with a constant term can R^2 be interpreted as the proportion of variation in y explained by variation in $\mathbf{x}$. An analogous computation can be done without computing deviations from means if the regression does not contain a constant term. Other purely algebraic artifacts will crop up in regressions without a constant, however. For example, the value of R^2 will change when the same constant is added to each observation on y, but it is obvious that nothing fundamental has changed in the regression relationship. One should be wary (even skeptical) in the calculation and interpretation of fit measures for regressions without constant terms.

3.6 SUMMARY AND CONCLUSIONS

This chapter has described the purely algebraic exercise of fitting a line (hyperplane) to a set of points using the method of least squares. We considered the primary problem first, using a data set of n observations on K variables. We then examined several aspects of the solution, including the nature of the projection and residual maker matrices and several useful algebraic results relating to the computation of the residuals and their sum of squares. We also examined the difference between gross or simple regression and correlation and multiple regression by defining "partial regression coefficients" and "partial correlation coefficients." The Frisch-Waugh Theorem (3.3) is a fundamentally useful tool in regression analysis which enables us to obtain in closed form the expression for a subvector of a vector of regression coefficients. We examined several aspects of the partitioned regression, including how the fit of the regression model changes when variables are added to it or removed from it. Finally, we took a closer look at the conventional measure of how well the fitted regression line predicts or "fits" the data.

Key Terms and Concepts

- Adjusted R-squared
- Analysis of variance
- Bivariate regression
- Coefficient of determination
- Disturbance
- Fitting criterion
- Frisch-Waugh theorem
- Goodness of fit
- Least squares
- Least squares normal equations
- Moment matrix
- Multiple correlation
- Multiple regression
- Netting out
- Normal equations
- Orthogonal regression
- Partial correlation coefficient
- Partial regression coefficient
- Partialing out
- Partitioned regression
- Prediction criterion
- Population quantity
- Population regression
- Projection
- Projection matrix
- Residual
- Residual maker
- Total variation

Exercises

1. **The Two Variable Regression.** For the regression model $y = \alpha + \beta x + \varepsilon$,
   a. Show that the least squares normal equations imply $\Sigma_i e_i = 0$ and $\Sigma_i x_i e_i = 0$.
   b. Show that the solution for the constant term is $a = \bar{y} - b\bar{x}$.
   c. Show that the solution for b is $b = [\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})]/[\sum_{i=1}^{n}(x_i - \bar{x})^2]$.

d. Prove that these two values uniquely minimize the sum of squares by showing that the diagonal elements of the second derivatives matrix of the sum of squares with respect to the parameters are both positive and that the determinant is $4n[(\sum_{i=1}^{n} x_i^2) - n\bar{x}^2] = 4n[\sum_{i=1}^{n}(x_i - \bar{x})^2]$, which is positive unless all values of x are the same.

2. **Change in the sum of squares.** Suppose that $\mathbf{b}$ is the least squares coefficient vector in the regression of $\mathbf{y}$ on $\mathbf{X}$ and that $\mathbf{c}$ is any other $K \times 1$ vector. Prove that the difference in the two sums of squared residuals is

$$(\mathbf{y} - \mathbf{Xc})'(\mathbf{y} - \mathbf{Xc}) - (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb}) = (\mathbf{c} - \mathbf{b})'\mathbf{X}'\mathbf{X}(\mathbf{c} - \mathbf{b}).$$

Prove that this difference is positive.

3. **Linear Transformations of the data.** Consider the least squares regression of $\mathbf{y}$ on K variables (with a constant) $\mathbf{X}$. Consider an alternative set of regressors $\mathbf{Z} = \mathbf{XP}$, where $\mathbf{P}$ is a nonsingular matrix. Thus, each column of $\mathbf{Z}$ is a mixture of some of the columns of $\mathbf{X}$. Prove that the residual vectors in the regressions of $\mathbf{y}$ on $\mathbf{X}$ and $\mathbf{y}$ on $\mathbf{Z}$ are identical. What relevance does this have to the question of changing the fit of a regression by changing the units of measurement of the independent variables?
4. **Partial Frisch and Waugh.** In the least squares regression of $\mathbf{y}$ on a constant and $\mathbf{X}$, to compute the regression coefficients on $\mathbf{X}$, we can first transform $\mathbf{y}$ to deviations from the mean $\bar{y}$ and, likewise, transform each column of $\mathbf{X}$ to deviations from the respective column mean; second, regress the transformed $\mathbf{y}$ on the transformed $\mathbf{X}$ without a constant. Do we get the same result if we only transform $\mathbf{y}$? What if we only transform $\mathbf{X}$?
5. **Residual makers.** What is the result of the matrix product $\mathbf{M}_1\mathbf{M}$ where $\mathbf{M}_1$ is defined in (3-19) and $\mathbf{M}$ is defined in (3-14)?
6. **Adding an observation.** A data set consists of n observations on $\mathbf{X}_n$ and $\mathbf{y}_n$. The least squares estimator based on these n observations is $\mathbf{b}_n = (\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'\mathbf{y}_n$. Another observation, $\mathbf{x}_s$ and y_s, becomes available. Prove that the least squares estimator computed using this additional observation is

$$\mathbf{b}_{n,s} = \mathbf{b}_n + \frac{1}{1 + \mathbf{x}_s'(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s}(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{x}_s(y_s - \mathbf{x}_s'\mathbf{b}_n).$$

Note that the last term is e_s, the residual from the prediction of y_s using the coefficients based on $\mathbf{X}_n$ and $\mathbf{b}_n$. Conclude that the new data change the results of least squares only if the new observation on y cannot be perfectly predicted using the information already in hand.

7. **Deleting an observation.** A common strategy for handling a case in which an observation is missing data for one or more variables is to fill those missing variables with 0s and add a variable to the model that takes the value 1 for that one observation and 0 for all other observations. Show that this 'strategy' is equivalent to discarding the observation as regards the computation of $\mathbf{b}$ but it does have an effect on R^2. Consider the special case in which $\mathbf{X}$ contains only a constant and one variable. Show that replacing missing values of x with the mean of the complete observations has the same effect as adding the new variable.
8. **Demand system estimation.** Let Y denote total expenditure on consumer durables, nondurables, and services and E_d, E_n, and E_s are the expenditures on the three

We also considered some practical problems that arise when data are less than perfect for the estimation and analysis of the regression model, including multicollinearity and missing observations.

The formal assumptions of the classical model are pivotal in the results of this chapter. All of them are likely to be violated in more general settings than the one considered here. For example, in most cases examined later in the book, the estimator has a possible bias, but that bias diminishes with increasing sample sizes. Also, we are going to be interested in hypothesis tests of the type considered here, but at the same time, the assumption of normality is narrow, so it will be necessary to extend the model to allow nonnormal disturbances. These and other 'large sample' extensions of the linear model will be considered in Chapter 5.

Key Terms and Concepts

- Assumptions
- Condition number
- Confidence interval
- Estimator
- Gauss-Markov Theorem
- Hat matrix
- Ignorable case
- Linear estimator
- Linear unbiased estimator
- Mean squared error
- Minimum mean squared error
- Minimum variance linear unbiased estimator
- Missing observations
- Multicollinearity
- Oaxaca's decomposition
- Optimal linear predictor
- Orthogonal random variables
- Principal components
- Projection matrix
- Sampling distribution
- Sampling variance
- Semiparametric
- Standard Error
- Standard error of the regression
- Statistical properties
- Stochastic regressors
- t ratio

Exercises

1. Suppose that you have two independent unbiased estimators of the same parameter θ, say $\hat{\theta}_1$ and $\hat{\theta}_2$, with different variances v_1 and v_2. What linear combination $\hat{\theta} = c_1\hat{\theta}_1 + c_2\hat{\theta}_2$ is the minimum variance unbiased estimator of θ?
2. Consider the simple regression $y_i = \beta x_i + \varepsilon_i$ where $E[\varepsilon \mid x] = 0$ and $E[\varepsilon^2 \mid x] = \sigma^2$
   a. What is the minimum mean squared error linear estimator of β? [Hint: Let the estimator be $[\hat{\beta} = \mathbf{c}'\mathbf{y}]$. Choose $\mathbf{c}$ to minimize $\text{Var}[\hat{\beta}] + [E(\hat{\beta} - \beta)]^2$. The answer is a function of the unknown parameters.]
   b. For the estimator in part a, show that ratio of the mean squared error of $\hat{\beta}$ to that of the ordinary least squares estimator b is
   $$\frac{\text{MSE}[\hat{\beta}]}{\text{MSE}[b]} = \frac{\tau^2}{(1+\tau^2)}, \quad \text{where } \tau^2 = \frac{\beta^2}{[\sigma^2/\mathbf{x}'\mathbf{x}]}.$$
   Note that τ is the square of the population analog to the "t ratio" for testing the hypothesis that $\beta = 0$, which is given in (4-14). How do you interpret the behavior of this ratio as $\tau \to \infty$?
3. Suppose that the classical regression model applies but that the true value of the constant is zero. Compare the variance of the least squares slope estimator computed without a constant term with that of the estimator computed with an unnecessary constant term.

4. Suppose that the regression model is $y_i = \alpha + \beta x_i + \varepsilon_i$, where the disturbances ε_i have $f(\varepsilon_i) = (1/\lambda)\exp(-\lambda\varepsilon_i)$, $\varepsilon_i \geq 0$. This model is rather peculiar in that all the disturbances are assumed to be positive. Note that the disturbances have $E[\varepsilon_i \mid x_i] = \lambda$ and $\text{Var}[\varepsilon_i \mid x_i] = \lambda^2$. Show that the least squares slope is unbiased but that the intercept is biased.
5. Prove that the least squares intercept estimator in the classical regression model is the minimum variance linear unbiased estimator.
6. As a profit maximizing monopolist, you face the demand curve $Q = \alpha + \beta P + \varepsilon$. In the past, you have set the following prices and sold the accompanying quantities:

Q	3	3	7	6	10	15	16	13	9	15	9	15	12	18	21
P	18	16	17	12	15	15	4	13	11	6	8	10	7	7	7

Suppose that your marginal cost is 10. Based on the least squares regression, compute a 95 percent confidence interval for the expected value of the profit maximizing output.

7. The following sample moments for $x = [1, x_1, x_2, x_3]$ were computed from 100 observations produced using a random number generator:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 100 & 123 & 96 & 109 \\ 123 & 252 & 125 & 189 \\ 96 & 125 & 167 & 146 \\ 109 & 189 & 146 & 168 \end{bmatrix}, \quad \mathbf{X}'\mathbf{y} = \begin{bmatrix} 460 \\ 810 \\ 615 \\ 712 \end{bmatrix}, \quad \mathbf{y}'\mathbf{y} = 3924.$$

The true model underlying these data is $y = x_1 + x_2 + x_3 + \varepsilon$.
   a. Compute the simple correlations among the regressors.
   b. Compute the ordinary least squares coefficients in the regression of y on a constant x_1, x_2, and x_3.
   c. Compute the ordinary least squares coefficients in the regression of y on a constant x_1 and x_2, on a constant x_1 and x_3, and on a constant x_2 and x_3.
   d. Compute the variance inflation factor associated with each variable.
   e. The regressors are obviously collinear. Which is the problem variable?
8. Consider the multiple regression of $\mathbf{y}$ on K variables $\mathbf{X}$ and an additional variable $\mathbf{z}$. Prove that under the assumptions A1 through A6 of the classical regression model, the true variance of the least squares estimator of the slopes on $\mathbf{X}$ is larger when $\mathbf{z}$ is included in the regression than when it is not. Does the same hold for the sample estimate of this covariance matrix? Why or why not? Assume that $\mathbf{X}$ and $\mathbf{z}$ are nonstochastic and that the coefficient on $\mathbf{z}$ is nonzero.
9. For the classical normal regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with no constant term and K regressors, assuming that the true value of $\boldsymbol{\beta}$ is zero, what is the exact expected value of $F[K, n-K] = (R^2/K)/[(1-R^2)/(n-K)]$?
10. Prove that $E[\mathbf{b}'\mathbf{b}] = \boldsymbol{\beta}'\boldsymbol{\beta} + \sigma^2 \sum_{k=1}^{K}(1/\lambda_k)$ where $\mathbf{b}$ is the ordinary least squares estimator and λ_k is a characteristic root of $\mathbf{X}'\mathbf{X}$.
11. Data on U.S. gasoline consumption for the years 1960 to 1995 are given in Table F2.2.
   a. Compute the multiple regression of per capita consumption of gasoline, G/pop, on all the other explanatory variables, including the time trend, and report all results. Do the signs of the estimates agree with your expectations?