Binary Logistic Regression Modeling on Net Income of Pagar Alam Coffee Farmers

Irmeilyana[1, *], Ngudiantoro[1] and Mukhlizar Nirwan Samsuri[1]

[1]*Department of Mathematics, Sriwijaya University,*
Indralaya, South-Sumatra, Indonesia
[*]*Corresponding Author: irmeilyana@unsri.ac.id*

Abstract

Pagar Alam Coffee is a Besemah coffee originating from the Smallholder Plantation in South Sumatra, Indonesia. The majority of Pagar Alam coffee farming is a hereditary business. Coffee farmers' income is very dependent on coffee production, production costs, and coffee prices. This study aims to obtain a probability model of Pagar Alam coffee farmers income based on the factors that influence it. The independent variables studied were the number of dependents, economic conditions, number of trees, age of trees, frequency of fertilizer used, frequency of pesticide used, production at harvest time, production outside harvest time, number of women workers outside the family, minimum price of coffee, maximum price of coffee, farmers' gross income, and land productivity. Modeling used binary logistic regression method on 179 respondents. There were three methods used, i.e. enter method, forward and backward methods. The model using enter method results the greatest prediction accuracy which is 87.7%. The factors that have a significant influence on the net income of Pagar Alam coffee farmers are gross income, land productivity, and the number of women workers from outside the family. The most influential variable is gross income.

Keywords: Net income, gross income, Pagar Alam coffee farmers, binary logistic regression

1   Introduction

Coffee is one of the mainstay export commodities of Indonesian plantation. Based on export value in 2017, Indonesia is among the 10 largest coffee exporting countries in the world. This can be seen from the results of the selection of leading export commodities (winning commodities) using several analytical methods, namely Computable General Equilibrium (CGE) and Export Product Dynamics (EPD) (Data source: Indonesia

Eximbank Institute and UNIED, 2019 in [1]). Indonesia is one of the fourth largest coffee producers in the world with an output of 6.84% of world coffee production. The provinces that contributed the most to Indonesia's coffee production were South Sumatra, Lampung, North Sumatra, Bengkulu, and Aceh.

In the WTO Series Webinars and Trade Policy Analysis on June 8-14, 2020, one of the speakers Dedi Budiman Hakim said that in 2017, the number of smallholder estates was estimated to have decreased by -0.07%, while state-owned and private estates rose by 0.07% and 0.18% respectively. The volume of coffee produced by smallholder plantations in 2017 is predicted to reach 599,902 tons, and the production growth will decrease by -0.37% compared to 2016. While state and private plantations' production volume were increased by 0.42% and 2.36% respectively. The factor that caused the decline in Indonesian coffee production was due to weather factors that did not support production activities. In addition, the factor of the lack of knowledge about coffee plantations by some farmers and the high price of fertilizers has caused Indonesia's coffee production to be at a maximum level.

The world coffee price is projected to increase due to the increasing demand for coffee for retail coffee shop needs, which is experiencing a positive trend due to a change in coffee drinking culture. In 2019, it is projected that Indonesian Coffee exports to the United States and several major destination countries will again grow positively in line with the increase in domestic production. Projections can be higher than expected if there is an increase in productivity of coffee plantations in Indonesia, which is caused by the application of better agricultural technology. Projections can be lower than expected if there is bad weather causing a decline in coffee yields, as well as an increase in domestic coffee consumption, so the amount of exported is decreasing (*Source: Trademap.org, Oxford Economics, LPEI* in [1]).

Coffee plantations in Pagar Alam, South Sumatra Province (Sum-Sel), are smallholder plantations that have existed since the Dutch colonial era. The majority of the types of coffee planted are Robusta. There is a portion of land at an altitude more than 1,000 masl producing Arabica coffee.

According to [2], Sum-Sel formed a cluster that its characteristics are the highest of land area, the highest of coffee production, the highest of area of TM-PR (Producing Plants on Smallholder Plantations), the highest of robusta area, and the highest of robusta coffee production. The area of TBM (Immature Plant), area of TR (Damaged Plants), and number of farmers in Sum-Sel are also high. Based on [3], Kota Pagar Alam, Banyuasin Regency and Lubuk Linggau formed a cluster that was formed not have dominant characteristics. All variable values researched are low. Those variables include: Area, Area of TM, Area of TBM, and Area of TR, Production, Average Production, and Number of Farmers.

Pagar Alam Coffee has its own unique taste, whose enjoyment is well known as Besemah coffee. This Besemah Coffee includes coffee which is also produced from Lahat Regency and Empat Lawang Regency [4]. These three regencies are bordering each other.

There are 31 variables (including 2 nominal scale variables) analyzed in the paper [5]. The correlation (*r*) between each of the 29 variables with the net income variable, the value is weak even very weak, ranging from -0.135 to 0.455. Only the gross income variable is strongly correlated with net income, which is 0.709. Variables that were negatively correlated were very weak, there were 6 variables, namely the age of the tree,

the number of women workers in the family, the number of male workers in the family, the length of harvest, and the maximum price of coffee.

Based on the results of the bivariate analysis with the chi-square test in [6], there were only 13 factors related to the farmers' net income, namely: the number of dependents, number of trees, age of trees, number of female workers from outside the family, frequency of fertilizer used, frequency of herbicide used, harvest production, outside harvest production, gross income, minimum price of coffee beans, maximum price of coffee beans, economic condition, and land productivity. Based on correspondence analysis in [7], the factors that have a relationship with land productivity are the area, number of trees, planting area of 1 tree, frequency of fertilization, frequency of herbicide use, harvest period, and harvest production.

The purpose of this paper is to form a Pagar Alam coffee farmers' net income probability model based on factors that are significantly related to net income. Data that is used base on [6]. Method used to determine the net income probability model is logistic regression. The estimation model in this paper can be used as a recommendation for parties from relevant agencies who need information about the factors that simultaneously influence the net income of coffee farmers in Pagar Alam. It can also be input for policy makers and related institutions to improve the welfare of coffee farmers through the optimization of production factors that affect farmers' incomes.

Logistic regression is a regression analysis that is used to describe the relationship between a dependent variable that is dichotomous (nominal or ordinal scale with 2 categories) or polychotomus (nominal or ordinal scale with more than 2 categories) and a set of predictor variables that are continuous or categorical [8]. If the dependent variable Y consists of two categories (binary), then the logistic regression is called binary logistic regression [9, 10].

One of application of binary logistic regression models was in [11]. The model can be used to analyse the characteristics of songket craftsmen in Ogan Ilir Regency, so that the potential of the craftsmen and the factors that directly influence the productivity of the craftsmen can be known. Factors that greatly influence productivity can be recommended for policy makers in improving the welfare of craftsmen. The same thing is also obtained from the results of the probability model on the factors that affect the income of coffee farmers.

2 Research Methodology

The subjects of this study were farmers in Pagar Alam, South Sumatra Province who run coffee farming. Respondents were chosen through purposive sampling technique. The data in this paper is the result of research on [6]. The variables used in this paper are variables that are significantly related to net income variable based on the results of correspondence analysis.

The method used is binary logistic regression analysis. The dependent variable is the net income of coffee farmers who are divided into 2 categories, namely 0: low net income and 1: high net income. There are 13 independent variables used, namely economic condition, number of dependents, number of coffee trees, age of trees, frequency of fertilizer used in one year, frequency of pesticides used in one year, production at harvest time (in quintals), production outside harvest time (in quintals), number of women

workers outside the family, the minimum price of coffee (in Rp / kg), the maximum price of coffee (in Rp / kg), gross income (in Rp), and land productivity (x 10^{-4} kg/m^2).

The application of binary logistics modeling is as follows:
1. Conducting descriptive analysis to find out the characteristics of the independent variables (can be seen in [5]);
2. Conducting a bivariate analysis to see the relationship between the independent variables with the dependent variable (that has been done on [6]);
3. Estimating model parameters using the Maximum Likelihood method;
4. Perform parameters testing simultaneously and partially;
5. Choosing the best model using the enter method, backward stepwise elimination, and forward stepwise elimination methods;
6. Interpreting the model and the results that have been obtained;
7. Make conclusions on the results of research.

Modeling is done with the help of Minitab 18 and SPSS 24 software.

3   Results and Discussions

The variables used in this paper are significantly related to the net income variable of Pagar Alam coffee farmers based on the results of correspondence analysis [6]. Data processing used binary logistic regression modeling. Dependent variable is the net income of coffee farmers divided into 2 categories, namely 0 as low net income and 1 as high net income. There are 13 independent variables used, all of which are divided into categories of variables and converted into ordinal scale. The division of each independent variable into categories based on the cut point method. These categories of independent variables and their descriptions can be seen in [2].

Table 1 shows variable notation, categories number of a variable, and category of a variable that has the highest percentage respondents.

Table 1. Defining variable notation and number of categories

No.	Variable	Notation	Number of Categories	Highest Category (in %)
1.	Net income	Y	2, i.e. 0: Low, and 1: High	1: High (63)
2.	Economic conditions	X_1	3	1 : Not enough (48) 2 : Enough (51)
3.	Number of dependents	X_2	5	3: 2 (36)
4.	Number of trees	X_3	5	3: (2,500; 4,000] (44)
5.	Age of tree	X_4	6	2: [10, 20] (45)
6.	Frequency of fertilizer used	X_5	6	2: 1 (46)
7.	Frequency of pesticides used	X_6	4	2: 1 (36)
8.	Production at harvest period	X_7	5	1: <1,000 (41) 2: [1,000; 2,000) (39)
9.	Production outside the harvest period	X_8	6	3: (50, 250] (41)
10.	Number of women workers outside the family (TKWL)	X_9	5	1: 0 person (79)
11.	Minimum price	X_{10}		1: < 19,333 (78)
12.	Maximum price	X_{11}	3	1: < 20,667 (63)
13.	Gross income	X_{12}	4	2: (10, 25] (49)
14.	Land productivity	X_{13}	4	2: [1,333; 2,667) (27) 3: [2,667; 5,000) (30)

Note: Suppose $X_{i(j)}$ as notation for variable X_i on category j. For example: $X_{2(3)}$ as notation for variable X_2 (number of dependents) on category 3 (2 person).

By defining the variables as in Table 1, the data matrix of 179 respondents was processed using SPSS version 24. Some of the outputs from data processing using the Enter method are as follows:

Table 2. Case Processing Summary

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	179	100.0
	Missing Cases	0	.0
	Total	179	100.0
Unselected Cases		0	.0
Total		179	100.0

a. If weight is in effect, see classification table for the total number of cases.

Based on Table 2, the number of respondents analyzed was 179 people and no data were missed.

Table 3. Dependent Variable Encoding

Dependent Variable Encoding

Original Value	Internal Value
0	0
1	1

Based on Table 3, the dependent variable code is 0 for low net income and 1 for high net income.

Table 4. Classification Table on Beginning Block

Block 0: Beginning Block

Classification Table[a,b]

		Predicted		
		Net Income		Percentage
	Observed	0	1	Correct
Step 0	Net Income 0	0	66	.0
	1	0	113	100.0
	Overall Percentage			63.1

a. Constant is included in the model.   b. The cut value is .500

Based on Table 4, in step 0, the logistic regression model only has a constant. There are no independent variables in the equation. The accuracy of the prediction is only 63.1%. With only a constant, it is significant at $\alpha = 5\%$ in influencing the net income of coffee farmers.

Table 5. Variables in Equation and not in Equation

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.538	.155	12.048	1	.001	1.712

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Economic conditions	5.699	1	.017
		Number of Trees	2.018	1	.155
		Age of Trees	.001	1	.978
		Freq. of Fertilizer	4.376	1	.036
		Freq. of Pesticides	1.034	1	.309
		Production at harvest	27.913	1	.000
		Prod. outside Harvest	7.694	1	.006
		TKWL	2.970	1	.085
		Minimum price	3.577	1	.059
		Maximum price	.670	1	.413
		Gross Income	87.195	1	.000
		Land Productivity	15.634	1	.000
		Number of Independents	1.729	1	.189
	Overall Statistics		94.330	13	.000

Based on Table 5, with the Wald test at $\alpha = 5\%$, we reject H0, so there are independent variables that affect farmers' net income.

Table 6. Block 1 in Enter Method

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	133.668	13	.000
	Block	133.668	13	.000
	Model	133.668	13	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	101.992[a]	.526	.719

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

The omnibus test in Table 6 presents a simultaneous test of all variable coefficients in the logistic regression model. Value $\chi^2 = 133.668$ is a distinction of -2 Log likelihood model with only constants and estimated models. Significant value represents that the independent variables affect net income.

The Cox & Snell R Square value is 0.526, which means that the independent variables in the model can explain the high or low net income of a farmer by 52.6%. Similarly,

Nagelkerke's value of 0.719 states that the independent variables in the model are able to explain the high or low net income of a farmer at 71.9%.

Table 7. Hosmer and Lemeshow Test

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	7.814	8	.452

Contingency Table for Hosmer and Lemeshow Test

		Net Income = 0		Net Income = 1		
		Observed	Expected	Observed	Expected	Total
Step 1	1	17	17.814	1	.186	18
	2	18	17.373	0	.627	18
	3	14	13.344	4	4.656	18
	4	7	6.760	11	11.240	18
	5	7	5.238	11	12.762	18
	6	1	3.499	17	14.501	18
	7	2	1.666	16	16.334	18
	8	0	.253	18	17.747	18
	9	0	.039	18	17.961	18
	10	0	.013	17	16.987	17

The Hosmer and Lemeshow test in Table 7 is based on the chi-square distribution test. The value of $\chi^2_{calculate}$ is 7.814. With df = 8 and $\alpha = 5\%$, the value of $\chi^2_{table\ (8;\ 0.05)}$ is 15.51, so $\chi^2_{calculate} < \chi^2_{(8;\ 0.05)}$. It means there is no difference between the observations with the model. The results of the chi-square test show no significance so the predicted probabilities correspond to the observed probabilities. In this case, the model formed can be said to be appropriate.

The Contingency Table for Hosmer and Lemeshow Test provides information that the data is divided into 10 groups. In each step, the number of farmers with high net income is raised. For example, in the first step, out of 18 cases, there were 17 low-net income farmers and 1 high-net income farmer. In the fifth step, out of 18 cases, there were 7 cases of low-net income farmers and 11 high-net income farmers.

Table 8. Classification Table on Step 1

Classification Table[a]

			Predicted		
			Net Income		
	Observed		0	1	Percentage Correct
Step 1	Net Income	0	48	18	72.7
		1	4	109	96.5
	Overall Percentage				87.7

a. The cut value is .500

The Classification Table in Table 8 shows how well the model classifies net income cases into 2 categories. The overall accuracy of model prediction is 87.7%. The accuracy value of this model is obtained from the corresponding column based on predictions divided by the amount of data (or the number of respondents). This shows that the model is better than the previous model with only a constant which is only 63.1% in predicting

the net income probability of Pagar Alam coffee farmers. While the accuracy of prediction of low and high net income farmers is respectively 72.7% and 96.5%.

Table 9. Variables in The Equation

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1ᵃ	Economic conditions	.223	.522	.182	1	.669	1.250
	Number of Trees	.501	.381	1.727	1	.189	1.651
	Age of Trees	-.154	.234	.435	1	.510	.857
	Freq. of Fertilizer	-.419	.368	1.295	1	.255	.658
	Freq. of Pesticides	.016	.281	.003	1	.955	1.016
	Production at harvest	-.512	.432	1.401	1	.236	.600
	Prod. outside Harvest	.403	.270	2.229	1	.135	1.497
	TKWL	.779	.394	3.919	1	.048	2.180
	Minimum price	.257	.730	.124	1	.724	1.294
	Maximum price	-.259	.642	.163	1	.686	.772
	Gross Income	4.663	.895	27.179	1	.000	105.998
	Land Productivity	.914	.400	5.214	1	.022	2.495
	Number of Independents	-.095	.232	.170	1	.680	.909
	Constant	-11.799	2.604	20.536	1	.000	.000

a. Variable(s) entered on step 1: Economic conditions, Number of Trees, Age of Trees, Freq. of Fertilizer Freq. of Pesticides, Freq. of Pesticides, Production at harvest, Prod. outside Harvest, TKWL, Minimum price, Maximum price, Gross Income, Land Productivity, Number of Independents.

Table 9 shows the Wald test which is a partial test of the significance of independent variables. Wald statistical value follows the chi-square distribution, so that if we see from the sig value, then for $\alpha = 5\%$, it is found that the independent variables of the number of women workers outside the family (TKWL), Gross Income, and Land Productivity significantly influence the significance of farmers' net income.

The value of Exp (B) is the odds ratio, all of which are positive. Exp (B) > 1 represents that for an increase in the independent variable by 1 unit, the ratio of the possibility of high net income to low net income also increases, by assuming the other independent variables are fixed.

In the economic condition variable whose ordinal scale data (i. e. 1, 2, and 3 consecutively states that the economic condition is not enough, sufficient, and more than enough), if the economic condition increases 1 level of category, then the likelihood ratio of high net income with low net income increases with a factor of 1.25, assuming the other independent variables are fixed.

Likewise, if the TKWL used rises by 1 level of category (i.e. category 1 for 0 TKWL, 2 for 1 person TKWL, and so on), then the likelihood ratio of high net income farmers with low net income rises by a factor of 2.18, assuming the other independent variables are fixed. If the farmer has 1 TKWL (notated by $X_{9(2)}$), then it is likely 2.18 times to increase the net income of the farmer, compared to if the farmer does not employ TKWL (notated by $X_{9(1)}$). An increase in 1 category of TKWL will cause an increase in net income from farmers by 2.18 times greater for each increase in the category.

For Exp (B) < 1 (or with a negative B value), it can represent if the independent variable goes up 1 level of the category, then the ratio of the possibility of high-net income farmers to low-net income will decrease by that factor, assuming the other independent variables are fixed. Examples of independent variables are the age of the tree, the frequency of fertilizer used, production at harvest, the maximum price, and the number of dependents. Every time there is an increase in the category of the age of trees, it is possible that the high net income of farmers will decrease.

Based on the significant factors affecting the net income of Pagar Alam coffee farmers, the binary logistic regression model that is formed is

$$\pi(x) = \frac{\exp\left(-11.799 + 0.779X_9 + 4.663X_1 + 0.914X_{13}\right)}{1 + \exp\left(-11.799 + 0.779X_9 + 4.663X_{12} + 0.914X_{13}\right)}$$

Because the coefficients of these significant variables are positive, so these variables can increase the probability value $\pi(x)$ of the model.

Calculation of the probability $\pi(x)$ of the model can be done in each of the available categories. For example, we can find probability model with TKWL $X_{9(1)}$, Gross Income $X_{12(2)}$, and Land Productivity $X_{13(3)}$. In this case, we have $X_9 = 1$, $X_{12} = 2$, and $X_{13} = 3$, so the probability $\pi(x)$ as follows:

$$\pi(x) = \frac{\exp\left(-11.799 + 0.779(1) + 4.663(2) + 0.914(3)\right)}{1 + \exp\left(-11.799 + 0.779(1) + 4.663(2) + 0.914(3)\right)}$$

Then, we get $\pi(x) = 0.7404$.

Based on the value $\pi(x)$ obtained, it can be seen that the probability for high net income for coffee farmers who do not employ TKWL have gross income of 10 till 25 million rupiahs, and have land productivity of 0.2667 - 0.5000 kg/m^2 is 74.04 %.

The results of calculating the probability of high net income from a combination of categories available from the three variables can be seen in Table 9

Table 9. Calculation Results $\pi(x)$ from the Combination of the Categories of Three Variables

No.	TKW-L (X_9) (in person)	Gross Income (X_{13}) (in million)	Land Productivity (X_{12})	Probability $\pi(x)$	No.	TKW-L (X_9) (in person)	Gross Income (X_{13}) (in million)	Land Productivity (X_{12})	Probability $\pi(x)$
1	1: 0	1: ≤ 10	1: <1333	0.0043			3: (25, 50]	1: <1333	0.9957
			2: [1333, 2667)	0.0107				2: [1333, 2667)	0.9983
			3: [2667, 5000)	0.0262				3: [2667, 5000)	0.9993
			4: ≥5000	0.0629				4: ≥5000	0.9997
		2: (10, 25]	1: <1333	0.3143			4: >50	1: <1333	1.0000
			2: [1333, 2667)	0.5334				2: [1333, 2667)	1.0000
			3: [2667, 5000)	**0.7404**				3: [2667, 5000)	1.0000
			4: ≥5000	0.8767				4: ≥5000	1.0000

		3: (25, 50]	1: <1333	0.9798
			2: [1333, 2667)	0.9918
			3: [2667, 5000)	0.9967
			4: ≥5000	0.9987
		4: >50	1: <1333	0.9998
			2: [1333, 2667)	0.9999
			3: [2667, 5000)	1.0000
			4: ≥5000	1.0000
2	2: 1 orang	1: ≤ 10	1: <1333	0.0093
			2: [1333, 2667)	0.0230
			3: [2667, 5000)	0.0554
			4: ≥5000	0.1276
		2: (10, 25]	1: <1333	0.4998
			2: [1333, 2667)	0.7136
			3: [2667, 5000)	0.8614
			4: ≥5000	0.9394
		3: (25, 50]	1: <1333	0.9906
			2: [1333, 2667)	0.9962
			3: [2667, 5000)	0.9985
			4: ≥5000	0.9994
		4: >50	1: <1333	0.9999
			2: [1333, 2667)	1.0000
			3: [2667, 5000)	1.0000
			4: ≥5000	1.0000
3	3: 2	1: ≤ 10	1: <1333	0.0201
			2: [1333, 2667)	0.0488
			3: [2667, 5000)	0.1133

4	4: 3	1: ≤ 10	1: <1333	0.0429
			2: [1333, 2667)	0.1005
			3: [2667, 5000)	0.2179
			4: ≥5000	0.4100
		2: (10, 25]	1: <1333	0.8259
			2: [1333, 2667)	0.9221
			3: [2667, 5000)	0.9672
			4: ≥5000	0.9866
		3: (25, 50]	1: <1333	0.9980
			2: [1333, 2667)	0.9992
			3: [2667, 5000)	0.9997
			4: ≥5000	0.9999
		4: >50	1: <1333	1.0000
			2: [1333, 2667)	1.0000
			3: [2667, 5000)	1.0000
			4: ≥5000	1.0000
5	5: ≥4	1: ≤ 10	1: <1333	0.0889
			2: [1333, 2667)	0.1958
			3: [2667, 5000)	0.3778
			4: ≥5000	0.6023
		2: (10, 25]	1: <1333	0.9118
			2: [1333, 2667)	0.9627
			3: [2667, 5000)	0.9847
			4: ≥5000	0.9938
		3: (25, 50]	1: <1333	0.9991
			2: [1333, 2667)	0.9996
			3: [2667, 5000)	0.9999

2: (10, 25]	4: ≥5000	0.2418	4: >50	4: ≥5000	0.9999
	1: <1333	0.6852		1: <1333	1.0000
	2: [1333, 2667)	0.8445		2: [1333, 2667)	1.0000
	3: [2667, 5000)	0.9312		3: [2667, 5000)	1.0000
	4: ≥5000	0.9713		4: ≥5000	1.0000
3: (25, 50]	1: <1333	0.9957			

Based on Table 9, if each of the independent variable categories is higher, then the probability value of the farmers' net income is higher. The most influential variable is gross income. While the variable with the smallest influence on the model is TKWL. In each TKWL category, if the gross income category is 4 (notated by $X_{12(4)}$), then for the land productivity in category 1 to 4 (starting from $X_{13(1)}$), the value $\pi(x) = 1$. In addition, in each TKWL category, if the gross income category is 3 (notated by $X_{12(3)}$), then for the land productivity in category 1 to 4 (starting from $X_{13(1)}$), the value $\pi(x) > 0.9$. Likewise, if the gross income category is 2 (notated by $X_{12(2)}$), then for the land productivity in category 4 (notated by $X_{13(4)}$), the value $\pi(x) \cong 0.9$.

In this case, the increasing net income of coffee farmers can be represented by high gross income, higher number of female workers outside the family, and high land productivity. In each TKWL category, if the gross income category is 1 (notated by $X_{12(1)}$), then regardless of the land productivity category (X_{13}), the probability value of net income $\pi(x)$ is very small.

The model results if data processing uses the forward and backward methods as follows:

Table 10. Some outputs by using the forward step method

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	115.827[a]	.488	.667
2	110.907[a]	.502	.686

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Classification Table[a]

			Predicted		
			Net Income		Percentage
	Observed		0	1	Correct
Step 1	Net Income	0	44	22	66.7
		1	2	111	98.2
	Overall Percentage				86.6
Step 2	Net Income	0	44	22	66.7
		1	2	111	98.2
	Overall Percentage				86.6

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Gross Income	4.315	.746	33.421	1	.000	74.790
	Constant	-7.508	1.448	26.882	1	.000	.001
Step 2[b]	TKWL	.624	.336	3.457	1	.063	1.867
	Gross Income	4.554	.812	31.436	1	.000	95.020
	Constant	-8.818	1.765	24.947	1	.000	.000

a. Variable(s) entered on step 1: Gross Income.
b. Variable(s) entered on step 2: TKWL.

Based on Table 10, the forward step method (as many as 2 steps) results two independent variables that have a significant effect on income, namely Gross Income and TKWL. The overall accuracy of predictions is 86.6%. This percentage is lower than the accuracy of the model of the enter method's result.

The model results if data processing uses the backward methods as follows:

Table 11. Some outputs by using the backward step method

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	101.992[a]	.526	.719
	...		
10	106.400[a]	.514	.703
11	108.059[a]	.510	.696

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Classification Table[a]

			Predicted		
			Net Income		Percentage
	Observed		0	1	Correct
Step 1	Net Income	0	48	18	72.7
		1	4	109	96.5
	Overall Percentage				87.7
	...				
Step 10	Net Income	0	44	22	66.7
		1	2	111	98.2
	Overall Percentage				86.6
Step 11	Net Income	0	44	22	66.7
		1	2	111	98.2
	Overall Percentage				86.6

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Tanggungan	-.095	.232	.170	1	.680	.909
	Economic conditions	.223	.522	.182	1	.669	1.250
	Number of Trees	.501	.381	1.727	1	.189	1.651
	Age of Trees	-.154	.234	.435	1	.510	.857
	Freq. of Fertilizer	-.419	.368	1.295	1	.255	.658
	Freq. of Pesticides	.016	.281	.003	1	.955	1.016
	Production at harvest	-.512	.432	1.401	1	.236	.600

	Prod. outside Harvest	.403	.270	2.229	1	.135	1.497
	TKWL	.779	.394	3.919	1	.048	2.180
	Minimum price	.257	.730	.124	1	.724	1.294
	Maximum price	-.259	.642	.163	1	.686	.772
	Gross Income	4.663	.895	27.179	1	.000	105.998
	Land Productivity	.914	.400	5.214	1	.022	2.495
	Constant	-11.799	2.604	20.536	1	.000	.000
...							
Step 10ᵃ	Prod. outside Harvest	.290	.230	1.585	1	.208	1.336
	TKWL	.675	.355	3.623	1	.057	1.965
	Gross Income	4.411	.809	29.699	1	.000	82.375
	Land Productivity	.439	.241	3.327	1	.068	1.552
	Constant	-10.343	2.017	26.286	1	.000	.000
Step 11ᵃ	TKWL	.618	.349	3.127	1	.077	1.855
	Gross Income	4.451	.805	30.597	1	.000	85.724
	Land Productivity	.394	.236	2.787	1	.095	1.483
	Constant	-9.522	1.832	27.018	1	.000	.000

Note: The outputs from step 2 to step 9 are not all displayed

Based on Table 11, using the backward step method (as many as 11 steps), obtained 3 independent variables that have a significant effect on net income, namely gross income, land productivity, and TKW-L. The overall accuracy of predictions is 86.6%. This percentage is the same as the accuracy of the model using the forward method, but lower than the model generated by the enter method.

The following Table 12 shows a recapitulation of data processing results based on 3 methods.

Table 12. Recapitulation of results from all three methods

Method	Accuracy of Model (%)	Significant Variables	Model
Enter	87.7	X_9, X_{12}, and X_{13}	$\pi(x) = \dfrac{\exp(-11.799 + 0.779X_9 + 4.663X_{12} + 0.914X_{13})}{1 + \exp(-11.799 + 0.779X_9 + 4.663X_{12} + 0.914X_{13})}$
Forward	86.6	X_9 and X_{12}	$\pi(x) = \dfrac{\exp(-8.818 + 0.624X_9 + 4.554X_{12})}{1 + \exp(-8.818 + 0.624X_9 + 4.554X_{12})}$
Backward	86.6	X_9, X_{12}, and X_{13}	$\pi(x) = \dfrac{\exp(-9.522 + 0.618X_9 + 4.451X_{12} + 0.394X_{13})}{1 + \exp(-9.522 + 0.618X_9 + 4.451X_{12} + 0.394X_{13})}$

Note: X_9: TKWL, X_{12}: Gross income, and X_{13}: Land productivity

Based on Table 12, the accuracy of the model resulting from the enter method is greatest. The coefficients of the independent variables on the model are also highest among three models.

The resulting model contains the gross income variable which has the highest effect on net income. Net income is gross income which has been reduced by production costs incurred by farmers. Production costs include land management, crop maintenance, labor costs, and other costs. Labor wages are usually issued for workers from outside the family, both men and women. Women workers are paid for picking coffee fruit. Plant maintenance includes the provision of fertilizers and weed control by herbicides. Tillage also supports crop maintenance. Crop maintenance costs also relate to the age of the tree

and the number of trees. Older trees need better maintenance, so that the roots remain sturdy and also need to be rejuvenated. The more trees, the greater the maintenance costs. Plant spacing that is too tight can reduce production. In this case, the frequency of fertilizer use, frequency of pesticide use, number of trees, and age of trees variables are related to production costs. Production costs are contained in the variable gross income.

4   Conclusions

The conclusions obtained from this study are the factors that have a significant influence on the net income of Pagar Alam coffee farmers are gross income, land productivity, and the number of women workers from outside the family.

Simultaneously, the gross income (X_{12}), land productivity (X_{13}), and the number of women workers from outside the family (X_9) variables affect net income with the probability model:

$$\pi(x) = \frac{\exp{(-11.799+0.779X_9+4.663X_{12}+0.914X_{13})}}{1+\exp{(-11.799+0.779X_9+4.663X_{12}+0.914X_{13})}}$$

All the coefficients of the variables that have a significant effect are positive, then these variables can increase the probability value $\pi(x)$ of the model. If each variable category gets higher, the probability value of the farmers' net income is higher. In each TKWL category, if the gross income category is 4 ($X_{12(4)}$) and land productivity category is starting from 1 ($X_{13(1)}$), then the value π (x) = 1.

This study does not describe variables related to land productivity. Because in this study it was found that land productivity has a significant effect on the binary logistic regression model, it is necessary to examine the indirect relationship between variables that have no significant effect on the model on farmer's net income. In this case, it needs to be further analyzed by path analysis regarding the indirect effects of the number of trees, frequency of fertilizer used, frequency of pesticides used, crop production, land area, area of 1 tree, and length of time of harvest.

Acknowledgements

References

[1] D. B. Hakim, "Analisis Kebijakan Perdagangan Internasional Pertanian dan Manufaktur," Webinar Berseri WTO dan Analisis Kebijakan Perdagangan, 8-14 Juni 2020, Cooperation ITABS, Dept. IE FEM IPB, ISEI Bogor Raya, Kementerian Perdagangan, and UNIED, 2020.

[2] Irmeilyana, Ngudiantoro, A. Desiani, and D. Rodiah, "Deskripsi hubungan luas areal dan produksi perkebunan kopi di Indonesia menggunakan analisis bivariat dan

analisis klaster," *Infomedia*, vol. 4, no. 1, 2019, pp. 21–27, doi: http://dx.doi.org/10.30811/jim.v4i2.1534.

[3] Irmeilyana, Ngudiantoro, A. Desiani, and D. Rodiah, "Deskripsi Hubungan Luas Areal dan Produksi Perkebunan Kopi di Provinsi Sumatra Selatan," in *Proseding SEMIRATA BKS PTN Indonesia Barat*, 2019, pp. 74–86.

[4] _____, "Kopi Robusta Besemah Sumatera Selatan," 2017. [Online]. Available: http://www.lintaskopi.com/kopi-robusta-besemah-sumatera-selatan/

[5] Irmeilyana, Ngudiantoro, and D. Rodiah, "Deskripsi Profil dan Karakter Usaha Tani Kopi Pagar Alam Berdasarkan Descriptive Statistics dan Korelasi," *Infomedia*, vol. 4, no. 2, 2019, pp. 60–68, doi: http://dx.doi.org/10.30811/jim.v4i2.1534.

[6] Irmeilyana, Ngudiantoro, and D. Rodiah, "Laporan Penelitian Hibah Kompetitif: Analisis Pengaruh Faktor-faktor Internal dan Eksternal Terhadap Produktivitas Petani Kopi dan Usaha Optimalisasi Produksi Kopi," Indralaya, 2019.

[7] Irmeilyana, Ngudiantoro, and D. Rodiah, "Application of Simple Correspondence Analysis to Analyze Factors that Influence Land Productivity of Pagar Alam Coffee Farming," in *International Conference on Mathematics, Statistics, and Their Applications (ICMSA) 2019*, 2019.

[8] A. Agresti, *Categorial Data Analysis*. New York: John Wiley & Sons Inc., 2009.

[9] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, 2nd Ed. New York: John Wiley & Sons Inc., 2000.

[10] A. Widarjono, *Analisis Multivariat Terapan*, Edisi Kedua. Yogyakarta: UPP STIM YKPN, 2015.

[11] M. N. Samsuri, Ngudiantoro, and Irmeilyana, "Pemodelan Regresi Logistik Produktivitas Pengrajin Songket di Kabupaten Ogan Ilir," in *Proceeding The 5th Annual Reasearch Seminar 2019*, 2020, pp. 66–71.