Let Them Eat Pie: Addressing Sample Selection in Multiparty Elections

Ali Kagalwala*
Thiago M. Q. Moreira†
Guy D. Whitten‡
Yongzhi Xu§

Abstract

Elections are central to the study of politics. When studying parties' vote shares across districts, scholars are encouraged to use compositional-outcome models in order to test their theories about what factors shape the dynamics of electoral support. Existing compositional modeling approaches incorrectly deal with scenarios in which not all parties compete in every electoral district. Because unobserved factors that affect a party's decisions to contest a district are likely correlated with its performance in districts where the party fielded candidates, failing to account for partial contestation is likely to result in sample selection bias. Addressing sample selection in a compositional setting is challenging because the outcomes are in log-ratio form, and thus the errors often deviate from normality. To deal with these issues, we introduce a novel maximum likelihood approach which accounts for this type of sample selection and demonstrate through simulations that our method outperforms commonly used solutions, including the conventional Heckman correction. We illustrate the utility of our approach by analyzing the 2017 and 2019 UK parliamentary elections in English constituencies.

Keywords: Sample Selection, Multiparty Elections, Compositional Outcomes, Maximum Likelihood Estimate

Word Count: 5,972

*Postdoctoral Researcher, Institute for Democracy, Journalism and Citizenship, Syracuse University, e-mail: akagalwa@syr.edu

†Assistant Professor, Department of Political Science, Louisiana State University, e-mail: tmoreira@lsu.edu

‡Professor of Political Science and Bob Bullock Chair in Public Policy and Finance, Department of Political Science, Texas A&M University, e-mail: g-whitten@tamu.edu

§Assistant Professor, School of Economics, University of International Business and Economics, e-mail: xuyongzhi6@gmail.com

Introduction

Elections are central to the study of politics. Scholars often analyze the vote shares of competing parties across districts to test theories about what influences election outcomes. It is well-established that results from multiparty elections are compositional data: vote shares range from 0 to 1, and one party's gains/losses are offset by changes in votes for other parties (Katz and King, 1999; Tomz, Tucker and Wittenberg, 2002). Despite this, few studies use compositional strategies to test the impact of predictors on the electoral performance of political parties. The main reason for this is the widespread phenomenon of *partial contestation*: some parties do not field candidates in every district. Existing compositional approaches to studying elections ignore that parties are rational actors which, when constrained by limited resources, strategically choose where to compete. Partial contestation is thus an example of sample selection, and failure to account for it is likely to result in biased and inconsistent estimators (Heckman, 1979).

Addressing sample selection in compositional models presents unique challenges because outcomes are expressed in log-ratio form, and thus errors are typically assumed to follow an extreme value type-I (Gumbel) distribution (Train, 2009, p. 34). This violates the crucial identification assumption of the classic selection approach proposed by Heckman (1979) that the errors are normally distributed. Under these conditions, the Heckman estimator is inefficient. A second identification assumption in the Heckman strategy is the exclusion restriction: at least one predictor in the sample selection equation (also known as an instrument) does not affect the outcome stage. Absent a valid instrument, the Heckman strategy results in severely biased estimates if the distributional assumption about the error terms is not met (Sartori, 2003; Wolfolds and Siegel, 2019; Cook, Lee and Newberger, 2021).

We introduce a novel approach that uses maximum likelihood estimation (MLE) to account for compositional trade-offs between parties' vote shares and sample selection due to partial contestation. MLE offers flexibility in modeling the distributional assumptions for identification and allows us to jointly model the selection process in which parties strategically decide the districts where they will compete and the outcome stage with the vote shares of parties that fielded candidates in a district. Using Monte Carlo experiments designed to reflect multiparty elections with partial contestation, we evaluate our approach against existing strategies to model compositional outcomes. Our analysis focuses on two key selection scenarios: one with a valid exclusion restriction and another in which it is violated. Across both settings, our proposed compositional model with selection (CMS) consistently outperforms competing approaches. Compositional methods that ignore partial contestation produce severely biased estimates, while the Heckman correction is inefficient under a valid exclusion and is both biased and inefficient without a valid instrument.

In the next section, we describe the existing approaches to studying elections and discuss the problem of partial contestation. We then present our strategy to account for sample selection, followed by sections on Monte Carlo experiments and an illustration of the utility of our approach analyzing the 2017 and 2019 elections in English constituencies. In the conclusion, we discuss the importance of correctly modeling sample selection in compositional outcomes and possible applications of our approach to other substantive domains.

Compositional Models & Partial Contestation

Vote shares are inherently compositional data. If v_{ij} represents party j's vote share ($j = 1, 2 \ldots, J$) in electoral district i ($i = 1, 2 \ldots, N$), it follows that

$$0 \leq v_{ij} \leq 1 \; \forall j \text{ and } \sum_{j=1}^{J} v_{ij} = 1. \tag{1}$$

To estimate models with vote shares as the outcome, v_{ij} terms can be expressed as $J - 1$ log-ratios to unbind them from a 0-1 scale to the real number line (Aitchison, 1982). The denominator of each log-ratio is the vote share of an arbitrarily chosen baseline party ($j = 1$) and each numerator is v_{ij} of one of the remaining $J - 1$ parties:[1]

$$s_{ij} = \log\left(\frac{v_{ij}}{v_{i1}}\right) \ \forall\, j \neq 1, \tag{2}$$

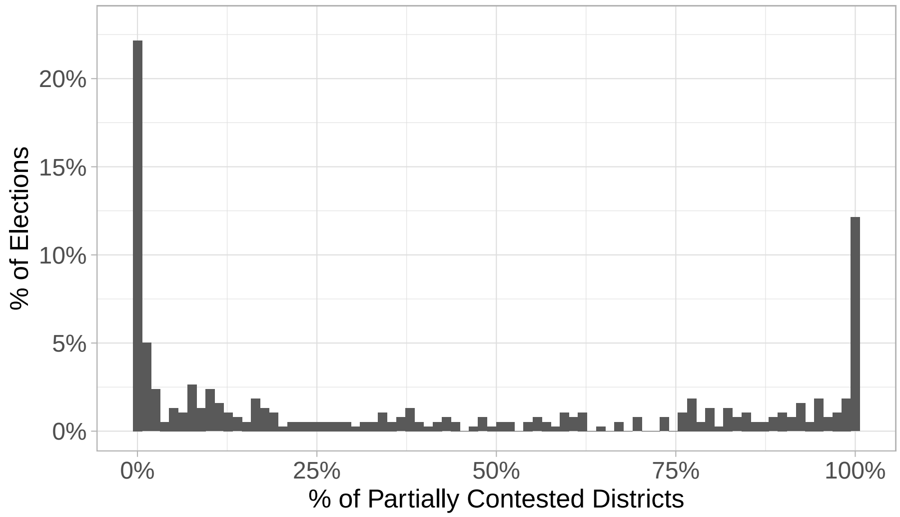
Katz and King (1999) introduced an additive logistic-t model to the study of election outcomes defined as s_{ij}. In 2002, Tomz, Tucker and Wittenberg (TTW) proposed a simpler approach in which they estimated the impact of predictors $\mathbf{x}_{ij}$ on s_{ij} using seemingly unrelated regressions (SUR) (Zellner, 1962). Estimated coefficients from the SUR equations, $\hat{\boldsymbol{\beta}}_j$, can be used for inferential hypothesis tests and to calculate quantities of interest such as predicted vote shares, $\hat{v}_{ij}$. The predictions ($\hat{s}_{ij}$) are translated back into $\hat{v}_{ij}$ using a multivariate logistic transformation (Philips, Rutherford and Whitten, 2016):

$$\begin{aligned} \hat{v}_{ij} &= \frac{\exp(\mathbf{x}'_{ij}\hat{\boldsymbol{\beta}}_j)}{1 + \sum_{j=2}^{J} \exp(\mathbf{x}'_{ij}\hat{\boldsymbol{\beta}}_j)} \ \forall\, j \neq 1, \\ \hat{v}_{i1} &= \frac{1}{1 + \sum_{j=2}^{J} \exp(\mathbf{x}'_{ij}\hat{\boldsymbol{\beta}}_j)} \end{aligned} \tag{3}$$

A straightforward solution to modeling vote shares as compositional outcomes has thus been available for over two decades. Despite this, scholars seldom estimate compositional models to study elections (Kagalwala, Moreira and Whitten, 2024*a*). Even though most theories involve trade-offs across specific parties or ideological families of parties, the dominant approach is to analyze variation in the vote share of one party, typically the incumbent, against all other categories at the national or district level (e.g., Powell Jr and Whitten, 1993; Duch and Stevenson, 2005; Kayser and Peress, 2012; Ward, 2020; Cavallaro, Pregliasco and Vassallo, 2018; Colantone and Stanig, 2018). The most common explanation offered for why scholars choose not to model election results as compositions is the

[1] Compositional outcomes can also incorporate abstention by calculating v_{ij} terms for the entire electorate and including an additional category j for abstention (Horiuchi and Kang, 2018; Moreira, 2025; Moreira, Goidel and Armstrong, 2025).

presence of partial contestation.



Source: Counstituency-Level Elections Archive (CLEA).

Figure 1: Contestation in 377 Elections Across 63 Countries (1970-2019)

In Figure 1, we show that partial contestation is a common occurrence around the world.[2] In a sample of 377 national elections across 63 countries between 1970 and 2019, in only 80 elections (21.1%) all parties contested all districts. More than 50% of districts were partially contested in 40.1% of all cases (152 elections).[3]

Partial contestation creates two major problems. First, the categories of the compositional outcome vary across districts.[4] Second, partial contestation is an example of sample selection. To address the first issue, Katz and King recommended estimating the "effective vote:" v_{ij} we would observe if all J parties had a candidate in all N districts (1999, p. 22). But as TTW noted, this insertion of a party into a race "awards that party more votes than it actually earned, at the expense of parties that truly competed" in that district (2002, p. 69). Instead, they proposed estimating separate SUR models for each pattern of

[2]We created Figure 1 using data on all parties that received at least 5% of the national vote in elections with 50 or more districts. We merged candidates without a party into a single category.

[3]41 elections (10.8%) had all districts partially contested. This happens mostly in countries with regional parties.

[4]Related to this, researchers must choose how to treat parties that are not viable contenders—those that received a tiny percentage of overall votes in the country and did not win a legislative seat. In our empirical example, we classify those parties as a residual category and combine their votes into a j category to calculate s_{ij}s. This is similar to decisions made in the ParlGov project (Döring and Manow, 2012).

contestation, a cumbersome strategy that can become unfeasible and leads to inconclusive analyses as the number of patterns of contestation increases and when there are few districts with a particular pattern of contestation. Kagalwala, Moreira and Whitten (2024*a*) showed that researchers do not follow this approach; rather, when using compositional models to study elections, they employed some form of data modification to address partial contestation. The most common strategy used in political science is to limit the sample to fully contested districts.

Other strategies to get around partial contestation also involve data modification. Pattie and Johnston (2009), for instance, replaced missing v_{ij} terms for parties that did not contest a district with a tiny value, 0.001 (0.1%), of the votes in the district:

$$s^*_{ij} = \begin{cases} \log\left(\frac{v_{ij}+0.001}{v_{i1}}\right) & \text{if } v_{ij} = 0 \\ s_{ij} & \text{if } v_{ij} > 0 \end{cases} \tag{4}$$

This small adjustment avoids zero values in Equation 2 that transforms v_{ij} into log ratios. The authors then use the SUR approach to estimate a single system of equations across all districts. However, this "tiny value" strategy "results in biased estimates, and the magnitude of the bias increases with the share of partially contested districts" (Kagalwala, Moreira and Whitten, 2024*a*, p. 7). This happens because the data adjustment in Equation 4 attenuates the relationship between a predictor of interest and s_{ij}. To address this, for each party that did not contest every district, Kagalwala, Moreira and Whitten (2024*a*) included a dummy variable indicating whether party j contested district i. The dummy variable for j is then interacted with other predictors in j's log-ratio equation in the SUR. This "dummy variable" approach is equivalent to estimating an SUR model with only fully contested districts and avoides the pitfalls of the TTW strategy created by data limitations such as small samples when many constituencies are partially contested. However, this relatively simple strategy does not address the second problem that stems from partially contested elections: sample selection.

Accounting for Sample Selection in Compositional Data

Political parties do not randomly select the districts where they field candidates. Parties are collective entities seeking to maximize their chances of winning legislative seats while constrained by a finite amount of campaign resources (Boix, 2009). Unobservable factors influencing their decision of where to compete are likely correlated with the party's expected performance. For example, a party's reputation and the quality of its candidate in a district should affect both its decision to contest the district and the vote share it receives when it does contest that district (e.g., Cox and Katz, 1996; Butler and Powell, 2014). This is a classic sample selection problem and failing to account for it results in biased parameter estimates (Heckman, 1976, 1979). Heckman (1979) showed that such situations are analogous to an omitted variable problem and can be addressed using a two-stage estimator with selection and outcome equations.

In elections, the first stage models the likelihood that party j will contest district i, and we only observe v_{ij} if j chooses to contest i. This selection mechanism can be defined as:

$$d^*_{ij} = \mathbf{z}'_{ij}\boldsymbol{\alpha}_j + \varepsilon_{ij}, \tag{5}$$

where d^*_{ij} is the latent utility of j contesting i, $\mathbf{z}_{ij}$ is a vector of predictors, and ε_{ij} is the random error. Then, the selection stage we observe is:

$$d_{ij} = \begin{cases} 1 & \text{if } d^*_{ij} \geq 0 \\ 0 & \text{if } d^*_{ij} < 0, \end{cases}$$

Using a single-outcome equation as a general case, the second stage consists of predictors that affect j's electoral performance in a given i. Let this outcome be:

$$v_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta}_j + u_{ij}, \tag{6}$$

where v_{ij} is observed if $d_{ij} = 1$, and is otherwise missing. Heckman (1979) showed that failing to account for sample selection in Equation 6 leads to biased estimates ($\hat{\boldsymbol{\beta}}_j$). His two-step solution is:

1. Estimate Equation 5 with a probit model.
2. Estimate Equation 6 with OLS controlling for the inverse Mills ratio, $\frac{\phi(\mathbf{z}'_{ij}\hat{\alpha}_j)}{\Phi(\mathbf{z}'_{ij}\hat{\alpha}_j)}$, evaluated at the set of observables that influence the selection stage.

This strategy, known as "Heckman correction," uses the inverse Mills ratio from the first stage to correct for sample selection in the second stage. This approach relies on two assumptions for identification: the error terms in both stages are bivariate normally distributed and there is a valid exclusion restriction—i.e., at least one explanatory variable in the selection equation is excluded from the outcome stage (Achen, 1986; Sartori, 2003; Wolfolds and Siegel, 2019; Cook, Lee and Newberger, 2021). Therefore, there must exist one variable in $\mathbf{z}_{ij}$ influencing d^*_{ij} that does not affect v_{ij}. When this exclusion restriction is violated, Heckman estimators are identified solely from distributional assumptions about the residuals, and the estimation procedure works "poorly in practice due to near collinearity" (Achen, 1986, p. 99).[5]

The Heckman correction is inappropriate for compositional models derived from discrete choices (e.g., vote choices when parties partially contest districts) because they typically violate the assumption of normally distributed errors. Instead of a single outcome

[5] To see this, note that if the same predictors in Equation 5 also affect the outcome in Equation 6, the second stage will be $v_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta}_j + \theta\frac{\phi(\mathbf{x}'_{ij}\hat{\alpha}_j)}{\Phi(\mathbf{x}'_{ij}\hat{\alpha}_j)} + u_{ij}$ (Sartori, 2003, p. 115).

equation, the compositional nature of elections implies that the outcome stage is:[6]

$$v_{ij} = \begin{cases} \frac{\exp(s_{ij})}{1+\sum_{j=2}^{J}(\exp(s_{ij}))} \; \forall j \neq 1 & \text{if } d_{ij} = 1 \\ \text{missing} & \text{if } d_{ij} = 0 \end{cases} \tag{7}$$

The log-transformation in Equation 2 imposes that the error terms in log-ratio equations are Gumbel distributed (Aitchison, 1982; Train, 2009).[7] The monotonic transformations of v_{ij} terms change the distribution of the random component (McFadden, 1974, p. 111). This is similar to assumptions about error terms commonly made with multinomial logistic models: the stochastic components are independently and identically Gumbel distributed (Train, 2009, p. 34).[8] Thus, accounting for sample selection by treating u_{ij} and ε_{ij} as jointly normal and using the inverse Mills ratio is no longer an adequate strategy. This will result in inefficiency and, without a valid instrument, in severe bias (Newey, Powell and Walker, 1990; Sartori, 2003; Wolfolds and Siegel, 2019).

To overcome this issue, we propose an MLE approach that accounts for sample selection in a compositional framework derived from a multinomial discrete choice process.[9] The conditional expectation of observing v_{ij} in Equation 7 is:

$$\mathbb{E}[s_{ij}|\mathbf{x}'_{ij}, d_{ij} = 1] = \mathbf{x}'_{ij}\boldsymbol{\beta}_j + \mathbb{E}[u_{ij}|\varepsilon_{ij} \geq -\mathbf{z}'_{ij}\boldsymbol{\alpha}_j], \tag{8}$$

[6]For $j = 1$,

$$v_{i1} = \frac{1}{1 + \sum_{j=2}^{J}(\exp(s_{ij}))}$$

[7]The compositional v_{ij} is a multinomial transformation of the latent utilities of all voters in i, representing the average probability that a share of voters in i chooses j over all other parties. See Appendix A.

[8]Our theory about the latent utility structure and vote-maximizing behavior informs our assumptions about the error distributions. We assume that latent vote shares follow a discrete process (multinomial). Assuming a Gumbel distribution for the errors is thus appropriate. However, assuming Gaussian distributions for the stochastic processes and using Heckman correction might be appropriate if the compositional data are not a function of a discrete-choice mechanism (e.g., cabinet portfolio and budgeting allocations). Our model allows for this type of flexibility.

[9]We are developing an R package to implement our CMS. In the meantime, practitioners can use the R function in our replication files following the instructions provided in Appendix E.

where $\mathbb{E}[u_{ij}|\varepsilon_{ij} \geq -\mathbf{z}'_{ij}\boldsymbol{\alpha}_j] \neq 0$.[10] If u_{ij} and ε_{ij} were jointly normal, then adding the inverse Mills ratio to the right-hand-side of outcome equations could result in consistent parameter estimates (Heckman, 1979). However, since compositional outcomes are transformed to log ratios, u_{ij} and ε_{ij} are assumed to follow a bivariate Gumbel distribution. Their marginal distributions are (Train, 2009, p. 34):

$$f(w) = e^{-\left(w+\gamma+e^{-(w+\gamma)}\right)}, \tag{9}$$

where γ is the Euler–Mascheroni constant. We can then estimate both stages simultaneously with the following likelihood function:

$$\mathcal{L}_j = \prod_{i=1}^{N} \mathbb{P}(\varepsilon_{ij} > -\mathbf{z}'_{ij}\boldsymbol{\alpha}_j)^{d_{ij}} f_{u|\varepsilon}(s_{ij} - \mathbf{x}'_{ij}\boldsymbol{\beta}_j|\varepsilon_{ij} > -\mathbf{z}'_{ij}\boldsymbol{\alpha}_j)^{d_{ij}} \mathbb{P}(\varepsilon_{ij} < -\mathbf{z}'_{ij}\boldsymbol{\alpha}_j)^{1-d_{ij}}, \tag{10}$$

where $\mathbb{P}(\varepsilon_{ij} > -\mathbf{z}'_{ij}\boldsymbol{\alpha}_j)^{d_{ij}}$ is the likelihood that j competes in district i, $f_{u|\varepsilon}(s_{ij} - \mathbf{x}'_{ij}\boldsymbol{\beta}_j|\varepsilon_{ij} > -\mathbf{z}'_{ij}\boldsymbol{\alpha}_j)^{d_{ij}}$ is the likelihood of observing v_{ji} given that j contested i, and $\mathbb{P}(\varepsilon_{ij} < -\mathbf{z}'_{ij}\boldsymbol{\alpha}_j)^{1-d_{ij}}$ is the likelihood that j does not contest i. The log-likelihood is then:[11]

$$\log(\mathcal{L}_j) = \sum_{i=1}^{N} d_{ij} \log\left(\int_{-\mathbf{z}'_{ij}\boldsymbol{\alpha}_j}^{\infty} f_{u\varepsilon}(s_{ij} - \mathbf{x}'_{ij}\boldsymbol{\beta}_j, \varepsilon_{ij}) \mathrm{d}\varepsilon_{ij} \right) + \sum_{i=1}^{N} (1 - d_{ij}) \log(F_{\varepsilon}(-\mathbf{z}'_{ij}\boldsymbol{\alpha}_j)), \tag{11}$$

where F_{ε} is ε_{ij}'s marginal CDF and $f_{u\varepsilon}$ is the joint density of u_{ij} and ε_{ij}. The CDF and PDF of the jointly Gumbel distributed u_{ij} and ϵ_{ij} are respectivelly:

$$F(u, \varepsilon; m) = \exp\left[-\left(e^{-mu} + e^{-m\varepsilon}\right)^{\frac{1}{m}} \right] \tag{12}$$

$$f(u, \varepsilon; m) = f_{u\varepsilon} = F(u, \varepsilon; m) e^{-m(u+\varepsilon)} \left(e^{-mu} + e^{-m\varepsilon}\right)^{\frac{1}{m}-2} . \left[m - 1 + \left(e^{-mu} + e^{-m\varepsilon}\right)^{\frac{1}{m}} \right], \tag{13}$$

where $m = \frac{1}{1-\rho} \in [1, \infty]$, ρ is the correlation between u_{ij} and ε_{ij}, and $m = 1$ when u_{ij}

[10] Appendix A shows the mathematical derivation.
[11] Appendix B shows the derivation of the log-likelihood.

and ε_{ij} are independently distributed.[12] This approach (CMS) allows scholars to fully test hypotheses about trade-offs across parties while correcting for sample selection arising from strategic partial contestation.[13]

Simulations: Sample Selection in Compositional Outcomes

To evaluate the efficacy of our proposed approach, we conduct a series of Monte Carlo experiments. We compare our CMS with the following four alternatives:

1. *Heckman correction*: including an inverse Mills ratio for parties that partially contest electoral districts on the right-hand-side of their respective log-ratio equations.

2. *TTW*: estimating an SUR model for only fully contested districts. Although TTW (2002) recommended estimating a separate SUR model for each pattern of contestation, restricting the sample to fully contested districts is the most common implementation of their strategy.

3. *Tiny value*: adding a tiny value (0.01%) for parties that do *not* contest in a district and estimating SUR models.

4. *Tiny value with dummy variables*: imputing a tiny value (0.01%) for parties that do not contest in a district *and* accounting for partial contestation by including dummy

[12] Although we assume that the errors are jointly Gumbel distributed, our MLE framework allows for other distributional assumptions based on substantive applications. In Appendix D, we provide two examples: one where the DGP follows a jointly normal distribution and we assume extreme value type-I (D.5) and one in which we correctly assume a jointly normal distribution (D.6).

[13] Currently, our MLE approach cannot account for correlated errors across equations in the outcome stage. Greene (2017, p. 333) notes that efficiency gains from an approach that accounts for correlated errors across equations (SUR-GLS) compared to one that does not (OLS) depends on 1) the size of error correlations across equations and 2) the degree of similarity in the matrix of predictors across equations. The larger the error correlations, the greater the efficiency gains from SUR-GLS. The less similar the matrix of predictors, the greater the efficiency gains from SUR-GLS. In Appendix D.8, we assess how SUR-GLS compares to OLS. OLS is the closest comparison to CMS because it 1) is numerically equivalent to MLE and 2) does not account for error correlations. We find that the only efficiency trade-offs are for the estimates of parameters on predictors that differ across equations and that these are relatively mild for the magnitude of error correlations found across equations in our empirical example. Figures S105-S107 are particularly useful for seeing this, showing a difference of only 0.01% in mean squared error (MSE) for a +0.5 error correlation, the highest correlation we found in our empirical applications.

variables for parties that partially contest districts and interacting these with other predictors (Kagalwala, Moreira and Whitten, 2024*a*).

Our data-generating process (DGP) includes three parties $j \in \{A, B, C\}$ contesting elections in N districts ($i = 1, \ldots, N$).[14] B and C contest all N districts, and A fields candidates only in some districts. A's contestation is determined by the following equation:

$$d^*_{iA} = \alpha_{0A} + \alpha_{1A} z_{1i} + \alpha_{2A} z_{2i} + \varepsilon_{iA}, \tag{14}$$

where A contests i if $d^*_{iA} \geq 0$. We vary α_{0A} depending on partially contested districts such that $\alpha_{0A} = 0.57$, 0, or -0.53 when A contests 33%, 50%, or 66% of districts, respectively. $\alpha_{1A} = 0.8$ and $\alpha_{2A} = 0.2$. C is the baseline category in the compositional outcomes:

$$s_{ij} = \log\left(\frac{v_{ij}}{v_{iC}}\right) = \beta_{0j} + \beta_{1j} z_{1i} + \beta_{2j} z_{2i} + u_{ij}, \tag{15}$$

where $\beta_{0A} = 1$, $\beta_{1A} = 1.2$, $\beta_{0B} = 1$, $\beta_{1B} = -1$, and $\beta_{2A} = \beta_{2B} = 0$ for our initial set of experiments. Additionally, u_{iA} and ε_{iA} are jointly generated from mean-zero extreme value type-I distributions (Train, 2009; McFadden, 1974).[15] As per Heckman (1979), failing to account for the correlation between u_{iA} and ε_{iA} can bias the estimate for β_{1A}. z_{2i} influences sample selection, but is unrelated to the outcome, thus satisfying the exclusion restriction. In another scenario, we examine the consequences of violating the exclusion restriction by setting $\beta_{2A} = 1$ and estimating models without z_{2i} in the outcome stage.[16]

We vary the number of electoral districts, $N \in \{100, 200, 500, 800\}$. The correlation between ε_{iA} and u_{iA} (ρ) varies from 0.1 to 0.9. Finally, the probability that A will partially contest districts is set to 33%, 50%, or 66%, which means that A contests $\frac{2}{3}$, $\frac{1}{2}$, or $\frac{1}{3}$ of all

[14] Appendix D.7 shows results with six parties.

[15] We also investigate what happens when errors are jointly normal. While Heckman's correction is unbiased, the CMS returns slightly biased results (overestimates β_{1A} by 4%-8%), but performs much better than approaches that ignore sample selection (Appendix D.5). In the package we are developing, practitioners will be able to choose whether the errors are assumed to be jointly normally distributed (Appendix D.6) or to follow a bivariate Gumbel distribution.

[16] In this Monte Carlo experiment, we wrongly assume that z_{2i} is a valid instrument and thus is wrongly omitted from the outcome equation. Since $\text{Cor}(z_{2i}, z_{1i}) = 0$, omitting z_{2i} in Equation 15 does not imply omitted variable bias.

districts in our simulations.[17] To assess how sample selection affects the performance of the relevant estimators, we present results for β_{1A}.[18]

Results: Valid Exclusion Restriction

Figure 2 shows results for bias across approaches when there is a selection process, z_{2i} is a valid instrument ($\beta_{2A} = 0$), and A does not contest 50% of the districts.[19] The four panels indicate different sample sizes, the horizontal axis is ρ ($\text{Cor}(\varepsilon_{iA}, u_{iA})$), and the vertical axis is the average bias. We find that the estimators for all three modeling strategies that

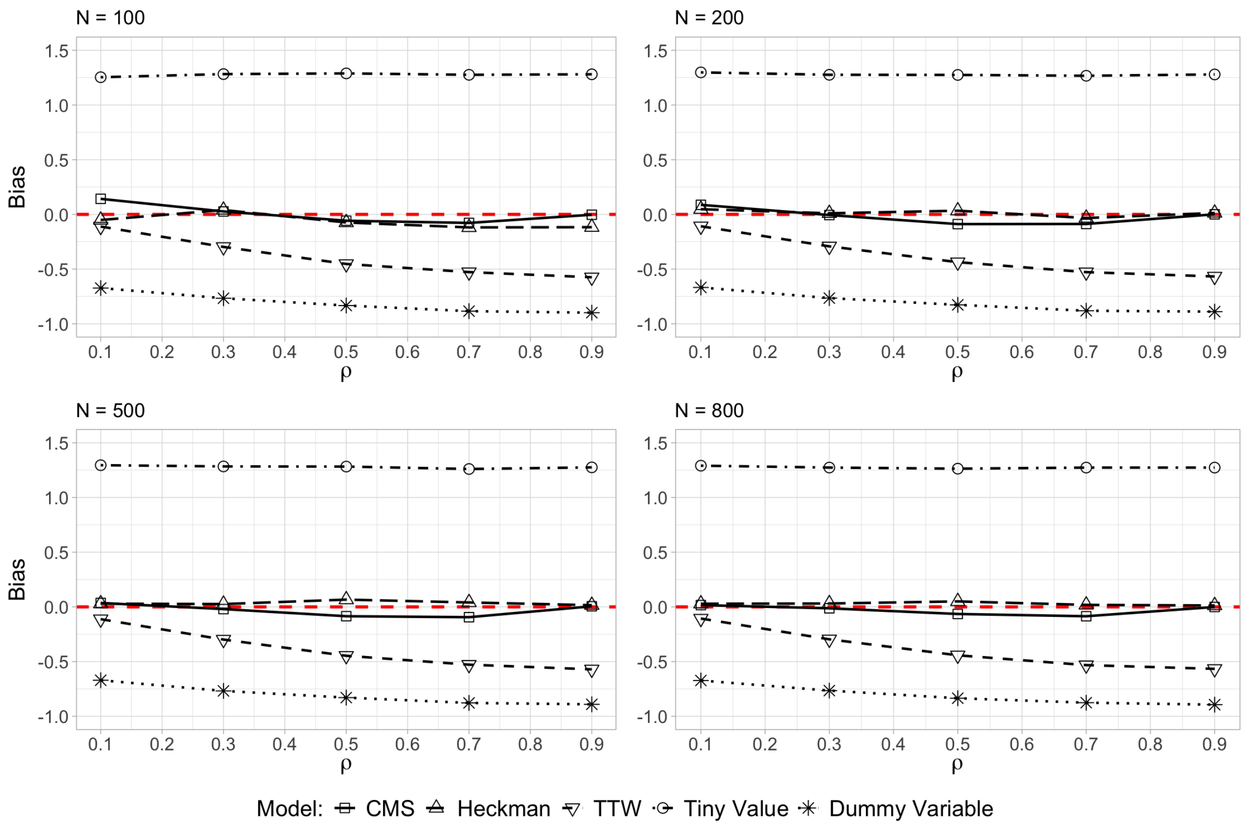


Figure 2: Bias in $\hat{\beta}_{1A}$—50% of Districts Partially Contested

ignore sample selection are severely biased: the TTW and dummy variable strategies

[17]We also ran simulations with random coefficients, in which the effect of z_{1i} on the trade-off between parties B and C (β_{1B}) depends on whether A has a candidate in i: $\beta_{1B} = -1$ if $d_{iA} = 1$, otherwise $\beta_{1B} = 0.5$ (Appendix D.3).

[18]We show results for other parameters (α_{1A}, α_{2A}, and β_{1B}) in Appendix D.4.

[19]Figures S20 and S30 show bias when A does not contest 33% and 66% of districts.

systematically underestimate the effect and the magnitude of this bias increases with ρ, while the tiny value strategy substantially overestimates the effect across all values of ρ. Both of the strategies that account for sample selection have minimal bias across the range of simulated circumstances.

Figure 3 presents the standard deviation of estimates from all five competing approaches. Across the board, the dummy variable and TTW strategies have the smallest standard deviation while the Heckman has the largest. The CMS and tiny value ap-

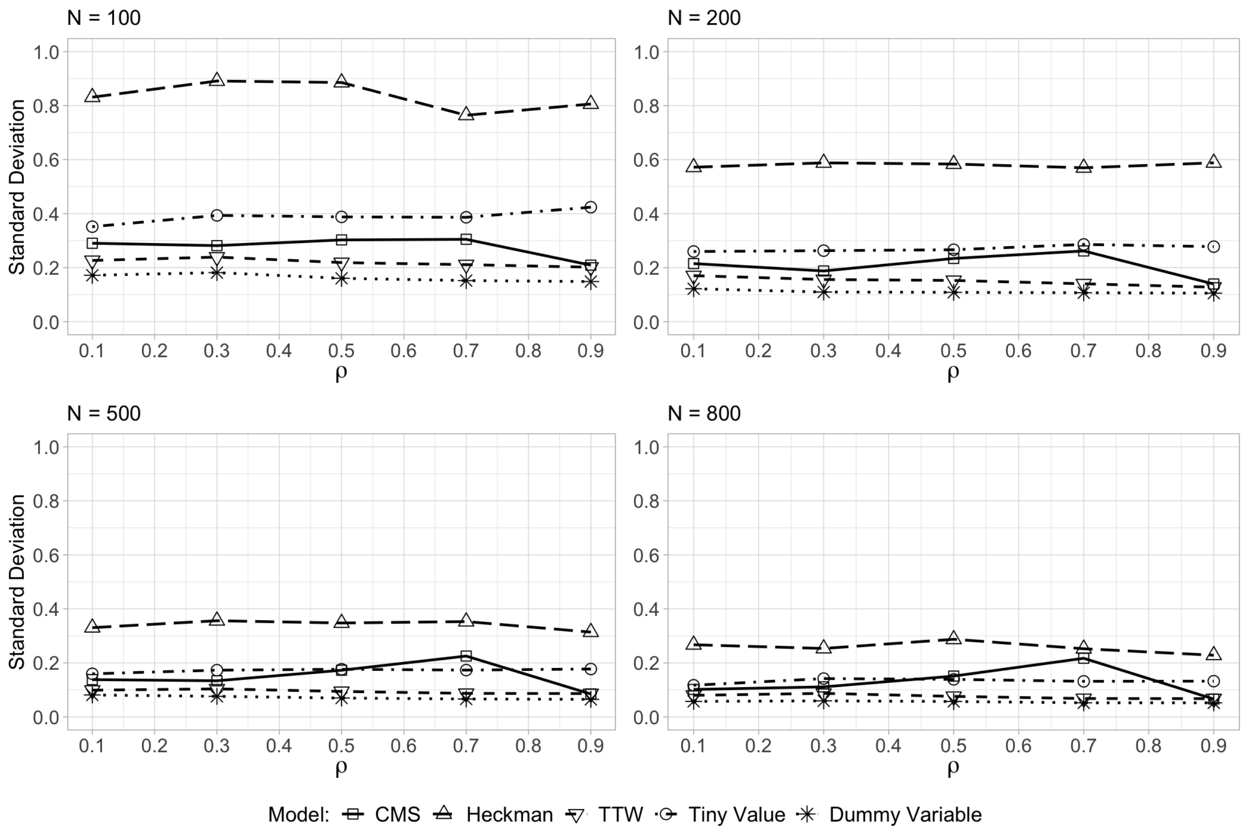


Figure 3: Standard Deviation of $\hat{\beta}_{1A}$s—50% of Districts Partially Contested

proaches perform much better than the Heckman and slightly worse than the other two. However, given the large amount of bias found for the TTW, dummy variable, and tiny value strategies, the main comparison of interest in this figure is between the Heckman and CMS strategies.[20] Because ε_{iA} and u_{iA} are not normally distributed, the standard de-

[20] Approaches that ignore selection perform much worse than the Heckman and CMS strategies on performance measures such as MSE that combine bias and efficiency. See results in Appendix D.

viation of Heckman estimates is larger, especially in relatively small samples. This shows that the CMS is more efficient than the Heckman. In Appendix D, we demonstrate that the Heckman strategy also results in overconfidence, higher MSE, and lower statistical power than the CMS.[21] Our approach consistently outperforms the Heckman model.

Results: Invalid Exclusion Restriction

To assess how the competing approaches perform when the exclusion restriction is violated, we repeat the experiment from the previous section, but set $\beta_{2A} = 1$ such that z_{2i} is no longer a valid instrument in Equation 14. Figure 4 presents bias in the estimates of β_{1A} under this condition. The Heckman estimator is severely biased across all values of ρ, per-

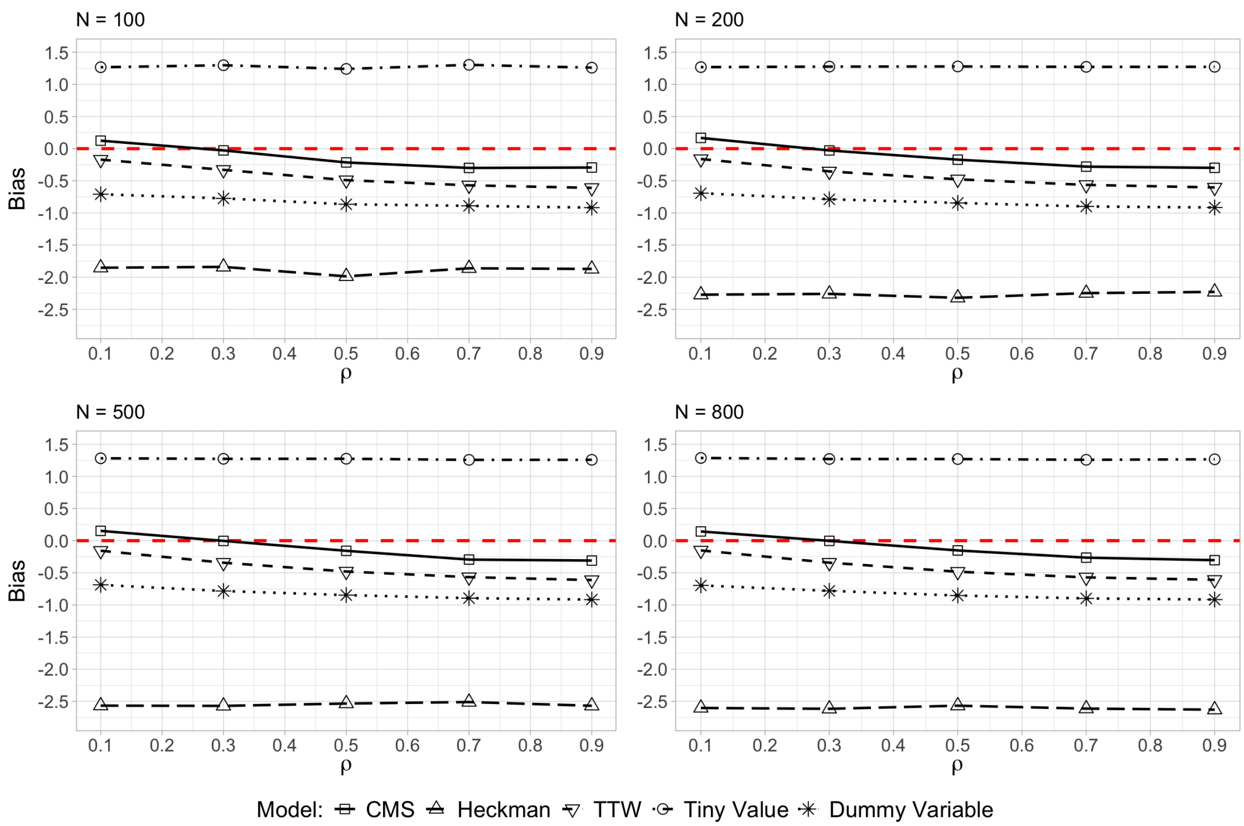


Figure 4: Bias in $\hat{\beta}_{1A}$—Invalid Instrument and 50% of Partial Contestation

forming much worse than models that ignore sample selection; the magnitude of this bias

[21] Following Hopkins et al. (2024), we compare the strategies across six performance statistics in Appendix D: bias, MSE, standard deviation, overconfidence, coverage probability, and statistical power.

worsens as the sample size increases. The Heckman estimator is biased because, without the exclusion restriction, the inverse Mills ratio is nearly collinear with other regressors, which makes the Heckman estimator potentially un-identifiable in practice (Honoré and Hu, 2024; Wolfolds and Siegel, 2019; Sartori, 2003; Newey, Powell and Walker, 1990).[22] In contrast, our CMS results in only slightly biased estimates when ρ is large, but it still outperforms alternatives that ignore sample selection.

In Appendix D, we show that without a valid instrument, the Heckman approach also performs poorly in terms of efficiency, overconfidence, coverage, and power, whereas the CMS performs well across all these performance statistics.[23] These results confirm that, under the distributional assumptions of error terms in compositional models derived from multinomial discrete choice processes, the Heckman correction is inappropriate if there is no valid instrument, a condition that is likely to happen in most empirical studies. Our CMS overcomes this limitation.

Application: Brexit & Vote Choice in England

We apply the CMS to analyze the English general elections in 2017 and 2019. Until the 2000s, elections in England yielded a suitable scenario to estimate compositional models as they revolved around two major parties (Labour and Conservatives) and a third force (Liberal Democrats, also known as Lib Dems), and all three parties fielded candidates in almost every district (Katz and King, 1999; Tomz, Tucker and Wittenberg, 2002). This situation has drastically changed as new parties, particularly the Greens and the UK Independence Party (UKIP), gained traction and became more competitive in some districts.[24]

[22] When the exclusion restriction is violated and the distributional error assumption is incorrect, our MLE approach is still identified, however, weakly, because it relies on the specification of error distribution. In Appendix D.9, we present results from a simulation where the errors are jointly normal, the exclusion restriction is violated, and the estimated CMS assumes jointly Gumbel errors. We find that our CMS outperforms other approaches under these circumstances.

[23] See Figures S45-S47 in Appendix D.

[24] UKIP and Brexit Party, the latter created in 2018, formed an electoral alliance in 2019, agreeing not to contest against one another in the same constituency. Hence, we treat them as one category in 2019 and refer to this alliance as UKIP.

The 2017 and 2019 elections followed the Brexit referendum in 2016. National politics was reshaped by the division between "leavers" (i.e., supporters of a quick solution for the UK's departure from the European Union) and "remainers" (i.e., those who opposed Brexit). In both elections, the Conservatives were the party of the prime minister, held the majority of seats, and were the main political force among the leavers.[25] During these elections, Brexit changed the competitive dynamics between parties (Fieldhouse et al., 2021). Rather than being viewed primarily as the incumbent party, the Conservatives were seen by most voters as the political force most able to implement Brexit. Hence, these two elections marked a transformation in long-term dynamics, such as economic voting and incumbency advantage, that had characterized English politics since the end of World War II (Kagalwala, Moreira and Whitten, 2024*b*). The leavers-remainers divide reduced the relevance of these previously stable voting patterns.

	2017		2019	
	Frequency	Percentage	Frequency	Percentage
Conservative	0	0%	0	0%
Labour	0	0%	0	0%
Liberal Democrats	2	0.37%	12	2.26%
Green	86	16.16%	78	14.68%
UKIP	197	37.03%	304	57.25%
Others	291	54.70%	239	45.01%
Constituencies	532		531	

Table 1: Partial Contestation by Party and Election

Partial contestation was high in 2017 and 2019 among small parties. Table 1 shows the frequency and the percentage of districts that parties did not contest.[26] The Greens did not have a candidate in about 15% of districts in both elections. UKIP contested 63% of districts in 2017, but only 48.75% in 2019. A residual category "Others"—comprised

[25] During the referendum, the Conservatives were neutral. After the "leave" victory, the party officially supported Brexit (Heath and Goodwin, 2017; Prosser, 2021).

[26] We exclude the constituency of the Speaker of the House of Commons, which is essentially uncontested. Because our models include a lagged dependent variable, we lose two districts in 2019 due to the Speaker changing shortly before the election. Hence, our data include 532 of 533 English districts in 2017 and 531 in 2019.

of parties that did not receive at least 1% of overall votes and did not win legislative seats—fielded candidates in 54.7% and 45% of districts in 2017 and 2019, respectively.[27]

Our CMS accounts for partial contestation as sample selection, allowing us to study how Brexit shaped trade-offs across all relevant parties. We use the Conservative Party as the baseline category of the log-ratios in Equation 2. Thus, the compositional outcome encompasses five log ratios, one for each $J-1$ category: the Labour Party, Liberal Democrats, Greens, UKIP, and Others. For each of these parties, we estimate the following model specification:

$$\begin{aligned} d_{ijt} &= \alpha_{0j} + \boldsymbol{z}_{ijt}\boldsymbol{\alpha}_j + \varepsilon_{ijt}, \\ s_{ijt} &= \beta_{0j} + \phi_j s_{ijt-1} + \boldsymbol{x}_{ijt}\boldsymbol{\beta}_j + u_{ijt}, \end{aligned} \tag{16}$$

where the first equation models j's decision to contest district i in election t, and the second equation is the compositional outcome with $J-1$ log ratios. We estimate these equations jointly using the log-likelihood function in Equation 11.[28] In the selection stage, d_{ijt} is an indicator for whether j had a candidate in i and $\boldsymbol{z}_{ijt}$ is a vector of the following district-level covariates:

- j's distance in vote share from the second place in the previous election $(t-1)$,
- the difference in vote shares between the top two parties at $t-1$,
- the effective number of parties (ENP) at $t-1$,
- the percentage of votes for leaving the European Union from the 2016 referendum,[29]
- the percentage of residents claiming unemployment benefits in the district,
- the log of the electorate size,

[27] As explained in Footnote 4, almost all elections have parties that are not viable competitors and thus cannot be included on their own in studies of results across districts.

[28] Since the Labour and Conservative Parties contested all districts (Table 1), we do not estimate a selection stage in the trade-off between them.

[29] Data from the House of Commons Library. In total, there are observed results for 169 constituencies. For the remaining 481 districts, Chris Hanretty estimated the results. As a robustness check, we follow Adler and Ansell (2020) and Ansell et al. (2022), using the average percentage change in house prices as a proxy for Brexit support. See Appendix C.6.

- the share of the rural population,
- dummy variables that indicate Conservative and Labour Party incumbency,
- regional fixed effects,[30] and
- election fixed effects.

The first three variables in $\boldsymbol{z}_{ijt}$ are tactical factors that a party j may consider when deciding whether to contest a district. We expect that smaller parties are less likely to employ scarce resources in localities where a single party dominates the contest. This is captured by the distance between the top two parties at $t-1$. Similarly, j may avoid areas where its past electoral performances were poor and districts crowded with many contenders. We measure these factors with j's distance from the runner-up and the ENP, both measured at $t-1$. Because these first three variables are expected to affect parties' contestation decisions but not the vote share of parties in districts that they contest, these variables are present in $\boldsymbol{z}_{ijt}$ but not $\boldsymbol{x}_{ijt}$. Hence, they also play the role of meeting the exclusion restriction. The remaining variables in the list above are all expected to affect both the decision of parties about whether or not to contest a district and the percentages of support for parties that do contest them. Our equations for s_{ijt} also include the temporally lagged dependent variable, s_{ijt-1}, which both reflects our expectation of persistence over time and controls for unobservable common factors of a district that shape the trade-off between j and the baseline category (Keele and Kelly, 2006).

In Figure 5, we display estimated coefficients for four predictors in the selection equations. Each panel shows the coefficient of each predictor for the partially contesting parties.[31] As expected, the tactical variables affect the likelihood that small parties will contest a district. As the difference between the top two parties in the previous elections increases, the Greens and UKIP are less likely to field candidates (Panel A). Similarly, as the distance between their previous vote and that of the runner-up (Panel B) increases,

[30] England's nine Government Office Regions.

[31] Table S2 in Appendix C shows full results.

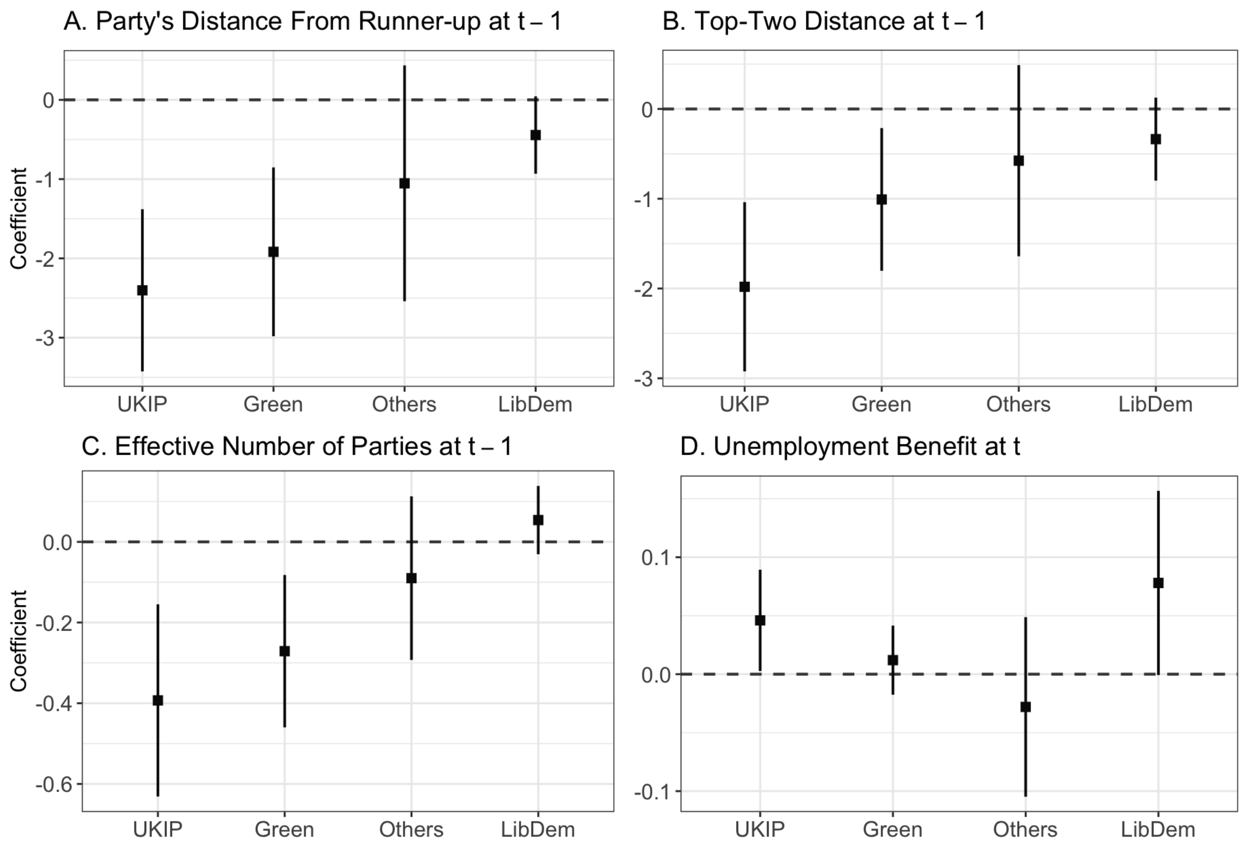


Figure 5: Coefficients in the Selection Stage

they are less likely to contest districts, and as the ENP (Panel C) at $t-1$ increases, they are less likely to field candidates. Finally, Panel D shows that covariates in the outcome stage can also affect a party's probability of fielding candidates. The unemployment variable is positively associated with the Liberals' and UKIP's decision to contest a district.

Figure S1 in Appendix C shows the estimated coefficients for Brexit votes and unemployment benefits across the $J-1$ outcome equations.[32] Consistent with Kagalwala, Moreira and Whitten (2024*b*), we find that support for Brexit increases the vote share of Conservatives to the detriment of parties that opposed leaving the European Union. Unemployment benefits, in turn, did not have the clear effect expected by economic voting theory.[33] These coefficients, though, are in log-ratio form. We can use them for hypoth-

[32]Figure S2 shows that competing compositional approaches estimate different coefficients in equations for parties that did not contest all districts in our empirical example.

[33]As an indicator of deteriorating conditions, unemployment is expected to negatively affect the prime

esis tests about the trade-offs between parties, assessing the direction of the relationship and its statistical significance.[34] To interpret the magnitude of these results, we transform them back into vote shares and calculate their associated confidence intervals using a non-parametric bootstrap in which we randomly sample the data 1,000 times with replacement. We calculate the difference in predicted vote share for each party after a one standard deviation increase (+1SD) holding all other variables at their observed values.[35]

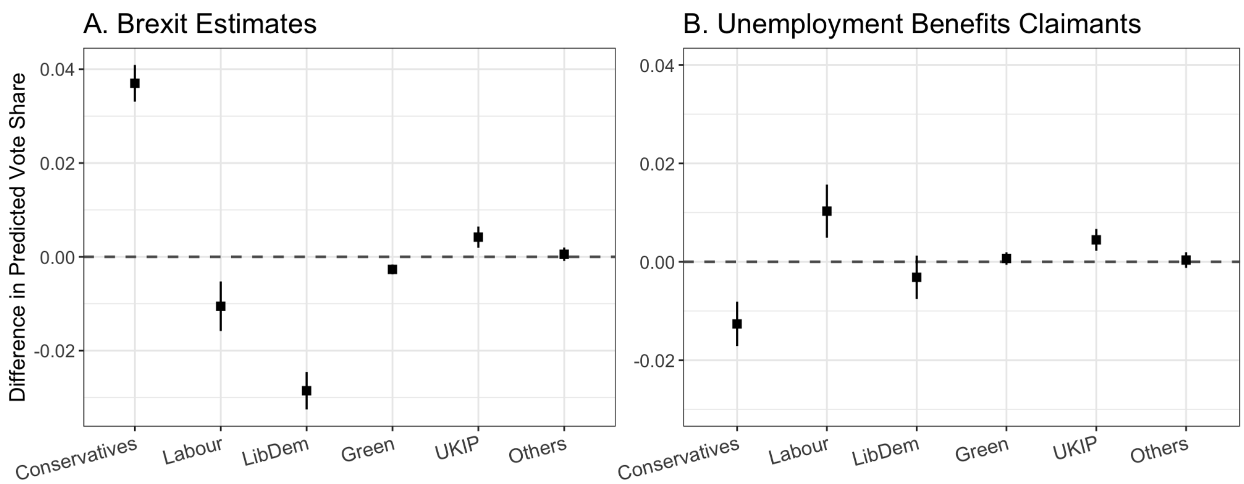


Figure 6: Change in Vote Share in Response to +1SD Shock

Figure 6 presents the effect of a +1SD increase in Brexit votes (Panel A) and unemployment benefits (Panel B) on the predicted vote share for each party. This indicates that the Conservatives, as the major force supporting Brexit in 2017 and 2019, would win 3.7% of votes due to a +1SD increase in estimated votes for Brexit in 2016. UKIP's votes would also increase, but by a small degree, 0.4%. The Labour Party and Liberal Democrats would lose 1.1% and 2.8% of votes respectively from this increase in Brexit support. Panel B shows that the estimated changes in vote shares after a +1SD increase in unemployment benefits are mostly smaller—1% and -1.2% for the Labour and Conservatives respectively. UKIP also electorally benefited from a shock in unemployment

minister's party across most log ratios.

[34] Negative values imply that a predictor benefits the baseline category at the expense of j in the numerator. Positive coefficients have the opposite interpretation.

[35] Table S1 in Appendix C shows descriptive statistics.

(0.4%). These results support the expectation that the Brexit debate dominated U.K. politics in the 2017 and 2019 elections.[36]

Table S2 in Appendix C shows that the correlation between errors across stages is high (> 0.8) and statistically significant for all parties, demonstrating the necessity to account for sample selection: A party's decision to contest a district and its vote share when it does contest that district are functions of correlated unobservables.

Conclusion

In this paper, we demonstrate that partial contestation constitutes a form of sample selection, which existing approaches to modeling election results fail to address. As a result, standard methods yield biased estimates. To address this, we introduce a novel MLE approach that jointly accounts for sample selection and the compositional nature of election results. Our approach yields mostly unbiased and efficient estimates, even if the exclusion restriction required by the traditional Heckman correction is violated.

In multiparty elections, outcomes are inherently interdependent—gains for one party necessarily entail losses for others. Compositional models are uniquely suited to capture these interrelationships and, most importantly, allow researchers to examine how shifts in support for one party influence the performances of others. This is particularly valuable in the study of vote switching and strategic voting. For example, rising support for radical parties does not uniformly affect all competitors (Colantone and Stanig, 2018; Cremaschi et al., 2024); identifying the factors behind voters' departure from specific parties and their movement toward others lies at the core of electoral studies. Yet, partial contestation complicates the use of the standard compositional framework. Our proposed approach addresses this challenge.

Although elections provide a clear application, the utility of our approach extends to

[36] In Appendices C.2-C.5, we present a variety of model fit diagnostics, including in-sample and out-of-sample predictions.

other substantive domains in which sample selection processes generate problems for models with compositional outcomes. These include analyses of government budget trade-offs, the allocation of portfolios among coalition partners, and the share of attention that parties devote to particular issues in their manifestos (Cox, 2021; Lipsmeyer, Philips and Whitten, 2023; Abou-Chadi and Krause, 2020). In all such examples, correlated unobservables may simultaneously influence both the strategic decision for a category to be included and its observed share. Addressing sample selection in a compositional framework is thus critical for accurate and consistent inferences.

Our MLE approach is flexible, accommodating different error distributions depending on the application. We encourage practitioners to carefully consider the DGP and exclusion restriction and to interpret their results as contingent on these assumptions when estimating compositional models with selection. Without a valid instrument, the CMS relies solely on functional-form identification. Future work may develop a CMS with other methods for constructing joint distributions such as a copula approach.

Supplementary Material

For supplementary material accompanying this paper visit

https://doi.org/

Data Availability

Replication code for this article has been published in the Political Analysis Harvard Dataverse at https://doi.org/10.7910/DVN/E8EOOD (Kagalwala et al., 2026).

Author ORCiDs

- Ali Kagalwala: https://orcid.org/0000-0002-0136-4581
- Thiago M.Q. Moreira: https://orcid.org/0000-0002-3891-572X

- Guy D. Whitten: https://orcid.org/0000-0003-1595-5507

- Yongzhi Xu: https://orcid.org/0009-0009-7780-7754

Authors Contributions

All authors made equal contributions to this manuscript, and the authors are listed in alphabetical order.

Financial Support

None.

Conflict of Interest

None.

References

Abou-Chadi, Tarik and Werner Krause. 2020. "The causal effect of radical right success on mainstream parties' policy positions: A regression discontinuity approach." *British Journal of Political Science* 50(3):829–847.

Achen, Christopher H. 1986. *The statistical analysis of quasi-experiments*. University of California Press.

Adler, David and Ben Ansell. 2020. "Housing and populism." *West European Politics* 43(2):344–365.

Aitchison, John. 1982. "The statistical analysis of compositional data." *Journal of the Royal Statistical Society: Series B (Methodological)* 44(2):139–160.

Ansell, Ben, Frederik Hjorth, Jacob Nyrup and Martin Vinæs Larsen. 2022. "Sheltering populists? House prices and the support for populist parties." *The Journal of Politics* 84(3):1420–1436.

Boix, Carles. 2009. The Emergence of Parties and Party Systems. In *The Oxford Handbook of Comparative Politics*, ed. Carles Boix and Susan C. Stokes. Oxford University Press pp. 499–521.

Butler, Daniel M and Eleanor Neff Powell. 2014. "Understanding the party brand: experimental evidence on the role of valence." *The Journal of Politics* 76(2):492–505.

Cavallaro, Matteo, Lorenzo Pregliasco and Salvatore Vassallo. 2018. "The electoral system and electoral geography-why the Rosato law appeared to be proportional while it is not." *Contemporary Italian Politics* 10(3):224–242.

Colantone, Italo and Piero Stanig. 2018. "The trade origins of economic nationalism: Import competition and voting behavior in Western Europe." *American Journal of Political Science* 62(4):936–953.

Cook, Jonathan, Joon-Suk Lee and Noah Newberger. 2021. "On identification and estimation of Heckman models." *The Stata Journal* 21(4):972–998.

Cox, Gary W. 2021. "Nonunitary parties, government formation, and gamson's law." *American Political Science Review* 115(3):917–930.

Cox, Gary W and Jonathan N Katz. 1996. "Why did the incumbency advantage in US House elections grow?" *American Journal of Political Science* pp. 478–497.

Cremaschi, Simone, Paula Rettl, Marco Cappelluti and Catherine E De Vries. 2024. "Geographies of discontent: Public service deprivation and the rise of the far right in Italy." *American Journal of Political Science* .

Döring, Holger and Philip Manow. 2012. "Parliament and government composition database (ParlGov)." *An infrastructure for empirical information on parties, elections and governments in modern democracies. Version* 12(10).

Duch, Raymond M and Randy Stevenson. 2005. "Context and the economic vote: a multilevel analysis." *Political Analysis* 13(4):387–409.

Fieldhouse, Edward, Jane Green, Geoffrey Evans, Jonathan Mellon, Christopher Prosser, Hermann Schmitt and Cees Van der Eijk. 2021. *Electoral shocks: The volatile voter in a turbulent world*. Oxford University Press.

Greene, William H. 2017. *Econometric Analysis*. Eight ed. New York, NY: Pearson.

Heath, Oliver and Matthew Goodwin. 2017. "The 2017 general election, Brexit and the return to two-party politics: An aggregate-level analysis of the result." *The Political Quarterly* 88(3):345–358.

Heckman, James J. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of economic and social measurement, volume 5, number 4*. NBER pp. 475–492.

Heckman, James J. 1979. "Sample selection bias as a specification error." *Econometrica: Journal of the Econometric Society* pp. 153–161.

Honoré, Bo E and Luojia Hu. 2024. "Sample selection models without exclusion restrictions: Parameter heterogeneity and partial identification." *Journal of Econometrics* 243(1-2):105360.

Hopkins, Vincent, Ali Kagalwala, Andrew Q Philips, Mark Pickup and Guy D Whitten. 2024. "How do we know what we know? Learning from Monte Carlo simulations." *The Journal of Politics* 86(1):36–53.

Horiuchi, Yusaku and Woo Chang Kang. 2018. "Why should the republicans pray for rain? electoral consequences of rainfall revisited." *American Politics Research* 46(5):869–889.

Kagalwala, Ali, Thiago M. Q. Moreira, Guy D. Whitten and Yongzhi Xu. 2026. "Replication Data for: Let Them Eat Pie: Addressing Sample Selection in Multiparty Elections." Harvard Dataverse, V1. https://doi.org/10.7910/DVN/E8EOOD.

Kagalwala, Ali, Thiago MQ Moreira and Guy D Whitten. 2024*a*. "A simple approach to dealing with partial contestation." *Social Science Quarterly* 105(4):948–964.

Kagalwala, Ali, Thiago MQ Moreira and Guy D Whitten. 2024*b*. "Tumultuous pies: Electoral dynamics in the Brexit era." *Electoral Studies* 90:102821.

Katz, Jonathan N and Gary King. 1999. "A statistical model for multiparty electoral data." *American Political Science Review* pp. 15–32.

Kayser, Mark Andreas and Michael Peress. 2012. "Benchmarking across borders: electoral accountability and the necessity of comparison." *American Political Science Review* 106(3):661–684.

Keele, Luke and Nathan J. Kelly. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis* 14:186–205.

Lipsmeyer, Christine S, Andrew Q Philips and Guy D Whitten. 2023. *The politics of budgets: Getting a piece of the pie*. Cambridge University Press.

McFadden, Daniel. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, ed. Paul Zarembka. New York, NY: Academic Press pp. 105–142.

Moreira, Thiago MQ. 2025. "Is It Still the Economy? Economic Voting in Polarized Politics." *British Journal of Political Science* 55:e45.

Moreira, Thiago MQ, Spencer Hamilton Goidel and Brenna Armstrong. 2025. "The Drug Crisis and Voting Behavior." *Perspectives on Politics* pp. 1–20.

Newey, Whitney K, James L Powell and James R Walker. 1990. "Semiparametric estimation of selection models: some empirical results." *The American Economic Review* 80(2):324–328.

Pattie, Charles and Ron Johnston. 2009. "Still Talking, But Is Anyone Listening? The Changing Face of Constituency Campaigning in Britain, 1997—2005." *Party Politics* 15(4):411–434.

Philips, Andrew Q, Amanda Rutherford and Guy D Whitten. 2016. "Dynamic pie: A strategy for modeling trade-offs in compositional variables over time." *American Journal of Political Science* 60(1):268–283.

Powell Jr, G Bingham and Guy D Whitten. 1993. "A cross-national analysis of economic voting: Taking account of the political context." *American Journal of Political Science* pp. 391–414.

Prosser, Christopher. 2021. "The end of the EU affair: the UK general election of 2019." *West European Politics* 44(2):450–461.

Sartori, Anne E. 2003. "An estimator for some binary-outcome selection models without exclusion restrictions." *Political Analysis* 11(2):111–138.

Tomz, Michael, Joshua A Tucker and Jason Wittenberg. 2002. "An easy and accurate regression model for multiparty electoral data." *Political Analysis* pp. 66–83.

Train, Kenneth E. 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press.

Ward, George. 2020. "Happiness and voting: Evidence from four decades of elections in Europe." *American Journal of Political Science* 64(3):504–518.

Wolfolds, Sarah E and Jordan Siegel. 2019. "Misaccounting for endogeneity: The peril of relying on the Heckman two-step method without a valid instrument." *Strategic Management Journal* 40(3):432–462.

Zellner, Arnold. 1962. "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias." *Journal of the American Statistical Association* 57(298):348–368.

Supplementary Materials Document for "Let Them Eat Pie"

A Appendix A: Detailed Explanation of the MLE Approach 2

B Appendix B: Derivation of the Log-Likelihood Function 5

C Appendix C: Empirical Example Results 6

C.1 Descriptive Information 6

C.2 Table of Results & Coefficient Plots 7

C.3 Predicted Vote Shares 10

C.4 Predictions in Fully & Partially Contested Districts 17

C.5 Out-of-Sample Prediction 19

C.6 Robustness Check—House Prices Change 26

D Appendix D: Monte Carlo Results 29

D.1 Valid Exclusion Restriction 30

D.2 Invalid Exclusion Restriction 47

D.3 Random Coefficients 65

D.4 Other Coefficients 77

D.5 Normal DGP & Extreme Value Type-I Assumption 83

D.6 Normal DGP & Normal Distribution Assumption 87

D.7 Monte Carlo Simulations with 6 Parties 91

D.8 Correlation Across Outcome Equations 97

D.9 Normal DGP, Invalid Exclusion Restriction, & Type-I Assumption 110

E Appendix E: R Code With Instructions 114

A Appendix A: Detailed Explanation of the MLE Approach

Suppose the latent utility for a voter who votes for party j in district i is

$$s_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta}_j + u_{ij},$$

where u_{ij} is distributed *iid* mean-zero extreme value type-I.[1]

Consider the probability that a voter in district i would vote for party j over all other parties in that district ($s_{ij'}$)

$$Pr(s_{ij} > s_{ij'}),$$

then the average probability that voters in district i vote for party j is some function of the average characteristics in the district (e.g., average house price). This average probability will equal the vote share of party j in district i, such that:

$$\begin{aligned} v_{ij} &= \frac{\exp(s_{ij})}{1 + \sum_{j=2}^{J}(\exp(s_{ij}))} \ \forall j \neq 1, \\ v_{i1} &= \frac{1}{1 + \sum_{j=2}^{J}\exp(s_{ij})} \end{aligned}$$

where the vote share of party j in district i is a multinomial transformation of the latent utilities of all voters in that district. This function is a multinomial transformation because we assumed the errors are distributed extreme value type-I. For identification purposes, we set the latent utility of voters who vote for party $j = 1$ to 0, such that $s_{i1} = 0$. Then,

$$\log\left(\frac{v_{ij}}{v_{i1}}\right) = \log\left(\frac{\frac{\exp(s_{ij})}{1+\sum_{j=2}^{J}\exp(s_{ij})}}{\frac{\exp(s_{i1})}{1+\sum_{j=2}^{J}\exp(s_{ij})}}\right) = \log\left(\frac{\exp(s_{ij})}{\exp(s_{i1})}\right) = \log(\exp(s_{ij})) = s_{ij}$$

This demonstrates that the latent utility for voting for party j can be expressed as a log ratio of the observed vote shares of some party, j, to that of an arbitrarily chosen baseline

[1] This is a standard assumption for utility functions of this type (e.g., McFadden, 1974; Train, 2009).

party, $j = 1$.[2]

Given the selection process in which parties choose which electoral districts to compete, we only observe the vote shares of parties when they contest elections. Let the selection procedure be given as

$$d_{ij}^* = \mathbf{z}_{ij}'\boldsymbol{\alpha}_j + \varepsilon_{ij},$$

where d_{ij}^* is latent utility of party j contesting in district i. The selection stage is then

$$d_{ij} = \begin{cases} 1 & \text{if } d_{ij}^* \geq 0 \\ 0 & \text{if } d_{ij}^* < 0 \end{cases}$$

and the outcome stage—the vote shares of parties that did contest—is given by

$$v_{ij} = \begin{cases} \frac{\exp(s_{ij})}{1+\sum_{j=2}^{J}(\exp(s_{ij}))} & \text{if } d_{ij}^* \geq 0 \\ \text{missing} & \text{if } d_{ij}^* < 0 \end{cases}$$

Using a single-outcome equation as a general case, Heckman demonstrated that failing to account for sample selection leads to biased parameter estimates. In the compositional-outcome stage, the conditional expectation of the observed vote share for party j in electoral district i is

$$\begin{aligned} \mathbb{E}[s_{ij}|\mathbf{x}_{ij}', d_{ij}^* \geq 0] &= \mathbb{E}[\mathbf{x}_{ij}'\boldsymbol{\beta}_j + u_{ij}|\mathbf{x}_{ij}', \varepsilon_{ij} \geq -\mathbf{z}_{ij}'\boldsymbol{\alpha}_j] \\ &= \mathbf{x}_{ij}'\boldsymbol{\beta}_j + \mathbb{E}[u_{ij}|\varepsilon_{ij} \geq -\mathbf{z}_{ij}'\boldsymbol{\alpha}_j] \end{aligned}$$

where $\mathbb{E}[u_{ij}|\varepsilon_{ij} \geq -\mathbf{z}_{ij}'\boldsymbol{\alpha}_j] \neq 0$. This quantity is biased as long as u_{ij} and ε_{ij} are correlated, which is a reasonable assumption here—e.g., a party's reputation in a district enters both the selection and outcome stages.

[2] In cases when there is abstention, the baseline in the log-ratio could also be the proportion of people who abstain from voting in a district i.

If u_{ij} and ε_{ij} are jointly normal, then adding an inverse Mills ratio, $\frac{\phi(\mathbf{z}_{ij}'\hat{\boldsymbol{\alpha}}_j)}{\Phi(\mathbf{z}_{ij}'\hat{\boldsymbol{\alpha}}_j)}$, to the right-hand-side of the outcome stage model would lead to unbiased parameter estimates. However, since our outcomes are in log-ratio form, we treat u_{ij} and ε_{ij} as distributed bivariate extreme value type-I and their marginal distributions are both extreme value type-I distributions with density

$$f(x) = e^{-\left(x+\gamma+e^{-(x+\gamma)}\right)},$$

where γ is the Euler–Mascheroni constant, then the joint CDF and PDF are given by

$$F(u,\varepsilon;m) = \exp\left[-\left(e^{-mu}+e^{-m\varepsilon}\right)^{\frac{1}{m}}\right] \tag{1}$$

$$f(u,\varepsilon;m) = F(u,\varepsilon;m)e^{-m(u+\varepsilon)}\left(e^{-mu}+e^{-m\varepsilon}\right)^{\frac{1}{m}-2}.\left[m-1+\left(e^{-mu}+e^{-m\varepsilon}\right)^{\frac{1}{m}}\right], \tag{2}$$

where $m = \frac{1}{1-\rho} \in [1,\infty]$, ρ is the correlation between u_{ij} and ε_{ij}, and $m = 1$ when u_{ij} and ε_{ij} are independently distributed.

The likelihood function is

$$\mathcal{L}_j = \prod_{i=1}^{N} \mathbb{P}(\varepsilon_{ij} > -\mathbf{z}_{ij}'\boldsymbol{\alpha}_j)^{d_{ij}} f_{u|\varepsilon}(s_{ij} - \mathbf{x}_{ij}'\boldsymbol{\beta}_j | \varepsilon_{ij} > -\mathbf{z}_{ij}'\boldsymbol{\alpha}_j)^{d_{ij}} \mathbb{P}(\varepsilon_{ij} < -\mathbf{z}_{ij}'\boldsymbol{\alpha}_j)^{1-d_{ij}},$$

where $\mathbb{P}(\varepsilon_{ij} > -\mathbf{z}_{ij}'\boldsymbol{\alpha}_j)^{d_{ij}}$ is the likelihood that party j competes in district i, $f_{u|\varepsilon}(s_{ij} - \mathbf{x}_{ij}'\boldsymbol{\beta}_j | \varepsilon_{ij} > -\mathbf{z}_{ij}'\boldsymbol{\alpha}_j)^{d_{ij}}$ is the likelihood of observing the vote share of party j in district i given that it participates in district i, and $\mathbb{P}(\varepsilon_{ij} < -\mathbf{z}_{ij}'\boldsymbol{\alpha}_j)^{1-d_{ij}}$ is the likelihood that party j does not participate in district i.

The log-likelihood is then given by

$$\log(\mathcal{L}_j) = \sum_{i=1}^{N} d_{ij}\log\left(\int_{-\mathbf{z}_{ij}'\boldsymbol{\alpha}_j}^{\infty} f_{u\varepsilon}(s_{ij} - \mathbf{x}_{ij}'\boldsymbol{\beta}_j, \varepsilon_{ij})\mathrm{d}\varepsilon_{ij}\right) + \sum_{i=1}^{N}(1-d_{ij})\log(F_{\varepsilon}(-\mathbf{z}_{ij}'\boldsymbol{\alpha}_j)),$$

where F_{ε} is the marginal CDF of ε_{ij} and $f_{u\varepsilon}$ is the joint density of u_{ij} and ε_{ij} in Equation 2.

B Appendix B: Derivation of the Log-Likelihood Function

Let event A be $d_{ij}^* = \mathbf{z}_{ij}'\boldsymbol{\alpha}_j + \varepsilon_{ij} \geq 0$ and event B be $s_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta}_j + u_{ij}$. According to Bayes' Theorem, the likelihood function for party j in district i is

$$\mathcal{L}_{ij} = \begin{cases} l_{ij}(AB) = l_{ij}(B|A)l_{ij}(A), & \text{if } d_{ij} = 1 \\ 1 - l_{ij}(A), & \text{if } d_{ij} = 0 \end{cases} \tag{3}$$

where $l(x)$ is the likelihood of event x, and

$$\begin{aligned} l_{ij}(A) &= \mathbb{P}(\varepsilon_{ij} > -\mathbf{z}_{ij}'\boldsymbol{\alpha}_j) \\ l_{ij}(B|A) &= f_{u|\varepsilon}(s_{ij} - \mathbf{x}_{ij}'\boldsymbol{\beta}_j | \varepsilon_{ij} > -\mathbf{z}_{ij}'\boldsymbol{\alpha}_j) \end{aligned} \tag{4}$$

So the likelihood function of party j in all districts is

$$\begin{aligned} \mathcal{L}_j = \sum_{i=1}^{n} \mathcal{L}_{ij} = &\prod_{i=1}^{n} \mathbb{P}(\varepsilon_{ij} > -\mathbf{z}_{ij}'\boldsymbol{\alpha}_j)^{d_{ij}} \mathbb{P}(\varepsilon_{ij} < -\mathbf{z}_{ij}'\boldsymbol{\alpha}_j)^{1-d_{ij}} \\ &\cdot f_{u|\varepsilon}(s_{ij} - \mathbf{x}_{ij}'\boldsymbol{\beta}_j | \varepsilon_{ij} > -\mathbf{z}_{ij}'\boldsymbol{\alpha}_j)^{d_{ij}} \end{aligned} \tag{5}$$

The log-likelihood is derived as follows

$$\begin{aligned} \log(\mathcal{L}_j) = &\sum_{i=1}^{n} d_{ij} \log(1 - F_\varepsilon(-\mathbf{z}_{ij}'\boldsymbol{\alpha}_j)) + \sum_{i=1}^{n} (1 - d_{ij}) \log(F_\varepsilon(-\mathbf{z}_{ij}'\boldsymbol{\alpha}_j)) \\ &+ \sum_{i=1}^{n} d_{ij} \log\left(\int_{-\mathbf{z}_{ij}'\boldsymbol{\alpha}_j}^{\infty} f_{u\varepsilon}(y_{ij} - \mathbf{x}_{ij}'\boldsymbol{\beta}_j, \varepsilon_{ij}) \mathrm{d}\varepsilon_{ij}\right) \\ &- \sum_{i=1}^{n} d_{ij} \log\left(1 - F_\varepsilon(-\mathbf{z}_{ij}'\boldsymbol{\alpha}_j)\right) \\ = &\sum_{i=1}^{n} (1 - d_{ij}) \log(F_\varepsilon(-\mathbf{z}_{ij}'\boldsymbol{\alpha}_j)) \\ &+ \sum_{i=1}^{n} d_{ij} \log\left(\int_{-\mathbf{z}_{ij}'\boldsymbol{\alpha}_j}^{\infty} f_{u\varepsilon}(s_{ij} - \mathbf{x}_{ij}'\boldsymbol{\beta}_j, \varepsilon_{ij}) \mathrm{d}\varepsilon_{ij}\right) \end{aligned}$$

C Appendix C: Empirical Example Results

C.1 Descriptive Information

Variable	Mean	St. Dev.	Min	Median	Max
Conservatives Vote Share	0.46	0.15	0.07	0.49	0.77
Labour Vote Share	0.39	0.18	0.04	0.38	0.86
LibDem Vote Share	0.10	0.10	0.00	0.06	0.56
Greens Vote Share	0.02	0.03	0.00	0.02	0.57
Ukip Vote Share	0.02	0.03	0.00	0.01	0.30
Others Vote Share	0.01	0.03	0.00	0.001	0.41
Conservatives Vote Share (lag)	0.42	0.15	0.05	0.46	0.70
Labour Vote Share (lag)	0.38	0.18	0.05	0.37	0.86
LibDem Vote Share (lag)	0.08	0.08	0.00	0.05	0.53
Greens Vote Share (lag)	0.03	0.03	0.00	0.02	0.52
UKIP Vote Share (lag)	0.08	0.07	0.00	0.05	0.44
Others Vote Share (lag)	0.01	0.02	0.00	0.002	0.36
$\log(\frac{\text{Labour}}{\text{Conservatives}})$	-0.22	0.91	-2.43	-0.30	2.35
$\log(\frac{\text{LibDem}}{\text{Conservatives}})$	$-Inf.$	0.95	$-Inf.$	-1.97	0.88
$\log(\frac{\text{UKIP}}{\text{Conservatives}})$	$-Inf.$	0.75	$-Inf.$	-3.39	0.35
$\log(\frac{\text{Greens}}{\text{Conservatives}})$	$-Inf.$	0.82	$-Inf.$	-3.10	1.18
$\log(\frac{\text{Others}}{\text{Conservatives}})$	$-Inf.$	1.18	$-Inf.$	-5.95	0.27
$\log(\frac{\text{Labour}}{\text{Conservatives}})$ (lag)	0.71	0.47	0.11	0.57	2.91
$\log(\frac{\text{LibDem}}{\text{Conservatives}})$ (lag)	0.17	0.19	0.00	0.11	1.49
$\log(\frac{\text{UKIP}}{\text{Conservatives}})$ (lag)	0.20	0.20	0.00	0.15	1.24
$\log(\frac{\text{Greens}}{\text{Conservatives}})$ (lag)	0.08	0.11	0.00	0.06	1.31
$\log(\frac{\text{Others}}{\text{Conservatives}})$ (lag)	0.02	0.06	0.00	0.01	0.88
% of Votes for Brexit in 2016	53.57	11.03	18.20	55.20	75.60
Avg. % Δ in House Prices	3.48	1.20	1.37	3.27	9.04
% Unemp. Benefits Claimants	2.33	1.31	0.44	2.04	8.13
Distance 1^{st} & 2^{nd} (lag)	0.26	0.15	0.001	0.26	0.77
Effective Number of Parties (lag)	2.54	0.44	1.34	2.48	4.29
Electorate (log)	11.21	0.10	10.92	11.21	11.64
Share Rural Population	0.18	0.24	0.00	0.05	0.99
Conservative Incumbent	0.58	0.49	0	1	1
Labour Incumbent	0.41	0.49	0	0	1
2019 Election	0.50	0.50	0	0	1
Distance LibDem & 2^{nd} (lag)	0.08	0.27	0	0	1
Distance Greens & 2^{nd} (lag)	0.004	0.06	0	0	1
Distance UKIP & 2^{nd} (lag)	0.11	0.31	0	0	1
Distance Others & 2^{nd} (lag)	0.004	0.06	0	0	1
LibDem Contest	0.99	0.11	0	1	1
Greens Contest	0.85	0.36	0	1	1
UKIP Contest	0.53	0.50	0	1	1
Others Contest	0.50	0.50	0	1	1
London	0.14	0.34	0	0	1
East	0.11	0.31	0	0	1
East Midlands	0.09	0.28	0	0	1
West Midlands	0.11	0.31	0	0	1
Yorkshire and the Humber	0.10	0.30	0	0	1
North East	0.05	0.23	0	0	1
North West	0.14	0.35	0	0	1
South East	0.16	0.36	0	0	1
South West	0.10	0.30	0	0	1

Table S1: Descriptive Information

C.2 Table of Results & Coefficient Plots

	Selection Stage			
	Party			
	Lib Dem	Green	UKIP	Others
Constant	1.901***	3.844***	0.807	3.455
	(0.731)	(1.465)	(2.036)	(2.245)
Brexit Vote in 2016	0.005**	0.001	-0.005*	0.007*
	(0.002)	(0.002)	(0.002)	(0.003)
Unemp. Benefits	0.078*	0.012	0.046**	0.028
	(0.040)	(0.015)	(0.022)	(0.039)
Log Electorate	-0.066	-0.162	0.174	0.429**
	(0.064)	(0.121)	(0.170)	(0.177)
Share Rural	0.104	-0.069	-0.013	0.086
	(0.128)	(0.092)	(0.119)	(0.122)
Top-Two Distance ($t-1$)	-0.336	-1.008**	-1.981***	-0.576
	(0.235)	(0.404)	(0.479)	(0.541)
Distance from 2^{nd} ($t-1$)	-0.444*	-1.917***	-2.403***	-1.053
	(0.248)	(0.541)	(0.520)	(0.756)
ENP ($t-1$)	0.054	-0.271**	-0.393***	-0.090
	(0.043)	(0.096)	(0.121)	(0.090)
Con. Inc.	0.232	0.322***	-0.571***	-0.239**
	(0.144)	(0.110)	(0.145)	(0.110)
Lab. Inc.	0.275 *	0.315***	0.005	0.029
	(0.163)	(0.112)	(0.136)	(0.115)
Election Fixed Effect	Yes	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes

	Compositional-Outcome Stage				
	Log Ratio				
	$\log\left(\frac{\text{Labour}}{\text{Conservative}}\right)$	$\log\left(\frac{\text{Lib Dem}}{\text{Conservative}}\right)$	$\log\left(\frac{\text{Green}}{\text{Conservative}}\right)$	$\log\left(\frac{\text{UKIP}}{\text{Conservative}}\right)$	$\log\left(\frac{\text{Others}}{\text{Conservative}}\right)$
Constant	-1.875*	-1.766	-10.631***	-12.051**	-10.474**
	(1.037)	(1.544)	(1.878)	(2.343)	(5.013)
Log Ratio ($t-1$)	1.386***	2.077***	3.109***	2.611***	4.293***
	(0.001)	(0.228)	(0.237)	(0.258)	(1.171)
Brexit Vote in 2016	-0.011***	-0.031***	-0.016***	0.004	-0.007
	(0.001)	(0.002)	(0.002)	(0.003)	(0.007)
Unemp. Benefits	0.051***	-0.001	0.031*	0.127***	0.012
	(0.011)	(0.016)	(0.017)	(0.028)	(0.057)
Log Electorate	0.094	0.103	0.698***	0.635***	0.495
	(0.092)	(0.135)	(0.167)	(0.203)	(0.439)
Share Rural	-0.339***	0.316***	0.229***	-0.232*	0.278
	(0.041)	(0.064)	(0.064)	(0.126)	(0.211)
Con Inc.	0.317***	0.138	0.213	0.460***	-0.050
	(0.062)	(0.200)	(0.131)	(0.151)	(0.305)
Lab. Inc.	0.482***	-0.007	0.319**	0.504***	0.118
	(0.066)	(0.204)	(0.130)	(0.137)	(0.302)
Election Fixed Effect	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes
Correlation		0.896***	0.889***	0.887***	0.824**
		(0.014)	(0.012)	(0.007)	(0.331)
Log-likelihood		-1164.805	-1259.518	-1089.786	-1444.538
AIC		2403.61	2593.035	2253.572	2963.076
BIC		2587.458	2776.883	2437.419	3146.923
Note:					*p<0.1;**p<0.05;***p<0.01

Table S2: CMS Results—2017 & 2019 Elections in England

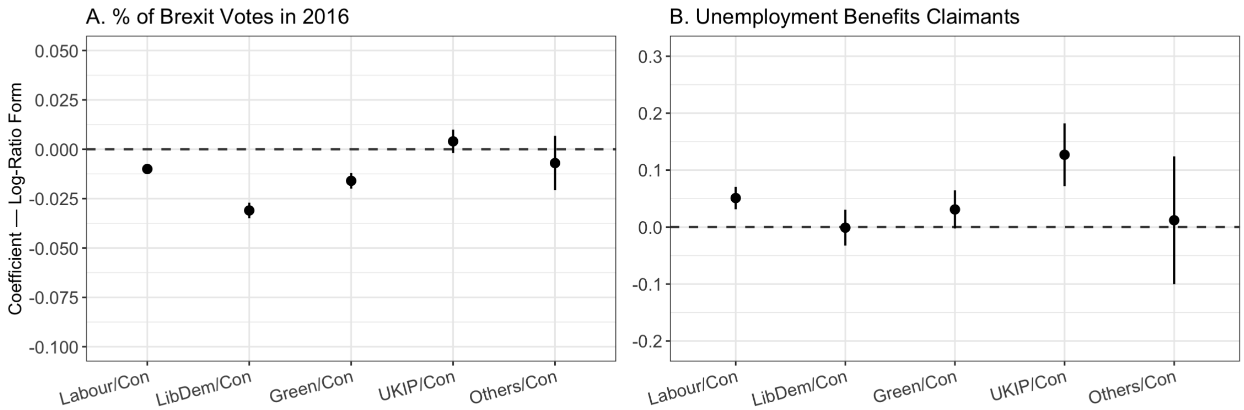


Figure S1: Coefficients in the Outcome Stage

In this section, we show results from our CMS using data from the 2017 and 2019 elections in England. Table S2 presents the coefiicient estimated in first and second stages across log-ratio equations. Figure S1 shows coefficients in the outcome stage for the percentage of votes for Brexit in the 2016 referendum (Brexit Vote) and the economic variable (Unemployment Benefits Claimants) across log-ratio equations. Finally, Figure S2 compares coefficients for both predictors across all five alternatives studied in our Monte Carlo experiment: CMS, Heckman, TTW, Tiny Value, and Dummy Variable. As the results demonstrate, these alternative strategies may return very different estimates with real-world data.

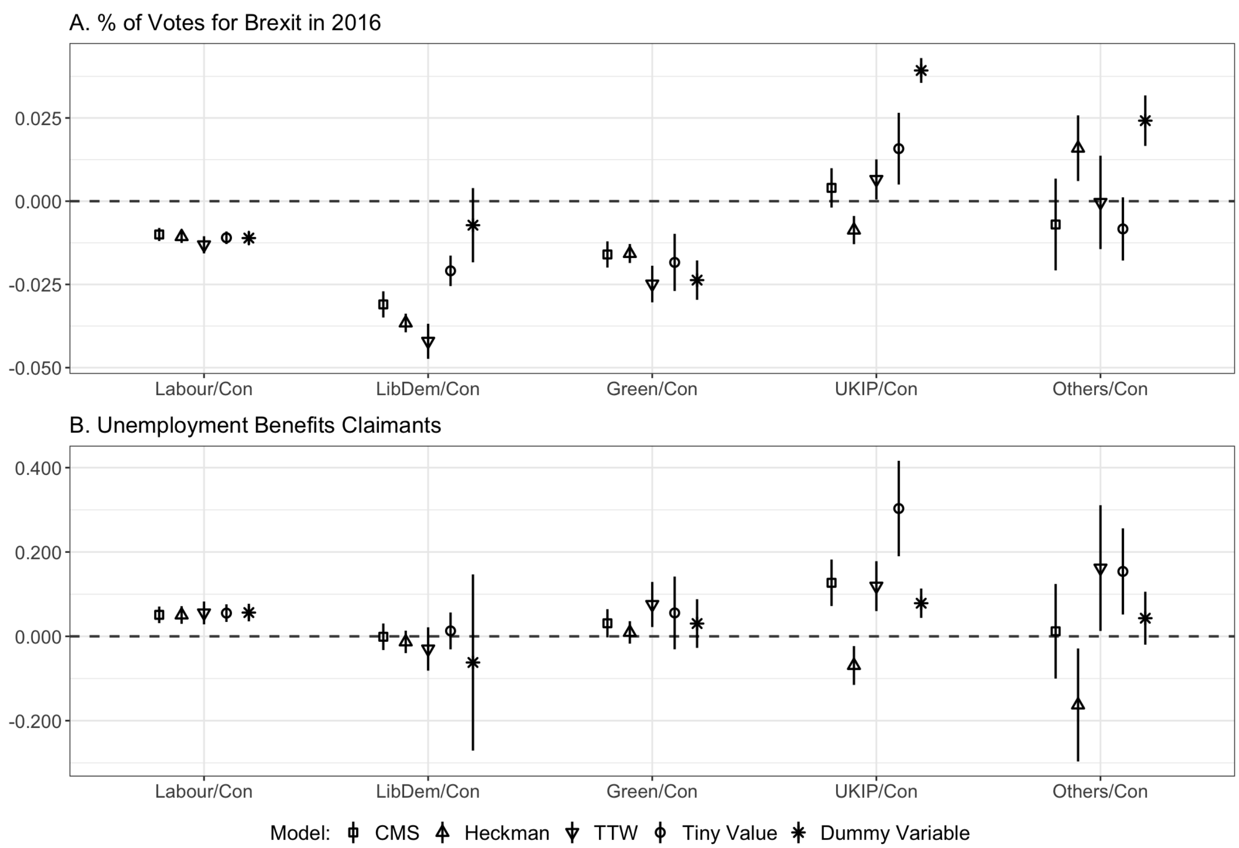


Figure S2: Coefficients From all Competing Approaches

C.3 Predicted Vote Shares

All figures in this section compare predicted vote shares for one party across all five alternative strategies (panels) discussed in our Monte Carlo experiment: CMS, Heckman, TTW, Tiny Value, and Dummy Variable. Each plot shows predicted (vertical axis) and observed (horizontal axis) vote shares such that good predictions are close to the 45 degree straight line. In this section, we only compare predictions in districts where a party indeed contested (i.e., vote shares were observed). The CMS's performances are similar to those from strategies that ignore sample selection, but Heckman can lead to extreme differences between predicted and observed vote shares of large parties (Conservatives, Labour, and Liberal Democrats).

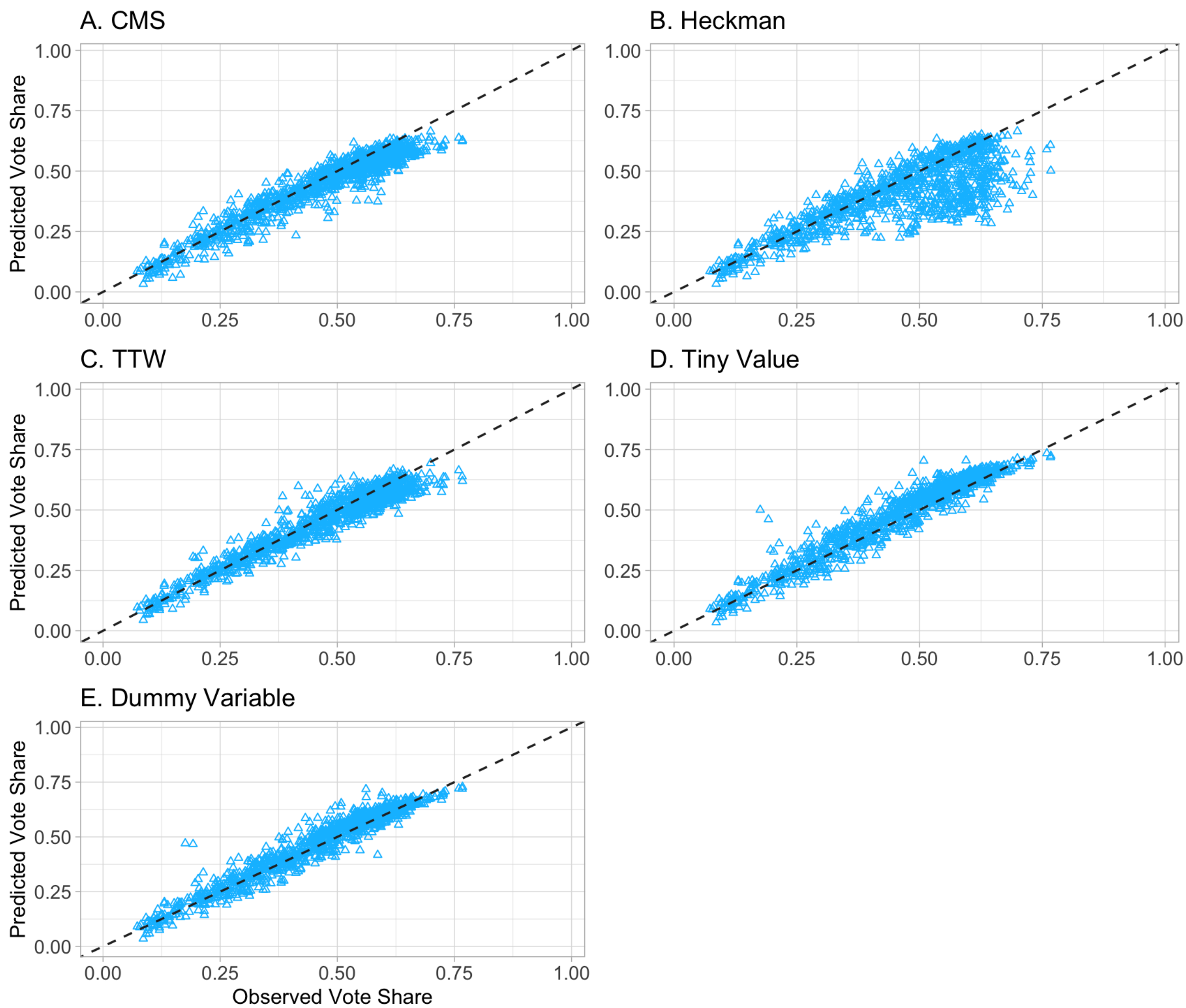


Figure S3: Predicted Vote Shares for Conservatives

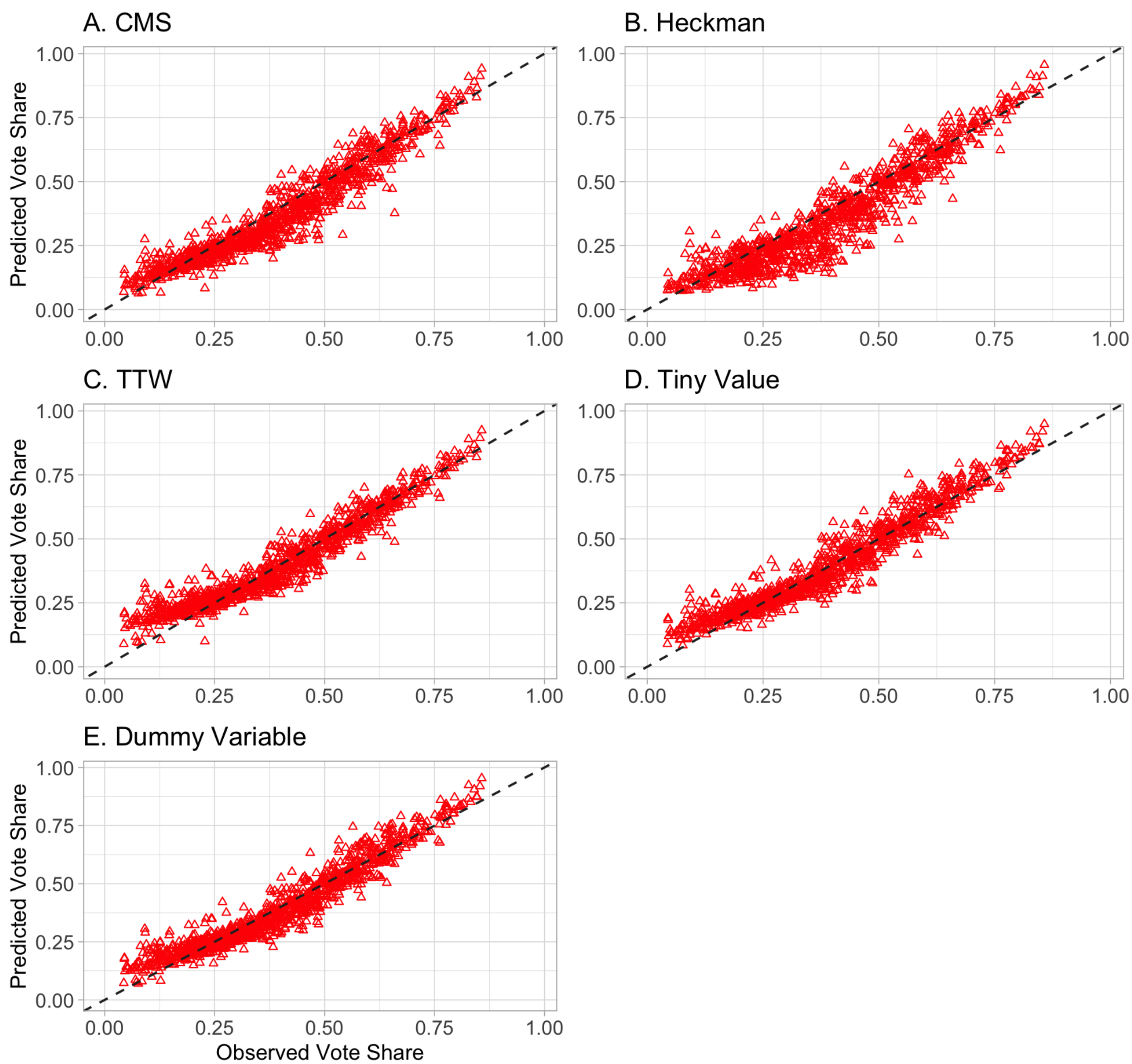


Figure S4: Predicted Vote Shares for the Labour Party

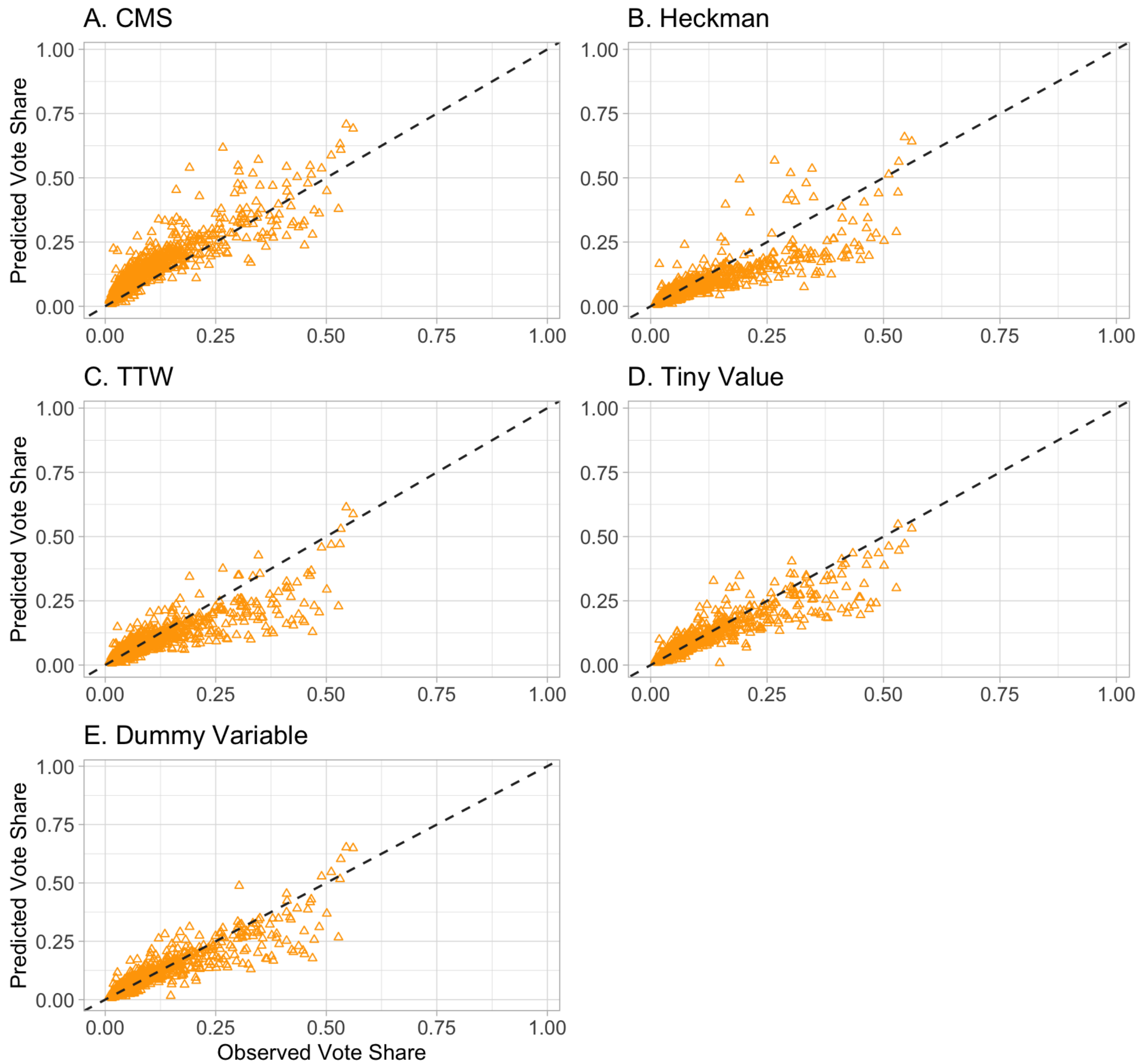


Figure S5: Predicted Vote Shares for Liberal Democrats

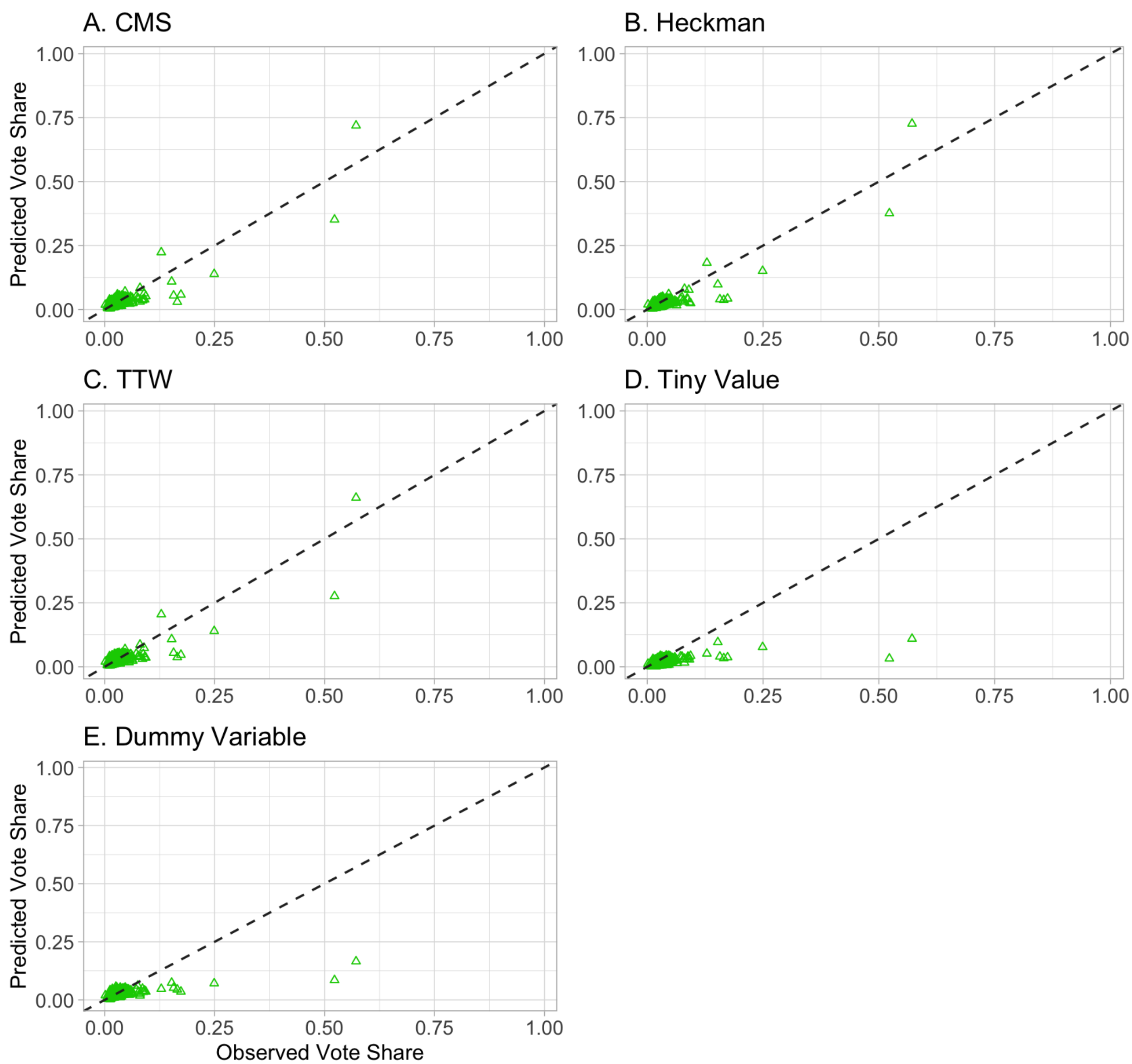


Figure S6: Predicted Vote Shares for the Green Party

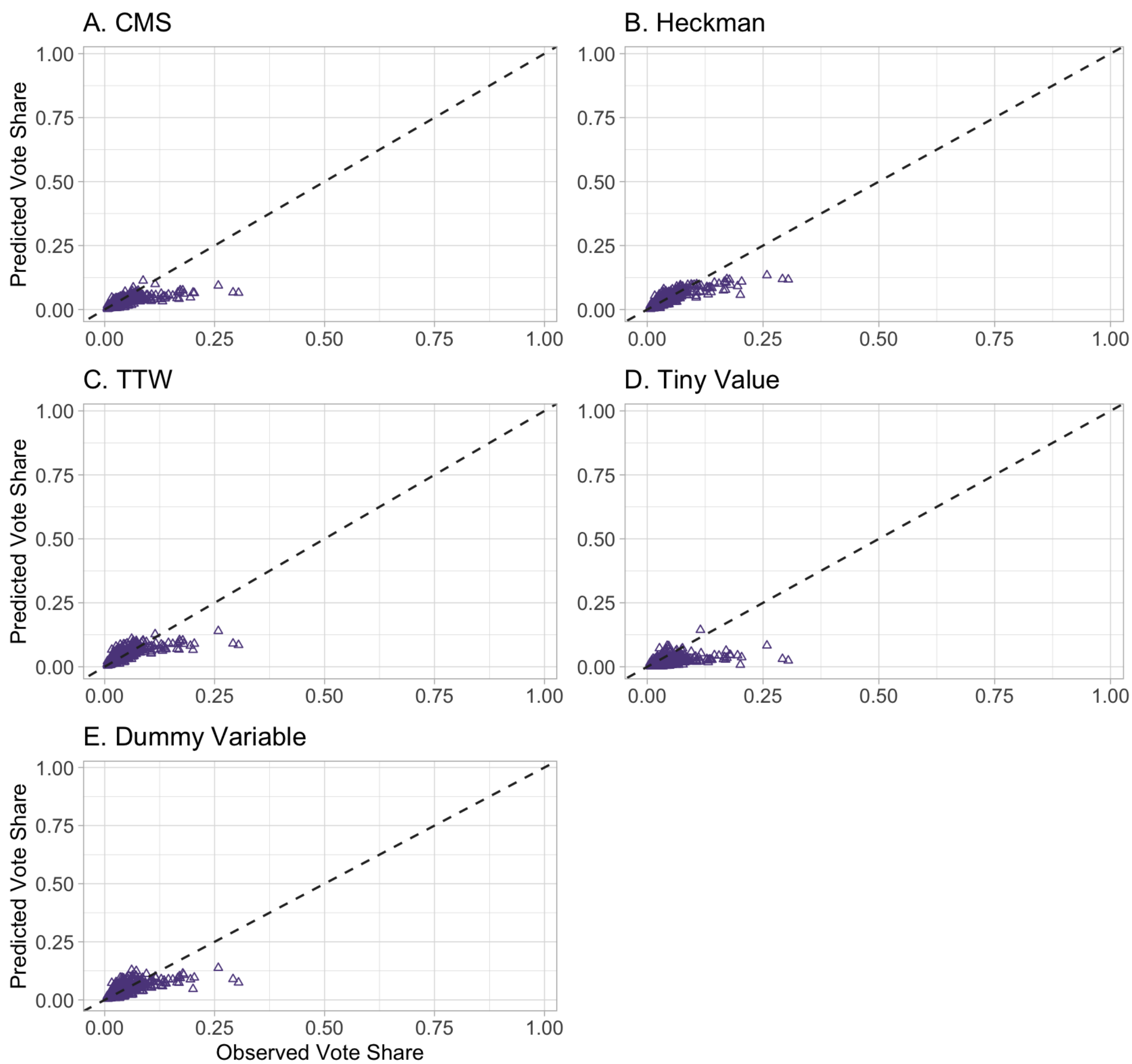


Figure S7: Predicted Vote Shares for the UKIP

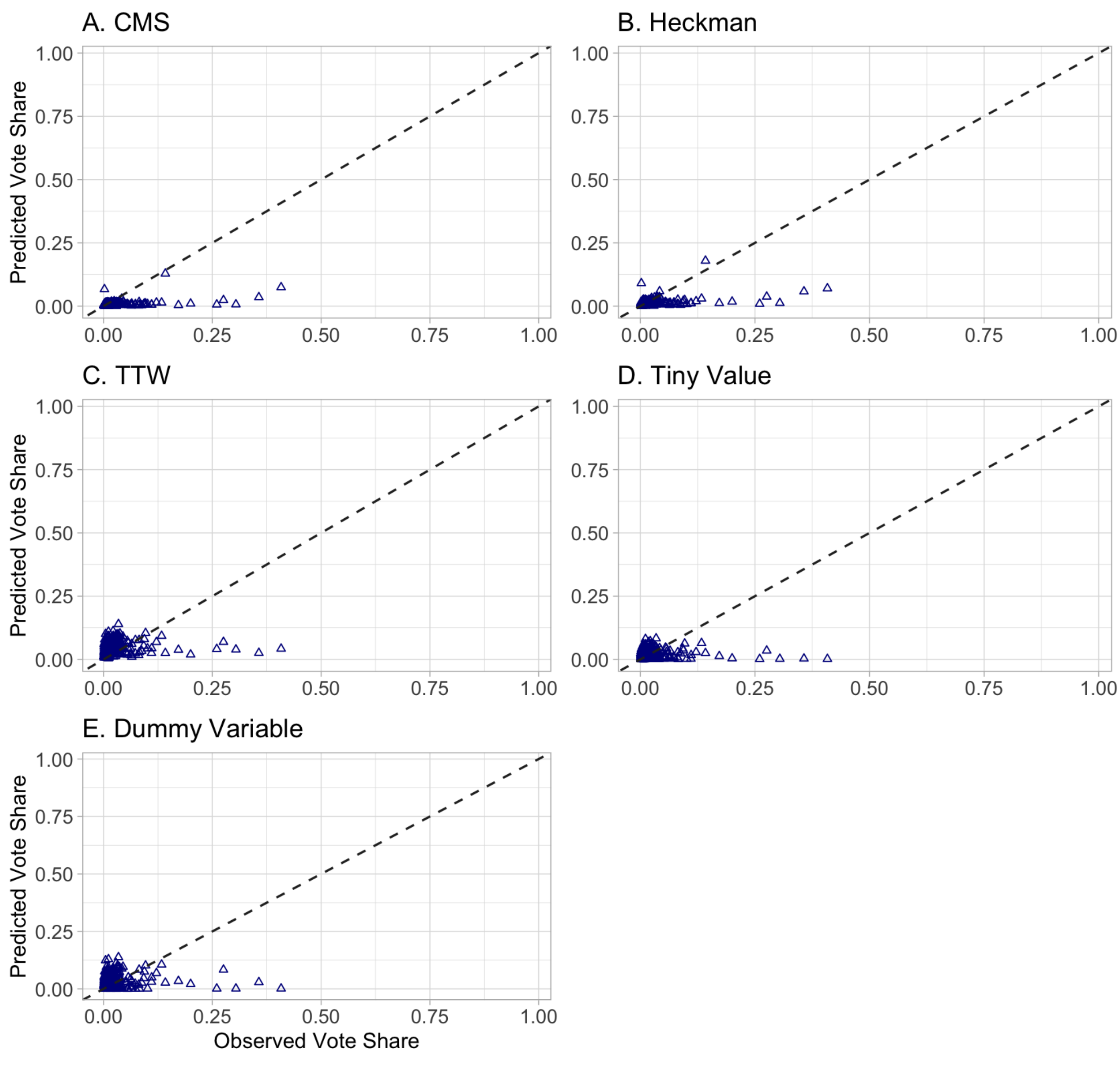


Figure S8: Predicted Vote Shares for Others

C.4 Predictions in Fully & Partially Contested Districts

Figures in this section compare predicted vote shares by CMS with observed vote shares in fully (Figure S9) and partially contested districts (Figure S10). We only compare predictions in districts where a party indeed consteted (i.e., its vote share was observed). As the results demonstrate, there is no systematic difference in how our CMS approach performs in fully and partially contested constituencies.

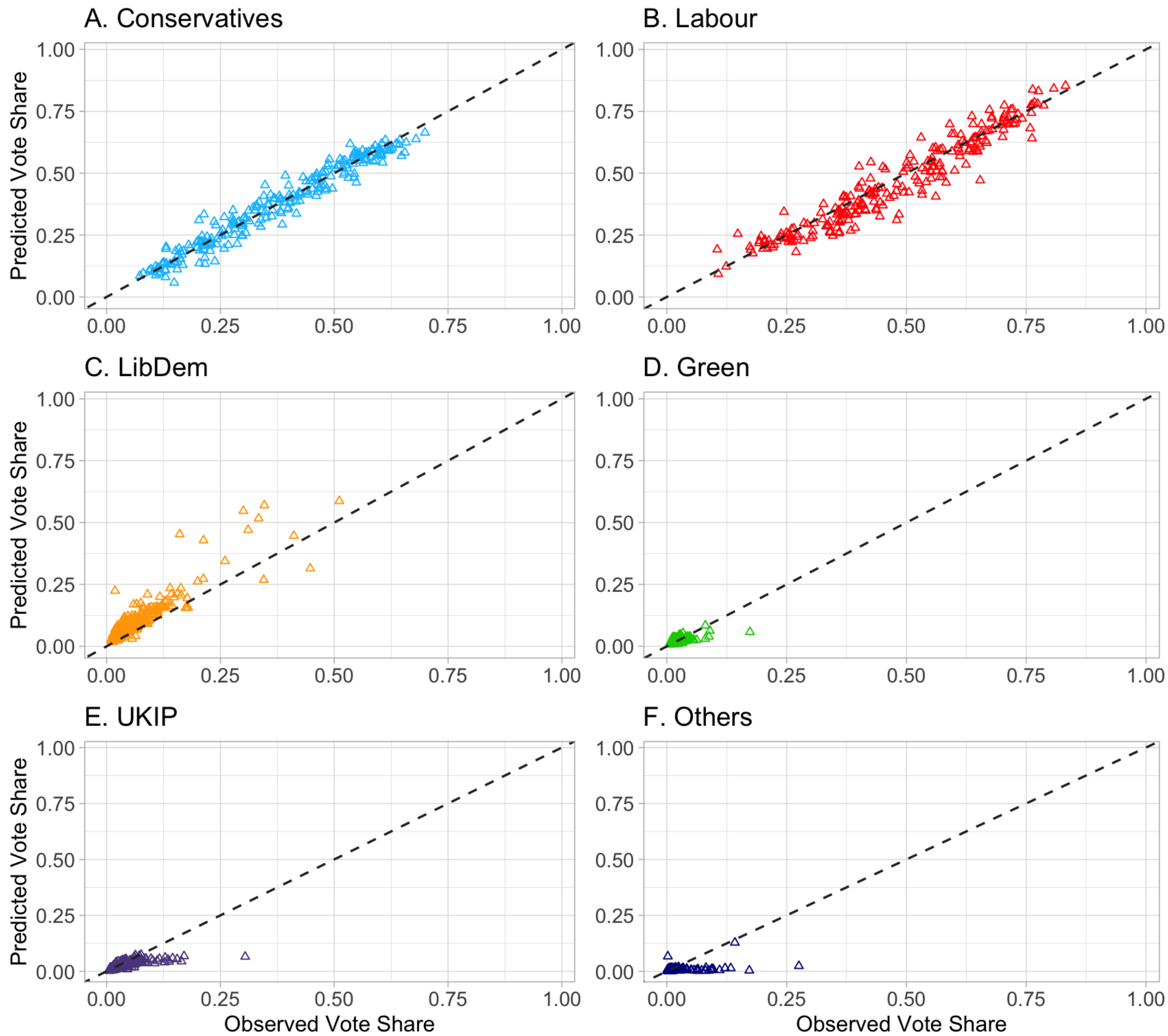


Figure S9: Vote Shares Predicted by the CMS—Fully Contested Districts

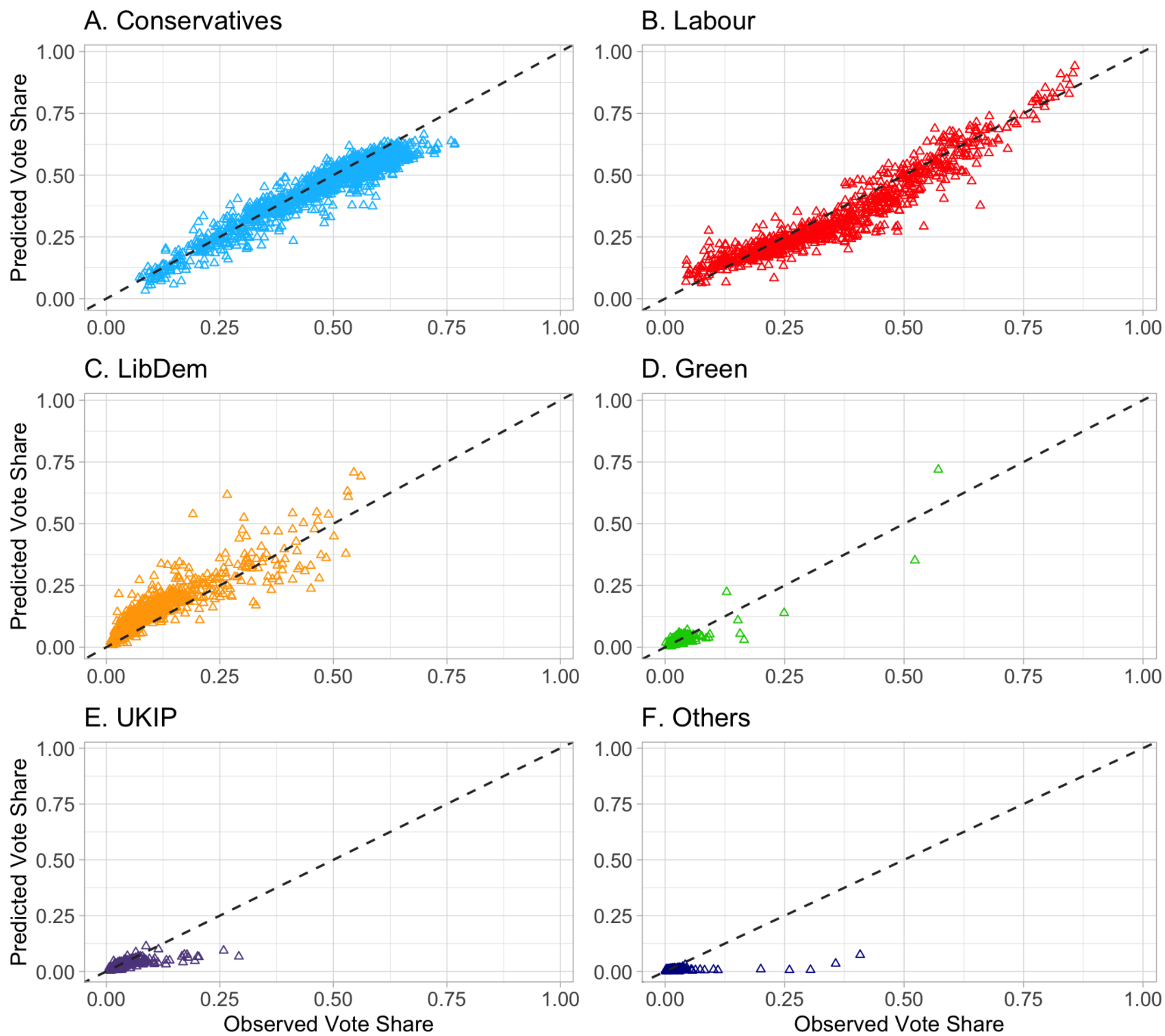


Figure S10: Vote Shares by the CMS—Partially Contested Districts

C.5 Out-of-Sample Prediction

In this section, we show predicted vote shares in out-of-sample data. To do so, we estimate the CMS with only 70% of all constituencies in the 2017 and 2019 elections in England. Then, we use the coefficients from this model to predict the vote shares of political parties in the remaining 30% of districts that were not used to estimate the model. Each figure compares predictions for a specific party across all the five alternatives (panels) examined in our Monte Carlo expertiment (CMS, Heckman, TTW, Tiny Value, and Dummy Variable). These results are very similar to those presented in Section C.3, with the CMS's performances being similar to those from strategies that ignore sample selection, but Heckman can lead to extreme differences between predicted and observed vote shares of large parties (Conservatives, Labour, and Liberal Democrats).

	RMSE Using Out-of-Sample Predictions and Observed Vote Shares						
Approach	Conservatives	Labour	Lib Dem	Greens	UKIP	Others	Total
CMS	0.054	0.061	0.068	0.021	0.036	0.022	0.047
Heckman	0.088	0.068	0.058	0.019	0.116	0.070	0.076
TTW	0.047	0.054	0.054	0.022	0.037	0.040	0.044
Tiny Value	0.049	0.050	0.043	0.044	0.033	0.026	0.042
Dummy Variable	0.047	0.051	0.042	0.040	0.024	0.035	0.041

Table S3: Root Mean Square Error (RMSE) Across Approaches

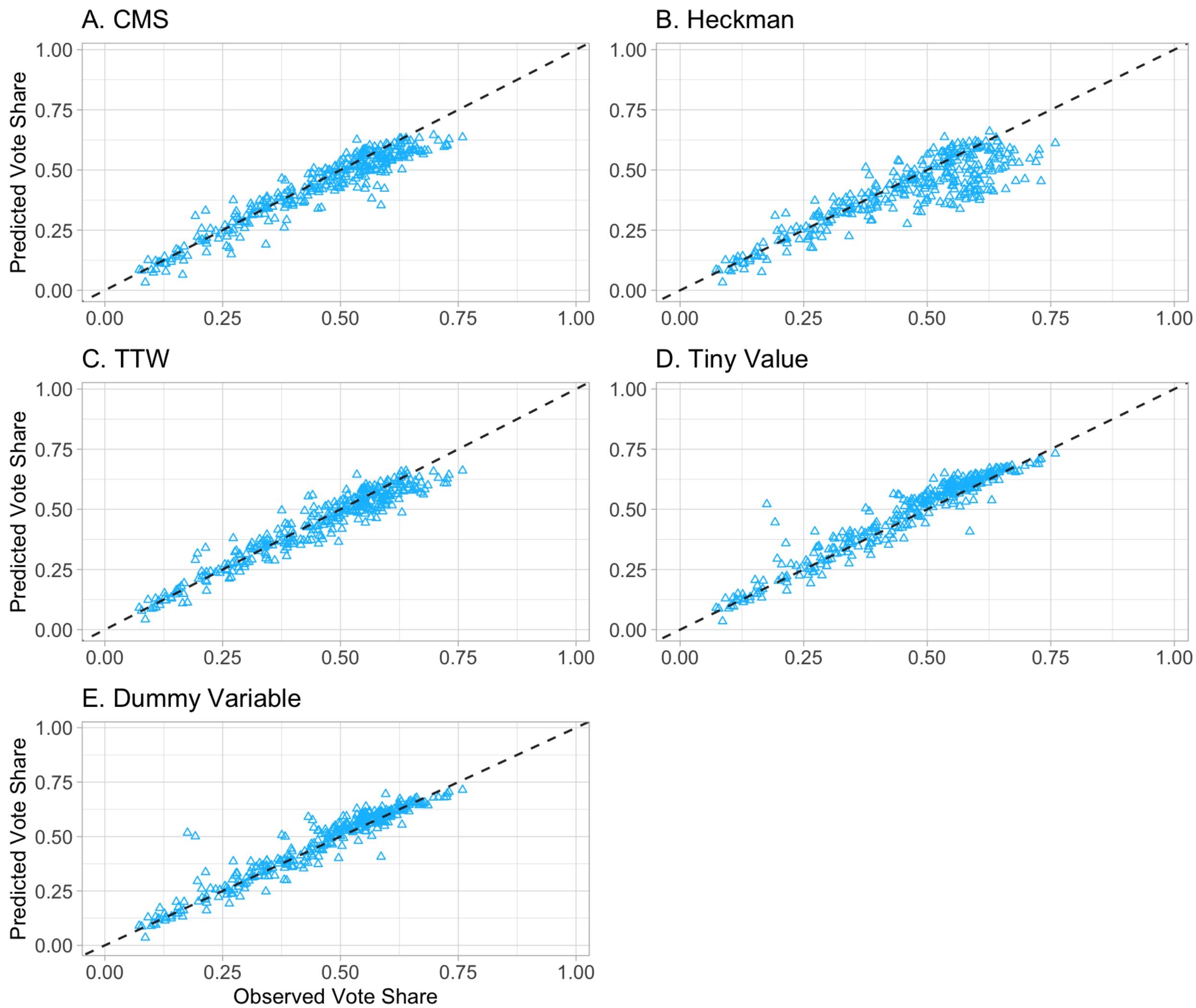


Figure S11: Out-of-Sample Predictions—Conservatives

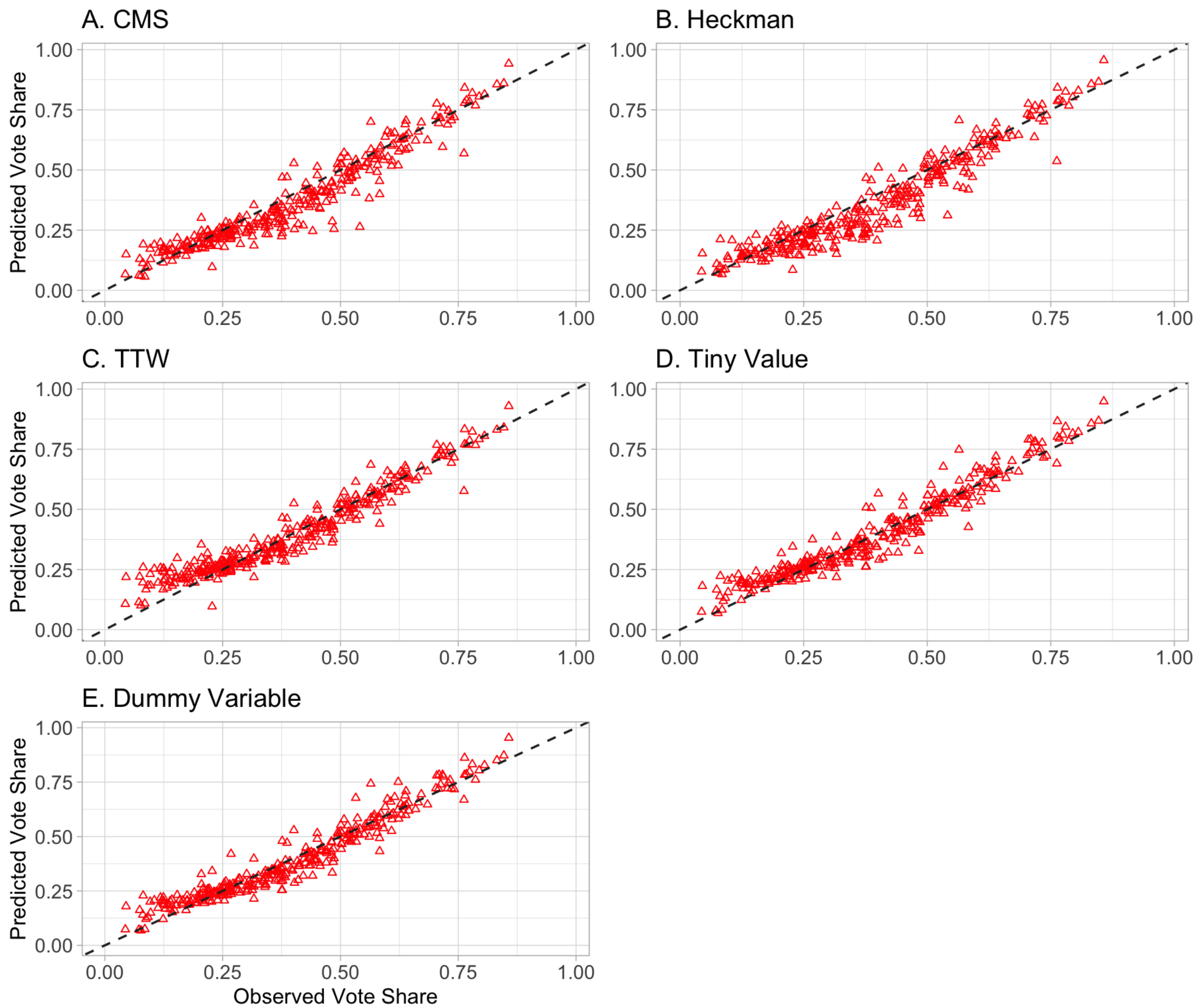


Figure S12: Out-of-Sample Predictions—Labour

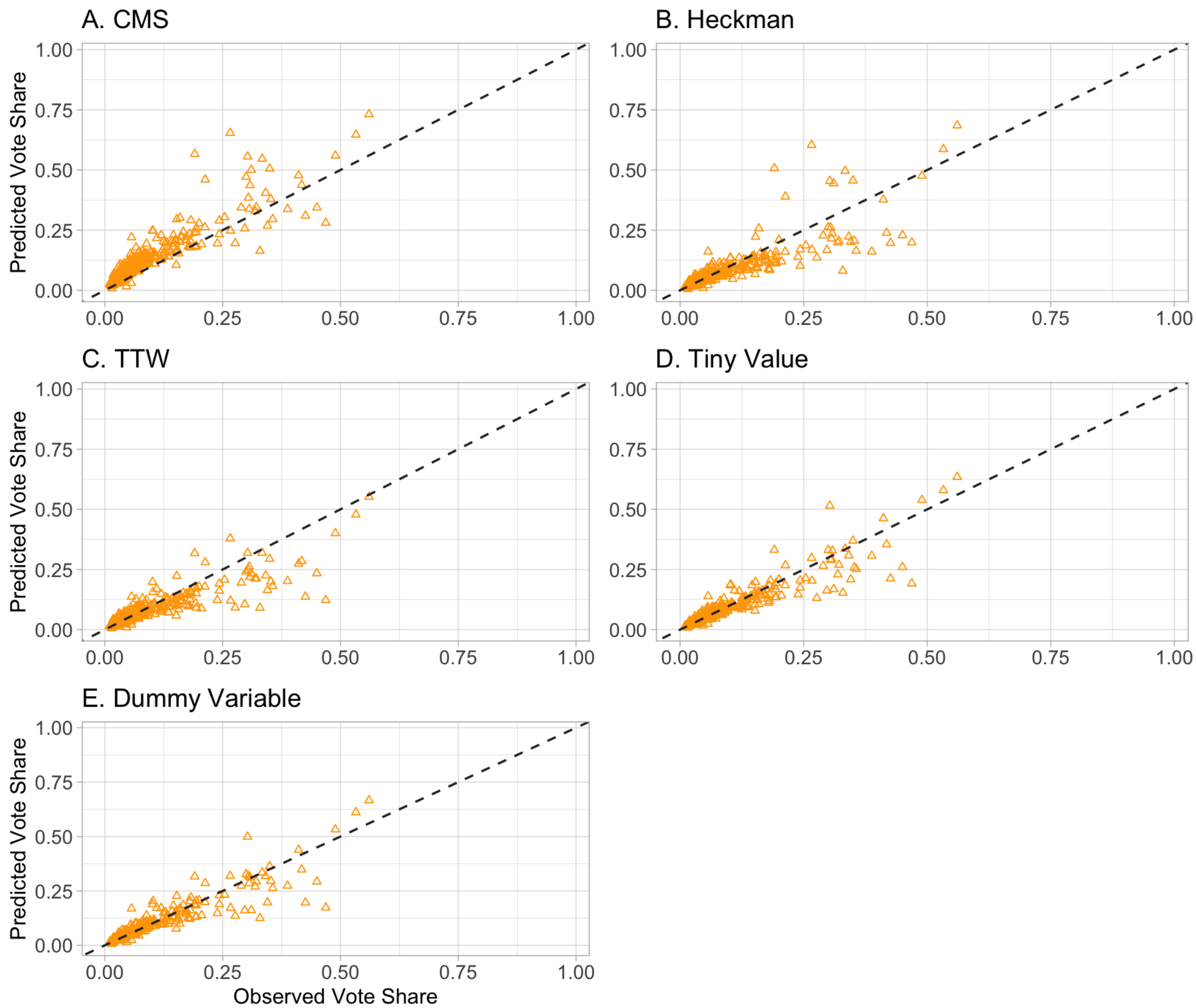


Figure S13: Out-of-Sample Predictions—Lib Dem

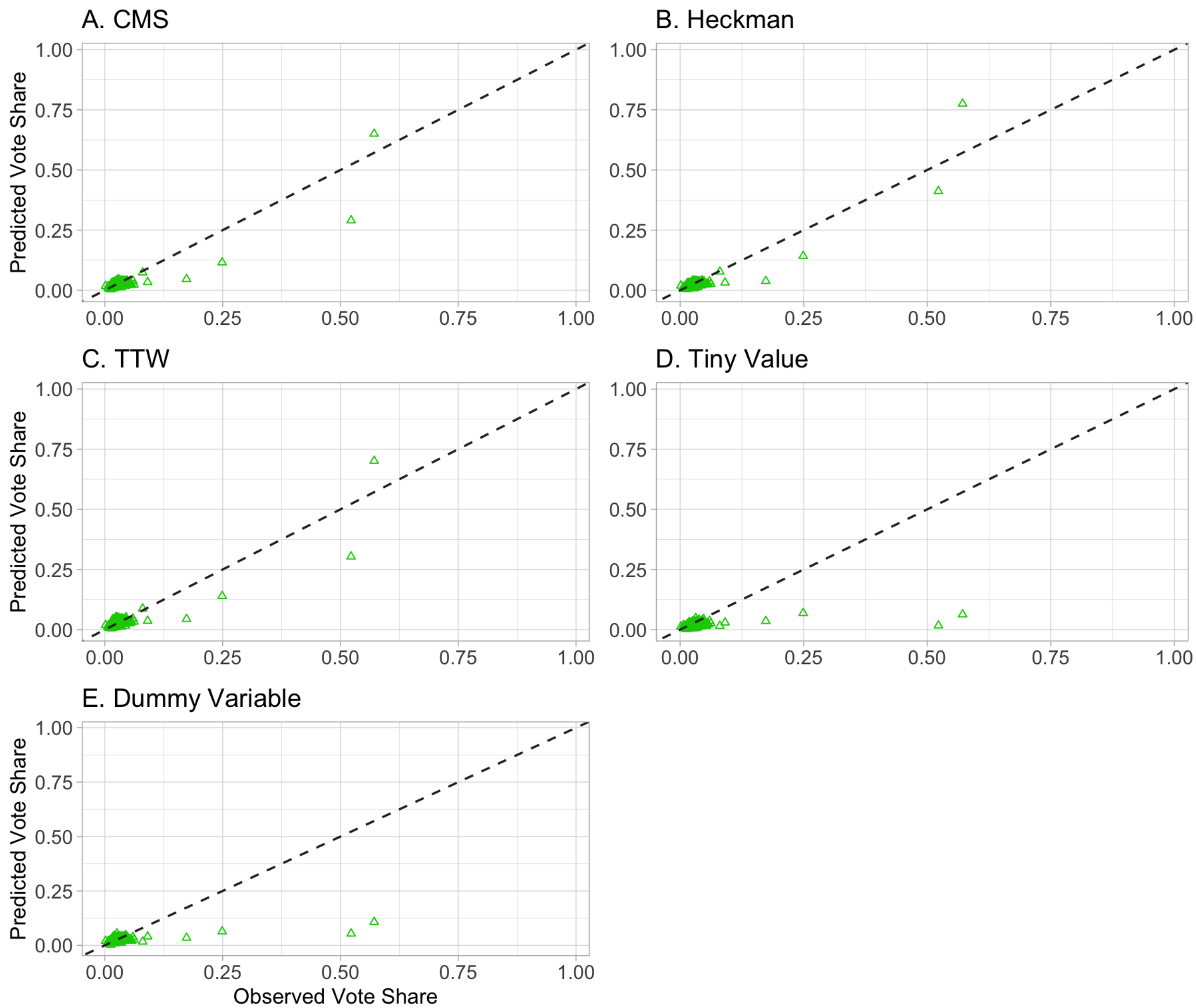


Figure S14: Out-of-Sample Predictions—Greens

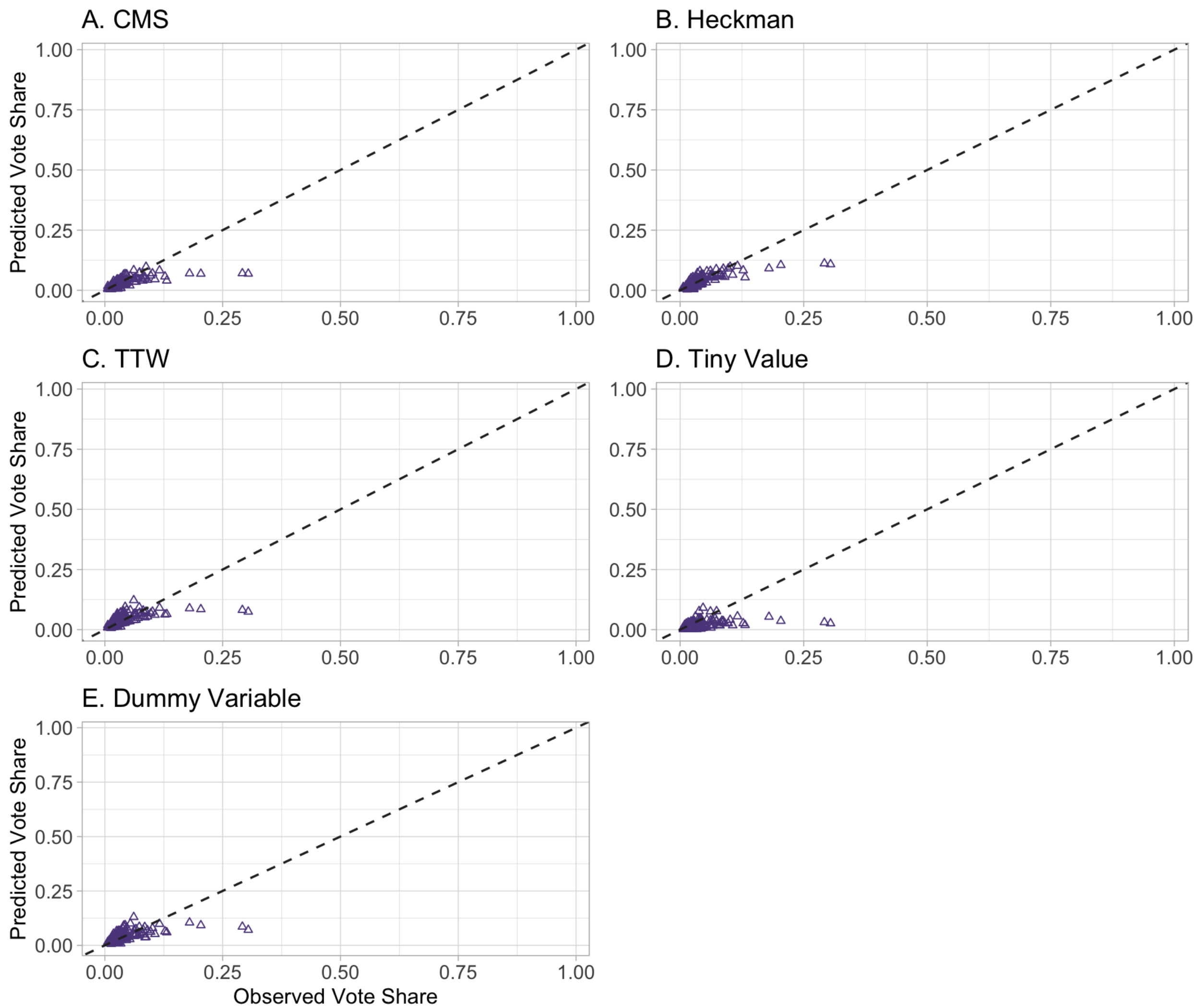


Figure S15: Out-of-Sample Predictions—UKIP

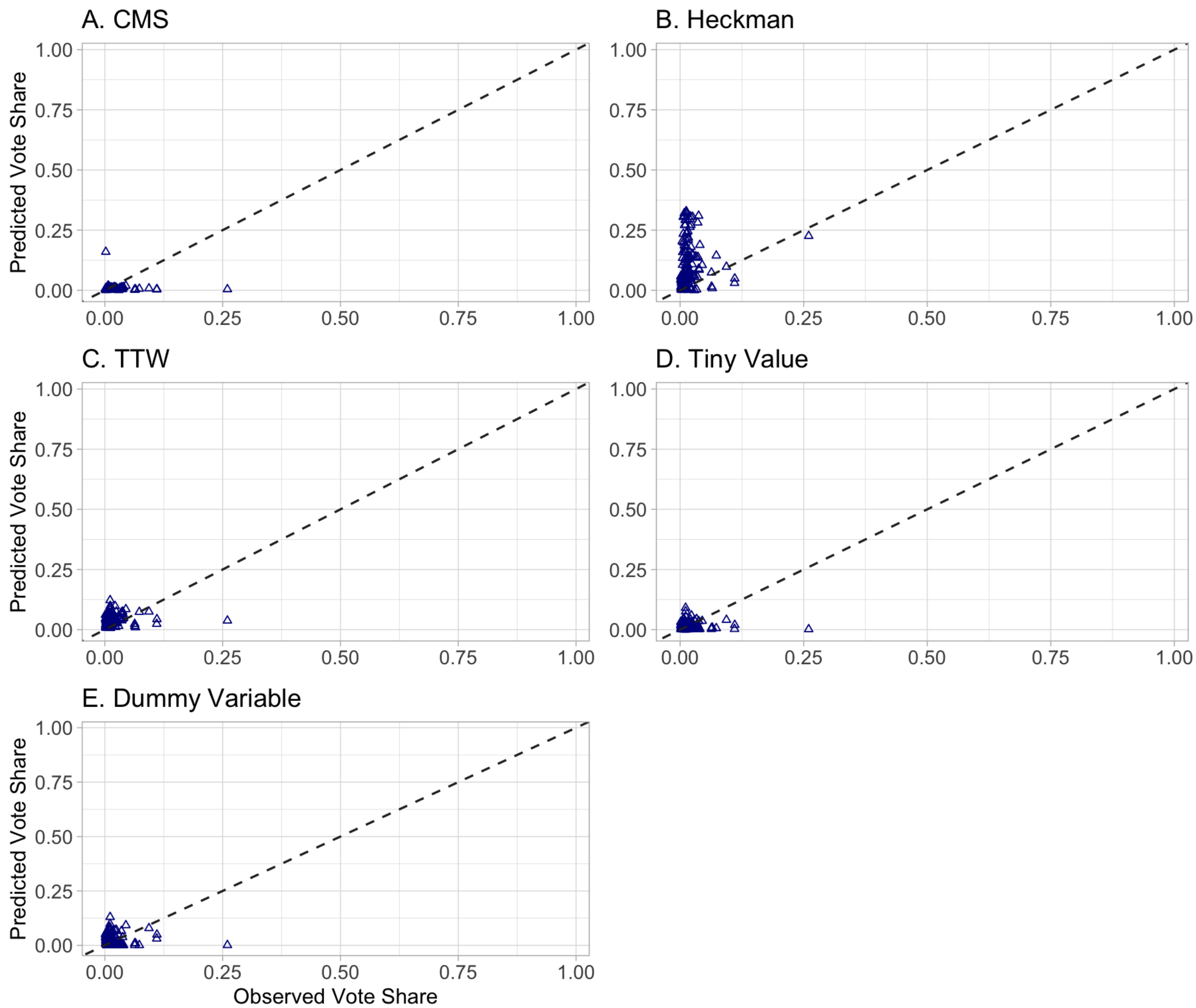


Figure S16: Out-of-Sample Predictions—Others

C.6 Robustness Check—House Prices Change

	Selection Stage			
	Party			
	Lib Dem	Green	UKIP	Others
Constant	2.050***	3.817***	-0.058	-3.288
	(0.735)	(1.446)	(1.878)	(2.302)
Change House Prices	0.039	0.017	-0.002	0.086*
	(0.037)	(0.023)	(0.032)	(0.046)
Unemp. Benefits	0.154**	0.019	0.042*	0.024
	(0.061)	(0.018)	(0.024)	(0.037)
Log Electorate	-0.090	-0.163	0.230	0.342*
	(0.063)	(0.124)	(0.163)	(0.176)
Share Rural	0.011	-0.044	0.122	0.100
	(0.144)	(0.080)	(0.110)	(0.149)
Top-Two Distance $(t-1)$	-0.247	-1.019**	-2.060***	-0.666
	(0.315)	(0.427)	(0.375)	(0.520)
Distance from 2^{nd} $(t-1)$	-0.153	-1.979***	-1.978***	-1.180
	(0.219)	(0.574)	(0.433)	(0.780)
ENP $(t-1)$	0.122	-0.282**	-0.341***	-0.079
	(0.077)	(0.104)	(0.094)	(0.118)
Con. Inc.	0.140	0.327***	-0.683***	-0.137
	(0.127)	(0.105)	(0.155)	(0.106)
Lab. Inc.	0.298*	0.340***	0.005	-0.023
	(0.171)	(0.112)	(0.152)	(0.115)
Election Fixed Effect	Yes	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes

	Compositional-Outcome Stage				
	Log Ratio				
	$\log\left(\frac{\text{Labour}}{\text{Conservative}}\right)$	$\log\left(\frac{\text{Lib Dem}}{\text{Conservative}}\right)$	$\log\left(\frac{\text{Green}}{\text{Conservative}}\right)$	$\log\left(\frac{\text{UKIP}}{\text{Conservative}}\right)$	$\log\left(\frac{\text{Others}}{\text{Conservative}}\right)$
Constant	-3.276***	-7.485***	-12.934***	-10.663***	-9.883**
	(1.063)	(1.731)	(1.914)	(2.149)	(4.923)
Log Ratio $(t-1)$	1.455***	2.565***	3.799***	2.431***	4.459***
	(0.035)	(0.266)	(0.271)	(0.218)	(1.207)
Change House Prices	0.100***	0.228***	0.126***	-0.043	-0.051
	(0.015)	(0.029)	(0.032)	(0.036)	(0.078)
Unemp. Benefits	0.017	-0.064***	-0.016	0.190***	-0.003
	(0.010)	(0.022)	(0.016)	(0.038)	(0.056)
Log Electorate	0.138	0.376**	0.775***	0.532***	0.427
	(0.100)	(0.156)	(0.170)	(0.192)	(0.453)
Share Rural	-0.346***	0.385***	0.189***	-0.265**	0.610***
	(0.043)	(0.093)	(0.058)	(0.129)	(0.201)
Con Inc.	0.255***	0.226	0.266**	0.489***	-0.075
	(0.065)	(0.216)	(0.131)	(0.147)	(0.302)
Lab. Inc.	0.434***	0.238	0.429***	0.470***	0.282
	(0.069)	(0.207)	(0.129)	(0.101)	(0.309)
Election Fixed Effect	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes
Correlation		0.870***	0.893***	0.894***	0.768**
		(0.024)	(0.014)	(0.009)	(0.350)
Log-likelihood		-1200.661	-1265.963	-1087.51	-1440.838
AIC		2475.322	2605.926	2249.02	2955.675
BIC		2659.169	2789.774	2432.868	3139.523
Note:					*p<0.1; **p<0.05; ***p<0.01

Table S4: CMS Results—2017 & 2019 Elections in England (House Prices)

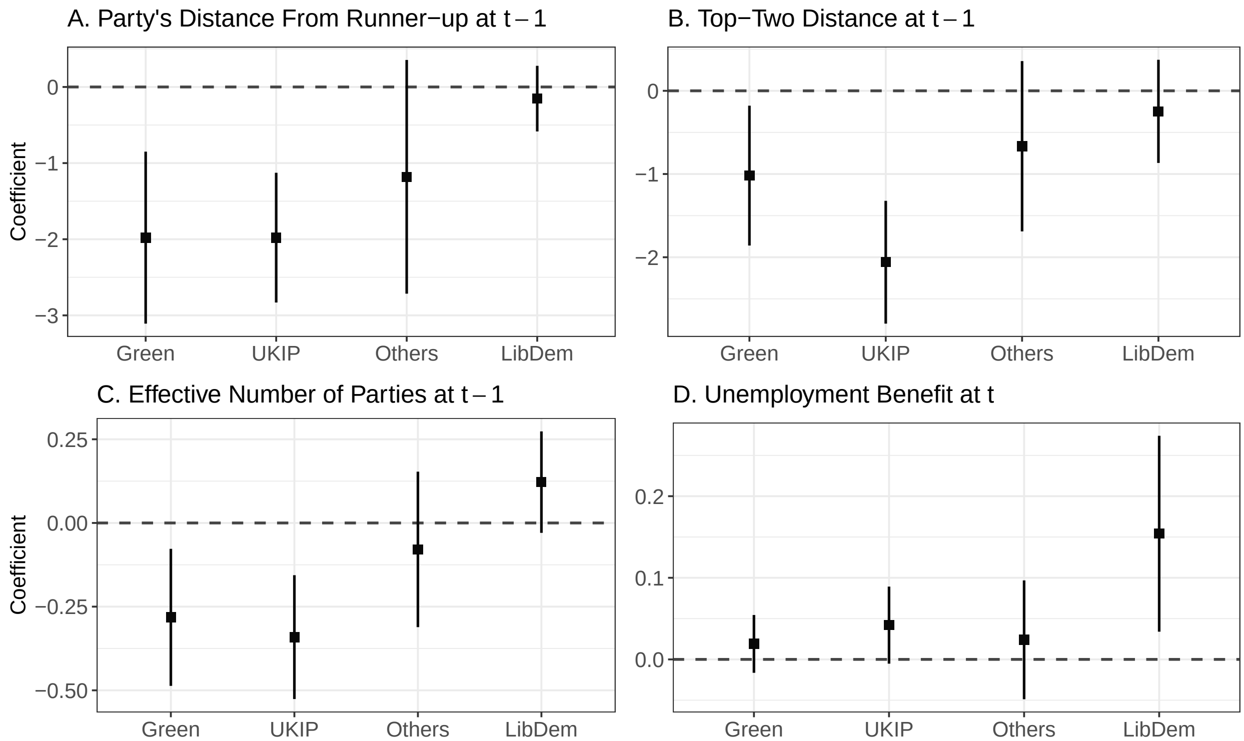


Figure S17: Coefficients in the Selection Stage—Model Estimated With House Prices

Figure S18: Coefficients in the Outcome Stage—Model Estimated With House Prices

In this section, we show results estimated by our CMS with the 2017 and 2019 elections in England using the average change in house prices as a proxy for Brexit support. Table S4 presents the coefiicient estimated in the first and second stages across log-ratio equations. Figure S17 displays coefficients in the selection stage. Figure S18 shows estimated

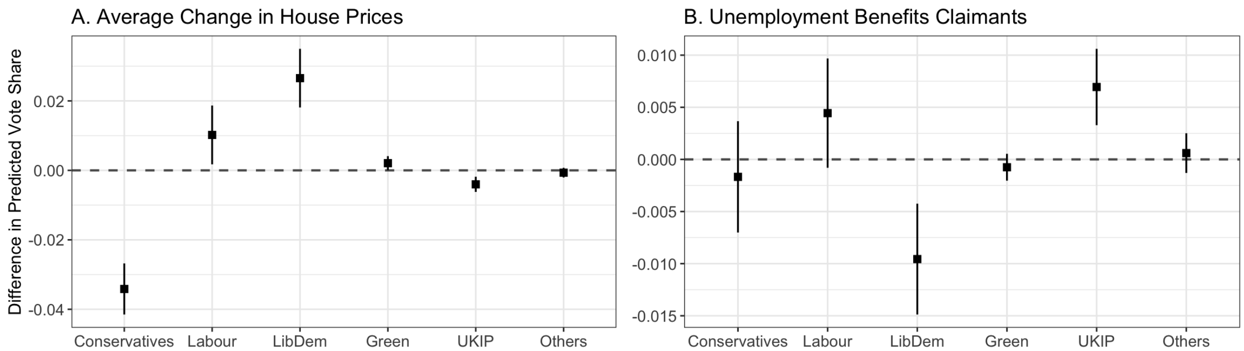


Figure S19: Change in Vote Share in Response to +1SD Shock—Model Estimated With House Prices

coefficients average changes in house prices (proxy for Brexit support) and the economic variable (Unemployment Benefits Claimants) in the outcome stage. Finally, Figure S19 presents difference in vote shares for all parties after a one standard deviation increase (+1SD shock) in either house prices or unemployment benefits, holding all other variables in their observed values.

D Appendix D: Monte Carlo Results

In this Appendix, we present additional comparisons of approaches in the Monte Carlo experiments. Following Hopkins et al. (2024), we compare the strategies across six performance statistics: bias, standard deviation of estimated parameters ($\hat{\beta}$), mean squared errors (MSE), confidence, coverage probability, and statistical power. Confidence is the ratio of the standard deviation of estimated parameters to the mean of standard errors, $\hat{\sigma}_{\beta}$:

$$\text{Confidence} : \frac{\sqrt{\frac{\sum_{i=1}^{N}(\hat{\beta}-\bar{\beta})^2}{N}}}{\text{mean}(\hat{\sigma}_{\beta})}$$

such that values larger than 1 indicate that $\hat{\sigma}_{\beta}$ is smaller than the standard deviation of $\hat{\beta}$s (overconfidence), making the approach more likely to reject a true null-hypothesis (type-I error). Values smaller than one, in turn, indicate underconfidence: the standard errors are larger than they should be, increasing the chance of failing to reject a false null-hypothesis (type-II error).

MSE evaluates the accuracy of an estimator by capturing the total error (i.e., how far predictions typically are from the truth) that stem from both bias and variance:

$$\text{MSE} = \mathbb{E}[\hat{\beta} - \beta]^2 + \mathbb{V}(\hat{\beta})$$

Coverage is calculated with the frequency with which 95% confidence intervals include the parameter:

$$\text{Coverage} = Pr(\hat{\beta} - 1.96 * \hat{\sigma}_{\beta} \leq \beta \leq \hat{\beta} + 1.96 * \hat{\sigma}_{\beta})$$

Finally, power is the frequency with which the approach rejects the false null-hypothesis:

$$\text{Power} = Pr(\text{Reject } H_0 \mid H_1 \text{ is true})$$

D.1 Valid Exclusion Restriction

This section compares the approaches when there is a valid instrument in the selection stage to estimate the inverse Mills ratio (i.e., the exclusion restriction holds). Strategies that do not account for sample selection are biased. Bias in TTW and dummy variable approaches increases with both ρ and the number of partially contested districts, and it does not decrease with the sample size, N. As a consequence, they also have low coverage. Heckman correction, in turn, is inefficient and overconfident, and both of these performance statistics get worse with partial contestation. The Heckman strategy's inefficiency stems from the fact that the errors ε_{ij} and u_{ij} are not specified as bivariate normal; rather they follow a bivariate extreme value type-I distribution. The difference between these distributions is small—"the extreme value distribution gives slightly fatter tails than a normal" (Train, 2009, p. 35)—but enough to make Heckman's estimators inefficient.

D.1.1 33% of Districts are Partially Contested

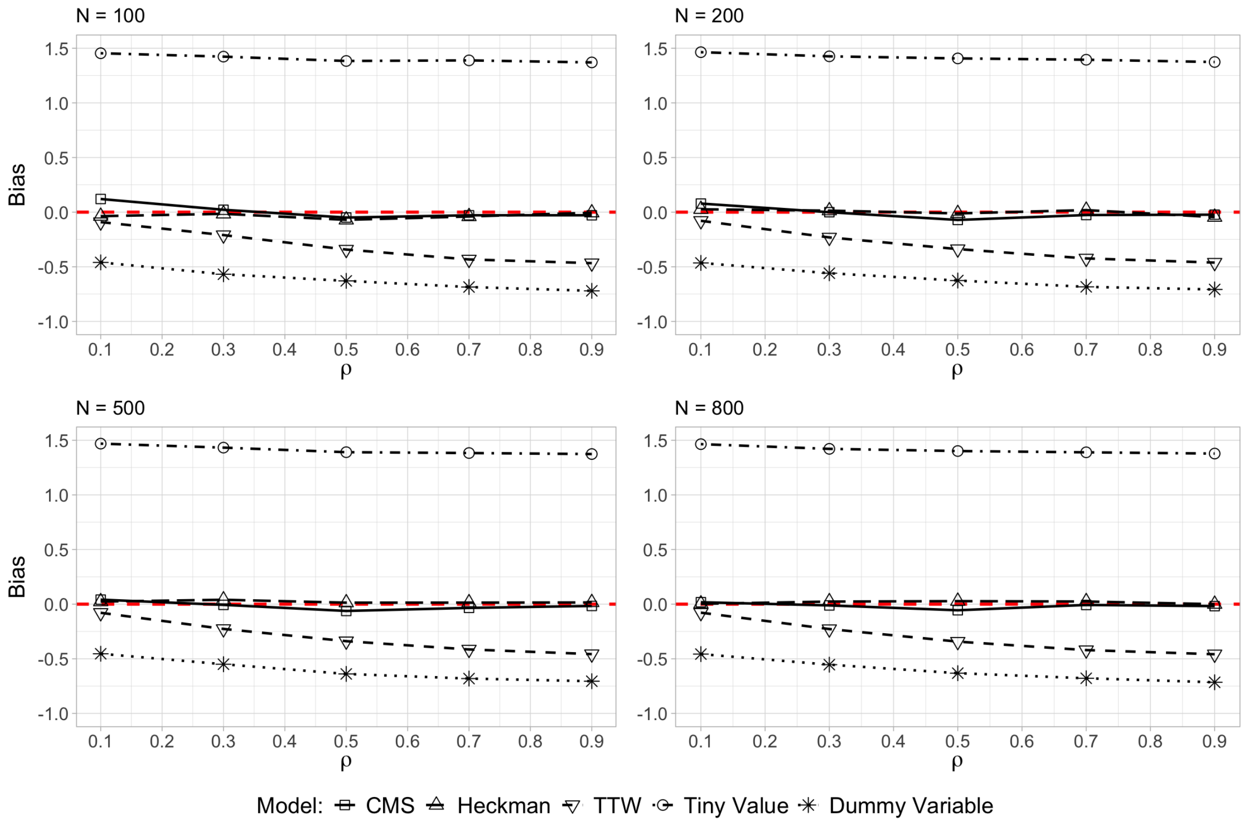


Figure S20: Bias in $\hat{\beta}_{1A}$—33% of Districts Partially Contested

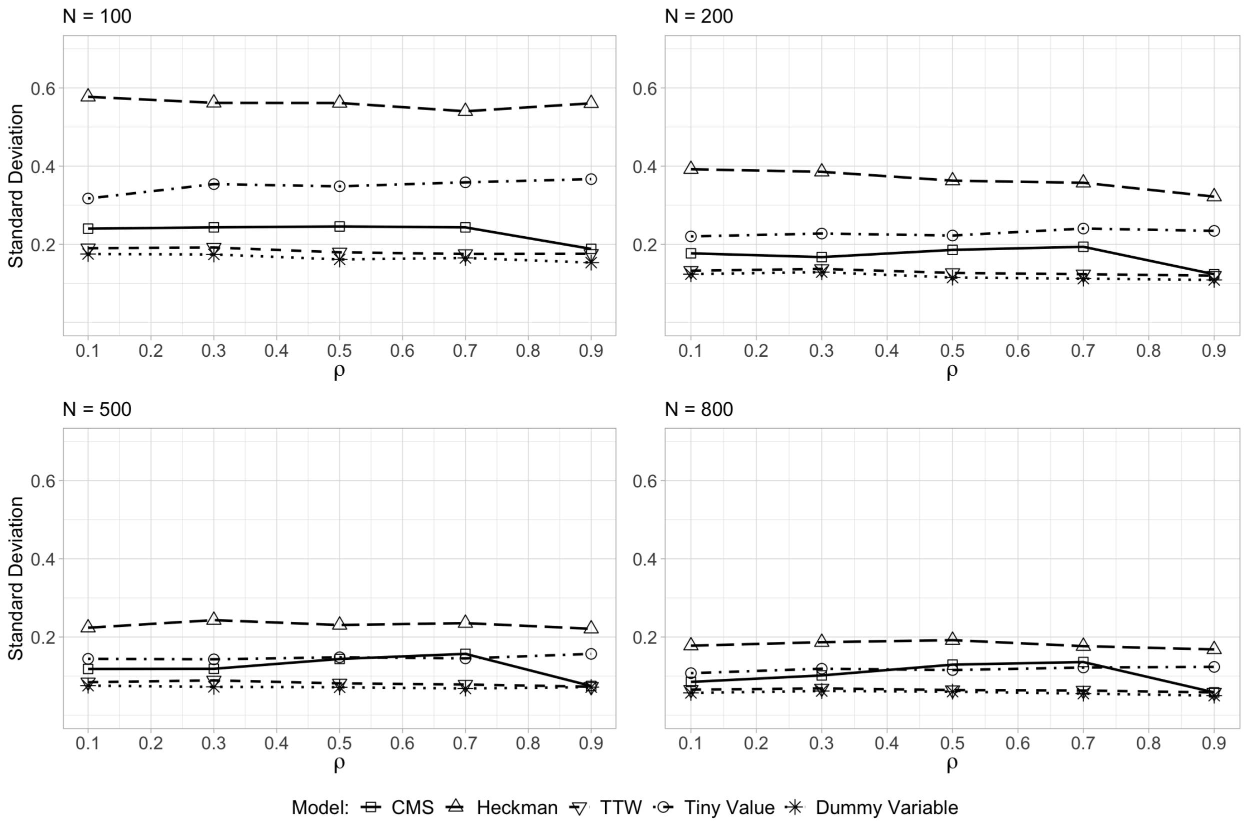


Figure S21: Standard Deviation of $\hat{\beta}_{1A}$—33% of Districts Partially Contested

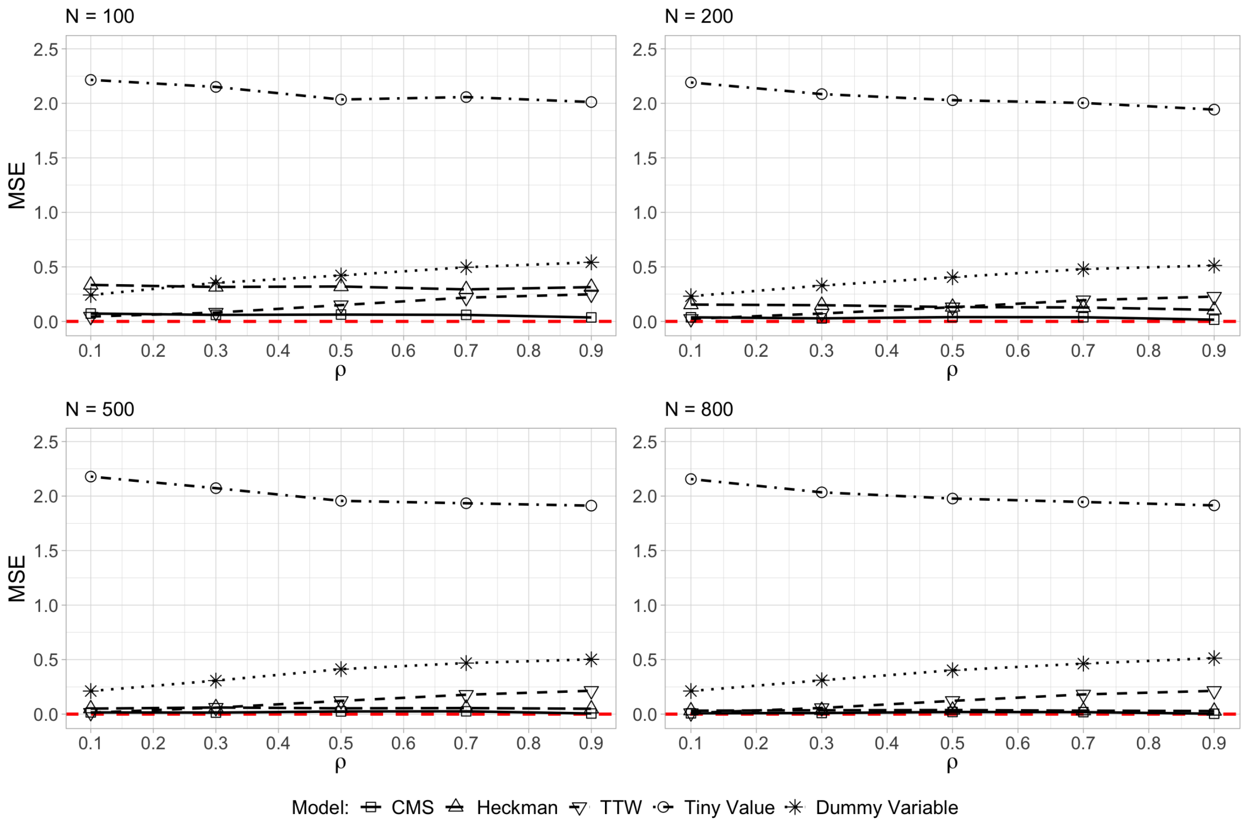


Figure S22: MSE of $\hat{\beta}_{1A}$—33% of Districts Partially Contested

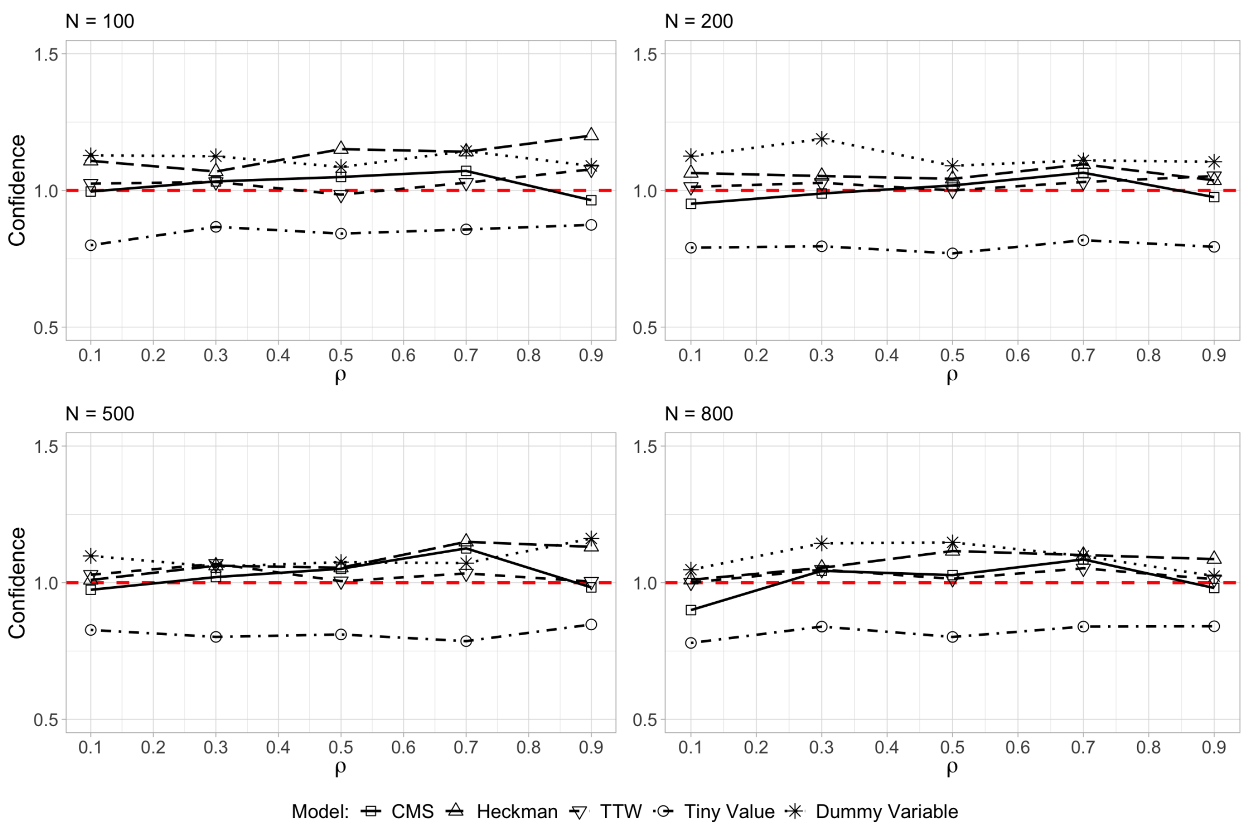


Figure S23: Confidence of $\hat{\beta}_{1A}$—33% of Districts Partially Contested

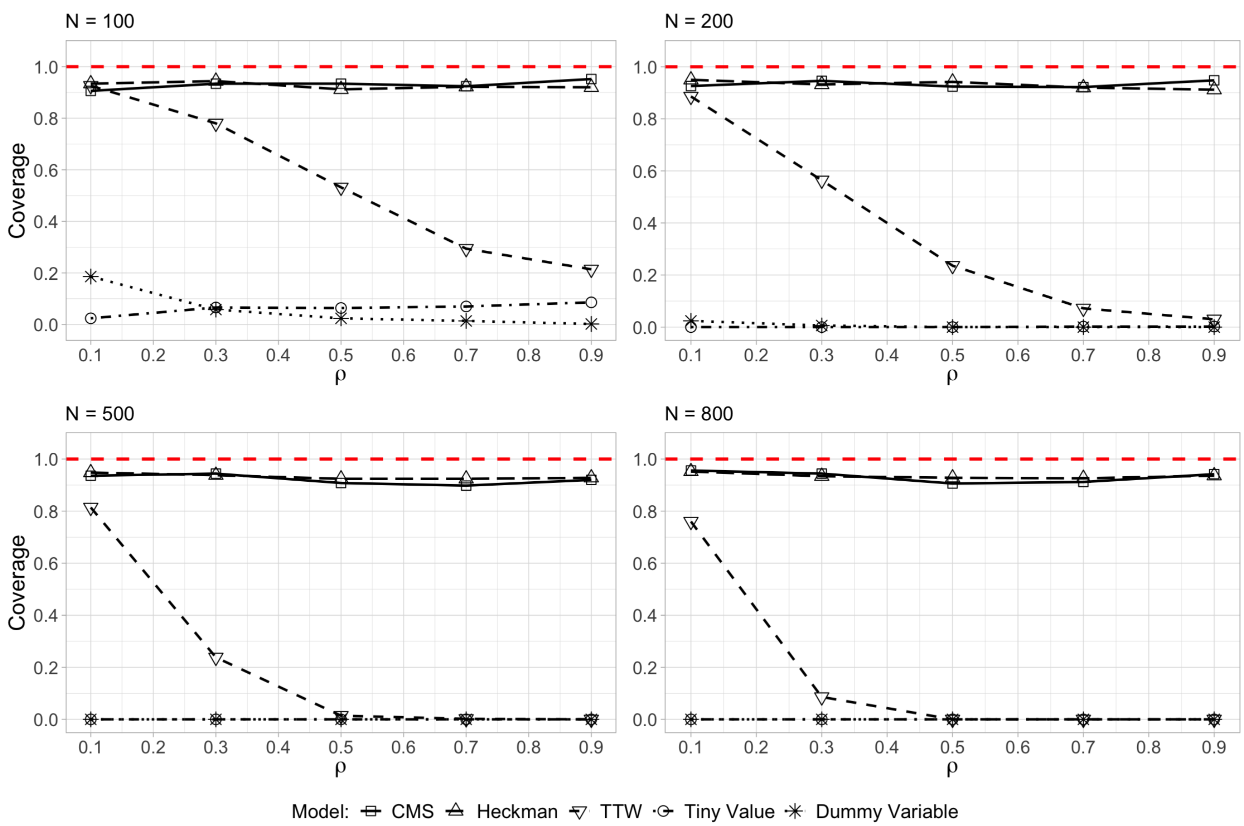


Figure S24: Coverage of $\hat{\beta}_{1A}$—33% of Districts Partially Contested

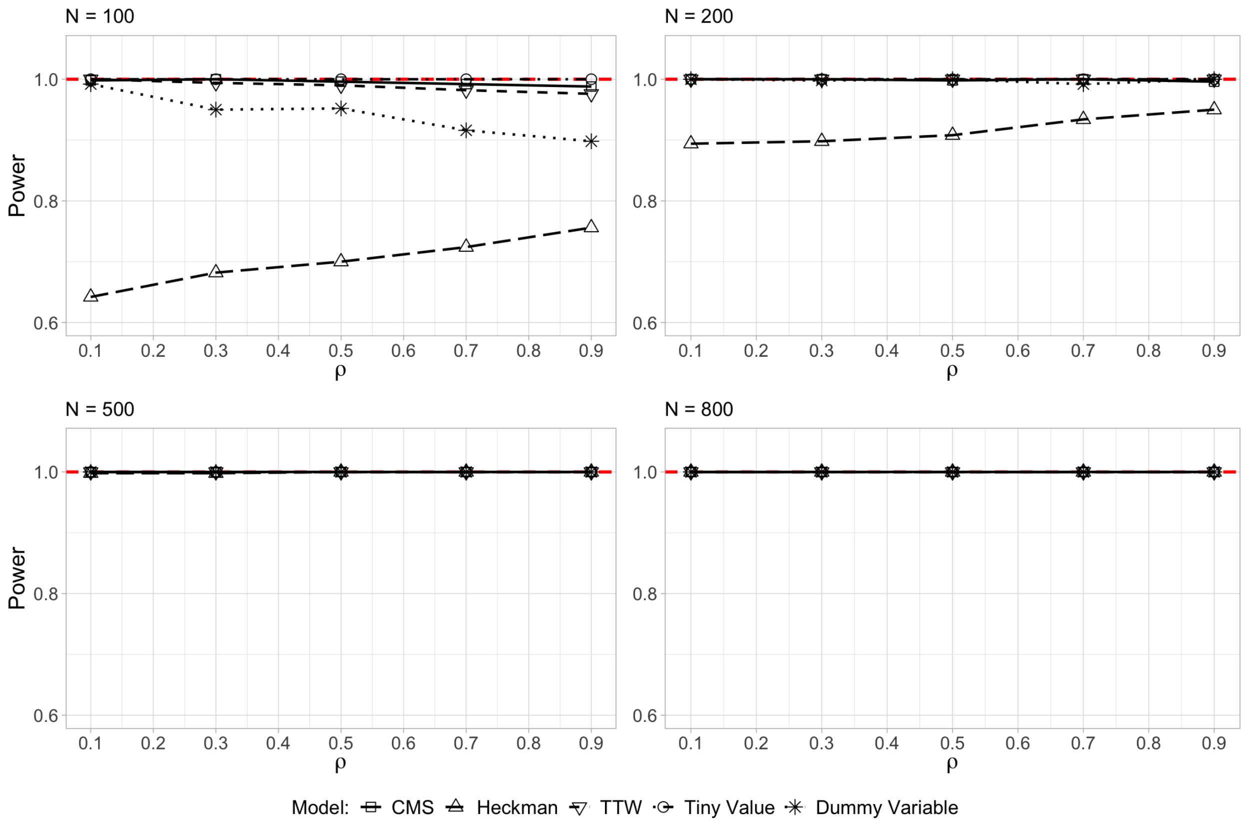


Figure S25: Power of $\hat{\beta}_{1A}$—33% of Districts Partially Contested

D.1.2 50% of Districts are Partially Contested

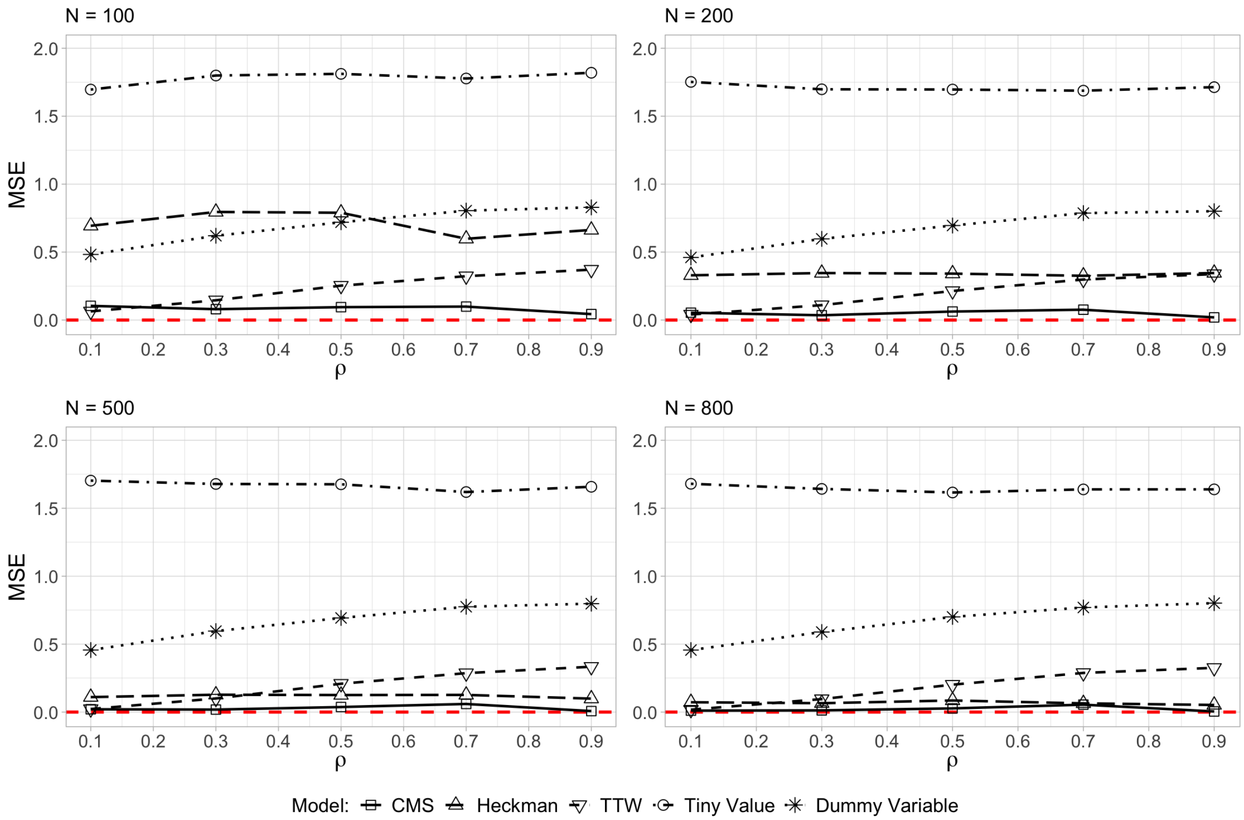


Figure S26: MSE of $\hat{\beta}_{1A}$—50% of Districts Partially Contested

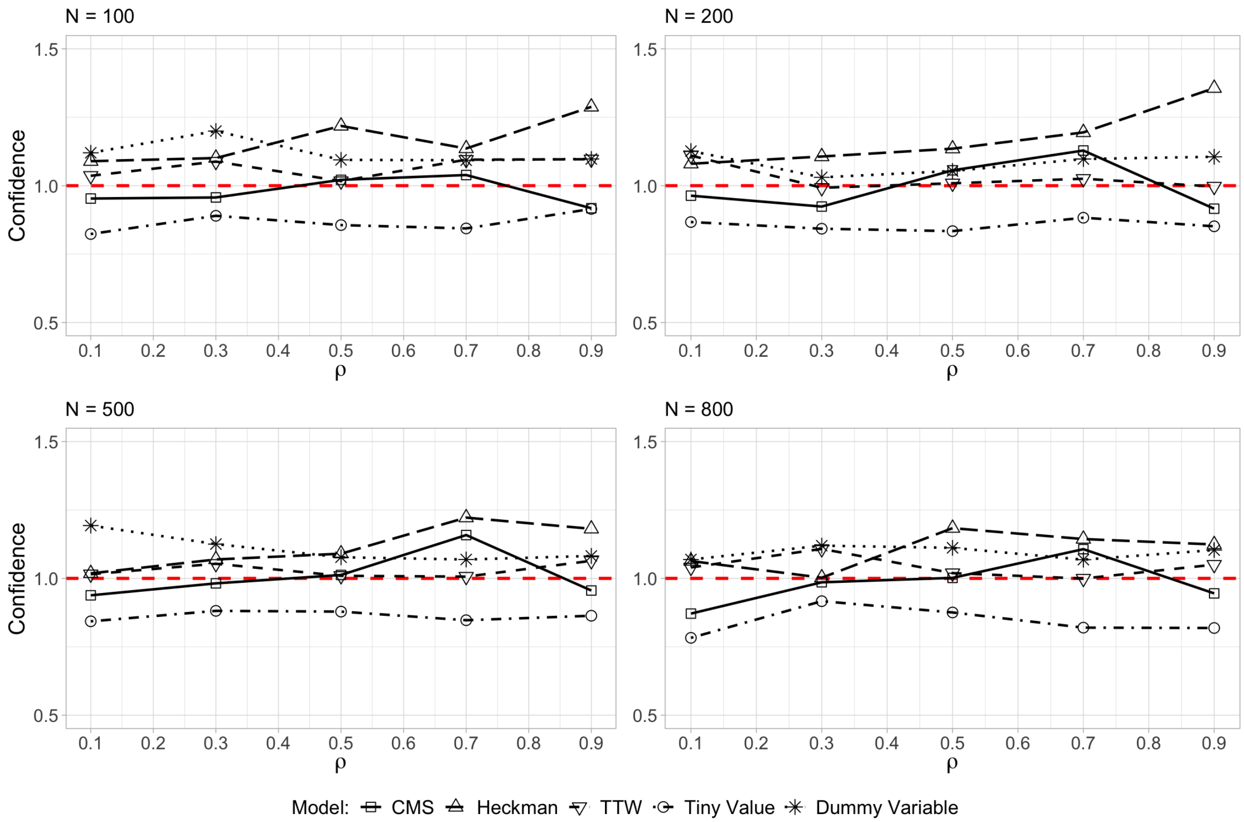


Figure S27: Confidence of $\hat{\beta}_{1A}$—50% of Districts Partially Contested

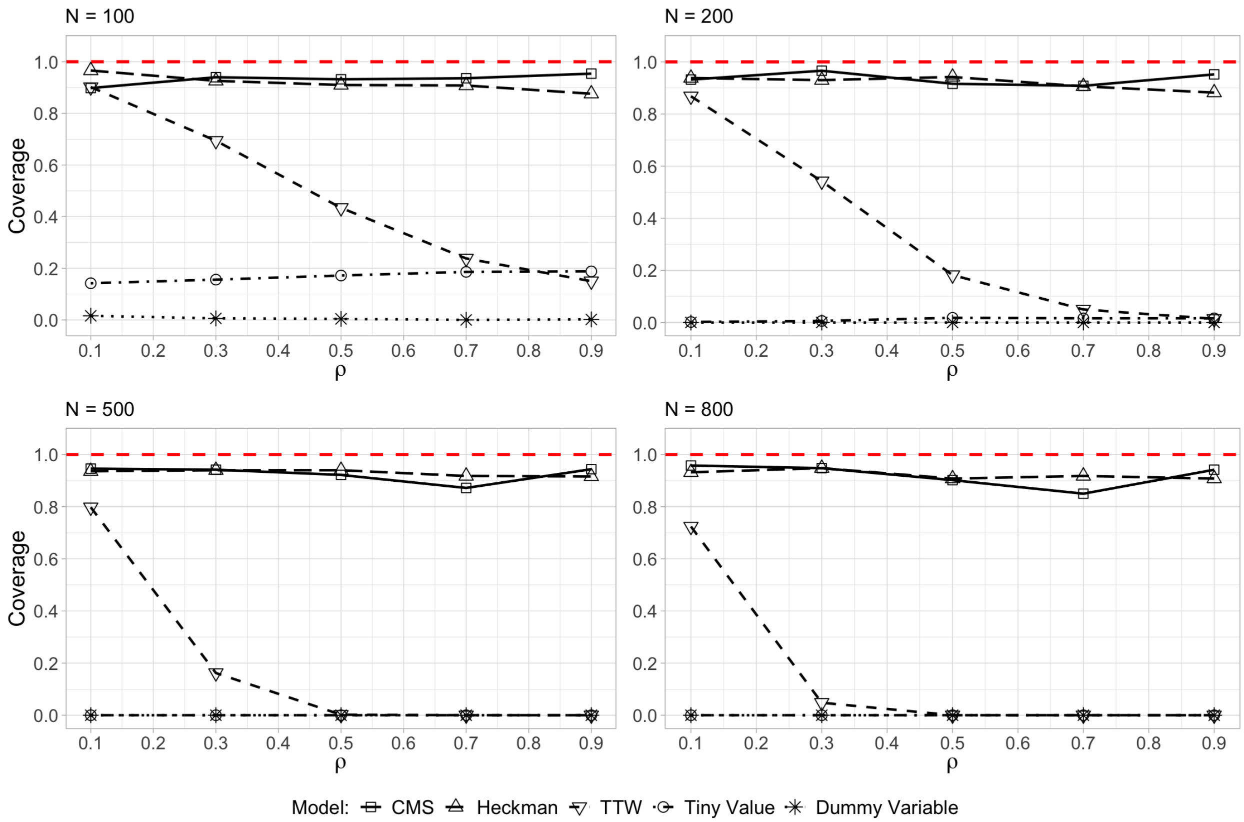


Figure S28: Coverage of $\hat{\beta}_{1A}$—50% of Districts Partially Contested

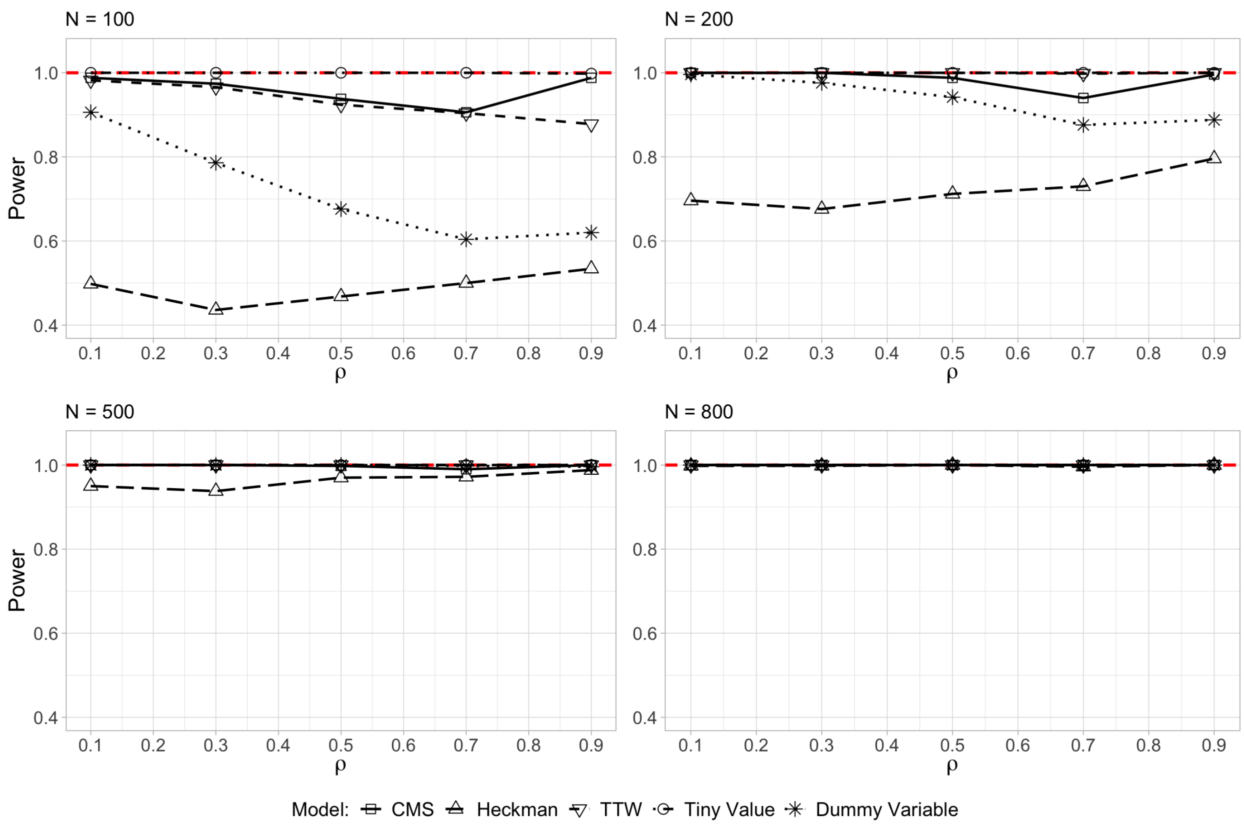


Figure S29: Power of $\hat{\beta}_{1A}$—50% of Districts Partially Contested

D.1.3 66% of Districts are Partially Contested

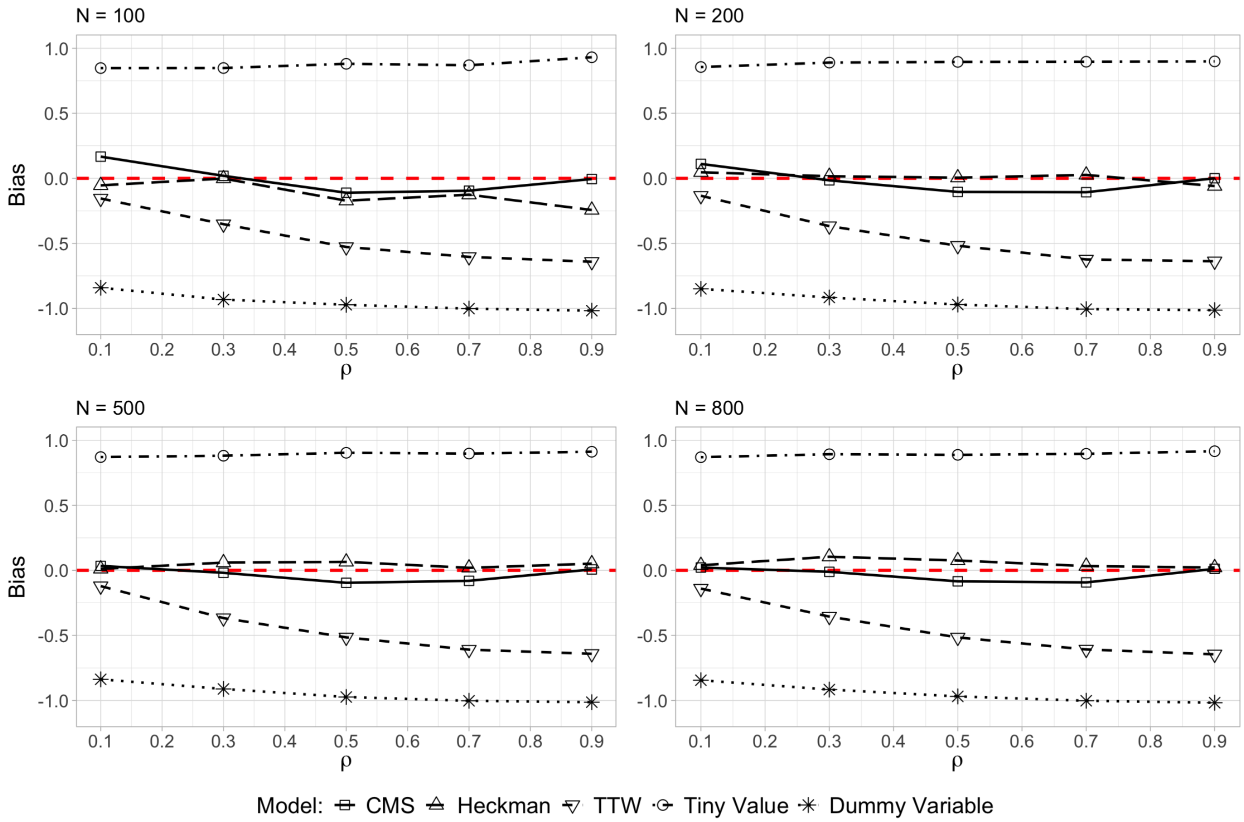


Figure S30: Bias in $\hat{\beta}_{1A}$—66% of Districts Partially Contested

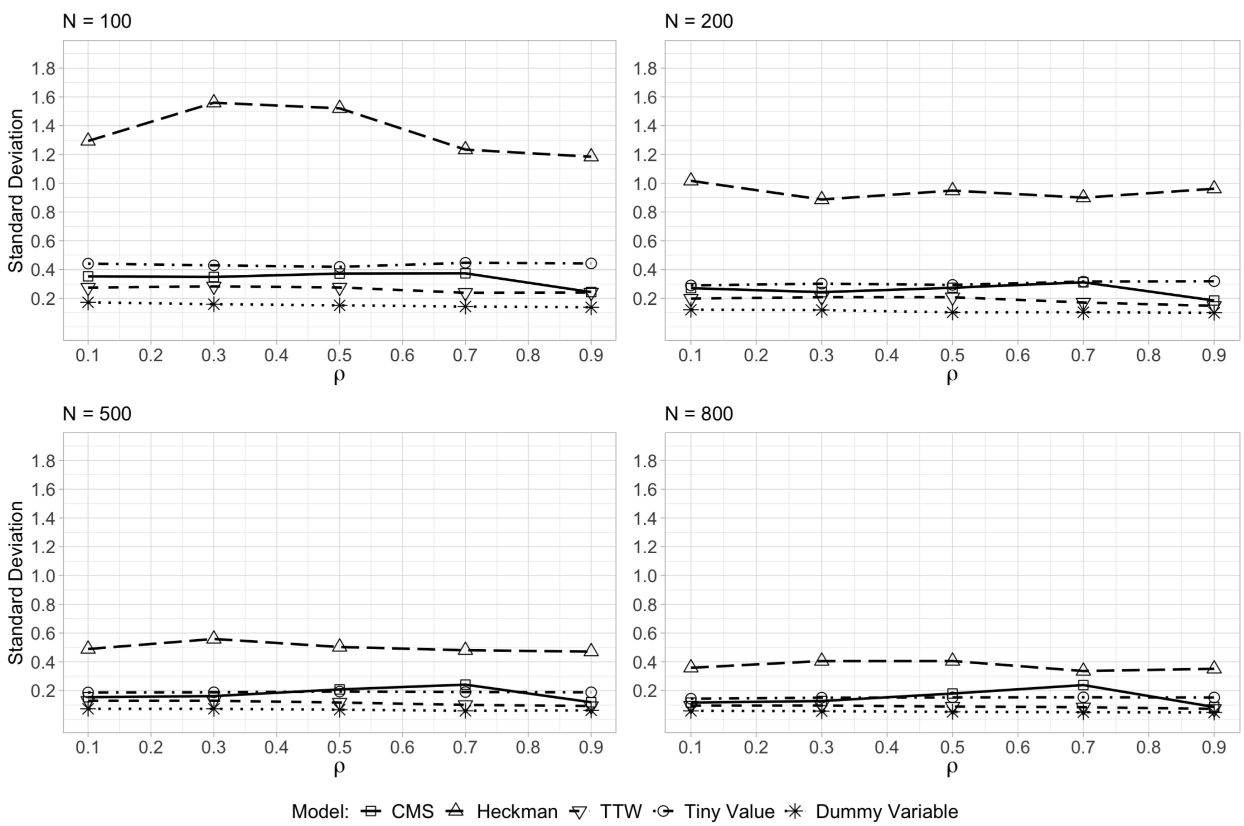


Figure S31: Standard Deviation of $\hat{\beta}_{1A}$—66% of Districts Partially Contested

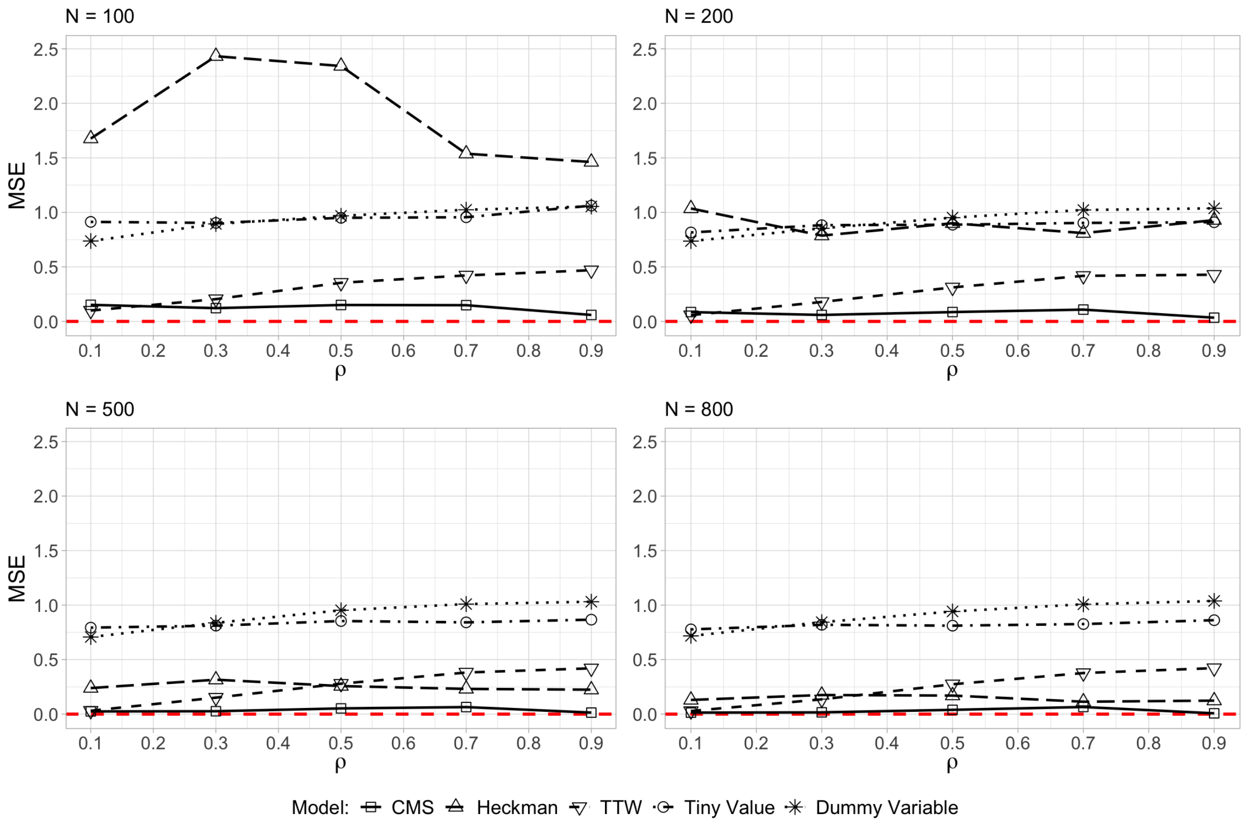


Figure S32: MSE of $\hat{\beta}_{1A}$—66% of Districts Partially Contested

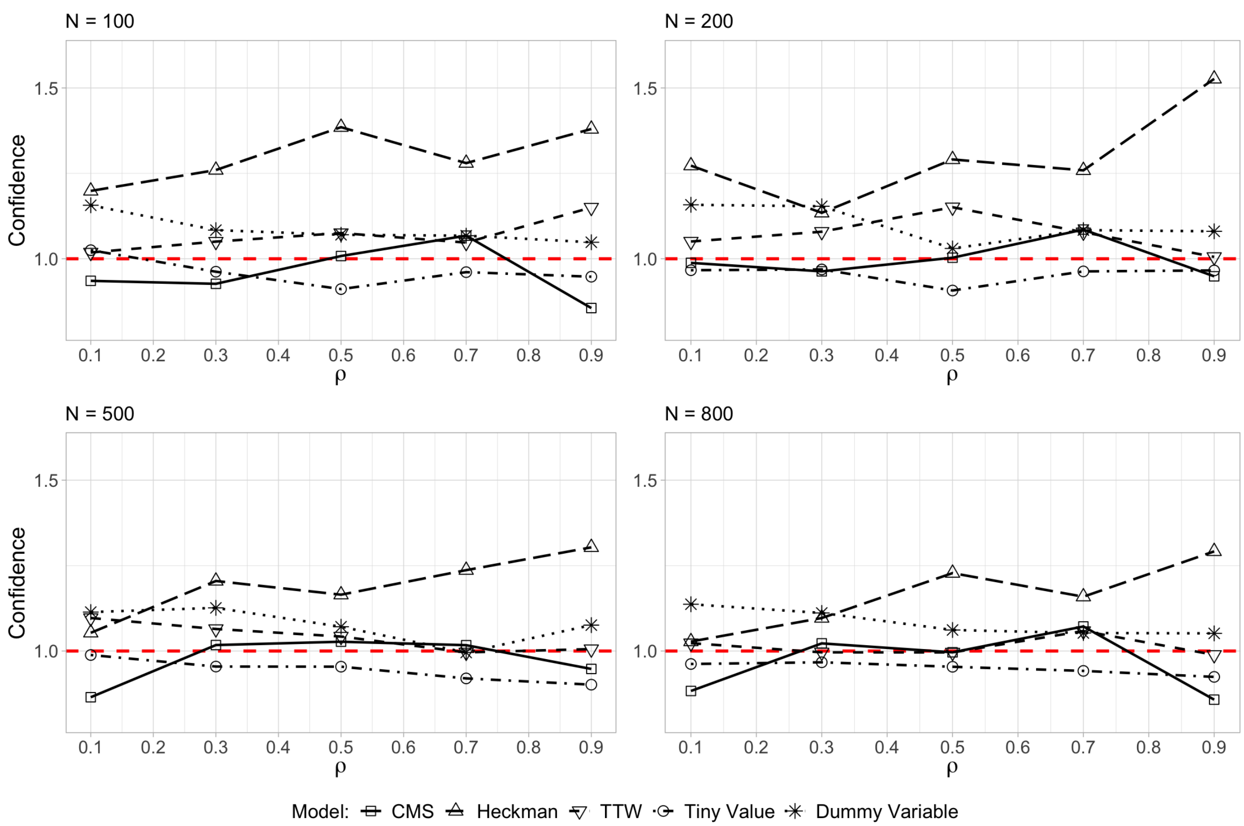


Figure S33: Confidence of $\hat{\beta}_{1A}$—66% of Districts Partially Contested

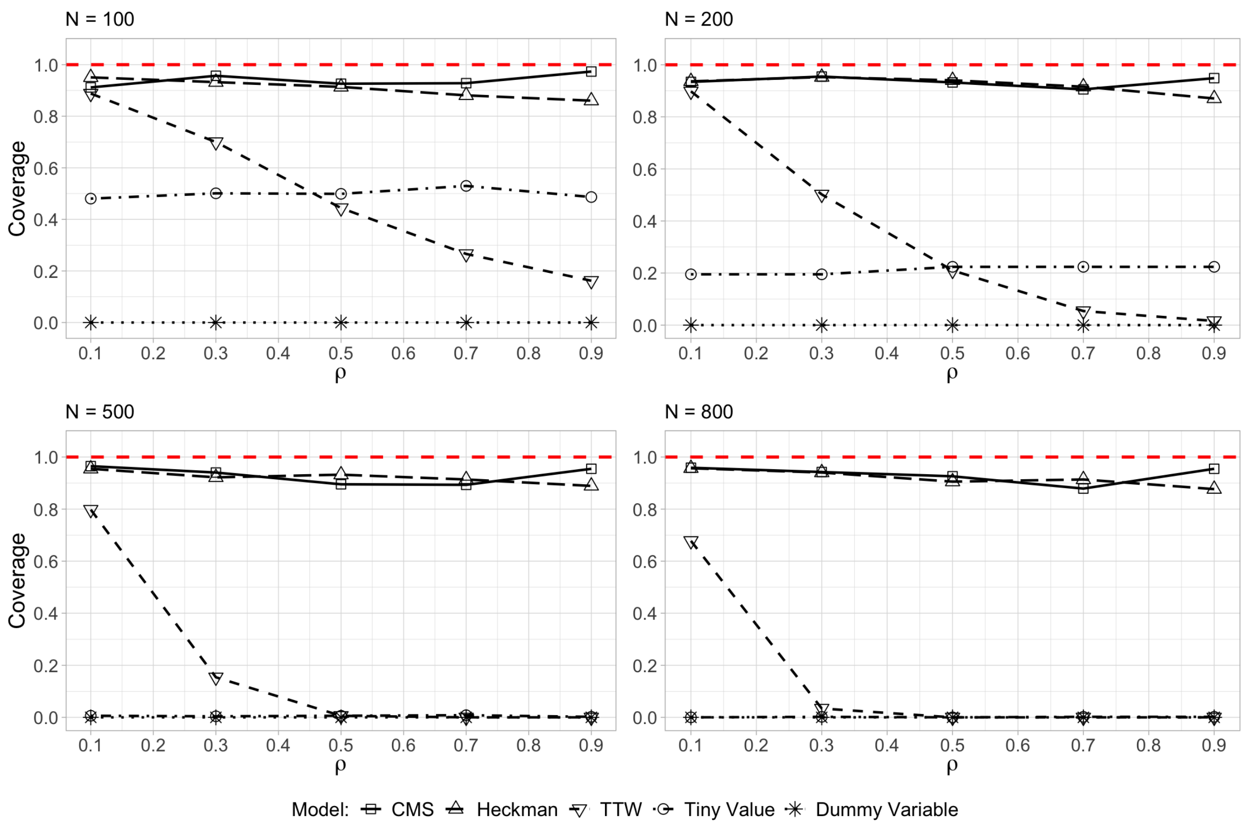


Figure S34: Coverage of $\hat{\beta}_{1A}$—66% of Districts Partially Contested

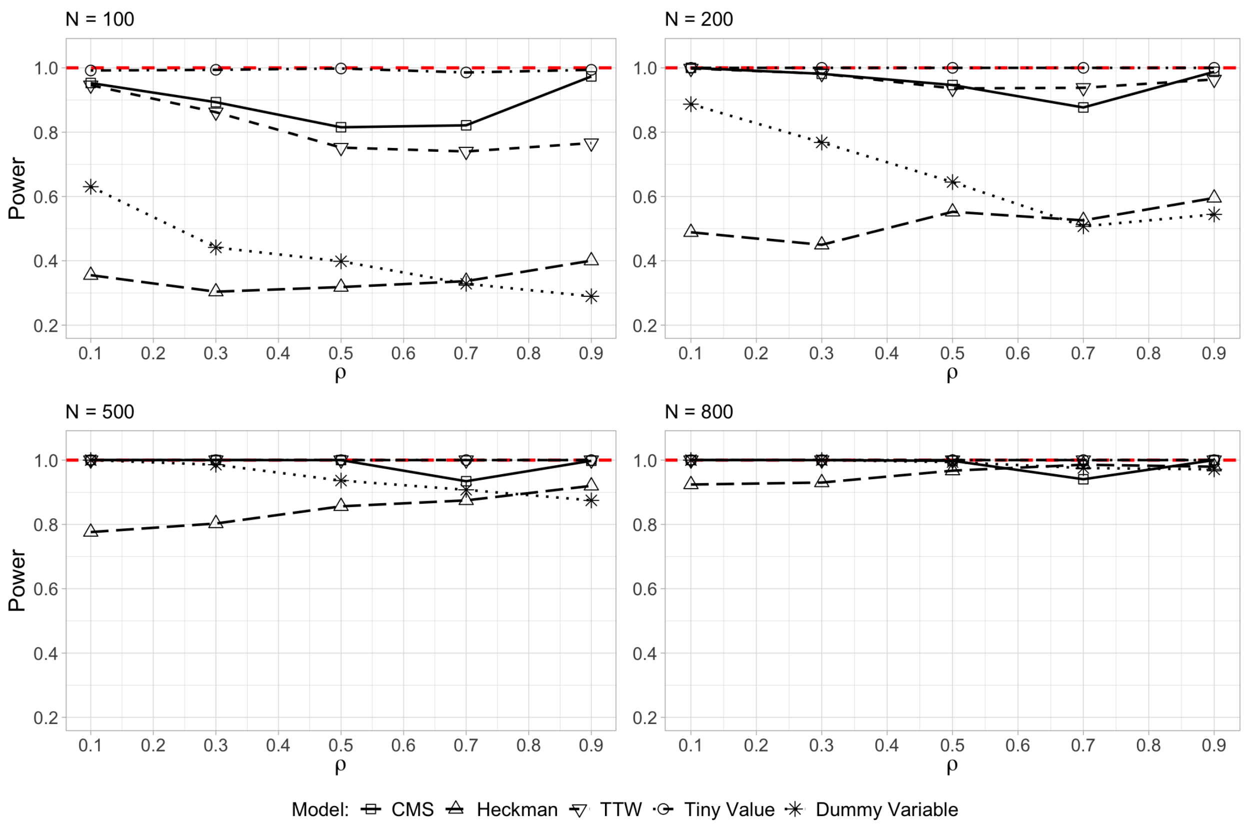


Figure S35: Power of $\hat{\beta}_{1A}$—66% of Districts Partially Contested

D.2 Invalid Exclusion Restriction

This section compares the approaches when there is no valid instrument in the selection stage to estimate the inverse Mills ratio (i.e., the exclusion restriction is violated). Under these conditions, Heckman approach performs poorer than models that do not account for sample selection in basically all statistical performances, whereas our approach (CMS) consistently performs well.

D.2.1 33% of Districts are Partially Contested

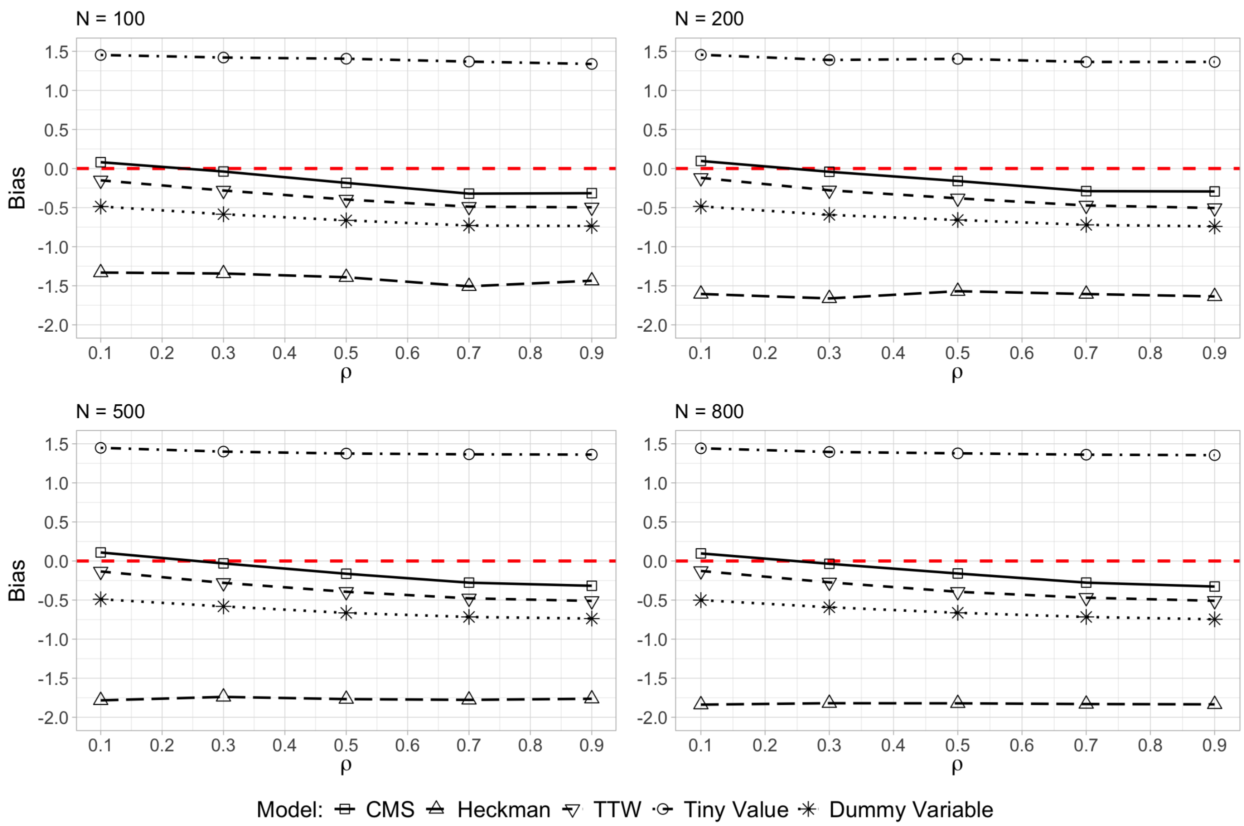


Figure S36: Bias in $\hat{\beta}_{1A}$—30% of Districts Partially Contested

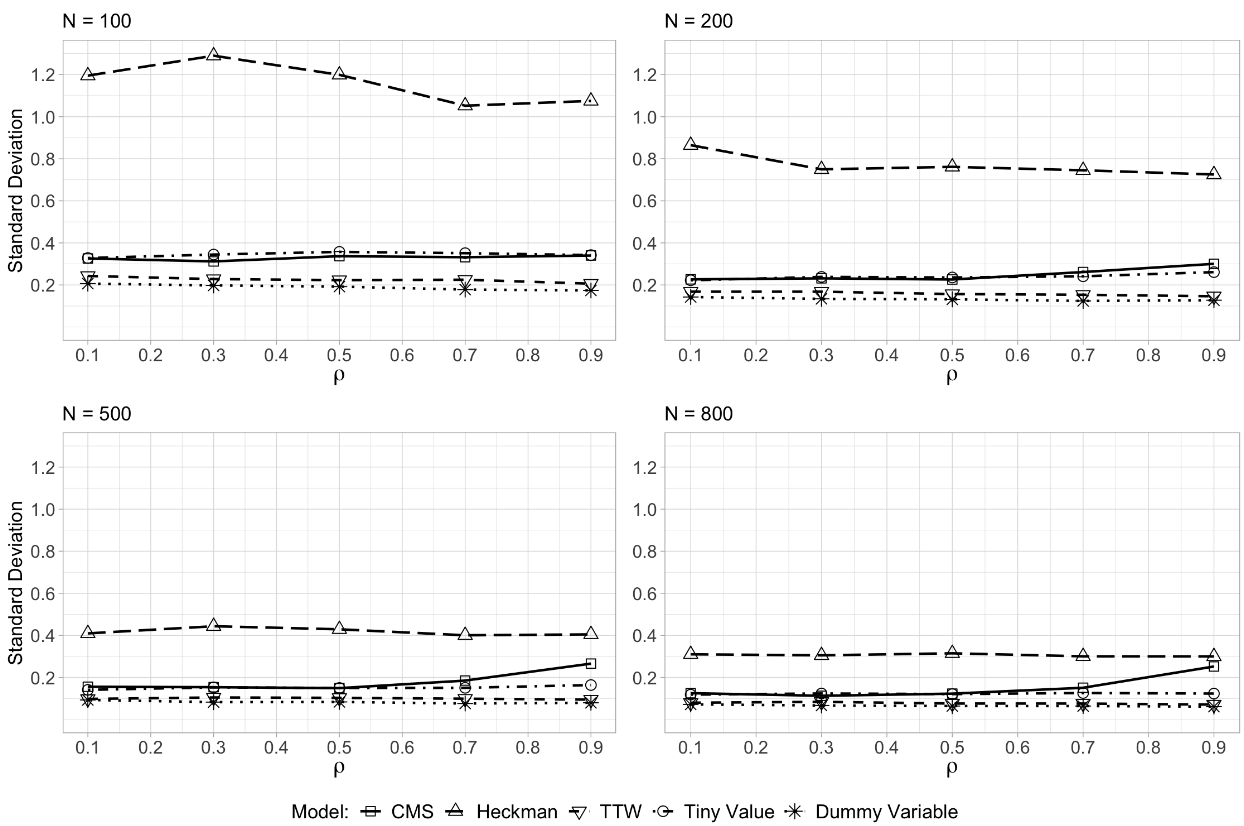


Figure S37: Standard Deviation of $\hat{\beta}_{1A}$—30% of Districts Partially Contested

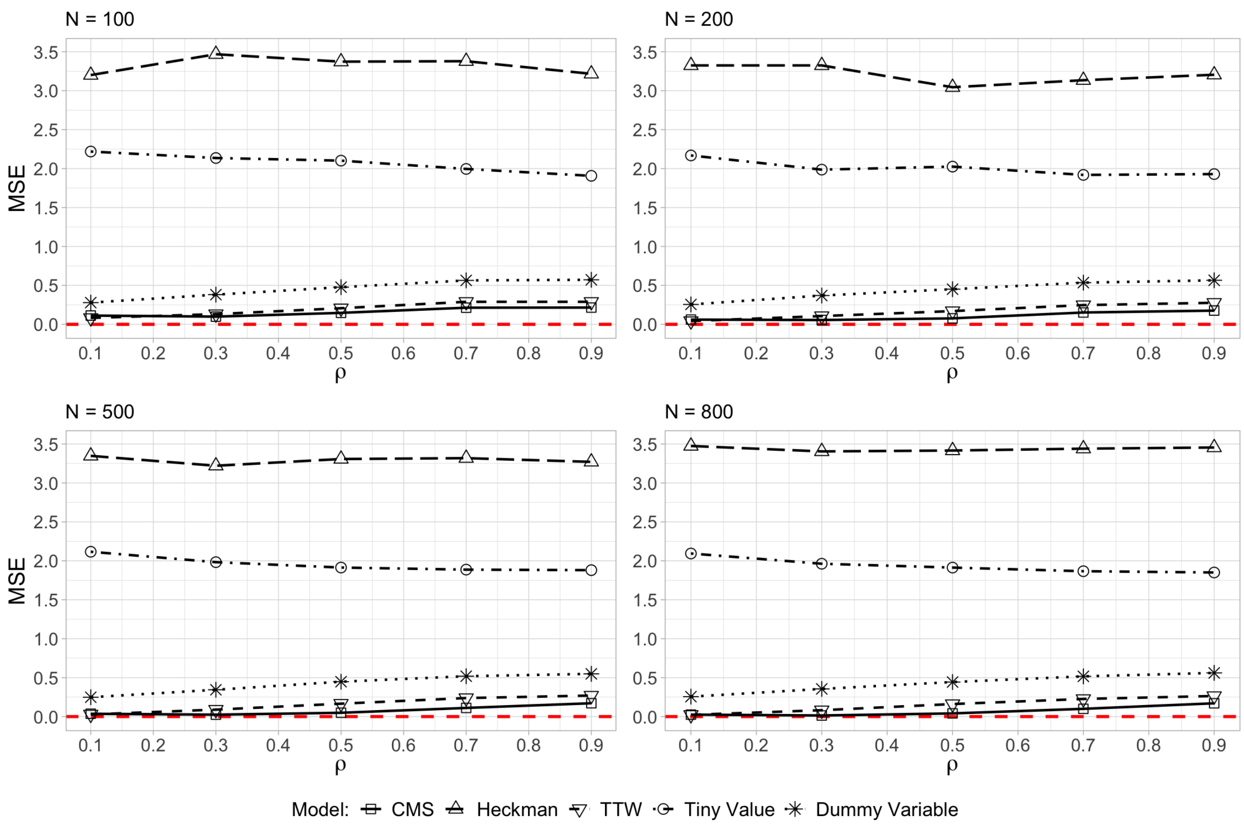


Figure S38: MSE of $\hat{\beta}_{1A}$—30% of Districts Partially Contested

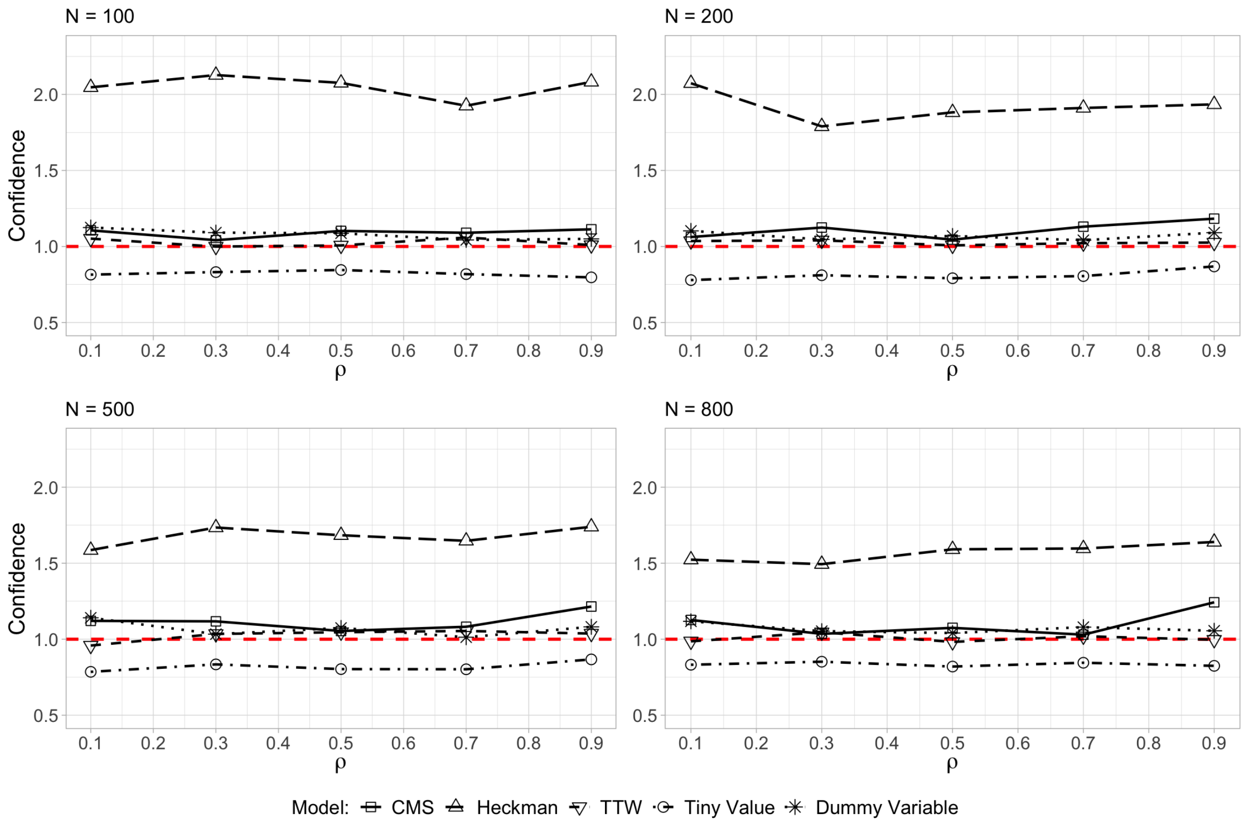


Figure S39: Confidence of $\hat{\beta}_{1A}$—30% of Districts Partially Contested

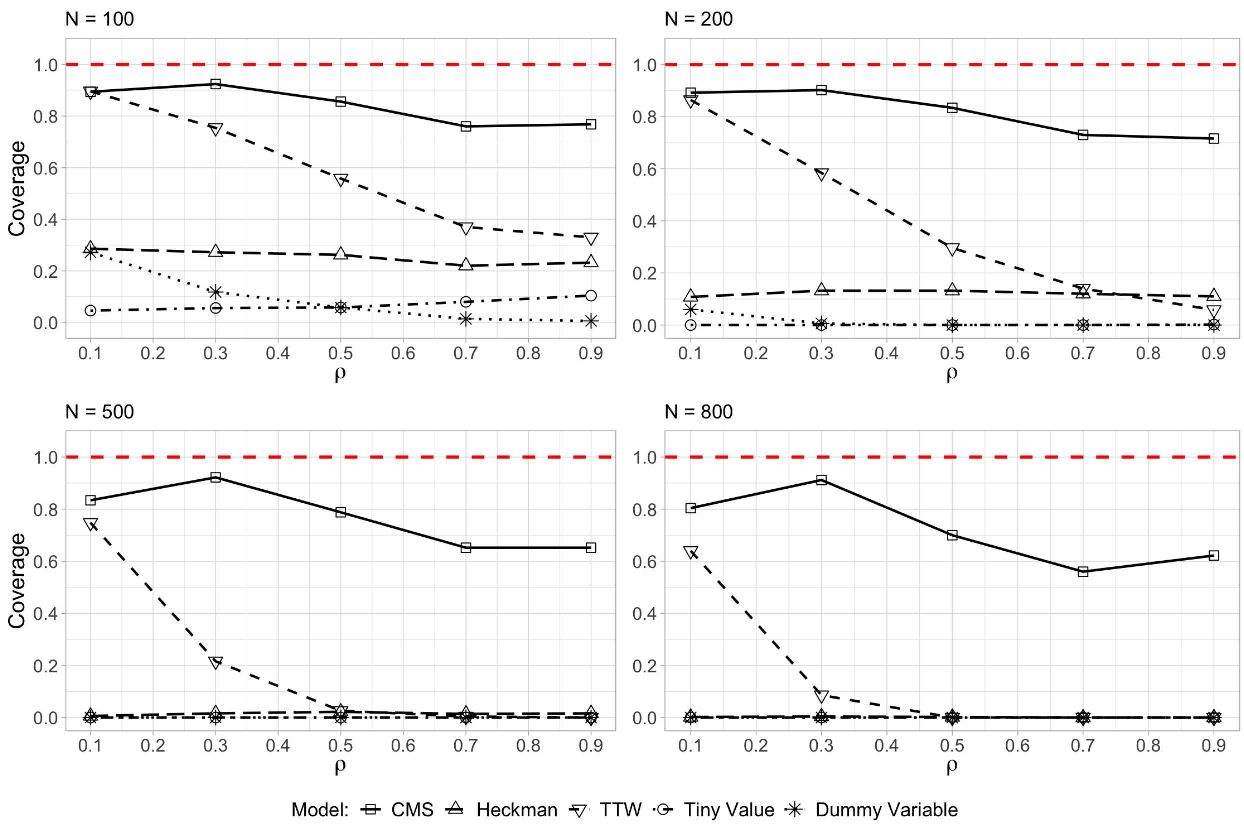


Figure S40: Coverage of $\hat{\beta}_{1A}$—30% of Districts Partially Contested

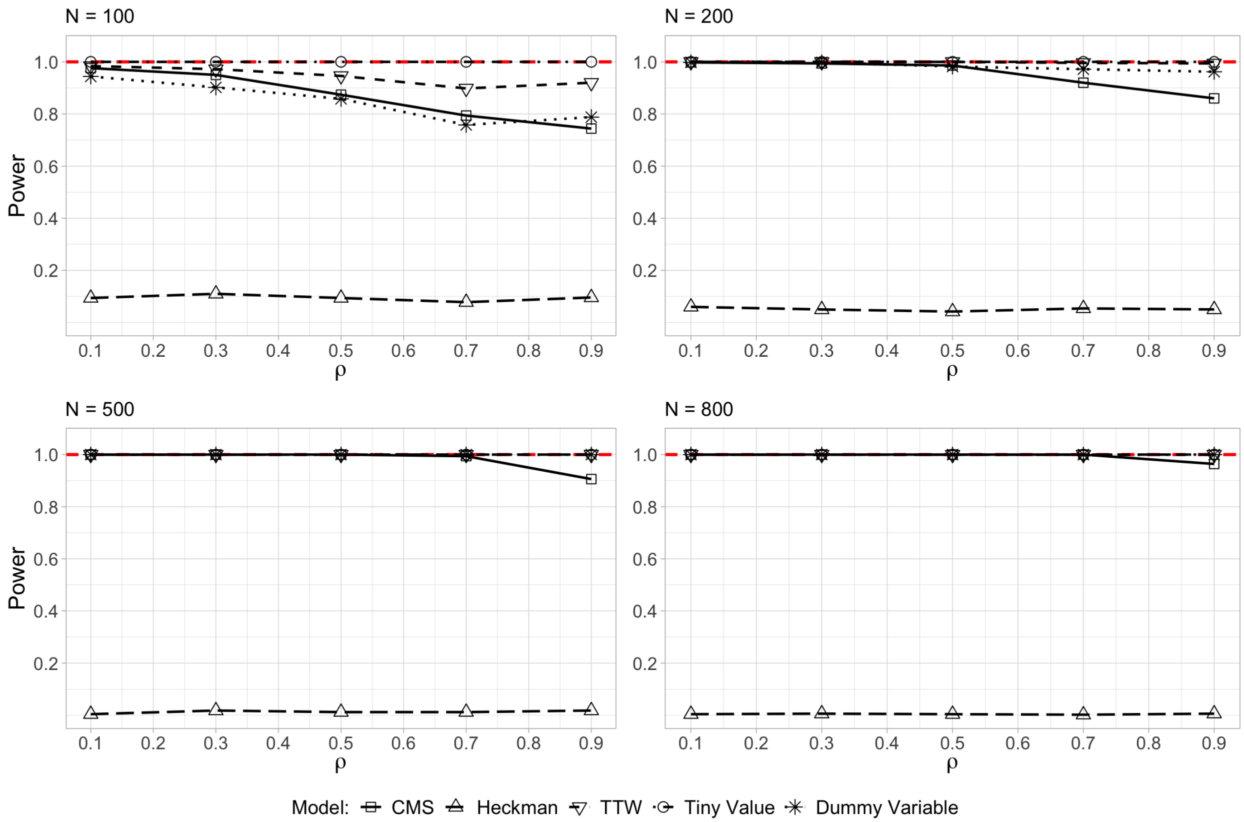


Figure S41: Power of $\hat{\beta}_{1A}$—30% of Districts Partially Contested

D.2.2 50% of Districts are Partially Contested

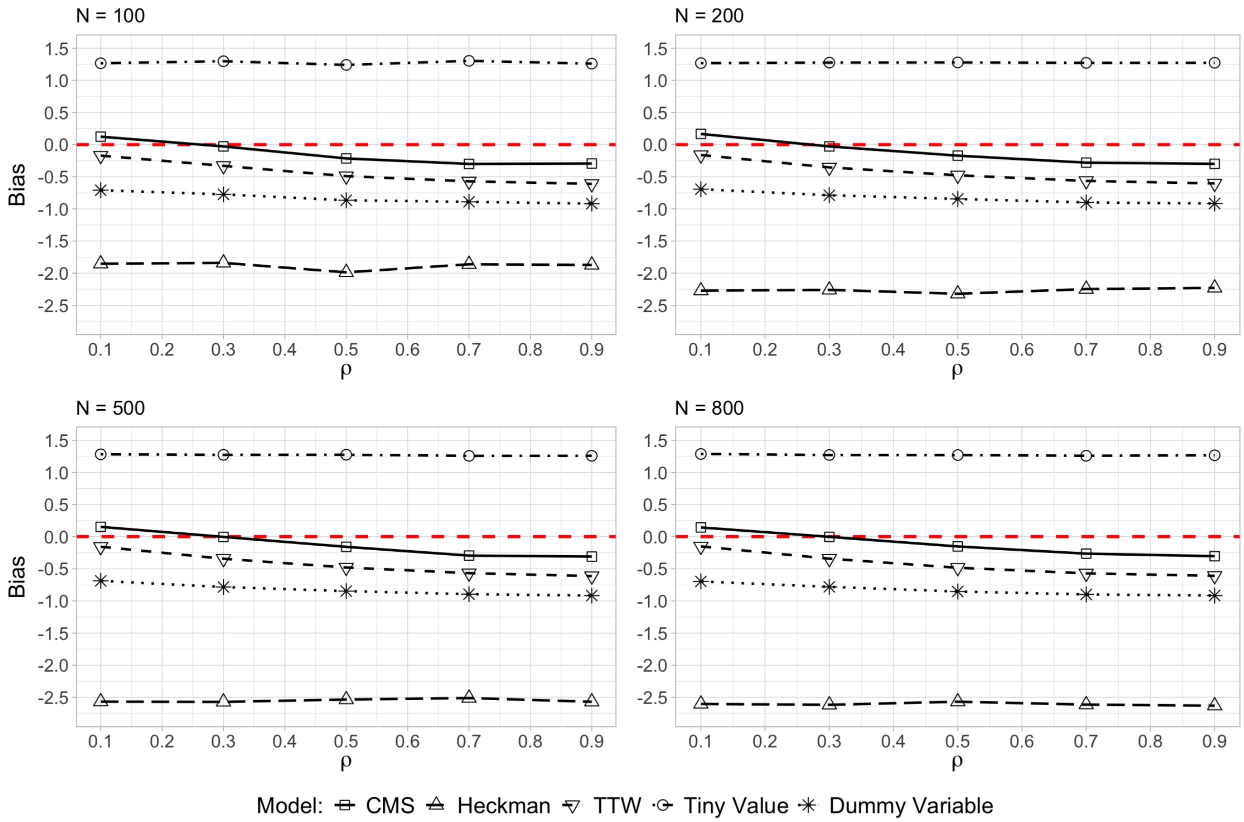


Figure S42: Bias in $\hat{\beta}_{1A}$—50% of Districts Partially Contested

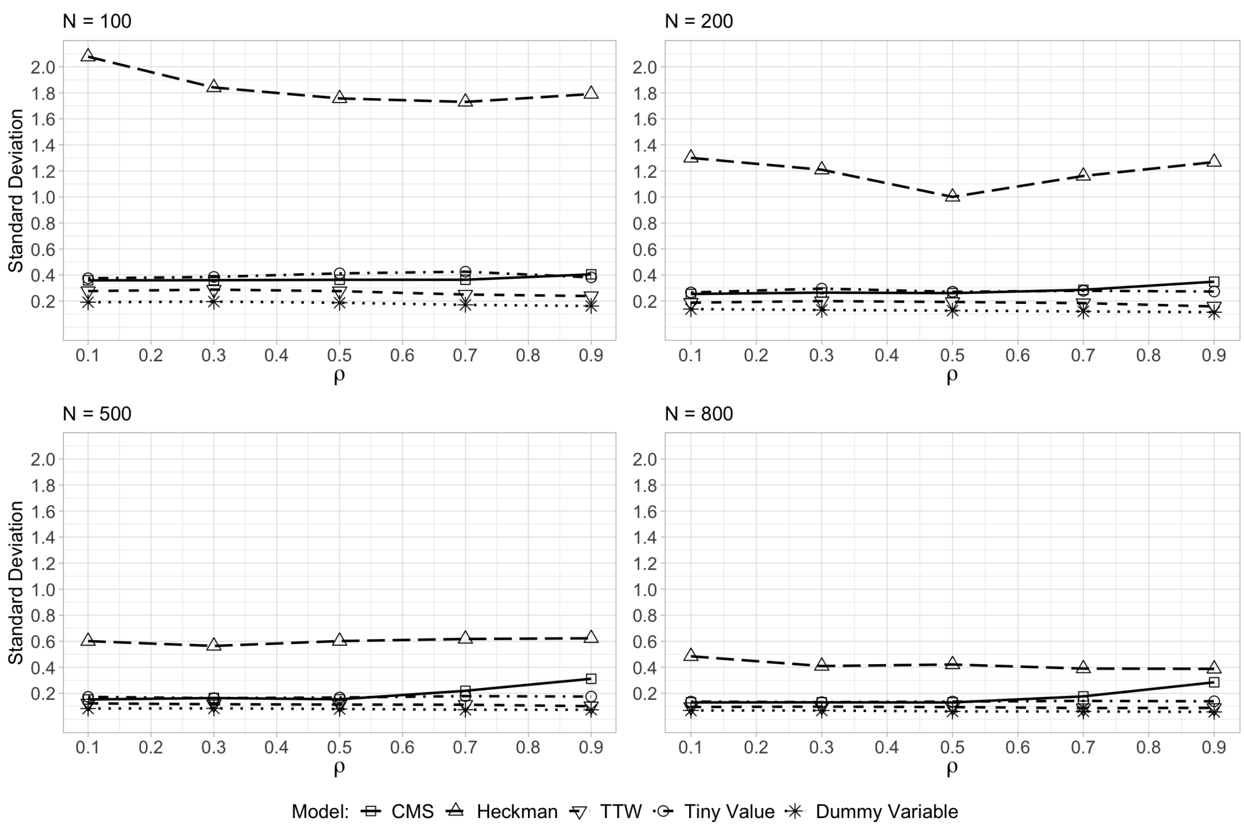


Figure S43: Standard Deviation of $\hat{\beta}_{1A}$—50% of Districts Partially Contested

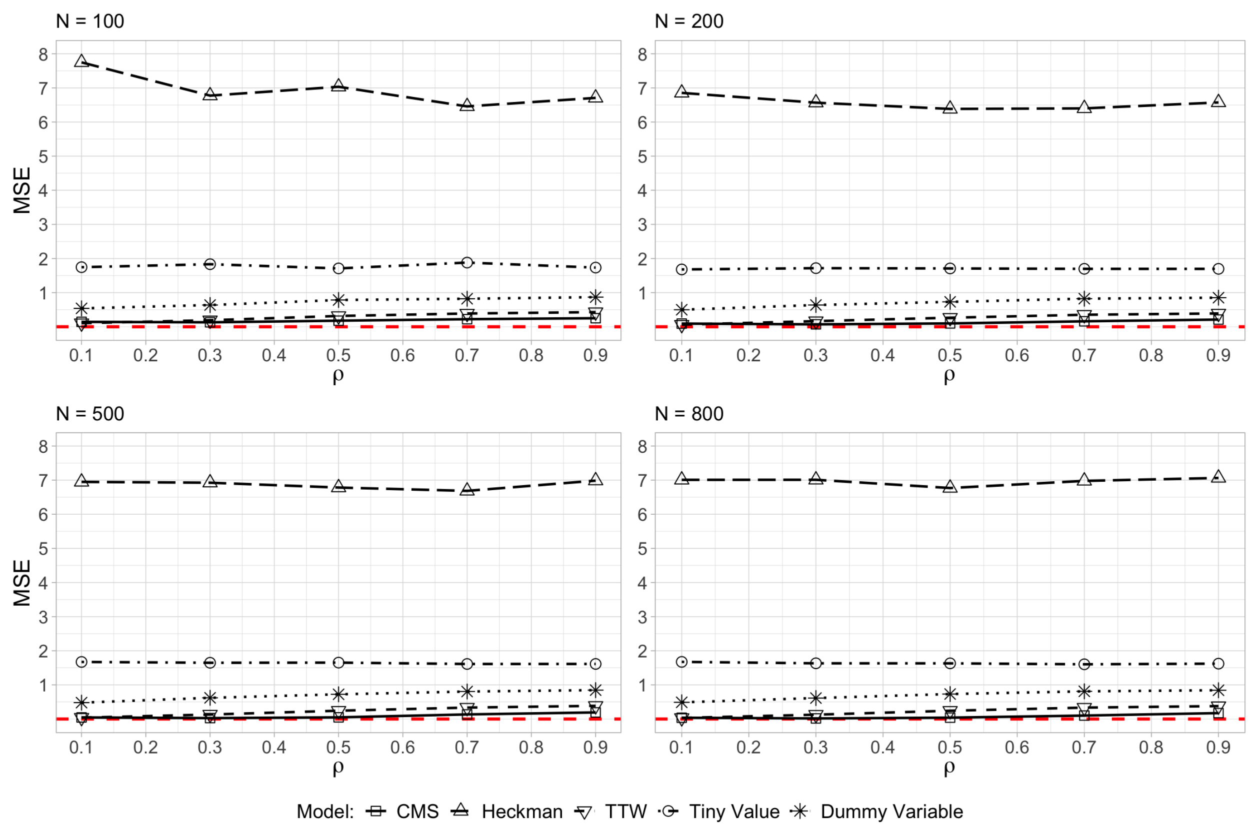


Figure S44: MSE of $\hat{\beta}_{1A}$—50% of Districts Partially Contested

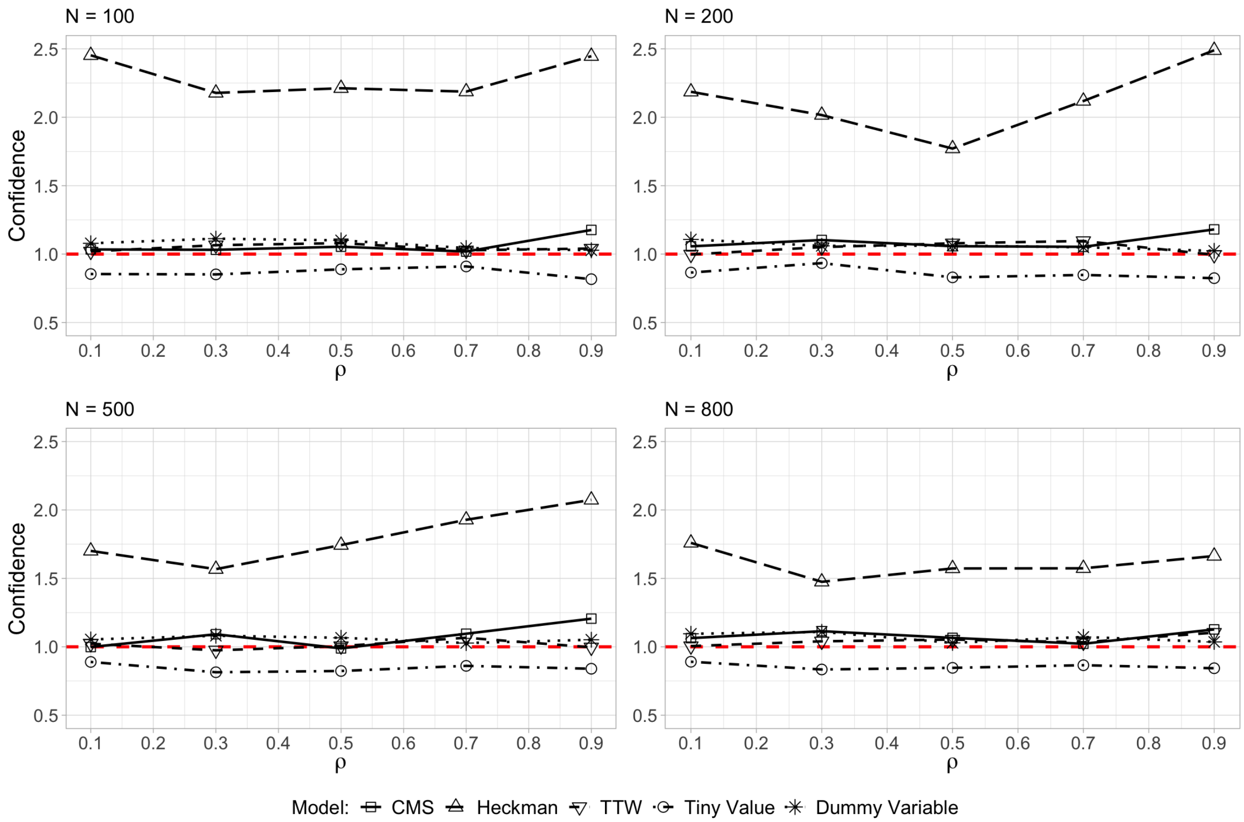


Figure S45: Confidence of $\hat{\beta}_{1A}$—50% of Districts Partially Contested

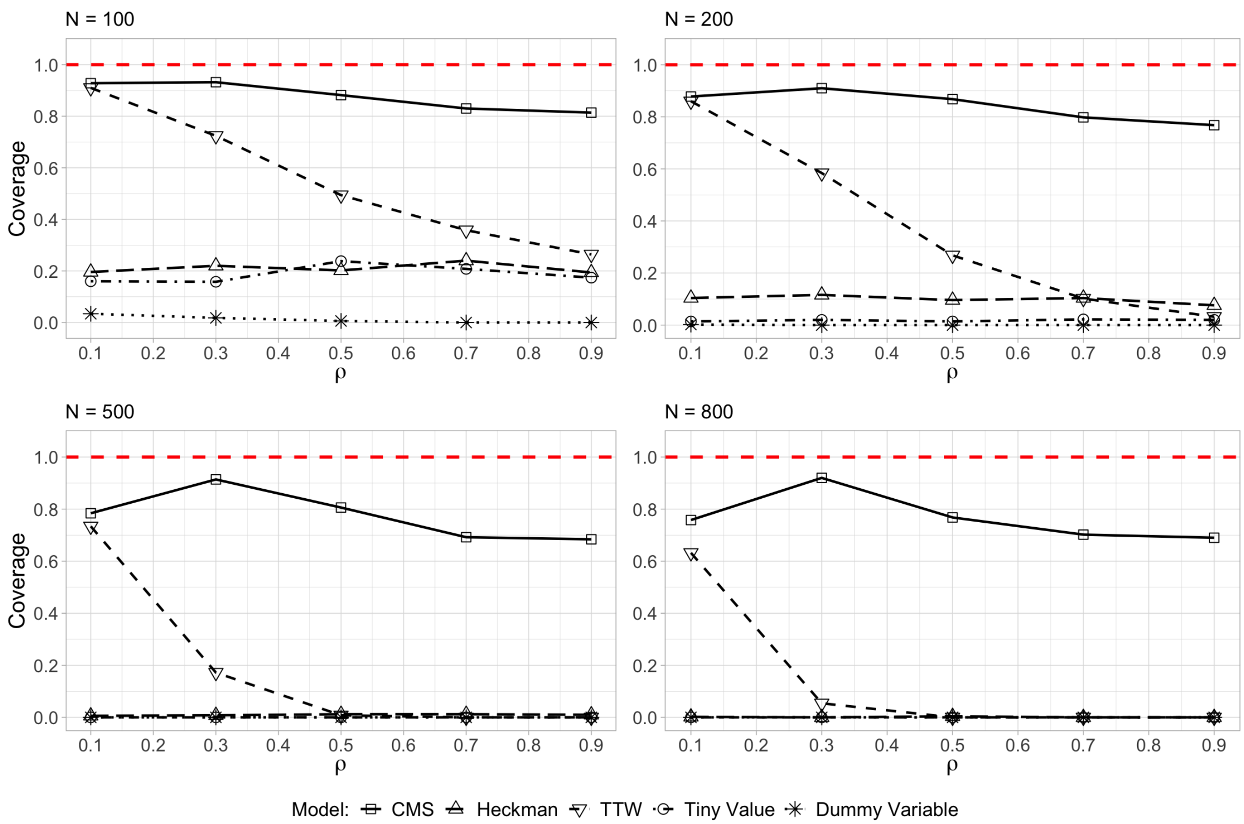


Figure S46: Coverage of $\hat{\beta}_{1A}$—50% of Districts Partially Contested

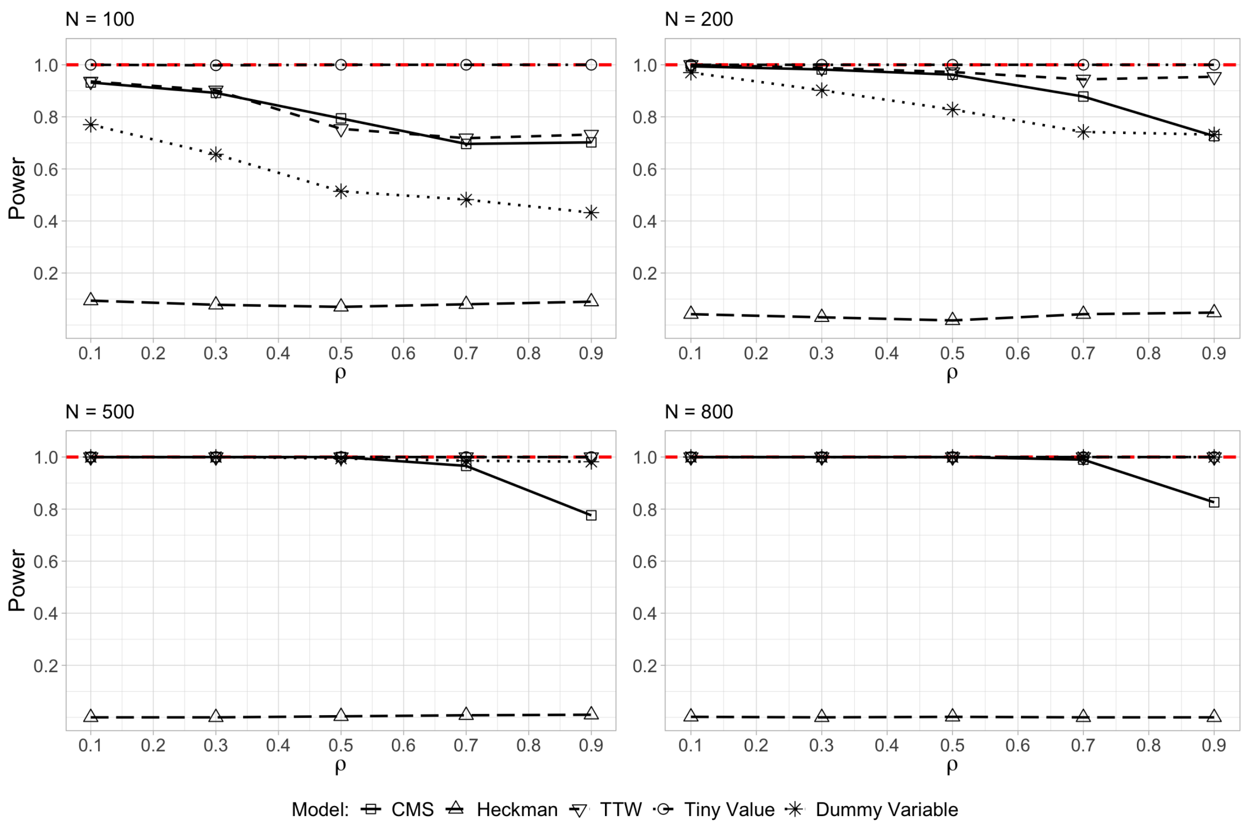


Figure S47: Power of $\hat{\beta}_{1A}$—50% of Districts Partially Contested

D.2.3 66% of Districts are Partially Contested

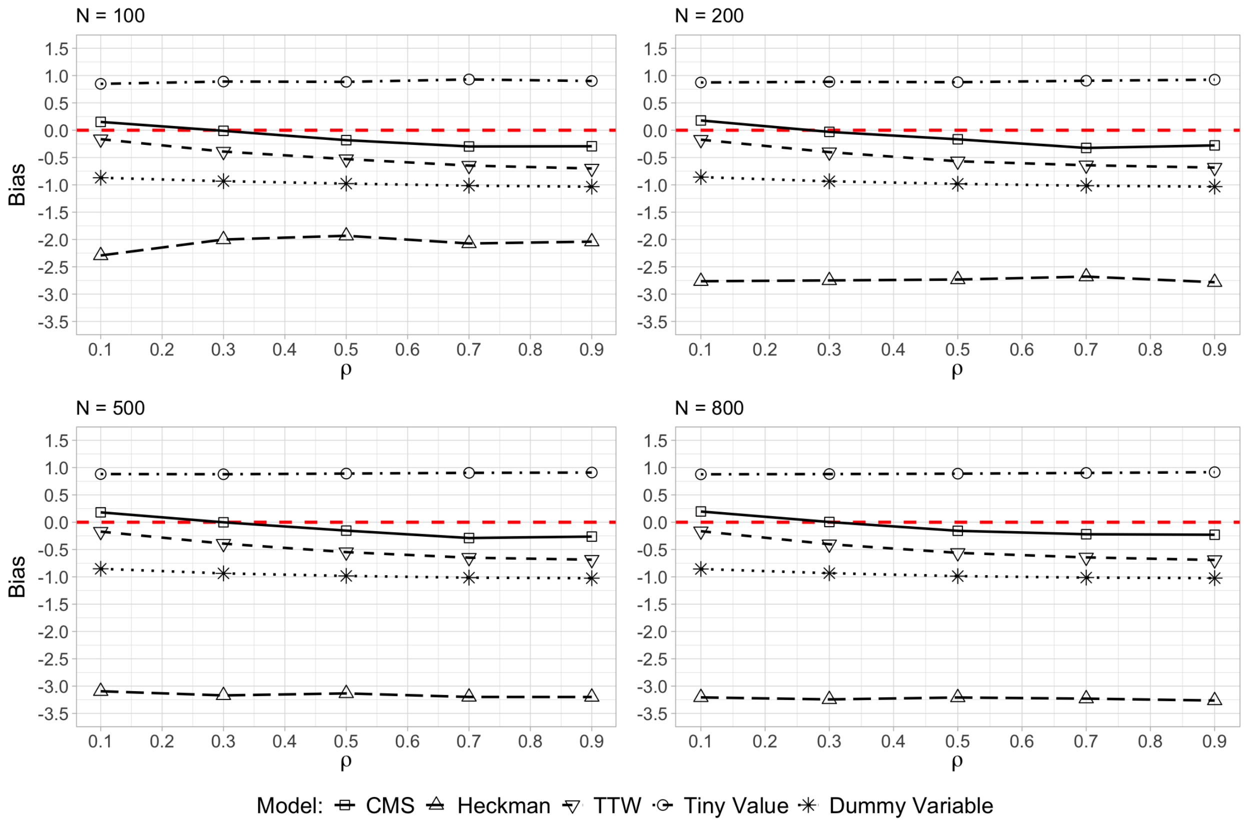


Figure S48: Bias in $\hat{\beta}_{1A}$—66% of Districts Partially Contested

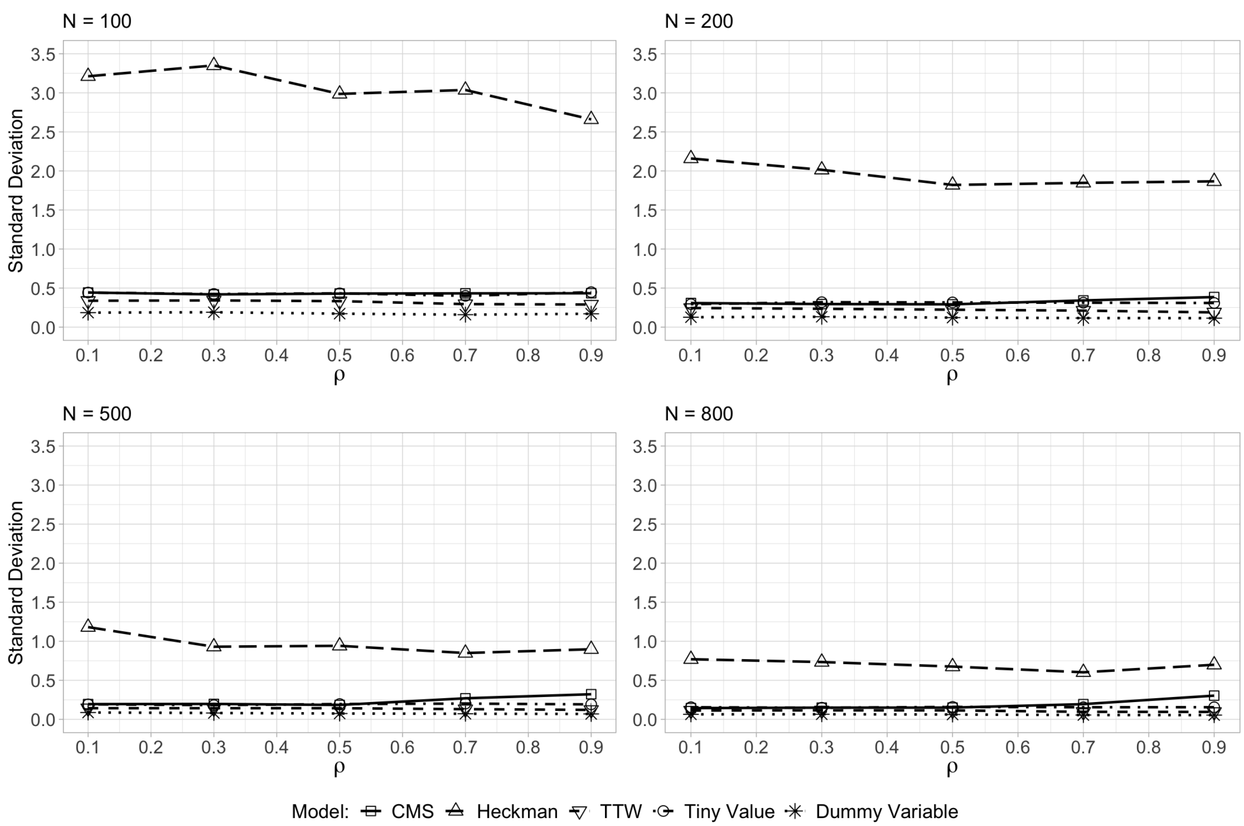


Figure S49: Standard Deviation of $\hat{\beta}_{1A}$—66% of Districts Partially Contested

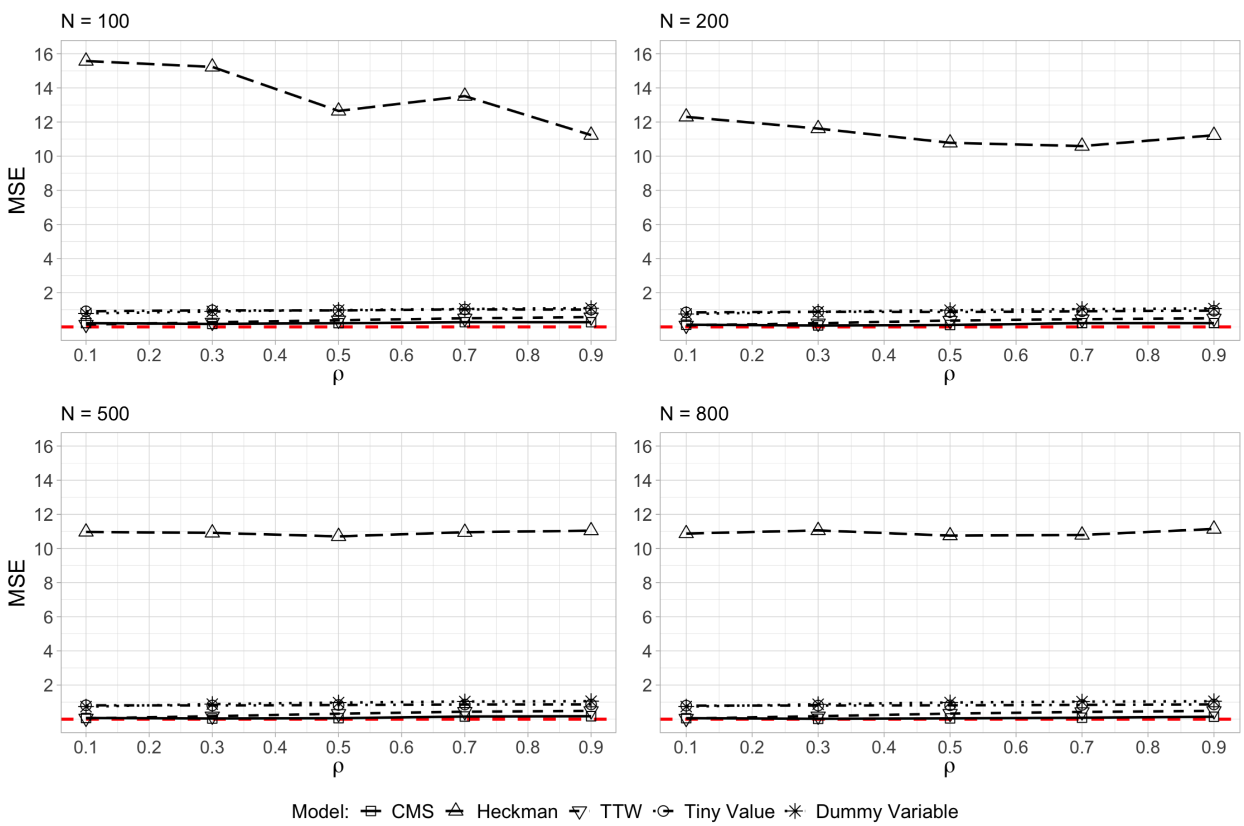


Figure S50: MSE of $\hat{\beta}_{1A}$—66% of Districts Partially Contested

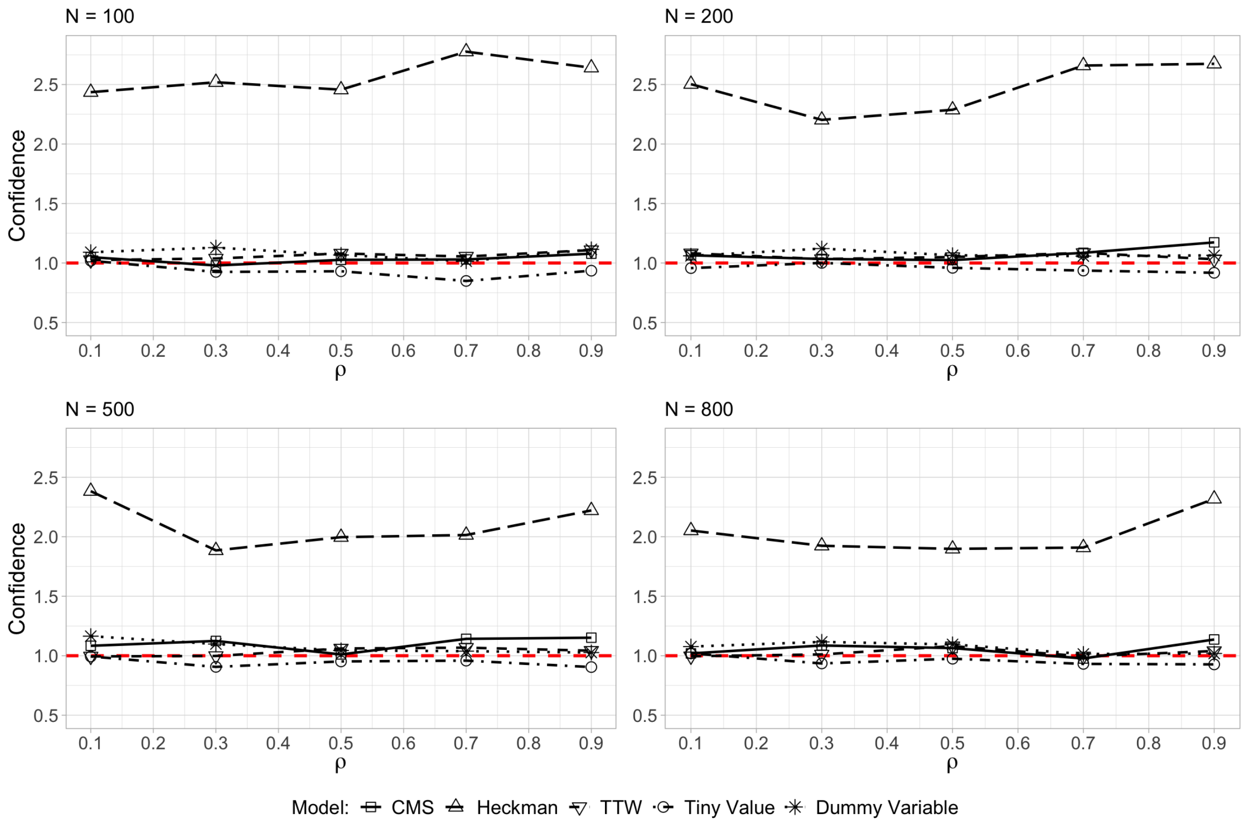


Figure S51: Confidence of $\hat{\beta}_{1A}$—66% of Districts Partially Contested

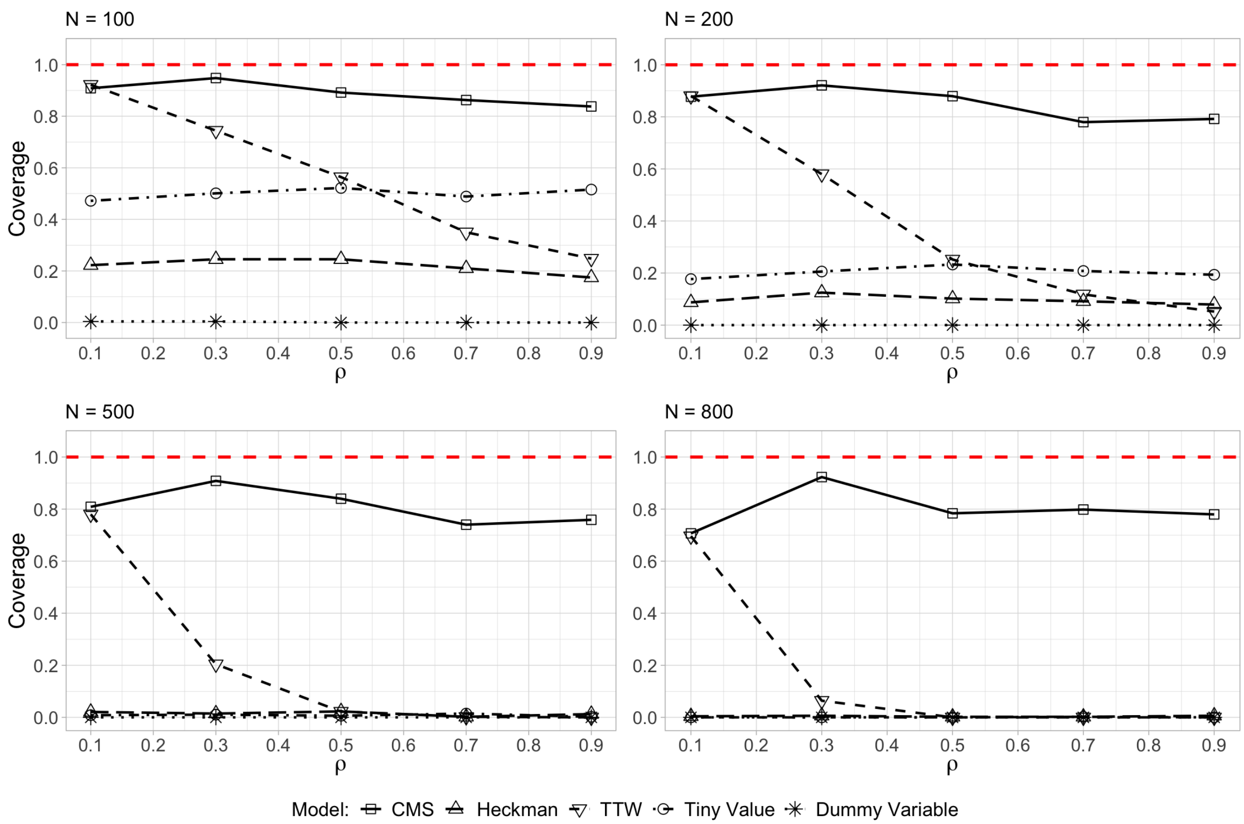


Figure S52: Coverage of $\hat{\beta}_{1A}$—66% of Districts Partially Contested

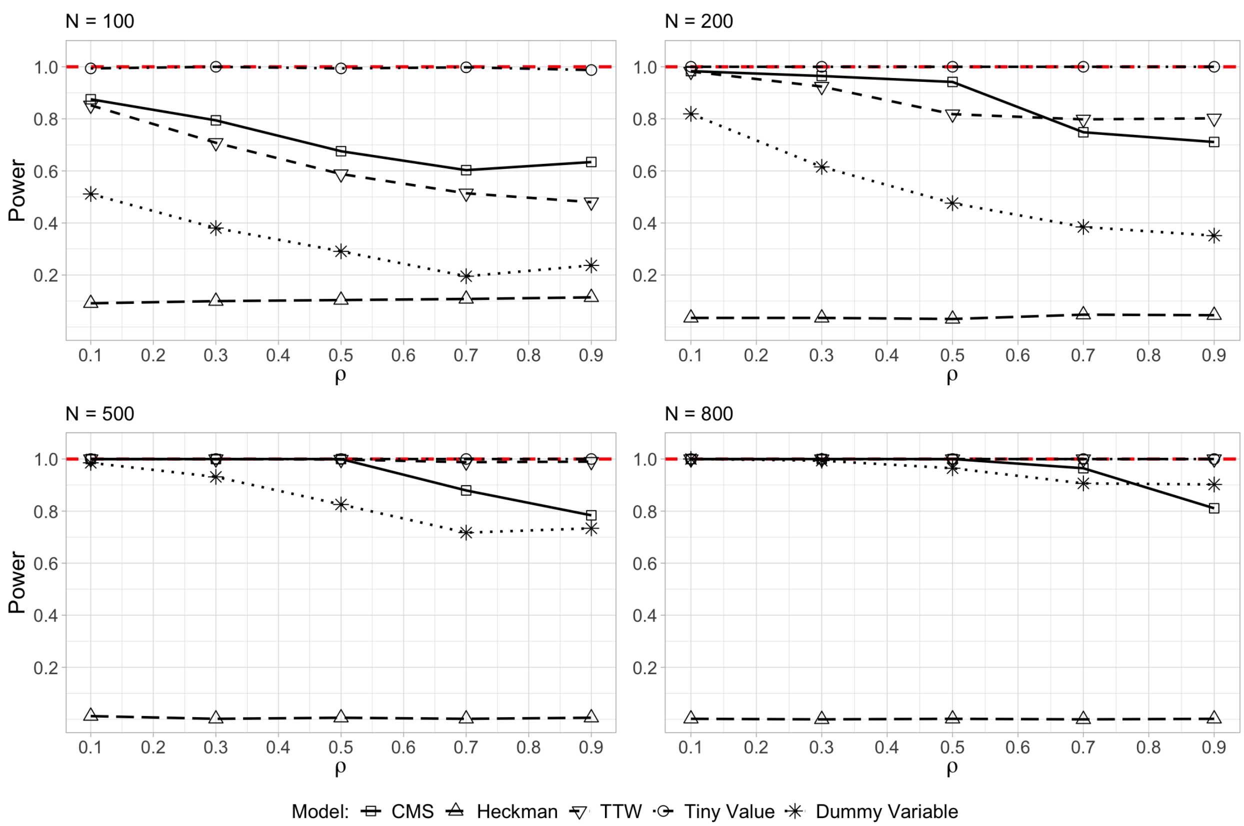


Figure S53: Power of $\hat{\beta}_{1A}$—66% of Districts Partially Contested

D.3 Random Coefficients

In this section, the DGP allows the effect of a predictor on the trade-off between parties that contest all districts (B and C) to vary with partially and fully contested districts: $\beta_{1B} = -1$ if $d_{iA} = 1$ (A contests i), otherwise $\beta_{1B} = 0.5$. We focus on bias and MSE. Estimating different effects across patterns of contestation might be a potential advantage of the TTW approach that estimates a separate SUR model for each distinct pattern of contestation. We show that our strategy (CMS) can also be used to estimate unbiased $\hat{\beta}_{1B}$s across patterns of contestation, but differently from other approaches, it returns mostly unbiased and efficient estimates of $\hat{\beta}_{1A}$ too.

D.3.1 33% of Districts are Partially Contested

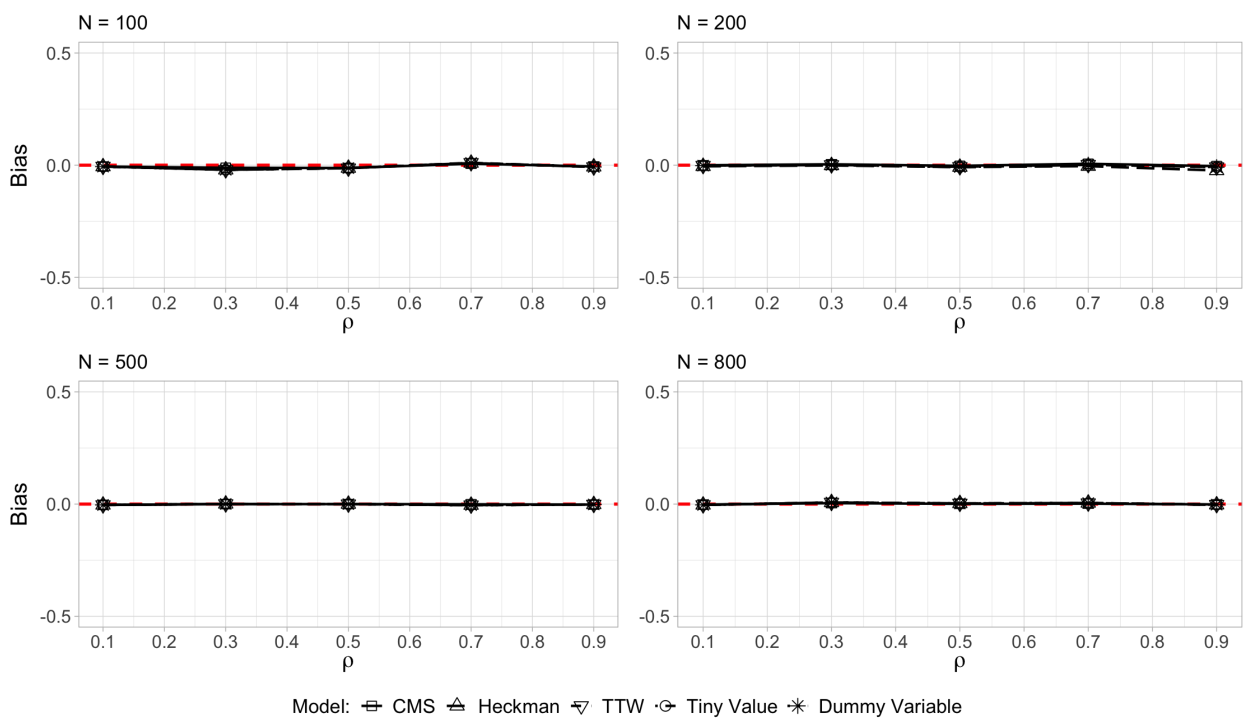


Figure S54: Bias in $\hat{\beta}_{1B}$ in Fully Contested Districts

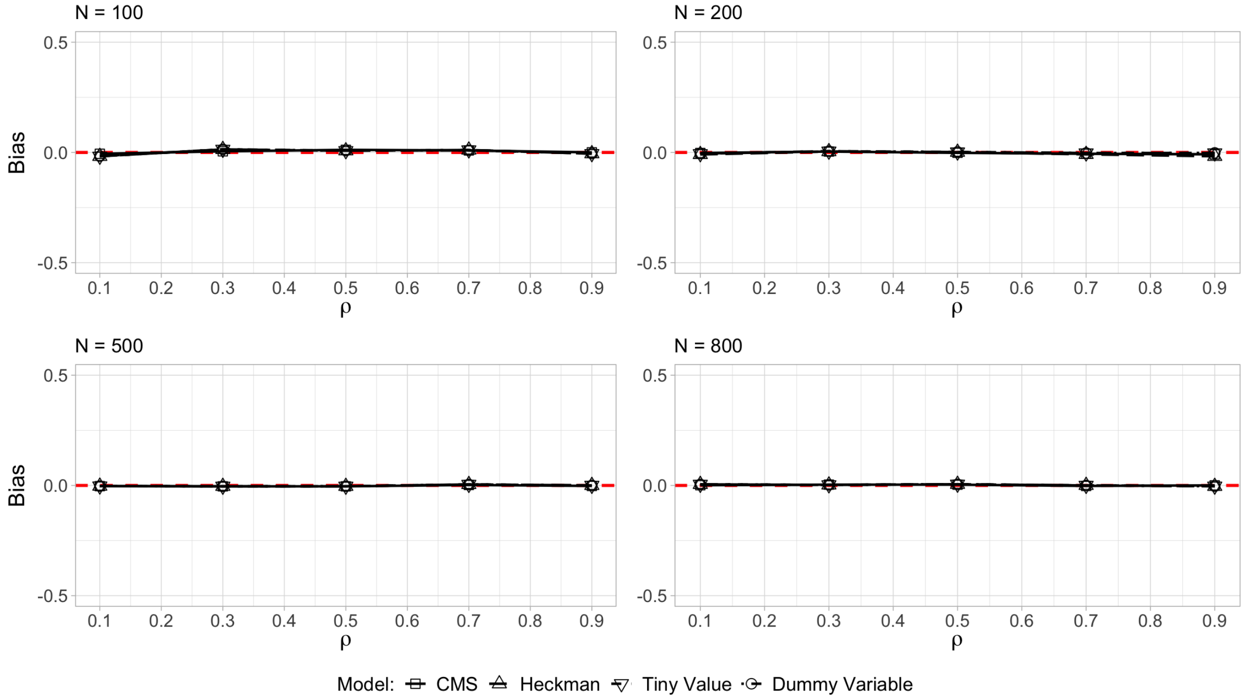


Figure S55: Bias in $\hat{\beta}_{1B}$ in Partially Contested Districts

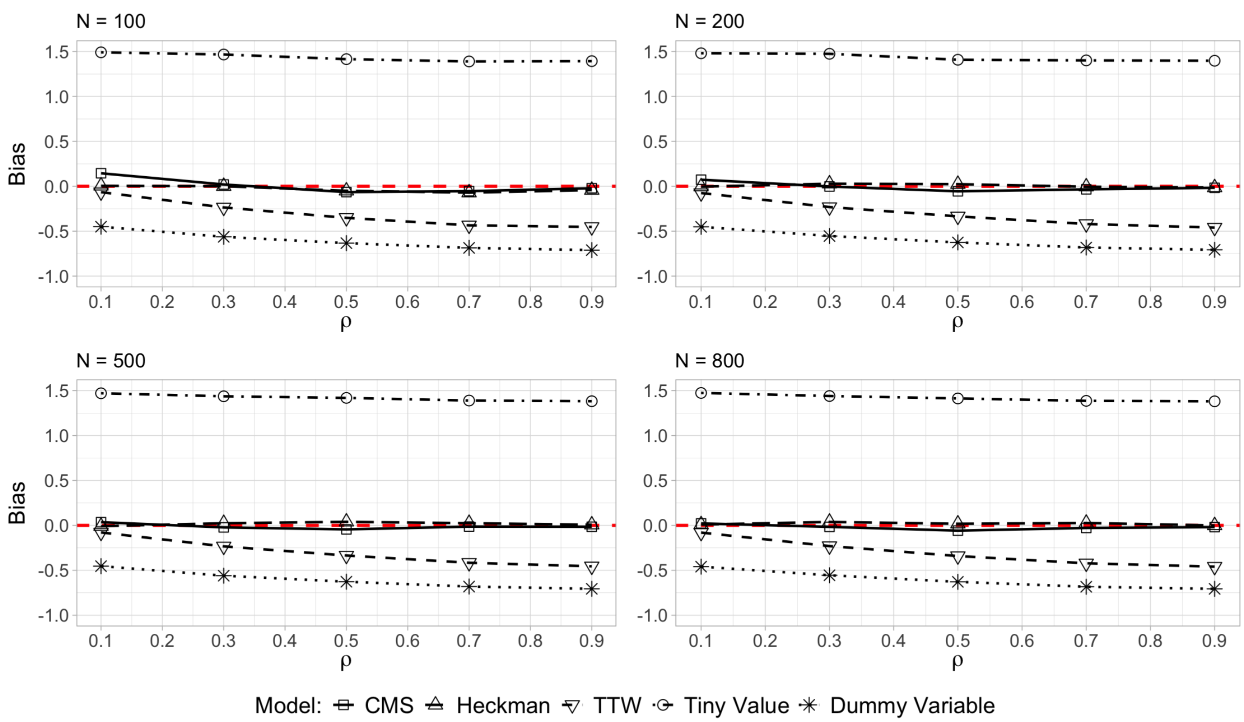


Figure S56: Bias in $\hat{\beta}_{1A}$—33% of Districts Partially Contested

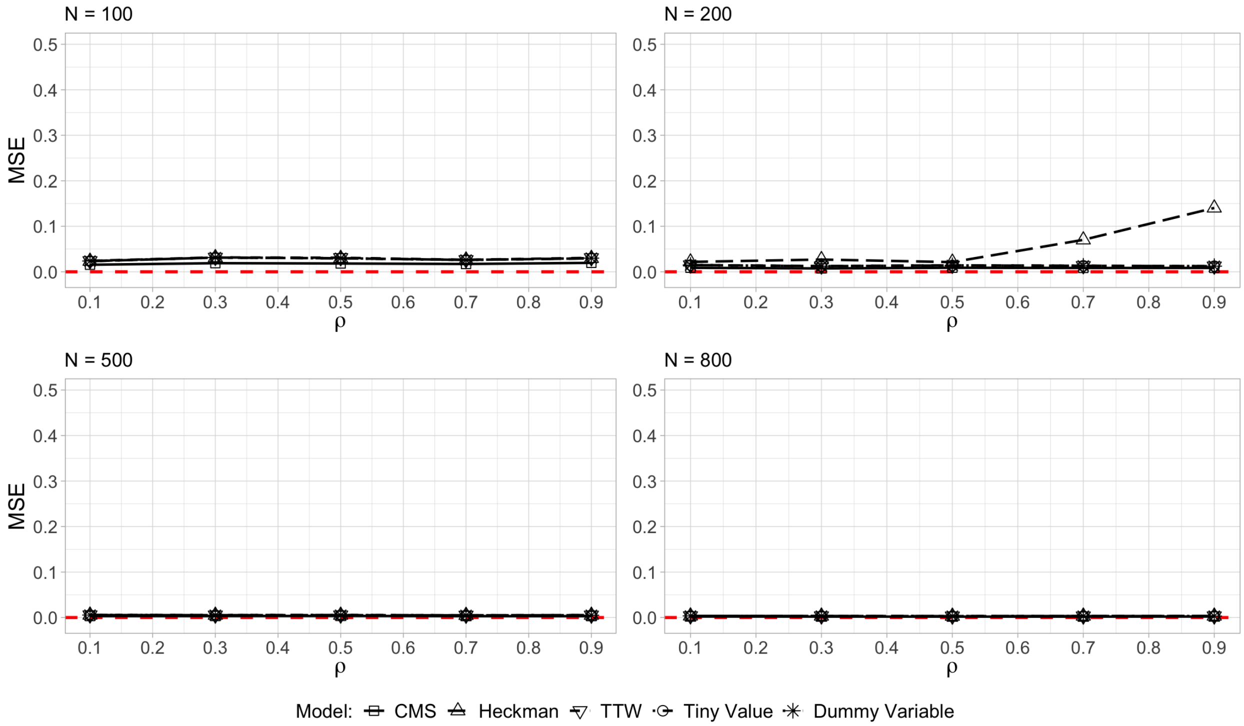


Figure S57: MSE of $\hat{\beta}_{1B}$ in Fully Contested Districts

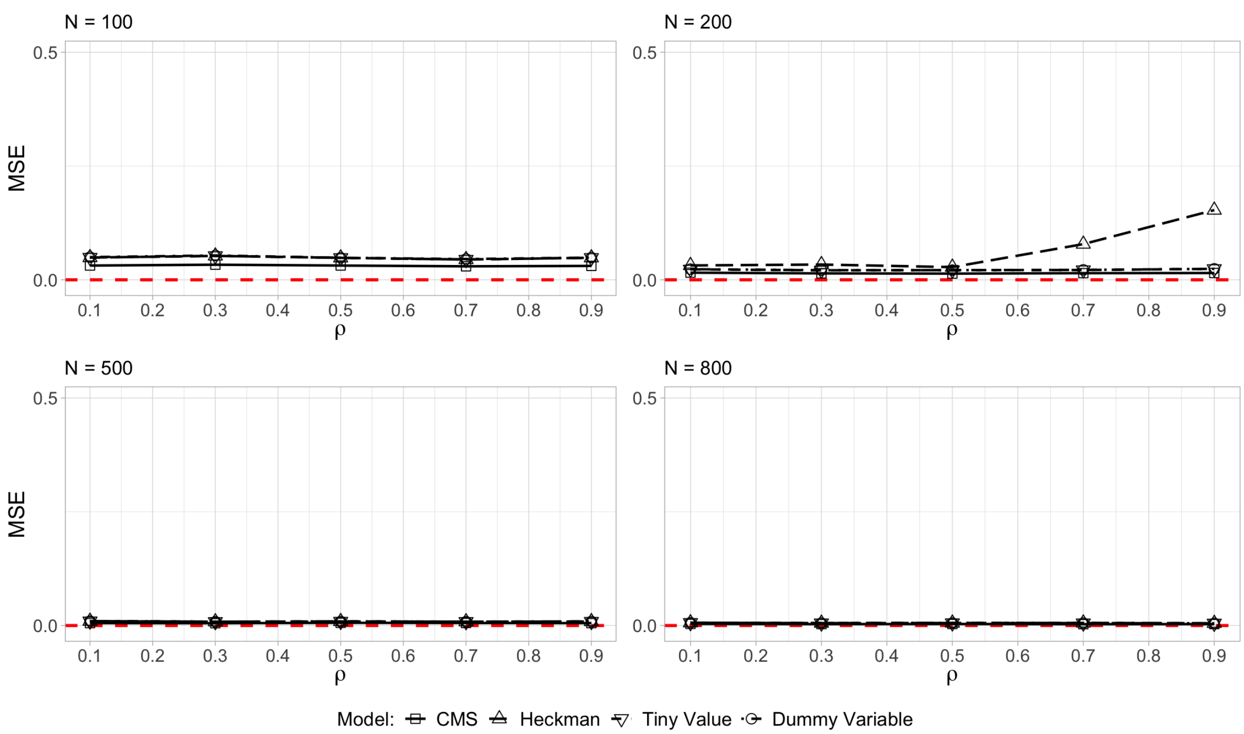


Figure S58: MSE of $\hat{\beta}_{1B}$ in Partially Contested Districts

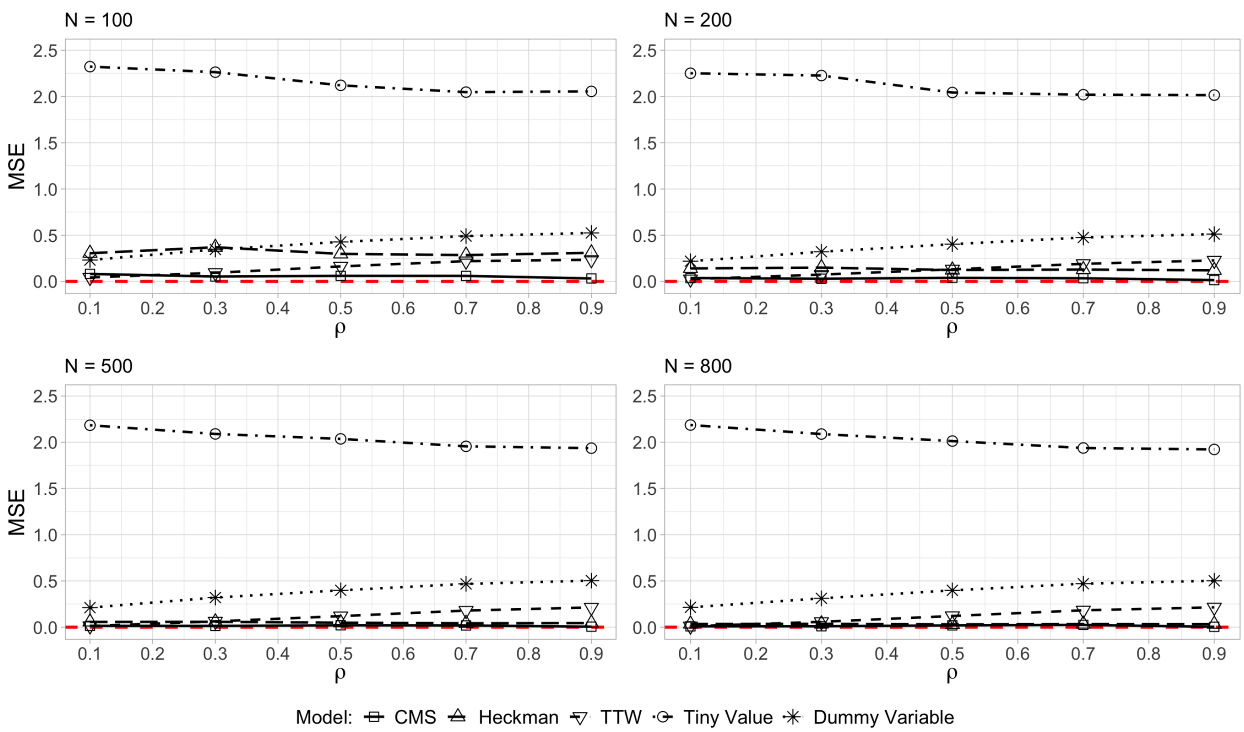


Figure S59: MSE of $\hat{\beta}_{1A}$—33% of Districts Partially Contested

D.3.2 50% of Districts are Partially Contested

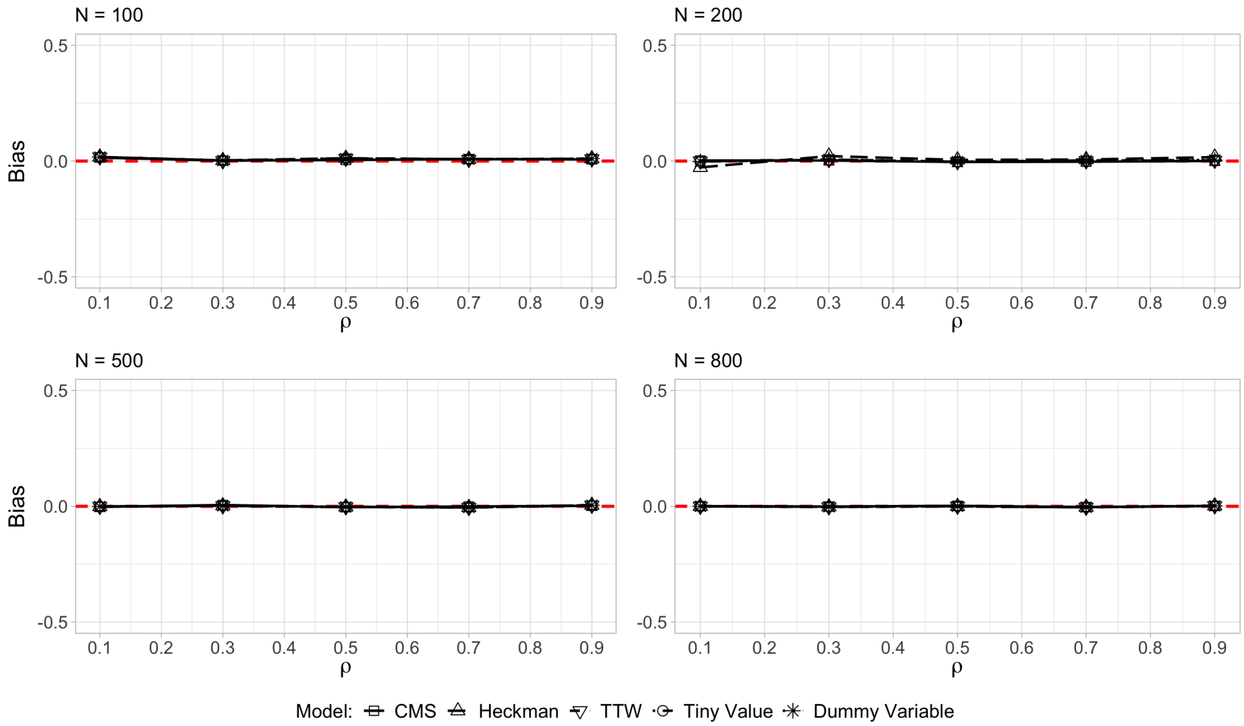


Figure S60: Bias in $\hat{\beta}_{1B}$ in Fully Contested Districts

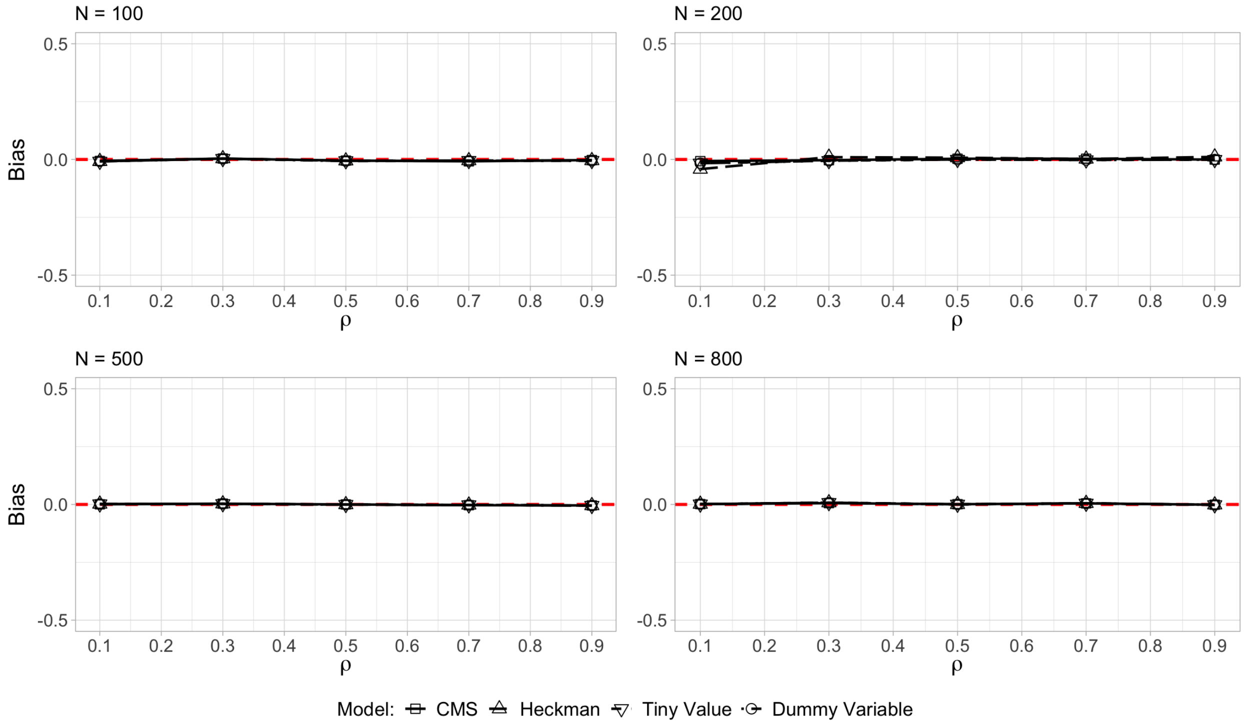


Figure S61: Bias in $\hat{\beta}_{1B}$ in Partially Contested Districts

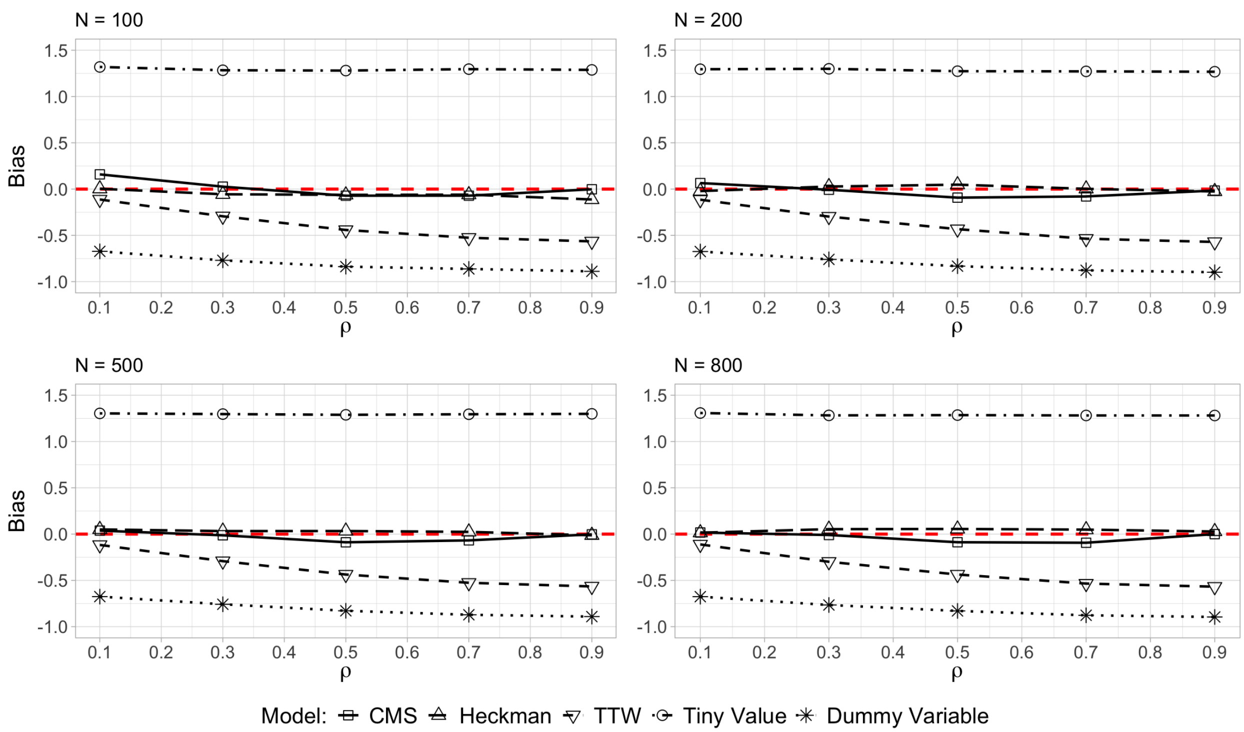


Figure S62: Bias in $\hat{\beta}_{1A}$—50% of Districts Partially Contested

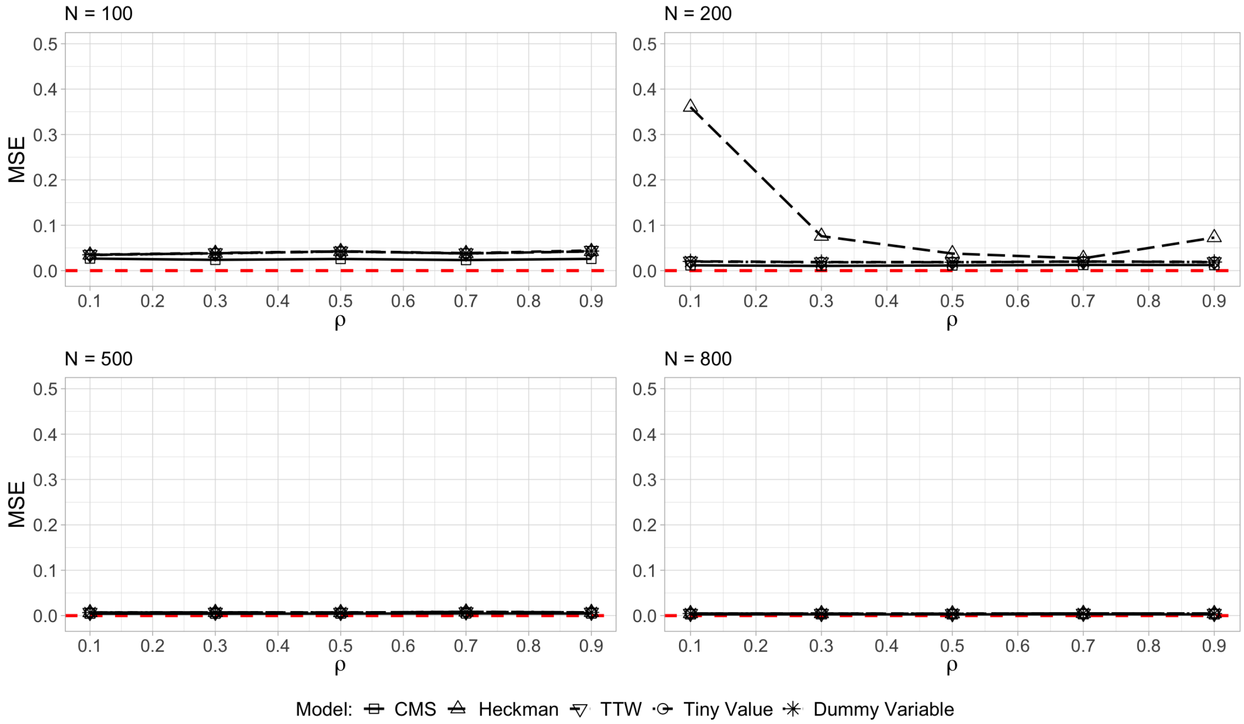


Figure S63: MSE of $\hat{\beta}_{1B}$ in Fully Contested Districts

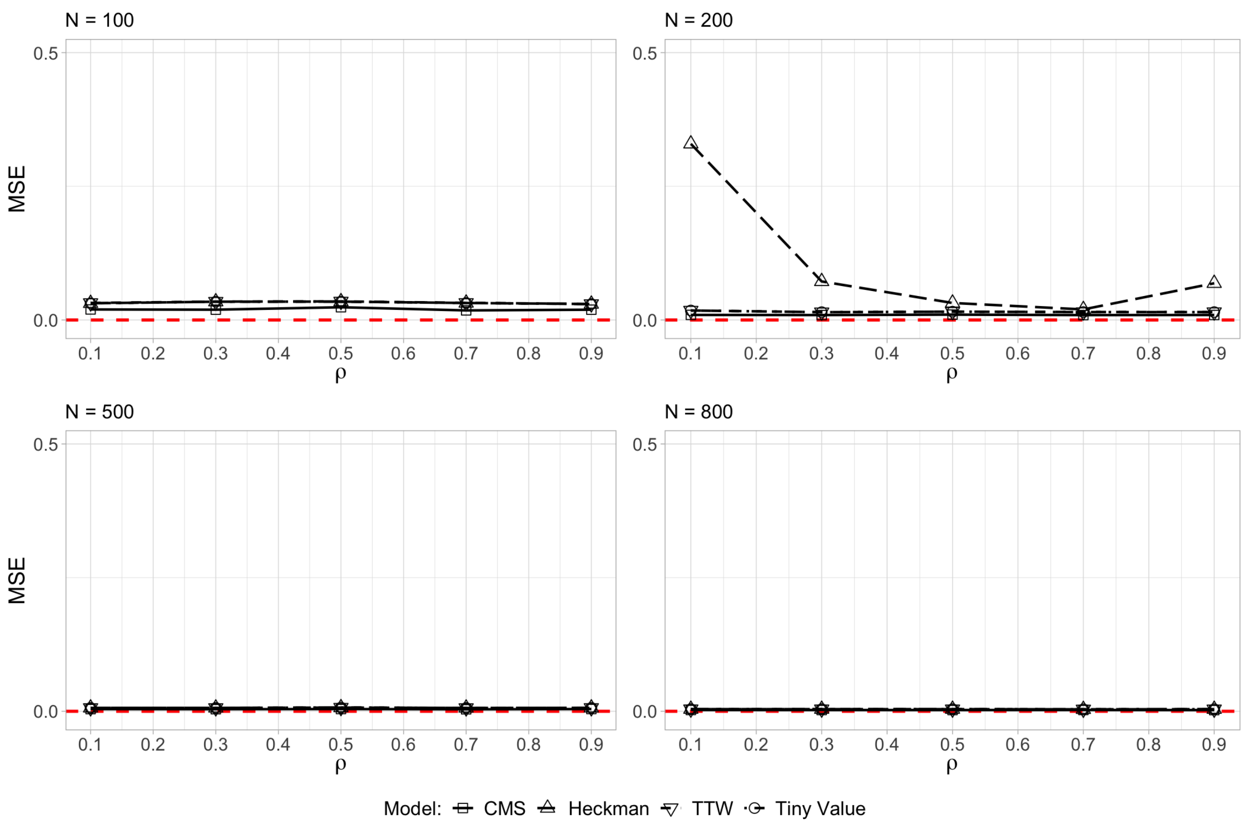


Figure S64: MSE of $\hat{\beta}_{1B}$ in Partially Contested Districts

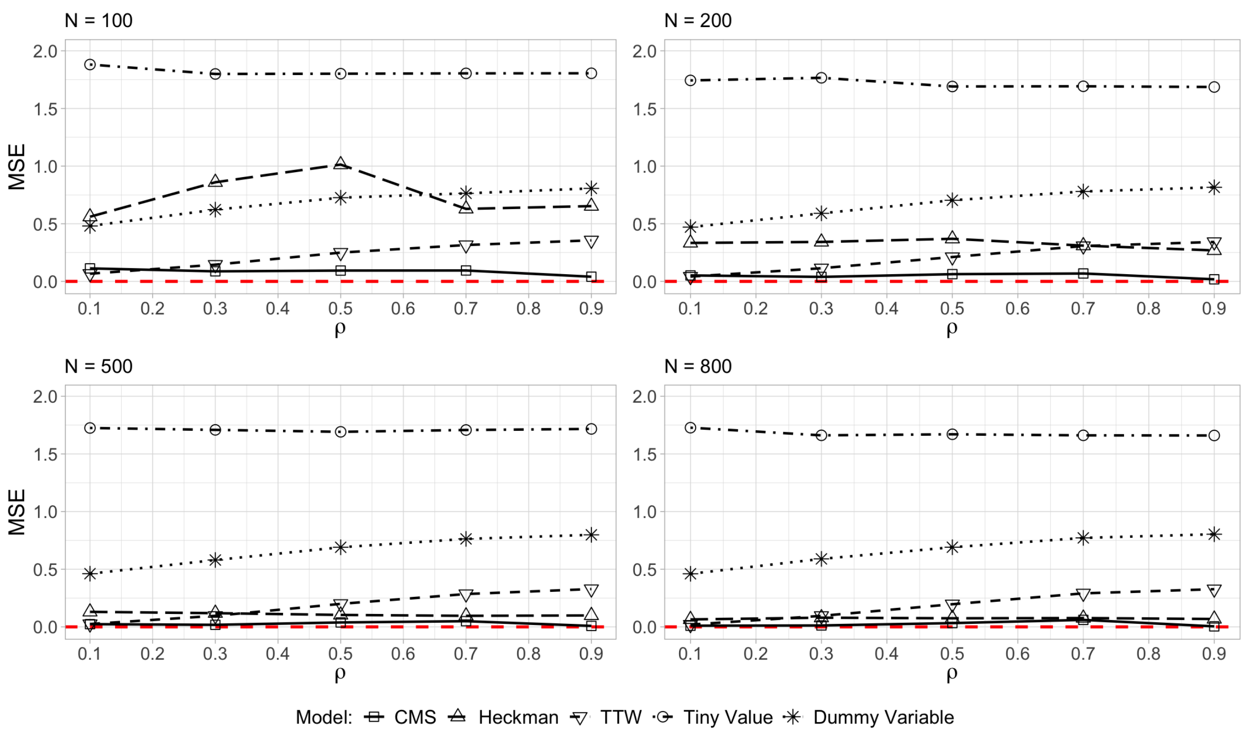


Figure S65: MSE of $\hat{\beta}_{1A}$—50% of Districts Partially Contested

D.3.3 66% of Districts are Partially Contested

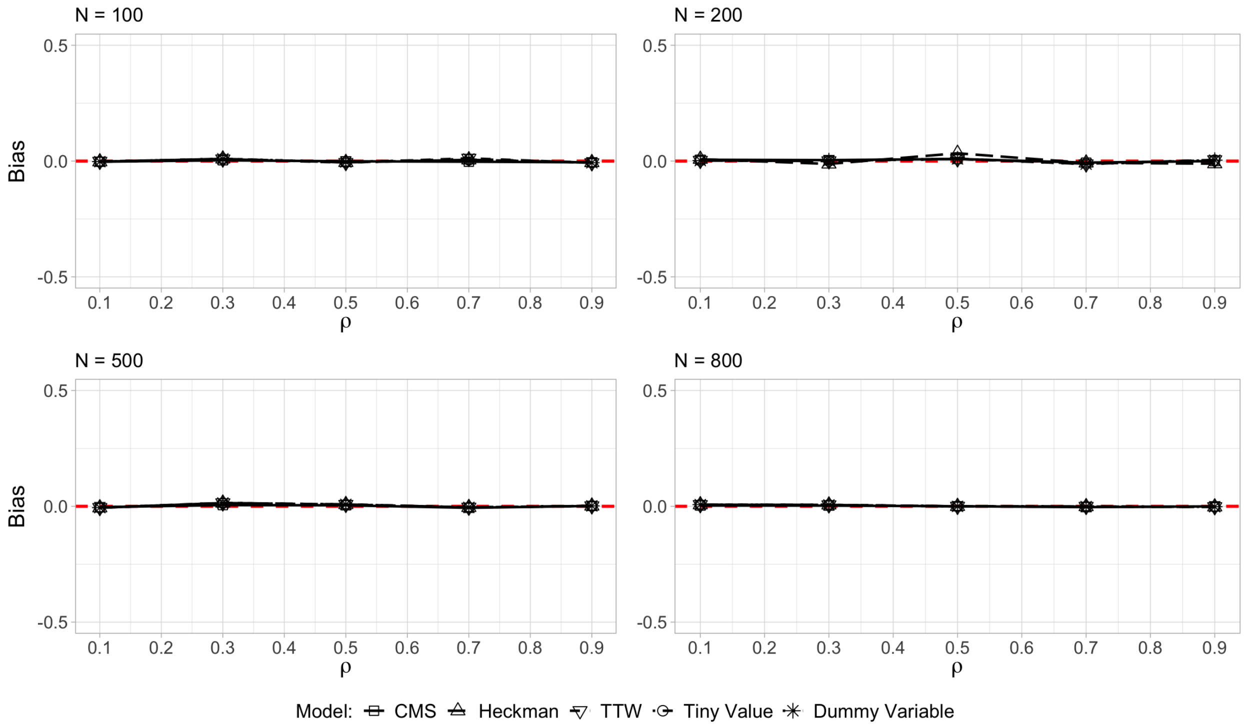


Figure S66: Bias in $\hat{\beta}_{1B}$ in Fully Contested Districts

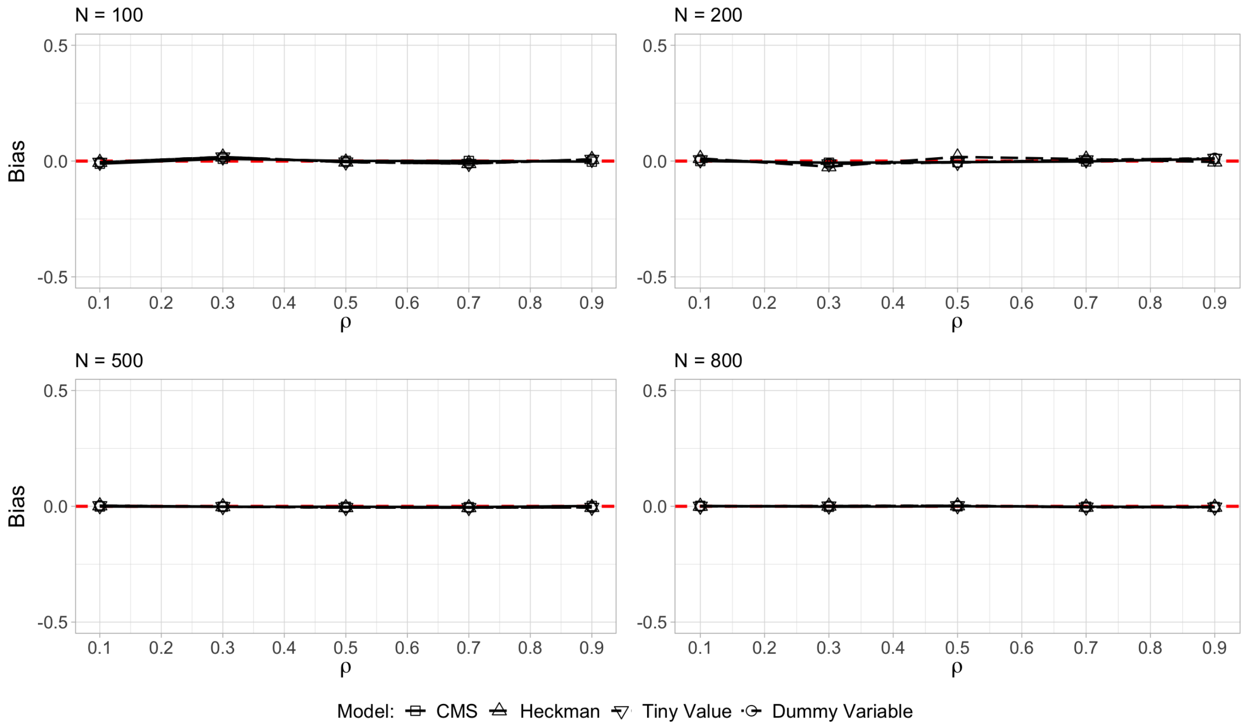


Figure S67: Bias in $\hat{\beta}_{1B}$ in Partially Contested Districts

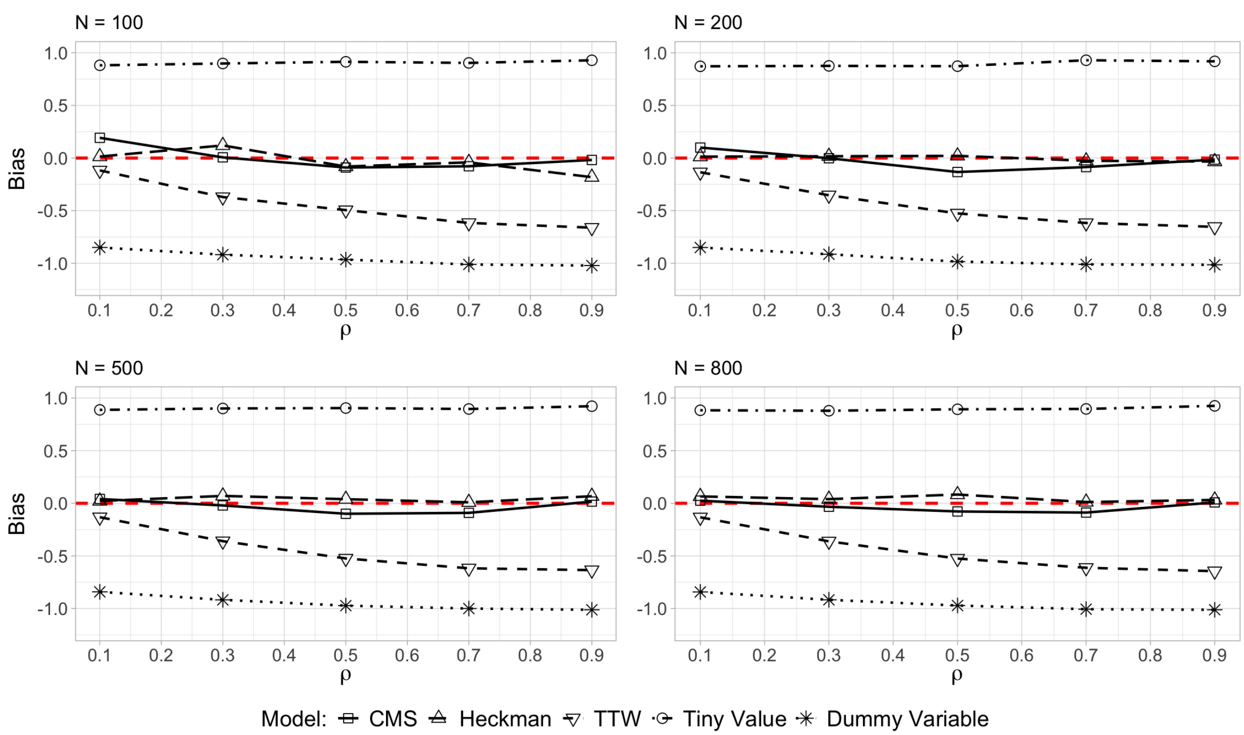


Figure S68: Bias in $\hat{\beta}_{1A}$—66% of Districts Partially Contested

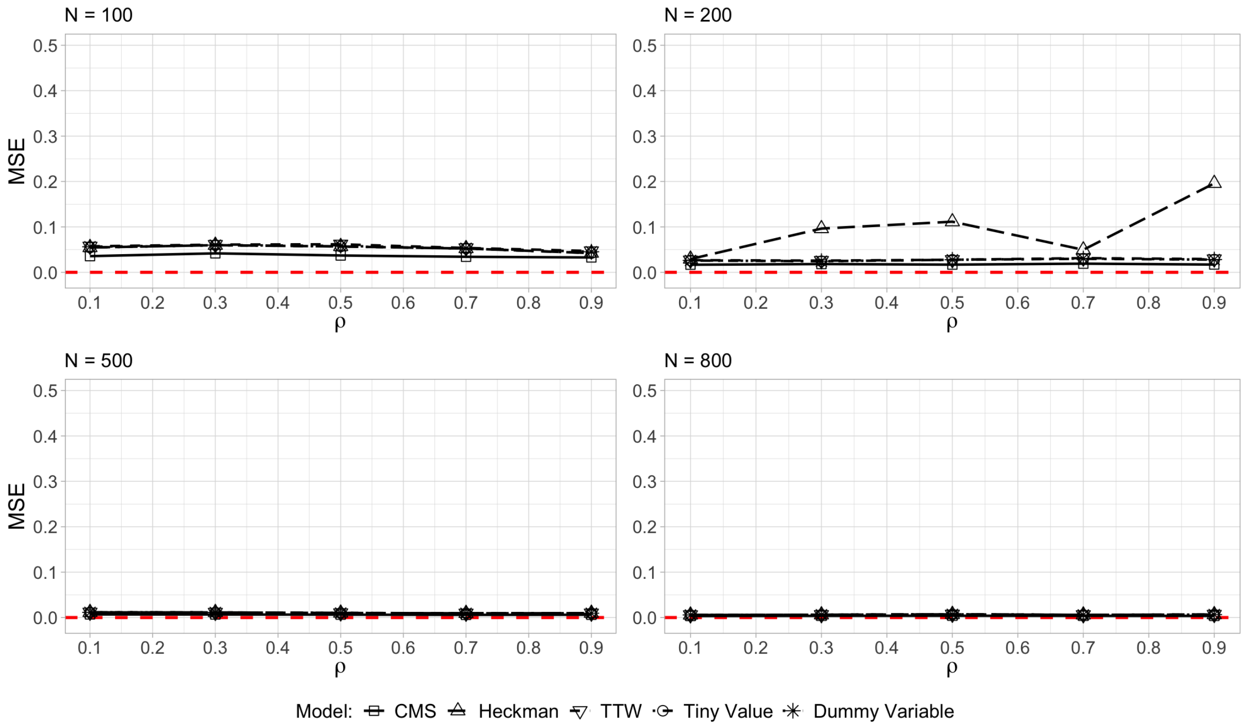


Figure S69: MSE of $\hat{\beta}_{1B}$ in Fully Contested Districts

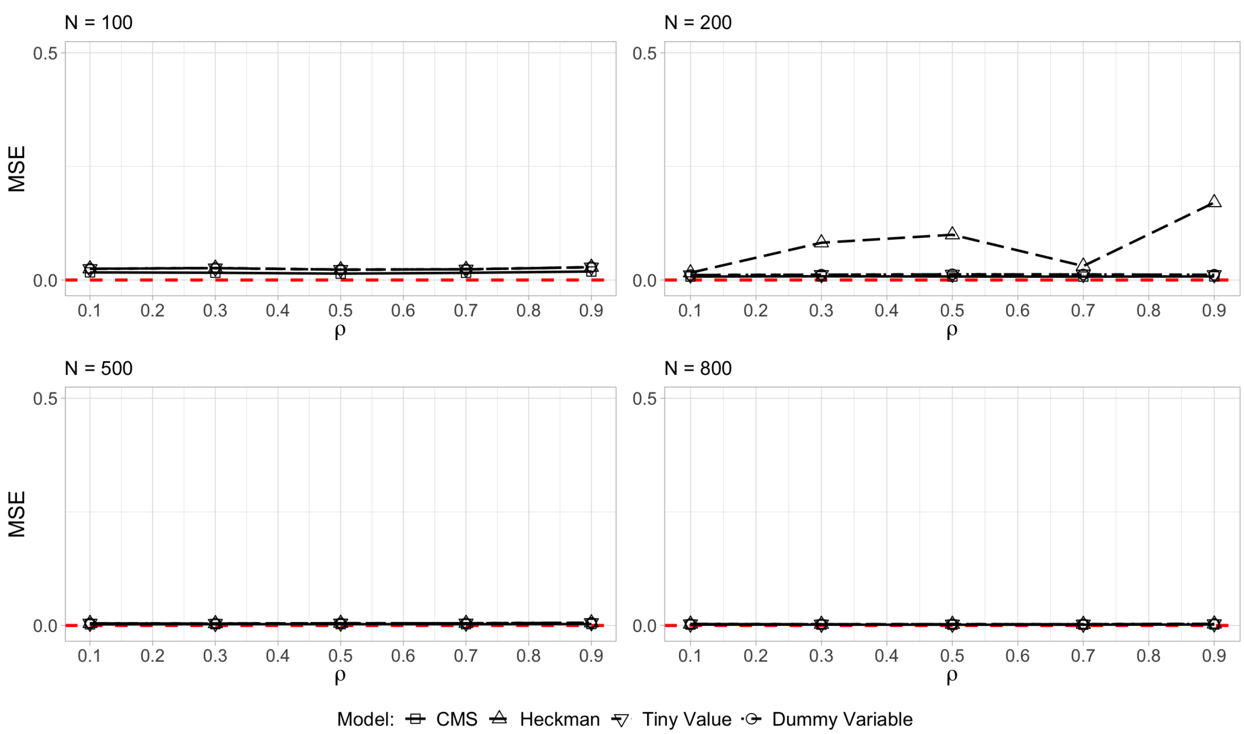


Figure S70: MSE of $\hat{\beta}_{1B}$ in Partially Contested Districts

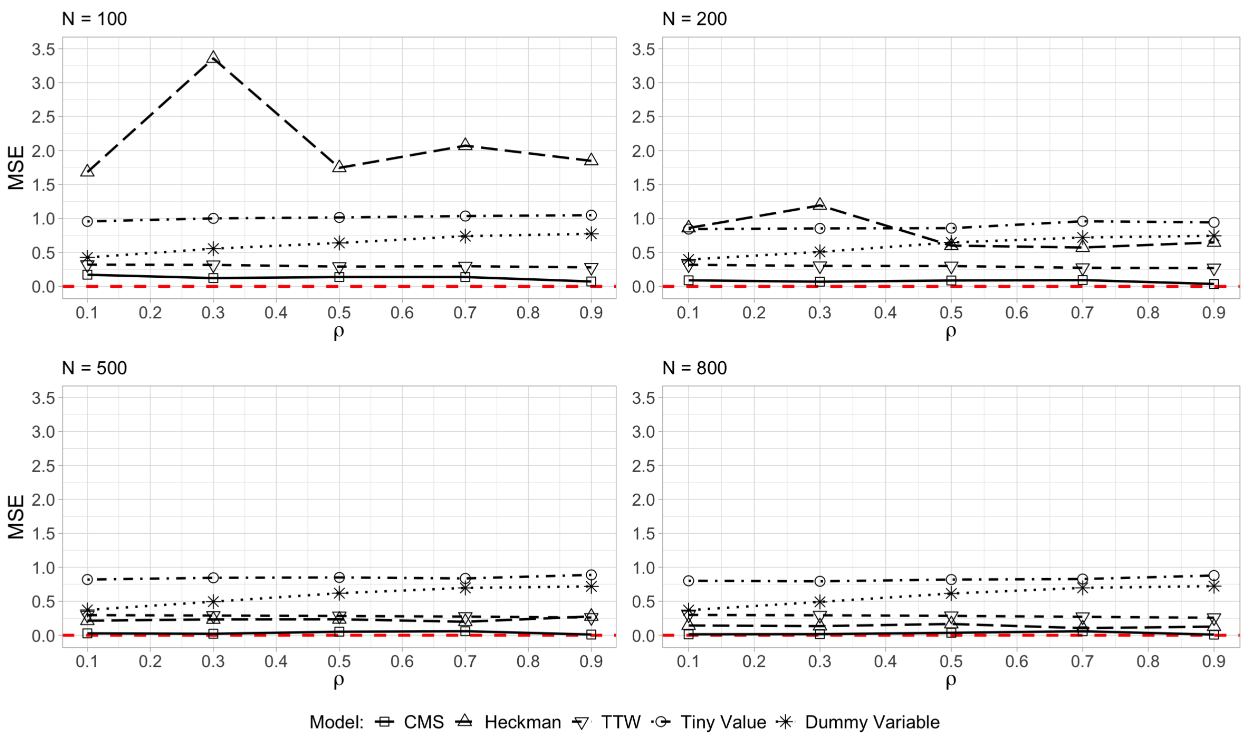


Figure S71: MSE of $\hat{\beta}_{1A}$—66% of Districts Partially Contested

D.4 Other Coefficients

In this section, we show the performance of our model to estimate other parameters in the selection (α_{1A} and α_{2A}) and outcome (β_{1B}) stages. We focus on bias and MSE in a DGP with 50% of districts partially contested, a valid instrument, and fixed coefficients.

D.4.1 β_{1B}—50% of Districts are Partially Contested & Fixed Coefficient

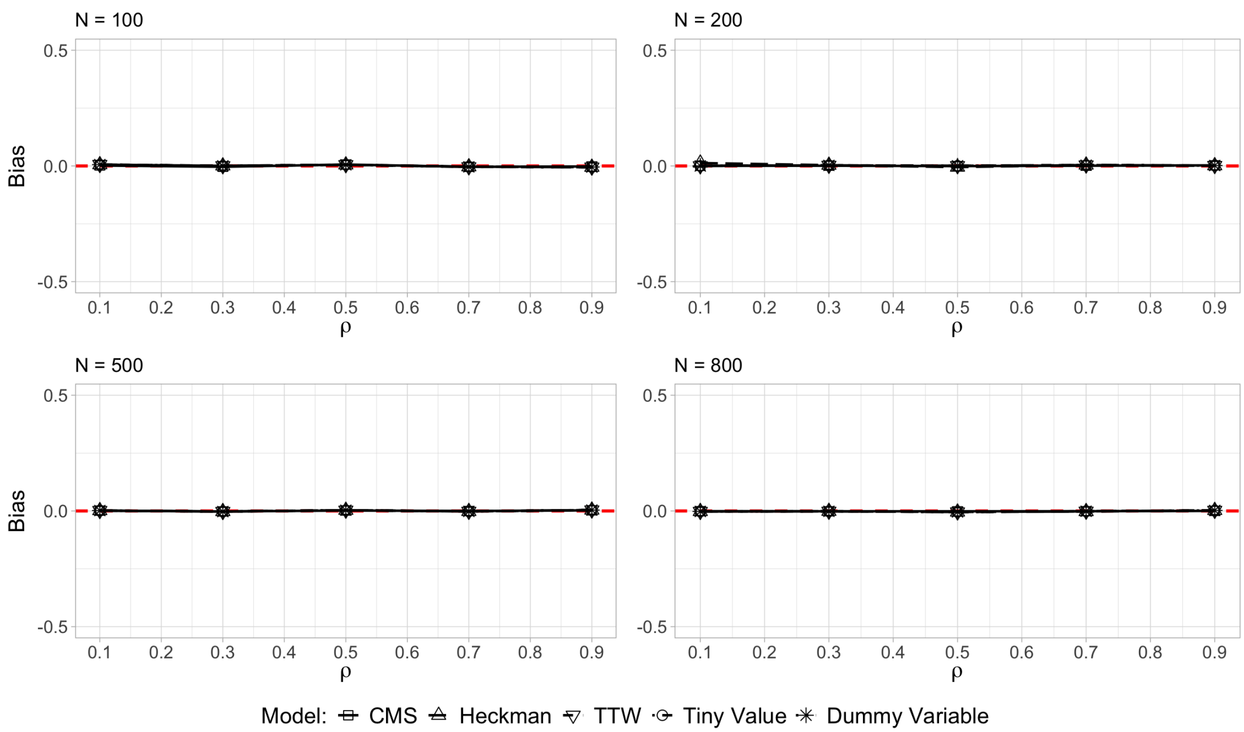


Figure S72: Bias in $\hat{\beta}_{1B}$—50% of Districts Partially Contested

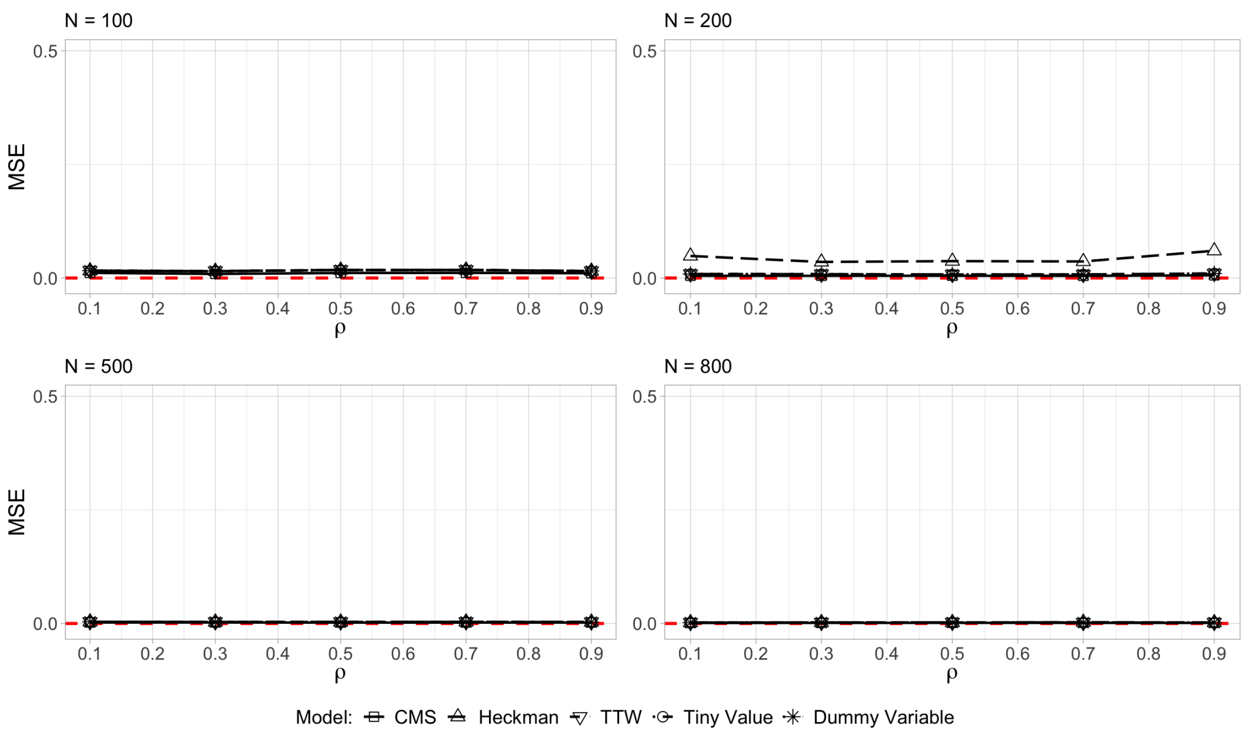


Figure S73: MSE of $\hat{\beta}_{1B}$ in Fully Contested Districts

D.4.2 α_{1A}—50% of Districts are Partially Contested & Fixed Coefficient

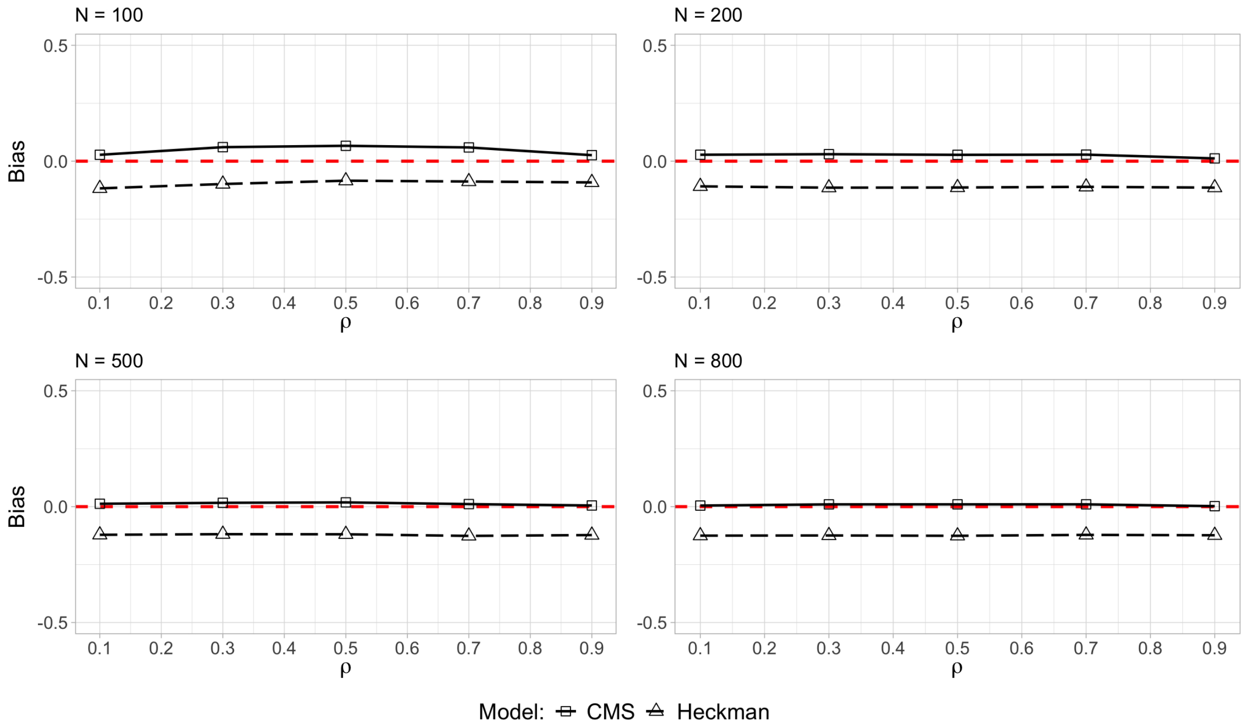


Figure S74: Bias in $\hat{\alpha}_{1A}$—50% of Districts Partially Contested

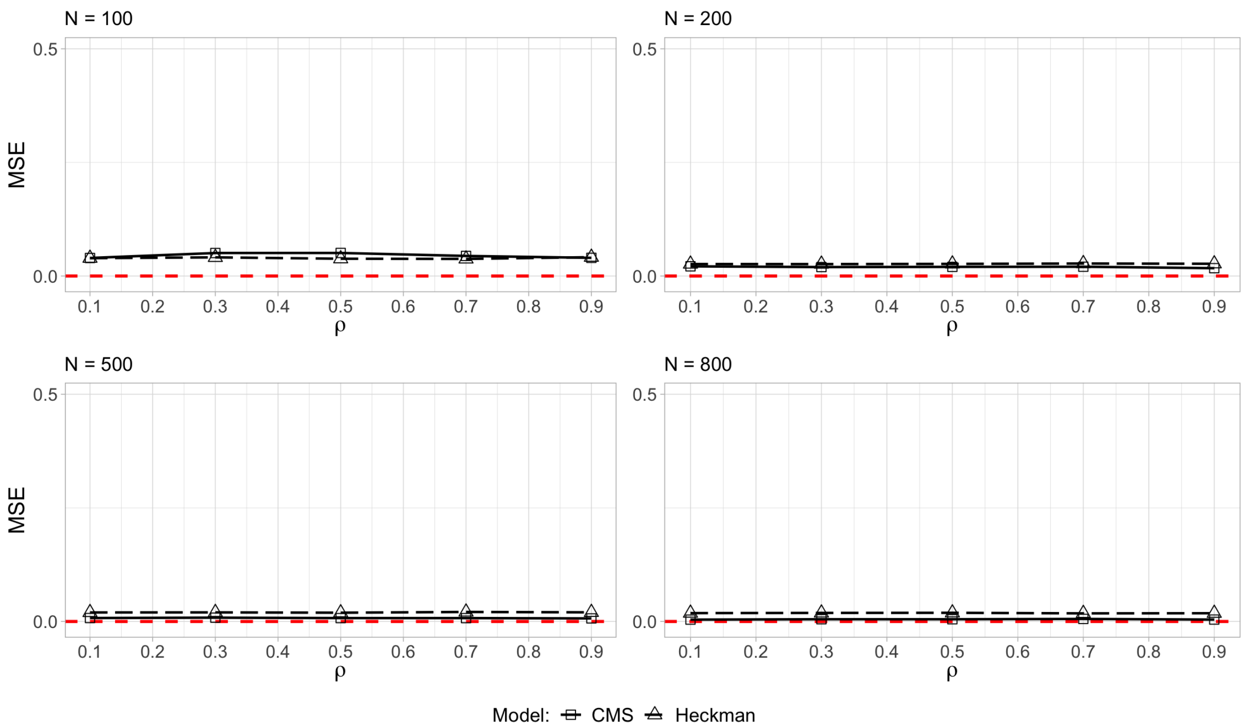


Figure S75: MSE of $\hat{\alpha}_{1A}$ in Fully Contested Districts

D.4.3 α_{2A}—50% of Districts are Partially Contested & Fixed Coefficient

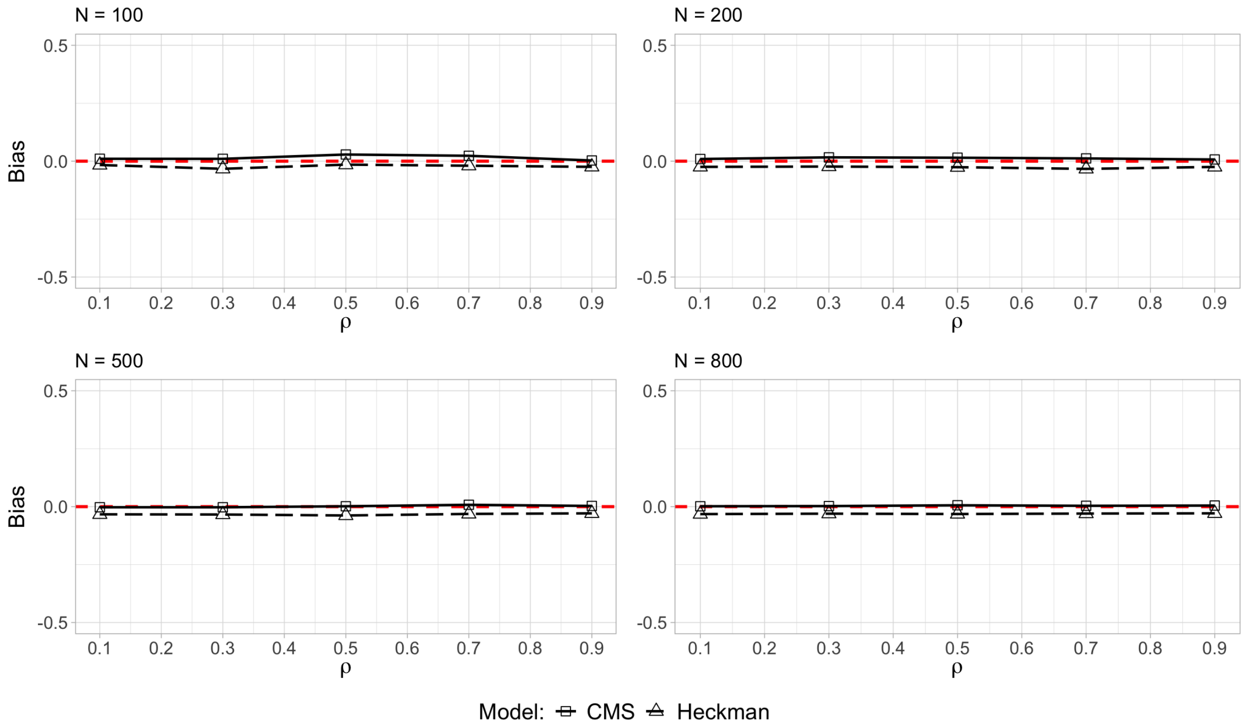


Figure S76: Bias in $\hat{\alpha}_{2A}$—50% of Districts Partially Contested

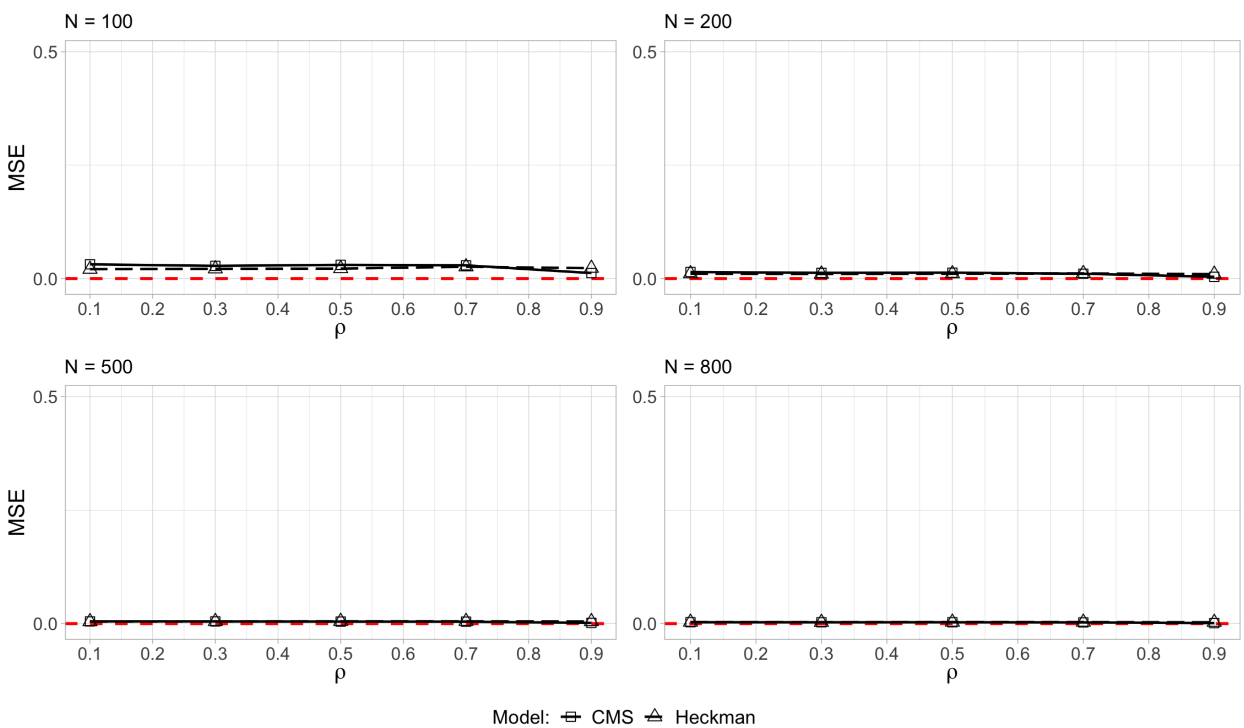


Figure S77: MSE of $\hat{\alpha}_{2A}$ in Fully Contested Districts

D.5 Normal DGP & Extreme Value Type-I Assumption

In this section, we show results from Monte Carlo simulations in which the error terms in the selection and outcome stages (u_{ij} and ε_{ij}) are jointly normally distributed, but the CMS still assumes that they would follow an extreme value type-I distribution. Party A does not contest 33% of the 500 districts. The results show that our model performs relatively well even when this assumption about the errors structures is violated. As Figure S78 shows, bias in our model is always small (between 4% and 8%) and decreases with the correlation across stages (ρ). Note that as our empirical example has demonstrated, $\hat{\rho}$ is typically high, with values above 0.8 in all equations of our empirical example (see Table S2).

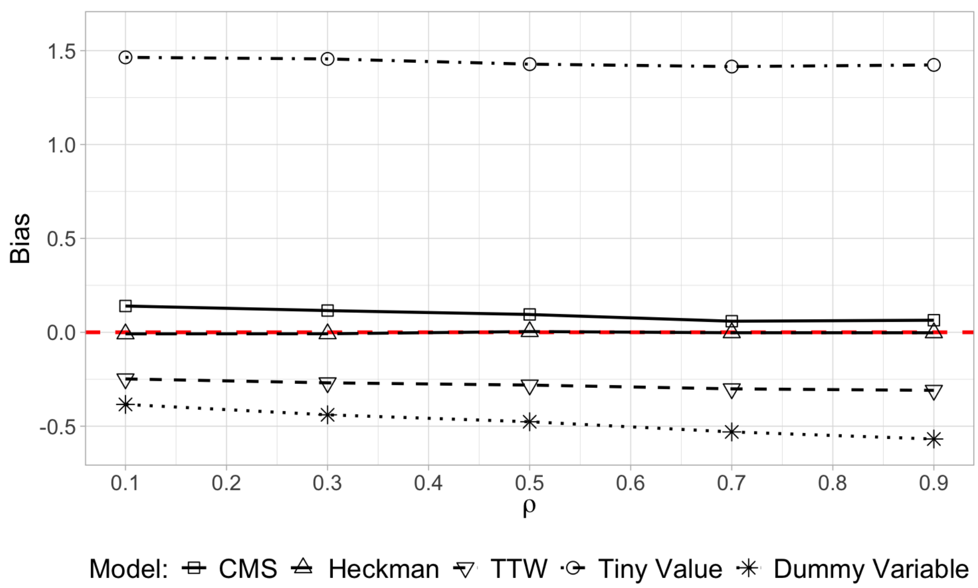


Figure S78: Bias of $\hat{\beta}_{1A}$

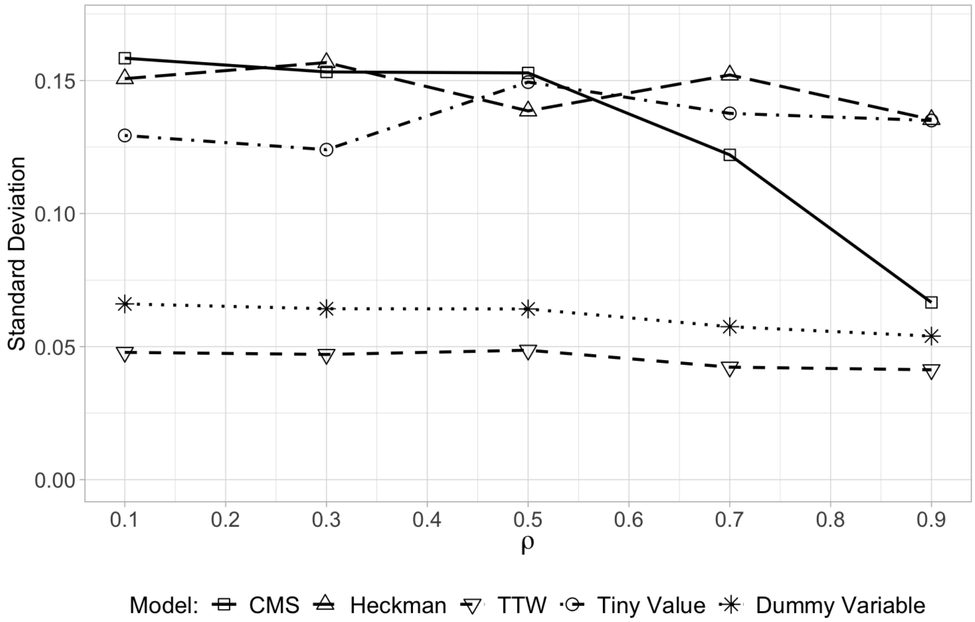


Figure S79: Standard Deviation of $\hat{\beta}_{1A}$

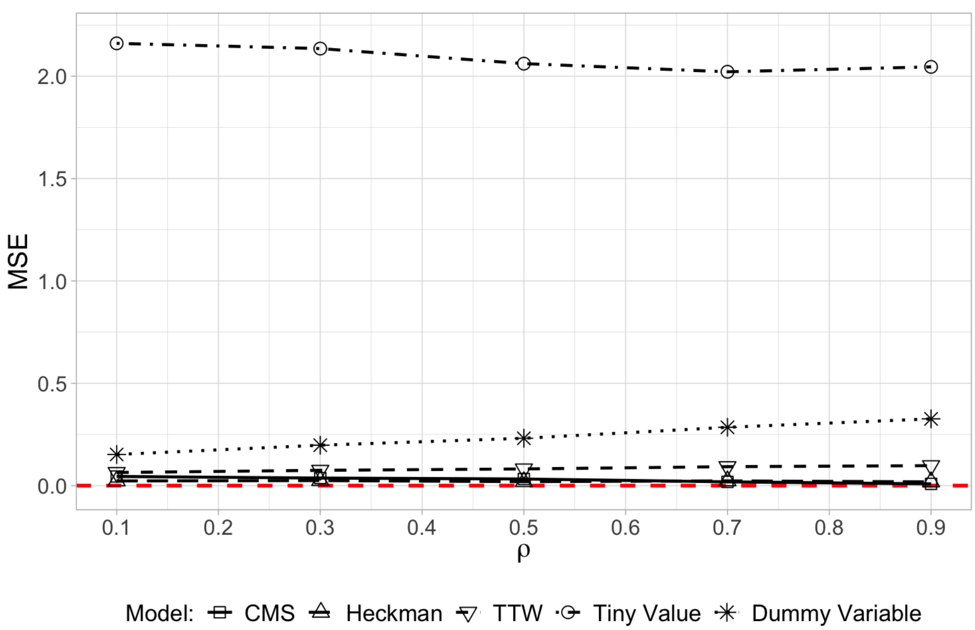


Figure S80: MSE of $\hat{\beta}_{1A}$

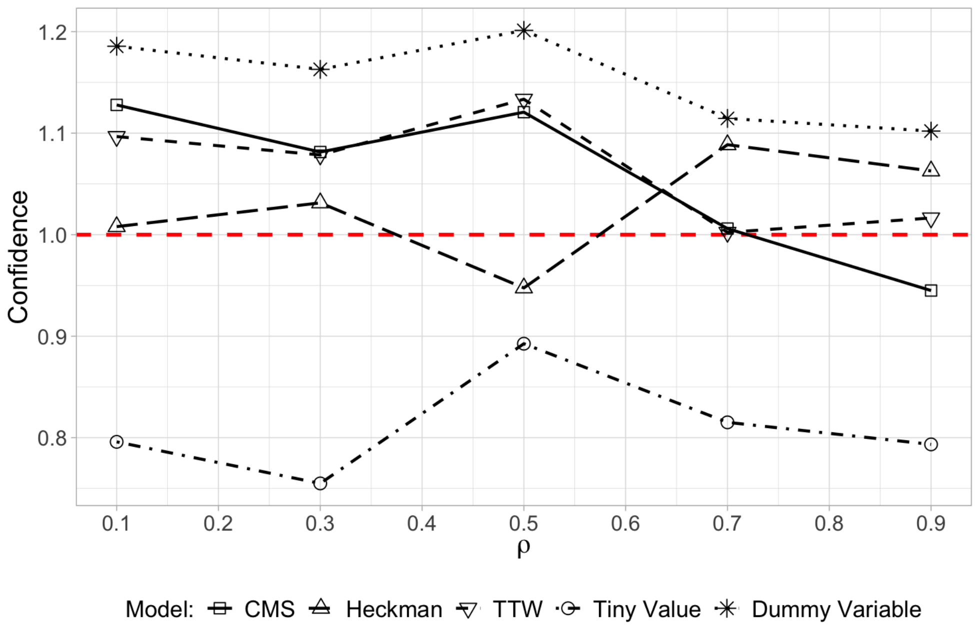


Figure S81: Confidence of $\hat{\beta}_{1A}$

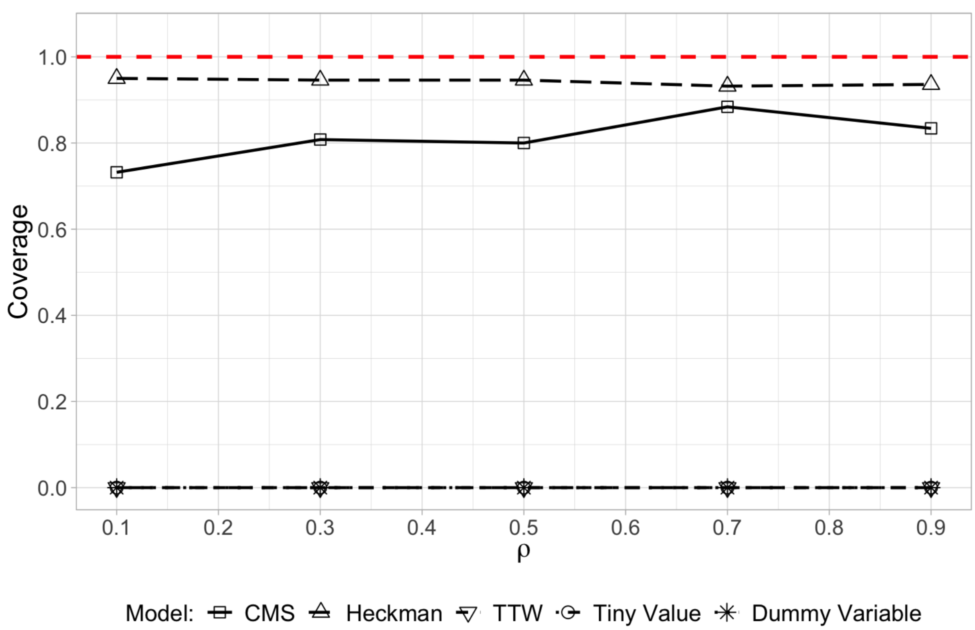


Figure S82: Coverage of $\hat{\beta}_{1A}$

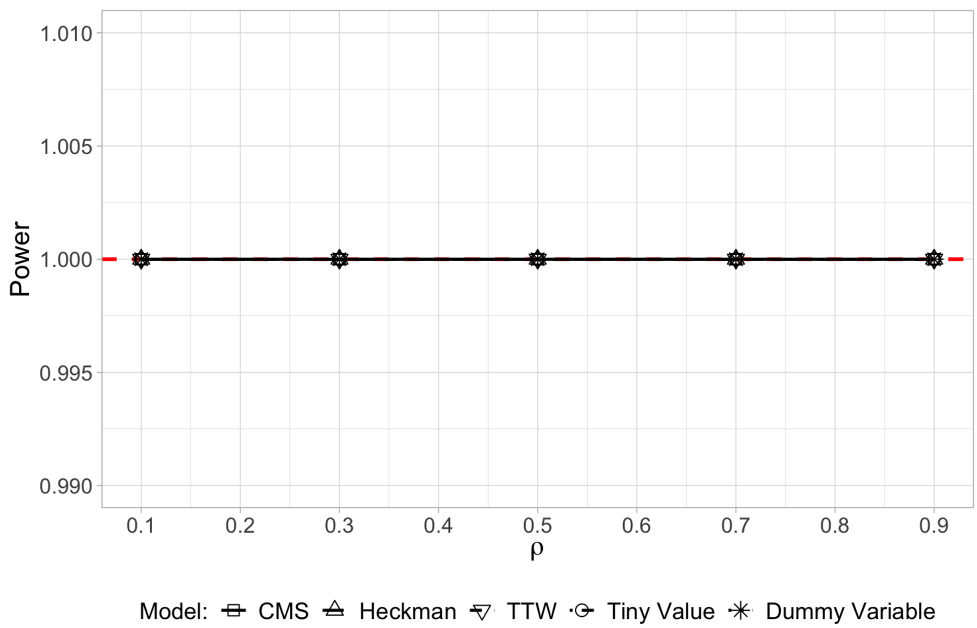


Figure S83: Power of $\hat{\beta}_{1A}$

D.6 Normal DGP & Normal Distribution Assumption

A crucial advantage of our CMS is its flexibility in that the error structure can be adjusted to reflect theoretical expectations about the DGP. In this section, we illustrate this advantage by simulating partially contested elections in which the errors follow a jointly normal distribution and the log-likelihood function of our CMS also assumes that u_{ij} and ε_{ij} are jointly normally distributed. Party A does not contest 33% of all the 500 districts. The results show that the CMS returns unbiased estimates and is more efficient than the Heckman correction.

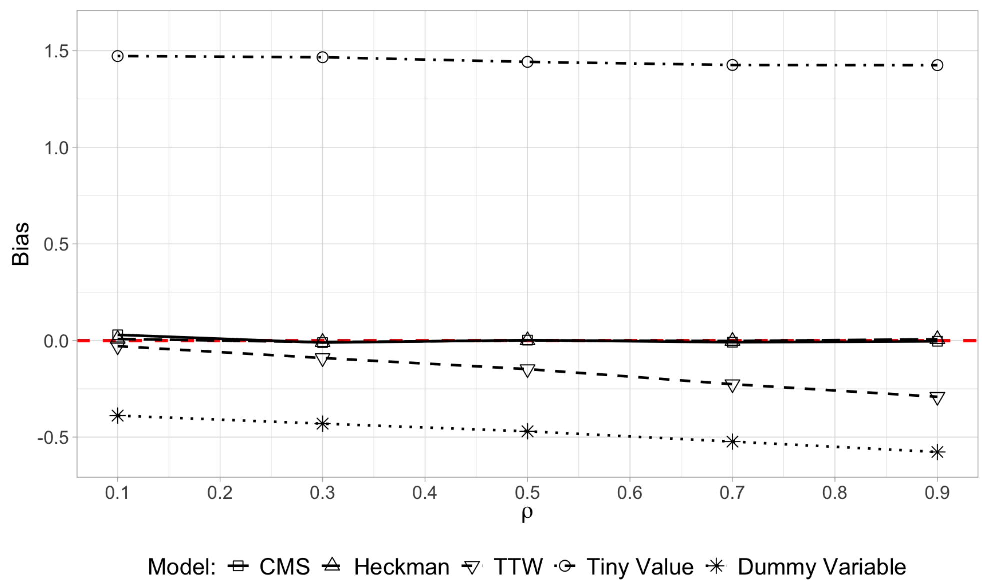


Figure S84: Bias of $\hat{\beta}_{1A}$

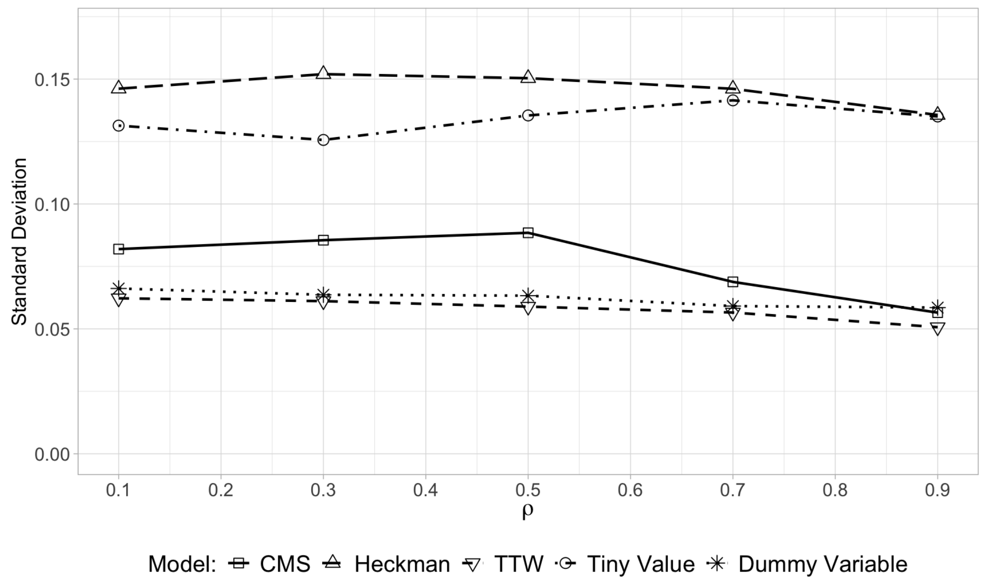


Figure S85: Standard Deviation of $\hat{\beta}_{1A}$

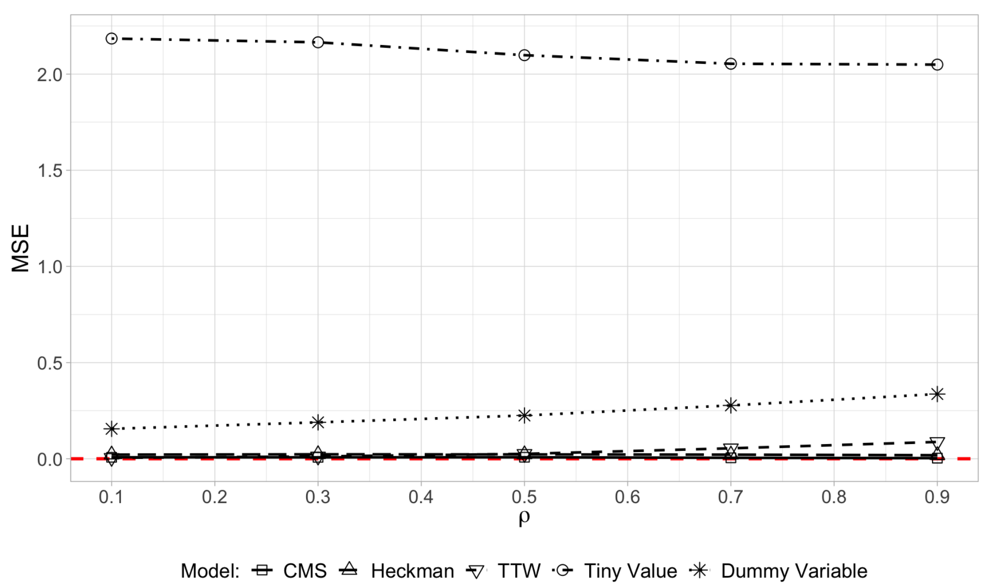


Figure S86: MSE of $\hat{\beta}_{1A}$

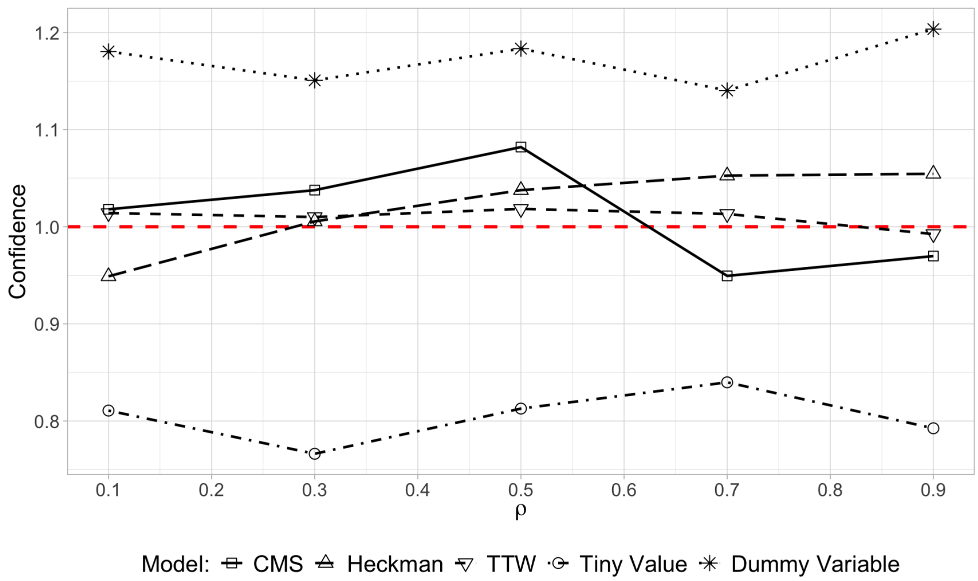


Figure S87: Confidence of $\hat{\beta}_{1A}$

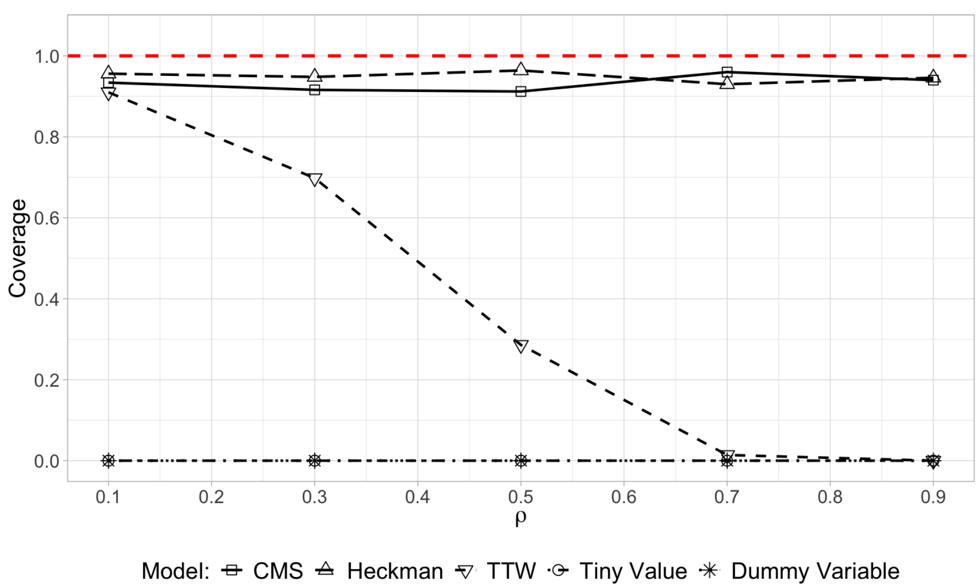


Figure S88: Coverage of $\hat{\beta}_{1A}$

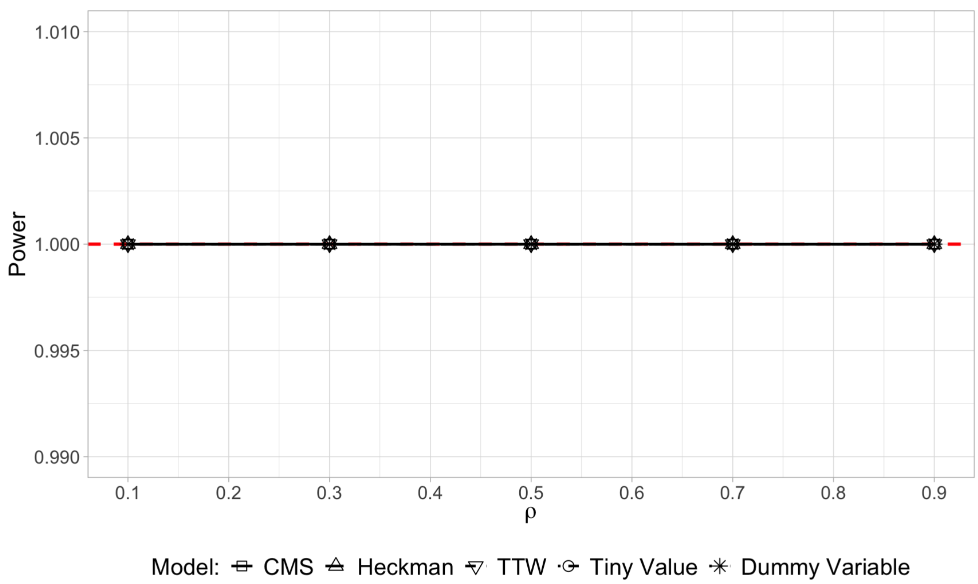


Figure S89: Power of $\hat{\beta}_{1A}$

D.7 Monte Carlo Simulations with 6 Parties

In this section, we show results from a Monte Carlo experiment with 6 parties (A, B, C, D, and E) and two of them partially contested districts (A and B). We set the parameters as follows: $\alpha_{0A} = 0.57$, $\alpha_{1A} = 0.8$, $\alpha_{2A} = 0.2$,$\alpha_{0B} = 0.59$, $\alpha_{1B} = 0.9$, $\alpha_{2B} = 0.1$ $\beta_{0A} = 1$, $\beta_{1A} = 1.2$, $\beta_{0B} = 1$, $\beta_{1B} = -1$, $\beta_{0C} = 2$, $\beta_{1C} = 3$, $\beta_{0D} = 0$, $\beta_{1D} = 1$, $\beta_{0E} = -1$, $\beta_{1E} = 2$ and $\beta_{2j} = 0 \forall j = A, B, C, D, E$. Strategies that ignore sample selection lead to biased results. Our approach, in turn, returns consistent estimates for both parties that partially contested districts. It is also more efficient than Heckman correction.

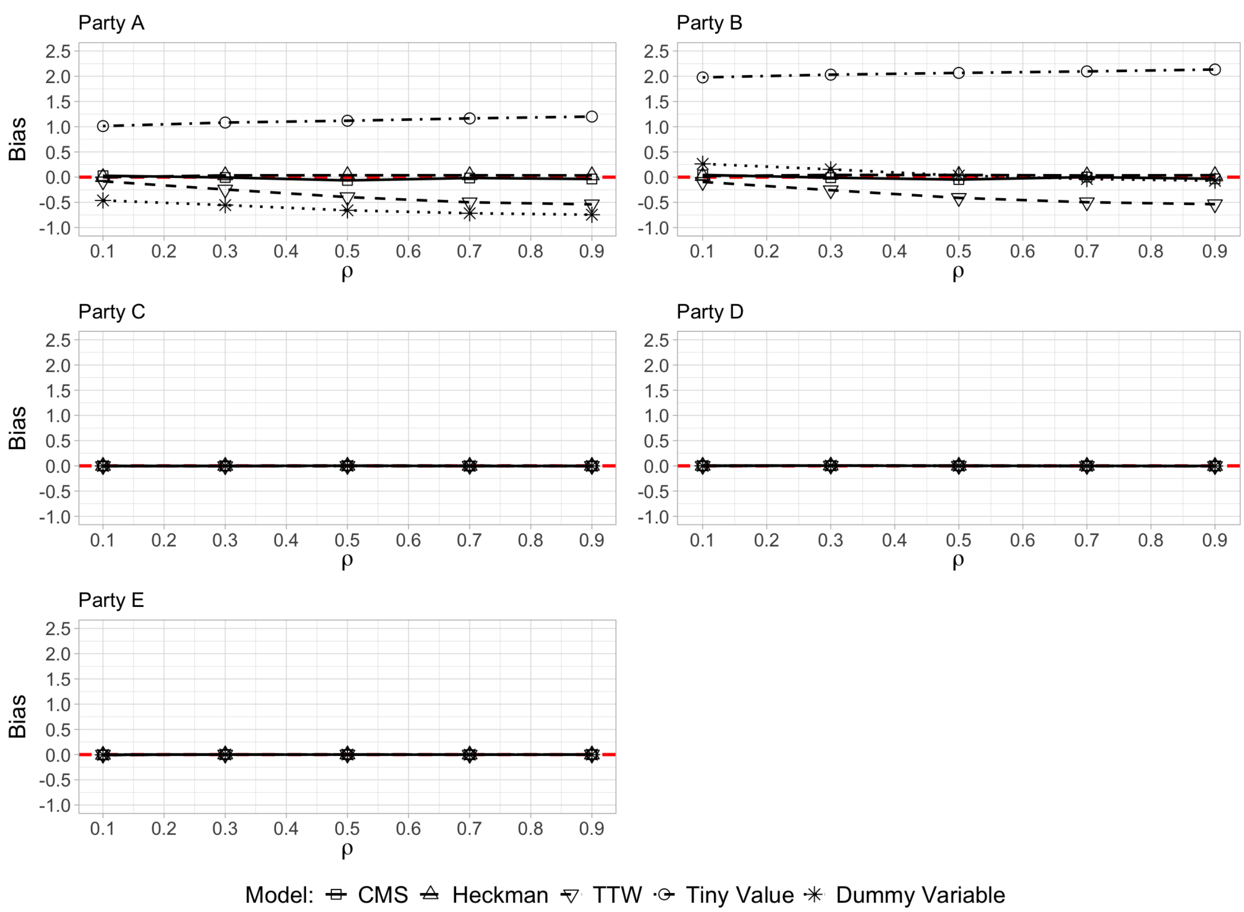


Figure S90: Bias in $\hat{\beta}_{1A}$ & $\hat{\beta}_{1B}$–Two Parties Contest 66% of Districts

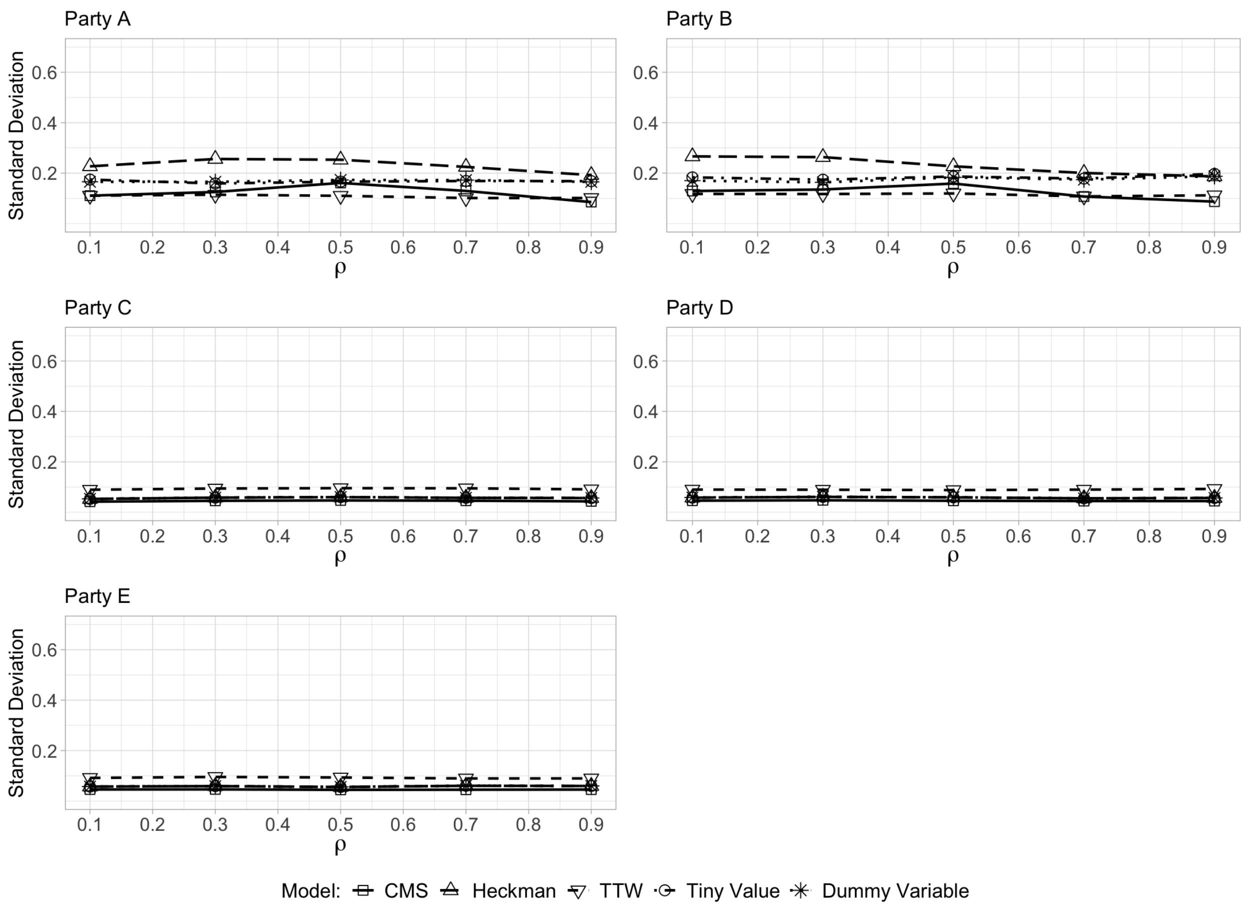


Figure S91: Standard Deviation of $\hat{\beta}_{1A}$ & $\hat{\beta}_{1B}$–Two Parties Contest 66% of Districts

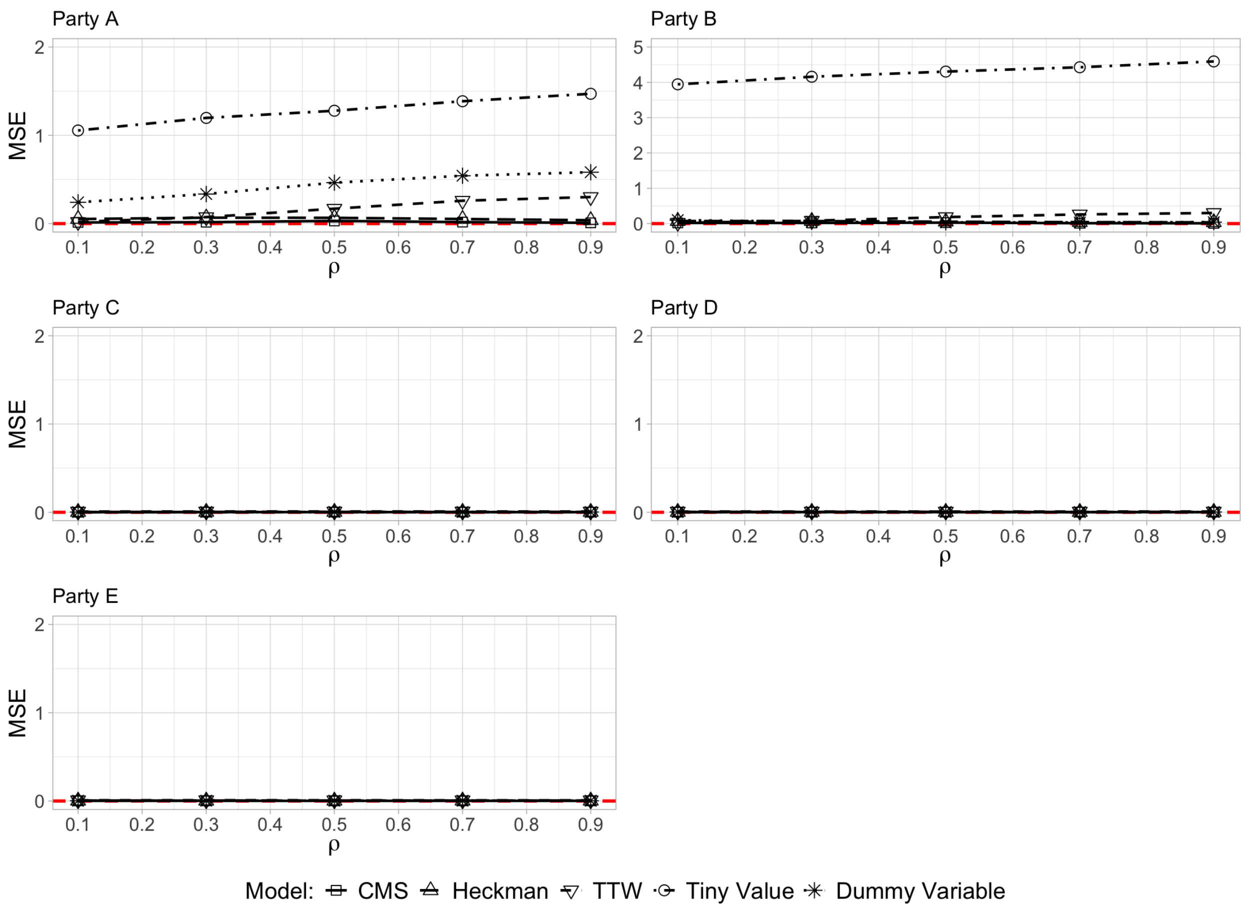


Figure S92: MSE of $\hat{\beta}_{1A}$ & $\hat{\beta}_{1B}$–Two Parties Contest 66% of Districts

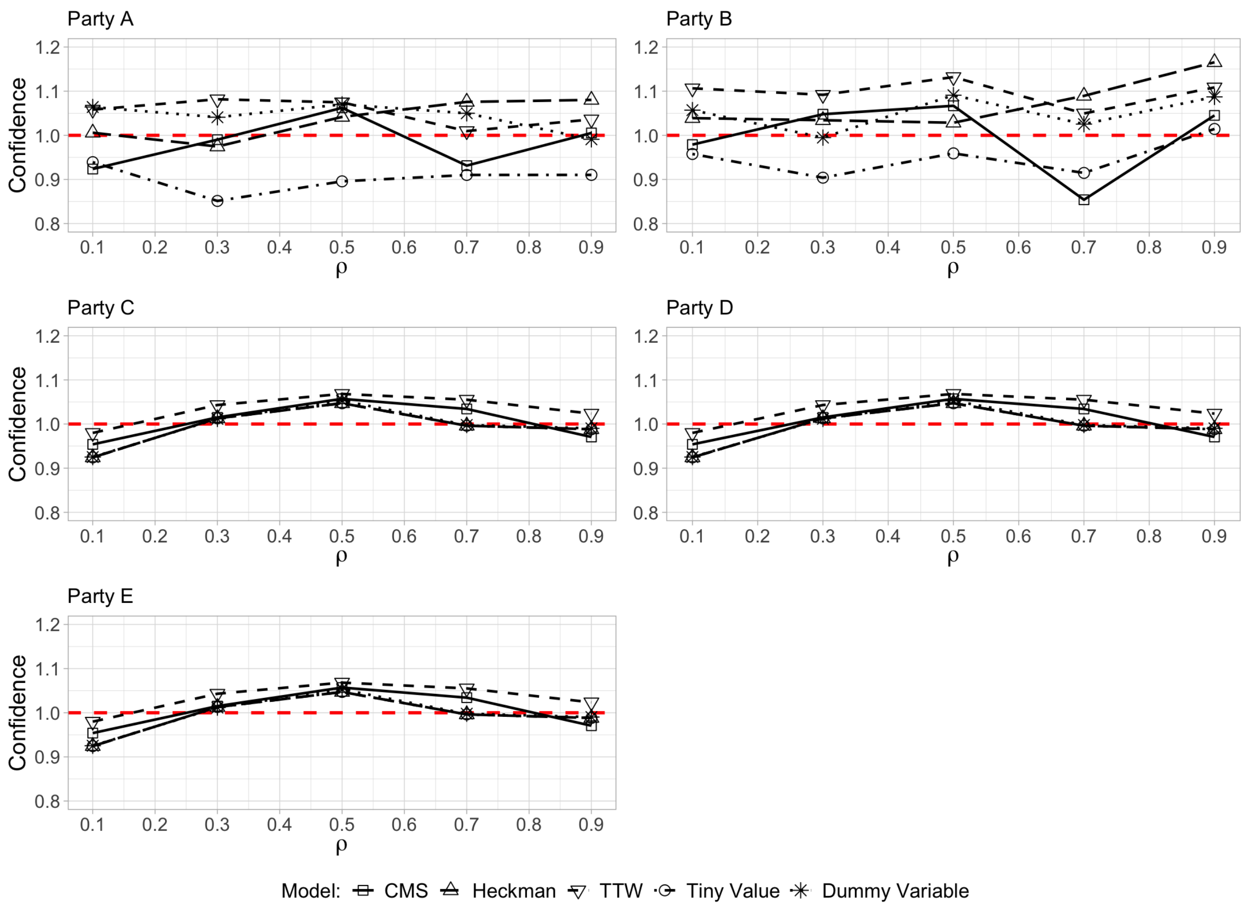


Figure S93: Confidence of $\hat{\beta}_{1A}$ & $\hat{\beta}_{1B}$–Two Parties Contest 66% of Districts

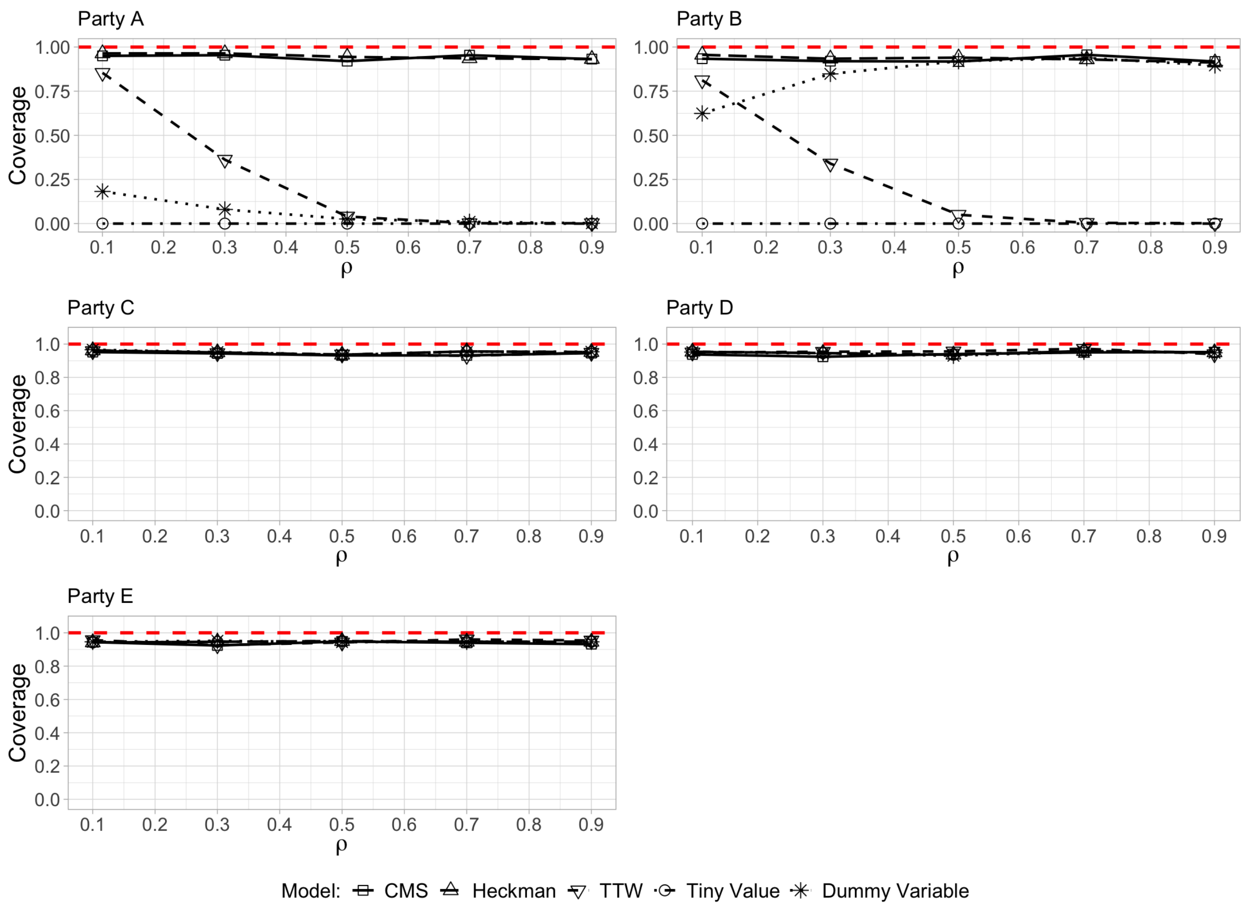


Figure S94: Coverage of $\hat{\beta}_{1A}$ & $\hat{\beta}_{1B}$–Two Parties Contest 66% of Districts

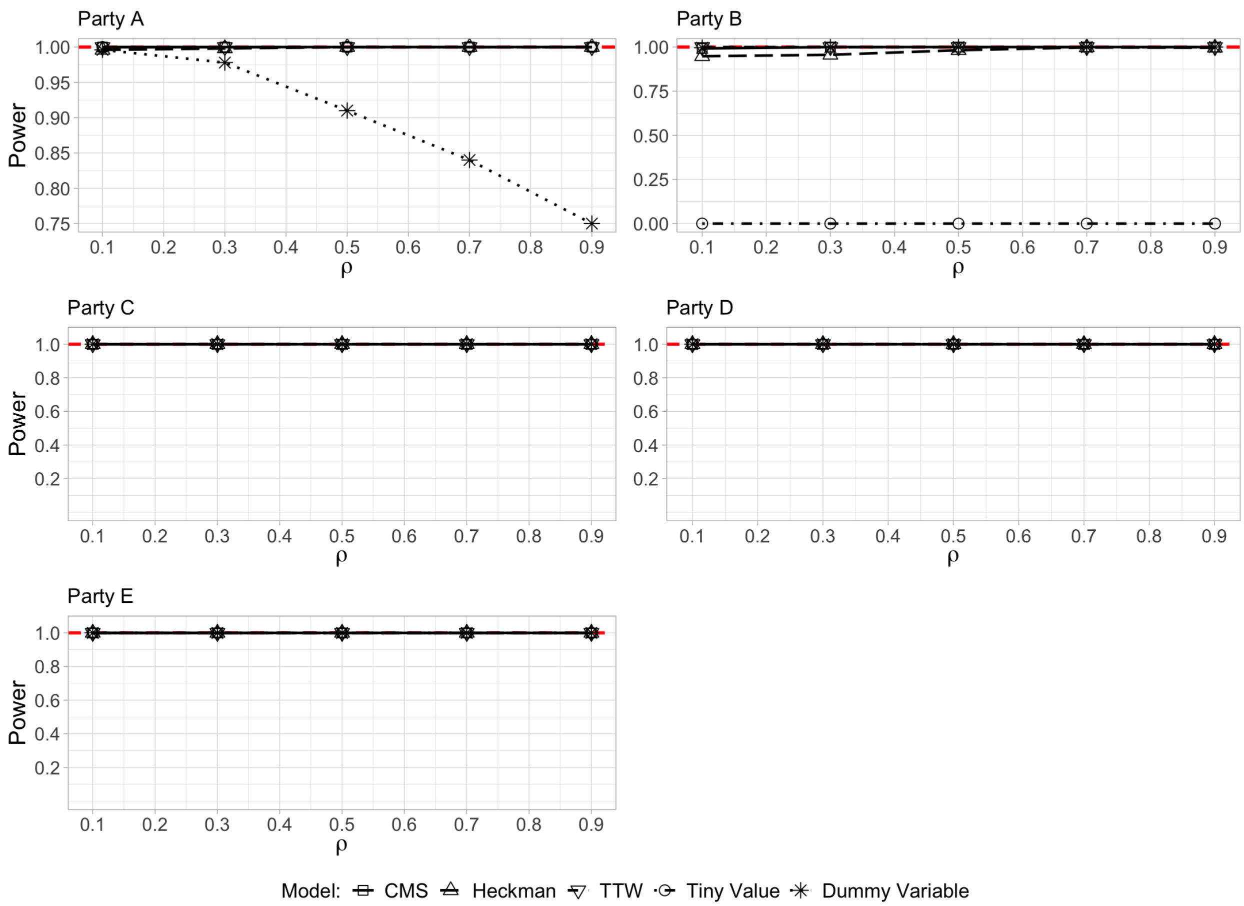


Figure S95: Power of $\hat{\beta}_{1A}$ & $\hat{\beta}_{1B}$–Two Parties Contest 66% of Districts

D.8 Correlation Across Outcome Equations

In this section, we examine the consequences of estimating models with election results as compositional outcomes while ignoring the likely correlation of error terms (u_{ij}) across the $J-1$ outcome equations ($\text{Cor}(u_{iA}, u_{iB}) \neq 0$). Our current CMS does not account for this cross-equation correlation, but we plan to incorporate this feature in future versions. Ignoring such correlation can lead to inefficient estimates. However, this issue is attenuated when the same (or very similar) set of parameters is estimated across all $J-1$ equations (Greene, 2017, p. 333)—a common practice in models of election results. In our empirical application, for example, all outcome equations include the same explanatory variables except for the lagged dependent variable. As we show below, this specification produces standard errors very similar to those from seemingly unrelated regressions (SUR). To demonstrate this, we simulate the following data:

$$s_{iA} = \beta_{0A} + \beta_{1A}x_{1A} + \beta_{2A}x_{2A} + \beta_{3A}x_{3A} + u_{iA}, \text{ (Party A)}$$
$$s_{iB} = \beta_{0B} + \beta_{1B}x_{1B} + \beta_{2B}x_{2B} + \beta_{3B}x_{3B} + u_{iB}, \text{ (Party B)}$$

where $\beta_{0A} = \beta_{0B} = \beta_{1A} = 1$, $\beta_{2A} = 2$, $\beta_{3A} = 3$, $\beta_{1B} = 3$, $\beta_{2B} = -1$, and $\beta_{3B} = -2$, $\text{Cor}(u_{iA}, u_{iB})$ varies between -0.9 and 0.9, and we also vary whether the predictors are equal in both equations according to the following scenarios:

1. $x_{1A} = x_{1B}$, $x_{2A} = x_{2B}$, and $x_{3A} = x_{3B}$.

2. $x_{1A} = x_{1B}$, $x_{2A} = x_{2B}$, and $x_{3A} \neq x_{3B}$.

3. $x_{1A} \neq x_{1B}$, $x_{2A} \neq x_{2B}$, and $x_{3A} \neq x_{3B}$.

In each scenario, we run 500 simulations with 500 observations per trial, estimating equations s_{iA} and s_{iB} separately using OLS and jointly using SUR. We compare these two approaches in terms of standard deviation (efficiency), confidence, bias, and MSE. As the results show, the strategy that ignores the correlation across equations (OLS) performs as well as the SUR approach in the first scenario, when all predictors are the same. In the second scenario, where one predictor differs across equations, OLS is slightly less efficient in estimating $\hat{\beta}_{1A}$, $\hat{\beta}_{2A}$, $\hat{\beta}_{1B}$, and $\hat{\beta}_{2B}$, but differences in standard deviations of both strategies is negligible. Inefficiency in $\hat{\beta}_{3A}$ and $\hat{\beta}_{3B}$ increases with the absolute value of the correlation when $x_{3A} \neq x_{3B}$ (scenario two) and for all estimates when all predictors differ (scenario three). Overall, these results suggest that correlation across equations does not lead to substantial inefficiency in the CMS when researchers use a similar set of predictors in all $J - 1$ equations.

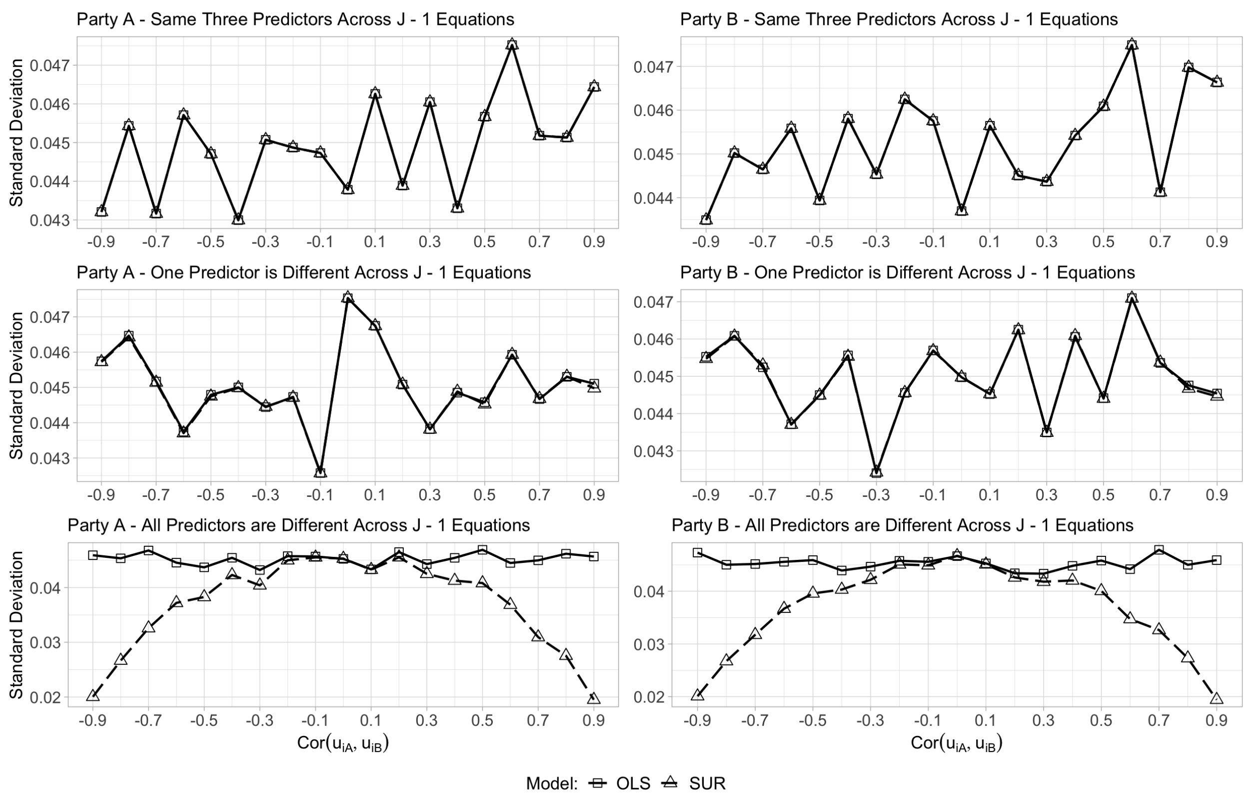


Figure S96: Standard Deviation of $\hat{\beta}_{1A}$ & $\hat{\beta}_{1B}$

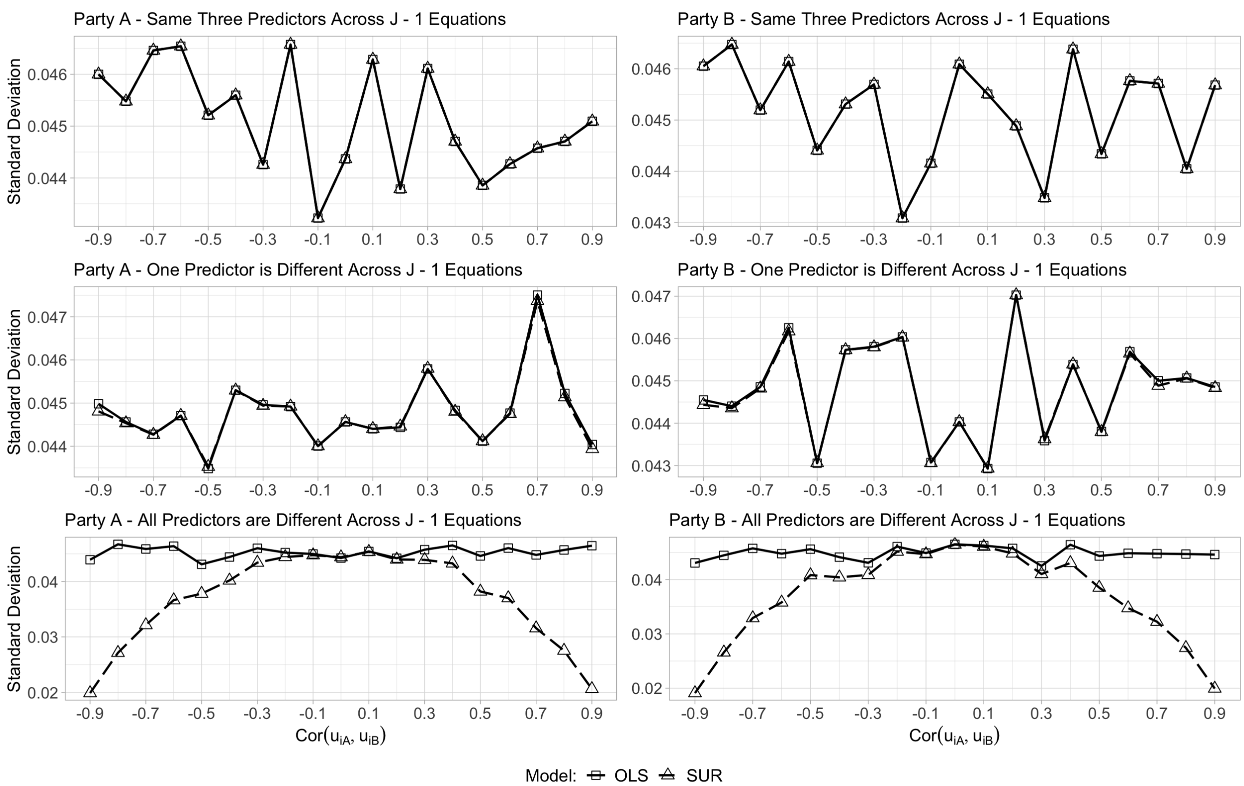


Figure S97: Standard Deviation of $\hat{\beta}_{2A}$ & $\hat{\beta}_{2B}$

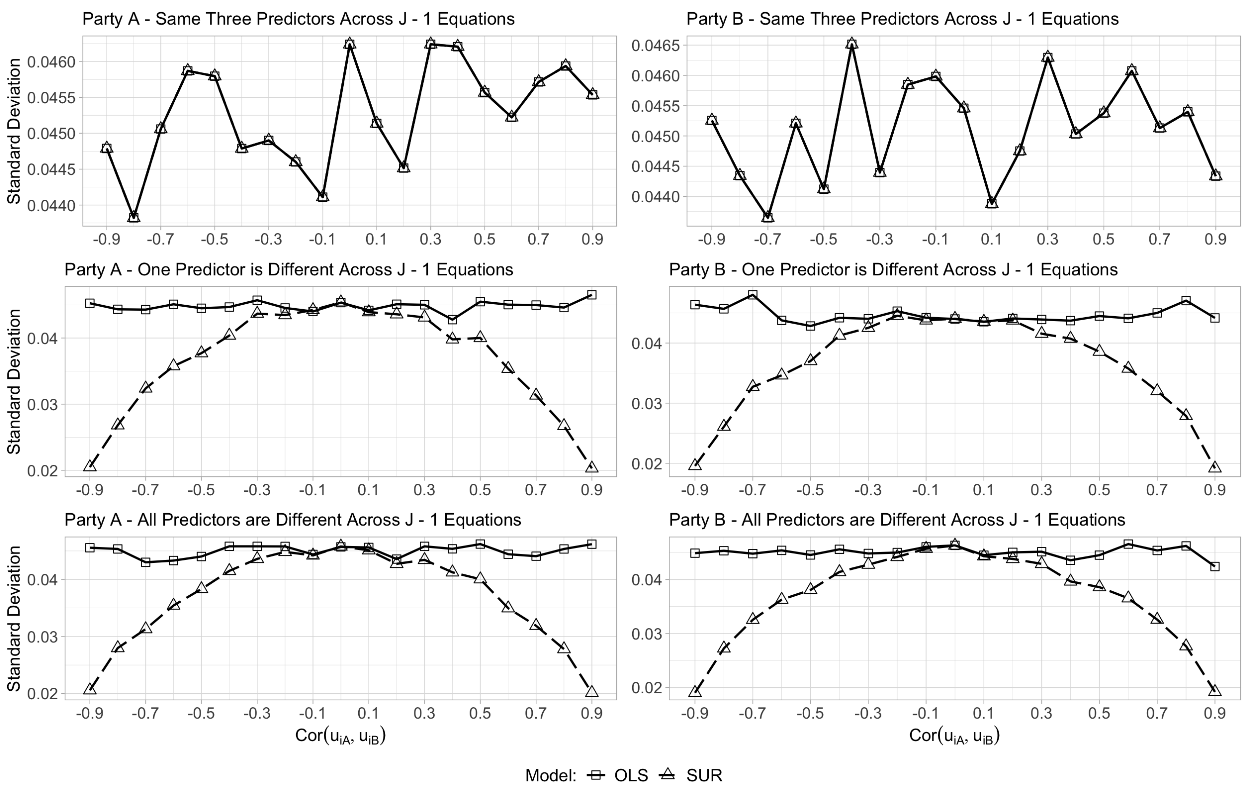


Figure S98: Standard Deviation of $\hat{\beta}_{3A}$ & $\hat{\beta}_{3B}$

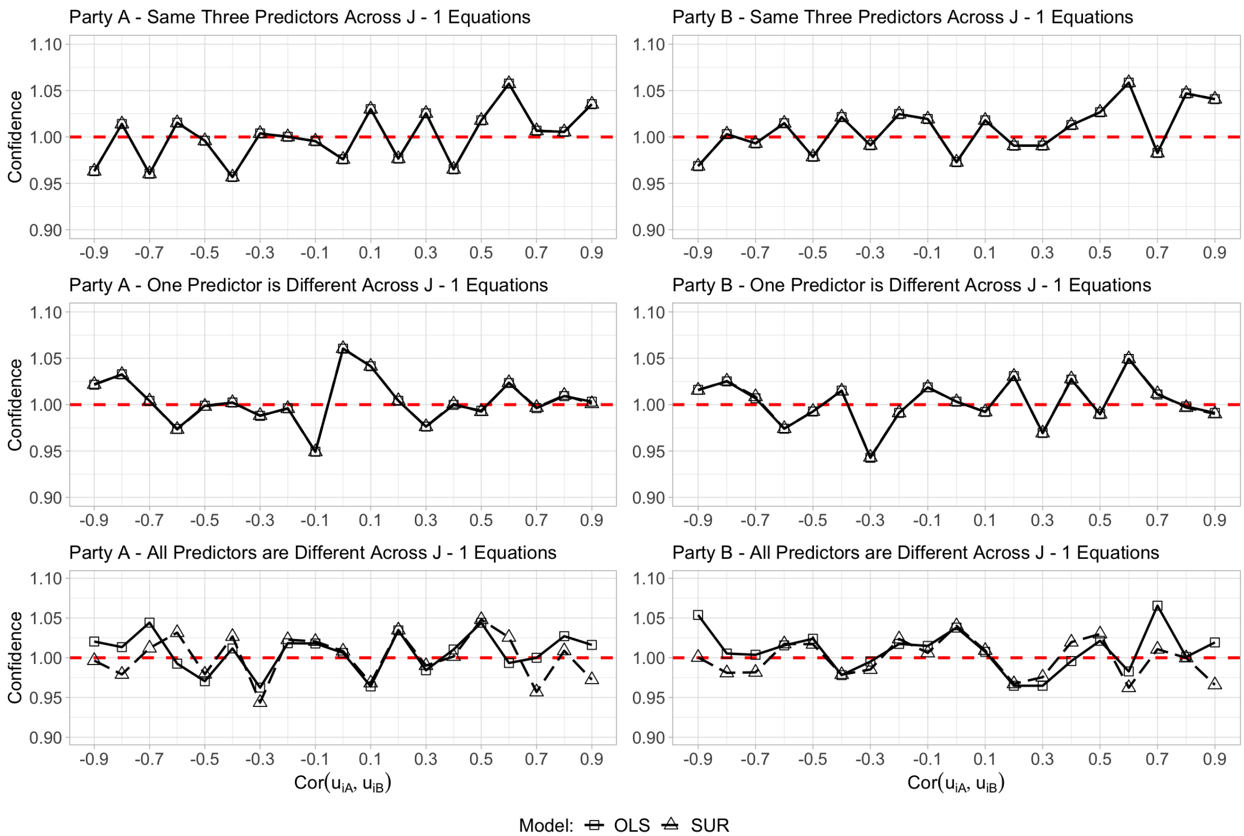


Figure S99: Confidence of $\hat{\beta}_{1A}$ & $\hat{\beta}_{1B}$

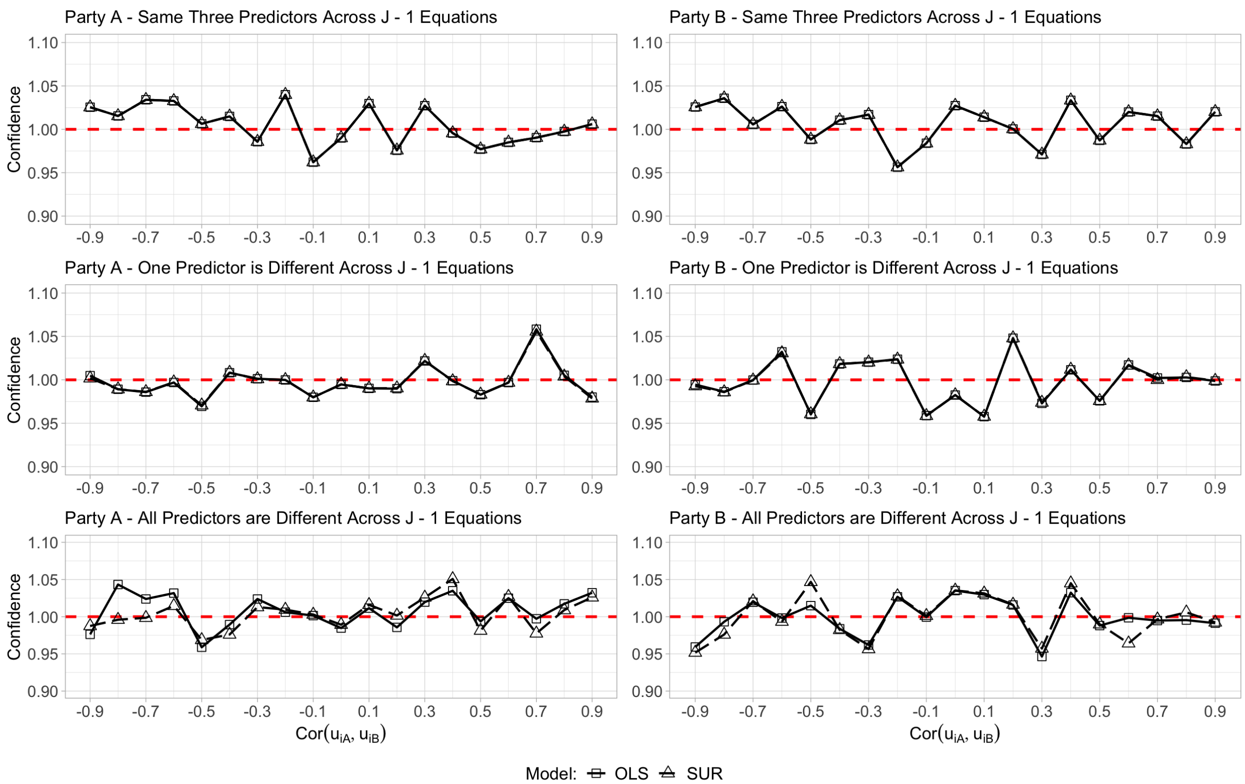


Figure S100: Confidence of $\hat{\beta}_{2A}$ & $\hat{\beta}_{2B}$

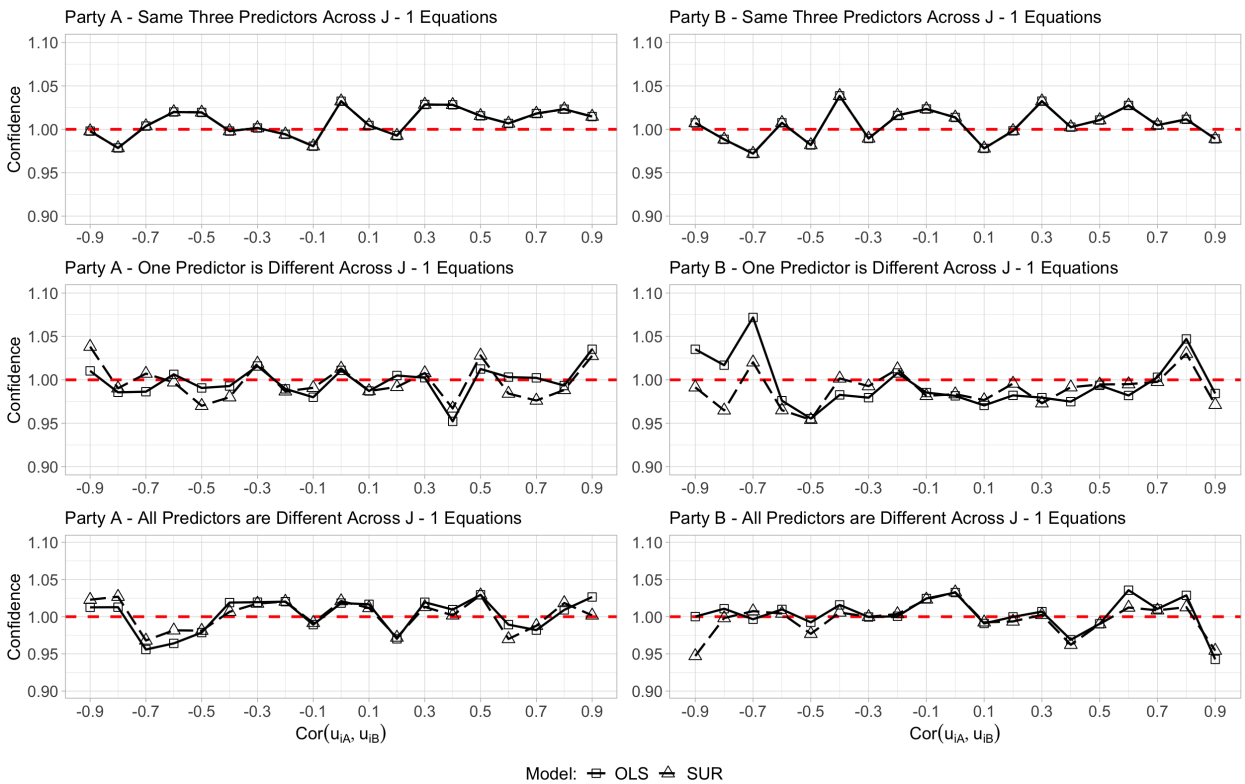


Figure S101: Confidence of $\hat{\beta}_{3A}$ & $\hat{\beta}_{3B}$

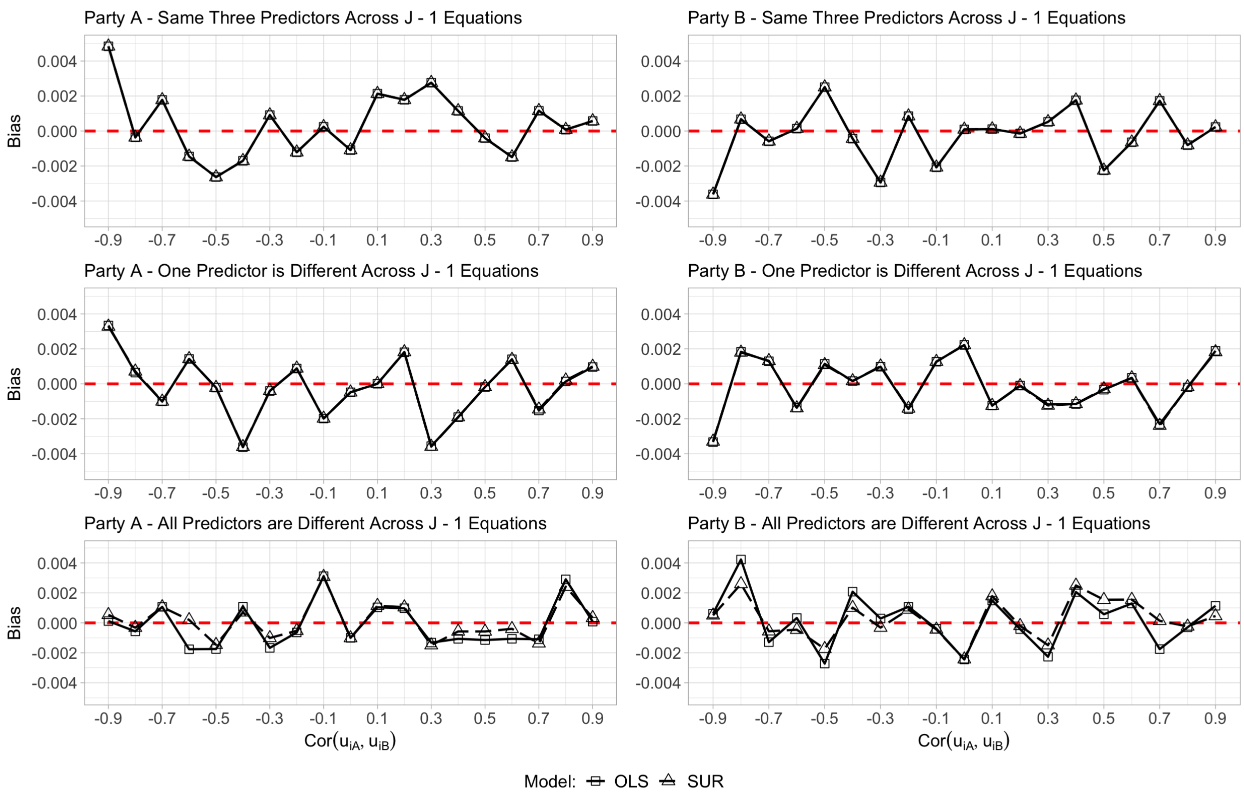


Figure S102: Bias in $\hat{\beta}_{1A}$ & $\hat{\beta}_{1B}$

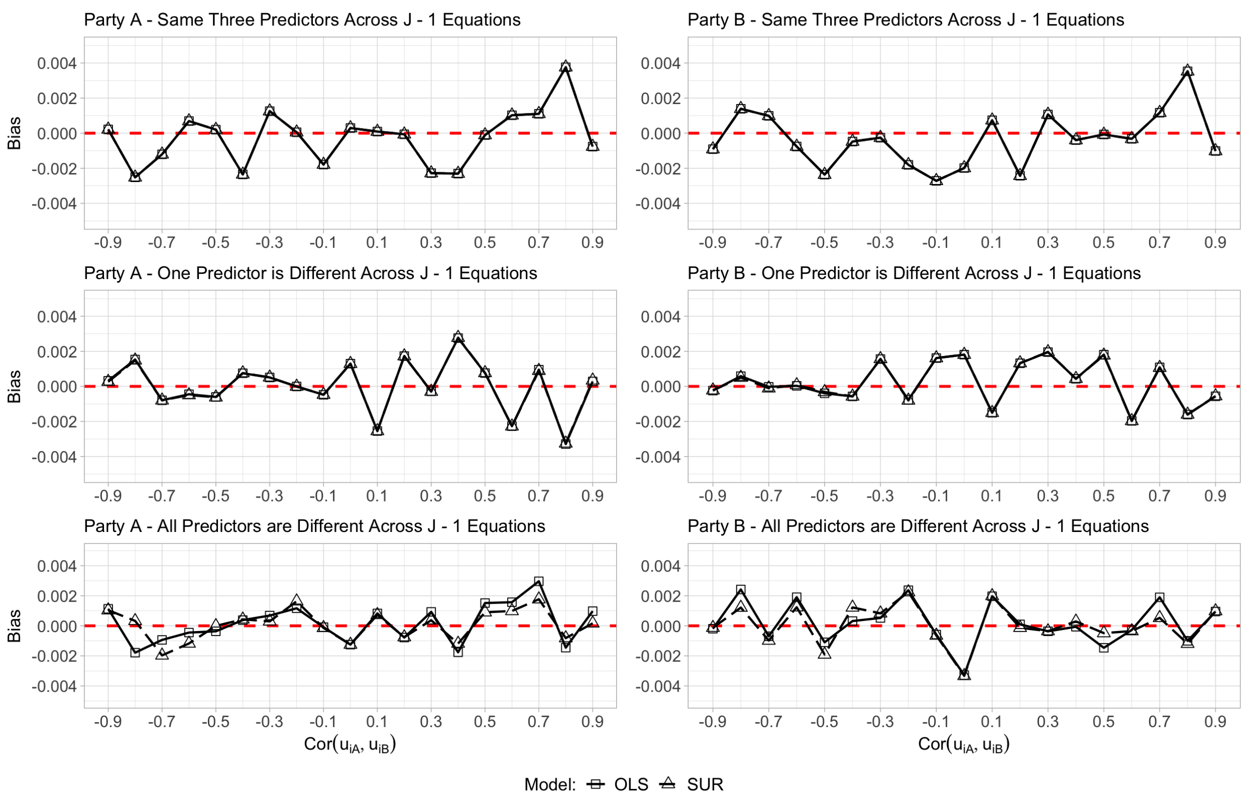


Figure S103: Bias in $\hat{\beta}_{2A}$ & $\hat{\beta}_{2B}$

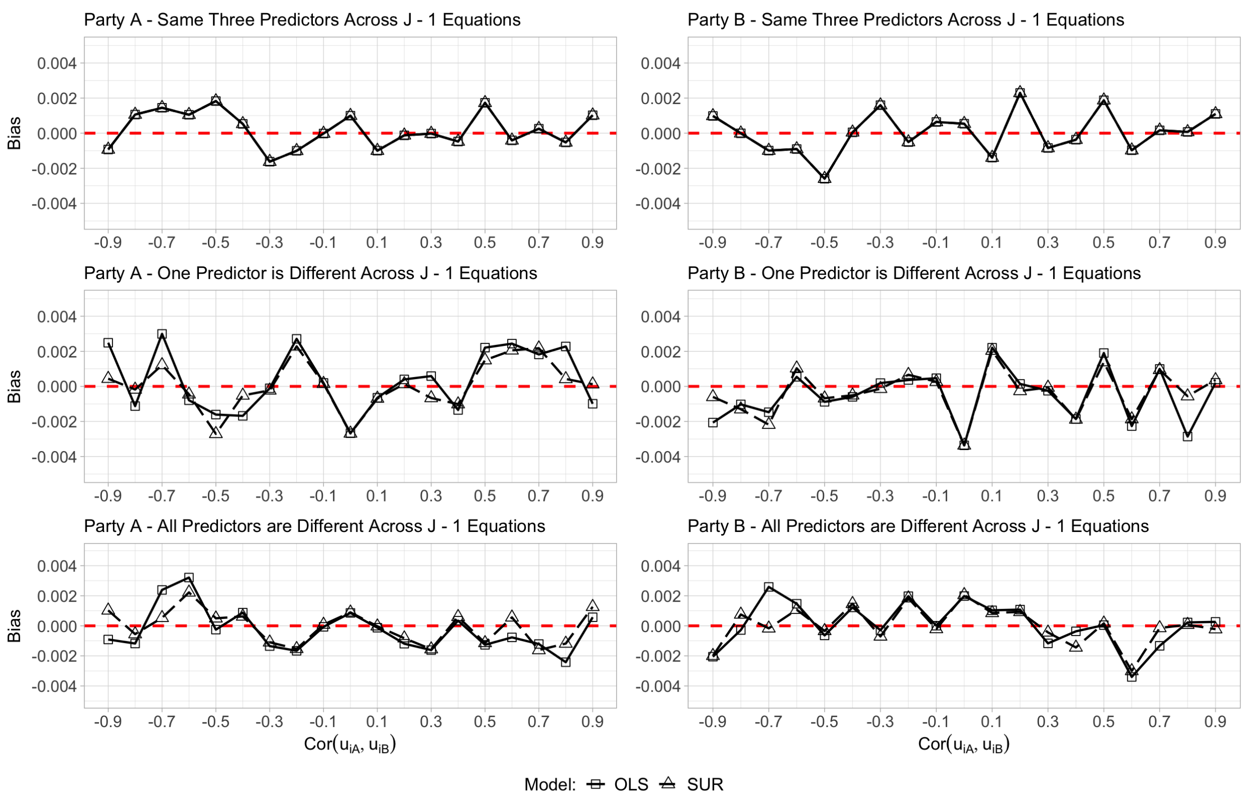


Figure S104: Bias in $\hat{\beta}_{3A}$ & $\hat{\beta}_{3B}$

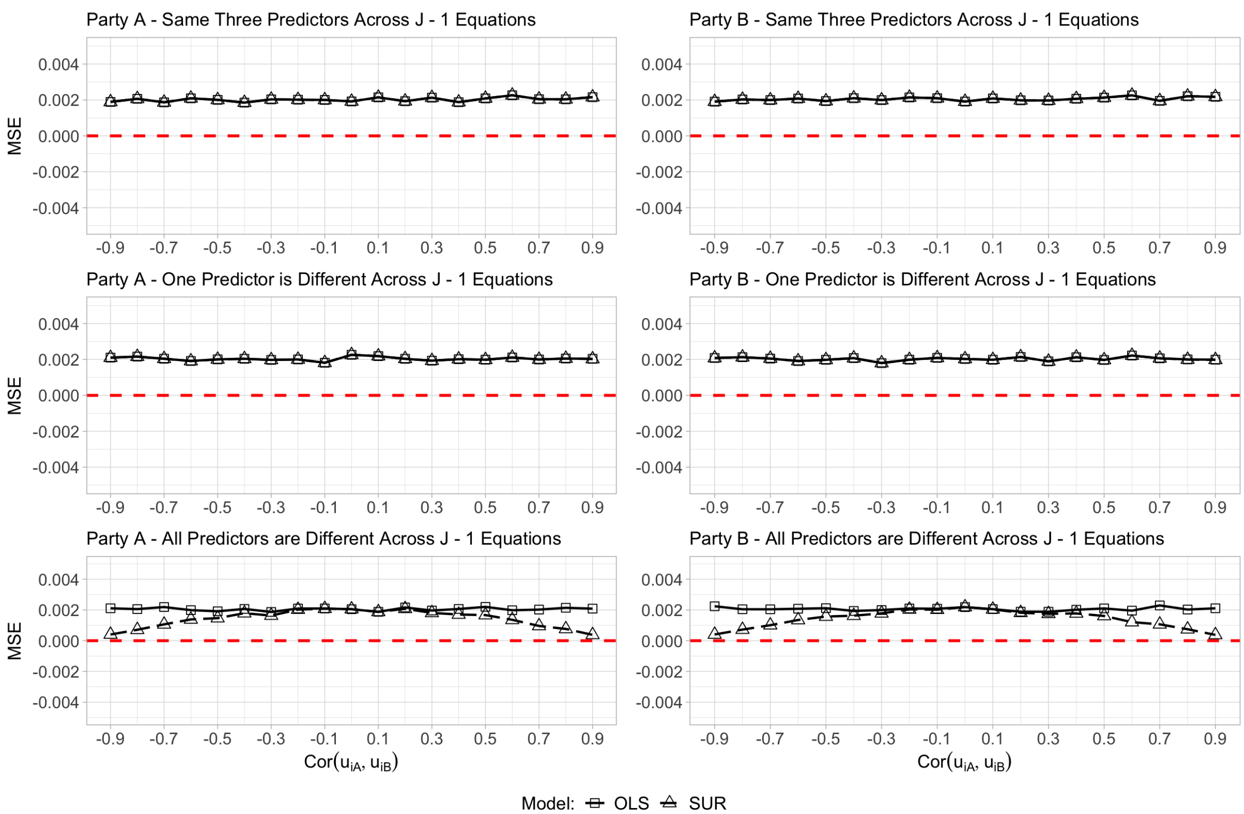


Figure S105: MSE of $\hat{\beta}_{1A}$ & $\hat{\beta}_{1B}$

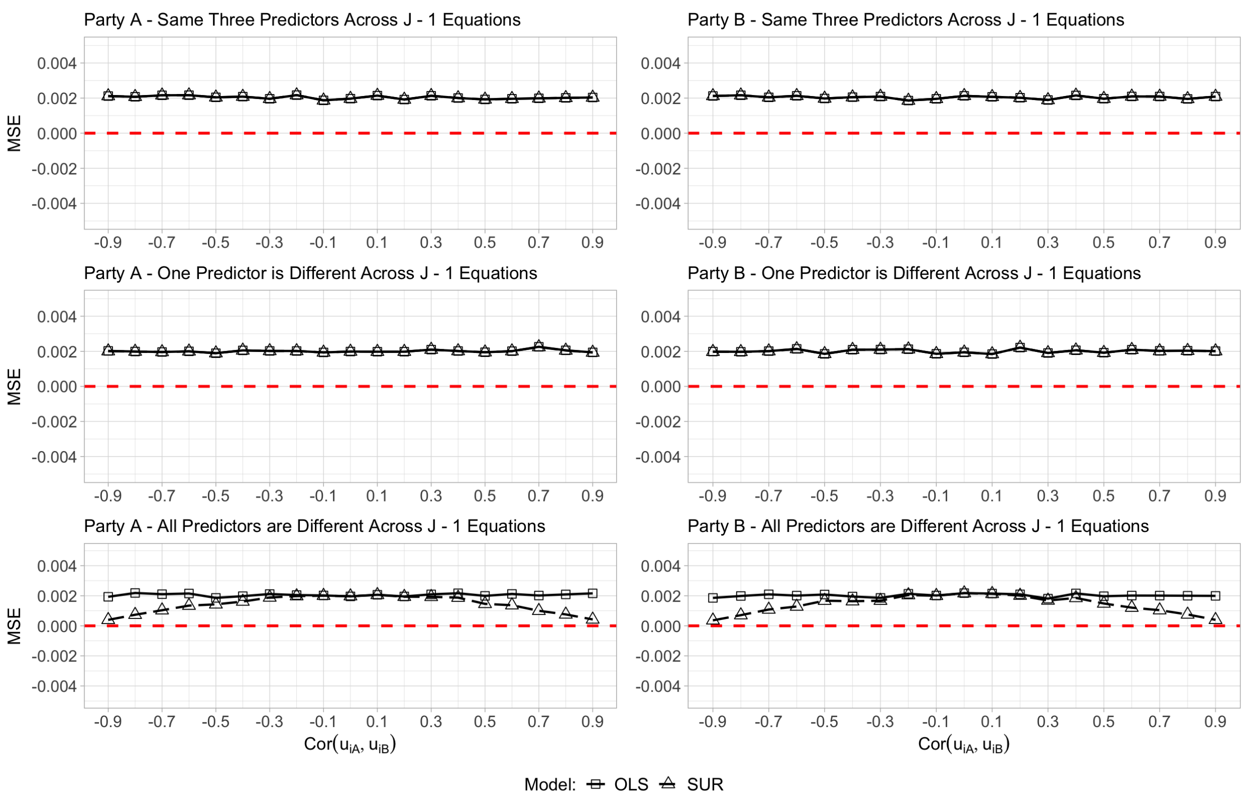


Figure S106: MSE of $\hat{\beta}_{2A}$ & $\hat{\beta}_{2B}$

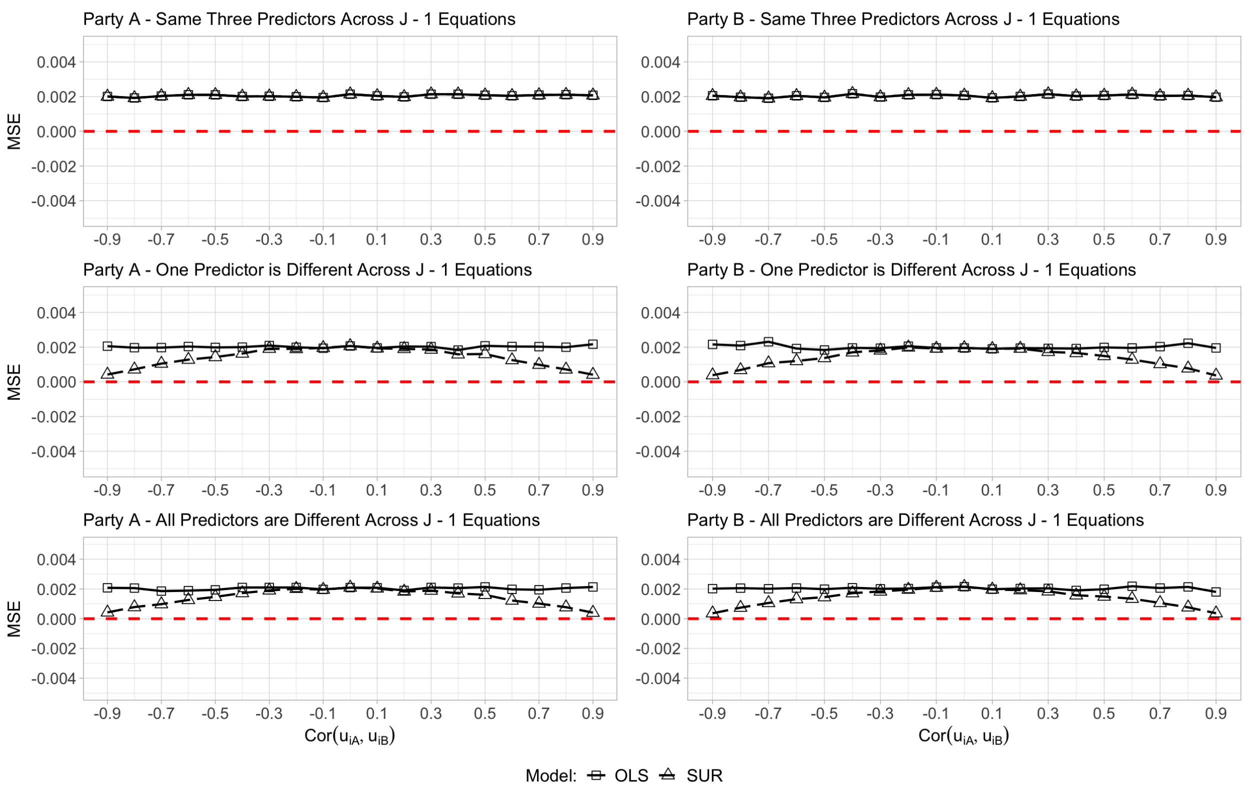


Figure S107: MSE of $\hat{\beta}_{3A}$ & $\hat{\beta}_{3B}$

D.9 Normal DGP, Invalid Exclusion Restriction, & Type-I Assumption

In this section, we show results from a Monte Carlo experiment in which the error terms in the selection and outcome stages (u_{ij} and ε_{ij}) are jointly normally distributed, z_{iA} affects both d_{iA} and s_{iA}, violating the exclusion restriction, and the CMS strategy assumes that u_{ij} and ε_{ij} are jointly extreme value type-I distributed. Party A does not contest 33% of the 500 districts. The results show that our model outperforms other strategies even under the wrong distributional assumption and an invalid instrument in terms of bias, MSE, and coverage. Under these conditions, the Heckman approach has the worst performance across all the six performance statistics.

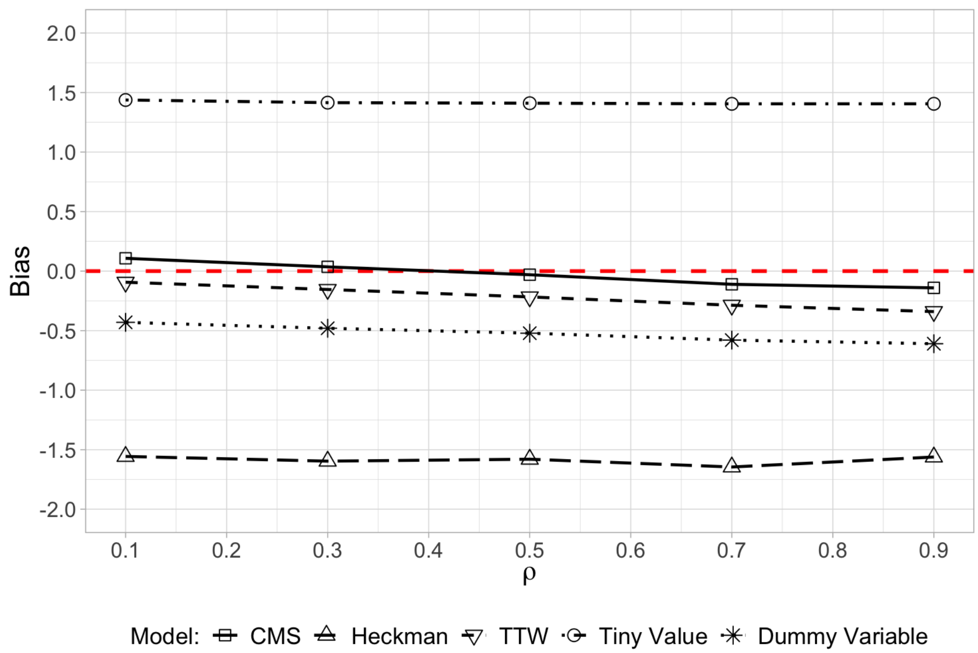


Figure S108: Bias in $\hat{\beta}_{1A}$

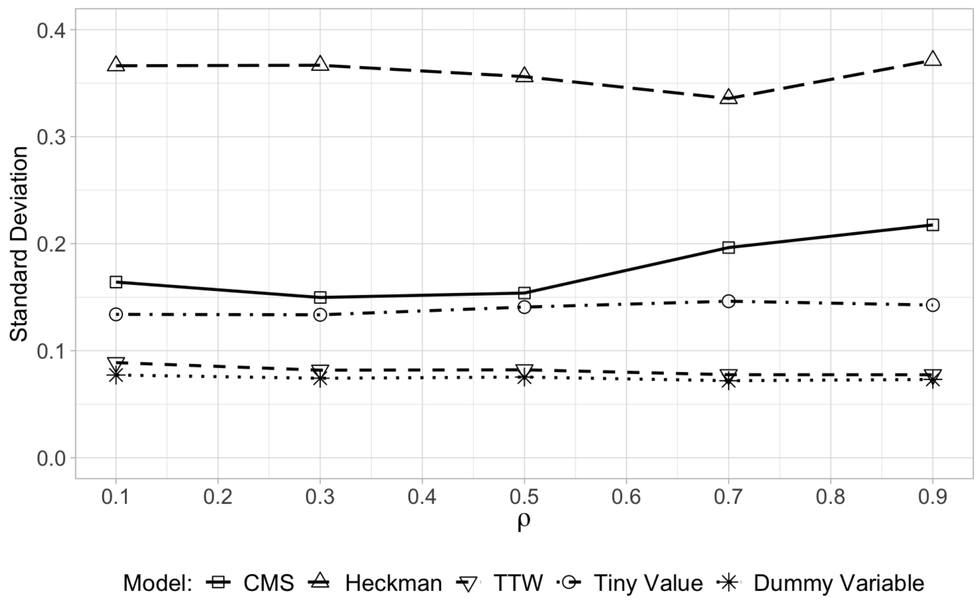


Figure S109: Standard Deviation of $\hat{\beta}_{1A}$

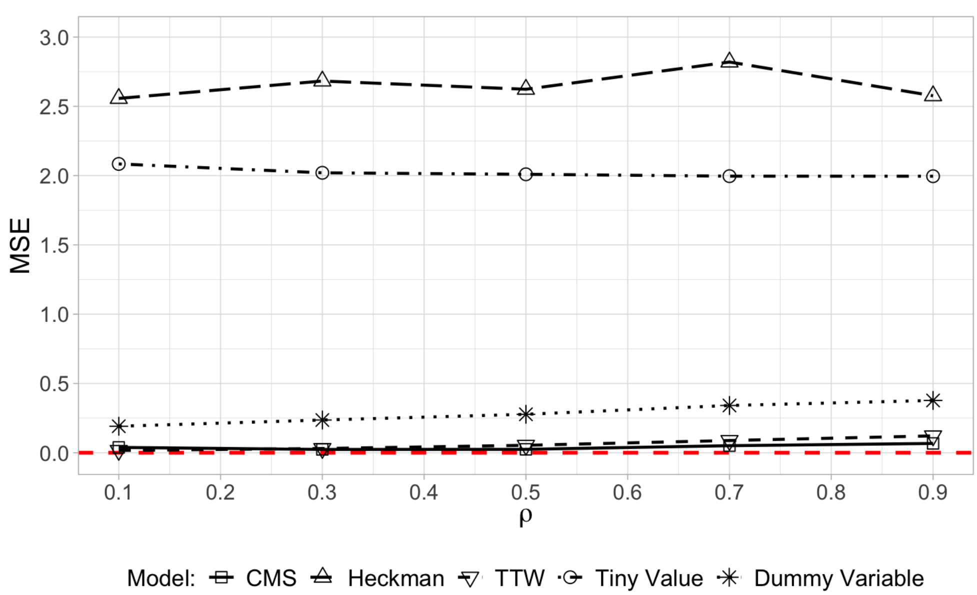


Figure S110: MSE of $\hat{\beta}_{1A}$

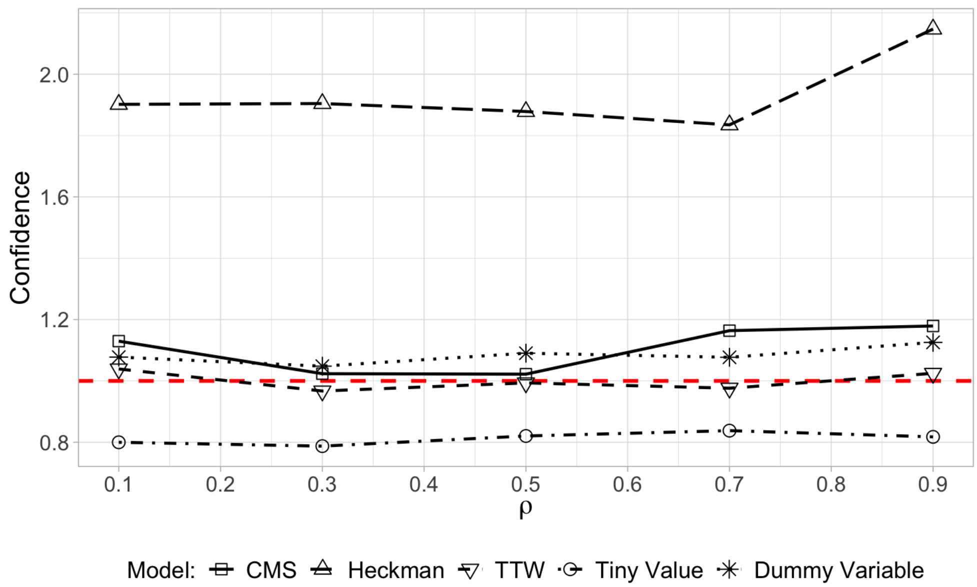


Figure S111: Confidence of $\hat{\beta}_{1A}$

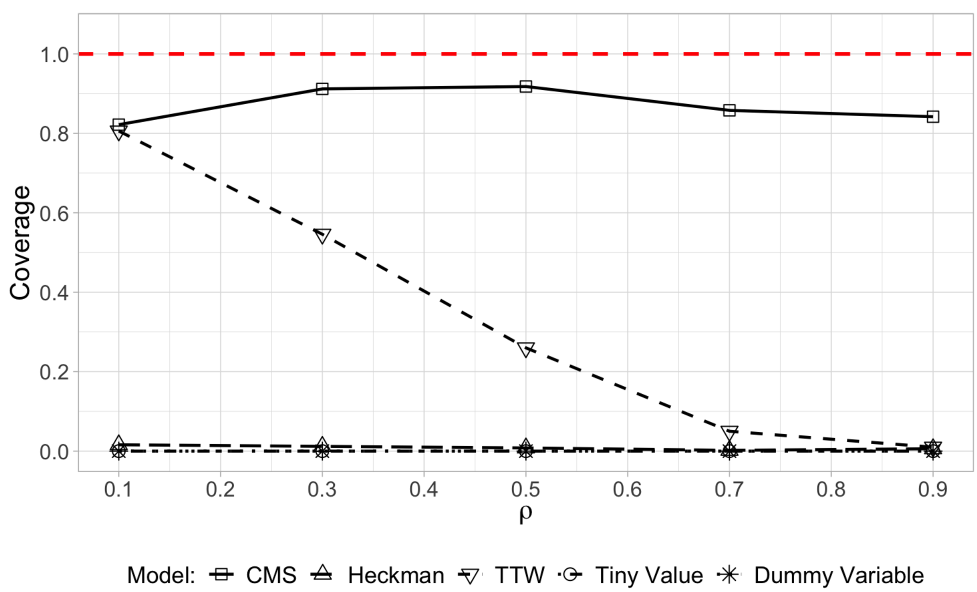


Figure S112: Coverage of $\hat{\beta}_{1A}$

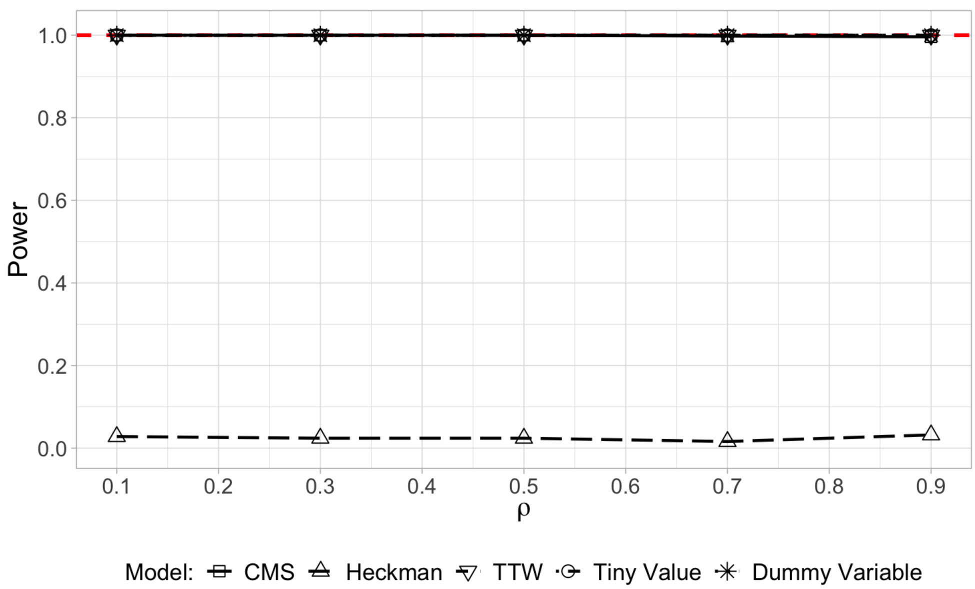


Figure S113: Power of $\hat{\beta}_{1A}$

E Appendix E: R Code With Instructions

```r
# ==========================
# Overview
# ==========================
# There are three sections. The outline is shown below:
# Section 1 contains helper functions used in estimation.
# Section 2 is the main function that computes maximum likelihood estimates of a
↪  two-stage model.
# Section 3 is an example of two parties, where one party is partially
↪  contesting.

remove(list = ls())

library(extRemes)
library(evd)
library(pracma)
library(dplyr)

# ==========================
# Section A: Helper Functions
# ==========================
# These functions support the estimation routine by defining distributional
↪  components
# and numerical integration methods.

# --------------------------
# XbetaFunc
# --------------------------
XbetaFunc = function(data, beta){
        # **Purpose:** Compute the linear predictor X*beta
        # **Input:**
        #   - data: a row or vector of covariates
```

```
        #   - beta: coefficient vector
        # **Output:** numeric value of X*beta
        as.numeric(data) %*% beta
}

# --------------------------
# jointevd
# --------------------------
jointevd = function(x, y, m){
        # **Purpose:** Compute the joint probability density function f(x, y; m)
        # **Input:**
        #   - x, y: points to evaluate the PDF
        #   - m: parameter
        # **Notes:** Adjust x and y by the Euler constant to center them; handle
        ↪  NaN/Inf
        x = x - digamma(1)
        y = y - digamma(1)
        jointdf = exp( -(exp(-m*x)+exp(-m*y))^(1/m) ) * exp(-m*(x+y)) *
        ↪  (exp(-m*x) + exp(-m*y))^(1/m-2) *
        (m-1 + (exp(-m*x)+exp(-m*y))^(1/m))
        jointdf[is.nan(jointdf)] = 0
        jointdf[is.infinite(jointdf)] = 0
        jointdf
}

# --------------------------
# cdfevd
# --------------------------
cdfevd = function(x){
        # **Purpose:** Compute marginal cumulative distribution function F(x)
        # **Input:** x: value to evaluate marginal CDF
        # **Output:** scalar CDF value
        x = x - digamma(1)
```

```
        exp(-exp(-x))
}

# --------------------------
# int_evd
# --------------------------
int_evd = function(data, m){
        # **Purpose:** Perform numerical integration of the joint PDF using
         ↪  standard
        #   `integrate` from package stats
        # **Input:**
        #   - data: vector [x_val, y_val, indicator]
        #   - m: parameter
        # **Output:** integration result or 1 if indicator != 1
        if(data[3]==1){
                result = integrate(jointevd, data[1], Inf, y=data[2], m=m)$value
        } else {
                result = 1
        }
        result
}

# --------------------------
# int_evd_quadinf
# --------------------------
int_evd_quadinf = function(data, m){
        # **Purpose:** Alternative numerical integration using quadinf from
         ↪  package pracma
        # **Used if the standard integrate fails**
        if(data[3]==1){
                result <- try(integrate(jointevd, data[1], Inf, y=data[2],
                 ↪  m=m)$value, silent = TRUE)
```

```
            if(class(result) == "try-error") {
                    result <- quadinf(jointevd, data[1], Inf, y=data[2],
                    ↪  m=m)$Q
            }
        } else {
            result = 1
        }
        result
}


# ==========================
# Section B: Main Function
# ==========================
# This section is the core estimation, which computes maximum likelihood
↪  estimates
# for a two-stage model with sample selection.

combinedMLE = function(data, y1, x1, y2, x2, valid = FALSE){
        # **Purpose:** Compute maximum likelihood estimates (MLE) for a bivariate
        ↪  sample.
        #
        # **Inputs:**
        #   - data: a data.frame containing all observations and covariates
        #   - y1: name (string) of the dependent variable (binary) in selection
        ↪  stage
        #   - x1: vector of strings, names of covariates for y1
        #   - y2: name (string) of the dependent variable (continuous) in outcome
        ↪  stage
        #   - x2: vector of strings, names of covariates for y2
        #   - valid: logical, if TRUE computes log-likelihood, AIC, BIC along
        ↪  with parameter estimates
        #
```

```
# **Outputs:**
#   - If valid = FALSE: numeric vector of parameter estimates (beta1,
↪  beta2, correlation)
#   - If valid = TRUE: list containing
#       - par: parameter estimates (beta1, beta2, correlation)
#       - loglik: log-likelihood at MLE
#       - AIC: Akaike Information Criterion
#       - BIC: Bayesian Information Criterion
#
# **Steps:**
#   1. Initialize parameters using OLS (for both y1 and y2)
#   2. Define objective functions (negative log-likelihood)
#   3. Optimize using optim (with alternative numerical integration if
↪  needed)
#   4. Transform correlation parameter to [0,1]
#   5. Optionally compute AIC, BIC, log-likelihood

length_beta1 = length(x1)
length_beta2 = length(x2)

# Get initial values from OLS on outcome stage for selected sample
OLSStage2 <- lm(as.formula(paste(y2, "~", paste(x2[2:length(x2)],
↪  collapse= "+"))),
data=data[data[y1]==1,])

# Get initial values from OLS on selection stage and correlation
if (length(table(unique(data[y1]))) == 1){
        corr_initial = -2
        para = c(rep(0, length_beta1), OLSStage2$coefficients,
         ↪  corr_initial)
        # initial m = 2, m = e^para + 1 = 1/(1-rho), rho =
         ↪  e^para(1+e^para)
} else{
```

```
        OLSStage1 <- lm(as.formula(paste(y1, "~",
        ↪  paste(x1[2:length(x1)], collapse= "+"))), data=data)

        corr_initial = log(1/(1-abs(cor(OLSStage2$residuals,
        ↪  OLSStage1$residuals[data[y1]==1]))) -1)
        if (corr_initial < -2){
                corr_initial = -2
        } else if (corr_initial > 2){
                corr_initial = 2
        }

        # The correlation is transformed from (0,1) to all real numbers
        para = c(OLSStage1$coefficients, OLSStage2$coefficients,
        ↪  corr_initial)
        # initial m = 2, m = e^para + 1 = 1/(1-rho), rho =
        ↪  e^para(1+e^para)
}

# --------------------------
# Objective function
# --------------------------
objFunc = function(para){
        data$Xbeta1_ = - apply(data[x1], 1, XbetaFunc, beta =
        ↪  para[1:length_beta1])
        data$Xbeta2 = apply(data[x2], 1, XbetaFunc,
        beta = para[(length_beta1+1):(length_beta1+length_beta2)])
        m = exp(para[length(para)]) + 1


        obj1 = (1-data[y1]) * log(cdfevd(data$Xbeta1_))
        obj1[data[y1] == 1] = 0
        obj1[obj1 < -600] = -600
```

```
        data$u = data[,y2] - data$Xbeta2
        obj2 = log(apply(data[c("Xbeta1_","u",y1)], 1, int_evd, m = m))
        obj2[obj2 < -600] = -600
        return(-sum(obj1+obj2))
}

# -------------------------
# Objective function with quadinf alternative
# -------------------------
objFunc_quadinf = function(para){
        data$Xbeta1_ = - apply(data[x1], 1, XbetaFunc, beta =
        ↪ para[1:length_beta1])
        data$Xbeta2 = apply(data[x2], 1, XbetaFunc,
        beta = para[(length_beta1+1):(length_beta1+length_beta2)])
        m = exp(para[length(para)]) + 1

        obj1 = (1-data[y1]) * log(cdfevd(data$Xbeta1_))
        obj1[data[y1] == 1] = 0
        obj1[obj1 < -600] = -600

        data$u = data[,y2] - data$Xbeta2
        obj2 = log(apply(data[c("Xbeta1_","u",y1)], 1, int_evd_quadinf,
        ↪ m = m))
        obj2[obj2 < -600] = -600
        return(-sum(obj1+obj2))
}

# -------------------------
# Optimization
# -------------------------
result = try(optim(para, objFunc), silent = TRUE)

if(class(result) == "try-error") {
```

```
        result <- optim(para, objFunc_quadinf)
}

# Retry with lower and upper correlation bounds
para[length(para)] = -2

result1 = try(optim(para, objFunc), silent = TRUE)

if(class(result1) == "try-error") {
        result1 <- optim(para, objFunc_quadinf)
}

para[length(para)] = 2

result2 = try(optim(para, objFunc), silent = TRUE)

if(class(result2) == "try-error") {
        result2 <- optim(para, objFunc_quadinf)
}

# Select the best optimization result
if(result1$value < result$value){
        result = result1
}

if(result2$value < result$value){
        result = result2
}

# Transform correlation parameter back to (0,1)
output = result$par
output[length(para)] = 1 - 1/(1+exp(output[length(para)]))
```

```
        # Optionally compute log-likelihood, AIC, BIC
        if (valid) {
                ll <- -result$value
                k <- length(result$par)
                n <- nrow(data)
                print(k)
                print(n)

                aic <- 2 * k - 2 * ll
                bic <- log(n) * k - 2 * ll

                return(list(
                par = output,
                loglik = ll,
                AIC = aic,
                BIC = bic
                ))
        } else {
                return(output)
        }

}
```

```
# =========================
# Section C: Example
# =========================
# In this example, party A is partially contesting with probablity 0.33,
# and party B is the baseline party that contests everywhere.

set.seed(2025)
a <- c(1, 1.2)           # coefficients for party A in the outcome equation
v <- c(1,-1)             # Coefficients for selection equation (excluding
↪  intercept)

n = 100                  # Sample size
censoring_prob = 0.33    # Target probability of selection=0 (censoring)
correlation = 0.5        # Correlation parameter for extreme value errors

# -------------------------
# Selection equation intercept
# -------------------------
v01 <- -(1+v[1]^2 + v[2]^2)^0.5*qnorm(censoring_prob)
# **Purpose:** Sets the intercept of the selection equation so that the
# probability of being selected (yA=1) matches censoring_prob.
# - sqrt(1 + sum(v^2)) adjusts for the scale of predictors
# - qnorm maps probability to standard normal quantile
v1 <- c(v01, v) # Complete coefficient vector for selection equation (intercept +
↪  slopes)

# -------------------------
# Generate covariates
# -------------------------
z1A = rnorm(n)
z2A = rnorm(n)
data = data.frame(z1A, z2A)
```

```
data$x1A = data$z1A

XA <- cbind(1, data$x1A)


# -------------------------
# Extreme value errors
# -------------------------
corr_err = rmvevd(n, dep = 1 - correlation)

data$e1 <- corr_err[, 1] + digamma(1)
data$u1 <- corr_err[, 2] + digamma(1)
data$const = 1


# -------------------------
# Binary outcome (selection)
# -------------------------
data$yA = v1[1] + v1[2]*data$z1A + v1[3]*data$z2A + data$e1
data$yA = data$yA > 0

# Check actual censoring rate in simulated data
censoring_rate = sum(1 - data$yA)/n



# -------------------------
# Second stage: compute choice probabilities (logit)
# -------------------------
data <- data %>%
mutate(denom = 1 + exp(XA%*%a + data$u1)*data$yA)

data <- data %>%
```

```
mutate(vsA = ifelse(data$yA == 0, 0, exp(XA%*%a + data$u1)/denom)) %>%
mutate(vsB =                          1/denom)

data$sA <- log(data$vsA/data$vsB)
data$sA[data$sA == -Inf] = 0



# --------------------------
# Combined MLE estimation
# --------------------------
# Estimate parameters using the combinedMLE function
paraA = combinedMLE(
data,
"yA",
c("const", "z1A", "z2A"),
"sA",
c("const", "x1A")
)
print(paraA)
paraA_valid = combinedMLE(
data,
"yA",
c("const", "z1A", "z2A"),
"sA",
c("const", "x1A"),
valid = TRUE
)

print(paraA_valid$par)
print(paraA_valid$loglik)
print(paraA_valid$AIC)
print(paraA_valid$BIC)
```

References

Greene, William H. 2017. *Econometric Analysis*. Eight ed. New York, NY: Pearson.

Hopkins, Vincent, Ali Kagalwala, Andrew Q Philips, Mark Pickup and Guy D Whitten. 2024. "How do we know what we know? Learning from Monte Carlo simulations." *The Journal of Politics* 86(1):36–53.

McFadden, Daniel. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, ed. Paul Zarembka. New York, NY: Academic Press pp. 105–142.

Train, Kenneth E. 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press.