Multiple Linear Regression
Part A: The Best Linear Predictor

Thomas Wiemann
University of Chicago

Econometrics
Econ 21020

Updated: May 16, 2022

Introduction

In lecture 7, we discussed the Selection on Observables (SO) assumption:

- ▷ Showed that $E[Y|W = w, X = x] = E[g(w, U)|X = x]$ under SO;
- ▷ Derived binning estimator for CATE and ATE for discrete (W, X).

But binning estimators are not versatile:

- ▷ For continuous/mixed (W, X), binning estimators are not applicable;
- ▷ Even for discrete (W, X), may run into the small bin problem.

Need an alternative estimator for the CEF $E[Y|W = w, X = x]$.

The alternative estimator we consider is *multiple* linear regression.

- ▷ Generalization of simple linear regression discussed in Lecture 6.

Introduction (Contd.)

Multiple linear regression has the same pros & cons discussed before:

- ▷ Easy to compute but difficult to interpret...
- ▷ Linear regression does not estimate the CEF directly!
- ▷ Linear regression estimates the *best linear approximation* of the CEF.

We again take two key steps:

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

B. Derive and characterize the linear regression estimator.

In contrast to Lecture 6, this time we focus on random *vectors*.

- ▷ Key results will be familiar, but proofs will be different.

Notation: *Throughout, vectors are always column vectors. Column vectors can be transformed to row vectors using the transpose-operator. In particular,* $x \in \mathbb{R}^p$, $p \in \mathbb{N}$ *is a column vector and* $x^\top$ *is a row vector.*

Outline

Outline

1. **Best Linear Predictor**
2. Properties of the BLP-Residual
3. Interpretation of the BLP-Coefficients
   - ▷ The Frisch-Waugh Theorem
   - ▷ Generalized Yitzhaki's Theorem
   - ▷ Causal Interpretation under Selection on Observables

Best Linear Predictor

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

- ▷ See Problem 5 of Problem Set 4 why this terminology is sensible.

Definition 1 (Best Linear Predictor; BLP)

Let Y be a random variable and $X = (1, X_1, \ldots, X_k)^\top$ be a random vector. The *best linear predictor* (BLP) of the conditional expectation $E[Y|X]$ is defined as

$$\mathrm{BLP}(Y|X) = X^\top \beta = \beta_0 + X_1\beta_1 + \ldots + X_k\beta_k, \tag{1}$$

where the BLP-coefficients $\beta = (\beta_0, \beta_1, \ldots, \beta_k)$ are such that

$$\beta \in \underset{\beta \in \mathbb{R}^{k+1}}{\arg\min}\ E\left[\left(E\left[Y|X\right] - X^\top \beta\right)^2\right]. \tag{2}$$

As before, the BLP is an *approximation* to the CEF:

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

BLP-Coefficients

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

Theorem 1

Let Y be a random variable and $X = (1, X_1, \ldots, X_k)^\top$ be a random vector. If $E[XX^\top]^{-1}$ exists, then

$$\begin{aligned} & \beta \in \underset{\beta \in \mathbb{R}^{k+1}}{\arg\min}\ E\left[\left(E\left[Y|X\right] - X^\top \beta\right)^2\right] \\ \Leftrightarrow \qquad & \beta = E[XX^\top]^{-1}E[XY]. \end{aligned} \tag{3}$$

Theorem 1 is hugely convenient:

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

Vector Differentiation Recap

As the objective in (2) is convex in β, FOCs are sufficient and necessary.

- Differentiate with respect to β, set to 0, then solve for β.

The difficulty: $\beta \in \mathbb{R}^{k+1}$ is a vector!

- Need vector differentiation rules (prerequisites?).

We only require the following rules, stated here without proof:

Lemma 1

Consider $x \in \mathbb{R}^p$, $A \in \mathbb{R}^{s,p}$, $B \in \mathbb{R}^{p,p}$ for $p, s \in \mathbb{N}$. Then

$$\frac{\partial}{\partial x} Ax = A, \qquad \frac{\partial}{\partial x^\top} x^\top A^\top = A, \\ \frac{\partial}{\partial x} x^\top Bx = x^\top (B^\top + B). \tag{4}$$

We're now equipped for the proof of Theorem 1.

Proof of Theorem 1

Proof. $R(\beta) = E[(E[Y|X] - X^T\beta)^2] = E[E[Y|X]^2 - 2E[Y|X]X^T\beta + \underbrace{(X^T\beta)^2}_{=\beta^T X X^T \beta}]$

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

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

FOC: $\frac{\partial R(\beta)}{\partial \beta} = -2E[YX^T] + \beta^T(E[XX^T]^T + E[XX^T])$

$= -2E[YX^T] + 2\beta^T E[XX^T] = 0^T_{k+1}$

$(XX^T)^T = (X^T)^T X^T = XX^T$

Transpose

$\Leftrightarrow \quad 0 = -2E[XY] + 2E[XX^T]\beta$

$\Leftrightarrow \quad E[XX^T]\beta = E[XY]$

$E[XX^T]^{-1}$ exists

$\Leftrightarrow \quad \underbrace{E[XX^T]^{-1}E[XX^T]}_{=I_{k+1}}\beta = E[XX^T]^{-1}E[XY]$

$\Leftrightarrow \quad \beta = E[XX^T]^{-1}E[XY]$

$\square$

Linear Conditional Expectation Functions

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

Corollary 1

Let Y be a random variable and $X = (1, X_1, \ldots, X_k)^\top$ be a random vector such that $E\left[XX^\top\right]^{-1}$ exists. If $E[Y|X]$ is linear, that is,

$$\exists \tilde{\beta} \in \mathbb{R}^{k+1} : \quad E[Y|X] = X^\top \tilde{\beta}, \tag{5}$$

$$= \tilde{\beta}_0 + X_1\tilde{\beta}_1 + \ldots + X_k\tilde{\beta}_k$$

then,

$$E[Y|X] = BLP(Y|X). \tag{6}$$

Proof. $E[Y|X] = X^\top\tilde{\beta}$

$\Rightarrow X E[Y|X] = XX^\top\tilde{\beta}$

$\Rightarrow E[XY|X] = XX^\top\tilde{\beta}$

$\Rightarrow \underbrace{E[E[XY|X]]}_{\substack{=E[XY] \\ \text{L.I.E.}}} = E[XX^\top]\tilde{\beta} \quad \Rightarrow \tilde{\beta} = E[XX^\top]^{-1}E[XY] = \beta$

BLP-coef. by Theorem 1

$\square$

Linear Conditional Expectation Functions (Contd.)

$\mathbb{1}_{x_1}(X)E[Y|X=x_1]+\mathbb{1}_{x_2}(X)E[Y|X=x_2]+\dots+\mathbb{1}_{x_{d_x}}(X)E[Y|X=x_{d_x}]$

$= \begin{bmatrix} \mathbb{1}_{x_1}(X) \\ \vdots \\ \mathbb{1}_{x_{d_x}}(X) \end{bmatrix}^T \begin{bmatrix} E[Y|X=x_1] \\ \vdots \\ E[Y|X=x_{d_x}] \end{bmatrix} = \tilde{X}^T\tilde{\beta}$ supp $X=\{x_1,\dots,x_{d_x}\}$

vector of indicators

Ex: $d_x=2$, $E[Y|X]=\mathbb{1}_{x_1}(X)E[Y|X=x_1]+\underbrace{\mathbb{1}_{x_2}(X)}_{=1-\mathbb{1}_{x_1}(X)}E[Y|X=x_2]$

As before, one should not generally believe that $E[Y|X]$ is linear.

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

Important exception: When X is discrete, then $E[Y|X]$ is linear in the set of indicators $\{\mathbb{1}_x(X)\}_{x\in\text{supp}\,X}$ w/o further restrictions: Let supp $X=\{x_1,\dots,x_{d_x}\}$

$$E[Y|X] = \mathbb{1}_{x_1}(X)E[Y|X=x_1]+\mathbb{1}_{x_2}(X)E[Y|X=x_2]+\dots+\mathbb{1}_{x_{d_x}}(X)E[Y|X=x_{d_x}]$$

$$= \begin{bmatrix} \mathbb{1}_{x_1}(X) \\ \vdots \\ \mathbb{1}_{x_{d_x}}(X) \end{bmatrix}^T \begin{bmatrix} E[Y|X=x_1] \\ \vdots \\ E[Y|X=x_{d_x}] \end{bmatrix} = \tilde{X}^T\tilde{\beta}$$

$\tilde{\beta}$ ← Vector of conditional expectations

$\tilde{X}$ ↑ Vector of indicators

E.g., $d_x=2$, then

$$E[Y|X]=\mathbb{1}_{x_1}(X)E[Y|X=x_1]+\mathbb{1}_{x_2}(X)E[Y|X=x_2]=\mathbb{1}_{x_1}(X)E[Y|X=x_1]+(1-\mathbb{1}_{x_1}(X))E[Y|X=x_2]$$

***Note**: Note that $E[Y|X]$ is not guaranteed to be linear in X even if X is discrete! It's important to transform X using indicators: $X=\sum_{x\in\text{supp}\,X}\mathbb{1}_x(X)x$.*

Outline

1. Best Linear Predictor
2. **Properties of the BLP-Residual**
3. Interpretation of the BLP-Coefficients
   - ▷ The Frisch-Waugh Theorem
   - ▷ Generalized Yitzhaki's Theorem
   - ▷ Causal Interpretation under Selection on Observables

BLP-Residual

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

- ▷ Convenient object in the analysis of the BLP.

Definition 2 (BLP-Residual)

Let Y be a random variable and $X = (1, X_1, \ldots, X_k)^\top$ be a random vector. The BLP-*residual* ε is defined as

$$\varepsilon = Y - \mathrm{BLP}(Y|X). \tag{7}$$

Properties of the BLP-Residual

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

- ▷ Importantly: This is not an assumption!

Lemma 2

Let Y be a random variable and $X = (1, X_1, \ldots, X_k)^\top$ be a random vector. If $\varepsilon = Y - \text{BLP}(Y|X)$, then

$$E[\varepsilon] = 0, \quad \text{and} \quad E[\varepsilon X] = 0. \tag{8}$$

$\Leftrightarrow E[\varepsilon X^T] = 0^T$

Proof.

$$E[\varepsilon X^T] = E[(Y - X^T\beta)X^T] = E[(Y - \beta^T X)X^T]$$

$$= E[YX^T] - \beta^T E[XX^T] = -\frac{1}{2}\frac{\partial R(\beta)}{\partial \beta} = 0^T$$

Take $e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

Then $0 = 0^T e_1 = E[\varepsilon X^T]e_1 = E[\varepsilon X^T e_1] = E[\varepsilon \cdot 1] = E[\varepsilon]$

$\square$

Properties of the BLP-Residual (Contd.)

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

In particular, if Y is a random variable, $X = (1, X_1, \ldots, X_k)^\top$ is a random vector, and $\varepsilon = Y - \text{BLP}(Y|X)$, then typically

$$E[\varepsilon|X] \neq 0, \tag{9}$$

except in very special cases (e.g., when the CEF is linear).

- ▷ See Problem 1e) of Problem Set 4.

Outline

Interpretation of the BLP-Coefficient β

Note that 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:

$$\text{BLP}(Y|X = x') - \text{BLP}(Y|X = x) = x'^{\top}\beta - x^{\top}\beta = (x' - x)^{\top}\beta \tag{10}$$

Note that x and x' are *vectors*. Interpretation:

- ▷ β captures the *approximate* expected change in Y *associated* with a change from $X = x$ to $X = x'$.

Terminology is very important to avoid confusion:

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

Solving for Subvectors of β

We're often interested in only a *sub*vector of the BLP-coefficient β.

- Often: The component of β corresponding to the policy variable.
- *Ceteris paribus*-principle.

Consider Y and $(X^\top, W) = (1, X_1, \dots, X_{k-1}, W)$.

- X is a random vector but W is a random variable.

Let $\beta = (\beta_0, \beta_1, \dots, \beta_{k-1}, \beta_W)^\top = (\beta_X^\top, \beta_W)^\top$ be the BLP$(Y|X, W)$-coefficient.

Suppose we're *only* interested in β_W.

- E.g., because W is the policy variable of interest;

How do we interpret β_W?

- β_W just the kth component of β...

Solving for Subvectors of β (Contd.)

Frisch and Waugh (1933) motivate an alternative interpretation of β_W.

Define

- $\tilde{Y} \equiv Y - BLP(Y|X)$;
- $\tilde{W} \equiv W - BLP(W|X)$.

Then the Frisch-Waugh Theorem shows

$$\beta_W = \frac{Cov(\tilde{W}, \tilde{Y})}{Var(\tilde{W})},$$

whenever $Var(\tilde{W}) > 0$.

Interpretation:

- β_W is the coefficient of W *controlling* for $X = (1, X_1, \ldots, X_{k-1})^\top$;
- But be very careful: *Controlling* is not *conditioning*!

BLP with De-Meaned Variables

We first consider simply de-meaning the variables under consideration.

Lemma 3

Let Y be a random variable and $X = (1, X_1, \ldots, X_k)^\top = (1, X_{1:k}^\top)^\top$ be a random vector. Let $\bar{Y} \equiv Y - E[Y]$ and $\bar{X} \equiv X_{1:k} - E[X_{1:k}]$. If $\beta = (\beta_0, \beta_1, \ldots, \beta_k)^\top$ are BLP$(Y|X)$-coefficients, then $\beta_{1:k} = (\beta_1, \ldots, \beta_k)$ are BLP$\left(\bar{Y}|\bar{X}\right)$-coefficients.

Proof. $Y = \beta_0 + X_{1:k}^\top \beta_{1:k} + \varepsilon$ w/ $\varepsilon \equiv Y - X^\top \beta$

$\Rightarrow Y - E[Y] = \beta_0 + X_{1:k}^\top \beta_{1:k} + \varepsilon - E[\beta_0 + X_{1:k}^\top \beta_{1:k} + \varepsilon]$

$= (X_{1:k} - E[X_{1:k}])^\top \beta_{1:k} + \varepsilon - \underbrace{E[\varepsilon]}_{=0}$

$\Rightarrow \bar{Y} = \bar{X}^\top \beta_{1:k} + \varepsilon$

$\Rightarrow \bar{X}\bar{Y} = \bar{X}\bar{X}^\top \beta_{1:k} + \bar{X}\varepsilon$

$\Rightarrow E[\bar{X}\bar{Y}] = E[\bar{X}\bar{X}^\top]\beta_{1:k} + \underbrace{E[\bar{X}\varepsilon]}_{=0}$

$\Rightarrow \beta_{1:k} = E[\bar{X}\bar{X}^\top]^{-1} E[\bar{X}\bar{Y}]$ $\square$

The Frisch–Waugh Theorem

Theorem 2 states a version of the result due to Frisch and Waugh (1933).

- ▷ Arguably one of the most important theorems in econometrics.

Theorem 2 (Frisch–Waugh Theorem)

Let Y be a random variable and $(X^\top, W) = (1, X_1, \ldots, X_{k-1}, W)$ be a random vector. Let $\tilde{Y} \equiv Y - BLP(Y|X)$ and $\tilde{W} \equiv W - BLP(W|X)$. If $Var(\tilde{W}) > 0$ and $\beta = (\beta_0, \beta_1, \ldots, \beta_{k-1}, \beta_W)^\top = (\beta_X^\top, \beta_W)^\top$ are $BLP(Y|X, W)$-coefficients, then

$$\beta_W = \frac{Cov(\tilde{W}, \tilde{Y})}{Var(\tilde{W})}. \tag{11}$$

Importantly: The Frisch–Waugh Theorem is a purely descriptive result!

- ▷ As before, the coefficient β_W is a purely descriptive parameter;
- ▷ Do not get fooled by fancy maths...

The Frisch–Waugh Theorem (Contd.)

Proof. $Y = X^T\beta_x + W\beta_w + \varepsilon$ w/ $\varepsilon \equiv Y - X^T\beta_x - W\beta_w$

$$\begin{aligned}
\Rightarrow Y - BLP(Y|X) &= X^T\beta_x + W\beta_w + \varepsilon - BLP(X^T\beta_x + W\beta_w + \varepsilon | X) \\
&= X^T\beta_x + W\beta_w + \varepsilon - X^T E[XX^T]^{-1} E[X(X^T\beta_x + W\beta_w + \varepsilon)] \\
&= X^T\beta_x + W\beta_w + \varepsilon - X^T E[XX^T]^{-1}\Big[E[XX^T]\beta_x + E[XW]\beta_w + \underbrace{E[X\varepsilon]}_{=0}\Big] \\
&= X^T\beta_x + W\beta_w + \varepsilon - X^T\beta_x - BLP(W|X)\beta_w \\
&= (W - BLP(W|X))\beta_w + \varepsilon
\end{aligned}$$

$$\Rightarrow \tilde{Y} = \tilde{W}\beta_w + \varepsilon$$

$$\Rightarrow \tilde{W}\tilde{Y} = \tilde{W}\tilde{W}\beta_w + \tilde{W}\varepsilon$$

$$\Rightarrow E[\tilde{W}\tilde{Y}] = E[\tilde{W}^2]\beta_w + \underbrace{E[\tilde{W}\varepsilon]}_{=0}$$

$$\Rightarrow \beta_w = \frac{Cov(\tilde{W}, \tilde{Y})}{Var(\tilde{W})}$$

□

Outline

1. Best Linear Predictor

2. Properties of the BLP-Residual

3. **Interpretation of the BLP-Coefficients**
   - ▷ The Frisch-Waugh Theorem
   - ▷ **Generalized Yitzhaki's Theorem**
   - ▷ Causal Interpretation under Selection on Observables

Interpretation of the BLP-Coefficient β (Contd.)

If $E[Y|X, W]$ is linear in both X and W, then

$$\frac{\partial}{\partial w} E[Y|X, W = w] \overset{(1)}{=} \frac{\partial}{\partial x} \text{BLP}(Y|X, W = w) = \beta_W, \tag{12}$$

where (1) follows from Corollary 1.

- Under linearity, β_W is the CEF derivative w.r.t. W.

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

Would like derivative-interpretation for β_W w/o functional assumptions...

- ... but we don't have one!

Generalized Yitzhaki's Theorem

Angrist and Krueger (1999) generalize Yitzhaki's Theorem (Lecture 6A):

- ▷ Don't restrict $E[Y|X, W]$ but assume $E[W|X]$ is linear.

Theorem 3 (Generalized Yitzhaki's Theorem)

Let Y and W be random variables and X be a random vector. Let β be the $BLP(Y|X, W)$-coefficient where β_W is the coefficient corresponding to W. If $E\left[Var(W|X)\right] > 0$ and $E[W|X]$ is linear, then

$$\beta_W = E\left[\int_{-\infty}^{\infty} \left(\frac{\partial}{\partial t} E[Y|W = t, X]\right) \omega(t, X) dt\right], \tag{13}$$

where

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

***Note**: Angrist and Krueger (1999) only provide formulas for a discrete variable of interest. Theorem 3 is a slight generalization of their result.*

Generalized Yitzhaki's Theorem (Contd.)

Proof.

Skipped. (Check back after the exam!)

Generalized Yitzhaki's Theorem (Contd.)

$\square$

Generalized Yitzhaki's Theorem (Contd.)

The generalized Yitzhaki weights are such that:

- $\forall x \in \operatorname{supp} X$, the weights $\omega(t,x)$ are s.t. $\omega(t,x) \geq 0, \forall t$, and $\int_{-\infty}^{\infty} \omega(t,x)dt = 1$.
- $\forall x \in \operatorname{supp} X$, maximum weight reached at $t = E[W|X = x]$ (if density exists at $E[W|X = x]$).

Similar to Yitzhaki's weights but now also w/ expectations w.r.t. X!

- Allows for precise interpretation as weighted average CEF derivative;
- But precise interpretation even more difficult w/ inclusion of X!

Are practitioners thinking of Theorem 3 when interpreting β_W?

- Recall: When linearity of $E[W|X]$ is not assumed, we don't even have a weighted-average derivative interpretation of β_W!

Outline

1. Best Linear Predictor
2. Properties of the BLP-Residual
3. **Interpretation of the BLP-Coefficients**
    - ▷ The Frisch-Waugh Theorem
    - ▷ Generalized Yitzhaki's Theorem
    - ▷ **Causal Interpretation under Selection on Observables**

Causal Interpretation under Random Assignment

Consider the all causes model discussed in previous lectures:

$$Y = g(W, U). \tag{14}$$

The *conditional average structural function* (casf) is

$$g_1(w, X) \equiv E_U[g(w, U)|X], \tag{15}$$

Conditional effects of marginal changes in the policy variable:

$$g_1'(w, X) \equiv \frac{\partial}{\partial w} g_1(w, X). \tag{16}$$

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

$$\overline{g}_1' \equiv E_{W,X}\left[g_1'(W, X)\right]. \tag{17}$$

- ▷ $\overline{g}_1'$ 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 lecture 7, we saw that under Assumption SO and CS, we have

$$E[g(w,U)|X] = E[Y|W = w, X]. \tag{18}$$

Then simply

$$g_1'(w,X) = \frac{\partial}{\partial w} E[Y|W = w, X]. \tag{19}$$

Under the conditions of Theorem 3, SO and CS, we then have

$$\beta_W = E\left[\int_{-\infty}^{\infty} g_1'(t,X)\omega(t)dt\right]. \tag{20}$$

- Under linearity of $E[W|X]$, SO, and CS, may interpret β as weighted average of the asf-derivative;
- But β_W is generally distinct from average asf-derivative $\overline{g}_1'$.

Causal Interpretation under Random Assignment (Contd.)

The Yitzhaki interpretation for β_W in Equation (20) is often challenging. We thus also discuss a weaker alternative.

$\text{BLP}(Y|W = w, X = x)$ is an approx./ to $E[Y|W = w, X = x]$.

- ▷ Under SO and CS, $E[Y|W = w, X = x] = E[g(w, U)|X = x]$.
- ▷ Hence, $\text{BLP}(Y|W = w, X = x)$ is an approx./ to $E[g(w, U)|X = x]$ whenever SO and CS are assumed.

SO and CS thus motivate an approximate causal interpretation of β_W:

- ▷ Under SO and CS, β_W captures the *approximate* expected change in Y *caused* by a unit-change in W.

Summary

Today, we generalized the BLP($Y|X$) for vector-valued X.

- ▷ Showed the BLP-coefficients are well-defined when $E[XX^\top]^{-1}$ exists;
- ▷ 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 BLP($Y|X$)-coefficients β challenging;
- ▷ Used Frisch-Waugh Theorem for analysis of sub-vector β_W;
- ▷ Used Theorem 3 to motivate a weighted-average derivative interpretation of β_W when $E[W|X]$ is linear;
- ▷ Discussed interpretation of β_W under SO and CS.

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

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

References

Angrist, J. D. and Krueger, A. B. (1999). Empirical strategies in labor economics. In *Handbook of Labor Economics*, volume 3, pages 1277–1366. Elsevier.

Frisch, R. and Waugh, F. V. (1933). Partial time regressions as compared with individual trends. *Econometrica*, pages 387–401.