Explaining Recruitment to Extremism: A Bayesian Hierarchical Case-Control Approach

Roberto Cerina*[1], Christopher Barrie[2], Neil Ketchley[3], and Aaron Y. Zelin[4]

[1]Maastricht University
[2]University of Edinburgh
[3]University of Oxford
[4]Brandeis University

Forthcoming in *Political Analysis*

Abstract

Who joins extremist movements? Answering this question is beset by methodological challenges as survey techniques are infeasible and selective samples provide no counterfactual. Recruits can be assigned to contextual units, but this is vulnerable to problems of ecological inference. In this article, we elaborate a technique that combines survey and ecological approaches. The Bayesian hierarchical case-control design that we propose allows us to identify individual-level and contextual factors patterning the incidence of recruitment to extremism, while accounting for spatial autocorrelation, rare events, and contamination. We empirically validate our approach by matching a sample of Islamic State (ISIS) fighters from nine MENA countries with representative population surveys enumerated shortly before recruits joined the movement. High status individuals in their early twenties with university education were more likely to join ISIS. There is more mixed evidence for relative deprivation. The accompanying `extremeR` package provides functionality for applied researchers to implement our approach.

*Author contributions: NK and CB conceived of the study. RC developed the models. RC, CB, and NK contributed to the analysis. RC, CB, and NK developed the R package and the documentation. CB, NK, and AZ contributed to the data collection. CB, NK, and AZ developed the literature review. RC, NK, CB, and AZ contributed to the writing. We received helpful feedback and advice from Sir David Cox, Thomas Hegghammer, Bjørn Høyland, Bent Nielsen, Jacob Aasland Ravndal, and Frank Windmeijer. Hertog et al. (2021) reached out to us as we were finalizing our manuscript. Their analysis also uses leaked ISIS recruitment data to analyze the socio-economic correlates of joining the movement. To implement the methods described in this paper, see the associated R package, as well as documentation and vignettes: `http://extremeR.info`

1 Introduction

Identifying who is more likely to join an extremist movement is a pressing issue for both political science and public policy. However, empirical research on this topic is beset by methodological challenges. Population surveys offer little insight into the phenomenon as recruits to extremism are tiny minorities in any society, and so are tiny minorities in samples. This is before obvious problems related to eliciting truthful responses to questions probing illicit actions. Recent innovations in survey and online digital trace methodologies have allowed researchers to obtain more accurate measures of support for extremism (Bail et al. 2018; Blair et al. 2013; Corstange 2009; Mitts 2019). However, these approaches capture attitudes rather than behaviour. For researchers interested in why some individuals join extremist movements and not others, the most common strategy is to collect a convenience sample of recruits. Using these data, scholars typically either: i) report sample proportions of a given characteristic, e.g. the percentage of recruits who have university education; or ii) assign recruits to meaningful contexts and use the characteristics of those places to explain variation in the recruitment rate. While the first approach is descriptively useful, it fails to account for population baselines and other confounding factors affecting the incidence of recruitment. The second approach does provide a counterfactual and allows for multivariate analysis but suffers from familiar problems of ecological inference (Robinson 1950).

The method we propose in this paper allows researchers to leverage both survey *and* contextual data to make robust inferences about the individual and ecological correlates of recruitment to extremism. To do so, we take inspiration from the case-control design used in epidemiology and show how it can be adapted to combine a convenience sample of cases (recruits to extremism) with controls (respondents from a representative survey). In this, we build on the recent introduction of case-control methods to political science by Rosenfeld (2017; 2018), who shows how this design can be used to study protest participation and other forms of rare political behavior. Several statistical challenges arising from the nature of extremism remain, however. In particular, popular approaches for modelling rare events (King and Zeng 2001) do not account for hierarchical data structures or spatial autocorrelation in the incidence of recruitment. We also have to account for potential separation issues and the possibility of contamination between cases and controls (Rosenfeld 2018).

Our approach offers a complete solution to these statistical problems and can be described as a hierarchical, Bayesian case-control design that is robust to rare events, contamination, and spatial autocorrelation patterning the incidence of recruitment (Rota et al. 2013). Following Rosenfeld (2018), the Bayesian approach is preferable for a number of reasons. First, it permits the use of informative priors to account for the true prevalence of the event of recruitment, as well as to regularize coefficient estimates to account for separation bias and instability when carrying out regressions (Heinze and Schemper 2002). Second, in the absence of prior knowledge of the overall propensity of being a recruit in a given context, the model can estimate the propensity from the data (Rota et al. 2013). Finally, Bayesian probabilistic programming software provides unique flexibility in the modeling of the complex hierarchical structures characterizing recruitment into extremism.

A great strength of our method — and the open source software that accompanies this paper — is that applied extremism researchers can choose those parameters most relevant for their case. When sampling from national populations, the risk of contamination between cases and controls may be sufficiently low such that it does not pose a threat to inference. On the other hand, recruitment may not qualify as a rare event when comparing recruits to certain sub-populations. So too, spatial autocorrelation in recruitment may not apply if sampling from a small area or closed context. Our modelling strategy is flexible to the inclusion or exclusion of these parameters, depending on the case at hand. In support of our approach, and to help guide the modelling decisions of future practitioners, we provide practical advice and an extensive simulation study that compares our model to alternative frameworks, and show its robustness and superiority in predicting the true underlying probability of recruitment under various bias-inducing scenarios.

To display some of the key properties of our modeling strategy, we analyze recruitment of Sunni Muslim males in nine MENA countries to the Islamic State in Iraq and Syria (ISIS). We focus our analysis on an individual's level of education and social status — two key factors associated with recruitment to extremism found in the literature on violent Islamist movements (Gambetta and Hertog 2016; Krueger and Maleckova 2003; Krueger 2017; Morris 2020; Mesquita 2005). We show how our approach can be used to perform two types of analyses. In the first, we leverage a multilevel regression model trained on a cross-national sample of ISIS recruits and non-recruits. This provides a robust descriptive analysis about the individual-level characteristics of recruits across countries and sub-national administrative units. A second analysis focuses on two countries for which we have rich contextual information: Egypt and Tunisia. This analysis adds value by adjusting for local heterogeneity with the addition of relevant ecological covariates, allowing us to ascertain the potential sensitivity of individual-level findings to unobserved contextual confounding.

For the purposes of illustration, we implement the complete solution described above, accounting for spatial autocorrelation in recruitment, the possibility of contamination, and separation in our regression coefficients. Overall, we find that high-status males with university education in their early twenties were more likely to join ISIS. We also find that relatively deprived males in Egypt were more likely to join ISIS, but not in Tunisia. This heterogeneity in the individual and contextual correlates of violent extremism demonstrates the importance of accounting for both individual and context specific factors.

2 Explaining recruitment to extremism

A common strategy available to researchers interested in the correlates of recruitment to extremism is to sample on the dependent variable, obtaining relevant demographic information on individual extremists or members of extremist movements. In the ideal scenario, researchers are able to obtain movement membership lists, which can reveal information on tens of thousands of individuals (e.g. Biggs and Knauss 2012), although in practice such complete data is rare. Absent such lists, a well-established strategy is to leverage data from arrests or killings to generate samples of participants (e.g. Ketchley and Biggs 2017; Krueger and Maleckova 2003; Skare 2022). Alternatively, researchers can look to collect demographic information on extremists by either interviewing former recruits (e.g. Bérubé et al. 2019; della Porta 2013), or by reconstructing the biographical profiles of prominent individuals from open source information (e.g. Gambetta and Hertog 2016; Jensen et al. 2020; Ketchley et al. 2021). Per Rosenfeld (2018), a principle limitation of these samples is that they do not provide information on individuals outside of the subpopulation of interest, meaning that it is not possible to compare recruits to the population from which they are drawn. To remedy this, researchers typically either confine attention to variation amongst recruits (e.g. Morris 2020), or else assign individuals to meaningful contexts, e.g. universities, cities, or countries, and then use the characteristics of those units to explain cross-sectional variation in the recruitment rate (e.g. Barrie and Ketchley 2018; Pape 2021). While this latter approach is undoubtedly superior to simply analyzing sample proportions, it inevitably relies on ecological inference.

2.1 A hierarchical Bayesian case-control design

In what follows, we suggest two new methods for analyzing recruitment to extremism. The first leverages a cross-national, multilevel regression model trained on a complete sample of recruits and survey respondents. This provides a robust descriptive analysis about the individual-level factors which characterize recruits across countries and subnational units. The model uses random effects to control for unobservable subnational heterogeneity; these are preferable to fixed effects due to potentially heavily imbalanced area-level sample sizes (Gelman and Hill 2006; Clark et al. 2015). The model further uses a conditionally auto-regressive prior (Besag et al. 1991; Morris et al. 2019) to account for spatial smoothing. The second analysis focuses on single country studies where rich contextual information is available. The added value of this analysis lies in controlling for local heterogeneity in order to ascertain the robustness of any individual-level findings to contextual confounding. Taken together, our proposed setup thus plots a way forward for researchers to combine survey and ecological information for the robust analysis of recruitment to extremism.

2.2 Simple case-control set-up

We begin by describing the backbone of our model, which is a logistic regression accounting for case-control sampling protocol via an offset. Borrowing from Rota et al. (2013), we define $r_i = \{0, 1\}$ as the set of states that observation i in our sample of size $n = n_0 + n_1$ can obtain, where $r_i = 1$ implies the observation is a 'case', $r_i = 0$ defines a control, $n_1 = \sum_i^n \mathbb{1}(r_i = 1)$ and $n_0 = \sum_i^n \mathbb{1}(r_i = 0)$. In our application, a 'case' would refer to a known extremist; a 'control' to a survey respondent. Recall that cases are selected entirely on the dependent variable while controls come from the population that cases are drawn from. Take N_1 to represent the number of cases in the population of interest, and N_0 the number of controls. The probability of being included in the sample ($s_i = 1$) conditional on the true state of any individual can hence be understood as $P_1 = \Pr(s_i = 1 \mid r_i = 1) = \frac{n_1}{N_1}$, while that of being sampled as a control is $P_0 = \Pr(s_i = 1 \mid r_i = 0) = \frac{n_0}{N_0}$. The log-ratio of these sampling probabilities can then be used as an 'offset' in a logistic regression, to account for the sampling protocol. The hierarchical specification of the model follows, with regression coefficients being assigned a very weakly informative prior;[1]

$$r_i \sim \text{Bernoulli}(\rho_i); \tag{1}$$

$$\text{logit}(\rho_i) = \log\left(\frac{P_1}{P_0}\right) + \sum_k x_{i,k}\beta_k; \tag{2}$$

$$\beta_k \sim N(0, 10). \tag{3}$$

The above hierarchical model thus contains three layers: layer (1) is a model of the true state of an observation, conditional on their latent propensity ρ; layer (2) describes this latent propensity, by accounting for systematic variation due to heterogeneity in covariates; layer (3) models the effects of each covariate by assigning a prior probabilistic model.

2.3 Contaminated controls

Recall that that the case-control setup as described above takes known recruits and combines them with 'controls' taken from survey respondents. While we know that our cases are correctly labeled, we do not know whether this is true of our controls. That is, our controls may be 'contaminated' as survey respondents may have become recruits (Lancaster and Imbens 1996; Rosenfeld 2018). This is especially concerning when re-

[1] The normal distribution in our model (and in `Stan`) is parameterized by mean and standard deviation. See `https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations` for prior-choice advice when using `Stan`.

searchers have access to biographical information on tens of thousands of extremists (e.g. Biggs and Knauss 2012) or are comparing recruits to small sub-populations (e.g. Ketchley and Biggs 2017; Ketchley et al. 2021). Rota et al. (2013) outline a 'latent variable' formulation of their contamination model. Below we present our version of that same model as a mixture, which we find more intuitive.

The 'label' of an observation, $y_i = \{0, 1\}$, is observed for all observations, while the true 'state' of an observation, $r_i = \{0, 1\}$, is only observed for cases. The implied probability distribution of labels conditional on being a control is:

$$\Pr(y_i = 1 \mid r_i = 0, s_i = 1) = 0 = \theta_0;$$
$$\Pr(y_i = 0 \mid r_i = 0, s_i = 1) = 1 = (1 - \theta_0);$$

Due to contamination, it is possible that observations characterized by $y_i = 0$ are actually in state $r_i = 1$; hence we need a probability distribution for $y \mid r_i = 1$. Let $\pi = \frac{N_1}{N_1+N_0}$ be the prevalence of recruits in the population of interest, and let $n_u = \sum_i^n \mathbb{1}(y_i = 0)$ be the number of unlabeled observations. We expect there to be πn_u cases amongst the unlabeled observations. We can then characterize the probability distribution of labels, conditional on being a case, as:

$$\Pr(y = 1 \mid r = 1, s = 1) = \frac{n_1}{n_1 + \pi n_u} = \theta_1;$$
$$\Pr(y = 0 \mid r = 1, s = 1) = \frac{\pi n_u}{n_1 + \pi n_u} = (1 - \theta_1).$$

Finally, our model for the latent state r_i must reflect the possibility of contamination. We do this by re-defining the relative-risk of being sampled as:

$$\frac{P_1}{P_0} = \frac{\frac{n_1+\pi n_u}{N_1}}{\frac{(1-\pi)n_u}{N_0}} = \frac{n_1}{\pi n_u} + 1.$$

The updated, hierarchical specification for the case-control model accounting for contaminated controls is then:

$$y_i \sim \text{Bernoulli}(\theta_{r_i}); \tag{4}$$

$$r_i \sim \text{Bernoulli}(\rho_i); \tag{5}$$

$$\text{logit}(\rho_i) = \log\left(\frac{n_1}{\pi n_u} + 1\right) + \sum_k x_{i,k}\beta_k; \tag{6}$$

$$\beta_k \sim N(0, 10). \tag{7}$$

In summary, we derive our labels via two distinct data generating processes, identified by a latent state $r_i = \{1, 0\}$. In the event that the latent state of a given record is that of a true control, $r_i = 0$, it is then impossible for this record to be labeled $y_i = 1$; conversely, if the latent state is that of a true case, $r_i = 1$, then it is still possible for a record to be labeled $y_i = 0$, with probability $(1 - \theta_1)$. This latter model describes the issue of contamination. Note that in our application, θ is always observed, and fed to the model as data.

2.4 Area-level random effects

Survey data and information on extremists often contain information on the origin or location of residence of individuals. And we can understand individuals as nested within geographical units of increasing sizes. Generalizing, we can exploit variance at three levels: the individual, some small-area, and some large-area.

These area effects could be incorporated in the model via fixed-effects, by expanding the design matrix to include relevant dummy-variables for each area of interest. We consider this strategy unwise when trying to explain recruitment to extremism and prefer a random-effects approach. In the case of rare forms of political behaviour, our geographical units at all levels of analysis will have relatively few observations (Gelman and Hill 2006). Additionally, for many units, we will have no cases. Finally, we know that lists of recruits are unlikely to be exhaustive; that is, we will not have data for every recruit hailing from every subnational unit or country. Here, a sample of recruitment data or similar can be treated as a non-probability sample — it is unlikely that we can have complete confidence the sample constitutes a complete or random sample of the population of interest. Given these concerns, a random effects approach is preferable as it means: 1) we are able to borrow strength across areas, which also increases efficiency, to produce more realistic estimates for the area-level coefficients (Clark et al. 2015; Baio 2012); 2) in the absence of more detailed knowledge about the data-generating process, the shrinkage effect obtained by partial pooling is more likely to shield our estimates from any systematic sampling bias among our cases (Gelman and Hill 2006).

We can also relax some of the theoretical bias associated with the shrinkage induced by random effects via incorporating observable area-level heterogeneity in the design-matrix as fixed effects (Gelman and Hill 2006). This is what we elect to do in single country analyses. Finally, it is worth highlighting that our goal is not to make inferences about area-level effects. Rather, we seek to strip our individual-level effects estimates of contested variance that may be associated with the provenance of the recruit. The resulting hierarchical model is as follows:

$$y_i \sim \text{Bernoulli}(\theta_{r_i}); \tag{8}$$

$$r_i \sim \text{Bernoulli}(\rho_i); \tag{9}$$

$$\text{logit}(\rho_i) = \log\left(\frac{n_1}{\pi n_u} + 1\right) + \sum_k x_{i,k}\beta_k + \phi_{l[i]} + \eta_{j[i]}; \tag{10}$$

$$\beta_k \sim N(0, 10); \tag{11}$$

$$\phi_l \sim N(0, \sigma_\phi); \tag{12}$$

$$\sigma_\phi = \frac{1}{\sqrt{\tau_\phi}}, \ \tau_\phi \sim \text{Gamma}(\epsilon, \epsilon); \tag{13}$$

$$\eta_j \sim N(0, \sigma_\eta); \tag{14}$$

$$\sigma_\eta = \frac{1}{\sqrt{\tau_\eta}}, \;\; \tau_\eta \sim \text{Gamma}(\epsilon, \epsilon); \tag{15}$$

where ϵ stands for some arbitrary number, chosen as a compromise to minimize the prior information and maximise the Markov chain Monte Carlo (MCMC) convergence speed and stability.

2.5 Spatial autocorrelation

The network ties connecting actors across space play an important role in recruitment to high-risk activism. Sometimes the ties connecting recruits will be available; more commonly this information will not be recoverable. In the absence of detailed network information, we propose controlling for network effects at levels of varying scale. We work on the assumption that network ties are more likely to form between individuals who are geographically proximate. Depending on the richness of the data on recruits, we may generate distance matrices between geographical units of varying size.

To account for area-level spatial autocorrelation, we incorporate a version of the conditional auto-regressive (CAR) model (Besag et al. 1991). This approach has been used in individual-level models of behaviour, enabling local smoothing of predictions according to behaviour observed in neighbouring areas (Selb and Munzert 2011). The key ingredients of a CAR model are $\boldsymbol{\omega}$, a distance-weight matrix; α, a parameter governing the degree of autocorrelation, where $\alpha = 0$ implies spatial independence, and $\alpha = 1$ implies an intrinsic conditional auto-regressive (ICAR) model (Besag and Kooperberg 1995); and σ_ψ, the standard deviation of the subnational unit effects. The resulting model for spatial random effect $\psi_l \; \forall \; l = \{1, ..., L\}$ is then:

$$\psi_l \mid \psi_{l'} \sim N\left(\alpha \sum_{l' \neq l} \omega_{ll'} \psi_{l'}, \sigma_\psi\right).$$

In practice, we implement the ICAR specification of the model, with $\alpha = 1$, and take $\boldsymbol{\omega}$ to be the neighbourhood matrix. The neighbourhood matrix has diagonals zero (a unit cannot neighbour itself) and off-diagonal zero or one depending on whether the given units are neighbours. We choose this specification of the distance matrix because of the efficiency gains it affords in a Bayesian context (Morris et al. 2019). This leads to:

$$\psi_l \mid \psi_{l'} \sim N\left(\frac{\sum_{l' \neq l} \psi_{l'}}{d_{l,l}}, \frac{\sigma_\psi}{\sqrt{d_{l,l}}}\right),$$

where $d_{l,l}$ is an entry of the diagonal matrix D of size $L \times L$, whose diagonal is defined as a vector of the number of neighbours of each area. The joint distribution of this model is simply a multivariate normal distribution $\boldsymbol{\phi} \sim N(0, [\tau_\psi(D - W)]^{-1})$, $\tau_\psi = \frac{1}{\sigma_\psi^2}$, which is conveniently proportional to the squared pairwise difference of neighbouring effects. Note that the sum-to-zero constraint is needed for identifiability, as in its absence any constant added to the ψs would cancel out in the difference.[2] Following Morris et al. (2019), setting the precision to 1 and centering the model such that $\sum_l^L \psi_l = 0$, we arrive at:

$$\log p(\boldsymbol{\psi}) \propto \exp\left\{-\frac{1}{2}\sum_{l' \neq l}(\psi_l - \psi_{l'})^2\right\};$$

The hierarchical model we implement to incorporate the spatial component is within the Besag-York-Mollié (BYM) family (Besag et al. 1991). For a given level of analysis, say the city or province in a cross-country analysis, BYM models are characterized by two random effects which explain unobserved heterogeneity: ϕ_l defines a non-spatial component while ψ_l defines systematic variance due to spatial dependency. The typical challenge with BYM is that the two areal effects cannot be identified without imposing

[2] This model has the disadvantage of being an improper-prior, as its density does not integrate to unity and is non-generative, though it serves our purposes within the context of a hierarchical model. The prior also encodes an intrinsic dependence between subnational units. It can no longer detect the degree of spatial autocorrelation supported by the data but instead assumes that areas are explicitly dependent, and estimates coefficients accordingly.

some structure since they are mutually dependent, meaning either component is capable of accounting for contested variance at the area-level. This leads to inefficient posterior exploration of any MCMC sample, and subsequent lack of convergence (Riebler et al. 2016). To overcome this, we implement a state-of-the-art solution leveraging penalized-complexity priors (Simpson et al. 2017), which proposes modelling the two effects as a scaled mixture such that:

$$\gamma_l = \sigma \left(\phi_l \sqrt{(1-\lambda)} + \psi_l \sqrt{(\lambda/s)} \right);$$

where ϕ and ψ are random effects scaled to have unitary variance and $\lambda \in [0,1]$ is a mixing parameter, defining the proportion of residual variation attributable to spatial dependency. In order for the spatial and unstructured effects to share σ, they must be on the same scale (Riebler et al. 2016). We must therefore scale the ICAR-distributed effects, as their original scale is defined by the local neighbourhood. A proposed scaling factor is chosen such that the geometric mean of the variance parameters over the areal units is 1, $\text{Var}(\psi_l) = 1$. Note that this scaling factor, s in the equation above, can be calculated directly from the adjacency matrix, and hence it is not to be estimated but passed to the model as data.

The resulting hierarchical specification of our model follows:

$$y_i \sim \text{Bernoulli}(\theta_{r_i}); \tag{16}$$

$$r_i \sim \text{Bernoulli}(\rho_i); \tag{17}$$

$$\text{logit}(\rho_i) = \log\left(\frac{n_1}{\pi n_u} + 1\right) + \sum_k x_{i,k}\beta_k + \gamma_{l[i]} + \eta_{j[i]}; \tag{18}$$

$$\beta_k \sim N(0, 10); \tag{19}$$

$$\gamma_l = \sigma \left(\phi_l \sqrt{(1-\lambda)} + \psi_l \sqrt{(\lambda/s)} \right); \tag{20}$$

$$\lambda \sim \text{Beta}(0.5, 0.5); \tag{21}$$

$$\phi_l \sim N(0, 1); \tag{22}$$

$$\psi_l \mid \psi_{l'} \sim N\left(\frac{\sum_{l' \neq l} \psi_{l'}}{d_{l,l}}, \frac{1}{\sqrt{d_{l,l}}}\right) \tag{23}$$

$$\sigma \sim \frac{1}{2}N(0,1); \tag{24}$$

$$\eta_j \sim N(0, \sigma_\eta); \tag{25}$$

$$\sigma_\eta = \frac{1}{\sqrt{\tau_\eta}}, \ \tau_\eta \sim \text{Gamma}(\epsilon, \epsilon); \tag{26}$$

where $\frac{1}{2}N$ denotes a half-normal distribution, which is the recommended prior for the variance of BYM effects (Morris et al. 2019).

2.6 Regularizing prior coefficients

Multiple contributions have highlighted problems with logistic regression coefficient estimates under rare-events (King and Zeng 2001). The intuition behind these challenges is typically described as some variation on the standard separation problem where any given covariate or simple combination thereof perfectly separates cases from controls. This leads to biased and unstable point-estimates with large associated uncertainty (Heinze 2017). A number of regularization techniques have been proposed to reduce bias and stabilize the coefficient estimates. Our preferred regularization method is that proposed by Gelman et al. (2008) and Ghosh et al. (2018). The approach assumes it should be unlikely to observe unit-changes in the (standardized) covariates that would lead to outcome changes as large as 5 points on the logit scale. Using a slight variation on this approach to ensure sufficient regularisation, we use a Cauchy prior with scale-parameter set to 1 for the regression coefficients, and a 'looser' scale of 10 logit points on the intercept to accomodate for the rarity of the event in the sample. The advantages of the Cauchy prior lie in its fat tails, which avoid over-shrinkage of large coefficients (Ghosh et al. 2018). We apply this prior to our fixed effects exclusively, as the likelihood of our random effects is already structured and penalized. Our final model specification is then as follows:

$$y_i \sim \text{Bernoulli}(\theta_{r_i}); \tag{27}$$

$$r_i \sim \text{Bernoulli}(\rho_i); \tag{28}$$

$$\text{logit}(\rho_i) = \log\left(\frac{n_1}{\pi n_u} + 1\right) + \sum_k x_{i,k}\beta_k + \gamma_{l[i]} + \eta_{j[i]}; \tag{29}$$

$$\beta_1 \sim \text{Cauchy}(0, 10); \tag{30}$$

$$\beta_k \mid k > 1 \sim \text{Cauchy}(0, 1); \tag{31}$$

$$\gamma_l = \sigma\left(\phi_l\sqrt{(1-\lambda)} + \psi_l\sqrt{(\lambda/s)}\right); \tag{32}$$

$$\lambda \sim \text{Beta}(0.5, 0.5); \tag{33}$$

$$\phi_l \sim N(0, 1); \tag{34}$$

$$\psi_l \mid \psi_{l'} \sim N\left(\frac{\sum_{l' \neq l}\psi_{l'}}{d_{l,l}}, \frac{1}{\sqrt{d_{l,l}}}\right) \tag{35}$$

$$\sigma \sim \frac{1}{2}N(0, 1); \tag{36}$$

$$\eta_j \sim N(0, \sigma_\eta); \tag{37}$$

$$\sigma_\eta = \frac{1}{\sqrt{\tau_\eta}},\ \tau_\eta \sim \text{Gamma}(\epsilon, \epsilon); \tag{38}$$

2.7 Simulation and practical advice

In the Supplementary Materials in Section D, we outline an extensive simulation study demonstrating the performance advantage of a hierarchical Bayesian case-control approach relative to competing strategies such as the King and Zeng model (2001), as well as a simple fixed-effects logistic regression. In the simulation study, we explicitly test the performance of our model under varying values for the following parameters: a) sample size (n); b) population prevalence (π); c) discrepancy between sample and population prevalence ($\pi - \hat{\pi}$); d) spatial auto-correlation (as measured by Moran's I). Two dimensions of our modeling framework remain untested: i) the sensitivity of the model to poor prior information about π, the population prevalence assumed for the contamination layer; ii) the model's ability to deal with non-probability samples resulting from exogenous selection effects (i.e. beyond the 'selection on the dependent variable' type). In Section E Supplementary Materials we provide actionable advice for researchers and discuss how these untested dimensions may affect the robustness of the model, in light of the results from the simulation study and the robust modeling framework we have adopted.

3 Who was more likely to join ISIS?

To illustrate our approach, we analyze a set of leaked border documents capturing recruitment to ISIS. This leak was widely covered in international news media and has been used to provide descriptive statistics on the geographical distribution and demographic characteristics of ISIS fighters from multiple MENA countries (Devarajan et al. 2016; Zelin 2018; Sterman and Rosenblatt 2018). For the case-control design, we combine individual-level ISIS recruitment data with a nationally representative sample of Muslim males from Wave III of the Arab Barometer (2014) survey. The fieldwork for the Arab Barometer surveys was completed *before* most recruits recorded in our border documents entered ISIS-held territory, and so may be vulnerable to contamination.[3]

Our choice of covariates to use from this survey is constrained by the information included in the border documents. We elect to include covariates for age, age squared, marital status, university education, and student status. We also combine two variables for unemployed and employment in agricultural or manual labor to create a composite variable designed to measure "low status" activity. An interaction between this variable and our university education variable is designed to capture relative deprivation; that is,

[3]See Supplementary Information for more information on these data.

whether highly educated individuals engaged in low status economic activity are more likely to become recruits. Full details of each covariate are listed in the Supplementary Materials.

A first model — which we refer to as the "Bird's Eye" approach — uses a multilevel regression model trained on the complete sample of $1,051$ recruits and $5,093$ unlabeled records. This first model provides a robust descriptive analysis of the individual-level factors characterizing recruits across countries and subnational units.

A second model — which we refer to as the "Worm's Eye" approach — incorporates contextual information for Egypt ($n_1 = 66, n_0 = 551$ complete records) and Tunisia ($n_1 = 426$, $n_0 = 589$ complete records) at the district level. We focus on these two countries due to the availability of contextual information at the district level that is not accessible for the other countries in our sample. The added value of this analysis lies in controlling for observable district-level heterogeneity in order to ascertain the robustness of any individual-level findings to contextual confounding. For both Egypt and Tunisia, we include variables to capture subnational differences in demographic and labor-market composition, employment opportunities, as well as more context-specific variables designed to capture support for Islamist political organizations and prehistories of contentious politics. Full details of all covariates are listed in the Supplementary Materials.

For the main analyses, we present: 1) the posterior density of fixed and random effects according to our models; 2) the posterior predictive distribution across potential recruitment profiles.[4]

[4]Convergence diagnostics are in the Supplementary Materials.

3.1 Fixed and Random Effects

Figure 1 presents the posterior density of the individual-level fixed effects in the Bird's Eye model; Figures 2a and 2b present the Worm's Eye equivalent. These plots contain the main results of our models. Note that all the covariates, including dummies, are centered and scaled, hence the coefficients are to be interpreted in terms of standard deviations from the mean of each covariate (Figures G.1, G.2 and G.3 in the Supplementary Materials are the individual-level posterior densities on the original, non-standardized scale). Since we are principally interested in the robust estimation of individual-level predictors, we display only the posterior density of individual fixed effects for all of our models.[5] To aid with interpretation, the mean and standard deviation of each covariate are reported in the legend of each plot.

The estimated intercepts for the three models are extremely low. For the Bird's Eye model, the log-odds are in the order of -11. For the Egypt Worm's Eye model, it is just over -13; in Tunisia it is -9. The size of the intercept is primarily driven by the size of the offset, which is in turn determined by the overall prevalence of recruitment. It is therefore not surprising that Egypt's intercept is so dramatically low, given the close-to-zero prevalence of recruitment when compared to population size ($\pi = \frac{4}{100{,}000}$) versus Tunisia where this prevalence is higher ($\pi = \frac{2}{1{,}000}$). For the Bird's Eye model, a different offset is provided for observations coming from different countries, to account for country-specific prevalence. The large and negative intercept underscores an important challenge in the explanation of why individuals join movements like ISIS: a linear combination of features capable of pushing an individual to become a recruit has to be extremely large, on the log-odds scale, to meaningfully affect the otherwise extremely low probability of recruitment.

[5]Figures G.4a and G.4b display the standardized district-level posterior densities, while G.5a and G.5b present district-level coefficients on the original, non-standardized scale.

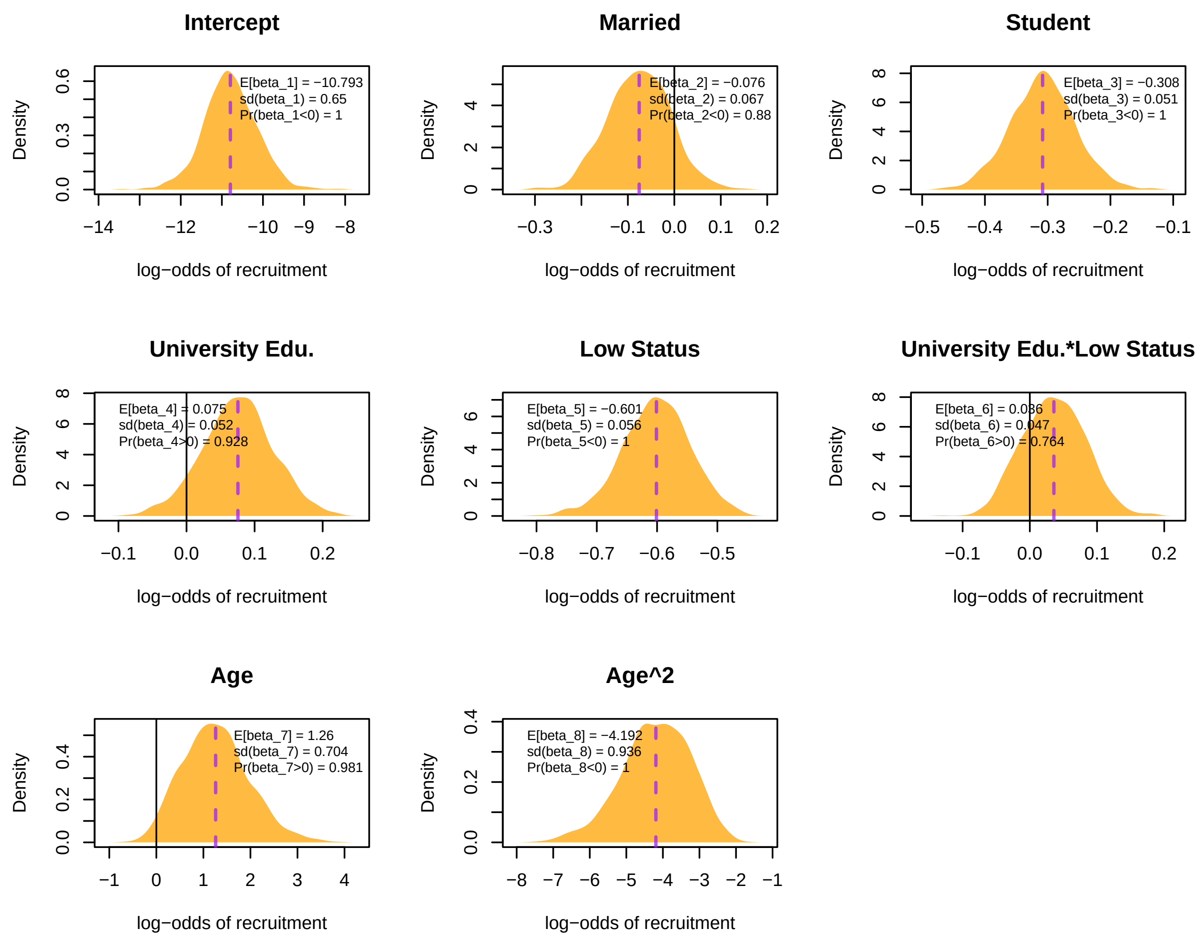
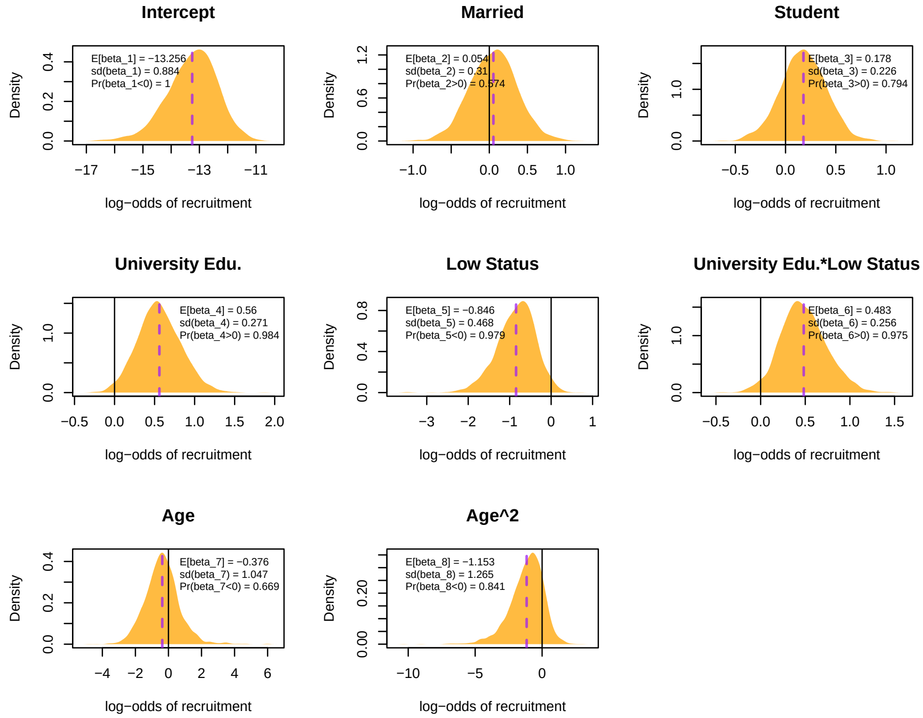
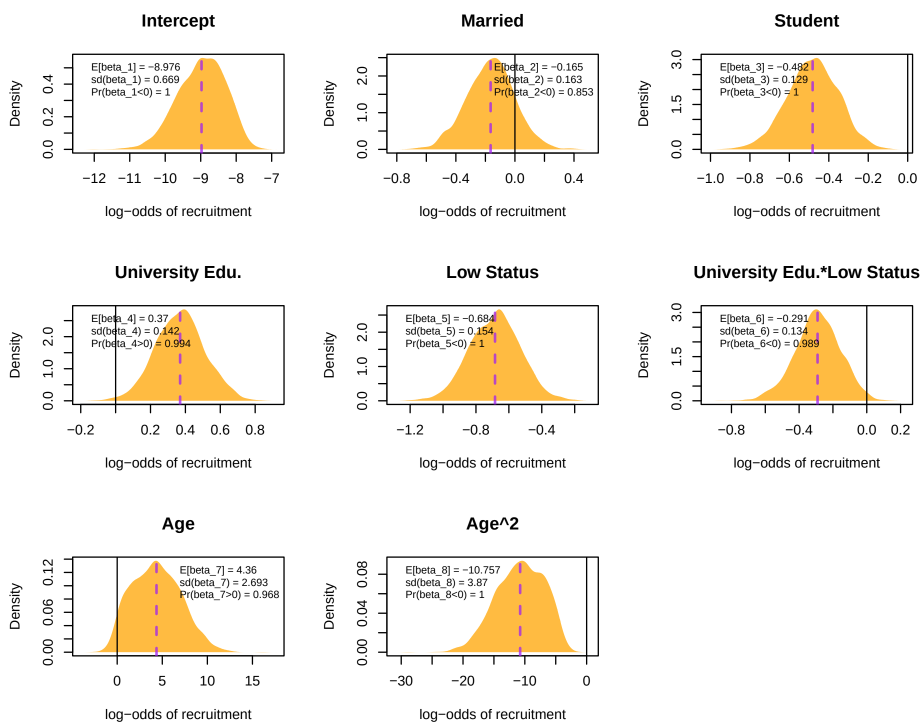


Figure 1: Posterior density of fixed-effect coefficients for the *Bird's Eye* model.

We focus primarily on testing the role of education and social status in an individual's decision to join ISIS. An individual who has university education and low-status is assumed to be relatively deprived. We compare predicted log-odds, as opposed to predicted probabilities, as these are scarcely comparable due to the powerful effect of the intercept, which drags probabilities of most profiles close to zero (though see Supplementary Figures H.1, H.2, and H.3 for predicted probabilities of recruitment relative to the 'average' profile, and Figures H.4, H.5, and H.6 for expected counts under different relative-deprivation scenarios). The total logit effects on probability of recruitment for different relative-deprivation profiles are shown in Figure 3 for the Bird's Eye model,



(a) Egypt



(b) Tunisia

Figure 2: Posterior density of fixed-effect coefficients for the *Worm's Eye* models.

and in Figures 4a and 4b for the Worm's Eye.

Relative deprivation finds mixed support: at the Bird's eye level, we find being high-status plays a key role in increasing propensity of being recruited, while having university education plays a more minor role. A similar pattern is evident in Tunisia, though the effect of being high-status and having university education is starker, meaningfully increasing the propensity to join ISIS by around 3 points on the log-odds scale compared to relatively deprived individuals. In Egypt the effects are more consistent with relative-deprivation, though note the large prediction intervals around the total effects of relatively-deprived individuals. There is also substantial overlap between the distributions in all plots. This is largely due to the uncertainty around the intercept, which plays a role in marginalising these effects. Note further that varying prediction intervals on the effects reflect the highly unbalanced prevalence of the groups in our study. All in all, the evidence from these analyses suggest that high-status individuals were more likely to be recruited by ISIS, and that being high status *and* having a university education further increases the likelihood of recruitment. The large prediction intervals, which result from uncertainty around the intercept, underscore that much remains unknown about the underlying systematic determinants of recruitment.

To fit the ICAR model, we implemented the fully-connected graph shown in Figure 5a. The spatially autocorrelated component dominates the governorate-level variance, as shown by the posterior of mixing parameter λ in Figure 5b, estimated via Monte-Carlo mean at close to 0.9, suggesting around 90% of the variance at the governorate level can be explained by the ICAR model.[6]

We repeat these analyses for Egypt and Tunisia. Figure 6 shows similar mixing among spatial and non-spatial components for the two countries, with around 15% of the district-level variance in Egypt being explained by spatial patterns, and 19% in

[6]The spatial distribution of point estimates for governorate and country-level random effects are presented in Figure I.3.

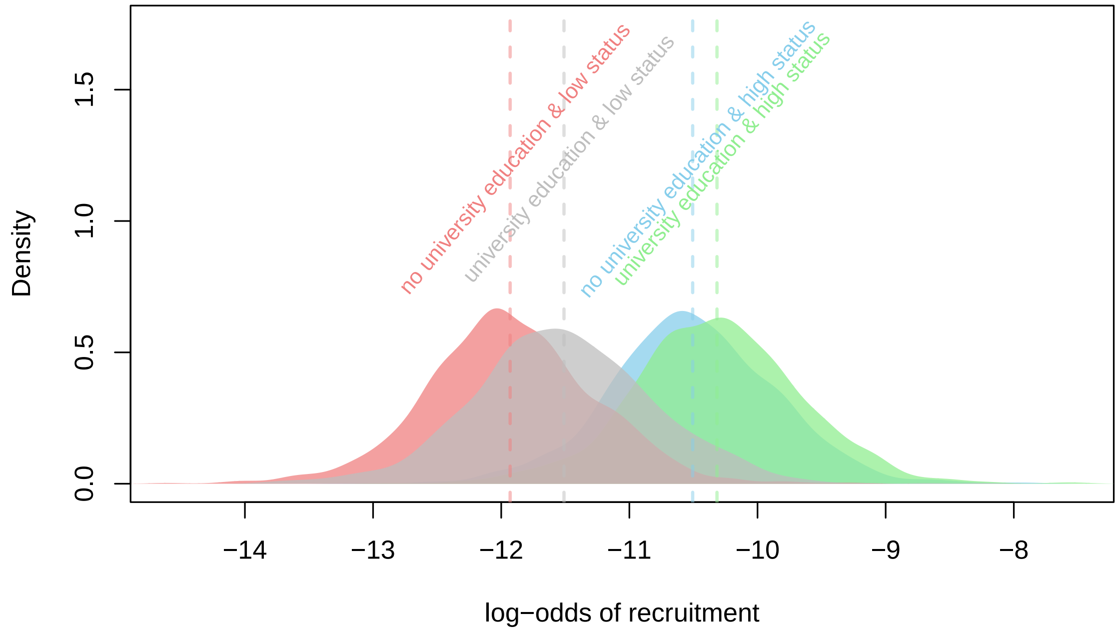
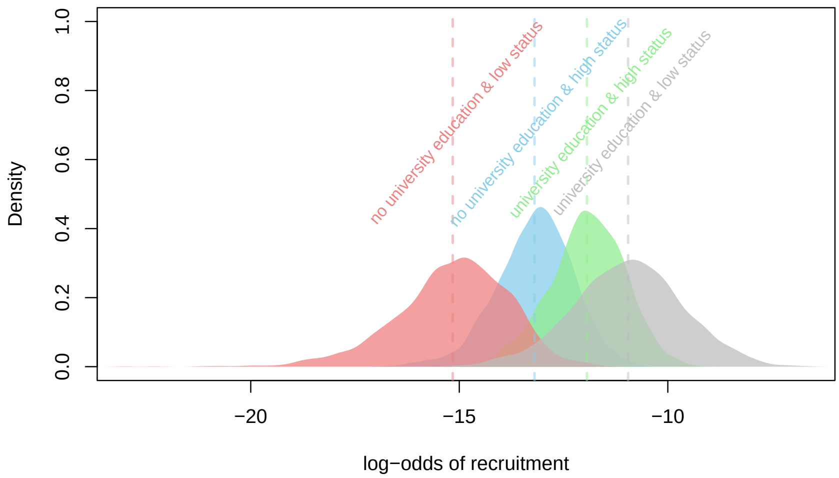


Figure 3: Predicted propensity of recruitment for relative-deprivation profiles according the *Bird's Eye* model.

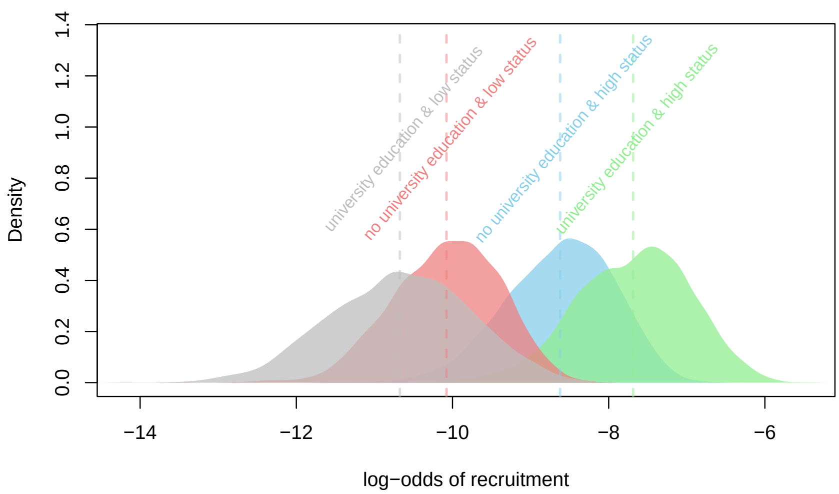
Tunisia. It is noteworthy that very few of our contextual variables have explanatory power for predicting recruitment. Coupled with the low percentage of variance being explained by the spatial components, our analysis suggests that, in spite of our best efforts to account for observable heterogeneity, there exist a vast array of unobserved, non-spatial district-level effects, which accounts for over 80% of the unexplained district-level variance in both Egypt and Tunisia. Hence this contextual variance, while properly accounted for, remains unexplained. In the Supplementary Materials we also describe Moran's I statistics for the Worm's Eye analysis as well as point estimates for the district and governorate effects in Egypt and Tunisia (Supplementary Figure I.1).

3.2 Predicted propensity of recruitment by profile

To conclude our analysis, we present inferences derived from the posterior predictive distribution of the out-of-sample probability of recruitment, focusing on individual-level

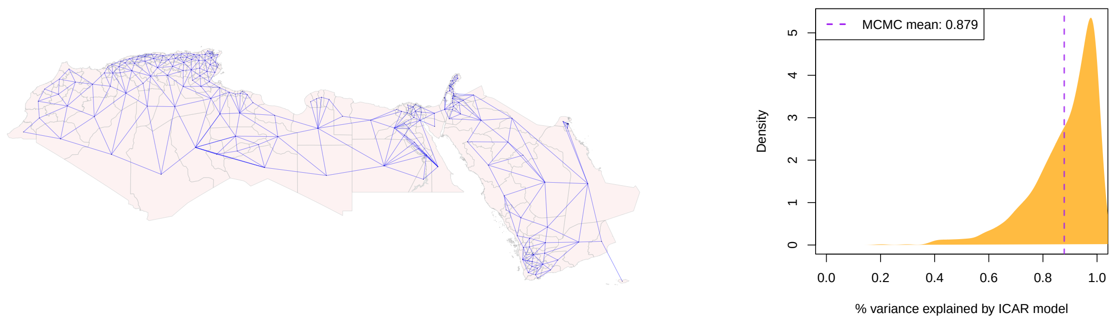


(a) Egypt



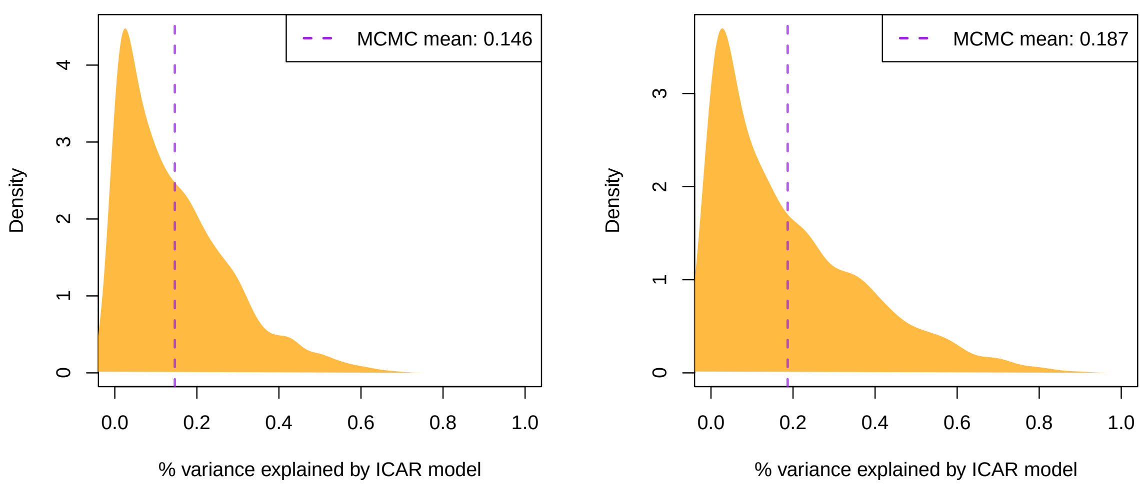
(b) Tunisia

Figure 4: Predicted propensity of recruitment for relative-deprivation profiles according to the *Worm's Eye* models.



(a) (b)

Figure 5: Fully-connected graph for the Bird's Eye model (a) and Governorate-level variance mixing parameter - λ (b).



(a) (b)

Figure 6: District-level variance mixing parameter - λ - for Egypt (a) and Tunisia (b).

characteristics.

What is the profile of individuals 'at risk' of recruitment to ISIS according to our models? We attempt to answer this question by analyzing the predicted probabilities of all possible theoretical profiles, defined by the individual-level characteristics available in our data. Every profile is assumed to come from a hypothetical 'average district'.

Figure 7 presents point estimates and prediction intervals for the log-odds of recruitment, over 160 possible profiles in the Bird's eye model. Similar plots displaying the absolute and relative probabilities of recruitment are available in Figures I.15 and I.14 in the Supplementary Materials. Table 1 presents the profiles of the top 10 most likely profiles to be recruited, providing four useful metrics to interpret the results: predicted probability; predicted rate per $10,000$ people; predicted odds relative to the average profile; and log-odds.

A note of caution on the interpretation of these visuals: these are useful summaries of the data, but the uncertainty around the point estimates tends to be relatively large. Taking 7 as an example, a qualitative interpretation of the uncertainty would be as follows: *it cannot be categorically ruled out that the most likely profile is actually ranked only* 30^{th} *(out of* 160*), though this would be very unlikely given the evidence implied by the data*. In general, we note that profiles which are at high-risk of recruitment are endowed with higher levels of certainty around their point estimates, suggesting that: i) it is possible to distinguish high-risk profiles from low-risk profiles (at least in Tunisia and in the Bird's eye view); ii) it is easier to distinguish between different high-risk profiles than it is between low-risk profiles. For Egypt, although we do observe a reduction in uncertainty at high levels of risk, we cannot entirely distinguish between low-risk and high-risk profiles, as a significant degree of overlap between posterior distributions is maintained across profiles. This is likely as a result of the relatively small sample of cases, and the large effect of the unexplained intercept.

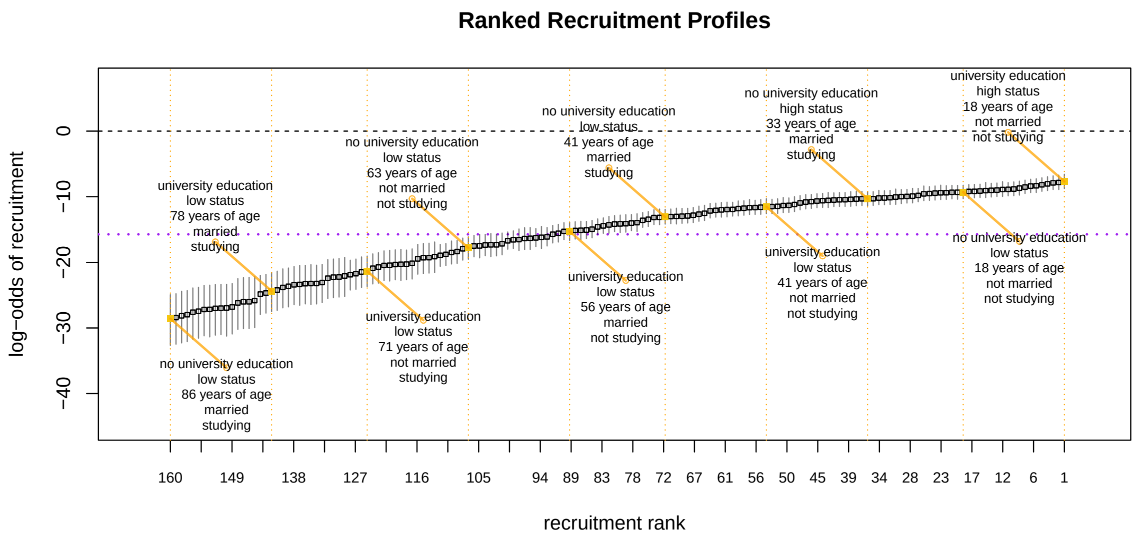


Figure 7: Distribution of the predicted probabilities prediction intervals, presented on the log-odds scale to aid cross-profile comparisons. The black dotted line highlights the the zero-log-odds point, while the purple dotted line notes a central estimate for the median recruitment propensity across profiles.

Rank	Married	Student	University Edu	Low-Status	Age	$\hat{\mathrm{P}}(r=1 \mid X=x)$	Predicted Rate	$\frac{\hat{\mathrm{P}}(r=1\mid X=x)}{\hat{\mathrm{P}}(r=1\mid X=\bar{X})}$	$\mathrm{logit}\left(\hat{\mathrm{P}}(r=1 \mid X=x)\right)$
1	0	0	1	0	18	0.000460	5/10000	22.561	-7.684
2	1	0	1	0	18	0.000398	4/10000	19.273	-7.828
3	0	0	0	0	18	0.000376	4/10000	18.632	-7.885
4	1	0	0	0	18	0.000324	3/10000	16.013	-8.036
5	0	0	1	0	26	0.000281	3/10000	13.745	-8.177
6	1	0	1	0	26	0.000241	2/10000	11.763	-8.331
7	0	0	0	0	26	0.000230	2/10000	11.287	-8.379
8	1	0	0	0	26	0.000197	2/10000	9.695	-8.530
9	0	1	1	0	18	0.000168	2/10000	8.295	-8.690
10	1	1	1	0	18	0.000147	1/10000	7.211	-8.828

Table 1: Top 10 recruitable theoretical profiles according to the Bird's eye model. Profiles are ordered by predicted probability of recruitment net of sampling protocol. 10 ages are evaluated, starting at 18 (to avoid non-existent profiles) and ending at the largest observed age (86). The last four columns represent respectively: i. the predicted probability of recruitment; ii. the predicted rate of recruitment per $10,000$ people; iii. the predicted odds of recruitment, relative to the 'average' profile; iv. the log-odds of recruitment.

From the Bird's Eye prediction intervals we notice that the predicted probability of recruitment is centered around -15 on the log-odds scale, again underscoring the rarity of becoming a recruit. A select number of profiles approach a predicted probability around -7, and translate to meaningful rates of recruitment; these are highlighted in the predicted probabilities table, which show the 10 most recruitable profiles. Looking at Table 1, we can say that the most likely recruit profile (loosely characterised as a young, high-status, Sunni male with some university education who is unmarried and not currently studying), is around 23 times as likely to be recruited as an average Sunni male from an average area in the MENA. For every $10,000$ members of the most recruitable profile across the region, we expect 5 to have joined ISIS. It is worthwhile to note that, consistent with Figure 3, all the most recruitable profiles are high-status individuals, and a majority of them has some university education. Unsurprisingly, all of these profiles are under-25, and not currently studying.

The Bird's Eye profiles are comparable to the Worm's Eye profiles for Tunisia (Figure 8 and Table 2) whereas the Egypt analysis points to stronger evidence for the relative deprivation hypothesis. In Egypt, a majority of the likely recruit profiles are relatively deprived (Figure 9 and Table 3).[7] The relative recruitment likelihood of the most susceptible profiles in Egypt and Tunisia is also greater. In Egypt, the most likely recruit profile (loosely characterised as a young, low-status, Sunni male with some university education who is married and is currently studying) is around 157 times as likely to be recruited as the average Egyptian Sunni male. The Egypt-specific recruitment propensity is dramatically lower than that of Tunisia, again highlighting the role of contextual effects. In Tunisia, the most likely recruit profile (loosely characterised as a young, high-status, Sunni male who has university education, is unmarried and is not currently studying) has a probability of recruitment equivalent to 0.04. This profile is

[7] For absolute and relative probabilities of recruitment from the Worm's eye models, see Supplementary Figures I.17, I.16, I.19, and I.18.

over 335 times as likely as the average Tunisian Sunni male to be recruited, highlighting that though recruitment is still relatively rare in the population, the probability of recruitment is far greater in the top recruitment profiles. Figure I.19 shows only a handful of profiles have predicted probabilities above $\frac{1}{100}$.

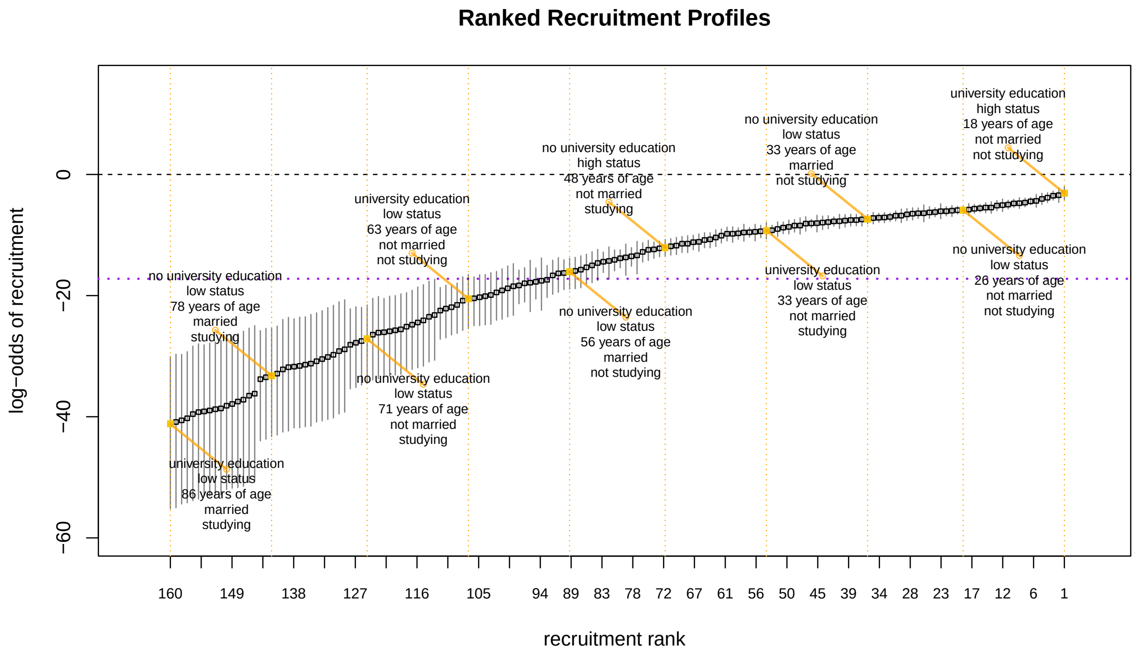


Figure 8: Worm's Eye (Tunisia) distribution of the predicted probabilities prediction intervals, presented on the log-odds scale to aid cross-profile comparisons. The black dotted line highlights the the zero-log-odds point, while the purple dotted line notes a central estimate for the median recruitment propensity across profiles.

Rank	Married	Student	University Edu	Low-Status	Age	$\hat{\mathrm{P}}(r=1 \mid X=x)$	Predicted Rate	$\frac{\hat{\mathrm{P}}(r=1 \mid X=x)}{\hat{\mathrm{P}}(r=1 \mid X=\bar{X})}$	$\mathrm{logit}\left(\hat{\mathrm{P}}(r=1 \mid X=x)\right)$
1	0	0	1	0	18	0.043680	437/10000	335.663	-3.086
2	1	0	1	0	18	0.031493	315/10000	242.502	-3.426
3	0	0	1	0	26	0.028925	289/10000	222.013	-3.514
4	1	0	1	0	26	0.021177	212/10000	159.752	-3.833
5	0	0	0	0	18	0.017397	174/10000	137.069	-4.034
6	1	0	0	0	18	0.012615	126/10000	98.539	-4.360
7	0	0	0	0	26	0.011647	116/10000	88.370	-4.441
8	0	0	1	0	33	0.009503	95/10000	73.103	-4.647
9	0	1	1	0	18	0.008662	87/10000	68.471	-4.740
10	1	0	0	0	26	0.008327	83/10000	64.331	-4.780

Table 2: Top 10 recruitable theoretical profiles according to the Tunisia 'Worm's Eye' model. Profiles are ordered by predicted probability of recruitment net of sampling protocol. 10 ages are evaluated, starting at 18 (to avoid non-existent profiles) and ending at the largest observed age (86). The last four columns represent respectively: i. the predicted probability of recruitment; ii. the predicted rate of recruitment per $10,000$ people; iii. the predicted odds of recruitment, relative to the 'average' profile; iv. the log-odds of recruitment

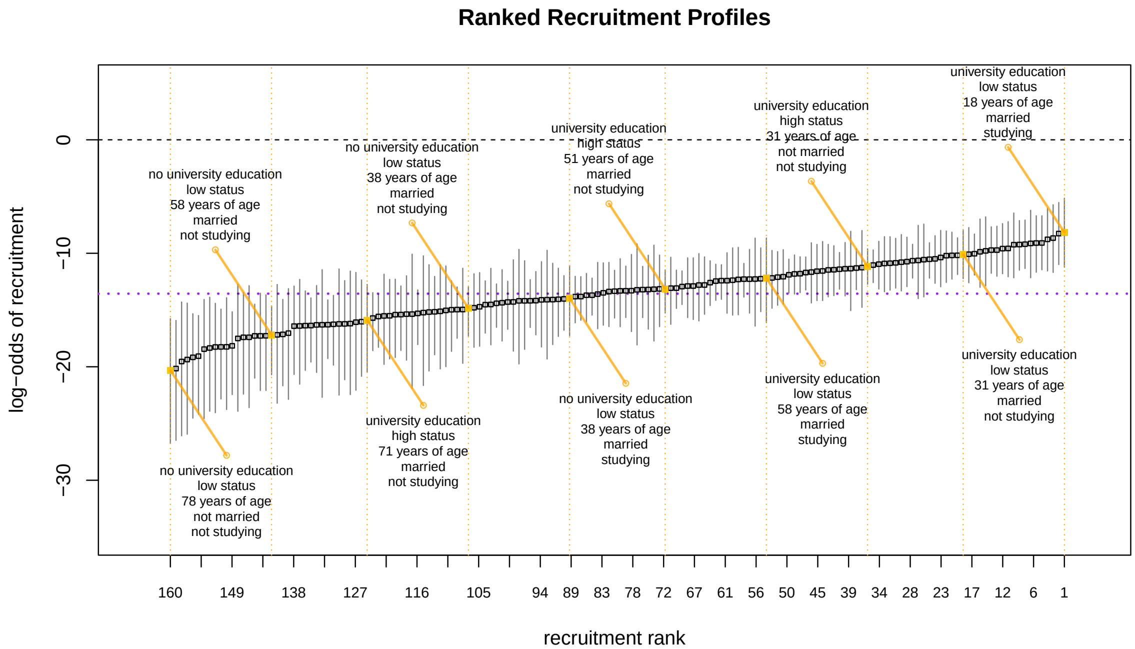


Figure 9: Worm's Eye (Egypt) distribution of the predicted probabilities prediction intervals, presented on the log-odds scale to aid cross-profile comparisons. The black dotted line highlights the the zero-log-odds point, while the purple dotted line notes a central estimate for the median recruitment propensity across profiles.

Rank	Married	Student	University Edu	Low-Status	Age	$\hat{\mathrm{P}}(r=1 \mid X=x)$	Predicted Rate	$\frac{\hat{\mathrm{P}}(r=1 \mid X=x)}{\hat{\mathrm{P}}(r=1 \mid X=\bar{X})}$	$\mathrm{logit}\left(\hat{\mathrm{P}}(r=1 \mid X=x)\right)$
1	1	1	1	1	18	0.000287	3/10000	156.721	-8.155
2	0	1	1	1	18	0.000258	3/10000	141.021	-8.264
3	1	1	1	1	25	0.000173	2/10000	97.486	-8.660
4	0	1	1	1	25	0.000155	2/10000	87.536	-8.772
5	1	0	1	1	18	0.000113	1/10000	62.422	-9.091
6	1	1	1	0	18	0.000112	1/10000	62.529	-9.096
7	1	1	1	1	31	0.000108	1/10000	59.223	-9.130
8	0	0	1	1	18	0.000101	1/10000	56.978	-9.199
9	0	1	1	0	18	0.000098	1/10000	53.336	-9.229
10	0	1	1	1	31	0.000097	1/10000	52.977	-9.244

Table 3: Top 10 recruitable theoretical profiles according to the Egypt 'Worm's Eye' model. Profiles are ordered by predicted probability of recruitment net of sampling protocol. 10 ages are evaluated, starting at 18 (to avoid non-existent profiles) and ending at the largest observed age (78). The last four columns represent respectively: i. the predicted probability of recruitment; ii. the predicted rate of recruitment per $10,000$ people; iii. the predicted odds of recruitment, relative to the 'average' profile; iv. the log-odds of recruitment

4 Conclusion

Extreme forms of political behaviour are rarely ever committed by more than a tiny subsection of any given national population. Despite their small size, these groups often have an outsized influence on state and international politics. *Because* of their small size, extremists are particularly hard to study using conventional statistical methods and research designs.

To address this, we propose that extremism researchers take inspiration from epidemiology and recent applications of case-control methods in political science (Rosenfeld 2018). Here, we propose a new variant of the case-control design that allows us to combine survey techniques with ecological forms of analysis, allowing for meaningful comparisons with the underlying populations from which recruits are drawn. To implement this, we solve a number of statistical problems when explaining rare and extreme forms of political behaviour. In particular, we demonstrate: 1) how best to incorporate area-level random effects when the number of recruits for a given unit is small; 2) how to account for spatial autocorrelation in this setup; 3) how to regularize coefficients to guard against separation. Simulations demonstrate the performance advantage of this new approach over alternatives.

While our analysis focuses on recruitment to ISIS, our hope is that this paper inspires social scientists to apply case-control methods to other instances of extremism where data on recruits and population surveys are available. Examples include participation in the 2021 attack on the Capitol Building in Washington D.C. (Pape 2021), recruitment to far-right movements and white supremacist groups (Klandermans and Nonna 2006; Simi et al. 2017), as well as other examples of violent extremism (della Porta 2013). It is in this spirit that we provide the `extremeR` software package so that extremism researchers working on a range of different cases can easily apply our models (see `http://extremeR.info`).

References

ArabBarometer (2014). Arab barometer wave iii technical report. Technical report, International Development Research Centre, United States Institute for Peace, University of Michigan, Princeton University.

Bail, C. A., F. Merhout, and P. Ding (2018, June). Using Internet search data to examine the relationship between anti-Muslim and pro-ISIS sentiment in U.S. counties. *Science Advances 4*(6), eaao5948.

Baio, G. (2012). *Bayesian methods in health economics.* CRC Press.

Barrie, C. and N. Ketchley (2018). Is protest a safety valve against ISIS in Tunisia?

Bérubé, M., R. Scrivens, V. Venkatesh, and T. Gaudette (2019). Converging Patterns in Pathways in and out of Violent Extremism. *Perspectives on Terrorism 13*(6), 73–89.

Besag, J. and C. Kooperberg (1995). On conditional and intrinsic autoregressions. *Biometrika 82*(4), 733–746.

Besag, J., J. York, and A. Mollié (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics 43*(1), 1–20.

Biggs, M. and S. Knauss (2012, October). Explaining Membership in the British National Party: A Multilevel Analysis of Contact and Threat†. *European Sociological Review 28*(5), 633–646.

Blair, G., C. C. Fair, N. Malhotra, and J. N. Shapiro (2013). Poverty and Support for Militant Politics: Evidence from Pakistan. *American Journal of Political Science 57*(1), 30–48.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell (2017). Stan: A probabilistic programming language. *Journal of statistical software 76*(1).

Cerina, R. and R. Duch (2020). Measuring public opinion via digital footprints. *International Journal of Forecasting 36*(3), 987–1002.

Clark, T. S., D. A. Linzer, et al. (2015). Should i use fixed or random effects. *Political Science Research and Methods 3*(2), 399–408.

Corstange, D. (2009). Sensitive Questions, Truthful Answers? Modeling the List Experiment with LISTIT. *Political Analysis 17*(1), 45–63. Publisher: Cambridge University Press.

della Porta, D. (2013). *Clandestine Political Violence*. Cambridge Studies in Contentious Politics. Cambridge: Cambridge University Press.

Devarajan, S., L. Mottaghi, Q.-T. Do, A. Brockmeyer, C. Joubert, K. Bhatia, and M. Abdel Jelil (2016). Economic and social inclusion to prevent violent extremism. Technical report, World Bank.

Dodwell, B., D. Milton, and D. Rassler (2016, April). The Caliphate's Global Workforce: An Inside Look at the Islamic State's Foreign Fighter Paper Trail. Technical report, Combating Terrorism Center.

Gambetta, D. and S. Hertog (2016). *Engineers of Jihad: The Curious Connection Between Violent Extremism and Education*. Princeton: Princeton University Press.

Gelman, A. and J. Hill (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Gelman, A., A. Jakulin, M. G. Pittau, Y.-S. Su, et al. (2008). A weakly informative default prior distribution for logistic and other regression models. *The annals of applied statistics 2*(4), 1360–1383.

Gelman, A. and D. B. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical science 7*(4), 457–472.

Ghosh, J., Y. Li, R. Mitra, et al. (2018). On the use of cauchy prior distributions for bayesian logistic regression. *Bayesian Analysis 13*(2), 359–383.

Grewal, S., A. Yildirim, and S. Williamson (2020). Counting the Black Flags: Measuring Support for ISIS through Endorsement Experiments. *Working paper*.

Hanretty, C., B. E. Lauderdale, and N. Vivyan (2018). Comparing strategies for estimating constituency opinion from national survey samples. *Political Science Research and Methods 6*(3), 571–591.

Heinze, G. (2017). Logistic regression with rare events: problems and solutions. *Medical University of Vienna: CeMSIIS-Section for Clinical Biometrics. https://www. eio. upc. edu/ca/seminari/docs/georg-heinze-logistics-regression-with-rare-eventsproblems-and-solutions. pdf*.

Heinze, G. and M. Schemper (2002). A solution to the problem of separation in logistic regression. *Statistics in medicine 21*(16), 2409–2419.

Hertog, S., P. Neumann, and S. Maher (2021). Great expectations or nothing to lose? socio-economic correlates of joining the islamic state.

Hoffman, M. D. and A. Gelman (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res. 15*(1), 1593–1623.

Jensen, M. A., A. Atwell Seate, and P. A. James (2020, July). Radicalization to Violence: A Pathway Approach to Studying Extremism. *Terrorism and Political Violence 32*(5), 1067–1090.

Ketchley, N. and M. Biggs (2017). The Educational Contexts of Islamist Activism: Elite Students and Religious Institutions in Egypt. *Mobilization 22*(1), 57–76.

Ketchley, N., S. Brooke, and B. Lia (2021). Who Supported the Early Muslim Brotherhood? *Politics and Religion*, 1–29.

Kim, A. Y. and J. Wakefield (2010). R data and methods for spatial epidemiology: The spatialepi package. *Dept of Statistics, University of Washington*.

King, G. and L. Zeng (2001). Logistic regression in rare events data. *Political analysis 9*(2), 137–163.

Klandermans, B. and M. Nonna (2006). *Extreme Right Activists in Europe: Through the Magnifying Glass*. London: Routledge.

Krueger, A. B. (2017). *What Makes a Terrorist: Economics and the Roots of Terrorism*. Princeton: Princeton University Press.

Krueger, A. B. and J. Maleckova (2003). Education, Poverty and Terrorism: Is There a Causal Connection? *Journal of Economic Perspectives 17*(4), 119–144.

Lancaster, T. and G. Imbens (1996). Case-control studies with contaminated controls. *Journal of Econometrics 71*(1-2), 145–160.

Lauderdale, B. E., D. Bailey, J. Blumenau, and D. Rivers (2020). Model-based pre-election polling for national and sub-national outcomes in the us and uk. *International Journal of Forecasting 36*(2), 399–413.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter (2000). Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing 10*(4), 325–337.

Mesquita, E. B. D. (2005). The Quality of Terror. *American Journal of Political Science 49*(3), 515–530.

Mitts, T. (2019, February). From Isolation to Radicalization: Anti-Muslim Hostility and Support for ISIS in the West. *American Political Science Review 113*(1), 173–194.

Morris, A. M. (2020, June). Who Wants to Be a Suicide Bomber? Evidence from Islamic State Recruits. *International Studies Quarterly 64*(2), 306–315.

Morris, M., K. Wheeler-Martin, D. Simpson, S. J. Mooney, A. Gelman, and C. DiMaggio (2019). Bayesian hierarchical spatial models: Implementing the besag york mollié model in stan. *Spatial and spatio-temporal epidemiology 31*, 100301.

Pape, R. A. (2021). Opinion | What an analysis of 377 Americans arrested or charged in the Capitol insurrection tells us. *Washington Post*.

Park, D. K., A. Gelman, and J. Bafumi (2004). Bayesian multilevel estimation with poststratification: State-level estimates from national polls. *Political Analysis 12*(4), 375–385.

Piironen, J. and A. Vehtari (2017). On the hyperprior choice for the global shrinkage parameter in the horseshoe prior. In *Artificial Intelligence and Statistics*, pp. 905–913. PMLR.

Plummer, M. et al. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, Volume 124, pp. 1–10. Vienna, Austria.

Riebler, A., S. H. Sørbye, D. Simpson, and H. Rue (2016). An intuitive bayesian spatial model for disease mapping that accounts for scaling. *Statistical methods in medical research 25*(4), 1145–1165.

Robinson, W. S. (1950). Ecological Correlations and the Behavior of Individuals. *American Sociological Review 15*(3), 351–357.

Rosenblatt, N. (2018). All Jihad is Local: What ISIS' Files Tell Us About Its Fighters. Technical Report I, New America.

Rosenfeld, B. (2017). Reevaluating the middle-class protest paradigm: A case-control study of democratic protest coalitions in russia. *American Political Science Review 111*(4), 637–652.

Rosenfeld, B. (2018). A case-control method for studying protest participation and other rare events: Application to ukraine's euromaidan. *Paper presented at the 2018 Annual Conference of the American Political Science Association*, 1–41.

Rota, C. T., J. J. Millspaugh, D. C. Kesler, C. P. Lehman, M. A. Rumble, and C. M. Jachowski (2013). A re-evaluation of a case–control model with contaminated controls for resource selection studies. *Journal of animal ecology 82*(6), 1165–1173.

Selb, P. and S. Munzert (2011). Estimating constituency preferences from sparse survey data using auxiliary geographic information. *Political Analysis*, 455–470.

Simi, P., K. Blee, M. DeMichele, and S. Windisch (2017, December). Addicted to Hate: Identity Residual among Former White Supremacists. *American Sociological Review 82*(6), 1167–1187.

Simpson, D., H. Rue, A. Riebler, T. G. Martins, S. H. Sørbye, et al. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical science 32*(1), 1–28.

Skare, E. (2022). Affluent and well-educated? analyzing the socioeconomic backgrounds of fallen palestinian islamist militants. *The Middle East Journal 76*(1), 72–92.

Sterman, D. and N. Rosenblatt (2018). All Jihad is Local: ISIS in North Africa and the Arabian Peninsula. Technical Report II, New America.

Vats, D. and C. Knudson (2021). Revisiting the gelman–rubin diagnostic. *Statistical Science 36*(4), 518–529.

Vehtari, A., A. Gelman, D. Simpson, B. Carpenter, and P.-C. Bürkner (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of mcmc (with discussion). *Bayesian analysis 16*(2), 667–718.

Wang, W., D. Rothschild, S. Goel, and A. Gelman (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting 31*(3), 980–991.

Zelin, A. Y. (2018). Tunisian foreign fighters in iraq and syria. Technical report, Washington Institute for Near East Policy.

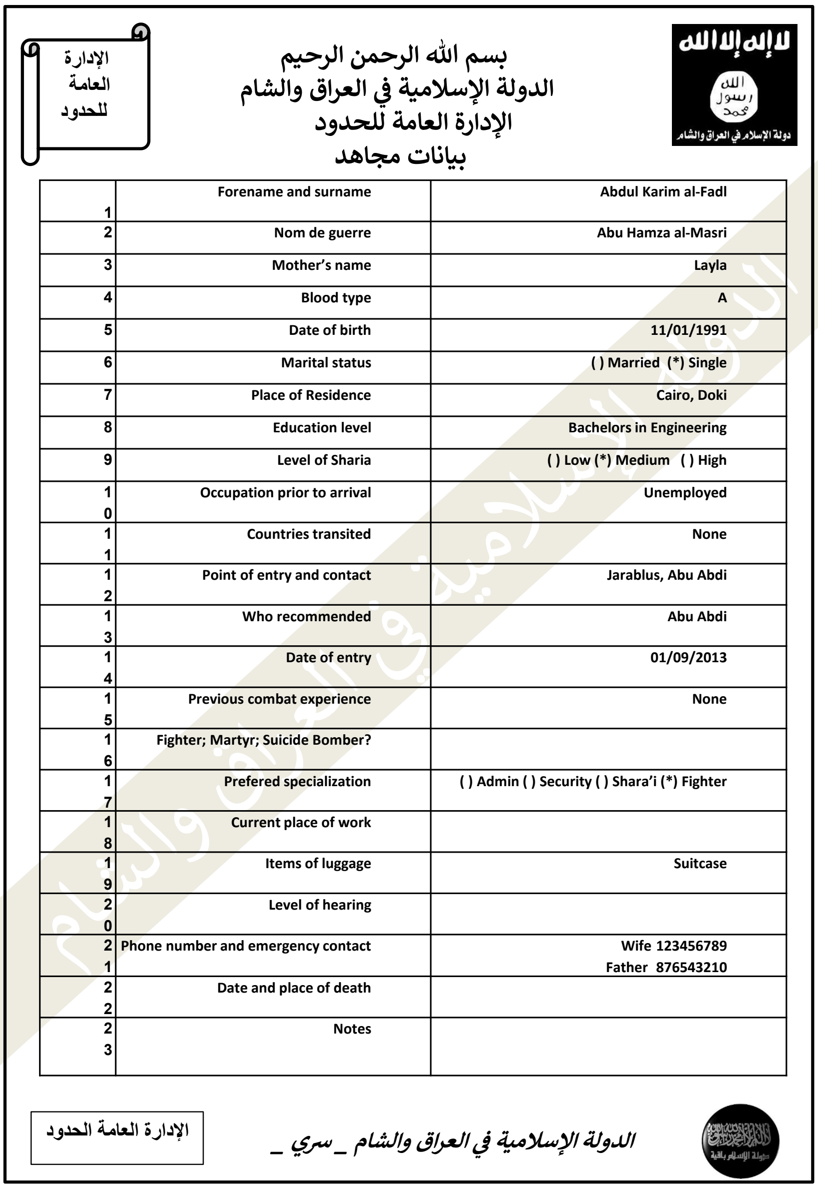
Supplementary Materials

Table of Contents

A Sources

It is important to consider the data-generating process for the leaked ISIS border documents. These documents contain detailed information on the home residence of each recruit, age, education, marital status, previous employment, employment status, previous combat experience, and date of entry into ISIS-controlled territory. They derive from a set of leaked documents recording the details of fighters who have crossed into ISIS-controlled territory with the intention of becoming a recruit.

Supplementary Figure A.1 provides an imitation of one of the border documents. We use data for nine countries in the MENA that were included in the leak. These are: Algeria, Egypt, Jordan, Kuwait, Lebanon, Libya, Morocco, Tunisia, and Yemen. In total, we have complete records for 1,051 recruits. It remains unclear whether these constitute a representative sample of recruits. Dodwell et al. (2016) demonstrate, however, that 98% of these individuals can be matched against records for ISIS recruits held by the U.S. Department of Defense. Further, the Bayesian case-control approach we detail below takes into account the non-probability nature of the data-generating process through its multilevel design.



الإدارة العامة للحدود

بسم الله الرحمن الرحيم
الدولة الإسلامية في العراق والشام
الإدارة العامة للحدود
بيانات مجاهد

1	Forename and surname	Abdul Karim al-Fadl
2	Nom de guerre	Abu Hamza al-Masri
3	Mother's name	Layla
4	Blood type	A
5	Date of birth	11/01/1991
6	Marital status	( ) Married (*) Single
7	Place of Residence	Cairo, Doki
8	Education level	Bachelors in Engineering
9	Level of Sharia	( ) Low (*) Medium ( ) High
10	Occupation prior to arrival	Unemployed
11	Countries transited	None
12	Point of entry and contact	Jarablus, Abu Abdi
13	Who recommended	Abu Abdi
14	Date of entry	01/09/2013
15	Previous combat experience	None
16	Fighter; Martyr; Suicide Bomber?	
17	Prefered specialization	( ) Admin ( ) Security ( ) Shara'i (*) Fighter
18	Current place of work	
19	Items of luggage	Suitcase
20	Level of hearing	
21	Phone number and emergency contact	Wife 123456789 Father 876543210
22	Date and place of death	
23	Notes	

الإدارة العامة الحدود

الدولة الإسلامية في العراق والشام _ سري _

Figure A.1: Example of border document (details changed)

A.1 Independent variable details

The Arab Barometer surveys were in the field at different times for each country: December, 2012-January, 2013 for Jordan; February, 2013 for Tunisia; March-April, 2013 for Egypt and Algeria; April-June, 2013 in Morocco; July 2013 in Lebanon; November-December 2013 in Yemen; and February-March 2014 in Kuwait (ArabBarometer 2014).

For both Egypt and Tunisia, we also include variables to capture subnational differences in demographic and labor-market composition, employment opportunities, as well as more context-specific variables designed to capture support for Islamist political organizations and prehistories of contentious politics. Our choice of contextual variables is based on existing research finding that lack of employment opportunities, prehistories of mobilization and repression, as well as support for political Islam, are predictive of ISIS recruitment (Devarajan et al. 2016; Rosenblatt 2018; Grewal et al. 2020; Barrie and Ketchley 2018).

Table A.1: Individual-level variable codings across border documents and survey data

Variable	Border Documents	ABIII
coledu	1 if Education level mentions "university"	1 if q1003 >5 (or >4 for Tunisia; >6 for Yemen)
age	Date of entry - Date of birth	q1001
married	1 if Marital status is "married"	q1010
student	1 if Occupation prior to arrival is "student"	q1004 = 3 (Student)
lowstat	1 if Occupation prior to arrival is agricultural or manual/unemployed	q1004 = 5 (Unemployed) or q1010 = 4/5 (Agricultural or manual worker)

Table A.2: Egypt district-level covariates

Variable	Details	Source
Population density	number of individuals in district/district area in kmˆ2	2006 Census
Population	number of individuals in district aged 10 or over	2006 Census
% Christian	percentage of individuals in district recorded as Christian	2006 Census
% University	percentage of individuals in district who are university educated	2006 Census
% Agriculture	percentage individuals employed in agriculture denominated by total active population	2006 Census
% Mursi	percent of total votes in district for Muhammad Mursi in the first round of the 2012 presidential election	El-Masry and Ketchley (2021)
Unemployment rate	number individuals aged without employment denominated by total active population	2006 Census
Killed at Rabaa	number of deaths of individuals from district at the 2013 Rabaa Massacre (square-rooted)	Ketchley and Biggs (2017)
Post-revolutionary protest	number of protests recorded in district in 12 months after Jan 25 Revolution (square-rooted)	Barrie and Ketchley (2019)

Table A.3: Tunisia district-level covariates

Variable	Details	Source
Population	number of individuals in district aged 10 or over	2014 Census
Population density	number of individuals in district aged 10 or over/district area in km^2	2014 Census
% University	percentage population with higher education certificate denominated by total population	2014 Census
% Agriculture	percentage individuals employed in agriculture denominated by total active population aged 15 or over	2014 Census
Unemployment rate	number individuals aged 18-59 without employment denominated by total active population aged 18-59	2014 Census
Graduate unemployment rate	number individuals with higher education certificate without employment denominated by total active population aged 18-59	2014 Census
% Ennahda 2011	percentage of total votes in district for Ennahdha in 2011 election	INS Tunisia
% Ennahdha 2014	percentage of total votes in district for Ennahdha in 2014 election	INS Tunisia
Post-revolutionary protests	number of protests recorded in district in 12 months after Jan 14 Revolution (square-rooted)	Barrie and Ketchley (2019)
Distance to Libya	distance to Libyan border from centroid of target district (square-rooted)	NA

B MCMC Convergence

Convergence diagnostics provide a first measure of the reliability of our parameter estimates for both the Bird's Eye and Worm's Eye models. Here, we follow Vehtari et al (2021) and implement multiple state-of-the-art tests.

Per Supplementary Figures F.1, F.5, and F.7, we examine four versions of the Gelman-Rubin statistic ($\hat{R}$) to verify convergence is obtained broadly, as well as when we encounter heteroskedasticity across chains, or when these are heavy-tailed. There exist various convergence-thresholds in the literature – the most stringent requires $\hat{R} < 1.01$, a medium-stringency threshold suggests $\hat{R} < 1.05$ (especially if we are estimating a large number of parameters), whilst the historical recommendation was $\hat{R} < 1.1$ (Gelman and Rubin 1992). Recent work demonstrates that this latter threshold prematurely diagnoses convergence in most cases (Vats and Knudson 2021). The parameters of all of our models are broadly convergent under the harshest 1.01 threshold for all of the measures of $\hat{R}$, with the exception of a very small number of spatial effects which are convergent under a slightly more laxed threshold, though still well below the 'premature convergence' threshold $\hat{R} < 1.1$. [8]

Supplementary Figures F.2, F.6, and F.8 present five measures of Effective Sample Size (ESS), which tell us about the true number of independent draws from the joint posterior distribution after accounting for auto-correlation within chains. The measures check that the independent sample is 'large enough' to ensure stability of summaries of the distribution at various moments (e.g. overall, at the median, at the tails, etc.). Mirroring the performance of the $\hat{R}$, the posterior samples for most of our estimates parameters are well above the recommended threshold ($ESS > 400$) for ensuring stability of the central and tail estimates.

We further explore convergence at different quantiles of the posterior distribution of our least-convergent parameters – those with the lowest bulk and tail ESS (Figure F.3 presents these measures for the Bird's Eye model). The inference is that if these relatively low-ESS parameters showcase satisfactory ESS at every quantile, we can be reassured that the whole model has converged. These plots suggest broad reliability of estimates at every section of the distribution. Finally, we explore the mixing properties of our chains for these least-convergent parameters (Figure F.4).[9] These plots broadly suggest good mixing properties of our model, even for these relatively inefficient posterior samples.

[8]Note that in the Bird's Eye model, this struggle is slightly exacerbated by the inclusion of governorates from Israel and Saud Arabia, for which we have no observations, and whose effects are fully interpolated via the spatial process.

[9]Here, we choose to include mixing diagnostics for the Bird's Eye model.

C Bayesian modeling in Stan

The model that we propose is amenable to Bayesian estimation via Monte Carlo Markov Chain (MCMC) methods. Previous contributions to the case-control literature [e.g., Rota et al. 2013; Rosenfeld 2017] have used WinBUGS (Lunn et al. 2000) or JAGS (Plummer et al. 2003) as software to implement some variations on a simple Gibbs sampler. Due to the heavy computational burden imposed by the spatial prior, we propose instead to innovate by estimating this model in Stan (Carpenter et al. 2017). Stan leverages a version of Hamiltonian Monte-Carlo (HMC) called the 'No U-Turn Sampler' (NUTS) (Hoffman and Gelman 2014), which dramatically improves the efficiency and speed of convergence of our Markov-Chains. A challenge we face is that Stan cannot handle the sampling of latent discrete parameters (r_i in our hierarchical model above), posing a problem for the estimation of mixture models. The state-of-the-art solution is to marginalize the latent parameter out. In practice this means replacing our model for the observed labels y with the following mixture of Bernoulli distributions:

$$f(y_i \mid \rho_i) = \rho_i \text{ Bernoulli}(y_i \mid \theta_1) + (1 - \rho_i) \text{ Bernoulli}(y_i \mid \theta_0). \tag{39}$$

Beyond allowing for model parameters to be informed by y_i according to the mixed structure above, marginalization provides significant advantages for posterior exploration and MCMC efficiency as it leverages expectations rather than sampling of discrete parameters. Listing 1 in the Supplementary Materials presents the Stan code for our final model. Note that fixed-effects covariates are standardized.[10] Estimates of regression coefficients on the original, unstandardised scale are computed and available in these Supplementary Materials.

C.1 Stan listings

Listing 1: Stan Data Declaration Block.

```
data{

     int<lower = 1> n;          // total number of observations
     int<lower = 1> p;          // number of covariates in design matrix
     int<lower = 0> y[n];          // vector of labels
     matrix[n, p] X;          // design matrix

     int<lower = 1> small_area_id[n];          // small-area id
     int<lower = 1> N_small_area;          // number of small areas

     int<lower = 1> N_small_area_edges;          // number of edges in the spatial process
     int<lower=1, upper=N_small_area> node1_small_area[N_small_area_edges]; // node1[i] adjacent to node2[i]
     int<lower=1, upper=N_small_area> node2_small_area[N_small_area_edges]; // node1[i] adjacent to node2[i]

     real scaling_factor;          // scaling factor derived from the adjacency matrix

     int<lower = 1> large_area_id[n];          // large-area ids
     int<lower = 1> N_large_area;          // number of large-areas

     vector[N_large_area] log_offset;          // log-scale offset

     matrix[2,N_large_area] theta;          // Pr(Y = 1 | r = 1, s = 1)

     }
```

[10] We standardize both dichotomous and continuous variables as this aids convergence.

Listing 2: `Stan` Parameters Declaration Block.

```
parameters{

           // cauchy prior for individual-level coefficients expressed as scale mixture of gaussian density
                functions
           vector[p] aux_a;         // central component
           vector<lower = 0>[p] aux_b;         // scale component


           vector[N_small_area] phi;         // small-area unstructured effects
           vector[N_small_area] psi;         // small-area spatial effect

           real<lower = 0,upper = 1> lambda;         // mixing prior on spatial component

           real<lower = 0> sigma_gamma;         // small-area effect scale

           vector[N_large_area] eta;         // large-area unstructured effect

           real<lower = 0> sigma_eta;         // large-area effect scale

           }
```

Listing 3: `Stan` Transformed Parameters Block.

```
transformed parameters{

                    vector[p] beta = aux_a ./ sqrt(aux_b);        // individual-effect prior

                    vector[n] mu;        // expected propensity of recruitment

                    vector[N_small_area] gamma = (sqrt(1-lambda) * phi + sqrt(lambda / scaling_factor) * psi)*
                        sigma_gamma;
                    // convolved small -area effect

                    mu = log_offset[large_area_id] + eta[large_area_id]*sigma_eta + gamma[small_area_id] + X *
                        beta;
                    // linear function of the logit -scale propensity to be a recruit

                    }
```

Listing 4: `Stan` Model Declaration Block.

```
model{

      aux_a ~ normal(0,1);            // prior on the centrality of the cauchy prior
      aux_b[1] ~ gamma(0.5,100*0.5);            // prior on intercept-scale
      aux_b ~ gamma(0.5,0.5);            // prior on individual covariate scales

      target += -0.5 * dot_self(psi[node1_small_area] - psi[node2_small_area]);
      // ICAR prior

      phi ~ normal(0,1);            // unstructured random effect on small -area
      sum(psi) ~ normal(0, 0.01 * N_small_area);
      // soft sum -to-zero , equivalent to mean(psi) ~ normal (0 ,0.01)

      lambda ~ beta(0.5,0.5);            // mixing weight prior

      sigma_gamma ~ normal(0,1);            // prior small-area scale

      eta ~ normal(0,1);            // prior large-area effect

      sigma_eta ~ normal(0,1);            // prior large-area scale

      // likelihood
      for (i in 1:n) {
            target += log_mix(1-inv_logit(mu[i]),
                              bernoulli_lpmf(y[i] | theta[1,large_area_id[i]]),
                              bernoulli_lpmf(y[i] | theta[2,large_area_id[i]]));
                              // labels distributed as mixture of bernoulli distributions
                  }
      }
```

D Simulation study

The model that we propose extremism researchers should adopt is significantly more complex than the standard case-control design using rare-events logistic regression and requires a substantial understanding of Bayesian methods to be fully appreciated. Moreover, the model's estimation becomes roughly exponentially more computationally challenging as the sample size increases. To provide evidence that our approach is nevertheless preferable to a more straightforward case-control design, we report the results of a comprehensive simulation study that compares the performance of our model against the King and Zeng model (2001), as well as a simple fixed-effects logistic regression. We score these models according to their ability to accurately predict the underlying latent propensity of recruitment, $\mu_i = \text{logit}(\rho_i)$. We further investigate these models' performance in accurately estimating the intercept, regression coefficients, and residual area effects.

We note that we did not test the 'coverage' properties of our models' estimates as part of the simulation study. To test coverage, we would have needed to run the chains of each of our simulations long enough for the parameters of our models to converge in their second-moment - this was not feasible under a simulation framework where we had to run the model 200 times. As pointed out in our discussion of model fitting strategy, the well-behaved models we use to derive our results took up to 48 hours to achieve posterior samples displaying satisfactory convergence. We therefore leave it for future work to formally quantify the coverage of or models.

Simulations show that our model is robust and general. The results suggest that in a rare-event scenario, our model outperforms King and Zeng's rare-events logistic regression thanks to its ability to account for spatial auto-correlation, while also remaining largely unbiased to discrepancies in sample and population prevalence. As prevalence increases, our model retains a degree of robustness that neither a simple fixed-effects logistic regression, nor the 'rare events logit', can offer – largely thanks to the contamination layer. This robustness extends not just to the ability to correctly estimate latent propensity $\boldsymbol{\mu}^{\star}$, but actively reduces bias and RMSE in the estimation of coefficients.

D.1 Data Generating Function

In what follows we present a more detailed view of the setup and results of the simulation study. First, we create a data-generating function to draw sample-datasets generated according to the mechanism implied by either the rare-events or contaminated case-control model. We reduce the data generating process to its essence for simplicity: a single continuous covariate x_i is considered, and large-area effects are dropped. Small-area effects are simulated according to a random intrinsic conditionally auto-regressive process from one of three widely-used maps.[11], available from the `R` package `SpatialEpi` (Kim and Wakefield 2010). This enables the random sampling of ICAR effects whilst

[11] $\mathcal{M} = \{$`scotland_lipcancer`, `newyork_lukemia`, `pennsylvania_lungcancer`$\}$

preserving a plausible geography (i.e. neighbourhood structure and distance between areal units). Pseudo-algorithm 1 describes the steps taken to generate the simulated data.

Algorithm 1 A pseudo algorithm displaying the steps taken by the data generating function to generate a random sample of data.

Require:

sample size:	$n \in [100, 2000]$
population prevalence:	$\pi \in \left[\frac{1}{1000000}, \frac{1}{2}\right]$
expected sample prevalence:	$\hat{\pi} \in [0.01, 0.99]$
global auto-correlation:	$I \in (0, 1)$
map:	$\mathcal{M} \in \{\texttt{scotland}, \texttt{newyork}, \texttt{pennsylvania}\}$

(0.) derive key quantities directly from inputs:

i. expected number of case-labelled records:
$n_1 \leftarrow n \times \hat{\pi}$

ii. expected number of unlabelled records:
$n_u \leftarrow n - n_1$

iii. relative prob. of sampling a case v. control:
$\frac{P_1}{P_0} \leftarrow \frac{(n_1 + \pi \times n_u)/\pi}{n_u}$

iv. prob. of sampling a case-labelled record conditional on being a true control:
$\theta_0 \leftarrow 0$

v. prob. of sampling a case-labelled record conditional on being a true case:
$\theta_1 \leftarrow \frac{n_1}{n_1 + \pi \times n_u}$

(1.) sample area effects on selected map: $\boldsymbol{\gamma} \sim \text{ICAR}(\mathcal{M})$
(2.) sample initial value for intercept: $\beta_1 \sim N(0, 1)$
(3.) sample covariate value: $x_i \sim N(0, 1)$
(4.) sample covariate effect: $\beta_2 \sim N(0, 1)$
(5.) optimise intercept to meet specified sample prevalence:
$\beta_1^\star \leftarrow \arg\max_{\beta_1} f(\hat{\pi}; \beta_1)$
(6.) calculate latent recruitment propensity: $\boldsymbol{\mu} \leftarrow \log\left(\frac{P1}{P0}\right) + \beta_1^\star + \boldsymbol{x}\beta_2 + \boldsymbol{\gamma}$
(7.) calculate recruitment propensity: $\boldsymbol{\rho} \leftarrow \texttt{inv\_logit}(\boldsymbol{\mu})$
(8.) sample recruitment status: $\boldsymbol{r} \sim \text{Bernoulli}(\boldsymbol{\rho})$
(9.) sample labels: $\boldsymbol{y} \sim \text{Bernoulli}(\theta_{\boldsymbol{r}})$

We simulate `n.sims` = 200 datasets[12] using the data-generating function. The inputs to the function (highlighted under the 'Required' header in the pseudo-code) are

[12] In practice we simulate datasets in two stages: first we examine the model performance by sampling 100 draws from a 'rare event' process ($\pi \in [1/1000000, 1/10]$); then, in a second stage, we sample another 100 draws from a process with less extreme prevalence ($\pi \in [1/10, 1/2]$). This is done to evaluate performance in two different scenarios – extreme (rare-event) v. non-extreme – and ensure a large-enough sample size to capture salient dynamics in both.

sampled at random from uniform distributions conforming to the specified range for each input - in the case of the maps, a map is chosen at random amongst the three candidates.

D.2 Candidate Models

A second step is to define the models analysed in this simulation study. The candidate models are: **m.1** – a simple fixed-effects logistic regression, where the area-effects are also estimated via fixed-effects; **m.2** – similar to **m.1**, but importantly augmented with the use on an offset (prior-correction) a-la (King and Zeng 2001); **m.3** – an essential version of our rare-events, Bayesian contaminated-controls model with a BYM2 area-effects prior. The models are detailed in Figure D.1.

(m.1): Fixed-effects logit

$$y_i \sim \text{Bernoulli}(\rho_i); \quad (40)$$
$$\text{logit}(\rho_i) = \beta_1 + x_i\beta_2 + \sum_l z_{i,l}\gamma_l; \quad (41)$$
$$\beta_1 \sim N(0, 10); \quad (42)$$
$$\beta_2 \sim N(0, 1); \quad (43)$$
$$\gamma_l \sim N(0, 1). \quad (44)$$

(m.2): King & Zeng

$$y_i \sim \text{Bernoulli}(\rho_i); \quad (45)$$
$$\text{logit}(\rho_i) = \log\left[\left(\frac{1-\pi}{\pi}\right)\left(\frac{\bar{y}}{1-\bar{y}}\right)\right] + \quad (46)$$
$$+ \beta_1 + x_i\beta_2 + \sum_l z_{i,l}\gamma_l; \quad (47)$$
$$\beta_1 \sim N(0, 10); \quad (48)$$
$$\beta_2 \sim N(0, 1); \quad (49)$$
$$\gamma_l \sim N(0, 1). \quad (50)$$

(m.3): Cerina et al.

$$y_i \sim \text{Bernoulli}(\theta_{r_i}); \quad (51)$$
$$r_i \sim \text{Bernoulli}(\rho_i); \quad (52)$$
$$\text{logit}(\rho_i) = \log\left(\frac{1}{\pi n_u} + 1\right) + \quad (53)$$
$$+ \beta_1 + x_i\beta_2 + \gamma_l; \quad (54)$$
$$\beta_1 \sim \text{Cauchy}(0, 10); \quad (55)$$
$$\beta_2 \sim \text{Cauchy}(0, 1); \quad (56)$$
$$\gamma_l = \sigma\left(\phi_l\sqrt{(1-\lambda)} + \psi_l\sqrt{(\lambda/s)}\right); \quad (57)$$
$$\lambda \sim \text{Beta}(0.5, 0.5); \quad (58)$$
$$\phi_l \sim N(0, 1); \quad (59)$$
$$\psi_l \mid \psi_{l'} \sim N\left(\frac{\sum_{l' \neq l}\psi_{l'}}{d_{l,l}}, \frac{1}{\sqrt{d_{l,l}}}\right) \quad (60)$$
$$\sigma \sim \frac{1}{2}N(0, 1). \quad (61)$$

Figure D.1: Hierarchical formulation of the three competing models considered in the simulation study.

D.3 Computational Constraints

In order to fit the 600 models necessary for this simulation study, we have to 'live dangerously'[13], and lower our expectations over the stringent convergence properties of any given model. What we are interested in is the stability of the simulation results, and this is an aggregate set of quantities which is relatively robust to the semi-convergence of any given model. We therefore run each model in `Stan`, with the following settings: `n.cores` = 4; `n.chains` = 4; `n.thin` = 4; `n.iter` = 4000; `n.warmup` = $\frac{2}{3}$`n.iter`;

[13] As others have done before when frequently fitting complex `Stan` models on large datasets – see (Lauderdale et al. 2020) for an example running multiple short chains.

all other settings are set to the `Stan` default. This gives us posterior samples which have relatively small effective sample-sizes, but are nevertheless able to give us reliable central-estimates for the parameters of interest – proof of this is that the results from the simulation study are replicable over multiple samples. Note that very rarely the chains will diverge for m.3 under high-levels of contamination. When this happens, we drop these simulations from the analysis and re-run the model.

D.4 Comparison Metrics

Finally, we define the parameters of the comparison. The simulations are intended to investigate the ability of competing models to estimate the following quantities of interest: $\boldsymbol{\mu}^{\star} = \boldsymbol{\mu} - \log\left(\frac{P_1}{P_0}\right)$, the latent propensity to be a recruit; β_1, the baseline propensity to be a recruit; β_2, the effect of simulated covariate $\boldsymbol{x}$; $\boldsymbol{\gamma}$, the set of area-level effects which contribute to the latent propensity. $\boldsymbol{\mu}^{\star}$ is a good summary metric of performance on all of these dimensions, so our primary inference refers to this quantity. The models are scored only on their point-estimates, as an evaluation of uncertainty is computationally unfeasible due to the large number of MCMC iterations necessary to obtain convergent estimates of the second-moment for all these parameters.

The models are generating parameter estimates $\hat{f}$ to approximate the true simulated parameters f; they are scored on three dimensions: i. bias$(\hat{f}) = \frac{1}{n}\sum_i \hat{f}_i - f_i$; ii. Root-mean-square-error RMSE$(\hat{f}) = \frac{1}{n}\sum_i(\hat{f}_i - f_i)^2$; iii. Pearson correlation coefficient $r(\hat{f}) = \frac{\sum_i(\hat{f}_i - \bar{\hat{f}}_i)(f_i - \bar{f}_i)}{\sqrt{\sum_i(\hat{f}_i - \bar{\hat{f}}_i)^2 \sum_i(f_i - \bar{f}_i)^2}}$. The bias tells us the average direction of the estimation error; the RMSE tells us about the average magnitude of the error, penalising large deviations more heavily than smaller-ones; the Pearson correlation tells us about the ability of the model to correctly order (rank) the parameters.

D.5 Results

Figure D.2 presents a comparison of m.2 and m.3 in their ability to estimate latent propensity $\boldsymbol{\mu}^{\star}$, for each scoring function (on the y-axis) across key characteristics of the data (on the x-axis). A comparison including m.1 is initially omitted here as the scale of the errors in m.1 is so large that it makes it visually impossible to distinguish between the (otherwise substantial) differences in m.2 and m.3 performance. A complete plot including m.1 is available below.

A visual analysis of Figure D.2 presents two clear dimensions in which our model advances the literature: i) m.3 is superior at moderate levels of prevalence ($\pi > 0.1$), a feat obtained thanks to the contamination layer of the model; ii) m.3 is superior under moderate-to-high levels of spatial auto-correlation ($I > 0.2$), due to the BYM2 spatial component. Related to the first advantage, we note that as the discrepancy between population and sample prevalence becomes positive ($\pi - \hat{\pi} > 0$), the upward bias

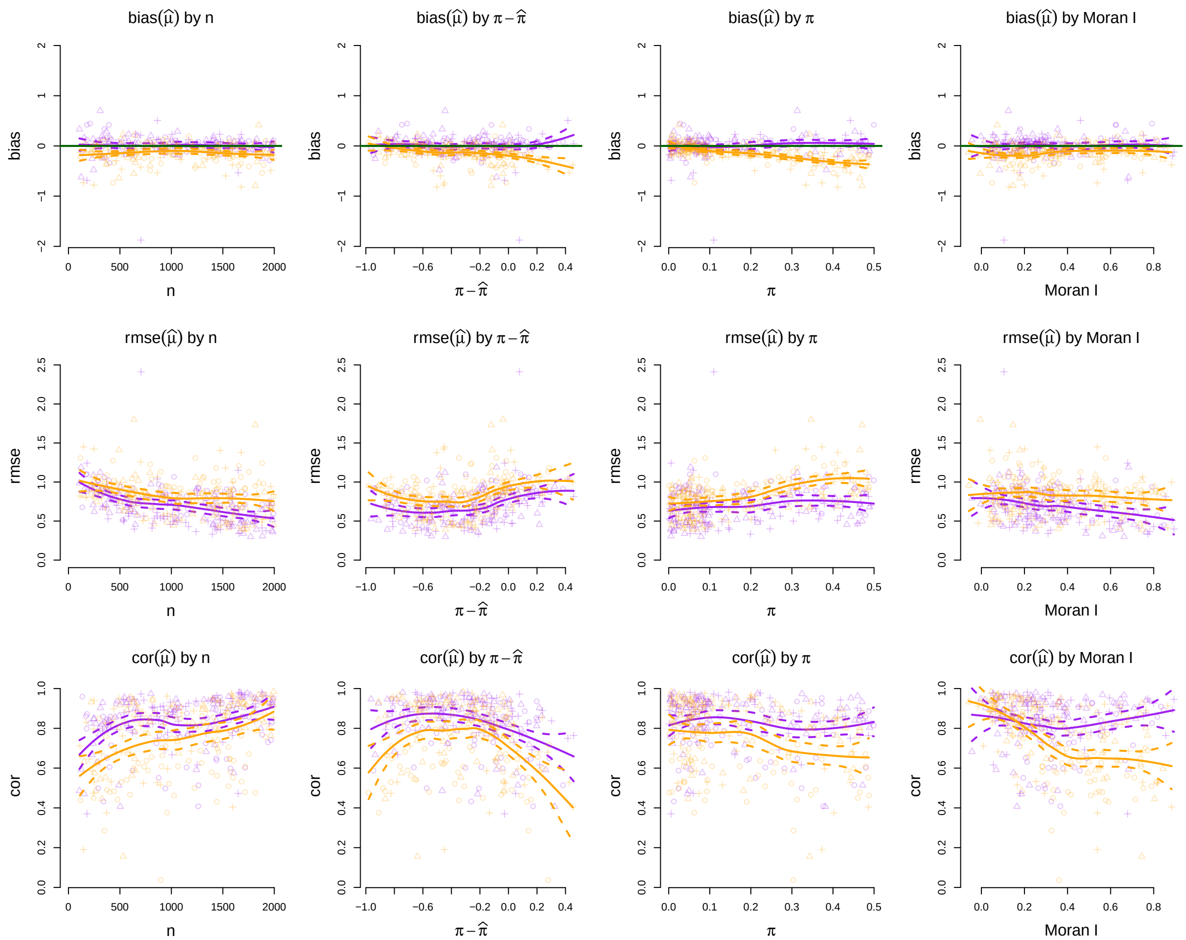


Figure D.2: Results of the simulation study, comparing the performance of our model (m.3, in purple) and the more traditional rare-events logistic regression with prior-correction for the intercept (King and Zeng 2001) (m.2, in orange) in estmating the true latent propensity $\mu^{\star} = \mu - \log(\frac{P_1}{P_0})$.

which m.3 suffers from as a result of contamination is significantly more contained than the downward bias which characterises m.2 as a result of a non-contaminated offset, again highlighting another robustness advantage, pertaining to the relationship between sample and population prevalence. Moreover, Figures D.5 and D.6, which present the ability of m.2 and m.3 to estimate respectively the correct intercept parameter β_1 and the covariate effect β_2, also paint a favourable picture. The ability of our model to perform under high levels of prevalence affords significant reductions in bias an RMSE, in both β_1 and β_2, already at moderate levels of contamination. Figure D.7 compares models in their ability to estimate the correct area-level effect. Though all three models are, unsurprisingly, unbiased, m.3 is clearly more precise (lower RMSE) and and better at ordering areas according to their propensity (higher Pearson correlation), in the presence of spatial auto-correlation.

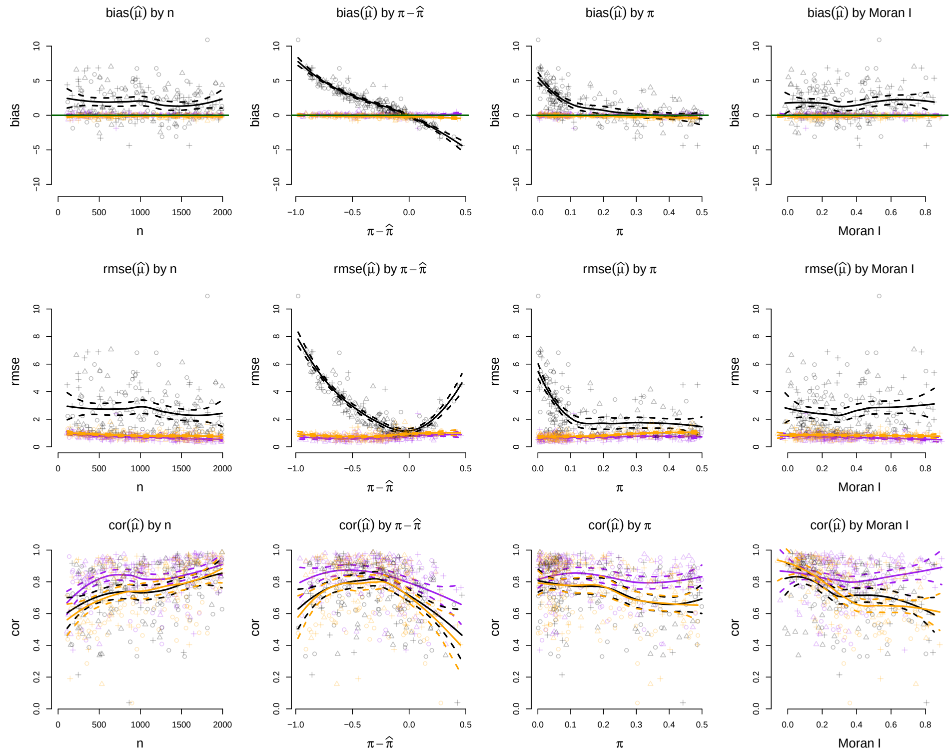


Figure D.3: Results from the simulation study, capturing the ability of the simple fixed effects model (m.1, in black), the King & Zeng model (m.2, in orange) and our proposed approach (m.3, in purple) to estimate the latent propensity of recruitment for each record in our sample $\boldsymbol{\mu}^{\star}$.

Figure D.3 presents the scoring of models in their ability to predict latent propensity $\boldsymbol{\mu}$; Figure D.4 displays the models' performance in estimating the baseline propensity β_1, with Figure D.5 zooming-in to a comparison between out proposed model and the rare-events logit by King & Zeng; Figure D.4 shows model performance in estimating covariate effect β_2; Figure D.7 presents a comparison with respect to the estimation of area-level effects $\boldsymbol{\gamma}$.

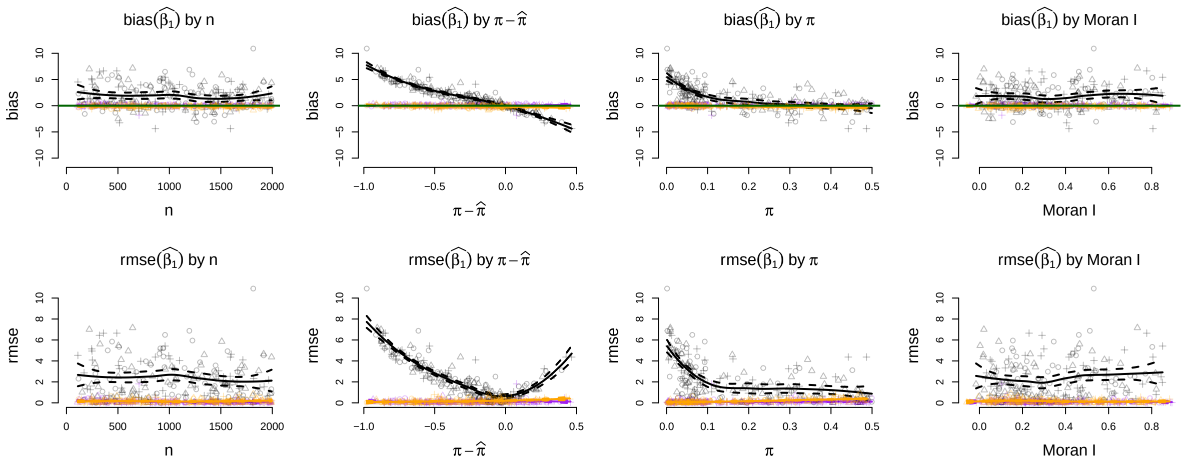


Figure D.4: Results from the simulation study, capturing the ability of the simple fixed effects model (m.1, in black), the King & Zeng model (m.2, in orange) and our proposed approach (m.3, in purple) to estimate the true intercept β_1.

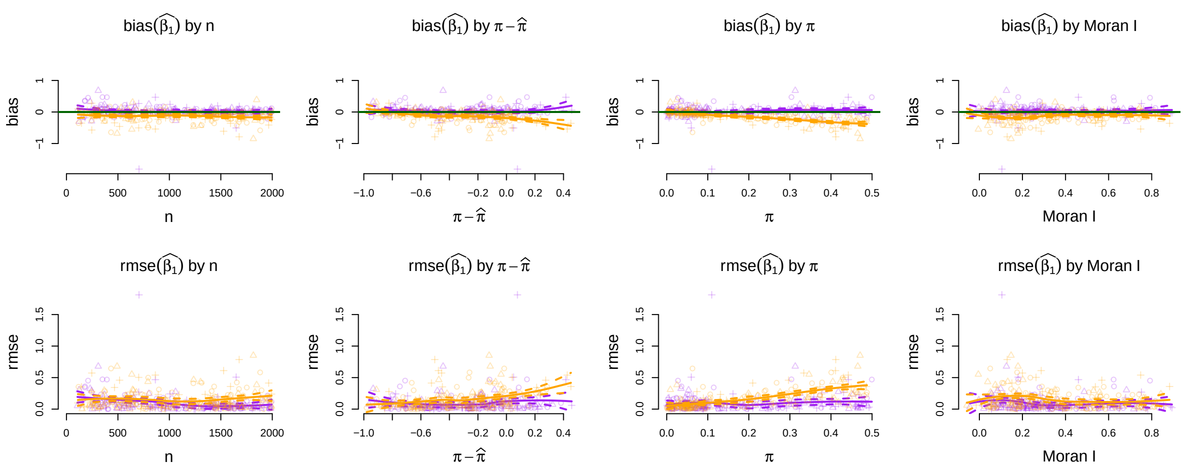


Figure D.5: Results from the simulation study, capturing the ability of the King & Zeng model (m.2, in orange) and our proposed approach (m.3, in purple) to estimate the true intercept β_1.

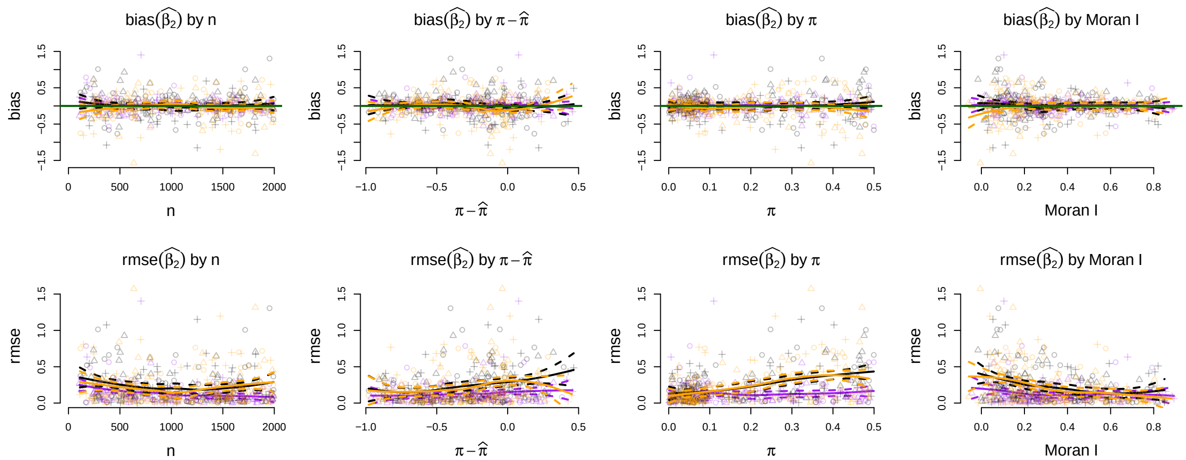


Figure D.6: Results from the simulation study, capturing the ability of the simple fixed effects model (m.1, in black), the King & Zeng model (m.2, in orange) and our proposed approach (m.3, in purple) to estimate the true covariate effect β_2.

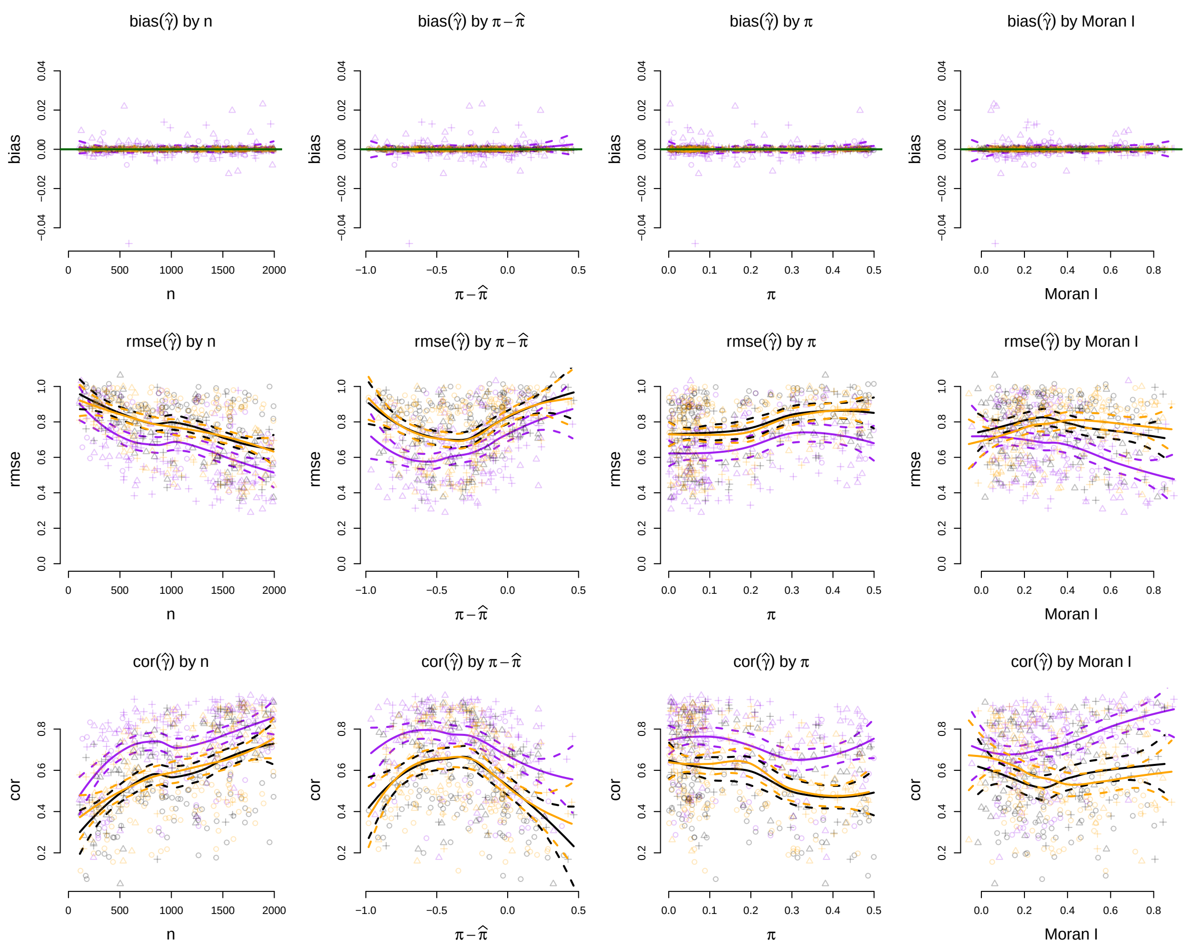


Figure D.7: Results from the simulation study, capturing the ability of the simple fixed effects model (m.1, in black), the King & Zeng model (m.2, in orange) and our proposed approach (m.3, in purple) to estimate the true area-level effects $\boldsymbol{\gamma}$.

E Practical Advice for Researchers

In this section we provide guidance to applied researchers who seek to analyse data on recruitment to extremist organisations - or any other sort of data which can plausibly be affected by selection on the dependent variable, contamination, spatial auto-correlation, etc. - using our modeling framework. We focus on specific features of a typical application, and leverage lessons from our simulation study as well as our experience from the ISIS recruitment example detailed in this paper. We also address features of the modeling framework which have not been tested, presenting our current understanding of their potential impact, and outlining ways in which future research could mitigate their influence. We refer to Rosenfeld (2018) for further treatment of the underlying assumptions of a contaminated case-control model.

i. **Sample size**: our proposed strategy is unbiased for estimating recruitment propensity μ starting at $n > 100$ (the minimum sample size tested in our simulation study). Compared to existing alternatives, our model affords greater returns on additional samples in terms of reduction of average error (RMSE). The advantage over other alternatives seems particularly evident for $n > 1000$ - where alternatives' RMSE tends to plateau, whilst our model's RMSE continues to decrease in seemingly linear fashion. A breakdown by estimated parameter tells us that our model affords some meaningful RMSE gains for every estimated parameter: for regression coefficients (both β_1 and β_2) we see a stable reduction of RMSE persisting after $n > 1000$, whilst alternative models plateau; the largest efficiency gains however are observed on the area-level effects γ.

   Our advice with respect to sample size is to use our model over any existing alternatives at any sample size. The recommended sample size to obtain the best results is $n > 2000$. We suspect decreasing returns will kick-in at some stage, but we have not tested where that point might be, and leave it for future research.

ii. **Population prevalence**: π plays a two-fold role in our model: it is used to calculate an optimal offset to account for selection, and it is used to account for contamination in the unlabeled controls. Our model is robust to any level of $\pi \in (0, 0.5)$. Under a contaminated data generating process, this is in stark contrast with the best available alternative - the King & Zeng model - which suffers from negative-bias as the contamination rate increases. The gains appear primarily as a result of estimating an unbiased intercept β_1, but also as a result of having a precise estimate of β_2 at any level of π - whilst the King & Zeng model suffers from roughly linearly-increasing RMSE at increased contamination rates.

   Our advice with respect to the true population prevalence is to use our model at any level of π. There is no recommended level of prevalence at which our model's ability to estimate either propensity μ or regression coefficients β_1 and β_2 underperforms - our model is simply robust to contamination, given knowledge of π is available.

iii. **Discrepancy between sample & population prevalence**: $\pi - \hat{\pi}$ captures the difference between the true prevalence and the sample-prevalence. This arises as a direct result of the stacking procedure, where whatever available cases are artificially appended to a sample of unlabeled controls. Note that this is a measure of the degree to which our sample ends up being a non-probability sample.

    Here, compared to the best available alternative, we have a trade-off: both models will tend to have biased intercepts at very high-levels of $\pi - \hat{\pi}$. The King & Zeng model will tend towards a negative bias, as the offset used here does not take into account contamination, and the relatively large numbers of unlabeled controls will be 'hiding' a very large number of cases, meaning the offset is inaccurately calibrated. This model essentially believes that there are less cases in the data than in reality, and therefore estimates a smaller intercept than it should. On the other hand, our model will tend to do the opposite: by accounting for contamination, it will tend to believe there are relatively more cases in the sample than there actually are in the population. This bias however tends to be smaller, and appear at higher levels of discrepancy, compared to that of the King & Zeng model.

    Our advice with respect to the size of the discrepancy between sample and population prevalence is as follows: researchers should prefer to account for contamination rather than not, due to the relatively smaller bias and RMSE on the intercept. They should however be aware that in a regime of large under-sampling of the cases, the bias will be in the positive direction under our contamination model, but in the negative direction if contamination is not accounted for. Researchers should attempt to create stacked-samples that have sample prevalence roughly equal to population prevalence; over-sampling of cases is not an issue, but under-sampling is, if the discrepancy is large. As a conservative guide, we advise researchers to ensure their stacked sample does not under-sample cases by more than 10 percentage points, relative to population prevalence, when using our model.

iv. **Spatial auto-correlation**: spatial auto-correlation in the area-level effects γ tends not to affect the bias of the model estimates, but it does impact efficiency (in terms of RMSE) and the ability of models to properly rank individuals according to their underlying recruitment propensity μ. Large discrepancies in RMSE and correlation of γ tend to kick-in around a Moran-I of 0.2; the comparison with other models increasingly favours our approach as I increases, generating an advantage as large as 0.2 correlation points at high-levels of spatial auto-correlation. Note that under low-levels of auto-correlation, our model performs at least equally well as any other alternatives.

    Our advice with respect to spatial auto-correlation is to explicitly account for it in the model. There are no drawbacks to doing so in terms of the metrics we test. There might be some issues relative to the 'coverage' ability of the model, which we could not test due to computational burden, in the sense that estimates of γ under our model will be greatly 'shrunk' and therefore tend to have smaller uncertainty than fixed effects. But regardless of the potential coverage issues, the gains in terms of ability to order and discriminate between profiles are sufficiently large we feel comfortable advising to use the ICAR model to account for spatial auto-correlation.

v. **Prior knowledge of population prevalence**: we encode in our model's contamination layer an expectation that the unlabeled cases will be recruited at a rate roughly equal to the population prevalence of recruitment. The validity of this assumption depends on the application at hand. For us, the known population propensity made for a good prior because our hypothetically-contaminated observations came from a random sample of the population, so the known prevalence, and the sampling design of the Arab Barometer, were at the same level. However this assumption becomes increasingly inappropriate as the sampling frame of the contaminated units veers further away from a random sample of the population of interest. The degree of error induced by misspecification is not tested in our simulation study - it is assumed that the correct prevalence is always known.

   Speculating on the potential effects of misspecification we can consider the nature of the bias we would be introducing: artificially increasing the contamination rate relative to the true population rate will positively bias the intercept of the model, by 'flipping' an unreasonable amount of unlabeled observations to 'cases'. This is similar to what we see in the effect of sample and population prevalence discrepancy $\pi - \hat{\pi}$.

   In a fully Bayesian model this parameter can be estimated from the data (Rota et al. (2013); Rosenfeld (2018)) though we have found, anecdotally in our experimenting for this paper, that under this fully-Bayesian approach the estimated posterior of π tends to be necessarily biased by the selection effects into the artificially-stacked samples. This is somewhat in contradiction with Rosenfeld (2018), and we merely point this out to encourage further exploration of this question.

   Our advice with respect to how to best use prior knowledge on the population prevalence is to 'use it with care'. If it is known with certainty, we advise to use it, and introduce it as an 'observed value' for the prevalence in the Bayesian model. If no knowledge on the true population prevalence is available, we do encourage researchers to perform a fully-Bayesian analysis and estimate the prevalence as part of the model parameters; however we would further advise, where possible, to use strongly informative priors to counteract the bias introduced by the artificial sampling design.

vi. **Exogenous Selection**: a dimension on which our model is untested is the degree to which non-random samples of cases could bias the analysis. Examples of this would be if the sample of recruits in our data was obtained via snowball sampling, or any other sampling design which is vulnerable to systematic distortions brought about by exogenous effects.

This paper takes the view that it is generally hard-to-impossible to obtain representative samples of recruits, and therefore builds-in a series of robustness measures - such as regularising priors, random effects, and mixture models, to limit the negative effects of non-representativity. More regularising approaches are worth exploring and introducing into the modeling framework - for instance the use of regularised horseshoe priors (Piironen and Vehtari (2017)) could contribute to excluding irrelevant covariates from the analysis, as well as regularising coefficients, hence encouraging the avoidance of over-fitting to potentially biased data. It is further worth noting that this model is amenable to post-stratification (Hanretty et al. (2018); Park et al. (2004)), which would allow for more representative estimation of average recruitment propensities at the small-area level. Enhancing this analysis with a post-stratification layer would enable the use of even more extreme unrepresentative samples of recruits, such as individuals observed to be extremists on social media or in other unconventional samples (Wang et al. (2015); Cerina and Duch (2020)). But this approach would not solve the bias in regression coefficients, and is only relevant if the small-area is indeed the desired level of analysis.

Our advice with respect to the potential impact of exogeneous selection effects is to build into the model reasonable protections against over-fitting to biased data. In our case, this is possible to some degree through the use of various regularising priors. In general the conventional wisdom stands: if it is at all possible to obtain a representative sample of cases, researchers should do whatever they can to obtain it. However, reality dictates that this is very rarely possible, especially in the context of extremist movements. Our regularised modeling approach therefore becomes the preferred solution.

F Convergence diagnostics

To ensure absolute convergence of all model parameters we run our model with extremely conservative settings: `n.iter` $> 10,000$, `n.warmup` $> 9,000$,[14] `n.chains` > 4; `n.thin` $=$ `n.cores` $=$ `n.chains`; `max_treedepth` $= 25$, `adapt_delta` $= 0.99$. Note that the Worm's Eye models take around 12 hours to run for Egypt, 24 hours for Tunisia, whilst the Bird's Eye model takes 48 hours. As a final note, it's worth highlighting that convergence of point estimates for the individual-level covariates happens under far more laxed estimates, and exploratory versions of this model can be fit under 1 hour in all cases.

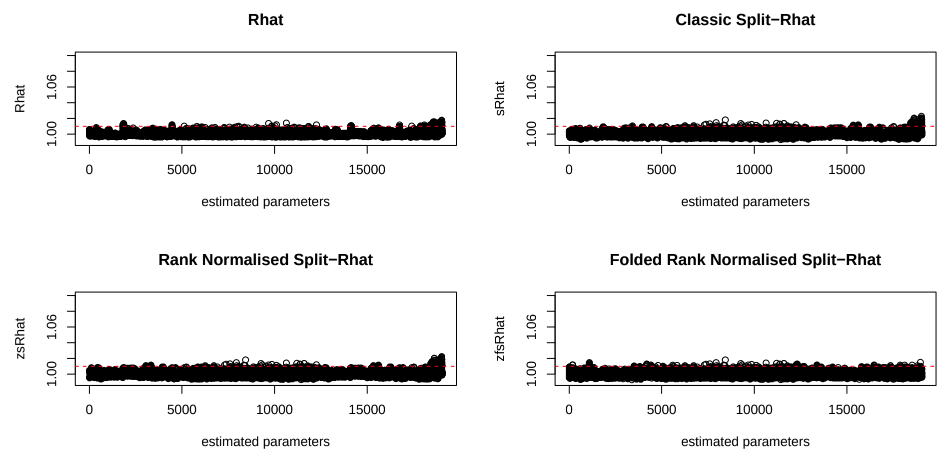


Figure F.1: Gelman-Rubin Statistics for the *Bird's Eye* model.

[14]For the 'Bird's Eye model, we set `n.iter` $= 10,000$ and `n.warmup` $= 9,000$, and ran the model over 8 chains spread over 8 cores, thnning by a factor of 8 – whilst for the Worm's Eye models we can afford a larger number of iterations – `n.iter`$= 25,000$ and `n.warmup` $= 22,500$, runnng 4 chains spread over 4 cores, and thinning by a factor of 4.

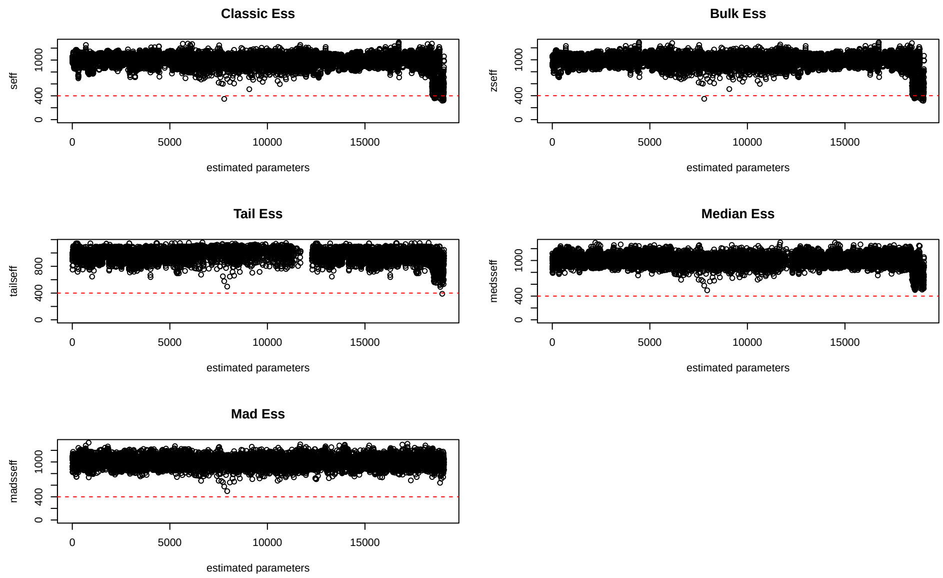


Figure F.2: Effective sample-size (ESS) for the parameters of the *Bird's Eye* model.

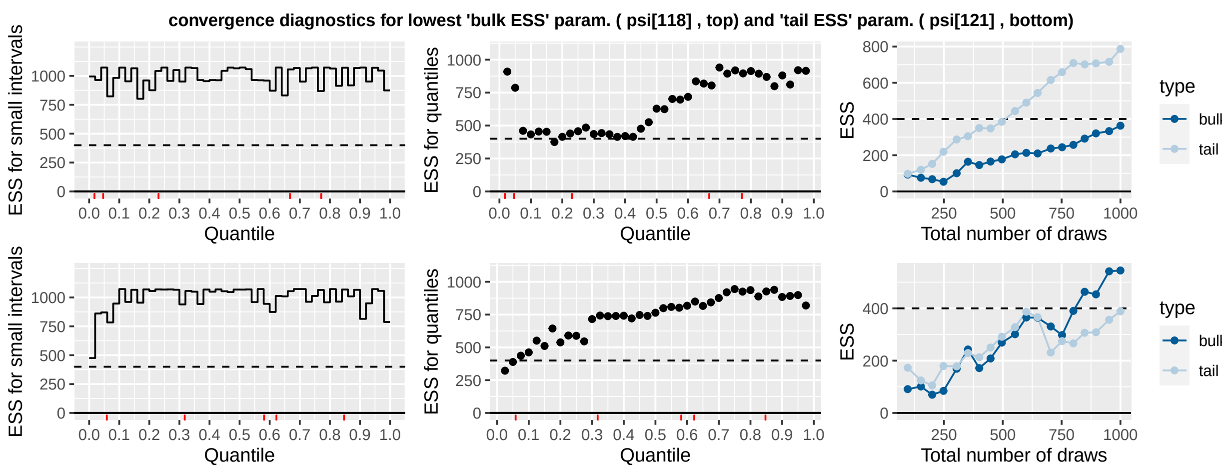


Figure F.3: Convergence dynamics for the parameters with the lowest bulk (top) and tail (bottom) ESS, for the *Bird's Eye* model. The quantile plots show satisfactory ESS for every section of the posterior distribution, whilst the positive and close-to-linear gradient in the 'total number of draws' plot suggests ESS would improve further by drawing more samples – a sign that the posterior is well-explored.

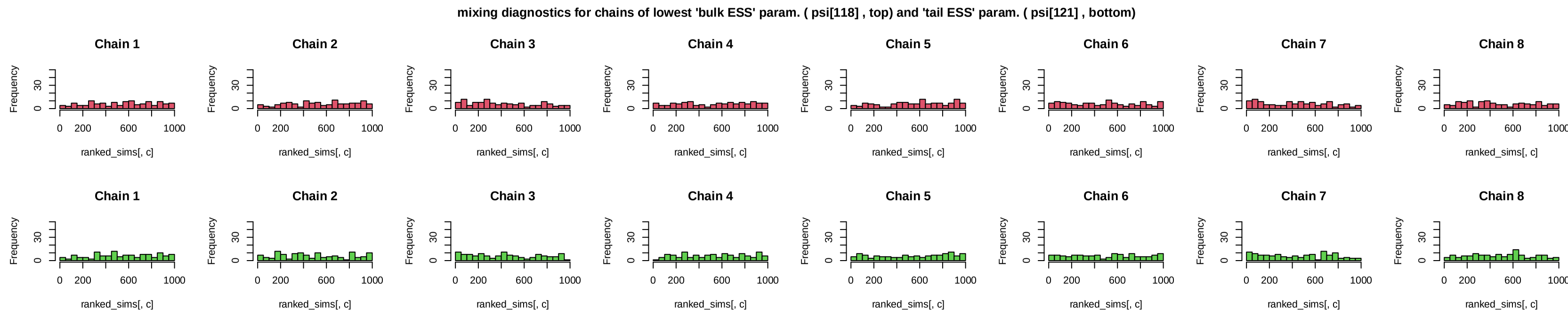


Figure F.4: Histogram of the ranked posterior draws for the parameters with the lowest bulk (top) and tail (bottom) ESS, for the *Bird's Eye* model. This plot is evidence of reasonably good mixing also for our 'least convergent' parameters, as the ranked draws from each chain could be reasonably drawn from a uniform distributions.

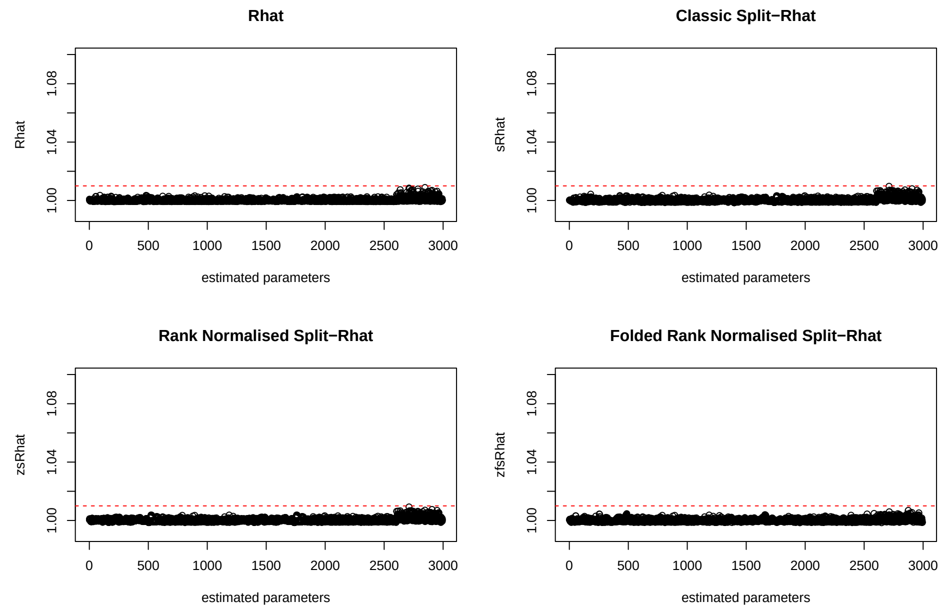


Figure F.5: Gelman-Rubin Statistics for the Egypt *Worm's Eye* model.

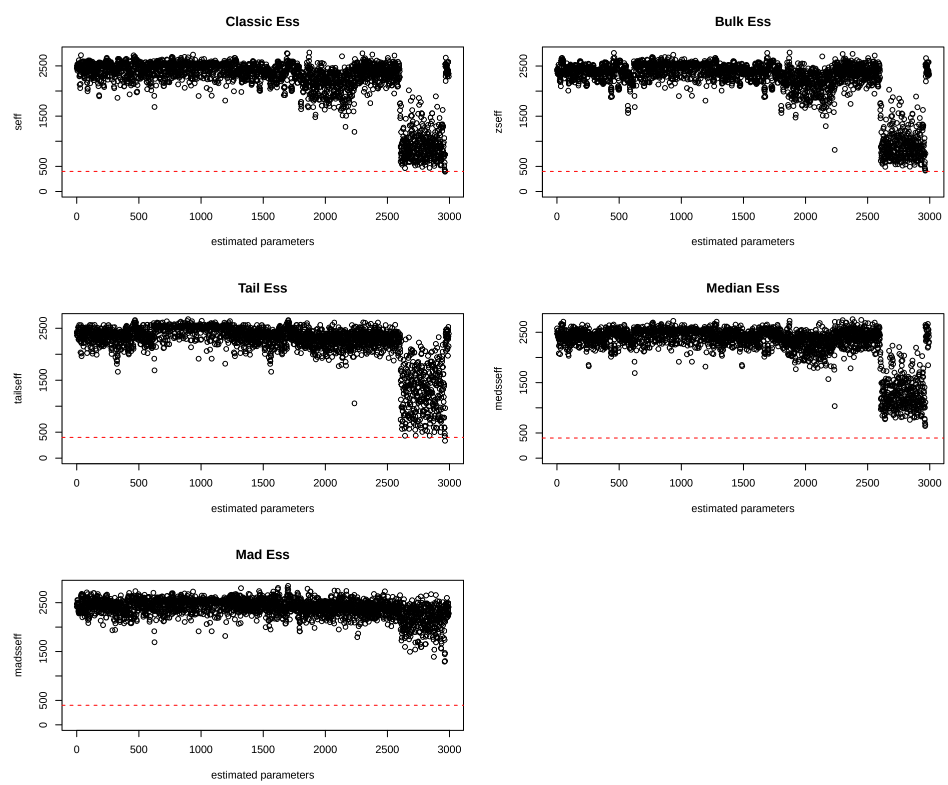


Figure F.6: Effective sample-size (ESS) for the parameters of the Egypt *Worm's Eye* model.

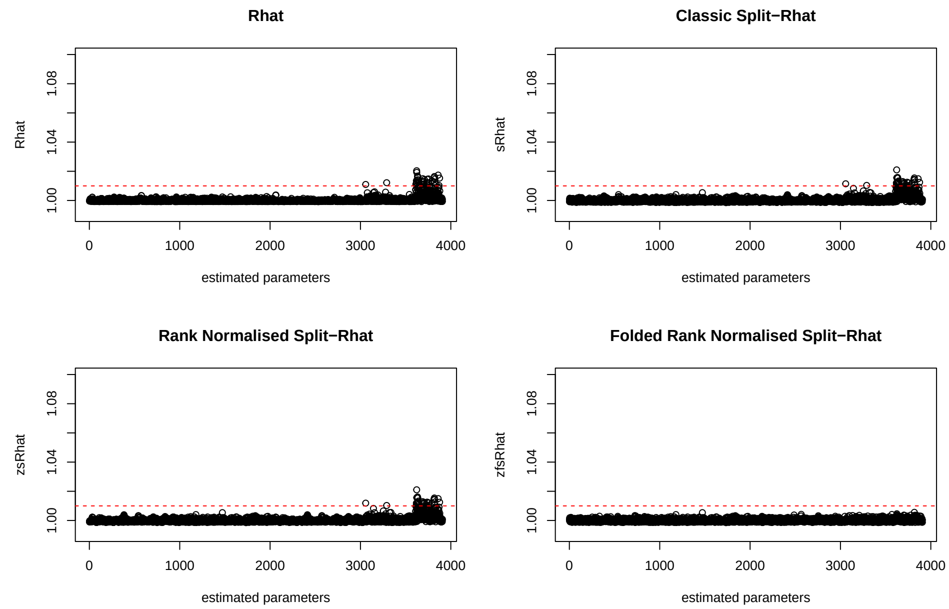


Figure F.7: Gelman-Rubin Statistics for the Tunisia *Worm's Eye* model.

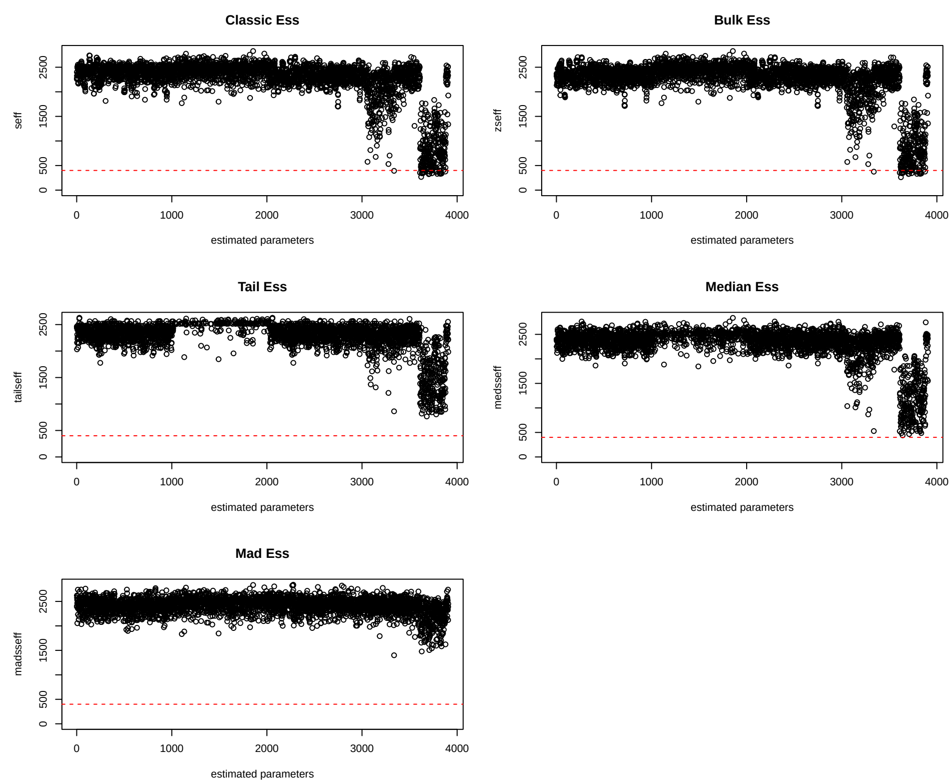


Figure F.8: Effective sample-size (ESS) for the parameters of the Tunisia *Worm's Eye* model.

G Posterior densities of regression coefficients

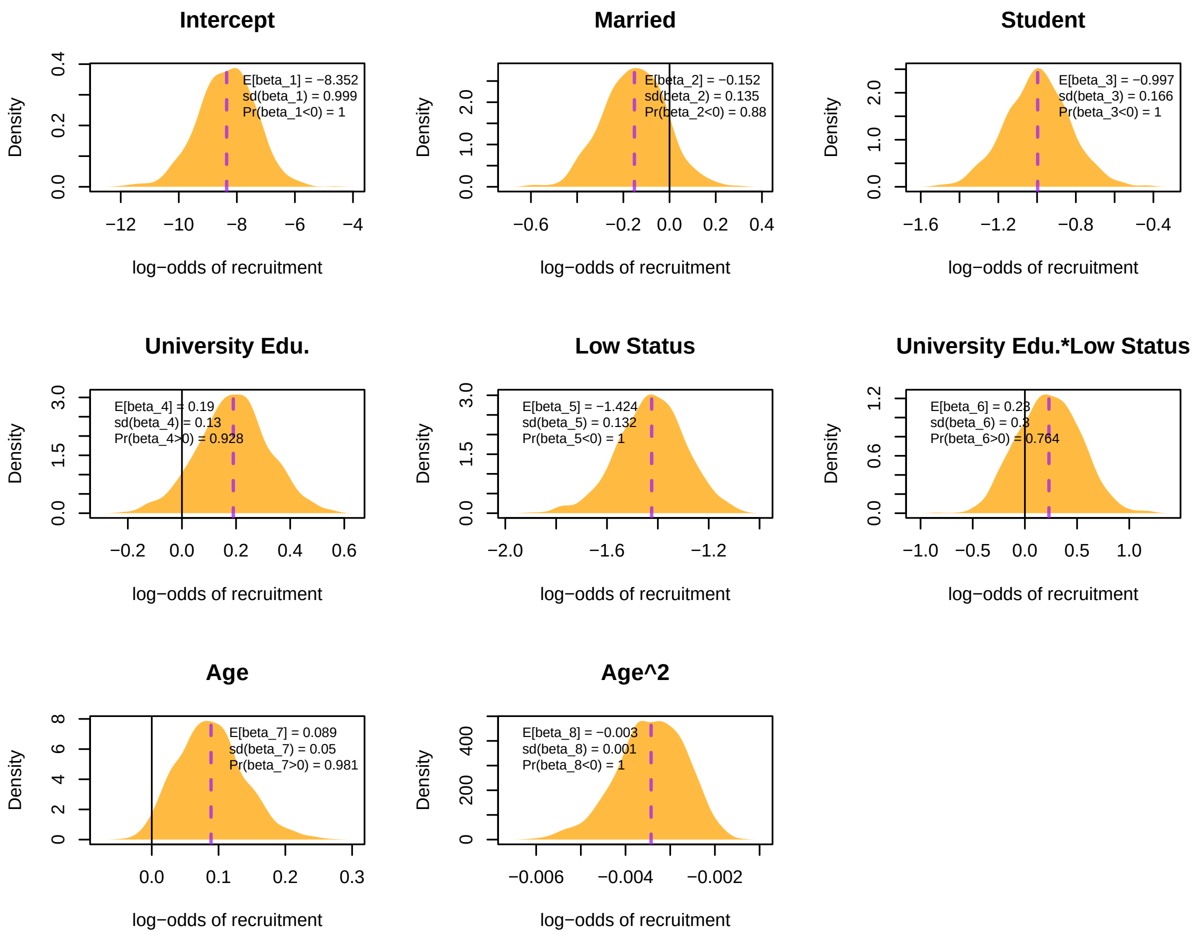


Figure G.1: Posterior density of individual-level fixed-effect coefficients for the 'Bird's Eye' model. These effects are presented on the original, non-standardized scale.

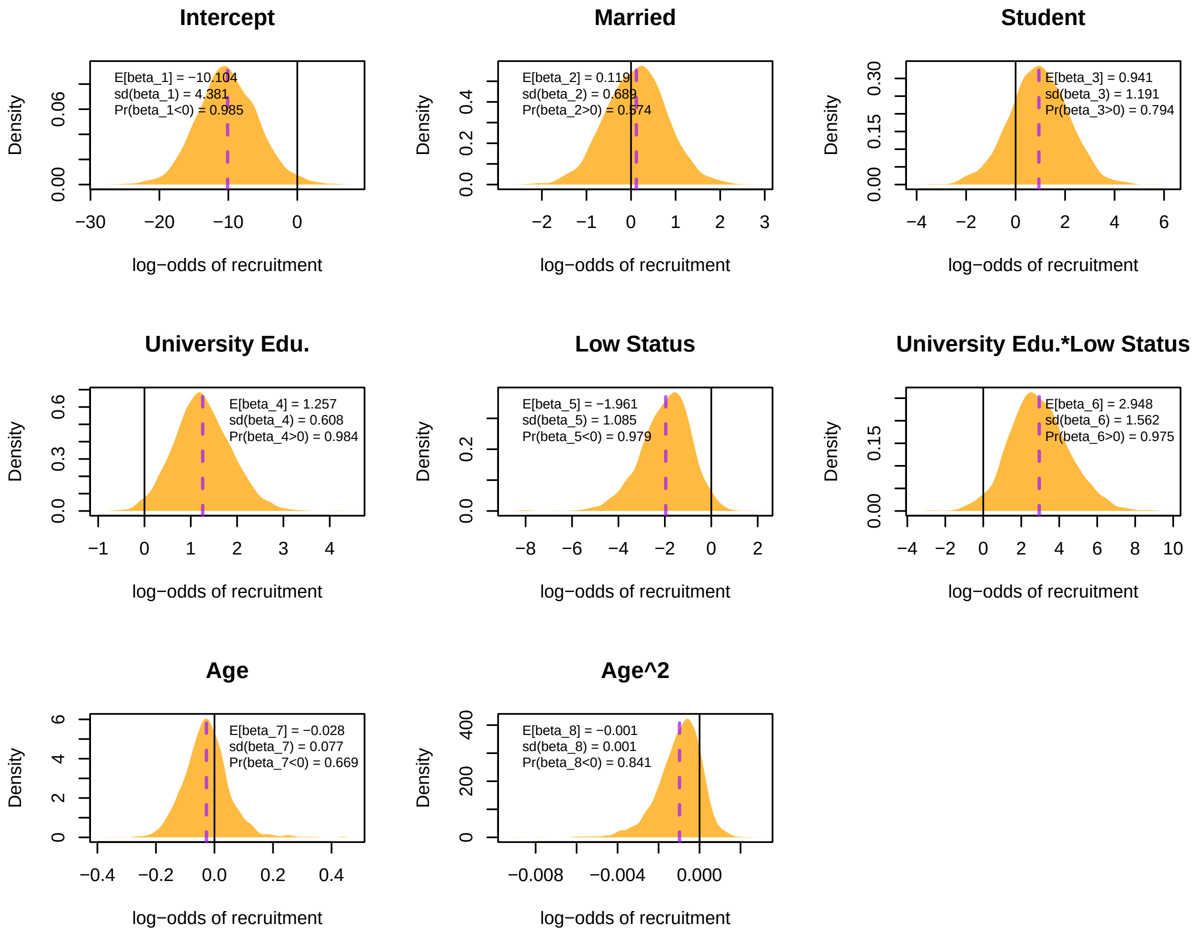


Figure G.2: Posterior density of individual-level fixed-effect coefficients for the Egypt 'Worm's Eye' model. These effects are presented on the original, non-standardized scale.

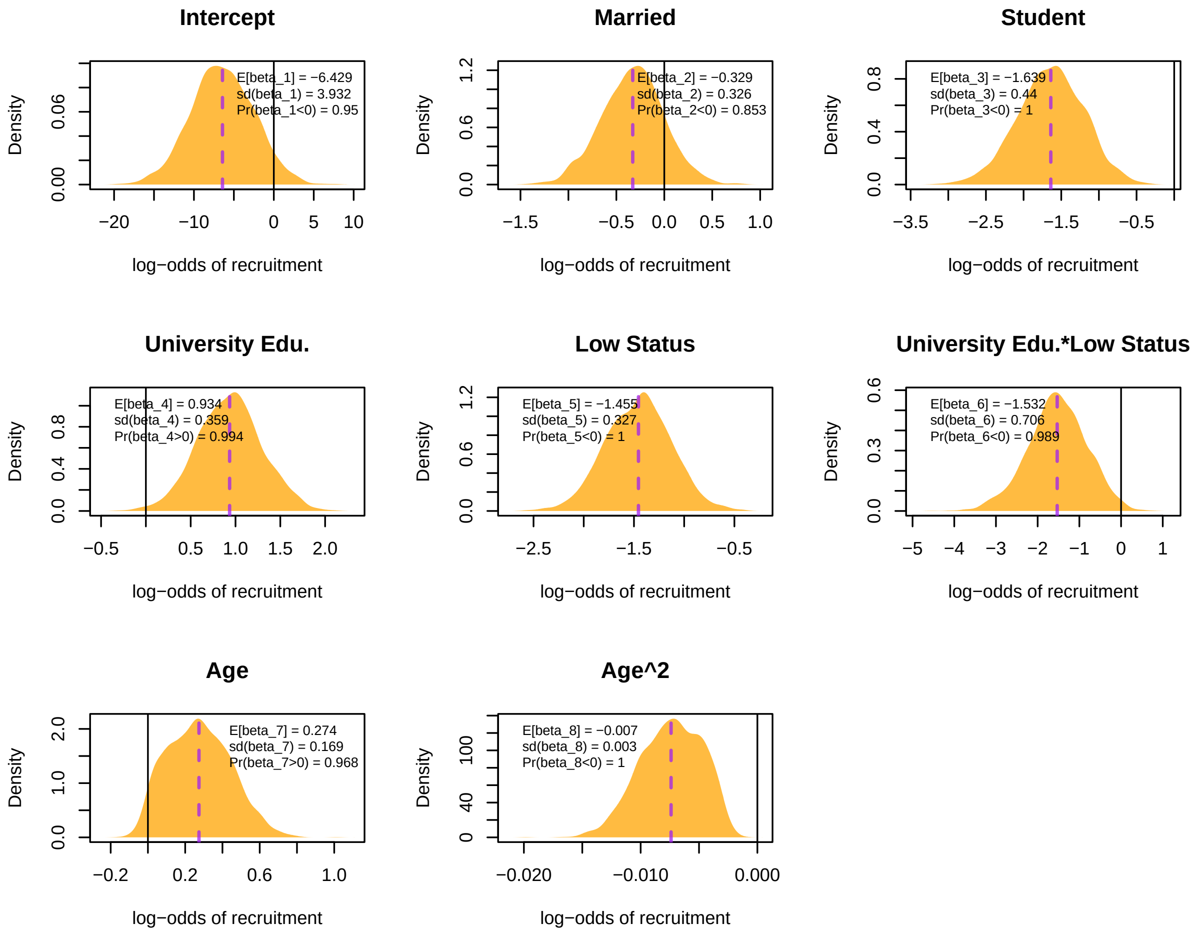
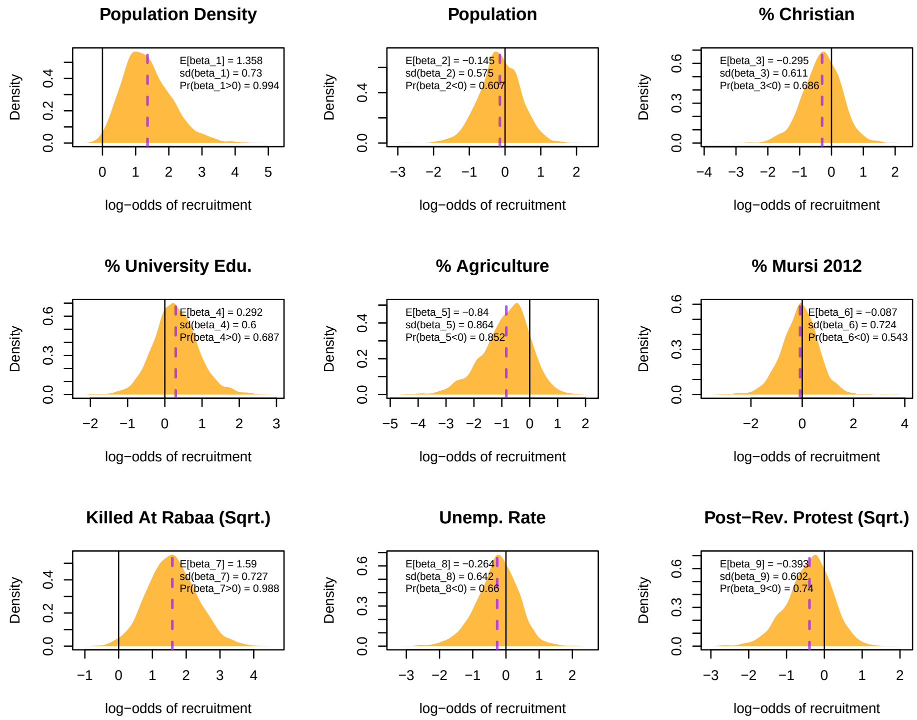
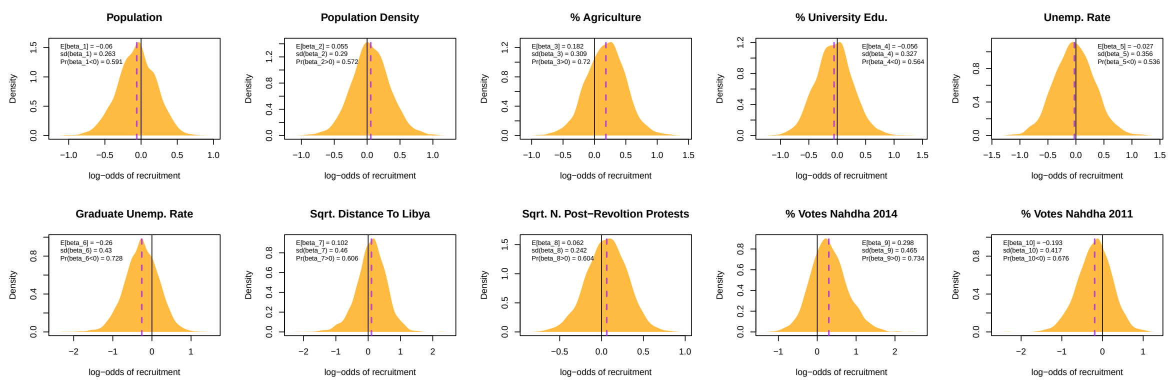


Figure G.3: Posterior density of individual-level fixed-effect coefficients for the Tunisia 'Worm's Eye' model. These effects are presented on the original, non-standardized scale.

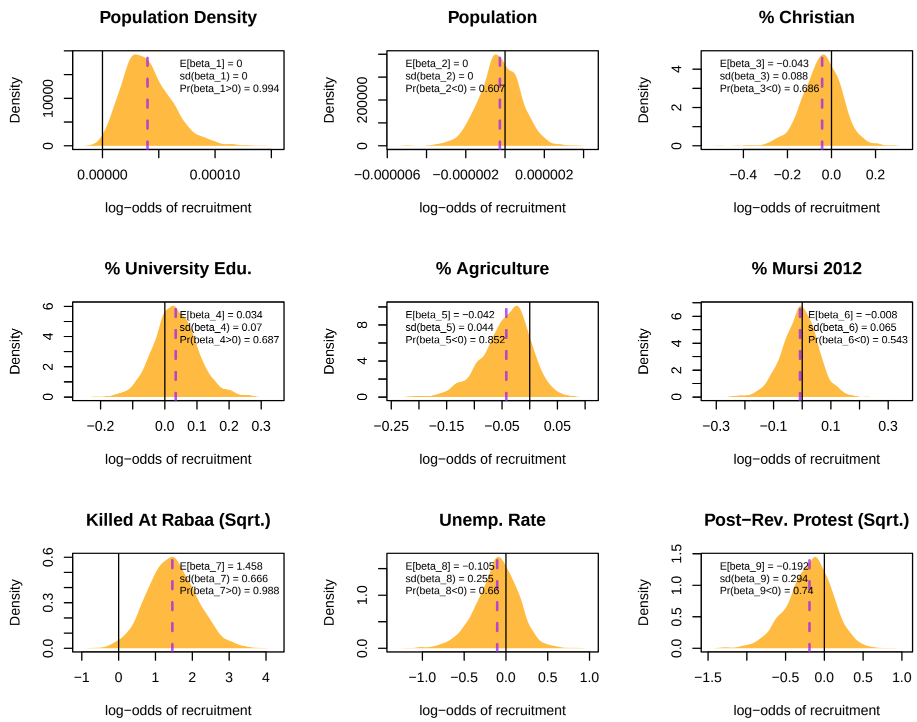


(a) Egypt

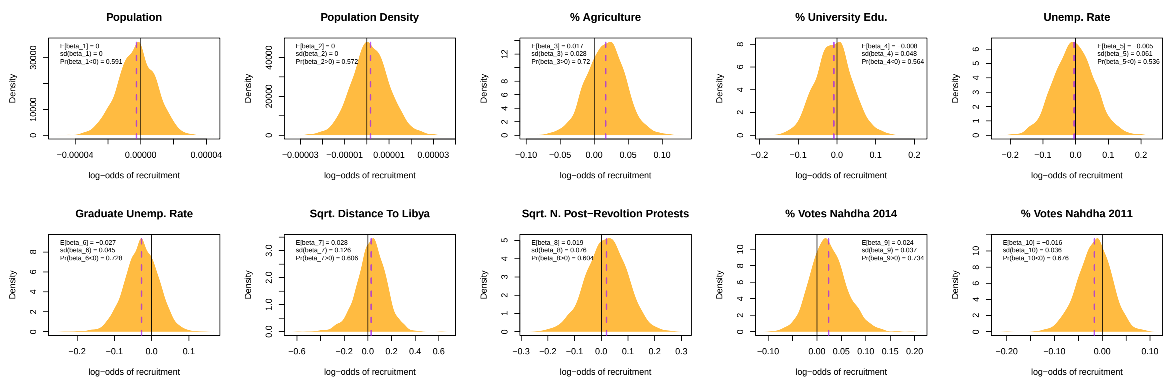


(b) Tunisia

Figure G.4: Posterior density of district-level fixed-effect coefficients for the *Worm's Eye* models.



(a) Egypt



(b) Tunisia

Figure G.5: Posterior density of district-level fixed-effect coefficients for the *Worm's Eye* models. These effects are presented on the original, non-standardized scale.

H Relative Deprivation Effects

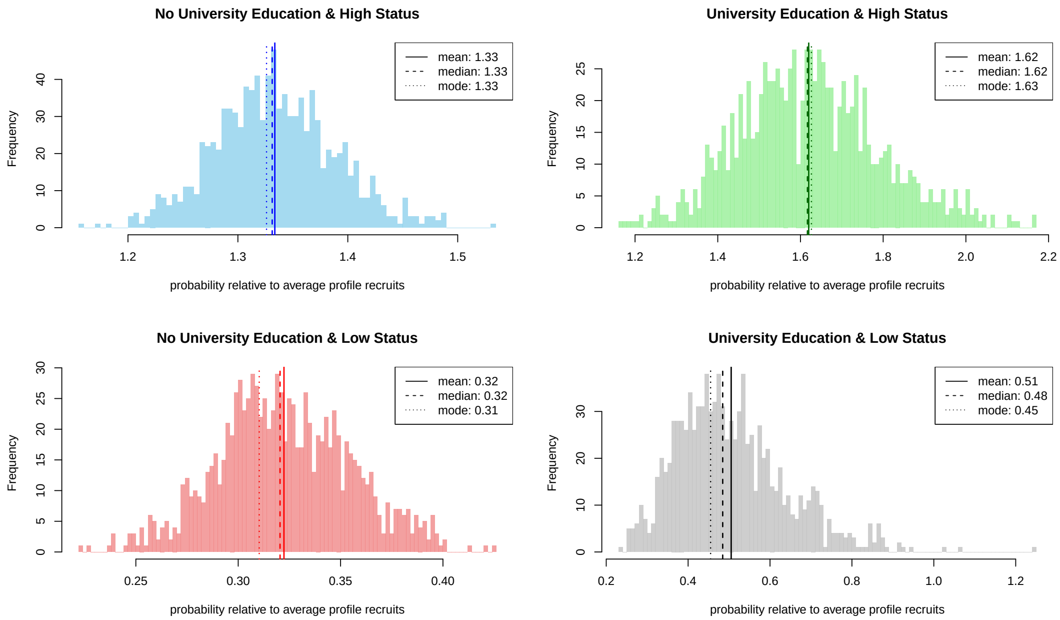


Figure H.1: Predicted propensity of recruitment for relative-deprivation profiles according to the 'Bird's Eye' model. The effects are presented as odds relative to the 'average' recruitment profile.

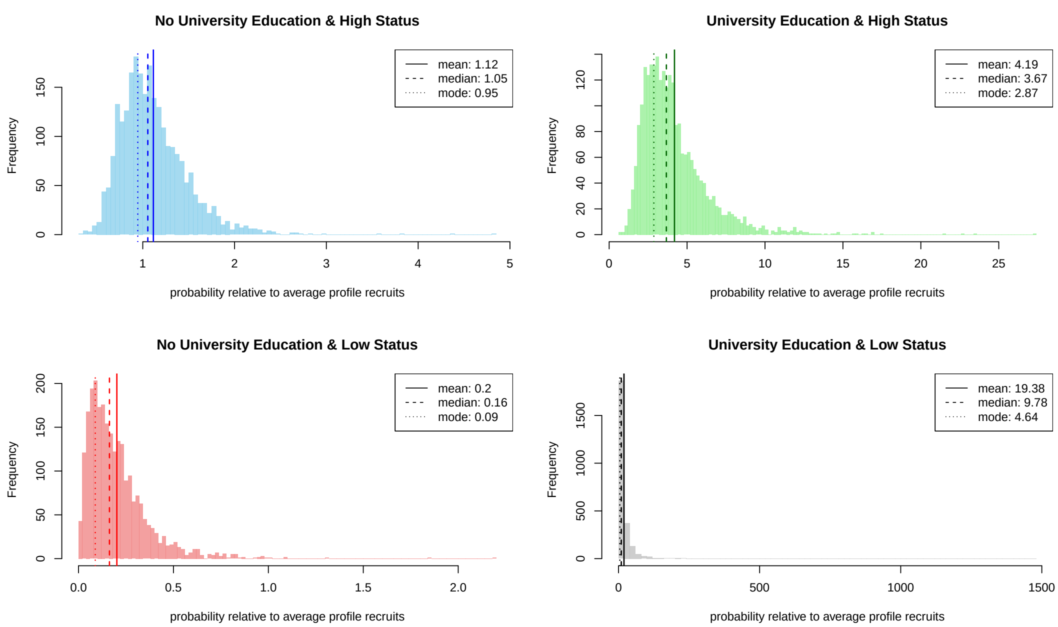


Figure H.2: Predicted propensity of recruitment for relative-deprivation profiles according to the Egypt 'Worm's Eye' model. The effects are presented as odds relative to the 'average' recruitment profile.

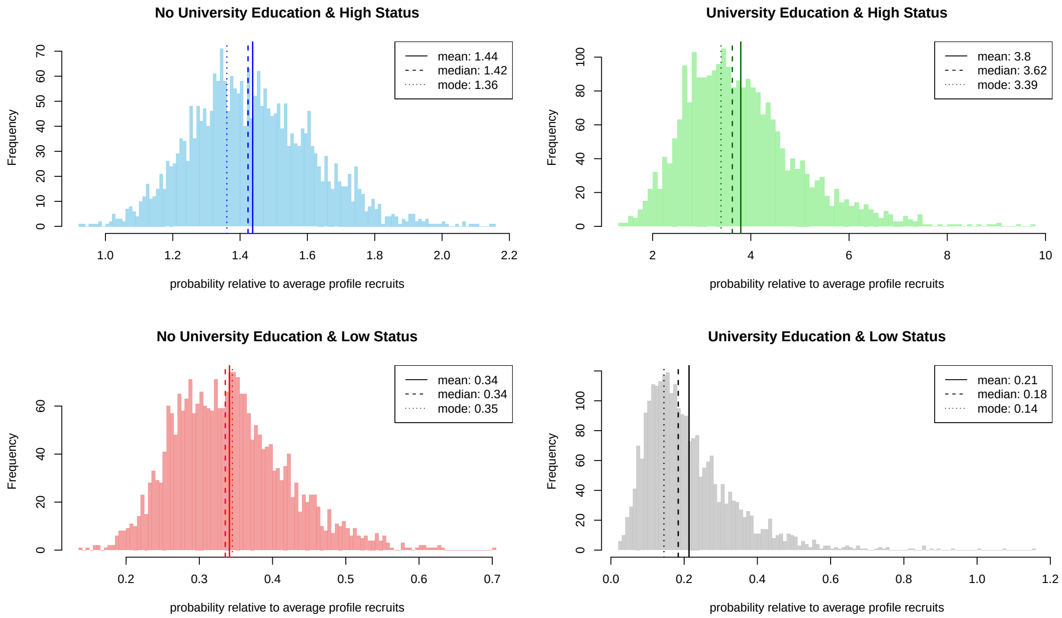


Figure H.3: Predicted propensity of recruitment for relative-deprivation profiles according to the Tunisia 'Worm's Eye' model. The effects are presented as odds relative to the 'average' recruitment profile.

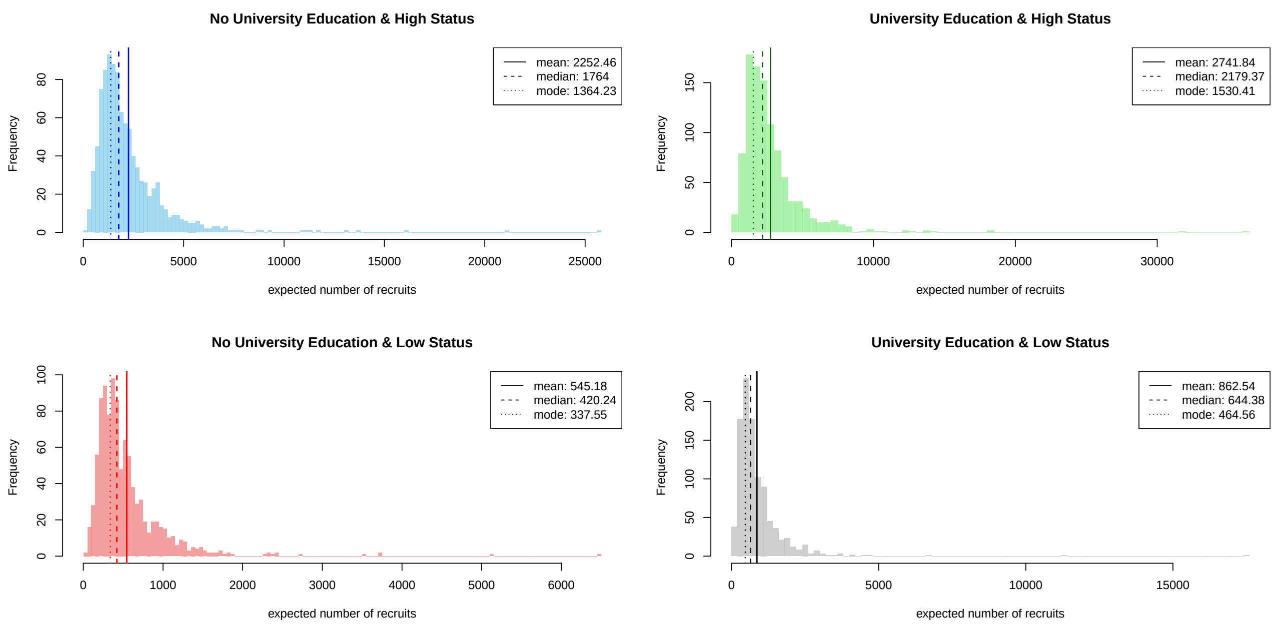


Figure H.4: Predicted propensity of recruitment for relative-deprivation profiles according to the 'Bird's Eye' model. The effects are presented as predicted counts under the assumption that everyone in the population is an 'average profile', and only changing the profile's relative deprivation status.

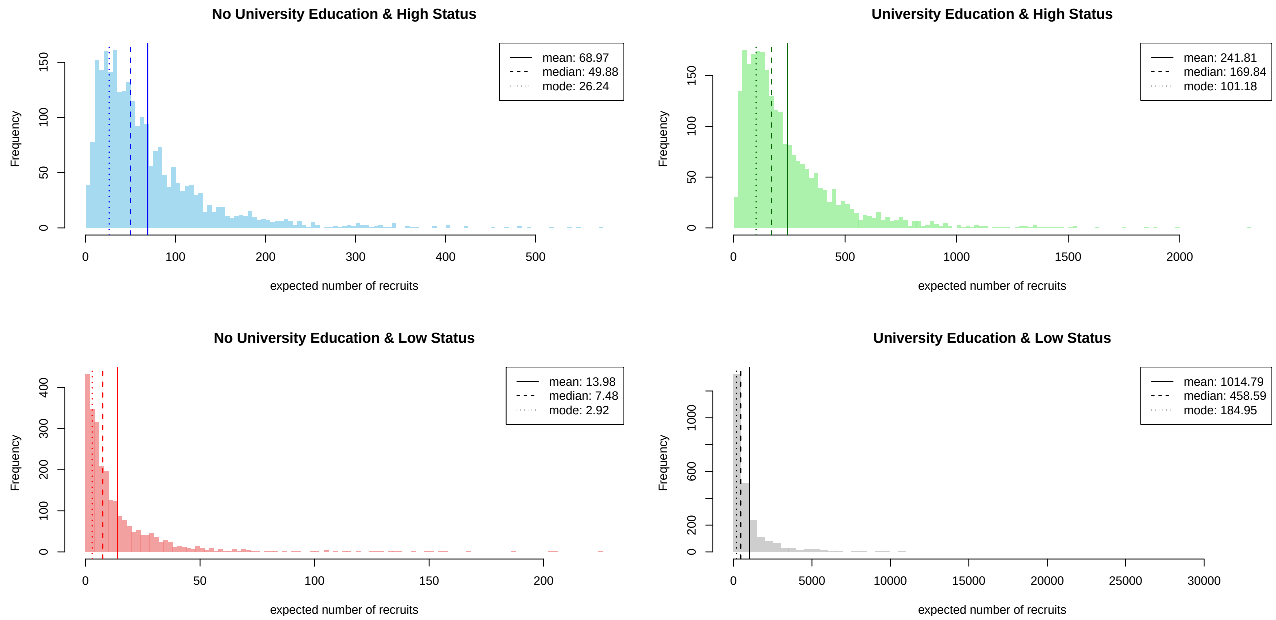


Figure H.5: Predicted propensity of recruitment for relative-deprivation profiles according to the Egypt 'Worm's Eye' model. The effects are presented as predicted counts under the assumption that everyone in the population is an 'average profile', and only changing the profile's relative deprivation status.

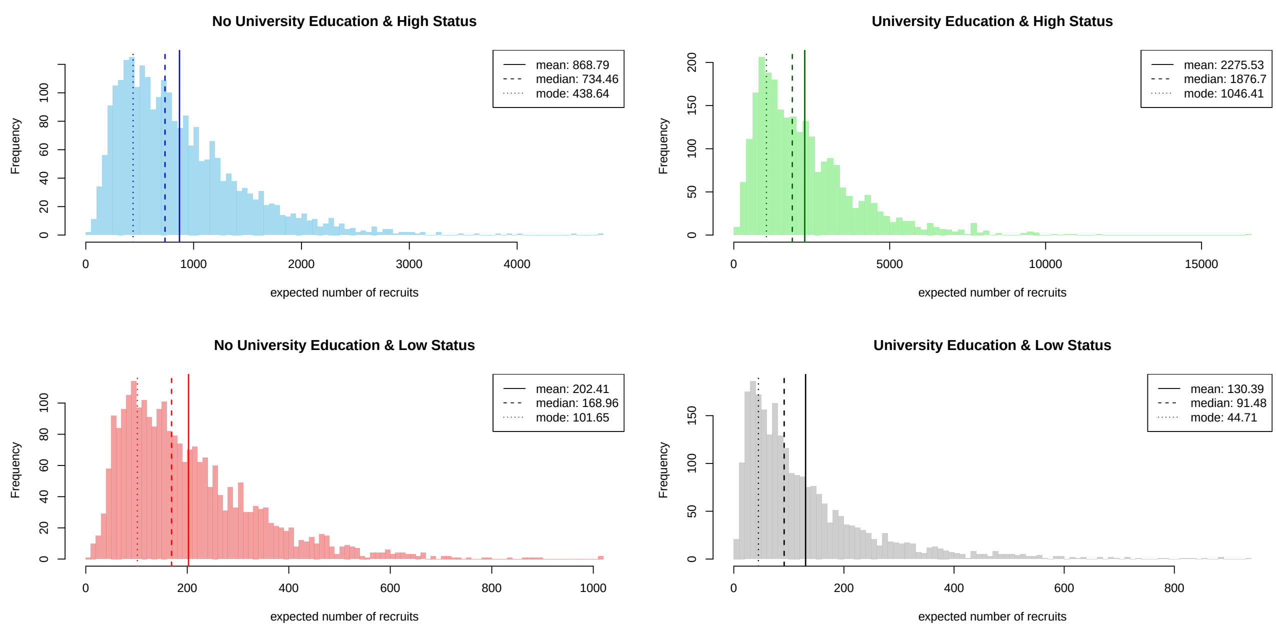


Figure H.6: Predicted propensity of recruitment for relative-deprivation profiles according to the Tunisia 'Worm's Eye' model. The effects are presented as predicted counts under the assumption that everyone in the population is an 'average profile', and only changing the profile's relative deprivation status.

I Residual Area-Level Analysis

In Figure 5a in the main manuscript, edges connect nodes identified by the centroids of governorates for each country. Minor adjustments were performed to ensure the absence of islands or sub-graphs, which would have made the analysis needlessly complicated. Note also that Israel and Saudi Arabia are included for the purpose of obtaining this fully-connected graph, but no observations were available for either of these countries in terms of recruits or Arab Barometer observations, and hence the direction of the estimates for their governorates is entirely driven by the spatial component. Supplementary Figure I.2 displays the observed number of recruits per area, along with the residual for each governorate.

We are satisfied that the spatial pattern implied by the adjacency matrix derived from the fully connected graph is completely extracted from the residuals, as shown by the relatively uniform color pallet of the rightmost map in Figure I.2, and most importantly the posterior distribution of the residuals' Moran's I in Figure I.1, which is normally distributed around the expected null value.

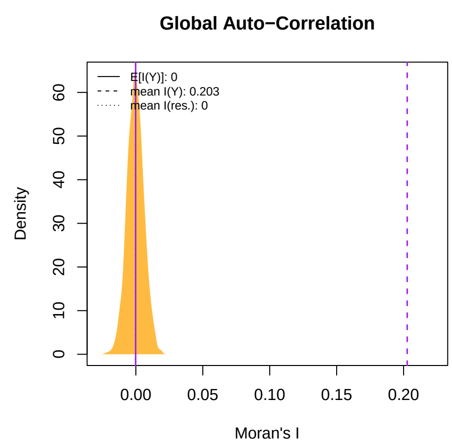
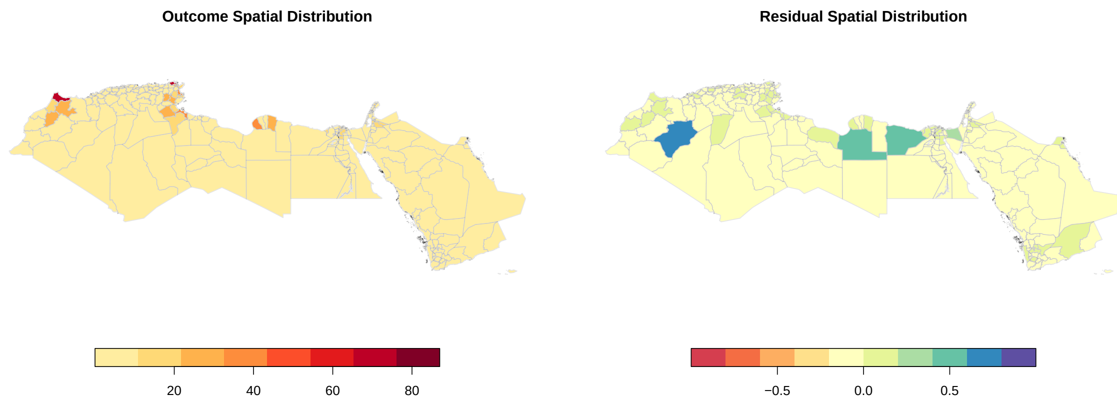


Figure I.1: Posterior distribution of Moran's I, a coefficient of global spatial auto-correlation. The adjacency matrix implied by Figure 5a is used as the weight matrix. $I(Y)$ indicates the coefficient value prior to spatial modeling; $I(res)$ shows the complete nullification of auto-correlation as a result of the ICAR prior. The expected value under the null distribution, $E[I(Y_{null})]$, is calculated as $\frac{-1}{(n_1+n_u)-1}$.

Figure I.2 displays the observed number of recruits per area, along with the residual for each governorate.

In I.2b the residual is calculated as follows: take $z = z_1, ..., z_L$ to be the subset of individuals $i \in z_l$, who belong to small-area l; take $s = 1, ..., S$ to be the index of posterior sample draws; then $\text{res}_l = \frac{1}{S}\sum_s \left[\frac{1}{\sum_i \mathbb{1}(i \in z_l)} \sum_{i \in z_l} (y_i - \hat{y}_{i,s})\right]$. A first concern is the presence of spatial autocorrelation in the recruitment data, which could bias individual-level coefficients. The spatial distribution in Figure I.2a seem to suggest the possibility of spillover effects around high-density coastal areas. This is confirmed by



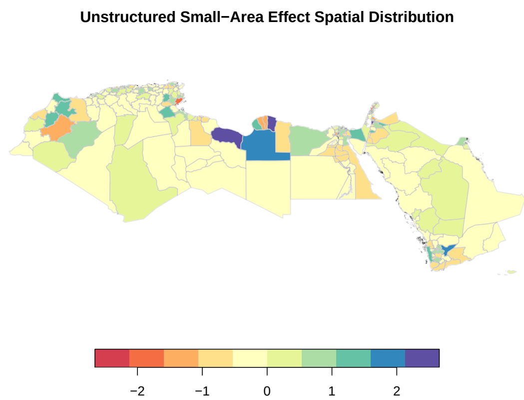
(a) (b)

Figure I.2: Spatial distribution of observations (a) and residuals (b) at the Governorate level.

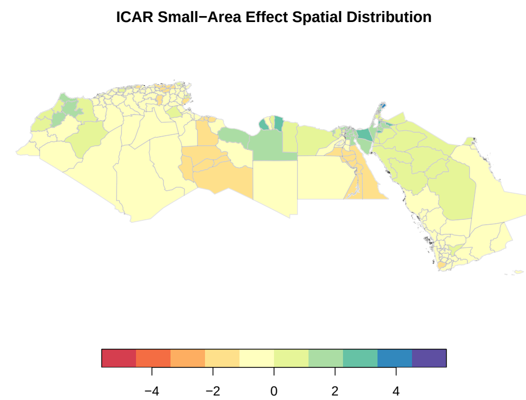
the Moran's I ($I(Y)$), which shows statistically significant spatial auto-correlation.[15]

We display below the spatial distribution of the point estimates for Governorate and Country-level random effects in Figure I.3. The corresponding prediction intervals for country and governorate effects are shown in Figure I.4 for country-effects and Figure I.5 for Governorate effects. It is worth noting that part of the reason for heightened recruitment propensity around the eastern Governorates could be the increasing proximity to Syria and the ISIS caliphate itself, as well as higher proportions of refugees from destabilized regions of Syria, and in general more potential for pro-ISIS unobservable network-dynamics. We see a strong unexplained effect in Tunisia, highlighting unobserved but systematic variance in favour of recruitment , while Algeria, Egypt and Yemen show significant unexplained negative effects on recruitment over and above their spatial and unstructured Governorate-level variance.

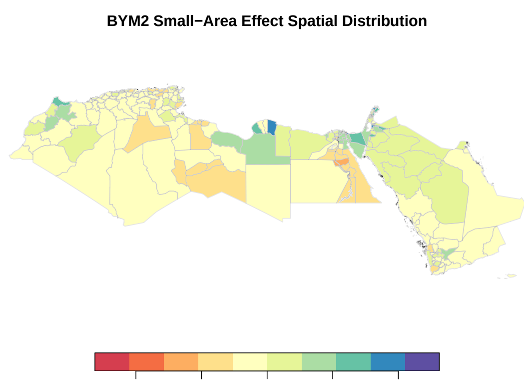
[15] As I(Y) is an observed, not modeled, quantity, it carries no uncertainty around it; it is reasonable to assume that the distribution of the $I(Y)$ would be the same as that of the $I(res)$ in terms of its shape and variance, and only differ as a result of the mean paramater. This is what is commonly assumed under standard hypothesis testing. Hence, it is easy to see that by applying the extremely narrow simulated variance around the $I(Y)$ dotted line, there would be a 0 probability of that distribution crossing the $E[I(Y)]$ line, and hence we can say the $I(Y)$ is highly significant. Calculating the significance of $I(Y)$ in frequentist terms, using the `ape` package, reveals a p-value of 0. We plot and describe our calculation for Moran's I in Figure I.1



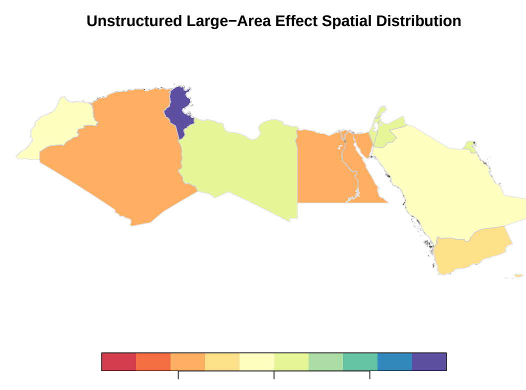
(a)



(b)



(c)



(d)

Figure I.3: Spatial distribution of: (a) the unstructured Governorate-level effect - ϕ; (b) the spatial Governorate level effect - ψ; (c) the total Governorate effect - $\gamma = \sigma(\phi\sqrt{1-\lambda} + \psi\sqrt{\lambda/s})$; (d) the unstructured Country effect - η.

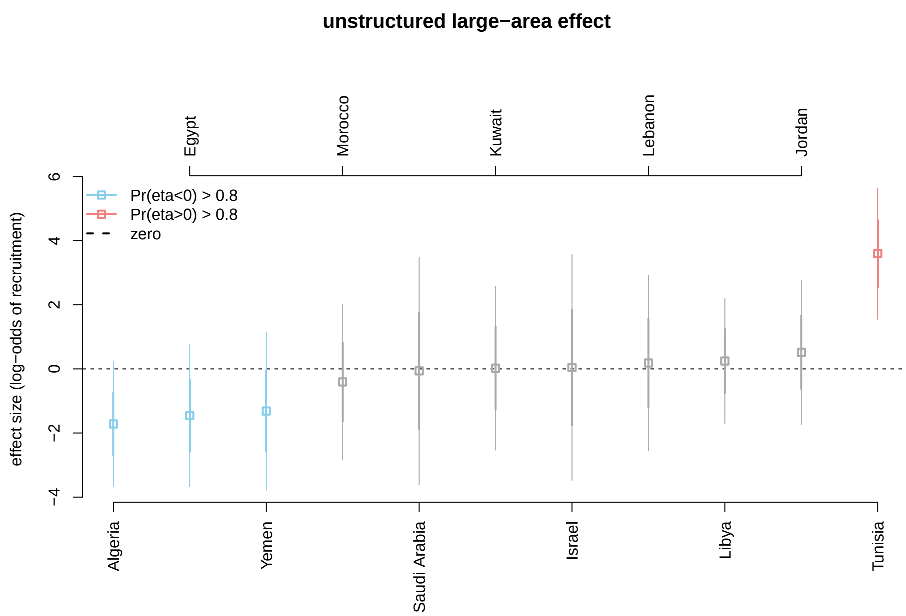


Figure I.4: Unstructured large-area effect η for the 'Bird's Eye' model.

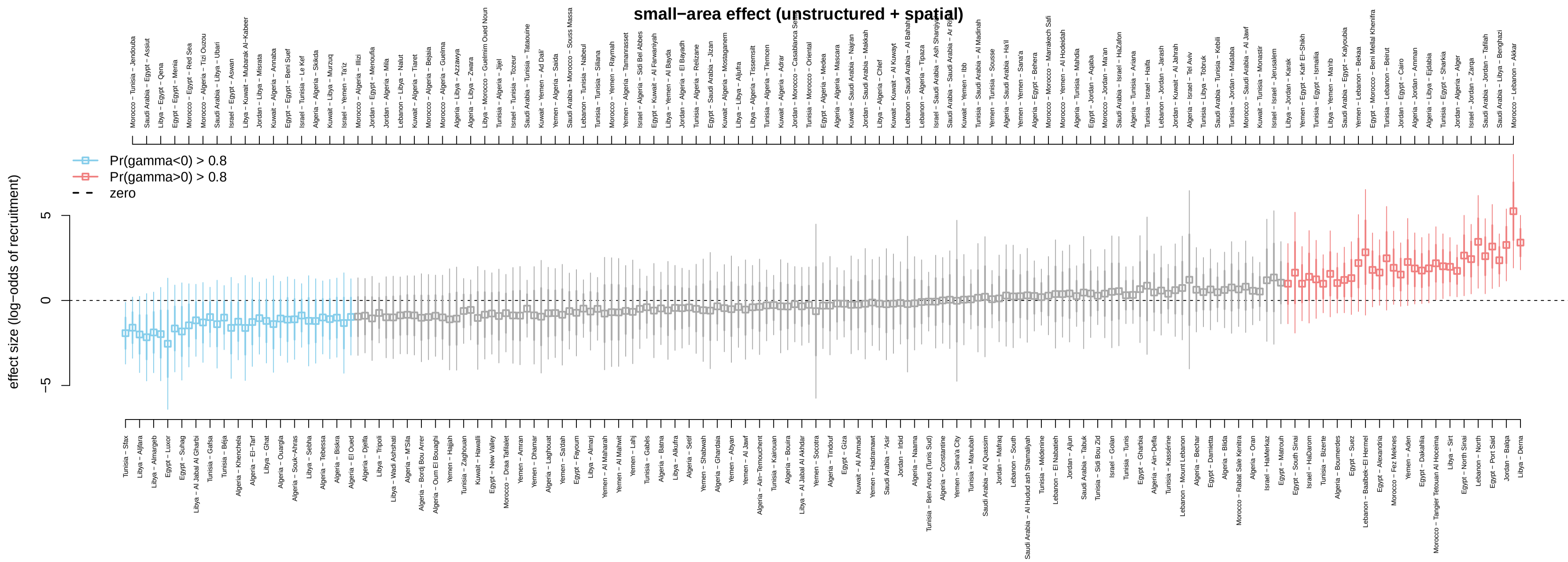
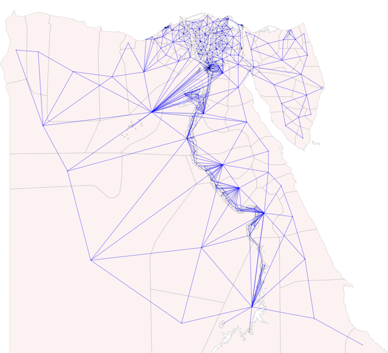


Figure I.5: BYM2 small-area effect γ for the ‘Bird’s Eye’ model.

Figure I.6 presents the Egypt and Tunisia fully-connected graphs used to derive the district-level adjacency matrices fed to the ICAR model. Again, a small number of adjustments were made to connect islands and ensure full-connectivity.



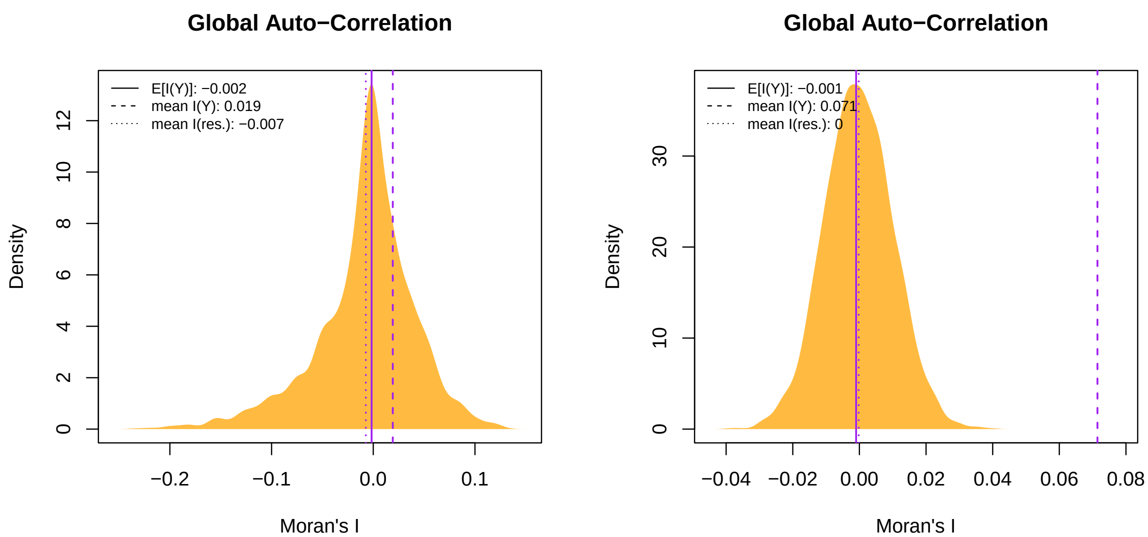
(a)

(b)

Figure I.6: Fully-connected graphs of (a) Egypt and (b) Tunisia at the District level.

The residual plots in Figure I.8, along with the Moran's I presented in Figure I.7, convincingly show we have extracted all spatial variance from the observations: the resulting Moran's Is are distributed around the null-value.

Figures I.9 and I.10 present the spatial distribution of point-estimates for the District and Governorate effects of Egypt and Tunisia respectively. The spatial distribution for Egypt indicates a substantially heightened propensity of recruitment in northeastern regions. No similar pattern is evident in Tunisia, though the mid-eastern costal areas do display systematically lower spatial recruitment effects than the rest of the country. Both countries estimate a number of highly significant district-level effects, which account for large portions of the variance in recruitment of both countries, with highly significant effects ranging from -5 to $+5$ log-odds points . In Tunisia, we also find evidence of a negative Sfax effect. Clearly, in order to be a recruit you must be subjected to unobserved area-level heterogeneity; individual-level covariates alone cannot counteract the underlying rarity of the event, as highlighted by the intercepts.



(a) (b)

Figure I.7: Posterior distribution of Moran's I for Egypt (a) and Tunisia (b). The adjacency matrices implied by Figure I.6 are used as the weight matrices.

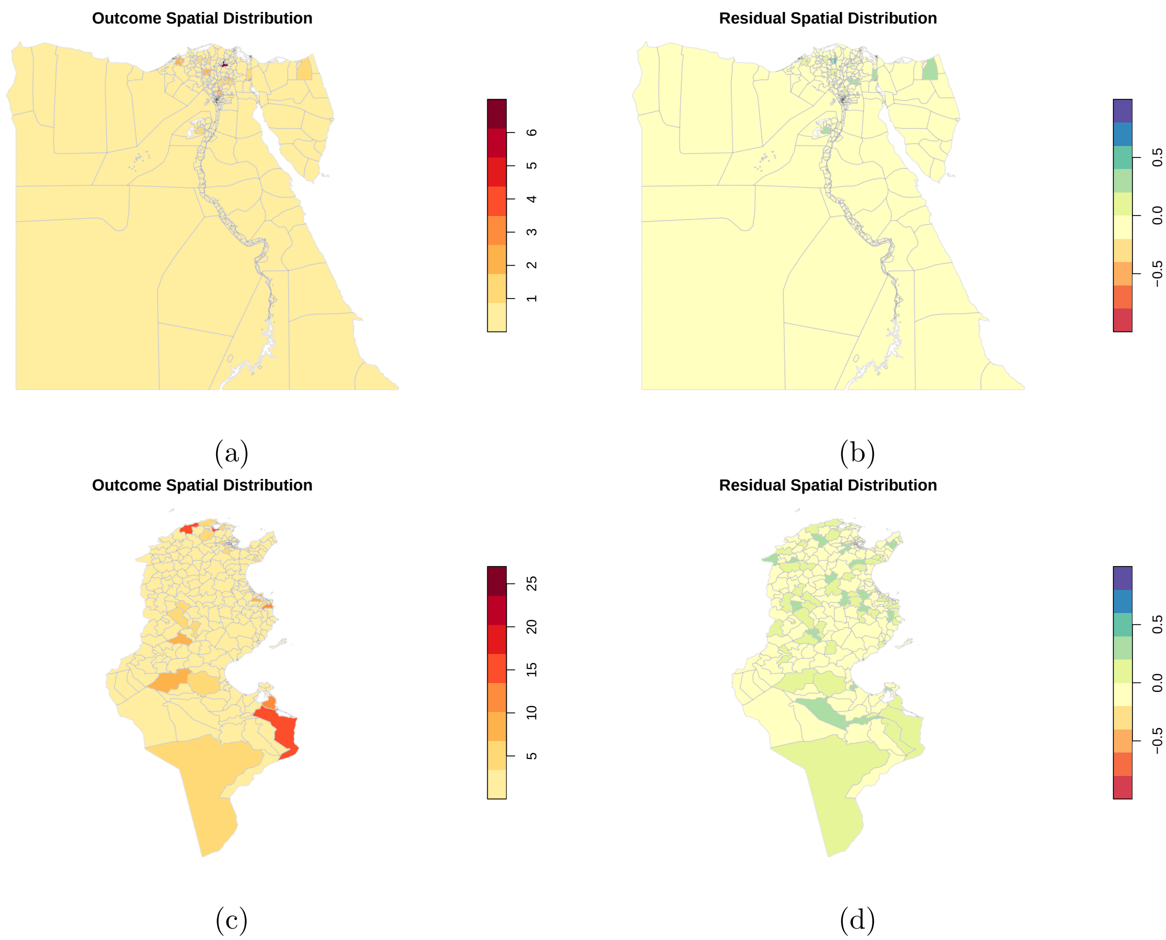


Figure I.8: Spatial distribution of Egyptian observations (a) and residuals (b); Tunisian observations (c) and residuals (d) at the District level. (a) and (c) present the spatial distribution of recruits.

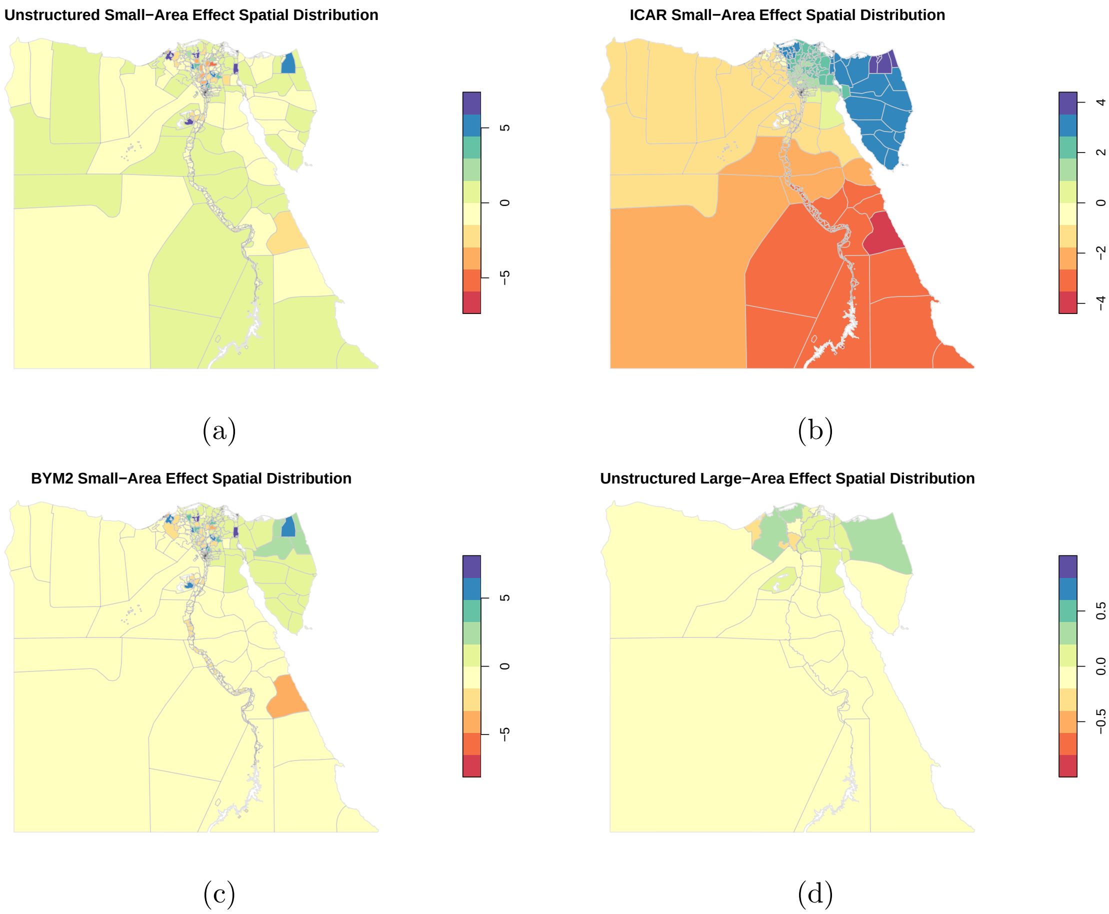


Figure I.9: Egypt's Spatial distribution of: (a) the unstructured Governorate-level effect - ϕ; (b) the spatial Governorate level effect - ψ; (c) the total Governorate effect - $\gamma = \sigma(\phi\sqrt{1-\lambda}) + \psi\sqrt{\lambda/s})$; (d) the unstructured Country effect - η.

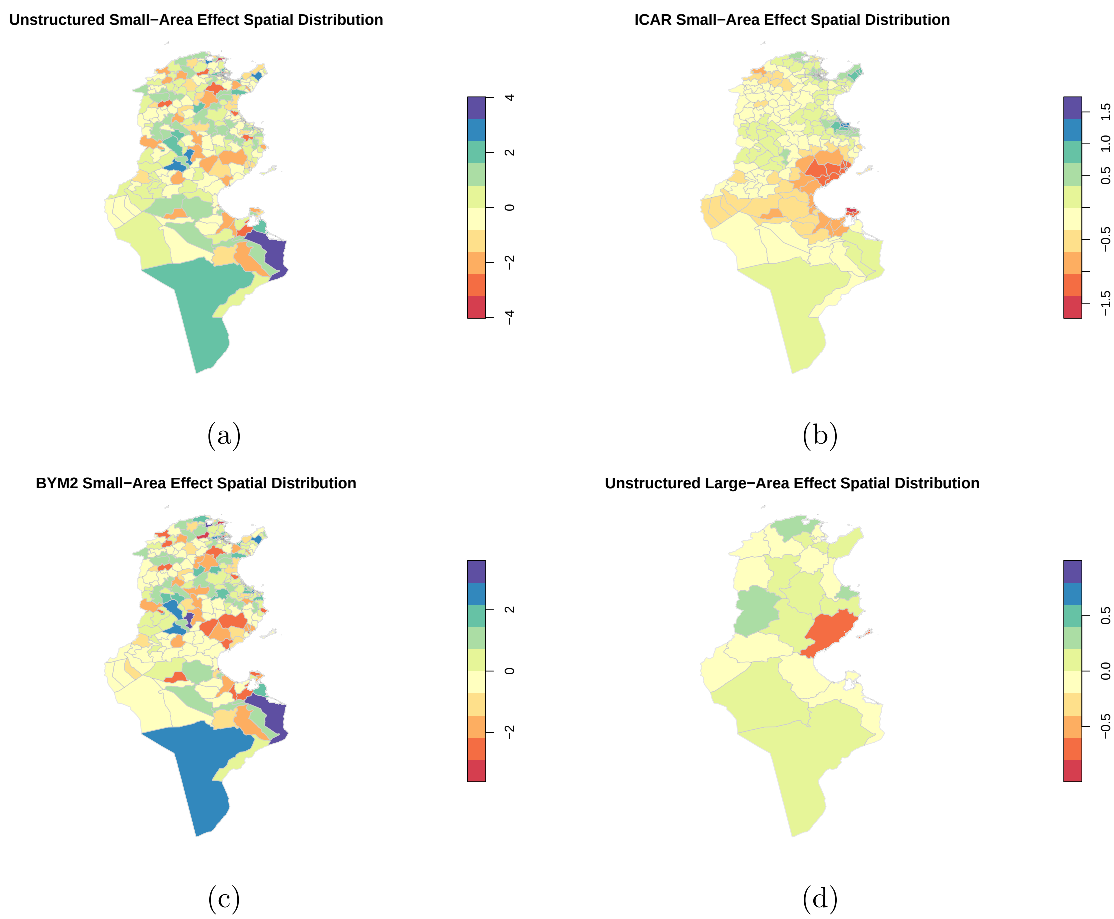
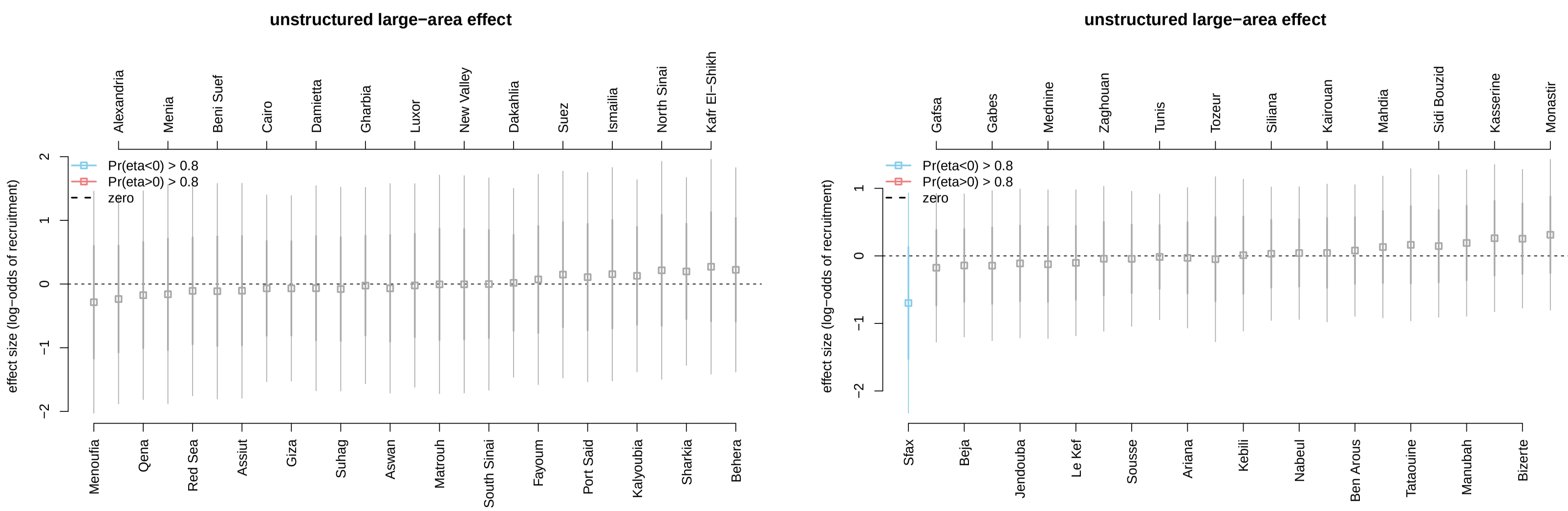


Figure I.10: Tunisia's Spatial distribution of: (a) the unstructured Governorate-level effect - ϕ; (b) the spatial Governorate level effect - ψ; (c) the total Governorate effect - $\gamma = \sigma(\phi\sqrt{1-\lambda}) + \psi\sqrt{\lambda/s})$; (d) the unstructured Country effect - η.



(a) Egypt

(b) Tunisia

Figure I.11: Governorate effect (η) ordered by proportion of posterior simulations above zero.

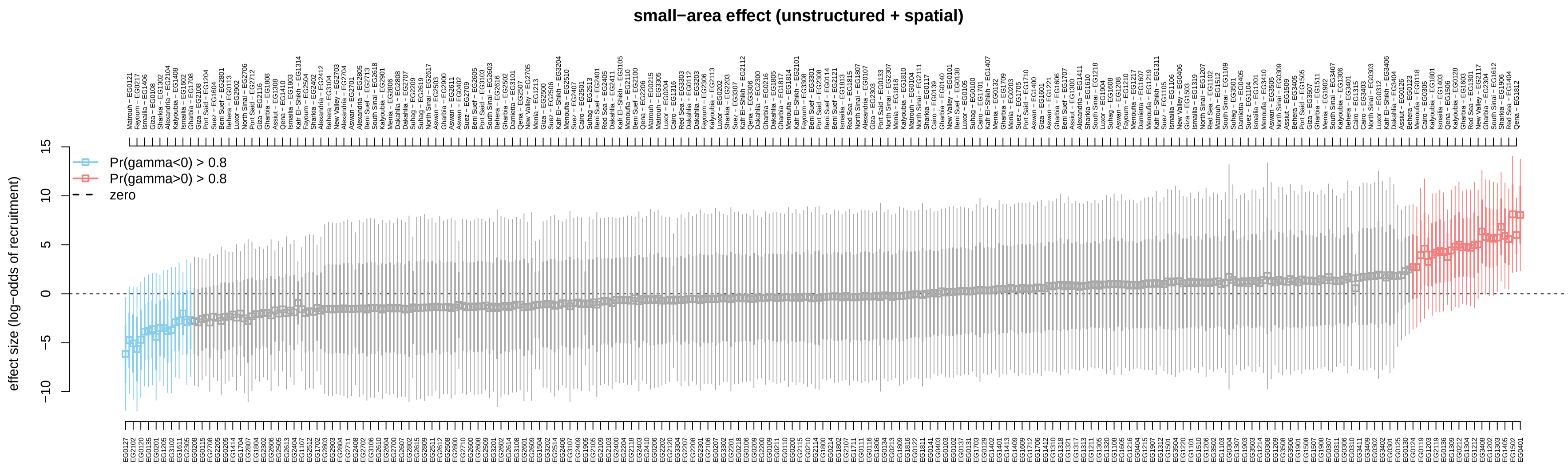


Figure I.12: Total (structured + spatial) residual District effects in Egypt, ordered by proportion of posterior simulations above zero.

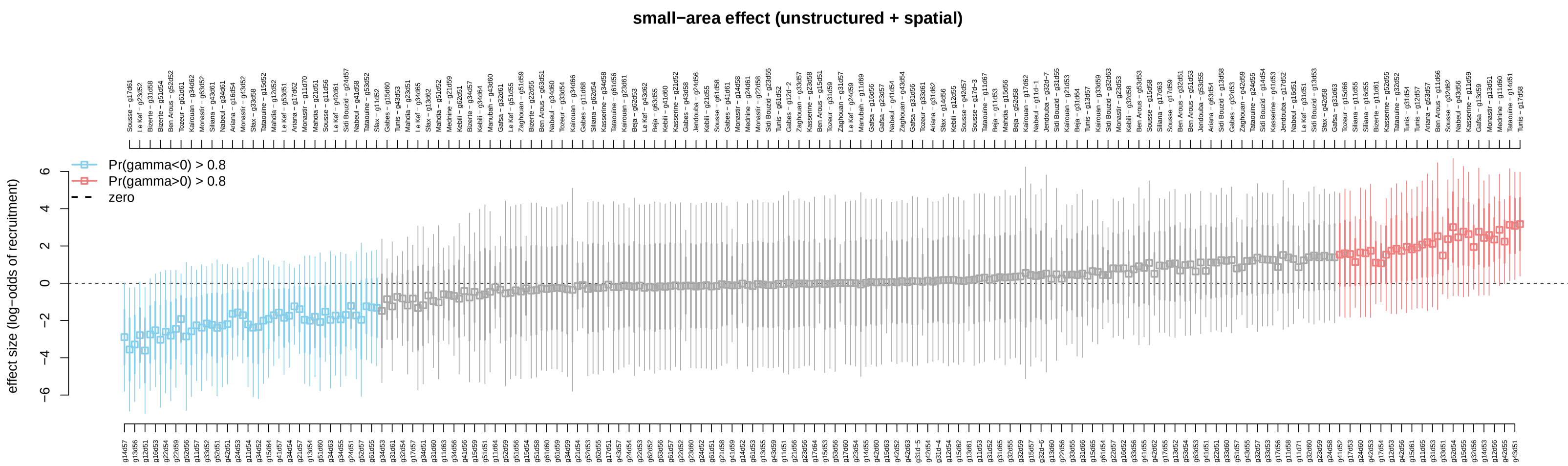


Figure I.13: Total (structured + spatial) residual District effects in Tunisia, ordered by proportion of posterior simulations above zero.

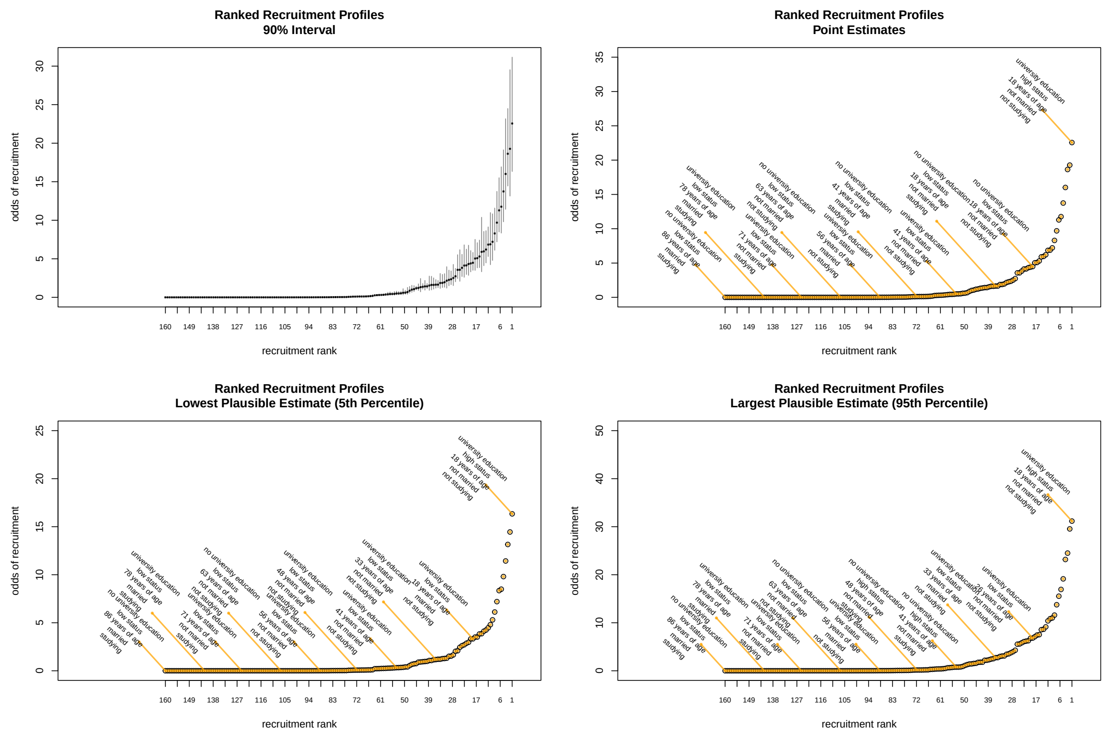


Figure I.14: Bird's Eye Distribution of predicted probabilities across a variety of hypothetical profiles. The distribution is presented on the odds relative to the average profile. To aid with interpretation, minimal and maximal estimates are presented separately. These plots help showcasing the sharp non-linearity across profile's recruitment propensities.

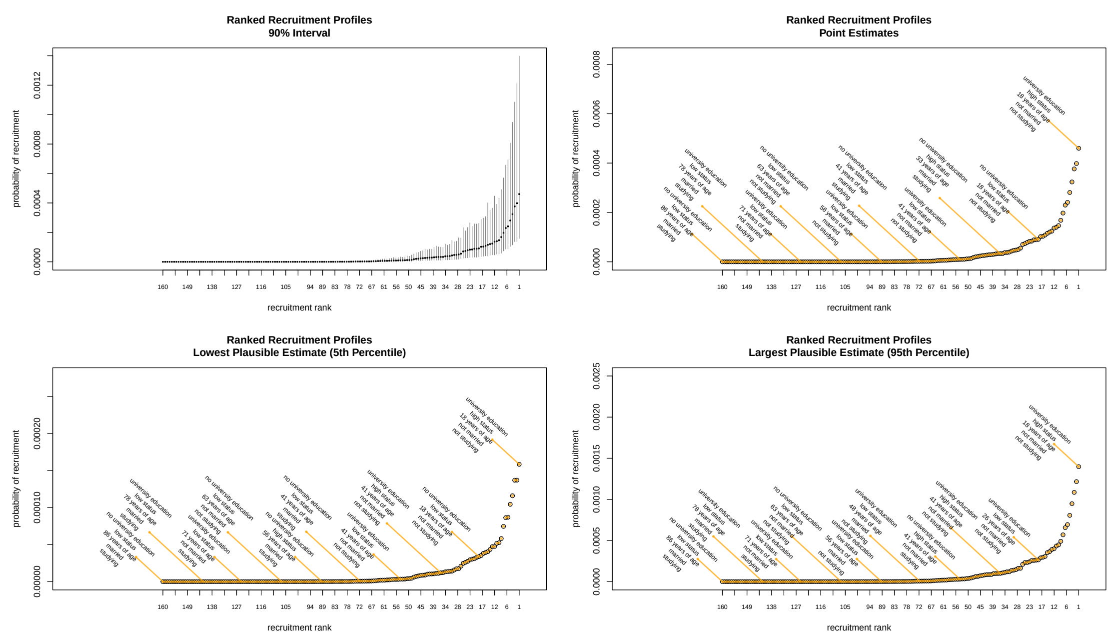


Figure I.15: Bird's Eye Distribution of predicted probabilities across a variety of hypothetical profiles. The distribution is presented on the probability scale. To aid with interpretation, minimal and maximal estimates are presented separately. These plots help showcasing the sharp non-linearity across profile's recruitment propensities.

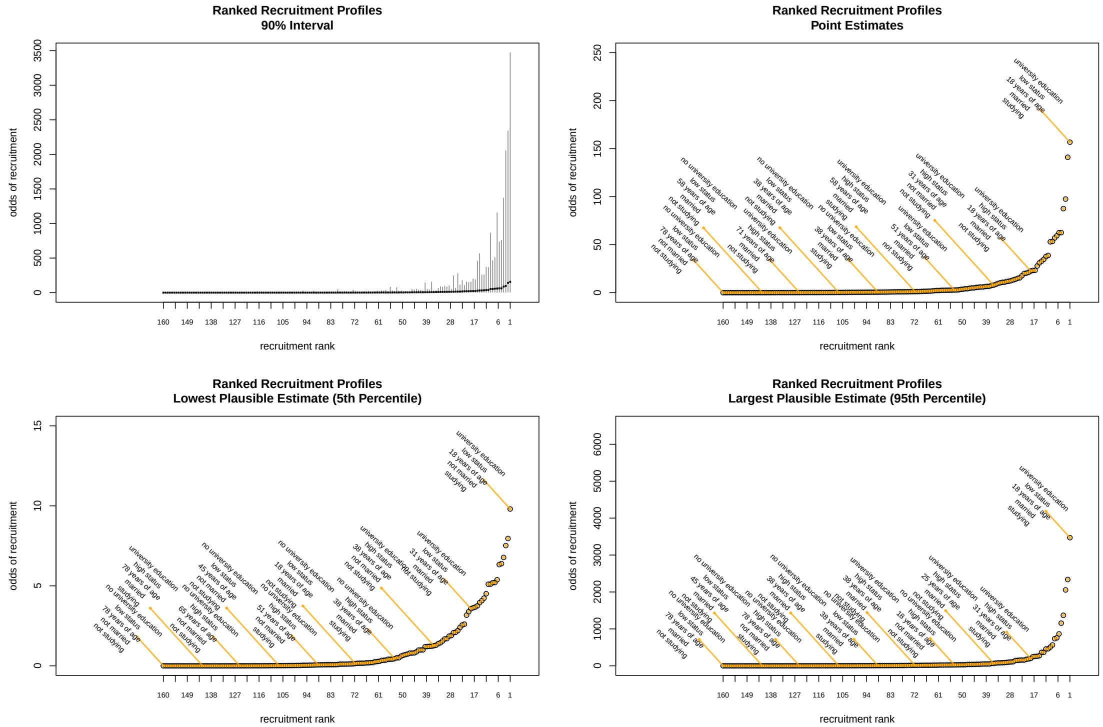


Figure I.16: Worm's Eye (Egypt) distribution of the predicted probabilities across a variety of hypothetical profiles. The distribution is presented on the odds relative to the average profile. To aid with interpretation, minimal and maximal estimates are presented separately. These plots help showcasing the sharp non-linearity across profile's recruitment propensities.

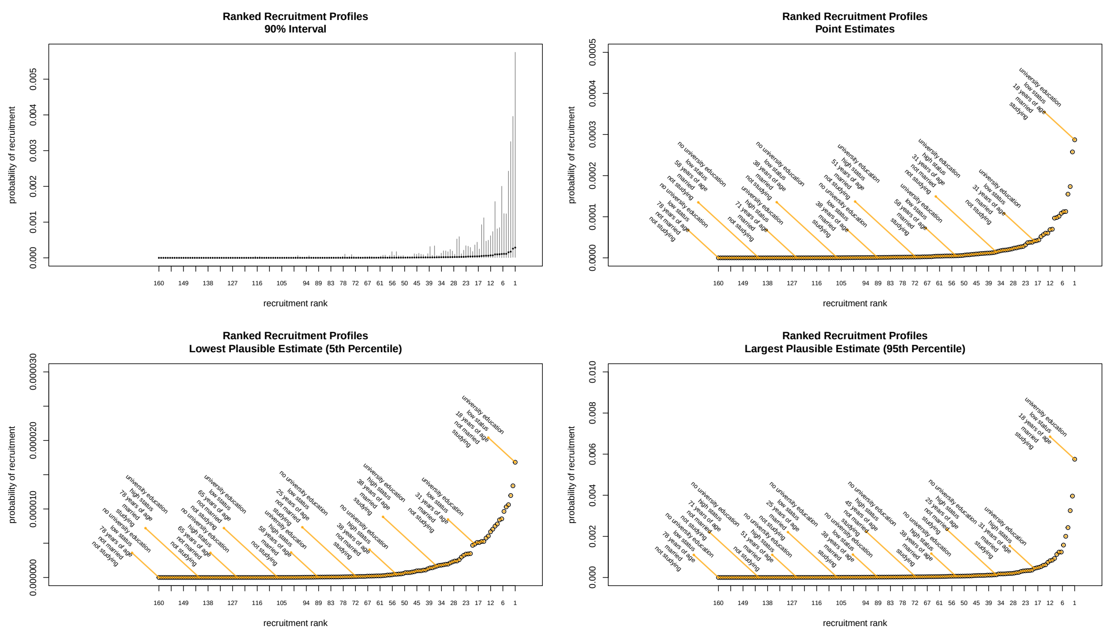


Figure I.17: Worm's Eye (Egypt) distribution of the predicted probabilities across a variety of hypothetical profiles. The distribution is presented on the probability scale. To aid with interpretation, minimal and maximal estimates are presented separately. These plots help showcasing the sharp non-linearity across profile's recruitment propensities.

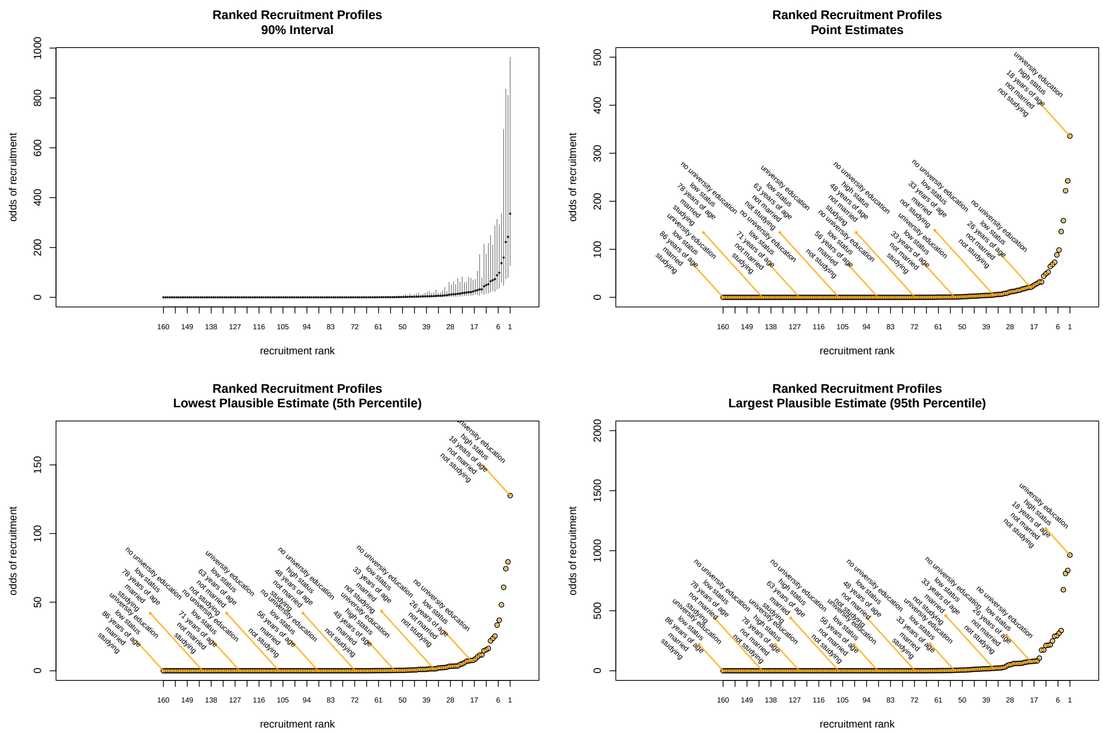


Figure I.18: Worm's Eye (Tunisia) distribution of the predicted probabilities across a variety of hypothetical profiles. The distribution is presented on the odds relative to the average profile. To aid with interpretation, minimal and maximal estimates are presented separately. These plots help showcasing the sharp non-linearity across profile's recruitment propensities.

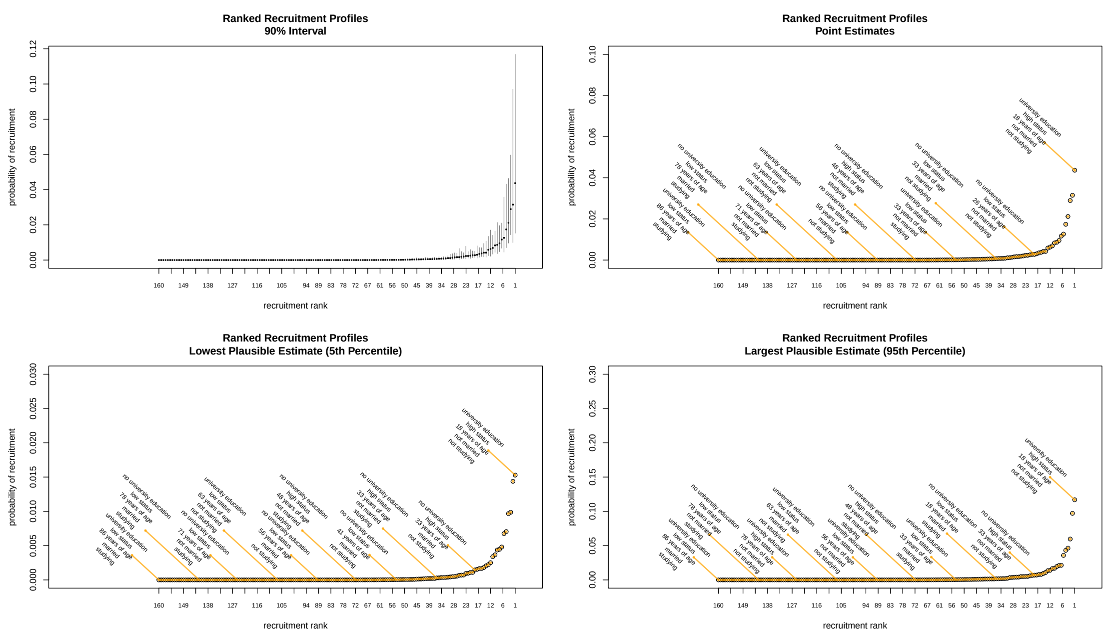


Figure I.19: Worm's Eye (Tunisia) distribution of the predicted probabilities across a variety of hypothetical profiles. The distribution is presented on the probability scale. To aid with interpretation, minimal and maximal estimates are presented separately. These plots help showcasing the sharp non-linearity across profile's recruitment propensities.