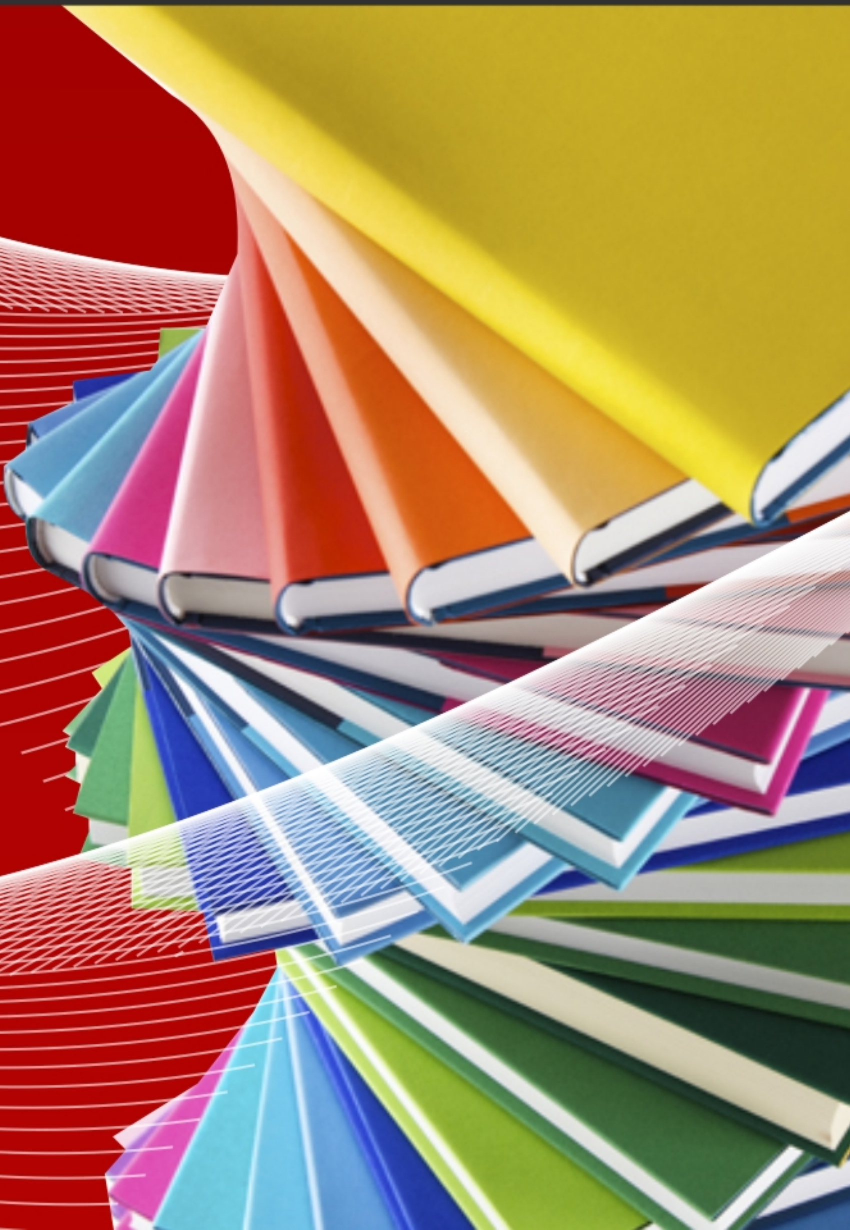
澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU

Outstanding Academic Papers by Students
學 生 優 秀 作 品



UNIVERSITY OF MACAU

FACULTY OF BUSINESS ADMINISTRATION

Essay on Tail Risk Measure with Application in Volatility Forecast

ZHAO YUPEI
M-B4-4632-9

Supervisor: Dr. KO, Iat Meng, Stanley

Thesis presented to the
Faculty of Business Administration
University of Macau
In partial fulfillment for granting the Master of Science in Finance Degree

2016

Essay on Tail Risk Measure with Application in Volatility Forecast

Abstract
In this paper, we study the tail risk measure in financial market. Recently, Kelly and Jiang (2014) propose a measure based on Hill (1975)'s method. They arbitrarily estimate the Hill tail risk by fixing the threshold at the 0.05 empirical quantile level. We show by various simulation experiments that their measure is very sensitive to the choice of thresholds. To endogenize the threshold choice, we propose a novel composite Pareto-Normal model for tail risk measure. Using the variance decomposition, our tail risk measure natually maps to the overall volatility. We show that the induced total volatility from our estimated tail risk measure matches the market volatility well, whereas that of Kelly and Jiang's Hill estimate deviates substantially from the market volatility. Finally, we investigate the predictive power of tail risk on realized volatility. The results show that our proposed method outperform the Hill estimate in volatility forecast.

Keywords: Tail risk; Hill estimator; Extreme value theory; Composite Pareto-Nomal model; Volatility forecast

1 Introduction

Tail risk, which describes the probability of extreme loss, is of interest in varieties of areas, such as finance and actuarial science. Practitioners in finance pay cautious attention to the downside tail risk for it represents the risk of large investment loss. It is well acknowledged that the heavy tails exists in financial data, see e.g. Jansen and De Vries (1991) and Mandelbrot (1997). Thus, traditional techniques that use normal or other approximations to estimate the distribution of assets' returns will underestimate the tail risk. The very early measurement of downside risk is the safety-first criterion of Roy (1952). Based on this criterion that minimizing the probability of the portfolio's return falling below a minimum desired threshold, investors can select a portfolio or asset over the others accordingly. Some scholars such as Bawa (1975) and Fishburn (1977) propose the lower partial moments, in which risk is defined as the probability weighted function of the deviations below a target return. After the financial crises in the 1990s, value-at-risk (VaR) is widely used as a risk measure. It estimates the least loss over a target period given a level of confidence. Later, Artzner (1997), Artzner, Delbaen, Eber, and Heath (1999) show that VaR has some shortcomings and propose the conditional value-at-risk (CVaR), also known as the expected shortfall (ES). The CVaR at a given level of confidence is defined as the expected loss under the condition that the loss is greater than the corresponding VaR.

It is commonly observed that the tail distribution of financial data entails power type behaviour such as the Pareto distribution. The Pareto distribution is a power law probability distribution named after Vilfredo Pareto. This distribution is originally used to describe the income distribution. Later, the Pareto distribution and its generalized form are recognized as a useful model for heavy tailed behavior. The Pareto distribution is parameterized by the shape coefficient α. The smaller the value of α, the thicker the tail captured by Pareto. Thus, the reciprocal $1/\alpha$ can be regarded as the tail risk index, i.e. the larger the value of $1/\alpha$, the higher the tail risk and vice versa.

The most popular estimator for the tail index is proposed by Hill (1975). The Hill estimator is semi-parametric and easy to calculate. Therefore, it is adopted by both practical risk management and academic studies. However, the application of the Hill estimator has its limit in financial data, because there are very few extreme returns observed for a particular firm. Kelly and Jiang (2014) (KJ hereafter) hypothesise that all firms expose to the same tail risk and use pooled firm returns to overcome the small sample size problem. They assume that each firm's tail risk consists of two parts. The first part is the firm-specific tail risk and the second part is time-varying global tail risk. Accordingly, they decompose the shape parameter α into two parts: a_i corresponds to firm-specific risk, and $1/\lambda_t$ corresponds to the dynamic common risk. Following this assumption, KJ's Hill estimator estimates a common tail risk component λ_t multiplied by the mean of $1/a_i$.

Besides the small sample problem, KJ's Hill estimator also suffer from the uncertainty of threshold choice. To implement KJ's method, we have to first specify a threshold, e.g. the 0.05 empirical quantile, then estimate the Pareto parameter using the data below it, discarding the rest data. For time varying tail risk, the threshold should also change over time according to different risk level. One method of threshold selection is to use various graphical tools, e.g. a log-log plot of the CDF (see for example Coles, Bawa, Trenner, and Dorazio (2001)). However, this may result in bias estimate due to its subjectivity. Clauset, Young, and Gleditsch (2007) propose a formal way to select the threshold. They suggest to choose the point which minimizes the Kolomogorov-Smirnov statistic. A comprehensive survey on the application of Pareto distributions can be found in Clauset, Shalizi, and Newman (2009).

Another approach is to construct a full parametric model to endogenize the threshold as one of the parameter to be estimated. De Melo Mendes and Lopes (2004) propose a simple composite distribution consists of normal and generalised Pareto distributions. However, in their paper, the threshold is not modeled as a parameter but as a by-product in fitting the best proportion of data in the tail. Behrens, Lopes, and Gamerman (2004) also consider the model with normal distribution for the non-tail part and generalised Pareto distribution for the tail part. And the threshold is treated as a parameter and estimated by Bayesian inference approach. Zhao, Scarrott, Oxley, and Reale (2010) develop a similar composite distribution and apply in the GARCH framework. They study univariate time series (e.g. S&P 100 return in their paper) and thus the small sample problem of extreme returns exists. Thus, they adopt the Bayesian method to overcome the problem. However, all the abovementioned papers do not consider the continuity and smooth conditions, so the composite density may be discontinuous at the threshold. Carreau and Bengio (2009) develop the hybrid Pareto distribution model composed of normal and generalised Pareto distribution with constraints of continuity and differentiability. And they apply their model to insurance data.

In this paper, we adopt the second approach and propose a composite Pareto-Normal model for tail risk. We also follow KJ's dynamic common risk setup with pooled cross-section data. Therefore, we avoid the problem of sparse extreme sample and at the same time endogenize the choice of threshold problem in KJ's method. The reminder of this paper is organized as follows. In Section 2, we review KJ's dynamic Hill estimator and propose our composite Pareto-Normal model. Monte Carlo simulations are conducted in Section 3. In Section 4, we apply our model to the monthly pooled return data and compare our tail risk estimate with KJ's method. Section 5 studies the predictive power of different tail risk estimators in volatility forecast. Finally, we conclude the paper in Section 6.

2 Methedology

2.1 The Hill estimator and dynamic power law

The Hill power law estimator estimates the shape parameter α of Pareto distribution which is referred to as tail risk measure and α may change over time (see Quintos, Fan, and Phillips (2001)) . Therefore, it is infeasible to capture the tail risk of one individual firm because of infrequency of extreme returns. KJ develops a panel estimation method which estimate the common part of the tail risk of firms over time. They pool together all firms' daily returns within a month to calculate the Hill estimator. Specifically, the lower tail of the return of asset i is assumed to follow a tail pareto distribution with the cumulative density function

$$F_{R_{i,t}}(x) = \left(\frac{x}{\theta_t}\right)^{-a_i/\lambda_t}, x < \theta_t < 0, \tag{1}$$

where θ_t is an extreme threshold in month t. Compared to the usual form of the power law, KJ decomposes the shape parameter α into two parts, a_i and $1/\lambda_t$, where a_i corresponds to the specific tail risk of asset i and $1/\lambda_t$ is the time-varying common tail risk for all firms. The Hill estimator is

$$\lambda_t^{Hill} = \frac{1}{K_t}\sum_{k=1}^{K_t} ln\left(\frac{R_{k,t}}{\theta_t}\right), \tag{2}$$

where $R_{k,t}$ is the daily return below the threshold θ_t in month t, and K_t is the total number of those exceeding returns in month t.

We can see from Equation 1 that $ln(R_{i,t}/\theta_t)$ is exponentially distributed with the rate parameter a_i/λ_t. Thus the expected value of $ln(R_{i,t}/\theta_t)$ is

$$E_{t-1}[ln(R_{i,t}/\theta_t)] = \lambda_t/a_i, \tag{3}$$

and together with Equation 2, the expected value of λ_t^{Hill} is

$$E_{t-1}(\lambda_t^{Hill}) = E_{t-1}\left[\frac{1}{K_t}\sum_{k=1}^{K_t} ln\left(\frac{R_{k,t}}{\theta_t}\right)\middle|\lambda_t, R_{k,t} < \theta_t\right] \approx \lambda_t \frac{1}{\bar{a}}, \tag{4}$$

where $\frac{1}{\bar{a}} \equiv \frac{1}{n}\sum_{n}^{i=1}\frac{1}{a_i}$. Therefore, the expected Hill estimator is equal to the common tail risk component λ_t multiplied by a constant bias term which is the mean of the reciprocal of a_i.

The threshold θ_t is the point below which returns are assumed to follow the Pareto Law. Thus a small threshold may lead to bias of the estimator because many legitimate tail data are discarded. On the other hand, a large threshold will also induce bias because non-tail data are included. KJ uses a simple rule advocated by Gabaix, Gopikrishnan, Plerou, and Stanley (2005) to determine θ_t. They fix θ_t at the fifth quantile of the pooled returns each month regardless of the risk level. We conjecture that the lower the true tail risk (small value of λ_t), the larger the bias of KJ's estimate, since more returns from the non-tail domain are included when the threshold is not adjusted accordingly.

2.2 Composite Pareto-Normal Model

In contrast to KJ's approach, we propose a composite pareto-normal model to estimate the tail risk. This model is constructed by stitching a Pareto tail to the left tail of a normal distribution. We use Pareto distribution instead of generalised Pareto distribution because the former has less parameters than generalized Pareto distribution, making it relatively simple in constructing the model and estimating the paremeters. We also impose continuity and smoothness restrictions as in Carreau and Bengio (2009).

Specifically, the Pareto-Normal (PN hereafter) density is

$$f_t(x) = \begin{cases} r_t \frac{1}{1-\Phi\left(\frac{\theta_t-\mu_t}{\sigma}\right)} f_1(x) & \text{if } x > \theta_t, \\ (1-r_t) f_2(x) & \text{if } x \le \theta_t, \end{cases} \tag{5}$$

where

$$\begin{aligned} f_1(x) &= \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left\{-\frac{(x-\mu_t)^2}{2\sigma_t^2}\right\}, \\ f_2(x) &= \frac{-\alpha_t \theta_t^{\alpha_t}}{x^{\alpha_t+1}}, \end{aligned} \tag{6}$$

$\Phi(\xi)$ is the cumulative distribution function of the standard normal distribution, $\theta_t < 0$ is the threshold splitting the domain into tail and non-tail parts, r_t is the corresponding quantile probability $Pr(x \le \theta_t)$; μ_t and σ_t^2 are the mean and variance of the non-tail part; and α_t is the shape parameter of the Pareto tail. Imposing a continuity requirement at θ such that $f(\theta_t-) = f(\theta_t+)$, we have

$$\begin{aligned} r &= \frac{f_2(\theta_t)}{\frac{f_1(\theta_t)}{1-\Phi\left(\frac{\theta_t-\mu_t}{\sigma_t}\right)} + f_2(\theta_t)} \\ &= \frac{\alpha_t\left[1-\Phi\left(\frac{\theta_t-\mu_t}{\sigma_t}\right)\right]}{\alpha_t\left[1-\Phi\left(\frac{\theta_t-\mu_t}{\sigma_t}\right)\right] - \frac{\theta_t}{\sqrt{2\pi}\sigma_t}\exp\left\{-\frac{(\theta_t-\mu_t)^2}{2\sigma_t^2}\right\}}. \end{aligned} \tag{7}$$

Further, by differentiability condition at θ such that $f'(\theta_t-) = f'(\theta_t+)$, we have

$$\frac{\theta_t-\mu_t}{\sigma_t} = \frac{\sigma_t}{\theta_t}(\alpha_t+1), \tag{8}$$

where $f'(x)$ is the first derivative of $f(x)$. These two constraints restrict the probability density function to be continuous and smooth. Our model is relatively simple with three unknown parameters, α_t, θ_t and σ_t. The threshold θ_t is one of the parameters to be estimated and thus we endogenize the choice of threshold. Note that threshold quantile probability r_t is also time varying in contrast to KJ's method which is always fixed at 0.05.

In the next section. we conduct various Monte Carlo simulations and demonstrate the performance of KJ's Hill estimator and PN model under different scenarios.

3 Monte Carlo Simulations

In this section, Monte Carlo simulations are conducted according to the composite PN distribution. The data generating process follows the PN distribution with $\mu = 0$ and $\sigma = 1$. For the tail parameter α, we choose nine different values of α such that $1/\alpha$ ranges from 0.2 to 0.6 with increment 0.05. Note that the corresponding θ and r are determined by Equation 7 and Equation 8 given μ, σ and α. Figure 1 shows the density plots of the composite PN distribution with different α. We simulate 1000 times for each values of $1/\alpha$ with three different sample sizes 10000, 50000 and 100000. The sample sizes are chosen in order to have similar magnitude of our pooled financial data. We choose ten different thresholds which are the first to tenth quantiles of the simulated data for the Hill estimation. Note that the Hill estimator is estimating the value of $1/\alpha$. We also calculate the percentage biases between the Hill estimators and the true $1/\alpha$ values.

[Figure 1]

In Table 1, panel A, B and C respectively show the simulation results in three different sample sizes. It is obvious that the Hill estimators have no significant differences in these three panels. This indicates that the sample sizes are large enough for estimation. Table 2 shows the percentage biases between Hill estimators and the true values. The result indicates that for a given risk level (fixed a row), the bias gets larger if the selected threshold is too high such that non-tail data are included. While for each threshold (fixed a column), the Hill estimator tends to overestimate when the tail risk level gets lower. This confirms our conjecture that lower tail risk level results in larger bias of the Hill estimator. We roughly draw the boarder separating the high and low biases region in the table. The diagonal pattern of the optimal threshold demonstrates that the threshold should adjust with different tail risk level. In KJ's paper, they choose the fifth percentile as the fixed threshold, which corresponds to the 5% threshold column in Table 1 and Table 2. We can see the biases are relatively large when the the tail risk is lower than 0.35 in our simulations.

[Table 1 and Table 2]

We also use the PN model to estimate the values of tail risk of the simulated data. The estimation results and the biases between estimates and the true values are shown in Table 3. It is not surprising that, for different values of $1/\alpha$, our PN model accurately estimates the true values in all sample sizes with small biases.

[Table 3]

4 The Market Tail Risk

4.1 Tail Risk Estimates: PN vs. KJ

Following KJ's paper, we collect daily data from CRSP for NYSE/AMEX/NASDAQ stocks with share codes 10 and 11. The time range of data is from January 1963 to December 2010. We calculate daily returns of each stock and pool all returns together within the same month. There are 58,047,910 returns in total and 100,780 returns on average in each month. We compare KJ's dynamic Hill estimator and the tail risk measure of our composite PN model (i.e. $1/\alpha$). The PN model is estimated by maximum likelihood method.

Figure 2 plots the KJ's Hill estimator and the PN tail risk measure during the period from January 1963 to December 2010. We can see that they share almost the same pattern but at different levels. The beginning of the series is just after a large drop in the U.S. stock market in the postwar era. Both KJ's estimator and PN's measure are high at beginning, and they start to decline rapidly until they hit the lowest level of the whole sample in the year 1968. This lowest tail risk level corresponds to the booming market in the late 1960s. The PN and KJ's tail risk series both rise afterward and the increasing momentum lasts for about ten years. After this rise, both KJ and PN show that the tail risk performs a slow and long slide during the 1990s. Later they go up to a relatively high level until 2003 in which there is a market trough. Though both the PN and JK's tail risks show the same trending pattern over the sample period, there are obvious discrepancies between their estimated risk level. The magnitudes of discrepancies are not merely a vertical location shift but changing over time. If we take the PN series as a benchmark, the differences may due to overestimate of KJ's method as shown in the simulations.

[Figure 2]

It may be a surprise that there is no increase in both tail risks during 2007 to 2009, in which the recent financial crisis happens. However, KJ and Brownlees, Engle, and Kelly (2011) propose an explaination that the tail risk remains stable while the market volatility increases dramatically. They argue that an expanding of the threshold may absorb the effect of changes in volatility, which keeps the estimates of the tail risk unaffected. Their arguement coincides with Figure 3 which plots the lower tail threshold series of both PN and JK's methods and the realized S&P 500 index volatility. In Figure 3, the thresholds of both methods go down strongly while the market volatility increases drastically to the peak, during the finacial crisis period. Note that the thresholds of JK are always the 5% quantile levels. And it is obvious that the threshold corresponding to the 5% level will be smaller in the year of heavy tail. In contrast, both the threshold and its corresponding confidence level of our method are adjusting over the years that our measure gives more freedom in assessing tail risk. Similar to the risk measure, our PN estimated thresholds are always disproportionately lower than KJ's fixed 5% thresholds.

[Figure 3]

4.2 Decomposition of Variance

Since the distribution is divided into tail part and non-tail part, total variance of returns can be decomposed according to the law of total variance, also known as variance decomposition formula, see Weiss, Holmes, and Hardy (2006). The law of total variance states that if $A_1, A_2, \ldots, A_n$ is a partition of the whole outcome space and these events are mutually exclusive and exhaustive, then

$$\begin{aligned} Var(X) = &\sum_{i=1}^{n} Var(X|A_i)P(A_i) \\ &+ \sum_{i=1}^{n} E(X|A_i)^2(1-P(A_i))P(A_i) \\ &- 2\sum_{i=2}^{n}\sum_{j=1}^{i-1} E(X|A_i)P(A_i)E(X|A_j)P(A_j). \end{aligned} \tag{9}$$

The first part in this formula is the expectation of the conditional variance and the other two components are the variance of the conditional expectation. In our current context, the corresponding A_1 is $\{X > \theta\}$ and A_2 is the tail $\{X \leq \theta\}$. And, thus,

$$\begin{aligned} Var(X) =& Var(X|X>\theta)P(X>\theta) + Var(X|X\leq\theta)P(X\leq\theta) \\ &+ E(X|X>\theta)^2(1-P(X>\theta))P(X>\theta) \\ &+ E(X|X\leq\theta)^2(1-P(X\leq\theta))P(X\leq\theta) \\ &- 2E(X|X>\theta)P(X>\theta)E(X|X\leq\theta)P(X\leq\theta), \end{aligned} \tag{10}$$

where $E(X|X > \theta)$ and $Var(X|X > \theta)$ are mean and variance of the non-tail part, and $E(X|X \leq \theta)$ and $Var(X|X \leq \theta)$ are mean and variance of the tail part.

In our PN model, the non-tail part of the PN model is a truncated normal distribution with mean and variance,

$$E(X|X > \theta) = \mu + \sigma \frac{\phi(\frac{\theta-\mu}{\sigma})}{1 - \Phi(\frac{\theta-\mu}{\sigma})}, \tag{11}$$

$$Var(X|X > \theta) = \sigma^2 \left[1 - \left(\frac{\phi(\frac{\theta-\mu}{\sigma})}{1 - \Phi(\frac{\theta-\mu}{\sigma})} \right)^2 + \frac{\frac{\theta-\mu}{\sigma}\phi(\frac{\theta-\mu}{\sigma})}{1 - \Phi(\frac{\theta-\mu}{\sigma})} \right], \tag{12}$$

where $\phi(\xi)$ is the probability density function of the standard normal distribution and $\Phi(\xi)$ is its cumulative distribution function. The tail part of PN model is a pareto distribution, of which the mean and variance are

$$E(X|X \leq \theta) = \begin{cases} \infty & \text{if } \alpha \leq 1, \\ \frac{\alpha\theta}{\alpha-1} & \text{if } \alpha > 1, \end{cases} \tag{13}$$

$$Var(X|X \leq \theta) = \begin{cases} \infty & \text{if } \alpha \in (1, 2], \\ \left(\frac{\theta}{\alpha-1}\right)^2 \frac{\alpha}{\alpha-2} & \text{if } \alpha > 2. \end{cases} \tag{14}$$

It can be easily derived from Equation 10 to Equation 12 that, when $\alpha > 2$, the variance of the whole PN model is

$$\begin{aligned} Var_{PN}(X) =& rVar(X|X > \theta) + (1-r)Var(X|X \leq \theta) \\ &+ r(1-r)[E(X|X > \theta) - E(X|X \leq \theta)]^2 \\ =& r\sigma^2 \left[1 - \left(\frac{\phi(\frac{\theta-\mu}{\sigma})}{1 - \Phi(\frac{\theta-\mu}{\sigma})} \right)^2 + \frac{\frac{\theta-\mu}{\sigma}\phi(\frac{\theta-\mu}{\sigma})}{1 - \Phi(\frac{\theta-\mu}{\sigma})} \right] + (1-r)\left(\frac{\theta}{\alpha-1}\right)^2 \frac{\alpha}{\alpha-2} \\ &+ r(1-r)\left[\left(\mu + \sigma \frac{\phi(\frac{\theta-\mu}{\sigma})}{1 - \Phi(\frac{\theta-\mu}{\sigma})} \right) - \left(\frac{\alpha\theta}{\alpha-1} \right) \right]^2. \end{aligned} \tag{15}$$

Thus the total variance, denoted as Var_{PN}, is composed of three parts: $rVar(X|X > \theta)$, the variance of the truncated normal distribution multiplied by its weight r; $(1 -$

$r)Var(X|X \leq \theta)$, the variance of the pareto distribution multiplied by $(1 - r)$; and $r(1 - r)\left[E(X|X > \theta) - E(X|X \leq \theta)\right]^2$.

Next, we consider using KJ's dynamic Hill estimators to construct the total variance based on the variance decomposition formula Equation 10, Equation 13 and Equation 14. To apply the Hill's formula, we choose four different thresholds θ which are the 1st, 3rd, 5th and 7th percentile of the data. Note that KJ's tail risk estimator corresponds to the 5th percentile. For the tail part, the mean and variance are computed by using Equation 13 and Equation 14 with α replaced by the reciprocal of the Hill estimator. The mean and variance of non-tail part are calculated by sample mean and sample variance using the data above the threshold. Finally, the total variance can be obtained according to Equation 10, which is denoted as Var_{KJ}. In the rest of this section, we compares the computed total variances and the ture sample variances of the stock returns in each month. Theoretically, the reconstructed total variance should match the market variance calculated from the return data if the model is correctly specified. Deviation between the two indicates potential misspecifications in the assumptions or calculations.

We plot the total variance computed from the two different methods as stated above. Figure 4 exhibits the total variance calculated by using KJ's Hill estimators with respect to four different thresholds. If the threshold is fixed at the first percentile of the data, the difference between the total variance Var_{KJ} and the market variance is tiny, which is plotted in Figure 4(a). However, it can be seen in Figure 4(b), (c) and (d) that as threshold goes up, Var_{KJ} deviates more from the market variance. The figure shows that the computed total variance by the 1% threshold is very close to the market variance. Figure 5 shows the total variance Var_{PN} computed by the composite PN model. We can see that Var_{PN} is nearly identical with the market variance. The figure illustrates that our PN model maps the tail risk to volatility through equation 15. This reflects our PN tail risk and the volatility both measure the same source of risk but in different angles.

[Figure 4 and Figure 5]

In Figure 6, we plot the three components of the total volitility in Equation 15, alongside with the market variance. As can be seen in the gragh, a large portion goes to the non-tail variance, and the tail variance contributes the least. However, the three components share the same cycling pattern over the years.

[Figure 6]

5 Application in Volatility Forecast

After the international stock market crash of 1987, modeling and forecasting volatility of the stock market has become one of the major tasks of financial regulators, investors and researchers. Volatility is often considered as a measure of asset's or portfolio's risk. Investors and fund managers are always cautious when the volatility of their portfolio surges. The volatilities of securities are also important for portfolio construction. Risk averting investors may give more weights on assets which is less volatile in order to maintain their portfolios at a low risk level. Meanwhile, volatility is critical for risk management such as stress-testing and calculating value at risk.

In option pricing, the most popular pricing model for options on equity is developed by Black and Scholes (1973). Other options pricing models are also developed after Black and Scholes' seminal work, e.g. Garman and Kohlhagen (1983)'s model for options on futures, Seidel and Ginsberg (1983)'s model for options on currency futures. Since the volatility of the underlying asset is an explicit parameter of any option pricing model, it is important to obtain the accurately forecasted volatility. Moreover, in volatility arbitrage, future realized volatility of the underlying assets must be predicted by traders. Volatility arbitrage is implemented by trading a delta-neutral portfolio containing an option and its underlying asset. The purpose of this arbitrage is to take advantage of the differences between the implied volatility of the option and a forecasted future realized volatility of the underlying.

Besides the essential role of volatility in option pricing and hedging, an accurately forecasted volatility is also useful to instruments which trade volatility directly, such as variance swap, volatility swap, VIX futures contract and exchange-listed VIX option. Therefore, volatility forecast is crucial in many financial pricing and trading activities.

The main stream volatility forecast model is the GARCH type model developed by Engle (1982) and Bollerslev (1986). Variants of the GARCH model include the EGARCH by Nelson (1991), J.P. Morgan's RiskMetrics, GJR-GARCH by Glosten, Jagannathan, and Runkle (1993), among others. Stochastic volatility models are also popular recently, see Andersen, Davis, Kreiss, and Mikosch (2009) for a comprehensive summary of volatility models and applications.

In this section, we investigate the predictive power of our proposed PN and JK's tail risk estimators in forecasting future volatilities. We follow Paye (2012) to focus on the realized volatility,

$$LVOL_t = ln(\sqrt{mRV_t}), \tag{16}$$

where $LVOL_t$ is the natural logarithm of annualized volatility in month t, and m corresponds to the number of periods within a year which is 12 here for monthly sampling. RV_t is defined as the realized monthly variance in month t, which is calculated by summing up the squared daily returns,

$$RV_t = \sum_{i=1}^{N_t} R_{i,t}^2, \tag{17}$$

where N_t is the number of trading days in month t and $R_{i,t}$ denotes the daily excess return on the S&P 500 index of the ith trading day in month t. In the following sections, we conduct both in-sample and out-of-sample analysis for realized volatility forecast.

5.1 In-Sample Analysis

For the in-sample regression, we regress log volatility on tail risk and its own lag values,

$$LVOL_t = \beta_0 + \beta_1 TailRisk_{t-1} + \sum_{k=1}^{K} \gamma_t LVOL_{t-k} + \epsilon_t \tag{18}$$

where $LVOL_t$ is defined as the natural logarithm of annualized volatility in month t, and $LVOL_{t-k}$ represents the lagged volatilities. The $TailRisk$ in the regression is approximated by our PN estimator and KJ's Hill estimator such that

$$\begin{aligned} LVOL_t &= \beta_0 + \beta_1 \left(\frac{1}{\alpha_{t-1}}\right) + \sum_{k=1}^{K} \gamma_t LVOL_{t-k} + \epsilon_t, \\ LVOL_t &= \beta_0 + \beta_1 \lambda_{t-1}^{KJ} + \sum_{k=1}^{K} \gamma_t LVOL_{t-k} + \epsilon_t, \end{aligned} \tag{19}$$

Table 4 shows the autocorrelation function (ACF) and partial autocorrelation function (PACF) of $LVOL_t$. It can be observed from Table 4 that the PACF falls below 0.1 after lag5. Therefore, we choose 5 lags in our regression models, i.e. K = 5.

[Table 4]

The results of these two regressions are shown in Table 5. According to Panel A, during the sample period from 1963 to 2010, the null hypothesis of no predictive power is rejected for PN estimator at 10% significance level. On the contrary, Panel B indicates that KJ's Hill estimator is insignificant in the in-sample regression. The regression confirms our previous result that the PN tail risk estimator maps well to the volatility whereas the JK's estimator falls short in reconstructing the overall volatility.

[Table 5]

5.2 Out-of-Sample Analysis

In this section, we investigate the out-of-sample predictive ability of the PN's measure and the KJ's Hill estimator by conducting the recursive regressions based on regression Equation 19. The initial estimating window is from January 1963 to December 1973, a 120 month period. We use the estimated regression coefficients to predict the volatility in the next month $t+1$. Then we expand the estimating window by one month to include the $t+1$ data, and forecast the next month's volatility using the new estimated coefficients. We repeat this process until the whole sample is exhausted.

Figure 7 plots the recursive estimated coefficients β and the 95% confidence intervals of both the PN estimator and the KJ's Hill estimator from January 1973 to December 2010. The results show that the recursive estimated coefficients of PN tail risk are significantly different from zero in the whole sample period. This demonstrates that our PN tail risk has predictive power in forecasting future volatilities from an out-of-sample perspective. On the contrary, Figure 7(b) shows that the estimated coefficients of KJ's Hill estimator are insignificant throughout the sample period. We can conclude that KJ's estimator has no predictive power in forecasting future volatilities in both in-sample and out-of-sample perspectives.

[Figure 7]

Next, we include other standard predictors for the volatility forecast for reference.[1] The summary statistics of $LVOL_t$, PN tail risk, KJ's tail risk and other predictors are calculated in Table 6. In each volatility forecasting regression, we use the tail risk together with one predictor as follow,

$$LVOL_t = \beta_0 + \beta_1 TailRisk_{t-1} + \beta_2 Z_{t-1} + \sum_{k=1}^{K} \gamma_t LVOL_{t-k} + \epsilon_t \tag{20}$$

where $TailRisk_{t-1}$ is λ_{t-1}^{KJ} or $1/\alpha_{t-1}$ and Z_{t-1} is the predictor in question. Similar to the previous out-of-sample regression, regression Equation 20 is conducted in the recursive manner as described before.

[Table 6]

We adopt the out-of-sample R^2 statistic to measure the forecasting perfomance,

$$R_{OoS}^2 = 1 - \frac{\sum(LVOL_{t+1} - \widehat{LVOL}_{t+1|t})^2}{\sum(LVOL_{t+1} - \overline{LVOL}_t)^2}, \tag{21}$$

where $\widehat{LVOL}_{t+1|t}$ is the out-of-sample prediction of the natural logarithm of annualized volatility in month $t+1$ and $\overline{LVOL}_t$ is the historical average volatility through month t. The higher the value of R_{OoS}^2, the more predictive power the model entails. Note that the R_{OoS}^2 can be negatively valued.

The R_{OoS}^2 are shown in Table 7. The first two rows are the forecasting models correspond to Equation 19 and Figure 7 with only the $TailRisk$ included. The rest of the row corresponds to different combinations of tail risks and predictors in regression Equation 20. The first two columns are the R_{OoS}^2 calculated using full sample data. We can see most of the results associated with the PN tail risk outperform that of the KJ's

[1] See the comprehensive description of the predictors in Welch and Goyal (2008). See also Paye (2012).

tail risk. For robust check, we split the sample into two sub-samples, the period from 1960 to 1986 and from 1987 to 2010. The forecasting results are shown in column 3 to column 6. In the first sub-sample, most of the forecasting models associated with the PN tail risk show higher R^2_{OoS} than that of KJ's tail risk. But there are four cases associated with the PN tail risk show poorer performances than the KJ's method. However, in the second sub-sample, our PN tail risk far outperform the KJ's method. The forecasting results show that our proposed PN tail risk measure contains useful information in forecasting volatility.

[Table 7]

6 Conclusion

This paper focuses on the tail risk measure in financial market. We conduct simulation experiments to find out that their measurement is very sensitive to the selection of threshold. To overcome the weakness of their method, we propose the composite PN model which is a combination of a Pareto distribution and a Normal distribution. Monte Carlo simulations show that the KJ's method overestimates the tail risk in various scenarios. We also examine the variance decomposition according to the law of total variance and prove that, form this perspective, the performance of our composite PN model is better than KJ's approach.

Volatility forecasting is crucial for many aspects of financial market such as asset allocation, asset pricing, and risk management. We try to link our tail risk model to volatility prediction. The results demonstrate that the PN tail risk predicts the market future volatilities, whereas KJ's tail risk is lack of predictive power in volatility forecast.

References

ANDERSEN, T. G., R. A. DAVIS, J.-P. KREISS, AND T. V. MIKOSCH (2009): *Handbook of financial time series*. Springer Science & Business Media.

ARTZNER, P. (1997): "Applebaum, D.(2004). Lévy Processes and Stochastic Calculus (Cambridge University Press). Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D.(1997). Thinking coherently, Risk 10, pp. 68–71.," *Risk*, 10, 68–71.

ARTZNER, P., F. DELBAEN, J.-M. EBER, AND D. HEATH (1999): "Coherent measures of risk," *Mathematical finance*, 9(3), 203–228.

BAWA, V. S. (1975): "Optimal rules for ordering uncertain prospects," *Journal of Financial Economics*, 2(1), 95–121.

BEHRENS, C. N., H. F. LOPES, AND D. GAMERMAN (2004): "Bayesian analysis of extreme events with threshold estimation," *Statistical Modelling*, 4(3), 227–244.

BLACK, F., AND M. SCHOLES (1973): "The pricing of options and corporate liabilities," *The journal of political economy*, pp. 637–654.

BOLLERSLEV, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, 31(3), 307–327.

BROWNLEES, C. T., R. F. ENGLE, AND B. T. KELLY (2011): "A practical guide to volatility forecasting through calm and storm," *Available at SSRN 1502915*.

CARREAU, J., AND Y. BENGIO (2009): "A hybrid Pareto model for asymmetric fat-tailed data: the univariate case," *Extremes*, 12(1), 53–76.

CLAUSET, A., C. R. SHALIZI, AND M. E. NEWMAN (2009): "Power-law distributions in empirical data," *SIAM review*, 51(4), 661–703.

Clauset, A., M. Young, and K. S. Gleditsch (2007): "On the frequency of severe terrorist events," *Journal of Conflict Resolution*, 51(1), 58–87.

Coles, S., J. Bawa, L. Trenner, and P. Dorazio (2001): *An introduction to statistical modeling of extreme values*, vol. 208. Springer.

De Melo Mendes, B. V., and H. F. Lopes (2004): "Data driven estimates for mixtures," *Computational statistics & data analysis*, 47(3), 583–598.

Engle, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica: Journal of the Econometric Society*, pp. 987–1007.

Fishburn, P. C. (1977): "Mean-risk analysis with risk associated with below-target returns," *The American Economic Review*, 67(2), 116–126.

Gabaix, X., P. Gopikrishnan, V. Plerou, and H. E. Stanley (2005): "Institutional investors and stock market volatility," Discussion paper, National Bureau of Economic Research.

Garman, M. B., and S. W. Kohlhagen (1983): "Foreign currency option values," *Journal of international Money and Finance*, 2(3), 231–237.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993): "On the relation between the expected value and the volatility of the nominal excess return on stocks," *The journal of finance*, 48(5), 1779–1801.

Hill, B. M. (1975): "A simple general approach to inference about the tail of a distribution," *The annals of statistics*, 3(5), 1163–1174.

Jansen, D. W., and C. G. De Vries (1991): "On the frequency of large stock returns: Putting booms and busts into perspective," *The review of economics and statistics*, pp. 18–24.

Kelly, B., and H. Jiang (2014): "Tail risk and asset prices," *Review of Financial Studies*, 27(10), 2841–2871.

Mandelbrot, B. B. (1997): "The variation of certain speculative prices," in *Fractals and Scaling in Finance*, pp. 371–418. Springer.

Nelson, D. B. (1991): "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica: Journal of the Econometric Society*, pp. 347–370.

Paye, B. S. (2012): "Déjà vol: Predictive regressions for aggregate stock market volatility using macroeconomic variables," *Journal of Financial Economics*, 106(3), 527–546.

Quintos, C., Z. Fan, and P. C. Phillips (2001): "Structural change tests in tail behaviour and the Asian crisis," *The Review of Economic Studies*, 68(3), 633–663.

Roy, A. (1952): "Safety first and the holding of assets, Economet rica 20: 431–449," .

Seidel, A. D., and P. M. Ginsberg (1983): *Commodities trading: foundations, analysis, and operations*. Prentice Hall.

Weiss, N. A., P. T. Holmes, and M. Hardy (2006): *A course in probability*. Pearson Addison Wesley Boston, Massachusetts, USA.

Welch, I., and A. Goyal (2008): "A comprehensive look at the empirical performance of equity premium prediction," *Review of Financial Studies*, 21(4), 1455–1508.

Zhao, X., C. Scarrott, L. Oxley, and M. Reale (2010): "Extreme value modelling for forecasting market crisis impacts," *Applied Financial Economics*, 20(1-2), 63–72.

Table 1: KJ's Hill estimators from simulated data based on PN distribution.

Panel A: n=10000

$1/\alpha$	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
0.2	0.2012	0.2054	0.2149	0.2264	0.2386	0.2516	0.2647	0.2784	0.2922	0.3063
	(0.0204)	(0.0145)	(0.0118)	(0.0106)	(0.0101)	(0.0096)	(0.0091)	(0.0088)	(0.0087)	(0.0086)
0.25	0.2490	0.2489	0.2520	0.2584	0.2668	0.2767	0.2875	0.2993	0.3115	0.3243
	(0.0240)	(0.0168)	(0.0140)	(0.0122)	(0.0113)	(0.0106)	(0.0102)	(0.0096)	(0.0094)	(0.0094)
0.3	0.2989	0.2994	0.2997	0.3015	0.3054	0.3113	0.3194	0.3286	0.3385	0.3495
	(0.0301)	(0.0219)	(0.0171)	(0.0149)	(0.0136)	(0.0128)	(0.0120)	(0.0113)	(0.0110)	(0.0109)
0.35	0.3496	0.3498	0.3494	0.3497	0.3508	0.3532	0.3578	0.3643	0.3717	0.3802
	(0.0347)	(0.0242)	(0.0202)	(0.0173)	(0.0155)	(0.0145)	(0.0136)	(0.0131)	(0.0126)	(0.0121)
0.4	0.4002	0.3999	0.4006	0.4004	0.4001	0.4003	0.4021	0.4055	0.4105	0.4171
	(0.0406)	(0.0274)	(0.0229)	(0.0201)	(0.0178)	(0.0162)	(0.0149)	(0.0140)	(0.0135)	(0.0129)
0.45	0.4507	0.4499	0.4497	0.4501	0.4500	0.4504	0.4503	0.4513	0.4539	0.4579
	(0.0457)	(0.0332)	(0.0267)	(0.0232)	(0.0209)	(0.0192)	(0.0173)	(0.0163)	(0.0155)	(0.0149)
0.5	0.4979	0.4993	0.4993	0.4990	0.4993	0.5000	0.4999	0.4998	0.5008	0.5027
	(0.0512)	(0.0360)	(0.0293)	(0.0251)	(0.0228)	(0.0201)	(0.0185)	(0.0177)	(0.0166)	(0.0157)
0.55	0.5515	0.5494	0.5506	0.5506	0.5510	0.5511	0.5510	0.5508	0.5508	0.5512
	(0.0541)	(0.0395)	(0.0330)	(0.0282)	(0.0252)	(0.0225)	(0.0211)	(0.0196)	(0.0185)	(0.0179)
0.6	0.5986	0.5983	0.5991	0.5994	0.5998	0.5994	0.5995	0.5997	0.5996	0.5997
	(0.0616)	(0.0444)	(0.0375)	(0.0319)	(0.0285)	(0.0259)	(0.0237)	(0.0223)	(0.0209)	(0.0196)

Panel B: n=50000

$1/\alpha$	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
0.2	0.1997	0.2040	0.2139	0.2259	0.2386	0.2516	0.2651	0.2787	0.2925	0.3066
	(0.0091)	(0.0066)	(0.0054)	(0.0050)	(0.0046)	(0.0043)	(0.0042)	(0.0040)	(0.0042)	(0.0040)
0.25	0.2501	0.2502	0.2526	0.2587	0.2673	0.2773	0.2882	0.3000	0.3123	0.3250
	(0.0112)	(0.0080)	(0.0066)	(0.0056)	(0.0052)	(0.0049)	(0.0046)	(0.0044)	(0.0043)	(0.0042)
0.3	0.2998	0.3000	0.2996	0.3013	0.3052	0.3115	0.3193	0.3284	0.3385	0.3494
	(0.0137)	(0.0099)	(0.0081)	(0.0070)	(0.0063)	(0.0058)	(0.0053)	(0.0051)	(0.0050)	(0.0049)
0.35	0.3503	0.3502	0.3501	0.3500	0.3506	0.3534	0.3581	0.3644	0.3720	0.3806
	(0.0156)	(0.0111)	(0.0092)	(0.0078)	(0.0069)	(0.0064)	(0.0061)	(0.0058)	(0.0055)	(0.0052)
0.4	0.3996	0.3999	0.4000	0.4001	0.4002	0.4004	0.4022	0.4055	0.4106	0.4170
	(0.0176)	(0.0128)	(0.0100)	(0.0089)	(0.0082)	(0.0075)	(0.0070)	(0.0065)	(0.0062)	(0.0059)
0.45	0.4508	0.4502	0.4500	0.4501	0.4501	0.4500	0.4502	0.4514	0.4540	0.4580
	(0.0198)	(0.0141)	(0.0115)	(0.0102)	(0.0094)	(0.0084)	(0.0078)	(0.0072)	(0.0068)	(0.0064)
0.5	0.5003	0.4996	0.4996	0.4996	0.4996	0.4996	0.4993	0.4995	0.5003	0.5022
	(0.0216)	(0.0152)	(0.0127)	(0.0108)	(0.0098)	(0.0089)	(0.0082)	(0.0076)	(0.0073)	(0.0071)
0.55	0.5486	0.5496	0.5497	0.5498	0.5498	0.5497	0.5498	0.5499	0.5501	0.5504
	(0.0243)	(0.0173)	(0.0141)	(0.0121)	(0.0107)	(0.0099)	(0.0090)	(0.0086)	(0.0082)	(0.0078)
0.6	0.5996	0.5997	0.5993	0.5997	0.6001	0.6001	0.6001	0.6002	0.6002	0.6001
	(0.0263)	(0.0184)	(0.0151)	(0.0135)	(0.0124)	(0.0111)	(0.0102)	(0.0096)	(0.0092)	(0.0089)

Panel C: n=100000

$1/\alpha$	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
0.2	0.2000	0.2042	0.2141	0.2259	0.2386	0.2516	0.2650	0.2785	0.2924	0.3065
	(0.0065)	(0.0045)	(0.0038)	(0.0033)	(0.0030)	(0.0029)	(0.0028)	(0.0027)	(0.0027)	(0.0027)
0.25	0.2500	0.2499	0.2524	0.2587	0.2673	0.2772	0.2882	0.2998	0.3120	0.3247
	(0.0081)	(0.0057)	(0.0047)	(0.0040)	(0.0037)	(0.0035)	(0.0033)	(0.0032)	(0.0030)	(0.0030)
0.3	0.3002	0.3001	0.3000	0.3014	0.3055	0.3117	0.3196	0.3286	0.3388	0.3496
	(0.0095)	(0.0068)	(0.0054)	(0.0047)	(0.0042)	(0.0038)	(0.0036)	(0.0035)	(0.0033)	(0.0032)
0.35	0.3501	0.3504	0.3504	0.3503	0.3510	0.3536	0.3583	0.3644	0.3720	0.3805
	(0.0112)	(0.0077)	(0.0063)	(0.0054)	(0.0049)	(0.0046)	(0.0042)	(0.0040)	(0.0038)	(0.0036)
0.4	0.4002	0.3999	0.4001	0.4001	0.4000	0.4003	0.4021	0.4055	0.4105	0.4169
	(0.0125)	(0.0091)	(0.0074)	(0.0063)	(0.0057)	(0.0052)	(0.0049)	(0.0045)	(0.0043)	(0.0042)
0.45	0.4492	0.4499	0.4499	0.4500	0.4499	0.4499	0.4500	0.4511	0.4537	0.4577
	(0.0143)	(0.0095)	(0.0079)	(0.0068)	(0.0062)	(0.0057)	(0.0054)	(0.0050)	(0.0047)	(0.0045)
0.5	0.4996	0.4996	0.4996	0.5000	0.4999	0.4999	0.4999	0.4999	0.5007	0.5027
	(0.0155)	(0.0109)	(0.0089)	(0.0077)	(0.0070)	(0.0064)	(0.0059)	(0.0056)	(0.0052)	(0.0050)
0.55	0.5496	0.5499	0.5503	0.5500	0.5501	0.5500	0.5501	0.5501	0.5501	0.5505
	(0.0172)	(0.0125)	(0.0104)	(0.0090)	(0.0079)	(0.0073)	(0.0066)	(0.0062)	(0.0057)	(0.0054)
0.6	0.5999	0.5994	0.5996	0.5996	0.5995	0.5996	0.5999	0.5998	0.6000	0.5999
	(0.0190)	(0.0137)	(0.0113)	(0.0096)	(0.0085)	(0.0077)	(0.0072)	(0.0066)	(0.0063)	(0.0060)

The table reports KJ's Hill estimators computed using simulated data based on PN distribution. Panel A, B and C respectively show the simulation results in three different sample sizes 10000, 50000 and 100000. For each $1/\alpha$ and in each sample size, KJ's Hill estimators are computed using ten different thresholds which are the first to tenth percentile of the simulated data. The reported values are sample average of the estimators over 1000 repetitions and the sample standard deviations are in parentheses.

Table 2: Biases of KJ's Hill estimators from simulated data based on PN distribution.

Panel A: n=10000

$1/\alpha$	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
0.2	0.60%	2.68%	7.43%	13.18%	19.30%	25.80%	32.36%	39.21%	46.09%	53.17%
0.25	-0.40%	-0.44%	0.82%	3.35%	6.71%	10.67%	14.99%	19.73%	24.61%	29.74%
0.3	-0.36%	-0.19%	-0.09%	0.50%	1.80%	3.78%	6.48%	9.53%	12.83%	16.50%
0.35	-0.10%	-0.05%	-0.18%	-0.07%	0.22%	0.92%	2.22%	4.08%	6.20%	8.63%
0.4	0.05%	-0.03%	0.14%	0.11%	0.03%	0.07%	0.51%	1.37%	2.63%	4.26%
0.45	0.17%	-0.03%	-0.07%	0.03%	0.00%	0.09%	0.07%	0.30%	0.86%	1.75%
0.5	-0.41%	-0.13%	-0.15%	-0.19%	-0.14%	-0.01%	-0.02%	-0.03%	0.16%	0.55%
0.55	0.27%	-0.11%	0.12%	0.12%	0.19%	0.21%	0.17%	0.14%	0.15%	0.22%
0.6	-0.24%	-0.28%	-0.16%	-0.09%	-0.04%	-0.10%	-0.09%	-0.05%	-0.06%	-0.05%

Panel B: n=50000

$1/\alpha$	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
0.2	-0.14%	2.01%	6.94%	12.94%	19.31%	25.82%	32.54%	39.36%	46.26%	53.32%
0.25	0.03%	0.06%	1.03%	3.50%	6.92%	10.91%	15.30%	19.98%	24.90%	29.98%
0.3	-0.06%	0.01%	-0.13%	0.43%	1.73%	3.82%	6.44%	9.47%	12.82%	16.47%
0.35	0.09%	0.06%	0.03%	-0.01%	0.17%	0.97%	2.30%	4.11%	6.27%	8.73%
0.4	-0.09%	-0.03%	0.00%	0.02%	0.05%	0.11%	0.55%	1.38%	2.65%	4.25%
0.45	0.17%	0.04%	0.00%	0.01%	0.02%	-0.01%	0.05%	0.32%	0.90%	1.78%
0.5	0.06%	-0.08%	-0.07%	-0.08%	-0.08%	-0.07%	-0.13%	-0.10%	0.05%	0.43%
0.55	-0.25%	-0.08%	-0.06%	-0.03%	-0.04%	-0.05%	-0.05%	-0.01%	0.01%	0.08%
0.6	-0.06%	-0.05%	-0.11%	-0.04%	0.02%	0.01%	0.02%	0.03%	0.04%	0.02%

Panel C: n=100000

$1/\alpha$	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
0.2	0.02%	2.10%	7.03%	12.96%	19.28%	25.82%	32.50%	39.26%	46.20%	53.27%
0.25	-0.01%	-0.04%	0.94%	3.46%	6.91%	10.89%	15.27%	19.93%	24.78%	29.88%
0.3	0.06%	0.02%	0.02%	0.47%	1.82%	3.91%	6.53%	9.55%	12.93%	16.55%
0.35	0.02%	0.11%	0.11%	0.09%	0.28%	1.04%	2.36%	4.12%	6.27%	8.72%
0.4	0.06%	-0.02%	0.03%	0.03%	0.00%	0.07%	0.53%	1.37%	2.63%	4.23%
0.45	-0.18%	-0.02%	-0.03%	-0.01%	-0.02%	-0.03%	-0.01%	0.24%	0.82%	1.72%
0.5	-0.08%	-0.08%	-0.08%	-0.01%	-0.01%	-0.03%	-0.03%	-0.02%	0.15%	0.53%
0.55	-0.08%	-0.02%	0.05%	0.00%	0.02%	0.00%	0.01%	0.03%	0.02%	0.08%
0.6	-0.01%	-0.10%	-0.07%	-0.07%	-0.09%	-0.06%	-0.02%	-0.03%	-0.01%	-0.01%

The talbe exhibits the percentage biases between the estimates in Table 1 and the true $1/\alpha$.

Table 3: PN estimators from simulated data based on PN distribution.

$1/\alpha$	n=10000		n=50000		n=100000	
	Estimates	Bias	Estimates	Bias	Estimates	Bias
0.2	0.1987 (0.0156)	0.65%	0.1994 (0.0071)	0.28%	0.1997 (0.0049)	0.16%
0.25	0.2496 (0.0140)	0.15%	0.2499 (0.0065)	0.06%	0.2499 (0.0044)	0.02%
0.3	0.2998 (0.0139)	0.07%	0.2997 (0.0061)	0.11%	0.2999 (0.0044)	0.03%
0.35	0.3499 (0.0139)	0.03%	0.3499 (0.0062)	0.04%	0.3500 (0.0042)	0.01%
0.4	0.3995 (0.0140)	0.12%	0.4001 (0.0063)	0.02%	0.4001 (0.0045)	0.02%
0.45	0.4496 (0.0147)	0.08%	0.4499 (0.0066)	0.03%	0.4499 (0.0045)	0.02%
0.5	0.4999 (0.0159)	0.02%	0.4995 (0.0113)	0.11%	0.4995 (0.0088)	0.11%
0.55	0.5471 (0.0273)	0.53%	0.5476 (0.0235)	0.44%	0.5486 (0.0188)	0.26%
0.6	0.5925 (0.0453)	1.25%	0.5910 (0.0489)	1.51%	0.5906 (0.0493)	1.57%

The table exhibits the sample averages and sample standard deviations (in parentheses) of the PN model under three sampler sizes. Percentage biases are also reported.

Table 4: Autocorrelation function and partial autocorrelation function of $LVOL_t$.

	lag1	lag2	lag3	lag4	lag5	lag6	lag7	lag8	lag9	lag10
ACF	0.748	0.669	0.610	0.549	0.552	0.520	0.496	0.471	0.458	0.436
PACF	0.748	0.250	0.118	0.037	0.156	0.043	0.033	0.021	0.058	0.008

The table reports the autocorrelation function (ACF) and partial autocorrelation function (PACF) of $LVOL_t$.

Table 5: In-sample analyses of realized volatility forecasts.

Panel A: In-sample analysis for PN tail risk

R^2=0.6026

	Coefficient	t-statistic	p-value
$1/\alpha_{t-1}$	0.5011	1.94	0.0530
$LVOL_{t-1}$	0.5088	10.59	0.0000
$LVOL_{t-2}$	0.1610	3.19	0.0010
$LVOL_{t-3}$	0.0726	1.4	0.1620
$LVOL_{t-4}$	-0.0385	-0.84	0.4030
$LVOL_{t-5}$	0.1548	3.67	0.0000

Panel B: In-sample analysis for KJ's Hill tail risk

R^2=0.6014

	Coefficient	t-statistic	p-value
λ_{t-1}^{KJ}	-0.2813	-1.34	0.1810
$LVOL_{t-1}$	0.5063	10.56	0.0000
$LVOL_{t-2}$	0.1638	3.29	0.0010
$LVOL_{t-3}$	0.0733	1.41	0.1590
$LVOL_{t-4}$	-0.0437	-0.96	0.3390
$LVOL_{t-5}$	0.1591	3.74	0.0000

The table reports the results of in-sample regressions of $LVOL_t$ on the PN and KJ's Hill tail risks. The sample period is from 1963 to 2010.

Table 6: Summary statistics of $LVOL_t$, PN tail risk, KJ's Hill tail risk and other predictors.

	mean	std	skewness	kurtosis	JB test p-value
$LVOL_t$	-2.0959	0.4715	0.4219	3.9553	<0.001
$1/\alpha$	0.2986	0.0447	-0.3157	4.1998	<0.001
λ^{KJ}	0.4275	0.0561	-0.4550	2.6337	<0.001
Log dividend price ratio	-3.5588	0.4107	-0.4052	2.3634	<0.001
Log dividend yield	-3.5536	0.4110	-0.4106	2.3858	<0.001
Log earning price ratio	-2.8236	0.4600	-0.7545	5.4924	<0.001
Log smooth earning price ratio	-3.0647	0.3542	-0.0957	2.5455	0.051
Log dividend payout ratio	-0.7351	0.3171	3.0316	19.1921	<0.001
Book to market ratio	0.5198	0.2726	0.5818	2.3683	<0.001
T-bill rate	0.0539	0.0291	0.8016	4.4686	<0.001
Term spread	0.0174	0.0153	-0.2316	2.6092	0.017
Default yield spread	0.0104	0.0047	1.6259	6.5521	<0.001
Default return spread	0.0001	0.0143	-0.3028	10.4726	<0.001
Inflation	0.0034	0.0035	-0.1228	7.1308	<0.001
Log total net payout yield	-2.2211	0.2242	-1.4142	6.2465	<0.001

The table reports the summary statistics of $LVOL_t$, PN tail risk, KJ's Hill tail risk and other predictors. Default return spread is the difference between long-term corporate bond and long-term government bond returns. Default yield spread is the difference between the yield on BAA-rated corporate bonds and the yield on long-term US government bonds. Term spread is the difference between the long-term yield on government bonds and the Treasury bill rate.

Table 7: Volatility forecast performances measured by out-of-sample R^2_{OoS}.

	Full sample		1963-1986		1987-2010	
	$1/\alpha$	λ^{KJ}	$1/\alpha$	λ^{KJ}	$1/\alpha$	λ^{KJ}
$1/\alpha$	0.4474	-	0.4955	-	0.3291	-
λ^{KJ}	-	0.3862	-	0.4025	-	0.1279
Log dividend price ratio	0.4272	0.3908	0.4368	0.4135	0.3024	0.1534
Log dividend yield	0.4427	0.3841	0.4811	0.4182	0.2797	0.1624
Log earning price ratio	0.3595	0.3994	0.1970	0.4373	0.3448	0.2047
Log smooth earning price ratio	0.3822	0.4086	0.2740	0.4509	0.3083	0.1457
Log dividend payout ratio	0.4158	0.4055	0.3799	0.4560	0.3474	0.1393
Book to market ratio	0.4094	0.4152	0.3766	0.4760	0.3173	0.1726
T-Bill rate	0.4577	0.3868	0.5169	0.3458	0.3284	0.0032
Term spread	0.4526	0.3943	0.5153	0.4291	0.3127	0.1122
Default yield spread	0.4565	0.3922	0.5193	0.4043	0.3589	0.1636
Default return spread	0.4682	0.4085	0.4960	0.4043	0.4173	0.2343
Inflation	0.4556	0.4073	0.5229	0.4451	0.3285	0.1172
Log total net payout yield	0.4421	0.3620	0.4731	0.3581	0.3443	0.1768

The table reports the out-of-sample R^2_{OoS} of Equation 20. The first two columns are the out-of-sample R^2_{OoS} calculated using the full sample data associated with PN tail risk and KJ's Hill tail risk. The R^2_{OoS} by sub-samples,1963-1986 and 1987-2010 are reported in the third to sixth columns.

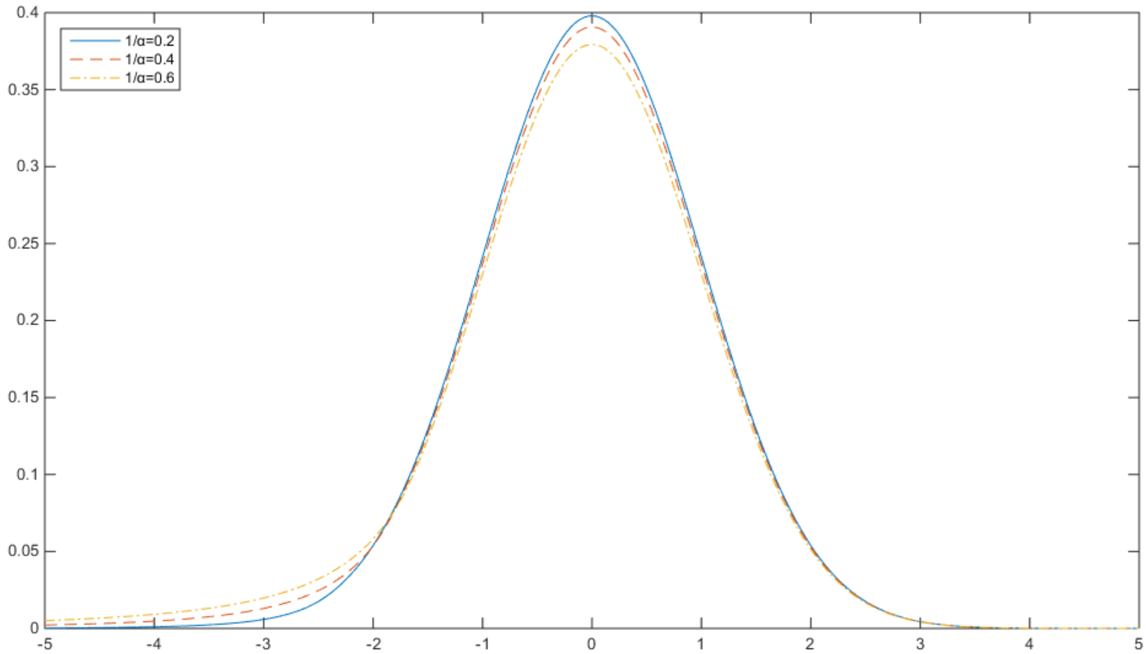


Figure 1: The density plots of the composite PN distribution with different α.

Plotted are the density of composite PN distributions with parameters $1/\alpha = 0.2$, $1/\alpha = 0.4$ and $1/\alpha = 0.6$, setting $\mu = 0$ and $\sigma = 1$.

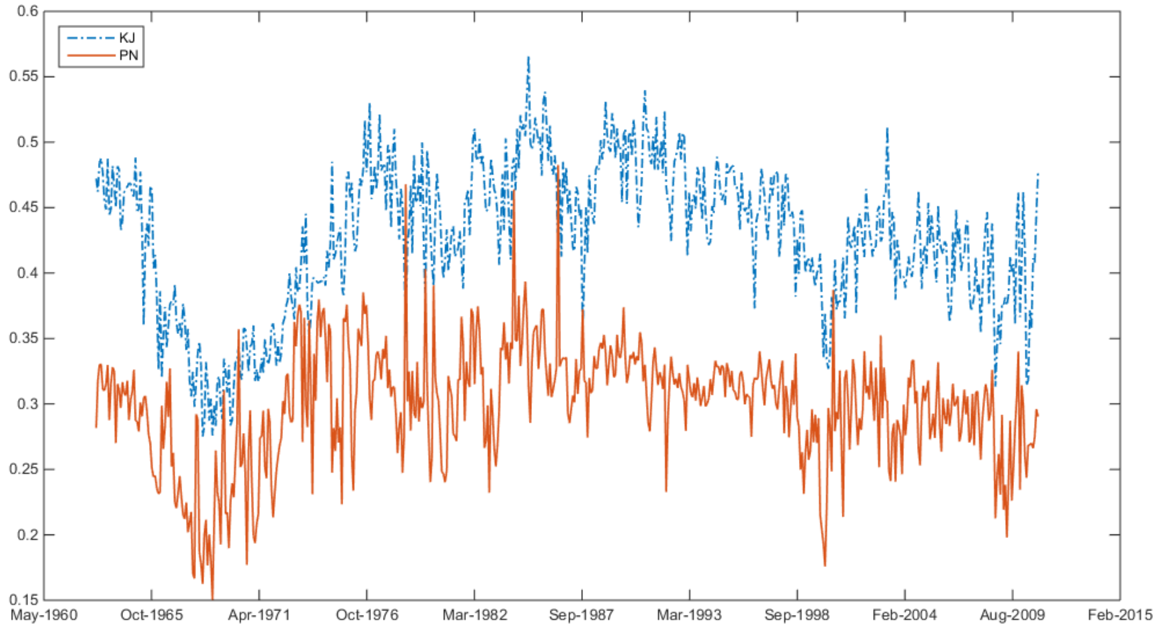


Figure 2: The KJ's Hill tail risk and the PN tail risk from January 1963 to December 2010.

Plotted are the KJ's Hill tail risk and the PN tail risk time series during a period from January 1963 to December 2010. KJ's Hill tail risk is calculated each month by using fifth percentile as threshold. The PN tail risk is a parameter in composite PN model and is estimated by conducting maximum likelihood estimation.

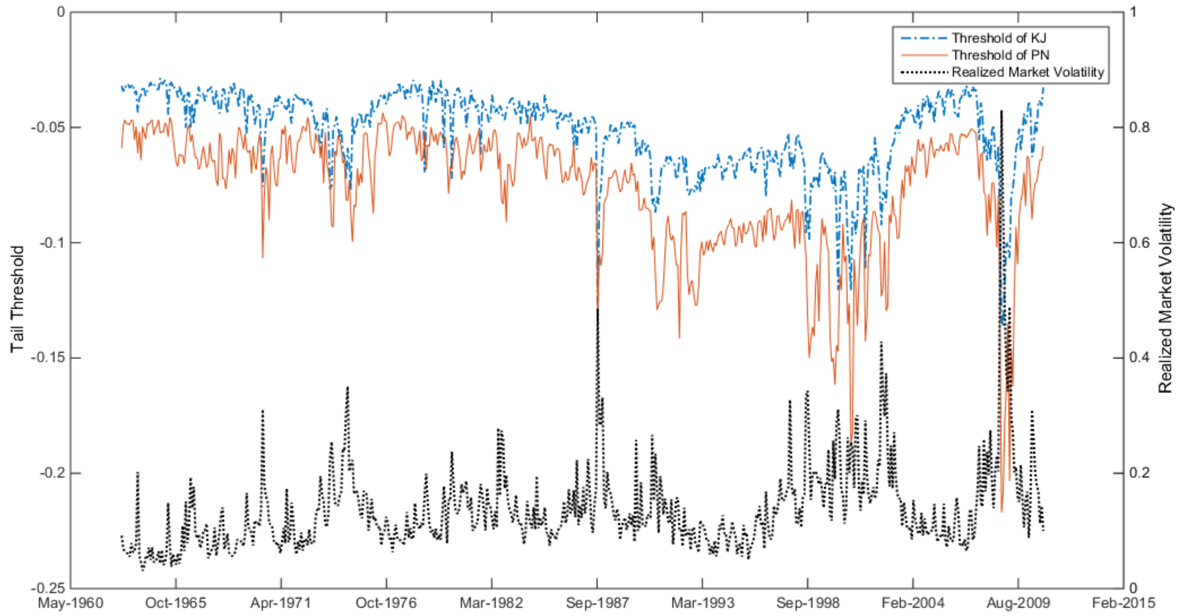
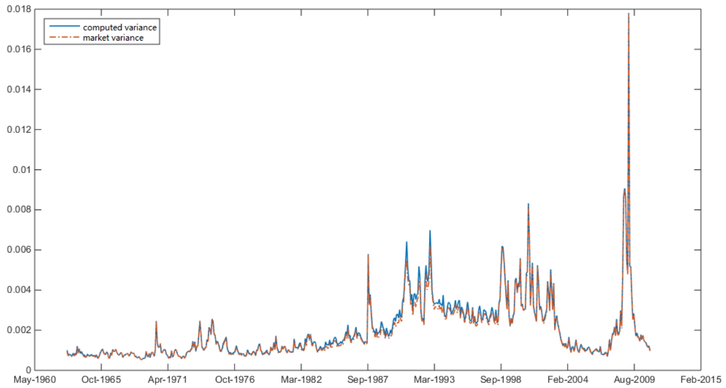
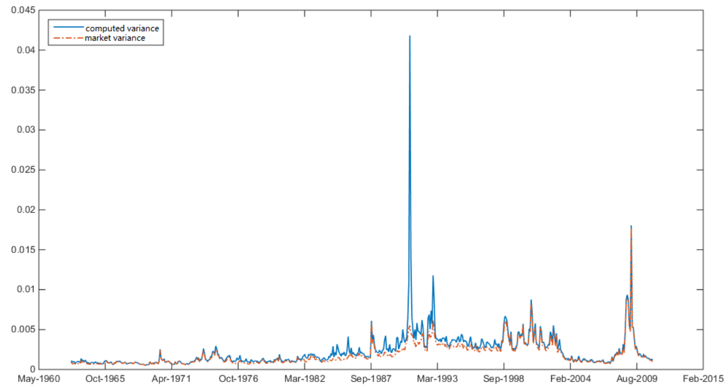


Figure 3: Threshold series of the KJ's Hill tail risk and the PN tail risk from January 1963 to December 2010.

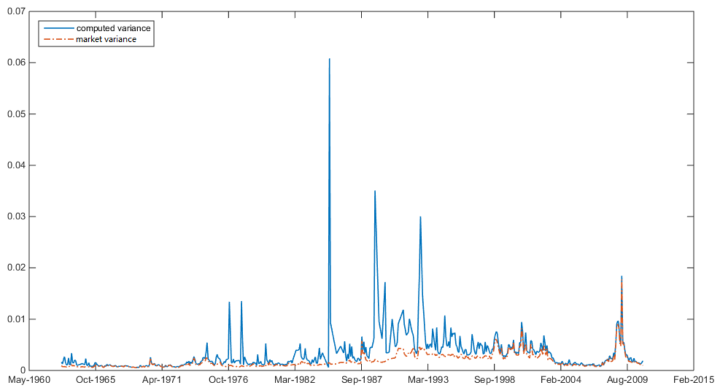
Plotted are the threshold series associated with the KJ's Hill tail risk and the PN model during the period from January 1963 to December 2010. Fifth percentile is used as threshold of KJ's Hill estimators. Threshold of composite PN model is estimated by the maximum likelihood method.



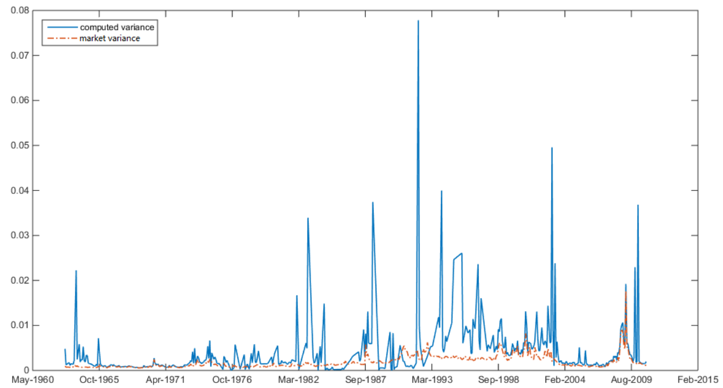
(a) 1% Threshold



(b) 3% Threshold



(c) 5% Threshold



(d) 7% Threshold

Figure 4: Total variance computed using KJ's method corresponding to four thresholds.

The plot exhibit the total variance calculated by using KJ's method with four different thresholds θ which corresponds to the 1st, 3rd, 5th and 7th percentile of the data.

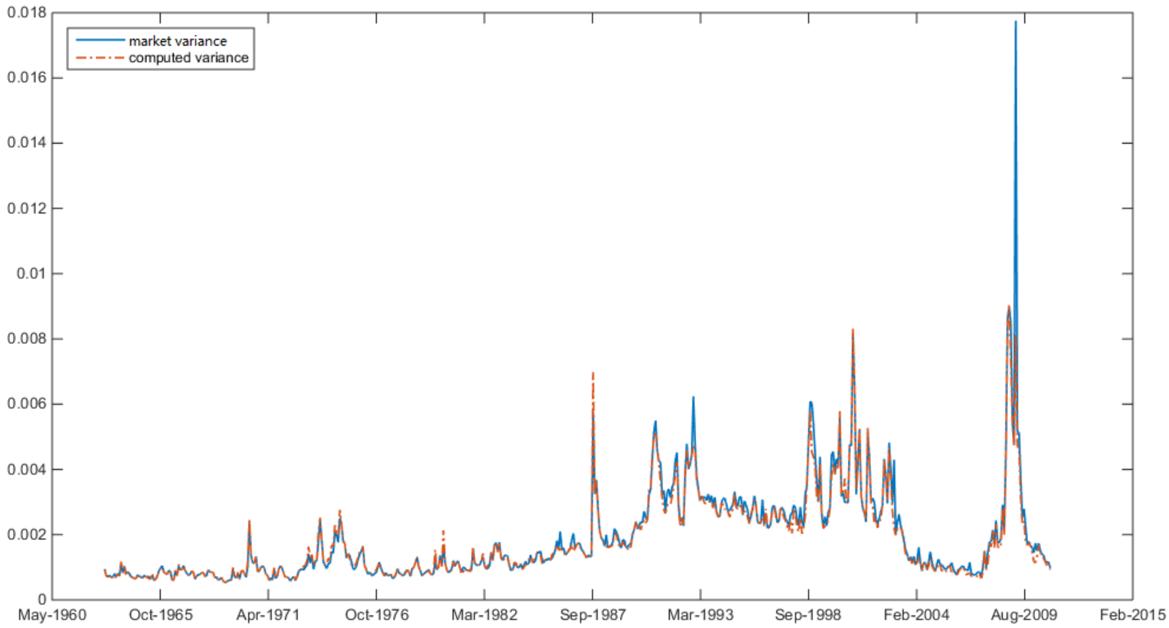


Figure 5: Total variance computed using the PN model.

Plotted is the total variance Var_{PN} computed from Equation 15 by using the composite PN model. Also plotted is the true variance calculated from original data.

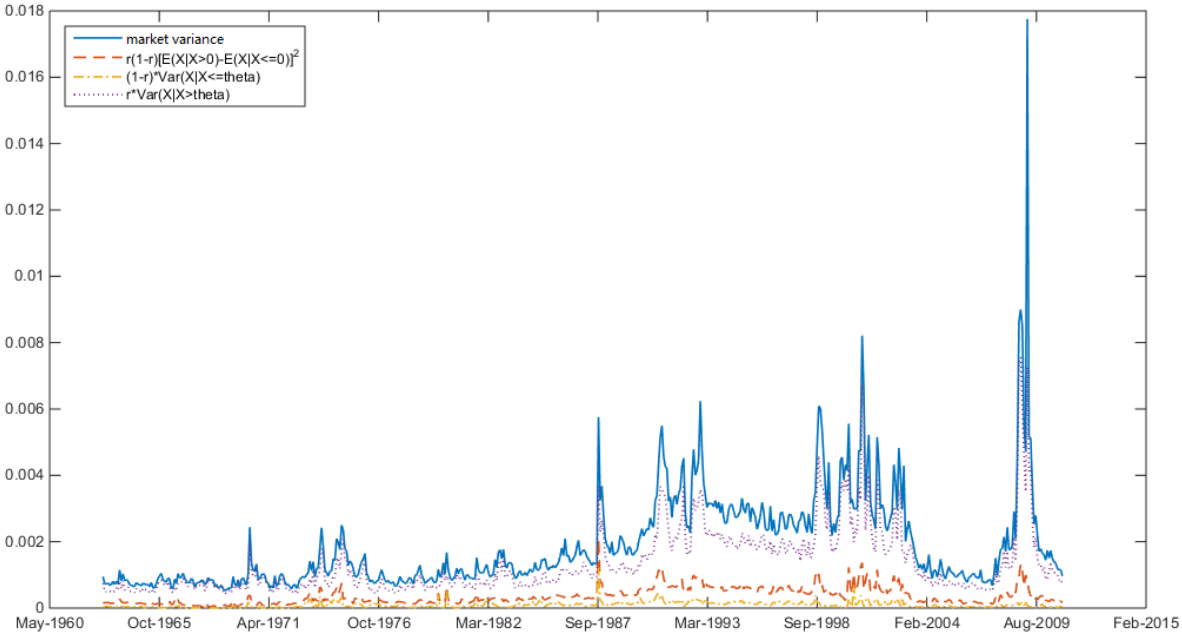
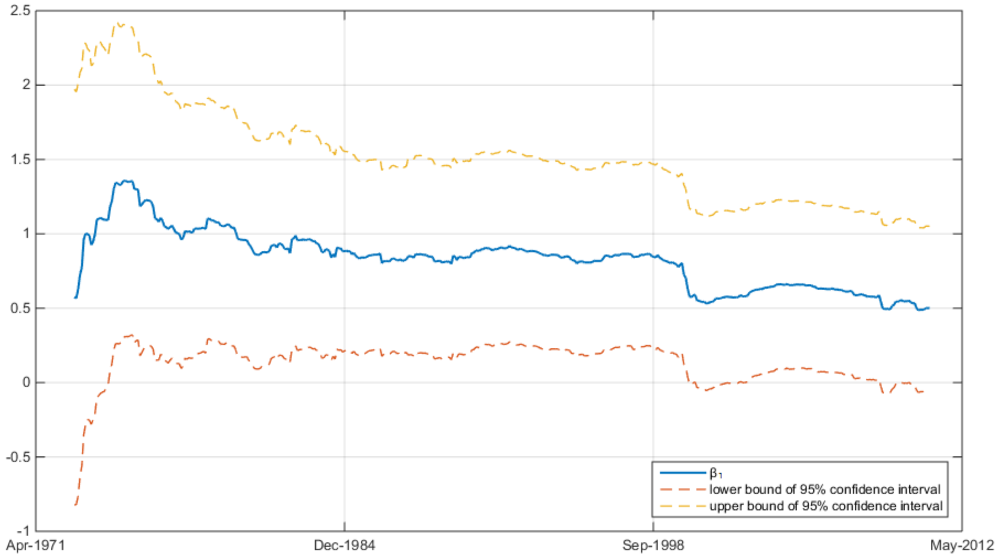
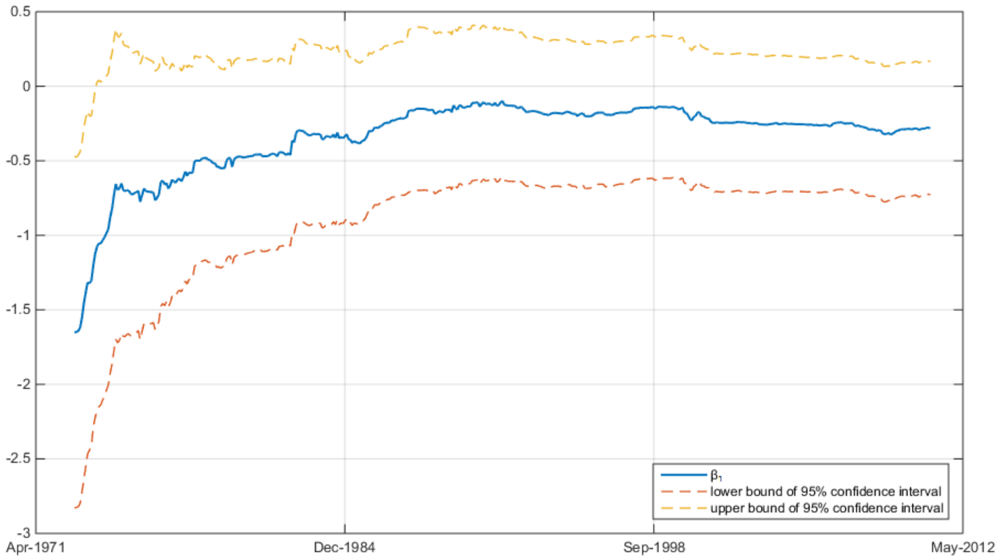


Figure 6: The three part decompositions of the total variance from the PN model.

Plotted are the three part decompositions of the total volitility, which are $rVar(X|X > \theta)$, variance of the truncated normal distribution multiplied by its weight r, $(1-r)Var(X|X \leq \theta)$, variance of the pareto distribution multiplied by $(1-r)$, and a component $r(1-r)\left[E(X|X > \theta) - E(X|X \leq \theta)\right]^2$. Also plotted is the market variance calculated from pooled monthly data.



(a) PN tail risk



(b) KJ's Hill tail risk

Figure 7: Recursively estimated coefficient β_1 and its 95% confidence interval in volatility forecasting regression.

This figure demonstrates the recursively estimated coefficients β_1 and the confidence intervals at 95% level of the out-of-sample regression of $LVOL_t$ on the PN tail risk and the KJ's Hill tail risk from January 1973 to December 2010.