DDHQ Report

Marc Etter

January 7, 2026

1 Introduction

Multilevel Regression and Post-stratification (MRP) is a widely-used method to estimate public opinion, with far-reaching applications in social science. However, the merits of this method for the specific task of election prediction are unclear. While notable organizations employ MRP in their electoral prognoses (YouGov), some election experts regard MRP as an unreliable tool (Gelman 2019). Is MRP a good method of choice for election prediction, or does a simpler alternative suffice?

To answer this question, this report compares the performance of a baseline multiple linear regression to MRP. The goal is to characterize the extent to which MRP improves upon the baseline model in forecasting victory margins in congressional elections. A secondary but related goal is to assess whether an ensemble of the baseline model and MRP may improve over both models.

2 Data and Methods

I seek to predict the victory margins in congressional districts in 2018 using three models. First, I create a baseline multiple linear regression model that I will use to benchmark all others; then, I employ a Multilevel Regression and Post-stratification model; lastly, I explore whether a weighted average of both of these models may yield superior predictions.

2.1 Baseline Model

Our baseline model is a multiple linear regression of the Republican minus Democratic vote margin in each congressional district. I use data on the geography of congressional districts, incumbency, past congressional district vote margins, and Partisan Voter Index which were provided to me by DecisionDeskHQ for the completion of this project. I impute missing data in the ‘geography‘ variable by a regression of ‘geography‘ on population density.

In addition to the included features, I also create additional features to capture the fundamentals of each House election: the Gallup presidential approval from the last survey fielded in September, Bureau of Labor Statistics Q3 inflation and GDP growth rates, incumbent party, and midterm versus general election indicator variables. Because the outcome variable is the Republican minus Democratic victory margin, I center my additional features (inflation, approval, GDP, etc.), and then multiply each by the incumbent party (-1 = Democrats; 1 = Republicans). The model regression coefficients, which include state fixed-effects, are reported in Table 1.

Table 1: Regression Coefficients of Baseline Model

	Dependent variable:
	Victory Margin (R - D)
Open Seat	13.899***
	(0.891)
Republican Incumbent Running	26.878***
	(0.785)
Extremely Urban	0.236
	(1.455)
Quite Rural	3.146**
	(1.397)
Quite Urban	2.651*
	(1.378)
District PVI	0.152***
	(0.041)
Population/Sq. Mile	−0.0003***
	(0.00005)
District African American %	−11.269***
	(3.142)
District Asian %	−8.226
	(6.675)
District College Educated %	3.643
	(3.712)
District Hispanic %	0.000
	(0.000)
District Non-hispanic White %	−0.023
	(2.463)
Last Pres. Candidate Vote %	−89.295***
	(4.920)
Last R House Candidate Vote %	24.444***
	(2.190)
Gallup Approval × Incumbent	−0.032
	(0.114)
Midterm × Incumbent	−14.416***
	(0.607)
Q3 Infation × Incumbent	−5.688***
	(0.917)
Q3 GDP × Incumbent	4.460***
	(0.369)
Constant	7.314
	(5.606)
Observations	2,280
R^2	0.891
Adjusted R^2	0.888
Residual Std. Error	11.215 (df = 2212)
F Statistic	270.287*** (df = 67; 2212)
Note:	*p<0.1; **p<0.05; ***p<0.01

2.2 MRP Model

I elected to build an MRP model due to the widespread use of this method in estimating sub-national opinion and its natural extension to constituency preferences (Hanretty 2019, YouGov 2024). In 2024, MRP models were used to predict the outcomes of both the US and UK General Elections.

The crux of MRP is the post-stratification of predicted survey response patterns to a target population (Gelman and Little 1997). One creates a model to predict public opinion based on demographics and other co-variates available in a survey (Multi-level Regression), and then weights these predictions by their frequency in sub-national units.

I begin with the Pew Research American Trends Panel Wave 38 (N = 10,682), which I will use as the basis of our predictions of vote choice. The Pew survey asks respondents which party's candidate in their district's upcoming election do they intend to vote for. I recode the responses in three categories: $\{\text{Republican}, \text{Democrat}, \text{Other/Unsure}\}$

In order to predict vote choice in each district, I use a multinomial regression to predict survey respondent vote intention in three categories: Democrat, Other/Unsure, Republican. Our model of vote choice may be expressed as follows:

$$P(y_p = 1) = logit^{-1}(\beta_{0,p} + \alpha^{\text{race}}_{r,p} + \alpha^{\text{age}}_{a,p} + \alpha^{\text{income}}_{i,p} + \alpha^{\text{region}}_{r,p} + \beta_{\text{sex,p}} \cdot \text{sex} + \beta_{\text{party,p}} \cdot \text{party}\ ) \tag{1}$$

where:

$\alpha^{race}_r \sim N(0, \sigma^{race}_r)$ for $r \in \{\text{Black}, \text{White}, \text{Other}\}$

$\alpha^{age}_a \sim N(0, \sigma^{age}_a)$ for $a \in \{< 30, \text{30-49}, \text{50-64}, \geq 65\}$

$\alpha^{income}_i \sim N(0, \sigma^{income}_i)$ for $i \in \{< \$30k, \text{\$30k-\$50k}, \text{\$50k-\$75k}, \text{\$75k-\$100k}, \text{\$100k-\$150k}, \geq \$150k\}$

$\alpha^{region}_r \sim N(0, \sigma^{region}_r)$ for $s \in \{\text{Northeast}, \text{North Central}, \text{South}, \text{West}\}$

In Equation 1, $P(y_p = 1)$ denotes the probability that a voter indicates the will vote for the specified party, for p $\in$ {Democrat, Republican, Other/Unsure}.

After fitting models to prediction vote choice, the MRP model post-stratifies public opinion estimates to the population of registered voters in districts. One typically post-stratifies responses to the general population using population frequencies obtained Census, but in recent years MRP predictions have been made among registered voters, by the use of voter files (Ghitza and Gelman 2020). My MRP uses the L2 Political data to post-stratify predictions according to sex (2 categories), age (4 categories), race (3 categories), income (6 categories), partisanship (3 categories), congressional district, and census region.

Since not all registered voters actually show up to vote, I create a second model that estimates turnout probability. By incorporating these turnout predictions into our MRP model, I hope to be able to better model differential patterns between segments of the registered voter population to better predict vote outcomes. In Equation 2, I predict $P(\text{turnout} = 1)$, denoting the probability that a Pew survey respondent indicates that they will vote in the election. The model for vote turnout follows:

$$P(\text{turnout} = 1) = logit^{-1}(\beta_0 + \alpha^{\text{race}}_r + \alpha^{\text{age}}_a + \alpha^{\text{income}}_i + \alpha^{\text{region}}_r + \beta_{\text{sex}} \cdot \text{sex} + \beta_{\text{party}} \cdot \text{party}\ ) \tag{2}$$

where:

$\alpha^{race}_r \sim N(0, \sigma^{race}_r)$ for $r \in \{\text{Black}, \text{White}, \text{Other}\}$

$\alpha^{age}_a \sim N(0, \sigma^{age}_a)$ for $a \in \{< 30, \text{30-49}, \text{50-64}, \geq 65\}$

$\alpha^{income}_i \sim N(0, \sigma^{income}_i)$ for $i \in \{< \$30k, \text{\$30k-\$50k}, \text{\$50k-\$75k}, \text{\$75k-\$100k}, \text{\$100k-\$150k}, \geq \$150k\}$

$\alpha^{region}_r \sim N(0, \sigma^{region}_r)$ for $s \in \{\text{Northeast}, \text{North Central}, \text{South}, \text{West}\}$

Using Equation 2, I make two adjustments to ensure better calibration with voting trends among registered voters. I adjust vote choice estimates from Equation 1 by multiplying the posterior vote choice probabilities by the posterior turnout probabilities from Equation 2. Additionally, I weight the resulting three MRP response probabilities so that it equals the projected House Popular Vote of the Pew Research survey (39% DEM - 29 % GOP - 32% UNSURE/OTHER).

After careful review of these two adjustments, it does not seem to matter whether I weight by the projected House Popular Vote, or attempt to model voter turnout. The correlation of MRP-projected victory margins with the actual victory margins does not manifestly differ between methods, and all model specifications have a largely similar correlation coefficients and Root Mean Squared Errors.

2.3 Ensemble Model

Lastly, I sought to create an ensemble model that could combine the strength of both models. Anticipating that the few predictors of the MRP model will cause poorly calibrated predictions at the congressional district level, I give one-third weight to the MRP model and two-thirds weight to the baseline model. This model outperforms both the baseline and MRP models, although by a slim margin.

3 Results

I compare all three models to each other, in view of the correlation of their predictions with the outcome, the number of incorrect forecasts, average error in congressional districts, and the projected House seat outcome.

Table 2 indicates that the MRP model performs poorly on its own, with substantially higher RMSE and numbers of incorrect forecasts than the baseline model. The model incorrectly forecasts 77 House districts, and its forecasted House seat count is off by more than forty. However, when I combined both models such that the baseline takes two-thirds weight and the MRP model one-third, the resulting prediction improves slightly over the baseline. These results seem to indicate that the MRP model adds predictive power to the baseline, despite its poor solo performance.

In Figure 1, I show the correlation between each model's forecasted victory margins and the actual margins. The outliers at the top of each scatterplot indicate races where a candidate dropped out at short notice, before our September modeling data could take these developments into account.

Figure 2 illustrates the distribution of GOP seat count from 1000 posterior draws under the MRP model, and the Ensemble model. In the MRP model, the modal simulated outcome is around forty seats away from the actual outcome. Whether this is due to over-representation of Democratic voters in the Pew survey, or an absence of location specific indicators to calibrate regression predictions to House districts, is unclear. On the right panel, we see the distribution of GOP seat count in the Ensemble model. These figures were determined by multiplying the posterior estimate for each congressional district by the corresponding baseline prediction.

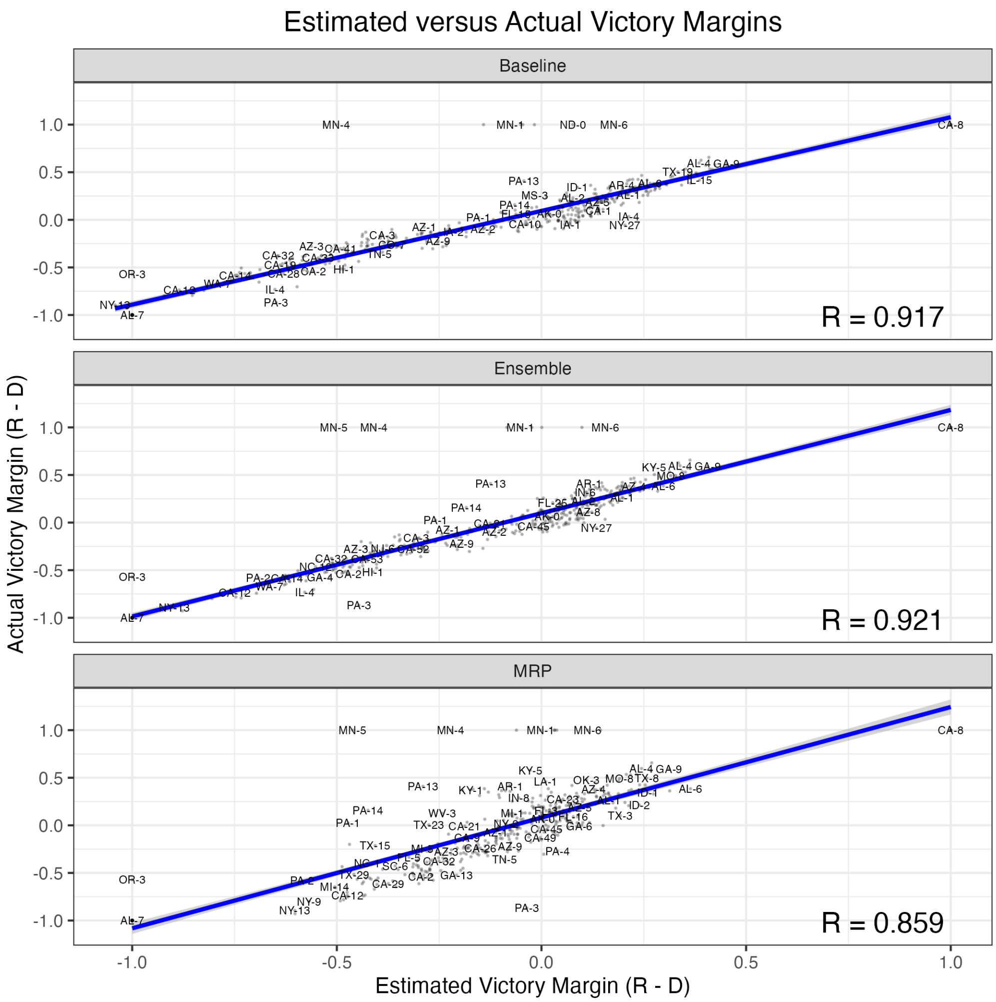

Figure 1: Comparison of the correlation between each model's forecasted victory margins and the actual margins.

Table 2: Model Performance on 2018 U.S. House Elections

Metric	MRP	Baseline	Ensemble
R	0.858	0.917	0.921
Incorrect Forecasts	77	42	40
Predicted GOP Seats	157	196	196
RMSE	0.243	0.203	0.196

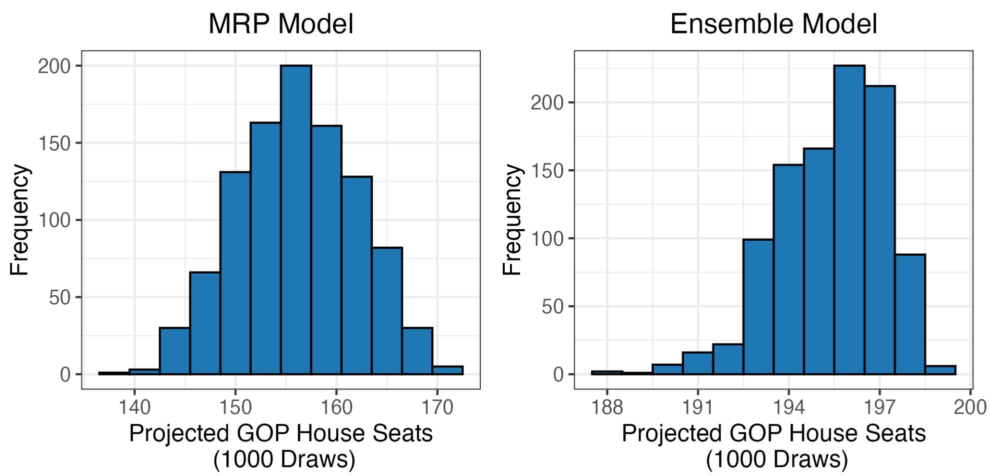

Figure 2: One thousand posterior draws from the MRP and ensemble models. Note that the range of the ensemble model predictions is smaller, due to the one-third weight of the MRP estimates in the ensemble.

4 Discussion

Our results indicate that although the MRP model possessed no information about individual races beyond the demographic composition of electorates, it was nevertheless able to improve predictions when combined with the baseline model.

Still, numerous avenues for improvement of the MRP model are evident. Notably, the primary limitation of the MRP model is its lack of covariates. Beyond demographic variables, only the Census Region of the respondent is available. An ideal model would record the congressional district of the respondents, as this would permit us to include congressional district-specific variables and target predictions at the congressional district level.

The absence of any location-specific variables in the Pew Research data outside of Census Region, places a more detailed model out of reach. In a future version of these models, it may be possible to make several beneficial amendments to the ensemble model. It is possible to acquire data that includes locations, including the Understanding America survey, which reports panelists' state of residence. Secondly, by making MRP predictions in other House elections beyond 2018, it would be possible to incorporate these MRP predictions as a co-variate in the baseline model, potentially improving results.

Although the MRP results I presented fall short of a reliable model, the limitations of the model owed to its data, not to the modeling strategy itself. Even with my limited data, MRP was able to improve upon the baseline model's results in ensemble. With better data, my MRP approach may become a promising method for House election forecasting.

References

[1] Gelman, Andrew. 2019. "MRP is the Carmelo Anthony of election forecasting methods"? So we're doing trash talking now??" *Statistical Modeling, Causual Inference, and Social Sci-*

ence. https://statmodeling.stat.columbia.edu/2019/05/15/we-see-mrp-as-a-way-to-combine-all-the-data-pre-election-voter-file-data-early-voting-precinct-results-county-results-polling-into-a-single-framework/

[2] Bailey, Delia and Douglas Rivers. 2024. "How YouGov's MRP model works for the 2024 U.S. presidential and congressional elections". https://today.yougov.com/politics/articles/50587-how-yougov-mrp-model-works-2024-presidential-congressional-elections-polling-methodology

[3] Ghitza, Yair and Andrew Gelman. 2020. "Voter Registration Databases and MRP: Toward the Use of Large-Scale Databases in Public Opinion Research". *Political Analysis*.

[4] Gelman, Andrew, and Thomas Little. (1997). "Poststratification into Many Categories using Hierarchical Logistic Regression". *Survey Methodology* 23: 127–135.

[5] Hanretty, C. (2019). An Introduction to Multilevel Regression and Post-Stratification for Estimating Constituency Opinion. Political Studies Review, 18(4), 630-645. https://doi.org/10.1177/1478929919864773 (Original work published 2020)