ELECTION MODELLING

DYNAMIC ELECTORAL MODEL USING BIOLOGICALLY INSPIRED PRINCIPLES

Alan Royce Gabriel
Shivan Ajay Iyer

BACKGROUND

Traditional Approaches

- Agent-based simulations
- Opinion dynamics models (DeGroot, voter models)
- Statistical models (historical data)

Limitations:

- High-dimensional parameters
- Low analytical tractability

A New Perspective

Inspired by computational systems biology:

- SIR model (disease spread)
- Lotka-Volterra equations (predator-prey dynamics)

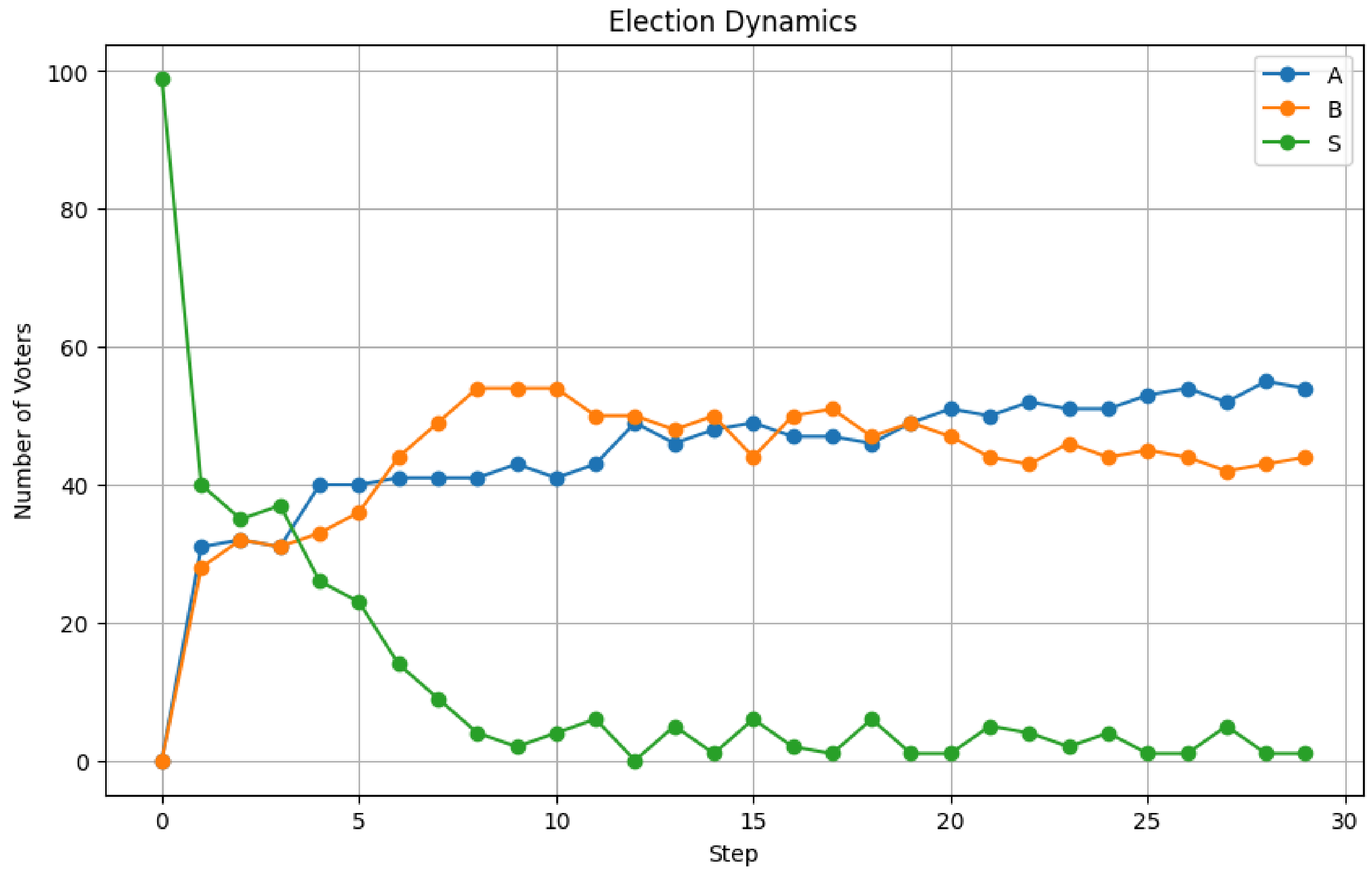
  → Capture complex, nonlinear population interactions elegantly.

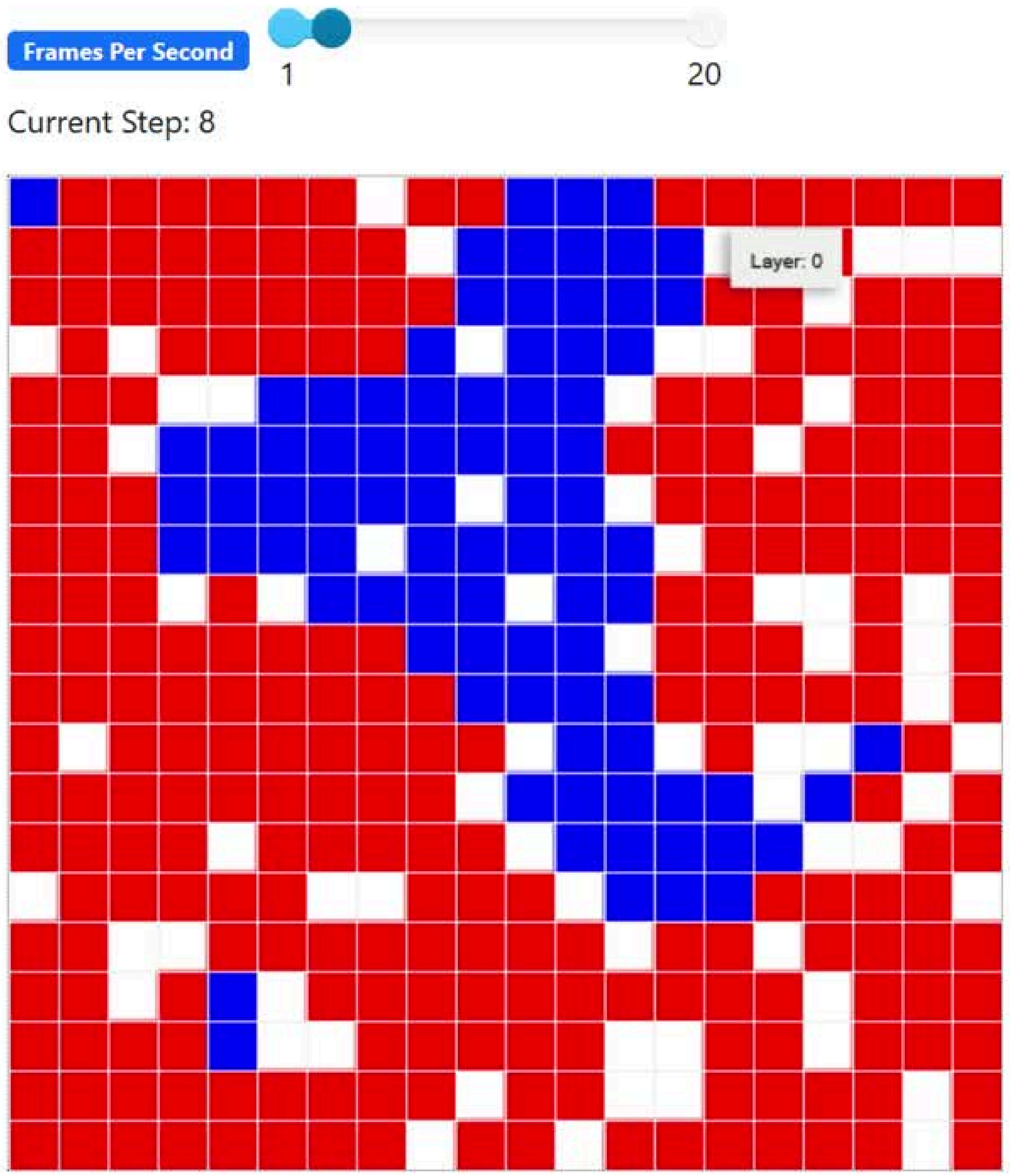
PROJECT GOALS

- Apply biological frameworks to model electoral dynamics
- Treat campaigning as a contagious process (opinion spread ~ disease transmission)
- Use predator-prey models to capture candidate-voter competition
- Simulate vote spread and stabilization over time
- Analyse candidate strategies and vote share dynamics
- Compare biologically-inspired models with traditional political models
- Demonstrate the power and interpretability of biological analogies for election modeling

AGENT BASED MODELLING

- Election dynamics with voter, candidate, and media agents.
- Opinion dynamics using a DeGroot-style averaging process.





ARIMA

- ARIMA(p, d, q)
  - where p is the order of the autoregressive (AR) component, d is the degree of differencing needed to make the time series stationary
  - q is the order of the moving average (MA) component

$$\phi(B)(1-B)^d y_t = \theta(B)\varepsilon_t$$

B is the backshift operator, ie,

$$By_t = y_{t-1}$$

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$$ AR polynomial

$$\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$$ MA polynomial

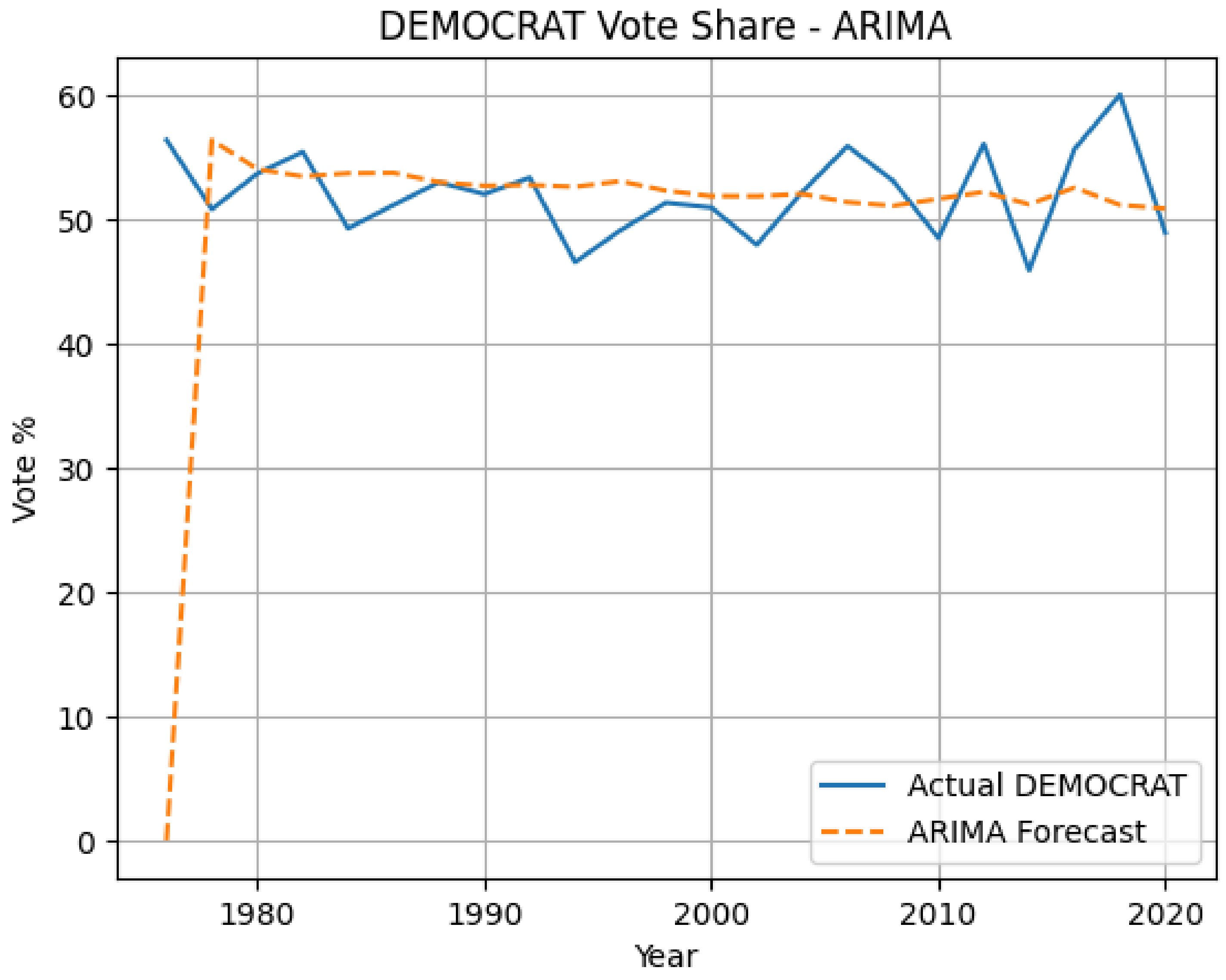
ARIMA(1, 1, 1)

$$(1-B)y_t = \phi_1(1-B)y_{t-1} + \varepsilon_t + \theta_1\varepsilon_{t-1}$$

$$y(t) = -0.1582 + (-0.1582) \cdot y(t-1) + (-0.9997) \cdot e(t-1)$$

ARIMA

ARIMA(p, d, q)



- **Autoregressive** (AR): Uses the dependency between an observation and some number of lagged observations.
- **Integrated** (I): Uses differencing of raw observations to make the time series stationary.
- **Moving Average** (MA): Uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.
- **Univariate**: Models a single time series independently.
- Assumptions: Requires data stationarity (or achieving it through differencing), assumes linear relationships

PINN

$$\frac{dy(t)}{dt} = f(t, y(t)) + \epsilon(t)$$

y(t) is the vote share at time t

f(t, y(t)) represents the dynamic function governing the evolution of support

ϵ(t) is the error term or residual capturing the discrepancy between the data and the model

Data Loss

$$L_{\text{data}} = \sum_i \left(y(t_i) - \hat{y}(t_i)\right)^2$$

Physics Loss

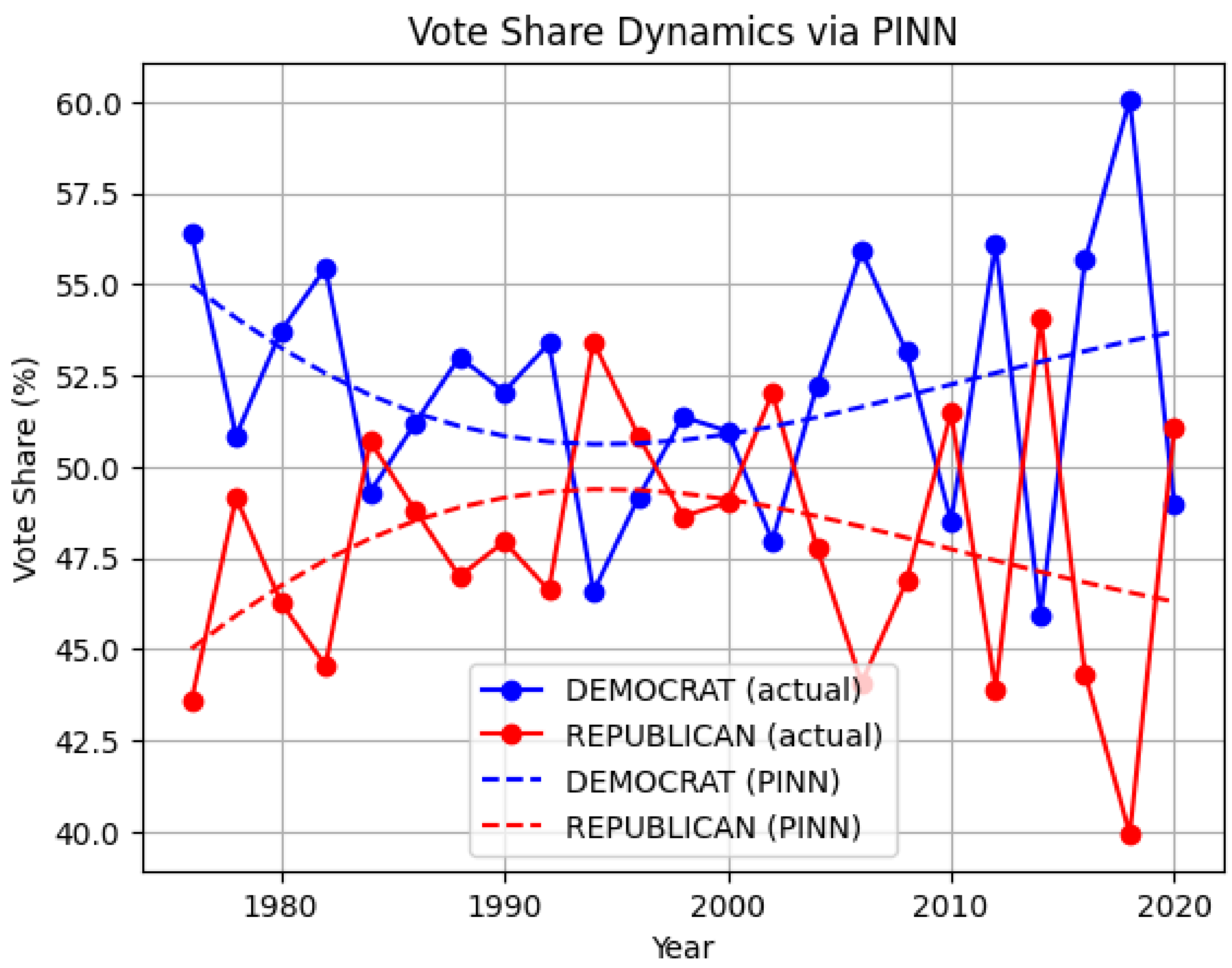
$$L_{\text{physics}} = \sum_j \left(\frac{dy(t_j)}{dt} - f(t_j, y(t_j))\right)^2$$

Total loss function for a PINN

$$L_{\text{PINN}} = L_{\text{data}} + \lambda L_{\text{physics}}$$

λ is a weight parameter that balances fitting the data versus respecting the dynamics

PINN



- **Hybrid Model**: Integrates data fitting (via NN) and physics/mechanism constraints (via ODEs in the loss).
- **Neural Network**: Uses a feed-forward neural network to map time t to vote shares [D, R].
- **Automatic Differentiation**: Leverages PyTorch's autograd to compute derivatives needed for the "physics" part of the loss function.
- **Physics Loss**: Incorporates terms that measure how well the NN output satisfies the assumed differential equations

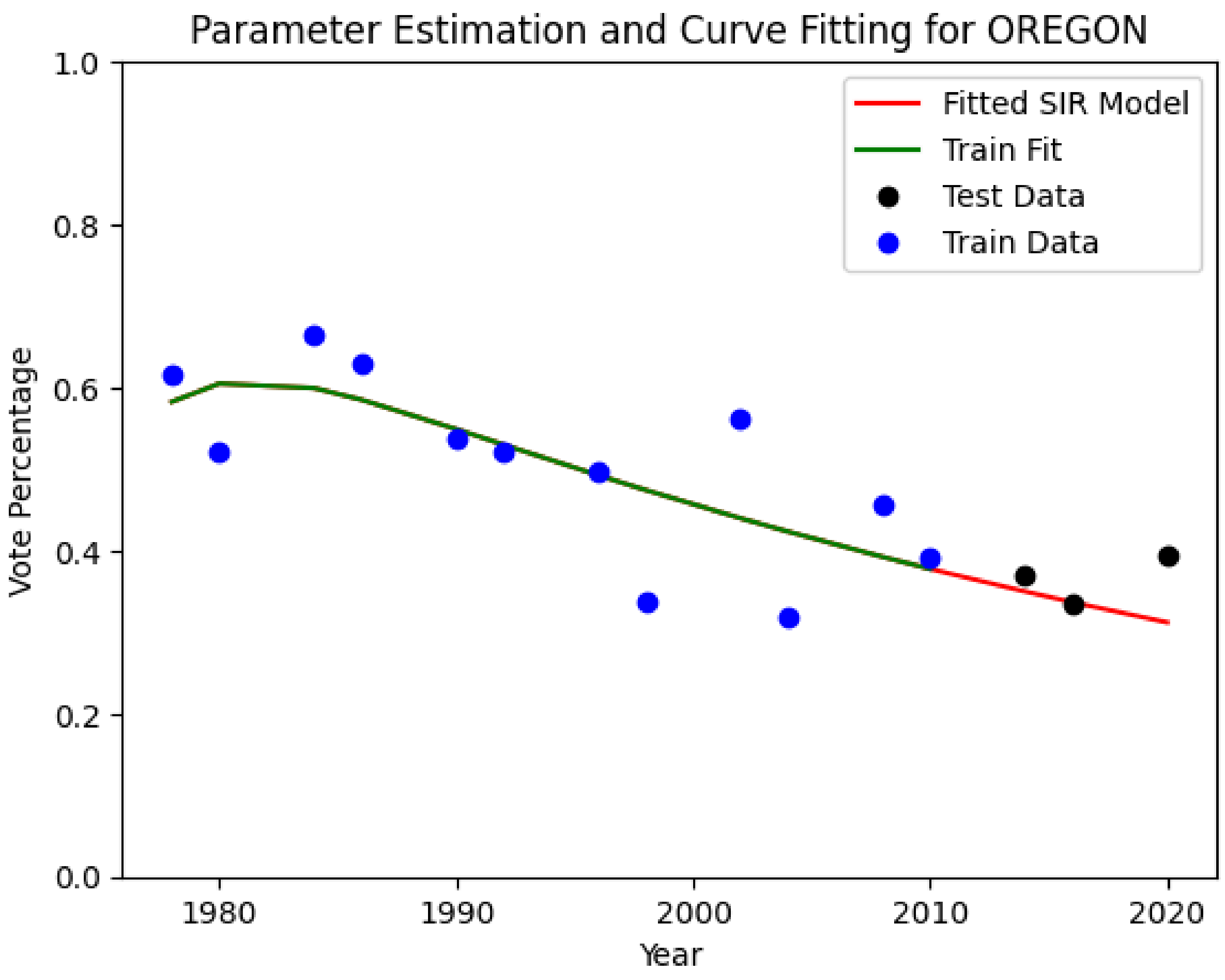
Model Comparision

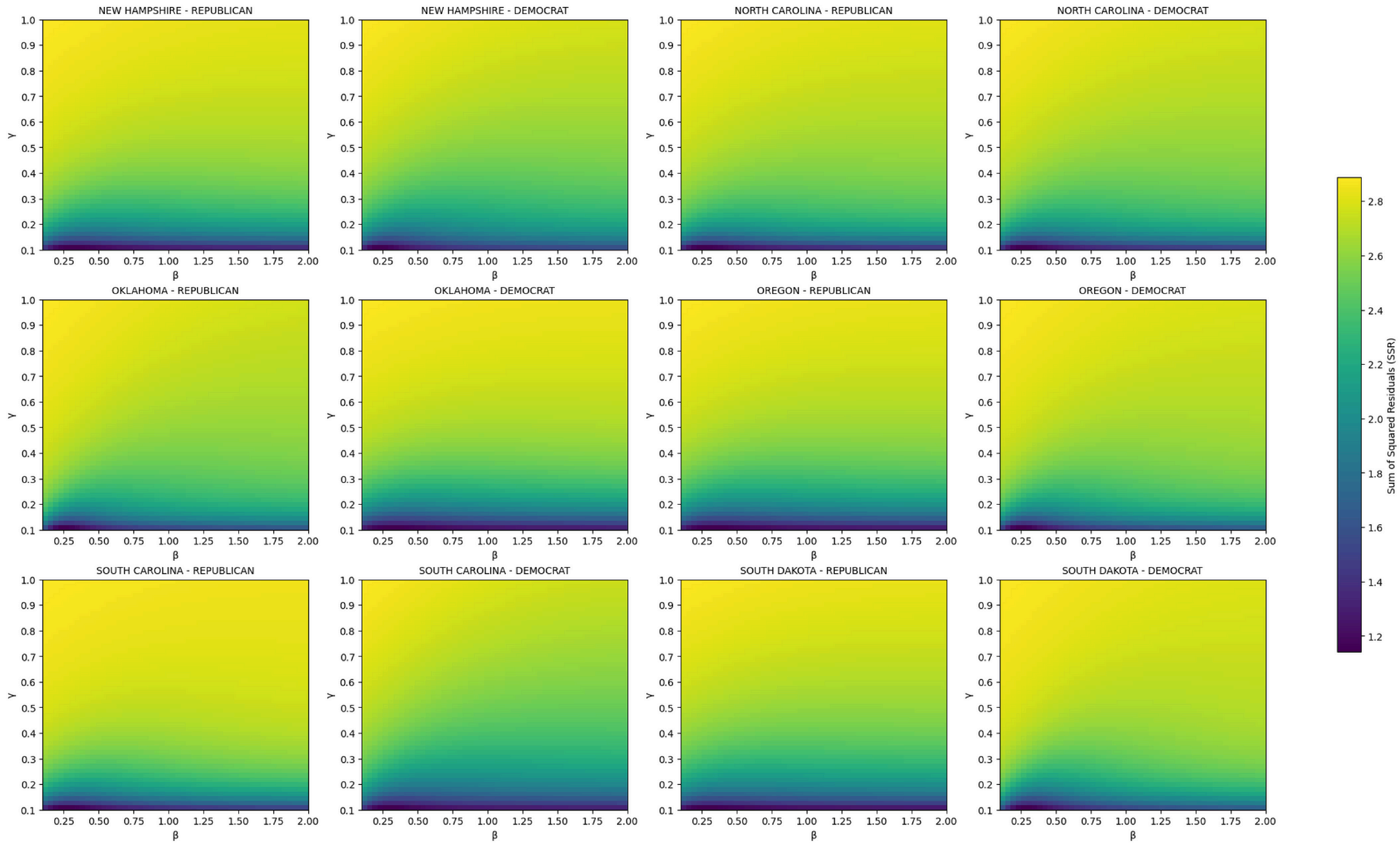
Feature	ARIMA	PPM (Lotka-Volterra)	PINNs	ABM
Approach	Statistical Time Series	ODE-based (Analogy)	Hybrid (NN + ODE)	Bottom-Up Simulation
Level	Aggregate (Univariate)	Aggregate (Multivariate)	Aggregate (Multivariate)	Micro/Individual
Mechanism	No (Data-driven)	Yes (Assumed ODE)	Yes (Assumed ODE)	Yes (Agent Rules)
Interactions	No (Implicit in lags)	Yes (Explicit in ODEs)	Yes (Explicit in ODEs)	Yes (Explicit Agent Rules)
Heterogeneity	No	No	No	Yes (Agent Attributes)
Interpretability	High (Stats)	Medium (ODE Params)	Low (NN Weights)	Medium (Rules → Emergence)
Flexibility	Low	Medium	Medium	High
Data Needs	Time Series Data	Time Series Data (Fit)	Time Series Data (Fit+Loss)	Rules/Params + Calibration
Advantage	Simple, Baseline	Models Interaction, Theory	Handles Non-linearity, Data	Heterogeneity, Emergence
Disadvantage	Ignores Interaction, Linear	Analogy? Homogeneity	Complex, Needs ODE, Debug?	Complex, Calibration, Costly

SIR

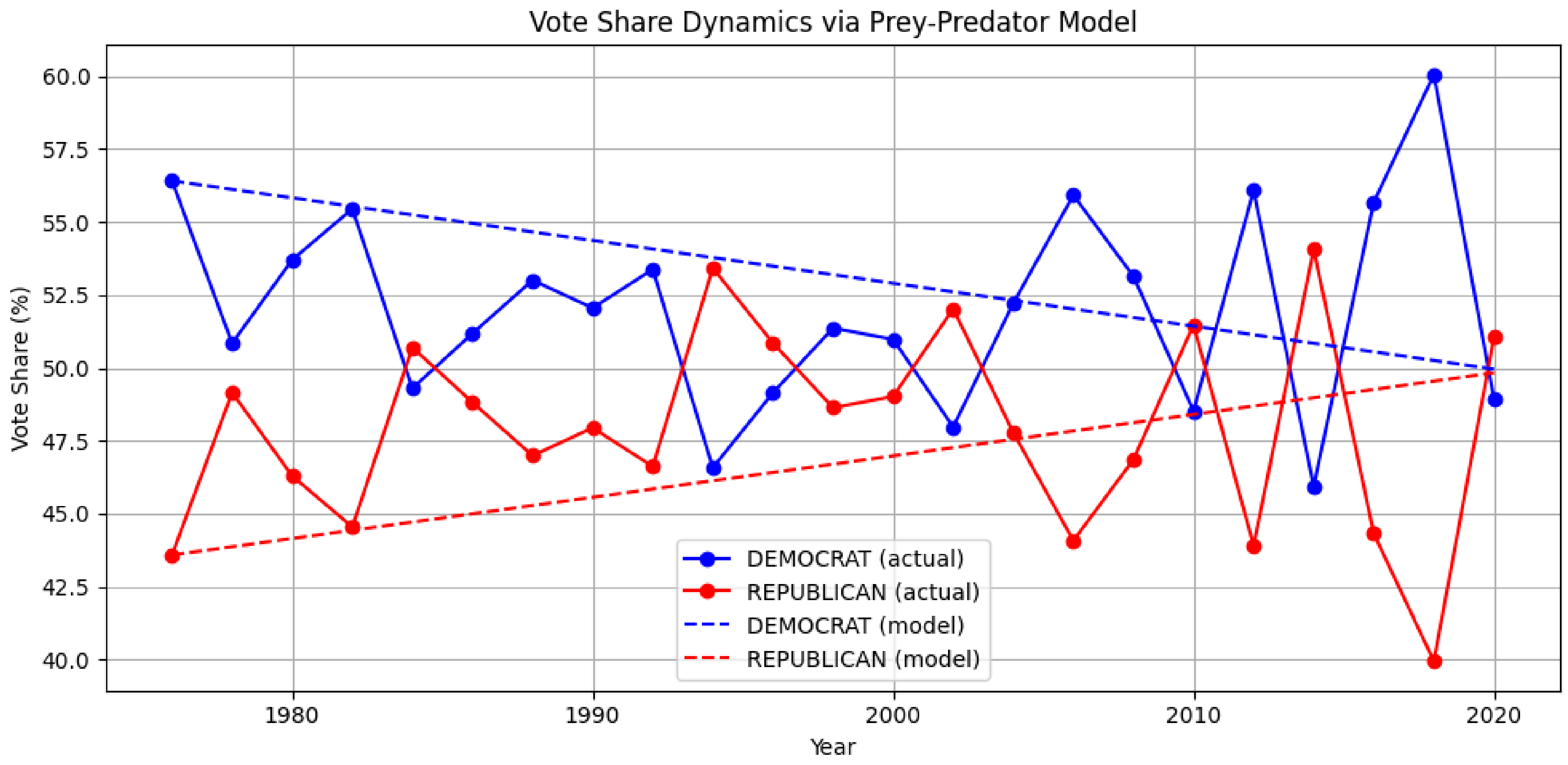
SIR(beta, gamma)

Mechanistic (Analogy): Uses an epidemic spread analogy. Voters are categorized as:
S (Susceptible): Undecided or persuadable voters.
I (Infected): Voters decided for the party being modeled (this is the vote_percentage data).
R (Recovered): Voters no longer susceptible or supporting this party (e.g., decided for the other party, became apathetic).
ODE System: Models the flow between these compartments using differential equations:. The total population S + I + R is implicitly normalized to 1.
Univariate Focus: Fits the model to a single time series (I = vote percentage for one party). S and R are derived relative to this.
Aggregate Model: Treats voters as homogeneous within the S, I, R compartments



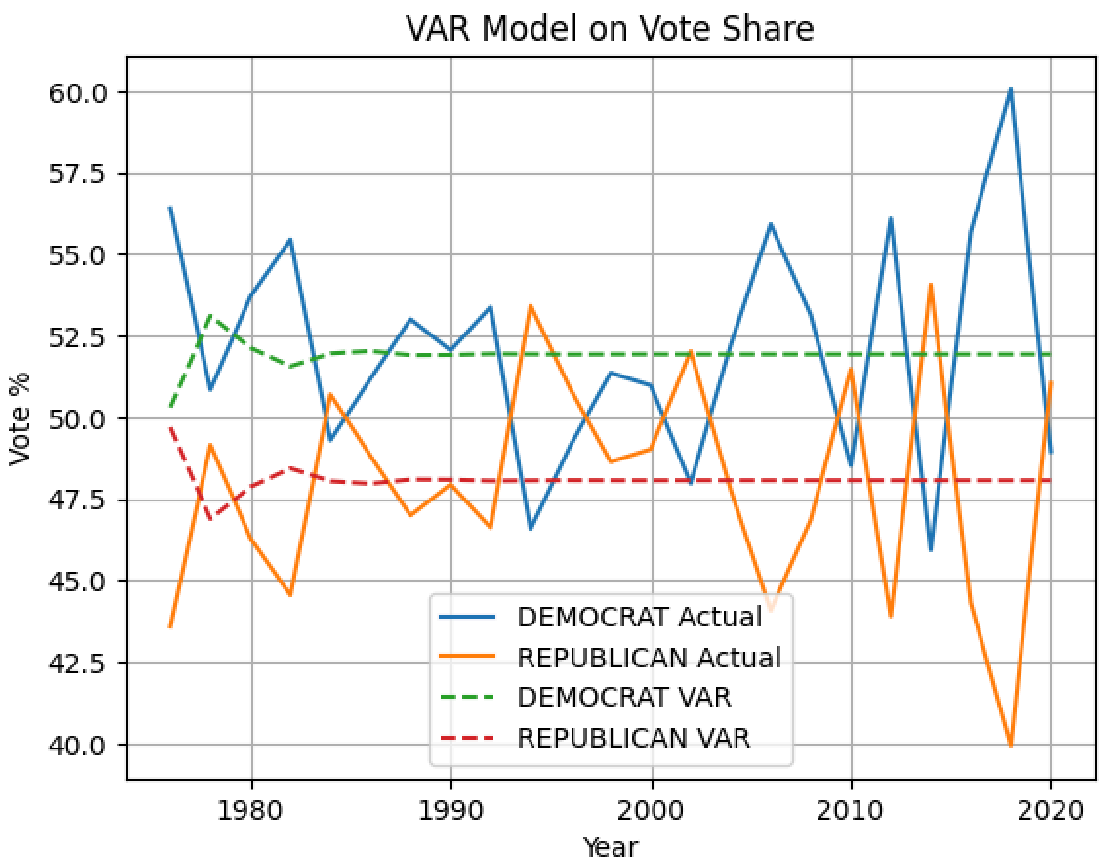

NEW HAMPSHIRE - REPUBLICAN
NEW HAMPSHIRE - DEMOCRAT
NORTH CAROLINA - REPUBLICAN
NORTH CAROLINA - DEMOCRAT
OKLAHOMA - REPUBLICAN
OKLAHOMA - DEMOCRAT
OREGON - REPUBLICAN
OREGON - DEMOCRAT
SOUTH CAROLINA - REPUBLICAN
SOUTH CAROLINA - DEMOCRAT
SOUTH DAKOTA - REPUBLICAN
SOUTH DAKOTA - DEMOCRAT
β
γ
Sum of Squared Residuals (SSR)

LOTKA VOLTERRA

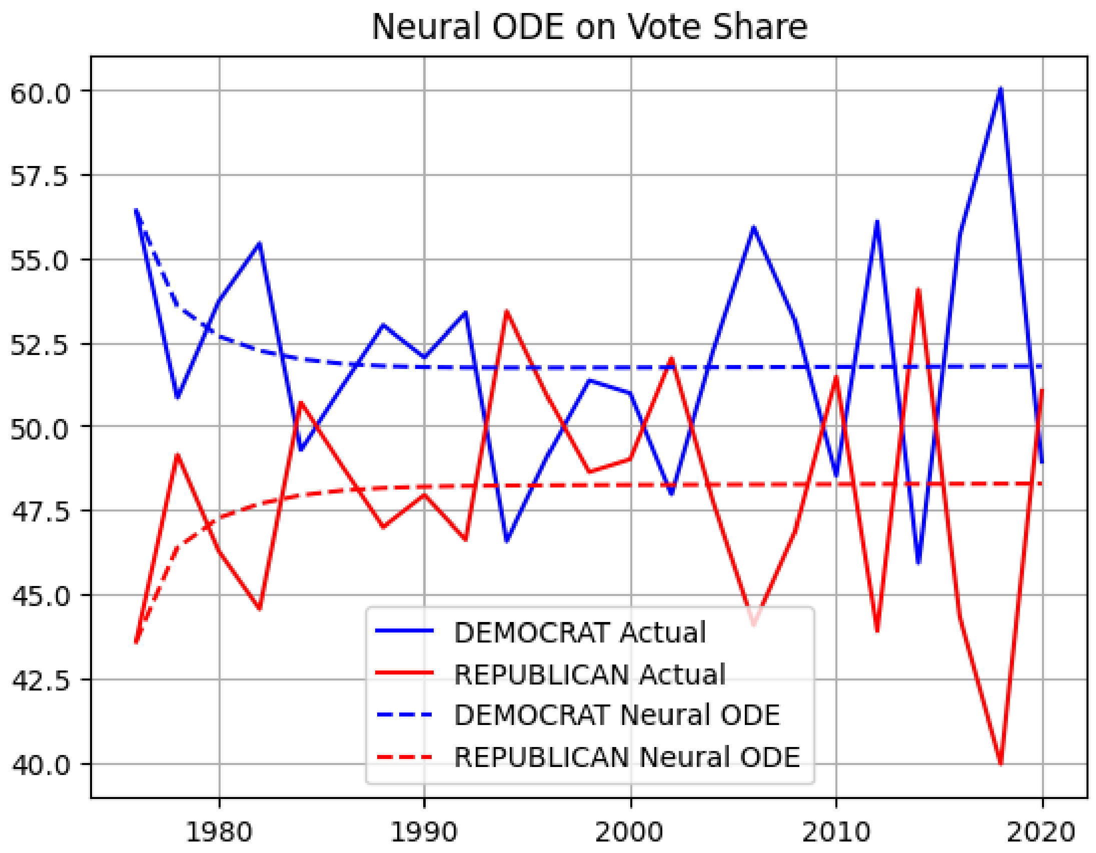


- **Mechanistic (Analogy)**: Assumes an underlying dynamic process governs the vote shares, analogous to species competition.
- **Coupled ODEs**: Models Democrat and Republican shares (x, y_val) simultaneously using differential equations that include growth terms (logistic-like) and interaction terms (beta * x * y_val, delta * x * y_val).
- **Parameter Estimation**: Uses optimization (scipy.optimize.minimize) to find ODE parameters (alpha, beta, delta, gamma) that best fit the historical data.
- **Aggregate Model**: Treats the entire vote share percentage as homogeneous entities following the ODEs.

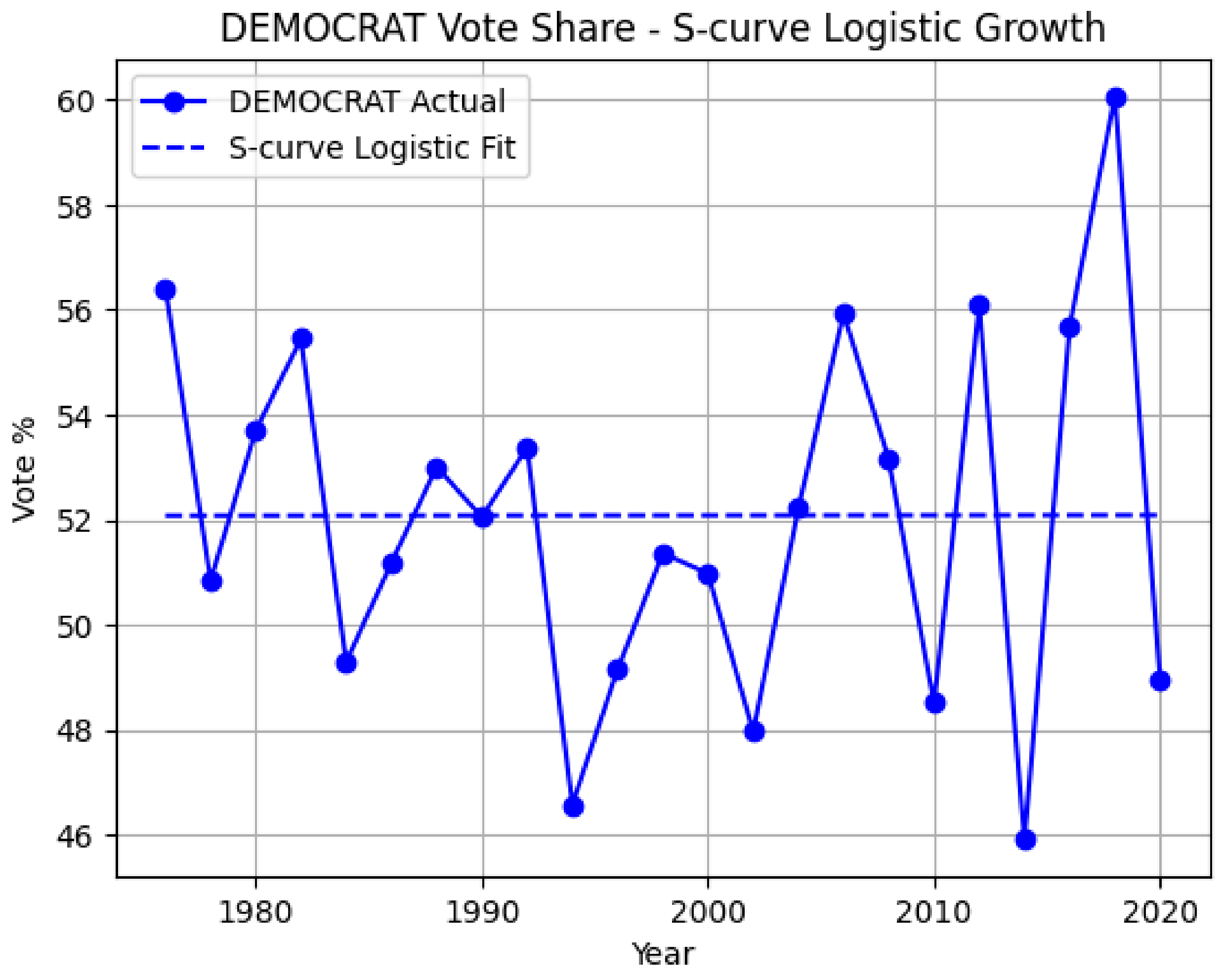
VAR



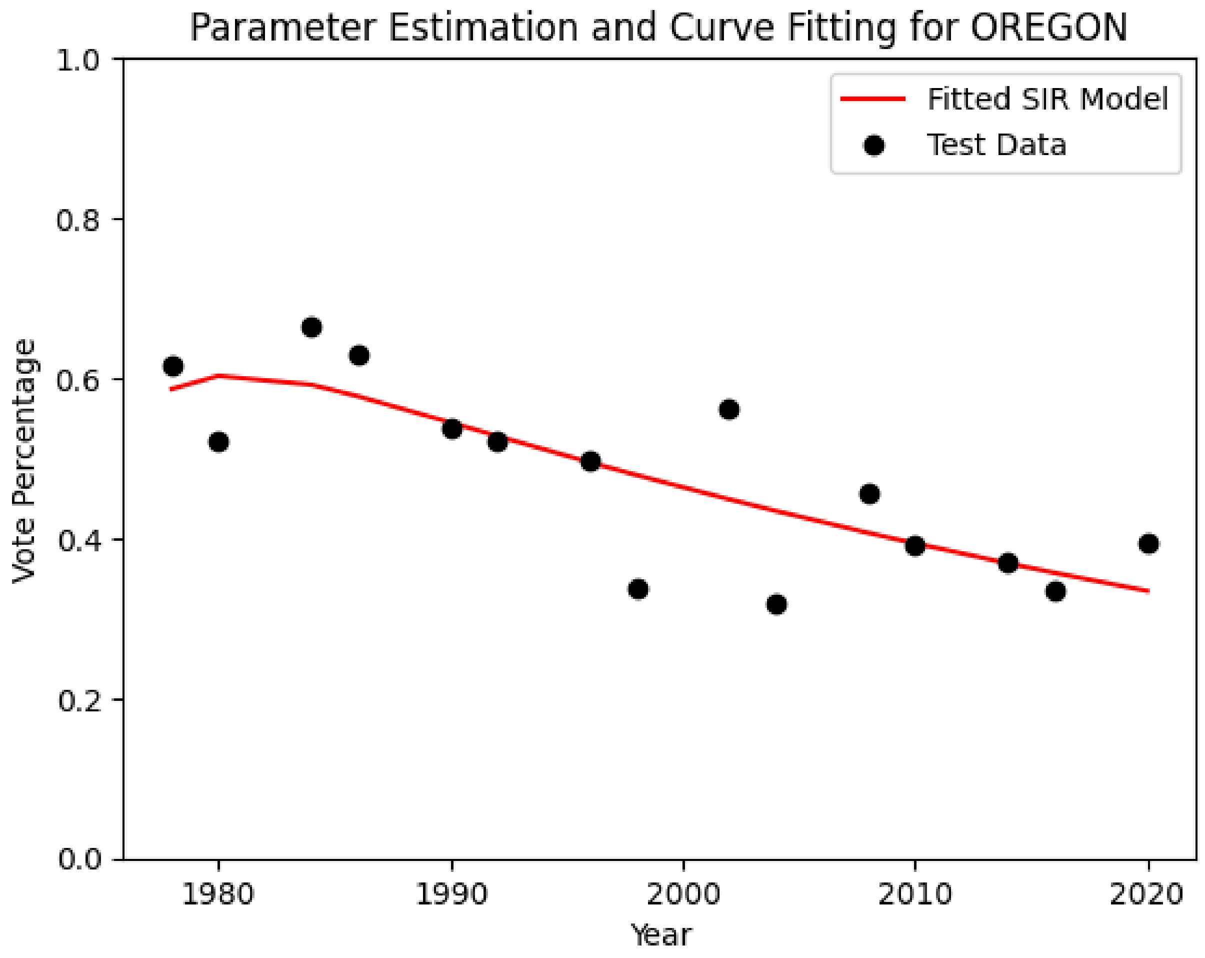
NEURAL ODE



S CURVE


DEMOCRAT Vote Share - S-curve Logistic Growth
DEMOCRAT Actual
S-curve Logistic Fit
Vote %
Year
60
58
56
54
52
50
48
46
1980
1990
2000
2010
2020

EA


Parameter Estimation and Curve Fitting for OREGON
Fitted SIR Model
Test Data
Vote Percentage
Year
1.0
0.8
0.6
0.4
0.2
0.0
1980
1990
2000
2010
2020

CONCLUSION

INTEGRATING BOTH THESE MODELS WITH AGENT BASED MODELS IS A VERY INTERESTING AND POWERFUL APPLICATION

Both the SIR and Lotka-Volterra models serve as useful conceptual tools with a large scope to be modified

Embedding the SIR contagion within realistic social and contact networks could facilitate studies on clustering phenomena
and echo chambers,

We propose a framework for a Multi-Compartment and Multi-Group Susceptible-Infected-(SIR) model, which opens several promising avenues for further research into electoral dynamics.

REFERENCES

- J. M. Epstein, Generative Social Science: Studies in Agent-Based Computational Modeling (Princeton University Press, 2006).
- G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, Time Series Analysis: Forecasting and Control (Wiley, 2008), 4th ed.
- C. A. Sims, Econometrica 48, 1 (1980).
- R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, "Neural ordinary differential equations," in Advances in Neural Information Processing Systems (NeurIPS), , vol. 31 (2018).
- 5. R. Pearl and L. J. Reed, Proc. National Acad. Sci. United States Am. 6, 175 (1920).
- M. Raissi, P. Perdikaris, and G. E. Karniadakis, J. Comput. Phys. 378, 686 (2019).
- W. O. Kermack and A. G. McKendrick, Proc. Royal Soc. A: Math. Phys. Eng. Sci. 115, 700 (1927).
- J. H. Holland, Adaptation in Natural and Artificial Systems (University of Michigan Press, 1975)