```
import pandas as pd
import statsmodels.formula.api as smf

# Load data set
df = pd.read_csv('swp_optimized.csv')

# Define and fit the model with state fixed effects
formula_a = 'employed ~ social_program + hsgrad + city + woman + C(region_id)'
model_a = smf.ols(formula=formula_a, data=df).fit()

# Print the summary of the model
print(model_a.summary())
```

C(region_id)[T.7]	0.0253	0.007	3.857	0.000	0.012	0.038
C(region_id)[T.8]	0.0521	0.007	7.718	0.000	0.039	0.065
C(region_id)[T.9]	−0.0150	0.007	−2.029	0.042	−0.029	−0.001
C(region_id)[T.10]	−0.0019	0.005	−0.386	0.700	−0.012	0.008
C(region_id)[T.11]	0.0519	0.006	8.342	0.000	0.040	0.064
C(region_id)[T.12]	0.0283	0.007	4.259	0.000	0.015	0.041
C(region_id)[T.13]	0.0487	0.006	7.700	0.000	0.036	0.061
C(region_id)[T.14]	0.0372	0.005	7.171	0.000	0.027	0.047
C(region_id)[T.15]	0.0430	0.006	6.946	0.000	0.031	0.055
C(region_id)[T.16]	0.0819	0.006	13.051	0.000	0.070	0.094
C(region_id)[T.17]	0.0471	0.006	7.407	0.000	0.035	0.060
C(region_id)[T.18]	−0.0038	0.006	−0.595	0.552	−0.016	0.009
C(region_id)[T.19]	−0.0211	0.007	−3.216	0.001	−0.034	−0.008
C(region_id)[T.20]	0.0096	0.006	1.489	0.137	−0.003	0.022
C(region_id)[T.21]	0.0508	0.007	7.724	0.000	0.038	0.064
C(region_id)[T.22]	0.0098	0.006	1.754	0.079	−0.001	0.021
C(region_id)[T.23]	0.0160	0.005	3.035	0.002	0.006	0.026
C(region_id)[T.24]	0.0829	0.006	13.115	0.000	0.071	0.095
C(region_id)[T.25]	−0.0014	0.006	−0.217	0.828	−0.014	0.011
C(region_id)[T.26]	0.0473	0.006	7.327	0.000	0.035	0.060
C(region_id)[T.27]	0.0019	0.007	0.298	0.766	−0.011	0.015
C(region_id)[T.28]	0.0911	0.006	14.459	0.000	0.079	0.103
C(region_id)[T.29]	0.0544	0.006	8.877	0.000	0.042	0.066
C(region_id)[T.30]	0.0645	0.007	9.615	0.000	0.051	0.078
C(region_id)[T.31]	0.0188	0.005	3.513	0.000	0.008	0.029
C(region_id)[T.32]	−0.0074	0.006	−1.205	0.228	−0.019	0.005
C(region_id)[T.33]	−0.0313	0.005	−6.362	0.000	−0.041	−0.022
C(region_id)[T.34]	0.0519	0.005	9.591	0.000	0.041	0.062
C(region_id)[T.35]	0.0461	0.006	7.188	0.000	0.034	0.059
C(region_id)[T.36]	0.0152	0.005	2.906	0.004	0.005	0.025
C(region_id)[T.37]	0.0026	0.006	0.409	0.682	−0.010	0.015
C(region_id)[T.38]	0.0221	0.006	3.405	0.001	0.009	0.035
C(region_id)[T.39]	−0.0106	0.005	−2.053	0.040	−0.021	−0.000
C(region_id)[T.40]	0.0119	0.007	1.820	0.069	−0.001	0.025
C(region_id)[T.41]	0.0346	0.006	5.363	0.000	0.022	0.047
C(region_id)[T.42]	0.0735	0.006	11.756	0.000	0.061	0.086
C(region_id)[T.43]	0.0272	0.006	4.237	0.000	0.015	0.040
C(region_id)[T.44]	0.0482	0.005	9.665	0.000	0.038	0.058
C(region_id)[T.45]	0.0722	0.006	11.262	0.000	0.060	0.085
C(region_id)[T.46]	0.0594	0.007	8.644	0.000	0.046	0.073
C(region_id)[T.47]	0.0489	0.006	7.979	0.000	0.037	0.061
C(region_id)[T.48]	0.0163	0.006	2.576	0.010	0.004	0.029
C(region_id)[T.49]	−0.0940	0.006	−14.998	0.000	−0.106	−0.082
C(region_id)[T.50]	0.0743	0.006	12.183	0.000	0.062	0.086
C(region_id)[T.51]	0.0505	0.007	7.655	0.000	0.038	0.063
social_program	−0.0367	0.001	−31.286	0.000	−0.039	−0.034
hsgrad	0.2723	0.001	242.616	0.000	0.270	0.274
city	0.0447	0.001	30.029	0.000	0.042	0.048
woman	−0.1311	0.001	−124.846	0.000	−0.133	−0.129

```
==============================================================================
Omnibus:                  4417537.097   Durbin-Watson:               1.997
Prob(Omnibus):                  0.000   Jarque-Bera (JB):        91439.248
Skew:                          -0.282   Prob(JB):                     0.00
Kurtosis:                       1.457   Cond. No.                     107.
==============================================================================

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

```
!pip install linearmodels
```

Requirement already satisfied: linearmodels in /usr/local/lib/python3.10/dist–packages (5.3)
Requirement already satisfied: numpy>=1.19.0 in /usr/local/lib/python3.10/dist–packages (from linearmodels) (1.23
Requirement already satisfied: pandas>=1.1.0 in /usr/local/lib/python3.10/dist–packages (from linearmodels) (1.5.
Requirement already satisfied: scipy>=1.5.0 in /usr/local/lib/python3.10/dist–packages (from linearmodels) (1.11.
Requirement already satisfied: statsmodels>=0.12.0 in /usr/local/lib/python3.10/dist–packages (from linearmodels)
Requirement already satisfied: mypy-extensions>=0.4 in /usr/local/lib/python3.10/dist–packages (from linearmodels
Requirement already satisfied: Cython>=0.29.34 in /usr/local/lib/python3.10/dist–packages (from linearmodels) (3.
Requirement already satisfied: pyhdfe>=0.1 in /usr/local/lib/python3.10/dist–packages (from linearmodels) (0.2.0)
Requirement already satisfied: formulaic>=0.6.5 in /usr/local/lib/python3.10/dist–packages (from linearmodels) (0
Requirement already satisfied: setuptools-scm[toml]<8.0.0,>=7.0.0 in /usr/local/lib/python3.10/dist–packages (fro
Requirement already satisfied: astor>=0.8 in /usr/local/lib/python3.10/dist–packages (from formulaic>=0.6.5–>line
Requirement already satisfied: interface-meta>=1.2.0 in /usr/local/lib/python3.10/dist–packages (from formulaic>=
Requirement already satisfied: typing-extensions>=4.2.0 in /usr/local/lib/python3.10/dist–packages (from formulai
Requirement already satisfied: wrapt>=1.0 in /usr/local/lib/python3.10/dist–packages (from formulaic>=0.6.5–>line
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist–packages (from pandas>=1.
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist–packages (from pandas>=1.1.0–>linea
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist–packages (from setuptools–scm[to
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist–packages (from setuptools-scm[toml]<8
Requirement already satisfied: tomli>=1.0.0 in /usr/local/lib/python3.10/dist–packages (from setuptools-scm[toml]
Requirement already satisfied: patsy>=0.5.2 in /usr/local/lib/python3.10/dist–packages (from statsmodels>=0.12.0–
Requirement already satisfied: six in /usr/local/lib/python3.10/dist–packages (from patsy>=0.5.2–>statsmodels>=0.

```python
import pandas as pd
from linearmodels import PanelOLS
# Equivalent to Stata's xtreg
import statsmodels.api as sm

# Set the index for panel data analysis
df = df.set_index(['region_id', 'year'])

# Define the model with state fixed effects
mod = PanelOLS.from_formula('employed ~ social_program + hsgrad + city + woman + EntityEffects', df)

# Fit the model
res = mod.fit()

# Print the summary of the model
print(res)
```

```
                         PanelOLS Estimation Summary
================================================================================
Dep. Variable:                employed   R-squared:                        0.0842
Estimator:                    PanelOLS   R-squared (Between):              0.3825
No. Observations:               812789   R-squared (Within):               0.0842
Date:                 Tue, Nov 21 2023   R-squared (Overall):              0.2528
Time:                         05:03:16   Log-likelihood                 -5.427e+05
Cov. Estimator:             Unadjusted
                                         F-statistic:                    1.867e+04
Entities:                           51   P-value                           0.0000
Avg Obs:                     1.594e+04   Distribution:                 F(4,812734)
Min Obs:                        6506.0
Max Obs:                      7.05e+04   F-statistic (robust):           1.867e+04
                                         P-value                           0.0000
Time periods:                       16   Distribution:                 F(4,812734)
Avg Obs:                     5.08e+04
Min Obs:                     3.991e+04
Max Obs:                     6.291e+04

                              Parameter Estimates
==============================================================================
                  Parameter  Std. Err.     T-stat    P-value    Lower CI   Upper CI
------------------------------------------------------------------------------
social_program     -0.0367     0.0012     -31.286     0.0000     -0.0390    -0.0344
hsgrad              0.2723     0.0011      242.62     0.0000      0.2701     0.2745
city                0.0447     0.0015      30.029     0.0000      0.0418     0.0476
woman              -0.1311     0.0010     -124.85     0.0000     -0.1331    -0.1290
==============================================================================

F-test for Poolability: 74.888
P-value: 0.0000
Distribution: F(50,812734)
```

```
    Included effects: Entity
```

```python
# Define the model with time fixed effects
mod = PanelOLS.from_formula('employed ~ social_program + hsgrad + city + woman + TimeEffects', df)

# Fit the model
res = mod.fit()

# Print the summary of the model
print(res)
```

```
                         PanelOLS Estimation Summary
    ================================================================================
    Dep. Variable:               employed   R-squared:                        0.0905
    Estimator:                   PanelOLS   R-squared (Between):              0.4981
    No. Observations:              812789   R-squared (Within):               0.0819
    Date:                Tue, Nov 21 2023   R-squared (Overall):              0.3192
    Time:                        05:03:49   Log-likelihood                   -5.42e+05
    Cov. Estimator:            Unadjusted
                                            F-statistic:                     2.021e+04
    Entities:                          51   P-value                           0.0000
    Avg Obs:                    1.594e+04   Distribution:                 F(4,812769)
    Min Obs:                       6506.0
    Max Obs:                    7.05e+04    F-statistic (robust):            2.021e+04
                                            P-value                           0.0000
    Time periods:                      16   Distribution:                 F(4,812769)
    Avg Obs:                    5.08e+04
    Min Obs:                    3.991e+04
    Max Obs:                    6.291e+04

                             Parameter Estimates
    ================================================================================
                  Parameter  Std. Err.     T-stat    P-value    Lower CI    Upper CI
    --------------------------------------------------------------------------------
    social_program   0.0094     0.0022     4.1962     0.0000      0.0050      0.0138
    hsgrad           0.2893     0.0011     253.60     0.0000      0.2871      0.2916
    city             0.0478     0.0014     33.788     0.0000      0.0450      0.0505
    woman           -0.1319     0.0010    -125.79     0.0000     -0.1340     -0.1299
    ================================================================================

    F-test for Poolability: 339.18
    P-value: 0.0000
    Distribution: F(15,812769)

    Included effects: Time
```

```python
# Define the model with both time and state fixed effects
mod = PanelOLS.from_formula('employed ~ social_program + hsgrad + city + woman + TimeEffects + EntityEffects', df)

# Fit the model
res = mod.fit()

# Print the summary of the model
print(res)
```

```
                         PanelOLS Estimation Summary
    ================================================================================
    Dep. Variable:               employed   R-squared:                        0.0889
    Estimator:                   PanelOLS   R-squared (Between):              0.4733
    No. Observations:              812789   R-squared (Within):               0.0825
    Date:                Tue, Nov 21 2023   R-squared (Overall):              0.3049
    Time:                        05:20:57   Log-likelihood                   -5.403e+05
    Cov. Estimator:            Unadjusted
                                            F-statistic:                     1.984e+04
    Entities:                          51   P-value                           0.0000
    Avg Obs:                    1.594e+04   Distribution:                 F(4,812719)
    Min Obs:                       6506.0
    Max Obs:                    7.05e+04    F-statistic (robust):            1.984e+04
                                            P-value                           0.0000
    Time periods:                      16   Distribution:                 F(4,812719)
    Avg Obs:                    5.08e+04
```

```
    Min Obs:                   3.991e+04
    Max Obs:                   6.291e+04

                          Parameter Estimates
    ====================================================================================
                Parameter  Std. Err.    T-stat    P-value    Lower CI    Upper CI
    ------------------------------------------------------------------------------------
    social_program   0.0014     0.0025    0.5541    0.5795    -0.0035      0.0063
    hsgrad           0.2880     0.0011    251.98    0.0000     0.2857      0.2902
    city             0.0429     0.0015    28.827    0.0000     0.0400      0.0458
    woman           -0.1313     0.0010   -125.44    0.0000    -0.1334     -0.1293
    ====================================================================================

    F-test for Poolability: 133.59
    P-value: 0.0000
    Distribution: F(65,812719)

    Included effects: Entity, Time
```