Political Science Research and Methods (2018), 1–14
doi:10.1017/psrm.2018.40

ARTICLE

Fixed effects in rare events data: a penalized maximum likelihood solution

Scott J. Cook[1*], Jude C. Hays[2] and Robert J. Franzese[3]

[1]Department of Political Science, Texas A&M University, 2010 Allen Building, College Station, TX 77843, USA,
[2]Department of Political Science, University of Pittsburgh, 4600 Wesley W. Posvar Hall Pittsburgh, PA 152603, USA and
[3]Department of Political Science, University of Michigan, 5700 Haven Hall, Ann Arbor, MI 48109, USA
*Corresponding author. Email: sjcook@tamu.edu

Abstract

Most agree that models of binary time-series-cross-sectional data in political science often possess unobserved unit-level heterogeneity. Despite this, there is no clear consensus on how best to account for these potential unit effects, with many of the issues confronted seemingly misunderstood. For example, one oft-discussed concern with rare events data is the elimination of no-event units from the sample when estimating fixed effects models. Many argue that this is a reason to eschew fixed effects in favor of pooled or random effects models. We revisit this issue and clarify that the main concern with fixed effects models of rare events data is not inaccurate or inefficient coefficient estimation, but instead biased marginal effects. In short, only evaluating event-experiencing units gives an inaccurate estimate of the baseline risk, yielding inaccurate (often inflated) estimates of predictor effects. As a solution, we propose a penalized maximum likelihood fixed effects (PML-FE) estimator, which retains the complete sample by providing finite estimates of the fixed effects for each unit. We explore the small sample performance of PML-FE versus common alternatives via Monte Carlo simulations, evaluating the accuracy of both parameter and effects estimates. Finally, we illustrate our method with a model of civil war onset.

To FE or not to FE?

Since the Dirty Pool symposium—a special issue on panel data estimation in *International Organization*—addressed the topic of fixed effects models more than a decade ago, little progress has been made on the problem of unobserved unit heterogeneity in rare-event binary time-series cross-sectional (re-BTSCS) data. While at the time, this debate helped to clarify many of the issues raised by various strategies for estimating models with these data (notably the summary from King 2001), the lasting legacy has been one of confusion rather than clarity. The inability of these scholars to reach a consensus leaves applied researchers unclear as to which approach they should adopt. While strategies for panel and TSCS estimation have the received renewed attention of late (Bell and Jones 2015; Clark and Linzer 2015), little of this work focuses on the unique problems posed by rare event binary TSCS data. We aim to advance this discussion, revisiting the debate on unit heterogeneity in re-BTSCS data and proposing a novel fixed effects estimation strategy for such data.

When analyzing re-BTSCS data researchers often select between pooling, random effects, and fixed effects models. Most opt for either pooling, which assumes no unobserved unit heterogeneity, or random effects, which assumes that any unit heterogeneity is orthogonal to the explanatory variables. Researchers often prefer these models to fixed effects estimation even when explicitly noting that these assumptions are violated (Nel and Righarts 2008; Wright 2009). While the orthogonality of predictors and unit effects is likely specious in most macro-level studies of comparative politics and international relations, these modeling strategies dominate because—drawing upon Beck and Katz (2001)—many adopt the view that it is never a good idea to estimate fixed effects models with re-BTSCS data.

In general, three reasons are given by researchers for avoiding fixed effects models. First, fixed effects estimation is assumed to be problematic because the unconditional estimator (i.e., unit dummy variables) suffers from the incidental parameters problem—generating biased estimates when T is small.[1] Second, researchers are unable to recover effect estimates of time-invariant predictors.[2] Finally, with rare events data, many units do not experience the event in small samples.[3] As a consequence, there is no overlap between the fixed effects and the outcome (i.e., separation), and these units do not enter the log-likelihood. That is, parameter estimates are produced using only the data from the event-experiencing set of units.

It is this final issue which seemed to most divide the participants of the Dirty Pool symposium and is the focus of our article, namely, what are the consequences of this sample selection and how should researchers proceed in light of it? We clarify that the primary consequence of this sample selection is not biased or inefficient estimates of predictor coefficients, but instead an inaccurate estimate of the baseline event risk (i.e., the unit effects).[4] Analyzing event-experiencing units alone produces an inflated average estimate of the event risk, as no-event units have a lower event probability than event-experiencing units on average. This, in turn, biases the marginal effect estimates of the predictors which are, in part, a function of these probabilities. This is the fundamental problem for most researchers with rare events BTSCS data: random effects is a biased estimator given non-zero correlation between predictors and unit effects, yet fixed effects estimators induce sample selection and resultant biases in marginal effects.

As a solution, we present an alternative estimation strategy for fixed effects which does not have this limitation. Our proposed penalized maximum likelihood fixed effects (PML-FE) estimator includes unit dummies, as in unconditional fixed effects, but retains the units that do not experience a rare event in the sample using a modified score function. In short, we recognize that this issue is a special case of separation and use a familiar strategy (i.e., Firth's logit) to address it. This is predicated on the theoretical belief that all units in the sample would, in time (i.e., with enough measures), experience the outcome. Modifying the score function to reflect this penalizes the fixed effect on no-event units away from negative infinity, retaining them in the log-likelihood. Our simulations show that this allows PML-FE to outperform pooled, random effects, and unconditional fixed effects in recovering marginal effect estimates when predictors and unit effects are correlated.

Rare events and fixed effects

To fix terms, consider the familiar latent-variable representation of BTSCS as

$$
\begin{aligned}
y^*_{it} &= \alpha_i + \mathbf{x_{it}}\boldsymbol{\beta} + \epsilon_{it}, \\
y^*_{it} &= 1(y^*_{it} > 0), \qquad (1)
\end{aligned}
$$

where y_{it} is an observed binary outcome for unit $i = 1,\ldots,N$ at time $t = 1,\ldots,T$, $\mathbf{x_{it}}$ is a vector of observed explanatory variables, α_i is an unobserved unit effect, and ε_{it} is an i.i.d. error. Assume, as most applied researchers have, that ε_{it} is distributed logistic, producing

[1]We briefly note that the incidental parameters problem will not typically be an issue with common TSCS sampling dimensions (see Beck 2011; Greene 2004). However, researchers should still be aware of the proliferating parameters problem that they may face with unconditional fixed effects estimation (Beck 2015).

[2]With fixed effect binary outcome models, methods for recovering the effects of time-invariant explanatory variables remain underdeveloped. There is not, for example, an analog to the fixed-effects variance decomposition estimator of Plümper and Troeger (2007, 2011) for limited dependent variable models.

[3]For our discussion, we mainly focus on instances where unit values in the dependent variable are all zeros, however, an equivalent and parallel problem arises where unit values are all ones (ex. democracy as an outcome).

[4]To clarify, by baseline event risk we mean the probability of the event when all covariates are held at zero.

$$\Pr(y_{it} = 1|\alpha_i, \mathbf{x_{it}}) = \frac{1}{1 + e^{-(\alpha_i + \boldsymbol{\beta}\mathbf{x_{it}})}}, \tag{2}$$

with different assumptions about α_i giving rise to the familiar pooled, random effects, and fixed effects models. In short, when a researcher assumes that α_i is α for all units i—that is, there is no heterogeneity in the unit intercepts—then the pooled model is supported and (a simplified version of) Equation 2 is estimated as in standard logistic regression.

However, when α_i is believed to vary across units, one of the panel models—random or fixed effects—should offer a better fit to the data. The main criterion usually considered when selecting between random and fixed effects is whether α_i is orthogonal to $\mathbf{x_{it}}$.[5] If yes, then the random effects model is often preferred.[6] First, it is more efficient than the fixed effects model, thereby producing smaller standard errors and superior estimates (assuming unbiasedness). Second, with random effects estimation the within-unit variance of the predictors is not partialed out, as with fixed effects, allowing researchers to obtain estimates on time-invariant predictors.

Where one assumes instead that α_i and $\mathbf{x_{it}}$ are correlated, then a fixed effects model is often more appropriate, as both pooled and random effects estimators will be biased. For logistic models, there are two fixed effects estimators: unconditional and conditional fixed effects. The former, unconditional fixed effects, includes dummy variables for each unit (except one) into the specification and maximizes the standard logit log-likelihood. The latter, conditional fixed effects, due to Chamberlain (1980) concentrates out the fixed effects (via $\Sigma_t y_{it}$) and maximizes a conditional log-likelihood function:

$$L^C = \Sigma_i ln\left(\frac{\exp(\beta'\Sigma_t x_{it} y_{it})}{\Sigma_{d_i} \exp(\beta'\Sigma_t x_{it} d_t)}\right)$$

where $d_i = \{d = (d_1,\ldots,d_T)|d_t = 0$ or 1 and $\Sigma_t d_t = \Sigma_t y_{it}\}$, that is, the individual combination of observed y's for each unit i.

While conditional fixed effects is often preferred due to concerns about the incidental parameters problem, there are several limitations.[7] First, it does not provide estimates of the individual unit effects so we are unable to estimate substantive quantities of interest such as marginal effects or first differences. Second, it is not currently possible to estimate the coefficients of time-invariant explanatory variables, as they are indistinguishable from the invariant unit-effect. Third, those units which do not realize an event in the sample are dropped from the analysis.

It is the final issue that is the main focus of our discussion and seemed to most divide the participants of the Dirty Pool symposium—arguably between those who prefer a more cautious approach to causal inference from observational data (Green et al. 2001), even when a significant portion of the data exhibits no within-unit variation, and those who prefer a more theoretically driven interpretation of all of the empirical evidence (Beck and Katz 2001; Oneal and Russett 2001). The latter group emphasizes that dummy variables are atheoretical and remove all the between-unit

[5]Other principled motivations are occasionally given for preferring one to the other—how the sample is drawn (and whether it is exhaustive), the distribution of the unit effects, etc.—however, here we focus on the correlation between the unit effects and the predictors.

[6]To clarify, the random effects model is estimated by integrating out the random intercept and incorporating the scale parameter into the likelihood for the the ith subject, is given by the following (marginalized) likelihood

$$L_i = \int \prod_{t=1}^{T} P_\sigma(Y_{it} = y_{it}|\beta, z_i)\phi(z_i)dz_i,$$

where $\alpha_i = \sigma z_i$, σ being the scale parameter, $z_i \sim N(0,1)$ and $\varphi(\cdot)$ is the standard normal density. The joint marginalized likelihood is

$$L = \prod_{i=1}^{n} (Y_i|\beta, \sigma),$$

which is maximized using a form of quadrature or MCMC. For more discussion of the logistic random-effects model, see Lesaffre and Spiessens (2001).

[7]Lancaster (2000) provides a useful summary of (and the history behind) the incidental parameters problem.

variation in the data. Their reaction to Green et al. (2001) was particularly strong, claiming that fixed effects models are rarely, if ever, justified in the case of binary dependent variables.

In part because of the veracity of the claims made by Beck and Katz (2001), researchers frequently eschew fixed effects. For example, Nel and Righarts (2008) write:

> … we do not run fixed effects models. Following Beck and Katz (2001), we consider the use of fixed effects models to control for the influences of unit idiosyncrasies in binary outcome time-series cross-sectional data as pernicious. There are many units with no other outcome than zero, and to control for their presumed effects on the parameter estimates does not make any sense.

Similarly, in research on democratization, Wright (2009) argues that:

> …[t]o address concerns of omitted variable bias, it would be ideal to include country fixed effects. However, there are many countries in the sample that do not experience a transition to democracy. Including fixed effects would entail dropping these countries from the sample because there is no variation in the dependent variable for those countries. Dropping these observations would induce severe sample selection bias by examining only countries with observed transitions to democracy. As a next-best approach, we can include random effects (RE) in the model.

It is repeated claims such as these which give us cause for concern. Erroneously assuming that the unit effects are absent or orthogonal to the included regressors in estimation, especially when one suspects these assumptions are problematic, risks bias. Estimating a random effects model is not a stand-in for a fixed effects, it is an alternative estimation strategy dependent on a fundamentally different assumption.

Moreover, researchers often seem unclear as to why sample censoring is a concern at all. While the loss of a large proportion of one's sample is displeasing, the particular consequences of this are rarely discussed. First, as noted by Heckman (1981) this is a small sample bias given that "as $T \to \infty$ this problem becomes unimportant." Asymptotics aside, naïve probability suggests that the extent of sample censoring should lessen with increases in time, as more units experience the event. Second, even in small sample analysis, the sample selection does *not* induce bias in the coefficient estimates on predictors.[8] The coefficient is now the within-unit effect of the predictor, for which only units with variation in the outcome provide information. Instead, it is those estimators that retain the complete sample yet fail to partial out the endogenous between-unit component of the predictors—that is, pooled and random effects—that risk bias in these coefficients.

This is not to say, however, that sample selection is without consequence, as it biases summary estimates of the base rate of the event. In short, by only retaining event-experiencing units, we overestimate the unconditional probability of the outcome.[9] This can be easily seen via the law of total probability (with Z an indicator for whether the units have experienced the event):

$$\underbrace{\Pr(Y=1)}_{\substack{\text{Population}\\ \text{relationship}}} = \underbrace{\Pr(Y=1 \mid Z=1)}_{\substack{\text{Uncensored}\\ \text{relationship}}}\underbrace{\Pr(Z=1)}_{\substack{\text{Probability}\\ \text{uncensored}}} + \underbrace{\Pr(Y=1 \mid Z=0)}_{\substack{\text{Censored}\\ \text{relationship}}}(1-\Pr(Z=1)),$$

Given that, on average, $\Pr(Y=1 \mid Z=1) > \Pr(Y=1 \mid Z=0)$—since we have observed realizations of the outcomes for these units—we will overestimate the base rate of the event if we only evaluate this subset. This not only affects any summary estimates of the event risk but any

[8]Unconditional fixed effects estimation may still be biased for reasons unrelated to sample censoring. For a discussion on this issue, see Beck (2015)

[9]This is not problematic if one only draws inferences on event-experiencing units, however, researchers are often interested in the complete sample (and the underlying population represented therein).

estimands which include these as inputs, such as marginal effects. Consider, for example, the average marginal effect (AME):

$$\text{AME} = \frac{1}{n}\sum_{i=1}^{n} \Pr\left(\overline{Y=1|\ x_i, \hat{\beta}, \hat{\alpha}_i}\right) \times \left(1 - \Pr\left(\overline{Y=1|x_i, \hat{\beta}, \hat{\alpha}_i)}\right) \times \hat{\beta} \tag{3}$$

which shows that even if with an accurate $\hat{\beta}$, inaccurate estimates of baseline probability (a function of $\hat{\alpha}$) induces bias in the marginal effect estimates. More specifically, overestimating the base rate due to the sample censoring causes researchers to *overestimate* the AME of **x**.[10] In short, with rare events, overestimating the unit effect shifts the location along the sigmoidal function of the logit to the right (closer to the inflection point), such that a $\hat{\beta}$ change now has a greater effect on the probability of obtaining ones.

Since these marginal effects are often the quantity of interest for researchers with binary outcomes, biased estimates here may be as or more problematic than biased coefficient estimates. To redress this, we present a fixed effects estimator in the next section that allows researchers to retain the full sample.

Penalized maximum likelihood—fixed effects

PML is a now common strategy for reducing estimation bias within a frequentist framework. Firth (1993) initially proposed PML estimation as a means of obtaining first-order unbiased estimates through a modified score function. Since PML achieves this bias prevention during iteration—rather than correcting the bias after estimation—it also produces finite parameter estimates even in the presence of quasi- or complete separation (Heinze and Schemper 2002; Zorn 2005). Separation occurs when a subvector $x_s \in x$ deterministically locates each observation in y, that is, a units' outcomes are perfectly predicted.[11] This *is* the problem confronted with conventional fixed effects models of rare events: no-event units have zero probability and therefore do not enter the log-likelihood. Given that this is a special instance of separation, we argue that PML can be utilized as a solution.

As noted above, Firth (1993) proposed modifying the score function as

$$U_r^*(\theta) = U_r(\theta) + A_r(\theta),$$

where $U_r(\theta)$ is the ordinary score and $A_r(\theta)$ is modification to the score derived from the data. For instance, in the exponential family of models this adjustment is given by

$$a_r = \frac{1}{2} tr\left\{ i^{-1}\left(\frac{\partial i}{\partial \theta_r}\right)\right\} = \frac{\partial}{\partial \theta_r}\left\{\frac{1}{2} log|i(\theta)|\right\}.$$

With the solution of $U_r^*(\theta) = 0$ locating a stationary point of

$$L^*(\theta) = L(\theta)\ \ |I(\theta)|^{\frac{1}{2}}.$$

That is, the ordinary likelihood L penalized by the square root of the determinant of the information matrix $|I(\theta)|^{\frac{1}{2}}$, which is equivalent to Jeffreys prior.[12] The determinant of the information matrix is maximized when $\beta = 0$, so the penalty function shrinks the estimates—including unit effect estimates that would otherwise be infinite—toward zero. This penalty supplies extra-empirical (or non-data-driven) information in estimation—that is, the assumption

[10]The same also holds when estimating marginal effects at means, given that the mean value of the constant from the censored sample larger than that from the complete sample.

[11]Though we focus on separation in a binary-outcome context here, it is a potential issue in all discrete outcome models (Cook et al. 2018).

[12]While other priors could be considered in a more traditional Bayesian setting (e.g., Gelman et al. 2008; Rainey 2016), we prefer Firth's logit as it is: 1) already familiar to political scientists as solution to separation and 2) it is easily implemented in Stata and R.

that with increases to T each unit would change state.[13] That we fail to observe this in any given sample, then, is just a small-sample artifact. To us, this assumption is often more valid than those implicitly made by many researchers currently—homogeneous unit effects or orthogonality between unit effects and predictors.

Therefore, in our PML-FE model, we include separate intercepts for each of the event-experiencing units, and a common intercept for those that do not

$$\alpha_i = \begin{cases} \alpha_i, & \text{if } \sum_{t=1}^{t} y_{it} \neq 0 \\ \alpha, & \text{otherwise.} \end{cases}$$

In conventional fixed effects estimation α—the fixed effect for the censored units—would tend to infinity.[14] With our penalized estimation, we are able to maintain the full sample and recover more accurate estimates of the baseline event risk.[15] As such, we believe PML-FE should offer superior performance to traditional fixed effects estimators (for the reasons given in Section 2).

Simulations

We explore the small-sample properties of the pooled, random effects, unconditional fixed effects, conditional fixed effects, and PML-FE logit estimators via Monte Carlo simulation. For our experiments, we employ a data-generation process similar to Beck (2015). The basic framework is given in Equations (1) and (2) above, with i and t indexing the unit and time, respectively, and the inputs drawn as follows:

$$x_{it} \sim \mathcal{N}(\overline{x}_i, \sigma)$$
$$\begin{pmatrix} \alpha_i \\ \overline{x}_i \end{pmatrix} \sim N\left[\begin{pmatrix} -4 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right],$$

where σ determines the within-unit variance in the predictor x_{it}, and ρ determines the correlation between the mean of each unit's predictor ($\overline{x}_i$) and the unit effect (α_i) which partly determines the overall level of endogeneity in x_{it}. The coefficient β is set to 1. The variances of $\overline{x}_i$ and α_i are also set to 1, which means that their covariance is also their correlation.

By varying ρ, we can manipulate the degree to which the orthogonality assumption of the random-effects estimator is violated. Reducing ρ should improve the performance of the random-effects estimator. By varying σ, we can manipulate the amount of within-unit variation available for estimation. This should benefit the fixed effects estimators since they rely on within-unit variation only. In fact, the conventional view is that fixed effects estimators should not be used when there is little within-unit variance (small σ). However, low within-unit variance also compounds the endogeneity problem, magnifying the strength of the relationship between x_{it} and α_i, and this is problematic for the random effects estimator. Therefore, the effect of σ on the relative performance of the estimators is ambiguous.

[13]There are other reasons why units may never change state: first, units may never fail because of an additional factor which ensures the outcome with certainty (e.g., infertile couples in studies of assisted reproductive technology); second, units may never fail because of misclassification or contamination in the data (e.g., non-reporting on events in a country so the outcome is always recorded as zeros even when true ones occur). In either of these cases increases to T, no matter how large, would never result in a change of state. See Copas (1988) and Cook et al. (2017) for discussion on these issues and possible solutions.

[14]One could also allow for each of the censored units to have their own intercepts, however, we prefer a common intercept for two reasons. First, little information is available to estimate separate intercepts on censored units—in particular when a panel is reasonably balanced. Second, reducing the number of separated parameters increases the computational efficiency of PML.

[15]Cook and McGrath (nd) consider a further refinement to this, allowing for random variation in the unit-level effects around the common fixed effect for uncensored units.

Table 1. RMSE Ratios for PML-FE Coefficient Estimates ($\rho = 0.5$, $\sigma = 1$)

	Pooled	Random Effects	Unconditional Fixed	Conditional Fixed
	$N = 50$	$N = 50$	$N = 50$	$N = 50$
	$N = 100$	$N = 100$	$N = 100$	$N = 100$
$T = 20$	0.86	0.73	0.83	0.99
	0.61	0.58	0.75	1.00
$T = 50$	0.54	0.65	0.92	1.00
	0.31	0.51	0.88	1.00

Note: RMSE ratios are relative to PML-FE results, such that values less than 1 indicate superior performance by PML-FE. RMSE = root-mean squared error; PML-FE = penalized maximum likelihood fixed-effects.

We report the results for four sets of experiments: 1) high endogeneity, low within variance (high-ρ, low-σ); 2) low endogeneity, low within variance (low-ρ, low-σ); 3) high endogeneity, high within variance (high-ρ, high-σ); and 4) low endogeneity, high within variance (low-ρ, high-σ). In the first set of experiments (Table 1), we set the within-unit variance σ to 1 (meaning the between-to-within, hereafter B/W, variation ratio is also 1) and the correlation between the predictor and the unit effect ρ to 0.50, varying the size of N (the number of units) and T (the number of time periods).[16] The high level of endogeneity should favor the fixed effects estimators in this case.

The results in Table 1 show that the fixed effects models dominate the alternatives in root-mean squared error (RMSE) terms. More specifically, the numbers in the table are RMSE ratios that express the relative performance of PML-FE to the alternatives (listed in the table's columns). Ratios less than 1 indicate the superiority of PML-FE. We see that the unconditional fixed effects estimator does consistently worse than the conditional fixed effects (and the PML-FE) estimator. Beck (2015) argues that the poor performance is attributable to the large number of parameters that are estimated in the unconditional fixed effects model. However, the PML-FE estimator, which the reader will recall is a modified estimation strategy for the unconditional fixed effects specification, does as well (and slightly better) than the conditional fixed effects model in recovering the coefficient estimates. The penalty-induced shrinkage in the PML-FE estimates seems to ameliorate the problem of proliferating parameters.[17]

In all but the linear-additive model, coefficient estimates are not effect estimates. Most of the time we are interested in the (average) marginal effect of a predictor on the probability of observing the outcome, yet the conditional fixed effects estimator is unable to produce this estimate. This is one reason to prefer the random effects or unconditional fixed effects estimators to conditional fixed effects, despite the fact that the latter approach estimates the coefficient β relatively accurately. Our results (see Table 2) suggest that this might be a poor justification, as these estimators fail to accurately recover these effects.[18]

Beginning with unconditional fixed effects, the performance of the estimator relative to PML-FE is abysmal and depends on the number of censored units in the sample, which, in turn, depends on T. For $T = 20$, the percentage of censored units is above 50 percent and the RMSE for PML-FE is about one-tenth the size as for the unconditional fixed effects estimator.[19] Since the censored units have systematically lower baseline probabilities of experiencing the event, the unconditional FE estimator produces a badly biased estimate of the AME. Additionally, the

[16]This value of ρ implies that the unobserved unit effect accounts for 25% of the between-unit variance in the predictor.

[17]In the remainder of the text, we focus on average marginal effect estimates, but this result holds across all of our experiments: PML-FE performs as well as or slightly better that conditional fixed effects when it comes to estimating the coefficient β.

[18]We calculate the average marginal effects as given in Equation (3).

[19]The original RSME results for each of the estimators are presented in Appendix I.

Table 2. RMSE Ratios for PML-FE Marginal Effects ($\rho = 0.5$, $\sigma = 1$)

	Pooled	Random Effects	Unconditional Fixed	Censoring
	$N = 50$	$N = 50$	$N = 50$	
	$N = 100$	$N = 100$	$N = 100$	(%)
$T = 20$	0.52	0.81	0.14	48
	0.40	0.73	0.09	50
$T = 50$	0.32	0.75	0.16	32
	0.25	0.71	0.14	33

Note: RMSE ratios are relative to PML-FE results, such that values less than 1 indicate superior performance by PML-FE. Censoring gives the percentage of units for which Y is never 1.

Table 3. RMSE Ratios for PML-FE Marginal Effects ($\rho = 0.25$, $\sigma = 1$)

	Pooled	Random Effects	Unconditional Fixed	Censoring
	$N = 50$	$N = 50$	$N = 50$	
	$N = 100$	$N = 100$	$N = 100$	(%)
$T = 20$	1.02	1.16	0.15	48
	0.71	1.17	0.10	50
$T = 50$	0.64	1.07	0.18	31
	0.42	1.16	0.16	32

Note: RMSE ratios are relative to PML-FE results, such that values less than 1 indicate superior performance by PML-FE. Censoring gives the percentage of units for which Y is never 1.

PML-FE AME estimates are superior to those produced by random effects. The endogeneity problem is compounded by the low within-unit variance in x_{it} to such an extent that that random effects is outperformed by PML-FE. Moreover, the size of this performance gap seems to be sensitive to N, with large N benefiting PML-FE relative to random effects. In sum, under these experimental conditions, the PML-FE estimator is able to recover coefficient estimates as accurate as conditional fixed effects and also produce the most accurate marginal effects estimates.

Decreasing the correlation between the predictor and the unit effect should reduce the MSE gains achieved from fixed effects, as random effects becomes less biased and more efficient. Therefore, in our second set of simulations, we reduce ρ to 0.25 holding all else constant. The results in Table 3 indicate that, as expected, random effects performs better and is now the preferred model. The RMSE of the sampling distribution for PML-FE estimator is between 27 percent and 47 percent larger than the RMSE for the random effects estimator, and the relative performance of random effects improves as N increases. Thus, when the overall level of endogeneity is high, increasing the number of units reduces the performance of the random effects estimator relative to PML-FE, but when the overall level of endogeneity is lowered, consistent with Beck's proliferating parameters problem, the relative performance of random effects improves.

A second issue is the extent to which predictors change over time—that is, covariates for which the within variation are small relative to the between variation—as coefficients on slowly changing variables can be poorly estimated with fixed effects models (Plümper and Troeger 2007). What happens when we increase the within unit variance for both the high and low endogeneity cases? For the experiment reported in Table 4, we increase the within-unit variance in the predictor for the high endogeneity case. This should benefit PML-FE by increasing the amount of information used to estimate the effects, but it also benefits the random effects estimator by reducing the overall degree of endogeneity in x_{it}. Comparing to Table 2, we see that the relative performance of random effects to PML-FE improves when we increase σ from 1 to 2.

Table 4. RMSE Ratios for PML-FE Marginal Effects ($\rho = 0.5$, $\sigma = 2$)

	Pooled	Random Effects	Unconditional Fixed	Censoring
	$N=50$	$N=50$	$N=50$	
	$N=100$	$N=100$	$N=100$	(%)
$T=20$	0.62	0.89	0.14	30
	0.44	0.78	0.09	32
$T=50$	0.37	0.85	0.22	15
	0.31	0.89	0.18	15

Note: RMSE ratios are relative to PML-FE results, such that values less than 1 indicate superior performance by PML-FE. Censoring gives the percentage of units for which Y is never 1.

Table 5. RMSE Ratios for PML-FE Marginal Effects ($\rho = 0.25$, $\sigma = 2$)

	Pooled	Random Effects	Unconditional Fixed	Censoring
	$N=50$	$N=50$	$N=50$	
	$N=100$	$N=100$	$N=100$	(%)
$T=20$	1.07	1.06	0.15	29
	0.74	1.03	0.10	32
$T=50$	0.69	1.01	0.26	13
	0.55	1.10	0.21	13

Note: RMSE ratios are relative to PML-FE results, such that values less than 1 indicate superior performance by PML-FE. Censoring gives the percentage of units for which Y is never 1.

The implication of the findings in Tables 2 and 4 runs counter to the conventional wisdom. When the level of endogeneity is high, estimating the coefficients on predictors with little within-unit variation (e.g., slowly-changing predictors) using the fixed effects model actually does relatively better in comparison random effects.

Comparing Tables 3 and 5, when the level of endogeneity is low ($\rho = 0.25$), we see the opposite result: the relative performance of the PML-FE estimator improves. Overall, it is clear that the marginal benefits to the random effects and PML-FE estimators from increasing the within unit variance of the predictor depend on the strength of the correlation between x_i and α_i.[20] When this correlation is strong, the improvements in the random effects estimator from reducing the endogeneity of the predictor (x_{it}) are greater than the improvements in the penalized fixed effects estimator from increasing the amount of information available to estimate the coefficient β.

In sum, our experiments suggest that the use of the pooled, unconditional, and conditional fixed effects estimators is rarely, if ever, justified with rare events data. With intercept heterogeneity, random effects typically outperforms the pooled estimator. If these unit effects are also correlated with the predictors, then PML-FE should often be preferred. In our simulations, PML-FE weakly dominated both unconditional and conditional fixed effects, providing coefficient estimates at least as good as the latter and more accurate marginal effects estimates than the former. The extent of these gains is both a function of the degree of censoring, as this biases the AME estimates from the unconditional fixed effects model.

The choice between random effects and PML-FE is less straight-forward and depends, not surprisingly, on the extent to which the orthogonality assumption of random effects is violated. As this is a theoretical choice that cannot be adjudicated empirically, there will always be a degree of subjectivity here. However, one bit of observable information that may be useful is the between-to-within variation in predictors. When one suspects a predictor is correlated with the unit effects, the lack of within-unit variation compounds this problem and weighs in favor of

[20]This is shown in Tables 2 and 4 of Appendix I.

PML-FE rather than random effects. Generally, researchers should report both values and discuss what deviant results may tell us.

Illustration—the determinants of civil war

One of the research areas most prominently affected by the issues raised in the Dirty Pool symposium has been peace studies. This makes sense given the initial substantive focus of that discussion (i.e., the democratic peace) and the abundance of binary outcome models in this literature (e.g., conflict onset, alliance formation, regime transition, etc…). Despite this, a reading of this literature suggests that there remains little conceptual clarity on how best to model unobserved unit heterogeneity. Researchers frequently seem to confuse the issues underlying the decision over which modeling strategy to prefer. In particular, they seem skeptical of, and occasionally hostile toward, fixed effects estimation. Stemming from this confusion, researchers have largely failed to adopt a uniform approach for estimating unit effects.

The civil war literature, in particular, demonstrates this variation. Numerous papers make no attempt to model unit effects—preferring pooled logit or probit estimation—while those that do often adopt different approaches. For example, Fearon and Laitin (2003) re-estimate their main model(s) using conditional fixed effects logit and indicate that their results are "virtually identical" to the pooled estimates.[21] Instead, Sambanis (2001) re-estimates his models using random effects probit, ultimately preferring the simple probit estimator—despite rejecting the null of independence—because of the similarity in the results. Finally, Collier and Hoeffler (2004) argue that fixed effects estimation is "very severe"—owing to sample censoring—yet they re-estimate their models with both random and fixed effects. Thus, even the canonical works in the civil war literature—with each of these articles having been cited more than 1000 times—disagree over the role of unobservables in the determination of civil war and how they should be addressed in empirical research.

This is troubling given that this issue is not simply one of taste or a methodological nuisance. Rather, it represents a significant theoretical belief, namely, whether we believe there are latent unobserved factors which cause some states to experience (avoid) civil conflict. More specifically, are there unobservables which may determine both conflict propensity and the included regressors? There are strong reasons to suspect this may be the case for some of the standard included determinants of civil war such as the level of development. In a different context, Acemoglu et al. (2008) have challenged the effect of income on democratization, arguing that (unobserved and therefore unmodeled) historical factors influence both political and economic development. Similarly, in the case of civil war and development, some states have likely developed better conflict-management processes and resultant institutions. The ability to resolve low-level disputes without resorting to violence aided development and reduced the risk of future fighting. That is, to the extent that there is a relationship between development and fighting it is one borne out over hundreds of years. Conversely, those states which were unable to resolve such disputes peaceably were set back in their development—forced to devote resources to security, dispute resolution, fighting, etc.—and had a greater probability of future conflict.

If correct, this would mean that the relationship between the level of GDP and civil war is spurious, with both sharing a common cause. Pooled and random effects models do not offer insurance against this, risking bias due to these unobservables. Instead, researchers should prefer fixed effects estimation which partials out these potential unobserved (time-invariant) characteristics from the analysis. To examine these issues, we reanalyze Fearon and Laitin's (2003) classic and oft-cited insurgency model which examines the determinants of civil war onset globally from 1945 to 1999. The dependent variable is binary, assuming a value of 1 for the first year of a civil war which is defined as conflict within a state in which both combatants (one

[21]However, Fearon and Laitin (2003) does not report these findings in the main text.

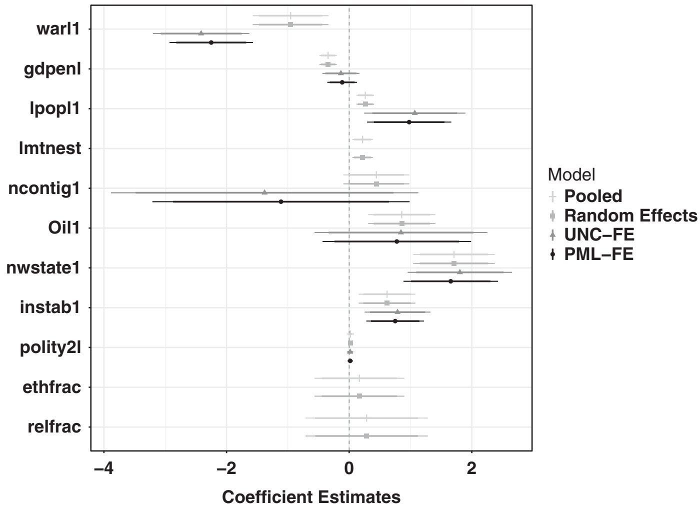

Figure 1. Unit heterogeneity in civil war onset (coefficients). Reanalysis of Fearon and Laitin (2003).

representing the government) are politically organized and engage in fighting resulting 500 or more battle-related deaths and 0 otherwise. We also include a battery of standard regressors that include both time-varying (lagged war, logGDP, logPopulation, Polity, Oil, Instability, Non-Contig) and time-invariant measures (logMountain, Ethnic Fractionalization, Religious Fractionalization).[22]

The results (presented in Figure 1) exhibit substantial differences between the various estimation strategies. There are notable differences between pooled/random effects and any of the fixed effects estimations.[23] While some measures are robust to all specification—instability and population—others lose (gain) significance from the alternative approaches included, most notably, GDP per capita (gdpenl). As in Fearon and Laitin (2003), the pooled results suggest a negative and significant effect of GDP on civil war. The same is true when random intercepts are including, produces results that are effectively identical to those from the pooled model.[24] The fixed effects models, however, suggest that the relationship between GDP and civil war is insignificant. This is substantial given that GDP is widely considered to be the most robust predictor of civil war onset (Hegre and Sambanis 2006). However, our results indicate that there is no within-unit effect of development on civil war, suggesting that these previous findings have been spurious.

This finding in itself, is not novel—several researchers have observed and most dismissed it—what is more important is establishing that it is not a statistical artifact but instead a meaningful theory-driven result. While others have argued that these fixed effects results are biased—from “treating as non-informative all countries where we do not observe variation in the response,” (Buhaug and Gleditsch 2008, 227)—here we have shown that it is not due to sample censoring.

[22]Variables and specification are identical to the main model in Fearon and Laitin (2003).

[23]A table providing the numeric values for all results is made available in Appendix II.

[24]On the similarity between the pooled and random effects results, we make two additional notes. First, this equivalence is also confirmed by the non-significant Likelihood Ratio test, $p = 0.497$. Second, this result is a consequence of the inclusion of *ncontig* in the model. As shown in Appendix II, this variable has a between-to-within ratio of nearly 8. As such, including it accounts for much of the between variation that would have otherwise been accounted for by the random intercepts. The main finding of our illustration—the difference in the estimate of GDP in the PML-FE versus RE/pooled—remains whether this variable is included or not. Given that it was part of the original analysis by Fearon and Laitin (2003), we choose to leave it in our replication.

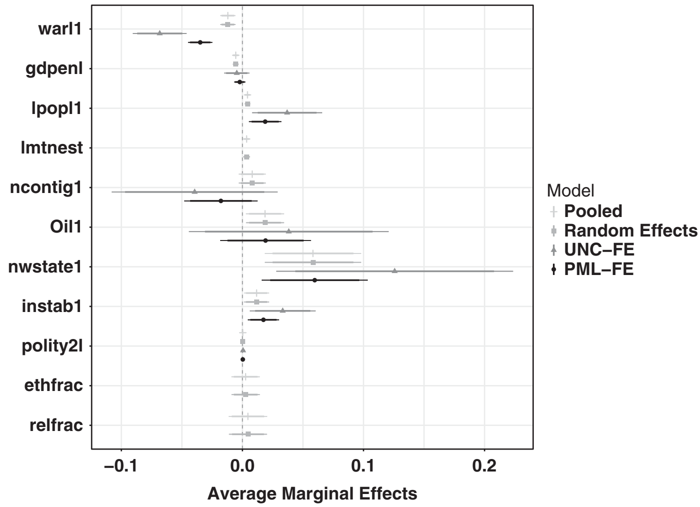

Figure 2. Unit heterogeneity in civil war onset (average marginal effects). Reanalysis of Fearon and Laitin (2003).

Instead, the results suggest that there is no evidence that within-unit changes to development affect the probability of civil war onset.[25]

Next, we calculate the marginal effects of the determinants of civil war for each of the models (presented in Figure 2).[26] Focusing on those predictors that are significant across all models, we observe quite different AMEs estimates (lpopl1, instab1, and warl1), with the fixed effects models suggesting larger effects. Comparing the unconditional fixed effects model and PML-FE, we see that unconditional fixed effects always returns a larger estimate. This is consistent with our expectation from Section 2 that because unconditional FE estimators will overestimate the baseline probability of conflict they will also overestimate the marginal effect of predictors. Here we see dramatic examples of that, as the effect of lagged war (*warl*) is nearly twice as large in the unconditional FE model (−0.068) as it is in the PML-FE model (−0.035). This is a dramatic shift given that the underlying specification of these two models is quite similar (i.e., the inclusion of unit dummies), with the only difference coming in how censored cases are treated in estimation.

Conclusion

While some may view these considerations as technical nuisance, we see the question of unit heterogeneity as deeply theoretical and central in any comparative analysis—engaging how and why units vary. Only by accurately understanding these features of their data can researchers draw credible inferences on the relationship of interest. That one's outcome happens to binary

[25]We note that this finding is actually consistent with formal models of rebellion, which have been unable to support a direct link between per capita income and conflict. Building on the canonical bargaining model of war, Chassang and Padro-i Miquel (2009) conclude that conflict is not a function of the productive capacity of the state. In short, while the opportunity cost of rebellion is diminished in poor states so too are the benefits of success. In this respect, the "costs and benefits from fighting move proportionately to the size of the economy, yielding no natural link" (Chassang and Padro-i Miquel (2009, 220). Fearon (2008) reaches a similar conclusion when model conflict as a contest model, finding that since the realized gains from fighting increase in proportion to the wealth of the state, there is no reason to suspect reduced violence (i.e., the bigger the pie the greater incentive to fight). He concludes that this result undermines support for poverty-based explanations of war commonly given in the empirical literature.

[26]Note that marginal effects for the random effects model are generated in Stata, since the margins function in R does not currently support glmer objects.

and rare does not alter these considerations, despite regular attempts by researchers in comparative politics and international relations to use these data features as a rationale for the adoption of estimation strategies inconsistent with their stated assumptions.[27]

One of the reasons most commonly mentioned for avoiding fixed effects is the censoring of units which do not experience the event of interest. Yet, how and why this should matter does not seem well understood. Here we have clarified that the main consequence is an overestimation of the average unit effect—that is, the baseline probability of the outcome—and, in turn, an overestimation of the AMEs of predictors. To redress this, we have suggested a novel estimation strategy for unconditional fixed effects—PML-FE—which uses a penalty to the score function to retain all sample units. Our simulation results suggest that with rare events BTSCS data PML-FE produces superior estimates to pooled, random effects, and conventional fixed effects estimators when there is a correlation between the unit effects and the predictors. As such, researchers should strongly consider this model when they believe such endogeneity concerns are present.

In future work, we plan to expand on this analysis in two ways. First, extending our simulations to explore the properties of PML-FE for dyadic data with standard interstate sampling dimensions, which contain tens or even hundreds of thousands of observations. This presents two challenges: first, given the dimensions of these data PML-FE is very computationally inefficient; second, Firth logit has been shown to risk incorrect inferences with very large data sets, a rare binary outcome, and a rare binary predictor (McGrath 2018). To overcome these issues, Cook and McGrath (n.d.) present an estimator using both a case-control sampling design and a penalized fixed effects model. In a second extension, we also intend to show the importance of retaining all units in spatial models of binary outcomes. Censoring the sample would bias estimates of cross-sectional dependence, forcing researchers to choose between fixed effects or (unbiased) spatial econometric models. Given that spatial dependence is a common feature of TSCS data, being able to simultaneously account for both is often necessary.

Supplementary Material. To view supplementary material for this article, please visit https://doi.org/10.1017/psrm.2018.40

Acknowledgments. For their helpful comments and suggestions, thanks to Neal Beck, Matthew Blackwell, Patrick Brandt, Ioannis Kosmidis, Jonathan Kropko, Jong Hee Park, Michael Peress, Carlisle Rainey, Curt Signorino, two anonymous reviewers, and the editor for their helpful suggestions. All remaining errors are our own. Inquiries should be directed to the corresponding author at sjcook@tamu.edu

References

Acemoglu D, Johnson S, Robinson JA and Yared P (2008) Income and Democracy. *American Economic Review* **98**, 808–842.

Beck N (2011) Is OLS with a Binary Dependent Variable Really OK? Estimating (Mostly) TSCS Models with Binary Dependent Variables and Fixed Effects. Working Paper, Annual Meeting of the Society of Political Methodology.

Beck N (2015) Estimating Grouped Data Models with a Binary Dependent Variable and Fixed Effects: What Are the Issues? Working Paper, Annual Meeting of the Society of Political Methodology. Available at https://pdfs.semanticscholar.org/65b4/5e49eff7f11a7ce0cb5780a8adcb9f311750.pdf.

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

Bell A and Jones K (2015) Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data. *Political Science Research and Methods* **3**, 133–153.

Buhaug H and Gleditsch KS (2008) Contagion or Confusion? Why Conflicts Cluster in Space1. *International Studies Quarterly* **52**, 215–233.

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

Chassang S and Padro-i Miquel G (2009) Economic Shocks and Civil War. *Quarterly Journal of Political Science* **4**, 211–228.

Clark TS and Linzer DA (2015) Should I Use Fixed or Random Effects? *Political Science Research and Methods* **3**, 399–408.

[27]Many literatures—e.g., inter- and intrastate war, conflict intervention, democratization, banking and currency crises, trade dispute initiation, the formation of alliances or regional trade pacts, etc.—in Comparative Politics and International Relations analyze rare event binary-outcome TSCS data.

Collier P and Hoeffler A (2004) Greed and Grievance in Civil War. *Oxford Economic Papers* **56**, 563–595.
Cook SJ, **Blas B**, **Carroll RJ and Sinha S** (2017) Two Wrongs Make a Right: Addressing Underreporting in Binary Data from Multiple Sources. *Political Analysis* **25**, 223–240.
Cook SJ and McGrath LF (n.d.) Unit Heterogeneity in Large Datasets with Rare Events. Working Paper, In Progress.
Cook SJ, **Niehaus J and Zuhlke S** (2018) A Warning on Separation in Multinomial Logistic Regression. *Research & Politics* **5**, 1–5.
Copas JB (1988) Binary Regression Models for Contaminated Data. *Journal of the Royal Statistical Society, Series B* **50**, 225–265.
Fearon J (2008) Economic Development, Insurgency, and Civil War. In E Helpman (ed), *Institutions and Economic Performance*. Cambridge, MA: Harvard University Press.
Fearon JD and Laitin DD (2003) Ethnicity, Insurgency, and Civil War. *American Political Science Review* **97**, 75–90.
Firth D (1993) Bias Reduction of Maximum Likelihood Estimates. *Biometrika* **80**, 27–38.
Gelman A, **Jakulin A**, **Pittau MG and** , **Su Y-S** *et al.* (2008) A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models. *The Annals of Applied Statistics* **2**, 1360–1383.
Green DP, **Kim SY and Yoon DH** (2001) Dirty Pool. *International Organization* **55**, 441–468.
Greene W (2004) The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects. *The Econometrics Journal* **7**, 98–119.
Heckman JJ (1981) The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process and Some Monte Carlo Evidence. In C Manski and D McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications*, 114–178. Cambridge, MA: MIT University Press.
Hegre H and Sambanis N (2006) Sensitivity Analysis of Empirical Results on Civil War Onset. *Journal of Conflict Resolution* **50**, 508–535.
Heinze G and Schemper M (2002) A Solution to the Problem of Separation in Logistic Regression. *Statistics in Medicine* **21**, 2409–2419.
King G (2001) Proper Nouns and Methodological Propriety: Pooling Dyads in International Relations Data. *International Organization* **55**, 497–507.
Lancaster T (2000) The Incidental Parameter Problem Since 1948. *Journal of Econometrics* **95**, 391–413.
Lesaffre E and Spiessens B (2001) On the Effect of the Number of Quadrature Points in a Logistic Random Effects Model: An Example. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **50**, 325–335.
McGrath LF (2018) Problems with Penalised Maximum Likelihood and Jeffrey's Priors to Account for Separation in Large Datasets with Rare Events. Available at https://www.dropbox.com/s/x5csn91nmt09bv8/mcgrath_separation.pdf, accessed 4 May 2018.
Nel P and Righarts M (2008) Natural Disasters and the Risk of Violent Civil Conflict. *International Studies Quarterly* **52**, 159–185.
Neyman J and Scott EL (1948) Consistent Estimates Based on Partially Consistent Observations. *Econometrica: Journal of the Econometric Society*, 1–32.
Oneal JR and Russett B (2001) Clear and Clean: The Fixed Effects of the Liberal Peace. *International Organization* **55**, 469–485.
Plümper T and Troeger VE (2007) Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects. *Political Analysis* **15**, 124–139.
Plümper T and Troeger VE (2011) Fixed-Effects Vector Decomposition: Properties, Reliability, and Instruments. *Political Analysis* **19**, 147–164.
Rainey C (2016) Dealing with Separation in Logistic Regression Models. *Political Analysis* **24**, 339–355.
Sambanis N (2001) Do Ethnic and Nonethnic Civil Wars Have the Same Causes?: A Theoretical and Empirical Inquire (Part 1). *Journal of Conflict Resolution* **45**, 259–282.
Wright J (2009) How Foreign Aid Can Foster Democratization in Authoritarian Regimes. *American Journal of Political Science* **53**, 552–571.
Zorn C (2005) A Solution to Separation in Binary Response Models. *Political Analysis* **13**, 157–170.

Cite this article: Cook SJ, Hays JC, Franzese RJ 2018. Fixed effects in rare events data: a penalized maximum likelihood solution. Political Science Research and Methods X: 1–14, doi:10.1017/psrm.2018.40