Supplementary Information: Estimating Substantive Effects in Binary Outcome Panel Models: A Comparison (Online Only)

Casey Crisman-Cox*

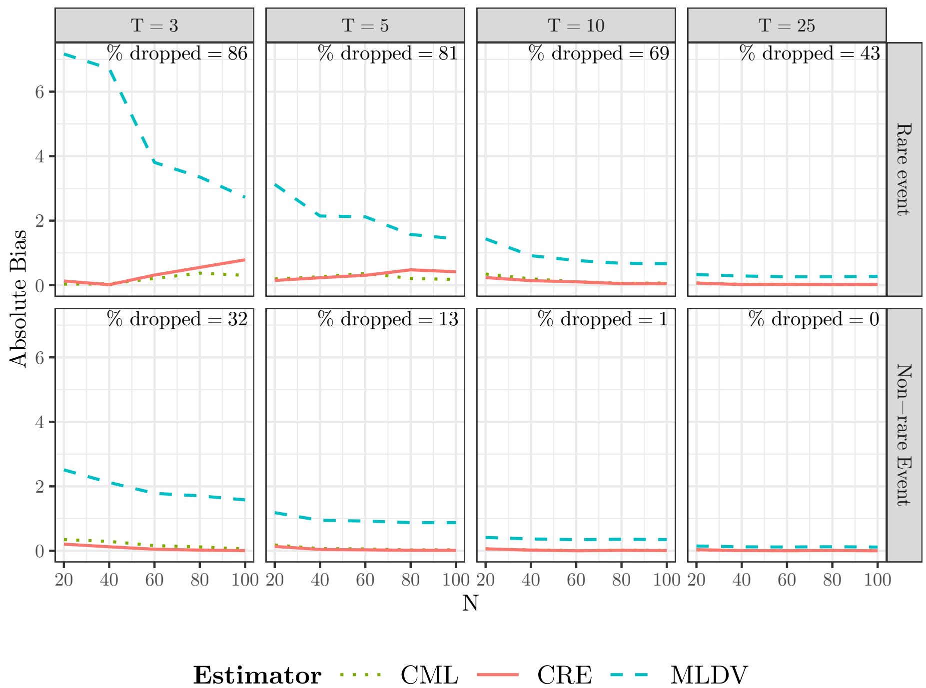
August 22, 2019

Contents

*Texas A&M. Email: c.crisman-cox@tamu.edu

A Main simulations: Bias results

Figure A.1: Bias in estimating β

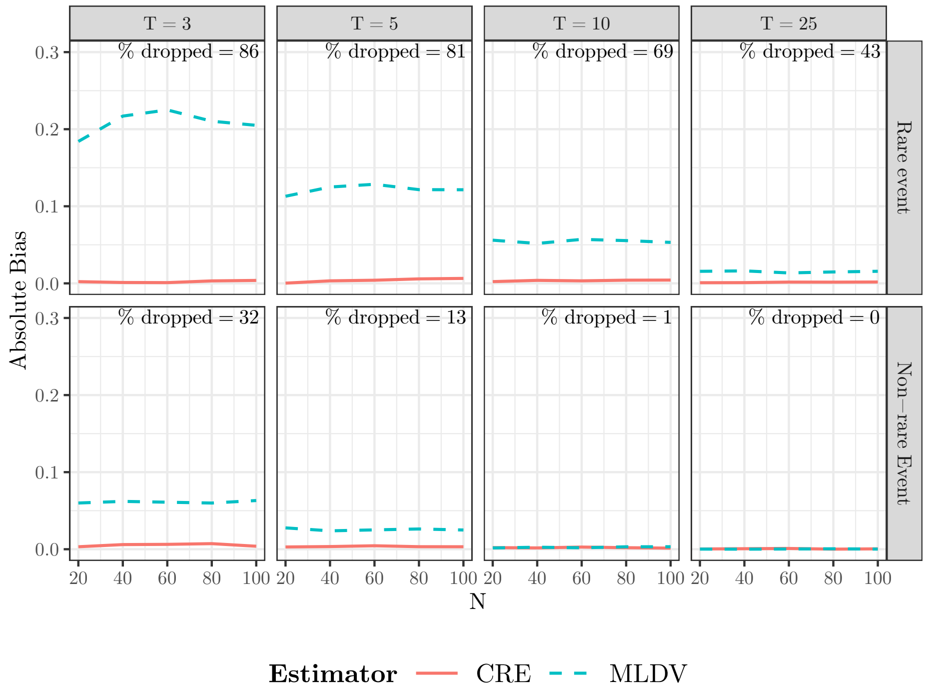


Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV and CML (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

In this appendix, I present additional results from the main Monte Carlo simulations. Figure A.1 considers the bias in the conditional maximum likelihood (CML), the maximum likelihood dummy variable (MLDV), and the correlated random effects (CRE) estimates of β. The first thing we see is that the advantage of the CRE over the MLDV persists. Again, both the CRE and the CML perform well and tend to strictly dominate the MLDV.

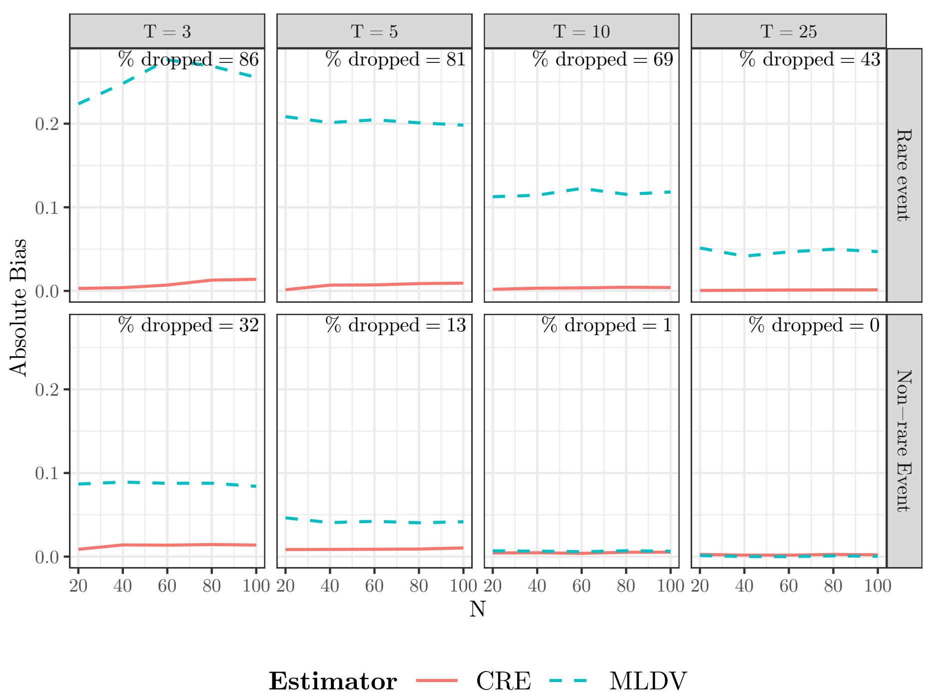
Turning to substantive effects, I first consider the estimators' performance at producing predicted probabilities. Across almost all panes in Figure A.2, the CRE is clearly less biased than the MLDV. These same trends appear in Figure A.3, which looks at bias in the AME. Again the CRE mostly dominates the MLDV. In both cases, when the data are the MLDV's best case scenario ($T \geq 10$ and non-rare events), the estimators are effectively identical. Taken together, these two plots clearly recommend the CRE over the MLDV with either rare events or small-T.

Figure A.2: Bias in estimating predicted probabilities



Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

Figure A.3: Bias in estimating the average marginal effect



Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

B Alternative approaches

Before proceeding, it's worth considering the performance of other panel-data estimators in this basic Monte Carlo setting. Specifically, I also consider a linear probability model with fixed effects, ordinary random effects (random intercepts) model, the two-step estimator from Beck (2015), and the penalized maximum likelihood (PML) estimator from Cook, Hays and Franzese (2018). Beck's two step works as follows. In the first step, the analyst uses the CML to estimate β. In the second step, the analyst estimates c by fitting the MLDV to the same data, but by fixing $\hat{\beta}$ to the first step value.

The PML works by creating a bias-reduced version of the MLDV. Cook, et al. note that the major problems with the MLDV stem from concerns that there may be separation bias in the constants associated with the homogeneous units. These units end up being dropped from the MLDV to avoid this issue, but doing so can introduce selection concerns into the estimated marginal effects. As they point out, Firth's (1993) bias-reduced estimator is promising in this context.[1] Firth-based bias reduction methods are an increasingly common tool for applied researchers looking to reduce the (typically separation) biases found in generalized linear models. The PML is created by adding a term onto the MLDV's log-likelihood function to "penalize" the estimates.[2] In practice the penalization works by imposing Jeffreys prior on all the parameters, which shrinks them toward zero. This shrinkage results from analyst-induced information (i.e., information from outside the data) and reflects a common understanding that estimates from a logistic regression should not be "too big" (Zorn 2005).

[1]Cook, et al. pool across the all-zero groups, i.e., set $c_i = c_0$ for all i where $\sum_{t=1}^{T} y_{it} = 0$. They note that having a common constant for all the groups without variation in y improves the feasibility of the estimator. I weaken this assumption to allow each of these groups to have their own constant and tackle the feasibility problem by exploiting the sparsity of the design matrix instead. I compare their original setup to the CRE in a later appendix.

[2]The PML can be estimated using Stata's `firthlogit` and the R's `brglm::brglm`.

I also considered a generalized estimation equations (GEE) approach. With the GEE there is a question of whether to include unit-dummies or just a single constant. In the case of the former, the results are nearly identical to the MLDV and with just a single constant, the results are nearly identical to a pooled logit (not good). This is expected as the GEE's main strength is it's ability to better model within unit dynamics. Given that this paper has deliberately ignored dynamic concerns in favor of considering the issues associated with unobserved heterogeneity, the GEE has almost nothing to offer over the MLDV. However, future work that addresses dynamic panels should consider the GEE as an important alternative. Given that the GEE results do not notably differ from the MLDV results, I omit them from this analysis.

Table B.1: Alternative panel-data estimators, best case conditions

	CML	PML	MLDV	RE	Beck	LPM
RMSE in $\hat{\beta}$	1.01	1.01	1.78	1.95	1.01	
RMSE in $\hat{p}$		1.49	1.53	1.49	1.52	2.09
RMSE in AME		1.24	0.98	2.48	1.42	1.10

Note: All values are relative to the CRE. Values greater than 1 imply that the CRE is the better performing estimator.

Table B.1 demonstrates the performance of these alternative panel estimators under the best-case conditions, $T = 25$ and non-rare events, with $N = 100$. Each cell reports the estimator's RMSE divided by the CRE's RMSE for the same quantity, meaning that values greater than 1 are evidence that the CRE is favored. As we can see, the PML, CML, and Beck's two-step all roughly match the CRE when it comes to finding β. Note that the CML values and the Beck values are identical when we consider β by construction. The MLDV does worse here than the others, but not as bad as the random effects model. The traditional random effects model's poor performance is expected given that it suffers from omitted variable bias by construction.

In terms of predicted values, we again see that all the alternatives perform worse than the CRE. The linear probability model estimates, in particular, have nearly twice the RMSE of the CRE estimates. The other estimators have an RMSE about about 1.5 times that of the CRE. Overall, the CRE is provides the best predicted probabilities in this experiment.

Moving on the marginal effects, we see that despite its issues with the other quantities, the MLDV is the best at finding the AME under these best case conditions. The CRE provides very similar RMSE as the ratio is very close to 1. The remaining estimators all do worse at finding the RMSE, with the traditional random effects estimator being far and away the worst choice.

Table B.2: Alternative panel-data estimators with rare events

	CML	PML	MLDV	RE	Beck	LPM
RMSE in $\hat{\beta}$	1.07	1.05	2.11	2.32	1.07	
RMSE in $\hat{p}$		1.67	2.00	1.63	1.87	2.43
RMSE in AME		2.86	8.86	2.32	7.73	1.07

Note: All values are relative to the CRE. Values greater than 1 imply that the CRE is the better performing estimator.

Overall, the results in Table B.1 suggest that even in conditions where alternative estimators should be at their best, the CRE outperforms them. This conclusion is unchanged if we were to consider a rare-events example with the same dimensions ($T = 25,\ N = 100$), and in fact the CRE's advantages tend to grow in that situation, as shown in Table B.2.

C Other parameter values

In this appendix, I expand the original Monte Carlos to consider more parameter values. The purpose here is to expand on the experiments to ensure that the main conclusions are not determined by specific values of α and β. The main setup of the Monte Carlos is unchanged except now $\alpha \in \{2, 2.25, 2.75, 3, 3.25\}$ and $\beta \in \{-3, -2, -1\}$. These values provide lots of variation in rareness with the probability that $y_{it} = 1$ ranging from about 0.003 to 0.8. In each of the figures, below, events get rarer as we move from the top to the bottom of the page. Figures C.1, C.2, and C.3 consider the RMSE in estimating β, predicted probabilities, and the AME, respectively. As in the main simulations, the MLDV does relatively well in the best-case conditions (the upper right quadrant of each figure) and performs less well than the CRE in the remaining three quadrants. Other alternative parameter settings can be found in Appendices F, E, and H.

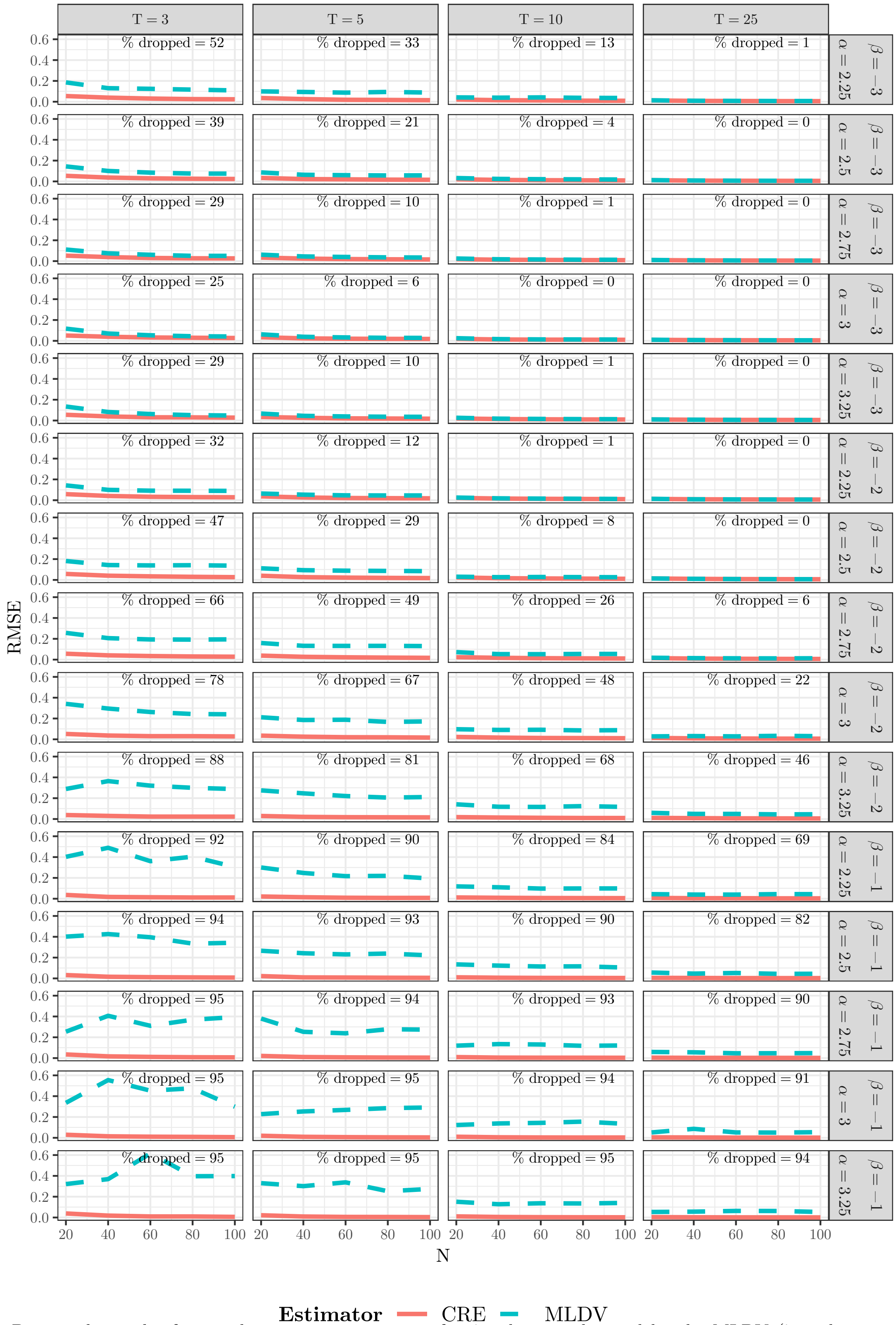
Figure C.1: RMSE in estimating β expanding the parameter space

Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV and CML (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

Figure C.2: RMSE in estimating p_{it} expanding the parameter space

Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

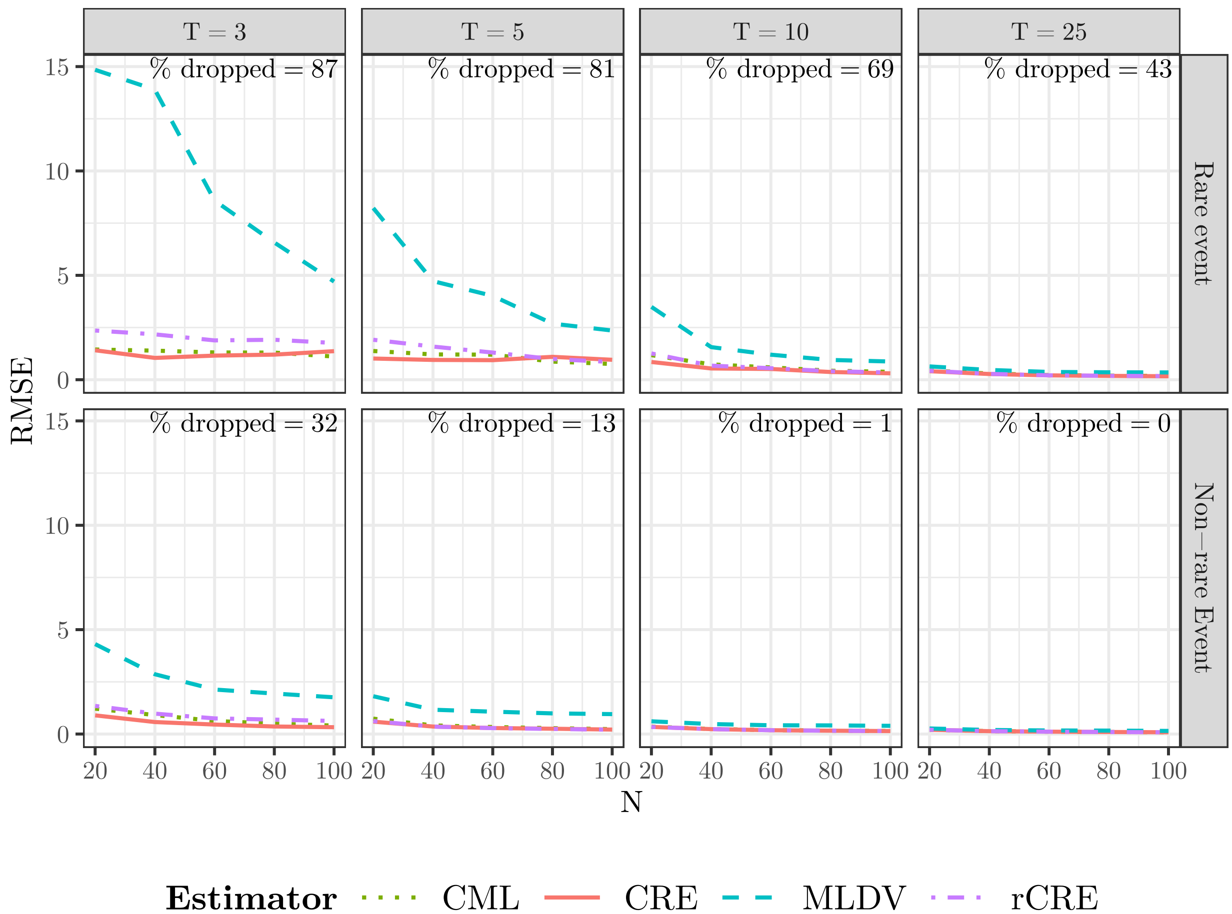
Figure C.3: RMSE in estimating the AME expanding the parameter space



Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

D Only heterogeneous units

Figure D.1: RMSE in estimating β in only heterogeneous units

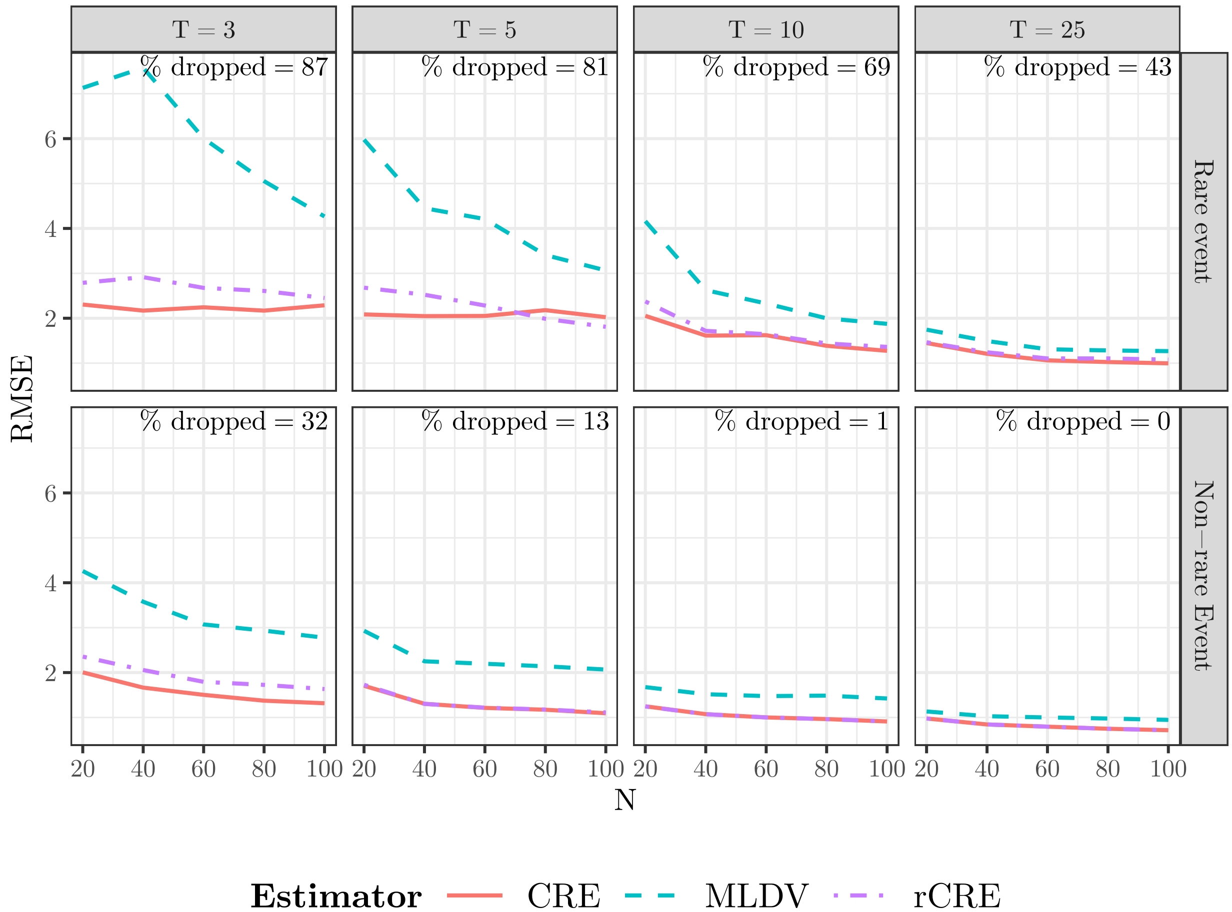


Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV, CML, and rCRE (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

Figures D.1 and D.2 consider the RMSE in estimating β and c. Throughout, the rCRE tends to be a little worse than the full-sample CRE. This difference suggests that when the all-zero units are relevant, their inclusion is helpful to the CRE. This improvement results from the partial pooling that occurs within the CRE framework, which allows it to use information from relevant homogeneous units.

We now turn our attention to how well the various estimators perform when they only consider the heterogeneous units. Specifically, we consider the following:

Figure D.2: RMSE in estimating c with only heterogeneous units



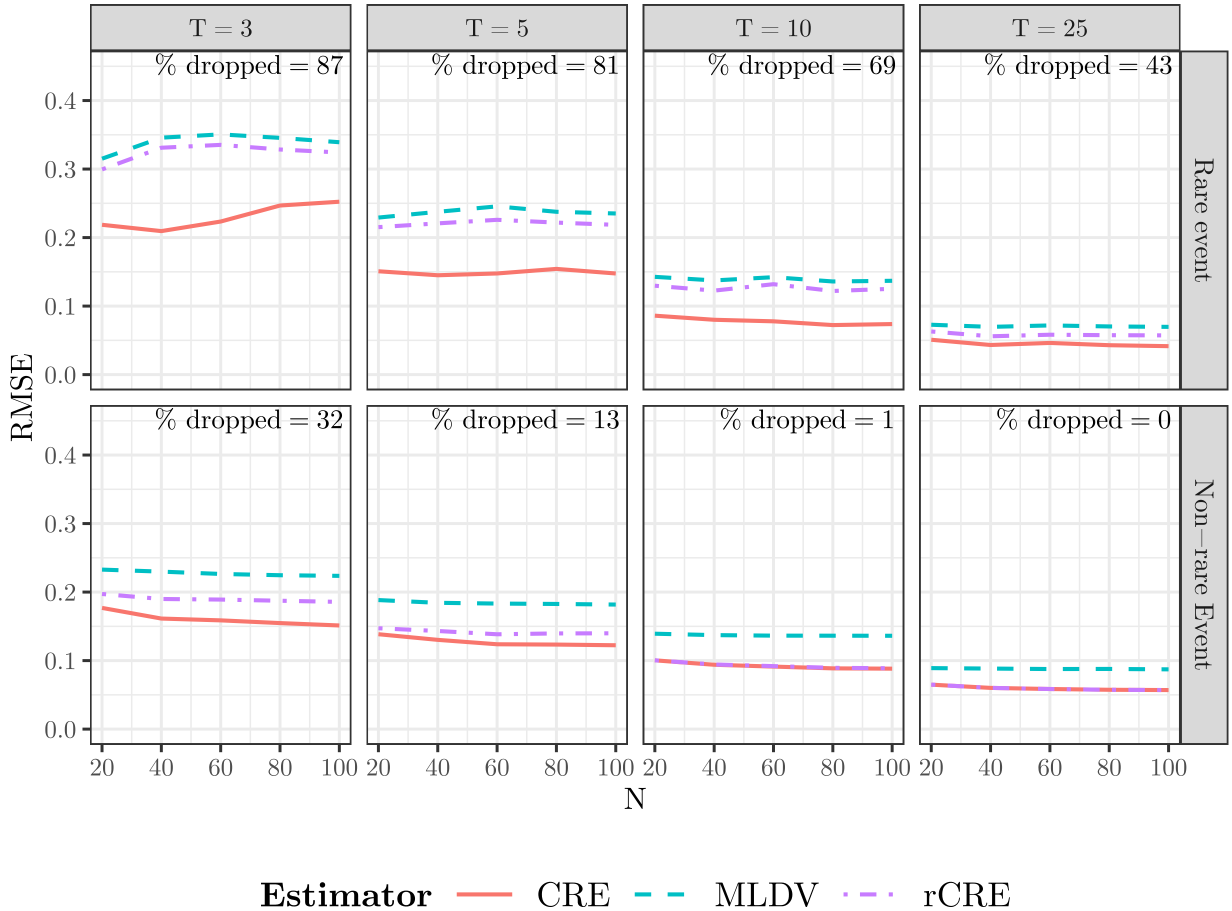
Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV and rCRE (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

- The CRE fit to the whole dataset, but we only consider the constants and substantive effects on the restricted sample considered by the CML and MLDV;
- A restricted-CRE (rCRE) fit to the restricted sample considered by the CML and MLDV;
- The MLDV.

All marginal effects here are conditional average marginal effects the (cAME) that only consider heterogeneous units. The data generating process is unchanged from the main text.

In Figures D.3 and D.4 we see how well the estimators perform at finding predicted probabilities and the cAME, respectively. The rCRE and MLDV struggle with both relative

Figure D.3: RMSE in estimating predicted probabilities in only heterogeneous units



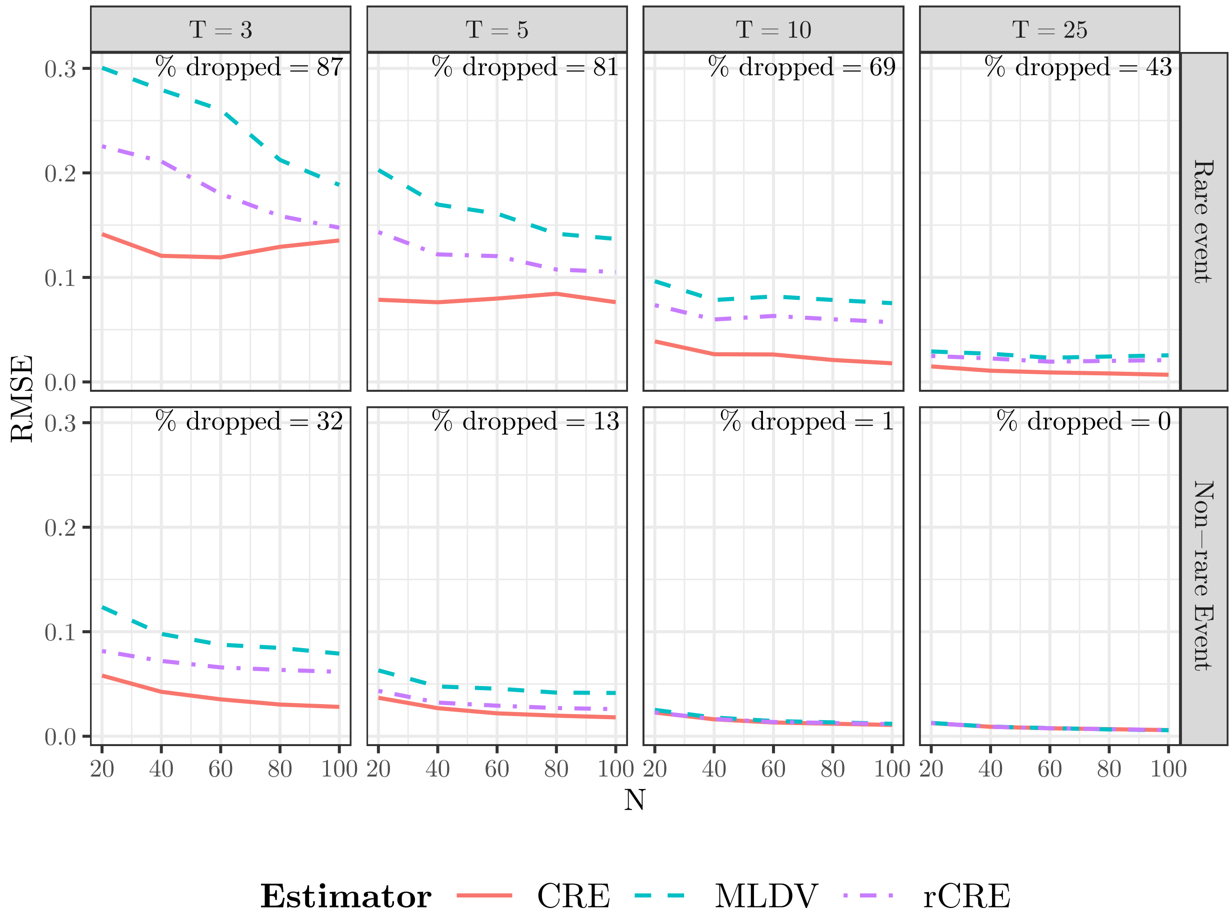
Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV and rCRE (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

to the CRE in all but the best-case conditions. In some cases, the rCRE is very similar to the MLDV, which is similar to the empirical applications.

However, in the above analysis all the homogeneous units are relevant to the data-generating process. In many cases, it will be unclear if all the all-zero units are in fact relevant. To the extent that there are relevant units, the full-sample CRE can use that information in ways that the other two approaches do not. This information helps the full-sample estimator so long as the units are relevant to the study.

Moving away from this ideal world, we now consider a situation where 90% of the sample is irrelevant by construction. To do this, I follow the same data-generating process as above, but add $9N$ additional units where $c_i = -\infty$. These units cannot ever experience the event.

Figure D.4: RMSE in estimating the cAME

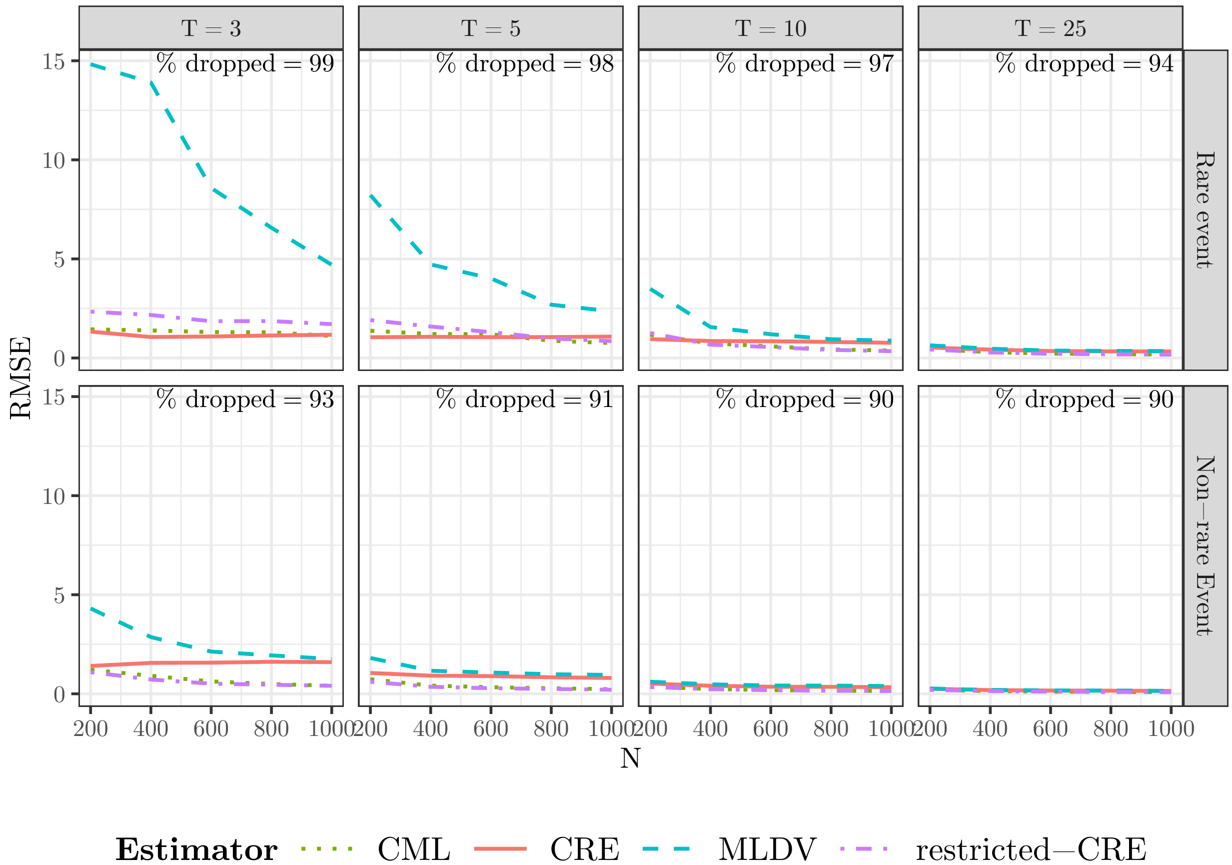


Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV and rCRE (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

The same three models are fit to this data and, as before, we are only interested in the substantive quantities associated with heterogeneous units: predicted probabilities and the cAME.

Looking at both the estimates of β (Figure D.5) and the predicted probabilities (Figure D.6), there appears to be a major split between the models depending on rareness. With rare-events there are some relevant all-zero (a minority of the total all-zero units, but more than in the non-rare events data), the full-sample CRE is still the best choice in many cases. As T increases there are fewer relevant all-zero units and the estimators perform more similarly. With more common data, the rCRE is the best performer, but the full-sample CRE still tends to outperform the MLDV.

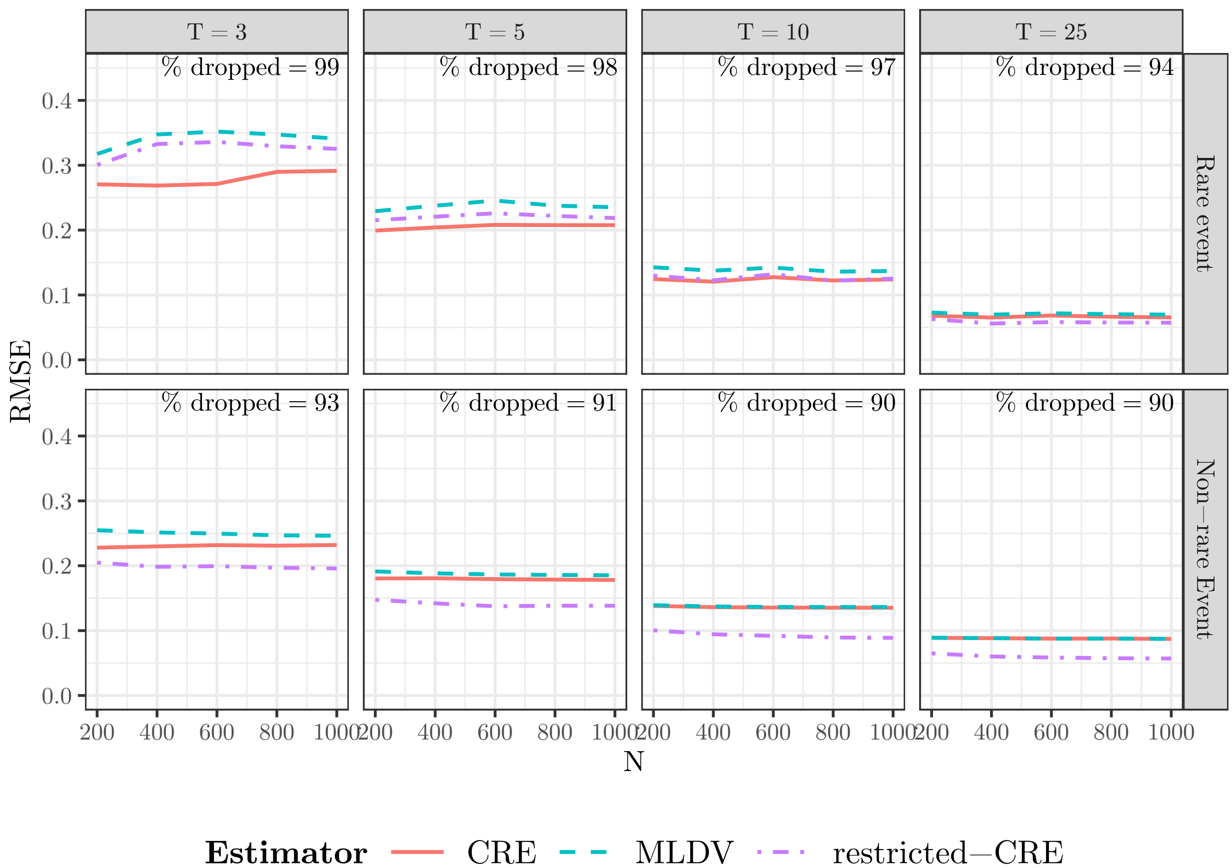
Figure D.5: RMSE in estimating β with only heterogeneous units and lots of irrelevant units.



Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV, CML, and rCRE (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

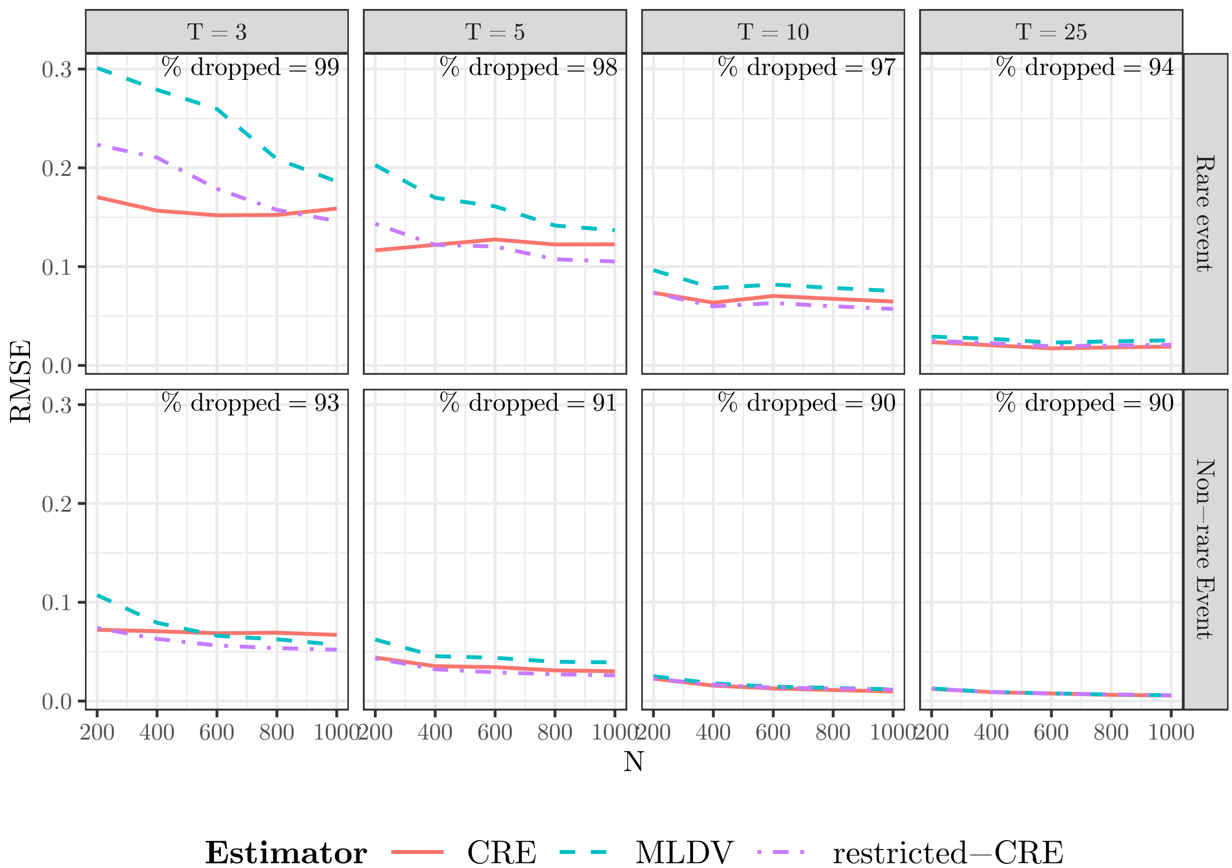
Turning to the marginal effects estimates (Figure D.7), the trends are similar but slightly different. Once again, the massive number of irrelevant observations worsens the CRE, but even with so many totally irrelevant units (90% of the sample), it tends to do better than the MLDV and very similar to the rCRE when it comes to finding the cAME. As the percentage of irrelevant units grows, the full-sample CRE will continue to worsen. Unfortunately, it is impossible to test for whether a unit is irrelevant. As such, the choice between using the full-sample, the restricted sample, or some hybrid is entirely theoretical. The safest course of action when the cAME is of interest is to estimate it multiple ways and note/discuss any differences that emerge.

Figure D.6: RMSE in estimating predicted probabilities with only heterogeneous units and lots of irrelevant units.



Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV and rCRE (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

Figure D.7: RMSE in estimating the cAME with only heterogeneous units and lots of irrelevant units.



Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV and rCRE (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

E Highly censored, very rare data

This next set of Monte Carlos considers an example inspired by Green, Kim and Yoon (2001). In this case, the data generating process is still the usual:

$$y_{it} = \mathbb{I}((z_i + x_{it})\beta + z_i\alpha + \varepsilon_{it} > 0),$$

where $x_{it} \sim N(0,1)$, $z_i \sim N(-4,1)$, $\beta = 2$, and $\varepsilon_{it} \sim \text{Logistic}(0,1)$, but now $\alpha = 5.2$.[3] This produces an extremely rare events data set (about 0.1% of the data are 1s) with lots of all-zero units (roughly 98% of units). Here we only consider $T = 30$ and $N = 1000$.

As in the previous appendix, we are interested in how (not) including these large numbers of homogeneous units affects estimator performance. Table E.1 displays the results. Even though there are many units that are effectively irrelevant (the 90th quartile for the probability that $y_{it} = 1$ is less than 0.0004), the full-sample CRE is easily the preferred choice. As in the empirical examples, the MLDV and rCRE tend to produce very similar estimates of predicted probabilities (in this case an average of 0.005–0.006, or about three times that of the CRE) and cAME (in this case an average of −0.07, or about twice that of the CRE).

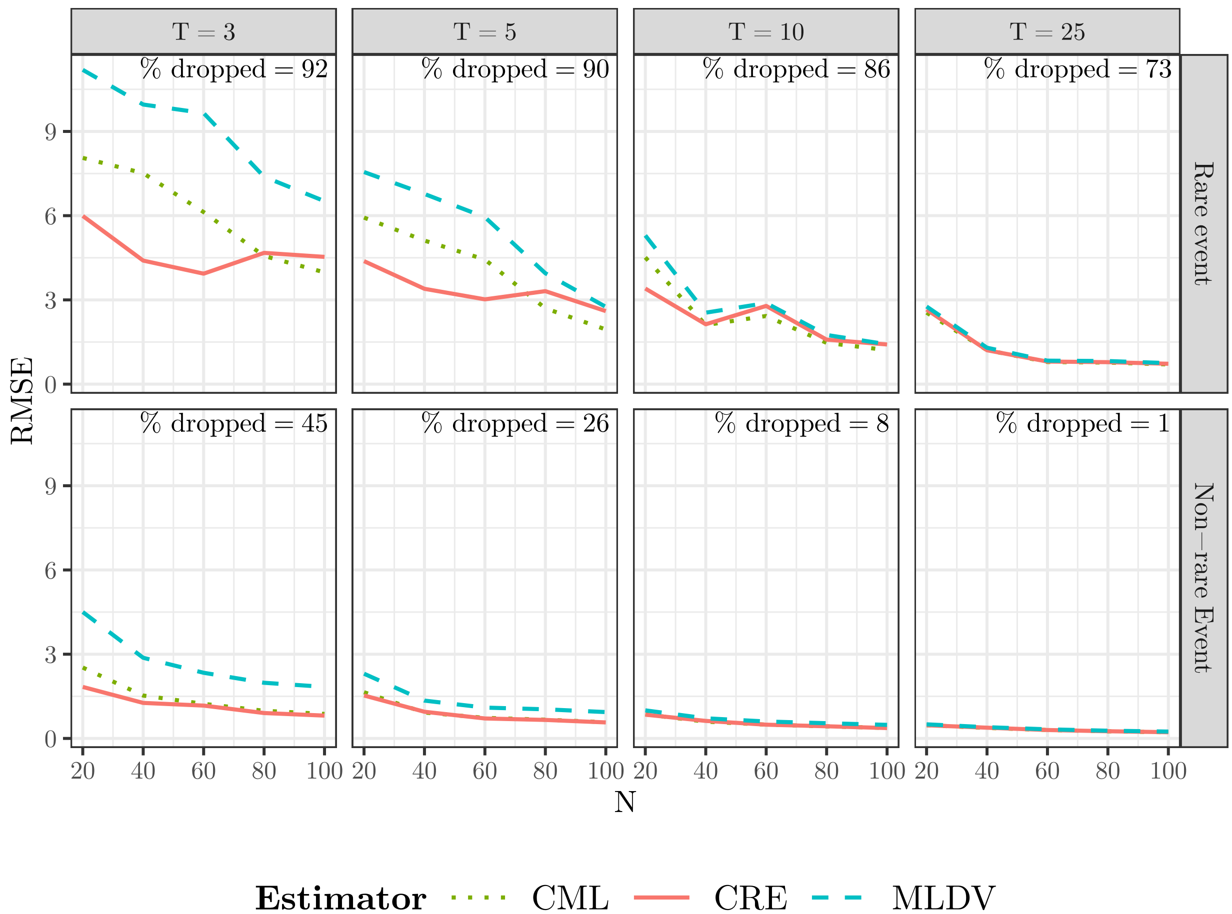
Table E.1: Estimator performance among heterogeneous units with highly censored, rare events data

	rCRE	CRE	MLDV
RMSE in $\hat{c}$	1.52	1.18	1.48
RMSE in $\hat{p}$	0.07	0.04	0.07
RMSE in cAME	0.03	0.01	0.07

[3]The constant terms are thus distributed normally with mean −20.8 and standard deviation 5.2. This creates a situation that is in between the settings considered in Appendix D with many homogeneous units that may be technically relevant in sense that they can experience event, but practically irrelevant in the sense that an event is *very* unlikely.

F Nonlinear relationship with unobservables

Figure F.1: RMSE in estimating β with a nonlinear relationship between the observed and unobserved variables



Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV and CML (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

A concern throughout is with the CRE's functional form assumption. The Mundlak specification used in this paper has the advantage of being simple and easy to use, but it may struggle in cases where there is a complicated relationship between the observed covariates x_{it} and unobserved heterogeneity z_i. Here, I examine one such case to get a handle on how bad the damage might be. The data generating process is still

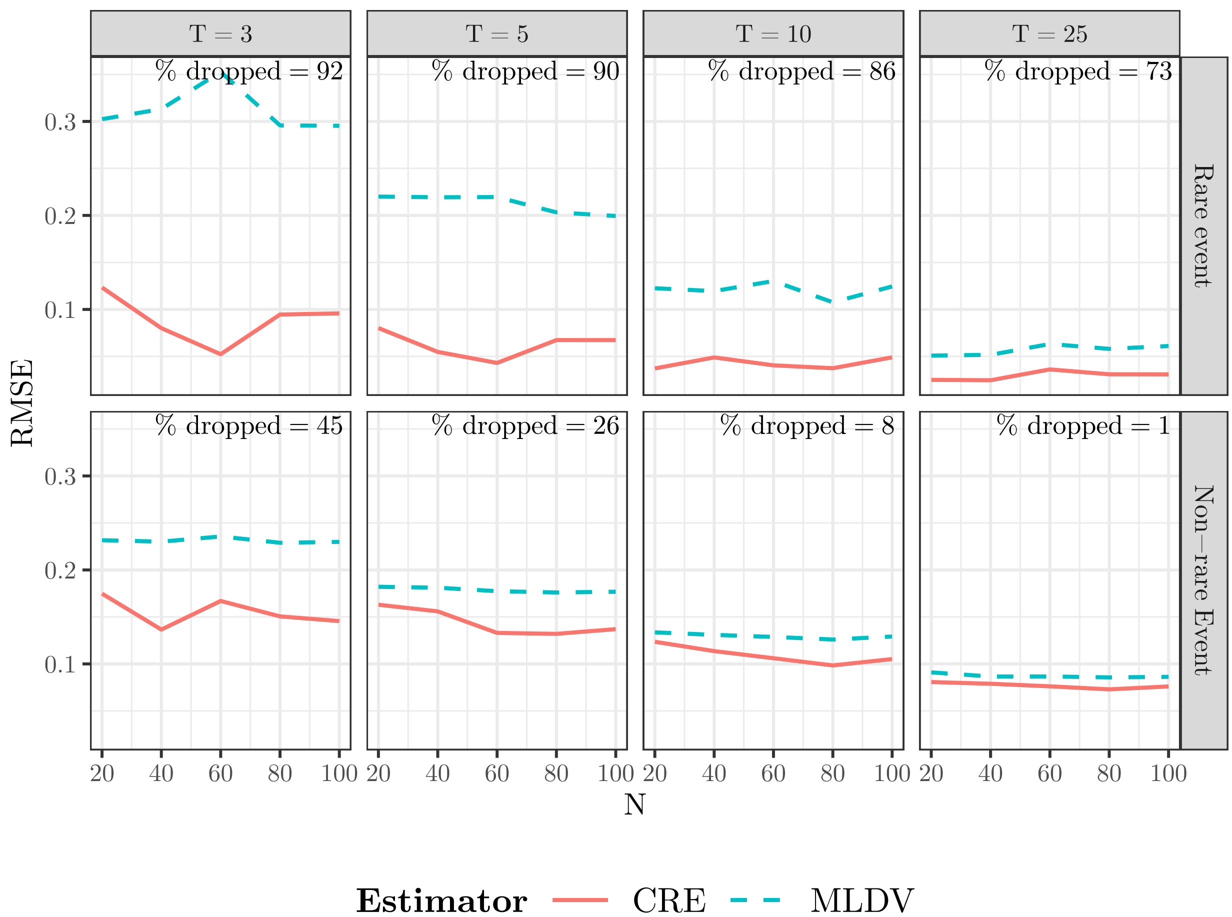
$$y_{it} = \mathbb{I}(x^*_{it}\beta + z_i\alpha + \varepsilon_{it} > 0).$$

However, now

$$x_{it}^* = \log(z_i^2 + 7z_i + 16 + x_{it}),$$

where $x_{it} \sim N(0,1)$, $z_i \sim N(-4,1)$, $\beta = 2$, and $\varepsilon_{it} \sim \text{Logistic}(0,1)$. As before, z_i is unobserved, and x_{it}^* is the only observable. Now α takes on values of 2 (rare event) and 1 (not rare). The relationship between x^* and z is now clearly nonlinear, so we may be concerned that the Mundlak specification is no longer up to the task.

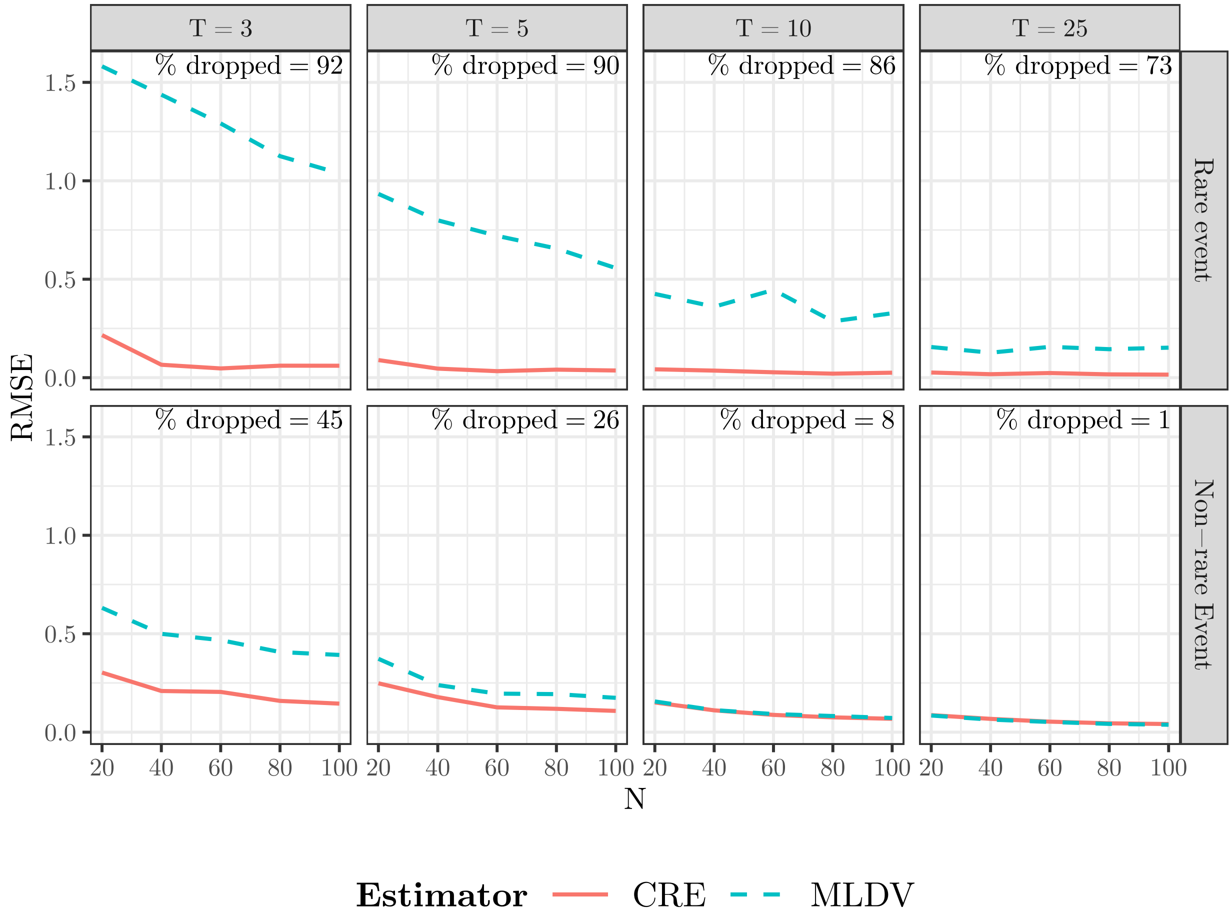
Figure F.2: RMSE in estimating predicted probabilities with a nonlinear relationship between the observed and unobserved variables



Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

Figure F.1 considers estimating β under these new settings. overall, the CRE still does very well in this situations, which is reassuring. Likewise as we look at the substantive

Figure F.3: RMSE in estimating the AME with a nonlinear relationship between the observed and unobserved variables

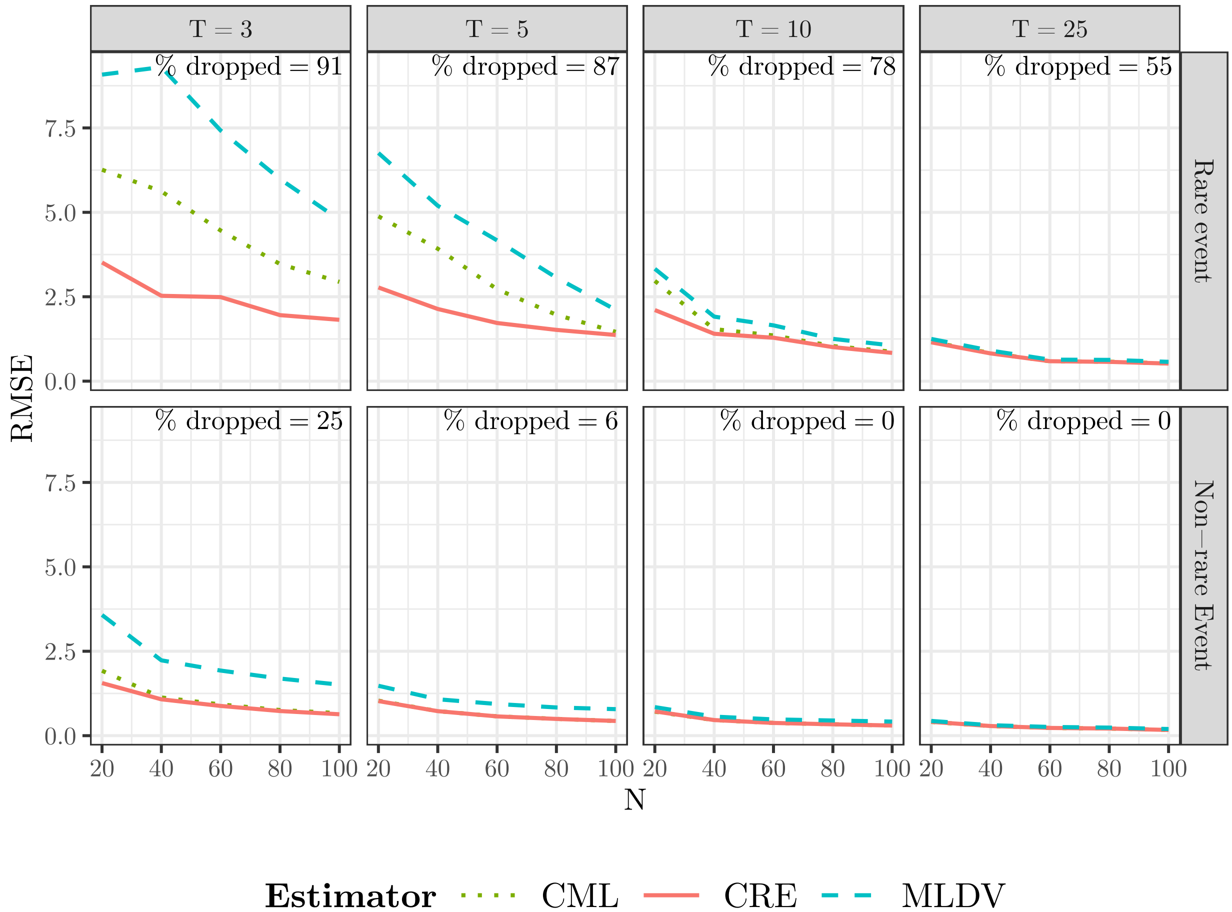


Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

quantities, we see that the CRE is still doing well relative to the MLDV when it comes to predicted probabilities (Figure F.2) and marginal effects (Figure F.3) .

G Little within variation

Figure G.1: RMSE in estimating β with low-amounts of within variation



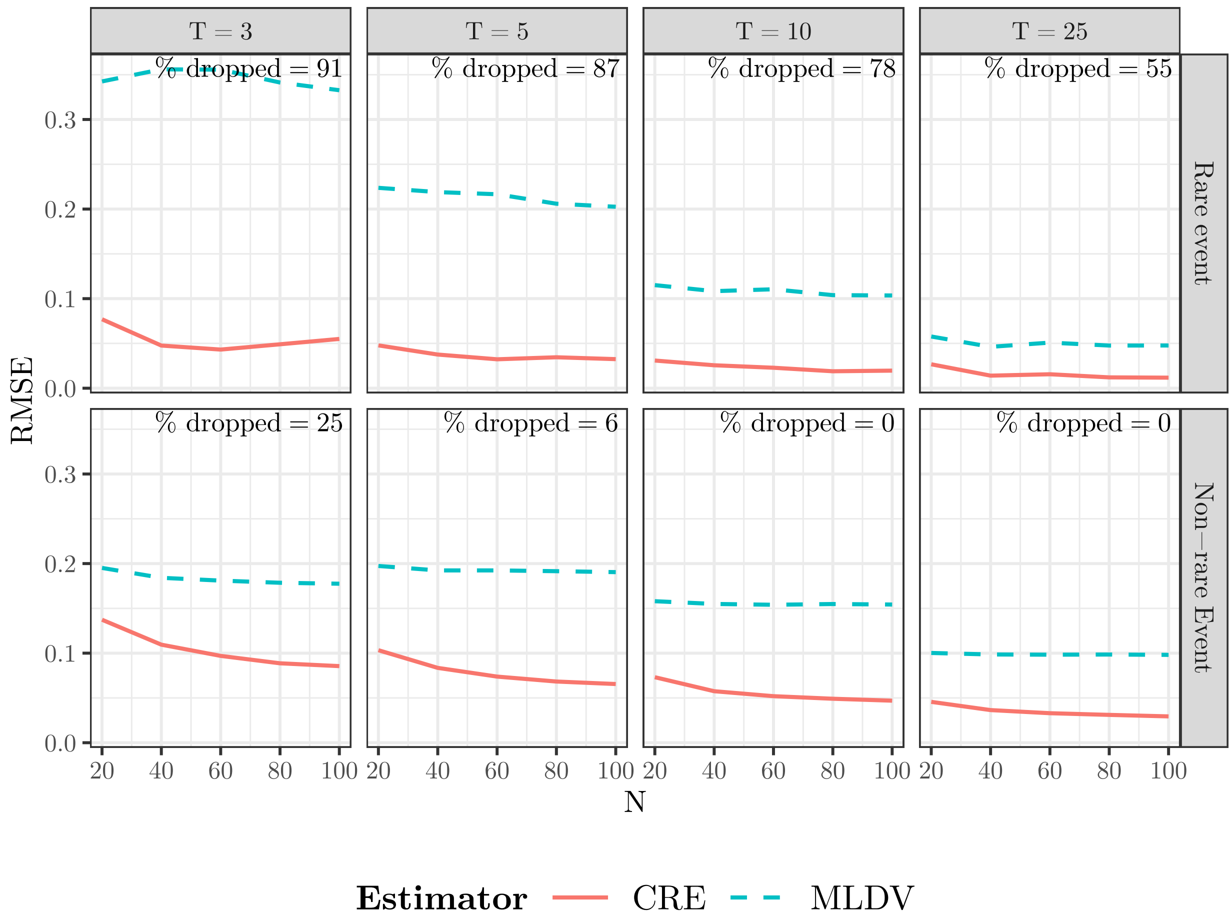
Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV and CML (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

We are now interested in how well the estimators perform with rarely-changing covariates. This is a common problem in a lot of IR and comparative TSCS datasets. Here, we return to the original data generating process presented in the main text, but now the within standard deviation of x^*_{it} is set to 0.25 instead of 1. Additionally, α now takes on the values of 3 and 2 to keep the proportion of 1s roughly similar to the main Monte Carlos. Tightening the within variation raises the correlation between x^* and $\bar{x}^*$ (up to above 0.95), which raises strong concerns about multicolinearity in the CRE.

Figure G.1 considers the point estimates in these conditions, and the results roughly match the main Monte Carlos. The CRE is still the most preferred over a range of situations.

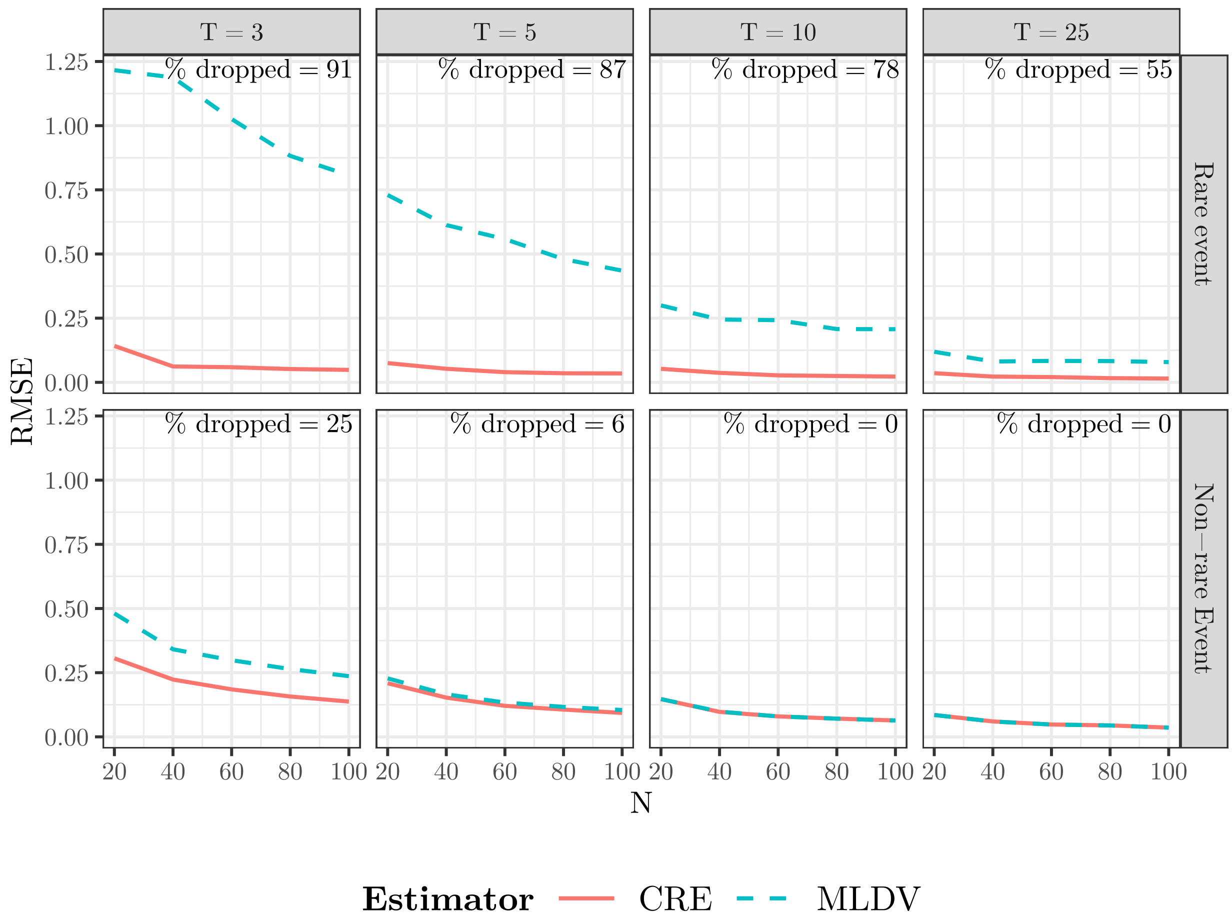
Similarly, when considering the substantive quantities, we see the usual trends: the CRE is almost always the best choice.

Figure G.2: RMSE in estimating predicted probabilities with low-amounts of within variation



Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

Figure G.3: RMSE in estimating the AME with low-amounts of within variation



Note: Percent dropped refers to the average percentage of units that are dropped by the MLDV (i.e., the percentage of all-zero/all-one units) across the simulated datasets within each pane.

H A direct comparison to PML

In this appendix, I provide a more direct comparison between the CRE and the PML by replicating the main Monte Carlos from Cook, Hays and Franzese (2018). Here, I follow their implementation of the PML and pool all the homogeneous units. Overall, this set of simulations provides two important results.

1. The CRE's improvement over the PML, as reported above, does not depend on how whether we pool the homogeneous units or not;

2. The CRE performs well under a different parameter setting.

In the below tables I rerun their main simulations and report how the CRE, CML, and MLDV compare to the PML, where values greater than 1 mean that the estimator is preferred to the PML

The data generating process follows Cook, et al. (2018):

$$
\begin{aligned}
y_{it} &= \mathbb{I}(x_{it}\beta + c_i + \varepsilon_{it} > 0)) \\
x_{it} &\sim N(\bar{x}_i, \sigma^2) \\
\begin{pmatrix} c_i \\ \bar{x}_i \end{pmatrix} &\sim N\left(\begin{pmatrix} -4 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),
\end{aligned}
$$

where $\beta = 1$ and ε_{it} remains i.i.d. standard logistic.

Tables H.1-H.5 reproduce Tables 1-5 from their manuscript, where the ordinary random effects model they consider is changed to be the CRE. Values greater than 1 signify that the estimator is preferred to their PML. Across all five of the above Tables, the CRE is at least as good (ratio of about 1) or noticeably preferred to the PML.

Table H.1: RMSE ratios in estimating β $(\rho = 0.5, \sigma = 1)$

	Pooled	CRE	MLDV	CML
T=20, N=50	0.88	1.02	0.84	0.99
T=20, N=100	0.59	1.04	0.74	0.99
T=50, N=50	0.55	1.00	0.92	1.00
T=50, N=100	0.32	1.00	0.87	1.00

Table H.2: RMSE ratios in estimating the AME $(\rho = 0.5, \sigma = 1)$

	Pooled	CRE	MLDV	Censoring (%)
T=20, N=50	0.52	1.06	0.14	48
T=20, N=100	0.41	1.12	0.10	50
T=50, N=50	0.33	1.09	0.17	32
T=50, N=100	0.26	1.28	0.14	33

Table H.3: RMSE ratios in estimating the AME $(\rho = 0.25, \sigma = 1)$

	Pooled	CRE	MLDV	Censoring (%)
T=20, N=50	1.07	1.04	0.15	49
T=20, N=100	0.73	1.15	0.11	50
T=50, N=50	0.66	1.14	0.18	31
T=50, N=100	0.44	1.30	0.16	32

Table H.4: RMSE ratios in estimating the AME $(\rho = 0.5, \sigma = 2)$

	Pooled	CRE	MLDV	Censoring (%)
T=20, N=50	0.60	1.02	0.14	30
T=20, N=100	0.43	1.00	0.09	32
T=50, N=50	0.35	1.04	0.21	15
T=50, N=100	0.31	1.11	0.18	15

Table H.5: RMSE ratios in estimating the AME $(\rho = 0.25, \sigma = 2)$

	Pooled	CRE	MLDV	Censoring (%)
T=20, N=50	1.06	1.01	0.16	29
T=20, N=100	0.72	1.00	0.10	32
T=50, N=50	0.66	1.04	0.25	13
T=50, N=100	0.56	1.12	0.21	13

I Fitting the CRE

As mentioned the CRE can be fit using Stata's `xtlogit` or R's `lme4::glmer`. Alternatively, a Bayesian implementation might be of interest, such as R's `rstanarm::stan_glmer`. For example

```r
library(data.table) #easy data aggregation
library(readstata13) #read the data
library(lme4) #Maximum likelihood RE
library(rstanarm) #Bayesian RE

emw <- data.table(read.dta13("temp-micro.dta"))
emw[,proUse:=mean(progov),by=dcode] #to match their coding

#Generate the x_bar terms using data table
emw[,`:=`(remit.bar = mean(remit,na.rm=T),
           cellphone.bar = mean(cellphone,na.rm=T),
           lage.bar = mean(lage,na.rm=T),
           education.bar = mean(education,na.rm=T),
           wealth.bar = mean(wealth,na.rm=T),
           male.bar = mean(male,na.rm=T),
           employment.bar = mean(employment,na.rm=T),
           travel.bar = mean(travel,na.rm=T),
           remitXpro.bar = mean(remitXpro,na.rm=T)),
      by = dcode] #grouping variable


# MLE
mle.cre.out <- glmer(protest01 ~ remit + remit:proUse + cellphone + lage +
                       education + wealth + male + employment+ travel +
                       remit.bar + remitXpro.bar +cellphone.bar + lage.bar +
                       education.bar + wealth.bar + male.bar + employment.bar +
                       travel.bar + (1|dcode),
                  data=emw, family=binomial(), nAGQ = 12,
                  control=glmerControl(optimizer = "bobyqa",
                                       optCtrl=list(maxfun=500000)))

# Bayesian
bayes.cre.out <- stan_glmer(protest01 ~ remit + remit:proUse + cellphone + lage +
                              education + wealth + male + employment+ travel +
                              remit.bar + remitXpro.bar + cellphone.bar + lage.bar +
                              education.bar + wealth.bar + male.bar + employment.bar +
                              travel.bar + (1|dcode),
                      data=emw, family=binomial(),
                      prior_intercept=normal(0,1), prior=normal(0,100),
                      chains=4, cores=4, seed=1234567, init="0",iter=1000, thin=2)
```

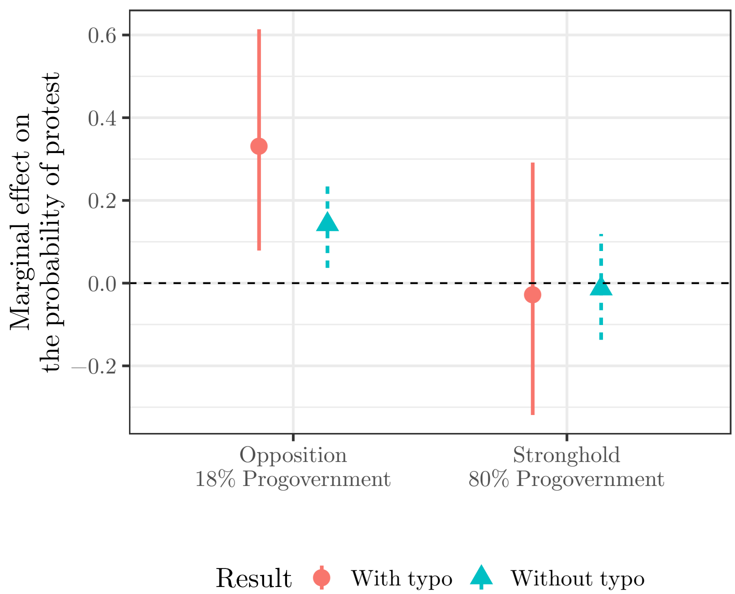
Listing 1: Fitting the CRE to EMW's data in R or in Stata

```
/*CRE*/
use temp-micro.dta, clear
xtset dcode
egen proUse = mean(progov), by(dcode) //match their use

/*Generate the group means */
egen remit_bar = mean(remit), by(dcode)
egen remitXpro_bar = mean(remitXpro), by(dcode)
egen cellphone_bar = mean(cellphone), by(dcode)
egen lage_bar = mean(lage), by(dcode)
egen education_bar = mean(education), by(dcode)
egen wealth_bar = mean(wealth), by(dcode)
egen male_bar = mean(male), by(dcode)
egen employment_bar = mean(employment), by(dcode)
egen travel_bar = mean(travel), by(dcode)

xtlogit protest01 remit c.remit#c.proUse cellphone lage education wealth male
    employment travel remit_bar remitXpro_bar cellphone_bar lage_bar education_bar
    wealth_bar male_bar employment_bar travel_bar, re
```

Listing 2: Fitting the CRE to EMW's data in Stata

J About Figure 2 in EMW

Figure J.1: Average marginal effects from EMW



Note that the marginal effects values I report for EMW (2018) method in the caption of Figure 5 do not match the published values in their Figure 2. Examining their code reveals a slight typo (lines 250 and 259 in their replication file) that explains the discrepancy. Specifically, in estimating the average marginal effect, given by

$$\begin{aligned}\widehat{\text{AME}}_{\text{Remit}} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T_i}\sum_{t=1}^{T_i}\Big[\Lambda\left(\text{Remit}_{it}\hat{\beta}_1 + (\text{Remit}_{it}\times\text{Progovernment}_i)\hat{\beta}_2 + x'_{it}\hat{\beta}_3 + \hat{c}_i\right)\\ \times\left(1-\Lambda\left(\text{Remit}_{it}\hat{\beta}_1 + (\text{Remit}_{it}\times\text{Progovernment}_i)\hat{\beta}_2 + x'_{it}\hat{\beta}_3 + \hat{c}_i\right)\right)\\ \times\left(\hat{\beta}_1 + \text{Progovernment}_i\hat{\beta}_2\right)\Big]\times 5,\end{aligned} \tag{J.1}$$

they accidentally omit the middle line giving them

$$\begin{aligned}\widehat{\text{AME}}_{\text{Remit}} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T_i}\sum_{t=1}^{T_i}\Big[\Lambda\left(\text{Remit}_{it}\hat{\beta}_1 + (\text{Remit}_{it}\times\text{Progovernment}_i)\hat{\beta}_2 + x'_{it}\hat{\beta}_3 + \hat{c}_i\right)\\ \times\left(\hat{\beta}_1 + \text{Progovernment}_i\hat{\beta}_2\right)\Big]\times 5.\end{aligned} \tag{J.2}$$

To verify this, consider Figure J.1.[4] The results in red match the printed results from EMW and are produced by me using Eq. J.2, while the results in blue are produced using Eq. J.1 and are presented in the main text. Obviously, this difference is incredibly minor does not change any of their conclusions.

References

Beck, Nathaniel. 2015. "Estimating Grouped Data Models with a Binary Dependent Variable and Fixed Effects: What are the Issues?" Unpublished manuscript.

Cook, Scott J., Jude C. Hays and Robert J. Franzese. 2018. "Fixed Effects in Rare Events Data: A Penalized Maximum Likelihood Solution." *Political Science Research and Methods*. Online first.

Escribà-Folch, Abel, Covadonga Meseguer and Joseph Wright. 2018. "Remittances and Protest in Dictatorships." *American Journal of Political Science* 62(4):889–904.

Firth, David. 1993. "Bias Reduction of Maximum Likelihood Estimates." *Biometrika* 80(1):27–38.

Green, Donald P., Soo Yeon Kim and David H. Yoon. 2001. "Dirty Pool." *International Organization* 55(2):441–468.

Zorn, Christopher. 2005. "A Solution to Separation in Binary Response Models." *Political Analysis* 13(2):157–170.

[4] Since $\Lambda(x)(1 - \Lambda(x)) = \Lambda'(x) = \lambda(x)$, the typo is easily understandable as accidentally using Stata's `invlogit` (Λ) in place of `logisticden` (λ).