CHAPTER 1

Introduction

1.1. What is econometrics?

Econometrics is a subject which study measurement problems in economics. Econometrics is the only systematic way we know now to examine the reality of the working of an economy. Science have made progress when a better way to examine reality is discovered. Telescope, radio telescope, microscope, electron microscope, X-ray, MRI, fMRI are examples. I believe the same can be said about econometrics. You will see the usefulness of econometrics throughout the course but at the same time you will also come to understand that there are many limitations in the methods. I hope those limitations do not discourage you. Rather, I hope you will take the limitations as challenges that are inviting your contribution.

1.2. Parameters of interest and its relation to a probability model parameter

Throughout the class we refer to an object of measurement as a *parameter*. Suppose you are measuring the income distribution. Then the parameter is a distribution. Suppose you are interested in knowing the shape of the business cycle or the demand function of gasoline. Then the parameter is a function.

You as an economist determine an economic phenomenon you want to study. That should define what parameter you are interested in measuring. Once a parameter of interest is defined, econometric methods should inform you how to measure it. Typically, the parameter of interest is modeled as a part of a probability model. A systematic knowledge that aids in this measurement process is called econometrics. This course is an introduction to this subject.

1.3. Applications

Let's think about some examples of parameters we may be interested in studying. A private firm may be involved in a sex discrimination suit. Lawyers working on both sides want to know if the firm did or did not systematically paid less to a group of women in the firm. The parameter of interest in this case may be the salary difference between comparable men and women in this firm. In order to proceed with this idea, we need to formalize what we mean by "comparable men and women" more precisely and how we summarize the salary of the comparable men and women.

In order to levy property tax we need to know what the property value is for a given property. In this case the parameter of interest is the value of the property given the characteristics of the property. In both cases, one could use the conditional mean function or the conditional median function as the parameters of interest.

1.4. Five ways economic theory and econometrics interacts

Economic theory and econometrics interact in at least five ways. First, economic theories suggests various parameters of interest. For example, a concept of demand function is created in economic theory.

Second, economic theories provide frameworks within which one can conduct measurement. In the context of demand function estimation, it is important to model how the price is determined. Economic theory provides the framework of price determination.

Third, economic theories help restrict the kind of values a parameter of interest can take and thereby help measure the parameter or they give us a set of relevant variables we should take into account. Continuing on the demand function example, economic theory tells us that a demand function is likely to slope downwards with respect to its own prices provided the income effect is not large and that the relevant variables are prices of related goods as well as its own price and variables that affect marginal utility for those goods.

Fourth, in turn, empirical results substantiate theoretical constructs. Deductive reasoning alone will not inform us the elasticity of a demand function, for example.

Fifth, empirical findings may help eliminate certain type of economic models as inconsistent with observations. For example, theoretically it is possible to reduce the tax rate and raise the tax revenue as suggested some time ago. Many, based on empirical results argued that that is not plausible. In this case, empirical results did not win the debate and as a result during the 1980's the federal debt more than doubled as a percentage of GDP (from 26.8% to 44.1%).

Fifth, empirical findings may lead to new theoretical models. For a long time during 1960's many economists believed that there was a trade-off between inflation rate and the unemployment rate in an economy. After we observed the shift in the trade-off some recognized the importance of the role individual's expectation play in the trade-off and led to explicitly modeling forms of expectation in macro economic models. Theory alone does not tell us anything about reality and we stressed the need to substantiate economic theories by empirical work. In addition to the roles of theory as suggesting some parameters of interest and an aid for measurement it is important to recognize that theory provides "explanation" of a phenomenon under study.

For example we observe higher incidences of lung cancers among smokers than among non-smokers. This is the best empirical study can provide without any theory. Although implausible, logically it is possible that people with lung cancer gene tend to like smoking so it may be the lung cancer gene that causes smoking. The empirical evidence alone cannot distinguish the two. If someone comes up with a mechanism under which smoking raises the incidence of lung cancer, then the theory may suggest some ways to distinguish the two hypotheses. In this way theories may provide a framework of "explaining" the phenomenon under study, in this case, of smoking and lung cancer relationship. Our explanation is as good as what the current theory is. In turn the theory needs to be substantiated by empirical evidences. We need both.

In conducting empirical studies, for which this course provides tools, it is important to remember this limitation of what theories and empirical studies can each accomplish. Empirical studies in themselves do *not* provide explanation. It is also important to take advantage of whatever information theories provide in conducting the measurements.

1.5. Exercises

(1) Think about what kind of economic phenomenon you want to study in your second year paper and explain what the main parameter of interest is. Is it a number, a function, or a stochastic process, or something else? It is likely that you haven't really thought about your thesis yet. In that case, please think of an economic phenomenon you are interested in studying.

CHAPTER 2

Conditional Probability Models

2.1. Probability model

In conducting an econometric measurement of a parameter of interest, like any other measurement problems, probability model is used. Thus we need to relate the parameter of interest with a particular parameter in a probability model. This is the uniquely econometric issue which we need to think carefully about. Since we typically do not address a completely new measurement issue, there is a standard framework literature uses for each of the measurement problem. We should make sure to think carefully whether the framework used is appropriate.

The parameter of interest which frequently arises in economics is how one variable affects another variable. As we saw earlier, the demand function relates its own price to its demand given other prices. Another example is a relationship between wage and factors affecting the human capital, such as education and experience given age and gender, industry, and occupation.

As these examples suggest, most cases in economics where we examine relationship between two variables require holding some other variables at some given values. However, if a randomized experiment is possible, we may not need to hold other variables at some given values, as we will see later in the course. Non-experimental data are usually referred to as observational data and distinguished from experimental data. Thus one may say that the requirement for holding other variables at some given values often arises because we often need to analyze observational data. More about this point later.

These relationships are studied using the concept of conditional distribution. Often, rather than studying the full conditional distribution, the conditional mean function is used to study the relationship. However, the conditional median function can be used, or more generally the conditional quantile function for different quantiles can be used. By doing so, a more complete understanding about the conditional distribution can be obtained.

2.2. Properties of conditional mean function

By far the most frequently used approach to studying a relationship between variables as a parameter of interest is the approach using the concept of conditional mean function. As discussed earlier, however, it should be recognized that the conditional mean function is only one aspect of the relationship between two or more random variables.

Below we will discuss various properties of conditional mean function. We denote the two random variables we focus on by Y and X_1 and denote the rest of random variables we take as given by X_2, which is a vector. We denote $X' =$

(X_1, X_2'), where X' denote the transpose of X. Let $m(x) = E(Y|X = x)$. We write $m(X)$ as $E(Y|X)$. Under this notation, we can show that

(1) $E(g(X)|X) = g(X)$.
(2) $E(g(X)Y|X) = g(X)E(Y|X)$.

To see the first relationship,

$$E(g(X)|X = x) = E(g(x)|X = x) = g(x)E(1|X = x) = g(x)$$

so that $E(g(X)|X) = g(X)$.

To see the second relationship,

$$\begin{aligned} E(g(X)Y|X = x) = E(g(x)Y|X = x) \\ = g(x)E(Y|X = x), \end{aligned}$$

so that $E(g(X)Y|X) = g(X)E(Y|X)$.

Here we list useful relationships we often use.

(1) Law of Iterated Expectations: $E(Y) = E[E(Y|W)]$.
(2) Its conditional version: $E(Y|X) = E[E(Y|X, Z)|X]$.
(3) If Y and X_1 are independent given X_2, then $E(Y|X_1 = x_1, X_2 = x_2) = E(Y|X_2 = x_2)$.
(4) If $U \overset{def}{=} Y - E(Y|X)$, then $E(U|X) = 0$ so that for any function $g(\cdot)$ for which $E(|g(X)U|) < \infty$, $E(g(X)U) = 0$. In particular, $E(U) = 0$ and $Cov(g(X), U) = 0$.
(5) If $c : R \to R$ is a convex function defined on R and $E(|X|) < \infty$. Then $c(E(X|Z)) \le E(c(X)|Z)$.
(6) $Var(Y) = E(Var(Y|X)) + Var(E(Y|X))$ where $Var(Y|X) = E[(Y - E(Y|X))^2|X]$.
(7) The conditional version is

$$Var(Y|X) = E(Var(Y|X, Z)|X) + Var(E(Y|X, Z)|X)$$

(8) When $E(Y^2) < \infty$, conditional mean function $m(X) = E(Y|X)$ can be characterized as a solution the following minimization problem

$$\min_{g(\cdot)} E[(Y - g(X))^2].$$

To see 3,

$$\begin{aligned} &\int yf(y|X_1 = x_1, X_2 = x_2)dy \\ &= \int yf_{Y,X_1,X_2}(y, x_1, x_2)/f_{X_1,X_2}(x_1, x_2)dy \\ &= \int yf_{Y,X_1|X_2}(y, x_1|x_2)f_{X_2}(x_2)/f_{X_1,X_2}(x_1, x_2)dy \\ &= \int yf_{Y|X_2}(y|x_2)f_{X_1|X_2}(x_1|x_2)f_{X_2}(x_2)/f_{X_1,X_2}(x_1, x_2)dy \\ &= \int yf_{Y|X_2}(y|x_2)dy. \end{aligned}$$

The last equality follows because $f_{X_1|X_2}(x_1|x_2)f_{X_2}(x_2) = f_{X_1,X_2}(x_1, x_2)$.

To see 4, note that $E(U|X) = E[Y - E(Y|X)|X] = E(Y|X) - E(Y|X) = 0$. Thus

$$\begin{aligned} E[g(X)U] &= E\{E[g(X)U|X]\} \\ &= E\{g(X)E(U|X)\} \\ &= 0. \end{aligned}$$

Claim 5 is the conditional version of the so called Jensen's inequality. To see this, note that when c is a convex function, one can find a linear function with slope a, $a(x - E(X|Z = z)) + c(E(X|Z = z))$, that touches the convex function $c(\cdot)$ at $(E(X|Z = z), c(E(X|Z = z)))$, but lies entirely below the convex function. Since $c(x) \geq a(x - E(X|Z = z)) + c(E(X|Z = z))$, $c(X) \geq a(X - E(X|Z = z)) + c(E(X|Z = z))$ so that

$$\begin{aligned} E(c(X)|Z = z) &\geq a(E(X|Z = z) - E(X|Z = z)) + c(E(X|Z = z)) \\ &= c(E(X|Z = z)). \end{aligned}$$

Thus $E(c(X)|Z) \geq c(E(X|Z))$.
This implies $E(Y^2) \geq E(Y)^2$ and $E(Y^2|X) \geq E(Y|X)^2$.

Claim 6 can be obtained using the properties of the conditional expectation operator and thus will be left as an exercise.

Claim 7 implies $E[Var(Y|X)] \geq E\{E[Var(Y|X, Z)|X]\}$ or via the law of iterated expectations, $E[Var(Y|X)] \geq E[Var(Y|X, Z)]$.

To see 8,

$$\begin{aligned} E[(Y - g(X))^2] &= E\{[(Y - m(X) + (m(X) - g(X))]^2\} \\ &= E\{[Y - m(X)]^2\} + E\{[m(X) - g(X)]^2\} \\ &\quad + 2E\{U[m(X) - g(X)]], \end{aligned}$$

where $U = Y - m(X)$. Using Claim 4 above $E\{U[m(X) - g(X)]] = 0$ so that

$$E\{[Y - g(X)]^2\} = E\{[Y - m(X)]^2\} + E\{[m(X) - g(X)]^2\}.$$

Clearly the left hand-side is minimized when $g(x)$ is chosen to be equal to $m(x)$.

2.3. Conditional mean function and the average treatment effect

Often we examine the conditional expectation function in order to study causal effect of one variable on another. Here we define the concept of causal effect and then discuss conditions under which the conditional mean function can be used to study causal effect.

In order to define the causal effect, we introduce a new notation to clearly describe the dependence of Y on X, so that

$$Y = Y(X, \omega).$$

The unobserved random variable ω captures the randomness in Y beyond the randomness driven by X.

We define the treatment effect of X_1 on Y when X_1 changes from x_1 to x_1' and when $X_2 = x_2$ as

$$Y(x_1', x_2, \omega) - Y(x_1, x_2, \omega).$$

Suppose we observe the value of Y for which $X_1 = x_1'$ and $X_2 = x_2$. Without knowing ω, we know $Y(x_1', x_2, \omega)$. However, because we do not know ω, we would not know $Y(x_1, x_2, \omega)$. Analogously, by observing the value of Y for which $X_1 = x_1$

and $X_2 = x_2$, without knowing ω, we know $Y(x_1, x_2, \omega)$ but, since we do not know ω, we would not know $Y(x_1', x_2, \omega)$. Either way, we only know either $Y(x_1', x_2, \omega)$ or $Y(x_1, x_2, \omega)$, but not both for the same ω. Therefore the treatment effect

$$Y(x_1', x_2, \omega) - Y(x_1, x_2, \omega)$$

we just defined cannot be directly obtained in data.

However, we show that the average treatment effect can be obtained under an additional assumption. To see this, we first define the average treatment effect of X_1 on Y when X_1 changes from x_1 to x_1' and when $X_2 = x_2$ as

$$E[Y(x_1', x_2, \omega) - Y(x_1, x_2, \omega)].$$

Note that

$$E[Y(x_1', x_2, \omega) - Y(x_1, x_2, \omega)] = E\{E[Y(x_1', x_2, \omega) - Y(x_1, x_2, \omega)|X_2]\},$$

one can obtain the average treatment effect if one can obtain

$$E[Y(x_1', x_2, \omega) - Y(x_1, x_2, \omega)|X_2 = x_2].$$

Since

$$E[Y(x_1', x_2, \omega) - Y(x_1, x_2, \omega)|X_2 = x_2] = E[Y(x_1', x_2, \omega)|X_2 = x_2] - E[Y(x_1, x_2, \omega)|X_2 = x_2],$$

we can obtain the average treatment effect if one can obtain

$$E[Y(x_1', x_2, \omega)|X_2 = x_2] \text{ and E[Y(}x_1\text{,}x_2\text{,}\omega\text{)|X}_2\text{=}x_2\text{]},$$

But, when Y and X_1 are independent given X_2, Claim 3 above implies that

$$\begin{aligned} E[Y(X_1, X_2, \omega)|X_1 = x_1', X_2 = x_2] &= E[Y(x_1', X_2, \omega)|X_1 = x_1', X_2 = x_2] \\ &= E[Y(x_1', X_2, \omega)|X_2 = x_2]. \end{aligned}$$

Analogously

$$E[Y(X_1, X_2, \omega)|X_1 = x_1, X_2 = x_2] = E[Y(X_1, X_2, \omega)|X_2 = x_2].$$

Therefore

$$\begin{aligned} &E[Y(X_1, X_2, \omega)|X_1 = x_1', X_2 = x_2] - E[Y(X_1, X_2, \omega)|X_1 = x_1, X_2 = x_2] \\ &\quad = E[Y(x_1', x_2, \omega)|X_2 = x_2] - E[Y(x_1, x_2, \omega)|X_2 = x_2] \\ &\quad\quad = E[Y(x_1', x_2, \omega) - Y(x_1, x_2, \omega)|X_2 = x_2]. \end{aligned}$$

Integrating over X_2 using the marginal distribution of X_2, we obtain the average treatment effect.

We have discussed the assumption under which the average treatment effect can be obtained using the conditional mean function. Note that it would be great if we can obtain the median or more generally a quantile of

$$Y(x_1', x_2, \omega) - Y(x_1, x_2, \omega).$$

However, that is not possible because we cannot observe

$$Y(x_1', x_2, \omega) - Y(x_1, x_2, \omega).$$

The average treatment effect is special in that we do not need a joint distribution of $Y(x_1', x_2, \omega)$ and $Y(x_1, x_2, \omega)$ due to the linearity in expectation operator:

$$E[Y(x_1', x_2, \omega) - Y(x_1, x_2, \omega)|X_2 = x_2] = E[Y(x_1', x_2, \omega)|X_2 = x_2] - E[Y(x_1, x_2, \omega)|X_2 = x_2].$$

When X_1 is a continuous random variable, one can consider taking a sequence of x_1' which converges to x_1:

$$\frac{E[Y(X_1,X_2,\omega)|X_1=x_1',X_2=x_2]-E[Y(X_1,X_2,\omega)|X_1=x_1,X_2=x_2]}{x_1'-x_1}$$
$$=E\left[\frac{Y(x_1',x_2,\omega)-Y(x_1,x_2,\omega)}{x_1'-x_1}|X_2=x_2\right]$$

A sufficient conditions, in addition to the assumption that Y and X_1 are independent given X_2, for the limit on both sides of the equality to exist and equal are that (1) $E[Y(X,\omega)]<\infty$, (2) $Y(x,\omega)$ is continuously partially differentiable with respect to x_1, (3) $|\partial Y(x,\omega)/\partial x_1|\le M(\omega)$ in a neighborhood of x with $E[M(\omega)]<\infty$.

Under these conditions

$$\frac{\partial E[Y|X=x]}{\partial x_1}=E\left[\frac{\partial Y}{\partial x_1}|X_2=x_2\right].$$

2.4. Exercises

(1) Consider the discrete random vector (Y,X) where Y and X takes on values y_j, $j=1,\dots,J$ and x_k, $k=1,\dots,K$, respectively with $(Y,X)=(y_j,x_k)$ with probability p_{jk} for $j=1,\dots,J$ and $k=1,\dots,K$.
   (a) What is the (marginal) distributions of Y and X, respectively?
   (b) Describe the random variable $E(Y|X)$.
   (c) In this example, show that $E(Y)=E(E(Y|X))$.

(2) For a random variable Y and a random vector X of length K, show that $Var(Y|X)=E[Var(Y|X,Z)|X]+Var[E(Y|X,Z)|X]$.

(3) For a random variable Y and a random vector X of length K, write out the elements of $E(XX')$, $Var(X)$, and $Cov(X,Y)$ using elements of X such as X_k for $k=1,\dots,K$.

(4) Denote the transpose of a matrix or a vector by the prime so that the transpose of a vector x is x' and the transpose of a matrix A is A'. Show that $A=(A')'$.

(5) Let A be an $m\times n$ matrix and its jth column is denoted as a_j. Let B be an $n\times m$ matrix and its jth row is denoted as b_j'. Show that $AB=\sum_{j=1}^{n}a_jb_j'$ and that

$$BA=\begin{pmatrix} b_1'a_1 & b_1'a_2 & \cdots & b_1'a_n \\ b_2'a_1 & b_2'a_2 & \cdots & b_2'a_n \\ \vdots & \vdots & \ddots & \vdots \\ b_n'a_1 & b_n'a_2 & \cdots & b_n'a_n \end{pmatrix}.$$

(6) For the same matrices A and B as above, show that $(AB)'=B'A'$.

(7) If $A'A=0$, then $A=0$.

(8) Define what the row rank and the column rank of an $m\times n$ matrix A are. Show that they are equal when either of the rank is 1. (Can you show that they are the same more generally? This result allows us to talk about the rank of a matrix.)

(9) Let A and B be an $m\times n$ matrix and an $n\times m$ matrix, respectively.

(a) Express column vectors of matrix AB in terms of column vectors of A to show that the column rank of AB does not exceed the column rank of A.

(b) Similarly, express the row vectors of matrix of AB in terms of row vectors of B to show that the row rank of AB does not exceed the row rank of B.

(c) Use the above results and the fact the row rank and the column rank are the same (Question 8), to prove that the rank of AB does not exceed the rank of A or the rank of B.

(10) Show that $trace(AB) = trace(BA)$.