CHAPTER 5

Overview of Ordinary Least Squares

5.1 GEOMETRIC THEORY

Starting with the concepts of

1. a vector space,
2. linear dependence, a basis, dimension of a vector space,
3. an inner product, length of a vector, and orthogonality,

we have developed the idea of a projection as the solution to a minimum-distance problem. The OLS problem

$$\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$$

is a minimum-distance problem in which we seek the element of the subspace $\mathrm{Col}(\mathbf{X})$ that is closest to the vector $\mathbf{y}$. The dimension of this subspace determines the uniqueness of the solution in $\boldsymbol{\beta}$. The optimal fitted values of $\mathbf{X}\boldsymbol{\beta}$, however, are always unique. They are given by

$$\hat{\boldsymbol{\mu}} = \mathbf{P}_{\mathbf{X}}\mathbf{y} = \underset{\boldsymbol{\mu}\in\mathrm{Col}(\mathbf{X})}{\mathrm{argmin}} \|\mathbf{y} - \boldsymbol{\mu}\|^2$$

where $\mathbf{P}_{\mathbf{X}}$ is the orthogonal projector onto $\mathrm{Col}(\mathbf{X})$, so that $\mathbf{y} - \hat{\boldsymbol{\mu}} \in \mathrm{Col}^{\perp}(\mathbf{X})$.

The orthogonal projector $\mathbf{P}_{\mathbf{X}}$ is a geometric concept; it is a one-to-one function of the *subspace* $\mathrm{Col}(\mathbf{X})$, not the *matrix* $\mathbf{X}$. If $\mathbf{X}$ is full-column rank, then *one* functional form for $\mathbf{P}_{\mathbf{X}}$ is

$$\mathbf{P}_{\mathbf{X}} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' \tag{5.1}$$

Thus we can interpret OLS as a two-step procedure. In the first step, one obtains the orthogonal projection $\mathbf{P}_{\mathbf{X}}\mathbf{y}$ of $\mathbf{y}$ onto $\mathrm{Col}(\mathbf{X})$. In the second step, if $\mathbf{X}$ is full rank, one decomposes this vector into the components determined by the basis in $\mathbf{X}$: $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{P_X}\mathbf{y}$. The two steps combine to yield $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$.

But if there is multicollinearity among the column vectors in $\mathbf{X}$, then given a basis for $\mathrm{Col}(\mathbf{X})$, say the column vectors of $\mathbf{X}_1$, we alter (5.1) to

$$\mathbf{P_X} = \mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1$$

No matter what basis we use, we obtain the same projector because it is unique. Indeed, we can even derive such a basis using the orthogonal projector, by recursively applying the projector to identify linearly independent vectors in $\mathrm{Col}(\mathbf{X})$. Furthermore, we can even make this basis orthonormal. Then we obtain

$$\mathbf{P_X} = \mathbf{P_R} = \mathbf{R}\mathbf{R}'$$

where the column vectors of $\mathbf{R}$ comprise the orthonormal basis.

We generalized orthogonal projection for the partitioned model, where $\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2$. We saw that

$$\hat{\boldsymbol{\mu}}_1 \equiv \mathbf{X}_1\hat{\boldsymbol{\beta}}_1 = \mathbf{X}_1\left(\mathbf{X}_{1\perp 2}'\mathbf{X}_1\right)^{-1}\mathbf{X}_{1\perp 2}'\mathbf{y}$$

where

$$\mathbf{X}_{1\perp 2} \equiv \left(\mathbf{I} - \mathbf{P}_{\mathbf{X}_2}\right)\mathbf{X}_1$$
$$\mathbf{P}_{\mathbf{X}_2} \equiv \mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'$$

The projector

$$\mathbf{P}_{12} = \mathbf{X}_1\left(\mathbf{X}_{1\perp 2}'\mathbf{X}_1\right)^{-1}\mathbf{X}_{1\perp 2}'$$

preserves elements of $\mathrm{Col}(\mathbf{X}_1)$ and annihilates $\mathrm{Col}^{\perp}(\mathbf{X}_{1\perp 2}) = \mathrm{Col}(\mathbf{X}_2) \oplus \mathrm{Col}^{\perp}(\mathbf{X})$, thereby isolating $\hat{\boldsymbol{\mu}}_1$. We can also write

$$\hat{\boldsymbol{\mu}}_1 = \mathbf{X}_1\left(\mathbf{X}_{1\perp 2}'\mathbf{X}_1\right)^{-1}\mathbf{X}_{1\perp 2}'\hat{\boldsymbol{\mu}}$$

because $\mathbf{y} - \hat{\boldsymbol{\mu}} \in \mathrm{Col}^{\perp}(\mathbf{X})$ so that $\mathbf{P}_{12}\left(\mathbf{y} - \hat{\boldsymbol{\mu}}\right) = 0$. In this case, the annihilation of $\hat{\boldsymbol{\mu}}_2$ corresponds to a movement onto $\mathrm{Col}(\mathbf{X}_1)$ along $\mathrm{Col}(\mathbf{X}_2)$. A general form for the projector onto $\mathrm{Col}(\mathbf{X})$ along $\mathrm{Col}^{\perp}(\mathbf{Z})$, denoted $\mathbf{P}_{\mathbf{X}\perp Z}$, is

$$\mathbf{P}_{\mathbf{X}\perp Z} = \mathbf{X}\left(\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{Z}' \tag{5.2}$$

if $\mathbf{Z}'\mathbf{X}$ is nonsingular. The orthogonal projector $\mathbf{P_X} \equiv \mathbf{P}_{\mathbf{X}\perp\mathbf{X}}$ is a special case.

Such projectors also arise in the restricted least-squares problem:

$$\hat{\boldsymbol{\beta}}_{\mathrm{R}} \equiv \underset{\boldsymbol{\beta}=\mathbf{S}\boldsymbol{\gamma}+\mathbf{s}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = \mathbf{P}_{\mathbf{S}\perp\mathbf{X}'\mathbf{X}\mathbf{S}}\left(\hat{\boldsymbol{\beta}} - \mathbf{s}\right) + \mathbf{s}$$

In this case, $\mathbf{XS}$ must be full rank. The general projector provides the unique solution to the generalized minimum-distance problem

$$\hat{\boldsymbol{\beta}}_{\mathrm{R}} = \underset{\boldsymbol{\beta}\in\mathrm{Col}(\mathbf{S})+\mathbf{s}}{\operatorname{argmin}} \left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|_{\mathbf{X}'\mathbf{X}}^2$$

where

$$\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|^2_{\mathbf{X}'\mathbf{X}} \equiv \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)' \mathbf{X}'\mathbf{X}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)$$

This is one example of a general solution that we can write as

$$\mathbf{P}_{\mathbf{X}\perp\mathbf{A}\mathbf{X}}\mathbf{y} = \underset{\boldsymbol{\mu}\in\text{Col}(\mathbf{X})}{\text{argmin}} \, \|\mathbf{y} - \boldsymbol{\mu}\|^2_{\mathbf{A}}$$

for any nonsingular matrix $\mathbf{A}$ such that $\mathbf{z}'\mathbf{A}\mathbf{z} > 0$ for all $\mathbf{z} \neq \mathbf{0}$, $\mathbf{z} \in \mathbb{R}^N$.

We will encounter the projector (5.2) in several new ways in later parts of this book. As we noted, if a generalization of $\mathbb{E}^N$ is constructed, then such projectors are the orthogonal projectors in that space. In the next part of this book, we will introduce yet another vector space, one consisting of vectors that are random variables, and projections in that vector space.

5.2 ECONOMETRIC SPECIFICATIONS

We also introduced, through our examples, several common, useful specifications for linear models.

1. **Indicators:** Indicator, or "dummy," RHS variables capture such discrete characteristics as the gender, race, or union status of an earner. We also used indicator variables to fit monthly seasonal variations in the national unemployment rate.

2. **Polynomial RHS variables:** Although the RHS function is linear in the coefficients and variables in $\mathbf{X}$, the RHS need not be linear in such a variable as experience in the labor force. One of the simplest ways to introduce nonlinearity is to include polynomial functions of such variables as RHS variables. For example, economists frequently include the square of experience as an RHS variable for the study of earnings. So-called "interactions" also introduce nonlinearity.

   Interactions with indicator variables also provide a method to provide differences in the RHS function for subsamples. We interacted (multiplied) an indicator for salaried earners with all of the RHS variables of the earnings function to permit changes in the coefficients for salaried earners and hourly-wage earners.

3. **Lagged dependent LHS variables:** In the study of such time series as the unemployment rate, so-called "lagged" values of the LHS variable serve as RHS variables to capture dynamics. Such specifications comprise a complex set of functions and we return to them in Chapter 20.

4. **Transformed LHS variables:** Just as one is not restricted to linear functions on the RHS, one can transform the LHS variable to obtain a new, nonlinear relationship with the RHS variables. Economists usually transform earnings with the natural logarithmic function. There are several reasons for this, and one is that the fitted coefficients can be interpreted as elasticities.

5.3 ECONOMETRIC METHOD

In this part, we have illustrated several informal uses for OLS fitted equations.

1. **Decomposition of variation:** Each OLS fitted coefficient provides a measure of the change in the LHS variable as the RHS variable changes among the observations, supposing that the values of the other RHS variables do not change. Generally, of course, the other RHS variables do change over a sequence of observations in a data set. In this sense, OLS offers a method of decomposing the overall changes in the LHS variable as several RHS variables change simultaneously from observation to observation.

   For example, we noted that men and women obtain different average levels of education and we used the OLS fit of log-earnings on an indicator for gender and a schooling variable (among other RHS variables) to describe the change in wages with schooling separately from changes in wages with gender.

2. **Exploring conjectures:** Besides summarizing patterns among observations, one may want to compare those patterns with conjectures about what one would find in the sample. Does the return in additional earnings to experience fall over a profile of earners of various ages, as some theories predict? Do unions raise earnings or do the characteristics of union members account for wage differentials?

3. **Forecasting:** OLS can be used as a naive forecasting tool.

All of these uses were informal in the sense that we left the goals of the data analysis vague and we gave the motivation for using OLS as convenience and intelligibility. If we have a more refined purpose, then we will want our method of analysis to serve that purpose. Any attempt to choose our method leads inevitably to making assumptions about the data we observe and the relationship of the data to our purpose. What does an OLS fit imply about gender discrimination? We begin to present formalizations of purposes and assumptions in the next part of this book.