Small Area Estimation for Survey Data:

A Hierarchical Bayes Approach

by

Milana Karaganis

A Thesis submitted to the Faculty of Graduate Studies of

The Univesity of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Statistics

University of Manitoba

Winnipeg

Abstract

Model-based estimation techniques have been widely used in small area estimation. This thesis focuses on the Hierarchical Bayes (HB) estimation techniques in application to small area estimation for survey data.

We will study the impact of applying spatial structure to area-specific effects and utilizing a specific generalized linear mixed model in comparison with a traditional Fay-Herriot estimation model. We will also analyze different loss functions with applications to a small area estimation problem and compare estimates obtained under these loss functions.

Overall, for the case study under consideration, area-specific geographical effects will be shown to have a significant effect on estimates. As well, using a generalized linear mixed model will prove to be more advantageous than the usual Fay-Herriot model. We will also demonstrate the benefits of using a weighted balanced-type loss function as opposed to the usual squared-loss function for the purpose of balancing the precision of estimates with their closeness to the direct estimates. In summary, methods considered in this thesis will be shown to provide estimates with better qualities than the usual HB models and methods used in small area estimation for survey data.

Acknowledgements

The completion of this thesis would not have been possible were it not for the support, patience, and generosity of many people. First and foremost, I am very grateful to Dr. Brad Johnson, my advisor, who has been very generous with his time and invaluable support and advice. He has also been very patient and has shown great faith in me by allowing me to proceed with this research on my own timetable.

I would also like to thank the faculty of the Department of Statistics, University of Manitoba, for all the help and support throughout my studies and for flexibility to complete studies at my own pace, for which, I am truly thankful.

The data for Chapters 3, 4, and 5 of this thesis came from Statistics Canada. I would like to thank everyone at Statistics Canada who believed in me, supported my returning to school, and helped me with their advice and time. A special thank-you goes to my supervisors at Statistics Canada, Ms. Carol Dunstone and Mr. C. Jerry Page, for encouraging me to go back to school and for accommodating my study schedule after I chose to do so.

Finally, I want to thank my family and especially my mother who came all the way from Russia to give me the moral support to persevere when the going was tough. A special thank-you belongs to my little son, who was born in the middle of my studies but was kind enough to allow me time to finish the thesis.

Table of contents

List of tables

List of figures

List of equations

Chapter 1

Introduction

1.1 Background and Motivation for Research

Obtaining estimates from survey data has long been recognized as the most cost efficient way to supply the information required for various purposes ranging from government funds allocation to tailoring marketing campaigns for new products. Survey estimates are typically obtained for the total population of interest as well as for smaller socio-demographic and geographic groups. In recent years, a new trend has developed with more users requiring more and more granular estimates, i.e. estimates for small domains defined as a combination of socio-demographic characteristics such as age range, sex, and family status within municipalities or even neighborhoods. Small businesses are interested in obtaining local data in order to tailor their business models. Large companies are looking to better estimate market potential of new products. Governments make decisions and allocate funds at the neighborhood and municipality level.

Historically, survey data have been utilized to provide reliable direct estimates at national and provincial levels. However, survey samples were not designed to account for an ever increasing need to produce estimates for smaller geographic entities or socio-demographic small domains. Following

Rao (2003, Ch. 1), we define a domain as large if “the domain sample size is large enough to yield direct estimates of adequate precision". If the domain sample size is not large enough to produce direct estimates of adequate precision, the domain is termed as a “small” domain. A typical example of a small domain in the context of national surveys conducted by Statistics Canada would be a domain defined as “females, aged between 20 and 29, living in specific municipalities or even neighborhoods”.

In response to an increasing demand for small area (domain) estimates, statistical methods have been developing in two major directions. In the first approach, the sample designs are adapted to produce domain estimates of adequate quality (precision) using collected survey data and standard design-based computational methods (so called “direct estimates"). Typically, domains of interest have to be specified during the survey design stage to allow for the allocation of a large enough sample to each domain. The drawback of this approach is an increase in the overall sample size, which leads to an increase in data collection costs. Another drawback of this approach is that there is no guarantee that users would be able to produce estimates for domains that were not specified during the survey design stage. For example, if initial domains were specified as a combination of age range, sex, and income status at a provincial level, there is no guarantee that the sample in each municipality within the province will be large enough to provide overall population estimates for each municipality.

The second major approach to small area estimation is not to rely solely on domain data from a particular survey but to "borrow strength" by using other available data for the domains of interest and related areas to produce indirect model-dependent estimators. In other words, "indirect" estimators are based not only on the collected survey data for small domains and variables of interest but on all other available data as well. The ultimate goal is to produce indirect estimators that would have adequate precision as opposed to the direct estimators that typically have unacceptably large standard errors. Obviously, availability of auxiliary data and proper linking models are prerequisite to producing reliable indirect estimators. This second approach constitutes the essence of the small area model-based estimation methods.

Small area model-based estimation methods can be broadly divided into two groups – methods based on implicit linking models and methods based on explicit linking models.

Indirect estimators produced by implicit linking models (synthetic and composite estimators) are based on the assumption that there is an adequate direct estimator for a larger area that one can "borrow strength" from to produce indirect estimators for the small areas. These estimators are typically design-based in the sense that survey weights are used and the sample design induces the probability distribution that is used for determination of confidence intervals and standard errors. The major drawback of implicit linking models is the assumption that small areas possess the same characteristics as larger areas. Typically this is not true and the resulting estimators will be exposed to bias.

Often, the bias can be significant, especially if the goal of the study is to locate domains with atypical characteristics (e.g., municipalities with high incidence level of a rare disease within a particular province).

Explicit linking models specify area-specific random effects and have an explicit mechanism to account for the between-area variation. Often, estimators obtained from the explicit linking models are called model-based estimators. Generally, model-based estimators are classified either as area-level models (where "area" is taken to mean a small domain of interest and models are specified at the area level rather than at the level of units constituting those areas) or unit-level models (where a collection of "units" constitutes a small domain of interest and models are specified at the unit level). Area-level models operate with area-level covariates and estimate area-level quantities while unit-level models use unit-level covariates and estimate unit-level quantities. In both cases, the assumed model induces a probability distribution that is used for the determination of confidence intervals and standard errors. Hence, model selection and validation are crucial to ensure that resulting estimators are not model-biased.

Using explicit linking models has proven to be more advantageous for both practitioners and theoreticians (*cf.* Rao, 2003, Ch. 1). Explicit models allow researchers to better understand and model sources of variability (random effects). Various models can be compared in terms of their fit to the sample data and model diagnostics can be performed on them. Models can operate with area-specific measures of precision and complex models can be enter-

tained (generalized linear mixed models, non-linear models, etc.). Finally, explicit linking models allow the smoothing of raw direct estimates while still differentiating small domains and identifying domains with abnormally high or low values of the characteristics of interest.

The three most popular model-based estimators are the empirical best linear unbiased prediction (EBLUP); empirical Bayes (EB); and hierarchical Bayes (HB) estimators. EBLUP estimators are obtained by minimizing the model mean-squared error (MSE) amongst the class of linear unbiased estimators while using empirical estimators of the variances and covariances of the random effects. HB and EB estimators are obtained by applying the Bayesian approach and Bayes' Theorem. EBLUP estimators are applicable to linear mixed models while EB and HB can be utilized with more general models (e.g., count data or binomial models) as well.

Until recently, computational limitations impacted the ease of applying Bayesian methods and using HB and EB estimators in complex models. With the development of Markov Chain Monte Carlo (MCMC) methods and software designed to handle MCMC computations more efficiently, HB and EB methods have become increasingly popular and have been applied to various practical problems in a large and expanding number of areas (several examples are provided in the next several paragraphs). At the same time, philosophical objections to the use of the Bayesian methodology as opposed to the frequentist methods have resulted, among other things, in a more careful utilization of prior distributions. In particular, utilization of prior distributions with little

informative content (e.g., non-informative priors) to allow the collected data to be the determining factor in impacting on the posterior distribution has become more widespread.

These recent developments have led to successful applications of HB and EB methods to various practical problems. Below, we provided several examples of such applications.

- Public health research has been utilizing HB and EB methods to map incidence rates of diseases such as cancer (disease mapping). The resulting estimates allow for the analysis of variation between different geographic areas to pinpoint areas with abnormally high / low incidence rates. The resulting maps can be used for resource allocation decisions as well as to determine intervention programs. Typically, however, sampling is not utilized in disease mapping applications; rather, administrative records are used and the related data quality considerations and concerns are reflected in the applied methods (e.g., record completeness and granularity as opposed to the quality of sampling frame and respresentativeness of a selected sample in survey setup).
- Various authors have applied HB methods to the estimation of unemployment rates in small areas (domains) - see Datta *et al*. (1999) (estimation of US unemployment rates); Chung *et al*. (2001) (estimation of Korean unemployment rates); and You *et al*. (2001) (estimation of Canadian unemployment rates).

- You and Rao (2002) applied HB methods to 1991 Canadian Census data to estimate the undercoverage rate.
- Bell (1999) applied HB methods to obtain estimates of a number of school-age children living in poverty at the US county and state levels.

Even this short list of studies performed by various authors shows the range of the goals and objectives that can be achieved by utilizing Bayesian methods. At the same time, it also illustrates the potential of cross-utilizing different estimation techniques depending on the objective of the study. For example, disease mapping applications have traditionally been concerned with the geographic aspects of disease distribution as well as with identifying areas with low or high incidence rates. Hence, great effort was put into the development of tools to capture and apply information about the geographical structure of small areas through spatio-temporal modeling techniques. On the model side, for rare diseases, either HB Normal models are used to fit the disease rates or HB Poisson models are applied to the disease counts (as an approximation to the binomial distribution).

On the other hand, one of the main objectives, when producing small area estimates based on survey data, has been to improve their precision (i.e. to reduce the standard error and coefficient of variation of estimates). At the same time, since survey data is often collected for specific needs (policy decision making or funds allocation, for example), there has been a practical concern by the data users to ensure that the model-based estimates are closely comparable with the frequentist estimates. Therefore, typically, basic models like the Fay-

Herriot model (*cf*. Fay and Herriot, 1979) or the Rao-Yu model (*cf*. Rao and Yu, 1992; and Rao and Yu, 1994) are applied. These models assume that the area-specific random effects are independent and identically distributed. This is a very strong assumption since, in practice, many characteristics in geographically adjacent areas typically exhibit a degree of correlation that can be quite high.

Overall, the needs of the user community have been evolving and there has been a steady and growing demand to use survey data to produce estimates for smaller and smaller geographic areas and socio-demographic domains as well as the identification of geographic patterns and domains with abnormal values of characteristics of interest. Analysis of spatially distributed data has become one of the most important analytical tools in many real-life applications such as targeted marketing campaigns; crop assessments; crime prevention programs; assessment of Census undercount rates (*cf*. Cressie, 1991); and investigation of the effects of rainfall on malaria incidence (*cf*. Nobre *et al*., 2005).

Hierarchical Bayes (HB) small area estimation methods have been adjusted to include spatial effects with the resulting models and techniques being extensively utilized in such applications as disease mapping problems. Porting these models and methods over the survey small area estimation problems would allow more realistic assumptions regarding spatial association between small areas and, hopefully, would lead to better estimates. At the same time, the nature of the many variables of interest (binary or count values, for example) would suggest that generalized linear mixed models may be more suitable

than the usual Fay-Herriot model or its extensions and can lead to better estimates.

For this thesis, we decided to apply HB spatial estimation models and methods as well as a specific generalized linear mixed model for the estimation of a particular variable in small socio-geographical domains. The resulting estimates will be compared with the results produced by the more traditional Fay-Herriot model.

Specifically, we decided to use data collected by the Canadian Community Health Survey (CCHS) 2.1 cycle for the analysis and to select "Number of people who had a flu shot more than two years ago" as a variable to be estimated for all age groups within each Health Region in Canada.

The choice of a dataset and a variable to estimate was motivated by the research objectives in the following sense. Generalized linear mixed models, such as Poisson, are more suitable for count data, for example. The CCHS dataset includes many different count variables to choose from. CCHS sample sizes are large enough to allow estimation at a Health Region level. Another big advantage of the CCHS data is that the survey included many basic variables that can be used as covariates in the small-area models (such as age, gender, occupation, income, etc.).

On the other hand, choice of a variable to be estimated had a two-fold motivation. As mentioned above, HB spatial models and estimation techniques have been extensively used in disease mapping applications. Disease mapping problems focus on estimating either the number of cases of a particular disease

(asthma, cancer, etc.) or a disease rate. By using a variable of a different nature (i.e., neither disease counts nor disease rates), the estimation techniques can be applied in a more general setting. At the same time, as Health Canada suggests (*cf.* www.hc-sc.gc.ca/dc-ma/influenza/index-eng.php): "The most effective way to protect yourself from the flu is to be vaccinated each year in the fall." Thus, estimating the number of people who had a flu shot more than two years ago can serve as an indicator of populations at risk (both with respect to geographic zones as well as socio-demographic domains) in order to prepare and prevent a potential flu (or any other related disease such as SARS or swine flu) outbreak. Therefore, developing estimates for such a variable (or similar variables) can also be considered in a more general setting as focusing on monitoring and prevention as opposed to rare disease counts and rates estimation.

In the second portion of the thesis, we will turn to another aspect of the Bayesian approach - utilization of different loss functions to reflect the problems at hand. Analyses in this area were motivated by the following considerations.

The traditional approach used by Hierarchical Bayes small area estimation models has been to assume a squared-error loss function (SEL) and to use the posterior mean as a parameter estimator. However, this approach, while attempting to mirror frequentist estimators as well as traditional measures of accuracy (MSE and CV, for example), fails to utilize an important aspect of the Baye-

sian philosophy, namely, the ability to structure a loss function in response to the nature of the problem.

While squared-error loss is considered to be the most reasonable choice to obtain estimates for the HB Normal models, estimators obtained from the other generalized linear mixed models (Poisson, for example) may not perform well with respect to squared-error loss. Additionally, a squared-error loss function does not assign much of a penalty when estimates are close to zero, which can be a problem in small-area estimation where counts and proportions are typically small.

Furthermore, surveys are typically conducted in response to specific needs. As a consequence, survey results are often used to make important decisions such as the allocation of transfer payments, program funds distribution, etc. Hence, assuming squared-error loss may not adequately reflect the needs of the data users and may fail to assign the proper penalty for underestimation or overestimation of the parameters of interest. In practice, however, it is difficult to come up with exact loss function specifications. Therefore, it may be necessary to obtain estimators with respect to different loss functions, to compare different estimators and to assess if any particular estimator would dominate the others with respect to estimated posterior loss.

1.2 Thesis Outline

As mentioned above, the main focus of this thesis is the Hierarchical Bayes estimation techniques in application to small area estimation based on survey data. We will apply spatial structure to the area-specific effects; utilize a generalized linear mixed model (Poisson) in comparison with the traditional Fay-Herriot model; and will obtain and compare several estimates under different loss functions.

This thesis consists of six chapters. Other than the current Chapter, which has provided a brief overview of small area estimation methods and motivation for the research presented in this thesis, the remainder of the thesis is organized as follows. Chapter 2 provides a more detailed introduction into model-based small area estimation. The main models utilized in small area estimation are described. Three major model-based approaches to the derivation of small area estimators - Empirical Best Linear Unbiased Prediction (EBLUP), Empirical Bayes (EB), and Hierarchical Bayes (HB) - are presented. While the three model-based estimation approaches are presented, in the later chapters, only the HB method will be utilized in our applications. Finally, Markov Chain Monte Carlo sampling methods and their application to the Bayesian estimation are introduced.

In Chapter 3, we present the problem and the dataset that was used to study and compare the different models as well as the estimators produced under the different loss functions. Four different models are proposed for the purpose of

comparing the results obtained by applying spatial structure to the area-specific effects with the results obtained by treating the area-specific effects as independent and identically distributed. As well, models are structured to allow comparison between an extension of the Fay-Herriot model and a generalized linear mixed model (Poisson). Chapter 3 also describes the covariate selection process for each model under consideration.

In Chapter 4, we describe the implementation of the four proposed models via Markov Chain Monte Carlo computational methods and derive the required conditional distributions and resulting estimators. Chain convergence and model fit assessment techniques are applied and the results are presented. At the end of Chapter 4, four proposed models and direct estimates are compared in terms of model fit and precision to assess the impact of applying spatial structure to the area-specific effects as well as the impact of a generalized linear mixed model.

In Chapter 5, we present a brief overview of different loss functions that have been utilized in various small area estimation applications. We propose several loss functions that can be utilized in the context of the case study analyzed in the thesis and derive corresponding estimators. We then apply these proposed loss functions to obtain numeric estimates. At the end of Chapter 5, estimates are compared to determine if any particular estimator would produce better results than the others with respect to posterior risk and with respect to estimated posterior loss.

Finally, in Chapter 6, we present and discuss some suggestions for future research.

Chapter 2

Introduction to Model-Based Small Area Estimation

As described in Section 1.1, small area model-based estimation methods utilize either explicit or implicit linking models. In this thesis, we will focus on the explicit linking models as these models have proven to be more advantageous (*cf*. Rao, 2003, Ch. 1).

In the next section, we will describe the most often used small area models that permit empirical best linear unbiased prediction (EBLUP) and Bayes prediction (Hierarchical and Empirical Bayes). Both area-level and unit-level models are introduced in this chapter; however, the rest of the thesis will focus on area-level models.

The brief overview of small area models presented in this section is not intended as an exhaustive literature review. The objective is to present the most often used models. For a more comprehensive overview, please see Rao (2003, Ch. 5).

2.1 Small Area Models

2.1.1 Area-Level Models

In this section, we will describe the most often used area-level models - the basic Fay-Herriot model (*cf.* Fay and Herriot, 1979); its extension to include temporal effects (*cf.* Rao-Yu model; Rao and Yu, 1992; Rao and Yu, 1994); and another extension to include spatial effects (*cf.* Besag *et al.*, 1991).

Throughout, we make use of the following notational conventions and assumptions.

- We assume that there are m small areas and parameters θ_i , $i= 1,...,m$ to be estimated (one for each small area).
- θ_i , $i= 1,...,m$ are functions of observed values obtained from survey data. Once the data is collected, we obtain direct estimates of the θ_i, which are denoted as $\hat{\theta}_i$.
- There are p auxiliary variables and $\boldsymbol{x}_i^T = (x_{1\,i},...,x_{p\,i})^T$ is a vector of auxiliary area-specific data for area i. Typically, it is assumed that the auxiliary variables are known. In practice, auxiliary variables can be estimated from survey data, and once estimates are obtained, they would be treated as known. We will follow this approach and will use auxiliary variables that will be estimated from survey data.
- $\boldsymbol{\beta} = (\beta_1,...,\beta_p)^T$ is a vector of regression coefficients.
- $\boldsymbol{\psi} = (\psi_1,...,\psi_m)^T$ is a vector of area-specific random effects.

- $\mathbf{e} = (e_1,...,e_m)^T$ is a vector of sampling errors.

Fay-Herriot Model

The Fay-Herriot model consists of a linking model and a sampling model. The linking model is specified as a linear model

$$\theta_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + b_i \psi_i, \quad i = 1,....,m,$$

where the b_i are assumed to be positive known constants, and the area-specific random effects ψ_i are assumed to be independent and identically-distributed with zero mean and constant variance; i.e., $E(\psi_i) = 0$ and $Var(\psi_i) = \sigma_\psi^2$ is assumed unknown.

The sampling model assumes independent sampling errors

$$\hat{\theta}_i = \theta_i + e_i, \quad i = 1,....,m,$$

where $E(e_i | \theta_i) = 0$ and $Var(e_i | \theta_i) = \sigma_{e_i}^2$ are known constants. In practice, $\sigma_{e_i}^2$ are estimated from data and treated as known. However, this approach does not take into account the variation in an estimate, and, hence, leads to underestimation of variance of $\hat{\beta}$. Different methods were proposed to estimate variance of sampling errors - for example, a common design effect approach proposed by You (*cf.* You, 2006; and You, 2008).

Combining the linking and sampling models, we obtain the Fay-Herriot model as the following

$$\hat{\theta}_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + b_i \psi_i + e_i, \quad i = 1,....,m \qquad (1)$$

where ψ_i and e_i are assumed to be independent. This model allows for design-induced errors (e_i) as well as model random effects (ψ_i).

This area-level model was first introduced by Fay and Herriot (1979) to estimate per capita income for small places (population less than 1,000) in the United States. It has been widely used ever since its introduction, for example, see Ericksen and Kadane (1985), Cressie (1989), Dick (1995), National Research Council (2000), Datta *et al*. (1991), Fay (1987), etc.

Rao-Yu Model

Many surveys are designed to track sampling units over time. For example, in the Canadian Labour Force Survey (LFS), sampled households are interviewed every month for six consecutive months. In the Canadian National Longitudinal Survey of Children and Youths (NLSCY), once an initial interview with a sampled household is completed and respondents are identified (both parents and children), respondents remain in the longitudinal sample and are interviewed every two years. In each survey, combining data over time can lead to an increase in the effective sample size, and, hence, to an increase in precision of small area estimates.

Rao and Yu (1992, 1994) considered the problem of handling time series in the context of the Fay-Herriot model and proposed the following extension

- a linking model:

$$\theta_{it} = \boldsymbol{x}_{it}^T \boldsymbol{\beta} + \psi_i + u_{it}, \; i = 1,....,m; \; t = 1,...,T, \qquad (2)$$

- a sampling model:

$$\hat{\theta}_{it} = \theta_{it} + e_{it}, \; i = 1,....,m; \; t = 1,...,T,$$

where

- $\hat{\theta}_{it}$ is a direct estimator for small area *i* at time *t*;
- e_{it} are sampling errors that are assumed to be normally distributed with zero means and block covariance matrix $\mathbf{\Psi}$ with known blocks $\mathbf{\Psi}_i$;
- $\boldsymbol{x}_{it}^T$ is a vector of area-specific covariates that may vary over time;
- the area-specific random effects ψ_i are assumed to be independent and identically distributed N (0 , σ_ψ^2) with σ_ψ^2 unknown;
- u_{it} are area-by-time effects that are assumed to follow a first order autoregressive process

$$u_{it} = \rho\, u_{i,t-1} + \varepsilon_{it},\ |\rho| < 1,\ \varepsilon_{it} \overset{\text{iid}}{\sim} \text{N}(0, \sigma^2);$$

  where σ^2 is unknown; and ρ is an autoregressive parameter that may be known or unknown.
- $e_{it}, \psi_i, \varepsilon_{it}$ are assumed to be independent.

Other authors have proposed different cross-sectional models with a temporal component for small area estimation - see Ghosh and Nangia (1993); Ghosh *et al*. (1996); Pfeffermann and Burck (1990); Datta *et al*. (1999).

Spatial Models

One of the assumptions used in the Fay-Herriot model is the assumption of area-specific random effects being independent and identically distributed with $\text{E}(\psi_i) = 0$ and $\text{Var}(\psi_i) = \sigma_\psi^2$ unknown. This is the assumption of geographically unstructured heterogeneity. In other words, underlying parameters are assumed to be exchangeable, there is no spatial dependency, and all individual area-specific estimates are displaced towards a global mean.

However, in the small area setup, geographically adjacent areas are often small enough to induce correlations among the area-specific random effects ψ_i. Hence, adding a spatial component is often required within the small area estimation setup to capture the assumption of geographically structured heterogeneity. In other words, individual parameters are assumed to be strongly influenced by the parameters in the neighborhood areas and much less influenced by the parameters in the remote areas. The individual area-specific parameter estimates are then displaced towards a local mean.

Typically, spatial dependency is modeled through random effects γ_i in the linking model (that are intended to represent geographically structured heterogeneity of the parameters θ_i)

$$\theta_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \gamma_i, \quad i= 1,...,m.$$

In the Fay-Herriot model (1), area-specific random effects ψ_i are assumed to be independent and identically distributed. In other words, it is assumed that ψ_i are not affected by geographic location and by neighboring areas. In the spatial models, the assumption of independency is lifted and area-specific random effects γ_i are assumed to be dependent and affected by geographic location and by neighboring areas.

Various spatial distributions and models have been proposed with the intrinsic Gaussian conditional autoregression (CAR) being one of the most often used distributions.

The intrinsic Gaussian CAR (also known as Gaussian Markov random field) was introduced by Besag *et al*. (1991) through the following joint distribution

$$\boldsymbol{\gamma} | \lambda \propto \lambda^{-m} exp \left\{ -\frac{1}{2\lambda^2} \sum_{i=1}^{m} \sum_{j<i} w_{ij} \left(\gamma_i - \gamma_j\right)^2 \right\}. \qquad (3)$$

Here, the parameter λ indicates the strength of the spatial correlation. Smaller values of λ are associated with stronger spatial correlations between adjacent areas. Setting $\lambda = 0$ shrinks conditional distributions to a common value. When λ approaches infinity, spatially structured variation increases accordingly. However, both situations are special cases of a model with exchangeable prior. The w_{ij} are the weights that reflect local spatial dependence between two areas. Typically, the weights w_{ij} are set as 1 if *i* and *j* are adjacent areas and 0 otherwise.

With this choice of weights, it is easy to obtain the full conditional distribution of $\gamma_i | \boldsymbol{\gamma}_{(-i)}, \lambda$ - it is given by the following normal distribution

$$\gamma_i | \boldsymbol{\gamma}_{(-i)}, \lambda \sim N(\bar{\gamma}_i, V_i) = N\left(\frac{\sum_{j \neq i} w_{ij} \gamma_j}{\sum_{j \neq i} w_{ij}}, \frac{\lambda^2}{\sum_{j \neq i} w_{ij}} \right) \qquad (4)$$

However, it has to be noted that this choice for the weights leads to an improper joint density (*cf.* Best *et al.*, 2005). Typically, the sum-to-zero constraint is added to resolve this issue and it is assumed that $\sum_{i=1}^{m} \gamma_i = 0$. While this approach can be difficult to justify from a theoretical point of view, it is easily implemented in Markov Chain Monte Carlo algorithms.

The intrinsic Gaussian CAR model can be rewritten to show that it is equivalent to specifying a multivariate normal model for the joint distribution of the area-specific random effects. An important feature of the model is the presence of the parameter λ that indicates the degree of spatial dependence - smaller values are associated with stronger spatial correlations between areas. Alternative models based on the conditional autoregressive principle were

suggested by Besag *et al*. (1991), MacNab (2002b), MacNab (2003b), Leroux *et al*. (1999), etc. Throughout the thesis, we will refer to the original intrinsic autoregressive model (3)-(4) as a CAR distribution for shortness.

In practice, though, it is often difficult to differentiate between a purely unstructured or a purely spatially structured model. Hence, Besag *et al*. (1991) suggested a compromise model that includes both of those components

$$\theta_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + b_i \psi_i + \gamma_i , \quad i = 1,...,m.$$

Here, the random effects γ_i are still intended to represent the spatial structure i.e. the structured heterogeneity of the parameters θ_i, and random effects ψ_i represent the unstructured pattern of heterogeneity. However, there is always a risk that ψ_i and γ_i cannot be separated and only the sum $b_i \psi_i + \gamma_i$ can be identified from the data.

One concern about the conditional autoregressive models is related to the fact that *all* areas are affected by the same parameter λ that determines the degree of smoothing (shrinkage towards local means). Semi-parametric spatial models were developed in response to this concern. These models operate by partitioning areas into clusters with each cluster having a different level of smoothing. The number of clusters is treated as a variable to be estimated. See Green and Richardson (2002), Knorr-Held and Raber (2000), and Denison and Holmes (2001) for examples of these models.

While the models described above typically operate with a spatial structure in terms of adjacency of the areas, another approach used to capture spatial structure is based on the inter-area distances $d_{ij} = |s_i - s_j|$, where the $\{s_i\}$

denote some reference points in each area (e.g., center points). Then, the joint distribution is given as

$$\boldsymbol{\gamma} \mid \lambda \sim N(\mu \ \mathbf{1}_m \ , \lambda^2 \ R_\lambda) \ ,$$

where

- $\mathbf{1}_m$ is a $(m \ x \ 1)$-vector of 1's;
- $R_\lambda = (r_{i\,j})$ is a matrix with elements $r_{i\,j}$ that are functions of $|s_i - s_j|$, i.e. $r_{i\,j} = f_\lambda(|s_i - s_j|)$ possibly dependent on a parameter λ. The most often used functions are exponential, power exponential, and spherical (*cf.* Cressie, 1993 and Stein, 1999 for more details).

As demonstrated by the brief summary presented above, various spatial distributions and models have been proposed and used in different applications. However, within the context of small area estimation, the intrinsic CAR models have been the most popular due to their flexibility and ease of implementation based on the adjacency matrix for the spatial structure and the conditional distributions of parameters.

2.1.2 Unit-Level Models

In this section, we will describe three unit-level models most often used in the small area estimation context.

For unit-level models, the following notations will be used.

- $y_{i\,j}$ denotes the variable of interest to be estimated for each population element j, $j = 1,...,N_i$ in each small area i, $i = 1,...,m$.

There are p auxiliary variables and $\boldsymbol{x}_{ij} = (x_{ij1},...,x_{ijp})^T$ is a vector of auxiliary unit-specific data for the unit j in the area i. Here, x_{ijp} represents the p-th auxiliary variable for element j in area i. Typically, it is assumed that auxiliary variables are known for each population element j in each small area i.

The simplest unit-level model used in small area estimation is a one-fold nested error linear regression model

$$y_{ij} = x_{ij}^T \boldsymbol{\beta} + \psi_i + e_{ij}, \quad j = 1,...,N_i; \quad i = 1,...,m; \tag{5}$$

where

- $\boldsymbol{\beta} = (\beta_1,...,\beta_p)^T$ is a vector of regression coefficients;
- The ψ_i's are area-specific random effects that are assumed to be independent and identically distributed with $E(\psi_i) = 0$ and $Var(\psi_i) = \sigma_\psi^2$ (unknown);
- The e_{ij}'s are residual errors that are assumed to satisfy the following condition - $e_{ij} = k_{ij}\,\tilde{e}_{ij}$ - with known constants k_{ij} and $\tilde{e}_{ij}$ being independent and identically distributed random variables with $E(\tilde{e}_{ij}) = 0$ and $Var(\tilde{e}_{ij}) = \sigma_e^2$ (could be known or unknown).
- ψ_i and $\tilde{e}_{ij}$ are assumed to be independent.

An underlying assumption here is that the model is valid for the entire population, i.e. for all population units. In other words, sampled units and non-sampled units follow the same model stated above. This assumption is satisfied for the simple random sampling design as well as for sampling designs that use x_{ij}^T as the auxiliary information for sample selection. However, if different covariates, z_{ij}^T, are used to determine selection probabilities, then the

distribution of sample data depends on z_{ij}^T, which introduces sample selection bias and we can not assume that model (5) holds for all units in the population. As well, model (5) is not correct for two-stage cluster sampling as this model does not include cluster effects. Solutions have been proposed to overcome both of these limitations (*cf.* Skinner, 1994; Stukel and Rao, 1999).

For example, if we consider a two-stage cluster sampling design, the model can be adjusted as the following to obtain a two-fold nested error regression model (*cf.* Stukel and Rao, 1999)

$$y_{ijl} = x_{ijl}^T \boldsymbol{\beta} + \psi_i + \vartheta_{ij} + e_{ijl}\,;\; l = 1,...,N_{ij};\; j = 1,...,M_i;\; i = 1,...,m;$$

where

- in the i-th small area, we assume M_i primary sampling units (clusters) and in the j-th primary sampling unit, we assume N_{ij} elements. The underlying assumption is that clusters are completely included in corresponding small areas and there is no clusters that would cross boundaries of small areas;
- y_{ijl} and $\boldsymbol{x}_{ijl}$ denote the variable of interest and a covariate vector (of length p) for the l-th element in the j-th primary unit from the i-th area;
- ψ_i are area-specific random effects that are assumed to be independent and identically distributed with $E(\psi_i) = 0$ and $Var(\psi_i) = \sigma_\psi^2$ (unknown);
- ϑ_{ij} are cluster effects that are assumed to be independent and identically distributed with $E(\vartheta_{ij}) = 0$ and $Var(\vartheta_{ij}) = \sigma_\vartheta^2$ (could be known or unknown);
- e_{ijl} are residual errors that are assumed to satisfy the following condition - $e_{ijl} = k_{ijl}\,\tilde{e}_{ijl}$ with known constants k_{ijl} and $\tilde{e}_{ijl}$ being independent and

identically distributed random variables satisfying $E(\tilde{e}_{ijl}) = 0$ and $Var(\tilde{e}_{ijl}) = \sigma_e^2$ (could be known or unknown); and

- ψ_i, ϑ_{ij}, and e_{ijl} are assumed to be mutually independent.

This model is valid for both sampled and unsampled units in the case of simple random sampling of clusters and subunits within sampled clusters. It will also hold for sampling designs that use $\boldsymbol{x}_{ijl}$ as auxiliary information to determine selection probabilities.

Finally, a general linear mixed model can also be considered (*cf.* Datta and Ghosh, 1991)

$$\boldsymbol{y}^P = \boldsymbol{X}^P \boldsymbol{\beta} + \boldsymbol{Z}^P \boldsymbol{\psi} + \boldsymbol{e}^P, \qquad (6)$$

where

- $(\cdot)^P$ denotes a population model;
- $\boldsymbol{y}^P$ is the vector of population values (Nx1);
- $\boldsymbol{\beta}$ is a vector of regression coefficients;
- $\boldsymbol{X}^P$ and $\boldsymbol{Z}^P$ are design matrices that are assumed to be known;
- $\boldsymbol{\psi} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{D}(\boldsymbol{\lambda}))$ is a vector of random effects with a positive definite variance-covariance matrix that has a known structural and functional form dependent on the parameters $\boldsymbol{\lambda}$;
- $\boldsymbol{\lambda}$ typically takes the form of ratios of variance components;
- $\boldsymbol{e}^P \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{\Psi}^P)$ is a vector of residual errors with a known positive definite variance-covariance matrix and is assumed to be independent of $\boldsymbol{\psi}$.

This concludes a brief overview of the unit-level models and we now turn to generalized linear mixed models.

2.1.3 Generalized Linear Mixed Models

In this section, we will describe the most often used generalized linear mixed models - a logistic regression model and a Poisson model. The first type of model, the logistic regression model, is suitable for binary data.

Let's assume that for unit j in area i, $y_{ij} = 0$ (with probability $1-p_{ij}$) or 1 (with probability p_{ij}) based on some criterion, N_i is the size of the i-th small area, and the small area proportion is defined as $P_i = \sum_j y_{ij} / N_i$ (proportion of unit that satisfy the criterion). Then the logistic regression model is defined as the following (with the objective of estimating the P_i's)

- $y_{ij} \mid p_{ij}$ are assumed to be independent Bernoulli(p_{ij}) variables;
- logit(p_{ij}) = log $\frac{p_{ij}}{1-p_{ij}} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \psi_i$; where $\psi_i \sim N(0, \sigma_\psi^2)$ are area-specific random effects; and $\boldsymbol{x}_{ij}^T$ are unit-specific covariates.

Once the estimator $\hat{p}_{ij}$ is obtained from this model, then P_i is estimated as

$$\hat{P}_i = \frac{\sum_{\text{sampled } j \text{ from area } i} y_{ij} + \sum_{\text{non-sampled } j \text{ from area } i} \hat{p}_{ij}}{N_i}.$$

Disease mapping applications have widely used another type of generalized linear mixed model. Disease mapping problems aim at portraying the geographical distribution of a disease and locating areas with abnormally high (or abnormally low rates) that would require intervention. In these applications, the variable of interest is often a count, y_i, of disease instances in the i-th area. The basic model can be formulated as

- y_i are assumed to be independent Poisson variables with $E(y_i|\ \theta_i) = n_i\theta_i$; where n_i is the number of exposed individuals in the *i*-th area and θ_i is a true disease rate;
- $\theta_i \sim$ Gamma (α , ν) - a gamma distribution (here, gamma distribution density is assumed to have the following parametrization - $f(x) = \frac{x^{\alpha-1} e^{-x/\nu}}{\Gamma(\alpha)\,\nu^{\alpha}}$; $x > 0$).

A different model that has been very popular in the context of disease mapping applications involves the log of true disease rates (*cf.* Besag *et al*., 1991; Bernardinelli and Montomoli, 1992; Clayton and Kaldor, 1987; Clayton and Bernardinelli, 1992; Nandram *et al*., 1999; Xia *et al*., 1997). In that model, we have

- y_i are assumed to be independent Poisson variables with $E(y_i|\ \theta_i) = n_i\theta_i$; where n_i is the number of exposed individuals in the *i*-th area and θ_i is a true disease rate;
- $\text{Log}(\theta_i) = \boldsymbol{x}_i^T\,\boldsymbol{\beta} + \psi_i + \gamma_i$, where $\boldsymbol{x}_i^T$ are covariates, ψ_i are independent and identically distributed random effects for the exchangeable model, and γ_i are random effects accounting for the spatial structure.

An extension of these models was proposed for exponential families of distributions with canonical parameters ς_{ij} by Ghosh *et al*. (1998)

$$\varsigma_{ij} = x_{ij}^T\,\boldsymbol{\beta} + \psi_i + \vartheta_{ij},\ j = 1,\ ...,\ n_i;\ \ i = 1,\ ...,\ m;$$

where ψ_is and ϑ_{ij}s are mutually independent random effects that follow normal distribution with zero mean and unknown constant variances σ_ψ^2 and σ_ϑ^2, respectively. The ς_{ij}s can take different forms depending on the specific

model under consideration - $\varsigma_{ij} = \log(\theta_{ij})$ for Poisson distribution; $\varsigma_{ij} = \text{logit}(p_{ij})$ for Binomial distribution, etc.

This concludes our brief overview of the models most often used in the small area estimation setup. We will now proceed with the review of the three estimation methods most often applied to small area models while focusing on area-level models only.

2.2 EBLUP Approach

The empirical best linear unbiased prediction (EBLUP) was developed within the classical frequentist framework and consists of two steps. First, the best linear unbiased estimator is obtained by minimizing MSE in the class of linear unbiased estimators. Then, variances (covariances) of random effects are estimated by using one of the three methods: method of moments, the maximum likelihood method, or the restricted maximum likelihood method. These estimates are used in the best linear unbiased estimator to obtain the EBLUP estimator.

The EBLUP method is utilized for linear mixed models. For the derivation of EBLUP point estimators, the assumption of normality of random effects and sampling errors is not necessary, but normality is typically required for MSE estimation.

Let us consider the following model

$$\mathrm{y} = \mathrm{X}\beta + \mathrm{Z}\psi + \mathrm{e}\,, \tag{7}$$

where

- $\boldsymbol{y}$ is a vector of observations $\boldsymbol{y} = (\boldsymbol{y}_1^T,...,\boldsymbol{y}_m^T)^T$ and $\boldsymbol{y}_i$ is an $(n_i \; x \; 1)$-vector;
- $\boldsymbol{X} = \text{col}_{1 \le i \le m}(\boldsymbol{X}_i)$ and $\boldsymbol{X}_i$ is an $(n_i \; x \; p)$-matrix

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ ... \\ \boldsymbol{X_m} \end{pmatrix};$$

- $\boldsymbol{Z} = \text{diag}_{1 \le i \le m}(\boldsymbol{Z}_i)$ and $\boldsymbol{Z}_i$ is an $(n_i \; x \; h_i)$-matrix

$$\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{Z_1} & 0 & ... & 0 \\ 0 & \boldsymbol{Z}_2 & ... & 0 \\ ... & ... & \boldsymbol{...} & ... \\ 0 & 0 & ... & \boldsymbol{Z}_m \end{pmatrix};$$

- both $\boldsymbol{X}$ and $\boldsymbol{Z}$ are assumed to be known matrices of full rank;
- $\boldsymbol{\beta} = (\beta_1,...,\beta_p)^T$ is a vector of regression coefficients;
- $\boldsymbol{\psi} = \text{col}_{1 \le i \le m}(\boldsymbol{\psi}_i)$ are random effects, where $\boldsymbol{\psi}_i$ is an $(h_i \; x \; 1)$-vector and $\boldsymbol{\psi}$ has $\boldsymbol{0}$ mean and covariance matrix $\boldsymbol{G} = \text{diag}_{1 \le i \le m}(\boldsymbol{G}_i(\boldsymbol{\delta}))$ that depends on variance parameters $\boldsymbol{\delta} = (\delta_1,...,\delta_q)^T$

$$\boldsymbol{\psi} = \begin{pmatrix} \boldsymbol{\psi}_1 \\ ... \\ \boldsymbol{\psi}_m \end{pmatrix} \text{ and } \boldsymbol{G} = \begin{pmatrix} \boldsymbol{G_1} & 0 & ... & 0 \\ 0 & \boldsymbol{G}_2 & ... & 0 \\ ... & ... & \boldsymbol{...} & ... \\ 0 & 0 & ... & \boldsymbol{G}_m \end{pmatrix};$$

- $\boldsymbol{e} = \text{col}_{1 \le i \le m}(\boldsymbol{e}_i)$ are sampling errors, where $\boldsymbol{e}_i$ is an $(n_i \; x \; 1)$-vector and $\boldsymbol{e}$ has $\boldsymbol{0}$ mean and covariance matrix $\boldsymbol{R} = \text{diag}_{1 \le i \le m}(\boldsymbol{R}_i(\boldsymbol{\delta}))$ that depends on variance parameters $\boldsymbol{\delta} = (\delta_1,...,\delta_q)^T$

$$\boldsymbol{e} = \begin{pmatrix} \boldsymbol{e}_1 \\ ... \\ \boldsymbol{e}_m \end{pmatrix} \text{ and } \boldsymbol{R} = \begin{pmatrix} \boldsymbol{R_1} & 0 & ... & 0 \\ 0 & \boldsymbol{R}_2 & ... & 0 \\ ... & ... & \boldsymbol{...} & ... \\ 0 & 0 & ... & \boldsymbol{R}_m \end{pmatrix};$$

- $\boldsymbol{\psi}$ and $\boldsymbol{e}$ are assumed to be independent (unconditionally).

This model can be thought of as a composition of m submodels

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\psi}_i + \boldsymbol{e}_i$$

and each $\boldsymbol{y}_i$ has variance

$$\text{Var}(\boldsymbol{y}_i) = \boldsymbol{V}_i(\boldsymbol{\delta}) = \boldsymbol{R}_i + \boldsymbol{Z}_i\boldsymbol{G}_i\boldsymbol{Z}_i^T .$$

Models (1), (2), and (3) introduced so far are special cases of this model (as will be shown below). The goal is to estimate *m* linear combinations of the following form

$$\mu_i = \boldsymbol{l}_i^T\boldsymbol{\beta} + \boldsymbol{m}_i^T\boldsymbol{\psi}_i,\ i = 1,...,m.$$

It has been shown that the EBLUP estimator for this model has the following form (*cf*. Rao, 2003, Ch. 6 and Ch. 7)

$$\hat{\mu}_i^{\text{EBLUP}} = \hat{\mu}_i^{\text{EBLUP}}(\hat{\boldsymbol{\delta}}) = \boldsymbol{l}_i^T\hat{\boldsymbol{\beta}} + \boldsymbol{m}_i^T\hat{\boldsymbol{\psi}}_i,$$

where

$$\hat{\boldsymbol{\psi}}_i = \boldsymbol{G}_i\boldsymbol{Z}_i^T\boldsymbol{V}_i^{-1}(\hat{\boldsymbol{\delta}})(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}});$$

$$\hat{\boldsymbol{\beta}} = \left(\sum\nolimits_i \boldsymbol{X}_i^T\boldsymbol{V}_i^{-1}(\hat{\boldsymbol{\delta}})\boldsymbol{X}_i\right)^{-1}\left(\sum\nolimits_i \boldsymbol{X}_i^T\boldsymbol{V}_i^{-1}(\hat{\boldsymbol{\delta}})\boldsymbol{y}_i\right);$$

and $\hat{\boldsymbol{\delta}}$ is an estimator of $\boldsymbol{\delta}$ obtained either by the method of moments, the maximum likelihood method, or the restricted maximum likelihood method.

Under the conditions specified above, it was shown that the EBLUP estimator is unbiased (*cf*. Kacker and Harville, 1981) regardless of whether $\boldsymbol{\delta}$ is known or not.

Turning to MSE estimation, an estimator of MSE of $\hat{\mu}_i^{\text{EBLUP}}$ is typically approximated and the resulting formulae depends on the method used to derive $\hat{\boldsymbol{\delta}}$ (hence, formulae for different estimation methods will be presented below).

For the restricted maximum likelihood (REML) estimators and moment estimators of $\boldsymbol{\delta}$, a second-order approximation to the estimator of MSE of $\hat{\mu}_i^{\text{EBLUP}}$ is given by

$$\text{mse}_1\ (\hat{\mu}_i^{\text{EBLUP}}) \approx g_{1i}(\hat{\boldsymbol{\delta}}) + g_{2i}(\hat{\boldsymbol{\delta}}) + 2\, g_{3i}(\hat{\boldsymbol{\delta}}),$$

where

(a) $g_{1i}(\hat{\boldsymbol{\delta}}) = \boldsymbol{m}_i^T \left(\boldsymbol{G}_i - \boldsymbol{G}_i \boldsymbol{Z}_i^T \boldsymbol{V}_i^{-1}(\hat{\boldsymbol{\delta}}) \boldsymbol{Z}_i \boldsymbol{G}_i\right) \boldsymbol{m}_i$

is an estimator of the variance of the best linear unbiased estimator (this estimator is obtained by minimizing MSE in the class of linear unbiased estimators under the assumption that all variances and covariances are known);

(b) $g_{2i}(\hat{\boldsymbol{\delta}}) = (\boldsymbol{l}_i^T - \boldsymbol{m}_i^T \boldsymbol{G}_i \boldsymbol{Z}_i^T \boldsymbol{V}_i^{-1}(\hat{\boldsymbol{\delta}}) \boldsymbol{X}_i)^T$

$\times (\sum \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1}(\hat{\boldsymbol{\delta}}) \boldsymbol{X}_i)^{-1}$

$\times (\boldsymbol{l}_i^T - \boldsymbol{m}_i^T \boldsymbol{G}_i \boldsymbol{Z}_i^T \boldsymbol{V}_i^{-1}(\hat{\boldsymbol{\delta}}) \boldsymbol{X}_i)$

accounts for the variability in the estimator $\hat{\boldsymbol{\beta}}$;

(c) $g_{3i}(\hat{\boldsymbol{\delta}}) = \text{tr}\left[\dfrac{\partial\{\boldsymbol{m}_i^T \boldsymbol{G}_i \boldsymbol{Z}_i^T \boldsymbol{V}_i^{-1}(\hat{\boldsymbol{\delta}})\}}{\partial \hat{\boldsymbol{\delta}}} \times \boldsymbol{V}_i(\hat{\boldsymbol{\delta}}) \right.$

$\left. \times \left(\dfrac{\partial\{\boldsymbol{m}_i^T \boldsymbol{G}_i \boldsymbol{Z}_i^T \boldsymbol{V}_i^{-1}(\hat{\boldsymbol{\delta}})\}}{\partial \hat{\boldsymbol{\delta}}}\right)^T \overline{V(\hat{\boldsymbol{\delta}})} \right],$

where $\overline{V(\hat{\boldsymbol{\delta}})}$ is the asymptotic covariance matrix of $\hat{\boldsymbol{\delta}}$, so that $g_{3i}(\hat{\boldsymbol{\delta}})$ is an approximation to the squared bias of the EBLUP estimator that is due to the estimation of the variance component $\boldsymbol{\delta}$, i.e. $\text{E}(\hat{\mu}_i^{\text{EBLUP}}(\hat{\boldsymbol{\delta}}) - \hat{\mu}_i^{\text{EBLUP}}(\boldsymbol{\delta}))^2$.

This second-order approximation is valid under certain regularity conditions and is approximately unbiased (*cf.* Rao, 2003, Ch. 6 and Ch. 7).

For the maximum likelihood (ML) estimator of $\boldsymbol{\delta}$, the following second-order approximation to the estimator of MSE of $\hat{\mu}_i^{\text{EBLUP}}$ is valid

$$\text{mse}_2(\hat{\mu}_i^{\text{EBLUP}}) \approx g_{1i}(\hat{\boldsymbol{\delta}}) - \boldsymbol{b}_{\hat{\boldsymbol{\delta}}}^T(\hat{\boldsymbol{\delta}}) \nabla g_{1i}(\hat{\boldsymbol{\delta}}) + g_{2i}(\hat{\boldsymbol{\delta}}) + 2g_{3i}(\hat{\boldsymbol{\delta}}),$$

where

- $g_{1i}(\hat{\boldsymbol{\delta}})$, $g_{2i}(\hat{\boldsymbol{\delta}})$, and $g_{3i}(\hat{\boldsymbol{\delta}})$ are defined in (a), (b), and (c) above;
- $\boldsymbol{b}_{\hat{\boldsymbol{\delta}}}^T(\hat{\boldsymbol{\delta}})$ is an approximation to the bias $E(\hat{\boldsymbol{\delta}}) - \boldsymbol{\delta}$;
- $\nabla(\cdot)$ denotes the vector of the first-order derivatives.

Note that $\text{mse}_2(\hat{\mu}_i^{\text{EBLUP}})$ is approximately unbiased as well. For more details and proofs, please see Prasad and Rao (1990), Harville and Jeske (1992) and Das *et al*. (2001).

Let us now consider the Fay-Herriot model (1) and apply EBLUP estimation techniques to this model. As mentioned above, the Fay-Herriot model is a special case of the linear mixed model (7)

- $\boldsymbol{y}_i = \hat{\theta}_i$; $\boldsymbol{X}_i = \boldsymbol{x}_i^T$; $\boldsymbol{Z}_i = b_i$; $\boldsymbol{\psi}_i = \psi_i$; $\boldsymbol{e}_i = e_i$; $i = 1,...,m$;
- $\boldsymbol{\beta} = (\beta_1,...,\beta_p)^T$; $\boldsymbol{G}_i = \sigma_\psi^2$; $\boldsymbol{R}_i = \sigma_{e_i}^2$ (assumed to be known);
- $\boldsymbol{V}_i = \sigma_{e_i}^2 + \sigma_\psi^2 b_i^{\,2}$;
- $\mu_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + b_i \psi_i$ so that $\boldsymbol{l}_i = \boldsymbol{x}_i$ and $\boldsymbol{m}_i^T = b_i$.

Then the EBLUP estimator is

$$\begin{aligned}\tilde{\mu}_i^{\text{EBLUP}} = \tilde{\mu}_i^{\text{EBLUP}}(\hat{\boldsymbol{\delta}}) &= \boldsymbol{x}_i^T \tilde{\boldsymbol{\beta}} + \hat{\kappa}_i(\hat{\theta}_i - \boldsymbol{x}_i^T \tilde{\boldsymbol{\beta}}) \\ &= \hat{\kappa}_i \hat{\theta}_i + (1 - \hat{\kappa}_i)\, \boldsymbol{x}_i^T \tilde{\boldsymbol{\beta}},\end{aligned}$$

where

- $\hat{\kappa}_i = \dfrac{b_i^{\,2} \hat{\sigma}_\psi^2}{\sigma_{e_i}^2 + \hat{\sigma}_\psi^2 b_i^{\,2}}$;

- $\tilde{\boldsymbol{\beta}} = \left[\sum_{i=1}^{m} \frac{\boldsymbol{x}_i \boldsymbol{x}_i^T}{\sigma_{e_i}^2 + \hat{\sigma}_\psi^2 b_i^2} \right]^{-1} \times \left[\sum_{i=1}^{m} \frac{\boldsymbol{x}_i \hat{\theta}_i}{\sigma_{e_i}^2 + \hat{\sigma}_\psi^2 b_i^2} \right]$.

As shown above, the EBLUP estimator is a linear combination of the direct estimator $\hat{\theta}_i$ and the regression estimator $\boldsymbol{x}_i^T \tilde{\boldsymbol{\beta}}$. Weights used in the estimator emphasize either the regression estimator in the case of a small model variance or the direct estimator in the case of a small design variance.

This estimator is design-consistent with design bias being zero on average if the linking model is appropriate. $\hat{\sigma}_\psi^2$ is obtained either through the method of moments, the maximum likelihood (ML) method, or the restricted maximum likelihood (REML) method. When $\hat{\sigma}_\psi^2$ is obtained through ML or REML methods, it typically has smaller asymptotic variance than when obtained through the method of moments. Once $\hat{\sigma}_\psi^2$ is obtained, the EBLUP estimator will be model-unbiased since the Fay-Herriot model typically requires ψ_i and e_i to be independent and normally distributed with zero means.

Regularity conditions were derived to justify the second-order approximation to the estimator of MSE of the EBLUP estimator (*cf*. Rao, 2003, Ch. 6 and Ch. 7). Then, for the REML and moment estimators of σ_ψ^2, the second-order approximation to the estimator of MSE of $\tilde{\mu}_i^{\text{EBLUP}}$ is given by

$$\text{mse}_1\ (\tilde{\mu}_i^{\text{EBLUP}}) \approx g_{1i}(\hat{\sigma}_\psi^2) + g_{2i}(\hat{\sigma}_\psi^2) + 2\, g_{3i}(\hat{\sigma}_\psi^2),$$

where

(d) $g_{1i}(\hat{\sigma}_\psi^2) = \hat{\kappa}_i\ \sigma_{e_i}^2$;

(e) $g_{2i}(\hat{\sigma}_\psi^2) = (1 - \hat{\kappa}_i)^2\ \boldsymbol{x}_i^T \left[\sum_i \frac{\boldsymbol{x}_i \boldsymbol{x}_i^T}{\sigma_{e_i}^2 + \hat{\sigma}_\psi^2 b_i^2} \right]^{-1} \boldsymbol{x}_i$;

(f) for REML, $g_{3i}(\hat{\sigma}_\psi^2) = 2 \times \frac{b_i^4\, \sigma_{e_i}^4}{\left(\sigma_{e_i}^2 + \hat{\sigma}_\psi^2 b_i^2\right)^3} \left[\sum_i \frac{b_i^4}{\left(\sigma_{e_i}^2 + \hat{\sigma}_\psi^2 b_i^2\right)^2} \right]^{-1}$;

(g) for a moment estimator,

$$g_{3i}(\hat{\sigma}_\psi^2) = \frac{2}{m^2} \times \frac{b_i{}^4 \sigma_{e_i}^4}{\left(\sigma_{e_i}^2 + \hat{\sigma}_\psi^2 b_i{}^2\right)^3} \times \sum_{i=1}^{m} \frac{\left(\sigma_{e_i}^2 + \hat{\sigma}_\psi^2 b_i{}^2\right)^2}{b_i{}^4}.$$

For the ML estimator of σ_v^2, the second-order approximation to the estimator of MSE of $\tilde{\mu}_i^{\text{EBLUP}}$ is given by (as mentioned above, certain regularity conditions need to hold to justify the second-order approximation to the estimator of MSE of the EBLUP estimator)

$$\text{mse}_2\ (\tilde{\mu}_i^{\text{EBLUP}}) \approx g_{1i}(\hat{\sigma}_\psi^2) - b_{\hat{\sigma}_v^2}^T(\hat{\sigma}_\psi^2)\, \nabla g_{1i}(\hat{\sigma}_\psi^2) + g_{2i}(\hat{\sigma}_\psi^2) + 2g_{3i}(\hat{\sigma}_\psi^2),$$

where:

- $g_{1i}(\hat{\sigma}_\psi^2)$, $g_{2i}(\hat{\sigma}_\psi^2)$, and $g_{3i}(\hat{\sigma}_\psi^2)$ are defined in (d), (e), (f), and (g) above;
- $\nabla g_{1i}(\hat{\sigma}_\psi^2) = b_i{}^2\ (1 - \hat{\kappa}_i)^2$;
- $b_{\hat{\sigma}_v^2}^T(\hat{\sigma}_\psi^2) = -\left[2\,\mathcal{I}(\hat{\sigma}_\psi^2)\right]^{-1} \text{tr}\left[\left\{\sum_{i=1}^{m} \frac{\boldsymbol{x}_i \boldsymbol{x}_i^T}{\sigma_{e_i}^2 + \hat{\sigma}_\psi^2 b_i{}^2}\right\}^{-1} \times \left\{\sum_{i=1}^{m} \frac{b_i{}^2}{\left(\sigma_{e_i}^2 + \hat{\sigma}_\psi^2 b_i{}^2\right)^2} \boldsymbol{x}_i \boldsymbol{x}_i^T\right\}\right]$;

  where $\mathcal{I}(\hat{\sigma}_\psi^2) = \frac{1}{2} \sum_{i=1}^{m} \frac{b_i{}^4}{\left(\sigma_{e_i}^2 + \hat{\sigma}_\psi^2 b_i{}^2\right)^2}$.

These approximations to the MSE estimators are unbiased to the order of m^{-1} (m is a number of small areas). It has also been shown that the approximations to the MSE estimators depend on the normality assumption for the ψ_i 's (*cf*. Rao, 2003, Ch. 6 and Ch. 7).

$\text{mse}_1\ (\tilde{\mu}_i^{\text{EBLUP}})$ and $\text{mse}_2\ (\tilde{\mu}_i^{\text{EBLUP}})$ can be further modified to use a term $g_{3i}^*(\hat{\sigma}_\psi^2)$ instead of $g_{3i}(\hat{\sigma}_\psi^2)$ that will depend on the area-specific data $\boldsymbol{y}_i$ - in

this case, area-specific estimators of MSE can be obtained (*cf.* Rao, 2003, Ch. 6 and Ch. 7).

Let us now turn to the Rao-Yu model (2) in the context of EBLUP estimation. If we consider $\hat{\boldsymbol{\theta}}_i = (\hat{\theta}_{i1},...,\hat{\theta}_{iR})^T$, $i = 1,...,m$ then the model (2) can be considered a special case of model (7) with the following notations

- $\boldsymbol{y}_i = \hat{\boldsymbol{\theta}}_i$;
- $\boldsymbol{X}_i = (\boldsymbol{x}_{i1},...,\boldsymbol{x}_{iR})^T$; $\boldsymbol{Z}_i = (\mathbf{1}_R, \boldsymbol{I}_R)$; where $\mathbf{1}_R$ is the Rx1 vector of 1's and $\boldsymbol{I}_R$ is the RxR identity matrix;
- $\boldsymbol{\psi}_i = (\psi_i, \boldsymbol{u}_i^T)$ and $\boldsymbol{u}_i^T = (u_{i1},...,u_{iR})^T$;
- $\boldsymbol{G}_i = \begin{pmatrix} \sigma_\psi^2 & \mathbf{0}^T \\ \mathbf{0} & \sigma^2 \boldsymbol{\Lambda} \end{pmatrix}$ and $\boldsymbol{\Lambda}$ is the covariance matrix of $\boldsymbol{u}_i$;
- $\boldsymbol{e}_i = (\boldsymbol{e}_{i1},...,\boldsymbol{e}_{iR})^T$;
- $\boldsymbol{R}_i = \boldsymbol{\Psi}_i$; $\boldsymbol{V}_i = \boldsymbol{\Psi}_i + \sigma^2 \boldsymbol{\Lambda} + \sigma_\psi^2 \mathbf{1}_R \mathbf{1}_R{}^T$;
- $\boldsymbol{\beta} = (\beta_1,...,\beta_p)^T$;
- $\theta_{iR} = \mu_i = \boldsymbol{x}_{iR}^T \boldsymbol{\beta} + \boldsymbol{m}_i{}^T \boldsymbol{\psi}_i$ so that $\boldsymbol{l}_i = \boldsymbol{x}_{iR}$ and
- $\boldsymbol{m}_i = (1, 0,..., 0, 1)^T$.

When ρ is considered to be known in either the AR(1) or the random walk model, the REML method (*cf.* Datta *et al*., 2002) or the simple transformation method (*cf.* Fuller and Battese, 1973) can be used to obtain unbiased estimators of σ_ψ^2 and σ^2. Once the estimators are obtained, the EBLUP estimator can be obtained

$$\tilde{\theta}_{iR}^{\text{EBLUP}} = \tilde{\theta}_{iR}^{\text{EBLUP}}(\hat{\boldsymbol{\delta}}) = \boldsymbol{x}_{iR}^T \tilde{\boldsymbol{\beta}} + \left(\hat{\sigma}_\psi^2 \mathbf{1}_R + \hat{\sigma}^2 \boldsymbol{\lambda}_R\right)^T \boldsymbol{V}_i^{-1} (\hat{\boldsymbol{\theta}}_i - \boldsymbol{X}_i \tilde{\boldsymbol{\beta}})$$

$$= \omega^*_{iR}\, \hat{\theta}_{iR} + (1 - \omega^*_{iR})\, \boldsymbol{x}_{iR}^T \tilde{\boldsymbol{\beta}} + \textstyle\sum_{t=1}^{R-1} \omega^*_{it} (\hat{\theta}_{it} - \boldsymbol{x}_{it}^T \tilde{\boldsymbol{\beta}}),$$

where

- $\boldsymbol{\lambda}_R{}^T$ is the R-th row of $\boldsymbol{\Lambda}$;
- $(\omega^*_{i\,1},...,\omega^*_{i\,R}) = \left(\hat{\sigma}^2_{\psi}\,\mathbf{1}_R + \hat{\sigma}^2\,\boldsymbol{\lambda_R}\right)^T\,\boldsymbol{V}_i^{-1}$; and
- $\tilde{\boldsymbol{\beta}} = \left(\sum_i \boldsymbol{X}_i^{\boldsymbol{T}}\,\boldsymbol{V}_i^{-1}\,\boldsymbol{X}_i\right)^{-1}\left(\sum_i \boldsymbol{X}_i^{\boldsymbol{T}}\,\boldsymbol{V}_i^{-1}\,\hat{\boldsymbol{\theta}}_i\right)$;

As shown above, the EBLUP estimator is a linear combination of a direct estimator $\hat{\theta}_{i\,R}$, a regression estimator $\boldsymbol{x}_{i\,R}^T\,\tilde{\boldsymbol{\beta}}$, and a weighted sum of the residuals ($\hat{\theta}_{i\,t}$ - $\boldsymbol{x}_{i\,t}^T\,\tilde{\boldsymbol{\beta}}$).

The MSE estimators for the case of a known ρ are obtained in a similar fashion from the formulae presented for the general model (7) - see Rao and Yu (1994); Datta *et al*. (2002); and You (1999, Chapter 8).

The case of an unknown ρ is more difficult to handle. The problem is to obtain a consistent and unbiased estimator of ρ that takes values within the (-1 , 1) range. Rao and Yu (1994) proposed a naive estimator of ρ that is inconsistent when sampling errors are present and is biased (typically, it underestimates ρ). Still, the EBLUP estimator will remain unbiased if the proposed naive estimator of ρ is used. The MSE estimators mse_1 and mse_2 though are not correct to terms of order $o(m^{-1})$ (m is a number of small areas).

Finally, switching to the last type of area-level models presented above, the spatial models, it can be shown that the spatial models are a special case of the linear mixed model (7). Therefore, the formulae stated above would apply to obtain an EBLUP estimator and an estimator of the MSE of the EBLUP estimator. Cressie and Chan (1989) used ML estimators of the parameter vector $\boldsymbol{\delta}$ but REML can also be applied. On the practical side, Cressie (1989) used the

EBLUP approach with spatial correlations to model the 1980 U.S. Census undercounts.

2.3 Bayesian Approach

Two other estimation methods used most often in small area model estimation are Empirical Bayes (EB) and Hierarchical Bayes (HB) methods.

The Bayesian approach uses sampling and linking models as well as prior beliefs about unknown parameters that are of interest (prior distributions) to build a small-area model. Likelihood is usually specified in the form of a probability distribution for the observed data given unknown parameters, $f(\mathbf{y}|\boldsymbol{\theta})$. Here, $\boldsymbol{\theta}$ is a vector of unknown parameters that are of interest and will be estimated through the Bayesian approach. The unknown parameters $\boldsymbol{\theta}$ are treated as random quantities and their probability distribution, given unknown hyperparameters (prior distribution), are stated, $\pi(\boldsymbol{\theta} \mid \boldsymbol{\eta})$.

Empirical Bayes analysis uses the conditional posterior distribution of $\boldsymbol{\theta}$ given hyperparameters $\boldsymbol{\eta}$

$$\pi(\boldsymbol{\theta} \mid \mathbf{y}, \boldsymbol{\eta}) = \frac{f(\boldsymbol{y} \mid \boldsymbol{\theta})\, \pi(\boldsymbol{\theta} \mid \boldsymbol{\eta})}{\int f(\boldsymbol{y} \mid \boldsymbol{u})\, \pi(\boldsymbol{u} \mid \boldsymbol{\eta})\, d\boldsymbol{u}},$$

to make inferences about $\boldsymbol{\theta}$. Observed data is utilized to derive the marginal distribution

$$m(\mathbf{y} \mid \boldsymbol{\eta}) = \int f(\boldsymbol{y} \mid \boldsymbol{u})\, \pi(\boldsymbol{u} \mid \boldsymbol{\eta})\, d\boldsymbol{u},$$

to estimate the unknown hyperparameters $\boldsymbol{\eta}$, $\hat{\boldsymbol{\eta}}(\mathbf{y})$. These estimates are then used in the posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{y}, \hat{\boldsymbol{\eta}})$ to make inferences about $\boldsymbol{\theta}$. In

essence, the EB approach simplifies computational complexity by eliminating integration over another set of unknown random quantities.

Hierarchical Bayes analysis assumes a hyperprior distribution $h(\boldsymbol{\eta})$ for hyperparameters $\boldsymbol{\eta}$ and then uses the following posterior distribution for inferences about $\boldsymbol{\theta}$

$$\pi(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{\int f(\boldsymbol{y} \mid \boldsymbol{\theta})\, \pi(\boldsymbol{\theta} \mid \boldsymbol{\eta})\, h(\boldsymbol{\eta})\, d\boldsymbol{\eta}}{\int\int f(\boldsymbol{y} \mid \boldsymbol{u})\, \pi(\boldsymbol{u} \mid \boldsymbol{\eta})\, h(\boldsymbol{\eta})\, d\boldsymbol{u}\, d\boldsymbol{\eta}}.$$

The HB approach is more straightforward and clear-cut than EB (from a theoretical standpoint). However, it requires the specification of a hyperprior distribution $h(\boldsymbol{\eta})$ and typically is more computationally intensive than EB (on the other hand, EB might require difficult numerical maximization). With the recent development of the Markov Chain Monte Carlo methods, HB methods are still computationally expensive but easily automated. Methods for prior and/or hyperprior distribution specification have also been under review and several options have been proposed (e.g., using expert opinion, specifying only a distributional family, using prior distributions with little informative content (non-informative priors), etc.).

These advancements have contributed to an increase in the popularity of HB methods which have recently been widely used and successfully applied to numerous small area estimation problems.

Let us now return to the Fay-Herriot model (1) and apply EB and HB estimation methods. The Fay-Herriot model (1) can be expressed as a Bayesian model under the assumption of the normality of random effects and sampling errors as

$$\hat{\theta}_i \mid \theta_i \overset{\text{ind}}{\sim} N\left(\theta_i,\ \sigma^2_{e_i}\right),\ \ i = 1,...,m,$$

$$\theta_i \overset{\text{ind}}{\sim} N\left(\boldsymbol{x}_i^T \boldsymbol{\beta},\ b_i{}^2\, \sigma^2_{\psi}\right),\ \ i = 1,...,m.$$

Then the conditional posterior distribution of $\theta_i \mid \hat{\theta}_i, \boldsymbol{\beta}, \sigma^2_{\psi}$ is given by

$$\theta_i \mid \hat{\theta}_i, \boldsymbol{\beta}, \sigma^2_{\psi} \ \sim N\left(\hat{\theta}_i^B,\ \kappa_i\, \sigma^2_{e_i}\right),$$

$$\text{where } \hat{\theta}_i^B = \kappa_i\, \hat{\theta}_i + (1 - \kappa_i)\ \boldsymbol{x}_i^T \boldsymbol{\beta},$$

$$\text{and } \kappa_i = \frac{b_i{}^2\, \sigma^2_{\psi}}{\sigma^2_{e_i} + \sigma^2_{\psi}\, b_i{}^2}.$$

Under squared-error loss, a Bayes solution is the mean of the posterior distribution of $\theta_i \mid \hat{\theta}_i, \boldsymbol{\beta}, \sigma^2_v$, and it takes the following form

$$\hat{\theta}_i^B = \kappa_i\, \hat{\theta}_i + (1 - \kappa_i)\ \boldsymbol{x}_i^T \boldsymbol{\beta}.$$

The Bayes solution $\hat{\theta}_i^B$ depends on unknown parameters $\boldsymbol{\beta}$ and σ^2_{ψ}. Under the EB approach, these parameters are estimated from the marginal distribution $\hat{\theta}_i \overset{\text{ind}}{\sim} N\left(\boldsymbol{x}_i^T \boldsymbol{\beta},\ b_i{}^2\, \sigma^2_{\psi} + \sigma^2_{e_i}\right)$ using the maximum likelihood (ML) or the restricted maximum likelihood (REML) methods. In this case, the EB estimator is given by $\hat{\theta}_i^{\text{EB}} = \hat{\kappa}_i\, \hat{\theta}_i + (1 - \hat{\kappa}_i)\ \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}$, which is identical to the EBLUP estimator given above.

Traditional survey estimation methods focus on the frequentist MSE as a measure of the precision of an estimator. The frequentist MSE of an estimator $\hat{\theta}^B$ is defined as

$$\text{MSE}(\hat{\theta}^B) = \text{E}_{\mathbf{y} \mid \theta}\ (\theta - \hat{\theta}^B)^2$$

where the expectation is taken with respect to the distribution of data (**y**) given the parameter θ.

Given that the EB and the EBLUP estimators are identical under the normality assumptions, typically, formulae for frequentist MSE estimation of the

EBLUP estimators are used in the EB case as well. These frequentist MSE estimators are nearly unbiased (to the order of m^{-1}). A jackknife method has been proposed by Jiang *et al.* (2002) to estimate the MSE of EB estimators.

However, the Bayesian approach offers a different method to estimate the precision of Bayes estimators. Namely, it uses the posterior variance with respect to an estimator $\hat{\theta}^B$. Carlin and Louis (2000) defined the posterior variance with respect to the estimator $\hat{\theta}^B$ as $E_{\theta|\mathbf{y}}\ (\theta - \hat{\theta}^B)^2$, where the expectation is taken over the posterior distribution, $\pi\ (\theta|\mathbf{y})$. This variance is sometimes called posterior MSE and can be decomposed as

$$\mathrm{E}_{\theta|\mathbf{y}}\left(\theta - \hat{\theta}^B\right)^2 = \mathrm{Var}_{\theta|\mathbf{y}}(\theta) + \left(\mathrm{E}_{\theta|\mathbf{y}}(\theta) - \hat{\theta}^B\right)^2, \qquad (8)$$

where $\mathrm{Var}_{\theta|\mathbf{y}}(\theta)$ is the posterior variance

$$\mathrm{Var}_{\theta|\mathbf{y}}(\theta) = \mathrm{E}_{\theta|\mathbf{y}}\ (\theta - \mathrm{E}_{\theta|\mathbf{y}}(\theta))^2),$$

$\mathrm{E}_{\theta|\mathbf{y}}(\theta)$ is the posterior mean, and $\mathrm{E}_{\theta|\mathbf{y}}(\theta) - \hat{\theta}^B$ is the posterior bias of the estimator $\hat{\theta}^B$.

If squared-error loss is used, the Bayes estimator is the mean of the posterior distribution and the posterior variance with respect to $\hat{\theta}^B$ equals the posterior variance $\mathrm{Var}_{\theta|\mathbf{y}}(\theta)$ (in other words, $\mathrm{E}_{\theta|\mathbf{y}}(\theta - \hat{\theta}^B)^2$ attains the minimum, $\mathrm{E}_{\theta|\mathbf{y}}(\theta - \hat{\theta}^B)^2 = \mathrm{Var}_{\theta|\mathbf{y}}(\theta)$).

In practical applications, using squared-error loss, the posterior variance may be used as an estimator of the frequentist MSE of the Hierarchical Bayes estimator. This is based on the assumption that the frequentist bias of the posterior variance as an estimator of the frequentist MSE($\hat{\theta}^B$) is small (*cf.* Rao, 2003, Ch. 10). In the case of EB estimation, though, the estimated posterior

variance depends on estimates of variance-covariance parameters, and, hence, can lead to severe underestimation of the true posterior variance. Therefore, it has to be adjusted to correct bias.

Returning to the Fay-Herriot model and EB estimation, if the parameters σ_ψ^2 and $\boldsymbol{\beta}$ are known, the posterior distribution is completely known and the posterior variance is used to measure the precision of $\hat{\theta}_i^{\text{EB}}$. It was shown (*cf.* Rao, 2003, Ch. 9) that in the case of known parameters σ_ψ^2 and $\boldsymbol{\beta}$, EBLUP and EB estimation approaches produce identical parameter and frequentist MSE estimators.

If σ_ψ^2 and $\boldsymbol{\beta}$ are unknown, using the estimated posterior density of $\hat{\theta}_i$ (namely, $N\left(\hat{\theta}_i^{\text{EB}},\ \hat{\kappa}_i\,\sigma_{e_i}^2\right)$) and its estimated posterior variance $\hat{\kappa}_i\,\sigma_{e_i}^2$ as a measure of the precision of $\hat{\theta}_i^{\text{EB}}$ leads to severe underestimation of the frequentist MSE ($\hat{\theta}_i^{\text{EB}}$). This is due to the fact that this approach treats hyperparameters as fixed and ignores the uncertainty associated with the estimation of hyperparameters. Two methods have been proposed to overcome this issue and to provide a better approximation to the posterior variance - the bootstrap method (*cf.* Laird and Louis, 1987; Butar and Lahiri, 2001) and the method due to Kass and Steffey (*cf.* Kass and Steffey, 1989).

Turning to the HB estimation methods, the Fay-Herriot model can be specified as the following

$$\hat{\theta}_i \,\Big|\, \theta_i,\ \sigma_{e_i}^2 \overset{\text{ind}}{\sim} N\left(\theta_i,\ \sigma_{e_i}^2\right),\quad i = 1,\ldots,m,$$

$$\theta_i \,\Big|\, \boldsymbol{\beta},\ \sigma_\psi^2 \overset{\text{ind}}{\sim} N\left(\boldsymbol{x}_i^T\boldsymbol{\beta},\ b_i^2\,\sigma_\psi^2\right),\quad i = 1,\ldots,m$$

and a prior distribution on $(\boldsymbol{\beta},\ \sigma_\psi^2)$ is given by

$$f(\boldsymbol{\beta}, \sigma_\psi^2) = f(\boldsymbol{\beta})\, f(\sigma_\psi^2) \propto f(\sigma_\psi^2).$$

In this case, σ_ψ^2 and $\boldsymbol{\beta}$ are considered as random variables and additional integration has to be performed over the unknown parameters σ_ψ^2 and $\boldsymbol{\beta}$. Hence, under squared-error loss, the Bayes estimator can be derived as follows

$$\hat{\theta}_i^{\mathrm{HB}} = E(\theta_i \mid \hat{\boldsymbol{\theta}}) = E_{\sigma_\psi^2}\left(\kappa_i \hat{\theta}_i + (1 - \kappa_i)\, \boldsymbol{x}_i^T \tilde{\boldsymbol{\beta}}\right),$$

where

$$\kappa_i = \frac{b_i^2 \sigma_\psi^2}{\sigma_{e_i}^2 + \sigma_\psi^2 b_i^2}, \text{ and}$$

$$\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}\left(\sigma_\psi^2\right) = \left(\sum\nolimits_i \frac{\boldsymbol{x}_i \boldsymbol{x}_i^T}{\sigma_{e_i}^2 + \sigma_\psi^2 b_i^2}\right)^{-1} \times \sum\nolimits_i \frac{\boldsymbol{x}_i \hat{\theta}_i}{\sigma_{e_i}^2 + \sigma_\psi^2 b_i^2}.$$

The expectation is taken with respect to the posterior distribution of σ_ψ^2, $f(\sigma_\psi^2 \mid \hat{\boldsymbol{\theta}})$ (as $\boldsymbol{\beta}$ is estimated and its estimate is dependent on σ_ψ^2).

The posterior variance of θ_i (measure of precision of $\hat{\theta}_i^{\mathrm{HB}}$) can be expressed as the following

$$\mathrm{Var}(\theta_i \mid \hat{\boldsymbol{\theta}}) = \mathrm{E}_{\sigma_\psi^2}\left[\mathrm{MSE}\{\kappa_i \hat{\theta}_i + (1 - \kappa_i)\, \boldsymbol{x}_i^T \tilde{\boldsymbol{\beta}}\}\right] + \\ + \mathrm{Var}_{\sigma_\psi^2}\left[\kappa_i \hat{\theta}_i + (1 - \kappa_i)\, \boldsymbol{x}_i^T \tilde{\boldsymbol{\beta}}\right];$$

where MSE denotes frequentist MSE of an estimator (since in the case of known parameters σ_ψ^2 and $\boldsymbol{\beta}$, EBLUP and EB estimation approaches produce identical estimators (*cf.* Rao, 2003, Ch. 9)) and the variance is taken with respect to the posterior distribution of σ_ψ^2, $f\left(\sigma_\psi^2 \mid \hat{\boldsymbol{\theta}}\right)$ (as mentioned above, $\boldsymbol{\beta}$ is estimated and its estimate is dependent on σ_ψ^2).

It is easy to verify that, in this setup, when σ_ψ^2 is known and $f(\boldsymbol{\beta}) \propto 1$, HB estimators are equivalent to EBLUP estimators.

In practice, it is often impossible to derive a closed-form solution to HB equations and the only way to proceed is through numerical computations. To

that end, Markov Chain Monte Carlo (MCMC) computational methods have been developed and are extensively used in HB estimation. In the next section, a brief introduction to MCMC methods will be presented.

2.4 Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) methods comprise a variety of computational algorithms intended for the iterative simulation of joint posterior distributions in Bayesian models (*cf.* Gilks *et al*., 1996). Algorithms are generic and can be applied to a variety of different models including very complex spatial and temporal Bayesian models. In fact, when studying complex models, it is often the only method available to sample from the posterior distribution in order to obtain Bayes estimates.

In essence, the MCMC methods could be illustrated by an example of a Markov chain constructed by Tanner and Wong (1987) (data augmentation algorithm). Let us assume that we are interested in a posterior distribution of X = (*z*, *y*). The marginal posterior densities of *z* and *y* could be derived as

$$p(z) = \int p(z \mid y)\, p(y)\, d\, y \quad \text{and} \quad p(y) = \int p(y \mid z)\, p(z)\, d\, z;$$

where $p(z \mid y)$ and $p(y \mid z)$ are known and we need to solve for $p(z)$.

Using a sampling-based approach, we can start with $p_0(z)$, an estimate of $p(z)$, and draw $Z^{(0)} \sim p_0(z)$. Then we would draw $Y^{(1)} \sim p(y \mid z^{(0)})$. The marginal distribution of $Y^{(1)}$ is $p_1(y) = \int p(y \mid z)\, p_0(z)\, d\, z$. Next we draw $Z^{(1)} \sim p(z \mid y^{(1)})$ with the marginal distribution

$$p_1(z) = \int p(z \mid y)\, p_1(y)\, d\, y$$

$$= \int p(z \mid y) \int p(y \mid z') \, p_0(z') \, dz' \, dy$$

$$= \int \left(\int p(z \mid y) \, p(y \mid z') \, dy \right) p_0(z') \, dz'$$

$$= \int h(z, z') p_0(z') \, dz' = I_h(p_0(z)).$$

Tanner and Wong (1987) showed that if we repeat the drawing process and produce a large enough number of pairs $(Z^{(i)}, Y^{(i)})$, the sequence $\{Z^{(i)}, Y^{(i)}\}$ will converge to *p(z,y)* (joint distribution) from which marginal distributions *p(z)* and *p(y)* could be obtained. In other words, for sufficiently large *i*, $\{Z^{(i)}\}$ could be considered to be a sample from the marginal density *p(z)*, and $\{Y^{(i)}\}$ could be considered to be a sample from the marginal density *p(y)*. The algorithm is iterative and requires a burn-in period before the produced samples can be assumed to come from the marginal densities.

Once the MCMC output for a model is generated (i.e. samples from the joint posterior distribution), posterior quantities of interest can be computed noting that samples obtained for each component of the vector of parameters can be considered as samples from the corresponding marginal posterior distributions.

The easiest way to obtain an estimate of a function of a particular parameter ($\phi = f(\theta_i)$, for example) is to use "ergodic averages", that is to consider

$$\hat{\phi} = \hat{f}(\theta_i) = \frac{1}{D} \sum_{k=d+1}^{d+D} \phi^{(k)},$$

where *d* is the length of a burn-in period that is used to achieve convergence (and hence these first *d* values are disregarded for the calculation purposes), D is the size of the retained sample and $\phi^{(k)} = f\left(\theta_i^{(k)}\right)$ with $\theta_i^{(k)}$ being the *k*-th draw from the MCMC sample for θ_i.

Turning to the estimation of the precision of estimators, the posterior variance of ϕ can be estimated as

$$\hat{\mathrm{V}}\mathrm{ar}(\phi \mid y) = \frac{1}{D-1} \sum_{k=d+1}^{d+D} \left(\phi^{(k)} - \hat{\phi}\right)^2.$$

If the mathematical form for the conditional expectation of ϕ given the data **y** and all the model parameters (let's denote the model parameter vector as $\boldsymbol{\xi}$) is known, we can use an improved Rao-Blackwellized estimator (*cf.* Gelfand and Smith, 1991)

$$\hat{\phi}^{\mathrm{RB}} = \frac{1}{\mathrm{D}} \sum_{\mathrm{k=d+1}}^{\mathrm{d+D}} \mathrm{E}\left(\phi \,\middle|\, \mathbf{y}, \boldsymbol{\xi}^{(\mathrm{k})}\right),$$

where $\{\boldsymbol{\xi}^{(k)}\}$ form an iid sample from the marginal posterior distribution of $\boldsymbol{\xi}|\mathbf{y}$ (it is assumed that the sample was obtained in such a manner that it could be considered quasi-iid for practical estimation purposes - see Section 2.4.3 for more details).

Then, the posterior variance can be estimated as

$$\begin{aligned}\tilde{\mathrm{V}}\mathrm{ar}(\phi \mid \boldsymbol{y}) = {} & \frac{1}{D} \sum_{k=d+1}^{d+D} \mathrm{Var}(\phi \mid \boldsymbol{y}, \boldsymbol{\xi}^{(k)}) \\ & + \frac{1}{D-1} \sum_{k=d+1}^{d+D} \left(\mathrm{E}\left(\phi \,\middle|\, \boldsymbol{y}, \boldsymbol{\xi}^{(k)}\right) - \hat{\phi}^{\mathrm{RB}}\right)^2.\end{aligned}$$

This formula will be valid if the generated sample values are approximately i.i.d. Therefore, it is recommended to use several parallel runs to generate the sample as well as to "thin" the generated samples to retain every 10th or 20th value only to ensure that this condition is met approximatively.

In practical applications, often, several parallel chains with overdispersed starting points are run (this approach will be discussed in greater length in Section 2.5.3. MCMC Implementation Issues). Let's assume that L parallel chains were run. Then, the ergodic formulae could be modified as follows

$$\hat{\phi} = \frac{1}{L\,D} \sum_{l=1}^{L} \sum_{k_l=d+1}^{d+D} \phi^{(k_l)} , \tag{9}$$

where d is the length of a burn-in period and D is the size of the retained sample (d and D are assumed to be the same for all chains); and $\phi^{(k_l)} = f\left(\theta_i{}^{(k_l)}\right)$ with $\theta_i{}^{(k_l)}$ being the k_l-th draw from the MCMC sample for θ_i from the l-th chain.

The posterior variance of ϕ can be estimated as

$$\hat{\mathrm{Var}}(\phi \mid y) = \frac{D-1}{D} W + \frac{1}{D} B , \tag{10}$$

where

- $\overline{\phi_l} = \frac{1}{D} \sum_{k_l=d+1}^{d+D} \phi^{(k_l)}$ – the average values in chain l;
- $\overline{\phi} = \frac{1}{L} \sum_{l=1}^{L} \overline{\phi_l}$ – the global average across all chains;
- $W = \frac{1}{L\,(D-1)} \sum_{l=1}^{L} \sum_{k_l=d+1}^{d+D} \left(\phi^{(k_l)} - \overline{\phi_l}\right)^2$ - within-chain variance;
- $B = \frac{D}{L-1} \sum_{l=1}^{L} \left(\overline{\phi_l} - \overline{\phi}\right)^2$ - between-chain variance.

As mentioned above, there are several different computational algorithms proposed within the framework of MCMC methods. The two most popular algorithms are the Gibbs sampling algorithm and the Metropolis-Hastings (M-H) algorithm. We will discuss these algorithms in the next two sections.

2.4.1 Gibbs Sampler

The Gibbs sampler is a MCMC algorithm with the transition kernel being constructed by the full conditional distributions. Let's assume that the distribution of interest is denoted as $\pi(\theta_1,\theta_2,...,\theta_m)$. The Gibbs sampler involves the following steps (*cf*. Geman and Geman, 1984; Gelfand and Smith, 1990)

1. Set initial values $\theta^{(0)} = (\theta_1^{(0)},...,\theta_m^{(0)})^T$.

2. Generate a new value $\theta^{(j)} = (\theta_1^{(j)},...,\theta_m^{(j)})^T$ as the following:

$\theta_1^{(j)} \sim \pi\left(\theta_1 \mid \theta_2^{(j-1)},...,\theta_m^{(j-1)}\right)$,

$\theta_2^{(j)} \sim \pi\left(\theta_2 \mid \theta_1^{(j)}, \theta_3^{(j-1)},...,\theta_m^{(j-1)}\right)$,

...,

$\theta_m^{(j)} \sim \pi\left(\theta_m \mid \theta_1^{(j)},...,\theta_{m-1}^{(j)}\right)$.

3. Change counter from j to j+1 and continue with step 2 until the convergence is reached. Once the convergence is reached, the resulting draws constitute a sample from the distribution $\pi(\theta_1,\theta_2,...,\theta_m)$. The marginal density $\pi(\theta_i)$ is approximated by the histogram of sampled values of θ_i.

The Gibbs sampler is applicable under the condition that the full conditional distributions have a standard and closed form. Then the samples can be generated directly from these distributions. If the conditional distributions do not have a closed form or are not standard, different methods (such as the Metropolis-Hastings algorithm) have to be used.

2.4.2 Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm (M-H) was introduced by Metropolis *et al*. (1953) and Hastings (1970). Let's assume that $p(u)$ is the true joint posterior distribution for a parameter U that is required to be estimated. First, one has to choose a symmetric function ($q(u,v) = q(v,u)$ for all u and v) and such that $q(\cdot, u)$ is a pdf for all u. $q(u,v)$ is called a candidate (proposal) density. The M-H algorithm can be summarized as the following

1. Draw v $\sim q(\cdot, U^{(t-1)})$, where $U^{(t-1)}$ is the current state of the Markov chain.

2. Calculate $r = p(v) / p(u)$.

3. Set $U^{(t)}$ as the following:

$$U^{(t)} = v, \text{ if } r \geq 1$$

$$= v \text{ with probability } r, \text{ if } r < 1$$

$$= u \text{ with probability } 1 - r, \text{ if } r < 1$$

Tierney (1994) has shown that $U^{(t)} \xrightarrow{d} U \sim p(u)$.

In the case when conditional distributions for parameters of interest θ_i's are available in a closed form for most of the θ_i's, the Gibbs sampler can be combined with a Metropolis subalgorithm. The Gibbs sampler would be executed for all parameters that have conditional distributions in a closed form. Once a parameter (θ_i , for example) that does not have a conditional distribution in a closed form is reached, a Metropolis subalgorithm would be run for T iterations and then the end state would be taken as the sample value $\theta_i^{(t)}$ to be used

in the Gibbs sampler for the rest of the parameters. This algorithm is referred to as Metropolis within Gibbs.

2.4.3 MCMC Implementation Issues

Even though the algorithms' convergence has been proven, practical implementation of MCMC algorithms may be quite challenging depending on the complexity of a model. Two major aspects that have to be monitored and analyzed during the implementation of MCMC algorithms are MCMC chain convergence and model adequacy.

Convergence of a MCMC algorithm can be defined as a point in a sample-generation process at which the output (sample) can be safely taken to truly represent the targeted posterior distribution. Convergence can be monitored through a variety of different tools and can be affected by many factors.

One of the most common problems affecting convergence of the Markov chains is overparametrization. Often, models have so many parameters that they can not be identified from the available data (in other words, typically, only a linear combination of parameters can be estimated but not separate parameters). Another negative aspect of overparametrization is the high correlation between parameters. Therefore, it is important to ensure that the parameters included in the model are necessary and can be identified from the data.

Another important convergence problem is how to define the length of a "burn-in" period. A "burn-in" period is required to ensure that the chain achieves convergence. Length of a "burn-in" period is affected by the chain starting point and chain convergence rate. Observations generated during a

"burn-in" period are disregarded. There is no clear-cut solution to specify the length of a "burn-in" period for a particular problem. Instead, it is recommended to monitor convergence and, depending on the convergence rate, determine a cut off point individually for each model.

Another factor affecting the convergence of an MCMC chain is autocorrelation. Given the nature of MCMC methods, a certain degree of autocorrelation between generated values is expected, especially, between consecutively generated sample values. However, autocorrelation can remain significant even with an increase in lag between two sample values.

Several approaches have been proposed to overcome the issue of autocorrelation within a MCMC chain. Gelfand and Smith (1990) have suggested running *n* parallel chains until convergence and taking *m* values from each chain once convergence is reached. Using chains initialized independently would allow for samples to be independent. Using over-dispersed starting points would also allow the entire posterior space to be covered.

Raftery and Lewis (1992) have suggested to "thin" a sample generated by a single chain by retaining every *k*-th observation only (after a "burn-in" period is finished and a chain has converged).

Gelman and Rubin (1992a) have recommended using a hybrid approach where a small number of parallel chains are run and the generated samples from each chain are "thinned" to retain every *k*-th observation only. Chains can be initiated independently and the initial values for each chain can be dispersed to cover the entire posterior space. Using every *k*-th observation

would allow for reduction in the autocorrelation within each chain. However, the hybrid approach can be computationally intensive and wasteful as each chain will have a burn-in period and many generated values will be discarded even after convergence is reached. Therefore, one may consider using a single chain if the convergence characteristics of the chain are well understood and there is no danger of missing secondary modes. Otherwise, using the hybrid approach would help safeguard against these issues.

Turning to convergence diagnostics, the proposed diagnostics can be categorized into two main groups. The first group includes more theoretical diagnostics that focus on the total variation distance between the limiting distribution and the distribution of the chain at iteration j. For more details about this approach, see Meyn and Tweedie (1994), Polson (1996), Roberts and Polson (1994), Roberts and Tweedie (1994) and Rosenthal (1993).

The second group includes diagnostics that use the data generated by a MCMC chain to come up with measures of convergence. The major drawback of this approach is that it does not guarantee convergence, rather, it simply summarizes the observed data. On the other hand, though, it is very practical and has been used successfully in many applications. We will describe several diagnostics within this group that are the most popular amongst practitioners.

The most popular method to assess convergence is due to Gelman and Rubin (1992b). Suppose we run m parallel chains with different starting points. The chains are run for $2N$ iterations each. The quantity used to monitor conver-

gence, denoted as R, is called a scale reduction factor and is calculated in the following way

$$R = \frac{V}{W} \ ,$$

where

- $V = \frac{N-1}{N} W + \frac{1}{N} B$;
- W is within-run variance (as specified in (10));
- B is between-run variance (as specified in (10)) for the last N observations in the chains.

If convergence is reached, the scale reduction factor will be close to 1.

Calculating and plotting a scale reduction factor is a very simple visual tool which is applicable to output from any MCMC chain. However, it has several drawbacks. The scale reduction factor has to be calculated separately for every parameter in the joint posterior distribution. Secondly, it requires the initial values to be over-dispersed, which is hard to verify in most practical applications.

Geyer (1992) and Raftery and Lewis (1992) suggested different convergence diagnostics. Geyer's approach focuses on the variance of values within a single chain while Raftery and Lewis concentrated on "thinning" the chain and assessing both bias and variance within the "thinned" chain. Both methods have their own drawbacks.

While a variety of convergence diagnostics have been proposed, Cowles and Carlin (1996) showed that all of them can fail to detect particular convergence issues. Therefore, in practice, the Gelman and Rubin method (scale reduction

factor) is often complemented with informal visual convergence techniques rather than with more sophisticated computational methods. One can run *n* parallel chains and plot a histogram of every *k*-th value for each chain separately. If the histograms are virtually indistinguishable, the chains are assumed to have reached convergence. Another visual technique requires monitoring a trajectory of a chain or of ergodic averages to visually inspect convergence and stationary behavior. Finally, a plot of autocorrelations with different lags can indicate possible convergence issues.

2.4.4 Model Validation And Comparison

Once convergence is established and a sample from the posterior distribution is generated, the next step is to analyze model adequacy and compare different models in the case of multiple competing models.

Often, reviewing residuals is considered to be the first step in model adequacy assessment. If there are two sets of data, a set $\mathbf{z} = (z_1,...,z_m)$ used to fit the model and a set $\mathbf{y} = (y_1,...,y_n)$ used as a validation sample, Bayesian residuals can be defined as $r_i = y_i - E(Y_i \mid \mathbf{z})$, $i = 1,...,n$. The expectation is taken with respect to the predictive distribution that assesses likelihood of each Y_i given **z**, the dataset used to build the model.

Following the usual approach to model adequacy assessment, Bayesian residuals

$$r_i = y_i - \mathrm{E}(Y_i \mid \mathbf{z})$$

can be plotted against the fitted values to detect any issues with the assumptions of normality and homogeneity of variance. Bayesian residuals can be

plotted against time as well to detect any deviations from the assumption of independence. Furthermore, the sum of standardized Bayesian residuals

$$d_i = \frac{y_i - \mathrm{E}(Y_i \mid \boldsymbol{z})}{\sqrt{\mathrm{Var}\,(Y_i \mid \boldsymbol{z})}}$$

can be used as an overall measure of fit. The variance and expectation are taken with respect to the predictive distribution described above.

If all data is used to fit the model and no validation sample is available, cross-validation residuals can be considered (*cf.* Gelfand *et al.*, 1992; Gelfand, 1996)

- $r_i' = y_i$ - $\mathrm{E}(Y_i \mid \boldsymbol{y}_{(-i)})$, where $\boldsymbol{y}_{(-i)}$ is the vector of all data except y_i, and
- $d_i' = \frac{y_i - \mathrm{E}(Y_i \mid \boldsymbol{y}_{(-i)})}{\sqrt{\mathrm{Var}\,(Y_i \mid \boldsymbol{y}_{(-i)})}}$ (standardized cross-validation residuals).

Here, mean and variance are computed with respect to the conditional predictive distribution which assesses the likelihood of each y_i given the rest of the data

$$f(y_i \mid \boldsymbol{y}_{(-i)}) = \frac{f(\boldsymbol{y})}{f(\boldsymbol{y}_{(-i)})} = \int f(y_i \mid \boldsymbol{\theta}, \boldsymbol{y}_{(-i)})\, p(\boldsymbol{\theta} \mid \boldsymbol{y}_{(-i)})\, d\boldsymbol{\theta}.$$

The density $f(y_i \mid \boldsymbol{y}_{(-i)})$ is also known as the conditional predictive ordinate (CPO) and can be used to detect outliers (when plotted vs. i) or to compare different models (larger values would indicate a better fit to the observed data). It has to be noted that this approach is computationally expensive as it requires the posterior to be re-evaluated for each i ($\boldsymbol{y}_{(-i)}$).

Another way to assess model fit is through the posterior predictive densities. The goal is to assess the degree of departure of the observed data from the assumed model (both likelihood and prior distributions) through some kind of a discrepancy measure $\mathbb{D}(\mathbf{y}, \boldsymbol{\theta})$. The discrepancy measure should be relevant to

the model and estimation objectives. For example, if the model fit in the lower tail of the distribution is the focus of the analysis, then the discrepancy measure can be chosen as $\mathbb{D}(\mathbf{y}, \boldsymbol{\theta}) = y_{\min}$. If an overall model fit is in question, then the usual measure of goodness-of-fit can be chosen

$$\mathbb{D}(\mathbf{y}, \boldsymbol{\theta}) = \sum_i \frac{[y_i - \mathrm{E}(Y_i \mid \boldsymbol{\theta})]^2}{\mathrm{Var}\,(Y_i \mid \boldsymbol{\theta})}.$$

Once a discrepancy measure is chosen and a sample from the posterior distribution of $\boldsymbol{\theta}$ is obtained, the distribution of $\mathbb{D}(\mathbf{y}, \boldsymbol{\theta})$ for the observed data can be compared with the distribution of $\mathbb{D}(\boldsymbol{y}^*, \boldsymbol{\theta})$ for future observations $\boldsymbol{y}^*$. Gelman and Meng (1996) proposed posterior predictive p-values as a summary comparative measure

$$\mathbb{p}_D = Prob\,\{\mathbb{D}(\boldsymbol{y}^*, \boldsymbol{\theta}) \geq \mathbb{D}(\boldsymbol{y}, \boldsymbol{\theta}) \mid \boldsymbol{y}\}$$

$$= \int Prob\,\{\mathbb{D}(\boldsymbol{y}^*, \boldsymbol{\theta}) \geq \mathbb{D}(\boldsymbol{y}, \boldsymbol{\theta}) \mid \boldsymbol{y}\}\; p(\boldsymbol{\theta} \mid \boldsymbol{y})\, d\boldsymbol{\theta}. \qquad (11)$$

If the model fit is adequate, then $\mathbb{D}(\boldsymbol{y}^*, \boldsymbol{\theta})$ and $\mathbb{D}(\mathbf{y}, \boldsymbol{\theta})$ should be close and $\mathbb{p}_D$ should be close to 0.5. Extreme values of $\mathbb{p}_D$ are considered to be an indication of poor fit. Posterior predictive p-values have an important drawback though. It uses the data twice, once to compute $\mathbb{D}(\mathbf{y}, \boldsymbol{\theta})$ and a second time to calculate $\mathbb{p}_D$. Several improvements have been suggested to counteract this issue, but the resulting comparative measures are much more difficult to implement and hence, they are not typically used in practical applications.

Turning to model comparison and selection, Laud and Ibrahim (1995) offered a way to compare different competing models based on the predictive densities

$$f(\boldsymbol{y}_{\text{new}} \mid \boldsymbol{y}_{\text{obs}}) = \int f(\boldsymbol{y}_{\text{new}} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \boldsymbol{y}_{\text{obs}})\, d\boldsymbol{\theta},$$

where $\boldsymbol{y}_{\text{new}}$ is a replicate of the observed data. First, a discrepancy measure $d(\boldsymbol{y}_{\text{new}}, \boldsymbol{y}_{\text{obs}})$ must be selected. A discrepancy measure typically depends on the distribution used in the models. Then, $E(d(\boldsymbol{y}_{\text{new}}, \boldsymbol{y}_{\text{obs}}) \mid \boldsymbol{y}_{\text{obs}}, \text{Model}_i)$ is calculated and the model with the smallest expectation is selected. Here, expectation is taken with respect to the predictive density for a given Model *i*. However, this measure also makes double-use of the observed data.

Spiegelhalter *et al*. (1998) offered another way to compare competing models. It is based on the posterior distribution of the deviance statistic defined as

$$\text{D}(\boldsymbol{\theta}) = -2 \log f(\mathbf{y} \mid \boldsymbol{\theta}) + 2 \log h(\mathbf{y}).$$

Here, $f(\mathbf{y} \mid \boldsymbol{\theta})$ is the likelihood function of the observed data given $\boldsymbol{\theta}$, and $h(\mathbf{y})$ is a function of the data alone and is chosen arbitrarily. Spiegelhalter *et al.* (1998) suggested the criterion called the Deviance Information Criterion (DIC) and defined as

$$\text{DIC} = E_{\theta \mid y}(D) + p_D = 2\overline{D} - D(\overline{\boldsymbol{\theta}}). \qquad (12)$$

Here, the fit of a model ("goodness-of-fit") is expressed by the posterior expectation of $\text{D}(\boldsymbol{\theta})$, $\overline{D} = E_{\theta \mid y}(\text{D})$. The complexity of a model (penalty for increasing a number of parameters) is determined by the effective number of parameters $p_D = E_{\theta \mid y}(\text{D}) - \text{D}(E_{\theta \mid y}(\boldsymbol{\theta}))$ = $\overline{D}$ - $\text{D}(\overline{\boldsymbol{\theta}})$. Following the same logic as the Akaike information criterion (DIC is a generalization of the Akaike information criterion), smaller values of DIC indicate a better model fit. DIC is not intended to determine the "correct" model, rather, it is widely used to compare

several models because differences in DIC values across models are more meaningful.

We will conclude this brief description of the MCMC implementation methods and issues by reviewing one last method used to compare different models, namely, Bayes factor comparison. A Bayes factor for comparison of two models, M_1 and M_2, is defined as the following

$$\text{BF} = \frac{p(\boldsymbol{y} \mid M_1)}{p(\boldsymbol{y} \mid M_2)},$$

where $p(\boldsymbol{y} \mid M_i) = \int f(\boldsymbol{y} \mid \theta_i, M_i)\, \pi_i(\theta_i)\, d\theta_i$. BF corresponds to the ratio of observed marginal densities for the two models.

It can be shown that

$$\text{BF} = \frac{f(M_1 \mid \boldsymbol{y})}{f(M_2 \mid \boldsymbol{y})} \Big/ \frac{f(M_1)}{f(M_2)},$$

where:

- $\frac{f(M_1 \mid \boldsymbol{y})}{f(M_2 \mid \boldsymbol{y})}$ - a ratio of posterior odds for model M_1 versus model M_2 ;
- $\frac{f(M_1)}{f(M_2)}$ - a ratio of prior odds for model M_1 versus model M_2 .

Hence, the Bayes factor can also be interpreted as a ratio of the posterior odds for model M_1 versus model M_2 to the prior odds for model M_1 versus model M_2. In other words, the Bayes factor represents a change in odds in favor of Model 1 once a prior distribution is transformed into a posterior distribution.

Kass and Raftery (1995) provide a detailed review of Bayes factors including various computational methods. They also suggest that a Bayes factor greater than 20 suggests strong evidence against M_2, a Bayes factor between 3

and 20 suggests substantial evidence against M_2, while a Bayes factor less than 3 suggests no difference at all.

Overall, Bayes factors used to be the most popular method to compare different models. Lately, serious criticism has been brought up against the use of Bayes factors (*cf.* Gelfand, 1996) mainly due to the fact that $p(\boldsymbol{y} \mid M_1)$ may be improper in the case of improper priors (non-informative priors, for example).

This concludes an introduction into the small area models and most popular model-based approaches to the derivation of small area estimators. We will now proceed with a description of the problem studied within the scope of this thesis.

Chapter 3

Case Study: Hierarchical Bayes Spatial and Generalized Linear Mixed Models

As mentioned in Chapter 1, the first research objective within the scope of this thesis is to apply HB spatial estimation models and methods as well as a specific generalized linear mixed model to the estimation of a particular variable in small socio-geographic domains. The resulting estimates will then be compared with the results produced by the more traditional Fay-Herriot model.

In this Chapter, we will first describe the dataset chosen for analysis in more details and then propose four HB small area estimation models that will be used to achieve the research objective stated above.

3.1 CCHS Dataset

The Canadian Community Health Survey (CCHS) is a cross-sectional survey that collects information related to health status, health care utilization and health determinants for the Canadian population. For more details, we refer the reader to the CCHS User Guide for the Public Use Microdata File, 2005. CCHS was designed to collect data at sub-provincial levels of geography (Health Regions) and to allow the aggregation of estimates at the provincial and national levels. As a key component of the Population Health Survey Program of Statistics Canada, CCHS provides support for the development of public policy. CCHS data is very useful for analytic studies because the survey collects economic, social, demographic, occupational and environmental variables related to the health issues being analyzed, thus, presenting a wide variety of covariates to choose from.

Until 2008, CCHS operated on a two-year cycle. In the first year (cycle .1), the sample size was large enough to allow estimation of various variables at the subprovincial level. In the second year (cycle .2), a smaller sample size was used to collect provincial data on specific focused health topics.

CCHS 2.1 was collected between January 2003 and December 2003, for 126 Health Regions that included all provinces and territories. The targeted population was individuals aged 12 or older, living in private occupied dwellings. Individuals living on Indian Reserves and on Crown Lands, institutional residents, full-time members of the Canadian Armed Forces, and residents of

certain remote regions were excluded from the sampling frame. The estimated frame coverage was approximately 98% of the target population.

Health authorities within each province use sub-provincial units called Health Regions (HRs) for administrative purposes. Statistics Canada, in consultation with local authorities, slightly modified boundaries of Health Regions to better correspond with the 2001 Census geography. As a result, CCHS used 123 Health Regions in 10 Canadian provinces. Also, each territory was designated as a single Health Region. Table 1 provides a summary breakdown of the Health Regions between different provinces and, in Figure 1, we display a map of considered Health Regions.

CCHS 2.1 collected data in all 126 Health regions. First, a list of households was sampled, and once a sampled household was contacted and a demographic roster of the household was completed, the survey was designed to choose one member of the household to complete the survey. Selection of individual respondents was designed to over-represent the population between the ages of 12 and 19 (youths).

A three-step sample allocation strategy was used for 10 Canadian provinces to derive a household-level sample. The objective was to give relatively equal importance to the HRs and the provinces to allow sub-provincial estimates. The first two steps were designed to allocate the sample at the provincial level based on their respective populations and the number of HRs they contained. In the last step, the provincial sample was allocated among its HRs proportionally to the square root of the estimated population in each HR (*cf.* CCHS User

Guide for the Public Use Microdata File, 2005). The targeted sample size was around 130,000 households and was allocated through this scheme. Then, sample sizes in each HR were increased to account for out-of-scope and anticipated non-response.

The three territories had their own allocation strategy that resulted in 850 sample units in Yukon; 900 in the Northwest Territory and 700 in Nunavut.

Three Health Regions in Quebec supplied extra funds to increase sample sizes and provide better quality of estimates for their regions. As well, a special study was conducted as a part of CCHS 2.1 collection to compare personal and telephone modes of interviewing and assess any potential bias due to the data collection mode. This resulted in further sample increases in different health regions.

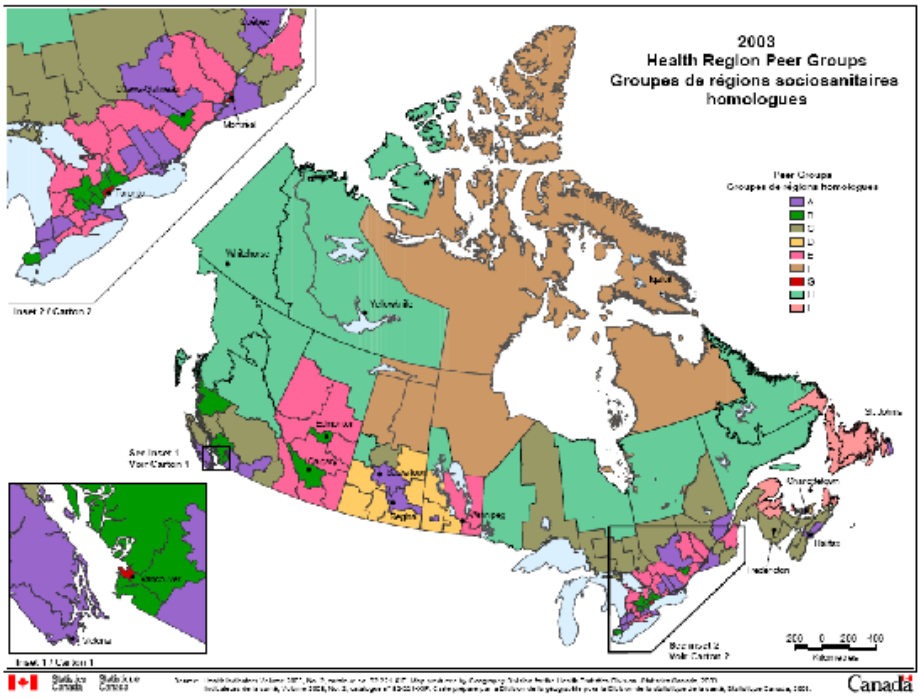
Finally, after removing the out-of-scope units, 166,222 households were chosen and a response was obtained for 144,836 of these households resulting in an overall household-level response rate of 87.1%.

Three frames were used to generate the sample - 48% of the sample of households came from an area frame, 50% came from a list frame of telephone numbers and the remaining 2% came from a Random Digit Dialling (RDD) sampling frame. Duplicates were identified during collection and every case was treated individually.

Table 1. Health Regions

Province or Territory	*Number of Health Regions*
Newfoundland and Labrador	6
Prince Edward Island	4
Nova Scotia	6
New Brunswick	7
Quebec	17
Ontario	37
Manitoba	10
Saskatchewan	11
Alberta	9
British Columbia	16
Yukon	1
Northwest Territories	1
Nunavut	1
TOTAL	126

Figure 1: Map of Health Regions.



Source: Health Indicators Volume 2003, No.2, catalogue no. 82-221-XE, Map produced by Geography Division for the Health Statistics Division, Statistics Canada, 2003.

CCHS used the Canadian Labour Force Survey (LFS) frame as a sampling area frame. LFS uses a multistage stratified cluster design with homogeneous strata formed at the first stage and independent samples of clusters drawn from each stratum (random sampling with probability proportional to size (PPS), the size being the number of households). Dwelling lists are prepared for each cluster in the second stage and dwellings, or households, are selected from the lists (systematic sampling). CCHS made modifications to the LFS sampling plan to derive targeted sample size in each HR and to account for boundary differences between HRs and geographic units used in LFS.

For the telephone frame, The Canada Phone directory, a commercially available CD-ROM, was linked to internal administrative conversion files to map telephone numbers to postal codes to HRs. There was one list frame stratum per HR. Within each stratum the required number of telephone numbers was selected using a simple random sampling.

In five HRs, a Random Digit Dialling (RDD) sampling frame of telephone numbers was used. Elimination of Non-Working Banks (ENWB) method was used to sample households from the RDD frame.

CCHS 2.1 was administered using computer-assisted interviewing (CAI). Sample units selected from the telephone list frame were interviewed from call centres using computer-assisted telephone interviewing. Units selected from the area frame were interviewed by field interviewers using computer-assisted personal interviewing. As a rule, field interviewers were not allowed to conduct interviews over the phone with some authorized exceptions.

Once the survey information was collected and transmitted to the central processing location, the data was edited and prepared for weighting. Weights were designed to take into account the three different frames used for sampling, survey non-response, seasonal effect, and post-stratification.

Using survey weights, direct estimates for various variables were obtained and included in the Public Use Microdata File. For the purposes of analysis presented in this thesis, the variable FLUC_162 was used. This variable measured the concept of when a respondent last had a flu shot by asking the following question: "When did you have your last flu shot?" Several possible answer were suggested

- LESS THAN 1 YEAR AGO;
- 1 YEAR TO LESS THAN 2 YEARS AGO;
- 2 YEARS AGO OR MORE;
- NOT APPLICABLE;
- DON'T KNOW;
- REFUSAL;
- NOT STATED.

Survey weights were applied to the number of answers in each category and direct estimates were obtained. By applying filters to define small domains based on the age group and Health Region, direct estimates for the “Number of people who had a flu shot more than two years ago” were obtained for each of these small domains. Correspondingly, the CV (coefficient of variation) were provided for estimates within each small domain. The resulting direct esti-

mates and CVs were used as direct estimates inputs into the HB models that will be described in the next section. This approach follows definition of small area-level models described in Chapter 2 and how these models use direct estimates $\hat{\theta}_i$. See Rao (2003, Ch. 9 and Ch. 10), for more details and examples.

3.2 Considered Models

Four different HB models were considered to obtain an estimate of the "Number of people who had a flu shot more than two years ago" for small domains defined as specific age groups within a Health Region.

CCHS 2.1 Public Use Microdata file supplied data for 103 Health Regions obtained after collapsing the initial 126 Health Regions (to protect confidentiality of respondents and to meet data quality standards established by Statistics Canada). Age groups were identified to coincide with the breakdown used in 2001 Census data, i.e., the following 14 age groups were specified

- 15-19 years old;
- 20-24 years old;
- 25-29 years old;
- 30-34 years old;
- 35-39 years old;
- 40-44 years old;
- 45-49 years old;
- 50-54 years old;

- 55-59 years old;
- 60-64 years old;
- 65-69 years old;
- 70-74 years old;
- 75-79 years old;
- and over 80 years old.

Thus, in total, 103 Health Regions × 14 age groups = 1,442 small domains were defined and direct estimates for each domain were obtained from the CCHS 2.1 Public Use Microdata File.

3.2.1 Model (13)

The Fay-Herriot normal model was used as the main model to allow comparison of outcomes from the other three models and to allow assessment of the impact of using spatial structures as well as generalized linear models. Typically, in various disease mapping applications, a normal HB model has been fit to the rates (proportions) of a particular event - this approach can be thought of as transforming the count data to the continuous scale to allow a better fit to the normal distribution. Using the same logic, a Fay-Herriot normal model was applied to estimate a proportion of the "Number of people who had a flu shot more than two years ago" within each small domain. The model was formulated as the following

$$(i) \quad P_{ij} \,\big|\, \mu_{ij} \overset{\text{ind}}{\sim} N\left(\mu_{ij},\ \sigma_{ij}^2\right);\ i = 1,\ ...,\ 14;\ j = 1,\ ...,\ 103; \qquad (13)$$

$$(ii) \quad \mu_{ij} \,\big|\, \beta_1, \beta_2, \beta_3, \beta_4, \psi_j, \sigma_v^2 \overset{\text{ind}}{\sim} N\left(\sum_{k=1}^{4} \beta_k x k_{ij} + \psi_j, \sigma_v^2\right);$$

$$(iii) \quad \psi_j \,\big|\, \alpha_\psi,\ \sigma_\psi^2 \overset{\text{iid}}{\sim} N\left(\alpha_\psi,\ \sigma_\psi^2\right);$$

with the following hyper-parameters

$\beta_1 \propto \text{const},\ \beta_2 \propto \text{const},\ \beta_3 \propto \text{const}, \beta_4 \propto \text{const},\ \alpha_\psi \propto \text{const};$

$\sigma_\psi^2 \sim \text{IG}\,(a,\ b),\ (\text{inverse gamma distribution});$

$\sigma_v{}^2 \sim \text{IG}\,(c,\ d);$

where σ_{ij}^2 are estimated and treated as known, and $a,\ b, c,\ d$ are known positive constants.

In this model

- P_{ij} - proportion of people who had their flu shot more than two years ago in age group i (i = 1,...,14) and Health Region j (j = 1,...,103). Following a typical approach to HB small-area models, we used CCHS direct estimates of P_{ij} as data points for this model (*cf.* Rao, 2003, Ch. 9 and Ch. 10; Fay and Herriot, 1979; Xia *et al.*, 1997; You and Rao, 2002; etc.). P_{ij} were modeled through a Normal distribution

$$f(P_{ij}) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp\left[-\frac{1}{2\sigma_{ij}^2}(P_{ij} - \mu_{ij})^2\right].$$

- x1, x2, x3, x4 - socio-demographic covariates that will be described in more details in the next section.

- σ_{ij}^2 - sampling variances that were modeled by using the equal design effects modeling approach suggested by You (*cf.* You, 2006; You, 2008).

Specifically, the design effect for the *i*-*j*-th area may be approximately written as

$$\text{deff}_{ij} = \frac{\sigma^2_{ij}}{\mu_{ij}(1-\mu_{ij})/n_{ij}},$$

where n_{ij} is the sample size.

If we denote

$$\tau_{ij} = \text{deff}_{ij}/n_{ij} = \sigma^2_{ij}/\mu_{ij}(1-\mu_{ij});$$

then we can estimate $\hat{\tau}_{ij}$ as

$$\hat{\sigma}^2_{ij} / P_{ij}(1-P_{ij}).$$

Once we average $\hat{\tau}_{ij}$ for each geographic area (using the assumption of a common design effect), we obtain $\bar{\tau}_j$ and the smoothed sampling variance σ^2_{ij} is estimated as

$$\mu_{ij}(1-\mu_{ij})\,\bar{\tau}_j$$

and treated as known.

- ψ_j - spatial effects that were modeled through a Normal distribution. This model utilizes an exchangeable prior and the small domain random effects ψ_j are assumed to be iid normal. This is the result of the assumption of geographical homogeneity (unstructured heterogeneity), and is termed as an "exchangeable" prior (in a geographic sense) in the disease mapping literature. In this case, individual estimates are expected to get displaced towards a global mean.
- The inverse gamma distribution (IG (p, q) in our notations) was chosen to model σ^2_{ψ} and σ^2_{v}

$$f(x) = \frac{q^p \, e^{-1/(qx)}}{\Gamma(p)\, x^{p+1}} \; ; \; x > 0 \; .$$

This decision was made to follow a commonly accepted practice of using a distribution conjugate to the normal distribution used to model the parameters themselves. Note that p is required to be greater than 1 for the expectation to exist and greater than 2 for the variance to exist.

- For parameters of inverse gamma distribution, we chose $a = b = c = d = 0.001$. This choice of constants permits using conjugate priors for hyper-parameters and ensures that the prior distributions are quite flat. In other words, the prior distributions remain proper and lead to a proper posterior distribution but are non-informative enough to allow the data to speak for themselves. Using non-informative (or diffuse) priors is quite common in small area estimation and 0.001 is one of the typical values assigned to the parameters of the inverse gamma distribution (*cf.* MacNab, 2003b; Best *et al*., 2005).
- At the same time, the prior distributions for β coefficients are flat and improper. This choice was made following typical modeling practices used in small area estimation (*cf.* Rao, 2003, Ch. 9 and Ch. 10). As Rao (2003, Ch. 10) points out, if the prior distributions for all hyper-parameters were improper, the Gibbs sampler could provide reasonable estimates for a non-existing posterior (but all Gibbs conditional distributions would be proper). To avoid such a situation, typically, prior distributions for variance hyper-parameters are modeled through diffuse proper distributions.

3.2.2 Model (14)

The second HB model was formulated on the basis of the Fay-Herriot model (13) but with the introduction of spatial structure in the spatial effects

$$(i) \quad P_{ij} \,\big|\, \mu_{ij} \overset{\text{ind}}{\sim} N\left(\mu_{ij},\ \sigma^2_{ij}\right); \tag{14}$$

$$(ii) \quad \mu_{ij} \,\big|\, \beta_0, \beta_1, \psi_j, \sigma_v^2 \overset{\text{ind}}{\sim} N\left(\beta_0 + \beta_1 \text{x1}_{ij} + \psi_j, \sigma_v^2\right);$$

$$(iii) \ \boldsymbol{\psi} \,\big|\, \tau \ \propto (\tau)^{M/2} \exp\left(-\tfrac{\tau}{2} \sum_{i=1}^{M} \sum_{j<i} w_{ij} (\psi_i - \psi_j)^2\right);$$

with the following hyper-parameters

$\beta_0 \propto$ const, $\beta_1 \propto$ const;

$\sigma_v^2 \sim \text{IG}(c, d)$;

$\tau \sim \text{IG}(e, f)$;

where σ^2_{ij} are estimated and treated as known, and c, d, e, f are known positive constants.

In this model

- P_{ij} - as in model (13), proportion of people who had their flu shot more than two years ago in the age group i and Health Region j, i = 1,...,14 and j = 1,...,103
- x1 - socio-demographic covariate that will be described in more detail in the next section. It has to be noted that adding more covariates to this model led to a deterioration of the model fit to the data, hence, only one covariate was retained (unlike the first model). We also retained an intercept for this model as it proved to be statistically significant in regression analysis performed during covariate selection.

- σ_{ij}^2 were estimated and treated as known. σ_{ij}^2 were modeled by using the equal design effects modeling approach introduced by You (2006; 2008) as described for model (13).

- ψ_j - spatial effects modeled through the conditional autoregressive model (CAR) (3)-(4). Using our model notations, full conditional distributions have the following form

$$\psi_j \left| \psi_{(-j)}, \ \tau \ \sim N\left(\sum_k \frac{w_{kj}\psi_k}{n_j}, \ \frac{1}{\tau n_j} \right), \right.$$

where $\psi_{(-j)}$ denotes all spatial effects except the *j*th; the w_{kj} are weights and equal to 1 if the *k*th area is adjacent to the *j*th area and 0 otherwise; and n_j is the number of areas adjacent to the *j*th area. To better align with Win BUGS definitions, we set $\tau = \frac{1}{\lambda^2}$ while λ was used in (3)-(4).

CAR model imposes a prior structure with local dependence in neighboring areas and incorporates the geographic structure of the map through an adjacency matrix (the matrix specifying which Health Regions are adjacent), which is equivalent to the assumption of geographically structured heterogeneity. Spatial random effects are assumed to be present and are separated from the small domain random effect / sampling error, hence, it is not "exchangeable". In this case, individual estimates are expected to get displaced towards a local mean, not towards a global mean. Spatial parametrization in the form of a CAR prior was chosen as it is often used in practical applications and is versatile enough to fit a variety of different situations. For instance, see Best *et al*. (2005) and Stern and Cressie (1999).

- The inverse gamma distribution was chosen to model σ_v^2 and τ to utilize the conjugate distributions - for exactly the same reasons as in model (13).
- For parameters of inverse gamma distribution, we chose $c = d = e = f = 0.001$ for the same reasons as in model (13) - to keep the distribution of the hyper-parameters non-informative and to maintain the propriety of the posteriors.

3.2.3 Model (15)

To apply a generalized linear mixed model in the context of this problem, a Poisson-normal model was proposed. Typically, in various disease mapping applications, a Poisson model is fit to the counts of a particular event using a log-linear linking function. Thus, the third model was designed to estimate the "Number of people who had a flu shot more than two years ago" within each small domain by utilizing a Poisson-log-normal HB structure.

We attempted to use the gamma distribution for the linking function (conjugate distribution with Poisson). However, convergence was slow compared to the Poisson-log-normal model as it took almost 25,000 iterations for convergence of the Poisson-gamma model as opposed to approximately 5,000 iterations in the Poisson-log-normal case. On top of that, model fit, as measured by the predictive posterior p-values, was very poor having a p-value around 0.02.

The Poisson-log-normal model was formulated as the following

$$(i) \quad Z_{ij} \,\Big|\, \beta_1, \gamma_j \overset{\text{ind}}{\sim} \text{Poisson}\,(E_{ij}\, \exp(\beta_1 \, x1_{ij} + \gamma_j)) \; ; \qquad (15)$$

$$(ii) \; \gamma_j \sim N\,(g,\ h); \;\; \beta_1 \sim N(k,\ l); \;\; g, h, k, l - \text{const} \; ;$$

where g, h, k, l are known constants.

In this model

- $Z_{ij} = \sqrt{Y_{ij}}$ and Y_{ij} is the number of people who had a flu shot more than two years ago in the age group i and Health Region j, i = 1,...,14 and j = 1,...,103. Z_{ij} were rounded to be integers. Data transformation was implemented after reviewing residuals and observing the corresponding patterns and is similar to the Freeman-Tukey transformation that is often used to stabilize variance. Various data transformations were attempted to further improve model fit (log, squared, etc.) but the square-root transformation led to the best results in terms of model fit.

  As mentioned above, following a typical approach to HB small-area models, we used CCHS direct estimates of Z_{ij} as data points for this model (*cf.* Rao, 2003, Ch. 9 and Ch. 10; Fay and Herriot, 1979; Xia *et al.*, 1997; You and Rao, 2002; etc.).

  Z_{ij} were modeled through a Poisson distribution

$$p\,(z_{ij}) = \frac{\exp[-\mu_{ij}] \times [\mu_{ij}]^{z_{ij}}}{z_{ij}!} .$$

- E_{ij} is the anticipated number of people who had a flu shot more than two years ago in the age group i and Health Region j, i = 1,...,14 and j = 1,...,103. First, the overall proportion of people in the country who had a flu shot more than two years ago was estimated as proportion =

$\sum\sum Y_{ij} / \sum\sum N_{ij}$, where N_{ij} is the number of people in the age group i and Health Region j, as estimated from CCHS data. Then, E_{ij} was calculated as E_{ij} = proportion × N_{ij}. E_{ij}'s are typically used in disease mapping applications as modulating constants to reflect contribution of the population size in each area into the expected disease counts in that area.

- x1 - socio-demographic covariate that will be described in more details in the next section. It has to be noted that adding more covariates to this model led to a deterioration of model fit to the data, hence, only one covariate was retained.
- γ_j - spatial effects modeled through a Normal distribution. Similar to model (13), this model utilizes an exchangeable prior (small domain random effects are assumed to be iid normal) - a hypothesis of unstructured heterogeneity is made.
- For parameters of normal distribution, we chose $g = k = 0$ and $h = l = 1/0.00001$ (to keep in line with parametrization of a normal distribution assumed above; however, in WinBUGS, normal distribution is programmed with a precision parameter, rather than with variance) for the same reasons as in models (13) and (14) - to keep the prior distribution non-informative and to maintain the propriety of the posteriors.

3.2.4 Model (16)

The last model was designed to include both spatial effects and the generalized linear model used in (15)

$$(i)\ Z_{ij} \,\big|\, \beta_1, \varphi_i, \gamma_j \overset{\text{ind}}{\sim} \text{Poisson}\,(E_{ij}\ \exp(\beta_1\, \text{x1}_{ij} + \varphi_i + \gamma_j))\ ; \qquad (16)$$

$$(ii)\ \boldsymbol{\gamma} \,\big|\, \tau \propto (\tau)^{-M/2} \exp\Big(-\tfrac{1}{2\tau} \textstyle\sum_{i=1}^{M} \sum_{i<j} w_{ij}\,(\gamma_i - \gamma_j)^2\Big)\ ;$$

$$(iii)\ \varphi_i \sim N(a, b);\ \beta_1 \sim N(c, d),\ \tau \sim \text{IG}\,(e,\ f);$$

where a, b, c, d, e, f are known constants.

In this model

- Z_{ij} , E_{ij} , and x1_{ij} - are defined as in the model (15).
- γ_j are spatial effects modeled through a conditional autoregressive model described in model (14) to allow spatial correlation between the Health Regions (structured heterogeneity).
- φ_i are random age effects modeled through a normal distribution. Model (15) did not permit splitting out the age effects - once implementation was attempted, chains did not converge (non-identifiable random effects). Using the CAR prior for the spatial effects in the fourth model helped overcome the problem and age effects were separated.
- For distribution parameters, we chose $a = c = 0,\ b = d = 1/0.00001$, and $e = f = 0.001$ for the same reasons as in models (13), (14), and (15) - to keep the prior distribution non-informative and to maintain the propriety of the posteriors.

In the next section, we will describe the covariate selection process for each of the four models (13), (14), (15), and (16).

3.3 Selection of Covariates

CCHS 2.1 dataset includes the following socio-demographic variables that can be used as covariates for the four models described above: Age; Sex; Marital status; Total household income; Total personal income; Education level; Labour force occupation (3 groups); and Race. The 2001 Census population estimates for all small domains were considered as another potential covariate for the four models.

To better identify the relationship between these variables and the response variable in each model, the following analysis was performed. First, a set of various linear regressions (forward linear regression, backward linear regression, forward linear regressions with several variables treated as entered variables) was fit to the data. Secondly, principal component analysis was completed. Please refer to Appendix A for detailed results.

For models (13) and (14) (the Fay-Herriot model and its extension to include spatial structure, the response variable is the proportion of people who had flu shot two or more years ago), the following results were obtained. After running linear regressions, the following subset of potential covariates was identified

- x1 = proportion of people who are single and have never been married;
- x2 = proportion of visible minority population;
- x3 = proportion of people with highest level of education less than a high school diploma;

x4 = proportion of people with group 1 labour occupations (occupations in Management, Business, Finance, Administration, Natural and Applied Sciences, Health, Social Sciences, Education, Religion, Art, Culture and Recreation);

- x5 = proportion of people with group 3 labour occupations (occupations in Trades, Transport and Equipment Operator, occupations Unique to Primary Industry, Processing, Manufacturing and Utilities);
- x6 = proportion of people with household income of 30-50K;
- x7 = proportion of people with household income of 50-80K;
- x8 = proportion of people with household income greater than 80K;
- x9 = population estimates from 2001 Census.

A linear regression of the proportion of people who had a flu shot more than two years ago against these nine variables resulted in an R-square of 0.402 and an adjusted R-square of 0.398 (please see Appendix A).

Next, principal component analysis was performed and produced the following subset of potential covariates

- x1 = proportion of people who are single and have never been married;
- x2 = proportion of people with group 1 labour occupations (occupations in Management, Business, Finance, Administration, Natural and Applied Sciences, Health, Social Sciences, Education, Religion, Art, Culture and Recreation);
- x3 = proportion of white population;

- x4 = proportion of people with highest level of education less than a high school diploma.

These are the variables with the most contribution to the first three principal components, which cumulatively explain approximately 57% of variation (please refer to Table 2).

A linear regression against these four variables resulted in R-square of 0.369 and adjusted R-square of 0.368. All four variables were significant.

Overall, the last set of covariates appeared to be important enough and despite some loss in model fit (compare to the final model identified through linear regression), the gap between R-square and adjusted R-square went down slightly indicating a less complex model. Typically, having less variables is always preferable to save degrees of freedom. Therefore, it was decided to use these four variables as covariates for models (13) and (14).

Once models (13) and (14) were fit through MCMC methods, using all four identified covariates in model (14) led to a deterioration of DIC and p-value (please see Chapter 4). Hence, only one covariate, x1, proportion of people who are single and have never been married, was retained for this model. Model (13) was fit with all four covariates.

Table 2. Normal models - principal component analysis, extraction sums of squared loadings.

Component	*Extraction Sums of Squared Loadings*	*Variance %*	*Cumulative %*
1	7.351	30.630	30.630
2	4.167	17.364	47.993
3	2.222	9.260	57.253
4	1.606	6.691	63.945

For models (15) and (16) (the Poisson model and its extension to include spatial structure, the response variable is the number of people who had flu shot two or more years ago), the following results were obtained. It has to be noted that while the best covariates for the proportion of the people who had flu shots more than two years ago (P_{ij}) were relatively distinct in the set of all possible covariates under consideration, this was not quite the case for the number of people who had flu shots more than two years ago (Y_{ij}). After running several linear regressions (forward linear regression, backward linear regression, forward linear regression with several variables treated as entered variables), the following two subsets of potential covariates was identified (please refer to Appendix A for more details)

Subset 1

- x1 = number of people who are single and have never been married;
- x2 = number of people with some post-secondary education;
- x3 = population estimates from 2001 Census.

This set of variables had the highest R-square (0.840) while VIF's for individual variables were still below 4 for the forward regression with population estimates from 2001 Census treated as an entered variable. Addition of any

other variables caused the VIF's for some variables to increase up to 22 - indicating severe multicollinearity (*cf*. Kutner *et al*., 2004).

Subset 2

- x4 = number of people with group 1 labour occupations;
- x5 = number of people with group 2 labour occupations;
- x2 = number of people with some post-secondary education;
- x6 = number of people with highest education level of a high school diploma.

This set of variables had the highest R-square (0.850) while the VIF's for individual variables were still below 5.1 for the forward linear regression. The addition of any other variables caused the VIF's for some variables to increase to over 10 - indicating multicollinearity.

After performing the principal component analysis, two other subsets of potential covariates were identified:

Subset 3 (based on extraction sums of squared loadings, see Table 3)

- x1 = number of people who are single and have never been married;
- x2= number of people with some post-secondary education;
- x5 = number of people with group 2 labour occupations;
- x7 = number of people with personal income 50-80K;
- x8 = number of people with personal income less than 15K.

These were the variables with the most contribution to the first two principal components, which cumulatively explained approximately 75% of variation.

Table 3. Poisson models - principal component analysis, extraction sums of squared loadings.

Extraction Sums of Squared Loadings			
Component	Total	Variance %	Cumulative %
1	16.428	65.712	65.712
2	2.371	9.484	75.196
3	1.291	5.165	80.361
4	1.023	4.090	84.451

Subset 4 (based on rotation sums of squared loadings, see Table 4)

- x1 = number of people who are single and have never been married;
- x8 = number of people with personal income less than 15K;
- x3 = population estimates from 2001 Census;
- x9 = number of people highest level of education of less than high school diploma.

These were the variables with the most contribution to the first two principal components, which cumulatively explained approximately 62% of variation (please refer to Table 4). Please note that an orthogonal rotation method was applied (varimax method) in order to minimize the number of variables that have high loadings on each factor with the objective to simplify the interpretation of the factors (hence, rotation sums of squared loadings were obtained). Rotation sums of squared loadings was considered for the Poisson case as the initial principal component analysis (extraction sums of squared loadings) did not conclusively identify a few variables with the most impact on the response variable. This was unnecessary in the Normal case where the initial principal

component analysis (extraction sums of squared loadings) clearly indicated covariates with the most impact on the response variable.

Table 4. Poisson models - principal component analysis, rotation sums of squared loadings.

Rotation Sums of Squared Loadings			
Component	Total	Variance %	Cumulative %
1	10.583	42.330	42.330
2	5.138	20.553	62.883
3	4.303	17.213	80.097
4	1.089	4.354	84.451

Comparing the four subsets of potential covariates that were stated above, the following results were obtained

a) Multicollinearity severity

All four subsets had VIF values greater than 2 but subsets 1 and 2 had the lowest VIF values. Subset 1 had a maximum VIF value of 3.453; subset 2 had a maximum VIF value of 5.005; subset 3 had a maximum VIF value of 12.466; and subset 4 had a maximum VIF value of 9.32. As well, some eigenvalues and tolerances for subsets 3 and 4 were too small. Overall, it appears that subsets 1 and 2 were least impacted by multicollinearity.

b) Importance of coefficients

Subsets 1 and 2 had all regression coefficients as statistically important, which was not the case for subsets 3 and 4 (as indicated by two-tailed significance level of t-value for coefficients being greater than 0.05).

c) Explanatory power

Table 5. Comparison of subsets of covariates - explanatory power

	*SSR**	*St. error***	*R-sq*	R-sq adj.
Subset 1	13,335,237,394	1330	0.84	0.84
Subset 2	13,492,838,718	1289	0.85	0.85
Subset 3	13,357,654,029	1325	0.841	0.841
Subset 4	13,156,323,556	1377	0.829	0.828

* SSR is the regression sum of squares and the higher SSR is the better is the explanatory power of a regression.

** St.error is a standard error of an estimate produced by a regression (square root of MSE) and is compared with the st. error of the response variable (calculated as a descriptive statistics before any regressions were fit) of 3320.33. A regression that produces estimates with smaller st. error is assumed to have better explanatory power.

Conclusions (models (15) and (16))

The principal components analysis did not conclusively identify a few variables with the most impact on the response variable (in fact, first component had correlation coefficients greater than 0.8 with 8 variables, which made it difficult to interpret components as well as to meaningfully use components as covariates for the response variable). Additionally, there were multicollinearity concerns with variables identified through the principal component analysis. On the other hand, the explanatory power of subset 2 appeared to be the best and multicollinearity did not appear to be significant for this subset. Therefore, it was decided to use subset 2 as four covariates for the Poisson models

(15) and (16) with a number of people who had a flu shot more than two years ago (Y_{ij}) being a response variable.

Once models (15) and (16) were fit through MCMC methods, though, using all four identified covariates led to deterioration of DIC and posterior predictive p-values p_D (please refer to formula (11) for the definition). Only one covariate, number of people with highest level of education of a high school diploma, was retained for both models. This was a result of running models with different combinations of covariates from subset 2 through MCMC procedures and assessing which model would yield a better fit and better reduce multicollinearity.

Furthermore, a square-root transformation was applied to this variable after reviewing the residuals (residuals were plotted against fitted values) and observing the corresponding curvature patterns. Initially, during the regression-fitting stage, transformations were attempted on all potential covariates (log, square, square-root, etc.) to improve model fit. Applying a square-root transformation to the variables in subset 2, and, consequently, to the only selected variable, proved to be the most advantageous in terms of model fit.

From the perspective of notations, the covariate used for the models (15) and (16) will be denoted as x1 and it stands for the square root of the number of people with highest level of education of a high school diploma.

This concludes a description of the four models under consideration. In the next Chapter, we will present a MCMC implementation of each model and the obtained results, and will conclude with model comparisons.

Chapter 4

Case Study: MCMC implementation

4.1 MCMC Implementation

For each of the four models stated in Chapter 3, estimates were derived numerically based on the samples from the marginal posterior distributions. These samples were generated with the help of Markov Chain Monte Carlo (MCMC). Specifically, the Gibbs sampler was run for each model using the WinBUGS software (*cf.* Spiegelhalter *et al*., 1995a; Spiegelhalter *et al*., 1995b; and Spiegelhalter *et al*., 2003). Appendix B contains the WinBUGS code used for each of the four models.

There is no general consensus as to what would be a preferable sample size to be generated through MCMC computations. In some applications, the overall sample size could be as low as 1,000 to 2,000 (*cf.* Carlin and Louis, 2000; case study on spatio-temporal mapping of lung cancer rates). You and Rao (2002) used 8 chains with 500 observations retained per chain for the total of 4,000 observations. Datta *et al*. (1999) used 10 chains with 500 observa-

tions retained per chain for the total of 5,000 observations. Carlin and Louis (2000) suggested to run 3 to 5 parallel chains and recommended that the length of the burn-in period as well as how much “thinning” is required should depend on the problem at hand. We followed these recommendations and adjusted the length of the burn-in period and “thinning” requirements based on the behavior exhibited by the chains.

For each model, three independent parallel sampling chains were run until convergence. Convergence was assessed with the help of the Gelman-Rubin diagnostics. As well, chain trajectories and autocorrelations were monitored to confirm convergence.

Each chain was run for 25,000 iterations with the first 5,000 iterations treated as a burn-in period and being discarded for estimation purposes. To further reduce autocorrelation, chains were thinned and only every 20th observation was retained. This resulted in a total of 1,000 observations from each chain for a grand total of 3,000 observations sampled from the posterior distribution for each model.

Model fit was assessed by using posterior predictive p-values as defined by (11). As well, the four models were compared in terms of model complexity and fit as measured by the Deviance Information Criterion (DIC). Finally, precision of derived estimates was estimated and compared against precision of direct estimates. For shortness, throughout the rest of the thesis, we will refer to "precision of estimates" when we are talking about estimated precision of estimates (i.e. estimated posterior variance and CV of derived estimates).

4.1.1 Full Conditional Distributions

As stated above, the Gibbs sampler was used to generate a sample from the posterior distribution. The Gibbs sampler works with the full conditional distributions to generate a sample from the posterior distribution through an iterative procedure. In this section, we will derive full conditional distributions for each model.

For the model (13), the full conditional distributions were obtained in the following manner. We start with the joint posterior distribution (here $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$)

$$
\begin{aligned}
& f\left(\mu_{ij}, \boldsymbol{\beta}, \psi_j, \sigma_v^2, \alpha_\psi, \sigma_\psi^2 \,\middle|\, P_{ij}\right) \propto \\
& \exp\left[-\sum_{i,j} \frac{(P_{ij}-\mu_{ij})^2}{2\sigma_{ij}^2}\right] \\
& \times \exp\left[-\sum_{i,j} \frac{1}{2\sigma_v^2}\left(\mu_{ij} - \sum_k \beta_k \mathrm{Xk}_{ij} - \psi_j\right)^2\right] \\
& \times \exp\left[-14\sum_j \frac{(\psi_j-\alpha_\psi)^2}{2\sigma_\psi^2}\right] \times \left[\left(\sigma_v^2\right)\left(\sigma_\psi^2\right)\right]^{-14\times 103/2} \\
& \times \exp\left[-\frac{14\times 103}{b\sigma_\psi^2}\right] \times \left[\frac{1}{\sigma_\psi^2}\right]^{(a+1)\times 14\times 103} \\
& \times \exp\left[-\frac{14\times 103}{d\sigma_v^2}\right] \times \left[\frac{1}{\sigma_v^2}\right]^{(c+1)\times 14\times 103}.
\end{aligned}
$$

Then, the full conditional distributions are

$$\mu_{i\,j}\,|\,\boldsymbol{\beta},\,\psi_j,\,\sigma_v^2,\;\alpha_\psi,\,\sigma_\psi^2\;,\,P_{i\,j}\;\sim$$

$$N\left(\frac{1}{\sigma_{i\,j}^2+\sigma_v^2}\,(P_{i\,j}\sigma_v^2+\left(\sum_k\,\beta_k\,\mathrm{Xk}_{i\,j}+\psi_j\right)\sigma_{i\,j}^2\,),\;\frac{\sigma_v^2\;\;\sigma_{i\,j}^2}{\sigma_{i\,j}^2+\sigma_v^2}\right);$$

$$\beta_q\,|\,\mu_{i\,j}\,,\,\boldsymbol{\beta}_{(-q)}\,,\,\psi_j,\,\sigma_v^2,\;\alpha_\psi,\,\sigma_\psi^2\;,\,P_{i\,j}\;\sim$$

$$N\left(\;\frac{1}{\sum_{i\,j}\;\mathrm{Xq}_{i\,j}^2}\left(\sum_{i\,j}\,\mathrm{Xq}_{i\,j}\left(\mu_{i\,j}-\sum_{k\,\neq q}\,\beta_k\,\mathrm{Xk}_{i\,j}-\psi_j\right)\right),\;\frac{\sigma_v^2}{\sum_{i\,j}\;\mathrm{Xq}_{i\,j}^2}\right);$$

$$\psi_j\,|\,\mu_{i\,j}\,,\,\boldsymbol{\beta}\,,\,\sigma_v^2,\;\alpha_\psi,\,\sigma_\psi^2\;,\,P_{i\,j}\;\sim$$

$$N\left(\left(14\,\sigma_v^2\,\alpha_\psi+\sigma_\psi^2\,\sum_i\left(\mu_{i\,j}-\sum_k\,\beta_k\,\mathrm{Xk}_{i\,j}\right)\right)\Big/\left(14\left(\sigma_\psi^2+\sigma_v^2\right)\right),\;\frac{\sigma_v^2\;\;\sigma_\psi^2}{14\left(\sigma_\psi^2\;+\;\sigma_v^2\right)}\right);$$

$$\sigma_v^2\,\Big|\,\mu_{i\,j}\,,\,\boldsymbol{\beta},\,\psi_j\,,\;\alpha_\psi,\,\sigma_\psi^2\;,\,P_{i\,j}\;\sim$$

$$\mathrm{IG}\,(\,14\times\,103\,(c+1.5)-1,$$

$$\frac{2\,d}{2\times 14\times 103+d\,\sum_{i\,j}(\mu_{i\,j}-\sum_k\,\beta_k\,\mathrm{Xk}_{i\,j}-\psi_j)^2}$$

$$\sigma_\psi^2\,\Big|\,\mu_{i\,j}\,,\,\boldsymbol{\beta},\,\psi_j,\,\sigma_v^2,\;\alpha_\psi\,,\,P_{i\,j}\;\sim$$

$$\mathrm{IG}\left(14\times\,103\,(a+1.5)-1,\;\frac{2\,b}{2\times\,14\times\,103+\;14\,\times\,b\,\sum_j\,(\psi_j-\alpha_\psi)^{\,2}}\right);$$

$$\alpha_\psi\;|\,\mu_{i\,j}\,,\,\boldsymbol{\beta},\,\psi_j,\,\sigma_v^2,\;\sigma_\psi^2\;,\,P_{i\,j}\;\sim$$

$$N\left(\frac{\sum_j\,\psi_j}{103},\;\frac{\sigma_\psi^2}{14\times 103}\right).$$

All full conditional distributions are standard normal or inverse gamma distributions and can be easily sampled from. Hence, the Gibbs sampler can be easily implemented to obtain a sample from the joint posterior distribution. Once a sample is generated, marginal point and interval summaries are obtained by the corresponding estimators based on the obtained sample for that component of the vector of parameters. Note that an obtained sample for each component can be considered as a sample from its marginal posterior distribution.

For model (14), the full conditional distributions were obtained in a similar manner. We start with the joint posterior distribution

$$f\left(\mu_{ij}, \beta_0, \beta_1, \psi_j, \sigma_v^2, \tau \,\middle|\, P_{ij}\right) \propto \exp\left[-\sum_{i,j} \frac{(P_{ij}-\mu_{ij})^2}{2\sigma_{ij}^2}\right]$$

$$\times \exp\left[-\sum_{i,j} \frac{(\mu_{ij}-\beta_0-\beta_1 \mathrm{x1}_{ij}-\psi_j)^2}{2\sigma_v^2}\right] \times \left(\sigma_v^2\right)^{-14\times 103/2}$$

$$\times \exp\left[-\frac{1}{2\tau}\sum_{i=1}^{M}\sum_{i<j} w_{ij}(\psi_i-\psi_j)^2\right] \times (\tau)^{-14\times 103/2}$$

$$\times \exp\left[-\frac{14\times 103}{f\tau}\right] \times \left[\frac{1}{\tau}\right]^{(e+1)\times 14\times 103}$$

$$\times \exp\left[-\frac{14\times 103}{d\sigma_v^2}\right] \times \left[\frac{1}{\sigma_v^2}\right]^{(c+1)\times 14\times 103}.$$

Then, the full conditional distributions are

$$\mu_{ij} \,|\, \beta_0, \beta_1, \boldsymbol{\psi}, \sigma_v^2, \tau, P_{ij} \sim$$

$$N\left(\frac{P_{ij}\,\sigma_v^2 + (\beta_0+\beta_1 \mathrm{x1}_{ij}+\psi_j)\sigma_{ij}^2}{\sigma_{ij}^2+\sigma_v^2}, \frac{\sigma_v^2\,\sigma_{ij}^2}{\sigma_{ij}^2+\sigma_v^2}\right);$$

$$\beta_0 \,|\, \mu_{ij},\, \beta_1,\, \boldsymbol{\psi},\, \sigma_v^2,\ \tau,\, P_{ij} \ \sim$$

$$N\left(\frac{\sum_{ij}\,(\mu_{ij} - \beta_1\, \mathrm{x1}_{ij} - \psi_j)}{14\times 103},\ \frac{\sigma_v^2}{14\times 103} \right);$$

$$\beta_1 \,|\, \mu_{ij},\, \beta_0,\, \boldsymbol{\psi},\, \sigma_v^2,\ \tau,\, P_{ij} \ \sim$$

$$N\left(\frac{\sum_{ij}\,\mathrm{x1}_{ij}\,(\mu_{ij} - \beta_0\, - \psi_j)}{\sum_{ij}\,\mathrm{x1}_{ij}^2},\ \frac{\sigma_v^2}{\sum_{ij}\,\mathrm{x1}_{ij}^2} \right);$$

$$\psi_j \,|\, \mu_{ij},\, \beta_0,\, \beta_1,\, \psi_{(-j)},\ \sigma_v^2,\ \tau,\, P_{ij} \ \sim$$

$$N\left(\frac{\sigma_v^2 \sum_{i\neq j} w_{ij}\,\psi_i \,+ \tau \sum_i\,(\mu_{ij} - \beta_0 - \beta_1\, \mathrm{x1}_{ij})}{14\,\tau \,+\, \sigma_v^2\, n_j},\ \frac{\sigma_v^2\ \ \tau}{14\,\tau \,+\, \sigma_v^2\, n_j} \right);$$

$$\sigma_v^2 \,\big|\, \mu_{ij},\, \beta_0,\, \beta_1,\, \boldsymbol{\psi},\, \tau,\, P_{ij} \ \sim$$

$$\mathrm{IG}\ (14\times 103\times (c + 1.5) - 1,$$

$$(2\,d) \Big/ \left(2\times 14\times 103 + \ d \sum_{ij}\,(\mu_{ij} - \beta_0 - \beta_1\, \mathrm{x1}_{ij} - \psi_j)^{\,2} \right);$$

$$\tau \,|\, \mu_{ij},\, \beta_0,\, \beta_1,\, \boldsymbol{\psi},\, P_{ij} \ \sim$$

$$\mathrm{IG}\left(14\times 103\times (e + 1.5) - 1,\ \frac{2\ f}{2\times 14\times 103 \ +\ f\ \sum_{i=1}^{M} \sum_{i<j} w_{ij}\,(\psi_i \ - \ \psi_j)^2} \right).$$

Again, all full conditional distributions are standard normal or inverse gamma distributions and can be easily sampled from. Hence, the Gibbs sampler can be easily implemented and comments made for model (13) are fully applicable in this case as well.

For model (15), the full conditional distributions were obtained in a similar manner. We start with the joint posterior distribution

$$f\ (\beta_1\,,\,\gamma_j\ |\,Z_{i\,j})\ \propto$$

$$\prod_{i\,j}\ \left\{\exp[-E_{i\,j}\ \exp(\beta_1\,\mathrm{x1}_{i\,j}\ +\gamma_j)]\ \times\ [\exp(\beta_1\,\mathrm{x1}_{i\,j}\ +\gamma_j)]^{Z_{i\,j}}\right\}$$

$$\times\ \exp\left[-\frac{14\times\ 103\,(\beta_1\ -\ c)^{\,2}}{2\,d}\right]\ \times\ \exp\left[-\sum_{j}\ \frac{14\,(\gamma_j-a)^{\,2}}{2\,b}\right]\ .$$

Then, the full conditional distributions are

$$\beta_1\,|\,\gamma_j,\ Z_{i\,j}\ \propto$$

$$\prod_{i\,j}\ \left\{\exp[-E_{i\,j}\ \exp(\beta_1\,\mathrm{x1}_{i\,j}\ +\gamma_j)]\ \times\ [\exp(\beta_1\,\mathrm{x1}_{i\,j})]^{Z_{i\,j}}\right\}$$

$$\times\ \exp\left[-\frac{14\times\ 103\,(\beta_1\ -\ c)^{2}}{2\,d}\right]\ ;$$

$$\gamma_j\,|\,\beta_1,\ Z_{i\,j}\ \propto$$

$$\prod_{i}\ \left\{\exp[-E_{i\,j}\ \exp(\beta_1\,\mathrm{x1}_{i\,j}\ +\gamma_j)]\ \times\ [\exp(\gamma_j)]^{Z_{i\,j}}\right\}$$

$$\times\exp\left[-\frac{14\,(\gamma_j-a)^{\,2}}{2\,b}\right]\ .$$

Full conditional distributions do not have a closed form (as expected, given that priors are modeled through Normal distributions, which are not conjugate with the Poisson distribution used for likelihood) and the Metropolis-Hastings algorithm within the Gibbs sampler needs to be used. Once a sample is generated, marginal point and interval summaries are obtained based on the sample for that component of the vector of parameters.

For model (16), a similar process was used to derive full conditional distributions. We start with the joint posterior distribution

$$f\ (\beta_1,\ \varphi_i,\ \gamma_j,\ \tau\ |\ Z_{i\,j})\ \propto$$

$$\prod_{i\,j}\left\{\exp[-E_{ij}\exp(\beta_1\,\mathrm{x1}_{ij}\ +\varphi_i+\gamma_j)]\ \times\ [\exp(\beta_1\,\mathrm{x1}_{i\,j}\ +\varphi_i+\gamma_j)]^{Z_{i\,j}}\right\}$$

$$\times\ \exp\left[-\frac{1}{2\,\tau}\sum_{i=1}^{M}\sum_{i<j}w_{i\,j}\,(\gamma_i\ -\ \gamma_j)^2\right]\times(\tau)^{-14\times 103/2}$$

$$\times\exp\left[-\frac{14\times 103\,(\beta_1\ -\ c)^{\,2}}{2\,d}\right]\times\exp\left[-\sum_{i}\ \frac{103\,(\varphi_i-a)^{\,2}}{2\,b}\right]$$

$$\times\exp\left[-\frac{14\times 103}{f\,\tau}\right]\times\left[\frac{1}{\tau}\right]^{(e+1)\times 14\times 103}.$$

Then, the full conditional distributions are

$$\beta_1\,|\,\varphi_i,\ \boldsymbol{\gamma},\ \tau,\ Z_{i\,j}\ \propto$$

$$\prod_{i\,j}\{\exp[-E_{i\,j}\ \exp(\beta_1\,\mathrm{x1}_{i\,j}\ +\varphi_i+\gamma_j)]$$

$$\times\ [\exp(\beta_1\,\mathrm{x1}_{i\,j})]^{Z_i\,j}\Big\}\times\exp\left[-\frac{14\times 103\,(\beta_1\ -\ c)^{\,2}}{2\,d}\right]\ ;$$

$$\gamma_j\,|\,\beta_1,\varphi_i,\ \gamma_{(-j)},\ \tau,\ Z_{i\,j}\ \propto$$

$$\prod_{i}\left\{\exp[-E_{i\,j}\ \exp(\beta_1\,\mathrm{x1}_{i\,j}\ +\varphi_i+\gamma_j)]\ \times\ [\exp(\gamma_j)]^{Z_{i\,j}}\right\}$$

$$\times\ \exp\left[-\frac{1}{2\,\tau}\sum_{i=1}^{M}w_{i\,j}\,(\gamma_i\ -\ \gamma_j)^2\right]\ ;$$

$$\varphi_i\,|\,\beta_1,\ \boldsymbol{\gamma},\ \tau,\ Z_{i\,j}\ \propto$$

$$\prod_{i\,j}\left\{\exp[-E_{i\,j}\ \exp(\beta_1\,\mathrm{x1}_{i\,j}\ +\varphi_i+\gamma_j)]\ \times\ [\exp(\varphi_i)]^{Z_{i\,j}}\right\}$$

$$\times\ \exp\left[-\sum_{i}\ \frac{103\,(\varphi_i-a)^{\,2}}{2\,b}\right]\ ;$$

$$\tau \mid \beta_1,\ \varphi_i,\ \boldsymbol{\gamma},\ Z_{ij} \propto$$

$$\mathrm{IG}\left(14\times 103\,(e+1.5)-1,\ \frac{2\,f}{2\times 14\times 103\ +\ f\ \sum_{i=1}^{M} w_{ij}\,(\gamma_i\ -\ \gamma_j)^2}\right).$$

With respect to marginal point and interval summaries, comments made for model (15) are fully applicable to model (16) as well.

4.1.2 Estimators

Once an MCMC output for each model (i.e. samples from the joint posterior distribution) is generated, posterior quantities of interest can be computed.

Since not all conditional distributions for the considered models admit a closed form expression, it could be next to impossible to obtain the conditional expectations for all models. For that reason, it was decided to proceed with the "ergodic averages" (as defined by (9) and (10)) to allow for a set of approximate estimates comparable across all models under consideration.

In the Bayesian approach, the form of an estimator used in a particular problem depends on the loss function used for the analysis. Typically, in small area estimation, the squared-error loss function (SEL) has been used.

Once a loss function $L\,(\theta\,,\ a)$ is defined, we wish to minimize posterior expected loss as defined by

$$\rho\,(\pi(\theta \mid y),\ a))\ =\ \int_{\Theta} L\,(\theta\,,\ a)\,d\,F^{\pi\,(\theta \mid y)}(\theta),$$

where $\pi(\theta \mid \mathrm{y})$ is the posterior distribution of θ given the sample y. Then, a Bayes estimator is any rule a that minimizes the posterior expected loss.

It can be shown that under SEL, the posterior mean is a Bayes estimator (*cf.* Berger, 1985) and a measure of accuracy of this estimator is the posterior variance (in other words, posterior variance with respect to an estimator, or posterior MSE as defined by (8), attains its minimum, posterior variance). Finally, the posterior expected loss equals posterior variance in the case of SEL.

Hence, if accuracy as measured by the posterior variance with respect to an estimator (or posterior MSE) is the only criterion of significance for the data users, using SEL and a corresponding Bayes estimator in the form of the posterior mean yields the best estimator. This approach has been widely used in survey small area estimation and in this Chapter, we will proceed with estimation under SEL.

In our case, using the "ergodic averages" approach and an assumption of independence of MCMC chains, posterior means, variances, and expected risks can be estimated as

$$(a)\ \hat{\mu}_{ij} = \frac{1}{Ld} \sum_{l=1}^{L} \sum_{k_l=D}^{D+d} \mu_{ij}^{(k_l)};$$

$$(b)\ \hat{\mathrm{Var}}(\hat{\mu}_{ij}) = \hat{\mathrm{Var}}_{\text{within}}(\hat{\mu}_{ij}) + \hat{\mathrm{Var}}_{\text{between}}(\hat{\mu}_{ij});$$

$$\text{where } \hat{\mathrm{Var}}_{\text{within}}(\hat{\mu}_{ij}) = \frac{1}{Ld} \sum_{l=1}^{L} \sum_{k_l=D}^{D+d} \left(\mu_{ij}^{(k_l)} - \overline{\mu_{ij}^{(l)}} \right)^2;$$

$$\hat{\mathrm{Var}}_{\text{between}}(\hat{\mu}_{ij}) = \frac{1}{L-1} \sum_{l=1}^{L} \left(\overline{\mu_{ij}^{(l)}} - \hat{\mu}_{ij} \right)^2.$$

Here, $\overline{\mu_{ij}^{(l)}}$ is an average of $\mu_{ij}^{(k_l)}$ for chain l; D is the length of the burn-in period; d is the size of the retained sample from the joint posterior distribution; L is the number of chains; and $\mu_{ij}^{(k)}$ is the kth draw from the posterior sample.

For the Normal models (13) and (14), the proportion of people who had a flu shot more than two years ago is estimated (denoted as μ_{ij}). To allow for comparison with the direct estimates, the resulting posterior sample will be transformed back to the count scale and estimates of the number of people who had a flu shot more than two years ago will be derived based on the transformed sample. The same logic will be applied to the Poisson models (15) and (16).

Now that the estimators and full conditional distributions have been derived, we will present the results of the MCMC implementation of each model.

4.2 Convergence And Fit

As stated above, MCMC chains were run for the four models using WinBUGS software. 3,000 posterior distribution sample values were obtained for each model and estimates were obtained. Refer to Appendices C and D for resulting estimates.

For all four models, chain convergence for all variables in all models appeared satisfactory judging by both visual observations as well as by Gelman-Rubin diagnostics. Autocorrelation was also minimal for all variables for all four models. Below, examples of trace charts, Gelman-Rubin diagnostic charts and autocorrelation charts are presented. The charts were very similar for all other variables in all four models.

- Model (13):

Figure 2: Trace chart for element mu [3, 24].

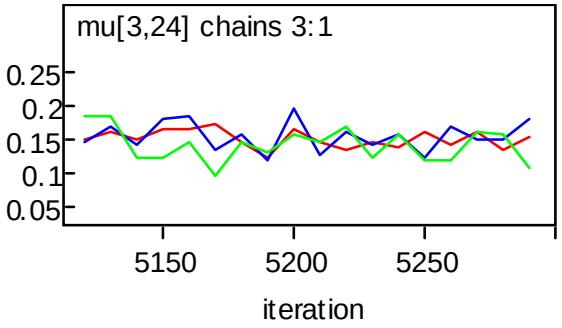


Figure 3: Gelman - Rubin convergence diagnostics for element mu [1, 2].

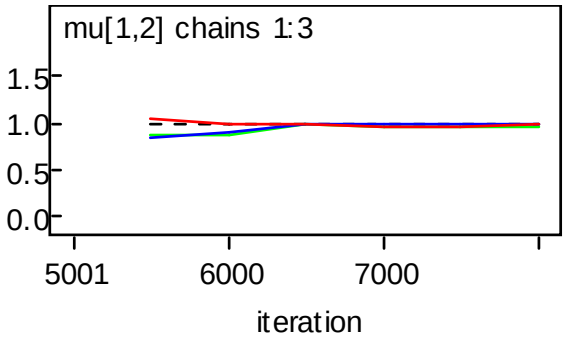
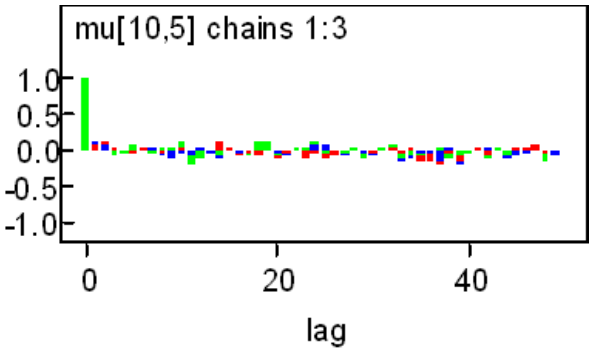


Figure 4: Autocorrelation diagram for element mu [10, 5].



- Model (14):

Figure 5: Trace chart for element mu [7, 74].

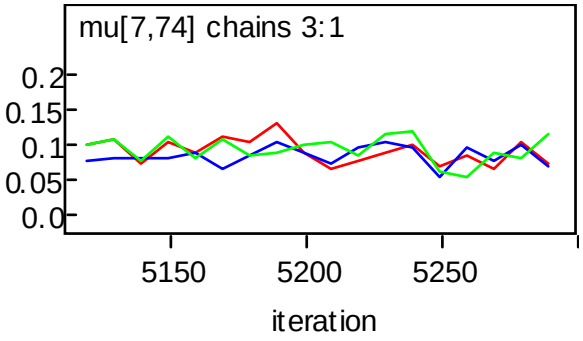


Figure 6: Gelman - Rubin convergence diagnostics for element mu [3, 7].

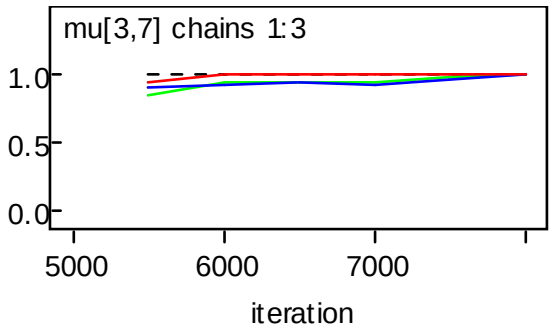
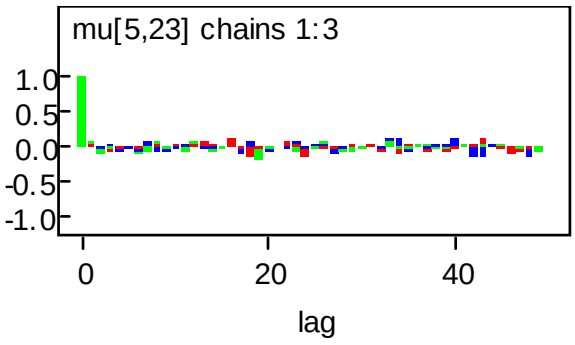


Figure 7: Autocorrelation diagram for element mu [5, 23].



- Model (15):

Figure 8: Trace chart for element mu [1, 7].

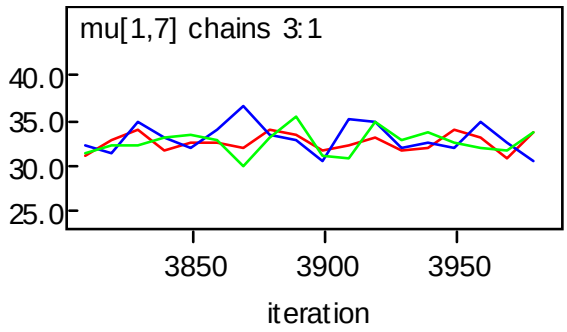


Figure 9: Gelman-Rubin convergence diagnostics for element mu [6, 13].

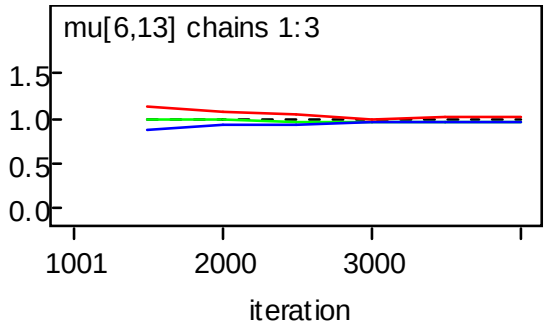
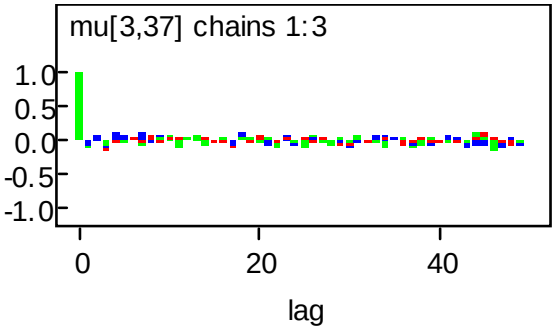


Figure 10: Autocorrelation diagram for element mu [3, 37].



- Model (16):

Figure 11: Trace chart for element mu [9, 9].

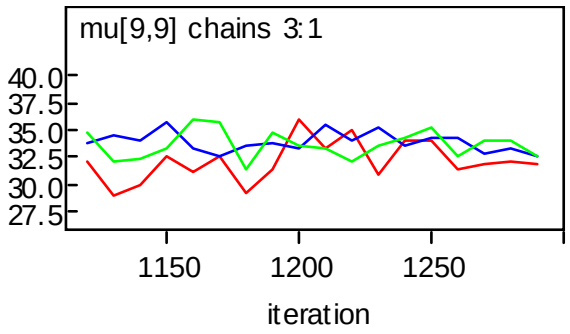


Figure 12: Gelman - Rubin convergence diagnostics for element mu [11, 58].

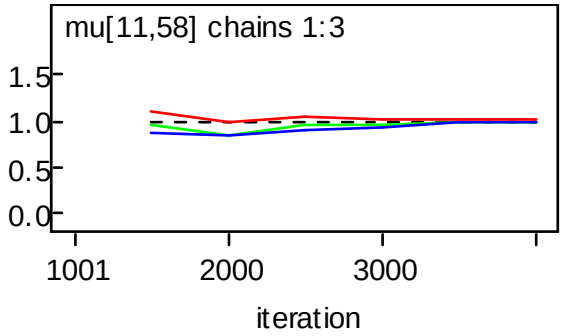
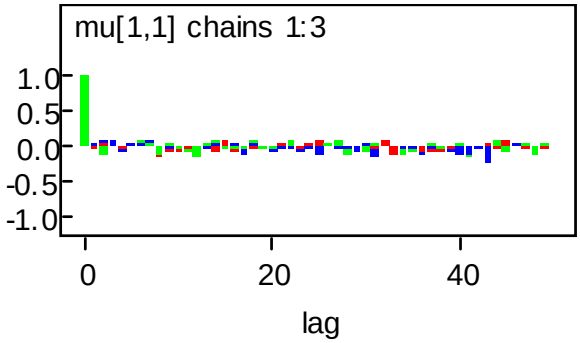


Figure 13: Autocorrelation diagram for element mu [1, 1].



Based on these results for all models, the Gibbs sampler appeared to have achieved convergence. It may be further suggested that the generated samples can be considered approximately independent, since thinning was applied and only every 20th element was retained.

Model fit was assessed by using posterior predictive p-values as defined in (11). The overall measure of goodness-of-fit that was selected as the discrepancy measure for calculations of posterior predictive p-values

$\mathbb{D}(\mathbf{P}, \boldsymbol{\mu}) = \sum_{i\,j} \frac{[P_{i\,j} - \mu_{i\,j}]^2}{\sigma^2_{i\,j}}$, for Normal models (13) and (14);

$\mathbb{D}(\mathbf{Z}, \boldsymbol{\mu}) = \sum_{i\,j} \frac{[Z_{i\,j} - \mu_{i\,j}]^2}{\mu_{i\,j}}$, for Poisson models (15) and (16).

Obtained posterior predictive p-values for the four models under consideration are presented in Table 6. To further assess model complexity and fit, Deviance Information Criterion (DIC) as defined by (12) was applied (please refer to Table 7).

In terms of model fit, the Poisson models (15) and (16) had posterior predictive p-values closest to 0.5 (therefore, they had better fit to the data). Using CAR priors improved posterior predictive p-values in both Normal and Poisson cases.

Table 6. P-values for models (13), (14), (15), and (16).

	p_D	*Conclusions*
Model (13)	0.81	Inadequate fit - this result is somewhat expected given the low R-square value for these covariates
Model (14)	0.64	Adequate fit
Model (15)	0.35	Adequate fit
Model (16)	0.39	Best fit (closest to 0.5)

Table 7. DIC values for models (13), (14), (15), and (16).

	$\bar{D}$	$\hat{D}$	p_D	*DIC*
Model (13)	-4967.7	-5239.5	271.8	-4695.9
Model (14)	-4915.4	-5220.3	304.9	-4610.4
Model (15)	12068.1	11964.2	103.9	12172
Model (16)	10896.6	10783.1	113.5	11010.1

In terms of model complexity and fit as measured by DIC values, the Poisson models (15) and (16) had the smallest effective number of parameters p_D indicating less complexity. The Normal models (13) and (14) had smaller DIC values than the Poisson models, though the direct comparison needs to be done with caution (due to different likelihood functions, and, hence, different deviance statistics). Using CAR priors also led to smaller DIC values in the Poisson case.

In Table 7, $\bar{D}$ is the posterior mean of the deviance statistic; $\hat{D}$ is a point estimate of the deviance; and p_D is an effective number of parameters (see Section 2.4.4 for the detailed definitions).

For the Poisson model (15), using the Normal prior to model spatial effects did not allow separation of age effects - three chains with different initial

values did not converge to the same values indicating an identifiability problem. On the other hand, using the CAR prior to model spatial effects in model (16) allowed separation of age effects, resulted in a decrease in DIC, and improved the posterior predictive p-value.

It should be noted that after completion of the residual analysis a square-root data transformation was implemented for the Poisson models (15) and (16) to stabilize variance and to mitigate a curvature effect in residuals. As a result, DIC for the models went down dramatically, from 150,000+ to 11,000+ and posterior predictive p-values improved from 0.003 to the 0.35 - 0.39 range. Other transformations were attempted as well but did not lead to such significant improvements.

For the Normal models (13) and (14), neither Normal nor CAR priors allowed for separation of age effects - three chains with different initial values did not converge to the same value indicating an identifiability problem.

As already mentioned above, for the models (14), (15), and (16), multicollinearity between covariates proved to be too high and caused deterioration in both DIC and posterior predictive p-values. Hence, only one covariate was retained in each of these models.

Figure 14: Model estimates for age group 1.

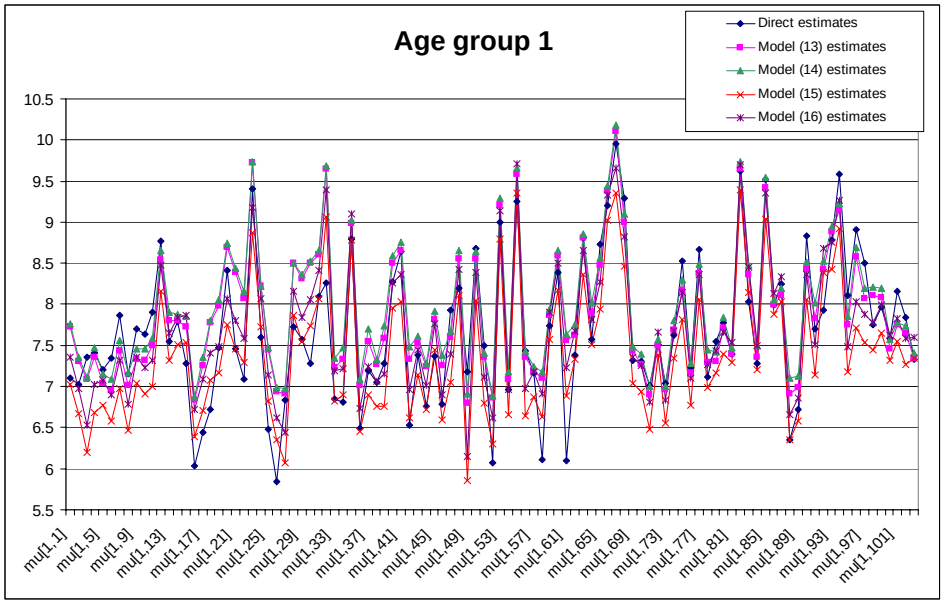


Table 8. CV of estimates, main statistics.

	Direct estimates	*Model (13) estimates*	*Model (14) estimates*	*Model (15) estimates*	*Model (16) estimates*
Ave	45.4	21.2	21.6	9.4	9.8
Min	15.7	8.7	6.2	5.0	5.8
Max	111.4	54.4	66.6	12.3	12.8

Figure 14 presents estimates obtained from all four models as well as CCHS direct estimates for the age group 1. Estimates are graphed on a log-scale to allow for a better representation by reducing the variance in the extreme estimates for different domains. See Appendix C for the charts for all age groups. Results were similar in all age groups.

Overall, the precision of the estimates was improved for all four models compared to the precision of the direct CCHS estimates (refer to Appendix D for the charts of CV of estimates by age group). On average, across all age

groups and Health Regions, the Poisson models (15) and (16) produced the best precision in terms of CV of estimates. Table 8 summarizes the main statistics (average, minimum and maximum values across all small domains) for CV of estimates for each model as well as for the direct estimates.

Another way to compare the precision of estimates across all models is to determine the number of small domains with CV less than 16% (good precision according to Statistics Canada's quality guidelines), the number of small domains with CV less than 33% (adequate precision according to Statistics Canada's quality guidelines), and the number of small domains with CV more than 33% (estimates can not be published due to low precision according to Statistics Canada's quality guidelines). Refer to Table 9 for estimate results.

According to Table 9, the Poisson models produced estimates with good quality in all small domains. The Normal models had 25-30% domains with estimates of good quality, 90-92% domains with estimates of acceptable quality, and 8-10% domains with estimates that still can not be published (according to specified guidelines).

Table 9. CV of estimates, domain counts*.

	Direct estimates	*Model (13) estimates*	*Model (14) estimates*	*Model (15) estimates*	*Model (16) estimates*
CV < 16%	1	364	433	1430	1430
CV < 33%	225	1319	1283	1430	1430
CV > 33%	1205	111	147	0	0

* 12 small domains were excluded from the analysis as the CCHS direct estimates were 0 and no CV was produced.

4.3 Conclusions

All four models have improved the precision of the estimates in comparison with the precision of the direct estimates (which is important if precision is the only criterion of interest). The Poisson models ((15) and (16)) provided the estimates with the lowest CV values. In both Normal and Poisson cases, precision of the estimates changed slightly with the introduction of the CAR prior.

In terms of fit to the data (as measured by posterior predictive p-values), the Normal model (13), where the geographic structure is modeled through a Normal distribution, had the worst fit. Using the CAR prior to model geographical structure and dependence between the Health Regions has improved the posterior predictive p-values, which is a good indicator of significance of a hypothesis of geographically structured heterogeneity.

One further argument in support of significance of a hypothesis of geographically structured heterogeneity can be presented. The CAR distribution depends on a parameter τ that can serve as an indicator of the strength of the

spatial correlations with larger values of τ indicating stronger spatial correlation. For the model (14), the value of τ was estimated to be around 5100 (partially due to the identifiability issue that did not allow us to split out the age effects from the pure geographic effects) and for the model (13), the value of τ was estimated as 3.853. These values are far enough from zero (greater than 1) to indicate a significance of the hypothesis of geographically structured heterogeneity.

On the other hand, the Poisson models (15) and (16) had a lower effective number of parameters p_D than the Normal models indicating less complexity, which is consistent with the overall structure of the models (in terms of a number of hierarchical levels in each model). However, the Normal models (13) and (14) had smaller DIC values than Poisson models, though the direct comparison between these models needs to be done with caution (due to different likelihood functions, and, hence, different deviance statistics).

Overall, it would appear that the Normal model (13) had the worst fit to the data, while the three other models ((14), (15) and (16)) resulted in similar levels of fit to the data. At the same time, the Poisson models (15) and (16) were less complex than the model (14). Moreover, utilization of the CAR prior in model (16) allowed for the separation of the age and geographic effects, which is very appealing for practical analysis and estimation purposes. As well, the Poisson models (15) and (16) resulted in a more significant improvement in the precision of the estimates compared to the Normal models (13) and (14). It might be argued though that the Poisson models overshrunk

the estimates towards the means and possibly introduced a significant bias. In the next Chapter, the bias - variance trade off will be examined in greater detail. But, if the main objective of a study is to improve precision of the direct estimates (which served as motivation for using squared-error loss function with corresponding Bayes estimators), the Poisson model (16) can be suggested as the best amongst the four proposed models to achieve this objective while providing a reasonable compromise in terms of fit to the data, utilization of geographic structuring, and composition / complexity of the model. Interestingly enough, the Fay-Herriot model (13) appeared to be the worst in terms of balancing the precision of estimates, fit to the data, and utilization of geographic structuring.

Comparison to the results obtained by Zhou and You

In the middle of our research, a paper was published by Zhou and You (2008) on Hierarchical Bayes estimation for the survey data collected by the Canadian Communities Health Survey (CCHS). In their paper, Zhou and You argued that spatial structures could have an impact on quality of HB estimates and considered four different HB models to demonstrate the impact. Zhou and You used data from Cycle 1.1 of CCHS and estimated the rate of asthma in 20 Health Regions in the province of British Columbia. Note that cycle 1.1 used different geographic boundaries for Health Regions than cycle 2.1 used for our research. Also, estimation results were published for 15 Health Regions in the province of British Columbia in cycle 2.1. First, they took a basic Fay-Herriot model with smoothed sampling variance estimates obtained through a

common design effect approach (*cf*. You, 2006; You, 2008). Then they modeled spatial structures through a special type of CAR distribution introduced by Leroux *et al*. (1999) and MacNab (2003a, 2003b). This type of CAR distribution reduces to the intrinsic autoregressive model that we used in our models. In the third model, Zhou and You used the Fay-Herriot model without spatial structures, but assumed sampling variance to be unknown (treated as an unknown random variable). And in the last model, they added spatial structures modeled through the CAR distribution to the Fay-Herriot model with unknown sampling variances.

Our approach expands and goes deeper than the paper by Zhou and You. First of all, Zhou and You focused on estimates at HR level while we considered lower level estimates at age group domains within HR. The corresponding HR-level direct estimates obtained from CCHS 1.1 data had CV below 18%, which is within the acceptable range of 33%. In contrast, small domains that we operated with had direct estimates with CV above 33% and as high as 70-80%. In other words, quality of direct estimates as expressed by CV was much lower in our case and using HB estimation to bring CV of estimates within the acceptable range served as an important demonstration of future uses of this methodology.

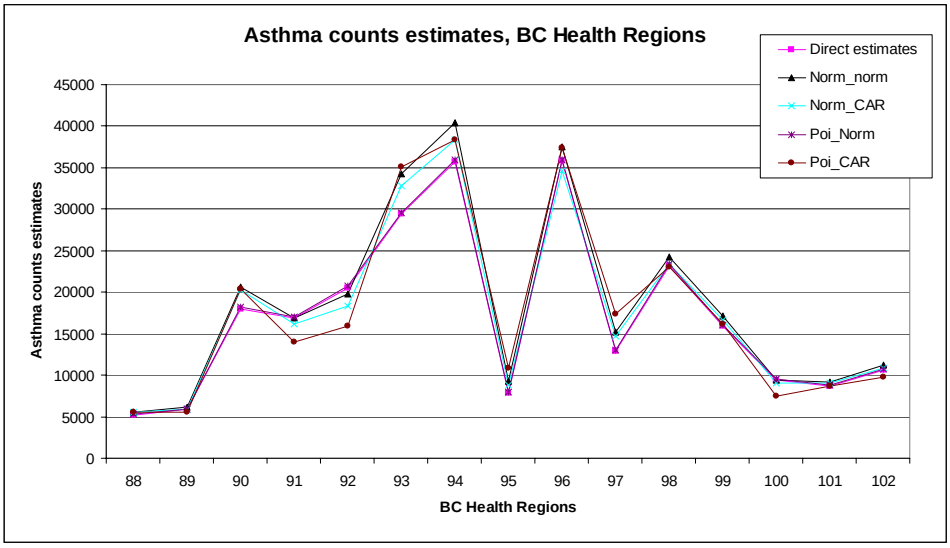
Zhou and You focused on the Fay-Herriot model and introduced spatial effects (modeled through a different distribution than in our models). In our research, we went further and argued that generalized linear models such as Poisson models could be more suitable for these estimation problems. Further-

more, as shown, Poisson models combined with spatial effects allowed for separation of random effects associated with small domain definition (age random effects and geographic random effects).

Finally, in the next chapter of this thesis, we look at utilizating different loss functions in the context of HB estimation and compared resulting estimates. This went above and beyond the scope of work completed by Zhou and You.

In an attempt to compare our results with the results of Zhou and You, we applied all four models introduced above to estimation of asthma counts in 15 Health Regions in the province of British Columbia. Direct comparison of obtained results is impossible due to several reasons. Zhou and You used a different data set (CCHS cycle 1.1) than us (CCHS cycle 2.1) - with all related implications (different sample design, different reference period, etc.). Health Regions had different geographic boundaries in these two cycles. Even questions changed between the cycles and we could not use the same area-specific covariates as Zhou and You because the same questions were not asked during cycle 2.1. Finally, Zhou and You used a different spatial distribution than us.

Figure 15: Asthma counts estimates.



Therefore, we compared estimation results obtained by Zhou and You with our results from a more generic perspective - whether the utilization of spatial structures improve the CV of estimates compared to direct and Fay-Herriot estimates. We also compared results obtained for all four models considered in our thesis to assess if we could derive conclusions similar to the conclusions described above for the original problem studied in the thesis.

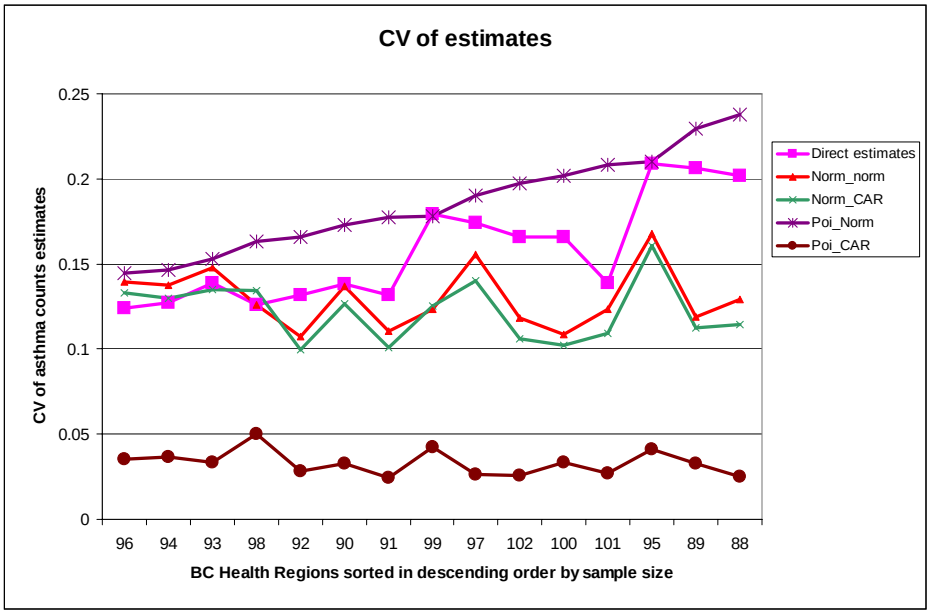
First, we followed the same process as described above to determine covariates for each model (regression analysis and principal component analysis). We identified one covariate for Normal models, specifically, proportion of people with household income between 50,000 and 80,000. For the Poisson models, we identified a different covariate, population estimates from CCHS 2.2 data, to be used in the model. However, when we ran the Poisson models, we observed that regression and spatial parameters could not be identified from the available data (chains were not converging for individual regression

and spatial parameters but the overall chains for the main parameters μ converged). As a result, we ended up using no covariates for the Poisson models.

Figure 15 represents the estimates obtained from all four models as well as direct estimates. Estimates from all models are very close to direct estimates and follow the same patterns.

Turning to CV estimation (please refer to Figure 16), we observed the same results as Zhou and You - CV of direct estimates increased with decrease in sample size; Fay-Herriot model (Norm_norm model in the chart) resulted in improvement in CV and even further improvement was achieved by applying a spatial model (model (14), Norm_CAR model in the chart). We also observed that Poisson-CAR model (model (16), Poi_CAR model in the chart) resulted in more significant improvements in CV than any other considered model. Interestingly enough, Poisson-Normal model (Poi_Norm in the chart) resulted in worst CV for all Health Regions and CV was increasing with decrease in sample size. Given that the Poisson models did not use any covariates, a significant improvement in CV demonstrated by the Poisson-CAR model over the Poisson-Normal model is a very good illustration of the importance of spatial structuring.

Figure 16: CV of asthma counts estimates.



In terms of model fit (as described by DIC criterion), models (14) and (16) (with CAR spatial structures) had better fit to the data than models (13) and (15) (with spatial structures modeled through Normal models). This is another indication of the importance of including spatial structures in estimation models.

Overall, our results were consistent with the results obtained by Zhou and You (greater improvement over the direct estimates in terms of smaller CVs achieved by HB models with CAR spatial components). As well, the results were consistent with conclusions summarized above for the problem considered in this thesis - Poisson-CAR model (16) can be suggested as the best amongst the four proposed models.

It is possible to further improve the precision of estimates produced by the Normal models by adjusting the loss function and placing more weight on

variance reduction. Using the same logic, the quality of the estimates produced by any models can be potentially improved by adjusting the loss function and placing weights to balance bias reduction with lower precision.

We will now proceed with the analysis of the impact of using different loss functions within the context of small area estimation and with application to the problem under consideration.

Chapter 5

Case Study: Loss Functions And Resulting Estimators

5.1 Literature Overview

A traditional approach used by Hierarchical Bayes small area estimation models has been to assume a squared-error loss function (SEL) and to use the posterior mean as a Bayes estimator. In fact, many small area models automatically assumed that the HB estimator is the posterior mean, which is the case under squared-error loss (*cf*. Rao, 2003, Ch. 10; You and Rao, 2002; Best *et al*., 2005; You *et al*., 2003; Datta *et al*., 1999; Ghosh *et al*., 1999; Xia *et al*., 1997; Carlin and Louis, 2000; etc.). However, this approach, while attempting to mirror the frequentist estimator as well as traditional measures of accuracy (MSE and CV, for example), fails to utilize an important aspect of the Bayesian approach, i.e. the ability to structure a loss function depending on the problem under consideration.

Typically, surveys are conducted in response to specific needs. As a consequence, survey results are used to make important decisions such as transfer

payments allocation, funds distribution, initiation of new programs etc. Hence, assuming squared-error loss may not adequately reflect the needs of the data users and may fail to assign a proper penalty for underestimation / overestimation of the parameters of interest. In practice, though, it is difficult to come up with exact loss function specifications. Hence, it may be necessary to obtain estimators under different loss functions and to compare different estimators to assess if any particular estimator would result in smaller losses under all loss functions. We will now present a brief overview of different loss functions and the corresponding Bayes estimators that have been used in small area estimation.

Sometimes, SEL is modified to include various weights - for example, weights designed to take spatial correlation into account (*cf*. Stern and Cressie, 1999) - and estimates under weighted squared-error loss are obtained.

A different loss function is used in situations when the objective of the study is to locate domains with unusually high values or rates of the variables of interest (*cf*. Stern and Cressie, 1999). This question is of a particular interest in epidemiological studies. Typically, the 0-1 loss function is used with the mode of the posterior distribution taken as a Bayes estimator (*cf*. Bernardo and Smith, 2000). Stern and Cressie (1999) proposed another extreme-value loss function that would allow for an estimate of the maximum order statistic and the index of the region attaining the maximum.

In some applications, the loss associated with errors in estimation can be different for different ranges of a parameter to be estimated. In these cases, a

threshold loss function is used. For example, the Small Area Income and Poverty Estimates project (*cf.* National Research Council, 2000) used a threshold loss function that assigns loss as

$$L(\theta, a) = \begin{cases} (\theta - a)^2 & \text{if } \theta > T \\ \\ a^2 & \text{if } \theta < T \end{cases} .$$

In this case, a Bayes estimator is obtained as a product of the posterior mean (conditional on $\theta > T$) and the posterior tail area.

In some cases, the loss function is approximately linear - for example, it might be determined that it is twice as harmful to underestimate as to overestimate. Typically, in these situations, a linear loss is utilized (*cf.* Carlin and Louis, 2000)

$$L(\theta, a) = \begin{cases} K_0 (\theta - a) & \text{if } \theta - a \geq 0 \\ \\ K_1 (a - \theta) & \text{if } \theta - a < 0 \end{cases}$$

Under this loss function, any $\frac{K_0}{K_0 + K_1}$ th percentile is a Bayes estimator of θ.

So far, we have been presenting various loss functions that have been utilized for small area point estimation. Some problems require estimating a histogram or an empirical distribution function of the parameter of interest. Other problems focus on estimating ranks. And in some cases, the triple-goal estimation problem is specified - it is required to produce good estimates of histograms, ranks, and parameters themselves. Shen and Louis (1998), Conlon and Louis (1999), and Louis and Shen (1999) studied the triple-goal estimation problem under weighted squared-error loss.

The last problem that we would like to mention in the context of this Chapter is the problem of a Constrained Bayes estimator. The Constrained Bayes estimator is derived by minimizing the posterior squared error ($E[\sum_i (\theta_i - t_i)| \hat{\boldsymbol{\theta}}]$) under two constraints - on the average and variance of t_i. Louis (1984) and Ghosh (1992) studied the problem and derived corresponding estimators.

5.2 Considered Loss Functions

In our case study, we have considered four different models - the Normal models (13) and (14) and the Poisson models (15) and (16). While SEL works well for the Normal models, estimators obtained from the Poisson models under SEL may not perform well. SEL does not penalize enough for estimation errors when estimates are close to zero, which can be a problem in small area estimation with its small counts and low proportions.

These considerations motivated an attempt to use different loss functions in our problem and to obtain and to compare estimates under different loss functions for all four models under consideration. Given the independence assumptions for P_{ij}'s and Z_{ij}'s, the problem was treated as a collection of 14×103 one-dimensional problems. Hence, no adjustments were made to introduce a multi-dimensional loss function.

The following three loss functions were considered

- Squared-error loss function (SEL)

  $L_1(\theta, a) = (\theta - a)^2$, where θ is a parameter of interest and a is its estimator.

- Normalized squared-error loss function (NSEL)

  $L_2(\theta, a) = (\theta - a)^2 / \theta$, where θ is a parameter of interest and a is its estimator.

  The NSEL penalizes errors in estimation more if θ is near 0 (in comparison with SEL), which is more suitable for small area estimation setup with domain counts expected to be small.

- Weighted balanced-type loss function (WBL) (*cf.* Jafari Jozani *et al.*, 2006)

  $$L_3(\theta, \delta) = \omega\, q(\theta)(\delta - \delta_0)^2 + (1 - \omega)\, q(\theta)(\delta - \theta)^2,$$

  where $0 \leq \omega \leq 1$ is a weight, $q(\theta)$ is a positive weight function, δ_0 is a "target" estimator, and δ is a current estimator. For our case study, we set $q(\theta)$ to be equal to $1/\theta$ and δ_0 to be equal to the direct CCHS estimates.

  WBL was designed to produce estimates that will reflect two objectives. On the one hand, estimates will stay close to the target estimate δ_0. Rubin-

Bleuer (2007) mentioned several concerns expressed by practitioners and theoreticians with respect to the production and use of model-based small-area estimators including the need for benchmarking to direct estimates for internal consistency as well as perceived design bias of the model-based estimators. Using direct estimates as the target estimate δ_0 would "shrink" HB estimates towards the direct estimates that are assumed to be design-unbiased. Thus, intuitively, this approach could be thought of as an attempt to address the concerns mentioned above. And to some extent, WBL can be thought of as another way to introduce constraints on estimators (in terms of closeness to δ_0) rather than utilizing specially designed priors. Another objective to be achieved by WBL and resulting estimates is to minimize NSEL or, in other words, the distance to the unknown parameters.

Let us now look at the three selected loss functions and obtain estimators under each

1) Squared-error loss function (SEL)

- As mentioned above, under SEL, the posterior mean is the Bayes estimator; a measure of the accuracy of this estimator is the posterior variance; and the posterior expected loss equals the posterior variance.
- Using the "ergodic averages" approach, these quantities can be estimated using

$$(a)\ \hat{\mu}_{ij} = \frac{1}{Ld}\sum_{l=1}^{L}\sum_{k_l=D}^{D+d}\mu_{ij}^{(k_l)};$$

$$(b)\ \hat{\mathrm{Var}}(\hat{\mu}_{ij}) = \hat{\mathrm{Var}}_{\text{within}}(\hat{\mu}_{ij}) + \hat{\mathrm{Var}}_{\text{between}}(\hat{\mu}_{ij});$$

where

$$\hat{\mathrm{Var}}_{\text{within}}(\hat{\mu}_{ij}) = \frac{1}{Ld}\sum_{l=1}^{L}\sum_{k_l=D}^{D+d}\left(\mu_{ij}^{(k_l)} - \overline{\mu_{ij}^{(l)}}\right)^2;$$

$$\hat{\mathrm{Var}}_{\text{between}}(\hat{\mu}_{ij}) = \frac{1}{L-1}\sum_{l=1}^{L}\left(\overline{\mu_{ij}^{(l)}} - \hat{\mu}_{ij}\right)^2;$$

and $\overline{\mu_{ij}^{(l)}}$ is an average of $\mu_{ij}^{(k_l)}$ for chain l;

$$(c)\ \hat{\rho}_{ij}^{\mu} = \frac{1}{Ld}\sum_{l=1}^{L}\sum_{k_l=D}^{D+d}\left(\mu_{ij}^{(k_l)} - \hat{\mu}_{ij}\right)^2,$$

where D is the length of a burn-in period; d is the size of a sample generated from the joint posterior distribution; L is the number of chains; and $\mu_{ij}^{(k)}$ is the *k*th draw from the posterior sample.

2) Normalized squared-error loss function (NSEL)

- It can be shown (*cf.* Berger, 1985) that a Bayes estimator under NSEL is the harmonic mean of the posterior (subject to existence of the first moment and inverse first moment)

$$\delta_2 = \frac{1}{\mathrm{E}\left(\frac{1}{\theta}\,|\,\mathbf{y}\right)} = 1\Big/\left(\int_{\Theta}(1/\theta)\,d\mathrm{F}^{\pi(\theta|\mathbf{y})}(\theta)\right).$$

- A measure of the accuracy of this estimator is given by posterior variance with respect to the estimator δ_2 (also known as posterior MSE), $E_{\theta|\boldsymbol{y}}\left(\theta - \delta_2\right)^2$, decomposed in (8) as follows

$$E_{\theta|\boldsymbol{y}}\left(\theta - \delta_2\right)^2 = \mathrm{Var}_{\theta|\boldsymbol{y}}(\theta) + \left(E_{\theta|\boldsymbol{y}}(\theta) - \delta_2\right)^2.$$

- Finally, posterior expected loss is defined as

$$\rho\left(\pi(\theta \mid \boldsymbol{y}),\ \delta_2\right) = \int_{\Theta} \frac{\left(\theta - \delta_2\right)^2}{\theta}\ dF^{\pi(\theta|y)}(\theta)\ .$$

- Using the "ergodic averages" approach, these quantities can be estimated as

$$(a)\ \hat{\delta}_{ij} = \frac{1}{L}\sum_{l=1}^{L}\frac{1}{\overline{\delta_{ij}^{(l)}}}\ ;$$

where $\overline{\delta_{ij}^{(l)}} = \frac{1}{d}\sum_{k_l=D}^{D+d}\frac{1}{\mu_{ij}^{(k_l)}}$;

$$(b)\ \widehat{E_{\theta|y}}\left(\theta - \hat{\delta}_{ij}\right)^2 = \frac{1}{L\,d}\sum_{l=1}^{L}\sum_{k_l=D}^{D+d}\left(\mu_{ij}^{(k_l)} - \hat{\delta}_{ij}\right)^2;$$

where $\mu_{ij}^{(k)}$ is the *k*th draw from the posterior sample and $\hat{\delta}_{ij}$ is the estimator obtained in (*a*);

$$(c)\ \hat{\rho}_{ij}^{\delta} = \frac{1}{L\,d}\sum_{l=1}^{L}\sum_{k_l=D}^{D+d}\left(\mu_{ij}^{(k_l)} - \hat{\delta}_{ij}\right)^2 \Big/ \mu_{ij}^{(k_l)};$$

where D is the length of a burn-in period; d is the size of a sample generated from the joint posterior distribution; L is the number of chains; and $\mu_{ij}^{(k)}$ is the *k*th draw from the posterior sample.

3) Weighted balanced-type loss function (WBL)

- In their paper, Jafari Jozani *et al*. (2006) showed that a Bayes estimator of θ under WBL can be obtained as

$$\delta_3 = \omega\,\delta_0(\boldsymbol{Y}) + (1 - \omega)\left(E(\theta\, q(\theta) \mid \boldsymbol{Y})\right) / \left(E(q(\theta) \mid \boldsymbol{Y})\right).$$

By using q(θ) = 1/θ and δ_0 as the direct estimates, we obtain a Bayes estimator for WBL as

$$\delta_3 = \omega\,\delta_0 + (1-\omega)\,\frac{1}{\mathrm{E}\left(\frac{1}{\theta}\mid \mathbf{Y}\right)} =$$

$$= \omega\,\delta_0 + (1-\omega)\,\delta_2.$$

We used ω = 0.5 so that δ_0 and δ_2 have an equal contribution to the new estimator. Specifically, equal weights were given to the target estimator δ_0 and the distance to the unknown parameter as measured by NSEL. We tried different values of ω to see how much weight could be given to the target estimator δ_0 while maintaining precision of the estimates at the minimum level of the direct CCHS estimates (precision measured by the coefficient of variation). We found that we could increase ω to 0.8 before precision of the estimates would deteriorate below the precision level of the direct CCHS estimates.

- A measure of the accuracy of this estimator is given by posterior variance with respect to the estimator δ_3 (also known as posterior MSE), $E_{\theta\mid \boldsymbol{y}}\left(\theta - \delta_3\right)^2$, decomposed in (8) as follows

$$E_{\theta\mid \boldsymbol{y}}\left(\theta - \delta_3\right)^2 = \mathrm{Var}_{\theta\mid \boldsymbol{y}}(\theta) + \left(E_{\theta\mid \boldsymbol{y}}(\theta) - \delta_3\right)^2.$$

- Posterior expected loss is defined as

$$\rho\left(\pi(\theta\mid \boldsymbol{y}),\,\delta_3\right)) = \int_{\Theta}\left[\frac{\omega}{\theta}\left((\delta_3-\delta_0)^2\right) + \frac{1-\omega}{\theta}\left((\delta_3-\theta)^2\right)\right] dF^{\pi(\theta\mid y)}(\theta).$$

- Using the "ergodic averages" approach, these quantities can be estimated as

$$(a)\ \hat{\psi}_{ij} = \frac{1}{L}\sum_{l=1}^{L}\left(\omega\,\delta_0 + (1-\omega)\,\frac{1}{\psi_{ij}^{(l)}}\right);$$

where $\overline{\psi_{ij}^{(l)}} = \frac{1}{d} \sum_{k_l = D}^{D+d} \frac{1}{\mu_{ij}^{(k_l)}}$;

$$(b)\ \hat{E}_{\theta|y} \left(\theta - \hat{\psi}_{ij}\right)^2 = \frac{1}{L\,d} \sum_{l=1}^{L} \sum_{k_l = D}^{D+d} \left(\mu_{ij}^{(k_l)} - \hat{\psi}_{ij}\right)^2;$$

where $\mu_{ij}^{(k)}$ is the *k*th draw from the posterior sample and $\hat{\psi}_{ij}$ is the estimator obtained in (*a*).

$$(c)\ \hat{\rho}_{ij}^{\psi} = \frac{1}{L\,d} \sum_{l=1}^{L} \sum_{k_l = D}^{D+d} \left[\omega \frac{\left(\hat{\psi}_{ij} - \delta_{0ij}\right)^2}{\mu_{ij}^{(k_l)}} + (1-\omega) \frac{\left(\hat{\psi}_{ij} - \mu_{ij}^{(k_l)}\right)^2}{\mu_{ij}^{(k_l)}} \right];$$

where D is the length of a burn-in period; d is the size of a sample generated from the joint posterior distribution; L is the number of chains; and $\mu_{ij}^{(k)}$ is the *k*th draw from the posterior sample.

5.3 Conclusions

The formulae derived in the previous section were applied to the MCMC outputs for all models (i.e., samples from the posterior distributions). The charts below represent estimates under different loss functions for all models for age group 14. Our conclusions were similar for all age groups (refer to Appendix E for the charts for all age groups). Estimates are graphed on a log-scale to allow for a better representation by reducing the variance in the extreme estimates for different domains.

The obtained estimates were compared with the direct estimates along three dimensions - bias, accuracy (as measured by coefficient of variation), and risk (as measured by posterior expected loss). The following results were obtained.

Figure 17: Model estimates under SEL, for all models, age group 14.

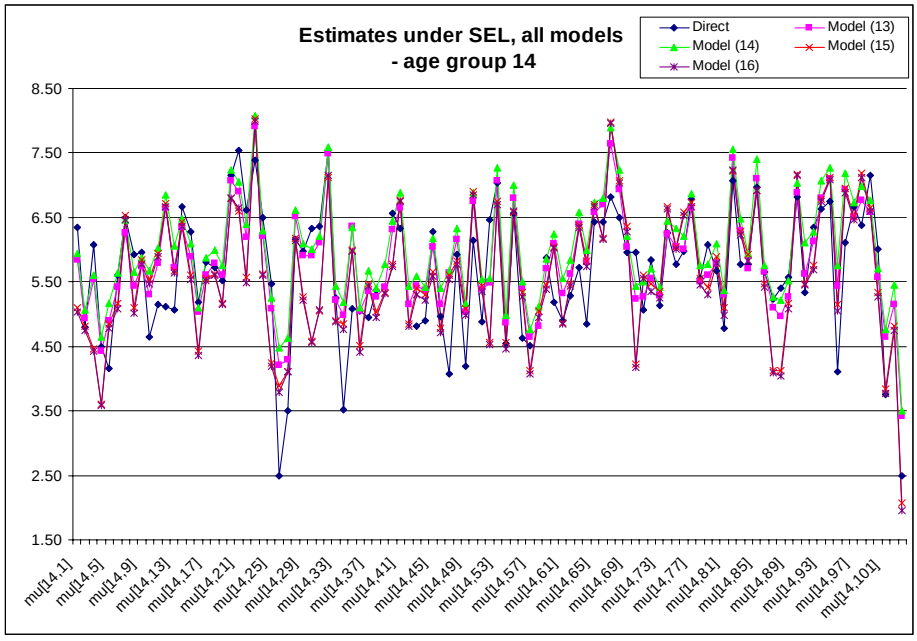


Figure 18: Model estimates under NSEL, for all models, age group 14.

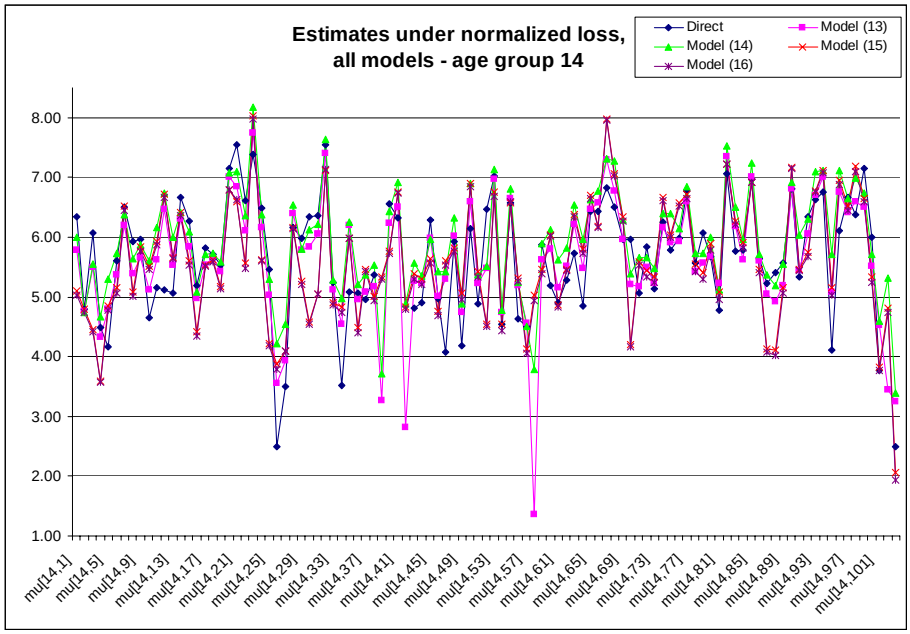


Figure 19: Model estimates under WBL, for all models, age group 14.

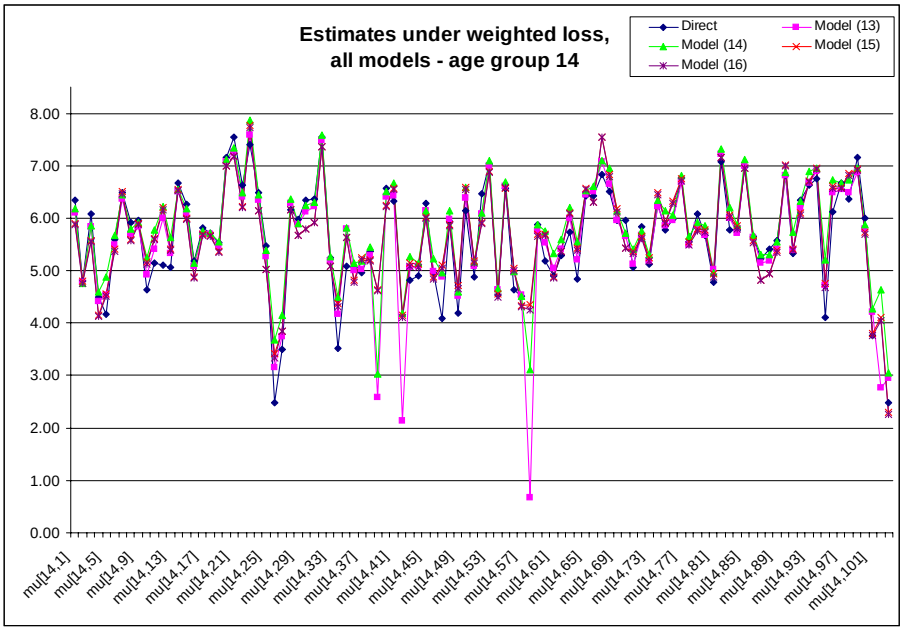


Figure 20: Model (13) estimates under all loss functions, age group 14.

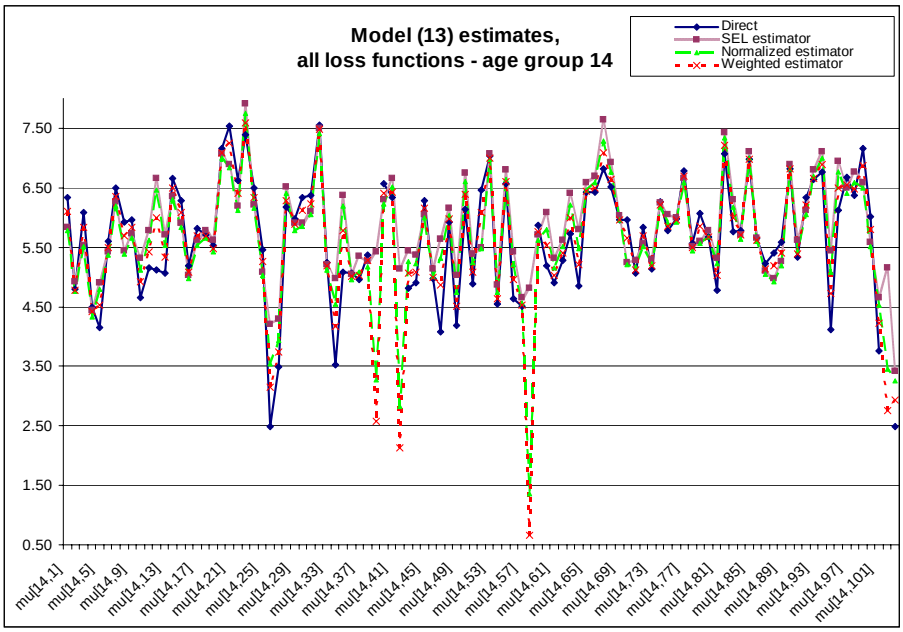


Figure 21: Model (14) estimates under all loss functions, age group 14.

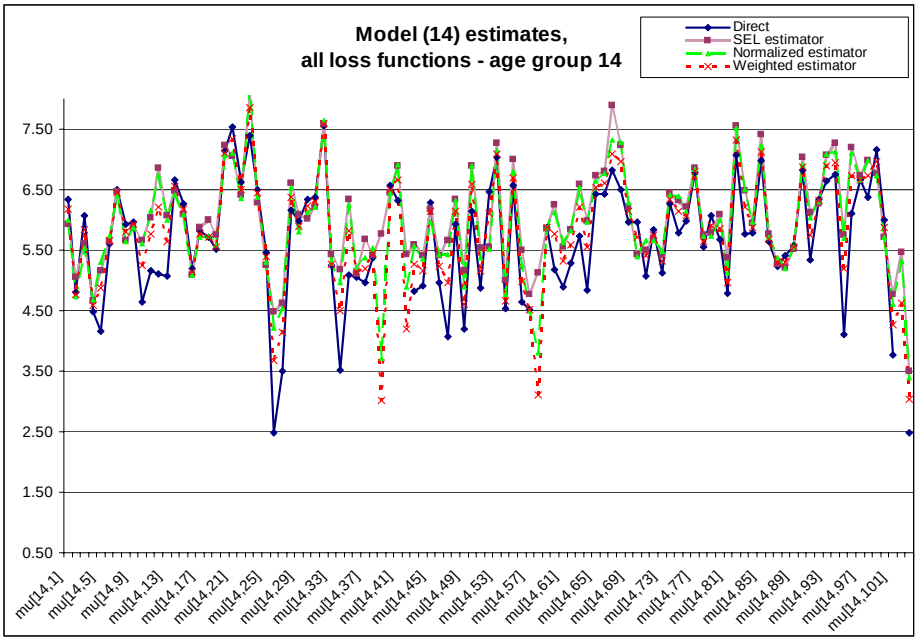


Figure 22: Model (15) estimates under all loss functions, age group 14.

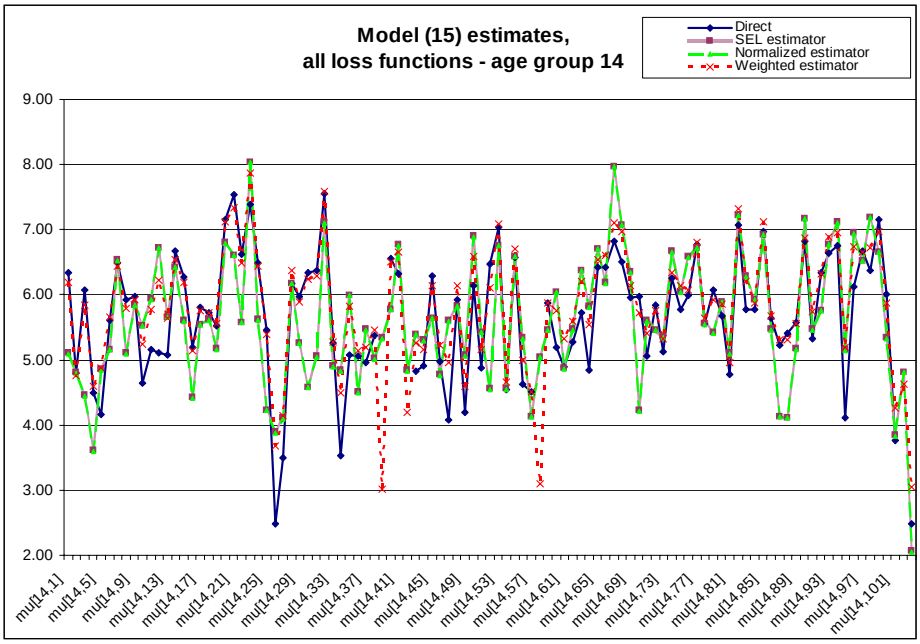
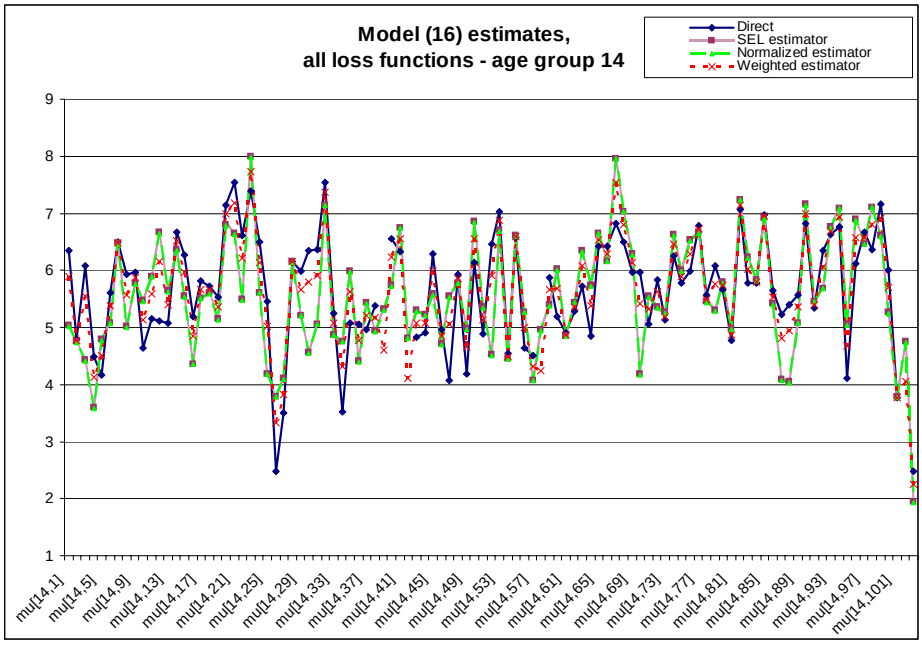


Figure 23: Model (16) estimates under all loss functions, age group 14.



a) Bias:

- To properly assess bias, the HB estimates need to be compared with the unknown parameters or with a set of estimates that are generally accepted as unbiased and valid estimates of these parameters. For example, Ghosh *et al*. (1996) used Census estimates as such reference estimates.

  In our case, there is no readily available set of reference estimates of the number of people who had a flu shot more than two years ago for all small domains under consideration. The only set of estimates that could be considered for comparison are the direct CCHS estimates. The direct estimates are assumed to be design-unbiased, however, there is always the possibility that the obtained set of direct estimates is an outlier and the corresponding confidence intervals will not include the true parameters.

Therefore, even though a comparison between the direct CCHS estimates and the HB estimates could be thought of as an approximation to bias assessment, it would be more prudent to use the term “relative difference” (*cf*. You and Rao, 2002) to reflect the uncertainty associated with how closely the direct estimates represent the true parameters.

- The relative difference between the direct CCHS estimates and the HB estimates was calculated as $|Y_{ij} - \hat{Y}_{ij}| / Y_{ij}$ (Y_{ij} - direct CCHS estimates, $\hat{Y}_{ij}$ - model estimates). Table 10 represents the averag relative differences (RD) for all models and loss functions.
- The Normal models (models (13) and (14)) had the lowest average RD under all three loss functions with model (13) producing the lowest average RD consistently for all three loss functions.
- Model (15) had the highest average RD under all three loss functions - under SEL, its average RD was 27% higher than the average RD of model (13); under NSEL, its average RD was 47% higher than the average RD of model (13); and under WBL, its average RD was 48% higher than the average RD of model (13).
- Models with a spatially structured prior (model (14) in the Normal case and model (16) in the Poisson case) had comparable values of an average RD that differed by 5-10% depending on the loss function.

Table 10. Average relative difference (RD).

	Model (13)	*Model (14)*	*Model (15)*	*Model (16)*
SEL estimates	0.470	0.491	0.599	0.513
NSEL estimates	0.403	0.459	0.593	0.507
WBL estimates	0.201	0.230	0.297	0.254

- As expected, using WBL allowed for a significant reduction in the average RD showing a 42-50% reduction across all the models. On the other hand, using NSEL instead of SEL had the biggest impact for model (13) showing a 15% average RD reduction.
- In the Poisson case, using a spatially structured prior (CAR) led to a decrease in the RD for all three loss functions. The opposite was observed in the Normal case where using a spatially structured prior led to an increase in the RD for all three loss functions.
- There was no consistent overestimation / underestimation in any model nor under any loss function. The differences between the model estimates and the direct estimates fluctuated around zero with approximately 47-51% of the differences being positive and 49-53% of the differences being negative for all combinations of four models and three loss functions (twelve sets of estimates). This would indicate that there was no consistent trend for any estimates from any combination of model / loss function to be always larger / smaller than the direct estimates. Rather, the model estimates were pulled towards local / global means with some model estimates being larger than the direct estimates, and some being smaller.

When comparing estimates under different loss functions for the same model (rather than against the direct estimates), as expected, estimates under SEL and estimates under NSEL were very close (1-6% average difference) with estimates under SEL being larger than the estimates under NSEL. In other words, NSEL slightly shrunk the estimates towards zero. This result was consistent for all four models.

- Further looking at the impact of using different loss functions, estimates under WBL pulled NSEL estimates towards the direct estimates, but there was no consistent trend for WBL estimates from any model to be always larger / smaller than SEL or NSEL estimates for the same model. This result was consistent for all four models.
- When comparing estimates from two Poisson models under the same loss function, estimates from model (16) were consistently smaller than the estimates from model (15). This effect was observed in 67% of small domains. The average difference between the estimates from two models was small though, a 3-5% difference. These results were consistent across all three loss functions.
- When comparing estimates from the two Normal models under the same loss function, there was no clear trend as there was with the Poisson models. 47-53% of small domains had estimates from model (14) larger than the estimates from model (13). An average difference between the estimates from the two models was negligible: 0-2% difference. These results were consistent across all three loss functions. As well, these results were consis-

tent with the observation that in the Poisson case, using the CAR prior led to a decrease in the RD for all loss functions, while this was not true for the Normal models.

- In summary, model (13) had the lowest RD across all loss functions. Model (15) had the highest RD across all loss functions. Models (14) and (16) with the CAR prior were comparable in terms of average RD. In the Poisson case, the introduction of the CAR prior had a favorable effect on the RD reduction, which did not happen in the Normal case. As expected, WBL allowed for a significant reduction in the average RD while SEL and NSEL produced estimates with comparable average RD levels. NSEL led to a slight shrinkage of the estimates towards zero as compared with the SEL estimates for every model.

b) Coefficient of variation (CV)

- In the HB case, the coefficient of variation was defined as the square root of the posterior variance with respect to an estimator (or posterior MSE) divided by the estimator.
- We calculated the posterior variance with respect to an estimator (or posterior MSE) based on the "ergodic averages" approach. On the other hand, Rao (2003, Ch. 10), used the following approach to obtain the posterior MSE of a constrained HB estimator (denoted as $\hat{\theta}^{CHB}$). Rao defined the posterior MSE of a constrained HB estimator as follows

$$E\left[\left(\theta - \hat{\theta}^{CHB}\right)^2 \mid \hat{\theta}\right] = E\left[(\theta - \hat{\theta}^{HB})^2 \mid \hat{\theta}\right] + (\hat{\theta}^{HB} - \hat{\theta}^{CHB})^2$$

$$= \mathrm{Var}(\theta \mid \hat{\theta}) + (\hat{\theta}^{HB} - \hat{\theta}^{CHB})^2,$$

where θ is an unknown parameter to be estimated; $\hat{\theta}$ are direct estimates; $\hat{\theta}^{HB}$ is a Hierarchical Bayes estimator of θ under squared-error loss (i.e. the posterior mean), and $\text{Var}(\theta \mid \hat{\theta})$ is the posterior variance. Rao then pointed out that the posterior MSE could be readily calculated using the posterior variance of θ, posterior mean, and constrained Hierarchical Bayes estimate.

Using the same logic, we could estimate the posterior variance with respect to an estimator δ_{ij} (or posterior MSE given by (8)) as

$$\hat{\text{Var}}(\hat{\delta}_{ij}) = \hat{\text{Var}}(\hat{\mu}_{ij}) + \left(\hat{\mu}_{ij} - \hat{\delta}_{ij} \right)^2;$$

where $\hat{\mu}_{ij}$ is the estimate of the posterior mean and $\hat{\text{Var}}(\hat{\mu}_{ij})$ is the estimate of the posterior variance (posterior variance is defined as $\text{Var}_{\theta|\boldsymbol{y}}(\theta) = \text{E}_{\theta|\boldsymbol{y}}(\theta - \text{E}_{\theta|\boldsymbol{y}}(\theta))^2)$). We will use the term "approximate" estimates for this type of estimates.

We applied "approximate" estimates formula to the estimates under NSEL and WBL and compared the obtained results with the estimates computed through the "ergodic averages" approach. In all four models and for both loss functions, the average difference (averaged across all small domains) between the "ergodic averages" estimates and the "approximate" estimates of the posterior variance with respect to an estimator was less than 0.05%. The maximum difference (in a small domain) was around 0.3%. This result can serve as an indicator that the "approximate" estimation of the posterior variance with respect to an estimator is appropriate for our case study. However, we decided to proceed with the "ergodic averages" estimates of the posterior variance with respect to an estimator so that the "ergodic

averages" approach is utilized consistently for all aspects of the analysis of our case study.

- Overall, the model estimates had smaller coefficients of variation than the direct estimates. Table 11 shows the average CV's for all models / loss function combinations as well as for the direct estimates.
- As expected, WBL resulted in estimates with higher CV than the estimates under the two other loss functions (WBL were constructed to bring model estimates closer to the direct estimates at the expense of increased posterior variance with respect to an estimator).
- As expected, estimates under SEL had the smallest CV followed by the estimates under NSEL and then by the estimates under WBL. The Normal models produced a more substantial increase in CV (12-13% increase) when comparing estimates under SEL with estimates under NSEL while the Poisson models resulted in a smaller increase in CV (2% increase). This is consistent with the argument that NSEL can be more suitable for the Poisson case. However, the Poisson models produced a more substantial increase in the CV (58-63% increase) when comparing estimates under NSEL to the estimates under WBL (the Normal models resulted in a 25-27% increase).
- The CV of the estimates under SEL and under NSEL are slightly higher (3-5% difference) for the models with the CAR prior when compared with the estimates for the models with the Normal prior (in both Poisson and Normal cases). However, in the case of WBL, models with the CAR prior

produced estimates with lower CV than the models with the Normal prior (2-9% difference). It must be noted though that the overall difference between the average CV of the estimates is quite small (when comparing estimates for the models (15) and (16) or the estimates for the models (13) and (14) produced under the same loss function).

- Overall, the Poisson models resulted in a smaller average CV and even WBL did not push the average CV above the acceptability threshold of 33%. The Normal models under WBL produced average CV near the acceptability threshold of 33%. Tables 12-14 present counts of small domains with estimates that had CV within specified ranges for all of the loss functions used to derive the estimates.

Table 11. Average CV of model and direct estimates.

	Model (13)	*Model (14)*	*Model (15)*	*Model (16)*	*Direct estimates*
Estimates under SEL	21.22	21.63	9.41	9.85	45.42
Estimates under NSEL	24.25	24.50	9.54	10.00	45.42
Estimates under WBL	32.83	32.49	25.59	23.36	45.42

Table 12. CV of estimates under SEL, domain counts.

	Model (13)	*Model (14)*	*Model (15)*	*Model (16)*	*Direct estimates*
CV < 16%	364	433	1430	1430	1
CV < 33%	1319	1283	1430	1430	225
CV > 33%	111	147	0	0	1205

Table 13. CV of estimates under NSEL, domain counts.

	Model (13)	*Model (14)*	*Model (15)*	*Model (16)*	*Direct estimates*
CV < 16%	344	276	1430	1430	1
CV < 33%	1194	1208	1430	1430	225
CV > 33%	236	222	0	0	1205

Table 14. CV of estimates under WBL, domain counts.

	Model (13)	*Model (14)*	*Model (15)*	*Model (16)*	*Direct estimates*
CV < 16%	136	350	481	567	1
CV < 33%	1007	1099	1053	1120	225
CV > 33%	423	331	377	310	1205

Table 15. Average expected posterior losses.

	Model (13)	*Model (14)*	*Model (15)*	*Model (16)*
Estimates under SEL	344013	368561	52436	69551
Estimates under NSEL	76	105	12	14
Estimates under WBL	89	101	157	91

- As shown above, the Poisson models (15) and (16) resulted in **all** small domains having a CV below 16% under SEL and NSEL. However, these are the estimates with the largest average RD. In other words, models (15) and (16) with SEL and NSEL resulted in over-shrinkage of estimates and significant bias as measured by RD.
- In summary, even though estimates under WBL had a bigger CV than the estimates under the two other loss functions, the CV of these estimates was still within an acceptable range (less than 33%) in approximately 75% of the small domains. This is in comparison with 15% of small domains that had the direct estimates with a CV of less than 33%. Additionally, the estimates under WBL had a significant decrease in the average RD for all models. The Poisson model (16) with the CAR prior had smaller average RD and better CV results as compared with the Poisson model (15) with the Normal prior.

c) Expected posterior loss

- Table 15 summarizes average expected posterior losses for all model / loss function combinations (refer to Appendix G for detailed charts).

Under SEL and NSEL, expected posterior losses for the models with the CAR prior (models (14) and (16)) were higher than expected posterior losses for the models with the Normal prior (models (13) and (15)). Under WBL, the Poisson model with the CAR prior, model (16), resulted in smaller expected posterior losses than the Poisson model with the Normal prior, model (15).

- In the Poisson case, under SEL, 50% of the domains had expected posterior losses smaller for model (15) as compared with model (16). Under NSEL, 58% of the domains had expected posterior losses smaller for model (15) as compared with model (16). Under WBL, 42% of the domains had expected posterior losses smaller for model (15) as compared with model (16).
- In the Normal case, the numbers were much more definite. Under SEL, 91% of small domains had expected posterior losses smaller in the case of the Normal prior (for model (13) as compared with model (14)). Under NSEL, this number increased to 98%. Under WBL, this number decreased to 68%.
- In summary, the models with the Normal priors (models (13) and (15)) resulted in lower levels of the expected posterior losses for each loss function. In the Normal case, this was the case for all three loss functions. In the Poisson case, the model with the CAR prior (model (16)) produced better expected posterior losses under WBL while the model with the Normal prior (model (15)) had better results under the two other loss functions.

Overall, under SEL and NSEL, the Normal models had the lower average RD. On the other hand, the Poisson models had better results for CV and the expected posterior losses. However, all four models offered a significant improvement over the direct estimates in terms of the CV of the estimates. Furthermore, introduction of the CAR prior in the Normal case did not appear to offer any significant advantages in terms of bias reduction and CV improvement. Hence, the basic Fay-Herriot model (13) appeared to be quite adequate under SEL and NSEL. Finally, using NSEL in the case of model (13) offered a slight reduction in the average RD but resulted in an increase in CV. Based on that, the usual SEL function in combination with model (13) appeared to be the best choice when balancing bias and CV trade-off (which offers a different set of criteria than the ones considered in the previous chapter).

The situation is quite different once we consider WBL. WBL allowed a significant average RD reduction for all models, but in the case of the Poisson models, a corresponding increase in CV was not as large as for the Normal models. Hence, the estimates from the Poisson models under WBL offered a better trade-off between bias (as measured by the average RD) and CV than the estimates from the Normal models. Furthermore, the CAR prior appeared to play a significant role in the Poisson case with model (16) offering the best set of estimates under WBL. Combined with the results obtained in the previous chapter (better fit to the data for the Poisson models; the Poisson models being less complex; utilization of the CAR prior in model (16) allowing separation of the age and geographic effects; better CV for the Poisson models), the

Poisson-CAR model (16) can be suggested as the best estimation model under WBL for the problem under consideration.

This result is encouraging as it shows a possible way to better reconcile HB models with the existing direct estimates while still allowing a significant reduction in the CV of estimates. WBL can be easily adjusted to handle any existing direct estimates, regardless whether they were produced from survey data or Census data or administrative records. On the other hand, WBL may also offer a way to compromise practical needs with research needs. On the practical end, estimates would be better balanced in terms of an RD-CV trade off. On the research end, a better separation of the various random effects can be entertained as a mechanism to investigate the contribution of different variables to the resulting estimates. The introduction of a geographic component can be one such example as intuitively it is very appealing and reasonable to have models with geographically structured components.

Chapter 6

Future Research

Analysis performed in the case study presented in this thesis indicated several areas of possible future research. First of all, using a geographically structured prior has proved to be beneficial, especially in the Poisson case. Thus, a more in-depth analysis of the impact of using different spatial models as well as different spatial priors is suggested. Moreover, the Poisson model appeared to be more sensitive to the use of a geographically structured prior (CAR prior), which can point to a need for further analysis of the interplay between the type of a model used and its sensitivity to the use of a geographically structured prior.

Utilization of WBL can be suggested as another area of research. As shown above, this loss function allowed for reduction in the average RD while retaining good precision of the estimates. Thus, using WBL can offer a way to combine practical and theoretical needs. From the practical perspective, WBL can be thought of as another way to benchmark model estimates with direct estimates (or any other estimates, for that matter), thus, allowing for protection against model failure. From the theoretical perspective, WBL will allow

utilization of a powerful Bayesian methodology. Thus, further investigation of how to better structure WBL within the context of small area estimation, what kind of estimates to use as δ_0, potential relationships between the type of model used and the type of loss function applied can be of interest. As well, there might be a connection between the Constrained Bayes estimators proposed by Louis (1984) and Ghosh (1992) and the estimators derived under WBL. Another interesting question to investigate with respect to WBL is to compare different methods of choosing the weights.

On the other hand, the analysis performed above would suggest that finding an appropriate HB model to produce model estimates can help not only better balance bias and variation, but can also offer a mechanism to study the impact of different random effects. For example, in the case study, the Poisson-CAR model (16) was designed to separate age and geographic effects. At the same time, the proposed Normal models did not formally include constraints on the range of values for the proportion under estimation (between 0 and 1). Once such constraints are included, performance of Normal models could improve. Traditionally, a log-linear linking model has been suggested to address this issue. However, there might be a different way of accounting for these types of constraints; through imposing constraints on the prior distribution, for example. Other fields of study (outside of the survey small area estimation) have utilized such methods and models in application to small area analysis and further analysis of applicability of these models to small area estimation can prove very beneficial.

In this thesis, we used three different loss functions applied to a specific problem. The results were encouraging, which may suggest studying other loss functions and their application to survey small area estimation. In the end, the goal should be to come up with a loss function that reflects the needs of data users and the risks that they place on estimation errors. While SEL and MSE have been traditionally utilized because of the connection with the traditional estimation methods, this way of thinking has artificially constrained the application of the HB methodology. Further analysis of various loss functions and their application to the survey small area estimation problems can help remove this constraint.

However, utilization of Bayesian estimation methods in this context requires reconciliation of the Bayesian measures of the precision of estimators (posterior variance, for example) with frequentist measures of precision (MSE). This issue was considered and analyzed for the usual Fay-Herriot model (*cf.* Singh *et al*., 1998) but more analysis is required for more general models like the ones used in this thesis.

Finally, while some of the research directions mentioned above can be quite extensive and effort-consuming, one immediate way to continue with the line of analysis presented in this thesis would be to apply the discussed methods to other variables in the CCHS 2.1 dataset and to assess if the results would be comparable with the results presented above.

References

1. Bell, W.R. (1999), Accounting for Uncertainty About Variances in Small Area Estimation, *Bulletin of the International Statistical Institute,* 52nd session, Helsinki.
2. Berger, J.O. (1985), *Statistical Decision Theory and Bayesian Analysis*, New York: Springer-Verlag.
3. Bernardinelli, L., and Montomoli, C. (1992), Empirical Bayes Versus Fully Bayesian Analysis of Geographical Variation in Disease Risk, *Statistics in Medicine*, **11**, 983-1007.
4. Bernardo, J.M., and Smith, A.F.M. (2000), *Bayesian Theory*, Chichester, UK: Wiley.
5. Besag, J., York, J., and Mollié, A. (1991), Bayesian Image Restoration With Two Applications in Spatial Statistics, *Annals of the Institute of Statistical Mathematics*, **43**, 1-59.
6. Best, N., Richardson, S., and Thomson, A. (2005), A Comparison of Bayesian Spatial Models for Disease Mapping, *Statistical Methods in Medical Research*, **14**, 35-59.

7. Butar, F.B., and Lahiri, P. (2001), On Measures of Uncertainty of Empirical Bayes Small-Area Estimators, Technical Report, Department of Statistics, University of Nebraska, Lincoln.
8. Carlin, B.P., and Louis, T.A. (2000), *Bayes and Empirical Bayes Methods for Data Analysis*, New York: Chapman & Hall.
9. CCHS User Guide for the Public Use Microdata File (2005), Statistics Canada, Ottawa, Canada.
10. Chung, Y.S., Lee, K-O., and Kim, B.C. (2001), Adjustment if Unemployment Estimates Based on Small Area Estimation in Korea, Technical Report, Department of Mathematics, KAIST, Taejun, Korea.
11. Clayton, D., and Kaldor, J. (1987), Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping, *Biometrics*, **43**, 671-681.
12. Clayton, D., and Bernardinelli, L. (1992), Bayesian Methods for Mapping Disease Risk, in P. Elliott, J. Cuzick, D. English, and R.Stern (Eds.), *Geographical and Environmental Epidemiology: Methods for Small-Area Studies*, Oxford: Oxford University Press.
13. Conlon, E.M., and Louis, T.A. (1999), Addressing Multiple Goals in Evaluating Region-Specific Risk Using Bayesian Methods, in A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J.F. Viel, and R. Bertollini (Eds.), *Disease Mapping and Risk Assessment for Public Health*, Chichester, UK: Wiley, pp 31-47.

14. Cowles, M.K., and Carlin, B.P. (1996), Markov Chain Monte Carlo Convergence Diagnostics: a Comparative Review, *Journal of the American Statistical Association*, **91**, 883-904.

15. Cressie, N. (1989), Empirical Bayes Estimation of Undercount in the Decennial Census, *Journal of the American Statistical Association*, 8**4**, 1033-1044.

16. Cressie, N. (1991), Small-Area Prediction of Undercount Using the General Linear Model, in *Proceedings of Statistics Symposium 90: Measurement and Improvement of Data Quality*, Ottawa: Statistics Canada, pp. 93-105.

17. Cressie, N. (1993), *Statistics for Spatial Data*, New York: Wiley.

18. Cressie, N., and Chan, N.H. (1989), Spatial Modelling of Regional Variables, *Journal of the American Statistical Association*, 8**4**, 393-401.

19. Das, K., Jiang, J., and Rao, J.N.K. (2001), Mean Squared Error of Empirical Predictor, Technical Report, School of Mathematics and Statistics, Carleton University, Ottawa, Canada.

20. Datta, G.S., Fay, R.E., and Ghosh, M. (1991), Hierarchical and Empirical 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.

21. Datta, G.S., and Ghosh, M. (1991), Bayesian Prediction in Linear Models: Applications to Small Area Estimation, *Annals of Statistics*, **19**, 1748-1770.

22. Datta, G.S., Lahiri, P., and Maiti, T. (2002), Empirical Bayes Estimation of Median Income of Four-Person Families by State Using Time Series and

Cross-Sectional Data, *Journal of Statistical Planning and Inference*, **102**, 83-97.

23. Datta, G.S., Lahiri, P., Maiti, T., and Lu, K.L. (1999), Hierarchical Bayes Estimation of Unemployment Rates for the U.S. States, *Journal of the American Statistical Association*, **94**, 1074-1082.

24. Denison, D.G.T., and Holmes, C.C. (2001), Bayesian Partitioning for Estimating Disease Risk, *Biometrics*, **57**, 143-149.

25. Dick, P. (1995), Modeling Net Undercoverage in the 1991 Canadian Census, *Survey Methodology*, **21**, 45-54.

26. Ericksen, E.P., and Kadane, J.B. (1985), Estimating the Population in Census year: 1980 and Beyond (with discussion), *Journal of the American Statistical Association*, **80**, 98-131.

27. 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*, New York: Wiley, pp 91-102.

28. Fay, R.E., and Herriot, R.A. (1979), Estimation of Income from Small Places: An Application of James-Stein Procedures to Census Data, *Journal of the American Statistical Association*, **74**, 269-277.

29. Fuller, W.A., and Battese, G.E. (1973), Transformations for Estimation of Linear Models with Nested-Error Structure, *Journal of the American Statistical Association*, **68**, 626-632.

30. Gelfand, A.E. (1996), Model Determination Using Sampling-Based Methods, in Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (Eds.) (1996),

Markov Chain Monte Carlo in Practice, London: Chapman and Hall, pp. 145-161.

31. Gelfand, A.E., Dey, D.K., and Chang, H. (1992), Model Determination Using Predictive Distributions with Implementation via Sampling-Based Methods (with discussion), in J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith (Eds), *Bayesian Statistics 4*, Oxford: Oxford University Press, pp. 147-167.
32. Gelfand, A.E., and Smith, A.F.M. (1990), Sample-Based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association*, **85**, 972-985.
33. Gelfand, A.E., and Smith, A.F.M. (1991), Gibbs Sampling for Marginal Posterior Expectations, *Communications in Statistics - Theory and Methods*, **20**, 1747-1766.
34. Gelman, S., and Rubin, D.B. (1992a), A Single Sequence from the Gibbs Sampler Gives a False Sense of Security, in J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith (Eds), *Bayesian Statistics 4*, Oxford: Oxford University Press, pp. 625-631.
35. Gelman, A., and Rubin, D.B. (1992b), Inferences from Iterative Simulation Using Multiple Sequences (with discussion), *Statistical Science*, **7**, 457-511.
36. Geman, S., and Geman, D. (1984), Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.

Gelman, A., and Meng, S.L. (1996), Model Chequing and Model Improvement, in Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (Eds.) (1996), *Markov Chain Monte Carlo in Practice*, London: Chapman and Hall, pp. 189-201.

38. Geyer, C.J. (1992), Practical Markov Chain Monte Carlo (with discussion), *Statistical Science*, **7**, 473-511.

39. Ghosh, M. (1992), Constrained Bayes Estimation with Applications, *Journal of the American Statistical Association*, **87**, 533-540.

40. Ghosh, M., Natarajan, K., Stroud, T.W.F., and Carlin, B.P. (1998), Generalized Linear Models for Small Area Estimation, *Journal of the American Statistical Association*, **93**, 273-282.

41. Ghosh, M., and Nangia, N., (1993), Estimation of Median Income of Four-person Families: A Bayesian Time Series Approach, Technical Report, Department of Statistics, University of Florida, Gainsville.

42. Ghosh, M., Nangia, N., and Kim, D. (1996), Estimation of Median Income of Four-person Families: A Bayesian Time Series Approach, *Journal of the American Statistical Association*, **91**, 1423-1431.

43. Ghosh, M., Natarajan, K., Waller, L.A., and Kim, D. (1999), Hierarchical Bayes GLMs for the Analysis of Spatial Data: an Application to Disease Mapping, *Journal of Statistical Planning and Inference*, **75**, 305-318.

44. Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (Eds.) (1996), *Markov Chain Monte Carlo in Practice*, London: Chapman and Hall.

45. Green, P., and Richardson, S. (2002), Hidden Markov Models and Disease Mapping, *Journal of the American Statistical Association*, **97**, 1055-1070.

46. Harville, D.A., and Jeske, D.R. (1992), Mean Squared Error of Estimation or Prediction Under General Linear Model, *Journal of the American Statistical Association*, **87**, 724-731.

47. Hastings, W.K. (1970), Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika*, **57**, 97-109.

48. Jafari Jozani, M, Marchand, E., and Parsian, A. (2006), On estimation with weighted balanced-type loss function, *Statistics & Probability Letters*, **76**, 773-780.

49. Jiang, J., Lahiri, P., and Wan, S.-M. (2002), A Unified Jackknife Theory, *Annals of Statistics*, **30**, 1782-1810.

50. Kackar, R.N., and Harville, D.A. (1981), Unbiasedness of Two-Stage Estimation and Prediction Procedures for Mixed Linear Models, *Communications in Statistics, Series A*, **10**, 1249-1261.

51. Kass, R.E., and Raftery, A. (1995), Bayes Factors, *Journal of the American Statistical Association*, **90**, 773-795.

52. Kass, R.E., and Steffey, D. (1989), Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models), *Journal of the American Statistical Association*, 8**4**, 717-726.

53. Knorr-Held, L., and Raber, G. (2000), Bayesian Detection of Clusters and Discontinuities in Disease Maps, *Biometrics*, **56**, 13-21.

54. Kutner, M.H., Nachtsheim, C.J., Neter, J., and Li, W. (2004), *Applied Linear Statistical Models*, New York: McGraw-Hill / Irwin.

55. Laird, N.M., and Louis, T.A. (1987), Empricial Bayes Confidence Intervals Based on Bootstrap Samples, *Journal of the American Statistical Association*, **82**, 739-750.

56. Laud, P., and Ibrahim, J. (1995), Predictive Model Selection, *Journal of the Royal Statistical Society, Series B*, **57**, 247-262.

57. Leroux, B. G., Lei, X., Breslow, N. (1999), Estimation of disease rates in small areas: a new mixed model for spatial dependence. In M.E. Halloran, D. Berry (Eds), *Statistical Models in Epidemiology, the Environment and Clinical Trials*, Springer-Verlag: New York, pp. 135-178.

58. Louis, T.A. (1984), Estimation a Population of Parameter Values Using Bayes and Empirical Bayes Methods, *Journal of the American Statistical Association*, **79**, 393-398.

59. Louis, T.A., and Shen, W. (1999), Innovations in Bayes and Empirical Bayes Methods: Estimating Parameters, Populations and Ranks, *Statistics in Medicine*, **18**, 2493-2505.

60. MacNab, Y. C. (2003a), Hierarchial Bayesian spatial modeling of small-area rates of non-rare disease, *Statistics in Medicine*, **22**, 1761-177.

61. MacNab, Y.C. (2003b), Hierarchical Bayesian Modeling of Spatially Correlated Health Service Outcome and Utilization Rates, *Biometrics*, **59**, 305-316.

62. Meyn, S.P., and Tweedie, R.L. (1994), *Markov Chains and Stochastic Stability*, London: Spinger-Verlag.

63. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953), Equations of state calculations by fast computing machines, *Journal of Chemical Physics*, **21**, 1087-1091.

64. Nandram, B., Sedransk, J., and Pickle, L. (1999), Bayesian Analysis of Mortality Rates from U.S. Health Service Areas, *Sankhyā : The Indian Journal of Statistics, Series B*, 61, 145-165.

65. National Research Council (2000), *Small-Area Estimates of School-Age Children in Poverty: Evaluation of Current Methodology*, C.F. Citro and G. Kalton (Eds.), Committee on National Statistics, Washington, DC: National Academy Press.

66. Nobre, A.A., Schmidt, A.M., and Lopes, H.F. (2005), Spatio-temporal models for mapping the incidence of malaria in Para, *Environmetrics*, **16**, 291-304.

67. Pfeffermann, D., and Burck, L. (1990), Robust Small Area Estimation Combining Time Series and Cross-Sectional Data, *Survey Methodology*, **16**, 217-237.

68. Polson, N.G. (1996), Convergence of Markov Chain Monte Carlo Algorithms (with discussion), in J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith (Eds), *Bayesian Statistics 5*, Oxford: Oxford University Press, pp. 297-322.

69. Prasad, N.G.N., and Rao, J.N.K. (1990), The Estimation of the Mean Squared Error of Small-Area Estimators, *Journal of the American Statistical Association*, **85**, 163-171.

70. Raftery, A.E., and Lewis, S. (1992), How Many Iterations in the Gibbs Sampler? in J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith (Eds), *Bayesian Statistics 4*, Oxford: Oxford University Press, pp. 763-773.

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

72. Rao, J.N.K., and Yu, M. (1992), Small Area Estimation by Combining Time Series and Cross-Sectional Data, *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 1-9.

73. Rao, J.N.K., and Yu, M. (1994), Small Area Estimation by Combining Time Series and Cross-Sectional Data, *Canadian Journal of Statistics*, **22**, 511-528.

74. Roberts, G.O., and Polson, N. (1994), On the Geometric Convergence of the Gibbs Sampler, *Journal of the Royal Statistical Society, series B*, **56**, 377-384.

75. Roberts, G.O., and Tweedie, R.L. (1994), Geometric Convergence and Central Limit Theorems, for Multidimensional Hastings and Metropolis Algorithms, *Biometrika*, **83**, 95-110.

76. Rosenthal, J.S. (1993), Rates of Convergence for Data Augmentation on Finite Sample Spaces, *Annals of Applied Probability*, **3**, 819-839.

77. Rubin-Bleuer, S. (2007), Current SAE Practice and Issues at Statistics Canada, Technical Report, Statistics Research and Innovation Division, Statistics Canada.

78. Shen, W., and Louis, T.A. (1998), Triple-Goal Estimates in Two-Stage, Hierarchical Models, *Journal of the Royal Statistical Society, Series B*, **60**, 455-471.

79. Singh, A. C., Stukel, D. M., and Pfeffermann, D. (1998), Bayesian versus Frequentist Measures of Error in Small Area Estimation, *Journal of the Royal Statistical Society, Series B*, **60**, 377-396.

80. Skinner, C.J. (1994), Sample Models and Weights, *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Washington, DC, pp. 133-142.

81. Spiegelhalter, D.J., Best, N., and Carlin, B.P. (1998), Bayesian Deviance, the Effective Number of Parameters, and the Comparison of Arbitrarily Complex Models, Research Report 98-009, Division of Biostatistics, University of Minnesota.

82. Spiegelhalter, D.J., Thomas, A., Best, N., and Gilks, W.R. (1995a), BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.50, Technical Report, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University.

83. Spiegelhalter, D.J., Thomas, A., Best, N., and Gilks, W.R. (1995b), BUGS Examples, Version 0.50, Technical Report, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University.

84. Spiegelhalter, D.J., Thomas, A., Best, N., and Lunn, D. (2003), *WinBUGS User Manual, Version 1.4.*

85. Stein, M.L. (1999), *Statistical Interpolation of Spatial Data: Some Theory for Krigging*, New York: Springer.

86. Stern, H., and Cressie, N. (1999), Inference for Extremes in Disease Mapping, in A. Lawson, A. Biggeri, D. Böhning, E. Lesaffre, J.F. Viel, and R. Bertollini (Eds.), *Disease Mapping and Risk Assessment for Public Health*, Chichester, UK: Wiley, pp 31-47.

87. Stukel, D.M., and Rao, J.N.K. (1999), Small-Area Estimation Under Two-fold Nested Errors Regression Models, *Journal of Statistical Planning and Inference*, **78**, 131-147.

88. Tanner, M.A., and Wong, W.H. (1987), The calculation of posterior distributions by data augmentation (with discussion), *Journal of the American Statistical Association*, **82**, 528-550.

89. Tierney, L. (1994), Markov Chains for Exploring Posterior Distributions (with discussion), *Annals of Statistics*, **22**, 1701-1762.

90. You, Y. (1999), *Hierarchical Bayes and Related Methods for Model-Based Small Area Estimation*, Unpublished Ph.D. Thesis, Carleton University, Ottawa, Canada.

91. You, Y. (2006), Model-Based Small Area Unemployment Rate Estimation for the Canadian Labour Force Survey, Methodology Branch Working Paper, HSMD-2006-004E, Statistics Canada, Ottawa, Canada.

92. You, Y. (2008), An Integrated Modeling Approach to Unemployment Rate Estimation for Sub-Provincial Areas of Canada, *Survey Methodology*, **34**, 19-27.

93. You, Y., Rao, J.N.K., and Gambino, J. (2001), Model-Based Unemployment Rate Estimation for the Canadian Labour Force Survey: A Hierarchical Bayes Approach, Technical Report, Household Survey Methods Division, Statistics Canada.

94. You, Y., and Rao, J.N.K. (2002), Small Area Estimation Using Unmatched Sampling and Linking Models, *Canadian Journal of Statistics*, **30**, 3-15.

95. You, Y., Rao, J.N.K., and Gambino, J. (2003), Model-Based Unemployment Rate Estimation for the Canadian Labour Force Survey: a Hierarchical Bayes Approach, *Survey Methodology*, **29**, 25-32.

96. Xia, H., Carlin, B.P., and Waller, L.A. (1997), Hierarchical Models for Mapping Ohio Lung Cancer Rates, *Environmetrics*, **8**, 107-120.

97. Zhou, Q.M., and You, Y. (2008), Hierarchical Bayes Small Area Estimation for the Canadian Community Health Survey, in *Proceedings of the Survey Methods Section*, SSC Annual Meeting, May 2008.

Appendix A

Covariate Selection

1) Models (13) and (14), response variable = proportion of people who had a flu shot more than two years ago.

a) Regression results for the original nine covariates:

R-square = 0.402; R-square adjusted = 0.398.

	Unst coeff	*Unst. coeff.*	*St. coeff*	*St. coeff*	*St. coeff*	*95% CI*	*95% CI*	*Colline arity*	*Colline arity*
	B	St.err	Beta	T	Sig	Lower	Upper	Toleran ce	VIF
Const	0.1	0.0		11.9	2.5E-03	0.1	0.14		
x1	0.12	0.01	0.53	23.2	1E-101	0.11	0.13	0.81	1.23
x3	-0.1	0.01	-0.2	-4.8	1.7E-06	-0.1	-0.03	0.32	3.12
x5	0.09	0.02	0.17	6.11	1.3E-09	0.06	0.12	0.53	1.9
x6	-0.1	0.02	-0.1	-4.9	1.4E-06	-0.2	-0.1	0.67	1.49
x7	-0.1	0.02	-0.1	-3.7	0.0	-0.1	0.0	0.46	2.19
x2	0.06	0.02	0.1	4.01	6E-05	0.03	0.09	0.71	1.4
x9	-2E-07	7E-08	-0.1	-3.4	0.0	-4E-07	-1E-07	0.71	1.41
x8	-0.06	0.02	-0.1	-3.4	0.0	-0.09	-0.03	0.32	3.11
x4	0.05	0.02	0.1	3.2	0.0	0.02	0.08	0.23	4.39

where:

x1 = proportion of people who are single and have never been married;

x2 = proportion of visible minority population;

x3 = proportion of people with highest level of education less than a high school diploma;

x4 = proportion of people with group 1 labour occupations (occupations in Management, Business, Finance, Administration, Natural and Applied Sciences, Health, Social Sciences, Education, Religion, Art, Culture and Recreation);

x5 = proportion of people with group 3 labour occupations (occupations in Trades, Transport and Equipment Operator, occupations Unique to Primary Industry, Processing, Manufacturing and Utilities);

x6 = proportion of people with household income of 30-50K;

x7 = proportion of people with household income of 50-80K;

x8 = proportion of people with household income greater than 80K;

x9 = population estimates from 2001 Census.

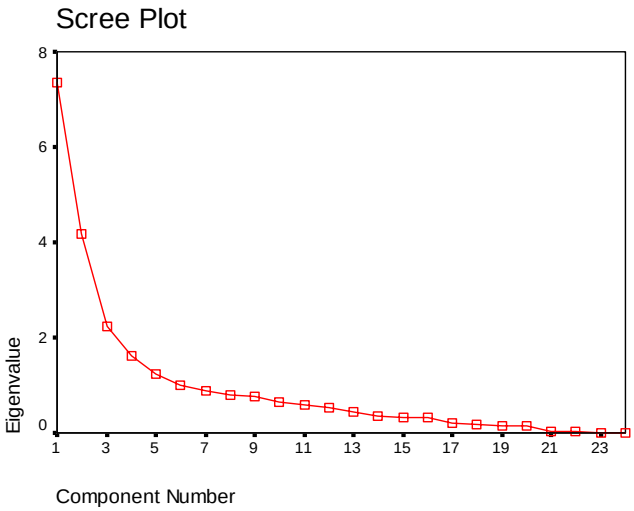
b) Principal component analysis

- Total variance explained:

	Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %
1	7.351	30.630	30.630
2	4.167	17.364	47.993
3	2.222	9.260	57.253
4	1.606	6.691	63.945
5	1.234	5.141	69.085
6	1.007	4.198	73.283

	Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %
1	7.080	29.498	29.498
2	3.783	15.761	45.259
3	2.182	9.092	54.351
4	2.099	8.748	63.098
5	1.356	5.648	68.746
6	1.089	4.537	73.283

Given little difference between extraction and rotation sums, it was decided to proceed with the original component matrix, not rotated.


Scree Plot
Eigenvalue
8
6
4
2
0
1
3
5
7
9
11
13
15
17
19
21
23
Component Number

- Component Matrix:

	Component					
	1	2	3	4	5	6
y1	0.453	0.023	0.385	0.366	0.521	0.148
y2	0.223	-0.827	-0.205	-0.190	-0.247	-0.027
y3	0.004	0.941	0.190	0.060	0.039	0.023
y4	-0.307	-0.251	-0.074	-0.141	0.080	-0.069
y5	-0.206	-0.337	0.751	-0.449	-0.003	0.153
y6	0.270	0.344	-0.704	0.452	-0.212	-0.035
y7	-0.213	-0.025	-0.156	-0.012	0.718	-0.393
y8	-0.861	0.130	-0.093	-0.273	0.059	0.144
y9	0.765	-0.423	-0.026	0.221	0.157	-0.004
y10	0.447	0.224	0.224	-0.043	-0.438	-0.265
y11	0.148	0.601	0.162	0.297	-0.117	-0.112
y12	-0.664	-0.163	-0.161	0.084	0.149	0.077
y13	0.796	0.194	-0.212	-0.262	0.011	-0.001
y14	-0.782	-0.366	0.020	0.189	-0.055	-0.004
y15	-0.081	-0.372	0.479	0.437	-0.121	0.112
y16	0.788	0.004	0.175	-0.108	0.039	0.002
y17	0.896	0.011	-0.050	0.052	0.135	0.019
y18	0.344	0.812	0.214	-0.065	0.009	0.075
y19	0.638	0.149	0.362	-0.162	-0.027	0.137
y20	-0.712	0.498	0.120	-0.055	0.028	0.089
y21	0.085	0.008	-0.325	0.134	0.055	0.798
y22	-0.205	-0.410	0.362	0.582	-0.160	-0.079
y23	0.712	-0.419	0.081	0.111	0.094	0.005
y24	0.768	-0.295	-0.230	-0.274	0.019	0.050

where:

y1 = proportion of people living Common law;

y2 = proportion of married people;

y3 = proportion of people who are single and have never been married;

y4 = proportion of people who are single (other);

y5 = proportion of white population;

y6 = proportion of visible minority population;

y7 = proportion of people who did not state their race;

y8 = proportion of people with highest level of education less than a high school diploma;

y9 = proportion of people with some post-secondary education;

y10 = proportion of people with highest level of education of a high school;

y11 = proportion of people with some post-high school education;

y12 = proportion of people with household income of less than 15K;

y13 = proportion of people with household income greater than 80K;

y14 = proportion of people with household income of 15-30K;

y15 = proportion of people with household income of 30-50K;

y16 = proportion of people with household income of 50-80K;

y17 = proportion of people with group 1 labour occupations (occupations in Management, Business, Finance, Administration, Natural and Applied Sciences, Health, Social Sciences, Education, Religion, Art, Culture and Recreation);

y18 = proportion of people with group 2 labour occupations (occupations in Sales and Services)

y19 = proportion of people with group 3 labour occupations (occupations in Trades, Transport and Equipment Operator, occupations Unique to Primary Industry, Processing, Manufacturing and Utilities);

y20 = proportion of people with personal income of less than 15K;

y21 = proportion of people with personal income greater than 80K;

y22 = proportion of people with personal income of 15-30K;

y23 = proportion of people with personal income of 30-50K;

y24 = proportion of people with personal income of 50-80K.

c) Regressions results for the selected four variables:

R-square = 0.369; R-square adjusted = 0.368.

	Unstand. Coefficients		*Standardized Coefficients*		
Data	B	Std. Error	Beta	t	Sig
(Const-ant)	0.112	0.014		8.057	1.6E-15
y17	0.022	0.013	0.071	2.051	0.041
y3	0.132	0.005	0.568	26.041	9E-123
y5	-0.026	0.013	-0.043	-1.954	0.051
y8	-0.049	0.011	-0.154	-4.468	8.5E-06

where:

y3 = proportion of people who are single and have never been married;

y17 = proportion of people with group 1 labour occupations (occupations in Management, Business, Finance, Administration, Natural and Applied Sciences, Health, Social Sciences, Education, Religion, Art, Culture and Recreation);

y5 = proportion of white population;

y8 = proportion of people with highest level of education less than a high school diploma.

2) Models (15) and (16), response variable = number of people who had a flu shot more than two years ago.

a) Forward linear regression with population estimates from 2001 Census treated as an entered variable

Selected variables:

- x1 = number of people who are single and have never been married
- x2 = number of people with some post-secondary education
- x3 = population estimates from 2001 Census

R-square = 0.840; R-square adjusted = 0.840.

	Unstand-d coeff.		*Stand-d coeff.*			*95% CI*		*Collinearity*	
Data	B	St. err	Beta	T	Sig.	Lower	Upper	Tolerance	VIF
Const	27.	44.1		0.62	0.54	-59.2	113.7		
x3	0.07	0.00	0.50	34.0	0.00	0.07	0.07	0.51	1.96
x2	0.36	0.02	0.31	15.9	0.00	0.32	0.41	0.29	3.44
x1	0.06	0.01	0.20	10.3	0.00	0.05	0.07	0.29	3.45

b) Forward linear regression

Selected variables:

- x4 = number of people with group 1 labour occupations;
- x5 = number of people who group 2 labour occupations;
- x2 = number of people with some post-secondary education;

- x6 = number of people with highest education level of a high school diploma.

R-square = 0.850; R-square adjusted = 0.850.

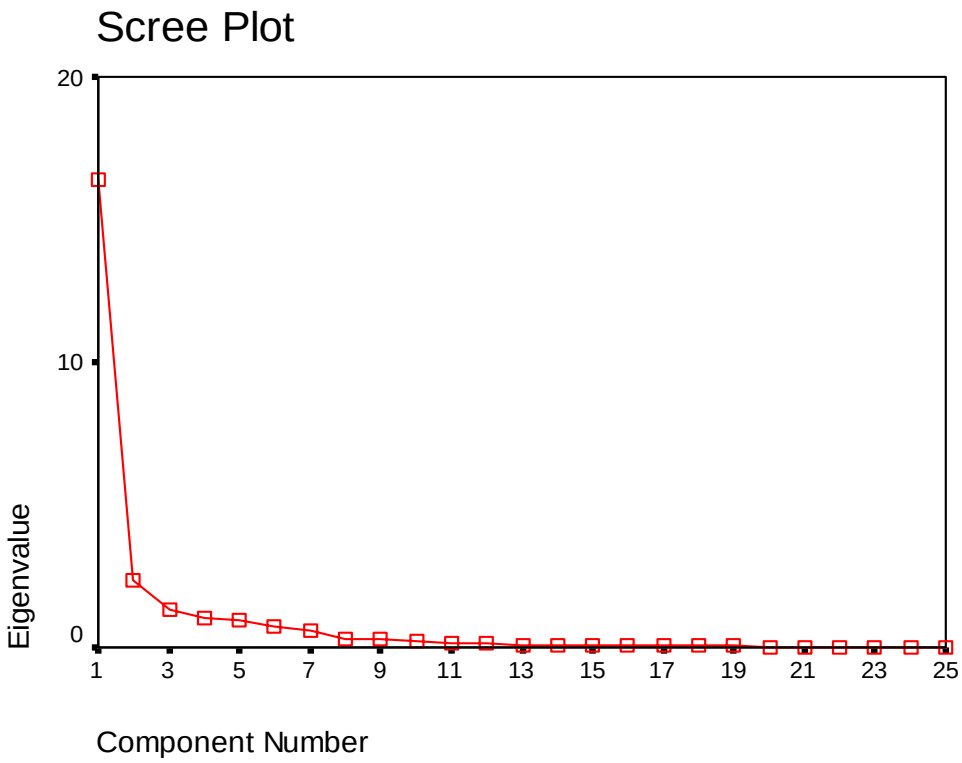
	Unstand-d coeff.		*Stand-d coeff.*			*95% CI*		*Collinearity*	
Data	B	Std. error	Beta	T	Sig.	Lower	Upper	Tolerance	VIF
Const	107.8	42.4		2.54	0.01	24.6	190.9		
x5	0.14	0.01	0.24	10.5	0.00	0.12	0.17	0.20	5.01
x6	0.15	0.02	0.20	8.98	0.00	0.11	0.18	0.22	4.60
x2	0.40	0.02	0.34	19.3	0.00	0.36	0.44	0.33	2.99
x4	0.07	0.01	0.25	12.8	0.00	0.06	0.08	0.27	3.67

c) Principal component analysis

- Total variance explained:

Extraction Sums of Squared Loadings			
Component	Total	% of Variance	Cumulative %
1	16.428	65.712	65.712
2	2.371	9.484	75.196
3	1.291	5.165	80.361
4	1.023	4.090	84.451

Rotation Sums of Squared Loadings			
Component	Total	% of Variance	Cumulative %
1	10.583	42.330	42.330
2	5.138	20.553	62.883
3	4.303	17.213	80.097
4	1.089	4.354	84.451


Scree Plot
20
10
0
Eigenvalue
1
3
5
7
9
11
13
15
17
19
21
23
25
Component Number

- Component Matrix:

	Component			
	1	2	3	4
y20	0.739	0.637	0.041	0.065
y3	0.676	0.628	-0.28	0.060
y8	0.545	0.528	0.333	0.108
y11	0.695	0.462	-0.303	0.097
y18	0.850	0.352	-0.291	0.096
y12	0.799	0.228	0.351	-0.083
y14	0.776	0.131	0.505	-0.151
y7	0.792	0.077	0.296	-0.212
y6	0.849	0.060	-0.131	0.151
y25	0.990	0.010	0.026	0.026
y15	0.947	0.003	0.154	-0.083
y10	0.907	-0.005	-0.148	0.146
y22	0.907	-0.024	0.154	-0.137
y5	0.958	-0.033	0.079	-0.013
y21	0.026	-0.113	-0.205	-0.268
y1	0.707	-0.158	0.058	-0.391
y16	0.950	-0.159	-0.083	0.025
y13	0.907	-0.162	-0.255	0.155
y19	0.872	-0.199	-0.231	0.024
y4	0.012	-0.227	0.383	0.754
y17	0.937	-0.248	-0.147	0.010
y9	0.940	-0.287	-0.011	-0.049
y23	0.903	-0.326	-0.027	-0.042
y2	0.822	-0.451	0.123	0.099
y24	0.810	-0.492	-0.085	0.079

where:

y1 = number of people living Common law;

y2 = number of Married people;

y3 = number of people who are single and have never been married;

y4 = number of people who are single (other);

y5 = number of people with race stated as “white";

y6 = number of people with race stated as “visible minorities";

y7 = number of people who did not state their race;

y8 = number of people with highest level of education less than a high school diploma;

y9 = number of people with some post-secondary education;

y10 = number of people with highest level of education of a high school;

y11 = number of people with some post-high school education;

y12 = number of people with household income of less than 15K;

y13 = number of people with household income greater than 80K;

y14 = number of people with household income of 15-30K;

y15 = number of people with household income of 30-50K;

y16 = number of people with household income of 50-80K;

y17 = number of people with group 1 labour occupations (occupations in Management, Business, Finance, Administration, Natural and Applied Sciences, Health, Social Sciences, Education, Religion, Art, Culture and Recreation);

y18 = number of people with group 2 labour occupations (occupations in Sales and Services)

y19 = number of people with group 3 labour occupations (occupations in Trades, Transport and Equipment Operator, occupations Unique to Primary Industry, Processing, Manufacturing and Utilities);

y20 = number of people with personal income of less than 15K;

y21 = number of people with personal income greater than 80K;

y22 = number of people with personal income of 15-30K;

y23 = number of people with personal income of 30-50K;

y24 = number of people with personal income of 50-80K;

y25 = population estimates from Census 2001.

- Rotated Component Matrix:

	Component			
	1	2	3	4
y1	0.579	0.098	0.475	-0.332
y2	0.891	0.010	0.308	0.126
y3	0.201	0.916	0.216	-0.091
y4	0.166	-0.154	-0.059	0.844
y5	0.739	0.384	0.483	0.006
y6	0.658	0.524	0.232	0.054
y7	0.480	0.260	0.682	-0.064
y8	0.052	0.543	0.585	0.243
y9	0.890	0.212	0.358	-0.070
y10	0.744	0.507	0.231	0.041
y11	0.323	0.816	0.153	-0.069
y12	0.400	0.383	0.713	0.077
y13	0.859	0.433	0.106	-0.003
y14	0.403	0.222	0.823	0.089
y15	0.686	0.360	0.572	-0.020
y16	0.842	0.358	0.310	-0.038
y17	0.896	0.307	0.238	-0.081
y18	0.507	0.798	0.205	-0.068
y19	0.835	0.350	0.153	-0.107
y20	0.184	0.823	0.492	0.065
y21	0.094	-0.050	-0.066	-0.335
y22	0.664	0.309	0.569	-0.067
y23	0.889	0.172	0.317	-0.070
y24	0.943	0.053	0.142	0.011
y25	0.751	0.461	0.452	0.015

where y variables are the same as for the Component Matrix.

- Extraction sum of squares - identified variables

Selected variables:

- x1 = number of people who are single and have never been married;

- x2= number of people with some post-secondary education;

- x5 = number of people with group 2 labour occupations;

- x7 = number of people with personal income 50-80K;

- x8 = number of people with personal income less than 15K.

R-square = 0.841; R-square adjusted = 0.841.

	Unstand-d coeff.		*Stand-d coeff.*			*95% CI*		*Collinearity*	
	B	Std. error	Beta	T	Sig.	Lower	Upper	Tolerance	VIF
Const	242.1	44.7		5.41	0.00	154.4	329.9		
x2	0.36	0.02	0.31	15.7	0.00	0.32	0.41	0.28	3.58
x1	0.11	0.01	0.38	10.9	0.00	0.09	0.12	0.09	10.8
x8	1.5E-02	0.01	0.03	1.16	0.25	-0.0	0.04	0.21	4.87
x7	0.28	0.01	0.37	22.0	0.00	0.25	0.30	0.39	2.51
x5	3.0E-02	0.02	0.05	1.35	0.18	-0.01	0.07	0.08	12.5

- Rotated sum of squares - identified variables

Selected variables:

- x1 = number of people who are single nad have never been married;

- x8 = number of people with personal income less than 15K;

- x3 = population estimates from 2001 Census;

- x9 = number of people with highest level of education less than a high school diploma.

R-square = 0.829; R-square adjusted = 0.828.

	Unstand-d coeff.	*Stand-d coeff.*	*95% CI*	*Collinearity*

	B	Std. error	Beta	T	Sig.	Lower	Upper	Tolerance	VIF
Const	215.1	47.3		4.55	0.00	122.3	307.8		
x3	8.6E-02	0.00	0.62	38.1	0.00	0.08	0.09	0.45	2.22
x1	0.120	0.01	0.43	17.5	0.00	0.11	0.13	0.2	5.08
x9	-0.10	0.01	-0.2	-10.2	0.00	-0.1	-0.1	0.34	2.91
x8	3.6E-02	0.02	0.07	1.95	0.05	0.00	0.07	0.11	9.32

Appendix B

WinBUGS code for the models under consideration

a) Model (13):

```
model{
  for( i in 1 : 14 )  {
    for( j in 1 : 103 ) {
     P[i , j] ~ dnorm(mu[i , j] , tau.c[i , j])
     mu[i , j] ~ dnorm(theta[i , j] , tau.v)
    theta[i , j] <- beta1× x1[i , j] + beta2 × x2[i , j] + beta3 × x3[i , j] + beta4 ×
x4[i , j] + psi[j]
     tau.c[i , j] <- 1 / ( tau1[j] × mu[i , j] × (1 - mu[i , j]))
                        }
                      }
      for( j in 1 : 103 ) {
       psi[j] ~ dnorm ( alpha.psi , sigma.psi)
                          }
  beta1 ~ dflat()
  beta2 ~ dflat()
  beta3 ~ dflat()
  beta4 ~ dflat()
  alpha.psi ~ dflat()
  sigma.psi ~ dgamma(0.001,0.001)
  tau.v ~ dgamma(0.001,0.001)
  sigma.v <- 1 / sqrt(tau.v)
          }
```

b) Model (14):

```
model{
  for( i in 1 : 14 )  {
    for( j in 1 : 103 ) {
       P[i , j] ~ dnorm(mu[i , j] , tau.c[i , j])
       mu[i , j] ~ dnorm(theta[i , j] , tau.v)
       theta[i , j] <- beta0 + beta1× x1[i , j] + psi[j]
       tau.c[i , j] <- 1 / ( tau1[j] × mu[i , j] × (1 - mu[i , j]))
                              }
                            }
     psi[1:103] ~ car.normal ( adj[],weights[],num[],tau )
     for( k in 1:sumNumNeigh ) {
           weights[k] <- 1
                                          }
   beta0 ~ dflat()
   beta1 ~ dflat()
   tau.v ~ dgamma(0.001,0.001)
   sigma.v <- 1 / sqrt(tau.v)
   tau  ~ dgamma(0.001, 0.001)
   sigma <- sqrt(1 / tau)
}
```

c) Model (15):

```
model{
 beta1 ~ dnorm(0 , 0.00001)
 for( i in 1 : 14 )  {
   for( j in 1 : 103 ) {
    Z[i , j] <- sqrt(Y[i , j])
    Z[i , j] ~ dpois(mu[i , j])
    log(mu[i , j]) <- log(E[i , j]) + theta[i , j]
    theta[i , j] <- beta1 × sqrt(x1[i , j]) + gamma[j]
                          }
                       }
    for( j in 1 : 103 )  {
      gamma[j] ~ dnorm(0 , 0.00001)
                              }
}
```

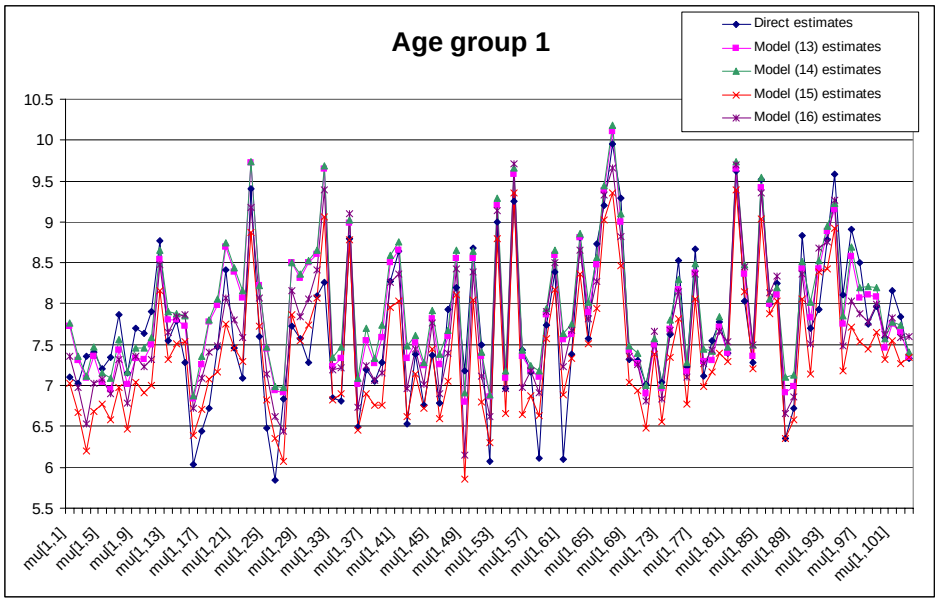
d) Model (16):

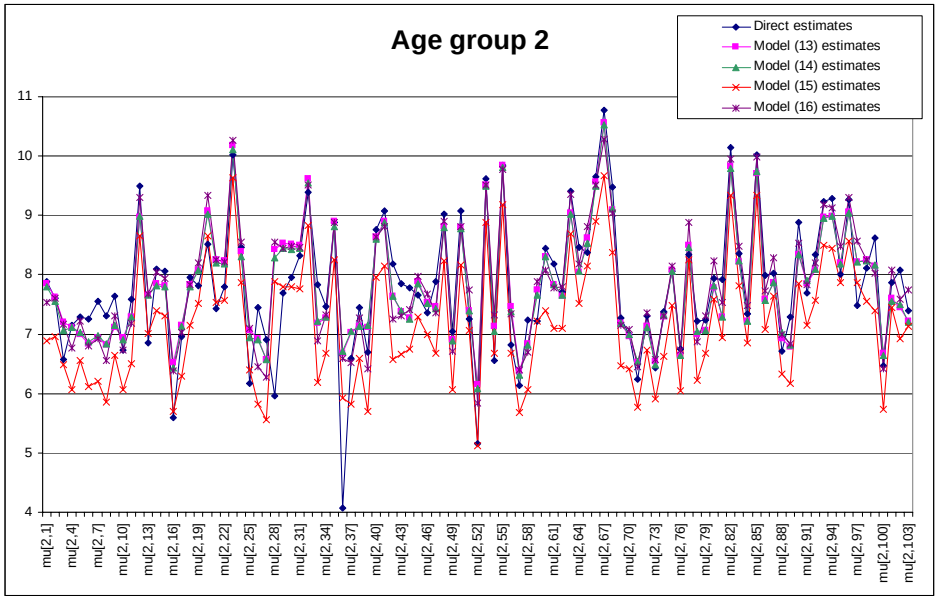
```
model{
  beta1 ~ dnorm(0 , 0.00001)
    for( i in 1 : 14 )  {
      for( j in 1 : 103 ) {
          Z[i , j] <- sqrt(Y[i , j])
          Z[i , j] ~ dpois(mu[i , j])
          log(mu[i , j]) <- log(E[i , j]) + theta[i , j]
          theta[i , j] <- beta1 × sqrt(x1[i , j]) + fi[i] + gamma[j]
                                }
                              }
     for( i in 1 : 14 )  {
           fi[i] ~ dnorm(0,0.00001)
                              }
     gamma[1:103] ~ car.normal(adj[],weights[],num[],tau)
          for(k in 1:sumNumNeigh) {
               weights[k] <- 1
                                          }
tau  ~ dgamma(0.001, 0.001)
sigma <- sqrt(1 / tau)
}
```

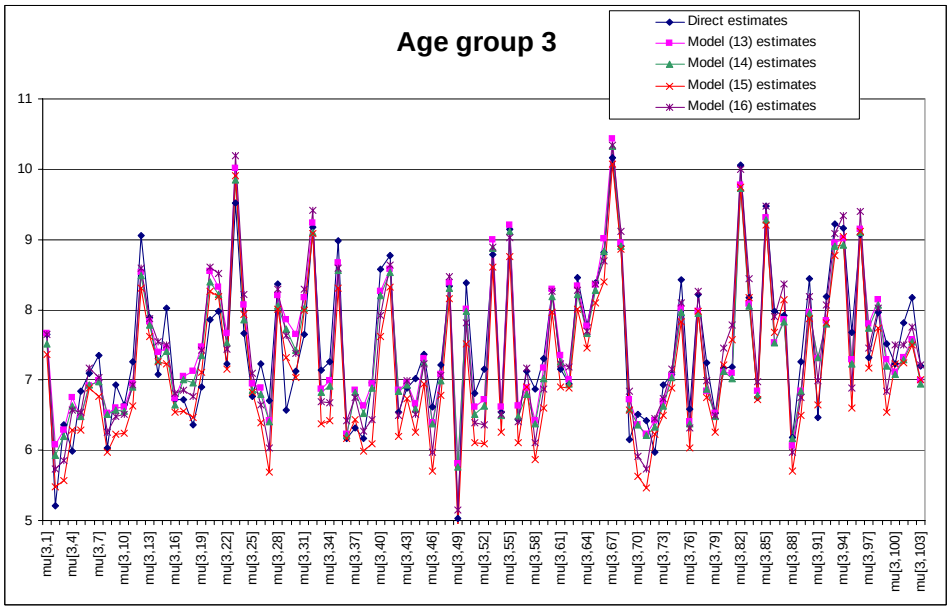
Appendix C

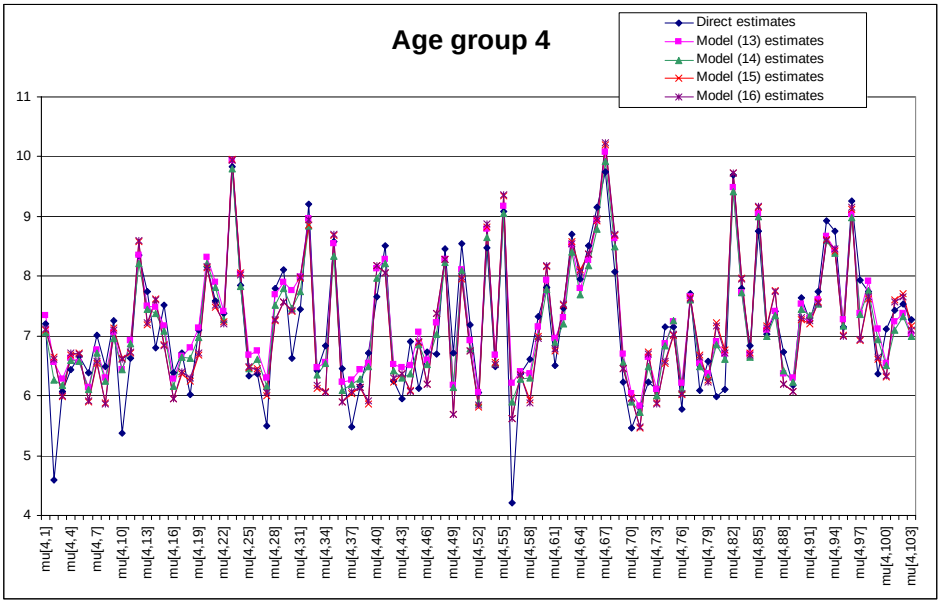
Model Estimates

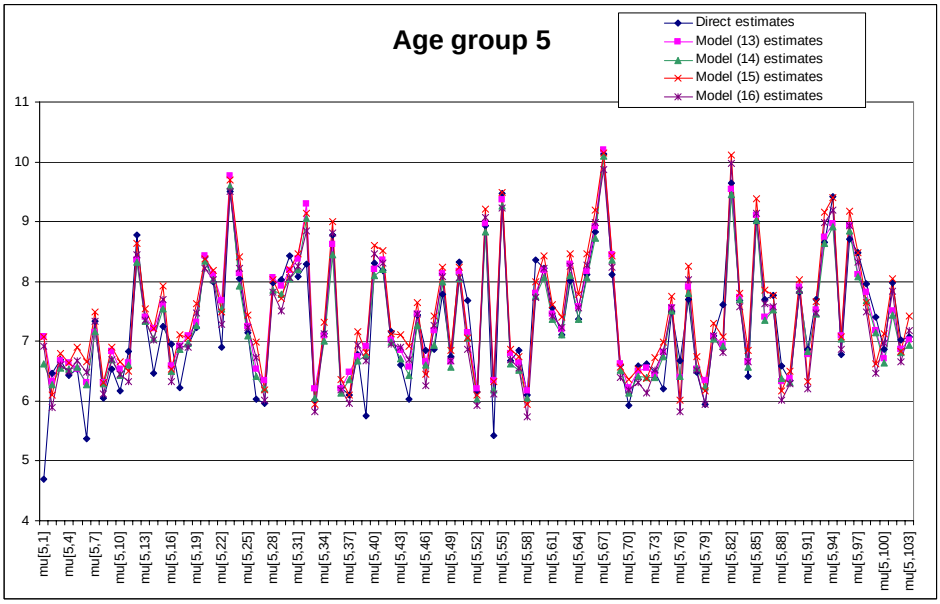
Model estimates for all four models are presented below, graphed by age group (estimates are graphed on a log-scale to allow for a better representation by reducing the variance in the extreme estimates for different domains):

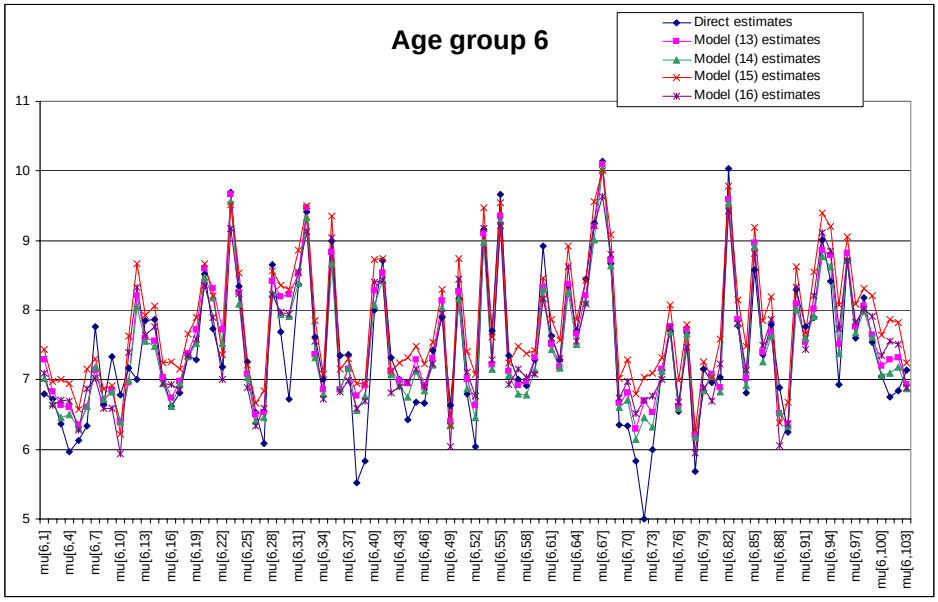


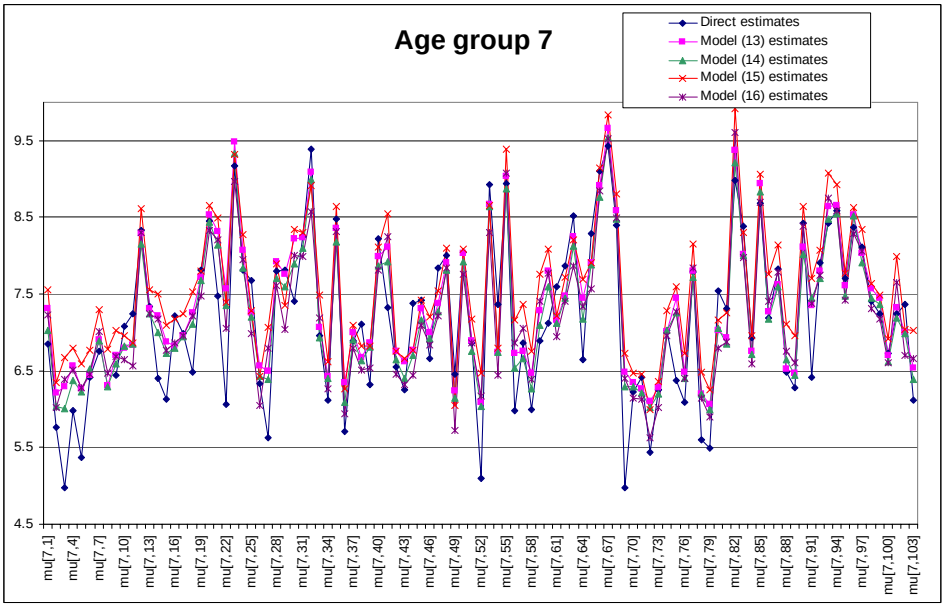

Age group 2
Direct estimates
Model (13) estimates
Model (14) estimates
Model (15) estimates
Model (16) estimates
11
10
9
8
7
6
5
4
mu[2,1]
mu[2,4]
mu[2,7]
mu[2,10]
mu[2,13]
mu[2,16]
mu[2,19]
mu[2,22]
mu[2,25]
mu[2,28]
mu[2,31]
mu[2,34]
mu[2,37]
mu[2,40]
mu[2,43]
mu[2,46]
mu[2,49]
mu[2,52]
mu[2,55]
mu[2,58]
mu[2,61]
mu[2,64]
mu[2,67]
mu[2,70]
mu[2,73]
mu[2,76]
mu[2,79]
mu[2,82]
mu[2,85]
mu[2,88]
mu[2,91]
mu[2,94]
mu[2,97]
mu[2,100]
mu[2,103]

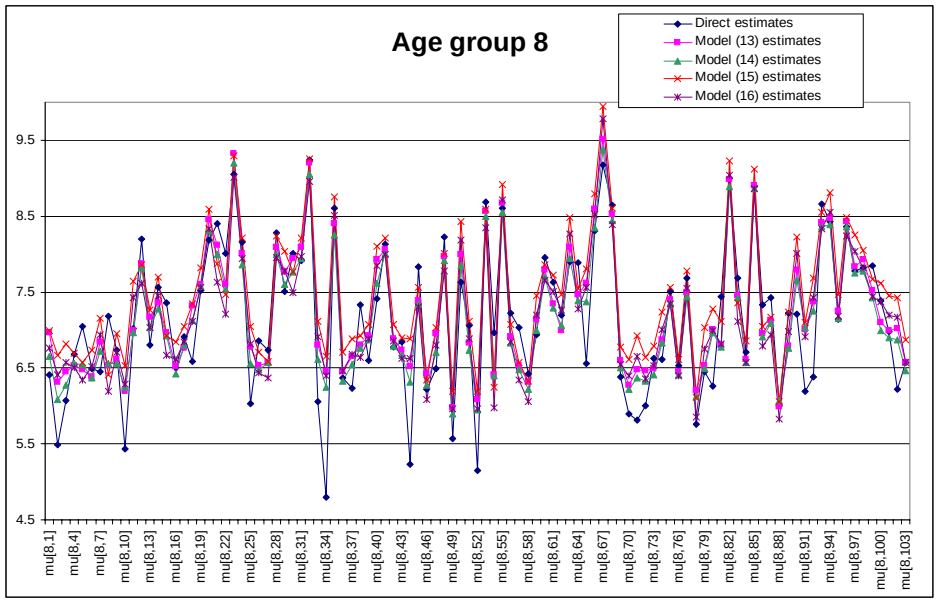

Age group 3
Direct estimates
Model (13) estimates
Model (14) estimates
Model (15) estimates
Model (16) estimates
11
10
9
8
7
6
5
mu[3,1]
mu[3,4]
mu[3,7]
mu[3,10]
mu[3,13]
mu[3,16]
mu[3,19]
mu[3,22]
mu[3,25]
mu[3,28]
mu[3,31]
mu[3,34]
mu[3,37]
mu[3,40]
mu[3,43]
mu[3,46]
mu[3,49]
mu[3,52]
mu[3,55]
mu[3,58]
mu[3,61]
mu[3,64]
mu[3,67]
mu[3,70]
mu[3,73]
mu[3,76]
mu[3,79]
mu[3,82]
mu[3,85]
mu[3,88]
mu[3,91]
mu[3,94]
mu[3,97]
mu[3,100]
mu[3,103]

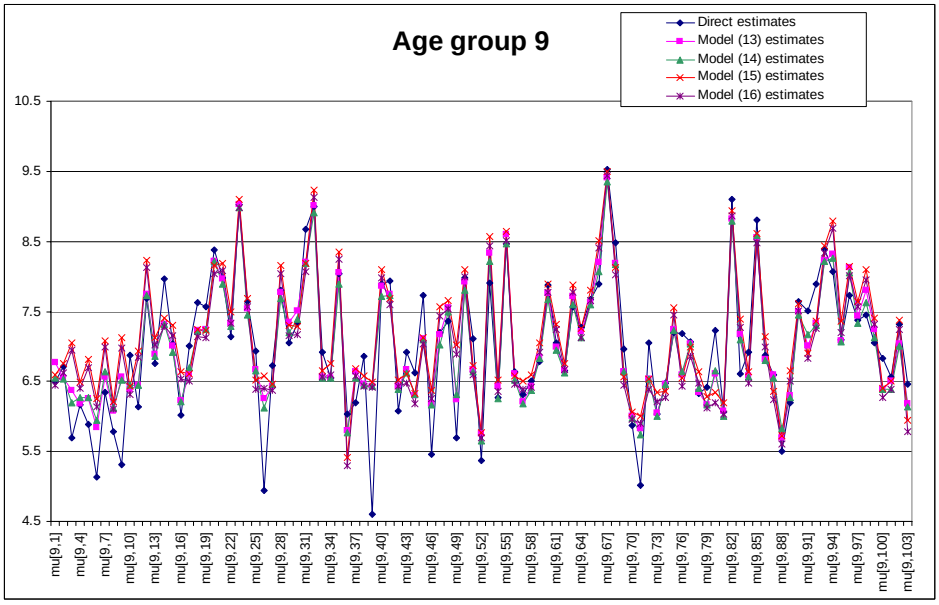

Age group 4
Direct estimates
Model (13) estimates
Model (14) estimates
Model (15) estimates
Model (16) estimates
11
10
9
8
7
6
5
4
mu[4,1]
mu[4,4]
mu[4,7]
mu[4,10]
mu[4,13]
mu[4,16]
mu[4,19]
mu[4,22]
mu[4,25]
mu[4,28]
mu[4,31]
mu[4,34]
mu[4,37]
mu[4,40]
mu[4,43]
mu[4,46]
mu[4,49]
mu[4,52]
mu[4,55]
mu[4,58]
mu[4,61]
mu[4,64]
mu[4,67]
mu[4,70]
mu[4,73]
mu[4,76]
mu[4,79]
mu[4,82]
mu[4,85]
mu[4,88]
mu[4,91]
mu[4,94]
mu[4,97]
mu[4,100]
mu[4,103]

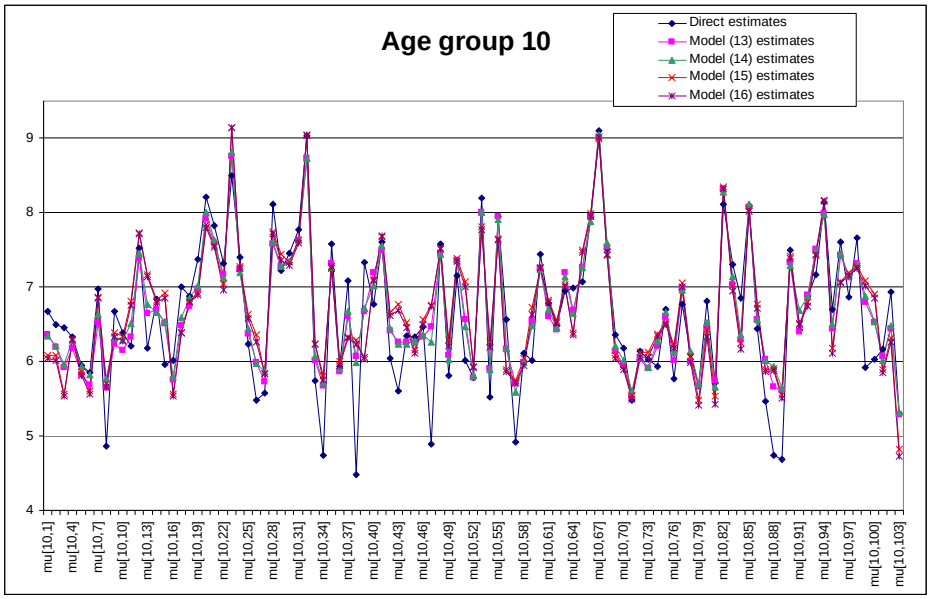

Age group 5
Direct estimates
Model (13) estimates
Model (14) estimates
Model (15) estimates
Model (16) estimates
11
10
9
8
7
6
5
4
mu[5,1]
mu[5,4]
mu[5,7]
mu[5,10]
mu[5,13]
mu[5,16]
mu[5,19]
mu[5,22]
mu[5,25]
mu[5,28]
mu[5,31]
mu[5,34]
mu[5,37]
mu[5,40]
mu[5,43]
mu[5,46]
mu[5,49]
mu[5,52]
mu[5,55]
mu[5,58]
mu[5,61]
mu[5,64]
mu[5,67]
mu[5,70]
mu[5,73]
mu[5,76]
mu[5,79]
mu[5,82]
mu[5,85]
mu[5,88]
mu[5,91]
mu[5,94]
mu[5,97]
mu[5,100]
mu[5,103]

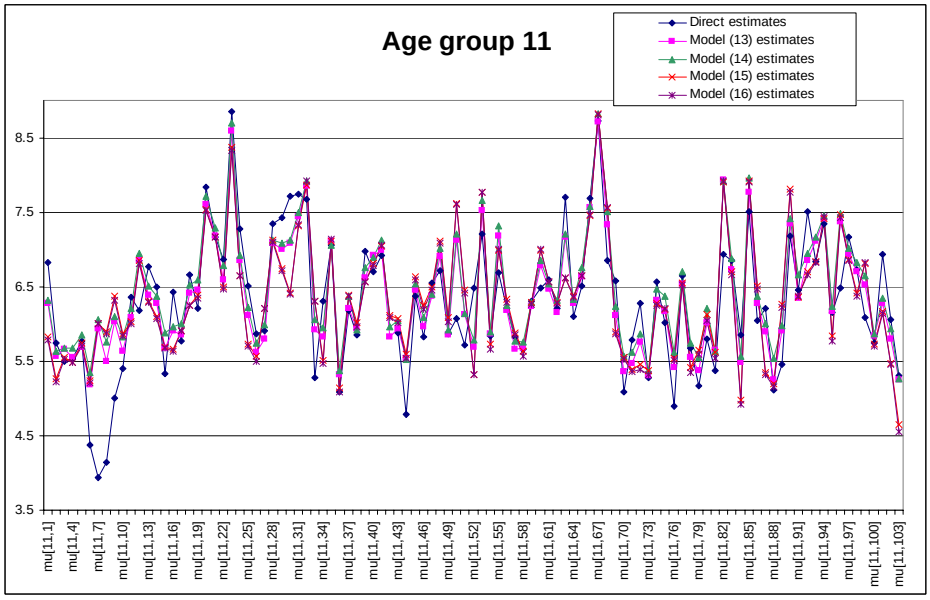

Age group 6
Direct estimates
Model (13) estimates
Model (14) estimates
Model (15) estimates
Model (16) estimates
11
10
9
8
7
6
5
mu[6,1]
mu[6,4]
mu[6,7]
mu[6,10]
mu[6,13]
mu[6,16]
mu[6,19]
mu[6,22]
mu[6,25]
mu[6,28]
mu[6,31]
mu[6,34]
mu[6,37]
mu[6,40]
mu[6,43]
mu[6,46]
mu[6,49]
mu[6,52]
mu[6,55]
mu[6,58]
mu[6,61]
mu[6,64]
mu[6,67]
mu[6,70]
mu[6,73]
mu[6,76]
mu[6,79]
mu[6,82]
mu[6,85]
mu[6,88]
mu[6,91]
mu[6,94]
mu[6,97]
mu[6,100]
mu[6,103]

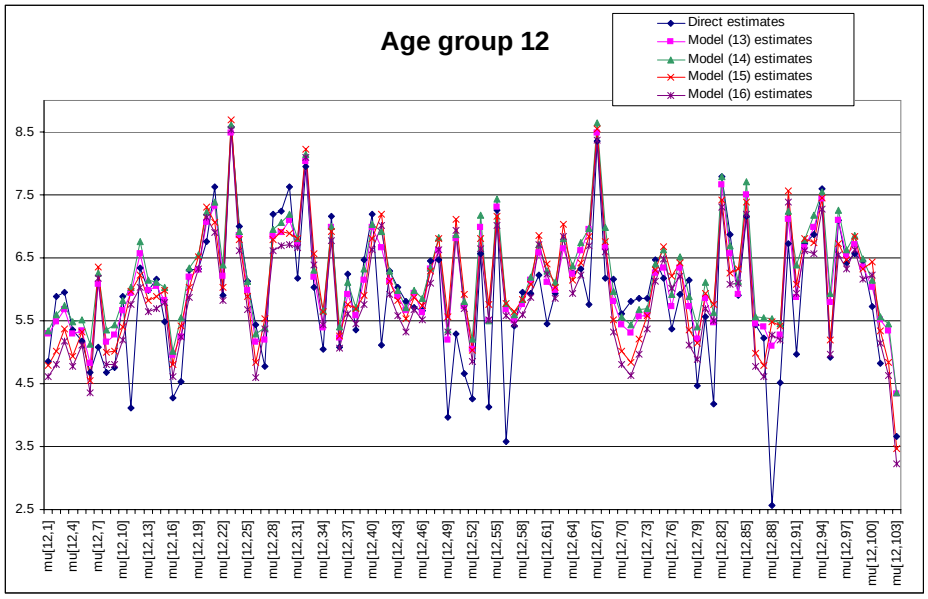

Age group 7
Direct estimates
Model (13) estimates
Model (14) estimates
Model (15) estimates
Model (16) estimates
9.5
8.5
7.5
6.5
5.5
4.5
mu[7,1]
mu[7,4]
mu[7,7]
mu[7,10]
mu[7,13]
mu[7,16]
mu[7,19]
mu[7,22]
mu[7,25]
mu[7,28]
mu[7,31]
mu[7,34]
mu[7,37]
mu[7,40]
mu[7,43]
mu[7,46]
mu[7,49]
mu[7,52]
mu[7,55]
mu[7,58]
mu[7,61]
mu[7,64]
mu[7,67]
mu[7,70]
mu[7,73]
mu[7,76]
mu[7,79]
mu[7,82]
mu[7,85]
mu[7,88]
mu[7,91]
mu[7,94]
mu[7,97]
mu[7,100]
mu[7,103]

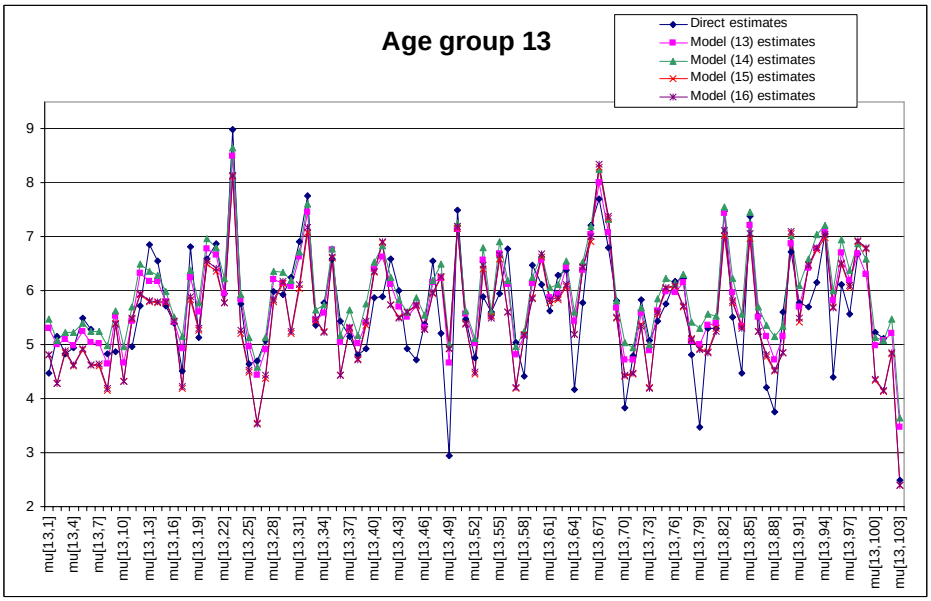

Age group 8
Direct estimates
Model (13) estimates
Model (14) estimates
Model (15) estimates
Model (16) estimates
9.5
8.5
7.5
6.5
5.5
4.5
mu[8,1]
mu[8,4]
mu[8,7]
mu[8,10]
mu[8,13]
mu[8,16]
mu[8,19]
mu[8,22]
mu[8,25]
mu[8,28]
mu[8,31]
mu[8,34]
mu[8,37]
mu[8,40]
mu[8,43]
mu[8,46]
mu[8,49]
mu[8,52]
mu[8,55]
mu[8,58]
mu[8,61]
mu[8,64]
mu[8,67]
mu[8,70]
mu[8,73]
mu[8,76]
mu[8,79]
mu[8,82]
mu[8,85]
mu[8,88]
mu[8,91]
mu[8,94]
mu[8,97]
mu[8,100]
mu[8,103]

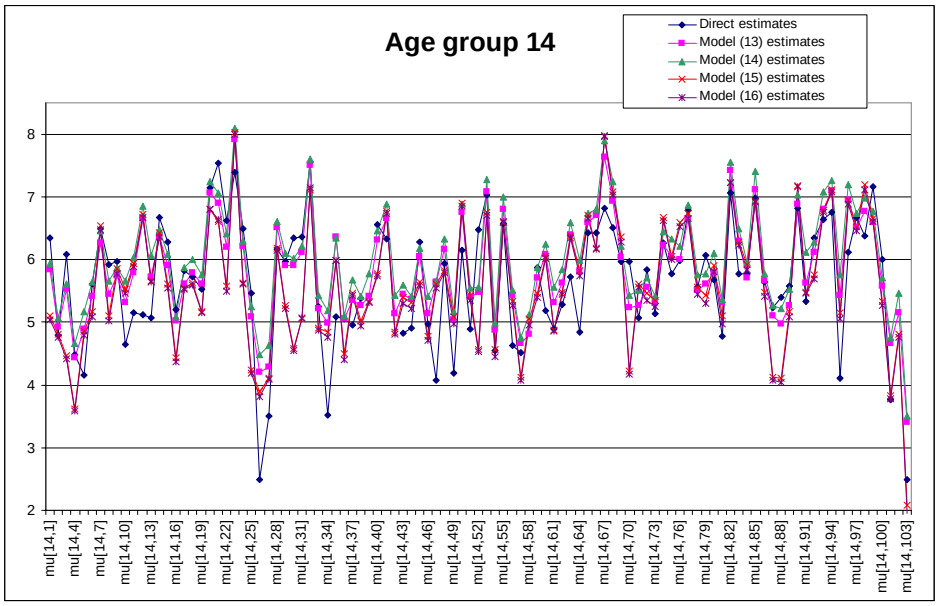

Age group 9
Direct estimates
Model (13) estimates
Model (14) estimates
Model (15) estimates
Model (16) estimates
10.5
9.5
8.5
7.5
6.5
5.5
4.5
mu[9,1]
mu[9,4]
mu[9,7]
mu[9,10]
mu[9,13]
mu[9,16]
mu[9,19]
mu[9,22]
mu[9,25]
mu[9,28]
mu[9,31]
mu[9,34]
mu[9,37]
mu[9,40]
mu[9,43]
mu[9,46]
mu[9,49]
mu[9,52]
mu[9,55]
mu[9,58]
mu[9,61]
mu[9,64]
mu[9,67]
mu[9,70]
mu[9,73]
mu[9,76]
mu[9,79]
mu[9,82]
mu[9,85]
mu[9,88]
mu[9,91]
mu[9,94]
mu[9,97]
mu[9,100]
mu[9,103]


Age group 10
Direct estimates
Model (13) estimates
Model (14) estimates
Model (15) estimates
Model (16) estimates
9
8
7
6
5
4
mu[10,1]
mu[10,4]
mu[10,7]
mu[10,10]
mu[10,13]
mu[10,16]
mu[10,19]
mu[10,22]
mu[10,25]
mu[10,28]
mu[10,31]
mu[10,34]
mu[10,37]
mu[10,40]
mu[10,43]
mu[10,46]
mu[10,49]
mu[10,52]
mu[10,55]
mu[10,58]
mu[10,61]
mu[10,64]
mu[10,67]
mu[10,70]
mu[10,73]
mu[10,76]
mu[10,79]
mu[10,82]
mu[10,85]
mu[10,88]
mu[10,91]
mu[10,94]
mu[10,97]
mu[10,100]
mu[10,103]


Age group 11
Direct estimates
Model (13) estimates
Model (14) estimates
Model (15) estimates
Model (16) estimates
8.5
7.5
6.5
5.5
4.5
3.5
mu[11,1]
mu[11,4]
mu[11,7]
mu[11,10]
mu[11,13]
mu[11,16]
mu[11,19]
mu[11,22]
mu[11,25]
mu[11,28]
mu[11,31]
mu[11,34]
mu[11,37]
mu[11,40]
mu[11,43]
mu[11,46]
mu[11,49]
mu[11,52]
mu[11,55]
mu[11,58]
mu[11,61]
mu[11,64]
mu[11,67]
mu[11,70]
mu[11,73]
mu[11,76]
mu[11,79]
mu[11,82]
mu[11,85]
mu[11,88]
mu[11,91]
mu[11,94]
mu[11,97]
mu[11,100]
mu[11,103]


Age group 12
Direct estimates
Model (13) estimates
Model (14) estimates
Model (15) estimates
Model (16) estimates
8.5
7.5
6.5
5.5
4.5
3.5
2.5
mu[12,1]
mu[12,4]
mu[12,7]
mu[12,10]
mu[12,13]
mu[12,16]
mu[12,19]
mu[12,22]
mu[12,25]
mu[12,28]
mu[12,31]
mu[12,34]
mu[12,37]
mu[12,40]
mu[12,43]
mu[12,46]
mu[12,49]
mu[12,52]
mu[12,55]
mu[12,58]
mu[12,61]
mu[12,64]
mu[12,67]
mu[12,70]
mu[12,73]
mu[12,76]
mu[12,79]
mu[12,82]
mu[12,85]
mu[12,88]
mu[12,91]
mu[12,94]
mu[12,97]
mu[12,100]
mu[12,103]

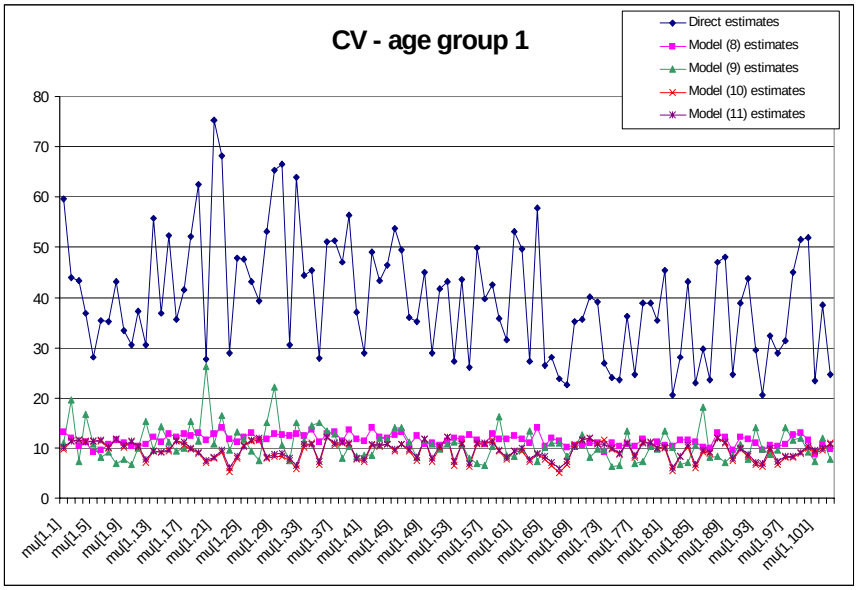

Age group 13
Direct estimates
Model (13) estimates
Model (14) estimates
Model (15) estimates
Model (16) estimates
9
8
7
6
5
4
3
2
mu[13,1]
mu[13,4]
mu[13,7]
mu[13,10]
mu[13,13]
mu[13,16]
mu[13,19]
mu[13,22]
mu[13,25]
mu[13,28]
mu[13,31]
mu[13,34]
mu[13,37]
mu[13,40]
mu[13,43]
mu[13,46]
mu[13,49]
mu[13,52]
mu[13,55]
mu[13,58]
mu[13,61]
mu[13,64]
mu[13,67]
mu[13,70]
mu[13,73]
mu[13,76]
mu[13,79]
mu[13,82]
mu[13,85]
mu[13,88]
mu[13,91]
mu[13,94]
mu[13,97]
mu[13,100]
mu[13,103]

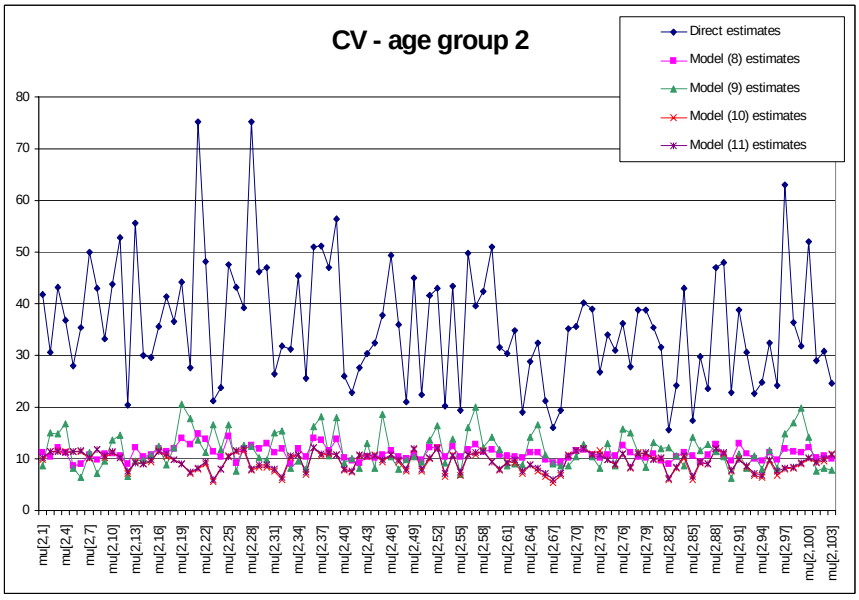

Age group 14
Direct estimates
Model (13) estimates
Model (14) estimates
Model (15) estimates
Model (16) estimates
8
7
6
5
4
3
2
mu[14,1]
mu[14,4]
mu[14,7]
mu[14,10]
mu[14,13]
mu[14,16]
mu[14,19]
mu[14,22]
mu[14,25]
mu[14,28]
mu[14,31]
mu[14,34]
mu[14,37]
mu[14,40]
mu[14,43]
mu[14,46]
mu[14,49]
mu[14,52]
mu[14,55]
mu[14,58]
mu[14,61]
mu[14,64]
mu[14,67]
mu[14,70]
mu[14,73]
mu[14,76]
mu[14,79]
mu[14,82]
mu[14,85]
mu[14,88]
mu[14,91]
mu[14,94]
mu[14,97]
mu[14,100]
mu[14,103]

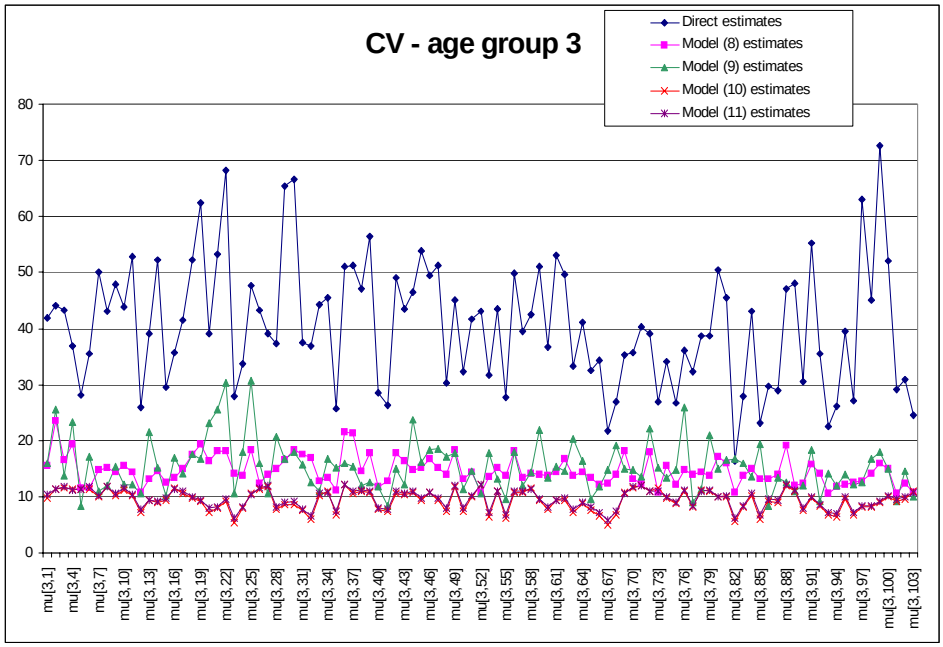
Appendix D

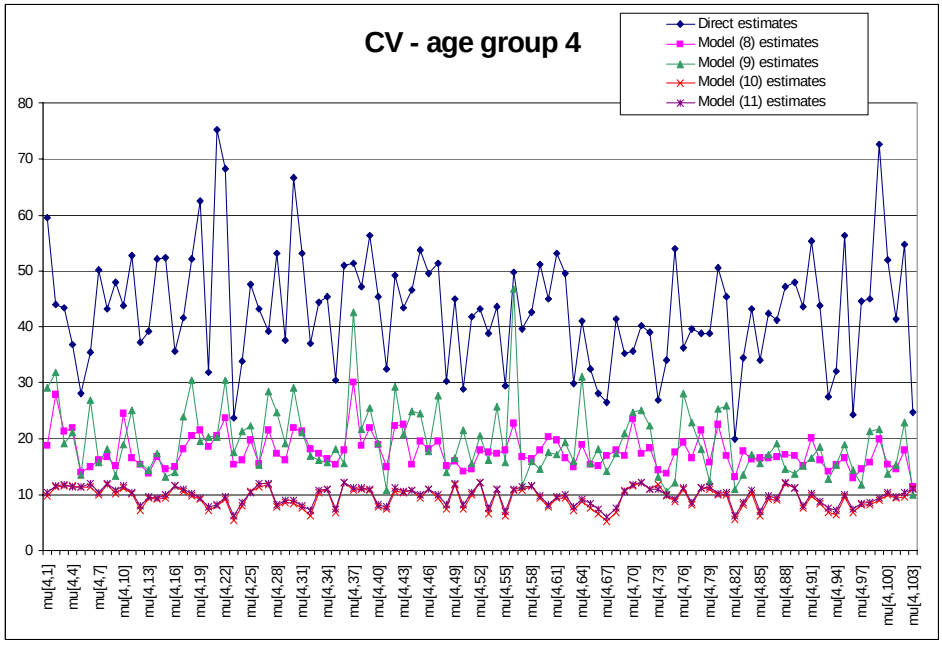
CV of Model Estimates

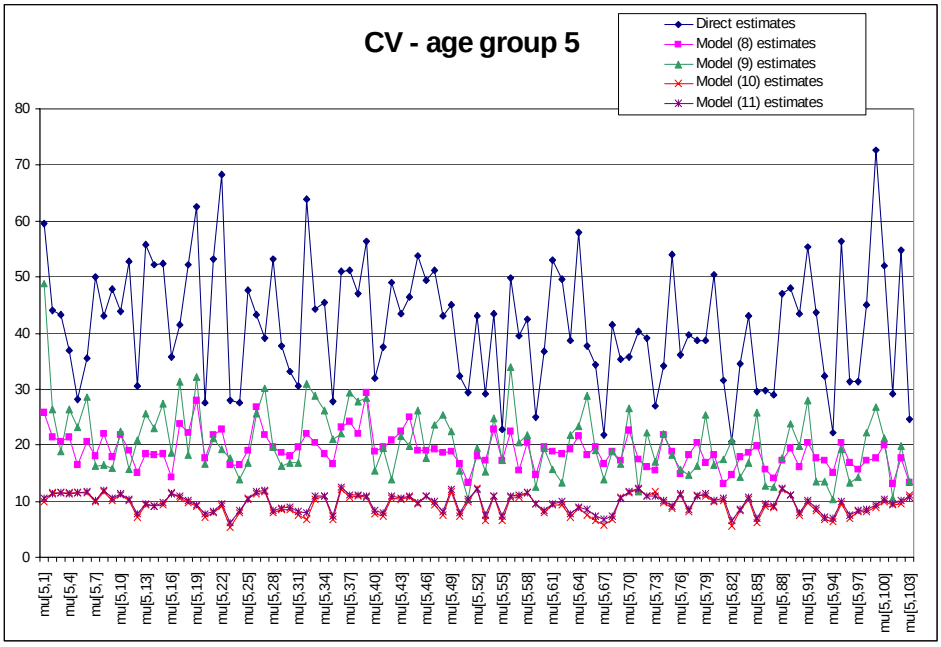
CV of model estimates from the four models are presented below, graphed by age group:

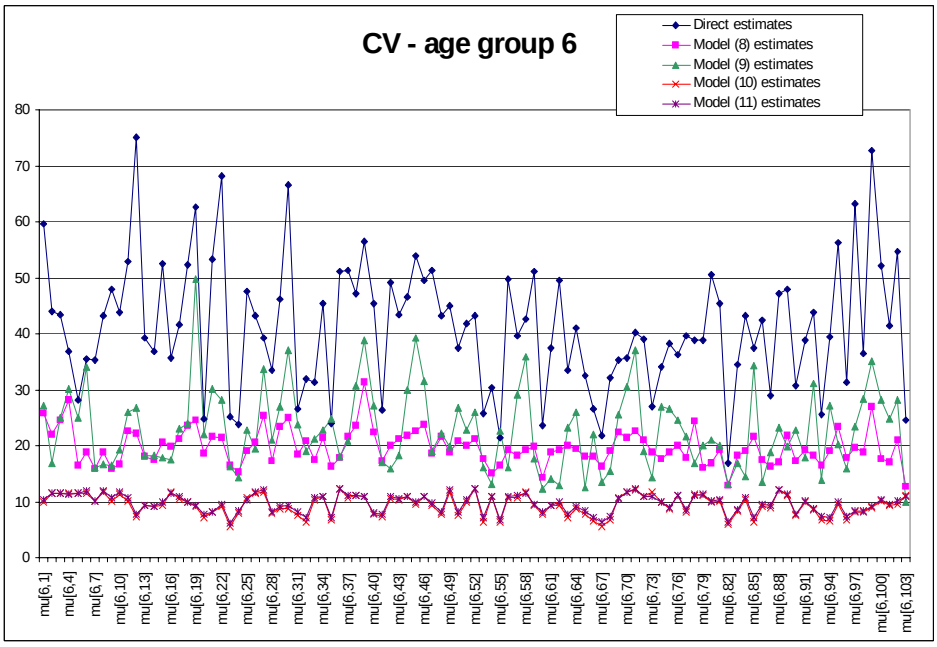


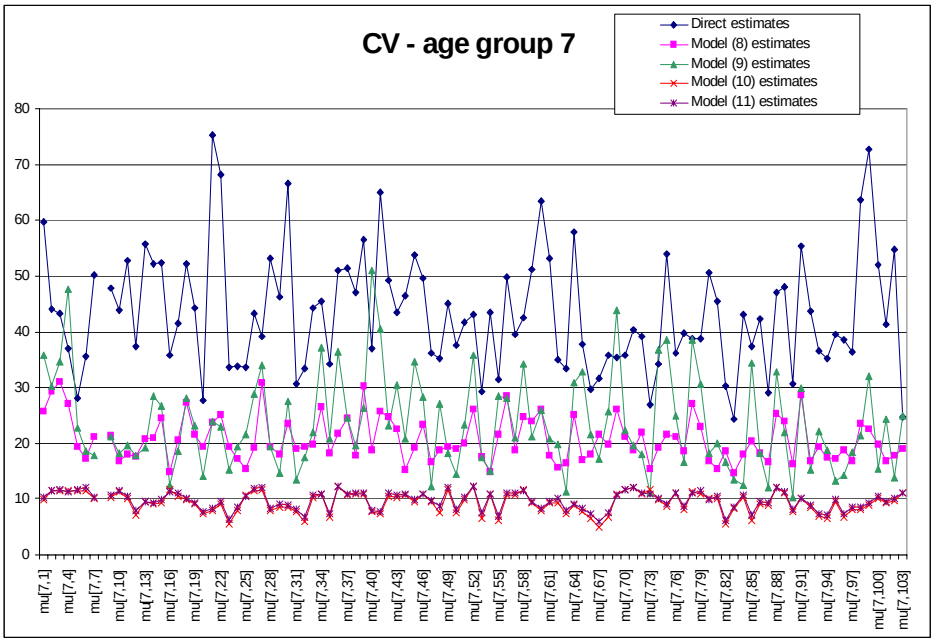


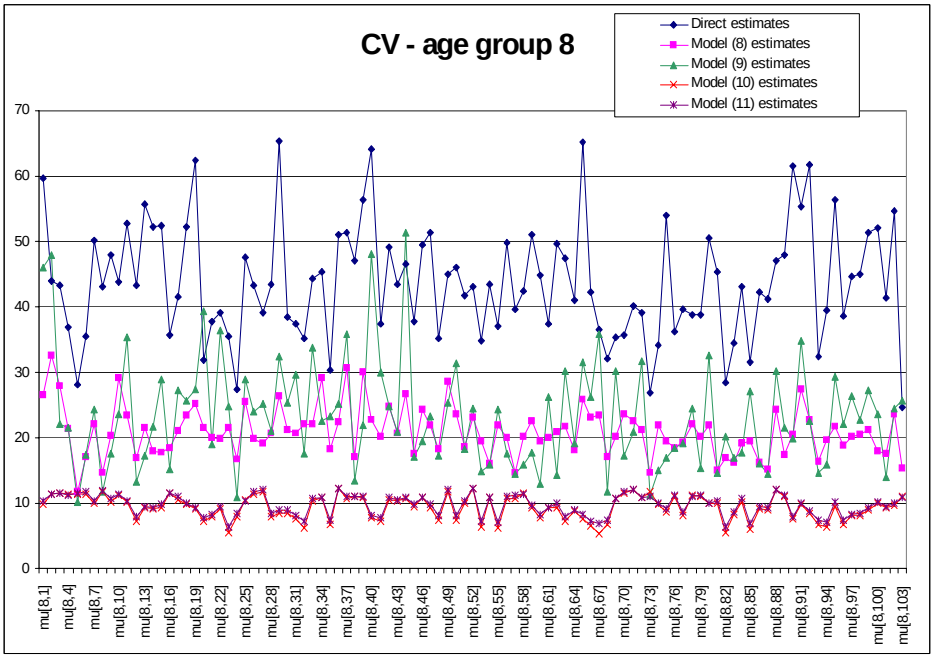

CV - age group 3
Direct estimates
Model (8) estimates
Model (9) estimates
Model (10) estimates
Model (11) estimates
0
10
20
30
40
50
60
70
80
mu[3,1]
mu[3,4]
mu[3,7]
mu[3,10]
mu[3,13]
mu[3,16]
mu[3,19]
mu[3,22]
mu[3,25]
mu[3,28]
mu[3,31]
mu[3,34]
mu[3,37]
mu[3,40]
mu[3,43]
mu[3,46]
mu[3,49]
mu[3,52]
mu[3,55]
mu[3,58]
mu[3,61]
mu[3,64]
mu[3,67]
mu[3,70]
mu[3,73]
mu[3,76]
mu[3,79]
mu[3,82]
mu[3,85]
mu[3,88]
mu[3,91]
mu[3,94]
mu[3,97]
mu[3,100]
mu[3,103]

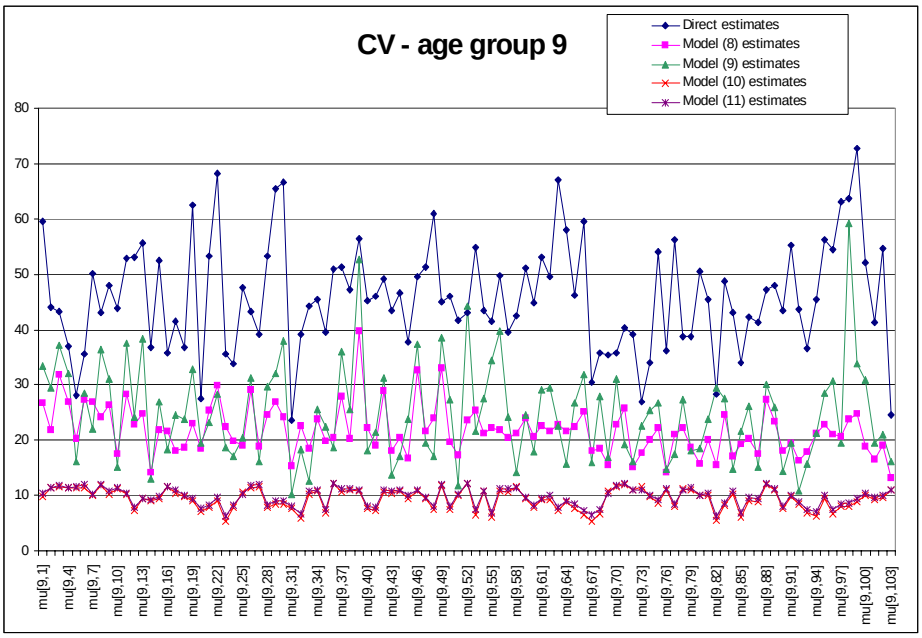

CV - age group 4
Direct estimates
Model (8) estimates
Model (9) estimates
Model (10) estimates
Model (11) estimates
0
10
20
30
40
50
60
70
80
mu[4,1]
mu[4,4]
mu[4,7]
mu[4,10]
mu[4,13]
mu[4,16]
mu[4,19]
mu[4,22]
mu[4,25]
mu[4,28]
mu[4,31]
mu[4,34]
mu[4,37]
mu[4,40]
mu[4,43]
mu[4,46]
mu[4,49]
mu[4,52]
mu[4,55]
mu[4,58]
mu[4,61]
mu[4,64]
mu[4,67]
mu[4,70]
mu[4,73]
mu[4,76]
mu[4,79]
mu[4,82]
mu[4,85]
mu[4,88]
mu[4,91]
mu[4,94]
mu[4,97]
mu[4,100]
mu[4,103]

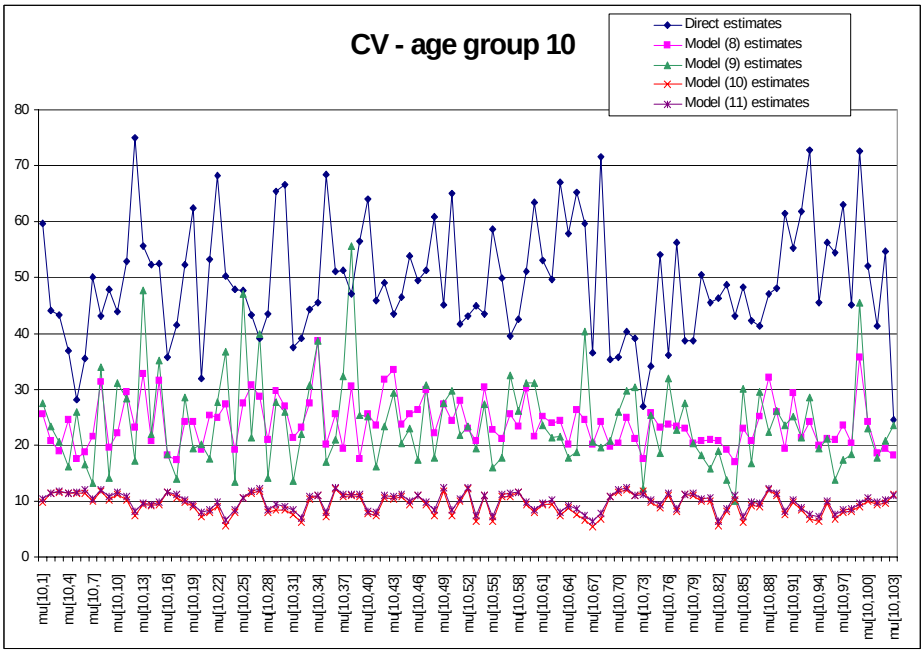

CV - age group 5
Direct estimates
Model (8) estimates
Model (9) estimates
Model (10) estimates
Model (11) estimates
0
10
20
30
40
50
60
70
80
mu[5.1]
mu[5.4]
mu[5.7]
mu[5.10]
mu[5.13]
mu[5.16]
mu[5.19]
mu[5.22]
mu[5.25]
mu[5.28]
mu[5.31]
mu[5.34]
mu[5.37]
mu[5.40]
mu[5.43]
mu[5.46]
mu[5.49]
mu[5.52]
mu[5.55]
mu[5.58]
mu[5.61]
mu[5.64]
mu[5.67]
mu[5.70]
mu[5.73]
mu[5.76]
mu[5.79]
mu[5.82]
mu[5.85]
mu[5.88]
mu[5.91]
mu[5.94]
mu[5.97]
mu[5.100]
mu[5.103]

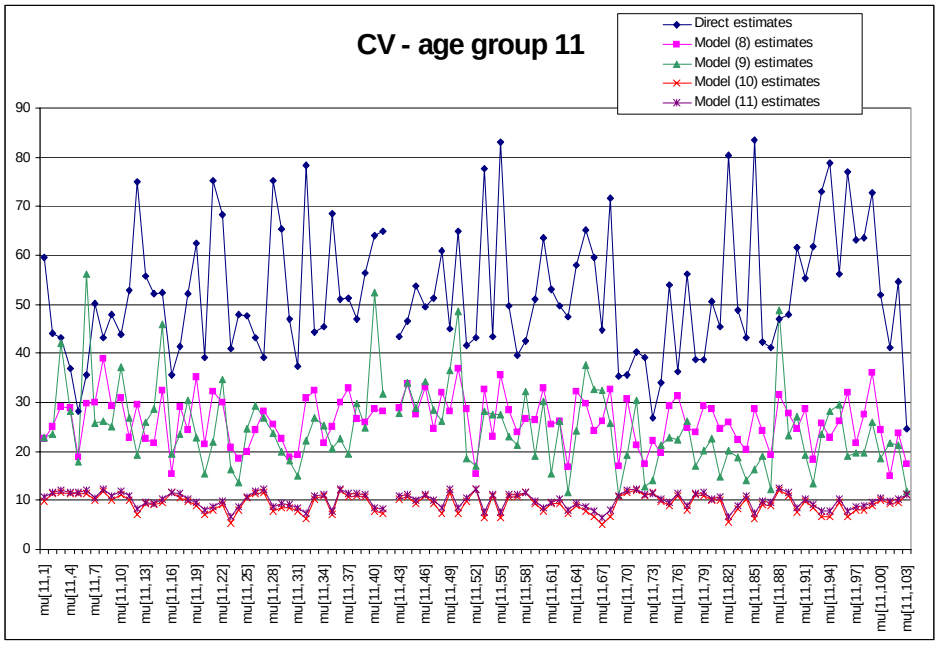

CV - age group 6
Direct estimates
Model (8) estimates
Model (9) estimates
Model (10) estimates
Model (11) estimates
0
10
20
30
40
50
60
70
80
mu[6.1]
mu[6.4]
mu[6.7]
mu[6.10]
mu[6.13]
mu[6.16]
mu[6.19]
mu[6.22]
mu[6.25]
mu[6.28]
mu[6.31]
mu[6.34]
mu[6.37]
mu[6.40]
mu[6.43]
mu[6.46]
mu[6.49]
mu[6.52]
mu[6.55]
mu[6.58]
mu[6.61]
mu[6.64]
mu[6.67]
mu[6.70]
mu[6.73]
mu[6.76]
mu[6.79]
mu[6.82]
mu[6.85]
mu[6.88]
mu[6.91]
mu[6.94]
mu[6.97]
mu[6.100]
mu[6.103]

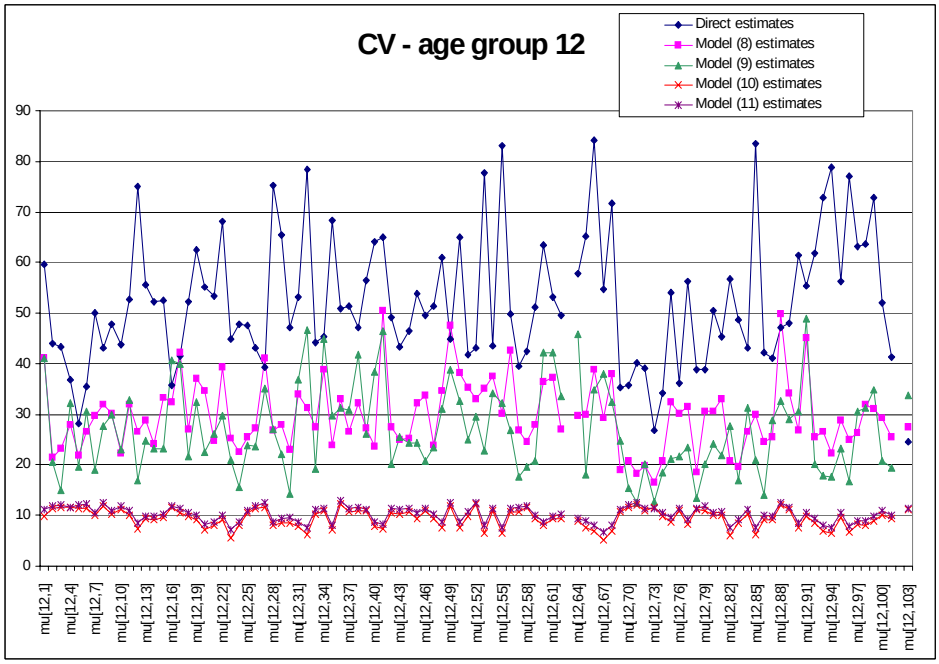

CV - age group 7
Direct estimates
Model (8) estimates
Model (9) estimates
Model (10) estimates
Model (11) estimates
80
70
60
50
40
30
20
10
0
mu[7,1]
mu[7,4]
mu[7,7]
mu[7,10]
mu[7,13]
mu[7,16]
mu[7,19]
mu[7,22]
mu[7,25]
mu[7,28]
mu[7,31]
mu[7,34]
mu[7,37]
mu[7,40]
mu[7,43]
mu[7,46]
mu[7,49]
mu[7,52]
mu[7,55]
mu[7,58]
mu[7,61]
mu[7,64]
mu[7,67]
mu[7,70]
mu[7,73]
mu[7,76]
mu[7,79]
mu[7,82]
mu[7,85]
mu[7,88]
mu[7,91]
mu[7,94]
mu[7,97]
mu[7,100]
mu[7,103]

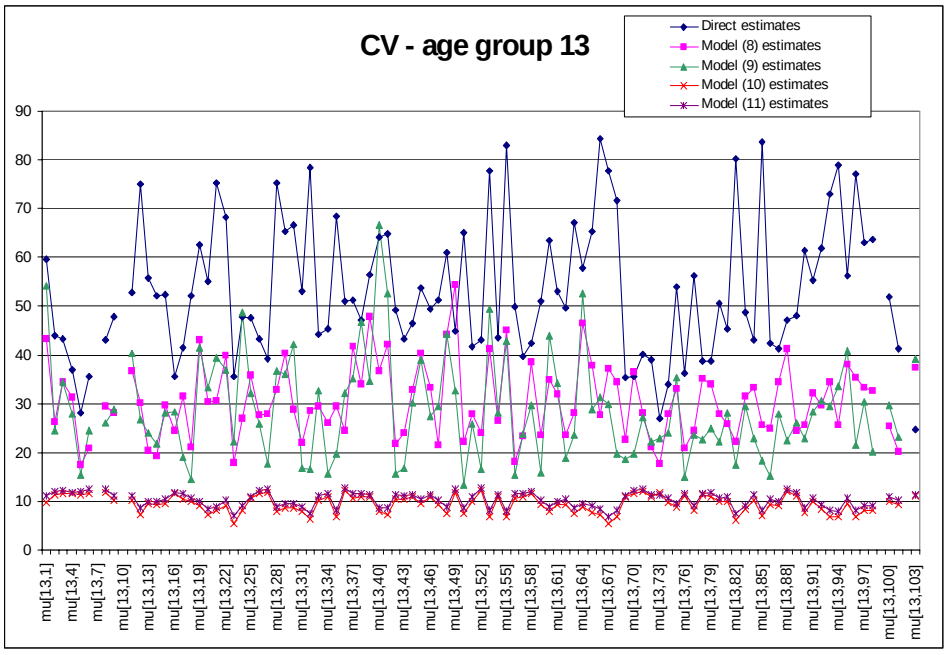

CV - age group 8
Direct estimates
Model (8) estimates
Model (9) estimates
Model (10) estimates
Model (11) estimates
70
60
50
40
30
20
10
0
mu[8,1]
mu[8,4]
mu[8,7]
mu[8,10]
mu[8,13]
mu[8,16]
mu[8,19]
mu[8,22]
mu[8,25]
mu[8,28]
mu[8,31]
mu[8,34]
mu[8,37]
mu[8,40]
mu[8,43]
mu[8,46]
mu[8,49]
mu[8,52]
mu[8,55]
mu[8,58]
mu[8,61]
mu[8,64]
mu[8,67]
mu[8,70]
mu[8,73]
mu[8,76]
mu[8,79]
mu[8,82]
mu[8,85]
mu[8,88]
mu[8,91]
mu[8,94]
mu[8,97]
mu[8,100]
mu[8,103]

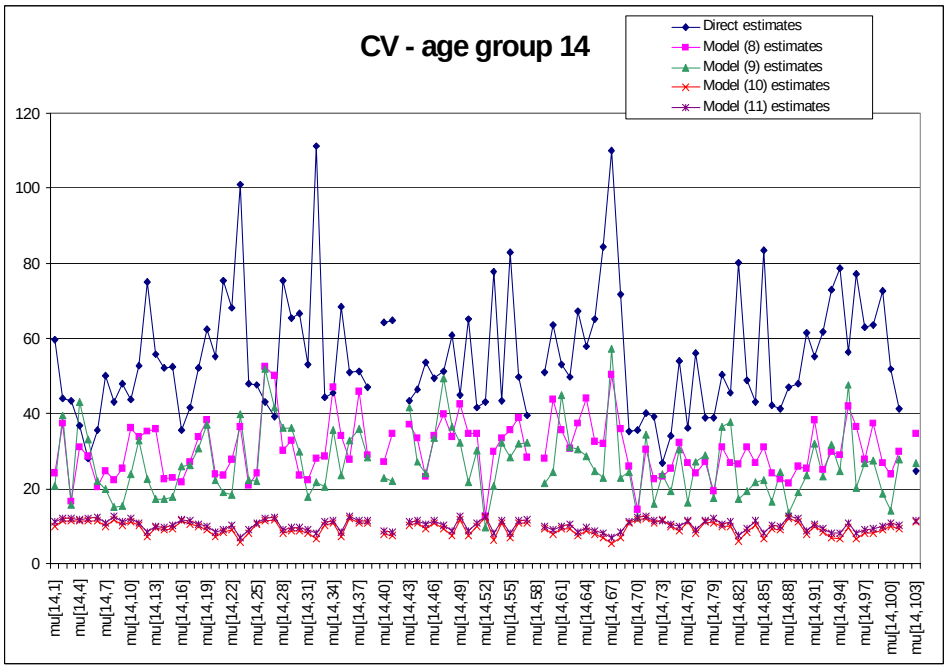

CV - age group 9
Direct estimates
Model (8) estimates
Model (9) estimates
Model (10) estimates
Model (11) estimates
80
70
60
50
40
30
20
10
0
mu[9,1]
mu[9,4]
mu[9,7]
mu[9,10]
mu[9,13]
mu[9,16]
mu[9,19]
mu[9,22]
mu[9,25]
mu[9,28]
mu[9,31]
mu[9,34]
mu[9,37]
mu[9,40]
mu[9,43]
mu[9,46]
mu[9,49]
mu[9,52]
mu[9,55]
mu[9,58]
mu[9,61]
mu[9,64]
mu[9,67]
mu[9,70]
mu[9,73]
mu[9,76]
mu[9,79]
mu[9,82]
mu[9,85]
mu[9,88]
mu[9,91]
mu[9,94]
mu[9,97]
mu[9,100]
mu[9,103]


CV - age group 10
Direct estimates
Model (8) estimates
Model (9) estimates
Model (10) estimates
Model (11) estimates
80
70
60
50
40
30
20
10
0
mu[10,1]
mu[10,4]
mu[10,7]
mu[10,10]
mu[10,13]
mu[10,16]
mu[10,19]
mu[10,22]
mu[10,25]
mu[10,28]
mu[10,31]
mu[10,34]
mu[10,37]
mu[10,40]
mu[10,43]
mu[10,46]
mu[10,49]
mu[10,52]
mu[10,55]
mu[10,58]
mu[10,61]
mu[10,64]
mu[10,67]
mu[10,70]
mu[10,73]
mu[10,76]
mu[10,79]
mu[10,82]
mu[10,85]
mu[10,88]
mu[10,91]
mu[10,94]
mu[10,97]
mu[10,100]
mu[10,103]


CV - age group 11
Direct estimates
Model (8) estimates
Model (9) estimates
Model (10) estimates
Model (11) estimates
90
80
70
60
50
40
30
20
10
0
mu[11,1]
mu[11,4]
mu[11,7]
mu[11,10]
mu[11,13]
mu[11,16]
mu[11,19]
mu[11,22]
mu[11,25]
mu[11,28]
mu[11,31]
mu[11,34]
mu[11,37]
mu[11,40]
mu[11,43]
mu[11,46]
mu[11,49]
mu[11,52]
mu[11,55]
mu[11,58]
mu[11,61]
mu[11,64]
mu[11,67]
mu[11,70]
mu[11,73]
mu[11,76]
mu[11,79]
mu[11,82]
mu[11,85]
mu[11,88]
mu[11,91]
mu[11,94]
mu[11,97]
mu[11,100]
mu[11,103]


CV - age group 12
Direct estimates
Model (8) estimates
Model (9) estimates
Model (10) estimates
Model (11) estimates
90
80
70
60
50
40
30
20
10
0
mu[12,1]
mu[12,4]
mu[12,7]
mu[12,10]
mu[12,13]
mu[12,16]
mu[12,19]
mu[12,22]
mu[12,25]
mu[12,28]
mu[12,31]
mu[12,34]
mu[12,37]
mu[12,40]
mu[12,43]
mu[12,46]
mu[12,49]
mu[12,52]
mu[12,55]
mu[12,58]
mu[12,61]
mu[12,64]
mu[12,67]
mu[12,70]
mu[12,73]
mu[12,76]
mu[12,79]
mu[12,82]
mu[12,85]
mu[12,88]
mu[12,91]
mu[12,94]
mu[12,97]
mu[12,100]
mu[12,103]

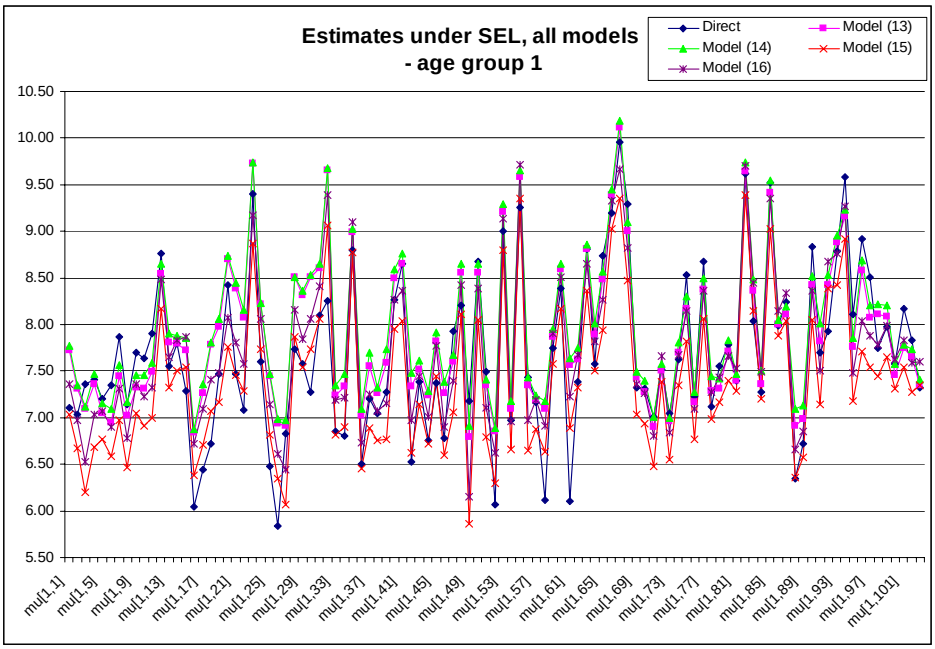

CV - age group 13
Direct estimates
Model (8) estimates
Model (9) estimates
Model (10) estimates
Model (11) estimates
90
80
70
60
50
40
30
20
10
0
mu[13,1]
mu[13,4]
mu[13,7]
mu[13,10]
mu[13,13]
mu[13,16]
mu[13,19]
mu[13,22]
mu[13,25]
mu[13,28]
mu[13,31]
mu[13,34]
mu[13,37]
mu[13,40]
mu[13,43]
mu[13,46]
mu[13,49]
mu[13,52]
mu[13,55]
mu[13,58]
mu[13,61]
mu[13,64]
mu[13,67]
mu[13,70]
mu[13,73]
mu[13,76]
mu[13,79]
mu[13,82]
mu[13,85]
mu[13,88]
mu[13,91]
mu[13,94]
mu[13,97]
mu[13,100]
mu[13,103]

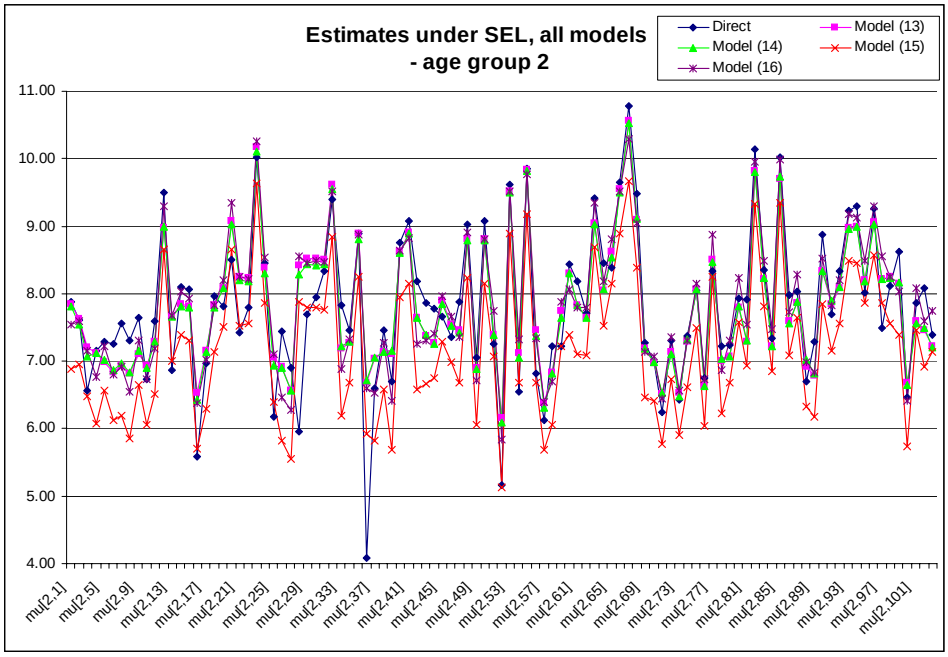

CV - age group 14
Direct estimates
Model (8) estimates
Model (9) estimates
Model (10) estimates
Model (11) estimates
120
100
80
60
40
20
0
mu[14,1]
mu[14,4]
mu[14,7]
mu[14,10]
mu[14,13]
mu[14,16]
mu[14,19]
mu[14,22]
mu[14,25]
mu[14,28]
mu[14,31]
mu[14,34]
mu[14,37]
mu[14,40]
mu[14,43]
mu[14,46]
mu[14,49]
mu[14,52]
mu[14,55]
mu[14,58]
mu[14,61]
mu[14,64]
mu[14,67]
mu[14,70]
mu[14,73]
mu[14,76]
mu[14,79]
mu[14,82]
mu[14,85]
mu[14,88]
mu[14,91]
mu[14,94]
mu[14,97]
mu[14,100]
mu[14,103]

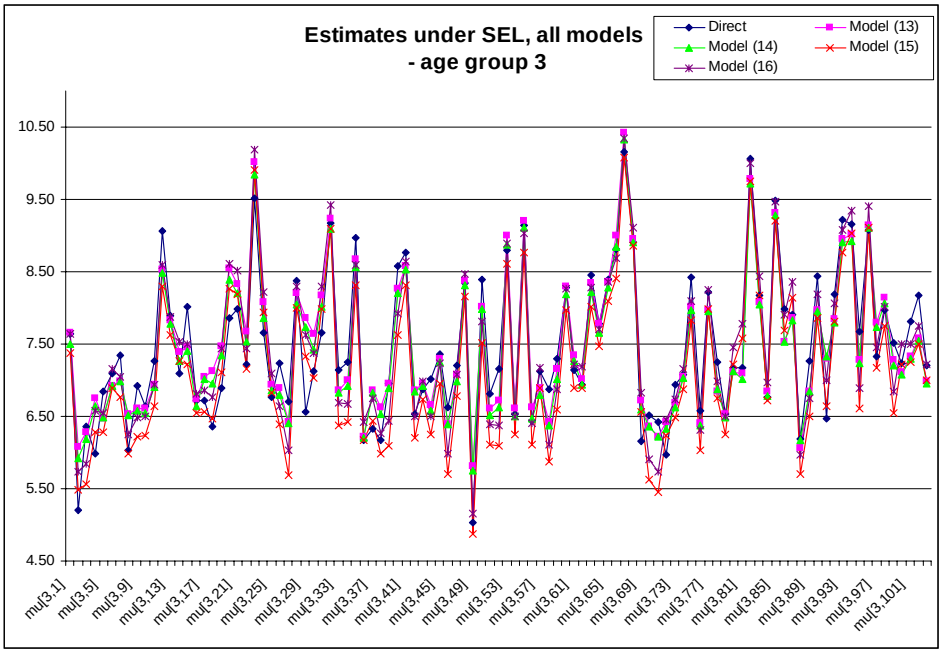
Appendix E

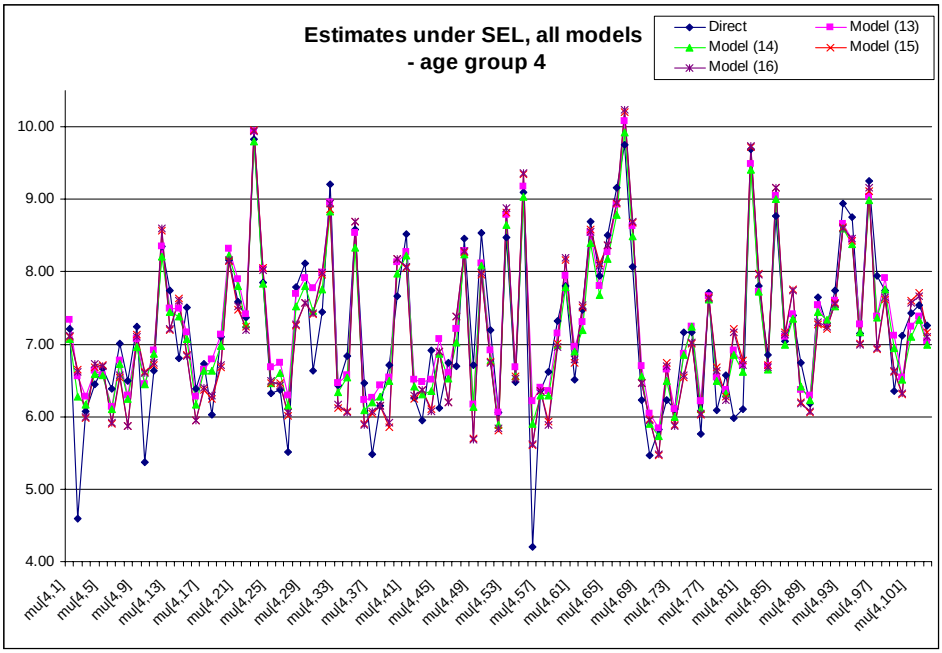
Estimates Under Different Loss Functions

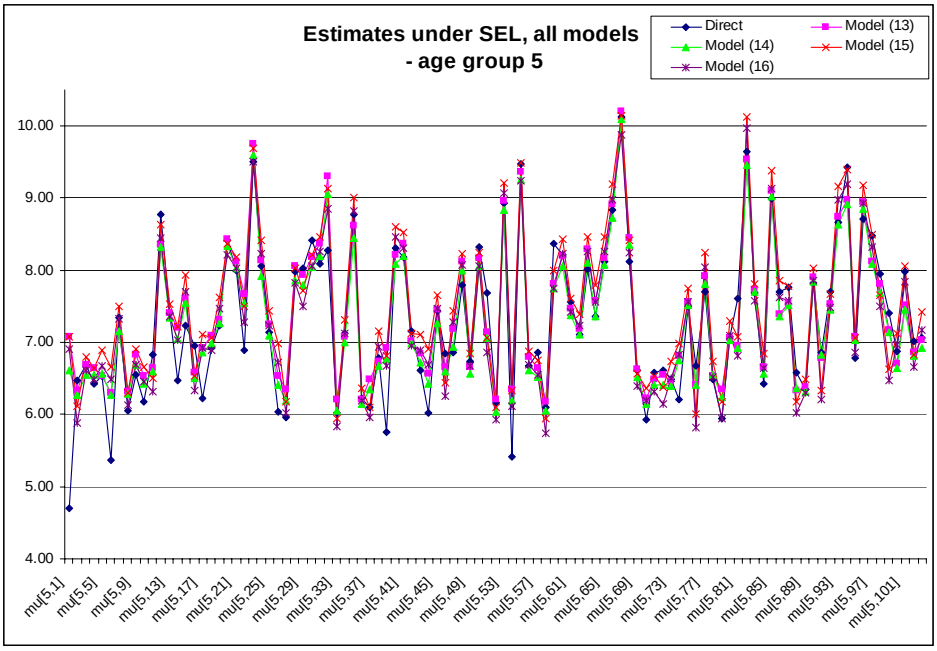
Estimates from all four models are presented below - graphed by age group as well as by the loss function. Under SEL, the following estimates were obtained (estimates are graphed on a log-scale to allow for a better representation by reducing the variance in the extreme estimates for different domains):

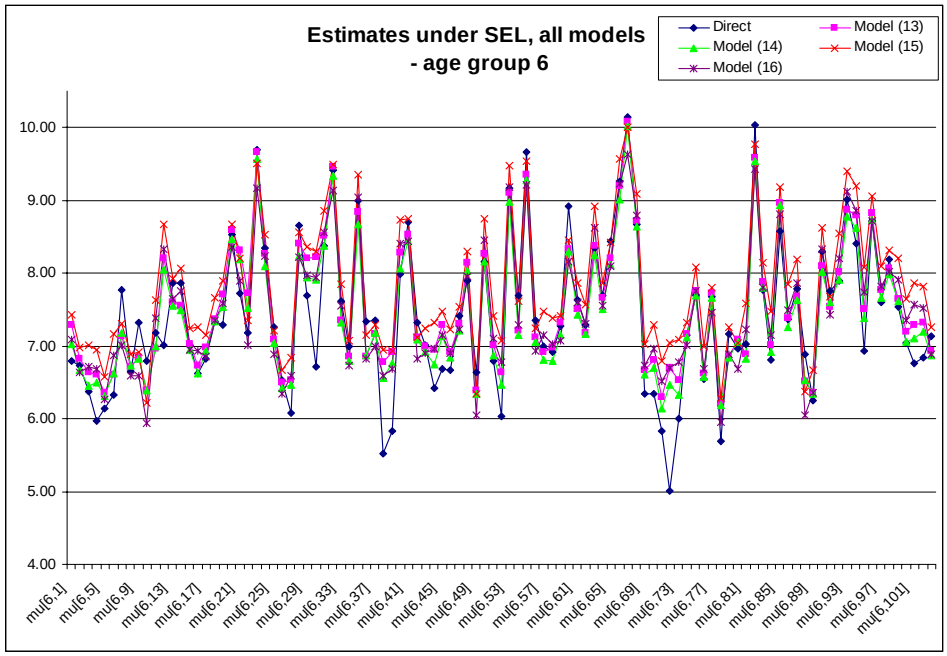


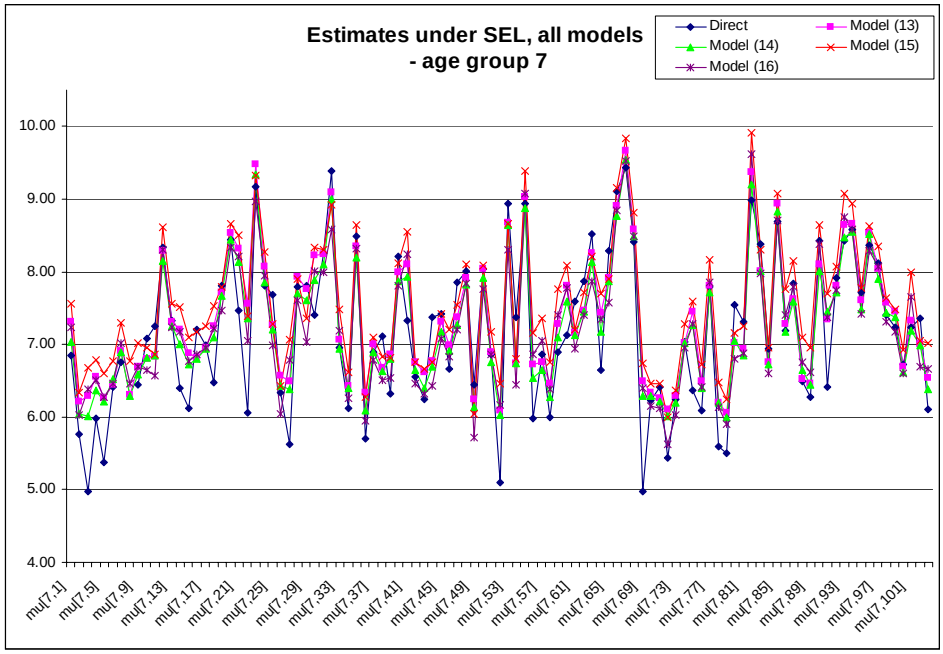

Estimates under SEL, all models
- age group 2
Direct
Model (13)
Model (14)
Model (15)
Model (16)
11.00
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[2,1]
mu[2,5]
mu[2,9]
mu[2,13]
mu[2,17]
mu[2,21]
mu[2,25]
mu[2,29]
mu[2,33]
mu[2,37]
mu[2,41]
mu[2,45]
mu[2,49]
mu[2,53]
mu[2,57]
mu[2,61]
mu[2,65]
mu[2,69]
mu[2,73]
mu[2,77]
mu[2,81]
mu[2,85]
mu[2,89]
mu[2,93]
mu[2,97]
mu[2,101]

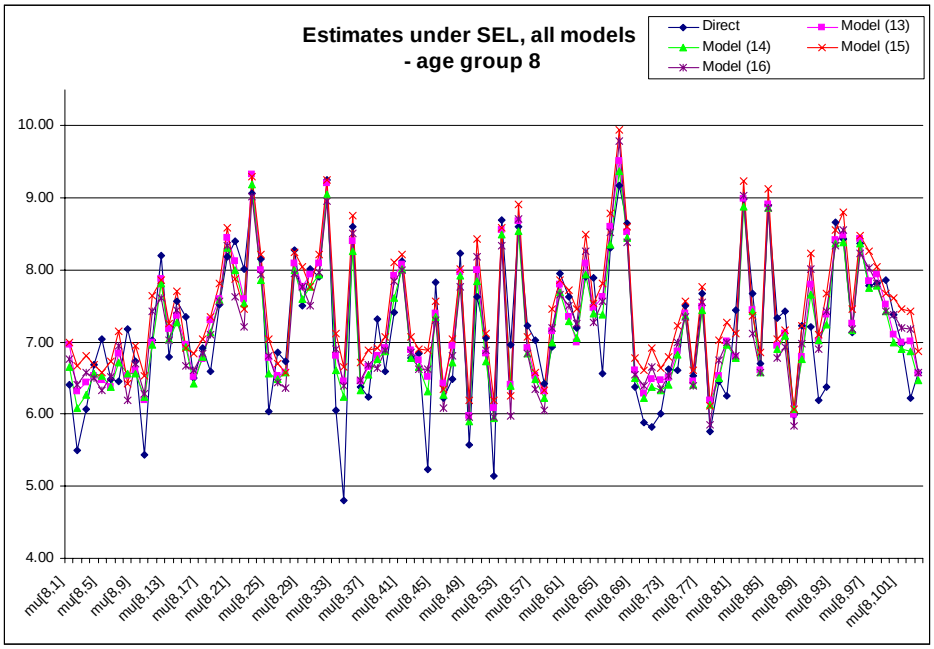

Estimates under SEL, all models
- age group 3
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.50
9.50
8.50
7.50
6.50
5.50
4.50
mu[3,1]
mu[3,5]
mu[3,9]
mu[3,13]
mu[3,17]
mu[3,21]
mu[3,25]
mu[3,29]
mu[3,33]
mu[3,37]
mu[3,41]
mu[3,45]
mu[3,49]
mu[3,53]
mu[3,57]
mu[3,61]
mu[3,65]
mu[3,69]
mu[3,73]
mu[3,77]
mu[3,81]
mu[3,85]
mu[3,89]
mu[3,93]
mu[3,97]
mu[3,101]

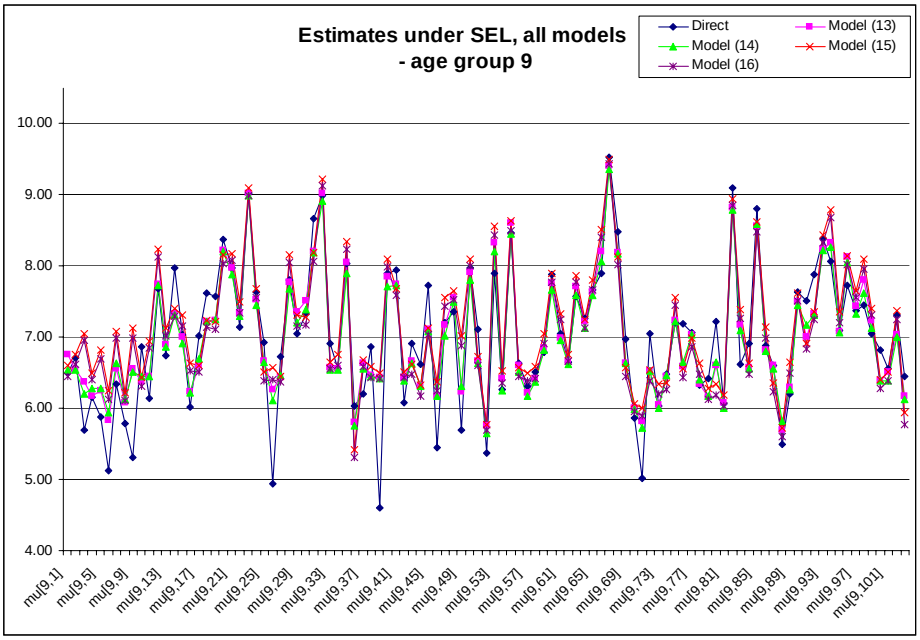

Estimates under SEL, all models
- age group 4
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[4,1]
mu[4,5]
mu[4,9]
mu[4,13]
mu[4,17]
mu[4,21]
mu[4,25]
mu[4,29]
mu[4,33]
mu[4,37]
mu[4,41]
mu[4,45]
mu[4,49]
mu[4,53]
mu[4,57]
mu[4,61]
mu[4,65]
mu[4,69]
mu[4,73]
mu[4,77]
mu[4,81]
mu[4,85]
mu[4,89]
mu[4,93]
mu[4,97]
mu[4,101]

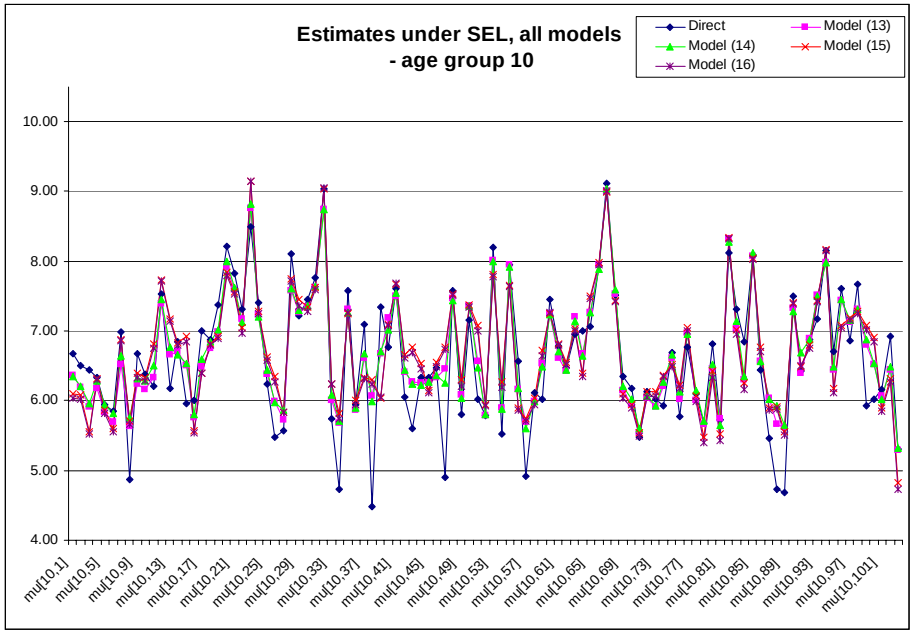

Estimates under SEL, all models
- age group 5
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[5,1]
mu[5,5]
mu[5,9]
mu[5,13]
mu[5,17]
mu[5,21]
mu[5,25]
mu[5,29]
mu[5,33]
mu[5,37]
mu[5,41]
mu[5,45]
mu[5,49]
mu[5,53]
mu[5,57]
mu[5,61]
mu[5,65]
mu[5,69]
mu[5,73]
mu[5,77]
mu[5,81]
mu[5,85]
mu[5,89]
mu[5,93]
mu[5,97]
mu[5,101]

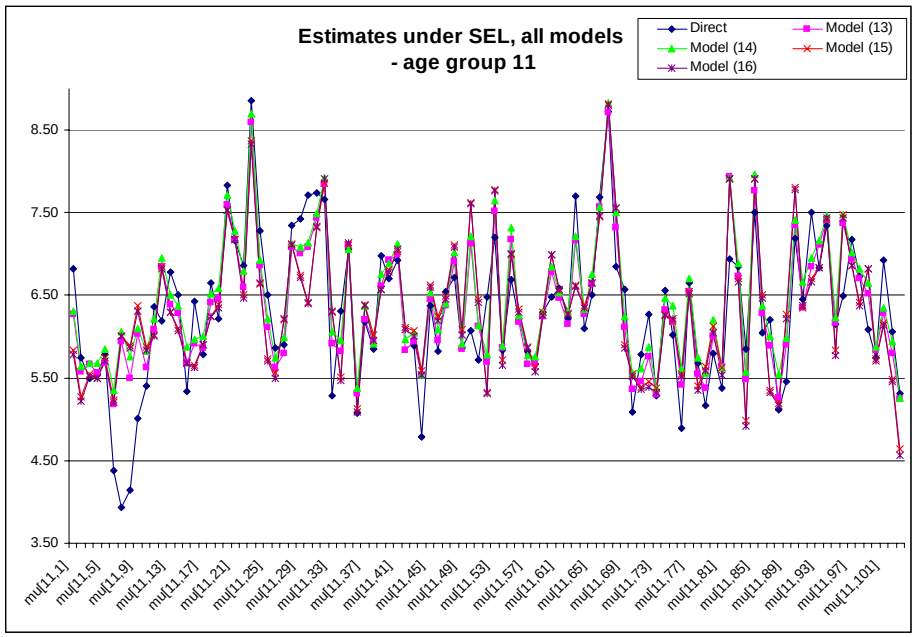

Estimates under SEL, all models
- age group 6
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[6,1]
mu[6,5]
mu[6,9]
mu[6,13]
mu[6,17]
mu[6,21]
mu[6,25]
mu[6,29]
mu[6,33]
mu[6,37]
mu[6,41]
mu[6,45]
mu[6,49]
mu[6,53]
mu[6,57]
mu[6,61]
mu[6,65]
mu[6,69]
mu[6,73]
mu[6,77]
mu[6,81]
mu[6,85]
mu[6,89]
mu[6,93]
mu[6,97]
mu[6,101]

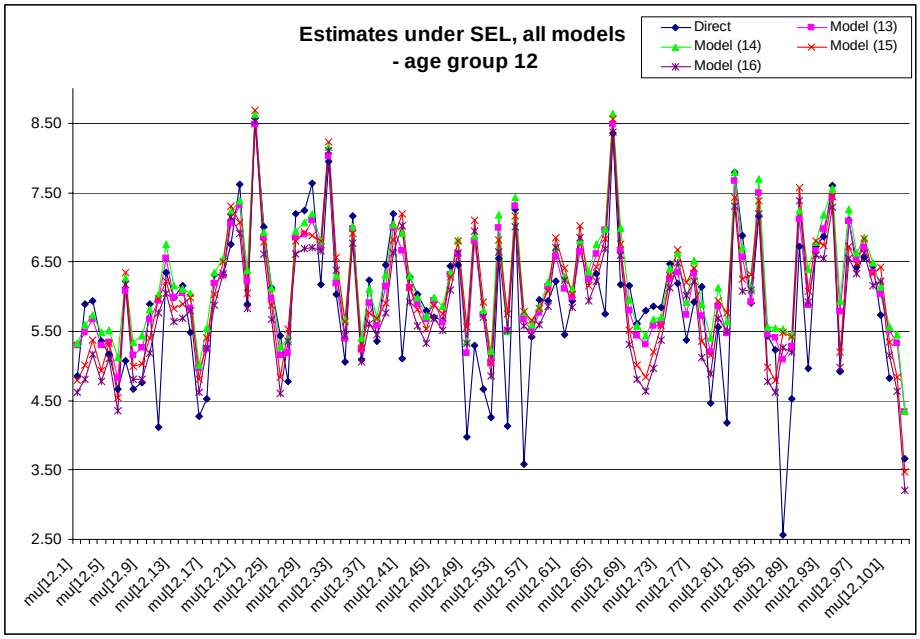

Estimates under SEL, all models
- age group 7
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[7,1]
mu[7,5]
mu[7,9]
mu[7,13]
mu[7,17]
mu[7,21]
mu[7,25]
mu[7,29]
mu[7,33]
mu[7,37]
mu[7,41]
mu[7,45]
mu[7,49]
mu[7,53]
mu[7,57]
mu[7,61]
mu[7,65]
mu[7,69]
mu[7,73]
mu[7,77]
mu[7,81]
mu[7,85]
mu[7,89]
mu[7,93]
mu[7,97]
mu[7,101]

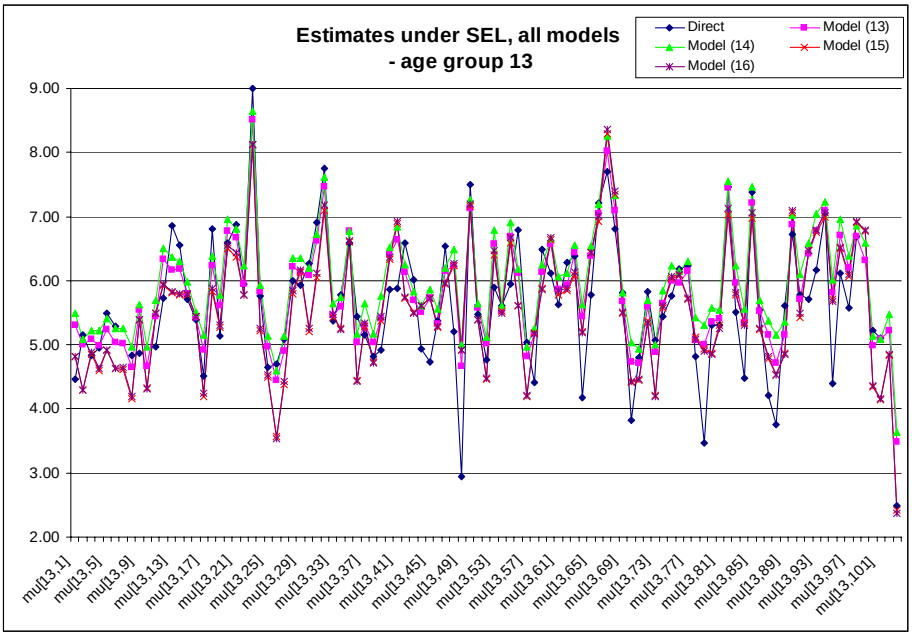

Estimates under SEL, all models
- age group 8
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[8,1]
mu[8,5]
mu[8,9]
mu[8,13]
mu[8,17]
mu[8,21]
mu[8,25]
mu[8,29]
mu[8,33]
mu[8,37]
mu[8,41]
mu[8,45]
mu[8,49]
mu[8,53]
mu[8,57]
mu[8,61]
mu[8,65]
mu[8,69]
mu[8,73]
mu[8,77]
mu[8,81]
mu[8,85]
mu[8,89]
mu[8,93]
mu[8,97]
mu[8,101]

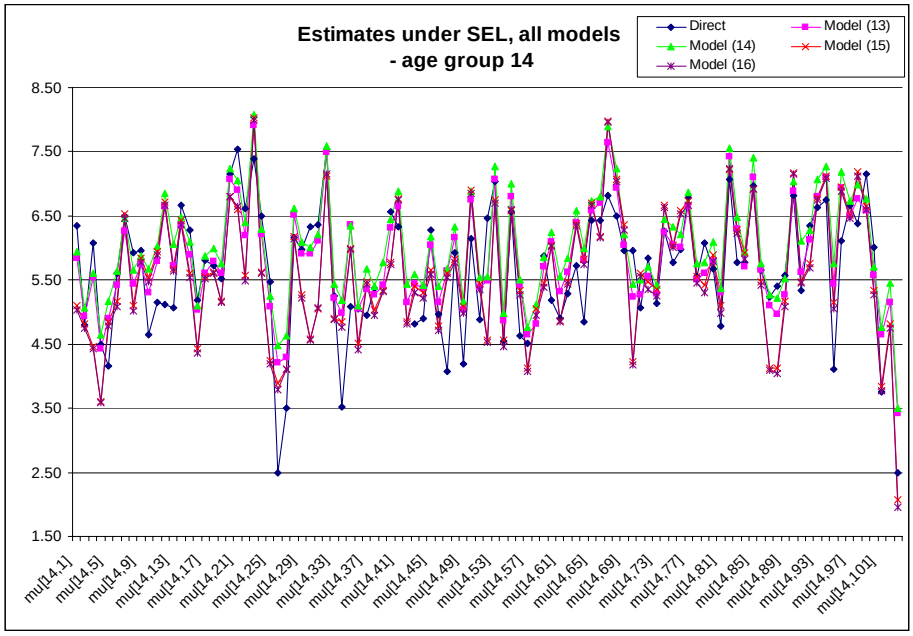

Estimates under SEL, all models
- age group 9
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[9,1]
mu[9,5]
mu[9,9]
mu[9,13]
mu[9,17]
mu[9,21]
mu[9,25]
mu[9,29]
mu[9,33]
mu[9,37]
mu[9,41]
mu[9,45]
mu[9,49]
mu[9,53]
mu[9,57]
mu[9,61]
mu[9,65]
mu[9,69]
mu[9,73]
mu[9,77]
mu[9,81]
mu[9,85]
mu[9,89]
mu[9,93]
mu[9,97]
mu[9,101]


Estimates under SEL, all models
- age group 10
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[10,1]
mu[10,5]
mu[10,9]
mu[10,13]
mu[10,17]
mu[10,21]
mu[10,25]
mu[10,29]
mu[10,33]
mu[10,37]
mu[10,41]
mu[10,45]
mu[10,49]
mu[10,53]
mu[10,57]
mu[10,61]
mu[10,65]
mu[10,69]
mu[10,73]
mu[10,77]
mu[10,81]
mu[10,85]
mu[10,89]
mu[10,93]
mu[10,97]
mu[10,101]

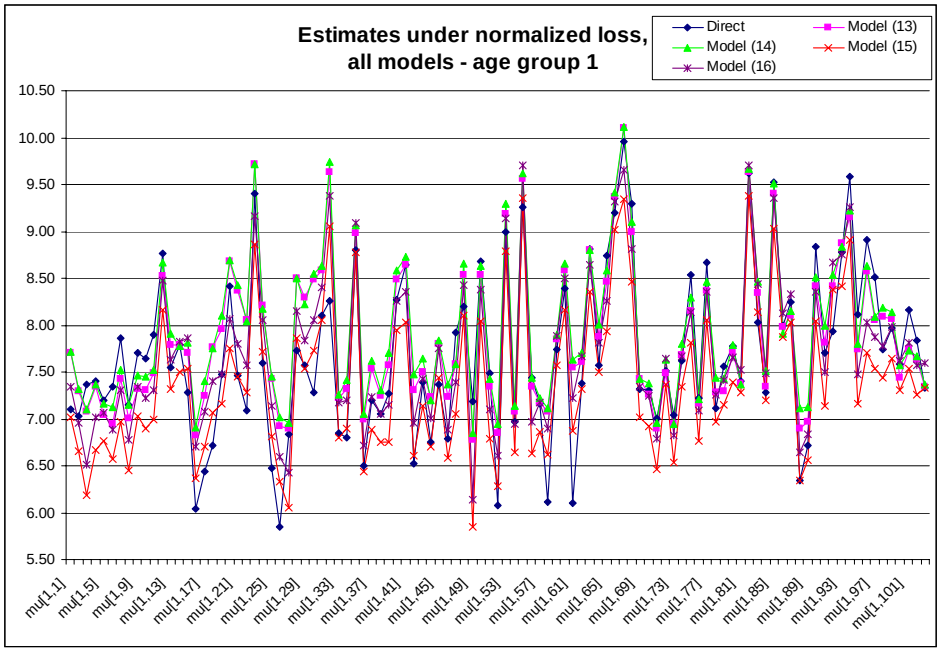

Estimates under SEL, all models
- age group 11
Direct
Model (13)
Model (14)
Model (15)
Model (16)
8.50
7.50
6.50
5.50
4.50
3.50
mu[11,1]
mu[11,5]
mu[11,9]
mu[11,13]
mu[11,17]
mu[11,21]
mu[11,25]
mu[11,29]
mu[11,33]
mu[11,37]
mu[11,41]
mu[11,45]
mu[11,49]
mu[11,53]
mu[11,57]
mu[11,61]
mu[11,65]
mu[11,69]
mu[11,73]
mu[11,77]
mu[11,81]
mu[11,85]
mu[11,89]
mu[11,93]
mu[11,97]
mu[11,101]

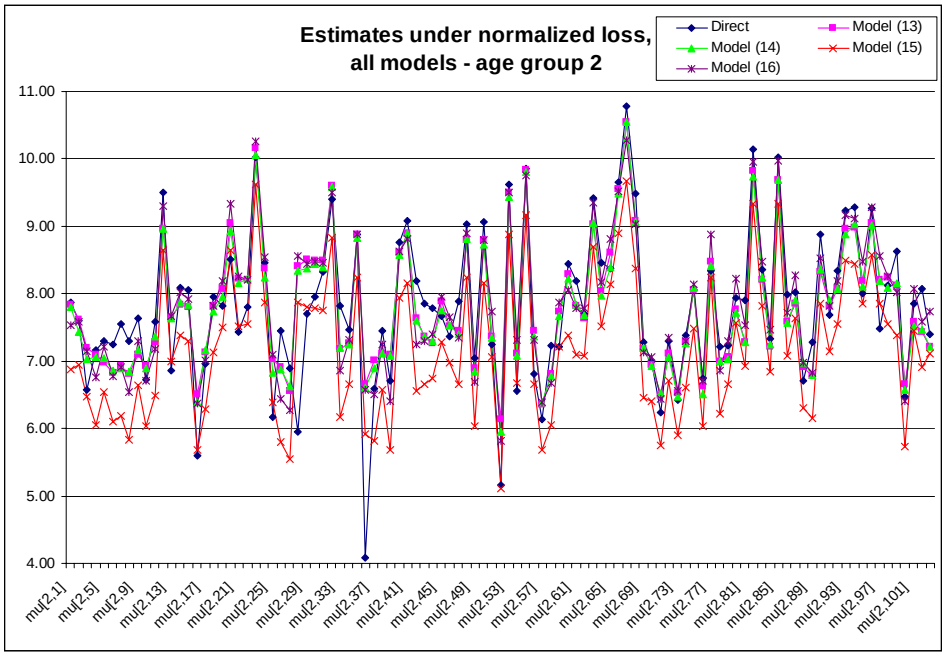

Estimates under SEL, all models
- age group 12
Direct
Model (13)
Model (14)
Model (15)
Model (16)
8.50
7.50
6.50
5.50
4.50
3.50
2.50
mu[12,1]
mu[12,5]
mu[12,9]
mu[12,13]
mu[12,17]
mu[12,21]
mu[12,25]
mu[12,29]
mu[12,33]
mu[12,37]
mu[12,41]
mu[12,45]
mu[12,49]
mu[12,53]
mu[12,57]
mu[12,61]
mu[12,65]
mu[12,69]
mu[12,73]
mu[12,77]
mu[12,81]
mu[12,85]
mu[12,89]
mu[12,93]
mu[12,97]
mu[12,101]

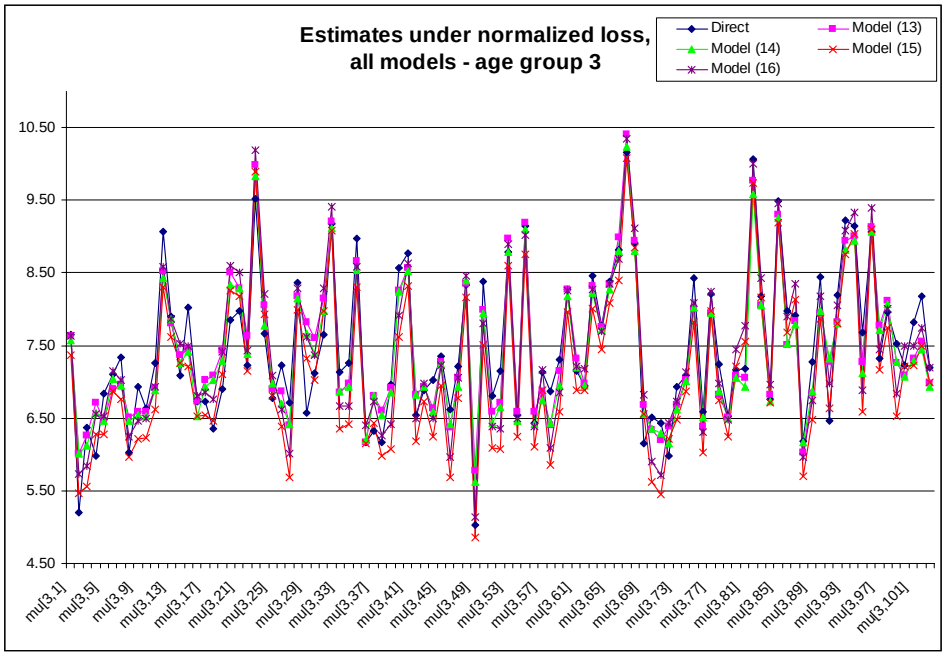

Estimates under SEL, all models
- age group 13
Direct
Model (13)
Model (14)
Model (15)
Model (16)
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
mu[13,1]
mu[13,5]
mu[13,9]
mu[13,13]
mu[13,17]
mu[13,21]
mu[13,25]
mu[13,29]
mu[13,33]
mu[13,37]
mu[13,41]
mu[13,45]
mu[13,49]
mu[13,53]
mu[13,57]
mu[13,61]
mu[13,65]
mu[13,69]
mu[13,73]
mu[13,77]
mu[13,81]
mu[13,85]
mu[13,89]
mu[13,93]
mu[13,97]
mu[13,101]

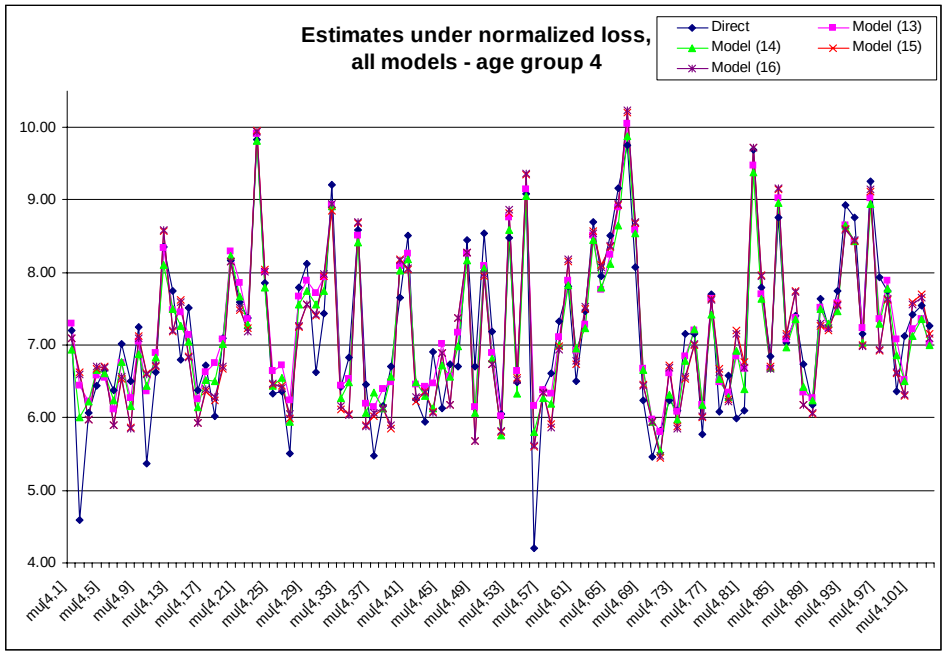

Estimates under SEL, all models
- age group 14
Direct
Model (13)
Model (14)
Model (15)
Model (16)
8.50
7.50
6.50
5.50
4.50
3.50
2.50
1.50
mu[14,1]
mu[14,5]
mu[14,9]
mu[14,13]
mu[14,17]
mu[14,21]
mu[14,25]
mu[14,29]
mu[14,33]
mu[14,37]
mu[14,41]
mu[14,45]
mu[14,49]
mu[14,53]
mu[14,57]
mu[14,61]
mu[14,65]
mu[14,69]
mu[14,73]
mu[14,77]
mu[14,81]
mu[14,85]
mu[14,89]
mu[14,93]
mu[14,97]
mu[14,101]

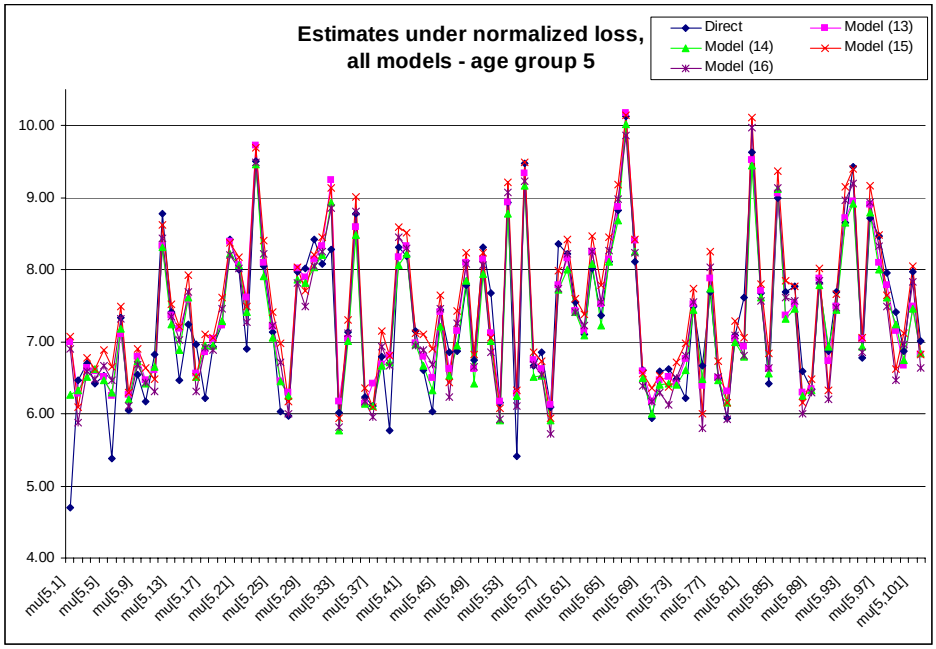
Estimates under NSEL (estimates are graphed on a log-scale to allow for a better representation by reducing the variance in the extreme estimates for different domains):

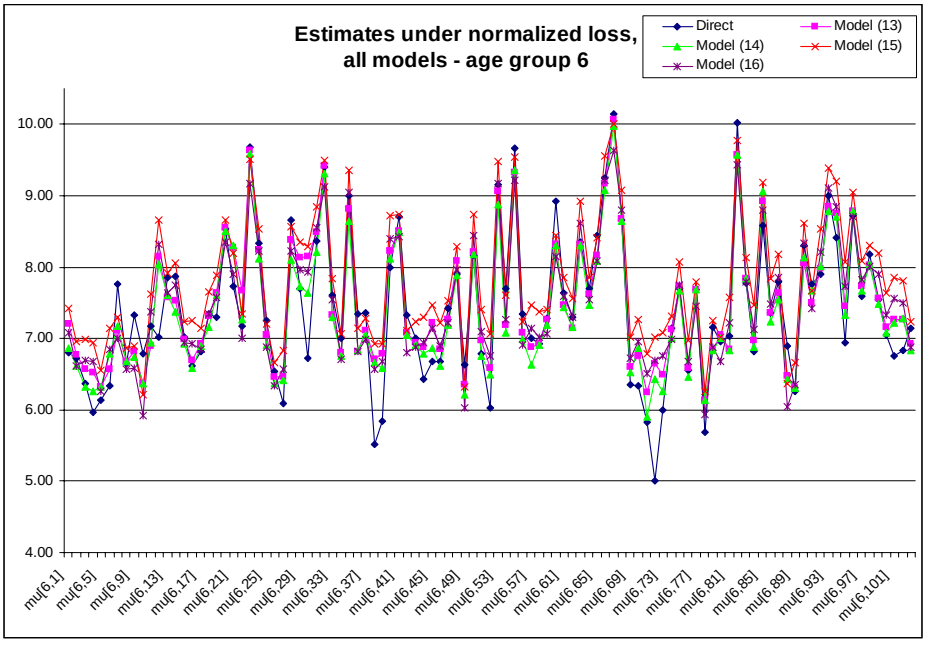


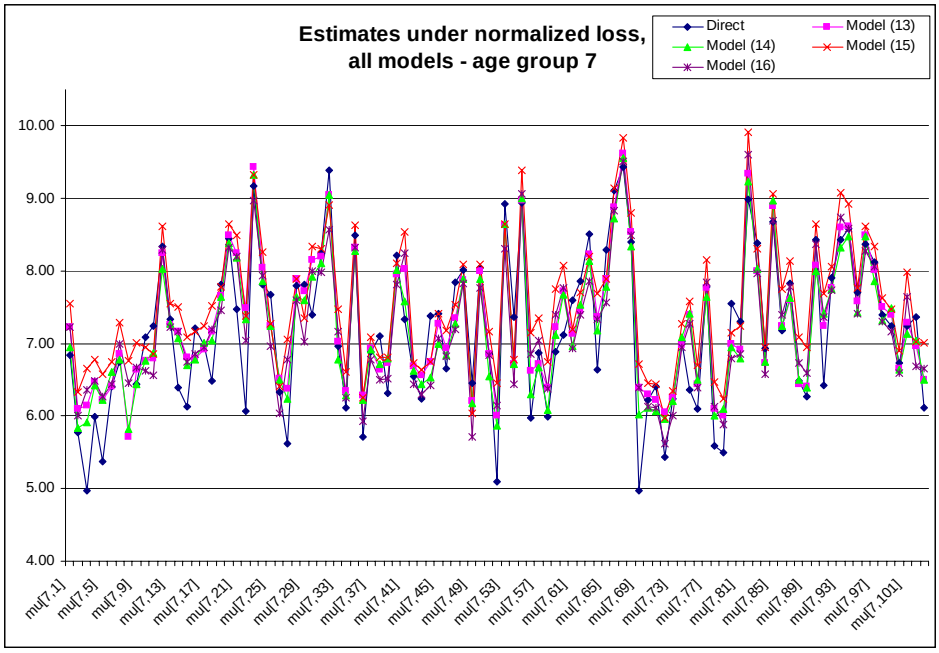


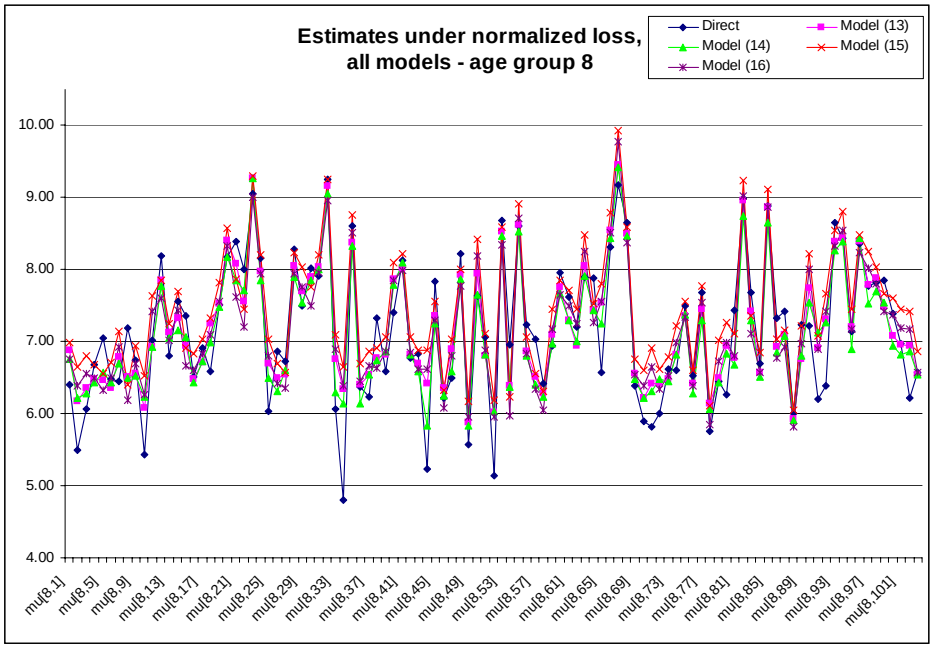

Estimates under normalized loss, all models - age group 3
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.50
9.50
8.50
7.50
6.50
5.50
4.50
mu[3,1]
mu[3,5]
mu[3,9]
mu[3,13]
mu[3,17]
mu[3,21]
mu[3,25]
mu[3,29]
mu[3,33]
mu[3,37]
mu[3,41]
mu[3,45]
mu[3,49]
mu[3,53]
mu[3,57]
mu[3,61]
mu[3,65]
mu[3,69]
mu[3,73]
mu[3,77]
mu[3,81]
mu[3,85]
mu[3,89]
mu[3,93]
mu[3,97]
mu[3,101]

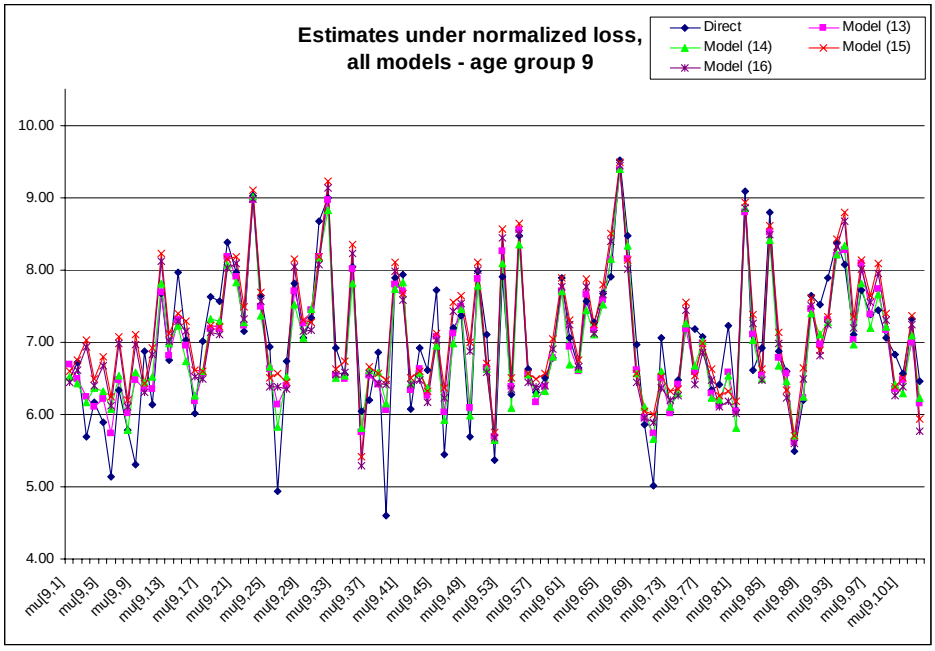

Estimates under normalized loss, all models - age group 4
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[4,1]
mu[4,5]
mu[4,9]
mu[4,13]
mu[4,17]
mu[4,21]
mu[4,25]
mu[4,29]
mu[4,33]
mu[4,37]
mu[4,41]
mu[4,45]
mu[4,49]
mu[4,53]
mu[4,57]
mu[4,61]
mu[4,65]
mu[4,69]
mu[4,73]
mu[4,77]
mu[4,81]
mu[4,85]
mu[4,89]
mu[4,93]
mu[4,97]
mu[4,101]

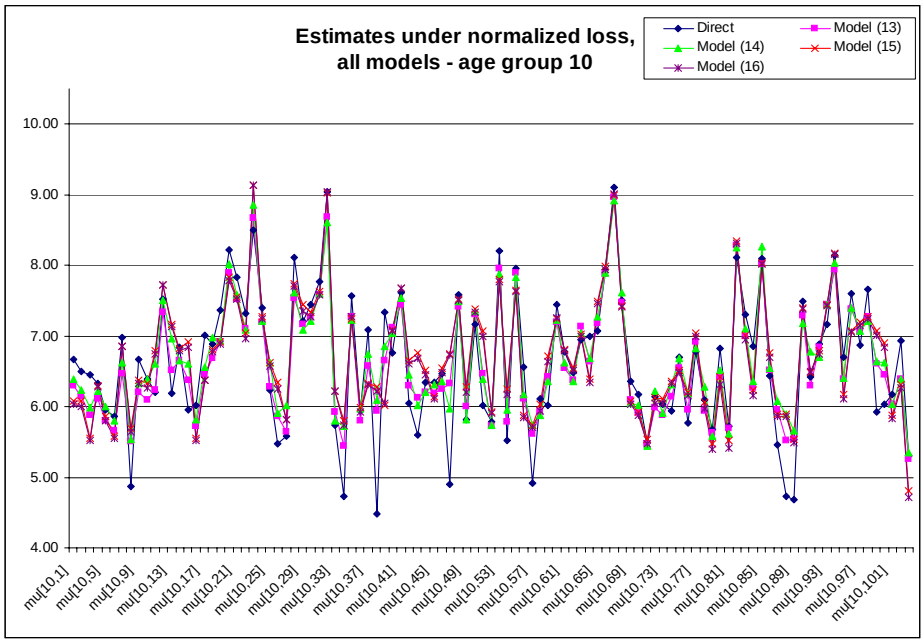

Estimates under normalized loss, all models - age group 5
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[5,1]
mu[5,5]
mu[5,9]
mu[5,13]
mu[5,17]
mu[5,21]
mu[5,25]
mu[5,29]
mu[5,33]
mu[5,37]
mu[5,41]
mu[5,45]
mu[5,49]
mu[5,53]
mu[5,57]
mu[5,61]
mu[5,65]
mu[5,69]
mu[5,73]
mu[5,77]
mu[5,81]
mu[5,85]
mu[5,89]
mu[5,93]
mu[5,97]
mu[5,101]

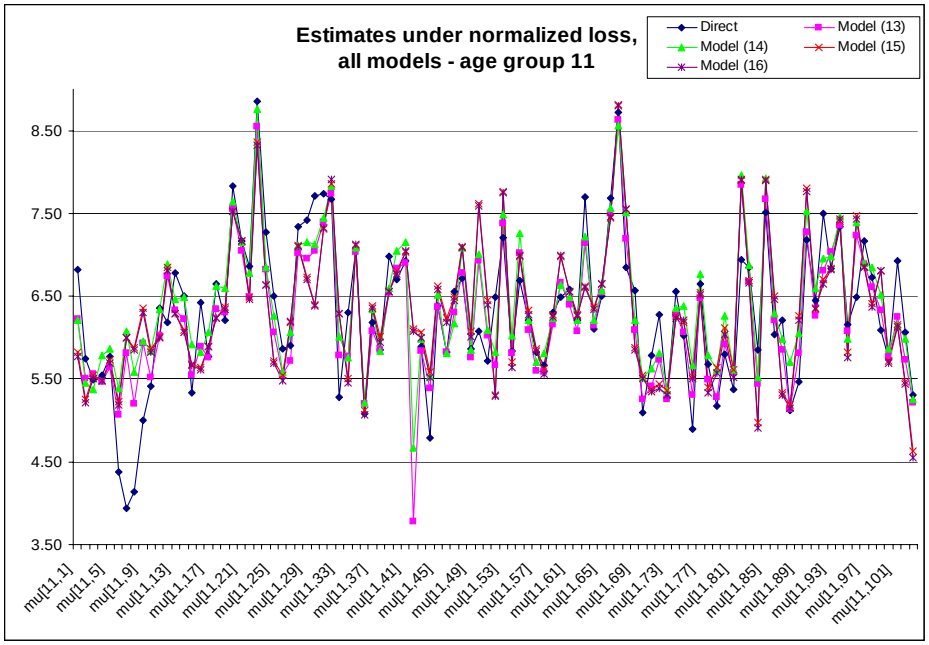

Estimates under normalized loss, all models - age group 6
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[6,1]
mu[6,5]
mu[6,9]
mu[6,13]
mu[6,17]
mu[6,21]
mu[6,25]
mu[6,29]
mu[6,33]
mu[6,37]
mu[6,41]
mu[6,45]
mu[6,49]
mu[6,53]
mu[6,57]
mu[6,61]
mu[6,65]
mu[6,69]
mu[6,73]
mu[6,77]
mu[6,81]
mu[6,85]
mu[6,89]
mu[6,93]
mu[6,97]
mu[6,101]

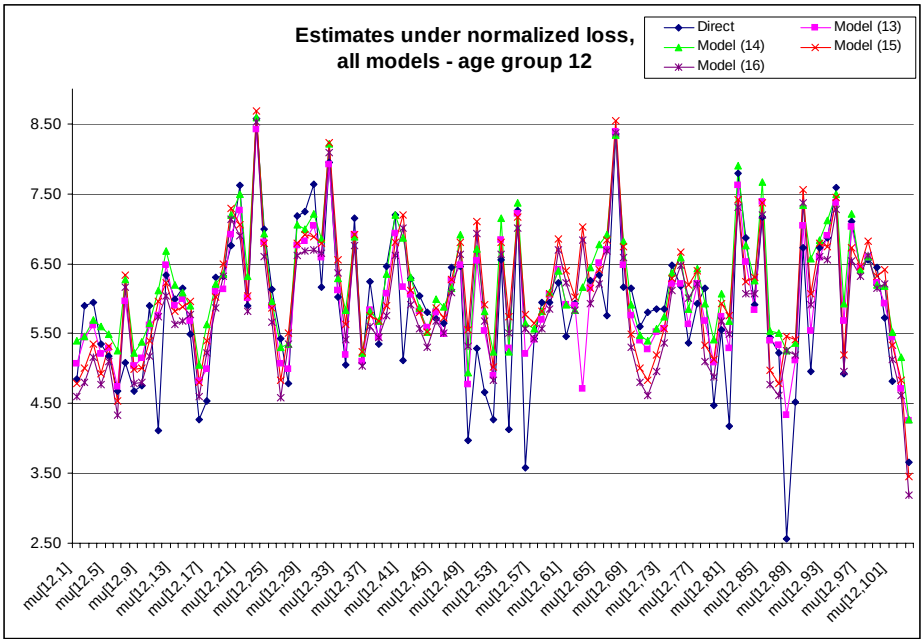

Estimates under normalized loss, all models - age group 7
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[7,1]
mu[7,5]
mu[7,9]
mu[7,13]
mu[7,17]
mu[7,21]
mu[7,25]
mu[7,29]
mu[7,33]
mu[7,37]
mu[7,41]
mu[7,45]
mu[7,49]
mu[7,53]
mu[7,57]
mu[7,61]
mu[7,65]
mu[7,69]
mu[7,73]
mu[7,77]
mu[7,81]
mu[7,85]
mu[7,89]
mu[7,93]
mu[7,97]
mu[7,101]

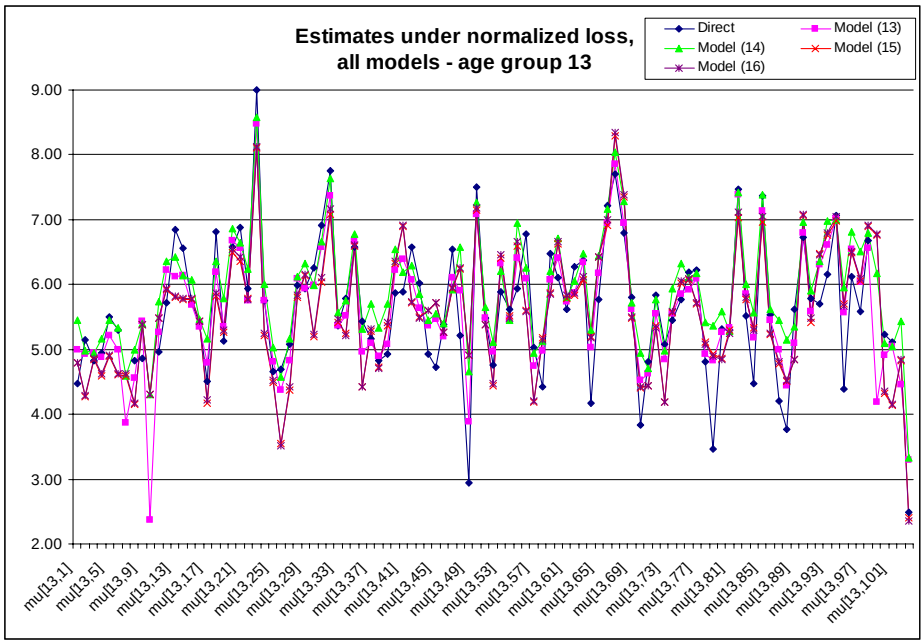

Estimates under normalized loss, all models - age group 8
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[8,1]
mu[8,5]
mu[8,9]
mu[8,13]
mu[8,17]
mu[8,21]
mu[8,25]
mu[8,29]
mu[8,33]
mu[8,37]
mu[8,41]
mu[8,45]
mu[8,49]
mu[8,53]
mu[8,57]
mu[8,61]
mu[8,65]
mu[8,69]
mu[8,73]
mu[8,77]
mu[8,81]
mu[8,85]
mu[8,89]
mu[8,93]
mu[8,97]
mu[8,101]

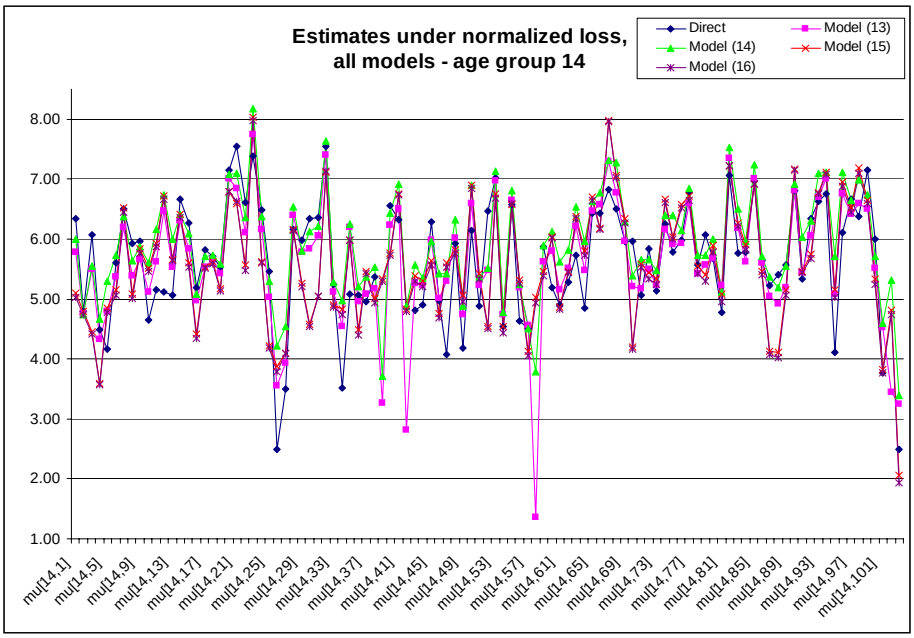

Estimates under normalized loss,
all models - age group 9
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[9,1]
mu[9,5]
mu[9,9]
mu[9,13]
mu[9,17]
mu[9,21]
mu[9,25]
mu[9,29]
mu[9,33]
mu[9,37]
mu[9,41]
mu[9,45]
mu[9,49]
mu[9,53]
mu[9,57]
mu[9,61]
mu[9,65]
mu[9,69]
mu[9,73]
mu[9,77]
mu[9,81]
mu[9,85]
mu[9,89]
mu[9,93]
mu[9,97]
mu[9,101]


Estimates under normalized loss,
all models - age group 10
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[10,1]
mu[10,5]
mu[10,9]
mu[10,13]
mu[10,17]
mu[10,21]
mu[10,25]
mu[10,29]
mu[10,33]
mu[10,37]
mu[10,41]
mu[10,45]
mu[10,49]
mu[10,53]
mu[10,57]
mu[10,61]
mu[10,65]
mu[10,69]
mu[10,73]
mu[10,77]
mu[10,81]
mu[10,85]
mu[10,89]
mu[10,93]
mu[10,97]
mu[10,101]

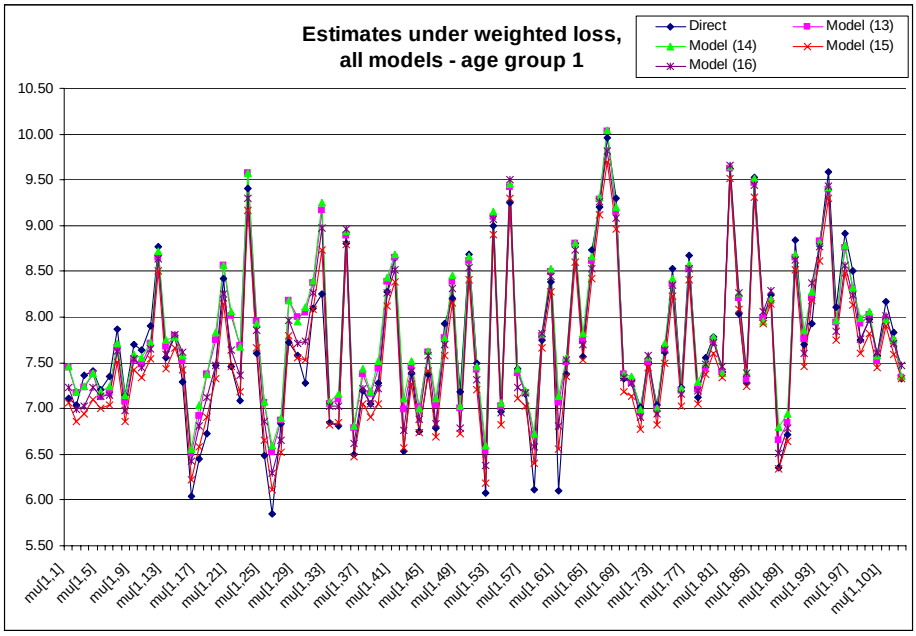

Estimates under normalized loss, all models - age group 11
Direct
Model (13)
Model (14)
Model (15)
Model (16)
8.50
7.50
6.50
5.50
4.50
3.50
mu[11,1]
mu[11,5]
mu[11,9]
mu[11,13]
mu[11,17]
mu[11,21]
mu[11,25]
mu[11,29]
mu[11,33]
mu[11,37]
mu[11,41]
mu[11,45]
mu[11,49]
mu[11,53]
mu[11,57]
mu[11,61]
mu[11,65]
mu[11,69]
mu[11,73]
mu[11,77]
mu[11,81]
mu[11,85]
mu[11,89]
mu[11,93]
mu[11,97]
mu[11,101]

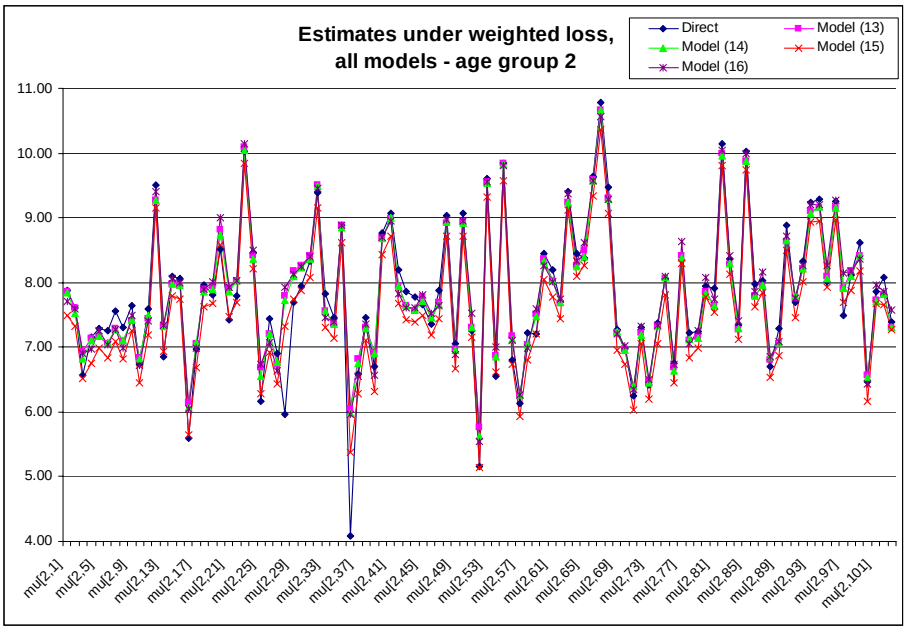

Estimates under normalized loss, all models - age group 12
Direct
Model (13)
Model (14)
Model (15)
Model (16)
8.50
7.50
6.50
5.50
4.50
3.50
2.50
mu[12,1]
mu[12,5]
mu[12,9]
mu[12,13]
mu[12,17]
mu[12,21]
mu[12,25]
mu[12,29]
mu[12,33]
mu[12,37]
mu[12,41]
mu[12,45]
mu[12,49]
mu[12,53]
mu[12,57]
mu[12,61]
mu[12,65]
mu[12,69]
mu[12,73]
mu[12,77]
mu[12,81]
mu[12,85]
mu[12,89]
mu[12,93]
mu[12,97]
mu[12,101]

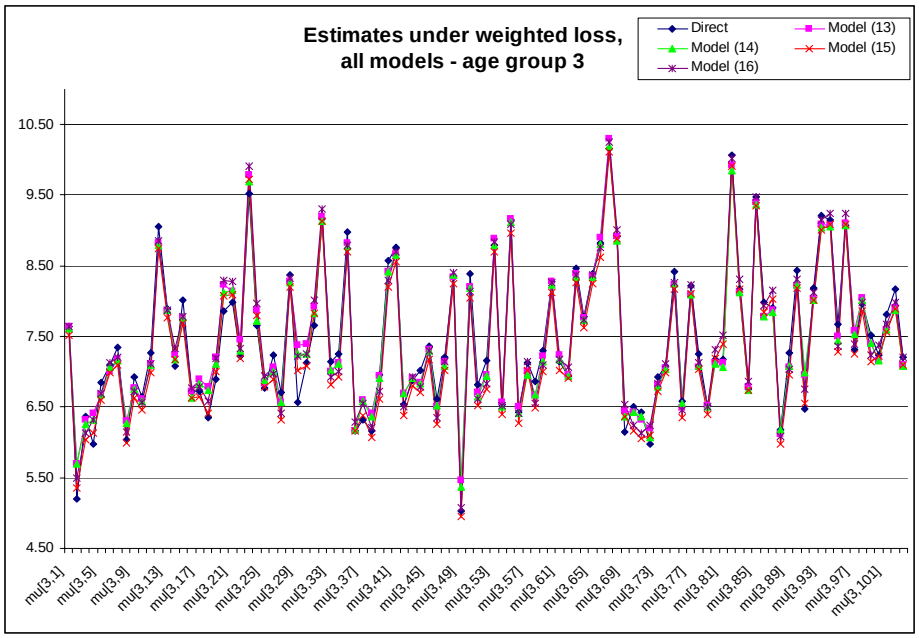

Estimates under normalized loss, all models - age group 13
Direct
Model (13)
Model (14)
Model (15)
Model (16)
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
mu[13,1]
mu[13,5]
mu[13,9]
mu[13,13]
mu[13,17]
mu[13,21]
mu[13,25]
mu[13,29]
mu[13,33]
mu[13,37]
mu[13,41]
mu[13,45]
mu[13,49]
mu[13,53]
mu[13,57]
mu[13,61]
mu[13,65]
mu[13,69]
mu[13,73]
mu[13,77]
mu[13,81]
mu[13,85]
mu[13,89]
mu[13,93]
mu[13,97]
mu[13,101]

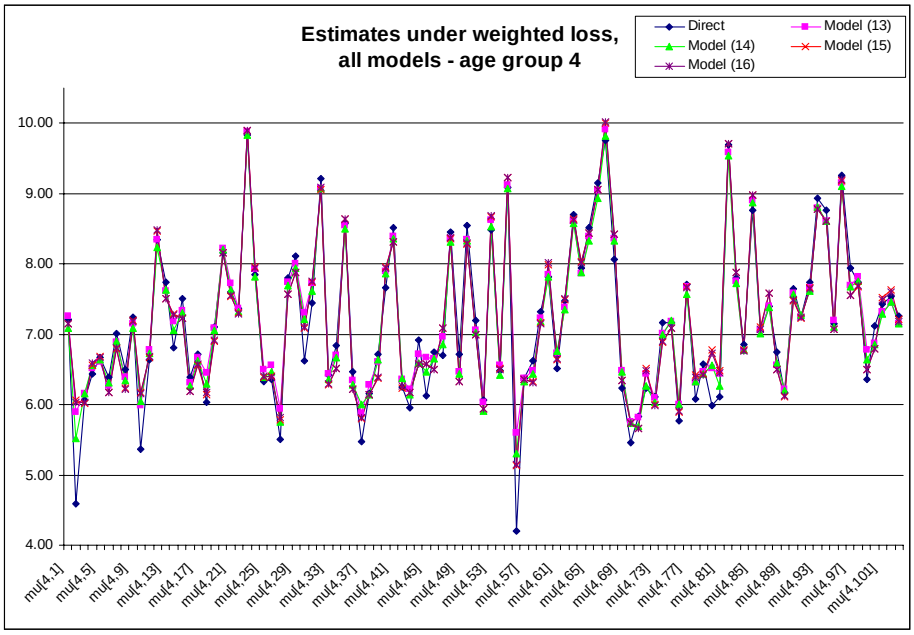

Estimates under normalized loss, all models - age group 14
Direct
Model (13)
Model (14)
Model (15)
Model (16)
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
mu[14,1]
mu[14,5]
mu[14,9]
mu[14,13]
mu[14,17]
mu[14,21]
mu[14,25]
mu[14,29]
mu[14,33]
mu[14,37]
mu[14,41]
mu[14,45]
mu[14,49]
mu[14,53]
mu[14,57]
mu[14,61]
mu[14,65]
mu[14,69]
mu[14,73]
mu[14,77]
mu[14,81]
mu[14,85]
mu[14,89]
mu[14,93]
mu[14,97]
mu[14,101]

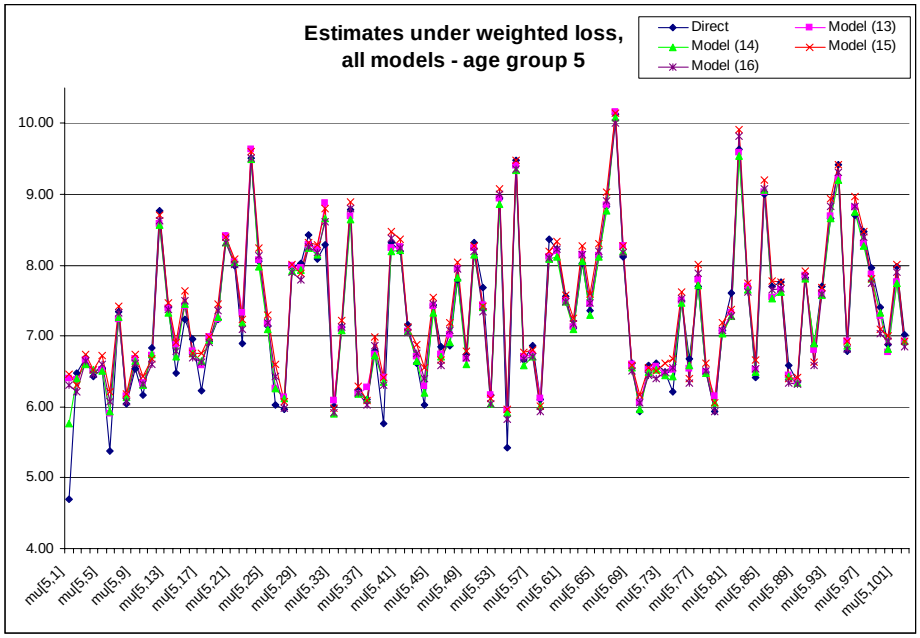
Estimates under WBL (estimates are graphed on a log-scale to allow for a better representation by reducing the variance in the extreme estimates for different domains):

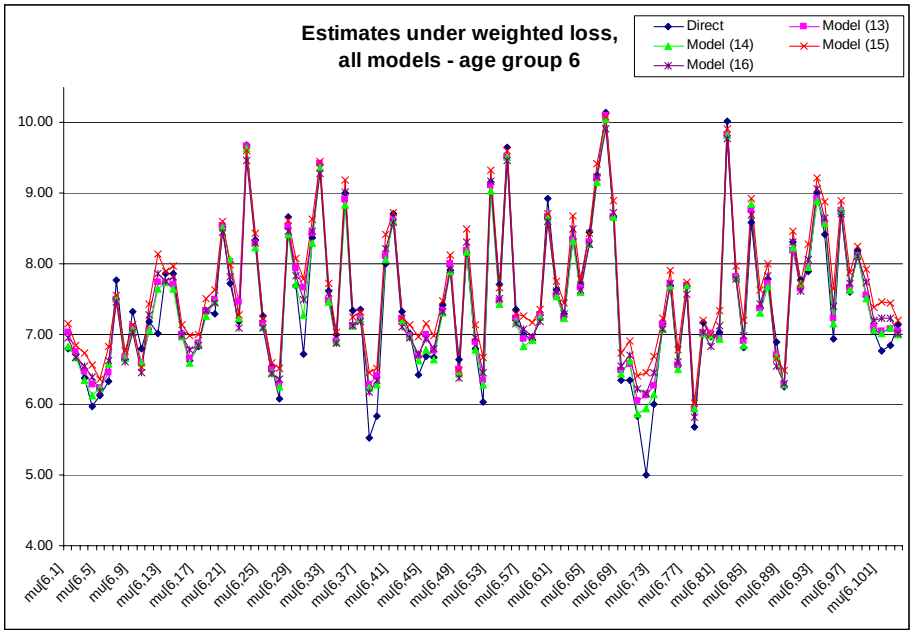


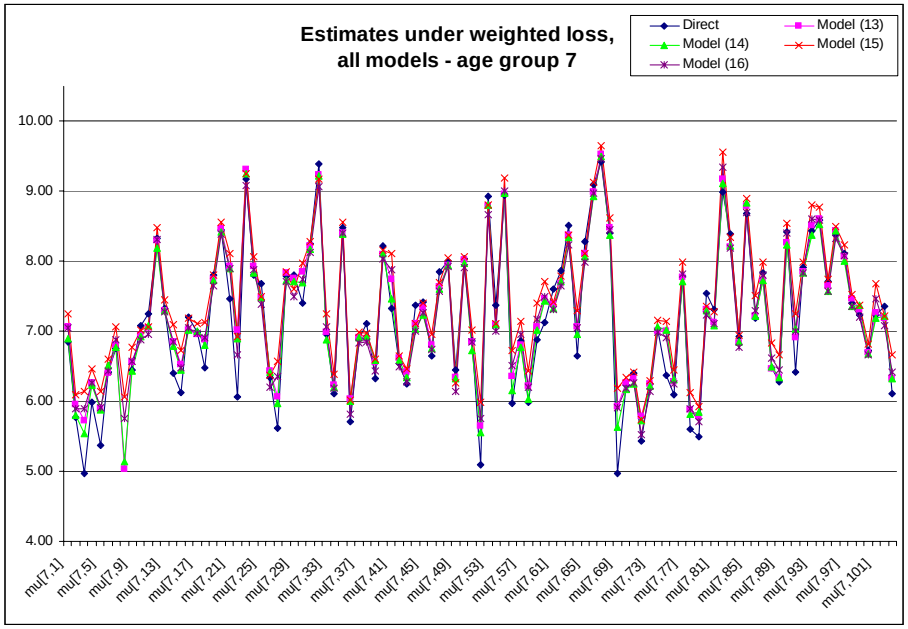


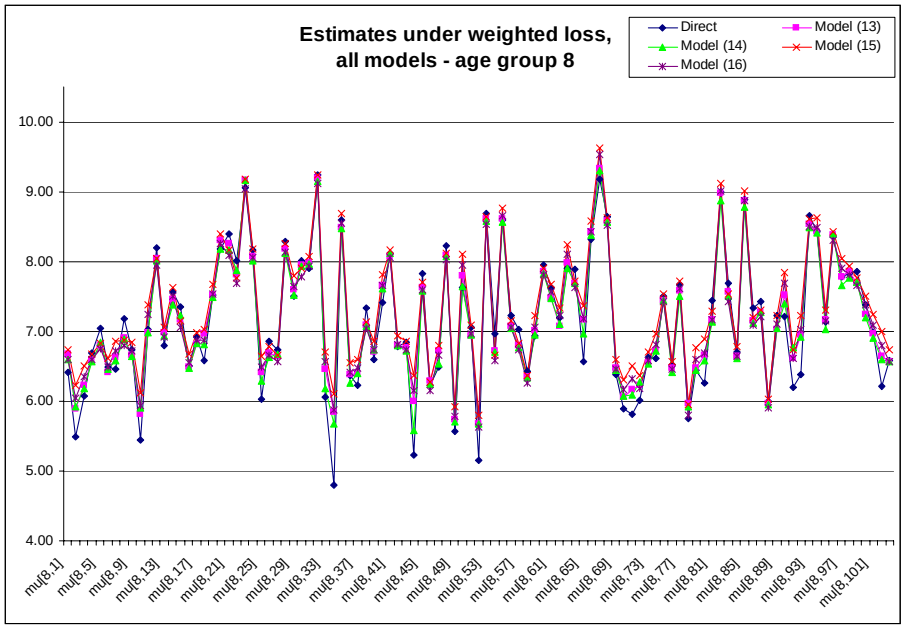

Estimates under weighted loss, all models - age group 3
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.50
9.50
8.50
7.50
6.50
5.50
4.50
mu[3,1]
mu[3,5]
mu[3,9]
mu[3,13]
mu[3,17]
mu[3,21]
mu[3,25]
mu[3,29]
mu[3,33]
mu[3,37]
mu[3,41]
mu[3,45]
mu[3,49]
mu[3,53]
mu[3,57]
mu[3,61]
mu[3,65]
mu[3,69]
mu[3,73]
mu[3,77]
mu[3,81]
mu[3,85]
mu[3,89]
mu[3,93]
mu[3,97]
mu[3,101]

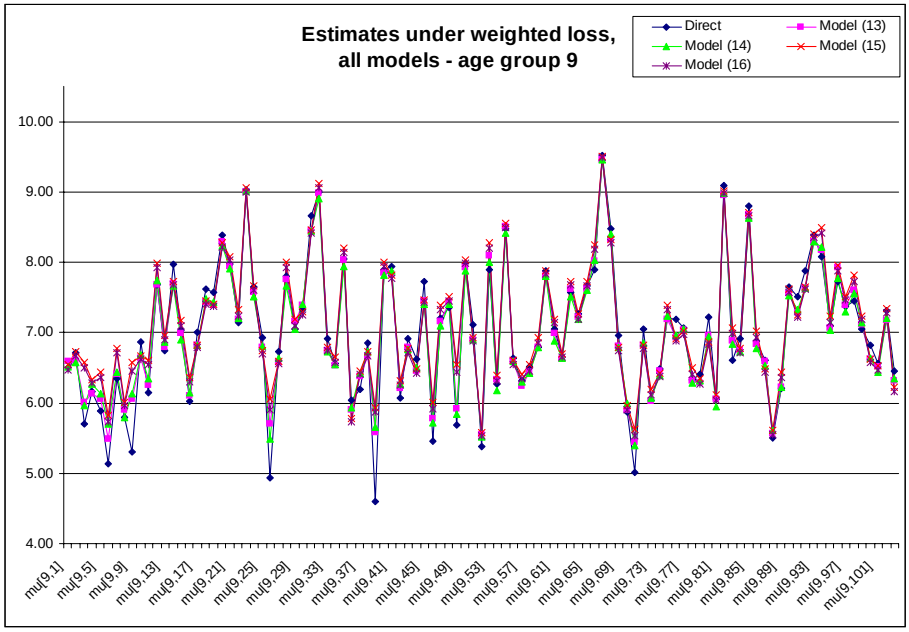

Estimates under weighted loss, all models - age group 4
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[4,1]
mu[4,5]
mu[4,9]
mu[4,13]
mu[4,17]
mu[4,21]
mu[4,25]
mu[4,29]
mu[4,33]
mu[4,37]
mu[4,41]
mu[4,45]
mu[4,49]
mu[4,53]
mu[4,57]
mu[4,61]
mu[4,65]
mu[4,69]
mu[4,73]
mu[4,77]
mu[4,81]
mu[4,85]
mu[4,89]
mu[4,93]
mu[4,97]
mu[4,101]

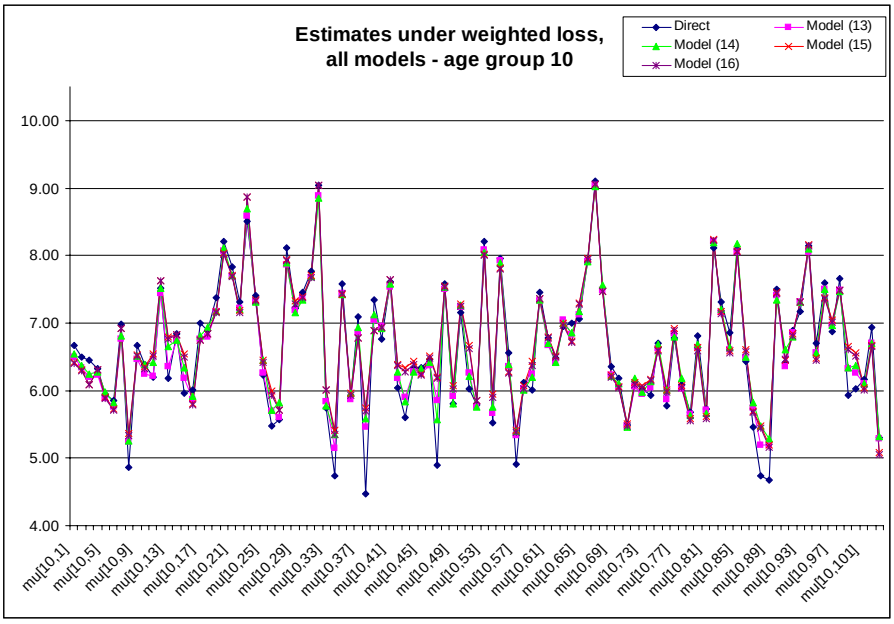

Estimates under weighted loss, all models - age group 5
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[5,1]
mu[5,5]
mu[5,9]
mu[5,13]
mu[5,17]
mu[5,21]
mu[5,25]
mu[5,29]
mu[5,33]
mu[5,37]
mu[5,41]
mu[5,45]
mu[5,49]
mu[5,53]
mu[5,57]
mu[5,61]
mu[5,65]
mu[5,69]
mu[5,73]
mu[5,77]
mu[5,81]
mu[5,85]
mu[5,89]
mu[5,93]
mu[5,97]
mu[5,101]

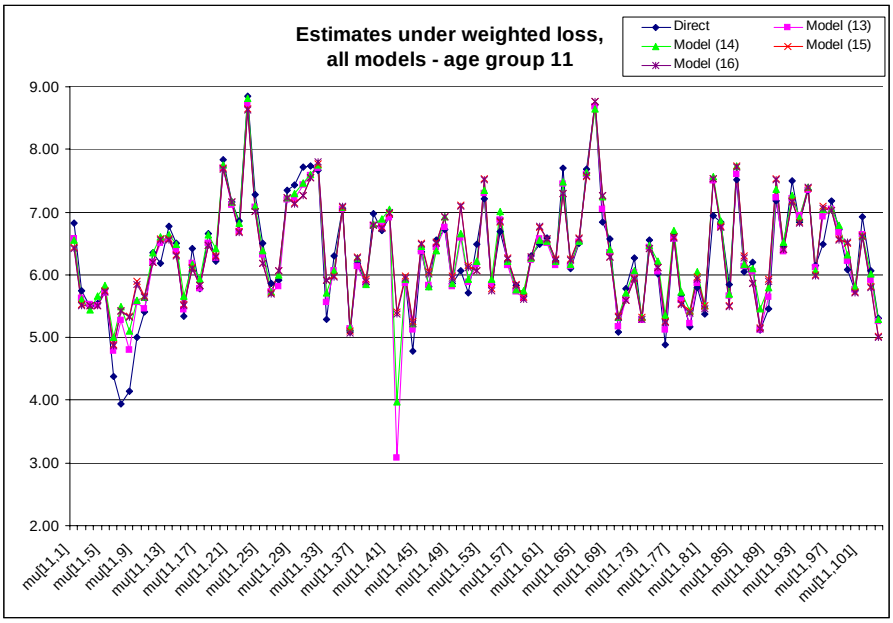

Estimates under weighted loss, all models - age group 6
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[6,1]
mu[6,5]
mu[6,9]
mu[6,13]
mu[6,17]
mu[6,21]
mu[6,25]
mu[6,29]
mu[6,33]
mu[6,37]
mu[6,41]
mu[6,45]
mu[6,49]
mu[6,53]
mu[6,57]
mu[6,61]
mu[6,65]
mu[6,69]
mu[6,73]
mu[6,77]
mu[6,81]
mu[6,85]
mu[6,89]
mu[6,93]
mu[6,97]
mu[6,101]

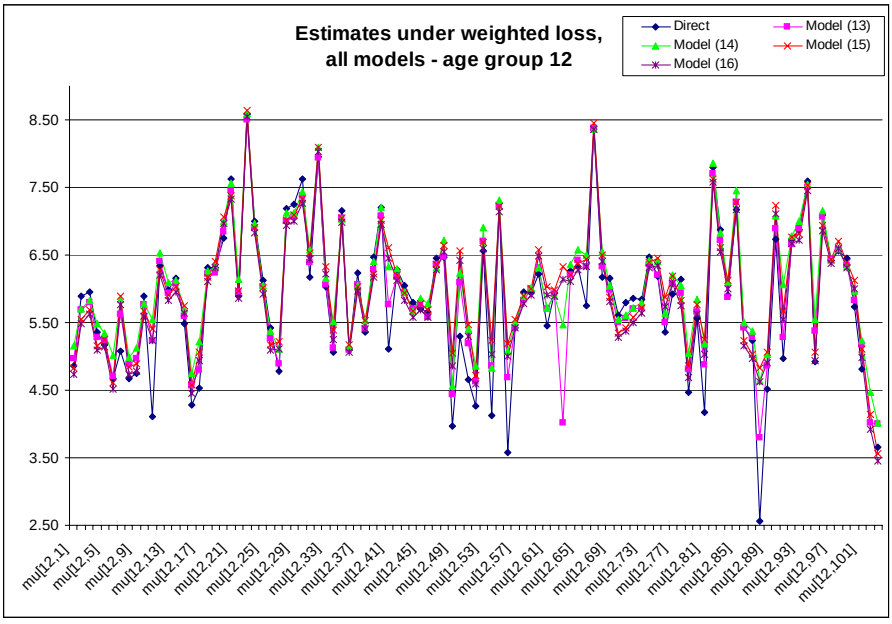

Estimates under weighted loss, all models - age group 7
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[7,1]
mu[7,5]
mu[7,9]
mu[7,13]
mu[7,17]
mu[7,21]
mu[7,25]
mu[7,29]
mu[7,33]
mu[7,37]
mu[7,41]
mu[7,45]
mu[7,49]
mu[7,53]
mu[7,57]
mu[7,61]
mu[7,65]
mu[7,69]
mu[7,73]
mu[7,77]
mu[7,81]
mu[7,85]
mu[7,89]
mu[7,93]
mu[7,97]
mu[7,101]

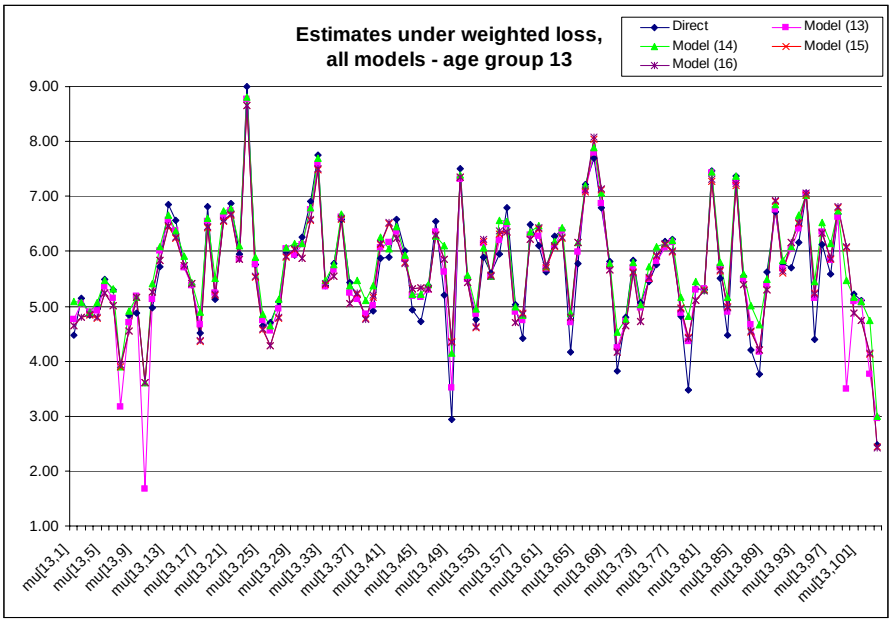

Estimates under weighted loss, all models - age group 8
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[8,1]
mu[8,5]
mu[8,9]
mu[8,13]
mu[8,17]
mu[8,21]
mu[8,25]
mu[8,29]
mu[8,33]
mu[8,37]
mu[8,41]
mu[8,45]
mu[8,49]
mu[8,53]
mu[8,57]
mu[8,61]
mu[8,65]
mu[8,69]
mu[8,73]
mu[8,77]
mu[8,81]
mu[8,85]
mu[8,89]
mu[8,93]
mu[8,97]
mu[8,101]

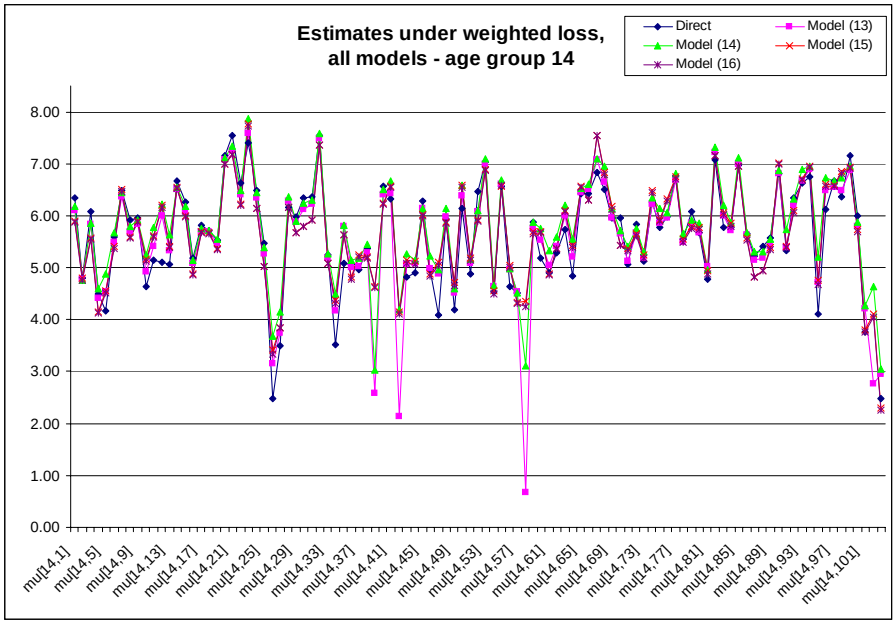

Estimates under weighted loss,
all models - age group 9
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[9,1]
mu[9,5]
mu[9,9]
mu[9,13]
mu[9,17]
mu[9,21]
mu[9,25]
mu[9,29]
mu[9,33]
mu[9,37]
mu[9,41]
mu[9,45]
mu[9,49]
mu[9,53]
mu[9,57]
mu[9,61]
mu[9,65]
mu[9,69]
mu[9,73]
mu[9,77]
mu[9,81]
mu[9,85]
mu[9,89]
mu[9,93]
mu[9,97]
mu[9,101]


Estimates under weighted loss,
all models - age group 10
Direct
Model (13)
Model (14)
Model (15)
Model (16)
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[10,1]
mu[10,5]
mu[10,9]
mu[10,13]
mu[10,17]
mu[10,21]
mu[10,25]
mu[10,29]
mu[10,33]
mu[10,37]
mu[10,41]
mu[10,45]
mu[10,49]
mu[10,53]
mu[10,57]
mu[10,61]
mu[10,65]
mu[10,69]
mu[10,73]
mu[10,77]
mu[10,81]
mu[10,85]
mu[10,89]
mu[10,93]
mu[10,97]
mu[10,101]

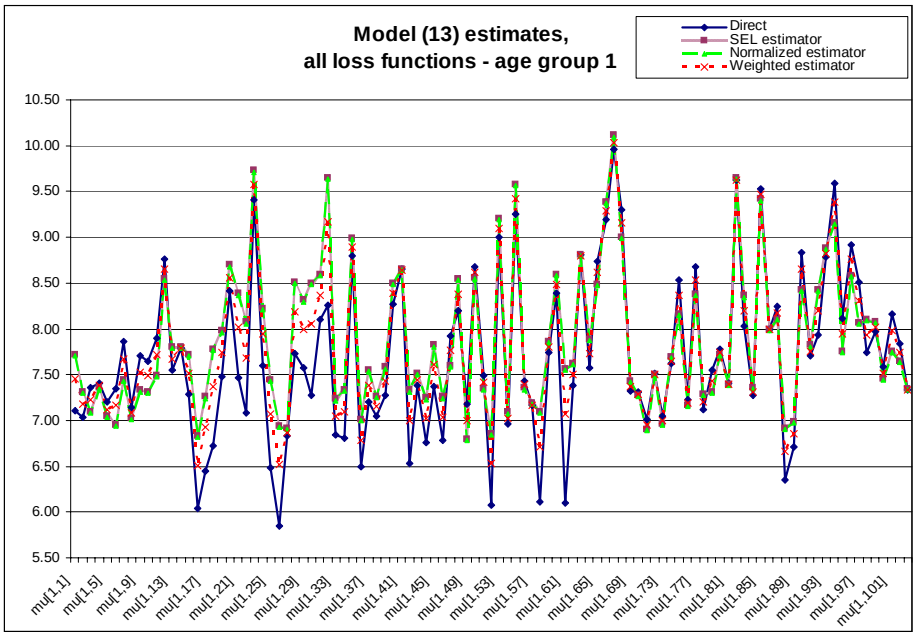

Estimates under weighted loss, all models - age group 11
Direct
Model (13)
Model (14)
Model (15)
Model (16)
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
mu[11,1]
mu[11,5]
mu[11,9]
mu[11,13]
mu[11,17]
mu[11,21]
mu[11,25]
mu[11,29]
mu[11,33]
mu[11,37]
mu[11,41]
mu[11,45]
mu[11,49]
mu[11,53]
mu[11,57]
mu[11,61]
mu[11,65]
mu[11,69]
mu[11,73]
mu[11,77]
mu[11,81]
mu[11,85]
mu[11,89]
mu[11,93]
mu[11,97]
mu[11,101]

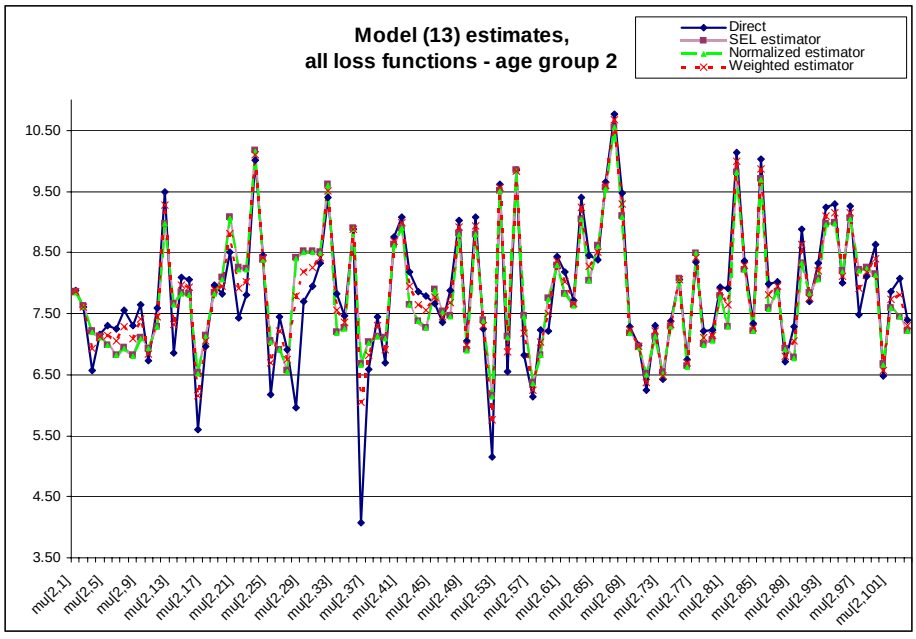

Estimates under weighted loss, all models - age group 12
Direct
Model (13)
Model (14)
Model (15)
Model (16)
8.50
7.50
6.50
5.50
4.50
3.50
2.50
mu[12,1]
mu[12,5]
mu[12,9]
mu[12,13]
mu[12,17]
mu[12,21]
mu[12,25]
mu[12,29]
mu[12,33]
mu[12,37]
mu[12,41]
mu[12,45]
mu[12,49]
mu[12,53]
mu[12,57]
mu[12,61]
mu[12,65]
mu[12,69]
mu[12,73]
mu[12,77]
mu[12,81]
mu[12,85]
mu[12,89]
mu[12,93]
mu[12,97]
mu[12,101]

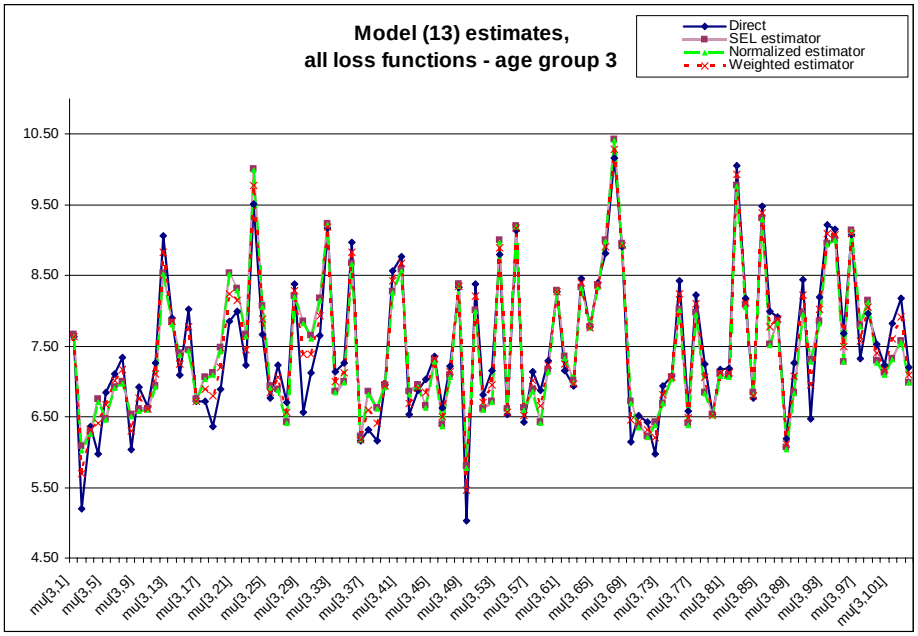

Estimates under weighted loss, all models - age group 13
Direct
Model (13)
Model (14)
Model (15)
Model (16)
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
mu[13,1]
mu[13,5]
mu[13,9]
mu[13,13]
mu[13,17]
mu[13,21]
mu[13,25]
mu[13,29]
mu[13,33]
mu[13,37]
mu[13,41]
mu[13,45]
mu[13,49]
mu[13,53]
mu[13,57]
mu[13,61]
mu[13,65]
mu[13,69]
mu[13,73]
mu[13,77]
mu[13,81]
mu[13,85]
mu[13,89]
mu[13,93]
mu[13,97]
mu[13,101]

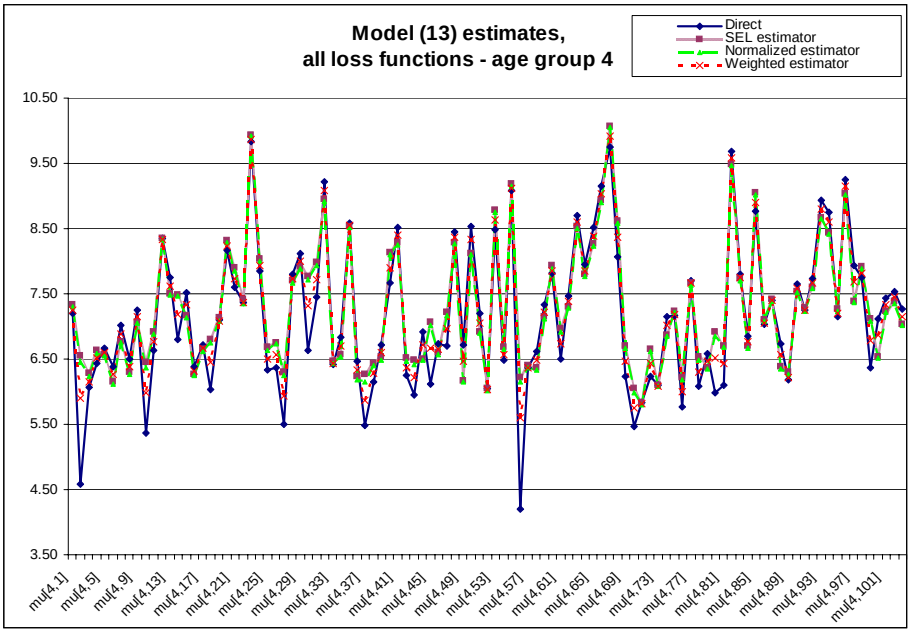

Estimates under weighted loss, all models - age group 14
Direct
Model (13)
Model (14)
Model (15)
Model (16)
8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
mu[14,1]
mu[14,5]
mu[14,9]
mu[14,13]
mu[14,17]
mu[14,21]
mu[14,25]
mu[14,29]
mu[14,33]
mu[14,37]
mu[14,41]
mu[14,45]
mu[14,49]
mu[14,53]
mu[14,57]
mu[14,61]
mu[14,65]
mu[14,69]
mu[14,73]
mu[14,77]
mu[14,81]
mu[14,85]
mu[14,89]
mu[14,93]
mu[14,97]
mu[14,101]

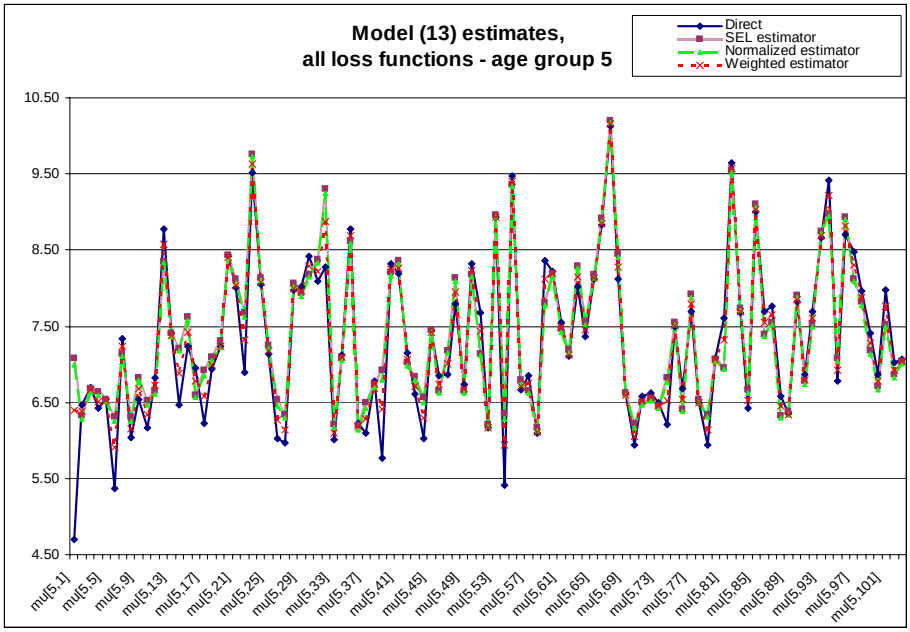
Below, estimates under all three loss functions are presented for one model at a time, graphed by age group. Model (13) estimates (estimates are graphed on a log-scale to allow for a better representation by reducing the variance in the extreme estimates for different domains):

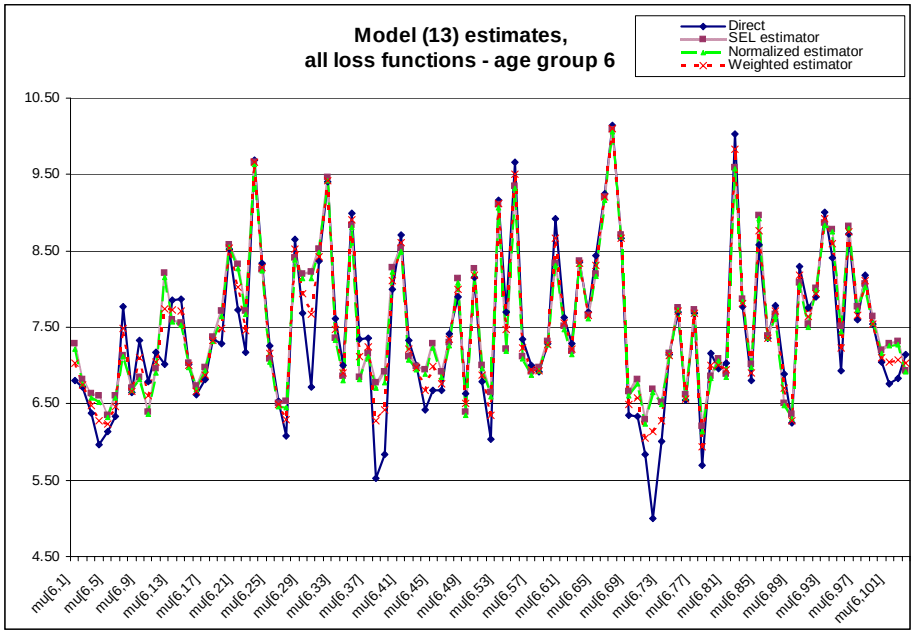


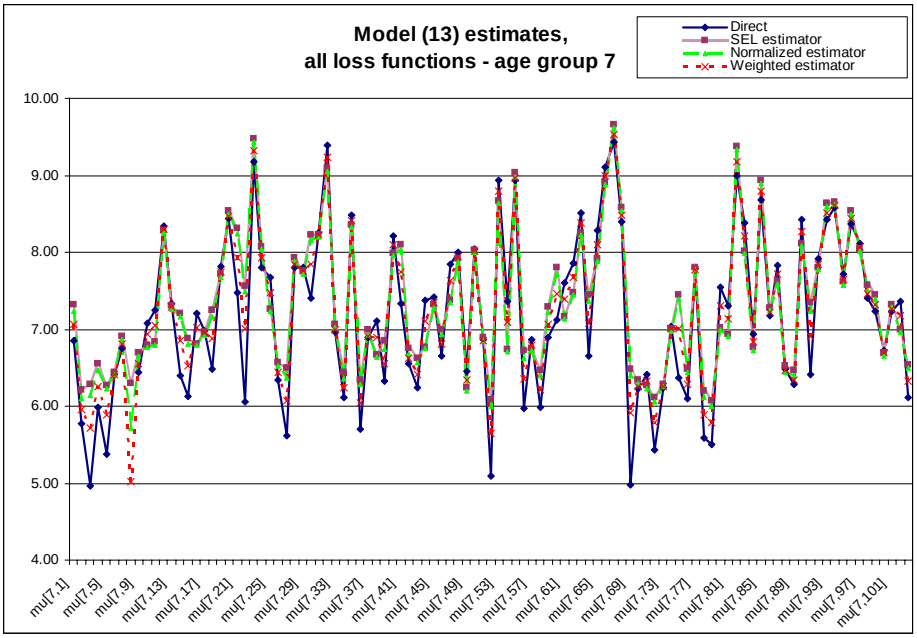


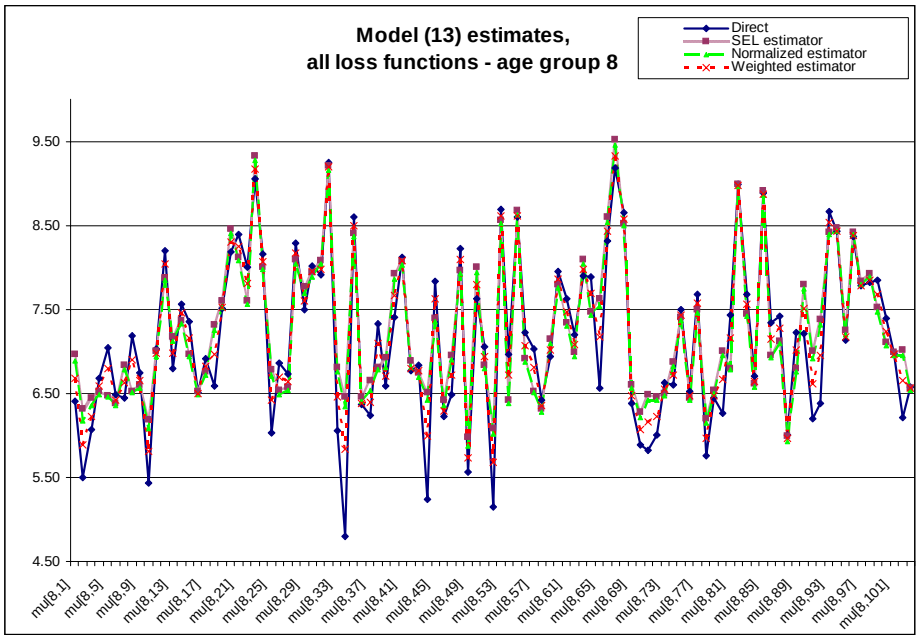

Model (13) estimates,
all loss functions - age group 3
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.50
9.50
8.50
7.50
6.50
5.50
4.50
mu[3,1]
mu[3,5]
mu[3,9]
mu[3,13]
mu[3,17]
mu[3,21]
mu[3,25]
mu[3,29]
mu[3,33]
mu[3,37]
mu[3,41]
mu[3,45]
mu[3,49]
mu[3,53]
mu[3,57]
mu[3,61]
mu[3,65]
mu[3,69]
mu[3,73]
mu[3,77]
mu[3,81]
mu[3,85]
mu[3,89]
mu[3,93]
mu[3,97]
mu[3,101]

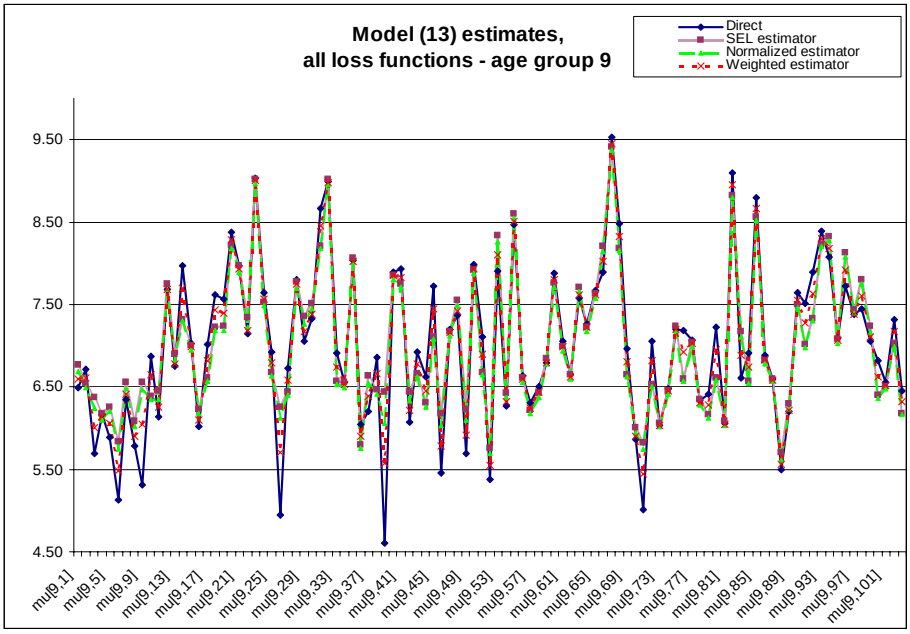

Model (13) estimates,
all loss functions - age group 4
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.50
9.50
8.50
7.50
6.50
5.50
4.50
3.50
mu[4,1]
mu[4,5]
mu[4,9]
mu[4,13]
mu[4,17]
mu[4,21]
mu[4,25]
mu[4,29]
mu[4,33]
mu[4,37]
mu[4,41]
mu[4,45]
mu[4,49]
mu[4,53]
mu[4,57]
mu[4,61]
mu[4,65]
mu[4,69]
mu[4,73]
mu[4,77]
mu[4,81]
mu[4,85]
mu[4,89]
mu[4,93]
mu[4,97]
mu[4,101]

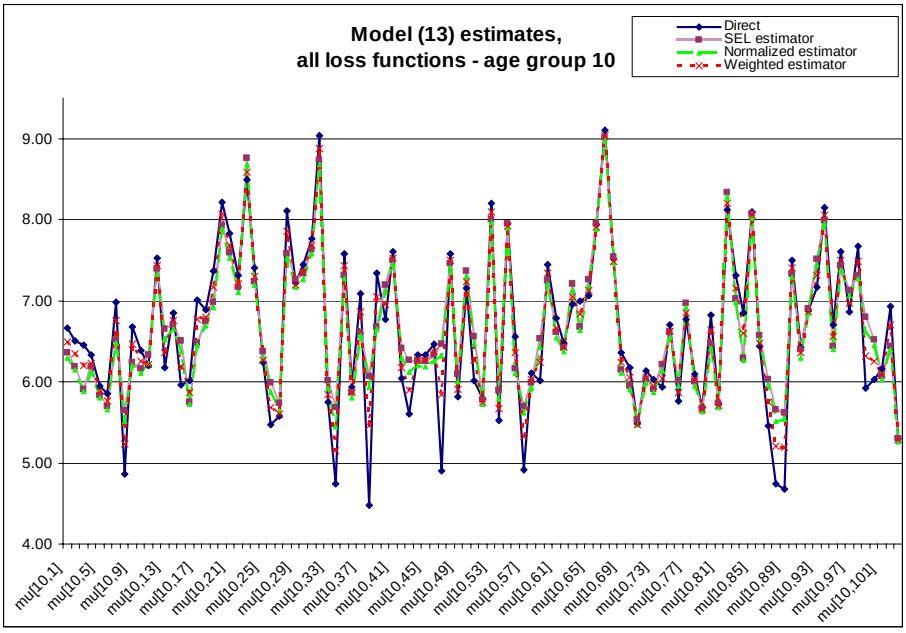

Model (13) estimates,
all loss functions - age group 5
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.50
9.50
8.50
7.50
6.50
5.50
4.50
mu[5,1]
mu[5,5]
mu[5,9]
mu[5,13]
mu[5,17]
mu[5,21]
mu[5,25]
mu[5,29]
mu[5,33]
mu[5,37]
mu[5,41]
mu[5,45]
mu[5,49]
mu[5,53]
mu[5,57]
mu[5,61]
mu[5,65]
mu[5,69]
mu[5,73]
mu[5,77]
mu[5,81]
mu[5,85]
mu[5,89]
mu[5,93]
mu[5,97]
mu[5,101]

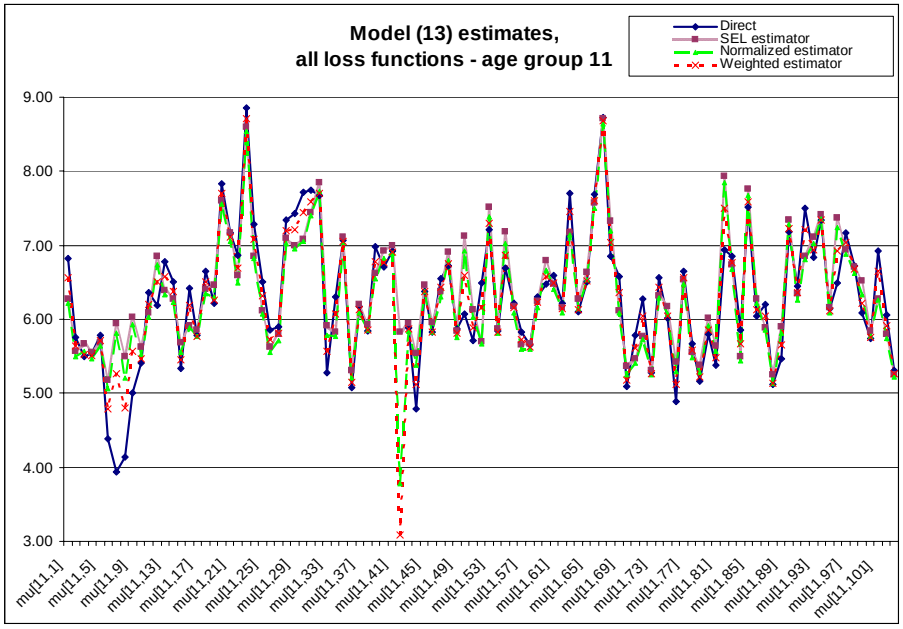

Model (13) estimates,
all loss functions - age group 6
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.50
9.50
8.50
7.50
6.50
5.50
4.50
mu[6,1]
mu[6,5]
mu[6,9]
mu[6,13]
mu[6,17]
mu[6,21]
mu[6,25]
mu[6,29]
mu[6,33]
mu[6,37]
mu[6,41]
mu[6,45]
mu[6,49]
mu[6,53]
mu[6,57]
mu[6,61]
mu[6,65]
mu[6,69]
mu[6,73]
mu[6,77]
mu[6,81]
mu[6,85]
mu[6,89]
mu[6,93]
mu[6,97]
mu[6,101]

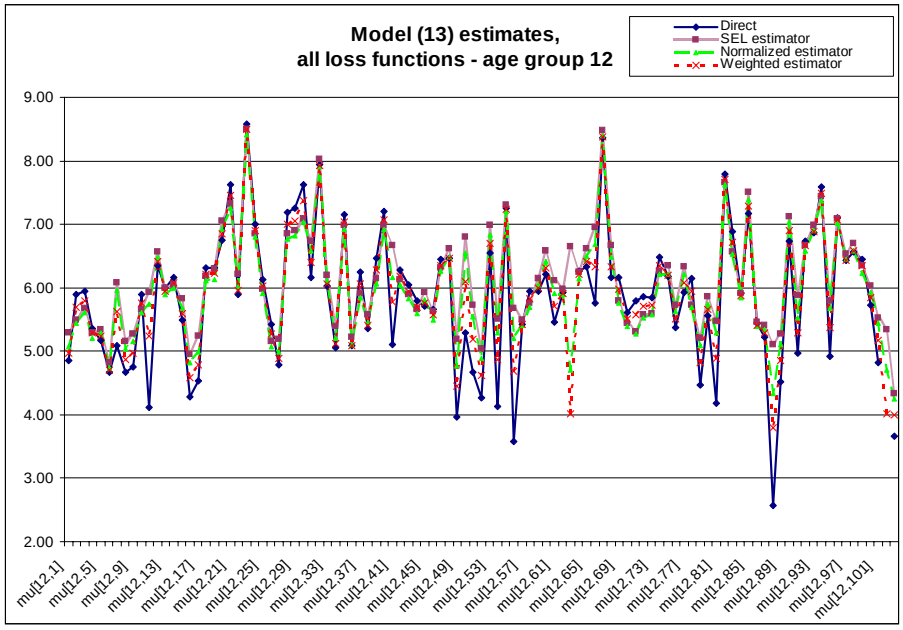

Model (13) estimates,
all loss functions - age group 7
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[7,1]
mu[7,5]
mu[7,9]
mu[7,13]
mu[7,17]
mu[7,21]
mu[7,25]
mu[7,29]
mu[7,33]
mu[7,37]
mu[7,41]
mu[7,45]
mu[7,49]
mu[7,53]
mu[7,57]
mu[7,61]
mu[7,65]
mu[7,69]
mu[7,73]
mu[7,77]
mu[7,81]
mu[7,85]
mu[7,89]
mu[7,93]
mu[7,97]
mu[7,101]

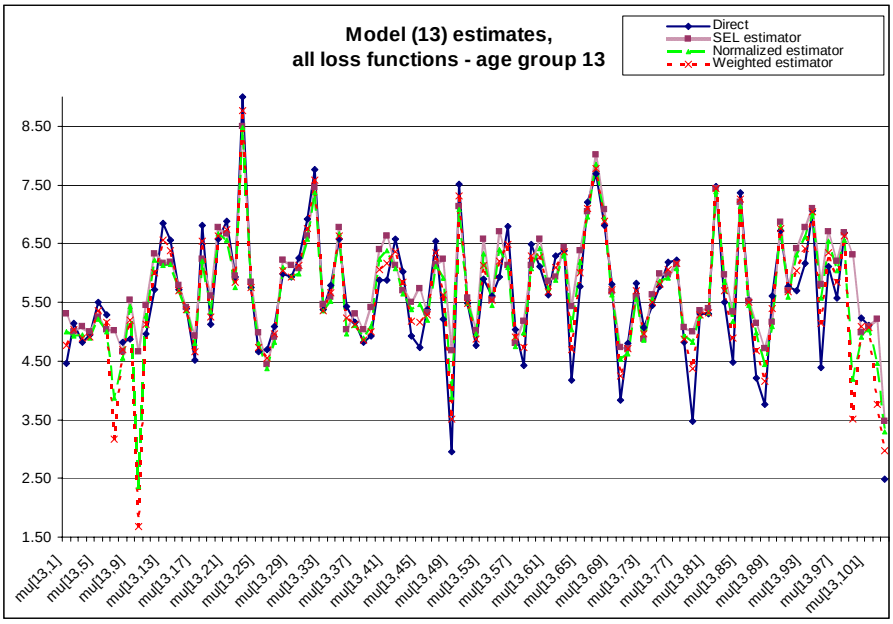

Model (13) estimates,
all loss functions - age group 8
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.50
8.50
7.50
6.50
5.50
4.50
mu[8,1]
mu[8,5]
mu[8,9]
mu[8,13]
mu[8,17]
mu[8,21]
mu[8,25]
mu[8,29]
mu[8,33]
mu[8,37]
mu[8,41]
mu[8,45]
mu[8,49]
mu[8,53]
mu[8,57]
mu[8,61]
mu[8,65]
mu[8,69]
mu[8,73]
mu[8,77]
mu[8,81]
mu[8,85]
mu[8,89]
mu[8,93]
mu[8,97]
mu[8,101]

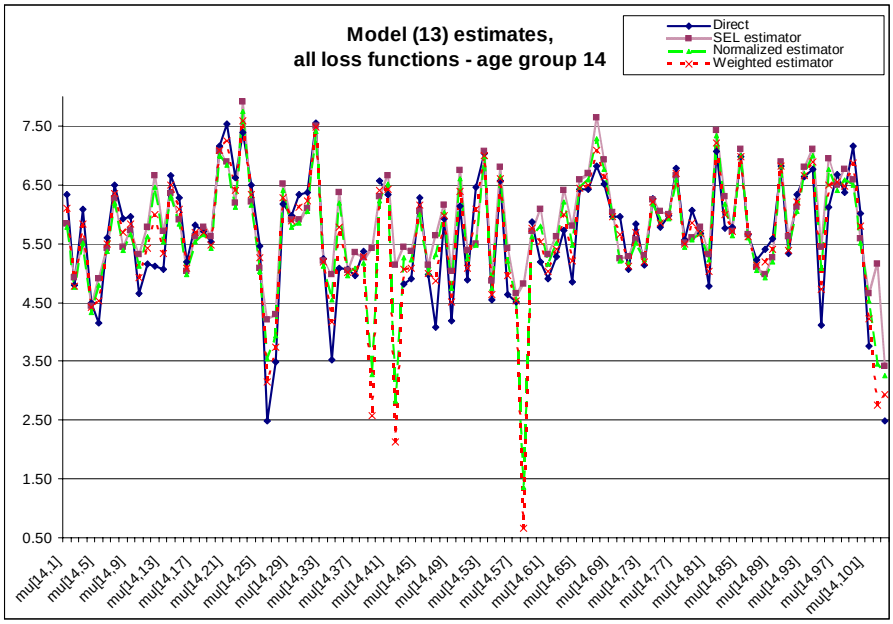

Model (13) estimates,
all loss functions - age group 9
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.50
8.50
7.50
6.50
5.50
4.50
mu[9,1]
mu[9,5]
mu[9,9]
mu[9,13]
mu[9,17]
mu[9,21]
mu[9,25]
mu[9,29]
mu[9,33]
mu[9,37]
mu[9,41]
mu[9,45]
mu[9,49]
mu[9,53]
mu[9,57]
mu[9,61]
mu[9,65]
mu[9,69]
mu[9,73]
mu[9,77]
mu[9,81]
mu[9,85]
mu[9,89]
mu[9,93]
mu[9,97]
mu[9,101]


Model (13) estimates,
all loss functions - age group 10
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.00
8.00
7.00
6.00
5.00
4.00
mu[10,1]
mu[10,5]
mu[10,9]
mu[10,13]
mu[10,17]
mu[10,21]
mu[10,25]
mu[10,29]
mu[10,33]
mu[10,37]
mu[10,41]
mu[10,45]
mu[10,49]
mu[10,53]
mu[10,57]
mu[10,61]
mu[10,65]
mu[10,69]
mu[10,73]
mu[10,77]
mu[10,81]
mu[10,85]
mu[10,89]
mu[10,93]
mu[10,97]
mu[10,101]

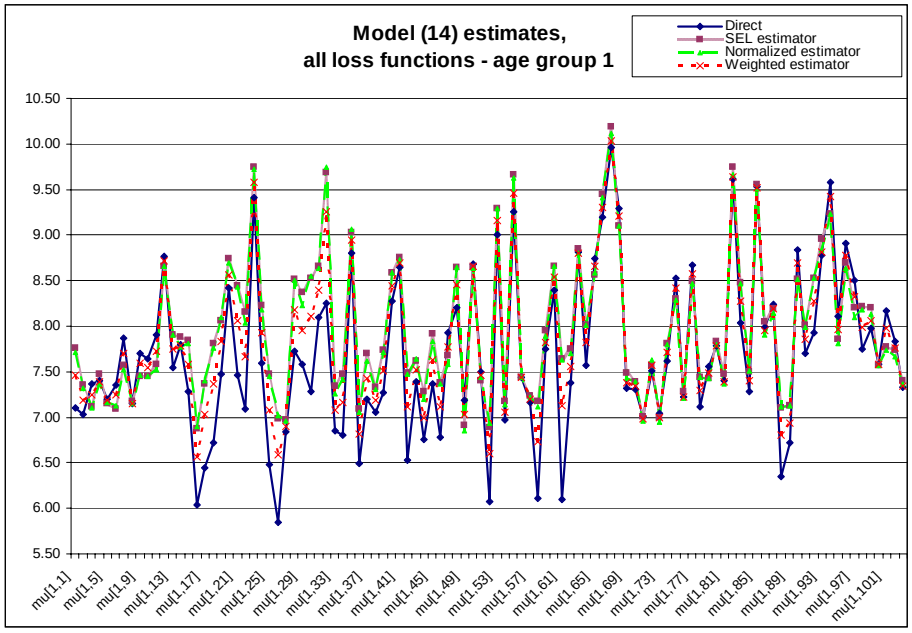

Model (13) estimates,
all loss functions - age group 11
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.00
8.00
7.00
6.00
5.00
4.00
3.00
mu[11,1]
mu[11,5]
mu[11,9]
mu[11,13]
mu[11,17]
mu[11,21]
mu[11,25]
mu[11,29]
mu[11,33]
mu[11,37]
mu[11,41]
mu[11,45]
mu[11,49]
mu[11,53]
mu[11,57]
mu[11,61]
mu[11,65]
mu[11,69]
mu[11,73]
mu[11,77]
mu[11,81]
mu[11,85]
mu[11,89]
mu[11,93]
mu[11,97]
mu[11,101]

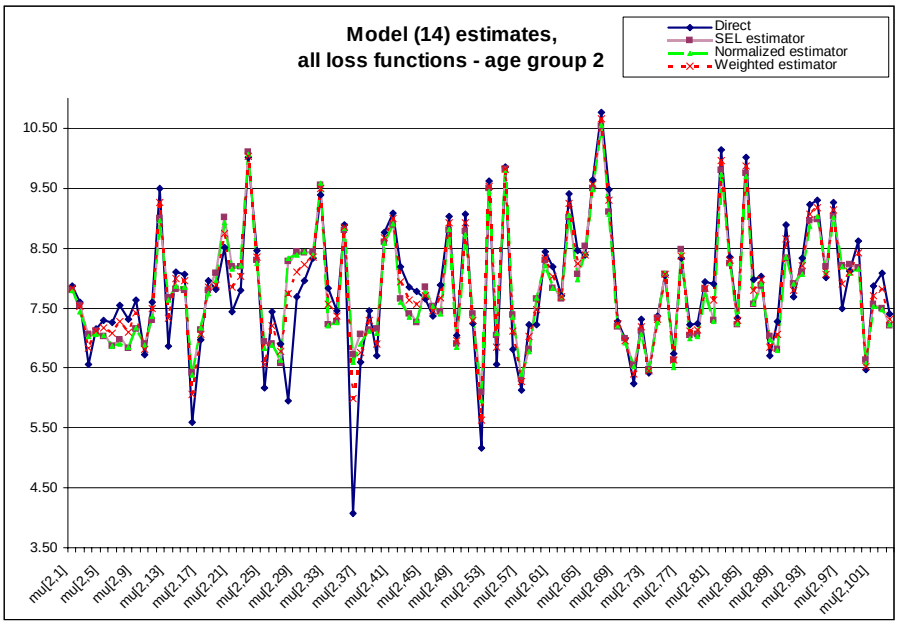

Model (13) estimates,
all loss functions - age group 12
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
mu[12,1]
mu[12,5]
mu[12,9]
mu[12,13]
mu[12,17]
mu[12,21]
mu[12,25]
mu[12,29]
mu[12,33]
mu[12,37]
mu[12,41]
mu[12,45]
mu[12,49]
mu[12,53]
mu[12,57]
mu[12,61]
mu[12,65]
mu[12,69]
mu[12,73]
mu[12,77]
mu[12,81]
mu[12,85]
mu[12,89]
mu[12,93]
mu[12,97]
mu[12,101]

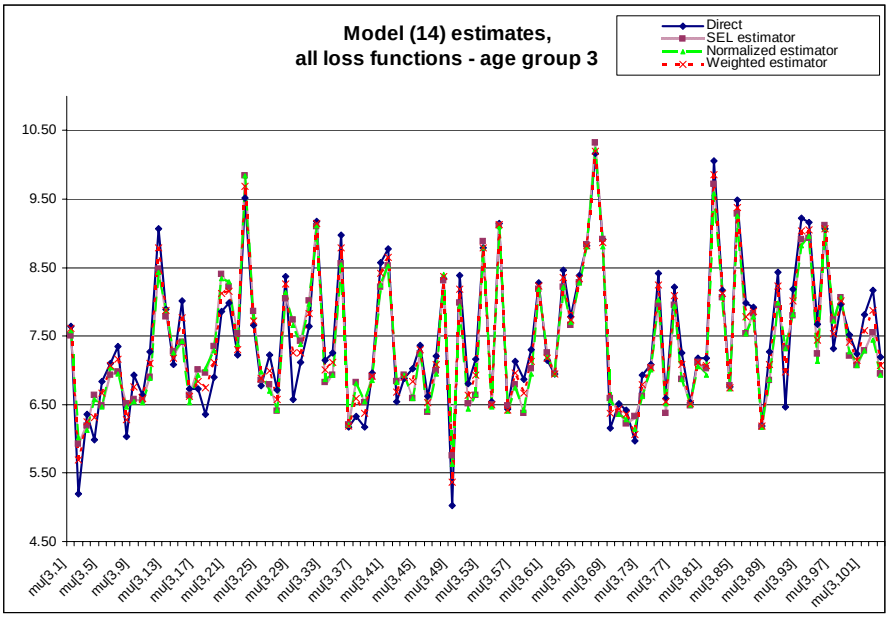

Model (13) estimates,
all loss functions - age group 13
Direct
SEL estimator
Normalized estimator
Weighted estimator
8.50
7.50
6.50
5.50
4.50
3.50
2.50
1.50
mu[13,1]
mu[13,5]
mu[13,9]
mu[13,13]
mu[13,17]
mu[13,21]
mu[13,25]
mu[13,29]
mu[13,33]
mu[13,37]
mu[13,41]
mu[13,45]
mu[13,49]
mu[13,53]
mu[13,57]
mu[13,61]
mu[13,65]
mu[13,69]
mu[13,73]
mu[13,77]
mu[13,81]
mu[13,85]
mu[13,89]
mu[13,93]
mu[13,97]
mu[13,101]

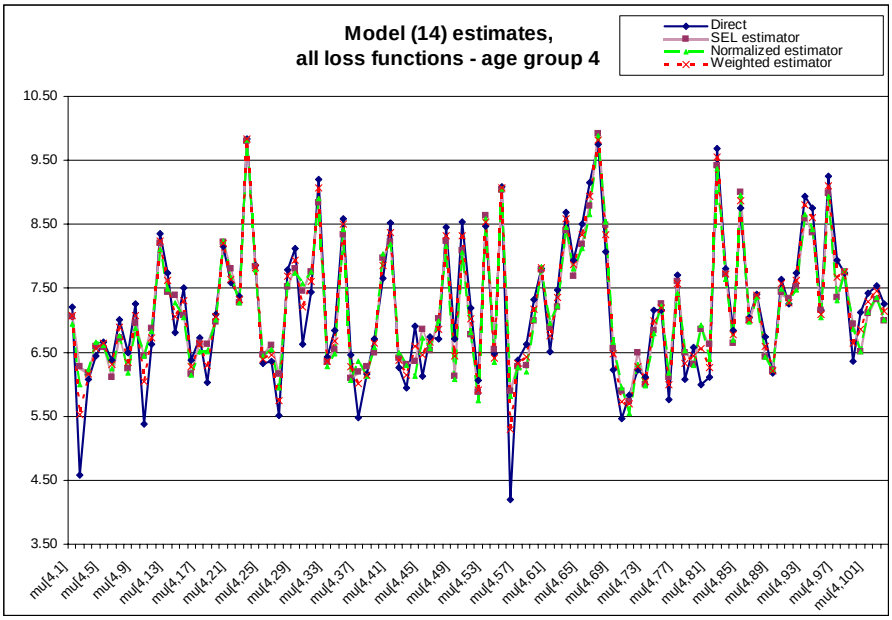

Model (13) estimates,
all loss functions - age group 14
Direct
SEL estimator
Normalized estimator
Weighted estimator
7.50
6.50
5.50
4.50
3.50
2.50
1.50
0.50
mu[14,1]
mu[14,5]
mu[14,9]
mu[14,13]
mu[14,17]
mu[14,21]
mu[14,25]
mu[14,29]
mu[14,33]
mu[14,37]
mu[14,41]
mu[14,45]
mu[14,49]
mu[14,53]
mu[14,57]
mu[14,61]
mu[14,65]
mu[14,69]
mu[14,73]
mu[14,77]
mu[14,81]
mu[14,85]
mu[14,89]
mu[14,93]
mu[14,97]
mu[14,101]

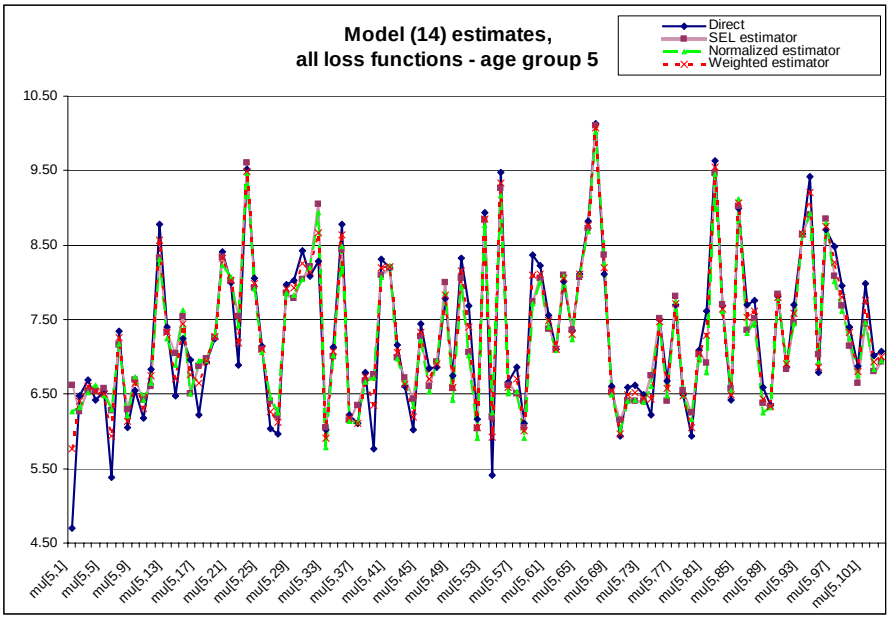
Model (14) estimates (estimates are graphed on a log-scale to allow for a better representation by reducing the variance in the extreme estimates for different domains):

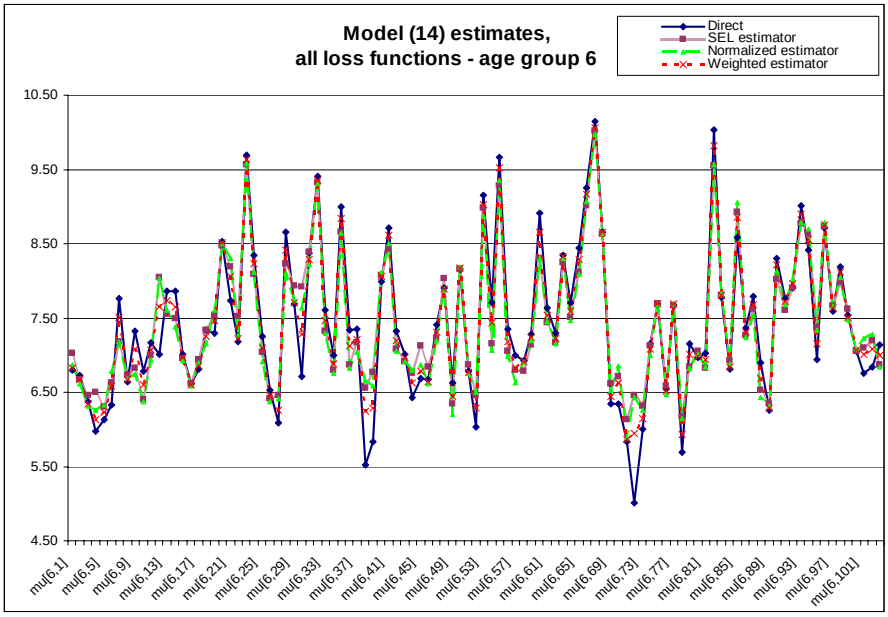


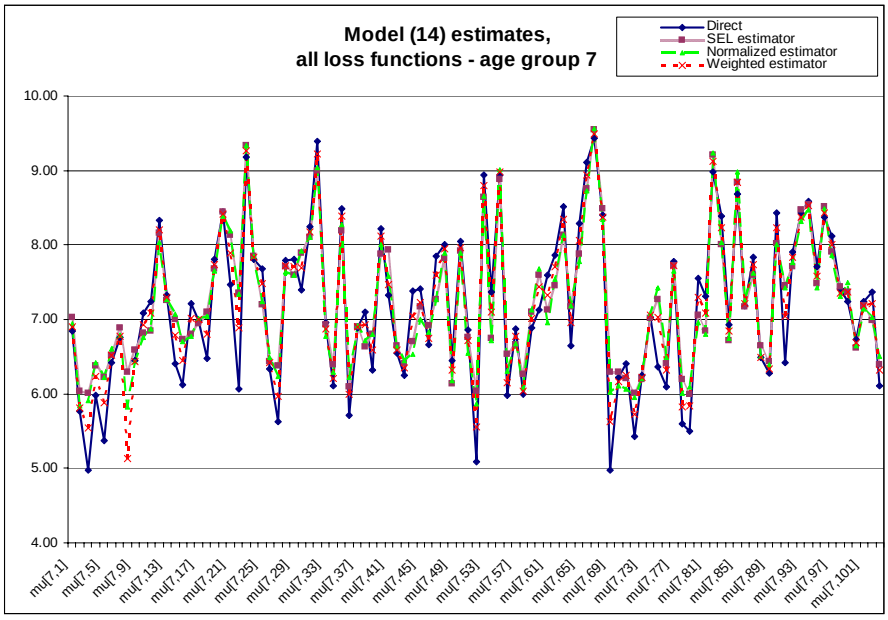


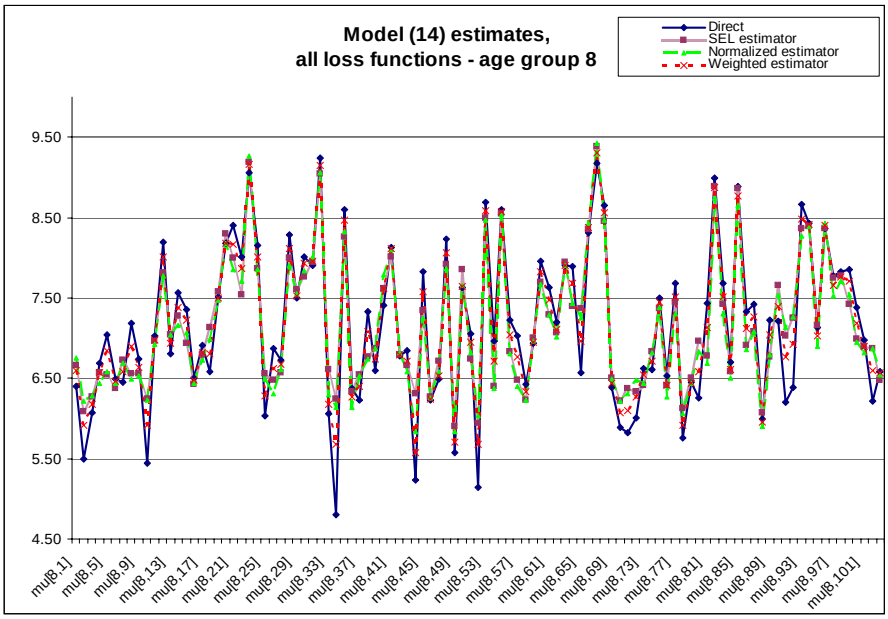

Model (14) estimates,
all loss functions - age group 3
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.50
9.50
8.50
7.50
6.50
5.50
4.50
mu[3,1]
mu[3,5]
mu[3,9]
mu[3,13]
mu[3,17]
mu[3,21]
mu[3,25]
mu[3,29]
mu[3,33]
mu[3,37]
mu[3,41]
mu[3,45]
mu[3,49]
mu[3,53]
mu[3,57]
mu[3,61]
mu[3,65]
mu[3,69]
mu[3,73]
mu[3,77]
mu[3,81]
mu[3,85]
mu[3,89]
mu[3,93]
mu[3,97]
mu[3,101]

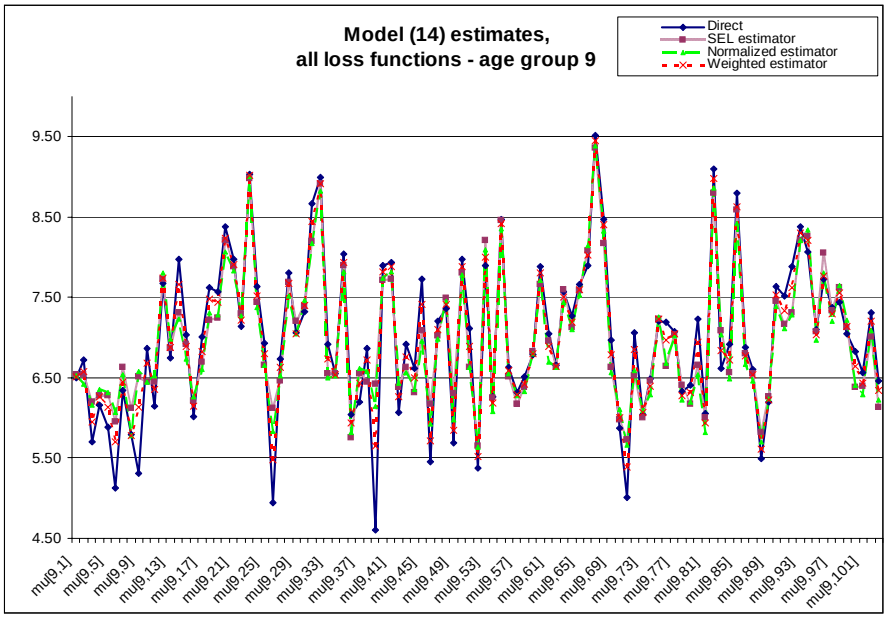

Model (14) estimates,
all loss functions - age group 4
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.50
9.50
8.50
7.50
6.50
5.50
4.50
3.50
mu[4,1]
mu[4,5]
mu[4,9]
mu[4,13]
mu[4,17]
mu[4,21]
mu[4,25]
mu[4,29]
mu[4,33]
mu[4,37]
mu[4,41]
mu[4,45]
mu[4,49]
mu[4,53]
mu[4,57]
mu[4,61]
mu[4,65]
mu[4,69]
mu[4,73]
mu[4,77]
mu[4,81]
mu[4,85]
mu[4,89]
mu[4,93]
mu[4,97]
mu[4,101]

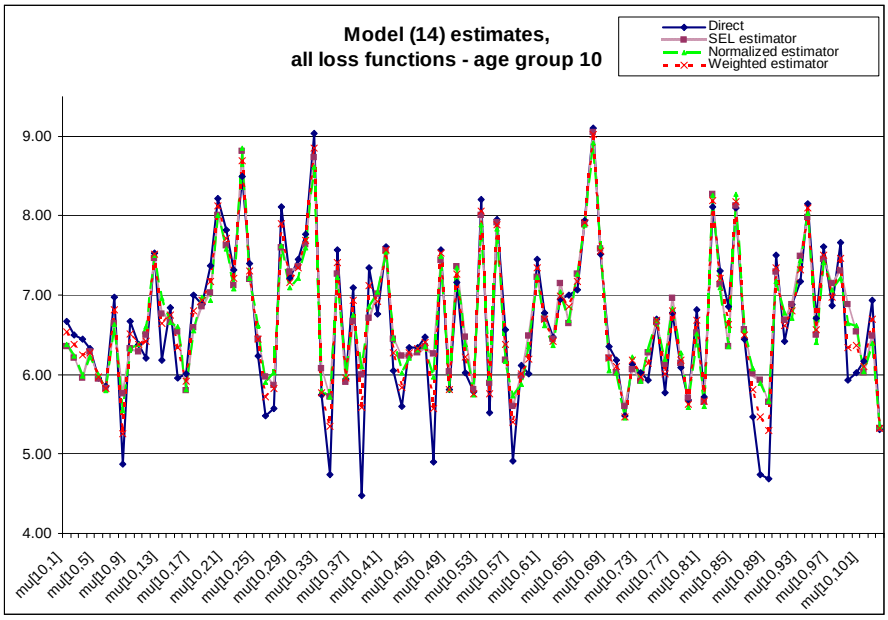

Model (14) estimates,
all loss functions - age group 5
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.50
9.50
8.50
7.50
6.50
5.50
4.50
mu[5,1]
mu[5,5]
mu[5,9]
mu[5,13]
mu[5,17]
mu[5,21]
mu[5,25]
mu[5,29]
mu[5,33]
mu[5,37]
mu[5,41]
mu[5,45]
mu[5,49]
mu[5,53]
mu[5,57]
mu[5,61]
mu[5,65]
mu[5,69]
mu[5,73]
mu[5,77]
mu[5,81]
mu[5,85]
mu[5,89]
mu[5,93]
mu[5,97]
mu[5,101]

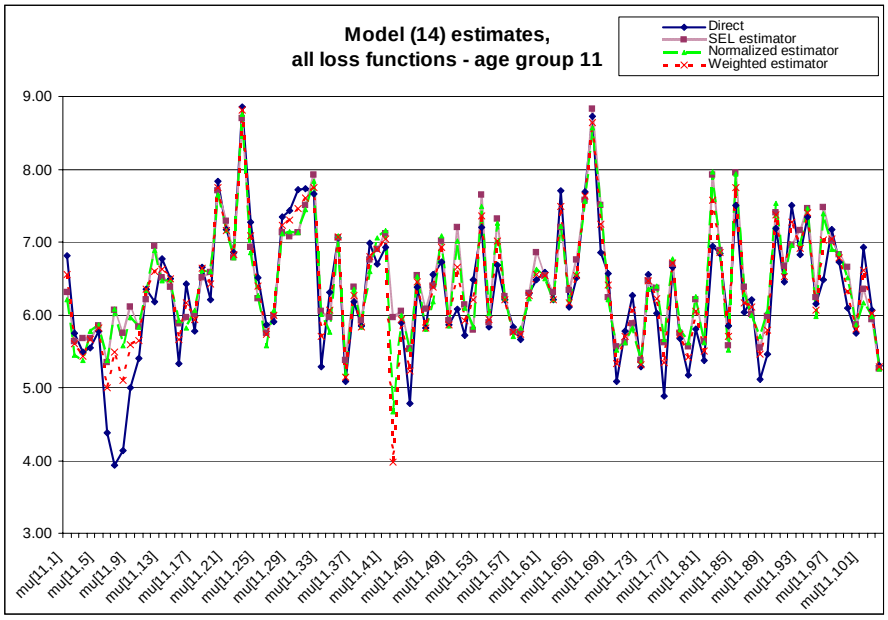

Model (14) estimates,
all loss functions - age group 6
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.50
9.50
8.50
7.50
6.50
5.50
4.50
mu[6,1]
mu[6,5]
mu[6,9]
mu[6,13]
mu[6,17]
mu[6,21]
mu[6,25]
mu[6,29]
mu[6,33]
mu[6,37]
mu[6,41]
mu[6,45]
mu[6,49]
mu[6,53]
mu[6,57]
mu[6,61]
mu[6,65]
mu[6,69]
mu[6,73]
mu[6,77]
mu[6,81]
mu[6,85]
mu[6,89]
mu[6,93]
mu[6,97]
mu[6,101]

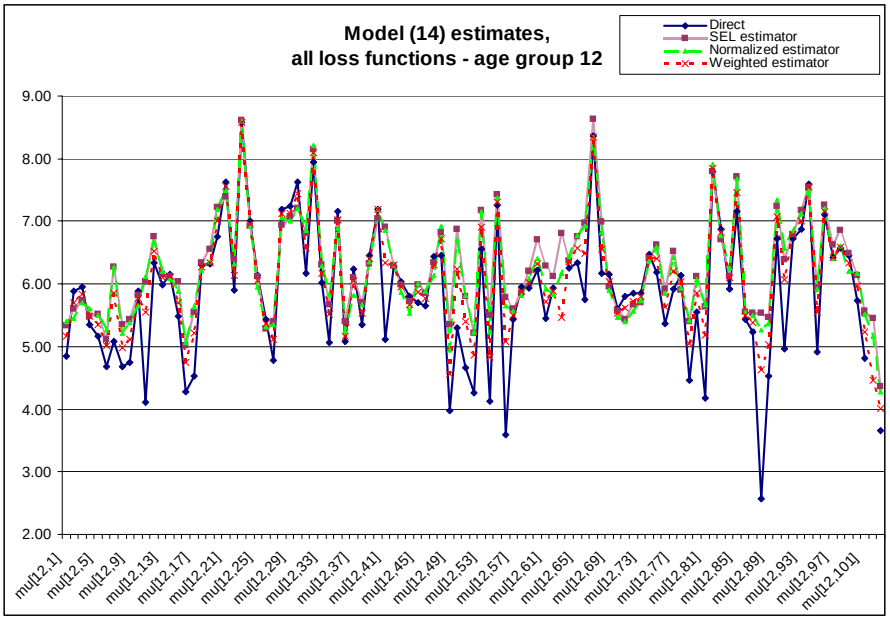

Model (14) estimates,
all loss functions - age group 7
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[7,1]
mu[7,5]
mu[7,9]
mu[7,13]
mu[7,17]
mu[7,21]
mu[7,25]
mu[7,29]
mu[7,33]
mu[7,37]
mu[7,41]
mu[7,45]
mu[7,49]
mu[7,53]
mu[7,57]
mu[7,61]
mu[7,65]
mu[7,69]
mu[7,73]
mu[7,77]
mu[7,81]
mu[7,85]
mu[7,89]
mu[7,93]
mu[7,97]
mu[7,101]

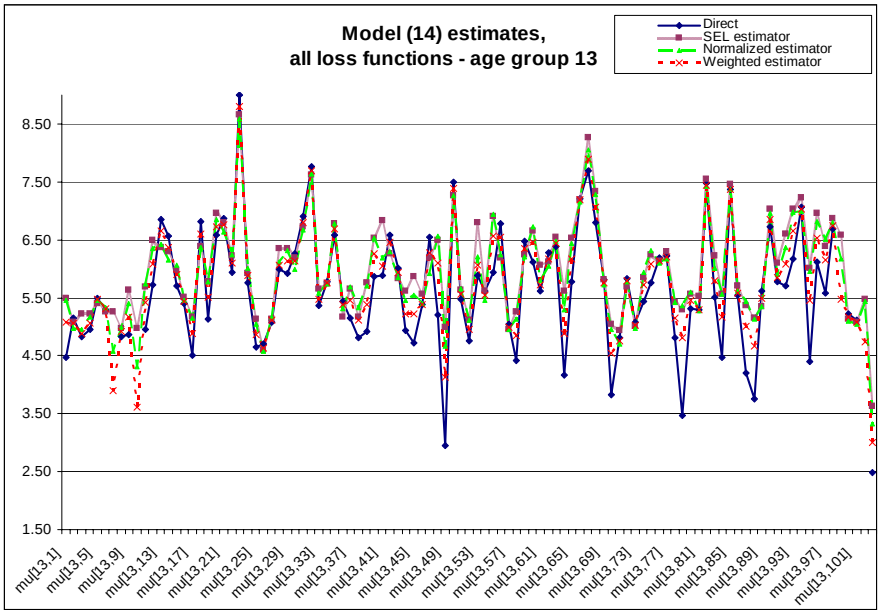

Model (14) estimates,
all loss functions - age group 8
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.50
8.50
7.50
6.50
5.50
4.50
mu[8,1]
mu[8,5]
mu[8,9]
mu[8,13]
mu[8,17]
mu[8,21]
mu[8,25]
mu[8,29]
mu[8,33]
mu[8,37]
mu[8,41]
mu[8,45]
mu[8,49]
mu[8,53]
mu[8,57]
mu[8,61]
mu[8,65]
mu[8,69]
mu[8,73]
mu[8,77]
mu[8,81]
mu[8,85]
mu[8,89]
mu[8,93]
mu[8,97]
mu[8,101]

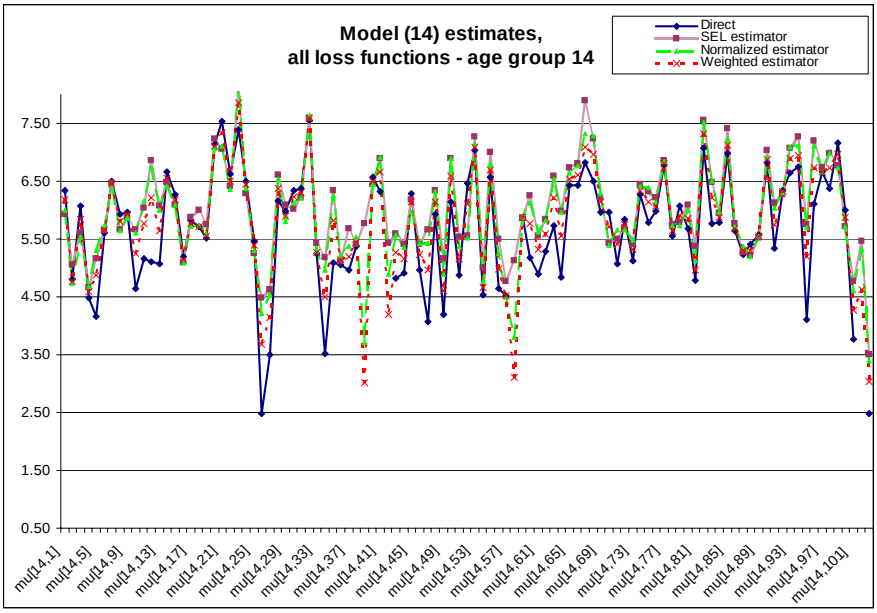

Model (14) estimates,
all loss functions - age group 9
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.50
8.50
7.50
6.50
5.50
4.50
mu[9,1]
mu[9,5]
mu[9,9]
mu[9,13]
mu[9,17]
mu[9,21]
mu[9,25]
mu[9,29]
mu[9,33]
mu[9,37]
mu[9,41]
mu[9,45]
mu[9,49]
mu[9,53]
mu[9,57]
mu[9,61]
mu[9,65]
mu[9,69]
mu[9,73]
mu[9,77]
mu[9,81]
mu[9,85]
mu[9,89]
mu[9,93]
mu[9,97]
mu[9,101]


Model (14) estimates,
all loss functions - age group 10
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.00
8.00
7.00
6.00
5.00
4.00
mu[10,1]
mu[10,5]
mu[10,9]
mu[10,13]
mu[10,17]
mu[10,21]
mu[10,25]
mu[10,29]
mu[10,33]
mu[10,37]
mu[10,41]
mu[10,45]
mu[10,49]
mu[10,53]
mu[10,57]
mu[10,61]
mu[10,65]
mu[10,69]
mu[10,73]
mu[10,77]
mu[10,81]
mu[10,85]
mu[10,89]
mu[10,93]
mu[10,97]
mu[10,101]

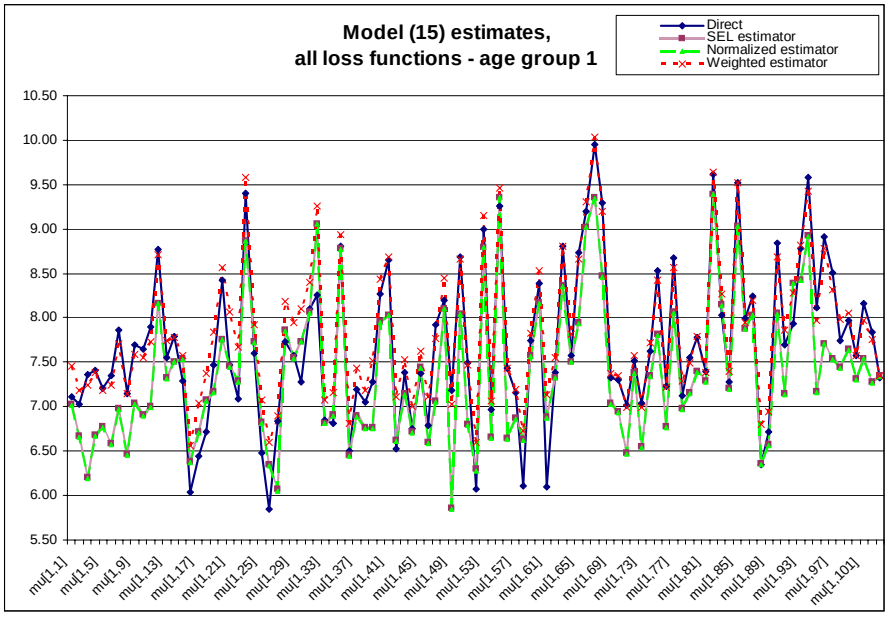

Model (14) estimates,
all loss functions - age group 11
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.00
8.00
7.00
6.00
5.00
4.00
3.00
mu[11,1]
mu[11,5]
mu[11,9]
mu[11,13]
mu[11,17]
mu[11,21]
mu[11,25]
mu[11,29]
mu[11,33]
mu[11,37]
mu[11,41]
mu[11,45]
mu[11,49]
mu[11,53]
mu[11,57]
mu[11,61]
mu[11,65]
mu[11,69]
mu[11,73]
mu[11,77]
mu[11,81]
mu[11,85]
mu[11,89]
mu[11,93]
mu[11,97]
mu[11,101]

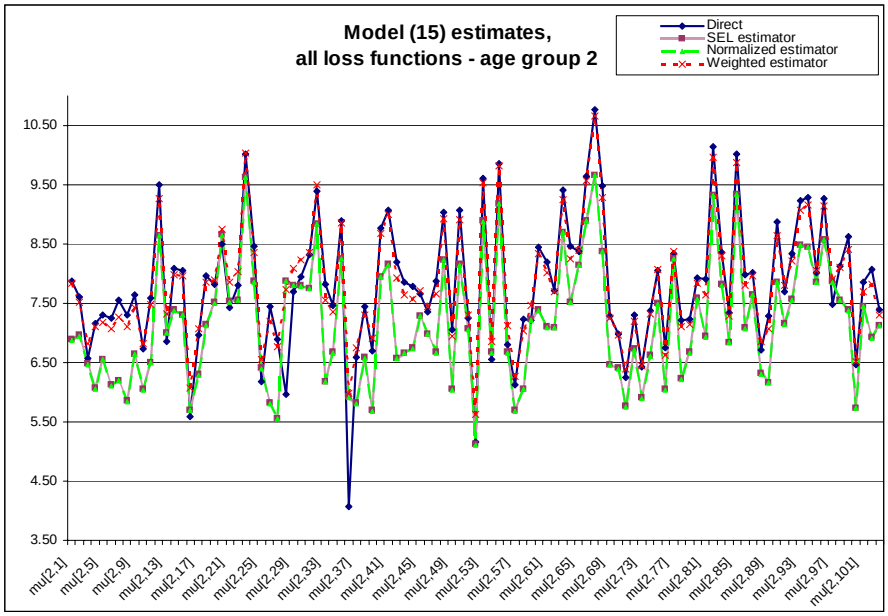

Model (14) estimates,
all loss functions - age group 12
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
mu[12,1]
mu[12,5]
mu[12,9]
mu[12,13]
mu[12,17]
mu[12,21]
mu[12,25]
mu[12,29]
mu[12,33]
mu[12,37]
mu[12,41]
mu[12,45]
mu[12,49]
mu[12,53]
mu[12,57]
mu[12,61]
mu[12,65]
mu[12,69]
mu[12,73]
mu[12,77]
mu[12,81]
mu[12,85]
mu[12,89]
mu[12,93]
mu[12,97]
mu[12,101]

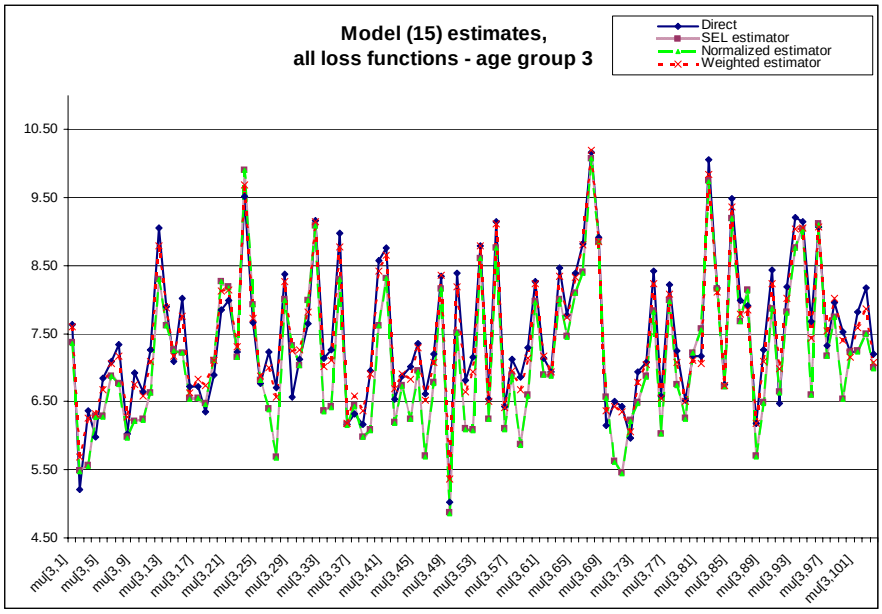

Model (14) estimates,
all loss functions - age group 13
Direct
SEL estimator
Normalized estimator
Weighted estimator
8.50
7.50
6.50
5.50
4.50
3.50
2.50
1.50
mu[13,1]
mu[13,5]
mu[13,9]
mu[13,13]
mu[13,17]
mu[13,21]
mu[13,25]
mu[13,29]
mu[13,33]
mu[13,37]
mu[13,41]
mu[13,45]
mu[13,49]
mu[13,53]
mu[13,57]
mu[13,61]
mu[13,65]
mu[13,69]
mu[13,73]
mu[13,77]
mu[13,81]
mu[13,85]
mu[13,89]
mu[13,93]
mu[13,97]
mu[13,101]

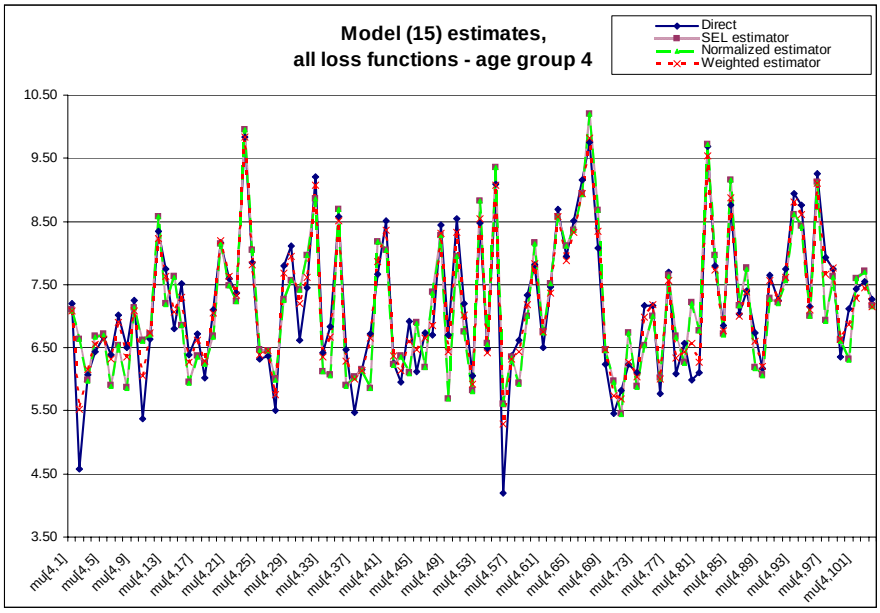

Model (14) estimates,
all loss functions - age group 14
Direct
SEL estimator
Normalized estimator
Weighted estimator
7.50
6.50
5.50
4.50
3.50
2.50
1.50
0.50
mu[14,1]
mu[14,5]
mu[14,9]
mu[14,13]
mu[14,17]
mu[14,21]
mu[14,25]
mu[14,29]
mu[14,33]
mu[14,37]
mu[14,41]
mu[14,45]
mu[14,49]
mu[14,53]
mu[14,57]
mu[14,61]
mu[14,65]
mu[14,69]
mu[14,73]
mu[14,77]
mu[14,81]
mu[14,85]
mu[14,89]
mu[14,93]
mu[14,97]
mu[14,101]

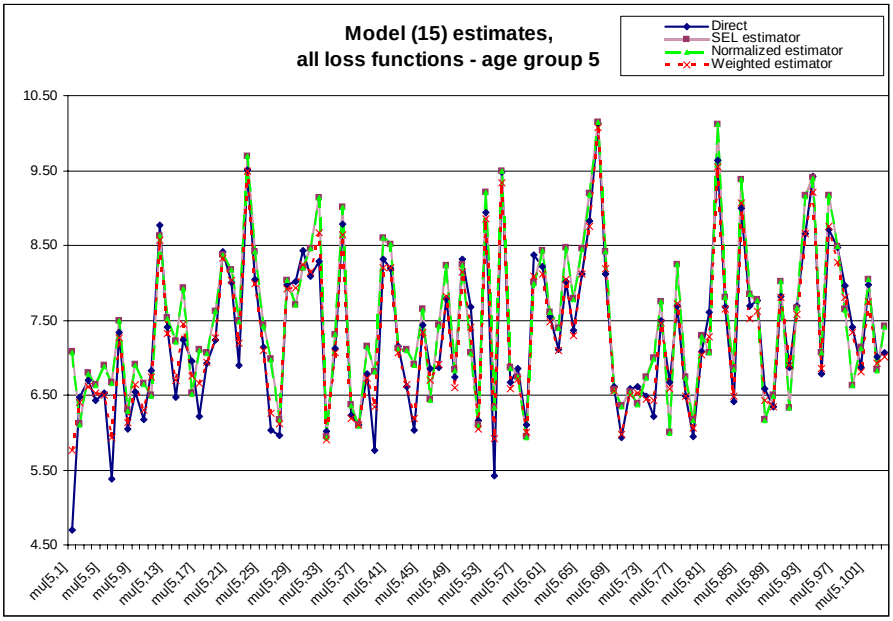
Model (15) estimates (estimates are graphed on a log-scale to allow for a better representation by reducing the variance in the extreme estimates for different domains):

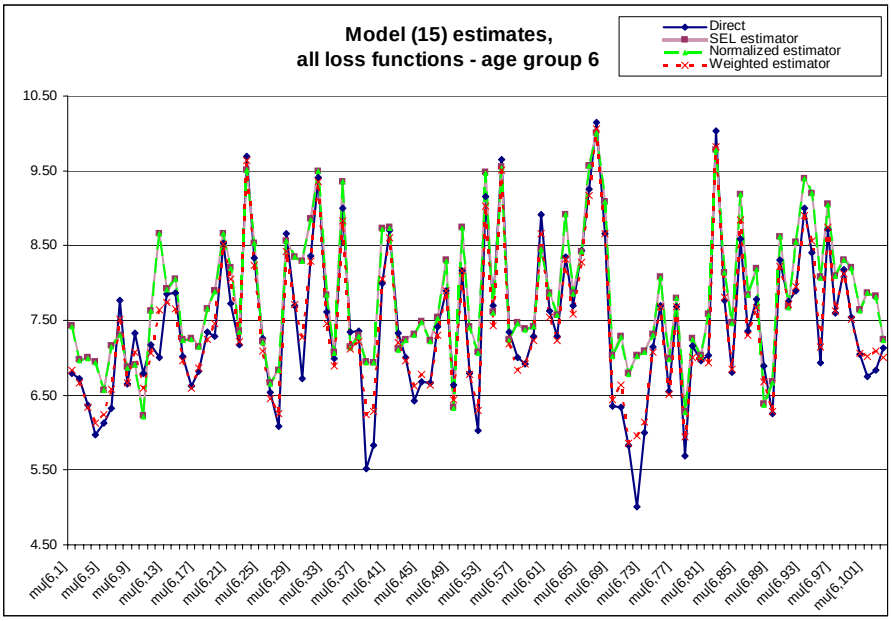


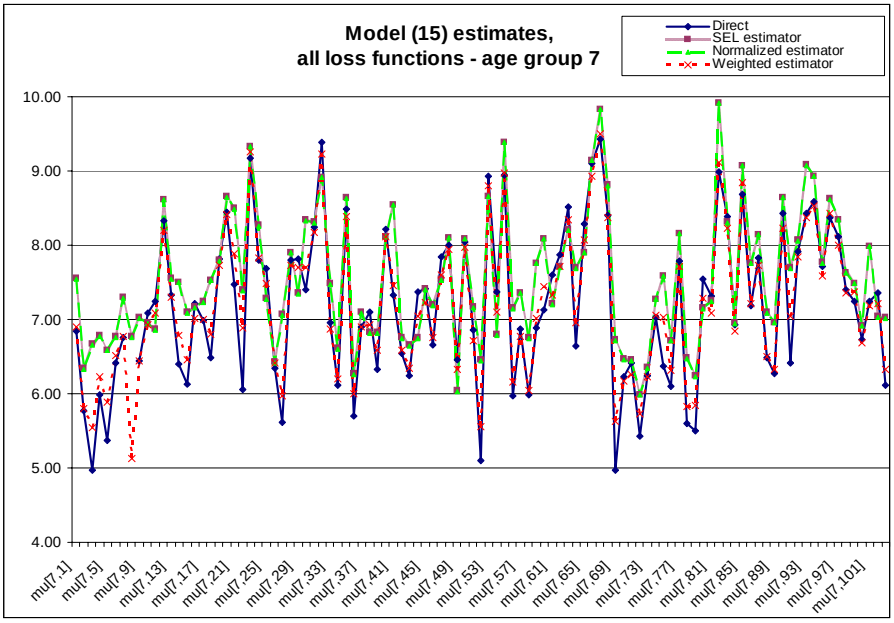


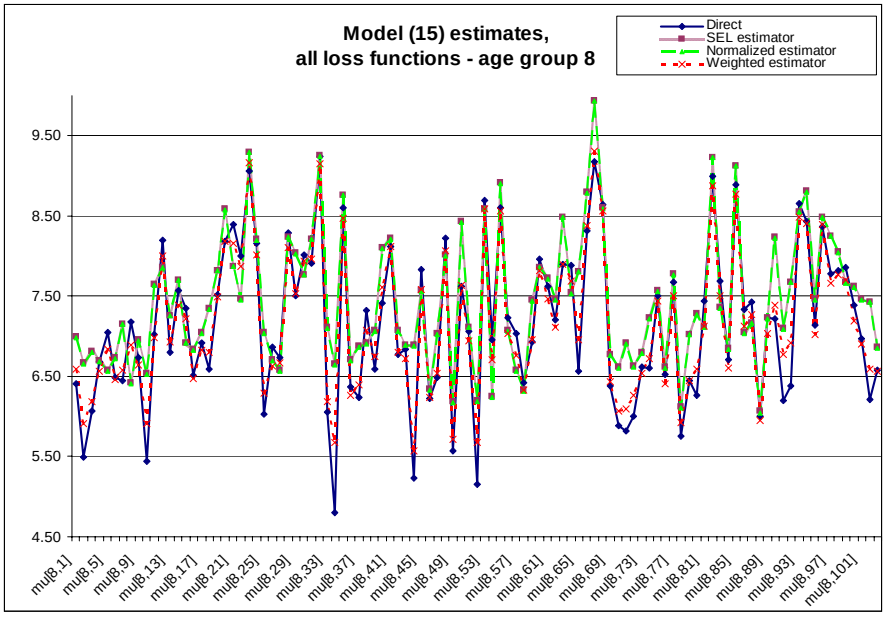

Model (15) estimates,
all loss functions - age group 3
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.50
9.50
8.50
7.50
6.50
5.50
4.50
mu[3,1]
mu[3,5]
mu[3,9]
mu[3,13]
mu[3,17]
mu[3,21]
mu[3,25]
mu[3,29]
mu[3,33]
mu[3,37]
mu[3,41]
mu[3,45]
mu[3,49]
mu[3,53]
mu[3,57]
mu[3,61]
mu[3,65]
mu[3,69]
mu[3,73]
mu[3,77]
mu[3,81]
mu[3,85]
mu[3,89]
mu[3,93]
mu[3,97]
mu[3,101]

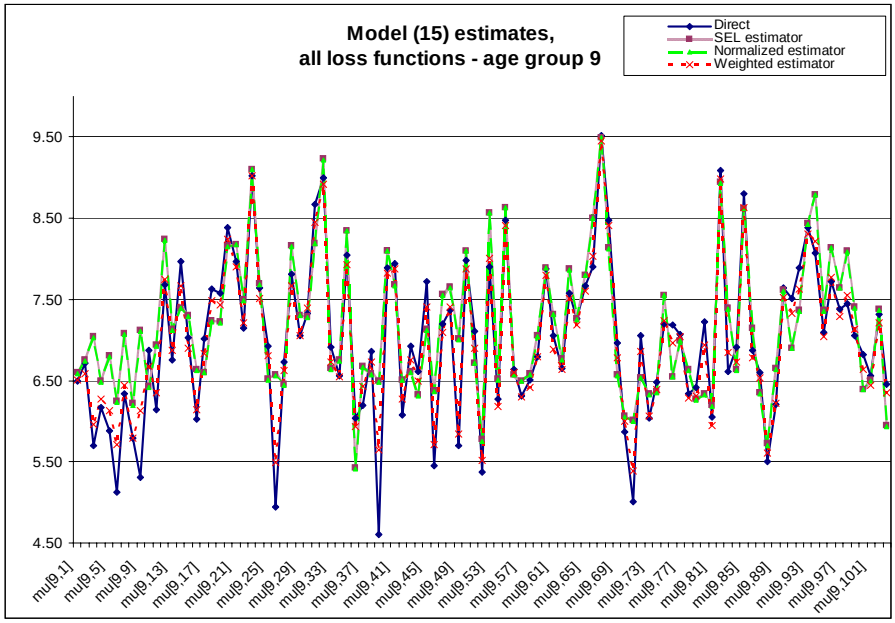

Model (15) estimates,
all loss functions - age group 4
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.50
9.50
8.50
7.50
6.50
5.50
4.50
3.50
mu[4,1]
mu[4,5]
mu[4,9]
mu[4,13]
mu[4,17]
mu[4,21]
mu[4,25]
mu[4,29]
mu[4,33]
mu[4,37]
mu[4,41]
mu[4,45]
mu[4,49]
mu[4,53]
mu[4,57]
mu[4,61]
mu[4,65]
mu[4,69]
mu[4,73]
mu[4,77]
mu[4,81]
mu[4,85]
mu[4,89]
mu[4,93]
mu[4,97]
mu[4,101]

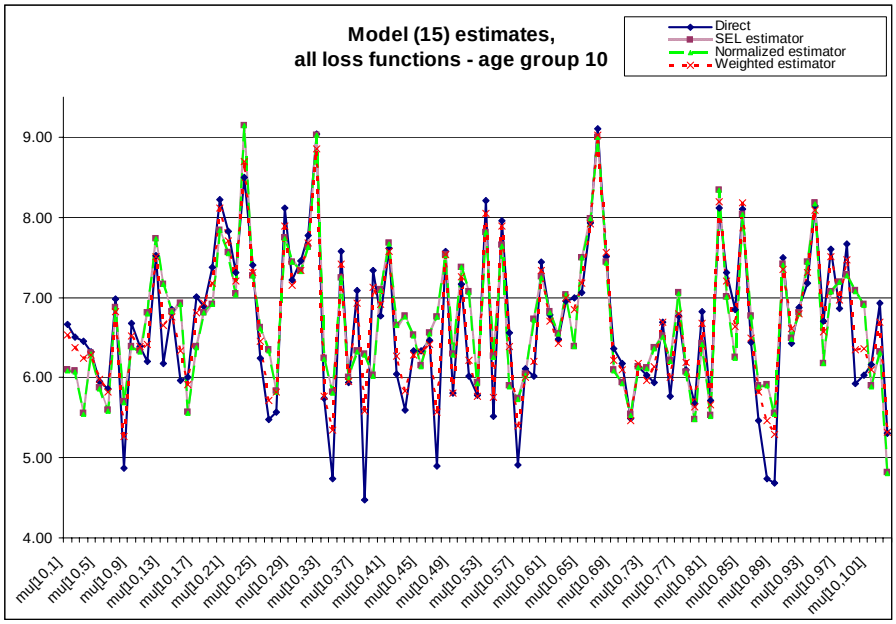

Model (15) estimates,
all loss functions - age group 5
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.50
9.50
8.50
7.50
6.50
5.50
4.50
mu[5,1]
mu[5,5]
mu[5,9]
mu[5,13]
mu[5,17]
mu[5,21]
mu[5,25]
mu[5,29]
mu[5,33]
mu[5,37]
mu[5,41]
mu[5,45]
mu[5,49]
mu[5,53]
mu[5,57]
mu[5,61]
mu[5,65]
mu[5,69]
mu[5,73]
mu[5,77]
mu[5,81]
mu[5,85]
mu[5,89]
mu[5,93]
mu[5,97]
mu[5,101]

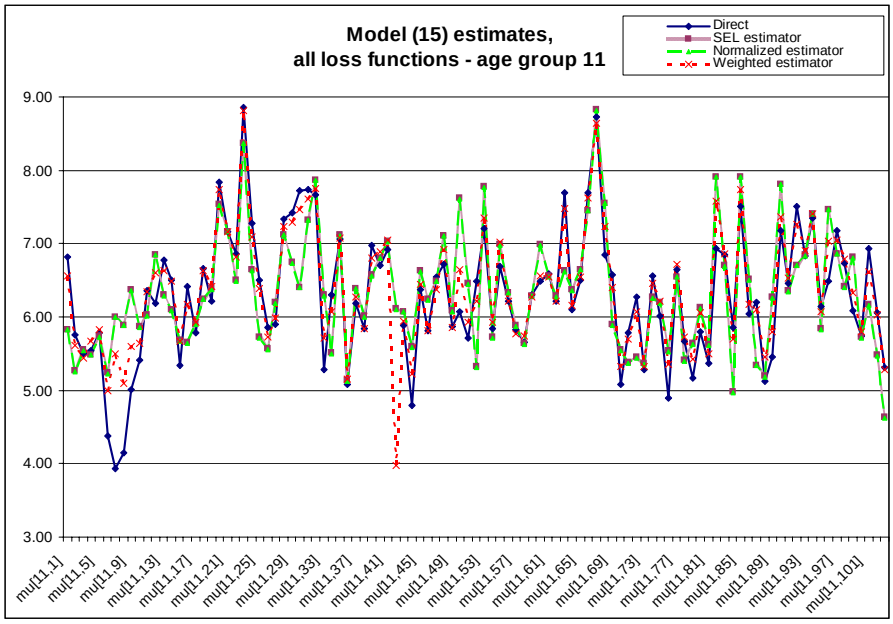

Model (15) estimates,
all loss functions - age group 6
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.50
9.50
8.50
7.50
6.50
5.50
4.50
mu[6,1]
mu[6,5]
mu[6,9]
mu[6,13]
mu[6,17]
mu[6,21]
mu[6,25]
mu[6,29]
mu[6,33]
mu[6,37]
mu[6,41]
mu[6,45]
mu[6,49]
mu[6,53]
mu[6,57]
mu[6,61]
mu[6,65]
mu[6,69]
mu[6,73]
mu[6,77]
mu[6,81]
mu[6,85]
mu[6,89]
mu[6,93]
mu[6,97]
mu[6,101]

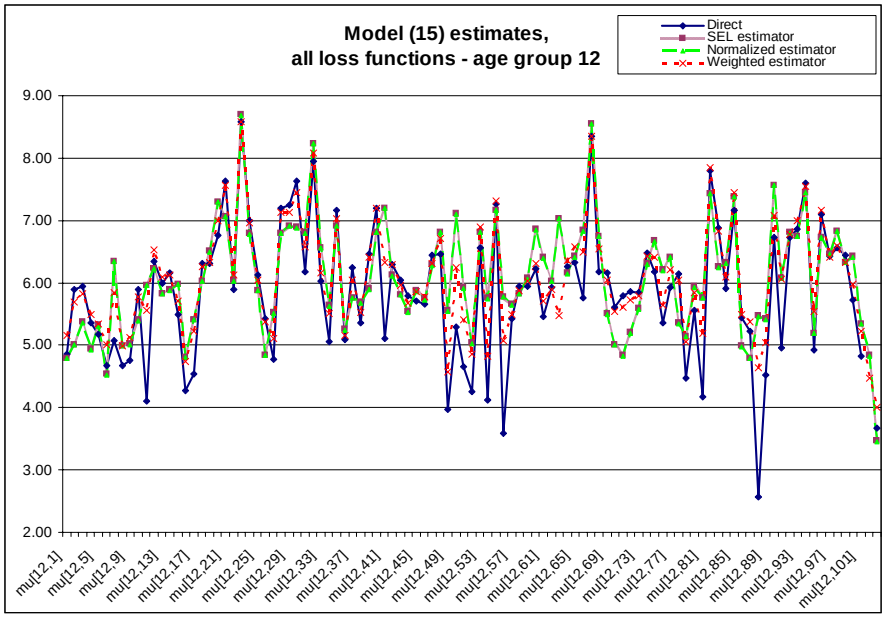

Model (15) estimates,
all loss functions - age group 7
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.00
9.00
8.00
7.00
6.00
5.00
4.00
mu[7,1]
mu[7,5]
mu[7,9]
mu[7,13]
mu[7,17]
mu[7,21]
mu[7,25]
mu[7,29]
mu[7,33]
mu[7,37]
mu[7,41]
mu[7,45]
mu[7,49]
mu[7,53]
mu[7,57]
mu[7,61]
mu[7,65]
mu[7,69]
mu[7,73]
mu[7,77]
mu[7,81]
mu[7,85]
mu[7,89]
mu[7,93]
mu[7,97]
mu[7,101]

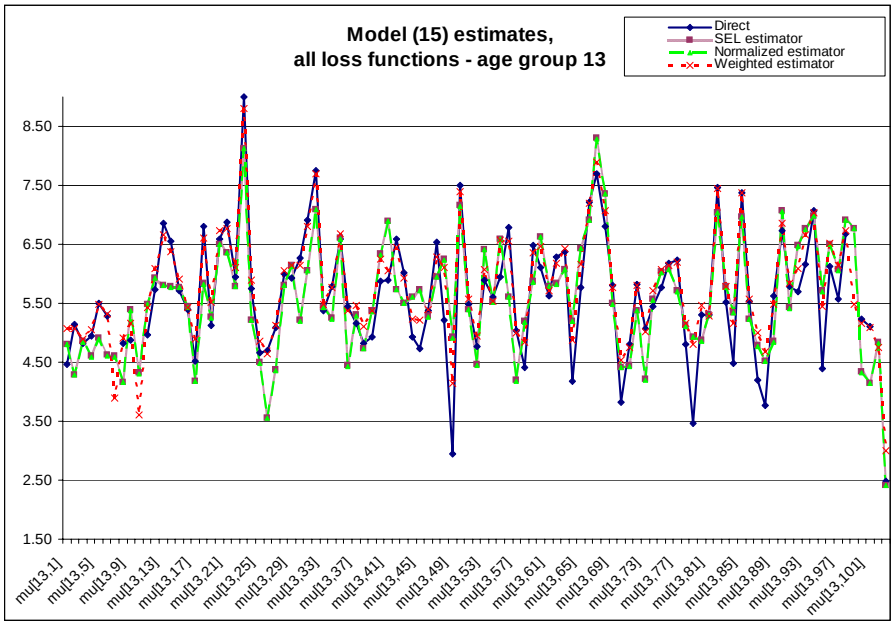

Model (15) estimates,
all loss functions - age group 8
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.50
8.50
7.50
6.50
5.50
4.50
mu[8,1]
mu[8,5]
mu[8,9]
mu[8,13]
mu[8,17]
mu[8,21]
mu[8,25]
mu[8,29]
mu[8,33]
mu[8,37]
mu[8,41]
mu[8,45]
mu[8,49]
mu[8,53]
mu[8,57]
mu[8,61]
mu[8,65]
mu[8,69]
mu[8,73]
mu[8,77]
mu[8,81]
mu[8,85]
mu[8,89]
mu[8,93]
mu[8,97]
mu[8,101]

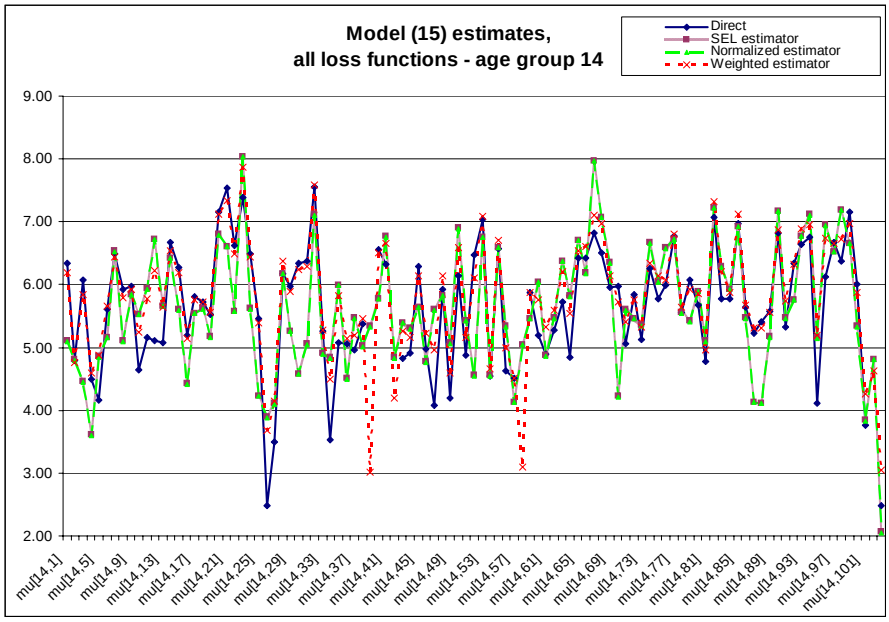

Model (15) estimates,
all loss functions - age group 9
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.50
8.50
7.50
6.50
5.50
4.50
mu[9,1]
mu[9,5]
mu[9,9]
mu[9,13]
mu[9,17]
mu[9,21]
mu[9,25]
mu[9,29]
mu[9,33]
mu[9,37]
mu[9,41]
mu[9,45]
mu[9,49]
mu[9,53]
mu[9,57]
mu[9,61]
mu[9,65]
mu[9,69]
mu[9,73]
mu[9,77]
mu[9,81]
mu[9,85]
mu[9,89]
mu[9,93]
mu[9,97]
mu[9,101]


Model (15) estimates,
all loss functions - age group 10
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.00
8.00
7.00
6.00
5.00
4.00
mu[10,1]
mu[10,5]
mu[10,9]
mu[10,13]
mu[10,17]
mu[10,21]
mu[10,25]
mu[10,29]
mu[10,33]
mu[10,37]
mu[10,41]
mu[10,45]
mu[10,49]
mu[10,53]
mu[10,57]
mu[10,61]
mu[10,65]
mu[10,69]
mu[10,73]
mu[10,77]
mu[10,81]
mu[10,85]
mu[10,89]
mu[10,93]
mu[10,97]
mu[10,101]

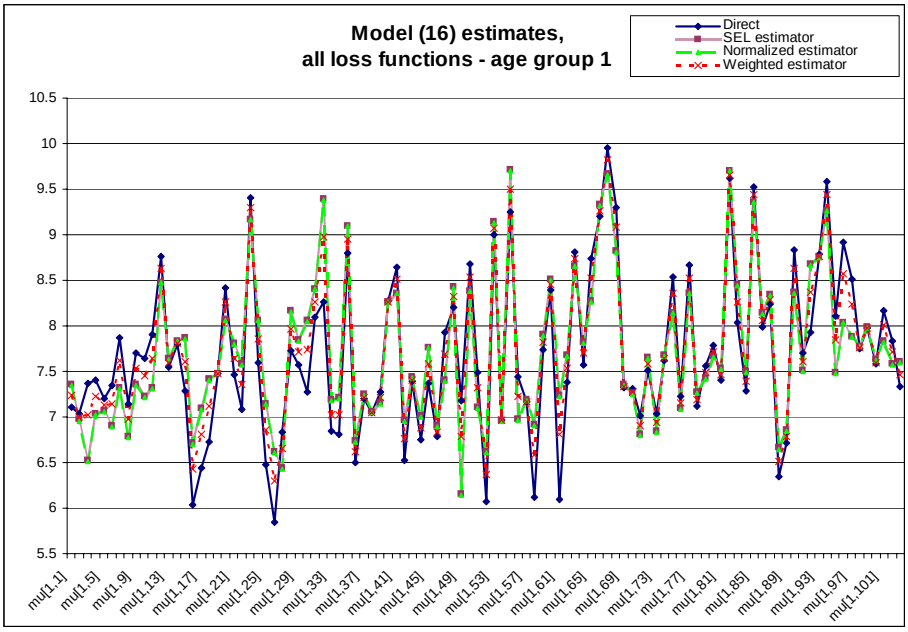

Model (15) estimates,
all loss functions - age group 11
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.00
8.00
7.00
6.00
5.00
4.00
3.00
mu[11,1]
mu[11,5]
mu[11,9]
mu[11,13]
mu[11,17]
mu[11,21]
mu[11,25]
mu[11,29]
mu[11,33]
mu[11,37]
mu[11,41]
mu[11,45]
mu[11,49]
mu[11,53]
mu[11,57]
mu[11,61]
mu[11,65]
mu[11,69]
mu[11,73]
mu[11,77]
mu[11,81]
mu[11,85]
mu[11,89]
mu[11,93]
mu[11,97]
mu[11,101]

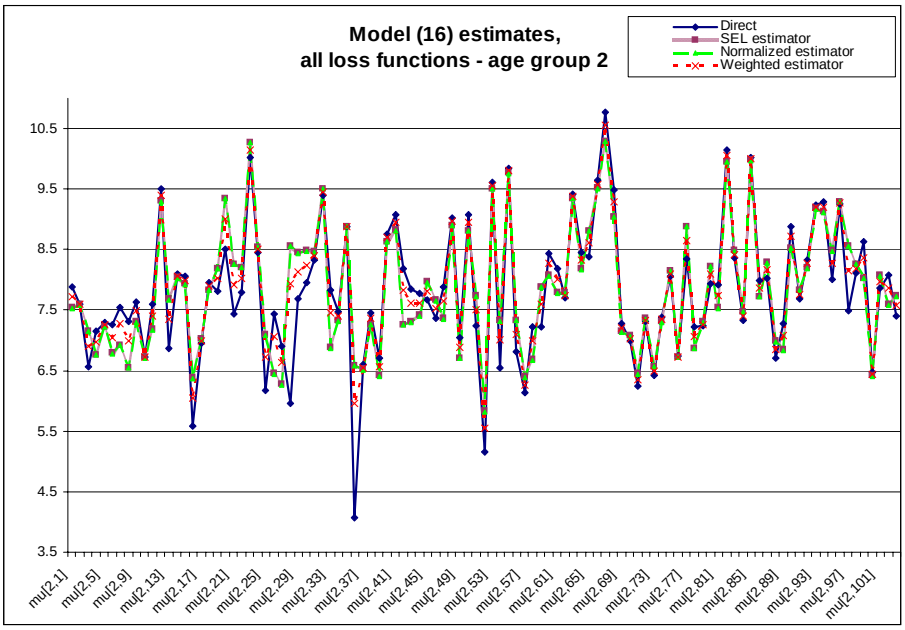

Model (15) estimates,
all loss functions - age group 12
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
mu[12,1]
mu[12,5]
mu[12,9]
mu[12,13]
mu[12,17]
mu[12,21]
mu[12,25]
mu[12,29]
mu[12,33]
mu[12,37]
mu[12,41]
mu[12,45]
mu[12,49]
mu[12,53]
mu[12,57]
mu[12,61]
mu[12,65]
mu[12,69]
mu[12,73]
mu[12,77]
mu[12,81]
mu[12,85]
mu[12,89]
mu[12,93]
mu[12,97]
mu[12,101]

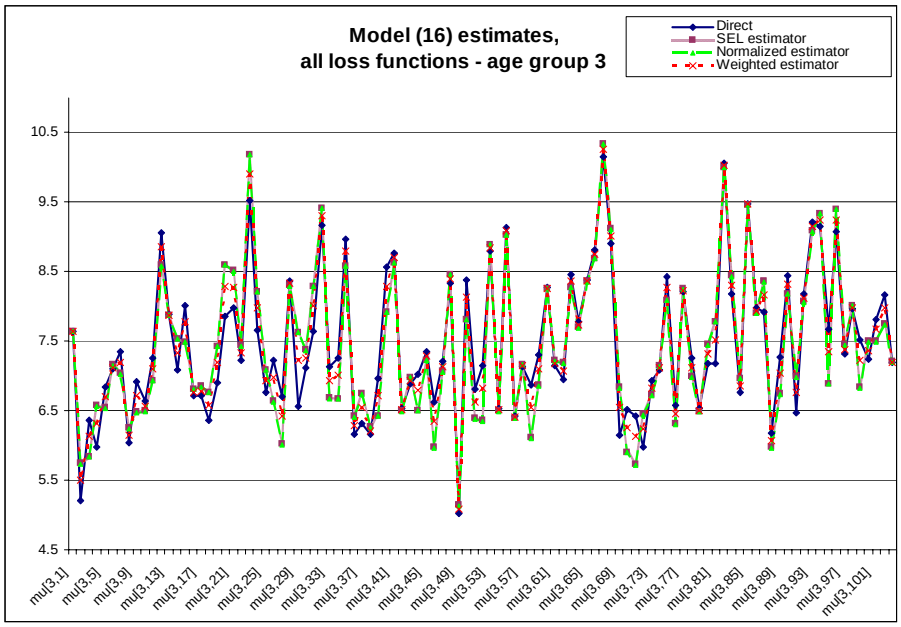

Model (15) estimates,
all loss functions - age group 13
Direct
SEL estimator
Normalized estimator
Weighted estimator
8.50
7.50
6.50
5.50
4.50
3.50
2.50
1.50
mu[13,1]
mu[13,5]
mu[13,9]
mu[13,13]
mu[13,17]
mu[13,21]
mu[13,25]
mu[13,29]
mu[13,33]
mu[13,37]
mu[13,41]
mu[13,45]
mu[13,49]
mu[13,53]
mu[13,57]
mu[13,61]
mu[13,65]
mu[13,69]
mu[13,73]
mu[13,77]
mu[13,81]
mu[13,85]
mu[13,89]
mu[13,93]
mu[13,97]
mu[13,101]

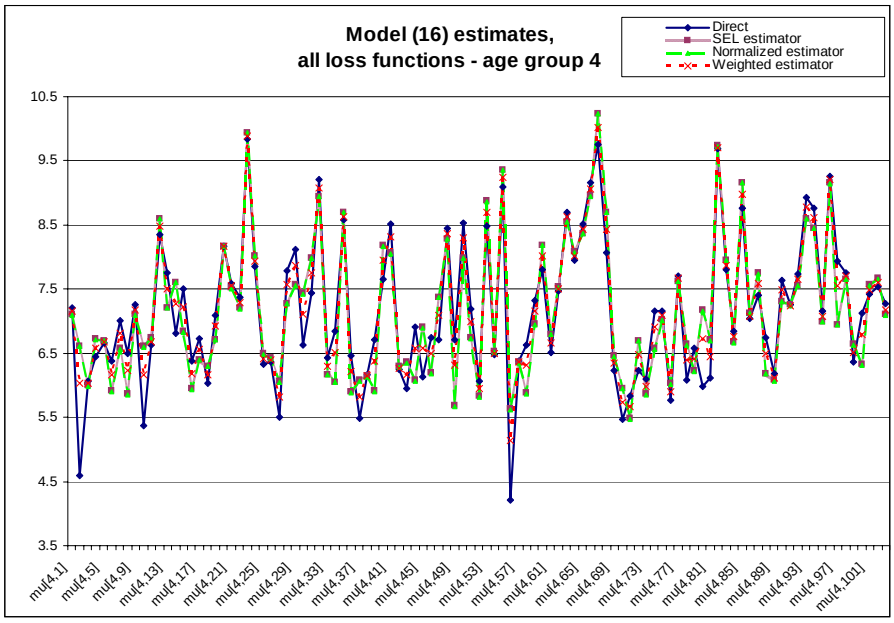

Model (15) estimates,
all loss functions - age group 14
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.00
8.00
7.00
6.00
5.00
4.00
3.00
2.00
mu[14,1]
mu[14,5]
mu[14,9]
mu[14,13]
mu[14,17]
mu[14,21]
mu[14,25]
mu[14,29]
mu[14,33]
mu[14,37]
mu[14,41]
mu[14,45]
mu[14,49]
mu[14,53]
mu[14,57]
mu[14,61]
mu[14,65]
mu[14,69]
mu[14,73]
mu[14,77]
mu[14,81]
mu[14,85]
mu[14,89]
mu[14,93]
mu[14,97]
mu[14,101]

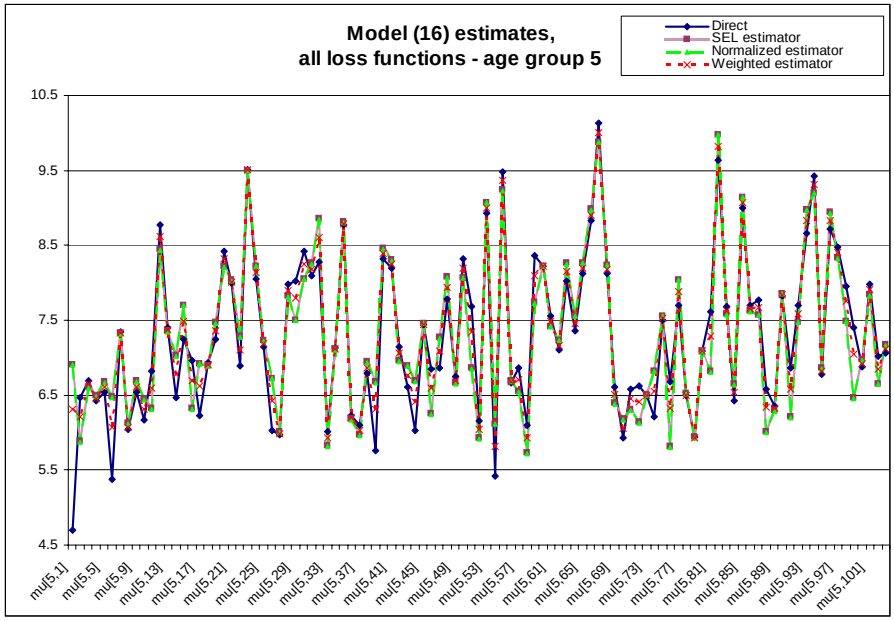
Model (16) estimates (estimates are graphed on a log-scale to allow for a better representation by reducing the variance in the extreme estimates for different domains):

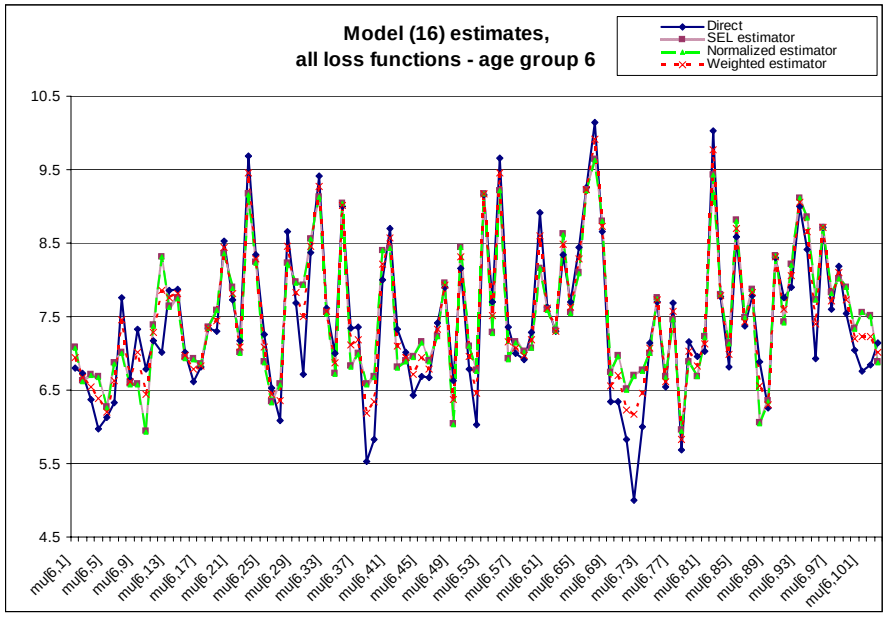


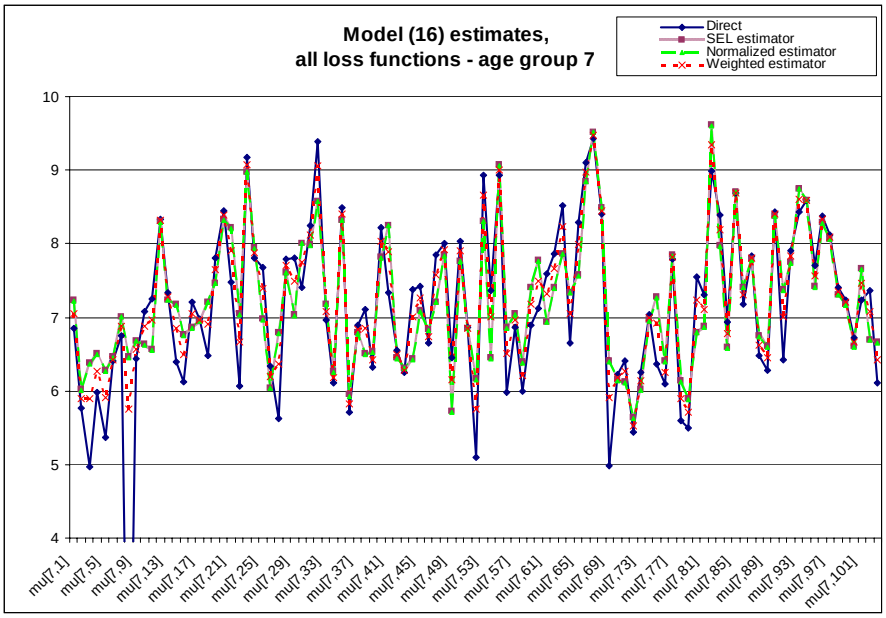


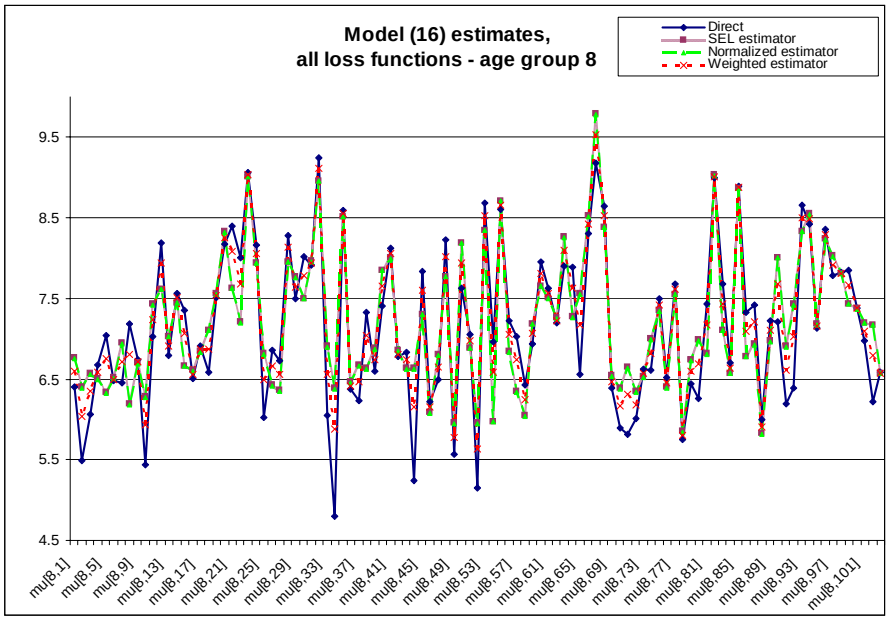

Model (16) estimates,
all loss functions - age group 3
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.5
9.5
8.5
7.5
6.5
5.5
4.5
mu[3,1]
mu[3,5]
mu[3,9]
mu[3,13]
mu[3,17]
mu[3,21]
mu[3,25]
mu[3,29]
mu[3,33]
mu[3,37]
mu[3,41]
mu[3,45]
mu[3,49]
mu[3,53]
mu[3,57]
mu[3,61]
mu[3,65]
mu[3,69]
mu[3,73]
mu[3,77]
mu[3,81]
mu[3,85]
mu[3,89]
mu[3,93]
mu[3,97]
mu[3,101]

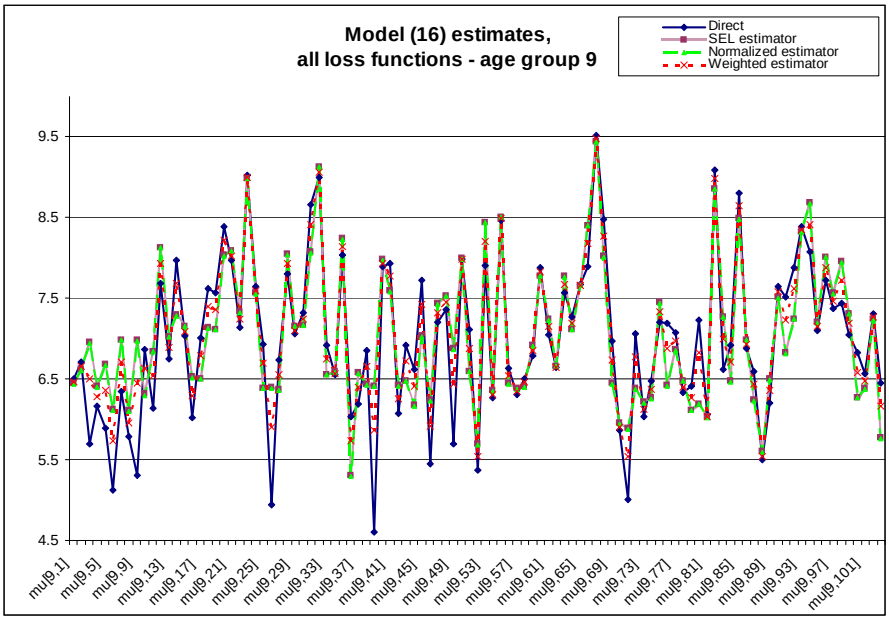

Model (16) estimates,
all loss functions - age group 4
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.5
9.5
8.5
7.5
6.5
5.5
4.5
3.5
mu[4,1]
mu[4,5]
mu[4,9]
mu[4,13]
mu[4,17]
mu[4,21]
mu[4,25]
mu[4,29]
mu[4,33]
mu[4,37]
mu[4,41]
mu[4,45]
mu[4,49]
mu[4,53]
mu[4,57]
mu[4,61]
mu[4,65]
mu[4,69]
mu[4,73]
mu[4,77]
mu[4,81]
mu[4,85]
mu[4,89]
mu[4,93]
mu[4,97]
mu[4,101]

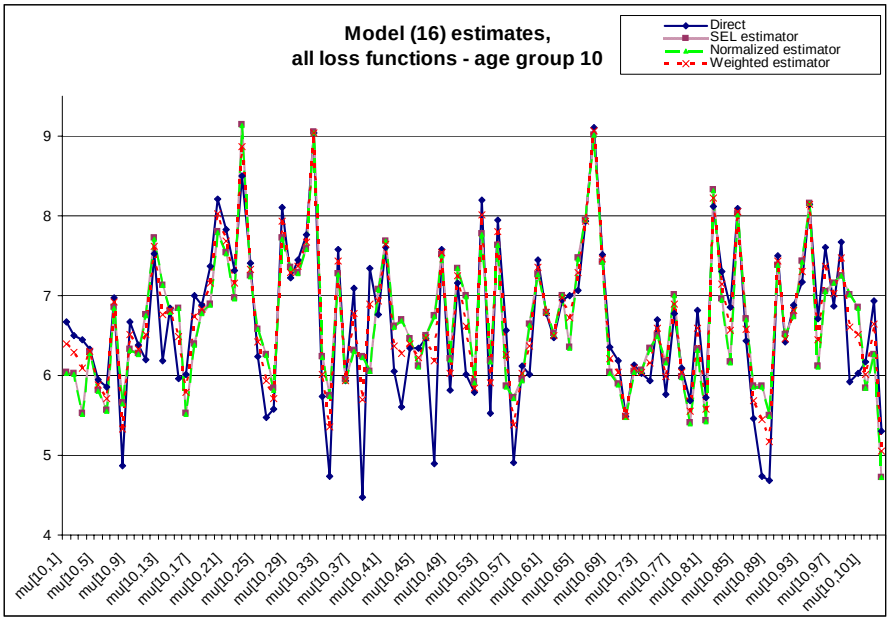

Model (16) estimates,
all loss functions - age group 5
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.5
9.5
8.5
7.5
6.5
5.5
4.5
mu[5,1]
mu[5,5]
mu[5,9]
mu[5,13]
mu[5,17]
mu[5,21]
mu[5,25]
mu[5,29]
mu[5,33]
mu[5,37]
mu[5,41]
mu[5,45]
mu[5,49]
mu[5,53]
mu[5,57]
mu[5,61]
mu[5,65]
mu[5,69]
mu[5,73]
mu[5,77]
mu[5,81]
mu[5,85]
mu[5,89]
mu[5,93]
mu[5,97]
mu[5,101]

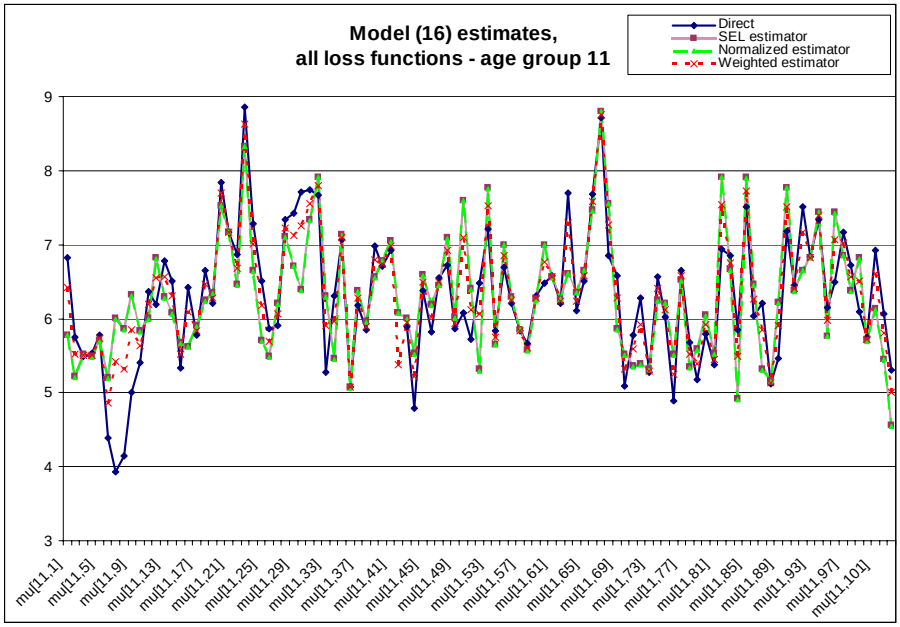

Model (16) estimates,
all loss functions - age group 6
Direct
SEL estimator
Normalized estimator
Weighted estimator
10.5
9.5
8.5
7.5
6.5
5.5
4.5
mu[6,1]
mu[6,5]
mu[6,9]
mu[6,13]
mu[6,17]
mu[6,21]
mu[6,25]
mu[6,29]
mu[6,33]
mu[6,37]
mu[6,41]
mu[6,45]
mu[6,49]
mu[6,53]
mu[6,57]
mu[6,61]
mu[6,65]
mu[6,69]
mu[6,73]
mu[6,77]
mu[6,81]
mu[6,85]
mu[6,89]
mu[6,93]
mu[6,97]
mu[6,101]

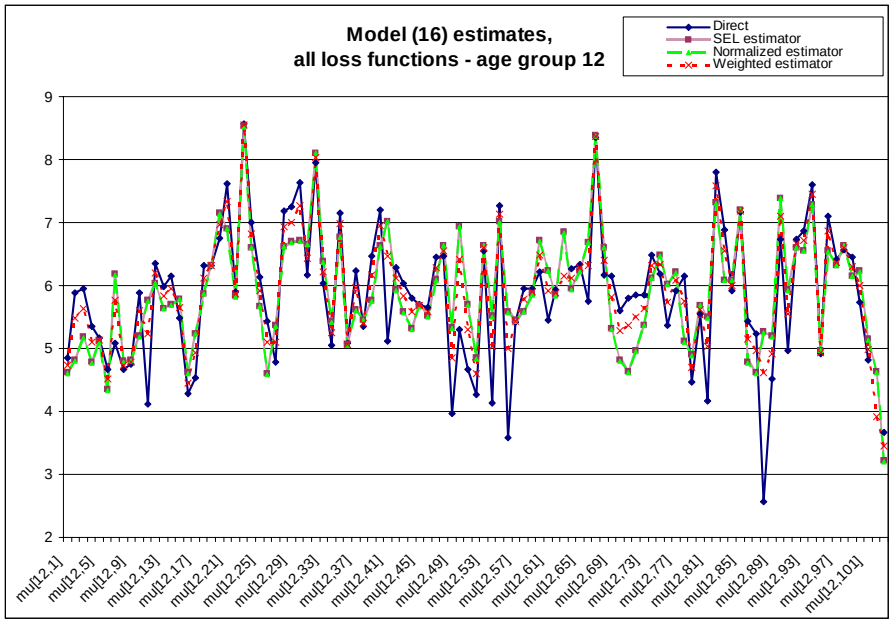

Model (16) estimates,
all loss functions - age group 7
Direct
SEL estimator
Normalized estimator
Weighted estimator
10
9
8
7
6
5
4
mu[7,1]
mu[7,5]
mu[7,9]
mu[7,13]
mu[7,17]
mu[7,21]
mu[7,25]
mu[7,29]
mu[7,33]
mu[7,37]
mu[7,41]
mu[7,45]
mu[7,49]
mu[7,53]
mu[7,57]
mu[7,61]
mu[7,65]
mu[7,69]
mu[7,73]
mu[7,77]
mu[7,81]
mu[7,85]
mu[7,89]
mu[7,93]
mu[7,97]
mu[7,101]

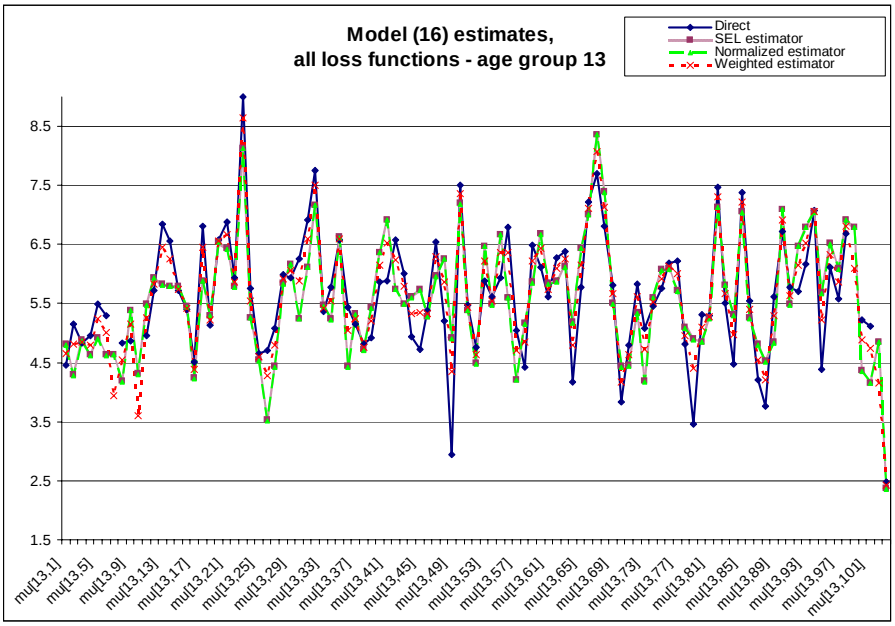

Model (16) estimates,
all loss functions - age group 8
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.5
8.5
7.5
6.5
5.5
4.5
mu[8,1]
mu[8,5]
mu[8,9]
mu[8,13]
mu[8,17]
mu[8,21]
mu[8,25]
mu[8,29]
mu[8,33]
mu[8,37]
mu[8,41]
mu[8,45]
mu[8,49]
mu[8,53]
mu[8,57]
mu[8,61]
mu[8,65]
mu[8,69]
mu[8,73]
mu[8,77]
mu[8,81]
mu[8,85]
mu[8,89]
mu[8,93]
mu[8,97]
mu[8,101]

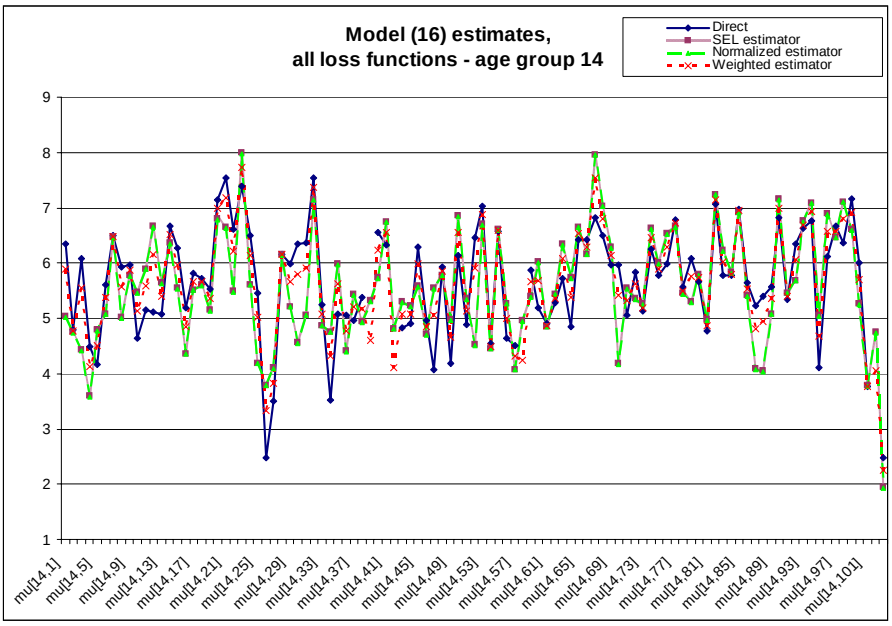

Model (16) estimates,
all loss functions - age group 9
Direct
SEL estimator
Normalized estimator
Weighted estimator
9.5
8.5
7.5
6.5
5.5
4.5
mu[9,1]
mu[9,5]
mu[9,9]
mu[9,13]
mu[9,17]
mu[9,21]
mu[9,25]
mu[9,29]
mu[9,33]
mu[9,37]
mu[9,41]
mu[9,45]
mu[9,49]
mu[9,53]
mu[9,57]
mu[9,61]
mu[9,65]
mu[9,69]
mu[9,73]
mu[9,77]
mu[9,81]
mu[9,85]
mu[9,89]
mu[9,93]
mu[9,97]
mu[9,101]


Model (16) estimates,
all loss functions - age group 10
Direct
SEL estimator
Normalized estimator
Weighted estimator
9
8
7
6
5
4
mu[10,1]
mu[10,5]
mu[10,9]
mu[10,13]
mu[10,17]
mu[10,21]
mu[10,25]
mu[10,29]
mu[10,33]
mu[10,37]
mu[10,41]
mu[10,45]
mu[10,49]
mu[10,53]
mu[10,57]
mu[10,61]
mu[10,65]
mu[10,69]
mu[10,73]
mu[10,77]
mu[10,81]
mu[10,85]
mu[10,89]
mu[10,93]
mu[10,97]
mu[10,101]


Model (16) estimates,
all loss functions - age group 11
Direct
SEL estimator
Normalized estimator
Weighted estimator
9
8
7
6
5
4
3
mu[11,1]
mu[11,5]
mu[11,9]
mu[11,13]
mu[11,17]
mu[11,21]
mu[11,25]
mu[11,29]
mu[11,33]
mu[11,37]
mu[11,41]
mu[11,45]
mu[11,49]
mu[11,53]
mu[11,57]
mu[11,61]
mu[11,65]
mu[11,69]
mu[11,73]
mu[11,77]
mu[11,81]
mu[11,85]
mu[11,89]
mu[11,93]
mu[11,97]
mu[11,101]


Model (16) estimates,
all loss functions - age group 12
Direct
SEL estimator
Normalized estimator
Weighted estimator
9
8
7
6
5
4
3
2
mu[12,1]
mu[12,5]
mu[12,9]
mu[12,13]
mu[12,17]
mu[12,21]
mu[12,25]
mu[12,29]
mu[12,33]
mu[12,37]
mu[12,41]
mu[12,45]
mu[12,49]
mu[12,53]
mu[12,57]
mu[12,61]
mu[12,65]
mu[12,69]
mu[12,73]
mu[12,77]
mu[12,81]
mu[12,85]
mu[12,89]
mu[12,93]
mu[12,97]
mu[12,101]

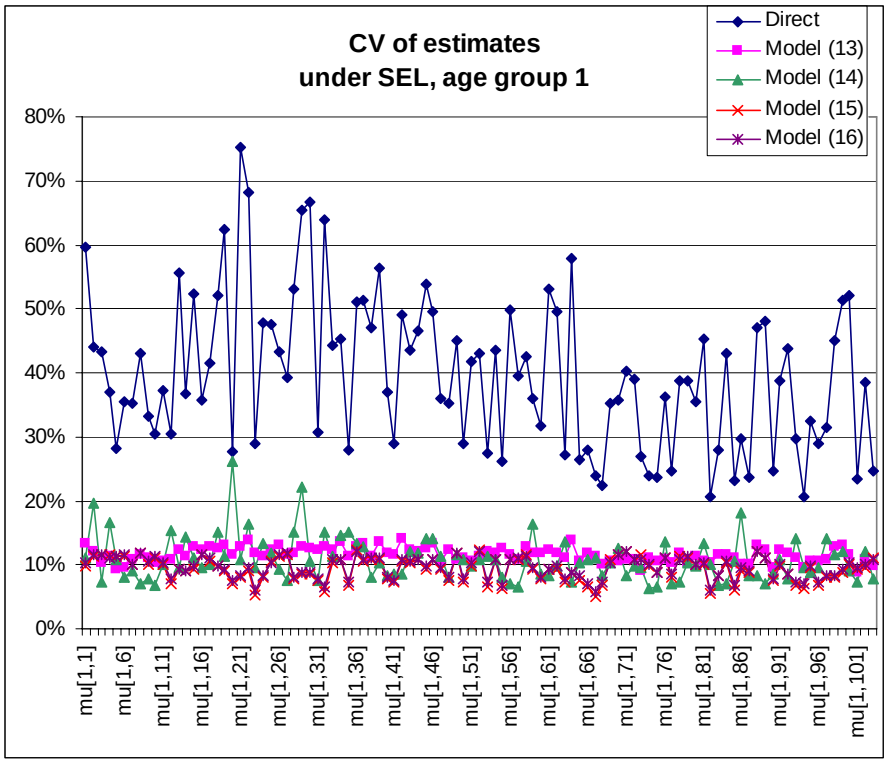

Model (16) estimates,
all loss functions - age group 13
Direct
SEL estimator
Normalized estimator
Weighted estimator
8.5
7.5
6.5
5.5
4.5
3.5
2.5
1.5
mu[13,1]
mu[13,5]
mu[13,9]
mu[13,13]
mu[13,17]
mu[13,21]
mu[13,25]
mu[13,29]
mu[13,33]
mu[13,37]
mu[13,41]
mu[13,45]
mu[13,49]
mu[13,53]
mu[13,57]
mu[13,61]
mu[13,65]
mu[13,69]
mu[13,73]
mu[13,77]
mu[13,81]
mu[13,85]
mu[13,89]
mu[13,93]
mu[13,97]
mu[13,101]

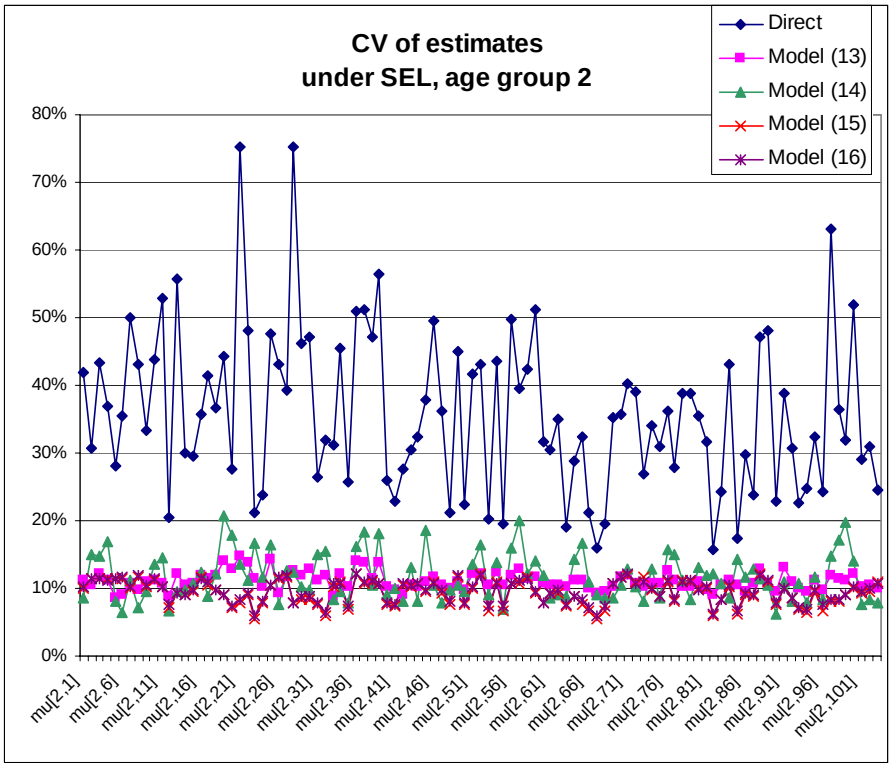

Model (16) estimates,
all loss functions - age group 14
Direct
SEL estimator
Normalized estimator
Weighted estimator
9
8
7
6
5
4
3
2
1
mu[14,1]
mu[14,5]
mu[14,9]
mu[14,13]
mu[14,17]
mu[14,21]
mu[14,25]
mu[14,29]
mu[14,33]
mu[14,37]
mu[14,41]
mu[14,45]
mu[14,49]
mu[14,53]
mu[14,57]
mu[14,61]
mu[14,65]
mu[14,69]
mu[14,73]
mu[14,77]
mu[14,81]
mu[14,85]
mu[14,89]
mu[14,93]
mu[14,97]
mu[14,101]

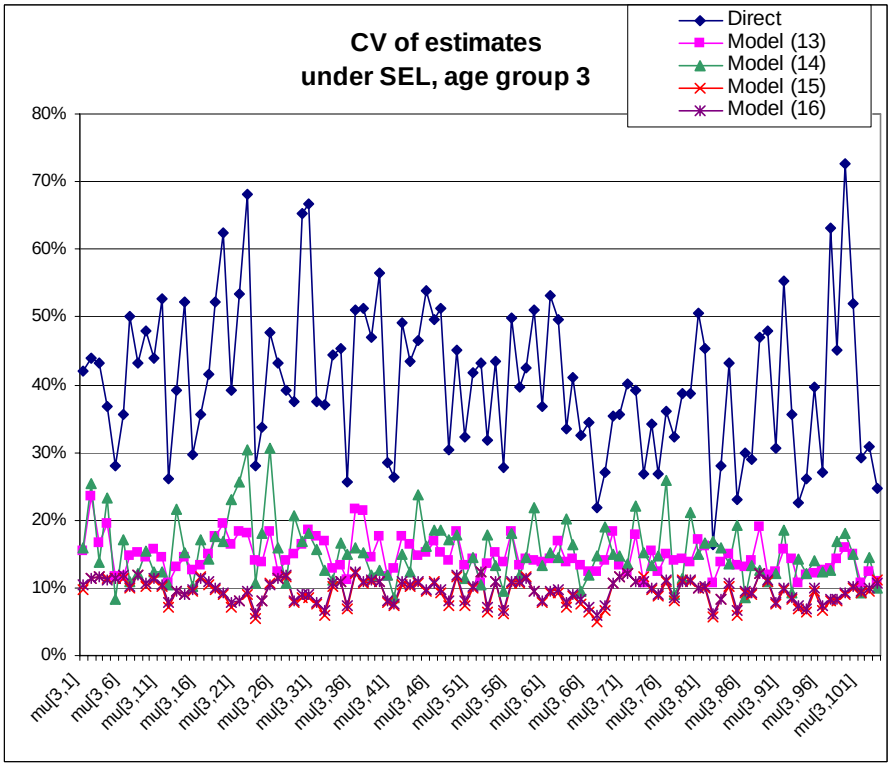
Appendix F

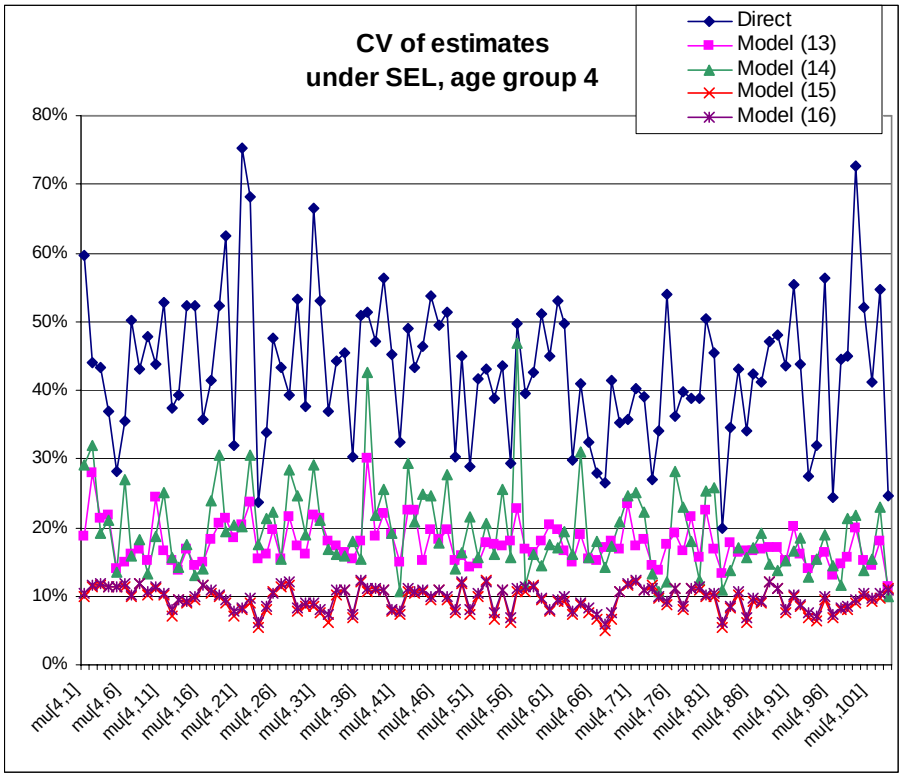
CV of Estimates Under Different Loss Functions

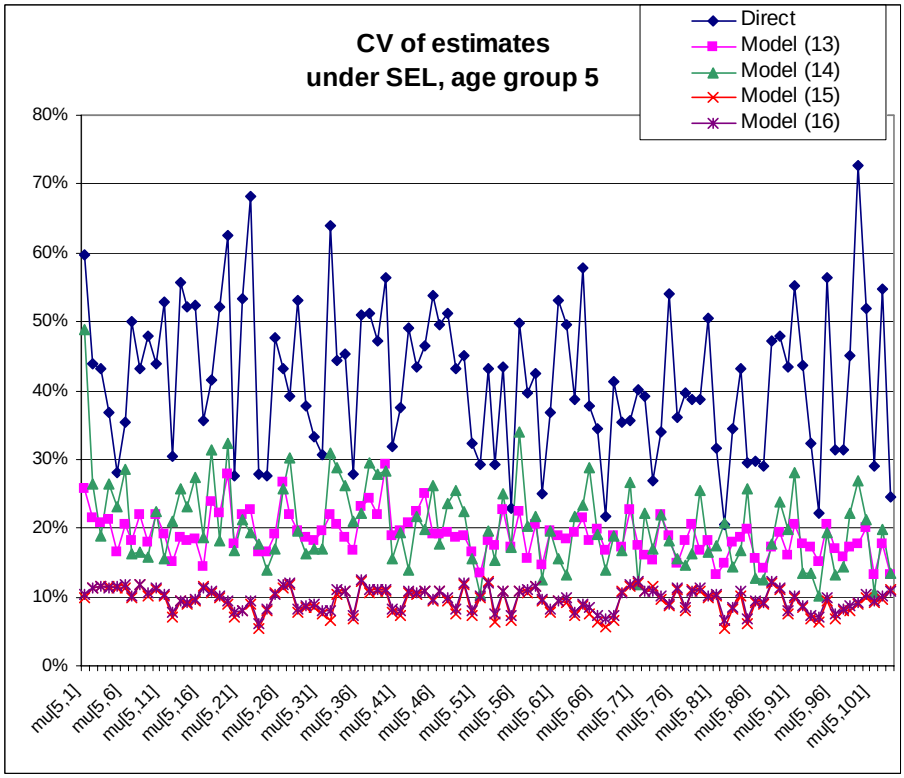
CV of estimates from all four models are presented below - graphed by age group as well as by the loss function. Under SEL, the following CV of estimates were obtained:

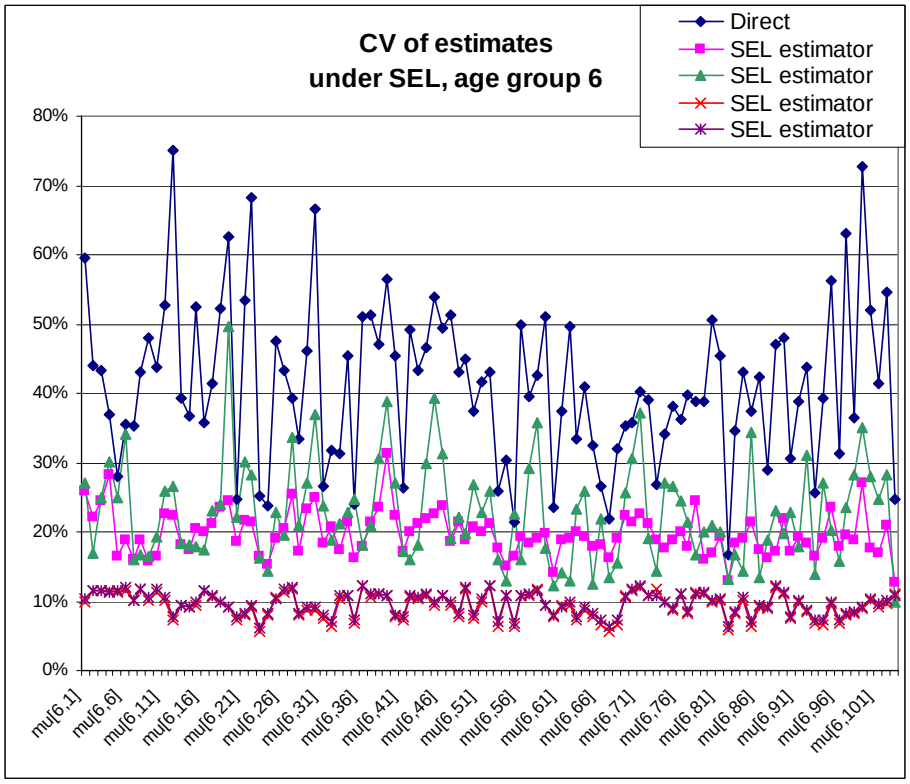


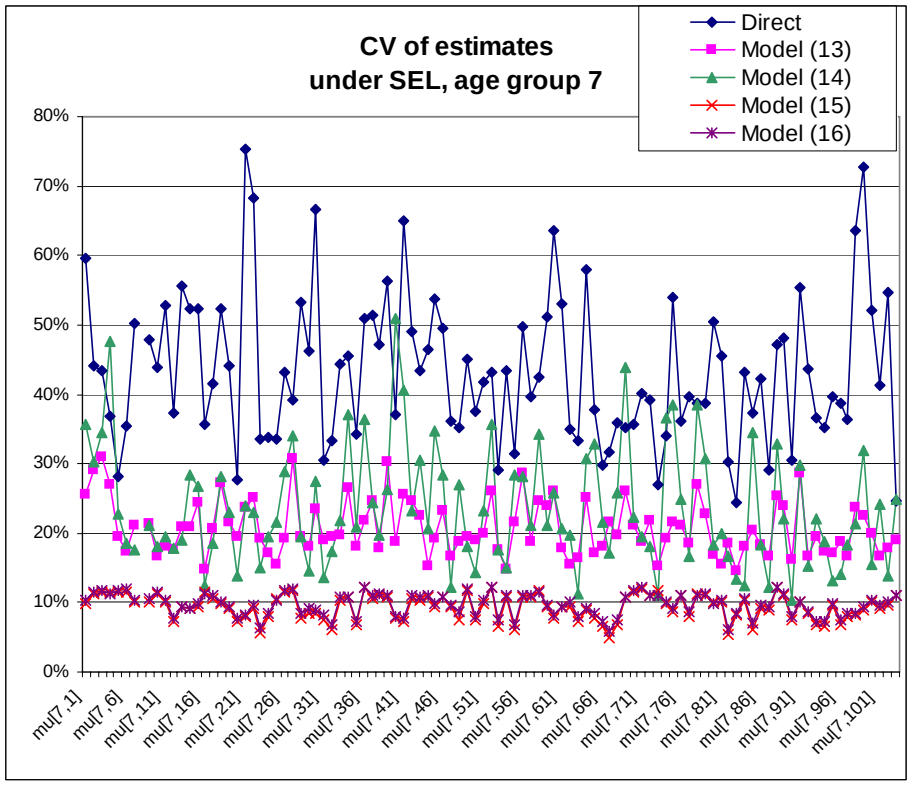

CV of estimates
under SEL, age group 2
Direct
Model (13)
Model (14)
Model (15)
Model (16)
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[2,1]
mu[2,6]
mu[2,11]
mu[2,16]
mu[2,21]
mu[2,26]
mu[2,31]
mu[2,36]
mu[2,41]
mu[2,46]
mu[2,51]
mu[2,56]
mu[2,61]
mu[2,66]
mu[2,71]
mu[2,76]
mu[2,81]
mu[2,86]
mu[2,91]
mu[2,96]
mu[2,101]

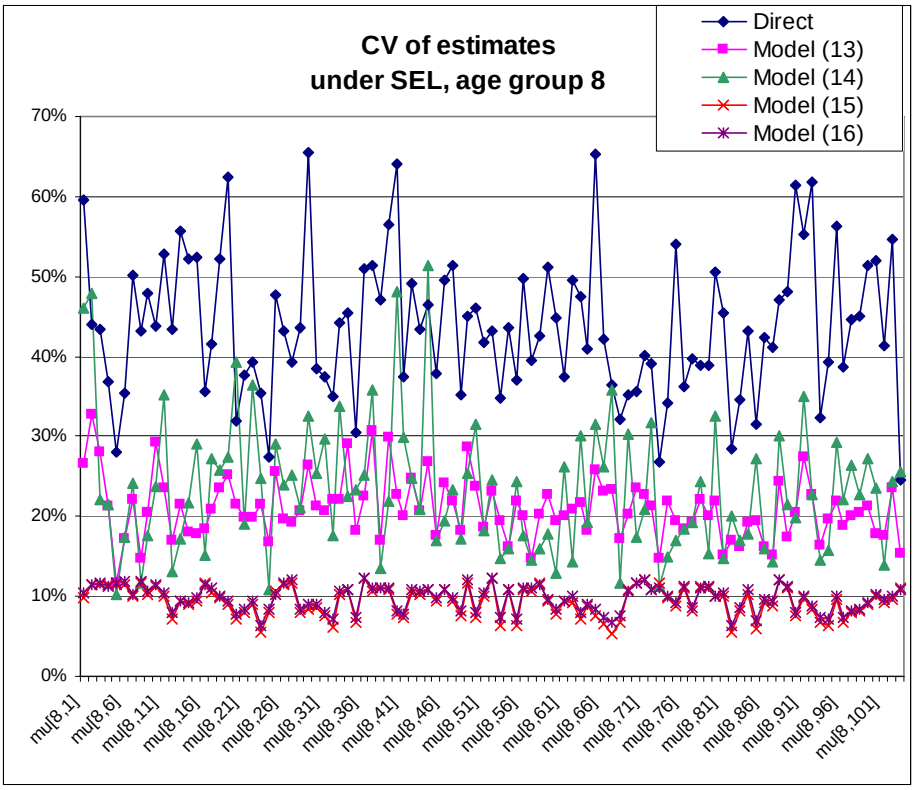

CV of estimates
under SEL, age group 3
Direct
Model (13)
Model (14)
Model (15)
Model (16)
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[3,1]
mu[3,6]
mu[3,11]
mu[3,16]
mu[3,21]
mu[3,26]
mu[3,31]
mu[3,36]
mu[3,41]
mu[3,46]
mu[3,51]
mu[3,56]
mu[3,61]
mu[3,66]
mu[3,71]
mu[3,76]
mu[3,81]
mu[3,86]
mu[3,91]
mu[3,96]
mu[3,101]

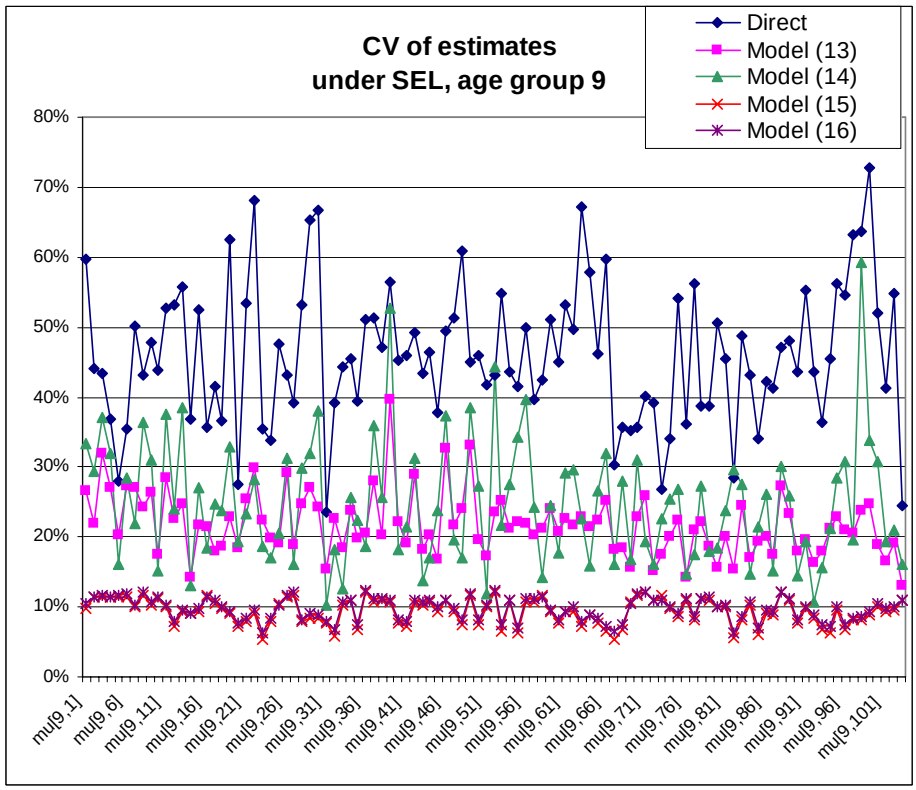

CV of estimates
under SEL, age group 4
Direct
Model (13)
Model (14)
Model (15)
Model (16)
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[4,1]
mu[4,6]
mu[4,11]
mu[4,16]
mu[4,21]
mu[4,26]
mu[4,31]
mu[4,36]
mu[4,41]
mu[4,46]
mu[4,51]
mu[4,56]
mu[4,61]
mu[4,66]
mu[4,71]
mu[4,76]
mu[4,81]
mu[4,86]
mu[4,91]
mu[4,96]
mu[4,101]

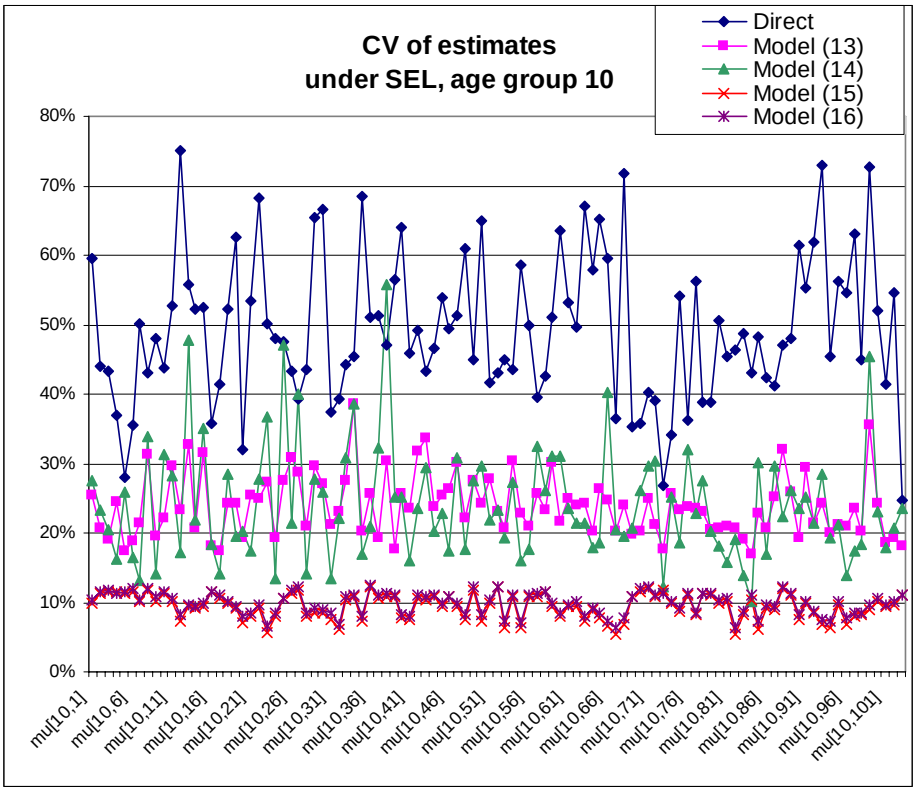

CV of estimates
under SEL, age group 5
Direct
Model (13)
Model (14)
Model (15)
Model (16)
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[5,1]
mu[5,6]
mu[5,11]
mu[5,16]
mu[5,21]
mu[5,26]
mu[5,31]
mu[5,36]
mu[5,41]
mu[5,46]
mu[5,51]
mu[5,56]
mu[5,61]
mu[5,66]
mu[5,71]
mu[5,76]
mu[5,81]
mu[5,86]
mu[5,91]
mu[5,96]
mu[5,101]

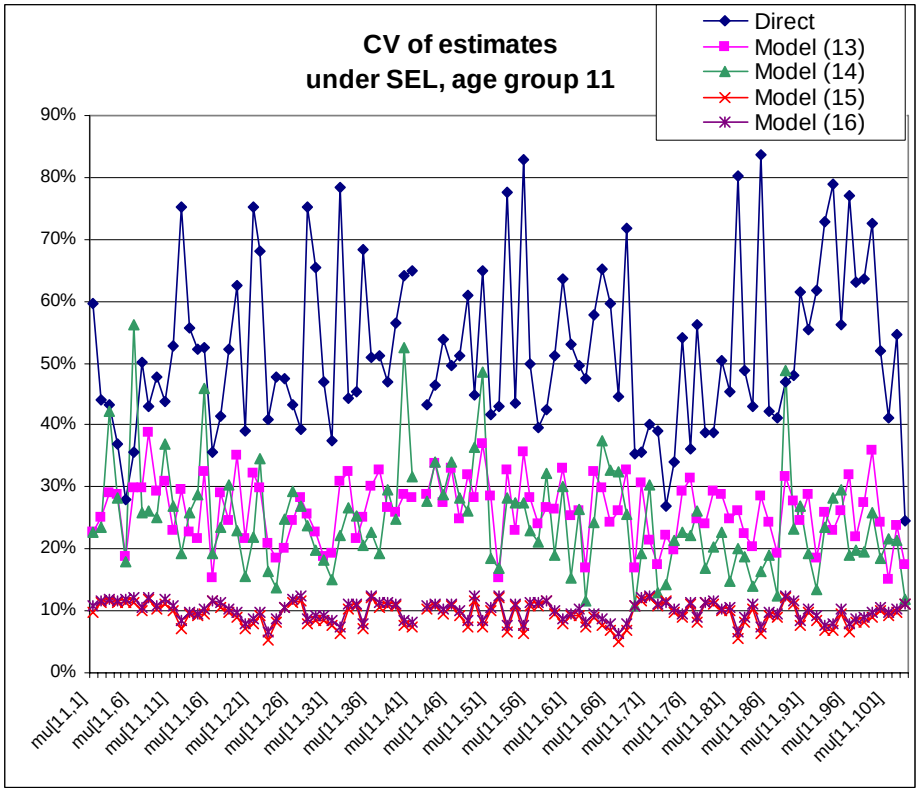

CV of estimates
under SEL, age group 6
Direct
SEL estimator
SEL estimator
SEL estimator
SEL estimator
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[6,1]
mu[6,6]
mu[6,11]
mu[6,16]
mu[6,21]
mu[6,26]
mu[6,31]
mu[6,36]
mu[6,41]
mu[6,46]
mu[6,51]
mu[6,56]
mu[6,61]
mu[6,66]
mu[6,71]
mu[6,76]
mu[6,81]
mu[6,86]
mu[6,91]
mu[6,96]
mu[6,101]

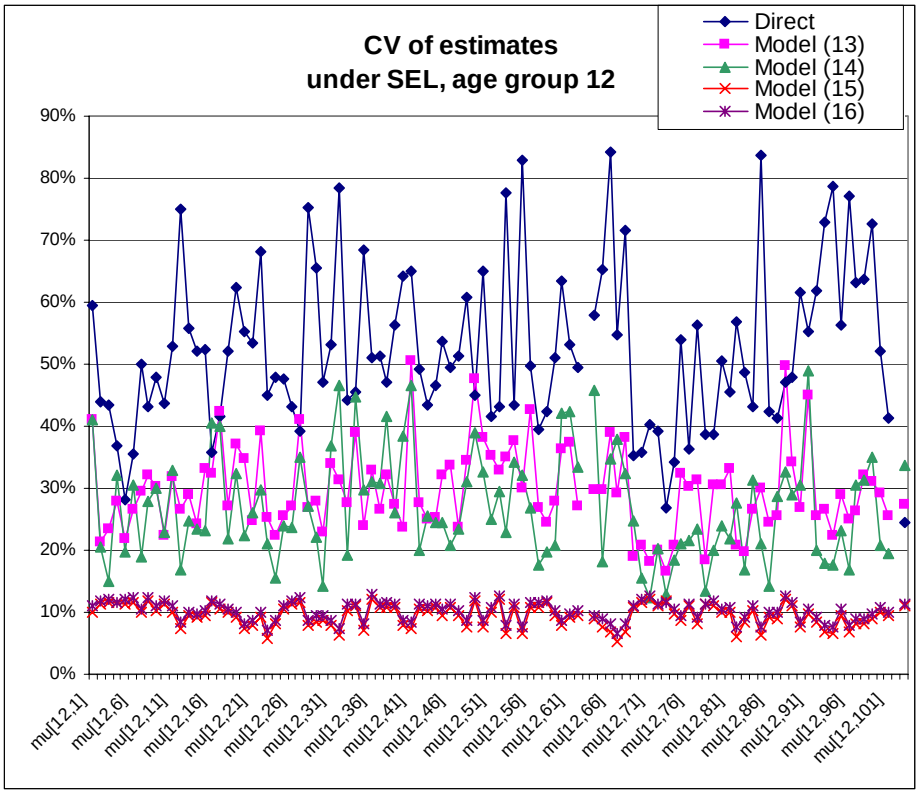

CV of estimates
under SEL, age group 7
Direct
Model (13)
Model (14)
Model (15)
Model (16)
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[7,1]
mu[7,6]
mu[7,11]
mu[7,16]
mu[7,21]
mu[7,26]
mu[7,31]
mu[7,36]
mu[7,41]
mu[7,46]
mu[7,51]
mu[7,56]
mu[7,61]
mu[7,66]
mu[7,71]
mu[7,76]
mu[7,81]
mu[7,86]
mu[7,91]
mu[7,96]
mu[7,101]

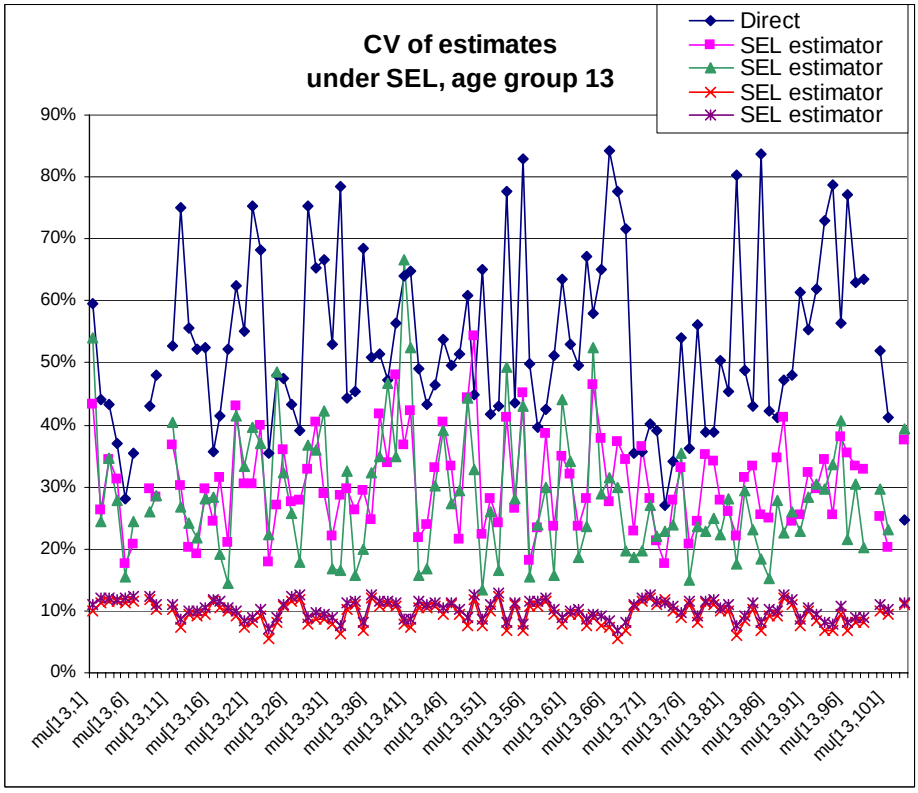

CV of estimates
under SEL, age group 8
Direct
Model (13)
Model (14)
Model (15)
Model (16)
70%
60%
50%
40%
30%
20%
10%
0%
mu[8,1]
mu[8,6]
mu[8,11]
mu[8,16]
mu[8,21]
mu[8,26]
mu[8,31]
mu[8,36]
mu[8,41]
mu[8,46]
mu[8,51]
mu[8,56]
mu[8,61]
mu[8,66]
mu[8,71]
mu[8,76]
mu[8,81]
mu[8,86]
mu[8,91]
mu[8,96]
mu[8,101]

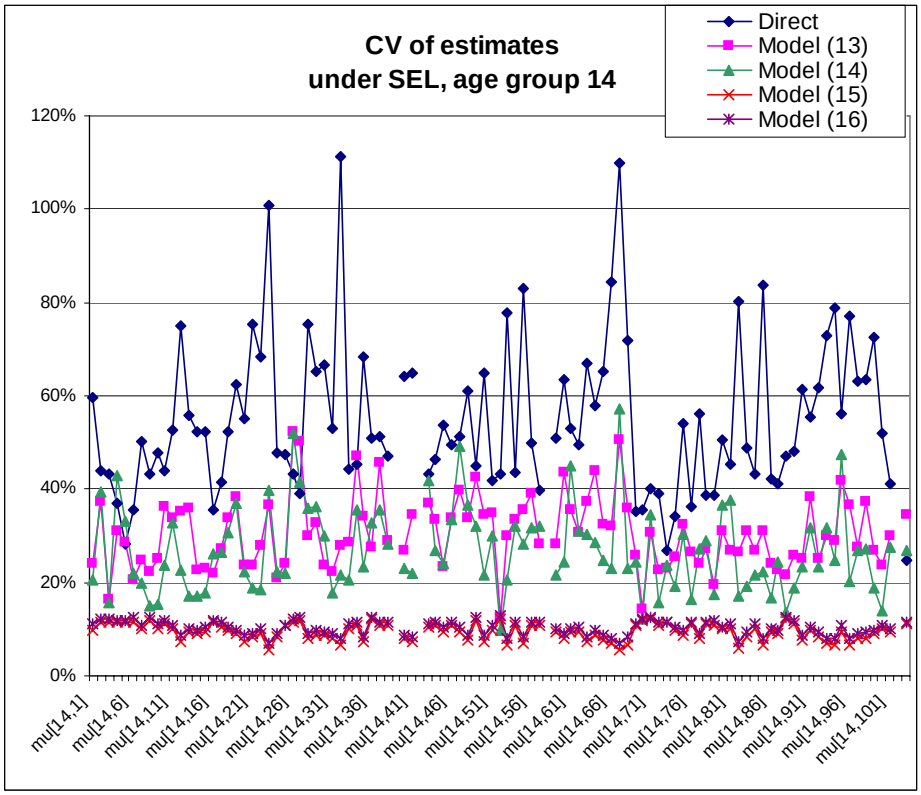

CV of estimates
under SEL, age group 9
Direct
Model (13)
Model (14)
Model (15)
Model (16)
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[9,1]
mu[9,6]
mu[9,11]
mu[9,16]
mu[9,21]
mu[9,26]
mu[9,31]
mu[9,36]
mu[9,41]
mu[9,46]
mu[9,51]
mu[9,56]
mu[9,61]
mu[9,66]
mu[9,71]
mu[9,76]
mu[9,81]
mu[9,86]
mu[9,91]
mu[9,96]
mu[9,101]


CV of estimates
under SEL, age group 10
Direct
Model (13)
Model (14)
Model (15)
Model (16)
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[10,1]
mu[10,6]
mu[10,11]
mu[10,16]
mu[10,21]
mu[10,26]
mu[10,31]
mu[10,36]
mu[10,41]
mu[10,46]
mu[10,51]
mu[10,56]
mu[10,61]
mu[10,66]
mu[10,71]
mu[10,76]
mu[10,81]
mu[10,86]
mu[10,91]
mu[10,96]
mu[10,101]

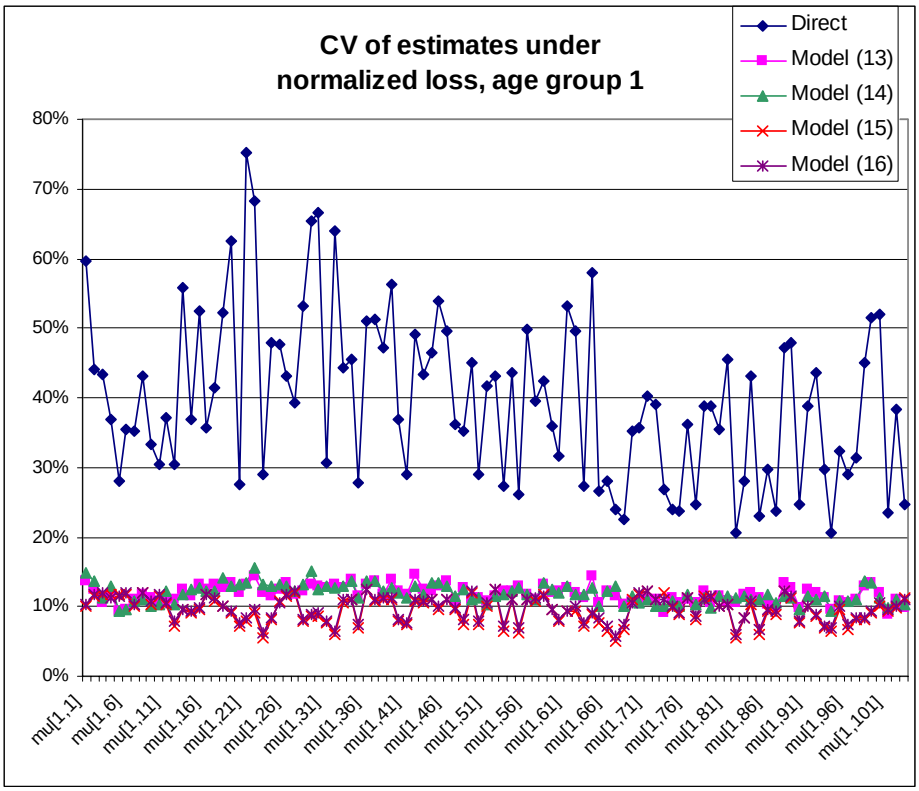

CV of estimates
under SEL, age group 11
Direct
Model (13)
Model (14)
Model (15)
Model (16)
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[11,1]
mu[11,6]
mu[11,11]
mu[11,16]
mu[11,21]
mu[11,26]
mu[11,31]
mu[11,36]
mu[11,41]
mu[11,46]
mu[11,51]
mu[11,56]
mu[11,61]
mu[11,66]
mu[11,71]
mu[11,76]
mu[11,81]
mu[11,86]
mu[11,91]
mu[11,96]
mu[11,101]

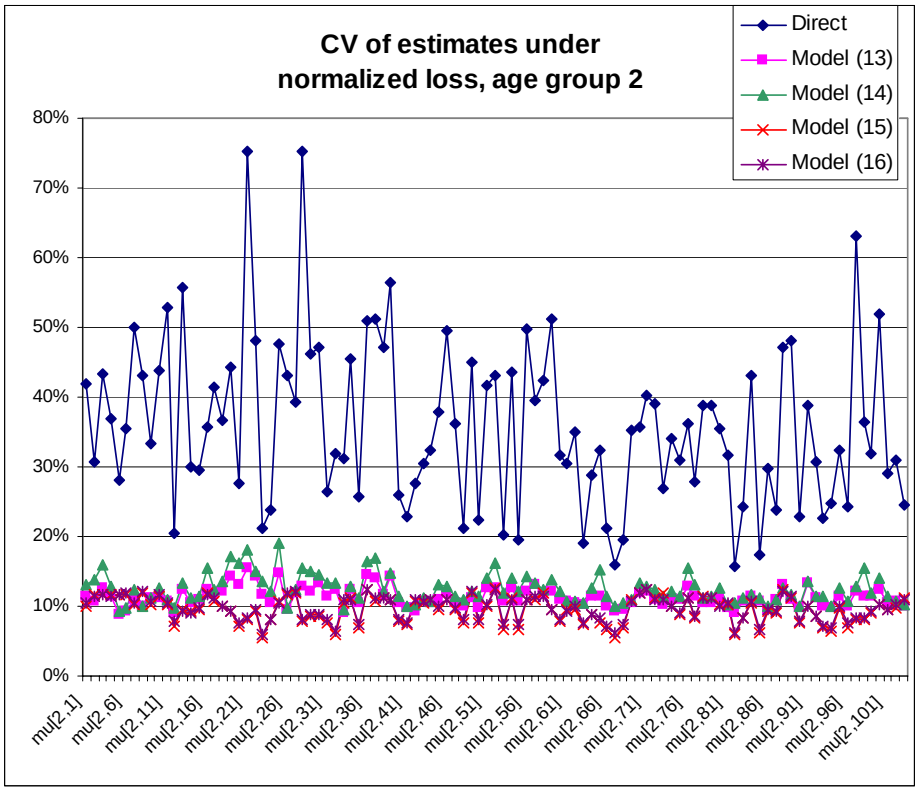

CV of estimates
under SEL, age group 12
Direct
Model (13)
Model (14)
Model (15)
Model (16)
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[12,1]
mu[12,6]
mu[12,11]
mu[12,16]
mu[12,21]
mu[12,26]
mu[12,31]
mu[12,36]
mu[12,41]
mu[12,46]
mu[12,51]
mu[12,56]
mu[12,61]
mu[12,66]
mu[12,71]
mu[12,76]
mu[12,81]
mu[12,86]
mu[12,91]
mu[12,96]
mu[12,101]

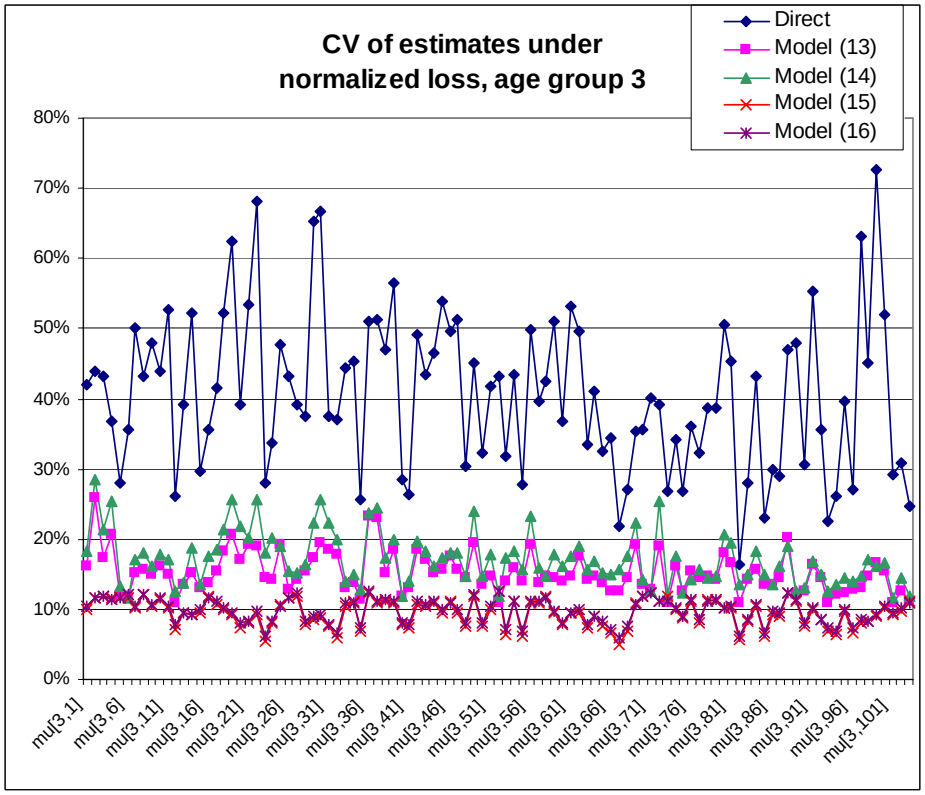

CV of estimates
under SEL, age group 13
Direct
SEL estimator
SEL estimator
SEL estimator
SEL estimator
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[13,1]
mu[13,6]
mu[13,11]
mu[13,16]
mu[13,21]
mu[13,26]
mu[13,31]
mu[13,36]
mu[13,41]
mu[13,46]
mu[13,51]
mu[13,56]
mu[13,61]
mu[13,66]
mu[13,71]
mu[13,76]
mu[13,81]
mu[13,86]
mu[13,91]
mu[13,96]
mu[13,101]

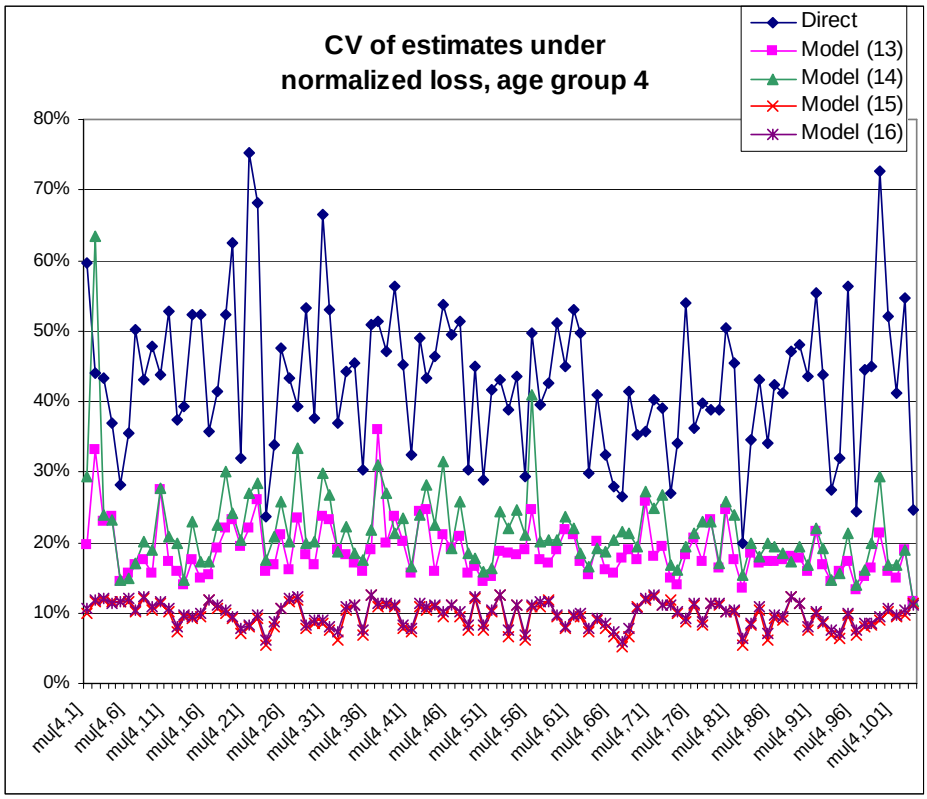

CV of estimates
under SEL, age group 14
Direct
Model (13)
Model (14)
Model (15)
Model (16)
120%
100%
80%
60%
40%
20%
0%
mu[14,1]
mu[14,6]
mu[14,11]
mu[14,16]
mu[14,21]
mu[14,26]
mu[14,31]
mu[14,36]
mu[14,41]
mu[14,46]
mu[14,51]
mu[14,56]
mu[14,61]
mu[14,66]
mu[14,71]
mu[14,76]
mu[14,81]
mu[14,86]
mu[14,91]
mu[14,96]
mu[14,101]

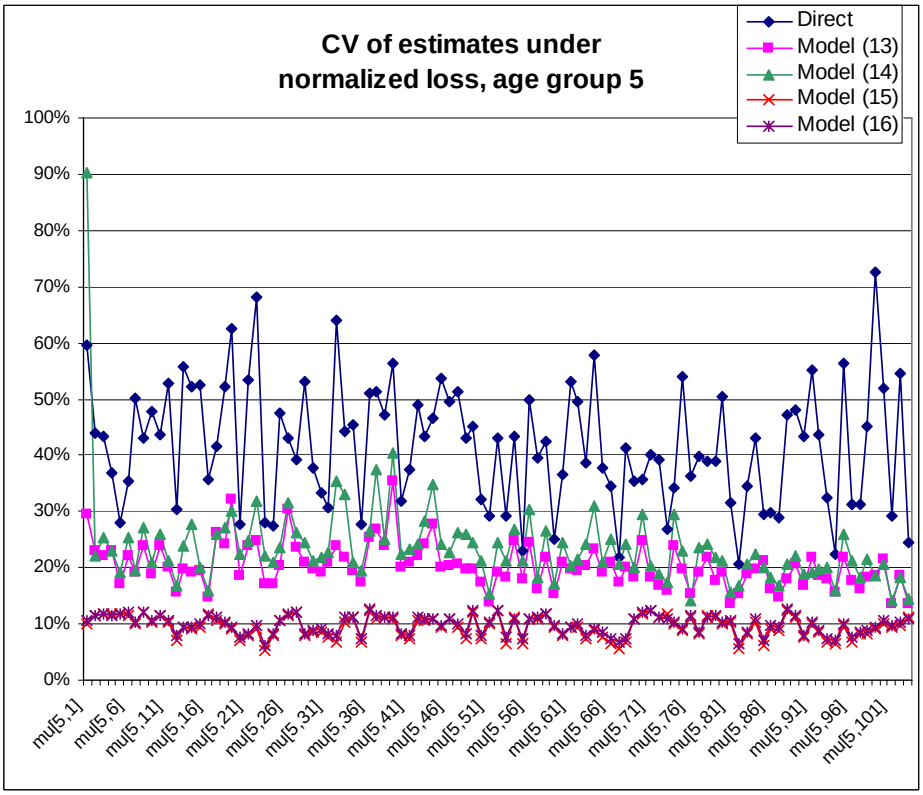
CV of estimates under NSEL:

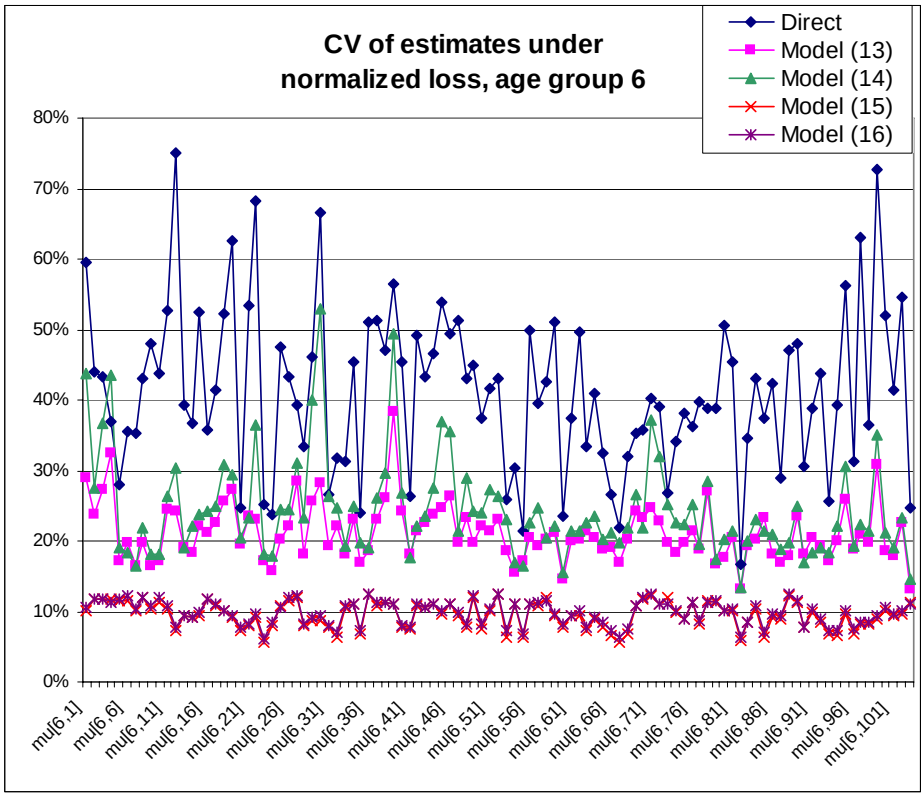


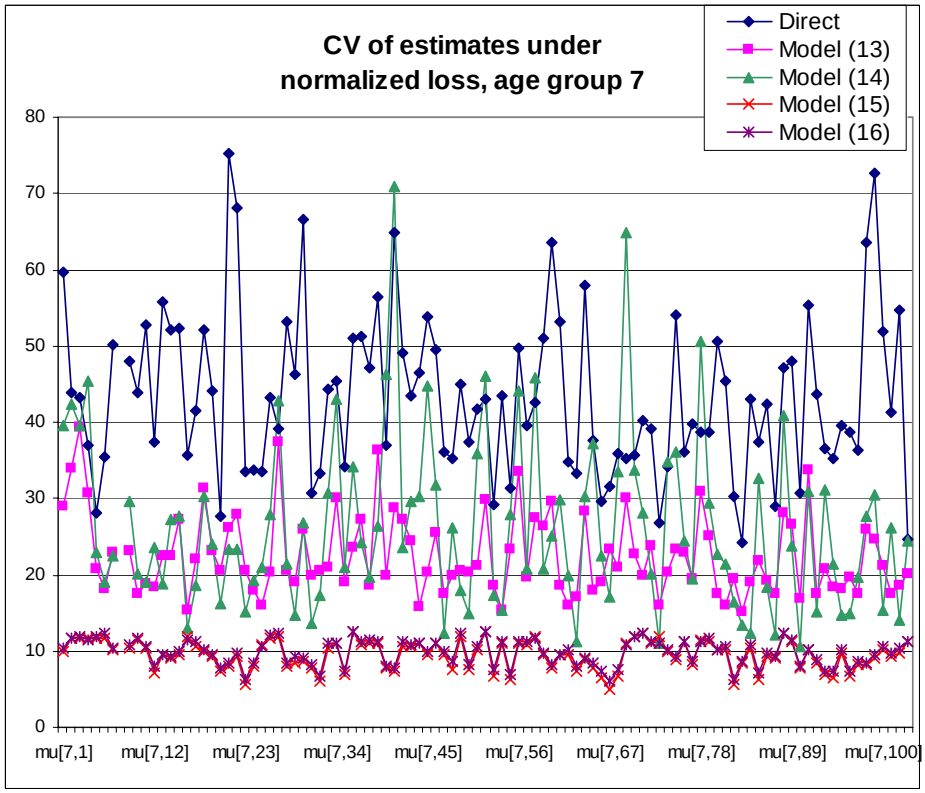

CV of estimates under normalized loss, age group 2
Direct
Model (13)
Model (14)
Model (15)
Model (16)
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[2,1]
mu[2,6]
mu[2,11]
mu[2,16]
mu[2,21]
mu[2,26]
mu[2,31]
mu[2,36]
mu[2,41]
mu[2,46]
mu[2,51]
mu[2,56]
mu[2,61]
mu[2,66]
mu[2,71]
mu[2,76]
mu[2,81]
mu[2,86]
mu[2,91]
mu[2,96]
mu[2,101]

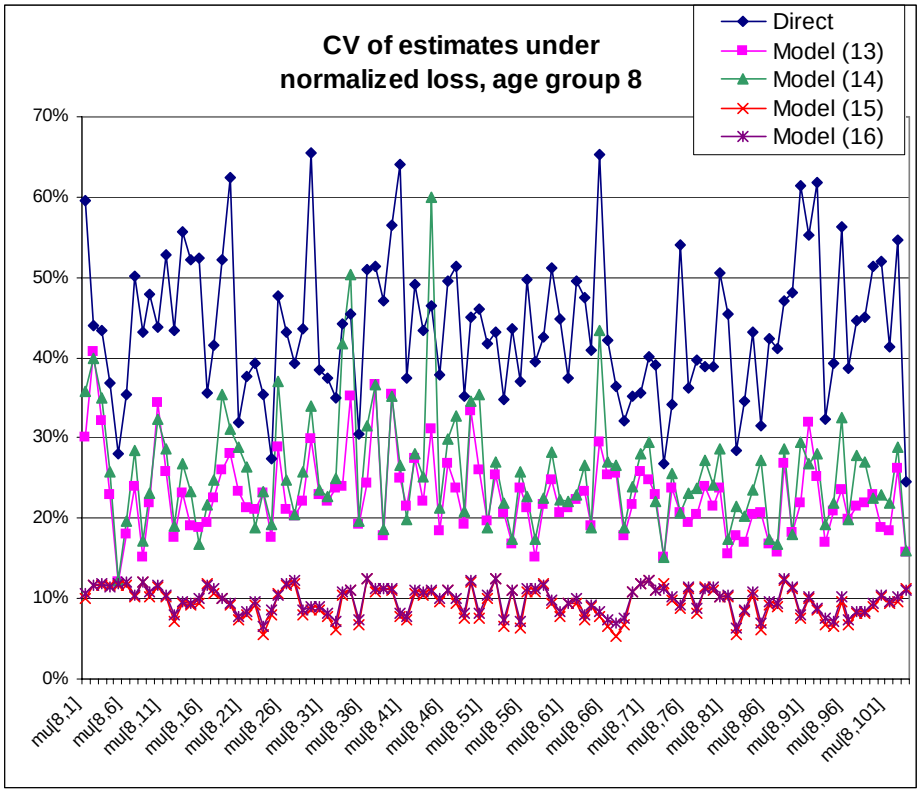

CV of estimates under normalized loss, age group 3
Direct
Model (13)
Model (14)
Model (15)
Model (16)
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[3,1]
mu[3,6]
mu[3,11]
mu[3,16]
mu[3,21]
mu[3,26]
mu[3,31]
mu[3,36]
mu[3,41]
mu[3,46]
mu[3,51]
mu[3,56]
mu[3,61]
mu[3,66]
mu[3,71]
mu[3,76]
mu[3,81]
mu[3,86]
mu[3,91]
mu[3,96]
mu[3,101]

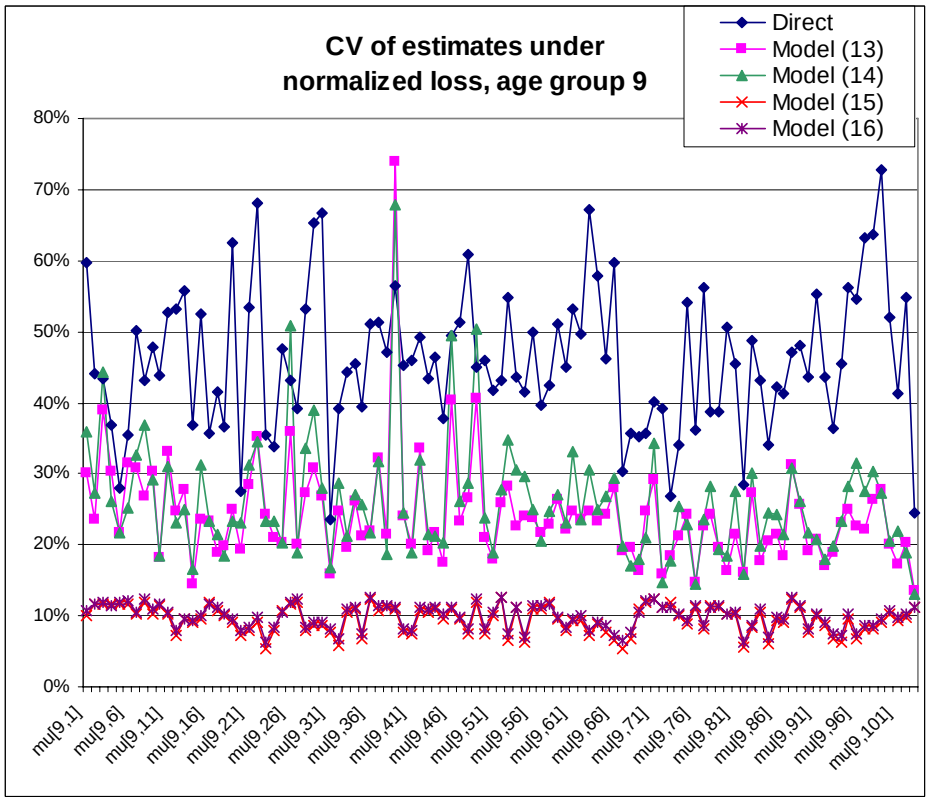

CV of estimates under
normalized loss, age group 4
Direct
Model (13)
Model (14)
Model (15)
Model (16)
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[4,1]
mu[4,6]
mu[4,11]
mu[4,16]
mu[4,21]
mu[4,26]
mu[4,31]
mu[4,36]
mu[4,41]
mu[4,46]
mu[4,51]
mu[4,56]
mu[4,61]
mu[4,66]
mu[4,71]
mu[4,76]
mu[4,81]
mu[4,86]
mu[4,91]
mu[4,96]
mu[4,101]

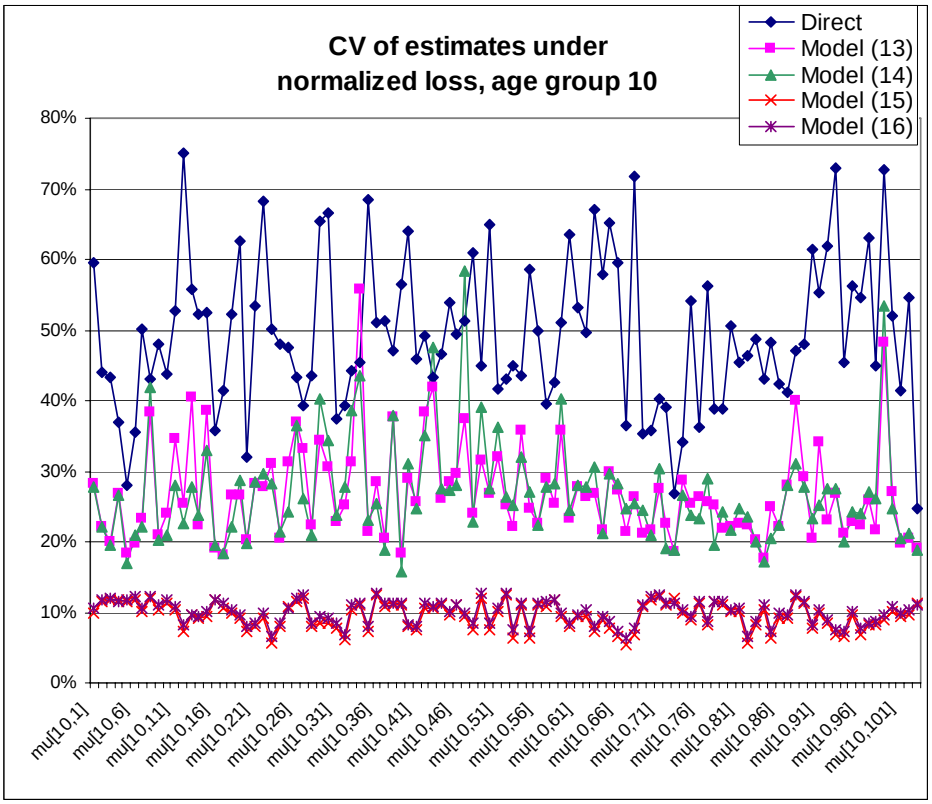

CV of estimates under
normalized loss, age group 5
Direct
Model (13)
Model (14)
Model (15)
Model (16)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[5,1]
mu[5,6]
mu[5,11]
mu[5,16]
mu[5,21]
mu[5,26]
mu[5,31]
mu[5,36]
mu[5,41]
mu[5,46]
mu[5,51]
mu[5,56]
mu[5,61]
mu[5,66]
mu[5,71]
mu[5,76]
mu[5,81]
mu[5,86]
mu[5,91]
mu[5,96]
mu[5,101]

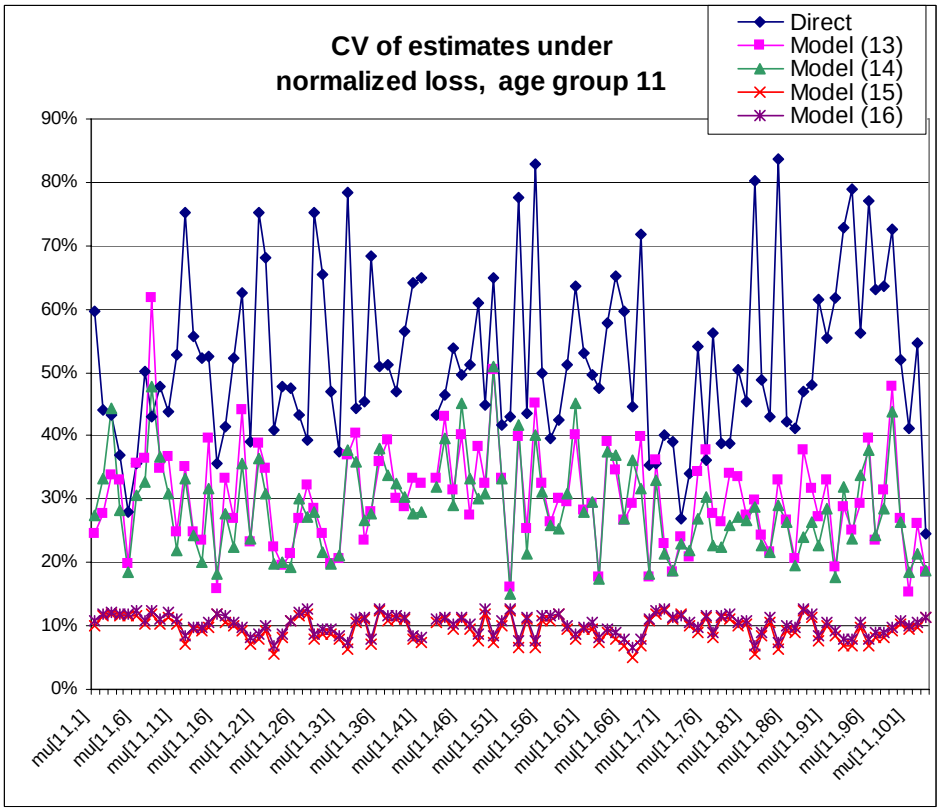

CV of estimates under normalized loss, age group 6
Direct
Model (13)
Model (14)
Model (15)
Model (16)
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[6,1]
mu[6,6]
mu[6,11]
mu[6,16]
mu[6,21]
mu[6,26]
mu[6,31]
mu[6,36]
mu[6,41]
mu[6,46]
mu[6,51]
mu[6,56]
mu[6,61]
mu[6,66]
mu[6,71]
mu[6,76]
mu[6,81]
mu[6,86]
mu[6,91]
mu[6,96]
mu[6,101]

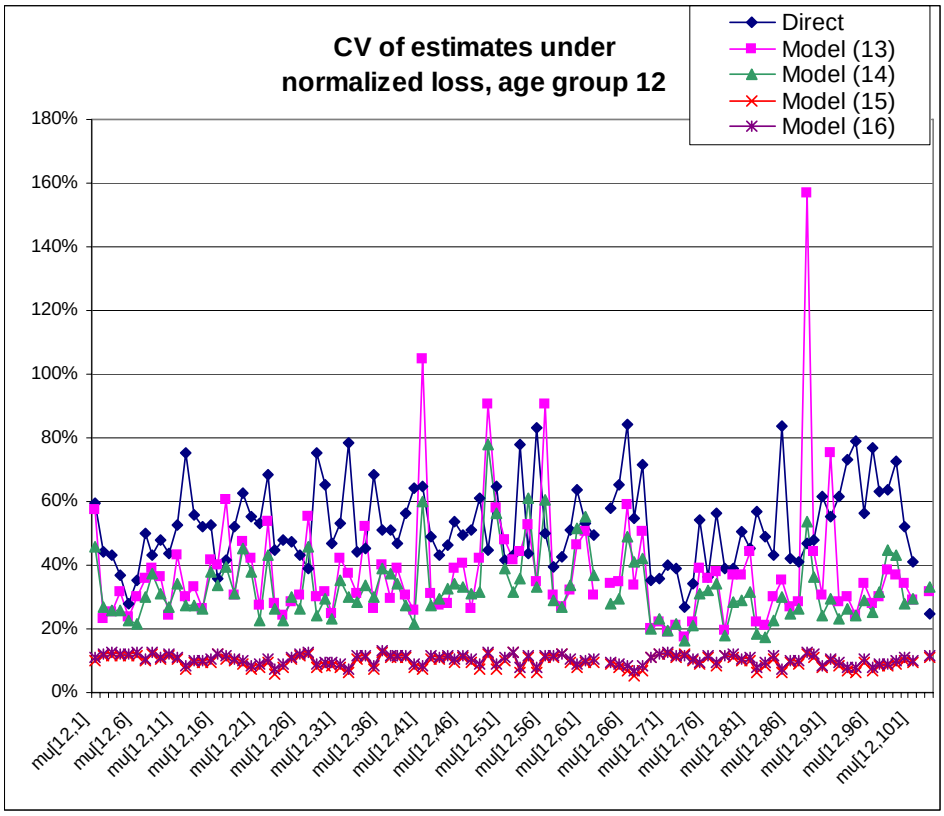

CV of estimates under normalized loss, age group 7
Direct
Model (13)
Model (14)
Model (15)
Model (16)
80
70
60
50
40
30
20
10
0
mu[7,1]
mu[7,12]
mu[7,23]
mu[7,34]
mu[7,45]
mu[7,56]
mu[7,67]
mu[7,78]
mu[7,89]
mu[7,100]

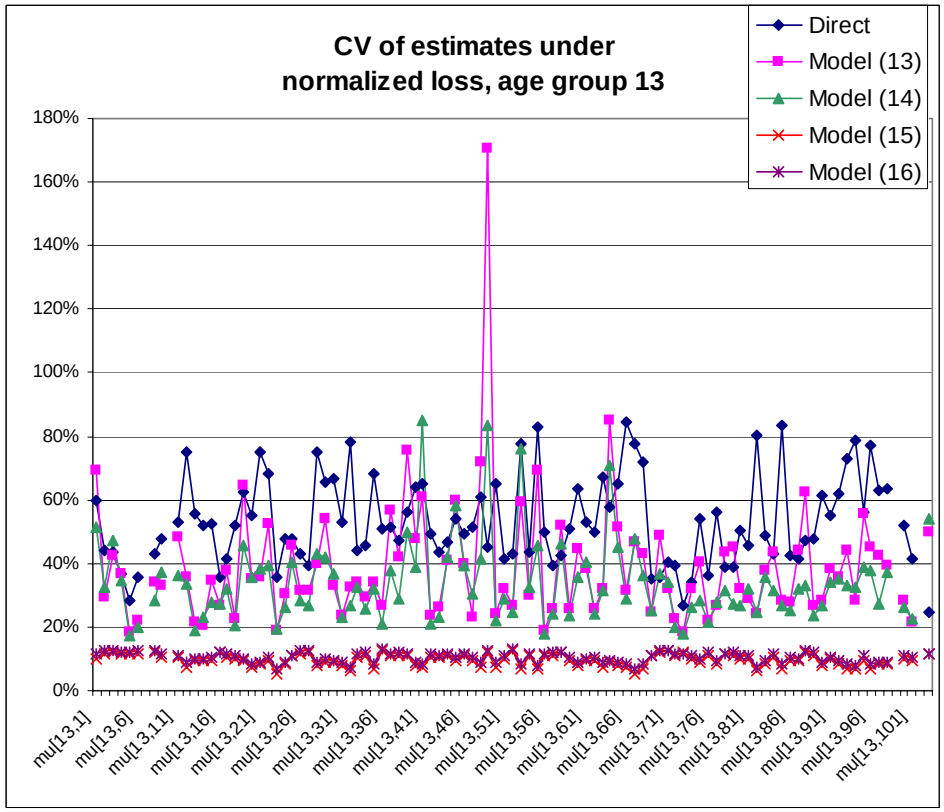

CV of estimates under
normalized loss, age group 8
Direct
Model (13)
Model (14)
Model (15)
Model (16)
70%
60%
50%
40%
30%
20%
10%
0%
mu[8,1]
mu[8,6]
mu[8,11]
mu[8,16]
mu[8,21]
mu[8,26]
mu[8,31]
mu[8,36]
mu[8,41]
mu[8,46]
mu[8,51]
mu[8,56]
mu[8,61]
mu[8,66]
mu[8,71]
mu[8,76]
mu[8,81]
mu[8,86]
mu[8,91]
mu[8,96]
mu[8,101]

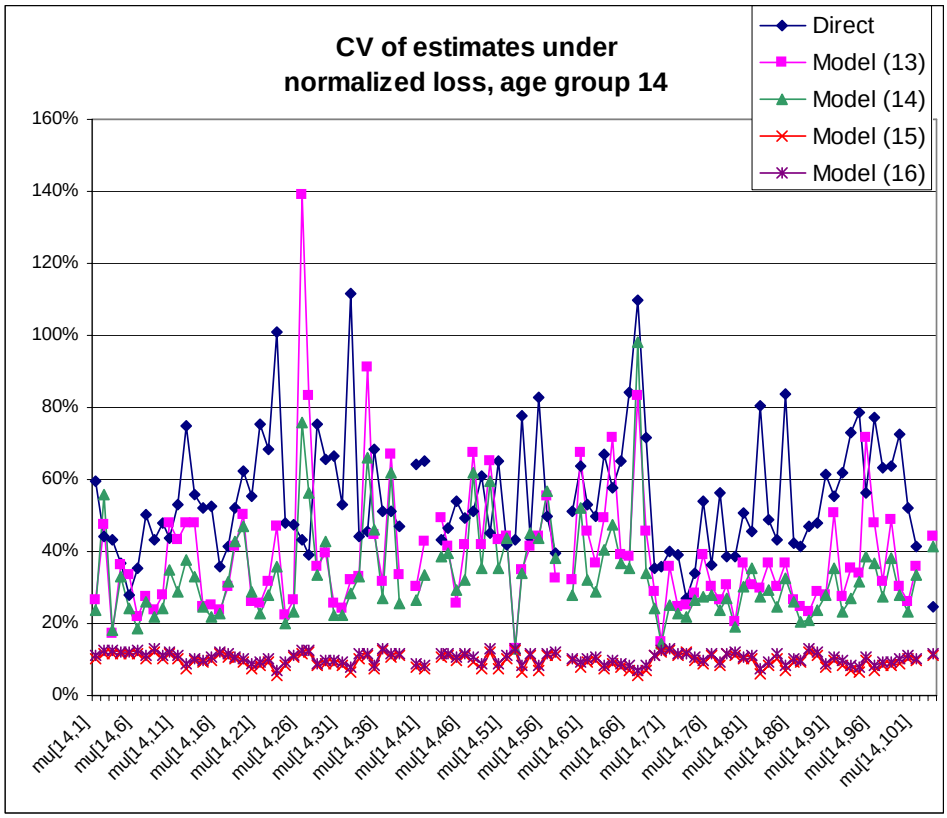

CV of estimates under
normalized loss, age group 9
Direct
Model (13)
Model (14)
Model (15)
Model (16)
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[9,1]
mu[9,6]
mu[9,11]
mu[9,16]
mu[9,21]
mu[9,26]
mu[9,31]
mu[9,36]
mu[9,41]
mu[9,46]
mu[9,51]
mu[9,56]
mu[9,61]
mu[9,66]
mu[9,71]
mu[9,76]
mu[9,81]
mu[9,86]
mu[9,91]
mu[9,96]
mu[9,101]


CV of estimates under
normalized loss, age group 10
Direct
Model (13)
Model (14)
Model (15)
Model (16)
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[10,1]
mu[10,6]
mu[10,11]
mu[10,16]
mu[10,21]
mu[10,26]
mu[10,31]
mu[10,36]
mu[10,41]
mu[10,46]
mu[10,51]
mu[10,56]
mu[10,61]
mu[10,66]
mu[10,71]
mu[10,76]
mu[10,81]
mu[10,86]
mu[10,91]
mu[10,96]
mu[10,101]

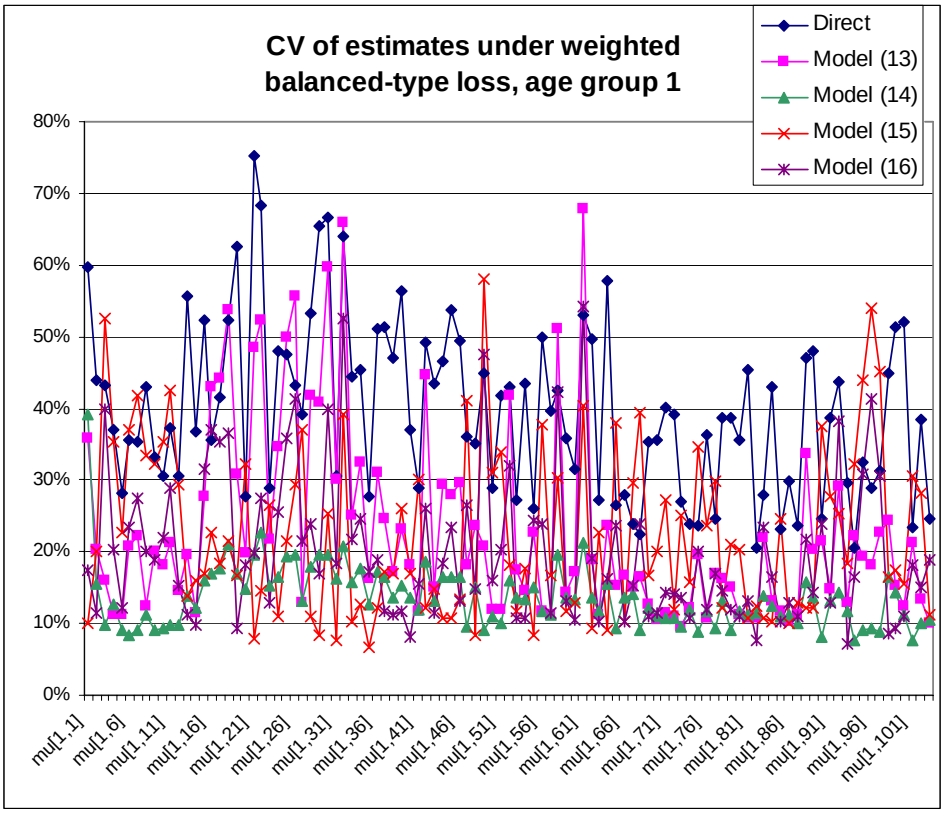

CV of estimates under
normalized loss, age group 11
Direct
Model (13)
Model (14)
Model (15)
Model (16)
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[11,1]
mu[11,6]
mu[11,11]
mu[11,16]
mu[11,21]
mu[11,26]
mu[11,31]
mu[11,36]
mu[11,41]
mu[11,46]
mu[11,51]
mu[11,56]
mu[11,61]
mu[11,66]
mu[11,71]
mu[11,76]
mu[11,81]
mu[11,86]
mu[11,91]
mu[11,96]
mu[11,101]

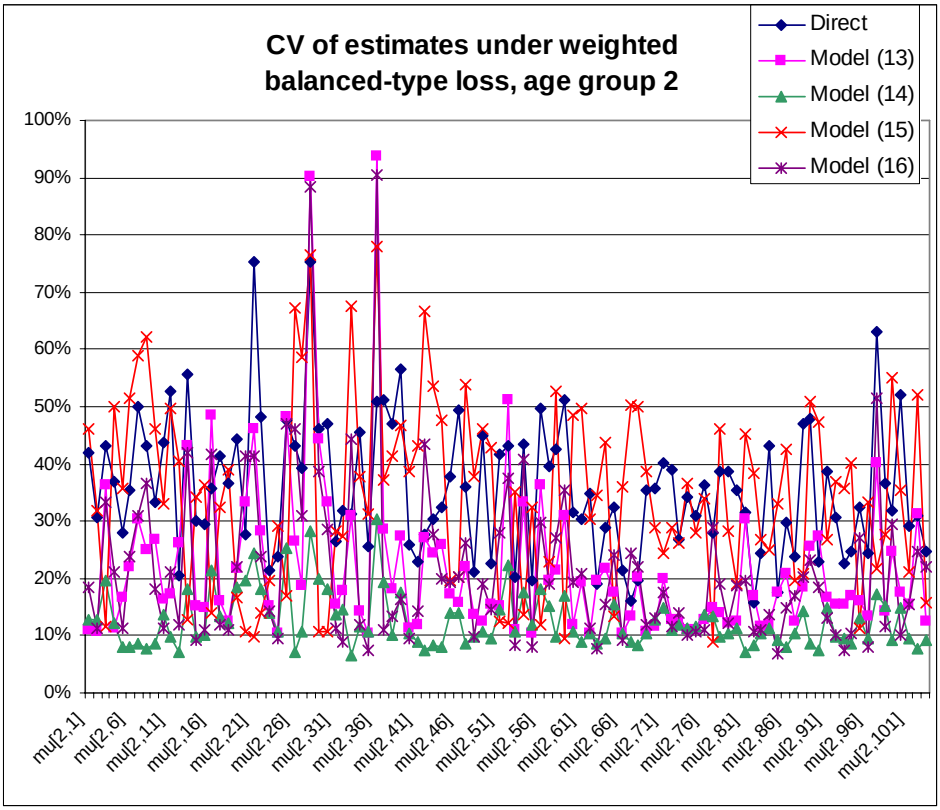

CV of estimates under
normalized loss, age group 12
Direct
Model (13)
Model (14)
Model (15)
Model (16)
180%
160%
140%
120%
100%
80%
60%
40%
20%
0%
mu[12,1]
mu[12,6]
mu[12,11]
mu[12,16]
mu[12,21]
mu[12,26]
mu[12,31]
mu[12,36]
mu[12,41]
mu[12,46]
mu[12,51]
mu[12,56]
mu[12,61]
mu[12,66]
mu[12,71]
mu[12,76]
mu[12,81]
mu[12,86]
mu[12,91]
mu[12,96]
mu[12,101]

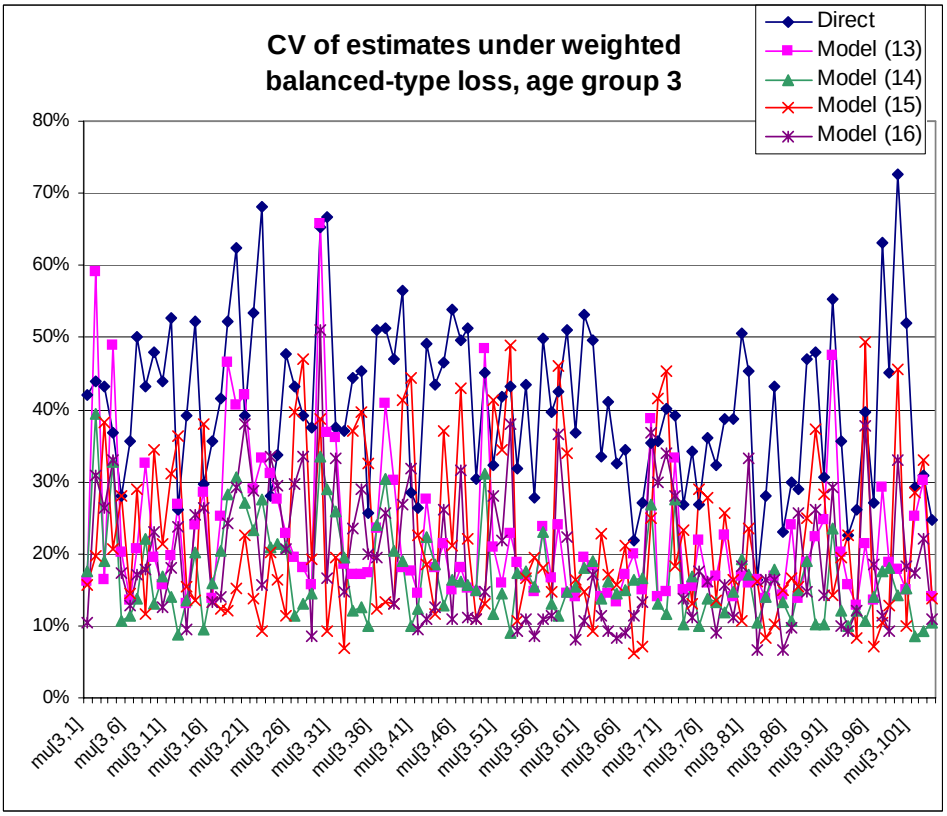

CV of estimates under
normalized loss, age group 13
Direct
Model (13)
Model (14)
Model (15)
Model (16)
180%
160%
140%
120%
100%
80%
60%
40%
20%
0%
mu[13,1]
mu[13,6]
mu[13,11]
mu[13,16]
mu[13,21]
mu[13,26]
mu[13,31]
mu[13,36]
mu[13,41]
mu[13,46]
mu[13,51]
mu[13,56]
mu[13,61]
mu[13,66]
mu[13,71]
mu[13,76]
mu[13,81]
mu[13,86]
mu[13,91]
mu[13,96]
mu[13,101]

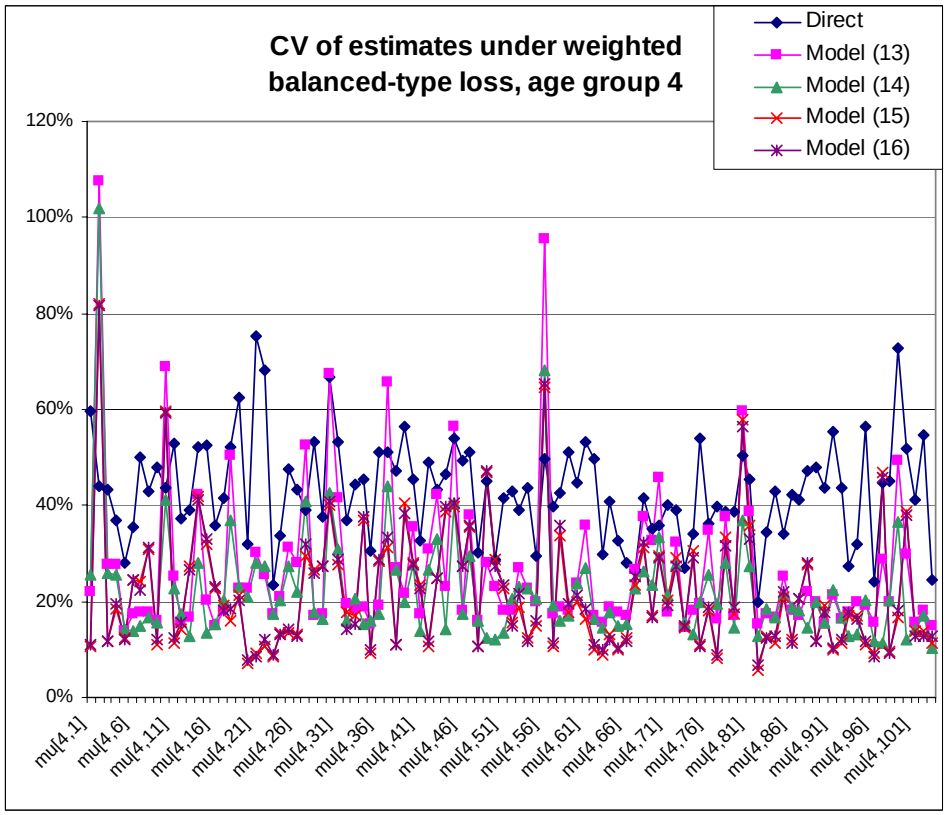

CV of estimates under
normalized loss, age group 14
Direct
Model (13)
Model (14)
Model (15)
Model (16)
160%
140%
120%
100%
80%
60%
40%
20%
0%
mu[14,1]
mu[14,6]
mu[14,11]
mu[14,16]
mu[14,21]
mu[14,26]
mu[14,31]
mu[14,36]
mu[14,41]
mu[14,46]
mu[14,51]
mu[14,56]
mu[14,61]
mu[14,66]
mu[14,71]
mu[14,76]
mu[14,81]
mu[14,86]
mu[14,91]
mu[14,96]
mu[14,101]

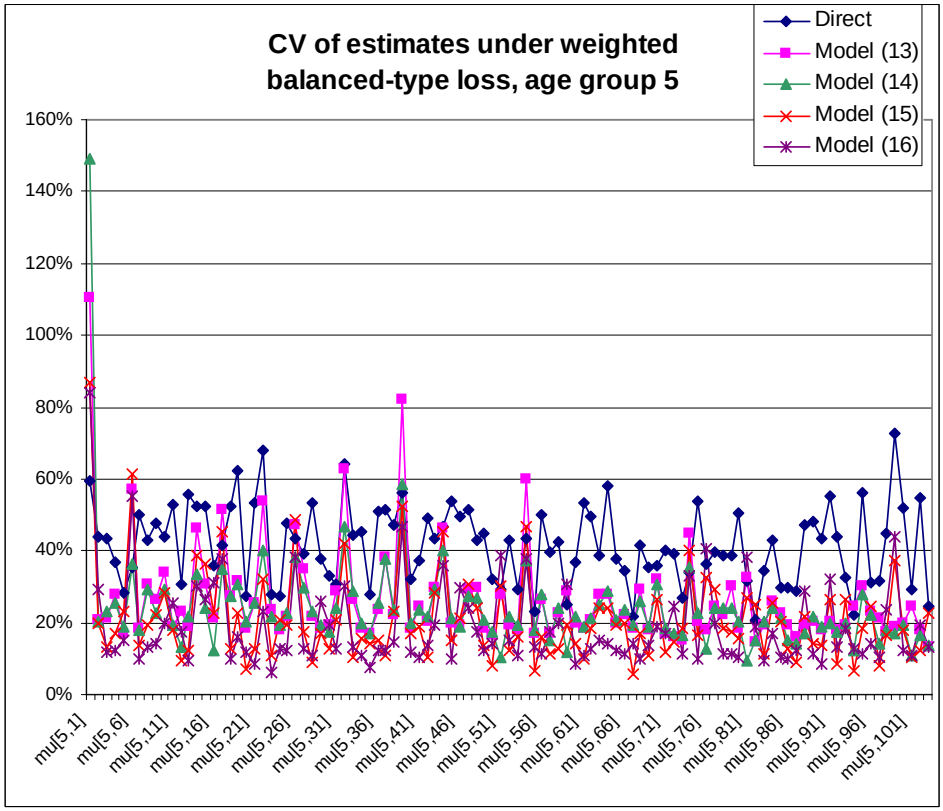
CV of estimates under WBL:

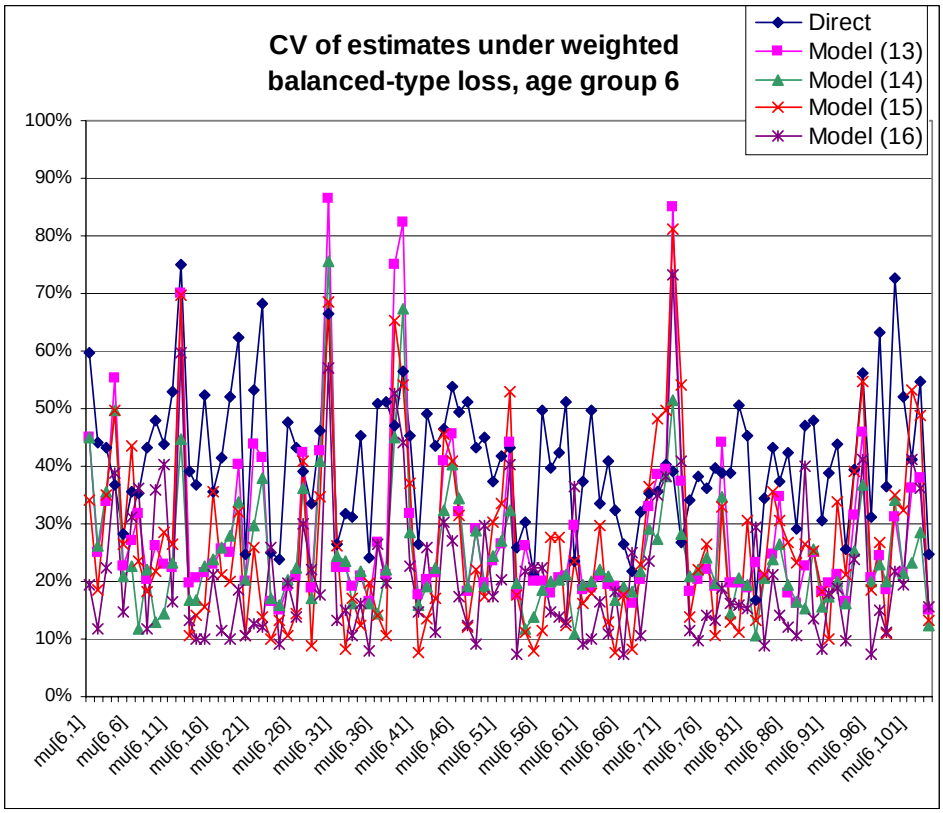


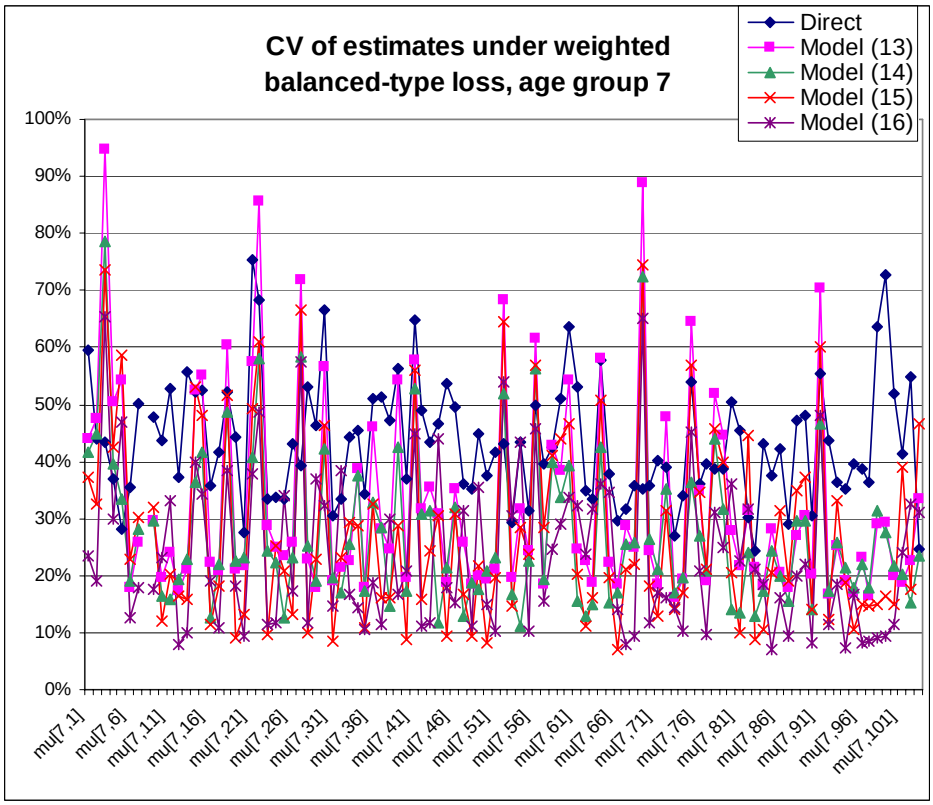

CV of estimates under weighted balanced-type loss, age group 2
Direct
Model (13)
Model (14)
Model (15)
Model (16)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[2,1]
mu[2,6]
mu[2,11]
mu[2,16]
mu[2,21]
mu[2,26]
mu[2,31]
mu[2,36]
mu[2,41]
mu[2,46]
mu[2,51]
mu[2,56]
mu[2,61]
mu[2,66]
mu[2,71]
mu[2,76]
mu[2,81]
mu[2,86]
mu[2,91]
mu[2,96]
mu[2,101]

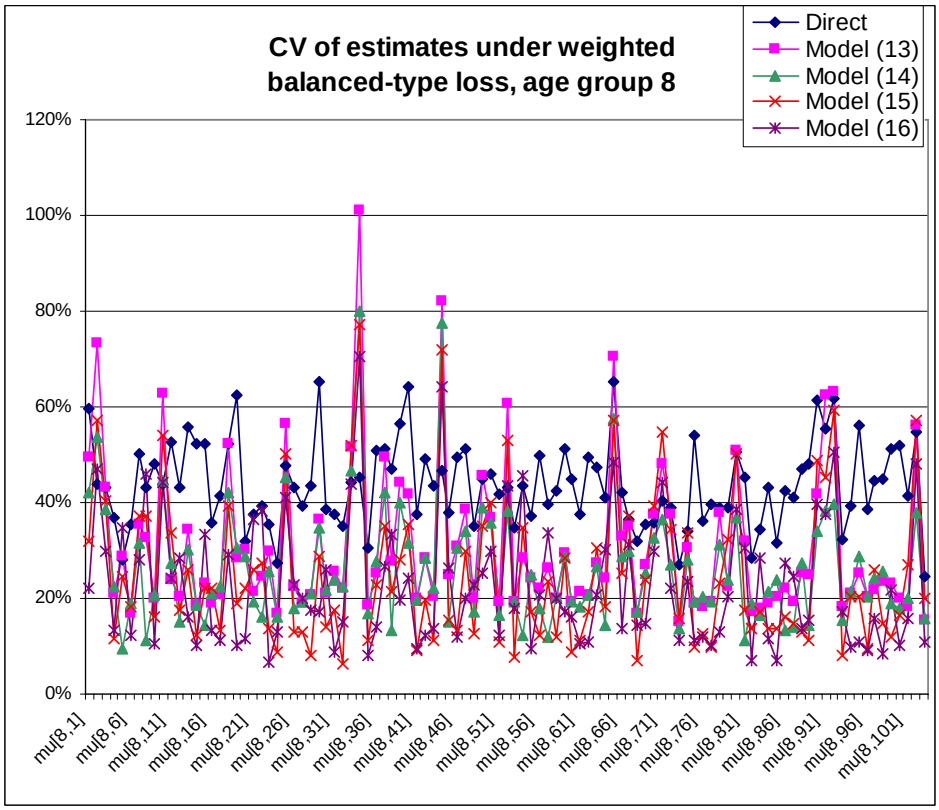

CV of estimates under weighted balanced-type loss, age group 3
Direct
Model (13)
Model (14)
Model (15)
Model (16)
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[3,1]
mu[3,6]
mu[3,11]
mu[3,16]
mu[3,21]
mu[3,26]
mu[3,31]
mu[3,36]
mu[3,41]
mu[3,46]
mu[3,51]
mu[3,56]
mu[3,61]
mu[3,66]
mu[3,71]
mu[3,76]
mu[3,81]
mu[3,86]
mu[3,91]
mu[3,96]
mu[3,101]

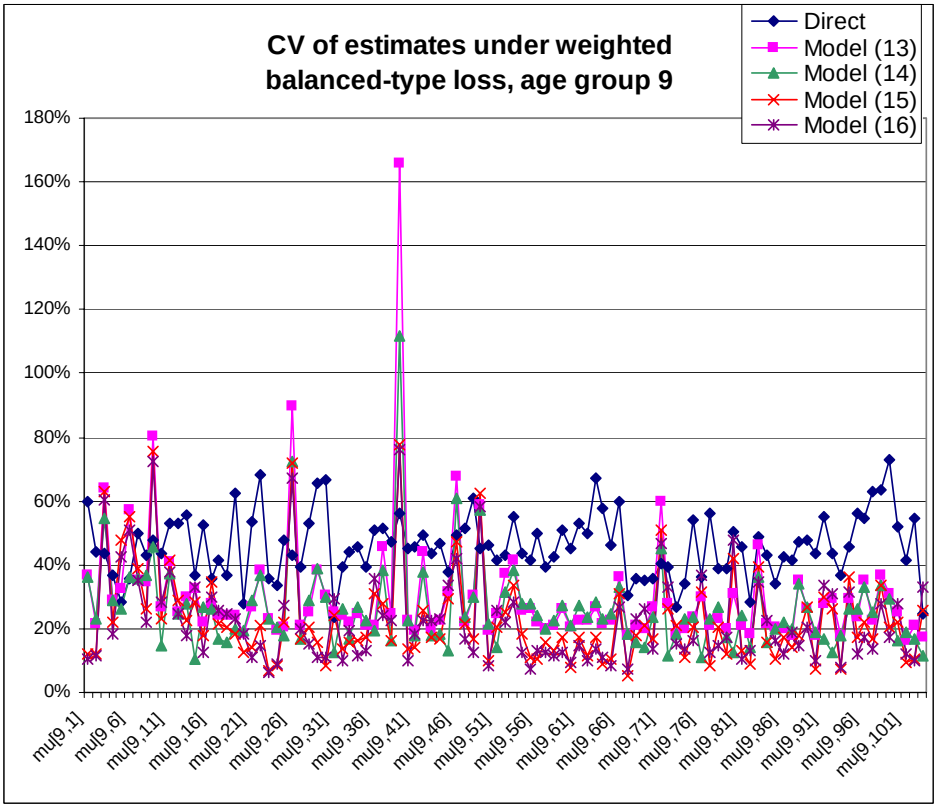

CV of estimates under weighted balanced-type loss, age group 4
Direct
Model (13)
Model (14)
Model (15)
Model (16)
120%
100%
80%
60%
40%
20%
0%
mu[4,1]
mu[4,6]
mu[4,11]
mu[4,16]
mu[4,21]
mu[4,26]
mu[4,31]
mu[4,36]
mu[4,41]
mu[4,46]
mu[4,51]
mu[4,56]
mu[4,61]
mu[4,66]
mu[4,71]
mu[4,76]
mu[4,81]
mu[4,86]
mu[4,91]
mu[4,96]
mu[4,101]

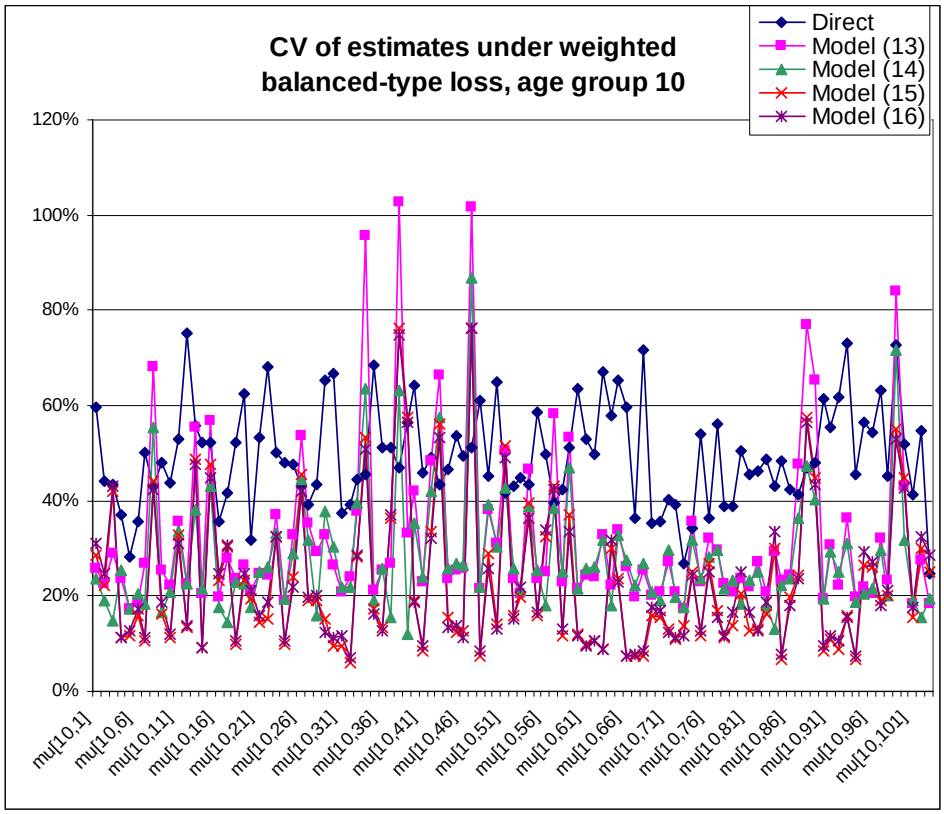

CV of estimates under weighted balanced-type loss, age group 5
Direct
Model (13)
Model (14)
Model (15)
Model (16)
160%
140%
120%
100%
80%
60%
40%
20%
0%
mu[5,1]
mu[5,6]
mu[5,11]
mu[5,16]
mu[5,21]
mu[5,26]
mu[5,31]
mu[5,36]
mu[5,41]
mu[5,46]
mu[5,51]
mu[5,56]
mu[5,61]
mu[5,66]
mu[5,71]
mu[5,76]
mu[5,81]
mu[5,86]
mu[5,91]
mu[5,96]
mu[5,101]

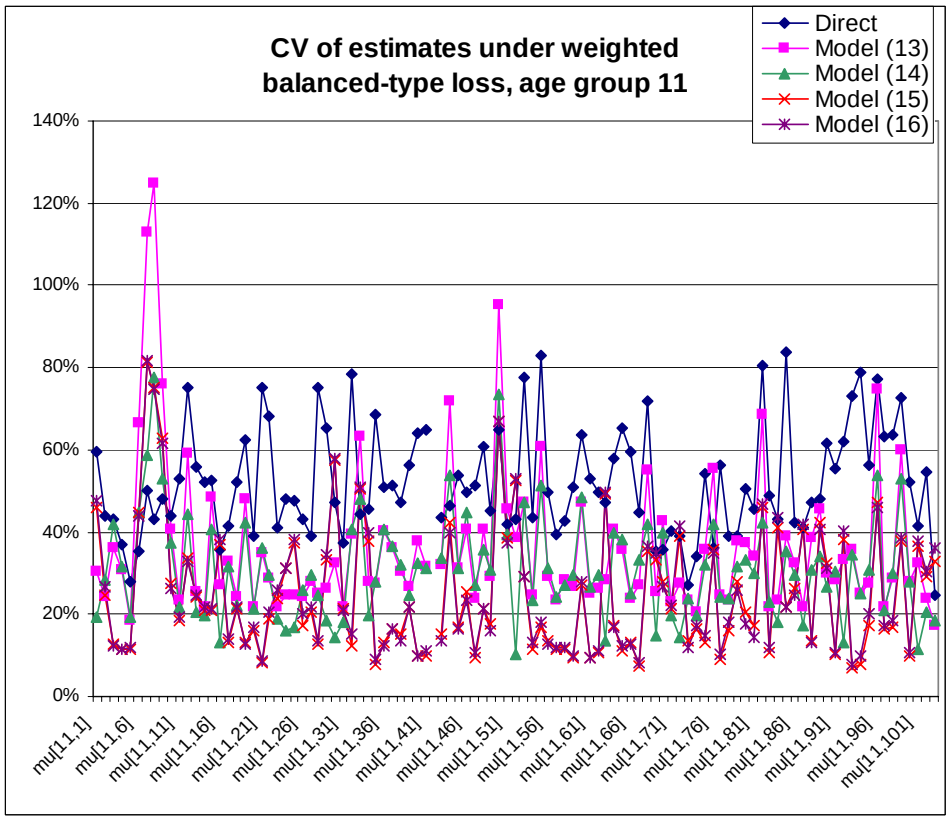

CV of estimates under weighted balanced-type loss, age group 6
Direct
Model (13)
Model (14)
Model (15)
Model (16)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[6,1]
mu[6,6]
mu[6,11]
mu[6,16]
mu[6,21]
mu[6,26]
mu[6,31]
mu[6,36]
mu[6,41]
mu[6,46]
mu[6,51]
mu[6,56]
mu[6,61]
mu[6,66]
mu[6,71]
mu[6,76]
mu[6,81]
mu[6,86]
mu[6,91]
mu[6,96]
mu[6,101]

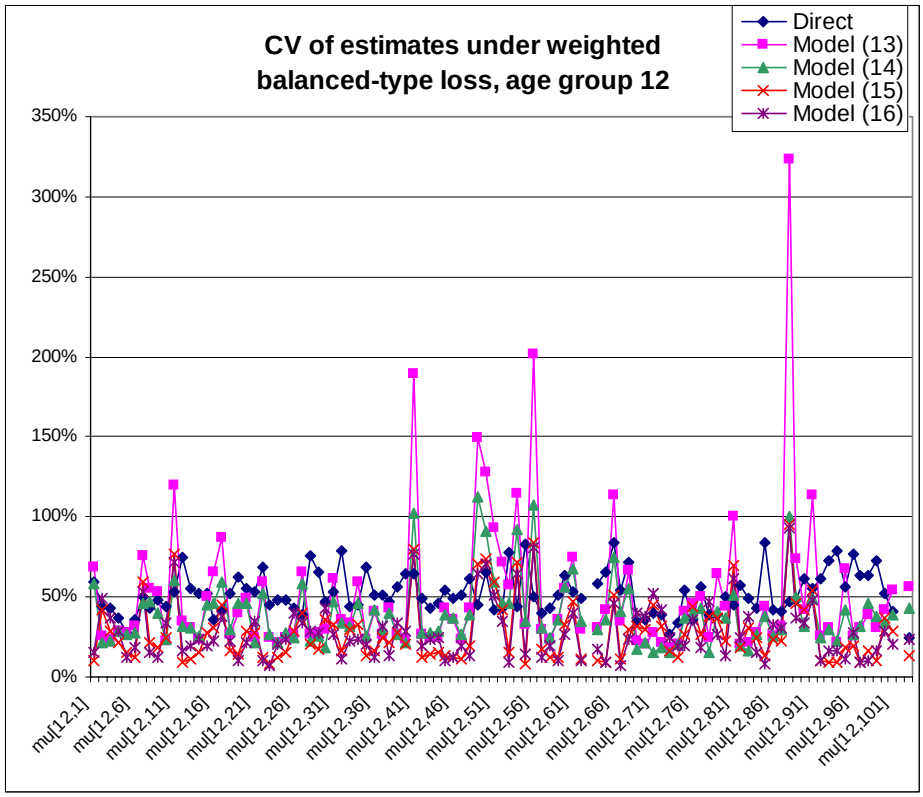

CV of estimates under weighted balanced-type loss, age group 7
Direct
Model (13)
Model (14)
Model (15)
Model (16)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[7,1]
mu[7,6]
mu[7,11]
mu[7,16]
mu[7,21]
mu[7,26]
mu[7,31]
mu[7,36]
mu[7,41]
mu[7,46]
mu[7,51]
mu[7,56]
mu[7,61]
mu[7,66]
mu[7,71]
mu[7,76]
mu[7,81]
mu[7,86]
mu[7,91]
mu[7,96]
mu[7,101]

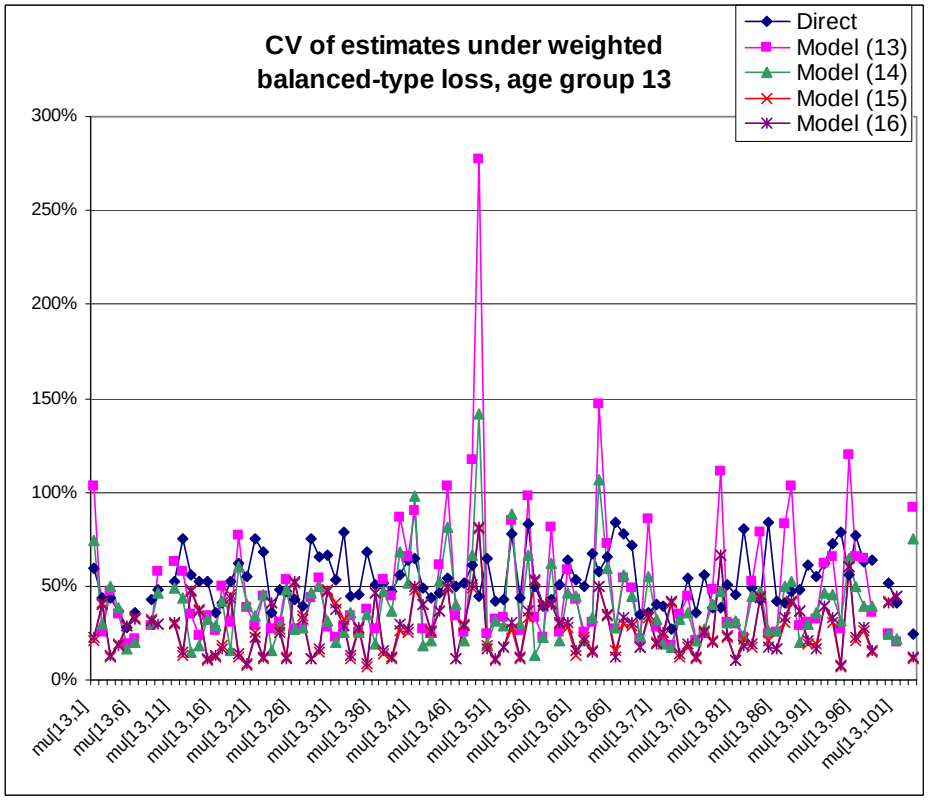

CV of estimates under weighted balanced-type loss, age group 8
Direct
Model (13)
Model (14)
Model (15)
Model (16)
120%
100%
80%
60%
40%
20%
0%
mu[8,1]
mu[8,6]
mu[8,11]
mu[8,16]
mu[8,21]
mu[8,26]
mu[8,31]
mu[8,36]
mu[8,41]
mu[8,46]
mu[8,51]
mu[8,56]
mu[8,61]
mu[8,66]
mu[8,71]
mu[8,76]
mu[8,81]
mu[8,86]
mu[8,91]
mu[8,96]
mu[8,101]

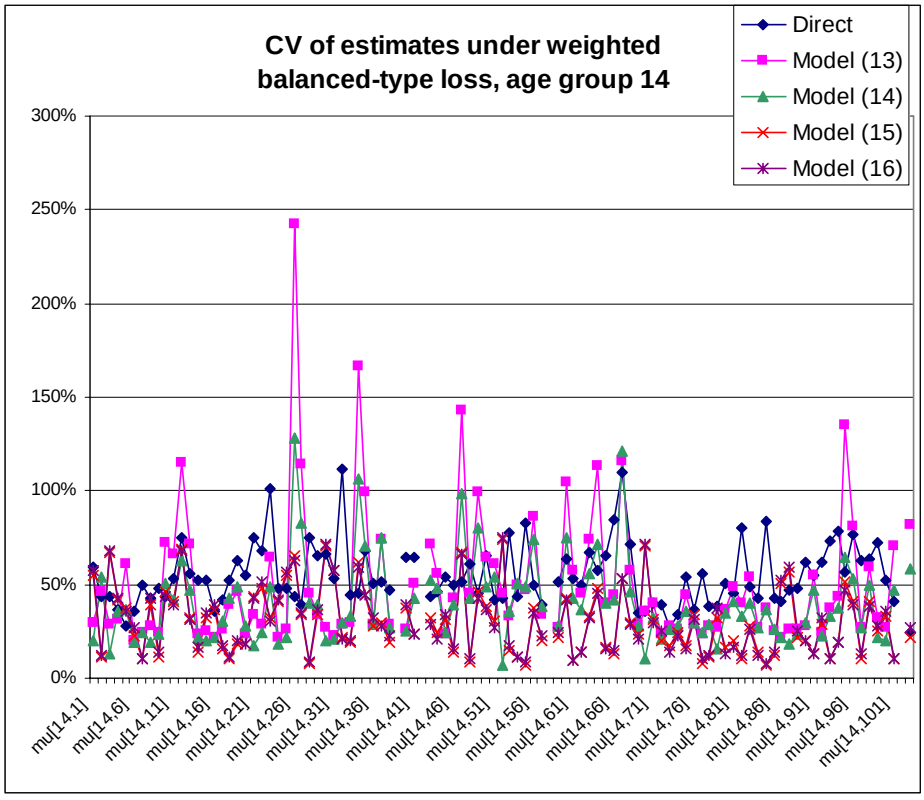

CV of estimates under weighted balanced-type loss, age group 9
Direct
Model (13)
Model (14)
Model (15)
Model (16)
180%
160%
140%
120%
100%
80%
60%
40%
20%
0%
mu[9,1]
mu[9,6]
mu[9,11]
mu[9,16]
mu[9,21]
mu[9,26]
mu[9,31]
mu[9,36]
mu[9,41]
mu[9,46]
mu[9,51]
mu[9,56]
mu[9,61]
mu[9,66]
mu[9,71]
mu[9,76]
mu[9,81]
mu[9,86]
mu[9,91]
mu[9,96]
mu[9,101]


CV of estimates under weighted balanced-type loss, age group 10
Direct
Model (13)
Model (14)
Model (15)
Model (16)
120%
100%
80%
60%
40%
20%
0%
mu[10,1]
mu[10,6]
mu[10,11]
mu[10,16]
mu[10,21]
mu[10,26]
mu[10,31]
mu[10,36]
mu[10,41]
mu[10,46]
mu[10,51]
mu[10,56]
mu[10,61]
mu[10,66]
mu[10,71]
mu[10,76]
mu[10,81]
mu[10,86]
mu[10,91]
mu[10,96]
mu[10,101]

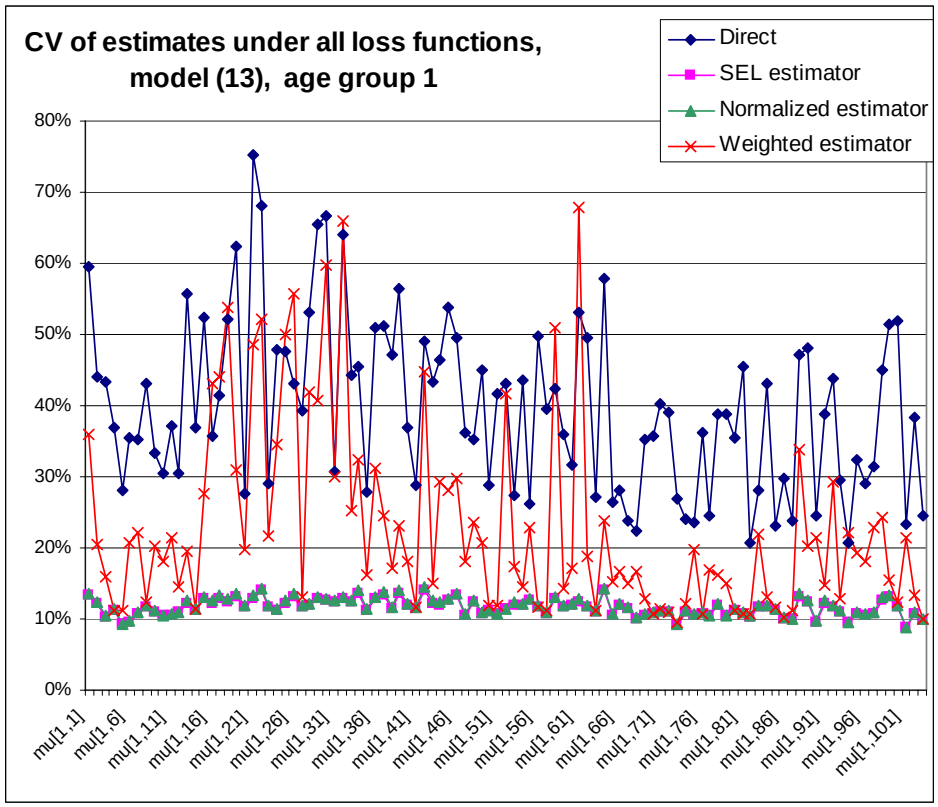

CV of estimates under weighted balanced-type loss, age group 11
Direct
Model (13)
Model (14)
Model (15)
Model (16)
140%
120%
100%
80%
60%
40%
20%
0%
mu[11,1]
mu[11,6]
mu[11,11]
mu[11,16]
mu[11,21]
mu[11,26]
mu[11,31]
mu[11,36]
mu[11,41]
mu[11,46]
mu[11,51]
mu[11,56]
mu[11,61]
mu[11,66]
mu[11,71]
mu[11,76]
mu[11,81]
mu[11,86]
mu[11,91]
mu[11,96]
mu[11,101]

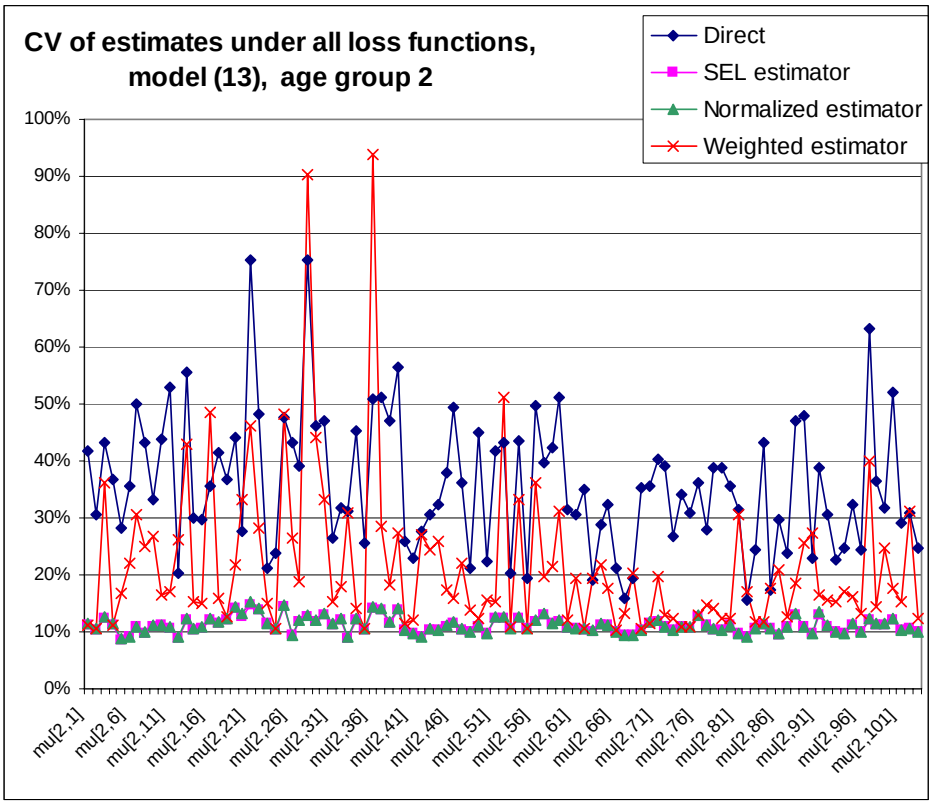

CV of estimates under weighted balanced-type loss, age group 12
Direct
Model (13)
Model (14)
Model (15)
Model (16)
350%
300%
250%
200%
150%
100%
50%
0%
mu[12,1]
mu[12,6]
mu[12,11]
mu[12,16]
mu[12,21]
mu[12,26]
mu[12,31]
mu[12,36]
mu[12,41]
mu[12,46]
mu[12,51]
mu[12,56]
mu[12,61]
mu[12,66]
mu[12,71]
mu[12,76]
mu[12,81]
mu[12,86]
mu[12,91]
mu[12,96]
mu[12,101]

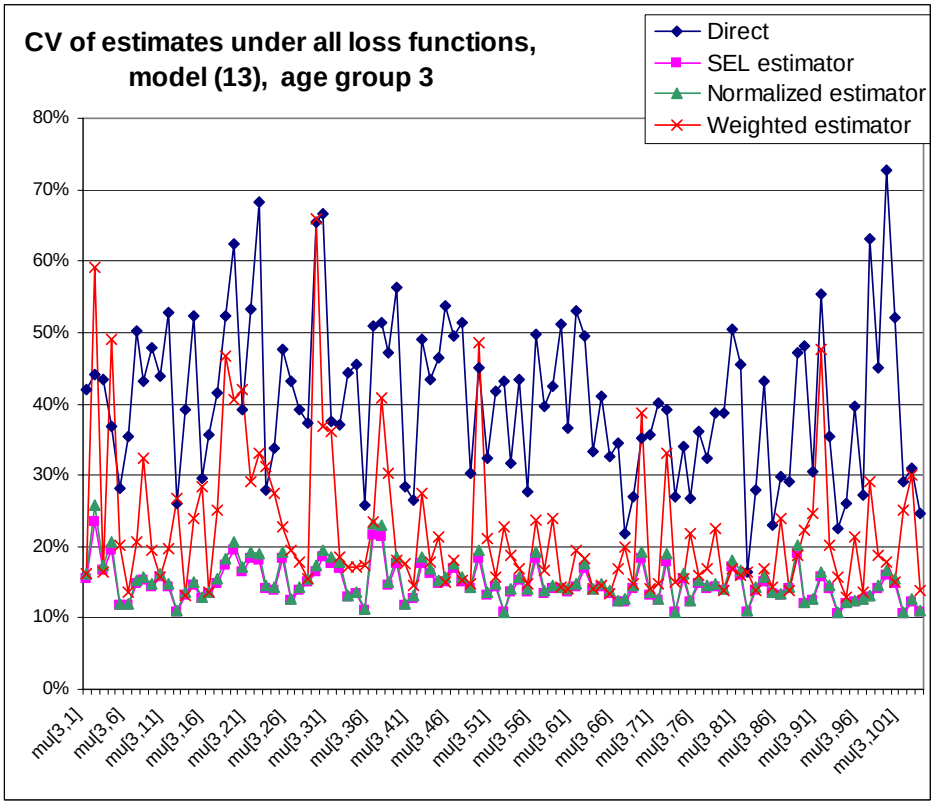

CV of estimates under weighted balanced-type loss, age group 13
Direct
Model (13)
Model (14)
Model (15)
Model (16)
300%
250%
200%
150%
100%
50%
0%
mu[13,1]
mu[13,6]
mu[13,11]
mu[13,16]
mu[13,21]
mu[13,26]
mu[13,31]
mu[13,36]
mu[13,41]
mu[13,46]
mu[13,51]
mu[13,56]
mu[13,61]
mu[13,66]
mu[13,71]
mu[13,76]
mu[13,81]
mu[13,86]
mu[13,91]
mu[13,96]
mu[13,101]

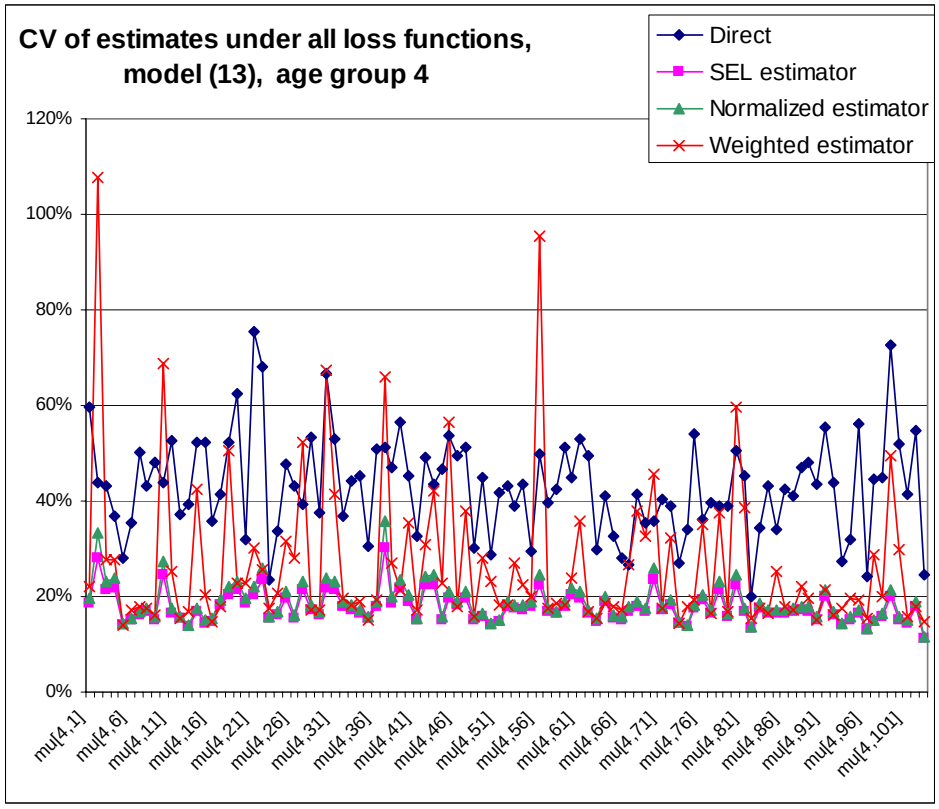

CV of estimates under weighted balanced-type loss, age group 14
Direct
Model (13)
Model (14)
Model (15)
Model (16)
300%
250%
200%
150%
100%
50%
0%
mu[14,1]
mu[14,6]
mu[14,11]
mu[14,16]
mu[14,21]
mu[14,26]
mu[14,31]
mu[14,36]
mu[14,41]
mu[14,46]
mu[14,51]
mu[14,56]
mu[14,61]
mu[14,66]
mu[14,71]
mu[14,76]
mu[14,81]
mu[14,86]
mu[14,91]
mu[14,96]
mu[14,101]

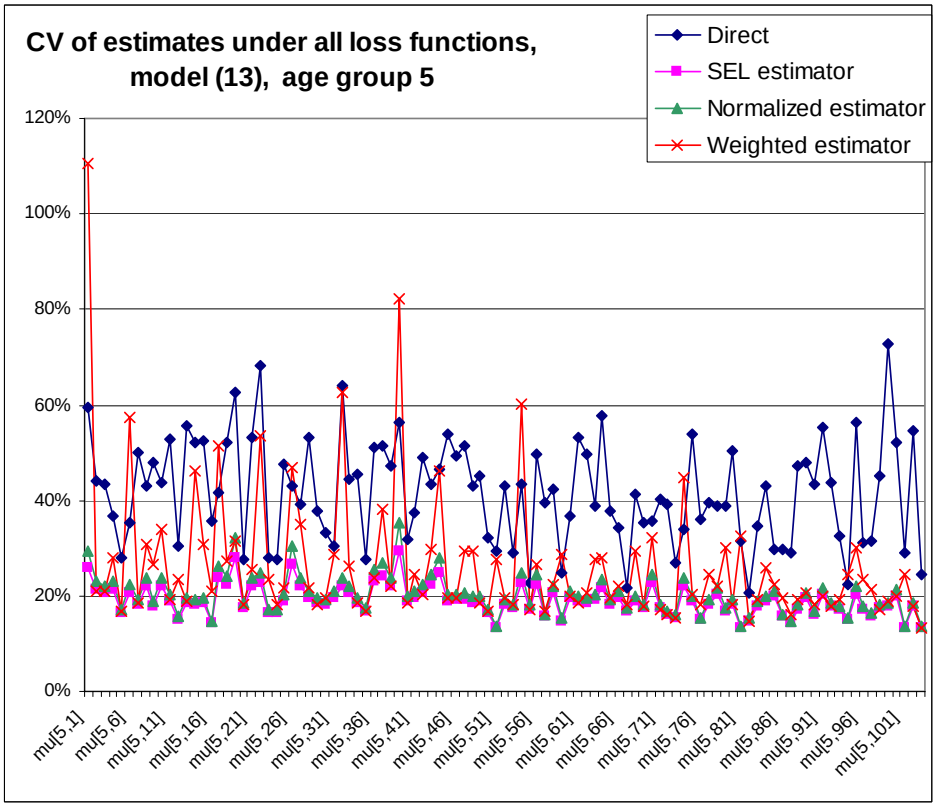
Below, CV of estimates under all three loss functions are presented for one model at a time, graphed by age group. CV of model (13) estimates:

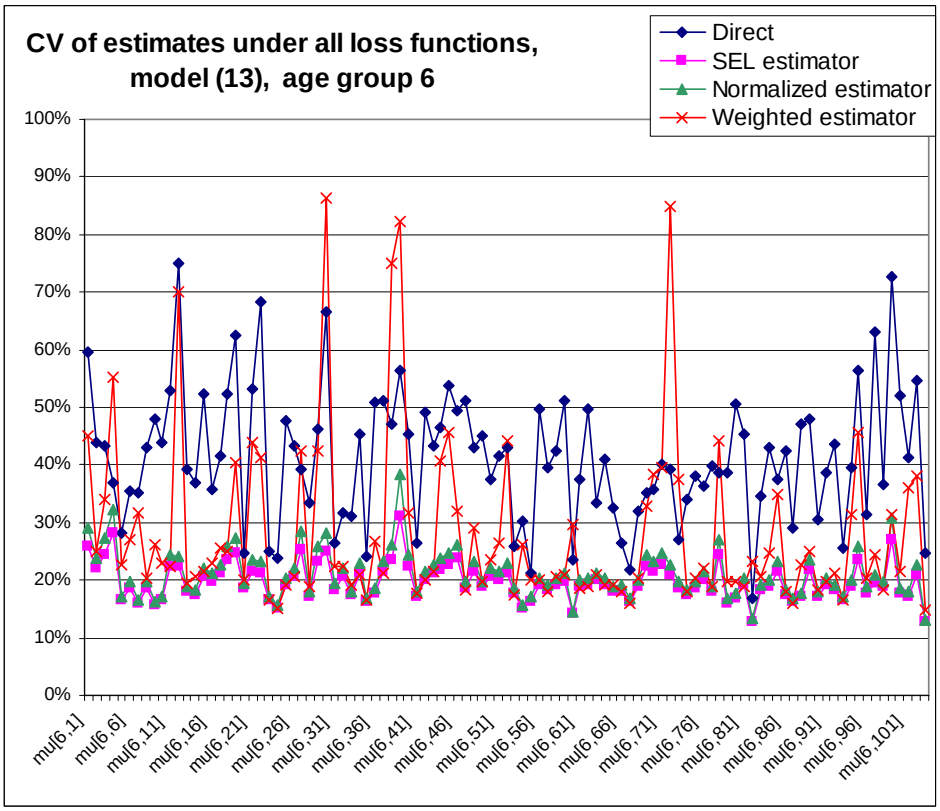


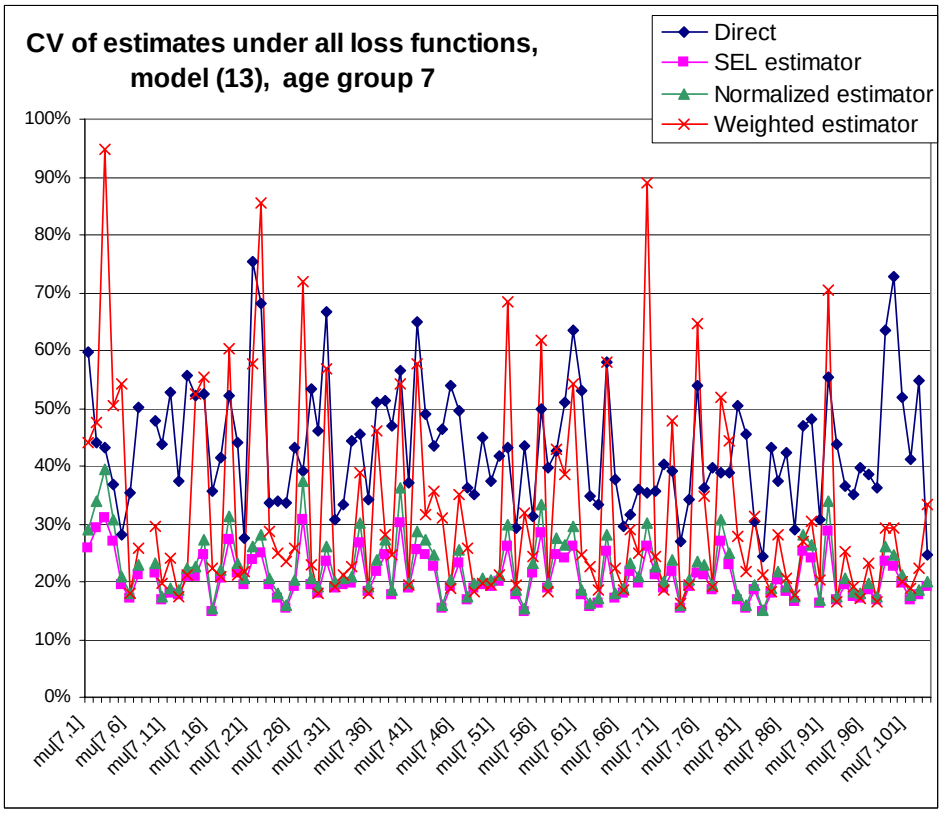

CV of estimates under all loss functions, model (13), age group 2
Direct
SEL estimator
Normalized estimator
Weighted estimator
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[2,1]
mu[2,6]
mu[2,11]
mu[2,16]
mu[2,21]
mu[2,26]
mu[2,31]
mu[2,36]
mu[2,41]
mu[2,46]
mu[2,51]
mu[2,56]
mu[2,61]
mu[2,66]
mu[2,71]
mu[2,76]
mu[2,81]
mu[2,86]
mu[2,91]
mu[2,96]
mu[2,101]

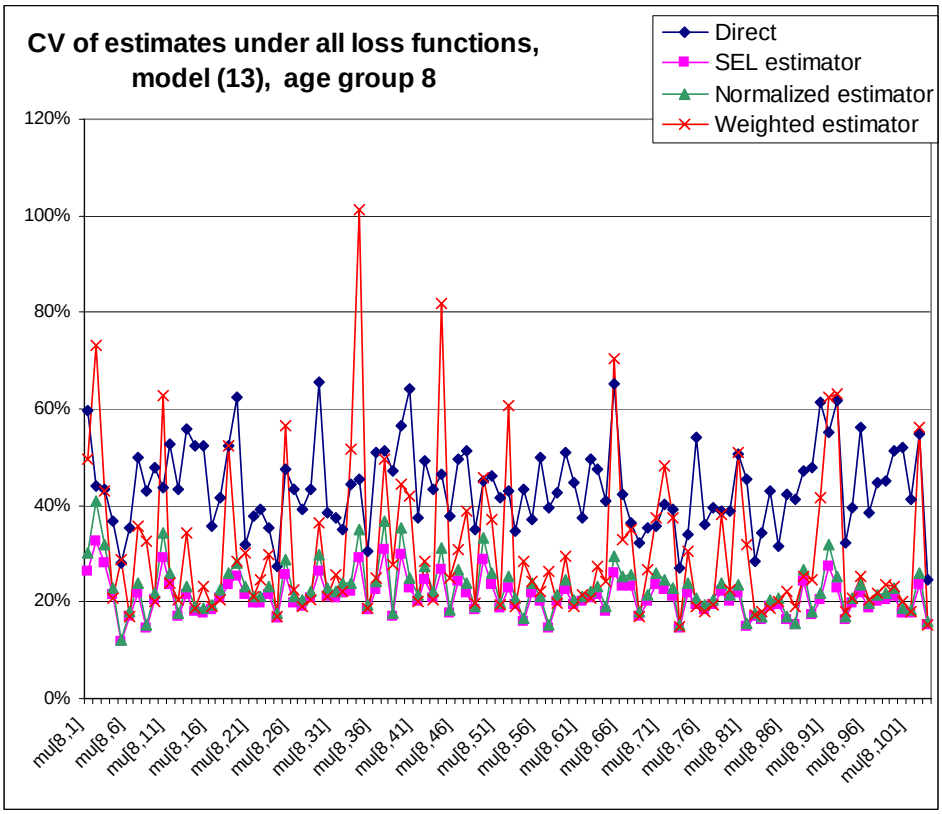

CV of estimates under all loss functions, model (13), age group 3
Direct
SEL estimator
Normalized estimator
Weighted estimator
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[3,1]
mu[3,6]
mu[3,11]
mu[3,16]
mu[3,21]
mu[3,26]
mu[3,31]
mu[3,36]
mu[3,41]
mu[3,46]
mu[3,51]
mu[3,56]
mu[3,61]
mu[3,66]
mu[3,71]
mu[3,76]
mu[3,81]
mu[3,86]
mu[3,91]
mu[3,96]
mu[3,101]

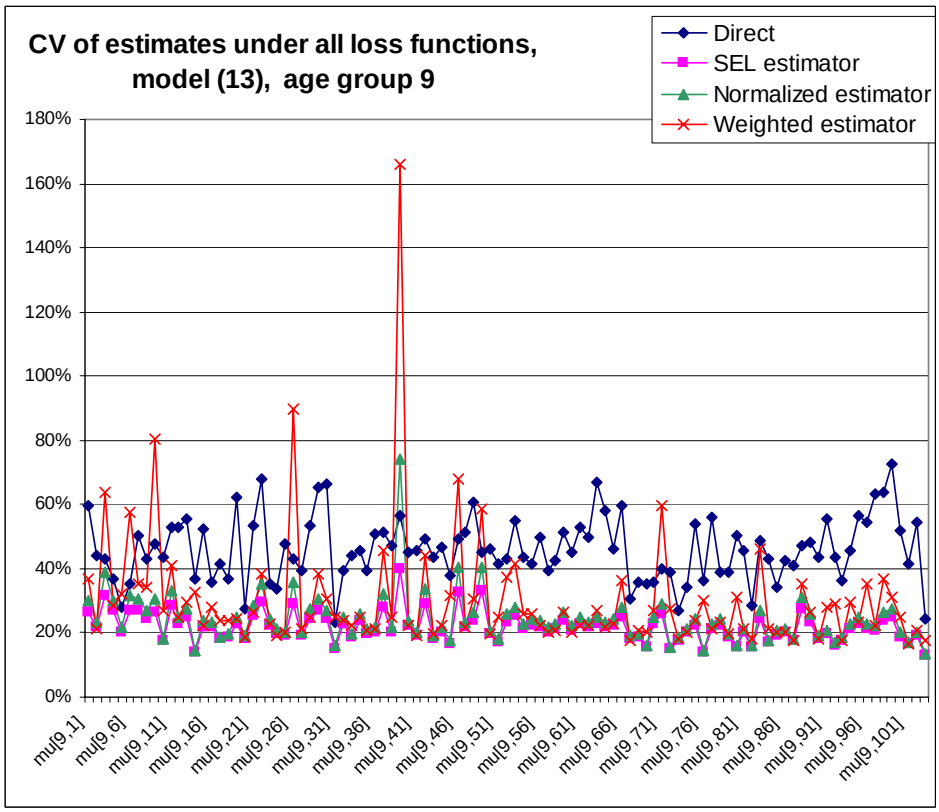

CV of estimates under all loss functions, model (13), age group 4
Direct
SEL estimator
Normalized estimator
Weighted estimator
120%
100%
80%
60%
40%
20%
0%
mu[4,1]
mu[4,6]
mu[4,11]
mu[4,16]
mu[4,21]
mu[4,26]
mu[4,31]
mu[4,36]
mu[4,41]
mu[4,46]
mu[4,51]
mu[4,56]
mu[4,61]
mu[4,66]
mu[4,71]
mu[4,76]
mu[4,81]
mu[4,86]
mu[4,91]
mu[4,96]
mu[4,101]

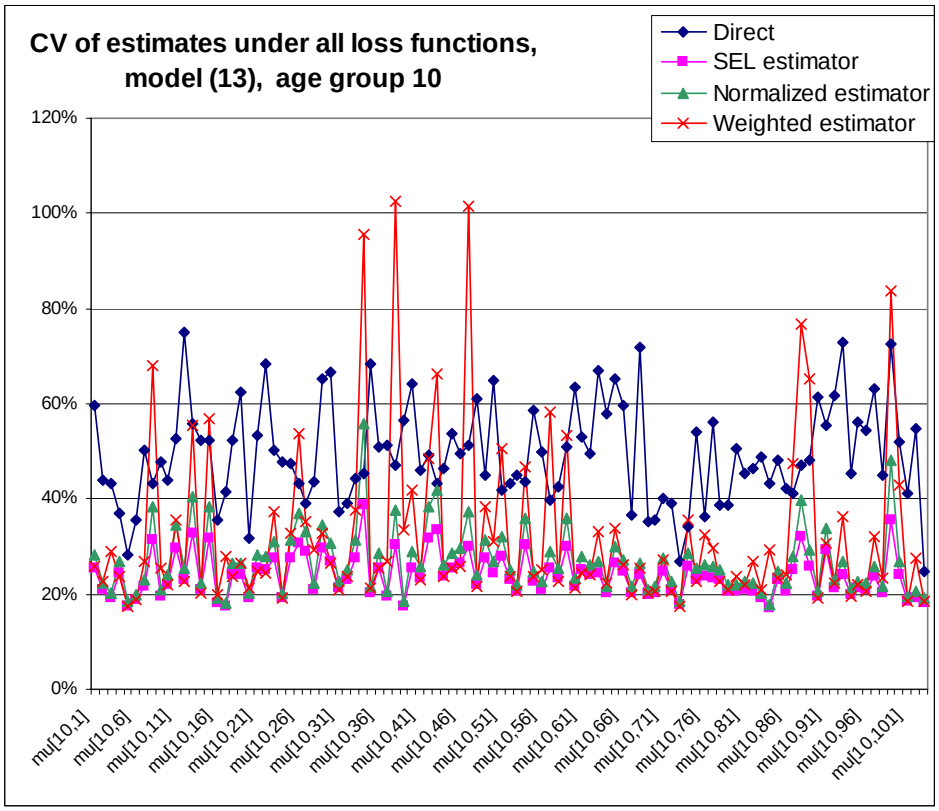

CV of estimates under all loss functions, model (13), age group 5
Direct
SEL estimator
Normalized estimator
Weighted estimator
120%
100%
80%
60%
40%
20%
0%
mu[5,1]
mu[5,6]
mu[5,11]
mu[5,16]
mu[5,21]
mu[5,26]
mu[5,31]
mu[5,36]
mu[5,41]
mu[5,46]
mu[5,51]
mu[5,56]
mu[5,61]
mu[5,66]
mu[5,71]
mu[5,76]
mu[5,81]
mu[5,86]
mu[5,91]
mu[5,96]
mu[5,101]

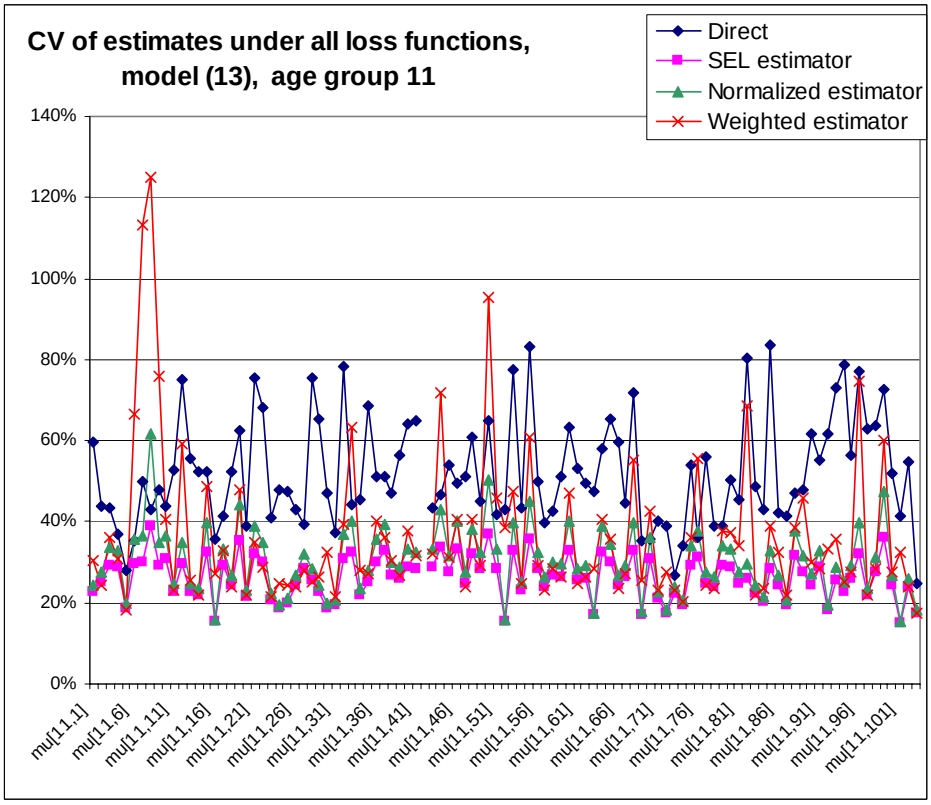

CV of estimates under all loss functions, model (13), age group 6
Direct
SEL estimator
Normalized estimator
Weighted estimator
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[6,1]
mu[6,6]
mu[6,11]
mu[6,16]
mu[6,21]
mu[6,26]
mu[6,31]
mu[6,36]
mu[6,41]
mu[6,46]
mu[6,51]
mu[6,56]
mu[6,61]
mu[6,66]
mu[6,71]
mu[6,76]
mu[6,81]
mu[6,86]
mu[6,91]
mu[6,96]
mu[6,101]

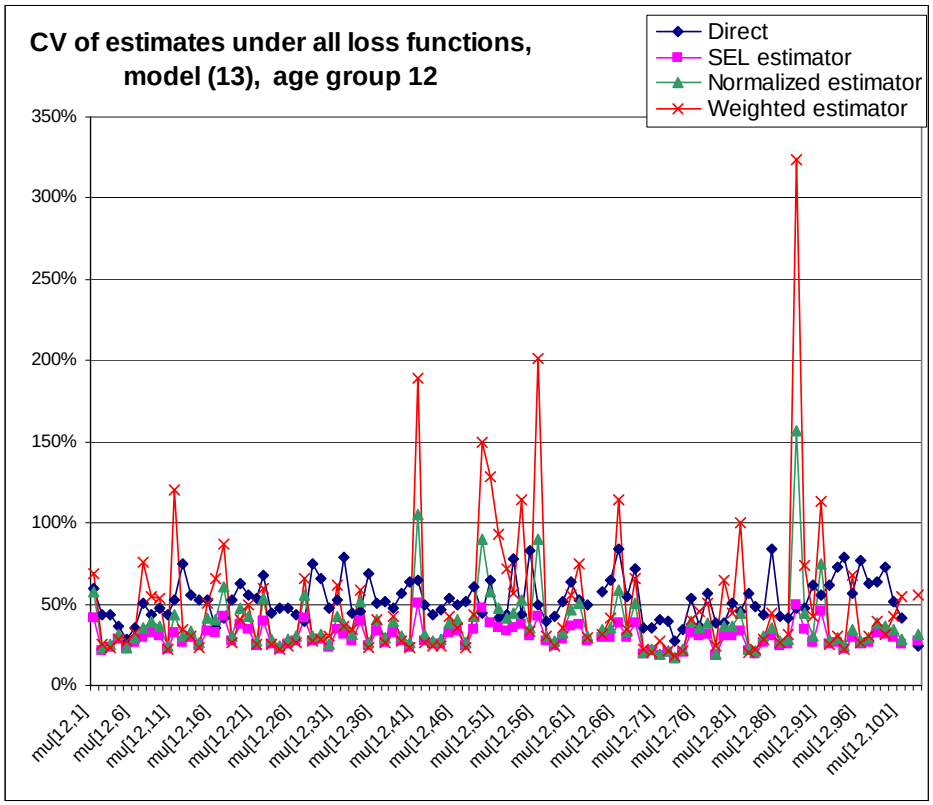

CV of estimates under all loss functions, model (13), age group 7
Direct
SEL estimator
Normalized estimator
Weighted estimator
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[7,1]
mu[7,6]
mu[7,11]
mu[7,16]
mu[7,21]
mu[7,26]
mu[7,31]
mu[7,36]
mu[7,41]
mu[7,46]
mu[7,51]
mu[7,56]
mu[7,61]
mu[7,66]
mu[7,71]
mu[7,76]
mu[7,81]
mu[7,86]
mu[7,91]
mu[7,96]
mu[7,101]

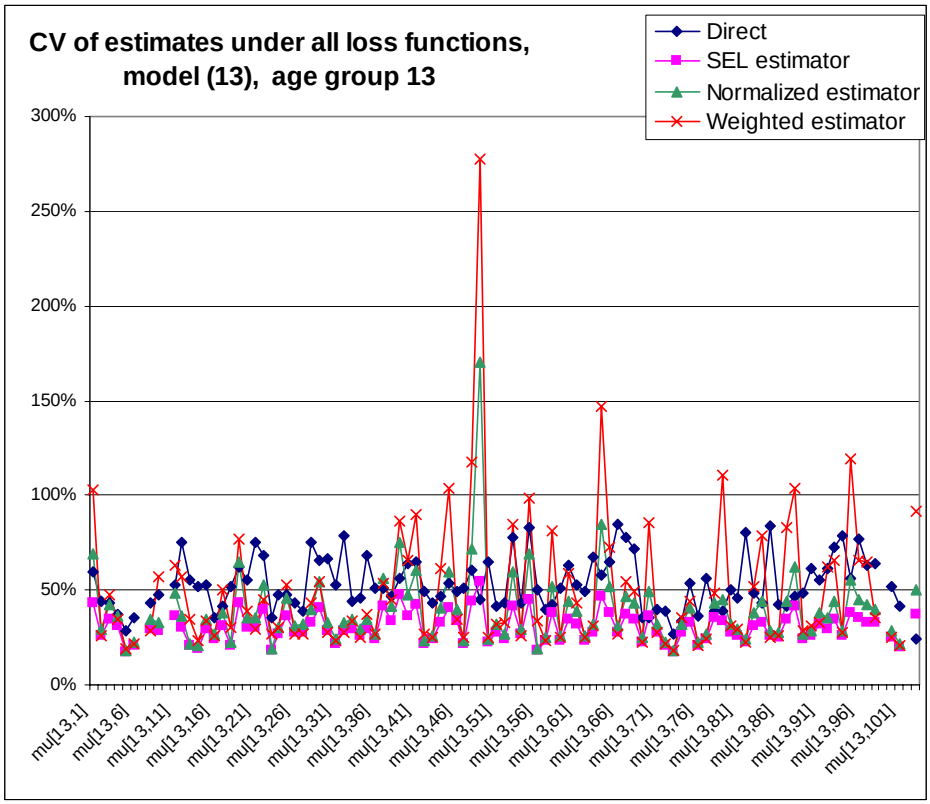

CV of estimates under all loss functions, model (13), age group 8
Direct
SEL estimator
Normalized estimator
Weighted estimator
120%
100%
80%
60%
40%
20%
0%
mu[8,1]
mu[8,6]
mu[8,11]
mu[8,16]
mu[8,21]
mu[8,26]
mu[8,31]
mu[8,36]
mu[8,41]
mu[8,46]
mu[8,51]
mu[8,56]
mu[8,61]
mu[8,66]
mu[8,71]
mu[8,76]
mu[8,81]
mu[8,86]
mu[8,91]
mu[8,96]
mu[8,101]

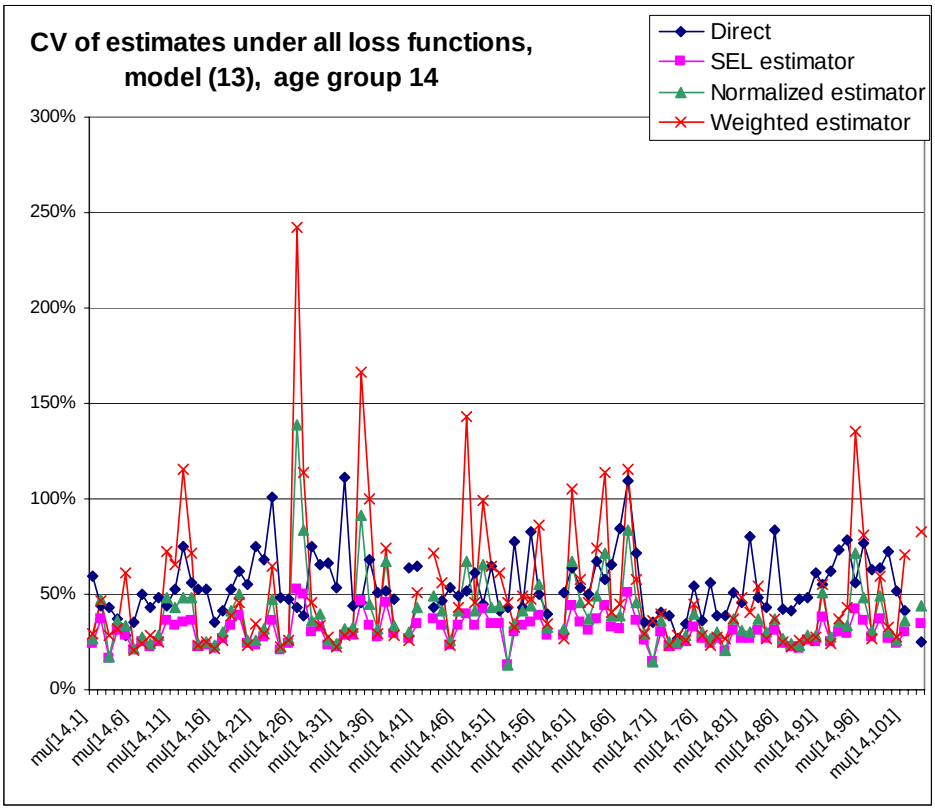

CV of estimates under all loss functions, model (13), age group 9
Direct
SEL estimator
Normalized estimator
Weighted estimator
180%
160%
140%
120%
100%
80%
60%
40%
20%
0%
mu[9,1]
mu[9,6]
mu[9,11]
mu[9,16]
mu[9,21]
mu[9,26]
mu[9,31]
mu[9,36]
mu[9,41]
mu[9,46]
mu[9,51]
mu[9,56]
mu[9,61]
mu[9,66]
mu[9,71]
mu[9,76]
mu[9,81]
mu[9,86]
mu[9,91]
mu[9,96]
mu[9,101]


CV of estimates under all loss functions, model (13), age group 10
Direct
SEL estimator
Normalized estimator
Weighted estimator
120%
100%
80%
60%
40%
20%
0%
mu[10,1]
mu[10,6]
mu[10,11]
mu[10,16]
mu[10,21]
mu[10,26]
mu[10,31]
mu[10,36]
mu[10,41]
mu[10,46]
mu[10,51]
mu[10,56]
mu[10,61]
mu[10,66]
mu[10,71]
mu[10,76]
mu[10,81]
mu[10,86]
mu[10,91]
mu[10,96]
mu[10,101]

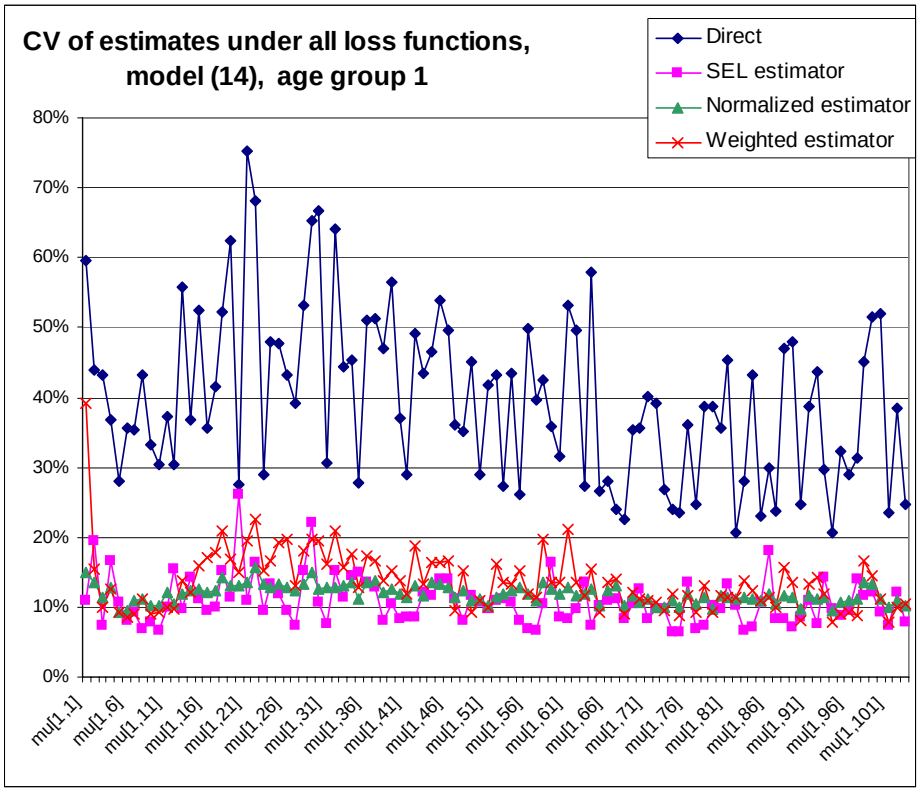

CV of estimates under all loss functions, model (13), age group 11
Direct
SEL estimator
Normalized estimator
Weighted estimator
140%
120%
100%
80%
60%
40%
20%
0%
mu[11,1]
mu[11,6]
mu[11,11]
mu[11,16]
mu[11,21]
mu[11,26]
mu[11,31]
mu[11,36]
mu[11,41]
mu[11,46]
mu[11,51]
mu[11,56]
mu[11,61]
mu[11,66]
mu[11,71]
mu[11,76]
mu[11,81]
mu[11,86]
mu[11,91]
mu[11,96]
mu[11,101]

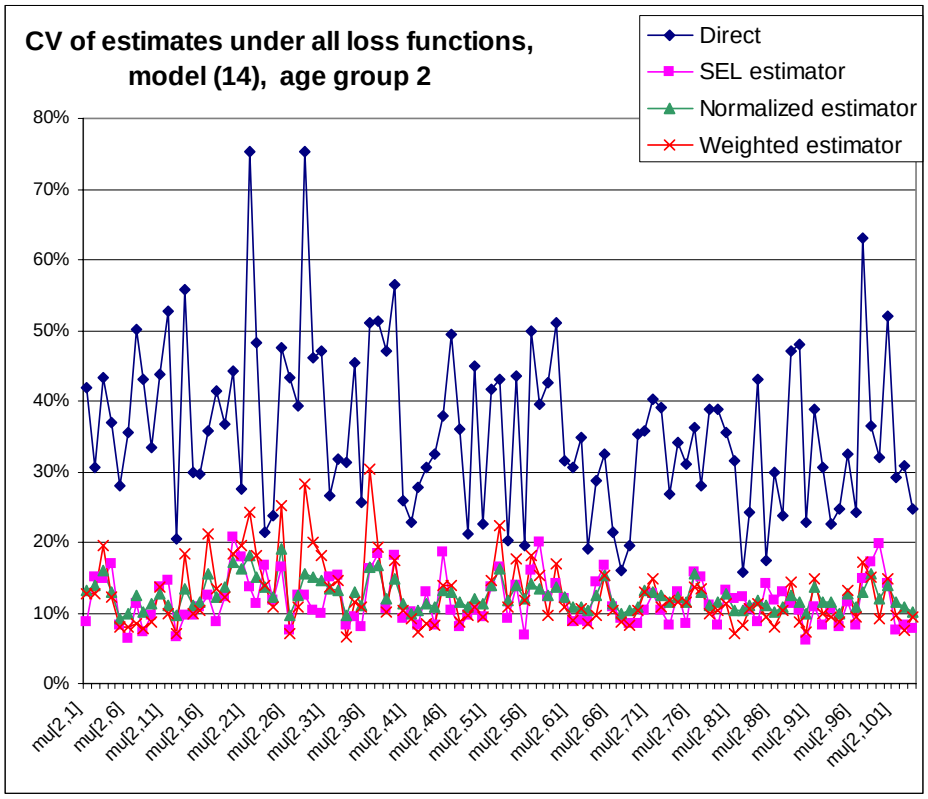

CV of estimates under all loss functions, model (13), age group 12
Direct
SEL estimator
Normalized estimator
Weighted estimator
350%
300%
250%
200%
150%
100%
50%
0%
mu[12,1]
mu[12,6]
mu[12,11]
mu[12,16]
mu[12,21]
mu[12,26]
mu[12,31]
mu[12,36]
mu[12,41]
mu[12,46]
mu[12,51]
mu[12,56]
mu[12,61]
mu[12,66]
mu[12,71]
mu[12,76]
mu[12,81]
mu[12,86]
mu[12,91]
mu[12,96]
mu[12,101]

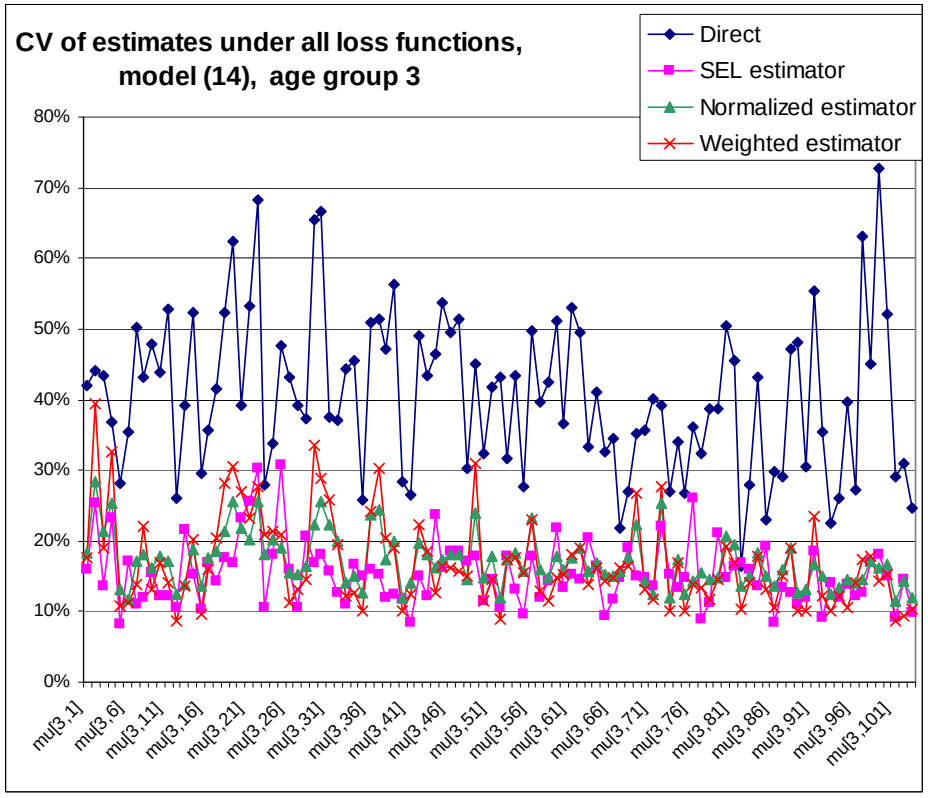

CV of estimates under all loss functions, model (13), age group 13
Direct
SEL estimator
Normalized estimator
Weighted estimator
300%
250%
200%
150%
100%
50%
0%
mu[13,1]
mu[13,6]
mu[13,11]
mu[13,16]
mu[13,21]
mu[13,26]
mu[13,31]
mu[13,36]
mu[13,41]
mu[13,46]
mu[13,51]
mu[13,56]
mu[13,61]
mu[13,66]
mu[13,71]
mu[13,76]
mu[13,81]
mu[13,86]
mu[13,91]
mu[13,96]
mu[13,101]

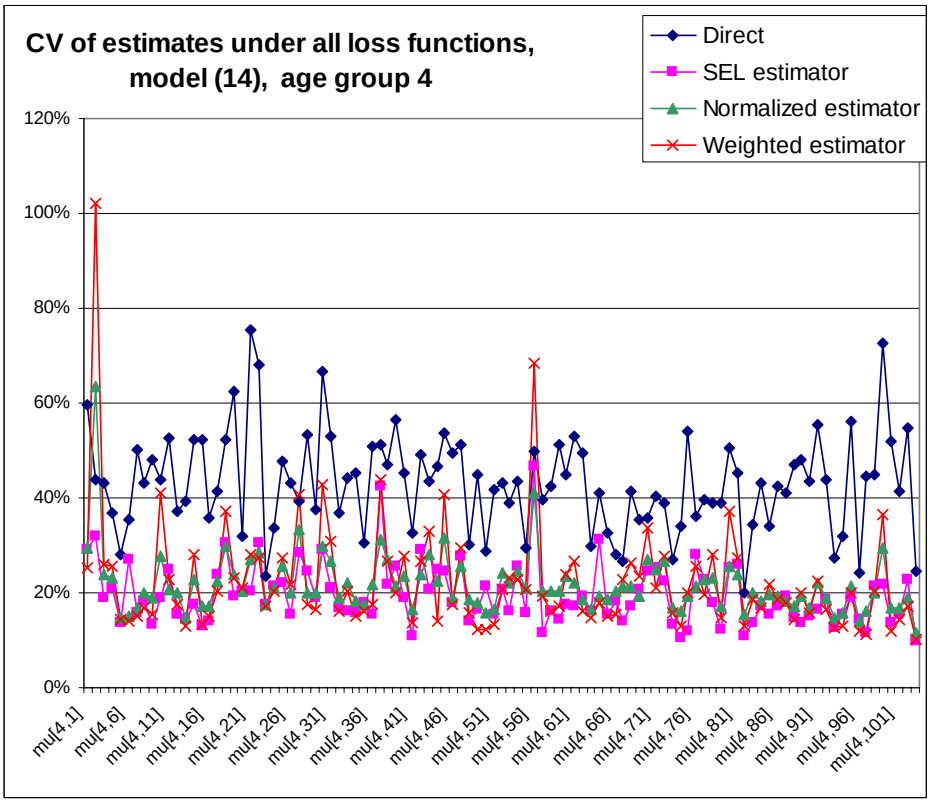

CV of estimates under all loss functions, model (13), age group 14
Direct
SEL estimator
Normalized estimator
Weighted estimator
300%
250%
200%
150%
100%
50%
0%
mu[14,1]
mu[14,6]
mu[14,11]
mu[14,16]
mu[14,21]
mu[14,26]
mu[14,31]
mu[14,36]
mu[14,41]
mu[14,46]
mu[14,51]
mu[14,56]
mu[14,61]
mu[14,66]
mu[14,71]
mu[14,76]
mu[14,81]
mu[14,86]
mu[14,91]
mu[14,96]
mu[14,101]

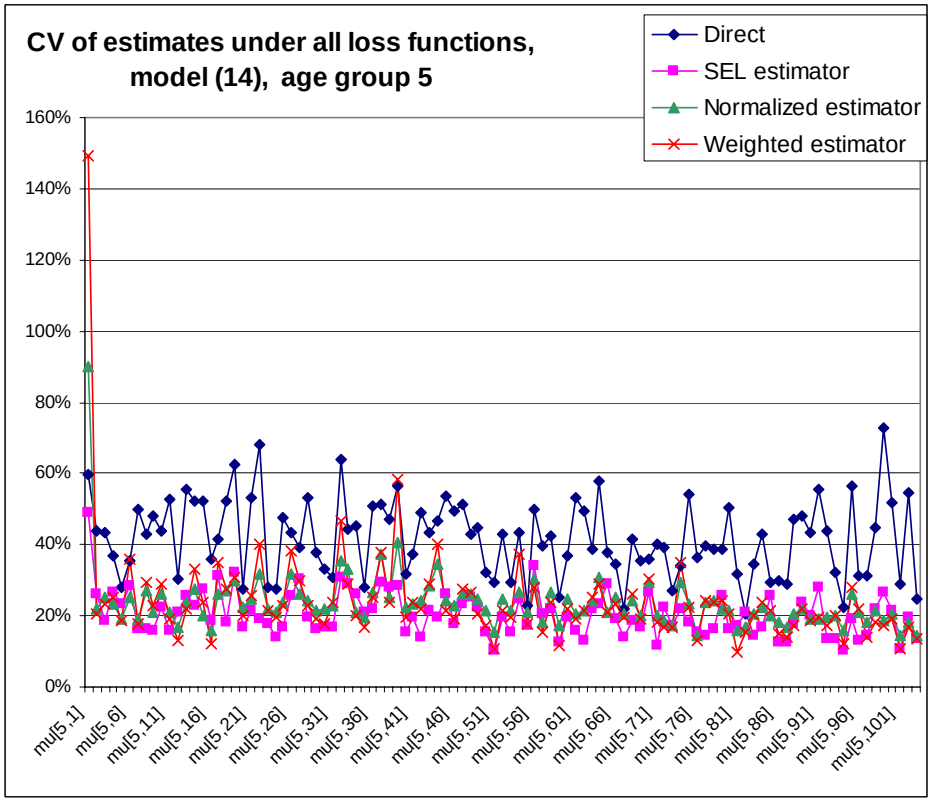
CV of model (14) estimates:

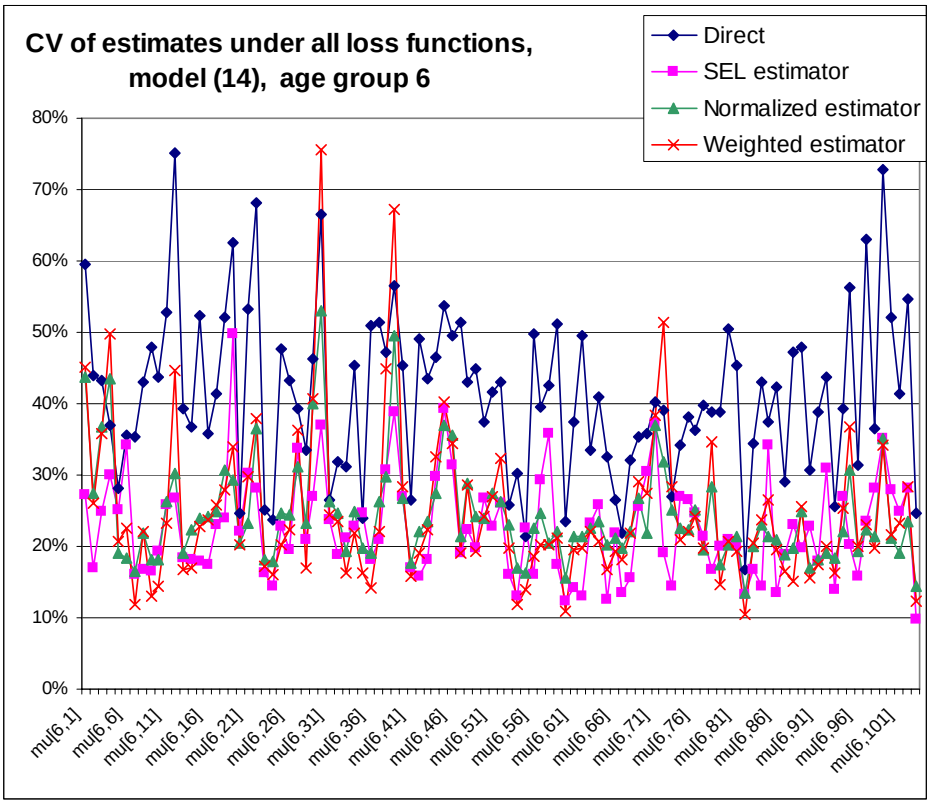


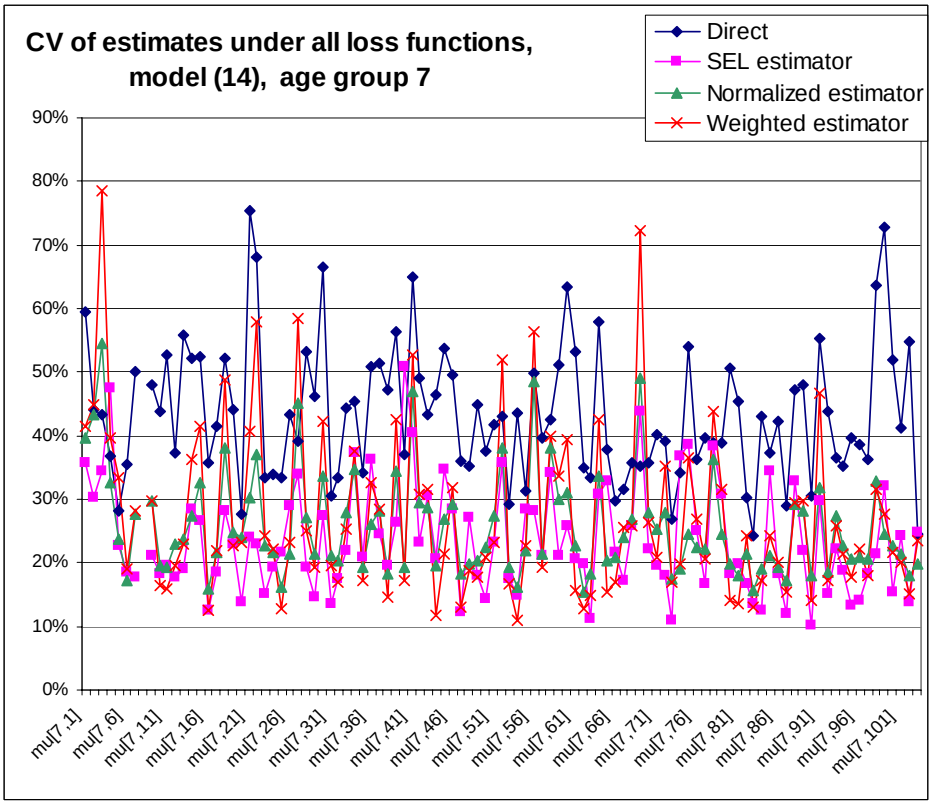

CV of estimates under all loss functions, model (14), age group 2
Direct
SEL estimator
Normalized estimator
Weighted estimator
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[2,1]
mu[2,6]
mu[2,11]
mu[2,16]
mu[2,21]
mu[2,26]
mu[2,31]
mu[2,36]
mu[2,41]
mu[2,46]
mu[2,51]
mu[2,56]
mu[2,61]
mu[2,66]
mu[2,71]
mu[2,76]
mu[2,81]
mu[2,86]
mu[2,91]
mu[2,96]
mu[2,101]

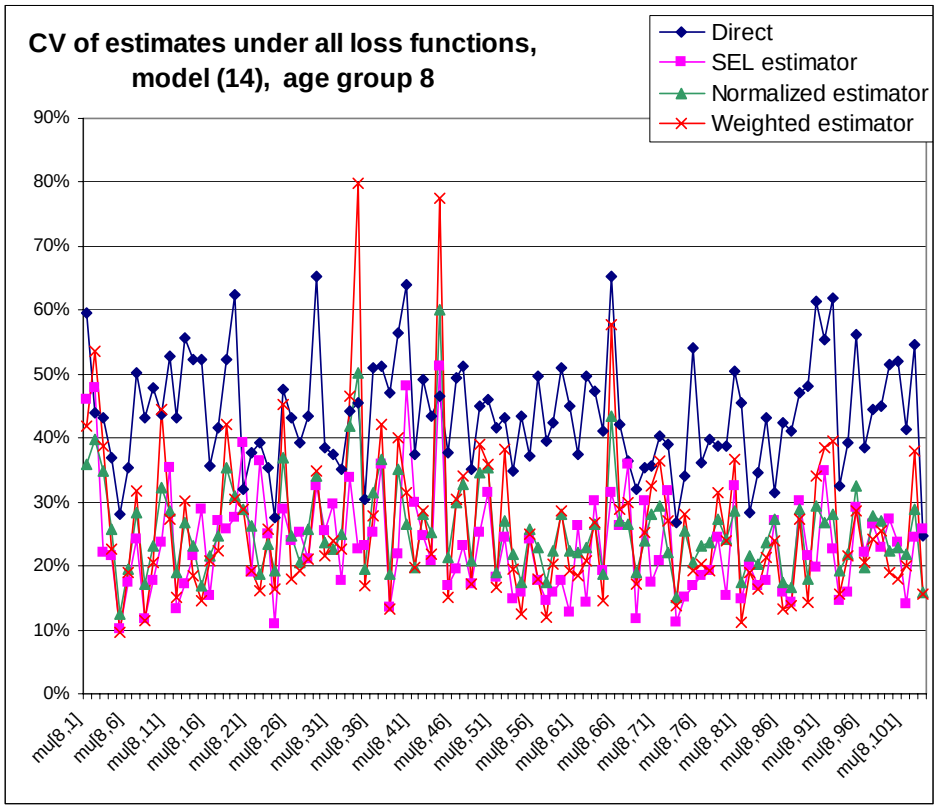

CV of estimates under all loss functions, model (14), age group 3
Direct
SEL estimator
Normalized estimator
Weighted estimator
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[3,1]
mu[3,6]
mu[3,11]
mu[3,16]
mu[3,21]
mu[3,26]
mu[3,31]
mu[3,36]
mu[3,41]
mu[3,46]
mu[3,51]
mu[3,56]
mu[3,61]
mu[3,66]
mu[3,71]
mu[3,76]
mu[3,81]
mu[3,86]
mu[3,91]
mu[3,96]
mu[3,101]

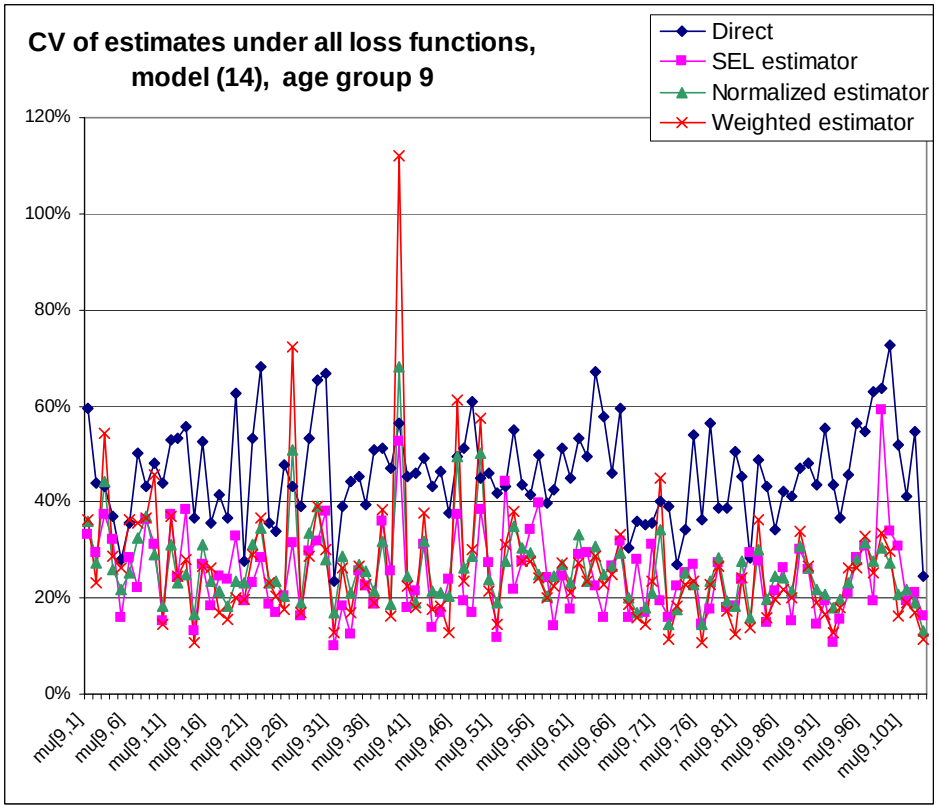

CV of estimates under all loss functions, model (14), age group 4
Direct
SEL estimator
Normalized estimator
Weighted estimator
120%
100%
80%
60%
40%
20%
0%
mu[4,1]
mu[4,6]
mu[4,11]
mu[4,16]
mu[4,21]
mu[4,26]
mu[4,31]
mu[4,36]
mu[4,41]
mu[4,46]
mu[4,51]
mu[4,56]
mu[4,61]
mu[4,66]
mu[4,71]
mu[4,76]
mu[4,81]
mu[4,86]
mu[4,91]
mu[4,96]
mu[4,101]

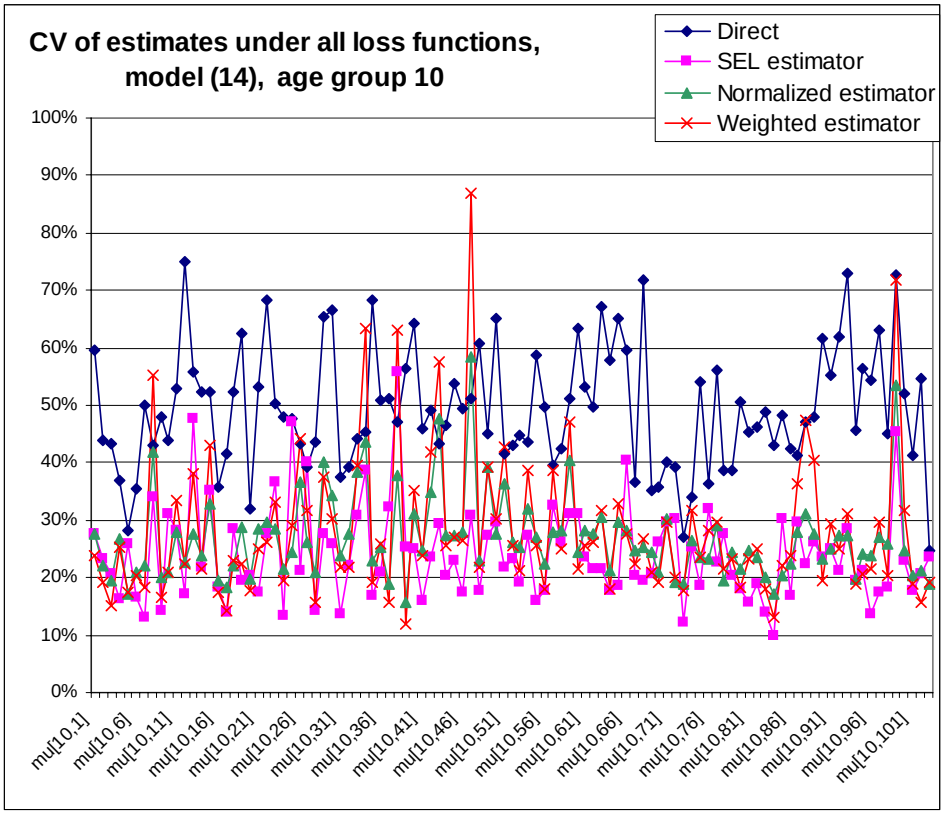

CV of estimates under all loss functions, model (14), age group 5
Direct
SEL estimator
Normalized estimator
Weighted estimator
160%
140%
120%
100%
80%
60%
40%
20%
0%
mu[5,1]
mu[5,6]
mu[5,11]
mu[5,16]
mu[5,21]
mu[5,26]
mu[5,31]
mu[5,36]
mu[5,41]
mu[5,46]
mu[5,51]
mu[5,56]
mu[5,61]
mu[5,66]
mu[5,71]
mu[5,76]
mu[5,81]
mu[5,86]
mu[5,91]
mu[5,96]
mu[5,101]

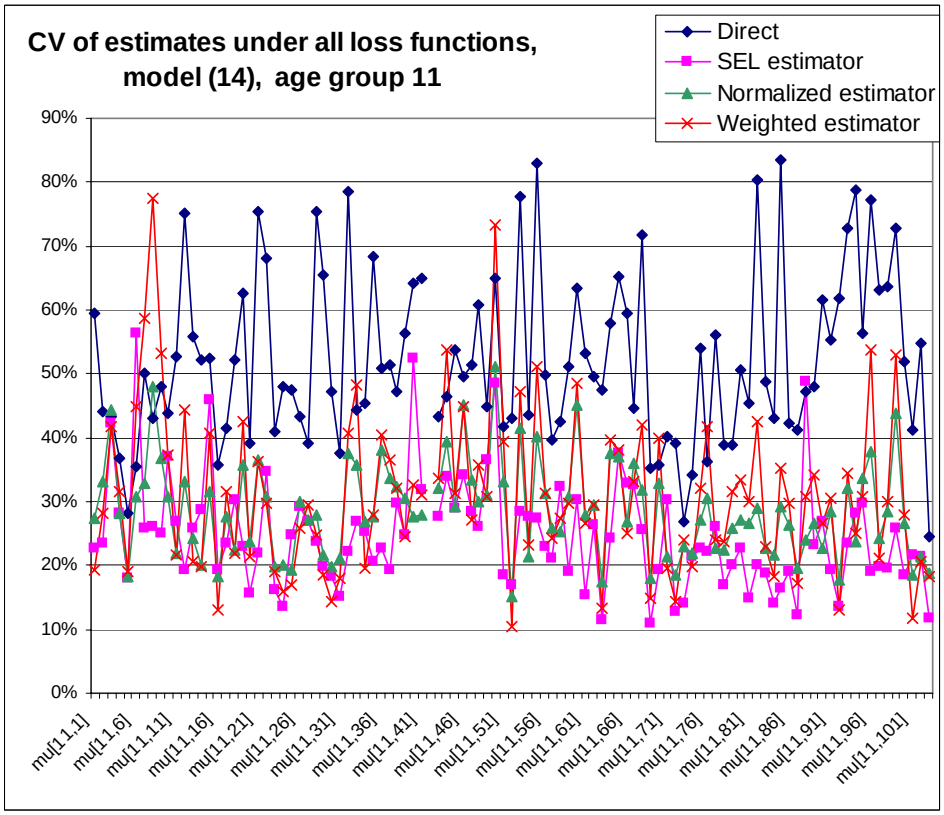

CV of estimates under all loss functions, model (14), age group 6
Direct
SEL estimator
Normalized estimator
Weighted estimator
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[6,1]
mu[6,6]
mu[6,11]
mu[6,16]
mu[6,21]
mu[6,26]
mu[6,31]
mu[6,36]
mu[6,41]
mu[6,46]
mu[6,51]
mu[6,56]
mu[6,61]
mu[6,66]
mu[6,71]
mu[6,76]
mu[6,81]
mu[6,86]
mu[6,91]
mu[6,96]
mu[6,101]

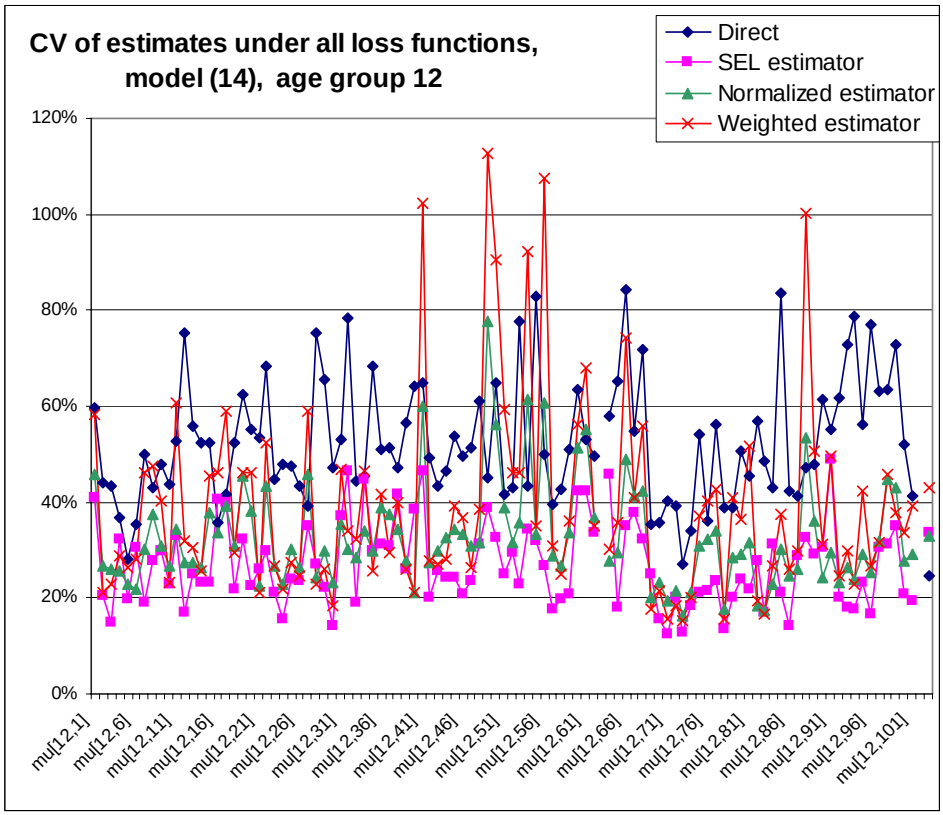

CV of estimates under all loss functions, model (14), age group 7
Direct
SEL estimator
Normalized estimator
Weighted estimator
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[7,1]
mu[7,6]
mu[7,11]
mu[7,16]
mu[7,21]
mu[7,26]
mu[7,31]
mu[7,36]
mu[7,41]
mu[7,46]
mu[7,51]
mu[7,56]
mu[7,61]
mu[7,66]
mu[7,71]
mu[7,76]
mu[7,81]
mu[7,86]
mu[7,91]
mu[7,96]
mu[7,101]

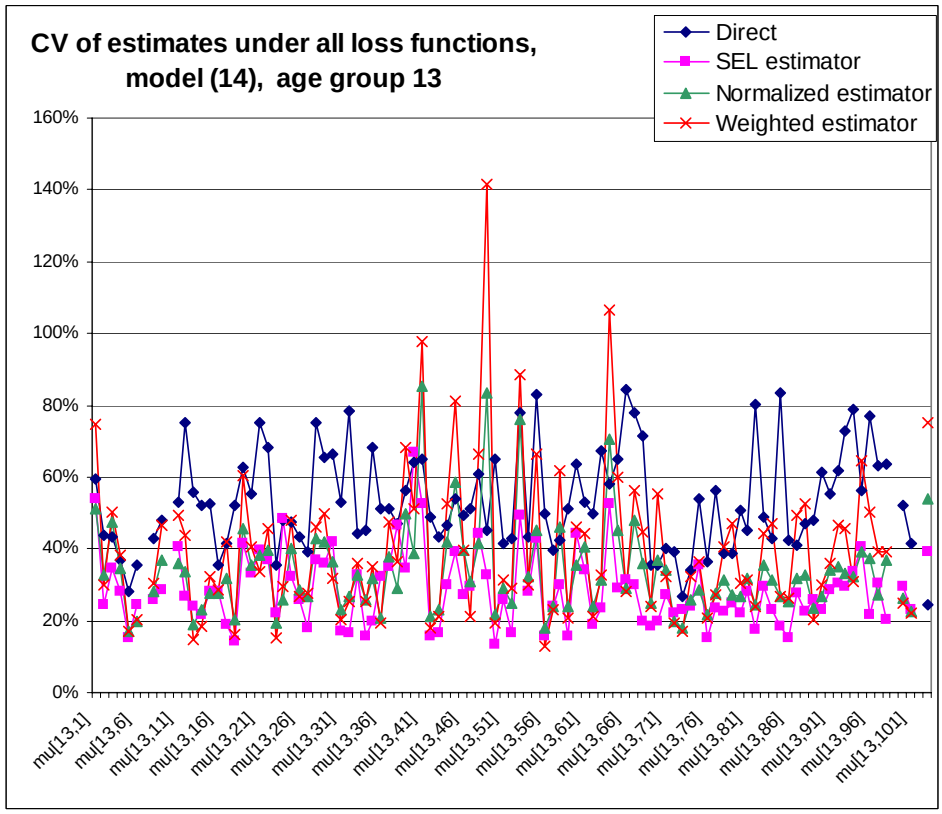

CV of estimates under all loss functions, model (14), age group 8
Direct
SEL estimator
Normalized estimator
Weighted estimator
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[8,1]
mu[8,6]
mu[8,11]
mu[8,16]
mu[8,21]
mu[8,26]
mu[8,31]
mu[8,36]
mu[8,41]
mu[8,46]
mu[8,51]
mu[8,56]
mu[8,61]
mu[8,66]
mu[8,71]
mu[8,76]
mu[8,81]
mu[8,86]
mu[8,91]
mu[8,96]
mu[8,101]

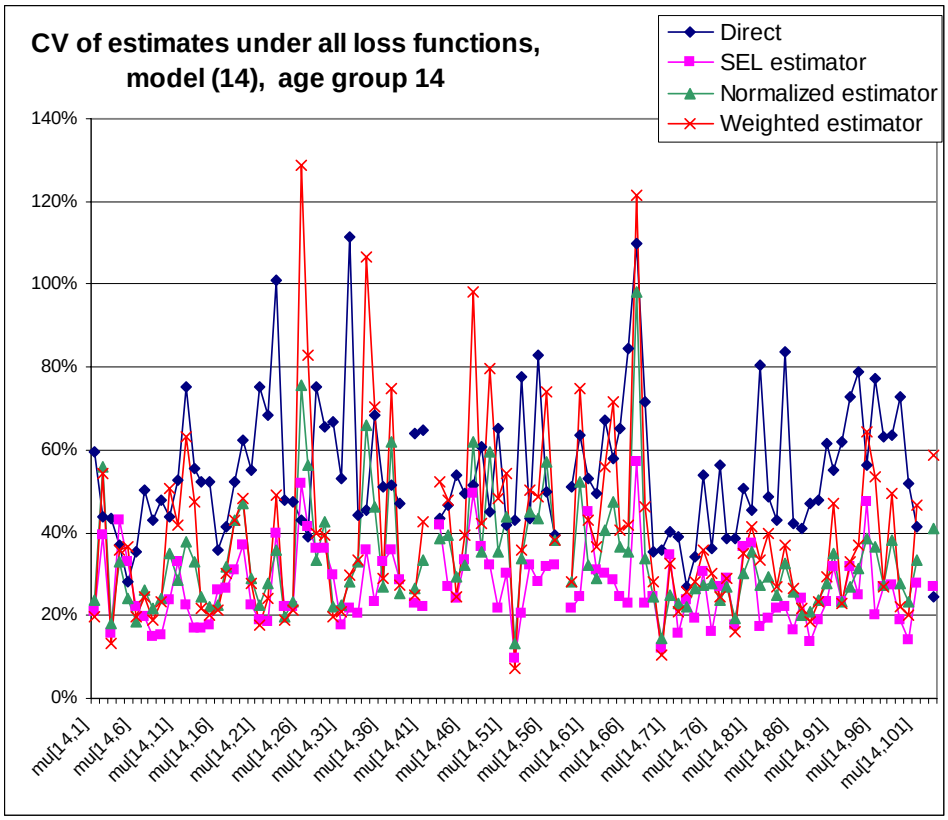

CV of estimates under all loss functions, model (14), age group 9
Direct
SEL estimator
Normalized estimator
Weighted estimator
120%
100%
80%
60%
40%
20%
0%
mu[9,1]
mu[9,6]
mu[9,11]
mu[9,16]
mu[9,21]
mu[9,26]
mu[9,31]
mu[9,36]
mu[9,41]
mu[9,46]
mu[9,51]
mu[9,56]
mu[9,61]
mu[9,66]
mu[9,71]
mu[9,76]
mu[9,81]
mu[9,86]
mu[9,91]
mu[9,96]
mu[9,101]


CV of estimates under all loss functions, model (14), age group 10
Direct
SEL estimator
Normalized estimator
Weighted estimator
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[10,1]
mu[10,6]
mu[10,11]
mu[10,16]
mu[10,21]
mu[10,26]
mu[10,31]
mu[10,36]
mu[10,41]
mu[10,46]
mu[10,51]
mu[10,56]
mu[10,61]
mu[10,66]
mu[10,71]
mu[10,76]
mu[10,81]
mu[10,86]
mu[10,91]
mu[10,96]
mu[10,101]

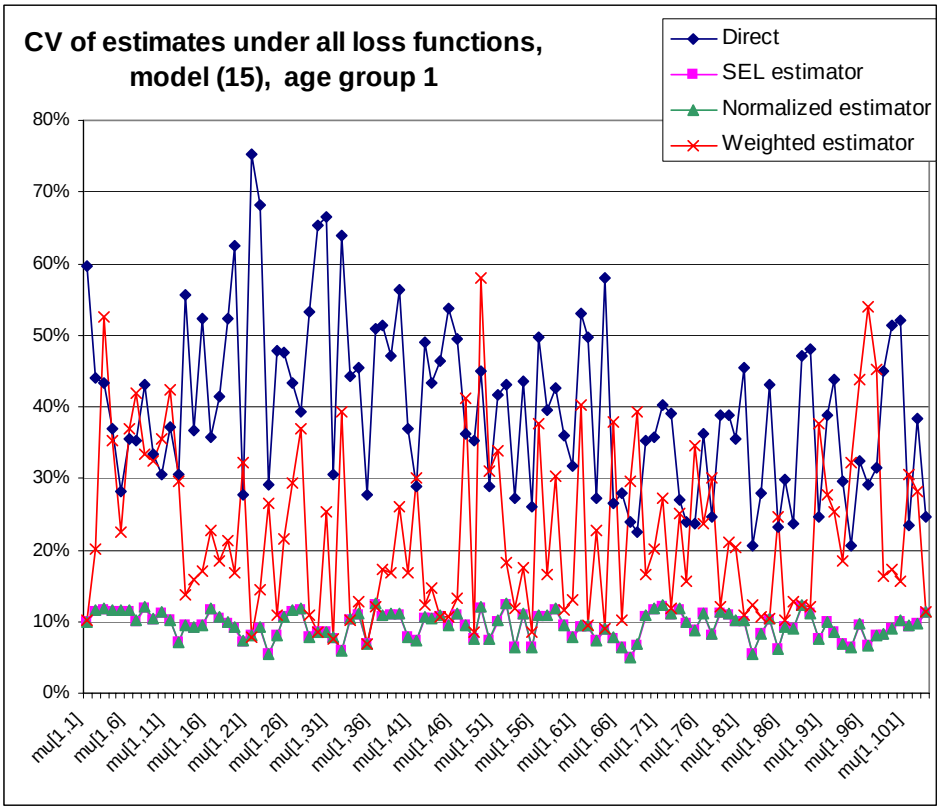

CV of estimates under all loss functions, model (14), age group 11
Direct
SEL estimator
Normalized estimator
Weighted estimator
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[11,1]
mu[11,6]
mu[11,11]
mu[11,16]
mu[11,21]
mu[11,26]
mu[11,31]
mu[11,36]
mu[11,41]
mu[11,46]
mu[11,51]
mu[11,56]
mu[11,61]
mu[11,66]
mu[11,71]
mu[11,76]
mu[11,81]
mu[11,86]
mu[11,91]
mu[11,96]
mu[11,101]

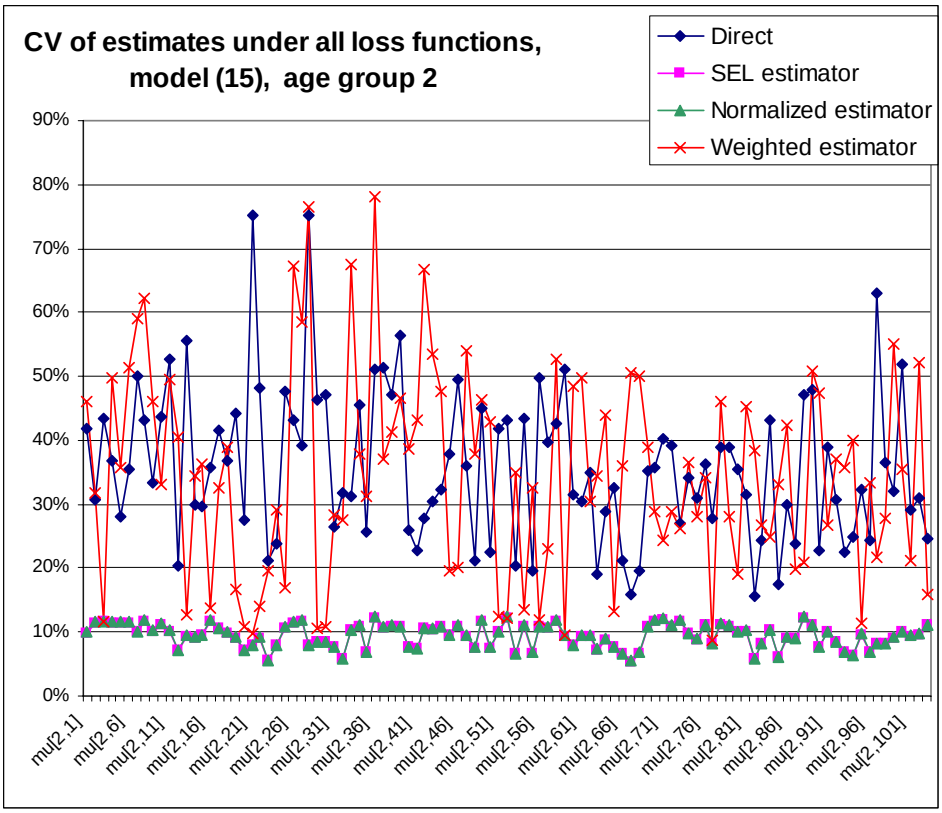

CV of estimates under all loss functions, model (14), age group 12
Direct
SEL estimator
Normalized estimator
Weighted estimator
120%
100%
80%
60%
40%
20%
0%
mu[12,1]
mu[12,6]
mu[12,11]
mu[12,16]
mu[12,21]
mu[12,26]
mu[12,31]
mu[12,36]
mu[12,41]
mu[12,46]
mu[12,51]
mu[12,56]
mu[12,61]
mu[12,66]
mu[12,71]
mu[12,76]
mu[12,81]
mu[12,86]
mu[12,91]
mu[12,96]
mu[12,101]

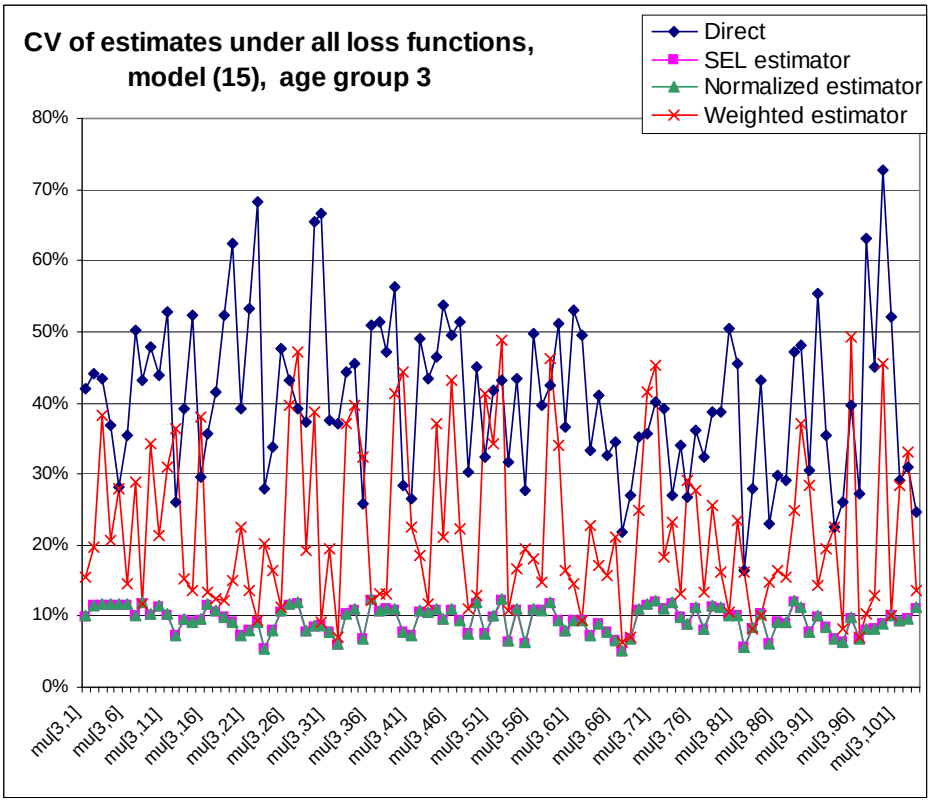

CV of estimates under all loss functions, model (14), age group 13
Direct
SEL estimator
Normalized estimator
Weighted estimator
160%
140%
120%
100%
80%
60%
40%
20%
0%
mu[13,1]
mu[13,6]
mu[13,11]
mu[13,16]
mu[13,21]
mu[13,26]
mu[13,31]
mu[13,36]
mu[13,41]
mu[13,46]
mu[13,51]
mu[13,56]
mu[13,61]
mu[13,66]
mu[13,71]
mu[13,76]
mu[13,81]
mu[13,86]
mu[13,91]
mu[13,96]
mu[13,101]

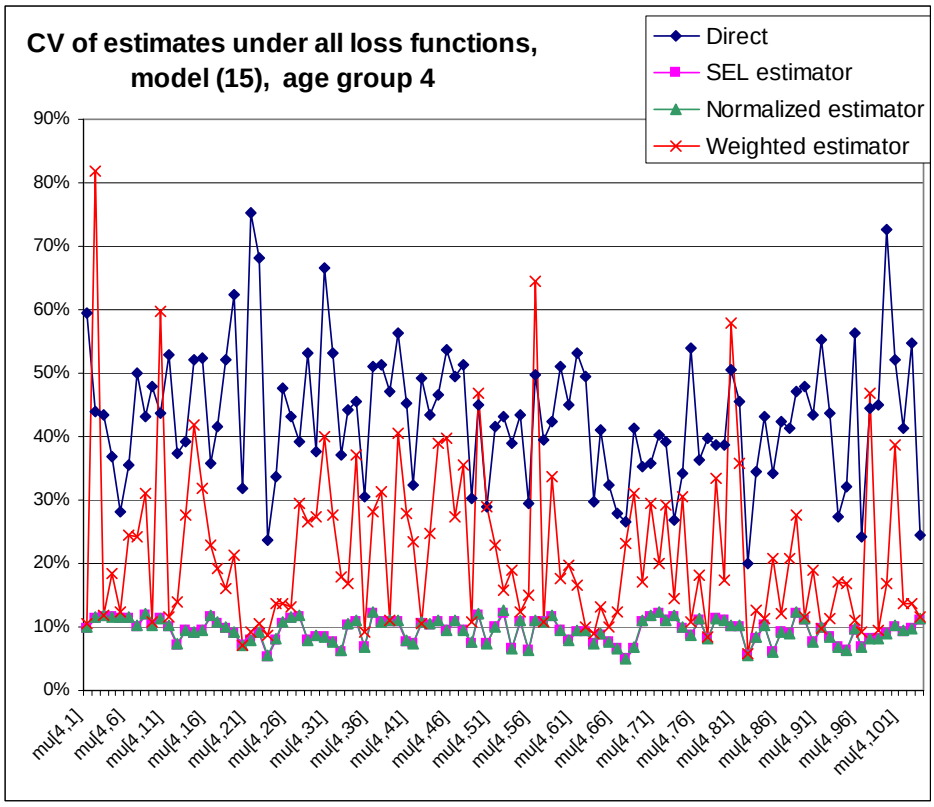

CV of estimates under all loss functions, model (14), age group 14
Direct
SEL estimator
Normalized estimator
Weighted estimator
140%
120%
100%
80%
60%
40%
20%
0%
mu[14,1]
mu[14,6]
mu[14,11]
mu[14,16]
mu[14,21]
mu[14,26]
mu[14,31]
mu[14,36]
mu[14,41]
mu[14,46]
mu[14,51]
mu[14,56]
mu[14,61]
mu[14,66]
mu[14,71]
mu[14,76]
mu[14,81]
mu[14,86]
mu[14,91]
mu[14,96]
mu[14,101]

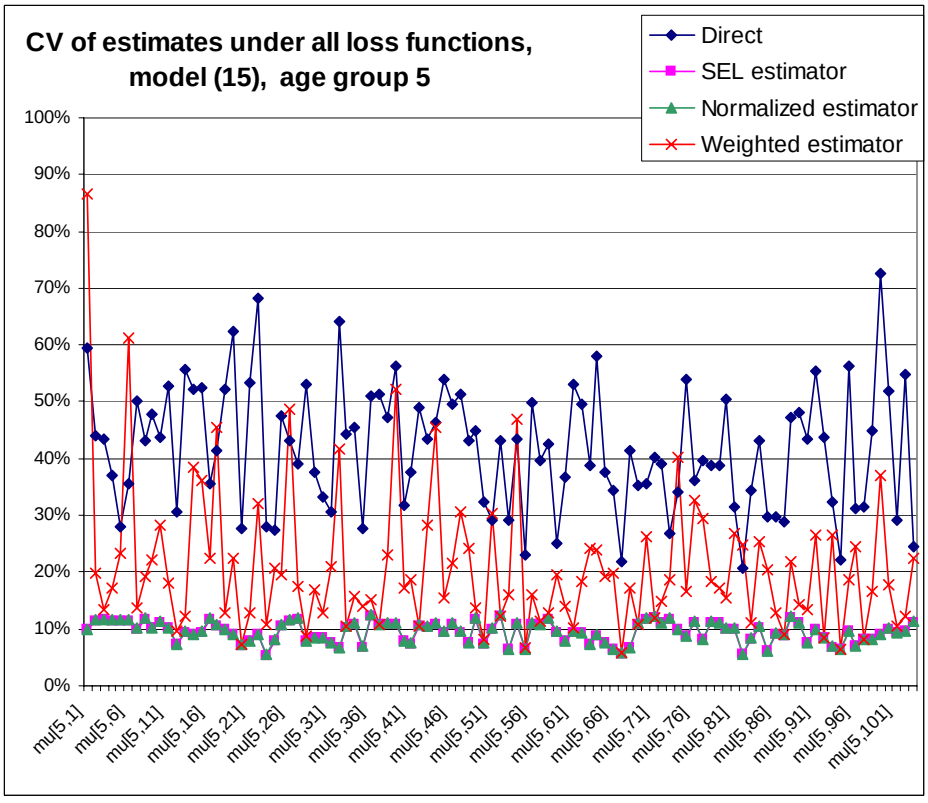
CV of model (15) estimates:

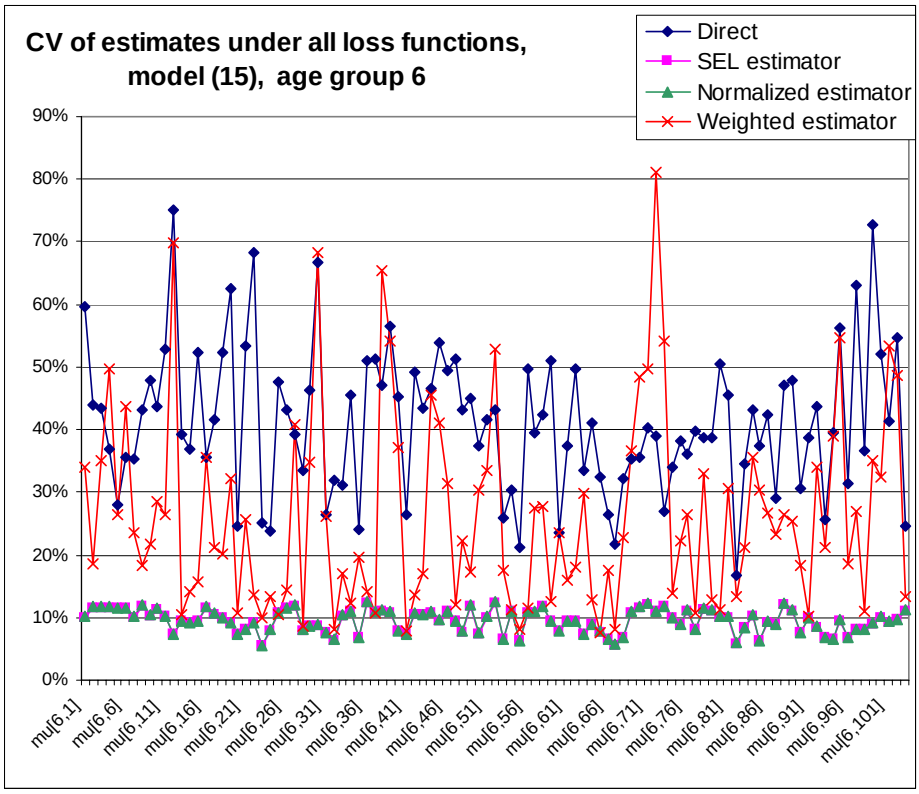


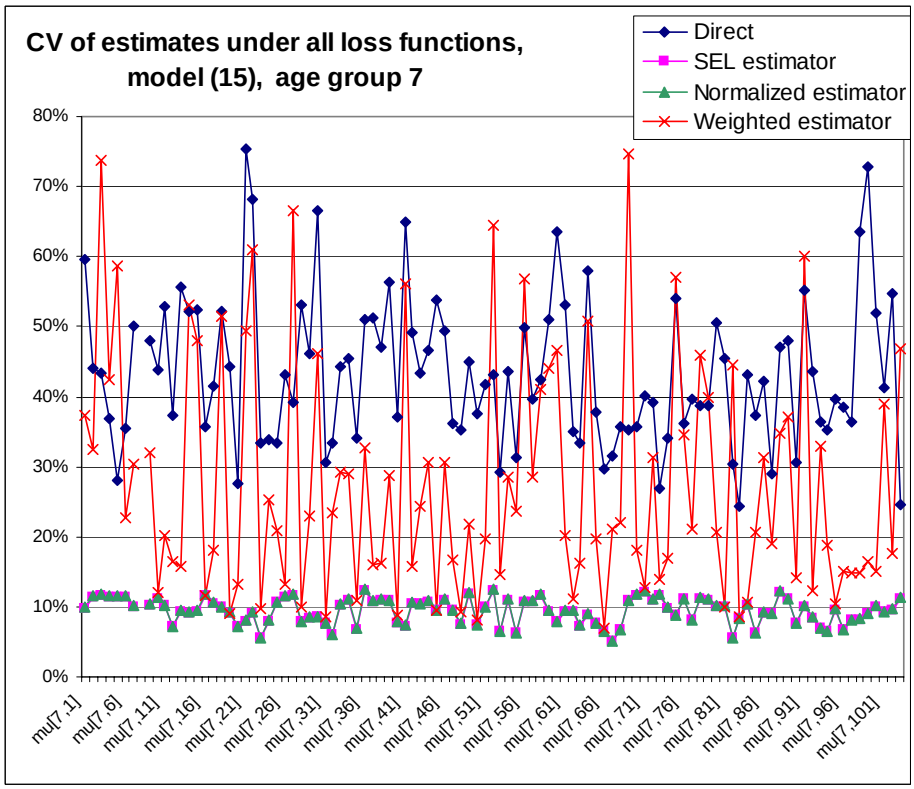

CV of estimates under all loss functions, model (15), age group 2
Direct
SEL estimator
Normalized estimator
Weighted estimator
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[2,1]
mu[2,6]
mu[2,11]
mu[2,16]
mu[2,21]
mu[2,26]
mu[2,31]
mu[2,36]
mu[2,41]
mu[2,46]
mu[2,51]
mu[2,56]
mu[2,61]
mu[2,66]
mu[2,71]
mu[2,76]
mu[2,81]
mu[2,86]
mu[2,91]
mu[2,96]
mu[2,101]

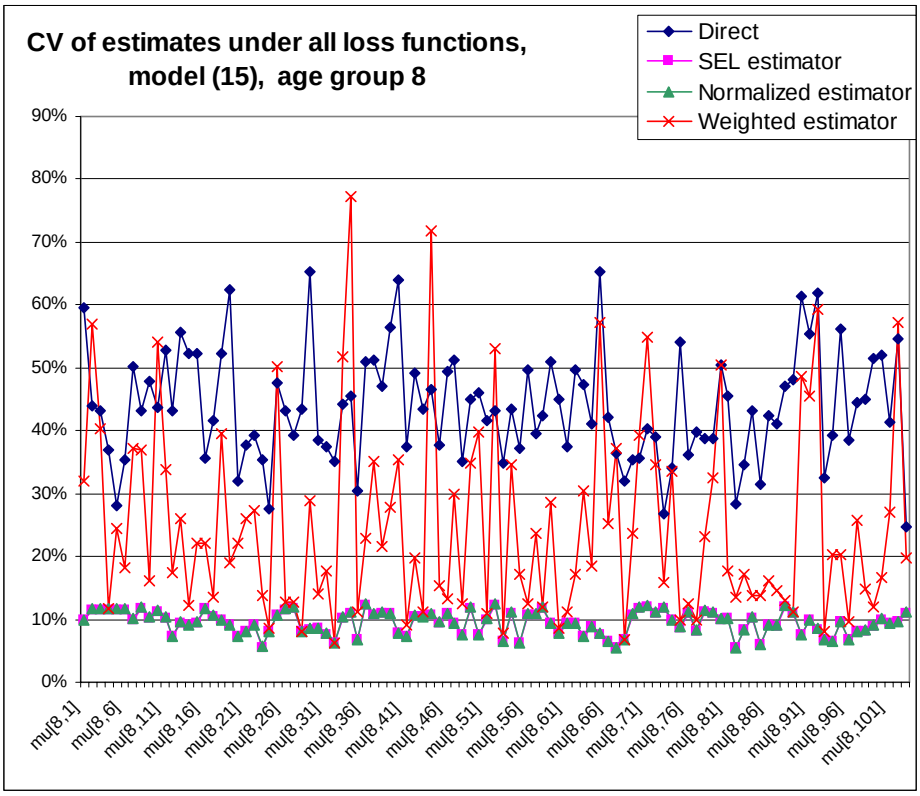

CV of estimates under all loss functions, model (15), age group 3
Direct
SEL estimator
Normalized estimator
Weighted estimator
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[3,1]
mu[3,6]
mu[3,11]
mu[3,16]
mu[3,21]
mu[3,26]
mu[3,31]
mu[3,36]
mu[3,41]
mu[3,46]
mu[3,51]
mu[3,56]
mu[3,61]
mu[3,66]
mu[3,71]
mu[3,76]
mu[3,81]
mu[3,86]
mu[3,91]
mu[3,96]
mu[3,101]

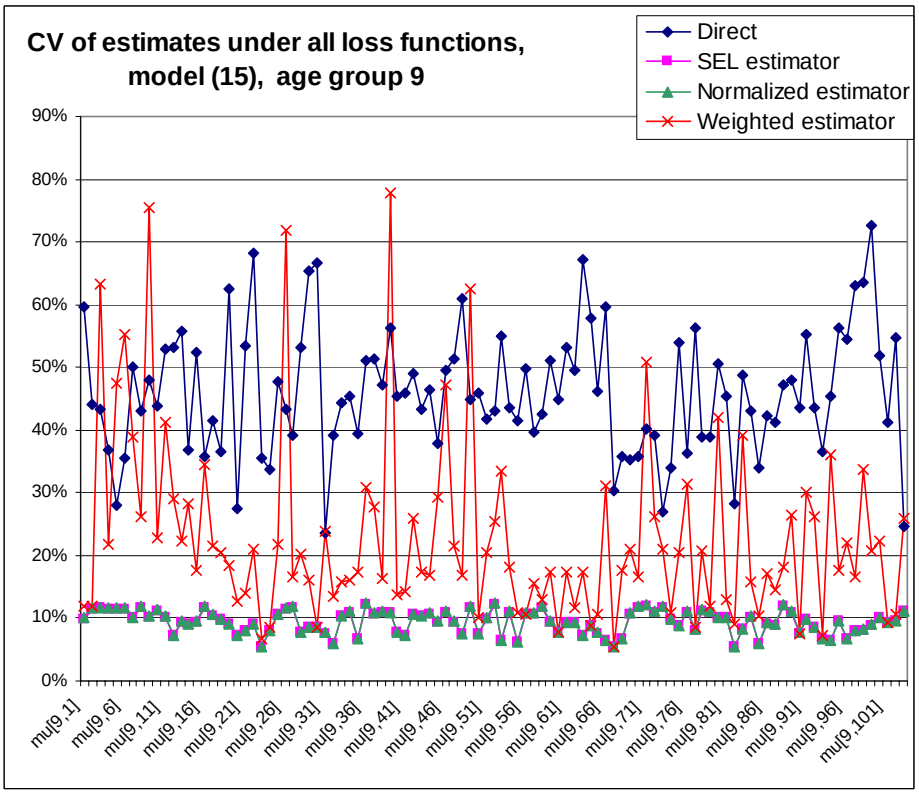

CV of estimates under all loss functions, model (15), age group 4
Direct
SEL estimator
Normalized estimator
Weighted estimator
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[4,1]
mu[4,6]
mu[4,11]
mu[4,16]
mu[4,21]
mu[4,26]
mu[4,31]
mu[4,36]
mu[4,41]
mu[4,46]
mu[4,51]
mu[4,56]
mu[4,61]
mu[4,66]
mu[4,71]
mu[4,76]
mu[4,81]
mu[4,86]
mu[4,91]
mu[4,96]
mu[4,101]

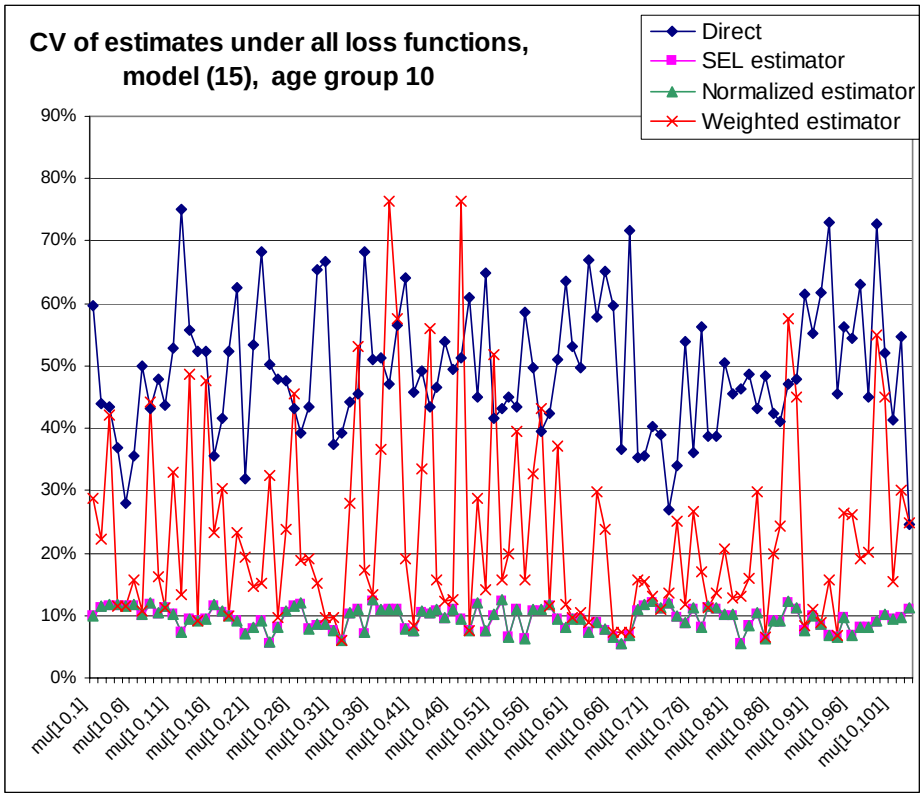

CV of estimates under all loss functions, model (15), age group 5
Direct
SEL estimator
Normalized estimator
Weighted estimator
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[5,1]
mu[5,6]
mu[5,11]
mu[5,16]
mu[5,21]
mu[5,26]
mu[5,31]
mu[5,36]
mu[5,41]
mu[5,46]
mu[5,51]
mu[5,56]
mu[5,61]
mu[5,66]
mu[5,71]
mu[5,76]
mu[5,81]
mu[5,86]
mu[5,91]
mu[5,96]
mu[5,101]

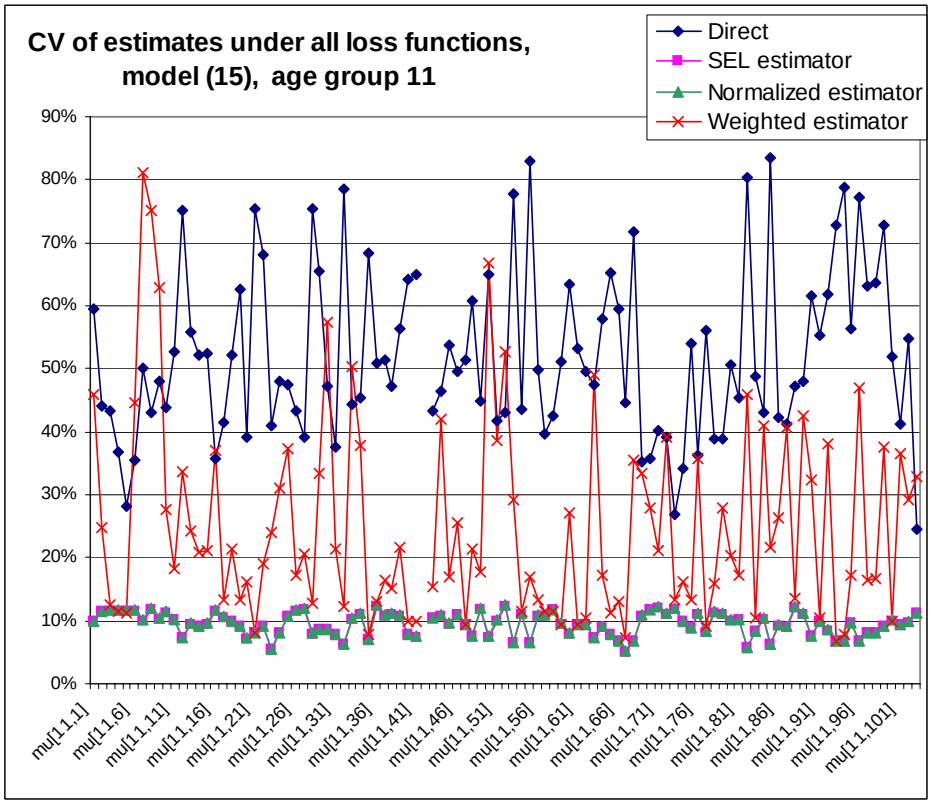

CV of estimates under all loss functions, model (15), age group 6
Direct
SEL estimator
Normalized estimator
Weighted estimator
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[6,1]
mu[6,6]
mu[6,11]
mu[6,16]
mu[6,21]
mu[6,26]
mu[6,31]
mu[6,36]
mu[6,41]
mu[6,46]
mu[6,51]
mu[6,56]
mu[6,61]
mu[6,66]
mu[6,71]
mu[6,76]
mu[6,81]
mu[6,86]
mu[6,91]
mu[6,96]
mu[6,101]

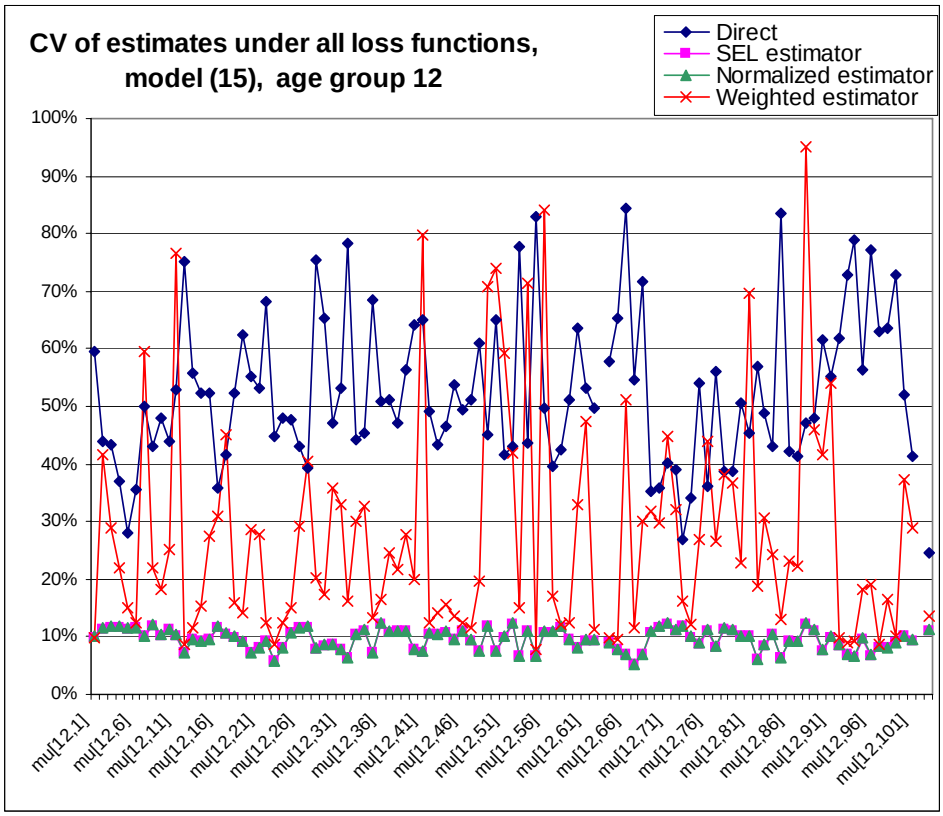

CV of estimates under all loss functions, model (15), age group 7
Direct
SEL estimator
Normalized estimator
Weighted estimator
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[7,1]
mu[7,6]
mu[7,11]
mu[7,16]
mu[7,21]
mu[7,26]
mu[7,31]
mu[7,36]
mu[7,41]
mu[7,46]
mu[7,51]
mu[7,56]
mu[7,61]
mu[7,66]
mu[7,71]
mu[7,76]
mu[7,81]
mu[7,86]
mu[7,91]
mu[7,96]
mu[7,101]

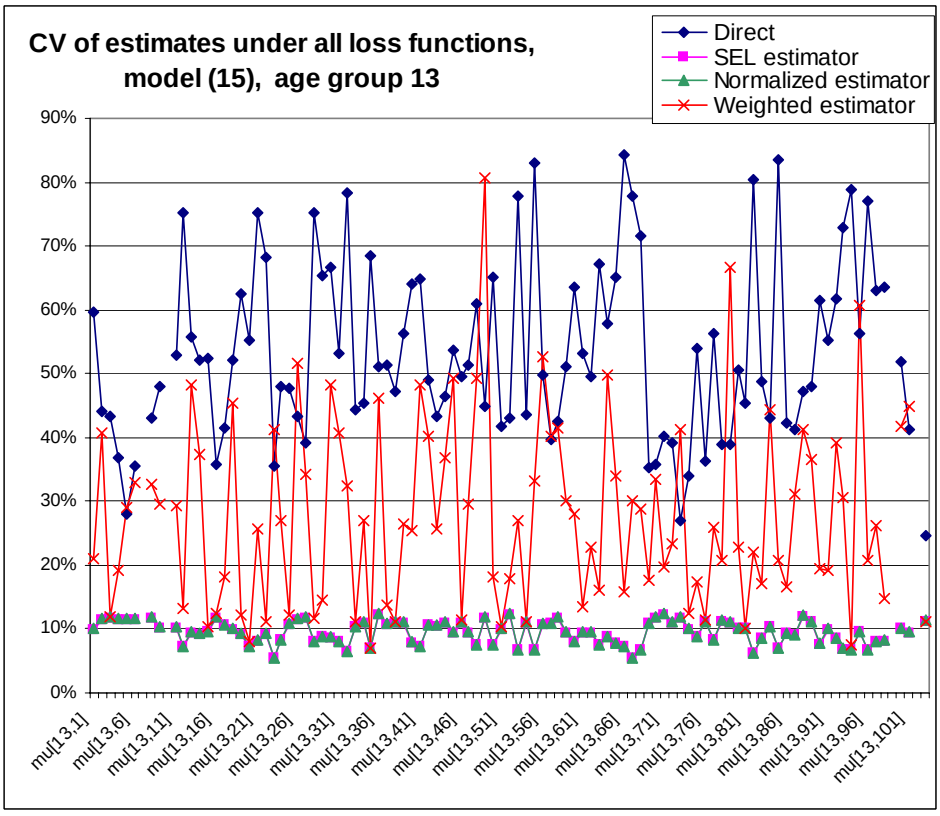

CV of estimates under all loss functions, model (15), age group 8
Direct
SEL estimator
Normalized estimator
Weighted estimator
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[8,1]
mu[8,6]
mu[8,11]
mu[8,16]
mu[8,21]
mu[8,26]
mu[8,31]
mu[8,36]
mu[8,41]
mu[8,46]
mu[8,51]
mu[8,56]
mu[8,61]
mu[8,66]
mu[8,71]
mu[8,76]
mu[8,81]
mu[8,86]
mu[8,91]
mu[8,96]
mu[8,101]

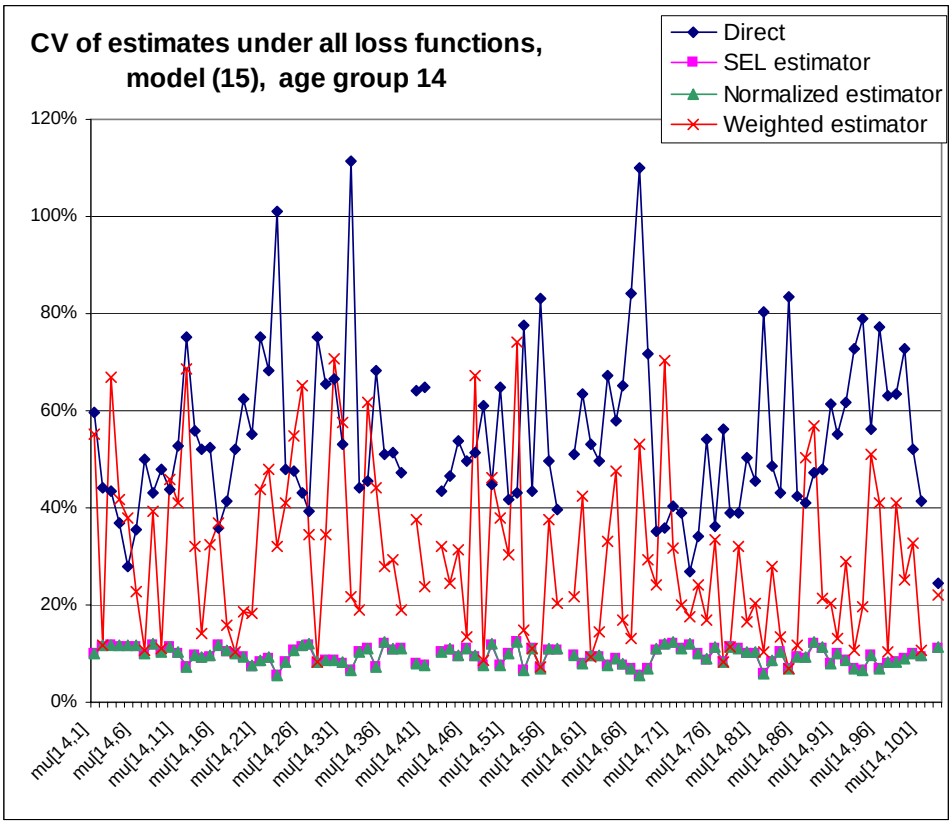

CV of estimates under all loss functions, model (15), age group 9
Direct
SEL estimator
Normalized estimator
Weighted estimator
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[9,1]
mu[9,6]
mu[9,11]
mu[9,16]
mu[9,21]
mu[9,26]
mu[9,31]
mu[9,36]
mu[9,41]
mu[9,46]
mu[9,51]
mu[9,56]
mu[9,61]
mu[9,66]
mu[9,71]
mu[9,76]
mu[9,81]
mu[9,86]
mu[9,91]
mu[9,96]
mu[9,101]


CV of estimates under all loss functions, model (15), age group 10
Direct
SEL estimator
Normalized estimator
Weighted estimator
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[10,1]
mu[10,6]
mu[10,11]
mu[10,16]
mu[10,21]
mu[10,26]
mu[10,31]
mu[10,36]
mu[10,41]
mu[10,46]
mu[10,51]
mu[10,56]
mu[10,61]
mu[10,66]
mu[10,71]
mu[10,76]
mu[10,81]
mu[10,86]
mu[10,91]
mu[10,96]
mu[10,101]

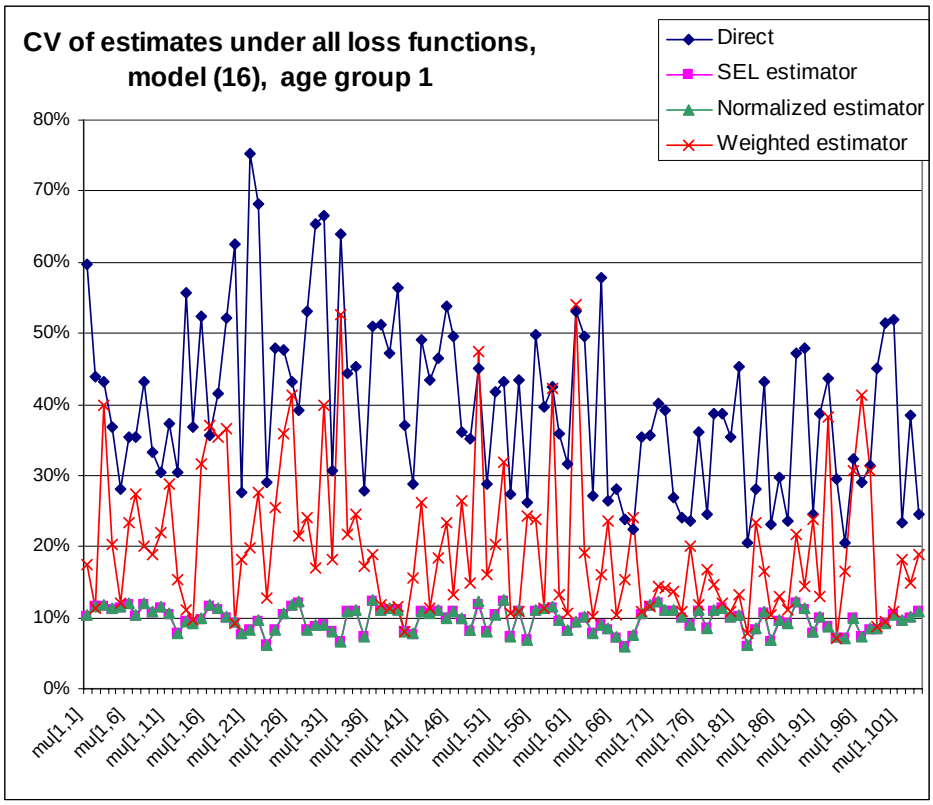

CV of estimates under all loss functions, model (15), age group 11
Direct
SEL estimator
Normalized estimator
Weighted estimator
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[11,1]
mu[11,6]
mu[11,11]
mu[11,16]
mu[11,21]
mu[11,26]
mu[11,31]
mu[11,36]
mu[11,41]
mu[11,46]
mu[11,51]
mu[11,56]
mu[11,61]
mu[11,66]
mu[11,71]
mu[11,76]
mu[11,81]
mu[11,86]
mu[11,91]
mu[11,96]
mu[11,101]

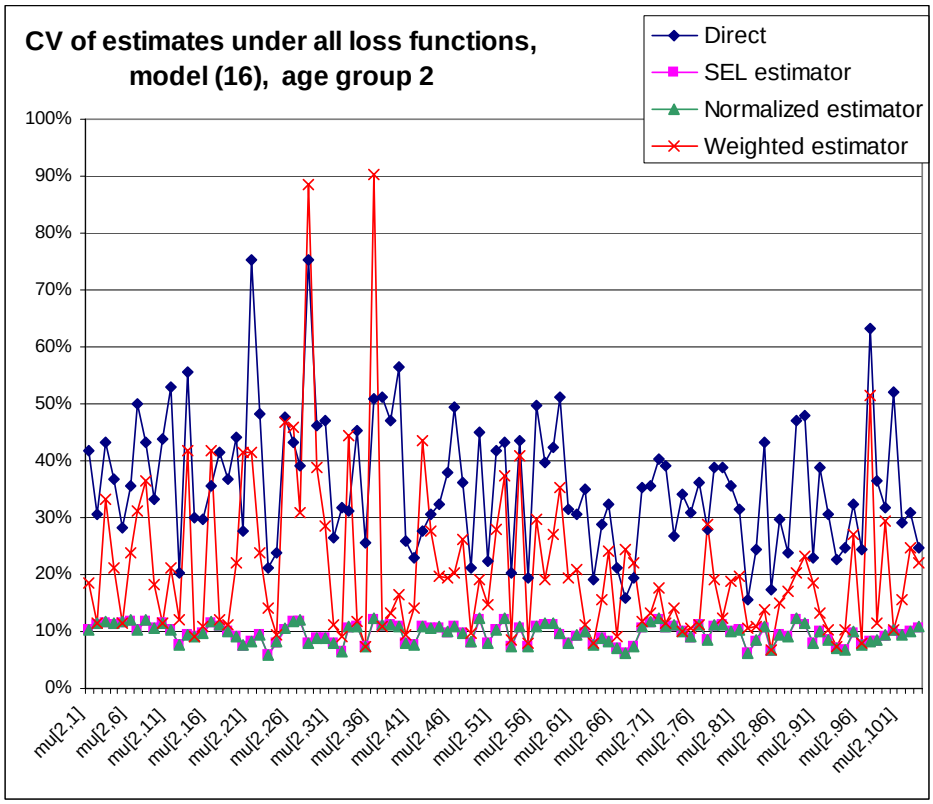

CV of estimates under all loss functions, model (15), age group 12
Direct
SEL estimator
Normalized estimator
Weighted estimator
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[12,1]
mu[12,6]
mu[12,11]
mu[12,16]
mu[12,21]
mu[12,26]
mu[12,31]
mu[12,36]
mu[12,41]
mu[12,46]
mu[12,51]
mu[12,56]
mu[12,61]
mu[12,66]
mu[12,71]
mu[12,76]
mu[12,81]
mu[12,86]
mu[12,91]
mu[12,96]
mu[12,101]

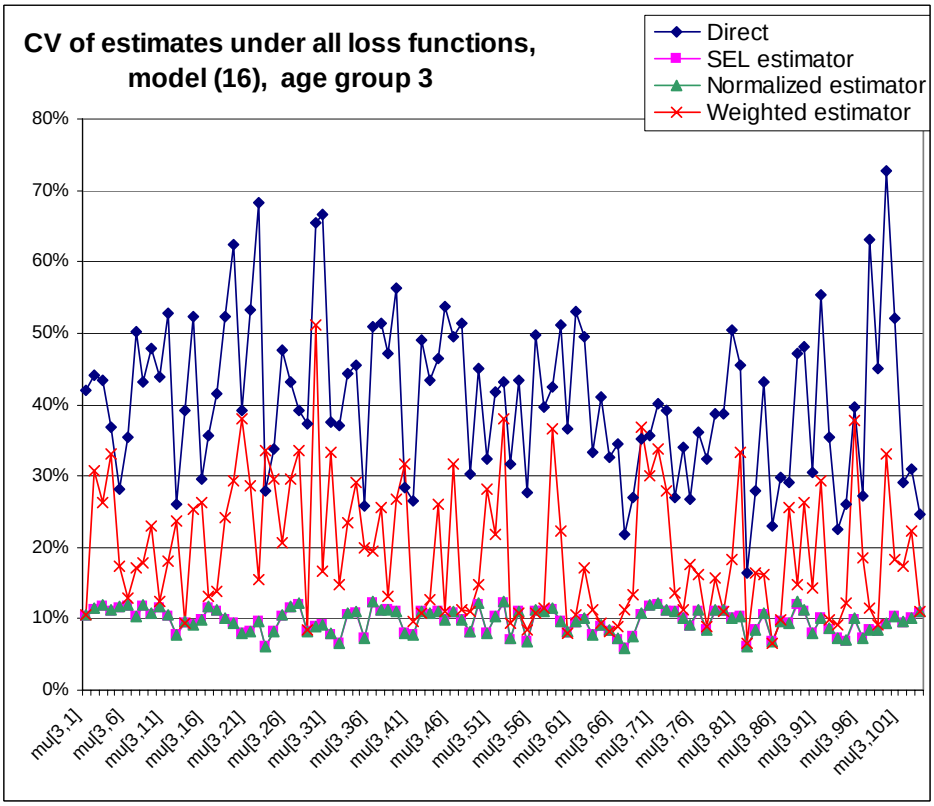

CV of estimates under all loss functions, model (15), age group 13
Direct
SEL estimator
Normalized estimator
Weighted estimator
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[13,1]
mu[13,6]
mu[13,11]
mu[13,16]
mu[13,21]
mu[13,26]
mu[13,31]
mu[13,36]
mu[13,41]
mu[13,46]
mu[13,51]
mu[13,56]
mu[13,61]
mu[13,66]
mu[13,71]
mu[13,76]
mu[13,81]
mu[13,86]
mu[13,91]
mu[13,96]
mu[13,101]

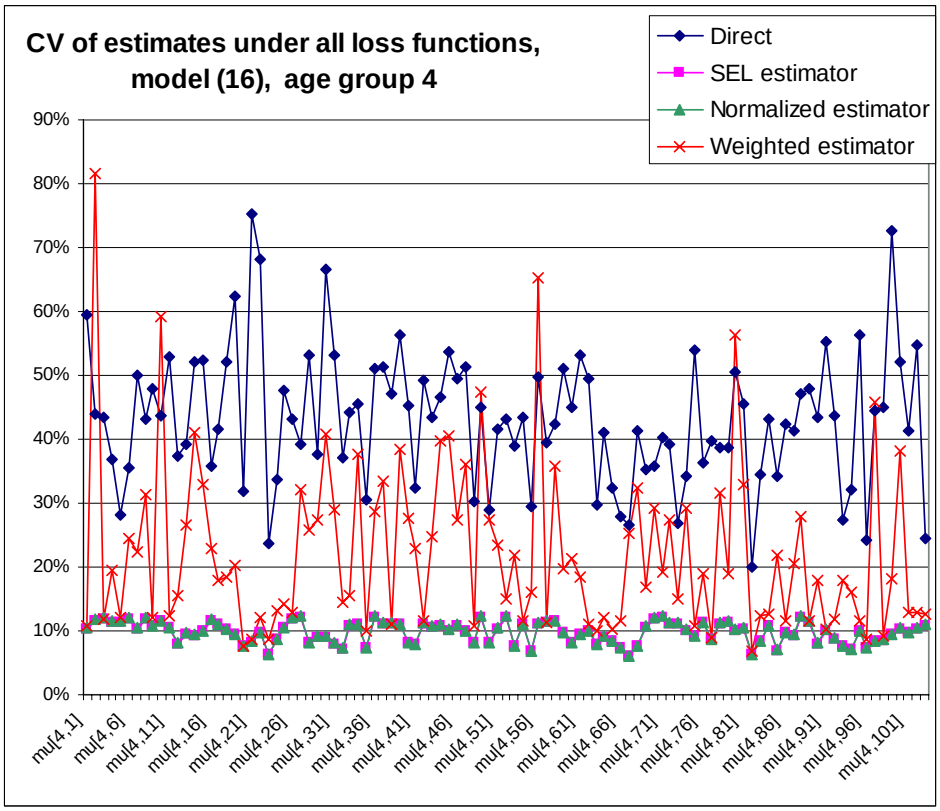

CV of estimates under all loss functions, model (15), age group 14
Direct
SEL estimator
Normalized estimator
Weighted estimator
120%
100%
80%
60%
40%
20%
0%
mu[14,1]
mu[14,6]
mu[14,11]
mu[14,16]
mu[14,21]
mu[14,26]
mu[14,31]
mu[14,36]
mu[14,41]
mu[14,46]
mu[14,51]
mu[14,56]
mu[14,61]
mu[14,66]
mu[14,71]
mu[14,76]
mu[14,81]
mu[14,86]
mu[14,91]
mu[14,96]
mu[14,101]

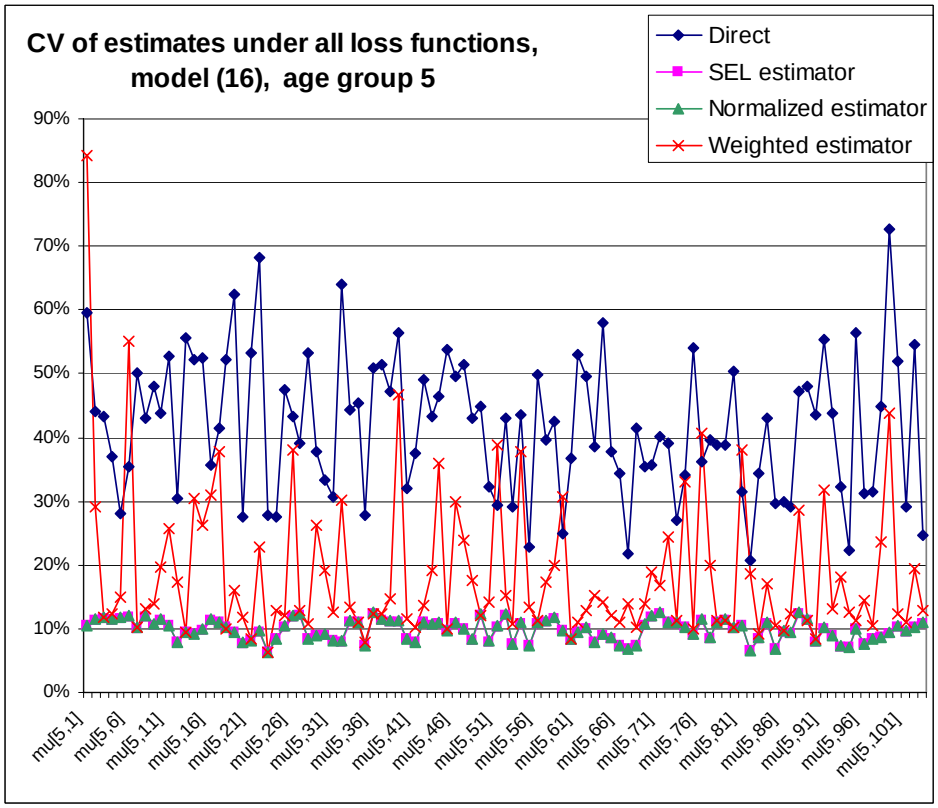
CV of model (16) estimates:

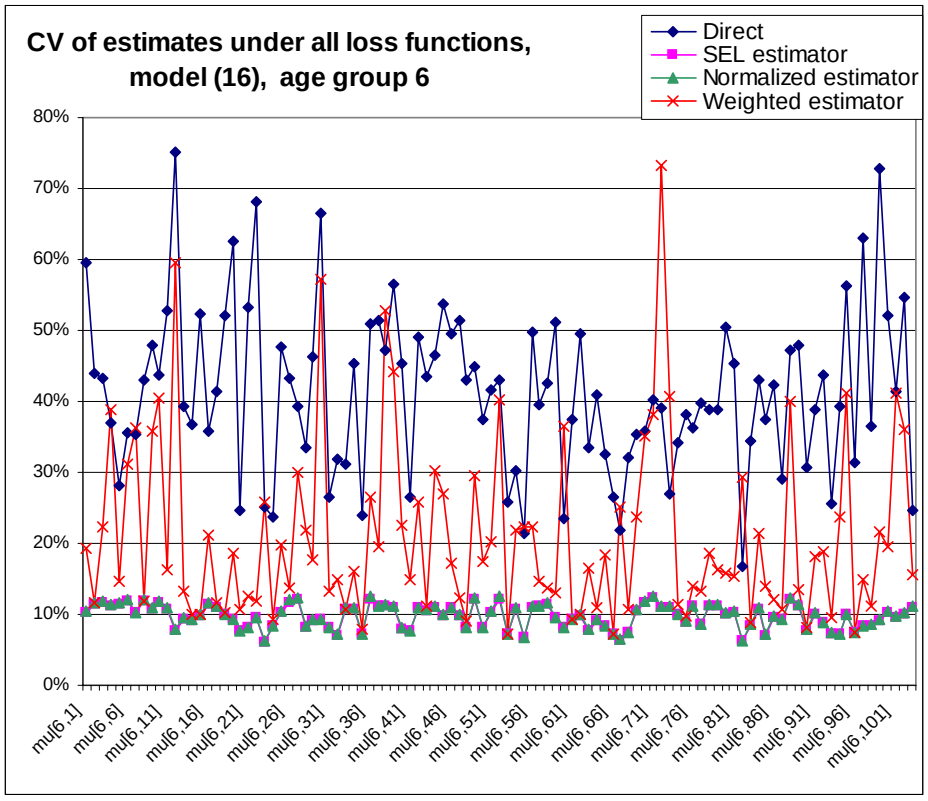


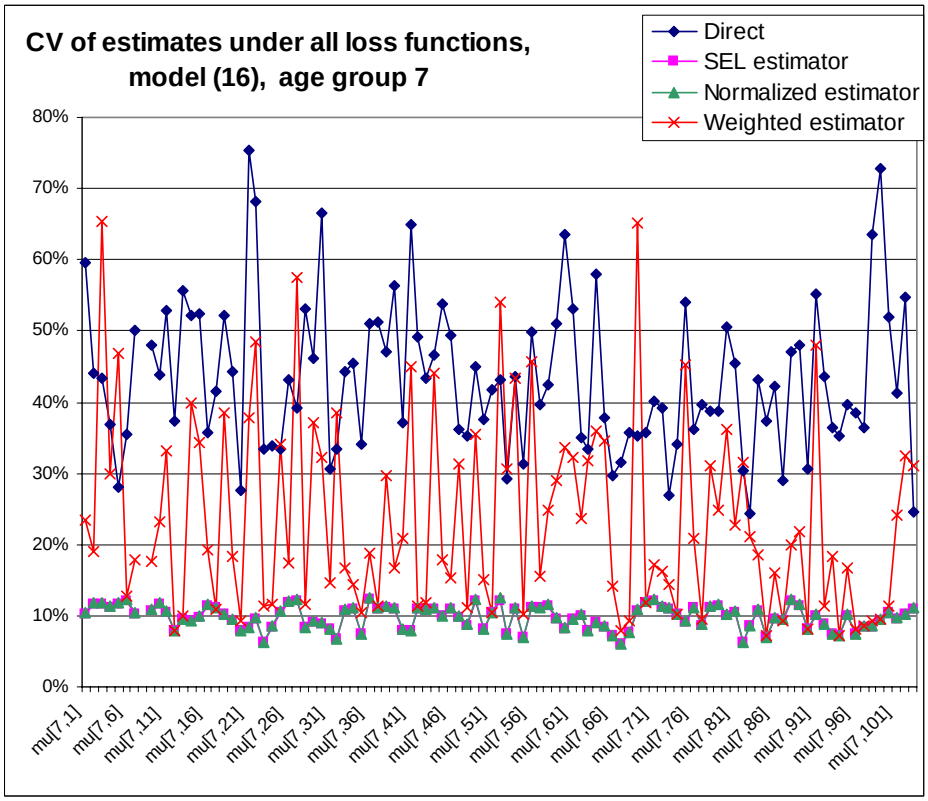

CV of estimates under all loss functions, model (16), age group 2
Direct
SEL estimator
Normalized estimator
Weighted estimator
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[2,1]
mu[2,6]
mu[2,11]
mu[2,16]
mu[2,21]
mu[2,26]
mu[2,31]
mu[2,36]
mu[2,41]
mu[2,46]
mu[2,51]
mu[2,56]
mu[2,61]
mu[2,66]
mu[2,71]
mu[2,76]
mu[2,81]
mu[2,86]
mu[2,91]
mu[2,96]
mu[2,101]

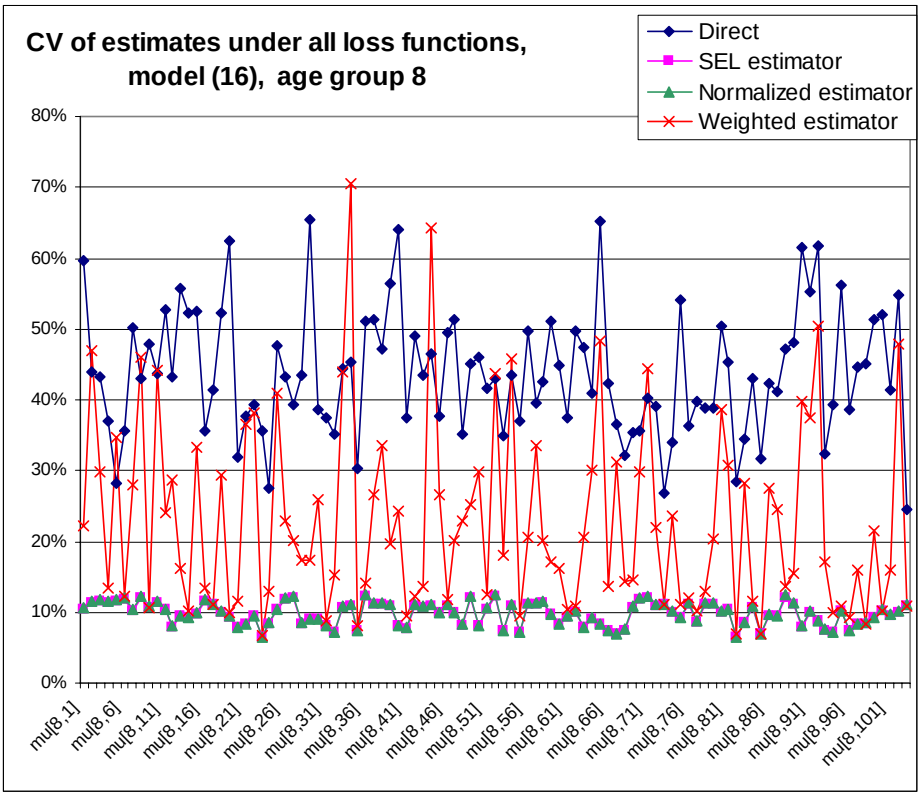

CV of estimates under all loss functions, model (16), age group 3
Direct
SEL estimator
Normalized estimator
Weighted estimator
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[3,1]
mu[3,6]
mu[3,11]
mu[3,16]
mu[3,21]
mu[3,26]
mu[3,31]
mu[3,36]
mu[3,41]
mu[3,46]
mu[3,51]
mu[3,56]
mu[3,61]
mu[3,66]
mu[3,71]
mu[3,76]
mu[3,81]
mu[3,86]
mu[3,91]
mu[3,96]
mu[3,101]

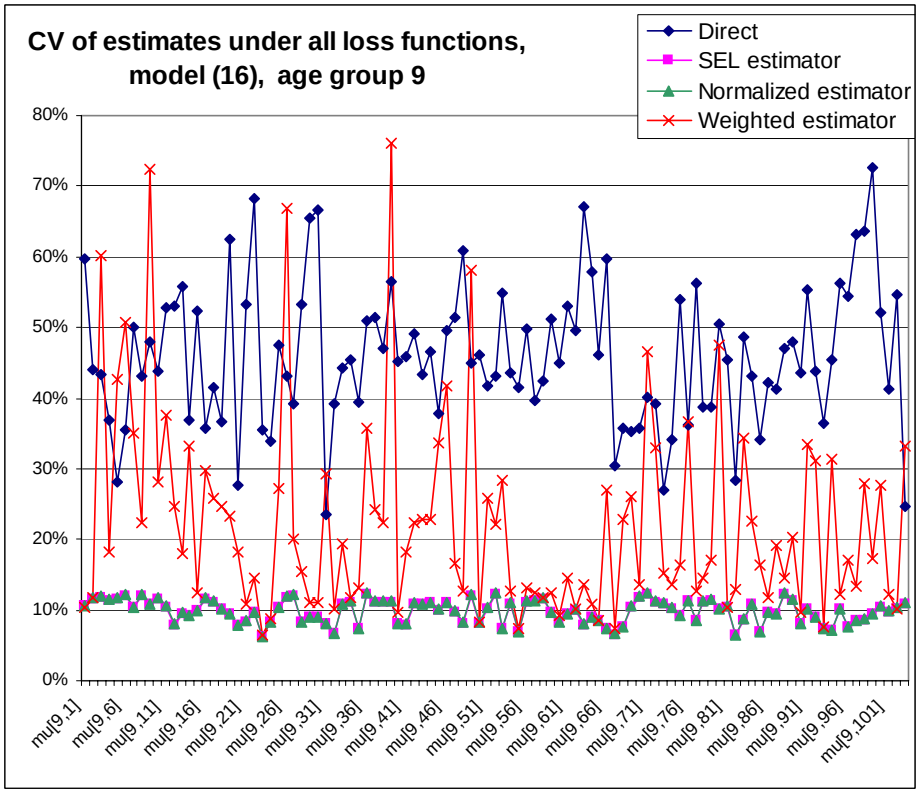

CV of estimates under all loss functions, model (16), age group 4
Direct
SEL estimator
Normalized estimator
Weighted estimator
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[4,1]
mu[4,6]
mu[4,11]
mu[4,16]
mu[4,21]
mu[4,26]
mu[4,31]
mu[4,36]
mu[4,41]
mu[4,46]
mu[4,51]
mu[4,56]
mu[4,61]
mu[4,66]
mu[4,71]
mu[4,76]
mu[4,81]
mu[4,86]
mu[4,91]
mu[4,96]
mu[4,101]

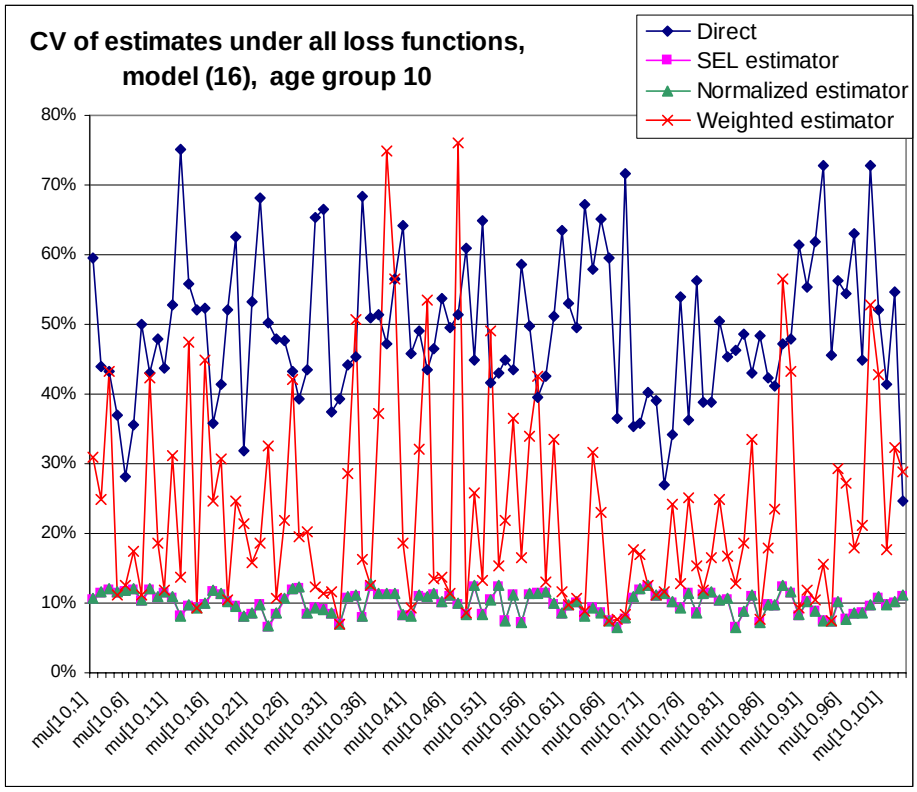

CV of estimates under all loss functions, model (16), age group 5
Direct
SEL estimator
Normalized estimator
Weighted estimator
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[5,1]
mu[5,6]
mu[5,11]
mu[5,16]
mu[5,21]
mu[5,26]
mu[5,31]
mu[5,36]
mu[5,41]
mu[5,46]
mu[5,51]
mu[5,56]
mu[5,61]
mu[5,66]
mu[5,71]
mu[5,76]
mu[5,81]
mu[5,86]
mu[5,91]
mu[5,96]
mu[5,101]

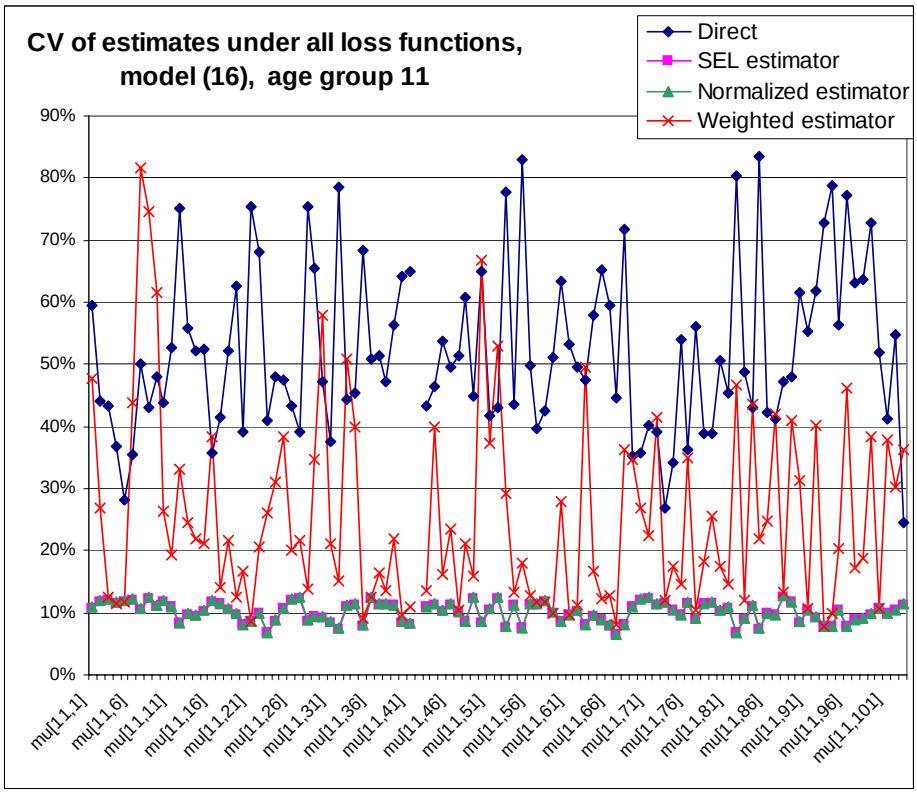

CV of estimates under all loss functions, model (16), age group 6
Direct
SEL estimator
Normalized estimator
Weighted estimator
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[6,1]
mu[6,6]
mu[6,11]
mu[6,16]
mu[6,21]
mu[6,26]
mu[6,31]
mu[6,36]
mu[6,41]
mu[6,46]
mu[6,51]
mu[6,56]
mu[6,61]
mu[6,66]
mu[6,71]
mu[6,76]
mu[6,81]
mu[6,86]
mu[6,91]
mu[6,96]
mu[6,101]

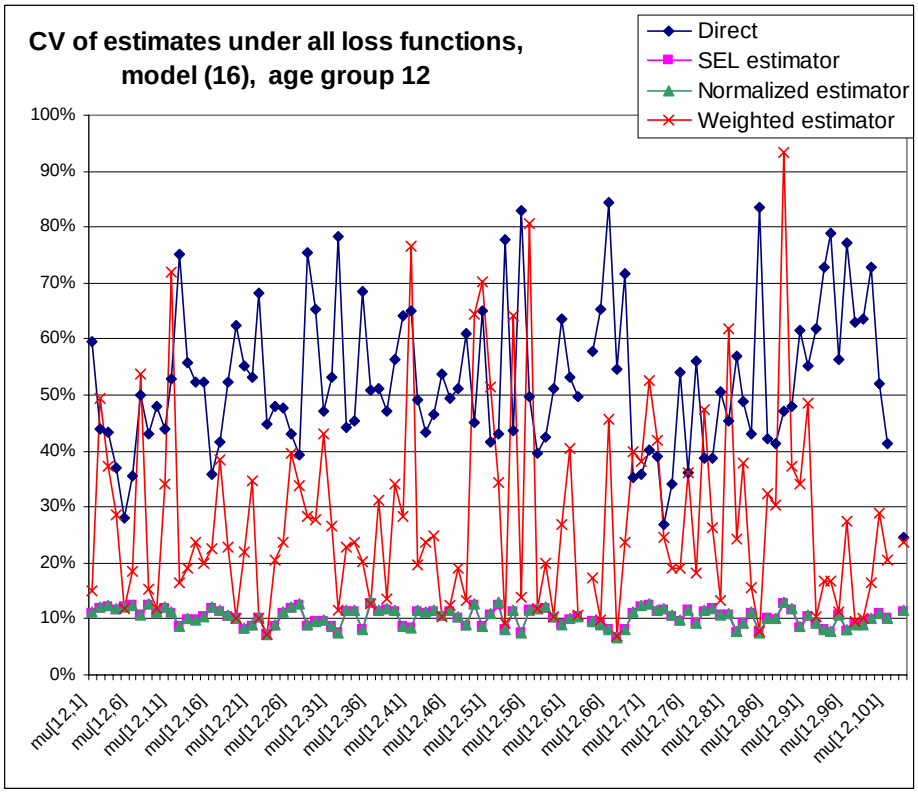

CV of estimates under all loss functions, model (16), age group 7
Direct
SEL estimator
Normalized estimator
Weighted estimator
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[7,1]
mu[7,6]
mu[7,11]
mu[7,16]
mu[7,21]
mu[7,26]
mu[7,31]
mu[7,36]
mu[7,41]
mu[7,46]
mu[7,51]
mu[7,56]
mu[7,61]
mu[7,66]
mu[7,71]
mu[7,76]
mu[7,81]
mu[7,86]
mu[7,91]
mu[7,96]
mu[7,101]

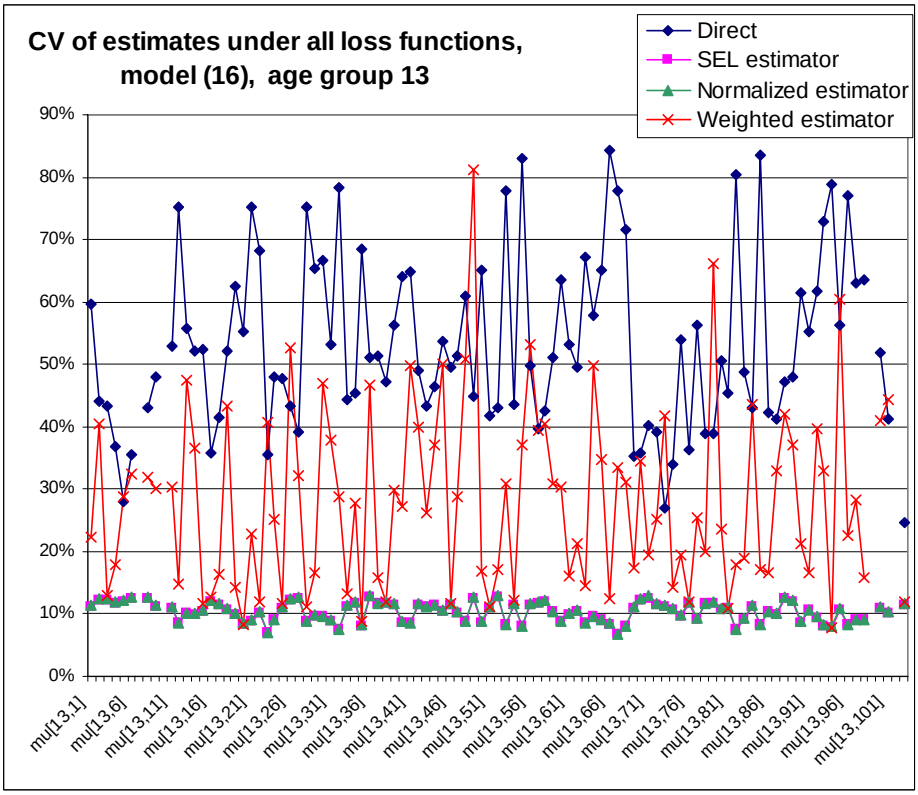

CV of estimates under all loss functions, model (16), age group 8
Direct
SEL estimator
Normalized estimator
Weighted estimator
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[8,1]
mu[8,6]
mu[8,11]
mu[8,16]
mu[8,21]
mu[8,26]
mu[8,31]
mu[8,36]
mu[8,41]
mu[8,46]
mu[8,51]
mu[8,56]
mu[8,61]
mu[8,66]
mu[8,71]
mu[8,76]
mu[8,81]
mu[8,86]
mu[8,91]
mu[8,96]
mu[8,101]

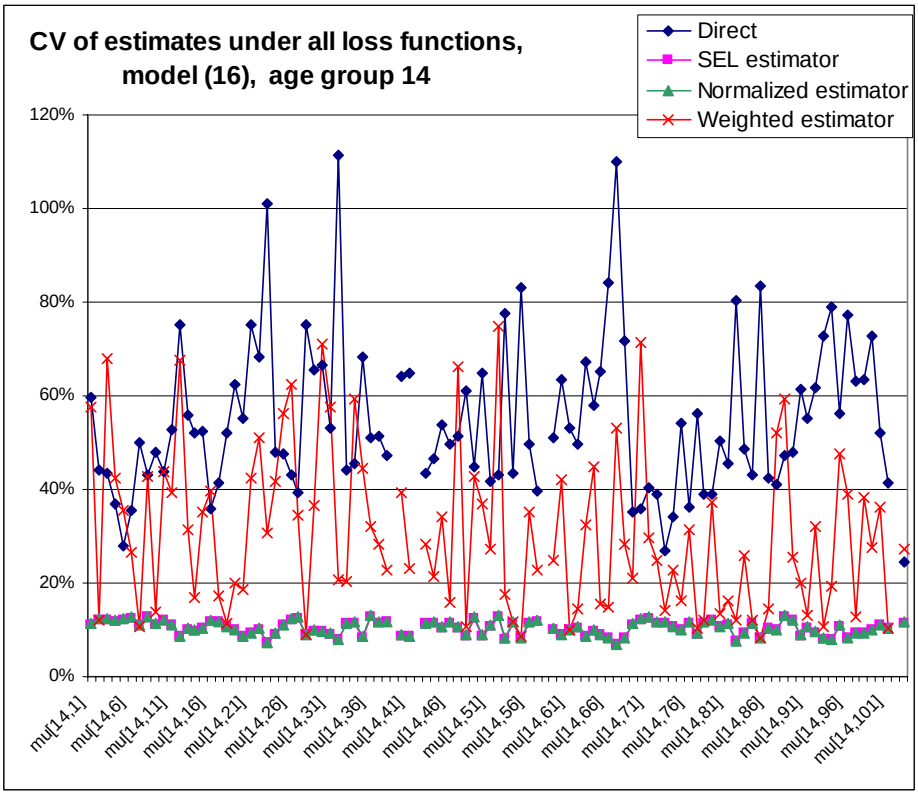

CV of estimates under all loss functions, model (16), age group 9
Direct
SEL estimator
Normalized estimator
Weighted estimator
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[9,1]
mu[9,6]
mu[9,11]
mu[9,16]
mu[9,21]
mu[9,26]
mu[9,31]
mu[9,36]
mu[9,41]
mu[9,46]
mu[9,51]
mu[9,56]
mu[9,61]
mu[9,66]
mu[9,71]
mu[9,76]
mu[9,81]
mu[9,86]
mu[9,91]
mu[9,96]
mu[9,101]


CV of estimates under all loss functions, model (16), age group 10
Direct
SEL estimator
Normalized estimator
Weighted estimator
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[10,1]
mu[10,6]
mu[10,11]
mu[10,16]
mu[10,21]
mu[10,26]
mu[10,31]
mu[10,36]
mu[10,41]
mu[10,46]
mu[10,51]
mu[10,56]
mu[10,61]
mu[10,66]
mu[10,71]
mu[10,76]
mu[10,81]
mu[10,86]
mu[10,91]
mu[10,96]
mu[10,101]


CV of estimates under all loss functions, model (16), age group 11
Direct
SEL estimator
Normalized estimator
Weighted estimator
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[11,1]
mu[11,6]
mu[11,11]
mu[11,16]
mu[11,21]
mu[11,26]
mu[11,31]
mu[11,36]
mu[11,41]
mu[11,46]
mu[11,51]
mu[11,56]
mu[11,61]
mu[11,66]
mu[11,71]
mu[11,76]
mu[11,81]
mu[11,86]
mu[11,91]
mu[11,96]
mu[11,101]


CV of estimates under all loss functions, model (16), age group 12
Direct
SEL estimator
Normalized estimator
Weighted estimator
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[12,1]
mu[12,6]
mu[12,11]
mu[12,16]
mu[12,21]
mu[12,26]
mu[12,31]
mu[12,36]
mu[12,41]
mu[12,46]
mu[12,51]
mu[12,56]
mu[12,61]
mu[12,66]
mu[12,71]
mu[12,76]
mu[12,81]
mu[12,86]
mu[12,91]
mu[12,96]
mu[12,101]

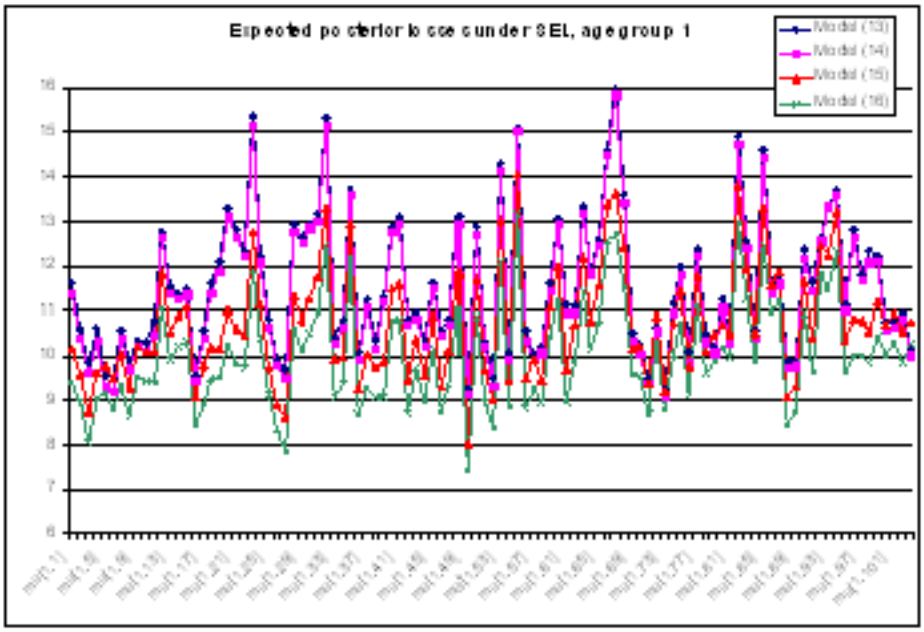

CV of estimates under all loss functions, model (16), age group 13
Direct
SEL estimator
Normalized estimator
Weighted estimator
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
mu[13,1]
mu[13,6]
mu[13,11]
mu[13,16]
mu[13,21]
mu[13,26]
mu[13,31]
mu[13,36]
mu[13,41]
mu[13,46]
mu[13,51]
mu[13,56]
mu[13,61]
mu[13,66]
mu[13,71]
mu[13,76]
mu[13,81]
mu[13,86]
mu[13,91]
mu[13,96]
mu[13,101]

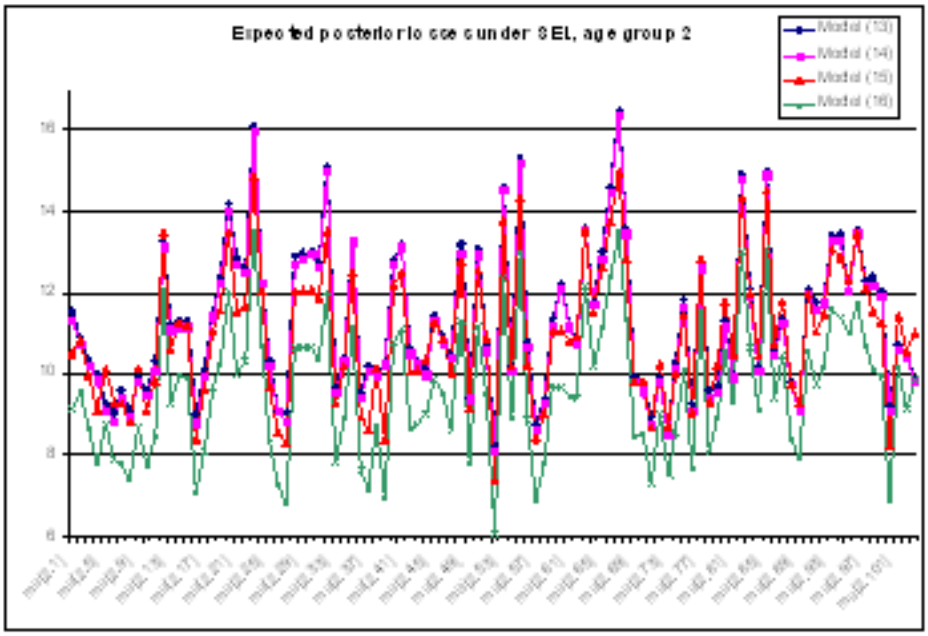

CV of estimates under all loss functions, model (16), age group 14
Direct
SEL estimator
Normalized estimator
Weighted estimator
120%
100%
80%
60%
40%
20%
0%
mu[14,1]
mu[14,6]
mu[14,11]
mu[14,16]
mu[14,21]
mu[14,26]
mu[14,31]
mu[14,36]
mu[14,41]
mu[14,46]
mu[14,51]
mu[14,56]
mu[14,61]
mu[14,66]
mu[14,71]
mu[14,76]
mu[14,81]
mu[14,86]
mu[14,91]
mu[14,96]
mu[14,101]

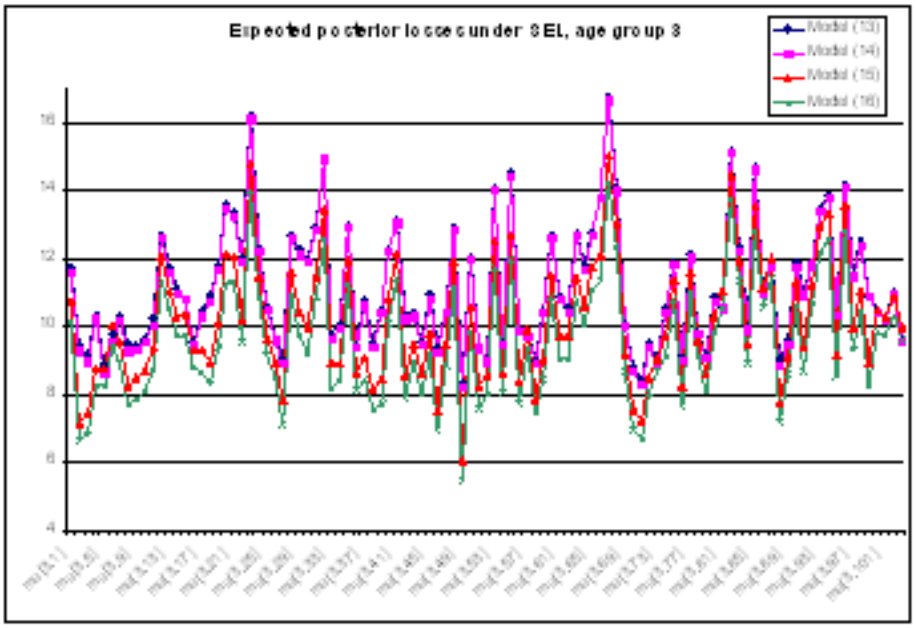
Appendix G

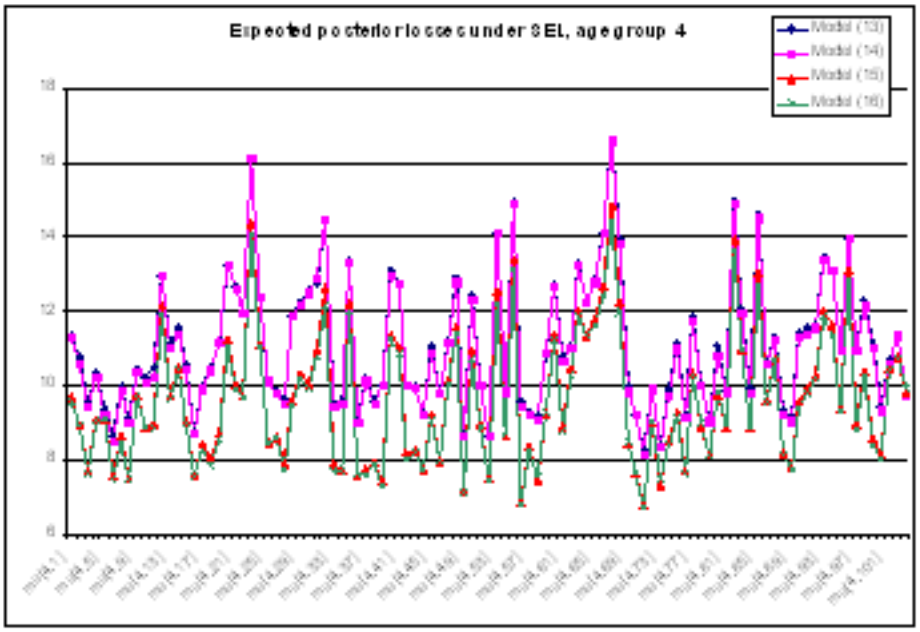
Expected Posterior Losses

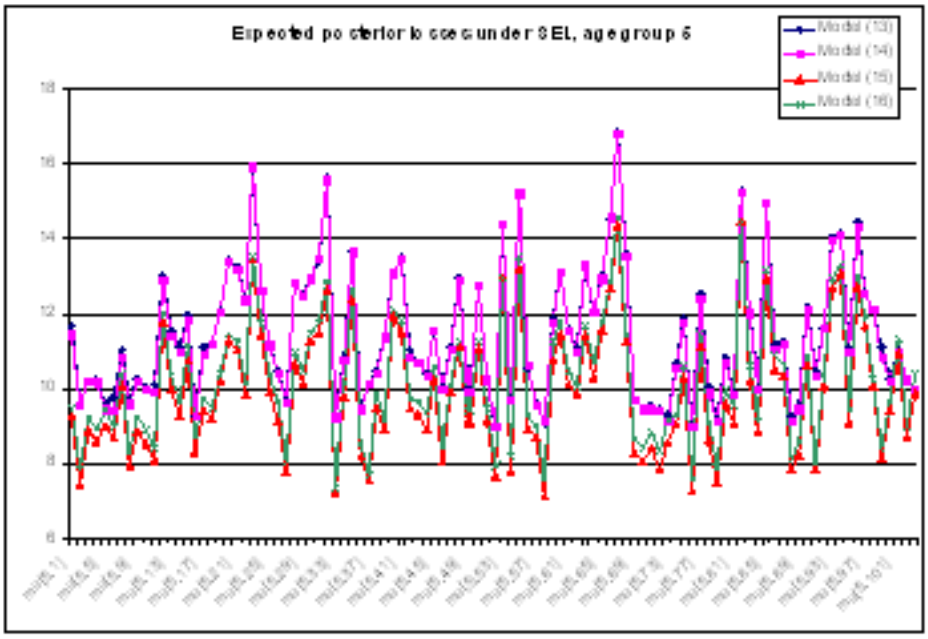
Expected posterior losses are presented below - graphed by age group as well as by the loss function. Under SEL, the following expected posterior losses were obtained (graphed on a log-scale to allow for a better representation by reducing the variance in the extreme estimates for different domains):

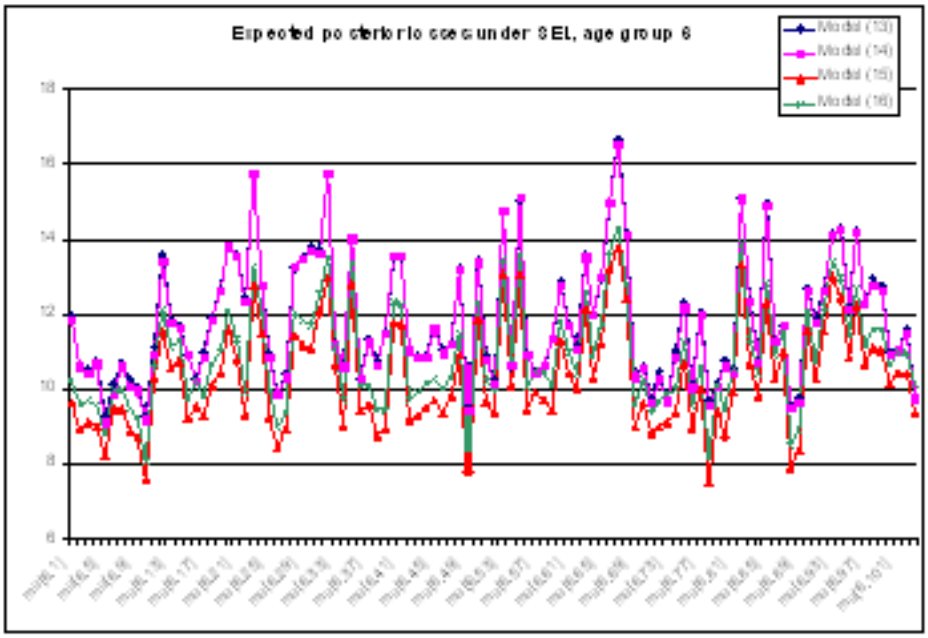


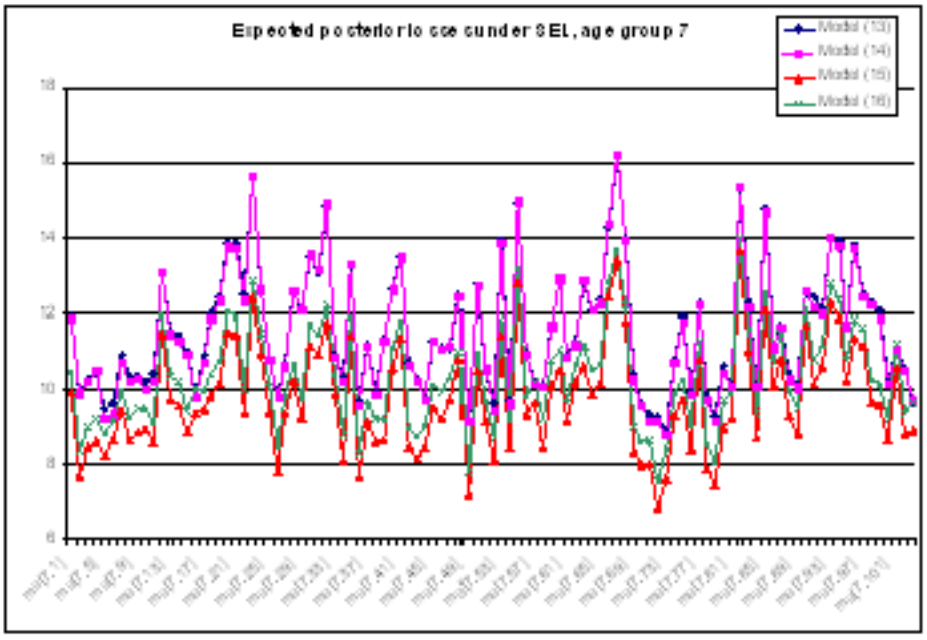

Expected posterior loss under SEL, age group 2
Model (13)
Model (14)
Model (15)
Model (16)

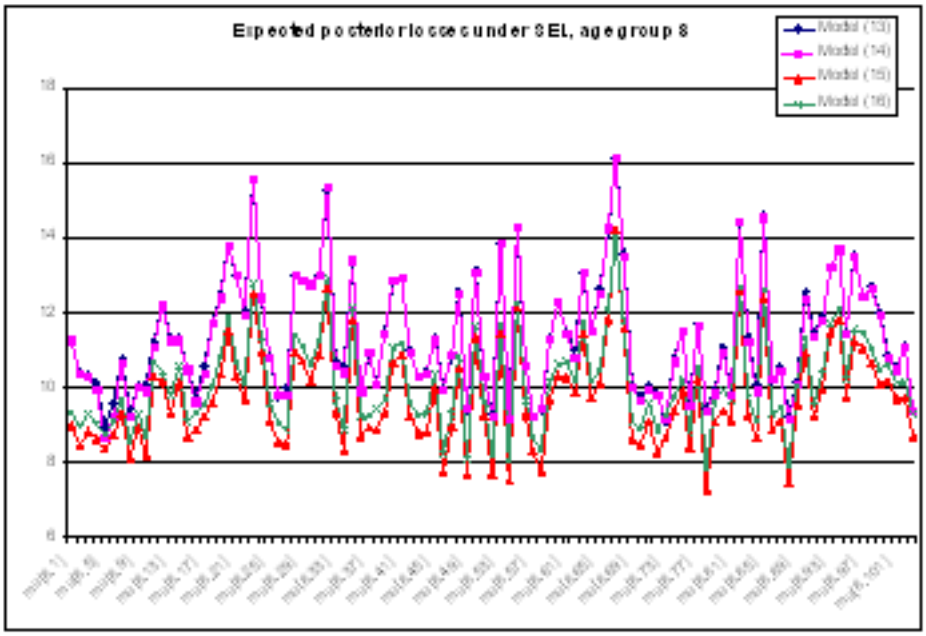

Expected posterior loss under SEL, age group 3
Model (13)
Model (14)
Model (15)
Model (16)

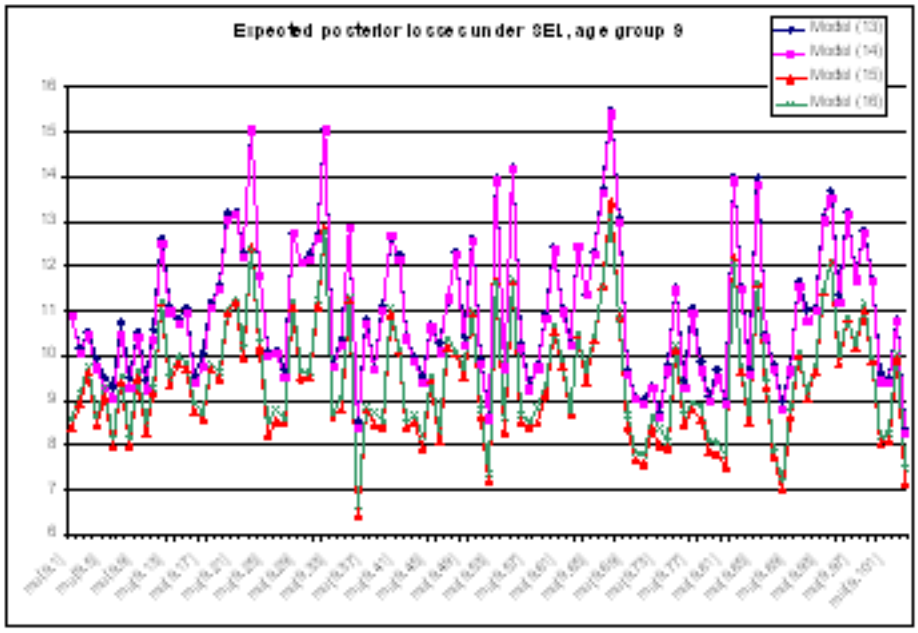

Expected posterior loss under SEL, age group 4
Model (13)
Model (14)
Model (15)
Model (16)
18
16
14
12
10
8
6

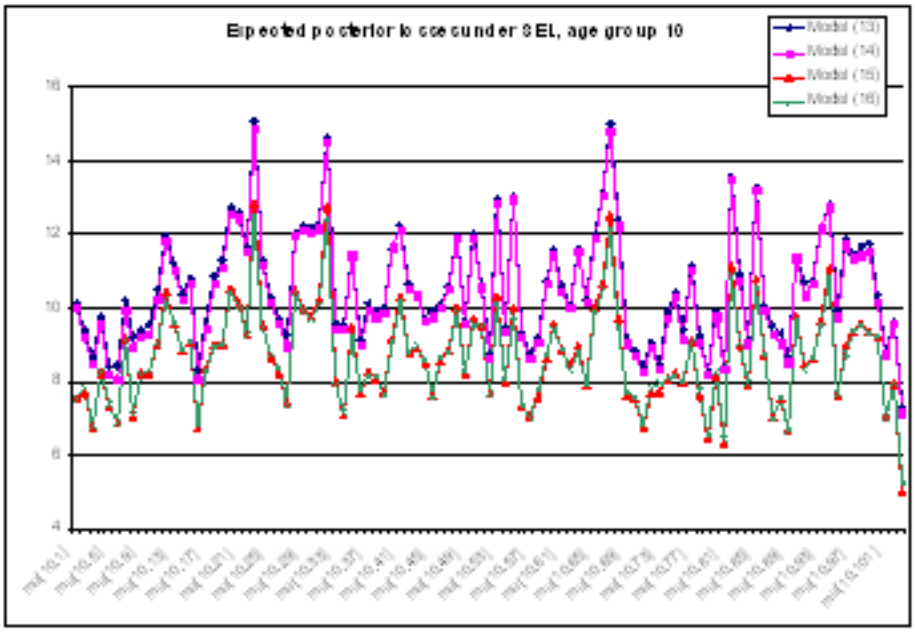

Expected posterior loss under SEL, age group 5
Model (13)
Model (14)
Model (15)
Model (16)
18
16
14
12
10
8
6

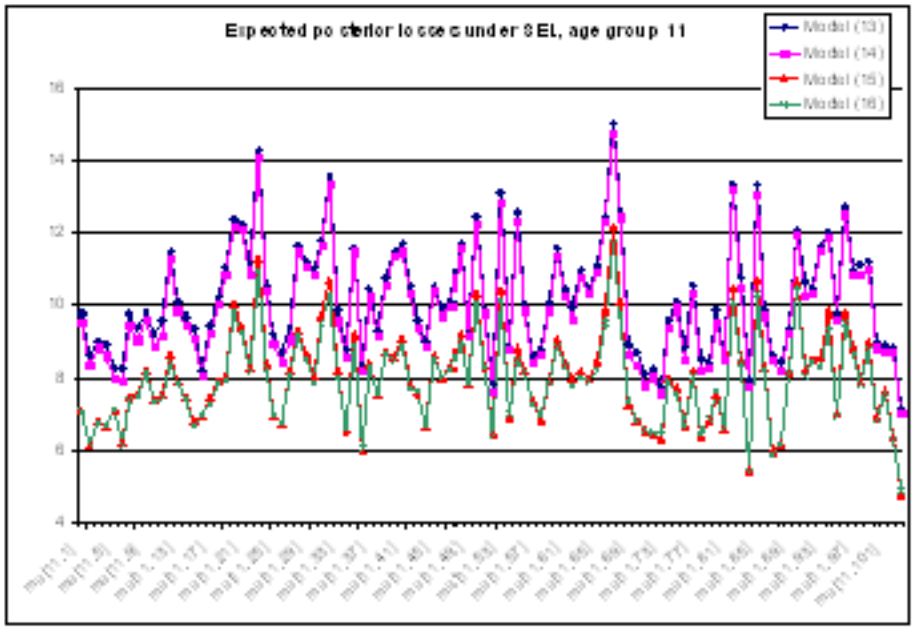

Expected posterior sse under SEL, age group 6
Model (13)
Model (14)
Model (15)
Model (16)
18
16
14
12
10
8
6

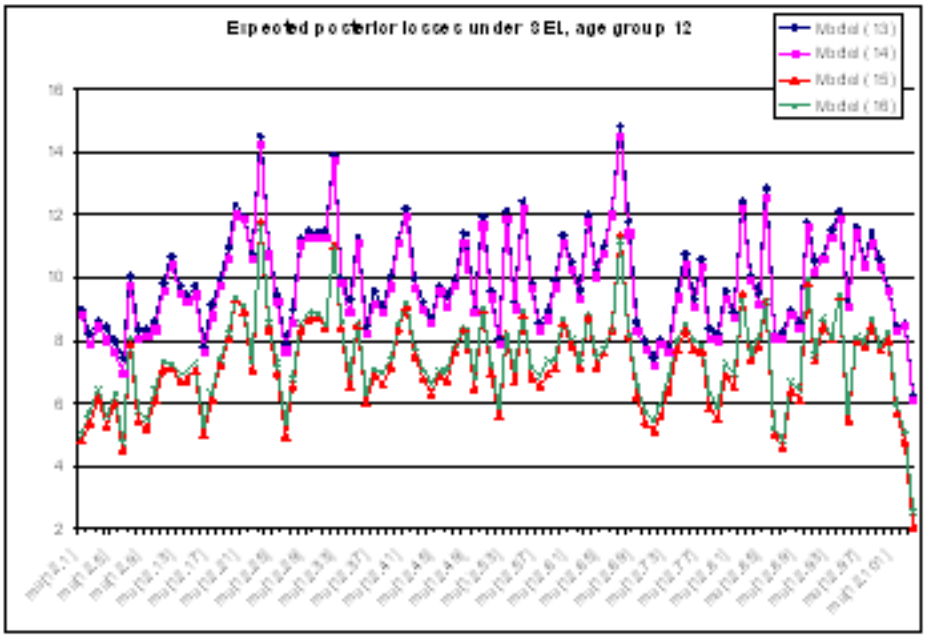

Expected posterior sse under SEL, age group 7
Model (13)
Model (14)
Model (15)
Model (16)
18
16
14
12
10
8
6

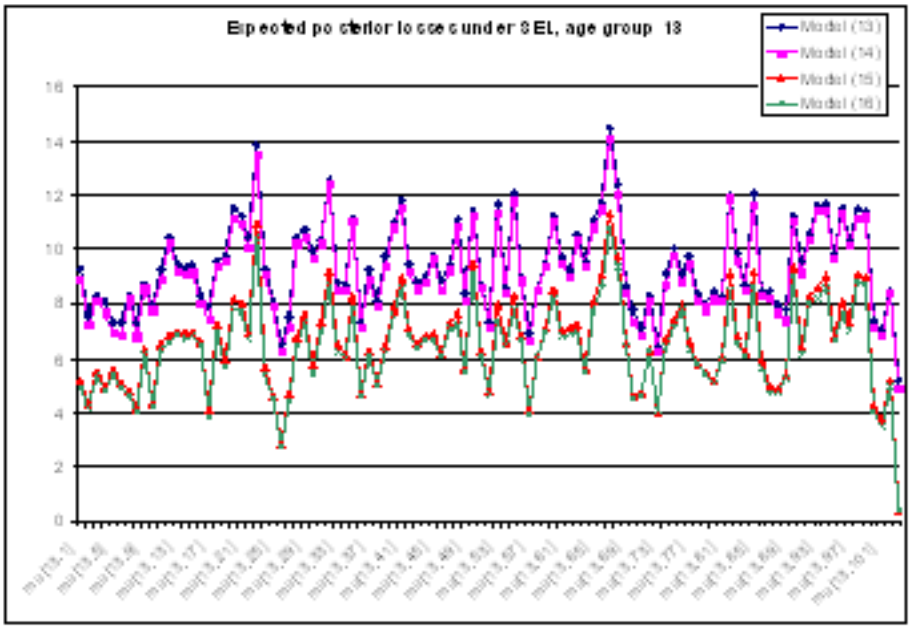

Expected posterior loss under SEL, age group 8
Model (13)
Model (14)
Model (15)
Model (16)

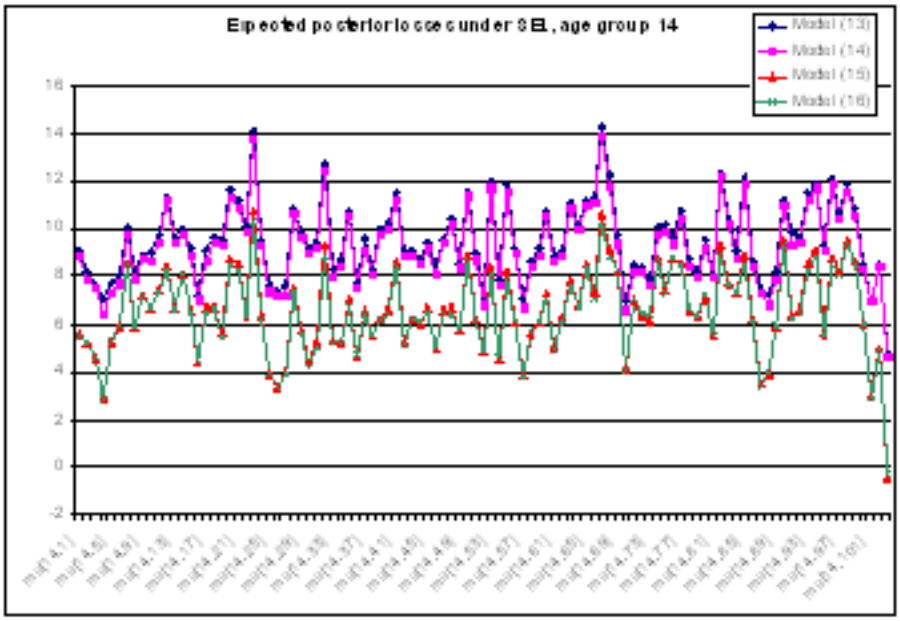

Expected posterior loss under SEL, age group 9
Model (13)
Model (14)
Model (15)
Model (16)


Expected posterior loss under SEL, age group 10
Model (13)
Model (14)
Model (15)
Model (16)
16
14
12
10
8
6
4

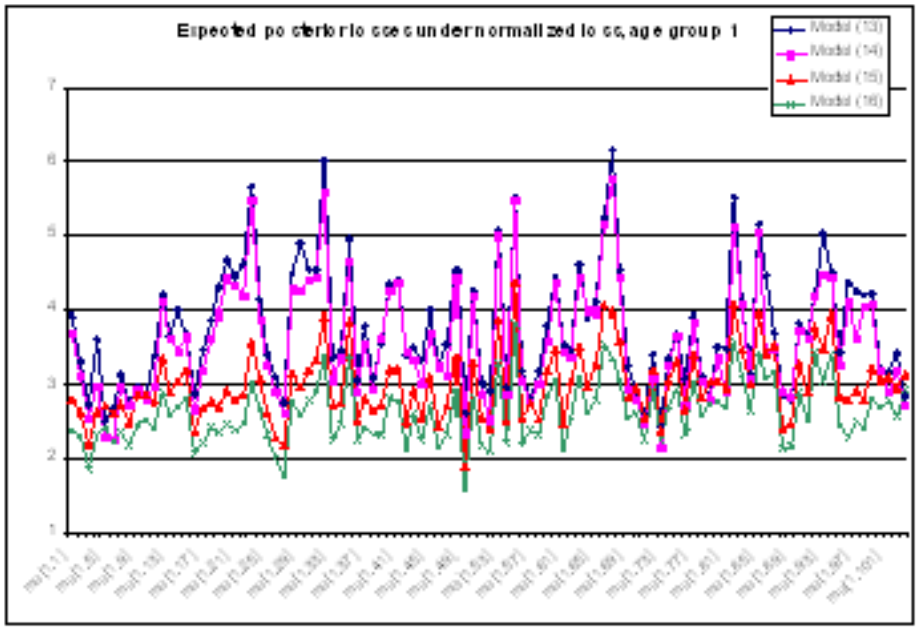

Expected posterior loss under SEL, age group 11
Model (13)
Model (14)
Model (15)
Model (16)
16
14
12
10
8
6
4

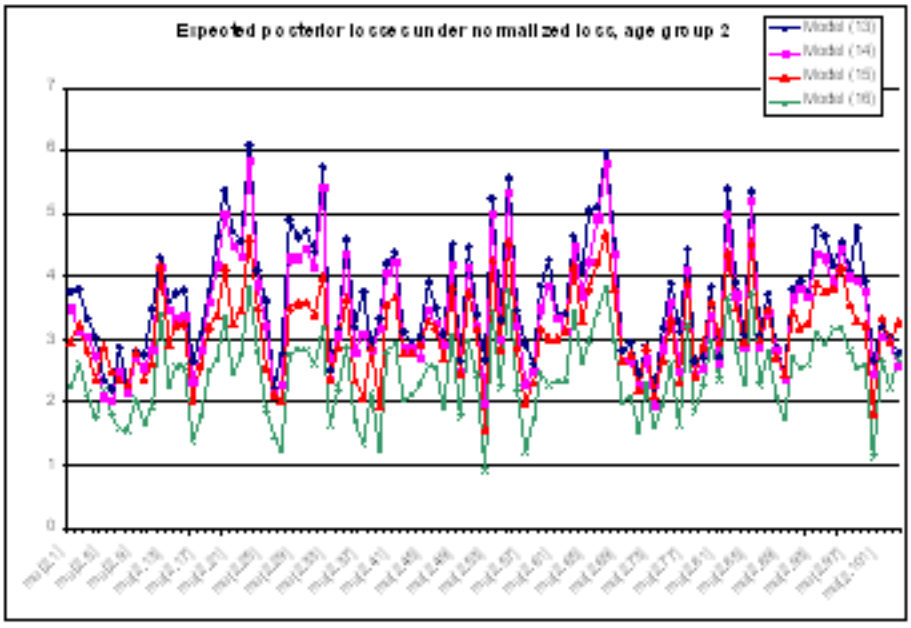

Expected posterior loccs under SEL, age group 12
Model (13)
Model (14)
Model (15)
Model (16)
16
14
12
10
8
6
4
2

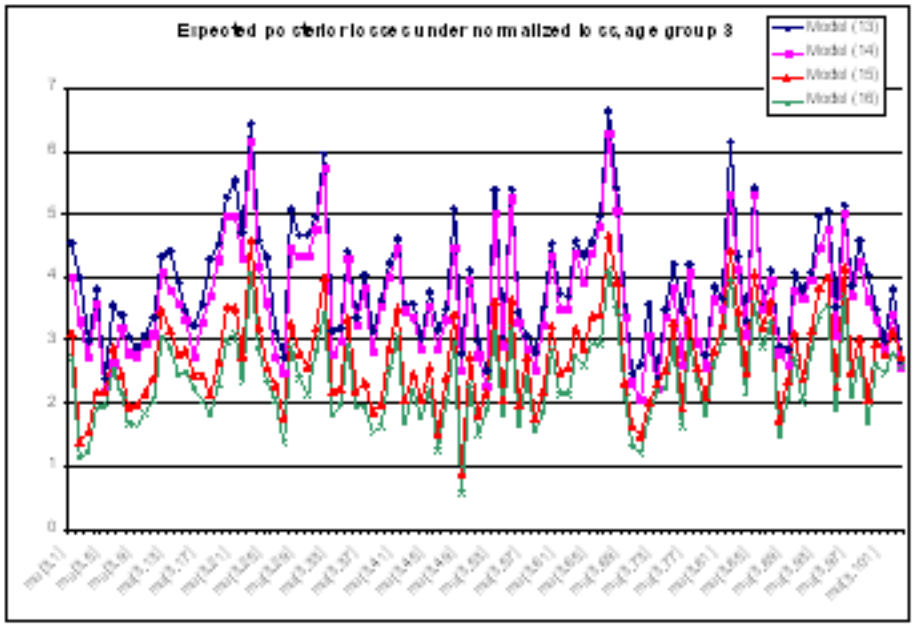

Expected posterior loccs under SEL, age group 13
Model (13)
Model (14)
Model (15)
Model (16)
16
14
12
10
8
6
4
2
0

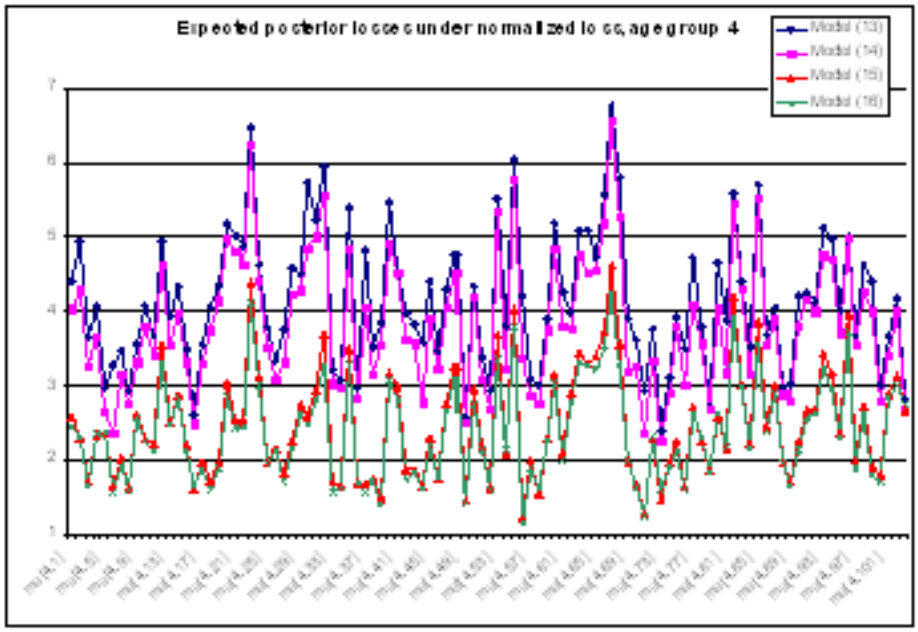

Expected posterior losses under SEL, age group 14
Model (13)
Model (14)
Model (15)
Model (16)
16
14
12
10
8
6
4
2
0
-2

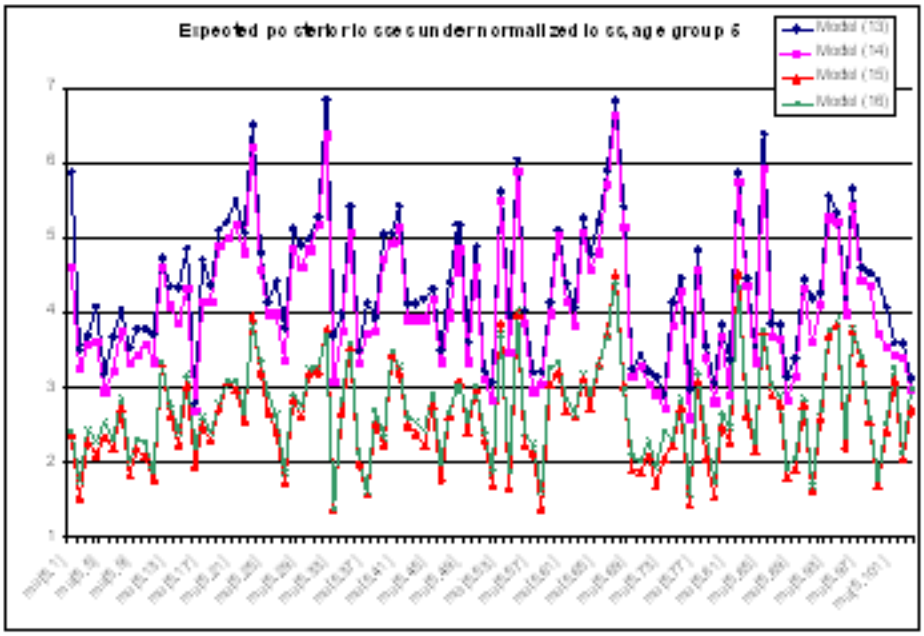
Under NSEL, the following expected posterior losses were obtained (graphed on a log-scale to allow for a better representation by reducing the variance in the extreme estimates for different domains):

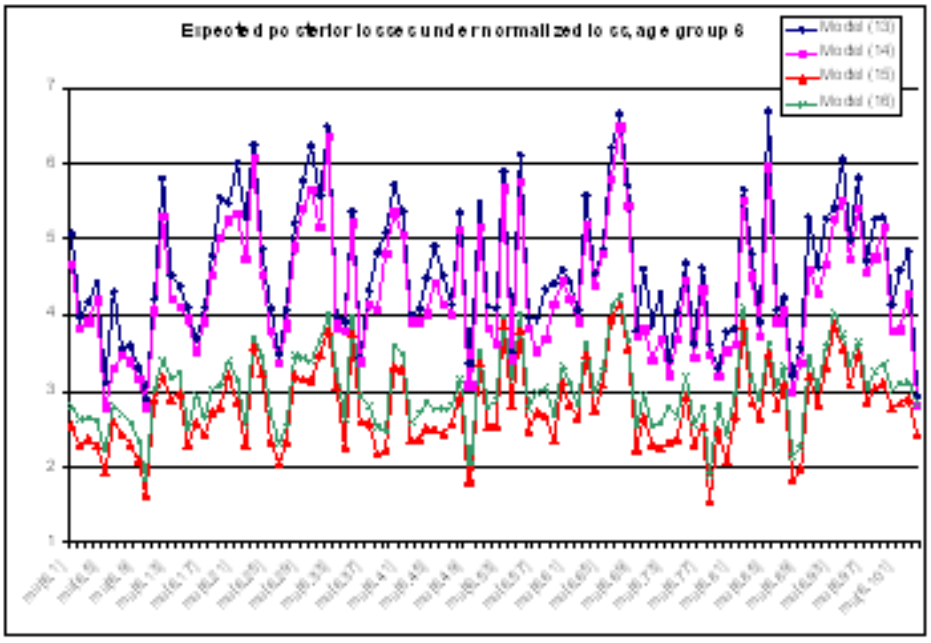


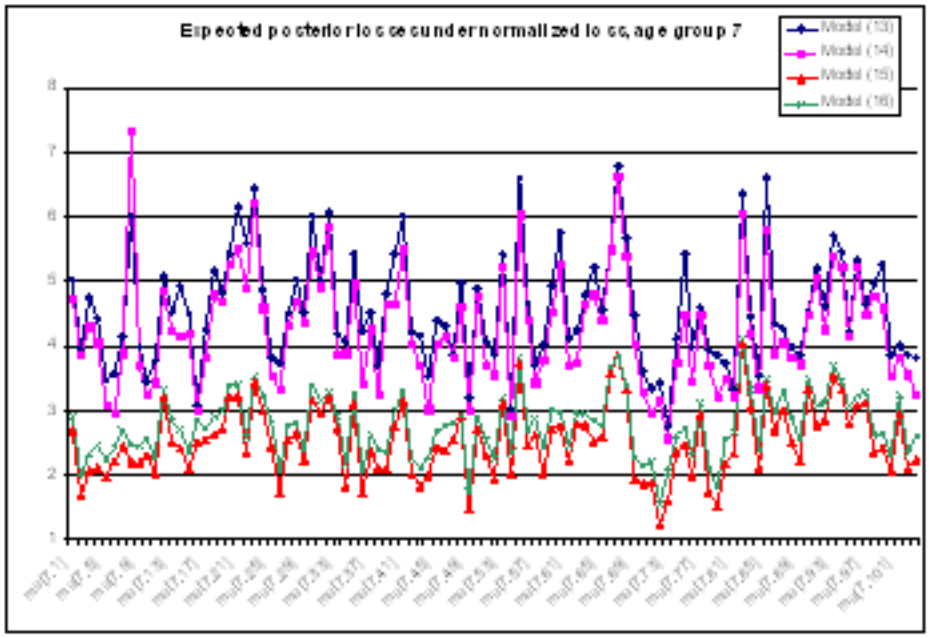


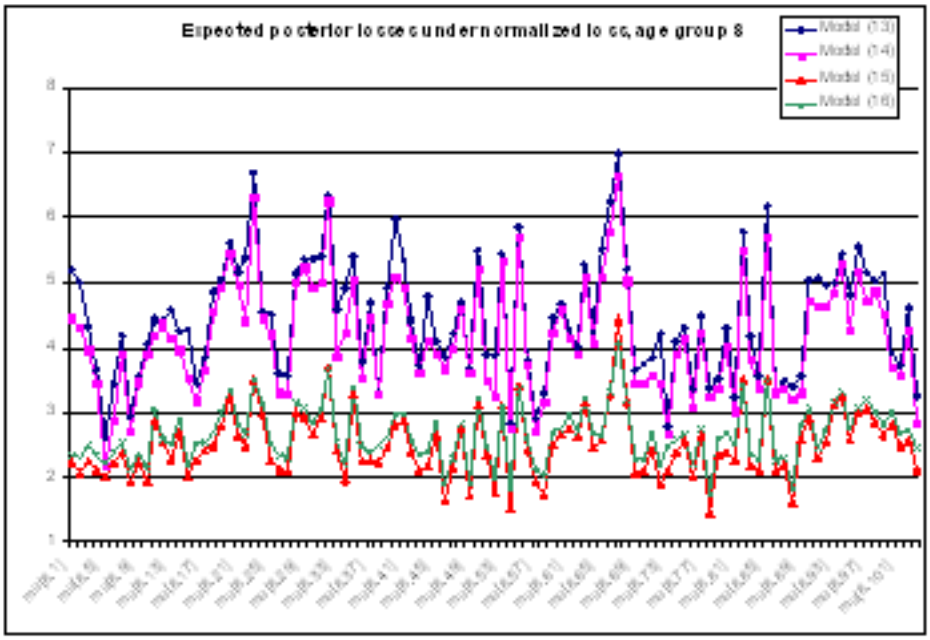

Expected posterior loss under normalized loss, age group 3
Model (13)
Model (14)
Model (15)
Model (16)

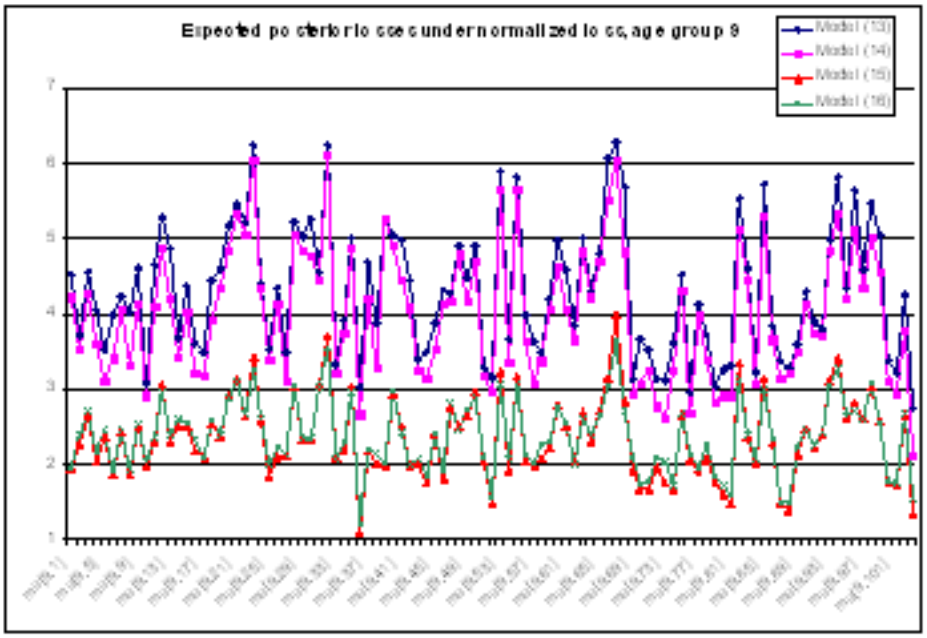

Expected posterior loss under normalized loss, age group 4
Model (13)
Model (14)
Model (15)
Model (16)

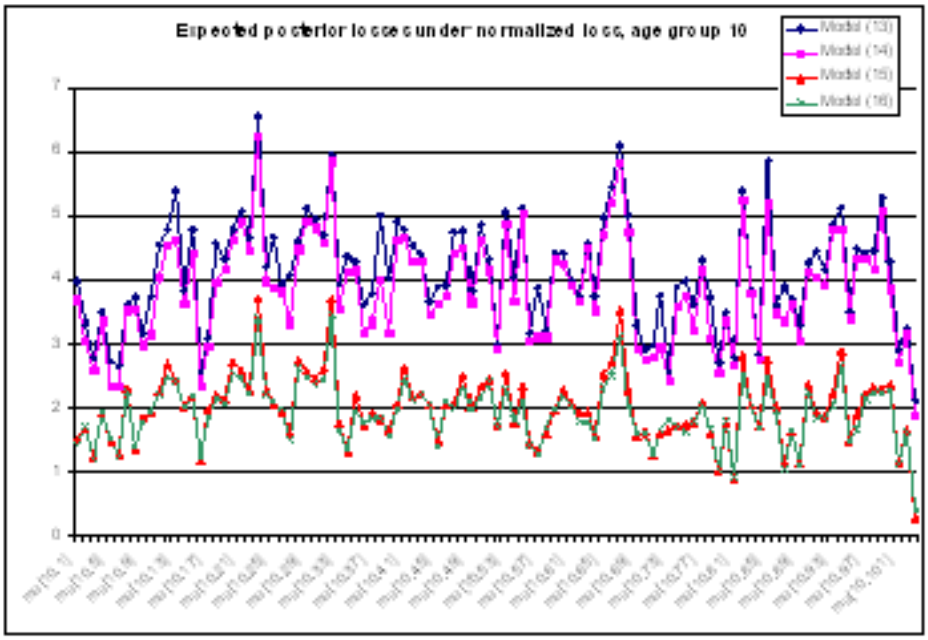

Expected posterior log odds under normalized log odds, age group 6
Model (13)
Model (14)
Model (15)
Model (16)
7
6
5
4
3
2
1

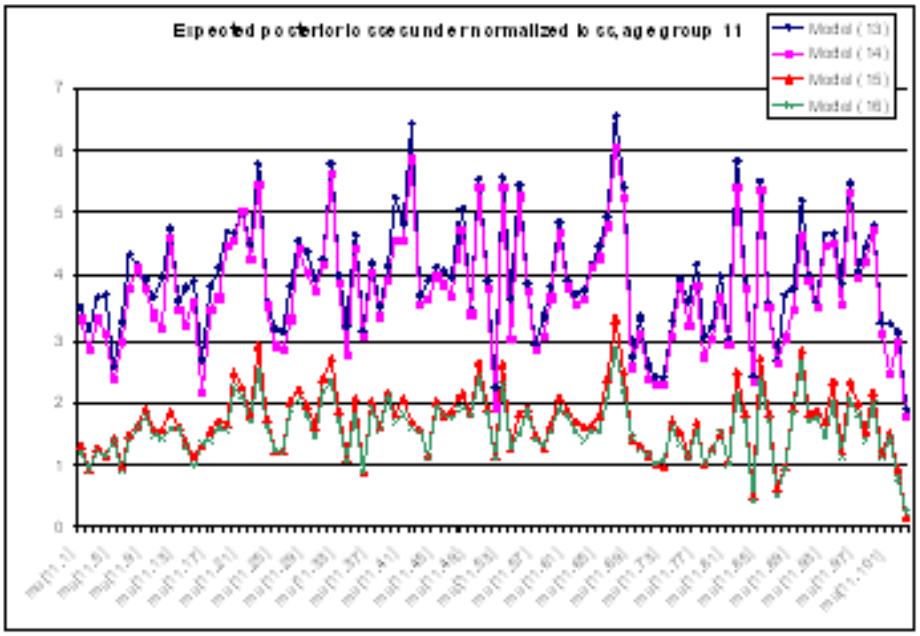

Expected posterior log odds under normalized log odds, age group 6
Model (13)
Model (14)
Model (15)
Model (16)
7
6
5
4
3
2
1

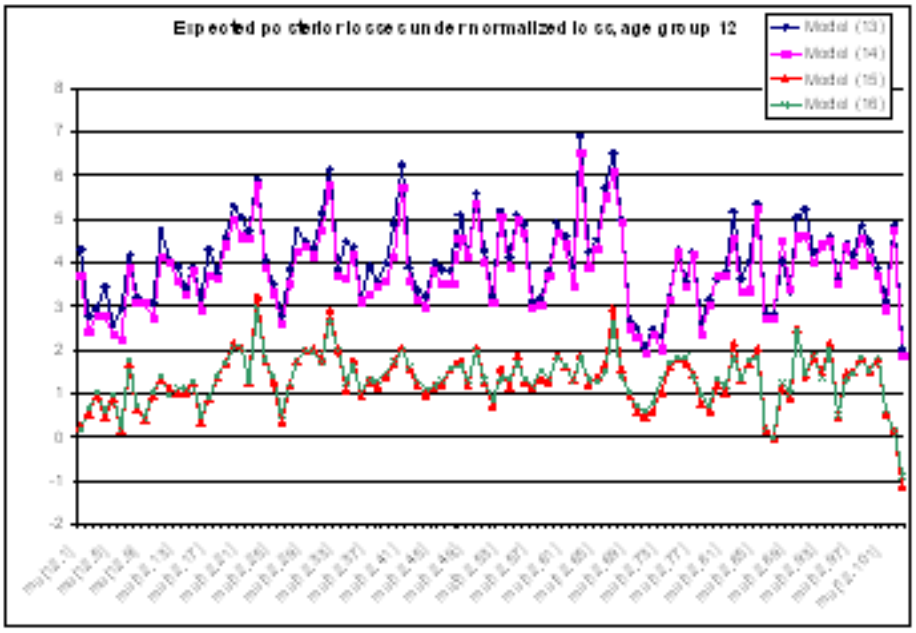

Expected posterior loss under normalized loss, age group 7
Model (13)
Model (14)
Model (15)
Model (16)

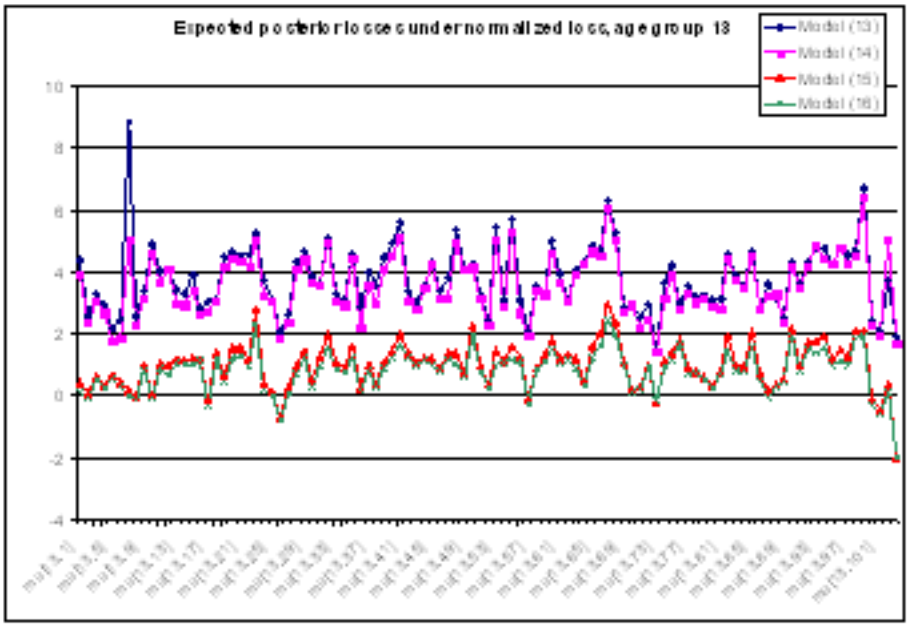

Expected posterior loss under normalized loss, age group 8
Model (13)
Model (14)
Model (15)
Model (16)

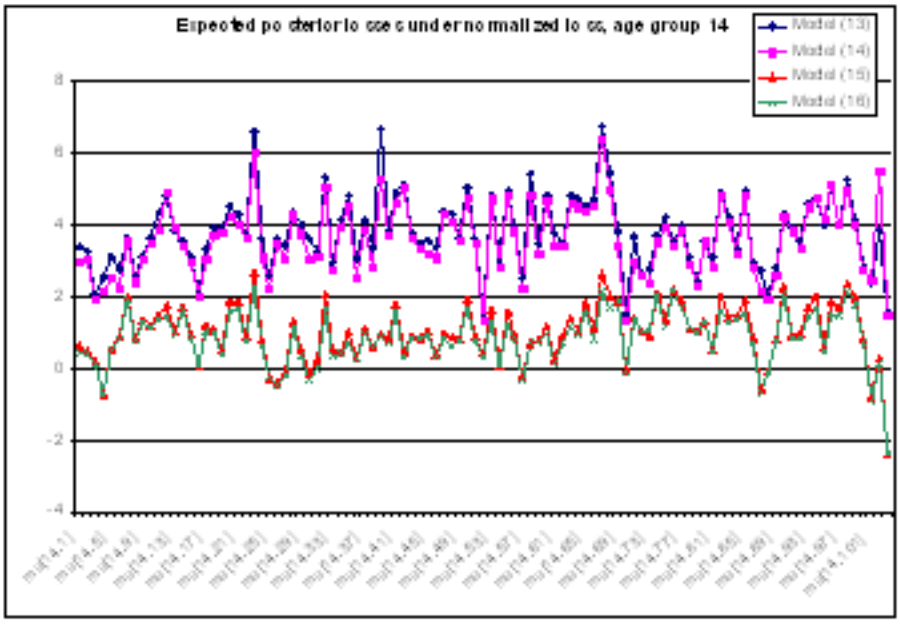

Expected posterior loss under normalized loss, age group 9
Model (13)
Model (14)
Model (15)
Model (16)


Expected posterior loss under normalized loss, age group 10
Model (13)
Model (14)
Model (15)
Model (16)

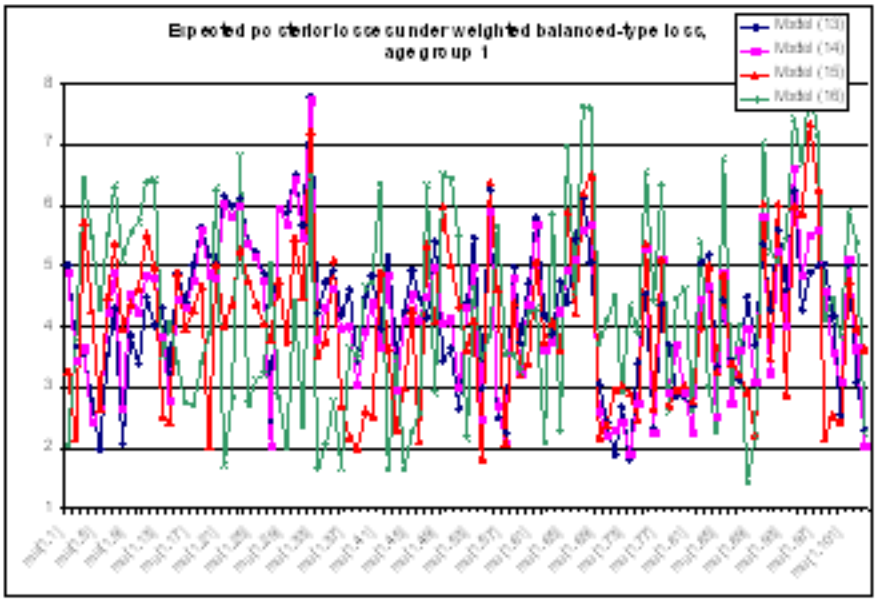

Expected posterior loss under normalized loss, age group 11
Model (13)
Model (14)
Model (15)
Model (16)

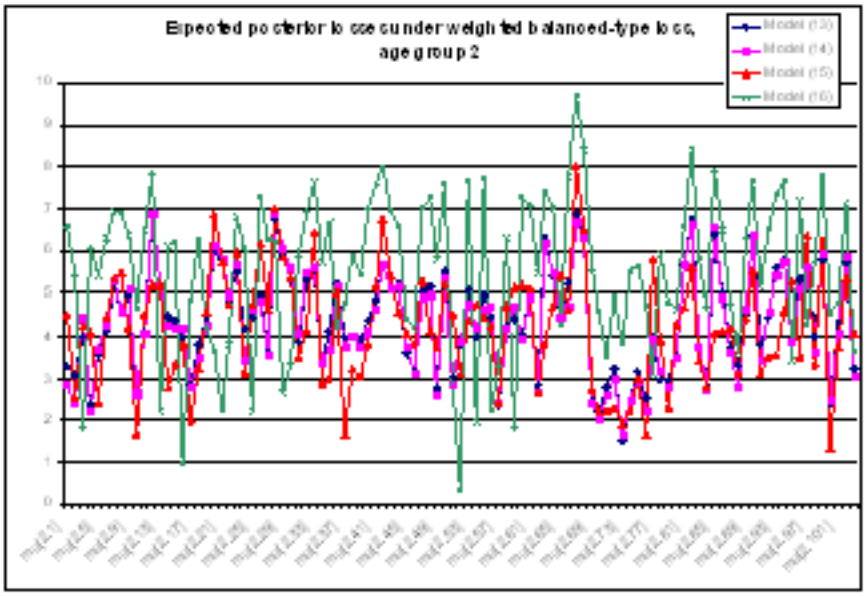

Expected posterior loss under normalized loss, age group 12
Model (13)
Model (14)
Model (15)
Model (16)

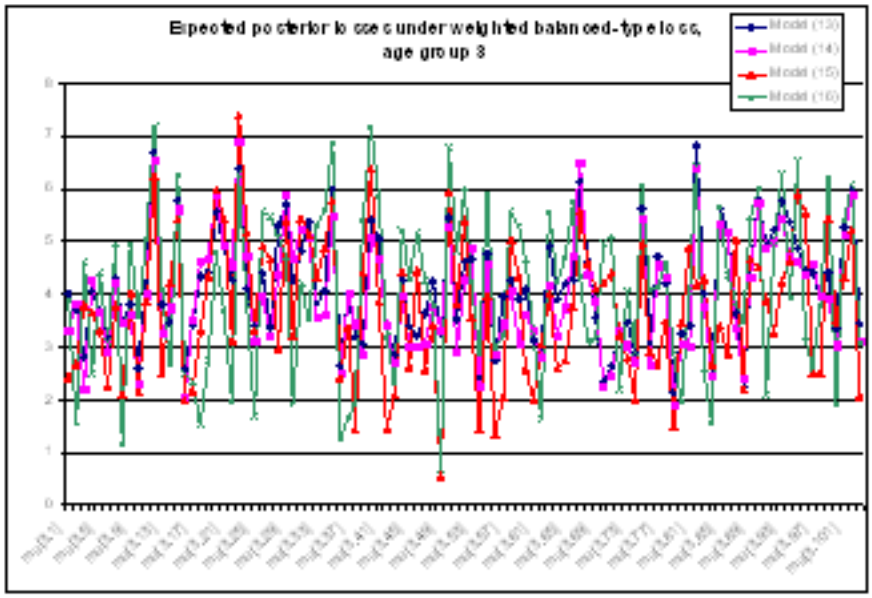

Expected posterior log cse under normalized loss, age group 13
Model (13)
Model (14)
Model (15)
Model (16)

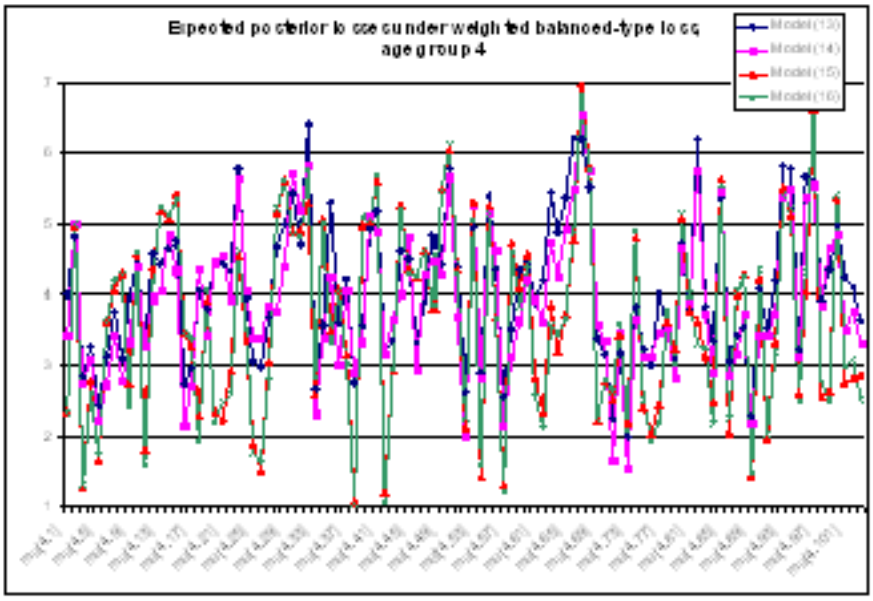

Expected posterior log cse under normalized loss, age group 14
Model (13)
Model (14)
Model (15)
Model (16)

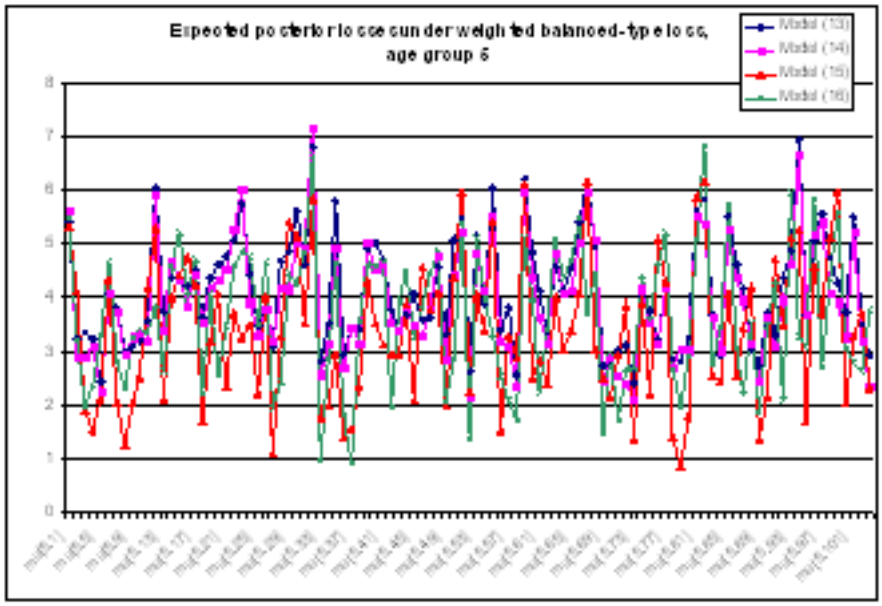
Under WBL, the following expected posterior losses were obtained (graphed on a log-scale to allow for a better representation by reducing the variance in the extreme estimates for different domains):

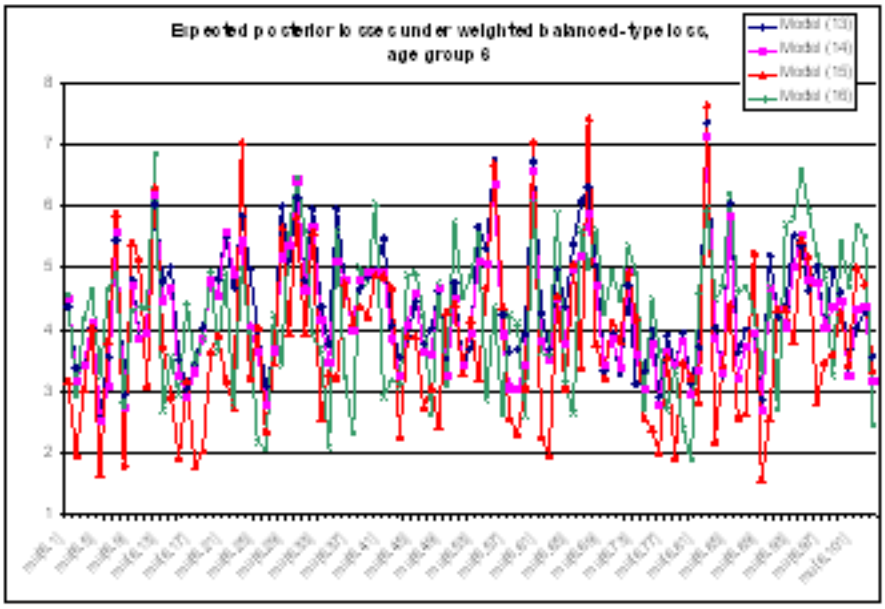


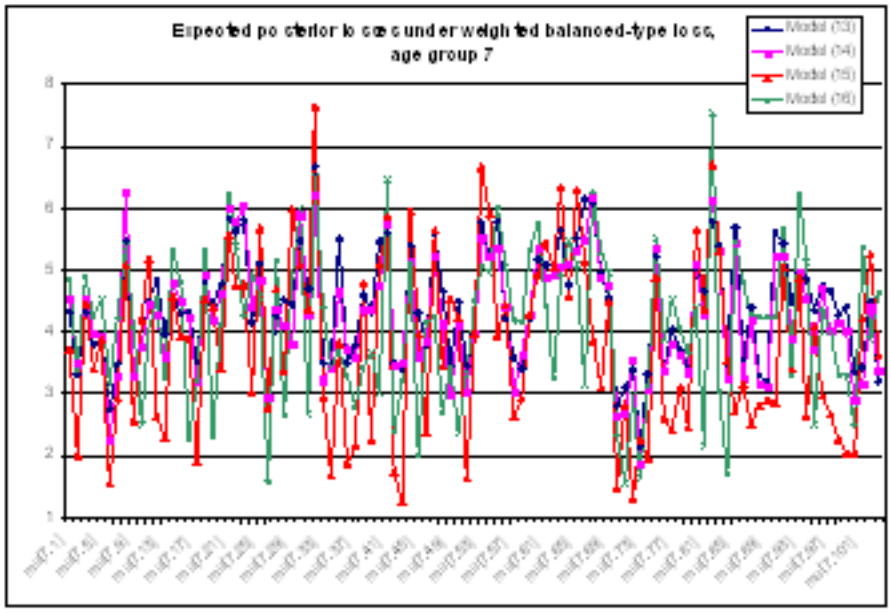


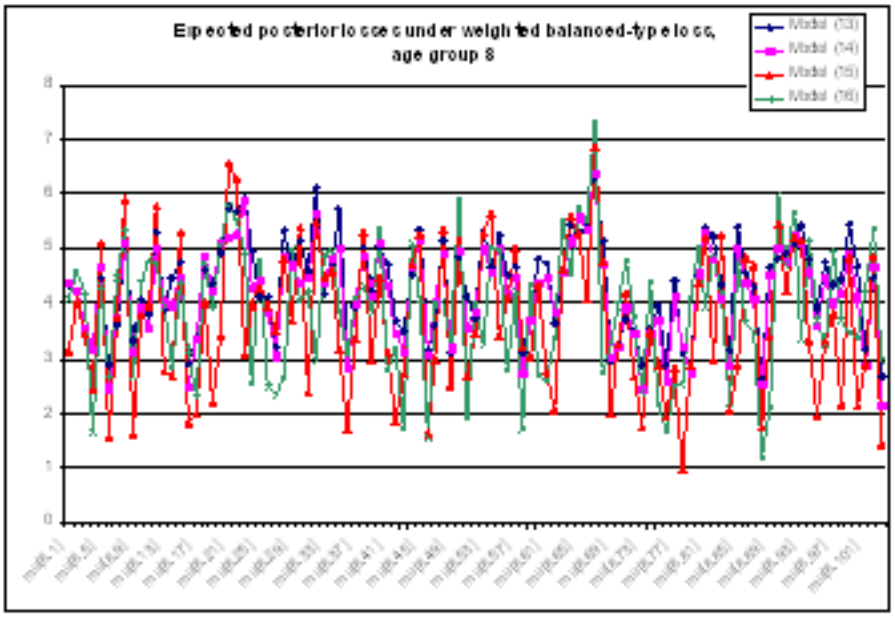

Expected posterior loss under weighted balanced-type loss, age group 3
Model (13)
Model (14)
Model (15)
Model (16)

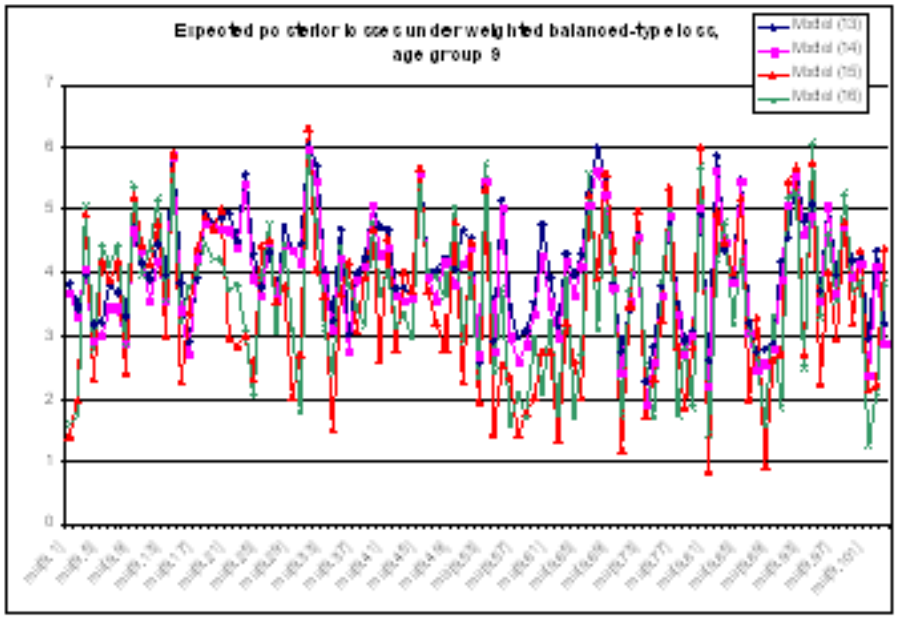

Expected posterior loss under weighted balanced-type loss, age group 4
Model (13)
Model (14)
Model (15)
Model (16)

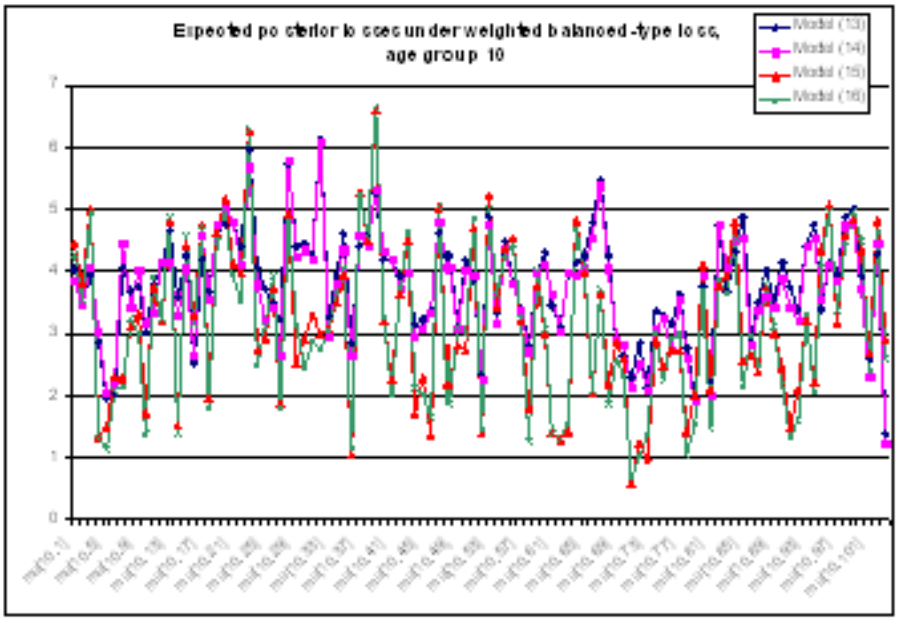

Expected posterior loss under weighted balanced-type loss, age group 5
Model (13)
Model (14)
Model (15)
Model (16)

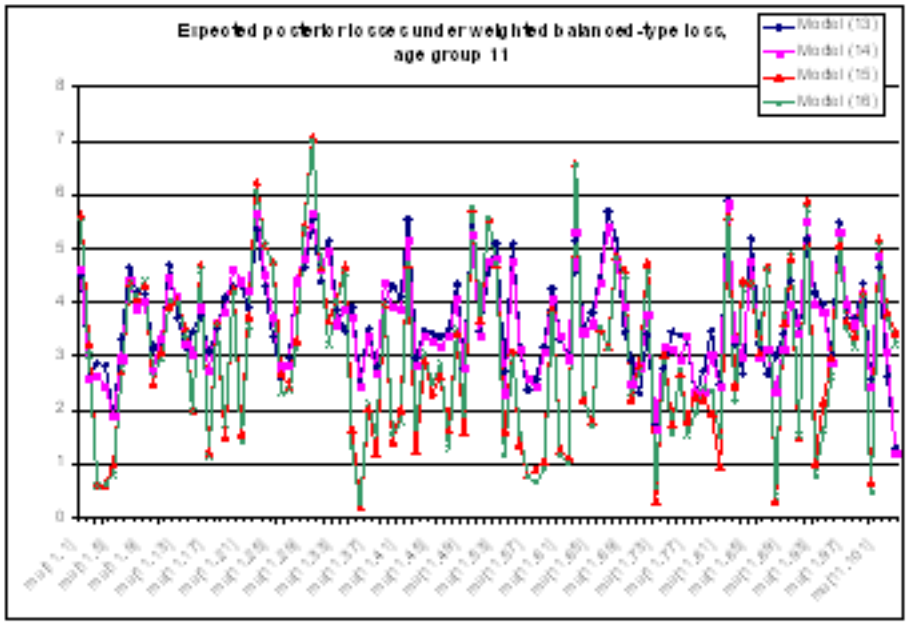

Expected posterior loss under weighted balanced-type loss, age group 6
Model (13)
Model (14)
Model (15)
Model (16)

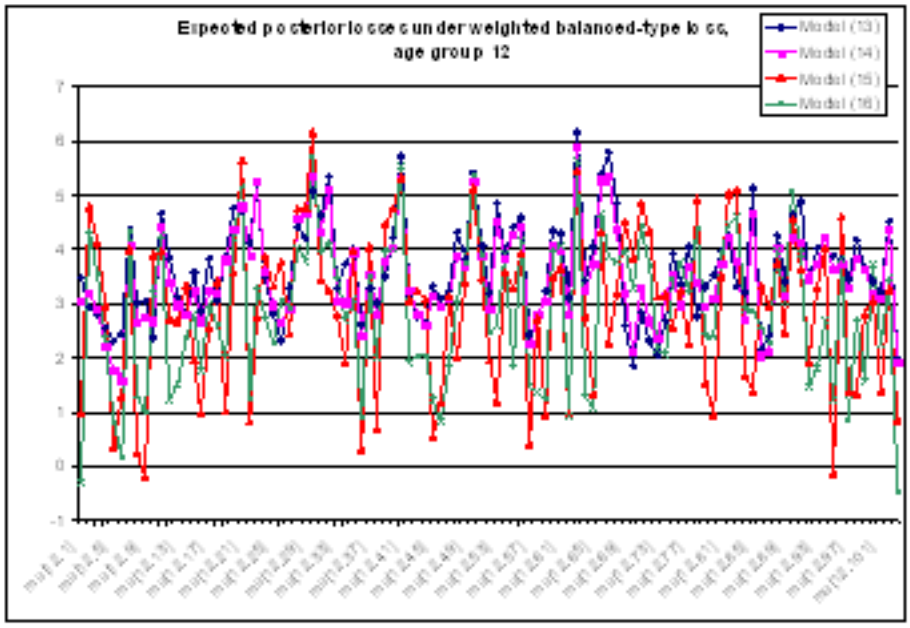

Expected posterior loss under weighted balanced-type loss, age group 7
Model (13)
Model (14)
Model (15)
Model (16)

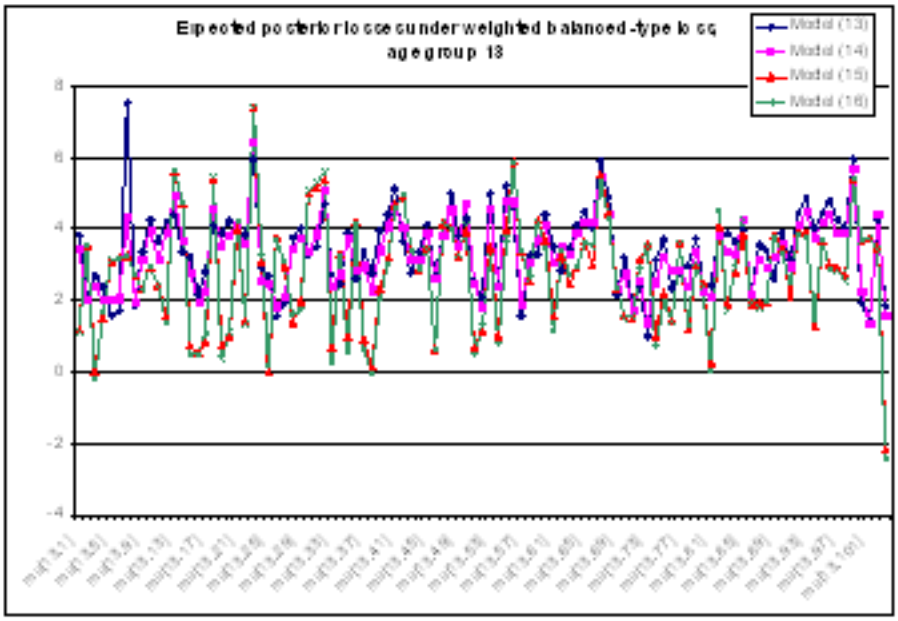

Expected posterior loss under weighted balanced-type loss, age group 8
Model (13)
Model (14)
Model (15)
Model (16)

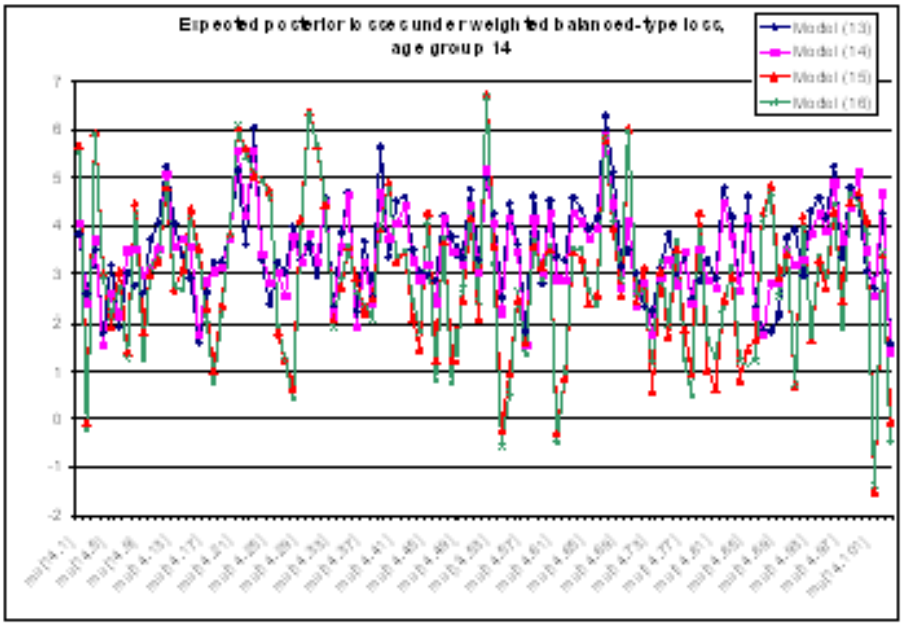

Expected posterior loss under weighted balanced-type loss, age group 9
Model (13)
Model (14)
Model (15)
Model (16)


Expected posterior loss under weighted balanced-type loss, age group 10
Model (13)
Model (14)
Model (15)
Model (16)


Expected posterior loss under weighted balanced-type loss, age group 11
Model (13)
Model (14)
Model (15)
Model (16)


Expected posterior loss under weighted balanced-type loss, age group 12
Model (13)
Model (14)
Model (15)
Model (16)


Expected posterior loss under weighted balanced-type loss,
age group 13
Model (13)
Model (14)
Model (15)
Model (16)


Expected posterior loss under weighted balanced-type loss,
age group 14
Model (13)
Model (14)
Model (15)
Model (16)