Two Conceptions of the Linear Model

Charlie Gibbons
ARE 212

Spring 2011

CEF decomposition

Let y be an outcome and x be a vector of predictors. Then,

$$y = \mathbb{E}[y \mid x] + \epsilon,$$

where $\mathbb{E}[\epsilon \mid x] = 0$:

$$\mathbb{E}[\epsilon \mid x] = \mathbb{E}[y - \mathbb{E}[y \mid x] \mid x] = \mathbb{E}[y \mid x] - \mathbb{E}[y \mid x] = 0.$$

$\mathbb{E}[y \mid x]$ is the *conditional expectation function* and the goal of econometrics is to estimate this function, that is, to understand how y changes with x.

The two conceptions

In class, we have discussed two ways of approaching the estimation of the CEF:

- *Linear conditional expectation function* model
  We claim that we know how the world works—specifically, the true model is linear in x'.
- *Linear projection* model
  We do *not* know how the world works, but we use a linear model that is the closest approximation to the true model. Note that this approximation
  - May be poor and
  - May not have economic meaning (*e.g.*, the derivatives of this model may not accurately give the impact of increasing x by one unit on y).

Conditional mean independence in the LCE

If we follow the LCE model, then the error $\epsilon \equiv y - x'\beta$ has 0 conditional expectation:

$$\begin{aligned}\mathbb{E}[\epsilon \mid x] &= \mathbb{E}[y - x'\beta \mid x] \\ &= x'\beta - x'\beta \\ &= 0.\end{aligned}$$

This is *not* an assumption; it follows from the assumption that we know the true model.

This property is called *conditional mean independence*, which is weaker than standard statistical independence.

Functions of x and the error

Note that this implies that any function of x is uncorrelated with the error:

$$\begin{aligned}
\mathrm{Cov}(h(x), \epsilon) &= \mathbb{E}[h(x)\epsilon] - \mathbb{E}[h(x)]\mathbb{E}[\epsilon] \\
&= \mathbb{E}\left[h(x)\mathbb{E}[\epsilon \mid x]\right] - \mathbb{E}[h(x)]\mathbb{E}\left[\mathbb{E}[\epsilon \mid x]\right] \\
&= 0.
\end{aligned}$$

Linear conditional expectation

Conditional mean independence for ϵ implies 0 correlation with x:

$$\begin{aligned} \text{Cov}(x, \epsilon) &= \mathbb{E}[x\epsilon] - \mathbb{E}[x]\mathbb{E}[\epsilon] \\ &= \mathbb{E}\left[x\mathbb{E}[\epsilon \mid x]\right] - \mathbb{E}[x]\mathbb{E}\left[\mathbb{E}[\epsilon \mid x]\right] \\ &= 0. \end{aligned}$$

Linear projection and the true model

Now, suppose that $y = m(x) + \epsilon$, where ϵ has 0 conditional mean.

We do not know $m(x)$ and approximate it using $x'\beta$. We have

$$\begin{aligned} y &= m(x) + x'\beta - x'\beta + \epsilon \\ &= x'\beta + \left[\big(m(x) - x'\beta\big) + \epsilon\right] \\ &= x'\beta + \tilde{\epsilon}. \end{aligned}$$

Conditional expectation of $\tilde{\epsilon}$

The conditional expectation of $\tilde{\epsilon}$ is

$$\begin{aligned}\mathbb{E}\left[\tilde{\epsilon} \mid x\right] &= \mathbb{E}\left[m(x) - x'\beta \mid x\right] + \mathbb{E}\left[\epsilon \mid x\right] \\ &= m(x) - x'\beta.\end{aligned}$$

This is only 0 if $m(x) = x'\beta$; *i.e.*, the linear conditional expectation model holds.

If the LCE model holds, then the linear projection is the linear conditional expectation function.

All of the properties of the LPM hold for the LCE, but knowing the true CEF gives us the ability to interpret parameters and we have conditional mean independence of our errors.

Meaning of β

In the LCE, β is some true parameter in the model of interest that we are attempting to recover.

In the LPM, β is a set of coefficients that approximates the true model and may or may not have economic meaning.

Specifically, in the LPM, β minimizes the expected squared distance from either y or $m(x)$ to $x'\beta$.

Partial effects of x

In the LCE, we see that

$$\frac{\partial \mathbb{E}[y \mid x]}{\partial x'} = \frac{\partial x'\beta}{\partial x'} = \beta;$$

the marginal effect of a unit increase in x is β.

In the LPM, we have

$$\frac{\partial \mathbb{E}[y \mid x]}{\partial x'} = \frac{\partial}{\partial x'} \left[x'\beta - \big(m(x) - x'\beta\big)\right] = \beta + \frac{\partial}{\partial x'} \left(m(x) - x'\beta\right).$$

We see that the marginal effect of x on y (or $\mathbb{E}[y \mid x]$) is *not* β. This is because the projection error may be a (non-linear) function of x.

Minimizing expected squared distance

In the LPM, we solve for β, which is a population parameter, according to

$$\begin{aligned} &\min_{\beta} \quad \mathbb{E}[(y - x'\beta)^2] \\ \Rightarrow\ & 0 = -2\mathbb{E}[x(y - x'\beta)] \\ & 0 = \mathbb{E}[xy] - \mathbb{E}[xx']\beta \\ & \beta = \left[\mathbb{E}[xx']\right]^{-1}\mathbb{E}[xy]. \end{aligned}$$

The model is said to be *identified* when this β is unique; specifically, it requires $\mathbb{E}[xx']$ to be invertible.

Alternate minimization

Note that our minimand can be rewritten

$$
\begin{aligned}
&\mathbb{E}[(y - x'\beta)^2] = \\
&\quad \mathbb{E}[(m(x) - x'\beta)^2] + 2\mathbb{E}[\epsilon(y - x'\beta)] + \mathbb{E}[\epsilon^2] \\
&\quad = \mathbb{E}[(m(x) - x'\beta)^2] + 2\mathbb{E}\left[(m(x) - x'\beta)\mathbb{E}[\epsilon \mid x']\right] + \mathbb{E}[\epsilon^2] \\
&\quad = \mathbb{E}[(m(x) - x'\beta)^2] + \mathbb{E}[\epsilon^2].
\end{aligned}
$$

Minimizing $\mathbb{E}[(m(x) - x'\beta)^2]$ would yield the same result as on the previous slide.

We can interpret the linear projection as minimizing the distance to y (as on the previous slide) or to the deterministic part of y.

$\mathbb{E}[x\tilde{\epsilon}] = 0$

Let the error of the LPM be $\tilde{\epsilon} = m(x) - x'\beta + \epsilon$. We see that

$$\begin{aligned}
\mathbb{E}[x\tilde{\epsilon}] &= \mathbb{E}\left[x(m(x) - x'\beta + \epsilon)\right] \\
&= \mathbb{E}[xm(x)] + \mathbb{E}[x\epsilon] - \mathbb{E}\left[xx'\left[\mathbb{E}[xx']\right]^{-1}\mathbb{E}[xy]\right] \\
&= \mathbb{E}[xy] - \mathbb{E}[xx']\left[\mathbb{E}[xx']\right]^{-1}\mathbb{E}[xy] \\
&= \mathbb{E}[xy] - \mathbb{E}[xy] \\
&= 0.
\end{aligned}$$

With an intercept, $\mathbb{E}[\tilde{\epsilon}] = 0$

Suppose that our x vector includes an intercept; *e.g.*, the element 1. Then, by the result of the previous slide,

$$\mathbb{E}[1 \times \tilde{\epsilon}] = \mathbb{E}[\tilde{\epsilon}] = 0.$$

Interpretation

As long as x contains a constant term, $\mathbb{E}[\tilde{\epsilon}] = 0$.

In this case, $\text{Cov}(x, \tilde{\epsilon}) = 0$.

This implies that there is no linear function of x that we could use to predict our model's error term.

We aren't able to rotate our line to get a better (linear) prediction.

Example

Suppose that $y = x^2 + \epsilon$. We use the LPM using $y = \beta_0 + \beta_1 x + \tilde{\epsilon}$.

Note: $y = x^2 + \epsilon$ *is* a linear function, but of the variable x^2. It is *not* a linear function of the variable x. Using the model $y = \beta_0 + \beta_1 x^2 + \tilde{\epsilon}$ would be a LCE model and we'd get the right answers (*i.e.*, $\beta_0 = 0$ and $\beta_1 = 1$; you should prove this yourself using the standard one predictor regression result derived below).

The first step is to calculate β_0 and β_1.

Calculating the coefficients

$$
\begin{aligned}
\beta &= \left[\mathbb{E}[xx']\right]^{-1}\mathbb{E}[xy] \\
&= \left[\mathbb{E}\begin{bmatrix} 1 & x \\ x & x^2 \end{bmatrix}\right]^{-1}\mathbb{E}\begin{bmatrix} y \\ xy \end{bmatrix} \\
&= \begin{bmatrix} 1 & \mathbb{E}[x] \\ \mathbb{E}[x] & \mathbb{E}\left[x^2\right] \end{bmatrix}^{-1}\begin{bmatrix} \mathbb{E}[y] \\ \mathbb{E}[xy] \end{bmatrix} \\
&= \frac{1}{\mathbb{E}\left[x^2\right] - \left(\mathbb{E}[x]\right)^2}\begin{bmatrix} \mathbb{E}\left[x^2\right] & -\mathbb{E}[x] \\ -\mathbb{E}[x] & 1 \end{bmatrix}\begin{bmatrix} \mathbb{E}[y] \\ \mathbb{E}[xy] \end{bmatrix} \\
&= \frac{1}{\text{Var}(x)}\begin{bmatrix} \mathbb{E}\left[x^2\right]\mathbb{E}[y] - \mathbb{E}[x]\mathbb{E}[xy] \\ \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y] \end{bmatrix} \\
&= \frac{1}{\text{Var}(x)}\begin{bmatrix} \text{Var}(x)\mathbb{E}[y] - \text{Cov}(x,y)\mathbb{E}[x] \\ \text{Cov}(x,y) \end{bmatrix}
\end{aligned}
$$

Interpretation of the coefficients

The intercept of the line is

$$\beta_0 = \mathbb{E}[y] - \beta_1 \mathbb{E}[x];$$

the intercept forces the point of expected values $(\mathbb{E}[x], \mathbb{E}[y])$ to be on the line.

The slope coefficient is

$$\beta_1 = \frac{\text{Cov}(x, y)}{\text{Var}(x)}.$$

Covariance is a measure of linear association between two variables, making it a natural component of the slope coefficient.

Economic interpretation

Suppose that x is distributed symmetrically around 0; *i.e.*, $f(x) = f(-x) \, \forall \, x$. Then, $\text{Cov}(x, y) = \text{Cov}(x, x^2) = 0$.

In this case, the LPM is $y = \mathbb{E}[y] + 0x + \tilde{\epsilon}$.

This seems to suggest that there is *no* relationship between x and y, *i.e.*,

$$\frac{d\mathbb{E}[y \mid x]}{dx} = 0.$$

This is false—there is no *linear relationship* between x and y.

Example: Dummy variable regression

Suppose that we are interested in whether small class sizes increase test scores. Let y be the score on a test and x be a dummy of whether the student was in a small class.

Under the regression

$$y = \beta_0 + \beta_1 x + \epsilon,$$

which conception applies: LCE or LPM?

Saturated models

Rewrite y using just the properties of conditional expectation:

$$\begin{aligned}\mathbb{E}[y] &= x \times \mathbb{E}[y \mid x = 1] + (1 - x) \times \mathbb{E}[y \mid x = 0] \\ &= \mathbb{E}[y \mid x = 0] + \left[\mathbb{E}[y \mid x = 1] - \mathbb{E}[y \mid x = 0]\right] x \\ &= \beta_0 + \beta_1 x;\end{aligned}$$

the LCE model is applicable here. In fact, the LCE model applies if you have only dummies in your model and you include all interactions—a "saturated" model.