

Lab 2 (QMSS5016 Time Series, Panel Data & Forecasting)

Submitted by: Gideon Tay

My UNI: gt2528

Contact me at: gideon.tay@columbia.edu

Option 1: Fixed/ Random Effects

Overview: in this lab, we will explore the question of whether stronger rule of law is associated with higher levels of foreign direct investment into a country (as a % of GDP). As we are using a panel dataset of countries' data over time, we use fixed and random effects models to account for possible unobservable differences across countries and years in our analysis. We also consider a third variable, corruption, which we think may account for the relationship between rule of law and foreign direct investment levels.

Import libraries, load in and prepare the data

For this lab, I will use the University of Gothenburg's Quality of Government (QoG) Institute's basic time series dataset ([more information here](#)). The codebook for the data can be [found here](#).

Citation: Dahlberg, Stefan, Aksel Sundström, Sören Holmberg, Bo Rothstein, Natalia Alvarado Pachon, Cem Mert Dalli, Rafael Lopez Valverde & Paula Nilsson. 2024. The Quality of Government Basic Dataset, version Jan24. University of Gothenburg: The Quality of Government Institute, <https://www.gu.se/en/quality-government> doi:10.18157/qogbasjan24

```
In [1]: # Import libraries needed for this lab
import pandas as pd
import statsmodels.formula.api as smf
from linearmodels.panel import PanelOLS
from linearmodels.panel import RandomEffects
from linearmodels.panel import compare
import numpy as np
from scipy import stats

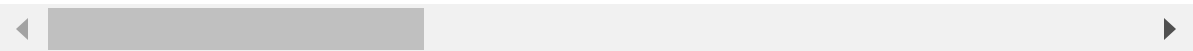
# Load in the data
url = 'https://www.qogdata.pol.gu.se/data/qog_bas_ts_jan24.xlsx'
df = pd.read_excel(url)

# View the first 5 rows of the data
df.head(5)
```

Out[1]:

	ccode	cname	year	ccode_qog	cname_qog	ccodealp	ccodecow	version
0	4	Afghanistan	1946	4	Afghanistan	AFG	700.0	QoGBasTSjan24
1	4	Afghanistan	1947	4	Afghanistan	AFG	700.0	QoGBasTSjan24
2	4	Afghanistan	1948	4	Afghanistan	AFG	700.0	QoGBasTSjan24
3	4	Afghanistan	1949	4	Afghanistan	AFG	700.0	QoGBasTSjan24
4	4	Afghanistan	1950	4	Afghanistan	AFG	700.0	QoGBasTSjan24

5 rows × 251 columns



Before we begin analysis, we prepare the data by dropping rows with missing values in columns we are interested in:

In [2]:

```
# Prepare data by dropping rows with missing values in key columns
df = df[['wbgi_rle', 'wdi_fdiin', 'vdem_corr', 'cname', 'year']].dropna()

# Check which years we have data for
unique_years = sorted(df['year'].unique().tolist())
print(unique_years)
```

[1996, 1998, 2000, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022]

We have data every year from 2002 to 2022, along with data for 1996, 1998, and 2000. To simplify things, we will only consider data for the continuous period of 2002 to 2022. Moreover, since the fixed effects model we will be using only helps to account for time-invariant but not time variant heterogenous factors across countries, using a smaller time period could be useful as well.

Let's drop the 1996, 1998, and 2000 data:

In [3]:

```
# Drop rows where the year is 1996, 1998, or 2000
df = df[~df['year'].isin([1996, 1998, 2000])]
```

(a) Run an OLS regression, including at least one independent variable and a time variable (as dummies). Explain how you think your independent variable relates to your dependent variable. Interpret your results. Did you find what you expected to find?

Dependent variable: Foreign direct investment, net inflows, as a percentage of GDP (QoG code: wdi_fdiin). It is defined as the investment inflows less disinvestment in the reporting economy from foreign investors, divided by GDP. Only consider investments to acquire a lasting management interest (10% or more of voting stock) in an enterprise operating in an economy other than that of the investor.

Independent variable: Rule of Law (QoG code: wbi_rle). It measures the extent to which agents have confidence in and abide by the rules of society. These include perceptions of the incidence of crime, the effectiveness and predictability of the judiciary, and the enforceability of contracts. Together, these indicators measure the success of a society in developing an environment in which fair and predictable rules form the basis for economic and social interactions and the extent to which property rights are protected.

Expected relationship: I expect that countries with stronger rule of law tend to have higher foreign direct investment (FDI) levels. With stronger rule of law, foreign investors have greater confidence that post-investment, their property rights and contracts will be well protected, and dispute resolution processes would be fair. This reduces the risk of investment, leading to higher FDI levels.

Let's run a naive OLS regression of FDI against rule of law and dummy year variables:

```
In [4]: # Run an OLS regression without country fixed effects
ols_model = smf.ols(
    formula='wdi_fdiin ~ wbi_rle + C(year)', data=df
).fit()

# Display the summary of the OLS regression
ols_model.summary()
```

Out[4]:

OLS Regression Results

Dep. Variable:	wdi_fdiin	R-squared:	0.021
Model:	OLS	Adj. R-squared:	0.015
Method:	Least Squares	F-statistic:	3.457
Date:	Tue, 29 Oct 2024	Prob (F-statistic):	1.65e-07
Time:	00:18:05	Log-Likelihood:	-15173.
No. Observations:	3487	AIC:	3.039e+04
Df Residuals:	3465	BIC:	3.053e+04
Df Model:	21		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
Intercept	3.7232	1.471	2.531	0.011	0.839	6.607
C(year)[T.2003]	0.4598	2.076	0.221	0.825	-3.611	4.531
C(year)[T.2004]	1.2738	2.080	0.613	0.540	-2.804	5.351
C(year)[T.2005]	3.2625	2.073	1.574	0.116	-0.803	7.328
C(year)[T.2006]	4.7766	2.073	2.304	0.021	0.712	8.842
C(year)[T.2007]	5.8207	2.067	2.816	0.005	1.768	9.874
C(year)[T.2008]	3.6214	2.067	1.752	0.080	-0.432	7.674
C(year)[T.2009]	2.5608	2.067	1.239	0.216	-1.492	6.614
C(year)[T.2010]	3.0657	2.067	1.483	0.138	-0.987	7.119
C(year)[T.2011]	3.3879	2.067	1.639	0.101	-0.665	7.441
C(year)[T.2012]	3.6010	2.070	1.739	0.082	-0.458	7.660
C(year)[T.2013]	1.5681	2.067	0.759	0.448	-2.485	5.621
C(year)[T.2014]	1.6877	2.067	0.816	0.414	-2.365	5.741
C(year)[T.2015]	2.3396	2.070	1.130	0.259	-1.719	6.399
C(year)[T.2016]	0.9475	2.073	0.457	0.648	-3.118	5.013
C(year)[T.2017]	0.3823	2.073	0.184	0.854	-3.683	4.447
C(year)[T.2018]	-1.4786	2.073	-0.713	0.476	-5.544	2.586
C(year)[T.2019]	2.7103	2.073	1.307	0.191	-1.355	6.775
C(year)[T.2020]	-1.0663	2.076	-0.514	0.608	-5.137	3.005
C(year)[T.2021]	-0.2908	2.083	-0.140	0.889	-4.374	3.793

C(year)[T.2022]	-2.1635	2.093	-1.034	0.301	-6.266	1.939
wbgi_rle	1.8847	0.326	5.783	0.000	1.246	2.524
Omnibus:	5030.207	Durbin-Watson:	0.957			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	8708497.190			
Skew:	7.828	Prob(JB):	0.00			
Kurtosis:	247.321	Cond. No.	23.1			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Interpretation of results:

- The coefficient on **wbgi_rle** is 1.8847. This suggests that every 1 unit increase in the Rule of Law index is associated with a 1.8847% increase in FDI as a percentage of GDP, net of time effects (controlled by dummy year variables).
- The coefficient on **wbgi_rle** is statistically significantly non-zero (p-value <0.001), and has a 97.5% confidence interval of [1.246, 2.524]. This strongly supports our initial expectation and hypothesis that Rule of Law and FDI have a positive relationship.
- The dummy year variables controls for time period effects. For example, with globalized markets, a period of high economic growth and investor exuberance could lead to high FDI levels for those years, across countries. 2002 is omitted and used as the reference year.
- The coefficient on dummy year variables indicate the average percentage point difference in FDI (% of GDP) between that specific year and the reference year 2002, net of Rule of Law index values.
- The coefficients on dummy year variables are all not statistically significantly non-zero (p-value>0.05), except for the years 2006 and 2007 (p-value<0.05), whose positive coefficients suggests that they have significantly higher FDI levels relative to 2002. 2006 and 2007 were the years just before the 2008 financial crisis, and was a period of high investor exuberance and associated high FDI levels globally.

(b) Run a fixed effect model version of that OLS model. Interpret your results. Did you find what you expected to find? Why? Why not?

Our previous OLS regression did not account for country-level fixed effects, the possibility that FDI levels may vary across countries for reasons other than Rule of Law. For example,

the underlying availability of investment opportunities may vary across countries and affect FDI. Natural resource endowments in some countries for instance may attract FDI.

To control for unobserved country-level heterogeneity, let's run a fixed effects model. We set the column `cname` as the entity index and `year` as the time index so that our fixed effects model controls for both country-level and time-level fixed effects. We also use clustered standard errors to account for potential correlation of errors within countries.

Let's run the fixed effects model:

```
In [5]: # Set the time (year) and entity (country) index
df2 = df.set_index(['cname', 'year'])

# Fit the fixed effects model controlling for time and country
# Clustered std errors for potential error correlation within clusters
fe_model = PanelOLS.from_formula(
    'wdi_fdiin ~ wbg_i_rle + EntityEffects + TimeEffects',
    data=df2
).fit(cov_type='clustered', cluster_entity=True)

# Display the summary of the regression
print(fe_model.summary)
```

PanelOLS Estimation Summary

```
=====
Dep. Variable:          wdi_fdiin    R-squared:                0.0061
Estimator:              PanelOLS     R-squared (Between):      -0.3977
No. Observations:       3487         R-squared (Within):       0.0047
Date:                   Tue, Oct 29 2024 R-squared (Overall):      -0.1031
Time:                   00:18:05      Log-likelihood            -1.469e+04
Cov. Estimator:         Clustered

                                F-statistic:          20.133
                                P-value                0.0000
Entities:                171         Distribution:            F(1,3295)
Avg Obs:                 20.392
Min Obs:                 4.0000
Max Obs:                 21.000
                                F-statistic (robust):    2.4827
                                P-value                0.1152
Time periods:           21         Distribution:            F(1,3295)
Avg Obs:                 166.05
Min Obs:                 160.00
Max Obs:                 168.00
```

Parameter Estimates

```
=====
Parameter  Std. Err.    T-stat    P-value    Lower CI    Upper CI
-----
wbg_i_rle   7.6555    4.8586    1.5756    0.1152    -1.8708    17.182
=====
```

```
F-test for Poolability: 5.8661
P-value: 0.0000
Distribution: F(190,3295)
```

Included effects: Entity, Time

Interpretation of results:

- For a given country, a 1 unit increase in the Rule of Law index over time is associated with a 7.6555% increase in FDI levels (as a % of GDP), net of time. This value is larger than our initial OLS model estimate, but it also comes with larger standard errors.
- The relationship between Rule of Law and FDI levels is no longer significant ($p\text{-value} > 0.1$), unlike our initial OLS model. After controlling for country-level and time-level fixed effects in the fixed effects model, the relationship between the two variables are no longer significant.
- This suggests that the initial relationship identified in simple OLS may have been spurious.
- One possibility is that the initial OLS model may have omitted some country-specific time-invariant factors that influence both Rule of Law and FDI levels. When these unobserved factors are controlled for in the fixed effects model, the previously significant relationship disappears.
- Another possibility is lack of within-country variation in Rule of Law over the time period studied. If `wbgi_rle` does not vary much over time within the same country, it becomes difficult to estimate its impact on FDI and could lead to higher standard errors as observed here. This higher standard errors than result in the loss of statistical significance in the relationship between Rule of Law and FDI levels.

(c) Include an additional predictor in your fixed effects model that you think might account for the initial relationship you found between your X and your Y. What effect does that new independent variable have in your new regression?

Additional predictor: Political corruption index (QoG Code: `vdem_corr`). This index runs from less corrupt (0) to more corrupt (1). The index is arrived at by taking the average of the public sector corruption index, executive corruption index, the indicator for legislative corruption, and the indicator for judicial corruption.

The level of political corruption could be a third variable that affects both the rule of law and FDI levels in a country:

- High political corruption likely negatively impacts the rule of law in a country. Corruption reduces the enforceability of contracts and fairness of dispute resolution mechanisms as judges or lawyers may be bribed to sway court decisions, and legislators may be bribed to unfairly change legislation for special interests.
- High political corruption also reduces FDI levels. Corruption increases the cost of doing business through bribery, bureaucratic red tape, and unofficial fees, which can deter

investors. Corruption also lowers the predictability of legal and regulatory environments, which deter investments.

Let us run the fixed effects model with this additional variable:

```
In [6]: # Fit the fixed effects model
fe_model2 = PanelOLS.from_formula(
    'wdi_fdiin ~ wbgi_rle + vdem_corr + EntityEffects + TimeEffects',
    data=df2
).fit(cov_type='clustered', cluster_entity=True)

# Display the summary of the regression
print(fe_model2.summary)
```

```

=====
                        PanelOLS Estimation Summary
=====
Dep. Variable:          wdi_fdiin    R-squared:                0.0065
Estimator:              PanelOLS     R-squared (Between):      -0.9245
No. Observations:       3487         R-squared (Within):       0.0043
Date:                   Tue, Oct 29 2024 R-squared (Overall):      -0.2522
Time:                   00:18:05     Log-likelihood            -1.468e+04
Cov. Estimator:         Clustered

                                F-statistic:                10.710
                                P-value                     0.0000
Entities:                171         Distribution:           F(2,3294)
Avg Obs:                  20.392
Min Obs:                   4.0000
Max Obs:                   21.000    F-statistic (robust):      1.2930
                                P-value                     0.2746
Time periods:             21         Distribution:           F(2,3294)
Avg Obs:                   166.05
Min Obs:                   160.00
Max Obs:                   168.00

```

```

=====
                        Parameter Estimates
=====

```

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
wbgi_rle	7.2827	4.6585	1.5633	0.1181	-1.8512	16.417
vdem_corr	-5.6280	4.1074	-1.3702	0.1707	-13.681	2.4254

```

=====

```

```

F-test for Poolability: 5.8630
P-value: 0.0000
Distribution: F(190,3294)

```

Included effects: Entity, Time

Interpretation of results:

- For a given country, a 1 unit increase in the political corruption index over time is associated with a -5.6280% decrease in FDI (as a % of GDP), net of the Rule of Law index. This is directionally in line with our initial hypothesis and expectation.

- However, the relation between the political corruption index and FDI is statistically insignificant (p-value>0.1). This suggests that political corruption does not significantly explain variations in FDI levels, net of rule of law.
- The coefficient on rule of law (`wbgi_rle`) decreases slightly compared to our previous fixed effects model (from 7.6555 to 7.2827). This coefficient is still statistically insignificant (p-value>0.1), suggesting that rule of law is still not significantly associated with FDI levels.
- A number of possibilities are still possible: that rule of law and FDI levels are truly not related after controlling for time and country-level effects, or that this is a data issue arising from lack of within-country variation in the rule of law variable that prevents precise estimates and produces statistically insignificant results.

(d) Run a random effects model equivalent to your fixed effects model. Interpret the results.

Next, we run a random effects model which assumes that the unobserved time and country-level heterogeneity is uncorrelated with the predictor or independent variables (`wbgi_rle` and `vdem_corr`).

```
In [7]: # Run the random effects model
re_model = RandomEffects.from_formula(
    'wdi_fdiin ~ wbgi_rle + vdem_corr',
    data=df2
).fit(cov_type='clustered', cluster_entity=True)

# Display the summary of the random effects model
print(re_model.summary)
```

RandomEffects Estimation Summary

Dep. Variable:	wdi_fdiin	R-squared:	0.0182
Estimator:	RandomEffects	R-squared (Between):	0.2449
No. Observations:	3487	R-squared (Within):	0.0027
Date:	Tue, Oct 29 2024	R-squared (Overall):	0.0749
Time:	00:18:05	Log-likelihood	-1.481e+04
Cov. Estimator:	Clustered		
		F-statistic:	32.275
Entities:	171	P-value	0.0000
Avg Obs:	20.392	Distribution:	F(2,3485)
Min Obs:	4.0000		
Max Obs:	21.000	F-statistic (robust):	63.884
		P-value	0.0000
Time periods:	21	Distribution:	F(2,3485)
Avg Obs:	166.05		
Min Obs:	160.00		
Max Obs:	168.00		

Parameter Estimates

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
wbgi_rle	4.8992	1.9064	2.5699	0.0102	1.1614	8.6369
vdem_corr	10.365	1.9286	5.3745	0.0000	6.5840	14.147

Interpretation of results:

- The **wbgi_rle** coefficient is halfway between the OLS and fixed effects estimate at 4.8992, and it is statistically significant (p-value<0.05). This is unlike the fixed effects model where the coefficient was not statistically significant.
- The **vdem_corr** coefficient is also now statistically significant (p-value<0.01). This is unlike the fixed effects model where the coefficient was not statistically significant. However, the coefficient is positive, suggesting that higher levels of corruption is associated with higher FDI levels, net of rule of law. This goes against our expectation.
- **Between-country variation drives the statistical significance:**
 - Random Effects leverage both between-country and within-country variation, while Fixed Effects only utilize within-country variation.
 - Hence, this result suggests that the relationship between rule of law and FDI is primarily driven by differences in the rule of law index across countries, rather than changes within each country over time.
- **Interpretation of Rule of Law's Impact on FDI:**
 - The lack of significance in the fixed effect model indicates that changes in the rule of law index within a given country over time do not strongly predict FDI in that country.

- This could be due to small changes in the rule of law within countries over the period where our data is available, or that FDI is less sensitive to incremental improvements in the rule of law within the same country.
- **Potential endogeneity in random effects model:**
 - Since the FE model controls for country-specific, time-invariant factors, the non-significant result could mean that the observed significant relationship in the RE model might be influenced by omitted factors that are correlated with rule of law and FDI but are constant within each entity over time (e.g., stable political institutions).
 - In that case, rule of law does not affect FDI, and they are only related due to confounding variables.

(e) Run a Hausman test to compare your fixed effects and your random effects models. What do you conclude?

To decide between the RE and FE models, a Hausman test can help determine if there's a systematic difference between the two models' estimates.

A significant result suggests that the RE model assumption (that unobserved country-specific effects are uncorrelated with the independent variables) is violated and the FE model is more appropriate. A non-significant result would support using the RE model, which produces more efficient estimates.

Firstly, let us produce a side-by-side comparison of the Random Effects and Fixed Effects models's key statistics to highlight their differences:

```
In [8]: # Compare the random and fixed effects models
result = compare({
    'Random Effects': re_model,
    'Fixed Effects': fe_model2
})
print(result)
```

Model Comparison

	Random Effects	Fixed Effects
Dep. Variable	wdi_fdiin	wdi_fdiin
Estimator	RandomEffects	PanelOLS
No. Observations	3487	3487
Cov. Est.	Clustered	Clustered
R-squared	0.0182	0.0065
R-Squared (Within)	0.0027	0.0043
R-Squared (Between)	0.2449	-0.9245
R-Squared (Overall)	0.0749	-0.2522
F-statistic	32.275	10.710
P-value (F-stat)	0.0000	0.0000
wbgi_rle	4.8992 (2.5699)	7.2827 (1.5633)
vdem_corr	10.365 (5.3745)	-5.6280 (-1.3702)
Effects		Entity Time

T-stats reported in parentheses

Now, let's build our own Hausman test:

```
In [9]: # Random Effects model
random_effects_model = RandomEffects.from_formula(
    'wdi_fdiin ~ wbgi_rle + vdem_corr', data=df2
).fit()

# Fixed Effects model
fixed_effects_model = PanelOLS.from_formula(
    'wdi_fdiin ~ wbgi_rle + vdem_corr + EntityEffects', data=df2
).fit()

# Extract the coefficients
b_fixed = fixed_effects_model.params
b_random = random_effects_model.params

# Extract the variance-covariance matrices
v_fixed = fixed_effects_model.cov
v_random = random_effects_model.cov

# Calculate the difference in coefficients
b_diff = b_fixed - b_random

# Calculate the variance of the difference
v_diff = v_fixed - v_random

# Hausman test statistic
hausman_stat = b_diff.T @ np.linalg.inv(v_diff) @ b_diff

# Degrees of freedom (number of coefficients being compared)
```

```

df = len(b_diff)

# Calculate p-value
p_value = 1 - stats.chi2.cdf(hausman_stat, df)

# Display the test statistic and p-value
print(f"Hausman test statistic: {round(hausman_stat,3)}")
print(f"P-value: {round(p_value,3)}")

```

Hausman test statistic: 7.328
P-value: 0.026

The Hausman Test Statistic is 7.328 (p-value<0.05), indicating that there is a statistically significant difference between the fixed effects and random effects models. This means that the assumption of the random effects model (that the unobserved heterogeneity is uncorrelated with the independent variables) is violated. The fixed effects model is more appropriate for this analysis.

In that case, as per our initial analysis of the fixed effects model, both rule of law (`wbgi_rle`) and political corruption (`vdem_corr`) are not significantly related to levels of foreign direct investment as a percentage of GDP (`wdi_fdiin`), after controlling for time and country-level fixed effects.