BUKTI KORESPONDENSI ARTIKEL JURNAL

JOURNAL OF SOUTHWEST JIAOTONG UNIVERSITY (JSJU) VOL.58, NO.3, JUNI 2023, HAL. 688-698 (TARNO)

Link Jurnal: Journal of Southwest Jiaotong University (jsju.org)

Link Artikel: https://www.jsju.org/index.php/journal/article/view/1694

Doi: **https://doi.org/10.35741/issn.0258-2724.58.3.57**

Link Indeks: Xinan Jiaotong Daxue Xuebao/Journal of Southwest Jiaotong University (scimagojr.com)

No.	Tanggal	Aktivitas Korespondensi	Halaman
1	28 Maret 2023	Call for paper submission (**La,mpiran 1**)	2
2	14 April 2023	Pengiriman paper (**Lampiran 2**)	4
3	14 April 2023	Email penerimaan paper submission (**Lampiran 3)**	18
4	11 Mei 2023	Pemberitahuan hasil telaah paper (**Lampiran 4**)	58
5	24 Mei 2023	Pengiriman paper yang sudah direvisi, bukti pembayaran, dan cover letters (**Lampiran 5)**	62
6	9 Juni 2023	Pemberitahuan jadwal publikasi secara online (**Lampiran 6**)	75
7	19 Juni 2023	Artikel dipublikasikan online (**Lampiran 7)**	80

Lampiran 1. Call for paper submission

Journal of Southwest Jiaotong University is accepting submissions for Volume 58, Issue 2, 2023.

editor@jsju.org <editor@jsju.org>

Sel 28/03/2023 11.34

Journal of Southwest Jiaotong University is accepting submissions for Volume 58, Issue 2, 2023. Deadline: April 15, 2023

As the world's leading open access journal, we are proud to provide scholars the opportunity to publish impactful insights faster.

We invite you to submit your papers. Please submit the topical previously unpublished papers through our Online Submission System http://jsju.org/index.php/journal/pages/view/papersubmission or directly to the Chief - Editor's e-mail: editor@jsju.org

The Journal of Southwest Jiaotong University charges Publication Fee at the rate of EUR 500.00. All articles published in our journal are open access and freely available online, immediately upon publication.

We recommend that the authors use the academic text editing service for the scientific articles, not just proofreading. Please use the American English option. We recommend the use of large, trusted companies with editors with a Ph.D. degree. You should also attach an editing certificate or use the editorial office services. Articles that native English speakers do not edit are not allowed to publication. The editorial team provides academic proofreading services at an additional cost of EUR 150–180 (word count depends) if the authors don't attach an official English editing certificate.

The Journal of Southwest Jiaotong University (ISSN 0258–2724) is covered by the following databases and archives:

Scopus 收录

EI Compendex

全国中文核心期刊

中国科技论文统计源期刊

中国科学引文数据库来源期刊.

The Journal of Southwest Jiaotong University is an international, interdisciplinary, peer-reviewed, open access journal, published bimonthly online by Science Press.

Open Access— free for readers, with article processing charges (APC) paid by authors or their institutions.

High visibility: indexed in many databases.

Subject area and category: multidisciplinary.

Rapid publication: manuscripts are peer-reviewed. The first decision is given to authors about 10–30 days after submission; acceptance for publication after revisions is done within 7–10 days (averages for articles published in this journal in 2022).

Recognition of reviewers: Reviewers who provide thorough review reports promptly receive vouchers that entitle them to a discount on the APC of their next publication.

Please read our published articles
http://jsju.org/index.php/journal/issue/archive.

If you have any questions, please do not hesitate to contact us via editor@jsju.org

Sincerely yours,
Editorial Office of Journal of Southwest Jiaotong University
http://jsju.org/index.php/journal/index

该杂志的出版商: Science Press
http://science-press.cn

Lamiran 2. Email Submit paper, Tanggal, 14 April 2023

MODELING REGRESSION ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM (RANFIS) FOR PANEL DATA

Tarno [a,] *, Di Asih I Maruddani [b], Yuciana Wilandari [c]

[a] Statistics Department, Faculty of Science and Mathematics, Universitas Diponegoro,

Semarang 50275, Indonesia, e-mail: tarno@lecturer.undip.ac.id

[b] Statistics Department, Faculty of Science and Mathematics, Universitas Diponegoro,

Semarang 50275, Indonesia, e-mail: maruddani@live.undip.ac.id

[c] Statistics Department, Faculty of Science and Mathematics, Universitas Diponegoro,

Semarang 50275, Indonesia, e-mail: yuciana.wilandari@gmail.com

Received: ▪ *Review:* ▪ *Accepted:* ▪ *Published*

Abstract

Panel data combines cross-sectional data and time-series data. Data on economic, business, social, and development issues are often presented in panel data. In constructing the panel data regression model, it is necessary to take various steps for testing the model specifications, including the Chow test and the Hausman test. The Chow test selects one of the two models, the Common Effect Model or Fixed Effect Model. Hausman test is used to compare Fixed Effect Model with Random Effect Model. This study aimed to construct a classical panel data regression model and the Regression Adaptive Neuro-Fuzzy Inference System (RANFIS). The RANFIS model is a regression model by applying fuzzy and Neural Network (NN) techniques expected to overcome the problem of uncertainty. The empirical study in this research is to construct a panel data regression model for the Human Development Index (HDI) in Central Java in 2017-2019. The variables involved were Junior High School Participation Rate, Senior High School Participation Rate, Number of Health Workers, Public Health Complaints, Population Growth Rate, Poverty Severity Index as predictor variables, and Human Development Index as response variable. Applying the classic panel data regression model, three factors that significantly affect HDI were obtained: the Junior High School Participation Rate, Public Health Complaints, and the Poverty Severity Index. These three variables were used as optimal inputs for the RANFIS modeling. Evaluation of model performance was measured based on the RMSE and MAPE values. Based on the ANFIS regression, the RMSE and MAPE values were 3.227 and 3.299, respectively.

Keywords: Panel Data Regression, Human Development Index, RANFIS

摘要 The authors may not translate the abstract and keywords into Chinese themselves.

关键词:

I. Introduction

Panel data combines cross-sectional data and time-series data [1]. Data on economic, business, social, and development issues are often presented in panel data. In constructing a suitable regression model for panel data, it is necessary to take various steps for model specification tests, including the Chow test, Hausman test, and the Lagrange Multiplier test. The Chow test selects one of two models, the Common Effect Model or the Fixed Effect Model. Hausman test is used to compare the models of Fixed Effect with Random Effect [2].

This study aimed to construct a suitable regression model for panel data. The regression model built was the classical panel data regression model and the Adaptive Neuro-Fuzzy Inference System (RANFIS) regression. The RANFIS model is a regression model applying fuzzy and Neural Network (NN) techniques expected to overcome the problem of uncertainty and nonlinearity in the data. The merging of these two methods aimed to obtain an accurate model. The fuzzy system is a universal approximator capable of classifying data with high uncertainty. At the same time, NN has good learning abilities on data.

The fuzzy system is a “universal approximator,” defined as techniques related to uncertainty based on fuzzy sets. The advantage of the system is that the developed model is characterized by linguistic interpretation abilities and rules that can be understood, verified, and developed [3], [4], [5]). Neural networks (NN) model is one example of a nonlinear model with a flexible functional form. It contains several parameters that cannot be interpreted as the parametric model. As a supervised machine learning method, NN provides a good framework for representing a relationship in data. Compared to other algorithms, NN has better adaptive ability, learning, and pattern non-stationary and nonlinear signals [6], [7].

The empirical study aimed to construct a panel data regression model, specifically to identify the factors that affect the Human Development Index (HDI). Human development intends to have more choices, especially in income, health, and education. HDI is a standard measure of human development set by the United Nations. HDI is formed through three essential variables: health, education, and decent living standards. According to the Central Bureau of Statistics Republic of Indonesia (2019a) [8], HDI is one way to measure the success of human development based on several fundamental components of life quality. To measure the health variable using the number of health workers and the percentage of people complaining about their health and seeking treatment. The education variable is measured by two indicators: the junior high school participation rate and the senior high school participation rate. The variable of decent living standard is measured by population growth and the severity of poverty.

To conduct a further study in this research, the variables identified for the empirical study were Junior High School Participation Rate, High School Participation Rate (Central Bureau of Statistic, 2019b) [9], Number of Health Workers, Public Health Complaints (Central Bureau of

Statistics, 2019c [10]), Growth Rate Population (Central Bureau of Statistics, 2019d [11]), and Poverty Severity Index as independent variables (predictors) (Central Bureau of Statistics, 2019e) [12], and Human Development Index (HDI) as response variables (dependent variable) (Central Bureau of Statistics, 2019a [8]). The data taken for the case study were from 35 regencies and cities in Central Java Province from 2017 to 2019. The modeling for panel data was carried out using the classical regression model and RANFIS. The estimation results using the two methods were compared with the level of accuracy based on the predicted MAPE value.

This study objective was to develop and apply a regression model for panel data: (1) Compile a classic panel data regression model for HDI data in Central Java, (2) Establish the ANFIS Regression model for HDI data in Central Java.

II. Theoretical framework

2.1.Panel Data Regression

Panel data is a combination of time series data and cross-section data. Regression using Panel data is called panel data regression model [1]. Baltagi (2005) developed panel data regression analysis with the following theoretical concepts.

- Panel Data Regression Model

Panel data combines cross-section data and time-series data, so the model can be written as follows.

$$Y_{it} = \alpha + \beta X_{it} + u_{it};$$
$$i = 1,2,\ldots,N;\ t = 1,2,\ldots,T.$$

where

i =1, 2,..., N are households, individuals, companies, or others showing the dimensions of cross-sectional data;

t =1,2 ,..., T represents the dimension of the time series data;

α : the scalar intercept coefficient

β: slope coefficient with dimensions K×1 where K is the number of independent variables

Y_{it}: dependent variable of individual i-th at time t

X_{it}: independent variable of individual i-th at time t

The residual component in the panel data regression model consists of a general residual component and a specific residual component. The general residual component is the residual component of the individual i-th and the general residual component of the time t. The specific residual component consists of the specific residual of individual i-th and time t. The specific residual component can be written as:

$$u_{it} = \mu_i + \lambda_t + \varepsilon_{it}$$

with

u_{it}: residual component for individual i-th at time t

μ_i: the specific influence of the individual i-th

λ_t: specific effect of time t

ε_{it}: residual for the individual i-th at time t

- Panel Data Regression Types

In estimating the panel regression model, there are three commonly used approaches: Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM) [1].

a. Common Effect Model (CEM)

The combined model is the simplest in panel data regression. The combined model ignores the individual-specific effect (μ_i) and the time-specific effect (λ_t) in the model. The model used follows the form of linear regression with the residual component u_{it} which only comes from the estimated residual component (ε_{it}). The parameter estimation method in this model is the same as the ordinary linear regression model, which uses the least-squares method (Gujarati 2004). The CEM model can be written as follows:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \varepsilon_{it};\ i = 1,2,\ldots N;\ t = 1,2\ldots,T \qquad (3)$$

b. Fixed Effect Model (FEM)

The fixed effect model is based on the assumption that the intercept between individual and time is different. However, the regression coefficient is constant for all individuals and time. In addition, this model assumes that there is a correlation between individual-specific effects (μ_i)and time-specific effects (λ_t) with independent variables. This assumption makes individual-specific effects (μ_i), and time-specific effects (λ_t) part of the intercept [1]. The FEM equation can be written as follows:

1. FEM with one-way residual component:

$$Y_{it} = \alpha + \mu_i + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \varepsilon_{it} \text{ with } \textstyle\sum_{i=1}^{N} \mu_i = 0,$$

or

$$Y_{it} = \alpha + \lambda_t + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \varepsilon_{it} \text{ with } \textstyle\sum_{t=1}^{T} \lambda_t = 0.$$

2. FEM model with two-way residual components:

$$Y_{it} = \alpha + \mu_i + \lambda_t + \beta_1 X_{1it} + \beta_2 X_{2it} + \cdots + \beta_k X_{kit} + \varepsilon_{it}$$

with $\sum_{i=1}^{N} \mu_i$=0 and $\sum_{t=1}^{T} \lambda_t$=0.

Intercept differences between the individual and time are caused by their different characteristics, so estimating parameters with these conditions uses the Least-Squares Dummy Variable (LSDV) method. The estimation results using the LSDV method produce an unbiased estimator. However, adding a large number of dummy variables will result in a significant loss of the degree of freedom resulting in the estimator inefficiency and multicollinearity due to too many predictable variables [1].

c. Random Effect Model (REM)

The random effect model assumes that there is no correlation between individual-specific effects (μ_i) and time-specific effects (λ_t) with independent variables. This assumption makes the residual component of the individual-specific effect (μ_i) and the time-specific effect (λ_t) included in the residual. The equation for the random effect model can be written as follows:

1. REM with one-way residual component:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \mu_i + \varepsilon_i$$

with $\mu_i \sim N(0, \sigma_i^2);\ cov(\mu_i, X_{it}) = 0$

or

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \lambda_t + u_t$$

with $\lambda_t \sim N(0, \sigma_t^2);\ cov(\lambda_t, X_{it}) = 0$

2. REM with two-way residual components:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_n X_{nit} + \mu_i + \lambda_t + w_{it}$$

with $\mu_i \sim N(0, \sigma_i^2);\ cov(\mu_i, X_{it}) = 0$ and

$$\lambda_t \sim N(0, \sigma_t^2);\ cov(\lambda_t, X_{it}) = 0$$

Panel Data Regression Estimation

In determining the estimation of the panel regression model, several tests were carried out to select the optimum estimation approach method. The first step in getting the desired model was the Chow test on the FEM estimation results; after proving that there was an individual effect, the Hausman test was carried out to determine between FEM and REM [1].

1. Chow Test

Chow test selects the two models between the Common Effect Model and the Fixed Effect Model. The assumption that each cross-sectional unit has the same behavior tends to be unrealistic, considering that each cross-sectional unit can have different behavior is the basis of the Chow test. In this test, the following hypotheses are carried out:

H_0: $\alpha_1 = \alpha_2 = \ldots = \alpha_N = \alpha$ (Common Effect Model)

H_1: there is at least one different intercept α_1 (Fixed Effect Model)

The basis for rejecting H_0 is to use F-statistics as follows (Baltagi, 2008):

$$Chow = \frac{(RSS1 - RSS2)/(N-1)}{RSS2/(NT - N - K)}$$

RSS1: residual sum of square of common effect model estimation results

RSS2: residual sum of square of fixed effect

model estimation results

N: number of cross-section unit

T: number of time series unit

K: number of independent variables

Chow Test statistics follow the distribution of F-statistics, namely $F_{(N-1,NT-N-K);\alpha}$. If the Chow statistic is greater than the F-table, there is sufficient evidence to reject H_0 and vice versa.

2. Hausman Test

Hausman test is used to compare Fixed Effect Model with Random Effect Model. The Hausman test is conducted when the Fixed Effect Model contains an element of trade-off, namely the loss of the degree of freedom element by including dummy variables and the Random Effect Model, which must heed the absence of assumptions violation of each component of the error. In this test, the following hypotheses are carried out:

H_0: corr(X_{it}, u_{it}) = 0 (Random Effect Model)

H_1: corr(X_{it}, u_{it}) ≠ 0 (Fixed Effect Model)

The basis for rejecting H_0 using Hausman Statistics is formulated as follows [13]:

$$\chi^2(K) = (b-\beta)'[Var(b-\beta)]-1(b-\beta)$$

with:

b: random effect coefficient

β: fixed effect coefficient

Hausman statistics spread Chi-Square, if the value of χ^2 is greater than $\chi^2_{(K, \alpha)}$ (K: number of independent variables) or P-Value < α, then there is sufficient evidence to reject H_0 and vice versa.

3. Lagrange Multiplier (LM) Test

This test is carried out to detect the presence of heteroscedasticity in the estimated model. The LM test hypotheses are as follows:

H_0: $\sigma_i^2 = \sigma^2$ (there is no heteroscedasticity)

H_1: $\sigma_i^2 \neq \sigma^2$ (ther is heteroscedasticity)

LM test statistics are as follows [13]:

$$LM = \frac{NT}{2(T-1)} + \sum_{i=1}^{N}\left(\frac{T^2\sigma_i^2}{\sigma^2}-1\right)^2$$

where:

T: number of time series unit

N: number of cross-section unit

σ_i^2: residual variance of the equation i

σ^2: residual variance of system equation

Conclusion H_0 is rejected if LM is greater than $\chi^2_{(1,\alpha)}$ which means heteroscedasticity occurs in the model. Thus, it must be estimated using the weight method: Cross-section weight.

4. Breusch Pagan Test

The Breusch Pagan test is an LM test to choose between a fixed effect model and a pooled regression model. The initial hypothesis is that the variance of the residuals in the fixed coefficient model is zero. The procedure is as follows [1]

Hypotheses

$H_0: \sigma_\mu^2 = 0$

$H_1: \sigma_\mu^2 \neq 0$

The test statistic used is the LM

$$LM = \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}(\sum_{t=1}^{T}\hat{u}_{it})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2}-1\right]$$

where

N: number of individuals

T: length of the time period

σ_μ^2: model residual variance

$\hat{u}_{it}$: residual estimation of the individual fixed coefficient model i $p$eriod t

If $LM > \chi^2_{(1,\alpha)}$ or p-value is less than the specified significance level, then H_0 is rejected. Thus, the random effect model is selected.

2.2. Adaptive Neuro-Fuzzy Inference System (ANFIS)

The materials and the sources used in this study cover all articles discussing ANFIS, which combines Neural Networks (NN) and Fuzzy Inference System (FIS). Before we discuss the procedure of ANFIS modeling, the most important material that should be described in this section is the structure of ANFIS networks. The NN architecture applied in ANFIS consists of five fixed layers [5], [14]. Without loss of generality, the architecture of ANFIS for modeling time-series data is given two input variables x_1, x_2 and single output variable y by assuming rule-base of Sugeno first order with two rules is as follows:

If $x_1\ is\ A_1\ and\ x_2\ is\ B_1$ then $y_1 = p_{11}x_1 + q_{12}x_2 + r_1$
If $x_1\ is\ A_2\ and\ x_2\ is\ B_2$ then $y_2 = p_{21}x_1 + q_{22}x_2 + r_2$

where

$x_i\ is\ A_j\ and\ x_2\ is\ B_1$; and $x_1\ is\ A_2\ and\ x_2\ is\ B_2$ as premise sections, whereas $y_1 = p_{11}x_1 + q_{12}x_2 + r_1$ and $y_2 = p_{21}x_1 + q_{22}x_2 + r_2$ as consequent sections; $p_{11}, q_{12}, r_1,\ p_{21}, q_{22}, r_2$ as linear parameters; A_1, B_1, A_2, B_2 as the nonlinear parameter. If the firing strength for two values y_1, y_2 are w_1, w_2 respectively then the output y can be expressed as in equation (1).

$$y = \overline{w}_1 y_1 + \overline{w}_2 y_2 \qquad (1)$$

where $\overline{w}_i = \frac{w_i}{\sum w_i}, i = 1,2.$

The structure of ANFIS networks (Figure 1) has five layers and can be explained as follows [5].

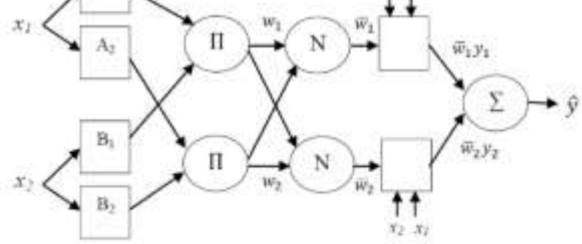

Figure 1. Structure of ANFIS Networks for Time Series Modeling [5]

Layer 1: Each neuron in this layer is adaptive to the parameters of an activation function. The output of each neuron is the membership degree of input. For example, the membership function of Generalized Bell is as follows:

$$\mu(x_i) = \frac{1}{1+\left|\frac{x_i - c_i}{a_i}\right|^{2b_i}}$$

where x_i is input and a_i, b_i and c_i are premise parameters [3], [4], [5].

Layer 2: Each neuron in this layer is a permanent neuron that is given the symbol Π, which is the product of all inputs in layer 1: $w_i = \mu_{Ai}(x_1) \times \mu_{Bi}(x_2),\ i = 1, 2.$

Each neuron output is called the firing strength of a rule [15], [16], [17], [18], [19].

Layer 3: Each neuron in this layer is a fixed neuron with the symbol N, which is the result of calculating the ratio of the i -firing strength to the total number of firing strengths in the second layer as follows: $\overline{w}_i = \frac{w_i}{\sum w_i}, i = 1, 2.$

The results of calculations at this layer are called normalized firing strength.

Layer 4: This layer is a neuron which is an adaptive neuron to an output:

$$\overline{w}_i y_i = \overline{w}_i(p_i x_1 + q_i x_2 + r_i)$$

where $\overline{w}_i$ is normalized firing strength in the third layer while p_i, q_i, and r_i are parameters in these neurons called consequent parameters.

Layer 5: This layer is a single neuron with the symbol Σ which is the sum of all outputs from the fourth layer, as follows:

$$y = \overline{w}_1 y_1 + \overline{w}_2 y_2, \text{ where}$$

$$\overline{w}_i = \frac{w_i}{\sum w_i}, i = 1,2.$$

3. Method of modeling

This research was based on a literature study. The initial step was to study in-depth and thoroughly from books and scientific articles that served as the basis for the new abstract system formation. We also examined supporting scientific articles that could be used in solving problems. At this stage, accuracy was needed in discussing supporting scientific articles, which were expected to solve the core problems. In addition to theoretical studies, applied studies were also carried out. In detail, this research method is described as follows.

3.1.Data Source

The data used in this study were the Human Development Index (%) and several factors that affect it in education, health, and population in 35 regencies and cities in Central Java Province from 2017 to 2019. All data were obtained from the Central Bureau of Statistics Central Java Province publications.

3.2.Research Variables

The variables used in this study were as follows:

a. Human Development Index (%) as response variable Y
b. Junior High School Participation Rate (%) as independent variable X1
c. High School Participation Rate (%) as independent variable X2
d. Number of Health Workers as independent variable X3
e. Public Health Complaints (%) as independent variable X4
f. Population Growth Rate as independent variable X5
g. Poverty Severity Index (%) as independent variable X6

3.3.Analysis Method

The data analysis method used in this research was modeling using panel data regression analysis, bootstrapping regression, and RANFIS. The following steps were taken to analyze the data.

3.3.1. Panel Data Regression Modeling

1. The general description of the data in data plots and descriptive statistics was seen.
2. The best panel data regression model to model the effect of the Junior High School Participation Rate, Senior High School Participation Rate, Number of Health Workers, Public Health Complaints, Population Growth Rate, and Poverty on the Human Development Index in Central Java was determined.
3. The Common Effect model, Fixed Effect model, and Random Effect model were estimated.
4. The best model was determined through the Chow test, Hausman test, and the Lagrange Multiplier (LM) test. If the Chow Test and Hausman Test showed the results of the Fixed Effect model, there was no need to proceed to the Lagrange Multiplier Test.
5. The classical assumptions of regression on the selected model were tested.
6. The significance of the panel data regression parameters, including Simultaneous Test (F-Test), Partial Test (t-Test), and the measure of the goodness of the model with R-Square, was tested.

3.3.2. ANFIS Regression Modeling

The estimation steps of the RANFIS model for panel data were as follows.

1. Preprocessing was performed by estimating the classical panel data regression model.
2. A new response data was formed based on the preprocessing results in step 1.
3. ANFIS modeling took new responses as targets with input variables as in panel data regression modeling.
4. Several clusters and membership functions for input variables were defined.
5. IF-THEN fuzzy rules were generated for output variables based on input, cluster, rule, and type of membership function. The IF-THEN fuzzy rules were formed using the First Order Sugeno model.
6. Fuzzy Inference System (FIS) training was conducted on an in-sample with a hybrid algorithm. The consequent parameters were estimated using a recursive LSE. The premise parameters were adjusted according to the backpropagation concept of gradient descent.
7. The predicted value in the in-sample was determined; the RMSE and MAPE were calculated.

4. Results and discussion

4.1. Regression of Panel Data

4.1.1. Common Effect Model

According to the data processing of Central Java HDI 2017-2019, the estimation of the combined model (Common Effect Model) was obtained as equation (2).

$$\hat{y}_{it} = 43.975 - 0.025x_{1it} + 0.270x_{2it} + 0.0008x_{3it} + 0.065x_{4it} + 3.122x_{5it} + 3.127x_{6it} \quad (2)$$

4.1.2. Fixed Effect Model

Fixed Effect Modeling of the Human Development Index was carried out with the RStudio program. The estimation result was obtained as equation (3):

$$\hat{y}_{it} = \hat{c}_i + 0.276x_{1it} + 0.031x_{2it} - 0.0002x_{3it} + 0.030x_{4it} - 2.785x_{5it} - 1.997x_{6it} \quad (3)$$

with the value of $\hat{c}_i$ owned by each region in Central Java presented in Table 1.

Tabel 1.

Intercept estimation $\hat{c}_i$ for Fixed Effect Model

i	Region	$\hat{c}_i$	i	Region	$\hat{c}_i$
1	Cilacap Regency	42.175	19	Kudus Regency	49.600
2	Banyumas Regency	46.306	20	Jepara Regency	46.815
3	Purbalingga Regency	44.943	21	Demak Regency	46.366
4	Banjarnegara Regency	41.898	22	Semarang Regency	48.394
5	Kebumen Regency	46.051	23	Temanggung Regency	43.084
6	Purworejo Regency	43.668	24	Kendal Regency	46.047
7	Wonosobo Regency	42.890	25	Batang Regency	42.462
8	Magelang Regency	44.018	26	Pekalongan Regency	44.475
9	Boyolali Regency	47.302	27	Pemalang Regency	40.184
10	Klaten Regency	46.261	28	Tegal Regency	40.945
11	Sukoharjo Regency	48.442	29	Brebes Regency	39.780
12	Wonogiri Regency	41.135	30	Magelang City	50.293
13	Karanganyar Regency	48.629	31	Surakarta City	53.419
14	Sragen Regency	45.788	32	Salatiga City	56.690
15	Grobogan Regency	43.474	33	Semarang City	59.217
16	Blora Regency	39.563	34	Pekalongan City	48.811
17	Rembang Regency	43.001	35	Tegal City	47.430
18	Pati Regency	44.184			

4.1.3. Random Effect Model

Random Effect Modeling of the Human Development Index was carried out with the help of the RStudio program. The estimation result was obtained as equation (4):

$$\hat{y}_{it} = \hat{c}_i + 33.515 + 0.337x_{1it} + 0.055x_{2it} - 0.00003x_{3it} + 0.029x_{4it} + 2.948x_{5it} - 2.172x_{6it} \quad (4)$$

with the value of $\hat{c}i$ owned by each region in Central Java presented in Table 2.

Tabel 2.

Intercept estimation $\hat{c}_i$ for Random Effect Model

i	Region	$\hat{c}_i$	i	Region	$\hat{c}_i$
1	Cilacap Regency	-2.375	19	Kudus Regency	49.600
2	Banyumas Regency	-0.652	20	Jepara Regency	46.815
3	Purbalingga Regency	1.986	21	Demak Regency	46.366
4	Banjarnegara Regency	-2.513	22	Semarang Regency	48.394

5	Kebumen Regency	-1.714	23	Temanggung Regency	43.084
6	Purworejo Regency	0.089	24	Kendal Regency	46.047
7	Wonosobo Regency	-0.615	25	Batang Regency	42.462
8	Magelang Regency	-2.817	26	Pekalongan Regency	44.475
9	Boyolali Regency	2.521	27	Pemalang Regency	40.184
10	Klaten Regency	2.058	28	Tegal Regency	40.945
11	Sukoharjo Regency	1.452	29	Brebes Regency	39.780
12	Wonogiri Regency	-2.467	30	Magelang City	50.293
13	Karanganyar Regency	1.612	31	Surakarta City	53.419
14	Sragen Regency	1.962	32	Salatiga City	56.690
15	Grobogan Regency	-0.795	33	Semarang City	6.609
16	Blora Regency	-4.377	34	Pekalongan City	2.183
17	Rembang Regency	-2.919	35	Tegal City	3.547
18	Pati Regency	-0.048			

4.2. Panel Data Model Selection

4.2.1. Selection of Common Effect Model and Fixed Effect Model with Chow Test

Calculation of the Chow test was carried out using RStudio program and obtained the value of F statistics is equal to 112.91 that is greater than F(0.05;5;98) = 2.30722 and p-value = 2.2e-16 is less than α =5%, so H_0 is rejected. Thus, there was an individual effect on Indonesia's energy consumption equation model, resulting in the Fixed Effect Model (FEM) as the appropriate model. Because the selected estimation model was the FEM model, the next test was the Hausman test, while the LM test did not need to be performed.

4.2.2 Selection of Fixed Effect Model and Random Effect Model with Hausman Test

The Hausman test calculation was carried out using the RStudio program and obtained p-value = 0.01444 that is less than α=5%, therefore, H_0 was rejected. Thus, the correct estimation of the regression model for the Human Development Index data in Central Java in 2017-2019 was to use the Fixed Effect Model.

4.2.3 Assumption Test

1. Residual Normality Assumption Test

The normality assumption test was done by using the Shapiro Wilk test. Using RStudio, a statistical p-value of 0.7562 was obtained because the p-value is greater than α=5%, the residuals of the Fixed Effect Model followed a normal distribution.

2. Autocorrelation Test

The non-autocorrelation assumption test was done by Run Test. Based on the results of using RStudio, the statistical value of the p-value test was 0.202. The p-value is greater than α=5%; hence, there was no serial correlation in the error component.

3. Heteroscedasticity Test

The Breusch Pagan Test is used to determine whether the residual covariance-variance of the Fixed Effect Model is homoscedastic or heteroscedastic. Based on the results using RStudio, the statistical value of the p-value test was 0.3768. The p-value is greater than α=5%, so the residual covariance structure of the Fixed Effect Model were homoscedastic.

4. Multicollinearity Test

Through the correlation test with the RStudio program, the correlation value between the independent variables was not too low. The value was less than 0.8, which H_0 was not rejected. It can be concluded that the resulting model did not contain elements of multicollinearity.

4.2.4 Parameter Significance Test

1. Simultaneous Test (F-Test)

This test is conducted to test the estimation of the Fixed Effect Model whether the independent variables together influence the dependent variable. Based on the RStudio program results, the F count value was 9.245e+04 with a p-value of 2.2e-16. Because the p-value is less than 0.05, the independent variables together significantly affected the dependent variable Human Development Index.

2. Partial Test (t-Test)

The t-test aims to see the significance of the influence of individual independent variables on the dependent variable by assuming other variables are constant. Based on the results of RStudio, the value of $|t\text{-statistic}|$ was obtained for variables c, x_1, x_4, and x_6 was greater than the value of t(0.025;103) which was 0.980103 or p-value is less than 0.05. So, it can be concluded that the variables c, x_1, x_4, and x_6 had a significant influence on the dependent variable Human Development Index in Central Java.

Based on the results of the RStudio program, the R-Squared value was 99.55%. The dependent variables were influenced by the Junior High School Participation Rate, Public Health Complaints, Poverty Severity Index, and regional factors with the equation (5):

$$\hat{y}_{it} = \hat{c}_i + 0.276x_{1it} + 0.030x_{4it} - 1.997x_{6it} \quad (5)$$

4.3 Modeling Human Development Index (HDI) Data with RANFIS

In order to obtain an estimate of HDI data regression parameters, the RANFIS method was used based on classical panel data regression preprocessing. In general, the stages of ANFIS regression modeling include: determining input variables, forming clusters (membership functions), and forming fuzzy rules. Preprocessing was done by applying classical panel data regression to determine the optimal input. The optimal input variables selected in the ANFIS regression modeling were: the HDI variable as the response, with the predictor variables being: Junior High School Participation Rate (X1), Public Health Complaints (X4), and Poverty Severity Index (X6). Based on the sample data, the following results were obtained.

In the preprocessing of panel data regression modeling on HDI data and its predictor variables, the predictor variables that had a significant effect on HDI were Junior High School Participation Rate, Public Health Complaints, and Poverty Severity Index. These predictor variables were then used as input in the ANFIS process.

After determining the input variables, the first step was to determine the membership function, the number of clusters, and the fuzzy rules that would be applied. This study determined clusters and rules using two methods, Fuzzy C-Means (FCM) and grid partition. Using a hybrid algorithm learning technique on in-sample data, the RMSE and MAPE values were obtained. To generate FIS using the FCM technique, the membership function (MF) used was the Gaussian function. In this technique, the number of rules was equal to the number of clusters determined. There were no combinations in the formation of the rule. Meanwhile, to generate FIS using the grid partition technique, each rule formed was a combination of the partition level for each input [20].

Optimal RANFIS modeling using FCM technique with two input variables x_1 and x_2 with two membership functions (clusters), two Sugeno rules of first-order can be formed as follows:

If x_1 *is* A_1 *and* x_2 *is* B_1 then $y_1 = p_{11}x_1 + q_{12}x_2 + r_1$

If x_1 *is* A_2 *and* x_2 *is* B_2 then $y_1 = p_{21}x_1 + q_{22}x_2 + r_2$

where A_1, B_1, A_2, B_2 as nonlinear parameters or premises, and $p_{11}, q_{12}, r_1, p_{21}, q_{22}, r_2$ as linear or consequent parameters.

If the firing strength for the two values of y_1 and y_2 is w_1, *and* w_2 then the output y could be determined as:

$$y = \frac{w_1y_1 + w_2y_2}{w_1 + w_2}.$$

In layer 1 in the RANFIS architecture, there are six groups of initial premise parameter values, with these values being used for the learning process. After obtaining the initial value of the premise parameters, the output generated in the first layer is the membership function of each input, $\mu_{A1}(x_1)$, $\mu_{A2}(x_1)$, $\mu_{B1}(x_2)$, and $\mu_{B2}(x_2)$. The membership function is used as input in layer 2, which produces the degree of activation of each rule. The optimal RANFIS has two rules, so layer 2 outputs are w_1 and w_2. Layer 2 output is used as input for layer 3, which will be normalized at the activation degree, then layer 3 output will be $\overline{w}_1$ and $\overline{w}_2$. The output of this layer is used as input in layer 4, which will produce linear parameters or consequent $p_{11}, q_{12}, r_1, p_{21}, q_{22}, r_2$ from the Recursive Least Squares Estimator (LSE) [20].

Based on Central Java HDI data as a case study, the RANFIS model obtained could be written as follows:

$$y = 1.059\overline{w}_{1,t}x_1 - 0.136\overline{w}_{1,t}x_4 + 0.166\overline{w}_{1,t}x_6 - 19.570\overline{w}_{1,t} + 0.506\overline{w}_{2,t}x_1 + 0.168\overline{w}_{2,t}x_4 - 4.9445\overline{w}_{2,t}x_6 + 16.034\overline{w}_{2,t}$$

where

$$\overline{w}_{1,t} = \frac{w_{1,t}}{w_{1,t}+w_{2,t}},$$

$$\overline{w}_{2,t} = \frac{w_{2,t}}{w_{1,t}+w_{2,t}},$$

$$w_{1,t} = exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-96.358}{2.073}\right)^2\right] + \left(\frac{x_4-56.833}{5.929}\right)^2 + \left(\frac{x_6-0.345}{0.226}\right)^2\right\},$$

$$w_{2,t} = exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-95.599}{2.241}\right)^2\right] + \left(\frac{x_4-45.907}{5.718}\right)^2 + \left(\frac{x_6-0.438}{0.209}\right)^2\right\}.$$

From the learning process using the hybrid algorithm, the RMSE, AIC, and BIC values were 3.227, respectively; 246.976; and 249.630; while the MAPE value was 3.299%.

5 Conclusion

Based on the panel data regression modeling procedure applied to the Human Development Index (HDI) data in Central Java in 2017-2019, an estimation of the panel data regression model of the Fixed Effect model was obtained. The Human Development Index variable could be explained from Junior High School Participation Rate, Public Health Complaints, and Poverty Severity Index. Using input variables selected through panel data regression, the optimal RANFIS model was obtained. The performance of the RANFIS model was evaluated using the RMSE and MAPE criteria. The RMSE and MAPE values were 3.227 and 3.299, respectively.

Acknowledgment

This paper is officially funded by grant of Faculty of Science and Mathematics, Universitas Diponegoro in 2021 with contract number 2164/UN7.5.8.2/PP/2021.

References

[1] BALTAGI, B.H. (2005) *Econometric Analysis of Panel Data*, third edition, John Wiley & Sons, Ltd.

[2] APRILIAWAN, D. TARNO, T., DAN YASIN, H. (2013) Pemodelan Laju Inflasi di Provinsi Jawa Tengah Menggunakan Regresi Data Panel, *Jurnal Gaussian*, vol. 2, no. 4, 311 – 321.

[3] JANG, J.-S, R. (1993) ANFIS: Adaptive-Network-Based Fuzzy Inference System, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 23, No.3, pp.665-68

[4] JANG, J.-S, R. (1996) Input Selection for ANFIS Learning, *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems*, Vol. 2, pp. 1493-1499.

[5] JANG, J.-S, R., SUN, C. –T., MIZUTANI, E. (1997) *Neuro-Fuzzy and Soft Computing: A computational Approach to Learning and Machine Intelligent*. Prentice Hall International.

[6] FAUSSET, L., (1994). Fundamental of Neural Networks: Architectures, Algorithms and Aplications, Prentice HallEnglewood Cliffs, New Jersey.

[7] HAYKIN, S., (1999). Neural Networks: A Comprehensive Foundation, Second Edition, Prentice Hall International, Inc.

[8] BPS JAWA TENGAH, (2019a). *Indeks Pembangunan Manusia (IPM) Menurut Jenis Kelamin dan Kabupaten/Kota di Jawa*

Tengah, 2014-2019. [Online] Available from https://jateng.bps.go.id/dynamictable/2019/01/23/58/indeks-pembangunan-manusia-ipm-menurut-jenis-kelamin-dan-kabupaten-kota-di-jawa-tengah-2014-2017.html [Accessed 14/10/20].

[9] BPS JAWA TENGAH, (2019b). *Angka Partisipasi Sekolah (APS) (Persen), 2017-2019.* [Online] Available from https://jateng.bps.go.id/indicator/28/71/1/angka-partisipasi-sekolah-aps-.html [Accessed 20/10/20].

[10] BPS JAWA TENGAH, (2019c). *Jumlah Tenaga Kesehatan Menurut Kabupaten/Kota di Provinsi Jawa Tengah, 2019.* [Online] Available from https://jateng.bps.go.id/ statictable/2020/07/20/1874/jumlah-tenaga-kesehatan-menurut-kabupaten-kota-di-provinsi-jawatengah-2019.html [Accessed 29/10/20].

[11] BPS JAWA TENGAH, (2019d). *Penduduk, Laju Pertumbuhan Penduduk, Distribusi Persentase Penduduk Kepadatan Penduduk, Rasio Jenis Kelamin Penduduk Menurut Kabupaten/Kota di Provinsi Jawa Tengah, 2010 dan 2019.* [Online] Available from https://jateng.bps.go.id/statictable/2020/06/11/1792/penduduk-laju-pertumbuhanpenduduk-distribusi-persentase-penduduk-kepadatan-penduduk-rasio-jenis-kelaminpenduduk-menurut-kabupaten-kota-di-provinsi-jawa-tengah-2010-dan-2019.html [Accessed 29/10/20].

[12] BPS JAWA TENGAH, (2019e). *Indeks Keparahan Kemiskinan (P2) (Persen), 2017-2019.* [Online] Available from https://jateng.bps.go.id/indicator/23/78/1/indeks-keparahan-kemiskinanp2-.html [Accessed 29/10/20].

[13] GREENE, Q. W. H., (2002). Econometric Analysis, Fifth Edition, New York University, Upper Saddle River, New Jersey 07458

[14] Tarno, T., Suparti, S., Ispriyanti, D. (2018). Modeling Cayenne Production Data in Central Java Using Adaptive Neuro Fuzzy Inference System (ANFIS) Model Assisted Statistics and Applications,13(1) p.45-52.

[15] FANI GKOUNTAKOU and BASIL PAPADOPOULOS. (2020) The Use of Fuzzy Linear Regression and ANFIS Methods to Predict the Compressive Strength of Cement, *Symmetry*, 12, 1295; doi:10.3390/sym12081295

[16] MOKARRAM, M., AMIN, H., AND KHOSRAVI, M.R. (2019) Using adaptive neuro‐fuzzy inference system and multiple linear regression to estimate orange taste, *Food Science & Nutrition, 7(10): 3176‐3184.*

[17] KARABOGA, D., KAYA, E. (2019). Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey. Artif Intell Rev 52, 2263–2293. https://doi.org/10.1007/s10462-017-9610-2

[18] SHAH, M.I.; ABUNAMA, T.; JAVED, M.F.; BUX, F.; ALDREES, A.; TARIQ, M.A.U.R.; MOSAVI, A. (2021). Modeling Surface Water Quality Using the Adaptive Neuro-Fuzzy Inference System Aided by Input Optimization. Sustainability, 13, 4576.

https:// doi.org/10.3390/su13084576

[19] HE, Z., WEN, X., LIU, H., & DU, J. (2014). A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. Journal of Hydrology. https://doi.org/10.1016/j.jhydrol.2013.11.054

[20] SARI. S.K., TARNO, SAFITRI, D. (2017). Pemilihan Input Model Regression Adaptive Neuro Fuzzy Inference System (RANFIS) Untuk Kajian Data IHSG, *Jurnal Gaussian,* Vol.6, No.3, 449-458

COVER LETTER

Title of the manuscript: Modeling Regression Adaptive Neuro-Fuzzy Inference System (RANFIS) For Panel Data
Abstract. Panel data combines cross-sectional data and time-series data. Data on economic, business, social, and development issues are often presented in panel data. In constructing the panel data regression model, it is necessary to take various steps for testing the model specifications, including the Chow test and the Hausman test. The Chow test selects one of the two models, the Common Effect Model or Fixed Effect Model. Hausman test is used to compare Fixed Effect Model with Random Effect Model. This study aimed to construct a classical panel data regression model and the Regression Adaptive Neuro-Fuzzy Inference System (RANFIS). The RANFIS model is a regression model by applying fuzzy
Keywords: Panel Data Regression, Human Development Index, RANFIS.

Type of manuscript (please specify):

- ☑ Research article
- ☐ Review article
- ☐ Brief report
- ☐ Short communication
- ☐ Research note

Full name and address of the Corresponding author: Tarno Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro Jl. Prof. Jacub Rais, Tembalang, Semarang 50275, Cental Java, Indonesia	
Telephone +6281325709047	Fax +622476486090
Email: tarno@lecturer.undip.ac.id	

☑ I hereby confirm that the manuscript was prepared in accordance with the instructions for authors of scientific publications, and that the content of this manuscript, or most of it, was not published in the journal indicated, and the manuscript was not submitted for publication elsewhere.

☑ I hereby confirm my consent to pay the Article Processing Charges (APC) EUR 450 in the case of the manuscript acceptance for publication. I am aware and hereby confirm that the APC is non-refundable.

☑ I hereby confirm my consent to make the payment for English Language Editing Services (EUR 150).

Publication fee payer details

____

____

14th April 2023

Signature of the Corresponding author **Date**

西南交通大学学报

第 58 卷 第 3 期
2023年 06 月

JOURNAL OF SOUTHWEST JIAOTONG UNIVERSITY

Vol. 58 No. 3
Jun. 2023

Copyright Agreement

Manuscript title:

Modeling Regression Adaptive Neuro-Fuzzy Inference System (RANFIS) For Panel Data

Full names of all authors: Tarno, Di Asih I Maruddani, and Yuciana Wilandari

Full name and address of the corresponding author:

Tarno
Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro
Jl. Prof. Jacub Rais, Tembalang, Semarang 50275, Cental Jawa, Indonesia

Telephone/Whatsap: +6281325709047   Fax: +622476480690

Email: tarno@lecturer.undip.ac.id

License Agreement

(1) Authors own all the copyright rights for the paper.
(2) Submitted manuscript is an original paper.
(3) Authors hereby grant the Issues of Journal of SWJTU with an exclusive, royalty-free, worldwide license to email the paper to all who will ask for it.
(4) All authors have made a significant contribution to the research and are ready to assume joint responsibility for the paper.
(5) All authors have seen and approved the manuscript in the final form as it is submitted for publication.
(6) This manuscript has not been published and also has neither been submitted nor considered for publication elsewhere
(7) The text, illustrations and any other materials, included into the manuscript, do not infringe any existing intellectual property rights or other rights of any person or entity.
(8) The editors of the Issues of Journal of SWJTU, its personnel or the Editorial Board members accept no responsibility for the quality of the idea expressed in this publication.

I am the Corresponding author and have full authority to enter into this agreement.

Full name, affiliation and position:

Tarno, Department of Statistics, Universitas Diponegoro, Lecturer

Signature: 

Date: 14th April 2023

Lampiran 3. Email penerimaan paper submission

Submission Received

jsjuorg@cpanel5.d.fozzy.com <jsjuorg@cpanel5.d.fozzy.com>
atas nama
Journal of Southwest Jiaotong University <editor@jsju.org>
Jum 14/04/2023 09.43

Kepada: Tarno <tarno@lecturer.undip.ac.id>

Thank you very much for uploading your manuscript to the Journal of Southwest Jiaotong University submission system. One of our editors will be in touch with you about the first decision within 10-14 days.

The following points were confirmed during submission:

1. Journal of Southwest Jiaotong University is an open access journal with publishing fees of EUR 500 for an accepted paper. This manuscript, if accepted, will be published under an open access Creative Commons CC BY license (https://creativecommons.org/licenses/by/4.0/), and I agree to pay the Article Processing Charges.

2. I hereby confirm my consent to make all payments for English Language Editing Services (EUR 150-200).

3. I understand that:

a. If previously published material is reproduced in my manuscript, I will provide proof that I have obtained the necessary copyright permission.

b. My manuscript is submitted on the understanding that it has not been published in or submitted to another peer-reviewed journal. Exceptions to this rule are papers containing material disclosed at conferences. I confirm that I will inform the journal editorial office if this is the case for my manuscript. I confirm that all authors are familiar with and agree with submission of the contents of the manuscript. The journal editorial office reserves the right to contact all authors to confirm this in case of doubt. I will provide email addresses for all authors and an institutional e-mail address for at least one of the co-authors, and specify the name, address and e-mail for invoicing purposes.

If you have any questions, please do not hesitate to contact the Editorial Office at editor@jsju.org

Kind regards,
Journal of Southwest Jiaotong University Editorial Office
http://jsju.org/index.php/journal/index
è¯¥æ□‚å¿—çš„å‡ºç‰ˆå•† (Our Publisher): Science Press
ç½‘å□€ http://science-press.cn/è¥¿å□—äº¤é€šå¤§å¦å¦æŠ¥/

*** This is an automatically generated email ***

Lampiran 4. Pemberitahuan hasil telaah paper, Tanggal, 11 Mei 2023

Re: Paper Submission

editor@jsju.org <editor@jsju.org>
Kam 11/05/2023 10.31
Kepada:Tarno <tarno@lecturer.undip.ac.id>

1 lampiran (4 MB)
Invoice 46_05-JS.pdf;

Dear Authors,

Greetings from Journal of Southwest Jiaotong University,
Thank you so much for being so interested in submitting your research to our Journal of Southwest Jiaotong University.

The independent experts in the field have reviewed your manuscript. We are pleased to inform you that the paper has been accepted for publication subject to minor revisions being made in response to the reviewers' comments given below.

You need to send the following to the editorial office via editor@jsju.org, if you want to publish in the Journal of Southwest Jiaotong University, Volume 58 (3), June 2023:
The article, strictly formatted according to the template recommendations for authors and correct the article on the reviewers' recommendations, responses to reviewers, cover letter and a proof of payment.

All revisions should be highlighted in the article in red. We hope you will do great.

Reviewers:
1 - Please add scientific novelty (100 words) in the abstract.
2 - Please present a schematic representation of the steps involved in conducting the study.
3 - Conclusion section needs to be described scientifically. Kindly frame it along the following lines:
i. Main findings of the present study (100 words)
ii. Comparison with other studies (100 words)
iii. Implications of study (50 words)
iv. Strengths and limitations (50 words).

Deadline for corrections and payment: May 15, 2023.

* Please write in the subject of the email: revised article, cover letter and payment proof.

We send you the invoice.

Note 1: Please note that APCs will not be refunded due to the retraction or correction of an article owing to author error or misconduct.

This is because has provided and author's and publishing services and is only able to recoup this investment through the APCs.

Note 2: Charges can be made by any of the following methods
Credit / debit card - charges can be made online using a secure payment gateway as soon as the manuscript has been editorially accepted. We send all the details through approval email.
Bank Transfer – charges may have remitted to the bank account.
Western Union – details will be given in approval email.

If it is convenient for you to pay by card/Western Union, please write us about it.

The Journal of Southwest Jiaotong University is covered by the following databases and archives:

收录

EI Compendex

全国中文核心期刊

中国科技论文统计源期刊

中国科学引文数据库来源期刊.

The Journal of Southwest Jiaotong University is an international, interdisciplinary, peer-reviewed, open access journal, published bimonthly online by Science Press.

Open Access— free for readers, with article processing charges (APC) paid by authors or their institutions.

High visibility: indexed in many databases.

Subject area and category: multidisciplinary.

Rapid publication: manuscripts are peer-reviewed. The first decision is given to authors about 10–30 days after submission; acceptance for publication after revisions is done within 7–10 days (averages for articles published in this journal in 2022).

Recognition of reviewers: Reviewers who provide thorough review reports promptly receive vouchers that entitle them to a discount on the APC of their next publication.

Please read our published articles
http://jsju.org/index.php/journal/issue/archive.

If you have any questions, please do not hesitate to contact us via
editor@jsju.org

Sincerely yours,
Editorial Office of Journal of Southwest Jiaotong University
http://jsju.org/index.php/journal/index

该杂志的出版商 (Our Publisher): Science Press
网址 http://science-press.cn

On 2023-04-14 18:43, JSJU wrote:
> Title of your paper: Modeling Regression Adaptive Neuro-Fuzzy Inference
> System (RANFIS) For Panel Data
> Corresponding Author's Email Address: tarno@lecturer.undip.ac.id
> Author(s): Tarno, Di Asih I Maruddani, and Yuciana Wilandari
> Keywords: Panel Data Regression, Human Development Index, RANFIS
> Abstract: Panel data combines cross-sectional data and time-series
> data. Data on economic, business, social, and development issues are
> often presented in panel data. In constructing the panel data
> regression model, it is necessary to take various steps for testing the
> model specifications, including the Chow test and the Hausman test. The
> Chow test selects one of the two models, the Common Effect Model or
> Fixed Effect Model. Hausman test is used to compare Fixed Effect Model
> with Random Effect Model.
> This study aimed to construct a classical panel data regression model
> and the Regression Adaptive Neuro-Fuzzy Inference System (RANFIS). The
> RANFIS model is a regression model by applying fuzzy and Neural Network
> (NN) techniques expected to overcome the problem of uncertainty.
> The empirical study in this research is to construct a Panel data
> regression model for the Human Development Index (HDI) in Central Java
> in 2017-2019. The variables involved were Junior High School
> Participation Rate, Senior High School Participation Rate, Number of
> Health Workers, Public Health Complaints, Population Growth Rate,
> Poverty Severity Index as predictor variables, and Human Development
> Index as response variables. Applying the classic Panel data regression
> model, three factors that significantly affect HDI were obtained: the
> Junior High School Participation Rate, Public Health Complaints, and
> the Poverty Severity Index. These three variables were used as optimal
> inputs for the RANFIS modeling. Evaluation of model performance was
> measured based on the RMSE and MAPE values. Based on the ANFIS
> regression, the RMSE and MAPE values were 3.227 and 3.299,
> respectively.

Lampiran 5. Pengiriman revised paper, paymenyt proof, and Cover letters, Tanggal, 25 Mei 2023

ISSN: 0258-2724 **DOI**：**10.35741/issn.0258-2724.57.6.1**

Category

Research Article

MODELING REGRESSION ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM (RANFIS) FOR PANEL DATA

Tarno [a, *], Di Asih I Maruddani [b], Yuciana Wilandari [c]

[a] Statistics Department, Faculty of Science and Mathematics, Universitas Diponegoro,

Semarang 50275, Indonesia, e-mail: tarno@lecturer.undip.ac.id

[b] Statistics Department, Faculty of Science and Mathematics, Universitas Diponegoro,

Semarang 50275, Indonesia, e-mail: maruddani@live.undip.ac.id

[c] Statistics Department, Faculty of Science and Mathematics, Universitas Diponegoro,

Semarang 50275, Indonesia, e-mail: yuciana.wilandari@gmail.com

Received: ▪ *Review:* ▪ *Accepted:* ▪ *Published*

Abstract

Panel data combines cross-sectional data and time-series data. Data on economic, business, social, and development issues are often presented in panel data. In constructing the panel data regression model, it is necessary to take various steps for testing the model specifications, including the Chow test and the Hausman test. This study aimed to construct a classical panel data regression model and the Regression Adaptive Neuro-Fuzzy Inference System (RANFIS). The RANFIS model is a regression model by applying fuzzy and Neural Network (NN) techniques expected to overcome the problem of uncertainty. One of the main problems in constructing an optimal RANFIS is selecting input variables. The input variables of RANFIS are selected based on the best classical regression. Those inputs are classified into optimal clusters which depend on the degree of fuzzy membership functions. The rule bases of RANFIS are determined based on optimal inputs and its clusters. The empirical study in this research is to construct a panel data regression model for the Human Development Index (HDI) in Central Java in 2017-2019. HDI is depend on several variables such as: the School Participation Rate, Number of Health Workers, Public Health Complaints, Population Growth Rate, Poverty Severity Index as predictor variables. Based on classical regression, three variables were used as optimal inputs for RANFIS modeling. Evaluation of model performance was measured based on the RMSE and MAPE values. Based on the RANFIS, the values of RMSE and MAPE were 3.227 and 3.299, respectively.

Keywords: Panel Data Regression, Human Development Index, RANFIS

摘要 The authors may not translate the abstract and keywords into Chinese themselves.

关键词:

I. Introduction

Panel data combines cross-sectional data and time-series data [1]. Data on economic, business, social, and development issues are often presented in panel data. In constructing a suitable regression model for panel data, it is necessary to take various steps for model specification tests, including the Chow test, Hausman test, and the Lagrange Multiplier test. The Chow test selects one of two models, the Common Effect Model or the Fixed Effect Model. Hausman test is used to compare the models of Fixed Effect with Random Effect [2].

This study aimed to construct a suitable regression model for panel data. The regression model built was the classical panel data regression model and the Adaptive Neuro-Fuzzy Inference System (RANFIS) regression. The RANFIS model is a regression model applying fuzzy and Neural Network (NN) techniques expected to overcome the problem of uncertainty and nonlinearity in the data. The merging of these two methods aimed to obtain an accurate model. The fuzzy system is a universal approximator capable of classifying data with high uncertainty. At the same time, NN has good learning abilities on data.

The fuzzy system is a “universal approximator,” defined as techniques related to uncertainty based on fuzzy sets. The advantage of the system is that the developed model is characterized by linguistic interpretation abilities and rules that can be understood, verified, and developed [3], [4], [5]). Neural networks (NN) model is one example of a nonlinear model with a flexible functional form. It contains several parameters that cannot be interpreted as the parametric model. As a supervised machine learning method, NN provides a good framework for representing a relationship in data. Compared to other algorithms, NN has better adaptive ability, learning, and pattern non-stationary and nonlinear signals [6], [7].

The empirical study aimed to construct a panel data regression model, specifically to identify the factors that affect the Human Development Index (HDI). Human development intends to have more choices, especially in income, health, and education. HDI is a standard measure of human development set by the United Nations. HDI is formed through three essential variables: health, education, and decent living standards. According to the Central Bureau of Statistics Republic of Indonesia (2019a) [8], HDI is one way to measure the success of human development based on several fundamental components of life quality. To measure the health variable using the number of health workers and the percentage of people complaining about their health and seeking treatment. The education variable is measured by two indicators: the junior high school participation rate and the senior high school participation rate. The variable of decent living standard is measured by population growth and the severity of poverty.

To conduct a further study in this research, the variables identified for the empirical study were Junior High School Participation Rate, High School Participation Rate (Central Bureau of Statistic, 2019b) [9], Number of Health Workers, Public Health Complaints (Central Bureau of Statistics, 2019c [10]), Growth Rate Population (Central Bureau of Statistics, 2019d [11]), and Poverty Severity Index as independent variables (predictors) (Central Bureau of Statistics, 2019e) [12], and Human Development Index (HDI) as response variables (dependent variable) (Central Bureau of Statistics, 2019a [8]). The data taken for the case study were from 35 regencies and cities in Central Java Province from 2017 to 2019. The modeling for panel data was carried out using the classical regression model and RANFIS. The estimation results using the two methods were compared with the level of accuracy based on the predicted MAPE value.

This study objective was to develop and apply a regression model for panel data: (1) Compile a classic panel data regression model for HDI data in Central Java, (2) Establish the ANFIS Regression model for HDI data in Central Java.

II. Theoretical framework

2.1.Panel Data Regression

Panel data is a combination of time series data and cross-section data. Regression using Panel data is called panel data regression model [1]. Baltagi (2005) developed panel data regression analysis with the following theoretical concepts.

- Panel Data Regression Model

Panel data combines cross-section data and time-series data, so the model can be written as follows.

$$Y_{it} = \alpha + \beta X_{it} + u_{it};$$
$$i = 1,2,\dots,N;\ t = 1,2,\dots,T.$$

where

i =1, 2,..., N are households, individuals, companies, or others showing the dimensions of cross-sectional data;

t =1,2 ,..., T represents the dimension of the time series data;

α : the scalar intercept coefficient

β: slope coefficient with dimensions K×1 where K is the number of independent variables

Y_{it}: dependent variable of individual i-th at time *t*

X_{it}: independent variable of individual i-th at time *t*

The residual component in the panel data regression model consists of a general residual component and a specific residual component. The general residual component is the residual component of the individual i-th and the general residual component of the time t. The specific residual component consists of the specific residual of individual i-th and time t. The specific residual component can be written as:

$$u_{it} = \mu_i + \lambda_t + \varepsilon_{it}$$

with

u_{it}: residual component for individual i-th at time t

μ_i: the specific influence of the individual i-th

λ_t: specific effect of time t

ε_{it}: residual for the individual i-th at time t

- Panel Data Regression Types

In estimating the panel regression model, there are three commonly used approaches: Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM) [1].

a. Common Effect Model (CEM)

The combined model is the simplest in panel data regression. The combined model ignores the individual-specific effect (μ_i) and the time-specific effect (λ_t) in the model. The model used follows the form of linear regression with the residual component u_{it} which only comes from the estimated residual component (ε_{it}). The parameter estimation method in this model is the same as the ordinary linear regression model, which uses the least-squares method (Gujarati 2004). The CEM model can be written as follows:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \varepsilon_{it};\ i = 1,2,\ldots N;\ t = 1,2\ldots,T \qquad (3)$$

b. Fixed Effect Model (FEM)

The fixed effect model is based on the assumption that the intercept between individual and time is different. However, the regression coefficient is constant for all individuals and time. In addition, this model assumes that there is a correlation between individual-specific effects (μ_i)and time-specific effects (λ_t) with independent variables. This assumption makes individual-specific effects (μ_i), and time-specific effects (λ_t) part of the intercept [1]. The FEM equation can be written as follows:

1. FEM with one-way residual component:

$$Y_{it} = \alpha + \mu_i + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \varepsilon_{it} \text{ with } \textstyle\sum_{i=1}^{N} \mu_i = 0,$$

or

$$Y_{it} = \alpha + \lambda_t + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \varepsilon_{it} \text{ with } \textstyle\sum_{t=1}^{T} \lambda_t = 0.$$

2. FEM model with two-way residual components:

$$Y_{it} = \alpha + \mu_i + \lambda_t + \beta_1 X_{1it} + \beta_2 X_{2it} + \cdots + \beta_k X_{kit} + \varepsilon_{it}$$

with $\sum_{i=1}^{N} \mu_i$=0 and $\sum_{t=1}^{T} \lambda_t$=0.

Intercept differences between the individual and time are caused by their different characteristics, so estimating parameters with these conditions uses the Least-Squares Dummy Variable (LSDV) method. The estimation results using the LSDV method produce an unbiased estimator. However, adding a large number of dummy variables will result in a significant loss of the degree of freedom resulting in the estimator inefficiency and multicollinearity due to too many predictable variables [1].

c. Random Effect Model (REM)

The random effect model assumes that there is no correlation between individual-specific effects (μ_i) and time-specific effects (λ_t) with independent variables. This assumption makes the residual component of the individual-specific effect (μ_i) and the time-specific effect (λ_t) included in the residual. The equation for the random effect model can be written as follows:

3. REM with one-way residual component:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \mu_i + \varepsilon_i$$

with $\mu_i \sim N(0, \sigma_i^2);\ cov(\mu_i, X_{it}) = 0$

or

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \lambda_t + u_t$$

with $\lambda_t \sim N(0, \sigma_t^2);\ cov(\lambda_t, X_{it}) = 0$

4. REM with two-way residual components:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_n X_{nit} + \mu_i + \lambda_t + w_{it}$$

with $\mu_i \sim N(0, \sigma_i^2);\ cov(\mu_i, X_{it}) = 0$ and

$$\lambda_t \sim N(0, \sigma_t^2);\ cov(\lambda_t, X_{it}) = 0$$

Panel Data Regression Estimation

In determining the estimation of the panel regression model, several tests were carried out to select the optimum estimation approach method. The first step in getting the desired model was the Chow test on the FEM estimation results; after proving that there was an individual effect, the Hausman test was carried out to determine between FEM and REM [1].

1. Chow Test

Chow test selects the two models between the Common Effect Model and the Fixed Effect Model. The assumption that each cross-sectional unit has the same behavior tends to be unrealistic, considering that each cross-sectional unit can have different behavior is the basis of the Chow test. In this test, the following hypotheses are carried out:

H_0: $\alpha_1 = \alpha_2 = ... = \alpha_N = \alpha$ (Common Effect Model)

H_1: there is at least one different intercept α_1 (Fixed Effect Model)

The basis for rejecting H_0 is to use F-statistics as follows (Baltagi, 2008):

$$Chow = \frac{(RSS1 - RSS2)/(N-1)}{RSS2/(NT - N - K)}$$

RSS1: residual sum of square of common effect model estimation results

RSS2: residual sum of square of fixed effect model estimation results

N: number of cross-section unit

T: number of time series unit

K: number of independent variables

Chow Test statistics follow the distribution of F-statistics, namely $F_{(N-1,NT-N-K);\alpha}$. If the Chow statistic is greater than the F-table, there is sufficient evidence to reject H_0 and vice versa.

2. Hausman Test

Hausman test is used to compare Fixed Effect Model with Random Effect Model. The Hausman test is conducted when the Fixed Effect Model contains an element of trade-off, namely the loss of the degree of freedom element by including dummy variables and the Random Effect Model, which must heed the absence of assumptions violation of each component of the error. In this test, the following hypotheses are carried out:

H_0: corr(X_{it}, u_{it}) = 0 (Random Effect Model)

H_1: corr$(X_{it}, u_{it}) \neq 0$ (Fixed Effect Model)

The basis for rejecting H_0 using Hausman Statistics is formulated as follows [13]:

$$\chi^2(K) = (b - \beta)'[Var(b - \beta)] - 1(b - \beta)$$

with:

b: random effect coefficient

β: fixed effect coefficient

Hausman statistics spread Chi-Square, if the value of χ^2 is greater than $\chi^2_{(K, \alpha)}$ (K: number of independent variables) or P-Value < α, then there is sufficient evidence to reject H_0 and vice versa.

3. Lagrange Multiplier (LM) Test

This test is carried out to detect the presence of heteroscedasticity in the estimated model. The LM test hypotheses are as follows:

H_0: $\sigma_i^2 = \sigma^2$ (there is no heteroscedasticity)

H_1: $\sigma_i^2 \neq \sigma^2$ (ther is heteroscedasticity)

LM test statistics are as follows [13]:

$$LM = \frac{NT}{2(T-1)} + \sum_{i=1}^{N} \left(\frac{T^2 \sigma_i^2}{\sigma^2} - 1 \right)^2$$

where:

T: number of time series unit

N: number of cross-section unit

σ_i^2: residual variance of the equation i

σ^2: residual variance of system equation

Conclusion H_0 is rejected if LM is greater than $\chi^2_{(1,\alpha)}$ which means heteroscedasticity occurs in the model. Thus, it must be estimated using the weight method: Cross-section weight.

4. Breusch Pagan Test

The Breusch Pagan test is an LM test to choose between a fixed effect model and a pooled regression model. The initial hypothesis is that the variance of the residuals in the fixed coefficient model is zero. The procedure is as follows [1]

Hypotheses

H_0: $\sigma_\mu^2 = 0$

H_1: $\sigma_\mu^2 \neq 0$

The test statistic used is the LM

$$LM = \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}(\sum_{t=1}^{T}\hat{u}_{it})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2} - 1\right]$$

where

N: number of individuals

T: length of the time period

σ_{μ}^2: model residual variance

$\hat{u}_{it}$: residual estimation of the individual fixed coefficient model i $p$eriod t

If $LM > \chi^2_{(1,\alpha)}$ or p-value is less than the specified significance level, then H_0 is rejected. Thus, the random effect model is selected.

2.2.Adaptive Neuro-Fuzzy Inference System (ANFIS)

The materials and the sources used in this study cover all articles discussing ANFIS, which combines Neural Networks (NN) and Fuzzy Inference System (FIS). Before we discuss the procedure of ANFIS modeling, the most important material that should be described in this section is the structure of ANFIS networks. The NN architecture applied in ANFIS consists of five fixed layers [5], [14]. Without loss of generality, the architecture of ANFIS for modeling time-series data is given two input variables x_1, x_2 and single output variable y by assuming rule-base of Sugeno first order with two rules is as follows:

If x_1 is A_1 and x_2 is B_1 then $y_1 = p_{11}x_1 + q_{12}x_2 + r_1$
If x_1 is A_2 and x_2 is B_2 then $y_2 = p_{21}x_1 + q_{22}x_2 + r_2$

where

x_i is A_j and x_2 is B_1; and x_1 is A_2 and x_2 is B_2 as premise sections, whereas $y_1 = p_{11}x_1 + q_{12}x_2 + r_1$ and $y_2 = p_{21}x_1 + q_{22}x_2 + r_2$ as consequent sections; $p_{11}, q_{12}, r_1,$ p_{21}, q_{22}, r_2 as linear parameters; A_1, B_1, A_2, B_2 as the nonlinear parameter. If the firing strength for two values y_1, y_2 are w_1, w_2 respectively then the output y can be expressed as in equation (1).

$$y = \overline{w}_1 y_1 + \overline{w}_2 y_2 \quad (1)$$

where $\overline{w}_i = \frac{w_i}{\sum w_i}, i = 1,2.$

The structure of ANFIS networks (Figure 1) has five layers and can be explained as follows [5].

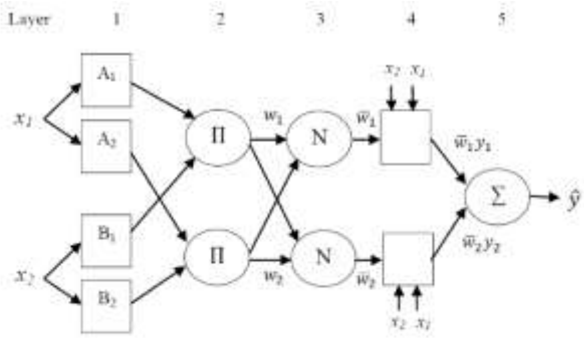

Figure 1. Structure of ANFIS Networks for Time Series Modeling [5]

Layer 1: Each neuron in this layer is adaptive to the parameters of an activation function. The output of each neuron is the membership degree of input. For example, the membership function of Generalized Bell is as follows:

$$\mu(x_i) = \frac{1}{1+\left|\frac{x_i - c_i}{a_i}\right|^{2b_i}}$$

where x_i is input and a_i, b_i and c_i are premise parameters [3], [4], [5].

Layer 2: Each neuron in this layer is a permanent neuron that is given the symbol Π, which is the product of all inputs in layer 1: $w_i = \mu_{Ai}(x_1) \times \mu_{Bi}(x_2),\ i = 1, 2.$

Each neuron output is called the firing strength of a rule [15], [16], [17], [18], [19].

Layer 3: Each neuron in this layer is a fixed neuron with the symbol N, which is the result of calculating the ratio of the i -firing strength to the total number of firing strengths in the second layer as follows: $\overline{w}_i = \frac{w_i}{\sum w_i}, i = 1, 2.$

The results of calculations at this layer are called normalized firing strength.

Layer 4: This layer is a neuron which is an adaptive neuron to an output:

$$\overline{w}_i y_i = \overline{w}_i(p_i x_1 + q_i x_2 + r_i)$$

where $\overline{w}_i$ is normalized firing strength in the third layer while p_i, q_i, and r_i are parameters in these neurons called consequent parameters.

Layer 5: This layer is a single neuron with the symbol Σ which is the sum of all outputs from the fourth layer, as follows:

$$y = \overline{w}_1 y_1 + \overline{w}_2 y_2, \text{ where}$$

$$\overline{w}_i = \frac{w_i}{\sum w_i}, i = 1,2.$$

3. Method of modeling

This research was based on a literature study. The initial step was to study in-depth and thoroughly from books and scientific articles that served as the basis for the new abstract system formation. We also examined supporting scientific articles that could be used in solving problems. At this stage, accuracy was needed in discussing supporting scientific articles, which were expected to solve the core problems. In addition to theoretical studies, applied studies were also carried out. In detail, this research method is described as follows.

3.1.Data Source

The data used in this study were the Human Development Index (%) and several factors that affect it in education, health, and population in 35 regencies and cities in Central Java Province from 2017 to 2019. All data were obtained from the Central Bureau of Statistics Central Java Province publications.

3.2.Research Variables

The variables used in this study were as follows:

a. Human Development Index (%) as response variable Y
b. Junior High School Participation Rate (%) as independent variable X1
c. High School Participation Rate (%) as independent variable X2
d. Number of Health Workers as independent variable X3
e. Public Health Complaints (%) as independent variable X4
f. Population Growth Rate as independent variable X5
g. Poverty Severity Index (%) as independent variable X6

3.3.Analysis Method

The data analysis method used in this research was modeling using panel data regression analysis, bootstrapping regression, and RANFIS. The following steps were taken to analyze the data.

3.3.1. Panel Data Regression Modeling

7. The general description of the data in data plots and descriptive statistics was seen.
8. The best panel data regression model to model the effect of the Junior High School Participation Rate, Senior High School Participation Rate, Number of Health Workers, Public Health Complaints, Population Growth Rate, and Poverty on the Human Development Index in Central Java was determined.
9. The Common Effect model, Fixed Effect model, and Random Effect model were estimated.
10. The best model was determined through the Chow test, Hausman test, and the Lagrange Multiplier (LM) test. If the Chow Test and Hausman Test showed the results of the Fixed Effect model, there was no need to proceed to the Lagrange Multiplier Test.
11. The classical assumptions of regression on the selected model were tested.
12. The significance of the panel data regression parameters, including Simultaneous Test (F-Test), Partial Test (t-Test), and the measure of the goodness of the model with R-Square, was tested.

The procedure of constructing data panel regression can be illustrated as figure 2.

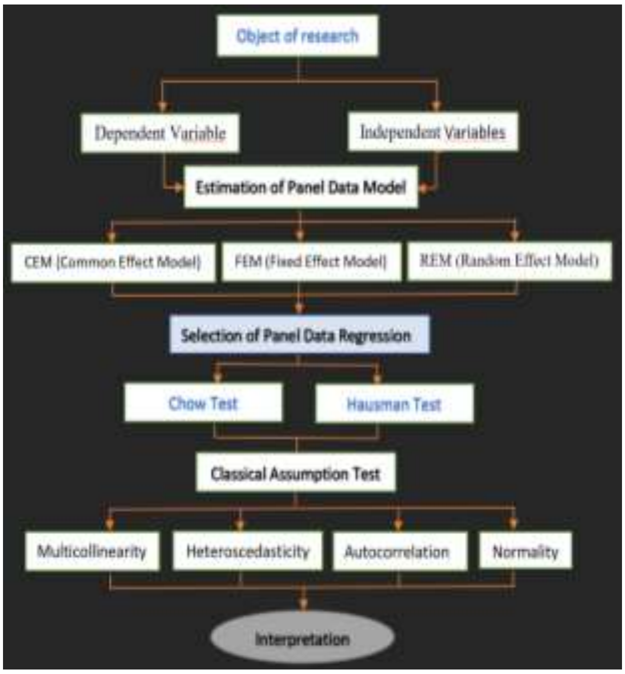

Figure 2. The procedure of constructing data panel regression

3.3.2. ANFIS Regression Modeling

The estimation steps of the RANFIS model for panel data were as follows.

8. Preprocessing was performed by estimating the classical panel data regression model.
9. A new response data was formed based on the preprocessing results in step 1.
10. ANFIS modeling took new responses as targets with input variables as in panel data regression modeling.
11. Several clusters and membership functions for input variables were defined.
12. IF-THEN fuzzy rules were generated for output variables based on input, cluster, rule, and type of membership function. The IF-THEN fuzzy rules were formed using the First Order Sugeno model.
13. Fuzzy Inference System (FIS) training was conducted on an in-sample with a hybrid algorithm. The consequent parameters were estimated using a recursive LSE. The premise parameters were adjusted according to the backpropagation concept of gradient descent.
14. The predicted value in the in-sample was determined; the RMSE and MAPE were calculated.

The procedure of constructing data panel RANFIS can be illustrated as figure 3.

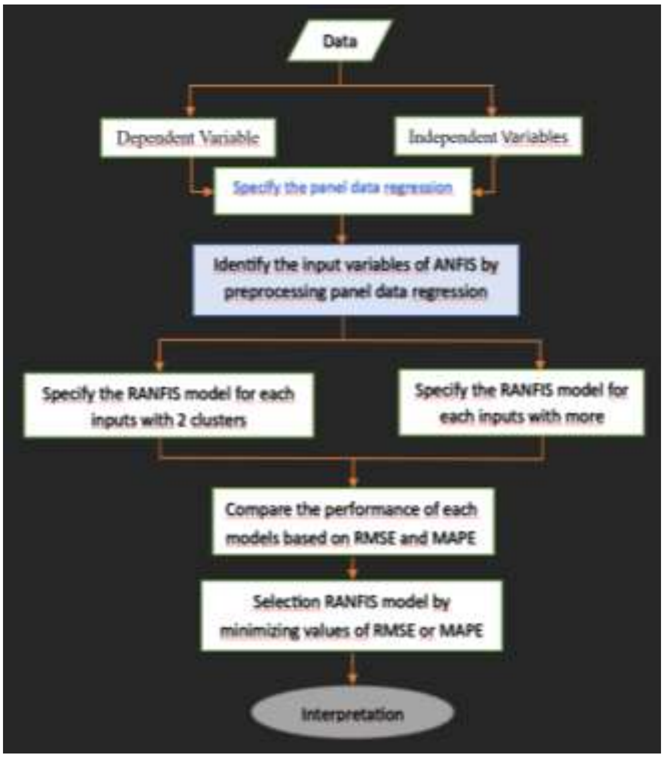

Figure 3. The procedure of constructing data panel RANFIS

4. Results and discussion

4.1. Regression of Panel Data

4.1.1. Common Effect Model

According to the data processing of Central Java HDI 2017-2019, the estimation of the combined model (Common Effect Model) was obtained as equation (2).

$$\hat{y}_{it} = 43.975 - 0.025x_{1it} + 0.270x_{2it} + 0.0008x_{3it} + 0.065x_{4it} + 3.122x_{5it} + 3.127x_{6it} \quad (2)$$

4.1.2. Fixed Effect Model

Fixed Effect Modeling of the Human Development Index was carried out with the RStudio program. The estimation result was obtained as equation (3):

$$\hat{y}_{it} = \hat{c}_i + 0.276x_{1it} + 0.031x_{2it} - 0.0002x_{3it} + 0.030x_{4it} - 2.785x_{5it} - 1.997x_{6it} \quad (3)$$

with the value of $\hat{c}_i$ owned by each region in Central Java presented in Table 1.

Tabel 1.

Intercept estimation $\hat{c}_i$ for Fixed Effect Model

i	Region	$\hat{c}_i$	i	Region	$\hat{c}_i$
1	Cilacap Regency	42.175	19	Kudus Regency	49.600
2	Banyumas Regency	46.306	20	Jepara Regency	46.815
3	Purbalingga Regency	44.943	21	Demak Regency	46.366
4	Banjarnegara Regency	41.898	22	Semarang Regency	48.394
5	Kebumen Regency	46.051	23	Temanggung Regency	43.084
6	Purworejo Regency	43.668	24	Kendal Regency	46.047
7	Wonosobo Regency	42.890	25	Batang Regency	42.462
8	Magelang Regency	44.018	26	Pekalongan Regency	44.475
9	Boyolali Regency	47.302	27	Pemalang Regency	40.184
10	Klaten Regency	46.261	28	Tegal Regency	40.945
11	Sukoharjo Regency	48.442	29	Brebes Regency	39.780
12	Wonogiri Regency	41.135	30	Magelang City	50.293
13	Karanganyar Regency	48.629	31	Surakarta City	53.419
14	Sragen Regency	45.788	32	Salatiga City	56.690
15	Grobogan Regency	43.474	33	Semarang City	59.217
16	Blora Regency	39.563	34	Pekalongan City	48.811
17	Rembang Regency	43.001	35	Tegal City	47.430
18	Pati Regency	44.184			

4.1.3. Random Effect Model

Random Effect Modeling of the Human Development Index was carried out with the help of the RStudio program. The estimation result was obtained as equation (4):

$$\hat{y}_{it} = \hat{c}_i + 33.515 + 0.337x_{1it} + 0.055x_{2it} - 0.00003x_{3it} + 0.029x_{4it} + 2.948x_{5it} - 2.172x_{6it} \quad (4)$$

with the value of $\hat{c}i$ owned by each region in Central Java presented in Table 2.

Tabel 2.

Intercept estimation $\hat{c}_i$ for Random Effect Model

i	Region	$\hat{c}_i$	i	Region	$\hat{c}_i$
1	Cilacap Regency	-2.375	19	Kudus Regency	49.600
2	Banyumas Regency	-0.652	20	Jepara Regency	46.815
3	Purbalingga Regency	1.986	21	Demak Regency	46.366
4	Banjarnegara Regency	-2.513	22	Semarang Regency	48.394
5	Kebumen Regency	-1.714	23	Temanggung Regency	43.084
6	Purworejo Regency	0.089	24	Kendal Regency	46.047
7	Wonosobo Regency	-0.615	25	Batang Regency	42.462
8	Magelang Regency	-2.817	26	Pekalongan Regency	44.475
9	Boyolali Regency	2.521	27	Pemalang Regency	40.184
10	Klaten Regency	2.058	28	Tegal Regency	40.945
11	Sukoharjo Regency	1.452	29	Brebes Regency	39.780
12	Wonogiri Regency	-2.467	30	Magelang City	50.293
13	Karanganyar Regency	1.612	31	Surakarta City	53.419
14	Sragen Regency	1.962	32	Salatiga City	56.690
15	Grobogan Regency	-0.795	33	Semarang City	6.609
16	Blora Regency	-4.377	34	Pekalongan City	2.183
17	Rembang Regency	-2.919	35	Tegal City	3.547

18	Pati Regency	-0.048

4.2.Panel Data Model Selection

4.2.1. Selection of Common Effect Model and Fixed Effect Model with Chow Test

Calculation of the Chow test was carried out using RStudio program and obtained the value of F statistics is equal to 112.91 that is greater than F(0.05;5;98) = 2.30722 and p-value = 2.2e-16 is less than α =5%, so H_0 is rejected. Thus, there was an individual effect on Indonesia's energy consumption equation model, resulting in the Fixed Effect Model (FEM) as the appropriate model. Because the selected estimation model was the FEM model, the next test was the Hausman test, while the LM test did not need to be performed.

5.2.2 Selection of Fixed Effect Model and Random Effect Model with Hausman Test

The Hausman test calculation was carried out using the RStudio program and obtained p-value = 0.01444 that is less than α=5%, therefore, H_0 was rejected. Thus, the correct estimation of the regression model for the Human Development Index data in Central Java in 2017-2019 was to use the Fixed Effect Model.

5.2.3 Assumption Test

5. Residual Normality Assumption Test

The normality assumption test was done by using the Shapiro Wilk test. Using RStudio, a statistical p-value of 0.7562 was obtained because the p-value is greater than α=5%, the residuals of the Fixed Effect Model followed a normal distribution.

6. Autocorrelation Test

The non-autocorrelation assumption test was done by Run Test. Based on the results of using RStudio, the statistical value of the p-value test was 0.202. The p-value is greater than α=5%; hence, there was no serial correlation in the error component.

7. Heteroscedasticity Test

The Breusch Pagan Test is used to determine whether the residual covariance-variance of the Fixed Effect Model is homoscedastic or heteroscedastic. Based on the results using RStudio, the statistical value of the p-value test was 0.3768. The p-value is greater than α=5%, so the residual covariance structure of the Fixed Effect Model were homoscedastic.

8. Multicollinearity Test

Through the correlation test with the RStudio program, the correlation value between the independent variables was not too low. The value was less than 0.8, which H_0 was not rejected. It can be concluded that the resulting model did not contain elements of multicollinearity.

5.2.4 Parameter Significance Test

3. Simultaneous Test (F-Test)

This test is conducted to test the estimation of the Fixed Effect Model whether the independent variables together influence the dependent variable. Based on the RStudio program results, the F count value was 9.245e+04 with a p-value of 2.2e-16. Because the p-value is less than 0.05, the independent variables together significantly affected the dependent variable Human Development Index.

4. Partial Test (t-Test)

The t-test aims to see the significance of the influence of individual independent variables on the dependent variable by assuming other variables are constant. Based on the results of RStudio, the value of $|t$-statistic$|$ was obtained for variables c, x_1, x_4, and x_6 was greater than the value of t(0.025;103) which was 0.980103 or p-value is less than 0.05. So, it can be concluded that the variables c, x_1, x_4, and x_6 had a significant influence on the dependent variable Human Development Index in Central Java.

Based on the results of the RStudio program, the R-Squared value was 99.55%. The dependent variables were influenced by the Junior High School Participation Rate, Public Health Complaints, Poverty Severity Index, and regional factors with the equation (5):

$$\hat{y}_{it} = \hat{c}_i + 0.276x_{1it} + 0.030x_{4it} - 1.997x_{6it} \quad (5)$$

5.3 Modeling Human Development Index (HDI) Data with RANFIS

In order to obtain an estimate of HDI data regression parameters, the RANFIS method was used based on classical panel data regression preprocessing. In general, the stages of ANFIS regression modeling include: determining input variables, forming clusters (membership functions), and forming fuzzy rules. Preprocessing was done by applying classical panel data regression to determine the optimal input. The optimal input variables selected in the ANFIS regression modeling were: the HDI variable as the response, with the predictor variables being: Junior High School Participation Rate (X1), Public Health Complaints (X4), and Poverty Severity Index (X6). Based on the sample data, the following results were obtained.

In the preprocessing of panel data regression modeling on HDI data and its predictor variables, the predictor variables that had a significant effect on HDI were Junior High School Participation Rate, Public Health Complaints, and Poverty Severity Index. These predictor variables were then used as input in the ANFIS process. After determining the input variables, the first step was to determine the membership function, the number of clusters, and the fuzzy rules that would be applied. This study determined clusters and rules using two methods, Fuzzy C-Means (FCM) and grid partition. Using a hybrid algorithm learning technique on in-sample data, the RMSE and MAPE values were obtained. To generate FIS using the FCM technique, the membership function (MF) used was the Gaussian function. In this technique, the number of rules was equal to the number of clusters determined. There were no combinations in the formation of the rule. Meanwhile, to generate FIS using the grid partition technique, each rule formed was a combination of the partition level for each input [20].

Optimal RANFIS modeling using FCM technique with two input variables x_1 and x_2 with two membership functions (clusters), two Sugeno rules of first-order can be formed as follows:

If x_1 $is\ A_1\ and\ x_2\ is\ B_1$ then $y_1 = p_{11}x_1 + q_{12}x_2 + r_1$

If x_1 $is\ A_2\ and\ x_2\ is\ B_2$ then $y_1 = p_{21}x_1 + q_{22}x_2 + r_2$

where A_1, B_1, A_2, B_2 as nonlinear parameters or premises, and $p_{11}, q_{12}, r_1, p_{21}, q_{22}, r_2$ as linear or consequent parameters.

If the firing strength for the two values of y_1 and y_2 is $w_1, and\ w_2$ then the output y could be determined as:

$$y = \frac{w_1y_1+w_2y_2}{w_1+w_2}.$$

In layer 1 in the RANFIS architecture, there are six groups of initial premise parameter values, with these values being used for the learning process. After obtaining the initial value of the premise parameters, the output generated in the first layer is the membership function of each input, $\mu_{A1}(x_1)$, $\mu_{A2}(x_1)$, $\mu_{B1}(x_2)$, and $\mu_{B2}(x_2)$. The membership function is used as input in layer 2, which produces the degree of activation of each rule. The optimal RANFIS has two rules, so layer 2 outputs are w_1 and w_2. Layer 2 output is used as input for layer 3, which will be normalized at the activation degree, then layer 3 output will be $\overline{w}_1$ and $\overline{w}_2$. The output of this layer is used as input in layer 4, which will produce linear parameters or consequent $p_{11}, q_{12}, r_1, p_{21}, q_{22}, r_2$ from the Recursive Least Squares Estimator (LSE) [20].

Based on Central Java HDI data as a case study, the RANFIS model obtained could be written as follows:

$$y = 1.059\overline{w}_{1,t}x_1 - 0.136\overline{w}_{1,t}x_4$$
$$+0.166\overline{w}_{1,t}x_6 - 19.570\overline{w}_{1,t}$$
$$+0.506\overline{w}_{2,t}x_1 + 0.168\overline{w}_{2,t}x_4$$
$$-4.9445\overline{w}_{2,t}x_6 + 16.034\overline{w}_{2,t}$$

where

$$\overline{w}_{1,t} = \frac{w_{1,t}}{w_{1,t}+w_{2,t}},$$

$$\overline{w}_{2,t} = \frac{w_{2,t}}{w_{1,t}+w_{2,t}},$$

$$w_{1,t} = exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-96.358}{2.073}\right)^2\right] + \left(\frac{x_4-56.833}{5.929}\right)^2 + \left(\frac{x_6-0.345}{0.226}\right)^2\right\},$$

$$w_{2,t} = exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-95.599}{2.241}\right)^2\right] + \left(\frac{x_4-45.907}{5.718}\right)^2 + \left(\frac{x_6-0.438}{0.209}\right)^2\right\}.$$

From the learning process using the hybrid algorithm, the RMSE, AIC, and BIC values were 3.227, respectively; 246.976; and 249.630; while the MAPE value was 3.299%.

6 Conclusion

Based on the panel data regression modeling procedure applied to the Human Development Index (HDI) data in Central Java in 2017-2019, an estimation of the panel data regression model of the Fixed Effect model was obtained. The HDI variable could be explained from Junior High School Participation Rate, Public Health Complaints, and Poverty Severity Index. Using the input variables selected through panel data regression, the optimal RANFIS model was obtained. The RANFIS optimal has three inputs with 2 clusters (membership functions). The performance of the RANFIS model was evaluated using the RMSE and MAPE criteria. The RMSE and MAPE values were 3.227 and 3.299, respectively. The RANFIS model performs well to apply for nonlinear data containing uncertainty.

Acknowledgment

This paper is officially funded by grant of Faculty of Science and Mathematics, Universitas Diponegoro in 2021 with contract number 2164/UN7.5.8.2/PP/2021.

References

[1] BALTAGI, B.H. (2005) *Econometric Analysis of Panel Data*, third edition, John Wiley & Sons, Ltd.

[2] APRILIAWAN, D. TARNO, T., DAN YASIN, H. (2013) Pemodelan Laju Inflasi di Provinsi Jawa Tengah Menggunakan Regresi Data Panel, *Jurnal Gaussian*, vol. 2, no. 4, 311 – 321.

[3] JANG, J.-S, R. (1993) ANFIS: Adaptive-Network-Based Fuzzy Inference System, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 23, No.3, pp.665-68

[4] JANG, J.-S, R. (1996) Input Selection for ANFIS Learning, *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems*, Vol. 2, pp. 1493-1499.

[5] JANG, J.-S, R., SUN, C. –T., MIZUTANI, E. (1997) *Neuro-Fuzzy and Soft Computing: A computational Approach to Learning and Machine Intelligent*. Prentice Hall International.

[6] FAUSSET, L., (1994). Fundamental of Neural Networks: Architectures, Algorithms and Aplications, Prentice HallEnglewood Cliffs, New Jersey.

[7] HAYKIN, S., (1999). Neural Networks: A Comprehensive Foundation, Second Edition, Prentice Hall International, Inc.

[8] BPS JAWA TENGAH, (2019a). *Indeks Pembangunan Manusia (IPM) Menurut Jenis Kelamin dan Kabupaten/Kota di Jawa Tengah, 2014-2019.* [Online] Available from https://jateng.bps.go.id/dynamictable/2019/01/23/58/indeks-pembangunan-manusia-ipm-menurut-jenis-kelamin-dan-kabupaten-kota-di-jawa-tengah-2014-2017.html [Accessed 14/10/20].

[9] BPS JAWA TENGAH, (2019b). *Angka Partisipasi Sekolah (APS) (Persen), 2017-2019.* [Online] Available from https://jateng.bps.go.id/indicator/28/71/1/angka-partisipasi-sekolah-aps-.html [Accessed 20/10/20].

[10] BPS JAWA TENGAH, (2019c). *Jumlah Tenaga Kesehatan Menurut Kabupaten/Kota di Provinsi Jawa Tengah, 2019.* [Online] Available from https://jateng.bps.go.id/ statictable/2020/07/20/1874/jumlah-tenaga-kesehatan-menurut-kabupaten-kota-

di-provinsi-jawatengah-2019.html [Accessed 29/10/20].
[11] BPS JAWA TENGAH, (2019d). *Penduduk, Laju Pertumbuhan Penduduk, Distribusi Persentase Penduduk Kepadatan Penduduk, Rasio Jenis Kelamin Penduduk Menurut Kabupaten/Kota di Provinsi Jawa Tengah, 2010 dan 2019.* [Online] Available from https://jateng.bps.go.id/statictable/2020/06/11/1792/penduduk-laju-pertumbuhanpenduduk-distribusi-persentase-penduduk-kepadatan-penduduk-rasio-jenis-kelaminpenduduk-menurut-kabupaten-kota-di-provinsi-jawa-tengah-2010-dan-2019.html [Accessed 29/10/20].
[12] BPS JAWA TENGAH, (2019e). *Indeks Keparahan Kemiskinan (P2) (Persen), 2017-2019.* [Online] Available from https://jateng.bps.go.id/indicator/23/78/1/indeks-keparahan-kemiskinanp2-.html [Accessed 29/10/20].

[13] GREENE, Q. W. H., (2002). Econometric Analysis, Fifth Edition, New York University, Upper Saddle River, New Jersey 07458

[14] Tarno, T., Suparti, S., Ispriyanti, D. (2018). Modeling Cayenne Production Data in Central Java Using Adaptive Neuro Fuzzy Inference System (ANFIS) Model Assisted Statistics and Applications,13(1) p.45-52.
[15] FANI GKOUNTAKOU and BASIL PAPADOPOULOS. (2020) The Use of Fuzzy Linear Regression and ANFIS Methods to Predict the Compressive Strength of Cement, *Symmetry*, 12, 1295; doi:10.3390/sym12081295
[16] MOKARRAM, M., AMIN, H., AND KHOSRAVI, M.R. (2019) Using adaptive neuro‐fuzzy inference system and multiple linear regression to estimate orange taste, *Food Science & Nutrition, 7(10): 3176‐3184.*
[17] KARABOGA, D., KAYA, E. (2019). Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey. Artif Intell Rev 52, 2263–2293. https://doi.org/10.1007/s10462-017-9610-2

[18] SHAH, M.I.; ABUNAMA, T.; JAVED, M.F.; BUX, F.; ALDREES, A.; TARIQ, M.A.U.R.; MOSAVI, A. (2021). Modeling Surface Water Quality Using the Adaptive Neuro-Fuzzy Inference System Aided by Input Optimization. Sustainability, 13, 4576.

https:// doi.org/10.3390/su13084576

[19] HE, Z., WEN, X., LIU, H., & DU, J. (2014). A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. Journal of Hydrology. https://doi.org/10.1016/j.jhydrol.2013.11.054

[20] SARI. S.K., TARNO, SAFITRI, D. (2017). Pemilihan Input Model Regression Adaptive Neuro Fuzzy Inference System (RANFIS) Untuk Kajian Data IHSG, *Jurnal Gaussian,* Vol.6, No.3, 449-458

参考文:

COVER LETTER

Title of the manuscript: Modeling Regression Adaptive Neuro-Fuzzy Inference System (RANFIS) For Panel Data

Abstract. Panel data combines cross-sectional data and time-series data. Data on economic, business, social, and development issues are often presented in panel data. In constructing the panel data regression model, it is necessary to take various steps for testing the model specifications, including the Chow test and the Hausman test. This study aimed to construct a classical panel data regression model and the Regression Adaptive Neuro-Fuzzy Inference System (RANFIS). The RANFIS model is a regression model by applying fuzzy and Neural Network (NN) techniques expected to overcome the problem of uncertainty. One of the main problems in constructing an optimal RANFIS is selecting input variables. The input variables of RANFIS are selected based on the best classical regression. Those inputs are classified into optimal clusters which depend on the degree of fuzzy membership functions. The rule bases of RANFIS are determined based on optimal inputs and its clusters. The empirical study in this research is to construct a panel data regression model for the Human Development Index (HDI) in Central Java in 2017-2019. HDI is depend on several variables such as: the School Participation Rate, Number of Health Workers, Public Health Complaints, Population Growth Rate, Poverty Severity Index as predictor variables. Based on classical regression, three variables were used as optimal inputs for RANFIS modeling. Evaluation of model performance was measured based on the RMSE and MAPE values. Based on the RANFIS, the

Keywords: Panel Data Regression, Human Development Index, RANFIS.

Type of manuscript (please specify):

- ☑ Research article
- ☐ Review article
- ☐ Brief report
- ☐ Short communication
- ☐ Research note

Full name and address of the Corresponding author: Tarno
Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro
Jl. Prof. Jacub Rais, Tembalang, Semarang 50275, Cental Java, Indonesia

Telephone +6281325709047 | Fax +622476486090

Email: tarno@lecturer.undip.ac.id

☑ I hereby confirm that the manuscript was prepared in accordance with the instructions for authors of scientific publications, and that the content of this manuscript, or most of it, was not published in the journal indicated, and the manuscript was not submitted for publication elsewhere.

☑ I hereby confirm my consent to pay the Article Processing Charges (APC) EUR 450 in the case of the manuscript acceptance for publication. I am aware and hereby confirm that the APC is non-refundable.

☑ I hereby confirm my consent to make the payment for English Language Editing Services (EUR 150).

Publication fee payer details

____

____

25th May 2023

Signature of the Corresponding author **Date**

Copyright Agreement

Manuscript title:

Modeling Regression Adaptive Neuro-Fuzzy Inference System (RANFIS) For Panel Data

Full names of all authors: Tarno, Di Asih I Maruddani, and Yuciana Wilandari

Full name and address of the corresponding author:

Tarno
Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro
Jl. Prof. Jacub Rais, Tembalang, Semarang 50275, Cental Jawa, Indonesia

Telephone/Whatsap: +6281325709047 Fax: +622476480690

Email: tarno@lecturer.undip.ac.id

License Agreement

(9) Authors own all the copyright rights for the paper.
(10) Submitted manuscript is an original paper.
(11) Authors hereby grant the Issues of Journal of SWJTU with an exclusive, royalty-free, worldwide license to email the paper to all who will ask for it.
(12) All authors have made a significant contribution to the research and are ready to assume joint responsibility for the paper.
(13) All authors have seen and approved the manuscript in the final form as it is submitted for publication.
(14) This manuscript has not been published and also has neither been submitted nor considered for publication elsewhere
(15) The text, illustrations and any other materials, included into the manuscript, do not infringe any existing intellectual property rights or other rights of any person or entity.
(16) The editors of the Issues of Journal of SWJTU, its personnel or the Editorial Board members accept no responsibility for the quality of the idea expressed in this publication.

I am the Corresponding author and have full authority to enter into this agreement.

Full name, affiliation and position:

Tarno, Department of Statistics, Universitas Diponegoro, Lecturer

Signature:

Date: 25th May 2023

Lampiran 6. Pemberitahuan jadwal publikasi secara online, 9 Juni 2023

Re: Bls: Bls: Paper Submission

editor@jsju.org <editor@jsju.org>
Jum 09/06/2023 06.02
Kepada:Tarno <tarno@lecturer.undip.ac.id>

Dear Tarno,

Many thanks for your support of open access publishing.

We acknowledge APC payment for paper received successfully.

Please wait for publication during June.

Sincerely yours,
Editorial Office of Journal of Southwest Jiaotong University

Lampiran 7. Publikasi secara online, 19 Juni 2023

Journal of Southwest Jiaotong University is accepting submissions for Volume 58, Issue 2, 2023.

editor@jsju.org <editor@jsju.org>

Sel 28/03/2023 11.34

Journal of Southwest Jiaotong University is accepting submissions for Volume 58, Issue 2, 2023. Deadline: April 15, 2023

As the world's leading open access journal, we are proud to provide scholars the opportunity to publish impactful insights faster.

We invite you to submit your papers. Please submit the topical previously unpublished papers through our Online Submission System http://jsju.org/index.php/journal/pages/view/papersubmission or directly to the Chief - Editor's e-mail: editor@jsju.org

The Journal of Southwest Jiaotong University charges Publication Fee at the rate of EUR 500.00. All articles published in our journal are open access and freely available online, immediately upon publication.

We recommend that the authors use the academic text editing service for the scientific articles, not just proofreading. Please use the American English option. We recommend the use of large, trusted companies with editors with a Ph.D. degree. You should also attach an editing certificate or use the editorial office services. Articles that native English speakers do not edit are not allowed to publication. The editorial team provides academic proofreading services at an additional cost of EUR 150–180 (word count depends) if the authors don't attach an official English editing certificate.

The Journal of Southwest Jiaotong University (ISSN 0258–2724) is covered by the following databases and archives:

Scopus 收录

EI Compendex

全国中文核心期刊

中国科技论文统计源期刊

中国科学引文数据库来源期刊.

The Journal of Southwest Jiaotong University is an international, interdisciplinary, peer-reviewed, open access journal, published bimonthly online by Science Press.

Open Access— free for readers, with article processing charges (APC) paid by authors or their institutions.

High visibility: indexed in many databases.

Subject area and category: multidisciplinary.

Rapid publication: manuscripts are peer-reviewed. The first decision is given to authors about 10–30 days after submission; acceptance for publication after revisions is done within 7–10 days (averages for articles published in this journal in 2022).

Recognition of reviewers: Reviewers who provide thorough review reports promptly receive vouchers that entitle them to a discount on the APC of their next publication.

Please read our published articles
http://jsju.org/index.php/journal/issue/archive.

If you have any questions, please do not hesitate to contact us via
editor@jsju.org

Sincerely yours,
Editorial Office of Journal of Southwest Jiaotong University
http://jsju.org/index.php/journal/index

该杂志的出版商: Science Press
http://science-press.cn

西南交通大学学报

第 57 卷 第 6 期
2022 年 12 月

JOURNAL OF SOUTHWEST JIAOTONG UNIVERSITY

Vol. 57 No. 6
Dec. 2022

ISSN: 0258-2724 DOI：10.35741/issn.0258-2724.57.6.1

Category
Research Article

MODELING REGRESSION ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM (RANFIS) FOR PANEL DATA

Tarno [a,] *, Di Asih I Maruddani [b], Yuciana Wilandari [c]

[a] Statistics Department, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang 50275, Indonesia, e-mail: tarno@lecturer.undip.ac.id

[b] Statistics Department, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang 50275, Indonesia, e-mail: maruddani@live.undip.ac.id

[c] Statistics Department, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang 50275, Indonesia, e-mail: yuciana.wilandari@gmail.com

Received: ▪ Review: ▪ Accepted: ▪ Published

Abstract

Panel data combines cross-sectional data and time-series data. Data on economic, business, social, and development issues are often presented in panel data. In constructing the panel data regression model, it is necessary to take various steps for testing the model specifications, including the Chow test and the Hausman test. The Chow test selects one of the two models, the Common Effect Model or Fixed Effect Model. Hausman test is used to compare Fixed Effect Model with Random Effect Model.

This study aimed to construct a classical panel data regression model and the Regression Adaptive Neuro-Fuzzy Inference System (RANFIS). The RANFIS model is a regression model by applying fuzzy and Neural Network (NN) techniques expected to overcome the problem of uncertainty.

The empirical study in this research is to construct a panel data regression model for the Human Development Index (HDI) in Central Java in 2017-2019. The variables involved were Junior High School Participation Rate, Senior High School Participation Rate, Number of Health Workers, Public Health Complaints, Population Growth Rate, Poverty Severity Index as predictor variables, and Human Development Index as response variable. Applying the classic panel data regression model, three factors that significantly affect HDI were obtained: the Junior High School Participation Rate, Public Health Complaints, and the Poverty Severity Index. These three variables were used as optimal inputs for the RANFIS modeling. Evaluation of model performance was measured based on the RMSE and MAPE values. Based on the ANFIS regression, the RMSE and MAPE values were 3.227 and 3.299, respectively.

Keywords: Panel Data Regression, Human Development Index, RANFIS

摘要 The authors may not translate the abstract and keywords into Chinese themselves.

关键词:

I. INTRODUCTION

Panel data combines cross-sectional data and time-series data [1]. Data on economic, business, social, and development issues are often presented in panel data. In constructing a suitable regression model for panel data, it is necessary to take various steps for model specification tests, including the Chow test, Hausman test, and the Lagrange Multiplier test. The Chow test selects one of two models, the Common Effect Model or the Fixed Effect Model. Hausman test is used to compare the models of Fixed Effect with Random Effect [2].

This study aimed to construct a suitable regression model for panel data. The regression model built was the classical panel data regression model and the Adaptive Neuro-Fuzzy Inference System (RANFIS) regression. The RANFIS model is a regression model applying fuzzy and Neural Network (NN) techniques expected to overcome the problem of uncertainty and nonlinearity in the data. The merging of these two methods aimed to obtain an accurate model. The fuzzy system is a universal approximator capable of classifying data with high uncertainty. At the same time, NN has good learning abilities on data.

The fuzzy system is a "universal approximator," defined as techniques related to uncertainty based on fuzzy sets. The advantage of the system is that the developed model is characterized by linguistic interpretation abilities and rules that can be understood, verified, and developed [3], [4], [5]). Neural networks (NN) model is one example of a nonlinear model with a flexible functional form. It contains several parameters that cannot be interpreted as the parametric model. As a supervised machine learning method, NN provides a good framework for representing a relationship in data. Compared to other algorithms, NN has better adaptive ability, learning, and pattern non-stationary and nonlinear signals [6], [7].

The empirical study aimed to construct a panel data regression model, specifically to identify the factors that affect the Human Development Index (HDI). Human development intends to have more choices, especially in income, health, and education. HDI is a standard measure of human development set by the United Nations. HDI is formed through three essential variables: health, education, and decent living standards. According to the Central Bureau of Statistics Republic of Indonesia (2019a) [8], HDI is one way to measure the success of human development based on several fundamental components of life quality. To measure the health variable using the number of health workers and the percentage of people complaining about their health and seeking treatment. The education variable is measured by two indicators: the junior high school participation rate and the senior high school participation rate. The variable of decent living standard is measured by population growth and the severity of poverty.

To conduct a further study in this research, the variables identified for the empirical study were Junior High School Participation Rate, High School Participation Rate (Central Bureau of Statistic, 2019b) [9], Number of Health Workers, Public Health Complaints (Central Bureau of Statistics, 2019c [10]), Growth Rate Population (Central Bureau of Statistics, 2019d [11]), and Poverty Severity Index as independent variables (predictors) (Central Bureau of Statistics, 2019e) [12], and Human Development Index (HDI) as response variables (dependent variable) (Central Bureau of Statistics, 2019a [8]). The data taken for the case study were from 35 regencies and cities in Central Java Province from 2017 to 2019. The modeling for panel data was carried out using the classical regression model and RANFIS. The estimation results using the two methods were compared with the level of accuracy based on the predicted MAPE value.

This study objective was to develop and apply a regression model for panel data: (1) Compile a classic panel data regression model for HDI data in Central Java, (2) Establish the ANFIS Regression model for HDI data in Central Java.

II. THEORETICAL FRAMEWORK

2.1. Panel Data Regression

Panel data is a combination of time series data and cross-section data. Regression using Panel data is called panel data regression model [1]. Baltagi (2005) developed panel data regression analysis with the following theoretical concepts.

- Panel Data Regression Model

Panel data combines cross-section data and time-series data, so the model can be written as follows.

$$Y_{it} = \alpha + \beta X_{it} + u_{it};$$
$$i = 1,2, \dots, N;\ t = 1,2, \dots, T.$$

where

i =1, 2,..., N are households, individuals, companies, or others showing the dimensions of cross-sectional data;

t =1,2 , ..., T represents the dimension of the time series data;

α : the scalar intercept coefficient

β: slope coefficient with dimensions K×1 where K is the number of independent variables

Y_{it}: dependent variable of individual i-th at time t

X_{it}: independent variable of individual i-th at time t

The residual component in the panel data regression model consists of a general residual component and a specific residual component. The general residual component is the residual component of the individual i-th and the general residual component of the time t. The specific residual component consists of the specific residual of individual i-th and time t. The specific residual component can be written as:

$$u_{it} = \mu_i + \lambda_t + \varepsilon_{it}$$

with

u_{it}: residual component for individual i-th at time t

μ_i: the specific influence of the individual i-th

λ_t: specific effect of time t

ε_{it}: residual for the individual i-th at time t

- Panel Data Regression Types

In estimating the panel regression model, there are three commonly used approaches: Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM) [1].

a. Common Effect Model (CEM)

The combined model is the simplest in panel data regression. The combined model ignores the individual-specific effect (μ_i) and the time-specific effect (λ_t) in the model. The model used follows the form of linear regression with the residual component u_{it} which only comes from the estimated residual component (ε_{it}). The parameter estimation method in this model is the same as the ordinary linear regression model, which uses the least-squares method (Gujarati 2004). The CEM model can be written as follows:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \varepsilon_{it};\ i = 1,2,\ldots N;\ t = 1,2\ldots,T \qquad (3)$$

b. Fixed Effect Model (FEM)

The fixed effect model is based on the assumption that the intercept between individual and time is different. However, the regression coefficient is constant for all individuals and time. In addition, this model assumes that there is a correlation between individual-specific effects (μ_i) and time-specific effects (λ_t) with independent variables. This assumption makes individual-specific effects (μ_i), and time-specific effects (λ_t) part of the intercept [1]. The FEM equation can be written as follows:

1. FEM with one-way residual component:

$$Y_{it} = \alpha + \mu_i + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \varepsilon_{it} \text{ with } \textstyle\sum_{i=1}^{N} \mu_i = 0,$$

or

$$Y_{it} = \alpha + \lambda_t + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \varepsilon_{it} \text{ with } \textstyle\sum_{t=1}^{T} \lambda_t = 0.$$

2. FEM model with two-way residual components:

$$Y_{it} = \alpha + \mu_i + \lambda_t + \beta_1 X_{1it} + \beta_2 X_{2it} + \cdots + \beta_k X_{kit} + \varepsilon_{it}$$

with $\sum_{i=1}^{N} \mu_i$=0 and $\sum_{t=1}^{T} \lambda_t$=0.

Intercept differences between the individual and time are caused by their different characteristics, so estimating parameters with these conditions uses the Least-Squares Dummy Variable (LSDV) method. The estimation results using the LSDV method produce an unbiased estimator. However, adding a large number of dummy variables will result in a significant loss of the degree of freedom resulting in the estimator inefficiency and multicollinearity due to too many predictable variables [1].

c. Random Effect Model (REM)

The random effect model assumes that there is no correlation between individual-specific effects (μ_i) and time-specific effects (λ_t) with independent variables. This assumption makes the residual component of the individual-specific effect (μ_i) and the time-specific effect (λ_t) included in the residual. The equation for the random effect model can be written as follows:

1. REM with one-way residual component:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \mu_i + \varepsilon_i$$

with $\mu_i \sim N(0, \sigma_i^2);\ cov(\mu_i, X_{it}) = 0$

or

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \lambda_t + u_t$$

with $\lambda_t \sim N(0, \sigma_t^2);\ cov(\lambda_t, X_{it}) = 0$

2. REM with two-way residual components:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_n X_{nit} + \mu_i + \lambda_t + w_{it}$$

with $\mu_i \sim N(0, \sigma_i^2);\ cov(\mu_i, X_{it}) = 0$ and $\lambda_t \sim N(0, \sigma_t^2);\ cov(\lambda_t, X_{it}) = 0$

Panel Data Regression Estimation

In determining the estimation of the panel regression model, several tests were carried out to select the optimum estimation approach method. The first step in getting the desired model was the Chow test on the FEM estimation results; after proving that there was an individual effect, the Hausman test was carried out to determine between FEM and REM [1].

1. Chow Test

Chow test selects the two models between the Common Effect Model and the Fixed Effect Model. The assumption that each cross-sectional unit has the same behavior tends to be unrealistic,

considering that each cross-sectional unit can have different behavior is the basis of the Chow test. In this test, the following hypotheses are carried out:
H_0: $\alpha_1 = \alpha_2 = ... = \alpha_N = \alpha$ (Common Effect Model)
H_1: there is at least one different intercept α_1 (Fixed Effect Model)

The basis for rejecting H_0 is to use F-statistics as follows (Baltagi, 2008):

$$Chow = \frac{(RSS1 - RSS2)/(N-1)}{RSS2/(NT-N-K)}$$

RSS1: residual sum of square of common effect model estimation results
RSS2: residual sum of square of fixed effect model estimation results
N: number of cross-section unit
T: number of time series unit
K: number of independent variables

Chow Test statistics follow the distribution of F-statistics, namely $F_{(N-1,NT-N-K);\alpha}$. If the Chow statistic is greater than the F-table, there is sufficient evidence to reject H_0 and vice versa.

2. Hausman Test

Hausman test is used to compare Fixed Effect Model with Random Effect Model. The Hausman test is conducted when the Fixed Effect Model contains an element of trade-off, namely the loss of the degree of freedom element by including dummy variables and the Random Effect Model, which must heed the absence of assumptions violation of each component of the error. In this test, the following hypotheses are carried out:
H_0: corr(X_{it}, u_{it}) = 0 (Random Effect Model)
H_1: corr(X_{it}, u_{it}) ≠ 0 (Fixed Effect Model)
The basis for rejecting H_0 using Hausman Statistics is formulated as follows [13]:

$$\chi^2(K) = (b-\beta)'[Var(b-\beta)] - 1(b-\beta)$$

with:
b: random effect coefficient
β: fixed effect coefficient
Hausman statistics spread Chi-Square, if the value of χ^2 is greater than $\chi^2_{(K,\ \alpha)}$ (K: number of independent variables) or P-Value < α, then there is sufficient evidence to reject H_0 and vice versa.

3. Lagrange Multiplier (LM) Test

This test is carried out to detect the presence of heteroscedasticity in the estimated model. The LM test hypotheses are as follows:
H_0: $\sigma_i^2 = \sigma^2$ (there is no heteroscedasticity)
H_1: $\sigma_i^2 \neq \sigma^2$ (ther is heteroscedasticity)

LM test statistics are as follows [13]:

$$LM = \frac{NT}{2(T-1)} + \sum_{i=1}^{N}\left(\frac{T^2\sigma_i^2}{\sigma^2} - 1\right)^2$$

where:
T: number of time series unit
N: number of cross-section unit
σ_i^2: residual variance of the equation i
σ^2: residual variance of system equation
Conclusion H_0 is rejected if LM is greater than $\chi^2_{(1,\alpha)}$ which means heteroscedasticity occurs in the model. Thus, it must be estimated using the weight method: Cross-section weight.

4. Breusch Pagan Test

The Breusch Pagan test is an LM test to choose between a fixed effect model and a pooled regression model. The initial hypothesis is that the variance of the residuals in the fixed coefficient model is zero. The procedure is as follows [1]
Hypotheses
$H_0: \sigma_\mu^2 = 0$
$H_1: \sigma_\mu^2 \neq 0$
The test statistic used is the LM

$$LM = \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}(\sum_{t=1}^{T}\hat{u}_{it})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2} - 1\right]$$

where
N: number of individuals
T: length of the time period
σ_μ^2: model residual variance
$\hat{u}_{it}$: residual estimation of the individual fixed coefficient model i period t
If $LM > \chi^2_{(1,\alpha)}$ or p-value is less than the specified significance level, then H_0 is rejected. Thus, the random effect model is selected.

2.2. Adaptive Neuro-Fuzzy Inference System (ANFIS)

The materials and the sources used in this study cover all articles discussing ANFIS, which combines Neural Networks (NN) and Fuzzy Inference System (FIS). Before we discuss the procedure of ANFIS modeling, the most important material that should be described in this section is the structure of ANFIS networks. The NN architecture applied in ANFIS consists of five fixed layers [5], [14]. Without loss of generality, the architecture of ANFIS for modeling time-series data is given two input variables x_1, x_2 and single output variable y by assuming rule-base of Sugeno first order with two rules is as follows:

If x_1 *is* A_1 *and* x_2 *is* B_1 then $y_1 = p_{11}x_1 + q_{12}x_2 + r_1$

If x_1 *is* A_2 *and* x_2 *is* B_2 then $y_2 = p_{21}x_1 + q_{22}x_2 + r_2$

where
x_i *is* A_j *and* x_2 *is* B_1; and x_1 *is* A_2 *and* x_2 *is* B_2 as premise sections, whereas $y_1 = p_{11}x_1 + q_{12}x_2 + r_1$ and $y_2 = p_{21}x_1 + q_{22}x_2 + r_2$ as consequent sections; $p_{11}, q_{12}, r_1,\ p_{21}, q_{22}, r_2$ as linear parameters; A_1, B_1, A_2, B_2 as the nonlinear

parameter. If the firing strength for two values y_1, y_2 are w_1, w_2 respectively then the output y can be expressed as in equation (1).

$$y = \bar{w}_1 y_1 + \bar{w}_2 y_2 \quad (1)$$

where $\bar{w}_i = \frac{w_i}{\sum w_i}, i = 1,2.$

The structure of ANFIS networks (Figure 1) has five layers and can be explained as follows [5].

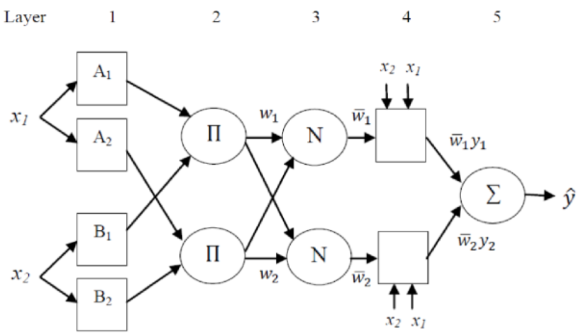

Figure 1. Structure of ANFIS Networks for Time Series Modeling [5]

Layer 1: Each neuron in this layer is adaptive to the parameters of an activation function. The output of each neuron is the membership degree of input. For example, the membership function of Generalized Bell is as follows:

$$\mu(x_i) = \frac{1}{1+\left|\frac{x_i-c_i}{a_i}\right|^{2b_i}}$$

where x_i is input and a_i, b_i and c_i are premise parameters [3], [4], [5].

Layer 2: Each neuron in this layer is a permanent neuron that is given the symbol Π, which is the product of all inputs in layer 1: $w_i = \mu_{Ai}(x_1) \times \mu_{Bi}(x_2),\ i = 1, 2$.

Each neuron output is called the firing strength of a rule [15], [16], [17], [18], [19].

Layer 3: Each neuron in this layer is a fixed neuron with the symbol N, which is the result of calculating the ratio of the i -firing strength to the total number of firing strengths in the second layer as follows: $\bar{w}_i = \frac{w_i}{\sum w_i}, i = 1, 2.$

The results of calculations at this layer are called normalized firing strength.

Layer 4: This layer is a neuron which is an adaptive neuron to an output:

$$\bar{w}_i y_i = \bar{w}_i(p_i x_1 + q_i x_2 + r_i)$$

where $\bar{w}_i$ is normalized firing strength in the third layer while p_i, q_i, and r_i are parameters in these neurons called consequent parameters.

Layer 5: This layer is a single neuron with the symbol Σ which is the sum of all outputs from the fourth layer, as follows:

$$y = \bar{w}_1 y_1 + \bar{w}_2 y_2, \text{ where}$$

$$\bar{w}_i = \frac{w_i}{\sum w_i}, i = 1,2.$$

3. METHOD OF MODELING

This research was based on a literature study. The initial step was to study in-depth and thoroughly from books and scientific articles that served as the basis for the new abstract system formation. We also examined supporting scientific articles that could be used in solving problems. At this stage, accuracy was needed in discussing supporting scientific articles, which were expected to solve the core problems. In addition to theoretical studies, applied studies were also carried out. In detail, this research method is described as follows.

3.1. Data Source

The data used in this study were the Human Development Index (%) and several factors that affect it in education, health, and population in 35 regencies and cities in Central Java Province from 2017 to 2019. All data were obtained from the Central Bureau of Statistics Central Java Province publications.

3.2. Research Variables

The variables used in this study were as follows:

a. Human Development Index (%) as response variable Y
b. Junior High School Participation Rate (%) as independent variable X1
c. High School Participation Rate (%) as independent variable X2
d. Number of Health Workers as independent variable X3
e. Public Health Complaints (%) as independent variable X4
f. Population Growth Rate as independent variable X5
g. Poverty Severity Index (%) as independent variable X6

3.3. Analysis Method

The data analysis method used in this research was modeling using panel data regression analysis, bootstrapping regression, and RANFIS. The following steps were taken to analyze the data.

3.3.1. Panel Data Regression Modeling

1. The general description of the data in data plots and descriptive statistics was seen.

2. The best panel data regression model to model the effect of the Junior High School Participation Rate, Senior High School Participation Rate, Number of Health Workers, Public Health Complaints, Population Growth Rate, and Poverty on the Human Development Index in Central Java was determined.
3. The Common Effect model, Fixed Effect model, and Random Effect model were estimated.
4. The best model was determined through the Chow test, Hausman test, and the Lagrange Multiplier (LM) test. If the Chow Test and Hausman Test showed the results of the Fixed Effect model, there was no need to proceed to the Lagrange Multiplier Test.
5. The classical assumptions of regression on the selected model were tested.
6. The significance of the panel data regression parameters, including Simultaneous Test (F-Test), Partial Test (t-Test), and the measure of the goodness of the model with R-Square, was tested.

3.3.2. ANFIS Regression Modeling

The estimation steps of the RANFIS model for panel data were as follows.

1. Preprocessing was performed by estimating the classical panel data regression model.
2. A new response data was formed based on the preprocessing results in step 1.
3. ANFIS modeling took new responses as targets with input variables as in panel data regression modeling.
4. Several clusters and membership functions for input variables were defined.
5. IF-THEN fuzzy rules were generated for output variables based on input, cluster, rule, and type of membership function. The IF-THEN fuzzy rules were formed using the First Order Sugeno model.
6. Fuzzy Inference System (FIS) training was conducted on an in-sample with a hybrid algorithm. The consequent parameters were estimated using a recursive LSE. The premise parameters were adjusted according to the backpropagation concept of gradient descent.
7. The predicted value in the in-sample was determined; the RMSE and MAPE were calculated.

4. RESULTS AND DISCUSSION

4.1. Regression of Panel Data

4.1.1. Common Effect Model

According to the data processing of Central Java HDI 2017-2019, the estimation of the combined model (Common Effect Model) was obtained as equation (2).

$$\hat{y}_{it} = 43.975 - 0.025x_{1it} + 0.270x_{2it} +0.0008x_{3it} + 0.065x_{4it} +3.122x_{5it} + 3.127x_{6it} \quad (2)$$

4.1.2. Fixed Effect Model

Fixed Effect Modeling of the Human Development Index was carried out with the RStudio program. The estimation result was obtained as equation (3):

$$\hat{y}_{it} = \hat{c}_i + 0.276x_{1it} + 0.031x_{2it} -0.0002x_{3it} + 0.030x_{4it} -2.785x_{5it} - 1.997x_{6it} \quad (3)$$

with the value of $\hat{c}_i$ owned by each region in Central Java presented in Table 1.

Tabel 1.
Intercept estimation $\hat{c}_i$ for Fixed Effect Model

i	Region	$\hat{c}_i$	*i*	Region	$\hat{c}_i$
1	Cilacap Regency	42.175	19	Kudus Regency	49.600
2	Banyumas Regency	46.306	20	Jepara Regency	46.815
3	Purbalingga Regency	44.943	21	Demak Regency	46.366
4	Banjarnegara Regency	41.898	22	Semarang Regency	48.394
5	Kebumen Regency	46.051	23	Temanggung Regency	43.084
6	Purworejo Regency	43.668	24	Kendal Regency	46.047
7	Wonosobo Regency	42.890	25	Batang Regency	42.462
8	Magelang Regency	44.018	26	Pekalongan Regency	44.475
9	Boyolali Regency	47.302	27	Pemalang Regency	40.184
10	Klaten Regency	46.261	28	Tegal Regency	40.945
11	Sukoharjo Regency	48.442	29	Brebes Regency	39.780
12	Wonogiri Regency	41.135	30	Magelang City	50.293
13	Karanganyar Regency	48.629	31	Surakarta City	53.419
14	Sragen Regency	45.788	32	Salatiga City	56.690
15	Grobogan Regency	43.474	33	Semarang City	59.217
16	Blora Regency	39.563	34	Pekalongan City	48.811
17	Rembang Regency	43.001	35	Tegal City	47.430
18	Pati Regency	44.184			

4.1.3. Random Effect Model

Random Effect Modeling of the Human Development Index was carried out with the help of the RStudio program. The estimation result was obtained as equation (4):

$$\hat{y}_{it} = \hat{c}_i + 33.515 + 0.337x_{1it} + 0.055x_{2it} - 0.00003x_{3it} + 0.029x_{4it} + 2.948x_{5it} - 2.172x_{6it} \quad (4)$$

with the value of $\hat{ci}$ owned by each region in Central Java presented in Table 2.

Tabel 2.
Intercept estimation $\hat{c}_i$ for Random Effect Model

i	Region	$\hat{c}_i$	i	Region	$\hat{c}_i$
1	Cilacap Regency	-2.375	19	Kudus Regency	49.600
2	Banyumas Regency	-0.652	20	Jepara Regency	46.815
3	Purbalingga Regency	1.986	21	Demak Regency	46.366
4	Banjarnegara Regency	-2.513	22	Semarang Regency	48.394
5	Kebumen Regency	-1.714	23	Temanggun g Regency	43.084
6	Purworejo Regency	0.089	24	Kendal Regency	46.047
7	Wonosobo Regency	-0.615	25	Batang Regency	42.462
8	Magelang Regency	-2.817	26	Pekalongan Regency	44.475
9	Boyolali Regency	2.521	27	Pemalang Regency	40.184
10	Klaten Regency	2.058	28	Tegal Regency	40.945
11	Sukoharjo Regency	1.452	29	Brebes Regency	39.780
12	Wonogiri Regency	-2.467	30	Magelang City	50.293
13	Karanganyar Regency	1.612	31	Surakarta City	53.419
14	Sragen Regency	1.962	32	Salatiga City	56.690
15	Grobogan Regency	-0.795	33	Semarang City	6.609
16	Blora Regency	-4.377	34	Pekalongan City	2.183
17	Rembang Regency	-2.919	35	Tegal City	3.547
18	Pati Regency	-0.048			

4.2. Panel Data Model Selection

4.2.1. Selection of Common Effect Model and Fixed Effect Model with Chow Test

Calculation of the Chow test was carried out using RStudio program and obtained the value of F statistics is equal to 112.91 that is greater than F(0.05;5;98) = 2.30722 and p-value = 2.2e-16 is less than α =5%, so H_0 is rejected. Thus, there was an individual effect on Indonesia's energy consumption equation model, resulting in the Fixed Effect Model (FEM) as the appropriate model. Because the selected estimation model was the FEM model, the next test was the Hausman test, while the LM test did not need to be performed.

4.2.2 Selection of Fixed Effect Model and Random Effect Model with Hausman Test

The Hausman test calculation was carried out using the RStudio program and obtained p-value = 0.01444 that is less than α=5%, therefore, H_0 was rejected. Thus, the correct estimation of the regression model for the Human Development Index data in Central Java in 2017-2019 was to use the Fixed Effect Model.

4.2.3 Assumption Test

1. Residual Normality Assumption Test

The normality assumption test was done by using the Shapiro Wilk test. Using RStudio, a statistical p-value of 0.7562 was obtained because the p-value is greater than α=5%, the residuals of the Fixed Effect Model followed a normal distribution.

2. Autocorrelation Test

The non-autocorrelation assumption test was done by Run Test. Based on the results of using RStudio, the statistical value of the p-value test was 0.202. The p-value is greater than α=5%; hence, there was no serial correlation in the error component.

3. Heteroscedasticity Test

The Breusch Pagan Test is used to determine whether the residual covariance-variance of the Fixed Effect Model is homoscedastic or heteroscedastic. Based on the results using RStudio, the statistical value of the p-value test was 0.3768. The p-value is greater than α=5%, so the residual covariance structure of the Fixed Effect Model were homoscedastic.

4. Multicollinearity Test

Through the correlation test with the RStudio program, the correlation value between the independent variables was not too low. The value was less than 0.8, which H_0 was not rejected. It can be concluded that the resulting model did not contain elements of multicollinearity.

4.2.4 Parameter Significance Test

1. Simultaneous Test (F-Test)

This test is conducted to test the estimation of the Fixed Effect Model whether the independent variables together influence the dependent variable. Based on the RStudio program results, the F count value was 9.245e+04 with a p-value of 2.2e-16. Because the p-value is less than 0.05, the independent variables together significantly affected the dependent variable Human Development Index.

2. Partial Test (t-Test)

The t-test aims to see the significance of the influence of individual independent variables on the dependent variable by assuming other variables are constant. Based on the results of RStudio, the value of $|t\text{-statistic}|$ was obtained for variables c, x_1, x_4, and x_6 was greater than the value of t(0.025;103) which was 0.980103 or p-value is less than 0.05. So, it can be concluded that the variables c, x_1, x_4, and x_6 had a significant influence on the dependent variable Human Development Index in Central Java.

Based on the results of the RStudio program, the R-Squared value was 99.55%. The dependent variables were influenced by the Junior High School Participation Rate, Public Health Complaints, Poverty Severity Index, and regional factors with the equation (5):

$$\hat{y}_{it} = \hat{c}_i + 0.276x_{1it} + 0.030x_{4it} - 1.997x_{6it} \quad (5)$$

4.3 Modeling Human Development Index (HDI) Data with RANFIS

In order to obtain an estimate of HDI data regression parameters, the RANFIS method was used based on classical panel data regression preprocessing. In general, the stages of ANFIS regression modeling include: determining input variables, forming clusters (membership functions), and forming fuzzy rules. Preprocessing was done by applying classical panel data regression to determine the optimal input. The optimal input variables selected in the ANFIS regression modeling were: the HDI variable as the response, with the predictor variables being: Junior High School Participation Rate (X1), Public Health Complaints (X4), and Poverty Severity Index (X6). Based on the sample data, the following results were obtained.

In the preprocessing of panel data regression modeling on HDI data and its predictor variables, the predictor variables that had a significant effect on HDI were Junior High School Participation Rate, Public Health Complaints, and Poverty Severity Index. These predictor variables were then used as input in the ANFIS process. After determining the input variables, the first step was to determine the membership function, the number of clusters, and the fuzzy rules that would be applied. This study determined clusters and rules using two methods, Fuzzy C-Means (FCM) and grid partition. Using a hybrid algorithm learning technique on in-sample data, the RMSE and MAPE values were obtained. To generate FIS using the FCM technique, the membership function (MF) used was the Gaussian function. In this technique, the number of rules was equal to the number of clusters determined. There were no combinations in the formation of the rule. Meanwhile, to generate FIS using the grid partition technique, each rule formed was a combination of the partition level for each input [20].

Optimal RANFIS modeling using FCM technique with two input variables x_1 and x_2 with two membership functions (clusters), two Sugeno rules of first-order can be formed as follows:

If x_1 *is* A_1 *and* x_2 *is* B_1 then $y_1 = p_{11}x_1 + q_{12}x_2 + r_1$

If x_1 *is* A_2 *and* x_2 *is* B_2 then $y_1 = p_{21}x_1 + q_{22}x_2 + r_2$

where A_1, B_1, A_2, B_2 as nonlinear parameters or premises, and $p_{11}, q_{12}, r_1, p_{21}, q_{22}, r_2$ as linear or consequent parameters.

If the firing strength for the two values of y_1 and y_2 is $w_1, and\ w_2$ then the output y could be determined as:

$$y = \frac{w_1y_1 + w_2y_2}{w_1 + w_2}.$$

In layer 1 in the RANFIS architecture, there are six groups of initial premise parameter values, with these values being used for the learning process. After obtaining the initial value of the premise parameters, the output generated in the first layer is the membership function of each input, $\mu_{A1}(x_1)$, $\mu_{A2}(x_1)$, $\mu_{B1}(x_2)$, and $\mu_{B2}(x_2)$. The membership function is used as input in layer 2, which produces the degree of activation of each rule. The optimal RANFIS has two rules, so layer 2 outputs are w_1 and w_2. Layer 2 output is used as input for layer 3, which will be normalized at the activation degree, then layer 3 output will be $\overline{\mathrm{w}}_1$ and $\overline{\mathrm{w}}_2$. The output of this layer is used as input in layer 4, which will produce linear parameters or consequent $p_{11}, q_{12}, r_1, p_{21}, q_{22}, r_2$ from the Recursive Least Squares Estimator (LSE) [20].

Based on Central Java HDI data as a case study, the RANFIS model obtained could be written as follows:

$$\begin{aligned} y = {} & 1.059\overline{\mathrm{w}}_{1,t}x_1 - 0.136\overline{\mathrm{w}}_{1,t}x_4 \\ & +0.166\overline{\mathrm{w}}_{1,t}x_6 - 19.570\overline{\mathrm{w}}_{1,t} \\ & +0.506\overline{\mathrm{w}}_{2,t}x_1 + 0.168\overline{\mathrm{w}}_{2,t}x_4 \\ & -4.9445\overline{\mathrm{w}}_{2,t}x_6 + 16.034\overline{\mathrm{w}}_{2,t} \end{aligned}$$

where

$$\overline{\mathrm{w}}_{1,t} = \frac{w_{1,t}}{w_{1,t}+w_{2,t}},$$

$$\overline{\mathrm{w}}_{2,t} = \frac{w_{2,t}}{w_{1,t}+w_{2,t}},$$

$$w_{1,t} = exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-96.358}{2.073}\right)^2\right] + \left(\frac{x_4-56.833}{5.929}\right)^2 + \left(\frac{x_6-0.345}{0.226}\right)^2\right\},$$

$$w_{2,t} = exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-95.599}{2.241}\right)^2 + \left(\frac{x_4-45.907}{5.718}\right)^2 + \left(\frac{x_6-0.438}{0.209}\right)^2\right]\right\}.$$

From the learning process using the hybrid algorithm, the RMSE, AIC, and BIC values were 3.227, respectively; 246.976; and 249.630; while the MAPE value was 3.299%.

5 CONCLUSION

Based on the panel data regression modeling procedure applied to the Human Development Index (HDI) data in Central Java in 2017-2019, an estimation of the panel data regression model of the Fixed Effect model was obtained. The Human Development Index variable could be explained from Junior High School Participation Rate, Public Health Complaints, and Poverty Severity Index. Using input variables selected through panel data regression, the optimal RANFIS model was obtained. The performance of the RANFIS model was evaluated using the RMSE and MAPE criteria. The RMSE and MAPE values were 3.227 and 3.299, respectively.

ACKNOWLEDGMENT

This paper is officially funded by grant of Faculty of Science and Mathematics, Universitas Diponegoro in 2021 with contract number 2164/UN7.5.8.2/PP/2021.

REFERENCES

[1] BALTAGI, B.H. (2005) *Econometric Analysis of Panel Data*, third edition, John Wiley & Sons, Ltd.

[2] APRILIAWAN, D. TARNO, T., DAN YASIN, H. (2013) Pemodelan Laju Inflasi di Provinsi Jawa Tengah Menggunakan Regresi Data Panel, *Jurnal Gaussian*, vol. 2, no. 4, 311 – 321.

[3] JANG, J.-S, R. (1993) ANFIS: Adaptive-Network-Based Fuzzy Inference System, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 23, No.3, pp.665-68

[4] JANG, J.-S, R. (1996) Input Selection for ANFIS Learning, *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems*, Vol. 2, pp. 1493-1499.

[5] JANG, J.-S, R., SUN, C. –T., MIZUTANI, E. (1997) *Neuro-Fuzzy and Soft Computing: A computational Approach to Learning and Machine Intelligent*. Prentice Hall International.

[6] FAUSSET, L., (1994). Fundamental of Neural Networks: Architectures, Algorithms and Aplications, Prentice HallEnglewood Cliffs, New Jersey.

[7] HAYKIN, S., (1999). Neural Networks: A Comprehensive Foundation, Second Edition, Prentice Hall International, Inc.

[8] BPS JAWA TENGAH, (2019a). *Indeks Pembangunan Manusia (IPM) Menurut Jenis Kelamin dan Kabupaten/Kota di Jawa Tengah, 2014-2019.* [Online] Available from https://jateng.bps.go.id/dynamictable/2019/01/23/58/indeks-pembangunan-manusia-ipm-menurut-jenis-kelamin-dan-kabupaten-kota-di-jawa-tengah-2014-2017.html [Accessed 14/10/20].

[9] BPS JAWA TENGAH, (2019b). *Angka Partisipasi Sekolah (APS) (Persen), 2017-2019.* [Online] Available from https://jateng.bps.go.id/indicator/28/71/1/angka-partisipasi-sekolah-aps-.html [Accessed 20/10/20].

[10] BPS JAWA TENGAH, (2019c). *Jumlah Tenaga Kesehatan Menurut Kabupaten/Kota di Provinsi Jawa Tengah, 2019.* [Online] Available from https://jateng.bps.go.id/ statictable/2020/07/20/1874/jumlah-tenaga-kesehatan-menurut-kabupaten-kota-di-provinsi-jawatengah-2019.html [Accessed 29/10/20].

[11] BPS JAWA TENGAH, (2019d). *Penduduk, Laju Pertumbuhan Penduduk, Distribusi Persentase Penduduk Kepadatan Penduduk, Rasio Jenis Kelamin Penduduk Menurut Kabupaten/Kota di Provinsi Jawa Tengah, 2010 dan 2019.* [Online] Available from https://jateng.bps.go.id/statictable/2020/06/11/1792/penduduk-laju-pertumbuhanpenduduk-distribusi-persentase-penduduk-kepadatan-penduduk-rasio-jenis-kelaminpenduduk-menurut-kabupaten-kota-di-provinsi-jawa-tengah-2010-dan-2019.html [Accessed 29/10/20].

[12] BPS JAWA TENGAH, (2019e). *Indeks Keparahan Kemiskinan (P2) (Persen), 2017-2019.* [Online] Available from https://jateng.bps.go.id/indicator/23/78/1/indeks-keparahan-kemiskinanp2-.html [Accessed 29/10/20].

[13] GREENE, Q. W. H., (2002). Econometric Analysis, Fifth Edition, New

York University, Upper Saddle River, New Jersey 07458

[14] Tarno, T., Suparti, S., Ispriyanti, D. (2018). Modeling Cayenne Production Data in Central Java Using Adaptive Neuro Fuzzy Inference System (ANFIS) Model Assisted Statistics and Applications,13(1) p.45-52.

[15] FANI GKOUNTAKOU and BASIL PAPADOPOULOS. (2020) The Use of Fuzzy Linear Regression and ANFIS Methods to Predict the Compressive Strength of Cement, *Symmetry*, 12, 1295; doi:10.3390/sym12081295

[16] MOKARRAM, M., AMIN, H., AND KHOSRAVI, M.R. (2019) Using adaptive neuro‐fuzzy inference system and multiple linear regression to estimate orange taste, *Food Science & Nutrition, 7(10): 3176‐3184.*

[17] KARABOGA, D., KAYA, E. (2019). Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey. Artif Intell Rev 52, 2263–2293. https://doi.org/10.1007/s10462-017-9610-2

[18] SHAH, M.I.; ABUNAMA, T.; JAVED, M.F.; BUX, F.; ALDREES, A.; TARIQ, M.A.U.R.; MOSAVI, A. (2021). Modeling Surface Water Quality Using the Adaptive Neuro-Fuzzy Inference System Aided by Input Optimization. Sustainability, 13, 4576. https:// doi.org/10.3390/su13084576

[19] HE, Z., WEN, X., LIU, H., & DU, J. (2014). A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. Journal of Hydrology. https://doi.org/10.1016/j.jhydrol.2013.11.054

[20] SARI. S.K., TARNO, SAFITRI, D. (2017). Pemilihan Input Model Regression Adaptive Neuro Fuzzy Inference System (RANFIS) Untuk Kajian Data IHSG, *Jurnal Gaussian*, Vol.6, No.3, 449-458

参考文:

COVER LETTER

Title of the manuscript: Modeling Regression Adaptive Neuro-Fuzzy Inference System (RANFIS) For Panel Data
Abstract. Panel data combines cross-sectional data and time-series data. Data on economic, business, social, and development issues are often presented in panel data. In constructing the panel data regression model, it is necessary to take various steps for testing the model specifications, including the Chow test and the Hausman test. The Chow test selects one of the two models, the Common Effect Model or Fixed Effect Model. Hausman test is used to compare Fixed Effect Model with Random Effect Model. This study aimed to construct a classical panel data regression model and the Regression Adaptive Neuro-Fuzzy Inference System (RANFIS). The RANFIS model is a regression model by applying fuzzy
Keywords: Panel Data Regression, Human Development Index, RANFIS.

Type of manuscript (please specify):

- [x] Research article
- [ ] Review article
- [ ] Brief report
- [ ] Short communication
- [ ] Research note

Full name and address of the Corresponding author: Tarno Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro Jl. Prof. Jacub Rais, Tembalang, Semarang 50275, Cental Java, Indonesia	
Telephone +6281325709047	Fax +622476486090
Email: tarno@lecturer.undip.ac.id	

- [x] I hereby confirm that the manuscript was prepared in accordance with the instructions for authors of scientific publications, and that the content of this manuscript, or most of it, was not published in the journal indicated, and the manuscript was not submitted for publication elsewhere.
- [x] I hereby confirm my consent to pay the Article Processing Charges (APC) EUR 450 in the case of the manuscript acceptance for publication. I am aware and hereby confirm that the APC is non-refundable.
- [x] I hereby confirm my consent to make the payment for English Language Editing Services (EUR 150).

Publication fee payer details

____

____

14th April 2023

Signature of the Corresponding author **Date**

Copyright Agreement

Manuscript title:
Modeling Regression Adaptive Neuro-Fuzzy Inference System (RANFIS) For Panel Data

Full names of all authors: Tarno, Di Asih I Maruddani, and Yuciana Wilandari

Full name and address of the corresponding author:

Tarno
Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro
Jl. Prof. Jacub Rais, Tembalang, Semarang 50275, Cental Jawa, Indonesia

Telephone/Whatsap: +6281325709047 Fax: +622476480690
Email: tarno@lecturer.undip.ac.id

License Agreement

(1) Authors own all the copyright rights for the paper.
(2) Submitted manuscript is an original paper.
(3) Authors hereby grant the Issues of Journal of SWJTU with an exclusive, royalty-free, worldwide license to email the paper to all who will ask for it.
(4) All authors have made a significant contribution to the research and are ready to assume joint responsibility for the paper.
(5) All authors have seen and approved the manuscript in the final form as it is submitted for publication.
(6) This manuscript has not been published and also has neither been submitted nor considered for publication elsewhere
(7) The text, illustrations and any other materials, included into the manuscript, do not infringe any existing intellectual property rights or other rights of any person or entity.
(8) The editors of the Issues of Journal of SWJTU, its personnel or the Editorial Board members accept no responsibility for the quality of the idea expressed in this publication.

I am the Corresponding author and have full authority to enter into this agreement.

Full name, affiliation and position:

Tarno, Department of Statistics, Universitas Diponegoro, Lecturer

Signature:

Date: 14th April 2023

Submission Received

jsjuorg@cpanel5.d.fozzy.com <jsjuorg@cpanel5.d.fozzy.com>
atas nama
Journal of Southwest Jiaotong University <editor@jsju.org>
Jum 14/04/2023 09.43

Kepada:Tarno <tarno@lecturer.undip.ac.id>

Thank you very much for uploading your manuscript to the Journal of Southwest Jiaotong University submission system. One of our editors will be in touch with you about the first decision within 10-14 days.

The following points were confirmed during submission:

1. Journal of Southwest Jiaotong University is an open access journal with publishing fees of EUR 500 for an accepted paper. This manuscript, if accepted, will be published under an open access Creative Commons CC BY license (https://creativecommons.org/licenses/by/4.0/), and I agree to pay the Article Processing Charges.

2. I hereby confirm my consent to make all payments for English Language Editing Services (EUR 150-200).

3. I understand that:

a. If previously published material is reproduced in my manuscript, I will provide proof that I have obtained the necessary copyright permission.

b. My manuscript is submitted on the understanding that it has not been published in or submitted to another peer-reviewed journal. Exceptions to this rule are papers containing material disclosed at conferences. I confirm that I will inform the journal editorial office if this is the case for my manuscript. I confirm that all authors are familiar with and agree with submission of the contents of the manuscript. The journal editorial office reserves the right to contact all authors to confirm this in case of doubt. I will provide email addresses for all authors and an institutional e-mail address for at least one of the co-authors, and specify the name, address and e-mail for invoicing purposes.

If you have any questions, please do not hesitate to contact the Editorial Office at editor@jsju.org

Kind regards,
Journal of Southwest Jiaotong University Editorial Office
http://jsju.org/index.php/journal/index
è¯¥æ,å¿—çš„å‡ºç‰ˆå•† (Our Publisher): Science Press
ç½‘å€ http://science-press.cn/è¥¿å—äº¤é€šå¤§å¦å¦æŠ¥/

*** This is an automatically generated email ***

Re: Paper Submission

editor@jsju.org <editor@jsju.org>
Kam 11/05/2023 10.31
Kepada:Tarno <tarno@lecturer.undip.ac.id>

1 lampiran (4 MB)
Invoice 46_05-JS.pdf;

Dear Authors,

Greetings from Journal of Southwest Jiaotong University,
Thank you so much for being so interested in submitting your research to our Journal of Southwest Jiaotong University.

The independent experts in the field have reviewed your manuscript. We are pleased to inform you that the paper has been accepted for publication subject to minor revisions being made in response to the reviewers' comments given below.

You need to send the following to the editorial office via editor@jsju.org, if you want to publish in the Journal of Southwest Jiaotong University, Volume 58 (3), June 2023:
The article, strictly formatted according to the template recommendations for authors and correct the article on the reviewers' recommendations, responses to reviewers, cover letter and a proof of payment.

All revisions should be highlighted in the article in red. We hope you will do great.

Reviewers:
1 - Please add scientific novelty (100 words) in the abstract.
2 - Please present a schematic representation of the steps involved in conducting the study.
3 - Conclusion section needs to be described scientifically. Kindly frame it along the following lines:
i. Main findings of the present study (100 words)
ii. Comparison with other studies (100 words)
iii. Implications of study (50 words)
iv. Strengths and limitations (50 words).

Deadline for corrections and payment: May 15, 2023.

* Please write in the subject of the email: revised article, cover letter and payment proof.

We send you the invoice.

Note 1: Please note that APCs will not be refunded due to the retraction or correction of an article owing to author error or misconduct.

This is because has provided and author's and publishing services and is only able to recoup this investment through the APCs.

Note 2: Charges can be made by any of the following methods
Credit / debit card - charges can be made online using a secure payment gateway as soon as the manuscript has been editorially accepted. We send all the details through approval email.
Bank Transfer – charges may have remitted to the bank account.
Western Union – details will be given in approval email.

If it is convenient for you to pay by card/Western Union, please write us about it.

The Journal of Southwest Jiaotong University is covered by the following databases and archives:

收录

EI Compendex

全国中文核心期刊

中国科技论文统计源期刊

中国科学引文数据库来源期刊.

The Journal of Southwest Jiaotong University is an international, interdisciplinary, peer-reviewed, open access journal, published bimonthly online by Science Press.

Open Access— free for readers, with article processing charges (APC) paid by authors or their institutions.

High visibility: indexed in many databases.

Subject area and category: multidisciplinary.

Rapid publication: manuscripts are peer-reviewed. The first decision is given to authors about 10–30 days after submission; acceptance for publication after revisions is done within 7–10 days (averages for articles published in this journal in 2022).

Recognition of reviewers: Reviewers who provide thorough review reports promptly receive vouchers that entitle them to a discount on the APC of their next publication.

Please read our published articles
http://jsju.org/index.php/journal/issue/archive.

If you have any questions, please do not hesitate to contact us via
editor@jsju.org

Sincerely yours,
Editorial Office of Journal of Southwest Jiaotong University
http://jsju.org/index.php/journal/index

该杂志的出版商 (Our Publisher): Science Press
网址 http://science-press.cn

On 2023-04-14 18:43, JSJU wrote:
> Title of your paper: Modeling Regression Adaptive Neuro-Fuzzy Inference
> System (RANFIS) For Panel Data
> Corresponding Author's Email Address: tarno@lecturer.undip.ac.id
> Author(s): Tarno, Di Asih I Maruddani, and Yuciana Wilandari
> Keywords: Panel Data Regression, Human Development Index, RANFIS
> Abstract: Panel data combines cross-sectional data and time-series
> data. Data on economic, business, social, and development issues are
> often presented in panel data. In constructing the panel data
> regression model, it is necessary to take various steps for testing the
> model specifications, including the Chow test and the Hausman test. The
> Chow test selects one of the two models, the Common Effect Model or
> Fixed Effect Model. Hausman test is used to compare Fixed Effect Model
> with Random Effect Model.
> This study aimed to construct a classical panel data regression model
> and the Regression Adaptive Neuro-Fuzzy Inference System (RANFIS). The
> RANFIS model is a regression model by applying fuzzy and Neural Network
> (NN) techniques expected to overcome the problem of uncertainty.
> The empirical study in this research is to construct a Panel data
> regression model for the Human Development Index (HDI) in Central Java
> in 2017-2019. The variables involved were Junior High School
> Participation Rate, Senior High School Participation Rate, Number of
> Health Workers, Public Health Complaints, Population Growth Rate,
> Poverty Severity Index as predictor variables, and Human Development
> Index as response variables. Applying the classic Panel data regression
> model, three factors that significantly affect HDI were obtained: the
> Junior High School Participation Rate, Public Health Complaints, and
> the Poverty Severity Index. These three variables were used as optimal
> inputs for the RANFIS modeling. Evaluation of model performance was
> measured based on the RMSE and MAPE values. Based on the ANFIS
> regression, the RMSE and MAPE values were 3.227 and 3.299,
> respectively.

Bls: Paper Submission

Tarno <tarno@lecturer.undip.ac.id>
Kam 25/05/2023 07.40
Kepada:editor@jsju.org <editor@jsju.org>

3 lampiran (1 MB)
RANFIS for Panel Data_revised.docx; payment proof2.jpg; payment proof1.jpg;

Dear editor jsju

I would like to send: revised article, cover letter and payment proof. The tittle of article: **MODELING REGRESSION ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM,(RANFIS) FOR PANEL DATA** (AUTHORS: TARNO, DI ASIH I MARUDDANI, YUCIANA WILANDARI)

Best regards,

Tarno
Department of Statistics
Universitas Diponegoro
Semarang, Indonesia 50275

Dari: editor@jsju.org <editor@jsju.org>
Dikirim: Kamis, 11 Mei 2023 10.31
Kepada: Tarno <tarno@lecturer.undip.ac.id>
Subjek: Re: Paper Submission

Dear Authors,

Greetings from Journal of Southwest Jiaotong University,
Thank you so much for being so interested in submitting your research to our Journal of Southwest Jiaotong University.

The independent experts in the field have reviewed your manuscript. We are pleased to inform you that the paper has been accepted for publication subject to minor revisions being made in response to the reviewers' comments given below.

You need to send the following to the editorial office via editor@jsju.org, if you want to publish in the Journal of Southwest Jiaotong University, Volume 58 (3), June 2023:
The article, strictly formatted according to the template recommendations for authors and correct the article on the reviewers' recommendations, responses to reviewers, cover letter and a proof of payment.

All revisions should be highlighted in the article in red. We hope you will do great.

Reviewers:

1 - Please add scientific novelty (100 words) in the abstract.
2 - Please present a schematic representation of the steps involved in conducting the study.
3 - Conclusion section needs to be described scientifically. Kindly frame it along the following lines:
i. Main findings of the present study (100 words)
ii. Comparison with other studies (100 words)
iii. Implications of study (50 words)
iv. Strengths and limitations (50 words).

Deadline for corrections and payment: May 15, 2023.

* Please write in the subject of the email: revised article, cover letter and payment proof.

We send you the invoice.

Note 1: Please note that APCs will not be refunded due to the retraction or correction of an article owing to author error or misconduct. This is because has provided and author's and publishing services and is only able to recoup this investment through the APCs.

Note 2: Charges can be made by any of the following methods
Credit / debit card - charges can be made online using a secure payment gateway as soon as the manuscript has been editorially accepted. We send all the details through approval email.
Bank Transfer – charges may have remitted to the bank account.
Western Union – details will be given in approval email.

If it is convenient for you to pay by card/Western Union, please write us about it.

The Journal of Southwest Jiaotong University is covered by the following databases and archives:

收录

EI Compendex

全国中文核心期刊

中国科技论文统计源期刊

中国科学引文数据库来源期刊.

The Journal of Southwest Jiaotong University is an international, interdisciplinary, peer-reviewed, open access journal, published bimonthly online by Science Press.

Open Access— free for readers, with article processing charges (APC) paid by authors or their institutions.

High visibility: indexed in many databases.

Subject area and category: multidisciplinary.

Rapid publication: manuscripts are peer-reviewed. The first decision is given to authors about 10–30 days after submission; acceptance for publication after revisions is done within 7–10 days (averages for articles published in this journal in 2022).

Recognition of reviewers: Reviewers who provide thorough review reports promptly receive vouchers that entitle them to a discount on the APC of their next publication.

Please read our published articles
http://jsju.org/index.php/journal/issue/archive.

If you have any questions, please do not hesitate to contact us via
editor@jsju.org

Sincerely yours,
Editorial Office of Journal of Southwest Jiaotong University
http://jsju.org/index.php/journal/index

该杂志的出版商 (Our Publisher): Science Press
网址 http://science-press.cn

On 2023-04-14 18:43, JSJU wrote:
> Title of your paper: Modeling Regression Adaptive Neuro-Fuzzy Inference
> System (RANFIS) For Panel Data
> Corresponding Author's Email Address: tarno@lecturer.undip.ac.id
> Author(s): Tarno, Di Asih I Maruddani, and Yuciana Wilandari
> Keywords: Panel Data Regression, Human Development Index, RANFIS
> Abstract: Panel data combines cross-sectional data and time-series
> data. Data on economic, business, social, and development issues are
> often presented in panel data. In constructing the panel data
> regression model, it is necessary to take various steps for testing the
> model specifications, including the Chow test and the Hausman test. The
> Chow test selects one of the two models, the Common Effect Model or
> Fixed Effect Model. Hausman test is used to compare Fixed Effect Model
> with Random Effect Model.
> This study aimed to construct a classical panel data regression model
> and the Regression Adaptive Neuro-Fuzzy Inference System (RANFIS). The
> RANFIS model is a regression model by applying fuzzy and Neural Network
> (NN) techniques expected to overcome the problem of uncertainty.
> The empirical study in this research is to construct a Panel data
> regression model for the Human Development Index (HDI) in Central Java
> in 2017-2019. The variables involved were Junior High School

> Participation Rate, Senior High School Participation Rate, Number of
> Health Workers, Public Health Complaints, Population Growth Rate,
> Poverty Severity Index as predictor variables, and Human Development
> Index as response variables. Applying the classic Panel data regression
> model, three factors that significantly affect HDI were obtained: the
> Junior High School Participation Rate, Public Health Complaints, and
> the Poverty Severity Index. These three variables were used as optimal
> inputs for the RANFIS modeling. Evaluation of model performance was
> measured based on the RMSE and MAPE values. Based on the ANFIS
> regression, the RMSE and MAPE values were 3.227 and 3.299,
> respectively.

Lampiran 5. Revised article, Cover letter, and payment proof

西 南 交 通 大 学 学 报

第 58 卷 第 3 期
2023 年 06 月

JOURNAL OF SOUTHWEST JIAOTONG UNIVERSITY

Vol. 58 No. 3
Jun. 2023

ISSN: 0258-2724 **DOI：10.35741/issn.0258-2724.57.6.1**

Category
Research Article

MODELING REGRESSION ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM (RANFIS) FOR PANEL DATA

Tarno [a,]*, Di Asih I Maruddani [b], Yuciana Wilandari [c]

[a] Statistics Department, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang 50275, Indonesia, e-mail: tarno@lecturer.undip.ac.id

[b] Statistics Department, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang 50275, Indonesia, e-mail: maruddani@live.undip.ac.id

[c] Statistics Department, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang 50275, Indonesia, e-mail: yuciana.wilandari@gmail.com

Received: ▪ *Review:* ▪ *Accepted:* ▪ *Published*

Abstract

Panel data combines cross-sectional data and time-series data. Data on economic, business, social, and development issues are often presented in panel data. In constructing the panel data regression model, it is necessary to take various steps for testing the model specifications, including the Chow test and the Hausman test. This study aimed to construct a classical panel data regression model and the Regression Adaptive Neuro-Fuzzy Inference System (RANFIS). The RANFIS model is a regression model by applying fuzzy and Neural Network (NN) techniques expected to overcome the problem of uncertainty. One of the main problems in constructing an optimal RANFIS is selecting input variables. The input variables of RANFIS are selected based on the best classical regression. Those inputs are classified into optimal clusters which depend on the degree of fuzzy membership functions. The rule bases of RANFIS are determined based on optimal inputs and its clusters. The empirical study in this research is to construct a panel data regression model for the Human Development Index (HDI) in Central Java in 2017-2019. HDI is depend on several variables such as: the School Participation Rate, Number of Health Workers, Public Health Complaints, Population Growth Rate, Poverty Severity Index as predictor variables. Based on classical regression, three variables were used as optimal inputs for RANFIS modeling. Evaluation of model performance was measured based on the RMSE and MAPE values. Based on the RANFIS, the values of RMSE and MAPE were 3.227 and 3.299, respectively.

Keywords: Panel Data Regression, Human Development Index, RANFIS

摘要 The authors may not translate the abstract and keywords into Chinese themselves.

关键词:

I. INTRODUCTION

Panel data combines cross-sectional data and time-series data [1]. Data on economic, business, social, and development issues are often presented in panel data. In constructing a suitable regression model for panel data, it is necessary to take various steps for model specification tests, including the Chow test, Hausman test, and the Lagrange Multiplier test. The Chow test selects one of two models, the Common Effect Model or the Fixed Effect Model. Hausman test is used to compare the models of Fixed Effect with Random Effect [2].

This study aimed to construct a suitable regression model for panel data. The regression model built was the classical panel data regression model and the Adaptive Neuro-Fuzzy Inference System (RANFIS) regression. The RANFIS model is a regression model applying fuzzy and Neural Network (NN) techniques expected to overcome the problem of uncertainty and nonlinearity in the data. The merging of these two methods aimed to obtain an accurate model. The fuzzy system is a universal approximator capable of classifying data with high uncertainty. At the same time, NN has good learning abilities on data.

The fuzzy system is a "universal approximator," defined as techniques related to uncertainty based on fuzzy sets. The advantage of the system is that the developed model is characterized by linguistic interpretation abilities and rules that can be understood, verified, and developed [3], [4], [5]). Neural networks (NN) model is one example of a nonlinear model with a flexible functional form. It contains several parameters that cannot be interpreted as the parametric model. As a supervised machine learning method, NN provides a good framework for representing a relationship in data. Compared to other algorithms, NN has better adaptive ability, learning, and pattern non-stationary and nonlinear signals [6], [7].

The empirical study aimed to construct a panel data regression model, specifically to identify the factors that affect the Human Development Index (HDI). Human development intends to have more choices, especially in income, health, and education. HDI is a standard measure of human development set by the United Nations. HDI is formed through three essential variables: health, education, and decent living standards. According to the Central Bureau of Statistics Republic of Indonesia (2019a) [8], HDI is one way to measure the success of human development based on several fundamental components of life quality. To measure the health variable using the number of health workers and the percentage of people complaining about their health and seeking treatment. The education variable is measured by two indicators: the junior high school participation rate and the senior high school participation rate. The variable of decent living standard is measured by population growth and the severity of poverty.

To conduct a further study in this research, the variables identified for the empirical study were Junior High School Participation Rate, High School Participation Rate (Central Bureau of Statistic, 2019b) [9], Number of Health Workers, Public Health Complaints (Central Bureau of Statistics, 2019c [10]), Growth Rate Population (Central Bureau of Statistics, 2019d [11]), and Poverty Severity Index as independent variables (predictors) (Central Bureau of Statistics, 2019e) [12], and Human Development Index (HDI) as response variables (dependent variable) (Central Bureau of Statistics, 2019a [8]). The data taken for the case study were from 35 regencies and cities in Central Java Province from 2017 to 2019. The modeling for panel data was carried out using the classical regression model and RANFIS. The estimation results using the two methods were compared with the level of accuracy based on the predicted MAPE value.

This study objective was to develop and apply a regression model for panel data: (1) Compile a classic panel data regression model for HDI data in Central Java, (2) Establish the ANFIS Regression model for HDI data in Central Java.

II. THEORETICAL FRAMEWORK

2.1. Panel Data Regression

Panel data is a combination of time series data and cross-section data. Regression using Panel data is called panel data regression model [1]. Baltagi (2005) developed panel data regression analysis with the following theoretical concepts.

- Panel Data Regression Model

Panel data combines cross-section data and time-series data, so the model can be written as follows.

$$Y_{it} = \alpha + \beta X_{it} + u_{it};$$
$$i = 1,2,\dots,N;\ t = 1,2,\dots,T.$$

where

i =1, 2,..., N are households, individuals, companies, or others showing the dimensions of cross-sectional data;

t =1,2 ,..., T represents the dimension of the time series data;

α : the scalar intercept coefficient

β: slope coefficient with dimensions K×1 where K is the number of independent variables

Y_{it}: dependent variable of individual i-th at time t

X_{it}: independent variable of individual i-th at time t

The residual component in the panel data regression model consists of a general residual component and a specific residual component. The general residual component is the residual component of the individual i-th and the general residual component of the time t. The specific residual component consists of the specific residual of individual i-th and time t. The specific residual component can be written as:

$$u_{it} = \mu_i + \lambda_t + \varepsilon_{it}$$

with

u_{it}: residual component for individual i-th at time t

μ_i: the specific influence of the individual i-th

λ_t: specific effect of time t

ε_{it}: residual for the individual i-th at time t

- Panel Data Regression Types

In estimating the panel regression model, there are three commonly used approaches: Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM) [1].

a. Common Effect Model (CEM)

The combined model is the simplest in panel data regression. The combined model ignores the individual-specific effect (μ_i) and the time-specific effect (λ_t) in the model. The model used follows the form of linear regression with the residual component u_{it} which only comes from the estimated residual component (ε_{it}). The parameter estimation method in this model is the same as the ordinary linear regression model, which uses the least-squares method (Gujarati 2004). The CEM model can be written as follows:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \varepsilon_{it};\ i = 1,2,\ldots N;\ t = 1,2\ldots,T \quad (3)$$

b. Fixed Effect Model (FEM)

The fixed effect model is based on the assumption that the intercept between individual and time is different. However, the regression coefficient is constant for all individuals and time. In addition, this model assumes that there is a correlation between individual-specific effects (μ_i) and time-specific effects (λ_t) with independent variables. This assumption makes individual-specific effects (μ_i), and time-specific effects (λ_t) part of the intercept [1]. The FEM equation can be written as follows:

1. FEM with one-way residual component:

$$Y_{it} = \alpha + \mu_i + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \varepsilon_{it} \text{ with } \textstyle\sum_{i=1}^{N} \mu_i = 0,$$

or

$$Y_{it} = \alpha + \lambda_t + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \varepsilon_{it} \text{ with } \textstyle\sum_{t=1}^{T} \lambda_t = 0.$$

2. FEM model with two-way residual components:

$$Y_{it} = \alpha + \mu_i + \lambda_t + \beta_1 X_{1it} + \beta_2 X_{2it} + \cdots + \beta_k X_{kit} + \varepsilon_{it}$$

with $\sum_{i=1}^{N} \mu_i$=0 and $\sum_{t=1}^{T} \lambda_t$=0.

Intercept differences between the individual and time are caused by their different characteristics, so estimating parameters with these conditions uses the Least-Squares Dummy Variable (LSDV) method. The estimation results using the LSDV method produce an unbiased estimator. However, adding a large number of dummy variables will result in a significant loss of the degree of freedom resulting in the estimator inefficiency and multicollinearity due to too many predictable variables [1].

c. Random Effect Model (REM)

The random effect model assumes that there is no correlation between individual-specific effects (μ_i) and time-specific effects (λ_t) with independent variables. This assumption makes the residual component of the individual-specific effect (μ_i) and the time-specific effect (λ_t) included in the residual. The equation for the random effect model can be written as follows:

1. REM with one-way residual component:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \mu_i + \varepsilon_i$$

with $\mu_i \sim N(0, \sigma_i^2);\ cov(\mu_i, X_{it}) = 0$

or

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_k X_{kit} + \lambda_t + u_t$$

with $\lambda_t \sim N(0, \sigma_t^2);\ cov(\lambda_t, X_{it}) = 0$

2. REM with two-way residual components:

$$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \ldots + \beta_n X_{nit} + \mu_i + \lambda_t + w_{it}$$

with $\mu_i \sim N(0, \sigma_i^2);\ cov(\mu_i, X_{it}) = 0$ and $\lambda_t \sim N(0, \sigma_t^2);\ cov(\lambda_t, X_{it}) = 0$

Panel Data Regression Estimation

In determining the estimation of the panel regression model, several tests were carried out to select the optimum estimation approach method. The first step in getting the desired model was the Chow test on the FEM estimation results; after proving that there was an individual effect, the Hausman test was carried out to determine between FEM and REM [1].

1. Chow Test

Chow test selects the two models between the Common Effect Model and the Fixed Effect Model. The assumption that each cross-sectional unit has the same behavior tends to be unrealistic,

considering that each cross-sectional unit can have different behavior is the basis of the Chow test. In this test, the following hypotheses are carried out:
H_0: $\alpha_1 = \alpha_2 = ... = \alpha_N = \alpha$ (Common Effect Model)
H_1: there is at least one different intercept α_1 (Fixed Effect Model)

The basis for rejecting H_0 is to use F-statistics as follows (Baltagi, 2008):

$$Chow = \frac{(RSS1 - RSS2)/(N-1)}{RSS2/(NT-N-K)}$$

RSS1: residual sum of square of common effect model estimation results
RSS2: residual sum of square of fixed effect model estimation results
N: number of cross-section unit
T: number of time series unit
K: number of independent variables

Chow Test statistics follow the distribution of F-statistics, namely $F_{(N-1,NT-N-K);\alpha}$. If the Chow statistic is greater than the F-table, there is sufficient evidence to reject H_0 and vice versa.

2. Hausman Test

Hausman test is used to compare Fixed Effect Model with Random Effect Model. The Hausman test is conducted when the Fixed Effect Model contains an element of trade-off, namely the loss of the degree of freedom element by including dummy variables and the Random Effect Model, which must heed the absence of assumptions violation of each component of the error. In this test, the following hypotheses are carried out:
H_0: corr(X_{it}, u_{it}) = 0 (Random Effect Model)
H_1: corr(X_{it}, u_{it}) ≠ 0 (Fixed Effect Model)
The basis for rejecting H_0 using Hausman Statistics is formulated as follows [13]:

$$\chi^2(K) = (b-\beta)'[Var(b-\beta)] - 1(b-\beta)$$

with:
b: random effect coefficient
β: fixed effect coefficient
Hausman statistics spread Chi-Square, if the value of χ^2 is greater than $\chi^2_{(K,\ \alpha)}$ (K: number of independent variables) or P-Value < α, then there is sufficient evidence to reject H_0 and vice versa.

3. Lagrange Multiplier (LM) Test

This test is carried out to detect the presence of heteroscedasticity in the estimated model. The LM test hypotheses are as follows:
H_0: $\sigma_i^2 = \sigma^2$ (there is no heteroscedasticity)
H_1: $\sigma_i^2 \neq \sigma^2$ (ther is heteroscedasticity)

LM test statistics are as follows [13]:

$$LM = \frac{NT}{2(T-1)} + \sum_{i=1}^{N}\left(\frac{T^2\sigma_i^2}{\sigma^2} - 1\right)^2$$

where:
T: number of time series unit
N: number of cross-section unit
σ_i^2: residual variance of the equation i
σ^2: residual variance of system equation
Conclusion H_0 is rejected if LM is greater than $\chi^2_{(1,\alpha)}$ which means heteroscedasticity occurs in the model. Thus, it must be estimated using the weight method: Cross-section weight.

4. Breusch Pagan Test

The Breusch Pagan test is an LM test to choose between a fixed effect model and a pooled regression model. The initial hypothesis is that the variance of the residuals in the fixed coefficient model is zero. The procedure is as follows [1]
Hypotheses
H_0: $\sigma_\mu^2 = 0$
H_1: $\sigma_\mu^2 \neq 0$
The test statistic used is the LM

$$LM = \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}(\sum_{t=1}^{T}\hat{u}_{it})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2} - 1\right]$$

where
N: number of individuals
T: length of the time period
σ_μ^2: model residual variance
$\hat{u}_{it}$: residual estimation of the individual fixed coefficient model i $p$eriod t
If $LM > \chi^2_{(1,\alpha)}$ or p-value is less than the specified significance level, then H_0 is rejected. Thus, the random effect model is selected.

2.2. Adaptive Neuro-Fuzzy Inference System (ANFIS)

The materials and the sources used in this study cover all articles discussing ANFIS, which combines Neural Networks (NN) and Fuzzy Inference System (FIS). Before we discuss the procedure of ANFIS modeling, the most important material that should be described in this section is the structure of ANFIS networks. The NN architecture applied in ANFIS consists of five fixed layers [5], [14]. Without loss of generality, the architecture of ANFIS for modeling time-series data is given two input variables x_1, x_2 and single output variable y by assuming rule-base of Sugeno first order with two rules is as follows:

If x_1 *is* A_1 *and* x_2 *is* B_1 then $y_1 = p_{11}x_1 + q_{12}x_2 + r_1$

If x_1 *is* A_2 *and* x_2 *is* B_2 then $y_2 = p_{21}x_1 + q_{22}x_2 + r_2$

where
x_i *is* A_j *and* x_2 *is* B_1; and x_1 *is* A_2 *and* x_2 *is* B_2 as premise sections, whereas $y_1 = p_{11}x_1 + q_{12}x_2 + r_1$ and $y_2 = p_{21}x_1 + q_{22}x_2 + r_2$ as consequent sections; p_{11}, q_{12}, r_1, p_{21}, q_{22}, r_2 as linear parameters; A_1, B_1, A_2, B_2 as the nonlinear

parameter. If the firing strength for two values y_1, y_2 are w_1, w_2 respectively then the output y can be expressed as in equation (1).

$$y = \overline{w}_1 y_1 + \overline{w}_2 y_2 \quad (1)$$

where $\overline{w}_i = \frac{w_i}{\sum w_i}, i = 1,2.$

The structure of ANFIS networks (Figure 1) has five layers and can be explained as follows [5].

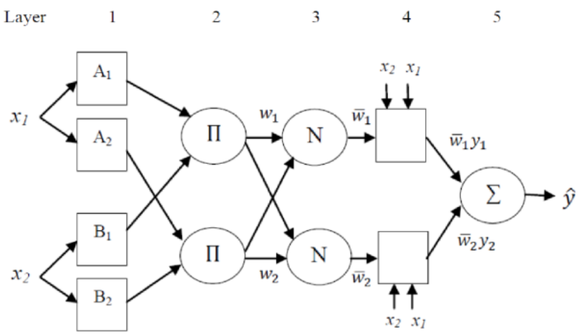

Figure 1. Structure of ANFIS Networks for Time Series Modeling [5]

Layer 1: Each neuron in this layer is adaptive to the parameters of an activation function. The output of each neuron is the membership degree of input. For example, the membership function of Generalized Bell is as follows:

$$\mu(x_i) = \frac{1}{1+\left|\frac{x_i - c_i}{a_i}\right|^{2b_i}}$$

where x_i is input and a_i, b_i and c_i are premise parameters [3], [4], [5].

Layer 2: Each neuron in this layer is a permanent neuron that is given the symbol Π, which is the product of all inputs in layer 1: $w_i = \mu_{Ai}(x_1) \times \mu_{Bi}(x_2),\ i = 1, 2$.

Each neuron output is called the firing strength of a rule [15], [16], [17], [18], [19].

Layer 3: Each neuron in this layer is a fixed neuron with the symbol N, which is the result of calculating the ratio of the i -firing strength to the total number of firing strengths in the second layer as follows: $\overline{w}_i = \frac{w_i}{\sum w_i}, i = 1, 2.$

The results of calculations at this layer are called normalized firing strength.

Layer 4: This layer is a neuron which is an adaptive neuron to an output:

$$\overline{w}_i y_i = \overline{w}_i (p_i x_1 + q_i x_2 + r_i)$$

where $\overline{w}_i$ is normalized firing strength in the third layer while p_i, q_i, and r_i are parameters in these neurons called consequent parameters.

Layer 5: This layer is a single neuron with the symbol Σ which is the sum of all outputs from the fourth layer, as follows:

$$y = \overline{w}_1 y_1 + \overline{w}_2 y_2, \text{ where}$$

$$\overline{w}_i = \frac{w_i}{\sum w_i}, i = 1,2.$$

3. METHOD OF MODELING

This research was based on a literature study. The initial step was to study in-depth and thoroughly from books and scientific articles that served as the basis for the new abstract system formation. We also examined supporting scientific articles that could be used in solving problems. At this stage, accuracy was needed in discussing supporting scientific articles, which were expected to solve the core problems. In addition to theoretical studies, applied studies were also carried out. In detail, this research method is described as follows.

3.1. Data Source

The data used in this study were the Human Development Index (%) and several factors that affect it in education, health, and population in 35 regencies and cities in Central Java Province from 2017 to 2019. All data were obtained from the Central Bureau of Statistics Central Java Province publications.

3.2. Research Variables

The variables used in this study were as follows:

a. Human Development Index (%) as response variable Y
b. Junior High School Participation Rate (%) as independent variable X1
c. High School Participation Rate (%) as independent variable X2
d. Number of Health Workers as independent variable X3
e. Public Health Complaints (%) as independent variable X4
f. Population Growth Rate as independent variable X5
g. Poverty Severity Index (%) as independent variable X6

3.3. Analysis Method

The data analysis method used in this research was modeling using panel data regression analysis, bootstrapping regression, and RANFIS. The following steps were taken to analyze the data.

3.3.1. Panel Data Regression Modeling

1. The general description of the data in data plots and descriptive statistics was seen.

2. The best panel data regression model to model the effect of the Junior High School Participation Rate, Senior High School Participation Rate, Number of Health Workers, Public Health Complaints, Population Growth Rate, and Poverty on the Human Development Index in Central Java was determined.
3. The Common Effect model, Fixed Effect model, and Random Effect model were estimated.
4. The best model was determined through the Chow test, Hausman test, and the Lagrange Multiplier (LM) test. If the Chow Test and Hausman Test showed the results of the Fixed Effect model, there was no need to proceed to the Lagrange Multiplier Test.
5. The classical assumptions of regression on the selected model were tested.
6. The significance of the panel data regression parameters, including Simultaneous Test (F-Test), Partial Test (t-Test), and the measure of the goodness of the model with R-Square, was tested.

   The procedure of constructing data panel regression can be illustrated as figure 2.

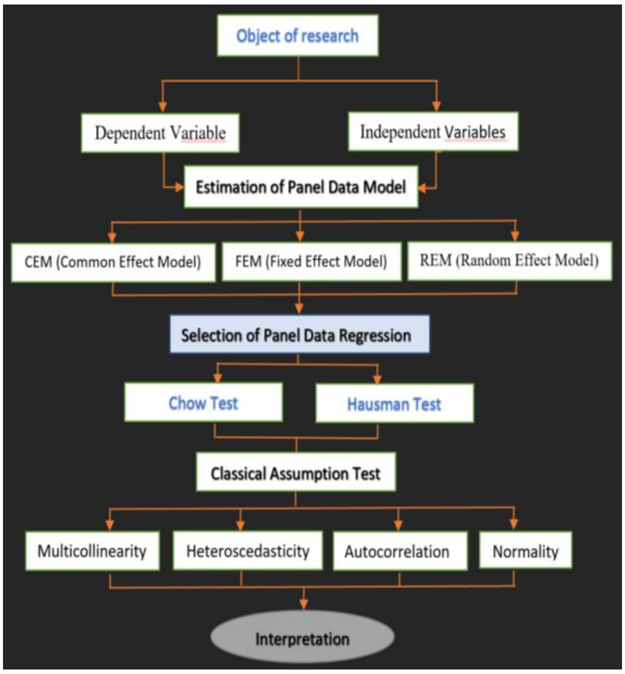

Figure 2. The procedure of constructing data panel regression

3.3.2. ANFIS Regression Modeling

The estimation steps of the RANFIS model for panel data were as follows.

1. Preprocessing was performed by estimating the classical panel data regression model.
2. A new response data was formed based on the preprocessing results in step 1.
3. ANFIS modeling took new responses as targets with input variables as in panel data regression modeling.
4. Several clusters and membership functions for input variables were defined.
5. IF-THEN fuzzy rules were generated for output variables based on input, cluster, rule, and type of membership function. The IF-THEN fuzzy rules were formed using the First Order Sugeno model.
6. Fuzzy Inference System (FIS) training was conducted on an in-sample with a hybrid algorithm. The consequent parameters were estimated using a recursive LSE. The premise parameters were adjusted according to the backpropagation concept of gradient descent.
7. The predicted value in the in-sample was determined; the RMSE and MAPE were calculated.

   The procedure of constructing data panel RANFIS can be illustrated as figure 3.

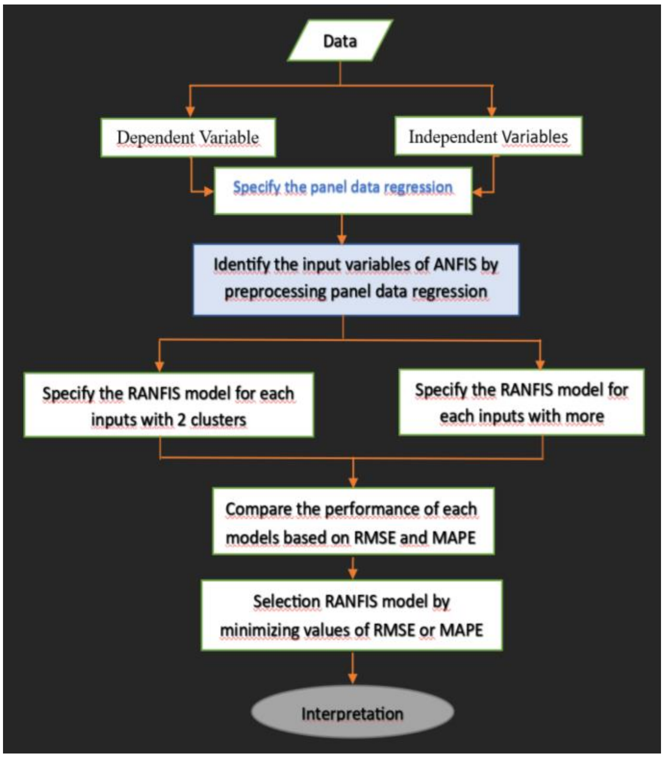

Figure 3. The procedure of constructing data panel RANFIS

4. Results and Discussion

4.1. Regression of Panel Data

4.1.1. Common Effect Model

According to the data processing of Central Java HDI 2017-2019, the estimation of the combined model (Common Effect Model) was obtained as equation (2).

$$\hat{y}_{it} = 43.975 - 0.025x_{1it} + 0.270x_{2it} + 0.0008x_{3it} + 0.065x_{4it} + 3.122x_{5it} + 3.127x_{6it} \quad (2)$$

4.1.2. *Fixed Effect Model*

Fixed Effect Modeling of the Human Development Index was carried out with the RStudio program. The estimation result was obtained as equation (3):

$$\hat{y}_{it} = \hat{c}_i + 0.276x_{1it} + 0.031x_{2it} - 0.0002x_{3it} + 0.030x_{4it} - 2.785x_{5it} - 1.997x_{6it} \quad (3)$$

with the value of $\hat{c}_i$ owned by each region in Central Java presented in Table 1.

Tabel 1.

Intercept estimation $\hat{c}_i$ for Fixed Effect Model

i	Region	$\hat{c}_i$	i	Region	$\hat{c}_i$
1	Cilacap Regency	42.175	19	Kudus Regency	49.600
2	Banyumas Regency	46.306	20	Jepara Regency	46.815
3	Purbalingga Regency	44.943	21	Demak Regency	46.366
4	Banjarnegara Regency	41.898	22	Semarang Regency	48.394
5	Kebumen Regency	46.051	23	Temanggung Regency	43.084
6	Purworejo Regency	43.668	24	Kendal Regency	46.047
7	Wonosobo Regency	42.890	25	Batang Regency	42.462
8	Magelang Regency	44.018	26	Pekalongan Regency	44.475
9	Boyolali Regency	47.302	27	Pemalang Regency	40.184
10	Klaten Regency	46.261	28	Tegal Regency	40.945
11	Sukoharjo Regency	48.442	29	Brebes Regency	39.780
12	Wonogiri Regency	41.135	30	Magelang City	50.293
13	Karanganyar Regency	48.629	31	Surakarta City	53.419
14	Sragen Regency	45.788	32	Salatiga City	56.690
15	Grobogan Regency	43.474	33	Semarang City	59.217
16	Blora Regency	39.563	34	Pekalongan City	48.811
17	Rembang Regency	43.001	35	Tegal City	47.430
18	Pati Regency	44.184			

4.1.3. *Random Effect Model*

Random Effect Modeling of the Human Development Index was carried out with the help of the RStudio program. The estimation result was obtained as equation (4):

$$\hat{y}_{it} = \hat{c}_i + 33.515 + 0.337x_{1it} + 0.055x_{2it} - 0.00003x_{3it} + 0.029x_{4it} + 2.948x_{5it} - 2.172x_{6it} \quad (4)$$

with the value of $\hat{ci}$ owned by each region in Central Java presented in Table 2.

Tabel 2.

Intercept estimation $\hat{c}_i$ for Random Effect Model

i	Region	$\hat{c}_i$	i	Region	$\hat{c}_i$
1	Cilacap Regency	-2.375	19	Kudus Regency	49.600
2	Banyumas Regency	-0.652	20	Jepara Regency	46.815
3	Purbalingga Regency	1.986	21	Demak Regency	46.366
4	Banjarnegara Regency	-2.513	22	Semarang Regency	48.394
5	Kebumen Regency	-1.714	23	Temanggung Regency	43.084
6	Purworejo Regency	0.089	24	Kendal Regency	46.047
7	Wonosobo Regency	-0.615	25	Batang Regency	42.462
8	Magelang Regency	-2.817	26	Pekalongan Regency	44.475
9	Boyolali Regency	2.521	27	Pemalang Regency	40.184
10	Klaten Regency	2.058	28	Tegal Regency	40.945
11	Sukoharjo Regency	1.452	29	Brebes Regency	39.780
12	Wonogiri Regency	-2.467	30	Magelang City	50.293
13	Karanganyar Regency	1.612	31	Surakarta City	53.419
14	Sragen Regency	1.962	32	Salatiga City	56.690
15	Grobogan Regency	-0.795	33	Semarang City	6.609
16	Blora Regency	-4.377	34	Pekalongan City	2.183
17	Rembang Regency	-2.919	35	Tegal City	3.547
18	Pati Regency	-0.048			

4.2. Panel Data Model Selection

4.2.1. *Selection of Common Effect Model and Fixed Effect Model with Chow Test*

Calculation of the Chow test was carried out using RStudio program and obtained the value of F statistics is equal to 112.91 that is greater than F(0.05;5;98) = 2.30722 and p-value = 2.2e-16 is less than α =5%, so H_0 is rejected. Thus, there was an individual effect on Indonesia's energy consumption equation model, resulting in the Fixed Effect Model (FEM) as the appropriate model. Because the selected estimation model was the FEM model, the next test was the Hausman test, while the LM test did not need to be performed.

4.2.2 Selection of Fixed Effect Model and Random Effect Model with Hausman Test

The Hausman test calculation was carried out using the RStudio program and obtained p-value = 0.01444 that is less than α=5%, therefore, H_0 was rejected. Thus, the correct estimation of the regression model for the Human Development Index data in Central Java in 2017-2019 was to use the Fixed Effect Model.

4.2.3 Assumption Test

1. Residual Normality Assumption Test

The normality assumption test was done by using the Shapiro Wilk test. Using RStudio, a statistical p-value of 0.7562 was obtained because the p-value is greater than α=5%, the residuals of the Fixed Effect Model followed a normal distribution.

2. Autocorrelation Test

The non-autocorrelation assumption test was done by Run Test. Based on the results of using RStudio, the statistical value of the p-value test was 0.202. The p-value is greater than α=5%; hence, there was no serial correlation in the error component.

3. Heteroscedasticity Test

The Breusch Pagan Test is used to determine whether the residual covariance-variance of the Fixed Effect Model is homoscedastic or heteroscedastic. Based on the results using RStudio, the statistical value of the p-value test was 0.3768. The p-value is greater than α=5%, so the residual covariance structure of the Fixed Effect Model were homoscedastic.

4. Multicollinearity Test

Through the correlation test with the RStudio program, the correlation value between the independent variables was not too low. The value was less than 0.8, which H_0 was not rejected. It can be concluded that the resulting model did not contain elements of multicollinearity.

4.2.4 Parameter Significance Test

1. Simultaneous Test (F-Test)

This test is conducted to test the estimation of the Fixed Effect Model whether the independent variables together influence the dependent variable. Based on the RStudio program results, the F count value was 9.245e+04 with a p-value of 2.2e-16. Because the p-value is less than 0.05, the independent variables together significantly affected the dependent variable Human Development Index.

2. Partial Test (t-Test)

The t-test aims to see the significance of the influence of individual independent variables on the dependent variable by assuming other variables are constant. Based on the results of RStudio, the value of $|t\text{-statistic}|$ was obtained for variables c, x_1, x_4, and x_6 was greater than the value of t(0.025;103) which was 0.980103 or p-value is less than 0.05. So, it can be concluded that the variables c, x_1, x_4, and x_6 had a significant influence on the dependent variable Human Development Index in Central Java.

Based on the results of the RStudio program, the R-Squared value was 99.55%. The dependent variables were influenced by the Junior High School Participation Rate, Public Health Complaints, Poverty Severity Index, and regional factors with the equation (5):

$$\hat{y}_{it} = \hat{c}_i + 0.276x_{1it} + 0.030x_{4it} - 1.997x_{6it} \quad (5)$$

4.3 Modeling Human Development Index (HDI) Data with RANFIS

In order to obtain an estimate of HDI data regression parameters, the RANFIS method was used based on classical panel data regression preprocessing. In general, the stages of ANFIS regression modeling include: determining input variables, forming clusters (membership functions), and forming fuzzy rules. Preprocessing was done by applying classical panel data regression to determine the optimal input. The optimal input variables selected in the ANFIS regression modeling were: the HDI variable as the response, with the predictor variables being: Junior High School Participation Rate (X1), Public Health Complaints (X4), and Poverty Severity Index (X6). Based on the sample data, the following results were obtained.

In the preprocessing of panel data regression modeling on HDI data and its predictor variables, the predictor variables that had a significant effect on HDI were Junior High School Participation Rate, Public Health Complaints, and Poverty Severity Index. These predictor variables were then used as input in the ANFIS process. After determining the input variables, the first step was to determine the membership function, the number of clusters, and the fuzzy rules that would be applied. This study determined clusters and rules using two methods, Fuzzy C-Means (FCM) and grid partition. Using a hybrid algorithm learning technique on in-sample data, the RMSE and MAPE values were obtained. To generate FIS using the FCM technique, the membership function (MF) used was the Gaussian function. In this technique, the number of rules was equal to the number of clusters determined. There were no combinations in the formation of the rule. Meanwhile, to generate FIS using the grid partition technique, each rule formed was a combination of the partition level for each input [20].

Optimal RANFIS modeling using FCM technique with two input variables x_1 and x_2 with two membership functions (clusters), two Sugeno rules of first-order can be formed as follows:

If x_1 *is* A_1 *and* x_2 *is* B_1 then $y_1 = p_{11}x_1 + q_{12}x_2 + r_1$

If x_1 *is* A_2 *and* x_2 *is* B_2 then $y_1 = p_{21}x_1 + q_{22}x_2 + r_2$

where A_1, B_1, A_2, B_2 as nonlinear parameters or premises, and $p_{11}, q_{12}, r_1, p_{21}, q_{22}, r_2$ as linear or consequent parameters.

If the firing strength for the two values of y_1 and y_2 is $w_1, and\ w_2$ then the output y could be determined as:

$$y = \frac{w_1 y_1 + w_2 y_2}{w_1 + w_2}.$$

In layer 1 in the RANFIS architecture, there are six groups of initial premise parameter values, with these values being used for the learning process. After obtaining the initial value of the premise parameters, the output generated in the first layer is the membership function of each input, $\mu_{A1}(x_1)$, $\mu_{A2}(x_1)$, $\mu_{B1}(x_2)$, and $\mu_{B2}(x_2)$. The membership function is used as input in layer 2, which produces the degree of activation of each rule. The optimal RANFIS has two rules, so layer 2 outputs are w_1 and w_2. Layer 2 output is used as input for layer 3, which will be normalized at the activation degree, then layer 3 output will be $\overline{w}_1$ and $\overline{w}_2$. The output of this layer is used as input in layer 4, which will produce linear parameters or consequent $p_{11}, q_{12}, r_1, p_{21}, q_{22}, r_2$ from the Recursive Least Squares Estimator (LSE) [20].

Based on Central Java HDI data as a case study, the RANFIS model obtained could be written as follows:

$$\begin{aligned} y = {} & 1.059\overline{w}_{1,t}x_1 - 0.136\overline{w}_{1,t}x_4 \\ & +0.166\overline{w}_{1,t}x_6 - 19.570\overline{w}_{1,t} \\ & +0.506\overline{w}_{2,t}x_1 + 0.168\overline{w}_{2,t}x_4 \\ & -4.9445\overline{w}_{2,t}x_6 + 16.034\overline{w}_{2,t} \end{aligned}$$

where

$$\overline{w}_{1,t} = \frac{w_{1,t}}{w_{1,t}+w_{2,t}},$$

$$\overline{w}_{2,t} = \frac{w_{2,t}}{w_{1,t}+w_{2,t}},$$

$$w_{1,t} = exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-96.358}{2.073}\right)^2\right] + \left(\frac{x_4-56.833}{5.929}\right)^2 + \left(\frac{x_6-0.345}{0.226}\right)^2\right\},$$

$$w_{2,t} = exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-95.599}{2.241}\right)^2\right] + \left(\frac{x_4-45.907}{5.718}\right)^2 + \left(\frac{x_6-0.438}{0.209}\right)^2\right\}.$$

From the learning process using the hybrid algorithm, the RMSE, AIC, and BIC values were 3.227, respectively; 246.976; and 249.630; while the MAPE value was 3.299%.

5 Conclusion

Based on the panel data regression modeling procedure applied to the Human Development Index (HDI) data in Central Java in 2017-2019, an estimation of the panel data regression model of the Fixed Effect model was obtained. The HDI variable could be explained from Junior High School Participation Rate, Public Health Complaints, and Poverty Severity Index. Using the input variables selected through panel data regression, the optimal RANFIS model was obtained. The RANFIS optimal has three inputs with 2 clusters (membership functions). The performance of the RANFIS model was evaluated using the RMSE and MAPE criteria. The RMSE and MAPE values were 3.227 and 3.299, respectively. The RANFIS model performs well to apply for nonlinear data containing uncertainty.

Acknowledgment

This paper is officially funded by grant of Faculty of Science and Mathematics, Universitas Diponegoro in 2021 with contract number 2164/UN7.5.8.2/PP/2021.

References

[1] BALTAGI, B.H. (2005) *Econometric Analysis of Panel Data*, third edition, John Wiley & Sons, Ltd.

[2] APRILIAWAN, D. TARNO, T., DAN YASIN, H. (2013) Pemodelan Laju Inflasi di Provinsi Jawa Tengah Menggunakan Regresi Data Panel, *Jurnal Gaussian*, vol. 2, no. 4, 311 – 321.

[3] JANG, J.-S, R. (1993) ANFIS: Adaptive-Network-Based Fuzzy Inference System, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 23, No.3, pp.665-68

[4] JANG, J.-S, R. (1996) Input Selection for ANFIS Learning, *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems*, Vol. 2, pp. 1493-1499.

[5] JANG, J.-S, R., SUN, C. –T., MIZUTANI, E. (1997) *Neuro-Fuzzy and Soft Computing: A computational Approach to Learning and Machine Intelligent*. Prentice Hall International.

[6] FAUSSET, L., (1994). Fundamental of Neural Networks: Architectures, Algorithms and Aplications, Prentice HallEnglewood Cliffs, New Jersey.

[7] HAYKIN, S., (1999). Neural Networks: A Comprehensive Foundation, Second Edition, Prentice Hall International, Inc.
[8] BPS JAWA TENGAH, (2019a). *Indeks Pembangunan Manusia (IPM) Menurut Jenis Kelamin dan Kabupaten/Kota di Jawa Tengah, 2014-2019.* [Online] Available from https://jateng.bps.go.id/dynamictable/2019/01/23/58/indeks-pembangunan-manusia-ipm-menurut-jenis-kelamin-dan-kabupaten-kota-di-jawa-tengah-2014-2017.html [Accessed 14/10/20].
[9] BPS JAWA TENGAH, (2019b). *Angka Partisipasi Sekolah (APS) (Persen), 2017-2019.* [Online] Available from https://jateng.bps.go.id/indicator/28/71/1/angka-partisipasi-sekolah-aps-.html [Accessed 20/10/20].
[10] BPS JAWA TENGAH, (2019c). *Jumlah Tenaga Kesehatan Menurut Kabupaten/Kota di Provinsi Jawa Tengah, 2019.* [Online] Available from https://jateng.bps.go.id/ statictable/2020/07/20/1874/jumlah-tenaga-kesehatan-menurut-kabupaten-kota-di-provinsi-jawatengah-2019.html [Accessed 29/10/20].
[11] BPS JAWA TENGAH, (2019d). *Penduduk, Laju Pertumbuhan Penduduk, Distribusi Persentase Penduduk Kepadatan Penduduk, Rasio Jenis Kelamin Penduduk Menurut Kabupaten/Kota di Provinsi Jawa Tengah, 2010 dan 2019.* [Online] Available from https://jateng.bps.go.id/statictable/2020/06/11/1792/penduduk-laju-pertumbuhanpenduduk-distribusi-persentase-penduduk-kepadatan-penduduk-rasio-jenis-kelaminpenduduk-menurut-kabupaten-kota-di-provinsi-jawa-tengah-2010-dan-2019.html [Accessed 29/10/20].
[12] BPS JAWA TENGAH, (2019e). *Indeks Keparahan Kemiskinan (P2) (Persen), 2017-2019.* [Online] Available from https://jateng.bps.go.id/indicator/23/78/1/indeks-keparahan-kemiskinanp2-.html [Accessed 29/10/20].
[13] GREENE, Q. W. H., (2002). Econometric Analysis, Fifth Edition, New York University, Upper Saddle River, New Jersey 07458
[14] Tarno, T., Suparti, S., Ispriyanti, D. (2018). Modeling Cayenne Production Data in Central Java Using Adaptive Neuro Fuzzy Inference System (ANFIS) Model Assisted Statistics and Applications,13(1) p.45-52.
[15] FANI GKOUNTAKOU and BASIL PAPADOPOULOS. (2020) The Use of Fuzzy Linear Regression and ANFIS Methods to Predict the Compressive Strength of Cement, *Symmetry*, 12, 1295; doi:10.3390/sym12081295
[16] MOKARRAM, M., AMIN, H., AND KHOSRAVI, M.R. (2019) Using adaptive neuro‐fuzzy inference system and multiple linear regression to estimate orange taste, *Food Science & Nutrition, 7(10): 3176‐3184.*
[17] KARABOGA, D., KAYA, E. (2019). Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey. Artif Intell Rev 52, 2263–2293. https://doi.org/10.1007/s10462-017-9610-2
[18] SHAH, M.I.; ABUNAMA, T.; JAVED, M.F.; BUX, F.; ALDREES, A.; TARIQ, M.A.U.R.; MOSAVI, A. (2021). Modeling Surface Water Quality Using the Adaptive Neuro-Fuzzy Inference System Aided by Input Optimization. Sustainability, 13, 4576. https:// doi.org/10.3390/su13084576
[19] HE, Z., WEN, X., LIU, H., & DU, J. (2014). A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. Journal of Hydrology. https://doi.org/10.1016/j.jhydrol.2013.11.054
[20] SARI. S.K., TARNO, SAFITRI, D. (2017). Pemilihan Input Model Regression Adaptive Neuro Fuzzy Inference System (RANFIS) Untuk Kajian Data IHSG, *Jurnal Gaussian*, Vol.6, No.3, 449-458

参考文:

COVER LETTER

Title of the manuscript: Modeling Regression Adaptive Neuro-Fuzzy Inference System (RANFIS) For Panel Data
Abstract. Panel data combines cross-sectional data and time-series data. Data on economic, business, social, and development issues are often presented in panel data. In constructing the panel data regression model, it is necessary to take various steps for testing the model specifications, including the Chow test and the Hausman test. This study aimed to construct a classical panel data regression model and the Regression Adaptive Neuro-Fuzzy Inference System (RANFIS). The RANFIS model is a regression model by applying fuzzy and Neural Network (NN) techniques expected to overcome the problem of uncertainty. One of the main problems in constructing an optimal RANFIS is selecting input variables. The input variables of RANFIS are selected based on the best classical regression. Those inputs are classified into optimal clusters which depend on the degree of fuzzy membership functions. The rule bases of RANFIS are determined based on optimal inputs and its clusters. The empirical study in this research is to construct a panel data regression model for the Human Development Index (HDI) in Central Java in 2017-2019. HDI is depend on several variables such as: the School Participation Rate, Number of Health Workers, Public Health Complaints, Population Growth Rate, Poverty Severity Index as predictor variables. Based on classical regression, three variables were used as optimal inputs for RANFIS modeling. Evaluation of model performance was measured based on the RMSE and MAPE values. Based on the RANFIS, the values of RMSE and MAPE were 3.227 and 3.299, respectively.
Keywords: Panel Data Regression, Human Development Index, RANFIS.

Type of manuscript (please specify):

- ☑ Research article
- ☐ Review article
- ☐ Brief report
- ☐ Short communication
- ☐ Research note

Full name and address of the Corresponding author: Tarno Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro Jl. Prof. Jacub Rais, Tembalang, Semarang 50275, Cental Java, Indonesia	
Telephone +6281325709047	Fax +622476486090
Email: tarno@lecturer.undip.ac.id	

☑ I hereby confirm that the manuscript was prepared in accordance with the instructions for authors of scientific publications, and that the content of this manuscript, or most of it, was not published in the journal indicated, and the manuscript was not submitted for publication elsewhere.

☑ I hereby confirm my consent to pay the Article Processing Charges (APC) EUR 450 in the case of the manuscript acceptance for publication. I am aware and hereby confirm that the APC is non-refundable.

☑ I hereby confirm my consent to make the payment for English Language Editing Services (EUR 150).

Publication fee payer details

25th May 2023

Signature of the Corresponding author **Date**

Copyright Agreement

Manuscript title:
Modeling Regression Adaptive Neuro-Fuzzy Inference System (RANFIS) For Panel Data

Full names of all authors: Tarno, Di Asih I Maruddani, and Yuciana Wilandari

Full name and address of the corresponding author:

Tarno
Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro
Jl. Prof. Jacub Rais, Tembalang, Semarang 50275, Cental Jawa, Indonesia

Telephone/Whatsap: +6281325709047 Fax: +622476480690
Email: tarno@lecturer.undip.ac.id

License Agreement

(1) Authors own all the copyright rights for the paper.
(2) Submitted manuscript is an original paper.
(3) Authors hereby grant the Issues of Journal of SWJTU with an exclusive, royalty-free, worldwide license to email the paper to all who will ask for it.
(4) All authors have made a significant contribution to the research and are ready to assume joint responsibility for the paper.
(5) All authors have seen and approved the manuscript in the final form as it is submitted for publication.
(6) This manuscript has not been published and also has neither been submitted nor considered for publication elsewhere
(7) The text, illustrations and any other materials, included into the manuscript, do not infringe any existing intellectual property rights or other rights of any person or entity.
(8) The editors of the Issues of Journal of SWJTU, its personnel or the Editorial Board members accept no responsibility for the quality of the idea expressed in this publication.

I am the Corresponding author and have full authority to enter into this agreement.

Full name, affiliation and position:

Tarno, Department of Statistics, Universitas Diponegoro, Lecturer

Signature: 

Date: 25th May 2023

Re: Bls: Bls: Paper Submission

editor@jsju.org <editor@jsju.org>
Jum 09/06/2023 06.02
Kepada:Tarno <tarno@lecturer.undip.ac.id>

Dear Tarno,

Many thanks for your support of open access publishing.

We acknowledge APC payment for paper received successfully.

Please wait for publication during June.

Sincerely yours,
Editorial Office of Journal of Southwest Jiaotong University

On 2023-06-09 10:22, Tarno wrote:
> Payer's full name: Nur 'Aini Hamada. She is my daughter. Thank You
>
> Best Regards
> Tarno
>
> Department of Statistics
> Universitas Diponegoro
> semarang, Indonesia
>
> ------------------------
>
> Dari: editor@jsju.org <editor@jsju.org>
> Dikirim: Kamis, 08 Juni 2023 03.48
> Kepada: Tarno <tarno@lecturer.undip.ac.id>
> Subjek: Re: Bls: Paper Submission
>
> Dear Authors,
>
> Please write the payer's full name.
>
> On 2023-05-25 16:40, Tarno wrote:
>> Dear editor jsju
>>
>> I would like to send: revised article, cover letter and payment
>> proof. The tittle of article: Modeling Regression Adaptive
> Neuro-Fuzzy
>> Inference System,(RANFIS) For Panel Data (authors: Tarno, Di Asih I
>> Maruddani, Yuciana Wilandari)
>>
>> Best regards,
>>
>> Tarno
>> Department of Statistics
>> Universitas Diponegoro

>> Semarang, Indonesia 50275
>>
>> -------------------------
>>
>> Dari: editor@jsju.org <editor@jsju.org>
>> Dikirim: Kamis, 11 Mei 2023 10.31
>> Kepada: Tarno <tarno@lecturer.undip.ac.id>
>> Subjek: Re: Paper Submission
>>
>> Dear Authors,
>>
>> Greetings from Journal of Southwest Jiaotong University,
>> Thank you so much for being so interested in submitting your
> research
>> to
>> our Journal of Southwest Jiaotong University.
>>
>> The independent experts in the field have reviewed your manuscript.
>> We are pleased to inform you that the paper has been accepted for
>> publication subject to minor revisions being made in response to the
>> reviewers' comments given below.
>>
>> You need to send the following to the editorial office via
>> editor@jsju.org, if you want to publish in the Journal of Southwest
>> Jiaotong University, Volume 58 (3), June 2023:
>> The article, strictly formatted according to the template
>> recommendations for authors and correct the article on the
> reviewers'
>> recommendations, responses to reviewers, cover letter and a proof of
>> payment.
>>
>> All revisions should be highlighted in the article in red. We hope
> you
>>
>> will do great.
>>
>> Reviewers:
>> 1 - Please add scientific novelty (100 words) in the abstract.
>> 2 - Please present a schematic representation of the steps involved
> in
>>
>> conducting the study.
>> 3 - Conclusion section needs to be described scientifically. Kindly
>> frame it along the following lines:
>> i. Main findings of the present study (100 words)
>> ii. Comparison with other studies (100 words)
>> iii. Implications of study (50 words)
>> iv. Strengths and limitations (50 words).
>>
>> Deadline for corrections and payment: May 15, 2023.
>>
>> * Please write in the subject of the email: revised article, cover

> > letter and payment proof.
> >
> > We send you the invoice.
> >
> > Note 1: Please note that APCs will not be refunded due to the
> > retraction
> > or correction of an article owing to author error or misconduct.
> > This is because has provided and author's and publishing services
> and
> >
> > is only able to recoup this investment through the APCs.
> >
> > Note 2: Charges can be made by any of the following methods
> > Credit / debit card - charges can be made online using a secure
> > payment
> > gateway as soon as the manuscript has been editorially accepted. We
> > send
> > all the details through approval email.
> > Bank Transfer – charges may have remitted to the bank account.
> > Western Union – details will be given in approval email.
> >
> > If it is convenient for you to pay by card/Western Union, please
> write
> >
> > us about it.
> >
> > The Journal of Southwest Jiaotong University is covered by the
> > following
> > databases and archives:
> >
> > 收录
> >
> > EI Compendex
> >
> > 全国中文核心期刊
> >
> > 中国科技论文统计源期刊
> >
> > 中国科学引文数据库来源期刊.
> >
> > The Journal of Southwest Jiaotong University is an international,
> > interdisciplinary, peer-reviewed, open access journal, published
> > bimonthly online by Science Press.
> >
> > Open Access— free for readers, with article processing charges
> (APC)
> >
> > paid by authors or their institutions.
> >
> > High visibility: indexed in many databases.
> >
> > Subject area and category: multidisciplinary.

>>
>> Rapid publication: manuscripts are peer-reviewed. The first decision
>> is
>> given to authors about 10–30 days after submission; acceptance for
>> publication after revisions is done within 7–10 days (averages for
>> articles published in this journal in 2022).
>>
>> Recognition of reviewers: Reviewers who provide thorough review
>> reports
>> promptly receive vouchers that entitle them to a discount on the APC
>> of
>> their next publication.
>>
>> Please read our published articles
>> http://jsju.org/index.php/journal/issue/archive.
>>
>> If you have any questions, please do not hesitate to contact us via
>> editor@jsju.org
>>
>> Sincerely yours,
>> Editorial Office of Journal of Southwest Jiaotong University
>> http://jsju.org/index.php/journal/index
>>
>> 该杂志的出版商 (Our Publisher): Science Press
>> 网址 http://science-press.cn
>>
>> On 2023-04-14 18:43, JSJU wrote:
>>> Title of your paper: Modeling Regression Adaptive Neuro-Fuzzy
>> Inference
>>> System (RANFIS) For Panel Data
>>> Corresponding Author's Email Address: tarno@lecturer.undip.ac.id
>>> Author(s): Tarno, Di Asih I Maruddani, and Yuciana Wilandari
>>> Keywords: Panel Data Regression, Human Development Index, RANFIS
>>> Abstract: Panel data combines cross-sectional data and time-series
>>> data. Data on economic, business, social, and development issues
> are
>>
>>> often presented in panel data. In constructing the panel data
>>> regression model, it is necessary to take various steps for testing
>> the
>>> model specifications, including the Chow test and the Hausman test.
>> The
>>> Chow test selects one of the two models, the Common Effect Model or
>>> Fixed Effect Model. Hausman test is used to compare Fixed Effect
>> Model
>>> with Random Effect Model.
>>> This study aimed to construct a classical panel data regression
>> model
>>> and the Regression Adaptive Neuro-Fuzzy Inference System (RANFIS).
>> The
>>> RANFIS model is a regression model by applying fuzzy and Neural
>> Network

>>> (NN) techniques expected to overcome the problem of uncertainty.
>>> The empirical study in this research is to construct a Panel data
>>> regression model for the Human Development Index (HDI) in Central
>> Java
>>> in 2017-2019. The variables involved were Junior High School
>>> Participation Rate, Senior High School Participation Rate, Number
> of
>>
>>> Health Workers, Public Health Complaints, Population Growth Rate,
>>> Poverty Severity Index as predictor variables, and Human
> Development
>>
>>> Index as response variables. Applying the classic Panel data
>> regression
>>> model, three factors that significantly affect HDI were obtained:
>> the
>>> Junior High School Participation Rate, Public Health Complaints,
> and
>>
>>> the Poverty Severity Index. These three variables were used as
>> optimal
>>> inputs for the RANFIS modeling. Evaluation of model performance was
>>> measured based on the RMSE and MAPE values. Based on the ANFIS
>>> regression, the RMSE and MAPE values were 3.227 and 3.299,
>>> respectively.

Lampiran 7. Artikel yang dipublikasikan secara online, 19 Juni 2023

西 南 交 通 大 学 学 报

第 58 卷 第 3 期
2 0 2 3 年 6 月

JOURNAL OF SOUTHWEST JIAOTONG UNIVERSITY

Vol. 58 No. 3
June 2 0 2 3

ISSN: 0258-2724 DOI：10.35741/issn.0258-2724.58.3.57

Research article

Mathematics

MODELING REGRESSION ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM (RANFIS) FOR PANEL DATA

面板数据的建模回归自适应神经模糊推理系统(兰菲斯)

Tarno*, Di Asih I. Maruddani, Yuciana Wilandari
Statistics Department, Faculty of Science and Mathematics, Universitas Diponegoro
Semarang, 50275, Indonesia, tarno@lecturer.undip.ac.id, maruddani@live.undip.ac.id, yuciana.wilandari@gmail.com

Received: March 9, 2023 ▪ Reviewed: April 13, 2023
▪ Accepted: May 24, 2023 ▪ Published: June 30, 2023

Abstract

Panel data combine cross-sectional and time-series data. Data on economic, business, social, and development issues are often presented in panel data. In constructing the panel data regression model, it is necessary to take various steps for testing the model specifications, including the Chow test and the Hausman test. This study constructed a classical panel data regression model and the regression adaptive neuro-fuzzy inference system (RANFIS). The RANFIS model is a regression model by applying fuzzy and neural network (NN) techniques expected to overcome the problem of uncertainty. One of the main problems in constructing an optimal RANFIS is selecting input variables. The input variables of RANFIS are selected on the basis of the best classical regression. These inputs are classified into optimal clusters, which depend on the degree of fuzzy membership functions. The rule bases of RANFIS are determined on the basis of optimal inputs and its clusters. The empirical study in this research is to construct a panel data regression model for the Human Development Index (HDI) in Central Java in 2017-2019. HDI depends on several variables such as the school participation rate, number of health workers, public health complaints, population growth rate, and poverty severity index as predictor variables. Based on classical regression, three variables were used as optimal inputs for RANFIS modeling. Evaluation of model performance was measured based on the RMSE and MAPE values. Based on the RANFIS, the values of RMSE and MAPE were 3.227 and 3.299, respectively.

Keywords: Panel Data Regression, Human Development Index, Regression Adaptive Neuro-Fuzzy Inference System

摘要 面板数据结合了横截面数据和时间序列数据。有关经济、商业、社会和发展问题的数据通常以面板数据的形式呈现。在构建面板数据回归模型时，需要采取各种步骤对模型规格进行检验，

包括松狮犬检验和豪斯曼检验。本研究构建了经典的面板数据回归模型和回归自适应神经模糊推理系统（兰菲斯）。兰菲斯模型是一种通过应用模糊和神经网络(神经网络)技术来克服不确定性问题的回归模型。构建最优兰菲斯的主要问题之一是选择输入变量。兰菲斯的输入变量是在最佳经典回归的基础上选择的。这些输入被分类到最佳集群中，这取决于模糊隶属函数的程度。兰菲斯的规则库是在最优输入及其簇的基础上确定的。本研究的实证研究是构建 2017-2019 年中爪哇人类发展指数（人类发展指数）的面板数据回归模型。人类发展指数取决于几个变量，例如学校参与率、卫生工作者数量、公共卫生投诉、人口增长率和贫困严重程度指数作为预测变量。基于经典回归，三个变量被用作兰菲斯建模的最佳输入。基于均方根误差和马佩值测量模型性能的评估。基于兰菲斯，均方根误差和马佩的值分别为 3.227 和 3.299。

关键词: 面板数据回归、人类发展指数、回归自适应神经模糊推理系统

I. INTRODUCTION

Panel data combine cross-sectional and time-series data [1]. Data on economic, business, social, and development issues are often presented in panel data. In constructing a suitable regression model for panel data, it is necessary to take various steps for model specification tests, including the Chow test, the Hausman test, and the Lagrange multiplier test. The Chow test selects one of two models, the common effect model or the fixed-effects model. The Hausman test is used to compare the models of fixed and random effects [2].

This study constructed a suitable regression model for panel data. The regression model built was the classical panel data regression model and the adaptive neuro-fuzzy inference system (RANFIS) regression. The RANFIS model is a regression model applying fuzzy and neural network (NN) techniques expected to overcome the problem of uncertainty and nonlinearity in the data. The merging of these two methods aimed to obtain an accurate model. The fuzzy system is a universal approximator capable of classifying data with high uncertainty. At the same time, NN has good learning abilities on data.

The fuzzy system is a “universal approximator” defined as techniques related to uncertainty based on fuzzy sets. The advantage of the system is that the developed model is characterized by linguistic interpretation abilities and rules that can be understood, verified, and developed [3], [4], [5]. The neural networks (NN) model is one example of a nonlinear model with a flexible functional form. It contains several parameters that cannot be interpreted as a parametric model. As a supervised machine learning method, NN provides a good framework for representing a relationship in data. Compared to other algorithms, NN has better adaptive ability, learning, and pattern non-stationary and nonlinear signals [6], [7].

The empirical study constructed a panel data regression model to identify the factors that affect the human development index (HDI). Human development intends to have more choices, especially in income, health, and education. HDI is a standard measure of human development set by the United Nations. HDI is formed through three essential variables: health, education, and decent living standards. According to [8], HDI is one way to measure the success of human development based on several fundamental components of quality of life. To measure the health variable using the number of health workers and the percentage of people complaining about their health and seeking treatment. The education variable is measured by two indicators: the junior high school participation rate and the senior high school participation rate. The variable of decent living standard is measured by population growth and the severity of poverty.

In this research, the variables identified for the empirical study were junior high school participation rate, high school participation rate [9], number of health workers, public health complaints [10], population growth rate [11], and poverty severity index as independent variables (predictors) [12] and human development index (HDI) as a response variable (dependent variable) [8]. The data taken for the case study were from 35 regencies and cities in Central Java Province from 2017 to 2019. Panel data modeling was carried out using the classical regression model and RANFIS. The estimation results using the two methods were compared with the level of accuracy based on the predicted MAPE value.

This study objective was to develop and apply a regression model for panel data: (1) compile a classic panel data regression model for HDI data in Central Java and (2) establish the ANFIS regression model for HDI data in Central Java.

II. THEORETICAL FRAMEWORK

A. Panel Data Regression

Panel data are a combination of time series data and cross-section data. Regression using panel data is called panel data regression model [1]. [1] developed panel data regression analysis with the following theoretical concepts.

Panel Data Regression Model

Panel data combine cross-section and time-series data, so the model can be written as follows.

$Y_{it} = \alpha + \beta X_{it} + u_{it};\ i = 1,2,\dots,N;\ t = 1,2,\dots,T.$

where

i = *1, 2,..., N* are households, individuals, companies, or others showing the dimensions of cross-sectional data;

t = *1,2,..., T* represents the dimension of the time series data;

α - the scalar intercept coefficient;

β - slope coefficient with dimensions K×1 where K is the number of independent variables;

Y_{it} - dependent variable of the i-th individual at time *t*;

X_{it} - independent variable of the i-th individual at time *t*.

The residual component in the panel data regression model consists of a general residual component and a specific residual component. The general residual component is the residual component of the *i*-th individual and the general residual component of time *t*. The specific residual component consists of the specific residual of the *i*-th individual and time t. The specific residual component can be written as

$u_{it} = \mu_i + \lambda_t + \varepsilon_{it}$

with:

u_{it} - residual component for the i-th individual at time *t*;

μ_i - the specific influence of the i-th individual;

λ_t - specific effect of time *t*;

ε_{it} - residual for the i-th individual at time *t*.

Panel Data Regression Types

In estimating the panel regression model, there are three commonly used approaches: common effect model (CEM), fixed-effects model (FEM), and random effect model (REM) [1].

a. Common Effect Model (CEM)

The combined model is the simplest in panel data regression. The combined model ignores the individual-specific effect (μ_i) and the time-specific effect (λ_t) in the model. The model used follows the form of linear regression with the residual component u_{it} which only comes from the estimated residual component (ε_{it}). The parameter estimation method in this model is the same as the ordinary linear regression model, which uses the least-squares method [21]. The CEM model can be written as follows:

$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \dots + \beta_k X_{kit} + \varepsilon_{it};\ i = 1,2,\dots N;\ t = 1,2\dots,T$

b. Fixed-Effects Model (FEM)

The fixed-effects model assumes that the intercept between an individual and time is different. However, the regression coefficient is constant for all individuals and time. In addition, this model assumes that there is a correlation between individual-specific effects (μ_i)and time-specific effects (λ_t) with independent variables. This assumption makes individual-specific effects (μ_i) and time-specific effects (λ_t) part of the intercept [1]. The FEM equation can be written as follows:

1. FEM with one-way residual component:

$Y_{it} = \alpha + \mu_i + \beta_1 X_{1it} + \beta_2 X_{2it} + \dots + \beta_k X_{kit} + \varepsilon_{it}$ with $\sum_{i=1}^{N} \mu_i = 0$ or $Y_{it} = \alpha + \lambda_t + \beta_1 X_{1it} + \beta_2 X_{2it} + \dots + \beta_k X_{kit} + \varepsilon_{it}$ with $\sum_{t=1}^{T} \lambda_t = 0$.

2. FEM model with two-way residual components:

$Y_{it} = \alpha + \mu_i + \lambda_t + \beta_1 X_{1it} + \beta_2 X_{2it} + \cdots + \beta_k X_{kit} + \varepsilon_{it}$ with $\sum_{i=1}^{N} \mu_i$= 0 and $\sum_{t=1}^{T} \lambda_t$= 0.

Intercept differences between the individual and time are caused by their different characteristics, so estimating parameters with these conditions uses the least squares dummy variable (LSDV) method. The estimation results using the LSDV method produce an unbiased estimator. However, adding a large number of dummy variables will result in a significant loss of the degree of freedom, resulting in the estimator inefficiency and multicollinearity due to too many predictable variables [1].

c. Random Effect Model (REM)

The random effect model assumes that there is no correlation between individual-specific effects (μ_i) and time-specific effects (λ_t) with independent variables. This assumption makes the residual component of the individual-specific effect (μ_i) and the time-specific effect (λ_t) included in the residual. The equation for the random effect model can be written as follows:

1. REM with one-way residual component:

$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \dots + \beta_k X_{kit} + \mu_i + \varepsilon_i$ with $\mu_i \sim N(0, \sigma_i^2);\ cov(\mu_i, X_{it}) = 0$ or

$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \dots + \beta_k X_{kit} + \lambda_t + u_t$ with $\lambda_t \sim N(0, \sigma_t^2);\ cov(\lambda_t, X_{it}) = 0$

2. REM with two-way residual components:

$Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \dots + \beta_n X_{nit} + \mu_i + \lambda_t + w_{it}$ with $\mu_i \sim N(0, \sigma_i^2);\ cov(\mu_i, X_{it}) = 0$ and $\lambda_t \sim N(0, \sigma_t^2);\ cov(\lambda_t, X_{it}) = 0$

Panel Data Regression Estimation

In determining the estimation of the panel regression model, several tests were conducted to select the optimum estimation approach method. The first step in getting the desired model was the Chow test on the FEM estimation results; after proving that there was an individual effect, the Hausman test was carried out to determine between FEM and REM [1].

1) The Chow Test

The Chow test selects the two models between the common effect and fixed-effects models. The assumption that each cross-sectional unit has the same behavior tends to be unrealistic, considering that each cross-sectional unit can have different behavior is the basis of the Chow test. In this test, the following hypotheses are tested:

H_0: $\alpha_1 = \alpha_2 = ... = \alpha_N = \alpha$ (common effect model).

H_1: There is at least one different intercept α_1 (fixed-effects model).

The basis for rejecting H_0 is to use F-statistics as follows [1]:

$$Chow = \frac{(RSS1 - RSS2)/(N-1)}{RSS2/(NT-N-K)}$$

RSS1 - residual sum of the square of common effect model estimation results;

RSS2 - residual sum of the square of fixed-effects model estimation results;

N - number of cross-section units;

T - number of time-series units;

K - number of independent variables.

The Chow test statistics follow the distribution of F-statistics, namely $F_{(N-1,NT-N-K);\alpha}$. If the Chow statistic is greater than the F-table, there is sufficient evidence to reject H_0 and vice versa.

2) The Hausman Test

The Hausman test was used to compare the fixed-effects model with the random effect model. The Hausman test is conducted when the fixed-effects model contains an element of trade-off, namely the loss of the degree of freedom element by including dummy variables and the random effect model, which must heed the absence of assumptions violation of each component of the error. In this test, the following hypotheses are tested:

H_0: $\text{corr}(X_{it}, u_{it}) = 0$ (random effect model)

H_1: $\text{corr}(X_{it}, u_{it}) \neq 0$ (fixed-effects model)

The basis for rejecting H_0 using the Hausman statistics is expressed as follows [13]:

$$\chi^2(K) = (b-\beta)'[Var(b-\beta)] - 1(b-\beta)$$

with:

b - random effect coefficient;

β − fixed effect coefficient.

The Hausman statistics spread Chi-square; if the value of χ^2 is greater than $\chi^2_{(K,\alpha)}$ (K - number of independent variables) or P-value < α, there is sufficient evidence to reject H_0 and vice versa.

3) Lagrange Multiplier (LM) Test

This test was carried out to detect heteroscedasticity in the estimated model. The LM test hypotheses are as follows:

H_0: $\sigma_i^2 = \sigma^2$ (there is no heteroscedasticity).

H_1: $\sigma_i^2 \neq \sigma^2$ (there is heteroscedasticity).

LM test statistics are as follows [13]:

$$LM = \frac{NT}{2(T-1)} + \sum_{i=1}^{N}\left(\frac{T^2\sigma_i^2}{\sigma^2} - 1\right)^2$$

where:

T - number of time-series units;

N - number of cross-section units;

σ_i^2 − residual variance of equation *i*;

σ^2 −residual variance of the system equation.

H_0 is rejected if LM is greater than $\chi^2_{(1,\alpha)}$, which means heteroscedasticity occurs in the model. Thus, it must be estimated using the weight method: cross-section weight.

4) The Breusch-Pagan Test

The Breusch-Pagan test is an LM test to choose between a fixed-effects model and a pooled regression model. The initial hypothesis is that the variance of the residuals in the fixed coefficient model is zero. The procedure is as follows [1]:

Hypotheses

$$H_0: \sigma_\mu^2 = 0$$
$$H_1: \sigma_\mu^2 \neq 0$$

The test statistic used is the LM.

$$LM = \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}(\sum_{t=1}^{T}\hat{u}_{it})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2} - 1\right]$$

where:

N - number of individuals;

T - length of the period;

σ_μ^2 - model residual variance;

$\hat{u}_{it}$ - residual estimation of the individual fixed coefficient model i for period t.

If $LM > \chi^2_{(1,\alpha)}$ or p-value is less than the specified significance level, H_0 is rejected. Thus, the random effect model is selected.

B. Adaptive Neuro-Fuzzy Inference System (ANFIS)

The materials and sources used in this study include all articles discussing ANFIS, which combines neural networks (NNs) and a fuzzy inference system (FIS). Before we discuss the procedure of ANFIS modeling, the most important material that should be described in this section is the structure of ANFIS networks. The NN architecture applied in ANFIS consists of five fixed layers [5], [14]. Without loss of generality, the architecture of ANFIS for modeling time-series data is given two input variables x_1 and x_2 and a single output variable y by assuming a rule base of Sugeno first order with two rules as follows:

If x_1 is A_1 and x_2 is B_1, $y_1 = p_{11}x_1 + q_{12}x_2 + r_1$.

If x_1 is A_2 and x_2 is B_2, $y_2 = p_{21}x_1 + q_{22}x_2 + r_2$.

where x_i is A_j, x_2 is B_1, x_1 is A_2, and x_2 is B_2 as premise sections, whereas $y_1 = p_{11}x_1 + q_{12}x_2 + r_1$ and $y_2 = p_{21}x_1 + q_{22}x_2 + r_2$ are consequent sections; p_{11}, q_{12}, r_1, p_{21}, q_{22}, and r_2 are linear parameters; A_1, B_1, A_2, and B_2 are the nonlinear parameters. If the firing strength for two values y_1 and y_2 is w_1 and w_2, respectively, the output y can be expressed as in Equation (1).

$$y = \bar{w}_1 y_1 + \bar{w}_2 y_2 \quad (1)$$

where $\bar{w}_i = \frac{w_i}{\sum w_i}, i = 1,2$.

The structure of ANFIS networks (Figure 1) has five layers and can be explained as follows [5].

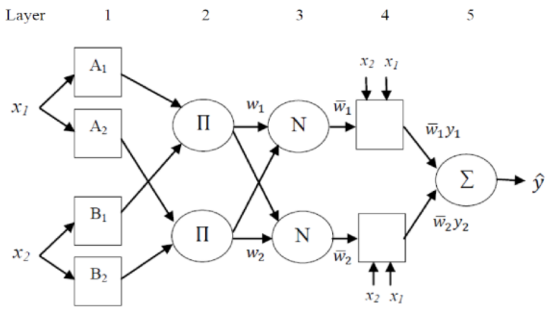

Figure 1. The structure of ANFIS networks for time series modeling [5]

Layer 1: Each neuron in this layer is adaptive to the parameters of an activation function. The output of each neuron is the membership degree of input. For example, the membership function of the generalized Bell is as follows:

$$\mu(x_i) = \frac{1}{1+\left|\frac{x_i - c_i}{a_i}\right|^{2b_i}}$$

where x_i is input, and a_i, b_i, and c_i are premise parameters [3], [4], [5].

Layer 2: Each neuron in this layer is a permanent neuron that is given the symbol Π, which is the product of all inputs in Layer 1: $w_i = \mu_{Ai}(x_1) \times \mu_{Bi}(x_2),\ i = 1, 2$.

Each neuron output is called the firing strength of a rule [15], [16], [17], [18], [19].

Layer 3: Each neuron in this layer is a fixed neuron with the symbol N, which is the result of calculating the ratio of the i-firing strength to the total number of firing strengths in the second layer as follows: $\bar{w}_i = \frac{w_i}{\sum w_i}, i = 1, 2$.

The results of calculations at this layer are called normalized firing strengths.

Layer 4: This layer is a neuron, which is an adaptive neuron to an output:

$$\bar{w}_i y_i = \bar{w}_i (p_i x_1 + q_i x_2 + r_i)$$

where $\bar{w}_i$ is the normalized firing strength in the third layer, while p_i, q_i, and r_i are parameters in these neurons called consequent parameters.

Layer 5: This layer is a single neuron with the symbol Σ, which is the sum of all outputs from the fourth layer, as follows:

$$y = \bar{w}_1 y_1 + \bar{w}_2 y_2,$$

where $\bar{w}_i = \frac{w_i}{\sum w_i}, i = 1,2$.

III. Method of Modeling

This research was based on a literature study. The initial step was to study in-depth and thoroughly the books and scientific articles that served as the basis for the new abstract system formation. We also examined supporting scientific articles that could be used in solving problems. At this stage, accuracy was needed in discussing supporting scientific articles, which were expected to solve the core problems. In addition to theoretical studies, applied studies were conducted. In detail, this research method is described as follows.

A. Data Source

The data used in this study were the human development index (%) and several factors that affect it in education, health, and population in 35 regencies and cities in Central Java Province from 2017 to 2019. All the data were obtained from the Central Bureau of Statistics of Central Java Province publications.

B. Research Variables

The variables used in this study were as follows:

a. Human development index (%) as response variable Y;

b. Junior high school participation rate (%) as an independent variable X1;

c. High school participation rate (%) as an independent variable X2;

d. Number of health workers as an independent variable X3;

e. Public health complaints (%) as an independent variable X4;

f. Population growth rate as an independent variable X5;

g. Poverty severity index (%) as an independent variable X6.

C. Analysis Method

The data analysis method used in this research was modeling using panel data regression analysis, bootstrapping regression, and RANFIS. The following steps were taken to analyze the data.

1) Panel Data Regression Modeling

1. A general description of the data was seen in data plots and descriptive statistics.
2. The best panel data regression model to model the effect of the junior high school participation rate, senior high school participation rate, number of health workers, public health complaints, population growth rate, and poverty on the human development index in Central Java was determined.
3. The common effect model, fixed-effects model, and random effect model were estimated.
4. The best model was determined through the Chow test, the Hausman test, and the Lagrange multiplier (LM) test. If the Chow test and the Hausman test showed the results of the fixed-effects model, there was no need to proceed to the Lagrange multiplier test.
5. The classical assumptions of regression on the selected model were tested.
6. The significance of the panel data regression parameters, including the simultaneous test (F-test), partial test (t-test), and measure of the goodness of the model with R-square, was tested.

The procedure of constructing panel data regression is illustrated in Figure 2.

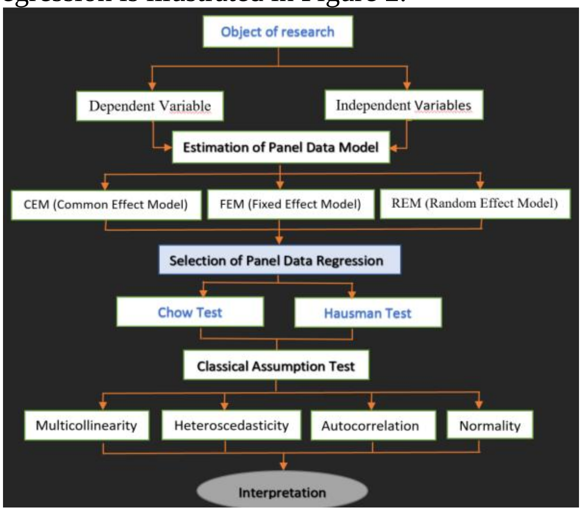

Figure 2. The procedure of constructing panel data regression

2) ANFIS Regression Modeling

The estimation steps of the RANFIS model for panel data were as follows.

1. Preprocessing was performed by estimating the classical panel data regression model.
2. A new response data was formed on the basis of the preprocessing results in Step 1.
3. ANFIS modeling took new responses as targets with input variables as in panel data regression modeling.
4. Several clusters and membership functions for input variables were defined.
5. Fuzzy if-then rules were generated for output variables based on input, cluster, rule, and the type of membership function. The fuzzy if-then rules were formed using the first-order Sugeno model.
6. Fuzzy inference system (FIS) training was conducted on an in-sample with a hybrid algorithm. The consequent parameters were estimated using a recursive LSE. The premise parameters were adjusted according to the backpropagation concept of gradient descent.
7. The predicted value in the in-sample was determined; the RMSE and MAPE were calculated.

The procedure of constructing the panel data RANFIS is illustrated in Figure 3.

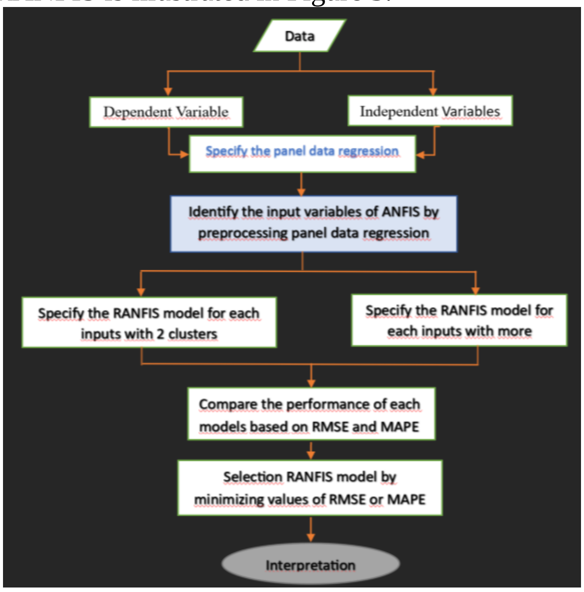

Figure 3. The procedure of constructing the panel data RANFIS

IV. Results and Discussion

A. Regression of Panel Data

1) Common Effect Model

According to the data processing of Central Java HDI 2017-2019, the estimation of the combined model (common effect model) was obtained as Equation (2).

$$\hat{y}_{it} = 43.975 - 0.025x_{1it} + 0.270x_{2it} + 0.0008x_{3it} + 0.065x_{4it} + 3.122x_{5it} + 3.127x_{6it} \quad (2)$$

2) Fixed-Effects Model

Fixed effect modeling of the human development index was carried out with the

RStudio program. The estimation result was obtained as Equation (3):

$$\hat{y}_{it} = \hat{c}_i + 0.276x_{1it} + 0.031x_{2it} - 0.0002x_{3it} + 0.030x_{4it} - 2.785x_{5it} - 1.997x_{6it} \quad (3)$$

with the value of $\hat{c}_i$ owned by each region in Central Java presented in Table 1.

Table 1.
Intercept estimation $\hat{c}_i$ for the fixed-effects model

i	Region	$\hat{c}_i$	i	Region	$\hat{c}_i$
1	Cilacap Regency	42.175	19	Kudus Regency	49.600
2	Banyumas Regency	46.306	20	Jepara Regency	46.815
3	Purbalingga Regency	44.943	21	Demak Regency	46.366
4	Banjarnegara Regency	41.898	22	Semarang Regency	48.394
5	Kebumen Regency	46.051	23	Temanggung Regency	43.084
6	Purworejo Regency	43.668	24	Kendal Regency	46.047
7	Wonosobo Regency	42.890	25	Batang Regency	42.462
8	Magelang Regency	44.018	26	Pekalongan Regency	44.475
9	Boyolali Regency	47.302	27	Pemalang Regency	40.184
10	Klaten Regency	46.261	28	Tegal Regency	40.945
11	Sukoharjo Regency	48.442	29	Brebes Regency	39.780
12	Wonogiri Regency	41.135	30	Magelang City	50.293
13	Karanganyar Regency	48.629	31	Surakarta City	53.419
14	Sragen Regency	45.788	32	Salatiga City	56.690
15	Grobogan Regency	43.474	33	Semarang City	59.217
16	Blora Regency	39.563	34	Pekalongan City	48.811
17	Rembang Regency	43.001	35	Tegal City	47.430
18	Pati Regency	44.184			

3) Random Effect Model

Random effect modeling of the human development index was carried out with the help of the RStudio program. The estimation result was obtained as Equation (4):

$$\hat{y}_{it} = \hat{c}_i + 33.515 + 0.337x_{1it} + 0.055x_{2it} - 0.00003x_{3it} + 0.029x_{4it} + 2.948x_{5it} - 2.172x_{6it} \quad (4)$$

with the value of $\hat{ci}$ owned by each region in Central Java presented in Table 2.

Table 2.
Intercept estimation $\hat{c}_i$ for the random effect model

i	Region	$\hat{c}_i$	i	Region	$\hat{c}_i$
1	Cilacap Regency	-2.375	19	Kudus Regency	49.600
2	Banyumas Regency	-0.652	20	Jepara Regency	46.815
3	Purbalingga Regency	1.986	21	Demak Regency	46.366
4	Banjarnegara Regency	-2.513	22	Semarang Regency	48.394
5	Kebumen Regency	-1.714	23	Temanggung Regency	43.084
6	Purworejo Regency	0.089	24	Kendal Regency	46.047
7	Wonosobo Regency	-0.615	25	Batang Regency	42.462
8	Magelang Regency	-2.817	26	Pekalongan Regency	44.475
9	Boyolali Regency	2.521	27	Pemalang Regency	40.184
10	Klaten Regency	2.058	28	Tegal Regency	40.945
11	Sukoharjo Regency	1.452	29	Brebes Regency	39.780
12	Wonogiri Regency	-2.467	30	Magelang City	50.293
13	Karanganyar Regency	1.612	31	Surakarta City	53.419
14	Sragen Regency	1.962	32	Salatiga City	56.690
15	Grobogan Regency	-0.795	33	Semarang City	6.609
16	Blora Regency	-4.377	34	Pekalongan City	2.183
17	Rembang Regency	-2.919	35	Tegal City	3.547
18	Pati Regency	-0.048			

B. Panel Data Model Selection

1) Selection of Common Effect and Fixed-Effects Models with the Chow Test

Calculation of the Chow test was carried out using the RStudio program; the value of F statistics was equal to 112.91, which is greater than F(0.05;5;98) = 2.30722, and p-value = 2.2e-16 is less than α =5%, so H_0 is rejected. Thus, there was an individual effect on Indonesia's energy consumption equation model, resulting in the fixed-effects model (FEM) as the appropriate model. Because the selected estimation model was the FEM model, the next test was the Hausman test, while the LM test did not need to be performed.

2) Selection of Fixed-Effects and Random Effect Models with the Hausman Test

The Hausman test calculation was carried out using the RStudio program and obtained p-value = 0.01444, which is less than α = 5%; therefore, H_0 was rejected. Thus, the correct estimation of the regression model for the human development index data in Central Java in 2017–2019 was to use the fixed-effects model.

3) Assumption Test

a) Residual Normality Assumption Test

The normality assumption test was done

using the Shapiro-Wilk test. Using RStudio, a statistical p-value of 0.7562 was obtained because the p-value is greater than α = 5%, and the residuals of the fixed-effects model followed a normal distribution.

b) Autocorrelation Test

The non-autocorrelation assumption test was done by run test. Based on the results of using RStudio, the statistical value of the p-value test was 0.202. The p-value is greater than α = 5%; hence, there was no serial correlation in the error component.

c) Heteroscedasticity Test

The Breusch Pagan Test is used to determine whether the residual covariance-variance of the fixed-effects model is homoscedastic or heteroscedastic. Based on the results using RStudio, the statistical value of the p-value test was 0.3768. The p-value is greater than α = 5%, so the residual covariance structure of the fixed-effects model was homoscedastic.

d) Multicollinearity Test

Through the correlation test with the RStudio program, the correlation value between the independent variables was not too low. The value was less than 0.8, which H_0 was not rejected. It can be concluded that the resulting model did not contain elements of multicollinearity.

4) Parameter Significance Test

1. *Simultaneous Test (F-Test)*

This test was conducted to estimate the fixed-effects model to determine whether the independent variables together influence the dependent variable. Based on the RStudio program results, the F count value was 9.245e+04 with a p-value of 2.2e-16. Because the p-value is less than 0.05, the independent variables together significantly affected the dependent variable Human Development Index.

2. *Partial Test (t-Test)*

The t-test aims to determine the significance of the influence of individual independent variables on the dependent variable by assuming that other variables are constant. Based on the results of RStudio, the value of $|t\text{-statistic}|$ for variables c, x_1, x_4, and x_6 was greater than the value of t(0.025;103), which was 0.980103, or p-value is less than 0.05. Thus, it can be concluded that the variables c, x_1 , x_4 , and x_6 had a significant influence on the dependent variable Human Development Index in Central Java.

Based on the results of the RStudio program, the R-squared value was 99.55%. The dependent variables were influenced by the junior high school participation rate, public health complaints, poverty severity index, and regional factors with Equation (5):

$$\hat{y}_{it} = \hat{c}_i + 0.276x_{1it} + 0.030x_{4it} - 1.997x_{6it} \quad (5)$$

C. Modeling Human Development Index (HDI) Data with RANFIS

To obtain an estimate of HDI data regression parameters, the RANFIS method was used based on classical panel data regression preprocessing. In general, the stages of ANFIS regression modeling include: determining input variables, forming clusters (membership functions), and forming fuzzy rules. Preprocessing was done by applying classical panel data regression to determine the optimal input. The optimal input variables selected in the ANFIS regression modeling were: the HDI variable as the response, with the predictor variables being junior high school participation rate (X1), public health complaints (X4), and poverty severity index (X6). Based on the sample data, the following results were obtained.

In the preprocessing of panel data regression modeling on HDI data and its predictor variables, the predictor variables that had a significant effect on HDI were the junior high school participation rate, public health complaints, and poverty severity index. These predictor variables were then used as inputs in the ANFIS process. After determining the input variables, the first step was to determine the membership function, the number of clusters, and the fuzzy rules that would be applied. This study determined clusters and rules using two methods: fuzzy C-means (FCM) and grid partition. Using hybrid algorithm learning technique on in-sample data, the RMSE and MAPE values were obtained. To generate FIS using the FCM technique, the membership function (MF) used was the Gaussian function. In this technique, the number of rules was equal to the number of clusters determined. There were no combinations in the formation of the rule. Meanwhile, to generate FIS using the grid partition technique, each rule formed combined the partition level for each input [20].

Optimal RANFIS modeling using the FCM technique with two input variables x_1 and x_2 with two membership functions (clusters), two first-order Sugeno rules can be formed as follows:

If x_1 is A_1 and x_2 is B_1, $y_1 = p_{11}x_1 + q_{12}x_2 + r_1$

If x_1 is A_2 and x_2 is B_2, $y_1 = p_{21}x_1 + q_{22}x_2 + r_2$

where A_1, B_1, A_2, and B_2 are nonlinear parameters or premises, and p_{11}, q_{12}, r_1 ,

p_{21}, q_{22}, and r_2 are linear or consequent parameters.

If the firing strength for the two values of y_1 and y_2 is w_1 and w_2, the output y could be determined as

$$y = \frac{w_1 y_1 + w_2 y_2}{w_1 + w_2}.$$

In Layer 1 of the RANFIS architecture, there are six groups of initial premise parameter values, with these values being used for the learning process. After obtaining the initial value of the premise parameters, the output generated in the first layer is the membership function of each input, $\mu_{A1}(x_1)$, $\mu_{A2}(x_1)$, $\mu_{B1}(x_2)$, and $\mu_{B2}(x_2)$. The membership function is used as input in Layer 2, which produces the degree of activation of each rule. The optimal RANFIS has two rules, so Layer 2 outputs are w_1 and w_2. Layer 2 output is used as input for Layer 3, which will be normalized at the activation degree, then Layer 3 output will be $\overline{\mathrm{w}}_1$ and $\overline{\mathrm{w}}_2$. The output of this layer is used as input in Layer 4, which will produce linear parameters or consequent p_{11}, q_{12}, r_1, p_{21}, q_{22}, r_2 from the recursive least squares estimator (LSE) [20].

Based on Central Java HDI data as a case study, the RANFIS model obtained can be written as follows:

$$y = 1.059\overline{\mathrm{w}}_{1,t}x_1 - 0.136\overline{\mathrm{w}}_{1,t}x_4 + 0.166\overline{\mathrm{w}}_{1,t}x_6 - 19.570\overline{\mathrm{w}}_{1,t} + 0.506\overline{\mathrm{w}}_{2,t}x_1 + 0.168\overline{\mathrm{w}}_{2,t}x_4 - 4.9445\overline{\mathrm{w}}_{2,t}x_6 + 16.034\overline{\mathrm{w}}_{2,t}$$

where

$$\overline{\mathrm{w}}_{1,t} = \frac{w_{1,t}}{w_{1,t}+w_{2,t}},$$

$$\overline{\mathrm{w}}_{2,t} = \frac{w_{2,t}}{w_{1,t}+w_{2,t}},$$

$$w_{1,t} = exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-96.358}{2.073}\right)^2\right] + \left(\frac{x_4-56.833}{5.929}\right)^2 + \left(\frac{x_6-0.345}{0.226}\right)^2\right\},$$

$$w_{2,t} = exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-95.599}{2.241}\right)^2\right] + \left(\frac{x_4-45.907}{5.718}\right)^2 + \left(\frac{x_6-0.438}{0.209}\right)^2\right\}.$$

From the learning process using the hybrid algorithm, the RMSE, AIC, and BIC values were 3.227, 246.976, and 249.630, respectively, while the MAPE value was 3.299%.

V. Conclusion

Based on the panel data regression modeling procedure applied to the Human Development Index (HDI) data in Central Java in 2017-2019, an estimation of the panel data regression model of the fixed-effects model was obtained. The HDI variable could be explained by the junior high school participation rate, public health complaints, and poverty severity index. Using the input variables selected through panel data regression, the optimal RANFIS model was obtained. The RANFIS optimal has three inputs with two clusters (membership functions). The performance of the RANFIS model was evaluated using the RMSE and MAPE criteria. The RMSE and MAPE values were 3.227 and 3.299, respectively. The RANFIS model performs well to apply to nonlinear data containing uncertainty.

Acknowledgment

This paper is officially funded by a grant from the Faculty of Science and Mathematics, Universitas Diponegoro in 2021 with contract number 2164/UN7.5.8.2/PP/2021.

References

[1] BALTAGI, B.H. (2005) *Econometric Analysis of Panel Data*. 3rd ed. Chichester; Hoboken, New Jersey: John Wiley & Sons.

[2] APRILIAWAN, D., TARNO, T., and YASIN, H. (2013) Pemodelan Laju Inflasi di Provinsi Jawa Tengah Menggunakan Regresi Data Panel. *Jurnal Gaussian*, 2 (4), pp. 311–321.

[3] JANG, J.-S.R. (1993) ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Transactions on Systems, Man, and Cybernetics*, 23 (3), pp. 665-668.

[4] JANG, J.-S.R. (1996) Input Selection for ANFIS Learning. In: *Proceedings of the 5th IEEE International Conference on Fuzzy Systems, New Orleans, Louisiana, September 1996*. Manhattan, New York: Institute of Electrical and Electronics Engineers, pp. 1493-1499.

[5] JANG, J.-S.R., SUN, C.–T., and MIZUTANI, E. (1997) *Neuro-Fuzzy and Soft Computing: A computational Approach to Learning and Machine Intelligence*. Upper Saddle River, New Jersey: Prentice Hall.

[6] FAUSSET, L. (1994) *Fundamentals of Neural Networks: Architectures, Algorithms and Aplications*. Englewood Cliffs, New Jersey: Prentice Hall.

[7] HAYKIN, S. (1999) *Neural Networks: A Comprehensive Foundation*. 2nd ed. Upper Saddle River, New Jersey: Prentice Hall.

[8] STATISTICS OF JAWA TENGAH PROVINCE (2019) *Indeks Pembangunan*

Manusia (IPM) Menurut Jenis Kelamin dan Kabupaten/Kota di Jawa Tengah, 2014-2019. [Online] Available from: https://jateng.bps.go.id/dynamictable/2019/01/23/58/indeks-pembangunan-manusia-ipm-menurut-jenis-kelamin-dan-kabupaten-kota-di-jawa-tengah-2014-2017.html [Accessed 14/10/20].

[9] STATISTICS OF JAWA TENGAH PROVINCE (2019) *Angka Partisipasi Sekolah (APS) (Persen), 2017-2019.* [Online] Available from: https://jateng.bps.go.id/indicator/28/71/1/angka-partisipasi-sekolah-aps-.html [Accessed 20/10/20].

[10] STATISTICS OF JAWA TENGAH PROVINCE (2019) *Jumlah Tenaga Kesehatan Menurut Kabupaten/Kota di Provinsi Jawa Tengah, 2019.* [Online] Available from: https://jateng.bps.go.id/statictable/2020/07/20/1874/jumlah-tenaga-kesehatan-menurut-kabupaten-kota-di-provinsi-jawatengah-2019.html [Accessed 29/10/20].

[11] STATISTICS OF JAWA TENGAH PROVINCE (2019) *Penduduk, Laju Pertumbuhan Penduduk, Distribusi Persentase Penduduk Kepadatan Penduduk, Rasio Jenis Kelamin Penduduk Menurut Kabupaten/Kota di Provinsi Jawa Tengah, 2010 dan 2019.* [Online] Available from: https://jateng.bps.go.id/statictable/2020/06/11/1792/penduduk-laju-pertumbuhanpenduduk-distribusi-persentase-penduduk-kepadatan-penduduk-rasio-jenis-kelaminpenduduk-menurut-kabupaten-kota-di-provinsi-jawa-tengah-2010-dan-2019.html [Accessed 29/10/20].

[12] STATISTICS OF JAWA TENGAH PROVINCE (2019) *Indeks Keparahan Kemiskinan (P2) (Persen), 2017-2019.* [Online] Available from: https://jateng.bps.go.id/indicator/23/78/1/indeks-keparahan-kemiskinanp2-.html [Accessed 29/10/20].

[13] GREENE, Q.W.H. (2002) *Econometric Analysis.* 5th ed. Upper Saddle River, New Jersey: New York University.

[14] TARNO, T., SUPARTI, S., and ISPRIYANTI, D. (2018) Modeling Cayenne Production Data in Central Java Using Adaptive Neuro Fuzzy Inference System (ANFIS). *Model Assisted Statistics and Applications*, 13 (1), pp. 45-52.

[15] GKOUNTAKOU, F. and PAPADOPOULOS, B. (2020) The Use of Fuzzy Linear Regression and ANFIS Methods to Predict the Compressive Strength of Cement. *Symmetry*, 12 (8), 1295.

[16] MOKARRAM, M., AMIN, H., and KHOSRAVI, M.R. (2019) Using adaptive neuro-fuzzy inference system and multiple linear regression to estimate orange taste. *Food Science & Nutrition,* 7 (10), pp. 3176–3184.

[17] KARABOGA, D. and KAYA, E. (2019) Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey. *Artificial Intelligence Review,* 52, pp. 2263–2293.

[18] SHAH, M.I., ABUNAMA, T., JAVED, M.F., BUX, F., ALDREES, A., TARIQ, M.A.U.R., and MOSAVI, A. (2021) Modeling Surface Water Quality Using the Adaptive Neuro-Fuzzy Inference System Aided by Input Optimization. *Sustainability,* 13 (8), 4576.

[19] HE, Z., WEN, X., LIU, H., and DU, J. (2014) A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *Journal of Hydrology,* 509, pp. 379-386.

[20] SARI, S.K., TARNO, T., and SAFITRI, D. (2017) Pemilihan Input Model Regression Adaptive Neuro Fuzzy Inference System (RANFIS) Untuk Kajian Data IHSG. *Jurnal Gaussian,* 6 (3), pp. 449-458.

[21] GUJARATI, D.N. (2004) *Basic Econometrics.* 4th ed. New York: McGraw-Hill Companies.

参考文:

[1] BALTAGI, B.H. (2005) 面板数据的计量经济学分析。第三版。奇切斯特；新泽西州霍博肯：约翰•威利父子公司。

[2] APRILIAWAN, D.、TARNO, T. 和 YASIN, H. (2013) 使用面板数据回归对中爪哇省的通货膨胀率进行建模。高斯期刊，2 (4)，第 311–321 页。

[3] JANG, J.-S.R. (1993) 情报局：基于自

适应网络的模糊推理系统。IEEE 系统、人和控制论汇刊，23 (3)，第 665-668 页。

[4] JANG, J.-S.R. (1996) 情报局学习的输入选择。载于：第 5 届 IEEE 国际模糊系统会议论文集，路易斯安那州新奥尔良，1996 年 9 月。纽约曼哈顿：电气和电子工程师协会，第 1493-1499 页。

[5] JANG, J.-S.R. 、 SUN, C.–T. 和 MIZUTANI, E. (1997) 神经模糊和软计算：学习和机器智能的计算方法。新泽西州上马鞍河：学徒霍尔。

[6] FAUSSET, L. (1994) 神经网络基础：架构、算法和应用。新泽西州恩格尔伍德悬崖：学徒霍尔。

[7] HAYKIN, S. (1999) 神经网络：综合基础。第二版。新泽西州上马鞍河：学徒霍尔。

[8] 中爪哇省统计数据（2019）2014-2019 中爪哇按性别和地区/城市划分的人类发展指数（人类发展指数）。[在线] 可从：https://jateng.bps.go.id/dynamictable/2019/01/23/58/indeks-pembangunan-manusia-ipm-menurut-jenis-kelamin-dan-kabupaten-kota-di- jawa-tengah-2014-2017.html [访问时间 14/10/20]。

[9] 中爪哇省统计数据 (2019) 入学率(APS)（百分比）统计数据，2017-2019。[在线] 可从：https://jateng.bps.go.id/indicator/28/71/1/angka-partisipasi-sekolah-aps-.html [20/10/20 访问]。

[10] 中爪哇省统计数据（2019）2019 年中爪哇省各地区/城市卫生工作者人数。[在线] 可从：https://jateng.bps.go.id/statictable/2020/07/20/1874/jumlah-tenaga-kesehatan-menurut-kabupaten-kota-di-provinsi-jawatengah-2019.html [2020 年 10 月 29 日访问]。

[11] 中爪哇省的统计数据（2019）中爪哇省人口、人口增长率、人口百分比、人口密度分布、按县/市划分的人口性别比，2010 年和 2019 年。[在线] 可从：https://jateng.bps.go.id/statictable/2020/06/11/1792/penduduk-laju-pertumbuhanpenduduk-distribusi-persentase-penduduk-kepadatan-penduduk-rasio-jenis-kelaminpenduduk-menurut-kabupaten-kota-di-provinsi-jawa-tengah-2010-dan-2019.html [访问时间 29/10/20]。

[12] 中爪哇省统计数据(2019) 贫困严重程度指数(P2)（百分比）统计数据，2017-2019。[在线] 可从以下网址获得：https://jateng.bps.go.id/indicator/23/78/1/indeks-keparahan-kemiskinanp2-.html [访问时间 29/10/20]。

[13] 格林，Q.W.H. (2002) 计量经济学分析。第 5 版。新泽西州上马鞍河：纽约大学.

[14] TARNO, T., SUPARTI, S. 和 ISPRIYANTI, D. (2018) 使用自适应神经模糊推理系统(情报局)对中爪哇的辣椒生产数据进行建模。模型辅助统计和应用，13 (1)，第 45-52 页。

[15] GKOUNTAKOU, F. 和 PAPADOPOULOS, B. (2020) 使用模糊线性回归和情报局方法预测水泥的抗压强度。对称性，12 (8), 1295。

[16] MOKARRAM, M., AMIN, H., 和 KHOSRAVI, M.R. (2019) 使用自适应神经模糊推理系统和多元线性回归来估计橙味。食品科学与营养，7 (10)，第 3176–3184 页。

[17] KARABOGA, D. 和 KAYA, E. (2019) 基于自适应网络的模糊推理系统 (ANFIS) 训练方法：综合调查。人工智能评论，52，第 2263–2293 页。

[18] SHAH, M.I., ABUNAMA, T., JAVED, M.F., BUX, F., ALDREES, A., TARIQ, M.A.U.R., 和 MOSAVI, A. (2021) 使用自适应神经模糊推理系统模拟地表水质量 借助输入优化。可持续性，13 (8), 4576。

[19] HE, Z., WEN, X., LIU, H., 和 DU, J. (2014) 人工神经网络、自适应神经模糊推理系统和支持向量机预测半干旱河流流量的比较研究 山区。水文学杂志，509，第 379-386 页。

[20] SARI, S.K., TARNO, T., 和 SAFITRI, D. (2017) 佩米利汉输入模型回归自适应神经模糊推理系统(兰菲斯) 联合会数据研究。高斯期刊，6 (3)，第 449-458 页。

[21] GUJARATI, D.N. (2004) 计量经济学基础。第 4 版。纽约：麦格劳-希尔公司。