ICIC Express Letters
Part B: Applications
Volume **11**, Number **3**, March **2020**

ICIC International ©2020 ISSN 2185-2766
pp. **221–227**

DETERMINING SYSTEM EXPANSION ON FACULTY GENDER PARITY BY USING ARIMA

Dian-Fu Chang[1] and Hui Hu[2,*]

[1]Graduate Institute of Educational Policy and Leadership
[2]Doctoral Program of Educational Leadership and Technology Management
Tamkang University
No. 151, Yingzhuan Road, Tamsui District, New Taipei City 25137, Taiwan
140626@mail.tku.edu.tw; *Corresponding author: 806760095@s06.tku.edu.tw

Received September 2019; accepted December 2019

Abstract. *Gender parity issues have been discussed widely in previous literature. While trend analysis for this phenomenon is still very limited. This study focuses on the effect of system expansion on faculty gender parity by using transfer function and ARIMA to tackle the series data. Gender parity index was transformed by Becker's discrimination coefficient (D) to detect the patterns of gender parity in the selected higher education system. To predict the future trend, the ARIMA (autoregressive integrated moving average) was used to build the fittest model to interpret the trends of D in next decade. The findings display the higher education system still favored male faculty, whereas following the post-expansion in next decade, the gender parity will reduce its discrepancy. This study provides an example to tackle the similar issue in different higher education settings.*
Keywords: ARIMA, Faculty members, Gender parity, Higher education, Time series

1. **Introduction.** According to the 2017 data of the United Nations Educational, Scientific and Cultural Organization (UNESCO), the gross entrance ratios (GERs) are 86% in Finland, 91% in Spain, and 93.78% in Korea [1]. The GER is usually as an indicator to reflect the development of a higher education. This indicator also relates to enrollment or participation in higher education. Typically, the higher education systems followed rational process to extend their participation and might cause structure change within the systems. Previous studies have shown that higher education expansion will transform the student gender parity. For example, when the higher education system goes into the later stage of gross entrance ratio (GER) 15%-50% or over 50%, the number of female students has gradually become greater than that of male students [2,3]. Furthermore, women are persistently under-represented in technology and engineering, but over-represented in other STEM (science, technology, engineering and mathematics) fields [4]. This phenomenon is also displayed in other countries [5,6]. This trend illustrates when student number increases in higher education, the system is going to favor female students. Most of previous studies have focused on the gender parity in student groups; however, the gender pattern of faculty groups is still unclear. This phenomenon might be interpreted by the current studies that only focused on the specific gender issues in higher education settings [7-10]. In this study, we tried to realize whether the system expansion has ensured faculty increasing or changed the patterns of faculty gender? We found this issue with long-term trend has been discussed very limited in current research literature. The finding will provide meaningful information for related policy makers.

Moreover, series data have been mining for different fields and provides a lot of forecasted experiences. For example, autoregressive integrated moving average (ARIMA) models have been applied for various fields in higher education settings, while the system itself

DOI: 10.24507/icicelb.11.03.221

with the issues still did not be addressed properly by using the predicted model. This study aims to realize the faculty gender pattern change by using ARIMA models to tackle the series data. We tried to transform the gender parity index for faculty in the target higher education system. The data transformation and ARIMA model could be a new approach in this field. Given this purpose, this study selected Taiwan as a research target to tackle the faculty gender parity issue in the system. Specifically, this study tries to answer the following research questions:

a. What trends are shown in gender of faculty with the system expansion?

b. Has there been reduction in faculty gender discrepancy in the expanded higher education system?

c. What trend regarding faculty gender parity is likely in the next decade?

The result will provide two patterns of faculty gender in previous and future trends in the selected higher education system. The structure of this paper is as follows. First, the method section addresses the characteristics of series data and the ARIMA models. Then, the results will be displayed and made further discussions. Finally, the conclusion and implication will be drawn.

2. **Method.** In this section, we define the crucial terms in this study, determine the cross correlation of both series, and build the fittest ARIMA model with Becker's discrimination coefficient (D) to predict the faculty gender parity in next decade.

2.1. **Definition of terms.** System expansion refers to open enrollment policy to widen the participation in higher education system. According to data from the UNESCO Institute for Statistics (2018), the GER in high income countries moved to a universal stage in 1993. Some countries reached 75% in 2011, while the GER in most of the middle-income countries moved to the mass stage in 2001 [4]. In our research target, the GER in this system increased from 15% (1976) to 50% (1999) during these 23 years. While the GER up to 85% happened in 2007, the system only spent another 8 years to reach the ceiling of GER [2]. Faculty gender parity refers to the parity working opportunity for male and female faculty in higher education institutions. In this study, the parity index will be calculated by using Becker's D based on annual data from Ministry of Education, Taiwan.

2.2. **Testing cross correlation with enrollment and faculty members.** This study considered a cross correlation to determine the relationship between two series variables contemporaneously and at various lagged values. Cross correlations help us understand if two variables are related to each other and, if so, whether movement in one variable tends to precede or follow movement in the other. Typically, the cross correlation function (CCF) is applying for identifying lags of the x-variable that might be useful predictors of y_t [5,6]. Pre-whitening solves this problem by removing the autocorrelation and trends with related series data SPSS can help to get CCF with its cross correlation command in forecasting. The CCF is defined as the set of correlations between x_{t+h} and y_t for $h = 0$, ± 1, ± 2, ± 3, and so on. A negative value for h is a correlation between the x-variable at a time before t and the y-variable at time t. For instance, consider $h = -2$. The CCF value would give the correlation between x_{t-2} and y_t.

- When one or more x_{t+h} with h *negative*, are predictors of y_t, it is sometimes said that x leads y.
- When one or more x_{t+h}, with h *positive*, are predictors of y_t, it is sometimes said that x lags y.

2.3. **Data transformation.** This study selected the data from the Ministry of Education as an example to realize the phenomenon in higher education settings [7,8]. In this specific field, the data were collected from 1968 to 2018 based on annual basis. Gender parity in faculty members was defined as the Becker's discrimination coefficient. The D is defined as follows [9,10]:

$$\mathrm{D}_i = (\mathrm{FM}_i/\mathrm{FF}_i) - 1$$

where, FM represents the male faculty, FF represents the female faculty in the higher education, and i represents the series data collection period from 1968 to 2017. The interpretation of D will follow the following rules:

- An increase in the calculated D implies that the system favors male faculty;
- The calculated D being negative means that system favors female faculty;
- A calculated D of zero or near zero represents equal opportunity for female and male faculty in the education system.

2.4. **Building ARIMA model.** The estimated patterns of faculty gender parity were conducted using ARIMA (autoregressive integrated moving average) model for further interpretation. Generally, in ARIMA model, the basic series data needed 50 periods or more. The selected data set with 51 periods fit the criterion. The fittest model was selected to interpret the data and the robustness of the forecasting model was evaluated for estimation from 2019-2028. This study followed the ARIMA(p, d, q) model building process as follows: 1) prepare target data to detect the data set is stationary or nonstationary series; 2) identify potential forecast models; 3) check the ACF (autocorrelation function)/PACF (partial autocorrelation function) of the residuals; and 4) forecast the series in next decade [11-14]. A non-seasonal ARIMA(p, d, q) model, where: p is the number of autoregressive terms, d is the number of non-seasonal differences needed for stationarity, and q is the number of lagged forecast errors in the prediction equation. Differenced series data provides more stationary types for forecasting, typically, two times difference is enough [15]. In this study, inspection of the ACF of a time series was used to determine whether a series is stationary or will require differencing [13,14]. The ACF and PACF should be considered together. For example, AR models have theoretical PACFs with non-zero values at the AR terms in the model and zero values elsewhere. The estimation of parameter for D will be presented to check the coefficient is significant. In this study, Box-Pierce Chi-square statistics were used to determine whether the model met the assumptions that the residuals were independent [16]. The calculations were listed as follows [16,17]:

$$Q^*(K) = (n-d) \cdot (n-d+2) \cdot \sum_{l=1}^{K} (n-d-l) \cdot r_l^2\left(\hat{a}\right)$$

where, n is the sample size, d is the degree of non-seasonal differencing used to transform the original series to a stationary one, $r_l^2\left(\hat{a}\right)$ is the sample autocorrelation at lag l for the residuals of the estimated model, K is the number of lags covering multiples of seasonal cycles, e.g., 12, 24, 36, $\ldots$, for yearly data.

In this process, we check the criteria for a white noise test on time series are described as follows: If the number is classified as white noise, then

a. the series averages are fixed;

b. the series variance is fixed; and

c. the series self-covariance is 0; that is, in this period, the last issue of the series is not related.

3. **Results.** Following the cross correlation function, ARIMA model selection, and forecasting, we present the results with related tables and figures.

3.1. **Cross correlation function between enrollment and faculty members.** Table 1 displays the lag 7 to lag −7 with higher cross correlation; it implies the enrollment and faculty members with relationship contemporaneously and at various lagged values. Faculty members increasing is related to the enrollment increasing. Figure 1 demonstrates the CCFs are significant. The effect of higher education expansion will impact on faculty members increasing. While how to determine the effect of expansion on faculty gender parity that might be transformed? It needs detecting the male and female faculty data to further interpretation. The next section will address the details of series D in ARIMA model.

TABLE 1. Cross correlation function with enrollment and faculty members

Lag	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7
Cross correlation	0.357	0.341	0.347	0.386	0.439	0.515	0.603	0.673	0.647	0.659	0.609	0.592	0.543	0.490	0.420
Standard error	0.152	0.151	0.149	0.147	0.146	0.144	0.143	0.141	0.143	0.144	0.146	0.147	0.149	0.151	0.152

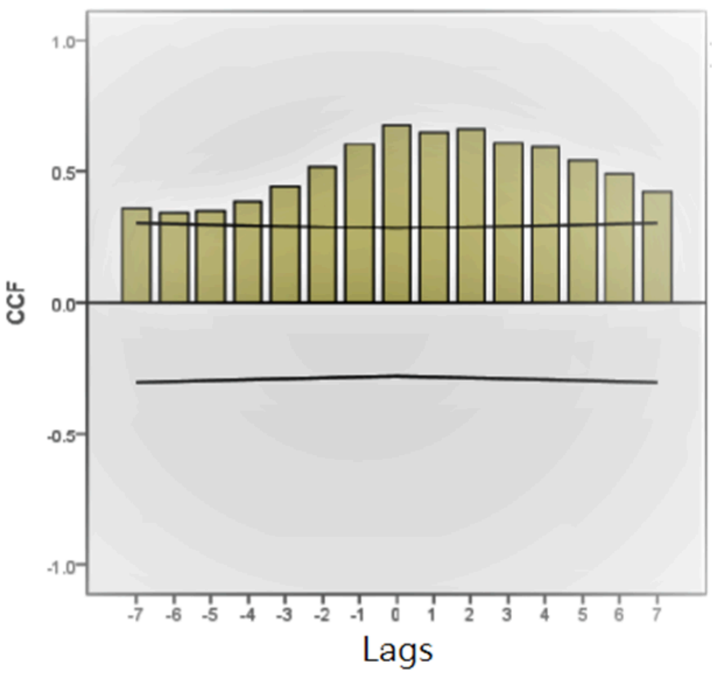


FIGURE 1. Cross correlation function between enrollment and faculty members from 1968 to 2018

TABLE 2. Selecting ARIMA models for predicting D

Predict D	AR	MA	Ljung-Box	Selection
ARIMA$(1, 2, 1)$	$\star$	$\bullet$	$\bullet$	
ARIMA$(1, 1, 1)$	$\bullet$	$\bullet$	$\bullet$	$\circledcirc$
ARIMA$(1, 1, 2)$	$\bullet$	$\times$	$\times$	

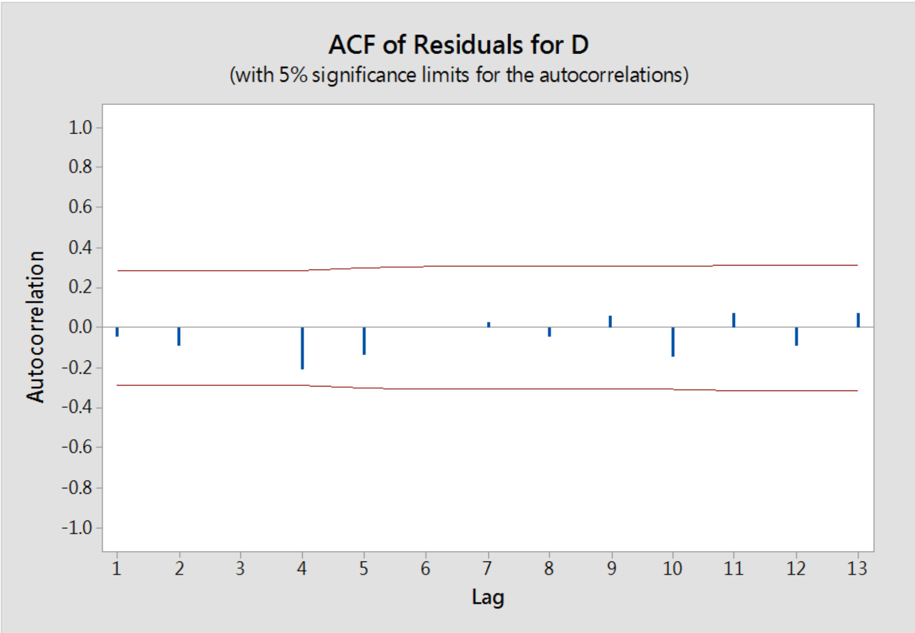
Note. $\bullet$ = excellent, $\star$ = fair, $\times$ = unacceptable.

3.2. **Selecting ARIMA models.** Based on one difference with the D series, the Minitab displays the ARIMA$(1, 1, 1)$ is the fittest model, see Table 2. The ACF, PACF, and modified Box-Pierce (Ljung-Box) Chi-square statistic all work well in ARIMA$(1, 1, 1)$ model, see Table 3 and Figure 2. The coefficients of AR$(1) = 0.95$ ($p = .00$) and MA$(1) = 0.60$ ($p = .00$) are significant at .05 level. In Box-Pierce (Ljung-Box) Chi-square statistic

Table 3. Final estimates of parameters and modified Box-Pierce (Ljung-Box) Chi-square statistic

Type	Coef	SE Coef	t-value	p-value
AR(1)	0.95	0.07	14.22	0.00
MA(1)	0.60	0.15	4.10	0.00
Constant	−0.003	0.003	−0.97	0.34

Lag	12	24	36	48
Chi-square	6.59	10.15	13.90	20.41
DF	9	21	33	45
p-value	0.679	0.977	0.999	0.999



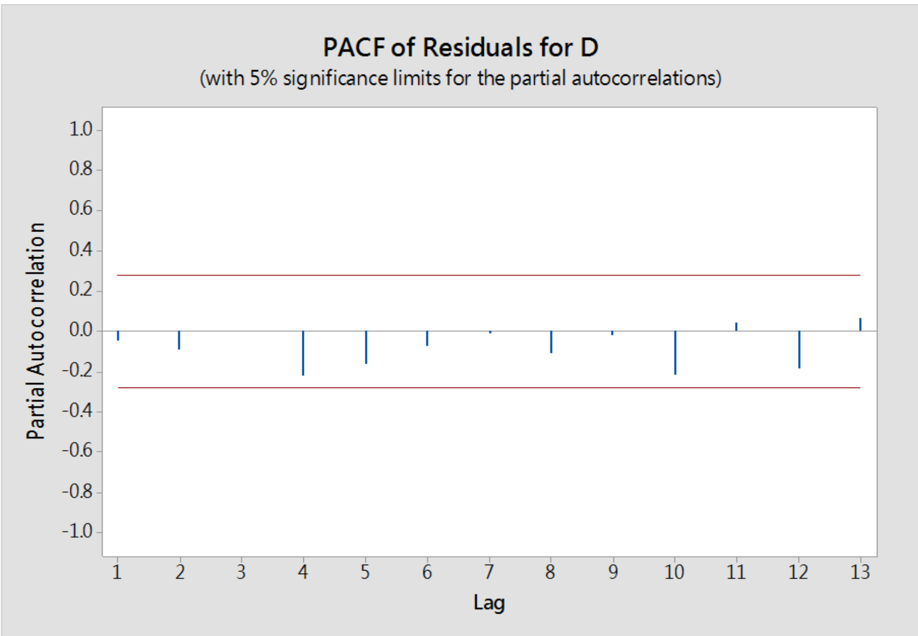


Figure 2. ACF and PACF for ARIMA$(1,1,1)$

test, we check the predicted series values with the number of lags 12, 24, 36, and 48 which are classified as white noise ($p > .05$). The proposed ARIMA$(1,1,1)$ model is robust. This study also applies two differences with the D series, the result reveals ARIMA$(1,2,1)$ did not work well due to its low coefficient of AR(1), see Table 2.

3.3. **Projecting faculty gender parity (D).** This study conducts ARIMA(1, 1, 1) model to predict the D in next decade. The result displays in Figure 3. Based on the prediction trend, the D will decline steadily in future. It implies the expanded higher education system will create a friendly environment for female faculty. Table 4 reveals, the forecast D will decrease from 0.734 in 2019 to 0.432 in 2028. The discrepancy of gender has shown diminishing.

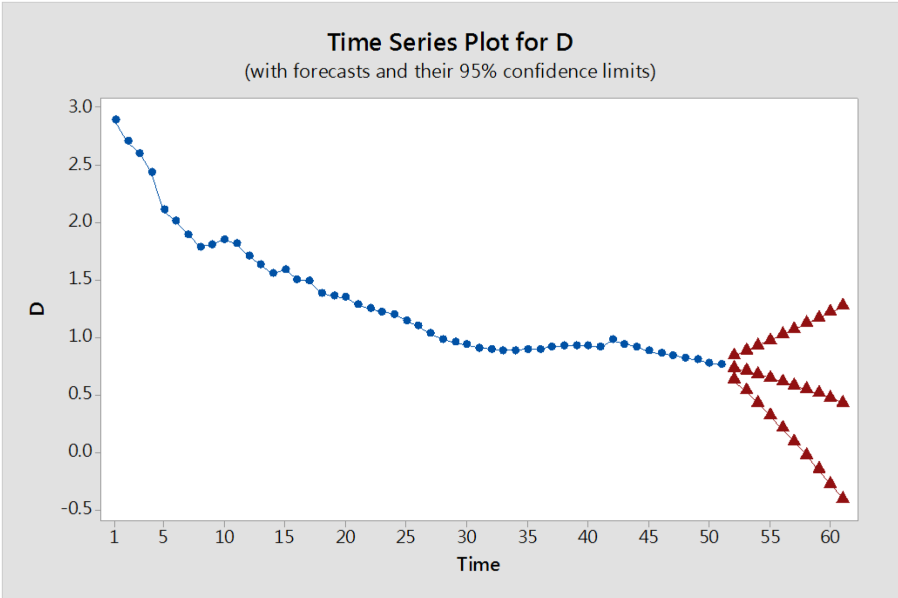


FIGURE 3. Time series plot for D from 1968 to 2028

TABLE 4. Forecasts of D from 2019 to 2028

Year	Forecast	95% Limits	
		Lower	Upper
2019	0.734	0.629	0.838
2020	0.706	0.530	0.882
2021	0.677	0.428	0.927
2022	0.646	0.320	0.972
2023	0.614	0.208	1.019
2024	0.580	0.092	1.068
2025	0.545	−0.028	1.118
2026	0.509	−0.151	1.168
2027	0.471	−0.278	1.219
2028	0.432	−0.406	1.270

4. **Conclusion.** This study conducts cross correlation function to tackle the series relationships in higher education expansion phenomenon. Becker's discrimination coefficient works well in this study to interpret the faculty gender parity transformation. According to the predictions with ARIMA model, in the next decade, the system may still favor male faculty in higher education institutions. However, according to the D, the situation is improving under the system development. Moreover, the trends reveal that gaps in faculty gender parity will decrease in the next decade.

This paper provides an example of an effective approach to addressing gender parity in an expanding higher education system. Although the quantitative approach was limited by the data set and transform model, the explanation of the trends provides a longitudinal perspective for reviewing the effect of expanded higher education on our concern issues. For further work, this paper proposes conducting cross-country comparisons with the researchers in the community to address similar issues.

REFERENCES

[1] UNESCO Institute for Statistics, *UIS Statistics – Gross Enrolment Ratio by Level of Education*, http://data.uis.unesco.org/, 2018.

[2] 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.

[3] S. C. Goy, Y. L. Wong, W. Y. Low et al., Swimming against the tide in STEM education and gender equality: A problem of recruitment or retention in Malaysia, *Studies in Higher Education*, DOI: 10.1080/03075079.2016.1277383, 2017.

[4] 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*, DOI: 10.1080/03075079.2018.1550479, 2018.

[5] S. Danica and D. Cathy, Gender bias in academe: An annotated bibliography of important recent studies, *CUNY Academic Works*, http://academicworks.cuny.edu/qc_pubs/163, 2017.

[6] I. de Melo-Martín, Patenting and the gender gap: Should women be encouraged to patent more?, *Science and Engineering Ethics*, vol.19, no.2, pp.491-504, 2013.

[7] Y.-L. Kim, Gender equity in higher education: Faculty salaries, career development and academic services, *Procedia – Social and Behavioral Sciences*, vol.29, pp.1274-1278, 2011.

[8] S. Bird, Unsettling universities' incongruous, gendered bureaucratic structures: A case-study approach, *Gender, Work and Organization*, vol.18, no.2, pp.202-230, 2011.

[9] P. Karen, Service and gender inequity among faculty, *PS: Political Science and Politics*, vol.44, pp.85-87, 2011.

[10] R. Curt, Why women leave academia and why universities should be worried, *The Guardian*, http://www.guardian.co.uk/higher-educationnetwork/blog/2012/may/24/why-women-leave-academia, 2012.

[11] 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.

[12] G. E. P. Box and G. M. Jenkins, *Times Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, CA, 1976.

[13] PennState, Cross correlation functions and lagged regressions, *Eberly College of Science*, https://newonlinecourses.science.psu.edu/stat510/node/74/, 2018.

[14] Ministry of Education, *Summary of Tertiary Education Institutes (2016-2017)*, http://stats.moe.gov.tw/files/main statistics/u.xls, 2017.

[15] Ministry of Education, *Main Statistical Table – Department of Statistics*, https://depart.moe.edu.tw/ED4500/cp.aspx?n=1B58E0B736635285&s=D04C74553DB60CAD, 2018.

[16] G. S. Becker, *The Economics of Discrimination*, 2nd Edition, The University of Chicago Press, Chicago, 1971.

[17] D.-F. Chang and H. Hu, Mining gender parity patterns in STEM by using ARIMA model, *ICIC Express Letters, Part B: Applications*, vol.10, no.2, pp.105-112, 2019.

[18] G. E. P. Box, G. M. Jenkins and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 4th Edition, Wiley & Sons, New York, 2008.

[19] S. Bisgard and M. Kulachi, *Time Series Analysis and Forecasting by Example*, Wiley & Sons, New York, 2011.

[20] R. Davies, T. Coole and D. Osipyw, The application of time series modelling and Monte Carlo simulation: Forecasting volatile inventory requirements, *Applied Mathematics*, no.5, pp.1152-1168, 2014.

[21] P. Rotela Jr., F. L. R. Salomon and E. de O. Pamplona, ARIMA: An applied time series forecasting model for the Bovespa stock index, *Applied Mathematics*, no.5, pp.3383-3391, 2014.

[22] D.-F. Chang, Y. H. Lin and F.-Y. Nyeu, Detecting the issues of population aging by using ARIMA model, *ICIC Express Letters, Part B: Applications*, vol.10, no.1, pp.39-45, 2019.

[23] G. Ljung and G. Box, On a measure of lack of fit in time series models, *Biometrika*, no.65, pp.297-303, 1978.

[24] R. Nau, *ARIMA Models for Time Series Forecasting*, https://people.duke.edu /~rnau/411arim.htm, 2014.