Generalized extreme value distribution for value at risk analysis on gold price

N Pratiwi¹, C Iswahyudi² and R I Safitri¹

¹ Department of Statistics, Institut Sains & Teknologi AKPRIND Yogyakarta, Yogyakarta, Indonesia
² Department of Informatics Engineering, Institut Sains & Teknologi AKPRIND Yogyakarta, Yogyakarta, Indonesia
E-mail: novianapratini@akprind.ac.id

Abstract. Gold investment has some benefits, it’s not only safe in value but also easily stored, durable, and the rate of return is relatively stable even several years tends to rise. Gold prices give us positive and negative side depending on the different event. To managing the negative side of holding gold, the investor has to measure the risk of gold at a given period. The most popular measurement of gold investment risk is Value at Risk (VaR). Determination of inappropriate methods will make VaR calculations inaccurate. The gold’s price from year to year is suspected to have fat tail distributed (heavy tail), the Extreme Value Theory (EVT) is considered as precise methods to find VaR. In this study, the Generalized Extreme Value (GEV) approach used in EVT estimated. GEV distribution identifies extreme values based on the maximum value of each block. Test results show that monthly block usage yield VaR value is more accurate than 0.899%. It means that for one coming period with a 95% confidence level the maximum loss that investors may experience is 0.899% of the total investment.

1. Introduction
In Indonesia, a lot has been said about the benefit of gold as an investment. 60% people put their funds in gold investments and no more than 5% in mutual fund investments. Financial planners argue that gold could be used to protect against inflation. Gold can enhance portfolio performance and complement alternative asset allocation for diversification [1]. Gold serves two main purposes: it protects – even improves – purchasing power, and it helps manage risk while there is an abundant opinion about the rationale for investing gold or what measures should be used to assess its effectiveness, its role in a portfolio (e.g., inflation protection, currency hedging, and safe haven).

Gold prices in Indonesia have increased as can be seen in figure 1. Gold prices from 1994 to 1997 tend not to experience movement. But after 1997 gold prices began to rise to 2015 although in the course of gold prices experienced fluctuating ups and downs in the long term, gold prices tended to rise.
Research on gold investment is mostly done by researchers in Indonesia and around the world. World gold prices are always increasing, and over the past five years, gold has become the highest investment demand compared to investments in industry or capital markets [2]. Gold can enhance portfolio performance and complement alternative asset allocation for diversification. Investment in gold is a good hedge against inflation in Pakistan [3].

Some different event makes gold prices have experienced both positive and negative side. Maximum gains in gold investment can be achieved when investor engage in risk management. The investor has to measure gold price risk for risk managing. Djuhar as Business Director II of PT Pegadaian (Persero) said that behind the benefits, gold investment also contains many risks, such as obtaining fake gold, potential loss and crime, also price fluctuations.

Gold price’s volatility is the emergence of market risk. To control and reduce the investment risk, investors can measure the volatility or risk level of the assets owned [4]. This risk measure sought will be used to control risk management. During two decades Value at Risk (VaR) became the most popular risk measurement technique in finance. VaR measures the maximum loss in value of assets/portfolio over a predetermined time for given internal confidence.

VaR models are useful only if they accurately predict future risk. VaR used to answer the question of “how much investor can lose during the investment period t with confidence level \( \alpha \)”. Time horizon, the confidence level and loss amount are important variables to calculate VaR [5]. The incorrect method will cause the calculation of VaR inaccurate. The gold price fluctuation is expected to approach extreme values. Value at Risk calculations in extreme condition can be measured by extreme value theory approach.

Extreme conditions occur, the Risk Metrics Technique cannot accurately calculate VaR [6]. Extreme Value theory is suitable for an extreme event in financial markets [7]. Chaithep [2] also uses extreme value theory on VaR search at American gold prices. Generalized Extreme Value (GEV) is a method used to estimate extreme values by identifying extreme values based on maximum values. This paper focuses on gold price risk and tail distribution of extreme event in the Indonesian financial market.

2. Methods
Value at risk with the Generalized Extreme Value Theory approach will be analyzed using the following steps:

2.1. Extreme Value identify with Block Maxima
There are two models used to identify extreme values in general, traditional and modern models [8]. Block Maxima is a traditional model which presented by Fisher and Tippett in 1928 that is used to identify extreme value based on the maximum value of observation data grouped by a certain period. In this method, data is divided into blocks within a certain period, for example, month, quarter, semester, or year.
On Generalized Extreme Value distribution, Block Maxima Modeling is used as an approach to model the maximum value of a series of random variables. For example, \( X_1, X_2, \ldots, X_{13} \) is a series of random variables, as shown in figure 2. Figure 2 shows that \( X_1, X_6, X_8, \) and \( X_{10} \) are extreme values on four observation blocks.

The limit for Block Maxima stated with \( M_n \) is \( \max(X_1, X_2, \ldots, X_s) \) which \( s \) is sample size in each block and \( n \) is number of block or the maximum number of values. In modeling the maximum value of a random variable, the extreme value theory resembles the central limit theorem. The modeling is given by the following theorem:

\[
M_n = \max(X_1, X_2, \ldots, X_s)
\]

\[
P(M_n \leq x) = P(X_1 \leq x, X_2 \leq x, \ldots, X_s \leq x) = F_n(x)
\]

\( x \in \mathbb{R}, n \in \mathbb{N} \) (1)

If there is a constant \( c_n > 0 \) and \( d_n \in \mathbb{R} \) converges in the distribution of extreme value \( H \), then:

\[
\lim_{n \to \infty} P\left( \frac{M_n - d_n}{c_n} \leq x \right) = \lim_{n \to \infty} F_n\left( c_n x + d_n \leq x \right) = H(x)
\]

With \( H \) is a non-degenerate distribution. The \( H \) distribution function will follow one of the three basic distribution of extremes values. The three types of distributions are Gumbel, Frechet, and Weibull distribution. The distribution equations follow:

Frechet: \( \Phi_\alpha(x) = \begin{cases} 0 & \text{if } x \leq 0, \alpha > 0 \\ e^{-\alpha x} & \text{if } x > 0, \alpha > 0 \end{cases} \) (3)

Weibull: \( \psi_\alpha(x) = \begin{cases} e^{-(-\alpha x)^\alpha} & \text{if } x \leq 0, \alpha > 0 \\ 1 & \text{if } x > 0, \alpha > 0 \end{cases} \) (4)

Gumbel: \( \Lambda(x) = e^{-e^{-x}}, x \in \mathbb{R} \) (5)

By taking \( \xi = \alpha^{-1} \) for Frechet distribution, \( \xi = -\alpha \) for Weibull and \( \xi = 0 \) for Gumbel distribution, the cumulative distribution (CDF) of the three distribution is Generalized Extreme Value (GEV) distribution as follow:

\[
H_\xi(x) = \begin{cases} e^{-(1+\xi x)^{-1/\xi}} & \text{if } \xi \neq 0 \\ e^{-e^x} & \text{if } \xi = 0 \end{cases}
\]

(6)

With \( x \) fulfill \( 1+\xi x > 0 \). If equation (6) is standardized, then the formula involving the parameters \( \xi, \mu \) and \( \sigma \) can be written in the form of:

\[
H_{\xi,\mu,\sigma}(x) = H_\xi\left( \frac{x-\mu}{\sigma} \right)
\]

(7)
So that cumulative distribution function of GEV with the parameters $\xi$, $\mu$ and $\sigma$ become:

$$H_\xi(x) = \begin{cases} e^{-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}} & \text{if } \xi \neq 0 \\ e^{-\exp\left(-\frac{x-\mu}{\sigma}\right)} & \text{if } \xi = 0 \end{cases}$$

(8)

The theory by Fisher and Tippett and Gnedenko prove that mathematically under certain conditions the distribution of the Gumbel, Frechet, and Weibull can approach the distribution of extreme values for sequence or random variables. The amalgamation of the Gumbel, Frechet and Weibull families is known as the Generalized Extreme Value (GEV).

2.2 Estimate Generalized Extreme Value (GEV)

The essence of the GEV distribution is data divided into blocks within a specified period. Each block period is determined by the maximum amount of data. Suppose there is a set of $n$ return $\{r_1, r_2, ..., r_n\}$. The minimum return from the set is $r_{(1)} = \min_{1 \leq j \leq n} \{r_j\}$ and maximum value is $r_{(n)} = \max_{1 \leq j \leq n} \{r_j\}$. The extreme value is focused on the minimum value because the minimum value is very relevant to VaR calculations [9].

Assume return $r_i$ is an independent sequence with the cumulative distribution function $F(x)$ with return range $[l, u]$. For log return $l = -\infty$ and $u = \infty$, cumulative distribution function $r_{(1)}$ is stated as $F_{n,1}(x)$ ie:

$$F_{n,1}(x) = \Pr[r_{(1)} \leq x] = 1 - \Pr[r_{(1)} > x]$$

$$= 1 - \prod_{j=1}^{n} \Pr(r_j > x)$$

$$= 1 - \prod_{j=1}^{n} \left[1 - \Pr(r_j \leq x)\right]$$

$$= 1 - \prod_{j=1}^{n} [1 - F(x)]$$

$$= 1 - [1 - F(x)]^n$$

(9)

In practice, the cumulative distribution function $F(x)$ of $r_{(1)}$ is unknown, so $F_{n,1}(x)$ defined as $F_{n,1}(x) \to 0$ if $x \leq 1$ and $F_{n,1}(x) \to 1$ if $x > 1$.

The extreme value theory has two parameters $\{\mu_n\}$ and $\{\sigma_n\}$ with $\sigma_n > 0$ distributed $r_{(t^*)} = \frac{r_{(t)} - \mu_n}{\sigma_n}$ converging on the non-generated distribution with $n$ towards infinity. Series $\{\mu_n\}$ are location factor and $\{\sigma_n\}$ are scale factor. Based on the assumption, the minimum normalized limit distribution of $r_{(t^*)}$ as follows

$$F^*(x) = \begin{cases} 1 - \exp\left[-(1 + \xi x)^{\frac{1}{\xi}}\right] & \text{if } \xi \neq 0 \\ 1 - \exp[-\exp(x)] & \text{if } \xi = 0 \end{cases}$$

(10)

For $x < -1/\xi$ if $\xi < 0$ and $x > -1/\xi$ if $\xi > 0$ where $*$ indicates the minimum. Parameter $\xi$ shows the shape parameter that fulfill the tail distribution boundary and $\alpha = -1/\xi$ is called tail index of the extreme statistical distribution [2].

Equation (10) is called Generalized Extreme Value Distribution (GEV) by Jenkison, 1955. Probability density function (pdf) of equation (10) can be defined by:

$$f^*(x) = \begin{cases} (1 + \xi x)^{\frac{1}{\xi}} \exp\left[-(1 + \xi x)^{1/\xi}\right] & \text{if } \xi \neq 0 \\ \exp[x - \exp(x)] & \text{if } \xi = 0 \end{cases}$$

(11)

where $-\infty < x < \infty$ for $\xi = 0$, $x < -1$ for $\xi < 0$, and $x > -1/\xi$ for $\xi > 0$. 

4
Maximum likelihood method will be used to estimate extreme value parameters. Assume that minimum sub-period \( \{r_{ni}\} \) is distributed extreme value, so that pdf of \( x_i = (r_{ni} - \mu_n) / \sigma_n \) given equation (11) can be transformed as follows:

\[
    f(r_{ni}) = \begin{cases} 
    \left( \frac{1}{\sigma_n} \right) \left( 1 + \frac{\xi_n(r_{ni} - \mu_n)}{\sigma_n} \right)^{- \frac{1}{\xi}} - 1 \exp \left[ - \left( 1 + \frac{\xi_n(r_{ni} - \mu_n)}{\sigma_n} \right)^{- \frac{1}{\xi}} \right] & \text{if } \xi \neq 0 \\
    \frac{1}{\sigma_n} \exp \left[ \frac{r_{ni} - \mu_n}{\sigma_n} - \exp \left( \frac{r_{ni} - \mu_n}{\sigma_n} \right) \right] & \text{if } \xi = 0
    \end{cases}
\]  

(12)

Likelihood function of minimum sub-period is

\[
    L(r_{n1}, \ldots, r_{ng} | \xi_n, \sigma_n, \mu_n) = \prod_{i=1}^{g} f(r_{ni})
\]  

(13)

2.3. Calculate Value at Risk with GEV estimations

Log-likelihood from equation (13) with conditions \( k \neq 0 \) is as follows:

\[
    \ln L = -g \ln \sigma_n - \left( 1 + \frac{1}{\xi_n} \right) \sum_{i=1}^{g} \ln \left( 1 + \frac{\xi_n(r_{ni} - \mu_n)}{\sigma_n} \right) - \sum_{i=1}^{g} \left( 1 + \frac{\xi_n(r_{ni} - \mu_n)}{\sigma_n} \right)^{- \frac{1}{\xi_n}}
\]  

(14)

Maximum likelihood is used to find VaR. Chaithep [2] formulated VaR with GEV approach as follows:

\[
    VaR = \mu_n - \frac{\sigma_n}{\xi_n} \{ 1 - [-n \ln(1 - p)]^{-1} \}
\]  

(15)

3. Results and discussions

3.1. Identify of the heavy tail and extreme values

This research uses secondary daily data from publication on the official website http://harga-emas.com from January 2, 2015, to December 31, 2017, with total 875 data.

| Table 1. Descriptive statistics of return |
|------------------------------------------|
| Mean 0.000202907 |
| Median 0 |
| Standard Deviation 0.006855404 |
| Skewness 0.1850253 |
| Kurtosis 3.128958 |
| Minimum -0.030776360 |
| Maximum 0.033989270 |
| n 875 |

Based on table 1, it can be seen that the average return on investment in gold is 0.000202907 with median 0, and standard deviation 0.006855404. Skewness has a positive value, i.e. 0.18502 which means that distribution of gold returns has a long right tail. While the value of high enough kurtosis is more than three (3.128958) which means that returns distribution is in leptokurtic type, that is the middle part of the data has a more pointed peak. Return volatility is seen based on the range or distance of maximum and minimum values that are relatively far enough so that the gold price fluctuates. Thus it can be said that returns have nature financial data, namely leptokurtic patterns and volatility.

The positive skewness and kurtosis which more than 3, indicate that data have heavy tail distribution. It can also be seen using a Q-Q plot as shown in figure 3.
Normal QQ Plot on figure 3 shows that the data aren’t in a straight line. It means that data don’t have the same distribution as the comparison distribution is a normal distribution. It has an “S” pattern which is down on the left and up on the right, so it can be said that data distribution is a more heavy tail than a normal distribution.

3.2. Blockmaxima
In this research, Blockmaxima methods are used to identify the extreme value based on the maximum value of the observation data based on a certain period. Weekly and monthly blocks were used on this research. With a total of 875 data, the block formed during the weekly period or 5 working days as many as 175 blocks, and for a monthly block with 21 working days as many as 42 blocks. Plot blockmaxima ad Histogram for each block (weekly and monthly) can be seen on figure 4a, 4b, 4c, and 4b.
Figure 4c. Histogram weekly blockmaxima

Figure 4d. Histogram monthly blockmaxima

Figure 4a, 4b, 4c, and 4d shows that the data is not normally distributed and the data contains a heavy tail.

3.3. GEV estimates and value at risk parameters

Blockmaxima result is used to estimate GEV parameters. Maximization of likelihood (equation 13) and that parameter are used to estimated VaR (equation 15) for these data leas to estimate:

| Parameters         | Weekly (175 Block) | Monthly (42 Block) |
|--------------------|--------------------|--------------------|
| location ($\mu$)   | 0.004792435        | 0.009296257        |
| scale ($\delta$)   | 0.004064818        | 0.005362288        |
| Shape($\xi$)       | 0.127146795        | 0.071190809        |
| VaR 95%            | 0.003740450        | 0.007898273        |
| VaR 99%            | 0.002095505        | 0.00569618         |

Table 2 shows that the shape parameter ($\xi$) is rthe ight tail because $\xi$ is greater than 0, so it can be concluded that return approaches by Fr aechet distribution where the probability of extreme value occurrence is greater than Gumbel or Weibull distribution. Location parameter ($\mu$) states that the location of the data center point. Monthly location block (0.009296257) has greater value than a weekly block (0.004792435). It means that monthly block has the greatest probability for the occurrence of risk in extreme event.

The calculation results of parameters in VaR equation (equation 15), obtained results 0.374045% in a weekly block with 95% confidence level. It means that tomorrow’s loss will be 0.374045% in a day at the confident level of 95% in extreme condition. Suppose an investor invests in gold of 100 million rupiahs, then the maximum loss that may be obtained in the next day is 374,000 rupiahs. For weekly block, VaR with 99% confident level is smaller than 95% confident level. And this also applies to monthly block, VaR with 99% confident level is smaller than the 95% confident level. The smallest value on VaR is 0.2095505% from the weekly block with 99% confident level.

4. Conclusions

Value at risk (VaR) measures the maximum loss in value of an asset/portfolio over a predetermined time for a given confidence level. Many methods can be used to evaluate VaR of the assets returns depending on data distribution. Generalized extreme value (GEV) distribution is one of the distributions that often occurs in a financial market condition. So Generalized Extreme Value Distribution method can use to evaluate VaR is. GEV used blockmaxima methods to identify extreme value on data distribution.
Holding Gold is the most interesting investment type in Indonesia for two decades, and this paper focuses on Gold investment risk and estimate Value at Risk. Empirical result shows that Value at Risk of Gold return with Generalized Extreme Value approach is equal to 0.2095505% at the confidence level 99% and it means a maximum loss that might occur to investors next day if he puts investment in gold of 0.2%. Suppose an investor invests in gold of 100 million rupiahs, then the maximum loss that may be obtained in the next day is Rp.209,550.00.

Acknowledgments
This research is part of the Hibah Penelitian Dosen Pemula (PDP) from Kemenristek Dikti Funding 2018. We thank to Kemenristek Dikti for the fund provided, to IST AKPRIND Yogyakarta which has provided research facilities and infrastructure.

References
[1] Abraham S 2016 Commerce Spectrum 4 41.
[2] Chaithep et al. 2012 The Empirical Econometrics and Quantitative Economics Letters 1 151.
[3] Shahbaz M, Tahir M I, Ali I and Rehman I 2014 The North American Journal of Economics and Finance 28 190.
[4] Pratiwi N 2015 Jurnal Teknologi Technoscientia 8.
[5] Pratiwi N 2014 Prosiding, Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014, Yogyakarta.
[6] Danielsson J 2000 Value-At-Risk and Extreme Return (London: Tinbergen Institute and Erasmus University Rotterdam).
[7] Neftci S N 2000 The Journal of Derivative 1-15.
[8] McNeil A, Frey R and Embrechts P 2005 Quantitative Risk Management (Princeton: Princeton University Press).
[9] Tsay R S 2005 Analysis of financial time series. Second Edition. Hoboken (New Jersey: John Wiley & Sons, Inc.)