

# **The Best Linear Predictor**

Gov 2001: Quantitative Social Science Methods I

Scott Cunningham

Harvard University

Spring 2026

# Today's Reading

## Required

- **Blackwell**, Ch. 5: Linear regression (pp. 99–118)
- **Aronow & Miller**, §2.2.4: BLP definition (pp. 80–88)
- **Angrist & Pischke**, §3.1.1–3.1.2: Regression and CEF

**Welcome to regression!** Everything we've learned comes together now.

# The Story So Far

**Week 4:** The CEF  $\mathbb{E}[Y|X]$  is the best predictor of  $Y$  given  $X$ .

**But:** The CEF can be any shape—nonlinear, wiggly, complicated.

**Today's question:** What if we *restrict* ourselves to linear functions?

Among all predictions of the form  $\alpha + \beta X$ , which is best?

**Answer:** The Best Linear Predictor (BLP).

This is what regression actually computes.

# Why Linear?

## Reasons to restrict to linear functions:

1. **Simplicity:** A line has only two parameters  $(\alpha, \beta)$
2. **Interpretability:** “A one-unit increase in  $X$  is associated with a  $\beta$ -unit change in  $Y$ ”
3. **Robustness:** Less prone to overfitting than flexible methods
4. **Often good enough:** If CEF is approximately linear,  $BLP \approx CEF$

Even when CEF is nonlinear, BLP gives a useful summary.

# The Best Linear Predictor: Definition

## Best Linear Predictor (BLP)

The BLP of  $Y$  given  $X$  is the linear function  $\alpha + \beta X$  that minimizes mean squared prediction error:

$$(\alpha^*, \beta^*) = \arg \min_{\alpha, \beta} \mathbb{E} [(Y - \alpha - \beta X)^2]$$

### Key distinction:

- CEF: Best predictor overall (any function)
- BLP: Best predictor *among linear functions*

If the CEF is linear, BLP = CEF. Otherwise, BLP approximates CEF.

## Finding the BLP: The Setup

**Goal:** Minimize  $\mathbb{E}[(Y - \alpha - \beta X)^2]$

**This is a calculus problem.** Take derivatives, set to zero.

**First-order conditions:**

$$\frac{\partial}{\partial \alpha} \mathbb{E}[(Y - \alpha - \beta X)^2] = 0$$

$$\frac{\partial}{\partial \beta} \mathbb{E}[(Y - \alpha - \beta X)^2] = 0$$

Let's solve these one at a time.

## Finding $\alpha$

**FOC for  $\alpha$ :**

$$\frac{\partial}{\partial \alpha} \mathbb{E}[(Y - \alpha - \beta X)^2] = -2 \mathbb{E}[Y - \alpha - \beta X] = 0$$

This gives us:

$$\mathbb{E}[Y] - \alpha - \beta \mathbb{E}[X] = 0$$

**Solving for  $\alpha$ :**

$$\alpha^* = \mathbb{E}[Y] - \beta^* \mathbb{E}[X]$$

**The intercept** ensures the line passes through  $(\mathbb{E}[X], \mathbb{E}[Y])$ .

## Finding $\beta$

**FOC for  $\beta$ :**

$$\frac{\partial}{\partial \beta} \mathbb{E}[(Y - \alpha - \beta X)^2] = -2 \mathbb{E}[X(Y - \alpha - \beta X)] = 0$$

This gives us:

$$\mathbb{E}[XY] - \alpha \mathbb{E}[X] - \beta \mathbb{E}[X^2] = 0$$

Substituting  $\alpha = \mathbb{E}[Y] - \beta \mathbb{E}[X]$ :

$$\mathbb{E}[XY] - (\mathbb{E}[Y] - \beta \mathbb{E}[X]) \mathbb{E}[X] - \beta \mathbb{E}[X^2] = 0$$

$$\mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y] - \beta(\mathbb{E}[X^2] - (\mathbb{E}[X])^2) = 0$$

$$\text{Cov}(X, Y) = \beta \text{Var}(X)$$

# The BLP Formula

## Best Linear Predictor

$$\beta^* = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

$$\alpha^* = \mathbb{E}[Y] - \beta^* \mathbb{E}[X]$$

**This is the regression coefficient!**

**The slope  $\beta^*$ :**

- Ratio of covariance to variance
- Positive if  $X$  and  $Y$  move together
- Larger when  $X$  and  $Y$  are more correlated

## Alternative Form: Using Correlation

**Recall:**  $\rho = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$

**Therefore:**

$$\beta^* = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{\rho \sigma_X \sigma_Y}{\sigma_X^2} = \rho \frac{\sigma_Y}{\sigma_X}$$

### Alternative Formula

$$\beta^* = \rho_{XY} \cdot \frac{\sigma_Y}{\sigma_X}$$

**Interpretation:** The slope is the correlation times the ratio of standard deviations.

## Example: Income and Education

### Population parameters:

- $\mathbb{E}[\text{Education}] = 13 \text{ years}$ ,  $\sigma_{\text{Ed}} = 3 \text{ years}$
- $\mathbb{E}[\text{Income}] = \$50,000$ ,  $\sigma_{\text{Inc}} = \$20,000$
- $\text{Corr}(\text{Ed}, \text{Inc}) = 0.4$

### BLP slope:

$$\beta^* = 0.4 \times \frac{20,000}{3} = \$2,667 \text{ per year of education}$$

### BLP intercept:

$$\alpha^* = 50,000 - 2,667 \times 13 = \$15,333$$

$$\text{BLP: } \mathbb{E}[\text{Income}|\text{Ed}] \approx 15,333 + 2,667 \times \text{Ed}$$

## Political Science Example: Campaign Spending

**Question:** How does campaign spending relate to vote share?

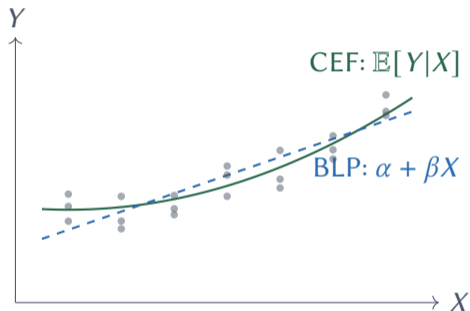
**BLP:**  $\text{Vote share} \approx \alpha^* + \beta^* \times \log(\text{Spending})$

**What  $\beta^*$  captures:**

- Best linear approximation to relationship between spending and votes
- Includes all reasons spending and votes correlate
- *Not* the causal effect of spending (confounders!)

**Key insight:** BLP is descriptive, not causal. A strong BLP relationship doesn't mean spending *causes* votes—it might be that strong candidates attract money.

## BLP vs. CEF: Visual



**The BLP is the best linear approximation to the CEF.**  
Even when CEF is curved, the line captures the overall trend.

## When BLP = CEF

**Key result:** If the CEF is linear, then BLP equals CEF.

### When is the CEF linear?

- $(X, Y)$  jointly normal
- Saturated model (dummy for each value of  $X$ )
- By assumption/modeling choice

### Angrist & Pischke's Theorem 3.1.2:

Even if the CEF is not linear, the BLP provides the **minimum MSE linear approximation** to the CEF.

The BLP is always doing something sensible.

# Properties of BLP Residuals

Define the BLP residual:

$$u = Y - \alpha^* - \beta^* X$$

## Key Properties

1.  $\mathbb{E}[u] = 0$
2.  $\text{Cov}(u, X) = 0$

**Compare to CEF residual:**  $\varepsilon = Y - \mathbb{E}[Y|X]$  has  $\mathbb{E}[\varepsilon|X] = 0$ .

BLP residual is *uncorrelated* with  $X$ ; CEF residual is *mean-independent* of  $X$ .  
Mean independence is stronger than uncorrelatedness.

## Proving $\text{Cov}(u, X) = 0$

From the FOC, we had:

$$\mathbb{E}[X(Y - \alpha - \beta X)] = 0$$

This is:

$$\mathbb{E}[Xu] = 0$$

And since  $\mathbb{E}[u] = 0$ :

$$\text{Cov}(u, X) = \mathbb{E}[Xu] - \mathbb{E}[u] \mathbb{E}[X] = 0 - 0 = 0$$

**This is a defining property of the BLP:** The residual is uncorrelated with  $X$  by construction.

# Blackwell's Perspective

## From Blackwell Ch. 5:

*“The BLP is defined as the linear function of  $X$  that minimizes the expected squared prediction error. Remarkably, this optimization problem has a closed-form solution.”*

**Key insight:** We don't need to know the full joint distribution of  $(X, Y)$ .

We only need:

- $\mathbb{E}[X]$  and  $\mathbb{E}[Y]$
- $\text{Var}(X)$  and  $\text{Cov}(X, Y)$

These are moments, not distributions. We can estimate them from data.

# From Population to Sample

## Population BLP:

$$\beta^* = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}, \quad \alpha^* = \mathbb{E}[Y] - \beta^* \mathbb{E}[X]$$

**We don't know these population quantities!**

## Sample analogs:

$$\hat{\beta} = \frac{\widehat{\text{Cov}}(X, Y)}{\widehat{\text{Var}}(X)} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$
$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}$$

**This IS ordinary least squares (OLS)!**

# The OLS Estimator

## OLS Coefficients

$$\hat{\beta} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$
$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

### The plug-in principle at work:

- Replace  $\mathbb{E}[X]$  with  $\bar{X}$
- Replace  $\text{Cov}(X, Y)$  with sample covariance
- Replace  $\text{Var}(X)$  with sample variance

By the LLN, these sample quantities converge to population quantities.

## Why is it Called “Least Squares”?

**Equivalent derivation:** Minimize the sum of squared residuals.

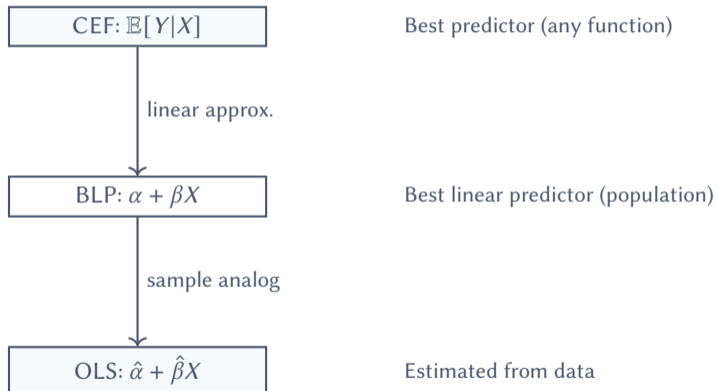
**In sample:**

$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{\alpha, \beta} \sum_{i=1}^n (Y_i - \alpha - \beta X_i)^2$$

**This gives the same answer!**

OLS minimizes squared errors in sample, which estimates the BLP that minimizes squared errors in population.

## Summary: The Big Picture



## Key Takeaways

1. **BLP** = best linear predictor = linear function minimizing MSE
2. **BLP slope:**  $\beta^* = \text{Cov}(X, Y) / \text{Var}(X)$
3. **BLP residual** is uncorrelated with  $X$
4. **If CEF is linear**,  $\text{BLP} = \text{CEF}$
5. **OLS** = sample analog of BLP
6. **OLS is consistent** for BLP by LLN

**Next:** OLS mechanics and how to interpret regression output.

# Looking Ahead

## Wednesday: OLS as Sample BLP

- The least squares derivation
- Fitted values and residuals
- Interpreting regression output
- R-squared

## Reading:

- Blackwell Ch. 5 (continue)
- A&M §4.1.1–4.1.2