Semiparametric Value-At-Risk Estimation of Portfolios

A replication study of Dias (*Journal of Banking & Finance*, 2014)

Jiahua Xu*

International Journal for Re-Views in Empirical Economics, Volume 3, 2019-6, DOI: 10.18718/81781.15

JEL: C51, G01, G11, G17
Keywords: Multi-asset portfolios, Risk management, Tail probability, Tail risk, Multivariate extreme value theory, Value-at-Risk, Replication study

Data Availability: The original data used by Dias and in this replication were downloaded from *Datastream* (paid content). The MATLAB-code to reproduce the results of this replication can be downloaded at IREE's data archive (DOI: 10.15456/iree.2018304.055009).

Please Cite As: Xu, Jiahua (2019). Semiparametric Value-At-Risk Estimation of Portfolios. A replication study of Dias (Journal of Banking & Finance, 2014). *International Journal for Re-Views in Empirical Economics*, Vol 3(2019-6). DOI: 10.18718/81781.15

Abstract

This paper aims to replicate the semiparametric Value-At-Risk model by Dias (2014) and to test its legitimacy. The study confirms the superiority of semiparametric estimation over classical methods such as mixture normal and Student-*t* approximations in estimating tail distribution of portfolios, which can be credited to the model's uniqueness in combining strengths of both extreme value theory (EVT) models and other multivariate models. The author however discovers, in one instance, the infeasibility of the Dias model, and suggests a modification.

*Research associate in the Institute of Insurance Economics at the University of St. Gallen in St. Gallen, Tannenstrasse 19, 9000 St.Gallen, Switzerland. E-mail: jiahua.xu@unisg.ch

Received October 31, 2018; Revised July 30, 2019; Accepted August 1, 2019; Published November 20, 2019.

1 Introduction

Diversification is a common strategy in risk management to reduce non-systematic risk by investing in a variety of, as opposed to single, assets (Sullivan 2003). This variety of assets as a whole is often called an investment portfolio. However, when carelessly selected, the assets in a portfolio can have considerable correlations with each other, which can impair the very purpose of diversification. Through the deepening of integration and interactions between industries, the financial performance of one company is increasingly dependent on the others. Global financial recessions have attested to the interconnection of business on an international scale. To avoid the occurrence of extreme losses, which are most likely caused by the synergy of correlated assets in their "down" times, correlations between companies are to be minimized.

Investors' goal is to maximize the return on their investments. Yet, stock prices are more frequently observed to crash than to jump, and the crashes are often too destructive to be compensated by jumps. Intuitively it implies that stock returns follow a heavy tailed distribution. To approximate distributions with heavy tails, Student-t and mixture normal (Newcomb, 1980) estimations are historically used as two classical univariate models. When we only concentrate on extreme losses, extreme value theory (EVT) models can be more appropriate as they are suited to fit only the tail part of a distribution. While such models are all reasonably suitable for individual assets, when estimating portfolio losses, they however treat a portfolio as a synthetic single asset, through which important information on assets' correlations are lost.

Motivated to improve distribution estimation for extreme losses of portfolios, Dias (2014) provides a multivariate semiparametric estimation of portfolio tail risk. The estimation of model parameters relies on empirical time series data of each component asset of the portfolio. As such, compared to the traditional EVT model, the model by Dias (2014) is more comprehensive and reliable in that it captures information on individual assets in a portfolio.

This paper aims to replicate the model by Dias (2014) and to test its legitimacy and executability. To do this, we apply the model to the sample collected in strict accordance with the description in Dias (2014). Through implementation of the model, we discover a potential deficiency in a calculation step that might skew the results. We then propose a simple algebraic transformation to eliminate the computational problem without altering the central idea of the model.

As Dias (2014) delivers an extensive theoretical background of the model, this paper can be seen as an extension of her paper as it examines the feasibility of the model, proposes new ideas on its execution and compares results using the same dataset.

2 Research design and methodology

2.1 Semiparametric approximation of extreme losses

The methodology of multivariate semiparametric approximation is based on extreme value theory (de Haan, 1977). A univariate extreme value distribution can take on different forms. When the distribution is heavy tailed, its cumulative distribution function takes the form of:

$$G(x) = \exp\left[-\left(1+\gamma\frac{x-b}{a}\right)^{-\frac{1}{\gamma}}\right],\ 1+\gamma\frac{x-b}{a} > 0.$$

We name γ, a, b the shape, location and scale parameters. The model is improved by multivariate semiparametric approximation in that it retains the dimensionality of a portfolio, which consequently demands the separate estimation of γ, a, b for each asset therein. If we define function I as:

$$I_X(x) = \begin{cases} 0, & x < X \\ 1, & x \geq X \end{cases} \tag{1}$$

we can write down the estimated survival function of extreme losses using the semiparametric method as:

$$\widehat{S(l)} := \frac{\sum_{t=1}^{n} I_l(\tilde{L}_t)}{nc(l)} \tag{2}$$

where

$$\tilde{L}_t = \hat{\boldsymbol{a}}\frac{\sqrt{\left[c(l)^{\hat{\gamma}}\left(1+\hat{\gamma}\frac{\boldsymbol{L_t}-\hat{\boldsymbol{b}}}{\hat{\boldsymbol{a}}}\right)-1\right]}}{\hat{\gamma}} + \hat{\boldsymbol{b}}$$

when the portfolio is average weighed, $\boldsymbol{L}_t = \{L_{it}\}$, the vector of losses of firms included in the portfolio at time t, $\hat{\boldsymbol{\gamma}} = \{\hat{\gamma}_i\}$ and likewise $\hat{\boldsymbol{a}} = \{\hat{a}_i\}$, $\hat{\boldsymbol{b}} = \{\hat{b}_i\}$.
It is now apparent that $\widehat{S(l)}$ necessitates information for individual assets. As a matter of fact, $\widehat{S(l)}$ can be seen as a derivation of the empirical survival function:

$$S(l) = P(\boldsymbol{L} > l) = \frac{\sum_{t=1}^{n} I_l(L_t)}{n},\ \text{where } \boldsymbol{L} = \{L_t\} \text{ is the time series of synthesized portfolio losses.}$$

Extreme losses are by their very nature rather occasional, and their distribution is rather difficult to model statistically. The “trick” to overcoming this obstacle using semiparametric approximation is to first scale up the losses such that more observations would be considered in the model, and then to scale down the probability that an observation exceeds the adjusted threshold value. That is why we use $\tilde{L}_t$ to estimate the distribution, which can be seen as the adjusted loss value scaled up from the original loss value L_t. $c(l)$ is the so-called scaling constant which varies dependent on l. Although this approach slightly deviates from the methodology in Dias (2014), the fundamental ideas are in accord with each other. The remainder of this section explains how the relevant parameters are estimated and presents the estimated results using our sample data. In Section 3, we highlight why an adaptation of the original method in Dias (2014) is necessary.

2.2 Sample description

To have comparable results, we collected data of 100 firms from Datastream as described in Dias (2014). As portfolio returns are studied here, these 100 firms will later be considered as an equally weighted 100-asset portfolio. Selected firms from those 100 were also used to form a 10-asset portfolio and a 50-asset portfolio. In order to control for variation in capital sizes, firms chosen are evenly distributed in terms of market capitalization. In total, 34 firms are from the then S&P LargeCap 500 list, 33 firms S&P MidCap 400 and 33 firms S&P SmallCap 600. In the 10-asset portfolio, the allocation is 4, 3, 3 and in the 50-asset portfolio 17, 17, 16. The Appendix lists all the firms whose price history is used in Dias (2014) and by design also in this paper. From Datastream, daily price data with Datatype "Price (Adjusted-Default)" covering the period from December 31, 1999 to January 1, 2014 were extracted. For each firm 3,654 adjusted price values were available. No filtering of price values was conducted.

2.3 Parameter estimation

If not otherwise specified in this paper, an asset (a firm) or a portfolio of assets is denoted by subscript i, and a specific day t. When $i = 1, 2, \ldots, 100$, a specific firm is referred to; when $i = \mathbf{10, 50, 100}$ (**bold**), a portfolio of 10-asset, 50-asset or 100-asset is referred to. The daily log return $R_{i,t}$ of firm i on day t is calculated as:[1]

$$R_{i,t} = 100 \ln(p_{i,t}/p_{i,t-1}),\ i = 1, 2, \ldots, 100,\ t = 1, 2, \ldots, 3653,$$

where $p_{i,t}$ is the price of firm i on day t which can be obtained directly from Datastream without any further manipulation. The daily log return of portfolio i is the arithmetic average of its component assets' daily log returns: $R_{i,t} = \overline{R_{i,t}}$, $i = \mathbf{10, 50, 100}$, $t = 1, 2, \ldots, 3653$.
In the instance of a 100-asset portfolio, $R_{\mathbf{100},t} = \overline{\boldsymbol{R}_{\mathbf{100},t}} = \frac{\sum_1^{100} R_{i,t}}{100}$, $\boldsymbol{R}_{\mathbf{100},t} = \{R_{1,t}, R_{2,t}, \ldots, R_{100,t}\}$, $t = 1, 2, \ldots, 3653$. Same goes for $\boldsymbol{R}_{\mathbf{10},t}$ and $\boldsymbol{R}_{\mathbf{50},t}$.

The four fundamental statistics: mean, standard deviation, skewness and kurtosis of each firm's daily log return $\boldsymbol{R}_i = \{R_{i,1}, R_{i,2}, \ldots, R_{i,3653}\}$ are presented in the Appendix. For portfolios, the summary statistics are presented separately in Table 1, in correspondence with Dias (2014). In this study, the distribution of losses is examined. Thus, from now on we employ daily loss[2] $L_{i,t}$ for parameter estimation. $L_{i,t}$ is simply the opposite of $R_{i,t}$, i.e.:
$L_{i,t} = -R_{i,t}$, $i \in \{1, 2, \ldots, 100\} \cup \{\mathbf{10, 50, 100}\}$, $t = 1, 2, \ldots, 3653$.

For any firm/portfolio i, the elements in its daily loss vector $\{L_{i,1}, L_{i,2}, \ldots, L_{i,3653}\}$ can be sorted in a descending way such that: $L_{i,(1)} \geq L_{i,(2)} \geq \ldots \geq L_{i,(3653)}$, $i \in \{1, 2, \ldots, 100\} \cup \{\mathbf{10, 50, 100}\}$.

[1] Although not clearly stated in Dias (2014), the daily log returns were scaled up 100 times according to the statistics summary in the paper.

[2] In Dias (2014), both return and "loss return" are denoted by R. To avoid confusion, we use different denotations for return and loss in this paper.

2.3.1 The moment parameter

The first- and second-order moment parameters are calculated using our sample data as:

$$M_i := \frac{1}{548}\sum_{j=1}^{548} \ln L_{i,(j)} - \ln L_{i,549}, \tag{3}$$

$$N_i := \frac{1}{548}\sum_{j=1}^{548} (\ln L_{i,(j)} - \ln L_{i,549})^2,\ i \in \{1, 2, \ldots, 100\} \cup \{\mathbf{10}, \mathbf{50}, \mathbf{100}\}, \tag{4}$$

where the number of upper order statistics 548 is determined by multiplying 3653 by 15%. This number represents the number of observations large enough to be included in the tail distribution and is denoted by k in Dias (2014). The 15% threshold is in line with Dias (2014). However, since the sample size itself is not provided by Dias (2014), the absolute value of upper order statistics k used in this paper might differ from those in Dias (2014). Those two moment estimators are intrinsic characteristics of assets (as well as portfolios when they are treated as single assets and their comprising assets' properties are ignored). They were first proposed by Dekkers et al. (1989) and widely used in ensuing literature (e.g. de Haan et al. (1993)) to estimate shape parameters, which will be discussed below.

2.3.2 The shape parameter

With the two moment parameters in hand, we are ready to calculate the shape parameter for each asset and portfolio, which according to Dekkers et al. (1989) is given by:

$$\hat{\gamma}_i := M_i + 1 - \frac{1}{2 \times (1 - M_i^2/N_i)},\ i \in \{1, 2, \ldots, 100\} \cup \{\mathbf{10}, \mathbf{50}, \mathbf{100}\}. \tag{5}$$

The shape parameters for our sample assets and portfolios are presented alongside other fundamental statistics in the Appendix and Table 1, respectively. In accordance with Dias (2014), the shape parameter is calculated based on losses $\boldsymbol{L}_i$, while other statistics are calculated based on returns $\boldsymbol{R}_i$.

Table 1 shows great similarities between data from the replicated study and the original. In both studies, firms with smaller capitalization tend to have higher returns on average. The price for that is higher volatility, represented by standard deviation. The overall negative skewness evidences that there are more often extreme losses than extreme positive returns. No substantial relationship between shape index and capitalization can be observed, which indicates rather coherent tail risk among portfolios with varying levels of market capitalization.

2.3.3 The scale parameter

From Table 1 and the Appendix we see that the shape parameters of our sample data are all positive. When $\gamma > 0$, the distribution is termed "heavy-tailed" in Dias (2014). In this case, the scale parameter of an asset is estimated as:

$$\hat{a}_i := L_{i,549}\sqrt{3M_i^2 - N_i},\ i = 1, 2, \ldots, 100. \tag{6}$$

Table 1: Descriptive statistics of portfolio (loss) returns

	Replicated				Original from Dias (2014)			
	Small Cap	Mid Cap	Large Cap	Portfolio	Small Cap	Mid Cap	Large Cap	Portfolio
10-asset								
Mean	0.0240	0.0164	0.0063	0.0147	0.0234	0.0152	0.0046	0.0134
Std. dev.	2.0943	1.9058	1.6340	1.5985	2.0948	1.9030	1.6337	1.5963
Skewness	−0.1852	−0.7714	−0.3243	−0.4481	−0.1822	−0.7643	−0.3239	−0.4488
Kurtosis	5.7730	13.5858	7.7763	7.6819	5.7359	13.5499	7.7370	7.6811
Shape	0.1366	0.3847	0.2847	0.2725	0.1679	0.2760	0.2379	0.2632
Min	−18.918	−23.136	−12.282	−13.317	−18.908	−23.109	−12.270	−13.320
Max	13.026	16.686	11.721	10.100	13.012	16.694	11.765	10.120
50-asset								
Mean	0.0253	0.0324	0.0294	0.0290	0.0257	0.0321	0.0301	0.0293
Std. dev.	1.6520	1.3260	1.3753	1.3727	1.6491	1.3239	1.3730	1.3699
Skewness	−0.2060	−0.3027	−0.3246	−0.3432	−0.2074	−0.3017	−0.3247	−0.3443
Kurtosis	4.9610	7.5096	9.8157	7.8821	4.9747	7.5197	9.8304	7.9114
Shape	0.2405	0.2741	0.2217	0.2421	0.1687	0.2296	0.2250	0.2114
Min	−12.800	−11.050	−12.247	−11.601	−12.810	−11.048	−12.247	−11.603
Max	9.010	10.363	13.211	10.871	9.001	10.362	13.209	10.867
100-asset								
Mean	0.0222	0.0217	0.0206	0.0215	0.0227	0.0214	0.0217	0.0219
Std. dev.	1.6470	1.4150	1.2923	1.3980	1.6435	1.4123	1.2798	1.3925
Skewness	−0.2934	−0.4649	−0.3020	−0.4235	−0.2951	−0.4637	−0.2803	−0.4200
Kurtosis	5.6043	8.1197	9.0355	7.6181	5.6308	8.1416	9.0186	7.6130
Shape	0.2320	0.2696	0.2475	0.2587	0.1959	0.2455	0.2509	0.2473
Min	−13.790	−12.353	−11.007	−11.846	−13.802	−12.359	−10.722	−11.854
Max	9.232	10.665	11.511	10.053	9.242	10.664	11.591	10.061

2.3.4 The location parameter

Following Dias (2014), the location parameter of firm i is the $(k+1)$th upper order statistic of $\boldsymbol{L}_i$, i.e.:

$$\hat{b}_i := L_{i,549},\ i = 1, 2, \ldots, 100. \tag{7}$$

2.3.5 The scaling constant

For each portfolio, we use vectors $\hat{\boldsymbol{\gamma}}_i$, $\hat{\boldsymbol{a}}_i$, $\hat{\boldsymbol{b}}_i$ to denote vectors of its component assets' estimated shape, scale and location parameters. The scaling parameter $c_i(l)$ of a portfolio is defined as the root of the equation:

$$\overline{\hat{\boldsymbol{a}}_i \frac{[c_l(l)\mathbf{1}]^{\hat{\boldsymbol{\gamma}}_i}}{\hat{\boldsymbol{\gamma}}_i} + \hat{\boldsymbol{b}}_i} = l,\ i = \mathbf{10}, \mathbf{50}, \mathbf{100}. \tag{8}$$

The right-hand side of the equation is calculated componentwise. Again, in the case of 100-asset portfolio:

$$\hat{\boldsymbol{\gamma}}_{\mathbf{100}} = \{\hat{\gamma}_1, \hat{\gamma}_2, ..., \hat{\gamma}_{100}\},\ \hat{\boldsymbol{a}}_{\mathbf{100}} = \{\hat{a}_1, \hat{a}_2, ..., \hat{a}_{100}\},\ \hat{\boldsymbol{b}}_{\mathbf{100}} = \{\hat{b}_1, \hat{b}_2, ..., \hat{b}_{100}\}, \frac{\sum_{i=1}^{100} \hat{a}_i \frac{c_{\mathbf{100}}(l)^{\hat{\gamma}_i}}{\hat{\gamma}_i} + \hat{b}_i}{100} = l.$$

3 An algorithmic improvement in transformation of losses

While we seek to stay as loyal as possible to the approach as described in Dias (2014), we were unable to complete the replication study without modifying, albeit slightly, one of the calculation steps. This section explains the necessity of the modification and how it is performed.

The adjustment concerns the transformation of losses. The function $I_l(\tilde{L}_{i,t})$ used in our replication exercise is essentially the adapted form of the indicator function $\mathbb{I}(\hat{\boldsymbol{L}}_{i,t} \in \hat{A}_i(l))$ in Dias (2014), where

$$\hat{\boldsymbol{L}}_{i,t} := \left(\mathbf{1} + \hat{\boldsymbol{\gamma}}_i \frac{\boldsymbol{L}_{i,t} - \hat{\boldsymbol{b}}_i}{\hat{\boldsymbol{a}}_i}\right)^{\frac{1}{\hat{\boldsymbol{\gamma}}_i}},$$

$$\hat{A}_i(l) := \frac{1}{c_i(l)}\left(\mathbf{1} + \hat{\boldsymbol{\gamma}}_i \frac{C_i(l) - \hat{\boldsymbol{b}}_i}{\hat{\boldsymbol{a}}_i}\right)^{\frac{1}{\hat{\boldsymbol{\gamma}}_i}},\ \hat{A}_i(l) \subseteq [0, \infty]^i \backslash \mathbf{0},$$

$$C_i(l) := \{\boldsymbol{L} \in \mathbb{R}^i | \overline{\boldsymbol{L}} \geq l\}\ i = \mathbf{10}, \mathbf{50}, \mathbf{100},\ t = 1, 2, ..., 3653. \tag{9}$$

In practice, however, the calculation can be problematic. For any $L_{i',t} \in \boldsymbol{L}_{i,t}$, as long as $1 + \hat{\gamma}_{i',t}\frac{L_{i',t} - \hat{b}_{i',t}}{\hat{a}_{i',t}} < 0$, then $\hat{\boldsymbol{L}}_{i,t} \notin \mathbb{R}^i$ and it can instantly be concluded that $\hat{\boldsymbol{L}}_{i,t} \notin \hat{A}_i(l)$. This does not, however, make intuitive sense. As is often the case, at time t the portfolio i as a whole can have a very large loss (enough to be considered in the tail distribution) but one particular asset i' within is not doing that badly and has a loss $L_{i',t}$ such that $1 + \hat{\gamma}_{i',t}\frac{L_{i',t} - \hat{b}_{i',t}}{\hat{a}_{i',t}} < 0$, and consequently the observation $\boldsymbol{L}_{i,t}$ would not be counted in the model. In other words, $\mathbb{I}(\hat{\boldsymbol{L}}_{i,t} \in \hat{A}_i(l))$ would return zero. To circumvent this problem, we simply take both $\hat{\boldsymbol{L}}_{(}i,t)$ and $\hat{A}_i(l)$ to the power of $\hat{\boldsymbol{\gamma}}_{i,t}$

(componentwise), i.e. we employ a new indicator $\mathbb{I}(\hat{\boldsymbol{L}}_{i,t}^{\hat{\gamma}_{i,t}} \in \hat{A}_i(l)^{\hat{\gamma}_{i,t}})$. This can be further rearranged through the steps below:

$$\hat{\boldsymbol{L}}_{i,t}^{\hat{\gamma}_{i,t}} \in \hat{A}_i(l)^{\hat{\gamma}_{i,t}} \Leftrightarrow \left(\mathbf{1} + \hat{\gamma}_i \frac{\boldsymbol{L}_{i,t} - \hat{\boldsymbol{b}}_i}{\hat{\boldsymbol{a}}_i}\right) \in \frac{1}{c_i(l)^{\hat{\gamma}_{i,t}}} \left(\mathbf{1} + \hat{\gamma}_i \frac{C_i(l) - \hat{\boldsymbol{b}}_i}{\hat{\boldsymbol{a}}_i}\right)$$
$$\Leftrightarrow \hat{\boldsymbol{a}}_i \frac{\left[c_i(l)^{\hat{\gamma}_i}\left(\mathbf{1} + \hat{\gamma}_i \frac{\boldsymbol{L}_{i,t} - \hat{\boldsymbol{b}}_i}{\hat{\boldsymbol{a}}_i}\right) - \mathbf{1}\right]}{\hat{\gamma}_i} + \hat{\boldsymbol{b}}_i \in C_i(l),$$

with the knowledge of Equation (9), this rearrangement leads us to:

$$\mathbb{I}(\hat{\boldsymbol{L}}_{i,t}^{\hat{\gamma}_{i,t}} \in \hat{A}_i(l)^{\hat{\gamma}_{i,t}}) \Leftrightarrow I_l(\tilde{L}_{i,t}),$$

where in the equivalent function $I_l(\tilde{L}_{i,t}), \tilde{L}_{i,t}$ as defined in Section 2.1 can be calculated as:

$$\tilde{L}_{i,t} = \overline{\hat{\boldsymbol{a}}_i \frac{\left[c_i(l)^{\hat{\gamma}_i}\left(\mathbf{1} + \hat{\gamma}_i \frac{L_{i,t} - \hat{\boldsymbol{b}}_i}{\hat{\boldsymbol{a}}_i}\right) - \mathbf{1}\right]}{\hat{\gamma}_i}} + \hat{\boldsymbol{b}}_i,\ i = \mathbf{10}, \mathbf{50}, \mathbf{100},\ t = 1, 2, \ldots, 3653.$$

The rearrangement does not alter the core idea of the model. It is meant to simplify the implementation of the model because it involves fewer parameters.

4 *Model evaluation*

In accordance with Dias (2014), the Student-t and a mixture of four normal distributions are used as benchmark models for the semiparametric approximation in this paper. The comparison between the three models will be conducted in this section.

4.1 *Survival probability plot*

To plot the semiparametric approximation, we need an array of $\{l\}$ (abscissa) and a corresponding array of $\{\widehat{S_i(l)}\}$ (ordinate) for which $\{c_i(l)\}$ is needed. To achieve this, we can either first determine an array of $\{l\}$ and then find out the corresponding array of $\{c_i(l)\}$ by solving Equation (8); or we can first write down an array of $\{c_i(l)\}$ and then get the $\{l\}$ by just calculating Equation (8) from left to right. In the end we always get a one-to-one mapping between $\{l\}$ and $\{\widehat{S_i(l)}\}$. As the second alternative is easier to execute, we first determine:

$$c_{\mathbf{10}}(l) \in [0.83, 108.43],\ c_{\mathbf{50}}(l) \in [0.70, 94.80],\ c_{\mathbf{100}}(l) \in [0.71, 89.77],$$

where the corresponding l of $\sup\{c_i(l)\}$ and $\inf\{c_i(l)\}$ roughly equal $L_{i,(1)}$ and $L_{i,(366)}$ (about 90%-quantile of $\{L_{i,1}, L_{i,2}, ..., L_{i,3653}\}$).

The plot of semiparametric approximation alongside the empirical data plot and the fitting of benchmark distributions are conducted using MATLAB. Figures 1-3 present replicated graphs on the left side, and the original ones on the right side. The overall resemblance of the two sets of graphs confirms the practicability of the modified approach described in Section 2.3.

The deviations between replicated graphs and original graphs can be explained by the differing ways of manipulating the raw data. Comparing the scatter plot of empirical data (Portfolio returns) between replicated and original graphs, we see that the ranges of the top 10% losses are identical, i.e. they are both $[1.63, 13.32]$, $[1.42, 11.6]$, $[1.44, 11.85]$ respectively for 10-asset, 50-asset and 10-asset portfolio. The survival probabilities, however, are quite different, with $S_i(\sup\{L_{i,n}\}) \approx 0.0005$ in the original graphs, and $S_i(\sup\{L_{i,n}\}) \approx 0.00027$ in the replicated ones. While $S_i(L_{i,n})$ is empirically just the inverse percentile rank of $L_{i,n}$ as in $\{L_{i,n}\}$, $S_i(\sup\{L_{i,n}\})$ can be calculated as $\frac{1}{n+1}$. As $n = 3653$ in our sample, $S_i(\sup\{L_{i,n}\})$ is subsequently equal to $\frac{1}{3653+1}$. Likewise, we can deduce that the sample size n in Dias (2014) is around 2000 ($\frac{1}{2000+1} \approx 0.0005$). Seemingly, the raw data have been filtered intensively by Dias (2014). As no filtering criteria are mentioned in Dias (2014), we do not eliminate any data points in our replication study[3]. As later shown in this paper, this unknown filtering does not appear to yield biased results, and deviation in input data does not cause material discrepancies between original and replicated results.

Despite differences in the raw data set, both the replicated and original graphs confirm that the performance of semiparametric approximation does not vary much across portfolios. That is to say, the semiparametric approach provides as good an approximation for the 10-asset portfolio as for the 50- and 100-asset portfolios. Compared with Student-t and mixture normal distributions, semiparametric approximation tends to overestimate losses more, which is consistent in both replicated and original graphs. In the original graphs, the Student-t curves are almost entirely below the empirical data scatter except for few observations. In the replicated graphs, the Student-t curves are “catching up” at the very end of tails, but are in general underestimating. Mixture normal approximation, on the contrary, provides a fairly good estimation at the start of tail distributions, but plummets at the end. This is the case in both replicated and original graphs.

In risk management, where conservativeness is encouraged, semiparametric estimation is more suitable than the benchmark models due to its slight overestimation of extreme value probability.

[3]From the sample size it can be noted that the weekends are already eliminated from Datastream by default. Some holidays are however not excluded from Datastream. Stock prices on those days are not filtered out because it is not described in Dias (2014), and holidays alone cannot explain the deviation either way as they only amount to around 200 days.

Figure 1: Loss survival probability plot of 10-asset portfolio

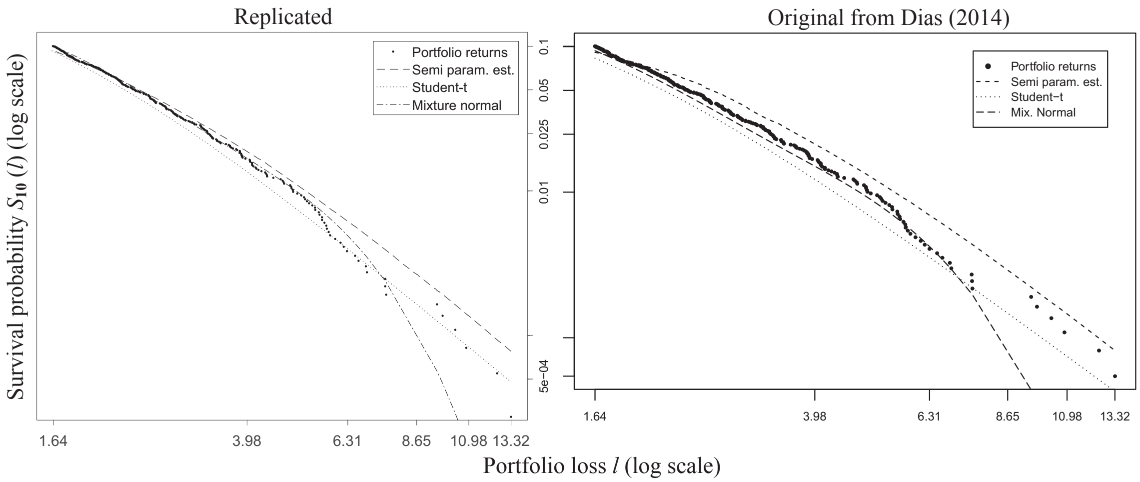


Figure 2: Loss survival probability plot of 50-asset portfolio

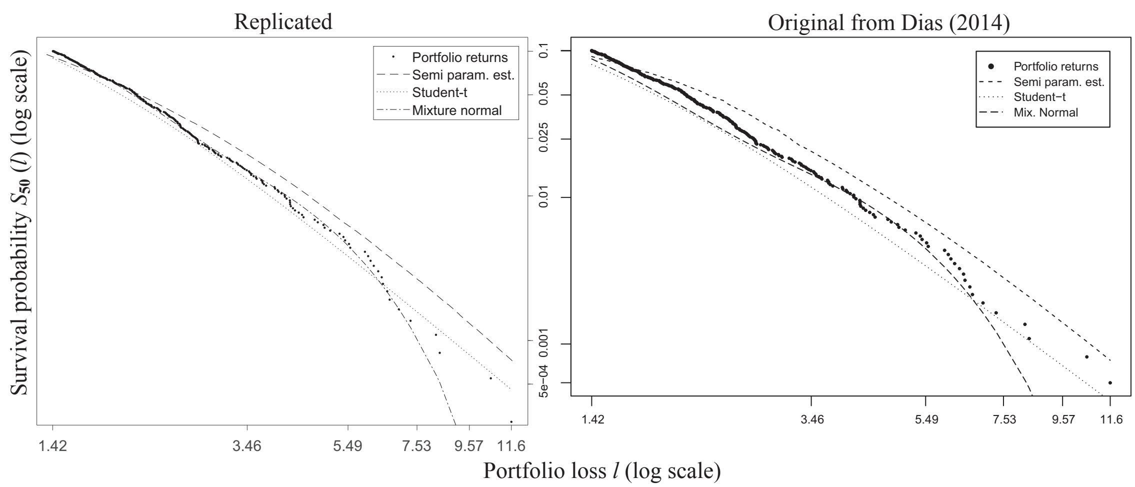
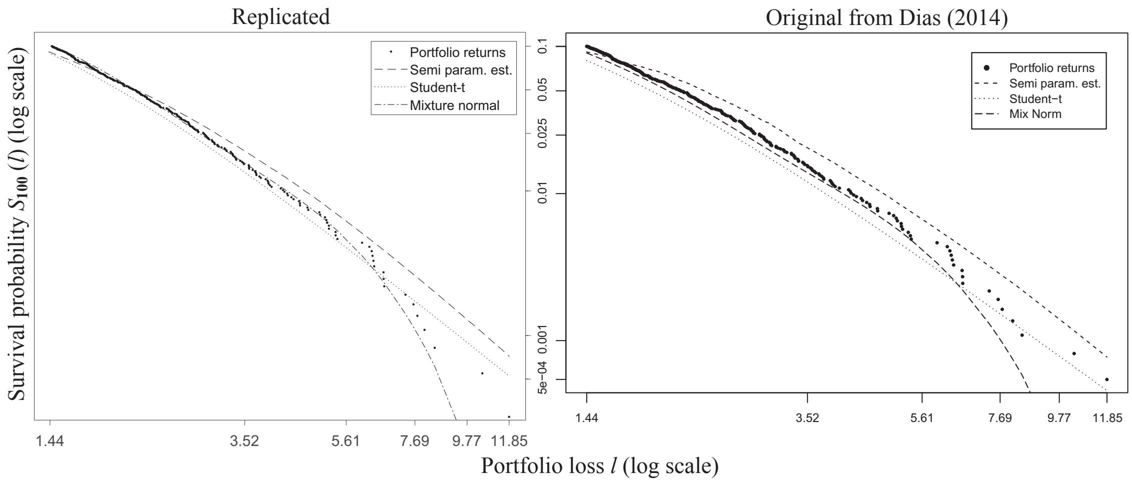


Figure 3: Loss survival probability plot of 100-asset portfolio



4.2 Value at risk estimation

In this section, value at risk of estimated semiparametric, Student-t and mixture normal distributions will be presented at four confidence levels $\alpha = 99\%, 97.5\%, 95\%, 90\%$. As defined in Artzner et al. (1999), a VaR of losses at confidence level α is the minimum value such that the probability of losses exceeding this value is no greater than $(1 - \alpha)$, i.e.:

$\mathrm{VaR}^{\alpha} = \inf\{l \in \mathbb{R} : P(L > l) \leq 1 - \alpha\}$, $\alpha = 99\%, 97.5\%, 95\%, 90\%$,

which slightly deviate from the function provided in Dias (2014) because VaR^{α}, or $\mathrm{VaR}^{\alpha}_{t}$ as denoted in Dias (2014), is reevaluated at each period t based on the information available at t. In our study however, we assume the estimated distributions as ex ante, hence only have one same VaR^{α} for one distribution throughout the sample period.

VaR^{α} can also be interpreted as the $(1 - \alpha)$ quantile of a loss distribution. That is to say, VaR^{α} is the solution to equation:

$$S(\widehat{\mathrm{VaR}^{\alpha}}) = 1 - \alpha, \alpha = 99\%, 97.5\%, 95\%, 90\%, \tag{10}$$

The percentage of violation of VaR^{α} is calculated as:

$$\%\mathrm{Viol.}^{\alpha} = 100 \frac{\sum_{t=1}^{3653} I_{\mathrm{VaR}^{\alpha}}(L_t)}{3653}, \alpha = 99\%, 97.5\%, 95\%, 90\%, \tag{11}$$

where function I is defined in Equation (1). We call a VaR^{α} underestimated when $\%\mathrm{Viol.}^{\alpha} > 100(1 - \alpha)$ and overestimated when $\%\mathrm{Viol.}^{\alpha} < 100(1 - \alpha)$. The result is presented on the left-hand side of Table 2, in comparison with the original data from Dias (2014) on the right-hand side.

In terms of violations, semiparametric estimation only underestimates at the beginning of the tail when $\alpha = 90\%$. This is the case in both the replicated table and the original table. Nevertheless, the model only underestimates once in Dias (2014) when it comes to the 10-asset portfolio, while in the replicated table, we note underestimation for all portfolios at the 90% level. In contrast, mixture normal in the replicated study underestimates 1, 2, 1, 0 time(s) respectively when $\alpha = 99\%, 97.5\%, 95\%, 90\%$, among which are 2 times for the 10-asset portfolio, 1 for the 50-asset portfolio, and 1 for the 100-asset portfolio. In the original study from Dias (2014), mixture normal underestimates 9 out of 12 cases. As for Student-t, if only those selected key values are considered, it constantly underestimates, which is the case in Dias (2014) as well.

Furthermore, to measure and compare the accuracy of VaR predictions produced by semiparametric estimation, mixture normal and Student-t, we apply three likelihood ratio (LR) tests: unconditional coverage test ($\mathrm{LR_{uc}}$), independence test ($\mathrm{LR_{ind}}$) and conditional coverage test ($\mathrm{LR_{cc}}$) – which is the combination of the former two (Christoffersen (1998)). $\mathrm{LR_{uc}}$ tests the null hypothesis of $\%\mathrm{Viol.}^{\alpha} = 100(1-\alpha)$ against the alternative $\%\mathrm{Viol.}^{\alpha} \neq 100(1-\alpha)$. $\mathrm{LR_{ind}}$ tests the null hypothesis that elements in $\{I_{\mathrm{VaR}}(L_t)\}$ are independently distributed.

Christoffersen (1998) is referred to for the proofs of the validity of these tests. In our study, we use the following equations to compute the values of $\mathrm{LR}^{\alpha}_{\mathrm{uc}}, \mathrm{LR}^{\alpha}_{\mathrm{ind}}, \mathrm{LR}^{\alpha}_{\mathrm{cc}}$ and their respective p-values:

$$\mathrm{LR}^{\alpha}_{\mathrm{uc}} = -2\ln\left\{\left[\frac{3652(1-\alpha)}{n^{\alpha}_{0}}\right]^{n^{\alpha}_{0}}\left(\frac{3652\alpha}{n^{\alpha}_{1}}\right)^{n^{\alpha}_{1}}\right\} \sim \chi^2(1), \tag{12}$$

$$\mathrm{LR}^{\alpha}_{\mathrm{ind}} = -2\ln\left\{\left[\frac{n^{\alpha}_{0}}{3652(1-\pi^{\alpha}_{01})}\right]^{n^{\alpha}_{00}}\left[\frac{n^{\alpha}_{0}}{3652(1-\pi^{\alpha}_{11})}\right]^{n^{\alpha}_{10}}\left(\frac{n^{\alpha}_{1}}{3652\pi^{\alpha}_{01}}\right)^{n^{\alpha}_{01}}\left(\frac{n^{\alpha}_{1}}{3652\pi^{\alpha}_{11}}\right)^{n^{\alpha}_{11}}\right\} \sim \chi^2(1), \tag{13}$$

$$\mathrm{LR}^{\alpha}_{\mathrm{cc}} = -2\ln\left[\left(\frac{1-\alpha}{1-\pi^{\alpha}_{01}}\right)^{n^{\alpha}_{00}}\left(\frac{1-\alpha}{1-\pi^{\alpha}_{11}}\right)^{n^{\alpha}_{10}}\left(\frac{\alpha}{\pi^{\alpha}_{01}}\right)^{n^{\alpha}_{01}}\left(\frac{\alpha}{\pi^{\alpha}_{11}}\right)^{n^{\alpha}_{11}}\right] \sim \chi^2(2), \tag{14}$$

where

$$n^{\alpha}_{x} = \sum\nolimits_{t=2}^{3653}\left[1 - x - (-1)^x I_{\mathrm{VaR}^{\alpha}}(L_t)\right],\ x = 0,1,$$

$$n^{\alpha}_{xy} = \sum\nolimits_{t=2}^{3653}\left\{\left[1 - x - (-1)^x I_{\mathrm{VaR}^{\alpha}}(L_{t-1})\right]\left[1 - y - (-1)^y I_{\mathrm{VaR}^{\alpha}}(L_t)\right]\right\},\ x = 0,1,\ y = 0,1,$$

$$n^{\alpha}_{xy} = \frac{n^{\alpha}_{xy}}{n^{\alpha}_{xx} + n^{\alpha}_{xy}},\ x = 0,1,\ y = 0,1.$$

The p-values are presented along with $\widehat{\mathrm{VaR}^{\alpha}}$ and $\%\mathrm{Viol.}^{\alpha}$ in Table 2. Overall, the test of unconditional coverage ($\mathrm{LR_{uc}}$) yields the highest p-value, followed by the test of conditional coverage ($\mathrm{LR_{cc}}$) and lastly the test of independence ($\mathrm{LR_{ind}}$). In the $\mathrm{LR_{uc}}$ test results of the replicated study, the number of p-values below 0.1 is 6 with semiparametric estimation, 0 with mixture normal and 5 with Student-t. In the $\mathrm{LR_{ind}}$ and $\mathrm{LR_{cc}}$ tests, p-values lie generally at the 0.1 level, apart from the three exceptions of $\mathrm{LR}^{99\%}_{\mathrm{cc}}$ for the mixture normal estimation. It can also be observed that p-values tend to decrease along with α. One interpretation is that high-percentile VaR estimates are more prone to be affected by clustering than low-percentile ones (Dias (2014)). With the ability to retain low p-values even at high percentile, semiparametric estimation is therefore more valuable when the focus is on extreme losses.

Table 2: Value at risk estimates and p-values of LR tests

	Replicated			Original from Dias (2014)		
	10-asset	50-asset	100-asset	10-asset	50-asset	100-asset
Semi param. est.						
$\widehat{\text{VaR}}^{99\%}$	5.201	4.590	4.637	4.961	4.432	4.514
% Viol.$^{99\%}$	0.849	0.684	0.766	0.952	0.735	0.816
$\text{LR}_{\text{uc}}^{99\%}$	0.346	0.042	0.140	0.772	0.090	0.249
$\text{LR}_{\text{ind}}^{99\%}$	0.028	0.011	0.018	0.045	0.015	0.023
$\text{LR}_{\text{cc}}^{99\%}$	0.057	0.005	0.020	0.128	0.012	0.039
Mix. normal						
$\widehat{\text{VaR}}^{99\%}$	4.893	4.122	4.288	4.857	4.314	4.247
% Viol.$^{99\%}$	1.013	0.958	0.931	1.061	0.762	0.952
$\text{LR}_{\text{uc}}^{99\%}$	0.937	0.799	0.672	0.709	0.130	0.772
$\text{LR}_{\text{ind}}^{99\%}$	0.058	0.046	0.041	0.070	0.017	0.045
$\text{LR}_{\text{cc}}^{99\%}$	0.165	0.133	0.113	0.180	0.019	0.128
Student-t						
$\widehat{\text{VaR}}^{99\%}$	4.476	3.838	3.964	4.468	3.825	3.938
% Viol.$^{99\%}$	1.232	1.150	1.068	1.225	1.143	1.089
$\text{LR}_{\text{uc}}^{99\%}$	0.174	0.373	0.683	0.185	0.392	0.592
$\text{LR}_{\text{ind}}^{99\%}$	0.019	0.013	0.008	0.018	0.012	0.009
$\text{LR}_{\text{cc}}^{99\%}$	0.025	0.030	0.027	0.025	0.029	0.028
Semi param. est.						
$\widehat{\text{VaR}}^{97.5\%}$	3.477	3.064	3.089	3.352	2.987	3.042
% Viol.$^{97.5\%}$	2.190	1.861	2.080	2.423	1.960	2.096
$\text{LR}_{\text{uc}}^{97.5\%}$	0.182	0.010	0.095	0.764	0.029	0.107
$\text{LR}_{\text{ind}}^{97.5\%}$	0.009	0.002	0.001	0.025	0.002	0.001
$\text{LR}_{\text{cc}}^{97.5\%}$	0.013	0.000	0.001	0.078	0.001	0.001
Mix. normal						
$\widehat{\text{VaR}}^{97.5\%}$	3.313	2.761	2.868	3.219	2.719	2.849
% Viol.$^{97.5\%}$	2.573	2.354	2.518	2.777	2.504	2.514
$\text{LR}_{\text{uc}}^{97.5\%}$	0.857	0.571	0.941	0.290	0.985	0.984
$\text{LR}_{\text{ind}}^{97.5\%}$	0.012	0.005	0.002	0.008	0.002	0.002
$\text{LR}_{\text{cc}}^{97.5\%}$	0.042	0.015	0.010	0.018	0.009	0.009
Student-t						
$\widehat{\text{VaR}}^{97.5\%}$	3.104	2.657	2.718	3.100	2.649	2.705
% Viol.$^{97.5\%}$	2.929	2.710	2.956	2.913	2.722	2.994
$\text{LR}_{\text{uc}}^{97.5\%}$	0.129	0.421	0.085	0.117	0.394	0.062
$\text{LR}_{\text{ind}}^{97.5\%}$	0.005	0.002	0.006	0.004	0.001	0.006
$\text{LR}_{\text{cc}}^{97.5\%}$	0.006	0.005	0.005	0.005	0.005	0.004

Table 2 continued on next page

Table 2 continued: Value at risk estimates and p-values of LR tests

	Replicated			Original from Dias (2014)		
	10-asset	50-asset	100-asset	10-asset	50-asset	100-asset
Semi param. est.						
$\widehat{\text{VaR}}^{95\%}$	2.433	2.114	2.135	2.379	2.198	2.211
% Viol.$^{95\%}$	4.845	4.818	4.955	4.982	4.301	4.492
$\text{LR}_{\text{uc}}^{95\%}$	0.614	0.614	0.843	0.960	0.046	0.151
$\text{LR}_{\text{ind}}^{95\%}$	0.015	0.000	0.000	0.026	0.000	0.000
$\text{LR}_{\text{cc}}^{95\%}$	0.047	0.000	0.001	0.086	0.000	0.001
Mix. normal						
$\widehat{\text{VaR}}^{95\%}$	2.420	2.056	2.121	2.345	1.977	2.101
% Viol.$^{95\%}$	4.845	5.201	4.982	5.145	5.581	4.982
$\text{LR}_{\text{uc}}^{95\%}$	0.614	0.577	0.903	0.686	0.112	0.960
$\text{LR}_{\text{ind}}^{95\%}$	0.015	0.000	0.000	0.047	0.000	0.000
$\text{LR}_{\text{cc}}^{95\%}$	0.047	0.000	0.001	0.128	0.000	0.001
Student-t						
$\widehat{\text{VaR}}^{95\%}$	2.272	1.942	1.973	2.270	1.936	1.966
% Viol.$^{95\%}$	5.393	5.831	5.612	5.417	5.799	5.690
$\text{LR}_{\text{uc}}^{95\%}$	0.314	0.024	0.110	0.251	0.030	0.060
$\text{LR}_{\text{ind}}^{95\%}$	0.053	0.001	0.000	0.058	0.000	0.000
$\text{LR}_{\text{cc}}^{95\%}$	0.053	0.001	0.000	0.058	0.000	0.000
Semi param. est.						
$\widehat{\text{VaR}}^{90\%}$	1.564	1.326	1.337	1.446	1.431	1.464
% Viol.$^{90\%}$	10.977	11.114	11.388	12.169	9.855	9.719
$\text{LR}_{\text{uc}}^{90\%}$	0.058	0.031	0.006	0.000	0.770	0.569
$\text{LR}_{\text{ind}}^{90\%}$	0.000	0.000	0.000	0.000	0.000	0.000
$\text{LR}_{\text{cc}}^{90\%}$	0.000	0.000	0.000	0.000	0.001	0.001
Mix. normal						
$\widehat{\text{VaR}}^{90\%}$	1.642	1.438	1.467	1.626	1.397	1.440
% Viol.$^{90\%}$	9.992	9.828	9.800	10.127	10.400	10.046
$\text{LR}_{\text{uc}}^{90\%}$	0.947	0.690	0.650	0.796	0.421	0.925
$\text{LR}_{\text{ind}}^{90\%}$	0.000	0.000	0.000	0.000	0.000	0.000
$\text{LR}_{\text{cc}}^{90\%}$	0.000	0.001	0.001	0.000	0.001	0.000
Student-t						
$\widehat{\text{VaR}}^{90\%}$	1.559	1.328	1.342	1.558	1.324	1.338
% Viol.$^{90\%}$	11.087	11.114	11.278	11.135	11.053	11.216
$\text{LR}_{\text{uc}}^{90\%}$	0.035	0.031	0.013	0.024	0.036	0.015
$\text{LR}_{\text{ind}}^{90\%}$	0.000	0.000	0.000	0.000	0.000	0.000
$\text{LR}_{\text{cc}}^{90\%}$	0.000	0.000	0.000	0.000	0.000	0.000

5 Criticism and conclusion

This paper reviews the estimation method to approximate multi-asset tail distribution provided by Dias (2014), namely semiparametric estimation. The paper attempts to rebuild the model following instructions in Dias (2014), and to generate comparable results by using the same sample as described in Dias (2014). The study reveals an instance of infeasibility in this model where the indicator function $\mathbb{I}(\hat{\boldsymbol{L}}_{i,t} \in \hat{A}_i(l))$ is estimated. Nevertheless, this can be readily sidestepped through a small transformation provided in the paper, without altering the core idea of the model. Due to lack of description on data filtering, the results in this replicated study slightly deviate from the ones in Dias (2014). Despite that, they are sufficiently close to confirm the superiority of semiparametric estimation over classical methods such as mixture normal and Student-t approximations in estimating tail distribution of portfolios. This is due to its uniqueness in combining the strengths of both EVT models and other multivariate models. As it only focuses on approximating the tail part of a distribution, and is comprehensive in terms of keeping the traits of individual portfolio components, the model performs very well.

The accuracy of the model is proven in both this paper and Dias (2014). However, the results should be taken with a pinch of salt. The ultimate goal of distribution estimation is to guide investors in portfolio selection. The performance of the estimation cannot be determined until the future performance of the corresponding portfolio is observed. In this sense, it is at best indicative of how well various models fit historical data when only in-sample estimation is performed. To attain a more convincing result, empirically estimated models should be compared also with regard to out-of-sample prediction.

The principal assumption underpinning the semiparametric estimation model is that the portfolio losses over the sample period of 14 years follow one single distribution and that the parameters remain the same. As for individual firms, intrinsic values such as γ, a, and b are highly likely to change over time, especially after company reforms. As a result, it is advisable to re-estimate those variables on a regular basis. As for portfolios, their losses probably follow a different distribution in economic boom years than in doom years. Within our sample period, stock performance between 2007 and 2009, during the time of the global financial crisis, is not representative. Therefore, we would recommend estimating the distribution on a more regular basis. This also applies to other estimation models.

In summary, Dias (2014) provides a fresh perspective on distribution estimation for large portfolio losses by innovatively developing the multivariate semiparametric estimation model. The model performs well in approximating historical data. There is, however, room for development, and it is left to be seen if the model is able to perform as well in forecasting.

References

Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber and David Heath (2001). "Coherent Measures of Risk." *Mathematical Finance* 9(3): 203–228. DOI: 10.1111/1467-9965.00068.

Christoffersen, Peter F. (1998). "Evaluating Interval Forecasts." *International Economic Review* 39(4): 841–862. DOI: 10.2307/2527341.

De Haan, Laurens and Sidney I. Resnick (1993). "Estimating the limit distribution of multivariate extremes." *Stochastic Models* 9(2): 275–309. DOI: 10.1080/15326349308807267.

De Haan, Laurens and Sidney I. Resnick (1977). "Limit theory for multivariate sample extremes." *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 40(4): 317–337. DOI: 10.1007/BF00533086.

Dekkers, Arnold L. M., John H. J. Einmahl and Laurens De Haan (1989). "A Moment Estimator for the Index of an Extreme-Value Distribution." *The Annals of Statistics* 17(4): 1833–1855. DOI: 10.1214/aos/1176347397.

Dias, Alexandra. (2014). "Semiparametric estimation of multi-asset portfolio tail risk." *Journal of Banking & Finance* 49: 398–408. DOI: 10.1016/j.jbankfin.2014.05.033.

Haas, Markus, Stefan Mittnik and Marc S. Paolella (2004). "Mixed Normal Conditional Heteroskedasticity." *Journal of Financial Econometrics* 2(2): 211–250. DOI: 10.1093/jjfinec/nbh009.

McNeil, Alexander J. and Rüdiger Frey (2000). "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach." *Journal of empirical finance* 7(3–4): 271–300. DOI: 10.1016/S0927-5398(00)00012-8.

Newcomb, Simon (1886). "A Generalized Theory of the Combination of Observations so as to Obtain the Best Result." *American Journal of Mathematics* 8(4): 343–366. DOI: 10.2307/2369392.

Øien, Beate (2011). "*Revisiting Foreign Direct Investment and Collective Labor Rights: Replicating "the positive case" of economic globalization.*" *(Master Thesis)* Trondheim: Norwegian Institute for Science and Technology. Handle: http://hdl.handle.net/11250/268407.

Sullivan, Arthur and Steven M. Sheffrin (2011). "*Economics: Principles in action.*" New Jersey: Pearson Prentice Hall.

Zhang, Zhengjun and Kazuhiko Shinki (2007). "Extreme co-movements and extreme impacts in high frequency data in finance." *Journal of Banking & Finance* 31(5): 1399–1415. DOI: 10.1016/j.jbankfin.2006.10.019.

Appendix: Descriptive statistics of asset (loss) returns

Tables 3-5 present the descriptive statistics of asset loss returns from large cap, mid cap and small cap firms, respectively. All the firms listed are treated as components of the 100-asset portfolio. Those chosen to be included in the 10-asset portfolio and 50-asset portfolio are labeled in column "Portfolio". The left part of the tables is based on the data from this duplicated study and the right part is directly extracted from Dias (2014). Refer to Section 2.2 for detailed sample description.

Table 3: Descriptive statistics of asset (loss) returns of large cap firms

		Replicated					Original from Dias (2014)				
Firma	Portfolio	Mean	Std. dev.	Skewness	Kurtosis	Shape	Mean	Std. dev.	Skewness	Kurtosis	Shape
ABT		0.0235	1.5337	−0.4244	9.3727	0.2227	0.0227	1.5437	−0.3569	9.3655	0.2443
AXP	50-asset	0.0171	2.4436	−0.0102	8.8774	0.2432	0.0192	2.4414	−0.0129	8.8581	0.1827
ADM	50-asset	0.0376	2.0822	−0.3255	9.9945	0.3166	0.0364	2.0844	−0.3138	9.9251	0.3119
BK	10-asset	−0.0053	2.5894	−0.1020	16.5778	0.2754	−0.0052	2.5959	−0.0930	16.3652	0.2697
BA	50-asset	0.0326	1.9806	−0.2574	5.6649	0.1683	0.0333	1.9815	−0.2573	5.6232	0.0993
CNP		0.0056	2.5060	−2.1810	123.5542	0.4462	0.0034	2.5007	−2.1758	124.0428	0.4746
CLX	50-asset	0.0167	1.5353	−0.3085	10.9304	0.3490	0.0198	1.5437	−0.2944	10.6538	0.2952
CSC		−0.0144	2.4743	−2.5501	53.7502	0.3209	−0.0047	2.4781	−2.5214	53.1690	0.3578
CVS	50-asset	0.0350	1.9911	−1.2400	19.5697	0.3381	0.0361	2.0050	−1.2025	19.0172	0.3049
DUK		0.0125	1.6536	−0.2221	11.2879	0.2548	0.0127	1.6528	−0.2182	11.2702	0.2086
EMR	50-asset	0.0245	1.8694	−0.0809	6.9553	0.2353	0.0245	1.8701	−0.0747	6.9138	0.1693
FDO		0.0378	2.2461	0.2481	7.3499	0.1821	0.0356	2.2509	0.2322	7.2681	0.2442
FTR	50-asset	−0.0305	2.1890	0.0029	10.7994	0.2600	−0.0266	2.1875	0.0012	10.7645	0.2900
GT	10-asset	−0.0045	3.3071	−0.3560	4.7396	0.2856	−0.0084	3.3057	−0.3493	4.7085	0.2691
HP	50-asset	0.0645	2.7446	−0.3810	5.8945	0.1581	0.0609	2.7504	−0.3638	5.8388	0.2374
IPG		−0.0323	2.8322	−0.3286	18.9865	0.3256	−0.0274	2.8322	−0.3301	18.8821	0.3198
IP†		−0.0039	2.4304	0.0167	8.0799	0.2523	0.0482	2.1650	−0.0074	4.8295	0.3217
JCI†	50-asset	0.0462	2.1638	0.0037	4.8323	0.1768	−0.0019	2.4318	0.0119	8.0199	0.1925
KR		0.0202	1.7863	−0.4587	6.4233	0.2541	0.0173	1.8605	−1.2994	19.8008	0.3186
LOW	50-asset	0.0328	2.1870	0.3407	4.4850	0.1555	0.0359	2.1865	0.3381	4.4605	0.1468
MCD		0.0240	1.5507	−0.1966	6.2088	0.1720	0.0208	1.5504	−0.1949	6.1721	0.2416
MUR	50-asset	0.0453	2.2281	−0.2673	6.8197	0.1314	0.0456	2.2258	−0.2650	6.8103	0.1745
JWN		0.0423	2.6835	0.1579	6.6596	0.1538	0.0419	2.6821	0.1587	6.6384	0.1715
OMC	50-asset	0.0109	1.9197	−0.3596	9.9608	0.2930	0.0137	1.9246	−0.3443	9.8132	0.3079
POM	10-asset	−0.0050	1.4977	−0.3603	10.4442	0.2721	−0.0067	1.5004	−0.3624	10.3311	0.2734
PBI	50-asset	−0.0194	1.9206	−1.9023	35.9885	0.3487	−0.0191	1.9245	−1.8795	35.4871	0.3480
RHI		0.0295	2.5213	0.7568	10.4213	0.2118	0.0303	2.5179	0.7535	10.4213	0.1908
SWN	50-asset	0.1059	3.0300	−0.5270	17.4886	0.2016	0.1034	3.0349	−0.5165	17.2872	0.2813
TE		−0.0020	1.8414	−1.0441	25.7491	0.3435	−0.0034	1.8381	−1.0450	25.8555	0.3468
TMO	50-asset	0.0590	1.9364	0.3214	6.5285	0.1942	0.0578	1.9369	0.3232	6.4942	0.2352
UTX		0.0343	1.8170	−1.6140	34.3436	0.1890	0.0372	1.8215	−1.5880	33.8222	0.1930
WBA	50-asset	0.0185	1.7445	−0.0254	7.1108	0.2790	0.0183	1.7550	−0.0197	6.9326	0.1778
WEC	10-asset	0.0399	1.2003	−0.0416	4.7243	0.1545	0.0387	1.1986	−0.0403	4.7297	0.2352
NBR	50-asset	0.0026	2.9343	−0.3810	4.0355	0.2309	0.0048	2.9403	−0.3758	3.9721	0.1571

†We believe the data from firm IP (INTERNATIONAL PAPER) and JCI (JOHNSON CONTROLS) are mistakenly switched in Dias (2014). This would explain the big statistical mismatching of those two firms between the replicated and original study. It might also contribute to the difference of results between the two studies, especially when a 50-asset portfolio is involved.

Table 4: Descriptive statistics of asset (loss) returns of large cap firms

		Replicated					Original from Dias (2014)				
Firma	Portfolio	Mean	Std. dev.	Skewness	Kurtosis	Shape	Mean	Std. dev.	Skewness	Kurtosis	Shape
AMD		−0.0361	4.0313	−0.4186	7.4068	0.2277	−0.0362	4.0280	−0.4161	7.3926	0.2305
WTR	50-asset	0.0359	1.6766	0.2157	5.7452	0.1063	0.0337	1.6804	0.2085	5.6556	0.0958
ATW	10-asset	0.0468	2.7880	−0.2494	6.9537	0.2061	0.0472	2.7851	−0.2462	6.9482	0.2359
BOH	50-asset	0.0315	1.9852	−0.7720	16.7466	0.3259	0.0297	1.9856	−0.7735	16.6601	0.2547
BRE⋆		0.0247	2.1463	0.0169	16.2520	0.3815	0.0251	2.1435	0.0174	16.2649	0.3670
CBT	50-asset	0.0397	2.3889	−0.4038	9.2036	0.2660	0.0410	2.3848	−0.4048	9.2235	0.3030
CYN		0.0240	2.2005	0.1115	10.4264	0.2904	0.0214	2.2019	0.1140	10.3376	0.2806
CBSH	50-asset	0.0264	1.6639	0.1327	11.8035	0.2478	0.0236	1.6649	0.1289	11.7220	0.2503
CNW		0.0039	2.5908	−0.2350	6.0463	0.2911	0.0061	2.5978	−0.2120	6.0404	0.2856
CR	50-asset	0.0334	2.1059	−0.3387	8.5960	0.2958	0.0367	2.1086	−0.3319	8.5065	0.2958
UFS		−0.0110	2.9222	−0.1654	9.9598	0.3129	−0.0111	2.9189	−0.1485	9.8629	0.3033
ESL	50-asset	0.0596	2.6566	−0.2831	8.9103	0.2904	0.0581	2.6547	−0.2817	8.8841	0.2230
FHN	10-asset	−0.0195	2.9089	−1.0517	29.5440	0.3548	−0.0223	2.9039	−1.0492	29.5786	0.3475
GGG	50-asset	0.0546	1.9462	0.0108	5.3723	0.2032	0.0560	1.9464	0.0123	5.3409	0.0717
HE		0.0162	1.2888	−0.3031	6.1455	0.1444	0.0150	1.2879	−0.3005	6.1186	0.2488
HSH⋆	50-asset	0.0070	1.6532	−0.0875	9.8952	0.3045	0.0051	1.6558	−0.0852	9.7576	0.3222
HUBB		0.0379	1.8206	0.1341	6.5452	0.1622	0.0374	1.8192	0.1353	6.5296	0.1775
ITT	50-asset	0.0563	1.8208	0.7003	11.9236	0.1638	0.0548	1.8188	0.7006	11.9182	0.1605
LPX		0.0075	3.5775	−0.4024	7.9196	0.2653	0.0071	3.5747	−0.3982	7.8902	0.2473
MDP	50-asset	0.0059	1.8828	0.0356	8.6251	0.2824	0.0080	1.8793	0.0333	8.6521	0.2258
NYT		−0.0309	2.5209	0.1227	9.6237	0.2451	−0.0258	2.5236	0.1344	9.5628	0.2325
OMI	50-asset	0.0497	1.9661	−0.4430	9.8327	0.2658	0.0492	1.9668	−0.4405	9.7646	0.2880
PNM	10-asset	0.0219	2.0472	−1.2679	22.2234	0.3182	0.0208	2.0465	−1.2548	22.1534	0.4000
STR	50-asset	0.0617	2.0690	−0.0673	13.7760	0.3437	0.0579	2.0700	−0.0684	13.6832	0.3147
RPM		0.0385	2.0492	−0.6451	8.8384	0.2800	0.0351	2.0467	−0.6402	8.8315	0.2107
SCI	50-asset	0.0263	3.0318	0.3748	11.5870	0.2968	0.0247	3.0312	0.3761	11.5410	0.2411
SPW		0.0247	2.6742	−1.3092	13.0814	0.3852	0.0266	2.6701	−1.3070	13.1028	0.3514
TER	50-asset	−0.0362	3.4810	−0.0805	4.6232	0.1058	−0.0278	3.5010	−0.0523	4.5457	0.1505
TR		0.0110	1.5615	0.3102	10.5322	0.0901	0.0123	1.5602	0.3096	10.5072	0.1960
UGI	50-asset	0.0494	1.4292	0.1563	6.2494	0.1917	0.0496	1.4289	0.1567	6.2257	0.1412
VMI		0.0610	2.6178	0.0920	6.8225	0.1643	0.0584	2.6408	0.1560	7.1372	0.1887
WR	50-asset	0.0176	1.6471	−1.1147	20.8651	0.3000	0.0149	1.6495	−1.1005	20.6945	0.3048
TDS		−0.0231	2.1813	−0.3313	10.1382	0.2416	−0.0248	2.1820	−0.3304	10.0671	0.2214

⋆ Firms currently delisted.

Table 5: Descriptive statistics of asset (loss) returns of large cap firms

		Replicated					Original from Dias (2014)				
Firma	Portfolio	Mean	Std. dev.	Skewness	Kurtosis	Shape	Mean	Std. dev.	Skewness	Kurtosis	Shape
NPBC	50-asset	−0.0032	2.7280	0.1137	10.5479	0.2988	−0.0039	2.7234	0.1128	10.5744	0.2912
AGYS		−0.0010	3.8853	−0.4877	16.9314	0.3166	0.0027	3.8834	−0.4842	16.8752	0.3774
AWR	50-asset	0.0239	1.8850	−0.1926	3.6551	0.1864	0.0216	1.8853	−0.1808	3.6511	0.1615
AIT		0.0518	2.3825	0.1660	5.3638	0.0448	0.0512	2.3772	0.1668	5.3939	0.1368
AZZ	50-asset	0.0758	2.7150	0.0629	7.6687	0.2647	0.0823	2.7201	0.0784	7.6082	0.2340
GBB	10-asset	−0.0057	2.2296	−0.2527	4.8021	0.1735	−0.0053	2.2265	−0.2511	4.8100	0.2361
CAS	50-asset	0.0063	3.6753	0.0469	11.0617	0.1532	0.0062	3.6689	0.0473	11.0821	0.2265
CBB		−0.0640	3.7307	−0.4100	14.1021	0.3488	−0.0582	3.7306	−0.4088	14.0249	0.3715
CMTL	50-asset	0.0430	3.1864	−0.7749	23.0896	0.2615	0.0362	3.2085	−0.7485	22.4287	0.2608
CUB		0.0541	2.6077	0.2973	5.4076	0.1975	0.0556	2.6070	0.2984	5.3877	0.2032
DY	50-asset	−0.0015	3.4834	−0.8884	13.8301	0.3219	0.0017	3.4838	−0.8851	13.7465	0.3219
SOGC		−0.0345	3.1195	−0.5828	8.7520	0.2900	−0.0301	3.1399	−0.4906	8.2967	0.2744
GK	50-asset	0.0179	2.4161	−0.8862	16.1963	0.1981	0.0169	2.4119	−0.8852	16.2248	0.2477
GTY		0.0136	2.3772	−3.9126	93.2781	0.3525	0.0103	2.3746	−3.9037	93.1506	0.3846
KAMN	50-asset	0.0308	2.8392	−0.5510	6.6255	0.2152	0.0327	2.8467	−0.5481	6.5598	0.1921
LZB	10-asset	0.0167	3.5916	−0.2515	16.2657	0.3692	0.0139	3.6007	−0.2298	16.0837	0.3517
MMI	50-asset	0.0000	2.9525	−0.0787	8.7069	0.1516	0.0003	2.9534	−0.0790	8.6487	0.2121
MYE		0.0220	3.1039	−0.4471	12.0394	0.2934	0.0233	3.1000	−0.4440	12.0395	0.2777
NEWP	50-asset	0.0046	4.1130	0.0982	6.3873	0.2990	0.0176	4.1282	0.1102	6.2831	0.1781
PKE		0.0132	2.8802	−1.3504	38.6522	0.2518	0.0073	2.9142	−1.5431	39.0104	0.3074
PEI	50-asset	0.0073	3.1404	−0.2223	19.7602	0.4716	0.0030	3.1374	−0.2195	19.7457	0.4097
PNK		0.0040	3.4428	0.4641	17.4193	0.1711	0.0059	3.4385	0.4701	17.5174	0.2057
RLI	50-asset	0.0478	1.6039	0.0289	7.5292	0.2039	0.0480	1.6001	0.0287	7.5701	0.2143
RYL		0.0553	3.2041	−0.0288	2.6393	0.1089	0.0570	3.2025	−0.0268	2.6307	0.1201
SWX	50-asset	0.0243	1.5778	−0.0421	5.9674	0.2205	0.0243	1.5849	−0.0092	5.9871	0.2090
SCL		0.0472	2.1701	0.2578	6.4786	0.1984	0.0473	2.1648	0.2587	6.5180	0.1280
RGR	50-asset	0.0577	2.8972	−1.5854	28.0865	0.2813	0.0572	2.8946	−1.5787	28.0398	0.3006
TXI⋆		0.0210	2.9650	−0.1413	8.5779	0.2559	0.0236	2.9616	−0.1442	8.5723	0.3095
UIL	50-asset	0.0063	1.5311	−0.6729	9.4401	0.2297	0.0054	1.5305	−0.6714	9.4320	0.2907
UNS⋆		0.0459	1.6342	1.6929	30.1949	0.1983	0.0454	1.6387	1.6822	30.0080	0.2701
WGO	50-asset	0.0276	3.3042	−0.0669	5.9011	0.1933	0.0285	3.2986	−0.0683	5.9392	0.1831
WWW	10-asset	0.0611	2.4018	−0.2798	14.0196	0.2056	0.0617	2.4130	−0.2417	13.8351	0.2266
POWL	50-asset	0.0623	3.0627	0.1013	6.6127	0.2052	0.0601	3.1043	−0.0270	7.3452	0.2543

⋆ Firms currently delisted.