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

December 30, 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 more than 10 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.

Political prediction markets 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-adjusted 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 develop a formal asset pricing framework to establish each contract's value as a function of state variables. We assume that state variables, which are a blend of latent and macro variables, follow vector autoregression (VAR) dynamics. This approach is similar to that employed in modeling of the term structure of interest rates (e.g., Ang and Piazzesi, 2003). What is different is that, in contrast to yields, probabilities are nonlinear functions of the state variables. This nonlinear connection between prices and state variables serves as an additional measurement equation. We rely on these observation equations 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% 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 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 Nevada and North Carolina appear somewhat distinct from the other two "Sun Belt" states of Arizona and Georgia, with whom they are often grouped. These clusters provide qualified contemporaneous empirical validation to these popularly cited groupings. They 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, less competitive, states have individually distinct patterns.

The informative and intuitive patterns we uncover cannot be revealed by applying PCA directly to polling data. Polling data are sparse for many states over significant portions of the sample, and their noisiness makes such an approach challenging. Even after addressing missing observations through forward filling, the resulting comovement patterns are markedly different, with no evidence of two distinct components characterizing comovements. At the other extreme, applying PCA to asset prices is challenging as well because they are highly nonlinear functions of the model states. The only feasible approach is to incorporate prediction market data alongside observed polls within a state-space framework, as in our analysis. This method is essential for estimating latent voter preferences and uncovering the described two-component structure.

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 baseline 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 25, while the Democratic candidate Kamala Harris is

equally likely to win North Carolina as she is to win Pennsylvania (approximately 11-12%), a win in Pennsylvania is more informative about her chances in other "Rust Belt" states. It essentially guarantees wins in Michigan and Wisconsin while a win in North Carolina merely makes them likely.

Next, 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 as of October 25 show the Republican candidate Donald Trump to be favored to win the election, with a probability of 84.2%. Because correlations increase the dispersion of outcomes, failing to take them into account would substantially raise Trump's odds of victory to 96.8%.

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.

To see how the model forecast evolves over time and responds to new information, we provide a series of 10 evenly spaced snapshots of the forecast as of different dates, starting on July 27 and ending on October 25. Each snapshot uses data on polls, prices, and fundamentals available through that date to estimate and simulate the model, with the last snapshot providing the forecast discussed throughout the paper. Our forecasts differ from prediction market prices both (1) because polls inform estimates of measurement error and risk premia contained in prices and (2) because the low-dimensional structure of the model raises the confidence of the forecast. The increased confidence turns out to be ex-post correct. For each of the ten states as well as for the overall election, the model's forecasted win probability is closer to the actual outcome than the

prediction market price.

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 consid-

erable 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. The October 2024 version of our paper introduced this forecasting model and produced real-time predictions of the U.S. presidential election.[5]

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

[5]This version is available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4998858.

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 sources are FiveThirtyEight and the Silver Bulletin. FiveThirtyEight is 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 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 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 had a value $270 million in October 2024. 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.

Table 1: Summary Statistics

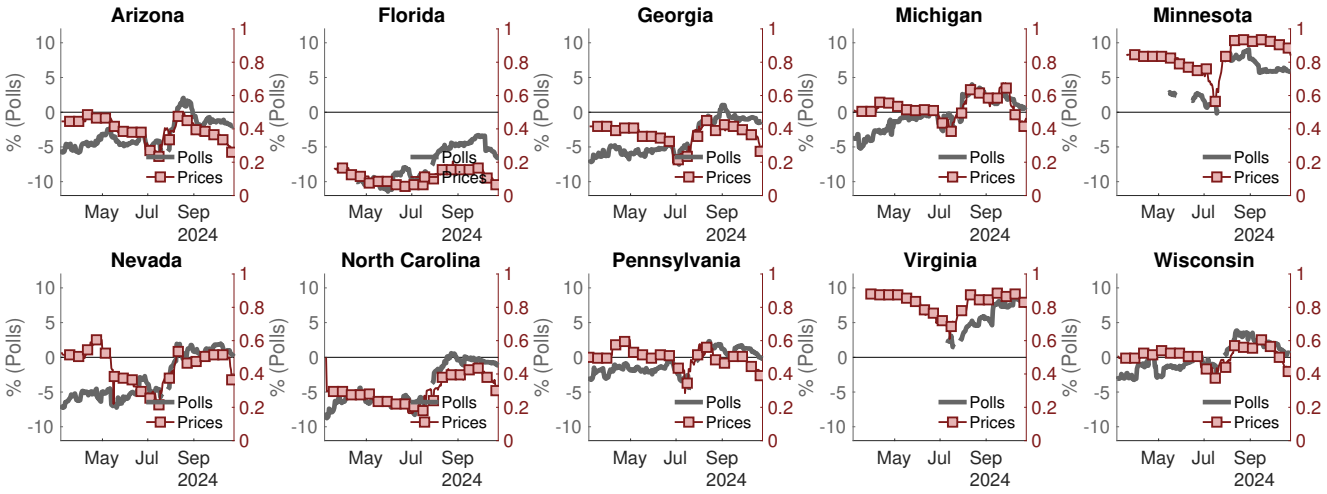
	2020 Results	Prediction Market Prices					Polling Averages				
	(Dem-Rep)	N	Start	End	Mean	Std	N	Start	End	Mean	Std
Arizona	0.30	233	3/07	10/25	39.5	6.5	238	3/01	10/24	-3.2	2.1
Florida	-3.36	221	3/19	10/25	11.3	4.1	191	4/18	10/25	-7.4	2.5
Georgia	0.23	233	3/07	10/25	37.1	6.0	238	3/01	10/24	-3.9	2.5
Michigan	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.

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, 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 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

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.

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

Stack the N-dimensional NVP vector $(d_{1,t}, \ldots, d_{N,t})'$ into $\mathbf{d}_t$ and the M-dimensional vector of fundamentals $(f_{1,t}, \ldots, f_{M,t})'$ into $\mathbf{f}_t$. It is assumed that the NVP process follows:

$$\mathbf{d}_t = \mathbf{a}_d + \Psi_d \mathbf{d}_{t-1} + B\mathbf{f}_t + \mathbf{u}_t, \quad \mathbf{u}_t \sim \mathcal{N}(0, \Sigma_u), \tag{1}$$

where $\mathbf{a}_d = (\alpha_1, \ldots, \alpha_N)'$, $B = (\beta_1', \ldots, \beta_N')'$ is a $N \times M$ matrix of factor loadings, and Ψ_d is a $N \times N$ matrix. The fundamentals process is described with a VAR(1):

$$\mathbf{f}_t = \Psi_f \mathbf{f}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(0, \Sigma_w). \tag{2}$$

For ease of exposition, we stack $\mathbf{d}_t$ and $\mathbf{f}_t$ into $\mathbf{x}_t = (\mathbf{d}_t', \mathbf{f}_t')'$, and write the state dynamics as:

$$\mathbf{x}_t = \underbrace{\begin{bmatrix} \mathbf{a}_d \\ \mathbf{0}_M \end{bmatrix}}_{\mathbf{a}} + \underbrace{\begin{bmatrix} \Psi_d & B\Psi_f \\ \mathbf{0}_{M\times N} & \Psi_f \end{bmatrix}}_{\Psi} \mathbf{x}_{t-1} + \underbrace{\begin{bmatrix} I_n & B \\ \mathbf{0}_{M\times N} & I_m \end{bmatrix} \begin{bmatrix} \mathbf{u}_t \\ \mathbf{w}_t \end{bmatrix}}_{\mathbf{e}_t} \tag{3}$$

where $\mathbf{e}_t$ is mean-zero and has a covariance matrix

$$\Sigma_e = \begin{bmatrix} I_n & B \\ \mathbf{0}_{M\times N} & I_m \end{bmatrix} \begin{bmatrix} \Sigma_u & \mathbf{0}_{N\times M} \\ \mathbf{0}_{M\times N} & \Sigma_w \end{bmatrix} \begin{bmatrix} I_n & B \\ \mathbf{0}_{M\times N} & I_m \end{bmatrix}'.$$

3.2 Prediction market asset price dynamics

A theoretical price of a contract that pays 1 at time T if the Democrat wins state i and 0 otherwise

$$p_{i,t} = E\left[M_{t,T} \mathbb{1}(d_{i,T} > 0) \mid \mathbf{x}_t\right], \tag{4}$$

where $M_{t,T}$ is a stochastic discount factor (SDF) and $\mathbb{1}(\cdot)$ denotes the indicator function. We posit that the one-period ahead SDF $M_{t,t+1}$ is log-affine in the innovations to the state vector $\mathbf{x}_t$:

$$\log M_{t,t+1} = -\frac{1}{2}\Omega_t'\Omega_t - \Omega_t'\Sigma_e^{-\frac{1}{2}}\mathbf{e}_{t+1}.$$

where the price of risk Ω_t is affine in the state vector itself:

$$\Sigma_e^{\frac{1}{2}}\Omega_t = \Omega_0 + \Omega_1 \mathbf{x}_t.$$

Note that this specification assumes $E[M_{t,t+1} \mid \mathbf{x}_t] = 1$, i.e., the short interest rate of zero. This simplifying assumption is consistent with the observation that prices on contracts for states in which a given party's candidate is overwhelmingly likely to win – essentially the risk-free assets in these markets – had a price of 100 cents on the dollar or close to it throughout the sample, without exhibiting a positive trend.

We introduce a risk-adjusted probability P^* associated with the SDF. Continuing with the valuation equation (4),

$$\begin{aligned} p_{i,t} &= E[M_{t,T} \mid \mathbf{x}_t] E\left[\frac{M_{t,T}}{E(M_{t,T} \mid \mathbf{x}_t)} \mathbb{1}(d_{i,T} > 0) \mid \mathbf{x}_t\right], \\ &= E[M_{t,T} \mid \mathbf{x}_t] E^*[\mathbb{1}(d_{i,T} > 0) \mid \mathbf{x}_t] = P^*[d_{i,T} > 0 \mid \mathbf{x}_t], \end{aligned}$$

where the last equality is obtained by imposing our assumption that $E[M_{t,t+1} \mid \mathbf{x}_t] = 1$. In other words, the prediction market price is the risk-adjusted probability of the Democrat winning the state conditional on all information available to investors as of that date.

To construct the risk-adjusted probability, define the state dynamics under P^* as:

$$\mathbf{x}_t = \mathbf{a}^* + \Psi^* \mathbf{x}_{t-1} + \mathbf{e}_t^*,$$

where we guess and verify that $\mathbf{e}^* \sim \mathcal{N}(0, \Sigma_e^*)$. Under this measure, it must be the case that $E[\exp(m_{t,t+1} + x_{t+1}^i) \mid \mathbf{x}_t] = E^*[\exp(x_{t+1}^i) \mid \mathbf{x}_t]$ for each element i of $\mathbf{x}_{t+1}$.

This condition yields

$$\mathbf{a} - \Omega_0 + (\Psi - \Omega_1)\mathbf{x}_t + \frac{1}{2}\Sigma_e = \mathbf{a}^* + \Psi^* \mathbf{x}_t + \frac{1}{2}\Sigma_e^*$$

which implies that

$$\mathbf{a}^* = \mathbf{a} - \Omega_0 \quad \text{and} \quad \Psi^* = \Psi - \Omega_1 \quad \text{and} \quad \Sigma_e^* = \Sigma_e.$$

Let ι_i be a $(N+M)\times 1$ vector with a 1 in the ith position and zeros elsewhere. Since the first N rows of $\mathbf{x}_t$ are NVPs, the time-t price for the i-state contract can be written as $P^*(\iota_i \mathbf{x}_T > 0 \mid \mathbf{x}_t)$. Iterating on the dynamics of $\mathbf{x}_t$:

$$\mathbf{x}_T = \sum_{j=0}^{T-t-1} (\Psi^*)^j \, \mathbf{a}^* + (\Psi^*)^{T-t} \, \mathbf{x}_t + \sum_{j=0}^{T-t-1} (\Psi^*)^j \, \mathbf{e}^*_{T-j}.$$

The first two terms give the conditional mean of $\mathbf{x}_T$ while the variance is given by:

$$\Sigma_{e,T-t} \equiv \text{Var}(\mathbf{x}_T \mid \mathbf{x}_t) = \sum_{j=0}^{T-t-1} (\Psi^*)^j \, \Sigma_e^* \left((\Psi^*)^j \right)'.$$

Vectorizing,

$$\text{vec}\left(\Sigma_{e,T-t}\right) = \sum_{j=0}^{T-t-1} (\Psi^* \otimes \Psi^*)^j \, \text{vec}(\Sigma_e^*)$$

where we use the identities $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$ and $(A \otimes B)^j = A^j \otimes B^j$. Taking the finite sum,

$$\text{vec}\left(\Sigma_{e,T-t}\right) = \left[I_{(N+M)^2} - (\Psi^* \otimes \Psi^*)\right]^{-1} \left[I_{(N+M)^2} - (\Psi^* \otimes \Psi^*)^{T-t}\right] \text{vec}(\Sigma_e^*).$$

Lastly,

$$\iota_i' \Sigma_{e,T-t} \iota_i = \text{vec}\left(\iota_i' \Sigma_{e,T-t} \iota_i\right) = (\iota_i' \otimes \iota_i') \text{vec}\left(\Sigma_{e,T-t}\right)$$

$(\iota_i' \otimes \iota_i')$ is just a row vector with a 1 in the $(M+N) * (i-1) + i$ position and zeros elsewhere.

Plugging into the pricing equation, we have:

$$p_{i,t} = \Phi\left(\frac{\iota_i'\left(\left(I_{N+M} - \Psi^*\right)^{-1}\left(I_{N+M} - \left(\Psi^*\right)^{T-t}\right)\mathbf{a}^* + \left(\Psi^*\right)^{T-t}\mathbf{x}_t\right)}{\sqrt{\text{vec}\left(\Sigma_{e,T-t}\right)_{(M+N)*(i-1)+i}}}\right). \tag{5}$$

3.3 Measurement errors

To take these equations to the data, we posit that polling averages (observed NVP) $\mathbf{d}_t^o$ and prices $\mathbf{p}_t^o$ have measurement errors:

$$\mathbf{d}_t^o = \mathbf{d}_t + \mathbf{v}_{d,t}, \tag{6}$$

$$\mathbf{p}_t^o = \Phi\big(\Phi^{-1}(\mathbf{p}_t) + \mathbf{v}_{p,t}\big), \tag{7}$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution applied to vectors on an element-by-element basis.[6] Polling averages can persist over time, so the measurement error is not necessarily classical. Its dynamics, $\mathbf{v}_t = [\mathbf{v}_{d,t}', \mathbf{v}_{p,t}']'$ are given by a VAR(1):

$$\mathbf{v}_t = \mu_v + \Psi_v \mathbf{v}_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon). \tag{8}$$

Persistent errors pose an identification challenge in the estimation, which the use of prediction market prices can resolve. In the absence of $\mathbf{v}_{p,t}$, the observed price accurately reflects the true risk-adjusted 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 (i.e., $\Omega_0 = \mathbf{0}_{(N+M)\times 1}, \Omega_1 = \mathbf{0}_{(N+M)\times(N+M)}$).

3.4 Estimation

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

[6] For an inverse cdf transformation applied to the data similar to (7), refer to Song and Tang (2023).

of $2N + M$, incorporating both polls and prices as well as M fundamental series. It is crucial to differentiate between the end of the estimation sample, $T_E < T$, and the final election date, T. Given that the election date is November 5, 2024, we set $T = T_E + 11$. The estimation is implemented via Gibbs sampler. See Appendix A.

3.5 Parameter restrictions

To aid identification and enhance the precision of our inference, we introduce the following assumptions about the model parameters.

First, the dynamics of the fundamentals are unaffected by the polls or asset prices, allowing us to estimate them independently from the joint estimation of polls and asset prices. This separation simplifies the process while maintaining the integrity of the overall framework. Consequently, we estimate (2) externally and treat Ψ_f and Σ_w as given. For simplicity, we exclude state-specific fundamental factors and use national ones instead. Specifically, we set $M = 3$ and estimate the fundamental dynamics 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 observe that macro fundamentals 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 the data to match the estimation sample used later. The remaining number of cross-sectional units to estimate internally is $2N = 20$.

Second, the persistence matrices Ψ_d in (1) and Ψ_v in (8) are assumed to be governed by a single scalar persistence parameter, consistent across all states. In particular, we assume that the measurement errors for asset prices are not serially correlated, or equivalently, the scalar governing serial correlation is set to zero for asset prices.

Third, with respect to restrictions on the price of risk, we assume that Ω_0 is governed by a single scalar parameter, $\Omega_{0,i} \equiv \omega_0$ $i = 1, \ldots, N$, and Ω_1 is set to zero, indicating a constant risk premium. The latter significantly simplifies the estimation algorithm, enabling the use of the

Gibbs sampler to sample directly from conditional posteriors.

Fourth, there are two important challenges to consider regarding the true NVP dynamics in (1). Due to the presence of serially correlated measurement errors, the poll-based estimate $\mathbf{d}_t^o$ of the net voter preference cannot alone identify the true NVP $\mathbf{d}_t$. Further, even if the poll-based estimate $\mathbf{d}_t^o$ were free from measurement errors, identifying the level $\mathbf{a}_d$ remains difficult when $\mathbf{\Psi}_d$ is highly persistent, i.e., when its eigenvalues are close to unity. This is where asset prices $\mathbf{p}_t^o$ in (7) become relevant, as they are assumed to be uncorrelated with the poll measurement errors $\mathbf{v}_{d,t}$. Although asset prices reflect risk-adjusted probabilities, the risk adjustment is identical across U.S. states because of our assumption about Ω_0. Consequently, asset prices aid in identifying both the true margin level $\mathbf{a}_d$ and the covariance matrix of shocks $\mathbf{u}_t$ affecting the true margin, Σ_u.

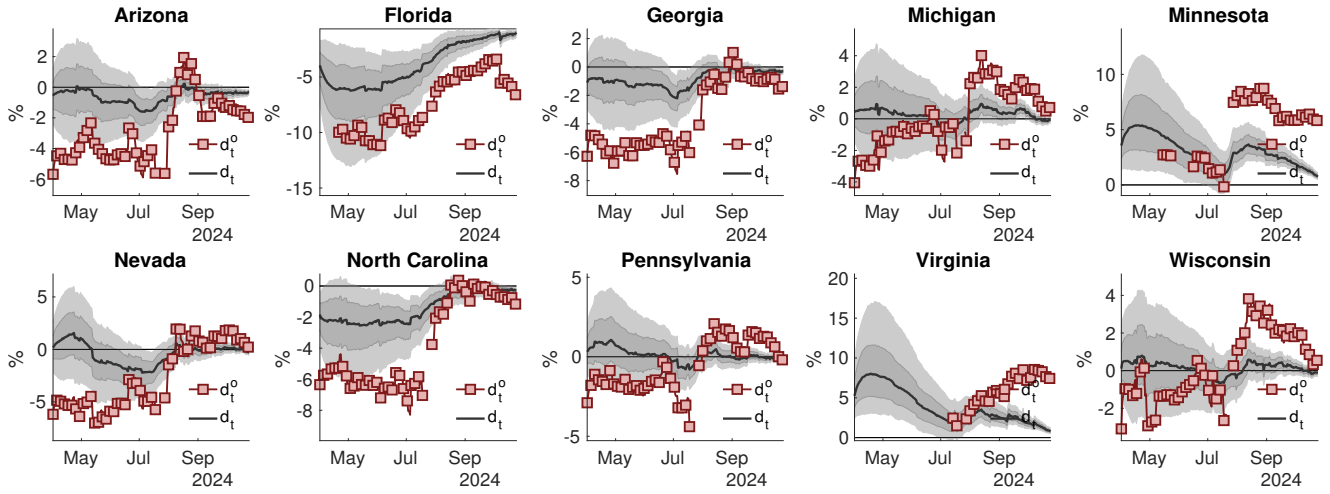
Fifth, there is a related, although not formally an identification problem: our relatively short sample complicates the task of estimating the measurement error component μ_v. These difficulties are exacerbated when the dynamics are highly persistent, as is the case here. To address these issues, we set μ_v to zero.

4 Results

We examine smoothed estimates of the latent variables for each state, potentially influenced by $\omega_0 = \Omega_{0,i}$ (a constant risk premium) and $\mathbf{v}_{p.t}$ (transient, possibly correlated biases). Transient biases across states are absorbed into $\mathbf{v}_{p,t}$. From June 1 to October 25, however, ω_0 is statistically indistinguishable from zero, indicating no evidence of a consistent bias across all ten states.

Figure 2 compares the estimated NVP, $\mathbf{d}_t$, with the observed poll estimates, $\mathbf{d}_t^o$, as described in (6). 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.

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 (6).

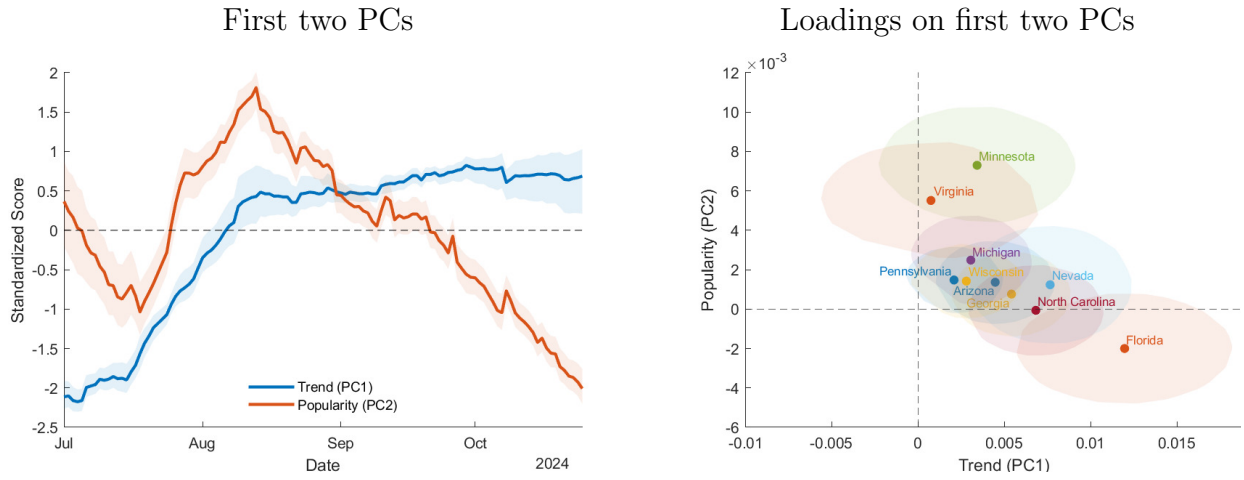
Our main research question is to identify the comovement patterns of voter preferences across states. The model allows for two sources of comovement. First, the B in (1) capture the comovement driven by macroeconomic fundamentals. Second, the $\mathbf{u}_t$ shocks in (1) capture the comovement driven by correlated political shocks.

Our estimates attribute a low weight to macroeconomic fundamentals in explaining the comovement of voter preferences. The posterior median of B in (1) 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, $\mathbf{u}_t$, 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 $a_{d,i}$. 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

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.

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 72.7% and 20.9% of the variability in d_t across the ten states ($N = 10$), respectively. 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 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 seven states with zero to 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 the four "Sun Belt" states of Arizona, Georgia, Nevada, and North Carolina. These seven states have broadly 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, with Michigan's response being strongest while North Carolina's being approximately zero.

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

Two of the states commonly included in the "Sun Belt" narrative are Nevada and North Carolina. However, our estimation shows them to be somewhat distinct from the other two of Arizona and Georgia. 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, both Nevada and North Carolina have a higher exposure to Trend than the other three states, meaning that the overall trend towards the

Democratic candidate has been more pronounced in those two states.

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. 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 forecasted state win probabilities – and ultimately, the overall election win probabilities – change conditional on a given state's

election outcome, i.e., Prob$[d_{i,T} > 0 \mid d_{j,T} > 0$ and $\mathbf{d}_t]$. While our simulation produces a forecast that only conditions on time t information and does not condition on other states' outcomes ("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:

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

where $\bar{d}_i$ indicate the time-series mean of $d_{i,t}$, $z_{j,t}$ 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 $z_{j,T}$, 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 $\mathbf{z}_t$ denote the 2×1 vector of the first two PCs at time t. Our goal is to construct the conditional distribution of $\mathbf{z}_T$ given a value of $\mathbf{z}_t$ on the last day in our sample. We assume that $\mathbf{z}_t$ follows a random walk with innovation covariance given by the sample covariance $\hat{\Sigma}_z = \text{Cov}[\mathbf{z}_t - \mathbf{z}_{t-1}]$. The choice to model $\mathbf{z}_t$ 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 $\mathbf{z}_T$ given $\mathbf{z}_t$ is just $\mathbf{z}_t$, and the conditional variance of $\mathbf{z}_T$ given $\mathbf{z}_t$ is $(T-t)\hat{\Sigma}_z$. Assuming the innovations are jointly normal, we can simulate the final PC values by

drawing from the multivariate normal distribution:

$$\mathbf{z}_T \sim \mathcal{N}(\mathbf{z}_t, (T-t)\hat{\Sigma}_z).$$

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

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

and then define the binary election outcome as $\mathbb{1}(\tilde{d}_{i,T} > 0)$.[7]

To construct the full simulation, we perform the above procedure 10,000 times with a fixed seed for each of the 20,000 draws from the posterior distribution of the model parameters. We then take the median across the 20,000 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., $E[\mathbb{1}(\tilde{d}_{iT} > 0) \mid \mathbf{d}_t]$ for each state. These probabilities are computed by dividing the number of simulation paths in which $\tilde{d}_{iT} > 0$ by the total number of simulation paths. 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 Virginia, while the Trump, Republican candidate, is a prohibitive favorite in Florida. The remaining eight states are all competitive, with Harris doing best in Michigan and worst in Arizona. Section 4.3 below compares these forecasts to prediction market prices.

The subsequent rows report the win probabilities of Harris in each state, additionally conditional on her winning the state labeled in the first column, i.e., $E[\mathbb{1}(\tilde{d}_{i,T} > 0) \mid \tilde{d}_{j,T} > 0 \text{ and } \mathbf{d}_t]$ where i indexes the column and j indexes the row. These probabilities are computed by first selecting only the simulation paths in which $\tilde{d}_{j,T} > 0$ and then calculating the fraction of *those* paths in which $\tilde{d}_{i,T} > 0$ as well. If states were independent, these rows would just repeat the top row values. However, the comovement patterns identified in the previous section imply that

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

Table 2: Democratic candidate state win probabilities

	AZ	FL	GA	MI	MN	NV	NC	PA	VA	WI
Unconditional	0.052	0.007	0.075	0.399	0.903	0.171	0.112	0.122	0.977	0.240
Arizona	1.000	0.102	0.956	1.000	1.000	1.000	0.859	0.973	1.000	1.000
Florida	0.746	1.000	0.944	0.831	0.958	0.986	1.000	0.634	0.958	0.859
Georgia	0.660	0.089	1.000	0.981	0.999	1.000	0.928	0.787	0.999	0.984
Michigan	0.130	0.015	0.184	1.000	1.000	0.406	0.243	0.305	1.000	0.602
Minnesota	0.057	0.008	0.083	0.442	1.000	0.189	0.122	0.135	1.000	0.266
Nevada	0.303	0.041	0.439	0.948	0.998	1.000	0.619	0.574	0.999	0.922
North Carolina	0.397	0.063	0.621	0.864	0.981	0.944	1.000	0.521	0.988	0.838
Pennsylvania	0.413	0.037	0.484	1.000	1.000	0.805	0.479	1.000	1.000	1.000
Virginia	0.053	0.007	0.077	0.408	0.924	0.175	0.113	0.125	1.000	0.246
Wisconsin	0.215	0.025	0.307	0.999	1.000	0.655	0.390	0.507	1.000	1.000

Notes: The top row ("Unconditional") reports the win probabilities for the Democratic candidate in each state (column) conditional only estimated voter preferences through the estimation date (Oct 25). The subsequent rows report the win probabilities for the Democratic candidate in each state (column), additionally conditioning on the Democratic candidate winning in a given state (row). Diagonal elements are 1.000 by definition. The probabilities are computed from 10,000 simulations of the model.

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 is 12.2%. But if she wins Wisconsin, her chances of winning Pennsylvania increase to 50.7%. 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 to win in all other states.

But there is an additional dimension that drives heterogeneity in the magnitude of conditional updates. Compare North Carolina to Pennsylvania, both states that Harris has a slightly higher than 1 in 10 chance of winning. A win in either state is relatively uninformative about her chances of winning in states where she is strong favorite (Virginia) or a massive underdog (Florida). Wins in either state boost Harris's chances of winning Arizona and Minnesota 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 North Carolina merely makes it likely. This is because, as shown in Figure 3, Pennsylvania is closer to Michigan and Wisconsin in PC loading space than North Carolina is. In contrast, a Harris win in North Carolina raises her chances of winning Nevada, a neighbor in PC space if not geographically, from 17.1% to 94.4%, making her favorite to win the desert state, while a win in Pennsylvania raises the probability only to 80.5%. The same is true for Georgia, where Harris's odds of winning are increased more by a win in North Carolina than they are by a win in Pennsylvania.

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

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

where EV_i is the number of electoral votes assigned to state i. We repeat this calculation for each

Table 3: Overall election win probabilities

	Baseline		Uncorrelated	
	Harris	Trump	Harris	Trump
Unconditional	0.158	0.842	0.032	0.968
Pennsylvania	1.000	0.959	0.195	0.990
Georgia	0.992	0.909	0.130	0.976
Arizona	1.000	0.888	0.094	0.972
Florida	0.972	0.848	0.568	0.972
North Carolina	0.845	0.928	0.101	0.977
Nevada	0.786	0.971	0.047	0.971
Wisconsin	0.654	0.999	0.077	0.983
Michigan	0.394	0.998	0.065	0.991
Minnesota	0.175	0.999	0.035	0.994
Virginia	0.162	0.996	0.033	1.000

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 win probabilities Harris and Trump, respectively, conditional only on $\mathbf{d}_t$ but not only outcomes in any particular state. They add up to 1 by construction. The subsequent rows report the win probabilities for each candidate, additionally conditioning on that candidate (column) winning 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.

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 conditional on voter preferences at time t, i.e., Prob[EV $\geq$ 270 | $\mathbf{d}_t$]. To additionally condition on election outcomes in a given state, i.e, to compute Prob[EV $\geq$ 270 | $\tilde{d}_{j,T} > 0$ and $\mathbf{d}_t$], we first select only the simulation paths in which $\tilde{d}_{j,T} > 0$ and then calculate the fraction of *those* paths in which EV $\geq$ 270.

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 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 240.2 votes and Trump wins 297.8. With 270 needed for a majority, these means forecast a fairly close election with Trump a likely victor.

However, correlated state win probabilities predictably lead to a much more dispersed distribution, with the standard deviation in the baseline case of 26.4 being double 13.4, 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 84.2% chance of winning, making him favorite but allowing for a substantial – 15.8% – chance of a Harris victory. For context, this is approximately the probability of a one-day 75 basis point drop in the S&P 500. In the uncorrelated case, Trump wins 96.8% of the time, and the outcome appears nearly 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 79% of the time. But if Trump wins Nevada, it is he who wins the election 97% 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 – 96% – of doing the same. Pennsylvania is pivotal for two 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.

The least pivotal states are the ones where the underdog, Harris, is most likely to win. A win in Virginia and Minnesota for Harris is expected, so it causes only small revisions to overall win probabilities. Of course, were Trump to 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.

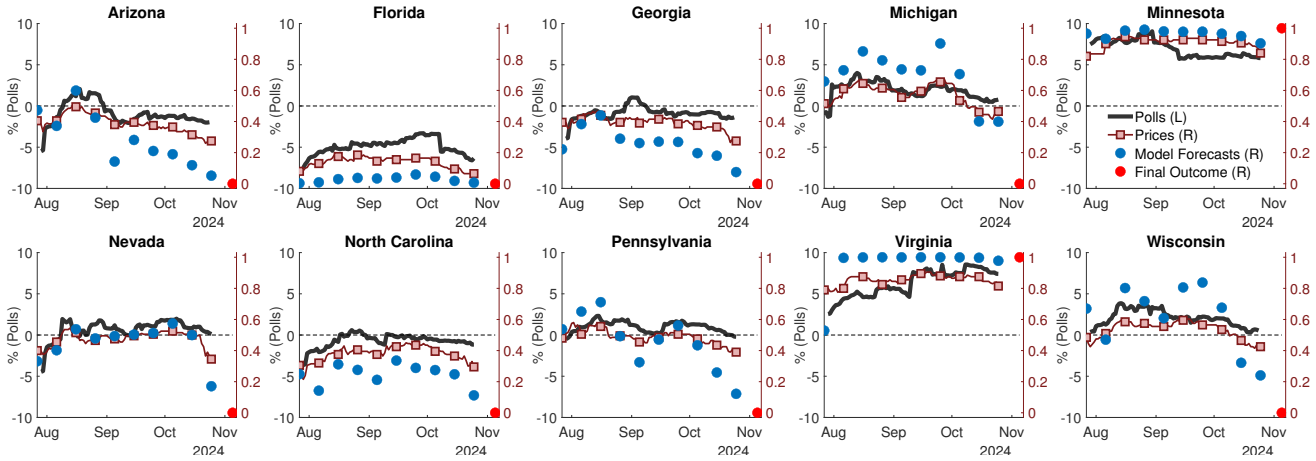
4.3 Assessing the forecasts

Our model uses polls and prediction market prices to generate forecasts for election outcomes in each battleground states as well as a forecast for the overall election outcome. What does the model add? In other words, can it predict the election "better" than alternative models directly based on its individual inputs?

To answer this question, one would ideally want to compare model forecasts to outcomes across both different states and different elections. Unfortunately, state-level prediction markets did not exist or were illiquid for elections prior to 2024, so we cannot re-estimate the model for previous elections. But we can still compare our model's forecasts with polls and prediction market prices over the course of the 2024 campaign.

To accomplish this, we consider 10 dates and re-estimate and re-simulate the model for each date using polling, pricing, and fundamentals data available through that date. The earliest date is July 27, 2024, approximately a week after Kamala Harris replaced Joe Biden as Democratic

Figure 4: Model forecasts vs. polls and prices: By state



Notes: Each panel plots polling averages (gray) and prediction market prices (maroon, squares) for a given state, repeating the information from Figure 1. Additionally, blue dots indicate the model forecast for the Democratic candidate's win probability estimated using data up to the given date. The final outcome of the election – either zero or one – is indicated by a red dot on election date.

nominee. This exercise yields Harris' state-level Prob$[\tilde{d}_{i,T} > 0 \mid \mathbf{d}_s]$ and national win probabilities Prob$[EV \geq 270 \mid \mathbf{d}_s]$ that our model would have generated given the information available as of $s =$ July 27. We repeat this exercise every 10 days, generating forecasts as of $s =$ August 6, August 16, etc., until Oct 25, which uses the full sample used elsewhere in the paper. We compare these forecasts to prediction market prices and polling averages over the same time period, as well as to the final outcome of the election.

Figure 4 plots the results for each state, where the graph of polls and prices from Figure 1 is augmented with model forecasts as well as the final election outcome. The values corresponding to the most recent forecast (Oct 25) are equal to the "Unconditional" row values in Table 2.

The model forecasts differ from prediction market prices on the same date for two reasons. First, prices are assumed to be measuring risk-adjusted probabilities with error, and the mapping from risk-adjusted to true probabilities depends on unobserved prices of risk. Despite also measuring voter preferences with error, polls act as an additional signal informing our estimates of voter preferences. Their contribution is particularly apparent when prediction market prices suggest a toss-up. For example, for much of the campaign, the market price of Harris winning

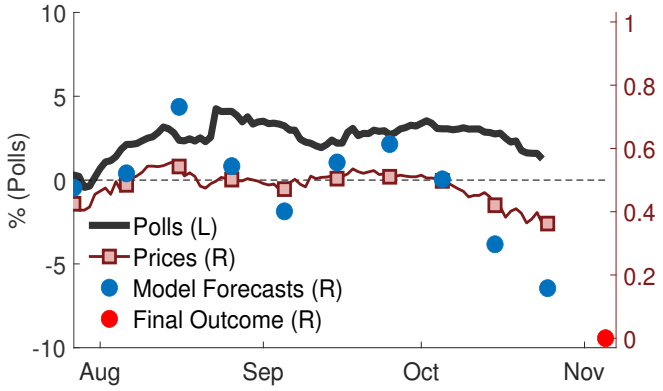
in Nevada was approximately 0.5, while polls gave her a slight lead. As a result, our forecast deviated from market prices slightly in her favor.

Second, the simulated model discards idiosyncratic components of state voter preference dynamics. Principal component analysis revealed that estimated latent voter preferences have essentially a two-factor structure. Remaining variation is small and largely idiosyncratic. While it could indeed reflect state-specific drivers of voter preferences, the presence of measurement error means that these state-specific drivers have a low signal-to-noise ratio. By excluding them from the simulation, we generate a more confident forecast than the prediction markets do in times when a relatively clear favorite emerges in a given state, e.g. in Florida, Minnesota, and Virginia, as well as in North Carolina, Georgia, and Arizona in the second half of the sample. As the election date approaches, the additional confidence resulting from imposing a two-factor structure in simulation increases.

Could this confidence be misplaced? In other words, by discarding higher order PCs, are we perhaps ignoring important state-specific components to voter preferences? With data on just one election, it is hard to answer this question reliably, but election outcomes suggest that we are not. In every instance, our final few forecasts deviate from market prices in the direction of the final outcome. The model gives Trump a higher probability than the market of winning the nine battleground states he won, while it gives Harris a higher probability than the market of winning Virginia, the sole battleground state that went in her favor. Most notably, starting in mid-October, the model gave Trump a substantially higher chance of winning Pennsylvania, the most pivotal state, than the markets did.

Figure 5 plots the result for the overall election outcome. The most recent (Oct 25) forecast value is equal to Harris's baseline model "unconditional" win probability from the top row of Table 3. In this figure, we compare forecasts with national polling data and national prediction market prices, which were *not* used as inputs into the model. National polls have larger sample sizes and smaller reported margins of error than state polls, while the national prediction market had higher liquidity and a substantial advantage in media attention. Therefore, it is possible that our model, based only on state polls and state-specific prediction markets, would be less informative about the final overall outcome, even if it generates robust estimates of comovement

Figure 5: Model forecasts vs. polls and prices: Overall



Notes: National polling averages are in gray and overall election winner prediction market prices are in maroon (squares). Blue dots indicate the model forecast for the Democratic candidate's overall election win probability estimated using data up to the given date. The final outcome of the election – either zero or one – is indicated by a red dot on election date.

between states.

However, that is not the case. First, national polls are relatively uninformative about the overall election outcome because of the Electoral Votes feature of U.S. presidential elections. Winning the majority of votes nationwide does not necessarily give the candidate the overall win, as illustrated by the outcomes of the 2000 and 2016 elections. Figure 5 shows that, while movements in national polls are directionally consistent with changes in market prices and model forecasts, their level consistently suggested a Harris win.

Second, the national prediction market prices suggested a very close election until the last few weeks of the campaign, with prices being between 0.4 and 0.6 for much of the sample. In contrast, our model responds more strongly to the dynamics of the two factors driving estimated voter preferences. For example, Harris's forecasted win probability peaks in the August 16 forecast, around the same time as the "popularity" factor peaks (see Figure 3). Like the state level forecasts, the last pair of national forecasts (ultimately correctly) predict Trump's victory more confidently than markets.

Intuitively, given the relative uninformativeness of national polls, investors trading in the national prediction market need a model of state comovement to map individual state signals into a national forecast. A lack of confidence in such a model would lead to large subjective variances

and bias their forecasts towards 0.5. As our paper argues, such models are hard to estimate, so investors' lack of confidence in their models would be justified. By using only state-level prediction market prices in our model, we are assuming that investors' assessments of voter preferences in a given state can inform forecasts, but that their covemoment models are inferior to the one we propose in this paper. At least in this election, this assumption produced a superior forecast.

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.[8] 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 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

[8]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.

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.

Ang, A. and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary economics*, 50(4):745–787.

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, see Appendix A.1. We use the state-space representation from Appendix A.2, with parameters reconfigured as described in Appendix A.3. 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.4.
- 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.6. In particular, coefficients subject to restrictions, such as those in Θ_ν and Θ_p are drawn based on Section A.6.1

A.1 Model recap

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

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

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

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

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

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 $[u_{1,t}, ..., u_{N,t}]' = u_t \sim N(0, \Sigma_u)$.

The observed macroeconomic factors are modeled using a VAR(1) process

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

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

$$\begin{aligned} p_{i,t} &\equiv P(d_{i,T} > 0 | \mathcal{I}_t), \quad \text{(A-4)} \\ &= P\left(u^e_{i,t+(T-t)} + \beta_i' w^e_{i,t+(T-t)} > -(\rho^{(T-t)}_{d,i} d_{i,t} + \alpha^*_i \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} + \beta_i' f^e_{i,t}) | \mathcal{I}_t \right), \\ &= \Phi\left(\frac{\rho^{(T-t)}_{d,i} d_{i,t} + \alpha^*_i \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} + \beta_i' f^e_{i,t}}{(\sigma^2_{u^e,i} + \beta_i' \sigma^2_{w^e,i} \beta_i)^{1/2}} \right), \end{aligned}$$

which is derived based on (A-5) and (A-6), as provided below, and $\alpha^*_i = \alpha_i - \omega_0$.

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

$$\begin{aligned} d_{i,t+(T-t)} = \rho^{(T-t)}_{d,i} d_{i,t} + \alpha^*_i \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} + \beta_i' \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} \rho^{(T-t-j)}_{f,i} f_{i,t}, \quad \text{(A-5)} \\ + \beta_i' \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} \sum_{k=0}^{(T-t-j)-1} \rho^k_{f,i} w_{i,t+(T-t-j)-k} + \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} u_{i,t+(T-t)-j}. \end{aligned}$$

The exposition in (A-4) relies on the following simplifying notations which are based on (A-5):

$$
\begin{aligned}
f^e_{i,t} &= \sum_{j=0}^{(T-t)-1} \rho^j_{d,i}\rho^{(T-t-j)}_{f,i} f_{i,t}, \qquad \text{(A-6)}\\
u^e_{i,t+(T-t)} &= \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} u_{i,t+(T-t)-j},\\
w^e_{i,t+(T-t)} &= \sum_{j=0}^{(T-t)-1} \rho^j_{d,i} \sum_{k=0}^{(T-t-j)-1} \rho^k_{f,i} w_{i,t+(T-t-j)-k},\\
\sigma^2_{u^e,i} \equiv var(u^e_{i,t+(T-t)}) &= \sum_{j=0}^{(T-t)-1} \rho^j_{d,i}(e_i'\Sigma_u e_i)(\rho^j_{d,i})',\\
\sigma^2_{w^e,i} \equiv var(w^e_{i,t+(T-t)}) &= \sum_{j=0}^{(T-t)-1} \left(\sum_{k=0}^{j} \rho_d^{j-k} I \rho^k_{f,i}\right)\Sigma_{w,i}\left(\sum_{k=0}^{j} \rho_d^{j-k} I \rho^k_{f,i}\right)',
\end{aligned}
$$

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

A.2 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^o_t and transformed asset prices $\Phi^{-1}(p^o_t)$, to the latent states. The representation is as follows:

$$
\begin{bmatrix} d^o_t \\ \Phi^{-1}(p^o_t) \end{bmatrix} = \begin{bmatrix} 0 \\ \delta_{p,t} - \omega_0\xi_{p,t} \end{bmatrix} + \begin{bmatrix} I & I & 0 & 0 \\ \gamma_{p,t} & 0 & 0 & I \end{bmatrix} \begin{bmatrix} d_t \\ \nu_{d,t} \\ u_t \\ \epsilon_{p,t} \end{bmatrix}, \qquad \text{(A-7)}
$$

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

where

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

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

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

$$\gamma_{p,t} = \begin{bmatrix} \frac{\rho_{d,1}^{(T-t)}}{(\sigma^2_{u^e,1} + \beta_1' \sigma^2_{w^e,1} \beta_1)^{1/2}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\rho_{d,N}^{(T-t)}}{(\sigma^2_{u^e,N} + \beta_N' \sigma^2_{w^e,N} \beta_N)^{1/2}} \end{bmatrix},$$

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

Here, ρ_d and ρ_{ν_d} denote the scalar parameters that fully parameterize the persistence matrices, respectively.

A.3 Reconfiguration

The state-space representation (A-7) can be expressed generically as follows:

$$\begin{aligned} y_t &= \Lambda_{0,t} + \Lambda_{1,t} s_t, \\ s_t &= \Gamma_{0,t} + \Gamma_1 s_{t-1} + \Omega \varepsilon_{s,t}, \quad \varepsilon_{s,t} \sim N(0, \Sigma_s). \end{aligned} \tag{A-8}$$

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

$$g^T = \{g_1, ..., g_T\}. \tag{A-9}$$

We collect parameters in

$$\Theta_f = \{\rho_f, \Sigma_w\}, \quad \Theta_d = \{\rho_d, \alpha, \beta, \Sigma_u\}, \quad \Theta_\nu = \{\mu_{\nu_d}, \rho_{\nu_d}, \Sigma_{\epsilon_d}\}, \quad \Theta_p = \{\omega_0, \Sigma_{\epsilon_p}\}. \tag{A-10}$$

A.4 Kalman filter and smoother

We rely on the state-space representation (A-8). 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.5 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 α_d and ρ_d where $\rho_d I_N = \Psi_d$, which merit further discussion. From the expression for the probability in (5), it is evident that as $T-t$ increases, $\rho_d^{(T-t)}$ 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 ρ_d 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_d should decrease with increasing persistence. Given that a tight near-unit root prior is imposed for ρ_d, it follows that α_d must be tightly shrunk toward zero. The true margin level, α_d, is centered around zero with a variance of 0.0001.

Finally, the prior for the price of risk ω_0 where $\Omega_0 = [\omega_0 \mathbf{1}_{1\times N}, \mathbf{0}_{1\times M}]'$ in the 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 parameter plays a crucial role in asset pricing; therefore, we impose a loose prior to allow the data to drive the results.

A.6 Posterior draws

We treat the smoothed estimates d^T, ν_d^T, u^T, ϵ_p^T from Appendix A.4 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.

$$c_t' = \underbrace{\begin{bmatrix} 1' & c_{t-1}' & g_t' \end{bmatrix}}_{w_t'} \underbrace{\begin{bmatrix} \phi_0' \\ \phi_1' \\ \phi_2' \end{bmatrix}}_{\phi'} + \epsilon_{c,t}', \quad \epsilon_{c,t} \sim N(\mathbf{0}, \Sigma). \tag{A-11}$$

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

$$\phi|\Sigma \sim MN\big(\underline{\phi}, \Sigma \otimes (\underline{V}_\phi \xi)\big), \quad \Sigma \sim IW(\underline{\Psi}, \underline{d}), \tag{A-12}$$

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

$$\phi|\Sigma \sim MN\big(\overline{\phi}, \Sigma \otimes \overline{V}_\phi\big), \quad \Sigma \sim IW(\overline{\Psi}, \overline{d}) \tag{A-13}$$

where

$$\begin{aligned} \overline{\phi} &= \big(W'W + (\underline{V}_\phi \xi)^{-1}\big)^{-1}\big(W'X + (\underline{V}_\phi \xi)^{-1}\underline{\phi}\big), \\ \overline{V}_\phi &= \big(W'W + (\underline{V}_\phi \xi)^{-1}\big)^{-1}, \\ \overline{\Psi} &= (X - W\overline{\phi})'(X - W\overline{\phi}) + (\overline{\phi} - \underline{\phi})'(\underline{V}_\phi \xi)^{-1}(\overline{\phi} - \underline{\phi}) + \underline{\Psi}, \\ \overline{d} &= T - k + \underline{d}, \quad k = dim(\phi). \end{aligned} \tag{A-14}$$

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-12). 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.6.1 Posterior draws under restrictions

When ϕ_1 is restricted by $\phi_1 = \rho_1 I$, where ρ_1 is a scalar, (A-11) can be redefined as

$$\tilde{c}_t' \equiv c_t' - \phi_0' - g_t'\phi_2' = c_{t-1}'\phi_1 + \epsilon_{c,t}', \quad \epsilon_{c,t} \sim N(\mathbf{0}, \Sigma). \tag{A-15}$$

Note that (A-15) can be arranged as

$$\underbrace{\begin{bmatrix} \tilde{c}_{1,2}' \\ \vdots \\ \tilde{c}_{1,T}' \\ \vdots \\ \tilde{c}_{N,2}' \\ \vdots \\ \tilde{c}_{N,T}' \end{bmatrix}}_{\tilde{X}} = \rho_1 \underbrace{\begin{bmatrix} c_{1,1}' \\ \vdots \\ c_{1,T-1}' \\ \vdots \\ c_{N,1}' \\ \vdots \\ c_{N,T-1}' \end{bmatrix}}_{\tilde{W}} + \underbrace{\begin{bmatrix} \epsilon_{c,1,2}' \\ \vdots \\ \epsilon_{c,1,T}' \\ \vdots \\ \epsilon_{c,N,2}' \\ \vdots \\ \epsilon_{c,N,T}' \end{bmatrix}}_{\tilde{\varepsilon}_c}, \quad \tilde{\varepsilon}_c \sim N(0, \Sigma \otimes I_{T-1}), \tag{A-16}$$

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

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-13) accordingly.

Remarks. We can apply the same technique for ω_0 in (A-7). Specifically, define

$$\tilde{c}_{i,t} = \Phi^{-1}(p_t^o) - (\delta_{p,t} + \gamma_{p,t} d_t) \tag{A-17}$$

to construct

$$\underbrace{\begin{bmatrix} \tilde{c}'_{1,1} \\ \vdots \\ \tilde{c}'_{1,T} \\ \vdots \\ \tilde{c}'_{N,1} \\ \vdots \\ \tilde{c}'_{N,T} \end{bmatrix}}_{\tilde{X}} = \omega_0 \underbrace{\begin{bmatrix} -\xi'_{p,1,1} \\ \vdots \\ -\xi'_{p,1,T} \\ \vdots \\ -\xi'_{p,N,1} \\ \vdots \\ -\xi'_{p,N,T} \end{bmatrix}}_{\tilde{W}} + \underbrace{\begin{bmatrix} \epsilon'_{p,1,1} \\ \vdots \\ \epsilon'_{p,1,T} \\ \vdots \\ \epsilon'_{p,N,1} \\ \vdots \\ \epsilon'_{p,N,T} \end{bmatrix}}_{\tilde{\varepsilon}}, \quad \tilde{\varepsilon} \sim N(0, \Sigma_{\epsilon_p} \otimes I_T), \tag{A-18}$$

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