Econ 534: Section Note 02

JJ Chen[*]

September 5, 2014

Of all the principles that can be proposed for this purpose, I think there is none more general, more exact, or easier to apply, than that which we have used in this work; it consists of making the sum of the squares of the errors a *minimum*. By this method, a kind of equilibrium is established among the errors which, since it prevents the extremes from dominating, is appropriate for revealing the state of the system which most nearly approaches the truth. — Legendre, 1805

The principle that the sum of the squares of the differences between the observed and computed quantities must be a minimum may, in the following manner, be considered independently of the calculus of probabilities. — Gauss, 1809 (For S3)

Dr. neyman advocated, wisely, in his opinion, the system which he ascribed to Markoff, though this was in essence the system of Gauss. It must be remembered that if the variances from the different populations were not, on a plausible expectation, to be considered equal, one seldom had prior knowledge or experimental evidence sufficient to make the P_i of Dr. Neyman's equation (8) properly speaking known numbers. — Fisher, 1934 (for S3)

Idea: Add Review Session to your Google calender. (How)

---

[*]jchen215@uic.edu

Bugs might exist. In today's section, we'll

1. review some basic concepts on Euclidean space (10 min),
2. discuss LS solution in Euclidean space (15 min),
3. see the partial regression graphically and numerically in Stata (15 min).

1   Quick Review

In the first two weeks, we have covered:

1. The Least Square (LS) solution $\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, and thus the predicted vector and residual vector: $\boldsymbol{y} = \hat{\boldsymbol{y}} + \hat{\boldsymbol{u}} = \boldsymbol{X}\hat{\beta} + \hat{\boldsymbol{u}}$.
2. Algebraic properties of LS solution:

    (1) P1 - P5 (See lecture notes. Notice that your data matrix should have a constant vector for these properties to be true).
    (2) Measure of goodness of fit: centered R^2 and $\bar{R}^2$.
    (3) Partial regression, projection, and the FWL theorem.

So far we only talked about the **algebric** (or **numerical**) properties of the LS solution. These properties exist simply because we apply the method of LS to data sets. Hence they hold regardless of whether you are talking about a population or a sample at hand. (To minimize confusion, I will adopt the superpopulation idea and use notations like $\hat{\beta}$ and $\hat{\boldsymbol{u}}$ to discuss their algebraic properties. These "hats" are not essential at the moment.)

2   Euclidean Space $(\mathbb{R}^n, \mathbb{R}, +, \cdot, \langle\,\rangle)$

Euclidean space is just our familiar world, in which we are allowed to do vector addition, scalar multiplication, and inner product in the vector space.

1. Vector addition (*visualize it!* If you see "*visualize it!*" hereafter, try to draw a two- or three-dimensional graph by yourself. If you couldn't do it and want to know it, ask your friends who attend the review sessions or come to my office hour).

     • Example: $\hat{\boldsymbol{y}} + \hat{\boldsymbol{u}} = \boldsymbol{y}$.

2. Scalar multiplication (*visualize it!*).

     • Example: $\hat{\beta} \cdot \boldsymbol{x} = \hat{\boldsymbol{y}}$.

3. Inner product (*visualize it!*).

     • Example: $\langle \hat{\boldsymbol{u}}, \hat{\boldsymbol{u}} \rangle \equiv \hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} \equiv \Sigma \hat{u}_i^2$. (Note that inner product defines sum of squares in this example.)

     • With inner product, we are able to define the length (or norm) or a vector (*visualize it!*), $\|\boldsymbol{y}\| = (\boldsymbol{y}'\boldsymbol{y})^{\frac{1}{2}} = \sqrt{\Sigma y_i^2}$, and the angle between two vectors (*visualize it!*), $\cos \theta = \frac{\langle \boldsymbol{y}, \hat{\boldsymbol{y}} \rangle}{\|\boldsymbol{y}\|\|\hat{\boldsymbol{y}}\|}$.

4. Combine vector addition and scalar multiplication, we can use column vectors to form a **subspace** in the N-dimensional Euclidean space.

     • Example: $\boldsymbol{x} = \boldsymbol{x}_1\beta_1 + \boldsymbol{x}_2\beta_2, \forall \beta_1, \beta_2 \in \mathbb{R}$. All linear combinations of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ form a whole 2-D plane. (*Visualize it!*) We call it $\text{span}(\boldsymbol{x}_1, \boldsymbol{x}_2)$, or $\text{col}(\boldsymbol{X})$, where $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2)$.

3   LS Solution in Euclidean Space

Let's now put the LS solution into a Euclidean space. Given a data matrix $\boldsymbol{X}_{nk}$, the columns of $\boldsymbol{X}_{nk}$ (assumed linear independent) form a "explained space", which is denoted $\text{col}(\boldsymbol{X})$, with $\dim = k$.

How about the other $n - k$ dimensions in the n-dimensional Euclidean space? The **orthogonal complement** of $\text{col}(\boldsymbol{X})$, which is denoted $\text{col}^{\perp}(\boldsymbol{X})$ or $\ker(\boldsymbol{X}')$ or $\text{null}(\boldsymbol{X}')$, form a "residual space" with $\dim = n - k$. (These dimensions will be handy when we discuss degrees of freedom.)

3.1   Visualize $\hat{y}$, $\hat{u}$, and their respective spaces

     • Example: Consider $n = 2$ and $k = 1$. Draw the graph for $\boldsymbol{X}, \hat{\beta}\boldsymbol{X}(= \hat{\boldsymbol{y}}), \hat{\boldsymbol{u}}, \boldsymbol{y}, \text{col}(\boldsymbol{X}), \text{null}(\boldsymbol{X}')$. (*Visualize it!*)

- Example: Consider $n = 3$ and $k = 2$. Do the same thing. (*Visualize it!*)

Form the graphs we drew, we should be able to "see" the algebraic properties P1 - P5.

3.2 Visualize TSS, RSS, ESS and R^2

With the graph we drew, we can clearly "see" the Pythagoras' theorem,

$$\|\boldsymbol{y}\|^2 = \|\hat{\boldsymbol{y}}\|^2 + \|\hat{\boldsymbol{u}}\|^2,$$

or

$$TSS = RSS + ESS,$$

and get the geometric meaning of R^2:

$$R^2 = (\frac{\|\hat{\boldsymbol{y}}\|}{\|\boldsymbol{y}\|})^2 = (\cos \theta)^2.$$

Naturally, $0 \leq \cos^2 \theta \leq 1$. Note that if we omit a constant vector in the data matrix, the R^2 here is the uncentered version and is different from the centered R^2 introduced in class. In some sense, the R^2 formula we see in class is one of the many application of the FWL theorem.

3.3 Projection, Partial Regression, and the FWL Theorem

It should be clear from the graph that the $\hat{\boldsymbol{y}}$ comes from a **projection** of $\boldsymbol{y}$ onto $\mathrm{col}(\boldsymbol{X})$ and $\hat{\boldsymbol{u}}$ comes from a projection of $\boldsymbol{y}$ onto $\mathrm{null}(\boldsymbol{X}')$. (Imagining using a "Flashlight" app. in your smart phone.) In linear algebra, a projection (or any transformation) can be represented as a matrix. Thus we have

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \hat{\boldsymbol{u}} = \boldsymbol{P_X} \boldsymbol{y} + \boldsymbol{M_X} \boldsymbol{y},$$

where $\boldsymbol{P_X} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and $\boldsymbol{M_X} = \boldsymbol{I}_n - \boldsymbol{P_X}$ are two projections.

The following properties should be intuitive:

1. $\boldsymbol{P_X}\boldsymbol{P_X} = \boldsymbol{P_X}$. *(Two ("vertical") "Flashlights" make one "Flashlight".)*

2. $M_X M_X = M_X$. *(Two ("horizontal") "Flashlights" make one "Flashlight".)*
3. $P_X M_X = O_n$. *(Two "Flashlights" cancel the effect of each other.)*
4. $P_X + M_X = I_n$. *(Decompose a vector into two and add them back.)*
5. $P'_X = P_X$. *(Projection preserves the norm of a vector.)*
6. $M'_X = M_X$. *(Projection preserves the norm of a vector.)*
7. $P_X P_{X_1} = P_{X_1}$. *(A big "Flashlight" and her little friend.)*
8. $M_X M_{X_1} = M_X$. *(A big "Flashlight" and his little friend.)*

With the projection matrices at hand, we can apply them to partial regression and get to know the FWL theorem: One can get (1) the same LS vectors ($\hat{\beta}_2$) and (2) the residual vectors from the following two regressions:

$$y = X_1\beta_1 + X_2\beta_2 + u, \qquad M_{X_1}y = M_{X_1}X_2\beta_2 + v.$$

There are many applications of the FWL theorem, such as visualize key regression result, deviation from means, detrending, adjustment for seasonal effects, OVB, FE estimation, IV estimation, and so on. I will use Stata to demonstrate some of these applications.

3.4   Stata Demo: The FWL Theorem

The FWL theorem can be described in a 3-step algorithm:

- Step 1: Project y onto the "residual space" $\text{null}(X'_1)$ and get the residual vector $M_{X_1}y$.
- Step 2: Project X_2 onto the "residual space" $\text{null}(X'_1)$ and get the residual vector $M_{X_2}X_2$.
- Step 3: Use two new residual vectors and apply the LS method to get the solution.

Note that Step 1 can be omitted if you are only interested in $\hat{\beta}_2$.

```
clear
set more off
cd "/Users/chan/Documents/Econ 534/"
```

```
* 1. Partial reg and "holding other var constant"

sysuse nlsw88

reg wage ttl_exp, noheader // bivariate case

reg wage ttl_exp age, nohe // partial effect

qui reg wage age
predict yres, res
qui reg ttl_exp age
predict xres, res // partial out the effect of age
reg yres xres, noconstant nohe // how about `reg wage xres, noc`

* 2. Regression Anatomy: Visualize Partial Effect

* ssc install reganat
reganat wage ttl_exp grade, dis(ttl_exp) biline

* 3. Deviation from means

su wage ttl_exp // check out the means
reg wage, noheader // reg with a constant
predict yres2, res // why `yres2` is the demeaned `wage`
reg ttl_exp, noheader // reg with a constant
predict xres2, res // why `xres2` is the demeaned `ttl_exp`

reg yres2 xres2, noc nohe
reg wage ttl_exp, nohe // what can you say about the slope coef.
```