Supplementary material for:
“Calendar effect and in-sample forecasting”

Enno Mammen

University of Heidelberg, Germany

mammen@math.uni-heidelberg.de

María Dolores Martínez-Miranda

University of Granada, Spain

mmiranda@ugr.es

Jens Perch Nielsen

Cass Business School, City, University of London, U.K.

Jens.Nielsen.1@city.ac.uk

Michael Vogt

University of Bonn, Germany

michael.vogt@uni-bonn.de

January 18, 2020

1 Introduction

Recently, Kuang et al. (2011) have studied age-period-cohort forecasting of outstanding liabilities in non-life insurance (also called claims reserving). Three different forecasters were defined and named as $I(0)$ (zero-times), $I(1)$ (one-time) and $I(2)$ (two-times) integrators. The authors argued that it is up to the analyst to decide which forecasting approach is most appropriate for the application at hand. In this document, we describe three new parametrizations of the three forecasting models considered by Kuang et al. (2011). Through these new parametrizations, we show that the main difference between the three forecasting models lies in the amount of data eventually used to estimate the slope of the calendar effect which is employed for extrapolation. The $I(2)$ approach uses the shortest possible past to estimate and forecast the slope of the calendar effect, whereas the $I(0)$ and $I(1)$ approaches use the longest possible past. In other words, the three models considered in Kuang et al. (2011) represent two extremes when selecting a forecasting strategy. In the following sections, we work under the discrete framework of Kuang et al. (2011), which can be seen as a discretization of the continuous framework considered in the paper.

2 A discrete model with piecewise constant functions

In claims reserving, the data are traditionally provided as run-off triangles. These show the observed numbers of claims according to the origin (accident period) and the delay (development period) of the claim. Similarly, in mortality studies, the available information consists of an array of numbers of deaths according to the age and the period of death. The traditional approach to work with this kind of aggregated data uses chain ladder-type models (see Kuang et al. (2008a,b, 2011), Martínez-Miranda et al. (2013) and Mammen et al. (2015)). Here, we briefly review these models and the related forecasting methods, and show how they relate to the continuous model and the methods described in the paper. To simplify the exposition, we restrict ourselves to the case of observations given in a run-off triangle, which is the case of the claims reserving problem considered by Kuang et al. (2011).

2.1 Chain ladder-type models

Suppose we have aggregated data in the form of a run-off triangle (see the case study in Section 6.1 of the paper for an example). The triangle can be written as $\aleph_m = \{N_{ij} : (i,j) \in \mathcal{I}_m\}$ with $\mathcal{I}_m = \{(i,j) : i = 1,\dots,m;\ j = 1,\dots,m;\ i+j-1 \leq m\}$. Here, N_{ij} is the total number of claims incurred in period (week, month, quarter, year) i and reported in period $i+j-1$, i.e., with $j-1$ periods delay. The index i thus denotes the accident period (i.e., the period when the accident happened) and j denotes the development period (i.e., the delay of $j-1$ periods from the accident time i). Moreover, m is the latest observed accident period. The aim is to predict the numbers of claims in the upper triangle $\mathcal{J}_m = \{(i,j) : i = 2,\dots,m;\ j = 1,\dots,m;\ i+j-1 > m\}$, which represent the outstanding liabilities in claims reserving. The traditional chain ladder model (see Kuang et al. (2009)) assumes that the entries N_{ij} in the triangle are Poisson distributed independent random variables with mean

$$\mathbb{E}[N_{ij}] = \alpha_i \beta_j \delta, \qquad (i,j) \in \mathcal{I}_m, \tag{1}$$

where α_i $(i = 1,\dots,m)$ are accident parameters, β_j $(j = 1,\dots,m)$ are development parameters, and δ gives the overall level. An extended chain-ladder model (see Kuang et al. (2008a,b, 2011)) considers an additional effect related to the calendar time, which is defined as $i+j-1$. Under this model, the mean parametrization is

$$\mathbb{E}[N_{ij}] = \alpha_i \beta_j \gamma_{i+j-1} \delta, \tag{2}$$

where γ_{i+j-1} denotes the calendar parameter. In mortality studies, this model is the well-known age-period-cohort model, where accident time corresponds to cohort, development time to age and calendar time to period. Taking logarithms, the model can be written as a log-additive model:

$$\mu_{ij} = a_i + b_j + g_{i+j-1} + d \tag{3}$$

with $\mu_{ij} = \log(\mathbb{E}N_{ij})$, $a_i = \log(\alpha_i)$, $b_j = \log(\beta_j)$, $g_{i+j-1} = \log(\gamma_{i+j-1})$ and $d = \log(\delta)$. One problem of this model is that it is not identified since the parametrization in (3) is not unique. As discussed in Kuang et al. (2008a,b), linear trends can be added to and subtracted from the parameters a_i, b_j, g_{i+j-1} and d such that the sum does not change. Thus, μ_{ij} is invariant to the group of transformations

$$\widetilde{\Psi} : \begin{pmatrix} a_i \\ b_j \\ g_{i+j-1} \\ d \end{pmatrix} \mapsto \left\{ \begin{array}{c} a_i + c_1 + (i-1)c_4 \\ b_j + c_2 + (j-1)c_4 \\ g_{i+j-1} + c_3 - (i+j-2)c_4 \\ d - c_1 - c_2 - c_3 \end{array} \right\} \tag{4}$$

for arbitrary constants c_1, c_2, c_3 and c_4, that is, $\mu(\theta) = \mu\{\widetilde{\Psi}(\theta)\}$ with $\theta = (a_i, b_j, g_{i+j-1}, d)$ and $\mu(\theta) = a_i + b_j + g_{i+j-1} + d$. A common way to deal with this identification problem is to impose identification constraints.

Once the identification problem is solved, the parameters in model (3) can be easily estimated by maximum likelihood. However, the calendar parameters will be estimated only up to the last observed calendar period, that is, $i+j-1 = m$. Therefore, forecasting the upper triangle requires to extrapolate these parameters. Kuang et al. (2008b) characterized forecasts of μ_{ij} that are invariant to arbitrary identifications of the parameters. The authors described three such forecasts which are inspired by robust forecasting techniques for non-stationary time series. These are the $I(0)$ (zero-times integrator) model, which is suited for the situation where the time series evolves in a stable way both in-sample and out-of-sample; the $I(1)$ (one-time integrator) model which fits the situation where there is a level shift in the forecast period; and the $I(2)$ (two-times integrator) model for the more extreme situation where there is a slope shift of the linear trend in the forecast period.

The rather conservative $I(0)$ forecast estimates the trend by the simple linear model $g_t = \eta + \nu t$ and extrapolates it into the future. If we denote by $\widehat{g}_1, \ldots, \widehat{g}_m$ the estimated calendar parameters from the data, the $I(0)$ forecast at calendar $m+h$ $(h = 1, \ldots, m)$ is

$$\widetilde{g}_{m+h} = \widehat{\eta} + (m+h)\widehat{\nu}, \tag{5}$$

where $\widehat{\nu} = \{\sum_{t=1}^{m}(\widehat{g}_t - \overline{g})(t - \overline{t})\}/\{\sum_{t=1}^{m}(t-\overline{t})^2\}$ and $\widehat{\eta} = \overline{g} - \widehat{\nu}\overline{t}$ with $\overline{t} = m^{-1}\sum_{t=1}^{m} t$ and $\overline{g} = m^{-1}\sum_{t=1}^{m}\widehat{g}_t$. The $I(1)$ model extrapolates the differences of the time series, $\Delta\widehat{g}_t = \widehat{g}_t - \widehat{g}_{t-1}$, using a random walk model of the type $\Delta g_t = \eta + \varepsilon_t$. Thus, the the $I(1)$ forecast at calendar $m+h$ $(h = 1, \ldots, m)$ is

$$\widetilde{g}_{m+h} = \widehat{g}_m + h\widehat{\eta}, \tag{6}$$

where $\widehat{\eta} = (m-1)^{-1}\sum_{t=2}^{m}\Delta\widehat{g}_t = (m-1)^{-1}(\widehat{g}_m - \widehat{g}_1)$. Finally, the $I(2)$ forecast proceeds by extrapolating the double differences, $\Delta^2\widehat{g}_t = \Delta\widehat{g}_t - \Delta\widehat{g}_{t-1}$, as $\Delta^2\widehat{g}_t = \varepsilon_t$. The h-step-ahead point forecast is then

$$\widetilde{g}_{m+h} = \widehat{g}_m + h\Delta\widehat{g}_m = \widehat{g}_m + h(\widehat{g}_m - \widehat{g}_{m-1}). \tag{7}$$

Kuang et al. (2008b, 2011) showed that either of these forecast approaches is invariant to the transformation in (4). Hence, any identification produces the same $I(0), I(1)$ and $I(2)$ forecast.

2.2 Forecasting methods and identification constraints

We now show that the $I(0)$, $I(1)$ and $I(2)$ forecasts can be reproduced by imposing particular identification constraints on the log-additive model (3) and by setting the extrapolated calendar parameters to zero. This result is closely related to and has inspired the new flexible forecasting method for the continuous chain ladder model in the paper.

The first two identification constraints, which are common to the three forecasting models, arise from the interpretation of the forecasting problem as the estimation of a multiplicatively structured density. Following the discussion in Martínez-Miranda et al. (2013), the parameters $\alpha_i = \exp(a_i)$ and $\beta_j = \exp(b_j)$ can be seen as the values of two histograms (with bin size equal to one) estimating the accident and the development density, respectively. This interpretation immediately yields the following two identification constraints:

$$\sum_{i=1}^{m}\exp(a_i) = \sum_{j=1}^{m}\exp(b_j) = 1. \tag{8}$$

The choice of the last two constraints defines how the extrapolation is performed. Let us start with the $I(1)$ model where the calendar parameters are extrapolated using the expression (6). Since we set the extrapolated calendar values to zero, the $I(1)$ forecast can be reproduced by imposing the two statements in (8) plus the constraint that

$$0 = \widetilde{g}_{m+h} = g_m + h(m-1)^{-1}(g_m - g_1), \quad h = 1, \dots, m-1,$$

which is satisfied for instance if

$$g_1 = g_m = 0, \tag{9}$$

or equivalently, $\gamma_1 = \gamma_m = 1$. Thus, the identification scheme defined by (8) and (9) leads to the $I(1)$ forecast. Similarly, the $I(2)$ method can be reproduced by imposing that

$$0 = \widetilde{g}_{m+h} = g_m + h(g_m - g_{m-1}), \quad h = 1, \dots, m-1,$$

which is satisfied for instance if

$$g_{m-1} = g_m = 0, \tag{10}$$

or equivalently, $\gamma_{m-1} = \gamma_m = 1$. The identification scheme defined by (8) and (10) thus leads to the $I(2)$ forecast. Finally, in order to reproduce the $I(0)$ method, we impose that

$$0 = \widetilde{g}_{m+h} = \widehat{\eta} + (m+h)\widehat{\nu}, \tag{11}$$

where $\widehat{\nu}$ and $\widehat{\eta}$ are respectively the slope and the intercept of the linear regression of the calendar parameters $\{g_1, \dots, g_m\}$ in (5). Note that (11) is satisfied if both slope and intercept vanish, which means that

$$\sum_{t=1}^{m} g_t = 0 \quad \text{and} \quad \sum_{t=1}^{m} t g_t = 0, \tag{12}$$

or equivalently, $\sum_{t=1}^{m} \log(\gamma_t) = \sum_{t=1}^{m} t \log(\gamma_t) = 0$. Thus, we can reproduce the $I(0)$ forecast by using the identification scheme in (8) and (12).

Explicit expressions for the parameters that satisfy the conditions (8), (9), (10) and (12) can be derived from any given vector of identified parameters $(a_1^\dagger, \dots, a_m^\dagger,\ b_1^\dagger, \dots, b_m^\dagger,\ g_1^\dagger, \dots, g_m^\dagger, d^\dagger)$, using the transformation Ψ in (4) for suitable numbers c_1, c_2, c_3 and c_4. Simple calculations yield the following equations for each identification scheme reproducing $I(0)$, $I(1)$ and $I(2)$: The two first constraints (8) are common for each case and yield the equations

$$c_1 = -\log\left\{\sum_{i=1}^{m}\exp(a_i^\dagger + (i-1)c_4\right\} \quad (13)$$

$$c_2 = -\log\left\{\sum_{j=1}^{m}\exp(b_j^\dagger + (j-1)c_4\right\}. \quad (14)$$

The last two equations are what defines the forecasting strategy. For the $I(1)$ case, we have that

$$c_3 = -g_1^\dagger \quad (15)$$

$$c_4 = (g_m^\dagger + c_3)/(m-1). \quad (16)$$

For the $I(2)$ case,

$$c_3 = (m-1)(g_m^\dagger - g_{m-1}^\dagger) - g_m^\dagger \quad (17)$$

$$c_4 = (g_m^\dagger + c_3)/(m-1), \quad (18)$$

and for the $I(0)$ case,

$$c_3 = c_4(m-1)/2 - \sum_{t=1}^{m} g_t^\dagger/m \quad (19)$$

$$c_4 = [12/(m(m-1)(m+1))]\sum_{t=1}^{m} t g_t^\dagger - [6/(m(m-1))]\sum_{t=1}^{m} g_t^\dagger. \quad (20)$$

The above derivations show that the so-called $I(0)$ and $I(1)$ methods of Kuang et al. (2011) use identification constraints on the entire past when modelling and estimating the slope of the calendar effect, while the identification constraint of the $I(2)$ method only concerns the most recent past. As a consequence, the slope produced by the $I(0)$ and $I(1)$ methods might be too biased and that of the $I(2)$ method too volatile. It seems appropriate to develop a methodology in the discrete setting which allows the identification to be based on some optimal amount of the past. However, the concrete full derivation of such a methodology in the discrete case is beyond the scope of this paper on continuous in-sample forecasting.

References

Kuang, D., Nielsen, B. and Nielsen, J. P. (2008a). Identification of the age-period-cohort model and the extended chain ladder model. *Biometrika* **95**, 979–986.

Kuang, D., Nielsen, B. and Nielsen, J. P. (2008b). Forecasting with the age-period-cohort model and the extended chain-ladder model. *Biometrika* **95**, 987–991.

Kuang, D., Nielsen, B. and Nielsen, J. P. (2009). Chain Ladder as Maximum Likelihood Revisited. *Annals of Actuarial Science* **4**(1), 105–121.

Kuang, D., Nielsen, B. and Nielsen, J. P. (2011). Forecasting in an extended chain-ladder-type model. *Journal of Risk and Insurance* **78**, 345–359.

Mammen, E., Martínez-Miranda, M. D. and Nielsen, J. (2015). In-sample forecasting applied to reserving and mesothelioma mortality. *Insurance: Mathematics and Economics* **61**, 76–86.

Martínez-Miranda, M. D., Nielsen, J. P Sperlich, S. and Verrall, R. (2013). Continuous Chain Ladder: Reformulating and generalising a classical insurance problem. *Expert Systems with Applications* **40**, 5588–5603.