{\beta}_I$, in this setting is equal to:

$$Bias\left(\hat{\beta}_I\right) = E\left[\hat{\beta}_I\right] - \beta = n^{-1}\sum_{i=1}^{n}E\left[\frac{\bar{X}_i}{S_i}\middle|\delta_i\right] - \beta = n^{-1}\sum_{i=1}^{n}E\left[\bar{X}_i\middle|\delta_i\right]E\left[S_i^{-1}\middle|\delta_i\right] - \beta =$$

$$n^{-1}\sum_{i=1}^{n}\left(\beta\sigma_i + \delta_i\right)\frac{1}{\sigma_i}\frac{\Gamma\left[(T-2)/2\right]}{\Gamma\left[(T-1)/2\right]}\sqrt{\frac{T-1}{2}} - \beta = \beta\frac{\Gamma\left[(T-2)/2\right]}{\Gamma\left[(T-1)/2\right]}\sqrt{\frac{T-1}{2}} +$$

$$\frac{\Gamma\left[(T-2)/2\right]}{\Gamma\left[(T-1)/2\right]}\sqrt{\frac{T-1}{2}}n^{-1}\sum_{i=1}^{n}\frac{\delta_i}{\sigma_i}-\beta=\beta o(T)+\left(1+o(T)\right)n^{-1}\sum_{i=1}^{n}\frac{\delta_i}{\sigma_i} \qquad (7.1)$$

Let $\sigma_i^{-1}=\xi_i, \quad \xi_i=n^{-1}\sum_{i=1}^{n}\xi_i+\nu_i=\bar{\xi}+\nu_i$, then the last summand in rhs of (7.1) can be written as follows: $n^{-1}\sum_{i=1}^{n}\sigma_i^{-1}\delta_i=n^{-1}\sum_{i=1}^{n}\xi_i\delta_i=n^{-1}\sum_{i=1}^{n}\left(\bar{\xi}+\nu_i\right)\delta_i=\bar{\xi}\,\bar{\delta}+n^{-1}\sum_{i=1}^{n}\nu_i\delta_i$,

where $\bar{\delta}\xrightarrow{p}0$ as $n\to\infty$, $\sum_{i=1}^{n}\delta_i=O(1)\Rightarrow\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i=o(1)$ and $\nu_i/\sqrt{n}\searrow 0$ as $n\to\infty$.

Thus, $n^{-1}\sum_{i=1}^{n}\nu_i\delta_i\to 0$ as $n\to\infty$. Thus, (7.1) limits:

$$Bias\left(\hat{\beta}_I\right)=\beta o(T)+\left(1+o(T)\right)o\left(n^{1/2}\right)\to 0 \text{ as } n\to\infty, T\to\infty. \qquad (7.2)$$

Further, the variance of $\hat{\beta}_I$ is equal to:

$$Var\left(\hat{\beta}_I\right)=V\left(n^{-1}\sum_{i=1}^{n}\frac{\bar{X}_i}{S_i}\middle|\boldsymbol{\delta}\right)=n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov\left(\frac{\bar{X}_i}{S_i},\frac{\bar{X}_j}{S_j}\middle|\delta_i,\delta_j\right)=$$
$$\frac{1}{Tn^2}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov\left(\frac{\bar{X}_i}{S_i/\sqrt{T}},\frac{\bar{X}_j}{S_j/\sqrt{T}}\middle|\delta_i,\delta_j\right)=\frac{1}{Tn^2}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov\left(t_i',t_j'\middle|\delta_i,\delta_j\right)=$$
$$\frac{1}{Tn^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{t_i',t_j'}\sqrt{V\left(t_i'|\delta_i\right)V\left(t_j'|\delta_j\right)} \qquad (7.3)$$

Due to the assumed underlying factor structure in the cross-section, the number of terms in the summation in (7.3) is reduced from n^2 to k^2 where k is assumed to be less or equal to $\sqrt{n}$. Thus, by expressing (7.3) in a matrix form and utilize the underlying factor structure, (7.3) becomes:

$$Var\left(\hat{\beta}_I\right)=\frac{1}{Tn^2}\mathbf{1}'\boldsymbol{\Sigma}_{\mathbf{t}'}\mathbf{1}=\frac{1}{Tn^2}\mathbf{1}'\left(\boldsymbol{\Gamma}\tilde{\boldsymbol{\Lambda}}\boldsymbol{\Gamma}'+\boldsymbol{\Gamma}\boldsymbol{\Xi}\boldsymbol{\Gamma}'\right)\mathbf{1}\le\frac{1}{Tn^2}\left(\sum_{i=1}^{\sqrt{n}}\lambda_i\mathbf{1}'\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}'_{(i)}\mathbf{1}+O(1)\right)\le$$
$$o(nT)Max\left\{\lambda_i\mathbf{1}'\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}'_{(i)}\mathbf{1}\right\}_{i=1}^{n}+o(nT)\to 0 \text{ as } n\to\infty, T\to\infty, \qquad (7.4)$$

where $\tilde{\boldsymbol{\Lambda}}_{(n\times n)}=\begin{bmatrix}\boldsymbol{\Lambda}_{(k\times k)} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}$, $\boldsymbol{\Lambda}=diag\left(\lambda_1,\cdots,\lambda_k\right)$ and $\lambda_1\ge\cdots\ge\lambda_k>0$ are the ordered eigenvalues related to $\boldsymbol{\Sigma}_{\mathbf{t}'}$.

Thus, from (7.1) and (7.4), we have that $\text{plim}\left(\hat{\beta}_I\right)=\beta$.

A6:

Under the same settings as in A5, we will derive the asymptotic properties of the third estimator (1.6). To simplify, we will first show convergence in probability of the numerator followed by convergence of the denominator, and finally by Slutsky's theorems we can derive the asymptotic properties of the third estimator. Before any derivation, note that the third estimator can be rewritten into the following form:

$$\hat{\beta}_{III} = \sum_{i=1}^{n} \bar{X}_i S_i \Big/ \sum_{i=1}^{n} S_i^2 = n^{-1}\sum_{i=1}^{n} \bar{X}_i S_i \Big/ n^{-1}\sum_{i=1}^{n} S_i^2 . \tag{8.1}$$

So, for the numerator in rhs of (8.1), we have the Bias is under assumption of multivariate normal distribution, simplified by independence between $\bar{X}_i, S_i$, to the following:

$$Bias\left(n^{-1}\sum_{i=1}^{n}\bar{X}_i S_i\right) = E\left[n^{-1}\sum_{i=1}^{n}\bar{X}_i S_i \middle| \boldsymbol{\delta}\right] - \beta\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 = n^{-1}\sum_{i=1}^{n}E\left[\bar{X}_i S_i \,|\delta_i\right] - \beta\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 =$$
$$n^{-1}\sum_{i=1}^{n}E\left[\bar{X}_i \,|\delta_i\right]E\left[S_i \,|\delta_i\right] - \beta\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 =$$
$$\frac{1}{n}\sum_{i=1}^{n}(\beta\sigma_i+\delta_i)\sigma_i\sqrt{\frac{2}{T-1}}\frac{\Gamma[T/2]}{\Gamma\left[(T-1)/2\right]} - \beta\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 = \left[1+o(T)\right]\beta\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 +$$
$$\left[1+o(T)\right]\frac{1}{n}\sum_{i=1}^{n}\delta_i\sigma_i - \beta\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 . \tag{8.2}$$

Let $\sigma_i = n^{-1}\sum_{i=1}^{n}\sigma_i + \nu_i = \bar{\sigma} + \nu_i$, then the second summand in rhs of (8.2) can be written as follows: $n^{-1}\sum_{i=1}^{n}\delta_i\sigma_i = n^{-1}\sum_{i=1}^{n}(\bar{\sigma}+\nu_i)\delta_i = \bar{\sigma}\,\bar{\delta} + n^{-1}\sum_{i=1}^{n}\nu_i\delta_i$, where $\bar{\delta}\xrightarrow{p}0$ as $n\to\infty$, $\sum_{i=1}^{n}\delta_i = O(1) \Rightarrow n^{-1/2}\sum_{i=1}^{n}\delta_i = o(1)$ and $\nu_i\big/\sqrt{n} \searrow 0$ as $n\to\infty$. Thus, $n^{-1}\sum_{i=1}^{n}\nu_i\delta_i \to 0$ as $n\to\infty$.

Thus, the asymptotic bias, (8.2), limits:

$$Bias\left(n^{-1}\sum_{i=1}^{n}\bar{X}_i S_i\right) = \left[1+o(T)\right]\beta\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 + \left[1+o(T)\right]o\left(n^{1/2}\right) - \beta\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 \to 0 \text{ as}$$
$$n\to\infty, T\to\infty . \tag{8.3}$$

The variance of the numerator in rhs of (8.1) is equal to:

$$V\left(n^{-1}\sum_{i=1}^{n}\bar{X}_i S_i \middle| \boldsymbol{\delta} \right) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov\left(\bar{X}_i S_i, \bar{X}_j S_j \left| \delta_i, \delta_j \right.\right) =$$

$$n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{i,j}\sqrt{V\left(\bar{X}_i S_i \left|\delta_i\right.\right)V\left(\bar{X}_j S_j\left|\delta_j\right.\right)} \le n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sqrt{V\left(\bar{X}_i S_i\left|\delta_i\right.\right)}\sqrt{V\left(\bar{X}_j S_j\left|\delta_j\right.\right)}, \tag{8.4}$$

where: $V\left(\bar{X}_i S_i\left|\delta_i\right.\right) = E\left[\bar{X}_i^2 S_i^2\left|\delta_i\right.\right] - E^2\left[\bar{X}_i S_i\left|\delta_i\right.\right] = E\left[\bar{X}_i^2\left|\delta_i\right.\right]E\left[S_i^2\left|\delta_i\right.\right] -$

$$E^2\left[\bar{X}_i\left|\delta_i\right.\right]E^2\left[S_i\left|\delta_i\right.\right] = \left(\frac{\sigma_i^2}{T} + \beta^2\sigma_i^2 + \delta_i^2\right)\left(\sigma_i^2\right) -$$

$$\left(\beta^2\sigma_i^2 + \delta_i^2\right)\left(\sigma_i^2 \frac{2}{T-1}\frac{\Gamma^2\left(T/2\right)}{\Gamma^2\left(\left(T-1\right)/2\right)}\right) = \frac{\sigma_i^4}{T} + \beta^2\sigma_i^4\left(1 - \frac{2}{T-1}\frac{\Gamma^2\left(T/2\right)}{\Gamma^2\left(\left(T-1\right)/2\right)}\right) +$$

$$\delta_i^2\sigma_i^2\left(1 - \frac{2}{T-1}\frac{\Gamma^2\left(T/2\right)}{\Gamma^2\left(\left(T-1\right)/2\right)}\right) = o\left(T\right)\left[\left(1-\beta^2\right)\sigma_i^4 + \delta_i^2\sigma_i^2\right]. \tag{8.5}$$

Substituting (8.5) into (8.4) gives the following expression:

$$V\left(n^{-1}\sum_{i=1}^{n}\bar{X}_i S_i \middle| \boldsymbol{\delta} \right) \le n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} o\left(T\right)\left[\left(1-\beta^2\right)\sigma_i^4 + \delta_i^2\sigma_i^2\right]^{1/2}\left[\left(1-\beta^2\right)\sigma_i^4 + \delta_i^2\sigma_i^2\right]^{1/2} =$$

$$o\left(Tn^2\right)\left(1-\beta^2\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_i^2\sigma_j^2 + o\left(Tn^2\right)\left(1-\beta^2\right)^{1/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_i^2\sigma_j\left|\delta_j\right| +$$

$$o\left(Tn^2\right)\left(1-\beta^2\right)^{1/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_j^2\sigma_i\left|\delta_i\right| + o\left(Tn^2\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_i\sigma_j\left|\delta_i\right|\left|\delta_j\right|. \tag{8.6}$$

Since $\delta \sim N\left(0,\sigma_\delta^2\right)$ then $\left|\delta_i\right|$ is half normal distributed with $E\left[\left|\delta_i\right|\right] = \sigma_\delta\sqrt{2/\pi}$ and $Var\left(\left|\delta_i\right|\right) = \sigma_\delta^2\left(1-2/\pi\right)$, so that $n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\delta_j\right| = O\left(1\right)$ and $\sigma_i^2\sigma_j\big/\sqrt{T} \searrow 0$, $\sigma_j^2\sigma_i\big/\sqrt{T} \searrow 0$, $\sigma_i\sigma_j\big/\sqrt{T} \searrow 0$ as $T \to \infty$, so that (8.6) limits zero as n and T limits infinity.

Thus we have:

$$V\left(n^{-1}\sum_{i=1}^{n}\bar{X}_i S_i \middle| \boldsymbol{\delta} \right) \le o\left(Tn^2\right)\left(1-\beta^2\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_i^2\sigma_j^2 + o\left(Tn^2\right)\left(1-\beta^2\right)^{1/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_i^2\sigma_j\left|\delta_j\right| +$$

$$o\left(Tn^2\right)\left(1-\beta^2\right)^{1/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_j^2\sigma_i\left|\delta_i\right| + o\left(Tn^2\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_i\sigma_j\left|\delta_i\right|\left|\delta_j\right| \to 0 \text{ as } n\to\infty,\ T\to\infty. \tag{8.7}$$

Further, the assumed factor structure in the cross-section, implies that we can expressing (8.4) in the following matrix form by using the spectral decomposition as follows:

$$V\left(n^{-1}\sum_{i=1}^{n}\bar{X}_iS_i\middle|\boldsymbol{\delta}\right)=\frac{1}{n^2}\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}=\frac{1}{n^2}\mathbf{1}'\left(\boldsymbol{\Gamma}\left(\tilde{\boldsymbol{\Lambda}}+\boldsymbol{\Xi}\right)\boldsymbol{\Gamma}'\right)\mathbf{1}=\frac{1}{n^2}\left(\sum_{i=1}^{\sqrt{n}}\lambda_i\mathbf{1}'\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}'_{(i)}\mathbf{1}+O(1)\right)\leq$$

$$o\left(n^{3/2}\right)Max\left\{\lambda_i\mathbf{1}'\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}'_{(i)}\right\}_{i=1}^{n}+o\left(n^2\right)=o\left(n^{3/2}\right)\to 0 \text{ as } n\to\infty,\ T\to\infty, \qquad (8.8)$$

where $\tilde{\boldsymbol{\Lambda}}_{(n\times n)}=diag\left(\lambda_1,\cdots,\lambda_k,0,\cdots,0\right)$, $tr\left(\boldsymbol{\Xi}\right)=O(1)$ and $\lambda_1\geq\cdots\geq\lambda_k>0$ are the ordered eigenvalues related to $\boldsymbol{\Sigma}$.

Thus, from (8.3) and (8.8) we have that $E\left|n^{-1}\sum_{i=1}^{n}\bar{X}_iS_i-\beta n^{-1}\sum_{i=1}^{n}\sigma_i^2\right|^2\to 0$ as $n\to\infty$, $T\to\infty$, which implies that $\operatorname{plim}\left(n^{-1}\sum_{i=1}^{n}\bar{X}_iS_i\right)=\beta n^{-1}\sum_{i=1}^{n}\sigma_i^2$. (8.9)

The bias of the denominator in (8.1) is by (5.16) equal to:

$$Bias\left(n^{-1}\sum_{i=1}^{n}S_i^2\right)=n^{-1}\sum_{i=1}^{n}E\left[S_i^2\right]-n^{-1}\sum_{i=1}^{n}\sigma_i^2=0. \qquad (8.10)$$

The denominator in (8.1) is in matrix form equal to:

$$n^{-1}\sum_{i=1}^{n}S_i^2=\frac{1}{nT}vec\left(\mathbf{X}\right)'\,\tilde{\mathbf{H}}\,vec\left(\mathbf{X}\right), \qquad (8.11)$$

where $\tilde{\mathbf{H}}$ is a block matrix consisting of the centering matrix $\mathbf{H}_{(T\times T)}$ in each diagonal block and $\mathbf{X}$ is the $(T\times n)$ data matrix .

The variance of the denominator in (8.1) is then by theorem 10.22 (Schott, 2005, p. 418) and by letting $\mathbf{y}=vec\left(\mathbf{X}\right)$, equal to:

$$V\left(n^{-1}\sum_{i=1}^{n}S_i^2\right)=V\left(\frac{1}{nT}\mathbf{y}'\tilde{\mathbf{H}}\,\mathbf{y}\right)=\frac{1}{n^2T^2}V\left(\mathbf{y}'\tilde{\mathbf{H}}\,\mathbf{y}\right)=\frac{2}{n^2T^2}\left[tr\left(\tilde{\mathbf{H}}\boldsymbol{\Omega}\tilde{\mathbf{H}}\boldsymbol{\Omega}\right)+2\tilde{\boldsymbol{\mu}}'\tilde{\mathbf{H}}\boldsymbol{\Omega}\tilde{\mathbf{H}}\tilde{\boldsymbol{\mu}}\right]=$$
$$\frac{2}{n^2T^2}\left[tr\left(\left(\mathbf{I}\otimes\mathbf{H}\right)\left(\boldsymbol{\Sigma}\otimes\mathbf{I}\right)\left(\mathbf{I}\otimes\mathbf{H}\right)\left(\boldsymbol{\Sigma}\otimes\mathbf{I}\right)\right)\right]+\frac{4}{n^2T^2}\left(\boldsymbol{\mu}'\otimes\mathbf{1}'\right)\left(\mathbf{I}\otimes\mathbf{H}\right)\left(\boldsymbol{\Sigma}\otimes\mathbf{I}\right)\left(\mathbf{I}\otimes\mathbf{H}\right)\left(\boldsymbol{\mu}\otimes\mathbf{1}\right)=$$
$$\frac{2}{n^2T^2}tr\left(\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}\right)\otimes\left(\mathbf{H}\mathbf{H}\right)\right)+\frac{4}{n^2T^2}\left(\boldsymbol{\mu}'\otimes\mathbf{1}'\right)\left(\boldsymbol{\Sigma}\otimes\mathbf{H}\right)\left(\boldsymbol{\mu}\otimes\mathbf{1}\right)=$$
$$\frac{2}{n^2T^2}tr\left(\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}'\right)tr\left(\mathbf{H}\right)+\frac{4}{n^2T^2}\left(\boldsymbol{\mu}'\boldsymbol{\Sigma}\boldsymbol{\mu}\otimes\mathbf{1}'\mathbf{H}\mathbf{1}\right)=\frac{2}{n^2T^2}tr\left(\boldsymbol{\Gamma}\boldsymbol{\Lambda}^2\boldsymbol{\Gamma}'\right)\left(T-1\right)=$$
$$\frac{2}{n^2T^2}\left(\sum_{i=1}^{\sqrt{n}}\lambda_i^{\,2}+O(1)\right)\left(T-1\right)\leq\frac{2}{n^{3/2}T}Max\left\{\lambda_i^{\,2}\right\}_{i=1}^{n}\to 0 \text{ as } n\to\infty, T\to\infty. \qquad (8.12)$$

Thus, from (8.10) and (8.12) we have that $E\left|\sum_{i=1}^{n} S_i^2 - \sum_{i=1}^{n} \sigma_i^2\right|^2 \to 0$ as $n \to \infty, T \to \infty$,

which implies that $\operatorname{plim}\left(\sum_{i=1}^{n} S_i^2\right) = \sum_{i=1}^{n} \sigma_i^2$. (8.13)

Finally, since both the denominator and the nominator in (8.1) converge in probability to (8.9) and (8.13) respectively, we have by Slutsky's theorems that the ratio converge in probability to β, hence: $\operatorname{plim}\left(\sum_{i=1}^{n} \bar{X}_i S_i \Big/ \sum_{i=1}^{n} S_i^2\right) = \beta \sum_{i=1}^{n} \sigma_i^2 \Big/ \sum_{i=1}^{n} \sigma_i^2 = \beta.$ (8.14)

A7:

The second estimator (1.5) is: $\hat{\beta}_{II} = n^{-1}\sum_{i=1}^{n} \bar{X}_i \Big/ n^{-1}\sum_{i=1}^{n} S_i$. (9.1)

The expected value of the numerator in (9.1) is:

$$E\left[n^{-1}\sum_{i=1}^{n} \bar{X}_i \middle| \boldsymbol{\delta}\right] = n^{-1}\sum_{i=1}^{n} E\left[\bar{X}_i \middle| \delta_i\right] = (Tn)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T} E\left[X_{i,t} \middle| \delta_i\right] = (Tn)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\beta\sigma_i + \delta_i\right) =$$
$$\beta n^{-1}\sum_{i=1}^{n}\sigma_i + n^{-1}\sum_{i=1}^{n}\delta_i = \beta\sum_{i=1}^{n}\sigma_i + \bar{\delta} \xrightarrow{p} \beta n^{-1}\sum_{i=1}^{n}\sigma_i \text{ as } n \to \infty, \tag{9.2}$$

where $\bar{\delta} \sim N\left(0, n^{-1}\sigma_\delta^2\right)$, and the bias of the numerator is:

$$Bias\left(n^{-1}\sum_{i=1}^{n} \bar{X}_i\right) = E\left[n^{-1}\sum_{i=1}^{n} \bar{X}_i\right] - \beta n^{-1}\sum_{i=1}^{n}\sigma_i \to 0 \text{ as } n \to \infty. \tag{9.3}$$

Hence, the asymptotic bias of the numerator is equal to zero.

The variance of the numerator in (9.1) by the assumed underlying factor structure equal to:

$$Var\left(n^{-1}\sum_{i=1}^{n} \bar{X}_i \middle| \boldsymbol{\delta}\right) = n^{-2}Var\left(\mathbf{1}_n'\bar{\mathbf{x}}\right) = n^{-2}\mathbf{1}_n' Var\left(\bar{\mathbf{x}}\right)\mathbf{1}_n = n^{-2}\mathbf{1}_n'\left(T^{-1}\boldsymbol{\Sigma}_{\mathbf{x}}\right)\mathbf{1}_n =$$
$$\left(Tn^2\right)^{-1}\mathbf{1}_n'\left(\left(\boldsymbol{\Gamma}\tilde{\boldsymbol{\Lambda}}\boldsymbol{\Gamma}' + \boldsymbol{\Gamma}\boldsymbol{\Xi}\boldsymbol{\Gamma}'\right)\right)\mathbf{1}_n = \left(Tn^2\right)^{-1}\left(\sum_{i=1}^{\sqrt{n}}\lambda_i\mathbf{1}_n'\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}_{(i)}'\mathbf{1}_n + O(1)\right) =$$
$$\left(Tn^2\right)^{-1}\sum_{i=1}^{\sqrt{n}}\lambda_i\mathbf{1}_n'\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}_{(i)}'\mathbf{1}_n + \left(Tn^2\right)^{-1}O(1) \le \left(Tn^{3/2}\right)^{-1}\lambda_{\max}\mathbf{1}_n'\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}_{(i)}'\mathbf{1}_n + \left(Tn^2\right)^{-1}O(1) \to 0 \text{ as}$$
$$n \to \infty,\ T \to \infty. \tag{9.4}$$

So from (9.2) and (9.4) $E\left|n^{-1}\sum_{i=1}^{n}\bar{X}_i - \beta n^{-1}\sum_{i=1}^{n}\sigma_i\right|^2 \to 0$ as $n \to \infty,\ T \to \infty$, which implies that $\operatorname{plim}\left(n^{-1}\sum_{i=1}^{n}\bar{X}_i\right) = \beta n^{-1}\sum_{i=1}^{n}\sigma_i.$ (9.5)

The asymptotic bias of the denominator in (9.1) is by the assumed underlying factor structure and by the fact that the sample covariance matrix converge in probability to the true covariance matrix with $o(T)$, equal to:

$$Bias\left(n^{-1}\sum_{i=1}^{n}S_i\right) = n^{-1}E\left[tr\left\{\mathbf{S}^{1/2}\right\}\right] - n^{-1}tr\left\{\mathbf{\Sigma}^{1/2}\right\} = \left[1+o(T)\right]n^{-1}tr\left\{\mathbf{\Sigma}^{1/2}\right\} - n^{-1}tr\left\{\mathbf{\Sigma}^{1/2}\right\} =$$

$$o(T)n^{-1}tr\left\{\mathbf{\Sigma}^{1/2}\right\} = o(T)n^{-1}tr\left\{\sum_{i=1}^{\sqrt{n}}\lambda_i^{1/2}\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}'_{(i)} + \sum_{i=1+\sqrt{n}}^{n}\lambda_i^{1/2}\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}'_{(i)}\right\} =$$

$$o(T)n^{-1}tr\left\{\sum_{i=1}^{\sqrt{n}}\lambda_i^{1/2}\boldsymbol{\gamma}_{(i)}\boldsymbol{\gamma}'_{(i)} + O(1)\right\} = o\left(Tn^{\delta}\right) \to 0 \quad \text{when} \quad n \to \infty, T \to \infty \quad \text{and}$$

$\delta \in \left]0,1/2\right[$. (9.6)

Further, by the assumed underlying factor structure and (6.10), the variance of the denominator of (9.1) limits:

$$Var\left(n^{-1}\sum_{i=1}^{n}S_i\right) \le o(T)\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2 + o(T)\frac{1}{n^2}\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n}\sigma_i\sigma_j \to 0 \text{ when } n \to \infty, T \to \infty.$$

(9.7)

Hence, by (9.6) and (9.7) we have $E\left|\frac{1}{n}\sum_{i=1}^{n}S_i - \frac{1}{n}\sum_{i=1}^{n}\sigma_i\right|^2 \to 0$ as $n \to \infty, T \to \infty$, which implies that $\operatorname{plim}\left(\sum_{i=1}^{n}S_i\right) = \sum_{i=1}^{n}\sigma_i.$ (9.8)

Finally, since both the denominator and the nominator in (9.1) converge in probability to (9.5) and (9.8) respectively, we have by Slutsky's theorems that the ratio converge in probability to β, hence: $\operatorname{plim}\left(\hat{\beta}_{II}\right) = \beta\sum_{i=1}^{n}\sigma_i \Big/ \sum_{i=1}^{n}\sigma_i = \beta.$ (9.9)

References

Basu, J. P.; Odell, P. L. and A. Kinderman. (1975). On Independence of Sample Mean and Translation Invariant Statistics of Samples from Multivariate Normal Populations. *Journal of American Statistical Association,* 70(350), p. 480-481.

Casella, G. and R. L. Berger. (1990). *STATISTICAL INFERENCE.* Duxbury.

Ferguson, T. H. (1996). *A COURSE IN LARGE SAMPLE THEORY*. Chapman & Hall/CRC.

Mardia, K.V.; Kent, J.T. and J.M. BIBBY. (1979). *Multivariate Analysis*. ACADEMIC PRESS.

Rohatgi, V. K. (1976). *An Introduction to Probability Theory and Mathematical Statistics*. WILEY.

Pearson, K. (1896). Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia. Philosophical Transactions of the Royal Society of London. Series A. 187, 253-318.

Roll, R. and Ross, S. A. (1980). An Empirical Investigation of the Arbitrage Pricing Theory. **Journal of Finance,** 35(5), 1073-1103.

Ross, S. A. (1976). The Arbitrage Theory of Capital Asset Pricing. Journal of Economic Theory, 13(3), 341-360.

Schott, R. J. (2005). *MATRIX ANALYSIS FOR STATISTICS*. WILEY.

Sharpe, W. F. (1966). Mutual Fund Performance. *Journal of Business*, 1(2) January, 119-138.