Economics 240A: Administrative Issues and Introduction to Regression Geometry

Michael R. Roberts
Department of Economics and
Department of Statistics
University of California at Berkeley

March 16, 2000

1 Administrative Junk

1.1 Personal Info (Nothing Interesting...Sorry)

My office is 416 Evans and my email is mroberts@econ.berkeley.edu. If you really need to get in touch with me you can call me at home (just ask for my number) anytime before midnight but if you call me before 9 AM, I will kill you. My office hours are TBA. I'd like to get a feel for when everyone would prefer to have them.

1.2 References

Some good references other than Goldberger include the following:

- Econometrics

    1. William Greene, *Econometric Analysis*, (Ch.2 contains a nice collection of useful linear algebra facts employed in econometrics. The rest of the book is a fairly easy read but not very geometrically oriented. The time series stuff is...not so good)
    2. Russell Davidson and James Mackinnon, *Estimation and Inference in Econometrics*, (Nice geometric perspective to regression. Very good reference for econometrics in general)
    3. Peter Kennedy, *A Guide to Econometrics* (A simple, heuristic approach to econometrics. This won't really help you with this or later courses but is nice for taking a step back from the math)

4. Paul Ruud, *An Introduction to Classical Econometrics* (I've only seen pieces of this "big" book while it was in production so I can't really give you the skinny on this. But, if it can explain concepts as well as he can, it's a must have.)

- Linear Algebra and Matrix Analysis

  1. Ben Noble and James Daniel, *Applied Linear Algebra* (undergraduate math text but quite thorough and easy to understand)
  2. Gilbert Strang, *Linear Algebra and It's Applications* (undergraduate math text)

- For a more general theory of vector spaces take a look at the following:

  1. Erwin Kreyszig, *Introductory Functional Analysis with Applications* (A great introductory book on the subject and a relatively easy read)
  2. John Conway, *A Course in Functional Analysis* (A more advanced and abstract take on the subject)

This list is obviously not exhaustive nor is it meant to be. These last two references are completely unnecessary for this course. I put them there only because I found them quite helpful in later courses (finance and probability). For references in any other subject area of econometrics, probability or statistics, please don't hesitate to ask.

1.3 Programming Language

There's no avoiding it, you're going to have to learn some language to work with data and perform analyses. Here's my personal take on some commonly used languages that I've had experience using:

1. Matlab. Fast, accurate, good memory management, and smart algorithms. The toolboxes provide a very comprehensive collection of pre-coded tools. The helpdesk is also quite nice. A lot of people in this department (and others) use it which means a lot of potential helping hands. Not really good for data manipulation though.

2. S-Plus. Sorry, but I think this is a dog. It's very slow and has poor memory management. It can't compete with Matlab except that it is more Statistics oriented.

3. Mathematica. Comparable to Matlab. Better in terms of symbolic math because of greater integration but Matlab's a cleaner matrix programming language.

4. SAS. The best language for data manipulation but the land of a thousand black boxes. I wouldn't recommend doing complex analysis in here because it is not always immediate what proc do-it-all is actually doing. SAS-IML is notorious for poor precision.

5. Stata. A lot of people like this. I think it's similar to SAS but easier to use and with less overhead in terms of getting up to speed.

6. GAUSS. I like to think of this as the precursor to Matlab. Same idea but, as of late, people seem to be choosing Matlab over GAUSS. Still, comparable to Matlab as I understand it.

7. TSP. A lot of nice pre-packaged routines. Not so good for data manipulation. Not really comparable to Matlab.

My final word. The choice is up to you. I think the ideal combination is SAS for data manipulation and Matlab for analysis. I can offer help and advice with these two languages (SAS more than Matlab). As for the others, you're on your own.

1.4 Homework

The word from the boss is you're welcome (even encouraged) to work together but everyone must turn in their own copy of the homework. I plan on going over homeworks in section after they've been turned in so the "unofficial" due date is really at the beginning of section before I start discussing them. There will be three homeworks that will come mainly from Goldberger's text.

2 Regression

This section of lecture notes presents an introduction to the geometry of regression. It is not meant to be comprehensive or rigorous but rather expresses my "simpleton" view of regression. I hope it is helpful and doesn't contain too many errors. Throughout, I try to stay consistent with Goldberger's notation and assume his classic regression (CR) framework which says

$$\begin{aligned} E(y) &= X\beta \\ V(y) &= \sigma^2 I \\ X &: \text{full rank and non-stochastic} \end{aligned}$$

2.1 The Picture

The image to keep in mind is shown on the following page and corresponds to the following regression equation:

$$y = X\beta + \varepsilon \tag{1}$$

for $y, \varepsilon \in R^3, X \in R^{3\times 2}$ and $\beta \in R^2$.

Figure 1:

The pieces of the picture are:

- $Span\,(X)$ is the vector space spanned by the columns of the X matrix. I loosely refer to this as the X-space. This is often called the "regression hyperplane".(the nonlinear version is called a "regression manifold")
- y is the vector of dependent variables.
- $\hat{y}$ is the vector of predicted values. This is equal to Xb where b is the vector of regression coefficients. The geometric interpretation is that $\hat{y} = Ny$ which is the projection of y into the linear space spanned by the columns of X. The matrix N is symmetric and idempotent and is sometimes referred to as the "hat matrix."
- e is the vector of OLS residuals and is defined as follows

$$e = (I - N)\,y = My = M\varepsilon = Y - Xb$$

The matrix $M = I - N$ projects into the orthocomplement of the column space of X, or "error space." Like N, M is symmetric and idempotent.

- θ, the angle between y and our estimate $\hat{y}$. This is going to be useful when discussing the fit of our regression, that is, how much of the variation in y are we able to capture with the regression.

Note that y can be decomposed into the sum of two orthogonal pieces, $\hat{y}$ and e. This follows from the fact that

$$N + M = N + (I - N) = I$$

so

$$y = (N + M)\, y = Ny + My = \hat{y} + e$$

2.2 R^2

Because of the orthogonality between $\hat{y}$ and e

$$\|y\|^2 = \|\hat{y} + e\|^2 = \|\hat{y}\|^2 + \|e\|^2 \tag{2}$$

where $\|x\| = \sqrt{x_1^2 + \ldots + x_n^2}$ is the Euclidean norm. This is just the Pythagorean theorem. Dividing (2) by $\|y\|^2$ gives

$$1 = \frac{\|\hat{y}\|^2 + \|e\|^2}{\|y\|^2}$$

and rearranging yields

$$\frac{\|\hat{y}\|^2}{\|y\|^2} = 1 - \frac{\|e\|^2}{\|y\|^2} \tag{3}$$

If a constant term is in the design matrix, then this quantity is called the centered coefficient of determination or R_C^2. Without a constant term in the design matrix, this is the uncentered coefficient of determination, R_U^2. I'll assume there's a constant in the X matrix and use R^2 to denote R_C^2. Equation (3) is sometimes written in words as

$$\frac{\text{Explained Sum of Squares}}{\text{Total Sum of Squares}} = 1 - \frac{\text{Error Sum of Squares}}{\text{Total Sum of Squares}}$$

An $R^2 = 1$ means the explained sum of squares equals the total sum of squares which is true if the length of $\hat{y}$ is equal to the length of y. This says that y lies in $Span\,(X)$. An R^2 of 0 means that y is orthogonal to $Span\,(X)$ so that the length of e is equal to the length of y.

If we take the square root of the left hand side of (3) we get $\|\hat{y}\| \big/ \|y\|$ which is the ratio of the length of the base of the triangle to the length of the hypotenuse in the picture above. But this is just another way of expressing the cosine of θ.

$$\cos\theta = \|\hat{y}\| \big/ \|y\|$$

So, R^2 is also equal to $\cos^2\theta$ implying that $R^2 \in [0,1]$. An $R^2 = 0$ means the angle between y and $\hat{y}$ is 90^o (i.e. they're orthogonal) while an $R^2 = 1$ implies that $\theta = 0$ so that y lies in the X-space.

By adding a regressor, R^2 will go up or stay the same. The latter occurs if the regressor we add is either a linear combination of existing covariates or is orthogonal to y. By similar reasoning, dropping a regressor will cause R^2 to either fall or stay the same.

2.3 Important Equations

These are just a couple of important equations you should know based on the classic regression framework

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

with

$$\begin{aligned} E(y) &= X\beta \\ V(y) &= \sigma^2 I \\ X &: \text{full rank and non random} \end{aligned}$$

1. OLS estimates $b = (X'X)^{-1} X'y$
2. covariance matrix of $b, V(b) = \sigma^2 (X'X)^{-1}$
3. Projection Matrices: $N = X(X'X)^{-1} X'; M = I - N$, and their properties (symmetric, idempotent, "annihilation" or "killing")
4. OLS estimate of $\hat{\sigma}^2 = e'e/T - k$, where e is the OLS residual, T is the number of observations and k is the number of parameters.

3 Problems

1. Show that $E(y) = E(\hat{y}) = X\beta$.
2. Show that $E(b) = \beta$. (i.e. unbiased)
3. Show that cov(b) is $\sigma^2 (X'X)^{-1}$
4. Show that $NX = X$ and $MX = 0$ where $N = X(X'X)^{-1} X'$ and $M = I - N$.

5. Under what conditions do the residuals sum to 0? If they don't, does this necessarily bias our estimate of σ^2?

   The constant vector must lie in the column space of X. This does not mean that we must have a constant. For example, we could have to dummies, indicating sex. Summing these two vectors would give us a constant vector implying the constant vector lies in $span\left(X\right).$

6. You have data from which you compute

$$\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} y_t \\ x_t \\ 1 \end{pmatrix}\begin{pmatrix} y_t & x_t & 1 \end{pmatrix} = \begin{pmatrix} 20 & 30 & 4 \\ 30 & 50 & 5 \\ 4 & 5 & 1 \end{pmatrix}$$

7. (a) Compute $b = \left(\hat{\beta}\right), \hat{\sigma}, V\left(b\right), R^2$ from the regression $y_t = x_t\beta + \varepsilon_t$

   (b) Compute $b = \left(\hat{\alpha}, \hat{\beta}\right), \hat{\sigma}, V\left(b\right), R^2$ from the regression $y_t = \alpha + x_t\beta + \varepsilon_t$

   (c) Find the covariance matrix of (y_i, x_i) under the assumption of i.i.d pairs.

   (d) Use the result from (c) to calculate $b, \hat{\sigma}, R_U^2$ from the regression $(y_t - \overline{y}) = b\left(x_t - \overline{x}\right) + \varepsilon_t$

8. From the model $y = X\beta + \varepsilon$ where $X \in R^{T\times K}$, let the least squares residual be e. Find the correlation coefficient between the residuals and y, and between the residuals and each of the columns of X.

9. Two series y and x have a correlation of 0.7. If you regress y on x and a constant, you get residuals u and an $R^2 =$?. If you regress x and y and a constant, you get an $R^2 =$? and residuals v. The correlation between u and v is ?. Draw a picture of this result and give a proof.