Time Series Analysis

Linear Regression Models

Binzhi Chen

Department of Economics
University of Birmingham

March 4, 2020

The Algebra of Linear Regression

Suppose that a scalar y_t, is related to a $(kx1)$ vector x_t and a disturbance term u_t according to the regression model:

$$y_t = \mathbf{x_t'}\beta + u_t \tag{1}$$

It is commonly written as the matrix form:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{u} \tag{2}$$

Where

$$\mathbf{Y}_{(Tx1)} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} \quad \mathbf{X}_{(TxK)} = \begin{bmatrix} \mathbf{x_1'} \\ \mathbf{x_2'} \\ \vdots \\ \mathbf{x_T'} \end{bmatrix} \quad \mathbf{u}_{(Tx1)} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix}$$

The Algebra of Linear Regression

The classical regression assumptions as follows:
1-Linearity:

$$y_t = \mathbf{x_t'}\beta + u_t$$

2-Conditional Mean:

$$E[u_i|\mathbf{X}] = 0$$

3-Rank:

$$P(rank(\mathbf{X}) = k) = 1$$

4-Conditional Homoskedasticity:

$$V[u_i|\mathbf{X}] = \sigma^2$$

5-Conditional Correlation:

$$E[u_i u_j|\mathbf{X}] = 0$$

6-Conditional Normality:

$$\mathbf{u}|\mathbf{X} \sim N(0, \sigma^2)$$

7-**X** is non-stochastic.

The Algebra of Linear Regression

The *ordinary least squares (OLS)* estimate of β (denoted $\hat{\beta}$) is the value of β that minimizes the residual sum of squares (RSS), denoted by Q:

$$
\begin{aligned}
Q &= \sum_{t=1}^{T}(y_t - \mathbf{x_t'}\beta)^2 \\
&= (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) \\
&= \mathbf{Y'Y} - \mathbf{2Y'X}\beta + \beta'\mathbf{X'X}\beta
\end{aligned}
$$

Taking first-order difference with respect to β:

$$\frac{\partial Q}{\partial \beta} = 0 = -\mathbf{2X'Y} + \mathbf{2X'X}\beta$$

The estimate of the value β is:

$$\hat{\beta} = (\mathbf{X'X})^{-\mathbf{1}}\mathbf{X'Y} \tag{3}$$

The Algebra of Linear Regression

The difference between the OLS estimate $\hat{\boldsymbol{\beta}}$ the true population parameter $\boldsymbol{\beta}$ is found by substituting equation(2) into equation(3):

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'[\mathbf{X}\boldsymbol{\beta} + \mathbf{u}] = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{u}$$

Remember that the assumptions of OLS are $E(\mathbf{u}) = 0$, $E(\mathbf{uu}') = \sigma^2\mathbf{I_T}$ and x is stochastic. Taking expectation of the above equation we get:

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'E(\mathbf{u}) = \boldsymbol{\beta}$$

$\hat{\boldsymbol{\beta}}$ is unbiased.

The variance-covariance matrix is given by:

$$\begin{aligned} E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'] &= E[(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}] \\ &= (\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'E[(\mathbf{u}\mathbf{u}')]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-\mathbf{1}} \end{aligned}$$

An inference based on $\hat{\boldsymbol{\beta}}$ about any linear combination of the elements of $\boldsymbol{\beta}$ will have a smaller variance than the corresponding inference based on any alternative linear unbiased estimator.

Thus, the preceding results imply:

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-\mathbf{1}})$$

The Algebra of Linear Regression

The OLS sample residual can be written as:

$$\hat{\mathbf{u}} = \mathbf{Y} - \mathbf{X}\hat{\beta} = \mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = [\mathbf{I_T} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y} = \mathbf{M_X}\mathbf{Y}$$

Where $\mathbf{M_X}$ is defined as the following (TxT) matrix:

$$\mathbf{M_X} = \mathbf{I_T} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

We can verify that $\mathbf{M_X}$ is symmetric, idempotent and orthogonal to the columns of $\mathbf{X}$:

$$\mathbf{M_X} = \mathbf{M_X}' \quad \mathbf{M_X}\mathbf{M_X} = \mathbf{M_X} \quad \mathbf{M_X}\mathbf{X} = 0$$

Hence population residuals is:

$$\hat{\mathbf{u}} = \mathbf{M_X}(\mathbf{X}\beta + \mathbf{u}) = \mathbf{M_X}\mathbf{u}$$

The Algebra of Linear Regression

The OLS estimate of the variance is:

$$\hat{\sigma}^2 = RSS/(T-K) = \hat{\mathbf{u}}'\hat{\mathbf{u}}/(T-K) = \mathbf{u}'\mathbf{M}_\mathbf{X}'\mathbf{M}_\mathbf{X}\mathbf{u}/(T-K)$$

Consider the numerator $\mathbf{u}'\mathbf{M}_\mathbf{X}'\mathbf{M}_\mathbf{X}\mathbf{u}$:

$$\mathbf{u}'\mathbf{M}_\mathbf{X}'\mathbf{M}_\mathbf{X}\mathbf{u} = \mathbf{u}'\mathbf{M}_\mathbf{X}\mathbf{u}$$

Since $\mathbf{M}_\mathbf{X}$ is symmetric, there exists a (TxT) matrix $\mathbf{P}$:

$$\mathbf{M}_\mathbf{X} = \mathbf{P}\Lambda\mathbf{P}' \tag{4}$$

and

$$\mathbf{P}'\mathbf{P} = \mathbf{P}\mathbf{P}' = \mathbf{I}_\mathbf{T}$$

Where Λ is a (TxT) matrix with the eigenvalues of $\mathbf{M}_\mathbf{X}$ along the principal diagonal and zeros elsewhere. In this equation, Λ contains k zeros and $(T-K)$ 1s along its principal diagonal.

The Algebra of Linear Regression

Substituting the equation(4) into the numerator:

$$\begin{aligned}\mathbf{u}'\mathbf{M_X}\mathbf{u} &= \mathbf{u}'\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'\mathbf{u} \\ &= (\mathbf{P}'\mathbf{u})'\boldsymbol{\Lambda}(\mathbf{P}'\mathbf{u}) \\ &= \mathbf{w}'\boldsymbol{\Lambda}\mathbf{w} \\ &= w_1^2 + w_2^2 + \cdots + w_{T-K}^2\end{aligned}$$

Where

$$\mathbf{w} = \mathbf{P}'\mathbf{u}$$

Furthermore:

$$E(\mathbf{w}\mathbf{w}') = E(\mathbf{P}'\mathbf{u}\mathbf{u}'\mathbf{P}) = \boldsymbol{\sigma}^2\mathbf{I_T}$$

Thus:

$$E(\mathbf{u}'\mathbf{M_X}\mathbf{u}) = (T-K)\sigma^2$$

Hence:

$$E(\hat{\sigma}^2) = \sigma^2$$

The Algebra of Linear Regression

The variance of $\hat{\sigma}^2$ follows:

$$RSS/\sigma^2 = \mathbf{u}'\mathbf{M_X}\mathbf{u}/\sigma^2 \sim \chi^2_{(T-K)}$$

Again, it is possible to show that under these assumptions, $\hat{\sigma}^2$ is efficient

The Algebra of Linear Regression

Suppose that we wish to test the null hypothesis that $\hat{\beta}_i$, the ith element of $\hat{\boldsymbol{\beta}}$ is equal to some particular value β_i^0. The OLS t-test is given by:

$$\frac{\hat{\beta}_i - \beta_i^0/\sqrt{\sigma^2 \xi^{ii}}}{RSS/\sigma^2/(T-K)} = \frac{\hat{\beta}_i - \beta_i^0/\sqrt{\sigma^2 \xi^{ii}}}{\sqrt{\hat{\sigma}^2/\sigma^2}}$$

Where ξ^{ii} is the row ith, column ith element of $(\mathbf{X}'\mathbf{X})$. The numerator is $N(0,1)$ and the denominator is the square root of $\chi^2_{(T-K)}$ divided by its degree of freedom

Hence, the t-test for the null hypothesis is:

$$t = \frac{\hat{\beta}_i - \beta_i^0}{\hat{\sigma}_{bi}}$$

The Algebra of Linear Regression

More generally, suppose we want a joint test of m different linear restrictions about β as represented by:

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

Here $\mathbf{R}$ is a known (mxk) matrix representing the particular linear combinations of $\boldsymbol{\beta}$ about which we entertain hypotheses and $\mathbf{r}$ is a known $(mx1)$ vector of the values that we believe these linear combinations take on.

For example, if we want to test $\beta_1 = \beta_2 = \beta_3 = 0$:

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \qquad \mathbf{r} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{5}$$

When using linear equality constraints, alternatives are specified as

$$H_1 : \mathbf{R}\beta \neq \mathbf{r}$$

Once both the null and the alternative hypotheses have been postulated, it is necessary to discern whether the data are consistent with the null hypothesis.

Three classes of statistics will be described to test these hypotheses: Wald, Lagrange Multiplier and Likelihood Ratio.

The Algebra of Linear Regression

Wald tests are perhaps the most intuitive: they directly test whether $\mathbf{R}\beta - \mathbf{r}$ is close to zero.

Lagrange Multiplier tests incorporate the constraint into the least squares problem using a lagrangian. If the constraint has a small effect on the minimized sum of squares, the Lagrange multipliers, often described as the shadow price of the constraint in economic applications, should be close to zero. The magnitude of these forms the basis of the LM test statistic.

Finally, likelihood ratios test whether the data are less likely under the null than they are under the alternative. If the null hypothesis is not rejected this ratio should be close to one and the difference in the loglikelihoods should be small

The Algebra of Linear Regression

In this lecture, we will focus on Wald test:

$$W = \frac{(\mathbf{R}\beta - \mathbf{r})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{R}']^{-1}(\mathbf{R}\beta - \mathbf{r})/m}{\hat{\sigma}^2} \sim F_{m,T-K}$$

The Wald test has a more common expression in terms of the *RSS* from

both the restricted and unrestricted models:

$$F_{m,T-K} = \frac{RRSS - URSS/m}{URSS/(T-K)}$$