



DUDLEY KNOX LIBRARY  
NAVAL POSTGRADUATE SCHOOL  
MONTEREY, CALIFORNIA 93943





# NAVAL POSTGRADUATE SCHOOL

## Monterey, California



# THESIS

REGRESSION MODELS  
OF QUARTERLY OVERHEAD COSTS  
FOR SIX GOVERNMENT AEROSPACE CONTRACTORS

by

David J. Jerabek

March 1986

Thesis Advisor:

D. C. Boger

Approved for public release; distribution is unlimited.

T226353



## REPORT DOCUMENTATION PAGE

REPORT SECURITY CLASSIFICATION <b>CLASSIFIED</b>		1b. RESTRICTIVE MARKINGS	
SECURITY CLASSIFICATION AUTHORITY		3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.	
DECLASSIFICATION / DOWNGRADING SCHEDULE		5 MONITORING ORGANIZATION REPORT NUMBER(S)	
PERFORMING ORGANIZATION REPORT NUMBER(S)		5 MONITORING ORGANIZATION REPORT NUMBER(S)	
NAME OF PERFORMING ORGANIZATION Naval Postgraduate School		6b OFFICE SYMBOL (if applicable) Code 55	
ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000		7a NAME OF MONITORING ORGANIZATION Naval Postgraduate School	
7b. ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
NAME OF FUNDING / SPONSORING ORGANIZATION		8b. OFFICE SYMBOL (if applicable)	
ADDRESS (City, State, and ZIP Code)		10 SOURCE OF FUNDING NUMBERS	
		PROGRAM ELEMENT NO	PROJECT NO
		TASK NO	WORK UNIT ACCESSION NO
TITLE (Include Security Classification) REGRESSION MODELS OF QUARTERLY OVERHEAD COSTS FOR SIX GOVERNMENT AEROSPACE CONTRACTORS			
PERSONAL AUTHOR(S) Grabek, David J.			
TYPE OF REPORT Master's Thesis	13b TIME COVERED FROM _____ TO _____	14 DATE OF REPORT (Year, Month, Day) 1986 March	15 PAGE COUNT 94
SUPPLEMENTARY NOTATION			
COSATI CODES		18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	Regression, Fourth-Order Autocorrelation, Overhead Costs, Quarterly Data	
	SUB-GROUP		
ABSTRACT (Continue on reverse if necessary and identify by block number)			
<p>Since overhead costs constitute a large percentage of total cost for aerospace contractors, it is important to be able to predict them accurately. The research performed in this thesis takes six government aerospace contractors and obtains regression models of their overhead costs that can be utilized for forecasting purposes. This method is preferable to some of the more commonly used methods because it estimates overhead costs directly, eliminating reliance upon predicted overhead rates. After the data were transformed to eliminate the effects of autocorrelation, excellent structural results were obtained for five of the six aerospace contractors. A Monte Carlo simulation was performed to compare various estimators of the autocorrelation. Two of the estimators were found to be superior. These two estimators are both two-stage estimators that are calculated utilizing Wallis's test statistic for fourth-order autocorrelation.</p>			
DISTRIBUTION / AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS <input type="checkbox"/>		21 ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
NAME OF RESPONSIBLE INDIVIDUAL John C. Boger		22b TELEPHONE (Include Area Code) (408) 646-2607	22c. OFFICE SYMBOL Code 54Bk

Approved for public release; distribution is unlimited.

Regression Models of Quarterly Overhead Costs  
For Six Government Aerospace Contractors

by

David J. Jerabek  
Lieutenant, United States Navy  
B.A., United States Naval Academy, 1979

Submitted in partial fulfillment of the  
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL  
March 1986



## ABSTRACT

Since overhead costs constitute a large percentage of total cost for aerospace contractors, it is important to be able to predict them accurately. The research performed in this thesis takes six government aerospace contractors and obtains regression models of their overhead costs that can be utilized for forecasting purposes. This method is preferable to some of the more commonly used methods because it estimates overhead costs directly, eliminating reliance upon predicted overhead rates. After the data were transformed to eliminate the effects of autocorrelation, excellent structural results were obtained for five of the six aerospace contractors. A Monte Carlo simulation was performed to compare various estimators of the autocorrelation. Two of the estimators were found to be superior. These two estimators are both two-stage estimators that are calculated utilizing Wallis's test statistic for fourth-order autocorrelation.

7-11  
2013  
01

TABLE OF CONTENTS

I.	INTRODUCTION . . . . .	8
II.	DATA SOURCES AND CHARACTERISTICS . . . . .	10
III.	MODELING QUARTERLY OVERHEAD COST . . . . .	12
	A. PRESENCE OF AUTOCORRELATION . . . . .	12
	B. THEORETICAL MODEL UTILIZED . . . . .	19
	C. ESTIMATORS OF FOURTH-ORDER AUTOCORRELATION . . . . .	20
	D. TRANSFORMATION FOR AUTOCORRELATION . . . . .	24
	E. PROCEDURE . . . . .	25
IV.	SMALL-SAMPLE PROPERTIES OF SEVERAL ESTIMATORS OF SAR(4) . . . . .	29
	A. GENERAL . . . . .	29
	B. RESULTS FROM PREVIOUS MONTE CARLO ANALYSIS . . . . .	29
	C. THE MODEL . . . . .	31
	D. DATA GENERATION . . . . .	32
	E. MEASURES OF EFFECTIVENESS . . . . .	34
	F. RESULTS . . . . .	35
	1. General . . . . .	35
	2. Estimation of Rho . . . . .	35
	3. Estimation of Slope Coefficient . . . . .	40
	4. EGLS Regression Quality . . . . .	40
	5. Summary . . . . .	49
V.	STRUCTURAL ANALYSIS . . . . .	51
	A. GENERAL . . . . .	51
	B. PROCEDURE . . . . .	51
	C. THE REMAINING MODELS . . . . .	62
	1. Contractor B . . . . .	62
	2. Contractor C . . . . .	63

3.	Contractor D . . . . .	64
4.	Contractor E . . . . .	64
5.	Contractor F . . . . .	66
D.	SUMMARY . . . . .	66
VI.	CONCLUSIONS . . . . .	71
	LIST OF REFERENCES . . . . .	74
	APPENDIX : COMPUTER PROGRAMS . . . . .	76
	INITIAL DISTRIBUTION LIST . . . . .	93

## LIST OF TABLES

1.	SIMULATION PARAMETERS . . . . .	32
2.	RMSE OF RHO FOR THE THREE SIMULATION RUNS . . .	36
3.	RELATIVE EFFICIENCIES FOR THE THREE SIMULATION RUNS . . . . .	41
4.	ADJUSTED R-SQUARED FOR THE THREE SIMULATION RUNS . . . . .	45
5.	RESULTS FOR CONTRACTOR A . . . . .	52
6.	RESULTS FOR CONTRACTOR B . . . . .	63
7.	RESULTS FOR CONTRACTOR C . . . . .	65
8.	RESULTS FOR CONTRACTOR D . . . . .	66
9.	RESULTS FOR CONTRACTOR E . . . . .	67
10.	RESULTS FOR CONTRACTOR F . . . . .	68

## LIST OF FIGURES

3.1a	Time Series Plots of the Dependent Variable . . . .	15
3.1b	Time Series Plots of the Dependent variable . . . .	16
3.2a	Time Series Plots of the Independent Variable . . . .	17
3.2b	Time Series Plots of the Independent Variable . . . .	18
4.1a	RMSE of Rho for Simulation for Contractor A . . . .	37
4.1b	RMSE of Rho for Simulation for Contractor B . . . .	38
4.1c	RMSE of Rho for Simulation for Contractor E . . . .	39
4.2a	Relative Efficiencies for Simulation for Contractor A . . . . .	42
4.2b	Relative Efficiencies for Simulation for Contractor B . . . . .	43
4.2c	Relative Efficiencies for Simulation for Contractor E . . . . .	44
4.3a	Adjusted R-squared for Simulation for Contractor A . . . . .	46
4.3b	Adjusted R-squared for Simulation for Contractor B . . . . .	47
4.3c	Adjusted R-squared for Simulation for Contractor E . . . . .	48
5.1	Autocorrelation Function of the Residuals for Contractor A . . . . .	53
5.2	Autocorrelation Function of the Residuals for Contractor A's model (estimator 3) . . . . .	54
5.3	Tests for Normality of Residuals for Contractor A's model (estimator 3) . . . . .	56
5.4	Tests for Homogeneity of Variance of the Residuals for Contractor A's model (estimator 3) . . . . .	57
5.5a	Autocorrelation Functions of the Residuals . . . .	60
5.5b	Autocorrelation Functions of the Residuals . . . .	61

## I. INTRODUCTION

The purpose of this thesis is to analyze the overhead costs of six government aerospace contractors, determine the best estimators of autocorrelation found in the residuals, and then obtain regression models for the overhead costs. The most frequently used method to estimate overhead costs is to use estimated overhead rates. These rates are applied to estimated labor hours or costs in several functional categories. Summing all of these category values gives total overhead. This method results in poor estimation in cases where the firm's output fluctuates significantly as in the case of government aerospace contractors. Consequently, it is unsatisfactory for use with these six aerospace contractors. Another approach, the one utilized in this thesis, is to estimate overhead costs directly, and hence eliminate reliance upon overhead rates.

For aerospace contractors, overhead cost comprises 30 to 50 percent of total cost. Therefore, it is important to be able to accurately predict these costs. It is also desirable that the predictive model be simple to utilize, without sacrificing predictive power, since personnel of various statistical backgrounds may be required to use it (Boger, 1983, pp.5-7).

The work performed in this thesis is a continuation of that presented in Boger (1984). The statistical methods and procedures used herein to obtain the predictive models have already been proven to give good, useful results. There are additions in two areas to the data used by Boger. First, data were obtained for two additional quarters and, secondly, data were obtained for one additional contractor. The major extension on Boger, and the emphasis of this thesis, is in determining the best method to use to estimate

autocorrelation. The focus of the analysis performed herein is to derive a simple and efficient regression model for overhead cost.

## II. DATA SOURCES AND CHARACTERISTICS

The data were obtained from six government aerospace contractors. To maintain confidentiality, all references to specific contractors will be with the labels A through F. Prior to obtaining the data, a specific format for data collection was selected to insure uniformity of data categories among the contractors. The overhead cost data were then collected on a quarterly basis for the selected categories from each of the contractors, starting with the first quarter of 1979 through the fourth quarter of 1984. The data for the last two quarters of 1984 were unavailable for contractor E. The data for the last two quarters of 1984 for contractor B were eliminated from further analysis because they were clearly outliers. In addition to the cost data, data pertaining to production and operating characteristics were obtained.

There are three major categories, two of which are made up of several subcategories, which comprise total overhead costs. The first major category, labor related-costs, has subcategories of indirect salaries and fringe benefits. The second major category, facilities costs, includes all facilities-related costs. The last major category, the mixed costs category, has three subcategories. These three subcategories are Independent Research and Development and Bid and Proposal (IR&D/B&P) costs, Electronic Data Processing (EDP) costs, and a subcategory that contains all other costs related to overhead.

The cost data were then converted from current dollars to constant fourth quarter 1984 dollars. This conversion was accomplished using Bureau of Labor Statistics (BLS) and Gross National Product Deflator (GNPD) indices for the appropriate categories. The labor-related costs were



converted using BLS SIC 372, the price index for production-worker average hourly wages for the aircraft and parts industry. In this case the only indices available were monthly indices. The monthly indices were then averaged to obtain quarterly values. The facilities costs were adjusted using the GNPD gross private domestic fixed nonresidential investment index, published by the Bureau of Economic Analysis. The mixed costs were adjusted using the GNPD personal consumption services expenditure index. As with all indices, those used here are imperfect. They were selected because they should provide the best adjustments for inflation among all readily available and relevant indices.

As previously mentioned, data in different production and operational categories were also obtained. The only one used in this analysis was the direct labor personnel category. Due to the nature of this data, no adjustment was necessary.

### III. MODELING QUARTERLY OVERHEAD COST

#### A. PRESENCE OF AUTOCORRELATION

Whenever a statistical model is based upon time series data, as is the case herein, the residuals can be expected to exhibit some form of autocorrelation. Further, "the shorter the periods of individual observations, the greater the likelihood of encountering autoregressive disturbances" (Kmenta, 1971, p.270). Thus the presence of autocorrelation is more likely with quarterly observations than when the data are smoothed by reporting annual values. The most common assumption about the form of the autocorrelation of the errors terms is that the errors are first-order autoregressive (Judge et. al., 1985, p.275). This model is called an AR(1) process and possesses the form

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + v_t, \quad (3.1)$$

where  $\varepsilon_t$  is the error term from a regression model corresponding to the observation at time  $t$ . The  $\rho_1$  term is the coefficient of correlation between the related error terms,  $\varepsilon_t$ , and  $\varepsilon_t$  of lag 1. The  $v_t$  are normally and independently distributed random variables with mean 0 and constant variance  $\sigma_v^2$ .

With autocorrelation the assumption that the error terms,  $\varepsilon_t$ , are independent identically distributed normal random variables with mean 0 and variance  $\sigma^2$  is not true. They are related to previous error terms and depend upon the form of autocorrelation present. Secondly, when the autocorrelation is AR(1), the variance of the error terms is  $\sigma_v^2 / (1 - \rho_1^2)$  (Kmenta, 1971, p.271).

When the error terms are autocorrelated the least squares estimators of the regression coefficients are still

unbiased and consistent, but they are no longer efficient or asymptotically efficient. The standard estimates of the variances of the coefficients are also biased. When the error terms are positively autocorrelated, as is most common for economic time series data, this variance will be biased downward (Kmenta, 1971, pp.278-283). This will cause the coefficient of determination, R-squared, and the t and F-statistics to be exaggerated (Maddala, 1977, p.283). The upward bias in each of these statistics leads to an unwarranted confidence in the regression model.

The adverse effect of an overestimated value of R-squared is obvious since the higher the R-squared value, the better fit the model is assumed to give. An upwardly biased t-statistic makes it easier to reject, for any given level of significance, the null hypothesis that the regression coefficient equals zero, thus, again giving the false impression of a more significant regression model. The effect of an exaggerated F-statistic is similar to that of the t-statistic. It is easier to reject the standard compound null hypothesis that all regression coefficients equal zero for the F-test.

So, in addition to obtaining unbiased estimates of the regression coefficients, it is equally important to obtain unbiased estimates of their standard errors. Only then can reliable statistics be obtained that can accurately assess the quality of the regression model.

As the AR(1) model indicates, the error terms in one period are related to those occurring one period prior. The AR(1) process is common in yearly economic data. When the data observations are quarterly, a special form of the fourth-order autoregressive, SAR(4), process will be present instead of the standard AR(4) process (Wallis, 1972, p.618).

This special SAR(4) process has the form

$$\varepsilon_t = \rho_4 \varepsilon_{t-4} + v_t. \quad (3.2)$$

With this form of autocorrelation the error terms are related to those occurring in the corresponding quarters of successive years. Using the notation of Box and Jenkins this is a seasonal model of order  $(1,0,0) \times (0,0,0)_4$ . The period equals four since the data is quarterly and can be expected to show seasonal effects within years (Box and Jenkins, 1970, pp.301-305). When this SAR(4) process is present the variance of the error terms is  $\sigma_v^2 / (1 - \rho_4^2)$  (Judge et. al., 1985, p.298). This special SAR(4) process should be distinguished from the general AR(4) process which possesses the form

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \rho_3 \varepsilon_{t-3} + \rho_4 \varepsilon_{t-4} + v_t. \quad (3.3)$$

The SAR(4) process used in this analysis considers the effect of the three previous quarters to be negligible compared to that of the same quarter of the previous year (Boger, 1983, p.16).

Time series plots of the dependent variable (Figures 3.1a and 3.1b) total overhead costs, suggest that the Wallis SAR(4) model is appropriate for this problem. As can be seen, the data show a seasonal effect within years. The data appear to follow a yearly cycle in which the quarterly values of successive years fall in the same relative position with respect to the remaining three quarters of their respective years.

The time series plots of the independent variable (Figures 3.2a and 3.2b), direct labor personnel, show that they all appear to follow some type of cycle or general trend. Contractors C, E, and F all appear to go through a

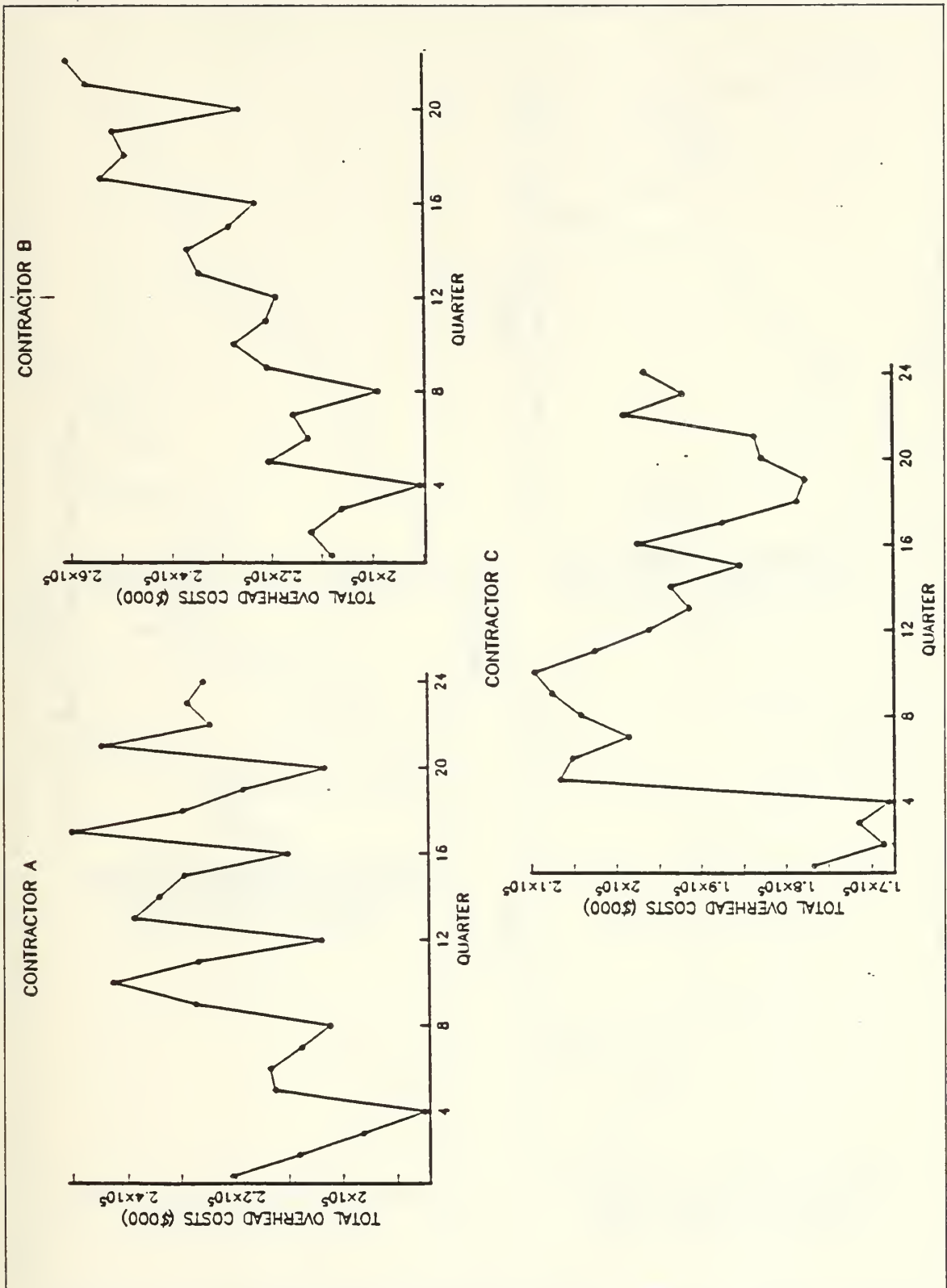


Figure 3.1a Time Series Plots of the Dependent Variable.

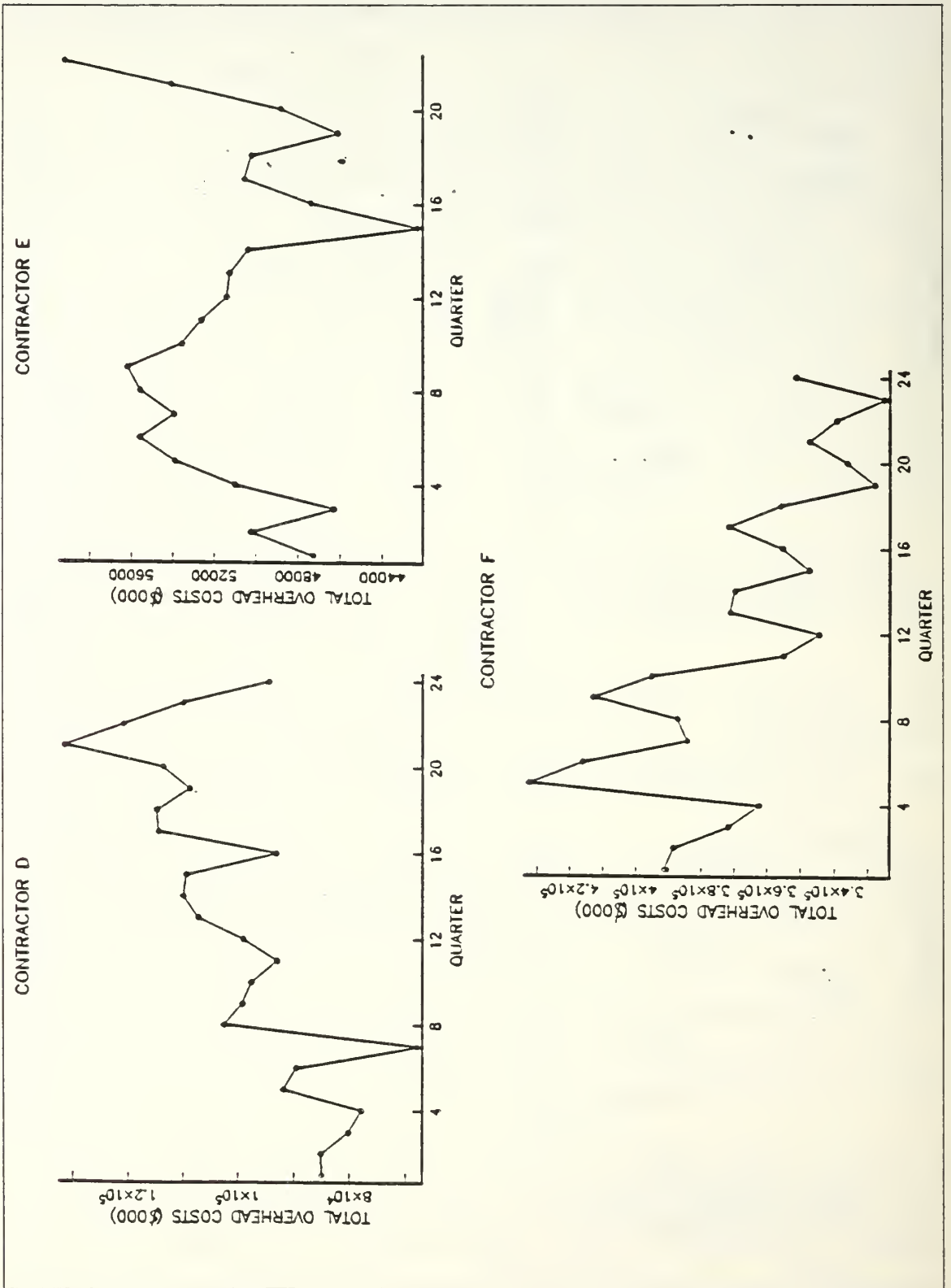


Figure 3.1b Time Series Plots of the Dependent variable.



Figure 3.2a Time Series Plots of the Independent Variable.

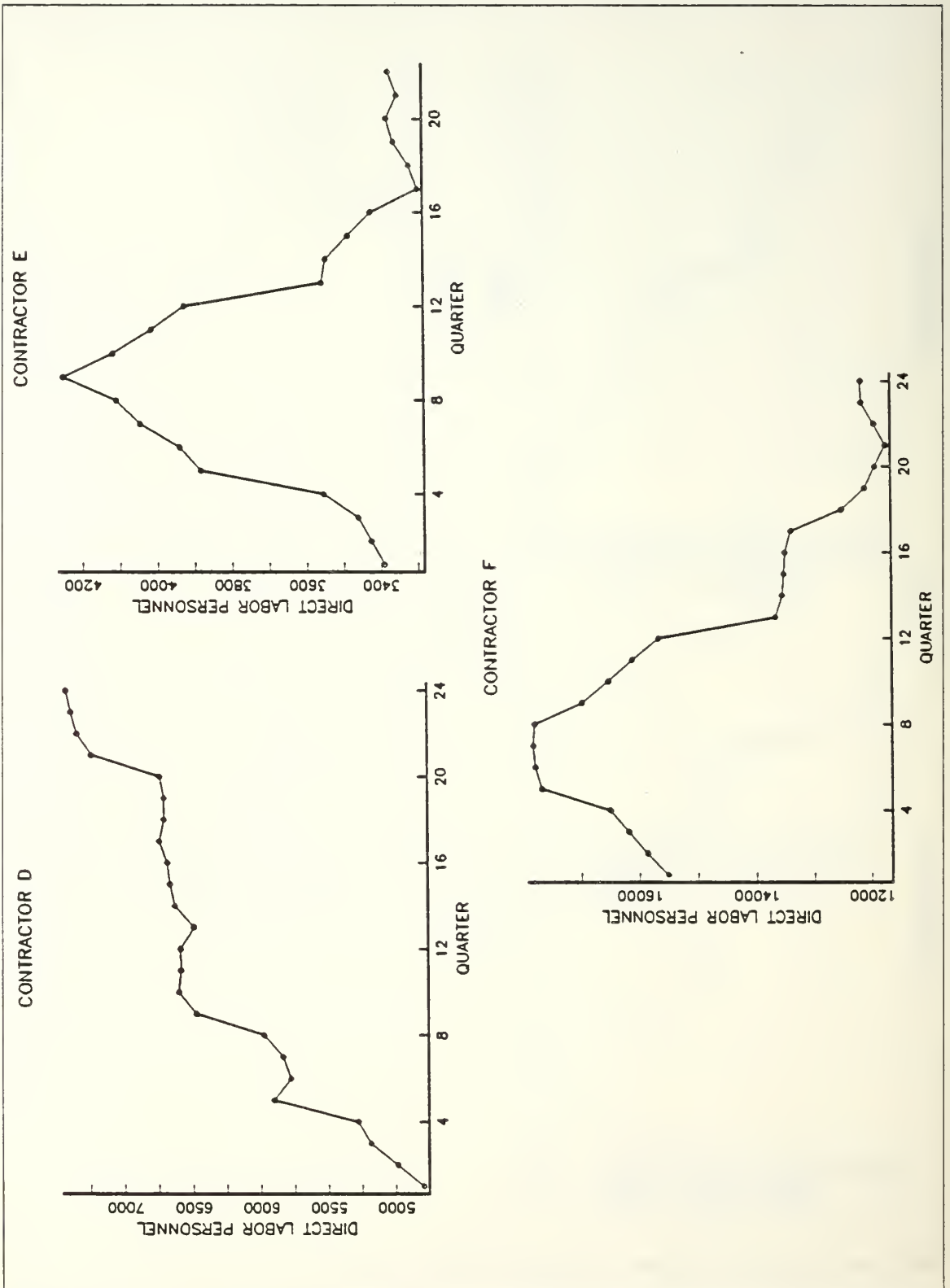


Figure 3.2b Time Series Plots of the Independent Variable.



single cycle. They start with an upward trend for approximately eight quarters, followed by a gradual decline for about twelve quarters, and end with the start of another upward swing. Contractor A appears to run through two shorter cycles. Contractors B and D both appear to follow a gradual, steady upward trend.

#### B. THEORETICAL MODEL UTILIZED

The general model utilized for overhead costs in this analysis is of the form

$$Y_t = X_t\beta + \varepsilon_t \quad (3.4)$$

$$\varepsilon_t = \rho_i\varepsilon_{t-i} + v_t, \quad (3.5)$$

where  $X_t$  is a  $t \times 2$  matrix and  $\beta$  is a  $2 \times 1$  vector.  $Y_t$  is the dependent variable, total overhead costs. The columns of  $X_t$  are the independent variable, direct labor personnel, preceded by a column of 1's for the constant term. Only one independent variable is utilized because it satisfactorily explains the dependent variable and minimizes the complexity of the model. The error term has the structure shown in equation (3.5) where  $i$  is 1 for an AR(1) process, or 4 for Wallis's special SAR(4) process. The  $v_t$  are normally and independently distributed random variables with mean 0 and constant variance.

As previously mentioned, when autocorrelation is present the estimators of the regression coefficients are not efficient and their variances are biased. If  $\rho$  is a known quantity, then the X and Y variables can be transformed to eliminate the effect of the autocorrelation. A regression of these new transformed variables, the Generalized Least Squares (GLS) solution, yields results that correct these deficiencies. However  $\rho$  is seldom, if ever, known. This difficulty can be overcome by estimating  $\rho$  from the data.

This yields the Estimated GLS (EGLS) solution that also possesses the desired properties (Kmenta, 1971, p.284).

### C. ESTIMATORS OF FOURTH-ORDER AUTOCORRELATION

One of the major thrusts of this analysis is to arrive at the most efficient estimate of  $\rho_4$ . All of the estimators considered fall into one of two categories, iterative or maximum likelihood procedures. All but one of the iterative estimators can be more specifically classified as two-stage estimators. The theory behind maximum likelihood estimation can be found in most intermediate level probability and statistics texts. The basic approach to performing the iterative, including the more specific two-stage, procedure is presented in Kmenta (1971, p.288). The first seven estimators are two-stage estimators, while the eighth estimator utilizes the full iterative procedure. The ninth estimator is the maximum likelihood estimator.

Since the AR(1) process is the most frequently encountered form of autocorrelation, most of the estimators of  $\rho_4$  used herein are simply adaptations of their corresponding  $\rho_1$  estimator. Judge et. al. (1985) derives six estimators for the AR(1) process case. Adaptations of the estimators of Judge are given below.

The first estimator of  $\rho_4$  is the standard Prais-Winsten estimator

$$\hat{\rho}_4 = \frac{\sum_{t=5}^T e_t e_{t-4}}{\sum_{t=1}^T e_t^2}, \quad (3.6)$$

where  $e_t = Y_t - X_t \hat{\beta}$ , the residual from the OLS regression in equation (3.4). This estimator is simply the sample correlation coefficient when the residuals possess the autocorrelation process shown in equation (3.5) for  $i=4$ . The

AR(1) process version of this estimator is known to give a downwardly biased estimate of  $\rho_1$  (Park and Mitchell, 1980, p.189). So we could expect this same result when using equation (3.6) to estimate  $\rho_4$ .

The second estimator of  $\rho_4$  is

$$\hat{\rho}_4 = \frac{(T-K) \sum_{t=5}^T e_t e_{t-4}}{(T-1) \sum_{t=1}^T e_t^2}, \quad (3.7)$$

where T is the number of observations and K the number of parameters that must be estimated, in this case 2. This is simply a modification to equation (3.6) derived by Theil to incorporate a degrees-of-freedom correction. As can be seen it will further increase the downward bias of equation (3.6).

The third estimator of  $\rho_4$  is

$$\hat{\rho}_4 = 1 - .5d_4, \quad (3.8)$$

where  $d_4$  is the Wallis test statistic for the special SAR(4) process. As presented in Wallis (1972) the equation for  $d_4$  is

$$d_4 = \frac{\sum_{t=5}^T (e_t - e_{t-4})^2}{\sum_{t=1}^T e_t^2}. \quad (3.9)$$

So equation (3.8) is easily computed once the test statistic,  $d_4$ , has been calculated.

The fourth estimator of  $\rho_4$  is

$$\hat{\rho}_4 = \frac{T^2(1 - .5d_4) + K^2}{T^2 - K^2}. \quad (3.10)$$

This estimator is a modification to equation (3.8) derived by Theil and Nagar. It is an improvement over equation (3.8) if the explanatory variables are smooth, "in particular, if the first and second differences of each explanatory variable are small in absolute value compared to the range of the variable itself" (Intriligator, 1978, p.164).

The fifth estimator of  $\rho_4$  is the Durbin estimator. It is the coefficient of  $Y_{t-4}$  obtained in a regression of  $Y_t$  on  $Y_{t-4}$ , a constant,  $X_t$ , and  $X_{t-4}$ :

$$Y_t = \rho_4 Y_{t-4} + \beta_1(1-\rho_4) + \beta_2 X_t - \beta_2 \rho_4 X_{t-4} + \varepsilon_t, \quad (3.11)$$

$t = 5, 6, \dots, T.$

The sixth estimator of  $\rho_4$  is obtained from an OLS regression of  $e_t$  on  $e_{t-4}$ :

$$e_t = \rho_4 e_{t-4} + v_t, \quad t = 5, 6, \dots, T. \quad (3.12)$$

The seventh estimator of  $\rho_4$  is an adaptation of the Park and Mitchell (1980) estimator:

$$\hat{\rho}_4 = \frac{\sum_{t=5}^T e_t e_{t-4}}{\sum_{t=5}^{T-1} e_t^2}. \quad (3.13)$$

As can be seen this is a modification of equation (3.6) where the summation in the denominator excludes the first four and the last observation. This will reduce the downward bias associated with equation (3.6).

The eighth estimator of  $\rho_4$  is the iterative Prais-Winsten estimator utilizing equation 3.13. This is the same as estimator seven except that the iterative procedure is carried out until convergence is achieved. For this estimator, as well as the maximum likelihood estimator that follows, convergence was defined to be consecutive estimates within  $|.00001|$  of each other. Each procedure was defined to be nonconvergent if convergence had not occurred within fifty-two iterations (Park and Mitchell, 1979, pp.2-6).

The maximum likelihood estimator, the ninth estimator of  $\rho_4$ , was obtained using the iterative algorithm derived by Beach and MacKinnon (1978). The specific procedure they derived was for the AR(1) process, so again minor adjustments had to be made to tailor it to the SAR(4) process. The only alterations necessary were in the calculation of the coefficients of the polynomial

$$f(\rho) = \rho^3 + A\rho^2 + B\rho + C = 0. \quad (3.14)$$

The coefficients are now computed as follows:

$$A = -(T-2) \sum_{t=5}^T e_t e_{t-4} / \text{DENOM} \quad (3.15)$$

$$B = [(T-1) \sum_{t=1}^4 e_t^2 - T \sum_{t=5}^T e_{t-4}^2 - \sum_{t=5}^T e_t^2] / \text{DENOM} \quad (3.16)$$

$$C = T \sum_{t=5}^T e_t e_{t-4} / \text{DENOM}, \quad (3.17)$$

where the common denominator is

$$\text{DENOM} = (T-1) \left( \sum_{t=5}^T e_{t-4}^2 - \sum_{t=1}^4 e_t^2 \right). \quad (3.18)$$

The remainder of the algorithm is the same as that presented in Beach and MacKinnon (1978). Beach and MacKinnon advertise that this algorithm should converge in four to seven iterations for five digit accuracy.

#### D. TRANSFORMATION FOR AUTOCORRELATION

If first or fourth-order autocorrelation is shown to exist in the residuals, then the data must be transformed to eliminate its presence. The transformation used for an AR(1) process is discussed in Judge et.al. (1985, p.285). It should be noted here that only one estimator for  $\rho_1$  was used, the two stage Prais-Winsten estimator derived in Park and Mitchell (1980)

$$\hat{\rho}_1 = \frac{\sum_{t=2}^T e_t e_{t-1}}{\sum_{t=2}^{T-1} e_t^2}. \quad (3.19)$$

This is the AR(1) process version of equation (3.13). This particular estimator was selected because it was found to perform better than any other commonly used alternative, including the more standard Prais-Winsten estimator (see Judge et. al., 1985, p.286), in Park and Mitchell's 1980 analysis.

For the SAR(4) process the transformation is

$$z_t^* = z_t (1 - \rho_4^2)^{1/2}, \quad t = 1, 2, 3, 4 \quad \text{and} \quad (3.20)$$

$$z_t^* = z_t - \rho_4 z_{t-4}, \quad t = 5, 6, \dots, T, \quad (3.21)$$

where  $\rho_4$  must be estimated using any of its known estimators. Some estimators, of course, are more efficient than others. This is one item to be resolved for the specific models used herein.

Note that both of these transformations utilize all T observations. An alternative approach is to omit a number of the initial observations in the transformation (see Cochrane and Orcutt, 1949, p.35). The number of omissions is dependent upon the type of autocorrelation found present in the data. The first observation is omitted when the data are being transformed to eliminate the presence of an AR(1) process. The transformation for an SAR(4) process would result in the omission of the first four observations.

It has been shown that the use of all T observations generally results in much more efficient results (Spitzer, 1979, and Park and Mitchell, 1980). It should be noted that these results were arrived at by studying the AR(1) process transformation where only the first observation is involved. The effects should be even more dramatic for the SAR(4) process transformation, since the first four observations are involved. It is also worth noting that the relative efficiency of these two alternative transformations is related to the specification of the independent variable (Maeshiro, 1979 and Taylor, 1981). Maeshiro found that in the case where the independent variable is trended and  $\rho=0$  (as is most commonly the case with economic data) the retention of all T observations greatly increased the efficiency of the estimator. He also found that retention of the initial observations was not as critical for the case of an untrended independent variable.

#### E. PROCEDURE

The general procedure followed herein was to first perform an Ordinary Least Squares (OLS) regression with direct labor personnel as the independent variable and total overhead cost as the dependent variable. Direct labor personnel was selected as the independent variable because it was shown to perform the best among numerous explanatory variables in a single variable regression with total

overhead cost in Boger's 1983 analysis. The residuals were then analyzed and tested for the presence of Wallis's special SAR(4) process, the AR(1) process, or a combination of both. To do this, the autocorrelation function of the residuals was first looked at to get an overall picture of the type of autocorrelation present. More formal testing was then performed.

The Durbin-Watson test was used to check for the presence of the AR(1) process (see Kmenta, 1971, pp.295-296). The Wallis test, a generalization of the Durbin-Watson test, was utilized to check for the presence of the SAR(4) process (see Wallis, 1972, pp.624-625). In both cases a two-sided test was performed using the null hypothesis  $H_0: \rho=0$ , versus the alternative  $H_1: \rho \neq 0$ . A significance level of size  $\alpha=.10$  was used to define the critical region.

One problem with both of these tests is the inconclusive region between the upper and lower significance points that determines the critical region. The size of this inconclusive region increases as the sample size decreases or as the number of regressors increases (Wallis, 1972, p.625). So in this analysis we are handicapped by the small sample size, but it is to our advantage here in keeping the number of regressors to a minimum. Maddala (1977, pp.285-286) presents numerous suggestions, derived by others, in dealing with this inconclusive region for the Durbin-Watson test. In this analysis we chose the statistically conservative approach of ignoring the lower significance point and using only the upper point to define the critical region. This rule was followed for both the Durbin-Watson and Wallis tests. This method is said to perform well in many situations for the Durbin-Watson test (Draper and Smith, 1981, p.167). As presented in Draper and Smith (1981) the rejection criteria for the two sided test for this rule are as follows; if  $d < d_u$  or  $4-d < d_u$ , reject  $H_0$  at level  $2\alpha$ . Any point that would have



fallen in the inconclusive region before, would now fall in the critical region and lead to the rejection of our null hypothesis. This treatment of the inconclusive region is also recommended by Hannan and Terrel (1968). They feel that the upper significance point is a good approximation for the bound on the critical region when the regressors are slowly changing. They further state that economic time series, as is the case here, are slowly changing so the upper significance point can be used as the lone significance point for the Durbin-Watson test (Maddala, 1977, p.285). In another study, Theil and Nagar (1961) computed significance points for the Von Neumann ratio of least-squares estimated regression disturbances, which is equivalent to the Durbin-Watson test statistic. Their calculated significance points were very close to the upper significance points for the Durbin-Watson test. So this also gives added credence to using the upper point as the sole significance point in performing these tests. Though all of the referenced results apply to the Durbin-Watson test, this rule was used on the Wallis test also because, as previously mentioned, the Wallis is a slight modification of the Durbin-Watson test.

Next, depending upon the form of autocorrelation found present, the data were transformed using the appropriate transformation. The EGLS solution was then obtained by reestimating the model using the transformed dependent and independent variables. Again the residuals of this regression were tested for the presence of autocorrelation. This cycle of reestimating the model and testing for autocorrelation was performed until a model was obtained where the residuals were free from any autoregressive process. Once this final model was obtained, the residuals were checked to insure that they were independent, identically distributed, Normal random variables with zero mean and constant

variance. In all cases this requirement was met. So the final models met all of the necessary assumptions required of a correct, reliable regression model.

#### IV. SMALL-SAMPLE PROPERTIES OF SEVERAL ESTIMATORS OF SAR(4)

##### A. GENERAL

As previously mentioned, one of the purposes of this thesis is to determine the best estimators of fourth-order autocorrelation. This is important because the performance of the EGLS regression is dependent upon the quality of the estimator (Rao and Griliches, 1969, p.258). In order to evaluate the nine estimators presented in Chapter 3, a Monte Carlo simulation was carried out to determine their relative performances. This chapter explains the simulation and presents the results. The three estimators that performed the best in this simulation were then used to obtain structural models for each contractor. These three models then provided another basis of comparison for the three final estimators. The computer programs utilized in the simulation are contained in the appendix.

##### B. RESULTS FROM PREVIOUS MONTE CARLO ANALYSIS

No previous simulations comparing estimators of fourth-order autocorrelation could be found. Therefore the results of this simulation could only be compared with results obtained from previous simulations that evaluated various estimators of first-order autocorrelation. Even though these past Monte Carlo simulations dealt with the AR(1), instead of SAR(4), process estimators, their results should still be useful in predicting the relative performance of the various estimators of  $\rho_4$ . Of the nine estimators evaluated in this thesis, results comparing the AR(1) process versions of only estimators one, five, seven, eight and nine could be found. As will be shown later in this chapter, estimators three and four proved to be the best of the nine estimators tested herein. It would have been interesting to

see how their first-order autocorrelation versions would have compared in these previous studies.

The Spitzer and Rao and Griliches studies compared the Durbin and standard Prais-Winsten estimators of first-order autocorrelation for Mean Squared Error (MSE) of  $\beta_2$ . In both analyses the Prais-Winsten estimator performed better for lower positive values ( $\rho_1 \leq .5$ ) while the Durbin estimator dominated for the higher values. The Spitzer study also included the maximum likelihood estimator and showed that it was better than both the Durbin and Prais-Winsten estimators for  $\rho_1 > .6$  for MSE of  $\beta_2$ , and better than the Durbin estimator for  $\rho_1 > .3$  for MSE of  $\rho_1$  (it wasn't compared to the Prais-Winsten estimator for MSE of  $\rho_1$ ). Park and Mitchell compared four estimators in their 1980 analysis for RMSE of  $\beta_1$  and  $\beta_2$ . The estimators they analyzed were the iterative Prais-Winsten, the maximum likelihood, their version of the Prais-Winsten (Equation (3.19)), and the standard Prais-Winsten estimators. Their maximum likelihood estimator was computed utilizing Beach and MacKinnon's algorithm. The iterative Prais-Winsten estimator was the best of the four with a slight edge over the maximum likelihood estimator. Since the iterative Prais-Winsten estimator outperformed its two-stage counterparts it was shown that iteration leads to a more efficient estimator. Of the two stage estimators, their version of the Prais-Winsten estimator was better than the standard version. Park and Mitchell's 1979 study showed that the iterative Prais-Winsten was also better than the maximum likelihood estimator for MSE of  $\rho_1$ . All of these studies showed that all of the estimators outperformed the OLS solution when a significant amount of autocorrelation was present in the residuals ( $\rho_1 > .2$ ).

### C. THE MODEL

The model utilized in the simulation was

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t \quad (4.1)$$

$$\varepsilon_t = \rho_4 \varepsilon_{t-4} + v_t, \quad (4.2)$$

where  $\varepsilon_t \sim N(0, \sigma_v^2 / (1 - \rho_4^2))$ ,  $v_t \sim N(0, \sigma_v^2)$ ,  $t = 1, 2, \dots, T$ , and  $|\rho_4| < 1.0$ . Three different independent variables were utilized in the simulation. They were the direct labor personnel for contractors A, B, and E. These were selected because they were three of the models where the SAR(4) process was found to be the most significant form of auto-correlation present in the residuals. A separate simulation run was performed for each of these so that the relative performance of the estimators could be compared for independent variables with different structures. The value of  $\sigma_v^2$  was particular to the contractor for which the simulation run was performed.

It was desired that the generated dependent variable,  $Y_t$ , be comparable in value and structure to the total overhead cost for the respective contractor (the dependent variable,  $Y_t$ , in equation (3.4)). Therefore the  $v_t$  terms of equation (4.2) had to be proportional to the  $v_t$  terms of equation (3.5). To accomplish this, the value of  $\sigma_v^2$  was the variance of the residuals,  $v_t$ , obtained from the OLS regression of  $e_t$  on  $e_{t-4}$ :

$$e_t = \rho_4 e_{t-4} + v_t. \quad (4.3)$$

The variables  $e_t$  and  $e_{t-4}$  were the (unlagged and lagged) residuals obtained from the OLS regression in equation (3.4). The sample size,  $T$ , was simply the number of data observations for each contractor, twenty-four for contractor A and twenty-two for contractors B and E. Each simulation was replicated one hundred times.

#### D. DATA GENERATION

For each simulation the independent variable,  $X_t$ , the regression coefficients,  $\beta_1$  and  $\beta_2$ , and  $\rho_4$  were predetermined, fixed values. As previously mentioned the three different independent variables were the direct labor personnel for contractors A, B, and E. Each simulation was run for values of  $\rho_4$  of .1, .2, .3, .4, .5, .6, .7, .8, .9 and .95. The value of  $\rho_4$  was restricted to positive values because this is the region most likely to be encountered for the SAR(4) process with economic data. As in the simulation performed by Rao and Griliches in 1969, the constant term,  $\beta_1$ , was set at zero. The value of the slope,  $\beta_2$ , was then set at a value that generated dependent variables,  $Y_t$ , proportional in size to the respective contractor's total overhead cost. The value of  $\beta_2$  for each simulation, is contained in Table 1.

TABLE 1  
SIMULATION PARAMETERS

Simulation run	$\beta_2$	$\sigma_v$
Contractor A	17	13454
Contractor B	17	8835
Contractor E	14	3444

With the ultimate goal of determining  $\hat{Y}_t$ , the data were generated using the following algorithm:

(1) Thirty-two  $v_t$  values were randomly generated from a normal distribution with mean 0 and variance  $\sigma_v^2$ . The value of the standard deviation,  $\sigma_v$ , for each simulation, is shown in Table 1. New  $v_t$  values were drawn for each of the 100 replications.

(2) The first four values of  $\varepsilon_t$  were computed to be

$$\varepsilon_t = v_t / (1-\rho_4^2)^{1/2}, \quad t = 1, 2, 3, 4. \quad (4.4)$$

This generated error terms from the normal distribution with mean 0 and variance  $\sigma_v^2 / (1-\rho_4^2)$ .

(3) The next twenty-eight (twenty-four for contractors B and E) values of  $\varepsilon_t$  were generated from

$$\varepsilon_t = \rho_4 \varepsilon_{t-4} + v_t, \quad t = 5, 6, \dots, T. \quad (4.5)$$

A total of  $T + 8$  values of  $\varepsilon_t$  were generated. The first eight of these values were dropped so that the first four values, generated by step (2), didn't excessively influence the sample (Spitzer, 1979, p.46). This left us with  $T$  values of  $\varepsilon_t$  that possess the SAR(4) process shown in equation (4.2).

(4) Using  $\beta_1=0$ , the respective  $\beta_2$ ,  $X_t$ , and the  $\varepsilon_t$  generated above, the dependent variable  $\hat{Y}_t$ , was generated as follows

$$\hat{Y}_t = \beta_1 + \beta_2 X_t + \varepsilon_t. \quad (4.6)$$

An OLS regression was then performed with direct labor personnel as the independent variable and  $Y_t$  the dependent variable. The residuals from this regression were then used to compute the nine estimators of  $\rho_4$ . As in Park and Mitchell's 1979 analysis any estimator that equaled or exceeded 1.0 was reset to .99999. For the two iterative algorithms, the iterative Prais-Winsten and the maximum likelihood, if two consecutive estimates exceeded 1.0 they were both reset to .99999 and convergence was declared. The results for these two estimators are only for the cases when convergence was attained.

## E. MEASURES OF EFFECTIVENESS

Three MOEs were used to determine the relative performances of the estimators, the RMSE of  $\rho_4$ , the RMSE of the slope coefficient,  $\beta_2$ , and the adjusted R-squared value. Each MOE was computed and averaged over the one hundred replications for each value of  $\rho_4$ .

To evaluate how accurately each estimator predicted  $\rho_4$  and its variance the RMSE of  $\rho_4$  was computed (Rao and Griliches, 1969). The equation is

$$\text{RMSE} (\rho_4) = \left[ \sum_{i=1}^{100} (\hat{\rho}_4 - \rho_4)^2 / 100 \right]^{1/2}. \quad (4.7)$$

The RMSE of  $\beta_2$  evaluated the estimators in terms of their performance for the slope coefficient,  $\beta_2$ . To make comparisons easier, the performance of each estimator (EGLS solution) relative to that of the OLS solution was computed. As presented in Park and Mitchell (1980) the relative efficiency is

$$\text{Rel. Eff.} (\beta_2, \text{estimator}_i) = \frac{\text{RMSE} (\beta_2, \text{OLS})}{\text{RMSE} (\beta_2, \text{EGLS}_i)}, \quad (4.8)$$

where

$$\text{RMSE} (\beta_2) = \left[ \sum_{i=1}^{100} (\hat{\beta}_2 - \beta_2)^2 / 100 \right]^{1/2}. \quad (4.9)$$

The last MOE utilized was the adjusted R-squared value from each EGLS regression. Though three MOE's were utilized, the RMSE of  $\rho_4$  was considered the most important. The other two were considered only if the RMSE of  $\rho_4$  was sufficiently close for alternative estimators.



## F. RESULTS

### 1. General

The results for the three simulations are summarized in Tables 2 through 4 and Figures 4.1a through 4.3c. The three different simulation runs are identified by the contractor label. Since each simulation was run with a different independent variable, each having its own unique structure, the results vary slightly between simulations.

The two iterative procedures had slight convergence problems at low values of  $\rho_4$  ( $\rho_4 \leq .3$ ). This could possibly have occurred because at low values of  $\rho_4$  the error terms,  $\varepsilon_t$ , still do not have a significant SAR(4) process structure. So the first estimate could be a poor one and the iterative procedure could proceed in the wrong direction (most likely toward negative values of  $\rho_4$ ). The convergence problem decreased as  $\rho_4$  increased, such that by  $\rho_4 = .7$  convergence occurred almost 100 percent of the time. For both estimators convergence generally occurred in four to six iterations.

### 2. Estimation of Rho

The RMSE of  $\rho_4$  was used to determine which estimators provided the best estimate of  $\rho_4$ . As can be seen in Table 2 and Figures 4.1a through 4.1c, no estimator was uniformly the best. Estimators three and four, the two estimators that utilized the Wallis test statistic, appeared to be vastly superior over the entire range of  $\rho_4$ . They were only outperformed at the extreme low end by estimators one and two, the two versions of the sample autocorrelation coefficient. Estimator three was the best in the range  $.2 \leq \rho_4 \leq .5$ . A crossover occurred at  $\rho_4$  equals  $.6$  and estimator four dominated for the upper range of  $\rho_4$ . An exception to this was for the contractor A simulation where estimator nine was the best for  $\rho_4 = .95$ . Estimator nine, the maximum likelihood estimator, appeared to be the third most

TABLE 2  
RMSE OF RHO FOR THE THREE SIMULATION RUNS

p4	Contractor A Estimator								
	1	2	3	4	5	6	7	8	9
.1	.198	.191	.195	.2	.218	.226	.244	.193	.224
.2	.183	.18	.17	.175	.214	.2	.22	.2	.217
.3	.204	.203	.171	.175	.209	.209	.227	.213	.226
.4	.209	.212	.154	.157	.198	.207	.209	.198	.214
.5	.233	.24	.153	.154	.199	.214	.218	.209	.21
.6	.286	.295	.184	.181	.235	.259	.253	.228	.213
.7	.33	.342	.207	.203	.26	.288	.284	.254	.236
.8	.286	.302	.162	.16	.217	.235	.217	.193	.167
.9	.283	.305	.134	.129	.195	.213	.184	.163	.126
.95	.243	.27	.097	.092	.133	.149	.134	.112	.070

p4	Contractor B Estimator								
	1	2	3	4	5	6	7	8	9
.1	.212	.204	.215	.22	.242	.248	.256	.216	.242
.2	.224	.219	.204	.208	.249	.255	.257	.201	.229
.3	.241	.239	.184	.188	.257	.256	.26	.206	.217
.4	.285	.284	.216	.217	.271	.286	.305	.249	.244
.5	.278	.283	.193	.193	.265	.283	.278	.24	.246
.6	.264	.274	.164	.164	.213	.228	.234	.225	.217
.7	.324	.337	.196	.192	.252	.275	.287	.26	.233
.8	.302	.319	.172	.169	.222	.245	.236	.207	.18
.9	.29	.314	.141	.137	.203	.216	.192	.176	.15
.95	.254	.283	.105	.101	.144	.153	.148	.131	.106

p4	Contractor E Estimator								
	1	2	3	4	5	6	7	8	9
.1	.21	.202	.228	.234	.265	.252	.263	.236	.264
.2	.213	.208	.213	.219	.263	.245	.253	.224	.253
.3	.23	.228	.185	.191	.253	.244	.258	.226	.24
.4	.271	.272	.204	.207	.286	.275	.292	.261	.268
.5	.26	.265	.174	.176	.269	.261	.262	.229	.237
.6	.252	.263	.154	.156	.212	.214	.22	.215	.216
.7	.314	.327	.174	.171	.254	.261	.27	.256	.23
.8	.302	.32	.159	.156	.228	.246	.229	.205	.179
.9	.286	.311	.125	.12	.196	.206	.169	.152	.12
.95	.259	.288	.105	.101	.15	.156	.148	.13	.106

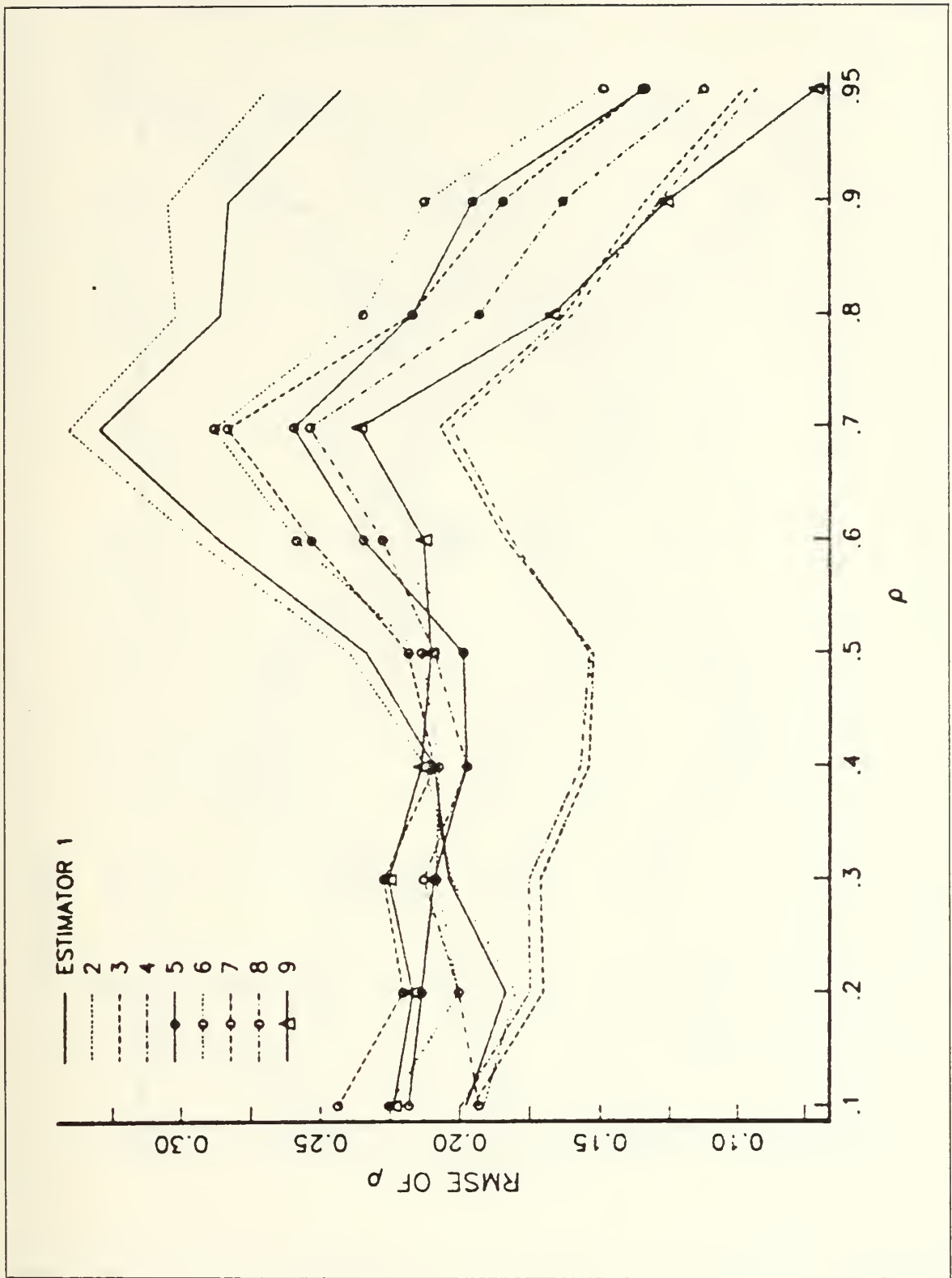


Figure 4.1a RMSE of Rho for Simulation for Contractor A.

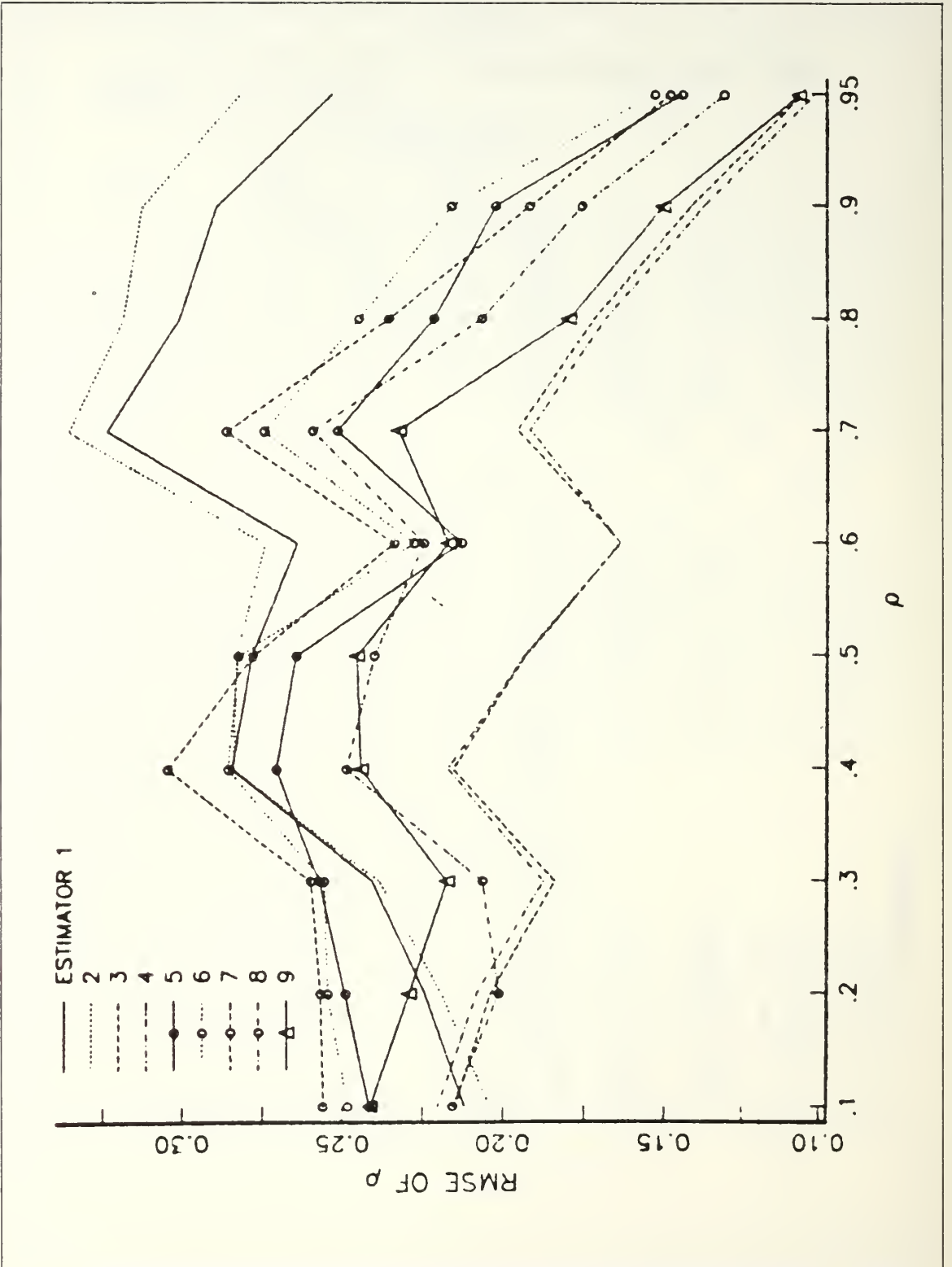


Figure 4.1b RMSE of Rho for Simulation for Contractor B.

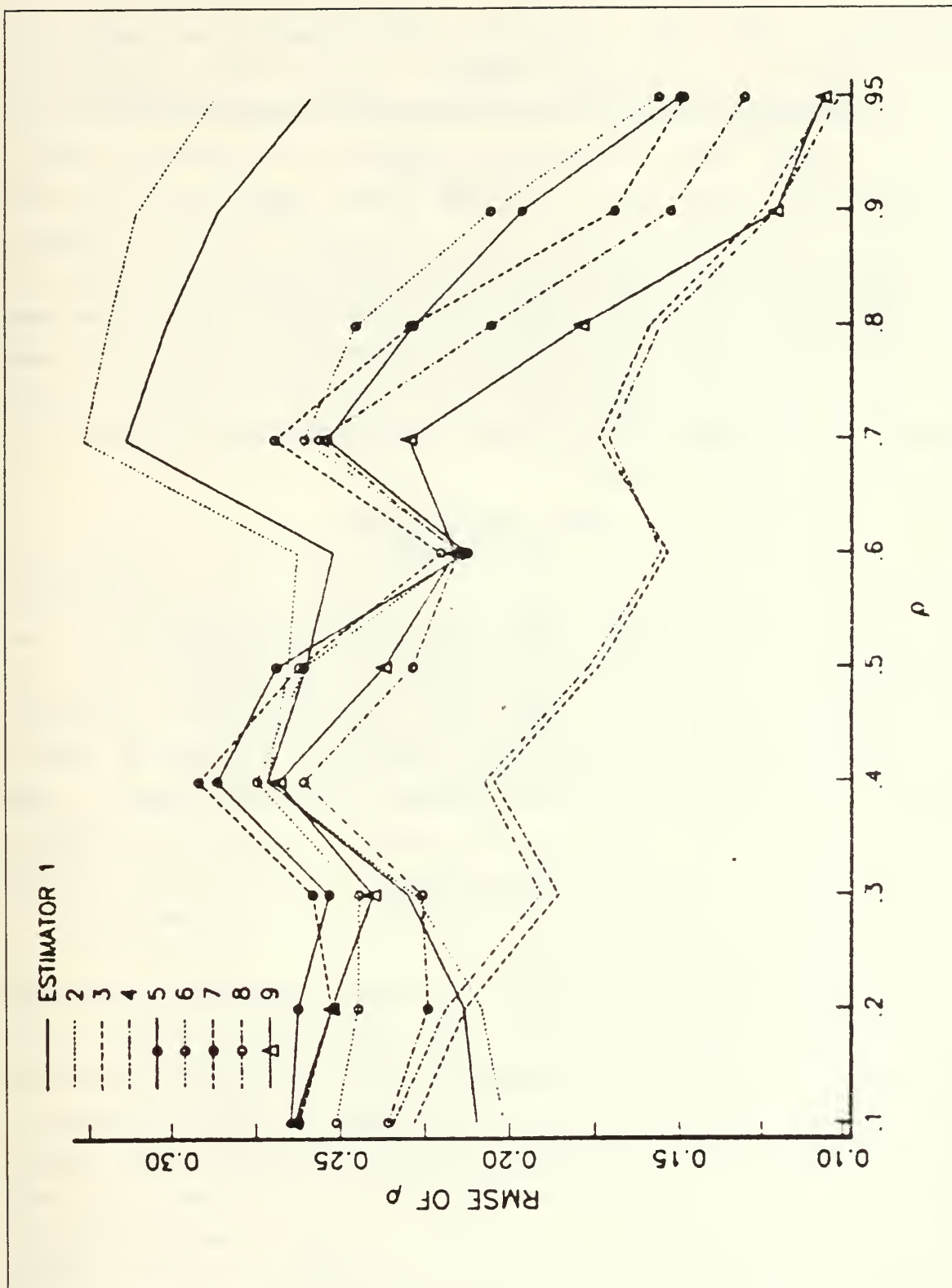


Figure 4.1c RMSE of Rho for Simulation for Contractor E.

efficient, with the iterative Prais-Winsten estimator a close fourth. It was interesting to compare the performance of the standard Prais-Winsten estimator, estimator one, with that of Park and Mitchell's version, estimator seven. Estimator one was better in the lower range ( $\rho_4 < .5$ ) with estimator seven better from that point on. Overall though, it appeared that estimator seven was the better of the two. It could also be observed that all estimators except one and two improved as  $\rho_4$  increase. Estimators one and two performed better at the lower end ( $\rho_4 = .1$ ). Recall that these two estimators are known to be downwardly biased in their AR(1) process forms.

### 3. Estimation of Slope Coefficient

Most noteworthy was the fact that all estimators provided more efficient estimation than did the OLS solution for  $\rho_4 > .2$ . As can be seen in Table 3 and Figures 4.2a through 4.2c, none of the estimators dominated over any significant range of  $\rho_4$ . The two best performers appeared to be estimators eight and nine. These two estimators possessed a slight edge over estimators three and four. The performances of the remaining estimators, except for estimators one and two which were clearly inferior, were very comparable and no significant distinctions could be drawn.

### 4. EGLS Regression Quality

The adjusted R-squared value was selected as the MOE to evaluate the quality of the EGLS regression. Again, as is shown in Table 4, all estimators provided better results than did the OLS solution. As can also be seen in Figures 4.3a through 4.3c estimators three and four were again vastly superior. This time estimator four performed the best until it was surpassed by estimator three at approximately  $\rho_4$  equal .8. Estimator nine again appeared to be the third best performer. Except for estimators one and two, which were again slightly inferior, the remaining estimators were

TABLE 3  
RELATIVE EFFICIENCIES FOR THE THREE SIMULATION RUNS

p4	Contractor A Estimator								
	1	2	3	4	5	6	7	8	9
.1	1.05	1.05	1.02	1.02	1.06	1.07	1.08	1.02	1.02
.2	1.07	1.07	1.02	1.02	1.04	1.08	1.07	1.08	1.07
.3	1.05	1.04	1.08	1.08	1.06	1.05	1.05	1.03	1.01
.4	1.1	1.09	1.08	1.08	1.11	1.11	1.1	1.09	1.08
.5	1.14	1.13	1.21	1.21	1.17	1.18	1.18	1.19	1.22
.6	1.25	1.24	1.33	1.33	1.33	1.28	1.29	1.35	1.35
.7	1.19	1.18	1.21	1.21	1.19	1.2	1.19	1.2	1.19
.8	1.31	1.3	1.37	1.37	1.37	1.34	1.36	1.38	1.38
.9	1.56	1.54	1.69	1.7	1.71	1.66	1.67	1.71	1.74
.95	1.38	1.37	1.41	1.41	1.41	1.4	1.41	1.42	1.43

p4	Contractor B Estimator								
	1	2	3	4	5	6	7	8	9
.1	0.93	0.93	0.97	0.97	0.96	0.95	0.98	0.98	0.97
.2	0.94	0.94	0.95	0.96	0.96	0.99	0.97	1.07	1.08
.3	1.03	1.03	1.01	1.01	1.03	1.04	1.04	1.05	1.04
.4	1.05	1.05	1	1	1.01	1.04	1.03	1.02	1.01
.5	1.02	1.02	1.04	1.04	1.04	1.02	1.02	1.06	1.05
.6	1.06	1.06	1.08	1.08	1.06	1.07	1.08	1.07	1.08
.7	1.03	1.03	1.06	1.06	1.06	1.04	1.03	1.08	1.09
.8	1.05	1.05	1.04	1.03	1.04	1.05	1.03	1.02	1.03
.9	1.12	1.12	1.14	1.14	1.15	1.15	1.12	1.13	1.14
.95	1.16	1.15	1.18	1.18	1.17	1.17	1.18	1.18	1.18

p4	Contractor E Estimator								
	1	2	3	4	5	6	7	8	9
.1	1.02	1.02	0.99	0.98	1.02	1.02	1.01	1.09	1.08
.2	1.06	1.06	1.01	1.01	1.06	1.07	1.07	1.04	1.04
.3	1	1	1	1	1.02	1.01	0.99	1.1	1.11
.4	1.03	1.02	1.08	1.08	1.07	1.04	1.04	1.06	1.07
.5	1.06	1.06	1.15	1.15	1.11	1.07	1.07	1.11	1.12
.6	1.13	1.13	1.15	1.16	1.15	1.14	1.14	1.15	1.15
.7	1.12	1.11	1.18	1.18	1.15	1.14	1.15	1.16	1.17
.8	1.1	1.1	1.12	1.12	1.12	1.1	1.1	1.09	1.09
.9	1.27	1.27	1.27	1.27	1.27	1.27	1.28	1.28	1.27
.95	1.37	1.37	1.36	1.36	1.36	1.36	1.36	1.36	1.36

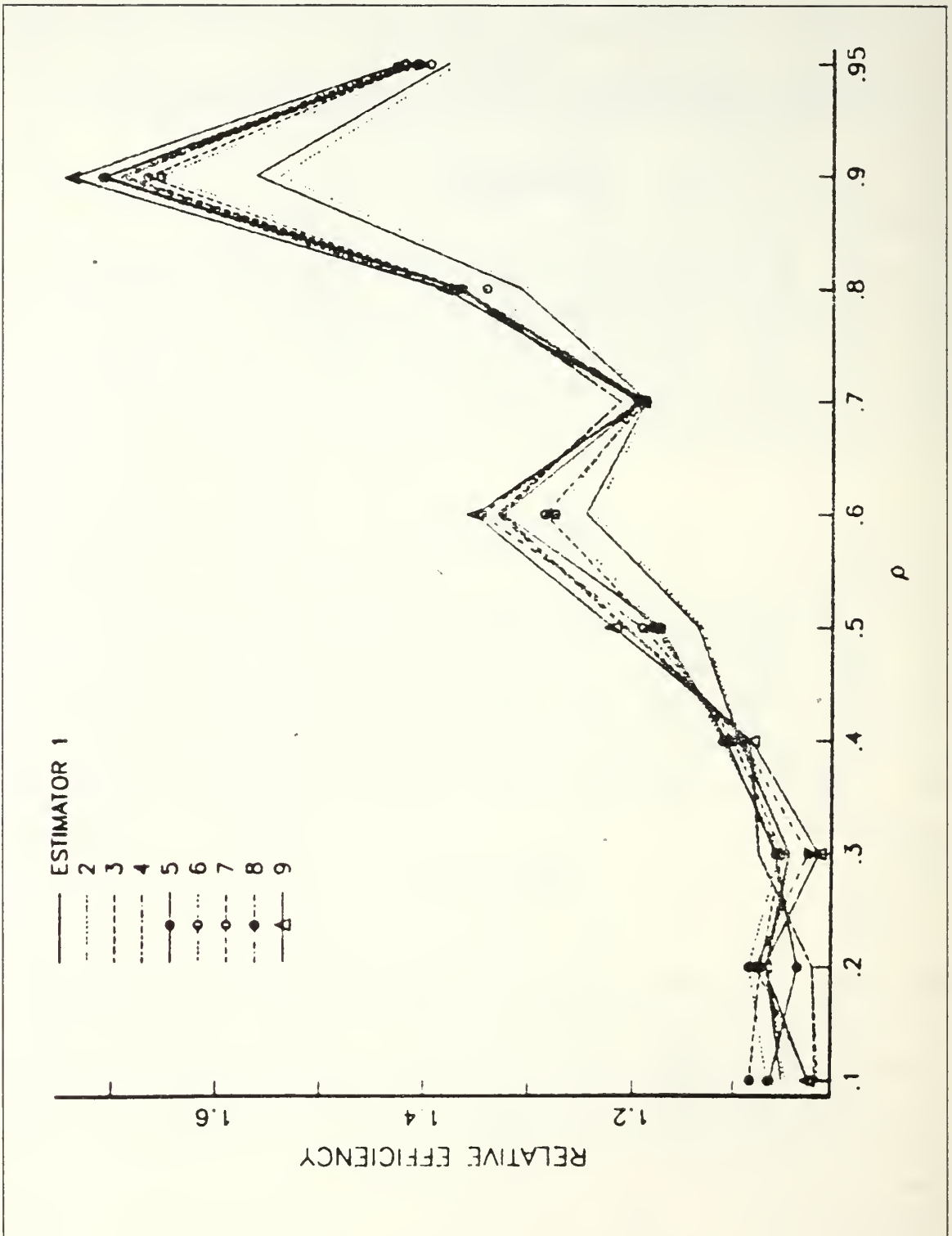


Figure 4.2a Relative Efficiencies  
for Simulation for Contractor A.



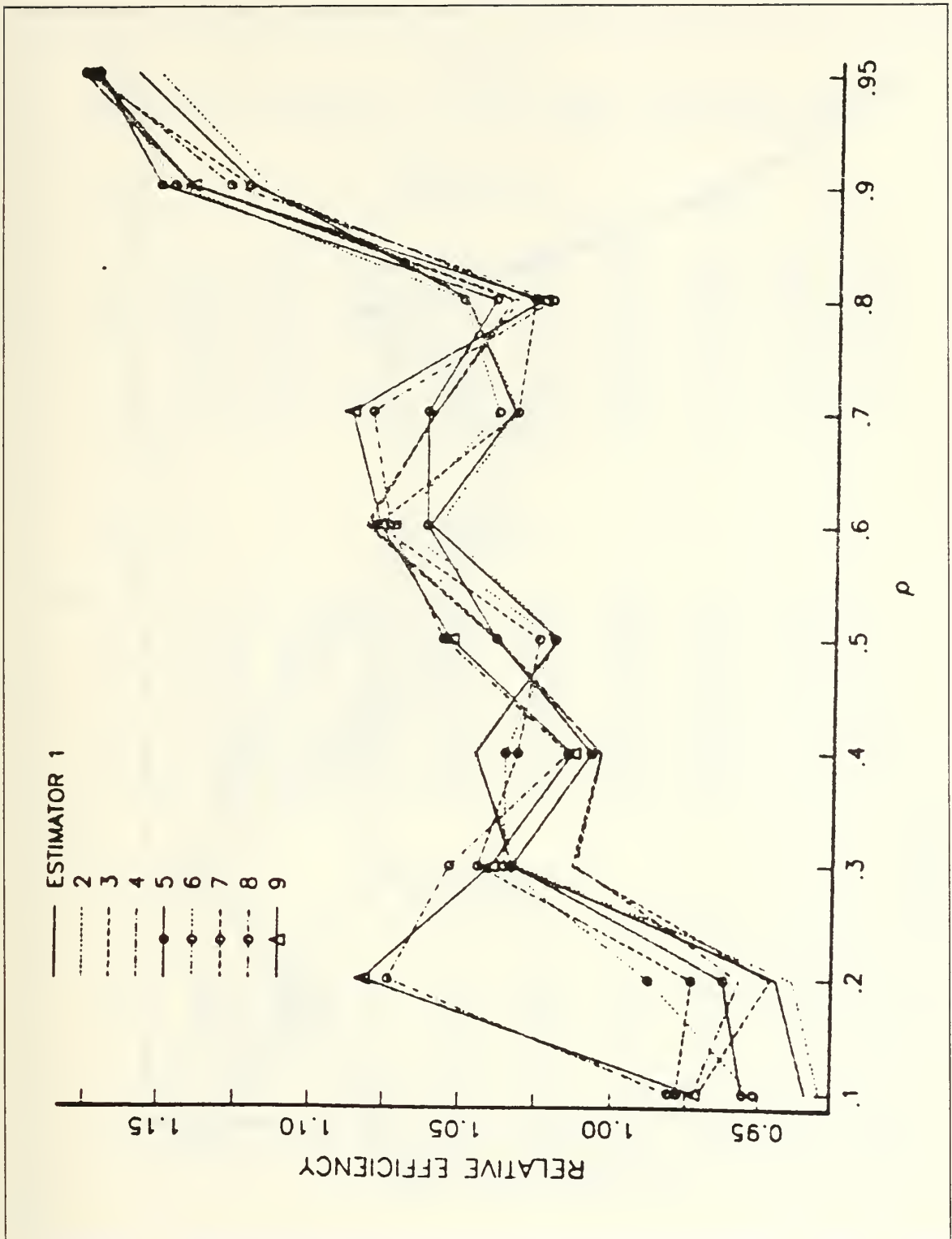


Figure 4.2b Relative Efficiencies  
for Simulation for Contractor B.

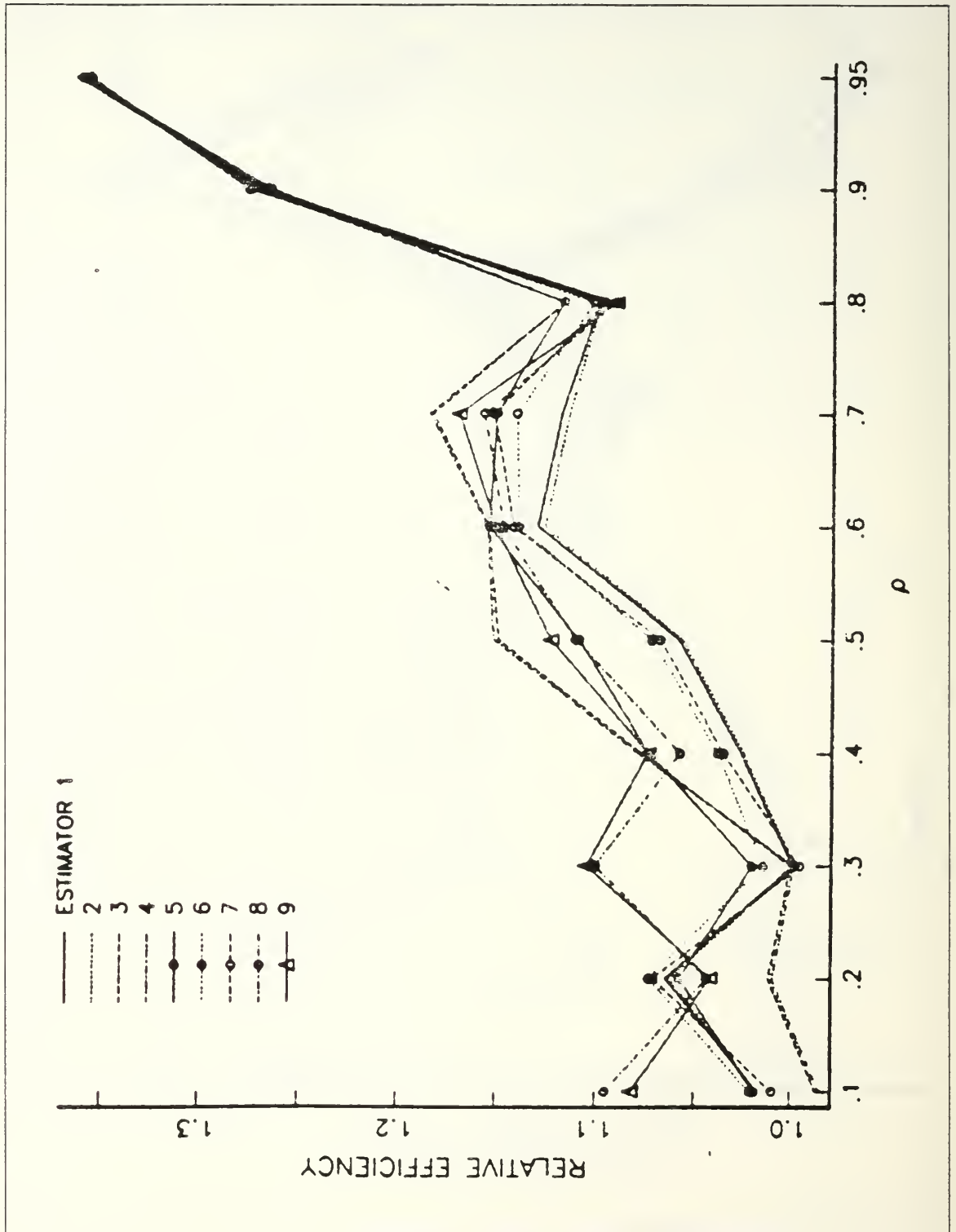


Figure 4.2c Relative Efficiencies  
for Simulation for Contractor E.

TABLE 4  
ADJUSTED R-SQUARED FOR THE THREE SIMULATION RUNS

$\rho_4$	Contractor A Estimator								
	1	2	3	4	5	6	7	8	9
.1	.41	.397	.509	.514	.443	.421	.422	.451	.473
.2	.433	.426	.596	.605	.519	.46	.466	.505	.53
.3	.521	.512	.669	.677	.59	.556	.569	.593	.615
.4	.604	.594	.745	.751	.701	.651	.655	.691	.721
.5	.644	.633	.764	.769	.716	.689	.702	.717	.741
.6	.667	.654	.771	.775	.741	.709	.721	.742	.768
.7	.705	.693	.795	.797	.758	.739	.744	.763	.777
.8	.765	.755	.805	.802	.762	.764	.767	.755	.769
.9	.768	.757	.779	.766	.703	.719	.7	.68	.697
.95	.751	.737	.722	.685	.656	.687	.592	.581	.6

$\rho_4$	Contractor B Estimator								
	1	2	3	4	5	6	7	8	9
.1	.574	.569	.611	.617	.609	.589	.594	.599	.62
.2	.563	.559	.693	.699	.609	.582	.588	.599	.62
.3	.624	.614	.755	.762	.705	.666	.68	.711	.74
.4	.658	.648	.816	.822	.742	.71	.71	.731	.766
.5	.717	.706	.847	.853	.772	.745	.761	.788	.808
.6	.804	.793	.891	.894	.849	.847	.855	.86	.876
.7	.819	.808	.901	.902	.864	.848	.821	.83	.875
.8	.865	.855	.908	.905	.843	.866	.847	.837	.881
.9	.882	.873	.885	.865	.762	.818	.703	.677	.798
.95	.873	.863	.838	.756	.666	.709	.572	.523	.683

$\rho_4$	Contractor E Estimator								
	1	2	3	4	5	6	7	8	9
.1	.648	.645	.691	.695	.662	.657	.666	.667	.676
.2	.658	.655	.735	.74	.682	.675	.68	.683	.699
.3	.671	.665	.75	.754	.694	.69	.707	.718	.731
.4	.692	.686	.792	.796	.73	.729	.733	.74	.764
.5	.725	.717	.822	.825	.766	.759	.771	.78	.806
.6	.76	.75	.841	.844	.81	.8	.813	.817	.835
.7	.77	.76	.845	.847	.814	.806	.805	.809	.83
.8	.801	.789	.854	.854	.827	.828	.833	.832	.842
.9	.801	.786	.842	.837	.809	.819	.804	.795	.814
.95	.789	.772	.822	.81	.796	.803	.784	.779	.786

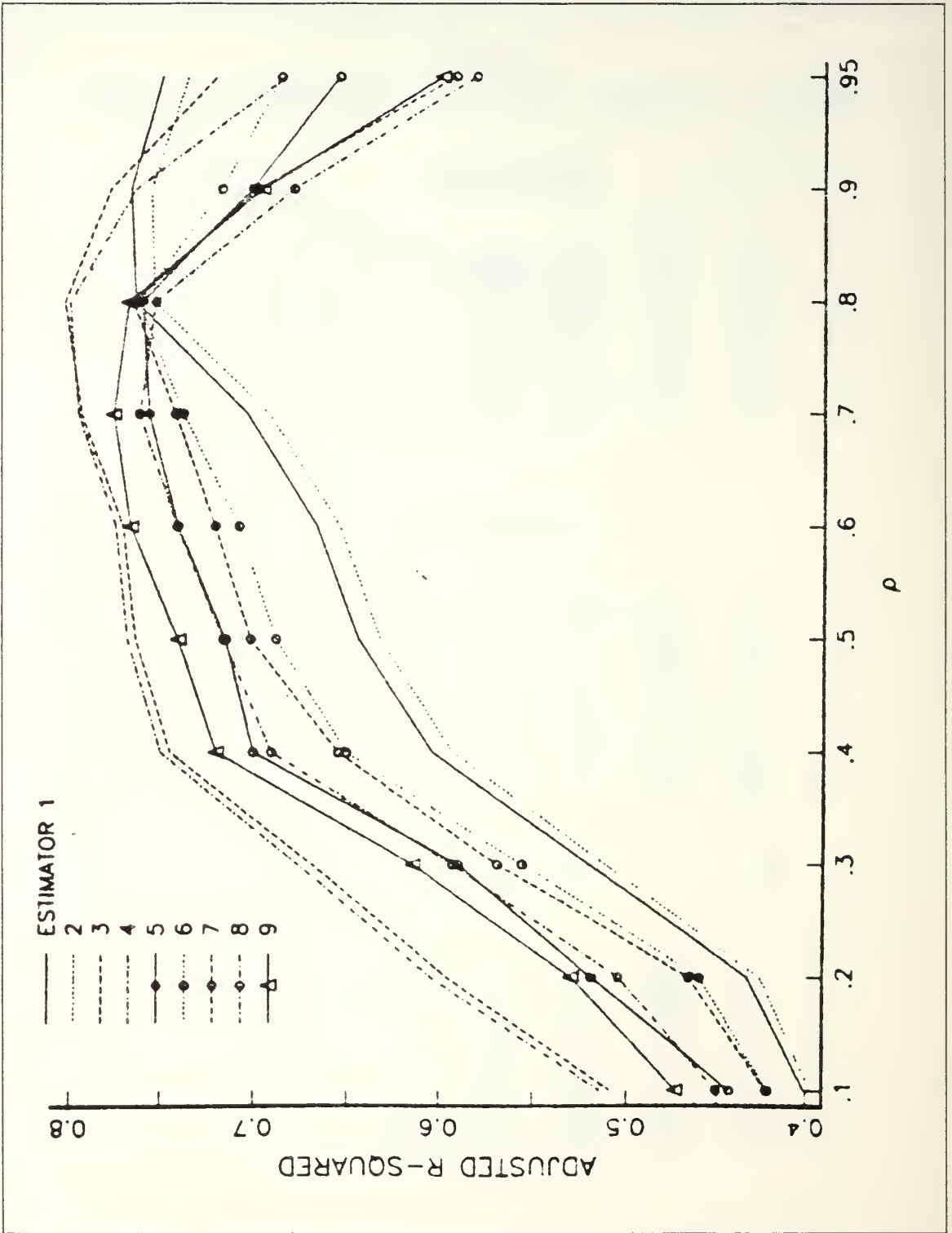


Figure 4.3a Adjusted R-squared for Simulation for Contractor A.

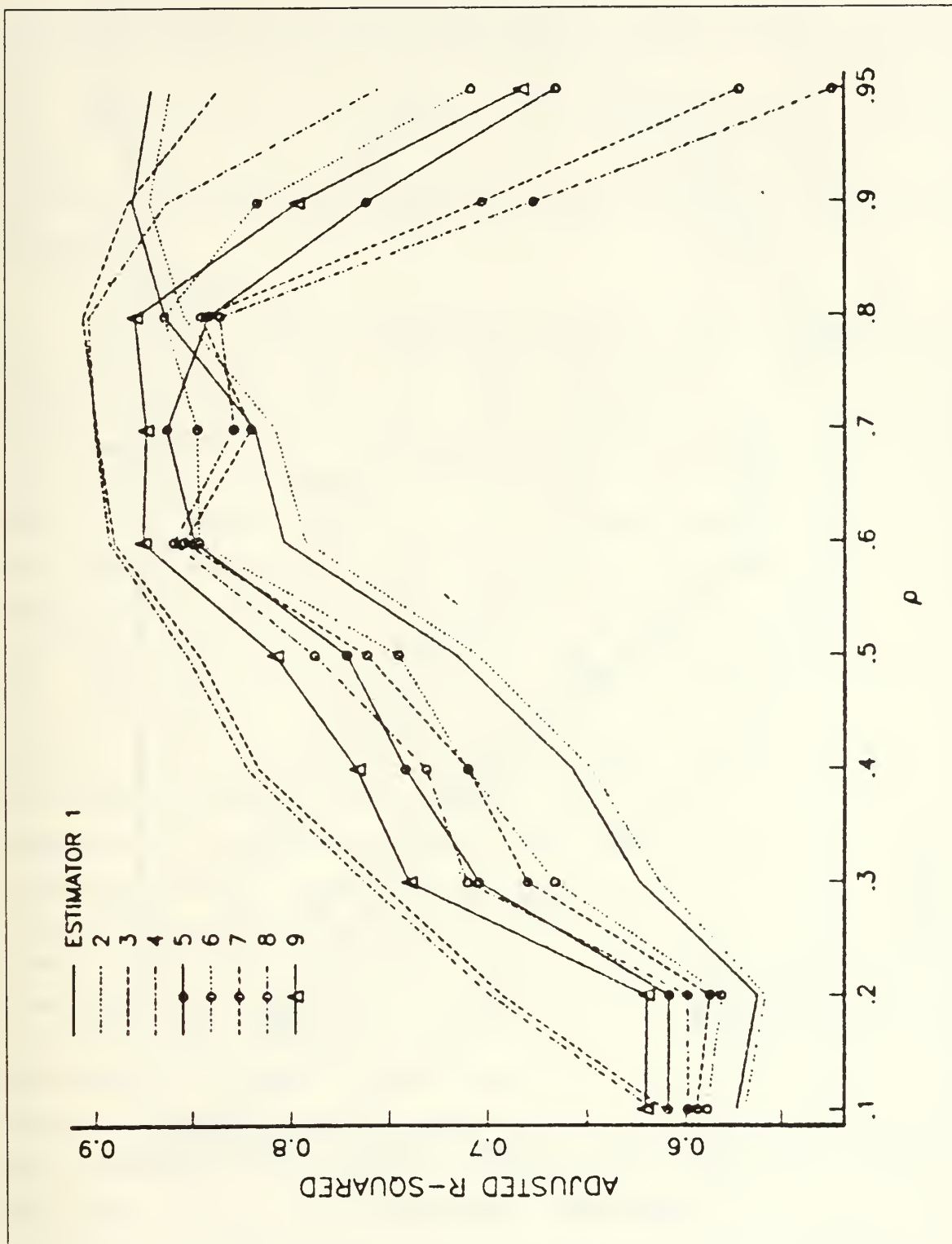


Figure 4.3b Adjusted R-squared for Simulation for Contractor B.

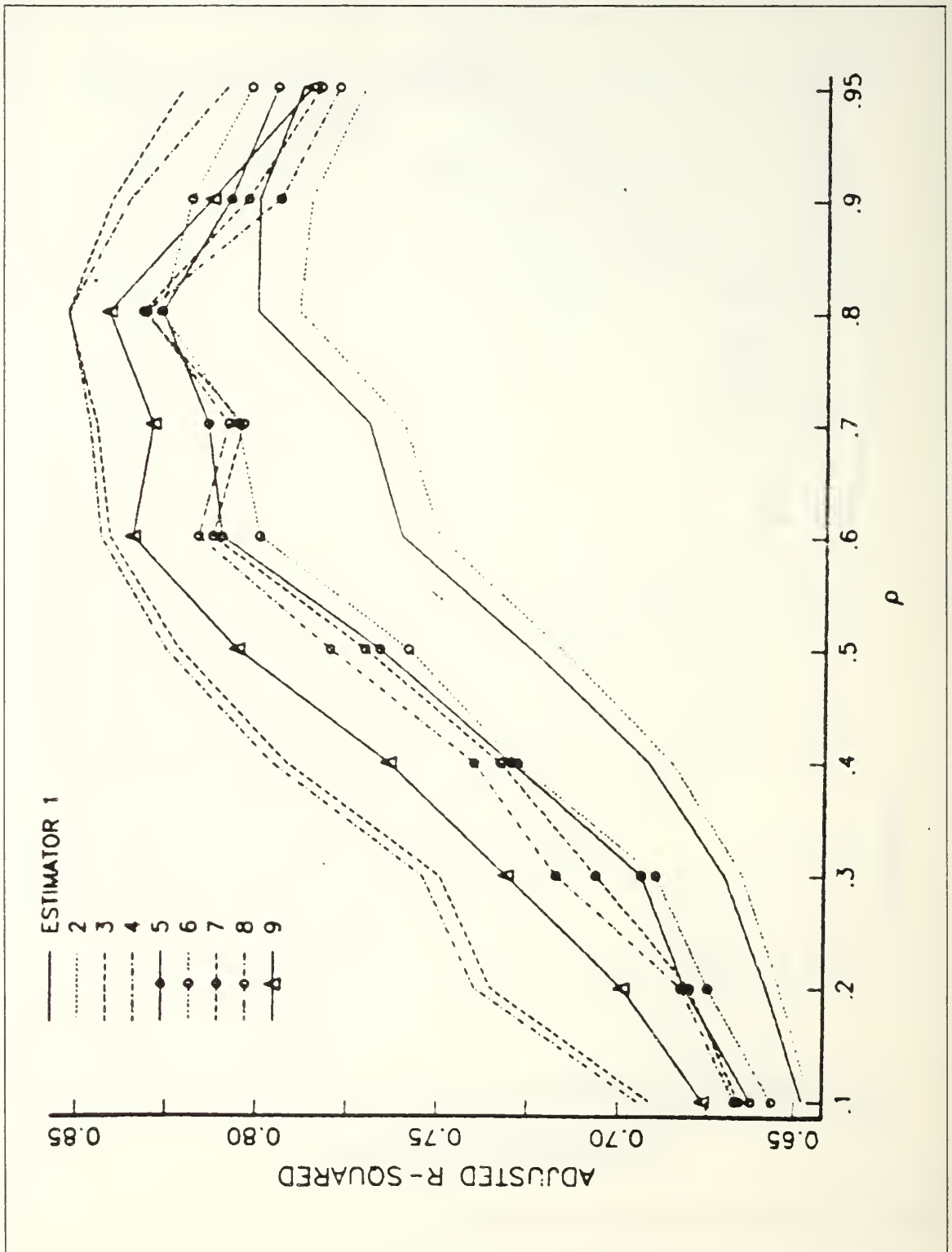


Figure 4.3c Adjusted R-squared for Simulation for Contractor E.

all relatively comparable in their overall performance. Estimators one and two were the best at the extreme high end ( $\rho_4 = .95$ ) in two of the simulations however. In comparing the two Prais-Winsten estimators, estimator seven performed better for  $\rho_4 \leq .7$ , with estimator one better for  $\rho_4 > .7$ . Again estimator seven appeared to be the better of the two estimators.

## 5. Summary

All of these estimators provided an improvement over the OLS solution when the SAR(4) process was present in a significant amount ( $\rho_4 > .2$ ). This mirrors the results of the previous Monte Carlo analyses for the AR(1) process cited earlier in the chapter. This simulation indicated that estimators three and four were the best estimators of  $\rho_4$ . Different results may be obtained for differently structured independent variables. The performances of estimators three and four were nearly identical. The selection of one over the other may be based upon the smoothness of the explanatory variable or the approximate value or range of  $\rho_4$ , if known. It was expected beforehand that the maximum likelihood and iterative Prais-Winsten estimators, the two most time costly estimators, would have outperformed all of the other estimators. They did in fact finish third and fourth, with estimator nine proving to be slightly better than estimator eight. It is worth noting again that none of the previous AR(1) process studies compared the AR(1) process versions of estimators three and four. Therefore it was believed that, like all of the other two-stage estimators, they would have been outperformed by the maximum likelihood and iterative Prais-Winsten estimators, as was shown to be the case in the previous studies. The result that estimator nine was better than estimator eight was the opposite of that obtained in the previous AR(1) process studies, but in both cases their performances were very close. As in the

previous studies it was proven that iteration leads to a more efficient estimator, as the iterative Prais-Winsten estimator was again shown to be better than its two-stage counterparts. Of the two Prais-Winsten estimators, estimator seven (Park and Mitchell's version) appeared to be better overall. This agreed with Park and Mitchell's findings in their analysis of first-order autocorrelation estimators. Knowledge of the approximate range of  $\rho_4$  beforehand would help in selecting the more appropriate of these two estimators. In the lower region estimator one was better, while estimator seven dominated in the upper region. The next chapter presents and compares the final regression models obtained for the three estimators, numbers three, four, and nine.



## V. STRUCTURAL ANALYSIS

### A. GENERAL

In this chapter the procedures outlined in chapter three were followed to obtain EGLS regression models for each contractor. A separate model was obtained using each of the three preferred estimators from the previous chapter, estimators three, four, and nine. All of the models are presented for comparison. Due to the large number of models obtained the entire procedure is illustrated in detail for only one, contractor A's EGLS model for estimator three. Only the final results of the other models are presented. In all cases direct labor personnel was utilized as the explanatory variable and total overhead costs as the dependent variable. The computer programs utilized in the structural analysis are contained in the appendix.

### B. PROCEDURE

Table 5 presents the results of these procedures applied to the regression of total overhead costs for contractor A (TOTOHA) on direct labor personnel for contractor A (DIRPERA). The results of this initial regression indicated very poor results. The adjusted R-squared value was very low and the F-statistic (not including constant term) was very close to its five percent critical value of 4.32 even though both were inflated due to the presence of autocorrelation. The low R-squared value indicated that the regression equation explained little beyond the mean of the dependent variable (Boger, 1983, p.21). Though biased downward, also due to the presence of autocorrelation, the standard errors of the regression coefficients were still large relative to the magnitude of the coefficients.

TABLE 5  
RESULTS FOR CONTRACTOR A

Model: TOTOHA = a + bDIRPERA

Untransformed Data

Standard Error of the Regression:	15000.	
Adjusted R-squared:	.1504	
F-statistic (degrees of freedom):	5.072	(1,22)
Estimate of a:	76270.	
Standard Error:	63930.	
Estimate of b:	12.06	
Standard Error:	5.355	
Durbin-Watson Test Statistic:	1.688	
Wallis Test Statistic:	.5480	
Estimator Three:	.7260	
Estimator Four:	.7381	
Estimator Nine:	.7843	

Transformed Data

Estimator Three

Standard Error of Regression:	9503.	
Adjusted R-squared:	.8660	
F-statistic:	149.7	
Estimate of a:	69090.	
Standard Error:	33980.	
Estimate of b:	12.18	
Standard Error:	2.787	

Estimator Four

Standard Error of Regression:	9468.	
Adjusted R-squared:	.8649	
F-statistic:	148.3	
Estimate of a:	68650.	
Standard Error:	33740.	
Estimate of b:	12.20	
Standard Error:	2.760	

Estimator Nine

Standard Error of Regression:	9361.	
Adjusted R-squared:	.8571	
F-statistic:	138.9	
Estimate of a:	67224.	
Standard Error:	33230.	
Estimate of b:	12.26	
Standard Error:	2.687	

The autocorrelation function of the residuals obtained from this regression (Figure 5.1), with its single significant spike at lag four, strongly suggested the presence of Wallis's SAR(4) process. Upon formally testing the residuals for the presence of this SAR(4) process, the null hypothesis that no fourth-order autocorrelation was present was clearly rejected. At this point the Durbin-Watson statistic was insignificant.

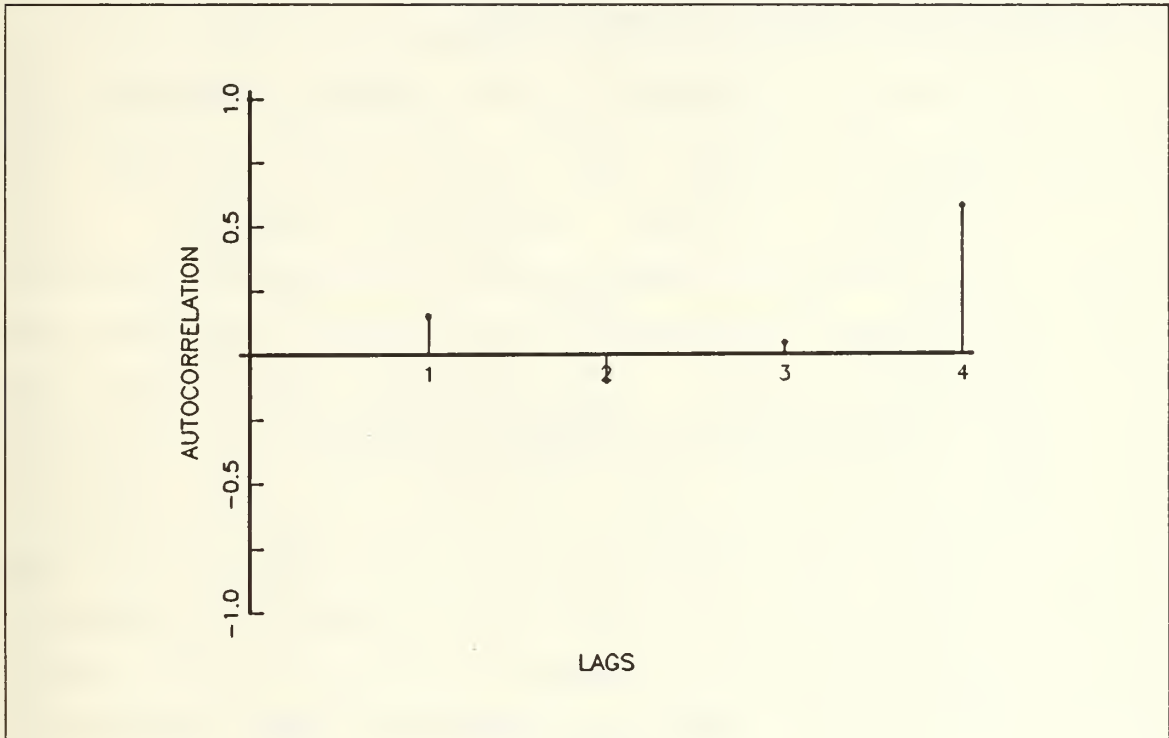


Figure 5.1 Autocorrelation Function of the Residuals for Contractor A.

As can be observed, the three calculated estimators of  $\rho_4$  were relatively close. The data were then transformed and the model reestimated. This step was performed three times in order to obtain a reestimated model for each of the estimators of  $\rho_4$ . The residuals of the new model were then analyzed for the presence of autocorrelation. In all three cases the Durbin-Watson statistic proved to be significant,

indicating the presence of first-order autocorrelation. The data were again transformed, this time to eliminate the AR(1) process (using the calculated estimator of  $\rho_1$ ), and the model was reestimated. Examination of the residuals from this model resulted in the finding that both the Durbin-Watson and Wallis test statistics were insignificant, indicating that neither the AR(1) nor SAR(4) processes were present in the residuals. The autocorrelation function of these residuals (Figure 5.2) also showed that no autoregressive process was present at any level.

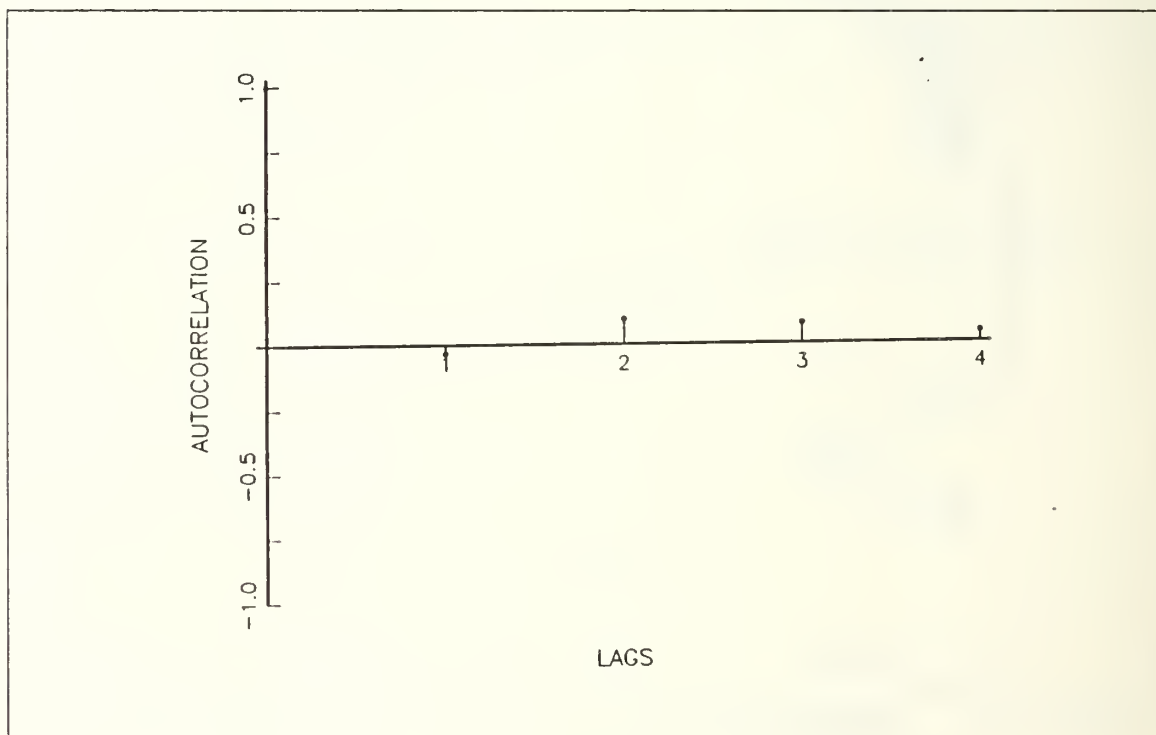


Figure 5.2 Autocorrelation Function of the Residuals for Contractor A's model (estimator 3).

The resulting model now had residuals that were free of autocorrelation. The residuals were then analyzed to determine if all of the assumptions required for the regression were satisfied. Note that this does not mean that it has been concluded that the assumptions are all necessarily

correct. It merely means that "on the basis of the data we have seen, we have no reason to say that they are incorrect" (Draper and Smith, 1981, p.142).

Three tests were performed to check the normality of the residuals. An empirical cumulative distribution function (CDF) was generated to compare the CDF of the residuals with that of the appropriate distribution. In this case the appropriate distribution was normal with mean zero and standard deviation 9099. A probability (Q-Q) plot of the residuals was also generated which plotted the quantiles of the residuals against the corresponding quantiles of the appropriate distribution. Each of these plots (Figure 5.3) were bounded by the ninety-five percent Kolmogorov-Smirnov (K-S) confidence boundaries. Both of these plots lie completely within the K-S boundaries, supporting the assumption that the residuals were distributed normally with mean zero and standard deviation 9099. To more formally test this hypothesis a Kolmogorov-Smirnov goodness-of-fit test was performed with null hypothesis  $H_0: F(x)=F^*(x)$ , versus  $H_1: F(x)\neq F^*(x)$ , where  $F^*(x)$  is the normal distribution with zero mean and standard deviation 9099 and  $F(x)$  is the unknown distribution function of the data. As presented in Conover (1980, p.347) the test statistic for the K-S test is simply the greatest distance between the cdf,  $F^*(x)$ , and empirical cdf of the data. It is possible to obtain this test statistic directly from the normal cdf plot in Figure 5.3. The test was performed using a significance level of size  $\alpha=.05$ . The K-S test statistic was significant at a level of .9693. Therefore, the null hypothesis could not be rejected.

Two tests were performed to test the constant variance (homogeneity of variance) assumption. The residuals were plotted against the predicted dependent variable (Figure 5.4) to see if any obvious abnormalities could be observed. The dispersion of the residuals appears to be fairly random

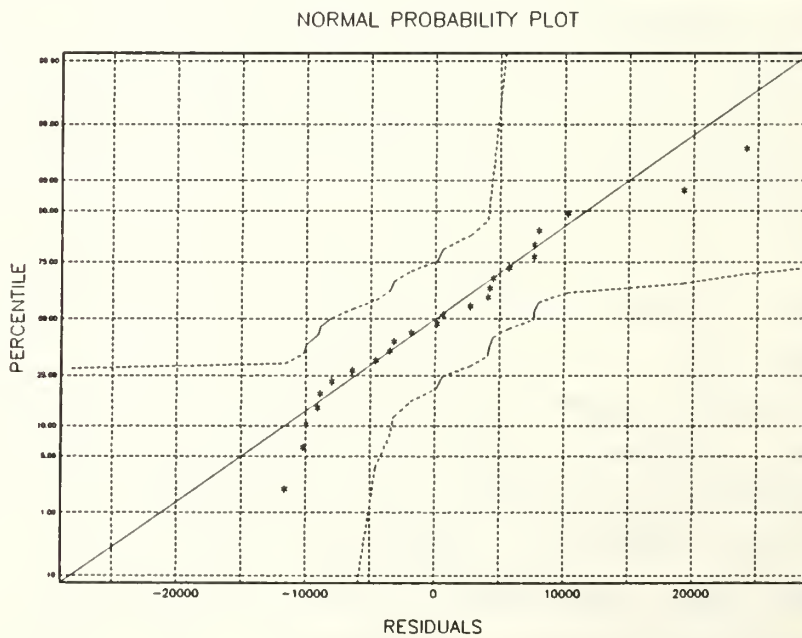
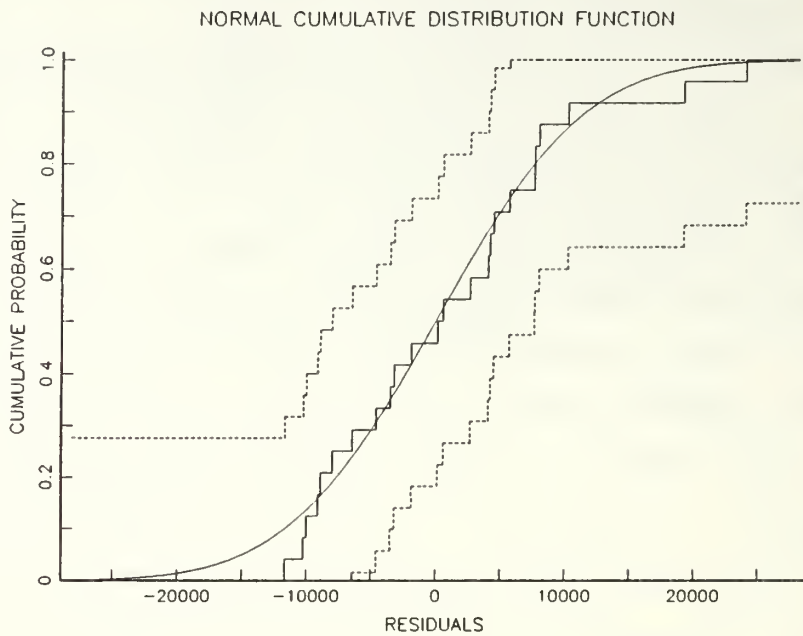


Figure 5.3 Tests for Normality of Residuals for Contractor A's model (estimator 3).

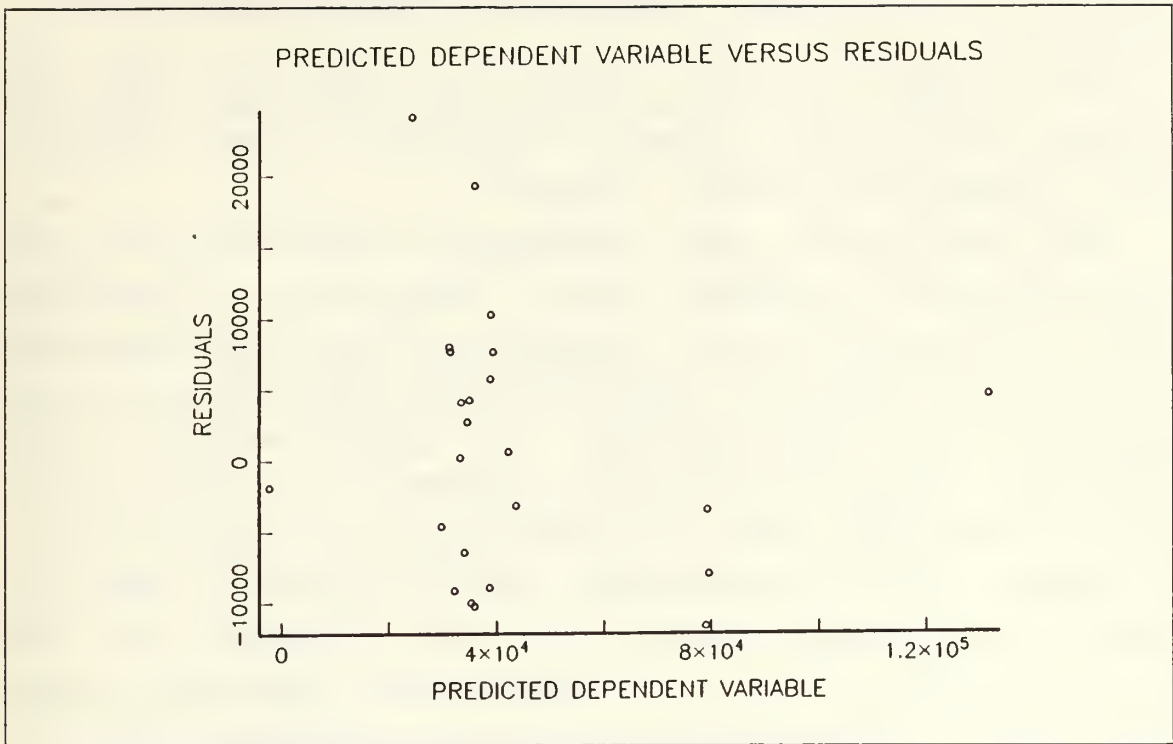


Figure 5.4 Tests for Homogeneity of Variance of the Residuals for Contractor A's model (estimator 3).

and the variance relatively constant throughout. An F test was then performed. The residuals were divided into two sets and the null hypothesis  $H_0: \sigma_1^2 = \sigma_2^2$  was tested against  $H_1: \sigma_1^2 \neq \sigma_2^2$ , where  $\sigma_1^2$  and  $\sigma_2^2$  were the variances of the two sets of residuals. As presented in Mood, et.al.(1974, p.438) the test statistic for this F-test is

$$R = \frac{(n_2 - 1)\Sigma(X_{1i} - \bar{X}_1)^2}{(n_1 - 1)\Sigma(X_{2i} - \bar{X}_2)^2}, \quad (5.1)$$

where  $\bar{X}_1$  and  $\bar{X}_2$  are the population expected values. This test statistic has the F distribution with degrees of freedom  $n_1-1$  and  $n_2-1$  when  $\sigma_1^2 = \sigma_2^2$ . This test was also performed using a significance level of size  $\alpha = .05$ . The test statistic was significant at a level of .5154.

Therefore the constant variance assumption could not be rejected.

These tests could not reject the assumption that the residuals were in fact normally distributed random variables with zero mean and constant variance.

The results for this regression indicated that the model contained a great deal of information on overhead costs. The R-squared value was significantly high indicating that the model contained much more information than just the mean of the dependent variable. The standard errors of the regression coefficients, especially that of the slope term, were now relatively small in comparison to the coefficients. In summary, after transforming the data numerous times to eliminate all autocorrelation from the residuals, an EGLS regression model was obtained that yielded excellent, reliable results. This model for the total overhead costs for contractor A was

$$\text{TOTOHA} = 69090 + 12.18 \text{ DIRPERA.}$$

Since all costs were measured in thousands of dollars, the model may be interpreted as indicating that there is a fixed cost component of approximately \$69,090,000 to overhead costs (when a function of direct labor personnel) with an additional \$12,180 per direct labor personnel to total overhead costs (Boger, 1983, pp.24-25).

The results of the final EGLS models for contractor A using estimators four and nine are also contained in Table 5. All required assumptions about the residuals in these models also appeared to be valid. All of the models yielded excellent results that were very comparable. The slope terms of all of the models were within eight one-hundredths, a spread of approximately three one-hundredths of one standard deviation. The model for estimator three was slightly superior however.



This same general procedure was carried out for all of the remaining models, but only their final results are reported. The only difference in the procedure for the different models was in the order that the various autoregressive processes were removed from the residuals. The normal sequence was to remove the AR processes by order of significance. To get an initial overview of the type of autocorrelation that was present in the residuals of the initial models, the autocorrelation functions of the residuals (Figure 5.5a and 5.5b) were examined. This, combined with the results of the Durbin-Watson and Wallis tests, gave the AR process that the data were to be adjusted for first. The remaining order was then determined from the results of the Durbin-Watson and Wallis tests performed on the residuals of the previous model.

The autocorrelation functions showed that only the residuals for contractor A clearly appeared to possess the pattern expected to be exhibited by residuals containing Wallis's SAR(4) process. It was also clearly visible that the models for contractors C and D did not possess this SAR(4) process at a significant level. As with all of the contractors they will be examined more thoroughly later in this chapter. Despite the uncharacteristic appearance of their autocorrelation functions it appeared that Wallis's SAR(4) process was also the most significant form of autocorrelation present in the initial models for the remaining contractors (B, E, and F).

A number of factors led to this conclusion. First, the amount of data was very small relative to the amount required to obtain an accurate portrayal of the autoregressive process from the autocorrelation function. It should be noted that in the cases of contractors A, B, E and F the spike for the lagged four residuals was generally most significant. Secondly, the data were quarterly and as was

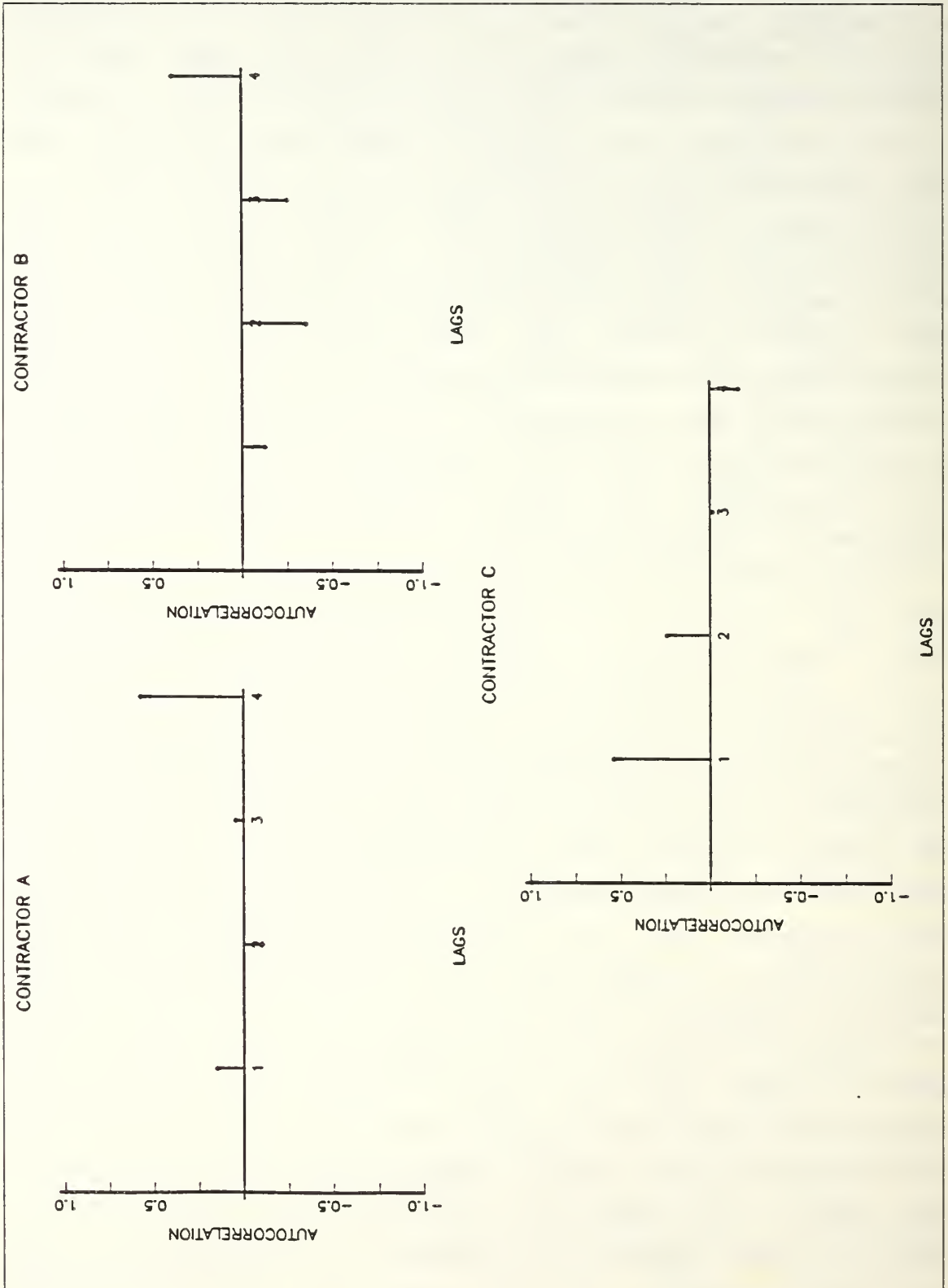


Figure 5.5a Autocorrelation Functions of the Residuals.

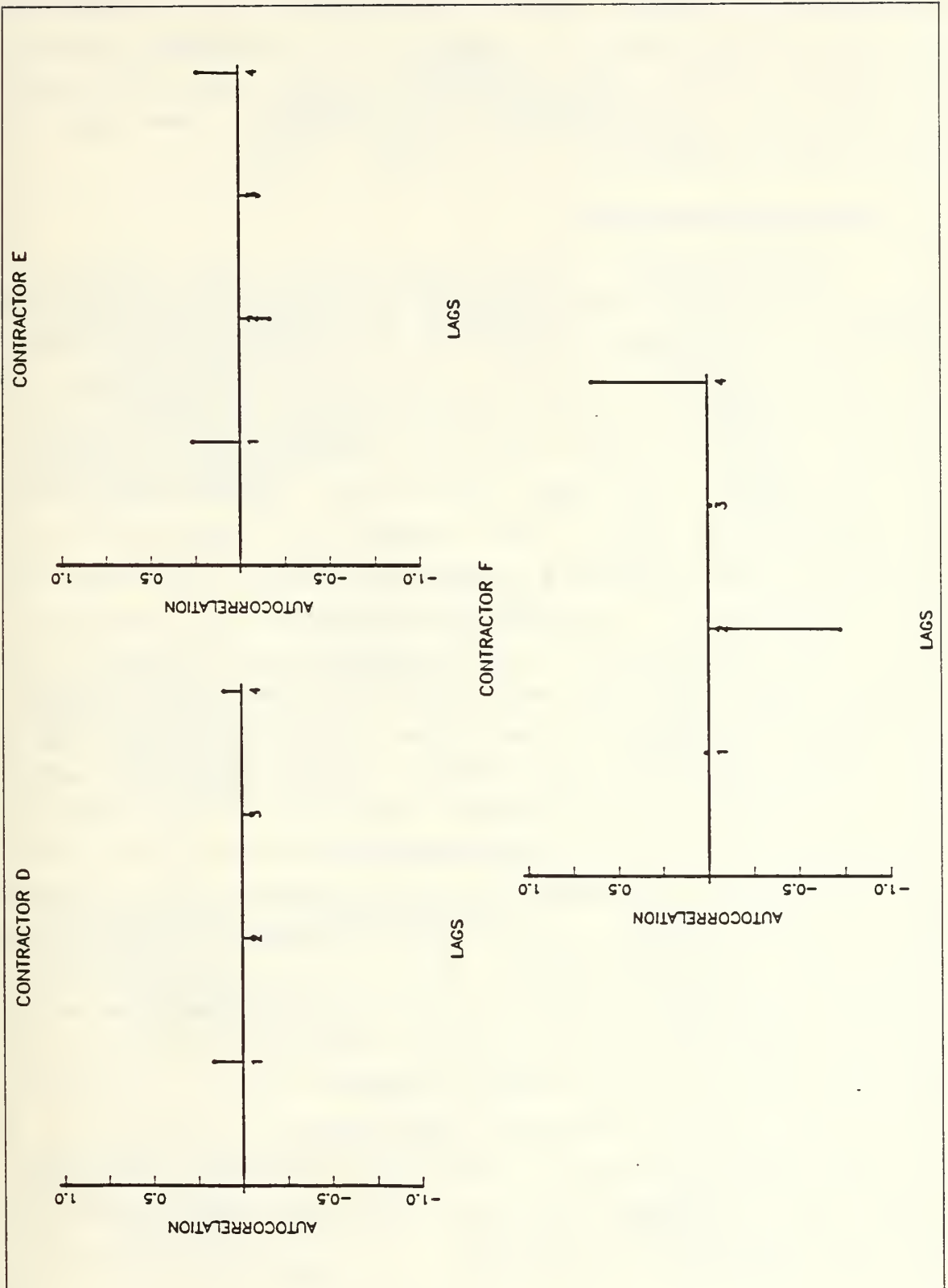


Figure 5.5b Autocorrelation Functions of the Residuals.

mentioned in chapter three, the plots of their dependent variables did appear to exhibit the seasonal pattern described by Wallis. Last, as will be shown, the Wallis test statistic was significant for the residuals from each of their initial models.

### C. THE REMAINING MODELS

#### 1. Contractor B

Table 6 contains the results for contractor B. The results of the initial regression of the untransformed data were fairly good. All of the statistical results, adjusted R-squared, standard error of regression, F-statistic, and standard errors of the regression coefficients, supported this conclusion. But these statistical results could have been misleading due to the presence of autocorrelation. Analysis of the residuals showed the Wallis test statistic to be significant, indicating the presence of the SAR(4), while the Durbin-Watson test statistic was shown to be insignificant. The data were then transformed and the model reestimated for each of the three estimators of  $\rho_4$ . Again all of the estimators were fairly close. These new models all provided excellent results. Analysis of the residuals showed no presence of any AR process, and none of the necessary assumptions pertaining to the residuals were violated in any of the models. Again the performance of all three models was very comparable, but the one obtained using estimator nine was slightly superior. It should be noted that all three models have negative intercepts. This implies negative fixed costs for those models. Though implausible, it is not totally infeasible. The models are being fit to data that are far from the Y-axis and, as with all models, these are only valid within the relevant range defined by the chosen explanatory variable.

TABLE 6  
RESULTS FOR CONTRACTOR B

Model: TOTOHB = a + bDIRPERB

Untransformed Data

Standard Error of the Regression:	9219.	
Adjusted R-squared:	.7657	
F-statistic (degrees of freedom):	69.61	(1,20)
Estimate of a:	-144300.	
Standard Error:	4450.	
Estimate of b:	27.98	
Standard Error:	3.354	
Durbin-Watson Test Statistic:	2.206	
Wallis Test Statistic:	.75	
Estimator Three:	.625	
Estimator Four:	.6385	
Estimator Nine:	.7256	

Transformed Data

Estimator Three

Standard Error of Regression:	7728.	
Adjusted R-squared:	.9219	
F-statistic:	248.7	
Estimate of a:	-110200.	
Standard Error:	50980.	
Estimate of b:	25.41	
Standard Error:	3.84	

Estimator Four

Standard Error of Regression:	7715.	
Adjusted R-squared:	.9235	
F-statistic:	254.3	
Estimate of a:	-109000.	
Standard Error:	51260.	
Estimate of b:	25.32	
Standard Error:	3.861	

Estimator Nine

Standard Error of Regression:	7651.	
Adjusted R-squared:	.9287	
F-statistic:	274.5	
Estimate of a:	-100500.	
Standard Error:	53240.	
Estimate of b:	24.68	
Standard Error:	4.006	

## 2. Contractor C

The results for contractor C are contained in Table 7. Very poor results were obtained for the initial regression. From the autocorrelation function of the residuals of this model (Figure 5.5a) it appeared that the AR(1) process was the most significant form of autocorrelation present. The Durbin-Watson test statistic supported this finding. It was significant, indicating the presence of first-order autocorrelation. As was expected, the Wallis test statistic was insignificant. The data were then transformed to eliminate the presence of the AR(1) process, and the model reestimated. Examination of the residuals of this model showed no autocorrelation present nor any required assumptions violated. The final model had been obtained without requiring a data transformation for the SAR(4) process. The results of this model, though inferior to the previous two, were fairly good.

## 3. Contractor D

Table 8 presents the results for contractor D. Again fairly poor results were obtained for the initial regression. From the autocorrelation function of the residuals of this model (Figure 5.5b) it appeared that no AR process was present in any significant amount. As expected, the Durbin-Watson and Wallis test statistics were both insignificant. So the final model, though rather poor, had been obtained without requiring any data transformations. Further analysis of the residuals indicated that again all necessary assumptions appeared to hold. The model for contractor D's total overhead costs is probably unreliable.

## 4. Contractor E

The results for contractor E are presented in Table 9. Again very poor results were obtained for the initial model. The sequence of steps for contractor E were the same as that for contractor A. First, it was necessary for the

TABLE 7  
RESULTS FOR CONTRACTOR C

Model: TOTOHC = a + bDIRPERC

		Untransformed Data
Standard Error of the Regression:		11260.
Adjusted R-squared:		.2123
F-statistic (degrees of freedom):		7.2 (1,22)
Estimate of a:		158100.
Standard Error:		12440.
Estimate of b:		2.148
Standard Error:		.8004
Durbin-Watson Test Statistic:		.7814
Wallis Test Statistic:		1.524
Estimator of $\rho_1$		.6263

Transformed Data

Estimator for  $\rho_1$

Standard Error of Regression:		9120.
Adjusted R-squared:		.6891
F-statistic:		51.97
Estimate of a:		149700.
Standard Error:		21420.
Estimate of b:		2.699
Standard Error:		1.389

data to be transformed to eliminate the presence of the SAR(4) process, and then the model was reestimated. As usual, a reestimated model was calculated for each of the three estimators of  $\rho_4$ . Estimators three and four were again very close while estimator nine was significantly larger. Next it was necessary (in all three cases) for the data to be transformed to eliminate the AR(1) process from the residuals. The EGLS regression of this transformed data proved to be the final reestimation required. Analysis of these residuals showed no presence of autocorrelation nor violation of any required assumptions. All three final models provided fairly good results and were again very comparable.

TABLE 8  
RESULTS FOR CONTRACTOR D

Model: TOTOHD = a + bDIRPERD

	Untransformed Data
Standard Error of the Regression:	10240.
Adjusted R-squared:	.5291
F-statistic (degrees of freedom):	26.84 (1,22)
Estimate of a:	5217.
Standard Error:	18400.
Estimate of b:	14.88
Standard Error:	2.873
Durbin-Watson Test Statistic:	1.450
Wallis Test Statistic:	1.366

The model for estimator nine was again slightly better, however.

#### 5. Contractor F

Table 10 presents the results for contractor F. This time the results for the OLS model were fairly good. But again, these statistical results could have been misleading due to the presence of autocorrelation. As with contractor E the data had to be transformed for the SAR(4) and then the AR(1) process, and the model reestimated before a final, "uncorrelated" model was obtained. Examination of these residuals showed that no type of autocorrelation was present and that all required assumptions appeared to hold. The results from all three final models were again very comparable, and provided very good results. The model for estimator three was slightly superior, however.

#### D. SUMMARY

Two things were performed in this chapter. First, regression models were obtained for each contractor that allowed for comparison of their total overhead costs.



TABLE 9  
RESULTS FOR CONTRACTOR E

Model: TOTOHE = a + bDIRPERE

Untransformed Data

Standard Error of the Regression:	3424.	
Adjusted R-squared:	.2427	
F-statistic (degrees of freedom):	7.73	(1,20)
Estimate of a:	27390.	
Standard Error:	8622.	
Estimate of b:	6.549	
Standard Error:	2.356	
Durbin-Watson Test Statistic:	1.043	
Wallis Test Statistic:	.8996	
Estimator Three:	.5502	
Estimator Four:	.5631	
Estimator Nine:	.7262	

Transformed Data

Estimator Three

Standard Error of Regression:	2897.	
Adjusted R-squared:	.7951	
F-statistic:	82.49	
Estimate of a:	19450.	
Standard Error:	12880.	
Estimate of b:	8.802	
Standard Error:	3.662	

Estimator Four

Standard Error of Regression:	2893.	
Adjusted R-squared:	.7969	
F-statistic:	83.39	
Estimate of a:	19240.	
Standard Error:	12790.	
Estimate of b:	8.864	
Standard Error:	3.637	

Estimator Nine

Standard Error of Regression:	2836.	
Adjusted R-squared:	.8022	
F-statistic:	86.15	
Estimate of a:	18580.	
Standard Error:	11820.	
Estimate of b:	9.08	
Standard Error:	3.358	

TABLE 10  
RESULTS FOR CONTRACTOR F

Model: TOTOHF = a + bDIRPERF

Untransformed Data

Standard Error of the Regression:	17470.	
Adjusted R-squared:	.6156	
F-statistic (degrees of freedom):	37.84	(1,22)
Estimate of a:	219600.	
Standard Error:	24470.	
Estimate of b:	10.14	
Standard Error:	1.648	
Durbin-Watson Test Statistic:	1.927	
Wallis Test Statistic:	.4977	
Estimator Three:	.7512	
Estimator Four:	.7634	
Estimator Nine:	.7872	

Transformed Data

Estimator Three

Standard Error of Regression:	9660.	
Adjusted R-squared:	.9664	
F-statistic:	663.3	
Estimate of a:	229700.	
Standard Error:	27490.	
Estimate of b:	9.175	
Standard Error:	1.847	

Estimator Four

Standard Error of Regression:	9660.	
Adjusted R-squared:	.9664	
F-statistic:	663.3	
Estimate of a:	229700.	
Standard Error:	27490.	
Estimate of b:	9.175	
Standard Error:	1.847	

Estimator Nine

Standard Error of Regression:	9641.	
Adjusted R-squared:	.9638	
F-statistic:	613.8	
Estimate of a:	227600.	
Standard Error:	28250.	
Estimate of b:	9.3	
Standard Error:	1.885	

Second, since final models were calculated for each of the three estimators of  $\rho_4$ , another means to compare the estimators was obtained.

The analysis indicated that excellent structural results were obtained for all but one contractor, contractor D. All of the others yielded reliable, useful regression models. For comparative purposes the model that provided the best results for each contractor is presented:

$$\begin{aligned} \text{TOTOHA} &= 69090 + 12.18 \text{ DIRPERA} \\ \text{TOTOHB} &= -100500 + 24.68 \text{ DIRPERB} \\ \text{TOTOHC} &= 149700 + 2.699 \text{ DIRPERC} \\ \text{TOTOHD} &= 5217 + 14.88 \text{ DIRPERD} \\ \text{TOTOHE} &= 18580 + 9.08 \text{ DIRPERE} \\ \text{TOTOHF} &= 229700 + 9.175 \text{ DIRPERF.} \end{aligned}$$

The model for contractor D was included in the comparison even though its results were unreliable.

It is possible to use these models to compare the overhead cost structures of the firms in the sample. The model for contractor E lies everywhere below the regressions for contractors A and F. In each case contractor E had both a significantly lower fixed cost and a lower (not significant) variable cost than the other contractor. Likewise contractor C's model was uniformly below that of contractors F. In this case though, the differences in the fixed and variable costs were both statistically significant. It should be noted that the comparisons imply only that, with the same number of direct personnel, one contractor experiences lower total overhead costs than another. They do not imply that the contractor possesses lower total overhead costs regardless of the circumstances (Boger, 1983, p.33).

For all relevant contractors the models for the three estimators were all very comparable. They were so comparable in fact, that no distinction could be drawn as to which

estimator was better. The superiority of estimators three and four over estimator nine in the Monte Carlo simulation was obviously not significant enough to manifest itself in this structural analysis.

## VI. CONCLUSIONS

The objectives of this thesis were twofold. The first objective was to determine the best estimators of autocorrelation found in the residuals. The second was to obtain simple and efficient regression models for overhead costs for the six aerospace contractors. The first objective was simply a means of achieving the second since the quality of the EGLS regression is dependent upon the quality of the estimator used.

Two methods were utilized to determine the best estimators of fourth-order autocorrelation, a Monte Carlo simulation and comparison of the EGLS regression models of the three preferred estimators (from the simulation). The results from the Monte Carlo simulation indicated estimators three and four, the two estimators that utilize the Wallis test statistic, to be the two best estimators of fourth-order autocorrelation. Their overall performances were nearly identical with preference determined by the value of  $\rho_4$ . Estimator three was superior to estimator four for  $\rho_4$  less than or equal to .5, while four was superior for  $\rho_4$  greater than .5. The maximum likelihood estimator, estimator nine, was shown to be the third best estimator. Another important result from the simulation was that all nine of the estimators provided an improvement over the OLS solution when the SAR(4) process was present. Comparison of the respective EGLS models showed no difference between the performances of the three estimators. The models were so comparable that no distinction could be drawn between the estimators.

Due to their superior performances in the Monte Carlo simulation, estimators three and four were chosen as the best estimators of fourth-order autocorrelation. However,

it may have appeared from the structural analysis that estimator nine was their equal. The Monte Carlo simulation was much more sensitive to slight differences in the performances of the estimators than the regression models. For this reason the simulation was the main criterion used in selecting the best estimator. Comparison of the regression models in the structural analysis would have been more useful if distinct disparities had arisen between the performances of the estimators.

From this analysis no distinct preference could be made between estimators three and four. Both estimators are fairly easy to calculate once the Wallis test statistic has been computed, so difficulty of computation can not be used to determine a preference between the two. The only criterion found to discriminate between the two is the amount of fourth-order autocorrelation present in the residuals (the value of  $\rho_4$ ) and the "smoothness" of the explanatory variables. If the value of  $\rho_4$  (if it can be speculated) is less than or equal to .5 then estimator three would be expected to perform better, while estimator four would be preferred for  $\rho_4$  greater than .5. Though not tested herein, if the explanatory variables are "smooth" then estimator four would again be expected to outperform estimator three. Though not as scrupulously investigated in this thesis, Park and Mitchell's version of the Prais-Winsten estimator was selected as the estimator of first-order autocorrelation. It was selected because it was shown to be the superior two-stage estimator for first-order autocorrelation in Park and Mitchell's 1980 study.

The structural analysis showed that excellent results could be obtained for four of the six aerospace contractors. The results of a fifth model were also more than adequate. These excellent results were only obtainable after the effects of autocorrelation were transformed out of the

dependent and independent variables. As was the case in Boger's 1984 analysis, the five superior structural models should provide excellent forecasting results. It can ultimately be concluded that, after eliminating the effects of autocorrelation through transformation, a simple, efficient model can be obtained to directly estimate overhead costs for five of these six aerospace contractors.

## LIST OF REFERENCES

- Beach, C.M., and J.G. MacKinnon, "A Maximum Likelihood Procedure for Regression with Autocorrelated Errors," *Econometrica*, Vol. 46, No. 1, pp. 52-58, January 1978.
- Boger, D.C., *Statistical Models For Estimating Overhead Costs*, Technical Report NPS-54-83-014, Department of Administrative Sciences, Naval Postgraduate School, 1983.
- Boger, D.C., *Statistical Modeling of Quarterly Contractor Overhead Costs*, Department of Administrative Sciences, Naval Postgraduate School, 1984.
- Box, G.E.P. and G.M. Jenkins, *Time Series Analysis Forecasting and Control*, Holden-Day, 1970.
- Cochrane, D. and G.H. Orcutt, "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms," *Journal of the American Statistical Association*, Vol. 44, No. 245, pp.32-61, March 1949.
- Conover, W.J., *Practical Nonparametric Statistics*, Second Edition, John Wiley and Sons, Inc., 1980.
- Draper, N.R. and H. Smith, *Applied Regression Analysis*, Second Edition, John Wiley and Sons, Inc., 1981.
- Hannan, E.J. and R.D. Terrell, "Testing for Serial Correlation After Least Squares Regression," *Econometrica*, Vol. 36, No. 1, pp. 133-150, January 1968.
- Intriligator, M.D., *Econometric Models, Techniques and Applications*, Prentice-Hall, Inc., 1978.
- Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lutkepohl, and T.-C. Lee, *The Theory and Practice of Econometrics*, Second Edition, John Wiley and Sons, 1985.
- Kmenta, J., *Elements of Econometrics*, Macmillan Company, 1971.
- Maddala, G.S., *Econometrics*, McGraw-Hill Book Company, 1977.
- Maeshiro, A., "On the Retention of the First Observations in Serial Correlation Adjustment of Regression Models," *International Economic Review*, Vol. 20, No. 1, pp. 259-265, February 1979.



Mood, A.M., F.A. Graybill and D.C. Boes, *Introduction to the Theory of Statistics*, Third Edition, McGraw-Hill Book Company, 1974.

Musgrave, G.L. and J.B. Ramsey, *APL-STAT*, Lifetime Learning Publications, 1981.

Park, R.E. and B.M. Mitchell, "Maximum Likelihood versus Minimum Sum-of-Squares Estimation of the Autocorrelated Error Model," Rand Note N-1325, Rand Corporation, November 1979.

Park, R.E. and B.M. Mitchell, "Estimating the Autocorrelated Error Model with Trended Data," *Journal of Econometrics*, Vol. 13, No. 2, pp. 185-201, June 1980.

Rao, P. and Z. Griliches, "Small-Sample Properties of Several Two-Stage Regression Methods in the Context of Auto-Correlated Errors," *Journal of the American Statistical Association*, Vol. 64, No. 325, pp. 253-272, March 1969.

Spitzer, J.J., "Small-Sample Properties of Nonlinear Least Squares and Maximum Likelihood Estimators in the Context of Autocorrelated Errors," *Journal of the American Statistical Association* Vol. 74, No. 365, pp. 41-47, March 1979.

Taylor, W.E., "On the Efficiency of the Cochrane-Orcutt Estimator," *Journal of Econometrics*, Vol. 17, No. 1, pp. 67-82, September 1981.

Theil, H. and A.L. Nager, "Testing the Independence of Regression Disturbances," *Journal of the American Statistical Association*, Vol. 56, No. 296, pp.793-806, December 1961.

Wallis, K.F., "Testing for Fourth-Order Autocorrelation in Quarterly Regression Equations," *Econometrica*, Vol. 40, No. 4, pp. 617-636, July 1972.

APPENDIX  
COMPUTER PROGRAMS

This appendix contains listings of the programs utilized in the analysis performed herein. All of the functions are written in APL and contain documentation. The programs utilized by the Monte Carlo simulation in Chapter 4 were SIM, RREGRESS, LAGS, DURBIN, CALL1, and TRANS. The following is a general description of what these programs do and the sequence of steps followed in the Monte Carlo simulation.

First, the function SIM generates the data required by the simulation. Next, an OLS regression of the generated dependent variable,  $Y_t$ , on the independent variable,  $X_t$ , is performed. The residuals of this OLS regression are then sent to the function LAGS which computes the nine estimators of  $\rho_4$  and the MSE of  $\rho_4$ . Next, the function TRANS transforms the data using the appropriate estimator. An EGLS regression is then performed on this transformed data. These last two steps are actually performed in the function CALL1 which calls the functions TRANS and RREGRESS in succession. The function LAGS then takes the results of this EGLS regression and computes the MSE of  $\beta_2$  and the Adjusted R-squared value. This cycle is replicated one hundred times for each value of  $\rho_4$ . The function RREGRESS is called to perform all of the regressions required in the simulation. The functions DURBIN, PWIT and MAX are called to compute the Durbin, iterative Prais-Winsten, and maximum likelihood estimators of fourth-order correlation.

The programs utilized in the structural analysis of Chapter 5 were REGRESS, TRANS, LAG, CHECKER, and VARTTEST. All regressions were performed by REGRESS (Musgrave and Ramsey, 1981, pp. 254-258). As can be seen from the program

listing, REGRESS outputs numerous statistical results. The function TRANS was utilized to transform the data for either an AR(1) or SAR(4) process. The functions LAG and CHECKER were called to compute numerous estimators and test statistics (for first and fourth-order autocorrelation) from the residuals. The F-test utilized in the analysis was performed using the function VARTEST.

### APL FUNCTION SIM

#### RHOS SIM REP

```
[1]  THIS PROGRAM PERFORMS A MONTE CARLO SIMULATION ON
[2]  THE NINE ESTIMATORS OF  $\rho_4$ . THE INPUT RHOS IS A
[3]  VECTOR OF  $\rho_4$  VALUES THAT YOU WANT TO RUN THE
[4]  SIMULATION FOR, IN THIS CASE IT IS A VECTOR OF
[5]  LENGTH TEN OF THE VALUES: .1, .2, .3, .4, .5, .6,
[6]  .7, .8, .9, AND .95. THE INPUT PARAMETER REP IS
[7]  THE NUMBER OF REPLICATIONS OF THE SIMULATION YOU
[8]  WISH TO RUN, IN THIS CASE 100.
[9]  PRINT← 3 9  $\rho_0$ 
[10] OUT←1
[11] RMSE $\rho_0$ ←AAR←EFF←RANK $\rho_0$ ←RANKAAR←RANKEFF
    ←(( $\rho_0$ RHOS),9) $\rho_0$ 
[12] L2:'PERFORMING TEST FOR  $\rho_4 \equiv$  ', $\rho_0$ RHOS[OUT]
[13]  $\rho_4$ ←RHOS[OUT]
[14] CREATION OF NECESSARY VECTORS.
[15] MSER $\rho_4$ ←MSER $\rho_4$ S←MSER $\rho_4$ L←MSEP $\rho_4$ ←MSEP $\rho_4$ S←MSEPLS←MSED←0
[16] MSEBOLS←MSEBR $\rho_4$ ←MSEBR $\rho_4$ S←MSEBR $\rho_4$ L←MSEBP $\rho_4$ ←MSEBP $\rho_4$ S←MSEBPLS
    ←MSEBD←0
[17] AAR2OLS←AAR2R $\rho_4$ ←AAR2R $\rho_4$ S←AAR2R $\rho_4$ L←AAR2P $\rho_4$ ←AAR2P $\rho_4$ S←AAR2PLS
    ←AAR2D←0
[18] MSEPW←MSEPPML←MSEBPW←MSEBPPML←AAR2PW←AAR2PPML←REPPW
    ←REPMAX←0
[19] PWITFAIL←0
[20] IN←1
[21] RANDOM GENERATION OF THE U(T). THEY ARE
```

[22] *ADISTRIBUTED NORMALLY WITH MEAN 0 AND VARIANCE*  
 [23] *ACCORDING TO THE CONTRACTOR.*  
 [24] *L1:U←32 NORRAND 0 13454*  
 [25] *ASET ALL E(T) TO U(T)÷((1-RHO4\*2)\*.5)*  
 [26] *E←U÷((1-RHO4\*2)\*0.5)*  
 [27] *I←5*  
 [28] *ASET THE LAST T-4 E(T)S TO (RHO4×E(T-4))+ U(T).*  
 [29] *IT:E[I]←(RHO4×E[I-4])+U[I]*  
 [30] *I←I+1*  
 [31] *→IT×<sub>1</sub>I≤32*  
 [32] *ASET B TO THE APPROPRIATE INTERCEPT AND SLOPE*  
 [33] *FOR THE CONTRACTOR BEING TESTED.*  
 [34] *B← 0 17*  
 [35] *XCURRENT←1,XCURRENT*  
 [36] *AGENERATE THE INDEPENDENT VARIABLE, Y*  
 [37] *YCURRENT←(XCURRENT+.×B)+ET←<sup>-24</sup>↑E*  
 [38] *XCURRENT←<sup>24</sup><sup>-1</sup>↑XCURRENT*  
 [39] *APERFORM THE OLS REGRESSION OF PRESET*  
 [40] *DEPENDENT VARIABLE, X, ON THE GENERATED*  
 [41] *INDEPENDENT VARIABLE, Y.*  
 [42] *YCURRENT RREGRESS XCURRENT*  
 [43] *MSEBOLS←MSEBOLS+((BER[2]-B[2])\*2)*  
 [44] *AAR2OLS←AAR2OLS+ADJR2*  
 [45] *ACALL THE FUNCTION LAGS WITH THE RESIDUALS FROM*  
 [46] *THE OLS REGRESSION AND THE VALUE OF RHO4.*  
 [47] *RHO4 LAGS UH*  
 [48] *IN←IN+1*  
 [49] *→L1×<sub>1</sub>IN≤REP*  
 [50] *ACOMPUTATION OF MOE'S AND OUTPUT*  
 [51] *MSERHO←MSER4,MSER4S,MSEP4,MSEP4S,MSED,MSEPLS,MSER4L,*  
*MSEPW,MSEPPML*  
 [52] *RMSEPHO[OUT;]←(MSERHO÷REP)\*0.5*  
 [53] *AR←AAR2R4,AAR2R4S,AAR2P4,AAR2P4S,AAR2D,AAR2PLS,*  
*AAR2R4L,AAR2PW,AAR2PPML*  
 [54] *AAR[OUT;]←AR÷REP*

```

[55] MS←MSEBR4 ,MSEBR4S ,MSEBP4 ,MSEBP4S ,MSEBD ,MSEBPLS ,
      MSEBR4L ,MSEBPW ,MSEBPPML
[56] RMSEBETA←(MS÷REP)*0.5
[57] EFF[OUT;]←((MSEBOLS÷REP)*0.5)÷RMSEBETA
[58] PRINT[1;]←RMSEBETA[OUT;]
[59] PRINT[2;]←AAR[OUT;]
[60] PRINT[3;]←EFF[OUT;]
[61] RANKRHO[OUT;]←ΔΔRMSEBETA[OUT;]
[62] RANKAAR[OUT;]←ΔΔAAR[OUT;]
[63] RANKEFF[OUT;]←ΔΔEFF[OUT;]
[64] 'FOR RHO4 ≡ ' ,ϕRHO4
[65] 'RMSE RHO,ADJUSTED RSQR AND EFF:'
[66] ' 1 2 3 4 5 6 7 8 9'
[67] PRINT
[68] 'AVG ADJ R2 FOR OLS ≡ ' ,ϕ(AAR2OLS÷REP)
[69] 'AVG REPS FOR PWIT ≡ ' ,ϕ(REPPW÷REP)
[70] 'PWIT FAILED TO CONVERGE ≡ ' ,ϕPWITFAIL
[71] 'AVG REPS FOR MAX ≡ ' ,ϕ(REPMAX÷REP)
[72] ' '
[73] OUT←OUT+1
[74] →L2×1OUT≤ρRHOS
[75] 'RMSE RHO'
[76] RMSEBETA
[77] ' '
[78] 'RANK RMSEBETA'
[79] RANKRHO
[80] 'AVG RANK RMSEBETA'
[81] (RANKRHO)÷ρRHOS
[82] ' '
[83] 'ADJUSTED R*2'
[84] AAR
[85] ' '
[86] 'RANK ADJUSTED R*2'
[87] RANKAAR
[88] 'AVG RANK ADJ R*2'

```

```

[89] (+/RANKAAR)÷ρRHOS
[90] ' '
[91] 'EFFICIENCY'
[92] EFF
[93] ' '
[94] 'RANK EFFICIENCY'
[95] RANKEFF
[96] 'AVG RANK EFF.'
[97] (+/RANKEFF)÷ρRHOS

```

### APL FUNCTION LAGS

#### RHO4 LAGS XX

```

[1] ρUSED IN THE MONTE CARLO SIMULATION. IT IS CALLED
[2] ρTO COMPUTE THE NINE ESTIMATORS OF ρ4 AND TO
[3] ρPERFORM THE EGLS REGRESSION AND CALCULATE
[4] ρTHE MOES FOR EACH ESTIMATOR. THE INPUT RHO4 IS THE
[5] ρVALUE OF ρ4 THAT THE SIMULATION IS BEING PERFORMED
[6] ρFOR THE INPUT XX IS A VECTOR OF THE RESIDUALS.
[7] I←1
[8] J←ρXX
[9] L←J+1
[10] K←J-4
[11] AA←EE←1K
[12] ρDETERMINE THE UNLAGGED AND LAGGED RESIDUALS.
[13] LOOP:ID←L-I
[14] AA[I]←XX[ID]
[15] EE[I]←XX[ID-4]
[16] I←I+1
[17] →LOOP×1I≤K
[18] T4←(ρXX)-4
[19] T5←T4-1
[20] A←+/XX*2
[21] A2←+/(0,0,0,0,(T5ρ1),0)/XX)*2

```

```

[22] D4←(÷A)×+/(((0,0,0,0,T4p1)/XX)-((T4p1),0,0,0,0)/XX)*2
[23] ρCALCULATE THE SEVEN ESTIMATORS OF ρ4
[24] R4←(÷A)×+/(((0,0,0,0,T4p1)/XX)×((T4p1),0,0,0,0)/XX)
[25] R4STAR←((J-2)×R4)÷(J-1)
[26] R4L←(÷A2)×+/(((0,0,0,0,T4p1)/XX)×((T4p1),0,0,0,0)/XX)
[27] P4←1-(0.5×D4)
[28] P4STAR←(((J*2)×P4)+4)÷((J*2)-4)
[29] PLS←AA⊞EE
[30] ρCALL THE FUNCTIONS DURBIN, PWIT, AND MAX TO
[31] ρCALCULATE THE DURBIN, ITERATIVE PRAIS-WINSTEN,
[32] ρAND MAXIMUM LIKELIHOOD ESTIMATORS
[33] DURBIN
[34] PWIT
[35] →MM×1(PWITF≡0)
[36] PWITFAIL←PWITFAIL+1
[37] →FF×1(PWITF≡1)
[38] MM:REPPW←REPPW+(NPW-1)
[39] FF:MAX
[40] →AA×1(MAXF≡0)
[41] AA:REPMAX←REPMAX+(NML-1)
[42] ρCALCULATE THE MSE OF ρ4 FOR EACH ESTIMATOR
[43] MSER4←MSER4+((R4-RHO4)*2)
[44] MSER4S←MSER4S+((R4STAR-RHO4)*2)
[45] MSER4L←MSER4L+((R4L-RHO4)*2)
[46] MSEP4←MSEP4+((P4-RHO4)*2)
[47] MSEP4S←MSEP4S+((P4STAR-RHO4)*2)
[48] MSEPLS←MSEPLS+((PLS-RHO4)*2)
[49] MSED←MSED+((PDURBIN-RHO4)*2)
[50] →MS×1(PWITF≡1)
[51] MSEPW←MSEPW+((PW-RHO4)*2)
[52] MS:MSEPPML←MSEPPML+((PPML-RHO4)*2)
[53] ρPERFORM THE EGLS REGRESSION AND CALCULATE MSE BETA
[54] ρAND ADJUSTED R-SQUARED FOR EACH OF THE ESTIMATORS
[55] CALL1 R4
[56] MSEBR4←MSEBR4+((BER[2]-B[2])*2)

```

```

[57] AAR2R4←AAR2R4+ADJR2
[58] CALL1 R4STAR
[59] MSEBR4S←MSEBR4S+((BER[2]-B[2])*2)
[60] AAR2R4S←AAR2R4S+ADJR2
[61] CALL1 R4L
[62] MSEBR4L←MSEBR4L+((BER[2]-B[2])*2)
[63] AAR2R4L←AAR2R4L+ADJR2
[64] CALL1 P4
[65] MSEBP4←MSEBP4+((BER[2]-B[2])*2)
[66] AAR2P4←AAR2P4+ADJR2
[67] CALL1 P4STAR
[68] MSEBP4S←MSEBP4S+((BER[2]-B[2])*2)
[69] AAR2P4S←AAR2P4S+ADJR2
[70] CALL1 PLS
[71] MSEBPLS←MSEBPLS+((BER[2]-B[2])*2)
[72] AAR2PLS←AAR2PLS+ADJR2
[73] CALL1 PDURBIN
[74] MSEBD←MSEBD+((BER[2]-B[2])*2)
[75] AAR2D←AAR2D+ADJR2
[76] →CC×1(PWITF≡1)
[77] CALL1 PW
[78] MSEBPW←MSEBPW+((BER[2]-B[2])*2)
[79] AAR2PW←AAR2PW+ADJR2
[80] CC:CALL1 PPML
[81] MSEBPPML←MSEBPPML+((BER[2]-B[2])*2)
[82] AAR2PPML←AAR2PPML+ADJR2

```

## APL FUNCTION TRANS

### P TRANS V

```

[1] ρTHIS FUNCTION TRANSFORMS THE RAW DATA FOR EITHER
[2] ρAN AR(1) OR AN SAR(4) PROCESS. THE INPUT P IS A
[3] ρVECTOR OF LENGTH 2 WHERE P[1] IS THE ESTIMATE OF ρ1
[4] ρAND P[2] IS THE ESTIMATE OF ρ4. THE INPUT V IS THE

```



```

[5]  ρVARIABLE YOU WANT TO BE TRANSFORMED. V CAN BE
[6]  ρEITHER A VECTOR OF LENGTH N OR A N×1 MATRIX.
[7]  DIM←ρρV
[8]  →VECTOR×1DIM≡1
[9]  ρRESHAPE V INTO A VECTOR IF IT WAS ENTERED
[10] ρAS A N×1 MATRIX
[11] CHECK←ρV
[12] CHECK1←CHECK[1]
[13] V←CHECK1ρV
[14] VECTOR:P1←P[1]
[15] P2←P[2]
[16] →NEXT×1P2>0
[17] ρLOOP TO BE PERFORMED FOR THE AR(1) TRANSFORMATION
[18] I1←(ρV)-1
[19] V1←(0,I1ρ1)/V
[20] VP1←((I1ρ1),0)/V
[21] VV1←V[1]×((1-(P1*2))*0.5)
[22] VV2←V1-(P1×VP1)
[23] VV←VV1,VV2
[24] →END×1DIM≡1
[25] ρRESHAPE THE VARIABLE INTO A N×1 MATRIX IF IT
[26] ρWAS ENTERED AS SUCH
[27] VV←(CHECK1,1)ρVV
[28] →END
[29] ρLOOP TO BE PERFORMED FOR THE AR(4) TRANSFORMATION
[30] NEXT:I4←(ρV)-4
[31] V1←(0,0,0,0,I4ρ1)/V
[32] VP4←((I4ρ1),0,0,0,0)/V
[33] ρCHECK IF THE ESTIMATOR IS < 1.0 AND IF NOT:
[34] →OK×1P2<1
[35] ρ 1) SET THE ESTIMATOR TO .99999
[36] P2←0.99999
[37] ρSAR(4) TRANSFORMATION COMPUTATION
[38] OK:VP1234←4↑V
[39] VV1234←VP1234×((1-(P2*2))*0.5)

```

```

[40] VV5←V1-(P2×VP4)
[41] VV←VV1234,VV5
[42] →END×1DIM≡1
[43] ρRESHAPE THE VARIABLE INTO A N×1 MATRIX IF IT
[44] ρWAS ENTERED AS SUCH
[45] VV←(CHECK1,1)ρVV
[46] END:

```

#### APL FUNCTION DURBIN

##### DURBIN

```

[1] ρTHIS FUNCTION CALCULATES THE DURBIN ESTIMATOR OF ρ4.
[2] DIMD←ρYCURRENT
[3] T4D←DIMD-4
[4] YDURBIN←(0,0,0,0,T4Dρ1)/YCURRENT
[5] XDURBIN←(T4D,4)ρ0
[6] XDURBIN[;1]←1
[7] XDURBIN[;2]←((T4Dρ1),0,0,0,0)/YCURRENT
[8] XCT←DIMDρXCURRENT
[9] XDURBIN[;3]←(0,0,0,0,T4Dρ1)/XCT
[10] XDURBIN[;4]←-1×(((T4Dρ1),0,0,0,0)/XCT)
[11] COE←YDURBIN⊖XDURBIN
[12] P5←COE[2]

```

#### APL FUNCTION PWIT

##### PWIT

```

[1] ρTHIS FUNCTION COMPUTES THE ITERATIVE PRAIS-WINSTEN
[2] ρESTIMATOR.
[3] TP←ρYCURRENT
[4] TPW←TP-4
[5] CHE←2
[6] PWITF←NPW←0

```

```

[7]   XPW←XCURRENT
[8]   YPW←YCURRENT
[9]   XPWT1←XPW1←(TP,1)ρ1
[10]  ITER:UHPW←(YCURRENT-(YHPW←(1,XCURRENT)+.×(BEPW←YPW
      (XPW←(XPWT1,XPW))))
[11]  DEN←+/(0,0,0,0,((TPW-1)ρ1),0)/UHPW)*2
[12]  PW←(÷DEN)×+/(0,0,0,0,TPWρ1)/UHPW)×((TPWρ1),0,0,0,0)
      /UHPW)
[13]  P←0,PW
[14]  ρTRANSFORM X AND Y
[15]  P TRANS XCURRENT
[16]  XPW←VV
[17]  P TRANS YCURRENT
[18]  YPW←VV
[19]  ρIF P>1 SET TO .99999
[20]  →CONPW×1(PW<1)
[21]  PW←0.99999
[22]  CONPW:XPWT1←XPW1×((TP,1)ρ(4ρ((1-PW*2)*0.5)),((TP-4)
      ρ(1-PW)))
[23]  NPW←NPW+1
[24]  →DD×1(NPW≤52)
[25]  PWITF←1
[26]  →0
[27]  DD:DELTA←CHE-PW
[28]  CHE←PW
[29]  ρPERFORM ANOTHER ITERATION IF THE ABSOLUTE
[30]  ρDIFFERENCE BETWEEN THE LAST TWO ESTIMATIONS
[31]  ρIS > .00001.
[32]  →ITER×1(|DELTA)>0.00001

```

## APL FUNCTION MAX

### MAX

```

[1]  ρTHIS FUNCTION COMPUTES THE MAXIMUMLIKELIHOOD

```

```

[2]  ρESTIMATOR OF ρ4 USING THE ALGORITHM DERIVED
[3]  ρBY BEACH AND MACKINNON.
[4]  NPML←ρYCURRENT
[5]  CH←2
[6]  T4ML←NPML- 4
[7]  XML←XCURRENT
[8]  YML←YCURRENT
[9]  XMLT1←XML1←(NPML, 1)ρ1
[10]  MAXF←NML←0
[11]  ρCOMPUTE THE RESIDUALS FOR THE REGRESSION OF X ON Y.
[12]  ML:UH←(YCURRENT- (YHML←(1, XCURRENT)+. ×(BEML←YML⊗(XML
      ←(XMLT1, XML))))))
[13]  A4←(1, 1, 1, 1, (T4MLρ0))/UH
[14]  AT4←((T4MLρ1), 0, 0, 0, 0)/UH
[15]  AT←(0, 0, 0, 0, T4MLρ1)/UH
[16]  DENOM←(NPML- 1)×((+/(AT4*2))- (+/(A4*2)))
[17]  ρCOMPUTE THE COEFFICIENTS OF THE POLYNOMIAL
[18]  AML←(¯1×(NPML- 2)×(+/(AT×AT4)))÷DENOM
[19]  BML←(((NPML- 1)×(+/(A4*2)))- ((NPML×(+/AT4*2))+
      (+/AT*2)))÷DENOM
[20]  CML←(NPML×(+/AT×AT4))÷DENOM
[21]  PMLP←BML- ((AML*2)÷3)
[22]  QML←CML+(2×(AML*3)÷27)- (AML×BML÷3)
[23]  PHI1←((QML×(27*0.5))÷(2×PMLP×((¯1×PMLP)*0.5)))
[24]  PHI←¯2◦PHI1
[25]  ρCOMPUTE THE ESTIMATOR
[26]  PPML←(¯2×(((¯1×PMLP)÷3)*0.5)×(2◦((PHI÷3)+◦(÷3))))-
      (AML÷3)
[27]  PPPML←◦, PPML
[28]  ρCALL THE FUNCTION TRANS TO TRANSFORM THE RAW DATA
[29]  PPPML TRANS XCURRENT
[30]  XML←VV
[31]  PPPML TRANS YCURRENT
[32]  YML←VV
[33]  ρIF P>1 RESET TO .99999

```

```

[34] →CONMAX×1(PPML<1)
[35] PPML←0.99999
[36] CONMAX:XMLT1←XML1×((NPML,1)ρ(4ρ((1-PPML*2)*0.5)),
      ((NPML-4)ρ(1-PPML)))
[37] NML←NML+1
[38] →MAF×1(NML≤52)
[39] MAXF←1
[40] →0
[41] MAF:DELT←CH-PPML
[42] CH←PPML
[43] ρIF THE ABSOLUTE DIFFERENCE BETWEEN THE LAST TWO
[44] ρESTIMATES IS > .00001 THEN PERFORM ANOTHER
[45] ρITERATION
[46] →ML×1(|DELT)>0.00001

```

#### APL FUNCTION CALL1

##### CALL1 EST

```

[1] ρTHIS FUNCTION CALLS THE TRANSFORMATION FUNCTION
[2] ρTRANS FOR AN ESTIMATOR OF  $\rho_4$  AND THEN CALLS THE
[3] ρFUNCTION RREGRESS AND PERFORMS AN EGLS REGRESSION
[4] ρON THE TRANSFORMED DATA.
[5] PEST←0,EST
[6] PEST TRANS XCURRENT
[7] XT←VV
[8] PEST TRANS YCURRENT
[9] YT←VV
[10] YT RREGRESS XT

```

#### APL FUNCTION RREGRESS

##### Y RREGRESS X;MS;SS

```

[1] ρTHIS IS A CONDENSED VERSION OF REGRESS LESS

```

```

[2]  ρTHE PRINTED OUTPUT TO BE USED IN CONJUCTION WITH
[3]  ρTHE SIMULATION SIM.
[4]  NP←ρX
[5]  K←NP[2]
[6]  SS←3ρ0
[7]  MS←((+/Y),MS←+/X)÷NP[1]
[8]  SS[2]←(SS[1]←+/Y*2)-NP[1]×MS[1]*2
[9]  UH←(Y-(YH←X+.×(BER←Y⊖(X←(1,X))))))
[10] SS[3]←+/UH*2
[11] R2←1-SS[3]÷SS[2]
[12] NMINUSK←NP[1]-(K+1)
[13] ADJR2←R2-((K÷NMINUSK)×(1-R2))

```

#### APL FUNCTION REGRESS

Y REGRESS X;MS;D;SS

```

[1]  ρTHIS FUNCTION PERFORMS AN OLS REGRESSION OF X ON Y.
[2]  ρTHE INPUT X IS THE INDEPENDENT VARIABLE AN IS
[3]  ρASSUMED TO BE AN N×K MATRIX. THE INPUT Y IS THE
[4]  ρDEPENDENT VARIABLE AND IS ASSULMED TO BE A VECTOR
[5]  ρOF LENGTH N. A DIAGNOSTIC IS PRINTED IF N≤K OR IF
[6]  ρRANK X<K. THE CONSTANT TERM IS ADDED BY THE PROGRAM.
[7]  ρTHIS FUNCTION WAS OBTAINED FROM RAMSEY AND
[8]  ρMUSGRAVE'S PL-ST T (SEE REFERENCES).
[9]  FLAG←1
[10] NP←ρX
[11] K←NP[2]
[12] SS←5ρ0
[13] →PREMEND1×1(NP[1]≤NP[2]+1)
[14] CM←(MM←(⊖X)+.×X)-(MS°.×MS←+/X)÷NP[1]
[15] CRM←D+.×CM+.×D←⊖(((1K)°.≡1K)×CM)*0.5
[16] MS←((+/Y),MS)÷NP[1]
[17] SS[2]←(SS[1]←+/Y*2)-NP[1]×MS[1]*2
[18] MAIN:UH←(Y-(YH←X+.×(BE←Y⊖(X←(XCON1,X))))))

```

```

[19] SS[4]←(SS[3]←+/YH*2)-NP[1]×MS[1]*2
[20] SS[5]←+/UH*2
[21] RSQ←1-SS[5]÷SS[2]
[22] NMINUSK←NP[1]- (K+1)
[23] ADJRSQ←RSQ- ((K÷NMINUSK)×(1- RSQ))
[24] VARU←SS[5]÷(- /NP)- 1
[25] STDERR←VARU*0.5
[26] F1←SS[4]÷SS[5]×K÷(- /NP)- 1
[27] F2←SS[3]÷SS[5]×(K+1)÷(- /NP)- 1
[28] COVBE←VARU× $\sum$ ( $\phi$ X)+.×X
[29] STDBE←(1 1  $\phi$ COVBE)*0.5
[30] TRATIO←BE÷STDBE
[31] →CONT
[32] PREMEND:FLAG←0
[33] CONT:→MAINEND×1FLAG
[34] 'ROUTINE ENDED DUE TO SINGULARITY OF X MATRIX'
[35] →0
[36] MAINEND:'THE COEFFICIENT ESTIMATES ARE, CONST.,X1:'
[37] BE
[38] 'THE CORRESPONDING STANDARD ERRORS ARE.'
[39] STDBE
[40] 'THE CORRESPONDING T RATIOS ARE:'
[41] TRATIO
[42] 'WITH DEGREES OF FREEDOM:'
[43] NMINUSK
[44] 'RSQ IS: ', $\phi$ RSQ
[45] 'ADJUSTED RSQ IS: ', $\phi$ ADJRSQ
[46] 'STANDARD ERROR OF REGRESSION IS: ', $\phi$ STDERR
[47] 'VAR OF ERROR TERM IS: ', $\phi$ VARU
[48] 'THE F STATISTIC INCLUDING THE CONSTANT TERM IS:'
[49] F2
[50] 'WITH DEGREES OF FREEDOM:'
[51] (K+1),( $\bar{1}$ + - /NP)
[52] 'THE F STATISTIC NOT INCLUDING CONSTANT TERM IS:'
[53] F1

```

```

[54] 'WITH DEGREES OF FREEDOM.'
[55] K, (1+ /NP)
[56] 'THIS ENDS THE OUTPUT FROM REGRESS.'
[57] →0
[58] PREMEND1: 'NO. OF OBS (N) IS TOO FEW RELATIVE THE '
[59] 'NO. OF REGRESSORS (K).'
[60] 'ROUTINE TERMINATED'

```

### APL FUNCTION LAG

#### LAG XX

```

[1]  GIVEN THE RESIDUALS THIS FUNCTION CALCULATES THE
[2]  DURBIN-WATSON AND WALLIS TEST STATISTICS, THE
[3]  SINGLE ESTIMATOR OF  $\rho_1$ , AND THREE PREFERRED
[4]  ESTIMATORS OF  $\rho_4$  (FROM THE SIMULATION). THE INPUT
[5]  XX IS A VECTOR OF LENGTH N OF THE REGRESSION
[6]  RESIDUALS.
[7]   $J \leftarrow \rho XX$ 
[8]   $T1 \leftarrow J - 1$ 
[9]   $T2 \leftarrow J - 2$ 
[10]  $T4 \leftarrow J - 4$ 
[11]  $A \leftarrow + / XX * 2$ 
[12]  $A1 \leftarrow + / ((0, (T2 \rho 1), 0) / XX) * 2$ 
[13]  CALCULATE THE TWO TEST STATISTICS
[14]  $D1 \leftarrow (\div A) \times + / (((0, T1 \rho 1) / XX) - ((T1 \rho 1), 0) / XX) * 2$ 
[15]  $D4 \leftarrow (\div A) \times + / (((0, 0, 0, 0, T4 \rho 1) / XX) - ((T4 \rho 1), 0, 0, 0, 0) / XX) * 2$ 
[16]  CALCULATE THE ESTIMATOR OF  $\rho_1$ 
[17]  $PP \leftarrow (\div A1) \times + / (((0, T1 \rho 1) / XX) \times ((T1 \rho 1), 0) / XX)$ 
[18]  CALCULATE THE THREE ESTIMATORS OF  $\rho_4$ 
[19]  ESTIMATOR NUMBER THREE IS
[20]  $P3 \leftarrow 1 - (0.5 \times D4)$ 
[21]  ESTIMATOR NUMBER FOUR IS
[22]  $P4 \leftarrow (((J * 2) \times P3) + 4) \div ((J * 2) - 4)$ 
[23]  ESTIMATOR NUMBER NINE IS THE MAXIMUM LIKELIHOOD

```



```

[24] ESTIMATOR
[25] MAX
[26] 'D1  $\equiv$  ',  $\phi$ D1
[27] 'DEGREES OF FREEDOM  $\equiv$  ',  $\phi$ T1
[28] 'PRAIS-WINSTEN 1ST ORDER ESTIMATE, P1  $\equiv$  ',  $\phi$ PP
[29] 'D4  $\equiv$  ',  $\phi$ D4
[30] 'DEGREES OF FREEDOM  $\equiv$  ',  $\phi$ T4
[31] 'DURBIN-WATSON 4TH ORDER ESTIMATE, P3  $\equiv$  ',  $\phi$ P3
[32] 'D-W THEIL-NAGAR MOD. ESTIMATE, P4  $\equiv$  ',  $\phi$ P4
[33] 'THE MAX LIKE ESTIMATOR IS, P9  $\equiv$  ',  $\phi$ PPML

```

### *APL FUNCTION CHECKER*

#### *CHECKER RESL*

```

[1] THIS FUNCTION COMPUTES THE DURBIN-WATSON AND
[2] WALLIS TEST STATISTICS ON THE RESIDUALS OF A
[3] REGRESSION. THE INPUT RESL IS A VECTOR OF LENGTH
[4] N OF RESIDUALS.
[5] DETERMINE DEGREES OF FREEDOM
[6]  $TC1 \leftarrow (\rho RESL) - 1$ 
[7]  $TC4 \leftarrow (\rho RESL) - 4$ 
[8]  $AC \leftarrow + / RESL * 2$ 
[9] DETERMINE TEST STATISTICS
[10]  $D1C \leftarrow (\div AC) \times + / (((0, TC1 \rho 1) / RESL) - ((TC1 \rho 1), 0) / RESL) * 2$ 
[11]  $D4C \leftarrow (\div AC) \times + / (((0, 0, 0, 0, TC4 \rho 1) / RESL) - ((TC4 \rho 1), 0, 0, 0, 0) / RESL) * 2$ 
[12] ' '
[13] 'AFTER TRANSFORMATION'
[14] 'D1  $\equiv$  ',  $\phi$ D1C
[15] 'DEGREES OF FREEDOM  $\equiv$  ',  $\phi$ TC1
[16] 'D4  $\equiv$  ',  $\phi$ D4C
[17] 'DEGREES OF FREEDOM  $\equiv$  ',  $\phi$ TC4

```

APL FUNCTION VARTEST

ALPHA VARTEST S

```

[1]  THIS FUNCTION PERFORMS THE F-TEST TO CHECK THE
[2]  HOMOGENEITY OF THE VARIANCE OF THE RESIDUALS.
[3]  NO←0.5×ρS
[4]  SA←SB←NOρ0
[5]  DOF←NO- 1
[6]  D←2ρDOF
[7]  SERARATE THE RESIDUALS INTO TWO GROUPS.
[8]  SA←NO↑S
[9]  SB←NO↓S
[10] ABAR←(+/SA)÷NO
[11] BBAR←(+/SB)÷NO
[12] CALCULATE THE VARIANCES OF THE TWO GROUPS.
[13] ASOS←+/(SA-ABAR)*2
[14] BSOS←+/(SB-BBAR)*2
[15] CALCULATE THE TEST STATISTIC R AND THE CRITICAL
[16] REGION.
[17] R←(ASOS↑BSOS)÷(ASOS↓BSOS)
[18] K1←D FQUAN(ALPHA÷2)
[19] K2←D FQUAN(1-(ALPHA÷2))
[20] DETERMINE THE LEVEL AT WHICH THE TEST STATISTIC IS
[21] SIGNIFICANT.
[22] ALPHAC←2×(1-(D FCENT R))
[23] 'FOR AN F TEST WITH ALPHA ≡ ',ⓧALPHA
[24] DETERMINE IF THE TEST STATISTIC R FALLS WITHIN
[25] THE CRITICAL REGION.
[26] →A×₁((R>K1)^(R<K2))
[27] 'REJECT H₀: VAR1 ≡ VAR2'
[28] →NEXT
[29] A:'ACCEPT H₀: VAR1 ≡ VAR2'
[30] NEXT:'F STATISTIC ≡ ',ⓧR
[31] 'OBSERVED LEVEL OF SIGNIF. IS ALPHA ≡ ',ⓧALPHAC

```

INITIAL DISTRIBUTION LIST

	No.	Copies
1. Defense Technical Information Center Cameron Station Alexandria, VA 22304-6145	2	
2. Defense Logistics Studies Information Exchange U.S. Army Logistics Management Center Fort Lee, VA 23801-6043	1	
3. Library Code 0142 Naval Postgraduate School Monterey, CA 93943-5002	2	
4. Associate Professor Dan C. Boger, Code 55Bo Department of Operations Research Naval Postgraduate School Monterey, CA 93943-5000	5	
5. Associate Professor F.R. Richards, Code 55Rh Department of Operations Research Naval Postgraduate School Monterey, CA 93943-5000	1	
6. David J. Jerabek 5408 Moore Rd. Muncie, IN 47302	5	







217465

Thesis

J4753

Jerabek

c.1

Regression models of  
quarterly overhead  
costs for six govern-  
ment aerospace contrac-  
tors.



thesJ4753

Regression models of quarterly overhead



3 2768 000 65929 6

DUDLEY KNOX LIBRARY