Online Appendix to:

The 2020 (Re)Election According to the Iowa Electronic Market: Politics, Pandemic, Recession and/or Protests?

Thomas S. Gruca and Thomas A. Rietz

August 2020

A.I Methods of estimating forecast errors

A. Cross-section method

First, we use the cross-section of forecasts from prior Presidential election markets to estimate average absolute prediction errors and standard deviations in errors at each horizon. This is similar in spirit to Erikson and Wlezien (2008) in that we assume that errors in prior elections help us estimate the size of errors in the current election. At horizon t, they estimate the directional error of polls in prior elections and adjust polls to forecast. Here, we do not make a directional adjustment, but simply use the average absolute error and standard deviation in errors to form confidence intervals around current vote share forecasts. The weakness in this method is that it assumes the distribution of errors is unchanged across elections.

Table A.1 shows the average absolute prediction errors for two-party vote shares by market averaged across various horizons. The average absolute error is quite small (overall daily average = 2.54%) and slightly larger at intermediate horizons than long and short horizons (consistent with Berg, Nelson and Rietz (2008) and Berg and Rietz (2019)).

Table A.1: Average Absolute Prediction Error by Market and Horizon

Market	Days Run	Horizon in Days				
		1 to 100	101 to 200	201 to 300	> 300	Overall
1988	158	3.26%	5.59%			4.11%
1992	294	2.76%	7.69%	4.24%		4.91%
1996	275	4.06%	4.28%	6.65%		4.85%
2000	307	2.62%	2.93%	3.05%	0.99%	2.82%
2004	618	0.88%	1.03%	1.21%	3.06%	2.08%
2008	880	1.10%	1.96%	1.23%	1.20%	1.28%
2012	494	2.03%	1.20%	0.93%	1.83%	1.56%
2016	636	5.16%	4.69%	5.73%	0.84%	2.89%
Weighted Average		2.73%	3.57%	3.79%	1.61%	2.54%

B. Time Series Method

Second, we use a time series method based on an efficient market random walk for VS prices. At each date, the normalized IEM VS forecast, $v_t := \frac{P_t^{UREP20\_VS}}{P_t^{UDEM20\_VS} + P_t^{UREP20\_VS}}$, reveals the forecast distribution mean where P_t = the price at time t for the designated contract. For example, P_t^{UREP20} is the price at time t of the vote share contract associated with the Republican candidate for President in the 2020 election.

In an efficient market, this mean should follow a random walk except for reflecting barriers at 0 and 1. To provide infinite support, define $V_t := \ln\left(\frac{v_t}{1-v_t}\right)$ and assume V_t evolves according to a random walk where $V_{t+1} - V_t = \varepsilon_t$, $\varepsilon_t \sim N(0, \sigma)$. Sigma (σ) is the daily standard deviation of the random walk. Logistic regressions use the same log-odds ratio transform.

At each date, we estimate daily volatility using a 100-day rolling window of past changes in V_t, then project the uncertainty in the *t*-step ahead forecast (assuming no trend and constant daily volatility). There are two weaknesses with this method. First, it assumes volatility (presumably driven by information flow) is constant over time. Second, it does not incorporate the information about the distribution from the WTA market.

C. Implied Volatility Method

A third method assumes consistent pricing across the vote-share and winner-takes-all markets. Again, VS prices reveal forecast distribution means while normalized WTA prices, $w_t := \frac{P_t^{REP20\_WTA}}{P_t^{DEM20\_WTA}+P_t^{REP20\_WTA}}$, reveal probabilities of exceeding 50%. For example, $P_t^{REP20\_WTA}$ is the price at time *t* of the winner-takes-all contract associated with the Republican candidate for President in the 2020 election.

We use v_t (defined above) and w_t to find the mean and implied standard deviation (i.e., "implied volatility") of the log-odds vote-share distribution and back-transform confidence intervals into vote-share space. The advantage of this method is that it is independent of prior markets and outcomes. The current WTA market reveals the uncertainty inherent in the VS forecast. However, there are several disadvantages. It assumes a parametric distribution and assumes the two markets integrate information about it efficiently. It also assumes symmetry, only working when $v_t < 0.5$ and $w_t < 0.5$ or $v_t > 0.5$ and $w_t > 0.5$.

D. Non-parametric method

Finally, we apply Berg, Geweke and Rietz's (2010) non-parametric method that uses bids and asks in both markets (instead of prices) to draw confidence bounds directly from estimated vote-share distributions, assuming informationally integrated markets. This numerical procedure starts with a prior distribution based on the historical distribution of vote-shares and creates a posterior distribution fit to be consistent with all IEM bids and asks while optimizing relative to smoothness and concentration criteria. This retains the advantage that it is independent of prior markets. It allows both the bid-ask spreads and the two markets to tell us about the uncertainty inherent in the forecasts. Further, it does not rely on a parametric distribution. In fact, over much of the 2020 vote share market to date (77% of days), it shows a two-peak forecast distribution. One disadvantage of this method is that, depending on bids and asks, the last-price-determined VS forecast may not lie in the confidence interval bounds.