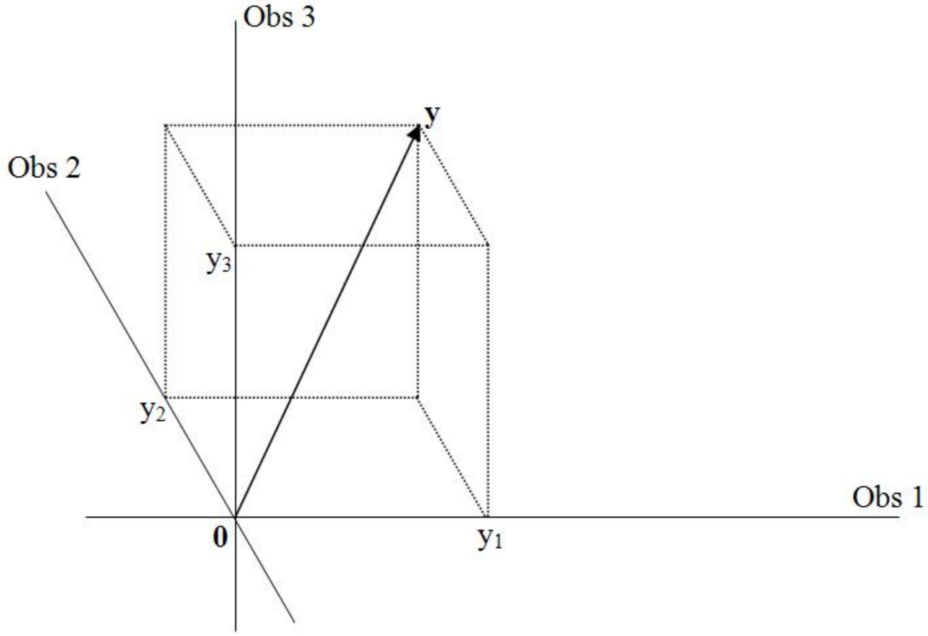
Lecture 2

Matthew Osborne

7 September 2006

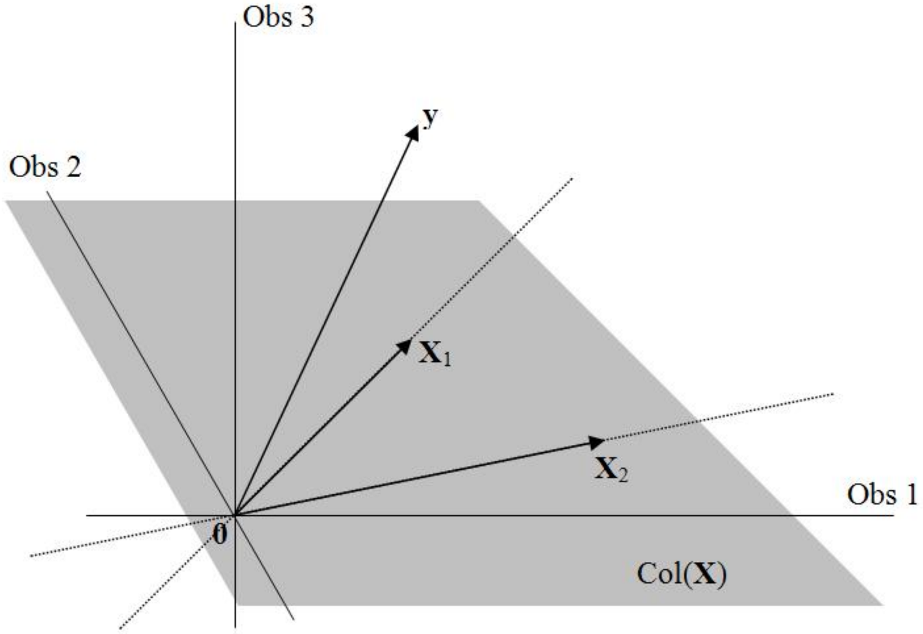
The Geometry of OLS

- **y**: N dimensional vector.



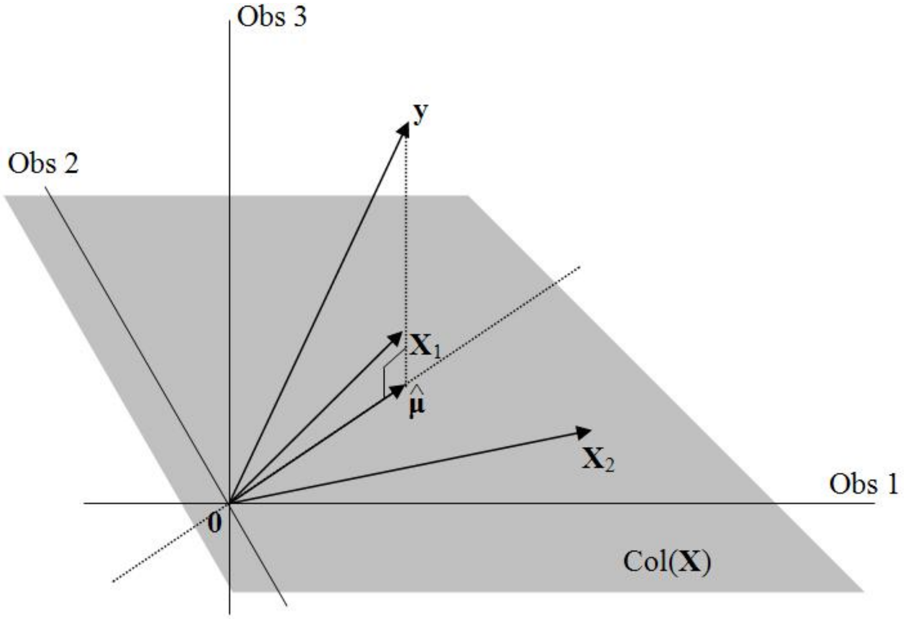
The Geometry of OLS

- Each column of $\mathbf{X}$ is also an N vector.
- Suppose $\mathbf{X} = [\mathbf{X}_1 \ \mathbf{X}_2]$.



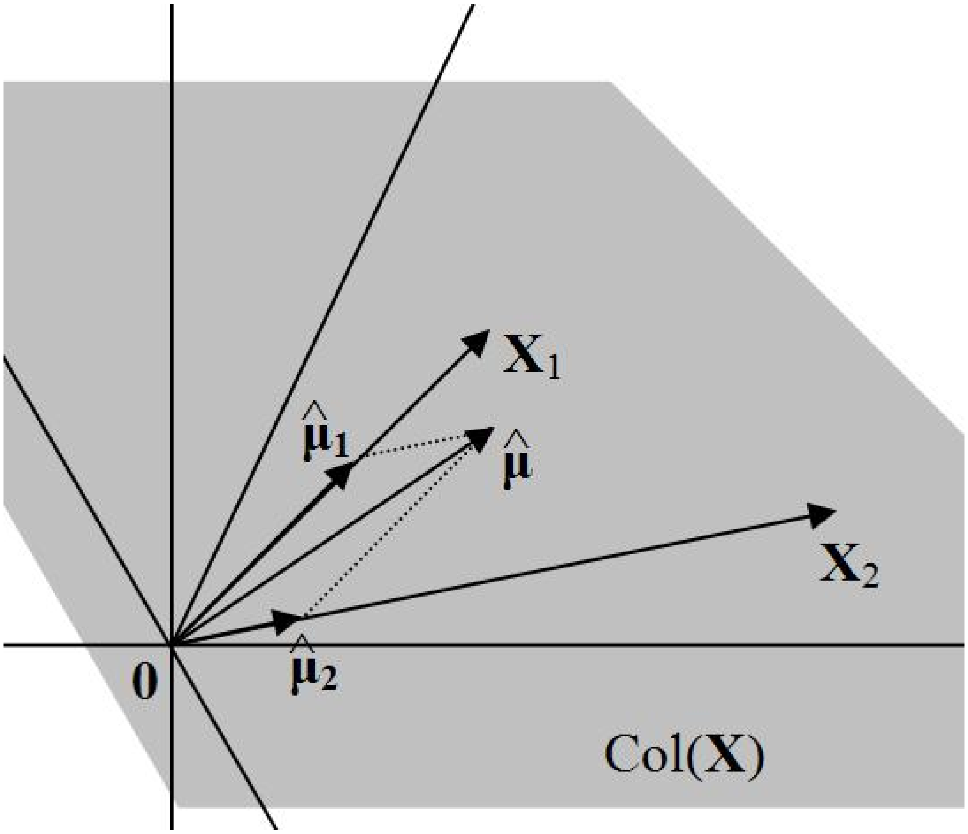
The Geometry of OLS

- $\hat{\boldsymbol{\mu}}$: orthogonal projection of $\mathbf{y}$ onto Cols($\mathbf{X}$).



The Geometry of OLS

- Coefficients: decompose $\hat{\boldsymbol{\mu}}$.
- $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}_1 + \hat{\boldsymbol{\mu}}_2 = \mathbf{X}_1\hat{\boldsymbol{\beta}}_1 + \mathbf{X}_2\hat{\boldsymbol{\beta}}_2$.



Two Steps to Doing OLS

$$\sum_{n=1}^{N}(y_n - \mathbf{x}_n'\boldsymbol{\beta})^2 \equiv ||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

$$\hat{\boldsymbol{\beta}} \equiv \underset{\boldsymbol{\beta}}{\text{argmin}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

1. Minimize $\hat{\boldsymbol{\mu}} \equiv \text{argmin}_{\boldsymbol{\mu} \in Col(\mathbf{X})} ||\mathbf{y} - \boldsymbol{\mu}||^2$.
2. Solve $\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}}$.

Proposition 1

Let $\hat{\boldsymbol{\beta}}$ be any solution to

$$\hat{\boldsymbol{\beta}} \equiv \underset{\boldsymbol{\beta}}{\text{argmin}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{1}$$

and let $\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}}$.

1. The vector of fitted values $\hat{\boldsymbol{\mu}}$ is the unique orthogonal projection of **y** onto Col(**X**).

2. The vector of fitted residuals **y** - $\hat{\boldsymbol{\mu}}$ is orthogonal to Col(**X**).

3. If dim[Col(**X**)]=K, then (1) has the unique solution

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\mu}}$$

Non Unique $\hat{\boldsymbol{\beta}}$

- Take our $K = 2$ examples, and make $\mathbf{X}_3 = \mathbf{X}_1 - 2\mathbf{X}_2$ the third column of $\mathbf{X}$.

- $\hat{\boldsymbol{\mu}}$ won't change.

$$\begin{aligned}
\hat{\boldsymbol{\mu}} &= \mathbf{X}_1\widehat{\beta}_1 + \mathbf{X}_2\widehat{\beta}_2 \\
&= \mathbf{X}_1\widehat{\beta}_1 + \frac{1}{2}(\mathbf{X}_1 - \mathbf{X}_3)\widehat{\beta}_2 \\
&= \mathbf{X}_1\left(\widehat{\beta}_1 + \frac{1}{2}\widehat{\beta}_2\right) + \mathbf{X}_3(-\frac{1}{2}\widehat{\beta}_2) \\
&= \mathbf{X}_1\widetilde{\beta}_1 + \mathbf{X}_3\widetilde{\beta}_3
\end{aligned}$$

- Infinitely many possible $\hat{\boldsymbol{\beta}}$'s.

Orthogonal Projection

- $\hat{\boldsymbol{\mu}}$ is an orthogonal projection of **y**.

- When cols of **X** are l.i., $\hat{\boldsymbol{\mu}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

- Orthogonal: $\mathbf{z}_1'\mathbf{z}_2 = 0$.

- Pythagorean Theorem: If $z_1 \perp z_2$ then

$$||z_1 + z_2||^2 = ||z_1||^2 + ||z_2||^2.$$

- Suppose that there is some $\hat{\boldsymbol{\mu}} \in \text{Col}(\mathbf{X})$ such that $\mathbf{X}$'($\mathbf{y}$ - $\hat{\boldsymbol{\mu}}$) = $\mathbf{0}$.

- Then for any other $\boldsymbol{\mu} \in \text{Col}(\mathbf{X})$, $\boldsymbol{\mu}'(\mathbf{y}-\hat{\boldsymbol{\mu}}) = 0$.

- Thus, $(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})'(\mathbf{y} - \hat{\boldsymbol{\mu}}) = 0$

$$\begin{aligned} ||\mathbf{y} - \boldsymbol{\mu}||^2 &= ||\mathbf{y} - \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}||^2 \\ &= ||\mathbf{y} - \hat{\boldsymbol{\mu}}||^2 + ||\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}||^2 \\ &\geq ||\mathbf{y} - \hat{\boldsymbol{\mu}}||^2 \end{aligned}$$

- Thus, $\hat{\boldsymbol{\mu}}$ is a sol'n to the minimum distance problem.

- $\hat{\boldsymbol{\mu}}$ is unique, since for any other sol'n $\tilde{\mu}$, $||\mathbf{y} - \tilde{\mu}||^2 = ||\mathbf{y} - \hat{\boldsymbol{\mu}}||^2$, and $||\tilde{\mu} - \hat{\boldsymbol{\mu}}||^2 = 0$.

- Sol'n to the minimization problem satisfies $\mathbf{X}$'($\mathbf{y}$ - $\hat{\boldsymbol{\mu}}$) $= \mathbf{0}$.

- So $\mathbf{X}$'($\mathbf{y}$ - $\mathbf{X}\hat{\boldsymbol{\beta}}$) $= \mathbf{0}$.

- $\mathbf{X}$'$\mathbf{X}\hat{\boldsymbol{\beta}}$- $\mathbf{X}$'$\mathbf{y}$ $= \mathbf{0}$.

- If $\mathbf{X}$'$\mathbf{X}$ is nonsingular then $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

- Can show that dim[Col($\mathbf{X}$)] $\equiv$ rank($\mathbf{X}$) $=$ K iff $\mathbf{X}$'$\mathbf{X}$ is nonsingular.

- Thus, $\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

The Projection Theorem

Let **y** $\in \mathbb{R}^N$ and let $\mathbb{S} \subseteq \mathbb{R}^N$ be a linear subspace. Then $\hat{\boldsymbol{\mu}} \in \mathbb{S}$ is a solution to the program

$$\min_{\boldsymbol{\mu} \in \mathbb{S}} ||\mathbf{y} - \boldsymbol{\mu}||^2$$

iff **y** - $\hat{\boldsymbol{\mu}} \perp \mathbb{S}$. Furthermore, $\hat{\boldsymbol{\mu}}$ is the unique solution and it exists.

Proof:

- ($\Leftarrow$) We showed this already using P.T.

Proof of Projection Theorem

- ($\Rightarrow$) Suppose bwoc that $\hat{\boldsymbol{\mu}}$ is a sol'n, but $\boldsymbol{\delta}'(\mathbf{y} - \hat{\boldsymbol{\mu}}) \neq 0$ for some $\boldsymbol{\delta} \in \mathbb{S}$.

- Then set

$$\boldsymbol{\mu} = \hat{\boldsymbol{\mu}} + \frac{\boldsymbol{\delta}'(\mathbf{y} - \hat{\boldsymbol{\mu}})}{\boldsymbol{\delta}'\boldsymbol{\delta}}\boldsymbol{\delta} \in \mathbb{S}$$

- Can show that $(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})'(\mathbf{y} - \boldsymbol{\mu}) = 0$.

- P.T.: $||\mathbf{y} - \hat{\boldsymbol{\mu}}||^2 = ||\mathbf{y} - \boldsymbol{\mu}||^2 + ||\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}||^2 > ||\mathbf{y} - \boldsymbol{\mu}||^2$. Contradiction!

Proof of Projection Theorem

- Uniqueness: we proved this already (P.T.).

- Existence: If $\mathbf{y} \in \mathbb{S}$ then $\hat{\boldsymbol{\mu}} = \mathbf{y}$.

- Suppose $\mathbf{y} \notin \mathbb{S}$. Let

$$\mathbb{B} = \{\boldsymbol{\mu} \in \mathrm{Col}(\mathbf{X}) \mid \quad ||\mathbf{y} - \boldsymbol{\mu}||^2 \leq ||\mathbf{y}||^2\}.$$

- $\mathbb{B} \neq \emptyset$ since $\mathbf{0} \in \mathbb{B}$.

- $||\mathbf{y}||^2 \geq ||\mathbf{y} - \hat{\boldsymbol{\mu}}||^2$ (P.T.), so $\hat{\boldsymbol{\mu}} \in \mathbb{B}$, if it exists.

Proof of Projection Theorem

- $\mathbb{B}$ is closed and bounded.
- $||\mathbf{y} - \boldsymbol{\mu}||^2$ is a cts function of $\boldsymbol{\mu}$.
- $||\mathbf{y} - \boldsymbol{\mu}||^2$ has a minimum on $\mathbb{B}$,so $\hat{\boldsymbol{\mu}}$ exists.

Orthogonal Projectors

- $\hat{\mu}$ is a linear transformation of $\mathbf{y}$: $\hat{\mu} = \mathbf{Py}$.

- If Col($\mathbf{X}$) are l.i., $\mathbf{P}_{\mathrm{X}} \equiv \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

- If $\mathrm{z} \in$ Col($\mathbf{X}$), $\mathbf{P}_{\mathrm{X}}\mathbf{z} = \mathbf{z}$.

- If $\mathrm{z} \perp$ Col($\mathbf{X}$), $\mathbf{P}_{\mathrm{X}}\mathbf{z} = \mathbf{0}$.

- Lemma: For every $\mathbf{z} \in \mathbb{R}^N$, we can decompose $\mathbf{z}$ uniquely into $\mathbf{z}_1 +$ $\mathbf{z}_2$ where $\mathbf{z}_1 \in$ Col($\mathbf{X}$) and $\mathbf{z}_2 \in$ Col$^{\perp}$($\mathbf{X}$).

Orthogonal Projection

- Def: Let $\mathbb{S}$ be a K-dimensional linear subspace of $\mathbb{R}^N$ so that for every $\mathbf{z} \in \mathbb{R}^N$ there is a unique $\mathbf{z}_1 \in \mathbb{S}$ and a unique $\mathbf{z}_2 \in \mathbb{S}^{\perp}$ such that $\mathbf{z} = \mathbf{z}_1 + \mathbf{z}_2$. Then the mapping of $\mathbb{R}^N$ into $\mathbb{S}$ that associates each $\mathbf{z}$ with its corresponding $\mathbf{z}_1$ is an orthogonal projection.

- If $\mathbb{S} = \mathrm{Col}(\mathbf{X})$, $\mathbf{P}_{\mathrm{X}}\mathbf{z} = \mathbf{z}_1$.

- Lemma: orthogonal projection from $\mathbb{R}^N$ onto $\mathbb{S}$ is a linear transformation.

- For $\mathbf{w}$, $\mathbf{z} \in \mathbb{R}^N$, take unique decomposition.

- $a\mathbf{w} + b\mathbf{z} = (a\mathbf{w}_1 + b\mathbf{z}_2) + (a\mathbf{w}_1 + b\mathbf{z}_2)$

Proposition 1 Revisited

- Lemma: If **P** is an orthogonal projector onto a subspace $\mathbb{S}$ of $\mathbb{R}^N$, **P** is unique.
- Part 1: consequence of projection theorem and Def of orthogonal projection.
- Part 2: consequence of projection theorem.
- Part 3: already done.