Account and Financial Management Journal e-ISSN: 2456-3374

Volume 4 Issue 04 April- 2019, (Page No.-1901-1912)

DOI: 10.31142/afmj/v4i4.02, I.F. - 4.614

The Fama-French Three-Factor Model under Uncertainty

Cong Wang

Department of Applied Mathematics, Nanjing University of Science and Technology, Nanjing 210094, P R China

Abstract: The present paper considers and restructures Fama-French three-factor model under the noise uncertainty based on the nonlinear expectation theory. Simultaneously, the identification and related statistical properties of the newly defined model are established.

Keywords: Fama-French three-factor model; Upper expectation regression; Weighted expectation regression

2010 Mathematics Subject Classification 60H15; 60G22; 91G80

1. Introduction

Linear regression is one of the most common and practical statistical models in statistics. For the assumptions in traditional statistical models, the most important hypothesis may be the certainty of the distribution, that is, each relevant variable has a certain probability distribution. However, in practical problems, the certainty of distribution is not completely guaranteed. For example, in the financial field, for the risk measurement problems, there are often situations where the distribution is uncertain. If there is a situation where the distribution is uncertain, then the expectation of the relevant variable is usually non-linear [5].

It is well known that in the real capital market, due to the incomplete information and rapid information update, the impact factors in the traditional statistical models cannot fully explain the response factors. Some of the impact factors are placed into the residual term for which are unobservable, unpredictable, or artificially neglected, resulting in the expected performance of the residual as a non-linear expectation. In order to the ease of processing, the residuals are still processed as constant expectation and fixed variance just like in the traditional statistical model, which will undoubtedly reduce the accuracy of the model.

According to the current situation of Chinese stock market, many scholars have studied the empirical effects of the Fama-French three-factor model in Chinese stock market. However, the results of the studies are not the same. Eun and Huang (2007)[2] confirmed that there are significant scale and book-to-market ratio effects in Chinese stock market. However, Chen etal.(2010)[1] found that the book-to-market ratio can effectively explain the change in crosssectional stock returns, but the scale effect does not.

Considering the current research situations, the authors believe that the difference in the empirical results of the three-factor model may be due to the assumption that the model itself is inconsistent with the complex economic status. In detail, the authors believe that there are other factors which can affect the pricing results of assets. Because other factors are unmeasurable or unknown, they are hidden in the noises. In addition, in order to facilitate the model processing, the error term is simply treated as a general assumption that the expectation is zero and the variance is determined.

Therefore, based on the uncertainty of the distributions of error terms, this paper considers the improvement of the specific linear regression model—Fama-French three-factor model under the condition of uncertain distribution. Under the k-sample hypothesis, this paper uses upper expectation regression method and weighted expectation regression method to deal with the uncertain distribution, and gets its optimal solution thorough genetic algorithm. Furthermore, this paper gives the identification and asymptotic normality properties.

An outline of this paper is as follows. In section 2 we show the necessary preliminary knowledge. In section 3 an improved Fama-French three-factor model based on the upper expectation regression method is given. In section 4 an improved Fama-French three-factor model based on the weighted expectation regression method is given. In section 5 an empirical example is given to show that the advantage of the new models.

2. Preliminaries

2.1 The Fama-French three-factor model

Sharp et al. (1964) [10] proposed the Capital Asset Pricing Model based on Markowitz's (1952) [6] portfolio theory, which laid the foundation of modern financial theory. Later, Ross(1976) [3] proposed the Arbitrage Pricing Theory (APT). Fama and French (1993) [9] introduceda kind of empirical factor research method in asset pricing research, and the specific research method is to build a simulation investment portfolio by sorting some characteristics of the company to obtain empirical factors, which is the Fama-French three-factor model:

$$R_{it} - R_{ft} = \beta(R_{Mt} - R_{ft}) + sSMB_t + hHML_t + e_{it}$$

R_{it} —— the yield of the securities or portfolio i in the t moment;

R_{ft} ——the risk-free rate of in the t moment;

R_{Mt} ——the market rate of return.;

SMB_t ——the scale factor;

HML_t ——the book market value ratio factor.;

β, s, and h ——factor loads;

e_{it} ——the disturbance term.

The contribution of the three-factor model is that the stock market value factor and the book-market value ratio can explain the price change and market return rate of most stocks well, and can replace other factors such as price-earnings risk factors.

2.2 Nonlinear expectation

Early researches of nonlinear expectations can be traced back to Huber's (1981)[8] study of the robustness of statistical models. However, in recent decades, scholars from various countries have made great progress in the study of theories and methods of nonlinear expectations. In some areas, such as financial risk measurement and economic direction, great achievements have been made. Peng (1997)[9] introduced g-expectation based on backward stochastic differential equations, and then extended G expectation in 2006, which are the representative research achievement.

It can be found that the research of nonlinear expectation theory in the direction of probability theory has developed rapidly, but few people have studied the related statistical models and their corresponding statistical inferences. The latest research on statistical models and statistical inferences about nonlinear expectations is only the k-sample expectation regression method which is proposed by LuLin et al. (2016)[1], in which the authors give the nonlinear expectation

$$\mathrm{E}[\varepsilon] = \sup_{F \in \mathfrak{I}} E_F(\varepsilon)$$

3. The Improved Fama-French three-factor model based on upper expectation

In this section, we present an improved Fama-French three-factor model based on the upper expectation regression method.

3.1 Model

The basic Fama-French three-factor model:

$$R_{it} - R_{ft} = \beta(R_{Mt} - R_{ft}) + sSMB_t + hHML_t + \varepsilon_{it}$$

Unlike traditional linear model assumptions, the author believes that some unmeasured and unknown factors are placed in the residual term, so the distribution is uncertain. Then, drawing on the ideas of Lu Lin et al. (2016), the distribution comes from such a collection:

$$\mathfrak{I} = \{F_1, \cdots, F_k\}$$

Among them, $F_1, \cdots, F_k$ are different distribution equations, meaning that in different situations, ε is of different distributions, and this distribution comes from the collection.

Based on the definition of the k sample, we get the following model:

$$E[\varepsilon] = \sup_{F \in \Im} E_F(\varepsilon)$$

$$R_{i\,t} - R_{f\,t} = \beta(R_{Mt} - R_{f\,t}) + s\,SMB_t + h\,HML_t + \varepsilon_{i\,t}$$

Among them, $E_F[\cdot]$ is the traditional expectation under the F distribution.

Let $E[Y_i] = E(R_i) - R_f$, $\bar{e} = E[\varepsilon] = \sup_{F \in \Im} E_F(\varepsilon)$, the author has obtained the improved Fama-French three-factor model under uncertainty:

$$E[Y|\Gamma] = (R_M - R_f, SMB, HML)\begin{pmatrix}\beta \\ s \\ h\end{pmatrix} + \bar{e}$$

where

$$\Gamma \square \begin{pmatrix} R_M - R_f \\ SMB \\ HML \end{pmatrix}$$

As a special linear regression model, the improved version of the three-factor model processes the information data contained in the residuals, so it conforms to all the basic assumptions of the linear model.

Theorem 3.1 *If* $E[\Gamma(R_M - R_f, SMB, HML)]$ *is a definite matrix, then,* $(\beta, s, h)'$ *is identifiable which can be expressed as*

$$(\beta, s, h)' = (E[\Gamma(R_M - R_f, SMB, HML)])^{-1} E\{\Gamma E[Y|\Gamma]\}$$
$$-\bar{e}\,(E[\Gamma(R_M - R_f, SMB, HML)])^{-1} E[\Gamma]$$

This property guarantees the certainty of the regression parameters of the new model and the certainty of the regression equation.

Proof The improved Fama-French three-factor model under uncertainty is

$$E[Y|\Gamma] = (R_M - R_f, SMB, HML)\begin{pmatrix}\beta \\ s \\ h\end{pmatrix} + \bar{e}$$

The both sides of the last equation lefty multiply by Γ simultaneously, then

$$\Gamma E[Y|\Gamma] = \Gamma(R_M - R_f, SMB, HML)\begin{pmatrix}\beta \\ s \\ h\end{pmatrix} + \Gamma\bar{e}$$

Seeking the expectations, then

$$E[\Gamma E[Y\,\Gamma]]$$
$$= E[\Gamma(R_M - R_f, SMB, HML)]\begin{pmatrix}\beta \\ s \\ h\end{pmatrix} + \bar{\mu}E[\Gamma]$$

Transposition, then

$$E\left[\Gamma\left(R_M - R_f, SMB, HML\right)\right]\begin{pmatrix}\beta\\ s\\ h\end{pmatrix}$$

$$= E\left[\Gamma E\left[Y|\Gamma\right]\right] - \bar{e}E\left[\Gamma\right]$$

The both sides of the last equation lefty multiply by

$$E\left[\Gamma\left(R_M - R_f, SMB, HML\right)\right]^{-1}$$

simultaneously, then

$$(\beta, s, h)' = \left(E\left[\Gamma\left(R_M - R_f, SMB, HML\right)\right]\right)^{-1} E\left\{\Gamma E\left[Y|\Gamma\right]\right\}$$
$$-\bar{e}\left(E\left[\Gamma\left(R_M - R_f, SMB, HML\right)\right]\right)^{-1} E\left[\Gamma\right]$$

This completes the proof of Theorem 3.1.

Theorem 3.1 guarantees the certainty of the regression parameters of the new model and the certainty of the regression equation.

3.2 Estimation and Prediction

In this section, the paper presents the algorithm and properties for estimating the improved Fama-French three-factor model. The improved Fama-French three-factor model can be expressed by vector:

$$\mathrm{E}\left[Y|\Gamma\right] = \left(R_M - R_f, SMB, HML\right)\begin{pmatrix}\beta\\ s\\ h\end{pmatrix} + \bar{e}$$

Let samples $\{(Y_i, (R_{Mi} - R_{fi}, SMB_i, HML_i)) : i = 1, 2, \cdots, N\}$ come from the original model, satisfy

$$Y_i = \left(R_M - R_{fi}, SMB_i, HML_i\right)\begin{pmatrix}\beta_i\\ s_i\\ h_i\end{pmatrix} + \varepsilon_i, i = 1, \cdots, N.$$

Because the distributions of $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N$ is uncertain, the distributions of $Y_1, Y_2, \cdots, Y_N$ is of uncertainty, which means its estimation method is different from the traditional model.

Refer to the mini-max two-stage method proposed by LinLu(2016), we can get (β, h, s) and $\bar{e}$. In the upper expectation framework, when $\left(R_M - R_f, SMB, HML\right)$ is given, the upper expectation of Y is fixed, which is $\left(R_M - R_f, SMB, HML\right)\begin{pmatrix}\beta\\ s\\ h\end{pmatrix} + \bar{e}$. So, similar to the traditional least squares approach, we find the values of (β, h, s) and $\bar{e}$ by minimizing the expected least squares loss:

$$\mathrm{E}\left[\left(Y - \left(R_M - R_f, SMB, HML\right)\begin{pmatrix}\beta\\ s\\ h\end{pmatrix} - \bar{e}\right)^2\right]$$

Based on the k-sample assumption, in order to complete the estimation steps, the paper gives the following k-sample condition:

C1. For $P_i, i = 1, 2, \cdots, k$, **when** $(ij) \in P_i$, $\varepsilon_{i1}, \cdots, \varepsilon_{in_i}$ **are identically distributed.**

For simplicity let $n_i = n, i = 1, \cdots, k$ Record F_i as the distribution of ε_{ij} , $(ij) \in I_i$, and the upper expected loss is:

$$\max_{1\le i\le k} \frac{1}{n}\sum_{j=1}^{n}\left[\left(Y_{ij} - (R_{Mj} - R_{fij}, SMB_{ij}, HML_{ij})\begin{pmatrix}\beta\\ s\\ h\end{pmatrix} - \bar{e}\right)\right]$$

Minimizing the last equation, the author gets the estimator of $\left((\beta, s, h)', \bar{e}\right)'$:

$$\left((\hat{\beta}, \hat{s}, \hat{h})', \hat{\bar{e}}\right)' = \arg\min_{(\beta,s,h)'\in B, \bar{\mu}\in U}\max_{1\le i\le k}\frac{1}{n}\sum_{j=1}^{n}\left[\left(Y_{ij} - (R_{Mj} - R_{fij}, SMB_{ij}, HML_{ij})\begin{pmatrix}\beta\\ s\\ h\end{pmatrix} - \bar{e}\right)\right]^2$$

In which, the parameter space of $(\beta, s, h)'$ and $\bar{e}$ is respectively B and U We can get the optimal solution of the last equation via genetic algorithm.

Denote $vol_i = E\left[\left(\varepsilon_{ij} - \bar{e}\right)^2\right]$ for $(ij) \in P_i$, $vol_{i_*} = \max_{1\le i\le k} vol_i$ and $e_* = E\left[\varepsilon_{i_* j}\right]$ for $(ij) \in P_{i_*}$.

Write $\Theta(X) = \begin{pmatrix}\Gamma(R_M - R_f, SMB, HML) & \Gamma\\ (R_M - R_f, SMB, HML) & 1\end{pmatrix}$, then the author get the asymptotic normality.

Theorem3.2. *Under the k-sample assumption which is condition C1, if* $E[\Gamma(R_M - R_f, SMB, HML)]$ *is a definite matrix and* $vol_{i_*} > vol_i$ *for all* $i \neq i_*$*, and* $n \to \infty$ *as* $N \to \infty$*, we get*

$$\sqrt{n}\left[\left(\begin{pmatrix}\hat{\beta}\\ \hat{s}\\ \hat{h}\end{pmatrix} - \begin{pmatrix}\beta\\ s\\ h\end{pmatrix}\right)', \hat{\bar{e}} - e_*\right] \xrightarrow{d} N\left(0, vol_*(E[\Theta(X)])^{-1}\right)$$

In which, $\xrightarrow{d}$ *means convergence in distribution and* $N\left(0, vol_*(E[\Theta(X)])^{-1}\right)$ *is a normal distribution.*

Theorem 3.1 guarantees the asymptotic normality and consistency of estimator $(\hat{\beta}, \hat{s}, \hat{h})'$. But the estimator $\hat{\bar{e}}$ tends to e_* instead of $\bar{e}$, so it is not always consistent. In order to get the consistent estimator of $\bar{e}$, the author uses a two stage procedure. With the last estimator $(\hat{\beta}, \hat{s}, \hat{h})'$, the estimator of $\bar{e}$ can be expressed as:

$$\hat{\bar{e}}_s = \max_{1\le i\le k}\frac{1}{n}\sum_{j=1}^{n}\left[Y_{ij} - \Gamma^{T}_{ij}\left(\hat{\beta}, \hat{s}, \hat{h}\right)'\right]$$

Write $e_i = E_i[\varepsilon_{ij}]$ for all $(ij) \in P_i$.

Denote $c = 1 - E[\Gamma]\left(E[\Gamma\Gamma^T]\right)^{-1}E[\Gamma]$, and

$\Psi^{-1}(X) = (E[\Gamma\Gamma^T])^{-1} + (E[\Gamma\Gamma^T])^{-1}E[\Gamma]E[\Gamma^T](E[\Gamma\Gamma^T])^{-1} / c$, then the author gets the consistency theorem.

Theorem3.3. *Under the condition of Theorem 3.1, if* $\{\varepsilon_{ij}, j = 1, \cdots, n\}$ *and* $\{\varepsilon_{sj}, j = 1, \cdots, n\}$ *for* $i \neq S$ *are independent, then the estimator* $\hat{\hat{e}}_S$ *satisfies*

$$\sqrt{n}\left[\hat{\hat{e}}_s - \hat{e}\right]' \xrightarrow{d} N\left(0, vol_{I_\cdot} + vol_{\Gamma}\left(E\left[\Gamma^T\right] E\left[\Psi^{-1}(X)\right]\right) E\begin{bmatrix} R_M - R_f \\ SMB \\ HML \end{bmatrix}\right)$$

In which, $\xrightarrow{d}$ *means convergence in distribution, and*

$$N\left(0, vol_{I_\cdot} + vol_{\Gamma}\left(E\left[\Gamma^T\right] E\left[\Psi^{-1}(X)\right]\right) E\begin{bmatrix} R_M - R_f \\ SMB \\ HML \end{bmatrix}\right)$$

is a normal distribution.

Theorem 3.2 guarantees the asymptotic normality and consistency of estimator $\hat{\hat{\hat{e}}}_S$.

Remark3.4. *The proof of the asymptotic normality and consistency of the method is similar to the proof of the asymptotic normality and consistency of weighted expectation regression method in Theorem 4.1, so they are omitted here.*

4. The Improved Fama-French three-factor model based on
weighted expectation regression method

In practice, the improved Fama-French three-factor model based on the upper expectation regression method behaves bad. The reason is that when using the upper expectation regression method in Chinese stock market, the method chooses the biggest error of the sample not the most representative ones. So the author develops a weighted expectation regression method.

In the following, the author presents the weighted expectation regression method.

Just like the improved Fama-French three-factor model based on the upper expectation regression method, the author believes that some unmeasured and unknown factors are placed in the residual term, so the distribution is uncertain. The distribution of the residual term comes from such a collection:

$$\Im = \{F_1, \cdots, F_k\}$$

Among them, $F_1, \cdots, F_k$ are different distribution equations, meaning that in different situations, ε is of different distributions, and this distribution comes from the collection.

Based on the definition of the k sample, we get the following model:

$$\mathbb{E}[\varepsilon] = \sum_{i=1}^{k} \frac{i}{1+\cdots+k} E_F(\varepsilon)$$

Among them, $E_F[\cdot]$ is the traditional expectation under the F distribution.

Let $\mathbb{E}[Y_i] = E(R_i) - R_f$,

$$\bar{v} = \mathbb{E}[\varepsilon] = \sum_{i=1}^{k} \frac{i}{1+\cdots+k} E_F(\varepsilon),$$

the author gets the improved Fama-French three-factor model under uncertainty:

$$E[Y|\Gamma] = (R_M - R_f, SMB, HML)\begin{pmatrix} \beta \\ s \\ h \end{pmatrix} + \bar{v}$$

As a special linear regression model, the improved version of the three-factor model processes the information data contained in the residuals, so it conforms to all the basic assumptions of the linear model.

Proposition 4.1 *If* $E\left[\Gamma\left(R_M - R_f, SMB, HML\right)\right]$ *is a definite matrix, then,* $(\beta, s, h)'$ *is identifiable which can be expressed as*

$$(\beta, s, h)' = \left(E\left[\Gamma\left(R_M - R_f, SMB, HML\right)\right]\right)^{-1} E\left\{\Gamma E\left[Y \mid \Gamma\right]\right\}$$
$$-\bar{v}\left(E\left[\Gamma\left(R_M - R_f, SMB, HML\right)\right]\right)^{-1} E\left[\Gamma\right]$$

This property guarantees the certainty of the regression parameters of the new model and the certainty of the regression equation. The proof is the same as Proposition 3.1, so it is omitted here.

Like the estimation of the upper expectation method, minimizing the expected least squares loss:

$$E\left[\left(Y - \left(R_M - R_f, SMB, HML\right)\begin{pmatrix}\beta\\ s\\ h\end{pmatrix} - \bar{v}\right)^2\right]$$

Under the condition C1, let $n_i = n, i = 1, \cdots, k$. Record F_i as the distribution of ε_{ij}, $(ij) \in I_i$ and the weighted expected loss is:

$$\sum_{i=1}^{k} \frac{i}{1+\cdots+k} \sum_{j=1}^{n}\left[\left(Y_{ij} - \begin{pmatrix}R_{Mij} - R_{fij}\\ SMB_{ij}\\ HML_{ij}\end{pmatrix} - \bar{v}\right)^2\right]$$

Minimizing the last equation, the author gets the estimator of $\left((\beta, s, h)', \bar{v}\right)'$:

$$\left(\left(\hat{\beta}, \hat{s}, \hat{h}\right)', \hat{\bar{v}}\right)' = \arg \min_{(\beta, s, h)' \in B, \bar{v} \in U} \sum_{1 \le i \le k} \frac{i}{1+\cdots+k} \frac{1}{n}$$
$$\times \sum_{j=1}^{n}\left[\left(Y_{ij} - \left(R_{Mij} - R_{fij}, SMB_{ij}, HML_{ij}\right)\begin{pmatrix}\beta\\ s\\ h\end{pmatrix} - \bar{v}\right)\right]^2$$

In which, the parameter space of $(\beta, s, h)'$ and $\bar{v}$ is respectively B and U. We can get the Optimal solution of the last equation via genetic algorithm.

Theorem4.2 *Under the k-sample assumption which is condition C1, if* $E\left[\Gamma\left(R_M - R_f, SMB, HML\right)\right]$ *is a definite matrix and* $Vol_{i_*} > Vol_i$ *for all* $i \neq i_*$, *and* $n \to \infty$ *as* $N \to \infty$, *we get*

$$\sqrt{n}\left[\left(\begin{pmatrix}\hat{\beta}\\ \hat{s}\\ \hat{h}\end{pmatrix} - \begin{pmatrix}\beta\\ s\\ h\end{pmatrix}\right)', \hat{\bar{v}} - v_*\right] \xrightarrow{d} N\left(0, vol_*\left(E\left[\Theta(X)\right]\right)^{-1}\right)$$

In which, $\xrightarrow{d}$ *means convergence in distribution and* $N\left(0, vol_*\left(E\left[\Theta(X)\right]\right)^{-1}\right)$ *is a normal distribution.*

Proof. Under the condition C1, we have

$$\frac{1}{n}\sum_{j=1}^{n}(\varepsilon_{ij}-\bar{v})^2 = vol_i + \delta_n$$

In which, δ_n is of $O_p\left(1/\sqrt{n}\right)$, and for $i \neq i_*, vol_{i_*} \geq vol_i$. Then we get

$$\max_{1\le i\le k}\frac{1}{n}\sum_{j=1}^{n}(\varepsilon_{ij}-\bar{v}) = \sigma_{i_*}^2 + \delta_n$$

When $n \to \infty$,

$$\max_{1\le i\le k}\frac{1}{n}\sum_{j=1}^{n}(\varepsilon_{ij}-\bar{v}) = \frac{1}{n}\sum_{j=1}^{n}(\varepsilon_{i_*j}-\bar{v}_{i_*})^2 + \delta_n$$

Then

$$\frac{1}{n}\sum_{j=1}^{n}\left[Y_{ij} - (R_{Mij} - R_{fij}, SMB_{ij}, HML_{ij})\begin{pmatrix}\beta\\ s\\ h\end{pmatrix} - \bar{v}\right]^2$$
$$\le \max_{1\le i\le k}\frac{1}{n}\sum_{j=1}^{n}(\varepsilon_{ij}-\bar{v}) = \frac{1}{n}\sum_{j=1}^{n}(\varepsilon_{i_*j}-\bar{v}_{i_*})^2 + \delta_n$$

Note the weighted expected loss as

$$M = \sum_{i=1}^{k}\frac{i}{1+\cdots+k}\frac{1}{n}\sum_{j=1}^{n}\left[\left(Y_{ij} - (R_{Mij} - R_{fij}, SMB_{ij}, HML_{ij})\begin{pmatrix}\beta\\ s\\ h\end{pmatrix} - \bar{v}\right)\right]^2$$

, the real value of $\begin{pmatrix}\beta\\ s\\ h\end{pmatrix}$, $\bar{v}$ is $\begin{pmatrix}\beta\\ s\\ h\end{pmatrix}^0$, $\bar{v}^0$, then note the it part of the weighted expected loss as M_i, which satisfies:

$$M_i = \frac{2i}{k(1+k)}\frac{1}{n}\sum_{j=1}^{n}\left[\left(Y_{ij} - (R_{Mij} - R_{fij}, SMB_{ij}, HML_{ij})\begin{pmatrix}\beta\\ s\\ h\end{pmatrix} - \bar{v}\right)\right]^2$$
$$= \frac{2i}{1+k}\frac{1}{n}\sum_{j=1}^{n}\left\{\left(\varepsilon_{ij} - \bar{v}^{\simeq 0}\right)^2 - 2\Xi_{ij}\left(\varepsilon_{ij} - \bar{v}^0\right) + (\Xi_{ij})^2\right\} + \delta_n$$

Where

$$\Xi_{ij} \square \left[\left(\begin{pmatrix}\beta\\ s\\ h\end{pmatrix} - \begin{pmatrix}\beta\\ s\\ h\end{pmatrix}^0\right)'\begin{pmatrix}R_{Mij} - R_{fij}\\ SMB_{ij}\\ HML_{ij}\end{pmatrix} + \left(\bar{v} - \bar{v}^0\right)\right]$$

Together with the last three equations, we get M_i, which satisfies

$$M_i = \frac{2i}{1+k}\frac{1}{n}\sum_{j=1}^{n}\left\{\left(\varepsilon_{ij} - \bar{v}^{\simeq 0}\right)^2 - 2\Xi_{ij}\left(\varepsilon_{ij} - \bar{v}^0\right) + (\Xi_{ij})^2\right\} + \delta_n$$

Notice that $\varepsilon_{ij} - \bar{v}^0, j = 1, \cdots, n$ is independent and identically distributed with zero mean and certain variance, then

$$\frac{1}{n}\sum_{j=1}^{n}\left[\left(\begin{pmatrix}\beta\\ s\\ h\end{pmatrix} - \begin{pmatrix}\beta\\ s\\ h\end{pmatrix}^0\right)'\begin{pmatrix}R_{Mij} - R_{fij}\\ SMB_{ij}\\ HML_{ij}\end{pmatrix} + \left(\bar{v} - \bar{v}^0\right)\right]\left(\varepsilon_{ij} - \bar{v}^0\right) = o_p(1)$$

It is clearly that:

$$\frac{1}{n}\sum_{j=1}^{n}\left[\left(\begin{pmatrix}\beta\\ s\\ h\end{pmatrix} - \begin{pmatrix}\beta\\ s\\ h\end{pmatrix}^0\right)'\begin{pmatrix}R_{Mij} - R_{fij}\\ SMB_{ij}\\ HML_{ij}\end{pmatrix} + \left(\bar{v} - \bar{v}^0\right)\right]^2 = o_p(1)$$

It is known that ε_{ij} is independent with δ_n and $\begin{pmatrix}\beta\\ s\\ h\end{pmatrix}$, then we can get that it is the same minimizing M_i and minimizing the following equation:

$$\sum_{j=1}^{n}\left\{-2\Xi_{ij}\left(\varepsilon_{ij}-\bar{v}^{0}\right)+\left(\Xi_{ij}\right)^{2}\right\}$$

Note that

$$\vartheta=\begin{pmatrix}\sqrt{n}\left(\begin{pmatrix}\beta\\ s\\ h\end{pmatrix}-\begin{pmatrix}\beta\\ s\\ h\end{pmatrix}^{0}\right)\\ -\sqrt{n}\left(\bar{v}-\bar{v}^{0}\right)\end{pmatrix},$$

$$\Theta\left(X_{ij}\right)=\begin{pmatrix}\Gamma_{ij}\left(R_M-R_f,SMB,HML\right)_{ij} & \Gamma_{ij}\\ \left(R_M-R_f,SMB,HML\right)_{ij} & 1\end{pmatrix}$$

We can write the last objective function as:

$$Z_n(\vartheta)=\sum_{j=1}^{n}\left[-\frac{2\left(\varepsilon_{ij}-\bar{v}^{0}\right)}{\sqrt{n}}\left(\begin{pmatrix}R_M-R_f\\ SMB\\ HML\end{pmatrix}_{ij}',-1\right)\vartheta+\frac{1}{n}\vartheta'\Theta\left(X_{ij}\right)\vartheta\right]$$

Obviously, $Z_n(\vartheta)$ is convex and can reach its minimum value at

$$\hat{\vartheta}=\sqrt{n}\begin{pmatrix}\left(\begin{pmatrix}\hat{\beta}\\ \hat{s}\\ \hat{h}\end{pmatrix}-\begin{pmatrix}\beta\\ s\\ h\end{pmatrix}^{0}\right)\\ \left(\hat{\bar{v}}-\bar{v}^{0}\right)\end{pmatrix}$$

According to the central limit theorem, the last function is with the following nature:

$$Z_n(\vartheta)\xrightarrow{d}Z_0(\vartheta)=-2\Upsilon'\vartheta+\vartheta'E\left(\Theta\left(X\right)\right)\vartheta$$

According to the central limit theorem, the last function is with the following nature:

$$Z_n(\vartheta)\xrightarrow{d}Z_0(\vartheta)=-2\Upsilon'\vartheta+\vartheta'E\left(\Theta\left(X\right)\right)\vartheta$$

In which, $\Upsilon \sim N\left(0,vol_E\left(\Theta\left(X\right)\right)\right)$ $Z_0(\vartheta)$ is of convexity, which assures the uniqueness of the minimum, then

$$\sqrt{n}\left[\left(\left(\hat{\beta},\hat{s},\hat{h}\right)-\left(\beta,s,h\right)^{0}\right),\left(\hat{\bar{v}}-\bar{v}^{0}_{i}\right)\right]'=\hat{\vartheta}_n=\arg\min \bar{Z}_n(\vartheta)\xrightarrow{d}\hat{\vartheta}_0=\arg\min \bar{Z}_0(\vartheta)$$

Then, the initial objective function M, which is the sum of all of the M_i:

$$M=\sum_{i=1}^{k}\frac{i}{1+\cdots+k}\sum_{j=1}^{n}\left[Y_{ij}-\left(R_{Mij}-R_{fij},SMB_{ij},HML_{ij}\right)\begin{pmatrix}\beta\\ s\\ h\end{pmatrix}-\bar{v}\right]^{2}+\delta_n$$

Obviously, the sum of convex functions is also a convex function, so M can get its Optimal solution. In order to get the asymptotic normality, we consider broadening the conditions:

$$\begin{aligned}M&=\sum_{i=1}^{k}\frac{i}{1+\cdots+k}\sum_{j=1}^{n}\left[Y_{ij}-\left(R_{Mij}-R_{fij},SMB_{ij},HML_{ij}\right)\begin{pmatrix}\beta\\ s\\ h\end{pmatrix}-\bar{v}\right]^{2}\\ &=\sum_{i=1}^{k}\frac{2ki}{1+k}\sum_{j=1}^{n}\left[Y_{ij}-\left(R_{Mij}-R_{fij},SMB_{ij},HML_{ij}\right)\begin{pmatrix}\beta\\ s\\ h\end{pmatrix}-\bar{v}\right]^{2}\\ &\le\frac{2k^{2}}{1+k}\frac{1}{n}\sum_{j=1}^{n}\left(Y_{ij}-\left(R_{Mij}-R_{fij},SMB_{ij},HML_{ij}\right)\begin{pmatrix}\beta\\ s\\ h\end{pmatrix}-\bar{v}\right)^{2}\\ &=\frac{2k^{2}}{1+k}\frac{1}{n}\sum_{j=1}^{n}\left\{\left(\varepsilon_{ij}-\bar{v}^{0}\right)^{2}-2\Xi_{ij}\left(\varepsilon_{ij}-\bar{v}^{0}\right)+\Xi_{ij}^{2}\right\}\end{aligned}$$

$$\sqrt{n}\left[\left(\left(\overline{\beta},\hat{s},\hat{h}\right)-\left(\beta,s,h\right)^{0}\right),\left(\hat{\overline{v}}-\overline{v}^{0}_{i,}\right)\right]' = \hat{\vartheta}_{n} = \arg\min \overline{Z}_{n}(\vartheta) \xrightarrow{d} \hat{\vartheta}_{0} = \arg\min \overline{Z}_{0}(\vartheta)$$

Then we get the asymptotic normality of the estimators of M, and the proof is finished

5. Empirical Analysis

In this section we have empirically demonstrated the new models in the Chinese Second-board Market. In order to get the clear results, we use the three model while regression. The three models are as follows:

$$R_{it} - R_{ft} = \beta(R_{Mt} - R_{ft}) + sSMB_t + hHML_t + \varepsilon_{it}$$

$$E\left[Y \mid \Gamma\right] = \left(R_M - R_f, SMB, HML\right)\begin{pmatrix}\beta\\ s\\ h\end{pmatrix} + \overline{e}$$

$$\mathbb{E}\left[Y \mid \Gamma\right] = \left(R_M - R_f, SMB, HML\right)\begin{pmatrix}\beta\\ s\\ h\end{pmatrix} + \overline{v}$$

Because the calculation process is too complicated, the genetic algorithm is used to obtain the parameter estimation value through python. Few paper focus on the application of the model in the Chinese Second-board Market, so the author uses stock data from February 2011 to November 2018 of the Chinese Second-board Market, and the empirical results are as follows:

Table 1: 300082.XSHE

Method	TSS	β	s	h	$\overline{v}$
LS	0.4076	0.9260	-0.1625	0.8408	-
Upper	0.4080	0.9960	0.1251	0.6208	0.0144
Weighted	0.3811	0.9409	-0.1074	0.9967	0.0186

Table 2: 300107.XSHE

Method	TSS	β	s	h	$\overline{v}$
LS	1.0462	1.0056	-0.2408	0.9033	-
Upper	1.7820	0. 2501	-0.4999	1.49e-03	-2.22e-04
Weighted	1.0221	0.8748	-0.1035	0.8711	0.0302

Table 3: 300127.XSHE

Method	TSS	β	s	h	$\overline{v}$
LS	1.004	1.0830	-0.0403	0.5450	-
Upper	1.0466	1.2385	0.0489	0.6203	0.0379
Weighted	0.9805	1.1232	-0.0729	0.8594	0.0184

Observing the performance of the above three tables, we have a lot of discoveries. Firstly, we find that the TSS of the weighted model is smaller than the LS's and the Upper's, which means the weighted model is better than the others. The reason is that the weighted model deal with the information in error, which makes the model more accurate. Secondly, the positive and negative signs of the estimators β, s, h of the Weighted model are the same as the LS's, meaning the SMB, HML and Market factors have the same nature to stock return like enhancement or attenuation. Moreover, the value of the

$\bar{v}$ in Weighted model is smaller than it in the Upper model, because the first one considers all the information of the k-sample, while the second considers only the maximum expected information of the k-sample. It means that the Upper model isn’t fit the stock market, and it may be useful in financial risk control field.

References

1. Chen et al., Fama-French-factor model,size and book-to-market effect: An empirical investigation of the Chinese stock market. China Economist, 2010, 5, 109-118.
2. C. S. Eun, W. Huang, Asset pricing in China’s domestic stock markets: Is there a logic?. Pacific-Basin Finance Journal, 2007, 15(5): 452-480.
3. E. F. Fama and K. R. French, Common risk factors in the returns on stocks andbonds. Journal of Financial Economics. 1993, 33, 3-56.
4. P. J. Huber Robust statistics. New York: JohnWiley and Sons; 1981.
5. L. Lin, Y. F. Shi and X.Wang, k-sample upper expectation linear regression. J Statist Plan Inference. 2016, 170:15-26.
6. H. Markowitz, Portfolio selection. The Journal of Finance, 1952, 7(1): 77-91.
7. S. Peng, Backward SDE and related g-expectations. In: El Karoui, Mazliak, (Eds.), Backward Stochastic Differential Equations. In: Pitman Research Notes in Math. Series, vol. 1997, 364. 141-159.
8. S. Peng, G-expectation, G-Brownian motion and related stochastic calculus of it’s type. In: Benth et al. (Eds.), The Abel Symposium 2005. In: Abel Symposia, vol. 2. Springer-Verlag, 541-567.
9. S. A. Ross The arbitrage theory of capital asset pricing. Journal of Economic Theory, 1976,13(3): 341-360.
10. W. F. Sharpe, Capital asset prices: A theory of market equilibrium under conditions of risk. The Journal of Finance, 1964, 19(3): 425-442.