Universität Trier

A Finite Mixture Fay Herriot-type model for estimating regional rental prices in Germany

Charlotte Articus
Jan Pablo Burgard

Research Papers in Economics
No. 14/14

A Finite Mixture Fay Herriot-type model for estimating regional rental prices in Germany

*Working Paper**

Charlotte Articus† and Jan Pablo Burgard‡

October 31, 2014

Abstract

In model-based small area estimation an explicit statistical model is used to enhance efficiency of estimation in case of small subsamples. This model assumes a fixed relationship between the statistic of interest and a set of covariates, which is valid for all areas under consideration and can, thus, be used to stabilize estimation. In some applications, there might, however, be different subgroups of areas with specific data-generating processes, i.e. specific relationships between response variable and auxiliary information. In this case, estimation of a distinct model for each subgroup would be more appropriate than one model for all observations. If so, the definition of subgroups becomes a crucial task in the estimation process.

We propose a Finite Mixture Fay Herriot-type model to account for unobserved heterogeneity in the data. More specifically, we assume that the statistic of interest stems from a mixture distribution with K components. The estimation of mixing proportions, area-specific probabilities of subgroup identity and the K sets of model parameters is then performed simultaneously. Eventually, the Finite Mixture Fay Herriot-type estimator is formulated as a weighted mean of predicts from model 1 to K, with weights given by the area-specific probabilities of subgroup identity.

The suggested method is tested in a model-based simulation study. It is then applied to the problem of estimating regional rental prices on district level in Germany.

*This is work in progress and contains preliminary results. Comments and questions are highly appreciated. The authors are grateful to Ralf Münnich for his valuable suggestions and comments. They also thank the Federal Statistical Office of Germany for providing the survey data on rental prices used in this study.

†University of Trier, Economic and Social Statistics Department, articus@uni-trier.de.

‡University of Trier, Economic and Social Statistics Department, burgardj@uni-trier.de.

1 Introduction

When analysing survey data in order to gain insight into social or economic phenomena, the aim might not only be to make statistical inferences about the entire target population but also to obtain reliable information for certain subentities. These can for example be smaller areas within a region under study or certain demographic subgroups in a population. Researchers are, however, usually confronted with the problem that the spatial or thematic disaggregation of the available sample results in small subsamples for the subentities of interest, which causes a lack of accuracy of conventional direct estimators.

Model-based Small Area Estimation (SAE) techniques, that are designed to produce reliable information even for small subsamples, might be a solution. These methods aim at improving efficiency of estimation in the case of small subsamples by means of an explicit statistical model. This model assumes a fixed relationship between a set of covariates and the statistic of interest, which is valid for all subentities under consideration. It can, thus, be estimated from the sampled information from all areas and then be used to stabilize estimation.

In many applications it might, however, be plausible to assume that the effects of given auxiliary information in some areas are different to those in others. Hence, it seems sensible to consider different subgroups of areas with specific data-generating processes and the estimation of subgroup-specific models might be more appropriate. This leads, however, to the open question of how to define suitable subgroups. In some cases, there might be a natural clustering variable. If this is not the case, finite mixture regression models might be a solution. In this framework, a set of two or more different models is specified and the estimation of model parameters is performed simultaneously to estimating subgroup identity or a probability of subgroup identity for each area.

We, therefore, propose a Finite Mixture Fay Herriot-type (FMFH) model to account for unobserved heterogeneity in the data. Our suggestion is inspired by an application of estimating regional rental prices in Germany. It might, however, be an appropriate method in any application where small area estimates for heterogeneous subentities are of interest. Furthermore, the suggested estimator can also be interpreted as a flexible approach when the distribution of the statistic of interest is unknown and the usual normality assumption of the basic small area models seems inappropriate.

SAE has gained much attention in the last decades and standard methods are nowadays well established. An extensive overview is provided with the standard work of Rao (2003). An account of newer developments is given by Jiang and Lahiri (2006) and Pfeffermann (2013). A study thematically related to the application at hand has been published by Pereira and Coelho (2013), who estimate average house prices in Portugal applying several different small area estimators.

Concerning the use of mixtures in SAE, some suggestions with different motivations have been made. Mixtures have been employed in order to relax restricting distributional assumptions (see Elbers & van der Weide, 2014; Maiti, 2003) as well as in robust SAE (see Datta & Lahiri, 1995; Gershunskaya, 2010). Recently, Datta and Mandal (n.d.) porposed a mixture of a degenerate distribution localized at zero and the usual normal distribution for the random effects in an

area-level mixed model. They, therewith, suggest a flexible strategy to small area modelling. Random effects are only included for those areas for which the statistic of interest is not sufficiently well explained by the covariates included in the fixed part of the model. All these approaches do not only differ from our proposal with respect to their motivation and underlying intuition but also in the form in which mixtures are included into the framework. None of these works considers a mixture of mixed-effects regression models employed for the prediction of the statistic of interest in model-based SAE.

To our knowledge, only Maiti, Ren, Dass, Lim, and Maier (in press) (see also the dissertation by Ren, 2011) have so far considered the integration of specific methods to account for the existence of different subgroups of areas into model-based SAE. They use a model-based clustering algorithm proposed by Booth, Casella, and Hobert (2008) to partition areas into subgroups. Small area estimates are then obtained employing cluster-specific Fay Herriot-models. In contrast to that, with suggesting a mixture model-based approach, we opted for a probabilistic assignment to subgroups instead of hard clustering.

Most notably since the introduction of the EM-algorithm by Dempster, Laird, and Rubin (1977), finite mixture model theory has received growing interest. See McLachlan and Peel (2000) for an extensive overview of theoretical and practical aspects of finite mixture modelling. Related to our approach, there are some applications – mainly in the field of biology and the health sciences – where mixtures of mixed-effects models have been utilized. See Yau, Lee, and Ng (2003), Celeux, Martin, and Lavergne (2005), Ng, McLachlan, Wang, Ben-Tovim Jones, and Ng (2006), McLachlan, Ng, and Wang (2008), Martinez, Lavergne, and Trottier (2009) as well as Martella et al. (2011). Verbeke and Lesaffre (1996) analysed linear mixed models where the random effects are distributed according to a mixture of normal distributions. Scharl, Grün, and Leisch (2010) compared the performance of mixtures of linear regression models with and without random effects in a simulation study. In a recent paper, Du, Kahili, Neslehova, and Steele (2013) proposed an approach for model selection for finite mixtures of linear mixed models. Note that in these applications and theoretical discussions of finite mixture models the interest usually either lies in clustering or, less frequently, in the interpretation of component-specific model coefficients, whereas the approach presented here focusses on predicting a statistic out of the estimated model.

The remainder of this paper is organized as follows: We start by introducing the two fields of basic theory which we are relying on in developing our proposal. More specifically we give a brief overview on model-based SAE and the specific model applied and extended in our approach. We then introduce the basic theory of finite mixture (regression) models. In section 3 we bring these two fields together in presenting the FMFH-model. The suggested estimator is evaluated in a simulation study. It is then applied to the problem of estimating regional rental prices for German districts. We conclude with a summarizing judgement of the proposed method and an outlook on remaining tasks.

2 Basic Theory

2.1 Model-based Small Area Estimation

When a survey is conducted, besides an interest in making statistical inferences about the entire population under study, there might also be the aim of obtaining reliable information for specified subentities. These subpopulations are called *areas* or *domains* depending on whether the disaggregation of the population is by region or by content. To specify the problem we consider the following setting (see Münnich, Burgard, & Vogt, 2013): A population $\mathcal{U}$ of size N is divided into m pairwise disjoint subpopulations $\mathcal{U}_i, i = 1, \ldots, m$. A sample $\mathcal{S}$ of size n is drawn, with $\mathcal{S}_i = \mathcal{S} \cap \mathcal{U}_i$ designating the sample realized in $\mathcal{U}_i$. Now the aim is to simultaneously estimate a vector of m area-specific parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$, e.g. means $\boldsymbol{\mu}$.

The disaggregation of the available sample often results in small sample sizes for the subentities. Under these circumstances, conventional direct estimators (which by definition only use the information from the area under consideration) lead to an unacceptably large standard error. Note, however, that such direct estimators are generally design-unbiased. Methods of model-based SAE are designed to produce reliable estimates even for very small subsamples. The strategy is to apply indirect techniques that make use of additional information as, for example, sampled information from other areas. This additional information is often included through a model. The intuition is, that part of the (inter-area) variation of the statistic of interest can be explained by a relationship between this variable and a certain set of covariates that is valid for all areas under consideration. A respective model can, thus, be estimated using the data points from all subsamples. The specified relationship can then be employed to stabilize estimation. In the literature this strategy is often referred to as "borrowing strength" (Ghosh & Rao, 1994).

A basic approach in SAE is to estimate a linear mixed model, either on level of the observation units or on aggregated level of the areas. The standard area level model as defined by Fay and Herriot (1979), which is the relevant model for the study at hand, is given by

$$\begin{aligned} \hat{\mu}_i^{Dir} &= \boldsymbol{x}_i^T \boldsymbol{\beta} + v_i + e_i \qquad \text{for} \quad i = 1, \ldots, m \\ v_i &\overset{i.i.d}{\sim} N(0, \sigma_v^2) \\ e_i &\overset{ind}{\sim} N(0, \sigma_{e,i}^2). \end{aligned} \tag{1}$$

$\hat{\mu}_i^{Dir}$ is the direct estimate, obtained as a weighted mean from the sample realized in area i. e_i is the error term, which in this case is the sampling error of the direct estimate with known variance $\sigma_{e,i}^2$. $\boldsymbol{x}_i^T$ denotes the vector of auxiliary information for area i with corresponding regression coefficients $\boldsymbol{\beta}$. The area-specific random effect is denoted by v_i. It is assumed that v_i and e_i are independently distributed.

When estimating this model, the interest is not as much in the model coefficients themselves as in the prediction of the target variable out of the model. The Empirical Best Linear Unbiased Predictor (EBLUP) for the parameter of interest

μ under this model (see Rao, 2003, p. 107-108, 116-118) is given by

$$\begin{aligned}\hat{\mu}_i^{FH} &= \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}} + \hat{v}_i \\ &= \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}} + \hat{\gamma}_i(\hat{\mu}_i^{Dir} - \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}) \\ &= \hat{\gamma}_i\hat{\mu}_i^{Dir} + (1-\hat{\gamma}_i)\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}\end{aligned} \tag{2}$$

with

$$\hat{\gamma}_i = \frac{\hat{\sigma}_v^2}{\sigma_{e,i}^2 + \hat{\sigma}_v^2}. \tag{3}$$

This estimator is called the Fay Herriot (FH) estimator $\hat{\mu}_i^{FH}$ for the parameter of interest. $\hat{\boldsymbol{\beta}}$ and $\hat{v}_i$ denote the BLUE for $\boldsymbol{\beta}$ and the BLUP for v_i, respectively (see Henderson, Kempthorne, Searle, and von Krosigk (1959) and Searle, Casella, and McCulloch (1992) for estimation and prediction of parameters for a mixed model). Note that $\hat{\mu}_i^{FH}$ can be expressed as a composite estimator of the synthetic estimator $\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}$, obtained from the fixed part of the model, and the direct estimator $\hat{\mu}_i^{Dir}$. Weights are given by the area-specific shrinkage factor $\hat{\gamma}_i$, that sets the model variance $\hat{\sigma}_v^2$ in relation to the total variance $\sigma_{e,i}^2 + \hat{\sigma}_v^2$. If the model-variance is small compared to the design variance, $\hat{\gamma}$ is close to zero and the synthetic estimator dominates. Intuitively, $\hat{\gamma}$ can be understood as a relative measure of confidence in the model- and the design-based estimator (see Rao, 2003, p. 116-117).

The aim, of course, is to stabilize the estimation, that is to yield estimates with a far smaller variance in the context of small sample sizes. Note, however, that this comes with the price of loosing the property of design-unbiasedness. There, hence, is a trade-off between bias and variance. The relevant measure to judge the quality of model-based small area estimates, therefore, is the mean square error (MSE), that is $\mathrm{MSE}(\hat{\mu}_i^{FH}) = \mathrm{E}(\hat{\mu}_i^{FH} - \mu_i)^2$.

2.2 Finite Mixture Models

Finite mixture models offer an intuitively appealing approach when it is plausible to assume that there is a certain number of – actually existing – subgroups in the population yet subgroup identity is unobserved for all observations. The aim then is to estimate a model for each subgroup as well as unconditional subgroup probabilities. Depending of the purpose of the application, there might also be an interest in attributing subgroup-identity or a probability of subgroup identity to specific observations. With this intuitive background, the framework of finite mixture modelling is sometimes conceptualized as a missing data problem, where the realizations of a multinomial variable indicating class membership is missing for all observations. Correspondingly, there are some parallels in the estimation of finite mixture models and the fitting of models in the case of missing values (see McLachlan & Peel, 2000, p. 7, 19-20).

Yet, the assumed components of the mixture distribution do not necessarily correspond with actually existing subgroups. Alternatively, finite mixture modelling can be interpreted as a flexible semi-parametric way to model unknown distributional shapes (see McLachlan & Peel, 2000, p. 7-8).

Consider m observations $y_i, i = 1, \ldots, m$ of an outcome variable y and a set of covariates $\boldsymbol{x}$. For a K-component finite mixture of regression models, the conditional probability density function of y given $\boldsymbol{x}$ can be written as

$$f(y|\boldsymbol{x}, \boldsymbol{\Psi}) = \sum_{k=1}^{K} \pi_k f_k(y|\boldsymbol{x}, \boldsymbol{\theta}_k). \tag{4}$$

Here, f_k are the K component densities, π_k are the mixing proportions with $\sum_{k=1}^{K} \pi_k = 1$ and $\pi_k > 0 \quad \forall \quad k = 1, \ldots, K$ and $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K$ are the K vectors of component-specific model parameters. $\boldsymbol{\Psi} = (\pi_1, \ldots, \pi_{K-1}, \boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_K^T)$ is a vector containing all the unknown parameters in the mixture distribution. The aim of mixture modelling is to specify K appropriate models and to estimate the corresponding model parameters $\boldsymbol{\theta}_k$ as well as the set of (unknown) mixing proportions or model probabilities π_k for each model.

As stated above, this framework is often conceptualized as a missing data problem (see McLachlan & Peel, 2000, p. 7, 19-20): It is assumed that each observation belongs to one of K (actually existing) classes with a specific data-generating process $y_i|\boldsymbol{x}_i, z_i \sim f_{z_i}(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_{z_i})$. Here, $z_i \in [1, \ldots, K]$ is an unobserved variable identifying the class, y_i belongs to, i.e. the data-generating process of y_i. The mixing proportions π_k are then interpreted as the unconditional probability that an observation belongs to class k, that is $\pi_k = Pr(z_i = k)$.

Additionally to the consideration of model probabilities, i.e. the unconditional probabilities for $z_i = k$, there might also be an interest in class-membership for a specific observations y_i. The conditional probability that observation i belongs to class k and is, thus, generated by the class-specific process is given by

$$\xi_{i,k} = (Pr(z_i = k)|y_i, \boldsymbol{x}_i, \boldsymbol{\Psi}) = \frac{\pi_k f_k(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_k)}{\sum_{j \in K} \pi_j f_j(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_j)}. \tag{5}$$

$\xi_{i,k}$ is, hence, an observation-specific measure of class-membership, which can also be interpreted as the degree to which observation i is consistent with the relationship between y and $\boldsymbol{x}$ implied by model k. $\text{argmax}_k(\hat{\xi}_{i,k})$ can be used to segment the observations into k clusters, i.e. to construct a list of observations that are best explained by model k (see McLachlan & Peel, 2000, p. 29).

A possible problem of finite mixture models is identifiability (see Titterington, Smith, & Makov, 1985). A finite mixture of regression models as specified above is said to be identifiable if, for a given set of covariates $\boldsymbol{x}$, for any two vectors of parameters $\boldsymbol{\Psi}$ the equality

$$\sum_{k=1}^{K} \pi_k f_k(y|\boldsymbol{x}, \boldsymbol{\theta}_k) = \sum_{k=1}^{K^*} \pi_k^* f_k(y|\boldsymbol{x}, \boldsymbol{\theta}_k^*) \tag{6}$$

implies that $K = K^*$ and that $\boldsymbol{\Psi}^*$ can be permuted[1] such that $\boldsymbol{\Psi} = \boldsymbol{\Psi}^*$. The identifiability of mixtures of linear regression models depends on the number of components K, the component-specific densities and the design matrices. See Hennig (2000) for details. In what follows, we assume that the mixture model under consideration is identifiable.

[1] To see the necessity of this amendment in the definition, note that the mixture density is invariant to the change of component labels in $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K)$ and $(\pi_1, \ldots, \pi_K)$ if the component densities belong to the same parametric family.

3 The Finite Mixture FH-type Model

As stated above, model based small area estimators use an explicit statistical model to improve estimation by exploiting the relationship between a set of covariates and the statistic of interest. In some applications, it may, however, be plausible to assume that the effects of given covariates in some areas are different to those in others. It might also be possible, that specific covariates that are important in some areas, do not play a role at all elsewhere.

Given a large enough number of areas, it may, then, be plausible to estimate different models for different types of areas. If so, next the questions arises of how to compose sensible subgroups. There might be a natural clustering variable, which can be used to segment areas into two or more subgroups. If this is not the case, finite mixture modelling might be a suitable framework to "let the data decide" on how to group the areas into K subgroups with specific data-generating processes. In the following sections we propose, analyse and apply an corresponding estimator for SAE in the presence of unobserved subgroups of areas.

3.1 Model and Estimator

Consider m areas divided into K unobserved classes. We assume that there is a specific data-generating process in each class. Thus, the statistic of interest for a given area i belonging to class k is appropriately modelled by a normal distribution with component-specific mean $\boldsymbol{x}^T\boldsymbol{\beta}_k$ and covariance matrix $\text{diag}_{i\in m}(\sigma^2_{e,i} + \sigma^2_{v,k})$, that is

$$\mu_i|(z_i = k) \sim N(\boldsymbol{x}_i^T\boldsymbol{\beta}_k, \sigma^2_{e,i} + \sigma^2_{v,k}). \tag{7}$$

As above, $\sigma^2_{e,i}$ denotes the variance of the direct estimate and is assumed to be known for all areas.

Alternatively, we can loosen the assumption of exclusive subgroup membership of areas to one single subgroup and instead draw on the more flexible notion that each area belongs to the different subgroups with a certain probability. This might be understood as an approach of fuzzy or soft clustering via mixture models, where instead of a hard assignment to clusters only a probabilistic statement about cluster membership is made (see Everitt, Landau, Leese, & Stahl, 2011, chap. 6 and p. 244-245). An intuitive interpretation is that there are different data-generating processes, i.e. different possible relationships between $\boldsymbol{x}$ and y, and each of them is valid in every area with a certain area-specific probability. Under this assumption, the statistic of interest for a given area is appropriately modelled by a weighted mean of the K component densities

$$\mu_i \sim \sum_{k=1}^{K} \xi_{i,k} N(\boldsymbol{x}_i^T\boldsymbol{\beta}_k, \sigma^2_{e,i} + \sigma^2_{v,k}) \tag{8}$$

where weights are given by $\xi_{i,k}$, i.e. the area-specific measures of consistency with model k.

Either way, we specify K appropriate models and estimate a finite mixture of K FH-models with component-specific fixed coefficients $\boldsymbol{\beta}_k$ and model variances $\sigma^2_{v,k}$:

$$\begin{aligned} \mu_i^{Dir} &= \sum_{k=1}^{K} \pi_k \left(\boldsymbol{x}_i^T \boldsymbol{\beta}_k + v_{i,k} + e_i \right), \quad i = 1, \ldots, m \\ v_{i,k} &\overset{i.i.d}{\sim} N(0, \sigma^2_{v,k}) \\ e_i &\overset{ind}{\sim} N(0, \sigma^2_{e,i}). \end{aligned} \tag{9}$$

As described above, simultaneously estimates for the K mixing proportions π_k and the $m \cdot K$ area-specific conditional probabilities $\xi_{i,k}$ are obtained.

Let $\hat{\mu}_i^{FH,k}$ denote the predict for μ_i derived from model k. Then – in accordance to the two possible interpretations we suggested above – we have two possibilities to define an estimator for μ_i that accounts for the existence of unobserved subgroups of areas. We can either use

$$\hat{\mu}_i^{FH,mix} = \hat{\mu}_i^{FH,k^*} \quad \text{where} \quad k^* = \text{argmax}_k(\hat{\xi}_{i,k}), \tag{10}$$

i.e. use the predict for the statistic of interest obtained from the model with the largest area-specific conditional probability that area i belongs to model k, or

$$\hat{\mu}_i^{FH,mix} = \sum_{k=1}^{K} \hat{\xi}_{i,k} \cdot \hat{\mu}_i^{FH,k}. \tag{11}$$

Here, we estimate the statistic of interest as a weighted mean of predicts from the K models, where the area-specific weights are given by the conditional probabilities that area i belongs to model k.

In the application presented below, we employed the more flexible notion of the second interpretation and opted for (11) for the following reasons:

If subgroups are not clearly separated such that two or more models are almost equally consistent with the observed information for a given area, the variance of k^* might be quite large. In this case the respective switching of prediction models in (10) might cause a considerable variance in the estimator. This variance is somehow unjustified in the sense that it is not due to the instability of estimated parameters. Furthermore, considering the basic notion of borrowing strength in SAE, it seems inappropriate not to exploit the explanatory power of all models that are, to some estimated degree, consistent with a given observation. This argument again is particularly valid if subgroups are overlapping and the conditional probabilities are similar for different components of the mixture. Eventually, the second estimator also is the appropriate choice when we drop the assumption of a true number of physically existent subgroups and interpret the finite mixture of FH-models as a semiparametric way to model unknown distributional shapes.

For these reasons, in the following, we draw on the estimator suggested in (11). We call it the Finite Mixture Fay Herriot-type (FMFH) estimator.

Note however, that the choice of one of these two estimators may depend on the specific application at hand. Given a clear separation of a small number of subgroups, each containing a large enough number of areas, it might be appealing to use the intuitively more straightforward estimator suggested in (10). It simplifies the interpretation and presentation of results. Besides, it yields clearly identified clusters which in itself might be a desired result and immediately justifies employing a mixture-based estimator.

3.2 Parameter Estimation

Parameter estimation is performed using the the frequentist approach of maximum likelihood estimation via the EM-algorithm. Alternatively, a Bayesian approach can be adopted. See McLachlan and Peel (2000, chap. 4) for an overview of Bayesian inference for finite mixture models.

The log-likelihood function for a mixture of K normal components is given by

$$\log L(\boldsymbol{\Psi}) = l(\boldsymbol{\Psi}) = \sum_{i=1}^{m} \log \left(\sum_{k=1}^{K} \pi_k f_k(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_k) \right) \tag{12}$$

Maximum likelihood estimates for the parameters $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k^T, \sigma_{v,k}^2)$ as well as the model probabilities π_k are obtained by maximizing this log-likelihood given the observed realizations for y and $\boldsymbol{x}$. However, optimization of (12) can not be done directly because the log of a sum in the function makes its derivative computationally intractable. Following a standard approach in fitting mixture models, maximum likelihood estimates are, therefore, obtained using the EM-algorithm, an iterative numerical optimization algorithm introduced by Dempster et al. (1977) as a method for maximum likelihood estimation in the case of incomplete data. See McLachlan and Peel (2000) for an overview of parameter estimation via the EM-algorithm for mixture models.

In this estimation context, the finite mixture model framework is, accordingly, usually interpreted as a missing data problem (see McLachlan & Peel, 2000, p. 48). As described above, this implies the notion that each observation y_i belongs to one of the K classes with z_i indicating the true class membership for observation y_i. The complete data set would, thus, contain m realizations of $y_i, \boldsymbol{x_i}$ and z_i and the complete-data likelihood l_c would be given by

$$l_c(\boldsymbol{\Psi}) = \sum_{i=1}^{m} \log \left(\sum_{k=1}^{K} I(z_i = k) \pi_k f_k(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_k) \right), \tag{13}$$

which can be rearranged[2] as

$$l_c(\boldsymbol{\Psi}) = \sum_{i=1}^{m} \sum_{k=1}^{K} I(z_i = k)(\log \pi_k + \log f_k\left(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_k\right)). \tag{14}$$

[2]For any i the indicator function equals 0 in $K-1$ cases and the inner sum $\sum_{k=1}^{K} I(z_i = k)(\pi_k f_k\left(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_k\right))$ reduces to a single term $\pi_k f_k(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_k)$ with corresponding log given by $\log \pi_k + \log f_k\left(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_k\right))$.

Maximizing of l_c with respect to the model parameters $\boldsymbol{\theta}_k$, the model probabilities π_k as well as the latent variable $\boldsymbol{z}$ is performed iteratively by altering between an

- Expectation-step *(E-step)*, where an expectation Q of l_c with respect to $I(z_i = k)$ is derived given the current estimates of the unknown parameters in $\boldsymbol{\Psi}$, and a
- Maximization-step *(M-step)*, where an updated estimate for $\boldsymbol{\Psi}$ is obtained by maximizing this expectation.

These two steps are performed alternately until convergence.

More detailed, in the context of finite mixture models the following procedure is applied:

- *Specification of starting values*
  To begin, an initial choice of starting values is made. This can either be an initial assumption for the parameters $\boldsymbol{\Psi}$ or a partition of observations into K groups. See McLachlan and Peel (2000, chap. 2.12) for a discussion of different initialization strategies and related convergence properties. As slow convergence was not a problem in the performed simulation studies and the real-data application presented below, we opted for a random assignment of z_i, i.e. for a random partition of areas into K subgroups.
- *E-step*
  The expectation of the complete-data log-likelihood l_c with respect to $I(z_i = k)$ is derived as

$$\begin{aligned} Q(\boldsymbol{\Psi}) &= E[l_c(\boldsymbol{\Psi})] \\ &= E\left[\sum_{i=1}^{m}\sum_{k=1}^{K} I(z_i = k)(\log \pi_k + \log f_k\left(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_k\right))\right] \\ &= \sum_{i=1}^{m}\sum_{k=1}^{K} \hat{\xi}_{i,k}^{(t)}(\log \pi_k + \log f_k\left(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_k\right)), \end{aligned} \tag{15}$$

  where $\hat{\xi}_{i,k}$ denotes the conditional expectation for $z_i = k$ given $y_i, \boldsymbol{x_i}$ and the current estimate for $\boldsymbol{\Psi}$. They are calculated using $\widehat{\boldsymbol{\Psi}}^{(t-1)}$, i.e. the estimates for $\boldsymbol{\theta}_k$ and π_k obtained in the last iteration step $(t-1)$ (or, in the first iteration, the chosen starting values):

$$\begin{aligned} \hat{\xi}_{i,k}^{(t)} &= Pr(z_i = k|\widehat{\boldsymbol{\Psi}}^{(t-1)}, y_i, \boldsymbol{x}_i) \\ &= \frac{\hat{\pi}_k^{(t-1)} f_k(y_i|\boldsymbol{x}_i, \hat{\boldsymbol{\theta}}_k^{(t-1)})}{\sum_{j\in K} \hat{\pi}_j^{(t-1)} f_j(y_i|\boldsymbol{x}_i, \hat{\boldsymbol{\theta}}_j^{(t-1)})} \end{aligned} \tag{16}$$

- *M-step*
  In the M-step an updated estimate $\widehat{\boldsymbol{\Psi}}^{(t)}$ for $\boldsymbol{\Psi}$ is obtained by maximizing Q as derived in the E-step. It is obvious from (15) that optimization for $\boldsymbol{\theta}_k$ can be done for each model separately by maximizing the weighted component-specific log-likelihood $\sum_{i=1}^{m} \hat{\xi}_{i,k}^{(t)} \log f_k\left(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}_k\right)$. For the FMFH-model this

was solved numerically via the BFGS-algorithm using the R-package maxLik (see Henningsen & Toomet, 2011).

Deriving an update for π_k requires maximizing $\sum_{i=1}^{m}\sum_{k=1}^{K}\hat{\xi}_{i,k}^{(t)}\log\pi_k$ under the constraint $\sum_{k=1}^{K}\pi_k = 1$. The result is

$$\hat{\pi}_k^{(t)} = \frac{1}{m}\sum_{i=1}^{m}\hat{\xi}_{i,k}^{(t)}, \tag{17}$$

i.e. $\hat{\pi}_k^{(t)}$ is obtained by taking the average of $\hat{\xi}_{i,k}^{(t)}$ over all observations.

- *Termination*
  Both steps are repeated until the likelihood improvement in a step is smaller than an *ex ante* specified threshold ϵ, that is until $L(\boldsymbol{\Psi}^{(t)}) - L(\boldsymbol{\Psi}^{(t-1)}) < \epsilon$.

The likelihood $L(\boldsymbol{\Psi})$ is never decreased after an iteration step so that – for a sequence of likelihood values bounded above – convergence of the algorithm is guaranteed. For multimodal distributions this might, however, be convergence to a local maximum (see Dempster et al., 1977; Wu, 1983). To overcome this issue, the algorithm is usually applied repeatedly with different starting values. In both the simulation study and the application presented below we adopted this simple strategy. For a detailed account of convergence properties of the EM-algorithm see Dempster et al. (1977), Wu (1983) and McLachlan and Krishnan (2008).

3.3 Simulation Study

To analyse the proposed estimator, a Monte Carlo (MC) simulation study was performed. Several populations were generated to study the performance of the suggested methods under different scenarios. The choice of scenarios and the design of the populations with respect to covariates, fixed model coefficients and variance components was guided by the characteristics of the application motivating our proposal (see section 3.4). In what follows we present results for two selected populations.

Population 1:

- $K = 1$
- $\boldsymbol{\beta}_{k=1} = (2.75, 0.15, 0.3)$
- $x_2 \sim N(5, 1.5)$ and $x_3 \sim N(4.5, 1)$
- $\sigma_u^2 = 0.07$
- $\sigma_{e,i}^2 \sim unif(0.02, 0.22)$

Population 2:

- $K = 2$ and $\pi_k = 1/K$ for all k
- $\boldsymbol{\beta}_{k=1} = (2.75, 0.15, 0.3)$ and $\boldsymbol{\beta}_{k=2} = (6.5, -0.15, 0.0)$
- $x_2 \sim N(5, 1.5)$ and $x_3 \sim N(4.5, 1)$
- $\sigma_{u,k}^2 = 0.07$ for all k

- $\sigma^2_{e,i} \sim unif(0.02, 0.22)$

The aim of population 1 was to analyse the consequences of a misspecification in the sense of assuming a mixture distribution when the population actually is homogeneous. With population 2 a scenario was considered where the areas actually are segmented into two equally-sized subgroups. In figure 3.3 the distribution of the true conditional probabilities that area i belongs to class 1 are depicted. The first boxplot illustrates the distribution of $\xi_{i,1}$ for areas actually belonging to class 1, whereas the second boxplot shows $\xi_{i,1}$ for areas belonging to class 2. It can be seen that the two subgroups, although indeed separated, do overlap to a certain degree. We consider this a more realistic scenario than a sharp separation of components.

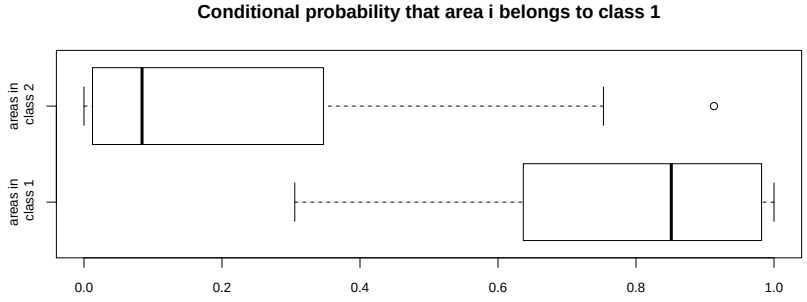

Figure 1: Conditional probabilities $\xi_{i,1}$ for class 1

We performed a MC simulation study with 10,000 runs, applying the following procedure: For each of the described scenarios we drew $\mu_i = \boldsymbol{x}_i^T \boldsymbol{\beta}_k + v_{i,k}, i = 1, \ldots, 200$ from the superpopulation once. The random effect was then held fixed over the simulation runs. Randomness was induced through the sampling variance of the direct estimate, that is $e_i, i = 1, \ldots, 200$ was drawn from $N(0, \sigma^2_{e,i})$ in each run.

In each MC-iteration, the specified model was estimated 30 times and the estimation with the largest likelihood was chosen as the final result. For each scenario the algorithm converged in all iterations. We obtained the proposed FMFH-estimator as well as the alternative mixture-model based estimator suggested in (10). Furthermore, we estimated the Standard FH-model and the FH-model with a correctly assigned dummy-variable for subgroup identity as benchmark estimators.

For the evaluation of results we considered the MC relative bias as well as the MC relative root mean square error:

$$RBias_i = \frac{\frac{1}{R}\sum_{r=1}^{R}(\hat{\mu}_{i,r} - \mu_i)}{\mu_i} \tag{18}$$

$$RRMSE_i = \frac{\sqrt{\frac{1}{R}\sum_{r=1}^{R}(\hat{\mu}_{i,r} - \mu_i)^2}}{\mu_i} \tag{19}$$

A crucial result of the simulation study was obtained from population 1. Although we do not gain accuracy compared to the standard FH-estimator, when

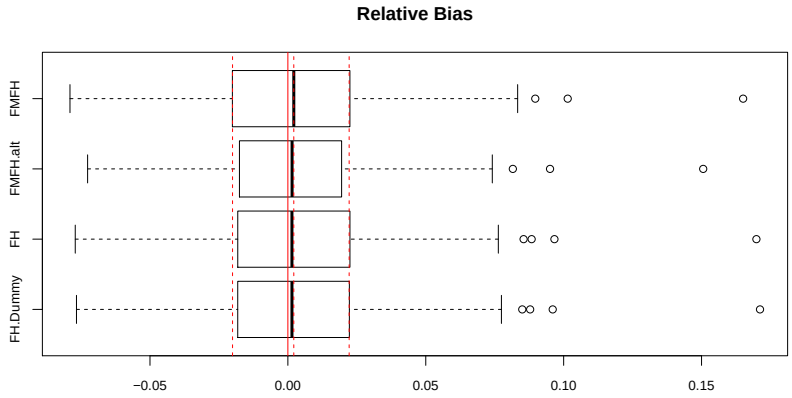

Figure 2: Relative Bias for Population 1

the assumption of the existence of subgroups is wrong), we do not lose it either (see figure 3). The improvement realized trough the application of SAE methods instead of direct estimation is retained. Even more importantly, this kind of misspecification does not cause a considerable amount of bias (see figure 2). As shown in figure 4, where the MC-expectations for the FH-estimator are plotted against those for the FMFH-estimator, the expectation for the predicts obtained from the FH-model and the FMFH-model is almost equal for all areas.

In population 2, the assumption of the existence of two distinct subgroups of areas actually was fulfilled. While the depiction of the relative bias shows almost no difference between the different estimators (see figure 5), figure 6 illustrates that the RRMSE can be reduced when employing the FMFH- instead of the standard FH-estimator. This improvement was realized despite the construction of the two subgroups as partly overlapping. A sharper separation of classes would further increase the superiority of the FMFH-estimator over the standard approach. The FH-estimator with dummy, included as a reference, clearly outperformed the proposed method. As it uses the true subgroup identity, this is an expected result. If subgroup identity were known, there would, however, be no need to employ a mixture model approach. The aim of the proposed method of course is, to come as close as possible to the results obtained with knowledge of subgroup identity in cases where this information is unobserved.

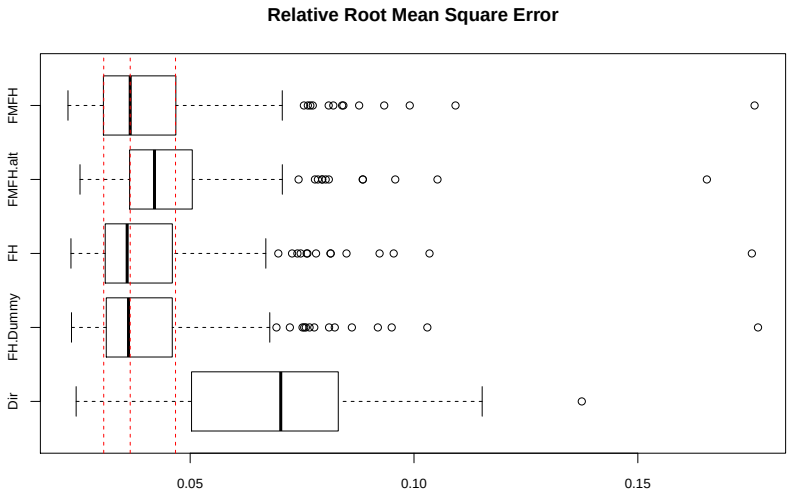

Figure 3: Relative Root Mean Square Error for Population 1

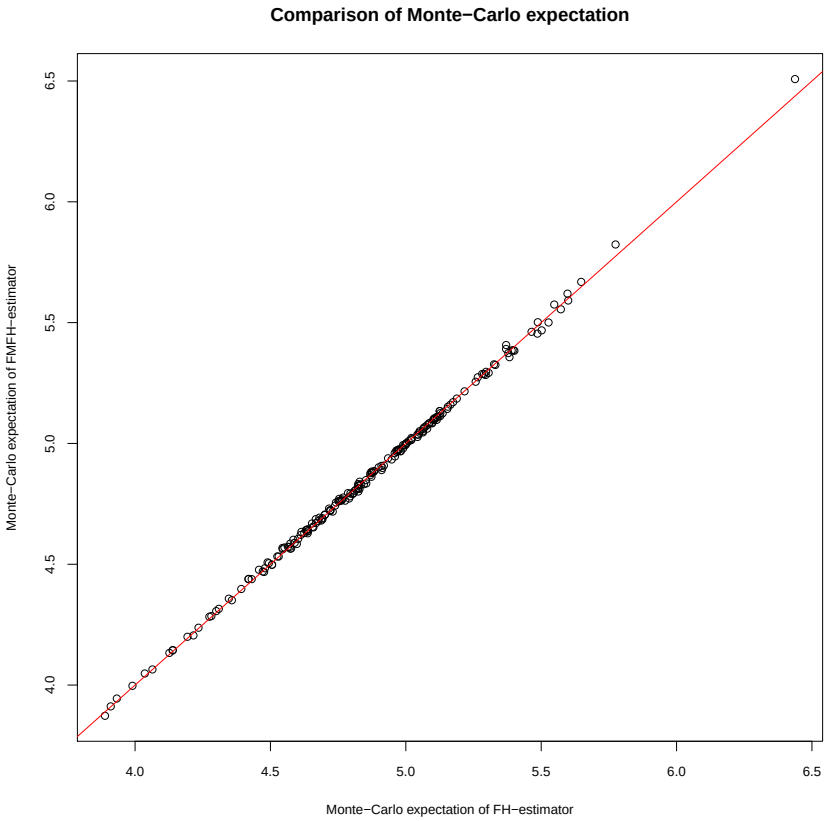

Figure 4: MC-expectations under Population 1

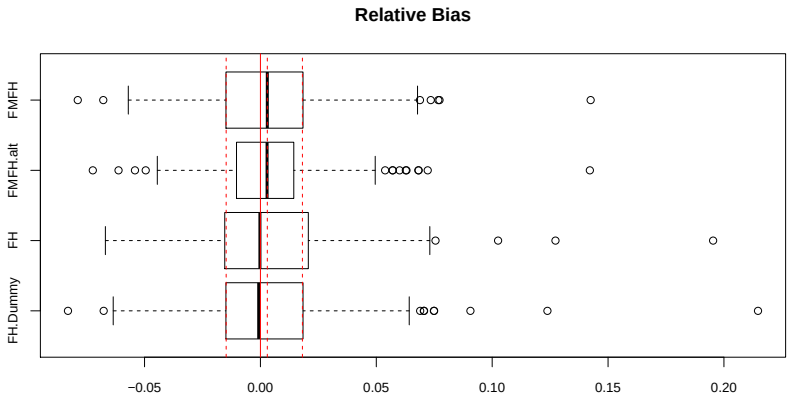

Figure 5: Relative Bias for Population 2

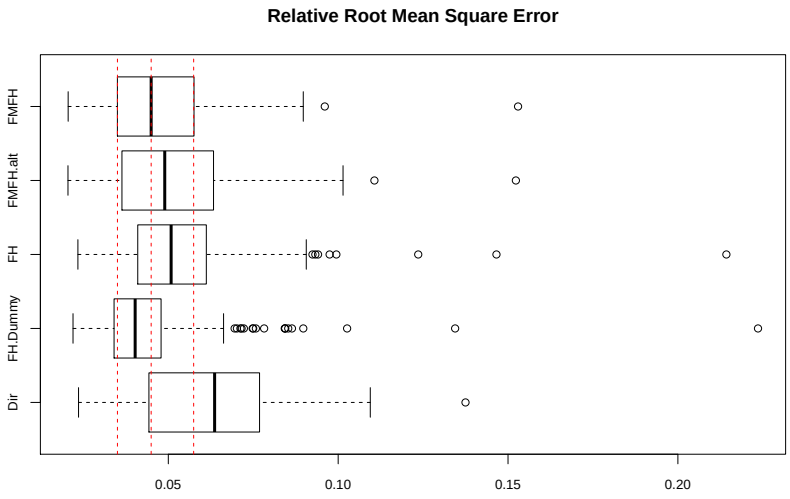

Figure 6: Relative Root Mean Square Error for Population 2

3.4 Application: Estimating Regional Rental Prices for German Districts

The suggested method was motivated by our intention to estimate rental prices for Germany on district level (NUTS-3). In the following we present this application and discuss the employment of standard SAE methods as well as the improvement realized when utilizing the proposed estimator.

Direct estimates for the estimation of regional rental prices were obtained from the German *Mikrozensus 2010*, an important 1%-household survey with a special section on housing every fourth year. The sample is drawn as a cluster sample with repeated stratification. For further information see Statistisches Bundesamt (2011, 2012). What is important to us is that results are usually not published on district level because they do not fulfill precision requirements. For the analysis at hand, a special evaluation was provided by the Federal Statistical Office. It contained average rents per square meter on district level for 246 of the 412 German districts.[3] Furthermore, district-specific sample sizes n_i as well as the estimated design variances of the direct estimates were provided. As frequently done in practical applications, we set $\hat{\sigma}^2_{e,i} = \sigma^2_{e,i}$, i.e. we used these estimates as the presumably known variances of the direct estimates, thereby ignoring the variability of the estimates (For a discussion of the implications of this assumption, see Bell (2008)).

We started by estimating the standard FH-model. Auxiliary information could be obtained from a broad range of regional indicators on district level provided by official statistics in Germany. Variable selection was performed by pursuing an literature-based analysis of important driving factors of rental prices as well as by applying simple stepwise selection procedures. As a model selection criterion we used the conditional AIC as suggested by Vaida and Blanchard (2005). Based on the results a model including indicators of tension on local rental markets as well as of purchasing power and attractiveness of a region was chosen. See table 1 for an overview.

Table 1: Auxiliary information

population growth rate	PGRO
relevance of rented housing	RENT
vacancy rate	VACR
employment rate	EMPL
net migration rate	MIGR
price of building land	LAND

The results were promising. The assumption of a common fixed part of the model for all districts, however, seemed inappropriate and was criticized when presenting the model to practitioners. We, therefore, adopted the proposed approach and estimated a finite mixture of FH-models with $K = 2$ and the same set of covariates as above for both components.

[3]Three of the federal states, namely Hesse, Bavaria, and Baden-Württemberg, did not give their approval to use the data.

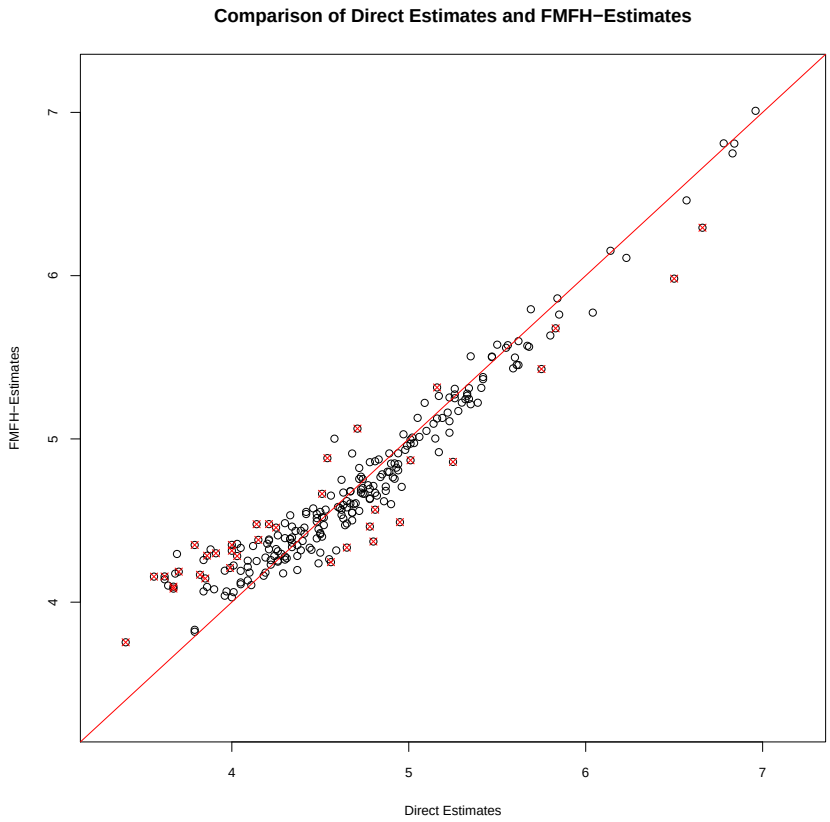

Figure 7: Comparison of Direct Estimates and FMFH-Estimates

We used the porposed estimator to calculate FMFH-estimates of average rental prices per square meter as a weighted mean of the predicts obtained from the two models. In figure 7, we plotted FMFH-estimates against the available direct estimates. Thus, the unbiased but imprecise direct estimates obtained from the sample are deployed to judge the bias of model-based FMFH-estimates. This simple plot has been suggested as a tool for bias diagnostic by Brown, Chambers, Heady, and Heasman (2001). To further analyse large deviations between direct and model-based estimates (that might either stem from a relatively high variance of the direct estimate or from a large difference between direct and synthetic estimate), data points for districts where $|\hat{\mu}_i^{FMFH} - \hat{\mu}_i^{Dir}| > \sigma_{e,i}$ are indicated by a cross. Note that deviations are acceptable if they are due to a large $\sigma_{e,i}$ yet problematic if the standard error of direct estimates is relatively small and deviations are caused by a considerable discrepancy between synthetic and direct estimates.[4]

The plot indicates a slight bias in the results for the cheapest districts. This affects approximately 10 of the 246 areas considered here and is a problem we observed for the standard FH-model as well. The results for the remaining regions appear to be fairly good. There are, however, few regions where the distance between direct estimate and FMFH-estimate is larger than the standard deviation of the

[4]Therefore, Fay and Herriot (1979) suggested a corresponding boundary to shrinkage towards the synthetic component in the calculation of FH-estimates.

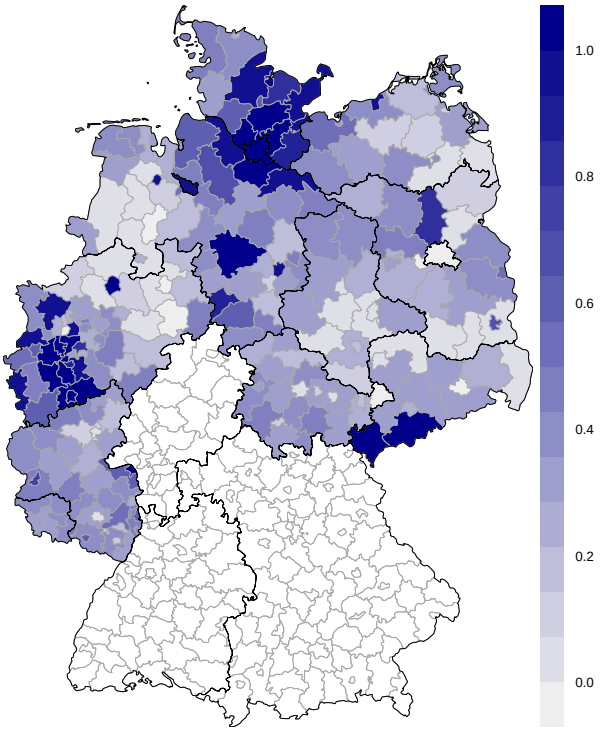

Figure 8: Conditional Probabilities for model 1

direct estimate.

A natural question arising is, what factors determine the conditional probabilities for subgroup identity. Put differently, the result of probabilistic clustering realized in the estimation process is of interest. We, therefore, plotted the estimated area-specific conditional probabilities $\hat{\xi}_{i,1}$ in a map (see figure 8). Note that $\hat{\xi}_{i,1} + \hat{\xi}_{i,2} = 1$, which makes the illustration of conditional probabilities for model 2 redundant. The spatial representation of $\hat{\xi}_{i,1}$ hints at some kind of agglomeration effect, as both the districts around Hamburg and the Rhineland-region are strongly assigned to model 2.

The results for rental prices on district level are illustrated in figure 9. Estimated prices range from approximately 3.70 to 7.00 Euro per square meter. As expected, the map clearly shows the particularly high prices in large cities such as Hamburg, Cologne, Düsseldorf, and Mainz. Rural districts in Eastern Germany and some areas in Rhineland-Palladium are identified as especially low-priced. As indicated above, Hesse, Bavaria, and Baden-Württemberg did not provide direct estimates for the analysis at hand.

The aim of applying SAE techniques is to realize a gain in accuracy in the context of small subsamples and, hence, large standard errors of traditional direct estimates. With figure 10 the performance of the suggested approach is evaluated in this regard. It depicts boxplots for the distribution of the estimated root mean square error (RMSE) of the proposed FMFH-estimates and the standard FH-estimates as well as of the Standard Deviation (SD) of the direct estimates. The red stars mark the respective average RMSE and SD over the areas. MSE

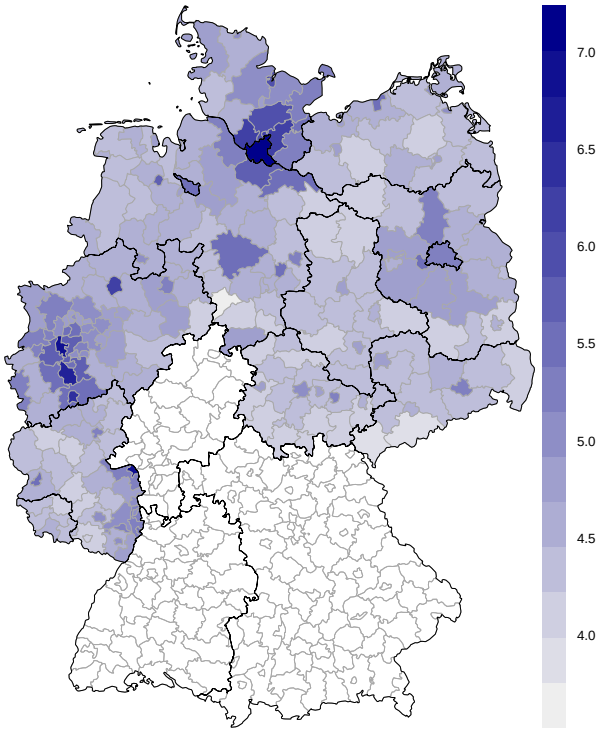

Figure 9: FMFH-estimates for regional rental prices

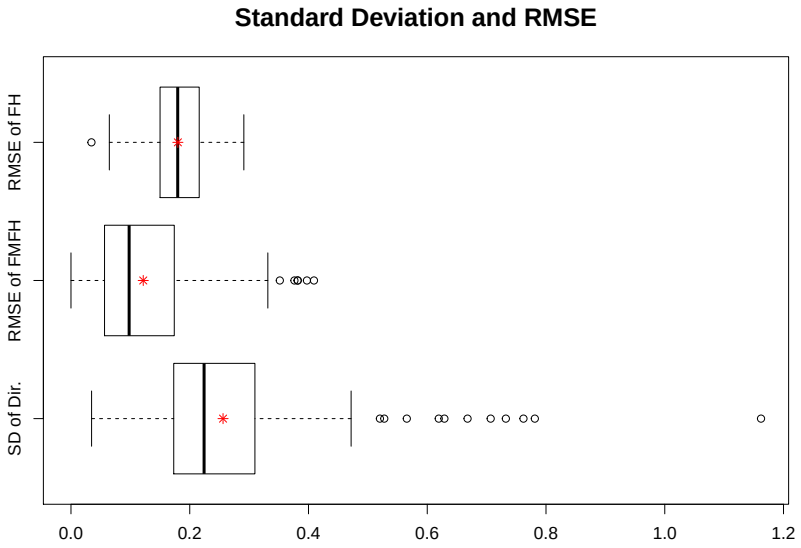

Figure 10: RMSE of Small Area Estimates and SD of Direct Estimates

estimation for the FMFH-estimator is performed applying a double-bootstrap as suggested by Chatterjee and Lahiri (2007). As expected, the plot shows that a significant gain in accuracy can be realized when applying SAE methods instead of direct estimation. Comparing the proposed approach and the standard FH-model, a considerably improvement can be made for almost all areas when applying the FMFH-estimator instead of standard methods. Correspondingly, the FMFH-estimator has a smaller average RMSE. There is, however, a small number of areas for which the standard method yielded better results.

4 Conclusion

Starting from the notion that the assumption of a common model for all areas might be inappropriate in many applications of SAE, we proposed a finite mixture of FH-models to account for the existence of unobserved or unobservable subgroups of areas. More specifically, we assumed that the statistic of interest is appropriately modelled by a mixture of K FH-models. As discussed above, this is not only a possible way to account for the existence of subgroups of areas, but can also be interpreted as a flexible semiparametric way to model unknown distributional shapes. We introduced and discussed an corresponding estimator, the FMFH-estimator, which is formulated as the weighted mean of predicts from the mixture components.

The simulation study, performed to analyse the suggested estimators, showed good results: The proposed method outperformed the standard FH-estimator in terms of RRMSE. Furthermore, the scenario considered with population 1 showed that this is even true, when the assumption of different subgroups is false. Finally, we obtained reasonable first results for the application to the problem of estimating regional rental prices for Germany.

Further work on mixture distributions in SAE is under progress. First, to carry on with the work presented here, we are working on the MSE estimation for the proposed estimator. As described above, first attempts with a double-bootstrap as suggested by Chatterjee and Lahiri (2007) were made and showed reasonable preliminary results. Calculation time, however, was daunting and so far prohibited the inclusion of MSE estimation into the simulation study. Second, we are currently considering a finite mixtures of unit-level models. A corresponding working paper is in progress. Finally, work remains to be done on model selection and diagnostics for the proposed estimator. Improved techniques of model selection will, hopefully, also lead to even better results for the application presented above.

References

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

Booth, J., Casella, G., & Hobert, J. (2008). Clustering using objective functions and stochastic search. *Journal of the Royal Statistical Society. Series B*, *70*, 119–139.

Brown, G., Chambers, R., Heady, P., & 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. achieving data quality in a statistical agency: A methodological perspective.*

Celeux, G., Martin, O., & Lavergne, C. (2005). Mixture of linear mixed models for clustering gene expression profiles from repeated microarray experiments. *Statistical Modelling*, *5*, 1–25.

Chatterjee, S., & Lahiri, P. (2007). A simple computational method for estimating mean squared prediction error in general small-area model. In *Proceedings of the section on survey research methods* (pp. 3486–3493).

Datta, G. S., & Lahiri, P. (1995). Robust hierarchical bayes estimation of small area characteristics in the presence of covariates and outliers. *Journal of Multivariate Analysis*, *54*, 310–328.

Datta, G. S., & Mandal, A. (n.d.). *Small area estimation with uncertain random effects.*

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–38.

Du, Y., Kahili, A., Neslehova, J. G., & Steele, R. J. (2013). Simultaneous fixed and random effects selection in finite mixture of linear mixed-effects models. *Canadian Journal of Statistics*, *41*(4), 596–616.

Elbers, C., & van der Weide, R. (2014). *Estimation of normal mixtures in a nested error model with an application to small area estimation of poverty and inequality* (Tech. Rep.). World Bank Group.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Chichester: John Wiley & Sons.

Fay, R. E., & Herriot, R. A. (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.

Gershunskaya, J. (2010). Robust small area estimation using a mixture model. In *Proceedings of the section on survey research methods* (pp. 2783–2796).

Ghosh, M., & Rao, J. N. K. (1994). Small area estimation: An appraisal. *Statistical Science*, *9*(1), 55–93.

Henderson, C. R., Kempthorne, O., Searle, S. R., & von Krosigk, C. M. (1959). The estimation of environmental and genetic trends from

records subject to culling. *Biometrics*, *15*(2), 192–218.

Hennig, C. (2000). Identifiablity of models for clusterwise linear regression. *Journal of Classification*, *17*(2), 273–296.

Henningsen, A., & Toomet, O. (2011). maxlik: A package for maximum likelihood estimation in R. *Computational Statistics*, *26*(3), 443–458.

Jiang, J., & Lahiri, P. (2006). Mixed model prediction and small area estimation. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, *15*(1), 1–96.

Maiti, T. (2003). Modelling small area effects using mixture of gaussians. *The Indian Journal of Statistics*, *65*, 612–625.

Maiti, T., Ren, H., Dass, S. C., Lim, C., & Maier, K. S. (in press). Clustering-based small area estimation: An application to meap data. *Calcutta Statistical Association Bulletin*.

Martella, F., Vermunt, J. K., Beekman, M., Westendorp, R. G. J., Slagboom, P. E., & Houwing-Duistermaat, J. J. (2011). A mixture model with random-effects components for classifying sibling pairs. *Statistics in Medicine*, *30*(27).

Martinez, M. J., Lavergne, C., & Trottier, C. (2009). A mixture model-based approach to the clustering of exponential repated data. *Journal of Multivariate Analysis*, *100*, 1938–1951.

McLachlan, G. J., & Krishnan, T. (2008). *The EM algorithm and extensions* (2nd ed.). Hoboken: Wiley.

McLachlan, G. J., Ng, S. K., & Wang, K. (2008). Clustering via mixture regression models with random effects. In *COMPSTAT. Proceedings in computational statistics.*

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models.* New York: John Wiley & Sons.

Münnich, R. T., Burgard, J. P., & Vogt, M. (2013). Small Area-Statistik: Methoden und Anwendungen. *AStA Wirtschafts- und Sozialstatistisches Archiv*, *6*(3), 149–191.

Ng, S. K., McLachlan, G. J., Wang, K., Ben-Tovim Jones, L., & Ng, S.-W. (2006). A mixture model with random-effects components for clustering correlated gene-expression profiles. *Bioinformatics*, *22*(14), 1745–1752.

Pereira, L. N., & Coelho, P. S. (2013). Estimation of house prices in regions with small sample sizes. *The Annals of Regional Science*, *50*(2), 603–621.

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, *28*(1), 40–68.

Rao, J. N. K. (2003). *Small area estimation.* New York: John Wiley & Sons.

Ren, H. (2011). *Some new models for small area estimation* (Unpublished doctoral dissertation). Michigan State University.

Scharl, T., Grün, B., & Leisch, F. (2010). Modelling time course gene expression data with finite mixtures of linear additive models. *Bioinformatics*, *26*(3), 370–377.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components*. New York et al.: John Wiley & Sons.

Statistisches Bundesamt. (2011). *Mikrozensus 2010. Qualitätsbericht*. Wiesbaden: Statistisches Bundesamt.

Statistisches Bundesamt. (2012). *Bauen und Wohnen. Mikrozensus-Zusatzerhebung 2010*. Wiesbaden: Statistisches Bundesamt.

Titterington, D. M., Smith, A. F. M., & Makov, E. (1985). *Statistical analysis of finite mixture distributions*. Chichester: John Wiley & Sons.

Vaida, F., & Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, *92*(2), 351–370.

Verbeke, G., & Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random effects population. *Journal of the American Statistical Association*, *91*(433), 217–221.

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *The Annaly of Statistics*, *11*(1), 95–103.

Yau, K. K. W., Lee, A. H., & Ng, S. K. (2003). Finite mixture regression model with random effects: Application to neonatal hospital length of stay. *Computational Statistics & Data Analysis*, *41*, 359–366.