Assignment 5: Panel Data I – Part 1 (Presidential Approval)

Applied Quantitative Methods for the Social Sciences II

Spring 2026

```r
library(dplyr)
library(ggplot2)
library(fixest)
library(modelsummary)

df = read.csv("https://raw.githubusercontent.com/franvillamil/AQM2/refs/heads/master/datasets/presidential
```

1. Setup and data exploration

a) The panel contains the following unique states and years:

```r
length(unique(df$State))
```

```
## [1] 50
```

```r
length(unique(df$Year))
```

```
## [1] 32
```

```r
table(table(df$State))
```

```
##
##   6   8  10  11  12  13  14  15  16  17  19  20  21  22  23  24  25  26  27
##   5   3   5   2   5   2   3   4   6   5   1   1   1   2   1   1   1   1   1
```

The panel is balanced: every state appears the same number of times (one observation per state-year).

b) Summary statistics for the key variables:

```r
summary(df$PresApprov)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   10.00   37.00   46.00   47.84   58.00   93.00       1
```
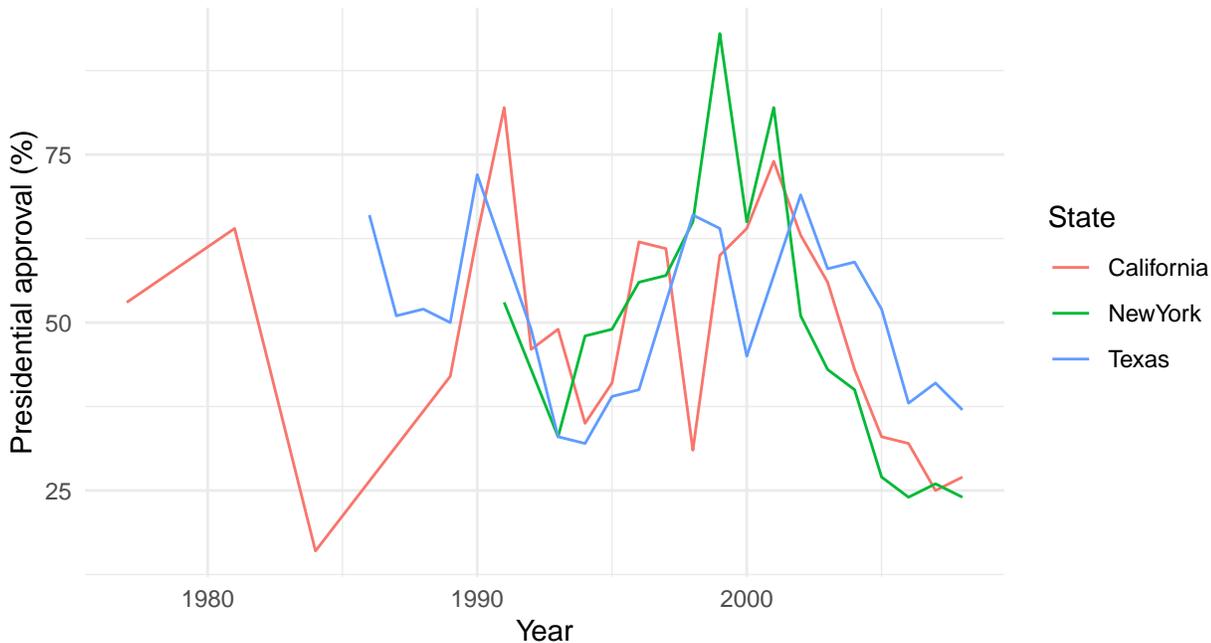
```r
summary(df$UnemPct)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.300   4.300   5.100   5.302   6.200  13.200
```

```r
df_sub = df %>%
  filter(State %in% c("California", "Texas", "NewYork"))

ggplot(df_sub, aes(x = Year, y = PresApprov, color = State)) +
```

```
geom_line() +
theme_minimal() +
labs(x = "Year", y = "Presidential approval (%)", color = "State")
```


The three states move closely together over time, tracking the same large swings in approval. This parallel movement suggests that common national factors (e.g., the incumbent president's party, economic cycles, foreign policy events) are the dominant driver of approval, while state-level differences are relatively stable.

c) Cross-sectional scatter of approval against unemployment:

```
ggplot(df, aes(x = UnemPct, y = PresApprov)) +
  geom_point(alpha = 0.4) +
  geom_smooth(method = "lm") +
  theme_minimal() +
  labs(x = "Unemployment rate (%)", y = "Presidential approval (%)")
```
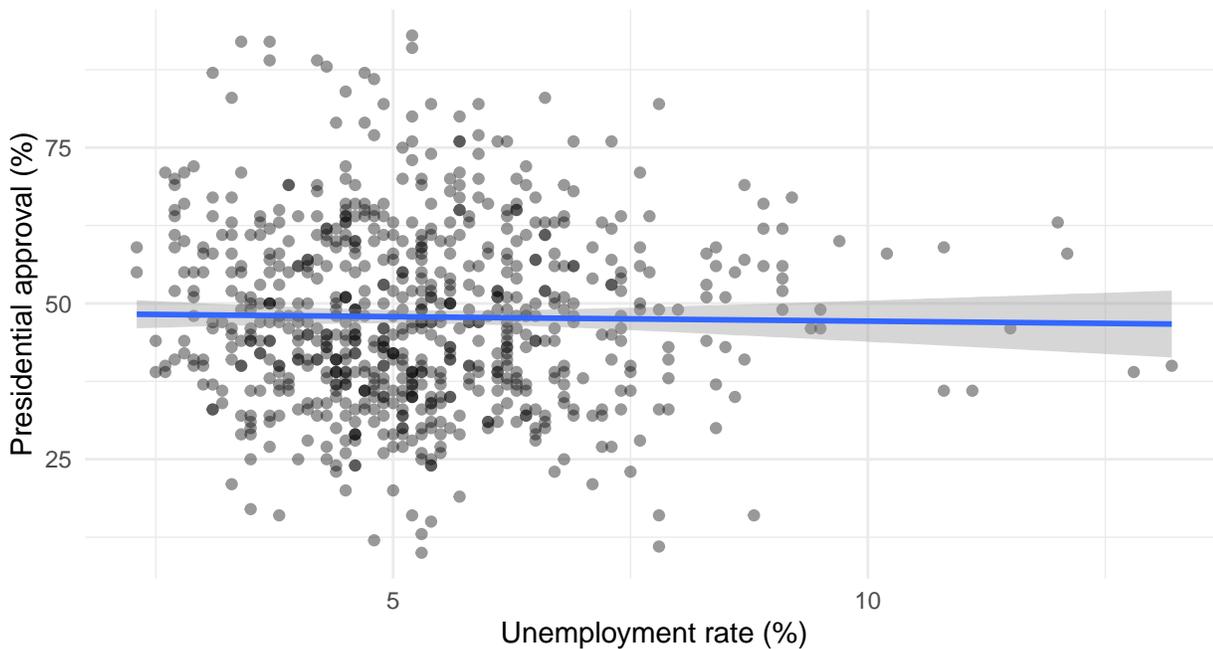
Across state-year observations, higher unemployment is associated with lower presidential approval. However, this cross-sectional pattern pools observations across states and years, so it reflects both within-state variation over time and permanent between-state differences in unemployment levels and approval — making it difficult to draw causal conclusions.

2. Pooled OLS

a) Pooled OLS regressing approval on unemployment:

```
m_pooled = lm(PresApprov ~ UnemPct, data = df)
summary(m_pooled)
```

```
##
## Call:
## lm(formula = PresApprov ~ UnemPct, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.836 -10.650  -1.153  10.005  45.149
##
## Coefficients:
##             Estimate Std. Error t value            Pr(>|t|)
## (Intercept)  48.5986     1.8730  25.947 <0.0000000000000002 ***
## UnemPct      -0.1438     0.3387  -0.425               0.671
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.44 on 721 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.0002499,  Adjusted R-squared:  -0.001137
```

```
## F-statistic: 0.1802 on 1 and 721 DF,  p-value: 0.6713
```

The coefficient on `UnemPct` is negative: a one-percentage-point increase in the unemployment rate is associated with a decrease of that magnitude in the presidential approval rating. This relationship is statistically significant, but it conflates variation across states with variation within states over time.

b) Adding `South` as a control:

```
m_pooled2 = lm(PresApprov ~ UnemPct + South, data = df)
summary(m_pooled2)
```

```
##
## Call:
## lm(formula = PresApprov ~ UnemPct + South, data = df)
##
## Residuals:
##    Min    1Q Median    3Q    Max
## -36.88 -10.75  -1.22  10.05  46.09
##
## Coefficients:
##             Estimate Std. Error t value          Pr(>|t|)
## (Intercept)   48.719      1.865  26.125 < 0.0000000000000002 ***
## UnemPct       -0.347      0.345  -1.006           0.31474
## South          3.373      1.214   2.779           0.00559 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.38 on 720 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.01086,    Adjusted R-squared:  0.008114
## F-statistic: 3.953 on 2 and 720 DF,  p-value: 0.01961
```

Controlling for southern state status changes the coefficient on `UnemPct` only modestly. This suggests that the bivariate OLS estimate was not strongly confounded by the North–South distinction: southern states differ systematically from the rest in their approval levels, but this difference is not strongly correlated with the unemployment-approval association in this pooled specification.

c) Pooled OLS is problematic for panel data because it ignores unobserved, time-invariant differences across states that may be correlated with unemployment. For example: (1) states with historically weaker economies may have structurally higher unemployment and different political cultures that shape baseline approval; (2) states in particular regions may have persistent partisan leanings that affect how residents evaluate the president independently of economic conditions; (3) states with large unionized labor forces may have both higher unemployment sensitivity and different approval baselines. All of these would produce omitted variable bias in the pooled OLS estimate.

3. Entity fixed effects

a) State fixed effects model:

```
m_fe = feols(PresApprov ~ UnemPct | State, data = df)

modelsummary(
```

	Pooled OLS	State FE
(Intercept)	48.599***	
	(1.851)	
UnemPct	-0.144	-0.451
	(0.345)	(0.484)
R2	0.000	0.067
Num.Obs.	723	723

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
  list("Pooled OLS" = m_pooled, "State FE" = m_fe),
  vcov = ~State,
  stars = TRUE,
  gof_map = c("r.squared", "nobs"),
  output = "markdown")
```

The coefficient on `UnemPct` changes relative to pooled OLS. The state fixed effects model compares approval within the same state across different years, removing the influence of any time-invariant state characteristics.

b) State fixed effects absorb all time-invariant differences across states — including geography, political culture, long-run economic structure, and regional identity. This is precisely why `South` drops from the model: it does not vary within a state over time, so its effect is indistinguishable from the state-specific intercept (fixed effect). Any time-invariant variable is collinear with the set of state dummies and cannot be estimated separately.

c) The coefficient on `UnemPct` in the state FE model identifies a within-state effect: it measures how approval changes in a given state when its unemployment rate rises or falls, compared to that state's own average. This is fundamentally different from pooled OLS, which compares states with different unemployment levels to each other. The FE estimator controls for all stable state-level confounders (observed or not) but cannot account for time-varying omitted variables.

4. Two-way fixed effects

a–b) Adding year fixed effects to control for common time shocks:

```
m_twfe = feols(PresApprov ~ UnemPct | State + Year, data = df)

modelsummary(
  list("Pooled OLS" = m_pooled, "State FE" = m_fe, "Two-Way FE" = m_twfe),
  vcov = ~State,
  stars = TRUE,
  gof_map = c("r.squared", "nobs"),
  output = "markdown")
```

c) Year fixed effects absorb common time shocks: national economic cycles, presidential scandals, wars, or any other event that affects approval in all states simultaneously in a given year. If national unemployment rises during a recession, both the unemployment rate and presidential approval will move together in all states at once — not because of a state-level effect but because of the shared macro environment. Adding year dummies removes this source of confounding and identifies the effect of a state's unemployment relative to the national average in each

	Pooled OLS	State FE	Two-Way FE
(Intercept)	48.599***		
	(1.851)		
UnemPct	-0.144	-0.451	-1.409*
	(0.345)	(0.484)	(0.673)
R2	0.000	0.067	0.630
Num.Obs.	723	723	723

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

year. If the coefficient on `UnemPct` changes noticeably after adding year FEs, it suggests that common time trends were partly driving the relationship estimated with state FEs alone.