Appendix 4A

Problems in Applying the Linear Regression Model

Solutions to Exercises

1. An appliance manufacturer uses quarterly time-series data:

   a. Possible *causes* of autocorrelation:

   i. *trends and cycles in economic variables*—the overall growth in the economy combined with business cycles causes most economic time series to have an overall upward trend with periodic upturns and downturns about this trend. As a result, time series data (and hence residuals) tend to exhibit positive autocorrelation.

   ii. *missing variables*—significant explanatory variables may be missing from this regression equation.

   iii. *nonlinearities*—nonlinear relationships may exist between the dependent and independent variables.

   b. Under autocorrelation, the least-squares procedure will tend to over or underestimate the sampling variances of the estimates of the regression parameters (*i.e.,* α and β's) depending on whether there is positive or negative autocorrelation. As a result, conclusions concerning the statistical significance of the individual explanatory variables by means of the t-test may not be correct. In short, the t-values are unreliable. Also, the coefficient of determination and F-test will no longer be reliable measures of the overall explanatory power of the regression model.

   c. Since the estimates of regression coefficients are unbiased even when autocorrelation is present, the forecasts of the dependent variable will be unbiased. However, the forecasts will have unnecessarily large sampling variances and hence will be less accurate than if autocorrelation were not present.

   d. Possible techniques for removing the autocorrelation:

   i. if the functional form of the dependence relationship among the successive values of the residuals is known, then the *original variables can be transformed* to remove this dependence. Popular transformations include logarithms, polynomials, and reciprocals.

ii. include a linear trend or "time" variable as one of the explanatory variables in the regression equation.
iii. calculate *first differences* in the time series of each of the variables (both dependent and explanatory) and then calculate the regression coefficients using these transformed variables.

2. Household cleaning products are examined by a product manager:

a. The linear model result is: Y = 1.210 +.0838X, where X is Selling Expense and Y is Sales, and has R^2 = .93

b. The doublé log model is: logY = log α + β logX + log μ

c. The double log result is: logY = 2.4747 + .6965logX , which is: Y = $298.3X^{.6965}$, with R^2 = .98

d. The logarithmic model explains more of the variability in the sales volume from region to region than does the linear model.

e. Since the logarithmic model fits the data better than the linear model, this indicates that a diminishing marginal returns relationship probably exists between sales and selling expenses. From calculus, we know that dY/dX = .6965(293.3) $X^{-.3035}$, which declines as X rises.

f. One might consider either a polynomial transformation, such as including a squared amount of Selling Expense, or a reciprocal transformation.

3. Sherwin Williams with logarithms:

a. Dep var: LSALES N: 10
Multiple R: .889 Squared multiple R: .791
Adjusted squared multiple R: .686
Standard error of estimate: 0.101

Variable	Coefficient	Std error	T
CONSTANT	6.952	1.443	4.818
LPROMEXP	–0.023	0.142	–0.164
LSELLPR	–1.037	0.366	–2.838
LDISPINC	0.388	0.375	1.034

Analysis of Variance

Source	Sum-of-squares	DF	Mean-square	F-ratio	P
Regression	0.232	3	0.077	7.558	0.018
Residual	0.061	6	0.010		

$$\text{Log } Y' = 6.952 - .023 \text{ Log } X_1 - 1.037 \text{ Log } X_2 + .388 \text{ Log } X_3$$

b. The negative sign of the promotional expenditures variable in the multiplicative model is inconsistent with the hypothesized sign (*i.e.*, positive). Selling price is statistically significant (at the .05 level) in both models. None of the other variables are statistically significant in either model. The overall measures of goodness of fit, namely R-square and F-ratio, are about equal in both models. These coefficients are also estimates of the related elasticities. Demand appears to be slightly elastic (-1.037); the income elasticity is normal but in the category of necessities (.388); and the promotional expense has an unexpected elasticity of –0.023, though not significantly different than zero.

4. Sales, Advertising, and Price data for 10 observations:

a. Dependent Variable: Sales in the linear model (i) with R^2 = 0. 0.9692 and an F-ratio of 110.122. S' = 247.644 + .3926A –. 7339P

Variable	Parameter Est.	Std. Error	t-ratio
Intercept	247.644	52.818078	3.942
Adverting	0.392558	0.029628	13.250
Price	-0.733859	.501514	-1.463

a. Dependent Variable: Log sales in the multiplicative model (ii) with R^2 =0.9659 and an F-ratio of 99.215. Log(S') = 2.4482 + .7296 Log A – 0.2406 Log P

Variable	Parameter Est.	Std. Error	t-ratio
Intercept	2.448206	0.616209	3.973
LG-Advertizing	0.729561	0.061676	11.829
LG-Price	–0.240607	0.158422	–1.519

b. From the computer output, only the advertising variable is significant at the 5% level (in both the linear and double log specifications). Price is not significant.

c. The R^2 and F-ratios are slightly higher for the linear model, but the difference is not significant.

5. It is likely that larger buildings also tend to have more rooms. Hence, it is reasonable to have a suspicion of multicollinearity.

   The simple correlation between size and the number of rooms is 0.590302. Nevertheless, multicollinearity gives unbiased estimates of the impact of each of the variables, and if the issue for the County Assessor is to give the best fitting price of prospective homes based on the sample, the R2 is higher using rooms, size, age, and the dummy variable for whether the garage is attached.

   Although large, this level of collinearity is unlikely to cause such an exaggeration of the standard errors as to render the t-scores a reliable test statistic.