Chapter 4 Lab - Allen Church

October 9, 2019

## Preparation

require(knitr)  
require(haven)  
require(car)  
library(readxl)  
  
opts\_chunk$set(echo = TRUE)  
options(digits = 3)

### Load the data in Chapter4\_Lab\_ChildPoverty.xlsx (or Chapter4\_Lab\_ChildPoverty.RData). The codebook for the data is in the “codebook” tab of Chapter4\_Lab\_ChildPoverty.xlsx.

load("Chapter4\_Lab\_ChildPoverty.RData")  
View(dta)

#### (a) Estimate a model in which public spending on family benefits explains child poverty. What is the coefficient family spending? Is it statistically significant?

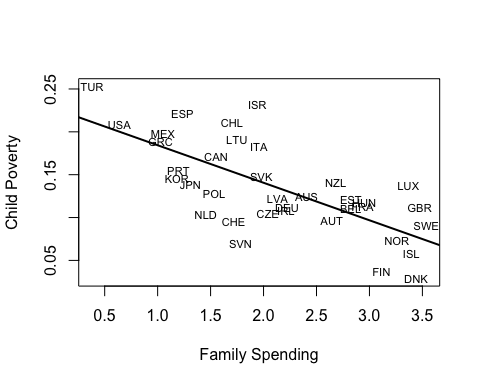
Statistical significance: t value > 2 means significant at 95% level

#Estimate first linear model and summarize results  
r1 <- lm(dta$child\_poverty ~ dta$family\_spending)  
summary(r1)

##   
## Call:  
## lm(formula = dta$child\_poverty ~ dta$family\_spending)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.0801 -0.0254 0.0056 0.0218 0.0888   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.22820 0.01724 13.24 5.7e-15 \*\*\*  
## dta$family\_spending -0.04379 0.00749 -5.85 1.4e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.0386 on 34 degrees of freedom  
## Multiple R-squared: 0.501, Adjusted R-squared: 0.487   
## F-statistic: 34.2 on 1 and 34 DF, p-value: 1.37e-06

From the model above, the coefficient of -0.04379 means that a 1 unit increase in family spending is associated with a decrease of 0.04379 units of child poverty units. However, we cannot call these results statistically significant, as the t value is -5.85.

#### (b) Create a scatterplot that corresponds to the above model. Include the OLS line, label the axes and label each point with the country name.



#### (c) Add the prediction and confidence intervals (as defined in class) to the above figure.

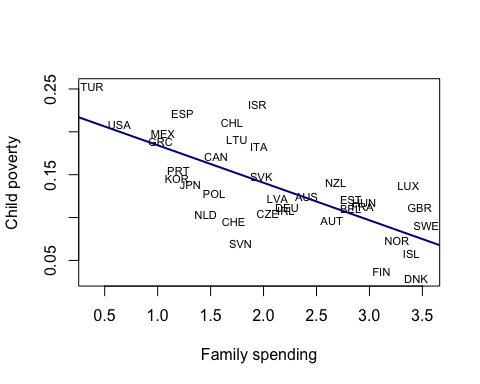
new.dat <- data.frame(family\_spending=seq(0, 4, length.out=15))  
Prediction.interval = predict(r1, newdata = new.dat, interval = 'prediction', level = 0.95)

## Warning: 'newdata' had 15 rows but variables found have 36 rows

Confidence.interval = predict(r1, newdata = new.dat, interval = 'confidence', level = 0.95)

## Warning: 'newdata' had 15 rows but variables found have 36 rows

plot(dta$child\_poverty ~ dta$family\_spending, xlab = "Family spending", ylab = "Child poverty", type="n",)  
text(dta$family\_spending, dta$child\_poverty, dta$country, cex = 0.7)  
abline(r1, col= "darkblue", lwd = 2)



# Add confidence intervals  
 #lines(x = new.dat$family\_spending, y = Confidence.interval[, "lwr"], col= "darkgreen")  
 #lines(x = new.dat$family\_spending, y = Confidence.interval[, "upr"], col= "darkgreen")  
 # Add prediction intervals  
 #lines(x = new.dat$family\_spending, y = Prediction.interval[, "lwr"], col= "red", lty = 2)  
 #lines(x = new.dat$family\_spending, y = Prediction.interval[, "upr"], col= "red", lty = 2)

#### (d) Estimate a model in which child poverty is a function of unemployment. Very briefly note the sign and significance of the coefficient on unemployment.

#Create linear model with child poverty as dependent variable and unemployment as independent variable. Summarize results  
result2 <- lm(dta$child\_poverty ~ dta$unemploy)  
summary(result2)

##   
## Call:  
## lm(formula = dta$child\_poverty ~ dta$unemploy)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.1027 -0.0286 -0.0089 0.0280 0.1101   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.10555 0.01767 5.97 9.3e-07 \*\*\*  
## dta$unemploy 0.00365 0.00193 1.89 0.067 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.052 on 34 degrees of freedom  
## Multiple R-squared: 0.0955, Adjusted R-squared: 0.0689   
## F-statistic: 3.59 on 1 and 34 DF, p-value: 0.0667

From the model above, the coefficient of 0.00365 means that a 1 unit increase in unemployment is associated with an increase of 0.00365 units of child poverty units. However, we cannot call these results statistically significant, as the t value is 1.89.

#### (e) What other factors should be included in multivariate model in order to help amerliorate endogeneity? Run a model with all other available and relevant independent variables. Very briefly comment on the coefficients on family\_spending and unemployment.

#Create multivariate model with child poverty as dependent variable. Independent variables are family spending, child vaccination rates, and unemployment. Summarize results  
result3 <- lm(dta$child\_poverty ~ dta$family\_spending + dta$child\_vac + dta$unemploy)  
summary(result3)

##   
## Call:  
## lm(formula = dta$child\_poverty ~ dta$family\_spending + dta$child\_vac +   
## dta$unemploy)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.08298 -0.02232 0.00126 0.02705 0.09040   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.41479 0.29155 1.42 0.16   
## dta$family\_spending -0.04185 0.00789 -5.30 8.3e-06 \*\*\*  
## dta$child\_vac -0.00213 0.00306 -0.70 0.49   
## dta$unemploy 0.00165 0.00155 1.06 0.30   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.039 on 32 degrees of freedom  
## Multiple R-squared: 0.521, Adjusted R-squared: 0.476   
## F-statistic: 11.6 on 3 and 32 DF, p-value: 2.62e-05

From the model above, the unemployment coefficient of 0.00165 suggests that a 1 unit increase in unemployment is associated with an increase of 0.00165 units of child poverty units. The child vaccination coefficient of -0.00213 suggests that a 1 unit increase in child vaccination units is associated with a decrease of 0.00213 of child poverty.

#### (f) Create a power curve for the *unemploy* variable for values of beta from 0 to 0.1. Use the actual standard error of the unemploy coefficient from the model explaining child poverty in the model you just ran. Sketch the power curve for one-sided . Approximately what is the probability of rejecting the null is if the true value of on unemploy is 0.003? What if the true beta on unemploy is 0.005?

#Create sequence with true betas  
TrueBeta <- seq(0.001, 0.01, by = 0.0001)  
  
#Calculate standard error by taking square root of variance  
se <- sqrt(vcov(result2)[2, 2])  
  
#Create quantile functions below  
qnorm(0.95, mean = 0, sd = 1)

## [1] 1.64

qnorm(0.975, mean = 0, sd = 1)

## [1] 1.96

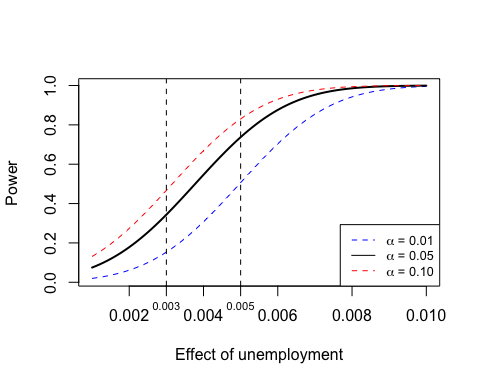
qnorm(0.995, mean = 0, sd = 1)

## [1] 2.58

#Curve with probability of normal distribution where mean is greater than 2.58  
power\_1 <- 1 - pnorm(2.58, mean = TrueBeta/se, sd = 1)  
#Curve with probability of normal distribution where mean is greater than 1.96  
power\_5 <- 1 - pnorm(1.96, mean = TrueBeta/se, sd = 1)  
#Curve with probability of normal distribution where mean is greater than 1.64  
power\_10 <- 1- pnorm(1.64, mean = TrueBeta/se, sd = 1)

Plotting code below

plot(TrueBeta, power\_1, type = "l", xlab = "Effect of unemployment", ylab = "Power", col = 4, lty=2)  
lines(TrueBeta, power\_5, col = 1, lwd = 2)  
lines(TrueBeta, power\_10, col = 2, lty = 2)  
abline(v = 0.003, col = 1, lty = 2)  
abline(v = 0.005, col = 1, lty = 2)  
legend("bottomright", c(expression(paste(alpha, " = 0.01")),  
 expression(paste(alpha, " = 0.05")),  
 expression(paste(alpha, " = 0.10"))),  
 lty = c(2,1,2), col = c(4,1,2), cex = 0.8)  
axis(1, at=0.003, labels = 0.003, cex.axis=0.7, mgp = c(2,.4,0)) #and label them on the x-axis  
axis(1, at=0.005, labels = 0.005, cex.axis=0.7, mgp = c(2,.4,0))



#### (g) Do the above exercise, but suppose that the standard error on the unemployment variable is half what it was above. See if you can create a single plot with two power curves, one for each standard error.

Plotting below

power\_5\_half <- 1- pnorm(1.96, mean = TrueBeta/(se/2), sd = 1)  
plot(TrueBeta, power\_5, type = "l", xlab = "Effect of unemployment", ylab = "Power",   
 col = 1)  
lines(TrueBeta, power\_5\_half, col = 2)  
legend("bottomright", legend = c("SD", "SD/2"), lty = c(1,1), col = c(1,2), cex = 0.8)  
abline(v = 0.003, col = 1, lty = 2)  
abline(v = 0.005, col = 1, lty = 2)  
axis(1, at=0.003, labels = 0.003, cex.axis=0.7, mgp = c(2,.4,0))  
axis(1, at=0.005, labels = 0.005, cex.axis=0.7, mgp = c(2,.4,0))

