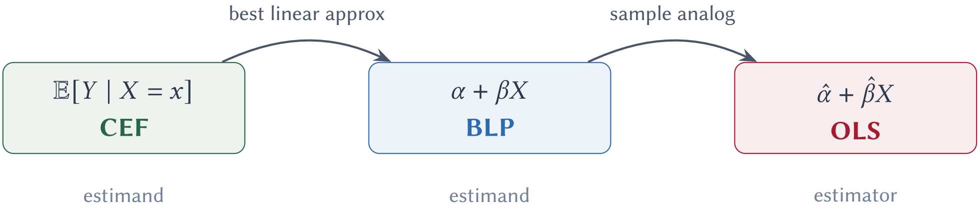
Multiple Regression

What covariates do to the BLP

Gov 2001 · Scott Cunningham · Spring 2026

Everything lives at one of three levels — keep them straight



Lecture 10a: we built this for one covariate. Today: extend to k covariates, ask what changes.

Presidential approval depends on both unemployment and inflation — simultaneously

Simple BLP

Approval $= \delta_0 + \delta_1 \cdot$ Unemployment

- δ_1: unemployment slope

↪ Inflation varies freely

↪ Effects are entangled

Multivariate BLP

Approval $= \beta_0 + \beta_1 \cdot$ Unemp $+ \beta_2 \cdot$ Infl

- β_1: unemployment effect, inflation fixed
- β_2: inflation effect, unemployment fixed

↪ Each holds the other fixed

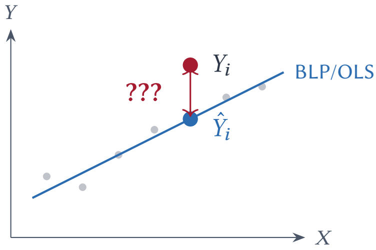
β_1 is a feature of the joint distribution of (Approval, Unemployment, Inflation) — no data required

Three questions for today — all at the population level

1. **What is the multivariate BLP?**
   Extend the population estimand from one X to k covariates
2. **What does each coefficient mean?**
   The partial derivative interpretation — and why it is not a causal claim
3. **What happens when you add or drop a covariate?**
   The OVB formula — pure algebra about two population objects

Estimation comes in Lecture 11 — today is still the estimand

This gap between observed and predicted — what do we call it?



In the sample: always **residual**

$e_i = Y_i - \hat{Y}_i$ universal across textbooks

What if we had all the data?

Frustratingly, too much creativity:

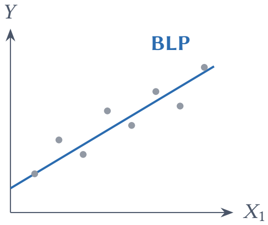
- *prediction error*
- *projection error*
- *error*
- *population residual*

All the same thing. Textbooks vary because not all insist on the population/sample distinction.

We say: prediction error

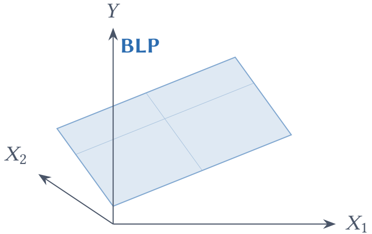
Adding a covariate extends the BLP from a line to a plane

One covariate: a line



$\alpha + \beta_1 X_1$ (a line)

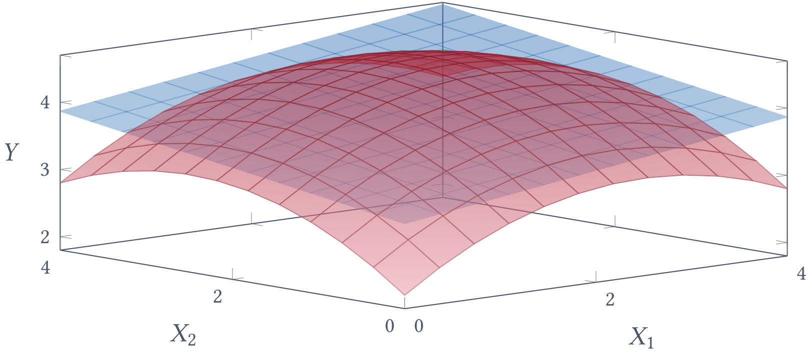
Two covariates: a plane



$\beta_0 + \beta_1 X_1 + \beta_2 X_2$ (a plane)

With k covariates: a hyperplane in $(k + 1)$ dimensions — same logic, impossible to draw

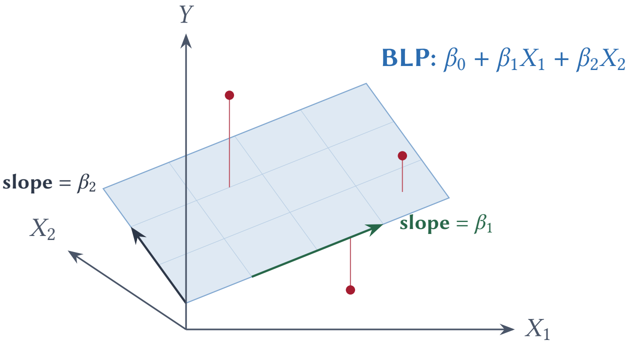
The BLP is the flat plane that best approximates the curved CEF



CEF: $\mathbb{E}[Y \mid X_1, X_2]$ — curved surface

BLP: $\beta_0 + \beta_1 X_1 + \beta_2 X_2$ — flat plane

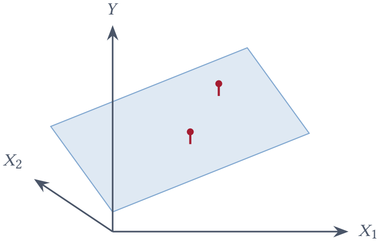
The BLP plane: X_1, X_2, and Y each get their own axis



Red lines: prediction errors — vertical distances from Y to the plane. BLP minimizes their squared average.

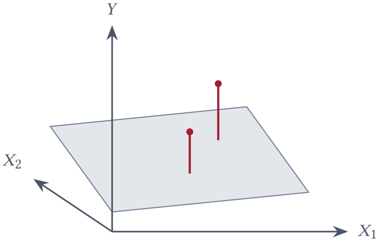
The BLP is the unique plane that minimizes squared prediction errors

BLP plane



$$\sum_i \hat{e}_i^2 = 0.13$$

A different plane (β_1 too small)



$$\sum_i \hat{e}_i^2 = 2.05$$

Same two data points, same β_0 and β_2 — only β_1 differs.

Choose β_0, β_1, β_2 to minimize expected squared prediction error

$$\min_{\beta_0, \beta_1, \beta_2} \mathbb{E}\left[\left(Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2\right)^2\right]$$

- Three free parameters: β_0, β_1, β_2
- Minimum $\Rightarrow$ set $\frac{\partial}{\partial \beta_j} = 0$ for each j
- Three equations, three unknowns

Differentiate with respect to β_0, set to zero: FOC (1)

Chain rule:

$$\frac{\partial}{\partial \beta_0} \mathbb{E}\Big[(Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2)^2\Big] = -2\, \mathbb{E}\Big[Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2\Big]$$

Set to zero, take expectations:

$$0 = \mathbb{E}[Y] - \beta_0 - \beta_1\, \mathbb{E}[X_1] - \beta_2\, \mathbb{E}[X_2]$$

$$\boxed{\mathbb{E}[Y] = \beta_0 + \beta_1\, \mathbb{E}[X_1] + \beta_2\, \mathbb{E}[X_2]}$$

Differentiate with respect to β_1, set to zero: FOC (2)

Chain rule:

$$\frac{\partial}{\partial \beta_1} \mathbb{E}\Big[(Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2)^2\Big] = -2\, \mathbb{E}\Big[(Y - \beta_0 - \beta_1 X_1 - \beta_2 X_2) \cdot X_1\Big]$$

Set to zero, distribute $\mathbb{E}[\,\cdot\, X_1]$, rearrange:

$$0 = \mathbb{E}[YX_1] - \beta_0\, \mathbb{E}[X_1] - \beta_1\, \mathbb{E}[X_1^2] - \beta_2\, \mathbb{E}[X_1 X_2]$$

$$\boxed{\mathbb{E}[YX_1] = \beta_0\, \mathbb{E}[X_1] + \beta_1\, \mathbb{E}[X_1^2] + \beta_2\, \mathbb{E}[X_1 X_2]}$$

Differentiate with respect to β_2, set to zero: FOC (3)

Chain rule:

$$\frac{\partial}{\partial \beta_2}\ \mathbb{E}\Big[(Y-\beta_0-\beta_1X_1-\beta_2X_2)^2\Big] \;=\; -2\ \mathbb{E}\Big[(Y-\beta_0-\beta_1X_1-\beta_2X_2)\cdot X_2\Big]$$

Set to zero, distribute $\mathbb{E}[\,\cdot\,X_2]$, rearrange:

$$0 = \mathbb{E}[YX_2] - \beta_0\,\mathbb{E}[X_2] - \beta_1\,\mathbb{E}[X_1X_2] - \beta_2\,\mathbb{E}[X_2^2]$$

$$\boxed{\mathbb{E}[YX_2] = \beta_0\,\mathbb{E}[X_2] + \beta_1\,\mathbb{E}[X_1X_2] + \beta_2\,\mathbb{E}[X_2^2]}$$

Three FOCs

Differentiating with respect to each β_j and setting to zero:

$$(1)\quad \mathbb{E}[Y] = \beta_0 + \beta_1\,\mathbb{E}[X_1] + \beta_2\,\mathbb{E}[X_2]$$

$$(2)\quad \mathbb{E}[YX_1] = \beta_0\,\mathbb{E}[X_1] + \beta_1\,\mathbb{E}[X_1^2] + \beta_2\,\mathbb{E}[X_1X_2]$$

$$(3)\quad \mathbb{E}[YX_2] = \beta_0\,\mathbb{E}[X_2] + \beta_1\,\mathbb{E}[X_1X_2] + \beta_2\,\mathbb{E}[X_2^2]$$

Equation (1) gives β_0 directly — substitute into (2) and (3)

Substituting β_0 into FOC (2) reveals the covariance structure

From (1): $\beta_0 = \mathbb{E}[Y] - \beta_1 \mathbb{E}[X_1] - \beta_2 \mathbb{E}[X_2]$

Substitute into FOC (2) and expand:

$$\begin{aligned}
\mathbb{E}[YX_1] &= \big(\mathbb{E}[Y] - \beta_1 \mathbb{E}[X_1] - \beta_2 \mathbb{E}[X_2]\big)\,\mathbb{E}[X_1] + \beta_1 \mathbb{E}[X_1^2] + \beta_2 \mathbb{E}[X_1 X_2] \\
&= \mathbb{E}[Y]\,\mathbb{E}[X_1] \;+\; \beta_1 \underbrace{\big(\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2\big)}_{\mathrm{Var}(X_1)} \;+\; \beta_2 \underbrace{\big(\mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\,\mathbb{E}[X_2]\big)}_{\mathrm{Cov}(X_1, X_2)}
\end{aligned}$$

Substituting β_0 into FOC (2) reveals the covariance structure

From (1): $\beta_0 = \mathbb{E}[Y] - \beta_1 \mathbb{E}[X_1] - \beta_2 \mathbb{E}[X_2]$

Substitute into FOC (2) and expand:

$$\begin{aligned}
\mathbb{E}[YX_1] &= \big(\mathbb{E}[Y] - \beta_1 \mathbb{E}[X_1] - \beta_2 \mathbb{E}[X_2]\big)\,\mathbb{E}[X_1] + \beta_1 \mathbb{E}[X_1^2] + \beta_2 \mathbb{E}[X_1 X_2] \\
&= \mathbb{E}[Y]\,\mathbb{E}[X_1] \;+\; \beta_1 \underbrace{\big(\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2\big)}_{\mathrm{Var}(X_1)} + \beta_2 \underbrace{\big(\mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\,\mathbb{E}[X_2]\big)}_{\mathrm{Cov}(X_1, X_2)}
\end{aligned}$$

Move $\mathbb{E}[Y]\,\mathbb{E}[X_1]$ to the left; recognize definitions:

$$(2') \quad \mathrm{Cov}(Y, X_1) = \beta_1 \mathrm{Var}(X_1) + \beta_2 \mathrm{Cov}(X_1, X_2)$$

$(3')$ is identical — replace X_1 with X_2 throughout:
$\mathrm{Cov}(Y, X_2) = \beta_1 \mathrm{Cov}(X_1, X_2) + \beta_2 \mathrm{Var}(X_2)$

Write equations (2′) and (3′) as a single matrix equation

$$\underbrace{\begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_1, X_2) & \mathrm{Var}(X_2) \end{pmatrix}}_{\text{covariate structure}} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \underbrace{\begin{pmatrix} \mathrm{Cov}(Y, X_1) \\ \mathrm{Cov}(Y, X_2) \end{pmatrix}}_{\text{signal}}$$

- **Covariate structure:** how X_1 and X_2 vary (and co-vary) with each other
- **Signal:** how each covariate co-moves with Y
- Unique solution exists when X_1 and X_2 are not perfectly collinear

The determinant of a 2×2 matrix: multiply the diagonal, subtract the off-diagonal

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

Simple example:

$$\det\begin{pmatrix} 3 & 1 \\ 2 & 4 \end{pmatrix} = (3)(4) - (1)(2) = 10$$

Our covariate matrix:

$$D = \det\begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_1, X_2) & \mathrm{Var}(X_2) \end{pmatrix}$$
$$= \mathrm{Var}(X_1)\,\mathrm{Var}(X_2) - \mathrm{Cov}(X_1, X_2)^2$$

Cauchy-Schwarz guarantees $\mathrm{Cov}(X_1, X_2)^2 \leq \mathrm{Var}(X_1)\,\mathrm{Var}(X_2)$, so $D \geq 0$ always — collinearity drives D to zero, never negative; $D > 0$ iff unique solution exists

Cramer's rule: each coefficient in closed form

Determinant: $D = \text{Var}(X_1)\,\text{Var}(X_2) - \text{Cov}(X_1, X_2)^2$

$$\beta_1 = \frac{\text{Cov}(Y, X_1)\ \text{Var}(X_2)\ -\ \text{Cov}(Y, X_2)\ \text{Cov}(X_1, X_2)}{D}$$

$$\beta_2 = \frac{\text{Cov}(Y, X_2)\ \text{Var}(X_1)\ -\ \text{Cov}(Y, X_1)\ \text{Cov}(X_1, X_2)}{D}$$

When $\text{Cov}(X_1, X_2) = 0$: $D = \text{Var}(X_1)\,\text{Var}(X_2)$ and $\beta_j = \text{Cov}(Y, X_j)/\text{Var}(X_j)$ — the simple bivariate slope

Vote share by district: plugging in the moments

Given moments:

Y = district vote share (%)

X_1 = avg. age (years)

X_2 = median income ($10k)

Vote share by district: plugging in the moments

Given moments:

Y = district vote share (%)

X_1 = avg. age (years)

X_2 = median income ($10k)

$$\mathbb{E}[Y] = 52,\ \mathbb{E}[X_1] = 40,\ \mathbb{E}[X_2] = 5$$
$$\text{Var}(X_1) = 4,\ \text{Var}(X_2) = 9$$
$$\text{Cov}(X_1, X_2) = 3$$
$$\text{Cov}(Y, X_1) = 10,\ \text{Cov}(Y, X_2) = 21$$

Vote share by district: plugging in the moments

Given moments:

Y = district vote share (%)

X_1 = avg. age (years)

X_2 = median income ($10k)

$$\mathbb{E}[Y] = 52,\ \mathbb{E}[X_1] = 40,\ \mathbb{E}[X_2] = 5$$
$$\text{Var}(X_1) = 4,\ \text{Var}(X_2) = 9$$
$$\text{Cov}(X_1, X_2) = 3$$
$$\text{Cov}(Y, X_1) = 10,\ \text{Cov}(Y, X_2) = 21$$

Calculate:

$$D = (4)(9) - (3)^2 = 27$$
$$\beta_1 = \frac{(10)(9) - (21)(3)}{27} = \frac{27}{27} = 1$$
$$\beta_2 = \frac{(21)(4) - (10)(3)}{27} = \frac{54}{27} = 2$$
$$\beta_0 = 52 - (1)(40) - (2)(5) = 2$$

Vote share by district: plugging in the moments

Given moments:

Y = district vote share (%)

X_1 = avg. age (years)

X_2 = median income ($10k)

$$\mathbb{E}[Y] = 52,\ \mathbb{E}[X_1] = 40,\ \mathbb{E}[X_2] = 5$$
$$\text{Var}(X_1) = 4,\ \text{Var}(X_2) = 9$$
$$\text{Cov}(X_1, X_2) = 3$$
$$\text{Cov}(Y, X_1) = 10,\ \text{Cov}(Y, X_2) = 21$$

Calculate:

$$D = (4)(9) - (3)^2 = 27$$
$$\beta_1 = \frac{(10)(9) - (21)(3)}{27} = \frac{27}{27} = 1$$
$$\beta_2 = \frac{(21)(4) - (10)(3)}{27} = \frac{54}{27} = 2$$
$$\beta_0 = 52 - (1)(40) - (2)(5) = 2$$

$$\text{VoteShare} = 2 + 1 \cdot \text{Age} + 2 \cdot \text{Income}$$

Simple bivariate slope on age: $10/4 = 2.5$ — controlling for income cuts it to 1

Each coefficient is the predicted change holding everything else fixed

$$\widehat{\text{VoteShare}} = 2 + 1 \cdot \text{Age} + 2 \cdot \text{Income}$$

- $\beta_1 = 1$: holding Income fixed, +1 yr of avg. age $\Rightarrow$ predicted vote share rises by exactly 1 pp
- $\beta_2 = 2$: holding Age fixed, +\$10k of median income $\Rightarrow$ predicted vote share rises by exactly 2 pp
- $\beta_0 = 2$: predicted vote share when Age = 0 *and* Income = 0

Each coefficient is the predicted change holding everything else fixed

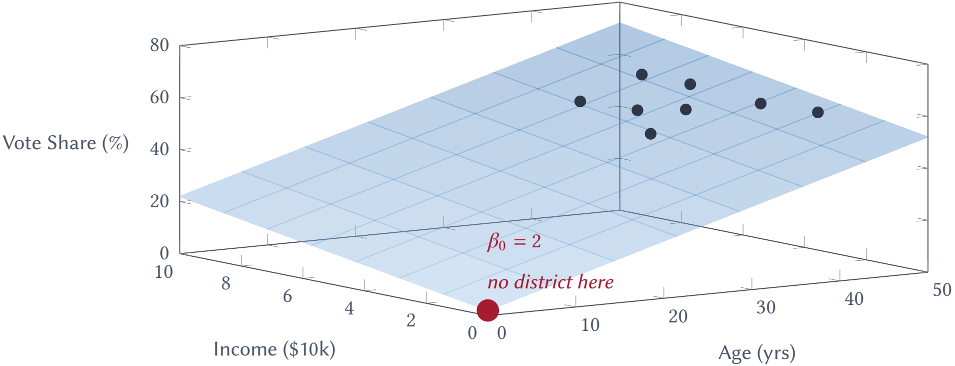
$$\widehat{\text{VoteShare}} = 2 + 1 \cdot \text{Age} + 2 \cdot \text{Income}$$

- $\beta_1 = 1$: holding Income fixed, +1 yr of avg. age $\Rightarrow$ predicted vote share rises by exactly 1 pp
- $\beta_2 = 2$: holding Age fixed, +\$10k of median income $\Rightarrow$ predicted vote share rises by exactly 2 pp
- $\beta_0 = 2$: predicted vote share when Age = 0 *and* Income = 0

Outside the support: no district has Age = 0 or Income = 0 — the BLP plane extends beyond the data and predicts there anyway

The BLP does not know where the data end. It is a plane. It predicts everywhere — correctly inside the support, freely outside it.

$\beta_0 = 2$ is a prediction for Age $= 0$, Income $= 0$ — no real district lives there



Districts: Age ≈ 34–48, Income ≈ \$30k–\$80k

BLP: $\hat{Y} = 2 + 1 \cdot \text{Age} + 2 \cdot \text{Income}$

The same logic extends to k covariates — a hyperplane in $(k+1)$ dimensions

Two covariates: a plane. The pattern generalizes to any k:

Population problem:

$$\min_{\beta_0, \beta_1, \ldots, \beta_k} \mathbb{E}\Big[\big(Y - \beta_0 - \beta_1 X_1 - \cdots - \beta_k X_k\big)^2\Big]$$

- One intercept + k slopes: $k + 1$ unknowns
- Setting $k + 1$ partial derivatives to zero: $k + 1$ equations
- Solution exists and is unique when no covariate is a linear combination of the others

k-covariate BLP: the FOC system generalizes the two-covariate solution

The same coupled structure — but now $k \times k$:

$$\beta_j = \frac{\partial}{\partial X_j} m(X_1, \ldots, X_k) \qquad \text{for each } j = 1, \ldots, k$$

$$\beta_0 = \mathbb{E}[Y] - \beta_1 \mathbb{E}[X_1] - \cdots - \beta_k \mathbb{E}[X_k]$$

- Each β_j holds all other covariates fixed — partial slope in direction j
- When covariates are correlated, all slopes are jointly determined
- The solution is a feature of the joint distribution of $(Y, X_1, \ldots, X_k)$

For $k > 2$, Cramer's rule is impractical — R solves via QR decomposition

General matrix form:

$$\underbrace{\Sigma}_{k\times k}\ \underbrace{\boldsymbol{\beta}}_{k\times 1} = \underbrace{\mathbf{c}}_{k\times 1} \implies \boldsymbol{\beta} = \Sigma^{-1}\mathbf{c}$$

Cramer's rule works for $k = 2$:

Ratio of two 2×2 determinants

For $k = 10$: compute 11 determinants of 10×10 matrices — impractical

For $k > 2$, Cramer's rule is impractical — R solves via QR decomposition

General matrix form:

$$\underbrace{\Sigma}_{k\times k} \underbrace{\boldsymbol{\beta}}_{k\times 1} = \underbrace{\mathbf{c}}_{k\times 1} \implies \boldsymbol{\beta} = \Sigma^{-1}\mathbf{c}$$

Cramer's rule works for $k = 2$:

Ratio of two 2×2 determinants

For $k = 10$: compute 11 determinants of 10×10 matrices — impractical

R's `lm()`: QR decomposition

Factors the $n \times (k+1)$ design matrix:

$\mathbf{X} = \mathbf{QR}$

$\mathbf{Q}$: orthogonal $\mathbf{R}$: upper triangular

$\hat{\boldsymbol{\beta}} = \mathbf{R}^{-1}\mathbf{Q}'\mathbf{y}$

Avoids forming $\mathbf{X}'\mathbf{X}$ directly — numerically stable

Lecture 11 bridges population to sample: Σ and $\mathbf{c}$ become sample moments, and `lm()` computes the rest

The multivariate CEF and BLP extend the same way

Multivariate CEF

Estimand: $\mathbb{E}[Y \mid X_1 = x_1, \ldots, X_k = x_k]$

Still: best predictor of Y given $X_1, \ldots, X_k$

Still: not assumed linear

Unrestricted: any function of $(X_1, \ldots, X_k)$ — could be nonlinear, interactive, discontinuous

Multivariate BLP

Estimand: $\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$

Still: best linear approximation to the CEF

Still: defined without any linearity assumption

Advantage: $k + 1$ parameters regardless of how many cells exist

The multivariate CEF and BLP extend the same way

Multivariate CEF

Estimand: $\mathbb{E}[Y \mid X_1 = x_1, \ldots, X_k = x_k]$

Still: best predictor of Y given $X_1, \ldots, X_k$

Still: not assumed linear

Unrestricted: any function of $(X_1, \ldots, X_k)$ — could be nonlinear, interactive, discontinuous

Multivariate BLP

Estimand: $\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$

Still: best linear approximation to the CEF

Still: defined without any linearity assumption

Advantage: $k + 1$ parameters regardless of how many cells exist

Adding covariates does not change what CEF and BLP *are* — it changes how many dimensions the surfaces live in

Theorem (A&M 2.3.7): each BLP coefficient is a partial derivative

In the multivariate BLP $m(X_1, \ldots, X_k) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$**:**

$$\beta_j = m(x_1, \ldots, x_j + 1, \ldots, x_k) \; - \; m(x_1, \ldots, x_j, \ldots, x_k)$$

In words: holding all other X's fixed, β_j = predicted change per unit of X_j

Theorem (A&M 2.3.7): each BLP coefficient is a partial derivative

In the multivariate BLP $m(X_1, \ldots, X_k) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$:

$$\beta_j = m(x_1, \ldots, x_j + 1, \ldots, x_k) \; - \; m(x_1, \ldots, x_j, \ldots, x_k)$$

In words: holding all other X's fixed, β_j = predicted change per unit of X_j

Bivariate BLP

$m(x) = \beta_0 + \beta_1 x$

β_1 = slope of the line

bivariate: no other X's to hold fixed

Multivariate BLP

$m(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

β_1 = slope *in the X_1 direction*

holding X_2 fixed at any value

Fix income at any level: the slope in age is always β_1 — parallel lines at different heights

$$\widehat{\text{VoteShare}} = 2 + 1 \cdot \text{Age} + 2 \cdot \text{Income}$$

Fix income at any level: the slope in age is always β_1 — parallel lines at different heights

$$\widehat{\text{VoteShare}} = 2 + 1 \cdot \text{Age} + 2 \cdot \text{Income}$$

Substitute Income = c and collect:

$$\widehat{\text{VoteShare}} = \underbrace{(2 + 2c)}_{\text{intercept shifts with } c} + \underbrace{1}_{\beta_1} \cdot \text{Age}$$

Fix income at any level: the slope in age is always β_1 — parallel lines at different heights

$$\widehat{\text{VoteShare}} = 2 + 1 \cdot \text{Age} + 2 \cdot \text{Income}$$

Substitute Income = c and collect:

$$\widehat{\text{VoteShare}} = \underbrace{(2 + 2c)}_{\text{intercept shifts with } c} + \underbrace{1}_{\beta_1} \cdot \text{Age}$$

$c = 3$ (\$30k)	$c = 5$ (\$50k)	$c = 8$ (\$80k)
$8 + 1 \cdot \text{Age}$	$12 + 1 \cdot \text{Age}$	$18 + 1 \cdot \text{Age}$

Fix income at any level: the slope in age is always β_1 — parallel lines at different heights

$$\widehat{\text{VoteShare}} = 2 + 1 \cdot \text{Age} + 2 \cdot \text{Income}$$

Substitute Income = c and collect:

$$\widehat{\text{VoteShare}} = \underbrace{(2 + 2c)}_{\text{intercept shifts with } c} + \underbrace{1}_{\beta_1} \cdot \text{Age}$$

$c = 3$ (\$30k)	$c = 5$ (\$50k)	$c = 8$ (\$80k)
$8 + 1 \cdot \text{Age}$	$12 + 1 \cdot \text{Age}$	$18 + 1 \cdot \text{Age}$

Three slices, three intercepts — same slope $\beta_1 = 1$ in every one

The next slide shows a slice like this drawn on the BLP plane

"Holding all else equal" is a mathematical statement — not a causal one

The geometry:

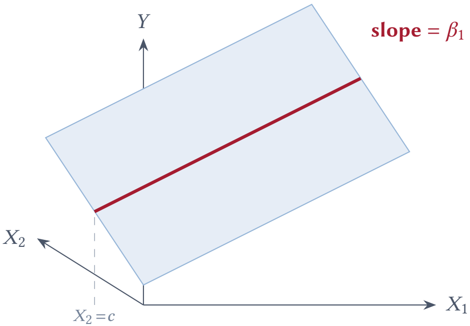
Two covariates: a plane in (X_1, X_2, Y)

β_1 = slope in the X_1 direction

β_2 = slope in the X_2 direction

Fix $X_2 = c \Rightarrow$ line with slope β_1

Same β_1 regardless of which slice c



Adding a covariate changes the estimand — not just the estimate

Bivariate BLP

$$m_S(X_1) = \delta_0 + \delta_1 X_1$$

Question: how does the BLP of Y shift with X_1?

δ_1 = slope of the best line through the (X_1, Y) scatter

X_2 free to vary

Multivariate BLP

$$m_L(X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Question: how does the BLP shift with X_1, *at fixed* X_2?

β_1 = slope in the X_1 direction of the fitted plane

X_2 fixed by construction

Adding a covariate changes the estimand — not just the estimate

Bivariate BLP

$$m_S(X_1) = \delta_0 + \delta_1 X_1$$

Question: how does the BLP of Y shift with X_1?

δ_1 = slope of the best line through the (X_1, Y) scatter

X_2 free to vary

Multivariate BLP

$$m_L(X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Question: how does the BLP shift with X_1, *at fixed* X_2?

β_1 = slope in the X_1 direction of the fitted plane

X_2 fixed by construction

$\delta_1 \neq \beta_1$ in general — they are different population objects, not an estimate and its corrected version

Multiple regression wraps up — Tuesday: FWL, OVB, and interactions

What we built today:

- Multivariate BLP: the flat plane minimizing squared prediction error
- FOC system: three equations, three unknowns, solved via Cramer's rule
- Partial slopes: β_1 holds X_2 fixed — different from the bivariate slope

Tuesday (Lecture 11a):

- **Frisch-Waugh-Lovell:** every partial slope is a residual-on-residual regression
- **OVB formula:** what omitting a variable does to the short BLP
- **Interactions and polynomials:** flexible BLPs for heterogeneous effects