Variance: Journal of Statistics and Its Applications
P-ISSN: 2685-8738 E-ISSN: 2685-872X
Volume 7 Issue 1 | April 2025
10.30598/variancevol7iss1page49-60

MULTIGROUP MODELING BASED PATH ANALYSIS ON PANEL DATA IN THE STUDY OF PRODUCTION PERFORMANCE IN SMALL AND MEDIUM INDUSTRIES IN MALANG

Meilina Retno Hapsari[1*], Nur Kamilah Sa'diyah[2]
[1,2]Department of Statistics, Faculty of Mathematics and Natural Sciences, Brawijaya University
Jl. Veteran No.10-11, Malang, East Java, 65145, Indonesia

Corresponding Author's Email: *meilinaretno@ub.ac.id*

Abstract: *This study aims to determine the effect of the number of workers on production performance through the amount of production between small and medium industries in Malang City. This study applies multigroup modeling based on path analysis on panel data on variables that affect industrial production performance in Malang and the application of the Hypothesis of Linear Parameter Function in the multigroup model, which is the second group. This data source is secondary data from Malang Industrial data. The sample in this study is all data from small and medium industries in Malang City over several years. The results showed that the Hypothesis of Linear Parameter Function can be applied to the two-group model, that is, in a medium industry, where three significant paths are obtained. In the small industry, there are two considerable paths. Based on the coefficient of total determination, the data diversity can explain the model of 96.0%. At the same time, the remaining 4.0% is affected by other variables. The small and medium industry model shows that the number of workers significantly affects production. The effect is greater in medium industries, which is 0.935, compared to small industries, which is 0.737.*

Keywords: *Production Performance, Multigroup, Path Analysis, Panel Data, Hypothesis of Linear Parameter Function.*

1. INTRODUCTION

The type of data in the study is not only cross-sectional data or only time series data, but can use panel data, namely combining cross-sectional and time series data [1]. Several approaches are commonly used in estimating panel data analysis or panel data regression, including pooling least squares (common effect), fixed effects, and random effects [2]. In addition, this study combines panel data analysis and path analysis because there are intervening variables in the model. In various studies, researchers rarely use only one endogenous variable observed. An analysis involving more than one endogenous variable is called a multivariate analysis. One of them is path analysis. Path analysis is a statistical modeling involving at least one exogenous variable, endogenous variable, and intervening variable [3]. The purpose is to determine the effect of exogenous variables on endogenous variables through intervening variables.

In some studies, there are those whose goal is to compare the groups. In the [4] study on The Interpersonal Power of Feminism: Is Feminism Good for Romantic Relationships? This study compared the results of multiple regression analysis on women's groups and men's groups separately. The research of [5] on Growth messages increases help-seeking and performance for women and men in STEM, using path analysis and comparing the two groups separately. This research has not accommodated a comparison of the results of testing the relationship between variables in the two groups. So, in this research, a modeling development was carried out to solve the problem of multigroup equation modeling.

Multigroup modeling uses path analysis to compare two groups, analyzed twice, and a comparison of model coefficients is carried out in both groups [6]. Multigroup solutions are identical to dummy regression analysis, where dummy variables with a value of 0 are added for group one and dummy variables with a value of 1 for group two. Multigroup modeling using path analysis can solve the problem of using multigroup data, but test results between groups cannot yet be compared. Therefore, a parameter linear function hypothesis test (HFLP) was carried out [7]. This analysis was carried out by [8], namely, applying multigroup equation modeling with application

 Submitted: 2024-11-20 *Accepted: 2025-05-08*

path analysis to student behavior in purchasing modern and traditional food. The type of data used is cross-sectional. The study [9] applied multigroup SEM-PLS analysis based on gender to Fintech services. The type of data used is also cross-sectional.

Sectoral economic growth and development are transforming. The agricultural sector, which previously made a significant contribution to the Indonesian economy, has begun to shift to the industrial sector along with technological developments and the entry of foreign capital into Indonesia. The development of the industrial sector can provide a strategic role in the economy. One of the efforts to improve industrial performance is identifying the factors that influence industrial performance. Industry must consider whether industrial products are optimal in production processes and activities to create optimal production value [10]. Based on the background above, the author analyzed the factors that influence industrial production performance as seen from the production value or income of Small and Medium Industries in Malang City. The factors suspected of influencing production performance are the number of workers and the amount of production as mediating variables. From this objective, the author wants to estimate the parameters of the multigroup model (small and medium industries) based on path analysis on panel data. Small industries are group one with dummy variable notation 0, and medium industries are group two with dummy variable notation 1.

2. METHODOLOGY

2.1. Data Sources and Research Variables

The data used in this study are secondary data obtained from the Malang City Industry data. The population in this study is all data from small and medium industries in Malang City. The sample in this study is small and medium-sized industries in Malang City from 2021 to 2023. The research variables used in this study are suspected of influencing industrial production performance, namely, the number of workers and the amount of production as intervening variables. Production performance is seen from the value of production or industrial income as an endogenous variable (Y). The number of workers is an exogenous variable (X). The amount of production is an intervening variable (Z).

2.2. Data Analysis Methods

This study uses a combination of panel data analysis and path analysis. Panel data analysis combines cross-sectional and time series data. Several approaches are often used when estimating regression model parameters using panel data, including pooling least squares (common effect), fixed effects, and random effects. This study combines panel data analysis with path analysis because there are intervening variables in two groups, namely, small and medium industries. So, the author conducted a multigroup model analysis (small and medium industries) based on path analysis using panel data. Small industries are group one with dummy variable notation 0, and medium industries are group two with dummy variable notation 1.

The steps of this research analysis are as follows.

1. Create a path diagram according to the theory.

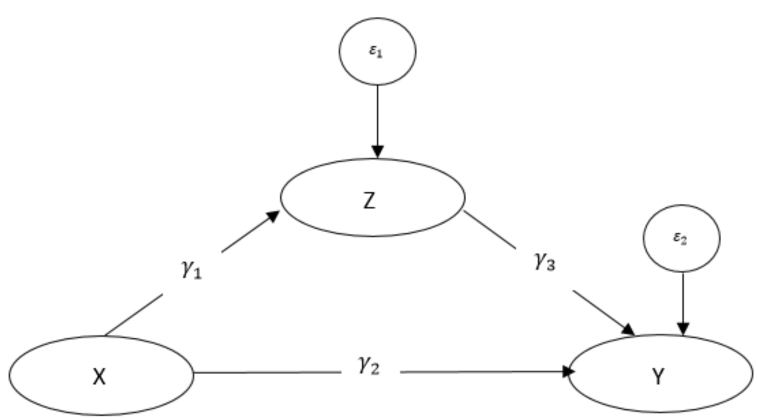

Figure 1. Path Diagram

Multigroup modeling using path analysis is an approach used to compare two groups. This study aims to compare small and medium industries in Malang. Multigroup path analysis is analyzed twice, and a comparison of the path model coefficients in both groups is carried out. The multigroup equation model is identical to dummy regression with groups one and two, which are group variables.

2. Create a path analysis model according to the path diagram.

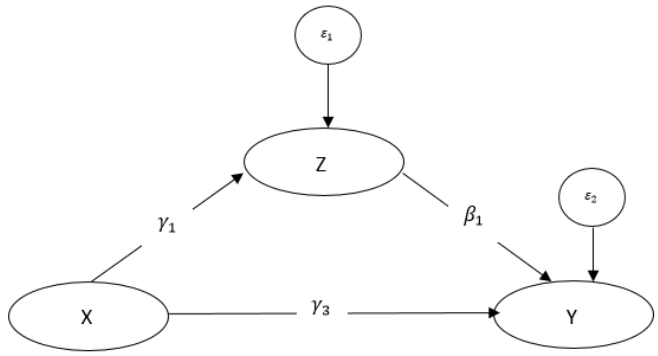

Figure 2. Path Diagram with Group One Dummy Variable

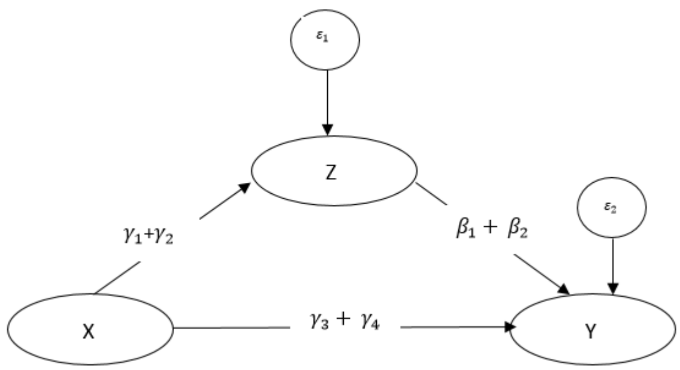

Figure 3. Path Diagram with Group Two Dummy Variable

$$Z = \gamma_1 X + \gamma_2 D_i X + \varepsilon_1$$
$$Y = \gamma_3 X + \gamma_4 D_i X + \beta_1 Z + \beta_2 DZ + \varepsilon_2 \quad (1)$$

For the dummy variable group one, if the notation 0 is entered in Equation (1):

$$Z = \gamma_1 X + \varepsilon_1$$
$$Y = \gamma_3 X + \beta_1 Z + \varepsilon_2 \quad (2)$$

For the dummy variable group two, if the notation 1 is entered in Equation (2):

$$Z = (\gamma_1 + \gamma_2) X + \varepsilon_1$$
$$Y = (\gamma_3 + \gamma_4) X + (\beta_1 + \beta_2) Z + \varepsilon_2 \quad (3)$$

3. Determine a panel data regression model. There are three model in panel data analysis, namely Common Effect Model (CEM), Fixed Effect Model (FEM) and Random Effect Model (REM). There are several tests to determine the panel data regression model. The Chow test is used to determine whether the common effect model is better than the fixed effect model. Hausman test to determine whether the fixed effect model is better than the random effect model. Lagrange Multiplier test to determine whether the random effect model is better than the common effect.

4. Estimate the path coefficient and the assumption test. Some assumptions that must be met are linearity, residual normality, and homoscedasticity. According to [11], the residual normality test uses Kolmogorov-Smirnov. To test the homogeneity of residual variance, the Breusch Pagan test is used [12]. One can use Ramsey's RESET Test [13] to test linearity.
5. Conduct a hypothesis test for group one path analysis with the t-test. Hypothesis testing using a t-test will only be done on group one, with the dummy variable notated as 0. For the testing criteria, if the t-statistic $\geq$ t -table or the p-value $< \alpha$, reject H_0. The path coefficient function provides a significant effect, meaning that one variable has a meaningful impact on another variable in the model formed in group two.
6. Conduct a hypothesis test for group two path analysis with the Hypothesis of Linear Function Parameter. The model coefficient values can be compared when solving multigroup modeling problems with path analysis. However, path analysis only accommodates testing of group one with a dummy variable notation of 0. The linear parameter function hypothesis will be used to complete the testing of group two with a dummy variable notation of 1. The second group in the modeling with a dummy notation 1 has a more complex equation, such as Equation (3).

   The hypothesis of linear function parameters according to Searle (1971): $H: k'\beta = m$

   The hypothesis of the linear function of the parameters formed based on Equation (5) is as follows:

$$\begin{aligned} &H_{01}: \gamma_1 + \gamma_2 = 0 && && H_1: \gamma_1 + \gamma_2 \neq 0 \\ &H_{02}: \gamma_3 + \gamma_4 = 0 && \text{vs} && H_1: \gamma_3 + \gamma_4 \neq 0 \\ &H_{03}: \beta_1 + \beta_2 = 0 && && H_1: \beta_1 + \beta_2 \neq 0 \end{aligned} \quad (5)$$

   can be written in a matrix:

$$H_{01}: k'_{(1)}\beta = m$$

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad (6)$$

$$H_{02}: k'_{(2)}\beta = m$$

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad (7)$$

$$H_{03}: k'_{(3)}\beta = m$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad (8)$$

For the testing criteria, if the F-statistic ≥ the F -table, reject H_0, or if the p-value $< \alpha$. The path coefficient function provides a significant effect, meaning that one variable directly impacts another variable in the model formed in group two.

7. Check the model evaluation. Model evaluation in estimating panel data path analysis parameters can be done by looking at the Total Determination Coefficient value. If the R^2 value > 0.75, 0.50 < R^2 < 0.75, or R^2 < 0.25 for the endogenous latent variable in the model, it can describe a strong, moderate or weak relationship, respectively [14].
8. Compare the magnitude of the path coefficients from the two groups.

3. RESULTS AND DISCUSSION

3.1. Descriptive Analysis

The following are the results of a descriptive analysis of the variables of the number of workers, the amount of production, and production performance in small and medium industries.

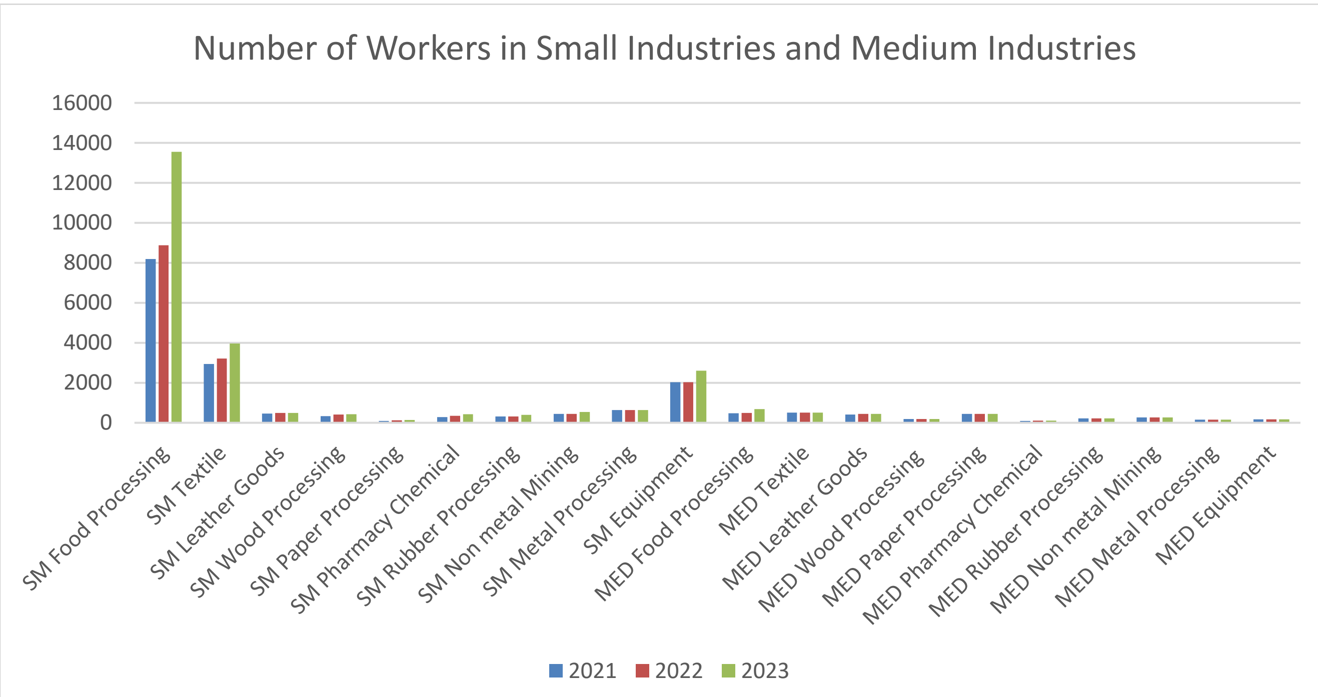

Figure 4. Number of Workers in Small Industries and Medium Industries

Based on Figure 4, it can be seen that the highest number of workers is in the small and medium industry, specifically the food processing industry. The second largest number of workers in small and medium industries is in the textile industry.

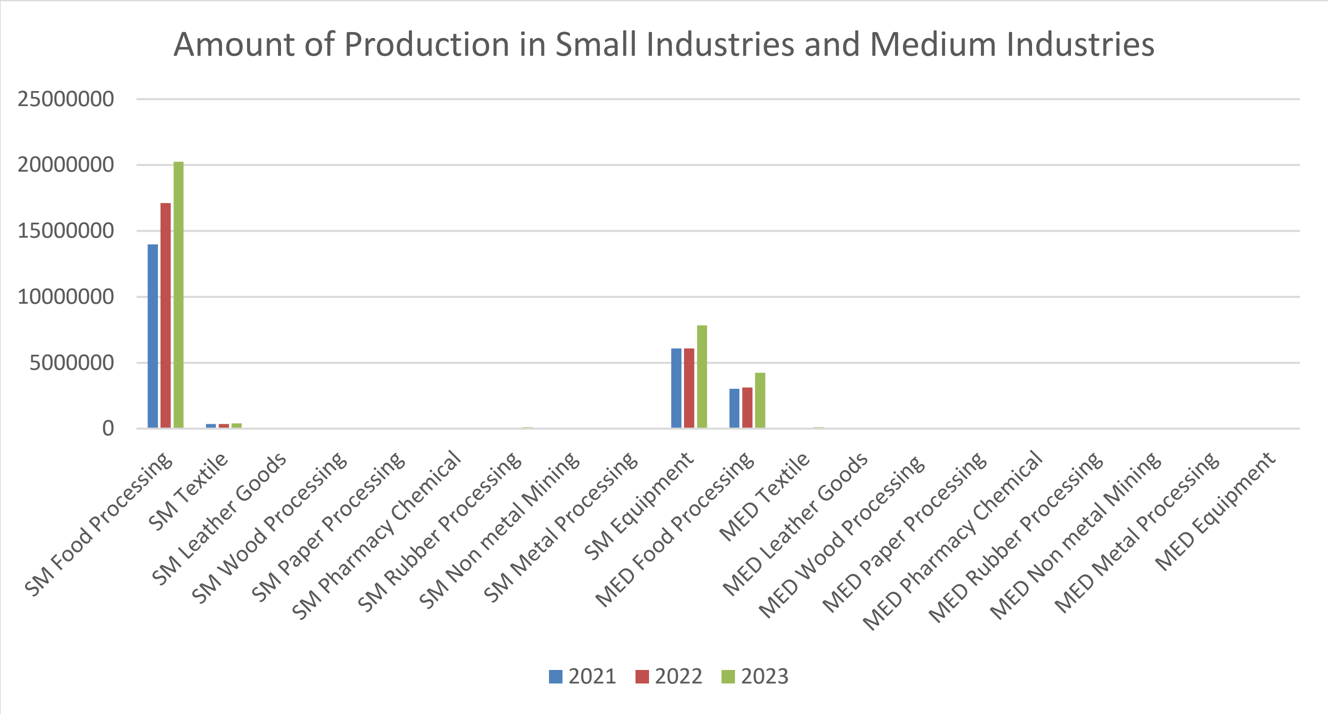

Figure 5. Amount of Production in Small Industries and Medium Industries

Based on Figure 5, it can be seen that the highest amount of production is in the small and medium industry, specifically the food processing industry. The second most significant production in the small industry is the equipment industry.

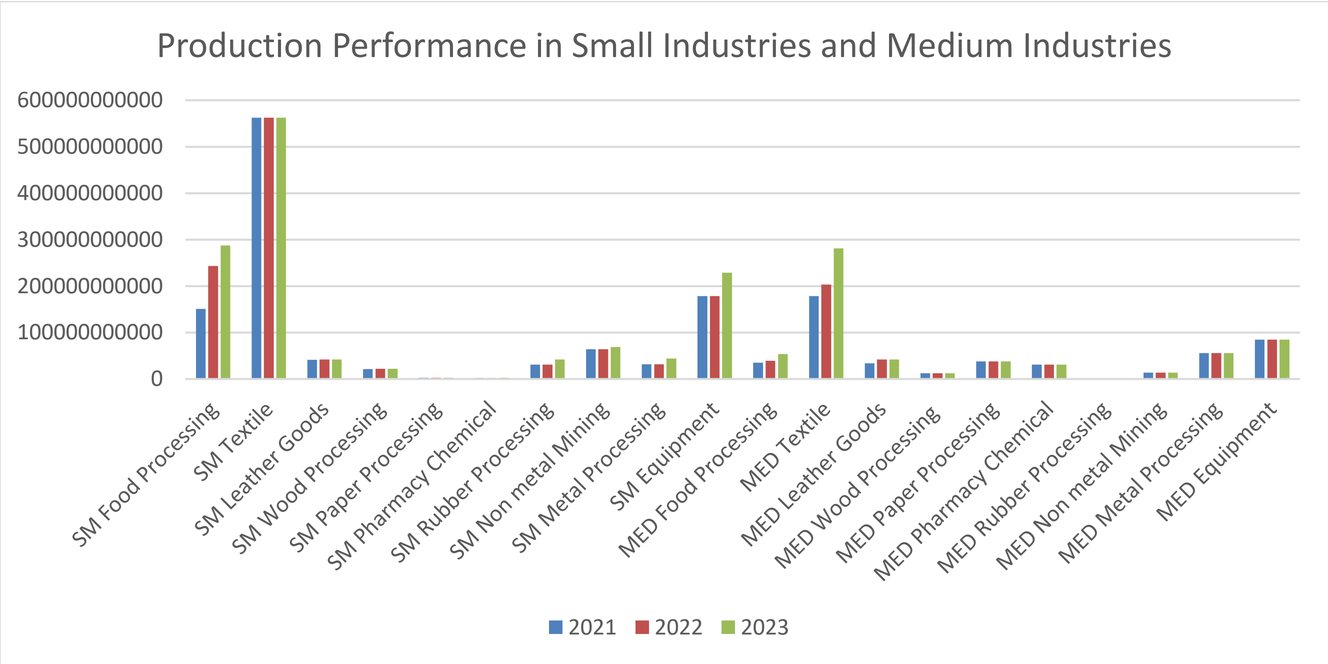

Figure 6. Production Performance in Small Industries and Medium Industries

Based on Figure 6, it can be seen that the highest production performance is in the small and medium industry, specifically the textile industry. The second largest production performance in the small industry is the food processing industry, and in the medium industry is the equipment industry.

3.1 Panel Regression Model Selection

Panel data with path analysis contributes significantly to understanding the relationship between variables, controlling for unobserved variables, and obtaining more accurate estimates than cross-sectional data. This is because panel data captures variability across time and between individuals, which cannot be done with cross-sectional data.

Common effect, fixed effect, and random effect models are three approaches that can be used to estimate model parameters when using panel data [15]. Furthermore, the Chow, Hausman, and Lagrange tests determine which panel data parameter estimates will be used. The selection of the panel regression model is carried out in each substructure. First, the Chow Test is carried out to determine the appropriate model between the Common Effect Model (CEM) and the Fixed Effect Model (FEM). The following are the results of the Chow Test:

Table 1. Results of Panel Regression Model Selection Test

Substructure	Test	p-value	Decision
I	Chow Test	0.000	Fixed Effect Model (FEM)
	Hausman Test	0.000	Fixed Effect Model (FEM)
II	Chow Test	0.000	Fixed Effect Model (FEM)
	Hausman Test	0.000	Fixed Effect Model (FEM)

Based on Table 1, the Chow test for substructures I and II obtained a p-value (0.000) < 0.05, so the appropriate model estimate applied to estimate parameters is the Fixed Effect Model (FEM) [15]. Furthermore, the Hausman Test is carried out which aims to determine the appropriate model between the Fixed Effect Model (FEM) and the Random Effect Model (REM). Based on the results of the Hausman substructure I and II tests in the table above, the p-value (0.0005 < 0.05) was obtained, so the estimation model used is the Fixed Effect Model (FEM).

3.2. Assumption Test

The assumptions that must be met are linearity, residual normality and homoscedasticity.

Table 2. The Result of Assumption Test

Assumption Test	Relationship between Variables	p-value
Linearity Test	X → Z	0.352
	X → Y	0.376
	Z → Y	0.412
Residual Normality Test	Amount of production	0.106
	Production performance	0.138
Homoscedasticity Test	Amount of production	0.814
	Production performance	0.525

Based on the results of the linearity test, the p-value of each relationship is above 0.05, so the relationship between variables is linear. Based on the residual normality test results, if the p-value of each residual is > 0.05, then the residuals are normally distributed. Based on the results of the homoscedasticity test, if the p-value > 0.05, then the residual variance is constant.

3.3. Path Analysis Parameter Estimation

The path analysis of Multigroup panel data in this study, the parameter estimation uses FEM with the Within Group approach. The following are two path diagrams for groups one (small industry) and two (medium industry).

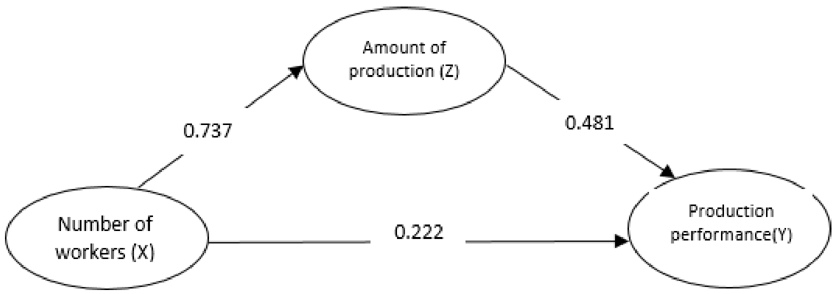

Figure 7. Path Coefficient of Group One Dummy Variable Notated 0 (Small Industry)

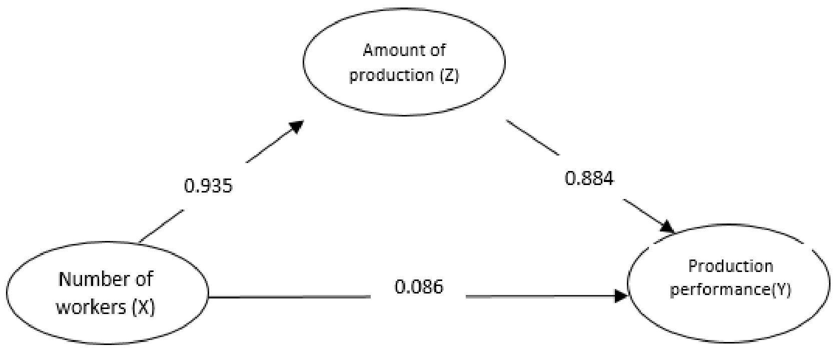

Figure 8. Path Coefficient of Group Two Dummy Variable Notated 1 (Medium Industry)

Based on Figure 7 and Figure 8, the existing path coefficients can be formed into path analysis equations for group one and group two as follows:

Equation for group one dummy variable notated 0:

$$Z = 0.737X$$
$$Y = 0.222\,X + 0.481\,Z \tag{9}$$

Equation for group two dummy variables notated 1:

$$Z = 0.935\,X$$
$$Y = 0.086\,X + 0.884\,Z \tag{10}$$

Based on the coefficients in Figure 7 and Figure 8, it can be explained that the number of workers and the amount of production have a positive effect on production performance in small and medium industries. The number of workers also positively affects the amount of production in small and medium industries. The effect can be seen in the hypothesis test result to determine whether or not it is significant.

3.4. Hypothesis Testing

In path analysis, hypothesis testing is carried out to determine whether the model formed is significant. The following are the results of hypothesis testing for Group One Dummy Variable Notated 0 (Small Industry).

Table 3. The Results of Hypothesis Testing of Group One Dummy Variable Notated 0 (Small Industry)

Relationship Between Variables	p-value	Decision
Number of workers (X) → Amount of production (Z)	0.004	Significant
Number of workers (X) → Production performance (Y)	0.428	Not Significant
Amount of production (Z) → Production performance (Y)	0.000	Significant

From the results of the analysis, it can be seen that the p-value of each variable relationship (0.000) <α (0.05), so the decision H_0 is rejected on all paths. It can be interpreted that in small industries, the number of workers significantly affects the amount of production. The number of workers does not have a significant effect on production performance. The amount of production has a substantial impact on production performance. In small industries, the number of workers does not significantly affect production performance for several reasons. In small industries, the number of workers is usually relatively smaller, and the production process is still very simple or limited to a small scale. Thus, the presence of additional workers may not significantly impact increasing productivity. In addition, small industries tend to rely more on manual methods or simple technology. Therefore, even though there is an increase in workers, not all aspects of production can be optimized simply by increasing the number of workers.

Hypothesis testing for Group Two Dummy Variables Notated 1 (Medium Industry) cannot be done with ordinary testing, but hypothesis testing is carried out with the Parameter Linear Function Hypothesis. The following are the results of hypothesis testing for Group Two Dummy Variables Notated 1 (Medium Industry).

Table 4. The Results of Hypothesis Testing of Group Two Dummy Variables Notated 1 (Medium Industry)

Relationship Between Variables	p-value	Decision
Number of workers (X) → Amount of production (Z)	0.000	Significant
Number of workers (X) → Production performance (Y)	0.000	Significant
Amount of production (Z) → Production performance (Y)	0.000	Significant

From the results of the analysis, it can be seen that the p-value of the three variable relationships (0.000) <α (0.05), then the decision H0 is rejected on two paths. It can be interpreted that in the medium industry, the number of workers significantly affects the amount of production. The number of workers has a significant effect on production performance. And the amount of production has a substantial impact on production performance. This study's results align with the statement of [10] that industry needs to consider the number of products and those carrying out industrial processes, such as labor, to create optimal production value.

3.5. Model Evaluation

Model evaluation in estimating panel data path analysis parameters can be done by looking at the Total Determination Coefficient value. The results of the determination coefficient calculation can be seen as follows.

Table 5. Coefficient of Determination

Endogen Variable	R^2	R^2 total
Amount of production	0.357	0.960
Production performance	0.939	

The calculation result of the total determination coefficient is 0.960. The value of $R^2 > 0.75$ means that the model is a good or strong in explaining the diversity of data contained in the model.

3.6. Trimming

Theory Trimming improves the path analysis model by removing insignificant path analysis coefficients. Based on the hypothesis test that has been carried out separately on both groups, the following path diagram is obtained:

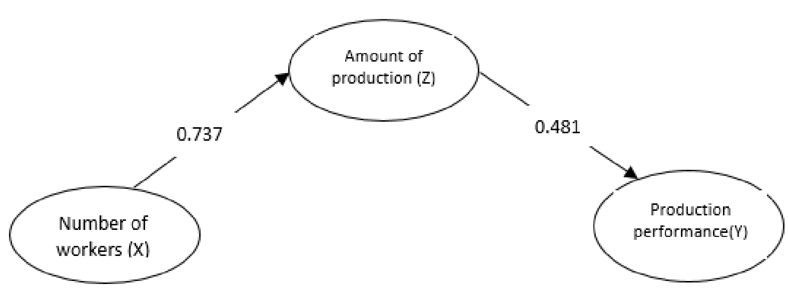

Figure 9. Path Coefficients after Trimming Group One Dummy Variable Notated as 0 (Small Industry)

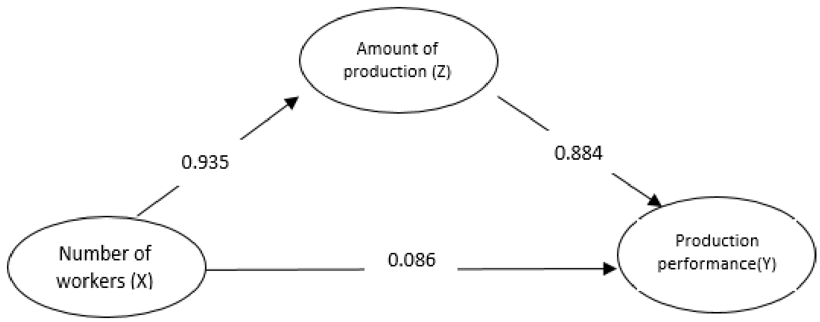

Figure 10. Path Coefficients after Trimming Group Two Dummy Variable Notated as 0 (Medium Industry)

Based on the significance test in group one using the t-test statistic, all paths were found to be significant. In the significance test in group two, utilizing the Hypothesis of Linear Parameter Function, there was an insignificant path: the effect of production volume on production performance. The insignificant path was then removed from the model to obtain the diagram in the Figure above. In summary, the path coefficients of the two groups are compared as follows.

Table 6. Comparison of path coefficients of the two groups

Relationship between variables	Group One (Small Industries)	Group Two (Medium Industries)
Number of workers→ Amount of production	0.737	0.935
Number of workers → Production performance	-	0.086
Amount of production → Production performance	0.481	0.884

Based on Table 6, the effect of the number of workers on the amount of production and the impact of the amount of output on production performance are higher in medium industries than in small industries.

3.7. Direct and Indirect Effect

After obtaining the effect of each variable from the significant model, the indirect effect and total effect on the results of the path analysis of the two groups can then be calculated.

Table 7. The Result of Direct Effect, Indirect Effect and Total Effect of Both Groups

Group	Relationship between Variables	Direct Effect	Indirect Effect	Total Effect
Small Industries	Number of workers → Amount of production	0.737		0.737
	Number of workers→ Production performance	-	0.737 x 0.481 = 0.256	0.256
	Amount of production → Production performance	0.481		0.481
Medium Industries	Number of workers → Amount of production	0.935		0.935
	Number of workers→ Production performance	0.086	0.935 x 0.884 = 0.826	0.912
	Amount of production → Production performance	0.884		0.884

Based on Table 7, it can be seen in group one, namely small industries, that the number of workers significantly affects the amount of production. In group two, namely medium industries, the number of workers substantially affects production. This is because the greater the number of workers available, the greater the company's capacity to produce goods or services. With more workers, the production process can be done faster, more efficiently, or on a larger scale. With enough labor, tasks in production can be divided and adjusted according to the workers' skills. This specialization increases efficiency and productivity, so that total production increases.

4. CONCLUSION

The conclusion of the analysis in this study is as follows:

1. Based on the results of comparing the large coefficient model between small and medium industry groups, the most significant effect of the number of workers on the amount of production is found in medium industries. The most important impact of the amount of output on production performance is found in medium industries.
2. Hypothesis Testing of Linear Function Parameters (HFLP) can be used to test the multigroup path analysis model with panel data in the second group with dummy variables notated as 1. This study's testing results with the second group, HFLP, obtained all significant paths.
3. These findings can be a reference for making strategic decisions for industry or business players to pay more attention to workforce allocation and production processes, increase output and optimal performance, and support the growth and sustainability of their business or industry.

REFERENCES

[1] D. N. Nachrowi and H. Usman, *Pendekatan Populer dan Praktis Ekonometrika untuk Analisis Ekonomi dan Keuangan*. Jakarta: Lembaga Penerbit FE UI, 2006.

[2] D. N. Gujarati and D. C. Porter, *Dasar-Dasar Ekonometrika*, 5th ed. Jakarta: Salemba Empat, 2015.

[3] I. Ghozali, *Aplikasi Analisis Multivariate dengan Program IBM SPSS 25*. Semarang: Badan Penerbit Universitas Diponegoro, 2018.

[4] L. A. Rudman and J. E. Phelan, "The Interpersonal Power of Feminism: Is Feminism Good for Romantic Relationships?," *Sex Roles*, vol. 57, no. 11–12, pp. 787–799, Dec. 2007, doi: 10.1007/s11199-007-9319-9.

[5] R. Covarrubias, G. Laiduc, and I. Valle, "Growth messages increase help-seeking and performance for women in STEM," *Group Process. Intergroup Relat.*, vol. 22, no. 3, pp. 434–451, Apr. 2019, doi: 10.1177/1368430218802958.

[6] Solimun, A. R. Fernandes, and Nurjannah, *Metode Statistika Multivariat Pemodelan Persamaan Struktural (SEM)*. Malang: UB Press, 2017.

[7] Searle, *Linear Model*. New York: John Wiley, 1971.

[8] R. Darmawan, "Pemodelan Persamaan Multigroup Berbasis Analisis Jalur (Studi pada Kepuasan dan Loyalitas Mahasiswa terhadap Makanan Tradisional dan Modern)," *Uiversitas Brawijaya*, 2018.

[9] B. Setiawan, "A multigroup analysis of gender in adopting Fintech services in Indonesia," *Enrich. J. Manag.*, vol. 14, no. 2, pp. 238–248, 2024.

[10] S. Sukirno, *Mikro Ekonomi Teori Pengantar*, 3rd ed. Jakarta: Rajawali Press, 2010.

[11] D. J. Sheskin, *Parametric and Nonparametric Statistikal Procedures*. 2000.

[12] M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, *Applied Linear Statistical Models*, 5th ed. New York: Mc-Graw-Hill, 2005.

[13] D. N. Gujarati, *Dasar-Dasar Ekonometrika*. Jakarta: Salemba Empat, 2004.

[14] J. F. Hair, C. M. Ringle, and M. Sarstedt, "PLS-SEM: Indeed a Silver Bullet," *J. Mark. Theory Pract.*, vol. 19, no. 2, pp. 139–152, Apr. 2011, doi: 10.2753/MTP1069-6679190202.

[15] D. N. Gujarati, *Dasar-Dasar Ekonometrika*. Jakarta: Salemba Empat, 2015.