Partial Effects for Binary Outcome Models with Unobserved Heterogeneity

Lucas Núñez *

California Institute of Technology

August 27, 2017

Abstract

Unobserved heterogeneity is ubiquitous in empirical research. In this paper, I propose a method for estimating binary outcome models with panel data in the presence of unobserved heterogeneity, called the *Penalized Flexible Correlated Random Effects* (PF-CRE) estimator. I show that this estimator produces consistent and efficient estimates of the model parameters. PF-CRE also provides consistent estimates of partial effects, which cannot be calculated with existing consistent estimators. Using Monte Carlo simulations, I show that PF-CRE performs well in small samples. To demonstrate that accounting for unobserved heterogeneity has important consequences for empirical analysis, I use PF-CRE to analyze tactical voting during the 2015 British Election. Ignoring the unobserved heterogeneity leads to an overestimation of the effects of being contacted by parties during the campaign. The suggested mechanism is that parties tend to contact voters who are already likely to cast tactical votes.

*Ph.D. Candidate, Division of the Humanities and Social Sciences, California Institute of Technology. Email: lnunez@caltech.edu

1 Introduction

The presence of unobserved heterogeneity is ubiquitous in observational studies in political science, and the social sciences in general. In studies of political behavior, this heterogeneity sometimes takes the form of voters' core beliefs, which are hard to define, let alone to measure. It can also take more mundane forms. For example, researchers rarely get to observe how political parties choose which voters to contact during electoral campaigns. Regardless of its origin and form, unobserved heterogeneity poses the same problem: ignoring it when it is correlated with the covariates of interest leads to biased and inconsistent estimates of the quantities of interest. Returning to the example, if a party contacts those voters who are already likely to support it (in a way the researchers do not observe), then the effect of party contact on the probability of voting for that party will be overestimated if the researchers does not account for unobserved heterogeneity.

There are three main estimation approaches for binary outcome models with panel data in the presence of unobserved heterogeneity: treat the heterogeneity as parameters to be estimated; use conditional maximum likelihood estimation (Rasch, 1961; Chamberlain, 1980) and related semiparametric techniques (e.g., Abrevaya, 2000); or use random or correlated random effects (Mundlak, 1978; Chamberlain, 1980). Each of these approaches suffers from one of three problems. They produce inconsistent and biased estimates, cannot produce estimates of the probability of the outcome nor partial effects of the covariates of interest, or they require making restrictive assumptions about how the individual heterogeneity relates to the observed covariates in the model.[1]

In this paper I develop an estimator that deals with unobserved heterogeneity in binary outcome models, the *Penalized Flexible Correlation Random Effects* (PF-CRE) estimator. In the PF-CRE estimator, I explicitly model the correlation between the observed and unobserved components of the model, using a large flexible specification.

[1]Making restrictive assumptions about the individual heterogeneity also leads to biased estimates. I distinguish the bias and inconsistency that arise from unrealistic assumptions from the one that arises from the estimation procedure itself.

I include a penalization step for variable selection to induce efficiency. This estimator addresses the three problems described above: it provides consistent estimates of the model parameters, allows for the estimation of partial effects, and makes weak assumptions about the unobserved heterogeneity.

The PF-CRE estimator builds upon the correlated random effects (CRE) approach by using a rich and *flexible* specification of the correlation between the unobserved heterogeneity and the observed covariates in the model. This flexible specification is composed of functions of the observed covariates (such as individual time-means and other exchangeable functions[2]), additional observed time-invariant characteristics, and higher order interactions between these terms. The flexible specification in PF-CRE requires making weaker assumptions about the unobserved heterogeneity than in the traditional CRE approach. Weaker assumptions mean that PF-CRE is more likely to capture the underlying heterogeneity correctly and lead to correct inferences.

The key challenge of the specification in PF-CRE is that it requires the estimation of additional parameters. When the number of covariates is small, this does not pose a major hurdle. However, the number of parameters grows exponentially with the number of covariates in the model. For example, with 3 observed covariates, a relatively simple specification that models the unobserved heterogeneity on the time-means of the covariates with up to three-way interactions requires the estimation of 25 parameters, which is manageable; with with 5 covariates, 63 parameters; with 10 covariates, 298 parameters.[3] Moreover, if the specification also includes additional time-invariant characteristics, the number of coefficients in the model becomes unmanageable very fast.[4]

To address the dimensionality issue, I estimate the model via *penalized* Maximum

[2]Exchangeable functions are those for which the order of their arguments does not change their value. For example, moments are exchangeable functions: an average does not change if we alter the order in which the terms enter the sum

[3]With three covariates, there are 3 coefficients associated with the covariates, a constant term, 3 associated with the time-means, 6 for two-way interactions, 10 for three-way interactions, and the variance of the random effect.

[4]For example, in the main application presented in this paper, I model the unobserved heterogeneity with 1797 parameters.

Likelihood using the Smoothly Clipped Absolute Deviation (SCAD) penalty. Importantly, the penalization is only applied to the terms that model the unobserved heterogeneity, but not to the covariates of interest. Like other penalized estimation methods, SCAD introduces a cost in the likelihood function for the size of each parameter to be estimated. Therefore, when maximizing the penalized likelihood, the coefficients of terms with little or no prediction power are shrunk to zero, performing variable selection. In the case of PF-CRE, the penalization selects the terms that are necessary to control for the unobserved heterogeneity and discards the rest. Since the main covariates of interest are not penalized in PF-CRE, no shrinkage is introduced to those parameters. The reduction of dimensionality is especially useful in small samples, as it can significantly reduce variance of the estimates.

The assumptions in the PF-CRE estimator may not always be sufficient to capture the unobserved heterogeneity present in the data.[5] Thus, for the logistic case, I present a model specification test to determine whether the PF-CRE approach is appropriate for the data at hand. This provides an indirect test of the assumptions in PF-CRE and a tool for researchers to decide when it is correct to use it.

I study the small sample performance of the PF-CRE estimator using Monte Carlo simulations. The simulations show that the asymptotic properties of PF-CRE hold in small samples, and that it performs better than alternative estimators. Additionally, the penalization step is key for reducing uncertainty around the estimates. For the logistic case, the simulations show that the rejection rate of the specification test is close to theoretical levels, although erring on the conservative side for small and moderate sample sizes.

I provide an application of PF-CRE to tactical voting during the 2015 United Kingdom General Election. The outcome of interest is whether a voter intends to cast a tactical vote, that is, a vote for a party that is not her most preferred one. I use six waves of the British Election Study Online Panel. The covariates of interest are an

[5]The underlying heterogeneity may be correlated with the observed covariates in a highly non-linear way that exchangeable functions of the covariates and interactions terms used by PF-CRE may fail to successfully approximate.

indicator of electoral viability of the most preferred party, feeling thermometers for the most preferred and most preferred viable parties, as well as indicators for whether the voters' were contacted by those parties in the four weeks prior to each survey wave. I model the unobserved heterogeneity using the time-means of these covariates together with several time-invariant demographic characteristics and (up to) three-way interaction terms. The specification test for PF-CRE shows that it provides consistent and efficient estimates of the parameters in the model, with the additional advantage that it allows for the estimation of partial effects.[6]

The results show that ignoring unobserved heterogeneity leads to significant overestimation of the effect of being contacted by the most preferred and most preferred viable parties on the probability of casting a tactical vote, almost doubling the effects. This overestimation is possibly the result of parties contacting those voters whom they think are more likely to cast a tactical vote. Controlling for the unobserved heterogeneity reduces or eliminates the bias by capturing each voter's overall characteristics, which are likely closely related to the information parties use to decide which voters to contact. The results also show that ignoring unobserved heterogeneity leads to an underestimation of the effect of party viability on voters' propensities to cast a tactical vote.

Finally, I provide two additional applications that show that the assumptions of the PF-CRE estimator hold in other political science applications. In particular, I show that PF-CRE provides consistent and efficient estimates of (1) the effect of preferences for immigration and economic fears on voting decisions for the 2016 Brexit Referendum in the U.K.; and (2) the effect of ideological preferences and candidate characteristics on vote choice during the 2012 U.S. Presidential Election. In both these cases, ignoring the unobserved heterogeneity leads to significant differences in the estimated partial effects of the covariates of interest and our understanding of voter behavior.

The rest of the paper is organized as follows. Section 2 introduces the PF-CRE

[6] I also show that the random effects and a traditional correlated random effects model do not pass the model specification test.

estimator and compares it to existing estimators. Section 3 presents the Monte Carlo simulations. Section 4 presents the application to tactical voting in the U.K. Section 5 presents a short discussion of the two additional applications. Section 6 concludes.

2 Penalized Flexible Correlated Random Effects

In this section I first provide a short introduction to binary outcome models with unobserved heterogeneity and define the quantities of interest. Second, I present the identification strategy, estimation, and asymptotic properties of PF-CRE. Third, I discuss alternative estimation approaches and compare them to PF-CRE. Finally, I present a model specification test for PF-CRE.

2.1 Binary Outcome Models with Unobserved Heterogeneity

A binary outcome model with unobserved heterogeneity consists of a binary response, y_{it}, and a k-dimensional vector of time-varying characteristics, x_{it}, such that the response for individual i at time t is generated by:

$$y_{it} = \mathbb{I}[\alpha + x_{it}\beta + c_i + \varepsilon_{it} > 0], \quad i = 1, .., n, \quad t = 1, .., T, \tag{1}$$

where $\mathbb{I}(A)$ is an indicator function that takes the value of one if A holds and zero otherwise, α is a constant, β is a k-dimensional parameter vector, c_i is the unobserved heterogeneity that is constant over time; and ε_{it} is an individual- and time-specific error.[7]

When the error terms are independently and identically distributed according to a known cumulative distribution $G(\cdot)$, equation (1) can be alternatively written as:

$$Prob(y_{it} = 1|x_{it}, c_i) = G(\alpha + x_{it}\beta + c_i). \tag{2}$$

[7]The focus on a balanced panel is for simplicity. One can allow for T to differ across individuals i.

Typical choices of $G(\cdot)$ are the normal distribution, which gives the probit model, or the logistic distribution, which gives the logit model.[8] Although some of the results and estimators discussed throughout this paper apply only to the logistic distribution, I use the general notation $G(\cdot)$ wherever more general cases are valid.

In some applications researchers may only be interested in the sign and relative sizes of the β coefficients. In many other cases, the interest lies in the partial effects that reflect how the probability of the outcome changes with respect to the covariates x. In the presence of unobserved heterogeneity these partial effects are calculated by taking expectations over c. The partial effects for the model in equation (2) are defined as:

$$PE_j(x) = E\left[\frac{\partial}{\partial x_j} G(\alpha + x\beta + c)\middle| x\right], \tag{3}$$

where x_j and β_j denote the jth elements of x and β. Additionally, researchers may be interested in the average partial effect, defined as:

$$APE_j = E[PE_j(x)] = E\left[\frac{\partial}{\partial x_j} G(\alpha + x\beta + c)\right], \tag{4}$$

where the last expectation is taken with respect to both x and c.[9]

2.2 Identification, Estimation, and Asymptotic Properties

The identification problem in the model of equation (2) lies in c_i being unobserved *and* correlated with x_{it}.[10] The identification strategy I use in this paper is to specify a distribution for c_i conditional on $(x_{i1}, .., x_{iT})$ without imposing excessively strong restrictions on the unobserved heterogeneity. I begin with the following assumption:

[8] The assumption of identically distributed errors implicitly imposes homoskedasticity.

[9] Note that some authors refer to equation (3) as the *average* partial effect, as it is averaging over the distribution of the unobserved heterogeneity. However, researchers also use the term average partial effect for equation (4). To avoid confusion, I reserve the term average partial effect of equation (4).

[10] When c is independent of x, it is known as a random effect.

Assumption 1 (Exchangeability).

$$f(c_i|x_{i1},..,x_{iT}) = f(c_i|x_{is_1},...,x_{is_T}) \text{ where } s_j \in \{1,...,T\},\ s_j \neq s_{j'}.$$

Assumption 1 requires that the distribution of the unobserved heterogeneity conditional on the observed covariates, $f(c_i|x_{i1},...,x_{iT})$, does not depend on the order in which x_{it} enters the function $f(c_i|\cdot)$. Under Assumption 1, without loss of generality, $f(c_i|x_{i1},...,x_{iT})$ can be written as a polynomial on $z_i^1,...,z_i^T$, where $z_i^t = \sum_{s=1}^{T}(x_{is})^t$ (Altonji and Matzkin, 2005).[11] Note that $(z_i^1,...,z_i^T)$ are in fact the first T non-central moments of $(x_{i1},...,x_{iT})$.

In most circumstances, researchers also observe time-invariant information, w_i, about each individual i, such as gender, race, and year of birth. These time-invariant characteristics can be added to the conditional distribution of c_i to improve fit. Moreover, the inclusion of these auxiliary variables can help the exchangeability assumption hold.

Assumption 1 is not sufficient for identification. The reason is that the first T non-central moments characterize the T observations per individual i, exhausting the degrees of freedom. Additional restrictions are necessary for identification:

Assumption 2 (Linear Index). *The conditional distribution $f(c_i|z_i^1,...,z_i^T,w_i)$ depends on a linear index of $(z_i^1,...,z_i^\tau,w_i)$ and interaction terms, for some $\tau < T$. That is,*

$$f(c_i|z_i^1,...,z_i^T,w_i) = f(c_i|z_i\gamma),$$

where z_i is the vector of the first τ moments, the observed time-invariant characteristics w_i, and interaction terms.

[11] The Weierstrass approximation theorem establishes that a function on bounded support can be uniformly approximated by a polynomial function. Because of exchangeability, this is a symmetric polynomial. By the fundamental theorem of symmetric polynomials, it may be written as a polynomial in the power functions (i.e. the moments). See Altonji and Matzkin (2005, p. 1062) and references therein for further details.

Under Assumption 2, I restrict attention to a linear index of the first τ moments of $(x_{i1}, ..., x_{iT})$, observed time-invariant characteristics, and interaction terms.[12] This implies a stronger condition than exchangeability alone, but it maintains sufficient flexibility to capture (or approximate) the conditional distribution of the unobserved heterogeneity.

Assumption 3 (Normality). *$f(c_i|\cdot)$ is a normal distribution with variance σ^2*

In order to obtain parametric identification, it is necessary to specify a distribution for the unobserved heterogeneity, hence Assumption 3. However, other distributions are possible, as long as they have finite moments.[13]

Estimation

Imposing Assumptions 1, 2, and 3 in the model in equation (2), results in the following specification:

$$Prob(y_{it}|x_{it}, c_i) = G(\alpha + x_{it}\beta + z_i\gamma + \eta_i), \text{ with } \eta_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \tag{5}$$

where z_i is a vector of moments of $(x_{i1}, ..., x_{iT})$, observed time-invariant characteristics, and interaction terms among these; while, η_i is a normally distributed random effect with variance σ^2, that is independent of the covariates in the model.[14]

In principle, the parameters β in equation (5) can be estimated via Maximum Likelihood. The log-likelihood function for this model is:

$$\log L(\beta, \alpha, \gamma, \sigma) = \sum_{t=1}^{T} \sum_{i=1}^{n} \Big[y_{it} \log(p_{it}) + (1 - y_{it}) \log(1 - p_{it}) \Big], \tag{6}$$

with

$$p_{it} \equiv Prob(y_{it}|x_{it}) = \int_{-\infty}^{\infty} G(\alpha + x_{it}\beta + z_i\gamma + \eta_i) \frac{1}{\sigma} \phi(\eta_i/\sigma) d\eta_i. \tag{7}$$

[12] Interaction terms, in fact, represent a polynomial function.
[13] Finite moments are required, because otherwise, expectations are not well defined.
[14] Independence follows from Assumptions 1 and 2, and normality from Assumption 3.

where $\phi(\cdot)$ is the standard normal density function.

The model in equation (5) represents a *Flexible* specification of a Correlated Random Effects (CRE) model. It is a CRE model as it assumes a specific correlation form between the unobserved heterogeneity and the covariates in the model (represented by $z_i\gamma$). It is flexible because, under Assumptions 1 and 2, it can accommodate a wide range of correlation forms.

The flexible specification derived from Assumptions 1 and 2 requires the estimation of additional coefficients (γ). When the number of covariates is small, γ is relatively low dimensional. However, the dimensionality of γ increases exponentially with the number of covariates in the model. With 3 covariates, a simple specification of z_i that includes the time-means of the covariates and up to three-way interactions requires the estimation of 20 additional parameters.[15] The same type of specification with 5 covariates requires the estimation of 56 additional parameters; with 10 covariates, 286 parameters. Moreover, the inclusion of time-invariant characteristics exacerbates this problem.

To address the dimensionality issue introduced by the flexible specification I use *penalized* Maximum Likelihood estimation. That is, I estimate β using *Penalized Flexible Correlated Random Effects* (PF-CRE), which is defined by:

$$(\hat{\beta}, \hat{\alpha}, \hat{\gamma}, \hat{\sigma}) = \arg \max_{(\beta,\alpha,\gamma,\sigma)} \quad \log L(\beta, \alpha, \gamma, \sigma) - \Pi_\lambda(\gamma), \tag{8}$$

where $\Pi_\lambda(\cdot)$ is a penalty function that penalizes only the terms used to model the unobserved heterogeneity (γ), but not the parameters associated with the observed covariates (β). I use the Smoothly Clipped Absolute Deviation (SCAD) penalty (Fan

[15] 3 time means, 6 two-way interactions, 10 three-way interaction, and the variance of the random effect.

and Li, 2001), defined as:

$$\Pi_\lambda(\gamma) = \begin{cases} \lambda|\gamma| & \text{if } |\gamma| \leq \lambda, \\ -\frac{|\gamma|^2 - 2a\lambda|\gamma| + \lambda^2}{2(a-1)} & \text{if } \lambda < |\gamma| \leq a\lambda, \\ \frac{(a+1)\lambda^2}{2} & \text{if } |\gamma| > a\lambda, \end{cases} \tag{9}$$

where a and λ are constants that govern the penalization. The SCAD penalty shrinks small values of γ towards zero, while leaving larger values of γ mostly unpenalized. This way, SCAD selects those terms in z_i that are most predictive of the outcome and discards those that are not. Importantly, the shrinkage introduced by the SCAD penalty does not affect the coefficients of interest, β, directly since they are left unpenalized.

Asymptotic Properties

The PF-CRE estimator with the SCAD penalty produces consistent, efficient, and asymptotically normal estimates of the model parameters, β. I state this result in the following Theorem 1 for easy reference:

Theorem 1. *Under assumptions 1, 2, and 3,*

$$(\hat{\beta} - \beta) \longrightarrow \mathcal{N}(0, I(\beta)^{-1}),$$

where $I(\beta)$ is the Fisher information matrix.

Theorem 1 follows from standard properties of Maximum Likelihood estimation and the Oracle property of the SCAD penalty. The Oracle property of SCAD establishes that the penalized estimator has the same asymptotic distribution as the underlying (and unknown) data generating process (Ibrahim et al., 2011; Hui et al., 2017). Consequently, it has the same asymptotic properties of the Maximum Likelihood estimator of the data generating process. Consistency, efficiency, and normality of the PF-CRE estimator thus follow from the properties of Maximum Likelihood estimators.[16]

[16]The asymptotic properties of Maximum Likelihood estimation hold under a number of regularity

The next result establishes that the PF-CRE estimates of partial effects are also consistent:

Corollary 1. *Under assumptions 1, 2, and 3, partial effects are identified, and for all x:*

$$\widehat{PE}_j(x) \equiv \int_{-\infty}^{\infty} g(\hat{\alpha} + x\hat{\beta} + z\hat{\gamma} + \eta)\frac{1}{\hat{\sigma}}\phi(\eta/\hat{\sigma})\hat{\beta}_j d\eta \xrightarrow{p} PE_j(x),$$

where $g(\cdot)$ is the probability density function of $G(\cdot)$.

The Oracle properties of SCAD guarantee that $z\hat{\gamma}$ is a consistent estimator of $z\gamma$. Corollary 1 follows from this and Theorem 1 by direct application of the continuous mapping theorem.[17] Standard errors for the partial effects can be obtained via the Delta method or bootstrap.

To estimate the partial effects, it is necessary to to specify a value of z. In principle, any value of z is valid for estimating the partial effects. However, a significant proportion (or all) of the terms in z are functions of x. For this reason, it is advisable to ensure that the values of x and z used to calculate the partial effects are consistent with one another, to avoid issues similar to those of extreme counterfactuals (King and Zeng, 2006). For example, suppose x represents individuals' ideology, and z corresponds to the average ideology of each individual across the panel waves. If we want to estimate the effect of changing x from liberal to very liberal, then the value z should also correspond to a liberal (or very liberal) individual. Although using a value of z corresponding to a conservative is technically correct, inferences in this case would rely heavily on extrapolation from the model.[18]

conditions, which the PF-CRE model satisfies.

[17]The continuous mapping theorem states that continuous functions are limit-preserving. Therefore, a continuous function evaluated at a consistent estimator is a consistent estimator of the value of that function.

[18]This is because individuals who report a liberal ideological position in a wave, but have generally reported to be a conservative in other waves, are rare or non-existent.

2.3 Relation to Existing Estimators

As previously mentioned, there are three main strategies for the estimation of binary outcome models with panel data in the presence of unobserved heterogeneity. I briefly discuss each of them and how they relate to the PF-CRE estimator I develop in this paper.[19]

The first approach is estimation via Fixed Effects (FE), where the c_is are treated as parameters to be estimated. This is operationalized through dummy variables for each individual in the sample. When the panel is short (small T), this requires estimating each dummy with a handful of observations, a problem known as the incidental parameters problem (first noted by Neyman and Scott, 1948). The incidental parameters problem implies that estimates from the FE approach are inconsistent for small T. This asymptotic bias can be substantial. For example, simulations in Greene (2004) show that with $T = 5$, this bias can be 40% of the true parameter value.[20,21]

In light of the inconsistency of the FE estimator, bias correction procedures have been proposed.[22] These corrections reduce the bias; however, they do not eliminate it.[23] A related strand of literature seeks to ameliorate the incidental parameters problem (as well as the computational burden of estimating $n + k$ parameters) by assuming that the individual heterogeneity is in fact group heterogeneity.[24] However, these grouped fixed-effects estimators also suffer from the incidental parameters problem and may not be appropriate for short panels.

[19]See Greene (2015) for a review of the literature on parametric estimation of discrete choice models.

[20]In the case of $T = 2$, Abrevaya (1997) shows that the maximum likelihood estimates of β using the FE approach converge to 2β. Thus, dividing the FE estimate by 2 results in a consistent estimate of β. However, the incidental parameters problem persists in the estimation of partial effects.

[21]The asymptotic bias is of order $O_p(T^{-1})$, meaning that it disappears as T tends to infinity. Monte Carlo evidence in Heckman (1981) suggests that this bias is negligible for a panel of size $T = 8$, although more recent studies in Coupe (2005) suggest that a larger size of $T = 16$ is preferable.

[22]See, for example, Fernandez-Val (2009); Fernandez-Val and Vella (2011); Hahn and Newey (2004); Dhaene and Jochmans (2015).

[23]In fact Dhaene and Jochmans (2015) show that the elimination of the leading term of the bias leads to larger magnitudes of the higher order terms of the bias in the bias-corrected estimator.

[24]See, for example, Bonhomme and Manresa (2015); Ando and Bai (2016); Su et al. (2016). Bonhomme et al. (2017) do not assume group heterogeneity, but assume that the heterogeneity can be coarsened into groups.

The second approach is estimation via Conditional Maximum Likelihood (CMLE), which results in consistent estimates of β (Rasch, 1961; Andersen, 1970; Chamberlain, 1984). This approach relies on conditioning the estimation only on those individuals with variation in the outcome across time. By restricting the estimation to these individuals, the conditional likelihood only depends on β and not the unobserved heterogeneity c_i, eliminating the incidental parameters problem. However, this property only holds for the logistic distribution.[25]

The CMLE approach has two main shortcomings. First, it does not provide estimates of the partial effects.[26] This is because location parameters are not estimated (in fact, β is estimated by eliminating the location parameters c_i from the likelihood function). The second shortcoming is inefficiency. The CMLE approach allows the heterogeneity to be completely unrestricted, which implicitly assumes that individuals with no variation in the outcome provide no information about β. However, if the heterogeneity has a less general form, conditioning on these individuals results in a loss of information, and consequently larger standard errors in the estimates.[27,28]

The third approach is estimation via Correlated Random Effects (CRE). This approach requires making explicit assumptions about the unobserved heterogeneity. The strongest restriction is assuming that the heterogeneity is independent of the covariates in the model, leading to the Random Effects (RE) model. Mundlak (1978) proposes to model the unobserved heterogeneity as a linear combination of the time-means of the covariates and a random effect, which allows for correlation between the model

[25]Chamberlain (2010) shows that if the support of the observed predictor variables is bounded, then identification is only possible in the logistic case. Moreover, if the support is unbounded, the information bound is zero unless the distribution is logistic. This means that consistent estimation at the standard asymptotic rates is only possible in the logistic case. For alternative semi-parametric estimators that require unbounded support, see Manski (1987); Abrevaya (2000).

[26]This is also a problem with semi-parametric alternatives to CMLE.

[27]Note that CMLE's conditoning on those individuals with variation in the outcome can also introduce errors if this subpopulation behaves differently than the overall population, beyond the unobserved heterogeneity. However, an implicit assumption in this paper is that despite the presence of unobserved heterogeneity, individuals' behavioral rules are the same. That is, they all have the same β.

[28]Note that the FE approach results in the same kind of information loss without discarding observations outright. The behavior of individuals with no variation in the outcome is fully explained by the dummy variables corresponding to these individuals. Thus, these individuals do not contribute to the estimation of the model parameters β (see, for example, Beck and Katz, 2001).

covariates and the unobserved heterogeneity.[29]

The main advantage of CRE is that, by providing an explicit model of the unobserved heterogeneity, it allows for the estimation of partial effects. However, it does so at the cost of severely restricting the unobserved heterogeneity. When this restriction is not satisfied by the data generating process (which is unobserved), CRE models are misspecified and provide incorrect estimates of the model parameters and partial effects.

The PF-CRE estimator represents a compromise between the unrestricted unobserved heterogeneity that FE and CMLE allow for and the restrictive assumptions underlying CRE models. I achieve this compromise through the exchangeability assumption proposed in Altonji and Matzkin (2005), which leads to a flexible specification of the unobserved heterogeneity. This flexible specification can capture a wide range of correlation forms between the unobserved heterogeneity and the observed covariates in the model.

If the exchangeability assumption holds, the PF-CRE estimator has several advantages relative to the FE and CMLE approaches. Unlike the FE approach, it does not suffer from the incidental parameters problem. It also allows for the estimation of probabilities and partial effects that the CMLE approach cannot provide. Finally, PF-CRE also provides more efficient estimates of the model parameters than FE and CMLE. This is because FE and CMLE account for every form of correlation between the covariates and the unobserved heterogeneity, even when it is not necessary. PF-CRE, on the other hand, selects the minimal specification for this correlation that is necessary to control for the unobserved heterogeneity, leading to efficiency gains. In other words, FE and CMLE assume there is no information in cross-sectional variation. PF-CRE allows cross-sectional variation to be informative of the parameter vector β when the estimated specification of the unobserved heterogeneity is sufficiently sparse (i.e. when

[29] Chamberlain (1980) proposes a more general version of Mundlak's model, modeling the unobserved heterogeneity by projecting the time dimension of the model into one dimension. This is akin to a weighted mean of the covariates across time.

few γ parameters are non-zero).

2.4 Specification Test

The method outlined in the previous section relies on the assumption that the unobserved heterogeneity in the data can be appropriately captured through the flexible correlation specification by the $z\gamma$ terms. This assumption will not necessarily hold in every application. Therefore, I present a model specification test for one of the most used models in applied research: the logistic case.

If the correlation between the observed and unobserved components of the model can be correctly captured by the $z\gamma$ terms, then the PF-CRE estimator proposed in this paper is both consistent and efficient. The Oracle property of the penalized estimator plays a crucial role here, as it ensures that the penalized model asymptotically attains the same information bound as the Oracle estimator, which is efficient.

For the logistic case, the CMLE estimator provides a consistent estimator of the model parameters. Under the null hypothesis that the unobserved heterogeneity can be sufficiently captured by the PF-CRE specification, the PF-CRE estimator is both consistent and efficient, whereas the CMLE estimator is consistent but inefficient. Under the alternative hypothesis, the PF-CRE estimator is inconsistent, but the CMLE estimator remains consistent.[30] Following Hausman (1978), I construct an specification test based on the standardized squared difference between these two estimators. That is, the test statistics is defined as:

$$\delta = d'V(d)^{-1}d, \quad \text{with } \; d = \hat{\beta}_{CMLE} - \hat{\beta}_{PF-CRE}, \tag{10}$$

where $V(d)$ is the variance of d.

Under the null hypothesis, δ is asymptotically distributed χ^2 with k degrees of

[30]The reason the test is restricted to the logistic case is that CMLE is consistent only for the logistic case. Semi-parametric alternatives to CMLE provide consistent estimates of the model parameters for any distribution. However, the convergence rate of these estimators is slower than $\sqrt{n}$. For this reason, asymptotic comparisons with the PF-CRE estimator, which converges at rate $\sqrt{n}$, are not well defined.

freedom. This is because both estimators are asymptotically normal with identical means under the null hypothesis, and therefore their difference d, is asymptotically normal with mean zero. The $\chi^2_{(k)}$ distribution follows from δ being the sum of k normally distributed terms.

The only challenging part of the specification test is the estimation of the variance, $V(d)$. Under the null hypothesis, this variance has a simple expression thanks to the efficiency of the PF-CRE estimator:[31]

$$V(d) = V(\hat{\beta}_{CMLE}) - V(\hat{\beta}_{PF-CRE}). \tag{11}$$

Hence, putting equations (11) and (10) together:

$$\delta = \left(\hat{\beta}_{CMLE} - \hat{\beta}_{PF-CRE}\right)' \left(V(\hat{\beta}_{CMLE}) - V(\hat{\beta}_{PF-CRE})\right)^{-1} \left(\hat{\beta}_{CMLE} - \hat{\beta}_{PF-CRE}\right). \tag{12}$$

3 Simulations

I conduct three simulation studies to analyze the performance of the PF-CRE estimator in small samples and compare it to that of alternative methods. I use the Oracle estimator as a benchmark for comparison. The Oracle estimator is the Maximum Likelihood estimate from the exact data generating process. In the first set of simulations I analyze the Root Mean Squared Error (RMSE) of PF-CRE and CMLE estimates of β relative to the Oracle. In the second set, I compare the estimates of the Partial Effects (PEs) from PF-CRE, the traditional CRE specification from Mundlak (1978), and an *un*penalized version of PF-CRE, that I call UF-CRE.[32] The final set of simulations studies the specification test for PF-CRE for different sample sizes.

[31] Hausman (1978) shows that the variance the difference between two consistent estimators when one of them is efficient, is the difference of the variances.

[32] Mundlak (1978)'s specification of CRE uses the time-means of the covariates to model the unobserved heterogeneity. The UF-CRE estimator uses the same specification as PF-CRE but without the penalized estimation step.

The data generating process in all simulations is the following:

$$
\begin{aligned}
&P(y_{it} = 1|x_{it}, c_i) = \Lambda(x_{it}\beta + c_i), \text{ with } x_{it} \in \mathbb{R}^5, \\
&c_i|\mathbf{x}_i \sim \mathcal{N}(\mu_i, \sigma_c^2), \\
&\mu_i = \alpha + \gamma_1 \bar{x}_{i1} + \gamma_2 \bar{x}_{i2} + \gamma_3 \overline{x_{i1} x_{i2}}, \\
&\beta = (0.7, 1.3, -0.4, 1.2, -0.1), \quad \alpha = 0.2, \quad \gamma = (0.5, 0.6, 1.2),
\end{aligned} \tag{13}
$$

where $\Lambda(\cdot)$ is the logistic cumulative distribution, $x_{it} \sim \mathcal{N}(0, \Sigma)$ with Σ non-diagonal, and $\bar{x}_{ij}$ denotes the time-mean of x_{itj}, where j denotes the jth variable in x_{it}. While this is a relatively simple DGP, it has several characteristics that make it useful to assess the properties of the PF-CRE estimator. First, the conditional distribution of the unobserved heterogeneity depends on an interaction term that the traditional CRE estimator does not include, making this estimator underspecified and inconsistent. Second, the conditional distribution of the unobserved heterogeneity is sparse. This implies that CMLE makes assumptions about the heterogeneity that are too general for this case, leading to inefficiency. The sparsity of the specification also helps illustrate the powerful gains from the penalization step. Finally, the covariates in the model are not independent from one another, and therefore neither are the terms in the conditional distribution of the unobserved heterogeneity, which imposes a certain level of difficulty for the variable selection step.[33]

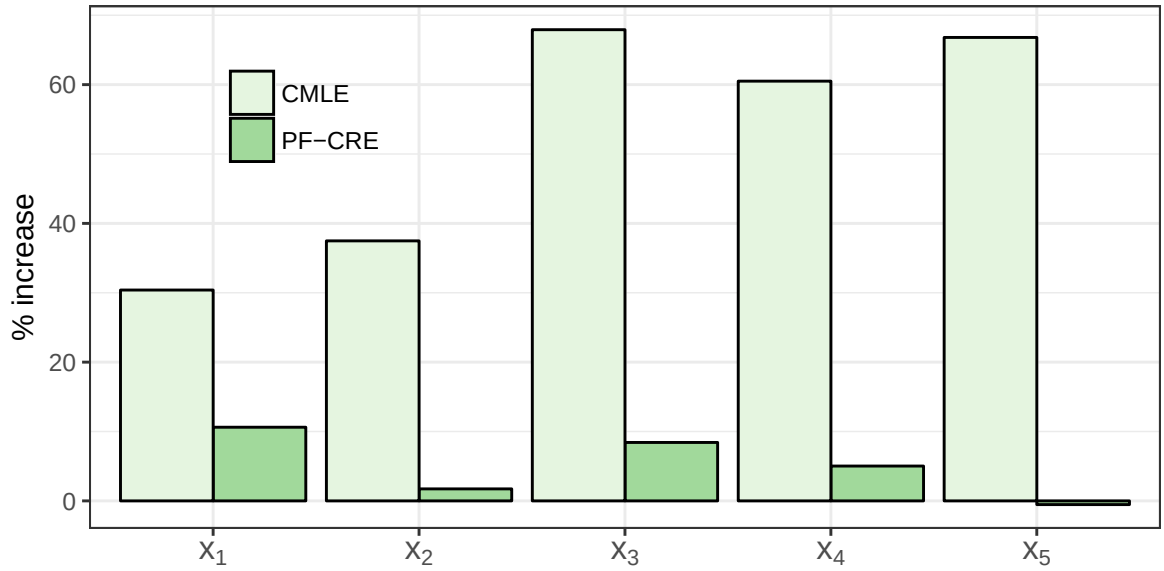
For all simulations $T = 2$. For the first two sets, $n = 1500$, whereas for the specification test simulations, I use an n size of 1500, 3000, 4500, and 6000. All results are based on 1000 draws from the model in equation (13).

[33]Intuitively, if there are two and only one of them belongs in the data generating process, the penalization procedure has a harder time determining which one of them predicts the outcome when they are correlated than when they are independent. The correlations used here are of moderate size (up to 0.5).

3.1 Parameter Estimates

The DGP in equation (13) satisfies the assumptions of both the CMLE and PF-CRE estimators, and therefore the estimates of β from both estimators are consistent. Figure 1 shows the RMSE of the CMLE and PF-CRE estimates of β relative to the RMSE of the Oracle estimator. Because both estimators are consistent, the differences in the relative RMSEs come mostly from the variance of the estimators.[34] As expected, given that the unobserved heterogeneity in the DGP is not completely unrestricted, the CMLE estimator produces less efficient estimates than the PF-CRE approach. In fact, the CMLE approach produces RMSEs that are 35% to almost 70% higher than those of the Oracle estimator. The RMSEs of the PF-CRE approach are at most 10% higher than those of the Oracle. This illustrates the efficiency gains of this estimator relative to the CMLE estimator, and the Oracle properties of PF-CRE.

Figure 1: *CMLE produces less efficient $\hat{\beta}$ than PF-CRE*



Bars represent the increase in RMSE from the PF-CRE and CMLE estimators relative to the RMSE of the Oracle estimator.

Figure A1 in the appendix shows the distribution of the estimates for the 5 parameters in β from PF-CRE, CRE, and the Oracle estimators. The distribution of the

[34]Both estimators have a slight bias in small samples. The simulations show that this bias is smaller for the PF-CRE than the CMLE estimator. See Figure A1 in the appendix.

CMLE estimates is generally centered around the true value of the coefficients, although there is some bias in some cases (due to the small sample size). The distribution of the estimates from the PF-CRE and Oracle estimators are almost identical, reflecting the Oracle properties of the PF-CRE estimator.[35]

3.2 Partial Effects

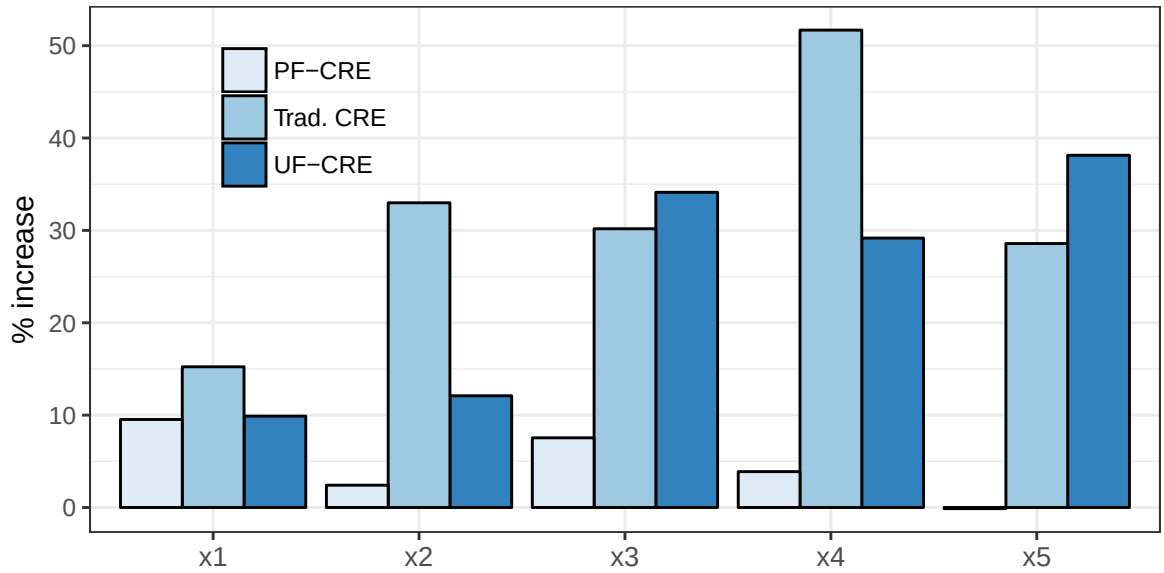
Here I compare the Partial Effects (PEs) for the DGP in equation (13) estimated via the PF-CRE approach, the traditional CRE approach, and the UF-CRE (i.e. the unpenalized PF-CRE).

Figure 2 shows the RMSE of the three approaches relative to that of the Oracle estimator. The RMSE of the PF-CRE approach is the lowest, and is at most 10% above the Oracle's. The traditional CRE, in turn, produces estimates with RMSEs that can be more than 50% higher than the Oracle's. This occurs because the CRE approach includes terms that do not belong in the data generating process for the unobserved heterogeneity (such as the time-means for x_3, x_4, and x_5) and it excludes terms that do belong there (the interaction term $\overline{x_1 x_2}$) which leads to inconsistent and inefficient estimates. The UF-CRE approach also produces estimates with a RMSE that can be 40% higher than the Oracle's. This is a reflection of the inefficiency of the unpenalized approach approach, as it includes 71 parameters in the model whereas the GDP has only 9. Via penalizetion, PF-CRE selects only relevant terms, which significantly reduces the variance of the estimator.

Figure A2 in the appendix shows that the distribution of the estimates from PF-CRE and UF-CRE are centered around the true value of the PEs. But UF-CRE has a larger variance, a consequence of including a plethora of terms that do not belong in the DGP. The traditional CRE, in turn, is generally biased, a consequence of missing the interaction term in the conditional mean equation.

[35]The Oracle property of the SCAD estimator holds asymptotically. Therefore, while asymptotically we should expect the RMSE of the PF-CRE approach to be identical to that of the Oracle, they will differ slightly in small samples.

Figure 2: *Average Partial Effects: PF-CRE v UF-CRE v Traditional CRE*



The bars represent the increase in RMSE of each estimator relative to the RMSE of the Oracle estimator. UF-CRE refers to estimation using all the variables included in the PF-CRE approach but without the variable selection step. Trad. CRE refers to the model that only includes the time-means of the covariates to model the unobserved heterogeneity.

3.3 Model Specification Test

Using the same setting as for the previous simulations, I calculated the rejection rate of the model specification test in equation (12) for four different sample sizes (1500, 3000, 4500, and 6000) at the 95% level. For each sample size, I drew 1000 samples of the DGP in equation (13). Table 1 shows that the rejection rate of the (true) null hypothesis that the PF-CRE estimator is consistent and more efficient than the CMLE estimator is close to the theoretical 5%, although it errs slightly on the conservative side for the smaller sample sizes.

Figure A3 in the appendix shows quantile-quantile plots, where the x-axis represents the quantiles from the simulations, and the y-axis the quantiles from the theoretical distribution of the test (in this case a $\chi^2_{(5)}$). The quantile-quantile plots show that the empirical quantiles of the test statistic are similar to their theoretical counterparts.[36]

[36]Deviations for the larger values are expected as many more simulations would be necessary for accurate representations of the tail of the distribution, as values above 10 occur with a probability of only 7.5%.

Table 1: *Simulations: Specification Test*

N	Rejection Rate
1500	0.056
3000	0.053
4500	0.052
6000	0.049

Rejection rate calculated as the percentage of p-values smaller than 5% *from* 500 *simulations for each sample size.*

Overall, the simulations show that the PF-CRE estimator produces estimates of the model parameters that are more efficient than those of the CMLE estimator when the data generating process for the unobserved heterogeneity satisfies the assumptions of PF-CRE. In addition, the simulations also illustrate the advantages of the PF-CRE estimator in the estimation of partial effects. They show that the flexibility of its specification gives it a significant advantage over the traditional correlated random effects, and that the penalization step can help to significantly reduce the uncertainty around the estimated quantities. Finally, the simulations also show that the specification test has rejection rates close to theoretical expectations, ensuring that the decision to use the PF-CRE estimator will be on solid ground.

4 Tactical Voting in the 2015 United Kingdom General Election

In elections with more than two candidates, voters often cast tactical votes. That is, when they believe their most preferred candidate is unlikely to win, they often vote for a less preferred candidate with chances of winning, if only to prevent their most disliked one from being elected (Duverger, 1954).[37]

The literature on tactical voting has generally focused on measuring its extent, but less on why some voters behave tactically while others do not. In this application, I

[37] I use the term tactical voting instead of strategic voting, as it is the common denomination used for this behavior in Britain.

focus on the effect that being contacted by political parties has on voters' propensity to cast a tactical vote. The empirical challenge lies in correctly identifying the effect of party contact itself, independent of the effect of unobserved confounders. In particular, parties possibly contact the voters that they believe are more likely to respond to the parties' message or appeals. However, researchers do not observe how parties decide which voters to contact.[38]

To address this challenge, I use a panel data survey collected prior to the 2015 United Kingdom General Election. Controlling for the unobserved heterogeneity using PF-CRE, allows me to reduce or eliminate the concerns outlined in the previous paragraph. In particular, the unobserved heterogeneity modeled by PF-CRE captures voters' overall characteristics and behavioral tendencies, which are likely closely related to what parties use to decide which voters to contact. My results show that voters who are contacted by their most preferred party are 7% less likely to cast a tactical vote, indicating that party contact can enforce loyalty and sincerity at the polling booth. However, being contacted by the most preferred viable party increases the probability of a tactical vote by 10%. These effects are overestimated when ignoring unobserved heterogeneity, almost by a factor of two.

In the rest of this section, I first provide a short review of the literature on tactical voting. I then describe the data and model specification. Finally, I present the estimates from the PF-CRE and CMLE estimators, together with estimates from pooled logit, which ignores unobserved heterogeneity.

4.1 Related Literature

The study of tactical voting has generally focused on establishing its existence and in estimating its extent. Available evidence from around the world shows that, typically, around 15 to 40 percent of voters who are in a position to cast a tactical vote actually

[38]Ideally, disentangling the effects of party contacts from that parties choosing who, to contact can be done by relying on field experiments, in the spirit of Gerber et al. (2008) for voter turnout. However, while an experimental intervention in a real election aimed at increasing voter turnout may be relatively uncontroversial, one aimed at altering voters' choices faces significant moral dilemmas.

do so (see Alvarez et al., 2018, for a review on this subject). In British elections, this figure is estimated at around 30 to 40%, with variations across elections (Alvarez et al., 2006; Kieweit, 2013). The importance of tactical voting for overall election outcomes has also been documented in literature. For example, Kieweit (2013) finds that as many as one in five Labour seats are won thanks to tactical votes by Liberal Democrat voters; and Spenkuch (2017) estimates that about one in ten seats in the German Bundestag would change hands if all voters were to cast sincere ballots.[39]

Despite the attention that measuring tactical voting has received by empirical researchers in the last two decades, there is less understanding on why and when some voters cast tactical votes when given the opportunity to do so, while others do not. The literature shows that the closeness of the race between the top two contenders has little or no effect on tactical voting (Lanoue and Bowler, 1992; Fisher, 2000; Kieweit, 2013; Elff, 2014; Núñez, 2016), despite theoretical expectations that it should (Cox, 1997); strong partisan and ideological attachments discourage tactical voting (Blais, 2002; Lanoue and Bowler, 1992; Karp et al., 2002)[40]; political sophistication and knowledge (sometimes proxied by education levels) are positively associated with tactical voting (Alvarez et al., 2006; Gschwend and van der Kolk, 2006; Karp et al., 2002)[41], as is experience with the electoral system (Spenkuch, 2017; Duch and Palmer, 2002). There is also evidence that the presence of a close ideological substitute to a non-viable preferred party encourages tactical voting (Karp et al., 2002), that the incumbency advantage interferes with the decision to cast a tactical vote (Moser and Scheiner, 2005), and that voters who believe the media influences the voting decisions of others are more likely to behave tactically (Cohen and Tsfati, 2009).

Most of the determinants of tactical voting that the literature identifies are relevant to the study of voter behavior, but they do not provide with actionable recommen-

[39]Note that Germany uses a compensatory mixed member system. Therefore, while seats would change hands if all voters were sincere, the partisan composition of the Bundestag would remain almost the same.

[40]Similar effects have been found in the study of split ticket voting in the U.S. (Burden and Kimball, 1998; Beck et al., 1992)

[41]However, Blais et al. (2006) find no evidence that the most informed were more likely to cast tactical coalition votes in Israeli elections than the least informed.

dations. In particular, they provide no guidance to a party or group who wants to encourage (or discourage) tactical voting behavior.[42] In this application, I focus on one actionable factor, the effect of party contact on the probability of casting a tactical vote. The goal of this application is not only to determine the existence of such effects, but also to quantify their magnitude.

4.2 Data and Model Specification

To study the effect of party contact on the probability of casting a tactical vote, I use data from six waves of the British Election Study Online Panel. These data were collected prior to the 2015 United Kingdom General Election.[43] I restrict the sample to respondents that reported vote intention and party preferences in at least two of the six waves. This leaves 2973 respondents for a total of 10014 observations.[44] I impute missing values for other variables using the package `mice` in `R` (Buuren and Groothuis-Oudshoorn, 2011).

The outcome of interest is whether a voter intends to cast vote that does not match her most preferred party. I define voters' most preferred party in the following way: (1) the party with the highest thermometer score; (2) if there are ties, these are broken by the thermometer scores for the leaders of the corresponding parties; (3) if ties remain, these are broken by party identification, only when identification is reported to be strong;[45] (4) all remaining ties are kept.[46,47] I define voters' most preferred *viable* party as the most preferred party from among the viable ones (note that when the most preferred party is viable, then the most preferred and the most preferred viable party

[42]That is, barring redistricting efforts aimed at selecting voters who are more sophisticated or have weak partisan attachments.

[43]The study covers England, Scotland, and Wales, but excludes Northern Ireland.

[44]1006 individuals are observed in two waves; 753 in three; 530 in four; 481 in five; and 203 in all six.

[45]Respondents are allowed to report different party identifications in each wave.

[46]In these cases, a tactical vote for these voters only occurs when the reported choice does not match either of their most preferred (tied) parties.

[47] There have been multiple criticisms to defining party preferences using feeling thermometers (see, for example Alvarez and Nagler, 2000). Data limitations preclude me from defining preferences in alternative ways.

is the same).

The covariates of interest are indicators for whether the voters' most preferred party or the voters' most preferred viable party contacted the voter during the four weeks prior to each survey wave. As in any study of tactical voting, I include an indicator for whether the voters' most preferred party, as reported in each wave, is considered to be out of the race in the corresponding constituency.[48] I also include as dependent variables the thermometer score for the most preferred and most preferred viable parties as reported by each respondent, measured on a scale from 1 to 10. Finally, I include a number of time-invariant characteristics that serve as control variables in pooled logit estimates and also as additional terms to model the conditional distribution of the unobserved heterogeneity in the PF-CRE estimator. Among these, I include self identification as middle or working class, educational attainment, employment status (full time, part time, retired, student, or unemployed), race (asian, black, mixed, or other non-white race), age, gender, and home ownership (owned outright, mortgaged, or rent).[49]

To model the correlation between the unobserved heterogeneity and the covariates of interest in the PF-CRE estimator, I use the time-means of the five main covariates of interest, plus the time-invariant characteristics, and three-way interactions among them, for a total of 1796 terms. Given that I use the logistic distribution in this application, I compare coefficient estimates from the PF-CRE estimator with those of the Conditional Maximum Likelihood estimator (CMLE). While both PF-CRE and CMLE account for unobserved heterogeneity, only PF-CRE allows for the estimation of partial effects. Additionally, I estimate a pooled logit that includes the time-invariant characteristics as additional controls, which does not account for unobserved heterogeneity. I compare

[48]Candidate viability can be defined in a number of ways. Here I define a party as not viable if it finished third or lower in the 2015 election in that constituency. Alternatively, the results of the previous election can be used for viability, research finds that voters rely on the electoral history heuristic (see, for example Lago, 2008). Finally, voters' own assessments of the chances of each party to win the seat in their constituency can be used to determine viability. Unfortunately, questions about this were only included in only two of the waves leading up to the 2015 election, strongly restricting data availability.

[49]Note that some of these variables might change in time. However, it is unlikely that they change within the short time frame analyzed in this application.

coefficient and partial effect estimates from the pooled logit and PF-CRE estimator to show the discrepancies that arise when ignoring the unobserved heterogeneity.

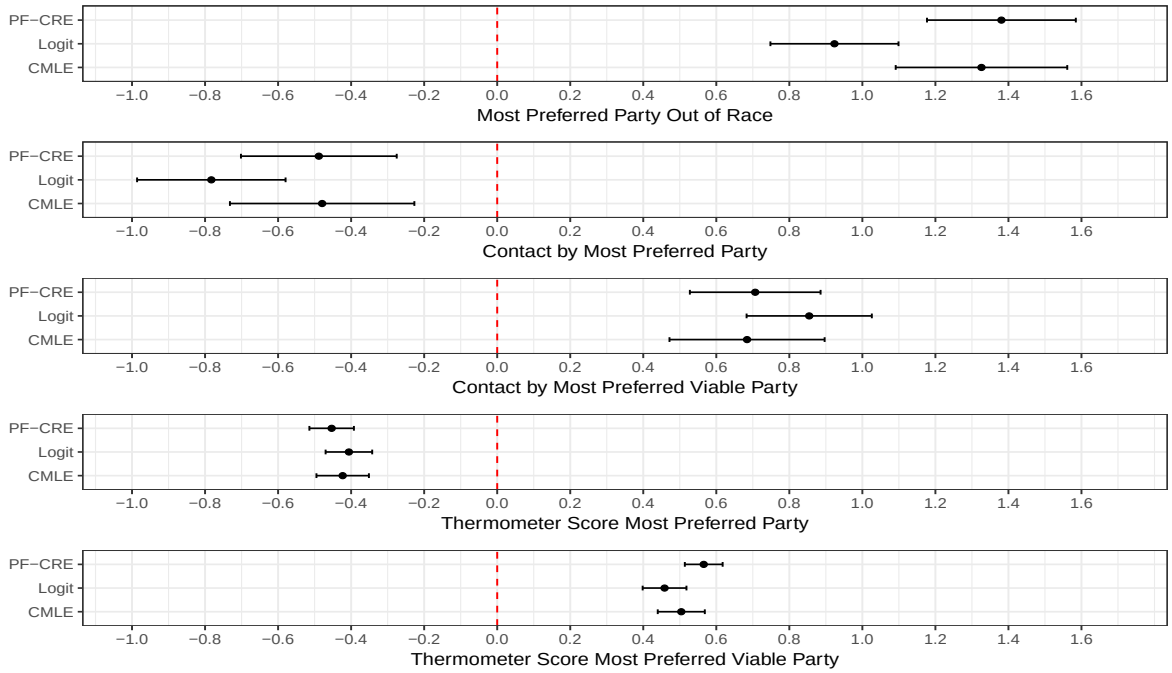
4.3 Results

Figure 3 shows that coefficient estimates from PF-CRE and CMLE are very similar to one another.[50] Indeed, the specification test does not reject the null hypothesis that PF-CRE is consistent and more efficient than CMLE, with a p-value of 0.59. This clearly establishes the validity of the PF-CRE approach in this case. Importantly, PF-CRE allows me to estimate partial effects that CMLE cannot estimate. It is also clear from Figure 3 that the pooled logit model underestimates the effect of viability of the most preferred party and overestimates the effects of being contacted by the most preferred and most preferred viable parties on the decision to cast a tactical vote. Estimates for the thermometer scores do not present significant differences between the three estimators.

Why does pooled logit overestimate the effects of party contacts, and underestimates the effect of viability of the most preferred party? In principle, unobserved heterogeneity is in fact unobserved, and researchers can only speculate as to its sources. In the case of party contacts, it is possible that candidates from viable parties in a given constituency tend to contact supporters of non-viable parties that they believe are likely to defect their preferred party and vote tactically. At the same time, candidates from non-viable parties may be more likely to contact potential defectors from among their supporters as a way to prevent their numbers from dropping. This implies that the voters that parties contact are those who are more likely to cast a tactical vote. However, researchers do not get to observe the way parties decide which voters to contact. Therefore, when ignoring the unobserved heterogeneity (like the pooled logit does) the coefficient estimates for party contacts capture both the effects of contact itself plus the selection effect just described. This selection effect can also explain why pooled logit underestimates the

[50] See Table B1 the appendix for details with the estimates from the three models, as well as a traditional CRE specification.

Figure 3: *Coefficient, Tactical Voting in 2015 U.K. General Election*



PF-CRE estimates were obtained using the SCAD penalty. The tunning parameter λ was obtained through 10-fold cross valuation using the Akaike information criterion. Logit standard errors are clustered by respondent.

effect of candidate viability on the probability of casting a tactical vote. As parties contact those voters who are more likely to switch, they are also contacting those voters whose preferred party is more likely to be out of the race. The overestimation of the effect of party contacts then necessarily implies that the effect of candidate viability is underestimated.[51]

Accounting for the unobserved heterogeneity, as the PF-CRE estimator does, controls for the selection effect introduced by the way parties choose voters for contact. This reduces or eliminates the bias introduced by this selection effect, as it captures voters overall characteristics, which are likely related to how parties choose which voters to contact.

The terms that the SCAD penalty selects to model the unobserved heterogeneity

[51]There are, of course, other possible explanations for these effects. For example, the underestimation of the effect of candidate viability might simply reflect the presence of standpatters, perhaps inspired by the same moral reasons against tactical voting that Pliny the Younger expected of Roman senators (see Farquharson, 1964, for an account of this episode).

provide some intuition as to how PF-CRE can account for this selection effect.[52] Among the surviving terms, there are the time-means of party contacts. When this mean is positive, it indicates that the voter was contacted by the party *at some point* during the time frame covered by the survey, but it does not indicate *in which wave* this happened. Therefore, the time-mean serves as a control for the party's decision to contact that voter, helping resolve the selection issue. The, the actual contact, in whichever wave occurs, is free from the selection effect.

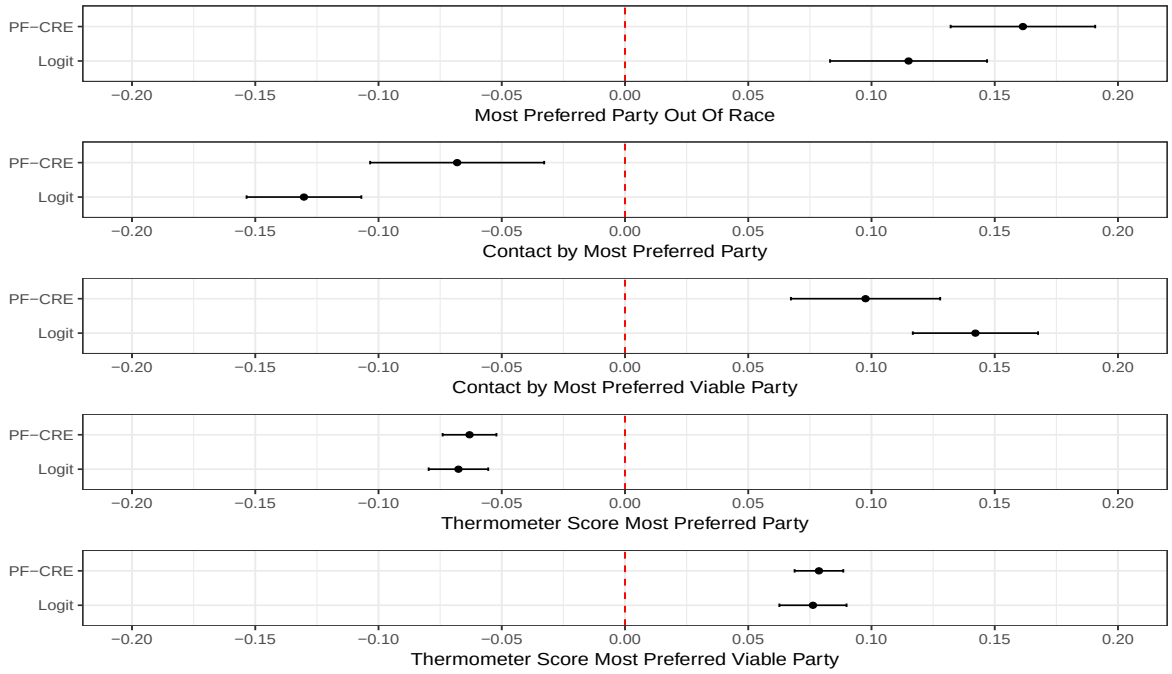
Figure 4 presents the partial effects for the pooled logit and PF-CRE estimators. While the CMLE and PF-CRE coefficient estimates are indistinguishable from one another, only the PF-CRE estimator provides estimates of probabilities and partial effects. To calculate the partial effects I use a baseline individual who is a 45 years old white male, with some college education or more, who is employed full time, identifies as middle class, whose most preferred party is out of the race, and who has thermometer scores and has been contacted by his most preferred and most preferred viable parties as much as is average for a respondent with these demographic characteristics.

The PF-CRE estimates show that when the baseline respondents' party is out of the race, he is 15.5% more likely to cast a vote for a less preferred viable party. Pooled logit underestimates this effect, at 11.5 percentage points, which is a significant underestimation of respondents' tactical motivations at the polling booth.

In terms of party contacts, PF-CRE estimates that respondents who are contacted by their most preferred party are 6.6% less likely cast a vote for a less preferred party, suggesting that party contact enforces party loyalty or sincerity in voters. Logit estimates this quantity at 13%, almost double the effect. Interestingly, being contacted by the most preferred viable party has a countervailing effect that is stronger than being contacted by the most preferred party, increasing the probability of casting a a vote for a less preferred party by 9.6%. Logit also overestimates this effect, in this case by 4.6 percentage points.

[52]For the full list of terms selected to model the unobserved heterogeneity, see Table B2 in the appendix.

Figure 4: *Partial Effects, Tactical Voting in 2015 U.K. General Election*



PArtial effects are calculated for a baseline individual. Baseline values for the conditional mean equation were chosen to be consistent with those of the observed characteristics in the baseline individual. Logit standard errors are clustered by respondent.

The results presented here show that unobserved heterogeneity is an important confounder in the study of tactical voting during the 2015 U.K. General Election./ This is evidenced by the significant over and underestimation of different effects when the heterogeneity is ignored. The PF-CRE estimator allows for the estimation of partial effects when accounting for the unobserved heterogeneity that other estimators cannot, and the results show that parties' efforts to contact voters during the pre-election season have significant effects on the probability that voters cast a tactical vote. In particular, voters are 9.6% more likely to cast a tactical vote when contacted by their most preferred viable party, and 6.6% less likely, when contacted by their most preferred party. This shows that parties can benefit significantly from contacting voters as a way to encourage or discourage them from voting tactically.

5 Additional Applications

In this section I present a very brief discussion of two additional applications: the effect of preferences for immigration and economic fears on voting decisions in the 2016 Brexit Referendum in the U.K.; and (2) the effect of ideological preferences and candidate personality perceptions on vote choice during the 2012 U.S. presidential election. The goal of this section is to show that unobserved heterogeneity matters in these contexts and that PF-CRE provides consistent estimates of the model parameters. Further details and discussion for both these applications are available in Appendix C.

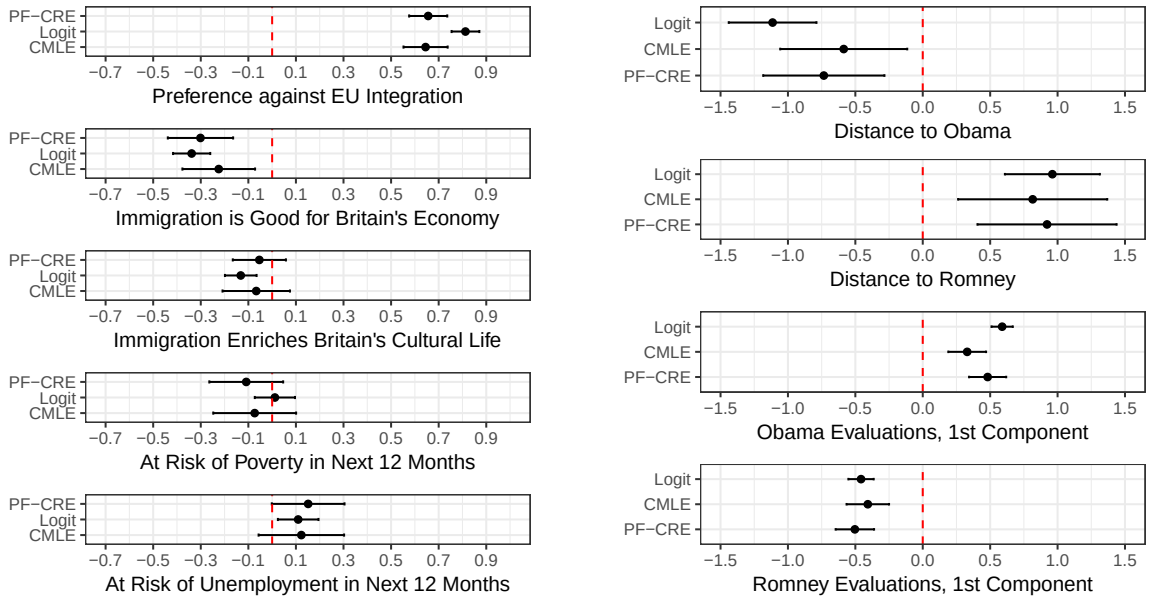
In the Brexit Referendum case, the outcome of interest is voting in favor of Brexit. The covariates of interest are preferences against European integration, views on immigration as it relates to British culture and economy, and fears of falling into poverty or unemployment in the coming year. I model the conditional distribution of the unobserved heterogeneity with a total of 1258 terms (of which 19 are selected by the penalization step). The specification test for PF-CRE returns a p-value of 0.074, which provides mixed statistical evidence for its validity. However, coefficient estimates from PF-CRE are similar to those of CMLE, and have smaller standard errors (see Figure 5a). Importantly, pooled logit, overestimates some effects, and provides excessively small confidence intervals for other variables.

In the case of the 2012 U.S. Presidential Election, the outcome of interest is voting for Barack Obama. The covariates are respondents' ideological distances to Barack Obama and Mitt Romney, and personality evaluations about the candidates. I model the conditional distribution of the unobserved heterogeneity in PF-CRE with 210 terms. The specification test supports the PF-CRE specification, with a p-value of 0.97. As Figure 5b shows, this is reflected in the similar coefficient estimates from PF-CRE and CMLE, with PF-CRE estimates generally having a slightly smaller variance. Pooled logit coefficients, on the other hand, overestimate the effects of personality evaluations and distance to Barack Obama.

Put together, the main application to tactical voting in Britain, plus the two ap-

Figure 5: *Coefficient Estimates*

(a) *Brexit Referendum* ***(b)*** *2012 Presidential Election*



PF-CRE estimates were obtained using the SCAD penalty. The tunning parameters was obtained through 5-fold cross validation using the Akaike information criterion. Logit standard errors are clustered by respondent.

plications briefly described in this section, show that PF-CRE is a valid alternative to estimating binary outcome models with unobserved heterogeneity. PF-CRE's value is two-fold: (1) it provides for consistent estimates of the model parameters *and* partial effects, which estimators like CMLE cannot estimate; and (2) it provides more efficient estimates (albeit sometimes only slightly more efficient).

6 Conclusion

Unobserved heterogeneity is pervasive in observational studies in political science, and the social sciences in general. Whatever its origin and form, all unobserved heterogeneity poses the same problem: if ignored, and correlated with the covariates of interest, it leads to biased and inconsistent estimates. One of the best ways to deal with unobserved heterogeneity is to use panel data. However, a standing problem in the case of binary outcomes (and discrete outcomes generally) is that consistent estimators of

model parameters do not allow for the estimation of partial effects, which are usually the quantity of interest to researchers.

In this paper, I develop the *Penalized Flexible Correlated Random Effects* (PF-CRE) estimator for binary outcome models with panel data. PF-CRE provides for consistent and efficient estimates of the model parameters and partial effects. It relies on adopting a flexible specification for the unobserved heterogeneity that is complemented with a penalization step for variable selection. The flexibility imposes weak assumptions on the unobserved heterogeneity, and the penalization induces a parsimonious model that results in efficiency gains. While the assumptions in PF-CRE restrict unobserved heterogeneity to some extent, I show in three different applications that these assumptions hold in these cases, using a model specification test.

The PF-CRE estimator has a number of advantages relative to alternative estimators. Unlike Fixed Effects, it does not suffer from the incidental parameters problem that leads to inconsistent estimates. PF-CRE allows for the estimation of partial effects that the Conditional Maximum Likelihood estimator does not provide. Finally, its assumptions are significantly less restrictive than those of traditional Correlated Random Effects models, meaning that PF-CRE's assumptions are more likely to hold in real world applications.

The main application I provide for the PF-CRE estimator is to tactical voting during the 2015 U.K. General Election. I show that ignoring unobserved heterogeneity leads to overestimation of the effect of being contacted by the most preferred and most preferred viable parties on the probability of casting a tactical vote, by as much as a factor of two. The intuition behind this overestimation is that parties possibly know something about voters, that researchers do not observe, that makes them more attractive for proselytizing. This makes party contacts correlated with these unobserved factors, leading to biased estimates.

I also provide two additional applications on electoral behavior, one to vote choice during the 2016 U.S. Presidential Election, and the other to vote choice during the

2016 Brexit Referendum in the U.K. In both these cases, the assumptions of the PF-CRE estimator hold, and alternative estimators produce upward or downward biased estimates of the partial effects of interest. While the validity of PF-CRE must be determined on a case by case basis, these results suggest that is is feasible in a number of applications.

PF-CRE can be applied in other areas of social science beyond political behavior. An area where PF-CRE can be a significant methodological contribution is to the study of comparative political institutions and international relations. In this type of data, most of the variation is usually across units; within unit variation is typically much smaller. For this reason, methods like CMLE and Fixed Effects tend to discard almost all of the information in the data, leading to mostly statistically non significant results. The appeal of PF-CRE in these cases is that, while it accounts for unobserved heterogeneity, it does not discard all cross-sectional variation. This is accomplished via the penalization step: if it selects a relatively sparse specification for the unobserved heterogeneity, a significant portion of cross-sectional variation will still be used to estimate the parameters of interest and partial effects.

Moving forward, a number of extensions of PF-CRE can be considered. The most natural, are extensions to discrete outcome models other than binary. Commonly used multinomial and ordered response models (like Conditional Logit and Ordered Probit) can incorporate unobserved heterogeneity in the form of correlated random effects (Wooldridge, 2010). However, the penalization step in these cases requires some refining. As in the binary case, allowing for a flexible specification made parsimonious with a penalization step can help these models realistically capture the unobserved heterogeneity, without leading to inefficient estimates or very restrictive assumptions.

Another extension is to allow for the model coefficients, and the coefficients in the conditional distribution of the unobserved heterogeneity, to vary by individual in the form of random coefficients. Random coefficients can be powerful tools to capture unobserved heterogeneity (independent of the covariates). An extension in this direction

can exploit recent developments in penalized estimation of generalized linear mixed models, the category under which these models fall (see Hui et al., 2017).

References

Abrevaya, J. A. (1997). The Equivalence of Two Estimators of the Fixed-Effects Logit Model. *Economic Letters*, 55:41–43.

Abrevaya, J. A. (2000). Rank Estimation of a Generalized Fixed Effects Regression Model. *Journal of Econometrics*, 95:1–23.

Aldrich, J. H. and McKelvey, R. D. (1977). A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections. *American Political Science Review*, 71(1):111–130.

Altonji, J. G. and Matzkin, R. L. (2005). Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors. *Econometrica*, 73(4):1053–1102.

Alvarez, R. M., Boehmke, F. J., and Nagler, J. (2006). Strategic Voting in British Elections. *Electoral Studies*, 25:1–19.

Alvarez, R. M. and Brehm, J. (2002). *Hard Choices, Easy Answers: Values, Information, and American Public Opinionb*. Princeton University Press, New Haven.

Alvarez, R. M., Kieweit, D. R., and Núñez, L. (2018). Preferences, Constraints, and Choices. In Fisher, J., Fieldhouse, E., Franklin, M. N., Gibson, R., Cantijoch, M., and Wlezien, C., editors, *Routledge Handbook of Elections, Voting Behavior and Public Opinion*. Elsevier.

Alvarez, R. M. and Nagler, J. (2000). A New Approach for Modelling Strategic Voting in Multiparty Elections. *British Journal of Political Science*, 30(1):57–75.

Andersen, E. B. (1970). Asymptotic Properties of Conditional Maximum-Likelihood Estimators. *Journal of the Royal Statistical Society, Series B (Methodological*, 32(2):283–301.

Ando, T. and Bai, J. (2016). Panel Data Models with Grouped Factor Structure under Unknown Group Membership. *Journal of Applied Econometrics*, 31(1).

Ansolabehere, S., Rodden, J., and Snyder, J. M. (2008). The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting. *American Politial Science Review*, 102:215–232.

Beck, N. and Katz, J. N. (2001). Throwing out the Baby with the Bath Water: A Comment on Green, Kim, and Yoon. *International Organization*, 55:487–495.

Beck, P. A., Baum, L., Clausen, A. R., and Smith, C. E. (1992). Patterns and Sources of Ticket Splitting in Subpresidential Voting. *American Political Science Review*, 86:916–928.

Becker, S. O., Fetzer, T., and Novy, D. (2017). Who Vote for Brexit? A Comprehensive District-Level Analysis. CEP Discussion Paper No. 1480.

Blais, A. (2002). Why is there So Little Strategic Voting in Canadian Plurality Rule Elections? *Political Studies*, 50:445–454.

Blais, A., Aldrich, J. H., Indridason, I. H., and Levine, R. (2006). Do Voters Vote For Government Coalitions. *Party Politics*, 12(6):691–705.

Bonhomme, S., Lamadon, T., and Manresa, E. (2017). Discretizing Unobserved Heterogeneity. Wokring paper.

Bonhomme, S. and Manresa, E. (2015). Grouped Patterns of Heterogeneity in Panel Data. *Econometrica*, 83(3):1147–1184.

Brinegar, A., Jolly, S., and Kitschelt, S. (2004). Varieties of Capitalism and POlitical Divides over European Integration. In Marks, G. and Steenbergen, M., editors, *European Integration and Political Conflict*, pages 62–89. Cambridge University Press, Cambridge.

Burden, B. C. and Kimball, D. C. (1998). A New Approach To the Study of Ticket Splitting. *American Political Science Review*, 45:533–544.

Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3):33–48.

Chamberlain, G. (1980). Analysis of Covariance with Qualitative Data. *Review of Economic Studies*, 47:225–238.

Chamberlain, G. (1984). Panel Data. In Griliches, Z. and Intriligator, M., editors, *Handbook of Econometrics*, volume II, chapter 22. North-Holland, Amsterdam.

Chamberlain, G. (2010). Binary Response Models for Panel Data: Identification and Information. *Econometrica*, 78(1):159–169.

Cohen, J. and Tsfati, Y. (2009). The Influence of Presumed Media Influence on Strategic Voting. *Communication Research*, 36:359–378.

Coupe, T. (2005). Bias in Conditional and Unconditional Fixed Effects Logit Estimation: A Correction. *Political Analysis*, 13:292–295.

Cox, G. W. (1997). *Making Votes Count*. Cambridge University Press, Cambridge, MA.

Curtice, J. (2016). A Question of Culture of Economics¿? Public Attitudes to the European Union in Britain. *The Political Quarterly*, 87(2):209–218.

Dhaene, G. and Jochmans, K. (2015). Split-Panel Jackknife Estimation of Fixed-Effect Models. *Review of Economic Studies*, 82:991–1030.

Duch, R. M. and Palmer, H. D. (2002). Strategic Voting in Post-Communist Democracy? *British Journal of Political Science*, 32:63–91.

Duverger, M. (1954). *Political Parties: Their Organization and Activity in the Modern State*. Wiley, New York.

Elff, M. (2014). Separating Tactical from Sincere Voting: A Finite Mixture Discrete Choice Modelling Approach to Disentangling Voting Calculi. Paper presented at the 2014 Annual Meeting of the Midwest Political Science Association, Chicago, April 3-6.

Fan, J. and Li, R. (2001). Variable Selection Via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association*, 96:1348–1360.

Farquharson, R. (1964). *A Theory of Voting*. Yale University Press, New Haven.

Feldman, S. (1988). Structure and Consistency in Public Opinion: the Role of Core Beliefs and Values. *American Journal of Political Science*, 32(2):416–440.

Fernandez-Val, I. (2009). Fixed Effects Estimation of Structural Parameters and Marginal Effects in Panel Probit Models. *Journal of Econometrics*, 150:71–85.

Fernandez-Val, I. and Vella, F. (2011). Bias Corrections for Two-Step Fixed Effects Panel Data Estimators. *Journal of Econometrics*, 163:144–162.

Fisher, S. (2000). Intuition versus Formal Theory: Tactical Voting in England: 1987-1997. Paper presented at the Annual Meeting of the American Political Science Association, Washington, DC.

Garry, J. and Tilley, J. (2015). Inequality, State Ownership and the EU: How Economic Context and Economic Ideology Shape Support for the EU. *European Union Politics*, 16(1):139–154.

Gerber, A. S., Green, D. P., and Larimer, C. W. (2008). Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment. *American Political Science Review*, 102(1):33–48.

Goren, P. (2005). Party Identification and Core Political Values. *American Journal of Political Science*, 49(4):881–896.

Greene, W. (2015). Panel Data Models for Discrete Choice. In Baltagi, B. H., editor, *The Oxford Handbook of Panel Data*. Oxford University Press.

Greene, W. H. (2004). The Behavior of the Fixed Effects Estimator in Nonlinear Models. *Econometrics Journal*, 7:98–119.

Gschwend, T. and van der Kolk, H. (2006). Split Ticket Voting in Mixed Member Proportional Systems: The Hypothetical Case of The Netherlands. *Acta Politica*, 41:163–179.

Hahn, J. and Newey, W. (2004). Jackknife and Analytical Bias Reduction for Nonlinear Panel Models. *Econometrica*, 72:1295–1319.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, 46(6):1251–1271.

Heckman, J. (1981). The Incidental Parameters Problem and the Problem of Initial Conditions in Discrete Time-Discrete Data Stochastic Process. In Manski, C. and McFadden, D., editors, *Structural Analysis of Discrete Data with Econometric Applications* . MIT Press, Cambridge.

Hui, F. K., Muller, S., and Welsh, A. H. (2017). Hierarchical Selection of Fixed and Random Effects in Generalized Linear Mixed Models. *Statistica Sinica*, 27:501–518.

Ibrahim, J. G., Zhu, H., Garcia, R. I., and Guo, R. (2011). Fixed and Random Effects Selection in Mixed Effects Models. *Biometrics*, 67(2):495–503.

Inglehart, R. F. and Norris, P. (2016). Trump, Brexit, and the Rise of Populism: Economic Have-Nots and Cultural Backlash. HKS Faculty Research Working Paper Series, RWP16-026.

Karp, J. A., Vowles, J., Banducci, S. A., and Donovan, T. (2002). Strategic Voting, Party Activity, and Candidate Effects: Testing Explanations for Split Voting in New Zealand's New Mixed System. *Electoral Studies*, 21:1–22.

Kieweit, D. R. (2013). The Ecology of Tactical Voting in Britain. *Journal of Elections, Public Opinion and Parties*, 23(1):86–110.

King, G. and Zeng, L. (2006). The Dangers of Extreme Counterfactuals. *Political Analysis*, 14:131–159.

Lago, I. (2008). Rational Expectations or Heuristics? Strategic Voting in Proportional Representation Systems. *Party Politics*, 14:31–49.

Lanoue, D. J. and Bowler, S. (1992). The Sources of Tactical Voting in British Parliamentary Elections, 1983-1987. *Political Behavior*, 14(2):141–157.

Manski, C. (1987). Semiparametric Analysis of Random Effects Linear Models From Binary Panel Data. *Econometrica*, 55:357–362.

McCann, J. A. (1997). Electoral Choices and Core Value Change: The 1992 Presidential Campaign. *American Journal of Political Science*, 41(2):564–583.

McLaren, L. M. (2002). Public Support for the European Union: Cost/Benefit Analysis or Perceived Cultural Threat? *Journal of Politics*, 87(2):209–218.

Moser, R. G. and Scheiner, E. (2005). Strategic Ticket Splitting and the Personal Vote in Mixed-Member Electoral Systems. *Legislative Studies Quarterly*, 30(2):259–276.

Mundlak, Y. (1978). On the Pooling of Time Series and Corss Section Data. *Econometrica*, 46:69–85.

Neyman, J. and Scott, E. (1948). Consistent Estimates Based on Partially Constistent Observations. *Econometrica*, 16:1–32.

Núñez, L. (2016). Expressive and Strategic Behavior in Legislative Elections in Argentina. *Political Behavior*, 38(4):899–920.

Poole, K., Rosenthal, H., Lewis, J., Lo, J., and Carroll, R. (2013). `basicspace`: A Package to Recover a Basic Space from Issue Scales.

Rasch, G. (1961). On General Laws and the Meaning of Measurement in Psychology. *The Danish Institute of Educational Research, Copenhagen.*

Shipman, T. (2016). *All Out War: The Full Story of How Brexit Sank Britain's Political Class.* William Collins, London.

Spenkuch, J. L. (2017). Expressive vs. Pivotal Voters: An Empirical Assessment. Wokring paper.

Su, L., Shi, Z., and Phillips, P. C. B. (2016). Identifying Latent Structures in Panel Data. *Econometrica*, 84(6):2215–2264.

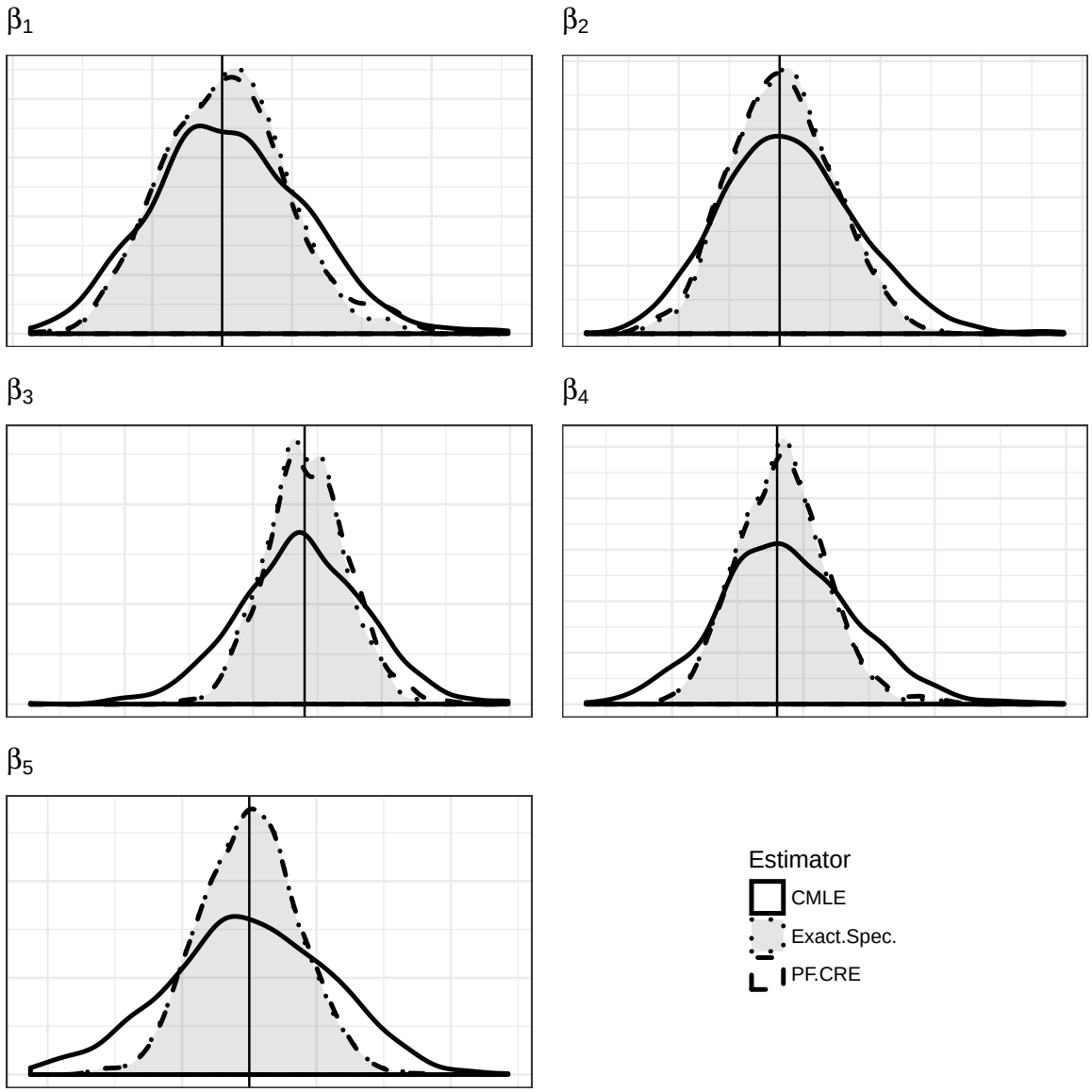
Tucker, J. A., Pacek, A. C., and Berinsky, A. J. (2002). Transitional Winners and Losers: Attitudes Toward EU Membership in Post-Communist Countries. *American Journal of Political Sciene*, 46(3):557–571.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data.* MIT Press, Cambridge, MA, second edition.

Appendices

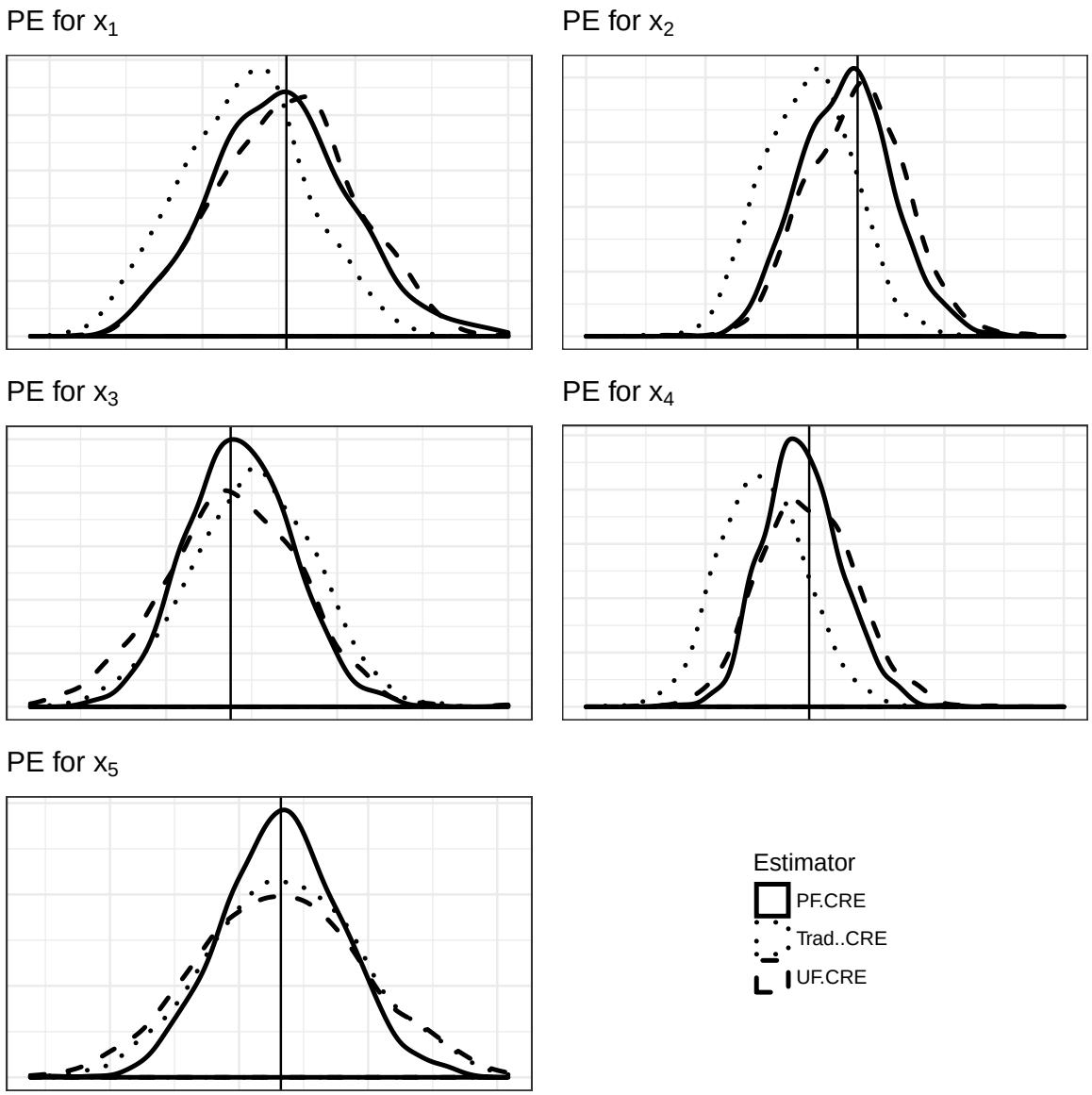
A Additional Figures from Simulations

Figure A1: *Simulations Distributions $\hat{\beta}$ PF-CRE v CMLE*



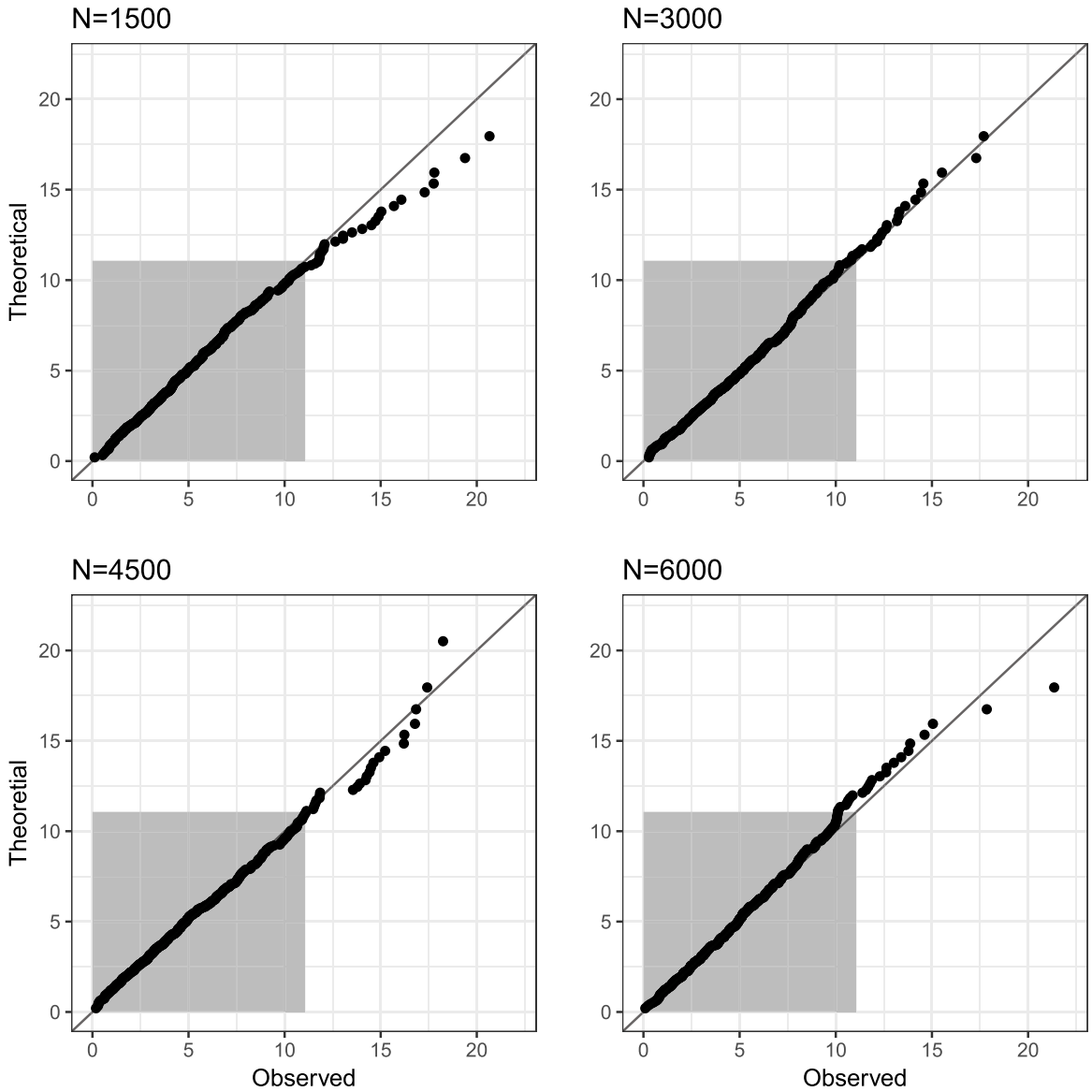
Lines represent the distribution of the estimates for each parameter and estimator from the Conditional Maximum Likelihood estimator (CMLE), the PF-CRE, and the exact data generating process. Vertical lines represent the true value of the parameters.

Figure A2: *Simulations PEs: PF-CRE v UF-CRE v Traditional CRE*



Lines represent the distribution of the estimates for each PE for each variable at the mean. Vertical lines represent the true value of the PEs.

Figure A3: *Simulations Specification Tests: Quantile-Quantile plots*



Observed represents the sample quantiles from the simulation runs. Theoretical represents the theoretical quantiles from a $\chi^2_{(5)}$ *distribution. Plots constructed from 500 simulations for each sample size. The shaded area represents 95% theoretical quantile.*

B Additional Tables and Figures, Tactical Voting

Table B1: *Coefficient Estimates for Tactical Voting in Britain*

	PF-CRE		CMLE		Trad. CRE		Logit	
	β	CI	β	CI	β	CI	β	CI
Contact Preferred	−0.49	(−0.70, −0.28)	−0.48	(−0.74, −0.23)	−0.51	(−0.77, −0.26)	−0.78	(−0.99, −0.58)
Contact Viable	0.71	(0.53, 0.89)	0.68	(0.49, 0.92)	0.53	(0.47, 0.59)	0.85	(0.68, 1.03)
Like Preferred	−0.45	(−0.51, −0.39)	−0.42	(−0.53, −0.38)	−0.45	(−0.52, −0.37)	−0.41	(−0.47, −0.34)
Like Viable	0.57	(0.51, 0.62)	0.50	(0.44, 0.57)	0.53	(0.46, 0.59)	0.46	(0.40, 0.52)
Out of Race	1.38	(1.17, 1.58)	1.33	(1.14, 1.62)	1.47	(1.23, 1.72)	0.92	(0.75, 1.10)
Controls	No		No		No		Yes	
N^o γ terms	1797		-		5		-	
Selected γs	36		-		5		-	
n ($n \times T_i$)	2973		2973		2973		2973	
Effective n	2973		958		2973		2973	
Observations ($n \times T_i$)	10014		10014		10014		10014	
Effective Obs	10014		3530		10014		10014	
$\chi^2_{(5)}$	5.04		-		-28.62		53.76	
p-value	0.59		-		NA		0.00	

All confidence intervals are at the 95%. Logit standard errors are clustered at the individual level. The effective n and effective number of observations refers to the number of actual observations used in estimation for CMLE. There is no χ^2 test reported to CMLE since this estimator is the basis for that test.

Table B2: *Terms Selected by SCAD Penalization*

Contact MP	**Contact MPV**
Thermometer MP	**Thermometer MPV**
Viable×**Thermometer MP**	**Viable**×Education (less than secondary)
Thermometer MP×**Contact MP**	**Thermometer MP**×Education (less than Secondary)
Thermometer MP×Education (GCSE)	**Thermometer MP**×Education (6th Form)
Thermometer MPV×**Contact MPV**	**Contact MP**×Education (less than Secondary)
Contact MP×Mixed Race	**Contact MP**×Mortgaged Home
Contact MPV×Middle Class	**Contact MPV**×Retired
Contact MPV×Education (6th Form)	Middle Class×Student
Middle Class×Education (GCSE)	Working Class×Employed Full Time
Working Class×Education (6th Form)	Working Class×Asian
Working Class×Own Home	Employed Full Time×Mixed Race
Employed Full Time×Other Race	Employed Part Time×Education (less than Secondary)
Employed Part Time×Mortgaged Home	Student×Education (6th Form)
Student×Black	Student×Own Home
Student×Mortgage Home	Education (GCSE)×Mixed Race
Education (GCSE)×Other Race	Education (GCSE)×Own Home
Middle Class×Other Race×Education (GCSE)	Middle Class×Retired×Own Home

Bold indicates time-means. MP denotes the most preferred party. MPV denotes the most preferred viable party.

C Additional Applications

Brexit Referendum

During the 2015 British General Election, internal struggles within the Conservative party lead Prime Minister David Cameron to promise a referendum of EU membership (Becker et al., 2017). In the run up to the Brexit Referendum, held on June 23rd, 2016, many arguments were presented for leaving the European Union. Some of them had to do with ensuring British independence from bureaucrats in Brussels or with preventing taxpayer money from lining up the Euro-coffers. In fact, the Leave campaign stressed that by leaving the EU, Britain would save £350 million each week.[53] Other arguments had to do with immigration (both from the EU and from other countries as a result of EU policy) and its effects on the British economy and culture. Research suggests that hostility towards the European Union has been fueled by the perception that EU membership represents a cultural threat (McLaren, 2002; Curtice, 2016; Inglehart and Norris, 2016). On the economic side, voters with fears of losing employment or of their economic wellbeing being negatively affected by EU policy, were expected to be more favorable towards the British exit from the European Union.[54]

In this application, I focus on whether fears of falling into poverty or unemployment affected voters' decision to support or oppose Brexit. Estimating these effects in a causal manner is not trivial, however. Notably, these economic fears may be more prevalent among certain groups of the population that, at the same time, are more (or less) likely to support Brexit for other reasons, some of which may be unobserved.

Data and Model Specification

I use panel data survey from the British Election Study Online Panel, collected prior to the Brexit referendum. These data allow me to study how changes in individuals'

[53]For an account of the Brexit referendum campaign, see Shipman (2016)

[54]There is a relatively large literature that focuses on an utilitarian approach to European integration (see, for example Tucker et al., 2002; Brinegar et al., 2004; Garry and Tilley, 2015).

fears of poverty and unemployment played out in their referendum vote decisions.

The main variables of interest indicate respondents' beliefs that in the next 12 months they will fall into poverty or unemployment, both on a scale from 1 to 5. I also include respondents' overall preferences against European integration, on a scale from 0 (unite fully with the European Union) to 10 (protest our independence).[55] I also include two additional questions about attitudes towards immigration. The first one, measures respondents' beliefs on whether immigration is good for Britain's economy, on a scale from 1 (bad) to 7 (good); the second, measures respondents' beliefs on whether immigration enriches Britain's cultural life, on a scale from 1 (undermines cultural life) to 7 (enriches cultural life). I also include a number of time-invariant characteristics: identification as middle or working class, age, gender, race, education level, employment status, household income, and indicator for whether the respondent has ever lived abroad, has friends from EU countries, and whether his/her parents were born in a foreign country.

I model the conditional distribution of the unobserved heterogeneity using PF-CRE with the time-means of the covariates of interest, plus the time-invariant characteristics, and up to three-way interactions among these terms, for a total of 1258 terms. I compare the estimates from PF-CRE to those of Conditional Maximum Likelihood (CMLE), a pooled logit estimator that includes the time-invariant characteristics as controls, and a traditional CRE approach that only uses the time-means of the covariates to model the conditional distribution of the unobserved heterogeneity.

Results

Table C1 presents the coefficient estimates from the four methods considered. The coefficient estimates from PF-CRE and CMLE are similar, with PF-CRE having smaller confidence intervals. The specification test of the null hypothesis that PF-CRE is con-

[55]Some respondents were assigned a different version of this question, on a scale from 0 (unification has already gone too far) to 10 (unification should be pushed further). Results change slightly when using this version of the question. However, the qualitative (and to a large extent) quantitative results remain the same. These results are available by request.

sistent and more efficient than CMLE returns a p-value of 0.074, which implies the validity of PF-CRE, albeit with weak statistical evidence. The traditional CRE approach overestimates some effects, but underestimates others. The specification test for CRE rejects the null hypothesis that it is a consistent estimator of the model parameters.[56] Pooled logit, that ignores unobserved heterogeneity, overestimates some effects, particularly the coefficient on preferences against European integration. Pooled logit also estimates a significant effect of believing that immigration enriches Britain's cultural life. However, this effect dissapears when accounting for the unobserved heterogeneity as in PF-CRE and CMLE.

The estimates of being at risk of unemployment highlight the efficiency advantage of PF-CRE relative to CMLE. While PF-CRE and CMLE provide very similar point estimates, the inefficiency of CMLE incorrectly leads to the conclusion that there is no statistically significant effect of fears of unemployment on voting for Brexit. PF-CRE, on the other hand, shows that this effect is statistically significant.

Table C1: *Coefficient Estimates for Brexit Referendum*

	PF-CRE		CMLE		Trad. CRE		Logit	
	β	CI	β	CI	β	CI	β	CI
Against Integration	0.66	(0.58, 0.74)	0.64	(0.55, 0.74)	0.68	(0.60, 0.77)	0.81	(0.75, 0.87)
Immigration, Cultural	−0.05	(−0.17, 0.06)	−0.07	(−0.21, 0.07)	−0.11	(−0.25, 0.02)	−0.13	(−0.20, −0.07)
Immigration, Economic	−0.30	(−0.44, −0.16)	−0.22	(−0.38, −0.07)	−0.25	(−0.39, −0.10)	−0.34	(−0.42, −0.26)
Risk Poverty	−0.11	(−0.26, 0.05)	−0.07	(−0.25, 0.10)	−0.20	(−0.36, −0.03)	0.01	(−0.07, 0.10)
Risk Unemployment	0.15	(0.00, 0.30)	0.12	(−0.06, 0.30)	0.06	(−0.10, 0.23)	0.11	(0.02, 0.19)
Controls	No		No		No		Yes	
N^o γ terms	1258		-		5		-	
Selected γs	19		-		5		-	
Observations	9466		9466		9466		9466	
Effective Obs	9466		1713		9466		9466	
$\chi^2_{(8)}$	10.04		-		-160.52		27.50	
p-value	0.07		-		NA		0.00	

All confidence intervals are at the 95%. Logit standard errors are clustered at the individual level. The effective number of observations refers to the number of actual observations used in estimation for CMLE. There is no χ^2 test reported to CMLE since this estimator is the basis for that test.

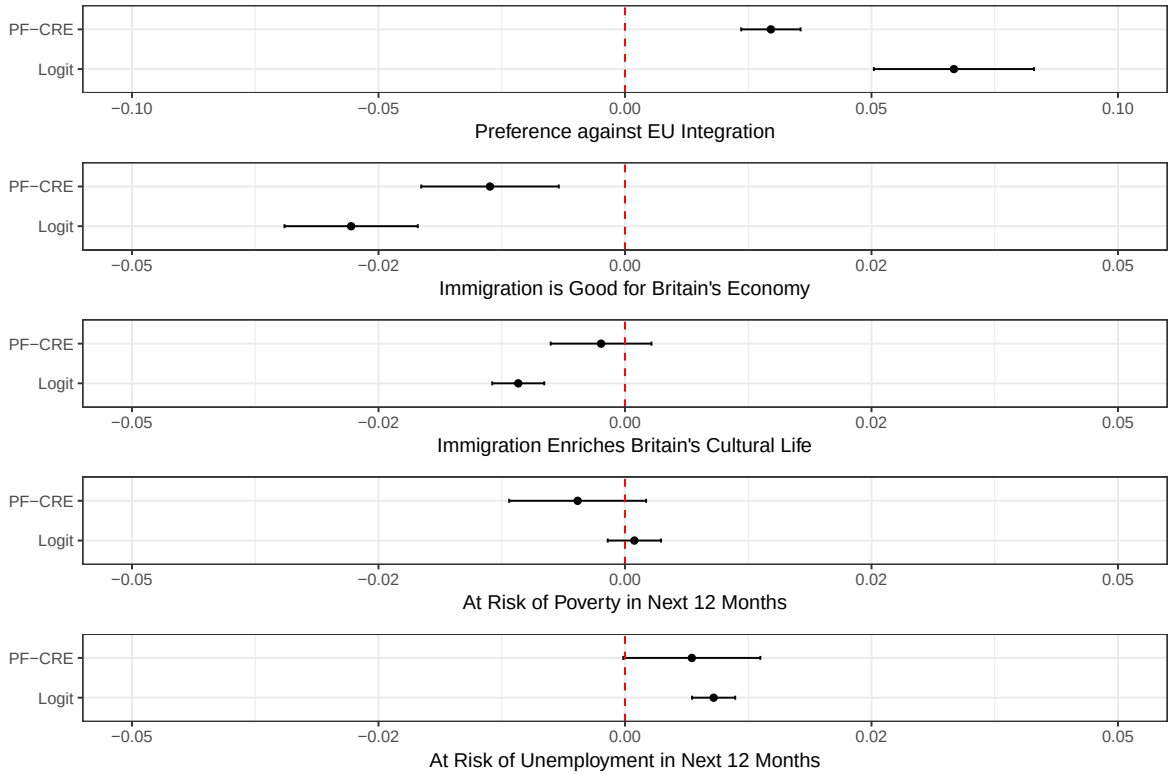
Figure C1 presents the partial effects estimated for a baseline individual.[57] The

[56]The χ^2 test in this case is negative. This means that the variance of CRE is larger than the variance of CMLE, which is evidence that CRE is not a consistent and efficient estimator of the model parameters.

[57]The baseline individual is a 45 years old white male, who is employed full time and has some college

PF-CRE estimates show that an increase in preferences against European integration are associated with a 2.95% increase in the probability of voting in favor of Brexit; logit overestimates this effect by 3.73 percentage points. In terms of respondents' views on immigration, the results show that those who find that immigration is good for the economy are 1.36% less likely to support Brexit, whereas there is not effect on the cultural side. Logit overestimates both these effects, by 1.42 and 0.84 percentage points, respectively. Finally, both estimators show that respondents' who consider themselves at risk of unemployment in the near future are more likely to support Brexit, by about 0.7 percentage points.

Figure C1: *Partial Effects from Brexit Referendum*



PF-CRE estimates were obtained using the SCAD penalty. The tunning parameters was obtained through 5-fold cross validation using the Akaike information criterion. Logit standard errors are clustered by respondent. Baseline values for the conditional mean equations in partial effects were chosen to be consistent with those of the observed characteristics of the baseline individual.

education. All other variables were set to the average value for an individual with those characteristics, as observed in the sample.

Overall, the results show that accounting for unobserved heterogeneity in the Brexit context has some important implications for our understanding voting behavior. Beyond the overestimation of various effects, there is no evidence that cultural fears actually drive support for Brexit. On the other hand, even after controlling for unobserved heterogeneity, the evidence shows that those voters with fears of losing their jobs in the near future are more likely to support the U.K.'s exit from the European Union. These results suggest that materialist concerns were the prime drivers of the referendum results, and that posmaterialist values related to Britain's culture did not play a significant role.

Ideology and Candidate Evaluations in the 2012 U.S. Election

The study of how voters make choices in elections has generally focused around two main axes: (1) ideological preferences, and (2) valence issues. The first axis is typically represented by the ideological distance between voters and the candidates, usually measured as part of standard political surveys.[58] The second axis is measured in surveys through questions, or batteries of questions, aimed at determining voters' opinion on different personal characteristics of the candidates, beyond their political positions: whether they think the candidates are moral, experienced, care about regular people, among others.

Unobserved heterogeneity is usually present in observational studies of vote choice. Important variables are not measured, hard to measure, or simply not available in the data at hand. For example, core values, which are hard to accurately capture in surveys, can be important motivators behind vote choices. The challenge they pose is that they are generally correlated with voters' ideological and personality evaluations about the candidates (see Alvarez and Brehm, 2002; Feldman, 1988). Therefore, ignoring them leads to biased inferences about these variables. Core values are generally thought of as

[58]Other focus on particular issue positions, sometimes in combination with overall ideological positions.

fixed, at least in the short and near term (Feldman, 1988; McCann, 1997).[59] Therefore, treating them as unobserved heterogeneity during the course of an election campaign is an appropriate course of action.

Beyond omitted variable bias, there are other challenges that accounting for unobserved heterogeneity can help ameliorate, like endogeneity or simultaneity. For example, positive evaluations of a candidate are usually associated with a higher probability of casting a vote for that candidate. However, a voter who has decided to cast a vote for a given candidate may then begin viewing that candidate's personality under a kinder light (even if just to diminish cognitive dissonance). Unobserved heterogeneity can alleviate this problem by accounting for individuals' general tendency to have positive (or negative) views about a candidate; the remaining variation in the data is more likely to reflect how changes in individuals' views about the candidates affect vote choices, than the other way around.

Data and Model Specification

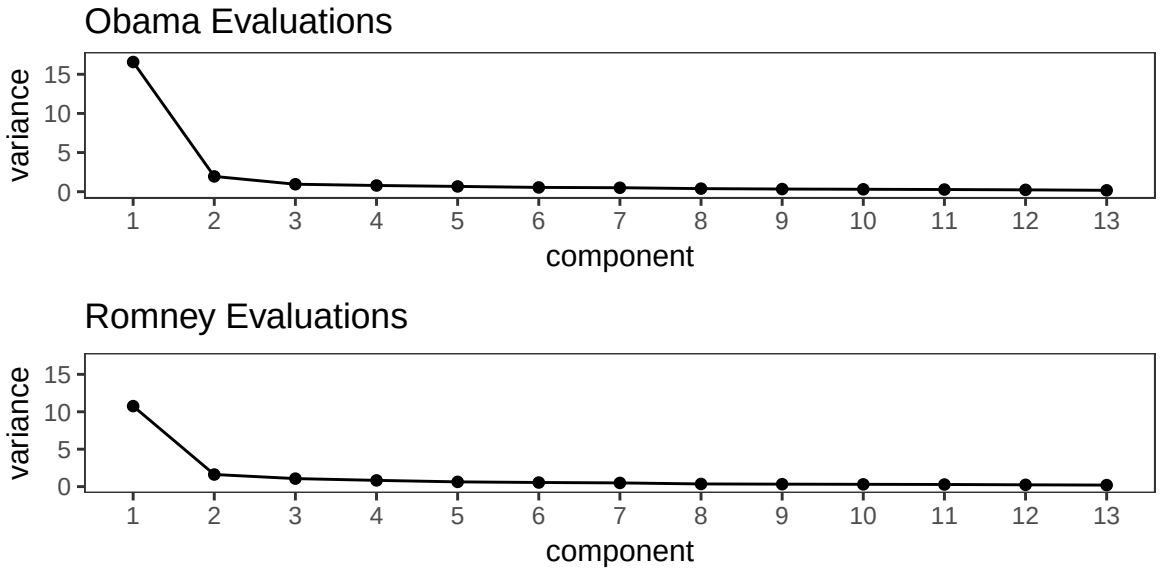
To study the effect of ideological distance and candidate evaluations on vote choice, I use data from three waves (February, June, and October) of The American Panel Study (TAPS) from the 2012 U.S. Presidential Election. The outcome of interest is whether a respondent intends to vote for Barack Obama during the General Election (Romney voters, non-voters, and third party voters are grouped together).

The variables of interest are the ideological distance of each respondent to Barack Obama and Mitt Romney, and individuals' perceptions about the candidates' personalities. I construct ideological distance as the absolute distance between the respondents' self-reported ideological position and their perceptions about the candidates' positions. Given the well-known problems of differential item functioning, self placements were adjusted using Aldrich-McKelvey rescaling (Aldrich and McKelvey, 1977), as implemented in the `basicspace` package in `R` (Poole et al., 2013).

[59] Goren (2005) challenges that core values are largely fixed, and posits that they are influenced by partisan identification.

Voters' perceptions of the candidates' personality are based on a battery of 10 questions.[60] These evaluations are very highly correlated with each other, and using them all together in a model introduces more noise than explanatory power (Ansolabehere et al., 2008, make a similar measurement error argument in the case of issue positions). For this reason, I simplify personality evaluations by replacing them with their first three principal components for each candidate, as additional dimensions do not contribute significantly to explaining the variance in candidate evaluations (see Figure C2).

Figure C2: *Principal Components Variances 2012 Presidential Election*



Principal components were calculated separately for each candidate. The evaluations for each candidate consist of 10 items, each ranging from 1 (disagree) to 7 (agree).

To model the conditional distribution of the unobserved heterogeneity in PF-CRE, I use the time-means of the covariates of interest, plus time-invariant characteristics, with up to three-way interactions, for a total of 210 terms. The time-invariant characteristics I include are race, income, year of birth, education, gender, and party identification from the first wave of the panel.[61]

[60]Respondents are asked to rate the following statements for each candidate: He is optimistic, He is partisan, He is fair, He is a strong leader, He is honest, He is trustworthy, He is experience, He is knowledgeable, He is inspiring, He is decisive, He cares about people like me, He is moral, He has a bad temper.

[61]Party identification in the TAPS data shows some variation across panel waves for some individuals. However, I choose to use the responses from the first wave, as subsequent variation is possibly a reflection of measurement error rather than actual changes in party identification.

I compare the estimates from PF-CRE to those of Conditional Maximum Likelihood (CMLE), a pooled logit estimator that includes the time-invariant characteristics as control variables, and a traditional CRE approach that only uses the time-means of the covariates to model the conditional distribution of the unobserved heterogeneity.

Results

Table C2 shows the coefficient estimates for the four main variables of interest in the model: ideological distance and the first three components of the candidate personality evaluations for Obama and Romney. The point estimates for PF-CRE and CMLE are similar to each other, with PF-CRE estimates generally having a slightly smaller variance. In fact, the specification test does not reject the null hypothesis that PF-CRE is consistent and more efficient than CMLE, with a p-value of 0.97. This test also rejects that the traditional CRE estimates are comparable to those of CMLE (p-value of 0.00). Finally, the pooled logit model, that does not account for unobserved heterogeneity, significantly overestimates the effect of ideological distance to Obama by a factor of two. Logit also significantly overestimates the effect of the Obama personality evaluations on the probability of voting for Obama. These differences highlight the importance of controlling for unobserved heterogeneity in the estimation of vote choice.

Figure C3 shows the partial effects estimated from PF-CRE and the pooled logit estimators. The baseline individual for the these partial effects is a 40 years old white woman, with median income, some college education, and with ideological distances and personality evaluations at the average for an individual with these demographic characteristics.

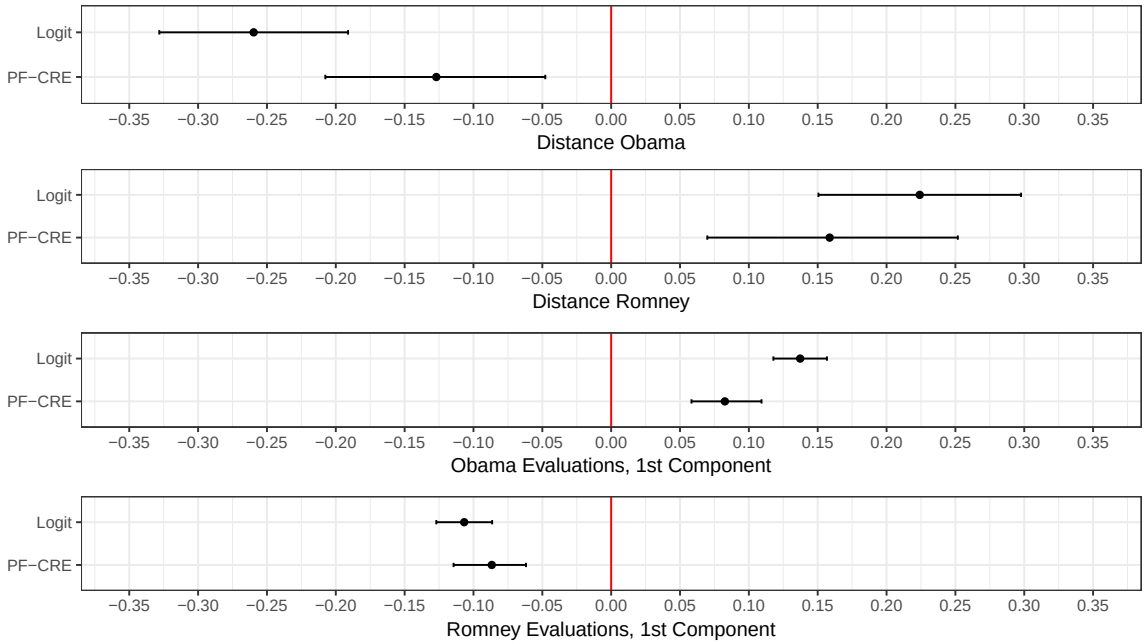
The partial effects from PF-CRE show that, for the baseline individual, increasing the ideological distance to Obama is associated with a 12.5% increase in the probability of voting for him. An increase in the ideological distance to Romney, increases the probability of voting for Obama by about 15%. Pooled logit, that ignores the unobserved heterogeneity, overestimates these effects by about 75 and 50 percent, respectively. A

Table C2: *Coefficient Estimates for 2012 U.S. Presidential Election*

	PF-CRE		CMLE		Trad. CRE		Logit	
	β	CI	β	CI	β	CI	β	CI
Distance BO	−0.73	(−1.18, −0.28)	−0.59	(−1.06, −0.12)	−0.76	(−1.22, −0.32)	−1.11	(−1.44, −0.79)
Distance MR	0.92	(0.41, 1.44)	0.82	(0.26, 1.37)	1.02	(0.49, 1.55)	0.96	(0.61, 1.31)
BO Eval, 1st	0.48	(0.34, 0.62)	0.33	(0.19, 0.47)	0.47	(0.33, 0.60)	0.59	(0.51, 0.67)
MR Eval, 1st	−0.50	(−0.65, −0.36)	−0.41	(−0.57, −0.25)	−0.48	(−0.63, −0.34)	−0.46	(−0.55, −0.36)
BO Eval, 2nd	−0.25	(−0.45, −0.04)	−0.23	(−0.51, 0.05)	−0.27	(−0.53, −0.07)	−0.30	(−0.51, −0.08)
MR Eval, 2nd	0.19	(−0.01, 0.39)	0.04	(−0.22, 0.31)	0.12	(−0.13, 0.37)	0.14	(−0.03, 0.32)
BO Eval, 3rd	0.10	(−0.13, 0.33)	0.18	(−0.16, 0.53)	0.19	(−0.12, 0.50)	0.07	(−0.12, 0.27)
MR Eval, 3rd	0.06	(−0.17, 0.28)	0.10	(−0.19, 0.40)	0.14	(−0.15, 0.43)	0.00	(−0.20, 0.20)
Controls	No		No		No		Yes	
N^o γ terms	200		-		8		-	
Selected γs	22		-		8		-	
Observations	3825		3825		3825		3825	
Effective Obs	3825		2175		3825		3828	
$\chi^2_{(8)}$	2.21		-		960.54		-143.69	
p-value	0.97		-		0.00		NA	

All confidence intervals are at the 95%. Logit standard errors are clustered at the individual level. The effective number of observations refers to the number of actual observations used in estimation for CMLE. There is no χ^2 test reported to CMLE since this estimator is the basis for that test.

Figure C3: *Partial Effects from 2012 U.S. Presidential Election*



PF-CRE estimates were obtained using the SCAD penalty. The tunning parameter λ was obtained through 5-fold cross validation using the Akaike information criterion. Logit standard errors are clustered by respondent. Baseline values for the conditional distribution of the unobserved heterogeneity in the partial effects were chosen to be consistent with those of the observed characteristics of the baseline individual.

similar picture arises from personality evaluations. PF-CRE estimates that a more positive evaluation of Obama is associated with a 7.5% higher probability of voting for him, whereas better personality evaluations of Romney are associated with a decrease of about 9% in the probability of voting for Obama. Pooled logit also overestimates these effects, by 75 and 25 percent, respectively.

Disussion

Overall, the partial effects from PF-CRE show that voters' perceptions on personality characteristics and ideological distance for both candidates have effects of similar size. While ideological distance to the candidates is an important predictor of vote choice, after controlling for unobserved heterogeneity, its partial effect is of comparable size to that of personality evaluations. Furthermore, the partial effects for ideological distance have a large degree of uncertainty relative to those of personality evaluations. These results suggest that ideological considerations are not the dominant axis along which votes intentions move, at least within the time-frame of an election year. Instead, candidate personality evaluations are of similar importance, and have a stronger statistical association with vote choice.

The difference between the pooled logit and PF-CRE estimates of the partial effects for ideological distance and personality evaluations point to two related conclusions, one methodological and the other substantive. On the methodological size, this difference is illustrative of the perils of ignoring unobserved heterogeneity. As the pooled logit shows, this leads to partial effects that can be twice as large as those of a model that controls for the unobserved heterogeneity. On the substantive size, the smaller partial effects of the PF-CRE model, and specifically those for ideological distance, are possibly an indication of the effects of political polarization, as they point to choices that are weakly responsve to changes in voters' perceptions during the campaign than would otherwise be expected.