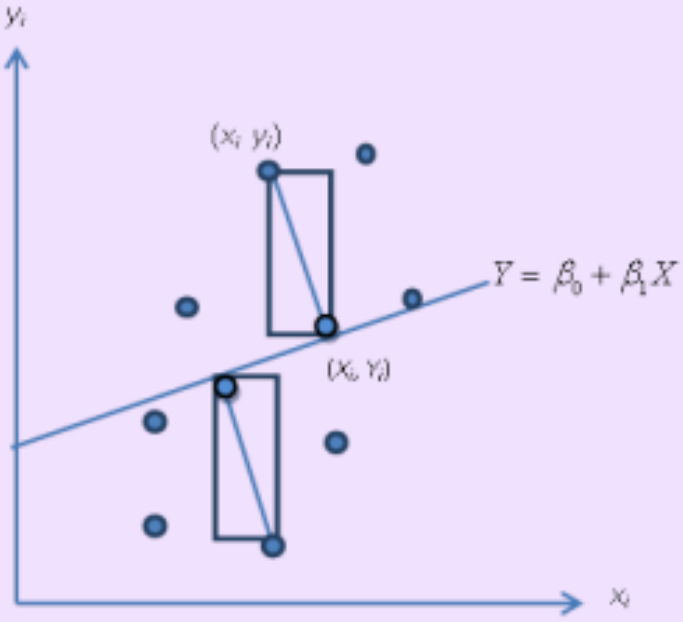
Flexible Functional Forms

Professor V.M. Auken, PhD

Reduced major axis regression method

The direct, reverse and orthogonal methods of estimation minimize the errors in a particular direction which is usually the distance between the observed data points and the line in the scatter diagram. Alternatively, one can consider the area extended by the data points in certain neighbourhood and instead of distances, the area of rectangles defined between corresponding observed data point and nearest point on the line in the following scatter diagram can also be minimized. Such an approach is more appropriate when the uncertainties are present in study as well as explanatory variables. This approach is termed as reduced major axis regression.



Reduced major axis method

Suppose the regression line is $Y_i = \beta_0 + \beta_1 X_i$ on which all the observed points are expected to lie. Suppose the points (x_i, y_i), $i = 1, 2, ..., n$ are observed which lie away from the line.

The area of rectangle extended between the i^{th} observed data point and the line is

$$A_i = (X_i \sim x_i)(Y_i \sim y_i) \quad (i = 1, 2, ..., n)$$

where (X_i, Y_i) denotes the i^{th} pair of observation without any error which lie on the line.

The total area extended by *n* data points is $\sum_{i=1}^{n} A_i = \sum_{i=1}^{n} (X_i \sim x_i)(Y_i \sim y_i).$

All observed data points $(x_i, y_i), (i = 1, 2, ..., n)$ are expected to lie on the line

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

and let

$$E_i^* = Y_i - \beta_0 - \beta_1 X_i = 0.$$

So now the objective is to minimize the sum of areas under the constraints E_i^* to obtain the reduced major axis estimates of regression coefficients. Using the Lagrangian multiplies method, the Lagrangian function is

$$L_R = \sum_{i=1}^{n} A_i - \sum_{i=1}^{n} \mu_i E_i^*$$

$$= \sum_{i=1}^{n} (X_i - x_i)(Y_i - y_i) - \sum_{i=1}^{n} \mu_i E_i^*$$

where $\mu_1, ..., \mu_n$ are the Lagrangian multipliers. The set of equations are obtained by setting

$$\frac{\partial L_R}{\partial X_i} = 0, \frac{\partial L_R}{\partial Y_i} = 0, \frac{\partial L_R}{\partial \beta_0} = 0, \frac{\partial L_R}{\partial \beta_1} = 0 \quad (i = 1, 2, ..., n).$$

Thus

$$\frac{\partial L_R}{\partial X_i} = (Y_i - y_i) + \beta_1 \mu_i = 0$$

$$\frac{\partial L_R}{\partial Y_i} = (X_i - x_i) - \mu_i = 0$$

$$\frac{\partial L_R}{\partial \beta_0} = \sum_{i=1}^{n} \mu_i = 0$$

$$\frac{\partial L_R}{\partial \beta_1} = \sum_{i=1}^{n} \mu_i X_i = 0.$$

Now

$$X_i = x_i + \mu_i$$

$$Y_i = y_i - \beta_1 \mu_i$$

$$\beta_0 + \beta_1 X_i = y_i - \beta_1 \mu_i$$

$$\beta_0 + \beta_1 (x_i + \mu_i) = y_i - \beta_1 \mu_i$$

$$\Rightarrow \mu_i = \frac{y_i - \beta_0 - \beta_1 x_i}{2\beta_1}.$$

Substituting μ_i in $\sum_{i=1}^{n} \mu_i = 0$, we get the reduced major axis regression estimate of β_0 is obtained as

$$\hat{\beta}_{0RM} = \overline{y} - \hat{\beta}_{1RM} \overline{x}$$

where $\hat{\beta}_{1RM}$ is the reduced major axis regression estimate of β_1. Using $X_i = x_i + \mu_i$, μ_i and $\hat{\beta}_{0RM}$ in $\sum_{i=1}^{n} \mu_i X_i = 0$,

we get

$$\sum_{i=1}^{n} \left(\frac{y_i - \overline{y} + \beta_1 \overline{x} - \beta_1 x_i}{2\beta_1} \right) \left(x_i - \frac{y_i - \overline{y} + \beta_1 \overline{x} - \beta_1 x_i}{2\beta_1} \right) = 0.$$

Reduced major axis regression method

Let $u_i = x_i - \overline{x}$ and $v_i = y_i - \overline{y}$, then this equation can be re-expressed as $\sum_{i=1}^{n}(v_i - \beta_1 u_i)(v_i + \beta_1 u_i + 2\beta_1 \overline{x}) = 0.$

Using $\sum_{i=1}^{n} u_i = \sum_{i=1}^{n} v_i = 0,$ we get

$$\sum_{i=1}^{n} v_i^2 - \beta_1^2 \sum_{i=1}^{n} u_i^2 = 0.$$

Solving this equation, the reduced major axis regression estimate of β_1 is obtained as

$$\hat{\beta}_{1RM} = sign(s_{xy})\sqrt{\frac{s_{yy}}{s_{xx}}}$$

where

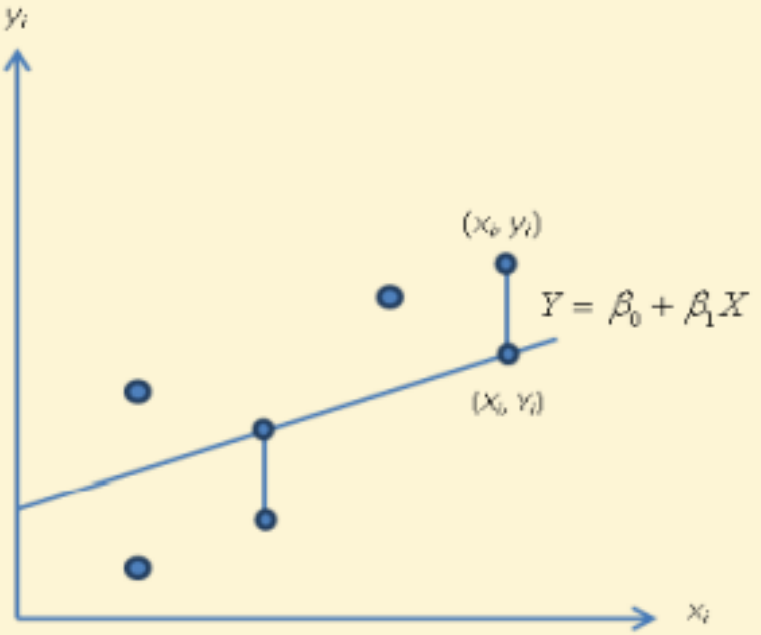
$$sign(s_{xy}) = \begin{cases} 1 & \text{if } s_{xy} > 0 \\ -1 & \text{if } s_{xy} < 0. \end{cases}$$

We choose the regression estimator which has same sign as that of s_{xy}.

Least absolute deviation regression method

The least squares principle advocates the minimization of sum of squared errors. The idea of squaring the errors is useful in place of simple errors because the random errors can be positive as well as negative. So consequently their sum can be close to zero indicating that there is no error in the model which can be misleading. Instead of the sum of random errors, the sum of absolute random errors can be considered which avoids the problem due to positive and negative random errors.

In the method of least squares, the estimates of the parameters β_0 and β_1 in the model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i.\ (i = 1, 2, ..., n)$ are chosen such that the sum of squares of deviations $\sum_{i=1}^{n} \varepsilon_i^2$ is minimum. In the method of least absolute deviation (LAD) regression, the parameters β_0 and β_1 are estimated such that the sum of absolute deviations $\sum_{i=1}^{n} |\varepsilon_i|$ is minimum. It minimizes the absolute vertical sum of errors as in the following scatter diagram:



Least absolute deviation regression

The LAD estimates $\hat{\beta}_{0L}$ and $\hat{\beta}_{1L}$ are the values β_0 and β_1, respectively which minimize $LAD(\beta_0,\beta_1)=\sum_{i=1}^{n}|y_i-\beta_0-\beta_1 x_i|$ for the given observations $(x_i,y_i)(i=1,2,...,n)$.

Conceptually, LAD procedure is simpler than OLS procedure because |e| (absolute residuals) is a more straightforward measure of the size of the residual than e^2 (squared residuals). The LAD regression estimates of β_0 and β_1 are not available in closed form. Rather they can be obtained numerically based on algorithms. Moreover, this creates the problems of non-uniqueness and degeneracy in the estimates. The concept of non-uniqueness relates to more than one best lines passing through a data point. The degeneracy concept describes that the best line through a data point also passes through more than one other data points. The non-uniqueness and degeneracy concepts are used in algorithms to judge the quality of the estimates. The algorithm for finding the estimators generally proceeds in steps. At each step, the best line is found that passes through a given data point. The best line always passes through another data point, and this data point is used in the next step. When there is non-uniqueness, then there are more than one best lines. When there is degeneracy, then the best line passes through more than one other data point. When either of the problem is present, then there is more than one choice for the data point to be used in the next step and the algorithm may go around in circles or make a wrong choice of the LAD regression line. The exact tests of hypothesis and confidence intervals for the LAD regression estimates can not be derived analytically. Instead they are derived analogous to the tests of hypothesis and confidence intervals related to ordinary least squares estimates.

Estimation of parameters when X is stochastic

In a usual linear regression model, the study variable is supped to be random and explanatory variables are assumed to be fixed. In practice, there may be situations in which the explanatory variable also becomes random.

Suppose both dependent and independent variables are stochastic in the simple linear regression model

$$y = \beta_0 + \beta_1 X + \varepsilon$$

where ε is the associated random error component. The observations (x_i, y_i), $i = 1, 2, ..., n$ are assumed to be jointly distributed. Then the statistical inferences can be drawn in such cases which are conditional on X.

Assume the joint distribution of X and y to be bivariate normal $N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ where μ_x and μ_y are the means of X and $y; \sigma_x^2$ and σ_y^2 are the variances of X and y, and ρ is the correlation coefficient between X and y. Then the conditional distribution of y given $X = x$ is univariate normal conditional mean

$$E(y \mid X = x) = \mu_{y|x} = \beta_0 + \beta_1 x$$

and conditional variance of y given $X = x$ is

$$Var(y|X = x) = \sigma_{y|x}^2 = \sigma_y^2(1 - \rho^2)$$

where

$$\beta_0 = \mu_y - \mu_x \beta_1$$

and

$$\beta_1 = \frac{\sigma_y}{\sigma_x} \rho.$$

When both X and y are stochastic, then the problem of estimation of parameters can be reformulated as follows. Consider a conditional random variable $y|X = x$ having a normal distribution with mean as conditional mean $\mu_{y|x}$ and variance as conditional variance $Var(y|X = x) = \sigma^2_{y|x}$. Obtain n independently distributed observation $y_i|x_i,\ i = 1, 2, ..., n$ from $N(\mu_{y|x}, \sigma^2_{y|x})$ with nonstochastic X. Now the method of maximum likelihood can be used to estimate the parameters which yields the estimates of β_0 and β_1 as earlier in the case of nonstochastic X as

$$\tilde{b} = \overline{y} - \tilde{b}_1 \overline{x}$$

and

$$\tilde{b}_1 = \frac{s_{xy}}{s_{xx}}$$

respectively.

Moreover, the correlation coefficient

$$\rho = \frac{E(y - \mu_y)(X - \mu_x)}{\sigma_y \sigma_x}$$

can be estimated by the sample correlation coefficient

$$\hat{\rho} = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

$$= \frac{s_{xy}}{\sqrt{s_{xx}}\sqrt{s_{yy}}}$$

$$= \tilde{b}_1 \sqrt{\frac{s_{xx}}{s_{yy}}}.$$

Thus

$$\hat{\rho}^2 = \tilde{b}_1^2 \frac{s_{xx}}{s_{yy}}$$

$$= \tilde{b}_1 \frac{s_{xy}}{s_{yy}}$$

$$= \frac{s_{yy} - \sum_{i=1}^{n} \hat{\varepsilon}_i^2}{s_{yy}}$$

$$= R^2$$

which is same as the coefficient of determination.

Thus R^2 has the same expression as in the case when X is fixed.

Thus R^2 again measures the goodness of fitted model even when X is stochastic.

QUESTIONS!