

## Article

# Model Error (or Ambiguity) and Its Estimation, with Particular Application to Loss Reserving

Greg Taylor <sup>1,\*</sup>  and Gráinne McGuire <sup>2</sup><sup>1</sup> School of Risk and Actuarial Studies, University of New South Wales, Randwick, NSW 2052, Australia<sup>2</sup> Taylor Fry, 45 Clarence Street, Sydney, NSW 2000, Australia; grainnemcguire@gmail.com

\* Correspondence: gregory.taylor@unsw.edu.au

**Abstract:** This paper is concerned with the estimation of forecast error, particularly in relation to insurance loss reserving. Forecast error is generally regarded as consisting of three components, namely parameter, process and model errors. The first two of these components, and their estimation, are well understood, but less so model error. Model error itself is considered in two parts: one part that is capable of estimation from past data (internal model error), and another part that is not (external model error). Attention is focused here on internal model error. Estimation of this error component is approached by means of Bayesian model averaging, using the Bayesian interpretation of the LASSO. This is used to generate a set of admissible models, each with its prior probability and likelihood of observed data. A posterior on the model set, conditional on the data, may then be calculated. An estimate of model error (for a loss reserve estimate) is obtained as the variance of the loss reserve according to this posterior. The population of models entering materially into the support of the posterior may turn out to be “thinner” than desired, and bootstrapping of the LASSO is used to increase this population. This also provides the bonus of an estimate of parameter error. It turns out that the estimates of parameter and model errors are entangled, and dissociation of them is at least difficult, and possibly not even meaningful. These matters are discussed. The majority of the discussion applies to forecasting generally, but numerical illustration of the concepts is given in relation to insurance data and the problem of insurance loss reserving.

**Keywords:** Bayesian model averaging; bootstrap; bootstrap matrix; forecast error; GLM; internal model structure error; LASSO; loss reserving; model ambiguity; model error; model uncertainty



**Citation:** Taylor, Greg, and Gráinne McGuire. 2023. Model Error (or Ambiguity) and Its Estimation, with Particular Application to Loss Reserving. *Risks* 11: 185. <https://doi.org/10.3390/risks11110185>

Academic Editor: Mogens Steffensen

Received: 16 August 2023

Revised: 3 October 2023

Accepted: 17 October 2023

Published: 25 October 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. Background and Literature Review

This paper is concerned with **forecast error**, particularly in relation to **actuarial or insurance loss reserving**. The majority of the concepts are fairly general, and might be applied in many forecasting environments, but the illustrations given here relate to loss reserving. To an extent, the paper is a sequel to [McGuire et al. \(2021\)](#), which dealt with estimation of a loss reserve by means of the LASSO. Both papers make applications of the LASSO, the first paper to point estimation, and this paper to estimation of forecast error, with a particular focus on model error.

When a forecast is made on the basis of a model of observations, it will inevitably contain an error relative to the true value yet to be observed. Estimation of the properties of this error will provide some understanding of the reliability of the forecast. This has been an issue in the loss reserving literature since it was raised by [Reid \(1978\)](#), [De Jong and Zehnwirth \(1983\)](#) and [Taylor and Ashe \(1983\)](#).

For example, the sensitivity of forecasts to individual observations within a claim array has been investigated by [Avanzi et al. \(2023\)](#). [Chukhrova and Johannssen \(2021\)](#) follow up [De Jong and Zehnwirth \(1983\)](#), mentioned above, in examining state space models for

loss reserving. Dahms (2018) makes comment on the estimation of forecast error when reserving at mid-year on the basis of models of the chain ladder type.

In recent years, forecast error has been decomposed into a number of components, most notably (in the terminology commonly used in loss reserving) parameter, process and model errors (see, e.g., Taylor (1988, 2021), O'Dowd et al. (2005), Taylor and McGuire (2016)). Estimation of the first two of these three has become well understood, but there has been limited development of the estimation of model error.

Notable exceptions are Shi (2013) and Zhang and Dukic (2013). Both of these contributions were concerned with the modelling of multivariate claim triangles, and used copulas for this purpose. In each case, allowance was made for uncertainty in the parameters describing both cell means and the copulas, and averaging estimated loss reserves over these dimensions of uncertainty gave estimates of model error. Comparative comment on these papers and the present one will be given in Section 7.3.

A further exception, though of a different type, is O'Dowd et al. (2005) and Risk Margins Task Force (2008), who provided a framework for the estimation of each component of forecast error using scorecards to score subjectively a range of factors identified as likely to influence the quantum of forecast error.

Where objective estimates are concerned, Taylor (2021) investigated model distribution error, a component of model error, and Bignozzi and Tsanakas (2016) consider model risk in the context of VaR estimation, which is of course relevant to loss reserving, but loss reserving models as such are beyond their scope. Blanchet et al. (2019) estimate the effect of model error on a performance statistic in terms of the extrema of that measure over the set of admissible models.

However, to the authors' knowledge, there has been no other progress in the actuarial loss reserving literature.

In the meantime, the subject has been addressed elsewhere in the economics and finance literature. Useful general overviews are given by Glasserman and Xu (2014) and Schneider and Schweizer (2015), in the context of financial risk management. The approach of Blanchet et al. (2019) is similar to the latter.

Huang et al. (2021) estimate the total of parameter error and model error (they call these data variability and procedural variability, respectively) in relation to deep neural networks. Their approach equates more or less to bootstrapping, but where the replications are obtained by random variation of the network initialization rather than data re-sampling.

The literature gives certain prominence to the **Bayesian model averaging ("BMA")** (Raftery 1995, 1996; Raftery et al. 1997; Hoeting et al. 1999; Clyde and George 2004; Clyde and Iversen 2013), and model confidence sets have also been considered (Hansen et al. 2011).

An example of this approach appears in the econometric literature in Loaiza-Maya et al. (2021), whose focus on Bayesian prediction is similar to the Bayesian approach followed in the present paper, but with conditional likelihood replaced by a scoring rule. Martin et al. (2022) give a non-Bayesian presentation of the same ideas.

This literature has been highly valuable at the fundamental conceptual level. However, the concepts are not easy to operationalize, and much of the literature does not address loss reserving specifically.

The present paper endeavours to fill some of the literature gaps identified above. It uses the LASSO (Hastie et al. 2009) to populate abstract concepts, such as the model set and its prior distribution, which occur in the more theoretical literature. Since the LASSO may also be used as the source of a loss reserving model, this creates a direct nexus between that model and the estimation of its model error.

## 1.2. Contribution

The paper establishes a rigorous approach to the estimation of model error, which has been largely absent from prior actuarial loss reserving literature. A typical difficulty in the application of BMA is the selection of a suitable prior distribution across models, which

is often approached subjectively. Here, the prior emerges from application of the LASSO, which adds rigour.

The paper uses the well-established techniques of LASSO and BMA, but confronts a number of practical hurdles in doing so. First, as the models of interest are predictive, extrapolating outside the range of the input data, some are found to fit past data extremely well yet produce highly dubious forecasts. A means of censoring these out of the BMA is required (Section 6.2.5).

Second, as in most Bayesian models, it is necessary to insert one or more parameters in the prior distributions “by hand”. This is the case for the LASSO’s Laplace dispersion parameter and the likelihood dispersion parameter. For the sake of rigour, evidence-based estimates of these quantities are formulated (Sections 6.1.2 and 6.2.3).

Although the LASSO produces a model set in principle for the BMA, it is bedevilled by its paucity, and the bootstrap is used to produce replicates of that set (Section 7). The use of the bootstrap comes with its own difficulties. One of these is that it creates some ambiguity between parameter error and model error, which requires consideration (Section 7.3).

The result is that, using the procedures described here, one may perform the following entire sequence of operations:

- Model a data set of claim observations and extract a point estimate of the loss reserve;
- Move on to estimating the distributions of several components of forecast error that constitute a major part of the total;
- Supplement these with the distributions of the missing components, derived from other sources;
- Apply these distributions to the calculation of loss reserve risk margins, or any other quantities of interest that depend on the distribution of forecast error.

### 1.3. Outline of Paper

The paper is arranged as follows. A high-level overview of the results is presented in Section 2. After the establishment of the necessary mathematical framework and notation in Section 3, the structure of the problem to be considered is established with a review of the components of forecast error, and thereafter the paper focuses on one particular component, **internal model structure error (“IMSE”)** (Section 4). Section 5 discusses the estimation of IMSE in the abstract, and the ingredients required for it. Then Section 6 puts these concepts to work in the specific context of the LASSO. This derives an estimate of the distribution of IMSE, but this estimate is then strengthened by bootstrapping the LASSO in Section 7. The whole procedure is then applied to several synthetic data sets, of varying complexity, with numerical results presented in Section 8. Finally, Section 9 summarizes and considers the successes and limitations of the paper in the attainment of its objectives and outlines areas for future work.

We have prepared a worked example to accompany this paper. The example and the accompanying code are available at <https://grainnemcguire.github.io/post/2023-05-04-model-error-example/> (accessed on 23 October 2023).

## 2. Overview of Results and Discussion

The paper formulates an estimate of model error in the context of insurance loss reserving. The relevance of this and the motivation for it are explained in Section 1.1. BMA forms the basis of the estimate (Section 5). Mathematically, this is a simple concept, but there are many practical problems to be addressed, most or all of which are context-dependent.

This paper aims to address them to the point where the estimation protocol becomes useable in a real-life setting, including that of meeting the statutory obligations of an insurance company. The current literature contains no objective version of this.

At the heart of BMA, as applied to model error estimation, is the identification of a model set, and the association of a prior probability distribution across this set. There may be various ways of achieving this, but this paper derives the prior by application of the LASSO to the data (Sections 5 and 6). The LASSO regularization parameter determines the

degree of model complexity that one is willing to tolerate, and hence the complexity of models in the model set. This requires a decision on the selection of a prior (Section 6.2.3).

This approach succeeds in generating the required model set with a prior and likelihood, and BMA is then applied to yield a posterior across the model set. Notwithstanding this success, the model set remains relatively sparse, and a bootstrap is used to add to its density.

Again, the concept of bootstrapping is simple, but leads to a number of practical difficulties. To begin with, some bootstrapped models fit poorly to the original data. While they are assigned low posterior probabilities, their forecasts may be sufficiently eccentric that they would contribute materially to the Bayesian average. These models require exclusion (Section 7.2).

A particular feature of loss reserving, especially in relation to a long-tail line of insurance business, is the need to forecast many periods into the future. This creates the possibility for seemingly reasonable models to go seriously awry in their longer-term forecasts. This risk is likely to be exacerbated by any machine learning procedure for generation of the model set, since the more black-box nature of ML models makes it difficult to impose guardrails on extrapolation such as those an actuary would normally include to ensure reasonable results. This is essentially unavoidable, as one cannot maintain full supervision over the contents of a sufficiently large model set.

As a machine learning procedure, the LASSO is subject to this risk and, indeed, models are observed that fit extremely well to past data yet produce outrageous forecasts. These seriously corrupt the BMA results, and it is necessary to prune the model set, both original and bootstrapped, in order to censor them (Section 6.2.5).

When BMA is applied with bootstrapping to realistic data, with attention paid to all these practicalities, reasonable estimates of model error are obtained (Section 8).

### 3. Reserving Framework and Notation

As far as possible, the notation here will follow that of McGuire et al. (2021). Accordingly, the analysis below will be concerned with the conventional **claims triangle**. Some random variable of interest  $Y$  is labelled by **accident period**  $i = 1, 2, \dots, I$  and **development period**  $j = 1, 2, \dots, I - i + 1$ . In this setup, a **cell** of the triangle refers to the combination  $(i, j)$ , and the observation  $Y$  in this cell denoted  $Y_{ij}$ . The **payment period** to which cell  $(i, j)$  relates will be denoted by  $t = i + j - 1$ .

Let  $\Delta$  denote the collection of cells (i.e., ordered pairs  $(i, j)$ ) of which the triangle consists, and let  $\mathcal{Y}$  denote the observations on these cells, i.e.,  $\mathcal{Y} = \{Y_{ij} : (i, j) \in \Delta\}$ . Accident and development periods will be assumed to be of equal duration, but not necessarily years. As further notation,  $E[Y_{ij}] = \mu_{ij}$ ,  $Var[Y_{ij}] = \sigma_{ij}^2$ . A realization of  $Y_{ij}$  will be denoted  $y_{ij}$ . For computational purposes, the data  $\mathcal{Y}$  will usually be represented by in vector form, denoted  $Y$ .

Let  $Z_{ij}$  be any random vector defined on  $\Delta$  and  $z_{ij}$  its realization. It will sometimes be useful to vectorize these quantities, and so  $Z$  will denote the column vector of all  $Z_{ij}$  listed in some defined order, and  $z$  the corresponding vector of all  $z_{ij}$ .

This paper will be concerned with **forecasts** produced by models fitted to the data  $Y$ . Forecasts are made in respect of the  $Y_{ij}$  for  $(i, j) \in \Delta^*$ , some set of cells, disjoint from  $\Delta$ , and relating to the future, i.e.,  $i + j > I + 1$ . These  $Y_{ij}$  will now be denoted  $Y_{ij}^*$ , and the vector of these will be  $Y^*$ .

The forecast of  $Y_{ij}^*$  by model  $M$  will be denoted  $\hat{Y}_{ij}^*(M)$  and is equal to  $f_M(Y; i, j)$  for some real-valued function  $f_M$  and future cell  $(i, j)$ . As the notation indicates, the forecast is  $M$ -dependent but, as this notation is cumbersome, the  $M$  will be suppressed and the forecast written as simply  $\hat{Y}_{ij}^*$  when this does not create any ambiguity. Other quantities derived from  $\hat{Y}_{ij}^*$  (e.g., those immediately below) will also be notated without explicit mention of  $M$ .

A model  $M$  will include a likelihood function  $L(\cdot|M)$  for the observations so that the likelihood of the data vector  $Y$  is  $L(Y|M)$ . Let  $\ell(Y|M) = \ln L(Y|M)$ , the log-likelihood function.

It will be convenient to denote a value fitted by the model to a past observation  $y_{ij}$  by  $\hat{\mu}_{ij} = f_M(Y; i, j)$  for past cell  $(i, j)$ . According to the vector notation given above,  $\hat{\mu}$  denotes the vector of  $\hat{\mu}_{ij}$  for  $(i, j) \in \Delta$ . Similarly, let  $\hat{Y}^*$  denote the vector of  $\hat{Y}_{ij}$  for  $(i, j) \in \Delta^*$ .

The modeller may select the model  $M$  but, in practice, will rarely know whether it is a correct representation of the data. Therefore, assume that  $M$  is selected from some collection  $\mathcal{M}$  of candidate models, hereafter called the **model set**. Assume further that the model set is equipped with a measure  $\pi$ .

Suppose that  $R = g(Y^*)$ , for some real-valued function  $g$ , and define its forecast as  $\hat{R} = g(\hat{Y}^*)$ . An example is  $R = 1^T Y^*$ , where  $1$  is a vector of the same dimension as  $Y^*$  and with all components equal to unity. If the  $Y_{ij}^*$  denote claim payments, then  $R$  is the amount of outstanding claim liability.

The **forecast error** associated with forecast  $\hat{R}$  will be defined as

$$e = R - \hat{R}. \quad (1)$$

Extensive use will be made of conditional expectations in this paper. For example, a typical claims forecast is conditional, usually tacitly, on a selected model. If the quantity under forecast is  $X$ , then typically the forecast is  $E[X|M, F]$ , where  $M$  denotes the algebraic structure of observation means assumed by the model, and  $F$  denotes the distribution of the random disturbance of observations. For example, if the model were a GLM, then  $M$  would denote the linear response transformed by the inverted link function, and  $F$  would be the GLM's selected error distribution.

Since a forecast  $X$  will be a function of data, a conditional expectation of the form  $E[X|M, F]$  is, in effect, a probability-weighted average of the forecast over all possible data sets.

If one averages over all models in an admissible set, the result is  $E_{M,F}E[X|M, F]$  where, in general, a subscript on the expectation operator indicates that expectation is taken with respect to the entity represented by that subscript. For brevity of notation, the convention is sometimes adopted whereby the absence of a conditioning entity from an expression implies that the entity has been integrated out. Thus, the above quantity  $E_{M,F}E[X|M, F]$  can be written equivalently as just  $E[X]$ .

In similar fashion, one might integrate out just one of  $M, F$ , e.g.,  $E_M E[X|M, F]$ , which would then be abbreviated to  $E[X|F]$ .

Conditional variances also appear below. Since a variance is just an expectation (of a squared residual), they are subject to the same notational conventions as above.

Later sections will make use of **open-ended ramp functions**. These are single-knot linear splines with zero gradient in the left-hand segment and unit gradient in the right-hand segment. In a machine learning context, one of these would be referred to as a **rectified linear unit ("ReLU")**. Let  $R_K(x)$  denote the open-ended ramp function with knot at  $K$ . Then

$$R_K(x) = \max(0, x - K) \quad (2)$$

For a given condition  $c$ , define the **indicator function**  $I(c) = 1$  when  $c$  is true, and  $I(c) = 0$  when  $c$  is false. Further, define the **Heaviside function**  $H_k(x) = I(x \geq k)$ .

#### 4. Components of Forecast Error

Some regulatory regimes require that a **capital margin** be associated with a technical reserve such that the total of the reserve plus a margin equals at least the **Value at Risk ("VaR")** of the liability at some high percentile, such as 99.5%. These regimes may also require the evaluation of a reserve **risk margin** within the capital margin. For example, the Australian prudential standards require a loss reserve (including the margin) to be at

least equal to the 75% VaR of the associated liability (Australian Prudential Regulatory Authority 2018).

IFRS17 requires a risk adjustment for non-financial risk. While there is no prescribed method for calculation, margins based on the VaR are likely to be widely used.

For the sake of definiteness, this paper proceeds on the basis that the specific liability forecast under consideration is the loss reserve. For some regulatory regimes (e.g., IFRS17) the quantity of interest might be the one-year-ahead forecast of claim payments, with an associated VaR. In such cases, the model error methodology set out in subsequent sections translates readily to this alternative situation.

The requirement of a VaR necessitates the estimation of the **distribution** of a forecast liability, as opposed to a simple point estimate. Often, especially in the case of low- or medium-percentile VaRs (some jurisdictions require reserving risk margins (as opposed to capital risk margins) at percentiles in the order of 75%), it is reasonable to assume that the required distribution is characterized by its mean (the point estimate) and variance. For estimation of higher VaRs, while the first and second moments are usually not sufficient, they do provide a starting point. Hence the need for examination of the variance of forecast error.

Forecast error consists of a number of identifiable components. Decomposition of forecast error is discussed by Taylor (2000, 2021), Taylor and McGuire (2016), McGuire et al. (2021), O’Dowd et al. (2005), Risk Margins Task Force (2008) and Hastie et al. (2009), and Huang et al. (2021), among others.

The different authors use slightly different terminology, and so Table 1 displays the correspondences between the different terminologies. A blank entry indicates that the component concerned is not explicitly considered by the authors in question. Correspondences are sometimes exact, but at other times are more approximate because of different approaches taken by different authors.

**Table 1.** Terminologies for components of forecast error.

Taylor and Co-Authors	O’Dowd et al. (2005)	Risk Margins Task Force (2008)	Hastie et al. (2009)	Machine Learning <sup>(a)</sup>
Model bias			Model bias (partial)	
Parameter error	Independent parameter risk	Independent parameter risk	Variance	Data variability
Process error	Independent process risk	Independent process risk	Irreducible error	Aleatoric uncertainty
Internal model structure error	Model specification risk; systemic parameter risk	Internal systemic risk	Model bias (partial)	Procedural variability
External model structure error	Future systemic risk	External systemic risk	Model bias (partial)	
Model distribution error				

Note: <sup>(a)</sup> Huang et al. (2021) bundle all model errors other than aleatoric under the term “epistemic uncertainty”.

The most comprehensive decomposition is given by Taylor (2021), on which the following is largely based. Henceforth, let the model  $M$  introduced in Section 3 consist of just its distribution-free part, i.e., the algebraic form of its non-stochastic part (e.g., the linear predictor in the case of a GLM), and let  $F$  denote the model distribution of the stochastic error.

Let  $\mu, \hat{\mu}$  denote  $E[R], E_{M,F}[\hat{R}|M, F]$ , respectively. Then the forecast error (1) may be expressed as

$$e = (\mu - \hat{\mu}) + (R - \mu) - (\hat{R} - E[\hat{R}|M, F]) - (E[\hat{R}|M, F] - \hat{\mu}). \tag{3}$$

The right side of the equation involves four members. They are, from left to right:

- The sampling error in the model forecast, averaged over all admissible models;
- The error in the model forecast, free of sampling error and averaged over all models, relative to the true (but unknown) reserve;
- The error in the reserve forecast by the selected model relative to its mean value;

- The error in the forecast reserve, free of sampling error, relative to its average over all admissible models.

It follows (Taylor 2021) that the **mean square error of prediction (“MSEP”)** of  $\hat{R}$ , allowing for variation over models, is

$$MSEP[e] = E_{M,F}E[e^2|M,F] = (\mu - \hat{\mu})^2 + E_F Var[R|F] + E_{M,F} Var[\hat{R}|M,F] + Var_{M,F}E[\hat{R}|M,F] \tag{4}$$

With a slight abuse of notation, the quantity  $MSEP[e]$  will also be referred to as forecast error when this usage is unambiguous. The four terms on the right may be recognised as, respectively, model bias, process error, parameter error (“PaE”) and model error.

Model error can be seen here to relate to possible misspecification of the model. This is referred to elsewhere in the literature as “model ambiguity” or “model uncertainty”. It may be further decomposed with a small amount of manipulation:

$$Var_{M,F}E[\hat{R}|M,F] = E_F Var_M E[\hat{R}|M,F] + Var_F E_M E[\hat{R}|M,F] \tag{5}$$

The two members on the right are labelled **model structure error** and **model distribution error** in Taylor (2021). If the latter is assumed away by assuming the distribution of  $F$  to be concentrated in a single point, then (5) collapses to just

$$Var_{M,F}E[\hat{R}|M,F] = Var_M E[\hat{R}|M] \tag{6}$$

O’Dowd et al. (2005) recognize that some of the variation over models  $M \in \mathcal{M}$  relates to misspecification of the model fitted to past data, and the rest to the extrapolation of that model in its forecasts.

In the case of a GLM, for example, the data will have been modelled using some linear response  $X\beta$ , and the data set augmented by the forecast experience will use an augmented linear response  $X^+\beta^+$ , where

$$X^+ = \begin{bmatrix} X & 0 \\ X^* & X^{**} \end{bmatrix}, \beta^+ = \begin{bmatrix} \beta \\ \beta^* \end{bmatrix}, \tag{7}$$

and the parameter sub-vector  $\beta^*$  is not estimated from data, but consists of assumptions about the future, which are combined with the design matrix subcomponent  $X^{**}$  to form the augmented part of the linear predictor. For example, **superimposed inflation (“SI”)**, a diagonal effect, might be estimated at one rate in past data, but a different rate assumed for the future. The past rate would be included in  $\beta$ , and the future in  $\beta^*$ .  $X^{**}$  will be 0 if there are no adjustments included in the model.

The forecast of future claims experience depends on the linear response  $X^*\beta + X^{**}\beta^*$ , in which the first of the two members relates to past observations, and the second to assumptions about the future. Forecast errors in relation to these two are referred to as **internal** and **external model errors**, respectively.

Hence, the first of the two members on the right side of (5) can be decomposed further:

$$E_F Var_M E[\hat{R}|M,F] = E_F Var_{M^{int}} E[\hat{R}|M,F] + E_F Var_{M^{ext}} E[\hat{R}|M,F] \tag{8}$$

where  $M^{int}, M^{ext}$  are those parts of model  $M$  generating internal and external model errors.

Substitution of (5) and (8) into (4) yields the full decomposition depicted in Figure 1.

The estimation of parameter and process error is covered by England and Verrall (1999, 2001, 2002), Taylor (2000), and Taylor and McGuire (2016). Model distribution error is discussed by Taylor (2021).

The present paper will be concerned with model structure error, and specifically internal model structure error (IMSE). As explained earlier in this section, this relates to any error in the selection of the algebraic form fitted to the data, and the objective here is to estimate this error on the basis of the data.

External model error, on the other hand, relates to future events and influences on future claim data and, by definition, there are no available data relevant to these. The estimation of this component of error must be performed by some means other than reference to past claim data (e.g., O’Dowd et al. 2005), and is not addressed here.

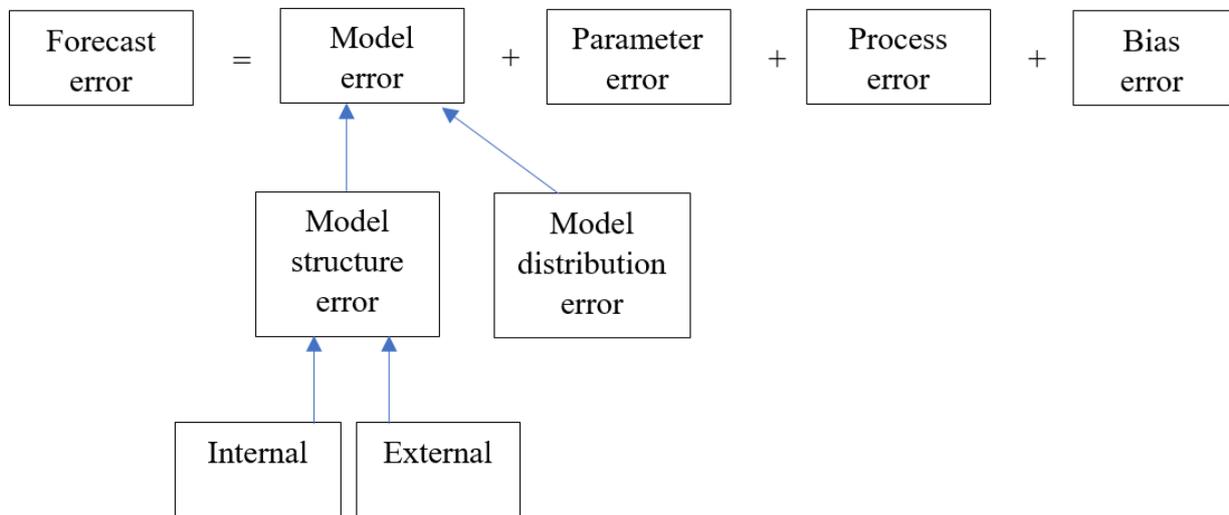


Figure 1. Full decomposition of model error.

### 5. Estimation of Internal Model Structure Error

#### 5.1. Fundamental Ingredients

The fundamentals of IMSE are discussed in Clyde and George (2004), Clyde and Iversen (2013), Hansen et al. (2011), and Schneider and Schweizer (2015).

As mentioned in Section 3, one contemplates a model set, models  $M \in \mathcal{M}$ , with probability measure  $\pi$  defined on  $\mathcal{M}$ . Each model  $M$  defines a loss reserve estimate  $\hat{R}(M) = 1^T \hat{Y}^*(M)$ . With this machinery, one is in a position to calculate summary statistics of the estimated loss reserve over candidate models. Of particular interest are

$$E_{\mathcal{M}}[\hat{R}] = \int_{\mathcal{M}} \hat{R}(M) d\pi(M), \tag{9}$$

$$Var_{\mathcal{M}}[\hat{R}] = \int_{\mathcal{M}} [\hat{R}(M) - E_{\mathcal{M}}[\hat{R}]]^2 d\pi(M). \tag{10}$$

Quantity (10) measures the dispersion of estimated loss reserve over the entire model set, and will be of interest in the formulation of model error in Section 6.2.

#### 5.2. Model Confidence Set

This is discussed by Hansen et al. (2011), Glasserman and Xu (2014) and Schneider and Schweizer (2015). In its most basic definition, a subset  $\mathcal{S} \subset \mathcal{M}$  is a  $100\alpha\%$  **model confidence set** relative to some reference model  $M_0$  if  $M_0 \in \mathcal{S}$  and  $\pi(\mathcal{S}) = \alpha$ . This is the approach of Hansen et al. (2011), but there are other possibilities. For example, Schneider and Schweizer (2015) are concerned with identification of the worst-case risk within a prescribed divergence radius of the reference model.

#### 5.3. Bayesian Model Averaging

The concepts of BMA are discussed in Raftery (1995, 1996), Raftery et al. (1997), Hoeting et al. (1999), Clyde and George (2004), and Clyde and Iversen (2013).

The BMA approach uses the structure of Section 5.1 but with  $\pi$  representing a Bayesian posterior distribution on  $\mathcal{M}$ . It is supposed that a prior distribution function  $\pi_{pr}$  is specified on  $\mathcal{M}$ . Then application of Bayes' theorem yields the posterior distribution

$$d\pi_{po}(M) = d\pi_{po}(M|Y = y) = \frac{L(y|M)d\pi_{pr}(M)}{\int_{M \in \mathcal{M}} L(y|M)d\pi_{pr}(M)}. \quad (11)$$

Consider the case in which  $\hat{R}(M)$  is any forecast of model  $M$ , not necessarily a loss reserve. Its posterior mean is

$$E[\hat{R}(M)|Y] = \int_{M \in \mathcal{M}} \hat{R}(M)d\pi_{po}(M), \quad (12)$$

which is a weighted average of the forecasts across all models, with weights  $d\pi_{po}(M)$ . The concept of averaging across models was introduced to the loss reserving context by Taylor (1985).

This may be noted as an estimator of ISME (6).

## 6. LASSO Estimation of Internal Model Structure Error and Other Forecast Errors

### 6.1. The LASSO

#### 6.1.1. Theoretical Background

Consider a model  $M$  of the data set  $Y$  that takes the form

$$Y_{ij} \sim F(i, j; \beta, \theta), \quad (13)$$

where  $F$  is a distribution function (d.f.) defined within some family by parameter vectors  $\beta, \theta$ , and subject to

$$E[Y_{ij}] = f_M(Y; i, j, \beta), \quad (14)$$

for some function  $f_M$ . Thus, it is supposed that the mean  $E[Y_{ij}]$  is defined by the parameters  $\beta$ , and the distribution around this mean defined by the parameters  $\theta$ .

The parameter vector  $\beta$  is left abstract in the present theoretical discussion, but will be given specific meaning in Section 6.1.2. In the meantime, a short commentary is provided as an aid to the reader's orientation.

It is supposed that cell means are subject to some model that is parametric but of unspecified form. The model may depend on  $i, j$ , but it may also depend on other covariates, particularly functions of  $i, j$ . For example, the chain ladder would be one such model, in which the right side of (14) consists of row and column effects, one for each row and column. A  $\beta$  parameter is associated with each row and column.

Another example might specify the development pattern (column effect) by means of a parametric curve rather than column-specific parameters. A sub-vector of  $\beta$ , which might contain only 2 or 3 components, would then describe the curve. A variation of this model might allow for the development pattern to change over rows. The model would then include row-column interaction terms described by further  $\beta$  parameters.

The LASSO, as used in McGuire et al. (2021) and here, makes extensive use of linear splines. For example, the log of the development pattern consists of a linear spline, each of whose linear segments is associated with a  $\beta$  parameter that stipulates its gradient.

A LASSO regression estimates  $\beta$ , of dimension  $p$ , as

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left( \delta(Y, \hat{\mu}(\beta)) + \sum_{r=1}^p \lambda_r |\beta_r| \right), \quad (15)$$

where the dependence of  $\hat{\mu}$  on  $\beta$  has been made explicit,  $\delta$  is some measure of the separation between  $Y$  and  $\hat{\mu}(\beta)$ , the  $\beta_r$  are the components of  $\beta$ , and the  $\lambda_r \geq 0, r = 1, \dots, p$  are fixed values selected by the modeller.

A choice of  $\delta$  such that the regression reduces to a GLM in the event that all  $\lambda_r = 0$ , is the negative log-likelihood function  $\delta(Y, \hat{\mu}(\beta)) = -\ell(Y|M)$ , in which case (15) becomes

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left( -\ell(Y|\beta) + \sum_{r=1}^p \lambda_r |\beta_r| \right), \quad (16)$$

where the model  $M$  has been represented by its parameter vector  $\beta$ .

There are a couple of fundamental derivations of the estimator (15), specifically as (Hastie et al. 2009):

- (a) The minimizer of the  $\delta$  quantity subject to an upper bound on each  $|\beta_r|$ ;
- (b) The maximizer of the *a posteriori* estimate of the  $\delta$  quantity when each coefficient  $\beta_r$  is random, subject to a Laplace density.

Case (b) will be of interest here, and so is explained in a little more detail. It is assumed that  $\beta_r$  is distributed with Laplace probability density

$$p(\beta_r) = 1/2 \lambda_r \exp(-\lambda_r |\beta_r|), \quad (17)$$

where  $\lambda_r$  is a scale parameter. This is a symmetric density about a zero mean, and

$$\operatorname{Var}(\beta_r) = 2/\lambda_r^2. \quad (18)$$

Then, the posterior likelihood of  $\beta$  is proportional to  $\exp\{-(-\ell(Y|\beta) + \lambda^T |\beta|)\}$ , where the modulus operates on a vector element-wise and  $\lambda$  denotes the column  $p$ -vector with the  $\lambda_r$  as components. This posterior likelihood may be compared with (16).

### 6.1.2. Implementation

The application of the LASSO to data follows the procedure set out in McGuire et al. (2021). The target variable  $Y$  denotes the observed claim payments over a  $40 \times 40$  triangle. The model takes the same form as a GLM with log link, specifically supplementing (13) and (14) with

$$F(i, j; \beta, \theta) = \operatorname{Gamma}(\mu_{ij}, \phi), \quad (19)$$

a Gamma d.f. with dispersion parameter  $\phi$ , and

$$\mu_{ij} = f_M(Y; i, j, \beta) = \exp(X\beta). \quad (20)$$

For discussion of this log-linear form see Section 6.1.1, just after (14).

In this formulation, the parameter  $\theta$  from (13) is set equal to the  $\phi$ , the over-dispersion parameter. This parameter requires estimation, which is discussed later in this sub-section.

The Gamma distribution has been chosen as simple, practical alternative that usually provides a reasonable representation of the data. It should be admitted, nonetheless, that, if one places emphasis on cell distributions, then the Gamma, and perhaps all commonly used 2-parameter distributions, may be found wanting.

Any error in the choice of Gamma would be accounted for as model distribution error in the decomposition (5) (see also Figure 1). However, as the focus of the present paper is model structure error, the matter of model distribution error is disregarded. In any event, the limited amount of available research on model distribution error (Taylor 2021) hints that it may be small.

A LASSO can be implemented by means of the *R* package *glmnet*, but only with Poisson ( $\phi = 1$ ), not the Gamma distribution. The solution is to do exactly this but, later, in the BMA of Section 6.2, adjust all Poisson distributions to Gamma with suitable  $\phi$ . A Gamma error term is preferred as being more realistic than Poisson but, although the *glmnet* package purports to provide the means for this, numerical difficulties precluded its application here.

McGuire et al. (2021) propose that the covariates represented by design matrix  $X$  be those of a basis set that linearly spans a sufficiently extensive function space. The choice there, retained here, was

- Ramp functions (2)  $R_K(i), R_K(j), R_K(t), K = 0, 1, \dots, 39$ ;
- Interactions of Heaviside functions  $H_k(i)H_\ell(j), H_k(i)H_g(t), H_g(t)H_\ell(j), k, g, \ell = 2, 3, \dots, 40$  for interactions.

All covariates were standardized, as described by McGuire et al. (2021). The orders of such basis sets are typically in the thousands or tens of thousands, depending on the dimensions of the data being used, but the LASSO will eliminate most of these from the model.

The vector  $\lambda$  is restricted to the form  $(0, \dots, 0, \lambda', \lambda', \dots, \lambda')^T$ , where  $\lambda' > 0$  and the leading zeros relate to specific data features that the modeller regards as of certain existence and wishes to force into the model—typically, an intercept term is fitted without penalty. Thus, the terms of the linear response corresponding to these covariates are always included without penalty, and all other members of the linear response carry penalty parameters that are positive and equal. Details appear in McGuire et al. (2021).

A sequence of values of  $\lambda'$  is chosen, covering a range from large to small. Let these be denoted by  $\lambda'^{(q)}, q = 1, \dots, Q$ , and let  $\lambda^{(q)}$  denote the vector  $(0, \dots, 0, \lambda'^{(q)}, \lambda'^{(q)}, \dots, \lambda'^{(q)})^T$ , where  $\lambda'^{(q)} \downarrow$  as  $q \uparrow$ . For each  $q$ , the LASSO is fitted to the data  $Y$ , i.e.,  $\beta$  is estimated according to (15), and the estimate denoted by  $\hat{\beta}_{(q)}$ . This induces a model  $M_q$ .

The fit of each  $M_q$  to data is assessed by 8-fold cross validation using the Poisson deviance as the loss function. The cross-validation loss in the  $s$ -th fold is denoted by  $CV_{qs}$ . The average of the  $CV_{qs}$  values across  $s$  is denoted by  $\overline{CV}_q$  and the standard error of the same  $CV_{qs}$  values by  $SE_q$ .

Two models that are conventionally regarded as “optimal” in the literature are selected from the collection  $\{M_q, q = 1, \dots, Q\}$ . These are:

- A “minCV” model  $M_{q_{min}}$  such that  $\overline{CV}_{q_{min}} = \min_{1 \leq q \leq Q} \overline{CV}_q$ ;
- A “1se” model  $M_{1se}$ , where  $M_{1se} = M_{q_{1se}}$  such that  $q_{1se} = \min\{q : \overline{CV}_q \leq \overline{CV}_{q_{min}} + SE_{q_{min}}\}$ .

At this point, the parameter  $\phi$  is estimated. The model structure (i.e., the covariates with non-zero coefficients) of the 1se model is extracted, and a GLM fitted to this structure. Henceforth, the preferred Gamma error is assumed, so the GLM is fitted using the  $R$  function  $glm()$  from the *stats* package with the same weights as in the LASSO fit. Since only a crude estimate of  $\phi$  is yielded by  $glm()$ , the maximum likelihood estimate of  $\phi$  is obtained from the  $gamma.dispersion()$  function from the *MASS* package.

This last value of  $\phi$  is used in conjunction with an assumed Gamma error for all calculations henceforth. It is not re-estimated in the bootstrap replications of Section 7. The GLM itself is of no particular interest other than the estimation of  $\phi$ .

The 1se and minCV models generate the LASSO estimates of outstanding claim liability chosen by McGuire et al. (2021). The extrapolation of a model of past data to the future requires decisions as to the extent to which any trends identified in the past are also extrapolated, and to what extent. Mathematically, this requires specification of  $X^{**}$  from (7).

For this purpose, the decisions of McGuire et al. (2021) have been followed, in that all past trends have been extended into the future without modification. In more precise terms, if the linear response included in the model of past data is  $f(i, j, t)$  (consistent with a linear combination of ramp functions and products of Heaviside functions), then the linear response for the future will also be taken as simply  $f(i, j, t)$ . In other words,  $X^{**}$  from (7) is assumed to be 0.

This has the benefit of even-handedness in the generation of a model set. There is no user intervention in the exclusion of models that fail to conform to preconceptions. On the other hand, however, it does admit some models that might be viewed as dubious. For

example, any payment quarter effect that is modelled to take effect in the last few payment quarters of experience will be extended indefinitely into the future. Ultimately, such exotic models may need to be pruned away (Section 6.2.5).

### 6.1.3. Model Bias

The LASSO’s shrinkage of parameter estimates induces bias in model estimates. For this reason, the LASSO is sometimes used just for selection of the model form, and then the selected model is refitted to the data without penalty, i.e., as a GLM. The result should be approximately unbiased.

McGuire et al. (2021) carried out some experimentation with this procedure in relation to the example data sets of Section 8.1, but commented that the results were not encouraging in that the refit caused a substantial deterioration in model performance in some cases.

See the further comment in Section 6.2.1.

## 6.2. Bayesian Model Averaging of LASSO

### 6.2.1. Generation of Model Set

The cell mean structure within the LASSO model is given by (14). Variation in  $\beta$  there will create a multiplicity of models  $\mathcal{M} = \{f_M(\cdot, \cdot, \cdot, \beta), \beta \in \mathbb{R}^p\}$ . It is of particular interest that, as  $\lambda_r$  increases to a certain threshold, the LASSO forces  $\beta_r$  to zero for that  $\lambda_r$  and all greater values. This amounts to dropping a covariate from the model, equivalent to changing the cell mean’s algebraic structure.

This is exactly one of the requirements of Section 5.1. In fact, the LASSO provides a rich model set that includes both variation in model algebraic structure and parametric variation within each distinct structure. The latter form of variation arises from variation in the penalty parameter, and is lost if a GLM is fitted to a LASSO model structure.

The BMA to be carried out in relation to  $\mathcal{M}$  is of the **M-open** type (Bernardo and Smith 1994), which is to say that the true model is not asserted to lie within  $\mathcal{M}$ , but rather that  $\mathcal{M}$  is a collection of models from which a reasonable proxy for the true one may be selected.

### 6.2.2. Model Likelihood

As explained in Section 6.1.2, the process of BMA about to be described assumes a Gamma distribution for each observation  $Y_{ij}$ . The log-likelihood is

$$\ell(Y|\beta) = \sum_{(i,j) \in \Delta} [\gamma \ln c_{ij} - \ln \Gamma(\gamma) + (\gamma - 1) \ln y_{ij} - c_{ij} y_{ij}], \tag{21}$$

where  $\gamma = 1/\phi$ ,  $c_{ij} = 1/\phi \mu_{ij}$ , with the  $\mu_{ij}$  given by (20) and the estimation of  $\phi$  is described in Section 6.1.2.

In the case of (20),  $X$  consists of columns representing all functions in the basis set defined in Section 6.1.2. In the row corresponding to cell  $(i, j)$ , each of the functions is evaluated for those values of  $i, j$ .

### 6.2.3. Prior Distributions of LASSO Models

From (17),

$$p(\beta) = 2^{-p} \left( \prod_{r=1}^p \lambda_r \right) \exp(-\lambda^T |\beta|), \tag{22}$$

assuming the random parameters  $\beta_r$  are independent, and this may be viewed as a prior distribution on the models  $M \in \mathcal{M}$ .

As appears in Section 6.1.2, the vector  $\lambda$  is defined by the scalar  $\lambda'$ . It is of especial relevance that, in (22),  $\lambda'$  functions as a dispersion parameter for the prior. There is a correspondence  $q \Leftrightarrow \lambda'^{(q)}$ , so that models may be labelled equivalently by  $q$  or  $\lambda'$ . As noted in Section 6.1.2, models are numbered in such a way that  $\lambda'^{(q)} \downarrow$  as  $q \uparrow$ . Decreasing

$\lambda^{(q)}$  implies lower penalty on parameters, and so model complexity increases steadily with increasing  $q$ . In the limit  $\lambda^{(q)} = 0$ , the LASSO degenerates to a GLM with the set of covariates equal to the entire basis set, i.e., an extremely complex model.

Similarly, high values of  $\lambda^{(q)}$  force the model toward simplicity. As  $\lambda^{(q)} \rightarrow \infty$ , the model tends to one containing only the unpenalized covariates.

Each value of  $\lambda^{(q)}$  also induces a posterior distribution  $\pi_{po} \left( M_q \mid y; \lambda^{(q)} \right)$ , where the dependence on  $q$  appears twice; first, the algebraic structure of the model  $M_q$  depends on  $q$  and, second, the posterior is influenced by the prior dispersion parameter. This sets up a sequence of relations  $q \Leftrightarrow \lambda^{(q)} \Rightarrow \pi_{po}^{(q)}$ , where the last term is written as an abbreviation for  $\pi_{po} \left( M_q \mid y; \lambda^{(q)} \right)$ .

Recall, however, that all of this applied in the context of Poisson likelihood (Section 6.1.2). Subsequently, this was replaced by a Gamma likelihood, with consequent change to the posterior. If  $\pi_{po[G]}^{(q)}$  denotes the posterior that replaces  $\pi_{po}^{(q)}$  when the Poisson likelihood is replaced by Gamma, then the last relation becomes  $q \Leftrightarrow \lambda'_{[G]}^{(q)} \Rightarrow \pi_{po[G]}^{(q)}$ , where  $\lambda'_{[G]}^{(q)}$  is now the Laplace prior dispersion parameter to “equate”  $\pi_{po[G]}^{(q)}$  to  $\pi_{po}^{(q)}$ , specifically to “equate” their modes. The quotes appear here to indicate that equality will usually be only approximate, due to discreteness of  $q$ .

The relation between  $\lambda^{(q)}$  and  $\lambda'_{[G]}^{(q)}$  is unknown. In principle, it could be derived by analytic comparison of the posteriors  $\pi_{po}^{(q)}$  and  $\pi_{po[G]}^{(q)}$ . Alternatively, and more simply, any required  $\lambda'_{[G]}^{(q)}$  may be found numerically as that value that, for any given  $q$ , produces the required “equality” between  $\pi_{po}^{(q)}$  and  $\pi_{po[G]}^{(q)}$ . This course has been followed here.

Section 6.1.2 identified the minCV and 1se models as significant special cases. The associated lambda values are denoted  $\lambda'_{[G]}^{(minCV)}$  and  $\lambda'_{[G]}^{(1se)}$ , respectively. Subsequent sections will rely mainly on these as relating to reasonable priors (equivalently, selections of  $\lambda'_{[G]}$ , which may not be specific to any LASSO model  $q$ ). Moreover, as each of these models will usually exhibit a high degree of compatibility with the data, the posterior distribution that follows from its prior will usually be centred somewhere in the vicinity of the posterior mode. In fact, the 1se model will be defined as the **primary model**, meaning that the actuary is assumed to use this as the source of an adopted loss reserve.

However, it will also be useful to investigate the extent to which  $\lambda'_{[G]}$  may credibly deviate from these selections. Therefore, two values of this dispersion parameter are identified as of special interest. These may be labelled “simple-model” and “complex-model” dispersion parameters, and are defined as follows:

- **Simple:** the parameter is  $\lambda'_{[G]} = \lambda'_{[G]}^{(simp)}$ , such that

$$\sum_{q=q_{1se}}^Q d\pi_{po[G]} \left( M_q \mid y, \lambda'_{[G]}^{(simp)} \right) = \varepsilon; \tag{23}$$

and

- **Complex:** the parameter is  $\lambda'_{[G]} = \lambda'_{[G]}^{(comp)}$ , such that

$$\sum_{q=1}^{q_{min}} d\pi_{po[G]} \left( M_q \mid y, \lambda'_{[G]}^{(comp)} \right) = \varepsilon; \tag{24}$$

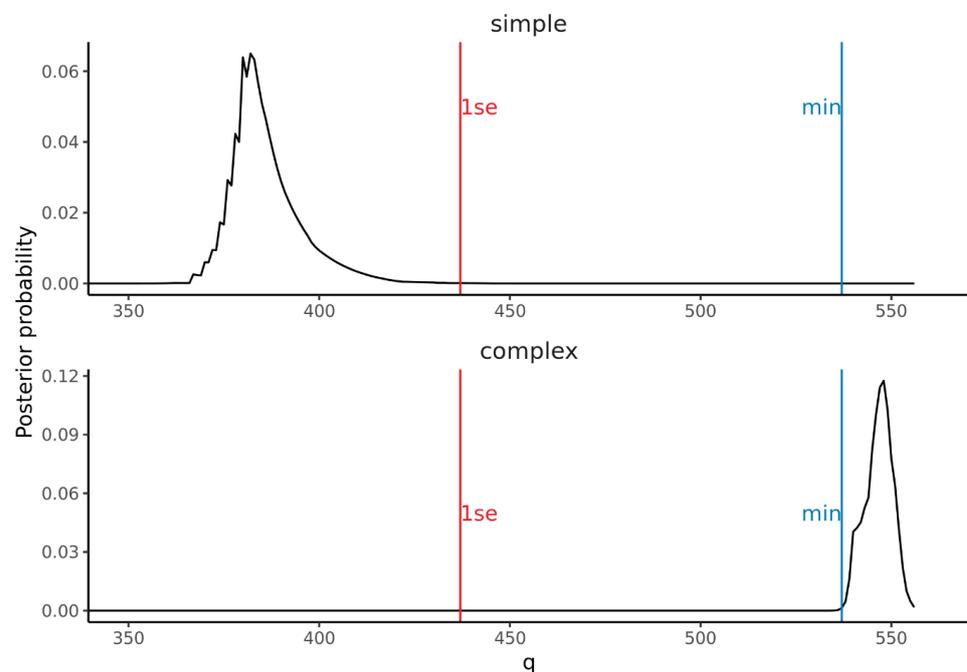
for some suitably small value  $\varepsilon > 0$ .

Thus,  $\lambda'_{[G]}^{(simp)}$  is the prior dispersion parameter which, when applied to all models  $M_q$ , is just high enough to force the aggregate prior probability of all models  $M_{1se}$  and those which are more complex below a threshold probability. This is the value of  $\lambda'_{[G]}$  that, when

applied across all  $M_q$ , generates the simplest models that are even remotely consistent with the data and model  $M_{1se}$ . The parameter value  $\lambda'_{[G]} = \lambda'^{(comp)}_{[G]}$  has a corresponding meaning; it selects out the most complex models acceptable. The threshold value used here is  $\varepsilon = 0.0005$ , so a good deal of liberty has been allowed in the range  $(\lambda'^{(comp)}_{[G]}, \lambda'^{(simp)}_{[G]})$ .

Sometimes no value  $\lambda'^{(simp)}_{[G]}$  or (particularly)  $\lambda'^{(comp)}_{[G]}$  can be found satisfying the required condition (23) or (24). In this case, there is no member of the family of LASSO models so simple (or complex) that it is inconsistent with the data for the stipulated threshold.

Examples of these simple-model and complex-model posteriors are given in Figure 2, where values of  $q$  are listed on the horizontal axis. Detailed analyses of four different data sets will be described in Section 8 and the specifics of each data set in particular in Section 8.1. The examples in Figure 2 are drawn from Data Set 4 described therein. The top half of the figure illustrates the simple model posterior probability, and the bottom half the complex. The minCV and 1se models are also marked on the horizontal axis.



**Figure 2.** Examples of simple-model and complex-model posterior distributions.

#### 6.2.4. Prior Distributions for Bayesian Model Averaging

The LASSO generates a family of models  $\mathcal{F} = \{M_q, q = 1, \dots, Q\}$ , each tagged with a prior distribution characterized by dispersion parameter  $\lambda'^{(q)}_{[G]}$ . The prior relates to the set of possible parameters of the model  $M_q$ .

The aim is now to conduct BMA over  $\mathcal{F}$ . For this, a single prior is required across all models in  $\mathcal{F}$ . This might reasonably be taken as defined by a sensibly selected value  $\lambda'_{[G]}$ . For example, the selection  $\lambda'_{[G]} = \lambda'^{(1se)}_{[G]}$  would take the prior distribution associated with LASSO model  $M_{1se}$ , and assume it to apply to all models in  $\mathcal{F}$ . When normalized across  $\mathcal{F}$ , this gives a prior distribution on  $\mathcal{F}$ , thus

$$Prob[M = M_q] = P_q = \frac{p_{1se}(\beta^{(q)})}{\sum_{q=1}^Q p_{1se}(\beta^{(q)})}, \tag{25}$$

where  $p_{1se}(\beta^{(q)})$  is obtained from (22) with the substitutions  $\lambda = \lambda_{[G]}^{(1se)}$ ,  $\beta = \beta^{(q)}$ , this last term representing the parameter vector of model  $M_q$ .

BMA of any forecast of any random quantity  $F$  then gives an average of

$$\bar{F}_{BMA} = \sum_{q=1}^Q P_q F_q, \tag{26}$$

where  $F_q$  is the forecast of  $F$  given by model  $M_q$ , as parameterized by the LASSO.

Though it may be reasonable, as suggested above, to select the prior  $p_{1se}(\beta_{1se})$  for use in (25), it is of interest to consider other possibilities. Therefore, re-write (25) in the more general form:

$$Prob[M = M_q] = P_q = \frac{p_0(\beta^{(q)})}{\sum_{q=1}^Q p_0(\beta^{(q)})}, \tag{27}$$

where  $p_0(\beta^{(q)})$  is constructed as for  $p_{1se}(\beta^{(q)})$  except for the choice  $\lambda = \lambda_{[G]}^{(0)}$ , which is some preferred but unspecified prior.

BMA will be carried out below under various selections of prior, namely (in order of increasing model complexity):

- Simple:  $\lambda_{[G]}^{(0)} = \lambda_{[G]}^{(simp)}$ ;
- 1se:  $\lambda_{[G]}^{(0)} = \lambda_{[G]}^{(1se)}$ ;
- minCV:  $\lambda_{[G]}^{(0)} = \lambda_{[G]}^{(minCV)}$ ;
- Complex:  $\lambda_{[G]}^{(0)} = \lambda_{[G]}^{(comp)}$ .

The middle two of these would be regarded as yielding realistic results. The other two indicate the extremes to which results may be taken before consistency with data is totally lost.

### 6.2.5. Pruning of Model Set

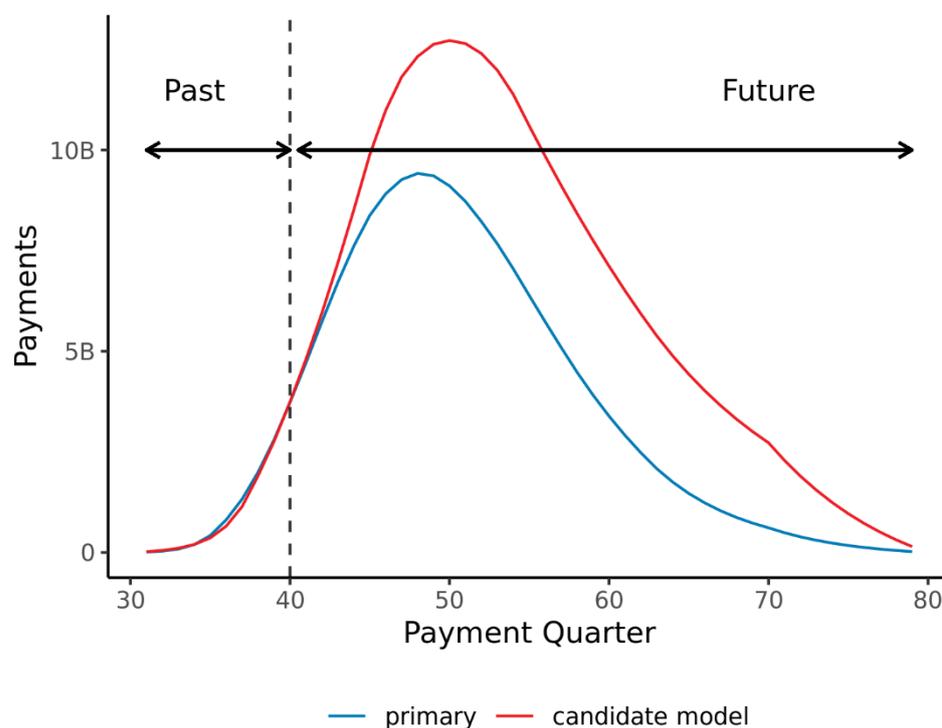
Loss reserve forecasting is essentially an exercise in extrapolation. Certain functions are fitted to past data, and then extrapolated to forecast future data.

In the highly supervised environment of conventional actuarial loss reserving, this rarely creates difficulty. The actuary maintains strict control over the explanatory functions, often assigning physical explanations to them. Supervision of extrapolation behaviour is part of this control; if it appears unreasonable, the actuary will modify the model.

In the case of the LASSO and other machine learning methods, on the other hand, the volume of models under test precludes this level of supervision. Moreover, the functions used to fit past data are assembled abstractly from a set of basis functions. Interpretation of the results may or may not be elementary.

It is in the nature of such exercises that strict control may be maintained over the fit of models to the data (e.g., by cross-validation error, as in Section 6.1.2), yet there be no control over the behaviour of these models in future. Hence, good quality fit of a model to data provides no assurance of reasonable forecast behaviour.

Figure 3 provides an example of this in relation to Data Set 2, as defined in Section 8.1. Claim payments, aggregated over the last 10 accident quarters, have been plotted in \$billion against payment quarter for both past and future quarters, in respect of both the primary model and an alternative candidate. The past includes payment quarters up to and including  $t = 40$ , as indicated in the figure. The future is represented by  $t > 40$ .



**Figure 3.** Example of high-quality fit to data but poor extrapolation.

It can be seen that the candidate model coincides with the primary almost perfectly over the past payment quarters, and its posterior probability is in fact 0.67. However, its forecast quickly diverges from the primary's over future payment quarters. Ultimately, it forecasts future liability (for the last 10 accident quarters) 56% higher than the corresponding primary forecast.

It is highly unlikely that an actuary, having invested in a primary model, would regard a model that diverges to this degree as suitable for inclusion in the model set. If surveillance over inclusions were maintained, this particular case would be censored. For practical purposes, this censorship must be automated.

Table 2 parameterizes the pruning used here. The forecasts of each model are confronted by a number of **inclusion gates**. If they fail to pass through all of these gates, the model is censored.

Examples of the meaning of the gates in the table are as follows. The row dealing with the last five accident quarters means that a model will be discarded if the ratio of its forecast of all claim payments in respect of the last five accident quarters prior to the valuation date to the corresponding forecast from the primary model fails to lie between 80% and 125%.

Similarly, the row dealing with the next five payment quarters means that a model will be discarded if the ratio of its forecast of future claim payments in respect of the five payment quarters (over all past accident quarters) succeeding the valuation date to the corresponding forecast from the primary model fails to lie between 87% and 115%.

It is found that the dominant mode of censorship relates to breaches of the accident, rather than payment, quarter limits. This seems unsurprising, as the limited experience of recent accident years provides little guide as to future development. It is easy for extrapolation functions to go awry.

The inclusion gates must be subjectively chosen, since they represent the actuary's judgement on the maximum acceptable deviation from the primary model. The gates are likely to vary from one reserving exercise to another, as the line of business and length of development tail, and other particulars change.

It is noteworthy that this introduces an unwelcome element of subjectivity into an otherwise data-driven estimation process. Unfortunately, we do not see an alternative.

Failure to address the issue causes delinquent models to lead to unduly large estimates of internal model error that lack credibility.

**Table 2.** Pruning of LASSO models.

Component of Forecast	Inclusion Gates (as Percentage of Primary Forecast)	
	Lower Limit	Upper Limit
<b>Accident quarters</b>		
Last 2	75	133
Last 5	80	125
Last 10	83	120
<b>Payment quarters</b>		
Next 2	91	110
Next 5	87	115
Next 10	83	120

### 6.2.6. Internal Model Structure Error from the Primary Model

Equations (9) and (10) give expressions for the mean and variance of a posterior distribution of loss reserve. Equation (12) is the counterpart of (9) when the posterior derives from BMA across LASSO models. It will often be convenient to convert the posterior variance to a coefficient of variation (“CoV”).

In this context, the posterior mean may be interpreted as an estimate of the loss reserve, and, as noted at the end of Section 5.3, the CoV as a measure of IMSE. Numerical examples of this are presented in Section 8.2.

## 7. Bootstrapping the LASSO

### 7.1. Motivation

Section 6.2.6 described the estimation of IMSE for the primary model. In this sense, the problem is solved at that point. However, the estimate may be regarded as a little “thin”, since the number of models with posterior probability greater than vanishingly small may be limited. It would not be unusual for this number to be in the range 10–30. This is likely to lead to under-estimation of the IMSE.

Bootstrapping of the LASSO generates multiple pseudo-data sets, leading to multiple posterior distributions, which add to the reliability of the IMSE estimate.

### 7.2. Form of Bootstrap

A semi-parametric form of bootstrap is used. This terminology is taken from Shibata (1997), and its implementation in relation to loss reserving (in relation to a GLM in that case) is explained in Taylor and McGuire (2016).

Pseudo-data sets are formed on the basis of the LASSO standardized residuals. Standardization requires an estimate of variance for each cell. In accordance with the assumption of a Gamma distribution in Section 6.1.2, the variance of the  $(i, j)$  cell is estimated as  $\hat{\phi}\hat{\mu}_{ij}^2$ , where  $\hat{\mu}_{ij}$  is the value fitted to the cell by the LASSO and  $\hat{\phi}$  is the estimate of  $\phi$  obtained as described at the end of Section 6.1.2.

The LASSO is known to be biased, and so the set of standardized residuals must be centred before use in the bootstrap. They are all shifted by a constant quantity, such that their arithmetic average is zero. One needs to check that the resulting centred standardized residuals appear iid, and they are then re-sampled to generate multiple pseudo-data sets  $Y^{[b]}$ ,  $b = 1, \dots, B$  in the usual way, i.e.,  $Y_{ij}^{[b]} = \rho_{ij}^{[b]}\hat{\mu}_{ij} + \hat{\phi}\hat{\mu}_{ij}^2$ , where the  $\rho_{ij}^{[b]}$  are the re-sampled residuals. Hereafter, all quantities superscripted by  $[b]$  will relate to this  $b$ -th bootstrap replication.

A brief pause at this point to consider the justification for use of the bootstrap is worthwhile. The key assumption in its use is that the centred standardized residuals  $\rho_{ij}$

from the LASSO described in Section 6.1 are (at least approximately) iid, so that, for each  $b$ , the  $\rho_{ij}^{[b]}$  may be considered a re-sampling of them.

Such checks are made by examination of sets of residual plots, where the  $\rho_{ij}$  are plotted against  $i, j, i + j - 1$  (payment quarter), and possibly other variables. These plots are assessed by eye for trendlessness in the means and homoscedasticity.

As with all forms of regression, the residuals from each model carry a small degree of dependency and, in line with standard bootstrapping practice, it is disregarded. In fact, since the models to which BMA is applied in Section 6.2 are of different design, their residuals would be likely to exhibit different patterns of dependency from one model to another, and hence their average would be likely to produce a lesser dependency than any single one of the models.

Each pseudo-data set is then submitted to the entire process described in Section 6, i.e., LASSO estimation, followed by BMA, with some modifications to censor unreasonable models and to centre the bootstrapped results to the original forecasts for each prior. Centring the bootstraps leads to a change in posterior probabilities, so the process followed is:

- Carry out LASSO estimation to produce models  $M_q^{[b]}, q = 1, \dots, Q^{[b]}$ , their liability estimates  $\hat{R}(M_q^{[b]} | Y^{[b]})$ , and associated posterior probabilities  $d\pi_{po}^{[b]}(M_q^{[b]})$ . Note that it may be advisable to use a less onerous level of censoring at this level than that set out in Table 2 to avoid eliminating models that may be acceptable after centring (next point). Hence, temporary inclusion gates are obtained by scaling those in the table by a factor of 1.4.
- Calculate the scaling factor required to scale the sample mean of the posterior means from each pseudo-data set to the original posterior means.
- Apply this scaling factor to the estimates of each model and recalculate the liability estimates,  $\hat{R}(M_q^{[b]} | Y^{[b]})$ .
- Apply censoring at the desired level, as discussed in Section 6.2.5.
- Calculate the associated posterior probabilities  $d\pi_{po}^{[b]}(M_q^{[b]})$ .
- Drop any bootstrap replication  $b$  for which posterior distribution over models  $M_q^{[b]}$  is concentrated on fewer than five models, in the sense of having fewer than five models with posterior probability masses each exceeding 0.0001.

The subsequent BMA is dealt with in Section 7.3.

### 7.3. The Bootstrap Matrix

Now form a  $B \times Q_{max}$  matrix (**the bootstrap matrix**), where  $Q_{max} = \max\{Q^{[b]}, b = 1, \dots, B\}$ , and the  $b$ -th row contains the liability estimates  $\hat{R}(M_q^{[b]} | Y^{[b]})$ , and contains missing values for columns  $q = Q^{[b]} + 1, \dots, Q_{max}$ . Some of the rows of the matrix will be empty when all models of those rows have been censored. In machine learning parlance, the bootstrap matrix might be considered an ensemble, or a ragged array.

Note that, for fixed  $q$ , the models  $M_q^{[b]}$  vary over  $b$ , i.e., a given column of the matrix might not contain the same model in all rows. Indeed, the models might be different in all rows of that column.

BMA produces an estimated liability  $E_{po}[\hat{R}^{[b]}(M) | Y^{[b]}]$  and estimated IMSE equal to  $CoV_{po}[\hat{R}^{[b]}(M) | Y^{[b]}]$  for data  $Y^{[b]}$ , where the subscripts  $po$  indicate that the moments are taken with respect to the posterior distribution (compare with Section 5.3). Denote these quantities by  $m^{[b]}$  and  $w_{ISME}^{[b]}$ , respectively. In addition, let  $s_{IMSE}^{2[b]} = (m^{[b]}w_{ISME}^{[b]})^2$ .

Impute a uniform distribution  $U$  across the rows of the matrix, i.e., all bootstrap replications are considered equally likely. Then more reliable estimators of liability and IMSE can be obtained by averaging over rows of the bootstrap matrix. These are, respectively,

$$m = E_U [m^{[b]}] = E_U E_{p_0} [\hat{R}^{[b]}(M) | Y^{[b]}], \quad (28)$$

$$s_{IMSE}^2 = E_U [s_{IMSE}^{2[b]}] = E_U Var_{p_0} [\hat{R}^{[b]}(M) | Y^{[b]}], \quad (29)$$

$$w_{IMSE}^2 = CoV_{IMSE}^2 = \frac{s_{IMSE}^2}{m^2}. \quad (30)$$

Conventional use of the bootstrap in loss reserving enables estimation of parameter error (PaE). This is achieved by means of a well-established procedure (England and Verrall 1999; Taylor and McGuire 2016). PaE is estimated from the variation in loss reserve estimates over bootstrap replications. Process error can be added quite separately, and does not require the bootstrap.

A parallel procedure can be followed in relation to the bootstrap matrix here. One can calculate

$$s_{Pa}^2 = Var_U [m^{[b]}] = Var_U E_{p_0} [\hat{R}^{[b]}(M) | Y^{[b]}], \quad (31)$$

$$w_{Pa}^2 = \frac{s_{Pa}^2}{m^2}$$

and this is an estimate of parameter error. Once again, the estimation of process error CoV, denoted  $w_{pr}$ , is separate and straightforward, using sampling from its known distribution (Section 7.2), as described by England and Verrall (1999, 2001, 2002), Taylor (2000), and Taylor and McGuire (2016).

Note that, by the law of total variance,

$$Var [\hat{R}^{[b]}(M)] = Var_U E_{p_0} [\hat{R}^{[b]}(M) | Y^{[b]}] + E_U Var_{p_0} [\hat{R}^{[b]}(M) | Y^{[b]}] = s_{Pa}^2 + s_{IMSE}^2, \quad (32)$$

asserted by (29) and (31), where the Var and E operators are understood as forming summary statistics from the sample values  $\hat{R}^{[b]}(M)$ , and where the variance on the left has been taken from over the entire bootstrap matrix, i.e., over rows and columns.

The combination of PaE and IMSE has squared CoV

$$w_{Pa+IMSE}^2 = \frac{s_{Pa}^2 + s_{IMSE}^2}{m^2} = w_{Pa}^2 + w_{IMSE}^2. \quad (33)$$

**Remark 1.** *The thinness of the primary estimate of IMSE, and the possibility of under-estimation, was noted in Section 7.1. Each bootstrap replication has the same stochastic properties as the primary, and so each row of the bootstrap matrix, and also the average of them, may lead to under-estimation.* □

**Remark 2.** *In view of Remark 1, the meanings of IMSE and PaE, as estimated by (29) and (31), merit some further consideration. The meaning of the latter is quite clear in the case of a GLM. Here, there is a fixed model structure across all bootstrap replications, and it is applied to the various data sets generated by those replications. Only the parameterization of that structure varies across replications. Quite clearly, this leads to genuine estimation of PaE.* □

When the LASSO is bootstrapped, the situation is quite different. Variation in the data set from one replication to another induces different sets of models. Hence, not only parameterization of models, but the models themselves, vary across replications. In this way, parameter and model errors become entangled, and the distinction between them becomes unclear. Indeed, there may even be a question as to whether the distinction is even meaningful.

From a practical viewpoint, it may be preferable to think in terms of just (33) as unambiguously the combination of parameter and IMSE errors, without concern for the decomposition into the two components. For most practical purposes (e.g., risk margins), only the combination will be required.

The estimates  $w_{ISME}$  and  $w_{Pa}$ , as defined by (29) and (31), are useful for the construction of (33), but no particular meaning may be assigned to the components themselves.

These estimates of model error may be compared with others in the literature. Studies by Shi (2013) and Zhang and Dukic (2013) were both concerned with models of multivariate claim triangles, whereas the present paper is concerned with only single triangles. Comparison therefore requires the multivariate component, represented in each case, by a copula, to be stripped away from those earlier papers.

Zhang and Dukic (2013) modelled a claim triangle using a log-linear structure whose development-year component was represented by a mixed model in which the fixed effects described a linear function and the random effects a cubic spline. The use of splines renders it somewhat similar to the LASSO structure set up in Section 6.1.2. It might be mentioned, however, that the adopted LASSO structure broadens the class of models considered by the inclusion of interaction terms, and also that the LASSO model selection enables the inclusion of a very large class of spline functions in the model set.

Uninformative priors were assigned to the random effects. This aspect of the model differs from that of the present paper, where informative priors are inferred from the LASSO (see Section 6.2.3).

Shi (2013) modelled a claim triangle according to a mixed model containing Tweedie distributed cells with chain ladder parameterization (individual row and column effects) as fixed effects and an added calendar-year effect and cell-specific effect as random effects. When copula parameters are disregarded, the only parameters other than row and column effects are the Tweedie  $p$  and dispersion parameter  $\phi$  (assumed constant across cells), as well as variances of the two random effects.

Shi estimates all model parameters, with standard errors, from data. Various model sets are considered. If, for consistency with the present paper, uncertainty in the row and column effect parameters is considered as parameter error, and uncertainty in the Tweedie  $p$  and dispersion parameter  $\phi$  is considered model distribution error, then model error relates to uncertainty in just the random effect parameters. In other words, model error contains no components relating to the algebraic structure used to represent cell means. Indeed, model error has limited meaning in models of a single claim triangle so long as one restricts oneself to saturated models of the chain ladder type.

## 8. Numerical Results

### 8.1. Data Sets

All data submitted to analysis in this paper are synthetic. Thus, all data sets contain known features. Four synthetic data sets were constructed and analysed. Each consisted of a  $40 \times 40$  quarterly triangle of incremental paid claims  $Y_{ij}$ ; thus, 810 cells. The data were simulated according to exactly the same specifications as those set out in McGuire et al. (2021).

Full details were given in that earlier paper. It suffices here to give just the briefest descriptions of the data sets, as below.

**Data Set 1:** Cell means consist of multiplicative row and column effects, and so this data set is a simple one, consistent with a chain ladder model.

**Data Set 2:** cell means are as for Data Set 1, except that a payment quarter effect has been added. This effect represents SI at a rate that is constant over the first 12 quarters, increases over the next 12, returns to constant over the next 8, then increases once more over the final 8. For any one payment quarter, the rate of SI is constant across development quarters.

**Data Set 3:** this is as for Data Set 2, except that one further complexity has been added. This consists of a sharp increase in claim costs in development quarters 21 and later within

accident quarters 17 and later. This last feature is of particularly difficulty for detection by a model, as it affects only 10 of the 810 cells of which the data set is composed.

**Data Set 4:** this is as for Data Set 2, except that greater complexity has been added to the rates of SI. The rates from Data Set 2 now apply to development quarter 1 and, for any payment quarter, they decline linearly to zero at development quarter 40.

These data sets might reasonably be listed 1, 2, 3, 4 in increasing order of complexity, though estimation might be more difficult in case 3 than case 4.

### 8.2. Primary Model

Each of the four data sets of Section 8.1 are modelled and extrapolated to produce an estimated loss reserve and IMSE on the basis of the primary model, as described in Section 6. The results are set out in Table 3. Since the data are synthetic, the true reserves are known, and these have also been included in the table.

**Table 3.** Forecast reserve and estimated IMSE for primary models.

Data Set	LASSO Model	Loss Reserve			Estimated IMSE (CoV)
		True	Forecast		
			Raw 1se	Posterior	
AUDB	AUDB	AUDB	%		
1	Simple	190		198	0.7
	<b>1se</b>	<b>190</b>	<b>194</b>	<b>194</b>	<b>0.4</b>
	<b>minCV</b>	<b>190</b>		<b>194</b>	<b>0.5</b>
	Complex	190		203	0.8
2	Simple	238		260	0.1
	<b>1se</b>	<b>238</b>	<b>261</b>	<b>260</b>	<b>0.1</b>
	<b>minCV</b>	<b>238</b>		<b>244</b>	<b>3.4</b>
	Complex	238		272	3.1
3	Simple	608		877	1.7
	<b>1se</b>	<b>608</b>	<b>778</b>	<b>777</b>	<b>6.8</b>
	<b>minCV</b>	<b>608</b>		<b>687</b>	<b>2.0</b>
	Complex	608		875	5.8
4	Simple	216		244	0.2
	<b>1se</b>	<b>216</b>	<b>247</b>	<b>247</b>	<b>0.3</b>
	<b>minCV</b>	<b>216</b>		<b>268</b>	<b>0.7</b>
	Complex	216		276	1.2

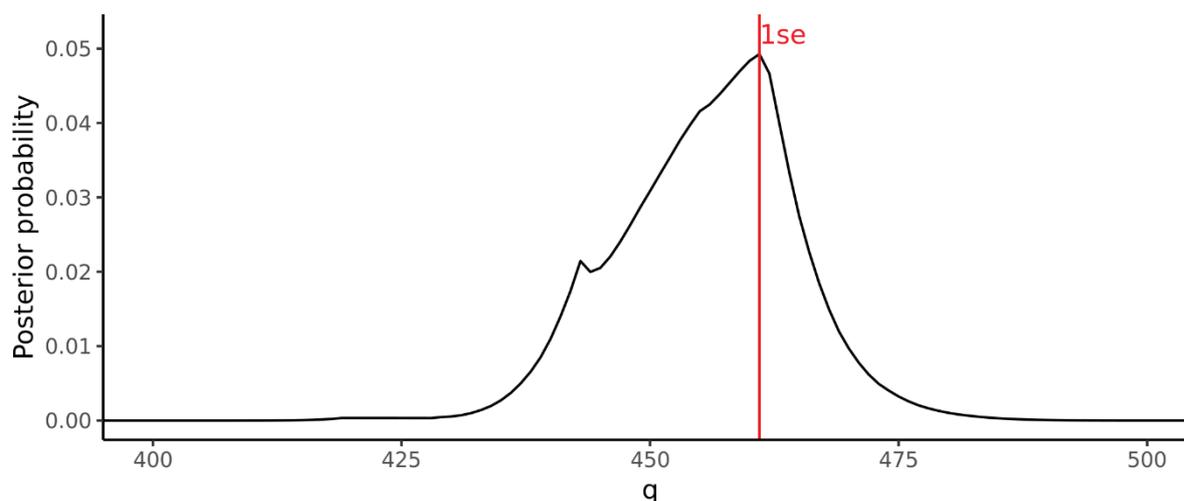
The table includes two sets of forecasts. The “raw 1se” is the 1se estimate from the primary model, obtained by a single application of the LASSO to the data set. The “posterior” is the mean of the posterior distribution obtained in the BMA described in Section 6.1.2.

The “1se” and “minCV” rows are bolded as the most relevant. The “simple and “complex” rows are also included to indicate the extent to which IMSE may vary before the models taken into account by it exhibit lack of fidelity to the data.

It may be noted that the posterior means usually exceed the true liabilities. This is not a product of the BMA, as the difference between the raw 1se estimate and its Bayesian average is small in all cases. The 1se models simply over-estimate liability in these cases.

The modelling of Data Set 1 behaves well, with small estimates of model error, as is to be expected since the data is compatible with a simple chain ladder model, reflecting just row and column effects.

It is to be emphasized that, although Table 3 focuses attention on the CoV of IMSE, the full posterior distribution of loss reserves is available for each row of the table. As an example, Figure 4 displays the posterior corresponding to the 1se prior in the case of Data Set 2. The convolution of this with the distributions of other components can be taken to obtain the distribution of full forecast error.



**Figure 4.** Posterior distribution of loss reserve for 1se prior in Data Set 2.

### 8.3. Bootstrap

The bootstrap procedure described in Section 7 was applied to all four data sets. In each case, the number of bootstrap replications was set to  $B = 500$ , although, as will be seen in the results below, some of these are discarded by the pruning regime of Section 6.2.5. Table 4 displays the results corresponding to the primary model results in Table 3.

**Table 4.** Forecast reserve and estimated IMSE for bootstrapped models.

Data Set	LASSO Prior	Loss Reserve		Estimated IMSE (CoV)	Number of Surviving Bootstraps
		True	Posterior Forecast		
		AUDB	AUDB	%	
1	Simple	190	197	1.1	427
	1se	<b>190</b>	<b>194</b>	<b>1.3</b>	<b>432</b>
	minCV	<b>190</b>	<b>191</b>	<b>2.1</b>	<b>448</b>
	Complex	190	192	2.3	422
2	Simple	238	260	1.3	449
	1se	<b>238</b>	<b>260</b>	<b>1.5</b>	<b>452</b>
	minCV	<b>238</b>	<b>266</b>	<b>2.0</b>	<b>323</b>
	Complex	238	259	2.7	448
3	Simple	608	678	2.1	134
	1se	608	<b>722</b>	<b>1.7</b>	<b>176</b>
	minCV	608	<b>715</b>	<b>2.7</b>	<b>155</b>
	Complex	608	668	2.1	110
4	Simple	216	245	1.0	399
	1se	<b>216</b>	<b>245</b>	<b>1.2</b>	<b>423</b>
	minCV	<b>216</b>	<b>248</b>	<b>2.3</b>	<b>430</b>
	Complex	216	247	2.7	399

A comparison of Table 4 with Table 3 reveals the effect of bootstrapping in supplementing the primary model. The “thinness” of the IMSE estimates in Table 3 becomes apparent in that they are typically lower than their Table 4 counterparts, perhaps indicating a wider range of model structures in the latter. Moreover, within each data set, the estimates from Table 3 progress less regularly as model complexity increases. This last characteristic is probably indicative of reduced sampling error in the results shown in Table 4.

Within each data set, the magnitudes of these estimates of IMSE vary over the four data sets, somewhat as expected; broadly, as model complexity increases, so does estimated IMSE. It is interesting that there is comparatively little difference between the estimates of IMSE for the four data sets, even though the complexity of the data varies between them. This suggests that the LASSO models perform about equally well in all cases.

The final column illustrates how the censorship rates increase with increasing complexity of the data set. In other words, as models are confronted with increasing data complexity, those models include additional features that render their extrapolation more prone to delinquency. This is particularly the case for Data Set 3, for which the forecasting difficulties have been discussed in Section 8.1.

As has been discussed in Section 7.3, the LASSO bootstrap provides an estimate of parameter error (PaE) in addition to ISME, and process errors may be estimated routinely. These are displayed in Table 5. Only the cases of 1se and minCV priors are included as those of greatest interest.

As noted in Remark 2, there is some ambiguity in the separate meanings of PaE and IMSE, but not in the meaning of their combined effects. Table 5, therefore, also contains a column displaying the CoV associated with the sum of these two components of forecast error.

The “Sub-total” error is the sum of parameter, process and IMSE. It is designated in this way to emphasize that it is not, in fact, the total forecast error because it omits the two components not discussed in this paper, namely

- External model structure error;
- Model distribution error.

**Table 5.** Forecast reserve and estimated forecast error components for bootstrapped models.

Data Set	LASSO Prior	Forecast						
		Liability			Error (CoV)			
		True	Posterior	IMSE	Parameter	IMSE+	Process	“Sub-Total”
		Parameter						
		\$billion	\$billion	%	%	%	%	%
1	1se	190	194	1.3	8.0	8.1	0.9	8.1
	minCV	190	191	2.1	11.4	11.6	1.3	11.7
2	1se	238	260	1.5	9.7	9.8	1.2	9.9
	minCV	238	266	2.0	9.8	10.0	1.5	10.1
3	1se	608	722	1.7	12.6	12.7	2.1	12.9
	minCV	608	715	2.7	12.7	13.0	1.4	13.1
4	1se	216	245	1.2	8.6	8.7	0.9	8.7
	minCV	216	248	2.3	11.5	11.8	1.3	11.8

It may be noted that, in the cases of Data Sets 2 and 3, the magnitude of the forecast “sub-total” error depends little on whether it is derived from a 1se or a minCV prior distribution. For Data Sets 1 and 4, a substantial margin appears in favour of the 1se priors, and this can be seen to be attributable mainly to estimated “parameter error”.

Investigation reveals that the minCV distribution of posterior means across bootstrap replications exhibits a markedly longer right tail than its 1se counterpart. The reasons for this have not been investigated further, but note the complex structure of SI for Data Set 4. Mis-estimation of SI as a result of undue model complexity could lead to volatile forecast along future diagonals.

There is some support for this explanation in Table 6, which repeats the parameter error CoVs from Table 5, but now flanked by those resulting from the simple and complex priors. The table displays a clear divide between the two simpler and the two more complex cases.

**Table 6.** Bootstrap estimates of parameter error for Data Set 4.

Prior	Parameter Error (CoV)
	%
simple	7.4
1se	8.6
minCV	11.5
complex	11.8

#### 8.4. Sensitivity to Inclusion Gate Width

All numerical results reported above are governed by the inclusion gates set out in Table 2. Since these gates were subjectively determined, it is natural to enquire into the sensitivity of the results to the gate widths.

Table 7 compares the “Sub-total” forecast errors of Table 5 with the errors that are obtained when all gates are widened by increasing the upper limit by a factor of 1.1 and decreasing the lower limit by the same factor. Thus, the first gate in Table 2 is widened from [0.75, 1.33] to [0.68, 1.46], and the other gates in that table are widened similarly.

**Table 7.** Sensitivity of forecast error to width of inclusion gates.

Data Set	Prior	Sub-Total Error (CoV)	
		Original Gates	Widened Gates
		%	%
1	1se	8.1	12.3
	minCV	11.7	17.1
2	1se	9.9	14.8
	minCV	10.1	18.0
3	1se	12.9	18.8
	minCV	13.1	18.8
4	1se	8.7	12.9
	minCV	11.8	17.0

It is seen that, in very broad terms, widening the inclusion gates by a factor of 1.1 increases forecast error by a factor of about 1.5. Thus, estimated forecast error can be sensitive to the selection of these gates, and they must be selected carefully.

#### 8.5. Reasonableness Check of Results

It is interesting to use a different methodology in an attempt to provide an independent check on the reasonableness of the above estimates of PaE plus IMSE. For this purpose, a GLM has been fitted to each data set, treating the model structure as known, i.e., containing the same explanatory variables as in the simulation specifications. Thus, the GLMs should have low levels of IMSE. The GLMs have then been bootstrapped to yield estimates of PaE. The GLM bootstrap forecasts were subject to the same set of inclusion gates as applied to the LASSO (see Table 2).

Before examining the results of these models, it is useful to consider reasonable expectations of them. As noted in Remark 2, some IMSE may leak into recorded PaE in the case of the LASSO. Accordingly, it would not be surprising if the PaE, as estimated by the LASSO, exceeded that estimated by the GLM.

Table 8 provides the comparison. It is certainly the case that the LASSO 1se estimates of PaE exceed the GLM estimates. Specific observations are that:

- The very simple Data Set 1 and the more complex Data Sets 2 and 3 both produce substantial differences;

- The discrepancy for Data Set 4 is also significant, but of lower magnitude.

**Table 8.** Comparison of LASSO and GLM estimates of parameter error.

Data Set	Prior	Forecast		
		True	Mean	Parameter Error (CoV)
		\$billion	\$billion	%
1	LASSO 1se	190	194	8.0
	GLM	190	214	4.8
2	LASSO 1se	238	260	9.7
	GLM	238	211	5.7
3	LASSO 1se	608	722	12.6
	GLM	608	631	8.3
4	LASSO 1se	216	245	8.6
	GLM	216	232	6.0

These results appear consistent with the expectations set out earlier in this sub-section.

## 9. Conclusions

Section 1 noted that this paper set out to study model error. This is all about forecasting under model uncertainty. Sections 6 and 7 have indeed set out a LASSO-based methodology for the estimation of a specific component of model error, namely internal model structure error, and the methodology has been illustrated numerically in Section 8. The methodology relies on a form of Bayesian model averaging.

There are at least two possible applications of the results. First, IMSE has been evaluated as a standard error in the numerical examples. As a component of forecast error, this standard error could be combined with the other components listed in Section 4, these having been derived by means described elsewhere in the literature.

The combination would usually take the form of simple addition of the IMSE variance to the variance of the sum of other components, since it would usually be reasonable to regard model error as stochastically independent of the other components.

In this way, the (total) standard error associated with a loss reserve forecast can be derived. If the reserve is to be supplemented with a risk margin such that the total lies at some prescribed percentile of the liability at which the reserve is aimed, then the standard error may be sufficient for computation of the percentile. This will usually be so if the percentile concerned is moderate (e.g., 75%) because in many cases the loss reserve forecast can be regarded as approximately normally distributed.

The second form of application attends more closely to the distribution of the loss reserve forecast, and may come into play at higher percentiles (perhaps above 95%). Note that the use of the bootstrap in the estimation of IMSE will yield a distribution of the latter. If a distribution of the sum of other components is also available (usually the case), the convolution of the two distributions may be obtained to yield the full distribution of the loss reserve forecast, and hence its percentiles.

Either of these two cases leads to the estimation of percentiles, and these can function as risk margins.

Neither LASSO nor BMA are new to the statistical literature. However, their application to the actuarial problem of reserving model error is new. This problem has been resistant to treatment in the past to the point of being often disregarded by actuaries or treated in a highly ad hoc fashion.

Although the application of LASSO and BMA is conceptually straightforward, the foregoing sections indicate clearly that their application in practical and realistic reserving situations is far from straightforward. The paper attempts to address the implementation

issues that arise, and it is hoped that this will render the proposed model error estimation of use to practising actuaries.

With one exception, noted in Section 6.2.5, the estimation of IMSE outlined here is rigorous, based on Bayesian model averaging over a space of admissible models. This provides an alternative to the subjective estimation found in prior literature. The literature also contains a couple of previous approaches to model error estimation. Section 7.3 discusses the current model in comparison with those earlier approaches, and notes where and how the current model adds unique contributions.

The exception discussed in Section 6.2.5 is concerned with the fact that some models that fit a data set extremely well may yet forecast abominably. The numerical studies in this paper suggest that this issue occurs with sufficient frequency to be of significant concern, particularly in data sets with a complicated structure. In such cases, there is no apparent data-driven means of excluding the delinquent models. This matter has been dealt with by the establishment of subjectively chosen inclusion gates. Models are rejected as unrealistic if they produce certain results that fail to pass through these gates.

The selection of these gates would be based on the reserving actuary's professional expertise. In one sense, the need to rely on this element of subjectivity is unfortunate. However, the issue cannot be disregarded, as delinquent models have the potential to corrupt the model error estimation completely and produce wildly inflated estimates.

The authors see no alternative to some form of subjective decision on the admissibility of individual models. There are, however, obvious risks associated with it in that under- (over-)estimation of model error will result from unduly strict (lax) selection of the gates. Moreover, the selection requires considerable care, as the estimate of model error is materially sensitive to it (Section 8.4).

The suggested estimation procedure is supplemented by bootstrapping in Section 7. This yields a bonus in that it produces an estimate of PaE as well as IMSE. However, as noted in Remark 2, the supposed estimates of PaE and IMSE are entangled, probably inextricably. Possibly, there is not even any clear meaning to be associated with these separate components.

But all is not lost here. Although there are not reliable estimates of PaE and IMSE (whatever they may be), Remark 2 points out that the estimate of their combined contribution to forecast error is indeed reliable, and this is all that is required for most practical purposes, such as loss reserve risk margins.

The estimates of PaE and IMSE obtained from the LASSO have been subjected to a sanity check in Section 8.5 by comparison with the PaE estimate obtained from a GLM, and appear generally reasonable.

It is important to note that, as pointed out in Section 4, IMSE is but one component of model error. The estimation of IMSE, PaE and process error are discussed in this paper, and illustrated numerically in Section 8, but this leaves model distribution error and external model structure error still to be brought to account.

The investigations of Taylor (2021) hint that the first of these would usually be small, but the second often would not. The only discussion of this component of forecast error in the actuarial literature occurs in O'Dowd et al. (2005) and Risk Margins Task Force (2008), where a subjective approach is taken.

This may well be the only option in making allowance for influences that are totally divorced from past experience and data. Be that as it may, the fact remains that any estimate of external model structure error must be derived by means quite different from those used for IMSE. If a subjective estimate is required, then, once obtained, it may be combined with the other components dealt with in this paper.

The limitations of this paper's methodology must also be considered. The major limitations appear to be twofold. First, as already mentioned just above, is the issue of subjective inclusion gates. The case for subjectivity has been presented here, but the discovery of any approach that could objectify them would be extremely useful.

Second, and to some extent related, the relatively large number of bootstrap replicates discarded for one reason or another may sometimes be a matter of concern. In Table 4, the rejection rate was broadly 10–20% in most cases, but was 60–80% for one of the more complex data sets (Data Set 3). There may be other devices that would generate a model set subject to lower rejection rates. Further research in this area might be useful.

Although external model structure error is beyond the scope of this paper, and possibly requires necessarily subjective estimation, it is worthwhile noting here that any research that would add objectivity would be valuable.

**Author Contributions:** Conceptualization, G.T.; methodology, G.T. and G.M.; software, G.M.; validation, G.T. and G.M.; formal analysis, G.T. and G.M.; investigation, G.T. and G.M.; resources, G.T. and G.M.; data curation, G.T. and G.M.; writing—original draft preparation, G.T.; writing—review and editing, G.T. and G.M.; visualization, G.M.; supervision, G.T.; project administration, G.T.; funding acquisition, G.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received financial supported under the Australian Research Council’s Linkage Projects funding scheme (project number LP130100723).

**Data Availability Statement:** The data sets used in this paper are synthetic. R code to generate them may be found at <https://institute-and-faculty-of-actuaries.github.io/mlr-blog/post/foundations/lasso/#generating-the-synthetic-data-set> (accessed on 23 October 2023). R code to run the analysis may be found at <https://grainnemcguire.github.io/post/2023-05-04-model-error-example/> (accessed on 23 October 2023).

**Acknowledgments:** The authors acknowledge valuable discussion with Gael Martin of Monash University, Benjamin Avanzi and David Yu, both of Melbourne University, and Bernard Wong of University of New South Wales. Finally, the authors are grateful for comments received from referees during the review process which led to improvements in the final version of the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Australian Prudential Regulatory Authority. 2018. Prudential Standard GPS 340: Insurance Liability Valuation. Available online: <https://www.legislation.gov.au/Details/F2018L00738> (accessed on 14 July 2022).
- Avanzi, Benjamin, Mark Lavender, Greg Taylor, and Bernard Wong. 2023. On the impact of outliers in loss reserving. *European Actuarial Journal* 1–40. [\[CrossRef\]](#)
- Bernardo, José M., and Adrian F. M. Smith. 1994. *Bayesian Theory*. New York: Wiley.
- Bignozzi, Valeria, and Andreas Tsanakas. 2016. Model Uncertainty in Risk Capital Measurement. *Journal of Risk* 18: 1–24. [\[CrossRef\]](#)
- Blanchet, Jose, Henry Lam, Qihe Tang, and Zhongyi Yuan. 2019. Robust actuarial risk analysis. *North American Actuarial Journal* 23: 33–63. [\[CrossRef\]](#)
- Chukhrova, Nataliya, and Arne Johannssen. 2021. Kalman Filter learning algorithms and state space representations for stochastic claims reserving. *Risks* 9: 112. [\[CrossRef\]](#)
- Clyde, Merlise, and Edward I. George. 2004. Model uncertainty. *Statistical Science* 19: 81–94. [\[CrossRef\]](#)
- Clyde, Merlise, and Edwin S. Iversen. 2013. Bayesian model averaging in the M-open framework. *Bayesian Theory and Applications* 14: 483–98.
- Dahms, René. 2018. Chain-ladder method and midyear loss reserving. *ASTIN Bulletin* 48: 3–24. [\[CrossRef\]](#)
- De Jong, Piet, and Ben Zehnwirth. 1983. Claims reserving, state-space models and the Kalman filter. *Journal of the Institute of Actuaries* 110: 157–81. [\[CrossRef\]](#)
- England, Peter D., and Richard J. Verrall. 1999. Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance: Mathematics and Economics* 25: 281–93. [\[CrossRef\]](#)
- England, Peter D., and Richard J. Verrall. 2001. A flexible framework for stochastic claims reserving. *Proceedings of the Casualty Actuarial Society* 88: 1–38.
- England, Peter D., and Richard J. Verrall. 2002. Stochastic claims reserving in general insurance. *British Actuarial Journal* 8: 443–544. [\[CrossRef\]](#)
- Glasserman, Paul, and Xingbo Xu. 2014. Robust risk measurement and model risk. *Quantitative Finance* 14: 29–58. [\[CrossRef\]](#)
- Hansen, Peter R., Asger Lunde, and James M. Nason. 2011. The model confidence set. *Econometrica* 79: 453–97. [\[CrossRef\]](#)
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *Elements of Statistical Learning: Data Mining, Inference and Prediction*. New York: Springer.
- Hoeting, Jennifer A., David Madigan, Adrian E. Raftery, and Chris T. Volinsky. 1999. Bayesian model averaging: A tutorial. *Statistical Science* 14: 382–401.

- Huang, Ziyi, Henry Lam, and Haofeng Zhang. 2021. Quantifying epistemic uncertainty in deep learning. *arXiv* arXiv:2110.12122.
- Loaiza-Maya, Ruben, Gael M. Martin, and David T. Frazier. 2021. Focused Bayesian prediction. *Journal of Applied Econometrics* 36: 517–43. [CrossRef]
- Martin, Gael M., Ruben Loaiza-Maya, Worapree Maneesoonthorn, David T. Frazier, and Andrés Ramírez-Hassan. 2022. Optimal probabilistic forecasts: When do they work? *International Journal of Forecasting* 38: 384–406. [CrossRef]
- McGuire, Gráinne, Greg Taylor, and Hugh Miller. 2021. Self-assembling insurance claim models using regularized regression and machine learning. *Variance* 14. Available online: <https://variancejournal.org/article/28366-self-assembling-insurance-claim-models-using-regularized-regression-and-machine-learning> (accessed on 23 October 2023). [CrossRef]
- O'Dowd, Conor, Andrew Smith, and Peter Hardy. 2005. A framework for estimating uncertainty in insurance claims cost. Paper presented at the Institute of Actuaries of Australia XVth General Insurance Seminar, Port Douglas, Australia, October 16–19.
- Raftery, Adrian E. 1995. Bayesian model selection in social research. *Sociological Methodology* 25: 111–63. [CrossRef]
- Raftery, Adrian E. 1996. Approximate Bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika* 83: 251–66. [CrossRef]
- Raftery, Adrian E., David Madigan, and Jennifer A. Hoeting. 1997. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* 92: 179–91. [CrossRef]
- Reid, D. H. 1978. Claim Reserves in General Insurance. *Journal of the Institute of Actuaries* 105: 211–96. [CrossRef]
- Risk Margins Task Force. 2008. A Framework for Assessing Risk. Paper presented at the Institute of Actuaries of Australia 16th General Insurance Seminar, Coolum, Australia, November 9–12.
- Schneider, Judith C., and Nikolaus Schweizer. 2015. Robust measurement of (heavy-tailed) risks: Theory and implementation. *Journal of Economic Dynamics and Control* 61: 183–203. [CrossRef]
- Shi, Peng. 2013. A copula regression for modeling multivariate loss triangles and quantifying reserving variability. *ASTIN Bulletin* 44: 85–102. [CrossRef]
- Shibata, Ritei. 1997. Bootstrap Estimate of Kullback-Leibler Information for Model Selection. *Statistica Sinica* 7: 375–94.
- Taylor, Greg. 1985. Combination of estimates of outstanding claims in non-life insurance. *Insurance: Mathematics and Economics* 4: 321–438. [CrossRef]
- Taylor, Greg. 1988. Regression models in claims analysis I: Theory. *Proceedings of the Casualty Actuarial Society* 74: 354–83.
- Taylor, Greg. 2000. *Loss Reserving: An Actuarial Perspective*. Dordrecht: Kluwer.
- Taylor, Greg. 2021. A special Tweedie sub-family with application to loss reserving prediction error. *Insurance: Mathematics and Economics* 101: 262–88. [CrossRef]
- Taylor, Greg, and Frank R. Ashe. 1983. Second moments of estimates of outstanding claims. *Journal of Econometrics* 23: 37–61. [CrossRef]
- Taylor, Greg, and Gráinne McGuire. 2016. *Stochastic Loss Reserving Using Generalized Linear Models*. CAS Monograph Series; Arlington: Casualty Actuarial Society.
- Zhang, Yanwei, and Vanja Dukic. 2013. Predicting multivariate insurance loss payments under the Bayesian copula framework. *Journal of Risk and Insurance* 80: 891–919. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.