

Time Series Arima Models of Undergraduate Grade-Point Average

Bruce R. Rogers

University of Northern Iowa

Abstract

The Box-Jenkins approach to time series analysis, a regression method analyzing sequential dependent observations, was used to select the appropriate stochastic model for describing undergraduate grade point averages. The technique, applied to approximately a half century of data from two universities, suggested that the moving average model provided the optimal fit. Suggestions were made for further exploration of GPA data.

Whenever a phenomena is observed over time, it is often useful to search for temporal patterns within the data. Economists have studied stock market prices, sociologists have examined population levels, and psychologists have investigated changes in the incidence of depression. For such purposes, a variety of time series analysis procedures have been developed, derived primarily from the theory of multiple regression. These techniques require data gathered from at least fifty time periods (McCleary and Hay, 1980, p. 20). Since archival data covering this many time periods is not as commonly collected in education as in some other fields, these mathematical approaches are not as widely used in educational research. It is the purpose of this paper to illustrate such an application, using undergraduate grade point averages.

Although educational institutions evaluate their students each term, a single group of pupils is not often evaluated fifty times on the same variable, as would be required for a time series analysis. However, a meaningful time series can be realized by obtaining the average grades given during each of the grading periods across a lengthy time span. For about the last half century, many universities and colleges have adopted a 5-point grading scale, using either the letters A through E or the numbers 1 through 5. Some of the institutions calculated, at each grading period, the average of grades awarded to their students, with the intent of maintaining reasonable consistency in their grading standards both among their departments and across time. Approximately fifteen years ago, reports began appearing that a conspicuous increase was occurring each year in the grading patterns at many institutions (Birnbaum, 1977). Although that pattern appears to have abated during the past few years (Suslow, 1977), grades remain at a noticeably higher level than

to the increase.

A variety of factors have been suggested to explain the phenomena of tutorial grade average fluctuation (Birnbaum, 1977), but there has been a lack of data that support the proposed explanations. Rogers (1983) tested several independent variables (demographic and economic) for the ability of explaining temporal variation over an extended time frame, and found each of them lacking in explanatory power.

Any "explanation" of a phenomena implies that the phenomena can be adequately described. Mathematical models, and regression models in particular, are appropriate for such a description, but an examination of the literature suggests that most authors rely solely on visual graphs rather than employing mathematical modeling. It was the purpose of this study to use a stochastic time series approach to generate mathematical models that might appropriately describe the entire sequence of grade point data.

Method

Sample

Grade point average data were collected from two midwestern universities over about a fifty year span. For the first, hereafter called University A, data was collected for each year from 1929 through 1982. This data is plotted as a time series plot in Figure 1. For the second institution, hereafter called University B, data was collected each year from 1932 to 1982, except for the years 1943 through 1946, when no data was available. This data is plotted in Figure 2.

Procedure

These data were analyzed with the time series analysis procedures brought together in 1970 by George E. P. Box and Gwilyn M. Jenkins, in their volume entitled Time Series Analysis: Forecasting and Control (revised

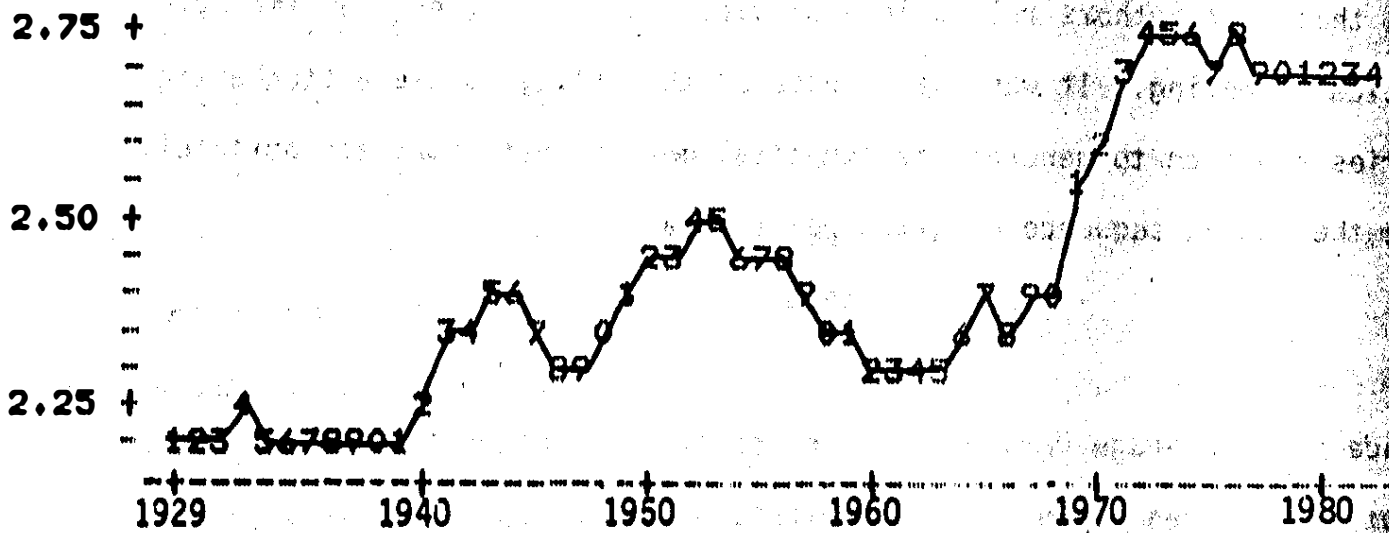


Figure 1. Grade Point Average (GPA) at University A, by year, from 1929 to 1982. (Prior to 1944 the data is for the whole year; afterward it is for fall term.)

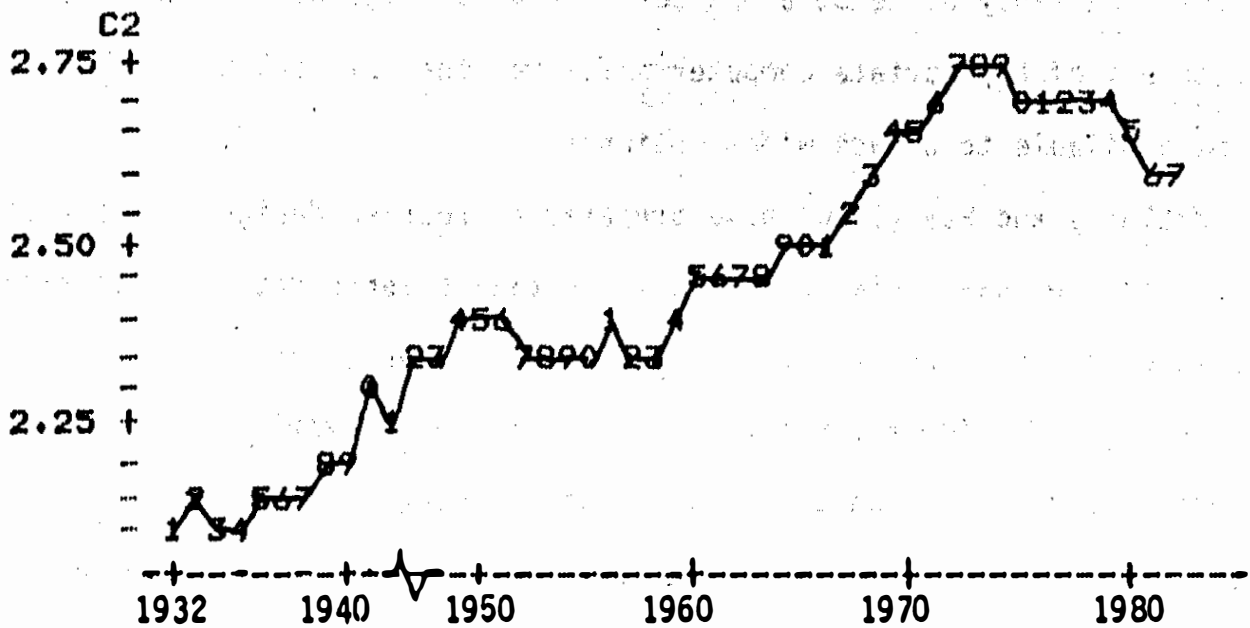


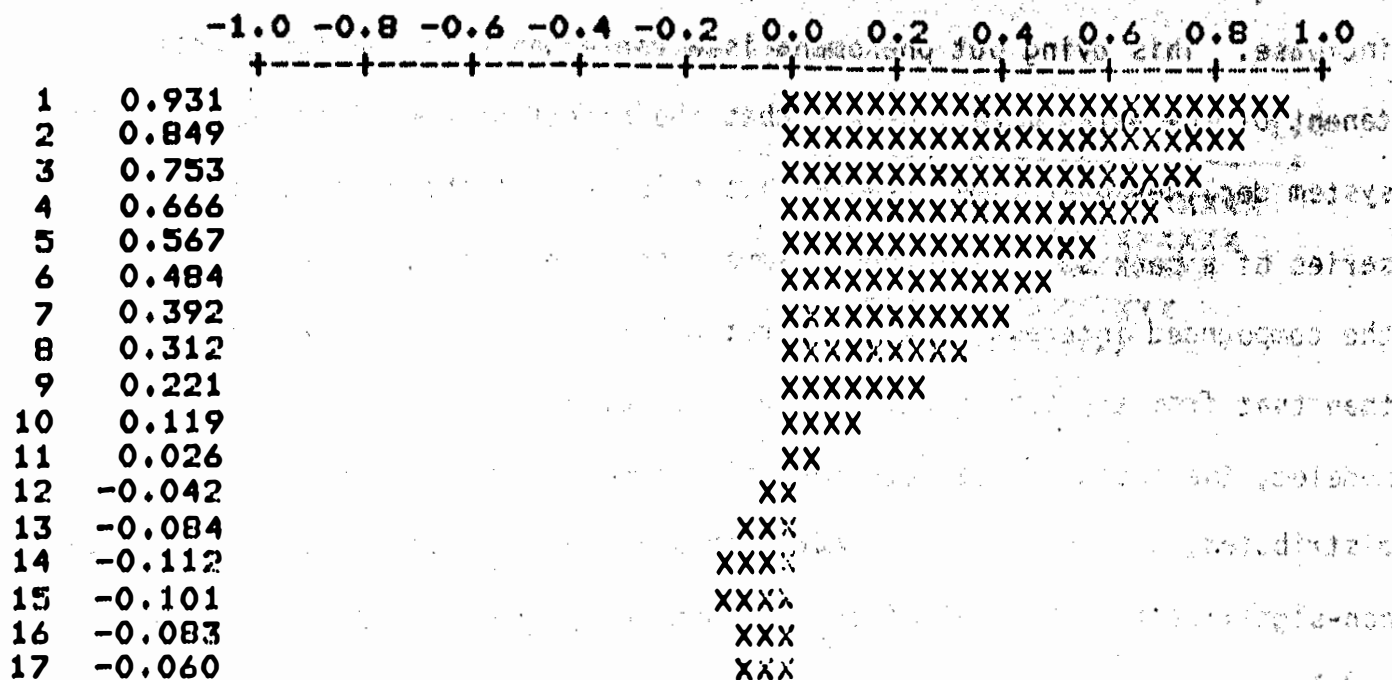
Figure 2. GPA at University 8, by year, from 1932 to 1982 (fall term).
 For 1943-1946, data are not available.

edition 1976). These Auto-Regressive Integrated Moving Average (ARIMA) models (often referred to as "Box-Jenkins" models) require a large amount of data. However, when data are collected over an extended time period, as in this study, there is the possibility that the social meaning of the data could change over time. Thus, it becomes difficult to assign the same interpretation to the data at the beginning and end of the series. Nonetheless, the study of temporal patterns is an intriguing one, and with the development of appropriate computer software, the Box-Jenkins methods have become available to a much wider audience.

McCleary and Hay (1980) have prepared a treatise designed to encourage the use of the Box-Jenkins analysis for social science data, and to explicate strategies for both analyzing the data on the computer and presenting the computer output. Their strategies undergird the analysis in this study. The data was processed on a Harris computer, using MINITAB (Ryan, et al., 1982). Other approaches and other computer programs could have been used, but this was the one available for this project. The reader will need to interpret the methodological procedure of this study in that light.

The empirical identification procedures recommended by Box and Jenkins require an analysis of the autocorrelation function (ACF) and the partial autocorrelation (PACF) of the time series. The graphed ACF and PACF for both of the University time series are shown in Figures 3 and 4. The ACF is a set of correlations, each one of which represents the correlation between the original sequence and itself when lagged k units. For observations close together, e.g., 1 or 2 lags, we most often find a higher correlation than for observations further apart, as is typified in Figures 1 and 2, where the correlations are slowly dying out as the lags

Autocorrelations



Partial Autocorrelations

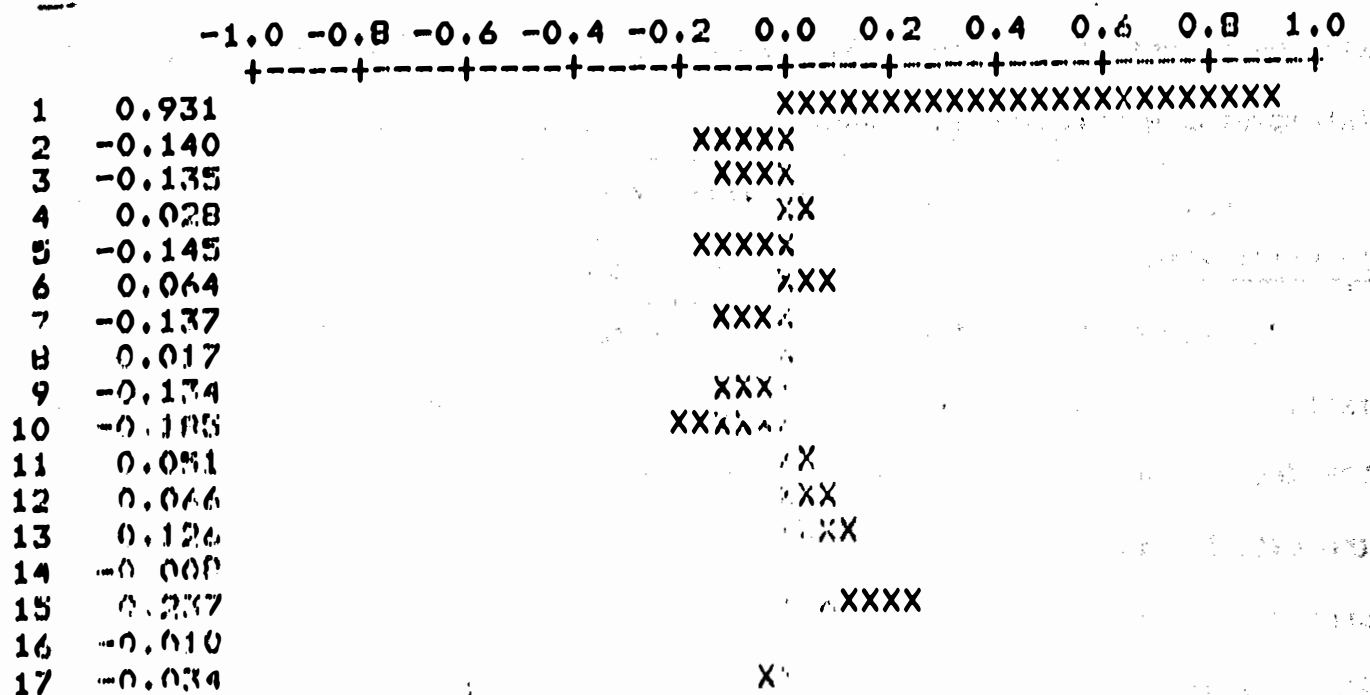


Figure 3. Estimated ACF and PACF for GPA. University A.

increase. This dying out phenomena is a consequence of the fundamental tenent of the ARIMA model, namely that the effect of any given input to the system declines over time. (Note that this is just the opposite of a time series of a bank savings account where, assuming a constant interest rate, the compounded interest from the first dollar invested is always larger than that from any subsequent dollar invested.) When the data is properly modeled, the residuals (errors resulting from the model) should be randomly distributed, and thus yield an ACF with with values that are all statistically non-significant. The goal of the Box-Jenkins approach is to find such a model.

The Box-Jenkins approach is a three stage procedure to build a model, consisting of Identification, Estimation, and Diagnosis. Each of these will be illustrated in the following analysis. The cycle iterates until an interpretable solution is found.

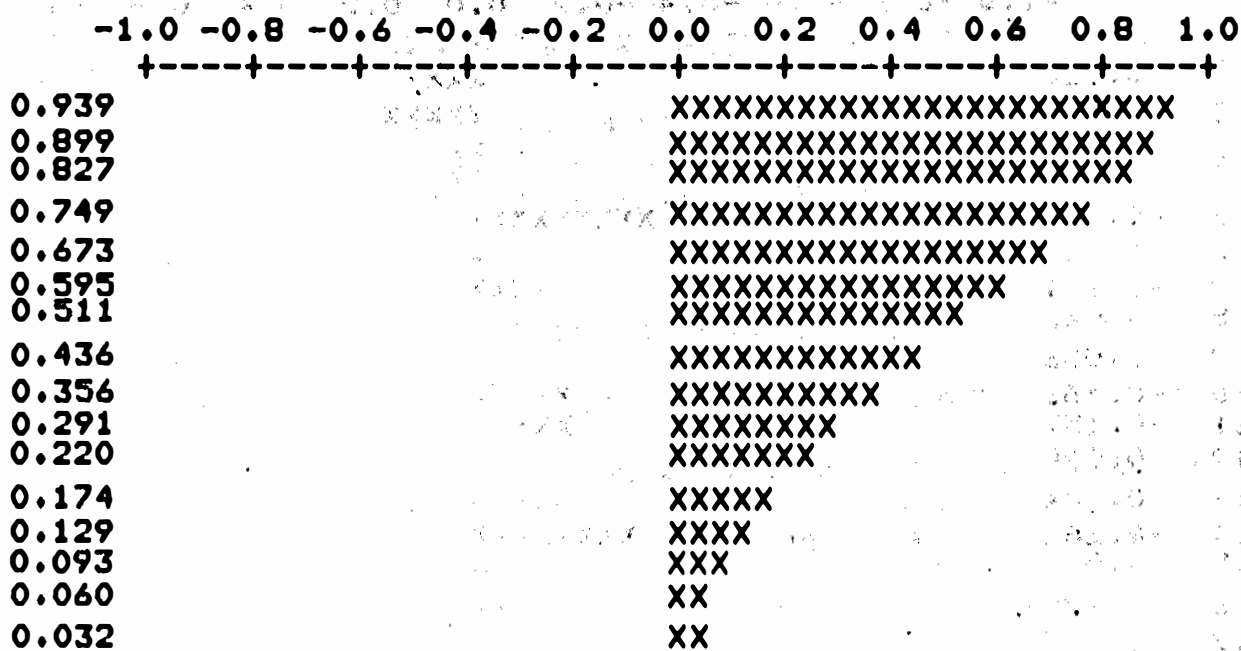
University A

Identification.

An examination of the ACF of the raw data (Figure 3) shows that the ACF falls to zero slowly, indicating that there is a strong systematic trend in the data. The most common method for removing this trend is to transform the data by replacing each observation with the difference between it and the preceding observation. When this differencing transformation is complete, the ACF is again computed. Figure 5 shows the ACF for the differences. The values are much smaller, indicating almost random data. However, there are some spikes, which may be due to sampling error or to some systematic process, so further analysis is required.

The PACF is interpreted similar to the ACF, except that each value is the correlation between observations k units apart after the correlation at

Autocorrelations



Partial Autocorrelations

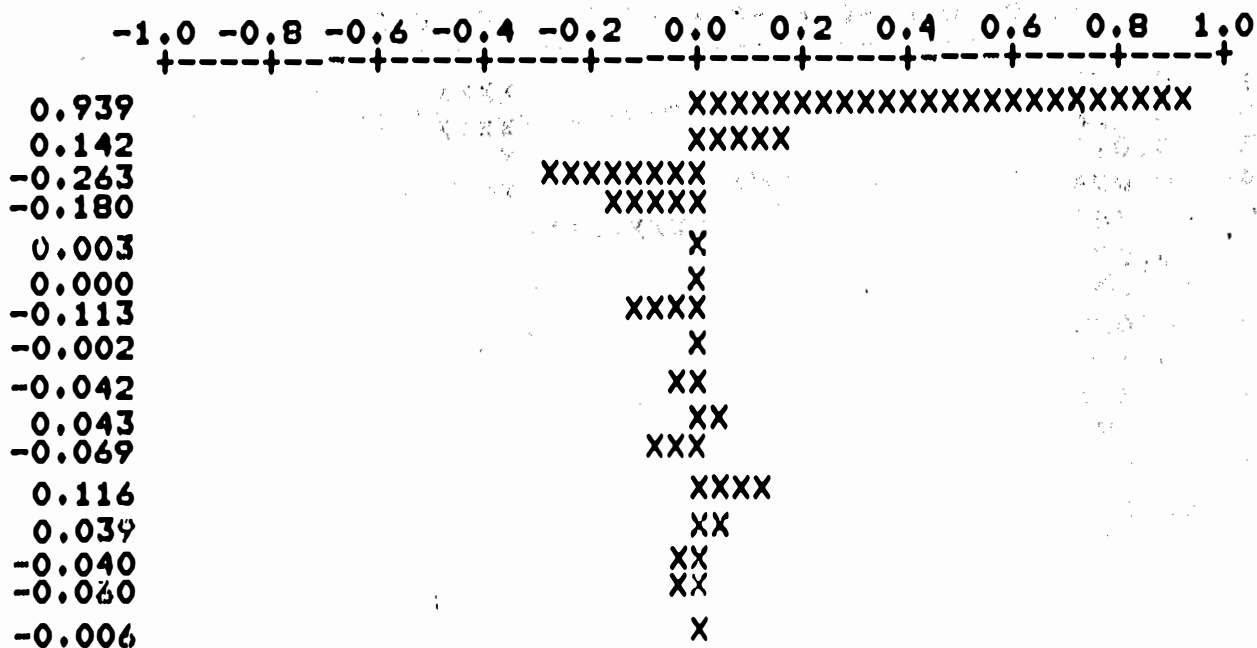
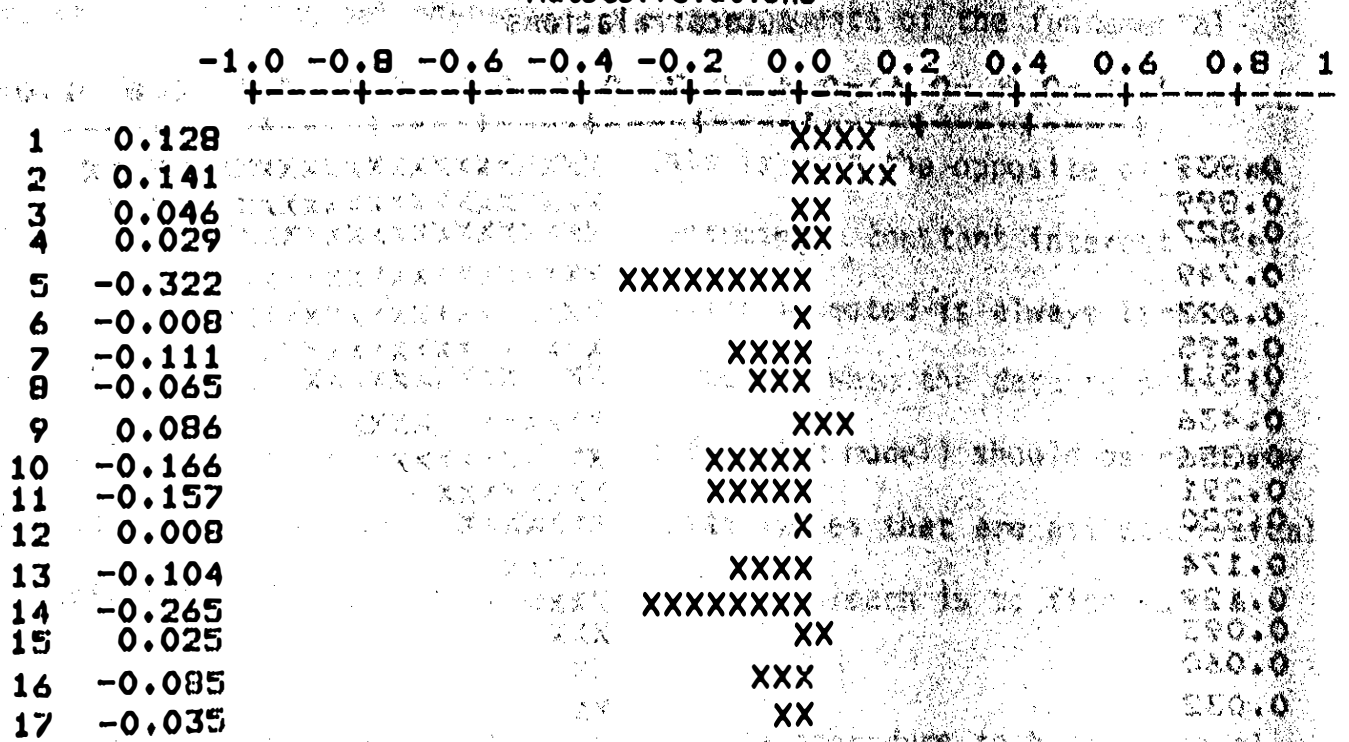


Figure 4. Estimated ACF and PACF for GPA. University B.

Autocorrelations



Partial Autocorrelations

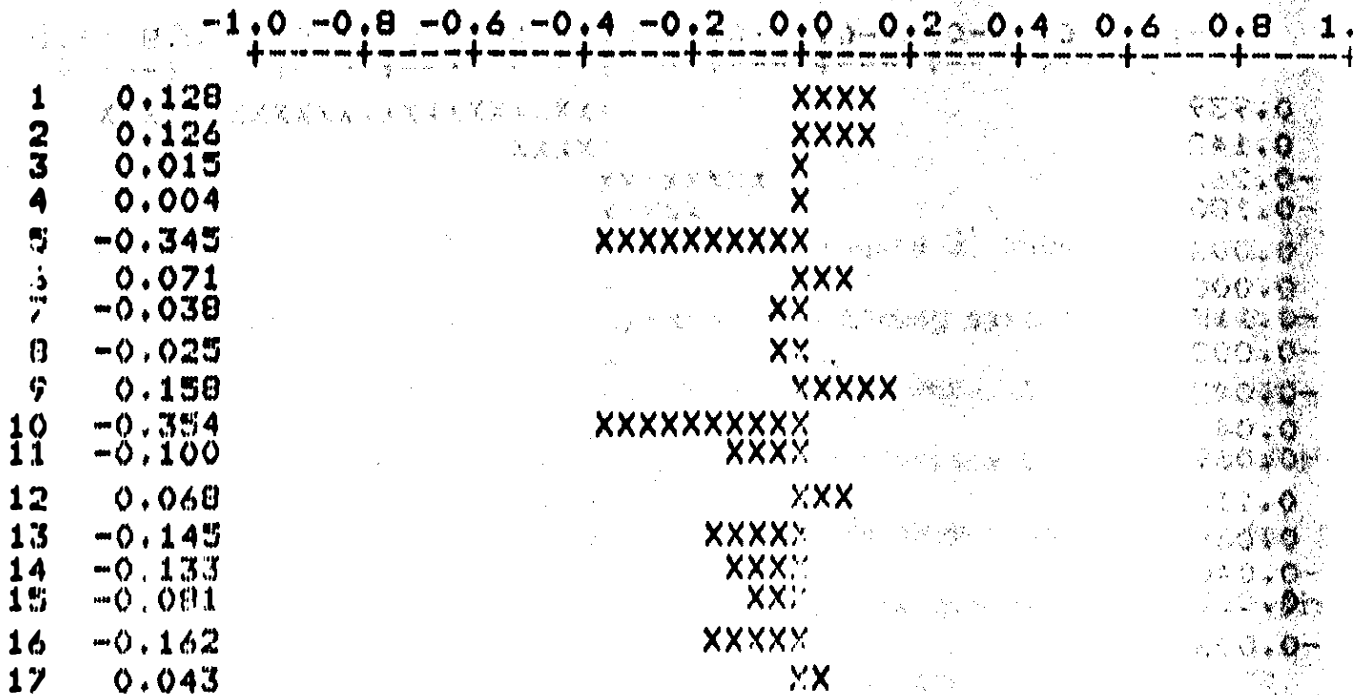


Figure 5. Estimated ACF and PACF for first differences. University A.

intermediate lags has been controlled or "partialled out". The PACF in Figure 3 shows a single spike, which may be the result of what is called a moving average (MA) component. This moving average component can be conceptualized as a random "shock" which is added to each observation to obtain the predicted value for the next observation.

The distinguishing characteristic of a moving average process is the finite duration of the shock. The shock persists for q observations and then is completely suppressed (McCleary and Hay, p. 61). Such a "shock" might be the result of the new grades that are added each term for each particular student. Since the majority of students will leave the institution after four years, the impact of any particular student will vanish when that individual leaves.

From the ACF and PACF we can now tentatively "identify" the model as an ARIMA (0, 1, 1). The zero indicates that there is no auto regressive (AR) term, the middle 1 indicates that differencing is to be used (this is the Integrative (I) term), and the last 1 indicates a moving average (MA) term.

Estimation.

When the estimates of the parameters were computed, it was found that the (0, 1, 1) model produced a t-value of only 1.23 for the MA term. Since this value was not statistically significant at the .05 level (nor anywhere near there), the model was rejected, and the procedure returned to the identification stage.

Identification.

It might be useful at this point to emphasize that since the estimated ACF and PACF are based on very small samples, they are subject to relatively large sampling errors. Consequently, any identification is very tentative.

Because the ACF and PACF for first differences appeared rough, it seemed appropriate to take second differences, i.e., differences between the difference scores. Figure 6 shows the resulting ACF and PACF. They appear more interpretable, suggesting a (0, 2, 1) model. An examination of Figure also suggested that the variance was not constant across time. To attempt to correct this, a logarithmic transformation of the data was performed.

Estimation.

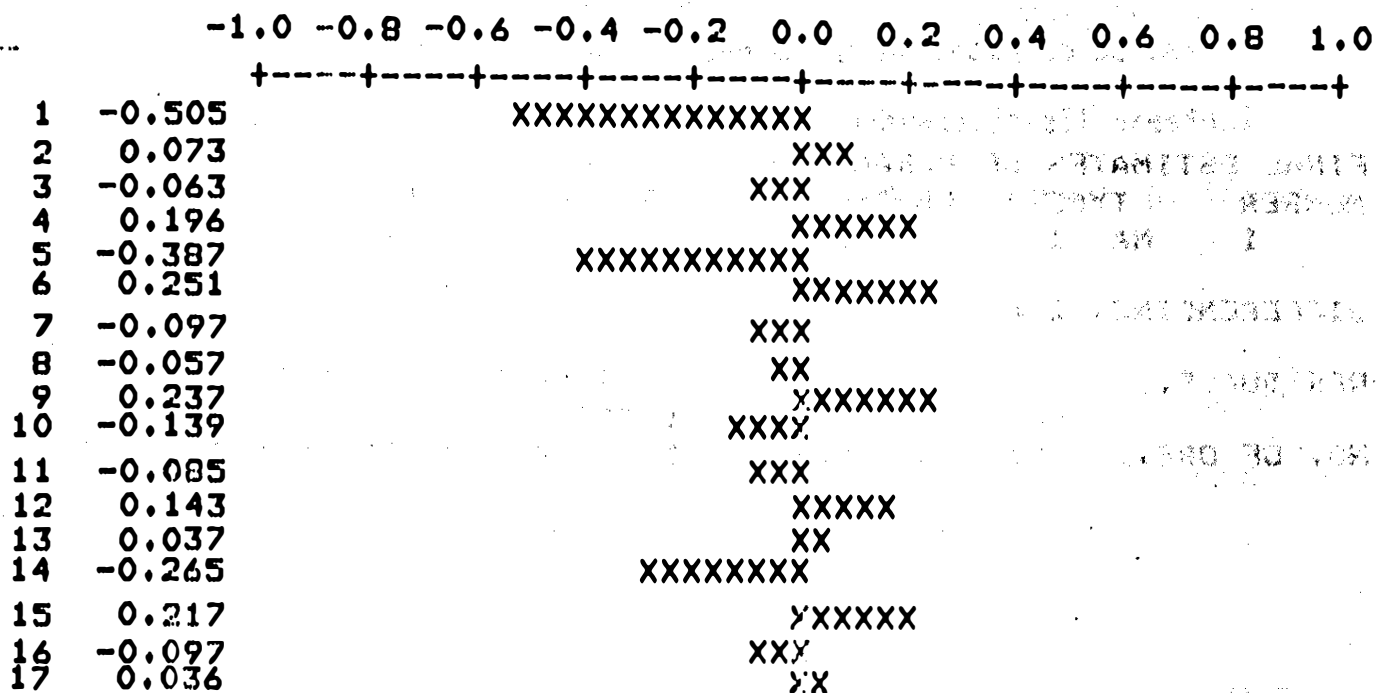
Table 1 shows the results of estimating the (0, 2, 1) model. The moving average parameter of .9767 satisfies the stationarity requirement that its absolute value be less than 1.0, and is also statistically significant at less than the .05 level.

Diagnosis.

The simplest diagnostic procedure is to compare the results of the given model and alternative models. In this way, it can be shown that a particular model is optimal in that neither a simpler nor a more complex model will suffice. The simpler model (0, 1, 1) was already shown to be inadequate. The more complex model (0, 2, 2) yielded a statistically insignificant second MA term, so it was rejected. The (1, 2, 1) model was also tested, but the AR term was insignificant. Thus, the ARIMA (0, 2, 1) model was accepted as the "best" fit.

The equation generated by this procedure can be conveniently written in the following form: $(1-B)^2 y_t = (1-.9767B)a_t$ where B is the backshift operator, and a_t is the random-shock element (McCleary and Hay, (1980), p. 464). The backshift operator is defined as $By_t = y_{t-1}$ and follows the usual algebraic rules. The operator (1-B) represents first differences and $(1-B)^2$ represents second differences.

Autocorrelations



Partial Autocorrelations

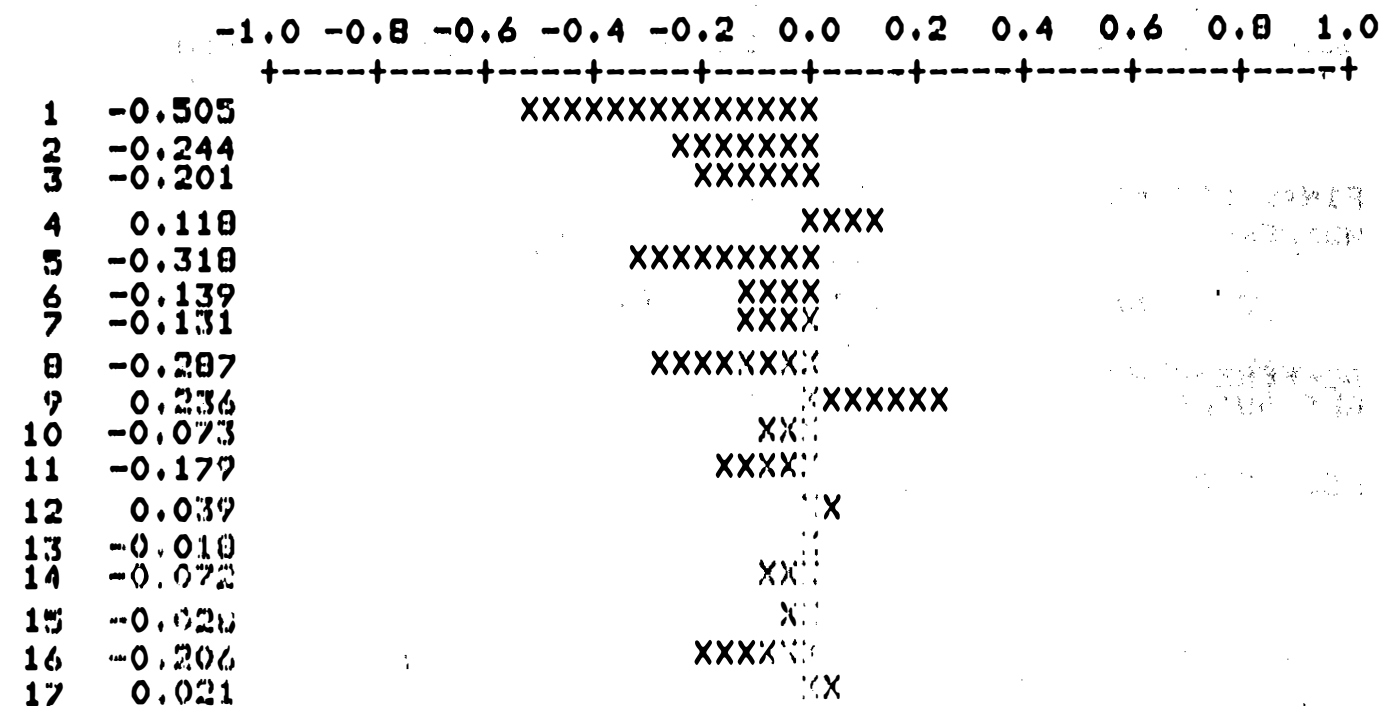


Figure 6. Estimated ACF and PACF for second differences. University A.

Table 1. Parameter estimates for the ARIMA (0, 2, 1) model: University A.

FINAL ESTIMATES OF PARAMETERS						
NUMBER	TYPE	ESTIMATE	ST. DEV.	T-RATIO		
1	MA 1	0.9767	0.0439	22.26		
DIFFERENCING. 2 REGULAR						
RESIDUALS.	SS =	0.0191286	(BACKFORECASTS EXCLUDED)			
	DF =	51	MS =	0.0003751		
NO. OF OBS.	ORIGINAL SERIES	54	AFTER DIFFERENCING	52		

Table 2. Parameter estimates for the ARIMA (0, 2, 2) model. University B.

FINAL ESTIMATES OF PARAMETERS						
NUMBER	TYPE	ESTIMATE	ST. DEV.	T-RATIO		
1	MA 1	1.1475	0.1224	9.38		
2	MA 2	-0.5302	0.1220	-4.35		
DIFFERENCING. 2 REGULAR						
RESIDUALS.	SS =	0.0429018	(BACKFORECASTS EXCLUDED)			
	DF =	43	MS =	0.0009977		
NO. OF OBS.	ORIGINAL SERIES	47	AFTER DIFFERENCING	45		

The random shock element a_t is the stochastic component in the equation. In the ARIMA model this moving average component can be shown to be mathematically equivalent to the exponentially weighted average of all previous observations (Pankratz, 1983, p. 49, 109; McCleary and Hay, (1980), p. 63).

University B

Identification.

An examination of the estimated ACF and PACF of the raw data (Figure 4) suggests that this data is also non-stationary and needs to be differenced.

A single spike on the PACF suggests a (0, 1, 1) model.

Estimation.

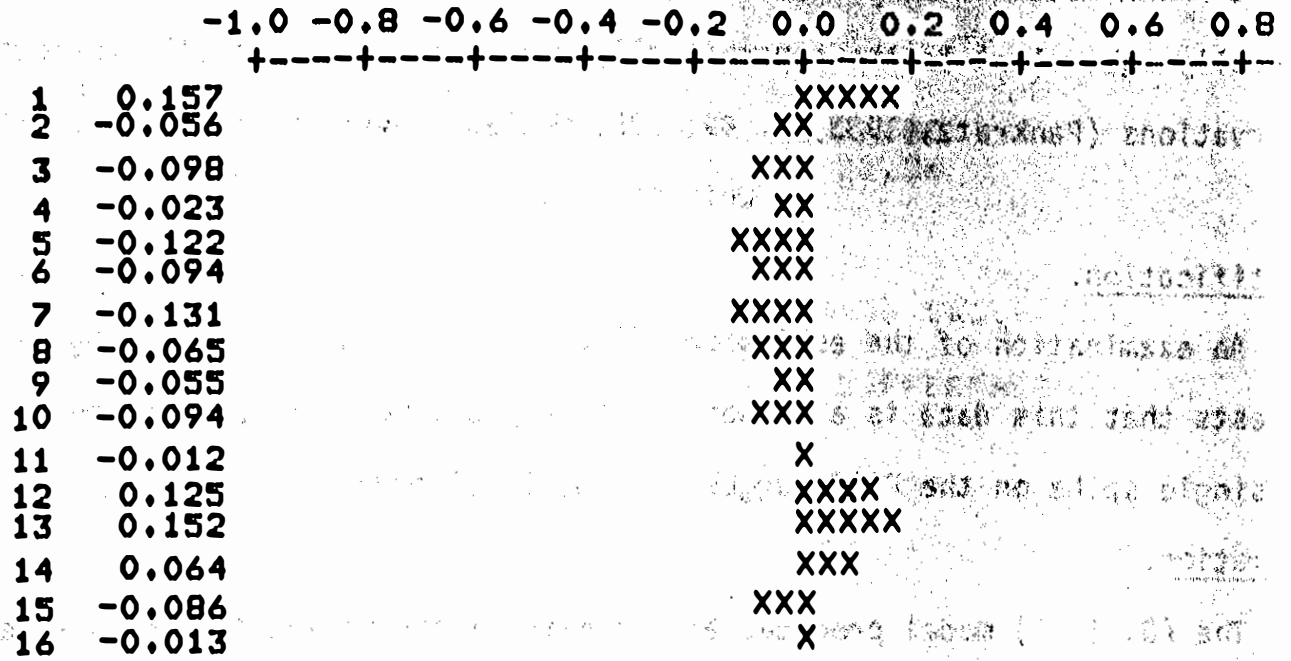
The (0, 1, 1) model produced an estimate of the Moving Average parameter with a t-value of .23. Since this was far from statistical significance, no adjustments needed to be made. Second differences were used, since the data appeared to approximate a quadratic trend. The (0, 2, 1) model produced an estimate of the MA parameter with a t-value of 11.12, which was highly significant.

Diagnosis.

The model was first diagnosed by comparing it with a more complex model. Accordingly, a (0, 2, 2) model was tested. It produced significant t-values for both MA parameters, as shown in Table 1. To compare the two models, the Mean Squares of the residuals was computed. The (0, 2, 1) model yielded $MSR = .0011274$, while the (0, 2, 2) model yielded $MSR = .0009977$. Finally, a (1, 2, 2) model (yet more complex) was tested, but it yielded $MSR = .011641$. Consequently, the (0, 2, 2) model was favored, since it yielded the smallest MSR.

The ACF and PACF for the Residuals of model (0, 2, 2) are shown in Figure 7. No spikes are shown at lag 1 or any other lags. The residuals appear to meet the diagnostic criteria, so the model is accepted.

Autocorrelations



Partial Autocorrelations

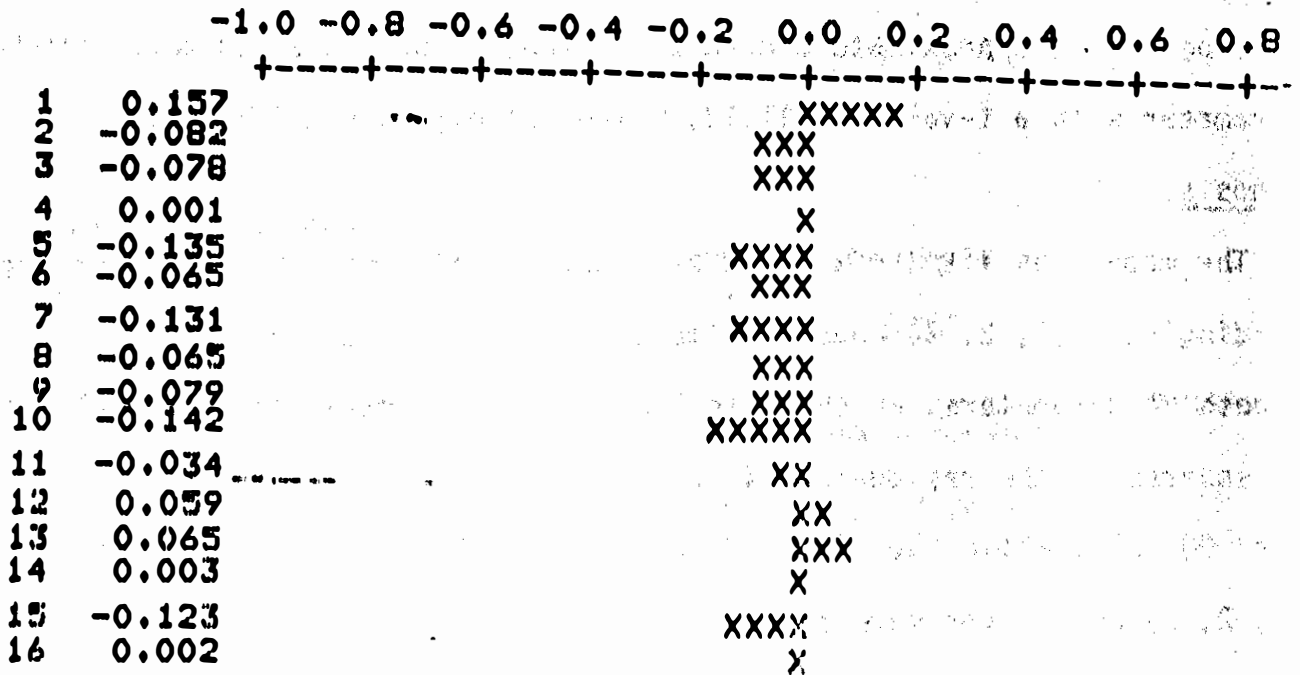


Figure 7. Estimated ACF and PACF for residuals from Arima (0, 2, 2) model. University B.

The model can be conveniently written as $(1-B)^2 y_t = (1 - 1.1475B + .5302B^2) a_t$.

Conclusion

This paper has suggested that meaningful mathematical models can be created to describe the time series of changes in the yearly grade point average at a university. The models are very tentative, partly because of the small number of available observations and also because of their relative complexity.

While this paper has not answered the questions about the so-called "grade inflation," it has indicated that a mathematical description of the time series of grades is sufficiently complex to suggest that no simple answer may suffice. The data is unstationary, as shown by the need for differencing. It further appears to be best modeled by an approach that postulates random shocks that persist for only a finite time, yet each of which can be represented as an exponentially weighted average of all previous observations. This perhaps reflects both the influx of new students and the persistent effects of traditional grading practices.

Data for this study was available for only two institutions of higher education, so the generalizability of the results is limited. Studies with data from other institutions would serve to indicate the existence of general patterns across institutions.

References

- Birnbaum, R. (1977). Factors relating to university grade inflation. *Journal of Higher Education*, 48, 519-539.
- Box, G. E. P. and Jenkins, G. M. (1976). Time series analysis: Forecasting and control. San Francisco, Holden-Day.
- McCleary, R. and Hay, R. A. (1980). Applied time series analysis for the social sciences. Beverly Hills: Sage Publications.
- Pankratz, A. (1983). Forecasting with univariate Box-Jenkins models. New York: John Wiley.
- Rogers, B. R. (1983). A time series approach to the longitudinal study of undergraduate grades, Paper presented at the annual meeting of the National Council of Measurement in Education, April 13, 1983, Montreal. (ERIC No. ED 235 228)
- Ryan, T. A., Joiner, B. L., and Ryan, B. F. (1982). Minitab reference manual. University Park, Pa: Pennsylvania State University.