A Probabilistic Stochastic Income Distribution Model of Coal Mining Industry

Muhammad Kamran¹
¹Bandung Institute of Technology, Bandung Indonesia

Abstract
Coal mining is a profitable enterprise. It creates job opportunities, generates revenue, and attract the foreign investment of a country. However, coal mining faces some challenges. To address the allocation of capital related to coal mining, an approach has been made to improve the impact of coal mining industry on the economy of one of Indonesia most important coal producing region, south Kalimantan. A total of seven households of large-scale and small-scale mining are analyzed in the study. Various copula-based prediction probability models were established, and the exponential distribution function of household income distribution was obtained with maximum range by utilizing the application of Monte Carlo simulation. Moreover, this research spells out the importance of income distribution of various household dynamics which will help the policy maker in economic analysis and financial decisions related to various household categories.

Abstrak
1. Introduction
Energy plays an important role for humanity to continue the economic growth and maintain high standard of living. International Energy Agency (IEA 2013; Nakicenovic 2007; Shahid and Jamal 2011) reported that the world will need 50% of more energy by 2030 than today. Around 81% of energy demand is fulfilled from the fossil fuel. Coal is an ordinary material transpire in two divergent stages, one is geochemical and the other is biochemical (Grainger and Gibson 2012). Coal is one of the most important, cheap, and easy source of world energy resources and its use is likely to be increase four times by 2020 (Mimuroto 2002). Although coal business plays profitable role for both individuals and businesses but the benefits of these enterprises to the region are uncertain (Roenker 2006). Indonesia's GDP per capita has steadily increased from US$ 665 in the year 1965 to US$ 3974 in the year of 2016 (World-bank 2017). Urbanization growth rate in Indonesia has soared more than four percent during last few years (UNDP 2017). The transformation of coal to electricity and then consumption of coal can increase the demand of coal. The usage of coal is more charismatic in various countries with adequate low-cost reserve located near to the consuming centers (Gordon and Gordon 1987). The relationship between the real consumption, real GDP and real gross fixed capital formation and the labor force has been established. The concomitant of real gross fixed capital formation and the labor force is positively correlated whereas coefficient for coal consumption is negatively correlated showing that the results are not likely to occur randomly or by chance (Apergis and Payne 2010). Hence there is a need for a proper distribution model of income distribution of various households. An approach has been made to categorize the probabilistic income distribution effects of coal mining on economic development of South Kalimantan Province, Indonesia and how the benefits of coal mines are shared by various communities within and outside the region. First, the data of various households of large-scale and small-scale mining is collected. Then Monte Carlo based simulations are employed to predict the model. Finally, the predictive performance of each model is comprehensively correlated and investigated.

2. Literature Review and Hypothesis
Following the growth, energy consumption has been increased, the residential sector is one of the largest consumers of energy in Indonesia. In South Kalimantan, difference of the welfare and incomes can be seen among various communities living nearby the mines (Adaro 2002). Several studies at home and abroad investigated the linkage between coal, and it effect on economic growth. Empirical analysis with modern data panel techniques implores that the coal consumption–GDP association is bidirectional (Li and Leung 2012). A relationship over almost four decades has been built between coal consumption and economics growth of Pakistan (Kumar and Shahbaz 2012). By enhancing the entirely performance and productivity of coal industry will continue preserve an important function in in perpetuation of sustainable growth (Bhattacharya et al. 2015). Coal industry has a lasting legacy of providing business and careers in remote areas, but it also assists with improving living standards from weaker long-term economic growth and high poverty (Betz et al 2015). The policy makers and the researchers must have to work side by side with the households to enhance a better awareness and insight to the crucial issue faced by them and to solve their problems in a timely manner (Huth et al 2018). Although these studies link the coal, and it effect on economic growth but the effect of income distribution on various household remains unchanged. This study elucidates the statistical characteristics of various income distribution probability of large-scale and small-scale coal mining respectively.

Moreover, this paper discusses following aspects of income distribution effects of coal mining including establishment of the range of various input parameters.
1. Construction of various probabilistic model by checking its performance employing to the input and output datasets.

2. Comparison of various simulation models to predict income distribution effect incorporating other input parameters.

The theoretical framework of proposed research process is shown in Fig. 1.

Figure 1. Theoretical framework of the study
Monte Carlo Simulation

Various researchers have utilized Monte Carlo (MC) simulation dealing with scarce information in different reliability analysis (Bird 1981; Beer et al. 2013; Dousilet et al. 1985; Nettavali et al. 1989; Ghasemi et al. 2010; Zhou 2020; Mahdiyar et al. 2020). This method is used as a computerized computation for identification of numerous problems such as the impact of risk and uncertainty of various models, including appraisal, estimation, evaluation, and prediction. To obtain the probabilistic evaluation Monte Carlo simulation is operated on repetitive random sampling (Solver 2010; EPA 1997) Each model that predicts outcomes requires an investigative set of hypotheses associated with real world issues and evaluation of the expected values build on the data (Solver 2010).

MC simulation has two important purposes: first is the quantitative testing of variation and uncertainty and the second is parameters investigation influencing uncertainty, variation and their proportion. (EPA 1997). In contrast to the conventional and traditional method MC simulation use a range of estimated values as an input and then the system returns a range of output values for a comprehensive data. Hence, a more realistic sketch of a simulation model can be generated in MC simulation. In this method, a random value is taken for each input based on the range of the outputs and then the output datasets are estimated based on these random values. In last the outcome of MC simulation is noted. This process is continuously repeated by using different random numbers. The simulation process is repeated 1000 times. After assessment and evaluation of simulation process, several consequences are generated as an output, which can be applicable to express the product (Ghasemi et al. 2012; Armaghani et al. 2006; Mahdiyar et al. 2017; Bianchini and Hewage 2012; Dunn and Shultis 2009; Kamran, 2021).

3. Data and Method
3.1 Setting range of Input variables

From the work of (Fatah 2008), two households categories income distribution including large-scale and small-scale mining effecting the coal mining industry are obtained. The income is received by two different factors of production (labor and capital) in all economic sectors in South Kalimantan. The distribution of the income in large-scale and small-scale mining in the overall dataset is given in Fig. 2 and Fig. 3. The complete income distribution data for large-scale and small-scale income distribution is presented in Table 1.

<table>
<thead>
<tr>
<th>Households</th>
<th>Large-scale mining</th>
<th>Small-scale mining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landless Framers</td>
<td>A1 3.62%</td>
<td>B1 3.48%</td>
</tr>
<tr>
<td>Small landowner farmers</td>
<td>A2 6.42%</td>
<td>B2 6.28%</td>
</tr>
<tr>
<td>Large landowner farmers</td>
<td>A3 12.84%</td>
<td>B3 12.49%</td>
</tr>
<tr>
<td>Low income non-farmers</td>
<td>A4 17.51%</td>
<td>B4 16.95%</td>
</tr>
<tr>
<td>Middle income nonfarmers</td>
<td>A5 15.84%</td>
<td>B5 15.35%</td>
</tr>
<tr>
<td>High income non-farmers</td>
<td>A6 21.35%</td>
<td>B6 20.86%</td>
</tr>
<tr>
<td>Very high income nonfarmers</td>
<td>A7 22.40%</td>
<td>B7 22.88%</td>
</tr>
</tbody>
</table>
The prediction probabilistic models for households income distribution model of coal mining industry is constructed based on large-scale and small-scale mining. The box plot of each indicator for the large-scale mining and small-scale mining is shown in Fig. 4 and Fig. 5 respectively. Except low income non-farmers, the household is positively correlated with each indicator. The larger the indicators values, the higher is the income effect of mining industry. Therefore, the effect of all indicators needs to be incorporated for better model.
3.2 Research Method

The prediction probabilistic models for income distribution model of coal mining industry is constructed for two households income categories including large-scale and small-scale mining. Each category is based on seven indicators i.e. landless farmers, small land-owner farmers, large land-owner farmers, low income non-farmers, middle income non-farmers, high income nonfarmers, very high income non-farmers represented by A₁, A₂, A₃, A₄, A₅, A₆, A₇ in large-scale mining and B₁, B₂, B₃, B₄, B₅, B₆, B₇ in small-scale mining respectively. The calculation model for households category is given as

$$H = t(C)$$  \hspace{1cm} (I)
Where \( H \) is the households category estimation, and \( t(C) \) is the calculation model of households prediction. Given that \( C \) is a random variable set, \( V \) is also a random variable. Therefore, the estimation probability \( P_f \) is given as

\[
P_f = P(V \leq V_{cri})
\]

(2)

\[
\int_{-\infty}^{+\infty} I(t(C) \leq V_{cri})f_a(C)dx
\]

(3)

where

\[
I(t(C) \leq C_{cri}) = \begin{cases} 
1 & t(C) \leq V_{cri} \\
1 & t(C) > V_{cri}
\end{cases}
\]

(4)

Where \( B_{cri} \) is the critical prediction value households estimation.

Based on Equation 1, it is very challenging to perform direct combination. Therefore, probability \( P_f \) to predict households estimation of Equation 1 can also be given by

\[
P_f = \frac{1}{N} \sum_{i=1}^{N} I(t(C) \leq V_{i,cri})
\]

(5)

Where \( N \) is the total number of samplings of the Monte Carlo simulation, \( C_i \) is the i-th parameter set of \( C \), and \( V_{i,cri} \) is the \( V_{cri} \) corresponding to the \( C_i \).

Therefore, as long as the above Equation 5 is obtained, the probability \( P_f \) of households category estimation can be obtained. Following are the calculation steps:

1. The parameter values of the specified number of times are simulated from the multidimensional joint probability distribution function of seven indicators in each large-scale and small-scale mining methods.
2. The obtained parameters are substituted into Equation 1 to compute the corresponding value for large-scale and small-scale mining methods.
3. Finally, the prediction values of each category are calculated by stochastic approach and the uncertainty and distribution function of each category is obtained. Then, the probability of prediction categories values was calculated using Equation 4.
4. It is also possible to compute the number of predictions less than the critical value households category and Equation 5 could be used to compute the occurrence probability of each households category estimation.

### 3.3 Prediction Models

In this section, the prediction models are carried to establish a proper relationship between different input indicators for various households categories. The performance of various prediction models is performed using squared coefficient of correlation index between the measured and predicted datasets (Rajesh and Sunoj 2019; Karkevandi-Talkhooncheh et al. 2018, Koopialipoor et al. 2019). The formulas for calculating this indicator is given by Equation 6.

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} (X-X')^2}{\sum_{i=1}^{n} (X-X'')^2}
\]

(6)

where \( X \) is the measured value, \( X' \) and \( X'' \) are predicted value and the mean value respectively. \( N \) represents the total number of datasets. Considering \( R^2 = 100\% \) for a model, the predictive model is defined as a perfect
Concerning determination of capability of the predicted models, the calculated performance indicator of the created dataset is specified.

4. Results and Discussion
To analyses the impact of coal mining on income distribution in South Kalimantan Province, various copula functions are utilized. The possible combinations are randomly selected in this approach from various defined distribution models. The relationship between input indicators has a great effect on Monte Carlo simulation modeling. Hence, to achieve improved MC modeling to simulate the income distribution range, the correlation of households between input variables of large-scale and small-scale mining should be considered as shown in Table 2. According to these figures, input data with different percentages are correlated to each other. If the aim is to attain a significant correlation instead of completely random sampling during input sampling, the combination cases presented in Table 2 should be considered as the goal of simulation.

Hence, the simulation process is given as: First data from coal mine is prepared and the initial analyses are done for them. Then the data was categorized by large-scale and small-scale mining methods. Probability distributions of seven parameters i.e. landless farmers, small land-owner farmers, large land-owner farmers, low income non-farmers, middle income nonfarmers, high income non-farmers, very high income non-farmers are inferred. The probability distribution from of the above parameters is unknown in advance, the distribution patterns commonly used, such as Normal distribution (normal for short), Lognormal distribution (lognormal for short) and exponential distribution (exponential for short short) are adopted. In the last, the overall distribution was tested. To obtain a statistical representation for income distribution range, various simulations were employed during the study. In this study, the simulation was carried out up to 1000 times to achieve high levels of simulation and the probability of this economical phenomenon. Fig. 6 indicates the normal probability distribution of households categories for long-scale and small-scale coal mining methods.

<table>
<thead>
<tr>
<th>Correlation of large-scale and small-scale mining methods</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="#" alt="Table 2" /></td>
</tr>
</tbody>
</table>

**Table 2. Correlation of large-scale and small-scale mining methods**

<table>
<thead>
<tr>
<th></th>
<th>Large-scale mining method</th>
<th>Small-scale mining method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large-scale mining method</td>
<td>1</td>
<td>0.997914</td>
</tr>
<tr>
<td>Small-scale mining method</td>
<td>0.997914</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 6. Normal income distribution probability of (a) large-scale coal mining (b) small-scale coal mining

The optimal multi-dimensional copula function is first resolved and then the joint probability distribution function is performed by Monte Carlo simulation sampling on large-scale and small-scale mining. The MCS simulation sampling is performed from the t copula function, and the $1,000 \times 7$-dimensional related nonnormal random variables generated are stored in the indicator's datasets. The generally applied multi-dimensional normal distribution model of various indicators are always not fit to get the optimal probability distribution model that characterized the relevant structures among various input indicators (Wu et al. 2019). Hence lognormal and exponential distribution model of above mentioned seven parameters for long-scale and small-scale mining coal methods are obtained. Fig. 7 indicates the lognormal probability distribution of households categories for long-scale and small-scale coal mining methods whereas Fig. 8 shows the exponential probability of households categories distribution for long-scale and small-scale coal mining methods.
Figure 7. Lognormal income distribution probability of (a) large-scale coal mining (b) small-scale coal mining

![Lognormal income distribution probability](image1)

Figure 8. Exponential income distribution probability of (a) large-scale coal mining (b) small-scale coal mining

![Exponential income distribution probability](image2)

Hence, simulation has been able to determine the income distribution with wide range and proper accuracy. The range of exponential large-scale mining method was obtained as the 0.01012% – 100.239% as smallest and the largest range of simulation by MC technique. This is while the actual measured values in the studied mine are in the range of 3% – 22%, indicating the acceptable level of simulation, and the simulation range is considered with simulation probability of 0.0001. Hence, all possible various situations are investigated by Monte Carlo simulation. This stochastic approach can be applied to probabilistic analysis income distribution models. Table 3. and Table 4. represents the statistical characteristics of various income distribution probability of large-scale and small-scale coal mining respectively.

Table 3. Statistical characteristics of various distribution for large-scale coal mining

<table>
<thead>
<tr>
<th>Distribution Models</th>
<th>Mean income distribution Mean</th>
<th>Mean income distribution Standard Deviation</th>
<th>Mean income distribution Variance</th>
<th>Mean income distribution Skewness</th>
<th>Mean income distribution Kurtosis</th>
<th>Mean income distribution Mode</th>
<th>Mean income distribution Minimum</th>
<th>Mean income distribution Maximum</th>
<th>Mean income distribution Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-normal income distribution</td>
<td>14.28</td>
<td>0.07149</td>
<td>0.00511</td>
<td>-0.012</td>
<td>-0.0103</td>
<td>14.2785</td>
<td>14.0246</td>
<td>14.5062</td>
<td>0.48154</td>
</tr>
<tr>
<td>Exponential income distribution</td>
<td>14.28</td>
<td>0.07145</td>
<td>0.00511</td>
<td>0.01489</td>
<td>-0.0336</td>
<td>14.2845</td>
<td>14.0495</td>
<td>14.5097</td>
<td>0.46027</td>
</tr>
<tr>
<td>Exponential income distribution</td>
<td>14.26</td>
<td>14.1771</td>
<td>200.991</td>
<td>1.89403</td>
<td>4.83291</td>
<td>0.26303</td>
<td>0.01012</td>
<td>100.239</td>
<td>100.229</td>
</tr>
</tbody>
</table>
Table 4. Statistical characteristics of various distribution for small-scale coal mining

<table>
<thead>
<tr>
<th>Distribution Models</th>
<th>Mean income distribution</th>
<th>Log-normal income distribution</th>
<th>Exponential income distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>14.04</td>
<td>14.04</td>
<td>13.56</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.71848</td>
<td>0.07155</td>
<td>12.6309</td>
</tr>
<tr>
<td>Variance</td>
<td>0.51622</td>
<td>0.00512</td>
<td>159.54</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.00821</td>
<td>-0.0025</td>
<td>1.46501</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.0113</td>
<td>0.0179</td>
<td>2.1811</td>
</tr>
<tr>
<td>Mode</td>
<td>13.986</td>
<td>14.0429</td>
<td>0.2130</td>
</tr>
<tr>
<td>Minimum</td>
<td>11.5845</td>
<td>13.7799</td>
<td>0.1477</td>
</tr>
<tr>
<td>Maximum</td>
<td>16.6006</td>
<td>14.2717</td>
<td>62.2757</td>
</tr>
<tr>
<td>Range</td>
<td>5.016</td>
<td>0.491</td>
<td>62.1228</td>
</tr>
</tbody>
</table>

The probability of best exponential probability large-scale and small-scale mining method is given by Equation 7. with 97% accuracy. The probability of occurrence is given in Equation 7. The optimal probability distribution function of the impact of income prediction X is exponential, and the probability distribution map is shown in Fig. 9. Thus, the mapping relationship between the impact of income prediction and the impact of income probability is established.

\[
P_f = 24.36 X - 25.413
\]  

Figure 9. Comparison of exponential large-scale and small-scale mining methods distributions
5. Conclusion
A novel income distribution probability model was proposed based on the statistical analysis of large-scale and small-scale mining methods. Seven quantifiable characteristic parameters were determined by the statistical analysis method, which obeyed normal, lognormal and exponential distributions, and the correlation among them was determined. The multi-dimensional joint probability distribution function of seven parameters was established based on the copula theory. The nonlinear mapping relationship between seven parameters and income distribution prediction ranges was established, and various copula income prediction probability model was constructed combined with MCS simulation, the optimal probability distribution function of the prediction the range of income distribution was exponential. The results of prediction probability model were validated, the maximum likelihood determined, and the occurrence probability was obtained, which was generally consistent with the actual situation. The model can help in realizing the effective fusion of multi-source characteristic parameter data of income distribution, and the probabilistic quantification of income prediction ranges results. The performance of the models signifies that the prediction probability obtained by this method could prove a valuable addition for policy makers in economic analysis and financial decisions related to various household categories. However, the prediction probability distribution function exponential was determined based on the data of existing income large-scale mining and small-scale mining cases, it should be combined with different engineering conditions to redetermine, re-analyze and re-examine the most optimal probability distribution function based on the method proposed to ensure the applicability and accuracy of the model with the richness of cases.

Acknowledgments
The author highly appreciates the constructive and precious comments of the anonymous reviewers. Their valuable ideas and scientific thoughts.

Funding

This is an academic research and no funding source is available.
References

Adaro (2002) Adaro’s community development fund provided to the community” (Dana Adaro diserahkan) 17 July 2002.


IEA 2013 Bioenergy Task 37 – Plant list <www.iea-biogas.net/_content/plant-list/plant-list.html> accessed


