ICIC Express Letters
Part B: Applications
Volume **14**, Number **4**, April **2023**

ICIC International ©2023 ISSN 2185-2766
pp. 349–357

HETEROGENEITY OF STUDENTS' PROGRAM PARTICIPATION IN EXPANDING HIGHER EDUCATION INTERPRETED BY THE BLAU INDEX AND ARIMA

Dian-Fu Chang[1], En-Yu Liu[1,*] and Bih-Yuh Chen[2]

[1]Department of Education and Futures Design
[2]Doctoral Program in Foresight for Educational Leadership and Technology Management
Tamkang University
No. 151, Yingzhuan Road, Tamsui District, New Taipei City 251301, Taiwan
140626@mail.tku.edu.tw; bihchen522@gmail.com
*Corresponding author: ninie00013@gmail.com

Received September 2022; accepted November 2022

Abstract. *Higher education expansion has become global concern issues, while a little study addressed whether the patterns of program participation have been changed and to what level the trend will be in future. This study focused on the heterogeneity of students' program participation in higher education. Taking Taiwan's higher education system as an example, we employed the concept of Blau index and conducted autoregressive integrated moving average (ARIMA) models to tackle the issue. The series data were aggregated from the government's data sets from 1950 to 2020. Blau index was used to transform the heterogeneity data of the program participation in the target higher education. ARIMA was used to select a fitted predicated model and project future trend for the target series. Based on the parameters and the residual test, the findings suggest the Blau index and ARIMA model work well to deal with this issue. Considering many countries have moved to over expanded higher education, the application of Blau index and ARIMA for detecting heterogeneity of program participation can be extended to solve similar issues in the other higher education systems.*

Keywords: ARIMA, Blau index, Enrollment, Higher education, Organizational heterogeneity

1. **Introduction.** Over the last 20 years, enrollment in higher education has experienced explosive growth across Asia and other countries in the world. According to data from the UNESCO Institute for Statistics 2018, the gross enrollment ratio (GER) in high-income countries moved to the universal system (GER over 50%) in 1993 [1]. The expansion has extended to most middle-income countries and to a significant number of low-income countries. The expansion phenomenon revealed some countries reached 75% in 2011 and most middle-income countries stepped into the mass stage (GER 15%-50%) in 2001 [2-4]. This phenomenon can be attributed to a higher birth rate, increased school participation, and the perception that higher education is important for subsequent employment opportunities [5,6]. It is commonly perceived that higher education expansion may change the patterns of student participation in different programs. Various studies address higher education expansion phenomena [4,7-10], whereas a little study concerns whether the patterns of program participation have been altered in higher education. Specifically, the data related to the program participation did not be reviewed cautiously in higher education settings. Moreover, when higher education is considered from a universal perspective and the pressure of balancing supply and demand is increasing, the question needs to be asked about patterns of program participation. If we persist neglecting student's program

DOI: 10.24507/icicelb.14.04.349

participation, it might cause negative effect of higher education expansion and result in management and unemployment problems. This is why we selected this topic to address.

Higher education in Taiwan has expanded dramatically from 1984 to 2014, and the number of students enrolled in higher education institutions has nearly quadrupled [12]. An overview of higher education in the past few decades revealed that the number of students increased from 299,486 (1976) to 576,623 (1999) and the GER rose from 15 to 50% within 23 years [11]. The expansion of higher education in Taiwan could be driven by the hard concept of managerialism [13], for example, the expanding of STEM (science, technology, engineering and mathematics) usually has an economic agenda. Whether the program expansion has shown that supply of the system is driven by the demands? If this phenomenon exists in the expansion process, what will happen in the future? This study aims to explore the patterns of student's program participation in higher education expansion process. This topic may include series data and their future trends. It is different from that of traditional cross-section method. Taking Taiwan's higher education expansion as an example, this study attempts an alternative way to explore the expansion phenomenon. Social researchers follow population series, such as birth rates or school enrollments [14,15]. Although time series analysis has been used for a long time, its specific application to higher education participation issues is still limited. Various time series studies have focused on ARIMA (autoregressive integrated moving average) models with serial data [16-18], while using series data is an emerging area in social science [19,20]. Tackling a new area with a novel study design, this study may fill a research gap in current higher education. Hopefully, the findings can contribute knowledge in the field. Bringing this purpose in mind, this study will explore the following research questions:

a) What are the trends of program participation in the higher education system?
b) What are the patterns of heterogeneity in the major program participation?
c) What is the trend of the heterogeneity index in next decade?

The rest parts of the paper are organized as follows. First, the method section will address the data collection, data transformation, and model building; Second, the result will demonstrate the trend of program participation with student numbers and the heterogeneity index; Finally, the conclusion will be drawn and suggestions will be addressed.

2. **Method.** This study employs the concept of Blau index and autoregressive integrated moving average (ARIMA) model to tackle program participation issue in higher education. First, we define the target data in the data warehouse (government's data set from 1950 to 2020 in Taiwan) to clarify and calculate the number of students participating in different programs [21]. Second, the index data sets were transformed and the fitted ARIMA models were selected. Third, the robustness of the forecasting model was evaluated from 2021 to 2030. Finally, the diversity of program participation with the Blau index was interpreted.

2.1. **Definition of explanation model.** Trow has defined higher education expansion phenomenon with three stages. The GER was used to classify the expansion stages. The GER in the elite stage is under 15%, the mass stage is from 15% to 50%, and the universal stage is over 50% [22]. This study employed the notion of three expanding stages to explore the program participation patterns. We focused on the effect of expansion on the heterogeneity of program participation in the target higher education system.

2.2. **Exploring program participation patterns.** The Blau index has been commonly used to assess organizational heterogeneity [23,24]. A high index value indicates a high degree of heterogeneity; theoretically, the minimum value is 0 and the maximum value is 1, depending on the number of categories, and is calculated as $(n-1)/n$. For example, the gender variable has two categories ($n = 2$), which implies that the maximum heterogeneity index is 0.5. Similarly, when a variable has four categories ($n = 4$), the maximum heterogeneity index is 0.75 [25]. The various categories can be standardized to compute the index. Previous studies have shown there are a couple of indicators used for evaluating the heterogeneity of organization. For instance, Blau index, standard deviation, and coefficient of variation are useful detecting criteria [26]. With a quadratic function, the Blau index can be more sensitive to smaller values and may thus be more appropriate to capture critical program effects. The standardized Blau index is defined as follows [27,28]:

$$1 - \sum_{i=1}^{R} p_i^2$$

Based on the formula, there are three categories in this study, so $R = 3$ and with program i, at year t. p is the proportion of the program participation in humanity, social science, and STEM. This study defines the Blau index in the system and the specific programs. In our data transformation, "Blau_Programs" refers to the Blau index for reflecting the students participated in humanity, social science, and STEM programs according to the change of participation ratios.

2.3. **Time series analysis.** Time series analysis has been used for wide settings, for example, environment, medicine, and social-related issues [29-31]. Various studies have discussed the innovative approach to transform the time series data [32-34]. This study applied time series analysis to predicting Blau index. We conducted time series analysis to explore the trend of Blau index and build fitted models for projecting. Before the model conducting, this study will verify the series data set which belongs to seasonal or non-seasonal data. If the series data set is non-seasonal pattern, we will follow the criteria of ARIMA(p, d, q) to select the fitted model. In the selecting process, we will consider the meaning of the parameters: the p as the order of the autoregressive part in terms of the number of autoregressive terms (AR) while d as the difference; the q as the order of the moving average, i.e., the number of lagged forecast errors in the prediction equation (in terms of MA) [29,35-37]. The white noise examination will use the Box-Pierce Chi-square statistic test to determine whether the model met the assumption that the residuals were independent [38,39]. Based on the Box-Pierce's suggestion, the Chi-square values should be no significant in the legs of 12, 24, 36 and 48.

3. Results.

3.1. **Student participation in major programs.** Figure 1 displays the number of students in the three major programs from 1950 to 2020. The STEM program attracted a large part of students, while the humanity program could be a disadvantaged one. Figure 1 reveals that the program expansion has stopped and the trend has declined in the last decade.

3.2. **Heterogeneity of program participation.** The term of "Blau_Programs" refers to Blau index in the three major programs from 1950 to 2020. Figure 2 shows the elite stage of the higher education system (from 1950 to 1975) with high Blau index in terms of more diversity of program participation. Similarly, the universal stage (after 1999) is also with high Blau index. The result indicates when higher education expanding from elite to mass stage, the program participation might become friendly for all the participants. However, over expansion phenomenon did not provide a better program balancing in the target system.

3.3. **Future trend of program participation.** We applied the Blau index with programs participation from 1950 to 2020 to predicting the series toward 2030. The candidate models include ARIMA(1,1,1), ARIMA(1,1,0) and ARIMA(2,1,1). The information of ACF, PACF, and modified Box-Pierce (Ljung-Box) Chi-square statistic is presented in Table 1. Based on one difference with the series of Blau index of program participation, the result of Minitab suggests that the ARIMA(2,1,1) is the fittest model.

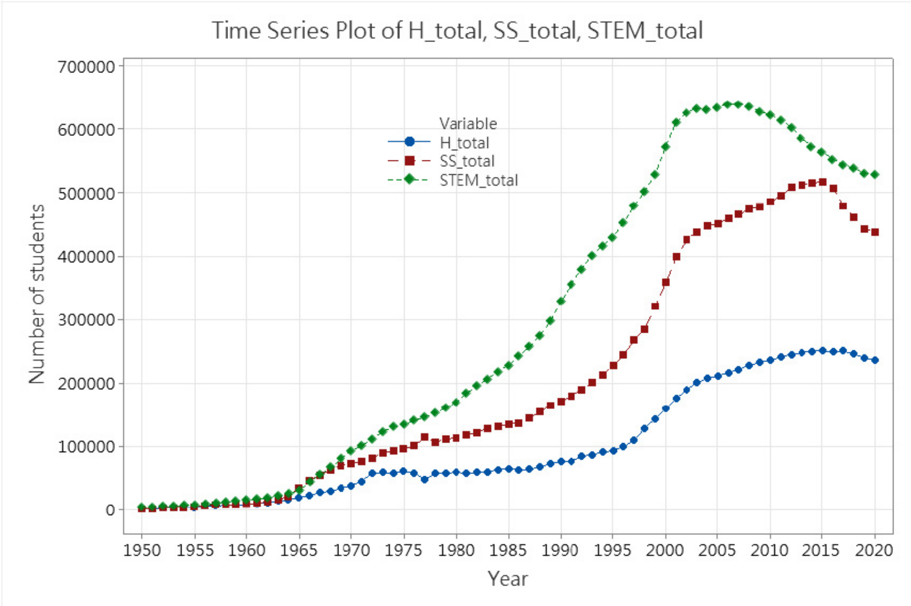

Figure 1. The number of students in the three major programs

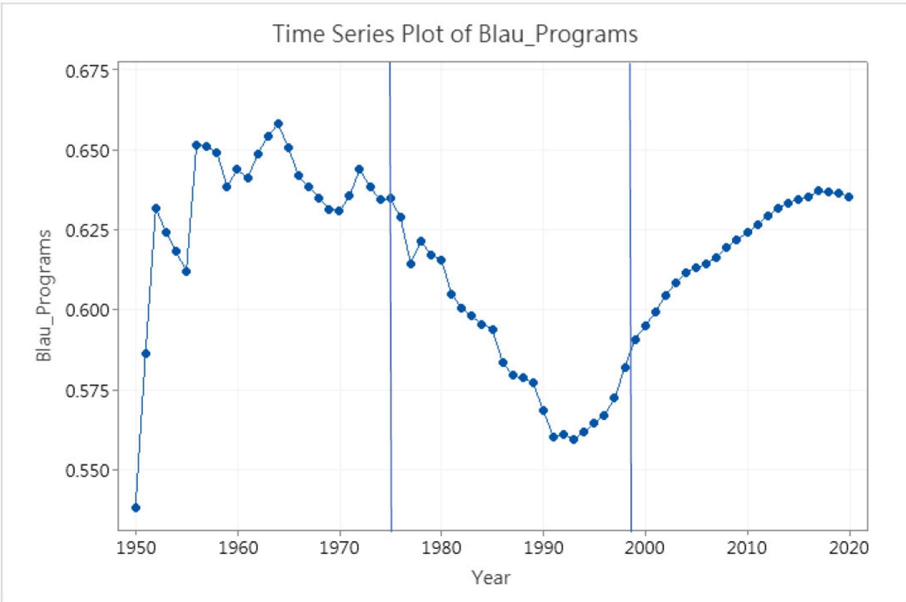

Figure 2. The heterogeneity of program participation

Table 1. Testing of the proposed ARIMA models

Models	AR	MA	Ljung-Box Chi-square statistic
ARIMA(1,1,1)	X	X	OK
ARIMA(1,1,0)	OK	OK	X
ARIMA(2,1,1)	OK	OK	OK

The coefficients of AR(1) $= 1.300$ ($p = 0.000$), AR(2) $= -0.306$ ($p = 0.008$) and MA(1) $= 1.010$ ($p = 0.000$) are significant at 0.05 level, see Table 2 (left). In Box-Pierce (Ljung-Box) Chi-square statistic test, we check the predicted series values with the number of lags 12, 24, 36, and 48 which are classified as white noise ($p > 0.05$), see Table 2 (right). The related residual plots display in Figure 3 and Figure 4. The proposed ARIMA(2,1,1) model is robust. Based on the prediction trend, the result suggests that the Blau index will increase steadily in future. Table 3 reveals that the forecasted Blau index with program participation will increase from 0.6358 in 2021 to 0.6424 in 2030. The time series plot of Blau index with program participation is displayed in Figure 5, which implies the series of Blau index will increase a little in next ten periods. The diversity of program participation will increase in future.

Table 2. Estimated parameters and Box-Pierce Chi-square statistic

ARIMA(2,1,1) estimates of parameters					ARIMA(2,1,1) Box-Pierce Chi-square statistic				
Type	Coef	SE Coeff	t-value	p-value	Lag	12	24	36	48
AR(1)	1.300	0.114	11.42	0.000	Chi-square	16.64	20.96	31.37	35.01
AR(2)	−0.306	0.113	−2.71	0.008	DF	9	21	33	45
MA(1)	1.010	0.000	10202.28	0.000	p-value	0.055	0.461	0.548	0.858

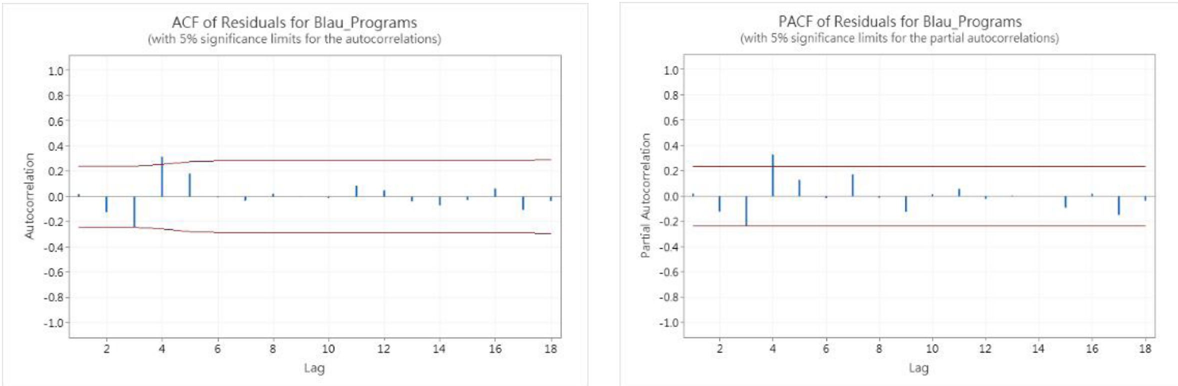

Figure 3. The plots of ACF and PACF

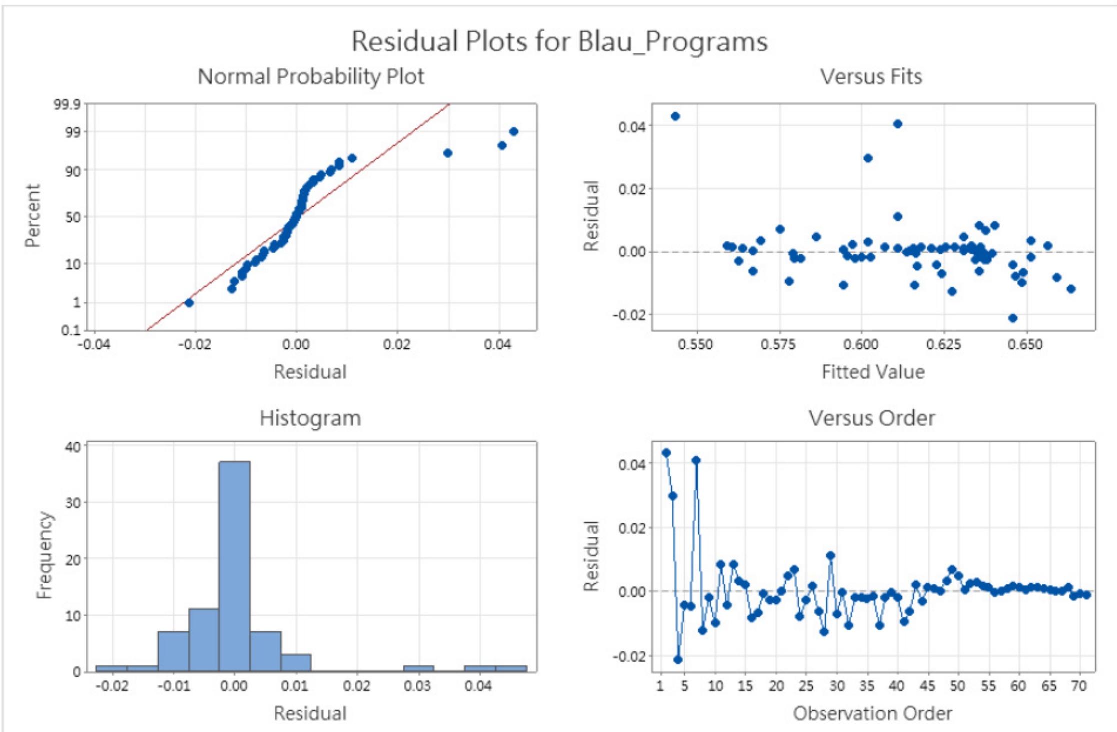

Figure 4. The residual plot of the Blau index reflecting the program participation

TABLE 3. Forecasts from period 72-81 for Blau index

Period (from 2021 to 2030)	Forecast	95% limits	
		Lower	Upper
72	0.635838	0.616583	0.655093
73	0.636461	0.605039	0.667884
74	0.637191	0.596283	0.678099
75	0.637948	0.589337	0.686559
76	0.638710	0.583617	0.693804
77	0.639469	0.578770	0.700169
78	0.640223	0.574579	0.705867
79	0.640970	0.570903	0.711038
80	0.641712	0.567645	0.715779
81	0.642448	0.564735	0.720160

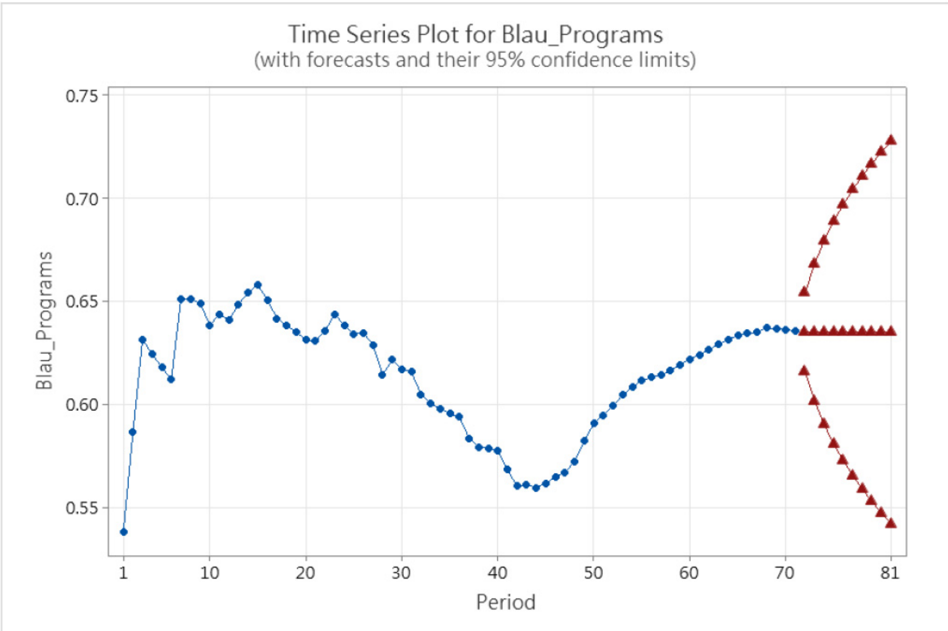

FIGURE 5. Predicating Blau index of program participation with ARIMA(2,1,1)

Figure 6 demonstrates the heterogeneity in terms of the Blau indices is relative high in both elite and universal stage than that of mass stage. It implies the higher education expansion steadily can diminish the diversity of program participation, while over expansion could be raised an issue of unbalancing. Since the GER has reached to 85% in 2014, the result reveals that the higher education did not reduce the diversity of program participation in the oversupply process.

4. **Conclusions.** This study demonstrates the Blau index can be used to tackle the issues of heterogeneity of program participation in higher education settings. The design of research can be extended to similar higher education systems. The three research questions have been answered. First, this study found the Blau index is sensitive to reflect the change of program participation, when the system is in low participation and high participation stage the diversity will increase. Second, the Blau index of program participation has shown decrease in the mass stage. A moderate higher education expanding may shape a health program participation pattern. Third, what is the trend of the heterogeneity index in next decade? In this study, we found that the 75% of GER may

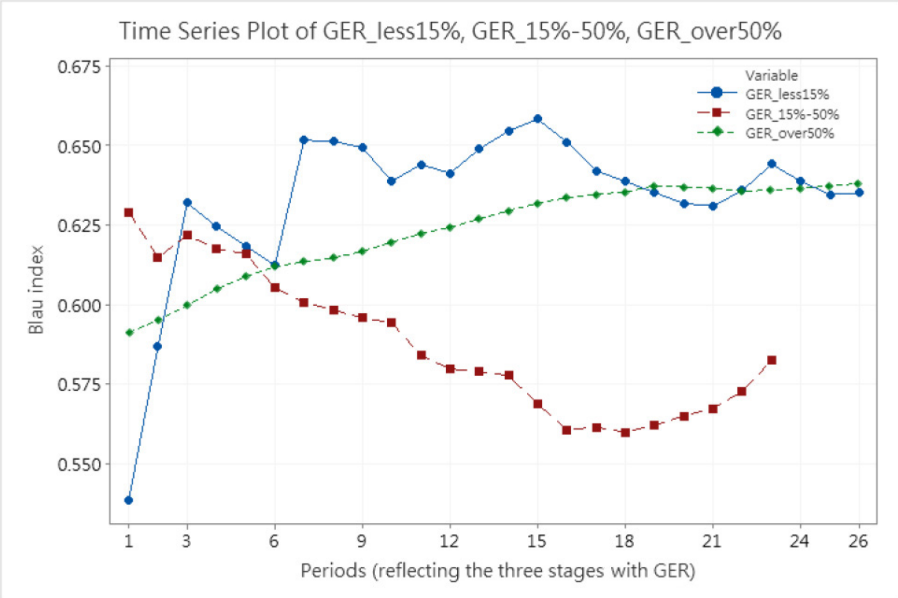

Figure 6. Comparing the heterogeneity of the three stages with GER

reflect the oversupply issue in the expanding system. Based on ARIMA model, the findings suggest that oversupply of higher education will widen the heterogeneity of program participation.

This study can detect that the increasing diversity might become an emerging issue in higher education. The design of study can provide a system-wide monitoring mechanism for related policy makers. Considering many countries that have moved to over expanded higher education systems, the application of Blau index for heterogeneity can be extended to detect the phenomenon. For further study, this study suggests that the Blau index is not only used to tackle the diversity issue, but considered the relationship between Blau index and related external factors in higher education, for example, the heterogeneity of program participation with economic growth or GDP per capita increasing. This direction may enrich the knowledge in this field.

REFERENCES

[1] UNESCO Institute for Statistics, *UIS Statistics-Gross Enrolment Ratio by Level of Education*, 2018, http://data.uis.unesco.org/, Accessed on May 20, 2022.

[2] S. Ilie and P. Rose, Is equal access to higher education in South Asia and sub-Saharan Africa achievable by 2030?, *Higher Education*, vol.72, no.4, pp.435-455, 2016.

[3] F. M. Msigwa, Widening participation in higher education: A social justice analysis of student loans in Tanzania, *Higher Education*, vol.72, no.4, pp.541-556, 2016.

[4] R. Schendel and T. McCowan, Expanding higher education systems in low- and middle-income countries: The challenges of equity and quality, *Higher Education*, vol.72, no.4, pp.407-411, 2016.

[5] D.-F. Chang and H.-C. ChangTzeng, Patterns of gender parity in the humanities and STEM programs: The trajectory under the expanded higher education system, *Studies in Higher Education*, vol.45, no.6, pp.1108-1120, 2020.

[6] D. W. Chapman and C.-L. Chien, Dilemmas of expansion: The growth of graduate education in Malaysia and Thailand, *Higher Education Studies*, vol.5, no.3, pp.1-10, 2015.

[7] E. Wright and H. Horta, Higher education participation in "high-income" universal higher education systems: "Survivalism" in the risk society, *Asian Education and Development Studies*, vol.7, no.2, pp.184-204, 2018.

[8] L. Yang, Higher education expansion and post-college unemployment: Understanding the roles of fields of study in China, *International Journal of Educational Development*, vol.62, pp.62-74, 2018.

[9] S. Marginson, High participation systems of higher education, *The Journal of Higher Education*, vol.87, no.2, pp.243-271, 2016.

[10] S. Marginson, The worldwide trend to high participation higher education: Dynamics of social stratification in inclusive systems, *Higher Education*, vol.72, no.4, pp.413-434, 2016.
[11] D.-F. Chang, Effects of higher education expansion on gender parity: A 65-year trajectory in Taiwan, *Higher Education*, vol.76, no.3, pp.449-466, 2018.
[12] Ministry of Education, *Educational Statistics 2015*, Taiwan, 2015, https://stats.moe.gov.tw/files/ebook/Education_Statistics/104/104edu_EXCEL.htm, Accessed on April 20, 2022.
[13] M. Trow, Trust, markets and accountability in higher education: A comparative perspective, *Higher Education Policy*, vol.9, no.4, pp.309-324, 1996.
[14] D.-F. Chang and A. Chang, Detecting the impacts of newborn babies and its effect on elementary school enrollment, *ICIC Express Letters, Part B: Applications*, vol.12, no.6, pp.499-506, 2021.
[15] C.-C. Chen, Y. Chen and C.-Y. Kang, Estimating student-teacher ratio with ARIMA for primary education in fluctuating enrollment, *ICIC Express Letters, Part B: Applications*, vol.12, no.6, pp.515-523, 2021.
[16] H. Alabdulrazzaq, M. N. Alenezi, Y. Rawajfih, B. A. Alghannam, A. A. Al-Hassan and F. S. Al-Anzi, On the accuracy of ARIMA based prediction of COVID-19 spread, *Results in Physics*, vol.27, 104509, 2021.
[17] C. Yuan, S. Liu and Z. Fang, Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1,1) model, *Energy*, vol.100, pp.384-390, 2016.
[18] B. Nath, D. Dhakre and D. Bhattacharya, Forecasting wheat production in India: An ARIMA modelling approach, *Journal of Pharmacognosy and Phytochemistry*, vol.8, no.1, pp.2158-2165, 2019.
[19] M. B. Chamlin and B. A Sanders, Social policy and crash fatalities: A multivariate time series analysis, *Journal of Crime and Justice*, vol.41, no.3, pp.322-333, 2018.
[20] N. Achille, S. Haberman and G. Consigli, *A Multivariate Approach to Project Common Trends in Mortality Indices*, 2018, http://dx.doi.org/10.2139/ssrn.3149989, Accessed on May 22, 2022.
[21] Ministry of Education, *School Basic Data*, https://depart.moe.edu.tw/ED4500/News.aspx?n=5A930C32CC6C3818&sms=91B3AAE8C6388B96, Accessed on May 10, 2022.
[22] M. Trow, *Problems in the Transition from Elite to Mass Higher Education*, ERIC (ED 091 983), 1973, http://files.eric.ed.gov/fulltext/ED091983.pdf, Accessed on May 11, 2022.
[23] K. H. Kim and A. A. Rasheed, Board heterogeneity, corporate diversification and firm performance, *Journal of Management Research*, vol.14, no.2, pp.121-139, 2014.
[24] E. Shehadeh, D. Aly and I. Yousef, The impact of boardroom internationalization on online disclosures of S&P 500, *Journal of Financial Reporting and Accounting*, vol.19, no.4, pp.596-614, 2021.
[25] A. Solanas, R. M. Selvam, J. Navarro and D. Leiva, Some common indices of group diversity: Upper boundaries, *Psychological Reports*, vol.111, no.3, pp.777-796, 2012.
[26] W.-C. Chou and D.-F. Chang, Gender diversity indices for interpreting the phenomena in expanding higher education, *Journal of Applied Research in Higher Education*, vol.14, no.1, pp.92-110, 2022.
[27] P. M. Blau, *Inequality and Heterogeneity: A Primitive Theory of Social Structure*, Free Press, New York, NY, 1977.
[28] A. L. Humbert and E. A. Günther, D3.1 The gender diversity index, preliminary considerations and results, *GEDII-Gender Diversity Impact – Improving Research and Innovation through Gender Diversity*, European Commission, 2017.
[29] S. Bisgard and M. Kulachi, *Time Series Analysis and Forecasting by Example*, Wiley & Sons, New York, NY, 2011.
[30] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, Springer, New York, NY, 2016.
[31] D.-F. Chang and K. L. Lai, Trajectory of the population dependency index by using ARIMA models, *ICIC Express Letters, Part B: Applications*, vol.10, no.3, pp.195-202, 2019.
[32] H. Qi et al., COVID-19 transmission in Mainland China is associated with temperature and humidity: A time-series analysis, *Science of the Total Environment*, vol.728, 2020.
[33] F. J. Elgar, A. Stefaniak and M. J. A. Wohl, The trouble with trust: Time-series analysis of social capital, income inequality, and COVID-19 deaths in 84 countries, *Social Science and Medicine*, vol.263, DOI: 10.1016/j.socscimed.2020.113365, 2020.
[34] D. Shan, W. Lu and J. Yang, The data-driven fuzzy cognitive map model and its application to prediction of time series, *International Journal of Innovative Computing, Information and Control*, vol.14, no.5, pp.1583-1602, 2018.
[35] G. E. P. Box and G. M. Jenkins, *Times Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, CA, 1976.
[36] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, OTexts, Melbourne, Australia, 2013.

[37] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, Springer, New York, NY, 2016.
[38] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, 5th Edition, Wiley, New York, NY, 2015.
[39] G. M. Ljung and G. E. P. Box, On a measure of lack of fit in time series models, *Biometrika*, vol.65, no.2, pp.297-303, 1978.