Chapter4
Predictions In Linear Regression Model

Prediction of values of study variable

An important use of linear regression modeling is to predict the average and actual values of the study variable. The term prediction of the value of study variable corresponds to knowing the value of $E(y)$ (in case of average value) and value of y (in case of actual value) for a given value of the explanatory variable. We consider both cases. The prediction of values consists of two steps. In the first step, the regression coefficients are estimated on the basis of given observations. In the second step, these estimators are then used to construct the predictor which provides the prediction of actual or average values of study variables. Based on this approach of construction of predictors, there are two situations in which the actual and average values of the study variable can be predicted- within sample prediction and outside sample prediction. We describe the prediction in both situations.

Within sample prediction in simple linear regression model

Consider the linear regression model $y = \beta_0 + \beta_1 x + \varepsilon.$ Based on a sample of n sets of paired observations (x_i, y_i) $(i = 1,2,...,n)$ following $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where ε_i's are identically and independently distributed following $N(0, \sigma^2)$. The parameters β_0 and β_1 are estimated using the ordinary least squares estimation as b_0 of β_0 and b_1 of β_1 as

$$b_0 = \overline{y} - b_1 \overline{x}$$

$$b_1 = \frac{s_{xy}}{s_{xx}}$$

where

$$s_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}), \quad s_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2, \quad \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \; \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

The fitted model is $y = b_0 + b_1 x$.

Case 1: Prediction of average value of *y*

Suppose we want to predict the value of $E(y)$ for a given value of $x = x_0$. Then the predictor is given by

$$p_m = b_0 + b_1 x_0.$$

Here *m* stands for mean value.

Predictive bias

The prediction error is given as

$$\begin{aligned} p_m - E(y) &= b_0 + b_1 x_0 - E(\beta_0 + \beta_1 x_0 + \varepsilon) \\ &= b_0 + b_1 x_0 - (\beta_0 + \beta_1 x_0) \\ &= (b_0 - \beta_0) + (b_1 - \beta_1) x_0. \end{aligned}$$

Then the prediction bias is given as

$$\begin{aligned} E\left[p_m - E(y)\right] &= E(b_0 - \beta_0) + E(b_1 - \beta_1) x_0 \\ &= 0 + 0 = 0. \end{aligned}$$

Thus the predictor p_m is an unbiased predictor of $E(y)$.

Predictive variance:

The predictive variance of p_m is

$$\begin{aligned} PV(p_m) &= Var(b_0 + b_1 x_0) \\ &= Var\left[\bar{y} + b_1 (x_0 - \bar{x})\right] \\ &= Var(\bar{y}) + (x_0 - \bar{x})^2 Var(b_1) + 2(x_0 - \bar{x}) Cov(\bar{y}, b_1) \\ &= \frac{\sigma^2}{n} + \frac{\sigma^2 (x_0 - \bar{x})^2}{s_{xx}} + 0 \\ &= \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_{xx}}\right]. \end{aligned}$$

Estimate of predictive variance

The predictive variance can be estimated by substituting σ^2 by $\hat{\sigma}^2 = MSE$ as

$$\begin{aligned} \widehat{PV}(p_m) &= \hat{\sigma}^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_{xx}}\right] \\ &= MSE \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_{xx}}\right]. \end{aligned}$$

Prediction interval :

The 100(1-α)% prediction interval for $E(y)$ is obtained as follows:

The predictor p_m is a linear combination of normally distributed random variables, so it is also normally distributed as

$$p_m \sim N\left(\beta_0 + \beta_1 x_0, PV\left(p_m\right)\right).$$

So if σ^2 is known, then the distribution of

$$\frac{p_m - E(y)}{\sqrt{PV(p_m)}}$$

is $N(0,1)$. So the 100(1-α)% prediction interval is obtained as

$$P\left[-z_{\alpha/2} \le \frac{p_m - E(y)}{\sqrt{PV(p_m)}} \le z_{\alpha/2}\right] = 1-\alpha$$

which gives the prediction interval for $E(y)$ as

$$\left[p_m - z_{\alpha/2}\sqrt{\sigma^2\left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{s_{xx}}\right]},\ p_m + z_{\alpha/2}\sqrt{\sigma^2\left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{s_{xx}}\right]}\right].$$

When σ^2 is unknown, it is replaced by $\hat{\sigma}^2 = MSE$ and in this case, the sampling distribution of

$$\frac{p_m - E(y)}{\sqrt{MSE\left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{s_{xx}}\right]}}$$

is t-distribution with $(n-2)$ degrees of freedom, i.e., t_{n-2}.

The 100(1-α)% prediction interval in this case is

$$P\left[-t_{\alpha/2,n-2} \le \frac{p_m - E(y)}{\sqrt{MSE\left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{s_{xx}}\right]}} \le t_{\alpha/2,n-2}\right] = 1-\alpha.$$

which gives the prediction interval as

$$\left[p_m - t_{\alpha/2,n-2}\sqrt{MSE\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{s_{xx}}\right)},\ \ p_m + t_{\alpha/2,n-2}\sqrt{MSE\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{s_{xx}}\right)}\right].$$

Note that the width of the prediction interval $E(y)$ is a function of x_0. The interval width is minimum for $x_0 = \bar{x}$ and widens as $|x_0 - \bar{x}|$ increases. This is also expected as the best estimates of y to be made at x-values lie near the center of the data and the precision of estimation to deteriorate as we move to the boundary of the x-space.

Case 2: Prediction of actual value

If x_0 is the value of the explanatory variable, then the actual value predictor for y is

$$p_a = b_0 + b_1 x_0 .$$

Here *a* means "actual". The true value of *y* in the prediction period is given by $y_0 = \beta_0 + \beta_1 x_0 + \varepsilon_0$ where ε_0 indicates the value that would be drawn from the distribution of random error in the prediction period. Note that the form of predictor is the same as of average value predictor, but its predictive error and other properties are different. This is the **dual nature of predictor**.

Predictive bias:

The predictive error of p_a is given by

$$\begin{aligned} p_a - y_0 &= b_0 + b_1 x_0 - (\beta_0 + \beta_1 x_0 + \varepsilon_0) \\ &= (b_0 - \beta_0) + (b_1 - \beta_1) x_0 - \varepsilon_0 . \end{aligned}$$

Thus, we find that

$$\begin{aligned} E(p_a - y_0) &= E(b_0 - \beta_0) + E(b_1 - \beta_1) x_0 - E(\varepsilon_0) \\ &= 0 + 0 + 0 = 0 \end{aligned}$$

which implies that p_a is an unbiased predictor of y.

Predictive variance

Because the future observation y_0 is independent of p_a, the predictive variance of p_a is

$$\begin{aligned} PV(p_a) &= E(p_a - y_0)^2 \\ &= E[(b_0 - \beta_0) + (x_0 - \bar{x})(b_1 - \beta_1) + (b_1 - \beta_1)\bar{x} - \varepsilon_0]^2 \\ &= Var(b_0) + (x_0 - \bar{x})^2 Var(b_1) + \bar{x}^2 Var(b_1) + Var(\varepsilon_0) + 2(x_0 - \bar{x}) Cov(b_0, b_1) + 2\bar{x} Cov(b_0, b_1) + 2(x_0 - \bar{x}) Var(b_1) \\ &\qquad \text{[rest of the terms are 0 assuming the independence of } \varepsilon_0 \text{ with } \varepsilon_1, \varepsilon_2, ..., \varepsilon_n] \\ &= Var(b_0) + [(x_0 - \bar{x})^2 + \bar{x}^2 + 2(x_0 - \bar{x})] Var(b_1) + Var(\varepsilon_0) + 2[(x_0 - \bar{x}) + 2\bar{x}] Cov(b_0, b_1) \\ &= Var(b_0) + x_0^2 Var(b_1) + Var(\varepsilon_0) + 2x_0 Cov(b_0, b_1) \\ &= \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{s_{xx}} \right] + x_0^2 \frac{\sigma^2}{s_{xx}} + \sigma^2 - 2x_0 \frac{\bar{x}\sigma^2}{s_{xx}} \\ &= \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_{xx}} \right]. \end{aligned}$$

Estimate of predictive variance

The estimate of predictive variance can be obtained by replacing σ^2 by its estimate $\hat{\sigma}^2 = MSE$ as

$$\widehat{PV}(p_a) = \hat{\sigma}^2\left[1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{s_{xx}}\right]$$
$$= MSE\left[1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{s_{xx}}\right].$$

Prediction interval:

If σ^2 is known, then the distribution of

$$\frac{p_a - y_0}{\sqrt{PV(p_a)}}$$

is $N(0,1)$. So the 100(1-α)% prediction interval for y_0 is obtained as

$$P\left[-z_{\alpha/2} \le \frac{p_a - y_0}{\sqrt{PV(p_a)}} \le z_{\alpha/2}\right] = 1-\alpha$$

which gives the prediction interval for y_0 as

$$\left[p_a - z_{\alpha/2}\sqrt{\sigma^2\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{s_{xx}}\right)},\ p_a + z_{\alpha/2}\sqrt{\sigma^2\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{s_{xx}}\right)}\right].$$

When σ^2 is unknown, then

$$\frac{p_a - y_0}{\sqrt{\widehat{PV}(p_a)}}$$

follows a t-distribution with $(n-2)$ degrees of freedom. The 100(1-α)% prediction interval for y_0 in this case is obtained as

$$P\left[-t_{\alpha/2,n-2} \le \frac{p_a - y_0}{\sqrt{\widehat{PV}(p_a)}} \le t_{\alpha/2,n-2}\right] = 1-\alpha$$

which gives the prediction interval for y_0 as

$$\left[p_a - t_{\alpha/2,n-2}\sqrt{MSE\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{s_{xx}}\right)},\ p_a + t_{\alpha/2,n-2}\sqrt{MSE\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{s_{xx}}\right)}\right].$$

The prediction interval is of minimum width at $x_0 = \bar{x}$ and widens as $|x_0 - \bar{x}|$ increases.

The prediction interval for p_a is wider than the prediction interval for p_m because the prediction interval for p_a depends on both the error from the fitted model as well as the error associated with the future observations.

Within sample prediction in multiple linear regression model

Consider the multiple regression model with k explanatory variables as

$$y = X\beta + \varepsilon,$$

where $y = (y_1, y_2, ..., y_n)'$ is a $n \times 1$ vector of n observation on study variable,

$$X = \begin{pmatrix} x_{11} & x_{12} \cdots x_{1k} \\ x_{21} & x_{22} \cdots x_{2k} \\ \vdots & \vdots \ddots \vdots \\ x_{n1} & x_{n2} \cdots x_{nk} \end{pmatrix}$$

is a $n \times k$ matrix of n observations on each of the k explanatory variables, $\beta = (\beta_1, \beta_2, ..., \beta_k)'$ is a $k \times 1$ vector of regression coefficients and $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)'$ is a $n \times 1$ vector of random error components or disturbance term following $N(0, \sigma^2 I_n)$. If the intercept term is present, take the first column of X to be $(1,1,...,1)'$.

Let the parameter β be estimated by its ordinary least squares estimator $b = (X'X)^{-1}X'y$. Then the predictor is $p = Xb$ which can be used for predicting the actual and average values of the study variable. This is the **dual nature of predictor**.

Case 1: Prediction of average value of y

When the objective is to predict the average value of y, i.e., $E(y)$, then the estimation error is given by

$$\begin{aligned} p - E(y) &= Xb - X\beta \\ &= X(b - \beta) \\ &= X(X'X)^{-1}X'\varepsilon \\ &= H\varepsilon \end{aligned}$$

where $H = X(X'X)^{-1}X'$.

Then

$$E\left[p - E(y)\right] = X\beta - X\beta = 0$$

which proves that the predictor $p = Xb$ provides an unbiased prediction for the average value.

The predictive variance of p is

$$\begin{aligned} PV_m(p) &= E\left[\{p - E(y)\}'\{p - E(y)\}\right] \\ &= E\left[\varepsilon' H.H\varepsilon\right] \\ &= E\left(\varepsilon' H\varepsilon\right) \\ &= \sigma^2 tr\, H = \sigma^2 k. \end{aligned}$$

The predictive variance can be estimated by $\widehat{PV}_m(p) = \hat{\sigma}^2 k$ where $\hat{\sigma}^2 = MSE$ is obtained from the analysis of variance based on OLSE.

When σ^2 is known, then the distribution of

$$\frac{p-E(y)}{\sqrt{PV_m(p)}}$$

is $N(0,1)$. So the 100(1-α)% prediction interval for $E(y)$ is obtained as

$$P\left[-z_{\alpha/2} \le \frac{p-E(y)}{\sqrt{PV_m(p)}} \le z_{\alpha/2}\right] = 1-\alpha$$

which gives the prediction interval for $E(y)$ as

$$\left[p - z_{\alpha/2}\sqrt{PV_m(p)},\ p + z_{\alpha/2}\sqrt{PV_m(p)}\right].$$

When σ^2 is unknown, it is replaced by $\hat{\sigma}^2 = MSE$ and in this case, the sampling distribution of

$$\frac{p-E(y)}{\sqrt{\widehat{PV}_m(p)}}$$

is t-distribution with $(n-k)$ degrees of freedom, i.e., t_{n-k}.

The 100(1-α)% prediction interval for $E(y)$ in this case is

$$P\left[-t_{\alpha/2,n-k} \le \frac{p-E(y)}{\sqrt{\widehat{PV}_m(p)}} \le t_{\alpha/2,n-k}\right] = 1-\alpha.$$

which gives the prediction interval for $E(y)$ as

$$\left[p - t_{\alpha/2,n-k}\sqrt{\widehat{PV}_m(p)},\quad p + t_{\alpha/2,n-k}\sqrt{\widehat{PV}_m(p)}\right].$$

Case 2: Prediction of actual value of y

When the predictor $p = Xb$ is used for predicting the actual value of the study variable y, then its prediction error is given by

$$\begin{aligned} p - y &= Xb - X\beta - \varepsilon \\ &= X(b-\beta) - \varepsilon \\ &= X(X'X)^{-1}X'\varepsilon - \varepsilon \\ &= -\left[I - X(X'X)^{-1}X'\right]\varepsilon \\ &= -\bar{H}\varepsilon. \end{aligned}$$

Thus

$$E(p-y) = 0$$

which shows that p provides unbiased predictions for the actual values of the study variable.

The predictive variance in this case is

$$\begin{aligned} PV_a(p) &= E\left[(p-y)'(p-y)\right] \\ &= E\left(\varepsilon'\bar{H}.\bar{H}\varepsilon\right) \\ &= E\left(\varepsilon'\bar{H}\varepsilon\right) \\ &= \sigma^2 tr\bar{H} \\ &= \sigma^2(n-k). \end{aligned}$$

The predictive variance can be estimated by

$$\widehat{PV}_m(p) = \hat{\sigma}^2(n-k)$$

where $\hat{\sigma}^2 = MSE$ is obtained from the analysis of variance based on OLSE.

Comparing the performances of p to predict actual and average values, we find that p in better predictor for predicting the average value in comparison to actual value when

$$\begin{aligned} & PV_m(p) < PV_a(p) \\ \text{or} \quad & k < (n-k) \\ \text{or} \quad & 2k < n. \end{aligned}$$

i.e. when the total number of observations is more than twice the number of explanatory variables.

Now we obtain the confidence interval for *y*.

When σ^2 is known, then the distribution of

$$\frac{p-y}{\sqrt{PV_a(p)}}$$

is $N(0,1)$. So the $100(1-\alpha)\%$ prediction interval for *y* is obtained as

$$P\left[-z_{\alpha/2} \le \frac{p-y}{\sqrt{PV_a(p)}} \le z_{\alpha/2}\right] = 1-\alpha$$

which gives the prediction interval for y as

$$\left[p - z_{\alpha/2}\sqrt{PV_a(p)},\ p + z_{\alpha/2}\sqrt{PV_a(p)}\right].$$

When σ^2 is unknown, it is replaced by $\hat{\sigma}^2 = MSE$ and in this case, the sampling distribution of

$$\frac{p-y}{\sqrt{\widehat{PV}_a(p)}}$$

is t-distribution with $(n-k)$ degrees of freedom, i.e., t_{n-k}.

The 100(1-α)% prediction interval of *y, in this case,* is obtained as

$$P\left[-t_{\alpha/2,n-k} \le \frac{p-y}{\sqrt{\widehat{PV}_a(p)}} \le t_{\alpha/2,n-k}\right] = 1-\alpha.$$

which gives the prediction interval for *y* as

$$\left[p - t_{\alpha/2,n-k}\sqrt{\widehat{PV}_a(p)},\quad p + t_{\alpha/2,n-k}\sqrt{\widehat{PV}_a(p)}\right].$$

Outside sample prediction in multiple linear regression model

Consider the model

$$y = X\beta + \varepsilon \qquad (1)$$

where y is a $n\times 1$ vector of n observations on study variable, X is a $n\times k$ matrix of explanatory variables and ε is a $n\times 1$ vector of disturbances following $N(0,\sigma^2 I_n)$.

Further, suppose a set of n_f observations on the same set of k explanatory variables are also available, but the corresponding n_f observations on the study variable are not available. Assuming that this set of observation also follows the same model, we can write

$$y_f = X_f\beta + \varepsilon_f \qquad (2)$$

where y_f is a $n_f \times 1$ vector of future values, X_f is a $n_f \times k$ matrix of known values of explanatory variables and ε_f is a $n_f \times 1$ vector of disturbances following $N(0,\sigma^2 I_{n_f})$. It is also assumed that the elements of ε and ε_f are independently distributed.

We now consider the prediction of y_f values for given X_f from model (2). This can be done by estimating the regression coefficients from the model (1) based on n observations and use it is formulating the predictor in the model (2). If ordinary least squares estimation is used to estimate β in the model (1) as

$$b = (X'X)^{-1}X'y$$

then the corresponding predictor is

$$p_f = X_f b = X_f(X'X)^{-1}X'y.$$

Case 1: Prediction of average value of study variable

When the aim is to predict the average value $E(y_f)$, then the prediction error is

$$\begin{aligned} p_f - E(y_f) &= X_f b - X_f \beta \\ &= X_f (b - \beta) \\ &= X_f (X'X)^{-1} X' \varepsilon. \end{aligned}$$

Then

$$\begin{aligned} E\left[p_f - E(y_f) \right] &= X_f \left(X'X \right)^{-1} X' E\left(\varepsilon \right) \\ &= 0. \end{aligned}$$

Thus p_f provides an unbiased prediction for the average value.

The predictive covariance matrix of p_f is

$$\begin{aligned} Cov_m(p_f) &= E\left[\left\{ p_f - E(y_f) \right\} \left\{ p_f - E(y_f) \right\}' \right] \\ &= E\left[X_f (X'X)^{-1} X' \varepsilon\varepsilon' X (X'X)^{-1} X_f' \right] \\ &= X_f \left(X'X \right)^{-1} X' E\left(\varepsilon\varepsilon' \right) X (X'X)^{-1} X_f' \\ &= \sigma^2 X_f (X'X)^{-1} X' X \left(X'X \right)^{-1} X_f' \\ &= \sigma^2 X_f (X'X)^{-1} X_f'. \end{aligned}$$

The predictive variance of p_f is

$$\begin{aligned} PV_m(p_f) &= E\left[\left\{ p_f - E(y_f \right\}' \left\{ p_f - E(y_f) \right\} \right] \\ &= tr\left[Cov_m(p_f) \right] \\ &= \sigma^2 tr\left[(X'X)^{-1} X_f' X_f \right]. \end{aligned}$$

If σ^2 is unknown, then replace σ^2 by $\hat{\sigma}^2 = MSE$ in the expressions of the predictive covariance matrix and predictive variance and their estimates are

$$\widehat{Cov}_m(p_f) = \hat{\sigma}^2 X_f (X'X)^{-1} X_f'$$

$$\widehat{PV}_m(p_f) = \hat{\sigma}^2 tr\left[\left(X'X \right)^{-1} (X_f' X_f) \right].$$

Now we obtain the confidence interval for $E(y_f)$.

When σ^2 is known, then the distribution of

$$\frac{p_f - E(y_f)}{\sqrt{PV_m(p_f)}}$$

is $N(0,1)$. So the $100(1-\alpha)\%$ prediction interval of $E(y_f)$ is obtained as

$$P\left[-z_{\alpha/2} \le \frac{p_f - E(y_f)}{\sqrt{PV_m(p_f)}} \le z_{\alpha/2}\right] = 1-\alpha$$

which gives the prediction interval for $E(y_f)$ as

$$\left[p_f - z_{\alpha/2}\sqrt{PV_m(p_f)},\ p_f + z_{\alpha/2}\sqrt{PV_m(p_f)}\right].$$

When σ^2 is unknown, it is replaced by $\hat{\sigma}^2 = MSE$ and in this case, the sampling distribution of

$$\frac{p_f - E(y_f)}{\sqrt{\widehat{PV}_m(p_f)}}$$

is t-distribution with $(n-k)$ degrees of freedom, i.e., t_{n-k}.

The 100(1-α)% prediction interval for $E(y_f)$ in this case is

$$P\left[-t_{\alpha/2,n-k} \le \frac{p_f - E(y_f)}{\sqrt{\widehat{PV}_m(p_f)}} \le t_{\alpha/2,n-k}\right] = 1-\alpha.$$

which gives the prediction interval for $E(y_f)$ as

$$\left[p_m - t_{\alpha/2,n-k}\sqrt{\widehat{PV}_m(p)},\ \ p_m + t_{\alpha/2,n-k}\sqrt{\widehat{PV}_m(p)}\right].$$

Case 2: Prediction of actual value of study variable

When p_f is used to predict the actual value y_f, then the prediction error is

$$\begin{aligned} p_f - y_f &= X_f b - X_f \beta - \varepsilon_f \\ &= X_f\ (b-\beta) - \varepsilon_f. \end{aligned}$$

Then

$$E\left(p_f - y_f\right) = X_f E\left(b-\beta\right) - E(\varepsilon_f) = 0.$$

Thus p_f provides an unbiased prediction for actual values.

The predictive covariance matrix of p_f in this case is

$$\begin{aligned} Cov_a\left(p_f\right) &= E\left[\left(p_f - y_f\right)\left(p_f - y_f\right)'\right] \\ &= E\left[\left\{X_f(b-\beta) - \varepsilon_f\right\}\left\{(b-\beta)'X_f' - \varepsilon_f'\right\}\right] \\ &= X_f V(b) X_f' + E(\varepsilon_f \varepsilon_f') \ \ (\text{Using } (b-\beta) = (X'X)^{-1}X'\varepsilon) \\ &= \sigma^2\left[X_f(X'X)^{-1}X_f' + I_{n_f}\right]. \end{aligned}$$

The predictive variance of p_f is

$$\begin{aligned} PV_a(p_f) &= E\left[\left(p_f - y_f\right)'\left(p_f - y_f\right)\right] \\ &= tr\left[Cov_a(p_f)\right] \\ &= \sigma^2\left[tr(X'X)^{-1}X_f'X_f + n_f\right]. \end{aligned}$$

The estimates of the covariance matrix and predictive variance can be obtained by replacing σ^2 by $\hat{\sigma}^2 = MSE$ as

$$\widehat{Cov}_a(p_f) = \hat{\sigma}^2\left[X_f(X'X)^{-1}X_f' + I_{n_f}\right]$$
$$\widehat{PV}_a(p_f) = \hat{\sigma}^2\left[tr(X'X)^{-1}X_f'X_f + n_f\right].$$

Now we obtain the confidence interval for y_f.

When σ^2 is known, then the distribution of

$$\frac{p_f - y_f}{\sqrt{PV_a(p_f)}}$$

is $N(0,1)$. So the $100(1-\alpha)\%$ prediction interval is obtained as

$$P\left[-z_{\alpha/2} \le \frac{p_f - y_f}{\sqrt{PV_a(p_f)}} \le z_{\alpha/2}\right] = 1-\alpha$$

which gives the prediction interval for y_f as

$$\left[p_f - z_{\alpha/2}\sqrt{PV_a(p_f)},\ p_f + z_{\alpha/2}\sqrt{PV_a(p_f)}\right].$$

When σ^2 is unknown, it is replaced by $\hat{\sigma}^2 = MSE$ and in this case, the sampling distribution of

$$\frac{p_f - y_f}{\sqrt{\widehat{PV}_a(p_f)}}$$

is t-distribution with $(n-k)$ degrees of freedom, i.e., t_{n-k}.

The $100(1-\alpha)\%$ prediction interval for y_f in this case is

$$P\left[-t_{\alpha/2,n-k} \le \frac{p_f - y_f}{\sqrt{\widehat{PV}_a(p_f)}} \le t_{\alpha/2,n-k}\right] = 1-\alpha.$$

which gives the prediction interval for y_f as

$$\left[p_f - t_{\alpha/2,n-k}\sqrt{\widehat{PV}_a(p_f)},\ \ p_f + t_{\alpha/2,n-k}\sqrt{\widehat{PV}_a(p_f)}\right].$$

Simultaneous prediction of average and actual values of the study variable

The predictions are generally obtained either for the average values of the study variable or actual values of the study variable. In many applications, it may not be appropriate to confine our attention to only to either of the two. It may be more appropriate in some situations to predict both the values simultaneously, i.e., consider the prediction of actual and average values of the study variable simultaneously. For example, suppose a firm deals with the sale of fertilizer to the user. The interest of the company would be in predicting the average value of yield which the company would like to use in showing that the average yield of the crop increases by using their fertilizer. On the other side, the user would not be interested in the average value. The user would like to know the actual increase in the yield by using the fertilizer. Suppose both seller and user, both go for prediction through regression modeling. Now using the classical tools, the statistician can predict either the actual value or the average value. This can safeguard the interest of either the user or the seller. Instead of this, it is required to safeguard the interest of both by striking a balance between the objectives of the seller and the user. This can be achieved by combining both the predictions of actual and average values. This can be done by formulating an objective function or target function. Such target function has to be flexible and should allow assigning different weights to the choice of two kinds of predictions depending upon their importance in any given application and also reducible to individual predictions leading to actual and average value prediction.

Now we consider the simultaneous prediction in within and outside sample cases.

Simultaneous prediction in within sample prediction

Define a **target function**

$$\tau = \lambda y + (1-\lambda)E(y);\ 0 \le \lambda \le 1$$

which is a convex combination of actual value y and average value $E(y)$. The weight λ is a constant lying between zero and one whose value reflects the importance being assigned to actual value prediction. Moreover $\lambda = 0$ gives the average value prediction and $\lambda = 1$ gives the actual value prediction. For example, the value of λ in the fertilizer example depends on the rules and regulation of the market, norms of society and other considerations etc. The value of λ is the choice of practitioner.

Consider the multiple regression model

$$y = X\beta + \varepsilon,\ E(\varepsilon) = 0,\ E(\varepsilon\varepsilon') = \sigma^2 I_n.$$

Estimate β by ordinary least squares estimation and construct the predictor

$$p = Xb.$$

Now employ this predictor for predicting the actual and average values simultaneously through the target function.

The prediction error is

$$\begin{aligned} p-\tau &= Xb-\lambda y-(1-\lambda)E(y) \\ &= Xb-\lambda(X\beta+\varepsilon)-(1-\lambda)X\beta \\ &= X(b-\beta)-\lambda\varepsilon. \end{aligned}$$

Thus

$$\begin{aligned} E(p-\tau) &= XE(b-\beta)-\lambda E(\varepsilon) \\ &= 0. \end{aligned}$$

So p provides an unbiased prediction for τ.

The variance is

$$\begin{aligned} Var(p) &= E(p-\tau)'(p-\tau) \\ &= E\left[\{(b-\beta)'X'-\lambda\varepsilon'\}\{X(b-\beta)-\lambda\varepsilon\}\right] \\ &= E\left[\varepsilon'X(X'X)^{-1}X'X\left(X'X\right)^{-1}X'\varepsilon+\lambda^2\varepsilon'\varepsilon-\lambda(b-\beta)'X'\varepsilon-\lambda\varepsilon'X(b-\beta')\right] \\ &= E\left[(1-2\lambda)\varepsilon'X(X'X)^{-1}X'\varepsilon+\lambda^2\varepsilon'\varepsilon\right] \\ &= \sigma^2(1-2\lambda)tr\left[(X'X)^{-1}X'X\right]+\lambda^2\sigma^2 trI_n \\ &= \sigma^2\left[(1-2\lambda)k+\lambda^2 n\right]. \end{aligned}$$

The estimates of predictive variance can be obtained by replacing σ^2 by $\hat{\sigma}^2 = MSE$ as

$$\widehat{Var(p)} = \hat{\sigma}^2\left[(1-2\lambda)k+\lambda^2 n\right].$$

Simultaneous prediction is outside sample prediction:

Consider the model described earlier under outside sample prediction as

$$\underset{n\times 1}{y} = \underset{n\times k}{X}\underset{k\times 1}{\beta}+\underset{n\times 1}{\varepsilon},\ E(\varepsilon)=0, V(\varepsilon)=\sigma^2 I_n$$

$$\underset{n_f\times 1}{y_f} = \underset{n_f\times k}{X_f}\underset{k\times 1}{\beta}+\underset{n_f\times 1}{\varepsilon_f};\ E(\varepsilon_f)=0, V(\varepsilon_f)=\sigma^2 I_{n_f}.$$

The target function, in this case, is defined as

$$\tau_f = \lambda y_f + (1-\lambda)E(y_f);\ 0\le\lambda\le 1.$$

The predictor based on OLSE of β is

$$p_f = X_f b;\quad b=\left(X'X\right)^{-1}X'y.$$

The predictive error of p is

$$\begin{aligned} p_f-\tau_f &= X_f b-\lambda y_f-(1-\lambda)E(y_f) \\ &= X_f b-\lambda(X_f\beta+\varepsilon_f)-(1-\lambda)X_f\beta \\ &= X_f(b-\beta)-\lambda\varepsilon_f. \end{aligned}$$

So

$$E(p_f - \tau_f) = X_f E(b-\beta) - \lambda E(\varepsilon_f)$$
$$= 0.$$

Thus p_f provides an unbiased prediction for τ_f.

The variance of p_f is

$$Var(p_f) = E(p_f - \tau)'(p_f - \tau)$$
$$= E\left[\left\{(b-\beta)'\varepsilon_f' - \lambda\varepsilon_f'\right\}\left\{X_f(b-\beta) - \lambda\varepsilon_f\right\}\right]$$
$$= E\left[\varepsilon' X(X'X)^{-1}X_f'X_f\left(X'X\right)^{-1}X'\varepsilon + \lambda\varepsilon_f'\varepsilon_f - 2\lambda\varepsilon_f'X_f(X'X)^{-1}X'\varepsilon\right]$$
$$= \sigma^2\left[tr\left\{X(X'X)^{-1}X_f'X_f(X'X)^{-1}X'\right\} + \lambda^2 n_f\right]$$

assuming that the elements in ε and ε_f are mutually independent.

The estimates of predictive variance can be obtained by replacing σ^2 by $\hat{\sigma}^2 = MSE$ as

$$\widehat{Var(p_f)} = \hat{\sigma}^2\left[tr\left\{X(X'X)^{-1}X_f'X_f(X'X)^{-1}X'\right\} + \lambda^2 n_f\right].$$