Simple Linear Regression
Part A: The Best Linear Predictor

Thomas Wiemann
University of Chicago

Econometrics
Econ 21020

Updated: April 26, 2022

Introduction

In lecture 5 discussed the Random Assignment (RA) assumption:

- ▷ Showed that $E[Y|W = w] = E[g(w, U)]$ under RA;
- ▷ Derived binning estimator for ATE for randomly assigned discrete W.

We maintain RA and discuss estimation of ATEs of the form

$$\text{ATE}_{w',w} = E[g(w', U) - g(w, U)], \quad (1)$$

where $w', w \in \text{supp}\, W$. Under RA,

$$\text{ATE}_{w',w} = E[Y|W = w'] - E[Y|W = w]. \quad (2)$$

- ▷ Can only construct binning estimator when $P(W = w') > 0$ and $P(W = w) > 0$: Not suitable for, e.g., continuous W.
- ▷ Even when W discrete, we showed that the sampling variance of the binning estimator is inversely related to $P(W = w')$, $P(W = w)$: May want alternative estimator due to Bias-Variance trade-off.

Introduction (Contd.)

We're in need of an alternative estimator for the CEF $E[Y|W = w]$.

The alternative estimator we consider is *linear regression*.

- ▷ *The* estimator in empirical economics.

Linear regression is easy to compute but *very* difficult to interpret.

- ▷ Linear regression does not estimate the CEF directly!
- ▷ Linear regression estimates the *best linear approximation* of the CEF.

To make this difficult topic approachable, we take two key steps:

A. Define, analyze and discuss the best linear approximation of the CEF.

B. Derive and characterize the linear regression estimator.

Throughout, we focus on (scalar-valued) random variables.

- ▷ Turn to regression with random vectors after the midterm.

Outline

1. Best Linear Predictor

2. Properties of the BLP-Residual

3. Interpretation of the BLP-Coefficient β
    - ▷ Descriptive Interpretation using Yitzhaki (1996)
    - ▷ Causal Interpretation under Random Assignment

These notes benefit greatly from the lecture notes of Prof. Alex Torgovitsky, Prof. James Heckman, and Francesco Ruggieri.

Outline

1. **Best Linear Predictor**
2. Properties of the BLP-Residual
3. Interpretation of the BLP-Coefficient β
   - ▷ Descriptive Interpretation using Yitzhaki (1996)
   - ▷ Causal Interpretation under Random Assignment

These notes benefit greatly from the lecture notes of Prof. Alex Torgovitsky, Prof. James Heckman, and Francesco Ruggieri.

Best Linear Predictor

The best linear approximation to the CEF w.r.t. the L^2-loss is commonly referred to as the *best linear predictor*.

- ▷ See Problem 4 of Problem Set 3 why this terminology is sensible.

Definition 1 (Best Linear Predictor; BLP)

Let Y and X be random variables. The *best linear predictor* (BLP) of the conditional expectation $E[Y|X]$ is defined as

$$\text{BLP}(Y|X) = \alpha + X\beta, \tag{3}$$

where the BLP-coefficients α and β are such that

$$(\alpha, \beta) \in \underset{\alpha,\beta\in\mathbb{R}}{\arg\min}\ E\left[\left(E\left[Y|X\right] - (\alpha + X\beta)\right)^2\right]. \tag{4}$$

Importantly, the BLP is an *approximation* to the CEF:

- ▷ $\text{BLP}(Y|X = x) \neq E[Y|X = x]$ except in very special cases!

Best Linear Predictor (Contd.)

The BLP is one of many possible approximations to the CEF.

Why do we care about the best *linear* approximation?

- Many years ago: Easy computation.
- Conciseness: Just one/two numbers necessary to communicate.
- ~~Easy interpretation?~~
- Mathematical convenience & path dependence.

Why do we care about the L^2-loss?

- Large deviations are penalized more heavily: Cautious approach?
- Mathematical convenience & path dependence.

There are many alternative approximation approaches considered in frontier research and industry...

- ... but you'll need to take a more advanced econometrics class to learn about them. (Hopefully this one motivates you to do so!)

BLP-Coefficients

BLP-coefficients are known functions of moments of (Y, X):

Theorem 1

Let Y and X be random variables. If $Var(X) > 0$, then

$$\begin{aligned} &(\alpha, \beta) \in \underset{\alpha,\beta\in\mathbb{R}}{\arg\min}\ E\left[\left(E\left[Y|X\right] - (\alpha + X\beta)\right)^2\right] \\ \Leftrightarrow \qquad &\beta = \frac{Cov(Y, X)}{Var(X)}, \quad \text{and} \quad \alpha = E[Y] - E[X]\beta. \end{aligned} \tag{5}$$

Theorem 1 is hugely convenient:

- ▷ Well equipped for analyzing moments of (Y, X);
- ▷ Immediately suggest sample analogue estimator (patience, for now).

BLP-Coefficients (Contd.)

Proof. $R(\alpha,\beta) \equiv E\left[\left(E[Y|X] - (\alpha + X\beta)\right)^2\right]$

$$= E\left[E[Y|X]^2 - 2E[Y|X](\alpha + X\beta) + (\alpha + X\beta)^2\right]$$

$$= E\left[E[Y|X]^2\right] - 2\alpha \underbrace{E\left[E[Y|X]\right]}_{=E[Y]} - 2\underbrace{E\left[E[Y|X]X\right]}_{=E[E[YX|X]] = E[YX]}\beta + \alpha^2 + 2\alpha E[X]\beta + E[X^2]\beta^2$$

FOC: $\frac{\partial}{\partial \alpha} R(\alpha,\beta) = -2E[Y] + 2\alpha + 2E[X]\beta = 0$

$$\Leftrightarrow \alpha = E[Y] - E[X]\beta$$

$$\frac{\partial}{\partial \beta} R(\alpha,\beta) = -2E[YX] + 2\alpha E[X] + 2E[X^2]\beta$$

$$= -2E[YX] + 2E[Y]E[X] - 2E[X]^2\beta + 2E[X^2]\beta$$

$$= -2\underbrace{\left(E[YX] - E[Y]E[X]\right)}_{Cov(Y,X)} + 2\underbrace{\left(E[X^2] - E[X]^2\right)}_{Var(X)}\beta = 0$$

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

□

Linear Conditional Expectation Functions

The next result gives the *special case* when the BLP is the CEF.

Corollary 1

Let Y and X be random variables. If $E[Y|X]$ is linear, that is,

$$\exists \tilde{\alpha}, \tilde{\beta} \in \mathbb{R}: \quad E[Y|X] = \tilde{\alpha} + X\tilde{\beta}, \tag{6}$$

then, whenever $Var(X) > 0$, we have

$$E[Y|X] = BLP(Y|X). \tag{7}$$

Proof. $u \equiv Y - E[Y|X]$ w/ $E[u|X] = 0 \Rightarrow E[u] = E[E[u|X]] = E[0] = 0$

$$\beta = \frac{Cov(Y,X)}{Var(X)} = \frac{Cov(\tilde{\alpha} + X\tilde{\beta} + u, X)}{Var(X)} = \frac{Cov(X\tilde{\beta}, X)}{Var(X)} = \frac{Cov(X,X)}{Var(X)}\tilde{\beta} = \tilde{\beta}$$

$$\alpha = E[Y] - E[X]\beta = E[\tilde{\alpha} + X\tilde{\beta} + u] - E[X]\beta = \tilde{\alpha} + E[X]\underbrace{(\tilde{\beta} - \beta)}_{=0} + \underbrace{E[u]}_{=0} = \tilde{\alpha}$$

□

Linear Conditional Expectation Functions (Contd.)

For general random variables Y and X, are there good reasons to believe that $E[Y|X]$ is linear? Most of the time: No!

- ▷ Economic theory rarely motivates severe *functional* form restrictions.

However, there is an important setting when $E[Y|X]$ is linear w/o further restrictions: When X is a binary random variable.

Corollary 2

Let Y and X be random variables. If X is binary, then $E[Y|X]$ is linear.

Proof. WLOG, supp $X = \{0, 1\}$.

$$E[Y|X] = X\,E[Y|X=1] + (1-X)\,E[Y|X=0]$$
$$= E[Y|X=0] + X\left(E[Y|X=1] - E[Y|X=0]\right)$$
$$= \tilde{\alpha} + X\tilde{\beta}$$

□

Outline

1. Best Linear Predictor
2. **Properties of the BLP-Residual**
3. Interpretation of the BLP-Coefficient β
   - ▷ Descriptive Interpretation using Yitzhaki (1996)
   - ▷ Causal Interpretation under Random Assignment

BLP-Residual

The BLP-residual is the error when predicting Y using $\text{BLP}(Y|X)$.

- ▷ Convenient object in the analysis of the BLP.

Definition 2 (BLP-Residual)

Let Y and X be random variables. The BLP-*residual* ε is defined as

$$\varepsilon = Y - \text{BLP}(Y|X). \tag{8}$$

Note that

$$\varepsilon = (Y - E[Y|X]) + (E[Y|X] - \text{BLP}(Y|X)) \, . \tag{9}$$

- ▷ Encapsulates minimal-prediction error & BLP-approximation error.

Note: *Recall that $E[Y|X]$ is the best predictor of Y given X. You showed this yourself in Problem 6 in Problem Set 1!*

Properties of the BLP-Residual

The BLP-residual is mean-zero and uncorrelated to X.

- ▷ Importantly: This is not an assumption!

Lemma 1

Let Y and X be random variables. If $\varepsilon = Y - \text{BLP}(Y|X)$, then

$$E[\varepsilon] = 0, \quad \text{and} \quad E[\varepsilon X] = 0. \tag{10}$$

Proof.

i) $E[\varepsilon] = E[Y - \alpha - X\beta] = E[Y] - (E[Y] - E[X]\beta) - E[X]\beta = 0$

ii) $E[\varepsilon X] = E[(Y - \alpha - X\beta)X] = E[YX] - \alpha E[X] - \beta E[X^2]$

$= -\frac{1}{2} \underbrace{\frac{\partial}{\partial \beta} R(\alpha, \beta)}_{=0 \text{ by FOC}} = 0$

□

Properties of the BLP-Residual (Contd.)

In general, the BLP-residual is *not* mean-independent of X.

Lemma 2

Let Y and X be random variables. Let $\varepsilon = Y - \text{BLP}(Y|X)$. If $E[Y|X]$ is linear, then

$$E[\varepsilon|X] = 0. \tag{11}$$

If $E[Y|X]$ is not linear, then (11) does not hold in general.

Proof.

$$E[\varepsilon|X] = E[Y - BLP(Y|X)|X] \overset{\text{Cor 1}}{=} E[Y - E[Y|X]|X] = E[Y|X] - E[Y|X] = 0$$

Counter example w/ $E[Y|X]$ not linear: $Y = X^2$, $X \sim N(0,1)$.

$$Cov(Y, X) = \underbrace{E[X^3]}_{=0} - \underbrace{E[X]}_{=0}E[X^2] = 0 \Rightarrow \beta = 0, \quad \alpha = E[Y] - E[X]\beta = E[X^2] = 1$$

$$E[\varepsilon|X] = E[Y-1|X] = E[X^2|X] - 1 = X^2 - 1 \neq 0$$

□

Outline

1. Best Linear Predictor
2. Properties of the BLP-Residual
3. **Interpretation of the BLP-Coefficient β**
   - ▷ **Descriptive Interpretation using Yitzhaki (1996)**
   - ▷ Causal Interpretation under Random Assignment

Interpretation of the BLP-Coefficient β

Note that $\text{BLP}(Y|X)$ is a feature of the joint distribution of (Y, X) :

- Purely descriptive;
- Captures the *approximate* expected level of Y *associated* with a level of X.

Practitioners often calculate the difference in BLPs:

$$\text{BLP}(Y|X = x') - \text{BLP}(Y|X = x) = (\alpha + x'\beta) - (\alpha + x\beta) = (x' - x)\beta \quad (12)$$

When $x' - x = 1$, we may thus interpret the BLP-coefficient β as follows:

- "β captures the *approximate* expected change in Y *associated* with a unit-change in X."

Terminology is very important to avoid confusion:

- Need "approximate" to highlight that $\text{BLP}(Y|X) \neq E[Y|X]$;
- Need "associated" to emphasize purely descriptive interpretation.

Interpretation of the BLP-Coefficient β (Contd.)

When $E[Y|X]$ is linear, then β has another interpretation:

$$\frac{\partial}{\partial x} E[Y|X=x] \overset{(1)}{=} \frac{\partial}{\partial x} \mathrm{BLP}(Y|X=x) = \beta, \tag{13}$$

where (1) follows from Corollary 1.

- If $E[Y|X]$ is linear, then β is its derivative w.r.t. X.

The interpretation is appealing but is appropriate only in special cases.

- Would like derivative-interpretation for β when $E[Y|X]$ is not linear.
- Yitzhaki (1996) shows that this is possible... *with qualifications*.

Yitzhaki (1996) shows that β admits a *weighted* average derivative interpretation.

Theorem 2 (Yitzhaki's Theorem)

Let Y and X be random variables. Let β satisfy (4). Then

$$\beta = \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial t} E[Y|X = t] \right) \omega(t) dt, \tag{14}$$

where

$$\omega(t) = \frac{(E[X|X \geq t] - E[X|X < t]) \, P(X \geq t) P(X < t)}{Var(X)} \tag{15}$$

Yitzhaki's Theorem (Contd.)

Proof. Proof for continuous X.

Define $u \equiv Y - E[Y|X]$. Then $Cov(Y,X) = Cov(E[Y|X] + u, X) = Cov(E[Y|X], X)$

$$h(x) \equiv E[Y|X], \quad h_{-\infty} = \lim_{t \to -\infty} h(t).$$

By FTC, $h(x) = h_{-\infty} + \int_{-\infty}^{x} h'(t)\,dt$.

$$Cov(Y,X) = Cov(h(X), X) = Cov\left(h_{-\infty} + \int_{-\infty}^{X} h'(t)\,dt, X\right) = Cov\left(\int_{-\infty}^{X} h'(t)\,dt, X\right)$$

$$= E\left[\left(\int_{-\infty}^{X} h'(t)\,dt - E\left[\int_{-\infty}^{X} h'(t)\,dt\right]\right)(X - E[X])\right]$$

$$= E\left[\int_{-\infty}^{X} h'(t)\,dt\,(X - E[X])\right] - E\left[\int_{-\infty}^{X} h(t)\,dt\right] \underbrace{E[X - E[X]]}_{=0}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{x} h'(t)\,dt\,(x - E[X]) f_X(x)\,dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{x} h'(t)(x - E[X]) f_X(x)\,dt\,dx = \int_{-\infty}^{\infty} \int_{t}^{\infty} h'(t)(x - E[X]) f_X(x)\,dx\,dt$$

$$= \int_{-\infty}^{\infty} h'(t) \underbrace{\int_{t}^{\infty} (x - E[X]) f_X(x)\,dx}_{\equiv \tilde{\omega}(t)}\,dt$$

Yitzhaki's Theorem (Contd.)

$$\tilde{\omega}(t) = \int_t^{\infty} (x - E[X]) f_X(x)\,dx = \int_{-\infty}^{\infty} x \underbrace{\mathbb{1}\{x \geq t\} \frac{f_X(x)}{P(X \geq t)}}_{f_{X|X\geq t}(x)} P(X \geq t)\,dx - E[X] \underbrace{\int_t^{\infty} f_X(x)\,dx}_{=P(X \geq t)}$$

$$= E[X|X \geq t]P(X \geq t) - E[X]P(X \geq t) = (E[X|X \geq t] - E[X])P(X \geq t)$$

$$= (E[X|X \geq t] - (E[X|X \geq t]P(X \geq t) + E[X|X < t]P(X < t)))P(X \geq t)$$

$$= (E[X|X \geq t]P(X < t) - E[X|X < t]P(X < t))P(X \geq t)$$

$$= (E[X|X \geq t] - E[X|X < t])P(X \geq t)P(X < t)$$

Thus, $\beta = \dfrac{Cov(Y, X)}{Var(X)} = \displaystyle\int_{-\infty}^{\infty} h'(t) \underbrace{\frac{\tilde{\omega}(t)}{Var(X)}}_{=\omega(t)} dt$ □

Example 1

Let $X \sim U(0,1)$. Then, for any $t \in [0,1]$, we have

$$E[X|X \geq t] = \frac{1}{2}(t+1) \quad , E[X|X < t] = \frac{1}{2}(0+t)$$
$$P(X|X \geq t) = 1-t \quad , P(X|X < t) = t \quad , Var(X) = \frac{1}{12}$$

Hence, for any $t \in [0,1]$, the Yitzhaki weights are given by

$$\omega(t) = 12\left(\frac{1}{2}(1-t)t\right) = 6(1-t)t \tag{16}$$

By Yitzhaki's Theorem, it follows that

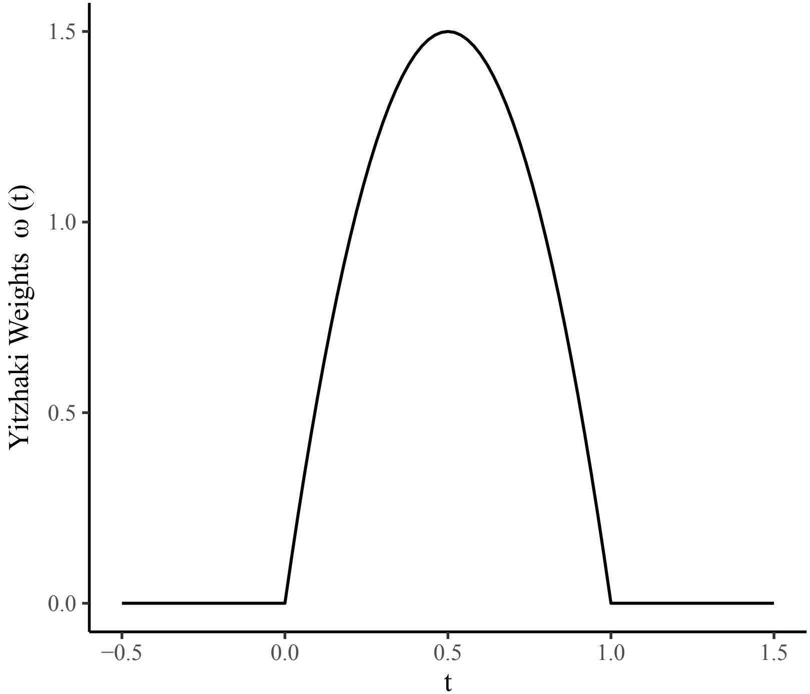
$$\beta = 6\int_0^1 \frac{\partial}{\partial x}E[Y|X=t](1-t)t\,dt \tag{17}$$

Which is distinct from the average derivative of $E[Y|X]$ given by

$$E\left[\frac{\partial}{\partial X}E[Y|X]\right] = \int_0^1 \frac{\partial}{\partial x}E[Y|X=t]\,dt \tag{18}$$

Yitzhaki's Theorem (Contd.)

Figure 1: Yitzhaki Weights for Standard Uniform X



Notes. Yitzhaki Weights for Standard Uniform X given by $\omega(t) = 6t(1-t)\mathbb{1}\{t \in [0,1]\}$. You can find the code generating the figure on GitHub: `lecture_plots.R`.

Yitzhaki's Theorem (Contd.)

The Yitzhaki weights are such that:

- The weights $\omega(t)$ are s.t. $\omega(t) \geq 0, \forall t$, and $\int_{-\infty}^{\infty} \omega(t) dt = 1$.
- Maximum weight reached at $t = E[X]$ (if density exists at $E[X]$). (See Problem 5 in Problem Set 3.)

Yitzhaki (1996) is remarkable:

- Relates β to a weighted average of the CEF derivative;
- Gives expression for the weights that allow for precise interpretation;
- Highlights that precise interpretation is... difficult!

Are practitioners thinking of Yitzahki's Theorem when interpreting β?

Outline

1. Best Linear Predictor
2. Properties of the BLP-Residual
3. **Interpretation of the BLP-Coefficient β**
   - ▷ Descriptive Interpretation using Yitzhaki (1996)
   - ▷ **Causal Interpretation under Random Assignment**

Causal Interpretation under Random Assignment

Consider the all causes model discussed in previous lectures:

$$Y = g(W, U). \tag{19}$$

When the policy variable W is continuous, a common parameter of interest is the *average structural function* (asf):

$$g_1(w) \equiv E_U[g(w, U)], \tag{20}$$

where w is fixed, not conditioned on! (E.g., Blundell and Powell, 2006)
To describe causal effects of marginal changes in the policy variables:

$$g_1'(w) \equiv \frac{\partial}{\partial w} g_1(w). \tag{21}$$

Practitioners are often content with a summary of $g_1'(w)$:

$$\overline{g}_1' \equiv E_W\left[g_1'(W)\right]. \tag{22}$$

- ▷ $\overline{g}_1'$ is the expected change in Y *caused* by a marginal change in W.

Causal Interpretation under Random Assignment (Contd.)

$\overline{g}'_1$ is a function (of the distribution) of U and is thus not identified.

- ▷ *Need* identifying assumption!

In the last lecture, we saw that under Assumption RA, we have

$$E[g(w, U)] = E[Y|W = w]. \tag{23}$$

Then simply

$$g'_1(w) = \frac{\partial}{\partial w} E[Y|W = w]. \tag{24}$$

From Yitzhaki's Theorem, it then follows that under RA, we have

$$\beta = \int_{-\infty}^{\infty} g'_1(t)\omega(t)dt. \tag{25}$$

- ▷ Under RA, may interpret β as weighted average of the asf-derivative;
- ▷ But β is generally distinct from average asf-derivative $\overline{g}'_1$.

Causal Interpretation under Random Assignment (Contd.)

The Yitzhaki interpretation for β in Equation (25) is often challenging. We thus also discuss a weaker alternative.

Recall that $\text{BLP}(Y|W = w)$ is an approximation to $E[Y|W = w]$.

- ▷ Under RA, $E[Y|W = w] = E[g(w, U)]$;
- ▷ Hence, $\text{BLP}(Y|W = w)$ is an approximation to $E[g(w, U)]$ whenever RA is assumed (but not generally!).

Assumption RA thus motivates a qualified causal interpretation of β:

- ▷ "Under RA, β captures the *approximate* expected change in Y *caused* by a unit-change in W."

Summary

Today, we introduced BLP($Y|X$) as approximation to $E[Y|X]$.

- ▷ Showed that the BLP-coefficients are well-defined when $Var(X) > 0$;
- ▷ Hopeful that this is a useful alternative to the direct analysis of $E[Y|X = x]$ when $P(X = x)$ is small.

But there is no free lunch...

- ▷ Approximation of $E[Y|X]$ makes interpretation of differences in BLP($Y|X$) challenging;
- ▷ Used Yitzhaki's Theorem to motivate a weighted-average derivative interpretation of β;
- ▷ Discussed interpretation of β under Assumption RA.

In Part B, we turn to estimating the BLP-coefficients

- ▷ Introduce the *ordinary least squares* estimator for (α, β);
- ▷ Analyze its statistical properties.

References

Blundell, R. and Powell, J. L. (2006). Endogeneity in nonparametric and semiparametric regression models. *Advances in Economics and Econometrics*, pages 312–357.

Yitzhaki, S. (1996). On using linear regressions in welfare economics. *Journal of Business & Economic Statistics*, 14(4):478–486.