ECON 423 - Regression Forecasting Lab

Introduction

Regression methods are useful tools to forecasters. They are more sophisticated than naive methods because regression models use more information, in the form of explanatory variables, to forecasting applications. Regression models also allow for the incorporation of economic theories into the forecasting process. They also require more data, additional computational burden, and the judgement of the forecaster.

This lab focuses on using regression methods to forecast demand for tickets at Major League Baseball games. You will estimate a simple demand model for tickets using OLS and use the estimated parameters from this to generate *ex post* and *ex ante* forecasts of ticket sales.

Goals

- 1. Estimate a demand function for Major League Baseball ticket sales
- 2. Understand how to use regression models to generate forecasts
- 3. Understand the relationship between regression diagnostic statistics and forecast evaluation tools
- 4. Gain more experience using the Excel Regression Wizard

Data

The Excel file baseball_data.xls contains data on attendance, ticket prices and other factors for two professional baseball teams - the Kansas City Royals and the Philadelphia Phillies - for the 1990 through 2001 baseball seasons. The file contains the following variables:

| Variable | Description |
|------------|--|
| year | Calendar year, 1990-2001 |
| teamname | Name of baseball team |
| avg_attend | Average attendance per game for season |
| price | Average ticket price |
| playoff | Dummy variable, equal 1 if team made the post season in that season |
| strike | Dummy variable, equal 1 if a baseball strike took place in that season |
| wins | Total number of games won in season |
| pct | Percent of games won in season |

Methods

Recall the general form of a simple regression model

$$Y_i = \beta_0 + \beta_1 X_i + e_i \tag{1}$$

where Y_i is the dependent variable, X_i is the independent or explanatory variable, e_i is the unobservable equation error that captures all factors except X_i that affect Y_i , and β_0 and β_1 are unknown parameters to be estimated. e_i is a random variable and by assumption

- 1. $E(e_i) = 0$
- 2. $var(e_i) = \sigma^2$
- 3. $cov(e_i, e_j) = 0$ for $i \neq j$

4.
$$e_i \sim N(0, \sigma^2)$$

Also recall the general form of a multiple regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + e_i.$$
(2)

In this lab, the economic theory underlying the simple regression model is the theory of consumer behavior. In particular, the demand function that emerges from the model of consumer behavior from microeconomics. According to this model, the quantity of any good demanded by consumers varies inversely with the price of that good, and demand for any good at any price also changes in response to changes in factors like income, the price of substitute and complementary goods, tastes and preferences, and other factors. Demand curves slope down and things like income and other prices shift demand curves to the left or right.

In this context, Y_i in equation (1) is demand for attendance at baseball games in season *i* and X_i is the average price of a ticket to games in season *i*.

Procedures

- 1. Create a time series plot of the annual average attendance for the Kansas City Royals over the period 1990-2001. What features do you observe on this graph?
- 2. Create a scatter plot of the annual attendance against the average ticket price for the Kansas City Royals over the period 1990-2001. Make sure that the ticket price variable is graphed on the vertical axis. Describe this graph in terms of the economic theory of consumer behavior.
- 3. Using data for the Kansas City Royals, estimate the demand function

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

```
The results are
SUMMARY OUTPUT
Regression Statistics
Multiple R
                     0.73
R Square
                     0.53
Adjusted R Square
                     0.48
Standard Error
                  2597.01
Observations
                       12
ANOVA
                     SS
            df
                                  MS
Regression
             1
                  74738768
                               74738768
Residual
            10
                  67444569
                                6744457
Total
            11
                142183337
            Coefficients
                             Standard Error t Stat
                 37434.53
                             4920.984
                                               7.607
Intercept
X Variable 1
                 -1611.77
                             484.175
                                              -3.329
```

- 4. Interpret the coefficients of the regression model. Identify the values of: the standard error of Estimate, and the Coefficient of determination.
- 5. The average ticket prices to Royals games were \$12.30 in 2002 and \$12.13 in 2003. Use these values to calculate *ex post* point forecasts for attendance at Royals games in 2002 and 2003. Recall that the formula for an *ex post* point forecast is

$$Y_i^* = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

6. Calculate an interval forecast for 2002. Recall that the formula for an interval forecast is

$$(\hat{\beta}_0 + \hat{\beta}_1 X_i) \pm t_c \sigma_e \sqrt{1 + \frac{1}{n} + \frac{(X_i - \overline{X})^2}{\sum x^2}}$$

- 7. Actual average attendance at Royals games was 16,334 in 2002 and 22,774 in 2003. Use these values to evaluate the *ex post* attendance forecasts. Discuss the results.
- 8. Suppose that you wanted to construct an *ex ante* forecast for 2004 attendance at Royals games. Describe a procedure for generating this forecast.
- 9. What variables other than the average ticket price might help to forecast demand for tickets at baseball games?