Journal of Mathematics and Modeling in Finance (JMMF)
Vol. 1, No. 2, Summer & Autumn 2021
Research paper

An application of stochastic approximation in simulated method of moments

Erfan Salavati[1], **Nazanin Mohseni**[2]

[1] Department of Mathematics, Amirkabir University of Technology, Tehran, Iran
erfan.salavati@aut.ac.ir

[2] Department of Mathematics, Amirkabir University of Technology, Tehran, Iran
n.mohseni@aut.ac.ir

Abstract:
Identifying the structures of dependence between financial assets is one of the interesting topics to researchers. However, there are challenges to this purpose. One of them is the modelling of heavy tail distributions. Distributions of financial assets generally have heavier tails than other distributions, such as exponential distributions. Also, the dependence of financial assets in crashes is stronger than in booms and consequently the skewed parameter in the left tail is more. To address these challenges, there is a function called Copula. So, copula functions are suggested for modelling dependency structure between multivariate data without any assumptions on marginal distributions, which they solve the problems of dependency measures such as linear correlation coefficient. Also, tail dependency measures have analytical formulas with copula functions. In general, the copula function connects the joint distribution functions to the marginal distribution of every variables. With regard, we have introduced a factor copula model that is useful for models where variables are based on latent factor structures. Finally, we have estimated the parameters of factor copula by Simulated method of Moment, Newton-Raphson method and Robbins-Monroe algorithm and have compared the results of these methods to each other.

Keywords: Crash; Heavy Tail; Factor Copula; Simulated Method of Moment; Newton-Raphson Method; Robbins-Monroe Algorithm.
MSC2010 Classifications: 91Gxx, 34L12

1 Introduction

The financial crisis of 2007-2008 was a threat for large financial institutions. In the early 2000s, the United States faced a severe financial credit crisis. They were faced with strange formulas for repaying their debts to the bank, which seemed those debates change over time. As interest rates on loans rose, many borrowers could no repay their debts, and some others were forced to relinquish their assets, including housing. As a result, this reduced the houses prices in the United States. This issue with other economic problems had negative effects on the US economy

[2]Corresponding author
Received: 2021-05-12 Accepted: 2021-07-13
http://dx.doi.org/10.22054/jmmf.2021.13840

rating and the world stock markets. When borrowers were not able to pay their debts, including high interest rates, the entire economic system ran into trouble. European stock prices also fell gradually. In fact, stocks that were even independent of each other fell together suddenly. This indicates that the dependence of random variables in tails is stronger than other regions.

As the history of financial crisis shows, one of the underlying causes of such events is the failure to model and understand the dependence structure of multi variate distributions of returns. A powerful tool for this purpose is Copula. Copulas are multivariate functions that capture the dependence structure of joint distributions [12]. Copulas allow us to abstract away the individual distributional properties and focus only on the dependence of multiple random variables.

One of the most important classes of multi-dimensional models used in economics and finance, are the factor models. Factor models assume that the target variables that we wish to study, depend on some latent variables which we do not directly observe. In modelling market prices, the latent factors can be interpreted as the common factors that drive the prices. One advantage of factor models is that they help us discover these common factors. One of the earliest and most influential factor models in finance has been the CAPM model [11].

The above two mentioned tools (copulas and factor models) can be combined in a single model called the factor copula model [2]. Factor copulas have the advantages of both models. On one hand they can discover the latent factors and on the other hand they focus only on the dependence structure. This comes with a price, and the price is that the estimation of such models becomes difficult. Indeed most of the factor copulas does not posses closed form densities which can be used for estimation. This has resulted in various methods in the literature in order to estimate the coefficients of a factor model. Because of the lack of closed form densities, most of the methods have resorted to simulations. Such methods should inevitably use some kind of stochastic approximation.

In this article, we consider three methods for estimation of factor copula models. All of these models use the idea of method of moments in combination with simulation and hence the name simulated method of moments. The first method comes from [1], and uses a minimization process in order to minimize the distance between the observed and the simulated moments. The second method, uses the straightforward idea of solving a nonlinear system consisting of simulated moments and applies the well-known Newton-Raphson method [3]. The third and final method uses the Robins-Monroe approximation method which is one the earliest and most used methods for stochastic approximation [8].

We have implemented the above three methods and applied them on the historical price data of 470 stocks included in S&P500 index. We have compared the performance of the estimation methods. All codes have been conducted by Python.

2 The copula of a latent factor model

Consider a vector of N variables, $\mathbf{Y}$, with joint distribution $\mathbf{F}_Y$, marginal distributions F_i, and copula $\mathbf{C}$:

$$[Y_1, \cdots, Y_N]' \equiv \mathbf{Y} \approx \mathbf{F}_Y = \mathbf{C}(F_1, \cdots, F_N) \tag{1}$$

The copula is a function that describes the dependence between the variables $Y_1, \cdots, Y_N$ as our observations. We will use existing models to estimate the marginal distributions F_i, and focus on constructing useful new models for the dependence between these variables, $\mathbf{C}$.

According to [1], we consider factor structure, based on a set of $N + K$ latent variables:

Let:

$$X_i = \sum_{k=1}^{K} \beta_{ik} Z_k + \varepsilon_i, \qquad i = 1, 2, \cdots, N \tag{2}$$

so

$$[X_1, \ldots, X_N]^T \equiv \mathbf{X} = \mathbf{BZ} + \varepsilon$$

where

$$\varepsilon_i \sim \text{iid } \mathbf{F}_\varepsilon(\gamma_\varepsilon),$$

$$Z_k \sim \text{inid } \mathbf{F}_{Z_k}(\gamma_k), \quad Z_k \perp \varepsilon_i \quad \forall i, k.$$

Then

$$\mathbf{X} \sim \mathbf{F}_X = \mathbf{C}(G_1(\theta), G_2(\theta), \ldots, G_N(\theta); \theta),$$

where $\theta \equiv [\text{vec}(\mathbf{B})^T, {\gamma_\varepsilon}^T, {\gamma_1}^T, \ldots, {\gamma_K}^T]^T$. The copula of the latent variables $\mathbf{X}$, denoted $\mathbf{C}()$, is used as the model for the copula of the observable variables $\mathbf{Y}$. An important point about the above construction is that the marginal distributions of X_i may be different from those of the original variables Y_i, so $F_i \neq G_i$ in general; on the other hand, due to the good features of copula function, instead of using marginal distribution and the presence of latent factors that complicate their distributions, we used vector structure, $\mathbf{X}$. The copula implied by equation 2 is generally not known in closed form. If $\{F_\varepsilon, F_{z_1}, \cdots, F_{z_K}\}$ are all Gaussian distributions, in which case the variable $\mathbf{X}$ is multivariate Gaussian, implying a Gaussian copula. For other choices of $\{F_\varepsilon, F_{z_1}, \cdots, F_{z_K}\}$ the joint distribution of $\mathbf{X}$, and the copula of $\mathbf{X}$, is generally not known in closed form [1]. However, it is simple to simulate from $\{F_\varepsilon, F_{z_1}, \cdots, F_{z_K}\}$ for many classes of distributions, and it can be extracted properties of the copula from simulated data , such as Rank correlation, Kendalls tau, and Quantile dependence. These simulated dependence measures can be used in the SMM estimation method of [2] Oh and Patton (2013), which is briefly described in 3.1.

Although most factor copulas do not have a closed-form density, we can use results from extreme value theory to obtain analytically results on the tail dependence

implied by a given factor copula model. According to the simple linear structure generating the factor copula, these results are relatively easy to obtain. Recall the definition of tail dependence for two variables X_i, X_j with marginal distributions G_i, G_j [2]:

$$\tau_{ij}^L = \lim_{q \searrow 0} \mathbf{P}[G_j(X_j) \leq q | G_i(X_i) \leq q] = \lim_{q \searrow 0} \frac{\mathbf{C}_{ij}(q,q)}{q} \tag{3}$$

$$\tau_{ij}^U = \lim_{q \nearrow 1} \mathbf{P}[F_j(X_j) > q | F_i(X_i) > q] = \lim_{q \nearrow 1} \frac{1 - 2q + \mathbf{C}_{ij}(q,q)}{1-q}.$$

These are lower and upper tail dependence measures, respectively and the probability of both variables lying below their q quantile, for q limiting to zero, scaled by the probability of one of these variables lying below their q quantile.

In this article, we used the Tau-Kendall rank dependency measurements to estimate the parameter θ. According to [2], rank dependence measures for the pair (X_i, X_j) are defined as:

$$\tau^{ij} \equiv 4\mathbf{E}[\mathbf{C}_{ij}(F_i(X_i), F_j(X_j))] - 1, \tag{4}$$

where $\mathbf{C}_{ij}$ is the copula of (X_i, X_j). The sample counterparts are defined as:

$$\hat{\tau}^{ij} = \frac{4}{T} \sum_{t=1}^{T} \hat{C}_{ij}(\hat{F}_i(\hat{X}_{it}), \hat{F}_j(\hat{X}_{jt})) - 1, \tag{5}$$

where

$$\hat{F}_i(x) \equiv \frac{1}{T+1} \sum_{t=1}^{T} 1\{\hat{X}_{it} \leq x\},\ \hat{C}_{ij}(u,v) \tag{6}$$

$$\equiv \frac{1}{T+1} \sum_{t=1}^{T} 1\{\hat{F}_i(\hat{X}_{it}) \leq u, \hat{F}_j(\hat{X}_{jt}) \leq v\}. \tag{7}$$

counterparts based on simulations are denoted by $\tilde{\tau}_{ij}(\theta)$.

In propositions 1 and 2 [1], we can see lower and upper tail dependence measures using factor models for single factor and multi-factor copula models. Beside, according to proposition 3 [1], if we know the distribution of latent variables and errors, we can obtain constant values of the formulas for the tail dependencies stated in propositions 1 and 2 using factor models. These propositions show that when the coefficients on the common factor have the same sign, and the common factor and idiosyncratic variables have the same tail index, the factor copula generates upper and lower tail dependence. If either $\mathbf{Z}$ or ε is asymmetrically distributed, then the upper and lower tail dependence coefficients can differ, which provides this model with the ability to capture differences in the probabilities of joint crashes and joint booms. When either of the coefficients on the common factor are zero, or if they have differing signs, then the upper and lower tail dependence coefficients

are both zero. These propositions consider the case that the common factor and idiosyncratic variables have the same tail index.

According to [1], we are likely to estimate parameter θ in below model:

$$\begin{gathered} X_i = \sum_{k=1}^{K} \beta_{ik} Z_k + \varepsilon_i, \qquad i = 1, 2, \cdots, N \\ [X_1, \ldots, X_N]' \equiv \mathbf{X} = \mathbf{BZ} + \varepsilon, \\ \varepsilon_i \sim \mathrm{N}(\mu = 0, \sigma^2 = 1), \\ Z_k \sim \text{Skew t}(\nu = \infty, \lambda), \qquad Z_k \perp \varepsilon_i \quad \forall i, k. \\ \mathbf{X} \sim \mathbf{F}_X = \mathbf{C}(G_1(\theta), G_2(\theta), \ldots, G_N(\theta); \theta), \\ \theta \equiv [\mathrm{vec}(\mathbf{B})', \lambda]', \end{gathered} \tag{8}$$

In this model, we considered the Normal distribution for errors and the latent variables had the Skew-t distribution with the degree of freedom infinity.

According to the proposition 4 [1], we estimated the number of latent factors in our model. To further explain, we should sort the eigenvalues of the matrix of $\hat{\mathbf{R}}_T^y$ and then consider the number of those whose values are greater than one, and if T tends to infinity, $\hat{K}_T$ will tend K in probability . This is obvious that the distributions and copula are continuous, and the *iid* assumption can be relaxed by invoking assumption 2 of Oh and Patton (2013) [2] and then we estimated standardized residuals from the original data. With assumptions (1)(2) of proposition 4 we find that $\hat{K}_T$ provides, asymptotically, a lower bound on the true number of factors; it will miss factors that are $\mathbf{g}_k(\mathbf{R}) \leq 1$ for $k \in [1, K]$. If N diverges with T then this cannot happen and assumption (3) will hold automatically (see Chamberlain and Rothschild 1983; Bai and Ng 2002) [10], while in this proposition setting of finite N this assumption may not hold. In such cases using a threshold of one provides a lower bound on the true number of factors.

3 Estimation of factor copula

As mentioned, joint density factor copula models do not have a closed-form and the methods that we already know, such as Maximum Likelihood estimation for estimating of parameters of copula are not efficient. According to [2], we use the Simulated Method of Moments estimation or SMM. This method is suitable for models that have a large number of unknown dependent parameters or their density functions which are not closed-form. Also, it can be estimated the marginal distributions using the EDF. Considering the following GARCH model (with p = q = 15) augmented with lagged market return information, that are used to filter each of the individual return series:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \cdots + \alpha_{\mathrm{q}} Y_{t-\mathrm{q}} + \varepsilon_t, \qquad t = 1, 2, \cdots, T \tag{9}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \cdots + \alpha_{\mathrm{q}}\varepsilon_{t-\mathrm{q}}^2 + \delta_1\sigma_{t-1}^2 + \cdots + \delta_{\mathrm{p}}\sigma_{t-\mathrm{p}}^2,$$
$$\alpha_0 \geq 0,\ \alpha_i, \delta_i \geq 0,\ i = \{1, 2, \cdots\}$$

In this model, ε_t represents error which is equal to $\sigma_t\eta_t$, and also η_t is residual variable. Indeed, residuals appear in errors. We estimated the distribution of the standardized residuals as the EDF using GARCH model and observations Y. From now on, we will estimate the factor copula parameters, $\hat{\theta}_0 \in \Theta$, with vector of residuals, $\hat{\eta}_t$. Let $\tilde{\mathbf{m}}_{\mathbf{S}}(\theta)$ be a $N \times N$ matrix of dependence measures computed using S simulations from $\mathbf{F}_{\mathbf{X}}(\theta)$, and $\{\mathbf{X}_s\}_{s=1}^{S}$ and let $\hat{\mathbf{m}}_{\mathbf{T}}$ be the corresponding vector of dependence measures computed using the standardized residuals $\{\eta_{\mathbf{t}}\}_{t=1}^{T}$.

We will now explain the three estimation methods used in this paper to obtain the values of the factor copula parameters.

3.1 Simulated Method of Moments

Define the difference between these as

$$\mathbf{g}_{T,S}(\theta) \equiv \hat{\mathbf{m}}_{\mathbf{T}} - \tilde{\mathbf{m}}_{\mathbf{S}}(\theta), \tag{10}$$

This SMM estimator is based on searching across $\theta \in \Theta$ to make 10 as small as possible. The estimator is defined as $\mathcal{Q}_{T,S}(\theta)$ where

$$\mathcal{Q}_{T,S}(\theta) \equiv \mathbf{g}'_{T,S}(\theta)\hat{\mathbf{W}}_T\mathbf{g}_{T,S}(\theta), \tag{11}$$

so

$$\hat{\theta}_{T,S} \equiv \arg\min_{\theta\in\Theta} \mathcal{Q}_{T,S}(\theta). \tag{12}$$

Weight matrix $\hat{\mathbf{W}}_T$ is a positive definite matrix, which may depend on the data. First, we use equation 5 for estimating parameter θ. We use the closing price of each stock to obtain daily returns and sort them by date. We used time series models and estimate unknown parameters using the same model. The reason why we used daily returns instead of stock prices is that time series analysis of returns is more appropriate than price time series and we expect the returns (not prices) change linearly. So, we obtained the residual variables using the GARCH model. We also estimated their marginal distributions empirically and finally obtained the Tau-Kendall rank dependency measures in the sample mood, it means $\hat{\mathbf{R}}_T^r$. We set $\hat{\mathbf{m}}_{\mathbf{T}}$ as the vector of the calculated dependency measurements of the standard residues $\{\hat{\eta}_t\}_{t=1}^{T}$. According to proposition 4 [1], let $\hat{\mathbf{m}}_{\mathbf{T}} = \hat{\mathbf{R}}_T^r$, Which is a $N \times N$ matrix. Then the number of latent factors returns the kth-largest eigenvalue of the matrix $\hat{\mathbf{R}}_T^r$ which is greater than 1. Moreover, we used data from S&P500. Our data, $\mathbf{Y}_i$, is related to the price of 500 US stocks from 2013 to 2018, which is 1258 workable days. This data includes stock name, stock date, open and close prices, highest and lowest price and volume. We sorted the stocks by date in a table whose columns bear the names of stocks and the rows of that date and stock yield in each sheet.

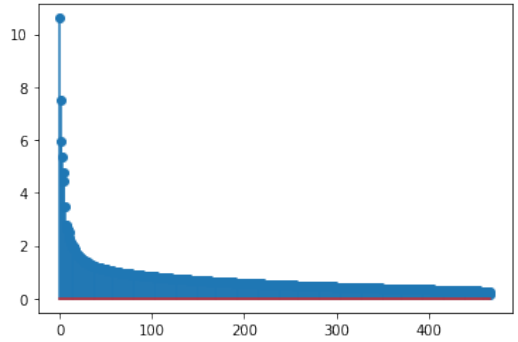

Figure 1: Number of latent factors with 470 returns

We have removed sheets whose yields have not been defined (price information is incomplete), and finally, the number of our stocks became equal to $N = 470$. The reason why we use returns instead of prices is that we expect returns to change linearly. Also, the dependence between returns is not very interesting for us, and what is important for us is the dependence between the residuals. The eigenvalue greater than 1 with $N = 470$ is equal to:

$$\hat{K}_T = \max\{k : \mathbf{g}_k(\hat{\mathbf{R}}_T^r) > 1\} = 67$$

109.595546	19.723693	10.608553	7.50174234	5.9589867	5.3865252
4.78211970	4.4297649	3.4951303	2.81086816	2.6780819	2.5997679
2.49934868	2.2946449	2.1583033	2.01111380	2.0072519	1.9677944
1.94252628	1.8877604	1.7358756	1.71122559	1.6496980	1.6233613
1.58726184	1.5445178	1.5050240	1.48155870	1.4568980	1.4422919
1.40280982	1.3870849	1.3687812	1.33594704	1.3214989	1.2980555
1.27601869	1.2611421	1.2521566	1.24031450	1.2291115	1.2244026
1.20202774	1.1876368	1.1780731	1.16482782	1.1592405	1.1562463
1.14476479	1.1336676	1.1163572	1.11225577	1.1042156	1.1009008
1.08945546	1.0793844	1.0679219	1.05886555	1.0532051	1.0491200
1.04006913	1.0279877	1.0198420	1.01858352	1.0134598	1.0065245
1.00231535					

We drew the Scree plot of eigenvalues. Since the first and the second eigenvalues are much larger than the third eigenvalue, we cut off them from the figure to have an obvious figure.

In this article, we took 10 stocks from the S&P500 index, and the number of latent factors being equal to one. Then, we can see its results and scree plot:

$$\hat{K}_T = \max\{k : \mathbf{g}_k(\hat{\mathbf{R}}_T^r) > 1\} = 1$$

3.2215	0.9858	0.8845	0.8331	0.7903
0.7454	0.6731	0.6534	0.6091	0.6035

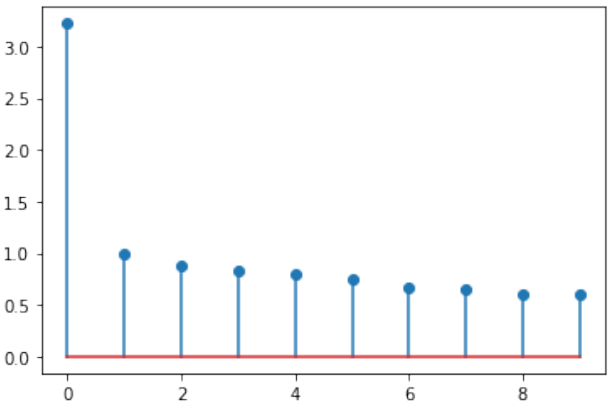

Figure 2: The number of latent factors with 10 returns

According to the SMM estimator, we also obtain $\tilde{\mathbf{m}}_{\mathbf{S}}(\theta)$. As we said, matrix $\tilde{\mathbf{m}}_{\mathbf{S}}(\theta)$ is the measure of dependence calculated from the 1000 simulations of $\mathbf{F}_{\mathbf{X}}(\theta)$ and $\{\mathbf{X}_s\}_{s=1}^{S}$. We calculate the Tau-Kendall dependency measure with the simulated latent variables, and let them, $\tilde{\mathbf{m}}_{\mathbf{S}}(\theta)$. Furthermore, This vector is a function with the parameter θ which includes the skewness λ and the latent factor coefficients matrix Z, which means β_i, $i = \{1, 2, \cdots, N\}$. The results are given in chapter 4.

3.2 Newton-Raphson Method

Suppose we had a family of distributions, $\mathrm{p}(x|\theta)$, and we wish to estimate the r-dimensional parameter vector θ by matching a m-dimensional vector of moments, $\mu(\theta) = \mathbf{E}[\tilde{\mathbf{m}}_{\mathbf{S}}(\mathbf{x})|\theta]$, to a fixed vector $\mu_0 = \hat{\mathbf{m}}_{\mathbf{T}}(\theta)$.

If $\mathbf{E}[\tilde{\mathbf{m}}_{\mathbf{S}}(\mathbf{x})|\theta]$ can be expressed analytically in closed form, we can obtain the moments estimate $\hat{\theta}$ using the Newton-Raphson method, as follows [3]. Start with a guessed value, θ_1. Then for $t = 1, 2, \cdots$, update the guess to

$$\theta_{t+1} = \theta_t + [\mathbf{E}[\tilde{\mathbf{m}}_{\mathbf{S}}(\mathbf{x})|\theta_t]']^{-1}(\hat{\mathbf{m}}_{\mathbf{T}}(\theta) - \mathbf{E}[\tilde{\mathbf{m}}_{\mathbf{S}}(\mathbf{x})|\theta_t]) \tag{13}$$

where $\mu'(\theta_t) = \mathbf{E}[\tilde{\mathbf{m}}_{\mathbf{S}}(\mathbf{x})|\theta_t]'$ is the matrix of derivatives of $\mu(\theta_t) = \mathbf{E}[\tilde{\mathbf{m}}_{\mathbf{S}}(\mathbf{x})|\theta_t]$ with respect to θ. Here we are concerned with problems for which $\mathbf{E}[\tilde{\mathbf{m}}_{\mathbf{S}}(\mathbf{x})|\theta_t]$ cannot be computed in closed form; instead, we can estimate it, for any given value of θ, by simulation of N draws of X from the distribution $\mathrm{p}(x|\theta)$.

Suppose now the problem is overdetermined, with more moments specified than parameters in the model, and we would like the θ that gives the best least-squares fit, minimizing $||\hat{\mathbf{m}}_{\mathbf{T}}(\theta) - \mathbf{E}[\tilde{\mathbf{m}}_{\mathbf{S}}(\mathbf{x})|\theta_t]||^2$. The normal equations are $\mathbf{E}[\tilde{\mathbf{m}}_{\mathbf{S}}(\mathbf{x})|\theta_t]'(\hat{\mathbf{m}}_{\mathbf{T}}(\theta) -$

$\mathbf{E}[\tilde{\mathbf{m}}_{\mathbf{S}}(\mathbf{x})|\theta_t]) = 0$, which we can again solve by Newton-Raphson, using iterative least squares. Starting out at a guess θ_1, for $t = 1, 2, \cdots$, the updated guess is

$$\theta_{t+1} = \theta_t + [\text{least squares regression of}\,(\mu_0 - \mu(\theta_t))]\,\text{on the matrix}\,\mu'(\theta_t)]. \quad (14)$$

After solving equation 14, θ_{t+1} will equal with $\theta_t + (\mu^T(\theta_t)\mu(\theta_t))^{-1}\mu'(t)^T(\mu_0 - \mu(\theta_t))$. One can apply the Monte Carlo method as before, using the estimates $\hat{\mu}(\theta)$ and $\hat{\mu}'(\theta)$ from the previous sections and converging to an approximate least squares fit by simulating a large number N of draws once the estimate θ_t is close to convergence [3].

As we know, obtaining $\mathbf{E}[\tilde{\mathbf{m}}_{\mathbf{S}}(\mathbf{x})|\theta_t]$ with closed-form which is random function, is impossible. Therefore, we cannot compute $\mathbf{E}[\tilde{\mathbf{m}}_{\mathbf{S}}(\mathbf{x})|\theta_t]'$. We used Secant method and as a result of Broyden's method that because of matrix form. The secant method is a root-finding algorithm that uses a succession of roots of secant lines to better approximate a root of a function f. The secant method can be thought of as a finite-difference approximation of Newton's method. The secant method is defined by the recurrence relation:

$$\theta_{t+1} = \theta_t + (\frac{\mu(\theta_t) - \mu(\theta_{t-1})}{\theta_t - \theta_{t-1}})^{-1}(\mu_0 - \mu(\theta_t)) \quad (15)$$

As can be seen from the recurrence relation, the Secant method requires two initial values, θ_0 and θ_1, which should ideally be chosen to lie close to the root [4,5]. Broyden's method is a generalization of the secant method to more than one dimension. In numerical analysis, Broyden's method is a quasi-Newton method for finding roots in r variables. Newton's method for solving $\mu(\theta) = 0$ uses the Jacobian matrix, J, at every iteration. However, computing this Jacobian is a difficult and expensive operation. The idea behind Broyden's method is to compute the whole Jacobian only at the first iteration and to do rank-one updates at other iterations [6, 7].

We supposed $\mu(\theta_t) \to \mu_t$ and $\mu'(\theta_t) \to J_t$ which is $m \times r$ matrix $(m = r)$. So

$$J_t = J_{t-1} + \frac{(\mu_t - \mu_{t-1}) - J_{t-1}(\theta_t - \theta_{t-1})}{||\theta_t - \theta_{t-1}||^2}(\theta_t - \theta_{t-1})^T, \quad (16)$$

and

$$\theta_{t+1} = \theta_t + J_t^{-1}(\mu_0 - \mu_t). \quad (17)$$

If $m > r$, we have equation 16 and

$$\theta_{t+1} = \theta_t + (J_t^T J_t)^{-1} J_t^T(\mu_0 - \mu_t). \quad (18)$$

Finally we used our data for calculating of parameters, θ, and MSE in chapter 4.

3.3 Robbins-Monroe algorithm

The RobbinsMonro algorithm, introduced in 1951 by Herbert Robbins and Sutton Monro, presented a methodology for solving a root problem, where the function is represented as an expected value.

Assume that we have a function $M(\theta)$, and a constant α, such that the equation $M(\theta) = \alpha$ has a unique root at θ_0. It is assumed that while we cannot directly observe the function $M(\theta)$, we can instead obtain measurements of the random variable $N(\theta)$ where $\mathbf{E}[N(\theta)] = M(\theta)$. The structure of the algorithm is to then generate iterates of the form:

$$\theta_{n+1} = \theta_n - a_n(N(\theta_n) - \alpha)$$

Here, $a_1, a_2, \ldots$ is a sequence of positive step sizes. Robbins and Monro proved [8] that θ_n converges in L^2 (also in probability) to θ, and Blum [9] later proved the convergence is actually with probability one, provided that:

- $N(\theta)$ is uniformly bounded,
- $M(\theta)$ is nondecreasing,
- $M'(\theta_0)$ exists and is positive, and
- The sequence a_n satisfies the following requirements:

$$\sum_{n=1}^{\infty} a_n = \infty, \ \sum_{n=1}^{\infty} a_n^2 < \infty.$$

In this algorithm using SMM method, we let:

(i) Moment of $\tilde{\mathbf{m}}_{\mathbf{S}}(\theta)$ as a function with random parameter θ.

(ii) A particular sequence of steps which satisfy these conditions, and was suggested by RobbinsMonro, have the form: $a_n = \frac{a}{n}$, for $a > 0$. Other series are possible but in order to average out the noise in $N(\theta)$, the above condition must be met.

(iii) Suppose we obtain $\hat{\mathbf{m}}_{\mathbf{T}}$ empirically, which is equal to α. Our purpose is that $\tilde{\mathbf{m}}_{\mathbf{S}}(\theta)$ equal to α as much as possible.

(iv) For updating θ in each step, we have:

$$\theta_{n+1} = \theta_n - a_n(\tilde{\mathbf{m}}_{\mathbf{S}}(\theta) - \hat{\mathbf{m}}_{\mathbf{T}}), \tag{19}$$

(v) Also we simulated with 1000 samples to 1000 times.

(vi) To estimate the parameters of λ, β, we used two elements of moments [0,1] and [1,2]. Then, we update the parameters every times by 19.

We obtain the optimal value of θ and MSE, and the results are given in chapter 4.

4 Numerical result

In this chapter, we fit the model with the three methods described in Chapter 3 with S&P500 index data. For simplicity, we consider N=10 to have one latent factor and a state that latent factor coefficients are common, it means: $\beta_1 = \cdots = \beta_{10} = \beta$. The reason why we used the model with one latent factor is that the first factor is significantly different from the other factors, so the one-factor model is a good model to estimate.**??** For simplicity, we assumed the weight matrix $\hat{\mathbf{W}}_T$ in Equation 32 is Identity matrix, $\hat{\mathbf{W}}_T = \mathbf{I}$. The results for simulated method of moment are shown in the below table:

Table 1: The result of parameters with the Simulation Method of Moment

Parameter	Initial value	Optimal value
β	-0.6	-0.5999985
λ	-2	-1.9991886
MSE $= 8.0508306e-09$		

On the other hand, we estimated the value of $\mathcal{Q}$ and that was 9.68755201. Although this method failed to reduce the value of objective function, it could be a optimal method to estimate the parameters of copula. We also fit the data on the Newton-Raphson method and we obtained results in blow table, when $m = 3, r = 2$:

Table 2: The result of parameters with the Newton-Raphson Method

Parameter	Initial value	Optimal value
β	-0.6	0.6158638421
λ	-2	-0.5999471132
MSE $= 0.000924684$		

For the Robbins-Monro algorithm, we have brought the results in the blow table:

Table 3: The result of parameters with the Robbins-Monro algorithm

Parameter	Initial value	Optimal value
β	-0.6	-0.55929763
λ	-2	-2.13245125
MSE = 0.05362494		

5 Conclusion

By comparing the results, we found that although the simulated method of moment was performed with more time, the MSE value decreased and the optimal parameters values became closer to the initial parameters values. Therefore, this method is more suitable than the other two methods.1

Also, despite the fact that Newton-Raphson method is faster to process, it cannot be a good method to estimate copula parameters because of calculating the values of parameters.2

It should be noted that the Robbins-Monro algorithm is a good estimator, too. In this method, the values of optimal parameters is approximately close to initial parameters. Therefore, it can be an appropriate method to estimate the factor copula parameters.3

Bibliography

[1] Oh, Dong Hwan and Patton, Andrew J *Modeling dependence in high dimensions with factor copulas* Journal of Business & Economic Statistics,Taylor & Francis, 35(1):139–154, 2017.

[2] Oh, Dong Hwan and Patton, Andrew J *Simulated method of moments estimation for copula-based multivariate models* Journal of the American Statistical Association,Taylor & Francis, 108(502):689–700, 2013.

[3] Gelman, Andrew *Method of moments using Monte Carlo simulation* Journal of Computational and Graphical Statistics,Taylor & Francis, 4(1):36–54, 1995.

[4] Avriel, Mordecai *Nonlinear programming: analysis and methods* Courier Corporation, 220–221 ,2003.

[5] Allen III, Myron B and Isaacson, Eli L *Numerical analysis for applied science* John wiley & sons, 35:188–195, 2011.

[6] Dennis Jr, John E and Schnabel, Robert B *Numerical methods for unconstrained optimization and nonlinear equations* SIAM, 168–193, 1996.

[7] Kvaalen, Eric *A faster Broyden method* BIT Numerical Mathematics, Springer, 31(2):369–372, 1991.

[8] Robbins, Herbert and Monro, Sutton *A stochastic approximation method* The annals of mathematical statistics ,JSTOR, 22(3):400–407, 1951.

[9] Blum, Julius R and others *Approximation methods which converge with probability one* The Annals of Mathematical Statistics, Institute of Mathematical Statistics, 25(2):382–386, 1954.

[10] Chamberlain, Gary *Funds, factors, and diversification in arbitrage pricing models* Econometrica: Journal of the Econometric Society, JSTOR, 1305–1323, 1983.

[11] French, Craig W. "The Treynor capital asset pricing model." Journal of Investment Management 1.2 (2003): 60-72.

[12] Cherubini, Umberto et al. Copula methods in finance. (2004).

How to Cite: Erfan Salavati[1], Nazanin Mohseni[2], *An application of stochastic approximation in simulated method of moments*, Journal of Mathematics and Modeling in Finance (JMMF), Vol. 1, No. 2, Pages:49–61, (2021).