POL 213 – Spring 2024

Quantitative Analysis in Political Science II

Lecture 3

Lauren Peritz

U.C. Davis

lperitz@ucdavis.edu

April 18, 2024

Table of Contents

Least Squares Estimator in R

Statistical Inference for Linear Models

Simulations of Gauss-Markov Violations

Least Squares Estimator

Last time, we showed that for multivariate regression, we are searching for the parameters $B_0, B_1, B_2, ...B_k$ that define the best linear unbiased model for the relationship between variables $X_1, X_2, ..., X_k$ and Y.

Recall, for k explanatory variables, we have $k + 1$ normal equations.

$$\begin{array}{rcl} B_0 n + B_1 \sum X_{i1} + B_2 \sum X_{i2} + ... + B_k \sum X_{ik} & = & \sum Y_i \\ B_0 \sum X_{i1} + B_1 \sum X_{i1}^2 + B_2 \sum X_{i1} X_{i2} + ... + B_k \sum X_{i1} X_{ik} & = & \sum X_{i1} Y_i \\ B_0 \sum X_{i2} + B_1 \sum X_{i1} X_{i2} + B_2 \sum X_{i2}^2 + ... + B_k \sum X_{i2} X_{ik} & = & \sum X_{i2} Y_i \\ \vdots & & \vdots \\ B_0 \sum X_{ik} + B_1 \sum X_{i1} X_{ik} + B_2 \sum X_{i2} X_{ik} + ... + B_k \sum X_{ik}^2 & = & \sum X_{ik} Y_i \end{array}$$

To write an explicit solution to the normal equations in scalar form would be impractical, even for small k.

Least Squares Estimator

For instance, if $k = 2$ we have the solutions to the system of equations :

$$\begin{aligned}
B_0 &= \bar{Y} - B_1\bar{X}_1 = B_2\bar{X}_2 \\
B_1 &= \frac{\sum X_1^* Y^* \sum X_2^{*2} - \sum X_2^* Y^* \sum X_1^* X_2^*}{\sum X_1^{*2} \sum X_2^{*2} - (\sum X_1^* X_2^*)^2} \\
B_2 &= \frac{\sum X_2^* Y^* \sum X_2^{*2} - \sum X_1^* Y^* \sum X_1^* X_2^*}{\sum X_1^{*2} \sum X_2^{*2} - (\sum X_1^* X_2^*)^2}
\end{aligned}$$

where asterisks denote mean-deviation form (i.e. $Y^* \equiv Y_i - \bar{Y}$).

Instead of dealing with this kind of mess, we moved to use the tools of linear algebra to express our system of equations.

Least Squares Estimator

This is particularly important for handling multivariate regression. Let's remind ourselves what the matrix notation means for the linear regression model:

$$\begin{array}{lcl} Y_1 & = & \beta_0 + \beta_1 x_{11} + \ldots + \beta_k x_{1k} + \varepsilon_1 \\ Y_2 & = & \beta_0 + \beta_1 x_{21} + \ldots + \beta_k x_{2k} + \varepsilon_2 \\ \vdots & = & \qquad\qquad\qquad \vdots \\ Y_n & = & \beta_0 + \beta_1 x_{n1} + \ldots + \beta_k x_{nk} + \varepsilon_n \end{array}$$

The system of equations is summarized as the $\boldsymbol{y}$ vector equal to the product of the $\boldsymbol{X}$ matrix of data times the $\boldsymbol{\beta}$ vector of parameters plus the vector of disturbances, $\boldsymbol{\varepsilon}$.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1, x_{11}, \ldots, x_{1k} \\ 1, x_{21}, \ldots, x_{2k} \\ \vdots \\ 1, x_{n1}, \ldots x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\begin{array}{cccc} \boldsymbol{y} = & \boldsymbol{X} & \boldsymbol{\beta} & +\boldsymbol{\varepsilon} \\ (n \times 1) & (n \times k + 1) & (k + 1 \times 1) & (n \times 1) \end{array}$$

The least squares estimator is:

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

Let's write out $\boldsymbol{X}$ and its transpose, $\boldsymbol{X}'$:

$$\boldsymbol{X} = \begin{bmatrix} 1, x_{11}, x_{12}, \ldots, x_{1k} \\ 1, x_{21}, x_{22}, \ldots, x_{2k} \\ \vdots \\ 1, x_{n1}, x_{n2}, \ldots, x_{nk} \end{bmatrix}, \boldsymbol{X}' = \begin{bmatrix} 1, & 1, & \ldots, & 1 \\ x_{11}, & x_{21}, & \ldots, & x_{n1} \\ x_{12}, & x_{22}, & \ldots, & x_{n2} \\ \vdots & & & \\ x_{1k}, & x_{2k}, & \ldots, & x_{nk} \end{bmatrix}$$

Conformability check?

$\boldsymbol{X}$ is $(n \times k+1)$. $\boldsymbol{X}'$ is $(k+1 \times n)$, so $\boldsymbol{X}'\boldsymbol{X}$ is $(k+1 \times k+1)$. The inverse of this product, $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ is thus also $(k+1 \times k+1)$.

$$\begin{array}{ccccc}
\hat{\beta} & = & (\boldsymbol{X}'\boldsymbol{X})^{-1} & \boldsymbol{X}' & \boldsymbol{y} \\
(k+1 \times 1) & = & (k+1 \times k+1) & (k+1 \times n) & (n \times 1) \\
(k+1 \times 1) & = & (k+1 \times n) & & (n \times 1) \\
(k+1 \times 1) & = & (k+1 \times 1) & &
\end{array}$$

Why check all the dimensions? We're going to implement the calculation in R.

Writing an OLS Estimator in R

Inverting a matrix by hand is tedious and involves many steps. Matrix multiplication is much easier but still bothersome. Let's use R to do each computation. Here are the key commands we need:

- Matrix multiplication: `%*%`
- Transpose: `t()`
- Matrix inverse: `solve()`

Let's use Congressional election data from Tennessee in 2010. We will model the Republican candidate's share of votes for the House (y) as a function of Obama's presidential vote share in 2008 (x_2) and the Republican candidate's financial standing relative to the Democrat in hundreds of thousands of dollars (x_3)

Writing an OLS Estimator in R

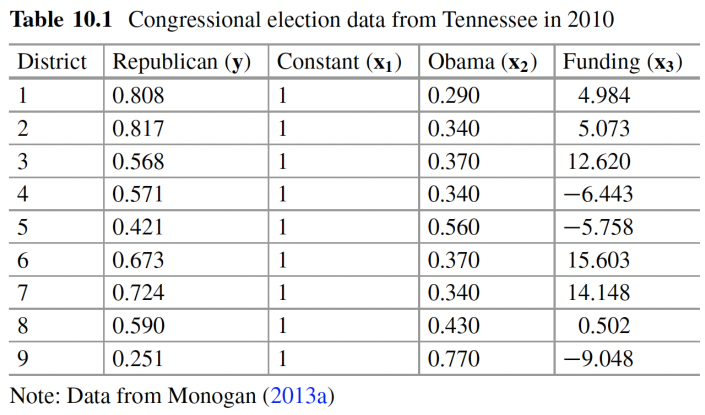

Table 10.1 Congressional election data from Tennessee in 2010

District	Republican ($\mathbf{y}$)	Constant ($\mathbf{x}_1$)	Obama ($\mathbf{x}_2$)	Funding ($\mathbf{x}_3$)
1	0.808	1	0.290	4.984
2	0.817	1	0.340	5.073
3	0.568	1	0.370	12.620
4	0.571	1	0.340	−6.443
5	0.421	1	0.560	−5.758
6	0.673	1	0.370	15.603
7	0.724	1	0.340	14.148
8	0.590	1	0.430	0.502
9	0.251	1	0.770	−9.048

Note: Data from Monogan (2013a)

Create the $\boldsymbol{X}$ and $\boldsymbol{y}$ in R:

```
Y<-c(.808,.817,.568,.571,.421,.673,.724,.590,.251)
X1 <- rep(1, 9)
X2<-c(.29,.34,.37,.34,.56,.37,.34,.43,.77)
X3<-c(4.984,5.073,12.620,-6.443,-5.758,15.603,14.148,0.502,
-9.048)
X<-cbind(X1,X2,X3)
```

To estimate OLS, we simply need to translate the estimator $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ into R syntax:

```
beta.hat<-solve(t(X)%*%X)%*%t(X)%*%Y
beta.hat
```

In class exercise:

Try the calculation with the file OLS_matrix.rmd. Does it match the canned `lm()` command for the coefficients?

Table of Contents

Least Squares Estimator in R

Statistical Inference for Linear Models

Simulations of Gauss-Markov Violations

Properties of Least Squares Estimator

With the matrix of data $\boldsymbol{X}$ fixed, the least squares coefficients result from a linear transformation of the response variable:

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

The least squares coefficient $\boldsymbol{b}$ is an unbiased estimator of β:

$$\mathbb{E}[\boldsymbol{b}] = \mathbb{E}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}] = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbb{E}[y] = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\beta = \beta$$

The covariance matrix of the least squares estimator is derived as follows:

$$\begin{aligned}
\mathbb{V}(\boldsymbol{b}) &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbb{V}(\boldsymbol{y})[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']' \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\sigma_\varepsilon^2\boldsymbol{I}_n[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']' \\
&= \sigma_\varepsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}
\end{aligned}$$

If the variable $\boldsymbol{y}$ is normally distributed, then it can be shown that the linear estimator is also normally distributed:

$$\boldsymbol{b} \sim N_{k+1}[\beta, \sigma_\varepsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}]$$

With these properties established, we can now turn to statistical inference from our multivariate regression model.

Inference for Individual Coefficients

Within $\boldsymbol{b}$, we have coefficients B_j normally distributed with expectation β_j and variance $\sigma_\varepsilon^2 \nu_{jj}$ where ν_{jj} is the jth diagonal entry of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

To test the hypothesis $H_0 : \beta_j = \beta_j^{(0)}$, in theory we could calculate the test statistic Z_0 but we don't know the true error variance.

$$Z_0 = \frac{B_j - \beta_j^{(0)}}{\sigma_\varepsilon \sqrt{\nu_{jj}}}$$

So instead, we use our unbiased estimate of the error variance: $S_E^2 = \boldsymbol{e}'\boldsymbol{e}/(n-k-1)$. Using this estimator, we can estimate the covariance matrix for $\boldsymbol{b}$:

$$\begin{aligned} \mathbb{V}(\hat{\boldsymbol{b}}) &= S_E^2 (\boldsymbol{X}'\boldsymbol{X})^{-1} \\ &= \frac{\boldsymbol{e}'\boldsymbol{e}}{(n-k-1)} (\boldsymbol{X}'\boldsymbol{X})^{-1} \end{aligned}$$

It can be shown that $(n-k-1)S_E^2/\sigma_\varepsilon^2 = \boldsymbol{e}'\boldsymbol{e}/\sigma_\varepsilon^2$ follows a chi-squared distribution with n-k-1 degrees of freedom.

illustration

Return to our example calculations in the file OLS_matrix.rmd.

Use the estimator for the covariance matrix for **b**. Then extract the corresponding standard errors by taking the square root of the diagonal terms.

Inference for Individual Coefficients

In order to estimate the theoretical quantity σ_ε with the observed quantity S_E, we want to account for the additional variability from our sample. So replace the normal distribution with the more spread out $t-$distribution to represent this greater variability.

To test the hypothesis $H_0 : \beta_j = \beta_j^{(0)}$, we calculate the test statistic:

$$t_0 = \frac{B_j - \beta_j^{(0)}}{SE(B_j)}$$

And then compare t_0 with the quantiles of $t_(n - k - 1)$. To find the $100(1 - a)\%$ confidnece interval for β_j, we calculate using critical value:

$$\beta_j = B_j \pm t_{a/2,n-k-1} SE(B_j)$$

Inference for Several Coefficients

Although we usually test regression coefficients individually, these tests may not be sufficient if the least squares estimators of different parameters are correlated.

We would see this in our data if the off-diagonal entries of $\mathbb{V}(\boldsymbol{b}) = \sigma_\varepsilon^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ giving the sample covariances of the LS coefficients have nonzero entries.

$$\mathbb{V}(\hat{\boldsymbol{b}}) = S_E^2 \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2 & \sigma_1\sigma_3 \\ \sigma_2\sigma_1 & \sigma_2^2 & \sigma_2\sigma_3 \\ \sigma_3\sigma_1 & \sigma_3\sigma_2 & \sigma_3^2 \end{bmatrix}$$

In this instance, we might ask about joint hypotheses. For instance, can we *simultaneously* test the hypothesis that all the regression parameters are zero? In other words, our model is junk?

The omnibus F-statistic for the hypothesis $H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$ is:

$$F_0 = \frac{RegSS/k}{RSS/(n-k-1)}$$

R Exercise

- Use Duncan's occupational prestige data by calling `library(carData)`
- Check out R code on Canvas.
- Calculate the standard error for each predictor

Inference for Several Coefficients

Another important approach is to test the null hypothesis that a subset of regression parameters is 0. We do this when we are comparing nested models and we want to compare a more complex versus a less complex specification.
Formally, let k be the number of coefficients in the full (complex) model and a subset $q < k$ be a subset in the simpler model. Then the null hypothesis corresponds to the model:

$$Y = \beta_0 + 0x_1 + 0x_2 + ... + 0x_q + \beta_{q+1}x_{q+1} + ... + \beta_k x_k + \varepsilon$$

such that $H_0 : \beta_1 = ...\beta_q = 0$.

Then we can similarly calculate an F-test comparing these two nested models:

$$F_0 = \frac{(RSS_0 - RSS)/q}{RSS/(n - k - 1)}$$

Let's go back to the R code and try it out.

Table of Contents

Least Squares Estimator in R

Statistical Inference for Linear Models

Simulations of Gauss-Markov Violations

Gauss-Markov Assumptions

Returning to the Gauss Markov Theorem, we have the following assumptions:

1. **Linearity**: The expected (mean) value of the disturbance term is 0.
2. **Nonstochastic regressors**: X values are independent of the error term (or X is exogenous). $cov(X_i, u_i) = 0$.
3. **Homoscedasticity**: constant error variance across values of X_i. Means OLS no longer efficient.
4. **Independence**: No autocorrelation between disturbances. $cov(u_i, u_j) = 0$ for $i \neq j$. Means OLS no longer efficient.

And further, in order to draw any statistical inferences, we also assume:

5. **Normality**: The disturbances are normally distributed, $\varepsilon \overset{\mathrm{d}}{\sim} N_n(\mathbf{0}, \sigma_\varepsilon^2 \boldsymbol{I}_n)$

Simulations

Let's turn to some simulations to study why violations of these assumptions renders our OLS estimator biased, inefficient, or inconsistent. These simulations are from:

https://www.econometrics-with-r.org/