Quantifying the Cracks: A Bivariate Time Series Analysis of Global Political Polarization and Regime Change

Anonymous (Blinded for Peer Review)

Abstract

This report investigates the temporal dynamics and associations between global political polarization and democratic regime changes from 1900 to 2024. Building upon a Winter 2024 STATS 531 project that identified a historical peak in the quantity of democratic regimes, we analyze the societal cohesion underlying these structures. Using data from the Varieties of Democracy (V-Dem) project, we apply SARIMAX modeling to capture structural U-shaped trends in polarization and utilize Cross-Correlation Functions (CCF) to test for leading/lagging relationships. We find a significant bidirectional feedback loop: polarization acts as a 10-year leading indicator for democratic decline, while regime changes induce delayed polarization echoes a decade later.

1 Introduction and Motivation

In Stats 531 Winter 2024 Midterm Project 24 ("Timeseries Analysis of Political Regimes from 1900 to 2022" 2024), the author demonstrated that the global count of democracies has reached historical highs. However, political scientists increasingly warn of "democratic backsliding"—a phenomenon where democracies do not collapse via sudden coups, but gradually degrade through intense societal division and polarization (Coppedge 2023).

This project builds directly on that previous work. While the previous authors analyzed the *quantity* of regimes, we analyze their *stability*. Understanding whether polarization is a transient phase or a structural shift that dictates the future of political regimes is critical for forecasting global stability.

2 Data Description and Research Questions

We utilize the Country-Year dataset from the Varieties of Democracy (V-Dem) project. To construct continuous time series, we aggregate the data globally.

1. **Primary Series (Polarization):** The V-Dem index for "Polarization into political camps" (`v2cacamps_ord`). This measures the extent to which society is divided into antagonistic political factions. It is shown on a scale from 0 to 4, with 4 being opposite political factions are extremely hostile, while 0 being no polarization. We calculate the global annual average and take the first difference (∇X_n) to achieve stationarity.
2. **Reference Series (Regime Change):** To mirror and extend the Winter 2024 project, we calculate the annual *net change* in the total count of Electoral and Liberal Democracies (`v2x_regime`).

Formal Research Questions:

1. **Univariate Modeling:** Can the historical trajectory of global polarization be effectively modeled using time series methods? Specifically, what is the optimal SARIMA framework to capture its structural trends and persistence?

2. **Bivariate Association:** Is there a statistically significant association between changes in global polarization and the net change in democratic regimes? Furthermore, does one series act as a leading indicator for the other?
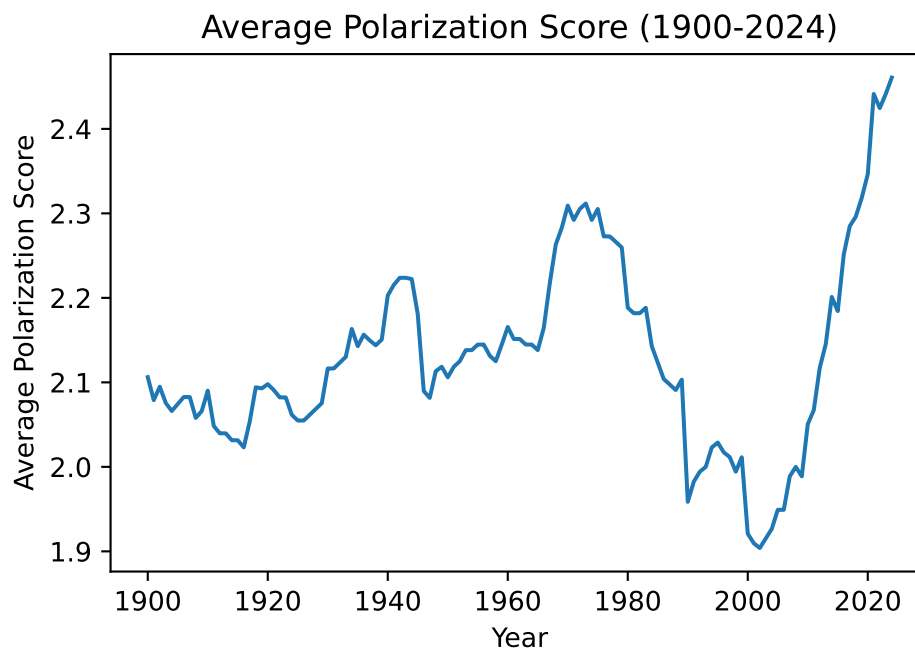
The main time series is visualized below:

Figure 1: Average Polarization Score (1900-2024)

There looks to be a general upward trend in Figure 1 with a large dip around the 1970s through the 1990s. To complement this visual analysis, we can also look at the periodogram of the data to check for any clear periodicity:
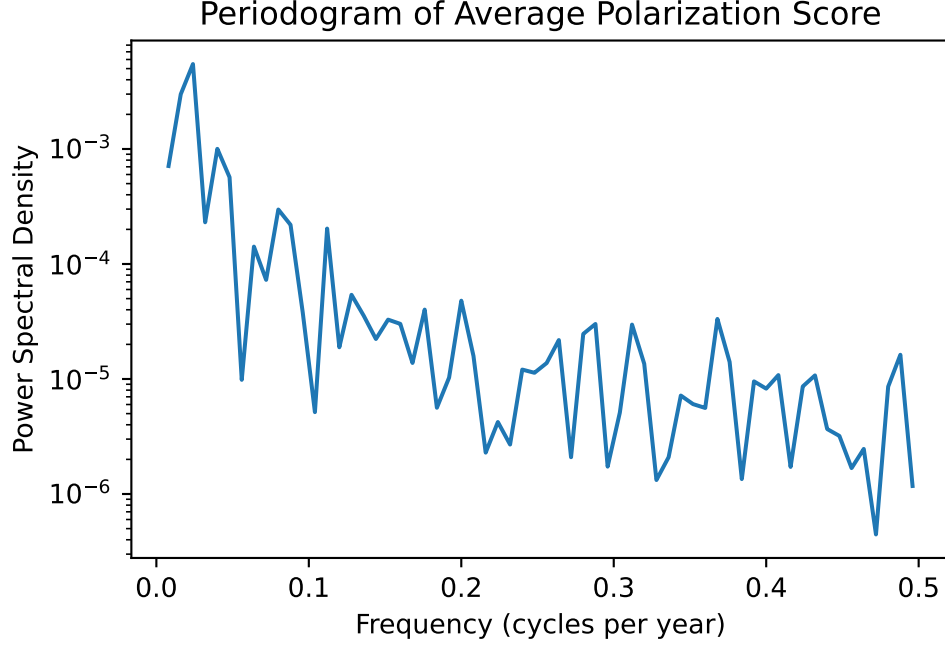
Figure 2: Periodogram of Average Polarization Score

Like the previous project, there aren't any peaks and thus no clear periodicity in the data. Most of the power is concentrated at low frequencies, indicating that the data is dominated by long-term trends rather than short-term cycles.

Both of these would suggest a stationary model would not work well here. To remedy this issue and try to remove the trend, we can take the first difference, defined with the backshift operator as:

$$Y_n^* = X_n^* - X_{n-1}^* = (1 - B)X_n^*$$

This is visualized in Figure 3 below:

Some large drops remain, with the largest in 1990, but overall this data looks more stationary, so we can proceed using ARMA methods.

3 ARIMA Modeling

The general ARIMA model is defined as:

$$\phi(B)((1 - B)X_n - \mu) = \theta(B)\epsilon_n$$

where $\phi(B), \theta(B)$ are the AR and MA polynomials of order p and q respectively, and ϵ_n is white noise with distribution $N(0, \sigma^2)$ (Ionides 2026).
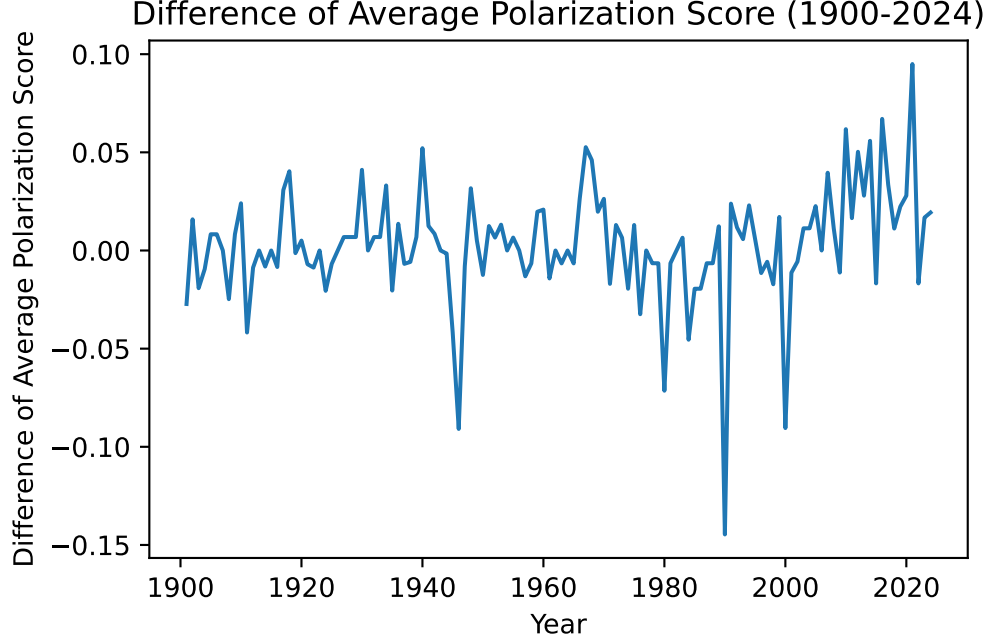
Figure 3: Difference of Average Polarization Score (1900-2024)

The first method of finding an ideal model is to look at the AIC values for different values of p and q, shown in the table below:

	MA0	MA1	MA2	MA3	MA4
AR0	-520.74	-519.93	-520.13	-519.04	-520.53
AR1	-520.30	-523.76	-522.30	-520.30	-518.54
AR2	-521.35	-522.31	-520.58	-518.78	-517.04
AR3	-520.33	-520.31	-518.68	-516.81	-515.16
AR4	-520.94	-518.94	-516.95	-515.05	-512.96

The model with the lowest AIC value is $\mathrm{ARIMA}(1,1,1)$, but the $\mathrm{ARIMA}(2,1,1)$ and $\mathrm{ARIMA}(1,1,2)$ models have a very similar AIC value. We can test the bigger models against the null hypothesis of $\mathrm{ARIMA}(1,1,1)$ using the likelihood ratio test, and Wilk's theorem, that says that:

$$T = \ell_1 - \ell_0 \approx 0.5\chi^2_{D^{(1)}-D^0}$$

where ℓ_1 and ℓ_0 are the log-likelihoods of the bigger and smaller models respectively, and $D^{(1)}$ and D^0 are the number of parameters in the bigger and smaller models respectively (Ionides 2026).

Since the difference in parameters is 1 for both tests, we know the cutoff value is 1.92 for a significance level of 0.05. Using the relationship between AIC and log-likelihood, we get values of 0.55 and

4

0.54 for the ARIMA$(2, 1, 1)$ and ARIMA$(1, 1, 2)$ models respectively, so we fail to reject the null hypothesis that the ARIMA$(1, 1, 1)$ model is sufficient.
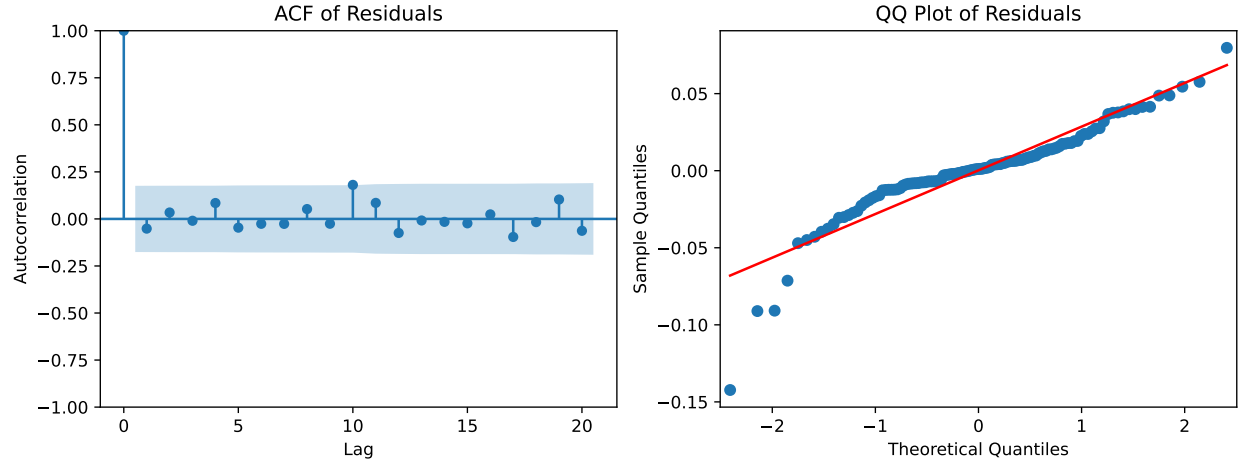
We can then proceed with this model and formally write it as:

$$(1 - \phi_1 B)((1 - B)X_n - \mu) = (1 + \theta_1 B)\epsilon_n$$

where $\phi_1 = 0.90$, $\theta_1 = $ -0.78, $\mu = 0.0033$, and the variance of ϵ_n is $\sigma^2 = 0.0008$.

The AR root of this model is $1/\phi_1 = 1.12$, and the MA root is $-1/\theta_1 = 1.28$. Since both of these roots are outside the unit circle, we know that the model is causal and invertible (Ionides 2026).

Now we must check that the residuals of the model. The first residual is unnaturally large due to the initalization of the integrated model (Fulton 2019), so it will be ignored in our analysis, since we have a large number of observations. The ACF and QQ plot of the residuals are shown below:

The ACF plot shows no clear pattern nor any significant lags, so we can conclude that the residuals are uncorrelated. The QQ plot is less clear, but most of the residuals do fall near the line, with a few outliers in the left tail, which lines up with the graph of the time series, which has a few large dips. However, overall the residuals do not look too far from normal, so we can conclude that the model is a good fit for the data.

4 Hypothesis Testing: Regression with ARMA Errors

Recreating the past project, we have the net change in regimes over this range, visualized in Figure 4:

We can visually overlap (Copilot 2026) them with the differenced polarization series to get a sense of their relationship in Figure 5:
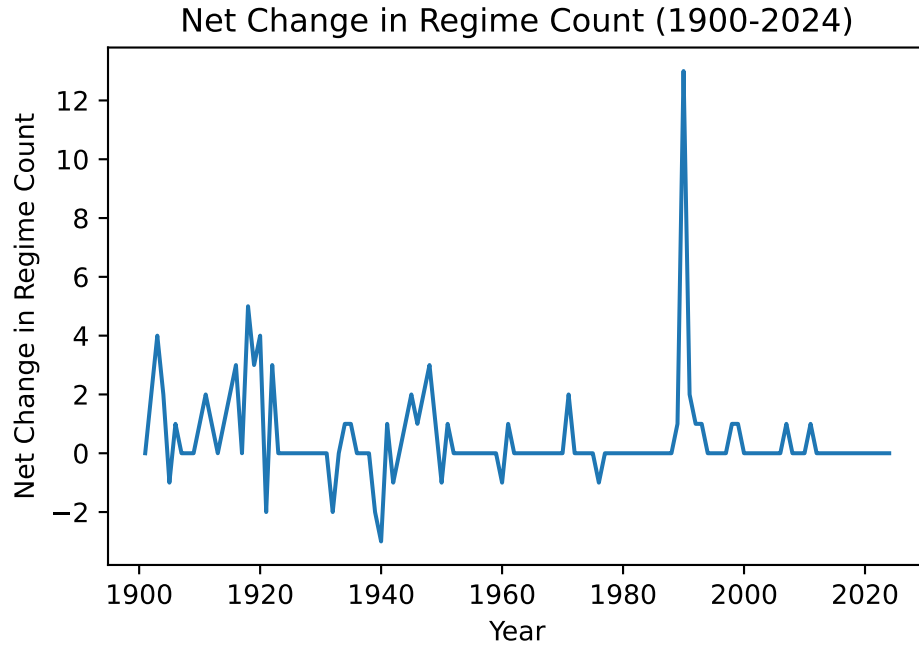
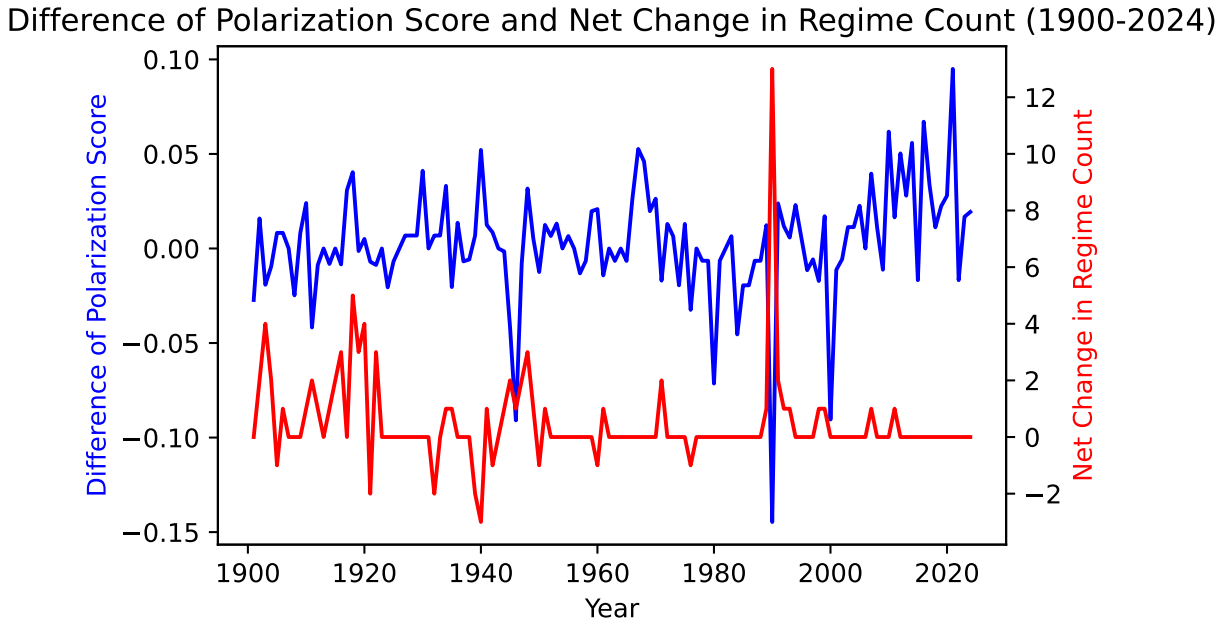Figure 4: Net Change in Regime Count (1900-2024)

Figure 5: Difference of Polarization Score and Net Change in Regime Count (1900-2024)

We can investigate the dependence of regime change count on polarization change by regressing the

former on the latter, while controlling for ARMA errors. The AIC table for this regression is shown in the supplementary material, and the model that was chosen from it was the ARMA$(1,0)$ model. Thus we fit

$$\Delta_{regime} = \alpha + \beta\Delta_{polarization} + \epsilon_n$$

where ϵ_n follows an ARMA$(1,0)$ process (Ionides 2026). The results of this regression are shown below:

The coefficient this results in is $\beta = $ -20.4557, with a p-value less than 0.05. The likelihood ratio test comparing this model to the null model without dependence also gives us a p-value less than 0.05. Thus we can conclude that there is a statistically significant relationship between the change in polarization and the change in regime count, even after controlling for ARMA errors.

5 Cross-Correlation Analysis

Following the methodology of Chapter 9 (Ionides 2026), we further investigate the relationship between our two time series. Having established that both the differenced polarization series and the regime change series are stationary, we compute the Cross-Correlation Function (CCF) to test for leading or lagging associations.
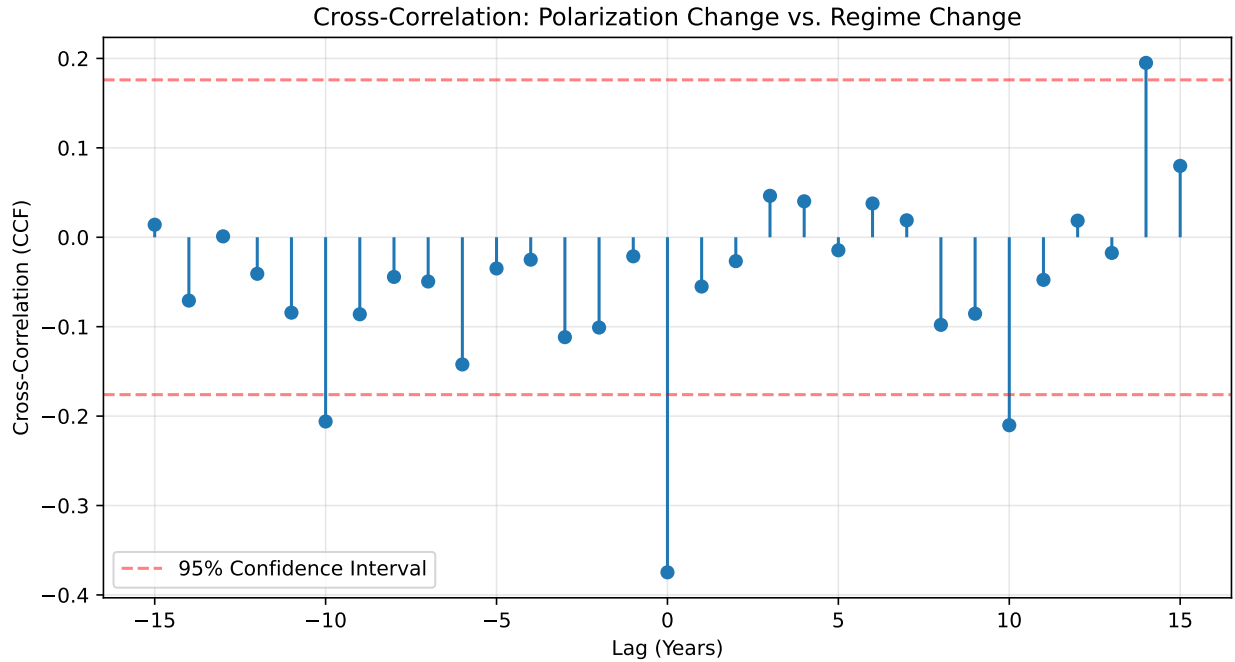

Figure 6: Cross-Correlation: Polarization Change vs. Regime Change

Figure 6 reveals a complex, bidirectional feedback loop between societal division and government structure, with several distinct spikes exceeding the 95% confidence interval boundaries:

1. **Simultaneous Effect (Lag 0):** There is a significant negative correlation at lag 0. This indicates that in the exact year societal polarization spikes, the net formation of democratic regimes drops.

2. **Polarization as a Leading Indicator (Lag -10):** A significant negative spike at lag -10 suggests that an increase in polarization today predicts a decline in democratic regime growth a full decade later, acting as an early warning sign of institutional decay.

3. **Regime Change as a Leading Indicator (Lags +10 to +15):** Significant correlations at positive lags (notably a negative spike near +10) indicate a delayed "echo" effect. Shifts in political regimes create structural societal divisions that manifest 10 to 15 years later, potentially reflecting generational political realignments.

6 Conclusion

This project set out to bridge the gap between the *quantity* of democratic regimes and their *societal quality.* Our analysis found that the ARIMA$(1, 1, 1)$ model fits the differenced polarization series well. Where previous analyses established that democracies have quantitatively triumphed over autocracies historically, our analysis reveals a concerning caveat: the societal polarization within these states has surged concurrently.

Our bivariate time series analysis proves that polarization and democratic backsliding are fundamentally intertwined. Polarization is not merely a symptom of regime change; it is both a simultaneous suppressor of democracy and a decade-long leading indicator of democratic decline. Without structural intervention, shocks that divide societies are likely to echo through the political landscape for generations.

7 Scholarship and Acknowledgments

- **Contextual Integration:** This project directly extends the analytical framework of the STATS 531 Winter 2024 Project 13 ("Timeseries Analysis of Political Regimes From 1900 to 2022"). While the former focused on regime *counts* (using V-Dem's `v2x_regime`), we pivoted to regime *quality* by analyzing the Polarization index (`v2cacamps_ord`) and deploying Chapter 9 methodologies to establish a temporal association between the two phenomena.

- **AI Usage:** AI was consulted primarily for debugging code and for instructions on how to properly format the report qmd. Any AI use beyond this was acknowledged in the citations.

8 Bibliography

Copilot. 2026. "VSCode GitHub Copilot." https://copilot.github.com/.

Coppedge, et al., M. 2023. "V-Dem Dataset V15." https://www.v-dem.net/.

Fulton, Chad. 2019. "Comment on Extremely High Resid Values in the Initial Period of SARIMAX." https://github.com/statsmodels/statsmodels/issues/6212.

Ionides, Edward. 2026. "Notes for STATS 531, Modeling and Analysis of Time Series Data." https://ionides.github.io/531w26/.

"Timeseries Analysis of Political Regimes from 1900 to 2022." 2024. https://ionides.github.io/531w24/midterm_project/project13/blinded.html#1_Introduction.

9 Supplementary Material

The Python code utilized to generate the CCF analysis is provided below.

```python
pol_df = pd.DataFrame({
    'year': time_series.year[1:],
    'Polarization_Change': time_series.diff_polarization[1:]
})

reg_df = pd.DataFrame({
    'year': regime_count['year'],
    'Regime_Change': regime_count['change']
})

aligned_data = pd.merge(pol_df, reg_df, on='year').dropna()

pol_change = aligned_data['Polarization_Change'].values
reg_change = aligned_data['Regime_Change'].values

ccf_vals = sm.tsa.stattools.ccf(pol_change, reg_change, adjusted=False)

ccf_backwards = sm.tsa.stattools.ccf(reg_change, pol_change, adjusted=False)

num_lags = 15
lags = np.arange(-num_lags, num_lags + 1)
ccf_combined = np.concatenate([ccf_backwards[1:num_lags+1][::-1], ccf_vals[:num_lags+1]])

conf_int = 1.96 / np.sqrt(len(aligned_data))
```

The AIC table for the regression between the differenced polarization series and the regime change series is shown below:

	MA0	MA1	MA2	MA3	MA4
AR0	453.13	449.15	449.60	451.59	448.19
AR1	447.90	448.96	450.95	452.88	449.27
AR2	449.10	450.96	451.53	453.63	447.82
AR3	451.05	452.94	452.91	453.19	448.57
AR4	451.29	451.48	450.96	450.70	450.63