

W271 Live Session 13: Analysis of Panel Data 2

Professor Jeffrey Yau

Spring 2019

Main topics covered in Week 12

- Fixed Effect Model
- A Digression: differencing when there are more than 2 time periods
- Random effect model
- Fixed effect vs. random effect models

Readings

W2016: Jeffrey Wooldridge. *Introductory Econometrics: A Modern Approach*. 6th edition. Cengage Learning

- Ch. 14.1 - 14.2
- [package plm] (<https://cran.r-project.org/web/packages/plm/plm.pdf>)
- [plm vignettes] (<https://cran.r-project.org/web/packages/plm/vignettes/plm.pdf>)

Agenda for this week's live session:

1. Discussion of Task 3 from Week 11's live session
2. An Example

Recap:

Remember that the observations in our (cross-section) linear regression models covered in w203 and discrete-response models covered in lecture 1 - 5 assume that observations are independent of each other. This is, nevertheless, a very strong assumption that is often not met in practice. Unfortunately, many practitioners still simply apply linear regression models to data whose observations are dependent due to temporal dependence. Studies collecting measures over time clearly cannot use the independence assumption.

Some start-up codes:

```
#sessionInfo()

# Insert the function to *tidy up* the code when they are printed out
library(knitr)
opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE)

# Clean up the workspace before we begin
rm(list = ls())

# Set working directory
wd <- "~/Documents/Teach/Cal/w271/course-main-dev/live-session-files/week12"
```

```

setwd(wd)

# Load libraries
library(car)

## Loading required package: carData
library(dplyr)

##
## Attaching package: 'dplyr'
## The following object is masked from 'package:car':
##
##      recode
## The following objects are masked from 'package:stats':
##
##      filter, lag
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
library(Hmisc)

## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2
##
## Attaching package: 'Hmisc'
## The following objects are masked from 'package:dplyr':
##
##      src, summarize
## The following objects are masked from 'package:base':
##
##      format.pval, units
library(gplots)

##
## Attaching package: 'gplots'
## The following object is masked from 'package:stats':
##
##      lowess
library(ggplot2)
library(stats)
library(foreign)
library(lattice)
library(plm)

```

```
##
## Attaching package: 'plm'

## The following objects are masked from 'package:dplyr':
##
##      between, lag, lead
```

An Example

Panel Data Exploratory Data Analysis (EDA)

In this example, we use a modified panel dataset on 7 countries. The *yIndex* is a measure of the well-being of the population in a country, considering that as some health measure. A bigger number reflects a better health status. The *xIndex* variable is a (normalized) measure of the economic activities of the country; a bigger number indicates “better” economic performance. *country* indicates, well, country, and *year* indicates the year in which the health and economic measurements were taken. The question of interest is “Does better economic activities, as measured by *xIndex*, improve the country’s health status, as measured by *yIndex*?”

Tasks: - Import the data `data_wk12.csv` as a dataframe. - Examine the basic structure of the data - At what level is this dataset? - Conduct a thorough EDA, including a growth curve analysis that was covered in the live session in week 12 - Discuss the EDA results

```
# YOUR CODE HERE
```

OLS Regression Models

Tasks: - Suppose you want to answer the question posted above “Does better economic activities, as measured by *xIndex*, improve the country’s health status, as measured by *yIndex*?” using an OLS regression model. - Is OLS regression model a reasonable approach in this case? Please explain. - Remember that there are different ways OLS regressions can be applied. One way is pooled-OLS. - What do you have to assume to make this approach reasonable? - Are these assumptions that you will have to make valid? - Interpret the estimation result! According to your model, what’s your answer to the question posted above?

```
# YOUR CODE HERE
```

Fixed-Effect Regression Model

Tasks: - Suppose you want to answer the question posted above using a Fixed-Effect regression model. - First, write down a fixed-effect regression model with an unobserved effect. - What assumptions are needed for the fixed-effect regression model? - Explain what a fixed effect is? - Think about why a fixed-effect framework would make sense for this question? - Estimate a fixed-effect model. - In your fixed-effect model, how many groups, time periods, and observations are there? - Don’t forget to interpret the result! So, according to your model, what’s your answer to the question posted above?

```
# YOUR CODE HERE
```

Random-Effect Regression Model

Tasks: - Suppose you want to answer the question posted above using a Random-Effect regression model. - First, write down a random-effect regression model with an unobserved effect. - What assumptions are needed for the random-effect regression model? - Explain what a random effect is? - Think about why a

random-effect framework would make sense for this question? - Estimate a fixed-effect model. - Don't forget to interpret the result! So, according to your model, what's your answer to the question posted above?

YOUR CODE HERE

Diagnostic Tests and Final Discussion

Tasks: - Conduct diagnostic tests to determine whether or not a fixed effects model or a random effects model is appropriate. - Discuss situations (and assumptions) under which fixed effect is more appropriate and situations under

- Hausman test: null hypothesis is that the preferred model is random effects vs. the alternative the fixed effects. It tests whether the unique errors u_i are correlated with the explanatory variables; the null hypothesis is they are not.
 - Estimate a fixed effects model and a random effect model. Then perform the test. If the p-value is significant (e.g. $p \leq 0.05$), then use fixed effects; otherwise, use random effects.

YOUR CODE HERE