Theory and Applications for Econometrics

Dr. Dimitrios D. Thomakos

October 19, 2009

Graduate Program in Economic Analysis
Department of Economics, University of Peloponnese

Primer on Conditional Expectations and Econometric Models

Theoretical economic models are meant to explain "average" behavior and cannot possibly account for idiosyncratic preferences and actions. However, many of these models are presented as deterministic without incorporating any notion of uncertainty or stochastic behavior from economic agents. This presents a mild(!) problem when we want to incorporate uncertainty and a stochastic component when formulating and estimating econometric models. The notion and workings of conditional expectations are both intuitively accessible and mathematically tractable to allow us to make a smooth transition from economic-theoretic models to estimable econometric models.

Consider thus a simple economic model with two continuous variables, Y and X. Both are observable random variables whose properties are completely determined by their joint probability density function $f(x, y; \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a $(p \times 1)$ vector of structural model parameters, not necessarily corresponding to moments of the random variables Y and X. If the form of $f(x, y; \boldsymbol{\theta})$ was known then we could easily model the joint and marginal behavior of Y and X, or the conditional behavior of Y *given* X. However, in almost all modeling instances we either do not know what the form of the joint density is or we are not willing to make a (strong) assumption about a particular functional form for the density. We are thus lead to simplifications and in particular to conditional modeling.

Suppose thus that economic theory suggests some functional relationship where the "average" behavior of Y depends on the evolution of X through a function $g(X)$, the function $g(\cdot)$ itself not necessarily known - maybe we know what the sign of the derivative $g'(\cdot)$ is. The problem we face is how to properly define "average" behavior and what form we might adopt for $g(X)$. To proceed let us first assume that we actually *know* what the form of $f(x, y; \boldsymbol{\theta})$ is. By appropriate integrations we can obtain the conditional distribution of Y given $X = x$ as:

$$f_y(y|x;\boldsymbol{\theta}) \stackrel{\text{def}}{=} \frac{f(x,y;\boldsymbol{\theta})}{f_x(x;\boldsymbol{\theta})} \tag{1}$$

where $f_x(\cdot)$ is the marginal distribution of X alone, i.e. $f_x(\cdot) \stackrel{\text{def}}{=} \int_{\mathcal{R}_y} f(x,y;\boldsymbol{\theta})dy$; remember that the marginalizing takes place over all values of Y, i.e. over $\mathcal{R}_y$. Now, for any given value of x the conditional distribution is a proper density function and therefore we can find its expected value: the conditional expected value of Y given $X = x$, for x fixed, as:

$$\mathsf{E}\left[Y|x;\boldsymbol{\theta}\right] \stackrel{\text{def}}{=} \int_{\mathcal{R}_y} y f_y(y|x;\boldsymbol{\theta})dy \tag{2}$$

again the integration taking place over all values of Y, i.e. over $\mathcal{R}_y$.

Under our assumption of knowing what the form of the joint density is we now have that the right-hand side of the above equation has a particular form. As an example, if we knew that the joint density of Y and X was multivariate normal then we could show, after doing the algebra, that the form of the conditional expected value would be given as:

$$\mathsf{E}\left[Y|x;\boldsymbol{\theta}\right] \stackrel{\text{def}}{=} \mathsf{E}\left[Y\right] + (x - \mathsf{E}\left[X\right])\frac{\mathsf{Cov}\left[Y,X\right]}{\mathsf{Var}\left[X\right]} \tag{3}$$

where now $\boldsymbol{\theta} \stackrel{\text{def}}{=} (\mathsf{E}\left[Y\right], \mathsf{E}\left[X\right], \mathsf{Cov}\left[Y,X\right], \mathsf{Var}\left[Y\right], \mathsf{Var}\left[X\right])$. Defining the composite parameters $\beta \stackrel{\text{def}}{=} \mathsf{Cov}\left[Y,X\right]/\mathsf{Var}\left[X\right]$ and $\alpha \stackrel{\text{def}}{=\!=} \mathsf{E}\left[Y\right] - \mathsf{E}\left[X\right]\beta$ we can re-write the above conditional expectation in the familiar regression format:

$$\mathsf{E}\left[Y|x;\boldsymbol{\theta}\right] = \alpha + \beta x \tag{4}$$

This example is an important and illuminating one for it (a) shows us that under the presence *or* assumption of multivariate normality the conditional expectation takes the form of a regression and fully justifies linear modeling - in fact we now have that $g(x) \equiv \alpha + \beta x$ and (b) shows us that if we knew what the joint density was we could construct an appropriate model for the conditional expectation - and then we could easily use that for modeling "average" behavior.

Unfortunately, however, all the above rarely holds in practice: we either have no idea what the joint density is or the assumption of multivariate normality is just too strong to be maintained. What are we to do then? We clearly understand that the conditional expected value is an appropriate way for modeling "average" behavior but how do we work with it if we have not other information?

Well, we can show that if all we care about is the "average" behavior of Y, as given by the conditional expectation, an important property holds: the conditional expectations is the "best possible approximation" to any unknown function $g(x)$ for any value of x. This is an important property with implications, which we will know prove. To begin with note that when we allow

x to take values over its domain of definition, say $\mathcal{R}_x$, (e.g. in the case of repeated sampling) we have that the conditional expectation is not a constant anymore but a random variable $\mathsf{E}[Y|X;\boldsymbol{\theta}]$. As such it has moments, such as an expected value and a variance. To find these moments we need to recognize that the source of uncertainty in $\mathsf{E}[Y|X;\boldsymbol{\theta}]$ is coming from the conditioning variable X and therefore we need the marginal density $f_x(x;\boldsymbol{\theta})$ for all calculations. We can thus have (by repeated substitutions):

$$\mathsf{E}[\mathsf{E}[Y|X;\boldsymbol{\theta}]] \overset{\text{def}}{=} \int_{\mathcal{R}_x} \mathsf{E}[Y|x;\boldsymbol{\theta}]\, f_x(x;\boldsymbol{\theta}) = \int_{\mathcal{R}_x}\int_{\mathcal{R}_y} yf(x,y;\boldsymbol{\theta})dydx = \int_{\mathcal{R}_y} yf_y(y;\boldsymbol{\theta})dy \equiv \mathsf{E}[Y] \tag{5}$$

that is, the expected value of a conditional expectation (over the range of values of the conditioning variable) is exactly equal to the unconditional expectation! This result is known as the *law of iterated expectations* or *LIE* - a result of much usefulness in econometric modeling.

We next turn to the issue of "best possible approximation" alluded to earlier. This relates to the relationship between the unknown approximating function $g(X)$, the variance of Y and the conditional expectation. Any such approximating function should not deviate a lot from the observed Y's, "on average". That is, we should choose such a function $g(X)$ whose "predictions" about Y should not deviate much from what is observed in practice (in our data). Consider thus the problem of "minimum average squared deviation" casted as:

$$\text{Find } g(X) \text{ such that you minimize the average squared deviation } Q[g(X)] = \mathsf{E}\left[(Y-g(X))^2\right] \tag{6}$$

To solve this problem we make immediate use of the *LIE* since we can re-write the expression above as:

$$Q[g(X)] = \mathsf{E}\left[\mathsf{E}\left[(Y-g(X))^2|X;\boldsymbol{\theta}\right]\right] \tag{7}$$

and note that we only need to minimize the inner conditional expectation with respect to $g(X)$. Squaring, taking conditional expectations and then taking the derivative with respect to $g(X)$ we obtain:

$$\frac{\partial \mathsf{E}\left[(Y-g(X))^2|X;\boldsymbol{\theta}\right]}{\partial g(X)} = 0 \Rightarrow 2g(X) - 2\mathsf{E}[Y|X;\boldsymbol{\theta}] = 0 \Rightarrow g(X) = \mathsf{E}[Y|X;\boldsymbol{\theta}] \tag{8}$$

which shows us that, in a *mean squared* sense, the conditional expectation is the "best possible approximation" among all candidate functions that we could propose for explaining the "average" behavior of Y given any value of X. This result forms the basis of econometric modeling. Note however, that it does not solve the problem of the functional form for $\mathsf{E}[Y|X;\boldsymbol{\theta}]$; the frequently used linear approximation $\mathsf{E}[Y|X;\boldsymbol{\theta}] = \alpha + \beta X$ is used for convenience and tractability and nothing more - and remember that it holds only if Y and X have a jointly normal distribution!

OK, now how are all the above useful to us? We already established that if economic theory suggests a model for "average" behavior for Y given X we should be approximating this model by $g(X) = \mathsf{E}[Y|X;\boldsymbol{\theta}]$, the conditional expectation of Y given any possible X. However, the observed values of Y need not exactly correspond to the "predictions" given by the "model", i.e. by the values taken by $\mathsf{E}[Y|X;\boldsymbol{\theta}]$. Various discrepancies, such as recording problems or (more importantly) incomplete knowledge about the mechanism that $g(X)$ works, lead us to have "errors" in our model. These errors should be fairly small, if our theory is close to actual observations, and they should have certain statistical properties if they are to be considered random. Suppose that such modeling errors are introduced additively as:

$$Y = \mathsf{E}[Y|X;\boldsymbol{\theta}] + u \tag{9}$$

where u has to be a *non-observable* (why?) random variable with certain properties. The above equation is a *correctly specified* econometric model.

What can be usefully said about these errors based on or analysis so far? An immediate consequence of the *LIE* is that these errors have conditional and unconditional expectation equal to zero since:

$$\mathsf{E}[u|X;\boldsymbol{\theta}] = \mathsf{E}[Y|X;\boldsymbol{\theta}] - \mathsf{E}[Y|X;\boldsymbol{\theta}] = 0 \Rightarrow \mathsf{E}[u] = \mathsf{E}[\mathsf{E}[u|X;\boldsymbol{\theta}]] = 0 \tag{10}$$

In addition, these errors must be uncorrelated with the conditioning variable since we have the following:

$$\mathsf{Cov}[X,u] = \mathsf{E}[Xu] - \mathsf{E}[X]\,\mathsf{E}[u] = \mathsf{E}[\mathsf{E}[Xu|X;\boldsymbol{\theta}]] = \mathsf{E}[X\,\mathsf{E}[u|X;\boldsymbol{\theta}]] = 0 \tag{11}$$

by the zero mean property of the modeling errors.

Note the importance of the above results for practical modeling: if the economic-theoretic model *or* the approximation used for $\mathsf{E}[Y|X;\boldsymbol{\theta}]$ is incorrect then we may not have the zero mean property $\mathsf{E}[u|X;\boldsymbol{\theta}] = 0$ for the modeling errors and, consequently, we will also not have the uncorrelatedness result about the conditioning variable and the errors. If these properties do not hold they reveal something about our econometric model, namely that it is not properly specified! When $\mathsf{E}[u|X;\boldsymbol{\theta}] \neq 0$ we must have that there is additional information ("hidden" in u) which we have not taken into account: this is the issue of "endogeneity" which is common in applied econometric modeling.