Section 3: CEF Properties and Regression Anatomy

Valentine Gilbert

September 23, 2021

Overview

Things You Should Know

Consider the **bivariate regression**: $Y_i = \beta_0 + \beta_1 X_i + u_i$

- $\beta_1 = cov(Y_i, X_i) / var(X_i)$
- $\hat{\beta}_1 = \widehat{cov}(Y_i, X_i) / \widehat{var}(X_i)$

Consider the **multivariate regression**: $Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} + u_i = X_i{}'\beta + u_i$

- $\beta = E[X_i X_i{}']^{-1} E[X_i Y_i]$
- $\hat{\beta} = \left[\frac{1}{N}\sum_{i=1}^{N} X_i X_i{}'\right]^{-1} \frac{1}{N}\sum_{i=1}^{N} X_i Y_i = [\boldsymbol{X'X}]^{-1}\boldsymbol{X'Y}$

In both cases, the regression residual is **uncorrelated** with the regressors

Motivation

- We'll prove the conditional expectation function $E[Y_i|X_i]$ is the **best predictor** of Y_i
- "Best" means the CEF is the function that minimizes the loss function $E[(Y_i - m(X_i))^2]$
- The CEF Prediction Property is a nice way to motivate linear regression: Regression gives us the best linear approximation to the CEF
- The proof of the CEF Prediction Property is pretty simple, but first we need to review some concepts and learn a new property of the CEF:
  - The Law of Iterated Expectations
  - The difference between independence, mean independence, and uncorrelatedness
  - The CEF Decomposition Property

The Law of Iterated Expectations

- The Law of Iterated Expectations: $E_X[E_Y[Y|X]] = E[Y]$
- Example: Mean income in MA is a population-weighted average of mean income in each county in MA
- The LIE is used in many econometric derivations

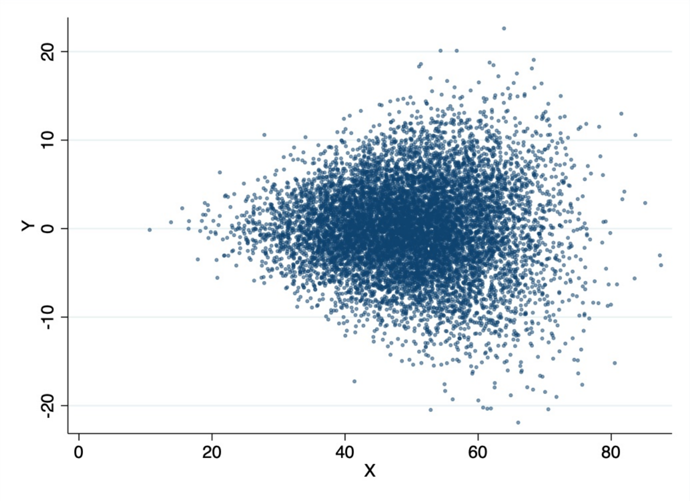
Proof.

$$\begin{aligned}
E_X[E_Y[Y|X]] &= E_X[\int y f_{Y|X}(y|x)dy] \\
&= \int \int y f_{Y|X}(y|x)dy f_X(x)dx \\
&= \int \int y f_{XY}(x,y)dydx \\
&= \int y(\int f_{XY}(x,y)dx)dy \\
&= \int y f_Y(y)dy \\
&= E[Y]
\end{aligned}$$

□

Independence, Mean Independence, and Uncorrelatedness

- We already talked about **independence**: Y and X are independent if $f(Y|X) = f(Y)$
- **Mean independence** is defined similarly: Y is mean independent of X if $E[Y|X] = E[Y]$
  - Note that Y can be mean independent of X even if X is not mean independent of Y
  - But $Y \perp X \iff X \perp Y$
- Mean independence implies that X and Y are **uncorrelated**
  - Correlation measures linear association and is weaker than mean independence
- In fact, if Y is mean independent of X, Y is uncorrelated with any function of X
  - Recall $cov(Y, h(X)) = E[Yh(X)] - E[Y]E[h(X)]$
  - By the LIE, $E[Yh(X)] = E[E[Yh(X)|X]] = E[h(X)E[Y|X]]$
  - By mean independence, $E[h(X)E[Y|X]] = E[h(X)E[Y]] = E[Y]E[h(X)]$
  - $\implies cov(Y, h(X)) = 0$



Are X and Y independent? Mean independent? Uncorrelated?

The CEF Decomposition Property

- The CEF Decomposition Property tells us we can decompose a random variable into the CEF and a residual with special properties
- $Y_i = E[Y_i|X_i] + \varepsilon_i$
  - ε_i is mean independent of X_i: $E[\varepsilon_i|X_i] = E[\varepsilon_i] = 0$
  - ε_i is uncorrelated with any function of X_i: $E[\varepsilon_i h(X_i)] = 0$
- Proof of mean independence:
  $E[\varepsilon_i|X_i] = E[Y_i - E[Y_i|X_i]|X_i] = E[Y_i|X_i] - E[Y_i|X_i] = 0$
- Mean independence of ε_i from X_i implies $cov(\varepsilon_i, h(X_i)) = E[\varepsilon_i h(X_i)] = 0$

The CEF Prediction Property

Now that we know the CEF decomposition property we're ready to prove the CEF prediction property!

- First, use the econometrician's favorite trick: Adding and subtracting the same thing in order to rewrite an expression

$$\begin{aligned}
(Y_i - m(X_i))^2 &= ((Y_i - E[Y_i|X_i]) + (E[Y_i|X_i] - m(X_i)))^2 \\
&= (Y_i - E[Y_i|X_i])^2 + (E[Y_i|X_i] - m(X_i))^2 + 2(E[Y_i|X_i] - m(X_i))(Y_i - E[Y_i|X_i]) \\
&= \varepsilon^2 + (E[Y_i|X_i] - m(X_i))^2 + 2\varepsilon(E[Y_i|X_i] - m(X_i))
\end{aligned}$$

- Taking expectations yields

$$E[(Y_i - m(X_i))^2] = var(\varepsilon) + E[(E[Y_i|X_i] - m(X_i))^2] + 2cov(\varepsilon, E[Y_i|X_i] - m(X_i))$$

- By mean independence, the last term is equal to zero. Thus the expected squared error is minimized when $m(X_i) = E[Y_i|X_i]$.

Variance Decomposition

- One last important property of the CEF is the Analysis of Variance theorem
- The ANOVA theorem says: $var(Y_i) = var(E[Y_i|X_i]) + E[var(Y_i|X_i)]$
- One application of this is to study income inequality in the U.S.
- Song et al. (2019 QJE) study the role of firms in the rise in earnings inequality in the U.S. from 1978 to 2013
  - Find that 1/3 of rise is due to increased earnings inequality *within* firms
  - Remaining 2/3 is due to increased inequality of average earnings *between* firms

Proof:

- From the CEF decomposition, $Y_i = E[Y_i|X_i] + \varepsilon_i$
- Since ε_i is uncorrelated with the CEF, $var(Y_i) = var(CEF) + var(\varepsilon_i)$
- $var(\varepsilon_i) = E[\varepsilon_i^2] = E[E[\varepsilon_i^2|X]] = E[var(Y_i|X_i)]$

Regression Anatomy: Theorem

- Recall that the slope coefficient from a bivariate regression of Y_i on X_i is

$$\beta_1 = \frac{cov(Y_i, X_i)}{var(X_i)}$$

- The key result from Wednesday's class is the regression anatomy result, which lets you think of the coefficients from a multivariate regression as the coefficients from a series of bivariate regressions
- This result is also known as the Frisch-Waugh-Lovell theorem

Theorem (Frisch-Waugh-Lovell)

Consider the multivariate regression equation, $Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} + \epsilon_i$.

$$\beta_j = \frac{cov(Y_i, \tilde{X}_{ji})}{var(\tilde{X}_{ji})} = \frac{cov(\tilde{Y}_i, \tilde{X}_{ji})}{var(\tilde{X}_{ji})},$$

where $\tilde{X}_i$ *is the residual from the regression of* X_{ji} *on the* $k-1$ *other regressors (and the constant).*

Regression Anatomy: Motivation

- FWL gives precise meaning to the idea that β_j is the relationship between X_j and Y "holding constant" the other covariates
- $\beta_j = cov(Y_i, \tilde{X}_{ji}) var(\tilde{X}_{ji})$ is also more intuitive than $\beta = E[X_i X_i']^{-1} E[X_i Y_i]$
- How can you test FWL in Stata or R?
    1. First, regress X_j on all other covariates and save the residuals
       reg x2 x1 x3 x4
       predict x2_squiggle, resid
    2. Then regress Y on $\tilde{X}_j$
       reg y x2_squiggle
- FWL also allows you to plot the relationship between Y and X_j, controlling for other covariates

Regression Anatomy: Proof

Let's go over the proof of FWL for the case where there are two covariates, X_1 and X_2:

- Consider the regression equation $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$
- Let $\alpha_1 = \frac{cov(X_{1i}, X_{2i})}{var(X_{2i})}$ and $\alpha_0 = E[X_{1i}] - \alpha_1 E[X_{2i}]$
- Define $\tilde{X}_{1i}$ as the residual from a regression of X_{1i} on X_{2i}: $X_{1i} = \alpha_0 + \alpha_1 X_{2i} + \tilde{X}_{1i}$

$$\begin{aligned} \implies \; cov(Y_i, \tilde{X}_{1i}) &= cov(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i, \tilde{X}_{1i}) \\ &= \beta_1 cov(X_{1i}, \tilde{X}_{1i}) + cov(\epsilon_i, \tilde{X}_{1i}) \end{aligned}$$

- Note: $cov(\epsilon_i, \tilde{X}_{1i}) = 0$, $cov(X_{1i}, \tilde{X}_{1i}) = cov(\alpha_0 + \alpha_1 X_{2i} + \tilde{X}_{1i}, \tilde{X}_{1i}) = var(\tilde{X}_{1i})$
- Therefore, $cov(Y_i, \tilde{X}_{1i}) = \beta_1 var(\tilde{X}_{1i})$
- Therefore, $\frac{cov(Y_i, \tilde{X}_{1i})}{var(\tilde{X}_{1i})} = \beta_1$

Regression Anatomy: Application

Let's consider an application of FWL: Exploring correlates of upward mobility

- I want to know the association between mean household income and upward mobility, so I run the bivariate regression

kfr_pooled_p25	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
hhinc_mean2000	.1331959	.0010175	130.91	0.000	.1312016	.1351901
_cons	23750.57	81.45435	291.58	0.000	23590.92	23910.22

- Does this relationship hold conditional on single parent share, college graduate share, and average test scores?

kfr_pooled_p25	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
hhinc_mean2000	.0129757	.001551	8.37	0.000	.0099357	.0160156
frac_coll_plus2000	10417.32	249.7108	41.72	0.000	9927.886	10906.75
singleparent_share2000	-25240.94	190.5843	-132.44	0.000	-25614.49	-24867.4
gsmn_math_g3_2013	994.8232	27.50184	36.17	0.000	940.9197	1048.727
_cons	35170	164.8833	213.30	0.000	34846.83	35493.17

Regression Anatomy: Application

- Recall Anscombe's quartet and the importance of visualizing your data
- But how can we visualize the data *holding covariates constant*?

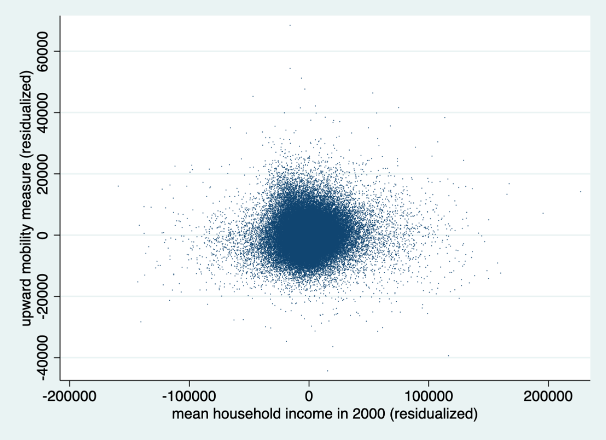
```
* Residualize y and x so we can visualize the relationship
reg hhinc_mean2000 frac_coll_plus2000 singleparent_share2000 gsmn_math_g3_2013,r
predict x_squiggle, resid

reg kfr_pooled_p25 frac_coll_plus2000 singleparent_share2000 gsmn_math_g3_2013,r
predict y_squiggle, resid
```

- First let's verify that we get the same regression coefficient

y_squiggle	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
x_squiggle	.0129755	.0015532	8.35	0.000	.0099312	.0160198
_cons	-.742224	21.38532	-0.03	0.972	-42.6574	41.17295

Regression Anatomy: Application

- Now we can visualize the relationship between our residualized y and x variables
- The raw scatter plot is hard to read. A binned scatter plot more clearly shows the expectation of Y conditional on X



Regression Anatomy: Application

- Now we can visualize the relationship between our residualized y and x variables
- The raw scatter plot is hard to read. A binned scatter plot more clearly shows the expectation of Y conditional on X