Multiple Linear Equation Models

Ethan Ligon

March 16, 2022

Beginning

The beauty of econometrics and modern ways of thinking about estimation and inference are easy to lose sight of when we get caught up in technical details. But we're developing tools that really help us to understand the world we live in, and in particular to learn about economic behavior. This is a profound endeavor!

Naïveté

You've spent the first half of the semester developing a toolkit of linear methods in econometrics. For today, I want you to set that toolkit aside, and to think about some very basic issues, starting from a very naïve perspective. You've probably seen *all* of this material before, but I want to be sure we're all aware of the forest as we walk along examining trees.

Notation for Random Variables

Setting	Scalar	Vector	Matrix
Statistics	X	$\boldsymbol{x}$	$\boldsymbol{X}$
Econometrics	x	$\boldsymbol{x}$	$\boldsymbol{X}$
PDF	X or x	$\boldsymbol{x}$	$\boldsymbol{X}$
Jupyter	$\mathtt{x}$	$\mathtt{x}$	$\mathtt{X}$
Handwriting			

Defining Random Variables in python

See random_variables0.ipynb on datahub.

The Fundamental Linear Regression Model

Start with

$$y = \boldsymbol{X}\beta + u.$$

- Allowing $(y, \boldsymbol{X}, u)$ to all be random.
- $\boldsymbol{X}$ has full column rank.

Compare Classical Approach

E.g., R.A. Fisher; Fisher Box 1980

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{u}.$$

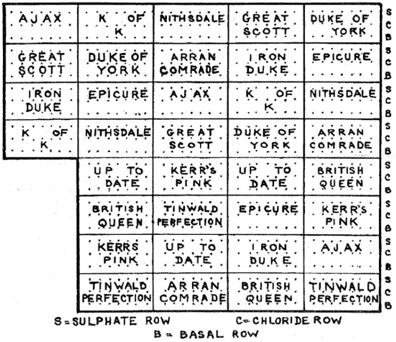
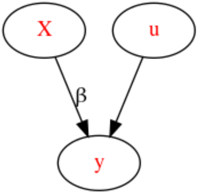


Diagram 1. Plan of experiment. Farmyard manure series.

Figure: "Triplicate Chessboard": Diagram I from Ronald A Fisher and Winifred A Mackenzie. 1923. Studies in crop variation. II. the manurial response of different potato varieties. *The Journal of Agricultural Science* 13 (3): 311–320

Classical Interpretation

The "dependent" variable y is determined by some random variables X with observations realized, and some random unobserved u. Critically, u and X are orthogonal; i.e., $\mathbb{E}(X^\top u) = \mathbf{0}$.



- The orthogonality of X and u is *not testable*, since u isn't observed.
- In general the causal diagram above imposes needed structure for interpreting regression.
- With this structure, β is "effect of variation in X on y."

Classical Regression in `python`

See `classical_regression.ipynb` on `datahub`.

Compare Bayes

$$y = \boldsymbol{X}\beta + u.$$

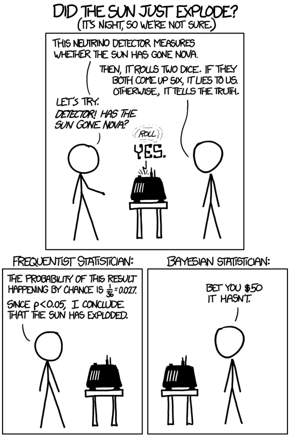
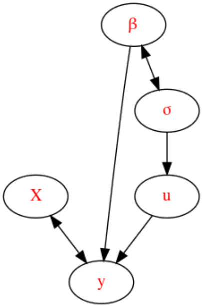


Figure: https://xkcd.com/1132/

Bayesian Interpretation

The disturbance u is independently distributed $Q(u|\sigma)$, and there's a "prior" $\Pr(\beta = \beta, \sigma = \sigma) = \Pr(\beta, \sigma)$ over these unknown vectors. The variables (β, σ) are ordinarily assumed to be distributed independently of $(\boldsymbol{X}, y)$.



Some Linear Estimation Problems with Multiple Equations

Let us develop some results for a broad class of linear estimators.
We've seen special cases of this before, so should look familiar.
Suppose we have an $n \times \ell$ random matrix $\boldsymbol{T}$.

Linear Model

$$y = \boldsymbol{X}\beta + u.$$

Premultiply by $\boldsymbol{T}$

$$\boldsymbol{T}^\top y = \boldsymbol{T}^\top \boldsymbol{X}\beta + \boldsymbol{T}^\top u.$$

We want to solve this equation for β. How should we proceed?

DISCUSS

Pretend you don't know any econometrics or statistics. What are issues? What if $\boldsymbol{T}$ is just a column of ones?

Aside on Moore-Penrose Inverse

`https://en.wikipedia.org/wiki/Moore%E2%80%93Penrose_inverse`

For any real matrix $\boldsymbol{A}$ (need not be square!), let $\boldsymbol{A}^+$ satisfy:

1. $\boldsymbol{A}\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{A}$; (generalizes idea that $\boldsymbol{A}\boldsymbol{A}^+ = \boldsymbol{I}$)
2. $\boldsymbol{A}^+\boldsymbol{A}\boldsymbol{A}^+ = \boldsymbol{A}^+$;
3. $(\boldsymbol{A}^+\boldsymbol{A})^\top = \boldsymbol{A}\boldsymbol{A}^+$; (form of symmetry)
4. $(\boldsymbol{A}\boldsymbol{A}^+)^\top = \boldsymbol{A}^+\boldsymbol{A}$.

Any such $\boldsymbol{A}^+$ satisfying these conditions is called the "Moore-Penrose Inverse" (or sometimes the pseudo-inverse).

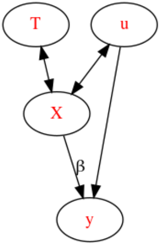
Facts about the Moore-Penrose Inverse:

1. $\boldsymbol{A}^+$ exists and is unique.
2. If the columns of $\boldsymbol{A}$ are linearly independent, then $\boldsymbol{A}^+ = (\boldsymbol{A}^\top\boldsymbol{A})^{-1}\boldsymbol{A}$ (this is sometimes called the "left inverse), and $\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{I}$.
3. If the rows of $\boldsymbol{A}$ are linearly independent, then $\boldsymbol{A}^+ = \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top)^{-1}$, and $\boldsymbol{A}\boldsymbol{A}^+ = \boldsymbol{I}$.

Linear Weighted Regression

A$_1$ $\mathbb{E}\boldsymbol{T}^\top u = \mathbf{0}$ (Orthogonality)

A$_2$ Some kind of full rank condition on $\boldsymbol{T}^\top \boldsymbol{X}$ (e.g., full column rank with probability one).



- The orthogonality of $\boldsymbol{T}$ and u is *not testable*, since u isn't observed.
- In general the causal diagram above imposes needed structure for interpreting regression.
- With this structure, β is "effect of variation in X on y."

Least Squares Estimator

Now:

$$\boldsymbol{T}^\top y = \boldsymbol{T}^\top \boldsymbol{X}\beta + \boldsymbol{T}^\top u$$
$$\Rightarrow \boldsymbol{T}^\top y - \boldsymbol{T}^\top u = (\boldsymbol{T}^\top \boldsymbol{X})\beta.$$

Now, using A_2:

$$(\boldsymbol{T}^\top \boldsymbol{X})^+(\boldsymbol{T}^\top y) - (\boldsymbol{T}^\top \boldsymbol{X})^+(\boldsymbol{T}^\top u) = \beta.$$

We only get a sample of realizations $(\boldsymbol{y}, \boldsymbol{X})$, and never even observe realizations of u. So to make further progress we take expectations and exploit A_1:

$$\mathbb{E}(\boldsymbol{T}^\top \boldsymbol{X})^+(\boldsymbol{T}^\top y) = \beta.$$

Analogy Principle (or Monte Carlo integration)

We have

$$\mathbb{E}(\boldsymbol{T}^\top \boldsymbol{X})^+(\boldsymbol{T}^\top y) = \beta.$$

We then apply the analogy principal, which allows us to substitute the mean of a sample for the expected value, yielding

$$b^* = (\boldsymbol{T}^\top \boldsymbol{X})^+(\boldsymbol{T}^\top \boldsymbol{Y}),$$

Then let we have $\mathbb{E}b^* = \beta$; i.e., this procedure results in an unbiased estimator of β for any $\boldsymbol{T}$ satisfying A_1 & A_2.

Linear Weighted Regression in python

See `weighted_regression.ipynb` on datahub.

Examples of $\boldsymbol{T}$

DISCUSSION on `bcourses`

Name an estimator we have encountered in the first half of the class that belong to this class of general weighted regressors. For each estimator what is the form of $\boldsymbol{T}$?