Stockholms universitet

# Generalized Bühlmann-Straub credibility theory for correlated data

Mikael Andblom

Masteruppsats i försäkringsmatematik
Master Thesis in Actuarial Mathematics

Masteruppsats 2023:2
Försäkringsmatematik
Februari 2023

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Master Thesis **2023:2**
http://www.math.su.se

# Generalized Bühlmann-Straub credibility theory for correlated data

Mikael Andblom*

February 2023

**Abstract**

In this thesis, we first go through classical results from the field of credibility theory. One of the most well-known models in the field is the Bühlmann-Straub model. The model is relatively straightforward to apply in practice and is widely used. A major advantage of the model is its simplicity and intuitive dependency on its model parameters. From our perspective, the main drawback is the assumption regarding uncorrelated data. We show that the correlation can be used to cancel observational noise and therefore obtain more accurate estimators. This leads to an extended credibility formula that contains the Bühlmann-Straub model as a special case. This comes at the cost of introducing singularities which may cause the estimators to behave unexpectedly under certain circumstances. Further research is needed to better understand how often the circumstances are met in practice and if transforming the optimal weights could be a way forward in such cases. Finally, through a simulation study based on real-world data, it is shown that the proposed model outperforms the Bühlmann-Straub model.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: mikaelandblom@hotmail.com. Supervisor: Filip Lindskog.

# Contents

# 1 Introduction

## 1.1 Key ratios and their estimators

In insurance, there are various key ratios that the actuary may need to consider. In this thesis, a key ratio observed at time $t$ for cohort $j$ is denoted by $\mu_{jt}$. The key ratio consists of a measurement, $x_{jt}$, as the numerator and a weight/volume, $w_{jt}$, as the denominator. Some common examples are the loss ratio (loss cost divided by premium), claim frequency (claim count divided by risk volume), claim severity (loss cost divided by claim count), and risk premium rate (expected loss cost divided by risk volume). Examples of risk volumes are turnover, number of insureds, sums insured, risk years, etc.

A cohort in this thesis refers to any group of insurance contracts. The cohort is free to be of any size, therefore it is also possible for a cohort to be a single contract, account, or client. In the context of credibility theory, the cohort may also be called group ([1] page 74) or risk ([2] page 81).

The key ratio can typically be aggregated in different levels of granularity. For instance, it can be aggregated by geographical region, insurance cover, customer segment, insurance product, various details about the insured, etc. The benefit of aggregating by wide groups is that the amount of data is larger and the volatility in the estimated key ratio will be reduced. On the other hand, using wide groups will assign less relevant key ratios to the individual cohorts in the group. In a pricing context, low relevancy implies a high likelihood of some individual cohorts cross-subsidizing others. This in turn can result in adverse selection and in a less robust pricing model. Less robust since the pricing would have to be revisited frequently to account for changes in the mix of the portfolio even if the underlying risk is unchanged.

It is possible to consider a key ratio being aggregated in a hierarchical structure. For a motor portfolio, the first level could be the car brand, the second level could be the country and the last level could be the sex of the driver. In this report, however, we will consider a two-level hierarchical structure. It will therefore be convenient to consider a portfolio-level key ratio $\mu$ and cohort-specific key ratios $\mu_j$.

It will be of interest to estimate both $\mu$ and $\mu_j$ using the estimators $\hat{\mu}$ and $\hat{\mu}_j$ respectively. The portfolio view and the individual cohort view are usually both useful on their own. As an example, the former can help to better understand the overall profitability while the latter could help find improvement areas or to adjust the risk appetite. Although one needs to keep in mind the corresponding drawbacks mentioned above.

A more stable and forward-looking view of the cohort-specific key ratio can be constructed by blending the two estimators $\hat{\mu}$ and $\hat{\mu}_j$ into $\hat{\mu}_j^b$ where

$$\hat{\mu}_j^b = z_j\hat{\mu}_j + (1 - z_j)\hat{\mu}, \tag{1}$$

and $z_j \in [0, 1]$ is the so called credibility factor. If $z_j = 1$ we say that we fully trust the cohort-specific estimate. On the other hand, for $z_j = 0$ we say that we fully trust the overall portfolio estimate.

Intuitively, factors that should influence the credibility are the uncertainty in the estimators $\hat{\mu}$ and $\hat{\mu}_j$, the homogeneity of the portfolio, and the correlation between the key ratios.

It is not necessarily the case that $\hat{\mu}_j^b$ should replace $\hat{\mu}_j$. One may still be interested in a purely experience-based and historical view. Furthermore, one may want to use expert judgment to manually tweak e.g. $z_j$ based on $\hat{\mu}_j$ and $\hat{\mu}$.

A summary of the available estimators can be seen in Table 1. The observed key ratio for cohort $j$ at time $t$ is denoted by $\mu_{jt}$. Note that we assume that the number of terms of data available for each cohort is the same. Thus we can have $T$ independent of $j$. This is sometimes referred to as the balanced model ([4] page 20) in credibility theory.

In this thesis, we will consider $\mu$ as unknown, and thus we are in need of the estimator $\hat{\mu}$. In some models, $\mu$ is considered known. A homogeneous credibility model includes an estimator for $\mu$ while an inhomogeneous credibility model does not ([2] pages 63-64).

As will be covered in more detail later, the Bühlmann-Straub model is the classical choice in credibility theory. It relies on assumptions regarding uncorrelated data. In the next section, we will motivate why this is not an appropriate assumption for us. The goal of this thesis can thus be said to extend the Bühlmann-Straub model to include correlated observations $\hat{\mu}_{jt}$.

Table 1: The set of estimators used to analyze the observed key ratios.

| Estimator | Target | View | Parameters | Data |
|---|---|---|---|---|
| $\hat{\mu}_j$ | $\mu_j$ | Historical | $\{a_{jt} \mid t \in [1, T]\}$ | $\{\mu_{jt} \mid t \in [1, T]\}$ |
| $\hat{\mu}$ | $\mu$ | Portfolio | $\{b_j \mid j \in [1, J]\}$ | $\{\hat{\mu}_j \mid j \in [1, J]\}$ |
| $\hat{\mu}_j^b$ | $\mu_j$ | Forward-looking | $z_j$ | $\{\hat{\mu}_j, \hat{\mu}\}$ |

### 1.2 Practical example: Travel insurance

To make the introduction more concrete we will in this subsection discuss some details in a practical example.

Consider a portfolio of $J$ payment cards each with a certain target customer group and travel insurance coverages. Each cohort thus corresponds to a particular payment card and the size of the cohort is the number of cardholders. Specifically, during underwriting year $t$ we have $w_{jt}$ cardholders that are covered by the $j$th card.

The task is to estimate the expected loss cost per cardholder, i.e. risk premium per cardholder, $\mu_j$. Since the price will be per cardholder the uncertainty in the number of future cardholders is not a problem. At least as long as we are able to construct accurate individual prices. As mentioned earlier, if the individual pricing is inaccurate we are subject to the risk of adverse selection.

The incurred losses are aggregated for each underwriting year and card which defines $c_{jt}$. The aggregated costs are divided by the number of cardholder years. The result is an observation $\mu_{jt} = c_{jt}/w_{jt}$ of the true risk premium $\mu_j$.

One aspect that most cards might have in common is the travel demand. Increased travel demand will likely increase the loss frequency (number of losses per cardholder year) across most cards. However, the increased travel demand might not impact the cards equally. Therefore, the measurements $\mu_{it}$ and $\mu_{jt}$ might be correlated and the correlation may be a function of $i$ and $j$. However, we assume no correlation across time, meaning $\mu_{it}$ and $\mu_{js}$ are independent for all $i$ and $j$ if $t \neq s$.

The correlation mentioned above extends the classical models used in credibility theory. This topic will be discussed in more technical detail within the literature study.

It is assumed that we a priori cannot tell the cards apart. In other words, we know that the cards are different but if we were to receive loss data without a way of identifying which losses belong to which card we could not price the cards more accurately than assigning the same risk premium per cardholder $\mu$ to all of $j$ cards. The risk premium $\mu$ may for example be estimated by

$$\hat{\mu} = \sum_{jt} c_{jt} \Bigg/ \sum_{jt} w_{jt}.$$

The assumption regarding a priori knowledge is important in the context of credibility theory. If we somehow knew that for a particular card $i$, the coverages were such that we should expect double the loss cost on average. Then we could transform the data points $\mu_{it}$ by dividing by 2 after which the a priori assumption can be assumed to hold. An even more trivial example would be if we knew that one card had its losses reported in a different currency. Without converting all losses into the same currency the portfolio would seem less homogeneous and therefore larger credibilities would be assigned to the individual estimators.

The problem can thus be summarized as follows. For the $j$th card - consider only losses from that card and construct a card-specific risk premium $\hat{\mu}_j$. Then estimate the underlying average $\mu$ by combining the estimates $\hat{\mu}_j$. The forward-looking risk premium is then constructed by blending the two risk premiums as in (1).

## 2 Mathematical framework

### 2.1 Problem formulation: The best unbiased estimator

Let $\hat{m}(\boldsymbol{\theta}, \boldsymbol{x})$ be an estimator of $m$ based on data $\boldsymbol{x}$ and parameter vector $\boldsymbol{\theta}$. The estimator is said to be unbiased if it has no expected error, i.e. $\mathbb{E}[\hat{m}(\boldsymbol{\theta}, \boldsymbol{x}) - m] = 0$. Additionally, the estimator is considered best if it minimizes the expected square error, i.e. the quantity $\mathbb{E}[(\hat{m}(\boldsymbol{\theta}, \boldsymbol{x}) - m)^2]$. Since $\mathrm{Var}(x) = \mathbb{E}[x^2] - \mathbb{E}[x]^2$ it is equivalent to say that the best unbiased estimator $\hat{m}(\boldsymbol{\theta}, \boldsymbol{x})$ minimizes the error variance, i.e. $\mathrm{Var}(\hat{m}(\boldsymbol{\theta}, \boldsymbol{x}) - m)$. We, therefore, define the best unbiased estimator as follows,

$$\hat{m}(\boldsymbol{x}) \coloneqq \hat{m}(\boldsymbol{\theta}^*, \boldsymbol{x}) : \boldsymbol{\theta}^* = \operatorname*{argmin}_{\boldsymbol{\theta}} \mathrm{Var}(\hat{m}(\boldsymbol{\theta}, \boldsymbol{x}) - m). \tag{2}$$

This definition will be used throughout to select model parameters.

Once we have designed unbiased estimators we will then have to consider the following three minimization problems.

$$\min_{\boldsymbol{a}_j} \mathrm{Var}(\hat{\mu}_j - \mu_j), \tag{3}$$

$$\min_{\boldsymbol{b}} \mathrm{Var}(\hat{\mu} - \mu), \tag{4}$$

$$\min_{z_j} \mathrm{Var}(\hat{\mu}_j^b - \mu_j). \tag{5}$$

Notice that the problems can be solved in any order and independently. This is because the estimators have no parameter in common as seen in Table 1. For instance, $a_{jt}$ is a parameter of $\hat{\mu}_j$ but not a parameter of $\hat{\mu}_j^b$ since for that estimator $\hat{\mu}_j$ is considered data and therefore fixed.

#### 2.1.1 The best credibility factor

Recall that $\hat{\mu}_j^b$ is an estimator of $\mu_j$. If we assume the estimator is unbiased we should, according to (5), select $z_j$ such that the error variance is minimized. Note that

$$\begin{aligned}\mathrm{Var}(\hat{\mu}_j^b - \mu_j) = {} & z_j^2 \mathrm{Var}(\hat{\mu}_j) + (1 - z_j)^2 \mathrm{Var}(\hat{\mu}) + \mathrm{Var}(\mu_j) \\ & + 2z_j(1 - z_j)\mathrm{cov}(\hat{\mu}_j, \hat{\mu}) - 2z_j \mathrm{cov}(\hat{\mu}_j, \mu_j) - 2(1 - z_j)\mathrm{cov}(\hat{\mu}, \mu_j).\end{aligned} \tag{6}$$

Differentiation yields

$$\begin{aligned}\frac{\partial}{\partial z_j} \mathrm{Var}(\hat{\mu}_j^b - \mu_j) = {} & 2z_j \mathrm{Var}(\hat{\mu}_j) - 2(1 - z_j)\mathrm{Var}(\hat{\mu}) \\ & + 2(1 - 2z_j)\mathrm{cov}(\hat{\mu}_j, \hat{\mu}) - 2\mathrm{cov}(\hat{\mu}_j, \mu_j) + 2\mathrm{cov}(\hat{\mu}, \mu_j).\end{aligned}$$

Rearranging by the credibility factor gives

$$\frac{\partial}{\partial z_j}\mathrm{Var}(\hat{\mu}_j^b - \mu_j) = 2z_j\left(\mathrm{Var}(\hat{\mu}_j) + \mathrm{Var}(\hat{\mu}) - 2\mathrm{cov}(\hat{\mu}_j,\, \hat{\mu})\right) \\ + 2\left(-\mathrm{Var}(\hat{\mu}) + \mathrm{cov}(\hat{\mu}_j,\, \hat{\mu}) - \mathrm{cov}(\hat{\mu}_j,\, \mu_j) + \mathrm{cov}(\hat{\mu},\, \mu_j)\right).$$

The root is given by

$$z_j = \frac{\mathrm{Var}(\hat{\mu}) - \mathrm{cov}(\hat{\mu}_j,\, \hat{\mu}) + \mathrm{cov}(\hat{\mu}_j,\, \mu_j) - \mathrm{cov}(\hat{\mu},\, \mu_j)}{\mathrm{Var}(\hat{\mu}) - 2\mathrm{cov}(\hat{\mu}_j,\, \hat{\mu}) + \mathrm{Var}(\hat{\mu}_j)}, \tag{7}$$

which is the general credibility factor for an arbitrary model.

Observe that $z_j$ converges to 1 and 0 asymptotically as $\mathrm{Var}(\hat{\mu})$ and $\mathrm{Var}(\hat{\mu}_j)$ tend to infinity respectively. This makes sense because as the uncertainty in one estimate increases we should put more trust into the other.

Additional properties of the credibility factor will be investigated once more concrete model assumptions have been presented.

## 2.2 Literature study

### 2.2.1 The random effects and Bühlmann-Straub models

In this section, we will give an introduction to classical results in credibility theory.

Generally speaking, a random effects model is a model with stochastic parameters. This type of model is well suited for studying credibility theory. This is because the first two moments of the individual cohort's key ratios can be seen as drawn from a common distribution. Thus the moments for cohort $j$ are determined by a random effect $V_j$. The moments can be seen as realized firstly and afterward the observable data. For this reason, the model is sometimes also referred to as a two-urn model.

It is not assumed which distribution the random effects belong to. Only their first two moments are assumed and they are independent and identically distributed. Following the notation used in [1] (pages 74-75) it is specifically assumed that

$$\begin{aligned} \mathbb{E}[V_j] &= \mu \\ \mathrm{Var}(V_j) &= \tau^2. \end{aligned} \tag{8}$$

When it comes to the observable key ratios it is only assumed that

$$\begin{aligned} \mathbb{E}[\mu_{jt} \mid V_j] &= V_j \\ \mathbb{E}\left[\mathrm{Var}(\mu_{jt} \mid V_j)\right] &= \sigma^2/w_{jt}. \end{aligned}$$

From the law of total variance, it follows that

$$\mathrm{Var}(\mu_{jt}) = \mathrm{Var}\left(\mathbb{E}[\mu_{jt} \mid V_j]\right) + \mathbb{E}\left[\mathrm{Var}(\mu_{jt} \mid V_j)\right] = \tau^2 + \sigma^2/w_{jt}. \tag{9}$$

In order to derive the credibility factor $z_j$ two sets of correlation assumptions need to be made.

Firstly, assume that the cohorts are independent, i.e.

$$\mathrm{cov}(V_i,\, V_j) = \mathrm{cov}(\mu_{it},\, V_j) = \mathrm{cov}(\mu_{it},\, \mu_{jt}) = 0, \tag{10}$$

for all $i \neq j$ and $t$. Secondly, the observed key ratios are conditionally independent, i.e.

$$\mathrm{cov}(\mu_{jt},\, \mu_{js} \mid V_j) = 0, \tag{11}$$

for all $t \neq s$ and $j$.

Under the above assumptions, one may derive the classical Bühlmann-Straub model, also known as the homogeneous credibility estimator ([2] page 89), which can be summarized as

$$\begin{cases} a_{jt}^B &= w_{jt}/w_{j*} \\ b_j^B &= z_j^B/z_*^B \\ z_j^B &= w_{j*}/(w_{j*} + \sigma^2/\tau^2), \end{cases} \tag{12}$$

where $*$ indicates summation over all elements in that dimension. E.g. $w_{j*} = \sum_j w_{ij}$.

The model parameters themselves are estimated using the following set of estimators ([1] page 78),

$$\begin{cases} \hat{\sigma}_j^2 &= \dfrac{1}{T-1} \displaystyle\sum_t w_{jt} \left(\mu_{jt} - \hat{\mu}_j\right)^2 \\ \hat{\sigma}^2 &= \dfrac{1}{T} \displaystyle\sum_j \hat{\sigma}_j^2 \\ \hat{\tau}^2 &= \dfrac{\sum_j w_{j*} \left(\hat{\mu}_j - \sum_i \frac{w_{i*}}{w_{**}} \hat{\mu}_i\right)^2 - (J-1)\hat{\sigma}^2}{w_{**} - \sum_j w_{j*}^2/w_{**}}. \end{cases} \tag{13}$$

**The suitability of the model for our purposes**

The Bühlmann-Straub model may be regarded as the traditional model for credibility theory. The model has been studied extensively and it is widely used. Using the model is fairly straightforward. The weights $a$ and $b$ are easy to compute. One could argue that there are only two parameters in the model, namely $\sigma$ and $\tau$, which is beneficial for the robustness of the model.

For our purposes, however, we will not be able to rely on the assumptions regarding uncorrelated cohorts. We are also interested in a more general variance structure. Although the latter is quite easily achieved already in the classical model.

Since the model is so widely used and known we will use it throughout for comparison with the proposed model.

### 2.2.2 Uncertainty-based credibility theory

An alternative formulation of credibility theory is discussed in [3]. We will later refer to the model as the Parodi-Bonche model. The model assumes the following structure

$$\begin{aligned}
\mu_j &= \mu + \tau\varepsilon_j \\
\hat{\mu}_j &= \mu_j + \sigma_{jj}\varepsilon'_j \\
\hat{\mu} &= \mu + s\varepsilon'.
\end{aligned}$$

The first equation describes the homogeneity of the portfolio. A smaller $\tau$ will imply a more homogeneous portfolio. I.e. the individual cohorts are more similar and their key ratios will be closer on average.

The last two equations describe the uncertainty in estimating the cohort-specific and portfolio key ratios respectively. The uncertainty in the estimates is measured by the standard deviations $\sigma_{jj}$ and $s$.

No distribution is assumed for the error terms. Instead, assumptions are made regarding the moments. Specifically

$$\begin{aligned}
\mathbb{E}[\varepsilon_j] = \mathbb{E}[\varepsilon'_j] = \mathbb{E}[\varepsilon'] &= 0 \\
\mathrm{Var}(\varepsilon_j) = \mathrm{Var}(\varepsilon'_j) = \mathrm{Var}(\varepsilon') &= 1 \\
\mathrm{cov}(\varepsilon_j,\, \varepsilon'_j) = \mathrm{cov}(\varepsilon_j,\, \varepsilon') &= 0 \\
\mathrm{cov}(\varepsilon'_j,\, \varepsilon') &= \rho_j.
\end{aligned}$$

Please note that $\varepsilon_j$ is conceptually different from $\varepsilon'_j$ and $\varepsilon'$. The former describes random noise when observing the key ratios. The latter also includes model and random errors which originates from constructing estimates based on a limited data set.

With the above assumptions we get

$$\begin{aligned}
\mathrm{Var}(\hat{\mu}) &= s^2 \\
\mathrm{Var}(\hat{\mu}_j) &= \tau^2 + \sigma_{jj}^2 \\
\mathrm{cov}(\hat{\mu}_j,\, \hat{\mu}) &= \sigma_{jj} s \rho_j \\
\mathrm{cov}(\hat{\mu}_j,\, \mu_j) &= \tau^2 \\
\mathrm{cov}(\hat{\mu},\, \mu_j) &= 0.
\end{aligned}$$

Insertion into (7) yields

$$z_j^P = \frac{s^2 - \sigma_{jj} s \rho_j + \tau^2}{s^2 - 2\sigma_{jj} s \rho_j + \tau^2 + \sigma_{jj}^2}, \tag{14}$$

which is the credibility factor presented in [3] (page 21).

#### The suitability of the model for our purposes

The uncertainty-based notation is straightforward to work with. It is also an improvement to the Bühlmann-Straub model for our purposes since it takes into account the correlation between the individual estimate and the portfolio estimate. The model also contains a more general variance structure.

However, the model is quite general and does not suggest how to actually construct the estimates $\hat{\mu}_j$ or $\hat{\mu}$ based on observed data. Therefore it does also not consider correlation in the observed key ratios $\mu_{jt}$.

In conclusion, the Parodi-Bonche model is well-suited as a foundation for our study.

### 2.2.3 Other literature on credibility theory with correlation

In [7] correlation is introduced to the classical Bühlmann-Straub model. Specifically, they let $\text{cov}(\mu_i, \mu_j) = \rho$ for all $i \neq j$. This is saying that the true cohort-specific key ratios are correlated. In this thesis, we are only interested in introducing a correlation between the observed data points. Hence we will have to consider other literature.

In [4] they try to capture correlation effects by introducing a common effect $\Lambda$ in addition to the random effects $V_j$. The common effect is similar to the random effect $V_j$ except that it is cohort-independent. It is assumed that $\mathbb{E}[\Lambda] = \mu$ and $\text{Var}(\Lambda) = \sigma_\lambda^2$. One way to summarize the impact of the assumed common effect is as follows

$$
\begin{aligned}
\mathbb{E}[\mu_{jt} \mid V_j, \Lambda] &= \mu(V_j, \Lambda) \\
\mathbb{E}[\mu(V_j, \Lambda) \mid \Lambda] &= \Lambda \\
\mathbb{E}[\text{Var}(\mu(V_j, \Lambda) \mid \Lambda)] &= \tau^2 \\
\mathbb{E}[\text{Var}(\mu_{jt} \mid V_j, \Lambda)] &= \sigma_{jt}^2.
\end{aligned}
$$

In the same paper, it is also shown that introducing a common effect like the above will only impact the inhomogeneous credibility estimator ([4] page 23). The homogeneous credibility estimator we are interested in is thus not impacted by a common effect.

## 2.3 Data model

We define the two-step data-generating model as

$$
\begin{aligned}
\mu_j &= \mu + \tau \varepsilon_j, \\
\mu_{jt} &= \mu_j + \sigma_{jt} \varepsilon_{jt},
\end{aligned}
\tag{15}
$$

where $\mu$, $\tau$ and $\sigma_{jt}$ are non-random scalars. The random variables $\varepsilon_j$ and $\varepsilon_{jt}$ are mutually uncorrelated with zero expectation and unit variance.

We do however assume that the correlation between $\varepsilon_{it}$ and $\varepsilon_{js}$ is non-zero for $t = s$. Specifically, let

$$\mathrm{cov}(\varepsilon_{it},\, \varepsilon_{js}) = \rho_{ij}\delta_{ts},$$

where $\delta_{ts}$ is the Kronecker delta and $\rho_{ij}$ is the correlation coefficient. The Kronecker delta is 1 if the indices are equal, otherwise 0. By definition we get $\rho_{jj} = 1$.

The entities $\mu$ and $\mu_j$ have already been covered in Section 1.1. The scalar $\tau$ relates to the homogeneity of the portfolio. Finally, the scalar $\sigma_{jt}$ relates to the noise in the observed data. It should generally be a decreasing function of the volume or weight $w_{jt}$.

Combining the two steps into one yield

$$\mu_{jt} = \mu + \tau\varepsilon_j + \sigma_{jt}\varepsilon_{jt}.$$

The homogeneity parameter $\tau$ is present in the variance of $\mu_{jt}$. In practice, the empirical variance will not include $\tau$ when looking at a single cohort. That is because when observing the key ratio for a particular cohort, the underlying true mean $\mu_j$ has already been realized. Therefore the observed variance will correspond to $\mathrm{Var}(\mu_{jt} \mid \mu_j) = \sigma_{jt}^2$.

If we let $\boldsymbol{\mu}$ be the vector with elements $\mu_j$ the law of total covariance gives us the following covariance tensor

$$\begin{aligned}
\bar{K}_{ij}^{ts} &:= \mathrm{cov}(\mu_{it},\, \mu_{js}) \\
&= \mathbb{E}[\mathrm{cov}(\mu_{it},\, \mu_{js} \mid \boldsymbol{\mu})] + \mathrm{cov}(\mathbb{E}[\mu_{it} \mid \boldsymbol{\mu}],\, \mathbb{E}[\mu_{it} \mid \boldsymbol{\mu}]) \\
&= \mathbb{E}[\mathrm{cov}(\sigma_{it}\varepsilon_{it},\, \sigma_{js}\varepsilon_{js}] + \mathrm{cov}(\mu_i,\, \mu_j) \\
&= \sigma_{it}\sigma_{js}\rho_{ij}\delta_{ts} + \tau^2\delta_{ij}.
\end{aligned} \tag{16}$$

### 2.3.1 Comparisson with the Bühlmann-Straub model

If we replace $V_j$ in the Bühlmann-Straub model with $\mu_j$ we can see that the two setups are consistent. For the random effect we have $\mathbb{E}[\mu_j] = \mu$ and $\mathrm{Var}(\mu_j) = \tau^2$ similarly to (8). For the observable key ratios, we also have $\mathrm{Var}(\mu_{jt}) = \tau^2 + \sigma_{jt}^2$ similarly to (9), but with a more general variance structure.

The correlation assumptions are however not all identical. The Bühlmann-Straub correlation assumptions can be mapped to assumptions regarding the error terms as follows

$$\begin{aligned}
\mathrm{cov}(V_i,\, V_j) = 0 &\iff \mathrm{cov}(\varepsilon_i,\, \varepsilon_j) = 0 \\
\mathrm{cov}(\mu_{it},\, V_j) = 0 &\iff \mathrm{cov}(\varepsilon_{it},\, \varepsilon_j) = 0 \\
\mathrm{cov}(\mu_{it},\, \mu_{jt}) = 0 &\quad\Rightarrow \mathrm{cov}(\varepsilon_{it},\, \varepsilon_{jt}) = 0 \\
\mathrm{cov}(\mu_{jt},\, \mu_{js} \mid V_j) = 0 &\iff \mathrm{cov}(\varepsilon_{jt},\, \varepsilon_{js}) = 0,
\end{aligned}$$

for all $i \neq j$ and $t \neq s$. The only assumption that is violated in the proposed model is the third one.

As already mentioned and similarly to the Parodi-Bonche model, we will later explicitly assume that there is a non-negligible correlation between $\hat{\mu}_j$ and $\hat{\mu}$. This is a consequence of the third correlation assumption in the previous equation. It is also a consequence of the fact that the two estimates share data.

### 2.4 Unbiased linear estimators

Weighted averages are an intuitive and straightforward approach to obtaining estimates of the key ratio. One way to construct the weighted averages is as follows

$$\begin{aligned} \hat{\mu}_j &= \sum_t a_{jt}\mu_{jt}, \\ \hat{\mu} &= \sum_j b_j \hat{\mu}_j. \end{aligned} \tag{17}$$

Note that these definitions satisfy the general setup in Table 1. By requiring that the weights sum up to one, i.e. $\sum_t a_{jt} = \sum_j b_j = 1$ the estimators become unbiased. For $\hat{\mu}_j$, this can be seen by expanding above:

$$\begin{aligned} \sum_t a_{jt}\mu_{jt} &= \sum_t a_{jt}(\mu_j + \sigma_{jt}\varepsilon_{jt}), \\ &= \mu_j + \sum_t a_{jt}\sigma_{jt}\varepsilon_{jt}, \end{aligned}$$

from which it follows that

$$\mathbb{E}[\hat{\mu}_j - \mu_j] = \mathbb{E}[\sum_t a_{jt}\sigma_{jt}\varepsilon_{jt}] = \sum_t a_{jt}\sigma_{jt}\mathbb{E}[\varepsilon_{jt}] = 0,$$

since $\varepsilon_{jt}$ has zero mean. Using the above we can also calculate the error variance as

$$\mathrm{Var}(\hat{\mu}_j - \mu_j) = \mathrm{Var}(\sum_t a_{jt}\sigma_{jt}\varepsilon_{jt}) = \sum_t a_{jt}^2\sigma_{jt}^2,$$

since $\varepsilon_{jt}$ and $\varepsilon_{js}$ are uncorrelated for $t \neq s$. For convenience, we define a more general relationship, namely,

$$\sigma_{ij}^2 \coloneqq \sum_t a_{it}a_{jt}\sigma_{it}\sigma_{jt}, \tag{18}$$

from which it follows that

$$\mathrm{Var}(\hat{\mu}_j - \mu_j) = \sigma_{jj}^2. \tag{19}$$

Note that $\sigma_{jj}^2$ is convex with respect to $a_{jt}$ since

$$\frac{\partial^2}{\partial a_{jt}^2}\sigma_{jj}^2 = 2\sigma_{jt}^2 > 0,$$

and therefore any local minimum is also the global minimum.

For $\hat{\mu}$ we have

$$\mathbb{E}[\hat{\mu}] = \sum_j b_j \mathbb{E}[\hat{\mu}_j] = \sum_j b_j \sum_t a_{jt} \mathbb{E}(\mu_{jt}) = \sum_j b_j \sum_t a_{jt} \mu = \mu,$$

which proves that the estimator is unbiased.

If we introduce a second covariance matrix $K$ with elements $K_{ij} := \mathrm{cov}(\hat{\mu}_i,\, \hat{\mu}_j)$ we may write the portfolio error variance as

$$\mathrm{Var}(\hat{\mu} - \mu) = \mathrm{Var}(\hat{\mu}) = \mathrm{Var}(\sum_j b_j \hat{\mu}_j) = \boldsymbol{b}^T K \boldsymbol{b} := s^2, \tag{20}$$

where $\boldsymbol{b}$ is a vector with elements $b_j$. Since $K$ is a covariance matrix it is positive definite. A quadratic form is convex if the corresponding matrix is positive definite. Therefore, any local minimum with respect to $b_j$ is also the global minimum.

Furthermore,

$$\begin{aligned}
K_{ij} &= \mathrm{cov}(\hat{\mu}_i,\, \hat{\mu}_j) \\
&= \mathrm{cov}\left(\sum_t a_{it}\mu_{it},\, \sum_t a_{jt}\mu_{jt}\right) \\
&= \sum_{ts} a_{it} a_{js} \mathrm{cov}(\mu_{it},\, \mu_{js}) \\
&= \sum_{ts} a_{it} a_{js} \bar{K}_{ij}^{ts} \\
&= \sum_{ts} a_{it} a_{js} (\sigma_{it}\sigma_{js}\rho_{ij}\delta_{ts} + \tau^2 \delta_{ij}), \\
&= \sigma_{ij}^2 \rho_{ij} + \tau^2 \delta_{ij}.
\end{aligned} \tag{21}$$

Note that we may equivalently write

$$K_{ij} = \boldsymbol{a}_i^T \bar{K}_{ij} \boldsymbol{a}_j,$$

where $\bar{K}_{ij}$ is a $T \times T$ covariance matrix with elements $\bar{K}_{ij}^{ts}$ and $\boldsymbol{a}_i$ is a $T$ dimensional column vector with elements $a_{it}$.

We will now investigate the correlation between the cohort-specific estimate and the portfolio estimate. We have

$$\begin{aligned}
\mathrm{cov}(\hat{\mu},\, \hat{\mu}_j) &= \mathrm{cov}(\sum_i b_i \hat{\mu}_i,\, \hat{\mu}_j) \\
&= \sum_i b_i \mathrm{cov}(\hat{\mu}_i,\, \hat{\mu}_j) \\
&= \sum_i b_i K_{ij}.
\end{aligned}$$

Using matrix notation, where $\upsilon_j := \text{cov}(\hat{\mu},\, \hat{\mu}_j)$ is the $j$th element in the column vector $\boldsymbol{\upsilon}$, we can write

$$\boldsymbol{\upsilon} = K\boldsymbol{b},$$

from which it follows that $s^2 = \boldsymbol{b}^T\boldsymbol{\upsilon}$.

The remaining entities needed for the credibility factor $z_j$ in (7) are $\text{cov}(\hat{\mu}_j,\, \mu_j)$, $\text{cov}(\hat{\mu},\, \mu_j)$ and $\text{Var}(\hat{\mu}_j)$. Note first that

$$\begin{aligned}\text{Var}(\hat{\mu}_j) &= \text{Var}(\mu_j + \sum_t a_{jt}\sigma_{jt}\varepsilon_{jt}) \\ &= \text{Var}(\mu + \tau\varepsilon_j + \sum_t a_{jt}\sigma_{jt}\varepsilon_{jt}) \\ &= \tau^2 + \sigma_{jj}^2,\end{aligned}$$

since $\varepsilon_j$ and $\varepsilon_{jt}$ are assumed uncorrelated. That $\varepsilon_j$ and $\varepsilon_{jt}$ are uncorrelated also imply that

$$\begin{aligned}\text{cov}(\hat{\mu}_j,\, \mu_j) &= \text{cov}(\mu_j + \sum_t a_{jt}\sigma_{jt}\varepsilon_{jt},\, \mu_j) \\ &= \text{cov}(\mu_j,\, \mu_j) \\ &= \tau^2.\end{aligned}$$

Finally,

$$\begin{aligned}\text{cov}(\hat{\mu},\, \mu_j) &= \text{cov}(\sum_i b_i\hat{\mu}_i,\, \mu_j) \\ &= \sum_i b_i \text{cov}(\hat{\mu}_i,\, \mu_j) \\ &= \sum_i b_i \text{cov}(\mu_i + \sum_t a_{it}\sigma_{it}\varepsilon_{it},\, \mu_j) \\ &= b_j\tau^2,\end{aligned}$$

since $\varepsilon_i$ and $\varepsilon_j$ are independent for $i \neq j$.

In summary, we now have

$$z_j = \frac{s^2 - \upsilon_j + \tau^2 - b_j\tau^2}{s^2 - 2\upsilon_j + \tau^2 + \sigma_{jj}^2}, \tag{22}$$

which is the credibility factor for any linear model of the form shown in (17).

Because $\upsilon_j = \sigma_{jj}s\rho_j$ in the Parodi-Bonche model the proposed credibility factor is very similar to the one in (14). The only difference is that we subtract the amount $b_j\tau^2$ from the numerator. This is because we do not assume $\text{cov}(\hat{\mu},\, \mu_j) = 0$.

### 2.5 The best unbiased linear estimators

We are now to select the weights $a_{jt}$ and $b_j$. The best choice is defined by (2).

By combining (3) and (19) we have

$$\begin{aligned} \min_{\boldsymbol{a}_j} \quad & \sigma_{jj}^2 \\ \text{s.t} \quad & \sum_t a_{jt} = 1, \end{aligned}$$

The corresponding Lagrangian becomes

$$\mathcal{L}(\boldsymbol{a}_j,\, \lambda) = \sum_t a_{jt}^2 \sigma_{jt}^2 - \lambda(\sum_t a_{jt} - 1),$$

which has the partial derivative

$$\frac{\partial}{\partial a_{jt}} \mathcal{L}(\boldsymbol{a}_j,\, \lambda) = 2 a_{jt} \sigma_{jt}^2 - \lambda.$$

The root is given by

$$a_{jt} = \frac{\lambda}{2\sigma_{jt}^2}.$$

The normal condition becomes

$$\frac{\lambda}{2} \sum_t \frac{1}{\sigma_{jt}^2} = 1.$$

The multiplier can thus be written as

$$\lambda = \frac{2}{\sum_t \frac{1}{\sigma_{jt}^2}},$$

which yields

$$a_{jt} = \frac{1}{\sigma_{jt}^2 \sum_s \frac{1}{\sigma_{js}^2}}. \tag{23}$$

This result is the so-called inverse-variance weighting which is sometimes used when aggregating multiple independent measurements measuring the same underlying entity. In our case, $\mu_{jt}$ for all $t$ can be considered an independent measurement of $\mu_j$. Finally, note that since the variance $\sigma_{jt}^2 > 0$ it follows that $0 < a_{jt} \leq 1$.

The next step is to select $\boldsymbol{b}$. By combining (4) and (20) we have

$$\begin{aligned} \min_{\boldsymbol{b}} \quad & s^2 \\ \text{s.t} \quad & \boldsymbol{u}^T \boldsymbol{b} = 1, \end{aligned}$$

where $\boldsymbol{u}$ is a column vector of ones. Note that the inverse-variance weighting does not apply here since the measurements $\hat{\mu}_j$ of $\mu$ are not independent. However, we can still apply the method of Lagrange multipliers.

Using (21) we have the following Lagrangian

$$\mathcal{L}(\boldsymbol{b},\, \lambda) = \boldsymbol{b}^T K \boldsymbol{b} - \lambda(\boldsymbol{u}^T \boldsymbol{b} - 1),$$

with gradient

$$\nabla_{\boldsymbol{b}}\, \mathcal{L}(\boldsymbol{b},\, \lambda) = 2K\boldsymbol{b} - \lambda \boldsymbol{u},$$

which is solved by

$$\boldsymbol{b} = \frac{\lambda}{2} K^{-1} \boldsymbol{u}.$$

Inserting into the normal condition yields

$$\frac{\lambda}{2} \boldsymbol{u}^T K^{-1} \boldsymbol{u} = 1,$$

which means that the multiplier must satisfy

$$\lambda = \frac{2}{\boldsymbol{u}^T K^{-1} \boldsymbol{u}}.$$

The weights are therefore calculated as

$$\boldsymbol{b} = \frac{K^{-1}\boldsymbol{u}}{\boldsymbol{u}^T K^{-1} \boldsymbol{u}}.$$

Since $\boldsymbol{v} = K\boldsymbol{b}$ it follows that

$$\boldsymbol{v} = \frac{\boldsymbol{u}}{\boldsymbol{u}^T K^{-1} \boldsymbol{u}},$$

i.e. $\boldsymbol{v}$ is a column vector with all elements equal to $1/K^{-1}_{**}$.

Consequently, since $s^2 = \boldsymbol{b}^T \boldsymbol{v}$ we have

$$s^2 = \frac{\boldsymbol{u}^T K^{-1} \boldsymbol{u}}{(\boldsymbol{u}^T K^{-1} \boldsymbol{u})^2} = \frac{1}{\boldsymbol{u}^T K^{-1} \boldsymbol{u}} = 1/K^{-1}_{**}, \tag{24}$$

from which we get

$$b_j = \frac{K^{-1}_{j*}}{K^{-1}_{**}}. \tag{25}$$

We also get $v_j = s^2$, or $\mathrm{cov}(\hat{\mu}_j,\, \hat{\mu}) = \mathrm{Var}(\hat{\mu})$. This means that we have

$$z_j = \frac{(1 - b_j)\tau^2}{\tau^2 + \sigma^2_{jj} - s^2} = \frac{\mathrm{cov}(\hat{\mu}_j - \hat{\mu},\, \mu_j)}{\mathrm{Var}(\hat{\mu}_j - \hat{\mu})} \tag{26}$$

The proposed model can now be summarized as follows.

$$\begin{cases} a_{jt} &= \sigma_{jt}^{-2} / \sum_s \sigma_{js}^{-2} \\ \sigma_{ij}^2 &= \sum_t a_{it} a_{jt} \sigma_{it} \sigma_{jt} \\ K_{ij} &= \sigma_{ij}^2 \rho_{ij} + \tau^2 \delta_{ij} \\ s^2 &= 1/K_{**}^{-1} \\ b_j &= K_{j*}^{-1} s^2 \\ z_j &= (1 - b_j)\tau^2 / (\tau^2 + \sigma_{jj}^2 - s^2). \end{cases} \tag{27}$$

Note that there is a singularity in the credibility factor at $\tau^2 + \sigma_{jj}^2 = s^2$. This is more easily understood in (26) from which we see that the singularity occurs when $\hat{\mu}_j = \hat{\mu}$. At this point, the blended estimate becomes $\hat{\mu}_j^b = \hat{\mu}$ for any $z_j$ and therefore there is no single optimal value for $z_j$.

## 2.6 Special case: The Bühlmann-Straub assumptions

In this section, we will study the special case when

$$\begin{cases} \sigma_{jt}^2 = \sigma^2 / w_{jt} \\ \rho_{ij} = \delta_{ij}, \end{cases} \tag{28}$$

which means that there is a cohort-independent base variance $\sigma^2$ and that the observations are independent.

Using (23) we have

$$\begin{aligned} a_{jt} &= \frac{1}{\sigma_{jt}^2 \sum_s \frac{1}{\sigma_{js}^2}} \\ &= \frac{1}{\frac{\sigma^2}{w_{jt}} \sum_s \frac{w_{js}}{\sigma^2}} \\ &= \frac{w_{jt}}{w_{j*}} \\ &= a_{jt}^B. \end{aligned} \tag{29}$$

Observe that this identity also holds for the more general case $\sigma_{jt} = \sigma_j^2 / w_{jt}$.

Furthermore, (18) gives us

$$\begin{aligned}
\sigma_{jj}^2 &= \sum_t a_{jt}^2 \sigma_{jt}^2 \\
&= \sum_t \frac{w_{jt}^2}{w_{j*}^2} \frac{\sigma^2}{w_{jt}} \\
&= \frac{\sigma^2}{w_{j*}^2} \sum_t w_{jt} \\
&= \frac{\sigma^2}{w_{j*}}.
\end{aligned} \tag{30}$$

The covariance matrix $K$ in (21) becomes a diagonal matrix with elements

$$\begin{aligned}
K_{jj} &= \tau^2 + \sigma_{jj}^2 \\
&= \tau^2 + \frac{\sigma^2}{w_{j*}} \\
&= \frac{w_{j*}\tau^2 + \sigma^2}{w_{j*}} \\
&= \tau^2 \frac{w_{j*} + \sigma^2/\tau^2}{w_{j*}} \\
&= \frac{\tau^2}{z_j^B}.
\end{aligned}$$

We also get

$$z_j^B = \frac{\tau^2}{\tau^2 + \sigma_{jj}^2}.$$

The inverse of a diagonal matrix is a diagonal matrix with elements equal to the reciprocal of the original diagonal. The inverted covariance matrix is therefore diagonal with elements

$$K_{jj}^{-1} = \frac{z_j^B}{\tau^2}.$$

(24) yields

$$s^2 = 1/K_{**}^{-1} = \frac{\tau^2}{z_*^B},$$

while (25) yields

$$b_j = K_{j*}^{-1}/K_{**}^{-1} = \frac{z_j^B}{z_*^B} = b_j^B.$$

Note that we can write

$$s^2 = \frac{\tau^2}{z_j^B} b_j = (\tau^2 + \sigma_{jj}^2) b_j,$$

which in turn gives

$$z_j = \frac{(1-b_j)\tau^2}{\tau^2 + \sigma_{jj}^2 - s^2} = \frac{(1-b_j)\tau^2}{(\tau^2 + \sigma_{jj}^2)(1-b_j)} = \frac{\tau^2}{\tau^2 + \sigma_{jj}^2} = z_j^B.$$

We have thus shown that the proposed model simplifies to the Bühlmann-Straub model in (12) if the assumptions in (28) are satisfied.

### 2.7 Special case: Two cohorts with time-constant weights

Finding the inverse of the covariance matrix $K$ in the general case is a non-trivial task. However, finding the inverse of a general two-by-two matrix is straightforward. Therefore we will in this subsection consider two cohorts and for additional simplicity, we will assume the weights are constant in time and that the cohorts have the same base variance. The assumptions correspond to $J = 2$, $w_{jt} = w_j$ and $\sigma_{jt}^2 = \sigma^2 / w_j$.

From (29) it follows that $a_{jt} = 1/T$ and consequently,

$$\begin{aligned}
\sigma_{ij}^2 &= \sum_t a_{it} a_{jt} \sigma_{it} \sigma_{jt} \\
&= \sum_t \frac{\sigma^2}{T^2 w_i w_j} \\
&= \frac{\sigma}{T w_i} \cdot \frac{\sigma}{T w_j} := \sigma_i \sigma_j.
\end{aligned}$$

The covariance matrix thus becomes

$$K = \begin{bmatrix} \sigma_1^2 + \tau^2 & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 + \tau^2 \end{bmatrix},$$

where $\rho = \rho_{12} = \text{cov}(\varepsilon_{1t},\, \varepsilon_{2t})$. The inverse is given by

$$K^{-1} = \begin{bmatrix} \sigma_2^2 + \tau^2 & -\sigma_1 \sigma_2 \rho \\ -\sigma_1 \sigma_2 \rho & \sigma_1^2 + \tau^2 \end{bmatrix} \cdot \frac{1}{(\sigma_1^2 + \tau^2)(\sigma_2^2 + \tau^2) - \sigma_1^2 \sigma_2^2 \rho^2}.$$

From above it follows that

$$s^2 = 1/K_{**}^{-1} = \frac{(\sigma_1^2 + \tau^2)(\sigma_2^2 + \tau^2) - \sigma_1^2 \sigma_2^2 \rho^2}{\sigma_1^2 + \sigma_2^2 + 2\tau^2 - 2\sigma_1 \sigma_2 \rho}.$$

For the first cohort, we have

$$1 - b_1 = b_2 = K_{2*}^{-1} s^2 = \frac{\sigma_1^2 + \tau^2 - \sigma_1 \sigma_2 \rho}{\sigma_1^2 + \sigma_2^2 + 2\tau^2 - 2\sigma_1 \sigma_2 \rho}, \tag{31}$$

which gives the credibility factor

$$
\begin{aligned}
z_1 &= \frac{\sigma_1^2+\tau^2-\sigma_1\sigma_2\rho}{\sigma_1^2+\sigma_2^2+2\tau^2-2\sigma_1\sigma_2\rho}\tau^2 \Big/ \left(\tau^2+\sigma_1^2-\frac{(\sigma_1^2+\tau^2)(\sigma_2^2+\tau^2)-\sigma_1^2\sigma_2^2\rho^2}{\sigma_1^2+\sigma_2^2+2\tau^2-2\sigma_1\sigma_2\rho}\right)\\
&= \frac{(\sigma_1^2+\tau^2-\sigma_1\sigma_2\rho)\tau^2}{(\sigma_1^2+\tau^2)(\sigma_1^2+\sigma_2^2+2\tau^2-2\sigma_1\sigma_2\rho)-(\sigma_1^2+\tau^2)(\sigma_2^2+\tau^2)+\sigma_1^2\sigma_2^2\rho^2}.
\end{aligned}
$$

If we introduce a generalized credibility coefficient

$$
\kappa_i := \frac{\sigma_i^2}{\tau^2} = \frac{\sigma^2}{Tw_i\tau^2},
$$

above simplifies into

$$
\begin{aligned}
z_1 &= \frac{\kappa_1+1-\sqrt{\kappa_1\kappa_2}\rho}{(\kappa_1+1)(\kappa_1+\kappa_2+2-2\sqrt{\kappa_1\kappa_2}\rho)-(\kappa_1+1)(\kappa_2+1)+\kappa_1\kappa_2\rho^2}\\
&= \frac{\kappa_1+1-\sqrt{\kappa_1\kappa_2}\rho}{(\kappa_1+1)^2-2(\kappa_1+1)\sqrt{\kappa_1\kappa_2}\rho+\kappa_1\kappa_2\rho^2}\\
&= \frac{\kappa_1+1-\sqrt{\kappa_1\kappa_2}\rho}{(\kappa_1+1-\sqrt{\kappa_1\kappa_2}\rho)^2}\\
&= \frac{1}{1+\kappa_1-\sqrt{\kappa_1\kappa_2}\rho}.
\end{aligned}
\tag{32}
$$

Note that $\kappa_i$ is not the same credibility coefficient as in the Bühlmann-Straub model where $\kappa = \sigma^2/\tau^2$. The generalization is consistent with the standard credibility coefficient. Intuitively, increasing $T$ or $w_j$ should increase the credibility factor $z_j$ and therefore reduce the credibility coefficient $\kappa_j$.

Again, we should get back the Bühlmann-Straub model if $\rho = 0$, which indeed is the case since

$$
z_1(\rho=0) = \frac{1}{1+\kappa_1} = \frac{Tw_1}{Tw_1+\sigma^2/\tau^2}.
$$

It is interesting to note that for the Bühlmann-Straub model, $z_1$ is only dependent on $\kappa_1$ and nothing else. While for the proposed model both $\kappa_2$ and $\rho$ plays a role.

If $k_2$ is sufficiently large, specifically if $\kappa_2 \geq (1+\kappa_1)^2/\kappa_1$, then there exists a singularity in the credibility factor in (32) at

$$
\rho^* = \frac{1+\kappa_1}{\sqrt{\kappa_1\kappa_2}},
$$

and

$$
\begin{cases} z_1 > 0 \text{ for } \rho < \rho^* \\ z_1 < 0 \text{ for } \rho > \rho^*. \end{cases}
$$

The question is what to do when $z_1 > 1$ or $z_1 < 0$. First we realize that $\rho$ has no impact on $\hat{\mu}_j$ but only on $\hat{\mu}$. Thus we are interested in learning how the accuracy of the estimator $\hat{\mu}$ depends on $\rho$. For this purpose, consider the scaled error variance

$$\frac{s^2}{\tau^2} = \frac{(\kappa_1+1)(\kappa_2+1) - \kappa_1\kappa_2\rho^2}{\kappa_1+\kappa_2+2-2\sqrt{\kappa_1\kappa_2}\rho}, \tag{33}$$

which is concave and maximized at $\rho = \rho^*$ since

$$\begin{aligned}
&\frac{\partial}{\partial\rho}\frac{s^2}{\tau^2}\bigg|_{\rho=\rho^*} = \\
&= \frac{-2\kappa_1\kappa_2\rho(\kappa_1+\kappa_2+2-2\sqrt{\kappa_1\kappa_2}\rho) - ((\kappa_1+1)(\kappa_2+1)-\kappa_1\kappa_2\rho^2)(-2\sqrt{\kappa_1\kappa})}{(\kappa_1+\kappa_2+2-2\sqrt{\kappa_1\kappa_2}\rho)^2} \\
&= \frac{2\kappa_1\kappa_2\rho(\sqrt{\kappa_1\kappa_2}\rho - (\kappa_1+1) - (\kappa_2+1)) + 2\sqrt{\kappa_1\kappa_2}(\kappa_1+1)(\kappa_2+1)}{(\kappa_1+\kappa_2+2-2\sqrt{\kappa_1\kappa_2}\rho)^2} \\
&= \frac{-2\sqrt{\kappa_1\kappa_2}(\kappa_1+1)(\kappa_2+1) + 2\sqrt{\kappa_1\kappa_2}(\kappa_1+1)(\kappa_2+1)}{((\kappa_2+1)-(\kappa_1+1))^2} \\
&= 0.
\end{aligned}$$

Therefore, the estimator $\hat{\mu}$ will be least accurate when $\rho = \rho^*$ and then gain in accuracy as $\rho$ moves away, in either direction, from $\rho^*$. Hence one might want to consider the following transformation of the credibility factor

$$\tilde{z}_i = \max(\min(z_i,\, 1),\, 0). \tag{34}$$

### 2.7.1 Numerical example

Since $w_i$ is a free variable it in turn means that both $\kappa_1$ and $\kappa_2$ are free. The covariance $\rho$ is also free. This means that we are able to vary the three variables in (32) freely and independently. The only restrictions are $\kappa_i > 0$ and $|\rho| \leq 1$.

As an example, we may consider $\kappa_1 = 1$ and $\kappa_2 = 16$ and investigate what happens as we vary $\rho$ from -1 to 1. This means that the singularity will occur at $\rho^* = 1/2$. It also means that $\kappa_2/\kappa_1 = w_1/w_2 = 16$ and that the first cohort has significantly more exposure than the second.

For comparison, we will consider the Bühlmann-Straub model where

$$\begin{cases}
z_1^B &= \dfrac{1}{1+\kappa_1} = 1/2 \\
z_2^B &= \dfrac{1}{1+\kappa_2} = 1/17 \\
b_1^B &= z_1^B/(z_1^B + z_2^B) = 17/19 \\
b_2^B &= 1 - b_1^B = 2/19
\end{cases}$$

As mentioned earlier, the accuracy of $\hat{\mu}_j$ is independent of $\rho$. Therefore, we will start by considering the accuracy of $\hat{\mu}$. Since $\hat{\mu}$ is unbiased the expected square error of the estimator is given by its error variance $s^2$. To get an expression in the credibility coefficients we will consider the scaled expected square error $s^2/\tau^2$. We therefore define

$$h_s(\rho) := \frac{1}{\tau^2}\mathbb{E}[(\hat{\mu} - \mu)^2] = s^2/\tau^2 = \boldsymbol{b}^T \left(K/\tau^2\right) \boldsymbol{b} = \boldsymbol{b}^T \begin{bmatrix} \kappa_1 + 1 & \sqrt{\kappa_1\kappa_2}\rho \\ \sqrt{\kappa_1\kappa_2}\rho & \kappa_2 + 1 \end{bmatrix} \boldsymbol{b}. \tag{35}$$

For the proposed model we have (33). For the Bühlmann-Straub model we can use

$$h_s^B(\rho) = \begin{bmatrix} 17/19 \\ 2/19 \end{bmatrix}^T \begin{bmatrix} 2 & 4\rho \\ 4\rho & 17 \end{bmatrix} \begin{bmatrix} 17/19 \\ 2/19 \end{bmatrix}.$$

The result has been visualized in Figure 1. It is clear that the proposed model has a lower expected square error for all $\rho$, ignoring the fact that the models are identical at $\rho = 0$. As expected, we see that for the proposed model the expected error is concave and maximized at $\rho = \rho^*$.

We will now move on to the credibility factor itself. We will only consider the first credibility factor. For the Bühlmann-Straub model, there is no dependency on $\rho$ and $z_1^B = 1/2$. For the proposed model we can use (32) and (34). The credibility factors can be viewed in Figure 2. We again see that the models intersect at $\rho = 0$. The singularity is clearly visible at $\rho = \rho^*$.

Finally, we will consider the scaled expected square error of $\hat{\mu}_j$. For simplicity, we will only consider the first cohort. We do this by first introducing

$$\begin{aligned} h_\sigma(\rho) &:= \frac{1}{\tau^2}\mathbb{E}[(\hat{\mu}_1 - \mu_1)^2] \\ &= z_1^2(\kappa_1 + 1) + (1 - z_1)^2 s^2/\tau^2 + 1 + 2z_1(1 - z_1)\nu_1/\tau^2 - 2z_1 - 2(1 - z_1)b_1 \end{aligned} \tag{36}$$

which follows from (6). To be able to evaluate this expression for the two models we need to calculate $\nu_1/\tau^2$. For the proposed model we have $\nu_1 = s^2$. For the Bühlmann-Straub model, we can use

$$\nu_1/\tau^2 = \boldsymbol{e}_1^T \left(K/\tau^2\right) \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}^T \begin{bmatrix} 2 & 4\rho \\ 4\rho & 17 \end{bmatrix} \begin{bmatrix} 17/19 \\ 2/19 \end{bmatrix},$$

which follows from $\boldsymbol{\nu} = K\boldsymbol{b}$. For the proposed model we also need to evaluate $b_1$ which we can do by slightly modifying (31) into

$$b_1 = \frac{\kappa_2 + 1 - \sqrt{\kappa_1\kappa_2}\rho}{\kappa_1 + \kappa_2 + 2 - 2\sqrt{\kappa_1\kappa_2}\rho}.$$

The result is presented in Figure 3 where we have also included the modified credibility factor which includes a floor and a ceiling. It is clear that the modified credibility factor performs worse than the unmodified one which should not be surprising considering it is a deviation from the optimum. It is however promising that the modified credibility factor is expected to perform considerably better than the Bühlmann-Straub model for strong positive correlations.

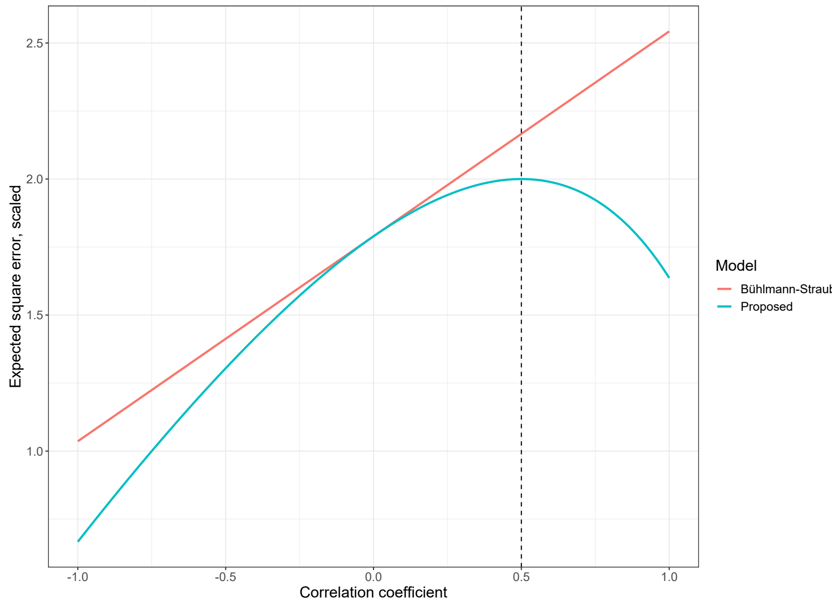


Figure 1: Scaled expected square error of $\hat{\mu}$, see (35), as a function of $\rho$ for the two models. A vertical line is drawn at $\rho = \rho^*$.

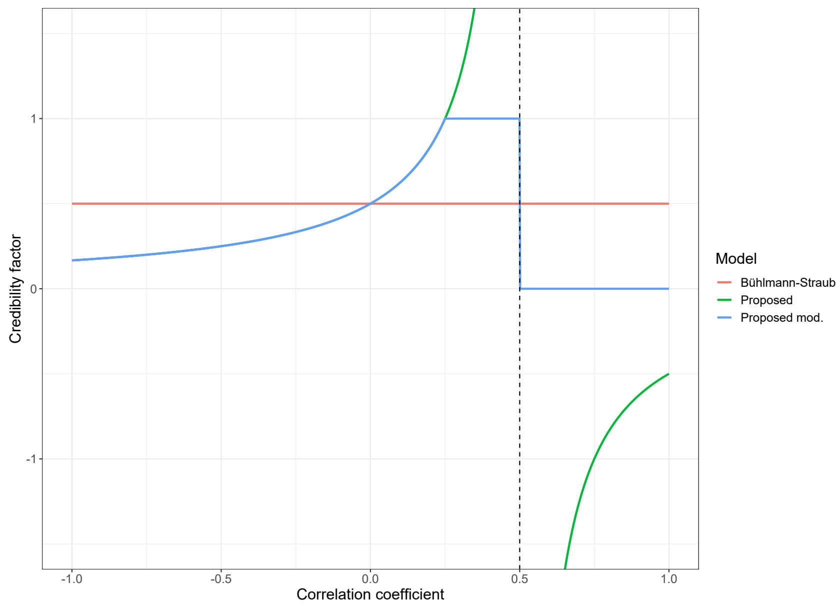


Figure 2: A comparison of the three candidates for the credibility factor. A vertical line is drawn at $\rho = \rho^*$.

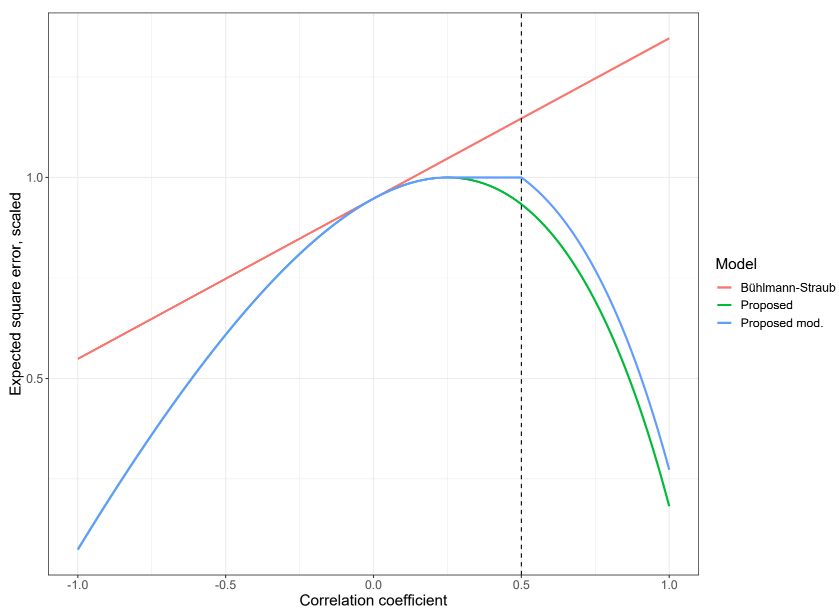


Figure 3: Scaled expected square error of $\hat{\mu}_j$, see (36), as a function of $\rho$ for the three models. A vertical line is drawn at $\rho = \rho^*$.

### 2.8 Parameter estimation

In this section, we will construct unbiased estimators of the model parameters. We will do this for the special case when

$$\sigma_{jt}^2 = \frac{\sigma_j^2}{w_{jt}}.$$

The special case is similar to the Bühlmann-Straub model but with the extension that the base variance depends on the cohort. It is also an extension due to that $\rho_{ij} \neq \delta_{ij}$ in general.

By slightly modifying (29) we see that the more general special case still implies $a_{jt} = w_{jt}/w_{j*}$. Since the weights $w_{jt}$ are known it follows that the weights $a_{jt}$ are also known.

For the covariance matrix $K$ we will need the product $\sigma_{ij}^2\rho_{ij}$ as seen in (21). By using (18) we can see that

$$\begin{aligned}\sigma_{ij}^2\rho_{ij} &= \sum_t a_{it}a_{jt}\sigma_{it}\sigma_{jt}\rho_{ij} \\ &= \rho_{ij}\sum_t \frac{w_{it}}{w_{i*}}\frac{w_{jt}}{w_{j*}}\frac{\sigma_i}{\sqrt{w_{it}}}\frac{\sigma_j}{\sqrt{w_{jt}}} \\ &= \frac{\sigma_i\sigma_j\rho_{ij}}{w_{i*}w_{j*}}\sum_t \sqrt{w_{it}w_{jt}}.\end{aligned}$$

Because the weights $w_{jt}$ are known, the above relation tells us that if we can find an unbiased estimator for the product $\sigma_i\sigma_j\rho_{ij}$ then we also have an unbiased estimator for $\sigma_{ij}^2\rho_{ij}$. We may therefore write

$$\widehat{\sigma_{ij}^2\rho_{ij}} = \frac{\sum_t \sqrt{w_{it}w_{jt}}}{w_{i*}w_{j*}}\,\widehat{\sigma_i\sigma_j\rho_{ij}}. \tag{37}$$

Note that we only need to construct the unbiased estimators $\widehat{\sigma_{ij}^2\rho_{ij}}$ and $\widehat{\tau^2}$. This can be seen by inspecting (27) and noting that $\sigma_{jj}^2$ can be written as $\sigma_{jj}^2\rho_{jj}$. I.e. we do not need to estimate $\sigma_{jj}^2$ separately.

We will start by searching for the unbiased estimator $\widehat{\sigma_{ij}^2\rho_{ij}}$. Note first that

$$\begin{aligned}\mathbb{E}[(\mu_{it}-\hat{\mu}_i)(\mu_{jt}-\hat{\mu}_j)] &= \mathrm{cov}(\mu_{it}-\hat{\mu}_i,\, \mu_{jt}-\hat{\mu}_j) + \mathbb{E}[\mu_{it}-\hat{\mu}_i]\mathbb{E}[\mu_{jt}-\hat{\mu}_j] \\ &= \underbrace{\mathrm{cov}(\mu_{it},\, \mu_{jt})}_{1} - \underbrace{\mathrm{cov}(\mu_{it},\, \hat{\mu}_j)}_{2} - \underbrace{\mathrm{cov}(\mu_{jt},\, \hat{\mu}_i)}_{3} + \underbrace{\mathrm{cov}(\hat{\mu}_i,\, \hat{\mu}_j)}_{4}.\end{aligned}$$

The expectations are zero since $\mathbb{E}[\mu_{jt}] = \mathbb{E}[\hat{\mu}_j] = \mu$ we only need to expand the covariance expression. The four terms can be expanded as follows

$$\begin{aligned}
(1) &= \bar{K}_{ij}^{tt} = \sigma_{it}\sigma_{jt}\rho_{ij} + \tau^2\delta_{ij} \\
(2) &= \sum_s a_{js}\mathrm{cov}(\mu_{it},\, \mu_{js}) = \sum_s a_{js}\bar{K}_{ij}^{ts} = \sum_s a_{js}(\sigma_{it}\sigma_{js}\rho_{ij}\delta_{ts} + \tau^2\delta_{ij}) \\
&= a_{jt}\sigma_{it}\sigma_{jt}\rho_{ij} + \tau^2\delta_{ij} \\
(3) &= a_{it}\sigma_{it}\sigma_{jt}\rho_{ij} + \tau^2\delta_{ij} \qquad \{\text{by } j \to i \text{ in } (2)\} \\
(4) &= K_{ij} = \sigma_{ij}^2\rho_{ij} + \tau^2\delta_{ij}.
\end{aligned}$$

The $\tau^2\delta_{ij}$ terms cancel and we get

$$\begin{aligned}
\mathbb{E}[(\mu_{it} - \hat{\mu}_i)(\mu_{jt} - \hat{\mu}_j)] &= \sigma_{it}\sigma_{jt}\rho_{ij} - a_{jt}\sigma_{it}\sigma_{jt}\rho_{ij} - a_{it}\sigma_{it}\sigma_{jt}\rho_{ij} + \sigma_{ij}^2\rho_{ij} \\
&= (1 - a_{it} - a_{jt})\sigma_{it}\sigma_{jt}\rho_{ij} + \sigma_{ij}^2\rho_{ij} \\
&= \left(1 - \frac{w_{it}}{w_{i*}} - \frac{w_{jt}}{w_{j*}}\right)\frac{\sigma_i\sigma_j\rho_{ij}}{\sqrt{w_{it}w_{jt}}} + \frac{\sigma_i\sigma_j\rho_{ij}}{w_{i*}w_{j*}}\sum_s \sqrt{w_{is}w_{js}} \\
&= \left(\frac{w_{i*}w_{j*} - w_{it}w_{j*} - w_{jt}w_{i*}}{\sqrt{w_{it}w_{jt}}} + \sum_s \sqrt{w_{is}w_{js}}\right)\frac{\sigma_i\sigma_j\rho_{ij}}{w_{i*}w_{j*}}.
\end{aligned}$$

It follows that

$$\begin{aligned}
& w_{i*}w_{j*}\mathbb{E}\left[\sum_t \sqrt{w_{it}w_{jt}}(\mu_{it} - \hat{\mu}_i)(\mu_{jt} - \hat{\mu}_j)\right] \\
&= \left(\sum_t (w_{i*}w_{j*} - w_{it}w_{j*} - w_{jt}w_{i*}) + \sum_t \sqrt{w_{it}w_{jt}}\sum_s \sqrt{w_{is}w_{js}}\right)\sigma_i\sigma_j\rho_{ij} \\
&= \left(Tw_{i*}w_{j*} - 2w_{i*}w_{j*} + \left(\sum_t \sqrt{w_{it}w_{jt}}\right)^2\right)\sigma_i\sigma_j\rho_{ij}.
\end{aligned}$$

Thus the following estimator is unbiased

$$\widehat{\sigma_i\sigma_j\rho_{ij}} = \frac{w_{i*}w_{j*}\sum_t \sqrt{w_{it}w_{jt}}(\mu_{it} - \hat{\mu}_i)(\mu_{jt} - \hat{\mu}_j)}{(T-2)w_{i*}w_{j*} + \left(\sum_t \sqrt{w_{it}w_{jt}}\right)^2}.$$

(37) then yields

$$\widehat{\sigma_{ij}^2\rho_{ij}} = \frac{\sum_t \sqrt{w_{it}w_{jt}}(\mu_{it} - \hat{\mu}_i)(\mu_{jt} - \hat{\mu}_j)}{\sum_t \sqrt{w_{it}w_{jt}} + (T-2)w_{i*}w_{j*}/\sum_t \sqrt{w_{it}w_{jt}}}.$$

What remains is to construct the unbiased estimator $\widehat{\tau^2}$. First, observe that

$$\mathbb{E}\left[\left(\hat{\mu}_j - \sum_i \frac{w_{i*}}{w_{**}}\hat{\mu}_i\right)^2\right] = \mathrm{Var}\left(\hat{\mu}_j - \sum_i \frac{w_{i*}}{w_{**}}\hat{\mu}_i\right) + \mathbb{E}\left[\hat{\mu}_j - \sum_i \frac{w_{i*}}{w_{**}}\hat{\mu}_i\right]^2.$$

The second term on the right-hand side is zero since $\mathbb{E}[\hat{\mu}_j] = \mu$ independently on the index $j$. Therefore, consider the following expansion of the variance.

$$\begin{aligned}
\operatorname{Var}\left(\hat{\mu}_j - \sum_i \frac{w_{i*}}{w_{**}}\hat{\mu}_i\right) &= \operatorname{Var}\left(\sum_i \left(\delta_{ij} - \frac{w_{i*}}{w_{**}}\right)\hat{\mu}_i\right) \\
&= \sum_{ik}\left(\delta_{ij} - \frac{w_{i*}}{w_{**}}\right)\left(\delta_{kj} - \frac{w_{k*}}{w_{**}}\right)K_{ik} \qquad (38) \\
&= \sum_{ik}\left(\delta_{ij} - \frac{w_{i*}}{w_{**}}\right)\left(\delta_{kj} - \frac{w_{k*}}{w_{**}}\right)\left(\sigma_{ik}^2\rho_{ik} + \tau^2\delta_{ik}\right).
\end{aligned}$$

The expression above will be expanded in two steps. Consider first

$$\begin{aligned}
&\sum_{ik}\left(\delta_{ij} - \frac{w_{i*}}{w_{**}}\right)\left(\delta_{kj} - \frac{w_{k*}}{w_{**}}\right)\sigma_{ik}^2\rho_{ik} \\
&= \sum_{ik}\left(\delta_{ij}\delta_{kj} - \delta_{ij}\frac{w_{k*}}{w_{**}} - \delta_{kj}\frac{w_{i*}}{w_{**}} + \frac{w_{i*}w_{k*}}{w_{**}^2}\right)\sigma_{ik}^2\rho_{ik} \\
&= \sum_i \left(\delta_{ij} - 2\frac{w_{i*}}{w_{**}}\right)\sigma_{ij}^2\rho_{ij} + \sum_{ik}\frac{w_{i*}w_{k*}}{w_{**}^2}\sigma_{ik}^2\rho_{ik}.
\end{aligned}$$

Observe that the second term is a constant. Therefore it follows that

$$\begin{aligned}
&\sum_j \frac{w_{j*}}{w_{**}}\sum_{ik}\left(\delta_{ij} - \frac{w_{i*}}{w_{**}}\right)\left(\delta_{kj} - \frac{w_{k*}}{w_{**}}\right)\sigma_{ik}^2\rho_{ik} \\
&= \sum_{ij}\left(\frac{w_{j*}}{w_{**}}\left(\delta_{ij} - 2\frac{w_{i*}}{w_{**}}\right) + \frac{w_{i*}w_{j*}}{w_{**}^2}\right)\sigma_{ij}^2\rho_{ij} \\
&= \sum_j \frac{w_{j*}}{w_{**}}\underbrace{\sum_i\left(\delta_{ij} - \frac{w_{i*}}{w_{**}}\right)\sigma_{ij}^2\rho_{ij}}_{:=c_j}.
\end{aligned}$$

Note that below is an unbiased estimator.

$$\hat{c}_j = \sum_i \left(\delta_{ij} - \frac{w_{i*}}{w_{**}}\right)\widehat{\sigma_{ij}^2\rho_{ij}}.$$

Now consider the weighted average of the second term in (38).

$$\begin{aligned}
&\sum_j \frac{w_{j*}}{w_{**}} \sum_{ik} \left(\delta_{ij} - \frac{w_{i*}}{w_{**}}\right) \left(\delta_{kj} - \frac{w_{k*}}{w_{**}}\right) \tau^2 \delta_{ik} \\
&= \sum_j \frac{w_{j*}}{w_{**}} \sum_i \left(\delta_{ij} - \frac{w_{i*}}{w_{**}}\right)^2 \tau^2 \\
&= \sum_j \frac{w_{j*}}{w_{**}} \sum_i \left(\delta_{ij}^2 + \frac{w_{i*}^2}{w_{**}^2} - 2\delta_{ij} \frac{w_{i*}}{w_{**}}\right) \tau^2 \\
&= \sum_j \frac{w_{j*}}{w_{**}} \left(1 + \sum_i \frac{w_{i*}^2}{w_{**}^2} - 2\frac{w_{j*}}{w_{**}}\right) \tau^2 \\
&= \left(1 + \sum_i \frac{w_{i*}^2}{w_{**}^2} - 2\sum_j \frac{w_{j*}^2}{w_{**}^2}\right) \tau^2 \\
&= \left(1 - \sum_i \frac{w_{i*}^2}{w_{**}^2}\right) \tau^2.
\end{aligned}$$

We now have

$$\mathbb{E}\left[\sum_j \frac{w_{j*}}{w_{**}} \left(\hat{\mu}_j - \sum_i \frac{w_{i*}}{w_{**}} \hat{\mu}_i\right)^2\right] = \sum_j \frac{w_{j*}}{w_{**}} c_j + \left(1 - \sum_i \frac{w_{i*}^2}{w_{**}^2}\right) \tau^2.$$

From this, it follows that below is an unbiased estimator of $\tau^2$.

$$\widehat{\tau^2} = \frac{\sum_j \frac{w_{j*}}{w_{**}} \left\{\left(\hat{\mu}_j - \sum_i \frac{w_{i*}}{w_{**}} \hat{\mu}_i\right)^2 - \hat{c}_j\right\}}{1 - \sum_i w_{i*}^2 / w_{**}^2}.$$

The results of this section can be summarized as follows.

$$\begin{cases}
\widehat{\sigma_{ij}^2 \rho_{ij}} & = \dfrac{\sum_t \sqrt{w_{it} w_{jt}} (\mu_{it} - \hat{\mu}_i)(\mu_{jt} - \hat{\mu}_j)}{\sum_t \sqrt{w_{it} w_{jt}} + (T-2) w_{i*} w_{j*} / \sum_t \sqrt{w_{it} w_{jt}}} \\
\hat{c}_j & = \displaystyle\sum_i \left(\delta_{ij} - \frac{w_{i*}}{w_{**}}\right) \widehat{\sigma_{ij}^2 \rho_{ij}} \\
\widehat{\tau^2} & = \dfrac{\sum_j w_{j*} \left\{\left(\hat{\mu}_j - \sum_i \frac{w_{i*}}{w_{**}} \hat{\mu}_i\right)^2 - \hat{c}_j\right\}}{w_{**} - \sum_i w_{i*}^2 / w_{**}} \\
\hat{K}_{ij} & = \widehat{\sigma_{ij}^2 \rho_{ij}} + \widehat{\tau^2} \delta_{ij}.
\end{cases} \tag{39}$$

As a final note, due to its construction, $\widehat{\tau^2}$ may come out negative. This is also a possibility in both the Bühlmann-Straub and Parodi-Bonche models. As suggested in [2] (page 95 and consequently also in [3] page 25) one may want to consider the following transformation instead

$$\widehat{\tau^2}^+ = \max\left(\widehat{\tau^2}, 0\right). \tag{40}$$

The transformed estimator will unfortunately be biased. Additionally, $\tau^2 = 0$ implies that there is no difference between the individual cohorts and we should therefore not consider them separately. This is in line with the fact that

$$\lim_{\tau^2 \to 0} z_j = 0,$$

in the Bühlmann-Straub model and the proposed model.

Notice that the estimators for $\sigma_j$ are consistent between the two models. This can be seen by realizing that

$$\widehat{\frac{\sigma_j^2}{w_{j*}}} = \widehat{\sigma_{jj}^2} = \widehat{\sigma_{jj}^2 \rho_{jj}} = \frac{\sum_t w_{jt}(\mu_{jt} - \hat{\mu}_j)^2}{w_{j*} + (T-2)w_{j*}} = \frac{\hat{\sigma}_j^2}{w_{j*}}, \tag{41}$$

where $\hat{\sigma}_j$ is the Bühlmann-Straub estimator in (13).

### 2.9 The balance property

A set of credibility blended estimates $\hat{\mu}_j^b$ is said to satisfy the balance property if the following condition is met.

$$\sum_{jt} w_{jt}\mu_{jt} = \sum_{jt} w_{jt}\hat{\mu}_j^b.$$

Observe that the property is satisfied in expectation for any unbiased model. This follows from $\mathbb{E}[\mu_{jt}] = \mathbb{E}[\hat{\mu}_j^b] = \mu$.

Note that $w_{jt}\mu_{jt}$ is not a key ratio. Instead, it is the measurement $x_{jt}$ in the numerator of the key ratio. The balance property then states that the blended estimate is unbiased with respect to $x_{**}$, i.e. the aggregation of measurements for the portfolio historically. As an example, suppose $x_{jt}$ is the aggregate loss amount for cohort $j$ over the term $t$. The blended estimate $\hat{\mu}_j^b$ then corresponds to a risk premium. The balance property states that if the insurer would have charged the risk premium historically, the net result would have been exactly 0.

We will now study the balance property closer for all models with $a_{jt} = w_{jt}/w_{j*}$. Consequently

$$\sum_t w_{jt}\mu_{jt} = w_{j*}\hat{\mu}_j.$$

Additionally, since $\hat{\mu}_j^b$ is independent on $t$ the balance property can be written as

$$\sum_j w_{j*}(\hat{\mu}_j - \hat{\mu}_j^b) = 0,$$

or

$$\sum_j w_{j*}(1 - z_j)(\hat{\mu}_j - \hat{\mu}) = 0.$$

This condition is not satisfied in general. It is however satisfied by the Bühlmann-Straub model regardless of whether the underlying assumptions of the model are satisfied or not. To see this, note first that for the Bühlmann-Straub model, we have

$$w_{j*}(1 - z_j) = w_{j*}\left(1 - \frac{w_{j*}}{w_{j*} + \kappa}\right) = \frac{w_{j*}\kappa}{w_{j*} + \kappa} = \kappa z_j,$$

and therefore also

$$\begin{aligned}\sum_j w_{j*}(1 - z_j)(\hat{\mu}_j - \hat{\mu}) &= \sum_j \kappa z_j(\hat{\mu}_j - \hat{\mu}) \\ &= \kappa \sum_j z_j \hat{\mu}_j - \kappa z_* \hat{\mu} \\ &= 0,\end{aligned}$$

since $\hat{\mu} = \sum_j z_j \hat{\mu}_j / z_*$. We have thus shown that the balance property is always satisfied for the Bühlmann-Straub model regardless of whether the assumptions in (28) are met or not.

# 3 Simulation study

## 3.1 Motivation

In this section, we will generate random but realistic data to which the proposed and Bühlmann-Straub models will be applied. The benefit of using simulated data is that the accuracy of the models can be easier assessed using the (mean) relative error since we have access to the true model parameters.

The basic idea is to run a large number of $N$ simulations and investigate the distribution of the relative errors in estimating $\mu$, $\mu_j$, and $\tau^2$. Since the proposed model does not satisfy the balance property in general the relative error for the aggregated measurements will be considered as well.

The simulations will be run for the special case

$$\begin{cases} \sigma_{jt}^2 & = \sigma_j^2 / w_{jt} \\ \rho_{ij} & = \rho + (1-\rho)\delta_{ij}. \end{cases}$$

The second assumption states that all pair-wise correlations are identical and equal to $\rho$. In other words, $\rho_{ij}$ is the elements of a so-called equicorrelation matrix [5]. In practice the correlations will most likely not be equal, however, using a single correlation parameter should be good enough to capture the average overall correlation. It will also enable us to easier study the error's dependency on the correlation.

The parameters $\mu$, $\tau$, $J$, and $T$ will be considered fixed and will therefore not change between the simulations. The entities $\mu_j$, $w_{jt}$, $\mu_{jt}$ and $\sigma_j$ will be randomly generated for each simulation. When it comes to $\rho$ it turns out we cannot select any $\rho \in [-1, 1]$. Since variance is non-negative a lower bound for $\rho$ can be constructed as follows.

$$\begin{aligned} \operatorname{Var}\left(\sum_j \varepsilon_{jt}\right) &\geq 0, \\ \sum_j \operatorname{Var}(\varepsilon_{jt}) + \sum_{i \neq j} \operatorname{cov}(\varepsilon_{it}, \varepsilon_{jt}) &\geq 0, \\ J + J(J-1)\rho &\geq 0, \\ \rho &\geq \frac{1}{1-J}. \end{aligned}$$

The same result is also mentioned in [5]. For simulation $n \in [1, N]$ we therefore set

$$\rho_n = \frac{N - 1 - J(n-1)}{(1-J)(N-1)}. \tag{42}$$

For later reference, we will also define the following relative errors

$$\mathrm{RE}_\mu = (\hat{\mu} - \mu)/\mu \tag{43}$$

$$\mathrm{RE}_{\mu_j} = \sum_j (\hat{\mu}_j^b - \mu_j)/\mu_j/J \tag{44}$$

$$\mathrm{RE}_\tau = (\hat{\tau}^2 - \tau^2)/\tau^2 \tag{45}$$

$$\mathrm{RE}_{\mathrm{BP}} = \left( \sum_j w_{j*}(\hat{\mu}_j^b - \hat{\mu}_j) \right) \Bigg/ \sum_j w_{j*}\hat{\mu}_j. \tag{46}$$

As a final note, we will use the "raw" estimators i.e. we will not use the transformations mentioned in (40) or in (34).

### 3.2 Parameter choice

The following parameter choice is based on actuarial expert judgment and is considered realistic for transformed loss ratio data. Each cohort is thus a sub-portfolio within a larger portfolio. The key ratio corresponds to a loss ratio and the weights correspond to earned premium. This description is intentionally vague in order to protect company information. The actual choice is summarized in Table 2. As can be seen, the parameter table states that we are studying 9 cohorts over 10 terms.

With the current parameter choice, it is possible that $\mu_{jt}$ is negative. Although this seems strange for a loss ratio it is not a problem for the model itself.

Table 2: Parameter choice for simulation study

| Variable | Sampled from, if random | Value, if fixed |
|---|---|---|
| $\mu$ | | 1 |
| $\tau$ | | 0.6 |
| $J$ | | 9 |
| $T$ | | 10 |
| $N$ | | $10^5$ |
| $\mu_j$ | $\Gamma(\mu^2/\tau^2,\ \mu/\tau^2)$ | |
| $\sigma_j^2$ | $\mathrm{Lognormal}(16,\ \sqrt{2})$ | |
| $w_j$ | $\mathcal{N}(3 \cdot 10^7,\ 3 \cdot 10^5)$ | |
| $w_{jt}$ | $\mathcal{N}(w_j,\ w_j/10)$ | |

### 3.3 Simulation procedure

In this subsection, we describe how the data, ultimately $\mu_{jt}$, is generated. We do this by defining a couple of simulation steps as follows.

1. Set simulation fixed parameters according to Table 2 and simulation counter $n$ to 1.
2. Construct the $n$th equicorrelation matrix by using (42).
3. Generate the random parameters according to Table 2.
4. Generate the matrix elements $\varepsilon_{jt}$ with columns drawn from a multivariate normal distribution with zero mean vector and covariance matrix equal to the equicorrelation matrix.
5. Generate the data points by the identity $\mu_{jt} = \mu_j + \sigma/\sqrt{w_{jt}} \cdot \varepsilon_{jt}$.
6. Estimate model parameters and evaluate the model. Calculate the relative errors $\mathrm{RE}_x$ defined in Section 3.1.
7. Go back to step 2 and increment $n$ by 1. Stop once $n = N$ is reached.

Once the data points have been generated (step 5) the proposed and Bühlmann-Straub models are evaluated using (27) and (12) respectively. The model parameters are estimated using (39) and (13) respectively.

### 3.4 A single simulation

For illustrative purposes, one out of the $N$ simulations will be investigated in closer detail. Specifically, we will use the first simulation, meaning that $\rho = \rho_1 = -0.125$. Remember that we are studying 9 cohorts over 10 insurance periods.

In Table 3.4 and in Table 3.4 the weights and key ratios can be seen respectively. In Table 3.4 the two models have been evaluated. A part of the table has been visualized in Figure 4 as well.

Due to (29) and (41) the estimators $\hat{\mu}_j$ and $\hat{\sigma}_{jj}$ are identical between the two models. The models are only different in $\hat{\mu}$, $z_j$, and $\hat{\mu}_j^b$. This explains the structure of Table 3.4.

The proposed model contains an estimator for the correlation matrix. The outcome of the estimator can be seen in Table 3.4. Notice that the estimator in (39) could be improved if the correlation was assumed to be identical between the cohorts. Since this is not the case here the proposed model will in a sense overfit the correlations. The average of the elements in the lower diagonal of the correlation matrix can be seen in Table 3.4 and we see that it is fairly close to the actual value nevertheless. Remember that the Bühlmann-Straub model has no estimator for the correlation.

The relative errors as defined in Section 3.1 are shown in Table 3.4. For the first simulation, the proposed model estimated $\tau$ noticeably better than the Bühlmann-Straub model. It also performed somewhat better at estimating the vector of elements $\mu_j$ when using the blended estimate, although both models performed quite badly in this simulation. No noticeable difference can be seen when estimating $\hat{\mu}$. Finally, a slight error in the balance property is seen for the proposed model as expected.

Table 3: Weights $w_{jt}$ in millions for the first simulation

| $j \setminus t$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2.9 | 2.5 | 2.1 | 2.3 | 2.3 | 2.5 | 2.5 | 2.6 | 2.5 | 2.6 |
| 2 | 3.7 | 3.5 | 3.1 | 3.1 | 2.7 | 3.2 | 3.4 | 2.7 | 3.1 | 3.8 |
| 3 | 3.7 | 2.8 | 3.1 | 3.1 | 3.2 | 2.4 | 2.7 | 2.9 | 3.1 | 2.6 |
| 4 | 2.9 | 3.0 | 2.9 | 1.9 | 3.0 | 3.2 | 2.8 | 2.2 | 2.9 | 2.6 |
| 5 | 3.5 | 3.5 | 3.2 | 2.8 | 3.2 | 3.3 | 3.6 | 3.1 | 3.0 | 3.7 |
| 6 | 3.0 | 2.9 | 3.0 | 2.9 | 2.5 | 2.5 | 2.8 | 3.4 | 2.4 | 2.9 |
| 7 | 3.2 | 3.5 | 2.9 | 3.5 | 3.1 | 3.8 | 3.3 | 3.4 | 3.2 | 3.1 |
| 8 | 2.4 | 2.1 | 2.6 | 2.9 | 2.7 | 2.8 | 2.8 | 2.9 | 2.1 | 2.4 |
| 9 | 3.5 | 2.7 | 3.4 | 3.1 | 3.1 | 3.2 | 3.1 | 3.6 | 3.1 | 3.2 |

Table 4: Data points $\mu_{jt}$ for the first simulation

| $j \setminus t$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 5.00 | -3.16 | -0.61 | -0.77 | 1.12 | 1.09 | -0.23 | 2.22 | -1.24 | 1.21 |
| 2 | 3.11 | 1.33 | 3.08 | 0.98 | -0.81 | 1.86 | -2.05 | 2.05 | 3.28 | 2.10 |
| 3 | -0.98 | -1.03 | -0.52 | -0.16 | 1.21 | 1.30 | 4.77 | -1.25 | 0.98 | 0.08 |
| 4 | 1.74 | 0.36 | 0.53 | 0.82 | 1.24 | -0.28 | 0.36 | 0.96 | 2.98 | 2.70 |
| 5 | -7.38 | -0.47 | 0.19 | 4.07 | 1.88 | 1.09 | -2.26 | 3.68 | 3.25 | 0.23 |
| 6 | -3.19 | 4.42 | 1.56 | 0.10 | -1.60 | -1.13 | -0.48 | -2.93 | -2.72 | -0.58 |
| 7 | 1.21 | 0.14 | 0.82 | 0.98 | 0.91 | 1.98 | 1.24 | 0.22 | -1.69 | 0.46 |
| 8 | 3.04 | 6.56 | 1.68 | 3.65 | 2.05 | 1.97 | 4.20 | 0.39 | 4.00 | -0.38 |
| 9 | 1.36 | 0.63 | 0.30 | -1.33 | 0.08 | -0.99 | 0.33 | 1.85 | -0.52 | -0.43 |

Table 5: Model evaluation for the proposed and Bühlmann-Straub models for the first simulation

| $j$ | $w_{j*} \cdot 10^{-5}$ | $\sigma_{jj}$ | $\hat{\sigma}_{jj}$ | $\mu_j$ | $\hat{\mu}_j$ | $\hat{\mu}$ | $\hat{\mu}^B$ | $z_j$ | $z_j^B$ | $\hat{\mu}_j^b$ | $\hat{\mu}_j^{b,B}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2487 | 0.65 | 0.73 | 1.05 | 0.55 | 0.7173 | 0.7180 | 0.38 | 0.48 | 0.65 | 0.64 |
| 2 | 3244 | 0.64 | 0.55 | 0.83 | 1.52 | 0.7173 | 0.7180 | 0.53 | 0.54 | 1.14 | 1.15 |
| 3 | 2968 | 0.54 | 0.55 | 1.08 | 0.37 | 0.7173 | 0.7180 | 0.52 | 0.52 | 0.54 | 0.54 |
| 4 | 2738 | 0.33 | 0.34 | 0.98 | 1.12 | 0.7173 | 0.7180 | 0.74 | 0.50 | 1.02 | 0.92 |
| 5 | 3271 | 0.76 | 1.08 | 0.34 | 0.24 | 0.7173 | 0.7180 | 0.23 | 0.54 | 0.61 | 0.46 |
| 6 | 2837 | 0.65 | 0.74 | 0.97 | -0.65 | 0.7173 | 0.7180 | 0.38 | 0.51 | 0.19 | 0.02 |
| 7 | 3302 | 0.30 | 0.31 | 0.15 | 0.65 | 0.7173 | 0.7180 | 0.77 | 0.55 | 0.66 | 0.68 |
| 8 | 2565 | 0.73 | 0.61 | 1.70 | 2.63 | 0.7173 | 0.7180 | 0.48 | 0.48 | 1.63 | 1.64 |
| 9 | 3196 | 0.39 | 0.32 | 0.10 | 0.16 | 0.7173 | 0.7180 | 0.76 | 0.54 | 0.29 | 0.42 |

Table 6: Estimated correlation matrix for the proposed model for the first simulation

| $j \setminus j$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.000 | 0.220 | -0.188 | 0.191 | -0.503 | -0.756 | 0.381 | -0.547 | 0.439 |
| 2 | 0.220 | 1.000 | -0.738 | 0.396 | -0.041 | -0.155 | -0.351 | -0.257 | 0.099 |
| 3 | -0.188 | -0.738 | 1.000 | -0.153 | -0.028 | -0.084 | 0.177 | 0.198 | -0.329 |
| 4 | 0.191 | 0.396 | -0.153 | 1.000 | 0.020 | -0.443 | -0.694 | -0.227 | -0.008 |
| 5 | -0.503 | -0.041 | -0.028 | 0.020 | 1.000 | 0.048 | -0.357 | -0.202 | -0.418 |
| 6 | -0.756 | -0.155 | -0.084 | -0.443 | 0.048 | 1.000 | 0.057 | 0.490 | -0.211 |
| 7 | 0.381 | -0.351 | 0.177 | -0.694 | -0.357 | 0.057 | 1.000 | -0.167 | -0.075 |
| 8 | -0.547 | -0.257 | 0.198 | -0.227 | -0.202 | 0.490 | -0.167 | 1.000 | -0.113 |
| 9 | 0.439 | 0.099 | -0.329 | -0.008 | -0.418 | -0.211 | -0.075 | -0.113 | 1.000 |

Table 7: Estimation outcome for the model parameters for the first simulation

| Parameter | True value | Proposed estimator | Bühlmann-Straub estimator |
|---|---|---|---|
| $\mu$ | 1 | 0.7173 | 0.7180 |
| $\tau$ | 0.6 | 0.617 | 0.652 |
| $\rho$ | -0.125 | -0.120 | 0 |

Table 8: Summary of the relative errors in percent for the first simulation

| Error type | Proposed model | Bühlmann-Straub model |
|---|---|---|
| $\mathrm{RE}_{\tau}$ | 5.66 | 18.02 |
| $\mathrm{RE}_{\mu_j}$ | 54.79 | 62.47 |
| $\mathrm{RE}_{\mu}$ | -28.27 | -28.20 |
| $\mathrm{RE}_{\mathrm{BP}}$ | 4.07 | 0 |

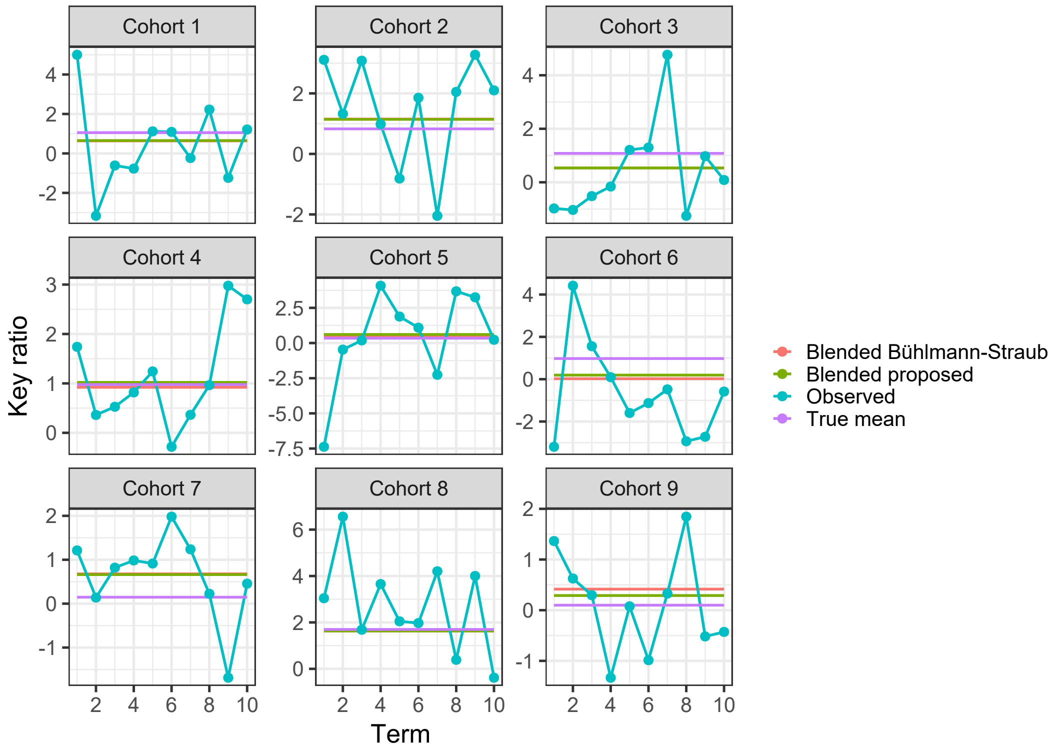


Figure 4: A visualization of the first simulation in the series of $N$ simulations.

### 3.5 All $N$ simulations

Now we will perform a similar analysis as in the previous subsection but averaged over all $N$ runs.

We will use a pseudo-log scale for the visualization of the relative error distributions. The pseudo-log transform is defined as

$$\texttt{pseudo\_log}(x) := \operatorname{arsinh}\left(\frac{x}{2}\right) = \ln\left(\frac{x}{2} + \sqrt{\left(\frac{x}{2}\right)^2 + 1}\right).$$

This transformation retains the sign of the original variable while at the same time behaving like the regular log transform for large $x$. Since the relative errors will span across multiple orders of magnitude a log-like transformation will help visualize the data. In Figure 5 the transformation has been visualized and can be compared to the natural logarithm and identity transformations.

The large variance in the proposed model is mainly due to singularities in $z_j$ and $b_j$. Similar to what we saw in the two-dimensional special case previously. If the estimated parameters happen to be close to the points of singularity the error may be magnified quite significantly.

We will start by inspecting density plots for the relative errors. Please see Figures 6, 7, 8 and 9. A summary is also available in Table 3.5.

For all targets, except for $\tau$, and for both models, the median relative error is lower than the average relative error. This indicates that outlier behavior is driving the average relative error. This likely originates from the heavy tail in the random data connected to the log-normal distribution assigned to $\sigma_j^2$. If one focuses on the mode of the density plots it is more clear that the proposed model outperforms the Bühlmann-Straub model in the simulations.

The blended estimate from the proposed model seems to outperform the Bühlmann-Straub model at estimating $\mu_j$. This comes at the cost of losing the balance property, but only by a couple of percents in the median.

The same data can also be plotted as a function of the correlation coefficient. Please see Figures 10, 11, 12 and 13. Since we already have the overall average from the summary statistics it would be of added interest to consider local averages. The local average is studied using the function `ggplot2::geom_smooth`() in `R`. By default, it uses a GAM with smoothing. The exact details of the trend line are not of interest here, as long as it is flexible enough.

The accuracy of the proposed model seems to be independent of $\rho$. This is good because otherwise, it would indicate there is still some bias left in the estimators. It is interesting to see that both models seem to be performing equally well at estimating $\mu$ when less focus is put on extreme values. The benefit of the proposed model seems to be a growing function of $\rho$ when estimating $\mu_j$ and $\tau$. Most notably so for $\mu_j$.

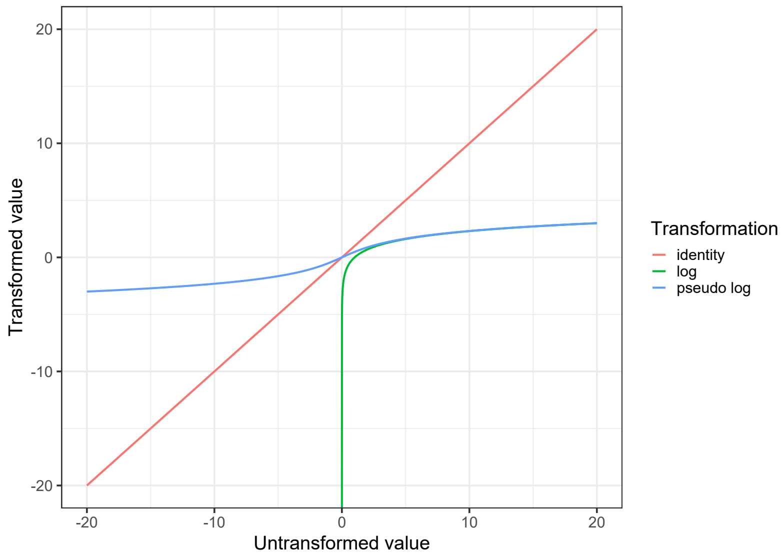


Figure 5: A comparison of the transformation candidates used for scaling the relative errors.

Table 9: Summary statistics in percent for the relative errors for the two models over $10^5$ simulations.

| Error type | Proposed model | | Bühlmann-Straub model | |
|---|---|---|---|---|
| | Average | Median | Average | Median |
| $\mathrm{RE}_{\mu}$ | -10.4 | 0.5 | 0.4 | -0.1 |
| $\mathrm{RE}_{\mu_j}$ | -22.3 | 16.7 | 55.0 | 39.4 |
| $\mathrm{RE}_{\tau}$ | -0.9 | -33.7 | -60.2 | -77.6 |
| $\mathrm{RE}_{\mathrm{BP}}$ | -15.1 | -2.7 | 0 | 0 |

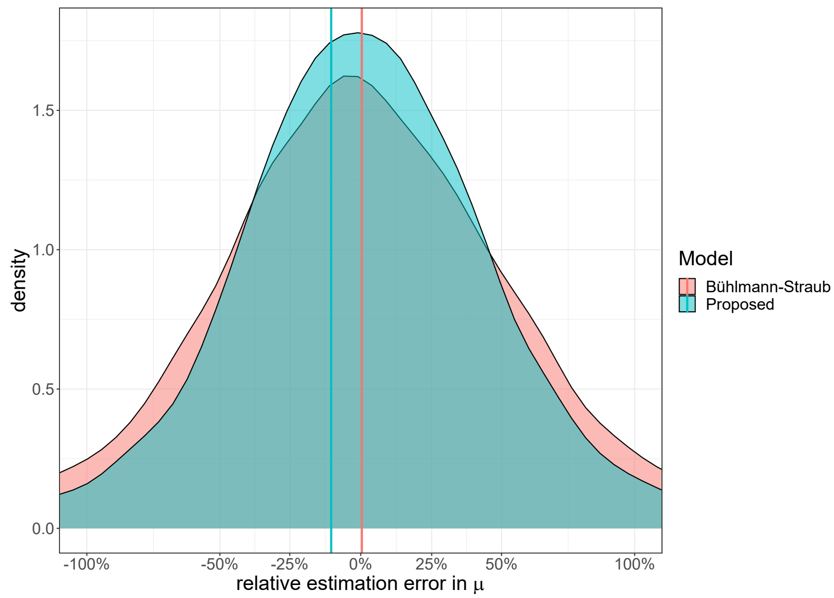


Figure 6: The distribution of relative error in the estimator $\hat{\mu}$ as defined in (43) over $10^5$ simulations for the two models. A vertical line is drawn at the average.

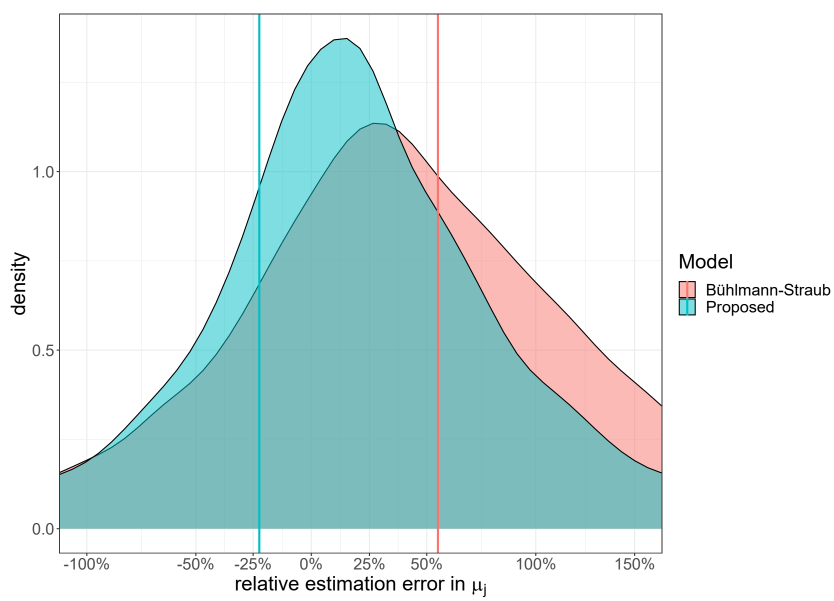


Figure 7: The distribution of relative error in the estimator $\hat{\mu}_j^b$ as defined in (44) over $10^5$ simulations for the two models. A vertical line is drawn at the average.

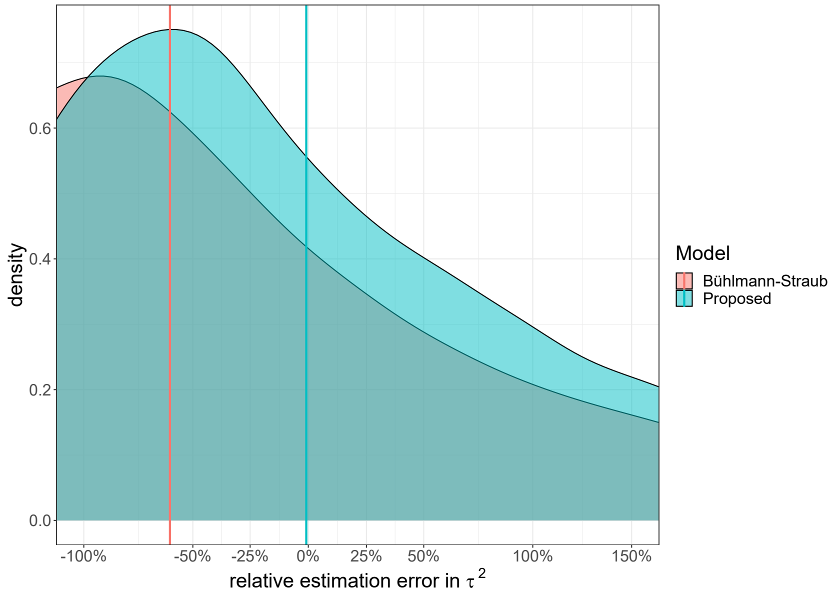


Figure 8: The distribution of relative error in the estimator $\hat{\tau}^2$ as defined in (45) over $10^5$ simulations for the two models. A vertical line is drawn at the average.

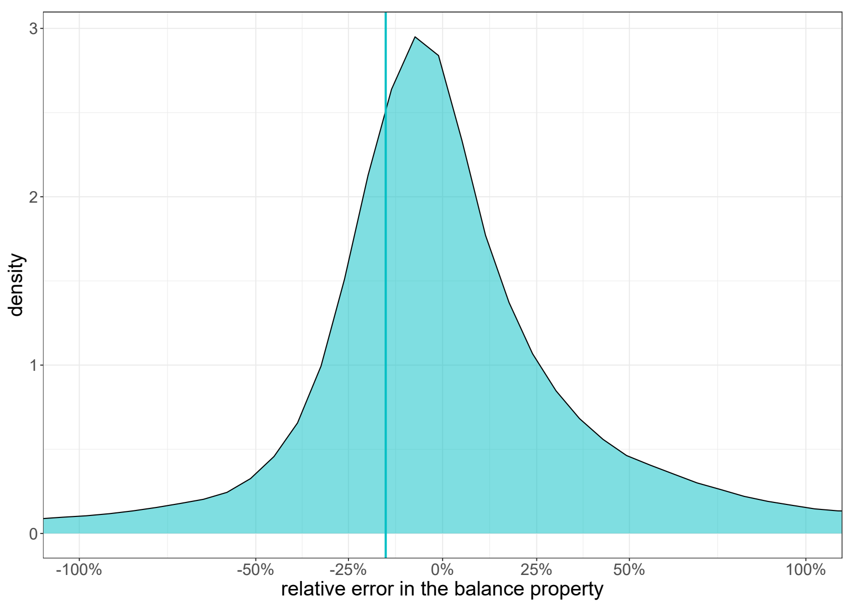


Figure 9: The distribution of relative error in the balance property as defined in (46) over $10^5$ simulations for the proposed model. A vertical line is drawn at the average.

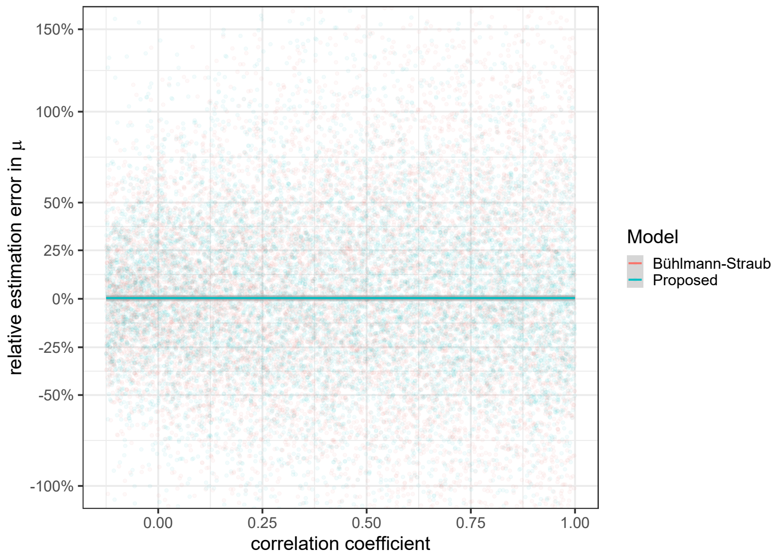


Figure 10: The $\rho$-dependency of the relative error in the estimator $\hat{\mu}$ as defined in (43) over $10^5$ simulations for the two models. A smooth trend line has been added to help identify local behavior.

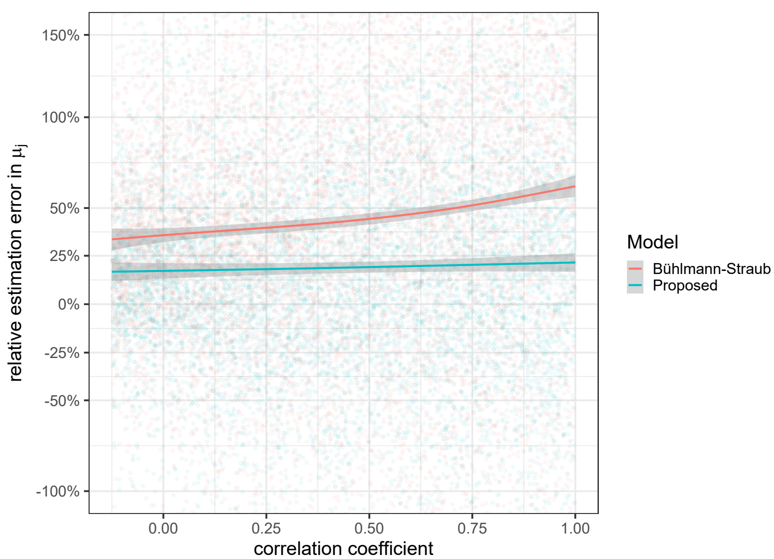


Figure 11: The $\rho$-dependency of the relative error in the estimator $\hat{\mu}_j$ as defined in (44) over $10^5$ simulations for the two models. A smooth trend line has been added to help identify local behavior.

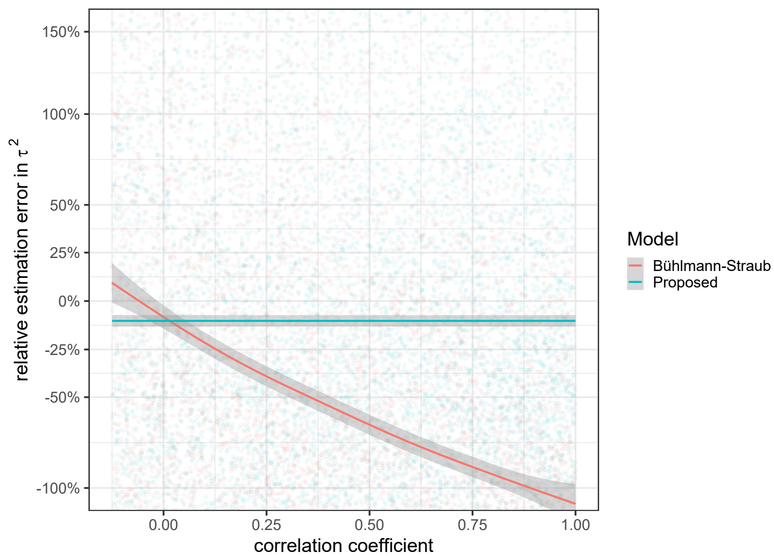


Figure 12: The $\rho$-dependency of the relative error in the estimator $\hat{\tau}^2$ as defined in (45) over $10^5$ simulations for the two models. A smooth trend line has been added to help identify local behavior.

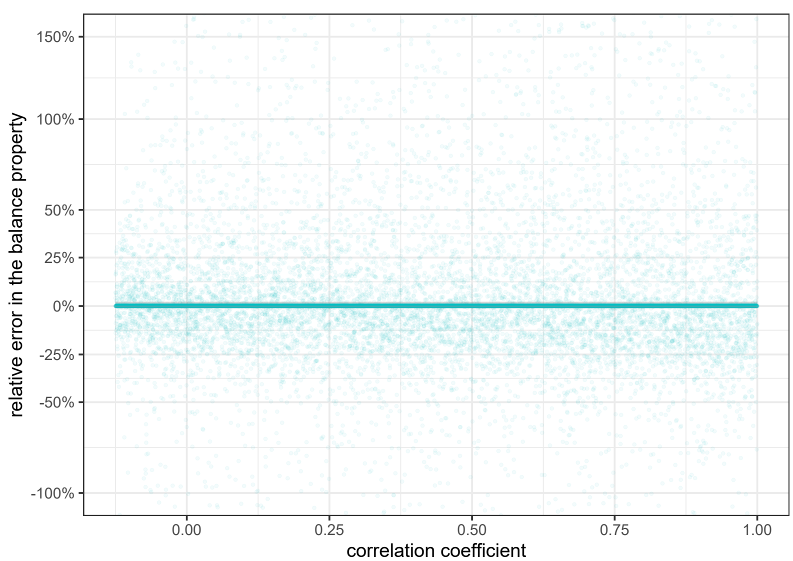


Figure 13: The $\rho$-dependency of the relative error in the balance property as defined in (46) over $10^5$ simulations for the proposed model. A smooth trend line has been added to help identify local behavior.

# 4 Discussion

The proposed model extends the homogeneous version of the Bühlmann-Straub model by allowing for correlation between the cohorts and also a more general variance structure. Remember that homogeneous here means that there is a built-in estimator for $\mu$ in the model. The inhomogeneous model takes $\mu$ as known. It was shown that the proposed model is an extension in the sense that the homogeneous Bühlmann-Straub model is included as a special case.

The proposed model is a so-called balanced model. Meaning that the number of terms with data is equal for all cohorts. In other words, $T$ is assumed to be independent of $j$. This assumption was made to simplify the notation and derivation of the estimators.

The basic idea behind the proposed model is to use the correlation between the cohorts to cancel noise and obtain more accurate estimators. This was especially clear in Section 2.7.1 where only two cohorts were considered. There we saw that a more extreme correlation could be used to gain superior estimators. It was also shown that when the size of the second cohort was increased the amount of noise that could be canceled when estimating the first cohort was increased. This is different from the classical model where only one credibility coefficient is present in each credibility factor.

It is worth noting that the extension introduced singularities in the credibility factor and weights $\boldsymbol{b}$. For the credibility factor, this happens when the optimal weights $b_j$ imply $\hat{\mu}_j = \hat{\mu}$. This point by itself is no problem since any $z_j$ would suffice here. However, it does mean that around this point the credibility factor is unbounded. If $z_j > 1$ then the credibility put into $\hat{\mu}$ is negative. Similarly for $b_j$ since we did not put any bounds on it, it may be optimal to choose $b_j > 1$. If this is optimal then it would give $z_j < 0$.

From a practical standpoint, it might seem strange to use negative weights in the calculation of a weighted average. Although from a mathematical standpoint, it is not actually a problem. Consider for instance short-selling in the context of minimum variance portfolios in finance mathematics. When there is a positive correlation, using a negative weight may reduce the variance of the weighted average. This is quite easy to see when looking at the weighting of two variables. It is, of course, possible to apply transformations to $z_j$ and $b_j$, such as ceilings and floors, at the cost of losing accuracy.

The simulation study seems to indicate that the proposed model and the Bühlmann-Straub model are equally well suited for estimating the portfolio average $\mu$. The largest difference between the models is seen when estimating $\mu_j$ using the blended estimator for large overall correlations. It was also noted that the result is quite volatile and ignoring extreme relative errors was needed to more easily compare the models.

It is important to note that the balance property in the homogeneous Bühlmann-Straub model is lost in the general case for the proposed model. The significance of this in practice is however unclear to the author.

The main concern with the proposed model from the author's perspective is the risk of over-parametrization. The bulk of the parameters will be captured by the covariance matrix $K$ meaning that the number of parameters in the model grows with the square of the number of cohorts. It is not fully understood how much benefit is gained from these added parameters compared to the Bühlmann-Straub model if the data set is limited. A somewhat limited data set was studied in the literature study but further studies are needed to better understand the practical ramifications.

# 5 Further research

It is unavoidable at some point to stumble upon a data set with cohort data that spans different ranges. Therefore there should be an alternative formulation of the proposed model that allows for a cohort-dependent time length, i.e. the extension $T \to T_j$. The extension should only have an impact on the estimator of $\sigma_{ij}^2 \rho_{ij}$.

Further research is needed to better understand how often unconventional weights (negative or larger than one) occur in practice and how reasonable the final blended estimates are in such cases. It is possible to receive extreme values of the estimators if the parameters are close to a singularity.

With the above in mind, it could be insightful to look into searching for the optimal weights $\boldsymbol{b}$ under additional constraints. One option is to add the constraint $b_j \geq 0$ for all $j$. Inequality-constrained optimization is unfortunately not as easily solved generally. Alternatives exist, such as optimizing with respect to a transformed variable $\tilde{b}_j = b_j^2$ which is always non-negative. These modifications would require an adjustment to the credibility factor since $\nu_j = s^2$ would no longer hold. It would also result in a worse expected square error, but it would likely still perform better near the singularities.

Further research is also needed to study the risk of over-parametrization. With this in mind, it would be interesting to study a simplified model with only a single correlation parameter. To pursue this, it is suggested to modify the estimator of the covariance matrix $K$ accordingly. Some kind of weighted average of the elements $\rho_{ij}$ seems intuitive.

## References


[1] Ohlsson, E., & Johansson, B. (2010). Non-life insurance pricing with generalized linear models. Berlin: Springer.

[2] Bühlmann, H., & Gisler, A. (2005). A course in credibility theory and its applications. Berlin: Springer.

[3] Parodi, P., & Bonche, S. (2010). Uncertainty-based credibility and its applications. Variance.

[4] Wen, L., Wu, X., & Zhou, X. (2009). The credibility premiums for models with dependence induced by common effects. Insurance: Mathematics and Economics.

[5] O'Neill, B. (2021). The Double-Constant Matrix, Centering Matrix and Equicorrelation Matrix: Theory and Applications. arXiv preprint arXiv:2109.05814. Page 18.

[6] Wen, L., & Deng, W. (2011). The credibility models with equal correlation risks. Journal of Systems Science and Complexity.

[7] Nurrohmah, S., & Fithriani, I. (2021). Bühlmann credibility model with correlated risk parameters. In Journal of Physics: Conference Series. IOP Publishing.