Working Paper 95-25
Statistics and Econometrics Series 07
July 1995

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

ON CREDIBILITY AND ROBUSTNESS WITH THE KALMAN FILTER

José Garrido and Rosario Romera*

Abstract

Bühlmann (1967) gave a formal Bayesian derivation of the credibility ratio estimators that actuaries had been using for many years. Since then various generalizations of Bühlmann's model have appeared in the literature, each relaxing the i.i.d. assumptions in its own way. The introduction of weights is due to Bülhmann & Straub (1970) and that the regressors to Hachemeister (1975), but the first comprehensive actuarial application of the Kalman filter is due to de Jong & Zehnwirth (1983).

More recent efforts have concentrated on the robustification of these estimators, as they proved to be extremely sensitive to large claims. Kremer (1991) studies a robust regression credibility model and Künsch (1992) tackles the weighted case. Following Kremer (1994) we propose here a robust Kalman filter credibility model.

Key Words
Credibility; Theory Robustness; Kalman Filter.

*Garrido, Concordia University; Romera, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid.

On Credibility and Robustness with the Kalman Filter

José Garrido
Concordia University
and
Rosario Romera
Universidad Carlos III de Madrid

June 30, 1995

Abstract

Bühlmann(1967) gave a formal Bayesian derivation of the credibility ratio estimators that actuaries had been using for many years. Since then various generalizations of Bühlmann's model have appeared in the literature, each relaxing the i.i.d. assumptions in its own way. The introduction of weights is due to Bühlmann & Straub(1970) and that of regressors to Hachemeister(1975), but the first comprehensive actuarial application of the Kalman filter is due to de Jong & Zehnwirth(1983).

More recent efforts have concentrated on the robustification of these estimators, as they proved to be extremely sensitive to large claims. Kremer(1991) studies a robust regression credibility model and Künsch(1992) tackles the weighted case. Following Kremer(1994) we propose here a robust Kalman filter credibility model.

1 Introduction

The lack of robustness of classical credibility estimators, i.e. their extreme sensitivity to large claims, is adressed in a series of recent works. Kremer(1991) studies a robust regression credibility model, duplicating, to a certain extent, the independent and prior work of Künsch, that appeared in 1992 and where he treats the weighted observations case. Following Kremer(1994), we propose here a robustification of the Kalman filter, a recursive estimator that can be given a Credibility interpretation. In the paper, we also implement the empirical version of this robust Kalman filter credibility estimator with a data set, and compare its sensitivity to large claims with that of other credibility estimators in the literature.

2 Classical Credibility

2.1 Bühlmann(1967): for i.i.d. observations.

- Assumptions:

  (i) For $j = 1, ..., k$ the pairs $(\Theta_j, \tilde{X}_j)$ are i.i.d.,

  (ii) Conditionally on Θ_j the r.v.'s $X_{j1}, ..., X_{jt}$ are i.i.d..

  Since the results only depend on the first two moments of the distribution of X_{jr}, the last assumption is usually replaced by

  (ii) $E(X_{jr}|\Theta_j) = \mu(\Theta_j)$ and $V(X_{jr}|\Theta_j) = \sigma^2(\Theta_j)$ for all $r = 1, ..., t$.

- The non-homogeneous linear Bayes rule which minimizes

$$E\{[\mu(\Theta_j) - c_0 - \sum_1^t c_r X_{jr}]^2\} \ , \ j = 1, ..., k \text{ fixed}$$

  is given by

$$\hat{\mu}(\Theta_j) = E(X_{jr}) + Z\ [\bar{X}_{j.} - E(X_{jr})]$$

  for

$$Z = \frac{V[\mu(\Theta_j)]\ t}{E[\sigma^2(\Theta_j)] + V[\mu(\Theta_j)]\ t}.$$

- The empirical Bayes rule is obtained by substituting the parameters $E(X_{jr})$, $V[\mu(\Theta_j)]$ and $E[\sigma^2(\Theta_j)]$ with unbiassed estimators.

2.2 Bühlmann & Straub(1970): the weighted case

- Assumptions: Here the observations are still assumed independent and with constant mean, but the variance is allowed to vary with the values of known weights.

  (i) For $j = 1, ..., k$ the vectors $(\Theta_j, \tilde{X}_j)$ are pair-wise independent and the Θ_j's are identically distributed,

  (ii) Conditionally on Θ_j the r.v.'s $X_{j1}, ..., X_{jt}$ are independent,

  (iii) $E(X_{jr}|\Theta_j) = \mu(\Theta_j)$ for all $r = 1, ..., t$,

  (iv) $Cov(X_{jr}, X_{ju}|\Theta_j) = \frac{\delta_{ru}}{w_{jr}}\sigma^2(\Theta_j)$ where the w_{jr} are known weights.

- The non-homogeneous linear Bayes rule in Bühlmann-Straub's case (B-S) is given by

$$\hat{\mu}(\Theta_j) = E(X_{jr}) + Z_j \, [\bar{X}_{jW} - E(X_{jr})]$$

for

$$Z_j = \frac{V[\mu(\Theta_j)]w_{j.}}{E[\sigma^2(\Theta_j)] + V[\mu(\Theta_j)]w_{j.}} ,$$

where

$$w_{j.} = \sum_{r=1}^{t} w_{jr} \text{ and } \bar{X}_{jW} = \sum_{r=1}^{t} \frac{w_{jr}}{w_{j.}} X_{jr} .$$

- The empirical Bayes rule is obtained by substituting the parameters $E(X_{jr})$, $V[\mu(\Theta_j)]$ and $E[\sigma^2(\Theta_j)]$ with unbiassed estimators.

2.3 Hachemeister(1975): the regression case

- Assumptions: This model departs from the i.i.d. case by allowing for different means, explained through regressors, and for different variances, which are again function of known weights, as in B-S's case. In addition it allows for possible covariances between observations and contracts.

  (i) For $j = 1, ..., k$ the vectors $(\Theta_j, \tilde{X}_j)$ are pair-wise independent and the Θ_j's are identically distributed,

(ii) Here $\tilde{\mu}(\Theta_j) = (\mu_1(\Theta_j), ..., \mu_t(\Theta_j))'$ with

$$\begin{aligned} E(\tilde{X}_j|\Theta_j) = \tilde{\mu}_j(\Theta_j) &= \mathbf{Y}\tilde{\beta}(\Theta_j) \\ &= \mathbf{Y}(\beta_1(\Theta_j), ..., \beta_n(\Theta_j))' \end{aligned}$$

where $\mathbf{Y}$ is a design matrix (of rank $n < t$) and $\tilde{\beta}(\Theta_j)$ is an unknown regression vector.

(iii) $Cov(\tilde{X}_j|\Theta_j) = \sigma^2(\Theta_j)\mathbf{V}_j$ where $\mathbf{V}_j$ is a positive semi-definite matrix of known weights and σ^2 a scalar function.

- The weighted least squares estimators are

$$\hat{\beta}(\Theta_j) = (\mathbf{Y}'\mathbf{V}_j^{-1}\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{V}_j^{-1}\tilde{X}_j$$

and

$$\hat{\mu}(\Theta_j) = \mathbf{Y}\hat{\beta}(\Theta_j) \ .$$

- The non-homogeneous linear Bayes rule is given by

$$E[\tilde{\mu}(\Theta_j)] + \mathbf{Y}Z_j \ [\hat{\beta}(\Theta_j) - E[\tilde{\beta}(\Theta_j)] \ ,$$

where

$$Z_j = Cov[\tilde{\beta}(\Theta_j)]\mathbf{Y}'\mathbf{V}_j^{-1}\mathbf{Y}\left[I + Cov[\tilde{\beta}(\Theta_j)]\mathbf{Y}'\mathbf{V}_j^{-1}\mathbf{Y}\right]^{-1} \ .$$

- The empirical Bayes rule is obtained by substituting the parameters $E[\tilde{\beta}(\Theta_j)]$, $Cov[\tilde{\beta}(\Theta_j)]$ and $E[\sigma^2(\Theta_j)]$ with unbiassed estimators.

2.4 de Jong & Zehnwirth(1983): the Kalman filter

- Assumptions: Here is a brief description of a Kalman filter model. More general versions exist.

  (i) For $r = 1, ..., t$ the row vectors of observations satisfy

  $$\tilde{X}_r = \mathbf{Y}_r\tilde{\beta}_r + \tilde{u}_r \ ,$$

  where $\mathbf{Y}_r$ is a known design matrix, $\tilde{\beta}_r$ is the parameter vector applicable at time r, and $\tilde{u}_r$ is a zero mean error vector with covariance matrix U_r.

(ii) Similarly, the row vectors of state parameters satisfy

$$\tilde{\beta}_r = \mathbf{H}_r \tilde{\beta}_{r-1} + \tilde{v}_r \ ,$$

where $\mathbf{H}_r$ is a known transition matrix, and $\tilde{v}_r$ is a zero mean vector of random shocks with covariance matrix V_r.

(iii) The random vectors $u(r), u(s), v(r), v(s)$ are uncorrelated.

- The least squares estimators of $\tilde{\beta}_r$ are obtained recursively from the observations $\tilde{X}_1, ..., \tilde{X}_{r-1}$, given a $\hat{\beta}_0$

$$\hat{\beta}_{r|r-1} = \mathbf{H}_r \hat{\beta}_{r-1} \text{ for } r = 1, ..., t$$

where

$$\hat{\beta}_r = \hat{\beta}_{r|r-1} + \mathbf{K}_r[\tilde{X}_r - \mathbf{Y}_r \hat{\beta}_{r|r-1}]$$

and given $\mathbf{C}_0$

$$\mathbf{K}_r = \mathbf{C}_{r|r-1}\mathbf{Y'}_r[\mathbf{Y}_r\mathbf{C}_{r|r-1}\mathbf{Y'}_r + \mathbf{U}_r]^{-1} \text{ for } r = 1, ..., t$$

where

$$\mathbf{C}_{r|r-1} = \mathbf{H}_r \mathbf{C}_{r-1} \mathbf{H'}_r + \mathbf{V}_r$$

- The filter can be given a Credibility interpretation, for instance in the Bühlmann & Straub case, fixing $j = 1, ..., k$

$$\mathbf{H}_r = 1 = \mathbf{Y}_r$$

$$\beta_r = \mu(\Theta_j) \ \ \mathbf{U}_r = \frac{\sigma^2(\Theta_j)}{w_{jr}} \text{ and } \mathbf{V}_r = 0$$

$$\beta_0 = E[\mu(\Theta_j)] \ , \mathbf{C}_0 = V[\mu(\Theta_j)] \ .$$

- The same is possible for Hachemeister's model introducing vectors in the above relations.

3 Robust Credibility

3.1 Motivation

- Before analysing robust credibility models we give here a simple empirical example that shows the sensitivity of classical credibility estimators to large claims.

  Consider Hachemeister's data set: 5 contracts, 12 periods (of 3 months). Available information:

  (i) Average bodily injury claims/period for an Automobile Insurance portfolio.

  (ii) Number of claims/period, used here as weights.

Table 1: Hachemeister's Data Set
Average Claims per Period (Number of Claims per period)

1,738 (7,861)	1,364 (1,622)	1,759 (1,147)	1,223 (407)	1,456 (2,902)
1,642 (9,251)	1,408 (1,742)	1,685 (1,357)	1,146 (396)	1,499 (3,172)
1,794 (8,706)	1,597 (1,523)	1,479 (1,329)	1,010 (348)	1,609 (3,046)
2,051 (8,575)	1,444 (1,515)	1,763 (1,204)	1,257 (341)	1,741 (3,068)
2,079 (7,917)	1,342 (1,622)	1,674 (998)	1,426 (315)	1,482 (2,693)
2,234 (8,263)	1,675 (1,602)	2,103 (1,077)	1,532 (328)	1,572 (2,910)
2,032 (9,456)	1,470 (1,964)	1,502 (1,277)	1,953 (352)	1,606 (3,275)
2,035 (8,003)	1,448 (1,515)	1,622 (1,218)	1,123 (331)	1,735 (2,697)
2,115 (7,365)	1,464 (1,527)	1,828 (896)	1,343 (287)	1,607 (2,663)
2,262 (7,832)	1,831 (1,748)	2,155 (1,003)	1,243 (384)	1,573 (3,017)
2,267 (7,849)	1,612 (1,654)	2,233 (1,108)	1,762 (321)	1,613 (3,242)
2,517 (9,077)	1,471 (1,861)	2,759 (1,121)	1,306 (342)	1,690 (3,425)

Now assume that the last claim amount of \$1,690 in Table 1 was miscoded or is replaced by a large claim. Table 2 gives the various components of the B-S premium estimators for the correct data value of \$1,690, as well as for three outlier values: \$5,000, \$6,000 and \$7,000. Notice the radical change in the individual premium $\bar{X}_{5W}$ for contract 5, affected by the outlier contamination.

A more serious problem, also apparent in Table 2, is the large effect a small contamination to a single data point of contract 5 has on the individual credibility factors, Z_j, and hence credibility premiums, $\hat{\mu}(\Theta_j)$, for all contracts.

The last row of Table 2 shows the effect on the credibility weighted portfolio average $\hat{X}_{ZW}$.

Table 2: Bühlmann-Straub's Premiums

Outlier	1,690	5,000	6,000	7,000
$\bar{X}_{1W}$	2,061	2,061	2,061	2,061
$\bar{X}_{2W}$	1,511	1,511	1,511	1,511
$\bar{X}_{3W}$	1,805	1,805	1,805	1,805
$\bar{X}_{4W}$	1,352	1,352	1,352	1,352
$\bar{X}_{5W}$	1,599	1,913	2,008	2,032
Z_1	0.9847	0.8130	0.6077	0.9653
Z_2	0.9276	0.4634	0.2353	0.8467
Z_3	0.8985	0.3740	0.1752	0.7922
Z_4	0.7279	0.1527	0.0603	0.5355
Z_5	0.9588	0.6105	0.3583	0.5355
$\hat{\mu}(\Theta_1)$	2,055	2,018	1,997	2,051
$\hat{\mu}(\Theta_2)$	1,524	1,684	1,806	1,551
$\hat{\mu}(\Theta_3)$	1,793	1,823	1,881	1,799
$\hat{\mu}(\Theta_4)$	1,443	1,760	1,864	1,548
$\hat{\mu}(\Theta_5)$	1,603	1,883	1,937	1,911
$\bar{X}_{ZW}$	1,684	1,833	1,897	1,772

The estimation is so unstable that when the outlier takes values larger than \$7,500, overflow occurs. One reason for this lack of robustness of B-S premiums is the sensitivity of the variance component estimators, $\hat{E}[\sigma^2(\Theta_j)]$ for the between contract variance (or heterogeneity) and $\hat{V}[\mu(\Theta_j)]$ for within contract variance. Table 3 shows how classical unbiassed estimators of these components wildly vary from exact to contaminated data sets.

Table 3: Variance Components Estimators

Outlier	1,690	5,000	6,000	7,000
$\hat{E}[\sigma^2(\Theta_j)]$	139,120,026	793,846,681	1,234,587,691	301,901,954
$\hat{V}[\mu(\Theta_j)]$	89,639	34,457	19,093	83,819

Similar problems occur when estimating premiums with the other classical methods such as Hachemeister's and the Kalman filter.

3.2 Künsch(1992), Gisler & Reinhard(1993)

- Assumptions: Here the robustification essentially takes B-S's model as a starting point.

  (i) In $\hat{\mu}(\Theta_j) = E(X_{jr}) + Z_j\ [\bar{X}_{jW} - E(X_{jr})]$ replace $\bar{X}_{jW}$ by a robust estimator: e.g. Künsch uses an M-estimator, $T_j(X_{j1}, ..., X_{jt})$, which is the implicit solution of

  $$\sum_{r=1}^{t} \chi(\frac{X_{jr}}{T_j}) = 0$$

  where

  $$\chi(z) = max\{-c_1,\ min(z-1, c_2)\} \text{ for } 0 < c_1 \leq 1 \text{ and } 0 < c_2\ .$$

  (ii) For the empirical credibility estimator, replace the unbiassed estimators of $E(X_{jr})$ by robust estimators, and the variance components $V[\mu(\Theta_j)]$ and $E[\sigma^2(\Theta_j)]$ by robust estimators: e.g. Künsch uses M-estimators again, while Gisler & Reinhard, which account for the large-claims provision in Swiss law, suggest another robust estimator defined implicitely according to an optimal truncation property.

3.3 Kremer(1991)

- Assumptions: As in Künsch(1992), a robust treatment of the B-S case is given, but in addition a robust regression model is discussed. Essentially the suggestion is to use an M-estimator to replace the weighted least squares estimation of $\hat{\beta}(\Theta_j)$. Gisler & Reinhard(1993) also hint to this idea but with using rather a general Huber estimator.

3.4 Kremer(1994)

Starting from the Kalman filter model of de Jong & Zehnwirth(1983), a robustification is used and interpreted in a Credibility context. Various robust versions of the Kalman filter can be found in the Statistics literature. Kremer uses that of Cipra & Romera(1991). Computational aspects of this robust Kalman filter can be found in Romera & Cipra(1995)

- Assumptions: Kremer uses a simple version of the robust Kalman filter of Cipra & Romera, estimating the one step ahead prediction $\hat{\beta}_{r|r-1}$ with an M-estimator. The model can be given a B-S, regression or general filtering interpretation in Credibility.

  (i) No robustification of the empirical credibility estimator is suggested.

  (ii) No comparison is made of the relative robustness of these Kalman credibility estimators. Other robust estimators, such as Künsch's M-estimator, Kremer's L-estimator or Gisler's optimal trimming estimator may be easier to implement and just as robust.

  (iii) The problems that arise in implementation with data are not treated.

The deficiencies in Kremer's study are addressed in this paper and a full study of the implementation phase carried out.

References

Bühlmann, H.(1967) "Experience rating and credibility", *ASTIN Bull.*, **4**, 3, pp.199-207.

Bühlmann, H. & Straub, E.(1970) "Glaubwürdigkeit für Schadensätze", *Mitt. Ver. Schweiz. Ver.*, **70**, 1, pp.111-133.

Cipra, T. & Romera, R.(1991) "Robust Kalman filter and its application in time series analysis", *Kybernetika*, **6**, pp.481-494.

de Jong, P. & Zehnwirth, B.(1983) "Credibility theory and the Kalman filter", *Insurance: Mathematics & Economics*, **2**, 4, pp.281-286.

Hachemeister, C.A.(1975) "Credibility for regression models with application to trend", in *Credibility, Theory and Applications*, Proc. Berkeley Act. Res. Conf. on Cred., Academic Press, New-York.

Künsch, H.R.(1992) "Robust methods for Credibility", *ASTIN Bulletin*, **22**, 1, pp.33-49.

Kremer, E.(1994) "Robust Credibility via Kalman filtering", Tech. Rep. **94-2**, *Inst. für Math. Stoc., U. of Hamburg.*

Kremer, E.(1991) "Large claims in Credibility", *Blat. Deut. Ges. Ver.*, pp.123-150.

Gisler, A. & Reinhard, P.(1993) "Robust Credibility", *ASTIN Bulletin*, **23**, 1, pp.117-144.

Romera, R. & Cipra, T.(1995) "On Practical Implementation of Robust Kalman Filtering", *Commun. Statist.- Simula.* **24-2**, pp 461-488.