Biometrika (2008), **95**, 4, *pp*. 987–991
doi: 10.1093/biomet/asn038
Advance Access publication 22 September 2008

Printed in Great Britain

Forecasting with the age-period-cohort model and the extended chain-ladder model

BY D. KUANG

Department of Statistics, University of Oxford, Oxford OX1 3TG, U.K.
di.kuang@some.ox.ac.uk

B. NIELSEN

Nuffield College, Oxford OX1 1NF, U.K.
bent.nielsen@nuffield.ox.ac.uk

AND J. P. NIELSEN

Cass Business School, City University London, 106 Bunhill Row, London EC1Y 8TZ, U.K.
Jens.Nielsen.1@city.ac.uk

SUMMARY

We consider forecasting from age-period-cohort models, as well as from the extended chain-ladder model. The parameters of these models are known only to be identified up to linear trends. Forecasts from such models may therefore depend on arbitrary linear trends. A condition for invariant forecasts is proposed. A number of standard forecast models are analysed.

Some key words: Age-period-cohort model; Chain-ladder model; Forecasting; Identification.

1. INTRODUCTION

Consider the age-period-cohort model used in epidemiology and demography. It describes the logarithm of the mortality in an additive form, involving three interlinked time scales,

$$\mu_{ij} = \alpha_i + \beta_j + \gamma_{i+j-1} + \delta, \tag{1}$$

where i is the cohort, j is the age, and $i + j - 1$ is the period. Thus, α_i is a cohort effect, β_j is an age effect, γ_{i+j-1} is a period effect, while δ determines the overall level. The indices i and j vary bivariately in an index set I, which, for simplicity is assumed to be triangular and given by $i, j = 1, \ldots, k$ so $i + j - 1 \leqslant k$. Given estimates $\widehat{\alpha}_i, \widehat{\beta}_j, \widehat{\gamma}_{i+j-1}$ and $\widehat{\delta}$ it is of interest to study properties of forecasts of μ_{ij} for the triangle J given by $i, j = 1, \ldots, k$ so $i + j - 1 > k$. To construct these forecasts it is necessary to extrapolate the γ-parameters, whereas the α-, β- and δ-parameters are readily available. It has long been appreciated that the parameterization in terms of $\alpha_i, \beta_j, \gamma_{i+j-1}$ and δ is not identified. Even when omitting the parameter δ or when letting $\alpha_1 = \beta_1 = \gamma_1 = 0$ the identification problem remains. In this paper, we discuss to what extent the identification problem has bearing on the forecasts.

Carstensen (2007) gave a group theoretic description of the identification problem showing that linear trends can be added to and subtracted from $\alpha_i, \beta_j, \gamma_{i+j-1}$ and δ, so that their sum μ_{ij} given in (1) is unchanged. Earlier, Clayton & Schifflers (1987) had suggested that the ratios of relative risks are identifiable. On a logarithmic scale, these ratios translate into second differences. Recently, Kuang et al. (2008) discussed the identification problem and suggested a canonical parameterization involving these second differences, which has a bijective correspondence with μ_{ij}, for all $(i, j) \in I$. Starting from these descriptions, this paper provides a simple condition, which ensures that the forecasts for the triangle J are invariant to these trends.

The forecasting problem has previously been studied by Berzuini & Clayton (1994). Building on the invariance of the second differences they proposed a latent model for the period effect, γ, in which the second differences are independent and identically distributed, and suggested estimation of the parameters α, β and δ jointly with the parameters of the latent model. With this approach, the γ-parameters can be extrapolated using the latent model in such a way that the identification problem is avoided. In this paper, we split this procedure into two stages. In the first stage, the parameters α, β, γ and δ are estimated. In the second stage, which is analysed here, a forecasting model is fitted to the estimated γ-parameters. This procedure gives more flexibility in formulating a forecasting model for the γ-parameters and thereby exploits experience on forecasting of nonstationary time series (Clements & Hendry, 1999).

The presented condition for invariant forecasting allows forecast models, which are based directly on the second differences of the canonical parameter. This relates to the proposal by Berzuini & Clayton (1994). However, it is also possible to construct invariant forecasts that are motivated by a forecasting model that appears to involve a particular identification of α_i, β_j, γ_{i+j-1} and δ. A number of examples are presented here. In particular, if the γ-parameters are extrapolated using a linear trend or a random walk with a drift then the forecast for μ_{ij} in the triangle J are invariant. Extrapolation using a constant level or a random walk without a drift will, in contrast, lead to non-invariant forecasts.

Kuang et al. (2008) gave a brief introduction to the typical applications for these forecast methods. In particular, age-period-cohort studies have been discussed by Keiding (1990), whereas the extended chain-ladder models used in non-life insurance were introduced by Zehnwirth (1994) and Barnett & Zehnwirth (2000) as an extention of the classical chain-ladder model discussed by for instance England & Verrall (2002).

In such applications two types of generalizations may be needed. First, the triangular index set I may have a more general form than the triangular form discussed here. In many cases it would be a generalized trapezoid (Kuang et al., 2008). Second, when forecasting outside the triangular set J it will be necessary also to extrapolate either the cohort effect, α_i, or the age effect, β_j, or both. Generalizations of these types are application specific, but would be covered by extending the arguments of this paper.

2. IDENTIFICATION

The parameters of (1) are

$$\theta = (\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_k, \gamma_1, \ldots, \gamma_k, \delta) \in R^{3k+1}.$$

As pointed out by Carstensen (2007), linear trends in α_i, β_j and γ_{i+j-1} can be added without changing the value of μ_{ij}. This can be expressed in terms of the group

$$g : \begin{pmatrix} \alpha_i \\ \beta_j \\ \gamma_k \\ \delta \end{pmatrix} \mapsto \begin{Bmatrix} \alpha_i + a + (i-1)d \\ \beta_j + b + (j-1)d \\ \gamma_{i+j-1} + c - (i+j-2)d \\ \delta - a - b - c \end{Bmatrix},$$

where a, b, c and d are arbitrary constants. The parameter μ is a function of θ, which is invariant to g; that is, $\mu(\theta) = \mu\{g(\theta)\}$.

Kuang et al. (2008) analysed this problem further and found the representation

$$\mu_{ij} = \mu_{11} + (i-1)(\mu_{21} - \mu_{11}) + (j-1)(\mu_{12} - \mu_{11}) + a_{ij},$$

for all $i, j \in I$, where

$$a_{ij} = \sum_{t=3}^{i} \sum_{s=3}^{t} \Delta^2 \alpha_s + \sum_{t=3}^{j} \sum_{s=3}^{t} \Delta^2 \beta_s + \sum_{t=3}^{i+j-1} \sum_{s=3}^{t} \Delta^2 \gamma_s,$$

$\Delta\alpha_i = \alpha_i - \alpha_{i-1}$ and $\Delta^2\alpha_i = \Delta\alpha_i - \Delta\alpha_{i-1}$. Their Theorem 1 shows that the canonical parameter vector

$$\xi = \left(\mu_{11}, \mu_{21}, \mu_{12}, \Delta^2\alpha_3, \ldots, \Delta^2\alpha_k, \Delta^2\beta_3, \ldots, \Delta^2\beta_k, \Delta^2\gamma_3, \ldots, \Delta^2\gamma_k\right) \in R^{3k-3}$$

gives a unique parameterization of μ, so that for $\xi^\dagger \neq \xi^{\dagger\dagger}$ then $\mu(\xi^\dagger) \neq \mu(\xi^{\dagger\dagger})$. The group g is maximal, so that $\theta^\dagger = g(\theta^{\dagger\dagger})$ if and only if $\xi(\theta^\dagger) = \xi(\theta^{\dagger\dagger})$. Invariance properties can then be investigated using g.

3. Forecasting

Suppose now that an estimate $\widehat{\theta}$ is available for a particular identification scheme for the original parameters θ. The aim is to forecast $\mu_{i,j}$ for some $(i, j) \in J$. The period coordinate for this point is $k + h = i + j - 1$, so an h-step ahead forecast is needed for the period factor, $\gamma_{k+h} = \gamma_{i+j-1}$. The overall forecast is then $\widetilde{\mu}_{i,j}(\widehat{\theta}) = \widehat{\alpha}_i + \widehat{\beta}_j + \widetilde{\gamma}_{i+j-1}(\widehat{\gamma}) + \widehat{\delta}$, where $\widetilde{\gamma}_{i+j-1}(\widehat{\gamma})$ is a forecast itself constructed by extrapolation from $\widehat{\gamma} = (\widehat{\gamma}_1, \ldots, \widehat{\gamma}_k)$. The question whether the forecast $\widetilde{\mu}_{i,j}(\widehat{\theta})$ depends on the chosen identification scheme for θ can be addressed as follows. Applying the group g to $\widehat{\theta}$ results in the forecast $\widetilde{\mu}_{i,j}\{g(\widehat{\theta})\} = \{\widehat{\alpha}_i + a + (i-1)d\} + \{\widehat{\beta}_j + b + (j-1)d\} + \widetilde{\gamma}_{i+j-1}\{g(\widehat{\gamma})\} + (\widehat{\delta} - a - b - c)$. The forecast is invariant to the group g if and only if $\widetilde{\mu}_{i,j}(\widehat{\theta}) = \widetilde{\mu}_{i,j}\{g(\widehat{\theta})\}$, so a condition for invariance is now easily derived.

Theorem 1. *The forecast $\widetilde{\mu}_{i,j}$ for $(i, j) \in J$ is invariant to the group g if and only if, for $k + h = i + j - 1$ and arbitrary $c, d \in R$,*

$$\widetilde{\gamma}_{k+h}\{g(\widehat{\gamma})\} = \widetilde{\gamma}_{k+h}(\widehat{\gamma}) + c - (k + h - 1)d. \tag{2}$$

The condition (2) for invariance of the forecast of $\mu_{i,j}$ allows the forecast of $\gamma_{i+j-1} = \gamma_{k+h}$ to be non-invariant as long as the arbitrarily chosen linear trend appears in an additive fashion. When forecasting outside the triangle J two, or even three, of the factors α_i, β_j and γ_{i+j-1} need to be extrapolated. The argument can be generalized to that situation.

The linear structure of condition (2) implies that forecasts for γ_{h+k} need to have a structure that involves the canonical parameter ξ. This structure is summarized in the next theorem, which is proved in the Appendix.

Theorem 2. *The forecast $\widetilde{\gamma}_{k+h}(\widehat{\gamma})$ satisfies the condition (2) if and only if, for some function f, it is given by*

$$\widetilde{\gamma}_{k+h}(\widehat{\gamma}) = \widehat{\gamma}_k + h\Delta\widehat{\gamma}_k + f\left(\Delta^2\widehat{\gamma}_3, \ldots, \Delta^2\widehat{\gamma}_k\right). \tag{3}$$

In order to interpret expression (3), note the telescopic formulas

$$\gamma_{k+h} = \gamma_k + \sum_{t=1}^{h} \Delta\gamma_{k+t}, \qquad \Delta\gamma_{k+t} = \Delta\gamma_k + \sum_{s=1}^{t} \Delta^2\gamma_{k+s}.$$

Inserting the second expression in the first and noting that estimates are available for γ_k and $\Delta\gamma_k$, only the $\Delta^2\gamma_{k+s}$-terms need to be forecasted implies

$$\widetilde{\gamma}_{k+h}(\widehat{\gamma}) = \widehat{\gamma}_k + h\Delta\widehat{\gamma}_k + \sum_{t=1}^{h}\sum_{s=1}^{t} \Delta^2\widetilde{\gamma}_{k+s}.$$

Theorem 2 therefore shows that the forecasts for $\sum_{t=1}^{h}\sum_{s=1}^{t} \Delta^2\widetilde{\gamma}_{k+s}$ should be based exclusively on the second-differences $\Delta^2\widehat{\gamma}_\ell$, which are part of the canonical parameter ξ.

In applications, the question is then how to choose the forecasts $\widetilde{\Delta^2\gamma}_{k+t}$. A regression model only involving the second-differences, $\Delta^2\widehat{\gamma}_\ell$, would clearly suffice. However, time series models that appear to involve first-differences, $\Delta\widehat{\gamma}_\ell$, or even levels, $\widehat{\gamma}_\ell$, can also be used as long as they eliminate any linear trend behaviour. It is interesting to study a few examples.

Table 1. *Invariance properties of various forecasting models*

Order of integration	Invariant forecasts	Non-invariant forecasts
I(0)	$x_t = \nu_c + \nu_l t + \varepsilon_t$	$x_t = \nu_c + \varepsilon_t$
	$x_t = \rho x_{t-1} + \nu_c + \nu_l t + \varepsilon_t$	$x_t = \rho x_{t-1} + \nu_c + \varepsilon_t$
I(1)	$\Delta x_t = \nu_c + \varepsilon_t$	$\Delta x_t = \varepsilon_t$
I(2)	$\Delta^2 x_t = \varepsilon_t$	
	$\Delta^2 x_t = \rho \Delta^2 x_{t-1} + \varepsilon_t$	

Consider first a simple forecasting model of the type $x_t = \nu + \varepsilon_t$, which will not produce invariant forecasts. Estimating ν by $\widehat{\nu} = k^{-1} \sum_{i=1}^{k} \widehat{\gamma}_i$ gives a point forecast of the form $\widetilde{\gamma}_{k+h}(\widehat{\gamma}) = \widehat{\nu}$. The formula (2) does not hold in this case since

$$\begin{aligned}\widetilde{\gamma}_{k+h}\{g(\widehat{\gamma})\} &= k^{-1} \sum_{i=1}^{k} \{\widehat{\gamma}_i + c - (i-1)d\} = k^{-1} \sum_{i=1}^{k} \widehat{\gamma}_i + c - (k-1)d/2 \\ &\neq k^{-1} \sum_{i=1}^{k} \widehat{\gamma}_i + c - (k+h-1)d = \widetilde{\gamma}_{k+h}(\widehat{\gamma}) + c - (k+h-1)d.\end{aligned}$$

This shows that the forecast $\widetilde{\mu}_{i,j}(\widehat{\theta})$ will have a linear trend component $\{(k-1)/2 + h\}d = \{(k+1)/2 + i + j - 1\}d$, depending on the arbitrarily chosen slope d. The forecasting model $x_t = \nu_c + \nu_l t + \varepsilon_t$ will, in contrast, produce invariant forecasts. This model is, however, tedious to analyse. In the following, two random walk models are therefore analysed to show the mechanics of the argument in greater detail.

Consider first a simple random walk forecasting model of the type $x_t = x_{t-1} + \varepsilon_t$. The point forecast is $\widetilde{\gamma}_{k+h}(\widehat{\gamma}) = \widehat{\gamma}_k$, which is not of the form (3).

Invariant forecasts can, however, be achieved from a random walk forecasting model with a drift where $x_t = x_{t-1} + \nu + \varepsilon_t$. Estimating ν by $\widehat{\nu} = (k-1)^{-1} \sum_{j=2}^{k} \Delta\widehat{\gamma}_j$, gives a point forecast of $\widetilde{\gamma}_{k+h}(\widehat{\gamma}) = \widehat{\gamma}_k + h\widehat{\nu}$. This is shown to be of the form (3) by noting that $\Delta\widehat{\gamma}_j = \Delta\widehat{\gamma}_k - \sum_{\ell=j+1}^{k} \Delta^2\widehat{\gamma}_\ell$, which implies that $\widehat{\nu} = \Delta\widehat{\gamma}_k - (k-1)^{-1} \sum_{j=2}^{k} \sum_{\ell=j+1}^{k} \Delta^2\widehat{\gamma}_\ell$. Alternatively, the condition (2) can be proved directly by noting that

$$\begin{aligned}\widetilde{\gamma}_{k+h}\{g(\widehat{\gamma})\} &= \{\widehat{\gamma}_k + c - (k-1)d\} + (k-1)^{-1} h \sum_{j=2}^{k} (\Delta\widehat{\gamma}_j - d) \\ &= \widetilde{\gamma}_{k+h}(\widehat{\gamma}) + c - (k+h-1)d.\end{aligned}$$

Density forecasts can also be analysed with these theorems. Suppose, in the random walk model with intercept, the innovations ε_t are independently, normally distributed with mean zero and variance σ^2. The innovation variance is then estimated by $\widehat{\sigma}^2 = (k-1)^{-1} \sum_{j=2}^{k} (\Delta\widehat{\gamma}_j - \widehat{\nu})^2$, which is a function of the second differences. The density forecast $\widetilde{\gamma}_{k+h}(\widehat{\gamma}) = \widehat{\gamma}_k + h\widehat{\nu} + \widetilde{\varepsilon}_{k:h}$, where $\widetilde{\varepsilon}_{k:h}$ is normally distributed with mean zero and variance $h\widehat{\sigma}^2$, therefore satisfies (3).

Turning to forecasting models based on the levels, $\widehat{\gamma}_\ell$, the same type of results can be found. A model with deterministic regressors needs to involve a linear trend to eliminate the arbitrary linear trend in the levels. Likewise, autoregressions should include linear trends. The results are summarized in Table 1.

Building on the econometrics literature the models in Table 1 are described in terms of integrated processes of order s, denoted $I(s)$. This notation indicates that the processes need to be differenced s times to achieve stationarity. Clements & Hendry (1999, § 5) discuss the merits of the different forecasting methods. The $I(0)$ methods tend to be preferable if they describe the sample variation in-sample and no structural changes are expected out-of-sample, whereas the higher order integrated methods tend to be more robust to structural changes out-of-sample.

ACKNOWLEDGEMENT

The second author received financial support from the U.K. Economic & Social Research Council.

APPENDIX

Proof of Theorem 2

Any function of $\widehat{\gamma}$ satisfying (2) can be written in the form

$$\widetilde{\gamma}_{k+h}(\widehat{\gamma}) = \widehat{\gamma}_k + h\Delta\widehat{\gamma}_k + F\left(\widehat{\gamma}_k, \Delta\widehat{\gamma}_k, \Delta^2\widehat{\gamma}_3, \ldots, \Delta^2\widehat{\gamma}_k\right),$$

for some function F. Since $g(\widehat{\gamma}_k + h\Delta\widehat{\gamma}_k) = \widehat{\gamma}_k + h\Delta\widehat{\gamma}_k + c - (k+h-1)d$, the condition (2) holds if and only if F is invariant to g, that is, for scalar x and y, and a $(k-2)$-vector z, then $F\{x + c - d(k-1), y - d, z\} = F(x, y, z)$ for all c and d. If $F(x, y, z) = f(z)$ as required by (3) this clearly holds. Conversely, setting first $d = 0$ it is seen that the function must be constant in its first argument, and setting then $c = 0$ it is seen that function must also be constant in its second argument. Thus $F(x, y, z) = f(z)$.

REFERENCES

BARNETT, G. & ZEHNWIRTH, B. (2000). Best estimates for reserves. *Proc. Casualty Actuar. Soc.* **87**, 245–321.
BERZUINI, C. & CLAYTON, D. (1994). Bayesian analysis of survival on multiple time scales. *Statist. Med.* **13**, 823–38.
CARSTENSEN, B. (2007). Age-period-cohort models for the lexis diagram. *Statist. Med.* **26**, 3018–45.
CLAYTON, D. & SCHIFFLERS, E. (1987). Models for temporal variation in cancer rates. II: Age-period-cohort models. *Statist. Med.* **6**, 469–81.
CLEMENTS, M. P. & HENDRY, D. F. (1999). *Forecasting Non-Stationary Time Series*. Cambridge, MA: MIT Press.
ENGLAND, P. D. & VERRALL, R. J. (2002). Stochastic claims reserving in general insurance. *Br. Actuar. J.* **8**, 519–44.
KEIDING, N. (1990). Statistical inference in the lexis diagram. *Phil. Trans. R. Soc.* A **332**, 487–509.
KUANG, D., NIELSEN, B. & NIELSEN, J. P. (2008). Identification of the age-period-cohort model and the extended chain-ladder model. *Biometrika* 10.1093/biomet/asn026.
ZEHNWIRTH, B. (1994). *Probabilistic Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals, and Risk-Based Capital*, 447–605. Casualty Actuarial Society Forum, 1994 Spring Forum.

[*Received January* 2008. *Revised May* 2008]