# PROJECT REPORT (sample)

**Fundamentals of Artificial Intelligence**

Applied Multivariate Regression model for Improvement of Performance in Labor Demand Forecast

Hai Pham Van1, Nguyen Dang Khoa2

1 School of Information Technology and Communication, Hanoi University of Science and Technology, Hai Ba Trung, Hanoi, Vietnam.

2 Hanoi University of Science and Technology, Hai Ba Trung, Hanoi, Vietnam.

**Abstract.** Forecasting labor demand play significant roles in the development of planning policies in both national labor markets with strategies of national human resources in order to meet the requirements of industrialization and modernization of the country. This paper has presented the proposed model using multivariate multiple regressions, dealing with time series for forecasting labor demand. It is also quantified influenced factors for prediction of labor demands. The contribution in this study is to propose multivariate regression approach that estimates a single regression model with outcome variable by dealing time series data. To confirm the proposed model, experimental results the proposed model in a real case study show that the multivariate regression model enhance significant performance for forecasting labor demand. The proposed model is evaluated on real world datasets to demonstrate this method’s effectiveness.

**Keywords:** Labor Demand Prediction, Labor Forecast, multiple regression, multivariate regression, intelligent labor forecast.

1. Introduction

Forecasting the labor demand in the future figures out the lack or redundancy of labor among sectors, provinces, locals, regions, degrees, for matching policies of human resources with its labor effectiveness. Labor forecast is an important task, emphasized through various works. In related works, Yas A. Alsultanny used data mining with Naïve Bayes Classifiers and created the Decision rules technique, which they recommended in predicting labor forecast [1]. Ross Gruetzemacheret et. al studied labor displacement in the advances of AI, with 90% and 99% of human tasks in the range between 10 and 15 years, respectively [2]. A. Luz et al. also studied effect from local labour demand on immigrant employment [3]. Studies of labor demand forecast impact using multivariate regression method for prediction of historical data [4,5]. Some studies combined with the time series prediction method [6,7,8] in order to forecast demand for labor economic sectors. Yalcinkaya, A. et. al studied Maximum likelihood estimators of the model parameters in multiple linear regression obtained genetic algorithm, which they proved to outperform traditional algorithms in most cases, and suggested using GA to obtain maximum likelihood estimators in specific cases [4]. Pan, Y. et. al used multiple linear regression and life-cycle cost analysis for cost-effective evaluation of pavement maintenance, showing the ability to apply multiple regression to establish decision – making systems [5].

Multivariate regression is an extension of multiple regression, which can allow more than one response for each input, making it better than multiple regression for predictions. Recent technological advancements and developments have led to a dramatic increase in the amount of high-dimensional data for proper and efficient multivariate regression methods. Xiaoxi H. et al. extended the scope of multivariate regression with sparse reduced rank regression and subspace assisted regression with row sparsity. The study has been enhanced model with improved interpretability of regression models [6]. Consonni, V. et al. (2021) described the regression toolbox for multivariate regression using MATLAB, which majorly contributed to the improvement of multivariate regression application in general [7]. L. Lucy and Z. Julie (2020) studied minimax D-optimal designs for multivariate regression models with multi-factors, hence using multivariate regression to introduce a design robust against small departures from an assumed error covariance matrix [8]. W. Yihe and Z.S. Dave used nonparametric empirical Bayes approach to large-scale multivariate regression, suggesting methods to improve multivariate regression with big data input [9]. Some approach was used by P.V. Hai to enhance uncertainty model [10]. In multiple variables in forecast under uncertainty. L.H. Son et al. proposed a new method for Hospital Cost Analysis using genetic algorithm and artificial neural network, suggesting the application of genetic algorithm in building decision – making systems that allow more variable and more accuracy [12]. P.V. Hai and N.T. Dong presented the Hybrid Louvain-Clustering model using a knowledge graph to cluster contents based on user behaviors in a social network [13]. The result is a model representing all multi-dimensional user relationships of contents based on users’ behaviors. D.X. Truong and P.V. Hai presented the Bayesian graph deep learning framework for the case of classified mixed node random block models to classify the topic of social posts as nodes by creating a homogeneous graph with links between them, showing improved performance of the Bayesian formulation in topic classification in social during the training process [14]. N.T. Dong and P.V. Hai also proposed a new graph deep learning model associated with knowledge graph with to prediction model the latent feature of user and item, supplying the principle of organizing interactions as a graph, combines information from social network and all kind of relations in the heterogeneous knowledge graph [15].

This paper proposes a model using multivariate multiple regressions, dealing with time series for forecasting labor demand. The approach is to propose multivariate regression approach by dealing time series data. To do an experiment in a real case study show that the multivariate regression model enhances significant performance for forecasting labor demand.

1. The proposed model
   1. Applied regression function to the Proposed model for labor forecast

In labor forecast, the relationships of labor demands are influenced by many factors [variables] such as GDP, export, import investment, science and technology, wages ... Hence, dependent variable *Y* depends on various explanatory variables. The random overall regression function of the proposed model with *k* variables can be expressed as follows:

(1)

Where is the cutting coefficient, *t* = 1,2,3, …, *n*

* are particular regression coefficient,
* is a parameter estimation
* *t* is tth observation
* *n* is the whole scale of overall

It supposes that *n* observations, each observation consists of *k* values () with *i* = 1 ÷ *n*. It is expressed by Eq.(2).

(2)

(3)

Hence, Eq. (1) is expressed by:

(4)

* 1. Hypothesis of multivariate regression models for Labor forecast

Hypothesis of multivariate regression models for Labor forecast is expressed by

(5)

+ Hypothesis 1:

+ Hypothesis 2:

Where *I* is the n levels matrix

+ Hypothesis 3: () is determined

+ Hypothesis 4: There is no multi-collinearity phenomenon among the explanatory variables or the rank of the matrix *X* by k: R (*X*) = *k*

* 1. Applied the proposed model for labor forecasting

**Step 1**- **Find the best regression function**

Calculate a regression function corresponding to each regression function which is expressed by

(6)

Optimize Min {}, which regression function has the smallest AIC is the selected function.

**Step 2- Calculate coefficients**

Vector is expressed by

(7)

**Step 3**-**Constructe a regression function for its estimation**

(8)

**Step 4- Forecast labor demand**

Calculate and Var

Calculate standard errors of and SE

Forecast the average confidence interval of is expressed by

E (9)

SE

* 1. Parameter estimation

The most common method used in order to estimate regression coefficients which is the smallest method of normal squares (MLS). The regression functions is expressed by

(10)

(11)

MLS estimates which can be expressed by

(12)

is the sum of square of remainder

The symbol is the matrix transposition of

Hence, it can be expressed by

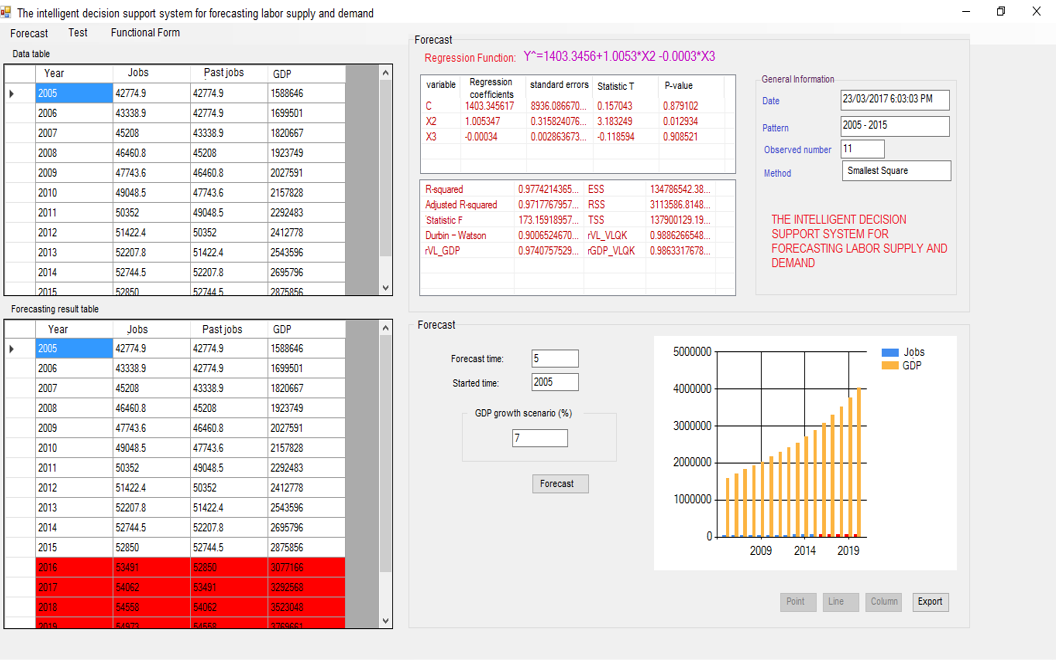
(13)

(14)

As referred from Eq. (9) we calculate the matrix of parameter estimates of .

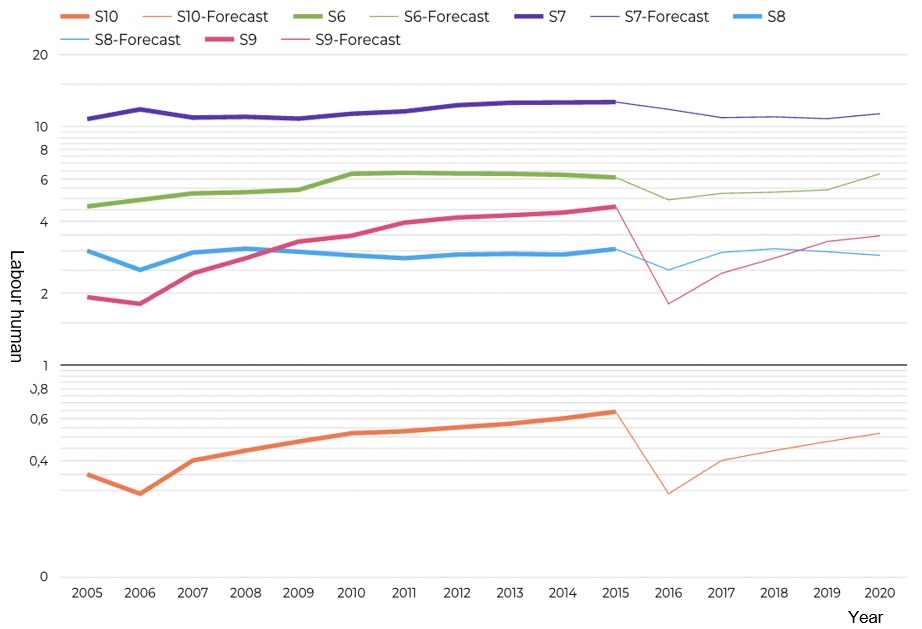
1. Experimental Results

In this paper, the proposed approach to has been designed to enable the creation of an effective model which provides a basis upon which the prediction of labor, as shown in Figure 1.



**Fig. 1** Results of labor forecast on the screen

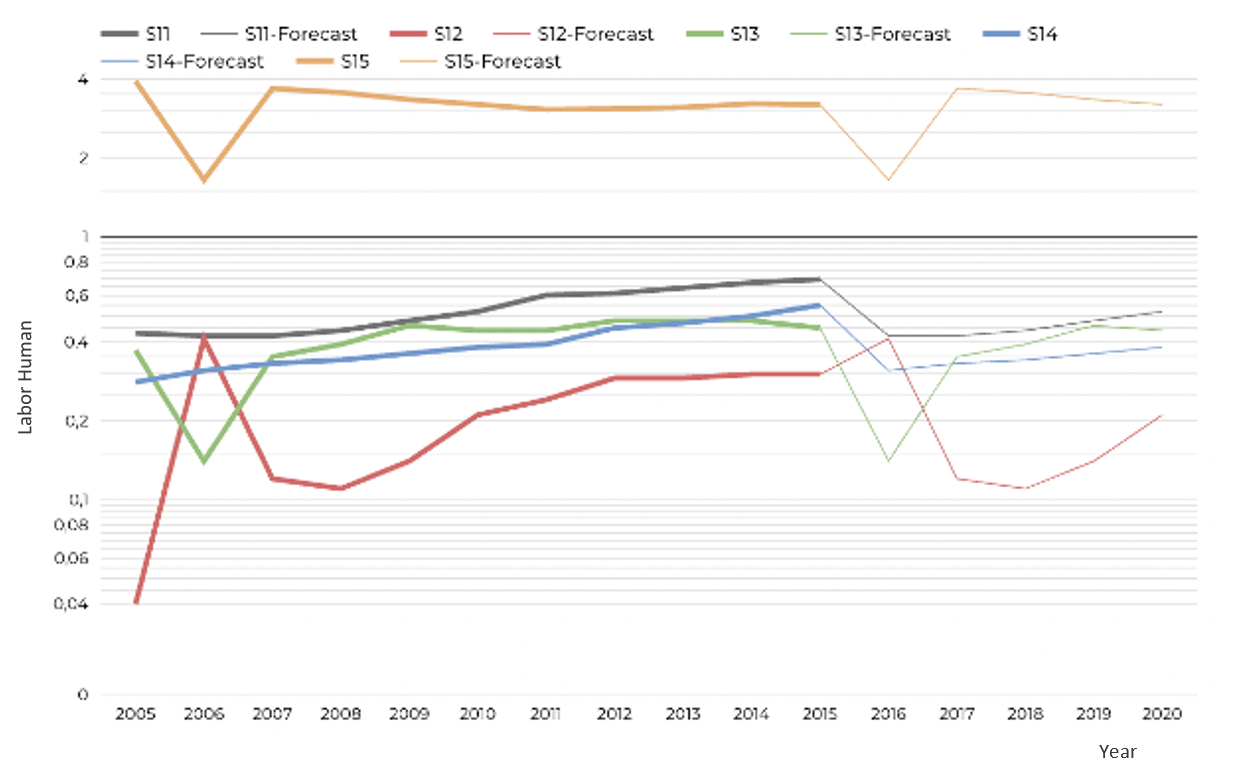
|  |
| --- |
| In research labor demand it is influenced by factors of human working jobs as follows [2,10]:  [Agriculture, fishery and forestry](https://en.wikipedia.org/wiki/Economy_of_Vietnam#Agriculture,_fishery_and_forestry)  Extractive  Processing, Making industry  Producing and distributing electricity, gas, hot water, steam and air conditioning  Water supply; waste and wastewater management and treatment activities  Construction  Wholesale and retail, car, motorcycle and motorbikes and other motor vehicles repair  Transportation, Warehouse  Accommodation and food services  Information and communication  Financial, bank and insurance activities  Estate business activities  Professional science and technology activities.  Administrative activities and support services  Activities of the Communist Party, Social and Political organizations; State management. national security; Compulsory social assurance  Education and training  Health and social assistance activities  Art, fun and entertainment  Other service activities  Activities of hiring jobs in households, production of material products and self-consumption services of households  Activities of international organizations and agencies  The proposed model has been tested using data sets from Statistic Government [10]. Experimental results show that labor forecast, as shown in Figure 2. |



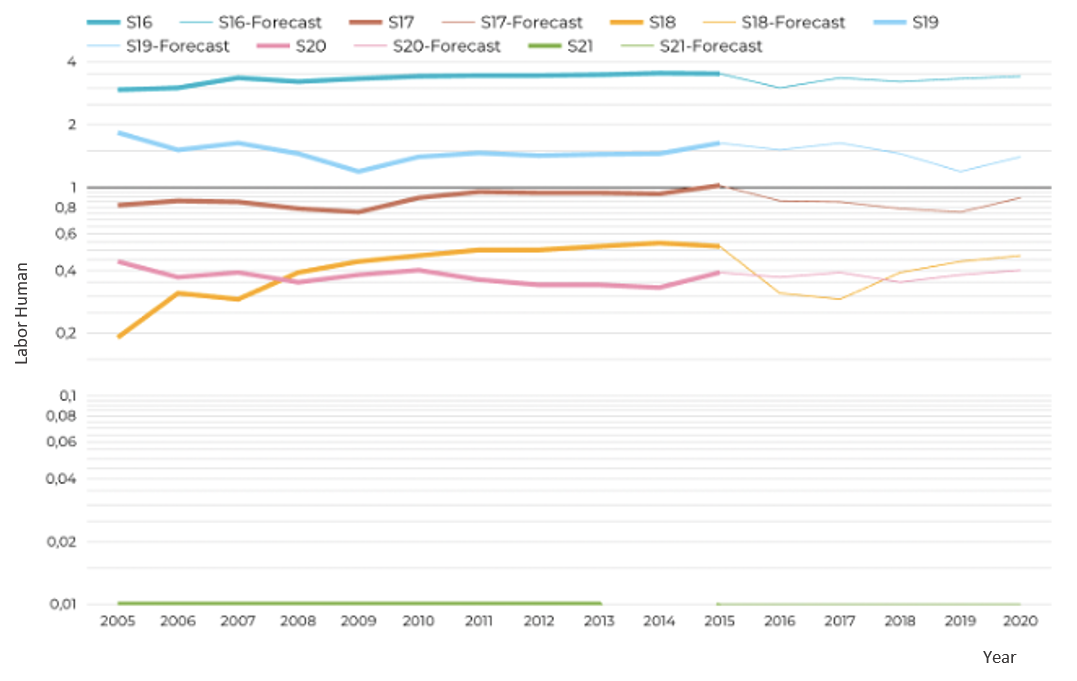
**Fig. 2.** Results of factors for labor forecast (S6-S10)

Figure 2 shows the annual data on percentage in the employed structure of 5 sectors. It is also reached at 6,34% in 2020. of wholesale and retail, car, motorcycle and motorbikes and other motor vehicles repair also accounted for a significant share of the employed structure and experienced little changes, reached at 12,66% in 2015 and expected at 11,31% in 2020. of labors in transportation and warehouse fluctuated around the mark of 3% to achieve 3,02% and 3,07% at the start and end of the 10 years. By the end of our forecast period, it was expected to fall back to 2,89%. of accommodation and food services rose sharply from 1,93% in 2005 to 4,62% in 10 years later, however it was expected to reach 3,49% at the end of 2020. of Information and communication is rather insignificant share around 0,5% by 2020.

Figure 3 shows the annual data on the percentage in the employers of 5 sectors from to . of labors of financial, bank and insurance activities, of estate business activities, of professional science and technology activities, and of administrative activities and support services overall rose from 2005 to 2015. of activities of the communist party, social and political organizations; state management. national security; compulsory social assurance held a larger share around 3 to 4%, with 3,19% in 2015 and expectedly 3,2% in 2020.



**Fig. 3.** Results of labor sectors’ forecast (S11-S15)



**Fig. 4.** Results of labor sectors’ forecast (S16-S21)

Figure 4 shows the annual data on percentage in the employed structure of 6 sectors from to . of labors in education and training experienced an overall increase in the data collecting period, from 2,94% to 3% and was expected to rise to 3,41% in 2020. of other service activities slightly fluctuated, finished at 1,63% in 2015 and was projected to reach 1,4% in 2020. of health and social assistance activities, of Art, fun and entertainment, and of activities of hiring jobs in households, production of material products and self-consumption services of households all accounted for approximately less than 1% of the employed structure. Their shares of the total employed in 2015 were 1,02%, 0,52% and 0,39% respectively, while these figures forecasted for 2020 were 0,89, 0,47 and 0,4%. Notably, from 2005 to 2015, S21 of Activities of international organizations and agencies always maintained at 0,01% with an exception of going to 0% in 2014. This labor line pattern was expected to continue going to 2020.

In evaluation, the proposed model has been tested with data sets [10], under the same conditions with the comparisons for the basic method of regression models. Forecast the average confidence interval of is given by Eq.(9). The results show that an average of the proposed model as labor rate through the strong correlation coefficient (R)% for 97,15%, respectively, as well as the percentage of accuracy (AA)% for 94,3%, respectively.

**Table 3.** Comparisons of the proposed model and MLRM performance

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **R%** | | **AA%** | |  | |
| **Multiple Linear Regression Method** | | 94,2 % | | 85% | |  | |
| **The proposed model** | | 97 % | | 94,3% | |  | |

|  |  |  |
| --- | --- | --- |
|  | **R%** | **AA%** |
| **Multiple Linear Regression Method** | 94,2 % | 85% |
| **The proposed model** | 97 % | 94,3% |

As shown in Table 3, the experimental results also show that the proposed approach performs better than MLRM method in term of correlation coefficient and accuracy.

1. Conclusion and direction of development.

In this paper, we have presented a novel method which targets improvements in labor forecast using real data sets. Our proposed approach uses multivariate regression approach that estimates a single regression model with outcome variable by dealing time series data. Experimental results demonstrate that the proposed model can provide enhanced forecast accuracy with respect to time series in data sets. Furthermore, the results derived from the testing [using real data sets] have confirmed that overall performance of the proposed method shows an improvement over the MLRM method. In summary, our reported results demonstrate the potential of our proposed model.

Further study for improvement in labor forecast accuracy, future research will address influenced factors, indicators to overall index of labor demand forecast.

References

1. Yas, A.A.: Labor Market Forecasting by Using Data Mining. In: Procedia Computer Science, vol. 18 (2013)
2. Ross, G., et al.: Forecasting extreme labor displacement: A survey of AI practitioners. In: Technological Forecasting and Social Change, vol. 161 (2020)
3. Luz, A., et al.: Local labour demand and immigrant employment. In: Labour Economics, vol. 63 (2020)
4. Yalcinkaya, A., et al: A new approach using the genetic algorithm for parameter estimation in multiple linear regression with long-tailed symmetric distributed error terms: An application to the Covid-19 data. In: Chemometrics and Intelligent Laboratory Systems, vol. 216 (2021)
5. Pan, Y., et al.: Cost-effectiveness evaluation of pavement maintenance treatments using multiple regression and life-cycle cost analysis. In: Construction and Building Materials, vol. 292 (2021)
6. Xiaoxi, H., et al.: Expanding the Scope of Multivariate Regression Approaches in Cross-Omics Research. In: Engineering (2021)
7. Consonni, V., et al.: A MATLAB toolbox for multivariate regression coupled with variable selection. In: Chemometrics and Intelligent Laboratory Systems, vol. 213 (2021)
8. Lucy, L., Julie, Z.: Minimax D-optimal designs for multivariate regression models with multi-factors. In: Journal of Statistical Planning and Inference, vol. 209, pp. 160-173 (2020)
9. Yihe, W., Dave, Z.S.: A nonparametric empirical Bayes approach to large-scale multivariate regression. In: Computational Statistics & Data Analysis, vol. 156 (2021)
10. General Statistics Office, <https://www.gso.gov.vn/en/statistical-data/>
11. Lan, L.T.H., et al.: A new complex fuzzy inference system with fuzzy knowledge graph and extensions in decision making. In: IEEE Access, vol.8, 164899-164921
12. Son, L.H., Ciaramella, A., Thu, D.T., Stanio, A., Tuan, T.M., Hai, P.V.: Predictive reliability and validity of hospital cost analysis with dynamic neural network and genetic algorithm. In: Neural Computing and Application, pp. 15237-15248 (2020)
13. Hai, P.V., Dong, N.T.: Hybrid Louvain-Clustering Model Using Knowledge Graph for Improvement of Clustering User’s Behavior on Social Networks. In: The International Conference on Intelligent Systems & Networks, pp. 126-133. Springer, Hanoi (2021)
14. Hai, P.V., Truong, D.X.: Social Network Analysis Based on Combining Probabilistic Models with Graph Deep Learning. In: Communication and Intelligent Systems, pp. 975-986. Springer, India (2021)
15. Hai, P.V., Dong, N.T.: Graph Neural Network Combined Knowledge Graph for Recommendation System. In: International Conference on Computational Data and Social Networks, pp. 59-70. Springer, Montreal (2020)