Discussion Paper

Hierarchical Bayesian bivariate Fay-Herriot model for estimating domain discontinuities

Jan van den Brakel
Harm Jan Boonstra

November 13, 2018

The design of long-standing repeated surveys is usually kept unchanged as long as possible with the purpose to construct consistent series. In a proper redesign of this kind of surveys, the old and the new design are conducted in parallel for some period of time to quantify the discontinuities, which are caused by the modifications in the survey process. However, there is often only limited budget or field work capacity available for parallel data collection, which hampers the application of direct estimators for estimating domain discontinuities. In this paper the direct estimates obtained under the regular and alternative approach are modelled with a bivariate hierarchical Bayesian Fay-Herriot model to obtain more precise predictions for domain discontinuities. This method is compared with a univariate Fay-Herriot model where the direct estimates under the regular approach are used as covariates in a Fay-Herriot model for the alternative approach conducted on a reduced sample size. The methods are applied to a redesign of the Dutch Crime Victimization Survey.

1 Introduction

Official statistics produced by national statistical institutes are generally based on repeated sample surveys. Much of their value lies in their continuity, enabling developments in society and the economy to be monitored, and policy actions decided. Survey samples contain besides sampling errors different sources of non-sampling errors that have a systematic effect on the outcomes of a survey. As long as the survey process is kept constant, this bias component is not visible. This is often an argument to keep survey processes of repeated surveys unchanged as long as possible. From time to time changes in surveys are needed to improve the efficiency, reduce the survey related costs, or meet new requirements, and this is seen strongly in the use of mixed-mode surveys including web-based questionnaires in official statistics. A redesign of the survey process generally has systematic effects on the survey estimates, since the biases induced by the aforementioned non-sampling errors are changed, disturbing comparability with figures published in the past.

Systematic differences in the outcomes of a repeated survey due to redesign of the survey process are called discontinuities. To avoid the implementation of a new survey process disturbing the comparability of estimates over time, it is important to quantify these discontinuities. This avoids confounding real change in the parameters of interest with changing measurement bias due to alteration in the survey process.

Several methods to quantify discontinuities are proposed in the literature (van den Brakel et al., 2008). A reliable and straightforward approach is to conduct the old and new approach alongside of each other at the same time for some period of time. Ideally this is based on a randomized experiment that can be embedded in the probability sample of the survey (van den Brakel, 2008). In this paper we consider the situation where the regular survey, used for the production of official figures, is conducted at the full sample size and is conducted in parallel with an alternative approach. Due to budget limitiations, the sample that is assigned to the alternative approach is often not sufficiently large to observe minimum detectable differences at prespecified significance and power levels using standard direct esitmators.

To obtain more precise domain estimates for the alternative approach, van den Brakel et al. (2016) proposed an hierarchical Bayesian univariate Fay-Herriot (FH) model (Fay and Herriot, 1979), where sample estimates of the regular survey are considered as potential auxiliary variables in a model selection procedure. This results in an area level model with measurement error (Ybarra and Lohr, 2008). The use of reliable direct estimates observed in the regular survey significantly increased the precision of the domain estimates for the alternative approach conducted at reduced sample size.

In the approach followed by van den Brakel et al. (2016), point estimates for discontinuities are obtained as the difference between the direct estimate obtained with the regular survey and the model based domain prediction obtained with the small sample assigned to the alternative approach. The use of direct estimates from the regular survey as auxiliary variables in the small domain predictions of the regular survey results in strong positive correlations between both estimators, which cannot be ignored in the standard errors for the discontinuities. To this end, two analytic approximations for the standard errors of the discontinuities are proposed. The first approach combines the design-based variance of the direct estimator of the regular survey with the posterior variance of the hierarchical Bayesian domain predictions of the alternative survey and a design-based estimator for the covariance between both point estimates. This approach is unstable in the sense that even negative variance estimates occur in the case of strong positive covariance estimates. A related issue is that design-based and model-based variance approximations are combined in one uncertainty measure for the discontinuities. Therefore a second analytic approximation was proposed, where a design-based estimator for the variance of the HB domain predictions is derived and combined with the design-based variance for the direct estimator for the regular survey and the design-based covariance between both point estimates.

The complications with variance estimation of domain discontinuities under a univariate FH model can also be circumvented by setting up a full Bayesian framework for the analysis of the domain discontinuities. Therefore this paper proposes a bivariate FH model to model the direct estimates under the regular and alternative approach simultaneously. The random component of this model accounts for the correlation between the domain parameters under the regular and alternative approach. The precision of the estimated discontinuities is improved by increasing the effective sample size within the domains with cross-sectional correlations. In addition a positive correlation between the random domain effects further decreases the standard error of the estimated discontinuities. The bivariate FH model is applied to a parallel run conducted to quantify domain discontinuities in the Dutch Crime Victimization Survey. The results are compared with the univariate FH model proposed in van den Brakel et al. (2016).

The FH model (Fay and Herriot, 1979) is frequently applied in the context of small area estimation (Rao and Molina, 2015). FH models are particularly appropriate if auxiliary information is available at the domain level. This is typically the case with a parallel run where reliable direct domain estimates from the regular survey are considered as auxiliary variables in models for small domain predictions for the alternative approach conducted at a reduced sample size or direct estimates from both surveys are combined in one multivariate model. Datta et al. (1996) employed a multivariate FH model fitted in in an HB framework to estimate median income. Multivariate FH models fitted in a frequentist framework are considered in Gonzales-Manteiga et al. (2008); Benavent and Morales (2016). Several authors provided time-series FH models to use sample infromation from previous editions of a survey as a form of small area estimation (Rao and Yu, 1994; Datta et al., 1999; You and Rao, 2000; Estaban et al., 2012; Marhuenda et al., 2013). Pfeffermann and Burck (1990); Pfeffermann and Tiller (2006); van den Brakel and Krieg (2016); Bollineni-Balabay et al. (2017) are some examples of FH time series models casted in a state-space framework.

The paper is structured as follows. In Section 2 the Crime Vicitimization Survey, the redesign and the set up of the parallel run are described. The bivariate FH model is explained in Section 3, including the hierarchical Bayesian framework and the model selection and evaluation approach. Results are presented in Section 4. The paper ends with a discussion in Section 5.

2 The Crime Victimization Survey

The Dutch crime victimization survey (CVS) is a long standing survey conducted by Statistics Netherlands at an annual frequency with the purpose to publish reliable figures about crime rates, safety feelings, and satisfaction about police performance in the Netherlands. The CVS is designed to provide reliable figures at the national level and at the level of police districts, which is a subdivision of the Netherlands in 25 regions. The CVS is based on a stratified simple random sampling design for people aged 15 years or older residing in the Netherlands. Strata are formed by police regions to control the precsion of these planned domain estimates. The sampling frame is based on the Dutch government's register of all residents in the Netherlands, called Municipal Basis Administration. The yearly sample of the regular CVS is designed such that about 19,000 respondents are observed. The sample is equally divided over the strata, such that about 760 observations are obtained in each stratum. The general regression (GREG) estimator (Särndal et al., 1992) is used to estimate population parameters at the national level and for police districts.

The CVS has undergone several redesigns in the past. This paper focusses on a redesign in 2008 where the data collection changed from a mixed-mode design via computer-assisted personal interviewing (CAPI) and computer-assisted telephone interviewing (CATI) to a sequential mixed-mode design that starts with web interviewing (WI) and a follow-up for nonrespondents with CAPI and CATI. In addition the questionnaire is changed to improve the wording as well as the order of the questions. To maintain uninterrupted series, discontinuities induced by this redesign need to be quantified. To this end the regular survey used for official publication purposes was conducted in parallel with the alternative survey approach with a sample size of about 6000 respondents. In this application, the regular approach was based on the new survey design using WI, CATI and CAPI and the alternative approach was based on the old design using CAPI and CATI data collection only.

The sample design for the parallel run is based on stratified simple random sampling where police districts are the strata, using proportional allocation. This results in a sample design that is optimal to estimate figures at the national level but suboptimal for domain estimation.

This survey reports on many different outcome variables. In the present study five key survey variables are considered, see Table 2.1. Estimates for these variables at the national level under the regular and alternative survey are specified in Table 2.2. The sample size allocated to the alternative approach is sufficiently large to estimate discontinuites at the national level using the GREG estimator but insufficient to estimate discontinuities at the domain level of the 25 police districts. The direct estimates for the disontinuities at the national level are indeed signifcantly different from zero, contrary to the average of the direct domain estimates. To obtain as precise as possible predictions for domain discontinuities a model-based small area estimation method based on area level models (Fay and Herriot, 1979) is proposed in the next paragraph.

variable	description
nuisance	perceived nuisance in the neighborhood on a ten point scale; this includes nuisance by drunk people, neigbours, or groups of youngsters, harassment, and drug related problems
unsafe	percentage of people feeling unsafe at times
propvict	percentage of people saying to have been victim to property crime in the last 12 months
offtot	total number of offenses per 100 people
satispol	percentage of people satisfied with police at their last contact (if contact in last 12 months)

Table 2.1 Five key NSM survey variables considered in the present study.

variable	regular		alternative		$\hat{\Delta}$	
	Average over 25 police districts					
offtot	42.29	(4.73)	33.28	(5.73)	9.01	(7.69)
unsafe	24.38	(2.03)	19.86	(2.87)	4.52	(3.57)
nuisance	1.61	(0.11)	1.28	(0.13)	0.33	(0.17)
satispol	60.61	(4.23)	55.58	(6.88)	5.04	(8.21)
propvict	12.55	(1.60)	9.78	(2.19)	2.78	(2.77)
	National level					
offtot	43.79	(1.07)	34.09	(1.04)	9.70	(1.49)
unsafe	25.07	(0.44)	20.48	(0.52)	4.59	(0.68)
nuisance	1.67	(0.02)	1.34	(0.02)	0.33	(0.03)
satispol	59.88	(0.92)	55.10	(1.25)	4.78	(1.55)
propvict	13.02	(0.36)	10.32	(0.39)	2.70	(0.53)

Table 2.2 GREG estimates regular and alternative survey approach averaged over districts and national level. Standard errors between brackets.

3 Methods

3.1 Small Area Estimation for domain discontinuities

Testing hypotheses about differences between estimates of a finite population parameter observed under different survey processes implies the existence of measurement errors. Therefore a measurement error model is required to explain systematic differences between survey estimates for the same population parameter observed under two different survey approaches. Let θ_i denote the population parameter of domain $i = 1, \dots, m$. Let y_i^r denote the observed value for θ_i in the case of a complete enumeration under the regular approach. In a similar way y_i^a denotes the observed value for θ_i in the case of a complete enumeration under the alternative approach. Direct estimates for y_i^r and y_i^a are obtained with GREG estimater based on two seperated samples and are denoted as $\hat{y}_i^r$ and $\hat{y}_i^a$ respectively.

The relation between the observed values under a complete enumeration and the real population parameter is:

$$y_i^q = \theta_i + \gamma_i^q, \;\; i = 1, \ldots, m, \;\; q = r, a,$$

with γ_i^q the real measurement bias if θ_i is measured with survey approach q. Without any external information, it is not possible to estimate γ_i^q. From the sample data only the relative bias, say $\Delta_i = y_i^r - y_i^a = \gamma_i^r - \gamma_i^a$ is identifyable. Direct estimates for these discontinuities are obtained from the survey data as the contrast between the GREG estimates, i.e. $\hat{\Delta}_i = \hat{y}_i^r - \hat{y}_i^a$.

In the case of the Dutch CVS the sample size of the regular survey is large enough to obtain sufficiently precise direct estimates for the planned domains, since the sample is designed to publish offcial statistics for these domains. The sample assigned to the alternative survey for the parallel run has only a size of one third of the regular sample size, which is insufficient to obtain precise direct estimates for the planned domains. In an earlier paper (van den Brakel et al., 2016) univariate FH models were developed to obtain as precise as possible predictions for the domain parameters observed with the small sample size assigned to the alternative survey approach using auxliary variables derived from three different sources in a step-forward variable selection procedure. The first source contains demographic variables derived from the Municipal Basis Administration (MBA). The second source contains related variables available in the Police Register of Reported Offences (PRRO). The third source, which is unique in the case of a parallel run, contains direct estimates for the same variables observed under the regular survey, which are sufficiently precise at least for the planned domains like police districts. The direct estimates from the regular survey are often selected as auxiliary variables for these univariate FH models. This comes not as a suprice since these are survey estimates for the same population parameters. Although measured with a different survey process, strong positive correlations can be expected. Strong improvements of the precision of small domain prediction are indeed found if the set of potential auxiliary variables, i.e. from MBA and PRRO, is extended with the direct estimates from the regular CVS.

In this application the sampling error in the auxiliary variables that come from the regular CVS can be ignored in FH model, since the sample size and therefore the sampling error for these domains is more or less equal for the domains (Ybarra and Lohr, 2008). This implies that the variance component of the random domain effects is inflated with the sampling error of the auxiliary variables, which is fine as long as the sampling error does not differ between domains. In most applications this is not the case and the methods proposed by Ybarra and Lohr (2008) should be used to account for sampling error in the auxiliary variables.

FH multilevel models can be fitted with frequentist approach using EBLUP or with an hierarchical Baysian (HB) approach (Rao and Molina, 2015). In van den Brakel et al. (2016) the HB approach is preferred above the EBLUP, since the strong auxiliary information in the fixed effect part of the model often results in zero estimates for the variance component of the random domain effects, giving too much weight to synthetic regression part and to little weight too the direct estimates in the EBLUP, (Bell, 1999; Rao and Molina, 2015).

Let $\tilde{y}_i^a$ denote the HB prediction for domain i under the alternative approach. Now domain discontinuites are obtained by $\tilde{\Delta}_i = \hat{y}_i^r - \tilde{y}_i^a$. Using direct estimates of the regular survey as auxiliary variables in the fixed part of the FH model for the alternative survey considerably increases the complexity of the variance estimation for the discontinuities. The variance of $\tilde{\Delta}_i$

can be expressed as $Var(\tilde{\Delta}_i) = Var(\hat{y}_i^r) + MSE(\tilde{y}_i^a) - 2cov(\hat{y}_i^r \tilde{y}_i^a)$. The use of $\hat{y}_i^r$ or related sample estimates as auxiliary variables to predict $\tilde{y}_i^a$, results in non-zero values for $cov(\hat{y}_i^r \tilde{y}_i^a)$ that cannot be ignored. To approximate $Var(\tilde{\Delta}_i)$, van den Brakel et al. (2016) proposed an approximately design-unbiased estimator for $cov(\hat{y}_i^r \tilde{y}_i^a)$ and $Var(\hat{y}_i^r)$, while the $MSE(\tilde{y}_i^a)$ is approximated with the posterior variance of the HB domain predictions. A major disadvantage of this approach is that model-based and design-based uncertainty measures are intertwined. On the one hand, the MSE's for $\tilde{y}_i^a$ are approximated with their posterior variances. On the other hand, the covariances between $\hat{y}_i^r$ and $\tilde{y}_i^a$ are approximated form a design-based perspective. Consequently, naive application of this approach may give negative variance estimates for the discontinuities. This drawback has been solved using a design-based approximation for the $MSE(\tilde{y}_i^a)$, resulting in a full design-based approximation for the uncertainty of the estimated domain discontinuities.

In this paper a full hierarchical Bayesian framework for estimating domain discontinuities is proposed as an alternative by developing a bivariate FH model for the domain parameters observed under both the regular and alternative approach. The advantage of this approach is that it improves the precision of both the direct estimates of the regular and alternative domain estimates by borrowing strenth from other domains and both surveys. Negative variance estimates for the estimated domain discontinuities are precluded by definition under this multivariate HB framework. Another advantage is that a bivariate FH model avoids the complications of accounting for sampling error in the auxilary variables, which is necessary if the survey estimates of the regular survey are used as covariates in univariate FH models and the sampling error differs between domains.

3.2 Bivariate Fay-Herriot model

A bivariate version of the FH model (Fay and Herriot, 1979) starts with a measurement model for the two GREG estimates observed in each domain:

$$\hat{y}_i = y_i + e_i, \quad i = 1, \ldots, m, \tag{1}$$

with $y_i = (y_i^r, y_i^a)^t$, $\hat{y}_i$ a vector containing the GREG estimates of y_i and $e_i = (e_i^r, e_i^a)^t$ a vector with the sampling errors of $\hat{y}_i$ for which it is assumed that

$$e_i \overset{ind}{\sim} N\left(0_2, \Psi_i\right), \quad i = 1, \ldots, m. \tag{2}$$

Here 0_2 is a 2 dimensional column vector with each element equal to zero. Since the sample for the regular and alternative survey are drawn independently, it is assumed that $\Psi_i = Diag(\psi_i^r, \psi_i^a)$ where ψ_i^q is the design variance of $\hat{y}_i^q$. It is also assumed that these design variances are known although they are replaced by their estimates in practice. The true domain parameters are modelled with a multilevel model. For the fixed effects it is assumed that the regular and alternative approach share the same covariates. In the most general case the regression coefficients for the fixed part are different for both variables i.e. $y_i^q = x_i^t \beta^q + v_i^q$, with $x_i = (x_{i1}, \ldots, x_{ip})^t$ a p-vector with covariates of domain i for y_i^q, β^q a p-vector of regression coefficients, which are equal over the domains but might be different between the two survey approaches. It is assumed that $x_{i1} = 1$ corresponds to the intercept. Furthermore

ν_i^q are random domain effects. This gives rise to the following bivariate multilevel model for the two domain parameters:

$$y_i = X_i\beta + \nu_i,\ \ i = 1, \dots, m, \tag{3}$$

where $X_i = I_2 \otimes x_i^t$, $\beta = (\beta^{r^t} \beta^{a^t})^t$, I_2 a 2 dimensional identity matrix, and $\nu_i = (\nu_i^r, \nu_i^a)^t$. For the random domain effects it is assumed that

$$\nu_i \overset{IID}{\sim} N(0_2, \Sigma),\ \ i = 1, \dots, m, \tag{4}$$

with Σ a general 2×2 covariance matrix for the random domain effects. Inserting (3) into (1) gives:

$$\hat{y}_i = X_i\beta + \nu_i + e_i,\ \ i = 1, \dots, m, \tag{5}$$

with model assumptions (2) and (4).

Since the number of domains in this application is small, it is important to select parsimonious models. One way to reduce model complexity is to assume that the regression coefficients are equal for both survey approaches. In this case a dummy indicator, say δ_i, is required wich is equal to zero for the regular survey and equal to one for the alternative survey. In this case $x_{i1} = 1$ corresponds to the overall intercept and $x_{i2} = \delta_i$ is the indicator whose coefficient measures the differences between intercepts of the variables observed under both surveys. So $y_i^q = x_i^t\beta + \nu_i^q$ and in (3) $X_i = x_i^t$, and β a vector with the corresponding regression coefficients. Two versions for the fixed effects are considered:

- FE_uq: A fixed effect model where the regular and alternative approach share the same covariates, but have different regression coefficients. In this case, domain discontinuities are given by
$$\Delta_i = \sum_{j=1}^{p} x_{i,j}(\beta_j^r - \beta_j^a) + (\nu_i^r - \nu_i^a). \tag{6}$$
- FE_eq: A more parsimoneous version for the fixed effect component by assuming that the regression coefficients are equal for the regular and alternative approach. In this case domain discontinuities are given by
$$\Delta_i = -\beta_2 + (\nu_i^r - \nu_i^a), \tag{7}$$
with β_2 the regression coefficient for $x_{i2} = \delta_i$.

The following covariance structures for the random domain effects are considered:

- RE_f: A full covariance matrix for the random domain effects, i.e.:
$$\Sigma = \begin{pmatrix} \sigma_r^2 & \rho\sigma_r\sigma_a \\ \rho\sigma_r\sigma_a & \sigma_a^2 \end{pmatrix}$$
Positive correlation between the random domain effects will further increase the precision of the estimates for the domian discontinuities since domain estimates borrow strenght not only from different domains but also accross the two surveys.

– RE_d: A diagonal covariance matrix with separate variances for the regular and alternative approach, i.e.: $\Sigma = Diag(\sigma_r^2, \sigma_a^2)$. This covariance structure in combination with model FE_uq comes down to applying a univariate FH model to both surveys separately. In this case models only use sample information from other domains within the same survey but not accross the two surveys to improve the precision of the estimates for domain discontinuities.
– RE_s: A diagonal covariance matrix with equal variances for the regular and alternative approach, i.e.: $\Sigma = \sigma^2 I_2$.

3.3 Estimation of the bivariate Fay-Herriot model

The model developed in Subsection 3.2 is fitted with a hierarchical Bayesian approach using Markov Chain Monte Carlo (MCMC) sampling. In particular the Gibbs sampler is used. Therefore the following priors are used for the model parameters and hyperparameters.

For the regression coefficients uniform improper priors are assumed:

$$\beta \sim 1. \tag{8}$$

For the random domain effects a redundant multiplicative parametrization is used, since this improves the convergence of the Gibbs sampler (Gelman et al., 2008) and yields more robust prior distributions for the variance parameters (Gelman, 2006). Let $\nu = (\nu_1^r, \nu_1^a, \dots, \nu_m^r, \nu_m^a)^t$ denote a vector with random effects. The random domain effects are defined as $\nu = \Delta_\xi \tilde{\nu}$, $\Delta_\xi = I_m \otimes Diag(\xi)$, and ξ a column vector of length 2 with scale parameters for the random domain effects. In the case of a full covariance matrix or a diagonal covariance matrix with unequal variance components, the random domain effects are scaled with two separate scale parameters $\xi = (\xi_r, \xi_a)^t$. In the case of a diagonal covariance matrix with equal variance components, only one scale parameter is used, i.e. $\xi = \xi(1,1)^t$. In the case of a full covariance matrix or a diagonal covariance matrix with unequal variance components, the prior on ξ is

$$\xi \quad \simeq \quad N(0_2, I_2). \tag{9}$$

In the case of a diagonal covariance matrix with equal variance components,

$$\xi \quad \simeq \quad N(0,1). \tag{10}$$

The prior on $\tilde{\nu}$ is multivariate normal;

$$\tilde{\nu}|\tilde{\Sigma} \quad \simeq \quad N(0_{2m}, I_m \otimes \tilde{\Sigma}). \tag{11}$$

In the case of a full covariance matrix for the random domain effects, the prior for $\tilde{\Sigma}$ is an inverse Wishart distribution:

$$\tilde{\Sigma} \quad \simeq \quad Inv - Wish(\nu_\nu, \Phi_\nu), \tag{12}$$

with $v_v = d + 1$ degrees of freedom, with d the dimension of $\tilde{\Sigma}$ which is equal to 2 in this application, and covariance matrix $\Phi_v = I_2$ by default. This comes down to a scaled inverse Wishart distribution for $\Sigma = \Delta_\xi \tilde{\Sigma} \Delta_\xi$, (O'Malley and Zaslavsky, 2008). In the case of a diagonal covariance matrix for the random domain effects, $\tilde{\Sigma} = Diag(\tilde{\sigma}_r^2, \tilde{\sigma}_a^2)$, the priors for $\tilde{\sigma}_q^2$ are independent inverse chi-squared distributions,

$$\tilde{\sigma}_q^2 \quad \simeq \quad Inv - \chi^2(v_q, S_q^2), \tag{13}$$

with v_q degrees of freedom equal to 1 and scale parameter S_q^2 equal to 1. In the case of a diagonal covariance matrix for the random domain effects with equal variances, $\tilde{\Sigma} = \tilde{\sigma}^2 I_2$ with the prior for $\tilde{\sigma}^2$ an inverse chi-squared distribution,

$$\tilde{\sigma}^2 \quad \simeq \quad Inv - \chi^2(v, S^2), \tag{14}$$

with v and S^2 equal to 1. Marginally the priors for $\sigma_q = \sqrt{\xi_q}\tilde{\sigma}_q$ and $\sigma = \sqrt{\xi}\tilde{\sigma}$ are half-Cauchy distributions with scale parameter S_q or S. The inverse chi-squared distributon is a more common prior for variance parameters. These priors, however, might be informative, even in the case of small scale and shape parameters. In addition convergence problems might occur with the Gibbs sampler. Both problems are largely avoided with the above described redundant multiplicative parametrization of the random effects, which is proposed by Gelman (2006); Gelman et al. (2008); Polson and Scott (2012).

All parameters are collected in a vector $\theta = (\tilde{v}^t, \xi^t, \beta, \tilde{\sigma}_r^2, \tilde{\sigma}_a^2, \rho)^t$. Let $\hat{y}$ denote the $2m$ column vector obtained by stacking the m column vectors $\hat{y}_i$, and X the matrix obtained by stacking the matrices X_i. In the case of unequal regression coefficients, X is a $2m \times 2p$ matrix. In the case of equal regression coefficients, X is a $2m \times p$ matrix. The likelihood function can be written as

$$p(\hat{y}|\theta) \quad \simeq \quad N(X\beta + \Delta_\xi \tilde{v}, \Psi), \tag{15}$$

with $\Psi = \bigoplus_{i=1}^m \Psi_i$ a $2m \times 2m$ diagonal matrix with the design variances of the direct estimates $\hat{y}$. The joint prior distribution of θ, say $p(\theta)$ equals the product of the priors assumed for $\tilde{v}$ in (11), ξ in (9) or (10), and $\tilde{\sigma}_r^2, \tilde{\sigma}_a^2$, and ρ in either (12), (13), or (14) The posterior distribution of θ is proportional to the joint density, i.e. $p(\theta|\hat{y}) \propto p(\theta)p(\hat{y}|\theta)$. The model is fitted using the Gibbs sampler Geman and Geman (1984); Gelfand and Smith (1990). The full conditional distributions used in the Gibbs sampler are specified in Appendix B.

For each model considered, the Gibbs sampler is run in three independent chains with randomly generated starting values. The length of each chain after the burn in period for each run is 10,000 iterations. This gives 30,000 draws to compute estimates and standard errors. The convergence of the MCMC simulation is assessed using trace and autocorrelation plots as well as the Gelman-Rubin potential scale reduction factor (Gelman and Rubin, 1992), which diagnoses the mixing of the chains. The diagnostics suggest that all chains converge well within 500 draws. The estimated Monte Carlo simulation errors are small compared to the posterior standard errors for all parameters, so that the number of draws are sufficient for our purposes.

The estimands of interest are expressed as functions of the parameters, and applying these functions to the MCMC output for the parameters results in draws from the posteriors for these estimands. Domain predictions for the target variables under the bivariate FH model are obtained as the posterior means from the Gibbs sampler output and are denoted as $\tilde{y}_i^{q,bFH}$. Domain predictions for the discontinuities are obtained as the posterior means of (6) or (7) from the Gibbs sampler output and are denoted as $\tilde{\Delta}_i^{bFH}$. Mean squared erors for $\tilde{y}_i^{q,bFH}$ and $\tilde{\Delta}_i^{bFH}$ are approximated with the posterior variance from the Gibbs sampler output.

The methods are implemented in R using the *mcmcsae* R-package (Boonstra, 2016).

3.4 Pooling design variances

Estimates for the design variances ψ_i^r and ψ_i^a are available from the GREG estimator and are used as if the true design variances are known. This is a standard assumption in small area estimation. Therefore it is important to provide reliable estimates for these design variances. For the regular survey the GREG estimates for the design variances are considered to be reliable enough to be used in the FH model. For the alternative survey the GREG estimates for the design variances are unreliable and therefore smoothed to improve the stability of the estimates of ψ_i^a. Under the assumption that the population variances of the GREG residuals under the alternative approach are equal accross domains, the analysis-of-variance type of pooled variance estimator is used:

$$\psi_i^a = \frac{1-f_i^a}{n_i^a}\frac{1}{n^a-m}\sum_{i=1}^{m} S_{i;GREG}^{a^2}$$

with f_i^a the sample fraction in domain i of the alternative survey, n_i^a the number of respondents in domain i under the alternative survey, $n^a = \sum_{i=1}^{m} n_i^a$, and $S_{i;GREG}^{a^2}$ the estimated population variance of the GREG residuals.

3.5 Model selection and evaluation

Frequently applied model selection criteria in hierarchical Bayesian settings are the Widely Applicable Information Criterion or Watanabe-Akaike Information Criteria (WAIC) (Watanabe, 2010, 2013) and the Deviance Information Criteria (DIC) (Spiegelhalter et al., 2002). They are popular because they are easy to compute from MCMC simulation output and because of their ability to make a reasonable tradeoff between model fit and model complexity. The WAIC is seen as an improvement on the DIC since the latter can produce negative estimates for the effective number of parameters and it is not defined for singular models (Vehtari et al., 2017). The WAIC is defined as $WAIC = -2(log\, p(\hat{y}|\theta) - p_{eff})$, with $p(\hat{y}|\theta)$ the likelihood (15) and estimated using the pointwise predictive density (Vehtari et al. (2017), equation (3)) and p_{eff} the effective number of model parameters estimated using the posterior variance of the log pointwise predictive density (Vehtari et al. (2017), equation (13)). The p_{eff} is used as a penalty for model complexity and is closely related to the effective number of parameters proposed by Hodges and Sargent (2001) for linear multilevel models where each fixed effect contributes one degree of freedom and the random effects contributes a value in the range between zero and m, depending on the size of the variance component. As follows from the definition of WAIC, models with lower WAIC values are preferred. The WAIC estimates contains uncertainty and an approximation for the standard error is provided by Vehtari et al. (2017) equation (23) and can be computed using R package *loo* (Vehtari et al., 2015).

Covariates are selected from the set of auxiliary variables listed in Appendix A using a step-forward selection procedure. Various models are compared using the aformentioned WAIC estimates. From the set of potential covariates, the covariate with the lowest WAIC value is selected in the model. This selection process is iteratively repeated as long as adding a new covariate further decreases the WAIC value. In this application, this step-forward selection procedure, further abbreviated as step-WAIC, often results in models with a large number of covariates. Since the WAIC values are estimates that contain error, it appears that it is not meaningful to minimize the WAIC by adding covariates to the model as long as it reduces the point estimates of the WAIC. As an alternative we applied a step-forward selection procedure where covariates are added to the model as long as a new covariate decreases the WAIC with a value that exceeds the estimated standard error of the WAIC, further abbreviated as step-WAIC-se.

The step-forward selection procedure is applied to each of the six different combinations of the two fixed effect versions (FE_uq and FE_eq) and the three covariance structures of the random component (RE_f, RE_d, and RE_s). From these six models the best is selected using WAIC. For the finally selected models, model adequacy is evaluated with posterior predictive checks. This implies that replicates from the posterior distribution are simulated and compared with the originally observed data to study systematic discrepancies and to evaluate how well the selected model fits the observed data (Gelman et al., 2004). Posterior predictive p-values are calculated for six different tests that evaluate particular aspects of the posterior predictive distribution. Posterior predictive p-values for the domain discontinuities are defined as $p = P(T(\hat{\Delta}^{sim}, \Delta) \geq T(\hat{\Delta}, \Delta)|\hat{\Delta})$, where $\hat{\Delta}^{sim}$ are replicates of the discontinuities under the posterior predicitive distribution, $\hat{\Delta}$ the observed direct estimates for the discontinuities and $T(\hat{\Delta}^{sim}, \Delta)$ a test statistic that depends on $\hat{\Delta}^{sim}$ and parameters Δ. Posterior predictive p-values are estimated from the Gibbs sampler output as the average over the Monte Carlo samples s

$$\hat{p} = \frac{1}{S} \sum_{s=1}^{S} I(T(\hat{\Delta}^{s}, \Delta^{s}) \geq T(\hat{\Delta}, \Delta^{s})),$$

with $I(A)$ the indicator function with value one if the condition A is fulfilled and zero otherwise. If a model fits the observed data adequately, then it is expected that $T(\hat{\Delta}, \Delta^{s})$ is in the bulk of the histogram of the replicates $T(\hat{\Delta}^{s}, \Delta^{s})$. Therefore p values close to zero or one are indications of a poor fit with respect to that test statistic. The following posterior predictive tests are defined (You, 2008):

1. A general goodness-of-fit test statistic $T_1 = \sum_{i=1}^{m} (\hat{\Delta}_i - \Delta_i)^2 / Var(\hat{\Delta}_i|\Delta_i)$. Here $Var(\hat{\Delta}_i|\Delta_i) = \psi_i^r + \psi_i^a$
2. $T_2 = max(\hat{\Delta}_i)$ and $T_3 = min(\hat{\Delta}_i)$ which are sensitive for deviations in the tails of the distribution
3. $T_4 = \frac{1}{m} \sum_{i=1}^{m} \hat{\Delta}_i \equiv \bar{\hat{\Delta}}$, i.e. the mean which is sensitive for bias in the domain predictions.
4. $T_5 = \frac{1}{m-1} \sum_{i=1}^{m} (\hat{\Delta}_i - \bar{\hat{\Delta}})^2$, i.e. the variance of the domain estimates, which is sensitive for e.g. overschrinkage
5. $T_6 = |max(\hat{\Delta}_i) - \bar{\Delta}| - |min(\hat{\Delta}_i) - \bar{\Delta}|$, with $\bar{\Delta} = \frac{1}{m} \sum_{i=1}^{m} \Delta_i$ which is sensitive to asymmetry in the distribution.

6. A graphical comparison of the simulated posterior predictive distribution with the direct estimate $\hat{\Delta}_i$ and the domain prediction $\tilde{\Delta}_i^{bFH}$, i.e. the posterior mean, for each domain separately.

4 Results

4.1 Model selection

In Subsection 3.2, two different versions for the fixed effects (FE_uq and FE_eq) and three different covariance structures of the random effects (RE_f, RE_d and RE_s) are considered. The step-forward selection procedure from Subsection 3.5 is applied to each combination separately to select covariates. For each covariance structure a choice between FE_uq and FE_eq is made based on the WAIC value. Results where covariates are added until the WAIC value is not further decreased are summarized in Table 4.1. Results where covariates are added as long as the decrease of the WAIC is larger than its standard error are summarized in Table 4.2. In these tables `mode` × `[other variable names]` refer to models with different regression coefficients for the covariates, i.e. FE_uq, and δ_i + `other variable names` refer to models with equal regression coefficients for the covariates, i.e. FE_eq. The tables summarize the selected models with their WAIC values, the standard errors of the WAIC and the posterior means and standard errors of variances and correlations of the random components.

For total offences, `offtot`, a full covariance matrix (RE_f) results in large random domain effects with a strong positive correlation of 0.98. In this case a model with equal regression coefficients results in the best WAIC value. Assuming diagonal covariance matrices for the random effects (RE_d and RE_s) result in models with a substantially larger number of covariates with unequal regression coefficients between both surveys. In this case the random components are small. Although the WAIC values of RE_d and RE_s are 8 points smaller compared to model RE_f there is, with only 25 domains, a substantial risk that RE_d and RE_s in combination with unequal regression coefficients result in models that overfit the data. If covariates are added to the model if the decrease in the WAIC value exceeds the standard error, then the same parsimoniuous model (RE_f, FE_eq) is found.

A first observation is that there seems to be a trade-off between the complexity of the random and the fixed component of the model; RE_f results in large random effects with strong correlation while RE_d and RE_s results in complex fixed effect models with small random effects. As a compromise, the covariates selected in RE_d and RE_s could be used in a model with equal regression coefficients. Assuming FE_eq in combination with RE_d or RE_s, results in the same set of covariates obtained with FE_uq (see Table 4.1). Since this set of covariates predict the domain variables quite well, the random effects are small. A model with equal regression coefficients and small random effects results in highly synthetic estimates for the discontinuities, since the regression coefficient β_2 for the dummy variable δ_i is the only variable from the fixed effect component that discriminates between the regular and alternative approach. Since the random effects are small, the domain discontinuity estimates are approximately equal to β_2. This is almost similar to assuming that the domain discontinuities are equal to the direct estimator for the discontinuity at the national level. This phenomenon is illustrated with figures 4.1, 4.2 and 4.3. Figure 4.1 shows boxplots of the posterior predictive distributions of domain discontinuities under model (RE_f, FE_eq) together with the direct estimates and the domain predictions for the discontinuities obtained with the bivariate FH model. In this model the variation in the prediction of the domain disontinuities comes from the random effects. In Figure 4.2 a similar comparison is shown for model (RE_d, FE_uq). In this case the variation in the prediction of the domain disontinuities comes from the differences in the regression coefficients of the fixed effects. Finally Figure 4.3 shows a similar comparison for model (RE_d, FE_eq) where the set of selected covariates is equal to the set selected for (RE_d, FE_uq). The WAIC value for this model equals 300. Since the random effects are small and regression coefficients are equal, the prediction for the domain discontinuities are almost equal to the regression coefficient for δ_i for each domain.

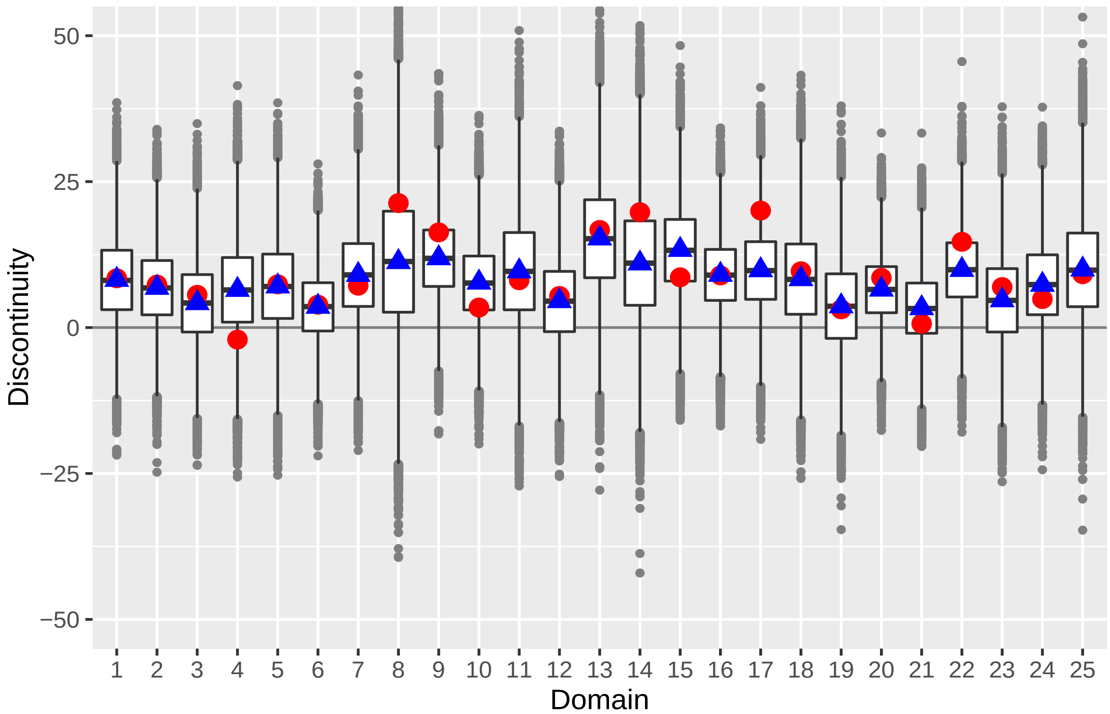


Figure 4.1 Boxplots posterior predicitive distribution domain discontinuities for offtot with selected model (RE_f and FE_eq) with direct estimates (red circles) and HB bivariate FH domain predictions (blue triangles).

The covariate selection for `unsafe` for all three covariance structures RE_f, RE_d, and RE_s results in relative large set of covariates with unequal regression coefficients (FE_uq). The random effects are small and the correlation between the random effects is also near zero. The selected set of covariates is similar for each of the three covariance structures, see Table 4.1. It is interesting to note that if a full covariance matrix for the random effects is fitted to the data with only an intercept, then the size of the random domain effects increases from 0.6 to 3 with a strong positive correlation of 0.89. This illustrates the aforementioned trade-off between the complexity of the covariance structure of the random effects and the fixed effects. The selected models under step-WAIC are large compared with the number of domains and therefore might overfitt. For domain discontinuities, a model with equal regression coefficients is not an alternative since this result in synthetic domain estimates, similar to `offtot`. The step-WAIC-se procedure results in more parsimonious models. Again the same set of covariates is selected for RE_f, RE_d and RE_s, see Table 4.2. The random effects are larger but the correlation between the random effects is weak (0.04). Therefore a model with equal regression coefficients (FE_eq) with a diagonal covariance matrix with equal variances (RE_s) is selected with covariates

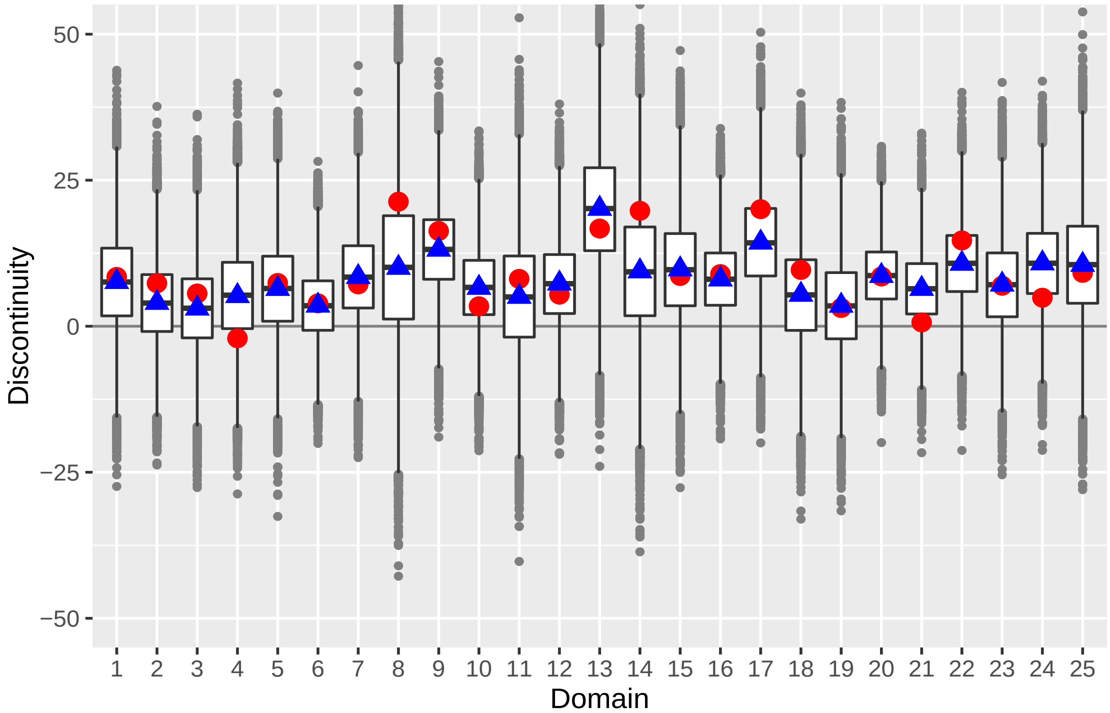


Figure 4.2 Boxplots posterior predicitive distribution domain discontinuities for offtot with model (RE_d and FE_uq) with direct estimates (red circles) and HB bivariate FH domain predictions (blue triangles).

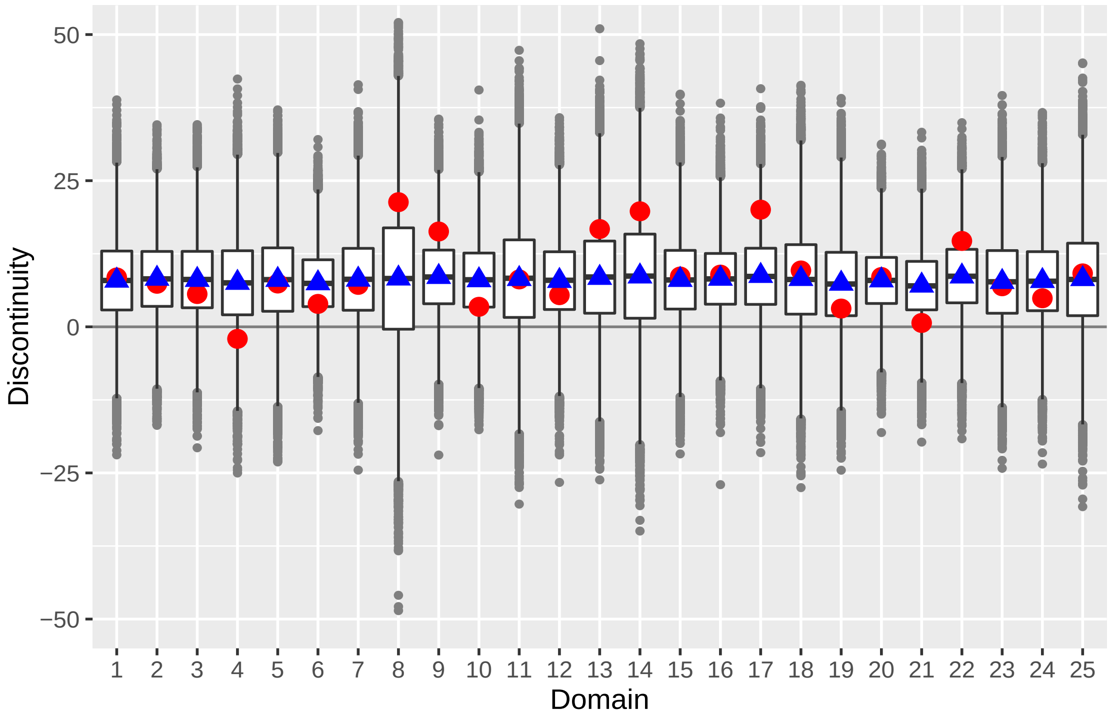


Figure 4.3 Boxplots posterior predicitive distribution domain discontinuities for offtot with model (RE_d and FE_eq) with direct estimates (red circles) and HB bivariate FH domain predictions (blue triangles).

specified in Table 4.2. Boxplots of the posterior predictive distributions under this model are provided in Figure 5.1 in Appendix C.1.

For nuisance the best model is obtained with a full covariance matrix for the random domain effects (RE_f) with a strong positive correlation of 0.81. With both selection procedures the same parsimonious set of covariates are selected (compare Table 4.1 with 4.2) assuming equal regression coefficients (FE_eq). The random effects are large enough to model the variation in the domain discontinuities. In the case of diagonal covariance structures, a substantial larger set of covariates is selected and the size of the random effects becomes small. Similarly to offtot this requires unequal regression coefficients to avoid synthetic domain discontinuity estimates. If for RE_d and RE_s covariates are selected with the step-WAIC-se procedure, then more parsimonious sets of covariates are obtained (Table 4.2). In this case the size of the random components increases, so models with equal regression coefficients give similar predictions for domain discontinuities compared to models with unequal regression coefficients. In Appendix C.1 boxplots of the posterior predictive distributions under the selected model (RE_f, FE_eq) are given in Figure 5.2 and compared with two alternative models. The first alternative is model (RE_d,FE_uq) with the large set of covariates from Table 4.1 in Figure 5.3. The second alternative is model (RE_d,FE_eq) with a parsimonious set of covariates from Table 4.2 in Figure 5.4.

In the case of propvict the model selection that minimizes the step-WAIC for RE_f, RE_d, and RE_s with unequal regression coefficients results in the same set of covariates (intercept plus two covariates). In the case of equal regression coefficients a very large model with 6 covariates is selected with small random effects. This results in synthetic predictions for domain discontinuities that do not vary between domains, similarly to offtot in Figure 4.3. Therefore a model with different regression coefficients is prefered. Model RE_f does not detect correlation between the random effects with the set of covariates specified in Table 4.1. A model with only mode as covariate, results in a model with larger random domain effects and a strong positive correlation (0.91). A compromise seems to be obtained with the step-WAIC-se selection procedure. This results in a model with a weak positive correlation of 0.1 and a smaller set of covariates. This model (RE_f, FE_uq) with the set of covariates specified in Table 4.2 is finally selected for propvic. A Boxplots of the posterior predictive distributions under this model are provided in Figure 5.5 in Appendix C.1.

The step-WAIC selection procedure for satispol results under RE_f, RE_d and RE_s in a model with a relatively large number of covariates with unequal regression coefficients, see Table 4.1. As in the case of offtot, unsafe, and propvict the random effects are small without correlation, since the set of covariates are strong predictors for the domain discontinuities. Model FE_eq is therefore not an alternative since it results in synthetic predictions for the domain discontinuities. If covariates are selected with the step-WAIC-se procedure, then more parismoneous models are selected, which appear to be an appropriate compromise to avoid overfitting. The finally selected model is (RE_s, FE_uq) with the covariates specified in Table 4.2. In Appendix C.1 boxplots of the posterior predictive distributions under the selected model are given in Figure 5.6. Boxplots of the posterior predictive distributions under model (RE_s, FE_uq) with the set of covariates from Table 4.1 are given in Figure 5.7.

variable	model	WAIC	df WAIC	SE WAIC	σ_r	SE σ_r	σ_a	SE σ_a	ρ	SE ρ
	Full covariance matrix random components									
nuisance	δ_i + adm_immigrnw	-61	15.3	5.1	0.20	0.04	0.14	0.04	0.81	0.01
unsafe	mode × [1+ pr_propcrim + pr_damage + adm_benefit + adm_immigrnw]	246	12.0	11.4	0.57	0.44	0.61	0.52	0.01	0.52
propvict	mode × [1+ pr_propcrim + adm_old]	199	7.8	8.5	0.54	0.37	0.36	0.28	0.00	0.49
offtot	δ_i + pr_weapon	307	12.8	5.4	8.77	1.66	5.32	1.43	0.98	0.03
satispol	mode × [1+ adm_immigr + adm_urban + pr_traffic]	305	6.6	7.5	0.68	0.62	0.65	0.60	0.01	0.537
	Diagonal covariance matrix random components, mode dependent random effects ($\sigma_r \neq \sigma_a$)									
nuisance	mode × [1+ adm_immigr + pr_damage + pr_threat + adm_urban + pr_violcrime + pr_weapon + pr_drugs]	-52	17.6	6.6	0.09	0.04	0.05	0.04	-	-
unsafe	mode × [1+ pr_propcrim + pr_damage + adm_benefit + adm_immigrnw]	246	13.7	10.6	0.86	0.57	1.08	0.78	-	-
propvict	mode × [1+ pr_propcrim + adm_old]	199	8.8	7.9	0.71	0.44	0.46	0.36	-	-
offtot	mode × [1+ adm_immigr + adm_young + pr_puborder + pr_propcrim]	299	8.3	5.5	1.51	1.07	1.02	0.81	-	-
satispol	mode × [1+ adm_immigr + adm_urban + pr_traffic]	305	7.12	7.3	1.27	0.94	1.25	1.01	-	-
	Diagonal covariance matrix random components, equal random effects ($\sigma_r = \sigma_a$)									
nuisance	mode × [1+ adm_immigr + pr_threat + adm_urban + pr_violcrime + pr_damage]	-49	18.2	6.6	0.08	0.03	0.08	0.03	-	-
unsafe	mode × [1+ pr_propcrim + pr_damage + adm_benefit + adm_immigrnw]	246	12.8	10.9	0.81	0.52	0.81	0.523	-	-
propvict	mode × [1+ pr_propcrim + adm_old]	199	7.5	8.7	0.46	0.31	0.46	0.31	-	-
offtot	mode × [1+ adm_immigr + adm_young + pr_puborder + pr_propcrim]	299	7.8	5.9	1.07	0.78	1.07	0.78	-	-
satispol	mode × [1+ adm_immigr + adm_urban + pr_traffic]	305	6.6	7.5	1.02	0.77	1.02	0.77	-	-

Table 4.1 Optimal models based on the step-WAIC procedure.

variable	model	WAIC	df WAIC	SE WAIC	σ_r	SE σ_r	σ_a	SE σ_a	ρ	SE ρ
	Full covariance matrix random components									
nuisance	δ_i + adm_immigrnw	-61	15.3	5.1	0.20	0.04	0.14	0.04	0.81	0.01
unsafe	δ_i + pr_propcrim	252	10.6	14.3	0.62	0.46	0.92	0.72	0.04	0.55
propvict	mode × [1+ pr_propcrim]	206	9.2	9.1	0.79	0.45	0.39	0.31	0.10	0.51
offtot	δ_i + pr_weapon	307	12.8	5.4	8.77	1.66	5.32	1.43	0.98	0.03
satispol	mode × [1+ adm_immigr]	308	4.5	8.0	0.69	0.62	0.67	0.63	0.02	0.55
	Diagonal covariance matrix random components, mode dependent random effects ($\sigma_r \neq \sigma_a$)									
nuisance	δ_i + adm_immigr	-43	21.0	8.9	0.18	0.04	0.10	0.04	-	-
unsafe	δ_i + pr_propcrim	251	14.7	13.2	0.97	0.59	1.68	0.90	-	-
propvict	mode × [1+ pr_propcrim]	206	10.1	8.1	1.04	0.48	0.50	0.38	-	-
offtot	mode × [1+ adm_immigr]	314	12.3	9.1	3.37	1.30	1.31	1.01	-	-
satispol	mode × [1+ adm_immigr]	308	6.2	7.7	0.90	0.77	0.95	0.90	-	-
	Diagonal covariance matrix random components, equal random effects ($\sigma_r = \sigma_a$)									
nuisance	δ_i + adm_immigr	-44	21.6	7.6	0.15	0.03	0.15	0.03	-	-
unsafe	δ_i + pr_propcrim	250	12.9	13.4	1.17	0.58	1.17	0.58	-	-
propvict	mode × [1+ pr_propcrim]	208	9.3	9.3	0.68	0.38	0.68	0.38	-	-
offtot	mode × [1+ adm_immigr]	317	12.9	9.9	2.48	1.05	2.48	1.05	-	-
satispol	mode × [1+ adm_immigr]	308	5.4	7.8	0.78	0.66	0.78	0.66	-	-

Table 4.2 Optimal models based on the step-WAIC-se procedure

The finally selected models are summarized in Table 4.3. The univariate FH models developed in van den Brakel et al. (2016) for the alternative survey approach are summarized in Table 4.4. The auxiliary variables starting with REG_xx are survey estimates obtained from the regular survey. A description of these variables is given in Table 5.3 in Appendix A.

Standard model diagnostics test the underlying assumptions that the random domain effects and the residuals are normally and independently distributed. Since the number of domains in this application is small, the power of the tests for normality are weak and do not indicate deviations from normality. Therefore the posterior predictive tests as summarised in Subsection 3.5 are used to evaluate the model adequacy. In Subsection 4.2 the domain predictions aggregated to the national level are compared with the direct estimates at the national level to evaluate the bias introduced with the small area estimation procedures. The posterior predictive p-values for the domain estimates of the target variables and the discontinuities are summarized in Table 4.5. The general measure for goodnes-of-fit (T1) indicates that the fit for the discontinuities of `offtot` is of reduced quality (other models considered had similar high values). The posterior predictive p-values for maximum (T2) and minimum (T3) values do not indicate problems with the tails of the distributions. The posterior predictive values for the mean (T4) and asymmetry of the distribution (T6) indicate that the distributions are symmetrically concenctrated arround their mean. The posterior predictive p-values for the variance (T5) indicate some undershrinkage for the discontinuities of `nuisance`, `propvict`, and `offtot`.

variable	model	covariance structure
nuisance	δ_i + adm_immigrnw	RE_f
unsafe	δ_i + pr_propcrim	RE_s
propvict	mode × [1+ pr_propcrim]	RE_f
offtot	δ_i + pr_weapon	RE_f
satispol	mode × [1+ adm_immigr]	RE_s

Table 4.3 Final models

variable	model
nuisance	1 + REG_nuisance + adm_old
unsafe	1 + REG_nuisance + adm_benefit + pr_propcrim + pr_drugs
propvict	1 + pr_propcrim + adm_old
offtot	1 + REG_victim
satispol	1 + REG_funcpol

Table 4.4 Final models univariate FH model from van den Brakel et al. (2016)

4.2 Estimation results

In this section, the HB domain predictions obtained with the multivariate FH model are compared with the direct estimates and with the domain predictions obtained with the univariate approach where the direct estimates of the regular approach are potential auxiliary variables in the model selection. First the domain predictions for the variables under the regular and alternative survey are discussed. Subsequently results for the domain discontinuities are discussed.

With model-based small area estimation, the design variance of the direct estimators is reduced at the cost of accepting a small amount of bias. To evaluate differences in the direct point

variable	T1	T2	T3	T4	T5	T6
	Discontinuities					
nuisance	0.927	0.940	0.034	0.345	0.988	0.465
unsafe	0.343	0.841	0.833	0.454	0.437	0.912
propvict	0.925	0.261	0.029	0.258	0.970	0.070
offtot	0.980	0.797	0.069	0.337	0.968	0.416
satispol	0.772	0.595	0.392	0.610	0.762	0.484
	Target variables					
nuisance	0.766	0.317	0.108	0.433	0.504	0.156
unsafe	0.308	0.779	0.492	0.420	0.474	0.708
propvict	0.695	0.339	0.168	0.379	0.655	0.194
offtot	0.859	0.249	0.024	0.317	0.524	0.089
satispol	0.742	0.929	0.584	0.457	0.875	0.797

Table 4.5 Posterior predictive p-values for the final models from Table 4.3

estimates and the small domain predictions, the following two measures are defined. The first one is the Mean Relative Difference (MRD), which summarizes the differences between the direct estimates and the domain predictions:

$$MRD = \frac{100\%}{m} \sum_{i=1}^{m} \frac{\hat{y}_i^q - \tilde{y}_i^q}{\hat{y}_i^q}, \quad q = r, a, \tag{16}$$

and $\tilde{y}_i^q$ is the domain prediction based on the bivariate FH model or the univariate FH model. The second measure is the Absolute Mean Relative Difference (AMRD) between the direct estimate and the domain prediction, which is defined as:

$$AMRD = \frac{100\%}{m} \sum_{i=1}^{m} \left| \frac{\hat{y}_i^q - \tilde{y}_i^q}{\hat{y}_i^q} \right|, \quad q = r, a, \tag{17}$$

The increased precision of the small domain predictions is measured with Mean Relative Difference of the Standard Errors (MRDSE) between the direct estimates and the domain predictions and is defined as

$$MRDSE = \frac{100\%}{m} \sum_{i=1}^{m} \frac{SE(\hat{y}_i^q) - SE(\tilde{y}_i^q)}{SE(\hat{y}_i^q)}, \quad q = r, a, \tag{18}$$

These measures are defined in a simialr way for the estimates and predictions of the domain discontiuities $\hat{\Delta}_i$ and $\tilde{\Delta}_i$.

In Table 4.6 the domain predictions and their standard errors averaged over the domains as well as the MRD, AMRD and MRDSE are given for the alternative survey under the univariate FH model with the models presented in Table 4.4. Results under the bivariate FH model are presented in Table 4.7 for the variables under the alternative survey and in Table 4.8 for the variables under the regular survey. Comparing Table 4.6 with Table 4.7 shows that the bivariate

FH model results in stronger reductions in the standard errors for all variables with the exception of `nuisance`. As a consequence the deviations between the direct estimates and the small area predictions are also larger under the bivariate FH model.

variable	HB est.	SE	MRD (%)	AMRD (%)	MRDSE (%)
`nuisance`	1.29	0.08	-0.74	5.02	37.96
`unsafe`	19.83	1.64	-0.96	7.58	41.16
`propvict`	9.85	0.84	-3.17	11.86	60.69
`offtot`	33.21	2.90	-0.44	7.03	47.74
`satispol`	55.09	2.54	-0.11	6.43	61.98

Table 4.6 Domain prediction alternative survey with univeriate FH model

variable	HB est.	SE	MRD (%)	AMRD (%)	MRDSE (%)
`nuisance`	1.28	0.08	-0.98	4.28	35.72
`unsafe`	19.82	1.21	-2.54	11.97	56.47
`propvict`	9.91	0.73	-4.81	14.70	65.35
`offtot`	33.26	2.82	-0.99	6.93	49.36
`satispol`	55.08	1.97	-0.49	8.97	70.06

Table 4.7 Domain prediction alternative survey with bivariate FH model

variable	HB est.	SE	MRD (%)	AMRD (%)	MRDSE (%)
`nuisance`	1.60	0.09	0.38	2.62	15.93
`unsafe`	24.22	1.07	-0.01	6.02	46.78
`propvict`	12.18	0.88	1.58	7.84	43.70
`offtot`	41.34	3.76	0.96	4.56	17.95
`satispol`	60.82	1.47	-0.77	5.38	64.83

Table 4.8 Domain prediction regular survey with bivariate FH model

variable	regular		alternative			discontinuity		
	GREG	biv. FH	GREG	biv. FH	uni. FH	GREG	biv. FH	uni. FH
`offtot`	43.79	42.47	34.09	34.02	34.09	9.7	8.45	9.7
`unsafe`	25.07	24.89	20.48	20.49	20.48	4.59	4.40	4.59
`nuisance`	1.67	1.66	1.34	1.34	1.34	0.33	0.32	0.34
`satispol`	59.88	60.36	55.10	55.06	55.12	4.78	5.29	5.04
`propvict`	13.02	12.76	10.32	10.33	10.32	2.70	2.43	2.70

Table 4.9 GREG estimates national level and aggregated HB predictions regular and alternative survey approach (19).

The direct estimates at the national level can be considered as accurate estimates since they are based on sufficient large sample sizes. Therefore the bias in model-based domain predictions is often assessed by comparing the direct estimates at the national level with the domain predictions aggregated to the national level. The target variables in this application are all defined as population means. Therefore the aggregated domain predictions are obtained as the average over the domains weighted with the domain size,

$$\tilde{y}^q = \sum_{i=1}^{m} \frac{N_i}{N} \tilde{y}_i^q \tag{19}$$

with N_i the population size of domain i and N the size of the total population.

Table 4.9 compares the weighted average of the domain predictions according to (19) with the national GREG estimates. For the univariate FH model, the aggregated domain predictions are almost exactly equal to the GREG estimates at the national level. For the bivariate FH model the differences are slightly larger but the aggregated domain predictions are still very close to the GREG estimates at the national level. The largest relative difference amounts to 3% and is observed for `offtot` under the regular survey.

For `offtot` and `unsafe` the domain predictions with the univariate and bivariate FH model are plotted against the GREG estimates in Figures 4.4 and 4.6. The graphs also contain the GREG estimate at the national level versus the domain predictions aggregated to the national level according to (19). Figure 4.4 shows that there is a limited amount of shrinkage for `offtot`. Figure 4.6 shows that the bivariate FH model shrinks the domain predictions for the alternative survey while the amount of shrinkage for the univariate FH model for the alternative survey and the bivariate FH model for the regular survey is smaller.

Figures 4.5 and 4.7 compare the standard errors of the GREG domain estimates, the univariate FH model, and the multivariate FH model for respectively `offtot` and `unsafe`. The precision of the domain predictions for `offtot` under the alternative survey is clearly improved compared with the GREG estimates. The differences between the univariate and bivariate FH model are small for this variable. The improvement in precision with the bivariate FH model for the regular survey is smaller, as expected since the sample size of the regular survey is larger. Also for `unsafe` the standard errors with the predictions of the FH models for the alternative survey are much smaller compared to the GREG estimates. In this case the standard errors of the predictions with the bivariate Fay-Herriot model are clearly smaller compared to the univariate FH model. In this case the bivariate FH model also clearly improves the precision of the domain predictions of the regular survey.

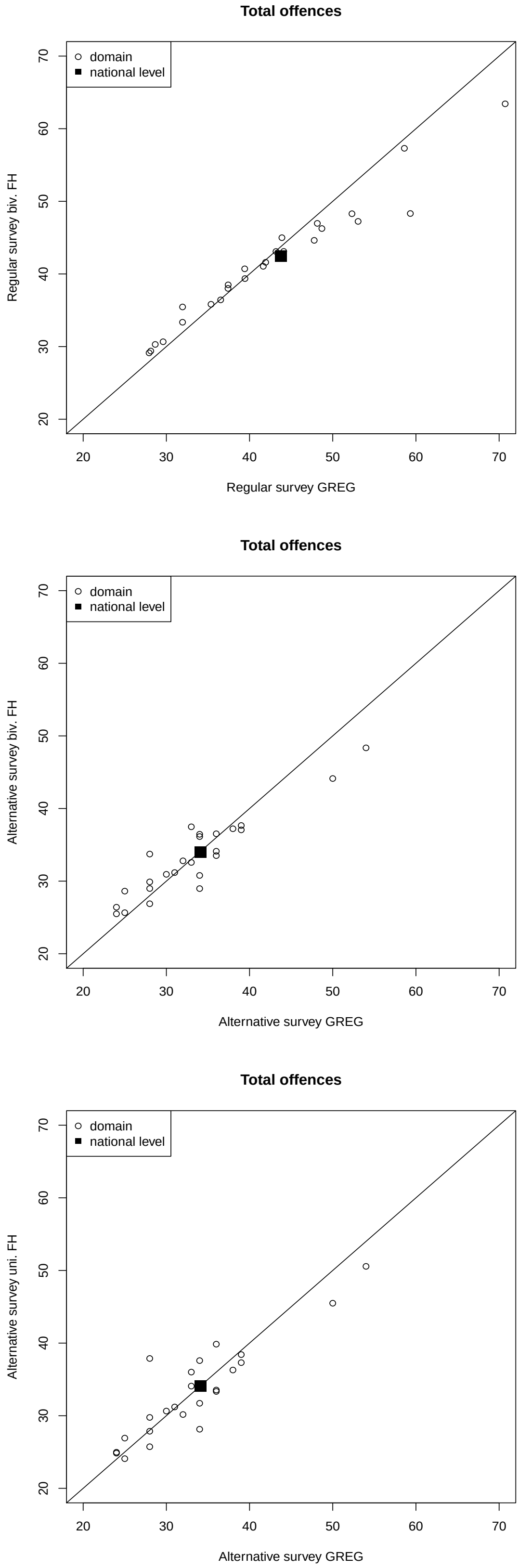


Figure 4.4 Domain estimates GREG versus HB predictions offtot. Upper panel: regular survey using bivariate FH model, middle panel: alternative survey using univariate FH model, lower panel alternative survey using bivariate FH model. Domain predictions are aggregated at the national level according to (19).

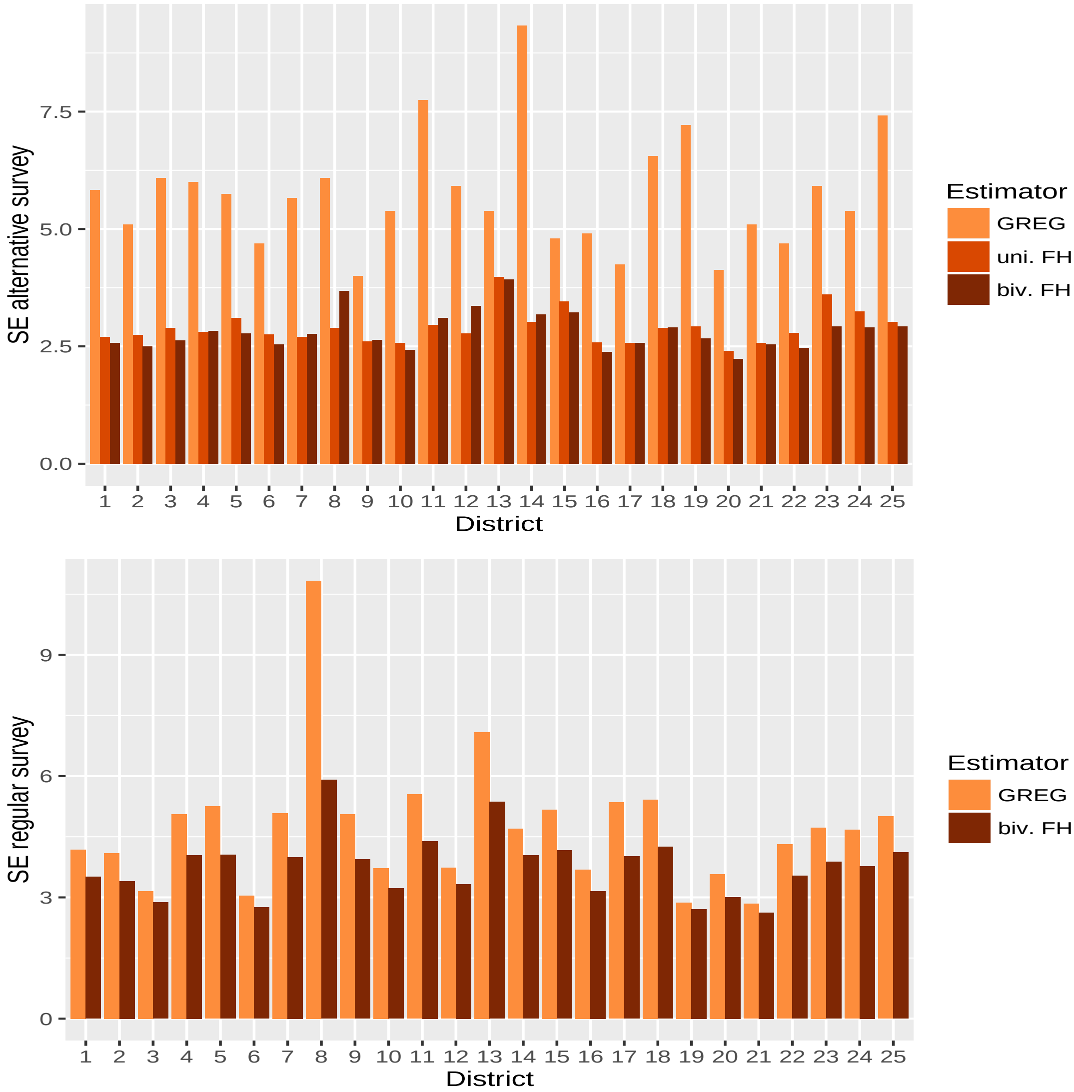


Figure 4.5 Standard errors domain estimates GREG and HB predictions offtot for alternative survey (upper panel) and regular survey (lower panel).

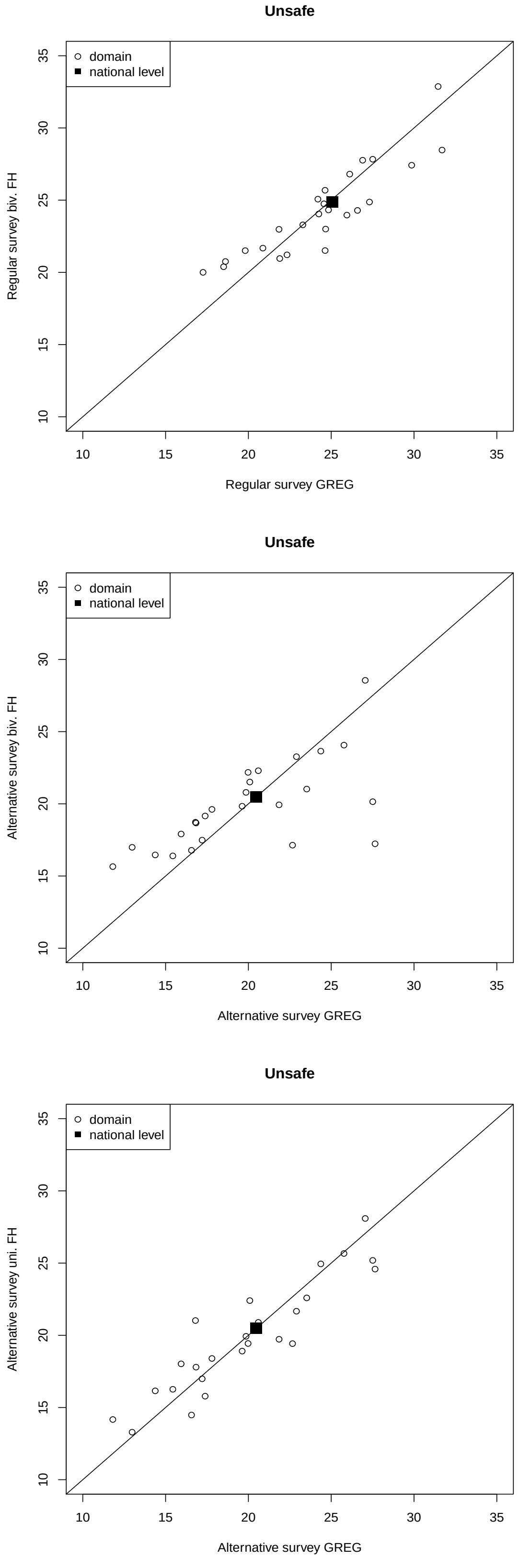


Figure 4.6 Domain estimates GREG versus HB predictions unsafe. Upper panel: regular survey using bivariate FH model, middle panel: alternative survey using univariate FH model, lower panel alternative survey using bivariate FH model. Domain predictions are aggregated at the national level according to (19).

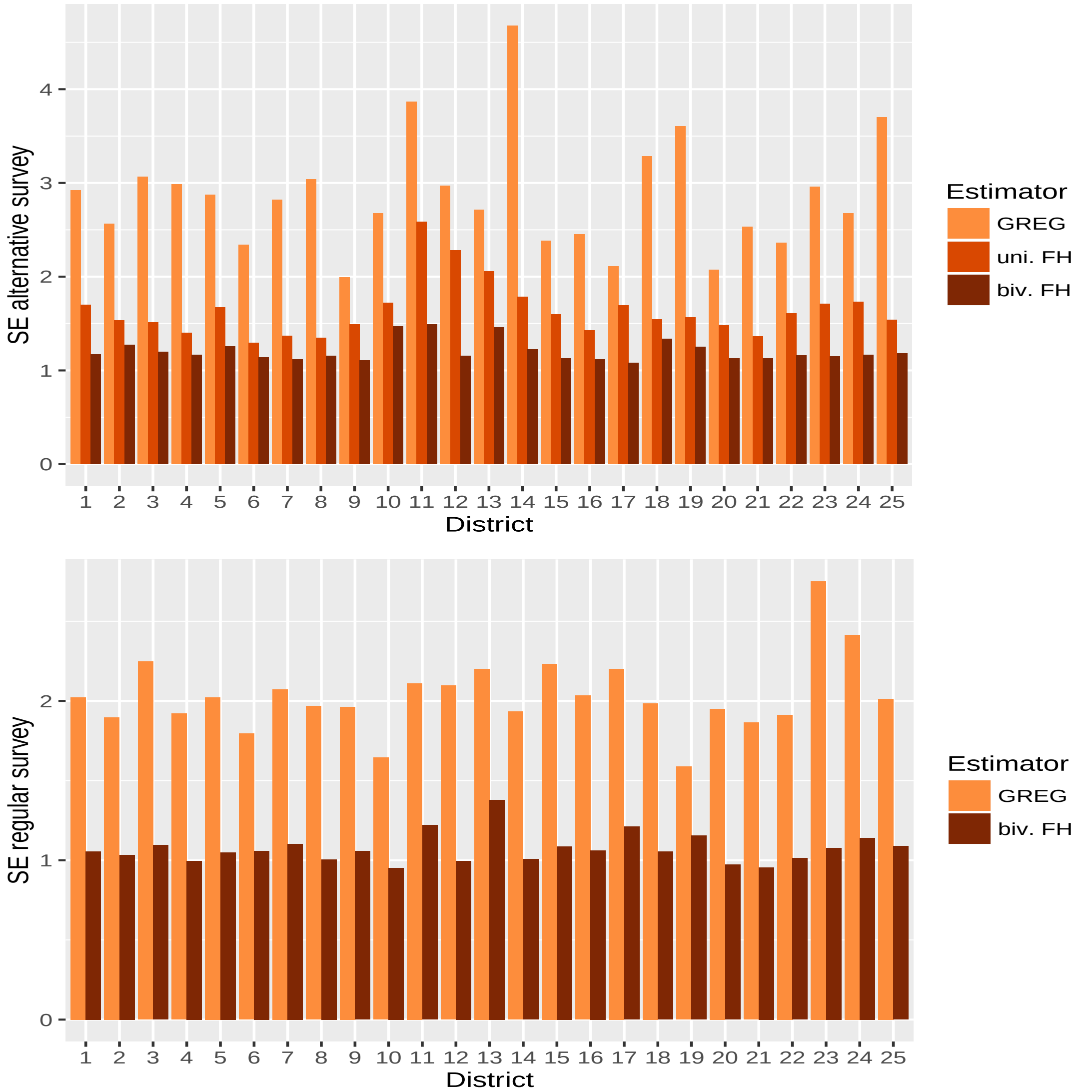


Figure 4.7 Standard errors domain estimates GREG and HB predictions unsafe for alternative survey (upper panel) and regular survey (lower panel).

In Tables 4.10 and 4.11 the domain predictions and their standard errors for the discontinuities averaged over the domains as well as the MRD, AMRD and MRDSE are summarized for the univariate FH model and bivariate FH model respectively. In the last three columns of Table 4.9, the GREG estimates for the discontinuities at the national level are compared with the domain predictions obtained with the univariate and bivariate FH model aggregated to the national level using (19). The MRD's and AMRD's are large because the GREG estimates for the discontinuities in the denominator of (16) and (17) frequently take values close to zero, which make these indicators instable. The differences between the GREG estimates for the discontinuities at the national level and the aggregated domain predictions are larger for the bivariate model. This can be expected since under the bivariate FH model the discontinuities are defined as the difference between domain predictions for the regular and alternative survey, while under the univariate FH model, domain discontinuities are obtained as the difference between the GREG estimates for the regular survey and the domain predictions for the alternative survey. So in the latter case the estimates for the regular survey are not adjusted contrary to the bivariate FH model. In addition the domain predictions for the alternative survey have larger MRD's and AMRD's under the bivariate FH model compared to the univariate FH model (compare Table 4.6 and 4.7).

With the exception of `nuisance` the standard errors for the domain predictions under the bivariate model are smaller compared to the univariate model. In the case of `propvict` and `offtot`, this is the result of slightly more precise domain predictions for the alternative survey with respect to the univariate FH model (compare Table 4.6 with 4.7), a clear improvement in precision of the domain predictions of the regular survey compared to the GREG estimators (Table 4.8) and the positve correlation between the random effects. In the case of `satispol` and `unsafe` this is mainly the result of a clear improvement of precision of the domain predictions with the bivariate FH model for the regular compared to the GREG estimators (Table 4.8) and also a clear improvement of the precision of the domain predictions with the bivariate FH model for the alternative survey compared to the univariate model (compare Table 4.6 with Table 4.7).

variable	HB est.	SE	MRD (%)	AMRD (%)	MRDSE (%)
`nuisance`	0.33	0.07	-5.80	18.97	57.49
`unsafe`	4.55	2.46	42.98	70.60	29.45
`propvict`	2.70	1.83	-142.60	170.76	32.75
`offtot`	9.08	3.92	-5.67	23.60	48.47
`satispol`	5.52	4.72	99.26	224.25	40.86

Table 4.10 Domain predictions for discontinuities univariate FH model

variable	HB est.	SE	MRD (%)	AMRD (%)	MRDSE (%)
`nuisance`	0.31	0.09	-4.01	22.96	44.47
`unsafe`	4.40	1.56	55.49	115.64	55.09
`propvict`	2.27	1.10	-113.00	151.22	59.02
`offtot`	8.07	2.68	1.94	25.83	63.60
`satispol`	5.74	2.46	228.50	437.25	68.75

Table 4.11 Domain predictions for discontinuities bivariate FH model

For `propvict` the point estimates for the regular and alternative survey are plotted against each other in Figure 4.8 for the GREG estimator, the univariate FH model and the bivariate FH model. The improved precision with the univariate and bivariate FH model is illustrated with the increased linear relationship between the point estimates. The deviation from the solid line with slope equal to one, illustrates the size of the discontinuity. For `propvict` the linear relationship suggests that the size of the discontinuity is proportional to the size of `propvict`. For `propvict` the discontinuities estimated with the GREG estimator, the univariate FH model and the bivariate FH model are plotted in Figure 4.9. Variation between size and sign of the point estimates is the largest for the GREG estimates, clearly improves with the univariate FH model and is the smallest under the bivariate FH model. The 95% confidence interval is clearly smaller for the univariate FH model compared to the GREG estimator. For the bivariate FH model the 95% confidence intervals are the smallest, resulting in the largest number of domains with discontinuity estimates that are significantly different from zero. Similar figures for the other variables are included in Appendix C.2. In a similar way as for `propvict`, the bivariate FH model results in a clear improvement of the predictions for the domain discontinuities of `offtot` and `unsafe`. For `nuisance` the standard errors for the discontinuities increase with the bivariate FH model compared to the univariate FH model. The bivariate FH model for `satispol` cannot adequately model the observations under the alternative survey with the auxiliary information from the two registers (MBA and PRRO). This results in overschrinkage of the domain predictions of `satispol` under the alternative approach. The univariate FH model indeed selects an auxiliary variable from the regular survey, see Table 4.4.

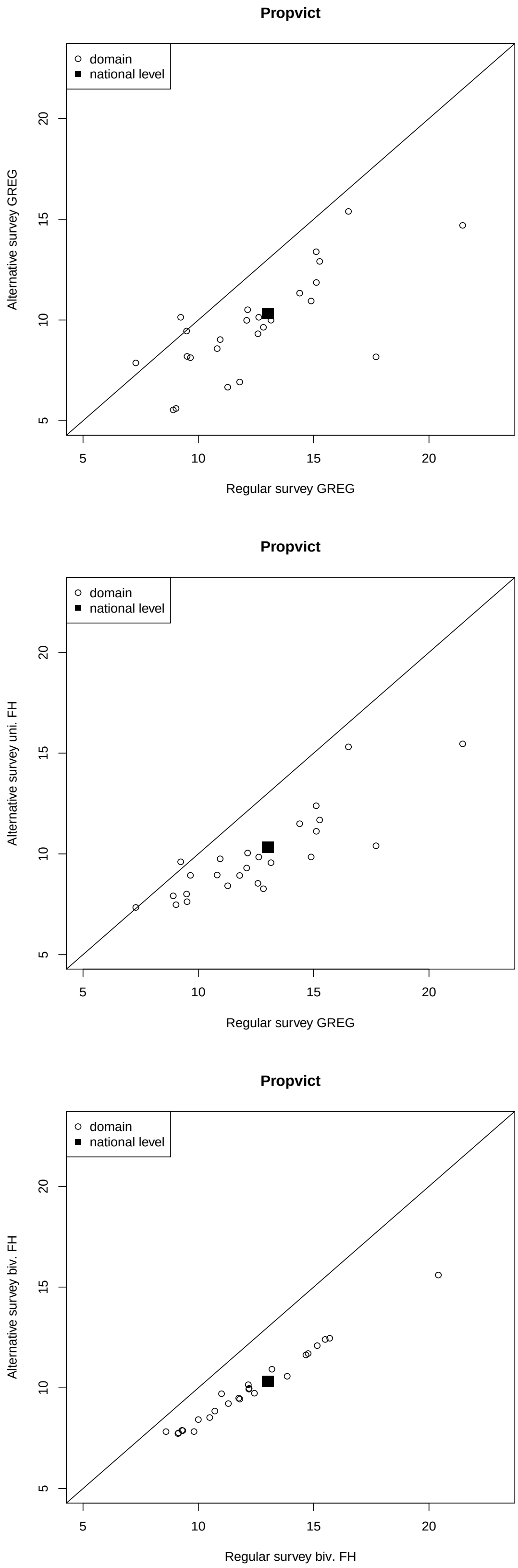


Figure 4.8 Domain estimates propvict alternative survey versus regular survey. Estimates at the national level are based on the GREG estimator.

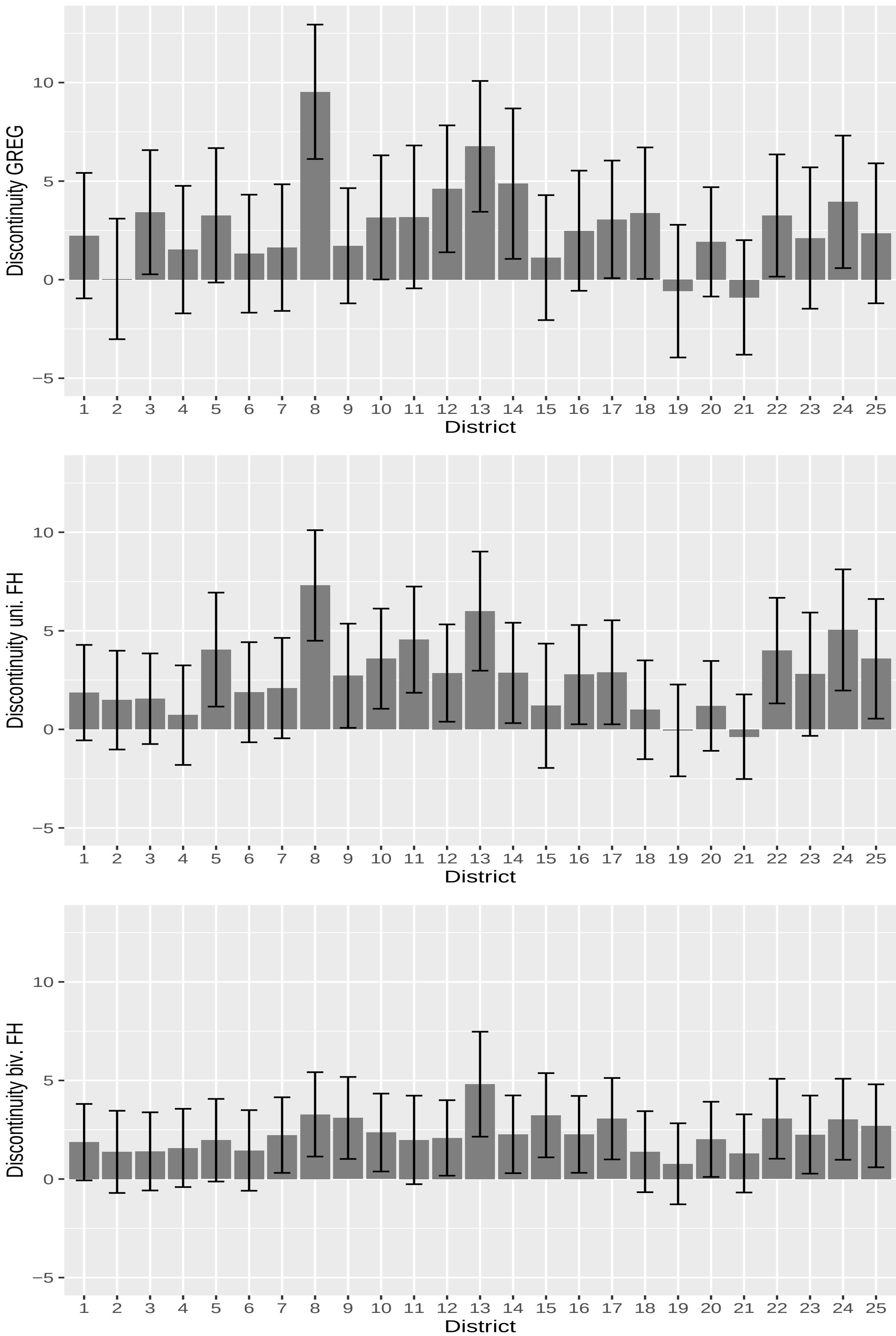


Figure 4.9 Discontinuities propvict based on the GREG estimator (upper panel), univariate FH model (middle panel) and bivariate FH model (lower panel) with a 95% confidence interval.

5 Discussion

Survey process redesigns often result in discontinuities that disturb the comparability of the outcomes over time obtained with a repeated survey. To avoid confounding real period-to-period change with differences in measurement bias, it is important that such discontinuities are quantified during the implementation of a new survey proces. A straightforward approach is to collect data under the old and new design in parallel to each other for some period of time. Available budgets for parallel data collection often do not meet the minimum required sample sizes that come from power calculations to detect minimum prespecified differences at certain significance and power levels. To obtain more precise predictions for the domain discontinuities a small area estimation approach based on hierarchical Bayesian Fay-Herriot (FH) models is proposed.

In an earlier paper (van den Brakel et al., 2016) a univariate FH model is proposed, where reliable direct domain estimates of the regular survey are considered as potential auxiliary variables in a step-forward model selection procedure to build adequate models for small domain prediction of the small sample assigned to the alternative survey. In this paper a bivariate FH model for the direct estimates obtained under both the regular and alternative survey is proposed as an alternative to obtain adequate predictions for domain discontinuities. Both methods are applied to a small scale parallel run conducted to quantify discontinuities in a survey process redesign of the Dutch Crime Victimization Survey (CVS).

Using direct estimates from the regular survey as auxiliary variables in models for small domains under the alternative approach results in a substantial improvement of precision, compared to univariate models that only use auxiliary variables from available registers. This can be expected since both surveys attempt to measure the same variables with a different survey approach. A drawback of the univariate approach is that the variance estimation procedure for the discontinuities is complex, since a non-ignorable covariance between the direct estimates from the regular design and the model based predictions for the alternative design arises. The method is complex since a model-based MSE is combined with a design-based variance of a direct estimator. This might even result in negative variance estimates for the discontinuities. These complications are partially circumvented by developing a design-based estimator for the MSE of the small domain predictions and the covariance component (van den Brakel et al., 2016).

With the bivariate FH model under a fully Bayesian framework the risk of negative variance estimates is avoided since the variances for disontinuities are derived from positive-definite covariance matrices of the bivariate model. The bivariate FH model improves the predictions for the domain discontinuities since the model improves the precision of the estimates of both the regular and alternative approach, and the strong positive correlation between the random domain effects further reduces the variance of the contrasts. For four out of five variables of the Dutch CVS the bivariate FH model indeed resulted in more precise predictions for domain discontinuities compared to the univariate FH model. Another advantage of the bivariate model is that it improves the domain prediction of both the regular and alternative model while the univariate model assumes that the sample size of the regular survey is sufficiently large to make reliable precise direct domain estimates. The bivariate model is therefore also appropriate in parallel runs where e.g. the sample size of the regular survey is reduced in order to increase the sample size for the alternative survey. Finally the bivariate FH model avoids the complications to account for sampling error in the covariates, which is often required if the direct estimates of the regular survey are used as covariates in a univariate FH model.

For one variable (satisfaction with police performance) no adequate model could be constructed with the available auxiliary variables from the registers only. For this variable the multivariate model seems to result in overshrinkage of the predictions for the domain discontinuities. The results of the univariate model are clearly better since the direct estimates from the regular survey are the only auxiliary variables that result in an adequate model for small domain predictions.

A general problem with the step-forward model selection procedure where covariates are included in the model as long as the WAIC value is reduced, is that this results in models with relatively large sets of covariates. With the limited number of domains in this application there is a real risk of overfitting the data. For some variables the covariates appear to be strong predictors for the domain variables, resulting in small random effects. Fitting a model without these covariates results in models with large random effects and strong positive correlations between the regular and alternative survey estimates. For other variables a model with a full covariance structure automatically results in parsimoneous models for the fixed effect part, probably because the set of available covariates are less strong predictors for these target variables.

Since the WAIC values are estimated from the Gibbs sampler output, these values are observed with some degree of uncertainty. This is an argument not to include covariates if they only result in a small reduction of the WAIC. In an alternative step-forward selection approach, covariates are only selected if the decrease in the WAIC value exceeds the estimated standard error of the WAIC. For variables where initially large sets of covariates were selected, this approach results in a reasonable compromise between model fit and model complexity. As an alternative, models with equal regression coefficients can be considered. Such models are, however, less appropriate for prediciting domain disontinuities if the random effects are small. In such situations the dummy indicator is the only model component that discrimates between the regular and alternative approach. This results in synthetic predictions for domain discontinuities that are almost equal over the domains, which are approximately equal to the direct estimator for the discontinuity at the national level. Depending on the type of changes in the survey process, it might be correct to assume that domain discontinuities are equal. In that case the best estimate is obtained with the directs estimator at the national level.

References

Bell, W. (1999). Accounting for uncertainty about variances in small area estimation. *Bulletin of the International Statistical Institute 52nd Session*.

Benavent, R. and D. Morales (2016). Multivariate fay-herriot models for small area estiamtion. *Computational Statistics and Data Analysis 94*, 372--390.

Bollineni-Balabay, O., J. van den Brakel, F. Palm, and H. Boonstra (2017). Multilevel hierarchical bayesian vs. state space approach in time series small area estimation: the dutch travel survey. *Journal of the Royal Statisitcal Society forthcoming*.

Boonstra, H. J. (2016). *mcmcsae: MCMC Small Area Estimation*. R package version 0.7.

Datta, G., M. Ghosh, N. Nangia, and K. Natarjan (1996). Estimation of median income in four-peron families: a bayesian approach. In D. Berry, K. Chaloner, and J. Geweke (Eds.), *Bayesian analysis in Statistics and Econometrics*, pp. 129--140. Wiley, New York.

Datta, G., P. Lahiri, T. Maiti, and K. Lu (1999). Hierarchical bayes estimation of unemployment rates for the states of the u.s. *Journal of the American Statistical Association 94*(448), 1074--1082.

Estaban, M., D. Morales, A. Perez, and L. Santamaria (2012). Small area estimation of poverty proportions under area-level time models. *Computational Statistics and Data Analysis 56*, 2840--2855.

Fay, R. and R. Herriot (1979). Estimates of income for small places: An application of james-stein procedures to census data. *Journal of the American Statistical Association 74*(366), 269--277.

Gelfand, A. and A. Smith (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*, 398--409.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis 1*(3), 515--533.

Gelman, A., J. Carlin, H. Stern, D. B. Dunson, A. Vehtari, and D. Rubin (2004). *Bayesian Data Analysis*. Chapman and Hall, New York.

Gelman, A., D. V. Dyk, Z. Huang, and W. Boscardin (2008). Using redundant parameterizations to fit hierarchical models. *Journal of Computational and Graphical Statistics 17*(1), 95--122.

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

Geman, S. and D. Geman (1984). Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Trans. Pattn Anal. Mach. Intell. 6*, 721--741.

Gonzales-Manteiga, W., M. Lombardia, I. Molina, D. Morales, and L. Santamaria (2008). Analytic and bootstrap approximations of prediction errors under a multivariate fay-herriot model. *Computational Statistics and Data Analysis 52*, 5242--5252.

Hodges, J. and D. Sargent (2001). Counting degrees of freedom in hierarchical and other richly-parameterised models. *Biometrika 88*(2), 367--379.

Marhuenda, Y., I. Molina, and D. Morales (2013). Small area estimation with spatio-temporal fay-herriot models. *Computational Statistics and Data Analysis 57*, 308--325.

O'Malley, A. and A. Zaslavsky (2008). Domain-level covariance analysis for multilevel survey data with structured nonresponse. *Journal of the American Statistical Association 103*(484), 1405--1418.

Pfeffermann, D. and L. Burck (1990). Robust small area estimation combining time series and cross-sectional data. *Survey Methodology 16*, 217--237.

Pfeffermann, D. and R. Tiller (2006). Small area estimation with state-space models subject to benchmark constraints. *Journal of the American Statistical Association 101*, 1387--1397.

Polson, N. and J. Scott (2012). On the half-cauchy prior for a global scale parameter. *Bayesian Analysis 7*(4), 887--902.

Rao, J. and I. Molina (2015). *Small Area Estimation*. Wiley-Interscience.

Rao, J. and M. Yu (1994). Small area estimation by combining time series and cross-sectional data. *The Canadian Journal of Statistics 22*, 511--528.

Särndal, C.-E., B. Swensson, and J. Wretman (1992). *Model Assisted Survey Sampling*. Springer.

Spiegelhalter, D., N. Best, B. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B 64*(4), 583--639.

van den Brakel, J. (2008). Design-based analysis of embedded experiments with applications in the dutch labour force survey. *Journal of the Royal Statisitcal Society 171*, 581--613.

van den Brakel, J., B. Buelens, and H. Boonstra (2016). Small area estimation to quantify discontinuities in repeated sample surveys. *Journal of the Royal Statisitcal Society 179*, 229--250.

van den Brakel, J. and S. Krieg (2016). Small area estimation with state-space common factor models for rotating panels. *Journal of the Royal Statisitcal Society 179*, 763--791.

van den Brakel, J., P. Smith, and S. Compton (2008). Quality procedures for survey transitions-experiments, time series and discontinuities. *Survey Research Methods 2*, 123--141.

Vehtari, A., A. Gelamn, and J. Gabry (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistical Computation 27*, 1413--1432.

Vehtari, A., A. Gelman, and J. Gabry (2015). *loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models*. R package version 0.1.3.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research 11*, 3571--3594.

Watanabe, S. (2013). A widely applicable bayesian information criterion. *Journal of Machine Learning Research 14*, 867--897.

Ybarra, L. and S. Lohr (2008). Small area estimation when auxiliary information is measured with error. *Biometrika 95*(4), 919--931.

You, Y. (2008). Small area estimation using area level models with model checking and applications. *SSC annual meeting, Proceedings of the Survey Methods Section*.

You, Y. and J. Rao (2000). Hierarchical bayes estimation of small area means using multi-level models. *Survey Methodology 26*(2), 173--181.

Appendix A: auxiliary data

variable	description
adm_immigr	percentage of immigrants in population
adm_immigrnw	percentage of non-western immigrants in population
adm_young	percentage of young people (aged 15 to 25)
adm_old	percentage of elderly people (aged over 65)
adm_urban	percentage of people living in urban regions
adm_house	mean house price
adm_benefit	percentage of social benefit claimants

Table 5.1 Auxiliary data from administrative registers. Data are at police district level.

variable	description
pr_propcrim	property crimes
pr_bicycle	bicycle thefts (subset of property crimes)
pr_violcrim	violent crimes
pr_assault	physical assaults (subset of violent crimes)
pr_threat	threats (subset of violent crimes)
pr_traffic	traffic offences
pr_drugs	illicit drug offences
pr_weapon	weapon offences
pr_damage	damage to public and private property
pr_puborder	disturbance of public order

Table 5.2 Auxiliary data from the Police Register of Reported Offences. Figures are reported offences per 100 inhabitants.

variable	description
REG_nuisance	Perceived nuisance in the neighbourhood, estimated with the regular survey
REG_victim	Percentage of people saying that they have been victim to a crime, estimated with the regular survey
REG_funcpol	Opinion on functioning of the police on a 10-point scale, estimated with the regular survey

Table 5.3 Auxiliary data from the regular survey used in the univariate FH models of Table 4.4.

Appendix B: Full conditional distributions for the Gibbs sampler

The Gibbs sampler iteratively samples from the full conditional distributions for each parameter b in $\theta = (\tilde{v}^t, \xi^t, \beta, \tilde{\sigma}_r^2, \tilde{\sigma}_a^2, \rho)^t$ for model (3). It is understood that b can also be a suitable block of parameters in θ. Let $\theta^{(-b)}$ denote the parameter vector where element b is deleted. Within the Gibbs sampler, the b-th parameter values are drawn conditionally on the data $\hat{y}$ and the rest of the parameter vector $\theta^{(-b)}$ (Geman and Geman, 1984; Gelfand and Smith, 1990).

The conditional posterior density for the regression coefficients β equals $p(\beta|\theta^{(-\beta)}, \hat{y}) \propto p(\beta)p(\hat{y}|\theta)$. Given the uninformative prior for β in (8) and the likelihood (15) for $\hat{y}$ it follows that the full conditional distribution for β is given by $p(\beta|\theta^{(-\beta)}, \hat{y}) \simeq N_\beta(E_\beta, V_\beta)$, with $E_\beta = (X^t\Psi^{-1}X)^{-1}X^t\Psi^{-1}[\hat{y} - \Delta_\xi\tilde{v}]$, and $V_\beta = (X^t\Psi^{-1}X)^{-1}$.

The conditional posterior density for the scale parameters ξ equals $p(\xi|\theta^{(-\xi)}, \hat{y}) \propto p(\xi)p(\hat{y}|\theta)$. Given the normal prior for ξ in (9) and the likelihood (15) it follows that $p(\xi|\theta^{(-\xi)}, \hat{y})$ $\propto N_\xi(0_2, I_2)N_{\hat{y}}(X\beta + \Delta_\xi\tilde{v}, \Psi)$ in the case of an unstructured covariance matrix or a diagonal covariance matrix with unequal variance components for the random effects. Denote $W = 1_m \otimes I_2$. For the random domain effects it holds that $v = \Delta_\xi\tilde{v} = Diag(W\xi)\tilde{v} = Diag(\tilde{v})W\xi \equiv \Delta_{\tilde{v}}\xi$. Using standard results on conjugate priors, it follows that the full conditional distribution for ξ is given by $p(\xi|\theta^{(-\xi)}, \hat{y}) \simeq N_\xi(E_\xi, V_\xi)$ with $E_\xi = (\Delta_{\tilde{v}}^t\Psi^{-1}\Delta_{\tilde{v}} + I_2)^{-1}\Delta_{\tilde{v}}^t\Psi^{-1}[\hat{y} - X\beta]$ and $V_\xi = (\Delta_{\tilde{v}}^t\Psi^{-1}\Delta_{\tilde{v}} + I_2)^{-1}$. In the case of a diagonal covariance matrix with equal variance components for the random effects, the full conditional distribution for the scale parameter ξ reduces to $p(\xi|\theta^{(-\xi)}, \hat{y}) \simeq N_\xi(E_\xi, V_\xi)$ with $E_\xi = (\tilde{v}^t\Psi^{-1}\tilde{v} + 1)^{-1}\tilde{v}^t\Psi^{-1}[\hat{y} - X\beta]$ and $V_\xi = (\tilde{v}^t\Psi^{-1}\tilde{v} + 1)^{-1}$.

The conditional distribution for the covariance matrix of the random effects $\tilde{\Sigma}$ equals $p(\tilde{\Sigma}|\theta^{(-\tilde{\Sigma})}, \hat{y}) \propto p(\tilde{\Sigma})p(\tilde{v}|\tilde{\Sigma})$. From (11) it follows that $p(\tilde{v}|\tilde{\Sigma}) \simeq N_v(0_{2m}, I_m \otimes \tilde{\Sigma})$. Three situations are distinguished:

1. The case where $\tilde{\Sigma}$ is a full covariance matrix with an inverse Wishart prior (12). In this case the conditional distribution for $\tilde{\Sigma}$ equals
$$\begin{aligned} p(\tilde{\Sigma}|\theta^{(-\tilde{\Sigma})}, \hat{y}) &\propto \left[|\tilde{\Sigma}|^{-(\nu_v+d+1)/2}e^{-\frac{1}{2}tr(\Phi_v\tilde{\Sigma}^{-1})}\right]\left[|I_m \otimes \tilde{\Sigma}|^{-1/2}e^{-\frac{1}{2}\tilde{v}^t(I_m\otimes\tilde{\Sigma})^{-1}\tilde{v}}\right] \\ &\propto |\tilde{\Sigma}|^{-(\nu_v+d+1+m)/2}e^{-\frac{1}{2}tr[(\Phi_v+\tilde{V}_m^t\tilde{V}_m)\tilde{\Sigma}^{-1}]}. \end{aligned}$$
Here $\tilde{V}_m$ is a $m \times 2$ matrix such that $\tilde{v} = vec(\tilde{V}_m^t)$. In the derivation the following relations are used:
$$\begin{aligned} |A \otimes B| &= |A|^{rank(B)}|B|^{rank(A)}, \\ tr(A^tB) &= vec(A)^tvec(B), \\ vec(ABC) &= (C^t \otimes A)vec(B). \end{aligned}$$
From these relations we have
$$\begin{aligned} \tilde{v}^t(I_m \otimes \tilde{\Sigma}^{-1})\tilde{v} &= vec(\tilde{V}_m^t)^t(I_m \otimes \tilde{\Sigma}^{-1})vec(\tilde{V}_m^t) \\ &= vec(\tilde{V}_m^t)^tvec(\tilde{\Sigma}^{-1}\tilde{V}_m^tI_m) = tr(\tilde{V}_m\tilde{\Sigma}^{-1}\tilde{V}_m^tI_m) \\ &= tr(\tilde{V}_m^t\tilde{V}_m\tilde{\Sigma}^{-1}). \end{aligned}$$

In the case of a full covariance matrix $\tilde{\Sigma}$, it follows that

$$
\begin{aligned}
p(\tilde{\Sigma}|\theta^{(-\tilde{\Sigma})}, \hat{y}) &\simeq Inv - Wish(v_{v1}, \Phi_{v1}),\\
& v_{v1} = v_v + m,\\
& \Phi_{v1} = \Phi_v + \tilde{V}_m^t \tilde{V}_m.
\end{aligned}
$$

Note that Σ is obtained from $\Sigma = Diag(\xi)\tilde{\Sigma}Diag(\xi)$ or alternatively by directly drawing from the inverse Wishart distribution with same degress of freedom and scale matrix $Diag(\xi)(\Phi_v + \tilde{V}_m^t \tilde{V}_m)Diag(\xi) = Diag(\xi)\Phi_v Diag(\xi) + V_m^t V_m$, where $vec(V_m) = \nu$.

2. The case where $\tilde{\Sigma}$ is a diagonal matrix with unequal variance components with independent chi-squared priors (13). In this case the conditional distribution for $\tilde{\Sigma}$ equals

$$
\begin{aligned}
p(\tilde{\sigma}_a^2, \tilde{\sigma}_r^2|\theta^{(-\tilde{\Sigma})}, \hat{y}) &\propto \prod_{q \in a,r} (\tilde{\sigma}_q^2)^{-\frac{v_q}{2}} e^{-\frac{v_q S_q^2}{2\tilde{\sigma}_q^2}} \times |I_m \otimes \tilde{\Sigma}|^{-1/2} e^{-\frac{1}{2}\tilde{\nu}^t (I_m \otimes \tilde{\Sigma})^{-1}\tilde{\nu}}\\
&\propto \prod_{q \in a,r} (\tilde{\sigma}_q^2)^{-\frac{v_q}{2}} e^{-\frac{v_q S_q^2}{2\tilde{\sigma}_q^2}} (\tilde{\sigma}_q^2)^{-\frac{m}{2}} e^{-\frac{1}{2\tilde{\sigma}_q^2}\sum_{i=1}^m \tilde{\nu}_{i,q}^2}.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
p(\tilde{\sigma}_a^2, \tilde{\sigma}_r^2|\theta^{(-\tilde{\Sigma})}, \hat{y}) &\simeq \prod_{q \in a,r} Inv - \chi^2(v_{q1}, S_{q1}^2),\\
& v_{q1} = v_q + m,\\
& S_{q1}^2 = \frac{1}{(v_q + m)}\left(v_q S_q^2 + \sum_{i=1}^m \tilde{\nu}_{i,q}^2\right).
\end{aligned}
$$

The original variance parameters in Σ are obtained from $\Sigma = Diag(\xi)\tilde{\Sigma}Diag(\xi)$ or alternatively by directly drawing independently from the inverse chi-squared distribution with the same degress of freedom and scale parameters

$$
\frac{1}{(v_q + m)}\left(v_q S_q^2 \xi_q^2 + \sum_{i=1}^m \nu_{i,q}^2\right).
$$

3. The case where $\tilde{\Sigma}$ is a diagonal matrix with equal variance components with a chi-squared priors (14), i.e. $\tilde{\Sigma} = \tilde{\sigma}^2 I_2$. In this case the conditional distribution for $\tilde{\Sigma}$ equals

$$
\begin{aligned}
p(\tilde{\sigma}^2|\theta^{(-\tilde{\sigma}^2)}, \hat{y}) &\propto (\tilde{\sigma}_q^2)^{-\frac{v}{2}} e^{-\frac{vS^2}{2\tilde{\sigma}^2}} \times |I_m \otimes \tilde{\sigma}^2 I_2|^{-1/2} e^{-\frac{1}{2}\tilde{\nu}^t (I_m \otimes \tilde{\sigma}^2 I_2)^{-1}\tilde{\nu}}\\
&\propto (\tilde{\sigma}_q^2)^{-\frac{(v+2m)}{2}} e^{-\frac{1}{2\tilde{\sigma}^2}(vS^2 + \tilde{\nu}^t\tilde{\nu})}.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
p(\tilde{\sigma}^2|\theta^{(-\tilde{\sigma}^2)}, \hat{y}) &\simeq Inv - \chi^2(v_1, S_1^2),\\
& v_1 = v + 2m,\\
& S_1^2 = \frac{1}{(v + 2m)}\left(vS^2 + \tilde{\nu}^t\tilde{\nu}\right).
\end{aligned}
$$

The original variance parameter is $\sigma^2 = \tilde{\sigma}^2\xi^2$.

The conditional posterior density for the scale parameters $\tilde{\nu}$ equals $p(\tilde{\nu}|\theta^{(-\tilde{\nu})}, \hat{y}) \propto p(\tilde{\nu})p(\hat{y}|\theta)$. Given the normal prior for $\tilde{\nu}$ in (11) and the likelihood (15) it follows that $p(\tilde{\nu}|\theta^{(-\tilde{\nu})}, \hat{y}) \propto N_{\tilde{\nu}}(0_{2m}, I_m \otimes \tilde{\Sigma})N_{\hat{y}}(X\beta + \Delta_\xi\tilde{\nu}, \Psi)$. This implies $p(\tilde{\nu}|\theta^{(-\tilde{\nu})}, \hat{y}) \simeq N_{\tilde{\nu}}(E_{\tilde{\nu}}, V_{\tilde{\nu}})$, with $E_{\tilde{\nu}} = \left(\Delta_\xi^t\Psi^{-1}\Delta_\xi + I_m \otimes \tilde{\Sigma}^{-1}\right)^{-1} \Delta_\xi^t\Psi^{-1}\left[\hat{y} - \Delta_\xi\tilde{\nu}\right]$ and $V_{\tilde{\nu}} = \left(\Delta_\xi^t\Psi^{-1}\Delta_\xi + I_m \otimes \tilde{\Sigma}^{-1}\right)^{-1}$. Since $\nu = \Delta_\xi\tilde{\nu}$ and $\Sigma = Diag(\xi)\tilde{\Sigma}Diag(\xi)$ it is also possible to obtain ν by drawing form $p(\nu|\theta^{(-\nu)}, \hat{y}) \simeq N_\nu(E_\nu, V_\nu)$ with $E_\nu = \left(\Psi^{-1} + I_m \otimes \tilde{\Sigma}^{-1}\right)^{-1}\Psi^{-1}\left[\hat{y} - \nu\right]$ and $V_\nu = \left(\Psi^{-1} + I_m \otimes \tilde{\Sigma}^{-1}\right)^{-1}$.

Appendix C: Results

C.1: Plots of posterior predictive distributions

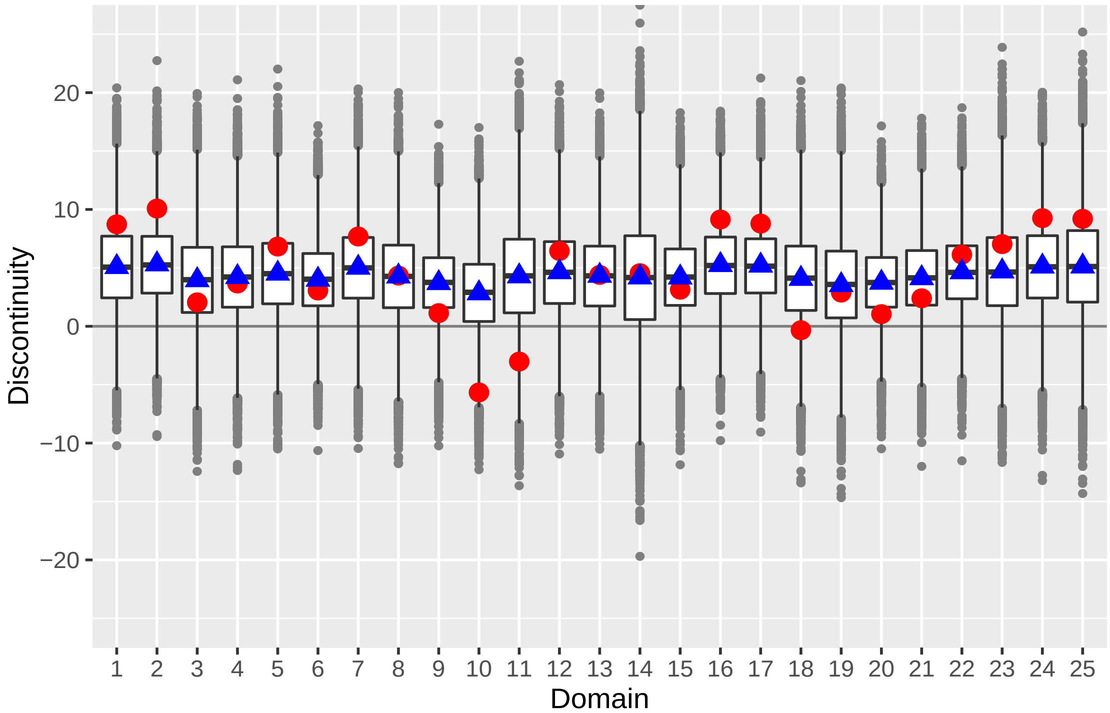


Figure 5.1 Boxplots posterior predicitive distribution domain discontinuities for unsafe selected model (RE_s and FE_eq) with direct estimates (red circles) and HB bivariate FH domain predictions (blue triangles). See Table 4.3 for the selected covariates.

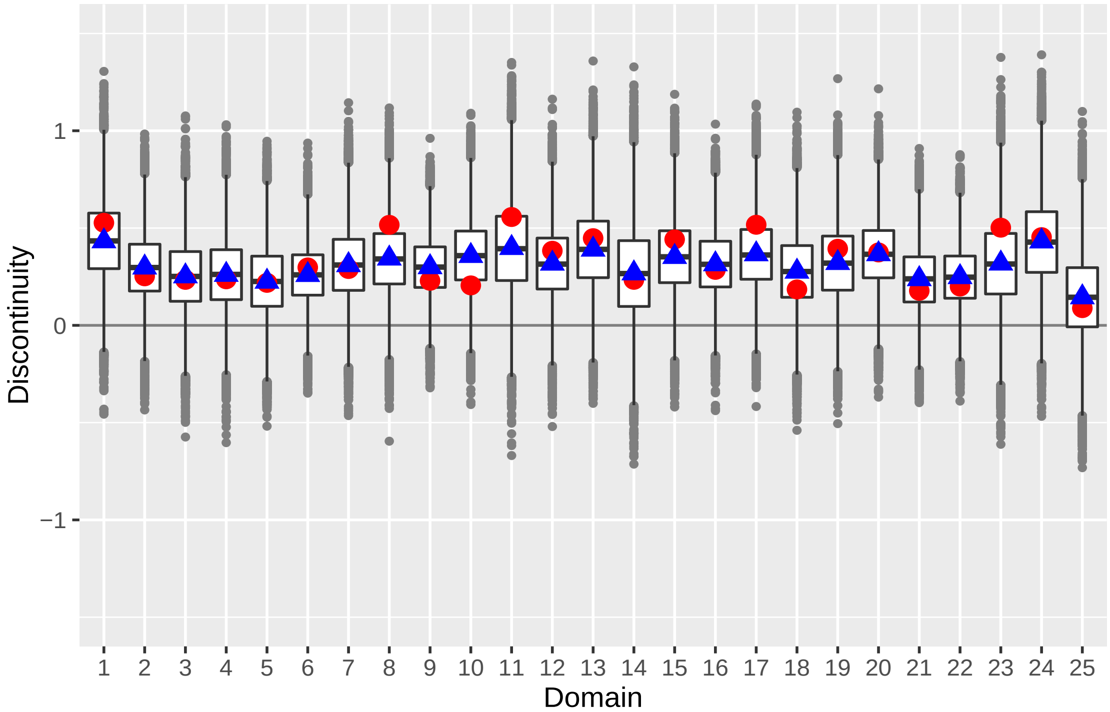


Figure 5.2 Boxplots posterior predicitive distribution domain discontinuities for nuisance under selected model (RE_f and FE_eq) with direct estimates (red circles) and HB bivariate FH domain predictions (blue triangles). See Table 4.3 for the selected covariates.

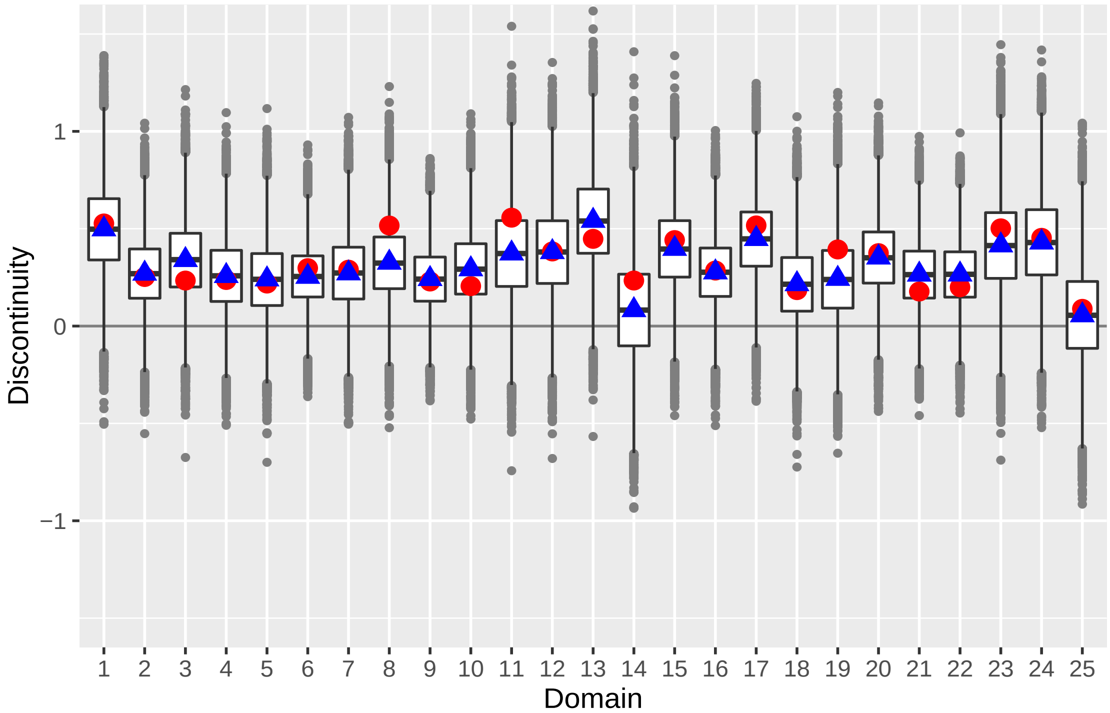


Figure 5.3 Boxplots posterior predicitive distribution domain discontinuities for nuisance under model (RE_d and FE_uq) with direct estimates (red circles) and HB bivariate FH domain predictions (blue triangles). See Table 4.1 for the selected covariates.

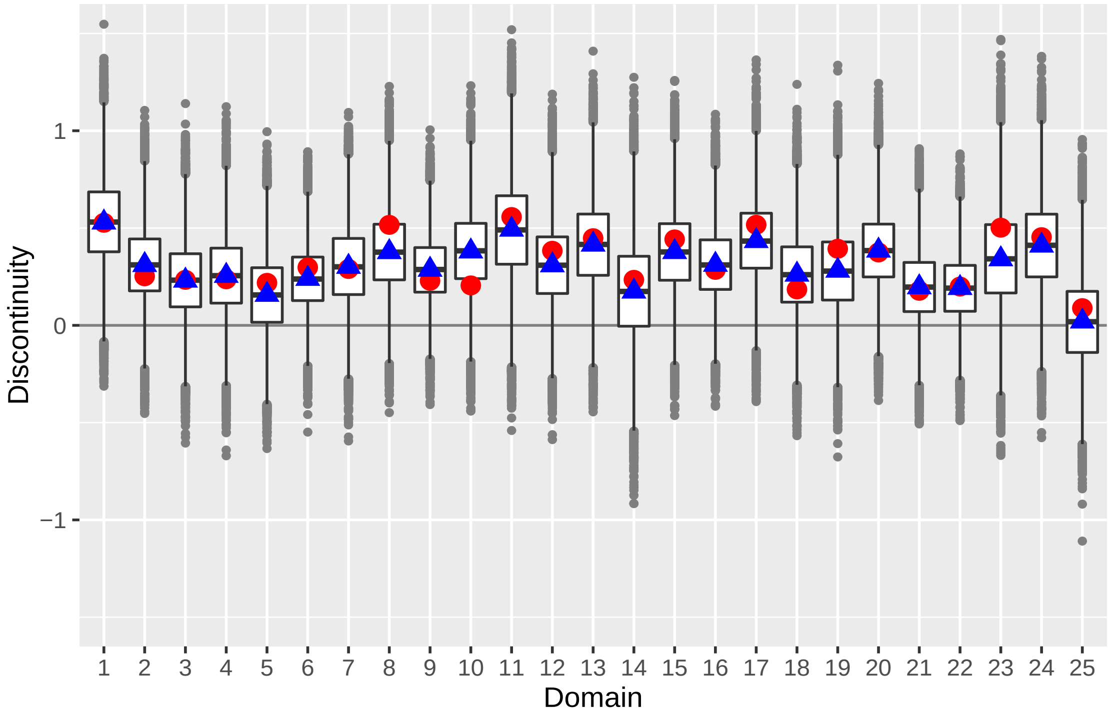


Figure 5.4 **Boxplots posterior predicitive distribution domain discontinuities for nuisance under model (RE_d and FE_eq) with direct estimates (red circles) and HB bivariate FH domain predictions (blue triangles). See Table Table 4.2 for the selected covariates.**

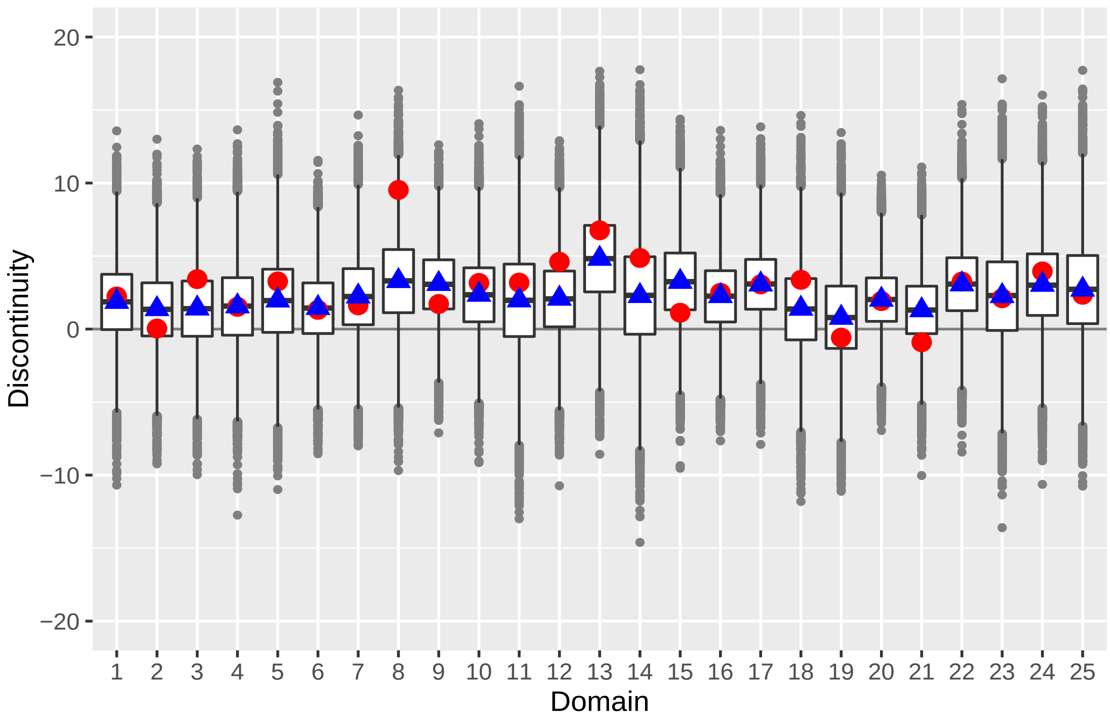


Figure 5.5 **Boxplots posterior predicitive distribution domain discontinuities for propvic under the finally selected model (RE_f and FE_uq) with direct estimates (red circles) and HB bivariate FH domain predictions (blue triangles). See Table 4.3 for the selected covariates.**

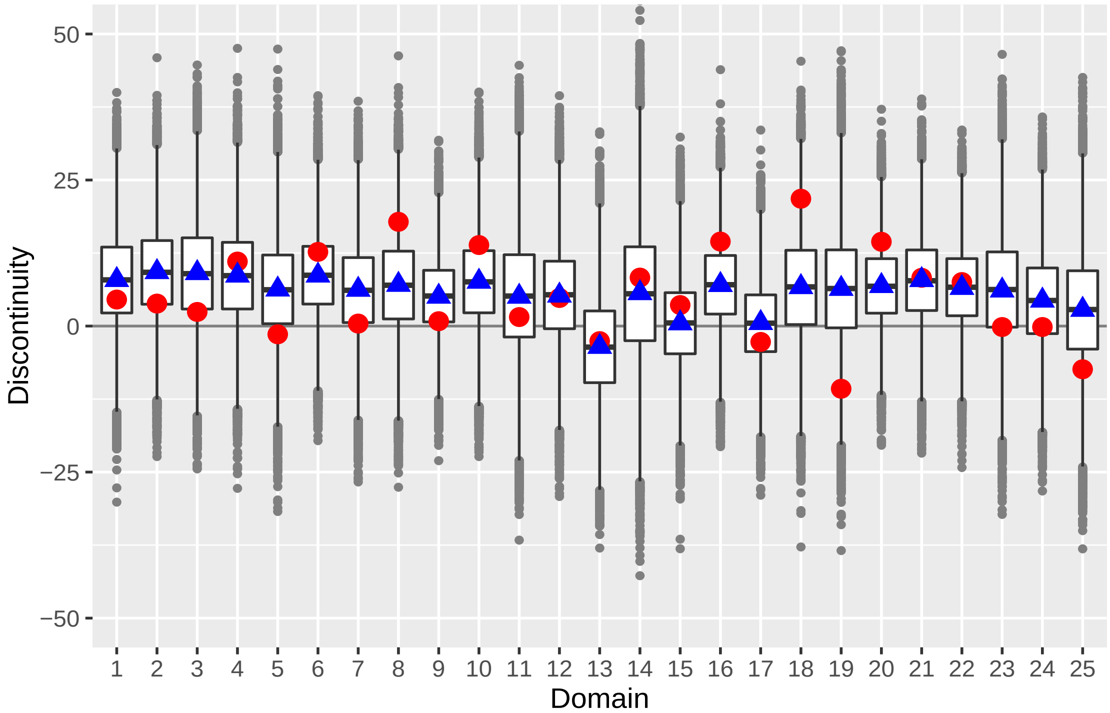


Figure 5.6 Boxplots posterior predicitive distribution domain discontinuities for satispol under the small model met allochtonen (RE_s and FE_uq) with direct estimates (red circles) and HB bivariate FH domain predictions (blue triangles). See Table 4.3 for the selected covariates.

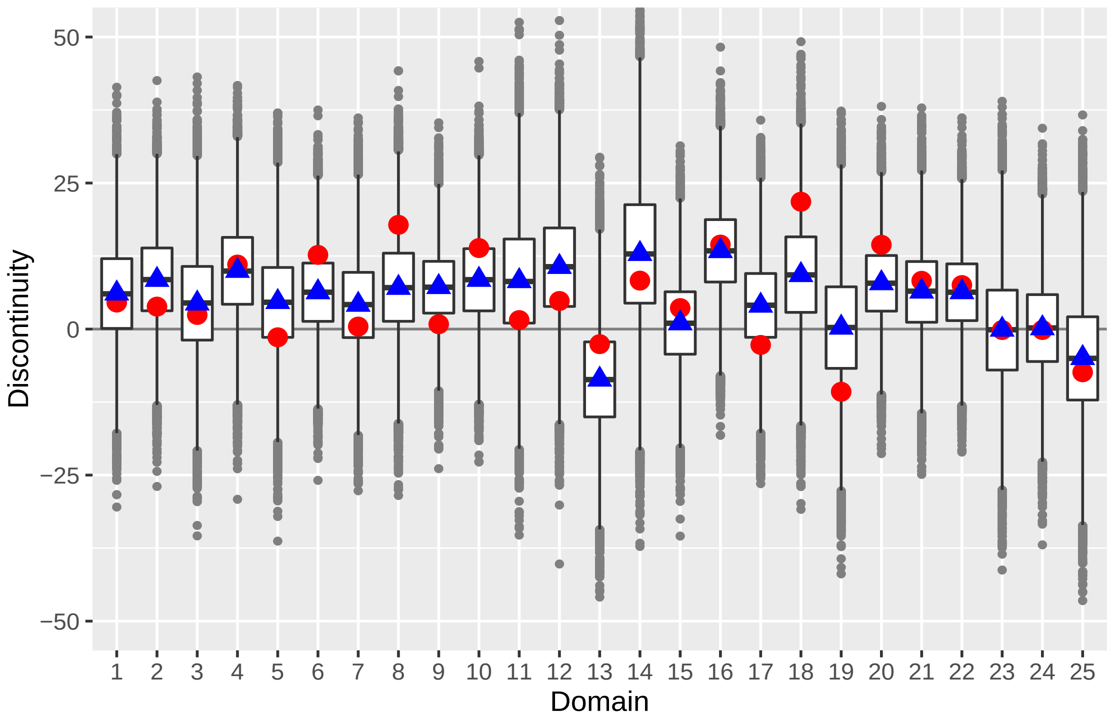


Figure 5.7 Boxplots posterior predicitive distribution domain discontinuities for satispol under the large model (RE_s and FE_uq) with direct estimates (red circles) and HB bivariate FH domain predictions (blue triangles). See Table 4.1 for the selected covariates.

C.2: Estimation results for domain discontinuities

For offtot and unsafe the point estimates for the regular and alternative survey are plotted against each other in Figures 5.8 and 5.10 for the GREG estimator, the univariate FH model and the bivariate FH model. The discontinuities estimated with the GREG estimator, the univariate FH model and the bivariate FH model are plotted in Figures 5.9 and 5.11. Compared with the GREG estimator, there is a clear increase in the linear relationship between the estimates under the regular and alternative survey under the univariate FH model. A further improvement is accomplished with the bivariate FH model. For offtot the plots suggest that the discontinuites are proportional with the level of the offtot. For unsafe this relation is less clear. The bar plots of the disontinuities clearly show that most stable and most precise estimates for discontinuities are obtained with the bivariate FH model. Under this estimator the domain predictions have the same sign and the smallest 95% confidence interval.

Similar plots are made for nuisance in Figures 5.12 and 5.13. The linear relationship between the estimates under the regular and alternative survey under the univariate and bivariate FH model clearly improves compared to the GREG estimator. For this variable the 95% confidence intervals are the smallest for the domain predictions under the univariate FH model. For none of the domains a discontinuity that is significantly different from zero is observed.

The plots for satispol are given in Figures 5.14 and Figure 5.15. Figure 5.14 illustrates overshrinkage for the domain predictions for the alternative survey under the bivariate FH model. This is the result of large standard errors for the GREG estimates for the alternative survey in combination with a bivariate FH model that has a scalar covariance matrix with unequal regression coefficients. The regression coefficient for the alternative survey equals zero while the regression coefficient for the regular survey has a significant negative value. Therefore there is hardly any information to discriminate between domain predictions for the alternative survey. A model with equal regression coefficients doesn't avoid overschrinkage of the domain predictions for the alternative survey and results in very synthetic predictions for the domain discontinuities for reasons explained in Subsection 4.1. A full covariance matrix for random domain effects also doesn't solve the problem, since the covariance estimates takes a value close to zero and the variance components are almost equal to the values obtained with a scalar covariance structure. It appears that the available auxiliary information is not adequate to model satispol under the alternative survey. The model for the univariate FH model indeed only selected are related variable estimated in the regular survey (REG_funcpol), see Table 4.4. For this variable the univariate FH model probably gives the best predictions for the domain discontinuities.

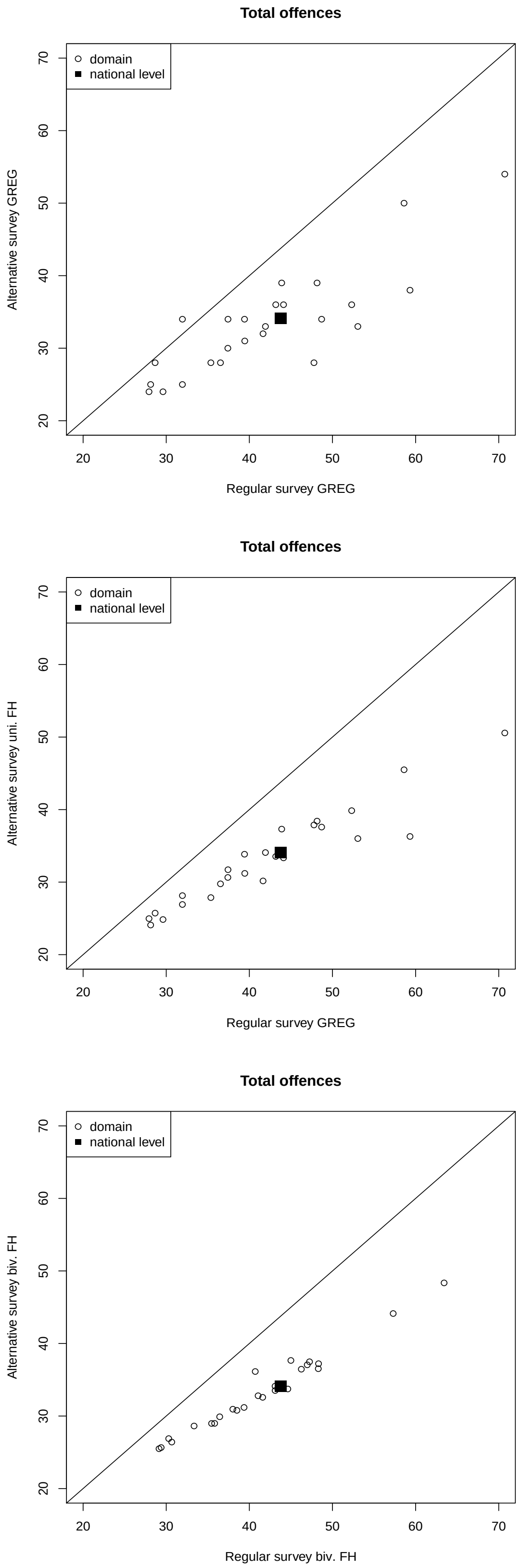


Figure 5.8 Domain estimates offtot alternative survey versus regular survey. Estimates at the national level are based on the GREG estimator.

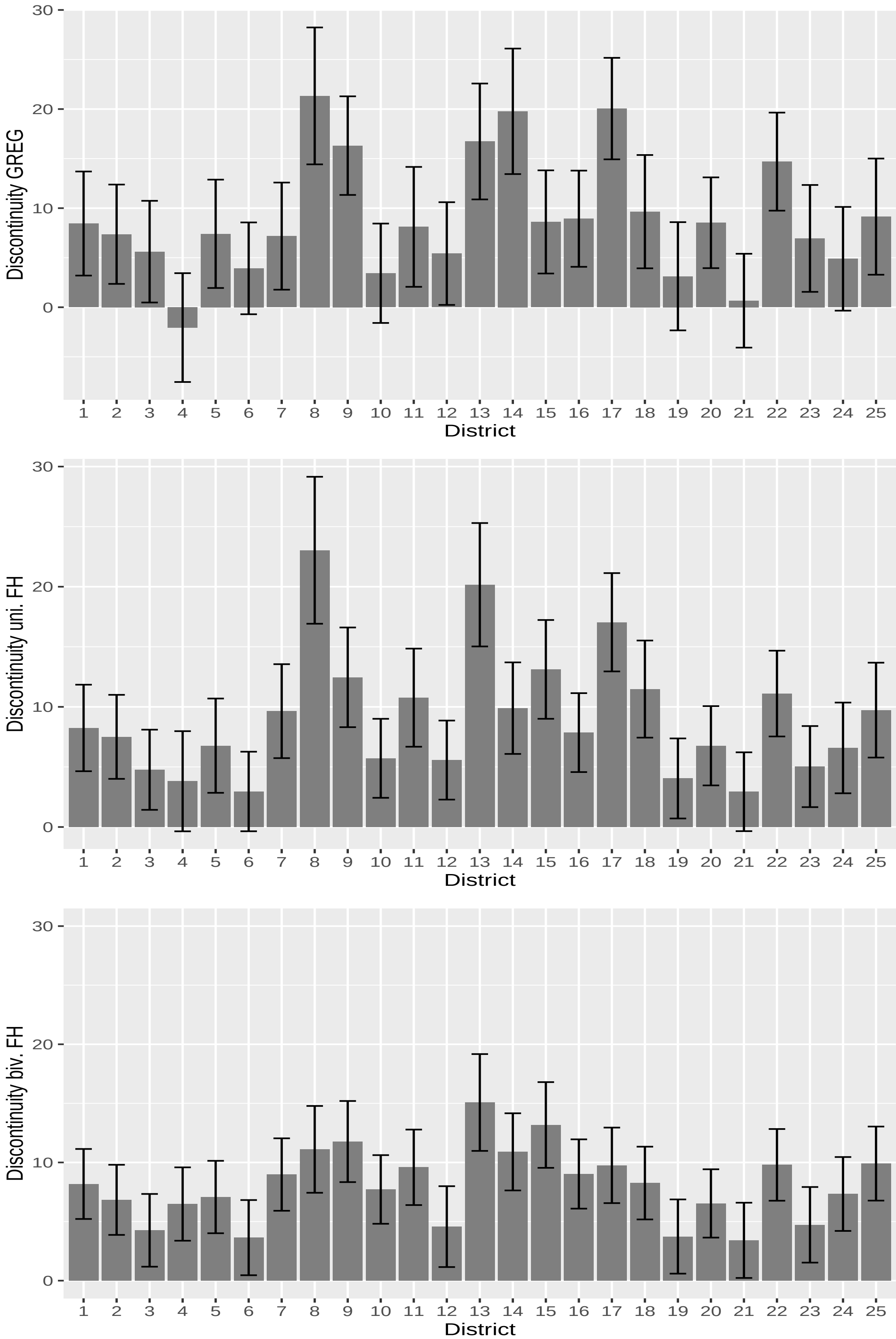


Figure 5.9 Discontinuities offtot based on the GREG estimator (upper panel), univariate FH model (middle panel) and bivariate FH model (lower panel) with a 95% confidence interval.

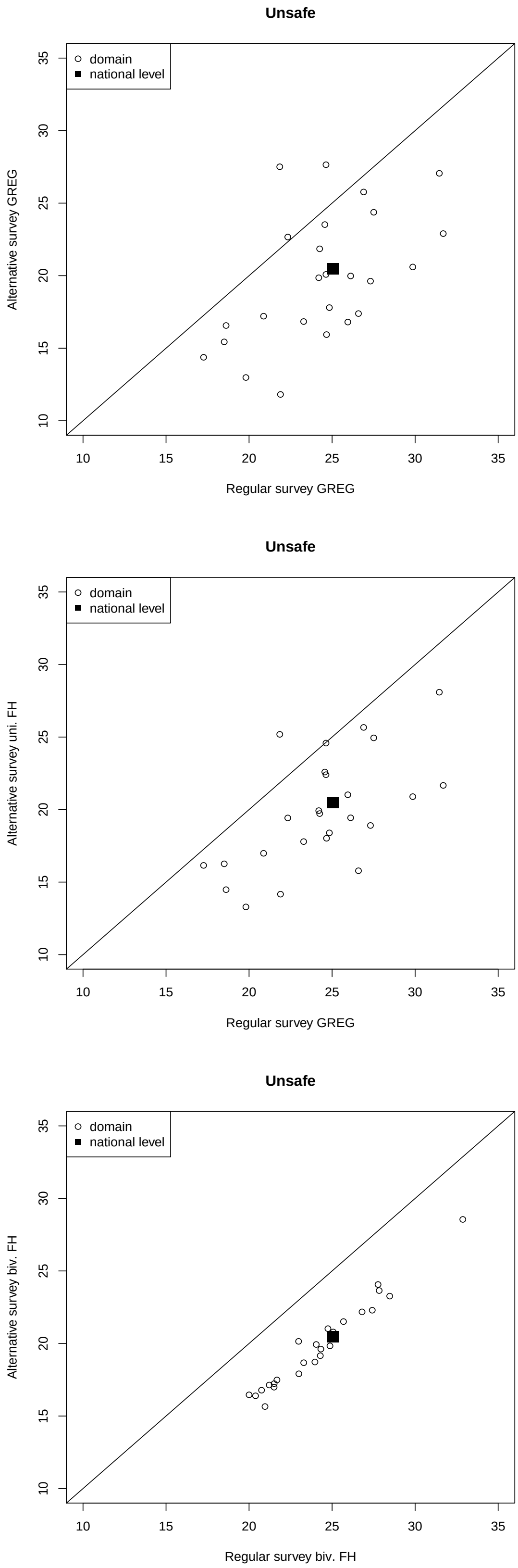


Figure 5.10 Domain estimates unsafe alternative survey versus regular survey. Estimates at the national level are based on the GREG estimator.

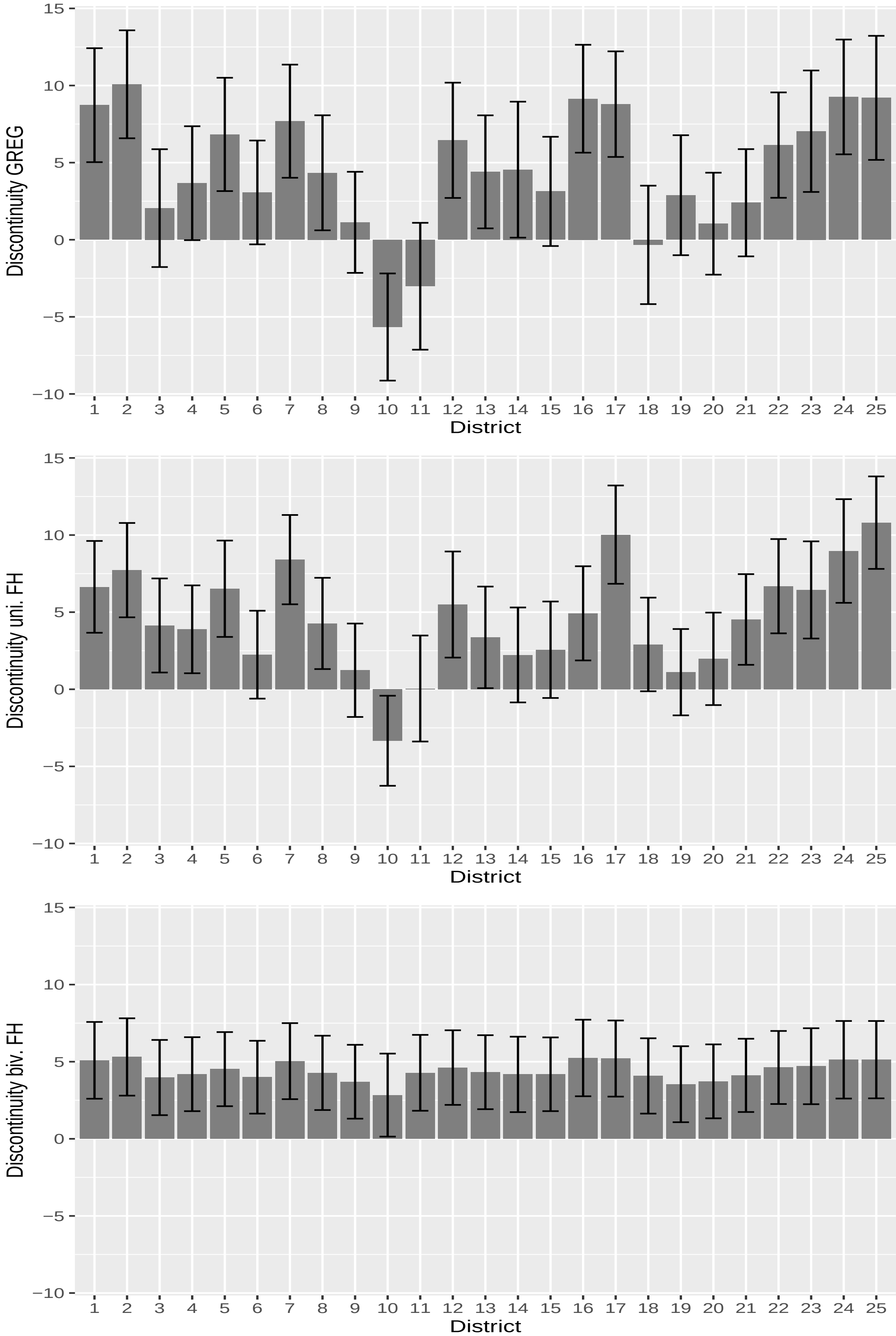


Figure 5.11 Discontinuities unsafe based on the GREG estimator (upper panel), univariate FH model (middle panel) and bivariate FH model (lower panel) with a 95% confidence interval.

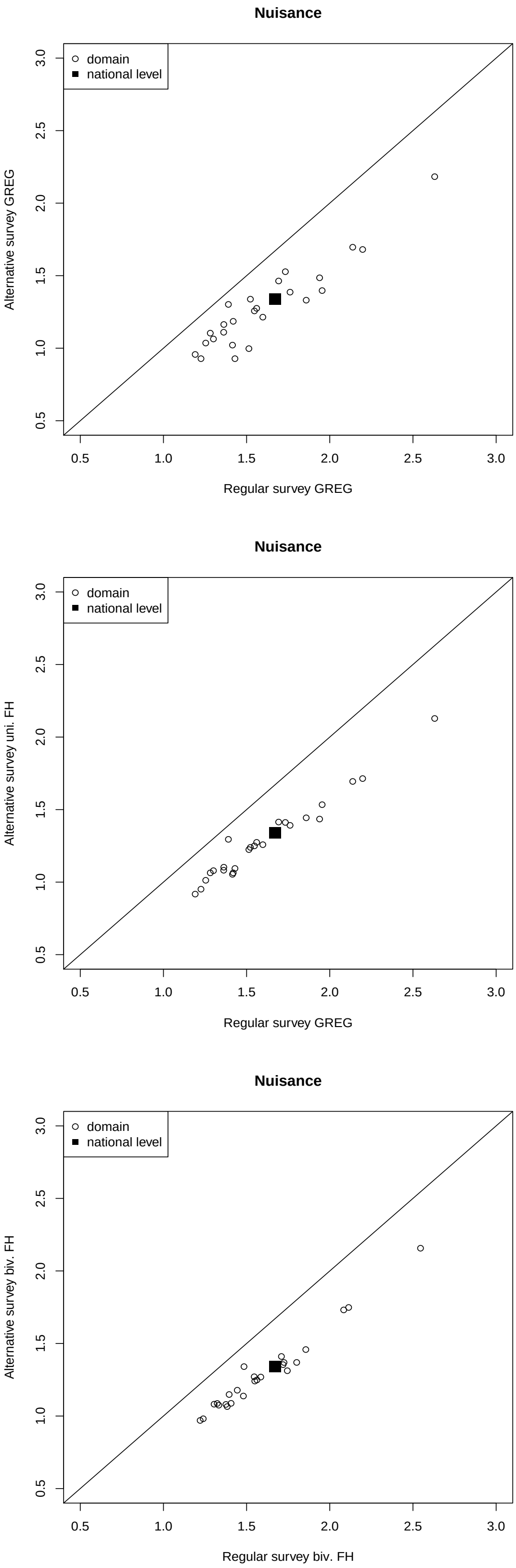


Figure 5.12 Domain estimates nuisance alternative survey versus regular survey. Estimates at the national level are based on the GREG estimator.

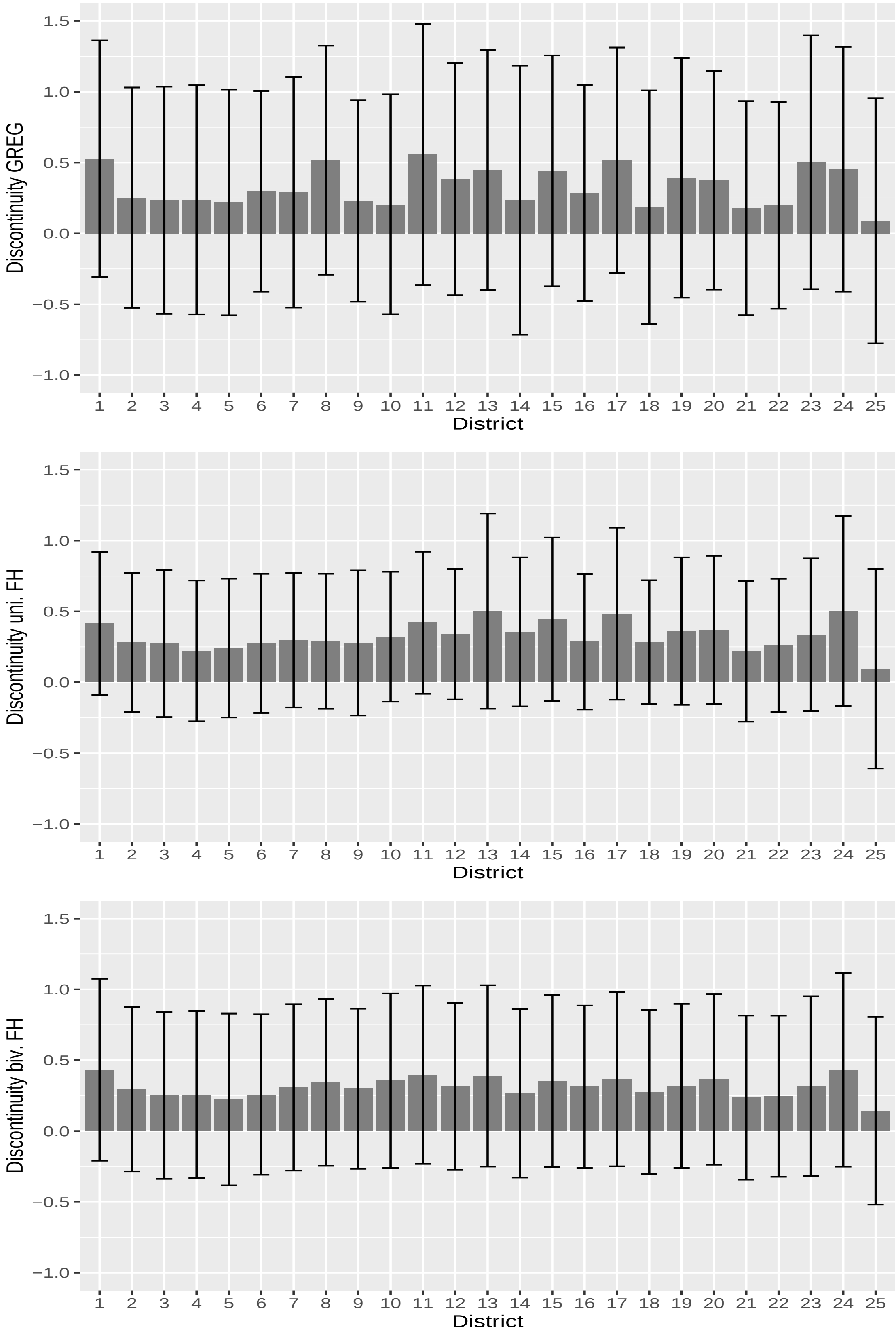
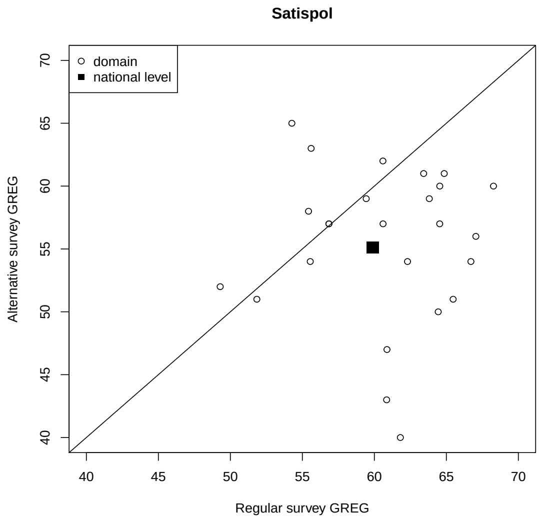
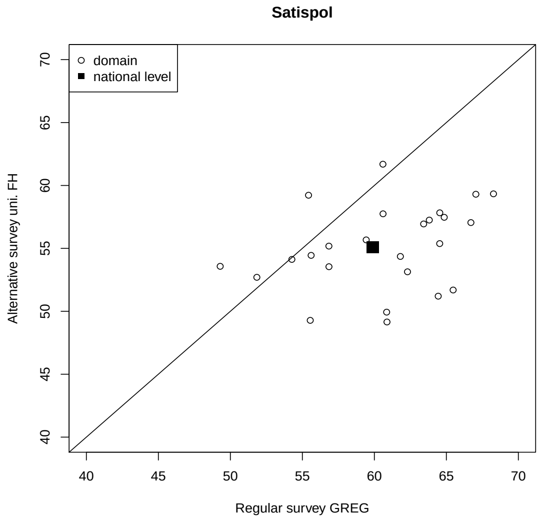


Figure 5.13 Discontinuities nuisance based on the GREG estimator (upper panel), univariate FH model (middle panel) and bivariate FH model (lower panel) with a 95% confidence interval.





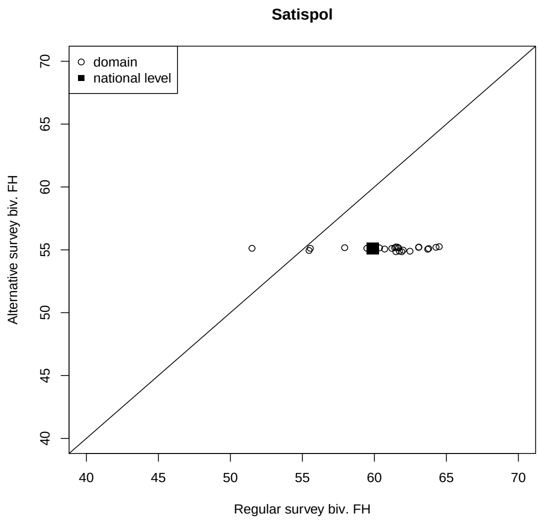


Figure 5.14 Domain estimates satispol alternative survey versus regular survey. Estimates at the national level are based on the GREG estimator.

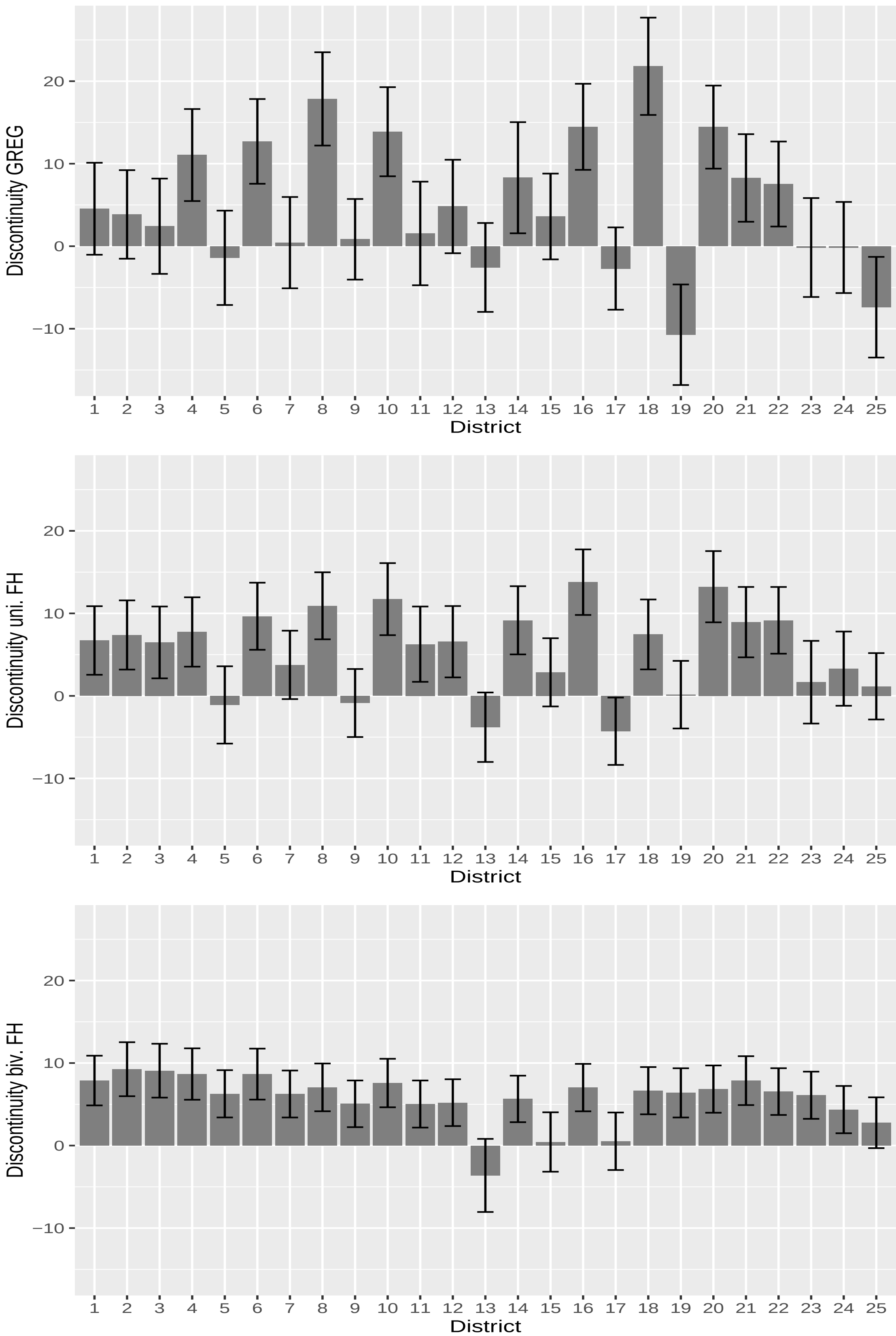


Figure 5.15 Discontinuities satispol based on the GREG estimator (upper panel), univariate FH model (middle panel) and bivariate FH model (lower panel) with a 95% confidence interval.

Colophon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands, Grafimedia

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information