Selection of Credibility Regression Models

by

Peter Bühlmann and Hans Bühlmann

Research Report No. 83
April 1998

Seminar für Statistik
Eidgenössische Technische Hochschule (ETH)
CH-8092 Zürich
Switzerland

Selection of Credibility Regression Models

Peter Bühlmann
Seminar für Statistik
ETH Zentrum
CH-8092 Zürich, Switzerland

and

Hans Bühlmann
Departement Mathematatik
ETH Zentrum
CH-8092 Zürich, Switzerland

April 1998

Abstract

We give a decision rule to select regression models in a credibility context. In contrast to usual model selection techniques on a collective level, our proposal allows to detect individual structures, even if they disappear in the collective.

We use the case of simple linear regression to illustrate the basic new ideas. The extension to other models may need some refinement but can be made on the basis of the same fundamental ideas.

Key words and phrases. Bias-variance trade-off, Empirical Bayes, Linear trend, Squared prediction loss.

Introduction

In the open market economy of today, one of the most challenging tasks of an insurer consists in the design of a rating system catching all relevant rating factors and omitting all irrelevant ones. Mathematically, this may be modelled as the endeavour of finding those covariates which lead to the best possible predictions, for example in a regression model.

In classical statistics, with a frequentist interpretation, this problem of model selection is widely discussed, cf. Akaike (1970), Mallows (1973), Linhart and Zucchini (1986). However, we do not know of any selection scheme for credibility models. It is hence the aim of this paper to start such a discussion. This is not possible without some rather general remarks to begin with. The reader of this article will, however, soon discover that, when it comes to technicalities, we prefer to restrict ourselves to the case of Simple Linear Regression. We believe that this case suffices to illustrate the basic ideas. The extension to the case of General Regression or other models (with many covariates) may need some refinement but can be made on the basis of the same fundamental ideas.

As often done in practice, a model for the collective data is selected with frequentist methods and for such a chosen model, credibility is then introduced in a second stage. This approach is missing individual structure which averages out in the collective; see also Figure 1 where there is an individual trend (slope) but no trend in the collective. In insurance, such collective decision making potentially leads to anti–selection which is very unpleasant and dangerous for the insurer. Our model selection criterion accounts for individual structure and is not mislead by the collective view.

Part I: The Decision Rule

We describe in this part the set-up of the problem and derive a rule for model selection.

1. The Full Regression Model

Consider a class of individual risks $r = 1, 2, \ldots, N$, each of them with risk parameter $\qquad \vartheta_r$,

risk observations $\qquad \begin{pmatrix} X_{1r} \\ X_{2r} \\ \vdots \\ X_{nr} \end{pmatrix} = \mathbf{X}_r$,

and probability distribution $\qquad dP(\mathbf{X}_r|\vartheta_r)$.

We assume that the individually correct pure premiums

$$\mu_i(\vartheta_r) = E[X_{ir}|\vartheta_r]\,, \qquad i = 1, 2, \ldots, n\,,$$

follow a regression pattern

$$\mu(\vartheta_r) = D_r\beta(\vartheta_r)\ ,$$

$$\mu(\vartheta_r) = \begin{pmatrix} \mu_1(\vartheta_r) \\ \mu_2(\vartheta_r) \\ \vdots \\ \mu_n(\vartheta_r) \end{pmatrix}\ , \quad \beta(\vartheta_r) = \begin{pmatrix} \beta_0(\vartheta_r) \\ \beta_1(\vartheta_r) \\ \vdots \\ \beta_{p-1}(\vartheta_r) \end{pmatrix}\ , \quad D_r = \left(D_{ij}^{(r)}\right)_{n\times p}\ .$$

In accordance with Bühlmann and Gisler (1997), we assume that

a) $\quad E[\mathrm{Cov}(\mathbf{X}_r|\vartheta_r)] \ = \ \Sigma_r$ is diagonal,

more precisely

$$\Sigma_r \ = \ \begin{pmatrix} \frac{\sigma^2}{V_1^{(r)}}\ , & 0 & , \ 0 & , \ldots, & 0 \\ 0 & , \ \frac{\sigma^2}{V_2^{(r)}} & , \ 0 & , \ldots, & 0 \\ & , & , & , & , \\ 0 & , \ 0 & , \ldots & , \ldots, & \frac{\sigma^2}{V_n^{(r)}} \end{pmatrix}\ ,$$

where $V_i^{(r)}$ is a volume measure of risk r in period i.

b) $\quad D_r$ has orthogonal columns, i.e.,

$$\frac{1}{V_\bullet^{(r)}} \sum_{i=1}^{n} D_{ik}^{(r)} D_{ij}^{(r)} V_i^{(r)} \ = \ 0 \quad \text{for } k \neq j\,,$$

where $V_\bullet^{(r)} \ = \ \sum_{i=1}^{n} V_i^{(r)}$.

Remark. Issue b) can always be achieved by an appropriate reparametrization, but possibly not simultaneously for all risks.

2. The General Problem and the General Philosophy

Let us ask the question how to possibly **reduce**
the set of regression parameters $\{\beta_0(\vartheta_r)\,,\ \beta_1(\vartheta_r)\,,\ \dots\,,\ \beta_{p-1}(\vartheta_r)\}$
to an "optimal subset" $\{\beta_{j_1}(\vartheta_r)\,,\ \dots\,,\ \beta_{j_m}(\vartheta_r)\}\quad (m \le p)$
or how to find an optimal subset regression model,

$$\mu_i(\vartheta_r) \;=\; \sum_{k=1}^{m} D_{ij_k}^{(r)}\beta_{j_k}(\vartheta_r)\,, \quad i = 1, 2, \dots, n\ .$$

It is worthwhile to remember that in frequentist statistics the fundamental argument for a selection is the following:

a) Each parameter β_{j_k} that we retain needs to be estimated, hence leading to a higher variance of the estimator for $\mu_i(\vartheta_r)\,,\quad i = 1, 2, \dots, n\,.$

b) On the other hand, any relevant parameter that we miss will cause a model bias in the estimator for $\mu_i(\vartheta_r)\,.$

All model selection procedures in frequentist statistics rely on an optimal compromise between a) and b) and on the basic principle that the compromise should be made on the basis of the observed data.

The selection of credibility models is motivated by a similar basic thought process, but there is one fundamental difference. Also in credibility models, each parameter β_{j_k} introduces an uncertainty into the prediction process, but the uncertainty is due to the fact that β_{j_k} is (in the Bayesian sense) considered as a **random variable** with a structural (prior) distribution. In addition for simplicity, we neglect any additional variance effects due to the **estimation** of structural parameters (whenever we take the Empirical Bayes viewpoint). This is heuristically justified by the fact that insurance collectives are typically of large size relative to individual sample sizes.

3. The Case of Simple Linear Regression

¿From here on we present the selection argument in the case of Simple Linear Regression. From a methodological point of view, it is sufficient to understand our basic ideas in this case. We even simplify further by assuming all volumes $V_j^{(r)}$ to be equal to 1.

The **Full Model** that we consider follows the regression equation (whenever there is no confusion, we also omit the index r from here on)

$$\mu_i(\vartheta) \;=\; \beta_0(\vartheta) + (i - \bar{i})\beta_1(\vartheta) \quad (i = 1, 2, \dots, n)\ ,$$

$$\text{where } \bar{i} \;=\; \frac{1}{n}\sum_{i=1}^{n} i \;=\; \frac{n+1}{2}$$

and for the observations we have

$$X_i \;=\; \mu_i(\vartheta) + \varepsilon_i(\vartheta) \quad (i = 1, 2, ..., n)$$

with

$$E[\varepsilon_i(\vartheta)|\vartheta] \;=\; 0\,, \quad \text{Cov}\,(\varepsilon_i(\vartheta), \varepsilon_j(\vartheta)|\vartheta) = \delta_{ij}\sigma^2(\vartheta) \quad (i, j = 1, 2, \ldots, n)$$

and structural parameters

$$E\left[\begin{pmatrix} \beta_0(\vartheta) \\ \beta_1(\vartheta) \end{pmatrix}\right] \;=\; \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}\,, \qquad E[\sigma^2(\vartheta)] = \sigma^2\,,$$

$$\text{Cov}\,(\beta(\vartheta)) \;=\; \begin{pmatrix} \tau_0^2 & 0 \\ 0 & \tau_1^2 \end{pmatrix}\,.$$

The **only Submodel** follows

$$\widetilde{\mu}_i(\vartheta) \;=\; \widetilde{\beta_0}(\vartheta)\,, \quad (i = 1, 2, \ldots, n)$$

and we have for the observations

$$X_i \;=\; \widetilde{\mu}_i(\vartheta) + \widetilde{\varepsilon}_i(\vartheta)$$

with

$$E[\widetilde{\varepsilon}_i|\vartheta] = 0\,, \quad \text{Cov}\;(\widetilde{\varepsilon}_i(\vartheta)\,,\,\widetilde{\varepsilon}_j(\vartheta)|\vartheta) \;=\; \delta_{ij}\widetilde{\sigma}^2(\vartheta)$$

and structural parameters

$$E[\widetilde{\beta}_0(\vartheta)] \;=\widetilde{b}_0 \qquad E[\widetilde{\sigma}^2(\vartheta)] = \widetilde{\sigma}^2\,,$$
$$\text{Var}\,[\widetilde{\beta}_0(\vartheta)] \;=\widetilde{\tau}_0^2\,.$$

Observe: Parameters in Full Model: no $\sim$

Parameters in Submodel: with $\sim$

This notation emphasizes the fact that the corresponding parameters may take different values in the Full Model and the Submodel.

The individual least squares estimates from the observations for fixed risk r (suppressing the index r)

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \;=\; \mathbf{X}$$

are the following:

Full Model

$$b_1^X \;=\; \frac{\sum\limits_{i=1}^{n} X_i(i-\bar{i})}{S_{II}} \;, \quad \text{where } S_{II} \;=\; \sum_{i=1}^{n}(i-\bar{i})^2 \;=\; \frac{n(n^2-1)}{12} \;,$$

$$b_0^X \;=\; \overline{X} \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i \;.$$

Submodel

$$\tilde{b}_0^X \;=\; \overline{X} \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i \;.$$

4. The One Collective Approach and its Decision Rule

Given the risks $r = 1, 2, \ldots, N$ and for each risk r the observations

$$\begin{pmatrix} X_{1r} \\ X_{2r} \\ \vdots \\ X_{nr} \end{pmatrix} \;=\; \mathbf{X}_r \;,$$

we pose ourselves the problem how we should model **all** these risks,

either by the Full Model?

or by the Submodel?

This is called the "One Collective Approach" since we **impose** that all risks must be described by the **same** model. In many insurance applications the imposition makes sense. Mostly, the class of those risks that should be considered as one collective is given by a priori reasons (e.g. all private car drivers of a particular geographic region).

We propose to use the following decision rule.

Times of observation: $1, 2, \ldots, n$,

next time point (for which the model should be used) : $n+1$,

$\widehat{\mu}_{n+1}(\vartheta)$: Credibility estimate for next year.

Calculate

$$\frac{1}{N}\sum_{r=1}^{N} E\left[\left(X_{n+1,r} - \widehat{\mu}_{n+1}(\vartheta_r)\right)^2\right] \;=\; \begin{cases} M^2 & \text{under the Full Model ,} \\ \widetilde{M}^2 & \text{under the Submodel .} \end{cases}$$

i.e. for M^2 — E is taken with respect to the probability distribution of the Full Model, $\widehat{\mu}_{n+1}(\vartheta)$ is the credibility estimator derived from the Full Model,

for $\widetilde{M}^2$ — E and $\widehat{\mu}_{n+1}(\vartheta)$ both relate to the Submodel,

and **choose** the Full Model if $M^2 < \widetilde{M}^2$

the Submodel if $\widetilde{M}^2 \leq M^2$.

We take in the following the Empirical Bayes point of view and estimate all structural parameters from **the same data**.

Remark. The reader familiar with selection procedures in frequentist statistics might be wondering why we do not take the probability distribution of the Full Model as well to calculate $\widetilde{M}^2$. As we show in Section 2 of Part II, this would always lead to $M^2 \leq \widetilde{M}^2$ and since we have chosen to neglect errors due to estimates of the structural parameters we would hence **always** choose the Full Model. This is in contrast to model selection in frequentist statistics.

The expected squared losses M^2 and $\widetilde{M}^2$ are of course unknown and have to be estimated from the data.

In the following we derive estimators $\widehat{M}^2$ and $\widehat{\widetilde{M}}^2$. For the latter, as the reader will see, we even propose a corrected version $\widehat{\widetilde{M}}^2_{\text{corr}}$.

We give here in summary our proposed **data–driven** model selection rule in computational form, ready for use.

$$\boxed{\widehat{M}^2 = \widehat{\sigma}^2\left(1 + \frac{\widehat{Z}_0}{n} + \frac{\widehat{Z}_1}{n}3\frac{n+1}{n-1}\right)}\,, \tag{1}$$

$$\widehat{\sigma}^2 = \frac{n}{n-2}S^2, \quad S^2 = \frac{1}{Nn}\sum_{r=1}^{N}\sum_{i=1}^{n}\left(X_{ir} - b_{0r}^X - b_{1r}^X(i-\bar{i})\right)^2 ,$$

$$\widehat{Z}_0 = \frac{n}{n + \frac{\widehat{\sigma}^2}{\widehat{\tau}_0^2}}, \quad \widehat{Z}_1 = \frac{n(n^2-1)}{n(n^2-1) + 12\frac{\widehat{\sigma}^2}{\widehat{\tau}_1^2}},$$

$$\widehat{\tau}_0^2 = \max\left(W_0 - \frac{\widehat{\sigma}^2}{n}, 0\right), \quad W_0 = \frac{1}{N-1}\sum_{r=1}^{N}\left(b_{0r}^X - \frac{1}{N}\sum_{t=1}^{N} b_{0t}^X\right)^2 ,$$

$$\widehat{\tau}_1^2 = \max\left(W_1 - \frac{12\widehat{\sigma}^2}{n(n^2-1)}, 0\right), \quad W_1 = \frac{1}{N-1}\sum_{r=1}^{N}\left(b_{1r}^X - \frac{1}{N}\sum_{t=1}^{N} b_{1t}^X\right)^2 ,$$

$$\boxed{\widehat{\widetilde{M}}^2_{\text{corr}} = \widehat{\widetilde{M}}^2 + \frac{3(n+1)}{n-1}\widehat{B} - \frac{n}{n-1}\left(1 + \frac{\widehat{\widetilde{Z}}_0}{n}\right)\widehat{B}}\,, \tag{2}$$

$$\widehat{\widehat{M}}^2 = \widehat{\widehat{\sigma}}^2 \left(1 + \frac{\widehat{\widehat{Z}}_0}{n}\right) ,$$

$$\widehat{\widehat{\sigma}}^2 = \frac{n}{n-1}\widetilde{S}^2 , \quad \widetilde{S}^2 = \frac{1}{Nn} \sum_{r=1}^{N} \sum_{i=1}^{n} \left(X_{ir} - \widetilde{b}_{0r}^X\right)^2 ,$$

$$\widehat{\widehat{Z}}_0 = \frac{n}{n + \frac{\widehat{\widehat{\sigma}}^2}{\widehat{\widehat{\tau}}_0^2}} ,$$

$$\widehat{\widehat{\tau}}_0^2 = \max\left(W_0 - \frac{\widehat{\widehat{\sigma}}^2}{n}, 0\right) ,$$

$$\widehat{B} = \widetilde{S}^2 - \frac{n-1}{n-2} S^2 .$$

$$\boxed{\begin{array}{ll} \text{Choose} & \\ \text{the Full Model if} & \widehat{M}^2 < \widehat{\widehat{M}}^2_{\mathrm{corr}} \\ \text{Submodel if} & \widehat{\widehat{M}}^2_{\mathrm{corr}} \leq \widehat{M}^2 \end{array}} .$$

The analytical development of this rule and its properties are given in the remaining part of the paper.

5. Explicit Calculations

Assuming that the structural parameters are known, we use the following credibility estimators, cf. Bühlmann and Gisler (1997). In the

Full Model

$$\widehat{\mu}_i(\vartheta) = \widehat{\beta}_0(\vartheta) + (i - \overline{i})\widehat{\beta}_1(\vartheta) \tag{3}$$

where

$$\widehat{\beta}_0(\vartheta) = \underbrace{\frac{n}{n + \frac{\sigma^2}{{\tau_0}^2}}}_{Z_0} \overline{X} + \underbrace{\frac{\frac{\sigma^2}{{\tau_0}^2}}{n + \frac{\sigma^2}{{\tau_0}^2}}}_{1-Z_0} b_0 ,$$

$$\widehat{\beta}_1(\vartheta) = \underbrace{\frac{S_{II}}{S_{II} + \frac{\sigma^2}{{\tau_1}^2}}}_{Z_1} b_1^X + \underbrace{\frac{\frac{\sigma^2}{{\tau_1}^2}}{S_{II} + \frac{\sigma^2}{{\tau_1}^2}}}_{1-Z_1} b_1 ,$$

where

$$S_{II} \;=\; \sum_{i=1}^{n} (i-\bar{i})^2 \;=\; \frac{n(n^2-1)}{12} \;,$$

and in the

Submodel

$$\widehat{\widetilde{\mu}}_i(\vartheta) \;=\; \underbrace{\frac{n}{n+\frac{\widetilde{\sigma}^2}{\widetilde{\tau}_0^2}}}_{\widetilde{Z}_0} \;\overline{X} + \underbrace{\frac{\frac{\widetilde{\sigma}^2}{\widetilde{\tau}_0^2}}{n \;+\; \frac{\widetilde{\sigma}^2}{\widetilde{\tau}_0^2}}}_{1-\widetilde{Z}_0} \;\widetilde{b}_0 \;, \tag{4}$$

Observe that, as we estimate in both models from the same data we have

$$\overline{X} \;=\; b_0^X \;=\; \widetilde{b}_0^X$$

and hence we assume from here on without much loss of generality

$$\beta_0(\vartheta) \;=\; \widetilde{\beta}_0(\vartheta) \;\text{ and }\; b_0 \;=\; \widetilde{b}_0 \,.$$

For the expected loss in the decision rule we get explicitly in the Full Model

$$\begin{aligned}
M^2 \;&=\; \sigma^2 + E\left[\Big(\mu_{n+1}(\vartheta) - \widehat{\mu}_{n+1}(\vartheta)\Big)^2\right] \;, \\
&=\; \sigma^2 + E\left[\Big(\beta_0(\vartheta) - \widehat{\beta_0}(\vartheta)\Big)^2\right] + E\left[\Big(\beta_1(\vartheta) - \widehat{\beta_1}(\vartheta)\Big)^2\right]\Big(n+1-\bar{i}\Big)^2 \;, \\
&=\; \sigma^2 + (1-Z_0){\tau_0}^2 + (1-Z_1){\tau_1}^2\Big(n+1-\bar{i}\Big)^2 \,, \\
&=\; \sigma^2\left[1 + \frac{1}{n+\frac{\sigma^2}{\tau_0^2}} + \frac{1}{S_{II}+\frac{\sigma^2}{\tau_1^2}}\Big(n+1-\bar{i}\Big)^2\right] \;.
\end{aligned}$$

Hence

$$M^2 \;=\; \sigma^2\left[1 + \frac{Z_0}{n} + \frac{Z_1}{n}\cdot 3\frac{n+1}{n-1}\right] \;. \tag{5}$$

The analogous calculation in the Submodel leads to

$$\widetilde{M}^2 \;=\; \widetilde{\sigma}^2\left[1 + \frac{\widetilde{Z}_0}{n}\right] \;. \tag{6}$$

All that needs to be done is to replace

$$\left.\begin{array}{ll} \sigma^2 \,,\, Z_0 \,,\, Z_1 & \text{by their estimators} \\ \widetilde{\sigma}^2 \,,\, \widetilde{Z}_0 & \text{by their estimators} \end{array}\right\} \quad \text{all from the } \textbf{same} \text{ data,}$$

and to compare the values of the estimated versions of (5) and (6) numerically.

6. Explicit Estimation Formulae

In both models the estimator for $\sigma^2 = E[\sigma^2(\vartheta)]$ or $\widetilde{\sigma}^2$, respectively, is derived from the average of squared residuals. The variance parameters in the collective can then be estimated as variance components of a properly defined statistic. Due to our assumptions made in Sections 1 and 3, the multidimensional estimation procedure for τ_0^2 and τ_1^2 can be separated into two independent one–dimensional procedures.

Full Model

Consider the residuals

$$S_{ir} \;=\; X_{ir} - b_{0r}^X - b_{1r}^X \left(i - \overline{i}\right)$$

and the average of the residual sum of squares

$$\begin{aligned} S_r^2 &= \frac{1}{n} \sum_{i=1}^{n} S_{ir}^2 \,, \\ S^2 &= \frac{1}{N} \sum_{r=1}^{N} S_r^2 \,. \end{aligned}$$

One finds (see Appendix)

$$E\left[S_r^2 | \vartheta\right] \;=\; (n-2)\frac{\sigma^2(\vartheta)}{n} \,, \quad E[S^2] \;=\; (n-2)\frac{\sigma^2}{n} \,.$$

Hence we find the unbiased estimator

$$\boxed{\widehat{\sigma}^2 \;=\; \frac{n}{n-2} S^2} \,. \tag{7}$$

The statistic

$$W_0 \;=\; \frac{1}{N-1} \sum_{r=1}^{N} \left(b_{0r}^X - \frac{1}{N} \sum_{t=1}^{N} b_{0t}^X \right)^2 \tag{8}$$

has expectation

$$E[W_0] \;=\; \tau_0^2 + \frac{\sigma^2}{n} \,.$$

Similarly

$$W_1 \;=\; \frac{1}{N-1} \sum_{r=1}^{N} \left(b_{1r}^X - \frac{1}{N} \sum_{t=1}^{N} b_{1t}^X \right)^2$$

has expectation

$$E[W_1] \;=\; \tau_1^2 + \frac{\sigma^2 12}{n(n^2-1)}$$

which leads to the estimators

$$\boxed{\begin{array}{ll} \widehat{\tau}_0^2 = W_0 - \dfrac{\widehat{\sigma}^2}{n} & = W_0 - \dfrac{S^2}{n-2} \\[2ex] \widehat{\tau}_1^2 = W_1 - \dfrac{\widehat{\sigma}^2 \cdot 12}{n(n^2-1)} & = W_1 - \dfrac{S^2}{n-2} \cdot \dfrac{12}{n^2-1} \end{array}} \; . \tag{9}$$

As usual for estimates of variance components, we set the estimate equal to zero if the right hand side in the formulae (9) should become negative.

Submodel

We proceed analogously with the residuals

$$\begin{aligned} \widetilde{S}_{ir} &= X_{ir} - \widetilde{b}_{0r}^X \; , \\ \widetilde{S}_r^2 &= \frac{1}{n} \sum_{i=1}^{n} \widetilde{S}_{ir}^2 \; , \\ \widetilde{S}^2 &= \frac{1}{N} \sum_{r=1}^{N} \widetilde{S}_r^2 \; , \end{aligned}$$

and obtain

$$E\left[\widetilde{S}_r^2 | \vartheta\right] = \frac{n-1}{n} \widetilde{\sigma}^2(\vartheta) \; , \quad E\left[\widetilde{S}^2\right] = \frac{n-1}{n} \widetilde{\sigma}^2 \; ,$$

which leads to the unbiased estimator

$$\boxed{\widehat{\widetilde{\sigma}}^2 = \frac{n}{n-1} \widetilde{S}^2} \; . \tag{10}$$

$\widetilde{W}_0$ can be taken identical to the statistic in formula (8) since $\widetilde{b}_{0r}^X = b_{0r}^X$. Hence we get

$$\boxed{\widehat{\widetilde{\tau}}_0^2 = W_0 - \frac{\widehat{\widetilde{\sigma}}^2}{n} = W_0 - \frac{\widetilde{S}^2}{n-1}} \; , \tag{11}$$

Again, we set the estimate to zero if the right hand side should become negative.

Remark. If we plug in all these estimators in formulae (7), (9), (10), (11), into the expressions in (5) and (6), we get estimates $\widehat{M}^2$ and $\widehat{\widetilde{M}}^2$, the first being exactly as in (1). We will see in Part II that $\widehat{\widetilde{M}}^2$ can in some sense be improved with $\widehat{\widetilde{M}}_{\mathrm{corr}}^2$ as given in (2). We can hence apply the decision rule as defined in Section 4. Observe that the estimates are not very sensitive to changes in the estimates $\widehat{\tau}_0^2$, $\widehat{\tau}_1^2$, or $\widehat{\widetilde{\tau}}_0^2$. They influence only the credibility weights Z_0, Z_1 or $\widetilde{Z}_0$ and all credibility weights can only vary between 0 and 1.

Part II: Testing the Decision Rule

In this part we want to find out how well the decision rule derived in Part I does perform.

1. Expected Performance: Right versus Wrong Decision

We calculate – approximately – the values of

$$E\left[\widehat{M}^2\right] \qquad \text{and} \qquad E\left[\widehat{\widetilde{M}}^2\right]$$

a) if the Full Model is correct. We denote this expected value by E.

b) if the Submodel is correct. We then denote the expected value by $\widetilde{E}$. Note that in the derivation of (10) and (11), we would now write $\widetilde{E}$ instead of E.

Such a calculation seems too difficult to be done in closed form. Let us therefore only do it approximately by treating the credibility weights – which vary little for different observed values – as if they were constant, i.e., if the structural parameters would be known. The key to our following discussion are the following formulae:

$$\begin{aligned}
\text{i)}\quad & E\left[S^2\right] &=& \frac{n-2}{n}\sigma^2 \,, \\
\text{ii)}\quad & \widetilde{E}\left[S^2\right] &=& \frac{n-2}{n}\widetilde{\sigma}^2 \,, \\
\text{iii)}\quad & E\left[\widetilde{S}^2\right] &=& \frac{n-1}{n}\sigma^2 + E\left[\beta_1^2(\vartheta)\right]\frac{n^2-1}{12} \,, \\
\text{iv)}\quad & \widetilde{E}\left[\widetilde{S}^2\right] &=& \frac{n-1}{n}\widetilde{\sigma}^2 \,.
\end{aligned} \tag{12}$$

These formulae are developed in the Appendix. Hence we have the following situation.

If the Full Model is correct, we get with the estimators in (7) and (10) plugged into (5) and (6)

$$\begin{aligned}
E\left[\widehat{M}^2\right] &\approx \sigma^2\left(1+\frac{Z_0}{n}+\frac{Z_1}{n}\,3\frac{n+1}{n-1}\right) \,, \\
E\left[\widehat{\widetilde{M}}^2\right] &\approx \left(\sigma^2+\frac{n}{n-1}E\left[\beta_1^2(\vartheta)\right]\frac{n^2-1}{12}\right)\left(1+\frac{\widetilde{Z}_0}{n}\right) \,.
\end{aligned} \tag{13}$$

We thus choose "on average" correctly, if $E\left[\beta_1^2(\vartheta)\right]$ is sufficiently different from zero.

If the Submodel is correct we have with the estimators in (7) and (10) plugged into (5) and (6)

$$\widetilde{E}\left[\widehat{M}^2\right] \approx \widetilde{\sigma}^2\left(1+\frac{Z_0}{n}+\frac{Z_1}{n}\,3\frac{n+1}{n-1}\right) ,$$
$$\widetilde{E}\left[\widehat{\widetilde{M}}\right] \approx \widetilde{\sigma}^2\left(1+\frac{\widetilde{Z}_0}{n}\right) .$$

Typically, Z_0 and $\widetilde{Z}_0$, estimated on the same data, will be approximately the same. This means that we will practically always take the right decision.

2. Expected Loss

Whereas we have studied the **frequency** of taking the **right decision** in Section 1, we now want to explicitly measure the **expected squared loss** of our credibility estimators. For this purpose, we define

$$L_{n+1} \;=\; \left(X_{n+1}-\widehat{\mu}_{n+1}(\vartheta)\right)^2$$

where $\widehat{\mu}_{n+1}(\vartheta)$ is the credibility estimator from the Full Model as in formula (3), and

$$\widetilde{L}_{n+1} \;=\; \left(X_{n+1}-\widehat{\widetilde{\mu}}_{n+1}\right)^2$$

where $\widehat{\widetilde{\mu}}_{n+1}(\vartheta)$ is the credibility estimator from the Submodel as in formula (4).

As in the previous section, we denote the expected values by either E or $\widehat{E}$ depending on whether they are taken with respect to the probability law of the Full Model or the Submodel, respectively. We assume in this section that the structural parameters are known.

If the **Full Model** is correct, then we have from formula (5)

$$E\left[L_{n+1}\right] \;=\; \sigma^2\left(1+\frac{Z_0}{n}+\frac{Z_1}{n}\,3\frac{n+1}{n-1}\right) , \tag{14}$$

On the other hand, by straightforward calculation,

$$\begin{aligned} E\left[\widetilde{L}_{n+1}\right] &= E\left[\left(\varepsilon_{n+1}(\vartheta)+\beta_0(\vartheta)+\beta_1(\vartheta)\left[n+1-\bar{i}\right]-\widehat{\widetilde{\beta}}_0(\vartheta)\right)^2\right] \\ &= \sigma^2+\widetilde{Z}_0^2\frac{\sigma^2}{n}+(1-\widetilde{Z}_0)^2\tau_0^2+E\left[\beta_1^2(\vartheta)\right]\left(\frac{n+1}{2}\right)^2 . \end{aligned} \tag{15}$$

If the **Submodel** is correct, one finds by straightforward calculation,

$$\begin{aligned} \widetilde{E}\left[L_{n+1}\right] &= \widetilde{E}\left[\left(\widetilde{\varepsilon}_{n+1}(\vartheta)+\widetilde{\beta}_0(\vartheta)-\widehat{\beta}_0(\vartheta)-\widehat{\beta}_1(\vartheta)\left[n+1-\bar{i}\right]\right)^2\right] \\ &= \widetilde{\sigma}^2+Z_0^2\frac{\widetilde{\sigma}^2}{n}+(1-Z_0)^2\widetilde{\tau}_0^2+E\left[\left(\widehat{\beta}_1(\vartheta)\right)^2\right]\left(\frac{n+1}{2}\right)^2 , \end{aligned} \tag{16}$$

and from formula (6),

$$\widetilde{E}\left[\widetilde{L}_{n+1}\right] \;=\; \widetilde{\sigma}^2\left(1+\frac{\widetilde{Z}_0}{n}\right) \,. \tag{17}$$

Remark. We argue that (14) is always smaller than (15) which proves a statement made already in the remark in Section 4 of Part I. And similarly, (17) is always smaller than (16).

This is seen as follows, looking first at (14) and (15).

Step 1

$$\widetilde{Z}_0^2\frac{\sigma^2}{n} + (1-\widetilde{Z}_0)^2\tau_0^2 \geq (1-Z_0)\tau_0^2 \;=\; \frac{Z_0}{n}\sigma^2 \,.$$

The first inequality holds since, among all possible values $\widetilde{Z}_0$, Z_0 is the one that minimizes the left side and achieves $(1-Z_0)\tau_0^2$.

Step 2

As $\frac{Z_1}{n}\sigma^2 \;=\; (1-Z_1)\frac{S_{II}}{n}\tau_1^2 < \frac{S_{II}}{n}E\left[\beta_1^2(\vartheta)\right]$,

we obviously have,

$$\sigma^2\frac{Z_1}{n}\,3\frac{n+1}{n-1} \;=\; \sigma^2\frac{Z_1}{n}\frac{(\frac{n+1}{2})^2}{\frac{n^2-1}{12}} \;=\; \sigma^2\frac{Z_1}{n}\frac{\left(\frac{n+1}{2}\right)^2 n}{S_{II}} \;<\; E\left[\beta_1^2(\vartheta)\right]\left(\frac{n+1}{2}\right)^2 \,.$$

Step 1 and Step 2 together prove the desired inequality for (14) and (15). The inequality between (16) and (17) is seen by a "Step 1–type" argument as above.

The discussion in the section needs **more insight** in as much we want to ask whether the inequalities just proved are **essential** or **inessential**. We call an inequality **essential** if it stays a strict inequality even if we assume that all estimators use the same correct structural parameters of their underlying probability model. Hence an inessential inequality can only be observed when structural parameters are **estimated.**

Comparison of (14) and (15)

Assuming $\widetilde{Z}_0 \;=\; Z_0$, Step 1 becomes an equality but Step 2 remains an inequality. Hence we have

$$E\left[L_{n+1}\right] \;\ll\; E\left[\widetilde{L}_{n+1}\right] \qquad \textbf{essential}\text{ inequality} \,,$$

Comparison of (16) and (17)

Assuming $Z_0 = \widetilde{Z}_0$, Step 1 becomes an equality and since $Z_1 = 0$, we have $\widehat{\beta}_1(\vartheta) = b_1 = 0$. Hence we have

$$\widetilde{E}\left[\widetilde{L}_{n+1}\right] \lessapprox \widetilde{E}\left[L_{n+1}\right] \qquad \textbf{inessential} \text{ inequality} . \tag{18}$$

The point of the discussion is that from an expected loss point of view, the worst situation is

if we use the Submodel when the Full Model is correct.

3. A Corrected Version of the Estimator for $\widetilde{M}^2$

Comparing formulae (13) and (15), we see that the bias terms are

$$\begin{aligned} \text{Bias in } E\left[\widehat{\widetilde{M}}^2\right] \quad &: \quad E\left[\beta_1^2(\vartheta)\right] \frac{n^2-1}{12} \cdot \frac{n}{n-1}\left(1+\frac{\widetilde{Z}_0}{n}\right) , \\ \text{most important Bias in } E\left[\widetilde{L}_{n+1}\right] \quad &: \quad E\left[\beta_1^2(\vartheta)\right]\left(\frac{n+1}{2}\right)^2 , \end{aligned}$$

where the notion of “most important” stems from the essential inequality in Step 2 above.

One discovers that our decision variable $\widehat{\widetilde{M}}^2$ underestimates this most important bias term in $E\left[\widetilde{L}_{n+1}\right]$. We hence propose a corrected estimator $\widehat{\widetilde{M}}^2_{\text{corr}}$ as given in (2),

$$\boxed{\widehat{\widetilde{M}}^2_{\text{corr}} = \widehat{\widetilde{M}}^2 + 3\tfrac{n+1}{n-1}\,\widehat{B} - \tfrac{n}{n-1}\left(1+\tfrac{\widehat{\widetilde{Z}}_0}{n}\right)\widehat{B}} ,$$

where

$$\widehat{B} = \widetilde{S}^2 - \frac{n-1}{n-2}S^2 \quad \text{with } S^2 , \widetilde{S}^2 , \text{ as in Section 6 of Part I.}$$

We see from formulae (12) that

$$E[\widehat{B}] = E\left[\beta_1^2(\vartheta)\right]\frac{n^2-1}{12} .$$

Hence $E\left[\widehat{\widetilde{M}}^2_{\text{corr}}\right] \approx \sigma^2\left(1+\frac{\widetilde{Z}_0}{n}\right) + E\left[\beta_1^2(\vartheta)\right]\left(\frac{n+1}{2}\right)^2$ has the same bias term as the most important bias in $E[\widetilde{L}_{n+1}]$. In addition, we also see from formulae (12) that

$$\widetilde{E}[\widehat{B}] = 0 ,$$

which means that if the Submodel is true, then $\widetilde{E}\left[\widehat{\widetilde{M}}^{2}_{\text{corr}}\right] \approx \widetilde{E}\left[\widehat{\widetilde{M}}^{2}\right]$. It is clear that $\widehat{\widetilde{M}}^{2}_{\text{corr}}$ has an additional but inessential noise term. The correction should nevertheless improve our decision as the **essential** bias term is more important.

Remark. In practical insurance applications it is very important to detect the trend if the Full Model is true. Our selection criterion would miss such a trend "on average" if $E\left[\widehat{\widetilde{M}}^{2}_{\text{corr}}\right] \leq E\left[\widehat{M}^2\right]$ which is essentially equivalent to $E\left[\beta_1^2(\vartheta)\right] = \tau_1^2 + b_1^2 \leq \sigma^2 \frac{12 Z_1}{n(n+1)(n-1)}$. But this is only possible if $\tau_1^2 + b_1^2$ is very small, i.e., if the Full Model is almost identical to the Submodel.

4. Simulation Results

We analyze our decision rule

a) if the Full Model is correct,

b) if the Submodel is correct.

In both situations, we want to find out

a) how often we make the right decision on average, i.e., the probability of making the right decision.

b) how big the incurred loss turns out on average:

$$L^*_{n+1} = \begin{cases} L_{n+1} \; if \;\; \widehat{M}^2 < \widehat{\widetilde{M}}^{2}_{\text{corr}} \\ \widetilde{L}_{n+1} \; if \;\; \widehat{\widetilde{M}}^{2}_{\text{corr}} \leq \widehat{M}^2 \end{cases} .$$

This is the average loss when using our selection criterion with (1) and (2).

c) how b) compares with the minimal loss if we choose **always** the correct model.

An important fact to note is that all parameters appearing in our decision variables are always **estimated** from the collective data: we are thus simulating the relevant practical situation, i.e., both past and future observations $X_1, \ldots, X_n$ and X_{n+1}, respectively, for each simulated individual risk. This allows us to test our pragmatic working hypothesis which we have used to **theoretically develop** our model selection criterion, namely that estimation errors in the structural parameters are of minor importance.

In all the simulations we have restricted ourselves to normally distributed random variables and parameters. We always use $\sigma^2(\vartheta) \equiv \sigma^2 = 1$, but vary the other parameters. The results are always based on 100 independent simulations each with $N(n+1)$ values of a given case (i.e., given values of structural parameters). The most interesting case to study is

$$\textit{Full Model} \quad \textit{correct} \quad \textit{with} \quad b_1 = 0 \ ,$$

that is, on average we have no trend but individually, there is a trend. A global frequentist analysis would miss the individual trends: in insurance, of course, anti–selection could happen in this case! Figure 1 shows eight individual realizations of the Full Model with $b_0 = 1$, $b_1 = 0$, $\tau_0 = 1$, $\tau_1 = 1/4$, $\sigma^2(\vartheta) \equiv 1$ and $n = 10$. It is visually very difficult to decide about a relevant trend for this data. However, in terms of expected loss, the fitted models with and without trend differ by more than a factor 2 (see Table 2) and our model selection criterion finds the better (the correct) model without a single misclassification error.

Figure 1 about here

We denote by Ave the simulated versions of the exceptions operator E (either E in the Full Model or $\widetilde{E}$ in the Submodel) and by MCR the misclassification rate

$$MCR \ = \ 1 - \ \text{rate of correctly identified models.}$$

In Tables 1–3 we show results when the Full Model is correct. We conclude the following:

a) The incurred average loss $Ave(L^*_{n+1})$ is at least as small as the average loss under the correct Full Model $Ave(L_{n+1})$ (there are two exceptions where $Ave(L^*_{n+1}) \ = \ Ave(L_{n+1}) + 0.01$, thus a negligible difference).

b) The misclassification rate MCR is low or even zero if the squared average loss of the correct Full Model $Ave(L_{n+1})$ is sufficiently smaller than $Ave(\widetilde{L}_{n+1})$ of the Submodel. In the cases where the misclassification rate is high, the incorrect Submodel is about as good as the correct Full Model: in these cases, misclassification does not necessarily imply a bad model choice with respect to the squared error loss.

c) The correct Full Model does not necessarily always yield the smaller loss: this because we are **estimating** structural parameters from the data (the analogy to model selection in frequentist statistics then applies). However, the possible gains in terms of expected loss by using an incorrect (but approximately correct) and lower dimensional model are in the case of simple regression only of small order. This might be different if a general regression model with many covariates is studied.

d) $\widehat{\widetilde{M}}^2_{\text{corr}}$ and $\widehat{M}^2$ are exhibiting only little bias for estimating $Ave(\widetilde{L}_{n+1})$ and $Ave(L_{n+1})$, respectively: even though the structural parameters are **estimated** and exact unbiasedness is not true anymore.

Tables 1 - 3 about here

In Table 4 we show one situation where the Submodel is correct. We only report the case with individual sample size $n = 10$; the cases with $n = 5$ and $n = 20$ gave similar results. We conclude for this situation the following. The average squared error losses in the Full Model and Submodel are similar and the choice of a model does not influence the accuracy of a predicted future value very much. This is consistent with the fact that

$$\widetilde{E}\left[\widetilde{L}_{n+1}\right] \overset{<}{\approx} \widetilde{E}\left[L_{n+1}\right]$$

is only an **inessential** inequality, see (18). But we would expect a bigger difference between $Ave(\widetilde{L}_{n+1})$ and $Ave(L_{n+1})$ when the difference in dimensionality between the Full Model and the Submodel would be much bigger than one, because of the effect of **estimated** structural parameters.

Table 4 about here

Summarizing the whole simulation study: in the case where the underlying model is unknown (which is always true in practice), the average loss $Ave(L^*_{n+1})$ in the model **chosen** by our criterion is very close to the optimal loss, which is most often, but not always, achieved with the correct unknown model.

Conclusions

We have developed a data–driven rule for choosing between a location and a linear trend credibility model. The question to decide whether there is or is not a linear trend does arise in practical insurance problems. Our proposed solution works very well, from a theoretical point of view and also when applied to simulated data. More than that, we believe that the general strategy which we have developed here will also be useful and successful in many other credibility models. For example, the general regression credibility model with different volumes, cf. Hachemeister (1975), the hierarchical models, cf. Jewell (1975) and Taylor (1979), or hierarchical regression models, cf. Sundt (1979) and Norberg (1986).

Appendix

In Section 6 of Part I we have defined for the

Full Model

$$S_{ir} \;=\; X_{ir} - b_{0r}^X - b_{1r}^X(i-\bar{i}) \qquad (i=1,2,\ldots,n)\ ,$$
$$S_r^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} S_{ir}^2 \ \text{and}\ S^2 = \frac{1}{N}\sum_{r=1}^{N} S_r^2\ ,$$

and for the

Submodel

$$S_{ir} \;=\; X_{ir} - b_{0r}^X\ ,$$
$$\widetilde{S}_r^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} \widetilde{S}_{ir}^2 \ \text{and}\ \widetilde{S}^2 \;=\; \frac{1}{N}\sum_{r=1}^{N} \widetilde{S}_r^2\ .$$

Before starting with our calculations it is worthwhile to observe the following table for the individual estimates:

Full Model

$$b_1^X \;=\; \beta_1(\vartheta) + \frac{\sum\limits_{j=1}^{n} \varepsilon_j (j-\bar{j})}{S_{II}}\ , \qquad b_0^X \;=\; \beta_0(\vartheta) + \frac{\sum\limits_{j=1}^{n} \varepsilon_j}{n}\ .$$

Submodel

$$b_1^X \;=\; \frac{\sum\limits_{j=1}^{n} \widetilde{\varepsilon}_j (j-\bar{j})}{S_{II}}\ , \qquad b_0^X \;=\; \beta_0(\vartheta) + \frac{\sum\limits_{j=1}^{n} \widetilde{\varepsilon}_j}{n}\ .$$

Expected Values for S_r^2

If the **Full Model** is correct we have for $i=1,2,\ldots,n$,

$$\begin{aligned} X_i - b_0^X - b_1^X(i-\bar{i}) \;&=\; \varepsilon_i - \frac{\sum\limits_{j=1}^{n}\varepsilon_j}{n} - \frac{\sum\limits_{j=1}^{n}\varepsilon_j(j-\bar{j})}{S_{II}}(i-\bar{i}) \;= \\ &=\; \varepsilon_i - \eta_0 - \eta_1(i-\bar{i}) \end{aligned}$$

and hence for the residual sum of squares (observe $\sum\limits_i \eta_0\eta_1(i-\bar{i}) = 0$)

$$\sum_{i=1}^{n}\Big(\varepsilon_i - \eta_0 - \eta_1(i-\bar{i})\Big)^2 = \sum_{i=1}^{n}\Big(\varepsilon_i^2 + \eta_0^2 + \eta_1^2(i-\bar{i})^2 - 2\varepsilon_i\eta_0 - 2\varepsilon_i\eta_1(i-\bar{i})\Big)\ .$$

One easily finds for the five sums on the right

$$E\left[\sum_{i=1}^{n} \varepsilon_i^2 \,|\, \vartheta\right] \;=\; n\sigma^2(\vartheta) \;,$$

$$E\left[\sum_{i=1}^{n} \eta_0^2 \,|\, \vartheta\right] \;=\; n \cdot \frac{\sigma^2(\vartheta)}{n} \;=\; \sigma^2(\vartheta) \;,$$

$$E\left[\sum_{i=1}^{n} \eta_1^2 (i-\bar{\imath})^2 \,|\, \vartheta\right] \;=\; \sum_{i=1}^{n} \frac{\sigma^2(\vartheta)}{S_{II}} (i-\bar{\imath})^2 \;=\; \sigma^2(\vartheta) \;,$$

$$-2E\left[\sum_{i=1}^{n} \varepsilon_i \eta_0 \,|\, \vartheta\right] \;=\; -2n\frac{\sigma^2(\vartheta)}{n} \;=\; -2\sigma^2(\vartheta) \;,$$

$$-2E\left[\sum_{i=1}^{n} \varepsilon_i \eta_1 (i-\bar{\imath}) \,|\, \vartheta\right] = -2\sum_{i=1}^{n} \frac{\sigma^2(\vartheta)}{S_{II}} (i-\bar{\imath})^2 = -2\sigma^2(\vartheta) \;.$$

Hence summing the five terms on the right side and dividing by n,

$$E\left[S_r^2 \,|\, \vartheta\right] \;=\; \frac{1}{n} E\left[\sum_{i=1}^{n} \left(X_i - b_0^X - b_1^X (i-\bar{\imath})\right)^2 |\, \vartheta\right] \;=\; \frac{n-2}{n}\sigma^2(\vartheta) \;. \qquad (A)$$

If the **Submodel** is correct, the calculations are formally exactly the same, $\widetilde{\varepsilon}_i$ taking the role of ε_i and $\widetilde{\sigma}^2(\vartheta)$ the role of $\sigma^2(\vartheta)$

$$\widetilde{E}\left[S_r^2 \,|\, \vartheta\right] \;=\; \frac{n-2}{n}\widetilde{\sigma}^2(\vartheta) \;. \qquad (B)$$

Expected Values for $\widetilde{S}_r^2$

If the **Full Model** is correct, we have for $i = 1, 2, \ldots, n$,

$$X_i - b_0^X \;=\; \beta_1(\vartheta)(i-\bar{\imath}) + \varepsilon_i - \frac{\sum\limits_{j=1}^{n} \varepsilon_j}{n} \;.$$

Hence

$$E\left[\left(X_i - b_0^X\right)^2 |\, \vartheta\right] \;=\; \beta_1^2(\vartheta)(i-\bar{\imath})^2 + E\left[(\varepsilon_i - \eta_0)^2 \,|\, \vartheta\right] \;.$$

$$E\left[\widetilde{S}_r^2 \,|\, \vartheta\right] \;=\; \frac{1}{n}\sum_{i=1}^{n} E\left[\left(X_i - b_0^X\right)^2 |\, \vartheta\right] \;=\; \beta_1^2(\vartheta)\frac{S_{II}}{n} + \frac{n-1}{n}\sigma^2(\vartheta) \;. \qquad (C)$$

If the **Submodel** is correct, one finds also very easily

$$\widetilde{E}\left[\widetilde{S}_r^2 \,|\, \vartheta\right] \;=\; \frac{1}{n}\sum_{i=1}^{n} \widetilde{E}\left[(\widetilde{\varepsilon}_i - \widetilde{\eta}_0)^2 \,|\, \vartheta\right] \;=\; \frac{n-1}{n}\widetilde{\sigma}^2(\vartheta) \;. \qquad (D)$$

Formulae (7) and (10) in Part I and formula (12) in Part II directly follow from formulae (A), (B), (C), and (D).

Acknowledgments: We thank Alois Gisler for constructive discussions.

References

[1] Akaike, H. (1970). Statistical predictor identification. Annals of the Institute of Statistical Mathematics 22, 203–217.

[2] Bühlmann, H. and Gisler, A. (1997). Credibility in the regression case revisited (A late tribute to Charles A. Hachemeister). ASTIN Bulletin 27, 83–98.

[3] Hachemeister, C.A. (1975). Credibility for regression models with application to trend. In Credibility, theory and application, Proc. of the Berkeley Actuarial Research Conference on Credibility, pp. 129–163. Academic Press.

[4] Jewell, W.S. (1975). The use of collateral data in credibility theory: a hierarchical model. Giornale dell'Instituto Italiano degli Attuari 38, 1–16.

[5] Linhart, H. and Zucchini, W. (1986). Model Selection. Wiley.

[6] Mallows, C.L. (1973). Some comments on C_p. Technometrics 15, 661–675.

[7] Norberg, R. (1986). Hierarchical credibility: Analysis of a random effect linear model with nested classification. Scandinavian Actuarial Journal, 204–222.

[8] Sundt, B. (1979). A hierarchical regression credibility model. Scandinavian Actuarial Journal, 107–114.

[9] Taylor, G.C. (1979). Credibility analysis of a general hierarchical model. Scandinavian Actuarial Journal, 1–12.

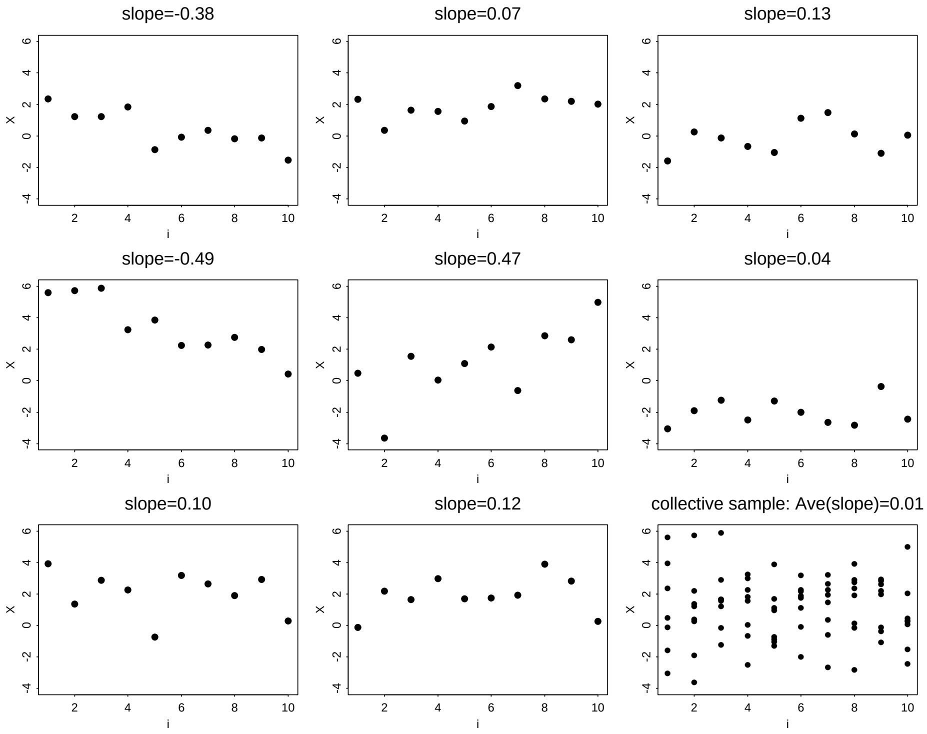

Figure 1: Eight individual samples and their joint representation as a collective sample: $n = 10, b_0 = 1, b_1 = 0, \tau_0 = 1, \tau_1 = 1/4, \sigma^2(\vartheta) \equiv 1$. The realized individual slopes $\beta_1(\vartheta_r)$ $(r = 1, \ldots, 8)$ and their average respectively are given on top of each picture.

τ_1, N	MCR	$Ave(L^*_{n+1})$	$Ave(\widetilde{L}_{n+1})$	$Ave(L_{n+1})$	$Ave\left(\widehat{\widetilde{M}}^2_{\mathrm{corr}}\right)$	$Ave(\widehat{M}^2)$
1, 100	0	1.92	8.96	1.92	10.26	1.97
1/2, 100	0	1.75	3.23	1.75	3.44	1.80
1/4, 100	0.01	1.49	1.67	1.49	1.74	1.50
1/8, 100	0.36	1.28	1.28	1.28	1.31	1.28
1/16, 100	0.53	1.19	1.18	1.20	1.20	1.23
1/32, 100	0.62	1.17	1.16	1.17	1.17	1.21
1, 20	0	2.02	9.02	2.02	9.92	2.00
1/2, 20	0	1.90	3.28	1.90	3.26	1.76
1/4, 20	0.22	1.65	1.72	1.64	1.63	1.43
1/8, 20	0.60	1.37	1.32	1.40	1.24	1.28
1/16, 20	0.64	1.28	1.22	1.31	1.15	1.25
1/32, 20	0.67	1.25	1.19	1.29	1.13	1.24

Table 1. Full Model is correct: $n = 5$, $b_0 = 1$, $b_1 = 0$, $\tau_0 = 1$.

τ_1, N	MCR	$Ave(L^*_{n+1})$	$Ave(\widetilde{L}_{n+1})$	$Ave(L_{n+1})$	$Ave\left(\widehat{\widetilde{M}}^2_{\mathrm{corr}}\right)$	$Ave(\widehat{M}^2)$
1, 100	0	1.42	27.77	1.42	31.23	1.45
1/2, 100	0	1.41	8.29	1.41	8.61	1.44
1/4, 100	0	1.37	2.93	1.37	2.96	1.40
1/8, 100	0	1.28	1.55	1.28	1.56	1.29
1/16, 100	0.08	1.18	1.20	1.18	1.20	1.17
1/32, 100	0.50	1.12	1.12	1.12	1.17	1.12
1, 20	0	1.40	28.36	1.40	31.16	1.47
1/2, 20	0	1.39	8.43	1.39	8.56	1.45
1/4, 20	0	1.36	2.97	1.36	2.93	1.40
1/8, 20	0.04	1.30	1.56	1.29	1.53	1.28
1/16, 20	0.40	1.18	1.19	1.18	1.19	1.17
1/32, 20	0.66	1.11	1.09	1.12	1.11	1.14

Table 2. Full Model is correct: $n = 10$, $b_0 = 1$, $b_1 = 0$, $\tau_0 = 1$.

τ_1, N	MCR	$Ave(L^*_{n+1})$	$Ave(\widetilde{L}_{n+1})$	$Ave(L_{n+1})$	$Ave\left(\widetilde{\widehat{M}}^2_{\text{corr}}\right)$	$Ave(\widehat{M}^2)$
1, 100	0	1.22	104.27	1.22	113.65	1.21
1/2, 100	0	1.22	28.22	1.22	29.19	1.21
1/4, 100	0	1.21	8.10	1.21	8.08	1.21
1/8, 100	0	1.20	2.80	1.20	2.81	1.20
1/16, 100	0	1.17	1.49	1.17	1.49	1.17
1/32, 100	0	1.12	1.16	1.12	1.16	1.11
1, 20	0	1.25	104.17	1.25	113.07	1.22
1/2, 20	0	1.25	28.00	1.25	29.08	1.21
1/4, 20	0	1.25	7.99	1.25	8.07	1.21
1/8, 20	0	1.24	2.82	1.24	2.81	1.20
1/16, 20	0	1.21	1.52	1.21	1.50	1.16
1/32, 20	0.16	1.15	1.19	1.15	1.16	1.11

Table 3. Full Model is correct: $n = 20$, $b_0 = 1$, $b_1 = 0$, $\tau_0 = 1$.

τ_0, N	MCR	$Ave(L^*_{n+1})$	$Ave(\widetilde{L}_{n+1})$	$Ave(L_{n+1})$	$Ave\left(\widetilde{\widehat{M}}^2_{\text{corr}}\right)$	$Ave(\widehat{M}^2)$
1, 100	0.38	1.12	1.11	1.12	1.09	1.11
1/2, 100	0.38	1.09	1.09	1.10	1.07	1.09
1/4, 100	0.37	1.05	1.05	1.06	1.04	1.05
1/8, 100	0.38	1.03	1.02	1.03	1.02	1.03
1/16, 100	0.38	1.02	1.01	1.02	1.01	1.02
1/32, 100	0.38	1.02	1.01	1.02	1.01	1.02
1, 20	0.33	1.18	1.17	1.18	1.08	1.11
1/2, 20	0.33	1.16	1.15	1.17	1.06	1.09
1/4, 20	0.33	1.13	1.11	1.13	1.03	1.06
1/8, 20	0.33	1.10	1.08	1.10	1.01	1.04
1/16, 20	0.33	1.09	1.07	1.09	1.00	1.03
1/32, 20	0.33	1.08	1.07	1.09	1.00	1.03

Table 4. Submodel is correct: $n = 10$, $b_0 = 1$.