

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

Applied Quantitative Methods for the Social Sciences II

Spring 2026

```
library(readstata13)
library(dplyr)
library(ggplot2)
library(fixest)
library(modelsummary)

df = read.dta13("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:

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

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

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

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

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

```
##
```

```
## 23 26 27 28 29 31 33 34 35 36 40 41 42 44 45 48 51 54 56 63
```

```
## 2 1 1 1 1 2 2 1 1 1 1 1 2 2 1 1 2 1 2 1
```

```
## 65 71 77 82 84 85 86 88 89 96 99 100 103 104 111 116 125 126 177 189
```

```
## 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1
```

```
## 209
```

```
## 1
```

The panel is unbalanced: while most states appear 73 times, a handful appear fewer times, so the number of time periods is not identical across all states.

b) Summary statistics for the key variables:

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

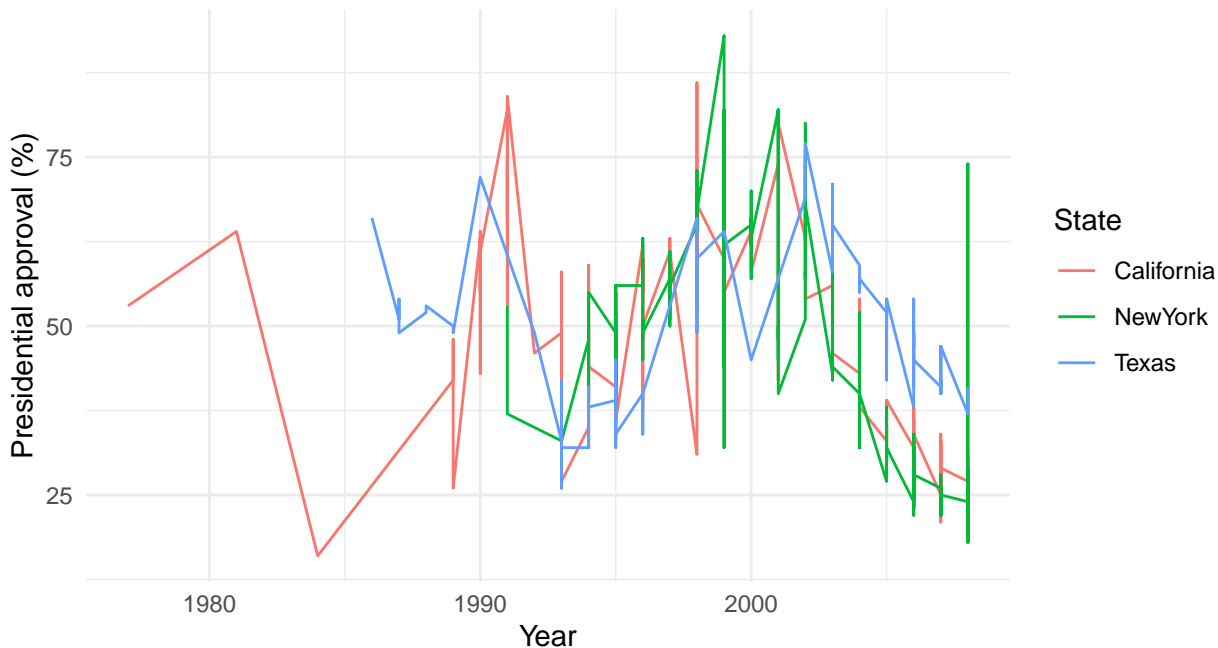
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      3.00   34.00   42.00   43.61   51.00   93.00     1
```

```
summary(df$UnemPct)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
## 2.300 4.300 4.900 5.131 5.900 13.200
```

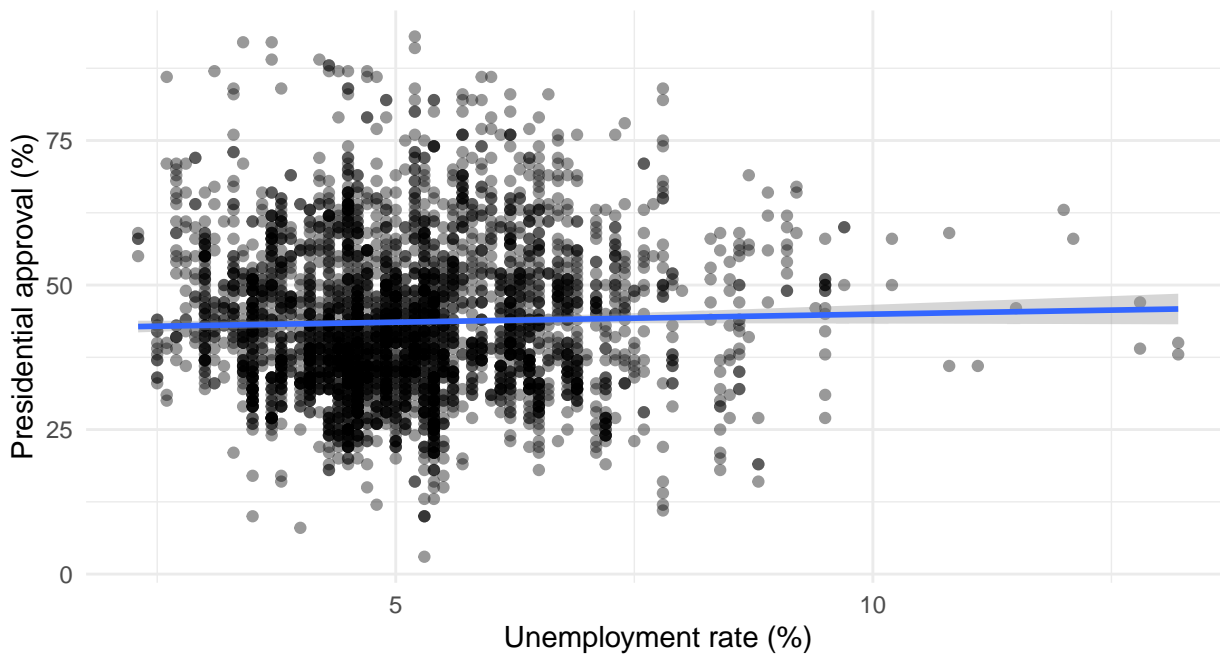
```
df_sub = df %>%  
  filter(State %in% c("California", "Texas", "NewYork"))  
  
ggplot(df_sub, aes(x = Year, y = PresApprov, color = State)) +  
  geom_line() +  
  theme_minimal() +  
  labs(x = "Year", y = "Presidential approval (%)", color = "State")
```



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
## -40.654  -9.320  -1.654   7.576  49.374
##
## Coefficients:
##              Estimate Std. Error t value      Pr(>|t|)
## (Intercept)  42.1773     0.8745  48.232 <0.0000000000000002 ***
## UnemPct       0.2786     0.1650   1.688    0.0914 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.97 on 3510 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.0008115, Adjusted R-squared:  0.0005269
```

```
## F-statistic: 2.851 on 1 and 3510 DF, p-value: 0.09142
```

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
## -40.015  -9.035  -1.915   7.766  50.005
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  41.9603     0.8722  48.108 < 0.0000000000000002 ***
## UnemPct       0.1990     0.1651   1.205     0.228
## South        2.7058     0.5193   5.210     0.0000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.92 on 3509 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.008482, Adjusted R-squared:  0.007917
## F-statistic: 15.01 on 2 and 3509 DF, p-value: 0.0000003232
```

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)	42.177*** (2.320)	
UnemPct	0.279 (0.447)	0.408 (0.609)
R2	0.001	0.083
Num.Obs.	3512	3512

+ 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)	42.177*** (2.320)		
UnemPct	0.279 (0.447)	0.408 (0.609)	-0.712 (0.792)
R2	0.001	0.083	0.670
Num.Obs.	3512	3512	3512
+ 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.