

ACE 564
Spring 2006

Lecture 2

Extensions of the Simple Linear Regression
Model II: Regression Through the Origin, Scaling and
Time Series Forecasting

by
Professor Scott H. Irwin

Readings:

Pindyck and Rubinfeld. “Forecasting with a Single-Equation Regression Model,” Chapter 8 in *Econometric Models and Economic Forecasts* (provided by instructor)

Kennedy. “Forecasting,” Chapter 18 in *A Guide to Econometrics*

Regression Through the Origin

There are some occasions where it is appropriate to specify the simple linear regression model without an intercept,

- Capital Asset Pricing Model
- Proportional farm-retail price spreads

In this case, the regression model is

$$y_t = \beta_2 x_t + e_t$$

where e_t and y_t are assumed to be *iid* with the following distributions,

$$e_t \sim N(0, \sigma^2) \quad \text{and} \quad y_t \sim N(\beta_2 x_t, \sigma^2)$$

This statistical model has two unknown parameters, β_2 and σ^2

The least squares estimators are,

$$b_2 = \frac{\sum_{t=1}^T x_t y_t}{\sum_{t=1}^T x_t^2} \quad \text{var}(b_2) = \frac{\sigma^2}{\sum_{t=1}^T x_t^2}$$

and σ^2 is estimated by

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^T \hat{e}_t^2}{T-1}$$

How do the above formulas differ from the case where an intercept is included in the model?

There are some unique features of this specification that should be noted

- Sum of the estimated errors does not have to equal zero
- R^2 can be negative

When regression is forced through the origin, many computer software packages print out a different measure of goodness of fit,

$$\text{raw } R^2 = \frac{\sum_{t=1}^T (x_t y_t)^2}{\sum_{t=1}^T x_t^2 \sum_{t=1}^T y_t^2}$$

The term “raw” refers to the fact that the sums of square are not mean-corrected

Because of these special features, a zero-intercept regression model should only be estimated when there is a strong [theoretical](#) justification

Sticking to the intercept model has two advantages

- If intercept is insignificantly different from zero, basically have a zero intercept model
- Avoids [specification](#) error if in fact an intercept should be included in the model

Gujarati and Mirer have excellent discussions of these issues

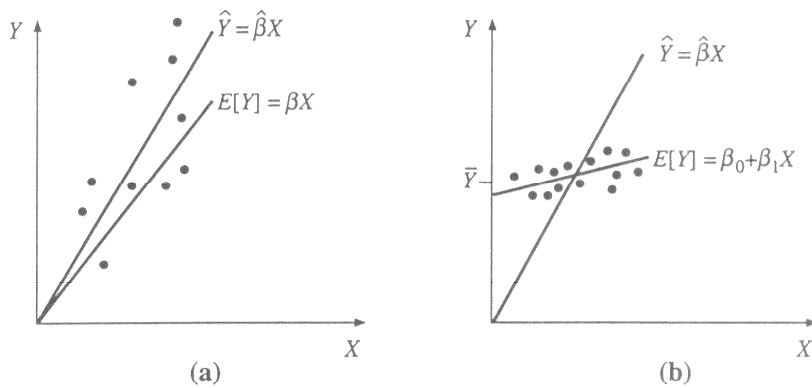


FIGURE 5.6 Part (a) shows the case of regression through the origin. The data are generated around a true regression that has a zero intercept, and the estimated regression is restricted to have a zero intercept also. In (b) the data are generated around a true regression that has a positive intercept, but an estimated regression through the origin is fit to the data. In such a situation, the estimated slope $\hat{\beta}$ tends to be larger than the true slope β_1 .

Mirer, Thad W. Economic Statistics and Econometrics, Third Edition. Prentice Hall, Englewood Cliffs, NJ. 1995.

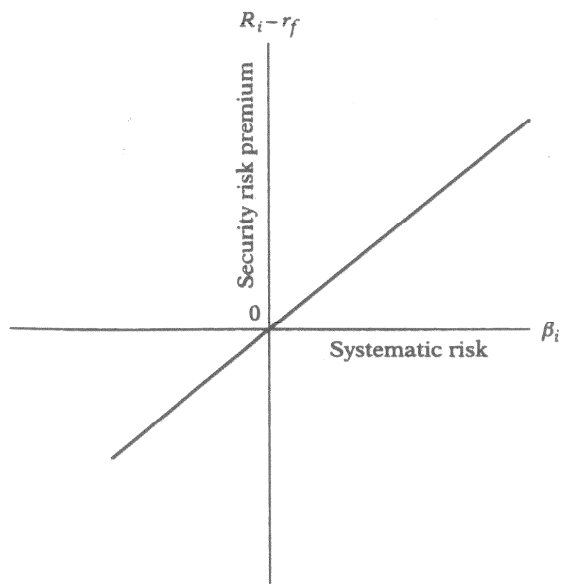


FIGURE 6.2
The Market Model of Portfolio Theory
(assuming $\alpha_i = 0$).

Gujarati, D.N. Basic Econometrics. McGraw-Hill, Inc. New York, NY. 1995.

TABLE 6.1
Annual rates of return
on Afuture Fund and on
the Fisher Index (market
portfolio), 1971–1980

Year	Return on Afuture Fund, % Y	Return on Fisher Index, % X
1971	67.5	19.5
1972	19.2	8.5
1973	-35.2	-29.3
1974	-42.0	-26.5
1975	63.7	61.9
1976	19.3	45.5
1977	3.6	9.5
1978	20.0	14.0
1979	40.3	35.3
1980	37.5	31.0

Source: Haim Levy and Marshall Sarnat, *Portfolio and Investment Selection: Theory and Practice*, Prentice-Hall International, Englewood Cliffs, N.J., 1984, pp. 730 and 738. These data were obtained by the authors from Weisenberg Investment Service, *Investment Companies*, 1981 edition.

Gujarati, D.N. *Basic Econometrics*. McGraw-Hill, Inc. New York, NY. 1995.

Scaling and Units of Measurement

The data we obtain for analysis may not always be in a convenient form for data entry or presentation in tables

- $x_t = \text{US corn production}$
- 1998, $x_t = 9,761,000,000$ bushels

If we divide x_t by 1 billion, then

$$x_t^* = \frac{x_t}{1,000,000,000} = 9.761 \text{ billion bushels}$$

Choice of scale is determined by the researcher in order to make interpretation meaningful and convenient

What is the effect of choice of scale on the simple linear regression model?

- Does not affect the estimation of the underlying relationship
- It may affect the interpretation of coefficient estimates and statistics

Start by considering the simple linear regression model in some "original" units,

$$y_t = \beta_1 + \beta_2 x_t + e_t$$

After estimation, we can write,

$$y_t = b_1 + b_2 x_t + \hat{e}_t$$

Let's now consider changing the scale of the dependent variable,

$$y_t^* = \frac{y_t}{c_1}$$

In order for the estimated regression model to remain valid we must divide all the other terms by c_1 ,

$$\frac{y_t}{c_1} = \frac{b_1}{c_1} + \frac{b_2}{c_1} x_t + \frac{\hat{e}_t}{c_1}$$

or,

$$y_t^* = b_1^* + b_2^* x_t + \hat{e}_t^*$$

where $y_t^* = \frac{y_t}{c_1}$, $b_1^* = \frac{b_1}{c_1}$, $b_2^* = \frac{b_2}{c_1}$ and $\hat{e}_t^* = \frac{\hat{e}_t}{c_1}$

Now consider changing the scale of the independent variable,

$$x_t^* = \frac{x_t}{c_2}$$

In order for the regression model to remain valid we must multiply the slope parameter by c_2 ,

$$y_t = b_1 + c_2 b_2 \frac{x_t}{c_2} + \hat{e}_t$$

or,

$$y_t = b_1 + b_2^* x_t^* + \hat{e}_t$$

where $b_2^* = c_2 b_2$ and $x_t^* = \frac{x_t}{c_2}$

Now consider changing the scale of both the dependent and independent variables,

$$y_t^* = \frac{y_t}{c_1} \text{ and } x_t^* = \frac{x_t}{c_2}$$

We can combine the previous results as follows,

$$\frac{y_t}{c_1} = \frac{b_1}{c_1} + \frac{c_2}{c_1} b_2 \frac{x_t}{c_2} + \frac{\hat{e}_t}{c_1}$$

or,

$$y_t^* = b_1^* + b_2^* x_t^* + \hat{e}_t^*$$

where $y_t^* = \frac{y_t}{c_1}$, $b_1^* = \frac{b_1}{c_1}$, $b_2^* = \frac{c_2}{c_1} b_2$, $x_t^* = \frac{x_t}{c_2}$ and

$$\hat{e}_t^* = \frac{\hat{e}_t}{c_1}$$

We can summarize our results as follows,

If the scaled variables and regressions are,

$$y_t^* = \frac{y_t}{c_1} \text{ and } x_t^* = \frac{x_t}{c_2}$$

$$y_t^* = b_1^* + b_2^* x_t^* + \hat{e}_t^* \quad y_t = b_1 + b_2 x_t + \hat{e}_t$$

Then the following relationships hold,

For parameter estimates

$$b_1^* = \frac{b_1}{c_1} \quad b_2^* = \frac{c_2}{c_1} b_2,$$

For the estimated variance of the regressions

$$\hat{\sigma}^{2*} = \frac{\hat{\sigma}^2}{c_1^2}$$

For the estimated standard error of parameter estimates

$$\hat{se}(b_1^*) = \frac{\hat{se}(b_1)}{c_1} \qquad \hat{se}(b_2^*) = \frac{c_2}{c_1} \hat{se}(b_2)$$

For t -statistics

$$t_{b_1}^* = t_{b_1} \qquad t_{b_2}^* = t_{b_2}$$

For the coefficient of determination

$$R^{2*} = R^2$$

Examples

GPDI (y_t) and GNP (x_t) in millions of dollars,

$$\hat{y}_t = -37,001.5205 + 0.17395x_t$$

GPDI (y_t) in billions and GNP (x_t) in millions of dollars,

$$\hat{y}_t = \underline{\hspace{2cm}} + \underline{\hspace{2cm}} x_t$$

GPDI (y_t) in millions and GNP (x_t) in billions of dollars

$$\hat{y}_t = \underline{\hspace{2cm}} + \underline{\hspace{2cm}} x_t$$

GPDI (y_t) in billions and GNP (x_t) in billions of dollars

$$\hat{y}_t = \underline{\hspace{2cm}} + \underline{\hspace{2cm}} x_t$$

TABLE 6.2
Gross private domestic investment (GPDI) and Gross National Product (GNP) in 1972 dollars, United States, 1974–1983

Year	GPDI (billions of 1972 dollars) (1)	GPDI (millions of 1972 dollars) (2)	GNP (billions of 1972 dollars) (3)	GNP (millions of 1972 dollars) (4)
1974	195.5	195,500	1246.3	1,246,300
1975	154.8	154,800	1231.6	1,231,600
1976	184.5	184,500	1298.2	1,298,200
1977	214.2	214,200	1369.7	1,369,700
1978	236.7	236,700	1438.6	1,438,600
1979	236.3	236,300	1479.4	1,479,400
1980	208.5	208,500	1475.0	1,475,000
1981	230.9	230,900	1512.2	1,512,200
1982	194.3	194,300	1480.0	1,480,000
1983	221.0	221,000	1534.7	1,534,700

Source: *Economic Report of the President*, 1985, p. 234 (for data expressed in billions of dollars).

Gujarati, D.N. Basic Econometrics. McGraw-Hill, Inc. New York, NY. 1995.

Time Series Forecasting

A cross-sectional prediction exercise was conducted in a previous section of the course (ACE 562: Lecture 8)

Most applications of regression in agricultural economics involve time series forecasts

- Forecast the price of hogs for the next month
- Forecast the size of the orange crop next year

Some specialized terminology has developed with respect to time series forecasting:

The in-sample period is the past time interval used to estimate a regression

- In-sample forecasts are generated for the period in which the model is estimated
- Out-of-sample forecasts are generated for periods outside of the estimation period

Two types of out-of-sample forecasts,

- An ex post forecast is made over some past time interval

Part of the original sample period in a study is “saved” for generating forecasts

Forecasts compared to actual values of dependent variable

Not a “true” forecast, but highly useful for analyzing the predictive accuracy of a regression model

- An ex ante forecast is made for some period in the future

True forecasts in the sense of making a statement about the unknown future

Forecasts are compared to actual values of dependent variable

May have to wait for realizations of dependent variable to occur before forecast error comparisons can be made

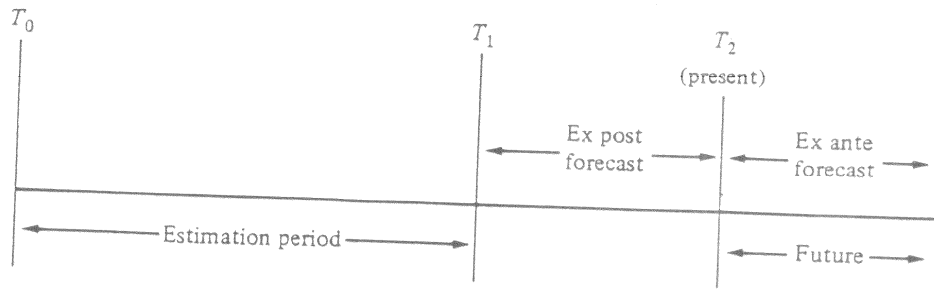


Figure 18.1 Types of Forecasts.

Johnson, A.C., Jr., M.B. Johnson and R.C. Buse. Econometrics: Basic and Applied. Macmillan Publishing Company, New York, NY, 1989.

Two more types of forecasts,

Unconditional forecast: value of the independent variable is known with certainty at the time forecast is generated

Conditional forecast: value of the independent variable is not known with certainty at the time forecast is generated

- Note, it is possible to generate unconditional or conditional *ex post* forecasts, depending on the specification of the forecast regression (e.g lags)
- It is also possible to generate unconditional or conditional *ex ante* forecasts, depending on the specification of the forecast regression (e.g lags)

When considering the properties of time series forecasts we will use a different notation that makes more explicit the time dimension

	Cross-Section	Time Series
Independent Variable	x_0	x_{t+1}
Dependent Variable	\hat{y}_0	\hat{y}_{t+1}

Case 1: β_1 , β_2 , σ^2 and x_{t+1} Known

Consider the following linear statistical model

$$y_t = \beta_1 + \beta_2 x_t + e_t$$

where e_t and y_t are assumed to be *iid* with the following distributions,

$$e_t \sim N(0, \sigma^2) \quad \text{and} \quad y_t \sim N(\beta_1 + \beta_2 x_t, \sigma^2)$$

It is important to note that we are assuming that β_1 , β_2 and σ^2 are known

Since the model holds for all t , we can write the following equation for one-period in the future

$$y_{t+1} = \beta_1 + \beta_2 x_{t+1} + e_{t+1}$$

and by assuming our best guess of the regression error in the future is zero, the least squares forecast for the next period, $t+1$, is

$$\hat{y}_{t+1} = \beta_1 + \beta_2 x_{t+1}$$

Even in this unrealistic case, the least squares forecast will not be exactly correct due to the error term in the regression

The forecast error is defined as the difference between the actual y in $t+1$ and the forecast of y made at t for $t+1$

$$f = y_{t+1} - \hat{y}_{t+1} = (\beta_1 + \beta_2 x_{t+1} + e_{t+1}) - (\beta_1 + \beta_2 x_{t+1}) = e_{t+1}$$

The forecast error in this case is exactly equal to the regression error term

Based on the above relationship, we can examine some important properties of the forecast error

Mean forecast error

We can examine the average value of the forecast error that we should expect,

$$E(f) = E(y_{t+1} - \hat{y}_{t+1}) = E(e_{t+1}) = 0$$

This shows that the least square forecast is an unbiased forecast, in the repeated sampling sense

Variance of the forecast error

While the least squares forecast is unbiased, it may still be wide of the mark for any particular forecast

The “reliability” of the forecast is measured by the variance of the forecast

$$\text{var}(f) = E(y_{t+1} - \hat{y}_{t+1})^2 = E(e_{t+1}^2) = \sigma^2$$

Shows that the variance of the forecast error is exactly equal to the variance of the regression error term (also assumed to be known)

Standard error of the forecast error

$$se(f) = \sqrt{\text{var}(f)} = \sqrt{\sigma^2} = \sigma$$

95% confidence interval for forecast

We can construct a standard normal random variable as follows,

$$Z_f = \frac{y_{t+1} - \hat{y}_{t+1}}{\sqrt{\text{var}(f)}} = \frac{f}{\sigma} \sim N(0,1)$$

Since Z_f is a standard, normal random variable, we can write

$$P[-1.96 \leq Z_f \leq 1.96] = 0.95$$

Substituting for Z_f

$$P[-1.96 \leq \frac{y_{t+1} - \hat{y}_{t+1}}{\sigma} \leq 1.96] = 0.95$$

Multiply the inequality in the brackets by σ

$$P[-1.96\sigma \leq y_{t+1} - \hat{y}_{t+1} \leq 1.96\sigma] = 0.95$$

Now, add \hat{y}_{t+1} to each term

$$P[\hat{y}_{t+1} - 1.96\sigma \leq y_{t+1} \leq \hat{y}_{t+1} + 1.96\sigma] = 0.95$$

Hence, the 95 percent forecast confidence interval is

$$\hat{y}_{t+1} \pm 1.96\sigma$$

We can generalize to any forecast confidence level, $1 - \alpha$, as follows

$$P[\hat{y}_{t+1} - Z_{\alpha/2}\sigma \leq y_{t+1} \leq \hat{y}_{t+1} + Z_{\alpha/2}\sigma] = 1 - \alpha$$

and

$$\hat{y}_{t+1} \pm Z_{\alpha/2}\sigma$$

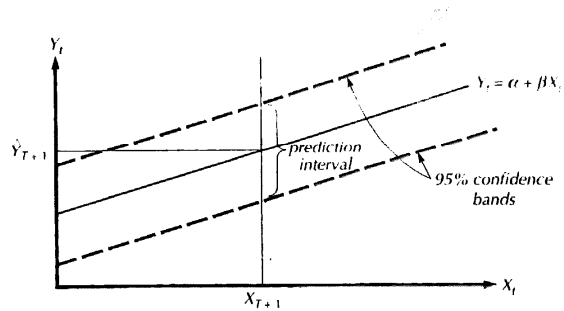


FIGURE 8.2
Forecast when equation parameters are known.

Source: Pindyck, R.S. and D.L. Rubinfeld. *Econometric Models and Economic Forecasts*, Fourth Edition, Boston: Irwin McGraw-Hill, 1998.

Case 2: β_1, β_2 Estimated; σ^2 and x_{t+1} Known

We start with the same linear statistical model

$$y_t = \beta_1 + \beta_2 x_t + e_t$$

But, since the parameters are unknown, we must estimate them with b_1 and b_2

Again assuming that the best guess of the error in the next period is zero, the least squares forecast is

$$\hat{y}_{t+1} = b_1 + b_2 x_{t+1}$$

The forecast error is defined as

$$f = y_{t+1} - \hat{y}_{t+1} = (\beta_1 + \beta_2 x_{t+1} + e_{t+1}) - (b_1 + b_2 x_{t+1})$$

which can be re-written:

$$f = y_{t+1} - \hat{y}_{t+1} = (\beta_1 - b_1) + (\beta_2 - b_2)x_{t+1} + e_{t+1}$$

Note that the forecast error in this case does not simply equal the regression error term

- The forecast error is now a function of three random variables, b_1 , b_2 , and e_{t+1}

Based on the above relationship, we can again examine important properties of the forecast error

Mean forecast error

The average value of the forecast error that we should expect is

$$\begin{aligned} E(f) &= E(y_{t+1} - \hat{y}_{t+1}) = E[(\beta_1 - b_1) + (\beta_2 - b_2)x_{t+1} + e_{t+1}] \\ &= [\beta_1 - E(b_1)] + [\beta_2 - E(b_2)]x_{t+1} + E(e_{t+1}) \\ &= [\beta_1 - \beta_1] + [\beta_2 - \beta_2]x_{t+1} + E(e_{t+1}) \\ &= E(e_{t+1}) = 0 \end{aligned}$$

This shows that even when the parameters have to be estimated the least square forecast is unbiased

Variance of the forecast error

While the least squares forecast is unbiased, it may still be wide of the mark for any particular forecast

The “reliability” of the forecast is measured by the variance of the forecast error

$$\text{var}(f) = E(y_{t+1} - \hat{y}_{t+1})^2 = E[(\beta_1 - b_1) + (\beta_2 - b_2)x_{t+1} + e_{t+1}]^2$$

Expanding the square

$$\begin{aligned}\text{var}(f) &= E[(\beta_1 - b_1)^2 + ((\beta_2 - b_2)x_{t+1})^2 + e_{t+1}^2 \\ &+ 2(\beta_1 - b_1)(\beta_2 - b_2)x_{t+1} + 2(\beta_2 - b_2)x_{t+1}e_{t+1} + (\beta_1 - b_1)e_{t+1}]\end{aligned}$$

Take the expectations through to each term

$$\begin{aligned}\text{var}(f) &= E[(\beta_1 - b_1)^2] + E[((\beta_2 - b_2)x_{t+1})^2] + E[e_{t+1}^2] \\ &+ 2E[(\beta_1 - b_1)(\beta_2 - b_2)x_{t+1}] + 2E[(\beta_2 - b_2)x_{t+1}e_{t+1}] + E[(\beta_1 - b_1)e_{t+1}]\end{aligned}$$

Which reduces to

$$\begin{aligned}\text{var}(f) &= E[(\beta_1 - b_1)^2] + E[((\beta_2 - b_2)x_{t+1})^2] + E[e_{t+1}^2] \\ &+ 2E[(\beta_1 - b_1)(\beta_2 - b_2)x_{t+1}]\end{aligned}$$

Now change the notation

$$\text{var}(f) = \text{var}(b_1) + \text{var}(b_2)x_{t+1}^2 + 2\text{cov}(b_1, b_2)x_{t+1} + \sigma^2$$

The next step is to substitute the definitions of $\text{var}(b_1)$, $\text{var}(b_2)$, and $\text{cov}(b_1, b_2)$ that we derived earlier

$$\begin{aligned} \text{var}(f) = & \sigma^2 \left[\frac{\sum_{t=1}^T x_t^2}{T \sum_{t=1}^T (x_t - \bar{x})^2} \right] + \sigma^2 \left[\frac{1}{\sum_{t=1}^T (x_t - \bar{x})^2} \right] x_{t+1}^2 \\ & + 2\sigma^2 \left[\frac{-\bar{x}}{\sum_{t=1}^T (x_t - \bar{x})^2} \right] x_{t+1} + \sigma^2 \end{aligned}$$

After some fairly tedious algebra, this can be reduced to

$$\text{var}(f) = \sigma^2 \left[1 + \frac{1}{T} + \frac{(x_{t+1} - \bar{x})^2}{\sum_{t=1}^T (x_t - \bar{x})^2} \right]$$

$$\text{var}(f) = \sigma^2 \left[1 + \frac{1}{T} + \frac{(x_{t+1} - \bar{x})^2}{\sum_{t=1}^T (x_t - \bar{x})^2} \right]$$

Key points:

- Since term in brackets must be greater than one, forecast error variance is larger than variance of the regression
- Reflects fact that forecast error is influenced not only by the regression error, but also that parameters must now be estimated
- The greater the distance between the mean of x and x_{t+1} , the greater the variance of the forecast error
- In other words, the more distant is the observation for the independent variable from its mean, the more uncertain is the forecast
- All else constant, the larger the sample, the smaller the variance of the forecast error

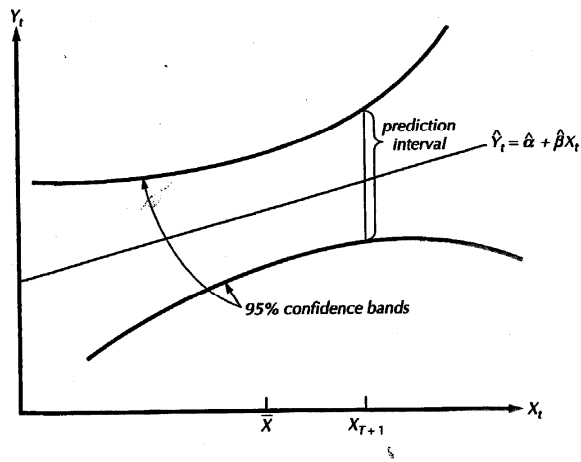
Standard error of the forecast

$$se(f) = \sqrt{\text{var}(f)}$$

95% confidence interval for forecast

$$\hat{y}_{t+1} \pm 1.96 \cdot se(f)$$

FIGURE 8.3
Forecast confidence intervals.



Source: Pindyck, R.S. and D.L. Rubinfeld. *Econometric Models and Economic Forecasts*, Fourth Edition, Boston: Irwin McGraw-Hill, 1998.

Case 3: β_1 , β_2 and σ^2 Estimated; x_{t+1} Known

In practice, the variance of the regression, σ^2 , is rarely known

We must replace σ^2 by its estimator $\hat{\sigma}^2$

The model and all other assumptions remain the same

Mean forecast error

$$\begin{aligned} E(y_{t+1} - \hat{y}_{t+1}) &= E[(\beta_1 - b_1) + (\beta_2 - b_2)x_{t+1} + e_{t+1}] \\ &= E(e_{t+1}) = 0 \end{aligned}$$

Variance of the forecast error

$$\hat{\text{var}}(f) = \hat{\sigma}^2 \left[1 + \frac{1}{T} + \frac{(x_{t+1} - \bar{x})^2}{\sum_{t=1}^T (x_t - \bar{x})^2} \right]$$

Standard error of the forecast

$$\hat{s}e(f) = \sqrt{\hat{\text{var}}(f)}$$

95% confidence interval for forecast

Previously, we constructed a standard normal random variable as follows

$$Z_f = \frac{y_{t+1} - \hat{y}_{t+1}}{\sqrt{\text{var}(f)}} \sim N(0,1)$$

But we now must replace $\text{var}(f)$ with its estimate $\hat{\text{var}}(f)$, which results in a t -distributed random variable

$$t_f = \frac{y_{t+1} - \hat{y}_{t+1}}{\sqrt{\hat{\text{var}}(f)}} = \frac{y_{t+1} - \hat{y}_{t+1}}{\hat{se}(f)} \sim t_{\alpha/2, T-2}$$

Since t_f is a t -distributed random variable, we can write

$$P[\hat{y}_{t+1} - t_{\alpha/2, T-2} \hat{se}(f) \leq y_{t+1} \leq \hat{y}_{t+1} + t_{\alpha/2, T-2} \hat{se}(f)] = 1 - \alpha$$

and,

$$\hat{y}_{t+1} \pm t_{\alpha/2, T-2} \hat{se}(f)$$

Case 4: β_1 , β_2 , σ^2 and x_{t+1} Estimated

Unfortunately, in most actual forecasting situations, circumstances are more difficult than in the previous three cases

We must estimate the parameters, regression variance and the value of the independent variable in $t+1$

Example: Assume a simple linear statistical model describes the annual demand for beef, where quantity is the dependent variable and price is the independent variable

⇒ To forecast the quantity demanded of beef in 2006, we have to forecast the price of beef in 2006

We start with the same linear statistical model

$$y_t = \beta_1 + \beta_2 x_t + e_t$$

Since the model holds for all t , we can write the following equation for one-period in the future ($t+1=2006$)

$$y_{t+1} = \beta_1 + \beta_2 x_{t+1} + e_{t+1}$$

By assuming our best guess of the regression error in the future is zero and using the least squares estimators for the intercept and slope, we would usually form the forecast for the next period as

$$\hat{y}_{t+1} = b_1 + b_2 x_{t+1}$$

But, x_{t+1} is not known at time t , when we are making the forecast

So, some type of forecast for x_{t+1} must be used

$$\hat{y}_{t+1} = b_1 + b_2 x_{t+1}^*$$

The uncertainty inherent in the independent variable can be modeled as

$$x_{t+1}^* = x_{t+1} + v_{t+1} \quad \text{where} \quad v_{t+1} \sim N(0, \sigma_v^2)$$

To summarize, in this case there are three sources of forecast uncertainty

- Regression error term
- Estimation of the parameters
- Forecast of the independent variable

Once again, the forecast error is defined as the difference between the actual y in $t+1$ and the forecast of y made at t for $t+1$

$$f = y_{t+1} - \hat{y}_{t+1} = (\beta_1 + \beta_2 x_{t+1} + e_{t+1}) - (b_1 + b_2 x_{t+1}^*)$$

which can be re-written

$$f = y_{t+1} - \hat{y}_{t+1} = (\beta_1 - b_1) + (\beta_2 x_{t+1} - b_2 x_{t+1}^*) + e_{t+1}$$

Substituting for x_{t+1}^*

$$f = y_{t+1} - \hat{y}_{t+1} = (\beta_1 - b_1) + (\beta_2 x_{t+1} - b_2 (x_{t+1} + v_{t+1})) + e_{t+1}$$

$$f = y_{t+1} - \hat{y}_{t+1} = (\beta_1 - b_1) + (\beta_2 - b_2)x_{t+1} - b_2 v_{t+1} + e_{t+1}$$

Mean forecast error

$$\begin{aligned} E(f) &= E(y_{t+1} - \hat{y}_{t+1}) = E[(\beta_1 - b_1) + (\beta_2 - b_2)x_{t+1} - b_2v_{t+1} \\ &\quad + e_{t+1}] \\ &= [\beta_1 - E(b_1)] + [\beta_2 - E(b_2)]x_{t+1} + \\ &\quad - b_2E(v_{t+1}) + E(e_{t+1}) \\ &= 0 \end{aligned}$$

So, despite having to predict the value of the independent variable, the forecast is still unbiased

- Of course, this depends on the forecast of the independent variable being unbiased itself!!

Variance of the forecast error

$$\begin{aligned} \text{var}(f) &= E(y_{t+1} - \hat{y}_{t+1})^2 = E[(\beta_1 - b_1) + (\beta_2 - b_2)x_{t+1} - b_2v_{t+1} \\ &\quad + e_{t+1}]^2 \end{aligned}$$

Which, after a very long series of algebraic manipulations, can be written as

$$\text{var}(f) = \sigma^2 \left[1 + \frac{1}{T} + \frac{(x_{t+1} - \bar{x})^2}{\sum_{t=1}^T (x_t - \bar{x})^2} + \frac{\sigma_v^2}{\sum_{t=1}^T (x_t - \bar{x})^2} \right] + \beta_2^2 \sigma_v^2$$

It is useful to re-state the variance of the forecast error from Case 2 (x_{t+1} known) for comparison

$$\text{var}(f) = E(y_{t+1} - \hat{y}_{t+1})^2 = \sigma^2 \left[1 + \frac{1}{T} + \frac{(x_{t+1} - \bar{x})^2}{\sum_{t=1}^T (x_t - \bar{x})^2} \right]$$

Note that the additional terms are all positive, so the variance of the forecast error in Case 4 is larger than for Case 2

This makes good sense, as there is an added source of uncertainty due to the necessity of forecasting the value of the independent variable

In Case 3, we recognized that the variance of the regression error term must be estimated

We can replace σ^2 with $\hat{\sigma}^2$ in the equation for the forecast error variance in Case 4 as well

$$\text{vâr}(f) = \hat{\sigma}^2 \left[1 + \frac{1}{T} + \frac{(x_{t+1} - \bar{x})^2}{\sum_{t=1}^T (x_t - \bar{x})^2} + \frac{\sigma_v^2}{\sum_{t=1}^T (x_t - \bar{x})^2} \right] + \beta_2^2 \sigma_v^2$$

However, we still cannot compute an estimate because the formula depends on both β_2 and the unknown variance of x_{t+1}^* (σ_v^2)

Since we have already established that b_2 is the “best” estimator of β_2 , we can simply replace β_2 in the formula with b_2

$$\text{vâr}(f) = \hat{\sigma}^2 \left[1 + \frac{1}{T} + \frac{(x_{t+1} - \bar{x})^2}{\sum_{t=1}^T (x_t - \bar{x})^2} + \frac{\sigma_v^2}{\sum_{t=1}^T (x_t - \bar{x})^2} \right] + b_2^2 \sigma_v^2$$

Unfortunately, σ_v^2 is a difficult parameter to estimate, which makes it quite difficult to complete the estimator of the forecast error variance in the present case (4)

In formal terms, no [solution](#) to this problem

One “second best” solution: somehow estimate σ_v^2

- Expert judgment, previous studies
- Highly difficult to implement even this rough approximation

The favored “second-best” solution: use the forecast error variance formula for Case 3 where x_{t+1} is assumed known

- Obviously, [understates](#) the true forecast error variance
- Does provide a [lower-bound](#) for the forecast error variance and resulting confidence bands
- Must remember to alert readers to this [approximation](#) and the caution that should be applied to any estimated forecast confidence intervals

TABLE 3.4
 U.S. coffee consumption (Y) in relation to
 average real retail price (X),* 1970–1980

Year	Y (cups per person per day)	X (\$ per lb)
1970	2.57	0.77
1971	2.50	0.74
1972	2.35	0.72
1973	2.30	0.73
1974	2.25	0.76
1975	2.20	0.75
1976	2.11	1.08
1977	1.94	1.81
1978	1.97	1.39
1979	2.06	1.20
1980	2.02	1.17

*Note: The nominal price was divided by the Consumer Price Index (CPI) for food and beverages, 1967 = 100.

Source: The data for Y are from *Summary of National Coffee Drinking Study*, Data Group, Elkins Park, Penn., 1981; and the data on nominal X (i.e., X in current prices) are from *Nielsen Food Index*, A. C. Nielsen, New York, 1981.

²⁷A formal test of the significance of r^2 will be presented in Chap. 8.

²⁸I am indebted to Scott E. Sandberg for collecting the data.

Coffee Forecasting Example

In-sample estimation results for 1970-1978,

$$\begin{array}{rcc} \hat{y}_t = 2.68 - 0.45x_t & R^2 = 0.65 & \\ (20.77) \quad (-3.63) & (t\text{-stat.}) & \\ [0.0000] \quad [0.0084] & [p\text{-value}] & \end{array}$$

*Out-of-sample forecasts for 1979 and 1980
(units: c./p./d.),*

1979:

$$\hat{y}_{t+1} = 2.68 - 0.45x_{t+1} = 2.68 - 0.45 \cdot 1.20 = 2.14$$

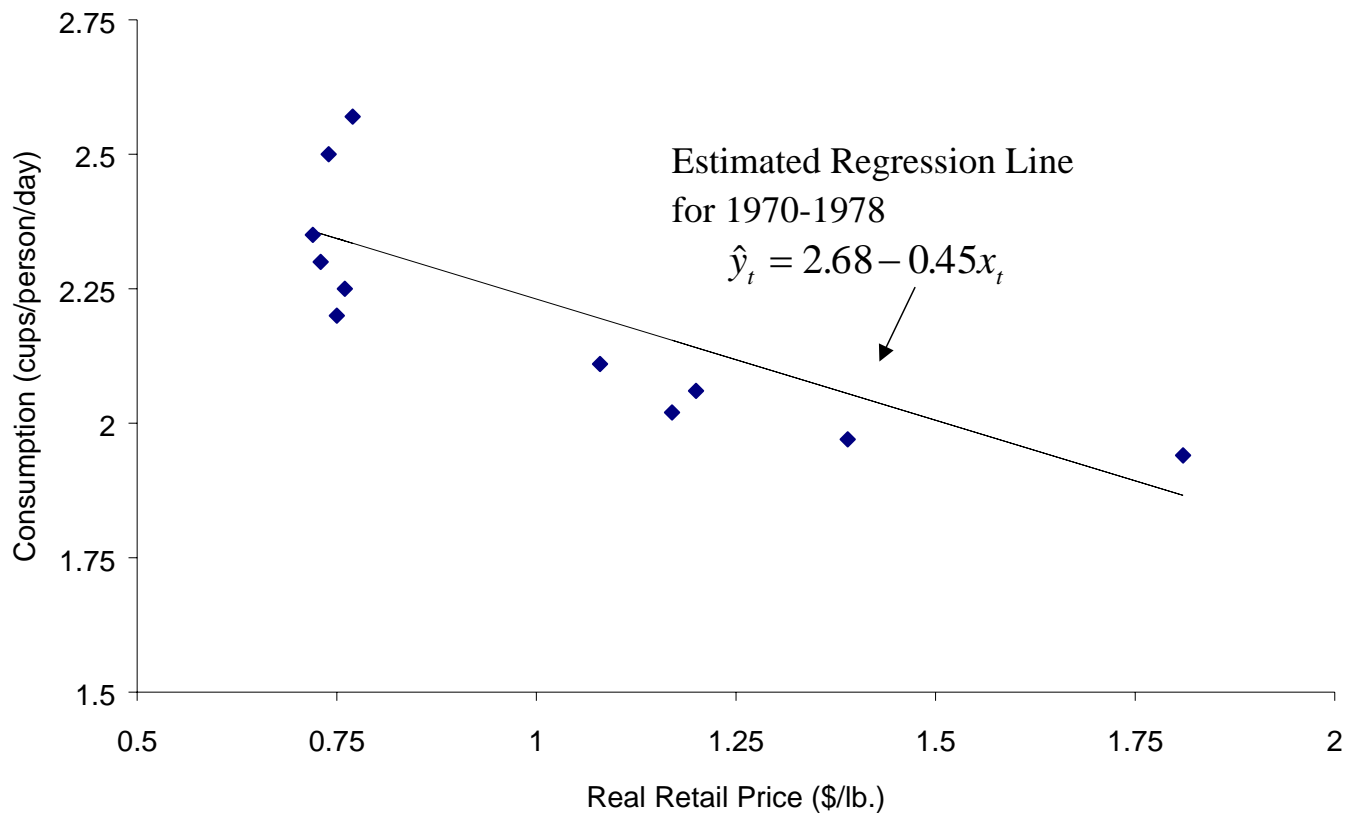
1980:

$$\hat{y}_{t+2} = 2.68 - 0.45x_{t+2} = 2.68 - 0.45 \cdot 1.17 = 2.15$$

Are these ex post or ex ante out-of-sample forecasts?

Are these unconditional or conditional forecasts?

US Coffee Consumption and Real Retail Price, 1970-1980



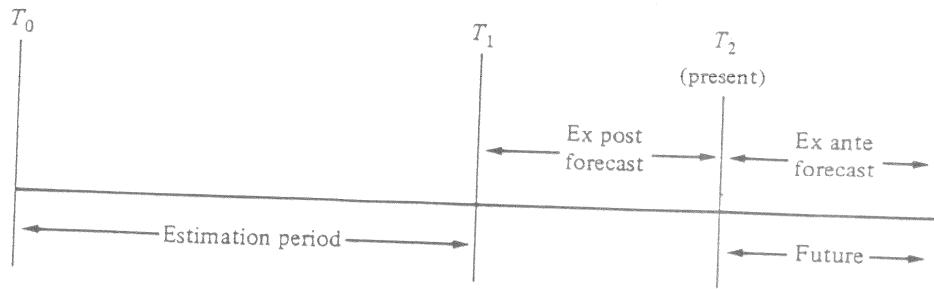


Figure 18.1 Types of Forecasts.

Johnson, A.C., Jr., M.B. Johnson and R.C. Buse. Econometrics: Basic and Applied. Macmillan Publishing Company, New York, NY, 1989.

Variance of forecast error (units: c./p./d.²),

1979:

$$\hat{\text{var}}(f) = 0.0185 \left[1 + 0.1111 + \frac{(1.20 - 0.9722)^2}{1.1996} \right] = 0.0214$$

1980:

$$\hat{\text{var}}(f) = 0.0185 \left[1 + 0.1111 + \frac{(1.17 - 0.9722)^2}{1.1996} \right] = 0.0212$$

Standard error of the forecasts (units: c./p./d.),

1979:

$$\hat{s}e(f) = \sqrt{\hat{\text{var}}(f)} = \sqrt{0.0214} = 0.1463$$

1980:

$$\hat{s}e(f) = \sqrt{\hat{\text{var}}(f)} = \sqrt{0.0212} = 0.1456$$

95% CI for forecasts (units: c./p./d.),

1979:

$$2.14 \pm 2.365 \cdot 0.1463$$

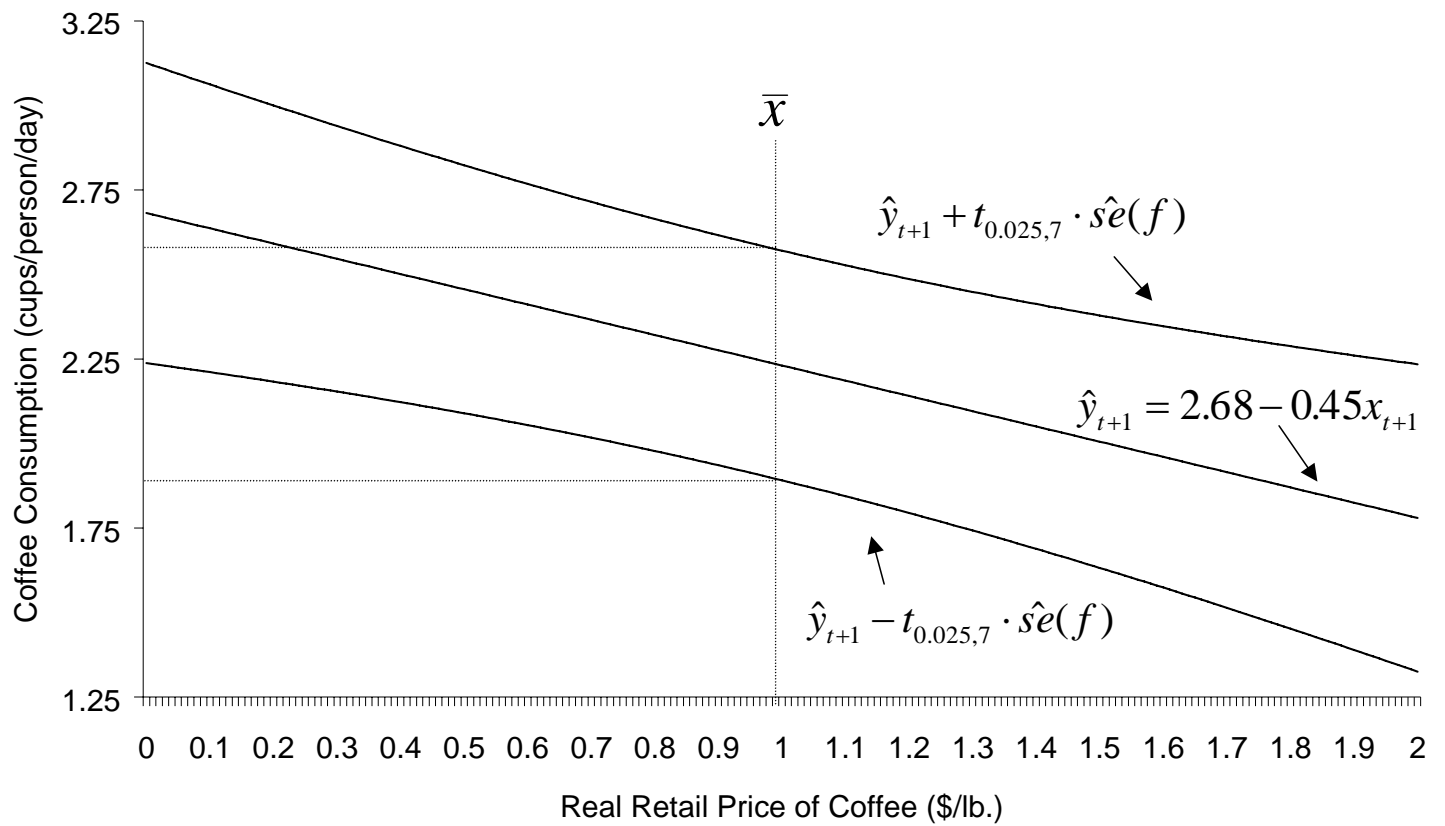
$$(1.79 \leq \hat{y}_{t+1} \leq 2.49)$$

1980:

$$2.15 \pm 2.365 \cdot 0.1456$$

$$(1.81 \leq \hat{y}_{t+1} \leq 2.50)$$

Forecast Confidence Intervals for Coffee Consumption



Forecast error (units: c./p./d.),

1979:

$$f = y_{t+1} - \hat{y}_{t+1} = 2.06 - 2.14 = -0.08$$

1980:

$$f = y_{t+1} - \hat{y}_{t+1} = 2.02 - 2.15 = -0.13$$

Measures of forecast accuracy,

Mean forecast error (units: c./p./d.):

$$ME = \frac{1}{N} \sum_{k=1}^N (y_{t+k} - \hat{y}_{t+k}) = \frac{1}{2} (-0.08 + -0.13) = -0.10$$

Mean square forecast error (units: c./p./d.²):

$$MSE = \frac{1}{N} \sum_{k=1}^N (y_{t+k} - \hat{y}_{t+k})^2 = \frac{1}{2} (-0.08^2 - 0.13^2) = 0.0116$$

Root mean square forecast error (units: c./p./d.):

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N (y_{t+k} - \hat{y}_{t+k})^2} = \sqrt{0.0116} = 0.1079$$