Statistics Netherlands

Discussion Paper

Model-based estimation of discontinuities for small domains in the Dutch Crime Victimization Survey

The views expressed in this paper are those of the author(s) and do not necesarily reflect the policies of Statistics Netherlands

2013 | 11

Jan van den Brakel
Bart Buelens
Harm-Jan Boonstra

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

Where to order
verkoop@cbs.nl
Fax +31 45 570 62 68
ISSN 1572-0314

60083 201311 X-10

Model-based estimation of discontinuities for small domains in the Dutch Crime Victimization Survey

Jan van den Brakel, Bart Buelens and Harm-Jan Boonstra

During redesigns of repeated surveys, the old and the new approach are often conducted in parallel to quantify discontinuities which are initiated by the modifications in the survey process. Due to budget limitations, the sample size allocated to the alternative approach is often considerably smaller, compared to the regular survey that is also used for official publication purposes. In this paper, small area estimation techniques are considered to improve the accuracy of domain estimates obtained under the alternative approach. Besides auxiliary information available from registrations, direct domain estimates available from the regular survey are useful auxiliary variables to construct model-based small area estimators. These methods are applied to a redesign of the Dutch Crime Victimization Survey.

Keywords: Area level models, Empirical Best Linear Unbiased Prediction, Hierarchical Bayes, Small area estimation, Survey Sampling

1 Introduction

Sample surveys conducted by national statistical institutes are generally conducted repeatedly over time with the purpose to construct time series that describe the evolution of finite population variables of interest. Modifications and redesigns of the underlying survey process generally affect the various sources of non-sampling errors in a survey and therefore have a systematic effect on the outcomes of a sample survey. Survey processes of long-standing surveys are generally kept unchanged as long as possible with the purpose to maintain uninterrupted series. It remains, however, inevitable to redesign the survey process from time to time. To avoid the confounding of real developments with systematic effects that are induced by the redesign, it is important to quantify the discontinuities that arise from a redesign of a survey process.

Several possibilities are available to quantify the discontinuity induced by a survey redesign. A straightforward approach is to conduct the old and new design in parallel for some period of time through a large scale field experiment to compare and analyze systematic differences between the direct estimates obtained under both approaches, see e.g. Van den Brakel (2008). Significance and power constraints necessary to establish prespecified contrasts generally require large sample sizes for both the old and the new survey in the parallel run. This is often not tenable due to budget constraints or limited field work capacity available to conduct an alternative approach in parallel with the regular survey. In Van den Brakel and Roels (2010) an intervention analysis approach using state-space models is proposed to quantify discontinuities in series for situations where no overlap or parallel run is available. In the intermediate case, there is only limited budget for a parallel run. In these cases the regular survey, used for official publication purposes, will be conducted in full scale while the alternative approach is conducted with a limited sample size. The design-based approach followed by Van den Brakel (2008) is appropriate to quantify discontinuities at the national level. The sample size allocated to the alternative approach, however, will often be insufficient to produce design-based estimates of discontinuities of adequate precision for subpopulations or domains.

In the case of small domain sample sizes, model-based estimators can be used. These estimators employ sample information observed in other domains through an explicit statistical model and thus increase the effective sample size in the separate domains. In survey methodology, this type of estimation techniques is known as small area estimation, see Rao (2003) for a comprehensive overview. In this paper small area estimation methods are considered to improve the precision of the domain estimates in a parallel run and thus the efficiency of estimating discontinuities, particularly in situations where the sample size allocated to the alternative approach is substantially smaller than the sample size of the regular survey.

In small area estimation, Empirical Best Linear Unbiased Prediction (EBLUP) estimators and Hierarchical Bayesian (HB) estimators are derived from a random effect model. In most cases they can be considered as a weighted average of a direct or design-based estimator based on the sample information observed in the domain for which an estimate is required and a synthetic estimator, where the weights are based on the model error variance and the sampling error variance. In a parallel run, the use of small area prediction can be considered as an intermediate form of the purely design-based approach for embedded experiments considered in Van den Brakel (2008) where inference about the discontinuities is completely based on sample information obtained in the parallel run, and the time series modeling approach considered in Van den Brakel and Roels (2010) where sample information from a parallel run is missing and discontinuities are estimated through an appropriate intervention model.

The extent to which model-based small area predictions result in a reduction of the mean squared error strongly depends on the information available to define models that explain the variation between the domains of interest. In most applications, models borrow strength from auxiliary information which is available from censuses and registrations. In the case of a parallel run, however, direct estimates for the same variables and the same domains are available from the regular survey. For the planned domains, these direct estimates will be sufficiently precise and are potential auxiliary variables to be used in the random effect model that is used to construct an EBLUP or an HB estimator. Since these variables are subject to sampling errors, this is an interesting application of the paper by Ybarra and Lohr (2008), where EBLUP estimators are considered for small areas using the basic area level model with auxiliary information measured with error.

In the Netherlands information on crime victimization, public safety and satisfaction with police performance is obtained by the Dutch Crime Victimization Survey (CVS). This is a long-standing survey that has been redesigned several times. Between 2006 and 2008 this survey was referred to as the Dutch National Safety Monitor (NSM). In 2008 the NSM was redesigned and changed to the Integrated Safety Monitor (ISM). In order to maintain consistent series, the NSM is conducted in parallel on a considerably smaller scale in 2008. With this sample reliable design-based estimates can be obtained at the national level but not for the most important planned domains of this survey. Therefore, model-based estimators are considered in this paper to construct estimates of sufficient accuracy for these domains.

The purpose of this paper is to discuss the use of small area estimation techniques in a parallel run aimed to quantify discontinuities induced by a redesign of the survey process. Questions addressed are which small area estimation techniques can be considered to take advantage of the additional auxiliary information typically available in a parallel run to construct model-based predictions for discontinuities and how to estimate the MSEs of these predictions. EBLUP and HB estimators for the area level model are applied to the parallel run that was part of the redesign of the Dutch CVS. Optimal models are obtained with a step forward model selection procedure based on conditional AIC. The HB approach is used to analyze discontinuities, since this approach gives the most realistic point and MSE estimates. The MSE of the HB predictions is estimated with the posterior variance of the HB estimator. To calculate variances of the discontinuities, the variance of the general regression estimator that is applied to the regular survey, has to be combined with the posterior variance of the HB estimator applied to the parallel run. To treat the uncertainty of both point estimates in a similar way, a design-based estimator for the MSE of the HB predictions is derived as an alternative. The performance of the two analytic MSEs for the HB estimator and the variances of the discontinuities are studied with a bootstrap procedure.

The paper starts with a description of the redesign of the Dutch CVS in section 2. Section 3 describes the different small area estimation procedures that can be applied to the data obtained in the parallel run of the redesign. Analytic expressions for the variance of the discontinuities are derived. The results of this application to the CVS data including the results of the bootstrap are presented in section 4. Section 5 concludes the paper with a discussion.

2 Redesign of the Crime Victimization Survey

The NSM is an annual survey conducted in the first quarter from 2006 through 2008 at full scale, with about 19,000 respondents. It is designed to provide sufficiently precise direct estimates at the national level and at the level of police districts, a subdivision of the Netherlands in 25 regions. In 2008 the NSM was redesigned and changed to the ISM, which is from that moment on conducted in the fourth quarter of each year with a sample size that is equal to that of the NSM. Besides the field work period, the questionnaire and the data collection modes changed during this redesign.

The NSM and the ISM are based on a stratified sample design of persons aged 15 years or older residing in the Netherlands, where the 25 police districts are used as the stratification variable. The sample is equally divided over the strata resulting in a target response of a minimum of 750 respondents in each stratum. Selection probabilities are chosen such that within each stratum a self-weighted sample is drawn. Since police districts have unequal population sizes, inclusion probabilities vary between police districts. This results in an allocation which is optimal for estimating parameters at the level of police districts. Under both survey designs, the generalized regression (GREG) estimator (Särndal et al., 1992) is used to estimate population parameters at the national level and for police districts.

In order to maintain consistent series, the NSM is conducted in parallel to the ISM with a size of about 6,000 respondents in 2008. The sample is designed to provide direct estimates with sufficient precision at the national level only. The selection probabilities are chosen such that a self-weighted sample design is obtained. This results in a proportional allocation of the sample over the strata, which is optimal for estimating parameters at the national level but suboptimal for estimating parameters at police district level. Users of the CVS, nevertheless, expect that discontinuities are also quantified at the level of the 25 police districts. This requires sufficiently precise estimates for the variables of interest based on the ISM and the NSM for the 25 police districts. Therefore, model-based estimators are considered to construct estimates of sufficient accuracy for these domains under the NSM.

This parallel run enables separating the total effect of the survey redesign and real developments of the variables of interest. Quantifying the separate effects of the modifications in the questionnaire and the data collection mode requires a randomized factorial experiment.

3 Methods

3.1 Auxiliary information

In small area estimation, two types of models are commonly used. The first one is the basic area level model, also known as the Fay-Herriot model (Fay and Herriot, 1979), where the input data for the model are the direct estimates for the domains. The second one is the nested error regression model of Battese et al. (1988), which is often referred to as the basic unit level model. In this model the input data are the observations obtained from the sampling units. Through these models, other relevant information can be used to improve the estimation of small domain parameters. In the case of one cross-sectional survey auxiliary information is generally available from registrations or censuses. An important source of auxiliary data in this study is the Police Register of Reported Offences (PRRO). This information is available at an aggregated level per police district. In addition, demographic information available or derived from the municipal administrations is used.

Long-standing surveys, like the Dutch CVS, are conducted repeatedly in time. In these cases there is also sample information available from preceding periods. One way to combine sample information from other domains and preceding periods is to allow for random domain and random time effects in a linear mixed model and apply an EBLUP estimator. Rao and Yu (1994) extended the area level model with an AR(1) model to combine cross-sectional data with information observed in preceding periods. In EURAREA (2004) linear mixed models that allow for spatial and temporal autocorrelation in the random terms are proposed for area and unit level models. A different approach is followed by Pfeffermann and Burck (1990) and Pfeffermann and Bleuer (1993). They combine time series data with cross-sectional data by modeling the correlation between the parameters of the separate domains in a multivariate structural time series model. Pfeffermann and Burck (1990) show how the Kalman filter recursions under particular state-space models can be restructured, like the EBLUP estimators, as a weighted average of a design-based estimator and a synthetic regression type estimator based on information observed in preceding sample surveys and other small domains.

If the redesign of the survey is accompanied by a parallel run, then there is a third source of auxiliary information. Indeed, the regular survey is conducted in parallel to the alternative approach to measure discontinuities due to the planned survey redesign. The regular survey is conducted with full sample size, since it is used for official publication purposes. Therefore sufficiently precise direct estimates are available at least for the planned domains, which are specified in the design stage of the sample survey. In situations were the sample allocated to the alternative approach in a parallel run is substantially smaller compared to the sample size allocated to the regular survey, the sample size will be too small to obtain sufficiently precise direct estimates even for the planned domains. In such cases EBLUP or HB estimators can be used as an alternative using the direct estimates for the same variables and the same domains from the regular design as auxiliary variables in the random effect model. Ybarra and Lohr (2008) consider EBLUP estimators for small areas using the basic area level model with auxiliary information which is measured with error. As an alternative the domain estimates of the regular and alternative approach can be modeled simultaneously in a bivariate random effect model. Fay (1987) and Datta et al. (1991), proposed a multivariate version of the area level model. This approach treats the available sample information in a more symmetric way. The correlation between the two domain parameters can result in more efficient small area predictions, particular for the alternative approach but also for the regular approach. This method is not applied in this paper since official statistics of the CVS are based on the GREG estimator. A bivariate area level model would alter the point estimates of the regular survey as well as the estimated discontinuities based thereon.

3.2 EBLUP for the area level model

Auxiliary information from other sample surveys is typically available at the level of domains. Also the auxiliary information from the police registration (PRRO) is available at the domain level only. Therefore the basic area level model (Fay and Herriot, 1979) is considered in this application. Another advantage of the area level model is that the complexity of the sample design is taken into account, since the dependent variables of the model are the design-based estimates derived from the probability sample and available auxiliary information used in the weighting model of the GREG estimator.

Let $\hat{\theta}_i$ denote the direct estimates of the target variables θ_i for the domains $i = 1, \ldots, m$. In this application $\hat{\theta}_i$ is the GREG estimator obtained under the alternative approach, i.e. the NSM. In the case of the area level model, the direct domain estimates are modeled with a measurement error model, i.e. $\hat{\theta}_i = \theta_i + e_i$, where e_i denotes the sampling error with design variance ψ_i. Furthermore, the unknown domain parameter is modeled with available covariates for the $i-$th domain, i.e. $\theta_i = z_i^t\beta + v_i$, with z_i a K-vector with the covariates $z_{i,k}$ for domain i, β the corresponding K-vector with fixed effects and v_i the random area effects with variance σ_v^2. For each variable a separate univariate model is assumed. Combining both components gives rise to the basic area level model, originally proposed by Fay and Herriot (1979):

$$\hat{\theta}_i = z_i^t\beta + v_i + e_i, \tag{1}$$

with model assumptions

$$v_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_v^2) \quad \text{and} \quad e_i \overset{\text{ind}}{\sim} \mathcal{N}(0, \psi_i). \tag{2}$$

Furthermore, it is assumed that v_i and e_i are independent and that ψ_i is known. Model (1) is a linear mixed model and estimation proceeds using Best Linear Unbiased Prediction (BLUP), Rao (2003). In the case of auxiliary information that is observed without sampling error, the BLUP estimator for θ_i based on model (1) is given by

$$\tilde{\theta}_i^{BLUP} = \gamma_i \hat{\theta}_i + (1 - \gamma_i) z_i^t \hat{\beta}, \tag{3}$$

$$\hat{\beta} = \left(\sum_{i=1}^{m} \gamma_i z_i z_i^t \right)^{-1} \sum_{i=1}^{m} \gamma_i z_i \hat{\theta}_i, \tag{4}$$

$$\gamma_i = \frac{\sigma_v^2}{\sigma_v^2 + \psi_i}, \tag{5}$$

$$V_i^{BLUP} = \gamma_i \psi_i + (1 - \gamma_i)^2 \sigma_v^2 z_i^t \left(\sum_{i=1}^{m} \gamma_i z_i z_i^t \right)^{-1} z_i. \tag{6}$$

The BLUP estimator can be viewed as the weighted average of the direct estimator $\hat{\theta}_i$ and the synthetic estimator $z_i^t \hat{\beta}$, where the weights are given in terms of their measures of uncertainty. Generally σ_v^2 is unknown. Several methods are available in the literature to estimate σ_v^2, see Rao (2003). In this paper Restricted Maximum Likelihood (REML) is used to estimate σ_v^2, see Rao (2003) section 6.2.4. Inserting the REML estimator for σ_v^2 into (5), results in the Empirical BLUP (EBLUP):

$$\tilde{\theta}_i^{EBLUP} = \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i) z_i^t \hat{\beta}, \tag{7}$$

where $\hat{\gamma}_i$ is obtained by (5) where σ_v^2 is replaced by the REML estimator $\hat{\sigma}_v^2$. An estimator for the MSE of the EBLUP estimator that accounts for the variability of estimating σ_v^2 requires an additional term in (6) that is given by Rao (2003), section 7.1.5, which is used in this paper.

Estimates for the design variances ψ_i are available from the GREG estimator but are used in the EBLUP estimator as if the true design variances are known, which is a standard assumption in small area estimation. Therefore it is important to provide reliable estimates for ψ_i, since the weights γ_i directly depend on them. The stability of the estimates for ψ_i can be improved using the following ANOVA-type pooled variance estimator

$$\psi_i = \frac{1 - f_i}{n_i} S_p^2,$$

$$S_p^2 = \frac{1}{n - m} \sum_{i=1}^{m} (n_i - 1) S_{i;GREG}^2,$$

with f_i the sample fraction in domain i, n_i the sample size in domain i, $n = \sum_i^m n_i$ and $S_{i;GREG}^2$ the estimated population variance of the GREG residuals. The sensitivity of an EBLUP estimator derived from the basic area level model for the variability due to estimating the sampling error variance is studied in Bell (2008). An estimator for the

MSE of an EBLUP estimator for area level models that accounts for the variability of estimating the design variance is proposed by Wang and Fuller (2003).

3.3 Hierarchical Bayesian approach for the area level model

Frequently applied methods to estimate the variance of random area effects, σ_v^2, are the Fay-Herriot moment estimator, maximum likelihood and restricted maximum likelihood estimation. A weakness of these methods is that in some situations the estimated model variance tends to zero, see e.g. Bell (1999) and Rao (2003). Zero estimates can occur when the number of areas is small, resulting in imprecise variance estimates. Zero estimates can also occur when the between area variation controlled for the covariates is small, for example in the case of strong auxiliary information. This is the case in the present application of the NSM. A zero or a significantly underestimated model variance leads to undesirable situations. In these cases the EBLUP estimator (7) gives too much weight to the synthetic regression part and too little weight to direct estimates, even in domains with larger sample sizes, and results in less plausible domain predictions. Furthermore there is a large risk of MSE underestimation because the area effects are estimated at zero and the variation between areas is associated with the variation in the auxiliary variables, which is in most cases not realistic.

These problems can be avoided with the Hierarchical Bayesian (HB) approach, Rao (2003), section 10.3. The basic area level model is expressed as an HB model for the case that σ_v^2 is known by (1), (2) and a flat prior on β. Subsequently the HB estimator for θ_i and its MSE are obtained as the posterior mean and variance of θ_i. The posterior density of θ_i given $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_m)^t$ and σ_v^2, i.e. $p(\theta_i|\hat{\theta}, \sigma_v^2)$, is normal with the mean equal to the BLUP estimator of θ_i and variance equal to the MSE of this BLUP estimator. As a result the HB estimator of θ_i given σ_v^2, say $\tilde{\theta}_i^{HB|\sigma_v^2}$, is given by (3) and the posterior variance of θ_i, say $V_i^{HB|\sigma_v^2}$, is given by (6).

If σ_v^2 is unknown, this uncertainty is taken into account by assuming a flat prior on σ_v^2. Other choices for a non-informative prior on σ_v^2 are possible. The choice of the prior is particularly essential for applications with a very small number of domains. In Gelman (2006) the use of a flat prior instead of the more common inverse-gamma is recommended as a less informative prior. The HB estimates for the domain parameters and its MSEs are obtained by averaging $p(\theta_i|\hat{\theta}, \sigma_v^2)$ over the posterior density of σ_v^2, i.e. $p(\theta_i|\hat{\theta}) = \int_0^\infty p(\theta_i|\hat{\theta}, \sigma_v^2)p(\sigma_v^2|\hat{\theta})\mathrm{d}\sigma_v^2$. As a result the HB estimator for θ_i is given by

$$\tilde{\theta}_i^{HB} = E_{\sigma_v^2}\left(\tilde{\theta}_i^{HB|\sigma_v^2}\right) = \int_0^\infty \tilde{\theta}_i^{HB|\sigma_v^2} p(\sigma_v^2|\hat{\theta})\mathrm{d}\sigma_v^2. \tag{8}$$

The posterior variance of θ_i is given by

$$\begin{aligned} V_i^{HB} &= E_{\sigma_v^2}\left(V_i^{HB|\sigma_v^2}\right) + V_{\sigma_v^2}\left(\tilde{\theta}_i^{HB|\sigma_v^2}\right) \\ &= \int_0^\infty \left(V_i^{HB|\sigma_v^2} + (\tilde{\theta}_i^{HB|\sigma_v^2} - \tilde{\theta}_i^{HB})^2\right) p(\sigma_v^2|\hat{\theta})\mathrm{d}\sigma_v^2 \end{aligned} \tag{9}$$

By Bayes' rule, the posterior density for σ_v^2 is $p(\sigma_v^2|\hat{\theta}) \propto p(\sigma_v^2)p(\hat{\theta}|\sigma_v^2)$. Since a flat prior on σ_v^2 is assumed, it follows that the posterior $p(\sigma_v^2|\hat{\theta})$ as a function of σ_v^2 is proportional to the marginal likelihood

$$p(\hat{\theta}|\sigma_v^2) \propto \left(\prod_{i=1}^{m}(\sigma_v^2+\psi_i)^{-\frac{1}{2}}\right)\left[\det\left(\sum_{i=1}^{m}\frac{z_i z_i^t}{\sigma_v^2+\psi_i}\right)\right]^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\frac{(\hat{\theta}_i - z_i^t\hat{\beta})^2}{\sigma_v^2+\psi_i}\right). \quad (10)$$

The determinant factor in this expression arises from the Gaussian integral over β as $p(\hat{\theta}|\sigma_v^2) = \int \mathrm{d}\beta p(\hat{\theta}|\beta, \sigma_v^2)p(\beta|\sigma_v^2)$. Expressions (8) and (9) are evaluated with $2m$ separate one-dimensional numerical integrations. This approach follows the standard convention that the design variances ψ_i are known. You and Chapman (2006) developed an HB estimator that accounts for the uncertainty of estimating the sampling error variances.

The HB estimate for σ_v^2 is obtained as the posterior mean

$$\hat{\sigma}_{vHB}^2 = \frac{\int_0^\infty \sigma_v^2 p(\hat{\theta}|\sigma_v^2)\mathrm{d}\sigma_v^2}{\int_0^\infty p(\hat{\theta}|\sigma_v^2)\mathrm{d}\sigma_v^2}. \quad (11)$$

Expression (11) is evaluated with two separate one-dimensional numerical integrations. Since this estimator is always unique and positive, the HB approach results in more plausible parameter and MSE estimates for the domain variables than the EBLUP estimator, Rao (2003), section 10.3.

3.4 Auxiliary information with sampling error

An important part of the available auxiliary information in a parallel run is based on a sample survey, as explained in section 3.1, resulting in auxiliary variables observed with sampling errors. Ybarra and Lohr (2008) developed the following measurement error EBLUP estimator under the basic area level model to account for sampling variability in the auxiliary variables

$$\tilde{\theta}_i^{ME} = \hat{\gamma}_i\hat{\theta}_i + (1-\hat{\gamma}_i)\hat{z}_i^t\hat{\beta}, \quad (12)$$

$$\hat{\gamma}_i = \frac{\hat{\sigma}_v^2 + \hat{\beta}^t C_i\hat{\beta}}{\hat{\sigma}_v^2 + \hat{\beta}^t C_i\hat{\beta} + \psi_i}, \quad (13)$$

with $\hat{z}_i$ a design-based estimator for the auxiliary variables z_i and C_i the covariance matrix of $\hat{z}_i$. They use an iterative least squares procedure to account for the variability in $\hat{z}_i$ for estimating β, and an estimator for σ_v^2 that is not inflated with the sampling error of the auxiliary variables. This is particularly important for situations where the C_is vary across the domains. Finally they propose MSE estimators that account for the variability of estimating β, σ_v^2 and z_i and follow the standard convention that C_i and ψ_i are known.

In the present application the sampling error in the auxiliary variables is approximately constant over the domains, since the allocation of the sample in the regular survey of the NSM and the ISM is designed such that about 750 respondents are obtained within each police district, as explained in section 2. If prediction is the purpose, it follows from Fuller (1987), p. 75, and Buonaccorsi (1995) that the measurement error in the auxiliary variables can be ignored in the estimator for β if the variance in the auxiliary information is the same for each unit. If the estimated auxiliary information is treated in the EBLUP or HB approach as if it is observed without error, then the sampling error in the auxiliary variables will inflate the estimate for the model variance σ_v^2, resulting in a larger weight $\hat{\gamma}_i$ for the direct estimate in the EBLUP or HB estimate as well as an increase of the estimated MSE. Since the sampling error in the auxiliary information is constant over the domains, there is no need to avoid the inflation of $\hat{\sigma}_v^2$ with the variance of the sampling error and correct the weights $\hat{\gamma}_i$ in (13) for the sampling variance separately for each domain. Moreover, Ybarra and Lohr (2008) concluded from a simulation study that in the case of constant sampling errors, the standard EBLUP estimator (7) performs better than the measurement error EBLUP estimator (12). Therefore the estimated auxiliary variables are substituted for their unknown values in the EBLUP and the HB estimator and treated as if observed without error in this application. The main advantage of this approach is that the standard HB estimator can be used to avoid the problems with zero estimates for the model variance.

3.5 Model selection

The covariates for the models are selected from a set of suitable auxiliary variables through a step forward variable selection procedure. The conditional Akaike Information Criterion (cAIC) is used as a comparison measure to select the most suitable models. The cAIC is proposed by Vaida and Blanchard (2005) for mixed models where the focus is on prediction at the level of clusters or areas. It is defined as cAIC $= -2\mathcal{L}+2p$, where $\mathcal{L}$ is the conditional log-likelihood and p a penalty based on a measure for the model complexity. In the case of a fixed effect model, p is the number of model parameters. The random part of a mixed model also contributes to the number of model degrees of freedom p with a value between 0 in the case of no domain effects (i.e. $\hat{\sigma}_v^2 = 0$) and the total number of domains m in the case of fixed domain effects (i.e. $\hat{\sigma}_v^2 \to \infty$). In the cAIC, p is the effective degrees of freedom of the mixed model and is defined as the trace of the hat matrix H, which maps the observed data to the fitted values, i.e. $\hat{y} = Hy$, see Hodges and Sargent (2001). In the case of the area level model, the conditional likelihood is given by

$$\begin{aligned}
\mathcal{L} &= \log p(\hat{\theta}|\hat{\beta},\hat{v}) = \log\left(\prod_{i=1}^{m}\frac{1}{\sqrt{2\pi\psi_i}}\exp\left(-\frac{1}{2}\frac{(\hat{\theta}_i - z_i^t\hat{\beta} - \hat{v}_i)^2}{\psi_i}\right)\right) \\
&= -\frac{m}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{m}\log(\psi_i) - \frac{1}{2}\sum_{i=1}^{m}\frac{\psi_i}{(\hat{\sigma}_{vHB}^2+\psi_i)^2}(\hat{\theta}_i - z_i^t\hat{\beta})^2,
\end{aligned}$$

where

$$\hat{v}_i = \frac{\hat{\sigma}_{vHB}^2}{(\hat{\sigma}_{vHB}^2+\psi_i)}(\hat{\theta}_i - z_i^t\hat{\beta}).$$

In this application $\hat{\sigma}_{vHB}^2$ is used for model selection. Note, however, that if $\sigma_v^2 << \psi_i$, then the choice between the REML estimate ($\hat{\sigma}_v^2$) and the HB estimate ($\hat{\sigma}_{vHB}^2$) hardly affects the conditional likelihood.

3.6 Analyzing discontinuities

Discontinuities can be calculated as the contrast between the direct estimate obtained with the regular survey and the EBLUP or HB estimate obtained under the alternative approach, and are defined as $\hat{\Delta}_i = \hat{y}_i^{reg} - \tilde{\theta}_i^a$, where $\hat{y}_i^{reg}$ denotes the GREG estimate obtained with the regular survey and $\tilde{\theta}_i^a$ the EBLUP estimate ($a = EBLUP$) or the HB estimate ($a = HB$) for the same parameter obtained under the alternative survey. The significance of this contrast can be tested with a standard t-statistic, but requires an expression for the variance of $\hat{\Delta}_i$. Since sample estimates from the regular survey are also used as auxiliary variables in the models for $\tilde{\theta}_i^a$, a non-zero covariance term arises in the expression for the variance of the contrast, i.e.

$$\mathrm{Var}(\hat{\Delta}_i) = \mathrm{Var}(\hat{y}_i^{reg}) + \mathrm{MSE}(\tilde{\theta}_i^a) - 2\mathrm{Cov}(\hat{y}_i^{reg}, \tilde{\theta}_i^a). \tag{14}$$

Expressions for $\mathrm{MSE}(\tilde{\theta}_i^a)$ are discussed in the preceding subsections. Expressions for the design variance, $\mathrm{Var}(\hat{y}_i^{reg})$, and design-based estimators $\widehat{\mathrm{Var}}(\hat{y}_i^{reg})$ can be found in Särndal et al. (1992).

If the fixed part in the multilevel model $\theta_i = \hat{z}_i^t \beta + v_i$ contains auxiliary variables that are estimated from the regular survey, then $\tilde{\theta}_i^a$ and $\hat{y}_i^{reg}$ will be correlated. Both $\hat{z}_i$ and

$$\hat{\beta} \;\; = \;\; \left(\sum_{i=1}^{m} \gamma_i \hat{z}_i \hat{z}_i^t \right)^{-1} \sum_{i=1}^{m} \gamma_i \hat{z}_i \hat{\theta}_i,$$

contain auxiliary variables that are estimated from the regular survey that are correlated with $\hat{y}_i^{reg}$. In addition, the fixed part $\hat{z}_i^t \hat{\beta}$ is nonlinear, since both components contain survey estimates from the regular survey. To account for the variability of $\hat{\beta}$ in $\hat{z}_i^t \hat{\beta}$, a Taylor approximation for $\hat{z}_i^t \hat{\beta}$ is derived. Let $\hat{\beta} = \hat{T}^{-1} \hat{t}$, with $\hat{T} = \sum_{i=1}^{m} \hat{\gamma}_i \hat{z}_i \hat{z}_i^t$ and $\hat{t} = \sum_{i=1}^{m} \hat{\gamma}_i \hat{z}_i \hat{\theta}_i$. In the appendix it is shown that a Taylor approximation for $\hat{z}_i^t \hat{\beta}$ around its true population values z_i and θ_i, $i = 1, \ldots, m$, and truncated at the first order terms, equals

$$\hat{z}_i \hat{\beta} \;\; \approx \;\; z_i^t \hat{\beta}_0 + \sum_{j=1}^{m} B_{i,j} (\hat{z}_j - z_j) + \sum_{j=1}^{m} C_{i,j} (\hat{\theta}_j - \theta_j), \tag{15}$$

with $B_{i,j}$ a K vector defined by

$$B_{i,j} = (\delta_{i,j} - \hat{\gamma}_j z_i^t T^{-1} z_j) \hat{\beta}_0^t + \hat{\gamma}_j (\theta_j - z_j^t \hat{\beta}_0) z_i^t T^{-1} \tag{16}$$

and $C_{i,j}$ a scalar, defined by

$$C_{i,j} = z_i^t T^{-1} z_j \hat{\gamma}_j. \tag{17}$$

Here $\delta_{i,j} = 1$ if $j = i$ and $\delta_{i,j} = 0$ if $j \neq i$, $\hat{\beta}_0$ is the GLS estimator for β in its real points z_i and θ_i, $i = 1, \ldots, m$, and $T = \sum_{i=1}^{m} \hat{\gamma}_i z_i z_i^t$.

Since domains coincide with the strata of the sampling design, only the auxiliary variables obtained form the regular survey from domain i are correlated with $\hat{y}_i^{reg}$. The direct estimates $\hat{\theta}_i$ are also uncorrelated with $\hat{y}_i^{reg}$, since they are based on two independent samples. As a result, the design covariance between $\tilde{\theta}_i^a$ and $\hat{y}_i^{reg}$ can be approximated with

$$\begin{aligned}\mathrm{Cov}(\hat{y}_i^{reg},\tilde{\theta}_i^a) &= \mathrm{Cov}(\hat{y}_i^{reg},\hat{\gamma}_i\hat{\theta}_i+(1-\hat{\gamma}_i)\hat{z}_i^t\hat{\beta}) = (1-\hat{\gamma}_i)\mathrm{Cov}(\hat{y}_i^{reg},\hat{z}_i^t\hat{\beta}) \\ &\approx (1-\hat{\gamma}_i)[(1-\hat{\gamma}_i z_i^t T^{-1}z_i)\hat{\beta}_0^t+\hat{\gamma}_i(\theta_i-\hat{\beta}_0^t z_i)z_i^t T^{-1}]\mathrm{Cov}(\hat{y}_i^{reg},\hat{z}_i),\end{aligned} \tag{18}$$

where $\mathrm{Cov}(\hat{y}_i^{reg},\hat{z}_i)$ denotes a K vector containing the covariances between $\hat{y}_i^{reg}$ and $\hat{z}_i$. For auxiliary variables that are obtained from other sources than the regular survey, the covariances in $\mathrm{Cov}(\hat{y}_i^{reg},\hat{z}_i)$ are zero. If one of the $\hat{z}_{i,k}=\hat{y}_i^{reg}$, then the corresponding component in $\mathrm{Cov}(\hat{y}_i^{reg},\hat{z}_i)$ simplifies to $\mathrm{Var}(\hat{y}_i^{reg})$. An estimator for (18) is obtained by replacing the unknown population values for their survey estimates and is given by

$$\widehat{\mathrm{Cov}}(\hat{y}_i^{reg},\tilde{\theta}_i^a)=(1-\hat{\gamma}_i)[(1-\hat{\gamma}_i\hat{z}_i^t\hat{T}^{-1}\hat{z}_i)\hat{\beta}^t+\hat{\gamma}_i(\hat{\theta}_i-\hat{\beta}^t\hat{z}_i)\hat{z}_i^t\hat{T}^{-1}]\widehat{\mathrm{Cov}}(\hat{y}_i^{reg},\hat{z}_i), \tag{19}$$

where $\widehat{\mathrm{Cov}}(\hat{y}_i^{reg},\hat{z}_i)$ is a K vector containing the estimates for the design covariances between $\hat{y}_i^{reg}$ and $\hat{z}_i$. For an expression and estimator of the design covariance $\mathrm{Cov}(\hat{y}_i^{reg},\hat{z}_{i,k})$, see Särndal et al. (1992), section 5.9.

The $\mathrm{Cov}(\hat{y}_i^{reg},\tilde{\theta}_i^a)$ depends on the weight $\hat{\gamma}_i$ that is attached to the direct estimator $\hat{\theta}_i$ in $\tilde{\theta}_i^a$. If $\tilde{\theta}_i^a$ is completely based on the direct estimate $\hat{\theta}_i$, then $\hat{\gamma}_i=1$ and (19) is equal to zero, since $\hat{\Delta}_i$ is completely based on the information of two separated samples in this situation. If $\tilde{\theta}_i^a$ is completely based on the synthetic part $z_i^t\hat{\beta}$, then $\hat{\gamma}_i=0$ and (19) reduces to $\hat{\beta}^t\widehat{\mathrm{Cov}}(\hat{y}_i^{reg},\hat{z}_i)$.

In (14), the MSE of $\tilde{\theta}_i^a$ is approximated from a model-based perspective, i.e. the posterior variance of the HB estimator or the MSE of the EBLUP, while the covariance between $\tilde{\theta}_i^a$ and $\hat{y}_i^{reg}$ is treated from a design-based perspective. A more symmetrical alternative is obtained if the model-based approximation for the MSE of $\tilde{\theta}_i^a$ is replaced by a design-based approximation. From (15) it follows that a Taylor approximation for $\tilde{\theta}_i^a$ is given by

$$\tilde{\theta}_i^a \approx \hat{\gamma}_i\hat{\theta}_i+(1-\hat{\gamma}_i)[z_i^t\hat{\beta}_0+\sum_{j=1}^m B_{i,j}(\hat{z}_j-z_j)+\sum_{j=1}^m C_{i,j}(\hat{\theta}_j-\theta_j)]. \tag{20}$$

An expression for the design variance of $\tilde{\theta}_i$ follows from (20) and is given by

$$\begin{aligned}\mathrm{Var}(\tilde{\theta}_i^a) = \hat{\gamma}_i^2\mathrm{Var}(\hat{\theta}_i)+(1-\hat{\gamma}_i)^2\left[\sum_{j=1}^m B_{i,j}\mathrm{Cov}(\hat{z}_j)B_{i,j}^t+\sum_{j=1}^m C_{i,j}^2\mathrm{Var}(\hat{\theta}_j)\right] \\ +2\hat{\gamma}_i(1-\hat{\gamma}_i)C_{i,i}\mathrm{Var}(\hat{\theta}_i).\end{aligned} \tag{21}$$

Here $\mathrm{Cov}(\hat{z}_j)$ is a $K\times K$ design-covariance matrix of the vector with auxiliary variables. This covariance matrix contains the design variances and covariances for auxiliary variables observed in the regular survey. Elements for corresponding auxiliary variables obtained from other sources are zero. Furthermore, $\mathrm{Var}(\hat{\theta}_i)$ are scalars that contain the design variances of the target variables observed in the parallel run. An estimator for (21) is

$$\widehat{\mathrm{Var}}(\tilde{\theta}_i^a)=\hat{\gamma}_i^2\widehat{\mathrm{Var}}(\hat{\theta}_i)+(1-\hat{\gamma}_i)^2\left[\sum_{j=1}^m \hat{B}_{i,j}\widehat{\mathrm{Cov}}(\hat{z}_j)\hat{B}_{i,j}^t+\sum_{j=1}^m \hat{C}_{i,j}^2\widehat{\mathrm{Var}}(\hat{\theta}_j)\right]$$

$$
\begin{aligned}
& +2\hat{\gamma}_i(1-\hat{\gamma}_i)\hat{C}_{i,i}\widehat{\mathrm{Var}}(\hat{\theta}_i), & (22)\\
\hat{B}_{i,j} &= (\delta_{i,j}-\hat{\gamma}_j\hat{z}_i^t\hat{T}^{-1}\hat{z}_j)\hat{\beta}^t+\hat{\gamma}_j(\hat{\theta}_j-\hat{z}_j^t\hat{\beta})\hat{z}_i^t\hat{T}^{-1},\\
\hat{C}_{i,j} &= \hat{z}_i^t\hat{T}^{-1}\hat{z}_j\hat{\gamma}_j,
\end{aligned}
$$

and $\widehat{\mathrm{Cov}}(\hat{z}_j)$ and $\widehat{\mathrm{Var}}(\hat{\theta}_i)$ are the estimators for the design variances and covariances in $\mathrm{Cov}(\hat{z}_j)$ and $\mathrm{Var}(\hat{\theta}_i)$ respectively.

The posterior variance of $\tilde{\theta}_i^{HB}$ and the MSE of $\tilde{\theta}_i^{EBLUP}$ account for the model uncertainty in $\hat{\beta}$ and $\hat{\sigma}_v^2$. The uncertainty of replacing auxiliary variables by design-based estimates is implicitly accounted for by an increase of the model variance $\hat{\sigma}_v^2$. A simulation confirms that $\hat{\sigma}_v^2$ increases proportionally with the sampling error in the auxiliary variables, see the appendix for details. The design-based approximation (22) ignores the variability of $\hat{\sigma}_v^2$ but explicitly accounts for the sampling error of using estimated auxiliary variables. Under this design-based approach, β is considered as a finite population parameter and (22) accounts for the uncertainty of using sample estimates for θ and z in $\hat{\beta}$ instead of their true population values.

An alternative approach is to estimate (14) by means of a bootstrap. This amounts to repeatedly drawing resamples r from the regular sample and the alternative sample. The resampling procedure must reflect the complexity of the sampling design used to draw the regular and the alternative sample from the finite target population, see e.g. Särndal et al. (1992), chapter 11 or Shao and Tu (1995), chapter 6. This is accomplished with the following non-parametric bootstrap. The ISM and the NSM both apply a stratified simple random sampling design. Let n_h^{reg} denote the size of the regular sample s_h^{reg} and n_h^{alt} the size of the alternative sample s_h^{alt} observed in stratum h. For each replicate, n_h^{reg} sampling units are selected from s_h^{reg} from each stratum h by means of simple random sampling with replacement for the regular survey. In a similar way, n_h^{alt} sampling units are selected from s_h^{reg} from each stratum for the alternative survey. The direct estimate for the target variable is calculated for the r-th resample from the regular survey, say $\hat{y}_i^{reg;r}$ and the alternative survey, say $\hat{\theta}_i^r$. Also the auxiliary variables as far as observed from the regular survey, $\hat{z}_i^r$, are re-estimated for each bootstrap sample. Finally the EBLUP or the HB estimate for the r-th resample from the alternative survey, say $\tilde{\theta}_i^{a;r}$ and the corresponding discontinuity $\hat{\Delta}_i^r = \hat{y}_i^{reg;r} - \tilde{\theta}_i^{a;r}$ are computed. This step is repeated R times and

$$
\hat{S}^2(\hat{\Delta}_i) = \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\Delta}_i^r - \bar{\hat{\Delta}}_i\right)^2, \tag{23}
$$

is an estimator for (14), with $\bar{\hat{\Delta}}_i$ the mean over the R bootstrap replicates $\hat{\Delta}_i^r$.

4 Results

4.1 Data

The NSM survey reports on many different outcome variables. In the present study five key survey variables are considered, see Table 1. Variables sourced from administrative registers are listed in Table 2. Table 3 shows the variables obtained from the Police Register of Reported Offences. Table 4 lists the variables from the ISM and past NSM surveys considered for use as covariates. These include the five target variables under consideration, extended with seven more variables of potential benefit. All auxiliary variables are named with prefixes indicating their source.

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 1. Five key NSM survey variables considered in the present study.

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`	level of urbanization (in 5 categories)
`adm\_house`	mean house price
`adm\_benefit`	percentage of social benefit claimants

Table 2. Auxiliary data from administrative registers. Data are at police district level.

4.2 Model selection

In the present application, cAIC is used for model selection. A cross validation (CV) measure was considered as an alternative, Boonstra et al. (2008). It was found that models determined using cAIC generally have a smaller number of covariates than models determined using CV. Furthermore, the cAIC models are often nested within the larger CV models. Since there are only 25 small areas in this application, the smaller cAIC models are preferred.

The cAIC criterion is used to assess the importance of various sets of covariates. Table 5 lists the cAIC values of optimal models. The value given in brackets is the average

variable	description
`prro\_propcrim`	property crimes
`prro\_bicycle`	bicycle thefts (subset of property crimes)
`prro\_violcrim`	violent crimes
`prro\_assault`	physical assaults (subset of violent crimes)
`prro\_threat`	threats (subset of violent crimes)
`prro\_traffic`	traffic offences
`prro\_drugs`	illicit drug offences
`prro\_weapon`	weapon offences
`prro\_damage`	damage to public and private property
`prro\_puborder`	disturbance of public order

Table 3. Auxiliary data from the Police Register of Reported Offences. Figures are reported offences per 100 inhabitants.

reduction in the coefficient of variation (cv) achieved under these optimal models. The average is taken over police districts, and the reduction reflects the gain of the HB estimates compared to the direct estimates of the NSM. Five sets of covariates are considered, corresponding to the columns in this table. The first and smallest set consists of administrative data only. This set is increasingly expanded by adding PRRO variables, then either ISM variables from the parallel runs, or NSM variables from the previous full scale NSM survey, held in the first quarter of 2008, and finally both ISM and NSM variables.

Models using only administrative data are generally worse than those including PRRO data, which in turn are worse than those also including ISM variables. Models involving only NSM variables from preceding editions of the survey are not as good as models with ISM variables. Including both as potential covariates does not lead to better models than using ISM variables only. The variable `propvict` benefits least from using data from other surveys. In this case PRRO variables seem to be the best predictors, which is not surprising since this variable is closely related to the police reported offenses.

Since the number of domains in this application is limited to 25 police districts, models with more than a few covariates will easily overfit the data. Therefore, principal component analysis is applied to reduce the dimension of the covariate space, see e.g. Hastie et al. (2003). The principal component models are never better than the models using the variables directly. A possible reason why principal components are not advantageous is that the set of covariates contains many variables with little predictive power. These variables do contribute to the principal components, but they do not get selected in the model selection process when using variables directly. In this way, the latter models make use of the covariates more efficiently. Alternative approaches such as partial least squares or canonical component analysis could prove useful, but are not investigated in the present research.

variable	description
surv_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
surv_unsafe	percentage of people feeling unsafe at times
surv_propvict	percentage of people saying to have been victim to property crime in the last 12 months
surv_violvict	percentage of people saying to have been victim to violent crime in the last 12 months
surv_satispol	percentage of people satisfied with police at their last contact (if contact in last 12 months)
surv_degrad	perceived degradation of the neighbourhood, on a ten point scale
surv_funcpol	opinion on functioning of the police, on a ten point scale
surv_victim	percentage of people saying to have been victim to crime
surv_offtot	number of offenses per 100 inhabitants (derived from victim reports)
surv_propcrim	number of property crimes per 100 inhabitants
surv_violcrim	number of violent crimes per 100 inhabitants
surv_bicycle	number of bicycle thefts per 100 inhabitants

Table 4. Auxiliary data from the parallel ISM surveys, and from the past NSM survey. The first five variables correspond to the NSM survey target variables. The prefix surv is either ism or nsm, reflecting the source of the variable.

variable	admin	+ PRRO	+ ISM	+ NSM	+ ISM + NSM
offtot	155.1 (47%)	152.1 (49%)	148.8 (56%)	153.1 (48%)	148.8 (56%)
unsafe	135.1 (24%)	134.7 (29%)	127.3 (37%)	127.7 (39%)	127.7 (39%)
nuisance	-20.7 (29%)	-22.5 (35%)	-34.8 (51%)	-30.2 (45%)	-34.8 (51%)
satispol	165.5 (50%)	164.4 (50%)	161.5 (55%)	164.2 (49%)	161.5 (55%)
propvict	108.9 (49%)	104.6 (49%)	105.0 (51%)	104.6 (49%)	105.0 (51%)

Table 5. cAIC values for optimal models based on different sets of covariates. Between brackets is the percentage improvement in coefficient of variation of the HB estimates compared to the direct estimates, averaged over districts.

Hence, the models to proceed with in the present application are those associated with the minimal cAIC values in Table 5. The models for `offtot`, `unsafe`, `nuisance` and `satispol` contain one or more variables from the ISM. In the model for `propvict` no ISM variables are selected. Here the PRRO variables play a more dominant role.

The optimal models for the five target variables of the NSM, that are finally selected for small area prediction are based on the original auxiliary variables and are specified in Table 6. No variables from the previous edition of the NSM were selected.

variable	cAIC-based model
`offtot`	`ISM\_victim`
`unsafe`	`ISM\_nuisance, adm\_benefit, prro\_propcrim, prro\_drugs`
`nuisance`	`ISM\_nuisance, adm\_old`
`satispol`	`ISM\_funcpol`
`propvict`	`prro\_propcrim, adm\_old`

Table 6. Optimal models for the five key survey variables. All models also include an intercept (not shown).

4.3 Model diagnostics

Small area estimates can be benchmarked to the national direct estimates. Benchmarking can be conducted for cosmetic purposes, since it restores the required consistency between the national estimates and the underlying domain estimates. Pfeffermann and Burck (1990) and Pfeffermann and Bleuer (1993) proposed benchmarking in the context of small area estimation with state space models as a build in mechanism against model miss-specification. The size of the calibration adjustments that are required can be viewed as a model diagnostic, see Brown et al. (2001). Small area estimates can be benchmarked to the direct estimates at the national level by formulating an appropriate Lagrange function such that the discrepancies between the direct estimate and the small area estimates are distributed proportional to the MSEs of the small area estimates. The benchmarked small area estimates have smaller or equal MSEs than the unadjusted small area estimates. The interpretation of this MSE reduction is that the restriction add additional information to the model that is applied to construct the small area estimates. See e.g. Van den Brakel and Roels (2010) for details and expressions of the Lagrange multiplier approach, including an expression for the MSE of the adjusted small area estimates. An alternative point of view is taken by You et al. (2003). They consider the differences between the original and the benchmarked estimates as a bias, whose squared value is added to the MSE of the original estimate.

The HB small area estimates, obtained with the optimal models described in Table 6, are benchmarked to the national direct estimates. In this application, the adjustments are very small, all less than 0.4% (see Table 7). This confirms the choice of models. Since the calibration adjustments are so small in the present application, no MSE adjustments are made. Table 7 shows the mean squared adjustments, the mean mse, and the percentage of the former to the latter: all but one are less than 0.1%.

variable	calibration (%)	mean bias	mean mse	percentage bias
`offtot`	-0.222%	5.61E-03	5.970	0.096%
`unsafe`	0.048%	9.39E-05	3.170	0.003%
`nuisance`	0.006%	5.63E-09	0.004	0.000%
`satispol`	0.132%	5.42E-03	9.103	0.061%
`propvict`	0.343%	1.23E-03	1.261	0.111%

Table 7. For each variable the percentage calibration adjustment required is shown in the third column. The next columns contain the mean squared calibration adjustment, the mean mse estimate, and the percentage of the former to the latter.

In Figure 2, HB estimates for `nuisance` and `propvict` are plotted against direct estimates. The dashed line is a linear regression line fitted to the data. As is commonly seen, the SAE estimates are smoothed compared to the direct estimates, as indicated by the slope of the regression line being smaller than 1. This can be expected since the direct estimates have larger variances than the HB estimates, while there should not be an overall bias as the HB estimates are calibrated to the direct estimates at national level. The difference between direct and HB estimates is smaller for `nuisance` than for `propvict`.

The residuals, i.e. the differences between the direct and HB estimates, are shown using a QQ-plot in Figure 3. The distributions do not deviate much from normal, not giving indications of model misspecification.

4.4 Small area estimates

The optimal models are used to produce estimates. The cv's of the direct estimates are compared with the HB estimates in Table 8. The values shown are averaged over police districts. For most variables reductions of between 40% and 50% are achieved, with a lower reduction for `unsafe`.

variable	cv direct	cv HB	percentage change
`offtot`	0.180	0.075	-56
`unsafe`	0.153	0.092	-37
`nuisance`	0.108	0.051	-51
`satispol`	0.125	0.055	-55
`propvict`	0.245	0.113	-49

Table 8. Coefficients of variation of the direct and HB estimates, and their percentage change. The coefficients shown are averages taken over all areas.

As discussed in section 3.3, REML estimation of the model variance σ_v^2 can be problematic in instances where the likelihood has its maximum at zero. Table 9 shows the REML and Bayesian estimates of σ_v^2 in the present application. The REML estimates are zero for all variables, which implies that the EBLUP estimates are completely based on the synthetic regression part. This demonstrates the advantage of the HB approach in this application, which always assigns a positive weight to the direct estimates in the HB prediction.

variable	Posterior mean	REML
`offtot`	4.995	0.000
`unsafe`	2.997	0.000
`nuisance`	0.003	0.000
`satispol`	7.725	0.000
`propvict`	0.805	0.000

Table 9. Posterior mean and REML estimates of the model variance σ_v^2.

For two variables, `nuisance` and `propvict`, some results are presented in more detail, as averaging over the areas may conceal important aspects of the applied methods. Figure 4 shows details at the area level for `nuisance` and `propvict`. The cv's of the direct and HB estimates are plotted for each police district, with the districts ordered by increasing population size. Since the NSM employs a proportional allocation over the strata, in this case the districts, the ordering is also according to NSM sample size. On average, the smaller areas benefit more than the larger areas from applying small area models, as their cv's are reduced more than the larger areas' cv. Nevertheless, there are gains too for the largest police districts. The cv's of the HB estimates do not vary a lot among areas, while the cv's of the direct estimates decrease with increasing area size.

4.5 Discontinuities

Discontinuities for the five variables from Table 1 are analyzed at the level of districts. For the ISM, domain estimates are based on the GREG estimator. For the NSM, the HB estimator developed in the preceding section is used. Two analytic variance estimates are compared with the bootstrap estimates. The first analytic estimator for (14) uses $\widehat{\mathrm{Var}}(\hat{y}_i^{reg})$ as an estimator for the variance of the ISM GREG estimates, the posterior variance for the MSE of $\tilde{\theta}_i^{HB}$ and (19) for the covariance between $\tilde{\theta}_i^{HB}$ and $\hat{y}_i^{reg}$. This estimator is referred to as the model-based standard error approximation. The second analytic estimator for (14) uses (22) as a design-based estimator for the MSE of $\tilde{\theta}_i^{HB}$ instead of the posterior variance. This estimator is referred to as the design-based standard error approximation. Bootstrap simulations are based on 2000 resamples. Plots of the bootstrap variance of the discontinuities (23) against the number of replications indicate that the bootstrap variance converges to a stable value after 1000 to 1500 resamples.

First the GREG and HB estimates are compared with their bootstrap estimates. Then results for the discontinuities are presented. Table 10 compares the analytic and bootstrap estimates for the GREG estimates and its standard errors for the five target variables of the ISM and the NSM, averaged over the 25 police districts. The bootstrap approximations are close to the analytic point and standard error estimates. To give an indication of the fluctuation of the differences between the districts, the mean of the absolute difference between analytic and bootstrap estimates over the 25 domains is also calculated and is called the absolute mean difference (AMD). It shows that differences between the NSM are somewhat larger than the ISM, which can be attributed to the smaller sample size of the NSM.

variable	Analytic NSM		Bootstrap NSM		AMD NSM	
	GREG	SE	GREG	SE	GREG	SE
`offtot`	33.28	5.73	33.29	5.50	0.24	0.84
`unsafe`	19.86	2.87	19.85	2.82	0.05	0.21
`nuisance`	1.28	0.13	1.28	0.13	0.00	0.01
`satispol`	55.58	6.88	55.56	6.86	0.29	0.43
`propvict`	9.78	2.19	9.79	2.10	0.04	0.23
variable	Analytic ISM		Bootstrap ISM		AMD ISM	
	GREG	SE	GREG	SE	GREG	SE
`offtot`	42.29	4.73	42.31	4.73	0.09	0.10
`unsafe`	24.38	2.03	24.38	1.98	0.04	0.07
`nuisance`	1.61	0.11	1.61	0.11	0.00	0.00
`satispol`	60.61	4.23	60.62	4.20	0.07	0.08
`propvict`	12.55	1.60	12.57	1.61	0.03	0.04

Table 10. Comparison analytic and bootstrap GREG estimates ISM and NSM averaged over districts.

Table 11 compares the analytic HB estimates, the model-based standard errors (SE1) and the design-based standard errors (SE2) with the bootstrap means of the HB estimates and the bootstrap approximations for the standard errors of the HB estimates averaged over the 25 districts for the NSM. The AMD refers to the average absolute differences between the analytic HB estimates and the bootstrap estimates. Figure 5 compares the design-based, the model-based and the bootstrap standard errors of the HB estimates for `nuisance` and `propvict` for the 25 districts. For `offtot` and `nuisance` the design-based standard errors are on average larger than the model-based standard errors. For the other three variables the model-based standard errors are on average larger. Recall from section 4.1 that `propvict` is the only variable that did not select an auxiliary variable in the optimal model from the ISM. The bootstrap standard errors are always larger than the two analytic approximations.

Table 12 compares the analytic point estimates for the discontinuities, the model-based

variable	Analytic			Bootstrap		AMD		
	HB est.	SE(1)	SE(2)	HB est.	SE	HB est.	SE(1)	SE(2)
`offtot`	33.21	2.43	2.90	33.29	3.13	0.88	0.74	0.39
`unsafe`	19.83	1.76	1.64	19.84	1.92	0.57	0.19	0.29
`nuisance`	1.29	0.06	0.08	1.28	0.08	0.03	0.02	0.01
`satispol`	55.09	3.00	2.54	55.29	3.58	1.14	0.59	1.05
`propvict`	9.85	1.09	0.84	9.88	1.12	0.25	0.11	0.28

Table 11. Comparison HB point and SE estimates with bootstrap results averaged over districts.

standard errors (SE1) and the design-based standard errors (SE2) with the bootstrap means of the discontinuities and the bootstrap approximations for the standard errors of the HB estimates, averaged over the 25 districts. The AMDs specify the average absolute differences between the analytic HB estimates and the bootstrap estimates. Figure 6 compares the design-based, the model-based and the bootstrap standard errors of the discontinuities for `nuisance` and `propvict` for the 25 districts. The results show the same pattern as for the HB estimates; for `offtot` and `nuisance` the design-based standard errors are on average larger, for the other variables the model-based standard errors are larger. The bootstrap standard errors are always larger than the two analytic approximations.

The last two columns of Table 12 contain the GREG estimates for the discontinuities, i.e. the difference between the GREG estimate of the ISM and the GREG estimate of the NSM and its standard error. The standard errors for the discontinuities that are based on the HB estimator are substantially smaller compared to the GREG estimates. The use of the HB estimator for the NSM variables increases the precision of the model predictions for θ_i substantially. Moreover, the use of direct estimates of the target variable and related variables from the regular survey as auxiliary variables in the model for the HB estimator results in a positive correlation between $\hat{y}_i^{reg}$ and $\tilde{\theta}_i^a$, which further reduced (14).

variable	Analytic			Bootstrap		AMD (Analytic)			GREG	
	Disc.	SE(1)	SE(2)	Disc.	SE	Disc.	SE(1)	SE(2)	Disc.	SE
`offtot`	9.08	3.54	3.92	9.02	4.92	0.89	1.38	1.00	9.01	7.69
`unsafe`	4.55	2.54	2.46	4.54	2.69	0.56	0.20	0.25	4.52	3.57
`nuisance`	0.33	0.05*	0.07	0.33	0.11	0.03	0.06	0.03	0.33	0.17
`satispol`	5.52	4.98	4.72	5.33	5.43	1.15	0.45	0.70	5.04	8.21
`propvict`	2.70	1.95	1.84	2.70	1.97	0.25	0.08	0.14	2.78	2.77

Table 12. Analysis results discontinuities averaged over districts. *: For nuisance 2 districts with negative variance estimates for the estimated discontinuity are truncated at zero.

For `nuisance` in two districts negative variance estimates for the discontinuities with the model-based standard error approximation are obtained. These problems do not occur with the design-based standard error approximation. The bootstrap standard errors, obtained with (23) are always positive by definition.

The bootstrap standard errors of the HB predictions and the discontinuities are always larger compared to the two analytic approximations. Recall that the posterior variance of the HB predictions accounts implicitly for the additional uncertainty of using auxiliary information with sampling error through an increase of $\hat{\sigma}_v^2$, see also the simulation in the appendix. The design-based standard error approximation of the HB predictions accounts explicitly for the sampling error in the auxiliary variables.

The main reason that the bootstrap standard errors are larger than the two analytic standard error approximations is an increase of the $\hat{\sigma}_v^2$ in the bootstrap resamples. A second reason why the bootstrap standard errors are larger is that the analytic standard error approximation of the HB estimator ignores the variability of the variance of the direct estimator for the domains. This variability is also captured by the bootstrap standard errors.

The variance estimates of the random effects (σ_v^2) and the cAIC based on the original sample are compared with their median in the bootstrap distributions in Table 13. The median of the REML estimate and the posterior mean of σ_v^2 and the cAIC in the bootstrap distribution are considerably larger compared to their values in the original sample. This is the result of two factors. First, the bootstrap tends to increase the between domain variance of the target variables, particularly in situations were the between domain variance is small compared to the within domain variance. This also applies to a parametric bootstrap, which therefore does not resolve this problem. Second, this is an indication that the data are over-fitted by the model, which can be a result of the small number of domains in combination with the extensive model selection procedure. This can result in a model that perfectly describes the particular structure in the original data of the NSM and ISM, but is on average less optimal for the resamples observed in the bootstrap. If within each bootstrap resample a new optimal model is selected, then different optimal models are selected (results not presented). These are indications that the data are over-fitted by the applied model selection procedure. Over-fitting of the data by the model results in smaller model variances $\hat{\sigma}_v^2$ and smaller weights $\hat{\gamma}_i$ on the direct estimate in the EBLUP or HB estimator. As a result, the contribution of the covariance term (19) in the estimator for (14) becomes too large. This increases the differences between the bootstrap standard errors for the discontinuities and the two analytic approximations.

The bootstrap standard errors of the discontinuities are finally used, since they are the most conservative approximations. They are still smaller compared to the standard errors of the direct estimates. Discontinuities with 95% confidence intervals are plotted for `nuisance` in Figure 7, `propvict` in Figure 8, `unsafe` in Figure 9, `offtot` in Figure 10, and `satispol` in Figure 11. Confidence intervals are based on bootstrap standard errors.

	Results initial sample			Bootstrap median		
variable	Post. mean	REML	cAIC	Post. mean	REML	cAIC
`offtot`	4.995	0.000	148.8	26.180	15.360	167.1
`unsafe`	2.997	0.000	127.7	10.870	6.967	137.6
`nuisance`	0.003	0.000	-34.8	0.019	0.012	-18.06
`satispol`	7.725	0.000	161.5	37.120	19.890	177.6
`propvict`	0.805	0.000	105.0	2.842	1.064	118.4

Table 13. Posterior mean and REML estimates of the model variance σ_v^2, and cAIC.

Estimated discontinuities are used to correct the NSM to the level of the ISM before the change-over using correction methods discussed in Van den Brakel et al. (2008).

4.6 Software

For the computation of the HB estimates, the R package `hbsae` (Boonstra, 2012) has been used. The R function uses one-dimensional deterministic numerical integration of the BLUP expressions over the posterior distribution of σ_v^2 to compute the HB estimates and their MSEs. This way, the full Bayesian approach is followed without relying on more time-consuming Monte Carlo methods. Alternatively, plug-in estimates of σ_v^2, such as its REML estimate or posterior mean can be used. The output of the R function also includes the estimated model parameters and a number of model selection measures including cross-validation and cAIC.

5 Discussion

To quantify the systematic effects induced by a redesign of repeatedly conducted surveys, the old and new approach are often run in parallel for some period. Due to budget constraints, the sample size for the alternative approach is often considerably smaller compared to the regular survey used for official publication purposes. In such cases, small area estimation procedures can be considered to obtain more precise domain estimates.

In the case of a parallel run there are, besides information from registers, sample estimates for the same target variable and related variables observed under the regular approach and generally also in preceding periods available. This information can be used for model building and selection, since strong correlations can be expected between sample estimates for the same variables, observed under two alternative survey designs.

In the SAE literature various approaches are available to incorporate this information in an appropriate estimation procedure. In this paper, an HB estimator for the area level model is developed to estimate discontinuities due to a redesign of the Dutch Crime Victimization Survey. The sample estimates from the regular survey are used as auxiliary information as if observed without sampling error. Ignoring the sampling error is legitimate in this application, since the sampling error of the estimated auxiliary information is approximately equal for the different domains. In this situation the sampling error of the auxiliary information is implicitly reflected by an increase of the variance of the random effects of the area level model. This allows the application of the full HB approach, which guarantees strictly positive estimates for the model variance. REML estimation indeed resulted in zero model variance estimates for all of the variables considered in this application.

The variance of the discontinuities contains the variance of the GREG estimator of the regular survey and the MSE of the HB estimator of the alternative approach. If sample estimates from the regular survey are used as auxiliary information in the area level model of the HB estimator, then the direct estimates from the regular survey will be correlated with the HB estimates for the alternative approach. An approximation for the design covariance between the direct estimator for the regular survey and the HB estimator for the alternative approach is derived. Using the posterior variance as an approximation for the MSE of the HB estimator has the conceptual drawback that the uncertainty of the HB estimator is measured from a model-based perspective, while the covariance between the HB estimator and the GREG estimator is evaluated from a design-based perspective. For one target variable this approach resulted in negative variance estimates for the discontinuities.

As an alternative, a design-based approximation for the MSE of the HB estimator is derived. In this application problems with negative variance estimates for the discontinuities are avoided with this MSE approximation. This design-based MSE approximation does not account for the uncertainty of using an estimator for the variance of the random effects of the area level model. The additional advantage of using the HB estimator in combination with this design-based MSE approximation is that a non-zero weight is attached to the direct estimator in the HB estimator for the domains.

The design-based MSE approximation accounts for the uncertainty of using sample estimates in the estimated regression coefficients of the fixed part of the area level model. It can be considered as a simplified alternative for the method proposed by Ybarra and Lohr (2008) to account for sampling error in the auxiliary variables of the area level model, but only for situations where the sampling error in the auxiliary variables is approximately constant over the domains. For situations where the sampling error in the auxiliary variables varies across the domains, the EBLUP proposed by Ybarra and Lohr (2008) gives the correct weight on the synthetic and the direct estimator in the EBLUP for the domain estimates.

Using sample estimates from the regular survey as well as from preceding periods as auxiliary information in the area level model significantly increases the precision of the HB estimates, compared to models that only use information available from registrations. The variance of the discontinuities is further decreased since selecting sample estimates from the regular survey as auxiliary variables in the area level model, results in strong positive correlations between the HB estimates for the alternative survey and the direct estimates of the regular survey.

There are no systematic differences between the posterior variance and the design-based MSE of the HB estimator. Estimating the MSE of the HB estimator with a bootstrap indicates that the MSE approximation of the HB estimator is an under estimation of the real MSEs. A possible reason is that for several variables the models obtained with the model selection procedure over-fit the data. This suggests further research in alternative model selection procedures that are more robust against selection of models that over-fit the data. There are also indications that the bootstrap tends to over-estimate the MSE of the HB estimator. In this application the bootstrap approximations are used as a conservative approximation for the standard error of the discontinuities, which are still considerably smaller compared to the standard errors of the discontinuities that uses the GREG estimator for both the regular and alternative survey approach.

A topic for further research is to consider a bivariate area level model for the two estimates observed in the regular survey and the parallel run. Under this approach two model-based SAE estimates for the variable under both the alternative and the regular approach are obtained. In this case standard errors of both estimates can mutually take advantage of their correlation. This approach is not followed in this application, since official statistics obtained with the regular survey are based on the GREG estimator and the approach used to approximate discontinuities should be as close as possible to the procedures followed in the regular survey.

References

Battese, G. E., Harter, R. M., and Fuller, W. A. (1988). An error components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association* 83, 28–36.

Bell, W. (2008). Examining sensitivity of small area inferences to uncertainty about sampling error variances. In *Proceedings of the joint statistical meetings 2008, section on survey research methods*, pp. 327-334.

Bell, W. R. (1999). *Accounting for uncertainty about variances in small area estimation.* Technical report, Bulletin of the International Statistical Institute.

Boonstra, H. J. (2012). *hbsae: Hierarchical Bayesian Small Area Estimation.* R package version 1.0.

Boonstra, H. J., Van den Brakel, J., Buelens, B., Krieg, S., and Smeets, M. (2008). Towards small area estimation at Statistics Netherlands. *METRON International Journal of Statistics* LXVI (1), 21–49.

Brown, G., Chambers, R., Heady, P., and Heasman, D. (2001). Evaluation of Small Area

Estimation Methods - An Application to Unemployment estimates from the UK LFS. In *Proceedings of Statistics Canada Symposium 2001*.

Buonaccorsi, J. (1995). Prediction in the presence of measurement error: General discussion and an example predicting defoliation. *Biometrics* 51, 1562–1569.

Datta, G. S., Fay, R. E., and Gosh, M. (1991). Hierarchical and emperical Bayes multivariate analysis in small area estimation. In *Proceedings of Bureau of the Census 1991 Annual Research Conference*, U.S. Bureau of the Census, Washington DC, pp. 63-79.

EURAREA (2004). Project reference volume. deliverable D7.1.4, Technical report, EURAREA consortium.

Fay, R. E. (1987). *Application of multivariate regression to small domain estimation.* in R. Platek, J.N.K. Rao, C.E. Särndal and M.P. Singh (Eds.) Small Area Statistics, John Wiley, New York, pp. 91-102.

Fay, R. E. and Herriot, R. A. (1979). Estimation of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association* 74, 268–277.

Fuller, W. (1987). *Measurement error models.* John Wiley, New York.

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

Hastie, T., Tibshirani, R., and Friedman, J. H. (2003). *The elements of statistical learning.* Springer-Verlag, New York.

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

Pfeffermann, D. and Bleuer, S. R. (1993). Robust Joint Modelling of Labour Force Series of Small Areas. *Survey Methodology* 19.

Pfeffermann, D. and Burck, L. (1990). Robust Small Area Estimation combining Time Series and Cross-sectional Data. *Survey Methodology* 16.

Rao, J. N. K. (2003). *Small Area Estimation.* John Wiley, New York.

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

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

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap.* Springer-Verlag, New York.

Vaida, F. and Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika* 92, 351–370.

Van den Brakel, J. (2008). Design-based analysis of embedded experiments with applications in the Dutch Labour Forcs Survey. *Journal of the Royal Statistical Society, Series A* 171, 581–613.

Van den Brakel, J. and Roels, J. (2010). Intervention analysis with state-space models to estimate discontinuities due to a survey redesign. *Annals of Applied Statistics* 4, 1105–1138.

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

Wang, J. and Fuller, W. (2003). The mean squared error of small area predictors constructed with estimated error variances. *Journal of the American Statistical Association* 98, 716–723.

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

You, Y. and Chapman, B. (2006). Small area estimation using area level models and estimated sampling variances. *Survey Methodology* 32, 97–103.

You, Y., Rao, J., and Dick, P. (2003). Benchmarking Hierarchical Bayes small area estimators in the Canadian census undercoverage estimation. *Statistics in transition* 6, 631–640.

Taylor approximation for small area predictions

In this section a Taylor approximation for the small area predictions based on the area level model, i.e.

$$\begin{aligned}
\tilde{\theta}_i^a &= \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i) \hat{z}_i^t \hat{\beta}, \\
\hat{\beta} &= \left(\sum_{i=1}^m \hat{\gamma}_i \hat{z}_i \hat{z}_i^t \right)^{-1} \sum_{i=1}^m \hat{\gamma}_i \hat{z}_i \hat{\theta}_i,
\end{aligned}$$

is derived. The auxiliary variables $\hat{z}_i$ can contain sampling error if they are based on a sample survey. A Taylor approximation for the synthetic component is obtained by considering $\hat{z}_i^t \hat{\beta}$ as a function in $\hat{z}_j$ and $\hat{\theta}_j$, $j = 1, ..., m$, that can be approximated around its real population values $z = (z_1, \ldots, z_m)$ and $\theta = (\theta_1, \ldots, \theta_m)$:

$$\begin{aligned}
f_i(\hat{z}, \hat{\theta}) \approx f_i(z, \theta) \quad &+ \quad \sum_{j=1}^m \sum_{k=1}^K \left. \frac{\partial f_i(\hat{z}, \hat{\theta})}{\partial \hat{z}_{j,k}} \right|_{(\hat{z}, \hat{\theta}) = (z, \theta)} (\hat{z}_{j,k} - z_{j,k}) \\
&+ \quad \sum_{j=1}^m \left. \frac{\partial f_i(\hat{z}, \hat{\theta})}{\partial \hat{\theta}_j} \right|_{(\hat{z}, \hat{\theta}) = (z, \theta)} (\hat{\theta}_j - \theta_j). \qquad (24)
\end{aligned}$$

Using standard rules from matrix differentiation for $f_i(\hat{z},\hat{\theta}) = \hat{z}_i^t\hat{\beta}$ gives

$$\begin{aligned}\frac{\partial f_i(\hat{z},\hat{\theta})}{\partial \hat{z}_{j,k}} &= \frac{\partial \hat{z}_i^t}{\partial \hat{z}_{j,k}}\hat{\beta} + \hat{z}_i^t\frac{\partial \hat{T}^{-1}}{\partial \hat{z}_{j,k}}\hat{t} + \hat{z}_i^t\hat{T}^{-1}\frac{\partial \hat{t}}{\partial \hat{z}_{j,k}} \\ &= \delta_{i,j}\hat{\beta}^t\lambda_k - \hat{z}_i^t\hat{T}^{-1}\hat{\Lambda}_{j,k}\hat{T}^{-1}\hat{\gamma}_j\hat{t} + \hat{z}_i^t\hat{T}^{-1}\lambda_k\hat{\theta}_j\hat{\gamma}_j, \end{aligned} \tag{25}$$

with λ_k a K vector with the k-th element equal to one and the other elements equal to zero, $\hat{T} = \sum_{i=1}^m \hat{\gamma}_i\hat{z}_i\hat{z}_i^t$ and $\hat{t} = \sum_{i=1}^m \hat{\gamma}_i\hat{z}_i\hat{\theta}_i$,

$$\hat{\Lambda}_{j,k} = \begin{pmatrix} 0 & 0 & \dots & \hat{z}_{j,1} & \dots & 0 \\ 0 & 0 & \dots & \hat{z}_{j,2} & \dots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ \hat{z}_{j,1} & \hat{z}_{j,2} & \dots & 2\hat{z}_{j,k} & \dots & \hat{z}_{j,K} \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \dots & \hat{z}_{j,K} & \dots & 0 \end{pmatrix}, \tag{26}$$

and $\delta_{i,j} = 1$ if $j = i$ and $\delta_{i,j} = 0$ if $j \neq i$. Further more,

$$\frac{\partial f_i(\hat{z},\hat{\theta})}{\partial \hat{\theta}_j} = \hat{z}_i^t\hat{T}^{-1}\hat{z}_j\hat{\gamma}_j, \tag{27}$$

Evaluating (25) and (27) in its real points gives:

$$\left.\frac{\partial f_i(\hat{z},\hat{\theta})}{\partial \hat{z}_{j,k}}\right|_{(\hat{z},\hat{\theta})=(z,\theta)} = \delta_{i,j}\hat{\beta}_0^t\lambda_k - z_i^tT^{-1}\Lambda_{j,k}\hat{\beta}_0\hat{\gamma}_j + z_i^tT^{-1}\lambda_k\theta_j\hat{\gamma}_j, \tag{28}$$

$$\left.\frac{\partial f_i(\hat{z},\hat{\theta})}{\partial \hat{\theta}_j}\right|_{(\hat{z},\hat{\theta})=(z,\theta)} = z_i^tT^{-1}z_j\hat{\gamma}_j, \tag{29}$$

with $\hat{\beta}_0$ the GLS estimator for β in its real points z_i and θ_i, i.e.

$$\hat{\beta}_0 = \left(\sum_{i=1}^m \hat{\gamma}_i z_i z_i^t\right)^{-1}\sum_{i=1}^m \hat{\gamma}_i z_i\theta_i,$$

and $T = \sum_{i=1}^m \hat{\gamma}_i z_i z_i^t$. Inserting (28) and (29) in (24) gives

$$\begin{aligned} f_i(\hat{z},\hat{\theta}) &\approx z_i^t\hat{\beta}_0 + \hat{\beta}_0^t(\hat{z}_i - z_i) - \sum_{j=1}^m \hat{\gamma}_j z_i^tT^{-1}z_j\hat{\beta}_0^t(\hat{z}_j - z_j) - \sum_{j=1}^m \hat{\gamma}_j z_j^t\hat{\beta}_0 z_i^tT^{-1}(\hat{z}_j - z_j) \\ &\quad + \sum_{j=1}^m z_i^tT^{-1}\theta_j\hat{\gamma}_j(\hat{z}_j - z_j) + \sum_{j=1}^m z_i^tT^{-1}z_j\hat{\gamma}_j(\hat{\theta}_j - \theta_j) \\ &= z_i^t\hat{\beta}_0 + \sum_{j=1}^m [(\delta_{i,j} - \hat{\gamma}_j z_i^tT^{-1}z_j)\hat{\beta}_0^t + \hat{\gamma}_j(\theta_j - z_j^t\hat{\beta}_0)z_i^tT^{-1}](\hat{z}_j - z_j) \\ &\quad + \sum_{j=1}^m z_i^tT^{-1}z_j\hat{\gamma}_j(\hat{\theta}_j - \theta_j) \\ &\equiv z_i^t\hat{\beta}_0 + \sum_{j=1}^m B_{i,j}(\hat{z}_j - z_j) + \sum_{j=1}^m C_{i,j}(\hat{\theta}_j - \theta_j), \end{aligned}$$

with $B_{i,j}$ a K vector defined by

$$B_{i,j} = [(\delta_{i,j} - \hat{\gamma}_j z_i^t T^{-1} z_j)\hat{\beta}_0^t + \hat{\gamma}_j(\theta_j - z_j^t \hat{\beta}_0) z_i^t T^{-1}],$$

and $C_{i,j}$ a scalar, defined by

$$C_{i,j} = z_i^t T^{-1} z_j \hat{\gamma}_j.$$

A Taylor approximation for $\tilde{\theta}_i$ is given by

$$\begin{aligned} \tilde{\theta}_i^a &= \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i) \hat{z}_i^t \beta \\ &\approx \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i)[z_i^t \hat{\beta}_0 + \sum_{j=1}^{m} B_{i,j}(\hat{z}_j - z_j) + \sum_{j=1}^{m} C_{i,j}(\hat{\theta}_j - \theta_j)]. \end{aligned}$$

Simulating auxiliary variables with survey error

In section 3.4 and 3.6 it is argued that auxiliary information that is measured with error can be treated as if it is measured without error in the situation that the error distribution does not vary between domains. The estimate of the model variance σ_v^2 will absorb the variance of the errors. A simulation is conducted to illustrate this effect.

The NSM variable propvict is considered, indicating the percentage of people that have been victim of property crime in the last 12 months. A strong predictor is the registered number of property crimes, which is taken to be the only covariate in the model used in this simulation study.

In the simulation, normally distributed errors are added to the covariate, with zero mean, and a standard deviation that increases from 0 to 50 in steps of 2. For each error distribution considered, 2,000 bootstrap iterations are run. In each iteration, the model is fit and the estimate of the model variance is retained. The values reported here are means taken over the 2,000 iterations. The goal of the simulation is an assessment of the extent to which the estimate of the model variance increases with increasing error variance. Considering the area level model, $z_i^t \beta + v_i + e_i$, adding errors ϵ_i to the covariate z_i can be expected to increase the variance of the random effects v_i by $\beta'^2 \sigma_\epsilon^2 + (\beta' - \beta)^2 \mathrm{Var}(z_i)$, with β' the regression coefficient of the covariates with error, and σ_ϵ^2 the variance of the errors. This increase is caused by the error variance itself, as well as by a reduction in the correlation between the covariate and the dependent variable.

Figure 1 shows how the estimate of the model variance increases when error is added to the covariate. On the horizontal axis the standard deviation of the errors is shown. The solid line shows the result of the simulation. The dashed line represents the increase one could expect. In the situation without errors added to the covariate, $\hat{\sigma}_v^2 = 1.30$. The standard deviation of the covariates without error is 10.8. With relatively small errors added, the estimate of the model variance is properly inflated to the expected levels. For large errors, the increase is even more than expected. This result confirms that the MSE of the EBLUP and HB predictions implicitly account for the additional uncertainty of

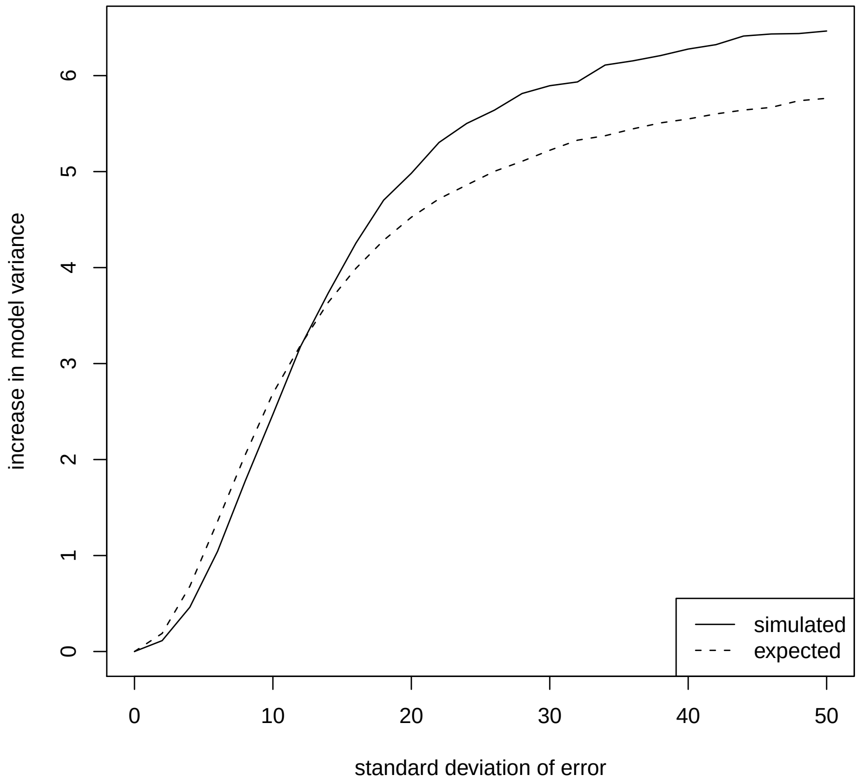

Figure 1. The effect of adding errors to the covariate on the estimation of the model variance.

covariates with errors if these errors are constant over the domains, which is inline with the results of Ybarra and Lohr (2008).

Figures

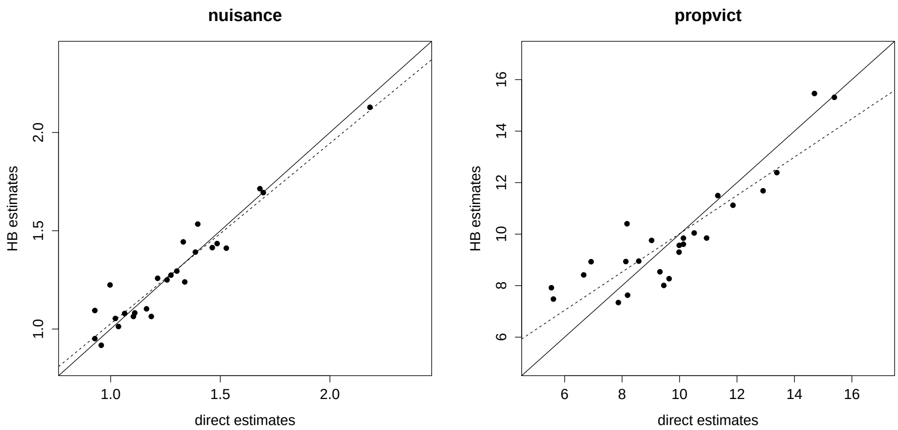

Figure 2. HB versus direct estimates for the two variables under consideration. The solid line is the diagonal, y = x, and the dashed line is a linear regression line fitted to the data.

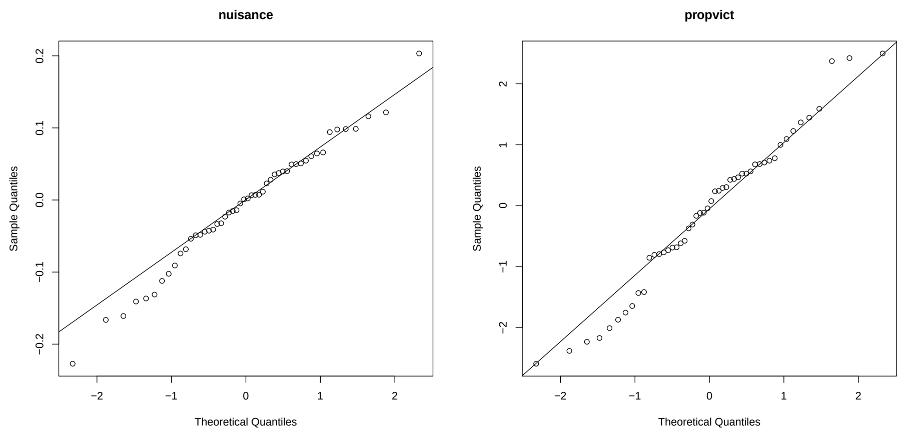

Figure 3. Normal QQ-plots of the residues of nuisance and propvict.

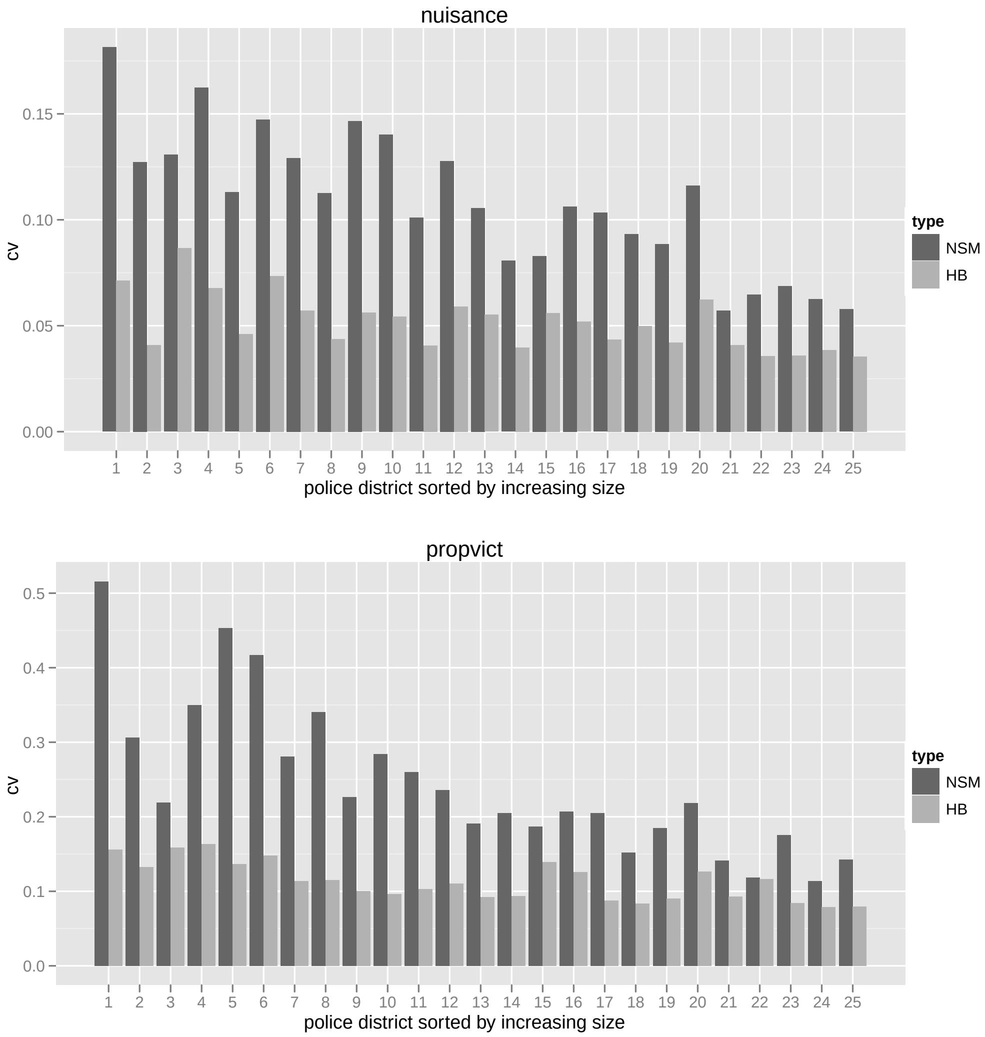

Figure 4. Coefficients of variation (cv) of direct (NSM) and HB estimates for two variables.

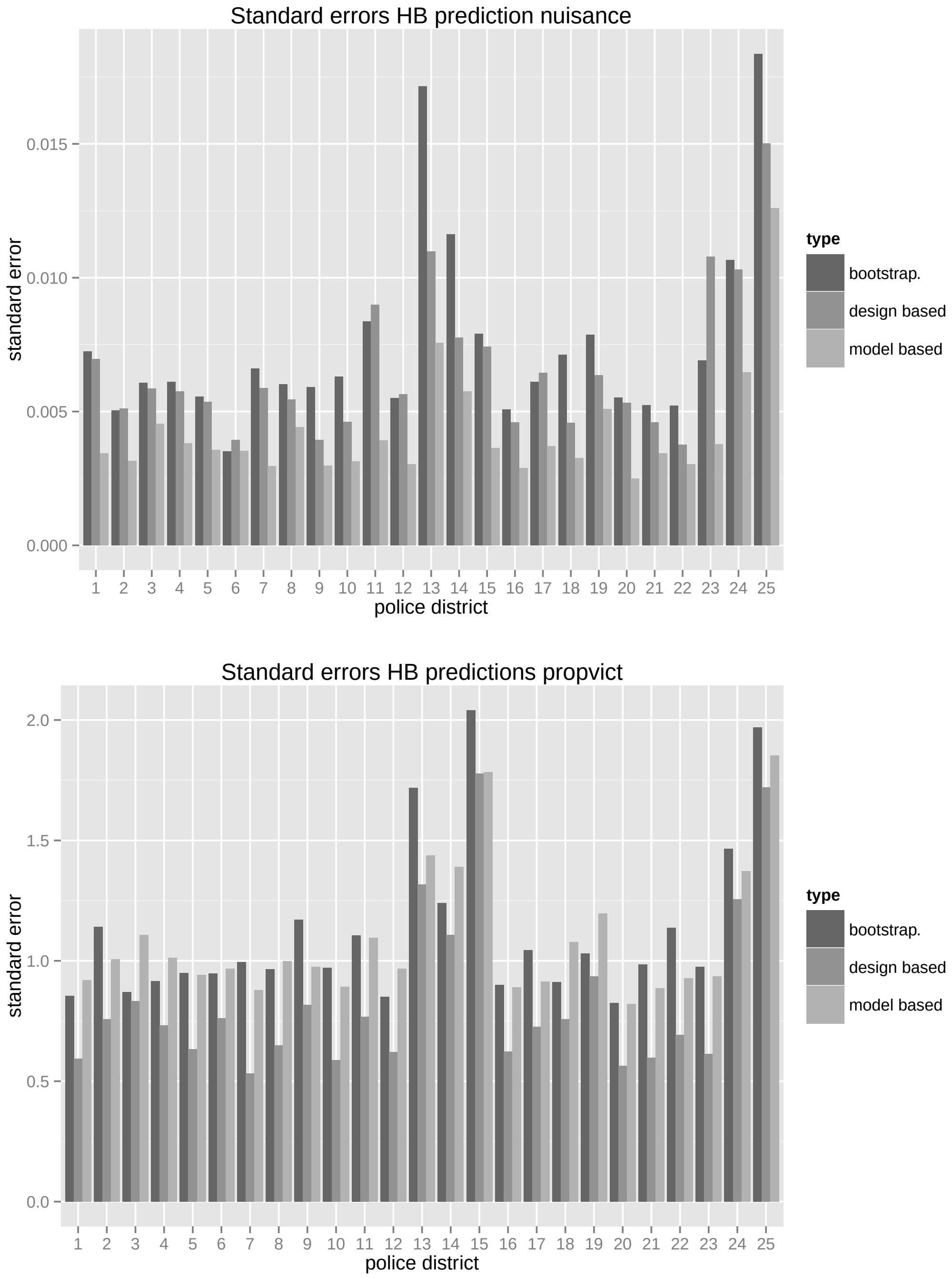

Figure 5. Comparison standard errors HB predictions nuisance and propvict.

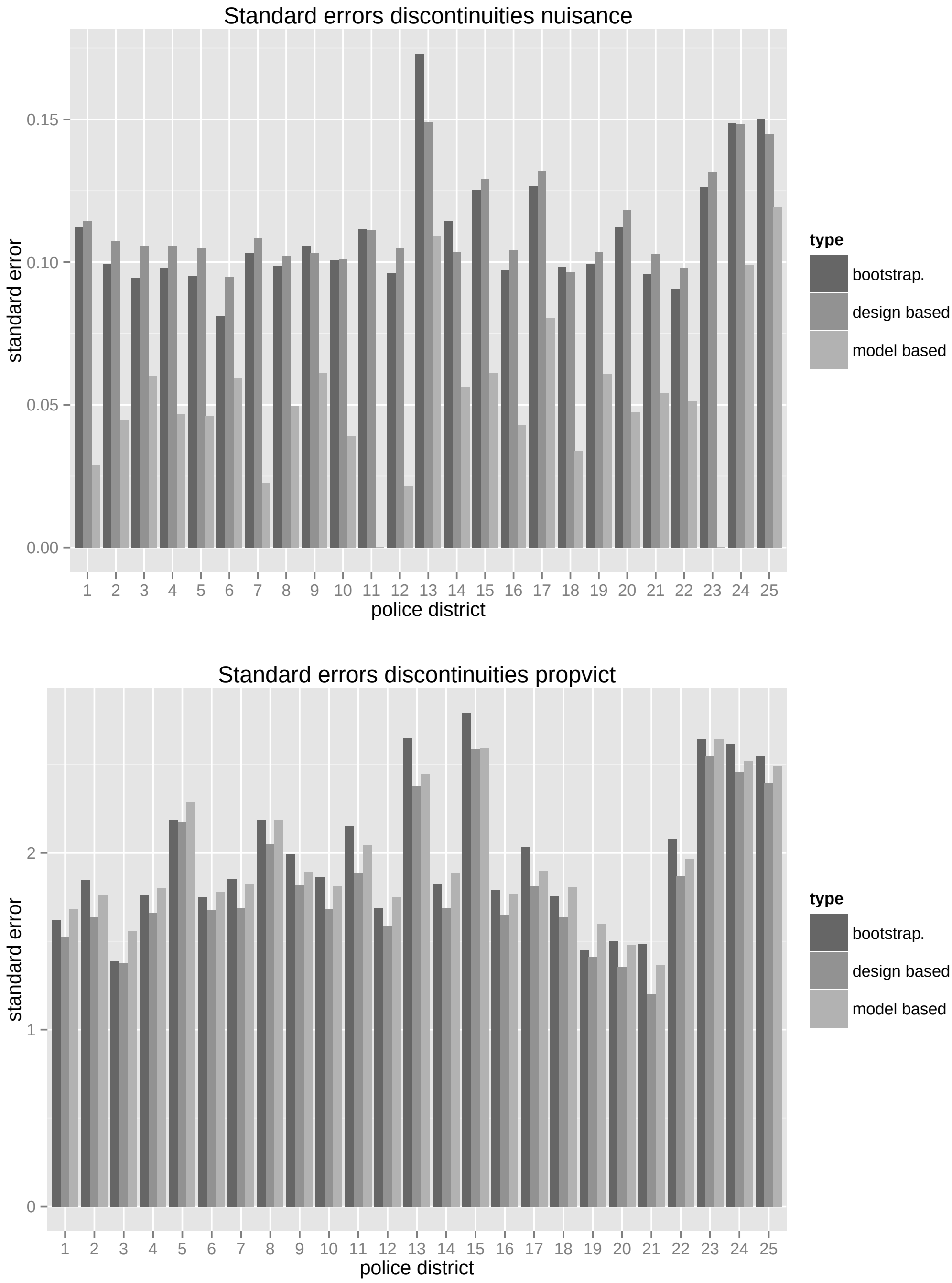

Figure 6. Comparison standard errors discontinuities nuisance and propvict.

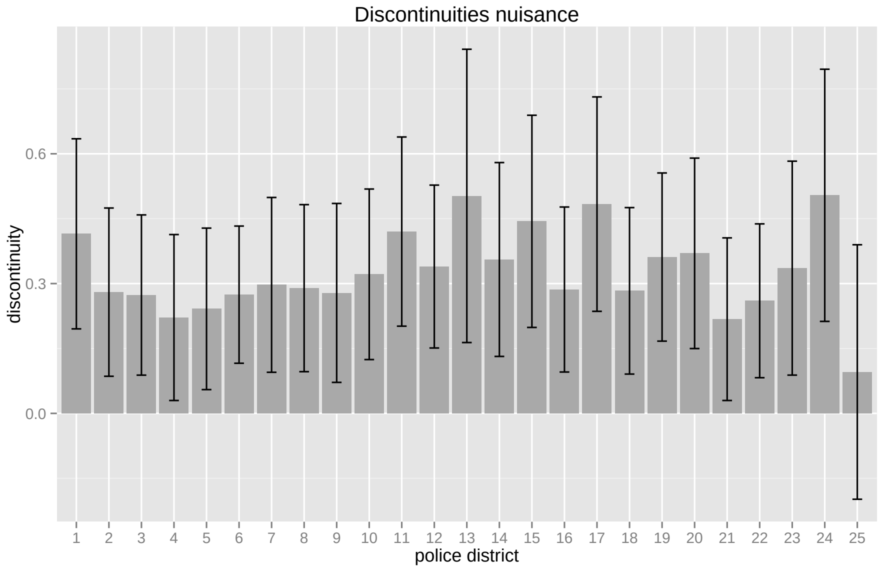

Figure 7. Discontinuities for nuisance with 95% confidence interval.

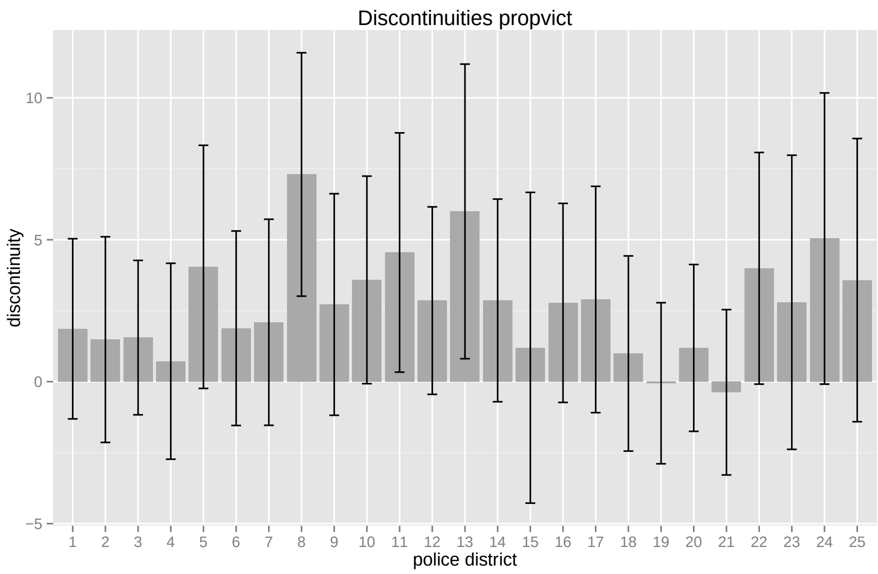

Figure 8. Discontinuities for propvict with 95% confidence interval.

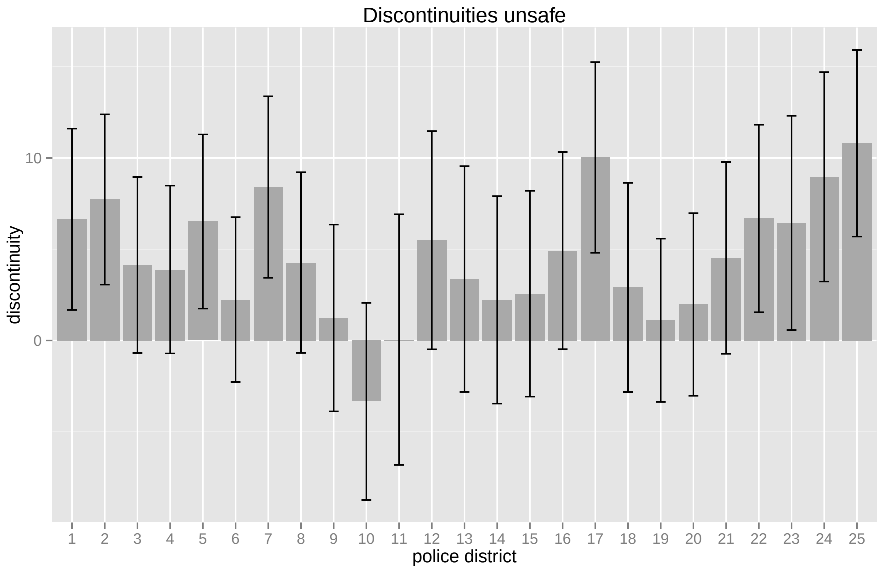

Figure 9. Discontinuities for unsafe with 95% confidence interval.

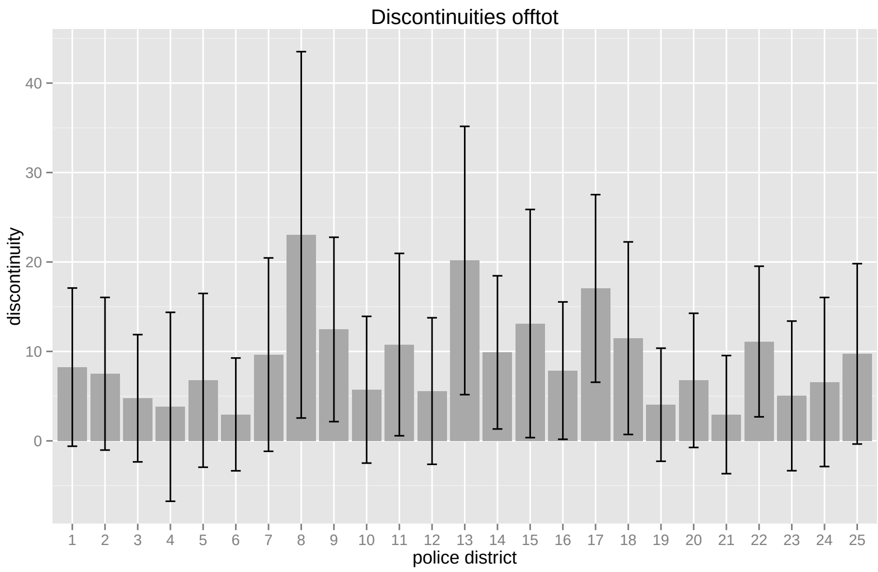

Figure 10. Discontinuities for offtot with 95% confidence interval.

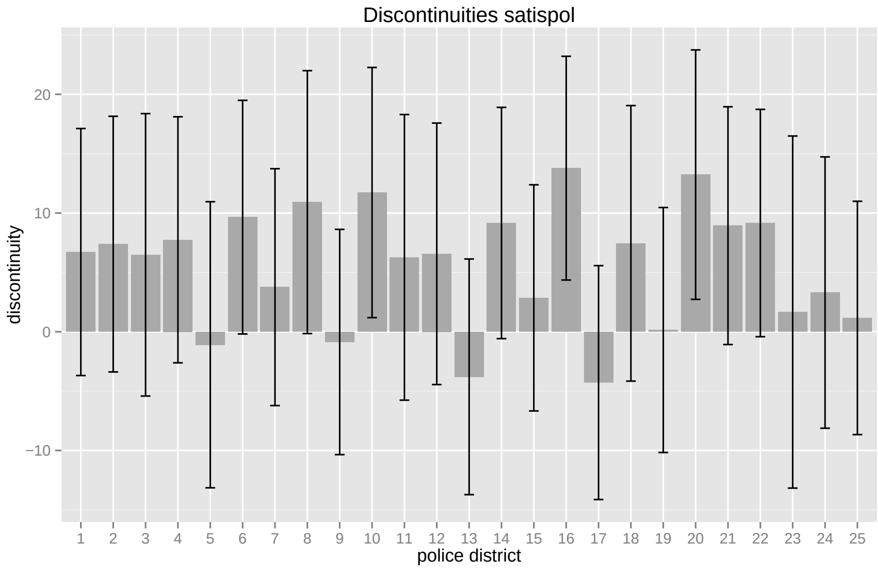

Figure 11. Discontinuities for satispol with 95% confidence interval.