The Effects of Foreign Aid on World Development Indicators

Charley Ferrari, Christina Taylor, David Stern

May 6, 2016

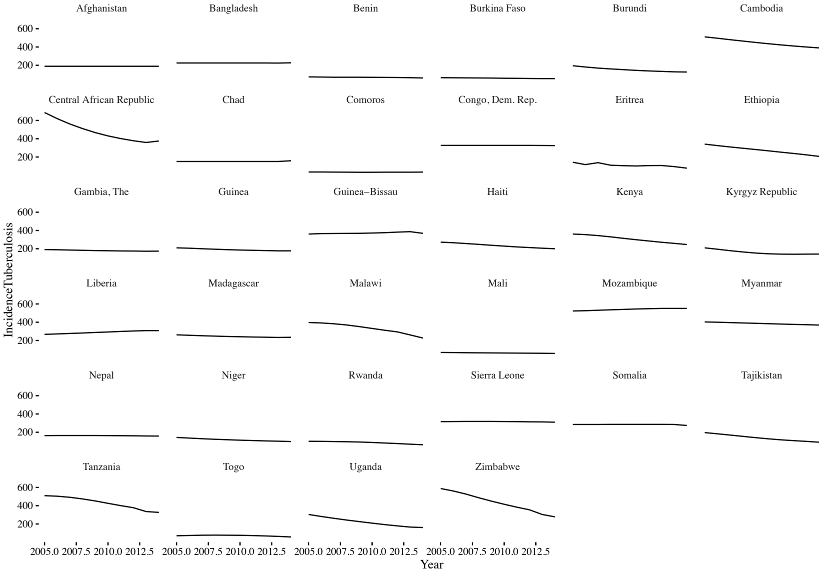
Panel Data Analysis

- Panel regression is different than the other methods studied in this class.
- Data is grouped in some way. In our case, we're looking at data across countries.
- The results of a panel regression speak more in terms of inference than prediction
- Because of this, it is related to methods not commonly thought of in terms of regression, like ANOVA Part of the analysis is determining how much varience is occurring between groups versus how much varience is occurring within groups.

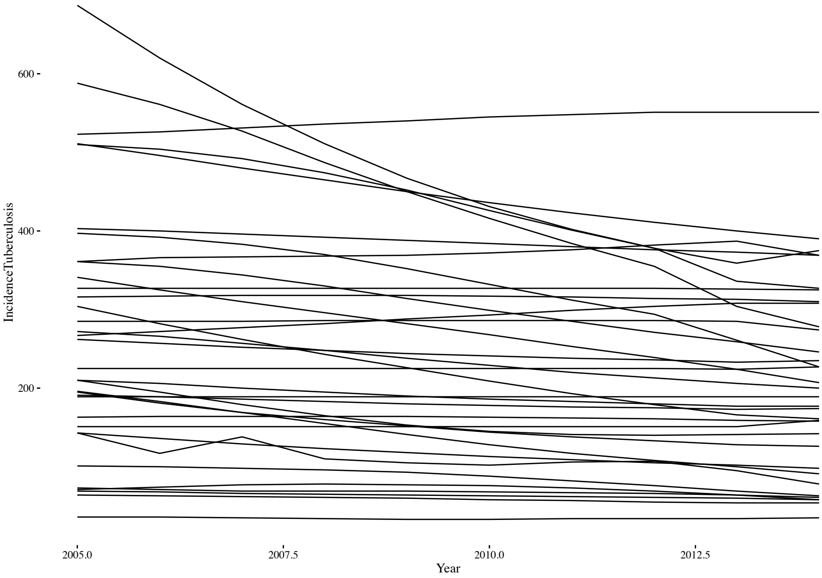
The World Bank Development Indicators

- The World Bank puts out a series of development indicators, both economic and societal, compiled from officially-recognized international sources.
- Data is presented on a global, regional, and national level. For the purposes of this project, we're looking at national level data.

Exploratory analysis of Tuberculosis Incidence



Exploratory analysis of Tuberculosis Incidence



Exploratory analysis of Tuberculosis Incidence

- The plot faceted by country allows us to scan for patterns between countries.
- Seeing everything together gives us a clearer idea of what sort of variation we have between groups.
- Tuberculosis seems to have higher variance within countries rather than between them.

Modeling by Country

- As a first step, we can just simply model Tuberculosis as a function of country. We can do this by using country as a categorical variable, and creating 49 dummy variables for our data.
- This method on its own is the same as an analysis of variance (ANOVA). In ANOVA, the central question is whether or not the means of different group are significantly different from eachother. In a linear regression, we can look at p-values of each dummy variable to decide the significance

ICC

- The ICC, or Intra-class correlation coefficient, is the ratio of the variance in y due to the fixed effects to the total variance in y. The ICC (ρ) is defined as:

$$\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2}$$

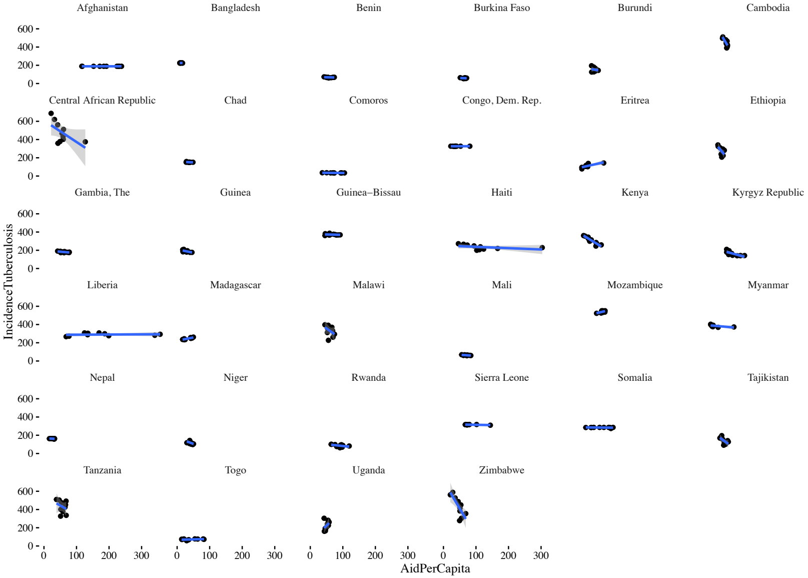
```
## [1] 0.9323198
```

Having an ICC this close to 1 implies what we saw in our plots earlier: The variation between groups is accounting for most of the variation of this particular dataset.

Looking at Aid Per Capita

- In looking at development indicators across low income countries, our main research question is how aid affects these outcomes.

Looking at Aid Per Capita



Naive Model

- This model won't take country into account. We will just consider countries and time periods to be independent observations. In this case, our model takes the form:

$$IncidenceTuberculosis = \beta_0 + AidPerCapita \times \beta_1$$

- R-Squared:

```
## [1] -0.001874993
```

Variables	Std.Error	t.value	pr.t
Intercept	243.84	19.34	0.000
AidPerCapita	-0.10	-0.60	0.546

Naive Model

- This model performs very poorly, it's pretty obvious we need more information. The negative sign in front of AidPerCapita is good, implying that as aid goes up Tuberculosis rates go down, but it's not significant, and there is a low r-squared.

Fixed Effects Model

- The form of the fixed effects model in this case is:

$$IncidenceTuberculosis = \mu + c_i + AidPerCapita \times \beta$$

- The c_i is the fixed effect, with i referring to the country.
- With just fixed effects, this model is equivalent to an lm with Country as a categorical variable (with n countries, n-1 dummy variables.)
- The Estimate for AidPerCapita is now -0.23, with a p-value of 0.0026.
- Still similar signs, which is good, along with a higher magnitude.

Random Effects Model

- The random effects model assumes that the effect of Country is due to a random variable.
- We wouldn't estimate the variables directly like in a fixed effect model, but would end up estimating the parameters of the random variable. The random effect has to have a mean of 0, so the important parameter being estimated is the variance.

Random Effects Model

- The coefficient for AidPerCapitaseems similar to what we had in our fixed effects model, at -0.229, with a similar p-value. The variance of the idiosyncratic random effects is 1278.03, whilet he individual random effects is 18649.36
- This model assumes however that the fixed effect isn't correlated with AidPerCapita. We can use a Hausman Test to find out if that's true.

Hausman Test

```
##
##  Hausman Test
##
## data:  IncidenceTuberculosis ~ AidPerCapita
## chisq = 0.078022, df = 1, p-value = 0.78
## alternative hypothesis: one model is inconsistent
```

- The Hausman test confirms that we should go with the alternative hypothesis. The random effects model is exhibiting omitted variable bias.

GLS

- Running a Panel GLS drastically improves the R-Squared: For this model it becomes 0.9375.
- More research would need to be done here.
- Bhargava et al has an example of adopting the Durbin Watson test for Panel data.