Estimation of conditional mean squared error of prediction for claims triangle reserving

Mathias Lindholm, Stockholm University

London October 10, 2018

The presented material is based on the preprint [6] co-authored with

- Filip Lindskog
- Felix Wahl

Aim with the talk:

- Provide a unified approach to conditional MSEP calculations
- Provide a computable "distribution-free" MSEP formula assuming only knowledge of lower order moments
- Give a number of examples on how to apply the method

Outline of the talk:

- Introduce Mack's definition of conditional MSEP and the general problem with frequentist "estimation error"
- Introduce a randomisation device for assessing the estimation error – conditional vs unconditional specification
- Derive a simple (semi-) analytical approximation formula for calculating conditional MSEP which only relies on lower order moments
- Examples: Mack's C-L, ultimo and CDR(k) calculations, ODP C-L etc.

If time allows:

- Relate different notions of parameter uncertainty – frequentist "estimation error" vs Bayesian "parameter uncertainty"
- Discuss model selection for auto-regressive reserving models – FPE and Mallows' C_p

Conditional MSEP

- Assume that we have a stochastic process S_t parametrised in terms of a parameter vector $\boldsymbol{\theta}$
- Let $\mathcal{F}_0 = \sigma\{S_t; t \leq 0\}$ denote the σ-algebra containing the information known up until today generated by the model S_t
- Let X denote a random quantity of interest which is a function of the evolution of S_t for $t > 0$
- A natural predictor of X is given by

$$\widehat{X} := h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0),$$

  where $h(\boldsymbol{\theta}; \mathcal{F}_0) := \mathbb{E}[X \mid \mathcal{F}_0]$, and where $\widehat{\boldsymbol{\theta}}$ corresponds to some $\mathcal{F}_0$-measurable estimator
- $\widehat{X}$ is often referred to as a "plug-in" estimator of X
- We will refer to the function $\boldsymbol{z} \mapsto h(\boldsymbol{z}; \mathcal{F}_0)$ as the "basis of prediction"

Conditional MSEP

Definition 1 (Conditional MSEP)

$$\mathrm{MSEP}_{\mathcal{F}_0}(X, \widehat{X}) := \mathbb{E}[(X - \widehat{X})^2 \mid \mathcal{F}_0]$$

Note that Definition 1 is an L^2-distance between a stochastic variable and its plug-in predictor.

We will only consider MSEP-calculations w.r.t. Definition 1

Conditional MSEP

Note that Definition 1 may be rewritten according to

$$\begin{aligned}
\mathrm{MSEP}_{\mathcal{F}_0}(X,\widehat{X}) &:= \mathrm{Var}(X \mid \mathcal{F}_0) + \mathbb{E}[(\mathbb{E}[\widehat{X} \mid \mathcal{F}_0] - \mathbb{E}[X \mid \mathcal{F}_0])^2 \mid \mathcal{F}_0] \\
&= \mathrm{Var}(X \mid \mathcal{F}_0) + \mathbb{E}[(h(\widehat{\boldsymbol{\theta}};\mathcal{F}_0) - h(\boldsymbol{\theta};\mathcal{F}_0))^2 \mid \mathcal{F}_0] \\
&= \mathrm{Var}(X \mid \mathcal{F}_0) + (h(\widehat{\boldsymbol{\theta}};\mathcal{F}_0) - h(\boldsymbol{\theta};\mathcal{F}_0))^2.
\end{aligned}$$

Moreover, note that

- "process variance": $\mathrm{Var}(X \mid \mathcal{F}_0) \geq 0$
- "estimation error": $\mathbb{E}[(\mathbb{E}[\widehat{X} \mid \mathcal{F}_0] - \mathbb{E}[X \mid \mathcal{F}_0])^2 \mid \mathcal{F}_0] \geq 0$
- if we let $H(\boldsymbol{\theta};\mathcal{F}_0) := \mathrm{MSEP}_{\mathcal{F}_0}(X,\widehat{X})(\boldsymbol{\theta})$, it follows that

$$\widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(X,\widehat{X}) := H(\widehat{\boldsymbol{\theta}};\mathcal{F}_0) = \widehat{\mathrm{Var}}(X \mid \mathcal{F}_0),$$

i.e. the estimation error is 0 – clearly an underestimation

Conditional MSEP

Although the plug-in process variance may be hard to calculate in practice, it is usually well defined

We will henceforth focus on the plug-in estimation error

Calculating conditional MSEP

- **Suggested solution:** introduce a randomisation device
- That is, replace the $\mathcal{F}_0$-measurable $\widehat{\boldsymbol{\theta}}$ with a randomised version $\widehat{\boldsymbol{\theta}}^*$ which is **not** $\mathcal{F}_0$-measurable
- $\widehat{\boldsymbol{\theta}}^*$ should be chosen to share key properties with the original estimator $\widehat{\boldsymbol{\theta}}$
- In this way we only estimate how the parameter uncertainty from $\widehat{\boldsymbol{\theta}}^*$ propagates through $\boldsymbol{z} \mapsto h(\boldsymbol{z}; \mathcal{F}_0)$ – **keeping the basis of prediction fix**

Calculating conditional MSEP

- The specification of $\widehat{\theta}^*$ should be based on the behaviour of $\widehat{\theta}$
- Recall that our observed data is generated by the process S_t
- Let $S_t^{\perp}$ correspond to an independent copy of S_t
  - $S_t^{\perp}$ can be thought of as a "parallel universe"
- It is then natural to assume that $\widehat{\theta}^*$ is defined in terms of $S_t^{\perp}$ and possibly S_t

Calculating conditional MSEP

Regarding the construction of $\widehat{\boldsymbol{\theta}}^*$ we will make the following assumption:

Assumption 1

Assume that

$$X \perp \widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0$$

Definition 2

$$\begin{aligned}\mathrm{MSEP}^*_{\mathcal{F}_0}(X, \widehat{X}) &:= \mathbb{E}[(X - \widehat{X}^*)^2 \mid \mathcal{F}_0] \\ &= \mathrm{Var}(X \mid \mathcal{F}_0) + \mathbb{E}[(h(\widehat{\boldsymbol{\theta}}^*; \mathcal{F}_0) - h(\boldsymbol{\theta}; \mathcal{F}_0))^2 \mid \mathcal{F}_0]\end{aligned}$$

Note that Assumption 1 is fulfilled if we let $\widehat{\boldsymbol{\theta}}^* = \widehat{\boldsymbol{\theta}}^\perp$, i.e. only resample based on $S_t^\perp$

Calculating conditional MSEP

Before we proceed any further, we will recall the definition of the "final prediction error" (FPE) introduced by Akaike in [1, 2]:

- Let $\mathcal{F}_0^\perp$ denote the filtration generated by $S_t^\perp$, and let $X^\perp$ denote the $S_t^\perp$ analog of X
- The **unconditional** FPE of $\widehat{X}$ is then given by

$$\text{FPE}(X, \widehat{X}) := \mathbb{E}[(X^\perp - h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0^\perp))^2],$$

  where $\widehat{\boldsymbol{\theta}}$ is calculated based on S_t being $\mathcal{F}_0$-measurable, but not $\mathcal{F}_0^\perp$-measurable
- Note that due to the exchangeability between S_t and $S_t^\perp$ it follows that

$$\text{FPE}(X, \widehat{X}) := \mathbb{E}[(X - h(\widehat{\boldsymbol{\theta}}^\perp; \mathcal{F}_0))^2]$$

Calculating conditional MSEP

- The **conditional** FPE of $\widehat{X}$ is then given by

$$\begin{aligned}\mathrm{FPE}_{\mathcal{F}_0^\perp}(X,\widehat{X}) &:= \mathbb{E}[(X^\perp - h(\widehat{\boldsymbol{\theta}};\mathcal{F}_0^\perp))^2 \mid \mathcal{F}_0^\perp] \\ &= \mathbb{E}[(X - h(\widehat{\boldsymbol{\theta}}^\perp;\mathcal{F}_0))^2 \mid \mathcal{F}_0]\end{aligned}$$

- That is, given that $\widehat{\boldsymbol{\theta}}^* = \widehat{\boldsymbol{\theta}}^\perp$ it holds that

$$\mathrm{MSEP}^*_{\mathcal{F}_0}(X,\widehat{X}) = \mathrm{FPE}_{\mathcal{F}_0}(X,\widehat{X}),$$

which is a well studied object for model selection for auto-regressive time series models!

Calculating conditional MSEP

- If we are interested in e.g. "distribution-free" reserving models the unconditional behaviour of $\widehat{\boldsymbol{\theta}}$, and hence $\widehat{\boldsymbol{\theta}}^{\perp}$, is not always computationally tractable
- Akaike circumvented this issue for auto-regressive time series models, by taking the limit of the sample size
- For conditional models an alternative is to use a conditional specification of $\widehat{\boldsymbol{\theta}}^{*}$ which differs from $\widehat{\boldsymbol{\theta}}^{\perp}$
- Moreover, we are often interested in situations where we only make assumptions on moments, not specifying the full distribution of $\widehat{\boldsymbol{\theta}}$ (and hence $\widehat{\boldsymbol{\theta}}^{*}$) completely

Calculating conditional MSEP

If we make **a single** first-order Taylor expansion of $h(\boldsymbol{\theta}; \mathcal{F}_0)$ we get

$$\begin{aligned}\mathrm{MSEP}^{*,\nabla}_{\mathcal{F}_0}(X, \widehat{X}) := \; & \mathrm{Var}(X \mid \mathcal{F}_0) \\ & + \nabla h(\boldsymbol{\theta}; \mathcal{F}_0)' \mathbb{E}[(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta})^2 \mid \mathcal{F}_0] \nabla h(\boldsymbol{\theta}; \mathcal{F}_0)\end{aligned}$$

Note that $\mathrm{MSEP}^{*,\nabla}_{\mathcal{F}_0}(X, \widehat{X})$ is

- expressed in terms of conditional moments $\Rightarrow$ "distribution-free"
- based on a **single** first-order Taylor expansion

Calculating conditional MSEP

We will henceforth

- focus on the calculation of $\mathrm{MSEP}^{*,\nabla}_{\mathcal{F}_0}(X, \widehat{X})$
- assume that $\nabla h(\boldsymbol{\theta}; \mathcal{F}_0)$ exists
- **make no further approximations!!**

We will continue with the specification of $\widehat{\boldsymbol{\theta}}^*$ in terms of assumptions on moments of $\widehat{\boldsymbol{\theta}}$

Specification of $\widehat{\boldsymbol{\theta}}^*$: moment assumptions

Assumption 2 (Unconditional structure)

Assume that

$$\mathbb{E}[\widehat{\boldsymbol{\theta}}] = \boldsymbol{\theta},$$
$$\mathrm{Cov}(\widehat{\boldsymbol{\theta}}) < \infty \quad a.s.$$

Note that given Assumption 2 it follows that

$$\mathbb{E}[\widehat{\boldsymbol{\theta}}^{\perp}] = \mathbb{E}[\widehat{\boldsymbol{\theta}}] = \boldsymbol{\theta},$$
$$\mathrm{Cov}(\widehat{\boldsymbol{\theta}}^{\perp}) = \mathrm{Cov}(\widehat{\boldsymbol{\theta}}) < \infty \quad a.s.$$

Specification of $\widehat{\boldsymbol{\theta}}^*$: moment assumptions

Assumption 3 (Conditional structure)

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p)$ and let $\mathcal{G}_k \subset \mathcal{F}_0$, $k = 1, \ldots, p$ and assume that $\mathcal{G}_j \subset \mathcal{G}_k, \quad j < k$ together with

$$\mathbb{E}[\widehat{\boldsymbol{\theta}}_k \mid \mathcal{G}_k] = \boldsymbol{\theta}_k,$$
$$C_{j,k}(\widehat{\boldsymbol{\theta}}) := \operatorname{Cov}(\widehat{\boldsymbol{\theta}}_j, \widehat{\boldsymbol{\theta}}_k \mid \mathcal{G}_j, \mathcal{G}_k) < \infty,$$

holds for all $j, k = 1, \ldots, p$ and that

$$\chi(\boldsymbol{\theta}; \mathcal{F}_0) := \begin{bmatrix} C_{1,1}(\widehat{\boldsymbol{\theta}}) & C_{1,2}(\widehat{\boldsymbol{\theta}}) & \ldots & C_{1,p}(\widehat{\boldsymbol{\theta}}) \\ C_{1,2}(\widehat{\boldsymbol{\theta}}) & \ddots & & \\ \vdots & & & \\ C_{1,p}(\widehat{\boldsymbol{\theta}}) & & & C_{p,p}(\widehat{\boldsymbol{\theta}}) \end{bmatrix},$$

corresponds to a proper covariance matrix.

Specification of $\widehat{\boldsymbol{\theta}}^*$: moment assumptions

Note that the distribution-free CL model fulfil Assumption 3

Given Assumption 3, let $\widehat{\boldsymbol{\theta}}^{*,c}$ be defined as a randomised version of $\widehat{\boldsymbol{\theta}}$ so that

$$\mathbb{E}[\widehat{\boldsymbol{\theta}}_k^{*,c} \mid \mathcal{F}_0] = \mathbb{E}[\widehat{\boldsymbol{\theta}}_k \mid \mathcal{G}_k] = \boldsymbol{\theta}_k,$$
$$\mathrm{Cov}(\widehat{\boldsymbol{\theta}}_j^{*,c}, \widehat{\boldsymbol{\theta}}_k^{*,c} \mid \mathcal{F}_0) = \mathrm{Cov}(\widehat{\boldsymbol{\theta}}_j, \widehat{\boldsymbol{\theta}}_k \mid \mathcal{G}_j, \mathcal{G}_k) < \infty,$$

holds for all $j, k = 1, \ldots, p$. That is,

$$\mathrm{Cov}(\widehat{\boldsymbol{\theta}}^{*,c} \mid \mathcal{F}_0) = \chi(\boldsymbol{\theta}; \mathcal{F}_0).$$

Specification of $\widehat{\boldsymbol{\theta}}^*$: moment assumptions

Note that

- $\widehat{\boldsymbol{\theta}}^{*,c}$ fulfilling Assumption 3 may be constructed using a mixture of S_t and $S_t^\perp$ — an example of such a construction for the distribution-free CL model is given in [3]
- the conditional construction of $\widehat{\boldsymbol{\theta}}^*$ is in general not meaningful without a specific naturally conditional/partitioned model structure or an ancillarity-like argument
- we have the following unbiasedness relations between $\widehat{\boldsymbol{\theta}}^{*,c}$ and $\widehat{\boldsymbol{\theta}}^\perp$:

$$\mathbb{E}[\mathbb{E}[\widehat{\boldsymbol{\theta}}_k^{*,c} \mid \mathcal{F}_0]] = \mathbb{E}[\widehat{\boldsymbol{\theta}}_k^\perp]$$

$$\mathbb{E}[\mathrm{Cov}(\widehat{\boldsymbol{\theta}}_j^{*,c}, \widehat{\boldsymbol{\theta}}_k^{*,c} \mid \mathcal{F}_0)] = \mathrm{Cov}(\widehat{\boldsymbol{\theta}}_j^\perp, \widehat{\boldsymbol{\theta}}_k^\perp)$$

Specification of $\widehat{\boldsymbol{\theta}}^*$: moment assumptions

Further, regardless of the specification of $\widehat{\boldsymbol{\theta}}^*$ it holds that

$$\begin{aligned}\mathrm{MSEP}^{*,\nabla}_{\mathcal{F}_0}(X,\widehat{X}) &:= H^{*,\nabla}(\boldsymbol{\theta};\mathcal{F}_0)\\ &= \mathrm{Var}(X \mid \mathcal{F}_0)\\ &\quad + \nabla h(\boldsymbol{\theta};\mathcal{F}_0)'\Lambda(\boldsymbol{\theta};\mathcal{F}_0)\nabla h(\boldsymbol{\theta};\mathcal{F}_0)\end{aligned}$$

where

$$\Lambda(\boldsymbol{\theta};\mathcal{F}_0) := \mathrm{Cov}(\widehat{\boldsymbol{\theta}}^* \mid \mathcal{F}_0).$$

That is

$$\widehat{\mathrm{MSEP}}_{\mathcal{F}_0}(X,\widehat{X}) := H^{*,\nabla}(\widehat{\boldsymbol{\theta}};\mathcal{F}_0) \qquad (1)$$

is a distribution-free computable MSEP-estimator using either of the two specifications of $\widehat{\boldsymbol{\theta}}^*$

Specification of $\widehat{\boldsymbol{\theta}}^*$: moment assumptions

- Note that the computable MSEP estimator given by (1) does not rely on that $\widehat{\boldsymbol{\theta}}^*$ is constructed explicitly!
- It is enough to have the above theoretical construction together with that the above moments in terms of $\widehat{\boldsymbol{\theta}}$ are computable!
  - This is the approach taken in [4] for the distribution-free CL model — "pseudo-estimators"
- For other more or less explicit constructions of (conditional) $\widehat{\boldsymbol{\theta}}^*$ for the distribution-free CL model, see e.g. [3, 8]
- See also [6] for explicit constructions of $\widehat{\boldsymbol{\theta}}^*$ for more general models

Applications

Applications: ultimo and CDR(k)

The outstanding claims reserve R_i for accident year i that is not yet fully developed when time is indexed in agreement with $\mathcal{F}_0$:

$$R_i := \sum_{j=J-i+2}^{J} I_{i,j} = C_{i,J} - C_{i,J-i+1}, \quad R := \sum_{i=2}^{J} R_i.$$

The ultimate claim amount U_i for accident year i that is not yet fully developed:

$$U_i := \sum_{j=1}^{J} I_{i,j} = C_{i,J}, \quad U := \sum_{i=2}^{J} U_i.$$

The amount of paid claims P_i for accident year i that is not yet fully developed:

$$P_i := \sum_{j=1}^{J-i+1} I_{i,j} = C_{i,J-i+1}, \quad P := \sum_{i=2}^{J} P_i.$$

Obviously, $U_i = P_i + R_i$ and $U = P + R$.

Applications: ultimo and CDR(k)

Warm up – still no models:

Ultimo:

- $h(\boldsymbol{\theta}; \mathcal{F}_0) := \mathbb{E}[U \mid \mathcal{F}_0]$
- $\widehat{U} := h(\widehat{\boldsymbol{\theta}}; \mathcal{F}_0)$
- $\mathrm{MSEP}_{\mathcal{F}_0}(U, \widehat{U})(\boldsymbol{\theta}) = \mathrm{MSEP}_{\mathcal{F}_0}(R, \widehat{R})(\boldsymbol{\theta})$
- $\widehat{MSEP}_{\mathcal{F}_0}(U, \widehat{U}) = \widehat{MSEP}^{*,\nabla}_{\mathcal{F}_0}(U, \widehat{U})(\widehat{\boldsymbol{\theta}})$

Applications: ultimo and CDR(k)

CDR(k):

- $h^{(0)}(\boldsymbol{\theta};\mathcal{F}_0) := \mathbb{E}[U \mid \mathcal{F}_0], \quad h^{(k)}(\boldsymbol{\theta};\mathcal{F}_k) := \mathbb{E}[U \mid \mathcal{F}_k]$
- $\mathrm{CDR}(k) := h^{(0)}(\widehat{\boldsymbol{\theta}}^{(0)};\mathcal{F}_0) - h^{(k)}(\widehat{\boldsymbol{\theta}}^{(k)};\mathcal{F}_k)$

$$\Longrightarrow h(\boldsymbol{\theta};\mathcal{F}_0) := \mathbb{E}[\mathrm{CDR}(k) \mid \mathcal{F}_0](\boldsymbol{\theta})$$
$$= h^{(0)}(\widehat{\boldsymbol{\theta}}^{(0)};\mathcal{F}_0) - \mathbb{E}[h^{(k)}(\widehat{\boldsymbol{\theta}}^{(k)};\mathcal{F}_k) \mid \mathcal{F}_0](\boldsymbol{\theta})$$
$$\Longrightarrow \widehat{\mathrm{CDR}}(k) := h(\widehat{\boldsymbol{\theta}}^{(0)};\mathcal{F}_0)$$

- $\widehat{\mathrm{MSEP}}_{\mathcal{F}_0}\left(\mathrm{CDR}(k), \widehat{\mathrm{CDR}}(k)\right) =$
  $\mathrm{MSEP}^{*,\nabla}_{\mathcal{F}_0}\left(\mathrm{CDR}(k), \widehat{\mathrm{CDR}}(k)\right)(\widehat{\boldsymbol{\theta}}^{(0)})$

Applications: ultimo and CDR(k)

Note that

- $h^{(0)}(\widehat{\boldsymbol{\theta}}^{(0)}; \mathcal{F}_0)$ is $\mathcal{F}_0$-measurable $\Leftrightarrow$ **part of the basis of prediction!!**
- MSEP for the ultimo claim amount is consistently defined with that of CDR(k)
- the definition of MSEP for the CDR(k) differ from those used in e.g. [3, 4] for the distribution-free CL model — we only resample w.r.t. $\boldsymbol{\theta}$ in $h(\boldsymbol{\theta}; \mathcal{F}_0) := \mathbb{E}[\mathrm{CDR}(k) \mid \mathcal{F}_0](\boldsymbol{\theta})$

Applications: the distribution-free CL model

Assume that, for $j = 1, \ldots, J-1$, there exist constants $f_j > 0$ and constants $\sigma_j^2 \geq 0$ such that

$$\mathbb{E}\big[C_{i,j+1} \mid C_{i,j}, \ldots, C_{i,1}\big] = f_j C_{i,j},$$
$$\mathrm{Var}(C_{i,j+1} \mid C_{i,j}, \ldots, C_{i,1}) = \sigma_j^2 C_{i,j},$$

where $i = i_0, \ldots, J$, and assume that,

$$\{C_{i_0,1}, \ldots, C_{i_0,J}\}, \ldots, \{C_{J,1}, \ldots, C_{J,J}\} \quad \text{are independent.}$$

We will focus on MSEP for the ultimo claim amount

Applications: the distribution-free CL model

By using the above defining relations together with the tower property of conditional expectations yields

$$h_i(\boldsymbol{f}; \mathcal{F}_0) := \mathbb{E}[U_i \mid \mathcal{F}_0] = C_{i,J-i+1} \prod_{j=J-i+1}^{J-1} f_j,$$

$$h(\boldsymbol{f}; \mathcal{F}_0) := \mathbb{E}[U \mid \mathcal{F}_0] = \sum_{i=2}^{J} h_i(\boldsymbol{f}; \mathcal{F}_0).$$

Note that the distribution-free CL model is defined conditionally — define $\widehat{\boldsymbol{f}}^*$ conditionally!!

Applications: the distribution-free CL model

That is,

$$\widehat{f}_j = \frac{\sum_{i=i_0}^{J-j} C_{i,j+1}}{\sum_{i=i_0}^{J-j} C_{i,j}},$$

and let B_j denote the j first columns of the claims triangle defining $\mathcal{F}_0$

Let $\widehat{\boldsymbol{f}}^*$ be defined so that

$$\mathbb{E}[\widehat{f}_j^* \mid \mathcal{F}_0] = \mathbb{E}[\widehat{f}_j \mid B_j] = f_j,$$

$$\mathrm{Var}(\widehat{f}_j^* \mid \mathcal{F}_0) = \{\boldsymbol{\Lambda}(\boldsymbol{\sigma};\mathcal{F}_0)\}_{j,j} = \mathrm{Var}(\widehat{f}_j \mid B_j) = \frac{\sigma_j^2}{\sum_{i=i_0}^{J-j} C_{i,j}},$$

where $\{\boldsymbol{\Lambda}(\boldsymbol{\sigma};\mathcal{F}_0)\}_{i,j} = 0$ for all $i \neq j$.

Applications: the distribution-free CL model

Want to calculate $\widehat{\text{MSEP}}^{*,\nabla}_{\mathcal{F}_0}(U, \widehat{U})$:

- Process variance? Given in [7]
- $\nabla h(\boldsymbol{f}; \mathcal{F}_0)$? Simple calculations!

$$\Rightarrow \{\nabla_{\boldsymbol{f}} h_i(\boldsymbol{f}; \mathcal{F}_0)\}_j = 1_{\{J-i+1 \leq j\}} C_{i,J-i+1} \frac{1}{f_j} \prod_{l=J-i+1}^{J-1} f_l,$$

for $i = 2, \ldots, J$ and $j = 1, \ldots, J-1$

Applications: the distribution-free CL model

By combining the above and simplifying gives us the following:

Proposition 1

$\widehat{\text{MSEP}}^{*,\nabla}_{\mathcal{F}_0}(U, \widehat{U})$ *from* (1) *is equal to the MSEP given in [7] for the distribution-free CL model.*

How about if we would use an unconditional specification of $\widehat{\boldsymbol{f}}^{*}$?

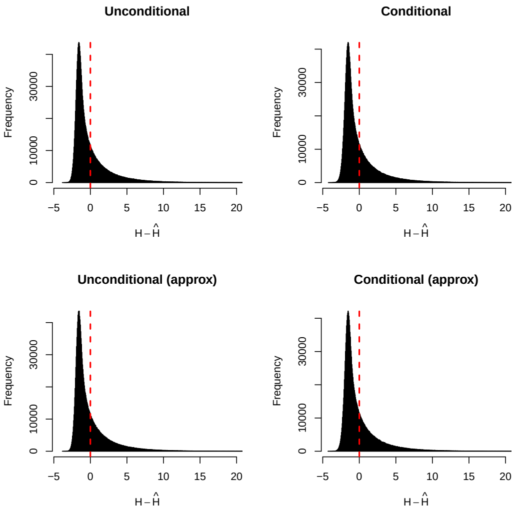

Figure: Illustration of conditional and unconditional specification of $\widehat{\theta}^*$ for Mack's C-L based on Taylor-Ashe '83 data.

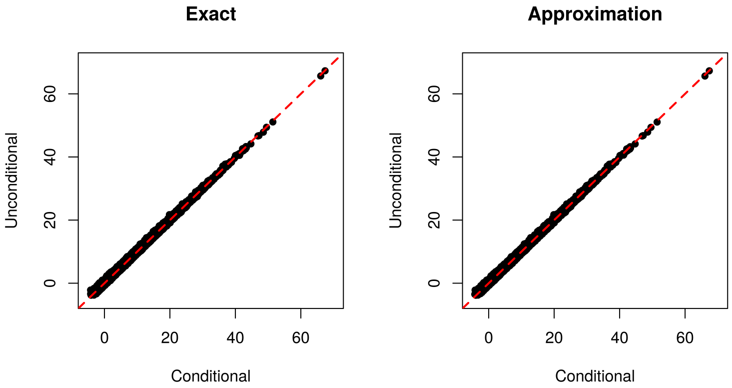

Figure: Illustration of conditional and unconditional specification of $\widehat{\theta}^*$ for Mack's C-L based on Taylor-Ashe '83 data.

Applications: the distribution-free CL model

- It is possible to show that the conditional specification of $\widehat{\boldsymbol{\theta}}^*$ in line with Assumption 3 is an **unbiased estimator of the unconditional dito** (in fact holds for a wider class of models)
- It is possible to show that both specifications will be close given a sufficient amount of data (asym. consistent)
- CDR(k) can be calculated in an analogous manner, but more involved calculations — skip!

Applications: over-dispersed CL model

Assume that

$$\mathbb{E}[I_{i,j}] = \mu_{i,j},$$
$$\mathrm{Var}(I_{i,j}) = \phi\mu_{i,j},$$

where

$$\log(\mu_{i,j}) = \eta + \alpha_i + \beta_j, \quad \alpha_1 = \beta_1 = 0.$$

Note that the above is a model defined **unconditionally**!

MSEP for the ultimo claim amount?

Applications: over-dispersed CL model

$$h_i(\boldsymbol{\theta}; \mathcal{F}_0) = \mathbb{E}[U_i \mid \mathcal{F}_0] = C_{i,I-i} + \sum_{j=I-i+1}^{J} \mu_{i,j} = C_{i,I-i} + g_i(\boldsymbol{\theta}),$$

where $\boldsymbol{\theta} = (\eta, \{\alpha_i\}, \{\beta_k\})$, from which it follows that

$$\nabla h(\boldsymbol{\theta}; \mathcal{F}_0) = \nabla g(\boldsymbol{\theta})$$

and similarly it follows that $\mathrm{Var}(U \mid \mathcal{F}_0) = \mathrm{Var}(R)$, where R **is indexed in time in agreement with** $\mathcal{F}_0$

Applications: over-dispersed CL model

In a bit more detail, we have that:

$$\frac{\partial}{\partial \eta} h_i(\boldsymbol{\theta}; \mathcal{F}_0) = \frac{\partial}{\partial \eta} g_i(\boldsymbol{\theta}) = \sum_{j=I-i+1}^{J} \mu_{i,j},$$

$$\frac{\partial}{\partial \alpha_k} h_i(\boldsymbol{\theta}; \mathcal{F}_0) = \frac{\partial}{\partial \alpha_k} g_i(\boldsymbol{\theta}) = \sum_{j=I-i+1}^{J} 1_{\{i=k\}} \mu_{i,j},$$

$$\frac{\partial}{\partial \beta_k} h_i(\boldsymbol{\theta}; \mathcal{F}_0) = \frac{\partial}{\partial \beta_k} g_i(\boldsymbol{\theta}) = \sum_{j=I-i+1}^{J} 1_{\{j=k\}} \mu_{i,j},$$

$$\mathrm{Var}(U_i \mid \mathcal{F}_0) = \mathrm{Var}(R_i) = \phi \sum_{j=I-i+1}^{J} \mu_{i,j}.$$

Applications: over-dispersed CL model

Thus, by simplifying the MSEP-estimator from (1) we get

$$\widehat{\text{MSEP}}_{\mathcal{F}_0}(U, \widehat{U}) = \widehat{\text{MSEP}}(R, \widehat{R}) \\ = \sum_{i=2}^{J} \text{Var}(R_i)(\widehat{\boldsymbol{\theta}}) + \sum_{i=2}^{J} \nabla g_i(\widehat{\boldsymbol{\theta}})' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}) \nabla g_i(\widehat{\boldsymbol{\theta}}) \\ + 2 \sum_{2 \leq i < k \leq j} \nabla g_i(\widehat{\boldsymbol{\theta}})' \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}) \nabla g_k(\widehat{\boldsymbol{\theta}}). \quad (2)$$

Applications: over-dispersed CL model

Note that

- $\mathbf{\Lambda}(\widehat{\theta})$ is not available analytically, but is approximated when using standard GLM-fitting procedures $\Rightarrow$ (2) is semi-analytical!
- the MSEP-estimator from (2) coincides with that from [5]
- the notion of MSEP reduces to the unconditional FPE — as wanted

Summary

- We have discussed how to calculate the conditional MSEP for (claims triangle based) reserving models
- We introduced a certain randomisation procedure in order to approach the problems with the estimation error part of these calculations
- The introduced randomisation procedure was related to Akaike's FPE and we showed that FPE $=$ MSEP when using an unconditional, complete resampling, procedure
- Based on the introduced randomisation procedure we also suggested a simple distribution-free formula for calculation of MSEP based on a single first order Taylor expansion
- The applicability of the approach was illustrated in a number of examples

References I

Hirotugu Akaike.
Fitting autoregressive models for prediction.
Ann. Inst. Statist. Math., 21:243–247, 1969.

Hirotugu Akaike.
Statistical predictor identification.
Ann. Inst. Statist. Math., 22:203–217, 1970.

Markus Buchwalder, Hans Bühlmann, Michael Merz, and Mario V Wüthrich.
The mean square error of prediction in the chain ladder reserving method (Mack and Murphy revisited).
ASTIN Bulletin, 36(2):521–542, 2006.

References II

Dorothea Diers, Marc Linde, and Lukas Hahn.
Addendum to the multi-year non-life insurance risk in the additive reserving model[insurance math. econom. 52 (3)(2013) 590–598]: Quantification of multi-year non-life insurance risk in chain ladder reserving models.
Insurance: Mathematics and Economics, 67:187–199, 2016.

Renshaw Arthur E.
On the second moment properties and the implementation of certain glim based stochastic claims reserving models.
City University Actuarial Research Paper No. 65, 1994.

Mathias Lindholm, Filip Lindskog, and Felix Wahl.
Estimation of conditional mean squared error of prediction for claims triangle reserving.
Preprint, 2018.

References III

Thomas Mack.
Distribution-free calculation of the standard error of chain ladder reserve estimates.
ASTIN Bulletin, 23(2):213–225, 1993.

Ancus Röhr.
Chain ladder and error propagation.
ASTIN Bulletin, 46(2):293–330, 2016.

Bayesian parameter uncertainty vs frequentist estimation error

- Let $\boldsymbol{\theta}^*$ denote a random parameter vector and let X be defined as above
- A Bayesian analog of the conditional MSEP is given by the predictive posterior variance of X:

$$\begin{aligned}
\mathrm{Var}(X \mid \mathcal{F}_0) &:= \mathbb{E}[(X - \mathbb{E}[X \mid \mathcal{F}_0])^2 \mid \mathcal{F}_0] \\
&= \mathbb{E}\left[\mathrm{Var}\left(X \mid \boldsymbol{\theta}^*, \mathcal{F}_0\right) \mid \mathcal{F}_0\right] \\
&\quad + \mathbb{E}\left[(h(\boldsymbol{\theta}^*; \mathcal{F}_0) - \mathbb{E}[h(\boldsymbol{\theta}; \mathcal{F}_0) \mid \mathcal{F}_0])^2 \mid \mathcal{F}_0\right] \\
&= \mathbb{E}_{\boldsymbol{\theta}^* \mid \mathcal{F}_0}\left[\mathrm{Var}\left(X \mid \boldsymbol{\theta}^*, \mathcal{F}_0\right)\right] \\
&\quad + \mathbb{E}_{\boldsymbol{\theta}^* \mid \mathcal{F}_0}\left[(h(\boldsymbol{\theta}^*; \mathcal{F}_0) - \mathbb{E}[h(\boldsymbol{\theta}; \mathcal{F}_0) \mid \mathcal{F}_0])^2\right]
\end{aligned}$$

Bayesian parameter uncertainty vs frequentist estimation error

- The estimation/uncertainty part in the two MSEP definitions are closely related
- Depending on assumptions concerning the a priori distribution of $\boldsymbol{\theta}^*$ you can, depending on model structure, affect the a posteriori distribution providing heuristic guidance on the specification of $\widehat{\boldsymbol{\theta}}^*$
- Still, $\boldsymbol{\theta}^*$ is **truly random** whereas $\widehat{\boldsymbol{\theta}}^*$ is a device to capture **uncertainty due to estimation of an unknown parameter vector**

Distribution-free model selection

- The concept of FPE was originally introduced as a device to be used for model selection for auto-regressive linear models
- These models are closely related to auto-regressive linear reserving models of the type introduced in e.g. [6]
- The above discussed techniques could hence be used for distribution-free model selection
- Other approaches to model selection that are closely connected to the above notion of MSEP is e.g. Mallows' C_p, which in our notation, is based on calculating the following quantity:

$$\mathbb{E}^t[(h^p(\widehat{\boldsymbol{\theta}}^p; \mathcal{F}_0) - h^t(\boldsymbol{\theta}^t; \mathcal{F}_0))^2 \mid \mathcal{F}_0]$$

again a situation which can be approached using the introduced techniques