```
!pip install linearmodels
import pandas as pd
import statsmodels.api as sm
import statsmodels.formula.api as smf
from linearmodels.panel import PanelOLS
from linearmodels.panel import compare
```

```
Collecting linearmodels
  Downloading linearmodels-6.1-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (7.9 kB)
Requirement already satisfied: numpy<3,>=1.22.3 in /usr/local/lib/python3.10/dist-packages (from linearmodels) (1.26.4)
Requirement already satisfied: pandas>=1.4.0 in /usr/local/lib/python3.10/dist-packages (from linearmodels) (2.2.2)
Requirement already satisfied: scipy>=1.8.0 in /usr/local/lib/python3.10/dist-packages (from linearmodels) (1.13.1)
Requirement already satisfied: statsmodels>=0.13.0 in /usr/local/lib/python3.10/dist-packages (from linearmodels) (0.14.4)
Collecting mypy-extensions>=0.4 (from linearmodels)
  Downloading mypy_extensions-1.0.0-py3-none-any.whl.metadata (1.1 kB)
Requirement already satisfied: Cython>=3.0.10 in /usr/local/lib/python3.10/dist-packages (from linearmodels) (3.0.11)
Collecting pyhdfe>=0.1 (from linearmodels)
  Downloading pyhdfe-0.2.0-py3-none-any.whl.metadata (4.0 kB)
Collecting formulaic>=1.0.0 (from linearmodels)
  Downloading formulaic-1.0.2-py3-none-any.whl.metadata (6.8 kB)
Collecting setuptools-scm<9.0.0,>=8.0.0 (from setuptools-scm[toml]<9.0.0,>=8.0.0->linearmodels)
  Downloading setuptools_scm-8.1.0-py3-none-any.whl.metadata (6.6 kB)
Collecting interface-meta>=1.2.0 (from formulaic>=1.0.0->linearmodels)
  Downloading interface_meta-1.3.0-py3-none-any.whl.metadata (6.7 kB)
Requirement already satisfied: typing-extensions>=4.2.0 in /usr/local/lib/python3.10/dist-packages (from formulaic>=1.0.0->linearmodels)
Requirement already satisfied: wrapt>=1.0 in /usr/local/lib/python3.10/dist-packages (from formulaic>=1.0.0->linearmodels) (1.16.0)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.4.0->linearmodels) (2.8
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.4.0->linearmodels) (2024.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.10/dist-packages (from pandas>=1.4.0->linearmodels) (2024.2)
Requirement already satisfied: packaging>=20 in /usr/local/lib/python3.10/dist-packages (from setuptools-scm<9.0.0,>=8.0.0->setuptools-s
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from setuptools-scm<9.0.0,>=8.0.0->setuptools-scm[
Requirement already satisfied: tomli>=1 in /usr/local/lib/python3.10/dist-packages (from setuptools-scm<9.0.0,>=8.0.0->setuptools-scm[tc
Requirement already satisfied: patsy>=0.5.6 in /usr/local/lib/python3.10/dist-packages (from statsmodels>=0.13.0->linearmodels) (1.0.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2->pandas>=1.4.0->linearmc
Downloading linearmodels-6.1-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.7 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.7/1.7 MB 17.1 MB/s eta 0:00:00
Downloading formulaic-1.0.2-py3-none-any.whl (94 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 94.5/94.5 kB 6.2 MB/s eta 0:00:00
Downloading mypy_extensions-1.0.0-py3-none-any.whl (4.7 kB)
Downloading pyhdfe-0.2.0-py3-none-any.whl (19 kB)
Downloading setuptools_scm-8.1.0-py3-none-any.whl (43 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 43.7/43.7 kB 2.6 MB/s eta 0:00:00
Downloading interface_meta-1.3.0-py3-none-any.whl (14 kB)
Installing collected packages: setuptools-scm, mypy-extensions, interface-meta, pyhdfe, formulaic, linearmodels
Successfully installed formulaic-1.0.2 interface-meta-1.3.0 linearmodels-6.1 mypy-extensions-1.0.0 pyhdfe-0.2.0 setuptools-scm-8.1.0
```

```
url = 'https://www.qogdata.pol.gu.se/data/qog_bas_ts_jan24.xlsx'
data = pd.read_excel(url)
data.head()
```

	ccode	cname	year	ccode_qog	cname_qog	ccodealp	ccodecow	version	cname_year	ccodealp_year	...	wdi_trade	wdi_unem
0	4	Afghanistan	1946	4	Afghanistan	AFG	700.0	QoGBasTSjan24	Afghanistan 1946	AFG46	...	NaN	
1	4	Afghanistan	1947	4	Afghanistan	AFG	700.0	QoGBasTSjan24	Afghanistan 1947	AFG47	...	NaN	
2	4	Afghanistan	1948	4	Afghanistan	AFG	700.0	QoGBasTSjan24	Afghanistan 1948	AFG48	...	NaN	
3	4	Afghanistan	1949	4	Afghanistan	AFG	700.0	QoGBasTSjan24	Afghanistan 1949	AFG49	...	NaN	
4	4	Afghanistan	1950	4	Afghanistan	AFG	700.0	QoGBasTSjan24	Afghanistan 1950	AFG50	...	NaN	

5 rows × 251 columns

```
#Prepare the data by dropping rows with missing values in relevant columns
regression_data = data[['wdi_birth', 'wdi_unempyfilo', 'year']].dropna()
```

1. Due to the complexity of variables such as birth rate and female unemployment, there is a high chance that we will end up not finding anything significant or that there are significant confounding factors that we cannot see due to the simplicity of the OLS model.

```
#Run a naive OLS regression on your time series data. Tell me how you expect your Xs to affect your Y and why. Interpret your results.
#naive OLS regression with year fixed effects
ols_model = smf.ols(formula='wdi_unempyfilo ~ wdi_birth + C(year)', data=regression_data).fit()
ols_model.summary()
```

OLS Regression Results

Dep. Variable:	wdi_unempyfilo	**R-squared:**	0.050
Model:	OLS	**Adj. R-squared:**	0.044
Method:	Least Squares	**F-statistic:**	9.095
Date:	Mon, 18 Nov 2024	**Prob (F-statistic):**	5.08e-41
Time:	22:30:17	**Log-Likelihood:**	-21920.
No. Observations:	5430	**AIC:**	4.390e+04
Df Residuals:	5398	**BIC:**	4.411e+04
Df Model:	31		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	23.7271	1.217	19.495	0.000	21.341	26.113
C(year)[T.1992]	-0.6397	1.532	-0.418	0.676	-3.643	2.364
C(year)[T.1993]	0.2892	1.524	0.190	0.849	-2.698	3.277
C(year)[T.1994]	0.6213	1.524	0.408	0.684	-2.367	3.609
C(year)[T.1995]	0.8799	1.525	0.577	0.564	-2.109	3.869
C(year)[T.1996]	1.2269	1.525	0.805	0.421	-1.762	4.216
C(year)[T.1997]	1.0214	1.525	0.670	0.503	-1.969	4.011
C(year)[T.1998]	0.7866	1.526	0.516	0.606	-2.204	3.777
C(year)[T.1999]	0.9716	1.526	0.637	0.524	-2.020	3.963
C(year)[T.2000]	0.7015	1.526	0.460	0.646	-2.290	3.693
C(year)[T.2001]	0.7971	1.527	0.522	0.602	-2.196	3.790
C(year)[T.2002]	0.9987	1.525	0.655	0.512	-1.990	3.988
C(year)[T.2003]	1.4179	1.525	0.930	0.353	-1.572	4.407
C(year)[T.2004]	1.2677	1.525	0.831	0.406	-1.722	4.258
C(year)[T.2005]	0.8276	1.525	0.543	0.587	-2.163	3.818
C(year)[T.2006]	0.5769	1.522	0.379	0.705	-2.406	3.560
C(year)[T.2007]	-0.2288	1.522	-0.150	0.881	-3.212	2.755
C(year)[T.2008]	-0.3235	1.522	-0.213	0.832	-3.307	2.660
C(year)[T.2009]	1.0962	1.522	0.720	0.471	-1.888	4.080
C(year)[T.2010]	1.5755	1.522	1.035	0.301	-1.409	4.560
C(year)[T.2011]	1.7286	1.521	1.137	0.256	-1.252	4.709
C(year)[T.2012]	1.9738	1.521	1.298	0.194	-1.008	4.955
C(year)[T.2013]	2.1458	1.521	1.410	0.158	-0.837	5.128
C(year)[T.2014]	1.9029	1.522	1.251	0.211	-1.080	4.886
C(year)[T.2015]	1.5734	1.522	1.034	0.301	-1.411	4.558
C(year)[T.2016]	1.2746	1.523	0.837	0.403	-1.711	4.260
C(year)[T.2017]	0.7870	1.524	0.517	0.605	-2.200	3.774
C(year)[T.2018]	0.1637	1.524	0.107	0.914	-2.824	3.152
C(year)[T.2019]	-0.2728	1.525	-0.179	0.858	-3.262	2.716
C(year)[T.2020]	2.5970	1.526	1.702	0.089	-0.394	5.588
C(year)[T.2021]	1.5426	1.530	1.008	0.313	-1.457	4.542
wdi_birth	-0.2528	0.016	-15.502	0.000	-0.285	-0.221

Omnibus:	1021.562	**Durbin-Watson:**	0.093
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	1742.050
Skew:	1.237	**Prob(JB):**	0.00
Kurtosis:	4.257	**Cond. No.**	899.

As expected, none of the years show a significant relationship between birth rate and female unemployment at the 5% level. The R-squared is also quite small which further highlights the model's inadequacies as well as the lack of a clear relationship.

```
# Create first differences for the dependent and independent variables
regression_data['diff_wdi_unempyfilo'] = regression_data['wdi_unempyfilo'].diff()
regression_data['diff_wdi_birth'] = regression_data['wdi_birth'].diff()

# Drop the first observation of each group (since diff creates NaN for the first row)
regression_data_diff = regression_data.dropna()

# Run the first difference model (no need for year dummies; they are differenced out)
fd_model = smf.ols(formula='diff_wdi_unempyfilo ~ diff_wdi_birth', data=regression_data_diff).fit()

# Display the summary
```

```
fd_model.summary()

```

OLS Regression Results

Dep. Variable:	diff_wdi_unempyfilo	**R-squared:**	0.078
Model:	OLS	**Adj. R-squared:**	0.078
Method:	Least Squares	**F-statistic:**	457.4
Date:	Mon, 18 Nov 2024	**Prob (F-statistic):**	1.68e-97
Time:	22:42:54	**Log-Likelihood:**	-15554.
No. Observations:	5429	**AIC:**	3.111e+04
Df Residuals:	5427	**BIC:**	3.113e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	-0.0022	0.058	-0.039	0.969	-0.115	0.111
diff_wdi_birth	-0.3494	0.016	-21.387	0.000	-0.381	-0.317

Omnibus:	1907.823	**Durbin-Watson:**	2.019
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	510075.723
Skew:	-0.368	**Prob(JB):**	0.00
Kurtosis:	50.480	**Cond. No.**	3.53

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

After using first-differencing for the model, we see that the relationship between the two is now significant in contrast to the previous naive OLS model. This is to a degree an expected outcome because first-differencing may account for serial autocorrelation or seasonality that may present itself in the data. The coefficient has also increased by a small amount, indicating that as birth rate increases, female unemployment rate decreases by around -0.35. This isn't a substantial amount but it is interesting because I would have expected an increase in birth rate to be associated with less women entering the workforce or at least being delayed in their entry.