The Comovement of Voter Preferences: Insights from U.S. Presidential Election Prediction Markets Beyond Polls*

Mikhail Chernov
UCLA, NBER, and CEPR

Vadim Elenev
Johns Hopkins University

Dongho Song
Johns Hopkins University

October 28, 2024

Abstract

We propose a novel time-series econometric framework to forecast U.S. Presidential election outcomes in real time by combining polling data, economic fundamentals, and political prediction market prices. Our model estimates the joint dynamics of voter preferences across states. Applying our approach to the 2024 Presidential Election, we find a two-factor structure driving the vast majority of the variation in voter preferences. We identify electorally similar state clusters without relying on historical data or demographic models of voter behavior. Our simulations quantify the correlations between state-level election outcomes. Failing to take the correlations into account can bias the forecasted win probability for a given candidate by nearly 20 percentage points. We find Pennsylvania to be the most pivotal state in the 2024 election. Our results provide insights for election observers, candidates, and traders.

Key words: Election forecasting, prediction markets, state correlations, voter preferences.
JEL codes: C32, C53, D72, P00.

*We thank Alexandre Poirier for providing useful comments. Correspondence: M. Chernov: mikhail.chernov@anderson.ucla.edu; V. Elenev: velenev@jhu.edu; D. Song: dongho.song@jhu.edu.

1 Introduction

Elections matter. The results of legislative or presidential elections in countries around the world have the potential to significantly and lastingly affect welfare (Nordhaus (1975), Persson, 2002, Besley and Case, 2003). Uncertainty about who will win constitutes a major source of political risk. Expectations about the outcome of an election may influence how firms invest, how households consume, how policymakers regulate, and how candidates act to maximize their chances of winning (see, e.g., Besley and Burgess, 2002, Drazen, 2008, Julio and Yook, 2012, Gulen and Ion, 2016, Jens, 2017, Bonaime et al., 2018, Meeuwis et al., 2022).

United States plays a central role in global affairs, meaning the U.S. Presidential election captures an especially high level of attention, (Ikenberry, 2001, Cha and Szechenyi, 2024, and Foerster and Schmitz, 1997). However, the U.S. electoral system makes forecasting the outcome of its election particularly challenging. There are 56 simultaneous regional elections being held in all 50 states, the District of Columbia, and individual congressional districts in Maine (two) and Nebraska (three).[1] The winner of each is awarded a number of "electoral votes" (EVs) approximately proportional to the region's population.[2] To become the next President, a candidate must secure the majority of these electoral votes ("win the electoral vote"). This stands in contrast to electoral systems in most other presidential republics, where a candidate wins the election by accumulating the most individual votes ("wins the popular vote").

As a result, to form expectations about the winner of the presidential election, one needs to model the joint distribution of all 56 election outcomes, and then consider all the combinations that produce an electoral vote majority for a given candidate. In other words, a forecaster must produce not just an expected probability of a candidate winning the election in every state ("marginal distribution") but also the changes in these probabilities given by the results in a particular state ("conditional distribution"). Historically, such models were informal and based on heuristic

[1] To simplify terminology, we will refer to all regions, including D.C. and the congressional districts, as "states."

[2] Each state's number of EVs is equal to the number of congressional districts plus two. The U.S. constitution gives each state the power to decide how they allocate their electoral votes. Most states allocate all of their EVs to the winner in that state. Maine and Nebraska give the statewide winner two EVs, while giving one EV to the winner of each district. For a comprehensive review of the U.S. presidential electoral system, see National Archives (2024), which explains the composition and procedures of the Electoral College and its implications for presidential elections.

assessments of similarity between a given pair of states. It is common to hear statements like, "Pennsylvania and Michigan vote the same because they're both part of the Rust Belt region." More recently, quantitative models have emerged that formally characterize the joint distribution as a correlated binomial.[3] But, as made salient by the failures of such models for mortgage default prediction during the Great Financial Crisis, they are especially sensitive to parameters governing the correlation structure, which are particularly challenging to estimate.[4]

A key data input into these models are polls, which measure current voter preferences. One approach could be to use polling data to estimate parameters of a joint distribution, governing both marginal distributions and the correlation structure. However, polls can change day to day either because some voters changed their mind or because the error, with which they measure voter preferences, changed (see, e.g., Kenett et al. (2018) and Jennings and Wlezien (2018) on polling inaccuracies). Only the former changes are relevant for disciplining the joint distribution, yet it is not possible to separately identify them from polling data.

In this paper, we propose a time-series econometric framework that allows us to identify the correlation structure between voter preferences in different states *encoded in the beliefs of asset market investors*. Arrow et al. (2008), Snowberg et al. (2008), and Snowberg et al. (2013) suggest that prediction markets can offer forecasts with greater accuracy than traditional methods.

Using polling data, data on economic fundamentals, and asset prices from political prediction markets for the 2024 U.S. Presidential Election, we construct a joint distribution of election outcomes in all states conditional on real-time polls and asset prices. To our knowledge, this is the first formal approach of its kind, representing the paper's main contribution.

Our estimates permit us to answer commonly asked questions in election forecasting. We formally group states into electorally similar clusters *without* relying on a prior model of similarity driven by, e.g., demographics. We identify pivotal states, i.e., states in which a victory is most likely to increase a candidate's overall odds of winning. Our estimates are useful not just to election observers, but also to candidates as they optimally allocate scarce campaign resources and to election night analysts and traders, as they continually update their forecasts as results

[3]Two prominent examples are FiveThirtyEight and the Princeton Election Consortium.
[4]See Salmon (2009) for a non-technical overview.

come in.

We set up the problem using state space representation. Each state's voter preferences are a latent variable that fluctuates stochastically around a state fixed effect. The fluctuations are persistent, respond to observable fundamental shocks (e.g., state of the economy), as well as to "political" shocks, that is, shocks to voter preferences that are orthogonal to fundamentals. Both fundamental and political shocks can be correlated across states. Pollsters observe and report each state's voter preferences with error. The error is also persistent and potentially correlated across states, making it a second latent state variable. The poll results, observable by the econometrician, are a sum of the two. With $M + 2$ latent variables (for M fundamentals) and $M + 1$ observable, identification is not possible absent additional data.

These data are provided by political prediction markets, which allow investors to bet on the outcome of each state's election. An investor can buy or sell an exchange-traded contract that pays out \$1 after and if candidate A wins a given election. The contracts need not be held to maturity. Under standard conditions, the market-clearing price on a given day before the election can then be interpreted as the marginal investor's risk-neutral probability of candidate A winning the election conditional on all information available to investors as of that date (Wolfers and Zitzewitz, 2006). This information contains polls and fundamentals, observable by the econometrician, but also many other variables that drive fluctuations in voter preferences. This means that changes in the price of this contract can shed additional light on changes in voter preferences beyond that which is contained in polls.

We solve for each state's contract price as a (nonlinear) function of the state variables and use it as an additional measurement equation. This enables us to estimate the parameters governing state dynamics as well as the values of the latent variables.

We use the estimated panel of voter preference across time and states to identify comovement patterns. The first two principal components of the panel account for 94.2% of the variation. The first principal component captures the overall trend in voter preferences towards the Democratic candidate. The second principal component has a down-up-down pattern, with reversals occurring around the times of key news events in the campaign, including the change in the Democratic candidate.

The states' loadings on the first two principal components capture relevant cross-sectional variation. Distances between states in the space formed by the first two principal components provide a measure of similarity. Given that the only inputs into our model are polls, fundamentals, and prediction market prices, this measure of similarity is driven entirely by 2024 data. We find two partially overlapping clusters that include the majority of competitive states. The first cluster includes the three "Rust Belt" states of Pennsylvania, Wisconsin, and Michigan, while the second includes three "Sun Belt" states of Arizona, Georgia, and Nevada, providing contemporaneous empirical validation to these popularly cited groupings. These clusters are similar in their response to the change in the Democratic candidate, but differ in their exposure to the overall trend towards the Democratic candidate. Other states have individually distinct patterns, including North Carolina, which is often grouped – erroneously, in our model's view – with the Sun Belt states.

The informative and intuitive patterns we uncover would not be revealed by a PCA performed on polls directly. Polling data is missing for some states for large parts of the sample, and the noisiness of the data would render it impossible to decompose comovement into two distinct components. Incorporating prediction market data in our analysis and using it to estimate the latent voter preferences is integral to uncovering the two-component patterns described above.

We use the estimated dynamics to simulate the election outcome in each state. These outcomes are significantly correlated, which we quantify by comparing the probability of a candidate winning one state conditional on winning another and contrasting it with the unconditional probability of winning that state. The magnitude of the update depends both on how surprising the win is, but also on how central (in PCA loading space) the won state is, and therefore how correlated other states are to it. For example, as of October 27, while the Democratic candidate Kamala Harris is equally likely to win Arizona as she is to win Georgia (approximately 5-7%), a win in Arizona – a more central state – raises her probability of winning the key battleground state of Pennsylvania from 23% to 98%, while a win Georgia raises it only to 88%.

Finally, we aggregate the state-level forecasts into a national election forecast and contrast the result with a counter-factual, in which individual state win probabilities are the same as in the baseline but the cross-sectional correlations are set to zero. Our results show the Republican

candidate Donald Trump to be favored to win the election, with a probability of 72.6%. Because correlations increase the dispersion of outcomes, failing to take them into account would substantially raise Trump's odds of victory to 90%.

A win in a given state raises a candidate's probability of winning the overall election both because receiving that state's electoral votes brings the candidate closer to the majority and also because the win in that state presages better electoral performance in other, correlated, states. We quantify the magnitude of the latter and show that it is substantial. We also use it to identify how "pivotal" each state is, i.e., how much a win in that state would increase either candidate's overall odds of winning. We find Pennsylvania to be the most pivotal state both because of its large number of electoral votes and because of its central position in the PC loading space. This result is consistent with the popular narrative that Pennsylvania is the most important state in the 2024 election. Somewhat surprisingly, a similarly large closely-contested state of Michigan is one of the least pivotal ones, because of its comparatively lower centrality.

Our paper contributes to the existing literature on time-series econometrics, political forecasting, and election prediction markets, building on foundational works in these areas.

Methodologically, we build on advances in state-space modeling, as outlined in Durbin and Koopman (2001), and Bayesian forecasting techniques discussed in West and Harrison (2006) and Del Negro and Schorfheide (2011). Additionally, we draw on Elliott and Timmermann (2013), which offers a comprehensive overview of forecasting methods used across macroeconomics and finance.

We contribute to the literature on election forecasting, which has seen contributions from a range of disciplines. Economists have focused on the role of economic variables for explaining election results in-sample and predicting them out-of-sample. Fair (1996) summarizes prominent examples, while Fair (2011) provides an updated and non-technical review of the forecasting literature. Prominent recent work connects broad macroeconomic trends such as globalization and risk appetites to election outcomes (see Autor et al., 2020, Pástor and Veronesi, 2020, and Pástor and Veronesi, 2021). The political science work on election forecasting is extensive. Its origins are reviewed by Lewis-Beck (2005). Prominent 21st century examples of U.S. presidential election forecasting include Abramowitz (2004) and Jennings et al. (2020). Many advances in election

forecasting occurred outside the academia. Silver (2012) reflects on the evolution of election forecasting, emphasizing the limitations of polling data and the advancements in predictive models over time.

Several studies examine political prediction markets. Snowberg et al. (2007) and Leigh and Wolfers (2006) examine prediction markets during the 2004 election in the U.S. and Australia, respectively. However, markets for outcomes in individual states have developed only recently, and we believe our paper is the first to use them for forecasting. Calvo et al. (2024) demonstrates that demographic trends fail to improve U.S. election forecasts, even with perfect knowledge of future shifts, which reinforces the value of focusing on alternative predictive factors like polling data and market prices, as we do in our model.

Lastly, our comovement estimates allow us to propose our own forecasting model. There are several contemporaneous models of the 2024 U.S. presidential election that have received considerable popular attention, e.g., The Economist (2024). FiveThirtyEight (2024) carries the brand name of Nate Silver's models for 2008-2020 elections, though Silver is no longer connected with the company. He maintains his own forecast at Silver (2024). We rely on both of these websites for poll aggregation, but our model is distinct in that it uses prediction market data to infer cross-sectional correlations.

In sum, we contribute to this literature by proposing a novel framework that integrates polling data, economic fundamentals, and prediction market prices to capture the dynamics of voter preferences across states, thereby enhancing political forecasting methodologies. This comprehensive approach offers valuable guidance for stakeholders in the upcoming U.S. presidential election, potentially improving upon current practices in both academia and the field.

The rest of the paper is organized as follows. Section 2 describes the data sources and the construction of our model's inputs. Section 3 explains the econometric framework. Section 4 presents the results, and Section 5 concludes.

2 Data

There are several types of data we need for this exercise. To estimate the model, we need to combine polling data with prediction market prices and data on fundamentals. Since we perform the exercise in real-time during the 2024 election campaign, our goal is to construct a state day panel for each variable.

A particular difficulty of working with 2024 election data is the unprecedented decision by incumbent President and Democratic nominee Joe Biden to drop out of the race and endorse his Vice President Kamala Harris instead. Essentially overnight, Harris emerged as the de facto Democratic nominee in place of Biden, and the contest between Biden and former President Donald Trump became a Harris-Trump race instead.

We address this by taking a party-level view of the race. Instead of considering individual candidates separately, we write our model in terms of voter preferences between a Democrat and a Republican. In this context, the overnight replacement of Biden with Harris shows up as (potentially large) shock to voter preferences for the Democratic candidate.

Polling data poses a particular challenge. Multiple polls may be released on the same day by different pollsters with different methodologies or samples. Some polls will include only the Democrat and Republican candidates, one of whom is overwhelmingly likely to win in each state, while others include third-party candidates as well. Pollsters may construct a sample of all adults, or all registered voters, or all likely voters, which is a subset of registered voters assessed likely to vote based on a self-report. On other days, there may not be any polls released. To avoid dealing with all the methodological issues involved in constructing an appropriate polling average, we use averages constructed by polling aggregators. Our main data source is FiveThirtyEight, a quantitative journalism and political forecasting website owned and operated by ABC News. It provides a daily polling average for each state starting in about March 2024. To address FiveThirtyEight's limited data availability immediately after Harris replaced Biden as Democratic nominee, we augment the data with polling averages from the Silver Bulletin, authored by Nate Silver, the founder and former head of FiveThirtyEight now operating independently. For each state-day, we define voter preferences as the difference between the polling average for the Democratic candidate and

Table 1: Summary Statistics

	2020 Results	Prediction Market Prices					Polling Averages				
	(Dem-Rep)	N	Start	End	Mean	Std	N	Start	End	Mean	Std
Arizona	0.30	233	3/07	10/25	39.5	6.5	238	3/01	10/24	-3.2	2.1
Florida	-3.36	221	3/19	10/25	11.3	4.1	191	4/18	10/25	-7.4	2.5
Georgia	0.23	233	3/07	10/25	37.1	6.0	238	3/01	10/24	-3.9	2.5
Michigan	2.78	233	3/07	10/25	53.2	6.2	239	3/01	10/25	-0.5	2.3
Minnesota	7.12	221	3/19	10/25	83.8	8.3	144	5/14	10/25	5.1	2.7
Nevada	2.39	233	3/07	10/25	44.5	9.8	239	3/01	10/25	-3.2	3.2
North Carolina	-1.34	232	3/08	10/25	30.3	7.8	239	3/01	10/25	-4.2	2.9
Pennsylvania	1.16	234	3/06	10/25	49.8	5.2	239	3/01	10/25	-0.9	1.7
Virginia	10.11	212	3/28	10/25	82.9	6.3	107	7/11	10/25	5.6	2.2
Wisconsin	0.63	233	3/07	10/25	51.1	4.9	238	3/01	10/24	-0.2	2.0

Notes: Our main data source is FiveThirtyEight, a quantitative journalism and political forecasting website owned and operated by ABC News. To address FiveThirtyEight's limited data availability immediately after Harris replaced Biden as Democratic nominee, we augment the data with polling averages from the Silver Bulletin, authored by Nate Silver. Political prediction market prices, quoted in cents, are taken from Polymarket.

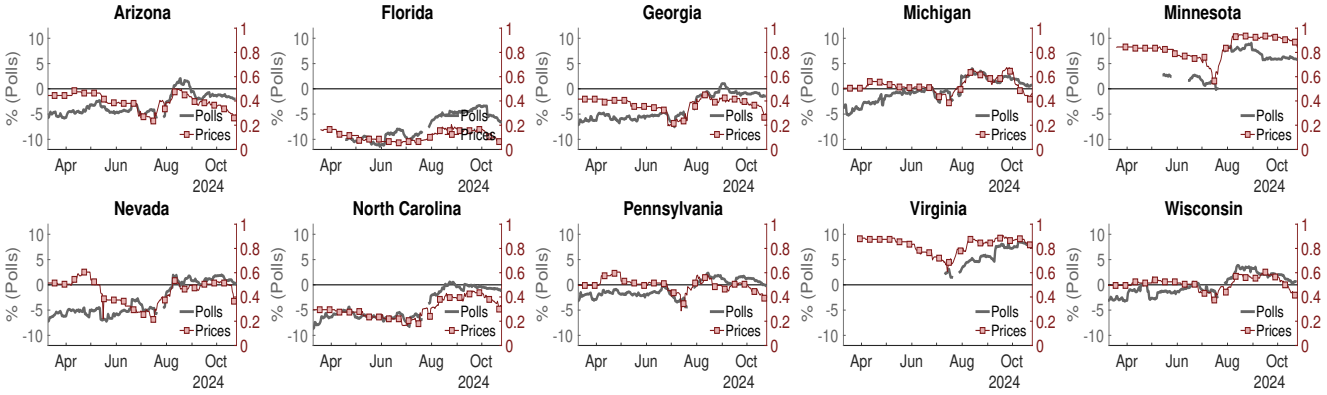
the polling average for the Republican candidate.

Political prediction market prices are taken at daily frequency from Polymarket. Polymarket is a platform for hosting prediction markets, and has the largest notional amounts outstanding of contracts on the U.S. Presidential election. For example, the market for the ultimate winner of the overall election has a value $270 million. We obtain daily end-of-day prices of contracts paying $1 if the Democratic candidate wins a given state for each traded state.

Many states are unlikely to be competitive. Past election results, current polls, and current prediction market prices all suggest, for example, that the Democrat candidate will almost certainly win California, while the Republican candidate will win Kansas. Markets for these states are not particularly liquid. And even if they were, their prices are not sensitive to changes in voter preference. For example, if Kansas' voters shift from preferring Republicans by 30 percent to preferring them by only 25, a Democrat's chance of winning the state will increase only insignificantly.

Such states are also polled less frequently, making polling averages more likely to be stale. As a result, we focus our analysis on the 10 "battleground," or most competitive, states, defined as Arizona, Florida, Georgia, Michigan, Minnesota, Nevada, North Carolina, Pennsylvania, Virginia,

Figure 1: Polling averages and prediction market prices



Notes: Our main data source is FiveThirtyEight, a quantitative journalism and political forecasting website owned and operated by ABC News. To address FiveThirtyEight's limited data availability immediately after Harris replaced Biden as Democratic nominee, we augment the data with polling averages from the Silver Bulletin, authored by Nate Silver. Political prediction market prices are taken from Polymarket.

Wisconsin, respectively. Table 1 presents summary statistics and Figure 1 plots the time-series of polling averages and prediction market prices and reports the date ranges for which the data is available. Figure 1 shows extensive missing data and notable discrepancies between polling averages and prediction market prices, particularly during April and May. Although our estimation method can handle missing observations, we aim to avoid the impact of this initially imperfect data on our inferences. Therefore, we restrict our estimation sample to the period from June 1, 2024, through the most recent data available as of October 25, 2024.

Finally, we consider three series measuring economic fundamentals, with our choice restricted by the requirement that the data be available at daily frequency. We obtain the Aruoba-Diebold-Scotti Business Conditions Index from the Philadelphia Fed, the level of the S&P 500 stock index and the Treasury yield spread between the 10-year and 3-month rates from the St. Louis Fed FRED database. All of these are national-level variables, affecting voter preferences in all states, potentially with their own beta. Because of data availability, we do not consider any state-specific fundamentals series.

3 Econometric Framework

We define $d_{i,t} = \text{Dem}_{i,t} - \text{Rep}_{i,t}$ as the net voter preference ("NVP") for the Democrat presidential candidate in state $i \in \{1, ..., N\}$ observed at time t during an election held at time T. The final election outcome is represented by $d_{i,T}$, where $d_{i,T} > 0$ signifies a victory for the Democratic presidential candidate in state i. We consolidate the dimensions of two polls into a single measure representing the net margin for the Democratic presidential candidate. This measure can be easily negated to derive the net margin for the Republican presidential candidate.

3.1 Democratic candidate's net percent margin dynamics

The poll-based estimate of the net percent margin, $d^o_{i,t}$, is specified as

$$\begin{bmatrix} d^o_{1,t} \\ \vdots \\ d^o_{N,t} \end{bmatrix} = \begin{bmatrix} d_{1,t} \\ \vdots \\ d_{N,t} \end{bmatrix} + \begin{bmatrix} \nu_{d,1,t} \\ \vdots \\ \nu_{d,N,t} \end{bmatrix}, \tag{1}$$

$$\begin{bmatrix} \nu_{d,1,t} \\ \vdots \\ \nu_{d,N,t} \end{bmatrix} = \begin{bmatrix} \mu_{\nu_d,1} \\ \vdots \\ \mu_{\nu_d,N} \end{bmatrix} + \begin{bmatrix} \rho_{\nu_d,1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \rho_{\nu_d,N} \end{bmatrix} \begin{bmatrix} \nu_{d,1,t-1} \\ \vdots \\ \nu_{d,N,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{d,1,t} \\ \vdots \\ \epsilon_{d,N,t} \end{bmatrix},$$

where $\nu_{d,i,t}$ represents measurement errors that are not mean-zero and include correlated shocks $[\epsilon_{d,1,t}, ..., \epsilon_{d,N,t}]' = \epsilon_{d,t} \sim N(0, \Sigma_{\epsilon_d})$. The dynamics of the true margin are as follows

$$\begin{bmatrix} d_{1,t} \\ \vdots \\ d_{N,t} \end{bmatrix} = \begin{bmatrix} \rho_{d,1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \rho_{d,N} \end{bmatrix} \begin{bmatrix} d_{1,t-1} \\ \vdots \\ d_{N,t-1} \end{bmatrix} + \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} + \begin{bmatrix} \beta_1' \\ \vdots \\ \beta_N' \end{bmatrix} \odot \begin{bmatrix} f_{1,t} \\ \vdots \\ f_{N,t} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ \vdots \\ u_{N,t} \end{bmatrix}, \tag{2}$$

where $\odot$ denotes element-wise multiplication, $f_{i,t}$ is a vector of fundamentals, potentially including state-specific factors, and the correlated shocks affecting the true polls are represented by $[u_{1,t}, ..., u_{N,t}]' = u_t \sim N(0, \Sigma_u)$.

3.2 Dynamics of fundamentals

The law of motion of the vector of (de-meaned) fundamentals $f_{i,t}$ follows

$$f_{i,t} = \rho_{f,i} f_{i,t-1} + w_{i,t}, \quad w_{i,t} \sim N(0, \Sigma_{w,i}). \tag{3}$$

We assume that these fundamentals are observable.

3.3 Prediction market asset price dynamics

An investor would pay $p^o_{i,t}$ for a contract that pays her 1 if the Democrat wins state i and 0 otherwise. We posit that

$$p^o_{i,t} = \Phi\Big(\lambda_p + \Phi^{-1}(p_{i,t}) + \epsilon_{p,i,t}\Big), \quad [\epsilon_{p,1,t}, ..., \epsilon_{p,N,t}]' = \epsilon_{p,t} \sim N(0, \Sigma_{\epsilon_p}), \tag{4}$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution and $p_{i,t}$ denotes the true probability of a Democratic presidential candidate's victory described in (5) below.[5] In essence, we assume that prices are observed with a bias—uniform across all states—and measurement errors that may be correlated across states. Specifically, λ_p represents a persistent bias that remains consistent across all states over the estimation sample, while $\epsilon_{p,i,t}$ captures biases that may be correlated across states but are transient.[6]

In the absence of λ_p and $\epsilon_{p,i,t}$, the observed price accurately reflects the true probability of a win. Wolfers and Zitzewitz (2006) argue that prediction market prices can be interpreted as the beliefs of the marginal trader about the probability of an event happening, under the assumption that markets are efficient and traders are risk-neutral.

Denote the information set at time t as $\mathcal{I}_t$. The probability that the Democratic presidential

[5] For a similar inverse cdf transformation applied to the data, refer to Song and Tang (2023).

[6] We do not rule out the possibility that this constant term includes a risk premium component. In such a case, it is straightforward to deduce that if $\lambda_p > 0$, the observed price $p^o_{i,t}$ exceeds the true price $p_{i,t}$, implying that investors' marginal utility is higher when the Democrat wins. This suggests that the risk premium is negative, and thus, the contract can be interpreted as a hedge. Conversely, if $\lambda_p \leq 0$, the interpretation is reversed.

candidate wins is

$$
\begin{aligned}
p_{i,t} &\equiv P(d_{i,T} > 0 | \mathcal{I}_t), \\
&= P\left(u^e_{i,t+(T-t)} + \beta_i' w^e_{i,t+(T-t)} > -(\rho^{(T-t)}_{d,i} d_{i,t} + \alpha_i \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} + \beta_i' f^e_{i,t}) | \mathcal{I}_t \right), \\
&= \Phi\left(\frac{\rho^{(T-t)}_{d,i} d_{i,t} + \alpha_i \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} + \beta_i' f^e_{i,t}}{(\sigma^2_{u^e,i} + \beta_i' \sigma^2_{w^e,i} \beta_i)^{1/2}} \right),
\end{aligned}
\tag{5}
$$

which is derived based on (6) and (7), as provided below.

First, we iterate the dynamics of the fundamentals in (3) forward to determine the true margin at time T for state i, based on (2), as follows:

$$
\begin{aligned}
d_{i,t+(T-t)} = \rho^{(T-t)}_{d,i} d_{i,t} + \alpha_i \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} + \beta_i' \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} \rho^{(T-t-j)}_{f,i} f_{i,t}, \\
+ \beta_i' \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} \sum_{k=0}^{(T-t-j)-1} \rho^k_{f,i} w_{i,t+(T-t-j)-k} + \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} u_{i,t+(T-t)-j}.
\end{aligned}
\tag{6}
$$

The exposition in (5) relies on the following simplifying notations which are based on (6):

$$
\begin{aligned}
f^e_{i,t} &= \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} \rho^{(T-t-j)}_{f,i} f_{i,t}, \\
u^e_{i,t+(T-t)} &= \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} u_{i,t+(T-t)-j}, \\
w^e_{i,t+(T-t)} &= \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} \sum_{k=0}^{(T-t-j)-1} \rho^k_{f,i} w_{i,t+(T-t-j)-k}, \\
\sigma^2_{u^e,i} \equiv var(u^e_{i,t+(T-t)}) &= \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} (e_i' \Sigma_u e_i) (\rho^j_{d,i})', \\
\sigma^2_{w^e,i} \equiv var(w^e_{i,t+(T-t)}) &= \sum_{j=0}^{(T-t)-1} \left(\sum_{k=0}^{j} \rho^{j-k}_d I \rho^k_{f,i} \right) \Sigma_{w,i} \left(\sum_{k=0}^{j} \rho^{j-k}_d I \rho^k_{f,i} \right)',
\end{aligned}
\tag{7}
$$

where e_i is a selection vector with a 1 in the ith position and 0s in all other positions.

3.4 Discussion

It is important to recognize from (1) that the poll-based estimate d_t^o of the net voter preference alone cannot identify its mean level α nor the shocks u_t influencing the true margins d_t, due to the presence of potentially biased measurement errors $\nu_{d,t}$.

This is precisely where asset prices p_t^o in (4) become relevant, as they are assumed to be uncorrelated with the poll measurement errors $\nu_{d,t}$. Even though asset prices may also exhibit bias, represented by λ_p, this bias is assumed to be identical across states. As a result, asset prices aid in the identification of both the true margin level α and the shocks u_t affecting the true margin.

Since the dynamics of fundamentals are unaffected by the polls or asset prices, we can estimate (3) separately from the joint estimation of polls and asset prices.

In summary, our goal is to understand the dynamics of the true net voter preference through the joint estimation of observed polls and asset prices. In the following section, we facilitate this process by introducing the state-space representation of our framework.

3.5 State-space representation

It is crucial to differentiate between the end of the estimation sample, $T_E < T$, and the final election date, T. For $t \in \{1, \ldots, T_E\}$, we cast the model described in (1), (2), (4), and (5) into a state-space representation.

$$\begin{bmatrix} d_t^o \\ \Phi^{-1}(p_t^o) \end{bmatrix} = \begin{bmatrix} 0 \\ \delta_{p,t} \end{bmatrix} + \begin{bmatrix} I & I & 0 & 0 \\ \gamma_{p,t} & 0 & 0 & I \end{bmatrix} \begin{bmatrix} d_t \\ \nu_{d,t} \\ u_t \\ \epsilon_{p,t} \end{bmatrix}, \tag{8}$$

$$\begin{bmatrix} d_t \\ \nu_{d,t} \\ u_t \\ \epsilon_{p,t} \end{bmatrix} = \begin{bmatrix} \delta_{d,t} \\ \mu_{\nu_d} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \rho_d & 0 & 0 & 0 \\ 0 & \rho_{\nu_d} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} d_{t-1} \\ \nu_{d,t-1} \\ u_{t-1} \\ \epsilon_{p,t-1} \end{bmatrix} + \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ I & 0 & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} u_t \\ \epsilon_{d,t} \\ \epsilon_{p,t} \end{bmatrix},$$

where

$$d_t^o = \begin{bmatrix} d_{1,t}^o \\ \vdots \\ d_{N,t}^o \end{bmatrix}, \Phi^{-1}(p_t^o) = \begin{bmatrix} \Phi^{-1}(p_{1,t}^o) \\ \vdots \\ \Phi^{-1}(p_{N,t}^o) \end{bmatrix}, \alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}, \beta' = \begin{bmatrix} \beta_1' \\ \vdots \\ \beta_N' \end{bmatrix}, f_t = \begin{bmatrix} f_{1,t} \\ \vdots \\ f_{N,t} \end{bmatrix}, \mu_{\nu_d} = \begin{bmatrix} \mu_{\nu_d,1} \\ \vdots \\ \mu_{\nu_d,N} \end{bmatrix}$$

$$\delta_{p,t} = \begin{bmatrix} \lambda_p + \frac{\alpha_1 \sum_{j=0}^{(T-t)-1} \rho_{d,1}^j + \beta_1' f_{1,t}^e}{(\sigma_{u^e,1}^2 + \beta_1' \sigma_{w^e,1}^2 \beta_1)^{1/2}} \\ \vdots \\ \lambda_p + \frac{\alpha_N \sum_{j=0}^{(T-t)-1} \rho_{d,N}^j + \beta_N' f_{N,t}^e}{(\sigma_{u^e,N}^2 + \beta_N' \sigma_{w^e,N}^2 \beta_N)^{1/2}} \end{bmatrix}, \quad \gamma_{p,t} = \begin{bmatrix} \frac{\rho_{d,1}^{(T-t)}}{(\sigma_{u^e,1}^2 + \beta_1' \sigma_{w^e,1}^2 \beta_1)^{1/2}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\rho_{d,N}^{(T-t)}}{(\sigma_{u^e,N}^2 + \beta_N' \sigma_{w^e,N}^2 \beta_N)^{1/2}} \end{bmatrix},$$

$\delta_{d,t} = \alpha + \beta' \odot f_t$, and ρ_d and ρ_{ν_d} denote the persistence matrices, which are assumed to be diagonal in (1) and (2), respectively.

3.6 Gibbs sampler

The state-space representation (8) can be expressed generically as follows:

$$\begin{aligned} y_t &= \Lambda_{0,t} + \Lambda_{1,t} s_t, \\ s_t &= \Gamma_{0,t} + \Gamma_1 s_{t-1} + \Omega \varepsilon_{s,t}, \quad \varepsilon_{s,t} \sim N(0, \Sigma_s), \quad t \in \{1, ..., T_E\}. \end{aligned} \tag{9}$$

We collect parameters in

$$\Theta_f = \{\rho_f, \Sigma_w\}, \quad \Theta_d = \{\rho_d, \alpha, \beta, \Sigma_u\}, \quad \Theta_\nu = \{\mu_{\nu_d}, \rho_{\nu_d}, \Sigma_{\epsilon_d}\}, \quad \Theta_p = \{\lambda_p, \Sigma_{\epsilon_p}\}. \tag{10}$$

Estimation sample

We set $N = 10$ as shown in Table 1, with the estimation sample spanning from June 1 to October 25, 2024. This results in a panel with a time series length of $T_E = 145$ and cross-sectional units of $2N = 20$, incorporating both polls and prices. Given that the election date is November 5, 2024, we set $T = T_E + 11$.

Parameter restrictions

We seek to enhance the precision of our inference. To further reduce the number of unknown parameters in the model, we assume that the two diagonal persistence matrices, ρ_d and ρ_{ν_d}, each contain a single scalar persistence parameter that remains consistent across all states.

Our estimation sample is relatively short, which complicates the task of identifying both the true margin level α and the measurement error component μ_{ν_d}. This challenge is particularly acute when ρ_d and ρ_{ν_d} are close to a unit root. Additionally, we aim to determine the bias in prediction market prices, denoted by λ_p. To address these challenges, we initially assume μ_{ν_d} is zero, based on the assumption that its true mean is zero, even though it may exhibit non-zero values intermittently in short samples due to ρ_{ν_d} being near one. Under this assumption, we can jointly identify λ_p from both poll estimates and prices, given that λ_p is a scalar and remains constant across states.

As explained earlier, Θ_f can be estimated independently based on the observed f^{T_E}. Hence, we treat f^{T_E} and $\{\Theta_f\}^{(j)}$, where $j \in \{1, ..., N_{sim}\}$, as given within this Gibbs sampler. For simplicity, we exclude state-specific fundamental factors and use national ones instead. Therefore, under this assumption, any variation in the true margin across states can be entirely attributed to their loadings β_i, along with α_i and $u_{i,t}$, as demonstrated in (2).

We estimate Θ_f using daily data from March 1, 2024 to October 18, 2024, incorporating the ADS index from Aruoba et al. (2009) (available from the Federal Reserve Bank of Philadelphia), the logarithmic S&P 500 index, and the Treasury yield spread between the 10-year and 3-month rates, all rescaled by a factor of 100, 10, and 100 for comparability, respectively.

We treat Θ_f as known in the Gibbs sampler used to estimate (8). We observe that macro fundamentals f^{T_E} are available only on business days, while polls and prediction market prices are accessible on non-business days as well. To make the most of the available data, we forward-fill macro fundamentals on non-business days (e.g., setting Saturday and Sunday values to those of Friday). Next, we de-mean f^{T_E} to match the estimation sample used in the Gibbs sampler.

In summary, the total number of unknowns—equivalent to the number of parameters that need to be estimated—is $1 + 4N + N(N-1)$ for Θ_d, $1 + N(N-1)$ for Θ_ν, and $1 + N(N-1)$ for Θ_p. This amounts to a total of $3N^2 + N + 3$ parameters, which equals 313 when $N = 10$.

Priors

Due to the large number of parameters we are estimating, the importance of setting appropriate priors is crucial. To ensure that the estimation results are not unduly influenced by our prior choices, we aim to use loose priors whenever possible. However, there are instances where we must impose tighter priors, which we detail below. Even in cases where we use loose priors, we still need to provide informative priors, as described further below.

Priors for the covariance matrices, which are crucial to our study, are set loosely while accounting for the fact that the magnitude of covariance for asset price measurement errors is significantly larger. To provide perspective, the variance of the inverse CDF level of observed asset prices, $\Phi^{-1}(p_t^o)$, is at least 100 times greater than that of the observed poll values, d_t^o.

Although we apply relatively loose priors overall, we impose more informative and restrictive priors specifically for ρ_d and α, which merit further discussion. From the expression for the probability in (5), it is evident that as $T-t$ increases, $\rho_{d,i}^{(T-t)}$ approaches zero, leading to the loss of the relationship between the probability $p_{i,t}$ and the true margin $d_{i,t}$. Given that we are working with daily polls and asset prices, we set priors for $\rho_{d,i}$ to be centered around 0.99, which implies a half-life of approximately 70 days, with a variance of 0.1. This necessitates a stringent imposition, as α should decrease with increasing persistence. Given that a tight near-unit root prior is imposed for $\rho_{d,i}$, it follows that α must be tightly shrunk toward zero. The true margin level, α, is centered around zero with a variance of 0.0001.

Finally, the prior for the bias term in the (transformed) asset prices is normally distributed with a mean of zero and a variance of 1. This prior accommodates both positive and negative values of significant magnitude, suggesting an equal likelihood of interpreting the contract as either a hedge or a risk. This is an important parameter with significant implications for asset pricing; therefore, we impose a loose prior to allow the data to drive the results.

Summary of procedures

We use the Gibbs sampler to estimate the model unknowns that are summarized in (9) and (10). For brevity, a detailed description is provided in the appendix. We provide a brief summary of

the Gibbs sampler here. For the jth iteration,

- Run Kalman smoother to generate $\{s^{T_E}\}^{(j)}$ conditional on $\{f^{T_E}, \Theta_f, \Theta_d, \Theta_\nu, \Theta_p\}^{(j-1)}$: This is explained in Appendix A.3.

- Obtain posterior estimates of $\{\Theta_d, \Theta_\nu, \Theta_p\}^{(j)}$ from the MNIW conditional on $\{s^{T_E}\}^{(j)}$ and f^{T_E}: This is explained in Section A.4. In particular, coefficients subject to restrictions, such as those in Θ_ν and Θ_p are drawn based on Section A.4.1

4 Results

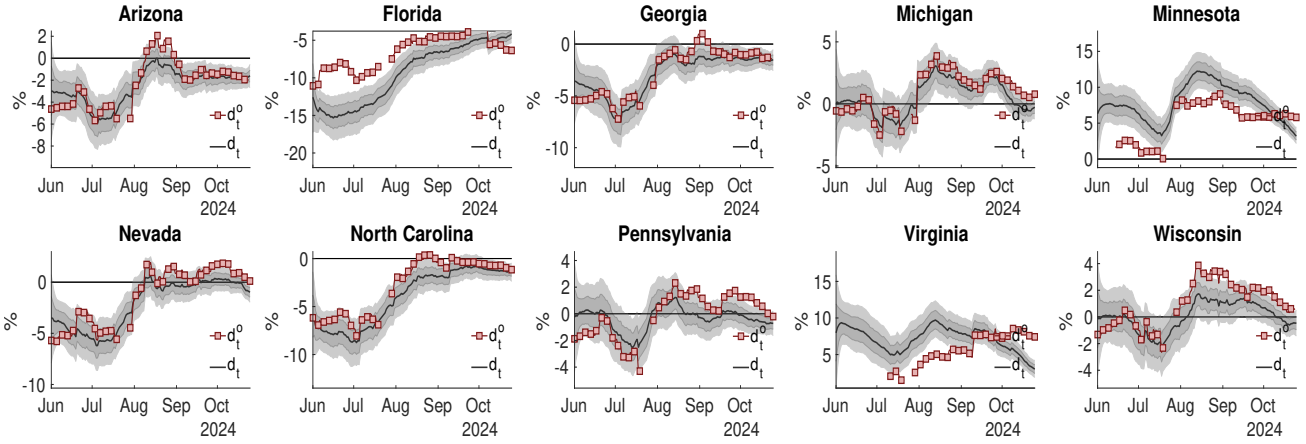
We examine smoothed estimates of the latent variables for each state, potentially influenced by λ_p (a constant bias) and $\epsilon_{p.t}$ (transient, possibly correlated biases). Transient biases across states are absorbed into $\epsilon_{p,t}$. From June 1 to October 25, however, λ_p is statistically indistinguishable from zero, indicating no evidence of a consistent bias across all ten states.[7]

Figure 2 compares the estimated net voter preferences ("NVP"), denoted as d_t, with the observed poll estimates, d_t^o, as described in (1). After accounting for measurement errors, we find that voters in battleground states are more undecided and less volatile in their preference than polling averages suggest, indicating a close race, and one that was less upended by the change in the Democratic candidate than implied by polls. Except for Minnesota and Virginia, which lean Democratic, the 90% credible intervals either include zero or are close to zero, highlighting the competitive nature of the presidential election.

Our main research question is to identify the comovement patterns of voter preferences across states. The model allows for two sources of comovement. First, the β_i in (2) capture the co-movement driven by macroeconomic fundamentals. Second, the u_t shocks in (2) capture the comovement driven by correlated political shocks.

[7]Expanding the sample to include April and May, when polling averages and prediction market prices diverged notably, λ_p becomes significantly positive. If λ_p reflects a risk premium, this result suggests investors may pay above fair probability, indicating a higher marginal utility in scenarios where the Democrat wins and suggesting a negative risk premium that positions the contract as a hedge. These outcomes imply time variation in the bias term, which could be further explored with distinct subsamples—an approach not pursued in this version.

Figure 2: Estimated net voter preferences versus observed poll estimates



Notes: We provide posterior median-smoothed estimates of d_t accompanied by 90% and 60% credible intervals. We overlay with observed poll estimates d_t^o. The relationship between the model-implied d_t and the observed one d_t^o is outlined in (1).

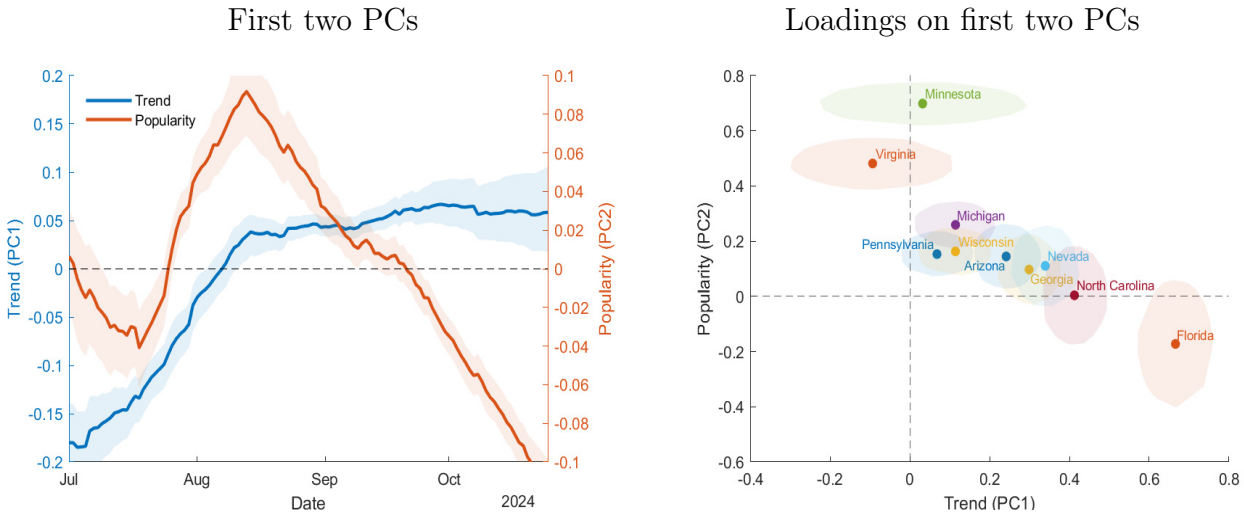
Our estimates attribute a low weight to macroeconomic fundamentals in explaining the co-movement of voter preferences. The posterior median of β_i in (2) is close to zero for all states, and the 60% credible intervals include zero for all states. This suggests that the comovement of voter preferences across states is largely driven by unobserved political shocks, u_t, that are orthogonal to macroeconomic changes.

4.1 Principal component analysis of comovement patterns

To uncover the comovement patterns, we perform a principal component analysis (PCA) on the estimated paths of $d_{i,t}$. The early part of the sample exhibits a high degree of estimation uncertainty and is most influenced by priors on α_i. To avoid our results being overly influenced by the priors, we focus on the period from July 1 onward, estimating the principal components and loadings for each simulated posterior panel of $d_{i,t}$. PCA decomposition is not unique. To aid interpretability, for each panel we flip the signs of the PCs and loadings to ensure that most states have positive loadings.

The right panel of Figure 3 shows the estimates of the first two principal components (PCs), which account for 66% and 27.8% of the variability in d_t across the ten states ($N = 10$), re-

Figure 3: Principal components of net voter preferences d_t



Notes: The left panel plots posterior median-smoothed estimates of the first two principal components (PCs), along with 60% credible intervals. The right panel displays the PC loadings, where the x-coordinate indicate the state's median loading on PC1 and the y-coordinate on PC2. The shaded regions around each point represent the 60% credible regions.

spectively. No other component accounts for more than 2%. The estimated PCs have intuitive interpretations in the context of the 2024 U.S. presidential election. The first PC ("trend") captures the overall trend in voter preferences towards the Democratic candidate. This trend has been consistently increasing, with a moderate flattening in early to mid August. Even though the macroeconomic variables that we chose had little explanatory power, this trend may still be driven by fundamentals and their perception, particularly the abating inflation that voters consistently cited as a top concern. The change in slope occurring in August follows the weaker-than-expected July jobs report released on August 5. In its aftermath, the media devoted considerable coverage to the cooling labor market and a likely Fed rate cut, which may have influenced voter preferences.

The second PC features more variation. It decreases until mid-July, then increases sharply until mid-August, and then decreases again, with reversals being sharp and pronounced. The timing of the sharp July reversal, close to the time when President Biden announced that he would not seek re-election and endorsed Vice President Kamala Harris, invites an interpretation of PC2 as capturing the personal popularity of the Democratic candidate, prompting us to label PC2 as "Popularity". The second reversal comes during a period of heightened media coverage,

occurring shortly after Harris announced her selection of Gov. Tim Walz as her running mate, days before the start of the Democratic National Convention, and around the same time as the release of her economic policy proposals.

The right panel plots the regression coefficients of each state's estimated $d_{i,t}$ value on Trend (x-axis) and Popularity (y-axis) principal components. The points represent median estimates, while the shaded regions around them represent the 60% credible regions. The visualization of PC loadings reveals clusters of closely related states and provides a measure of state similarity that is not driven by a priori assumptions about regional similarities or demographic models of voter preferences.

Consider first the partially overlapping cluster of six states with moderately positive Popularity loadings. It includes the states where the election outcome is most uncertain – the three "Rust Belt" states of Pennsylvania, Wisconsin, and Michigan, along with some "Sun Belt" states of Arizona, Georgia, and Nevada. These six states have similar loadings on Popularity (PC2). In other words, voters in these states were like-minded in their response to the change in the Democratic candidate. One slight exception is Michigan, whose exposure to Popularity is slightly higher than the other five states.

Where these states do differ is in their exposure to Trend (PC1). The three "Rust Belt" states have lower loadings than the three "Sun Belt" states, meaning that the they have been less affected by the overall trend of the election towards the Democratic candidate.

A fourth state commonly included in the "Sun Belt" narrative is North Carolina. However, our estimation shows it to be somewhat distinct from the other three of Arizona, Georgia, and Nevada. First, North Carolina essentially has no exposure to Popularity, meaning that dynamics associated with the personal popularity of the Democratic candidate have not been a driver of voter preferences in the state. Second, North Carolina has a higher exposure to Trend than the other three states, meaning that the overall trend towards the Democratic candidate has been more pronounced in North Carolina.

Finally, the three outliers constitute the three states in our analysis where the election is least competitive. Democrats have pronounced leads in Virginia and Minnesota, while Republicans are likely to win Florida. But not only do these three states have diverging preferences, the

dynamics of these preferences are also distinct. Virginia and Minnesota responded strongly to the personal popularity dynamics captured by PC2, while exhibiting none of the overall drift towards Democrats (or, in Virginia's case, potentially even a contrarian pattern). In contrast, Florida appears to be slightly contrarian with respect to PC2, while also exhibiting a strong overall trend towards the Democratic candidate (albeit from a strongly Republican starting point).

Overall, our estimation effectively uncovers important comovement patterns that reflect fundamental voter preferences, rather than being driven by common shocks to polling errors.

These patterns would not be revealed by PCA performed on polls directly. First, polling data is missing for some states for large parts of the sample. State polls are not conducted frequently enough early in the campaign. Second, their noisiness would render it impossible to decompose comovement into two distinct components – a PCA on forward-filled polling data produces a PC1 that explains 90% of the variation, while subsequent components are essentially noise, all in the low single digits and scattered across states.

Incorporating prediction market data in our analysis and using it to estimate the latent voter preferences is integral to uncovering the two-component patterns described above.

4.2 Simulating possible election outcomes

What are the practical implications of the comovement patterns identified in the previous section?

To answer this question, we consider a simulated distribution of possible election outcomes. The Democratic (Republican) candidate wins a given state when the value of $d_{i,T}$ on election day T is positive (negative). The common factor structure in $d_{i,t}$ implies that the outcomes in different states are correlated, and that the states of the world in which, e.g., the Democrat wins in Pennsylvania are also likely to be the states of the world in which the Democrat wins in Wisconsin.

The goal of our simulation is to assess quantitatively how the state win probabilities – and ultimately, the overall election win probabilities – change conditional on a given state's election outcome. While our simulation produces an unconditional forecast, we are mainly interested in this forecast as a baseline with which to compare the conditional forecasts.

Given our focus, we construct the simulation using the principal components estimated in the

previous section. We can write the principal components rotation as:

$$d_{i,t} = \bar{d}_i + \sum_{j=1}^{10} a_{i,j} x_{j,t}$$

where $\bar{d}_i$ indicate the time-series mean of $d_{i,t}$, $x_{j,t}$ as the j'th principal component, and $a_{i,j}$ denotes the loading of state i on PC j. Our goal will be to simulate $x_{j,T}$, i.e., the value of the principal component j on election day T, and then map it to a distribution of $d_{i,T}$ using the estimated loadings.

In the previous section, we found that the first two principal components account for the vast majority of the variation in $d_{i,t}$. To gauge the importance of comovement, we will consider the two polar cases. First, we will model $d_{i,T}$ as being *entirely* explained by the first two PCs, and later we will contrast these results with a counter-factual in which $d_{i,T}$ are cross-sectionally orthogonal.

Let X_t denote the 2×1 vector of the first two PCs at time t. Our goal is to construct the conditional distribution of X_T given a value of X_t on the last day in our sample. We assume that X_t follows a random walk with innovation covariance given by the sample covariance $\hat{\Sigma}_X = \text{Cov}[X_t - X_{t-1}]$. The choice to model X_t is a random walk as opposed to, say, a VAR, is motivated by two observations. First, the plot of the first two PCs in Figure 3 shows that the PCs are highly persistent and trending, suggesting that a stationary process would likely be misspecified. Second, the martingale property of a random walk yields the simplest possible model for the *average* path of the PCs, avoiding extrapolation from a relatively short sample over which the PCs are computed.

The conditional mean of X_T given X_t is just X_t, and the conditional variance of X_T given X_t is $(T-t)\hat{\Sigma}_X$. Assuming the innovations are jointly normal, we can simulate the final PC values by drawing from the multivariate normal distribution:

$$X_T \sim \mathcal{N}(X_t, (T-t)\hat{\Sigma}_X).$$

To obtain the election outcome in a particular state i, we first construct the final restricted net

Table 2: Democratic candidate state win probabilities

	AZ	FL	GA	MI	MN	NV	NC	PA	VA	WI
Unconditional	0.052	0.004	0.074	0.567	0.981	0.244	0.147	0.232	0.998	0.416
Arizona	1.000	0.064	0.975	1.000	1.000	1.000	0.932	0.983	1.000	1.000
Florida	0.805	1.000	0.976	0.927	1.000	1.000	1.000	0.610	1.000	0.951
Georgia	0.675	0.054	1.000	0.995	1.000	1.000	0.970	0.876	1.000	0.999
Michigan	0.091	0.007	0.130	1.000	1.000	0.407	0.232	0.409	1.000	0.733
Minnesota	0.053	0.004	0.076	0.579	1.000	0.249	0.149	0.237	1.000	0.424
Nevada	0.211	0.017	0.304	0.946	0.999	1.000	0.587	0.646	1.000	0.929
North Carolina	0.327	0.028	0.491	0.894	0.995	0.976	1.000	0.593	0.999	0.878
Pennsylvania	0.218	0.011	0.281	1.000	1.000	0.680	0.375	1.000	1.000	1.000
Virginia	0.052	0.004	0.075	0.568	0.982	0.245	0.147	0.232	1.000	0.417
Wisconsin	0.124	0.009	0.178	0.999	1.000	0.546	0.310	0.558	1.000	1.000

Notes: The top row reports the unconditional win probabilities for the Democratic candidate in each state (column). The subsequent rows report the conditional win probabilities for the Democratic candidate in each state, given that the Democratic candidate wins in a given state (row). Diagonal elements are 1.000 by definition. The probabilities are computed from 10,000 simulations of the model.

voter preference $\tilde{d}_{i,T}$:

$$\tilde{d}_{i,T} = \bar{d}_i + \sum_{j=3}^{10} a_{i,j} x_{j,T} + \sum_{j=1}^{2} a_{i,j} x_{j,T}$$

and then define the binary election outcome as $\mathbb{1}_{\tilde{d}_{i,T}>0}$.[8]

To construct the full simulation, we perform the above procedure 10,000 times with a fixed seed for each draw from the posterior distribution of the model parameters. We then take the median across parameter estimates to get a 10,000 element simulation of election outcomes in each of the 10 modeled states.

Table 2 presents the results. The top row reports the unconditional win probabilities, i.e., $E[\mathbb{1}_{\tilde{d}_{iT}>0}]$. for each state. These results are broadly consistent with the final values of prediction-market prices in the data. Harris, the Democratic candidate, is a strong favorite in Minnesota and Virginia, while the Trump, Republican candidate, is a strong favorite in Florida. The remaining

[8]The $\sum_{j=3}^{10} a_{i,j} x_{j,t}$ term ensures that the starting value of the restricted NVP $\tilde{d}_{i,t}$ is equal to the unrestricted value $d_{i,t}$.

seven states are all competitive, with Harris doing best in Michigan and worst in Arizona. They differ from the latest prediction market prices because the estimated model expects some mean reversion to the 2020 election outcome, while the PC-based simulation, designed for maximum transparency, does not.

The subsequent rows report the win probabilities of Harris in each state, conditional on her winning the state labeled in the first column. If states were independent, these rows would just repeat the top row values. However, the comovement patterns identified in the previous section imply that knowing who won a given state changes expectations of who will win in other states. For example, consider Pennsylvania. Unconditionally, Harris's chances of winning there are 23.2%. But if she wins Wisconsin, her chances of winning Pennsylvania increase to 55.8%. This change is because the two states are similar in their loadings on Trend and Popularity loadings, as shown in Figure 3.

Intuitively, a win in a state that the candidate was already expected to win does not change the overall probabilities much. For instance, the probabilities of Harris winning a state conditional on her winning Virginia do not differ from the unconditional probabilities. In contrast, a win in a state where Harris is a clear underdog, like Florida, leads to a large update. The unlikely states of the world that would need to materialize for her to win Florida would also make her a favorite, or even a lock to win, in all other states.

But there is an additional dimension that drives heterogeneity in the magnitude of conditional updates. Compare Nevada to Pennsylvania, both states that Harris has approximately a 1 in 4 chance of winning. A win in either state is relatively uninformative about her chances of winning in states where she is strong favorite (Minnesota, Virginia) or a massive underdog (Florida). Wins in either state boost Harris's chances of winning Arizona and Georgia by approximately equal amounts, but they are differentially informative about outcomes elsewhere. A win for Harris in Pennsylvania essentially guarantees a win in Wisconsin and Michigan, while a win Nevada merely makes it very likely. This is because, as shown in Figure 3, Pennsylvania is closer to Michigan and Wisconsin in PC loading space than Nevada is. In contrast, a Harris win in Nevada raises her chances of winning North Carolina, a neighbor in PC space if not geographically, from 14.7% to 58.7%, making her favorite to win the Tar Heel state, while a win in Pennsylvania raises the

probability only to 37.5%.

A win in a given state alters the overall election outcome. This happens for two reasons. First, a win in a state increases the number of electoral votes the candidate has, making it more likely that the candidate crosses the threshold of 270 needed to win the overall election. This would be the case even if states were independent. Second, the analysis in this section shows that a win in a state raises the conditional probabilities of winning other states, winning which would also deliver electoral votes to the candidate.

To see how much role comovement plays in the conditional probabilities of overall election outcomes, we aggregate our results. To start, we must take a stand on election outcomes in the 46 states and electoral districts outside of our analysis. Given the choices we made when constructing the sample, these states have the least uncertainty about their eventual winner. Therefore, we assign them either to the Democratic or to the Republican candidate in all simulations based on their latest prediction market prices. If the price of the Harris contract is above 0.8, we assume that she wins it with certainty, and if it is below 0.2, we assume that Trump wins it with certainty. This yields a starting value of 203 electoral votes for Harris, which includes DC and a congressional district in Nebraska where the state is otherwise won by Trump, and 189 for Trump, which includes a congressional district in Maine, a state otherwise won by Harris.

We then compute the total number of electoral votes won by Harris as

$$203 + \sum_{i=1}^{10} \mathbb{1}_{\tilde{d}_{i,T}>0} \times \text{EV}_i$$

where EV_i is the number of electoral votes assigned to state i. We repeat this calculation for each simulation and then compute the fraction of simulations in which each of the candidates wins at least 270 electoral votes. This fraction is the probability of the candidate winning the overall election.

We compare two scenarios in Table 3. The first two columns represent the baseline scenario, in which the simulated election outcomes in each state are entirely driven by the two common factors identified earlier, and hence are correlated. The next two columns present an alternate scenario, in which state outcomes are uncorrelated. To construct this counter-factual, we take

Table 3: Overall election win probabilities

	Baseline		Uncorrelated	
	Harris	Trump	Harris	Trump
Unconditional	0.274	0.726	0.102	0.898
Pennsylvania	1.000	0.945	0.362	0.977
Georgia	1.000	0.784	0.300	0.914
Arizona	1.000	0.765	0.216	0.904
Florida	1.000	0.729	0.769	0.901
Nevada	0.818	0.902	0.132	0.908
North Carolina	0.879	0.830	0.266	0.926
Wisconsin	0.658	1.000	0.184	0.957
Michigan	0.482	0.999	0.159	0.973
Minnesota	0.280	1.000	0.103	0.975
Virginia	0.275	1.000	0.102	0.983

Notes: The table reports overall election win probabilities for each candidate in two scenarios. The Baseline (first two) columns aggregate the state-level results using the PC-based simulation, while the Uncorrelated (last two) columns report the results for a counter-factual simulation in which every state's marginal probabilities are the same as in Baseline, but where realizations are uncorrelated. The top row reports the unconditional win probabilities Harris and Trump, respectively. They add up to 1 by construction. The subsequent rows report the conditional win probabilities for each candidate, given that candidate won in a given state (row). Because they condition on different outcomes (a Harris win vs. a Trump win in the same state), they do not add up to 1. The states are ordered by their "pivotalness," i.e., the degree to which a win in that state alters the overall outcome. The probabilities are computed from 10,000 simulations of the model.

the individual state win probabilities from the baseline model and simulate each state's outcome separately. In effect, each simulation is then a random draw from a binomial distribution with heterogeneous probabilities.

By construction, the two scenarios produce identical numbers of average electoral votes for each candidate. On average, Harris wins 248.6 votes and Trump wins 289.4. With 270 needed for a majority, these means forecast a close election with Trump a more likely victor.

However, correlated state win probabilities predictably lead to a much more dispersed distribution, with the standard deviation in the baseline case of 26.8 being almost double 14.6, its value in the uncorrelated scenario.

A wider distribution yields closer win probabilities. Consider the results reported in Table 3. The first row reports unconditional win probabilities for each of the two candidates under each scenario. In the baseline case, Trump has a 72.6% chance of winning, making him favorite but

allowing for a substantial – 27.4% – chance of a Harris victory. These probabilities are close to the final value from Nate Silver's FiveThirtyEight forecast in the 2016 election, which gave a Trump victory – the realized outcome – a 28.6% chance of occurring. In the uncorrelated case, Trump wins almost 9 out of 10 times, and the outcome appears much more certain.

Subsequent rows report aggregate win probabilities conditional on the given candidate winning the given state. For example, in the baseline scenario, conditional on winning Nevada, Harris wins the election 82% of the time. But if Trump wins Nevada, it is he who wins the election 90% of the time.

Rather than order states alphabetically, we order them in decreasing order by their baseline "pivotalness," which is the sum of the two conditional probabilities for that state. Intuitively, a state is most pivotal if winning it is sufficient to clinch the election. Such a state would have a conditional probability of 1 for each candidate, and therefore a "pivotalness" of 2. The least pivotal state would be one in which a win does not alter the unconditional probabilities at all, leading to a "pivotalness" of 1. Of course, for this to occur in practice, not only would the state outcomes need to be uncorrelated, the state would also need to award 0 electoral votes.

According to this measure, the most pivotal state is Pennsylvania. In our simulation, winning it would be sufficient for Harris to win the overall election, while a Trump win there would give him high odds – 89% – of doing the same. Pennsylvania is pivotal for three reasons. First, it awards 19 electoral votes, a large number. That makes it fairly pivotal even in the uncorrelated scenario, where this reason is the sole contributor. Second, it is close to many other states in PC loading space as shown in Figure 3, making their outcomes highly correlated. Third, it is a somewhat close state, so a win there leads to a large update in win probabilities for correlated states.

The least pivotal states are the least competitive ones. A win in Virginia and Minnesota for Harris is expected, so it causes only small revisions to overall win probabilities. Of course, were the underdog candidate win in these states, the realization of that unlikely scenario would lead to a large revision, but the fact that this only happens for one, not both candidates, keeps the overall "pivotalness" measure low.

Perhaps surprisingly, the third least pivotal state is Michigan. At first glance, it seems similar

to Pennsylvania, the most pivotal state. It also has a large number of electoral votes, 15, is competitive, and is close to Pennsylvania in PC loading space. However, it is unusually favorable to Harris and is more sensitive to the Popularity PC than all other competitive states, meaning that a Harris win there does not lead to as large an update about her aggregate odds of winning in other states as a Pennsylvania win does.

Overall, these results highlight the importance of modeling comovement for forecasting election outcomes and identifying states that are most likely to determine the overall election result.

5 Conclusion

In this paper, we developed a time-series econometric model that integrates polling data, economic fundamentals, and prediction market prices to estimate voter preferences across U.S. states for the 2024 Presidential Election. Our framework allows for the identification of pivotal states and electorally similar clusters, providing insights into how state-level election outcomes are correlated and how they shape the national forecast. We demonstrate that failing to account for these correlations results in meaningfully different predictions, and our model highlights the importance of states like Pennsylvania as key battlegrounds.

A key advantage of our method is that it uses contemporaneous and forward-looking data, which allows us to quantify the similarity of voter preferences across states in the current election without taking a stand on the source of this similarity and without assuming that it persists across election cycles.

While our results offer a robust tool for U.S. presidential election forecasting, there are important caveats. Prediction markets, while rich in information, may still reflect investor biases or manipulation attempts, unrelated to voter behavior.[9] Our model is robust to constant or highly persistent biases across time and states, but we could not consistently identify voter preferences if the biases vary. Our simulation exercises consider two polar cases, in which the dynamics of

[9]In October 2024, there have been media reports of unusually large trades in the Polymarket prediction market for the overall election winner, whose price impact may be evidence of market manipulation. Our model only uses data from the state prediction markets, not the national one. But if the alleged manipulation was sophisticated enough to move state prices in a way consistent with the national market (overall election outcome) through the lens of our model, our estimates would be affected.

voter preferences are either perfectly explained by the two significant principal components with no idiosyncratic shocks, or are entirely idiosyncratic. The true dynamics likely lie somewhere in between. Finally, we found little explanatory power for three specific measures of macroeconomic fundamentals, but other measures or their lags could be more informative. We think that all of these offer promising opportunities for future work.

References

Abramowitz, A. I. (2004). When good forecasts go bad: The time-for-change model and the 2004 presidential election. *PS: Political Science & Politics*, 37(4):745–746.

Arrow, K. J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J. O., Levmore, S., Litan, R., Milgrom, P., Nelson, F. D., Neumann, G. R., Ottaviani, M., Schelling, T. C., Shiller, R. J., Smith, V. L., Snowberg, E., Sunstein, C. R., Tetlock, P. C., Tetlock, P. E., Varian, H. R., Wolfers, J., and Zitzewitz, E. (2008). The promise of prediction markets. *Science*, 320(5878):877–878.

Aruoba, S. B., Diebold, F. X., and Scotti, C. (2009). Real-time measurement of business conditions. *Journal of Business and Economic Statistics*, 27(4):417–427.

Autor, D., Dorn, D., Hanson, G., and Majlesi, K. (2020). Importing political polarization? the electoral consequences of rising trade exposure. *American Economic Review*, 110(10):3139–3183.

Besley, T. and Burgess, R. (2002). The political economy of government responsiveness: Theory and evidence from india. *The quarterly journal of economics*, 117(4):1415–1451.

Besley, T. and Case, A. (2003). Political institutions and policy choices: Evidence from the united states. *Journal of Economic Literature*, 41(1):7–73.

Bonaime, A., Gulen, H., and Ion, M. (2018). Does policy uncertainty affect mergers and acquisitions? *Journal of financial economics*, 129(3):531–558.

Calvo, R., Pons, V., and Shapiro, J. M. (2024). Pitfalls of demographic forecasts of us elections. Working Paper 33016, National Bureau of Economic Research.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81(3):541–553.

Cha, V. and Szechenyi, N. E. (2024). The global impact of the 2024 u.s. presidential election. Contributors: Jon B. Alterman, Ryan C. Berg, Max Bergmann, Jude Blanchette, and others.

Del Negro, M. and Schorfheide, F. (2011). Bayesian Macroeconometrics. In *The Oxford Handbook of Bayesian Econometrics*. Oxford University Press.

Drazen, A. (2008). Political business cycles. *The new Palgrave dictionary of economics*, 2.

Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.

Elliott, G. and Timmermann, A. (2013). *Handbook of economic forecasting*. Newnes.

Fair, R. (2011). *Predicting presidential elections and other things*. Stanford University Press.

Fair, R. C. (1996). Econometrics and presidential elections. *Journal of Economic Perspectives*, 10(3):89–102.

FiveThirtyEight (2024). 2024 election forecast: https://projects.fivethirtyeight.com/2024-election-forecast/. Accessed: 2024-09-27.

Foerster, S. R. and Schmitz, J. J. (1997). The transmission of us election cycles to international stock returns. *Journal of International Business Studies*, 28:1–13.

Giannone, D., Lenza, M., and Primiceri, G. E. (2015). Prior Selection for Vector Autoregressions. *The Review of Economics and Statistics*, 97(2):436–451.

Gulen, H. and Ion, M. (2016). Policy uncertainty and corporate investment. *The Review of financial studies*, 29(3):523–564.

Ikenberry, G. J. (2001). *After Victory: Institutions, Strategic Restraint, and the Rebuilding of Order after Major Wars*. Princeton University Press.

Jennings, W., Lewis-Beck, M., and Wlezien, C. (2020). Election forecasting: Too far out? *International Journal of Forecasting*, 36(3):949–962.

Jennings, W. and Wlezien, C. (2018). Election polling errors across time and space. *Nature Human Behaviour*, 2(4):276–283.

Jens, C. E. (2017). Political uncertainty and investment: Causal evidence from u.s. gubernatorial elections. *Journal of Financial Economics*, 124(3):563–579.

Julio, B. and Yook, Y. (2012). Political uncertainty and corporate investment cycles. *The Journal of Finance*, 67(1):45–83.

Kenett, R. S., Pfeffermann, D., and Steinberg, D. M. (2018). Election polls—a survey, a critique, and proposals. *Annual Review of Statistics and Its Application*, 5(Volume 5, 2018):1–24.

Leigh, A. and Wolfers, J. (2006). Competing approaches to forecasting elections: Economic models, opinion polling and prediction markets. *Economic Record*, 82(258):325–340.

Lewis-Beck, M. S. (2005). Election forecasting: Principles and practice. *The British Journal of Politics and International Relations*, 7(2):145–164.

Meeuwis, M., Parker, J. A., Schoar, A., and Simester, D. (2022). Belief disagreement and portfolio choice. *The Journal of Finance*, 77(6):3191–3247.

National Archives (2024). Electoral college: https://www.archives.gov/electoral-college. Accessed: 2024-09-27.

Nordhaus, W. D. (1975). The political business cycle. *The Review of Economic Studies*, 42(2):169–190.

Pástor, L. and Veronesi, P. (2020). Political cycles and stock returns. *Journal of Political Economy*, 128(11):4011–4045.

Pástor, L. and Veronesi, P. (2021). Inequality aversion, populism, and the backlash against globalization. *The Journal of Finance*, 76(6):2857–2906.

Persson, T. (2002). Do political institutions shape economic policy? *Econometrica*, 70(3):883–905.

Salmon, F. (2009). Recipe for disaster: The formula that killed wall street. *Wired*. Accessed: 2024-09-27.

Silver, N. (2012). *The Signal and The Noise: Why so Many Predictions Fail–But Some Don't*. Penguin Group.

Silver, N. (2024). Nate silver's 2024 presidential election polls and model: https://www.natesilver.net/p/nate-silver-2024-president-election-polls-model. Accessed: 2024-09-27.

Snowberg, E., Wolfers, J., and Zitzewitz, E. (2007). Partisan Impacts on the Economy: Evidence from Prediction Markets and Close Elections*. *The Quarterly Journal of Economics*, 122(2):807–829.

Snowberg, E., Wolfers, J., and Zitzewitz, E. (2008). Prediction Markets: From Politics to Business (and Back). In Hausch, D. and Ziemba, W., editors, *Handbook of Sports and Lottery Markets*, pages 385–402. Elsevier: Handbooks in Finance series.

Snowberg, E., Wolfers, J., and Zitzewitz, E. (2013). Prediction Markets for Economic Forecasting. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting, Volume 2*. Elsevier: Handbooks in Economics series.

Song, D. and Tang, J. (2023). News-driven uncertainty fluctuations. *Journal of Business & Economic Statistics*, 41(3):968–982.

The Economist (2024). How this works: https://www.economist.com/interactive/us-2024-election/prediction-model/president/how-this-works. Accessed: 2024-09-27.

West, M. and Harrison, J. (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media.

Wolfers, J. and Zitzewitz, E. (2006). Interpreting prediction market prices as probabilities.

Appendix

A Gibbs Sampler

We use the Gibbs sampler to estimate the model unknowns. We use the state-space representation from Appendix A.1, with parameters reconfigured as described in Appendix A.2. For the jth iteration,

- Run Kalman smoother to generate $\{d^T, \nu_d^T, u^T, \epsilon_p^T\}^{(j)}$ conditional on $\{f^T, \Theta_f, \Theta_d, \Theta_\nu, \Theta_p\}^{(j-1)}$: This is explained in Appendix A.3.
- Obtain posterior estimates of $\{\Theta_f, \Theta_d, \Theta_\nu, \Theta_p\}^{(j)}$ from the MNIW conditional on $\{d^T, \nu_d^T, u^T, \epsilon_p^T\}^{(j)}$ and f^T: This is explained in Section A.4. In particular, coefficients subject to restrictions, such as those in Θ_ν and Θ_p are drawn based on Section A.4.1

A.1 State-space representation

In the state-space representation of our model, the state-transition equation defines the law of motion for the unobserved factors, while the measurement equation links the observable variables, such as polls d_t^o and transformed asset prices $\Phi^{-1}(p_t^o)$, to the latent states. The representation is as follows:

$$\begin{bmatrix} d_t^o \\ \Phi^{-1}(p_t^o) \end{bmatrix} = \begin{bmatrix} 0 \\ \delta_{p,t} \end{bmatrix} + \begin{bmatrix} I & I & 0 & 0 \\ \gamma_{p,t} & 0 & 0 & I \end{bmatrix} \begin{bmatrix} d_t \\ \nu_{d,t} \\ u_t \\ \epsilon_{p,t} \end{bmatrix}, \tag{A-1}$$

$$\begin{bmatrix} d_t \\ \nu_{d,t} \\ u_t \\ \epsilon_{p,t} \end{bmatrix} = \begin{bmatrix} \delta_{d,t} \\ \mu_{\nu_d} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \rho_d & 0 & 0 & 0 \\ 0 & \rho_{\nu_d} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} d_{t-1} \\ \nu_{d,t-1} \\ u_{t-1} \\ \epsilon_{p,t-1} \end{bmatrix} + \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ I & 0 & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} u_t \\ \epsilon_{d,t} \\ \epsilon_{p,t} \end{bmatrix},$$

where

$$d_t^o = \begin{bmatrix} d_{1,t}^o \\ \vdots \\ d_{N,t}^o \end{bmatrix}, \Phi^{-1}(p_t^o) = \begin{bmatrix} \Phi^{-1}(p_{1,t}^o) \\ \vdots \\ \Phi^{-1}(p_{N,t}^o) \end{bmatrix}, \alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}, \beta' = \begin{bmatrix} \beta_1' \\ \vdots \\ \beta_N' \end{bmatrix}, f_t = \begin{bmatrix} f_{1,t} \\ \vdots \\ f_{N,t} \end{bmatrix}, \mu_{\nu_d} = \begin{bmatrix} \mu_{\nu_d,1} \\ \vdots \\ \mu_{\nu_d,N} \end{bmatrix}$$

$$\sigma_{u^e,i}^2 = \sum_{j=0}^{(T-t)-1} \rho_{d,i}^j (e_i' \Sigma_u e_i)(\rho_{d,i}^j)', \quad \sigma_{w^e,i}^2 = \sum_{j=0}^{(T-t)-1} \left(\sum_{k=0}^{j} \rho_d^{j-k} I \rho_{f,i}^k \right) \Sigma_{w,i} \left(\sum_{k=0}^{j} \rho_d^{j-k} I \rho_{f,i}^k \right)',$$

$$\delta_{p,t} = \begin{bmatrix} \lambda_p + \frac{\alpha_1 \sum_{j=0}^{(T-t)-1} \rho_{d,1}^j + \beta_1' f_{1,t}^e}{(\sigma_{u^e,1}^2 + \beta_1' \sigma_{w^e,1}^2 \beta_1)^{1/2}} \\ \vdots \\ \lambda_p + \frac{\alpha_N \sum_{j=0}^{(T-t)-1} \rho_{d,N}^j + \beta_N' f_{N,t}^e}{(\sigma_{u^e,N}^2 + \beta_N' \sigma_{w^e,N}^2 \beta_N)^{1/2}} \end{bmatrix}, \quad \gamma_{p,t} = \begin{bmatrix} \frac{\rho_{d,1}^{(T-t)}}{(\sigma_{u^e,1}^2 + \beta_1' \sigma_{w^e,1}^2 \beta_1)^{1/2}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\rho_{d,N}^{(T-t)}}{(\sigma_{u^e,N}^2 + \beta_N' \sigma_{w^e,N}^2 \beta_N)^{1/2}} \end{bmatrix},$$

$$\delta_{d,t} = \alpha + \beta' \odot f_t, \quad u_t \sim N(0, \Sigma_u), \quad \epsilon_{d,t} \sim N(0, \Sigma_{\epsilon_d}), \quad \epsilon_{p,t} \sim N(0, \Sigma_{\epsilon_{p,t}}).$$

Here, ρ_d and ρ_{ν_d} denote the persistence matrices in (1) and (2), respectively. The observed macroeconomic factors are modeled using a VAR(1) process

$$f_{i,t} = \rho_{f,i} f_{i,t-1} + w_{i,t}, \quad w_{i,t} \sim N(0, \Sigma_{w,i}), \quad i \in \{1, ..., N\}. \tag{A-2}$$

A.2 Reconfiguration

The state-space representation (A-1) can be expressed generically as follows:

$$\begin{aligned} y_t &= \Lambda_{0,t} + \Lambda_{1,t} s_t, \\ s_t &= \Gamma_{0,t} + \Gamma_1 s_{t-1} + \Omega \varepsilon_{s,t}, \quad \varepsilon_{s,t} \sim N(0, \Sigma_s). \end{aligned} \tag{A-3}$$

The historical time series of a generic variable g_t is expressed as follows:

$$g^T = \{g_1, ..., g_T\}. \tag{A-4}$$

We collect parameters in

$$\Theta_f = \{\rho_f, \Sigma_w\}, \quad \Theta_d = \{\rho_d, \alpha, \beta, \Sigma_u\}, \quad \Theta_\nu = \{\mu_{\nu_d}, \rho_{\nu_d}, \Sigma_{\epsilon_d}\}, \quad \Theta_p = \{\lambda_p, \Sigma_{\epsilon_p}\}. \tag{A-5}$$

A.3 Kalman filter and smoother

We rely on the state-space representation (A-3). Conditional on $f^T, \Theta_d, \Theta_d, \Theta_\nu, \Theta_p$, we apply the standard Kalman filter. Suppose that the distribution of

$$s_{t-1}|\{y^{t-1}, f^{t-1}, \Theta_d, \Theta_d, \Theta_\nu, \Theta_p\} \sim N(s_{t-1|t-1}, P_{t-1|t-1}).$$

Then, the Kalman filter forecasting and updating equations take the form

$$\begin{aligned}
s_{t|t-1} &= \Gamma_{0,t} + \Gamma_1 s_{t-1|t-1} \\
P_{t|t-1} &= \Gamma_1 P_{t-1|t-1} \Gamma_1' + \Omega\Sigma\Omega' \\
s_{t|t} &= s_{t|t-1} + (\Lambda_{1,t} P_{t|t-1})'(\Lambda_{1,t} P_{t|t-1} \Lambda_{1,t}')^{-1} \left(y_t - \Lambda_{0,t} - \Lambda_{1,t} s_{t|t-1}\right) \\
P_{t|t} &= P_{t|t-1} - (\Lambda_{1,t} P_{t|t-1})'(\Lambda_{1,t} P_{t|t-1} \Lambda_{1,t}')^{-1} (\Lambda_{1,t} P_{t|t-1}).
\end{aligned}$$

In turn,

$$s_t|\{y^t, f^t, \Theta_d, \Theta_d, \Theta_\nu, \Theta_p\} \sim N(s_{t|t}, P_{t|t}).$$

Next, the backward smoothing algorithm developed by Carter and Kohn (1994) is applied to recursively generate draws from the distributions $s_t|(s_{t+1:T}, y^T, \Theta_d, \Theta_d, \Theta_\nu, \Theta_p)$ for $t = T-1, T-2, \ldots, 1$. The last element of the Kalman filter recursion provides the initialization for the simulation smoother:

$$\begin{aligned}
s_{t|t+1} &= s_{t|t} + P_{t|t} \Gamma_1' P_{t+1|t}^{-1} \left(s_{t+1} - \Gamma_{0,t} - \Gamma_1 s_{t|t}\right) \\
P_{t|t+1} &= P_{t|t} - P_{t|t} \Gamma_1' P_{t+1|t}^{-1} \Gamma_1 P_{t|t} \\
s_t^j &\sim N(s_{t|t+1}, P_{t|t+1}), \quad t = T-1, T-2, ..., 1.
\end{aligned}$$

A.4 Posterior draws

We treat the smoothed estimates d^T, ν_d^T, u^T, ϵ_p^T from Appendix A.3 as data points. Since all cases can be represented within a VAR framework, we adopt the following generic notation for the VAR model and outline the process for obtaining posterior coefficients.

$$c_t' = \underbrace{\begin{bmatrix} 1' & c_{t-1}' & g_t' \end{bmatrix}}_{w_t'} \underbrace{\begin{bmatrix} \phi_0' \\ \phi_1' \\ \phi_2' \end{bmatrix}}_{\phi'} + \epsilon_{c,t}', \quad \epsilon_{c,t} \sim N(\mathbf{0}, \Sigma). \tag{A-6}$$

Define $X = [c_2, ...c_T]'$, $W = [w_2, ..., w_T]'$, and $\epsilon_c = [\epsilon_{c,2}, ..., \epsilon_{c,T}]'$ conditional on the initial observations. If the prior distributions for ϕ and Σ_c are

$$\phi|\Sigma \sim MN\big(\underline{\phi}, \Sigma \otimes (\underline{V}_\phi \xi)\big), \quad \Sigma \sim IW(\underline{\Psi}, \underline{d}), \tag{A-7}$$

then because of the conjugacy the posterior distributions can be expressed as

$$\phi|\Sigma \sim MN\big(\overline{\phi}, \Sigma \otimes \overline{V}_\phi\big), \quad \Sigma \sim IW(\overline{\Psi}, \overline{d}) \tag{A-8}$$

where

$$\begin{aligned} \overline{\phi} &= \big(W'W + (\underline{V}_\phi \xi)^{-1}\big)^{-1}\big(W'X + (\underline{V}_\phi \xi)^{-1}\underline{\phi}\big), \\ \overline{V}_\phi &= \big(W'W + (\underline{V}_\phi \xi)^{-1}\big)^{-1}, \\ \overline{\Psi} &= (X - W\overline{\phi})'(X - W\overline{\phi}) + (\overline{\phi} - \underline{\phi})'(\underline{V}_\phi \xi)^{-1}(\overline{\phi} - \underline{\phi}) + \underline{\Psi}, \\ \overline{d} &= T - k + \underline{d}, \quad k = dim(\phi). \end{aligned} \tag{A-9}$$

We follow the exposition in Giannone et al. (2015) in which ξ is a scalar parameter controlling the tightness of the prior information in (A-7). For instance, prior becomes more informative when $\xi \to 0$. In contrast, when $\xi = \infty$, then it is easy to see that $\overline{\phi} = \hat{\phi}$, i.e., an OLS estimate.

A.4.1 Posterior draws under restrictions

When ϕ_1 is restricted by $\phi_1 = \rho_1 I$, where ρ_1 is a scalar, (A-6) can be redefined as

$$\tilde{c}_t' \equiv c_t' - \phi_0' - g_t'\phi_2' = c_{t-1}'\phi_1 + \epsilon_{c,t}', \quad \epsilon_{c,t} \sim N(\mathbf{0}, \Sigma). \tag{A-10}$$

Note that (A-10) can be arranged as

$$\underbrace{\begin{bmatrix} \tilde{c}_{1,2}' \\ \vdots \\ \tilde{c}_{1,T}' \\ \vdots \\ \tilde{c}_{N,2}' \\ \vdots \\ \tilde{c}_{N,T}' \end{bmatrix}}_{\tilde{X}} = \rho_1 \underbrace{\begin{bmatrix} c_{1,1}' \\ \vdots \\ c_{1,T-1}' \\ \vdots \\ c_{N,1}' \\ \vdots \\ c_{N,T-1}' \end{bmatrix}}_{\tilde{W}} + \underbrace{\begin{bmatrix} \epsilon_{c,1,2}' \\ \vdots \\ \epsilon_{c,1,T}' \\ \vdots \\ \epsilon_{c,N,2}' \\ \vdots \\ \epsilon_{c,N,T}' \end{bmatrix}}_{\tilde{\varepsilon}_c}, \quad \tilde{\varepsilon}_c \sim N(0, \Sigma \otimes I_{T-1}), \tag{A-11}$$

$$L^{-1}\tilde{X} = L^{-1}\tilde{W} + L^{-1}\tilde{\varepsilon}_c, \quad LL' = \Sigma \otimes I_{T-1}.$$

Then, we can consider $X = L^{-1}\tilde{X}$, $W = L^{-1}\tilde{W}$, and $var(L^{-1}\tilde{\varepsilon}_c) = I_N \otimes I_{T-1}$ and derive (A-8) accordingly.

Remarks. We can apply the same technique for λ_p in (A-1). Let $\tilde{\lambda}_p$ represent the previously drawn posterior. Specifically, define

$$\tilde{c}_{i,t} = \Phi^{-1}(p_t^o) - (\delta_{p,t} - \tilde{\lambda}_p + \gamma_{p,t} d_t) \tag{A-12}$$

to construct

$$\underbrace{\begin{bmatrix} \tilde{c}'_{1,1} \\ \vdots \\ \tilde{c}'_{1,T} \\ \vdots \\ \tilde{c}'_{N,1} \\ \vdots \\ \tilde{c}'_{N,T} \end{bmatrix}}_{\tilde{X}} = \lambda_p \underbrace{\begin{bmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{\tilde{W}} + \underbrace{\begin{bmatrix} \epsilon'_{p,1,1} \\ \vdots \\ \epsilon'_{p,1,T} \\ \vdots \\ \epsilon'_{p,N,1} \\ \vdots \\ \epsilon'_{p,N,T} \end{bmatrix}}_{\tilde{\varepsilon}}, \quad \tilde{\varepsilon} \sim N(0, \Sigma_{\epsilon_p} \otimes I_T), \tag{A-13}$$

$$L^{-1}\tilde{X} = L^{-1}\tilde{W} + L^{-1}\tilde{\varepsilon}, \quad LL' = \Sigma \otimes I_{T-1}.$$

B Simulation of Election Outcomes

We construct the simulation using the principal components estimated in the previous section. We can write the principal components rotation as:

$$d_{i,t} = \bar{d}_i + \sum_{j=1}^{10} a_{i,j} x_{j,t}$$

where $\bar{d}_i$ indicate the time-series mean of $d_{i,t}$, $x_{j,t}$ as the j'th principal component, and $a_{i,j}$ denotes the loading of state i on PC j. Our goal will be to simulate $x_{j,T}$, i.e., the value of the principal component j on election day T, and then map it to a distribution of $d_{i,T}$ using the estimated loadings.

In the paper, we found that the first two principal components account for the vast majority of the variation in $d_{i,t}$. To gauge the importance of comovement, we will consider the two polar cases. First, we will model $d_{i,T}$ as being *entirely* explained by the first two PCs, and later we will contrast these results with a counter-factual in which $d_{i,T}$ are cross-sectionally orthogonal.

Let X_t denote the 2×1 vector of the first two PCs at time t. Our goal is to construct the conditional distribution of X_T given a value of X_t on the last day in our sample. We assume that X_t follows a random walk with innovation covariance given by the sample covariance $\hat{\Sigma}_X = \text{Cov}[X_t - X_{t-1}]$. The choice to model X_t is a random walk as opposed to, say, a VAR, is motivated by two observations. First, the plot of the first two PCs shows that the PCs are highly persistent and that PC1 trends upwards, suggesting that a stationary process would likely be misspecified. Second, the martingale property of a random walk yields the simplest possible model for the *average* path of the PCs, avoiding extrapolation from a relatively short sample over which the PCs are computed.

The conditional mean of X_T given X_t is just X_t, and the conditional variance of X_T given X_t is $(T-t)\hat{\Sigma}_X$. Assuming the innovations are jointly normal, we can simulate the final PC values by drawing from the multivariate normal distribution:

$$X_T \sim \mathcal{N}(X_t, (T-t)\hat{\Sigma}_X).$$

To obtain the election outcome in a particular state i, we first construct the final restricted net voter preference $\tilde{d}_{i,T}$:

$$\tilde{d}_{i,T} = \bar{d}_i + \sum_{j=1}^{2} a_{ij} X_{j,T}$$

and then define the binary election outcome as $\mathbb{1}_{\tilde{d}_{i,T}>0}$.

To see how much role comovement plays in the conditional probabilities of overall election outcomes, we aggregate our results. To start, we must take a stand on election outcomes in the 46 states and electoral districts outside of our analysis. Given the choices we made when constructing the sample, these states have the least uncertainty about their eventual winner. Therefore, we assign them either to the Democratic or to the Republican candidate in all simulations based on their latest prediction market prices. If the price of the Harris contract is above 0.8, we assume that she wins it with certainty, and if it is below 0.2, we assume that Trump wins it with certainty. This yields a starting value of 203 electoral votes for Harris, which includes DC and a congressional district in Nebraska where the state is otherwise won by Trump, and 189 for Trump, which includes a congressional district in Maine, a state otherwise won by Harris.

We then compute the total number of electoral votes won by Harris as

$$203 + \sum_{i=1}^{10} \mathbb{1}_{\tilde{d}_{i,T}>0} \times \mathrm{EV}_i$$

where EV_i is the number of electoral votes assigned to state i. We repeat this calculation for each simulation and then compute the fraction of simulations in which each of the candidates wins at least 270 electoral votes. This fraction is the probability of the candidate winning the overall election.