

Merging Expert and Empirical Data for Rare Event Frequency Estimation: Pool Homogenisation for Empirical Bayes Models

John Quigley*, Gavin Hardman, Tim Bedford, Lesley Walls
University of Strathclyde, Department of Management Science, Graham Hills Building, 40 George Street,
Glasgow, G1 1QE, Scotland

Abstract

Empirical Bayes provides one approach to estimating the frequency of rare events as a weighted average of the frequencies of an event and a pool of events. The pool will draw upon, for example, events with similar precursors. The higher the degree of homogeneity of the pool, then the Empirical Bayes estimator will be more accurate. We propose and evaluate a new method using homogenisation factors under the assumption that events are generated from a Homogeneous Poisson Process. The homogenisation factors are scaling constants which can be elicited through structured expert judgement and used to align the frequencies of different events, hence homogenising the pool. The estimation error relative to the homogeneity of the pool is examined theoretically indicating that reduced error is associated with larger pool homogeneity. The effects of misspecified expert assessments of the homogenisation factors are examined theoretically and through simulation experiments. Our results show that the proposed Empirical Bayes method using homogenisation factors is robust under different degrees of misspecification.

Key words: Empirical Bayes; pairwise comparison; PRA; poisson processes

* Corresponding Author

Email addresses: j.quigley@strath.ac.uk (John Quigley); gavinhardman@btinternet.com (Gavin Hardman); lesley.walls@strath.ac.uk (Lesley Walls); tim.bedford@strath.ac.uk (Tim Bedford).

Submitted to Reliability Engineering and System Safety December, 2010

Nomenclature

m	Number of processes in the pool
Λ_i	Rate of occurrence of process i modelled as a random variable described by the prior distribution
λ_i	Realisation of the rate of occurrence of process i
α	Shape parameter of prior distribution when not explicitly modelling heterogeneity
β	Scale parameter of prior when not explicitly modelling heterogeneity
a	Shape parameter of prior distribution when explicitly modelling heterogeneity
b	Scale parameter of prior distribution when explicitly modelling heterogeneity
\hat{a}	Estimator of a
\hat{b}	Estimator of b
n_i	Number of events in process i
\underline{n}	Vector of number of events each process
t	Exposure time for each process (i.e. common for all processes)
h_i	Homogenisation factor for process i
\bar{h}	Mean of homogenisation factors
\tilde{h}_i	Subjective expert assessment of homogenisation factor for process i
$\overline{h^2}$	Mean of squared homogenisation factor
r	Ratio of $\bar{h}/\overline{h^2}$
U	Adjusted rate of occurrence of events over all processes
W	Information about the second moment within the pool of all processes
S_i	Multiplicative random error in subjective assessment of homogenisation factor h_i
θ	Mean of multiplicative error in subjective assessment of homogenisation factor h_i
σ^2	Variance of multiplicative error in subjective assessment of homogenisation factor h_i
μ_i	Expected rate of occurrence of process i , given parameter values of prior distribution are known
$\mu_{h,i}$	Expected rate of occurrence of process i , given homogenisation factors are

	used
$\hat{\mu}_i$	Estimator of expected rate of occurrence of process i when parameter values of prior distribution are known and so homogenisation factor not used
$\hat{\mu}_{h,i}$	Estimator of expected rate of occurrence of process type i when homogenisation factors are used
k_i	Logarithm (base 10) of the homogenisation factor for process i
\tilde{k}_i	Logarithm (base 10) of the subjective expert assessment of homogenisation factor for process i
e_i	Arithmetic error in estimating the rate of occurrence of process i

1 Introduction

The estimation of the frequency of rare events is common in Probabilistic Risk Assessment (PRA). Indeed our motivation for this work has been driven by PRA projects in, for example, explosive storage, railway system and power plants (Quigley et al, 2007, Hutchison et al, 2008). Consider the explosives storage application where we might anticipate a small number of observed events spread over a larger number of incident categories and explosive types. It is unlikely that we will regularly observe data for every combination of incident and explosive type. For example, (Merrifield and Moreton 1998) reported 79 incidents on the UK mainland between 1950 and 1997. Of these 79, only 16 occurred in storage, a rate of one every 3 years (and none of these major incidents).

A variety of inference approaches can be used to estimate the frequency of rare events, including classical statistical methods, direct subjective expert judgement, and both fully Bayesian and Empirical Bayes methods. Each has its strengths and limitations. For example, the classical approach is to estimate the rate as a ratio of the number of observed events to the exposure time. For situations where no events are observed, a variety of ways are proposed to adjusting the otherwise overoptimistic point estimate of zero. These include: substituting the zero estimate of the rate by the value of the reciprocal of the exposure time (Bailey 1997); using the upper bounds of Chi-Squared (Edmonson 1992) or Normal (Williams and Thorne 1997) confidence intervals for the rate; or minimising the maximum expected squared error (Quigley and Revie 2011). See Quigley and Revie

(2011), Bailey (1997) and Williams and Thorne (1997) for further review and comparison of alternative classical approaches for this problem. Generally we believe that classical statistical estimators for rare event frequencies may not be very useful because the uncertainty associated with the estimates obtained will be relatively large.

Bayesian methods are widely used in PRA (Kelly and Smith 2009) and allow the empirical data to be balanced against the subjective beliefs of experts. Practical examples of the use of full Bayes methods are given by (Parent and Bernier 2003, (Bunea et al.2005) and (Meel et al. 2007). A Bayesian paired comparison method for estimating rare event probabilities is considered by (Szwed et al. 2006). Parent and Bernier (2003) discuss common objections to the Bayesian approach. In our context, fully Bayesian methods require the specification of a subjective prior distribution, which may not only be difficult to achieve but which will largely determine the outcome of the estimation process because of the relatively small amount of data available with which to update the prior. Hence the estimates will be highly influenced by the initial subjective engineering judgement. This limitation is also relevant to the direct use of expert judgement to obtain a point estimate (Bedford et al 2008).

In (Quigley et al, 2007) we have previously explored the use of Empirical Bayes, a hybrid of classical and Bayesian methods in which the prior is found using empirical methods. The Empirical Bayes method allows the pooling of observed data across multiple events to estimate an overall rate. Individual occurrence rates are then calculated as deviations from this overall rate. See (Carlin and Louis 2000) for a general overview of Empirical Bayes. Unlike Bayesian methods, Empirical Bayes is not a fully subjective approach, because it uses the pooled data to estimate the prior distribution parameters. It can be argued that Empirical Bayes possesses some of the benefits of Bayesian methods while avoiding the need for subjective specification of a prior distribution. By relying on empirical data to form the prior, we can retain some of the benefits of using historical data when estimating rare event frequencies as discussed by (Gelman et al. 1998).

Empirical Bayes methods have been applied in the fields of reliability (Sarhan 2003) and risk analysis (Martz et al. 1999) and are regularly used to analyse accident occurrence patterns in road safety applications (Persaud and Lyon 2007). Empirical Bayes models have been shown to perform

well against full Bayes models (Camara and Tsokos 1999), and even favourably when there are few observed data (Vaurio and Jankala 2006, Srivastava and Kubokawa 2007). However, the Empirical Bayes method also has limitations. For example, it has not been established how to formally incorporate quantitative expert judgement and the accuracy of the estimates obtained depend on the degree of homogeneity of the pool of events used to form the prior; for example, the more homogeneous the pool of events, then we might expect the estimate obtained to be more accurate.

In this paper we discuss a novel integration of quantitative expert judgement into Empirical Bayes to homogenise the pool of events used to form the prior and so meet the goal of obtaining more accurate estimates of the rate of occurrence of rare events. Our method does not require an expert to assess the absolute values of frequencies, but merely to assess relative rates. Hence the method could be operationalised by using methods such as pairwise comparison (Park and Lee 2008, Goosens et al 2008) that naturally give ratios rather than absolute values.

Our proposed approach aims to use expert information to rescale data by effectively choosing a natural time scale for each event type in order to improve the behaviour of the Empirical Bayes estimator. Expert judgement is introduced to assess so-called homogenisation factors, which are scaling constants to bring the frequencies of different event types to approximately the same value. By using such homogenisation factors the estimates produced by pooling different event types using Empirical Bayes should become more accurate, in the sense that a more homogenous pool will attach more weight to the pooled average rather than the individual event experience. A reduction in error will be a consequence of greater reliance on a representative estimate derived from a larger sample size.

In this paper we develop the proposed method under the assumption that the events are generated from a Homogeneous Poisson Process (HPP), which is a not unreasonable model for the case where the rate of events can be treated as constant. Although the homogenisation factors are unknown constants, we consider the epistemic uncertainty in assessing them as random variables. We examine the estimation error relative to the homogeneity of the pool to provide an assessment of the accuracy of our proposed estimators. Given that we advocate that the homogenisation factors are obtained using expert judgement, we evaluate the impact of poor subjective assessments on the

robustness of our estimates. Therefore, in summary, this paper contributes a new approach to Empirical Bayes estimation for rare event frequencies and examines the properties of the proposed method both theoretically and through simulation experiments.

The modelling framework adopted for this problem and the development of the new Empirical Bayes estimator are described in detail in Section 2. Section 3 reports a theoretical investigation of the impact of misspecification of homogenisation factors on inference when the Method of Moments is used for parameter estimation. Section 4 presents the design and results of a simulation study to explore the key aspects of the model set-up under Maximum Likelihood and controlled pool sizes. Section 5 presents concluding remarks and discusses further work.

2 Model Formulation and Inference

We consider a set of processes, each generating data according to a Homogeneous Poisson Process (HPP) but not necessarily at the same rate of occurrence. Using an Empirical Bayes approach for inference we seek a model to describe the variability of the rate of occurrences within the pool of processes, which will be referred to as the prior distribution and whose parameter values will be estimated empirically. While Bayesian priors are usually constructed from subjective beliefs about the value of an event probability (Bedford et al 2008), Empirical Bayes (Carlin and Louis 2000) provides a means of pooling observed data to form an empirical prior.

A common and convenient parametric form of the prior associated with the HPP is the Gamma distribution because it is a conjugate prior. We denote the rate of occurrence of events from process i by Λ_i , which we assume has been realised from a Gamma distribution, and denote possible realisations as λ_i .

If prior to observing any data we have no expert judgement that would suggest a ranking of the rates of occurrence of events then we may be content to assign them with the same prior distribution, which we express in (1):

$$\pi(\lambda_i) = \frac{\beta^\alpha \lambda_i^{\alpha-1} e^{-\beta\lambda_i}}{\Gamma(\alpha)}, \alpha > 0, \beta > 0, i = 1, 2, \dots, m \quad (1)$$

Applying (1) as the prior for the pool would result in the Negative Binomial expressed in (2) as the predictive distribution representing the number of events that will be realised in an interval of time of length t for process i :

$$\begin{aligned} P(N_i = n_i) &= \int_0^\infty P(N_i = n_i | \lambda_i) \pi(\lambda_i) d\lambda_i \\ &= \frac{\Gamma(\alpha + n_i)}{\Gamma(\alpha) n_i!} \left(\frac{\beta}{\beta + t} \right)^\alpha \left(\frac{t}{\beta + t} \right)^{n_i}, t > 0, n_i = 0, 1, \dots \end{aligned} \quad (2)$$

The predictive distribution can be used to estimate the parameter values of the prior as they describe the variability of observation within the pool and are a function of the unknown parameters (i.e. α and β). Common approaches for inference in such cases are Maximum Likelihood Estimates or Moment Matching.

We presume experts can distinguish rates of occurrence between events in the pool to within orders of magnitude, and propose a more appropriate prior distribution expressed in (3), where h_i represents a homogenisation factor to re-scale the rate of occurrence to a common base:

$$\pi(\lambda_i | h_i) = \frac{\left(\frac{b}{h_i} \right)^a \lambda_i^{a-1} e^{-\frac{b}{h_i} \lambda_i}}{\Gamma(a)}, a > 0, b > 0, h_i > 0, i = 1, 2, \dots, m \quad (3)$$

Accounting for variation as in (3) would result in a predictive distribution expressed in (4):

$$P(N_i = n_i | h_i) = \frac{\Gamma(a + n_i)}{\Gamma(a) n_i!} \left(\frac{b}{h_i t + b} \right)^a \left(\frac{h_i t}{h_i t + b} \right)^{n_i}, t > 0, i = 1, 2, \dots, m, n_i = 0, 1, \dots \quad (4)$$

It is worth noting that the same model can be obtained by using the prior as given in (1) but using a Poisson Likelihood as given in (5). In short, we are re-scaling exposure to homogenise:

$$P(N_i = n_i | \lambda_i, h_i) = \frac{(\lambda_i t h_i)^{n_i} e^{-\lambda_i t h_i}}{n_i!}, t > 0, \lambda_i > 0, h_i > 0, i = 1, 2, \dots, m, n_i = 0, 1, \dots \quad (5)$$

Of course (3) and (1) are not consistent with each other in the sense that if (3) represents the ‘true’ variability of the rates of occurrence when we formally account for the homogenisation factor, then the distribution that would account for the variability in the rates of occurrence when homogeneity is not explicit would be one where a mixing distribution is used to average (3) over the variability of the homogeneity factors. Such a mixing would remove the prior from the family of Gamma distributions. However, from a parsimony perspective we could approximate the true distribution with a Gamma distribution as in (1).

2.1 *Effects on Estimation By Formally Accounting for Heterogeneity*

If we compare the relationship between the parameters of the prior in (1) with (3) when the first two moments of the pool are identical then we obtain the following relationship.

Result 1: There is a linear relationship between the pool parameters without homogenisation and those with homogenisation.

Consider using a Gamma distribution as in (1), rather than using a Gamma distribution as in (3), to describe the variability of the rate of occurrence within a pool. If both models have identical first two moments, then a linear relationship exists between the parameters as in the following:

$$\alpha = ar\bar{h} \leq a$$

$$\beta = br \leq \frac{b}{\sqrt{\bar{h}^2}}$$

where:

m the number of processes in the pool

$$r = \frac{\bar{h}}{h^2}$$

$$\bar{h} = \frac{\sum_{i=1}^m h_i}{m}$$

$$\bar{h}^2 = \frac{\sum_{i=1}^m h_i^2}{m}$$

As such, if the homogenising factors were chosen such that $\bar{h} = 1$ then:

$$\alpha = \frac{a}{h^2}$$

$$\beta = \frac{b}{h^2} \tag{6}$$

Proof: See Appendix

As a consequence of *Result 1*, the choice of formally modelling the heterogeneity will impact on the point estimator obtained through an Empirical Bayes process. Expressed in (7) are the point estimators for the rate of occurrence of an event i assuming the parameters of the prior distribution that are known to have experienced n_i events in a time period t with heterogeneity accounted (i.e. using prior (4)) denoted by $\mu_{h,i}$ and without (i.e. using prior (1)) denoted by μ_i .

$$\mu_{h,i} = \frac{a+n_i}{b+th_i} = \frac{a}{b} \left(\frac{b}{b+th_i} \right) + \frac{n_i}{th_i} \left(1 - \frac{b}{b+th_i} \right), a, b, t, h_i > 0, i = 1, 2, \dots, m, n_i = 0, 1, 2, \dots \quad (7)$$

Assuming (6) holds then we have the following:

$$\mu_i = \frac{a+n_i \bar{h}^2}{b+t \bar{h}^2} = \frac{a}{b} \left(\frac{b}{b+t \bar{h}^2} \right) + \frac{n_i}{t} \left(1 - \frac{b}{b+t \bar{h}^2} \right) \quad (8)$$

Upon inspection of (7) and (8) we see that by ignoring the heterogeneity we have less weight on the pooled mean, (i.e. a/b), if $h_i > \bar{h}^2$ but more on the individual experience (i.e. n_i/t , recalling that n_i is the number of events realised and t is the exposure time) or the individual experience adjusted for heterogeneity (i.e. n_i/th_i). Assuming $\bar{h} = 1$ then $\bar{h}^2 \geq 1$ so this condition will only be realised for very large homogenisation factors, which would correspond to processes with very high rates of occurrence relative to the pool.

2.2 Summary

So far we have been concerned with developing an Empirical Bayes inference procedure that explicitly accounts for heterogeneity within the pool and then comparing this new inference method with Bayesian updating of the empirical prior only as this implicitly takes account of heterogeneity. We establish a relationship between the pool parameters and show that, except for processes with high rates of occurrence, more weight would be applied to the pooled mean when heterogeneity is explicitly modelled. This will result in a smaller standard error, because the sample size used to estimate to pooled average is larger than that for an individual process.

3 Theoretical Examination of Misspecification of Homogenisation Factor

We use the Method of Moments approach to parameter estimation to investigate theoretically the impact of misspecification of the homogenisation factors on inference. As such we use (9) and (10) as estimation equations for pool parameters a and b respectively:

$$\hat{a} = \frac{U^2}{W - U^2} \quad (9)$$

$$\hat{b} = \frac{U}{W - U^2} \quad (10)$$

where:

$$U = \frac{\sum_{i=1}^m n_i}{\sum_{i=1}^m \tilde{h}_i t}$$

$$W = \frac{\sum_{i=1}^m n_i^2 - \sum_{i=1}^m n_i}{\sum_{i=1}^m (\tilde{h}_i t)^2}$$

U represents the overall adjusted rate of occurrence of events; n_i corresponds to the number of events realised by process i during an exposure time of duration t ; \tilde{h}_i is the homogenisation factor assessed by subjective expert judgment; and W captures information about the second moment within the pool of events. In order to use the moment matching method under the assumption of a Negative Binomial distribution, the sample variance must exceed the sample mean since this is a property of this distribution.

We seek to investigate the impact of misspecification of the prior by assuming a ‘true’ homogenisation factor, which is denoted by h_i for process i . We further assume that experts provide an assessment of the homogenisation factor subject to a random multiplicative error; that is, $\tilde{h}_i = h_i S_i$ where S_i represents the error in factor for the i -th process. The homogenisation factors (i.e. h_i) are unknown constants which are assessed by the expert (i.e. \tilde{h}_i) and so modelled as a random variable to describe the epistemic uncertainty. We denote the mean and variance as in (11) and assume independence between expert assessments:

$$\begin{aligned} E[S_i] &= \theta \\ \text{Var}[S_i] &= \sigma^2 \\ \text{Cov}(S_i, S_j) &= 0, i \neq j \end{aligned} \tag{11}$$

3.1 Key Findings

The following three results summarise our findings with respect to inference in the presence of misspecification of the homogenisation factors. The proofs for each result are given in the Appendix.

Result 2: As pool size increases towards infinity misspecification of the homogenisation factor has the following effect on the parameter estimators:

$$\begin{aligned} \lim_{m \rightarrow \infty} \hat{a} &\xrightarrow{P} a \frac{\sigma^2 + \theta^2}{\theta^2 - a\sigma^2} \geq a \\ \lim_{m \rightarrow \infty} \hat{b} &\xrightarrow{P} b\theta \frac{\sigma^2 + \theta^2}{\theta^2 - a\sigma^2} \geq b\theta \end{aligned}$$

Result 3: The empirical prior estimator of the aggregate rate of occurrence of events in the pool does not depend on the degree of misspecification of the homogenisation factor as the pool size tends to infinity:

$$\lim_{m \rightarrow \infty} \sum_{i=1}^m \frac{E[\Lambda_i]}{m} \xrightarrow{P} \frac{a}{b}$$

Result 4: The empirical posterior estimator of the rate of occurrence of events for a process i has the following limit:

$$\lim_{m \rightarrow \infty} E[\Lambda_i | \underline{n}] \xrightarrow{P} \frac{a \tilde{h}_i}{b \theta} \left(1 - \frac{t}{b \frac{\theta}{\tilde{h}_i} \left(\frac{\sigma^2 + \theta^2}{\theta^2 - a\sigma^2} \right) + t} \right) + \frac{n_i}{t} \left(\frac{t}{b \frac{\theta}{\tilde{h}_i} \left(\frac{\sigma^2 + \theta^2}{\theta^2 - a\sigma^2} \right) + t} \right)$$

where \underline{n} is the vector of number of occurrences for each process.

As a consequence of *Result 4* it is evident that as the variability of misspecification decreases then the estimator converges to the point estimator which is obtained for the exact homogeneity factor. That is, the mean bias has no influence:

$$\lim_{\substack{m \rightarrow \infty \\ \sigma^2 \rightarrow 0}} E[\Lambda_i | \underline{n}] \xrightarrow{P} \frac{a}{b} h_i \left(1 - \frac{th_i}{b + th_i} \right) + \frac{n_i}{t} \left(\frac{th_i}{b + th_i} \right)$$

This is an interesting result because we have presumed in the derivations that the scaling subjective assessments were independent. However this result shows that if assessments are

perfectly correlated (i.e. the multiplicative error associated with each assessment is identical so that the variance is 0) then for pools with a large number of events such bias has no influence.

3.2 *Summary*

In this Section we have investigated the effects of the misspecification of the homogenisation factors on the estimates of the pool characteristics (i.e. the point estimates of the pool parameters) as well as on individual events. We show that the effects on pool characteristics would be insignificant for large pool sizes. This is due to the reduction in uncertainty associated with the misspecifications and not because the prior has less influence since we have not considered the asymptotics associated with the exposure period. As the number of events increases towards infinity then we become more confident that we have sampled the entire population of possible misspecifications. For individual processes, the point estimator will be sensitive to misspecification and we have derived an expression to explicate this effect.

4 Simulation Investigation of Misspecification of the Homogenisation Factors for MLE under Controlled Pool Sizes

We now investigate the impact of misspecification when Maximum Likelihood Estimators (MLE), rather than Moment Matching methods, are used to estimate the pool parameters. While MLE's have many desirable inferential properties, for this model no closed form solutions are available for the MLE's, which has motivated the earlier theoretical work towards Moment Matching methods. It is worth noting that searching for such MLE can be computationally demanding and convergence can be an issue. Instead we have designed and conducted a simulation study. The simulation allows us to assess the impact of different types of expert misspecification for a selection of different pool sizes. We consider five different types of expert behaviour as summarised in Table 1.

INSERT TABLE 1 ABOUT HERE

The precise numerical specifications for expert types ET1 - ET5 are given in Table 2. These expert types have been chosen to represent plausible expert behaviours and provide baseline scenarios for comparison. The expert cannot perform better than correctly specifying all homogenisation factors (as in ET1) and ET5 provides information on minimum performance level that we can expect from an expert (a guide to when an expert type can be described as ‘no better than guessing’). Each ET is reviewed for three pool sizes to assess if the type of misspecification has different effects according to pool size. We also consider the effect of changing the base occurrence rate to establish if the observed errors under ET1-5 vary as event frequency increases.

INSERT TABLE 2 ABOUT HERE

4.1 Simulation Design

For the model described in Section 2, we assume that each event can be modelled as a homogeneous Poisson process as in (5). Therefore, the number of events that we observe in exposure period t is Poisson with mean $\lambda_i h_i t$, where the homogenisation factor can be seen to be playing a similar role to the exposure t .

We are interested in the effects of misspecification of h_i , so to focus simulation results, we constrain $b=1$ and assume $t=1$ for all i . Therefore, the pool mean will be equal to a , the weight for the observed frequency depends only on h_i , and the point estimator of the rate of occurrence of a process becomes (12):

$$\hat{\mu}_{h,i} = \hat{a} \frac{1}{1+h_i} + \frac{n_i}{h_i} \frac{h_i}{1+h_i} \quad (12)$$

The set of potential h_i values is defined to be $\{1, 10, 100, 1000\}$, and the ‘true’ pool is constructed from 4 sub-pools as follows: for initiators $1, \dots, 2\ell$, $h_i = 1$; for initiators $2\ell + 1, \dots, 4$

ℓ , $h_i = 10$; for initiators $4\ell + 1, \dots, 6\ell$, $h_i = 100$; for initiators $6\ell + 1, \dots, 8\ell$, $h_i = 1000$. There are 2ℓ initiators in each subgroup to allow the implementation of ET4 (partial accuracy). Under ET4, for each h_i value, ℓ initiators are defined by $\tilde{h}_i = h_i$, and for the other ℓ , $\tilde{h}_i \neq h_i$. The objective of ET4 is to test if the effect of the misspecification is mitigated by the accuracy of specification elsewhere. The splitting of each sub-pool means the precise numeric specification of each expert type (as shown in Table 2) requires us to specify \tilde{h}_i values for 8 groups, one for every ℓ initiators, $\ell = \{1, 6, 12\}$, giving pool sizes of 8 (small), 48 (medium) and 96 (large). Data are randomly generated from a Negative Binomial distribution parameterised by $(a, 1)$, where a takes values of 0.1 and 10 to represent situations in which realisations are either less or more frequent. Having generated the data, we then estimate \hat{a} based on the data and the \tilde{h}_i values using numerical Maximum Likelihood methods. As measures of the impact of misspecifying homogenisation factors, we consider the behaviour of the error and squared error. We know that for the simulation parameterisation, the correct point estimator for the occurrence rate is:

$$\mu_i = \frac{a + n_i}{1 + h_i}$$

and that the point estimator of the occurrence rate based on expert assessment of the \tilde{h}_i values is:

$$\hat{\mu}_{h,i} = \frac{\hat{a} + n_i}{1 + \hat{h}_i}$$

so the quantities we are interested in are:

$$e_i = \hat{\mu}_{h,i} - \mu_i = \frac{\hat{a} + n_i}{1 + \hat{h}_i} - \frac{a + n_i}{1 + h_i} \quad (13)$$

and the square of the relative error, i.e. $(e_i/\mu_i)^2$.

4.2 *Simulation algorithm*

The simulation algorithm can be summarised by the following steps:

1. Define quantities under test for a, h_i, \hat{h}_i ;
2. Generate data from $\text{NBIN}(a, 1)$;
3. Estimate \hat{a} from ℓ using \tilde{h}_i ;
4. Calculate e_i and $(e_i/\mu_i)^2$;
5. Repeat for specified number of realisations.

4.3 *Simulation Results*

Simulations are organised as 30 separate blocks. Each corresponds to a different combination of expert type, a value and pool size as described previously. We consider the absolute error and percentage mean squared error (PMSE) in estimating the ‘true’ occurrence rates as measures of expert accuracy. ET1 can be used as a way of calibrating the size of sampling error that we can expect in our simulations. The accuracy of assessment means there are no deviations from the ‘true’ occurrence rates except those introduced by sampling variation. Figure 1 shows that the accuracy increases as both the pool size and the value of h_i increase. Increasing pool size increases the number of realisations and increasing h_i increases the value of the denominator in (13), so both observed effects should be expected.

INSERT FIGURE 1 ABOUT HERE

Figure 2 shows the changes in the distribution of the error, e_i as pool size increases for ET1, ET2 and ET3. Each expert's error distribution has a distinctive shape: for total accuracy, there is a narrow spread either side of zero; for overestimation there is a heavy positive skew; for underestimation there is a negative skew, although the range is greater than for overestimation. In all cases the majority of e_i values are concentrated at 0. However, the shapes of the distribution remain reasonably unaffected by increased pool size.

INSERT FIGURE 2 ABOUT HERE

Figure 3 shows the changes in variation of the PMSE values as the difference between h_i and \tilde{h}_i varies. All h_i (or \tilde{h}_i) can be written as 10^{k_i} (or $10^{\tilde{k}_i}$) where $k = 0, \dots, 3$. The values on the x-axis of Figure 3 are defined as $k_i - \tilde{k}_i$ for each pair of values as defined in the simulation. There is a lack of data for values outside $\{-1, 0, 1\}$, but the chart appears to support the idea that errors for underestimation (< 0) and overestimation (> 0) are distributed differently.

INSERT FIGURE 3 ABOUT HERE

Figures 4, 5 and 6 allow us to compare the performance of expert types ET1-ET5. These show that ET3 (underestimation of risk) consistently performs poorly, realising higher error values than the other expert types with the exception of some ET5 values. ET2 (overestimation), consistently performs well in relation to ET1, indicating that under this formulation, excessive caution has less impact on estimates than undue optimism. Partial accuracy (ET4) and random assignment (ET5) perform similarly. Both appear less effective as pool size is increased (contrast the changing pattern between Figures 5 and 6 for these expert types against the relative consistency of the others). For both ET4 and ET5, there is a large discrepancy in performance between the accurately and inaccurately specified h_i values - correctly estimated terms performing in line with ET1 and

incorrectly estimated yielding large errors. Figures 4, 5 and 6 also clearly show the heteroscedasticity of the error values. The largest errors are associated with the larger h_i values. Consequently, the experts that are inaccurate at this end of the scale appear to perform worse.

INSERT FIGURES 4, 5 AND 6 ABOUT HERE

5 Conclusions and Future Work

This paper has considered the problem of frequency estimation for rare events, which is an intrinsically difficult problem requiring a variety of approaches. We have considered the role of Empirical Bayes inference which is based on pooling data arising from event types with different underlying generation processes and characterising the variability between processes with an empirical prior distribution, which is updated using Bayes' Theorem to obtain a posterior for each process. This results in an estimate for each process that is influenced both by its own data and the experience of the pool.

We have introduced a novel method of using expert assessment to rescale each process and thus homogenise the data pool. We have derived theoretical results for inference based on Moment Matching that show the error in estimation will decrease as the homogeneity of the pool increases. By selectively rescaling the Poisson processes using expert judgement inputs, it is possible to make the data more homogeneous. This novel way of using expert input therefore reduces uncertainties.

We have investigated the possible effect of misspecified expert judgements; both theoretically under the Method of Moments and through a simulation study under Maximum Likelihood estimation. The main conclusion of our theoretical study is that the proposed Empirical Bayes approach is robust under different degrees of misspecification. The simulation results show that - under our chosen model - pessimistic homogenisation, in which the expert assumes occurrence rates to be higher than they actually are, performs better than the expert underestimating occurrence rates,

or those in which the expert is partially correct. Related to this is the result that misspecification relating to large homogenisation factors yields larger errors.

Our study has been confined to the case where events are generated from HPP, and it would be interested to examine the properties of our proposed approach under alternative event generation processes. Further the simulation study has been restricted to consideration of 5 expert types - there are obviously many more and investigating these should allow us to improve understanding of the effects of expert optimism/pessimism in homogenisation problems. Further investigation of the trends seen in this simulation study is required, but a detailed understanding of the consequences of misspecification of homogenisation factors for this model would allow the development of procedures for data homogenisation, and the incorporation of a greater range of evidence into occurrence rate estimates for rare events.

Acknowledgements

The authors acknowledge the support and funding of the UK MOD DOSG in the preparation of work leading to this paper.

References

- [1] Bailey, RT. Estimation from zero failure data. *Risk Analysis*. 1997; 17: 375–380.
- [2] Bedford T, Quigley J and Walls L. Expert Elicitation for Reliable System Design (with discussion). *Statistical Science*. 2007; 42: 428-462.
- [3] Bienz A. The new Swiss model for the probability of an explosion in an ammunition storage. Technical report. NATO/Bienz, Kummer & Partner Ltd. 2007.
- [4] Bunea C, Charitos T, Cooke RM, and Becker G. Two stage Bayesian models application to ZEDB project. *Reliability Engineering and System Safety*. 2005; 90: 123–130.

- [5] Camara VAR and Tsokos CP. The effect of loss functions on empirical Bayes reliability analysis. *Mathematical Problems in Engineering*. 1999; 4: 539–560.
- [6] Carlin B and Louis T. *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd Edition, Chapman and Hall/CRC; 2000.
- [7] Edmonson JN. Consolidated proposals for determining basic level 1 accident initiation frequencies for storage and process buildings using the results from the scaling analysis studies. Technical report, MOD; 1992.
- [8] Elvik R. The predictive validity of empirical Bayes estimates of road safety. *Accident Analysis and Prevention*. 2008; 40: 1964–1969.
- [9] Gelman A, King G, and Boscardin WJ. Estimating the probability of events that have never occurred: When is your vote decisive? *Journal of the American Statistical Association*. 1998; 91: 1–9.
- [10] Goossens LHJ, Cooke RM, Hale AR, Rodic-Wiersma Lj. Fifteen years of expert judgement at TU Delft. *Safety Science*. 2008; 46(2). Occupational Accident Scenarios, and Accident Analysis Papers selected from the third international conference Working on Safety (WOS2006), September 12-15th, 2006, Zeewolde, The Netherland, Pages 234-244.
- [11] Greenwood M and Yule GU. An inquiry into the nature of frequency distributions representative of multiple happening with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal Statistical Society A*. 1920; 83: 255–279.
- [12] Hutchison K, Quigley J, Raza M and Walls L. Empirical Bayes methodology for estimating equipment failure rates with application to power generation plants. *Proc. IE Engineering Management Conference*; 2008.

- [13] Johnson N, Kotz S and Kemp A. Univariate discrete distributions (2nd ed.). New York: Wiley; 1993.
- [14] Kahneman D, Slovic P and Tversky A (Eds.). Judgment and uncertainty: Heuristics and biases. Cambridge University Press; 1982.
- [15] Kelly DL and Smith CL. Bayesian inference for probabilistic risk assessment – The current state of the art. *Reliability Engineering and System Safety*. 2009; 94: 628–643.
- [16] Martz HF, Parker RL, and Rasmuson DM. Estimation of trends in the scram rate at nuclear power plants. *Technometrics*. 1999; 41: 352–364.
- [17] Meel A, O’Neill LM, Levin JH, Seider WD, Oktem U, and Keren N. Operational risk assessment of chemical industries by exploiting accident databases. *Journal of Loss Prevention in the Process Industries*. 2007; 20: 113–127.
- [18] Merrifield R and Moreton P. An examination of the major-accident record for explosives manufacturing and storage in the UK. *Journal of Hazardous Materials*. 1998: A63: 107–111.
- [19] Parent E. and Bernier J. Encoding prior experts judgement to improve risk analysis of extreme hydrological events via POT modeling. *Journal of Hydrology*. 2003; 283: 1–18.
- [20] Park KS and in Lee J. A new method for estimating human error probabilities: AHP-SLIM. *Reliability Engineering & System Safety*. 2008; 93(4): 578-587.
- [21] Persaud B. and Lyon C. Empirical Bayes before-after studies: Lessons learned from two decades of experience and future directions. *Accident Analysis and Prevention*. 2007; 39: 546–555.
- [22] Quigley J, Bedford T and Walls L. Estimating rates of occurrence of rare events with empirical Bayes: A railway application. *Reliability Engineering and System Safety*. 2007; 92: 619–627.

- [23] Quigley J and Revie M. Estimating the Probability of Rare Events: Addressing Zero Failure Data. *To appear in Risk Analysis*, 2011.
- [24] Sarhan AM. Empirical Bayes estimates in exponential reliability model. *Applied Mathematics and Computation*. 2003; 135: 319–332.
- [25] Srivastava MS and Kubokawa T. Empirical Bayes regression analysis with many regressors but fewer observations. *Journal of Statistical Planning and Inference*. 2007; 137: 3778– 3792.
- [26] Szwed P, van Dorp JR, Merrick JRW, Mazzuchi TA, and Singh A. A Bayesian paired comparison study for relative accident probability assessment with covariate information. *European Journal of Operations Research*. 2006; 196: 157–177.
- [27] Vaurio JK and Jankala KE. Evaluation and comparison for estimation methods of failure rates and probabilities. *Reliability Engineering and System Safety*. 2006; 91: 209–221.
- [28] Wharton RK and Bagley MJ. Review of threshold quantities for the control of explosives in health and safety legislation in the Netherlands. *Safety Science*. 2004; 42: 621–626.
- [29] Williams MMR. and Thorne MC. Estimation of failure rates for low probability events. *Progress in Nuclear Energy*. 1997; 31: 373–476.

Author Biographies

John Quigley: John Quigley is a Professor in the department of Management Science at the University of Strathclyde. He has a BMath in Actuarial Science from the University of Waterloo, Canada and a PhD in Management Science from the University of Strathclyde, Scotland. He is an Associate of the Society of Actuaries, a Chartered Statistician and a Member of the Safety and Reliability Society. His current research interests are risk and reliability analysis.

Gavin Hardman: Gavin Hardman was awarded a MMath in Mathematics (2003) and a PhD in Mathematics and Statistics (2007) by the University of Durham. He worked as a Research Associate at the University of Strathclyde Management Science Department between 2006 and 2009 investigating the use of expert judgement methods for low probability events with applications to MoD projects. Gavin currently works as an analyst for NHS West Sussex developing statistical models for forecasting healthcare demand and methods for identifying shifts in healthcare provider behaviour.

Tim Bedford: Tim Bedford studied mathematics at Warwick University and later worked in Cambridge and Delft University of Technology in the Netherlands. He is now Professor of Decision and Risk Analysis at Strathclyde University in Glasgow. He is currently a member of the Board of Directors of the European Safety Reliability and Data Association ESReDA, a Fellow of the UK Safety and Reliability Society and a member of the Mathematics Programme Strategic Advisory Team of the UK Engineering and Physical Sciences Research Council. He was recently awarded an Honorary Doctorate by the Faculté Polytechnique de Mons for his work in probabilistic risk analysis. Current research interests are in probabilistic risk analysis, uncertainty modelling and decision analysis.

Lesley Walls: Lesley Walls has a PhD in Applied Statistics and is a Professor in the Department of Management Science at the University of Strathclyde. She is a Chartered Statistician and a Fellow of

the UK Safety and Reliability Society. She is an Associate Editor of IEEE Transactions on Reliability and a UK expert to the International Electrotechnical Commission. Current research interests are reliability analysis and operations risk modelling.

Appendix

Proof of Result 1: There is a linear relationship between the pool parameters without homogenisation and those with homogenisation.

We assume that the first two moments of the pooled rate of occurrence of events are equal. This implies the equating of $E\left[\sum_{i=1}^m \Lambda_i\right]$ and $Var\left[\sum_{i=1}^m \Lambda_i\right]$ for both models. As such we have the following:

$$m \frac{\alpha}{\beta} = \frac{a}{b} \sum_{i=1}^m h_i = m \frac{a}{b} \bar{h}$$

$$m \frac{\alpha}{\beta^2} = \frac{a}{b^2} \sum_{i=1}^m h_i^2 = m \frac{a}{b^2} \overline{h^2}$$

This is easily manipulated to give:

$$\alpha = a \frac{(\bar{h})^2}{\bar{h}^2} = ar\bar{h}$$

$$\beta = b \frac{\bar{h}}{\bar{h}^2} = br$$

By the Cauchy-Schwarz inequality $\overline{h^2} \geq (\bar{h})^2$. Therefore, $\alpha \leq a$

Since the means and variances of the pools are equal we find:

$$\frac{\alpha}{\beta} = \frac{a}{b} \bar{h} \geq \frac{\alpha}{b} \bar{h} \rightarrow \beta \leq \frac{b}{\bar{h}}$$

$$\frac{\alpha}{\beta^2} = \frac{a}{b^2} \overline{h^2} \geq \frac{\alpha}{b^2} \bar{h}^2 \rightarrow \beta \leq \frac{b}{\sqrt{\overline{h^2}}}$$

The second of which is the lower upper bound.

Proof of Result 2: As pool size increases towards infinity misspecification of the homogenisation factors has the following effect on parameter estimators

Asymptotically, the moments of the adjusted exposure time becomes the following:

$$\lim_{m \rightarrow \infty} \frac{\sum_{i=1}^m \tilde{h}_i t}{m} \xrightarrow{P} \theta \bar{h} t$$

$$\lim_{m \rightarrow \infty} \frac{\sum_{i=1}^m (\tilde{h}_i t)^2}{m} \xrightarrow{P} (\sigma^2 + \theta^2) \bar{h}^2 t^2$$

Asymptotically, the moments of the number of events is:

$$\lim_{m \rightarrow \infty} \frac{\sum_{i=1}^m n_i}{m} \xrightarrow{P} \frac{a}{b} \lim_{m \rightarrow \infty} \frac{\sum_{i=1}^m h_i}{m} t = \frac{a}{b} \bar{h} t$$

$$\lim_{m \rightarrow \infty} \frac{\sum_{i=1}^m n_i^2}{m} \xrightarrow{P} \frac{a}{b} \lim_{m \rightarrow \infty} \frac{\sum_{i=1}^m h_i}{m} t + \frac{a(a+1)}{b^2} \lim_{m \rightarrow \infty} \frac{\sum_{i=1}^m (h_i)^2}{m} t^2 = \frac{a}{b} \bar{h} t + \frac{a(a+1)}{b^2} \bar{h}^2 t^2$$

Substituting these limits into the expressions for W and U results in the following:

$$\lim_{m \rightarrow \infty} W = \frac{\sum_{i=1}^m n_i^2 - \sum_{i=1}^m n_i}{\sum_{i=1}^m (\tilde{h}_i t_i)^2} \xrightarrow{P} \frac{\frac{a}{b} \bar{h} t + \frac{a(a+1)}{b^2} \bar{h}^2 t^2 - \frac{a}{b} \bar{h} t}{(\sigma^2 + \theta^2) \bar{h}^2 t^2} = \frac{a(a+1)}{b^2 (\sigma^2 + \theta^2)}$$

$$\lim_{m \rightarrow \infty} U = \frac{\sum_{i=1}^m n_i}{\sum_{i=1}^m \tilde{h}_i t_i} \xrightarrow{P} \frac{\frac{a}{b} \bar{h} t}{\theta \bar{h} t} = \frac{a}{b\theta}$$

Substitution into the equations for the parameter estimates results in the following:

$$\lim_{m \rightarrow \infty} \hat{a} = \frac{U^2}{W - U^2} \xrightarrow{P} \frac{\left(\frac{a}{b\theta}\right)^2}{\frac{a(a+1)}{b^2(\sigma^2 + \theta^2)} - \left(\frac{a}{b\theta}\right)^2} = \frac{a(\sigma^2 + \theta^2)}{\theta^2(a+1) - a(\sigma^2 + \theta^2)} = a \frac{\sigma^2 + \theta^2}{\theta^2 - a\sigma^2}$$

$$\lim_{m \rightarrow \infty} \hat{b} = \frac{U}{W - U^2} \xrightarrow{P} \frac{\frac{a}{b\theta}}{\frac{a(a+1)}{b^2(\sigma^2 + \theta^2)} - \left(\frac{a}{b\theta}\right)^2} = \frac{b\theta(\sigma^2 + \theta^2)}{(a+1)\theta^2 - a(\sigma^2 + \theta^2)} = b\theta \frac{\sigma^2 + \theta^2}{\theta^2 - a\sigma^2}$$

Proof of Result 3: The empirical prior estimator of the aggregate rate of occurrence of events in the pool does not depend on misspecification of the homogenisation factors as the pool size tends to infinity.

Given the experts subjective assessment the prior mean for the i -th process is:

$$E[\Lambda_i] = \frac{\hat{a}}{\hat{b}} \tilde{h}_i$$

We know from Result 2 that the estimate of the ratio of the pool parameter converges in probability to the following:

$$E[\Lambda_i] \xrightarrow{P} \frac{a \tilde{h}_i}{b \theta}$$

Therefore, we have the following result:

$$\lim_{m \rightarrow \infty} \sum_{i=1}^m \frac{E[\Lambda_i]}{m} \xrightarrow{P} \frac{a}{b}$$

Proof of Result 4: The empirical posterior estimator of the rate of occurrence of events for process i have the following limit

Using Result 2 and re-arranging we have the following:

$$\begin{aligned} \lim_{m \rightarrow \infty} E[\Lambda_i | n_i] &\xrightarrow{P} \frac{a \frac{\sigma^2 + \theta^2}{\theta^2 - a\sigma^2} + n_i}{b\theta \frac{\sigma^2 + \theta^2}{\theta^2 - a\sigma^2} + t\tilde{h}_i} \tilde{h}_i \\ &= \frac{a \tilde{h}_i}{b \mu} \left(1 - \frac{t}{b \frac{\theta}{\tilde{h}_i} \left(\frac{\sigma^2 + \theta^2}{\theta^2 - a\sigma^2} \right) + t} \right) + \frac{n_i}{t} \left(\frac{t}{b \frac{\theta}{\tilde{h}_i} \left(\frac{\sigma^2 + \theta^2}{\theta^2 - a\sigma^2} \right) + t} \right) \end{aligned}$$

Table 1: Types of expert behaviour considered in simulation study

Expert Code	Expert Label	Description of Behaviour
ET1	Good Expert	Expert correctly assigns homogenisation to all processes so all resulting errors can be attributed purely to sampling error
ET2	Pessimistic Expert	Expert consistently overestimates the risk and assigns larger homogenisation factors than are appropriate
ET3	Optimistic Expert	Expert consistently underestimates the risk and assigns smaller homogenisation factors than are appropriate
ET4	Partial Expert	Expert correctly assesses half of the processes for each sub-pool, but otherwise underestimates high risk processes and overestimates low risk
ET5	Erratic Expert	Expert assigns homogenisation factors at random

Table 2: Homogenisation factors under test

Expert Type	$1, \ell$	$\ell+1, 2\ell$	$2\ell+1, 3\ell$	$3\ell+1, 4\ell$	$4\ell+1, 5\ell$	$5\ell+1, 6\ell$	$6\ell+1, 7\ell$	$7\ell+1, 8\ell$
True	1	1	10	10	100	100	1000	1000
ET1	1	1	10	10	100	100	1000	1000
ET2	10	10	100	100	1000	1000	1000	10000
ET3	1	1	1	1	10	10	100	100
ET4	1	10	10	100	100	10	1000	100
ET5	10	1	1	1000	1000	1	100	10

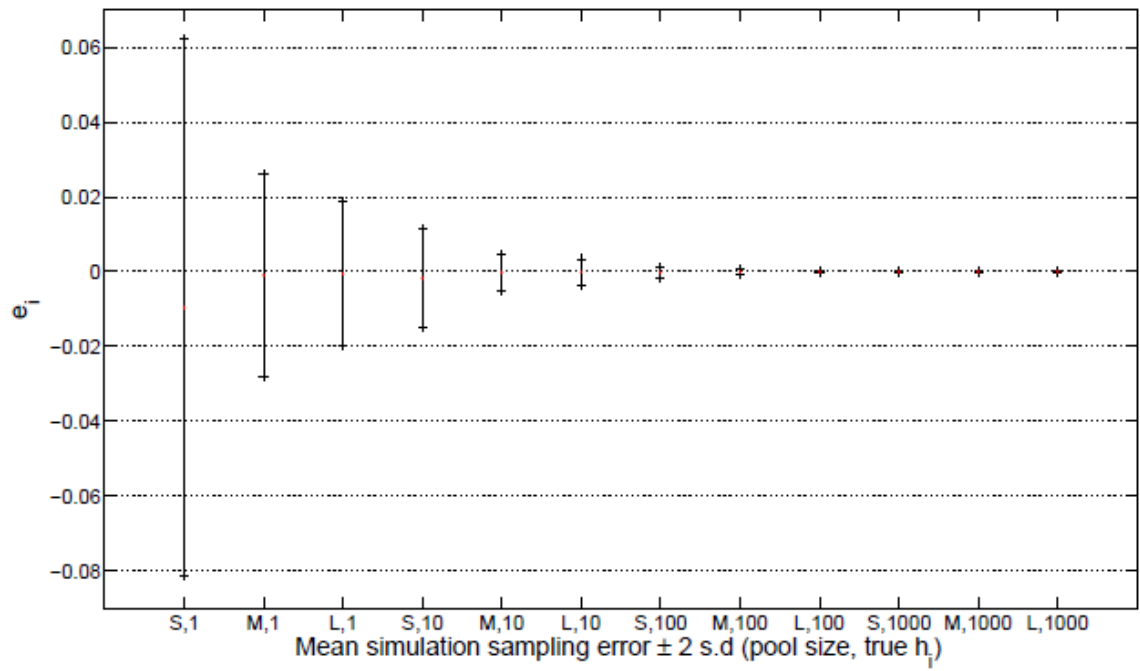


Figure 1: Mean simulation sampling error +/- 2 standard deviations and classed by pool size (small - S, medium - M large - L), true value of h_i (1, 10, 100, 1000)

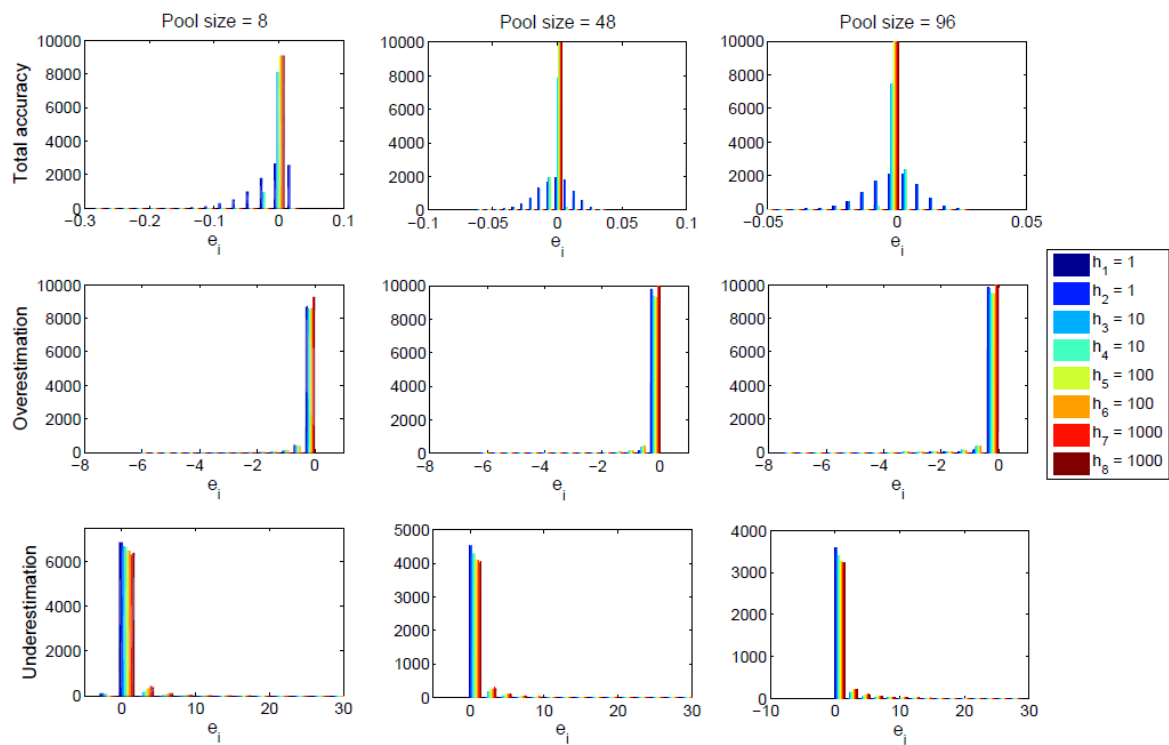


Figure 2: Pool size effects on error

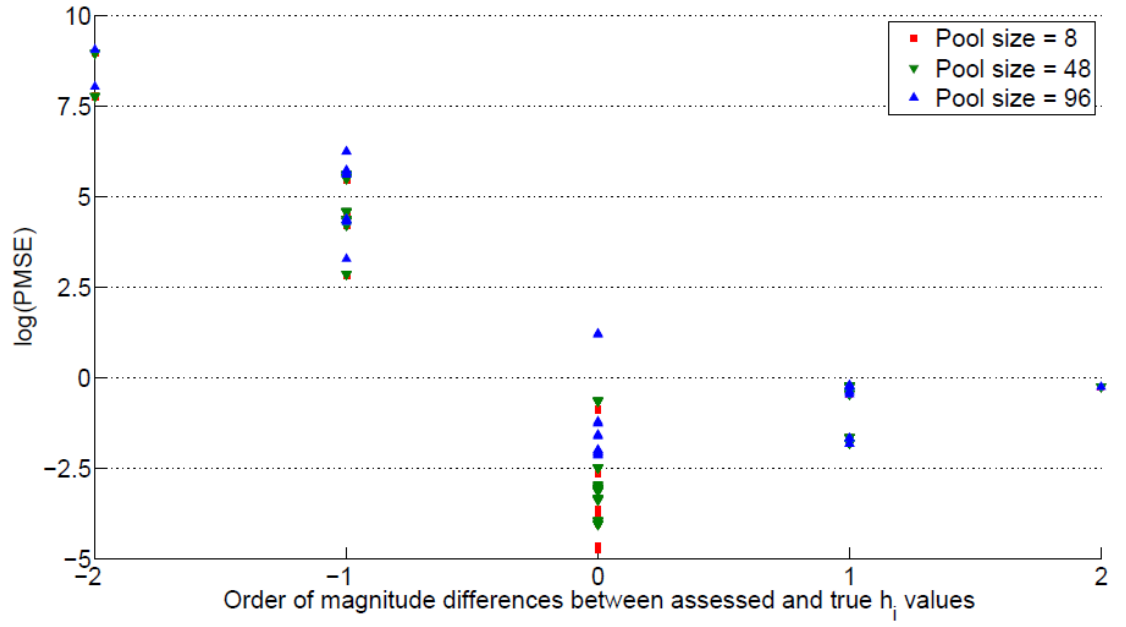


Figure 3: Dependence of observed error on subjective assessment error

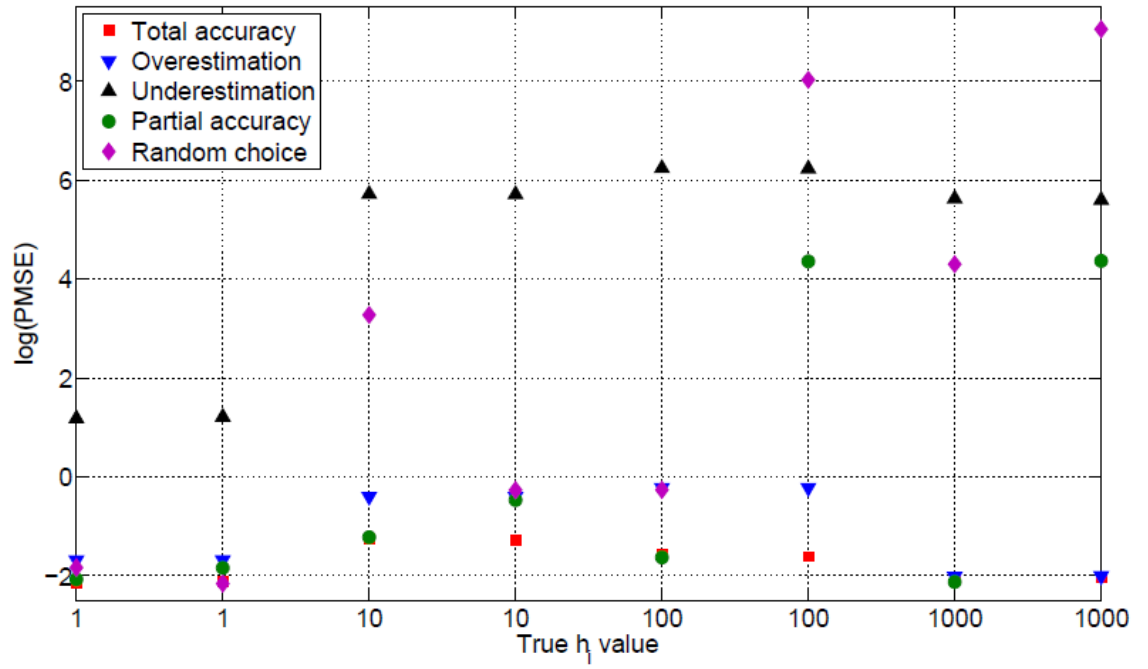


Figure 4: Comparison of expert types for pool size = 8

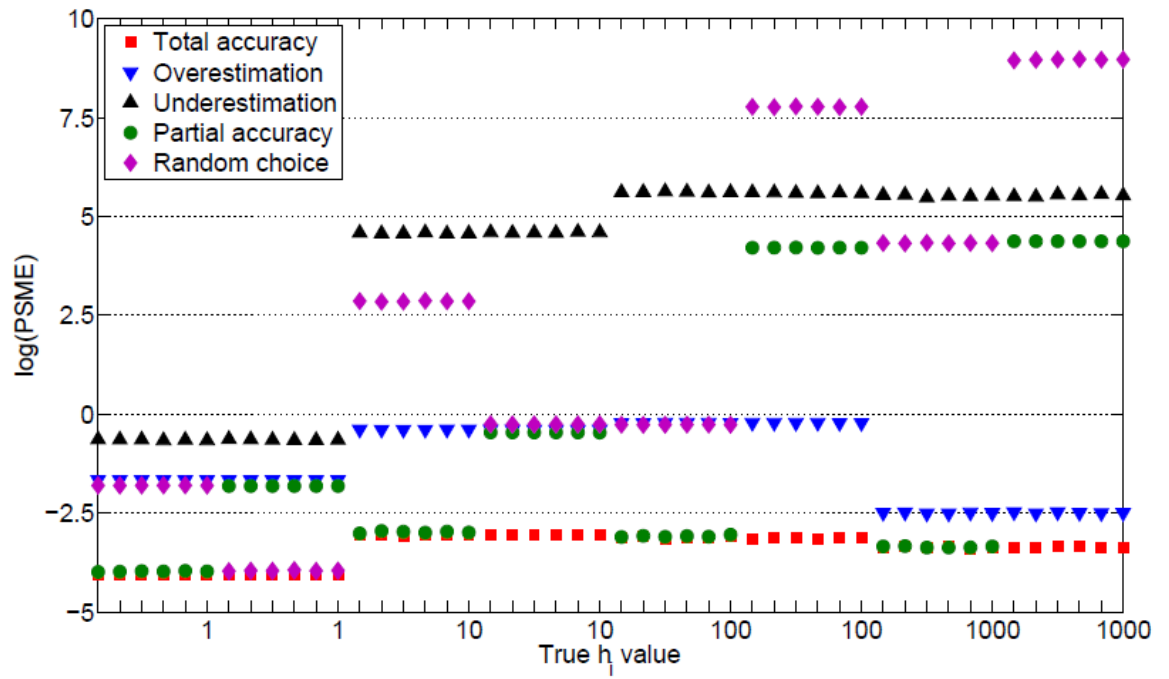


Figure 5: Comparison of expert types for pool size = 48

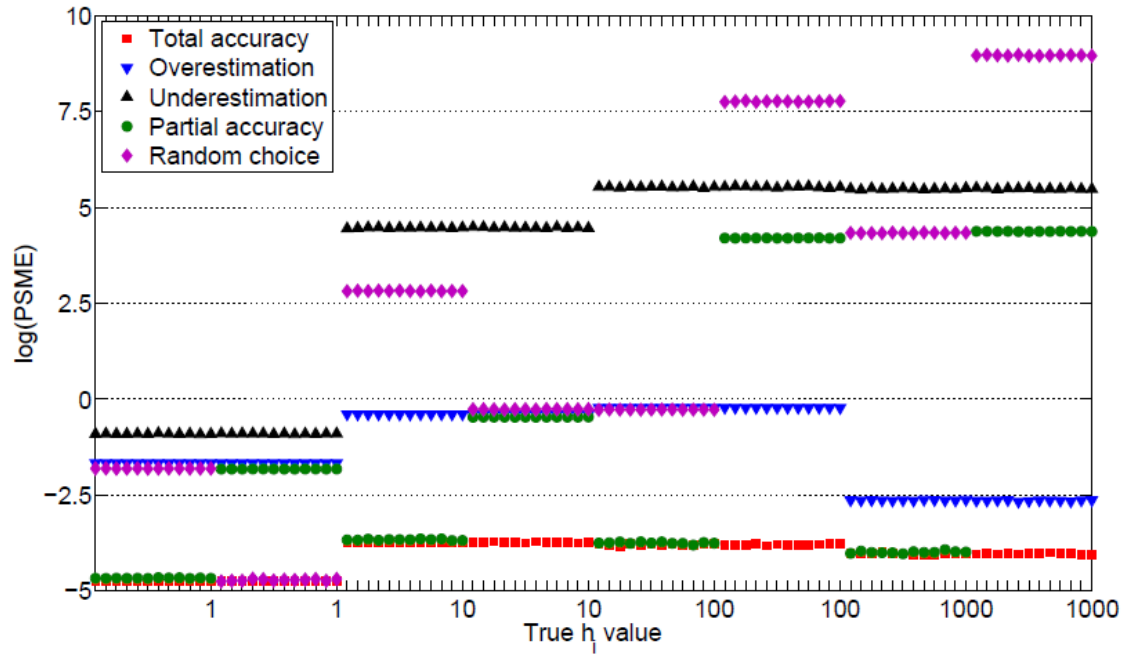


Figure 6: Comparison of expert types for pool size = 96