LEAST-SQUARES ESTIMATORS IN LINEAR MODELS

DAVAR KHOSHNEVISAN

1. The General Linear Model

Let \( Y \) be the response variable and \( X_1, \ldots, X_m \) be the explanatory variables. The following is the linear model of interest to us:

\[
Y = \beta_1 X_1 + \cdots + \beta_m X_m + \varepsilon,
\]

where \( \beta_1, \ldots, \beta_m \) are unknown parameters, and \( \varepsilon \) is “noise.”

Now we take a sample \( Y_1, \ldots, Y_n \). The linear model becomes

\[
Y_i = \beta_1 X_{i1} + \cdots + \beta_m X_{im} + \varepsilon_i \quad i = 1, \ldots, n.
\]

Define

\[
X = \begin{pmatrix} X_{11} & \cdots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{nm} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}.
\]

Note that

\[
X\beta = \begin{pmatrix} \beta_1 X_{11} + \cdots + \beta_m X_{1m} \\ \vdots \\ \beta_m X_{n1} + \cdots + \beta_m X_{nm} \end{pmatrix}.
\]

Therefore, the linear model (2) can be written more neatly as

\[
Y = X\beta + \varepsilon,
\]

where \( \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)' \). To be sure, \( Y \) is \( n \times 1 \), \( X \) is \( n \times m \), \( \beta \) is \( m \times 1 \), and \( \varepsilon \) is \( n \times 1 \).

The matrix \( X \) is treated as if it were non-random; it is called the “design matrix” or the “regression matrix.”

2. Least Squares

Let \( \theta = X\beta \), and minimize, over all \( \beta \), the following quantity:

\[
\| Y - \theta \|^2 = (Y - \theta)'(Y - \theta) = \varepsilon'\varepsilon = \sum_{i=1}^n \varepsilon_i^2.
\]

Note that

\[
\theta = \beta_1 X_1 + \cdots + \beta_m X_m,
\]

Date: August 30, 2004.
Figure 1. The projection $\hat{\theta}$ of $Y$ onto the subspace $\mathcal{C}(X)$

where $X_i$ denotes the $i$th column of the matrix $X$. That is, $\theta \in \mathcal{C}(X)$—the
column space of $X$. So our problem has become: Minimize $\|Y - \theta\|$ over all $\theta \in \mathcal{C}(X)$.

A look at Figure 1 will convince you that the closest point $\hat{\theta} \in \mathcal{C}(X)$
is the projection of $Y$ onto the subspace $\mathcal{C}(X)$. To find a formula for this
projection we first work more generally.

3. Some Geometry

Let $S$ be a subspace of $\mathbb{R}^n$. Recall that this means that:

(1) If $x, y \in S$ and $\alpha, \beta \in \mathbb{R}$, then $\alpha x + \beta y \in S$; and
(2) $0 \in S$.

Suppose $v_1, \ldots, v_k$ forms a basis for $S$; that is, any $x \in S$ can be represented
as a linear combination of the $v_i$'s. Define $V$ to be the matrix whose $i$th
column is $v_i$; that is,

$$V = [v_1, \ldots, v_k].$$

Then any $x \in \mathbb{R}^n$ is orthogonal to $S$ if and only if $x$ is orthogonal to every $v_i$;
that is, $x'v_i = 0$. Equivalently, $x$ is orthogonal to $S$ if and only if $x'V = 0$.

In summary,

$$x \perp S \iff x'V = 0.$$  

Now the question is: If $x \in \mathbb{R}^n$ then how can we find its projection $u$
onto $S$? Consider Figure 3. From this it follows that $u$ has two properties.

(1) First of all, $u$ is perpendicular $S$, so that $(u - x)'V = 0$. Equiva-

ently, $u'V = x'V$. 

(2) Secondly, \( u \in S \), so there exist \( \alpha_1, \ldots, \alpha_k \in \mathbb{R}^k \) such that \( u = \alpha_1 v_1 + \cdots + \alpha_k v_k \). Equivalently,

\[
(\text{10}) \quad u = V\alpha.
\]

Plug (2) into (1) to find that \( \alpha'V'V = x'V \). Therefore, if \( V'V \) is invertible, then \( \alpha' = (x'V)(V'V)^{-1} \). Equivalently,

\[
(\text{11}) \quad u = V(V'V)^{-1}Vx.
\]

Define

\[
(\text{12}) \quad P_S = V(V'V)^{-1}V'.
\]

Then, \( u = P_Sx \) is the projection of \( x \) onto the subspace \( S \).

4. **Application to Linear Models**

Let \( S = \mathcal{C}(X) \) be the subspace spanned by the columns of \( X \)—this is the *column space* of \( X \). Then, provided that \( X'X \) is invertible,

\[
(\text{13}) \quad P_{\mathcal{C}(X)} = X(X'X)^{-1}X'.
\]

Therefore, the LSE \( \hat{\theta} \) of \( \theta \) is given by

\[
(\text{14}) \quad \hat{\theta} = P_{\mathcal{C}(X)}Y = X(X'X)^{-1}X'Y.
\]

This is equal to \( X'\hat{\beta} \). So \( X'X\hat{\beta} = X'Y \). Equivalently,

\[
(\text{15}) \quad \hat{\beta} = (X'X)^{-1}X'Y.
\]