Section 1

Econ 240A - Second Half

Ingrid Haegele*

University of California, Berkeley

October 11, 2018

*E-mail: inha@berkeley.edu

What is Econ 240A - Second Half about?

In the first half of 240A you were introduced to statistical and probabilistic concepts that form the basis for the rest of this course sequence. In this part we will apply some of those concepts to explain the relationship among different variables.

We usually have some explanatory variable X (e.g. education) and another variable Y which we try to explain (e.g. earnings). Our task is to estimate the parameters or the moments that characterize this relationship. For instance, an extremely interesting object is the conditional expectation of Y given $X = x$, denoted as $\mathrm{E}(Y|X = x)$.

Notice that an object such as $\mathrm{E}(Y|X = x)$ is a *population* object. It is often convenient in Econometrics to use a "population first" approach where the objects of interest are first defined (and possibly identified) in terms of moments of the population. This is exactly how we are going to proceed during the first part of the course. Generally, we don't have at hand the entire distribution of the population. Hence we need to make use of data in order to *infer* something about some objects of the population, such as $\mathrm{E}(Y|X = x)$. Inference represents a crucial part of Econometrics. In the second part of this course, we will learn how to use different approaches to conduct inference about parameters of the linear regression model under a random sampling scheme.

Besides these topics, the goal of this half of Econ 240A is to get you started using Python. Most of the problem sets will have coding exercises in Python and I encourage you to work in groups to teach each other how to best use Python. For a short introduction to Python, please check out the notebook I uploaded to bCourses.

1 Definitions concerning Vector Spaces

The goal of this sectin is to explain how orthogonal projections can be used to find least squares predictions of random variables. We will start this section by recapping some basic concepts of vector spaces. The definitions provided below are meant to summarize the material covered in class in an intutive way. For a formal discussion, please refer to the lecture notes or to Luenberger (1969).

Definition 1: Dot product
The dot product takes two vectors, x and y, and produces a real number x·y. If x and y are represented in Cartesian coordinates, for example R^3, the dot product is defined by

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = x_1 y_1 + x_2 y_2 + x_3 y_3$$

Another key feature of the dot product is its geometric interpretation that links the dot product between two vectors x and y to the length of the vectors and the angle between them:

$$x \cdot y = \|x\| \|y\| cos\theta$$

If the angle between x and y is 90 degree and the vectors are perpendicular, the dot product will be zero since cos(90)=0.

Definition 2: Inner product
The inner product is a generalization of the dot product. In a vector space, it provides the possibility to multiply vectors with the result being a scalar.

An inner product satisfies three properties:

1. Bi-linearity

2. Symmetry

3. Positivity

Definition 3: Inner product space

A vector space that is equipped with an inner product is called an inner product space. Examples for inner product spaces are the real numbers with the standard multiplication as inner product ($\langle x, y\rangle$:=xy) or the Euclidean space with the dot product as inner product ($\langle x, y\rangle := x \cdot y$).

The inner product allows the introduction of intuitive geometrical notions, such as the length of a vector or the angle between two vectors. It makes it also possible to define orthogonality between vectors based on a zero inner product. An inner product naturally induces a norm, thus an inner product space is also a normed vector space.

Definition 4: Hilbert space

A complete vector space with an inner product is called a Hilbert space. The concept of a Hilbert space generalizes the notion of the Euclidean space (such as the notion of distance or orthogonality) to spaces with any finite or infinte number of dimensions.

Definition 5: Norm

Every inner product naturally induces a norm of the form $\|x\| = \sqrt{\langle x, x\rangle}$. The norm can be thought of as the length of the vector x.

In the Euclidean space for example, the norm implies that the length of the vector X $\in R^3$ may be written as $\|x\| = \sqrt{\langle x, x\rangle} = \sqrt{x_1{}^2 + x_2{}^2 + x_3{}^2}$.

When working with inner product spaces, the following three (in)equalities are very useful:

Cauchy-Schwarz Inequality:

$$|\langle X, Y\rangle| \leq \|X\|\|Y\|$$

Triangle Inequality:

$$\|X+Y\| \leq \|X\| + \|Y\|$$

Pythagorean Theorem:
If X $\perp$ Y, then $\|X+Y\|^2 = \|X\|^2 + \|Y\|^2$.

2 Projection Theorem

We can use the above introduced concept of hilbert spaces in order to formulate the projection theorem. The underlying logic stems from the basic geometry insight that one can find the shortest distance between a point and a line by dropping the perpendicular.

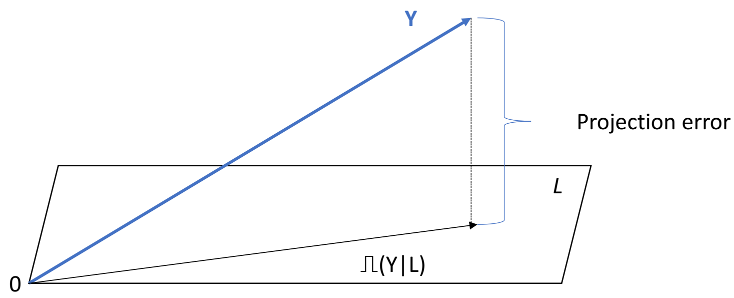
Let $\mathcal{H}$ be a vector space with an inner product and associated norm and $\mathcal{L}$ a closed linear subspace of $\mathcal{H}$. First, we define the projection operator that selects the element of $\mathcal{L}$, which is closest to Y and therefore best approximates Y measured by the corresponding norm. This projection operator can be written as $\Pi(\cdot|\mathcal{L}) : \mathcal{H} \to \mathcal{L}$ where $\Pi(Y|\mathcal{L})$ is the element $\hat{Y} \in \mathcal{L}$ that achieves $min_{\hat{Y}\in\mathcal{L}}\|Y - \hat{Y}\|$.

The first part of the projection theorem states the uniqueness of such a projection operator. The second part of the theorem states the orthogonal condition related to the first order condition.

Projection Theorem:
Let $\mathcal{H}$ be a vector space with an inner product and associated norm and $\mathcal{L}$ a subspace of $\mathcal{H}$., then for Y, an arbitrary element of $\mathcal{H}$., if there exists a vector $\hat{Y} \in \mathcal{L}$ such that $\|Y - \hat{Y}\| \leq \|Y - \tilde{Y}\|$ for all $\tilde{Y} \in \mathcal{L}$, then

1. $\hat{Y} = \Pi(Y|\mathcal{L})$ is unique
2. A necessary and sufficient condition for $\hat{Y}$ to be the uniquely minimizing vector in $\mathcal{L}$ is the orthogonality condition $\langle Y - \hat{Y}, \tilde{Y}\rangle = 0$ for all $\tilde{Y} \in \mathcal{L}$ or $Y - \Pi(Y|\mathcal{L}) \perp \tilde{Y}$ for all $\tilde{Y} \in \mathcal{L}$.

Figure 1: Projection Visualization



3 Applications of Projection Theorem

3.1 Least Squares

Let $\mathcal{L}$ be the linear span of $\mathbf{1}$, $\mathbf{X}$ (i.e., of the form $1\alpha + \mathbf{X}\beta$). We can find the projector $\Pi(Y|\mathcal{L})$ by computing $\hat{\alpha}$ and $\hat{\beta}$ as solutions to

$$\min_{\alpha,\beta \in R^2} \|\mathbf{Y} - \alpha\mathbf{1} - \beta\mathbf{X}\|^2 = \min_{\alpha,\beta \in R^2} \frac{1}{N}\sum_{i=1}^{N}(Y_i - \alpha - \beta X_i)^2$$

This is exactly the least squares fit of Y onto a constant and X that we can get by OLS.

3.2 Best Linear Predictor

Let $\mathcal{L}$ consist of all linear functions of $\mathbf{X}$ and let us use the norm associated with the L^2 Hilbert space. Then our projector $\Pi(Y|\mathcal{L})$ corresponds to computing α_0 and β_0 as solutions to

$$\min_{\alpha,\beta \in R^2} \|\mathbf{Y} - \alpha - \beta\mathbf{X}\|^2 = \min_{\alpha,\beta \in R^2} E[(Y_i - \alpha - \beta X_i)]^2$$

This corresponds to the best linear predictor of Y given X.

3.3 Projection in R^n

You can generalize these findings to computing least square fits in multiple dimensions.

3.4 Projection Matrices

We can use the notion of projection matrices as an alternative to usual regression algebra. To do so, we define two projection operators, $\mathbb{P}_X$ and $\mathbb{M}_X$ as:

$$\mathbb{P}_X = X(X'X)^{-1}X'$$

$$\mathbb{M}_X = I - \mathbb{P}_X$$

To provide intuition for an application, let's consider the case of the OLS estimator. The OLS estimator can be written as $\hat{\beta} = (X'X)^{-1}X'Y$, while the predicted value (that corresponds to our predictor in the notation from above) can be written as $\hat{Y} = X\hat{\beta} = P_X Y$. This is why $\mathbb{P}_X$ is often referred to as the "hat" matrix because it puts a hat on Y.

OLS residuals can be written as $\hat{u} = Y - X'\hat{\beta} = M_X Y$. $\mathbb{M}_X$ is sometimes called the "residual maker" matrix since it converts a vector of outcome data Y into a vector of residuals $\hat{u}$.

As projection operators, $\mathbb{P}_X$ and $\mathbb{M}_X$, are symmetric and idempotent. In addition, the product of these two matrices $\mathbb{P}_X\mathbb{M}_X$ is zero since $\mathbb{P}_X\mathbb{M}_X = \mathbb{P}_X - \mathbb{P}_X\mathbb{P}_X = 0$. We can use these properties to derive many useful properties, such as the Frisch-Waugh-Lovell (FWL) Theorem.