Pruscha:

Residual and forecast methods in time series models with covariates

Sonderforschungsbereich 386, Paper 33 (1996)

RESIDUAL AND FORECAST METHODS IN TIME SERIES MODELS WITH COVARIATES

BY HELMUT PRUSCHA

Mathematical Institute, University of Munich, Theresienstr. 39
D-80333 Munich, Germany

SUMMARY

We are dealing with time series which are measured on an arbitrary scale, e.g. on a categorical or ordinal scale, and which are recorded together with time varying covariates. The conditional expectations are modelled as a regression model, its parameters are estimated via likelihood- or quasi-likelihood-approach. Our main concern are diagnostic methods and forecasting procedures for such time series models. Diagnostics are based on (partial) residual measures as well as on (partial) residual variables; l-step predictors are gained by an approximation formula for conditional expectations. The various methods proposed are illustrated by two different data sets.

Some key words: Categorical time series; Conditional regression models; Forecast methods; Ordinal responses; Partial residuals.

1 INTRODUCTION

We are concerned with time series data $(Z_t, Y_t), t \geq 1$, where Y_t is a response variable measured on a scale which is not necessarily metrical, and Z_t is a vector of covariates. The evolution of the Y_t-process is assumed to be driven by its own history as well as by the covariate process Z_t. The conditional expectation of Y_t is modelled in the form $h(\eta_t)$, where h is a suitable response function and η_t a regression term containing the actual covariates Z_t as well as former observations (sec. 2 and 3). Such models were already investigated by Kaufmann (1987), Zeger and Qaqish (1988), Pruscha (1993), Lee (1991, 1994). Their statistical analysis is based on (quasi-) likelihood methods (sec. 4). The main concern of the present paper is to carry over two classic time series topics to these more general models. The first is the partial residual analysis (sec. 5) which can be used to assess the relevance of subsets of covariates as well as to remove the influence of covariate subsets (see Fuller, 1976, sec. 9.3 for the latter). Our methods are inspired by linear model theory and can be found in the special case of a binary logistic model in Landwehr et al (1984) and of a cumulative logistic model in Pruscha (1994). The second topic deals with forecasting future outcomes $Y_{T+1}, Y_{T+2}, \ldots$, if the process has been observed up to time T. Here, some recent work on the cumulative model (Pruscha, 1995) is continued and generalized: We will arrive at forecast formulas covering the well-known recursive equations of Box and Jenkins (1976) as well as the l-step transition laws for finite Markov chains (sec. 6).

In the following we assume that the response variable is m-dimensional,

$$Y_t = (Y_{t,1}, \ldots, Y_{t,m})^{\mathrm{T}} \ ,$$

and that the covariates Z_t form an r-dimensional vector process. Formally the continuous case, where the $Y_{t,j}$ are metrically scaled, is also covered. Most emphasis, however, lies on the discrete case, where Y_t is e.g. $M_m(1, \pi_t)$ distributed, i.e. multinomially distributed with parameters 1 and π_t,

$$\pi_t = (\pi_{t,1}, \ldots, \pi_{t,m})^{\mathrm{T}}, \quad \pi_{t,j} > 0, \quad \sum_{j=1}^{m} \pi_{t,j} < 1.$$

Within this case special attention is given to an ordinally scaled Y_t, where it is useful to introduce cumulative probabilities

$$\pi_{t(j)} = \pi_{t,1} + \ldots + \pi_{t,j} \ .$$

Throughout we will use subscripts in parenthesis to indicate an increasing order.

2 MODELLING

The collection of variables, observed earlier than Y_t, is denoted by

$$\mathcal{H}_t = (Z_1, Y_1, \ldots, Z_{t-1}, Y_{t-1}, Z_t) \ .$$

The m-dimensional conditional expectation vector

$$\mu_t = E(Y_t | \mathcal{H}_t)$$

is modelled in the form

$$\mu_t = h(\eta_t), \quad t = 1, 2, \ldots, \tag{1}$$

where $h : \mathbb{R}^m \to \mathbb{R}^m$ is an appropriate response function and $\eta_t = (\eta_{t,1}, \ldots, \eta_{t,m})^{\mathrm{T}}$ the linear regression term of the model. This term is written as

$$\eta_t = X_t \vartheta \ , \tag{2}$$

where $\vartheta \in \mathbb{R}^p$ comprises the unknown parameters and the entries of the $m \times p$ -matrix X_t are functions of $\mathcal{H}_t$.

Typically, the dependence on the last response Y_{t-1} [or on the last responses $Y_{t-1}, \ldots, Y_{t-k}$] and on the present covariates Z_t are separated in (2), leading to

$$\eta_{t,j} = \alpha_j + \gamma^{\mathrm{T}} H_{t-1,j} + \lambda^{\mathrm{T}} \Lambda(Y_{t-1})_j + \beta^{\mathrm{T}} Z_t \quad . \tag{3}$$

Here, for each $j = 1, \ldots, m$, α_j is an intercept term, $H_{t-1,j}$ is an $s \times 1$ -vector, the components of which being functions of $\mathcal{H}_{t-1}$, and $\Lambda(y)_j$ is a $q \times 1$ -vector expressing the dependence on the last response $y = Y_{t-1}$ [the term $\lambda^{\mathrm{T}} \Lambda(Y_{t-1})$ can be expanded to $\lambda^{\mathrm{T}} \Lambda(Y_{t-1}, \ldots, Y_{t-k})$ or to $\sum_{i=1}^{k} \lambda_i^{\mathrm{T}} \Lambda(Y_{t-i})$]. For model (1) - (3), the vector $\vartheta \in \mathbb{R}^p$, where

$$\vartheta^{\mathrm{T}} = (\alpha^{\mathrm{T}}, \gamma^{\mathrm{T}}, \lambda^{\mathrm{T}}, \beta^{\mathrm{T}}), \quad \alpha \in \mathbb{R}^m, \quad \gamma \in \mathbb{R}^s, \quad \lambda \in \mathbb{R}^q, \quad \beta \in \mathbb{R}^r,$$

is the unknown parameter vector of dimension $p = m+s+q+r$, and $H_0, \Lambda(Y_0)$ must be given in advance.

3 EXAMPLES

Let us consider examples, some of them already existing in the literature.

1. Y_t multinomially distributed

Let, conditionally on $\mathcal{H}_t$, the variable Y_t be $M_m(1, p_t)$ - distributed. In this case $p_t \equiv \mu_t$. We have $m+1$ alternatives, the last one being $Y_{t,m+1} = 1 - (Y_{t,1} + \ldots + Y_{t,m})$, occurring with probability

$$p_{t,m+1} = 1 - (p_{t,1} + \ldots + p_{t,m}) \quad .$$

A suitable response function is, e.g., $h_j(\eta) = \exp(\eta_j)/(1 + \sum_{k=1}^m \exp(\eta_k))$ which leads to a multivariate logistic regression model. Together with the general form (2) of the regression term multinomial models were analysed by Kaufmann (1987). Let us mention two specifications of $\Lambda(y)$ occurring in (3). In the first, so-called lagged variable dummies are used, i.e. we put

(i) $\quad q = m, \Lambda(y)_j = y$ for all j.

In the second , an $(m+1) \times (m+1)$ -transition matrix $P(\cdot, \cdot)$ is employed. Letting $y_{m+1} = 1 - (y_1 + \ldots + y_m)$ and $w = \sum_{j=1}^{m+1} j y_j$, we put

(ii) $\quad q = m+1, \Lambda(y)_{ij} = P(i,j)$, if $\quad w = i \; [\Lambda(y)_{ij} = 0 \; else]$,

such that $\lambda^{\mathrm{T}} \Lambda(y)_j = \lambda_w P(w,j)$, see Goettlein and Pruscha (1992) for an application. Assuming case (ii) and $h = id, \alpha = \gamma = \beta = 0, \lambda = 1$, model (1) - (3) describes a simple Markov chain with known transition matrix $P(\cdot, \cdot)$. A Markov chain with unknown transition matrix can be obtained from (1)-(3), if one allows $\lambda_j^{\mathrm{T}} \Lambda(y)_j$ instead of $\lambda^{\mathrm{T}} \Lambda(y)_j$ and takes case (i) above, but with $q = m+1$ instead of $q = m$.

2. Ordinally scaled response

Let Y_t as in 1. and define the $J = \{1, 2, \ldots, m+1\}$ -valued ordinal variable

$$W_t = \sum_{j=1}^{m+1} jY_{t,j} \quad . \tag{4}$$

Introducing $H_{(j)}(\eta) = h_1(\eta)+\ldots+h_j(\eta)$, we have from (1) for $j = 1, \ldots, m+1$

$$p_{t(j)} \equiv pr(W_t \leq j|\mathcal{H}_t) = H_{(j)}(\eta_t) \ . \tag{5}$$

If F denotes a (cumulative) distribution function we define a cumulative regression model (Mc Cullagh, 1980) by setting

$$H_{(j)}(\eta_t) = F(\eta_{t(j)}) \ . \tag{6}$$

Putting $s = 1$ and $H_{t-1} = p_{t-1}$ in (3), the regression term η_t can be written in the specific form

$$\eta_{t(j)} = \alpha_{(j)} + \gamma p_{t-1(j)} + \lambda^{\mathrm{T}}\Lambda(Y_{t-1})_{(j)} + \beta^{\mathrm{T}} Z_t \ . \tag{7}$$

Note that model (5) - (7) has an inherent recursive structure and a side condition on the parameters to ensure $\eta_{t(j)} - \eta_{t(j-1)} > 0$. Besides the specifications (i) and (ii) above we can here also choose

(iii) $\quad q = 1, \quad \Lambda(Y_{t-1})_j = W_{t-1}$ for all j ,

i.e. we can employ the lagged ordinal variables. Model (7) can easily extended to a higher order: instead of γp_{t-1} and $\lambda^{\mathrm{T}}\Lambda(Y_{t-1})$ sums of the form

$$\sum_{i=1}^{s_0} \gamma_i p_{t-i} \quad \text{and} \quad \sum_{i=1}^{q_0} \lambda_i^{\mathrm{T}}\Lambda(Y_{t-i})$$

can be used.

3. Metrically scaled responses

If the response vector $Y_t = (Y_{t,1}, \ldots, Y_{t,m})^{\mathrm{T}}$ consists of metrically scaled variables $Y_{t,j}$, the regression term η_t can be put as

$$\eta_t = \alpha + \sum_{i=1}^{s_0} \gamma_i \mu_{t-i} + \sum_{i=1}^{q_0} \lambda_i Y_{t-i} + 1\beta^{\mathrm{T}} Z_t \ , \tag{8}$$

with $1 = (1, \ldots, 1)^{\mathrm{T}} \in \mathrm{I\!R}^m$ and γ_i, λ_i scalars or $m \times m$ -matrices, see Zeger and Qaqish (1988) and Li (1994). Here we often have a function $h_0 : \mathrm{I\!R} \to \mathrm{I\!R}$, e.g. $h_0 = \exp$, such that

$$h(\eta) = (h_0(\eta_1), \ldots, h_0(\eta_m))^{\mathrm{T}} \ . \tag{9}$$

4 (QUASI-) LIKELIHOOD

For the following we will assume that the evolution of the process Z_t $t \geq 1$, is not influenced by the process Y_t , $t \geq 0$; precisely

$$pr(Z_{t+1} \in \cdot \ |\mathcal{H}_t, Y_t) = pr(Z_{t+1} \in \ \cdot \ |Z_1, \ldots, Z_t) \ . \tag{10}$$

Then a full likelihood approach is possible if (conditional) densities $f_t(y, \vartheta_t), y \in \mathrm{I\!R}^m$, of the conditional distributions $pr(Y_t \in \cdot \ |\mathcal{H}_t)$ are available, as in Ex. 1 and 2 above.

1. First, let us assume that the density f_t belongs to an m-parametric exponential family, i.e.

$$f_t(y, \vartheta_t) = \exp\left\{y^{\mathrm{T}} \vartheta_t - b(\vartheta_t)\right\} g_t(y). \tag{11}$$

Then we obtain $\mu_t(\vartheta_t) = b'(\vartheta_t)$, and, via (1), $\vartheta_t = u(\eta_t)$, with $u = (b')^{-1} \circ h$ and $\eta_t = X_t \vartheta$. Based on an observation $(Z_1, Y_1, \ldots, Z_n, Y_n)$, the log-likelihood function $l_n(\vartheta)$, the $p \times 1$ -score vector $U_n(\vartheta) = (d/d\vartheta) l_n(\vartheta)$ and the $p \times p$ -Hessian matrix $W_n(\vartheta) = (d^2/d\vartheta d\vartheta^{\mathrm{T}}) l_n(\vartheta)$ are given by

$$l_n(\vartheta) \ = \ \sum_{t=1}^{n} \{Y_t^{\mathrm{T}} \vartheta_t - b(\vartheta_t)\}, \ \vartheta_t = u(\eta_t), \ \eta_t = X_t \vartheta,$$

$$
\begin{aligned}
U_n(\vartheta) &= \sum_{t=1}^{n} X_t^{\mathrm{T}} D_t(\vartheta) \Sigma_t^{-1}(\vartheta)(Y_t - \mu_t(\vartheta)) \qquad (12) \\
W_n(\vartheta) &= R_n(\vartheta) - \sum_{t=1}^{n} X_t^{\mathrm{T}} D_t(\vartheta) \Sigma_t^{-1}(\vartheta) D_t^{\mathrm{T}}(\vartheta) X_t,
\end{aligned}
$$

where we have set $D_t(\vartheta) = (d/d\eta)h^{\mathrm{T}}(\eta_t), \Sigma_t(\vartheta) = (d^2/d\vartheta d\vartheta^{\mathrm{T}})b(\vartheta_t), \vartheta_t = u(\eta_t)$, $R_n(\vartheta)$ as in Fahrmeir and Tutz (1994, App. A1), and where we have neglected additive terms of $l_n(\vartheta)$ not depending on ϑ. In the case of an $M_m(1,p_t)$-distribution, where $p_t \equiv \mu_t$, Kaufmann (1987) gave conditions under which there exists a consistent m.l. estimate $\hat{\vartheta}_n$ for ϑ, which is asymptotically normally distributed in the sense that the distributional convergence

$$
\Gamma_n^{-1}(\hat{\vartheta}_n - \vartheta) \to N_p(0, V^{-1}(\vartheta)) \qquad (13)
$$

takes place for $n \to \infty$, where

$$
-V(\vartheta) = pr_\vartheta\text{-}\lim\ \Gamma_n^{\mathrm{T}} W_n(\hat{\vartheta}_n) \Gamma_n,
$$

if the latter limit exists for a sequence $\Gamma_n = \Gamma_n(\vartheta), n \geq 1$, of invertible norming matrices tending towards 0.

2. Secondly, let us assume an ordinally scaled response variable W_t as in Ex. 3.2 . Then $p_{t,j}(\vartheta) = pr_\vartheta(W_t = j|\mathcal{H}_t)$ and we can write

$$
\begin{aligned}
l_n(\vartheta) &= \sum_{t=1}^{n} \log p_{t,w_t}(\vartheta), \\
U_n(\vartheta) &= \sum_{t=1}^{n} u_{t,w_t}(\vartheta),\ u_{t,j}(\vartheta) = (d/d\vartheta)p_{t,j}(\vartheta)/p_{t,j}(\vartheta),
\end{aligned}
$$

and $W_n(\vartheta)$ similarly. Assuming the model (5)-(7) we can make use of recurrence relations of $p_t(\vartheta)$ and its derivatives (instead of exploiting the exponential family structure of the $M_m(1,p_t)$-distribution). Using a distance-diminishing theory for certain iterative function systems (Norman, 1972), (13) can be proved under the assumptions that $|sup_\eta F'(\eta)\ \gamma| < 1$ and that

$Z_t, t \geq 1$, forms a Markov process (of some order) with compact state space and with Lipschitz-bounded transition kernels (Pruscha, 1993).

3. Now we assume that, contrary to 1. and 2., a (conditional) density for $pr(Y_t \in \cdot \,|\mathcal{H}_t)$ cannot be given, but that (conditional) first and second moments

$$\mu_t(\vartheta) \;\equiv E_\vartheta(Y_t \mid \mathcal{H}_t) = h(\eta_t), \quad cov_\vartheta(Y_t \mid \mathcal{H}_t) = \Sigma_t(\eta_t), \quad \eta_t = X_t\vartheta,$$

can be specified. One still use $U_n(\vartheta) = 0$ as estimation equation with $U_n(\vartheta)$ as in (12). If the response variables $Y_t, t = 1, 2, \ldots$ are m-dimensional and independent, we are in the case of longitudinal data and the asymptotic covariance in (13) is of the form $V^{-1}(\vartheta)S(\vartheta)V^{-1}(\vartheta)$, where

$$S(\vartheta) = pr_\vartheta\text{-}lim\ \Gamma_n^{\mathrm{T}}\ \{\sum_{t=1}^{n} X_t^{\mathrm{T}} D_t(\hat{\vartheta}_n)\ \Sigma_t^{-1}(\hat{\vartheta}_n)\ cov(Y_t)\ \Sigma_t^{-1}(\hat{\vartheta}_n) D_t^{\mathrm{T}}(\hat{\vartheta}_n) X_t\}\ \Gamma_n\ ,$$

see Liang and Zeger (1986).

5 RESIDUAL ANALYSIS

5.1 Linear model residuals

The following derivation of global and partial residuals is inspired by linear model theory. In a linear regression model of the form

$$Y_t = \eta_t + e_t,\ \eta_t = X_t\vartheta, \quad t = 1, 2, \ldots,$$

global residuals are defined by

$$\hat{e}_t = Y_t - X_t\hat{\vartheta}, \tag{14}$$

$\hat{\vartheta}$ l.s. estimator for ϑ. Partial residuals from regression on X_1, where

$$X_t = (X_{t1}, X_{t2}) \quad \text{and} \quad \vartheta^{\mathrm{T}} = (\vartheta_1^{\mathrm{T}}, \vartheta_2^{\mathrm{T}})$$

are partitions, are given by

$$\hat{e}_t^{(par)} = \; Y_t - X_{t1}\hat{\vartheta}_1 = \; \hat{e}_t + X_{t2}\hat{\vartheta}_2. \tag{15}$$

Note that $\hat{e}_t^{(par)}$ can be gained from the "true partial residual" $e_t^{(par)} \equiv \Psi_t(Y_t) = \; Y_t - X_{t1}\vartheta_1$ by plugging in the estimator $\hat{\vartheta}_1$ for ϑ_1. Let the e_t's now $N(0, \sigma^2)$ -distributed, i.e. let $pr(Y_t \leq y) \; = \; \Phi_{\eta_t, \sigma^2}(y)$, with Φ_{μ, σ^2} being the $N(\mu, \sigma^2)$ -distribution function. Then $e_t^{(par)} \equiv \Psi_t(Y_t)$ can be gained from

$$pr(\Psi_t(Y_t) \leq y) = \; \Phi_{\eta_t^*, \sigma^2}(y), \quad \eta_t^* = \eta_t - X_{t1}\vartheta_1 = X_{t2}\vartheta_2. \tag{16}$$

In more general models like ours two different (partial) residual methods can be established (falling together in the normal linear case above)
a) residual measures for diagnostic purposes which will be defined in analogy with (15)
b) residual variables, which have values on the same scale as the Y_t-data and which can be submitted to further time series analysis. They will be gained in analogy with (16).

5.2 Partial residual measures

On the basis of the general model (1),(2) we build the global GLM-residuals (cf. Fahrmeir and Tutz, 1994, p. 98)

$$\hat{e}_t = \; D_t^{-T}(\hat{\vartheta}) \; (Y_t - \mu_t(\hat{\vartheta})), \tag{17}$$

with D_t as in 4.1. Using (17) we define partial residuals from regression on X_1 as in (15) by

$$\hat{e}_t^{(par)} = \hat{e}_t + X_{t2}\hat{\vartheta}_2. \tag{18}$$

Often it is desirable to summarize the m-components $\hat{e}_{t,j}^{(par)}$ of (18) into a one-dimensional measure. In the following two examples we will weight the components $\hat{e}_{t,j}$ of (17) by the diagonal elements $d_{t,j}(\hat{\vartheta}) \equiv \hat{d}_{t,j}$ of $D_t(\hat{\vartheta})$, i.e. we will build

$$\tilde{e}_t = \sum_{j=1}^{m} \hat{d}_{t,j}\ \hat{e}_{t,j}\ / \sum_{j=1}^{m} \hat{d}_{t,j}. \tag{19}$$

Then, defining $\tilde{X}_{t2}$ in analogy,

$$\hat{e}_t^{(par)} = \tilde{e}_t + \tilde{X}_{t2}\hat{\vartheta}_2 \tag{20}$$

is a one-dimensional partial residual measure.

Ex. 1. If we have a response function h of the form (9), then $\hat{d}_{t,j} = h_0'(\hat{\eta}_{t,j})$, and we obtain from (17) and (19)

$$\tilde{e}_t = \sum_{j=1}^{m}(Y_{t,j} - \mu_{t,j}(\hat{\vartheta}))\ /\ \sum_{j=1}^{m} \hat{d}_{t,j}\ ,$$

which is a kind of average of the residual components.

Ex. 2. In the case of an ordinal response with response function h of the form (6), i.e. $h_j(\eta) = F(\eta_{(j)}) - F(\eta_{(j-1)})$, we have $\hat{d}_{t,j} = F'(\hat{\eta}_{t(j)})$, and $D_t^{-T}(\vartheta)$ is a lower triangular matrix with j-th row

$$(1/F'(\eta_{t(j)}), \ldots, 1/F'(\eta_{t(j)}),\ 0, \ldots, 0)\ .$$

Hence the j-th component of (17) turns out to be

$$\hat{e}_{t,j} = (1/\hat{d}_{t,j}) \sum_{k=1}^{j} (Y_{t,k} - p_{t,k}(\hat{\vartheta})),$$

and (19) takes the form

$$\tilde{e}_t = -\ (1/\sum_j \hat{d}_{t,j})\ (W_t - m_t(\hat{\vartheta})), \tag{21}$$

with the $\{1, \ldots, m+1\}$ -valued ordinal variable $W_t = \sum_{j=1}^{m+1} jY_{t,j}$ as in (4) and with the mean category $m_t = \sum_{j=1}^{m+1} jp_{t,j}$ (see Pruscha, 1994, sec 3.1). Note that $W_t - m_t$ is a really ordinal residual.

Partial residual measures like (18), (20) are usually plotted over the regression term $X_{t2}\hat{\vartheta}_2$ or $\tilde{X}_{t2}\hat{\vartheta}_2$ to assess the significance of the regressor set X_2.

5.3 Partial residual variables

We restrict ourselves to univariate response variables with (up to parameter ϑ) known distribution function. Putting

$$\eta_t = X_t\vartheta, \quad \eta_t^* = \eta_t - X_{t1}\vartheta_1 = X_{t2}\vartheta_2,$$

we will call $\hat{Y}_t^* = \Psi_t(Y_t, \hat{\vartheta})$ partial residual variable (from regression on X_1), if we have for $Y_t^* = \Psi_t(Y_t, \vartheta)$, in analogy with (16),

$$pr_t(Y_t^* \leq y, \eta_t) \ = \ pr_t(Y_t \leq y, \eta_t^*), \text{ for all } \ y \in \mathbb{R}, \tag{22}$$

where $pr_t(\cdot) = pr(\cdot \mid \mathcal{H}_t)$, and $pr_t(\cdot\ , \eta)$ means that $pr_t(\cdot)$ is evaluated under regression term $\eta, \eta = \eta_t$ or $\eta = \eta_t^*$. Further we require

$$pr_t(Y_t = Y_t^*,\ \eta_t) = 1, \text{ if } \ \eta_t^* = \eta_t\ . \tag{23}$$

We will use the notation $F_t(y, \eta) = pr_t(Y_t \leq y, \eta)$ and will distinguish the cases where F_t is continuous in $y \in \mathbb{R}$ or not.

a) Let $F_t(y, \eta)$, $y \in \mathbb{R}$, continuous for each η . Put $F^{-1}(x, \eta) = inf\{y : F(y, \eta) \geq x\}$ and define $Y_t^* = \Psi_t(Y_t, \vartheta)$ via

$$\Psi_t(y,\vartheta) \;=\; F_t^{-1}(F_t(y,\eta_t),\;\eta_t^*)\;.$$

Then, since $F_t(Y_t,\eta_t)$ is $U[0,1]$-distributed under $pr_t(\cdot\;,\eta_t)$, one gets

$$\begin{aligned} pr_t(Y_t^* \le y,\;\eta_t) \;&=\; pr_t(F_t(Y_t,\eta_t) \le F_t(y,\eta_t^*),\;\eta_t) \\ &=\; F_t(y,\eta_t^*)\;, \end{aligned}$$

that is (22). Further, relation (23) is fulfilled, since $F_t^{-1}(F_t(Y_t,\eta_t),\;\eta_t) < Y_t$ occurs with $pr_t(\cdot\;,\;\eta_t)$-probability zero. In the example $F(y,\eta) = \Phi_{\eta,\sigma^2}(y)$, we have

$$\Psi_t(y,\vartheta) \;=\; \Phi^{-1}_{\eta_t^*,\sigma^2}(\Phi_{\eta_t,\sigma^2}(y)) \;=\; y - X_{t1}\vartheta_1,$$

such that $\Psi_t(y,\hat{\vartheta})$ as in (15) above.

b) Let $Y_t \in J = \{1,2,\ldots,m+1\}$ ordinally scaled (formerly denoted by W_t; for a purely categorical response, partial residual variables don't seem to be meaningful). Then we have $F_t(j,\eta) = pr_t(Y_t \le j,\;\eta),\;\; j \in J$, and (22) cannot be satisfied with a function $\Psi_t : J \to J$. Instead, we will define transition probabilities $\Psi_t(k|j),\;\; \sum_{k\in J}\Psi_t(k|j) = 1$, such that

$$\begin{gathered} Y_t^* = k \text{ is selected with probability } \Psi_t(k|j), \text{ if} \\ Y_t = j \text{ is the observed category} \end{gathered} \tag{24}$$

Let pr_t^* denote the (conditional) probability law governing the observed process as well as the random experiment (24). Then, making use of the model equation $pr_t(Y_t = j) \;=\; h_j(\eta_t)$, we have

$$pr_t^*(Y_t = j,\;Y_t^* = k) \;=\; h_j(\eta_t)\;\Psi_t(k|j)\;.$$

Putting $\Psi_t(j,k) \;=\; h_j(\eta_t)\Psi_t(k|j)$, we have to define Ψ_t in such a way that

$$\begin{array}{c} \sum_{k\in J}\Psi_t(j,k) = h_j(\eta_t) \\ \sum_{j\in J}\Psi_t(j,k) = h_k(\eta_t^*), \end{array} \tag{25}$$

for the second equation see (22), and that, with respect to (23),

$$\sum_{j\in J}\Psi_t(j,j) = 1, \text{ if } \eta_t = \eta_t^*. \tag{26}$$

To this end, let $H_{(j)}(\eta) \;=\; h_1(\eta) + \ldots + h_j(\eta)$ as in 3.2, $H_{(0)} = 0$, and introduce for $j,k \in J$ the intervals

$$I_t(j,k) = \;(H_{(k-1)}(\eta_t^*),\; H_{(k)}(\eta_t^*)] \;\cap\; (H_{(j-1)}(\eta_t),\; H_{(j)}(\eta_t)].$$

Then we define $\Psi_t(j,k)$ as the length of $I_t(j,k)$, i.e.

$$\Psi_t(j,k) \;=\; |I_t(j,k)|\;. \tag{27}$$

Since $\sum_k |I_t(j,k)| \;=\; H_{(j)}(\eta_t) - H_{(j-1)}(\eta_t)$ and $\sum_j |I_t(j,k)| \;=\; H_{(k)}(\eta_t^*) - H_{(k-1)}(\eta_t^*)$, equations (25) are fulfilled. Since $|I_t(j,j)| \;=\; H_{(j)}(\eta_t) - H_{(j-1)}(\eta_t)$, if $\eta_t = \eta_t^*$, (26) is also satisfied.

Partial residual variables Y_t^* are constructed with the intension to remove the influence of the regressor set X_1 on the response variable. A typical application is the removal of a trend in a time series.

6 FORECASTING METHODS

6.1 General method

Based on the model (1),(2) , i.e. $\mu_t \;\equiv\; E(Y_t \mid \mathcal{H}_t) \;=\; h(\eta_t),\;\; \eta_t = X_t\vartheta$, and given an observation up to time T , i.e.

$$\mathcal{F}_T = (Z_1, Y_1, \ldots, Z_T, Y_T) \;=\; (\mathcal{H}_T, Y_T),$$

we define the l-step predictor for μ_{T+l} by

$$\hat{\mu}_T(l) = E(\mu_{T+l} \mid \mathcal{F}_T), \quad l \geq 1 \ . \tag{28}$$

In the following we will write

$$E_T(\cdot) = E(\cdot \mid \mathcal{F}_T), \quad var_T(\cdot) = E_T(\ \cdot - E_T(\cdot))^2,$$

and we will use a similar definition for the conditional covariance matrix $cov_T(\cdot)$. Due to $\mathcal{F}_T \subset \mathcal{H}_{T+l}$ we also have

$$\hat{\mu}_T(l) = E_T(Y_{T+l}).$$

We will compute $\hat{\mu}_T(l)$ by the approximation $\check{\mu}_T(l) + B_T(l)$, where $\check{\mu}_T(l)$ is gained by interchanging conditional expectation and response function h , i.e. by

$$\check{\mu}_T(l) = h(\hat{\eta}_T(l)), \quad \hat{\eta}_T(l) = E_T(\eta_{T+l}) \ , \tag{29}$$

and $B_T(l)$ is a correction term, to be developped below. The l-step predictor $\hat{\eta}_T(l)$ is, contrary to $\hat{\mu}_T(l)$, computable for many models, especially for models with a recursive structure. For the derivation of such computation formulas we will distinguish between a continuous and a discrete response. In any case we have to assume that l-step predictors

$$\hat{Z}_T(l) = E_T(Z_{T+l})$$

are available for the covariate process $Z_t, t \geq 1$. This is the case, e.g., if Z_t forms an r-dimensional autoregressive process of some fixed order, see Brockwell and Davies (1987, sec. 11.4).

6.2 Recursive forecast formulas

a) Continuous response. For the m-variate model (8), with $s_0 = q_0$, one starts with

$$\hat{\eta}_T(1) = \alpha + \sum_{i=1}^{q_0} \gamma_i \mu_{T+1-i} + \sum_{i=1}^{q_0} \lambda_i Y_{T+1-i} + 1\beta^{\mathrm{T}} \hat{Z}_T(1) \tag{30}$$

and has, with $\hat{\mu}_T(1) = h(\hat{\eta}_T(1)) + B_T(1)$, the 2-step predictor

$$\hat{\eta}_T(2) = \alpha + \sum_{i=2}^{q_0} \gamma_i \mu_{T+2-i} + \sum_{i=2}^{q_0} \lambda_i Y_{T+2-i} + (\gamma_1 + \lambda_1)\hat{\mu}_T(1) + 1\beta^{\mathrm{T}} \hat{Z}_T(2)$$

and so on, until for $l > q_0$

$$\hat{\eta}_T(l) = \alpha + \sum_{i=1}^{q_0} (\gamma_i + \lambda_i)\hat{\mu}_T(l-i) + 1\beta^{\mathrm{T}} \hat{Z}_T(l). \tag{31}$$

The classical Box and Jenkins (1976, p. 129f) forecast formulas are contained as a special case. Indeed, one has to write $(\gamma + \lambda)\mu + \lambda(Y - \mu)$ instead of $\gamma\mu + \lambda Y$, and has to interpret $\mu_t, Y_t - \mu_t$ as their z_t and a_t, respectively; see also Lee (1994, p. 507).

b) Discrete response. We will use the extended version of model (7), i.e.

$$\eta_t = \alpha + \sum_{i=1}^{q_0} \gamma_i p_{t-i} + \sum_{i=1}^{q_0} \lambda_i^{\mathrm{T}} \Lambda(Y_{t-i}) + 1\beta^{\mathrm{T}} Z_t \ . \tag{32}$$

While $\hat{\eta}_T(1)$ is similar to (30), we have, with the probability vector $\hat{p}_T(1) = h(\hat{\eta}_T(1)) + B_T(1)$, the 2-step predictor

$$\hat{\eta}_T(2) = \alpha + \sum_{i=2}^{q_0} \gamma_i p_{T+2-i} + \sum_{i=2}^{q_0} \lambda_i^{\mathrm{T}} \Lambda(Y_{T+2-i}) + \gamma_1 \hat{p}_T(1) + \lambda_1^{\mathrm{T}} \hat{\Lambda}_T(1) + 1\beta^{\mathrm{T}} \hat{Z}_T(2),$$

where $\hat{\Lambda}_T(1) = \sum_{j=1}^{m+1} \hat{p}_{T,j}(1)\Lambda(e_j)$, $e_{m+1} = 0 \in \mathbb{R}^m$. Finally, for $l > q_0$ and with $\hat{\Lambda}_T(k) = \sum_{j=1}^{m+1} \hat{p}_{T,j}(k)\Lambda(e_j)$,

$$\hat{\eta}_t(l) = \alpha + \sum_{i=1}^{q_0} \gamma_i \hat{p}_T(l-i) + \sum_{i=1}^{q_0} \lambda_i^{\mathrm{T}} \hat{\Lambda}_T(l-i) + 1\beta^{\mathrm{T}} \hat{Z}_T(l) \ . \qquad (33)$$

In the special case of a simple Markov chain as in 3.1 we obtain $\hat{p}_{T,j}(l) = P^l(i,j)$, if $W_T \equiv \sum_{k=1}^{m+1} kY_{T,k} = i$, with P^l the l-th power of the known or estimated transition matrix.

6.3 Correction term

To give an estimate $\hat{B}_T(l)$ of the bias

$$B_T(l) = \hat{\mu}_T(l) - \check{\mu}_T(l)$$

produced by the approximation $E_T(h(\eta_{T+l})) \approx h(E_T(\eta_{T+l}))$, we start with expanding $h(\eta_{T+l})$ and $h(\hat{\eta}_T(l))$ at $\hat{\eta}_T(l-1)$, $\hat{\eta}_T(0) = \eta_T$, up to the order 2. To do this, we write h instead of h_j for some fixed j, assume twice continuous differentiability of h and introduce the abbreviations

$$\hat{\eta} = \hat{\eta}_T(l-1), \quad x_T(l) = \eta_{T+l} - \hat{\eta}, \quad \hat{x}_T(l) = \hat{\eta}_T(l) - \hat{\eta}.$$

Then, with remainder terms $R_T(l)$ etc.,

$$h(\eta_{T+l}) = h(\hat{\eta}) + x_T^{\mathrm{T}}(l)\ h'(\hat{\eta}) + \frac{1}{2}\ x_T^{\mathrm{T}}(l)\ h''(\hat{\eta})\ x_T(l) + R_T(l) \qquad (34)$$

$$h(\hat{\eta}_T(l)) = h(\hat{\eta}) + \hat{x}_T^{\mathrm{T}}(l)\ h'(\hat{\eta}) + \frac{1}{2}\ \hat{x}_T^{\mathrm{T}}(l)\ h''(\hat{\eta})\ \hat{x}_T(l) + \hat{R}_T(l). \qquad (35)$$

Applying conditional expectation to (34) we obtain

$$E_T(h(\eta_{T+l})) = h(\hat{\eta}) + \hat{x}_T^{\mathrm{T}}(l)\ h'(\hat{\eta}) + \frac{1}{2}\ E_T\{x_T^{\mathrm{T}}(l)\ h''(\hat{\eta})\ x_T(l)\} + \tilde{R}_T(l). \ (36)$$

Subtracting (35) from (36), neglecting remainder terms and introducing the omitted subscript j again, we arrive at

$$\begin{aligned}\hat{B}_{T,j}(l) &= \tfrac{1}{2}\ E_T\{\eta_{T+l}^{\mathrm{T}}\ h_j''(\hat{\eta})\ \eta_{T+l}\} - \tfrac{1}{2}\ \hat{\eta}_T^{\mathrm{T}}(l)\ h_j''(\hat{\eta})\ \hat{\eta}_T(l) \\ &= \tfrac{1}{2}\ E_T\{(\eta_{T+l} - \hat{\eta}_T(l))^{\mathrm{T}}\ h_j''(\hat{\eta})\ (\eta_{T+l} - \hat{\eta}_T(l))\}.\end{aligned}$$

Denoting the eigenvalues of the $m \times m$ -matrix $h_j''(\hat{\eta})$ by $\lambda_j^{(k)},\ k = 1, \ldots, m$, and writing

$$h_j''(\hat{\eta}) = A_j^{\mathrm{T}}(\hat{\eta})\ Diag(\lambda_j^{(k)})\ A_j(\hat{\eta}),$$

we finally obtain for $j = 1, \ldots, m$

$$\hat{B}_{T,j}(l) = \frac{1}{2}\sum_{k=1}^{m} \lambda_j^{(k)} A_j^{(k)T}(\hat{\eta})\ cov_T(\eta_{T+l})\ A_j^{(k)}(\hat{\eta}), \quad \hat{\eta} = \hat{\eta}_T(l-1) \tag{37}$$

where $A_j^{(k)T}$ stands for the k-th row of the $m \times m$ -matrix A_j and $cov_T(\eta)$ for the conditional covariance matrix of η.

For categorical responses, we have to define $\hat{B}_{T,m+1} = -\sum_{j=1}^{m} \hat{B}_{T,j}$ and (possibily) have to recalculate the $\hat{B}_{T,j}(l),\ \ j = 1, \ldots, m+1$, in such a way that the $\check{p}_{T,j}(l) + \hat{B}_{T,j}(l),\ \ j = 1, \ldots, m+1$, form a probability vector.

To use correction term (37) estimates for $cov_T(\eta_{T+l})$ must be available. Noting that $cov_T(\eta_{T+l}) = cov_T(\eta_{T+l} - \eta_T)$, one is led to build empirical variances/covariances of the vectors $\eta_{t+l} - \eta_t,\ \ t = 1, \ldots, T-l$.
Equation (37) will now be specialized in two examples.

Ex.1. For a response function h of the form (9) one obtains

$$\hat{B}_{T,j}(l) = \frac{1}{2}\ h_0''(\hat{\eta}_{T,j}(l-1))\ var_T(\eta_{T+l,j})\ .$$

Ex.2. In the case of an ordinal response with response function h as in (5), (6) we obtain for the bias $B_{T(j)}(l) = \hat{p}_{T(j)}(l) - \check{p}_{T(j)}(l)$ the estimate

$$\hat{B}_{T(j)}(l) = \frac{1}{2}\ F''(\hat{\eta}_{T(j)}(l-1))\ var_T(\eta_{T+l(j)})\ . \tag{38}$$

Here, we are also interested in the mean category

$$m_t = \sum_{j=1}^{m+1} j p_{t,j} = \sum_{j=0}^{m} (1 - p_{t(j)})$$

and its l-step prediction

$$\hat{m}_T(l) = \sum_{j=0}^{m} (1 - \hat{p}_{T(j)}(l))\ .$$

For the bias $B_{T,m}(l) = \hat{m}_T(l) - \check{m}_T(l)$ we obtain from (38) the estimate

$$\hat{B}_{T,m}(l) = -\frac{1}{2}\sum_{j=1}^{m} F''(\hat{\eta}_{T(j)}(l-1))\ var_T(\eta_{T+l(j)})\ . \tag{39}$$

6.4 Monte Carlo solution

Let us now assume that we are in the (most informative) situation, where the conditional probability laws $pr(Y_t \in \cdot \mid \mathcal{H}_t)$ and $pr(Z_{t+1} \in \cdot \mid Z_1, \ldots, Z_t)$ are explicitely given (see condition (10) above) and where estimates of all model parameters are available. Then one can gain the following computer-intensive solution $\overline{\mu}_T(l)$ for the l-step predictor $\hat{\mu}_T(l) = E_T(\mu_{T+l})$. Given the observation $(Z_1, Y_1, \ldots, Z_T, Y_T)$ the succeeding outcomes are simulated M times by the Monte-Carlo method. From

$$Z^{(i)}_{T+1}, Y^{(i)}_{T+1}, \ldots, Z^{(i)}_{T+l}, Y^{(i)}_{T+l}, \quad i = 1, \ldots, M, \tag{40}$$

one builds $\eta^{(i)}_{T+l},\ i = 1, \ldots, M$, and the average

$$\overline{\mu}_T(l) = \frac{1}{M}\sum_{i=1}^{M} \mu^{(i)}_{T+l}, \quad \mu^{(i)}_{T+l} = h(\eta^{(i)}_{T+l})\ ,$$

as the Monte-Carlo solution for $\hat{\mu}_T(l)$. The mean squared error

$$V_{T,\mu}(l) = E_T(\Delta_{T,\mu}(l)\ \Delta_{T,\mu}^{\mathrm{T}}(l)) = cov_T(\mu_{T+l}),$$

where $\Delta_{T,\mu}(l) = \mu_{T+l} - \hat{\mu}_T(l)$, can be estimated from (40) by

$$\hat{V}_{T,\mu}(l) = \frac{1}{M-1}\sum_{i=1}^{M}(\Delta_{T,\mu}^{(i)}(l)\ \Delta_{T,\mu}^{(i)\mathrm{T}}(l)),$$

where $\Delta_{T,\mu}^{(i)}(l) = \mu_{T+l}^{(i)} - \overline{\mu}_T(l)$.

Up to now there seems to exist no approach to estimate $\hat{V}_{T,\mu}(l)$ and related mean squared errors outside the Monte-Carlo method.

7 Applications

7.1 Data sets

We will use two different data sets for illustrating the various methods proposed above.
The first is a longitudinal data set on damages in beech, oak and pine trees, gathered by Dr. A. Goettlein, University of Bayreuth, during the last years in a forest district of the Spessart (Bavaria). The longitudinal structure of the data is determined by the observation period of 12 years (1983 - 1994) and by N sites (N = 80/25/14 sites with beech/oak/pine trees). The response variable W_t measures the percentage of leaves/needles lost on an ordinal scale of m+1=8 categories. For each site and each year t a vector Z_t of r=20 covariates were recorded concerning the trees (age, canopy), the site (gradient, height, exposition), the soil (moisture, pH-values) and the climate, see Goettlein and Pruscha (1992) and (1996) for detailed information. The parameters of the cumulative logistic regression model (5)-(7) were estimated by the m.l. method for each tree species separately. Concerning the Λ- function we made the special choice of lagged ordinal variables, see case (iii) in 3.2. Further we put $\gamma = 0$. The covariate process Z_t is assumed- as far as

forecasting methods are employed- to be driven by an AR(1)-equation.
The second data set concerns the aftershock series of the Friuli earthquake (May-Sept. 1976), which were placed at my disposal by Dr. H. Gebrande, University of Munich. The response variable W_t gives the number of shocks at day t, for $t = 1, \ldots, 115$ (corresponding to the period from 19^{th} May to 10^{th} Sept.), the covariates Z_t are the magnitude ML of the shocks (daily averages) and - in connection with trend analyses - the terms $t^{-1/2}, t^{-1/4}$. The cumulative logistic model (5)-(7) was applied, with the m+1=6 categories $0, \ldots, 5$ (i.e., instead w we took min(5,w) as response value), and with a preselected Λ-matrix. The first two, middle two and last two rows of the corresponding transition matrix $P(\cdot, \cdot)$, see case (ii) in 3.1, was chosen as $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0)$, $(0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0)$ and $(0, 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, respectively (actually, the F^{-1}-transformed cumulative probability vectors entered regression equation (7)).

7.2 Partial residuals

For the forest damage data we want to plot one-dimensional partial residual measures (20) over the regression term $X_{t2}\hat{\vartheta}_2$, where $\tilde{e}_t$ is calculated on the basis of a cumulative logistic model via formula (21). For the oak and pine tree, the partial residual plot for the regressor set X_2 = topography = (height, gradient, upper/lower part of slope) shows a clear upward trend (Fig. 1) and gives evidence for the significance of this covariates in the model equation. This is different with the beech tree, where the plot gives no hint to a relevance of the topography (in agreement with related test results, see Goettlein and Pruscha, 1996). The partial residual plot for $X_{2t} = W_{t-1}$, the lagged ordinal response variable (Fig. 2), reveals the strong dependence of the damage value W_t on the value W_{t-1} of the last year.

The method 5.3 b) of building ordinally scaled partial residual variables (from regression on X_1) is demonstrated for the Friuli earthquake data. The time-series plot of the number of shocks per day shows a decreasing tendency (see Fig. 3a). A trend function $a(t) = \alpha_1 t^{-1/2} + \alpha_2 t^{-1/4}$ was incorporated into the regression term η_t, and $\hat{a}(t)$ was plotted in the form trend(t) = $c_0 + c_1\hat{a}(t)$, with appropriate scaling rates c_0, c_1. Letting in the cumulative model with 6 categories, as described above, $X_{1t} = (t^{-1/2}, t^{-1/4})$ and $X_{2t} = (p_{t-1}, \Lambda(Y_{t-1}), \mathrm{ML}_t)$, we can calculate partial residual variables $\hat{Y}_t^*$ according to (24) and (27). The time-series plot of the $\hat{Y}_t^*$-values no longer reveals an obvious trend (Fig. 3b), in agreement with the test result that a trend component would no longer be a significant part of the regression term.

7.3 Forecasting

Fixing the observations of the Spessart data within the period 1983-1994 as known, we try to forecast the damage values for the years 1995-2002. That is, we put T=12, and we are interested in the l-step predictors $\hat{m}_T(l), l = 1, \ldots, 8$, of the mean damage category $m_t = \sum_{j=1}^{m+1} j p_{t,j}$, for each of the three tree species separately. The calculations of Fig. 4 were performed for each site $i = 1, \ldots, N$, followed by an average over the N sites of the species. On the basis of the cumulative logistic regression model with regression term $\eta_{t(j)} = \alpha_{(j)} + \lambda W_{t-1} + \beta^T Z_t$, the l-step prediction $\hat{m}_T(l)$ was computed by $\check{m}_T(l) + \hat{B}_{T,m}(l)$, with $\hat{B}_{T,m}(l)$ as in (39) and with

$$\check{m}_T(l) = \sum_{j=0}^{m} (1 - \check{p}_{T(j)}(l)), \;\; \check{p}_{T(j)}(l) = F(\hat{\eta}_{T(j)}(l))$$

$$\hat{\eta}_{T(j)}(l) = \alpha_{(j)} + \lambda \sum_{k=1}^{m+1} k \hat{p}_{T,k}(l-1) + \beta^T \hat{Z}_T(l).$$

The correction terms $\hat{B}_{T,m}(l)$ were calculated in two different ways: The term $var_T(\eta_{T+l})$ in (39) was estimated by empirical variances, as indicated in 6.3, and by the Monte-Carlo method 6.4, denoted by $\hat{B}_T^{(A)}(l)$ and $\hat{B}_T^{(M)}(l)$, respectively.

To come close to the correct forecast $E(m_{T+l}|\mathcal{F}_T)$, the forthcoming paths (40) were simulated M=600 times, assuming gaussian errors in the AR(1)-law of the covariate process. As in 6.4 averages $\overline{p}_{T(j)}(l)$ were built as well as $\overline{m}_T(l) = \sum_{j=0}^{m} (1 - \overline{p}_{T(j)}(l))$, together with the 95% confidence limits

$$\overline{m}_T(l) \pm \sqrt{\hat{V}_{T,m}(l)}\ 1.960/\sqrt{N},$$

where $\hat{V}_{T,m}(l)$ was calculated for each of the N sites as indicated in 6.4 and then averaged.
The approximations $\check{m}_T(l), l = 1, \ldots, 8$ run within these confidence limits $\overline{m}_T(l) \pm s$, see Fig. 4, the corrected forecasts $\check{m}_T(l) + \hat{B}_{T,m}(l)$ comes close to $\overline{m}_T(l)$ and hence close to the correct forecast of m_{T+l}. In the case of the beech tree the correction term $\hat{B}_T^{(A)}$ performs bad in the period 1997-2000.

Acknowledgement: This work is partially supported by the Deutsche Forschungsgemeinschaft (SFB 386).

References

Box,G.E.P and Jenkins,G.M. (1976) *Time Series Analysis: Forecasting and Control.* rev.ed. Holden-Day, San Francisco

Brockwell,P.J. and Davis,R.A. (1987). *Time Series: Theory and Methods.* Springer, New York

Fahrmeir,L. and Tutz,G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models.* Springer, New York

Fuller,W.A. (1976). *Introduction to Statistical Time Series.* Wiley, New York.

Goettlein,A. and Pruscha,H. (1992). Ordinal time series models with application to forest damage data. In: *Lecture Notes in Statistics* , **78**. Springer, New York, 113-118.

Goettlein,A. and Pruscha,H. (1996). The influence of stand characteristics, topography, site and weather on tree health in the forest district Rothenbuch. Forstw. Cbl., **115** (to appear)

Kaufmann,H. (1987). Regression models for nonstationary categorical time series: asymptotic estimation theory. *Ann. Statist.* , **15**, 79-98.

Landwehr,J.M., Pregibon,D. and Shoemaker,A.C. (1984). Graphical methods for assessing logistic regression models. *J.Am.Statist.Ass.,* **79**, 61-71.

Li,W.K. (1991). Testing model adequacy for some Markov regression models for time series. *Biometrika*, **78**, 83-89.

Li,W.K. (1994). Time series models based on generalized linear models: some further results. *Biometrics*, **50**, 506-511.

Liang,K.Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.

McCullagh,P. (1980). Regression models for ordinal data (with discussion). *J.R.Statist.Soc. B*, **42**, 109-142.

Norman,M.F. (1972). *Markov Processes and Learning Models.* Academic Press, New York.

Pruscha,H. (1993). Categorical time series with a recursive scheme and with covariates. *statistics*, **24**, 43-57

Pruscha,H. (1994). Partial residuals in cumulative regression models for ordinal data. *Statist. Papers*, **35**, 273-284.

Pruscha,H. (1995). Forecast methods in regression models for categorical time series. In: *Lecture Notes in Statistics* **104**. Springer, New York, 233-240.

Zeger,S.L. and Qaqish,B. (1988). Markov regression models for time series: a quasi likelihood approach. *Biometrics*, **44**, 1019-1031.

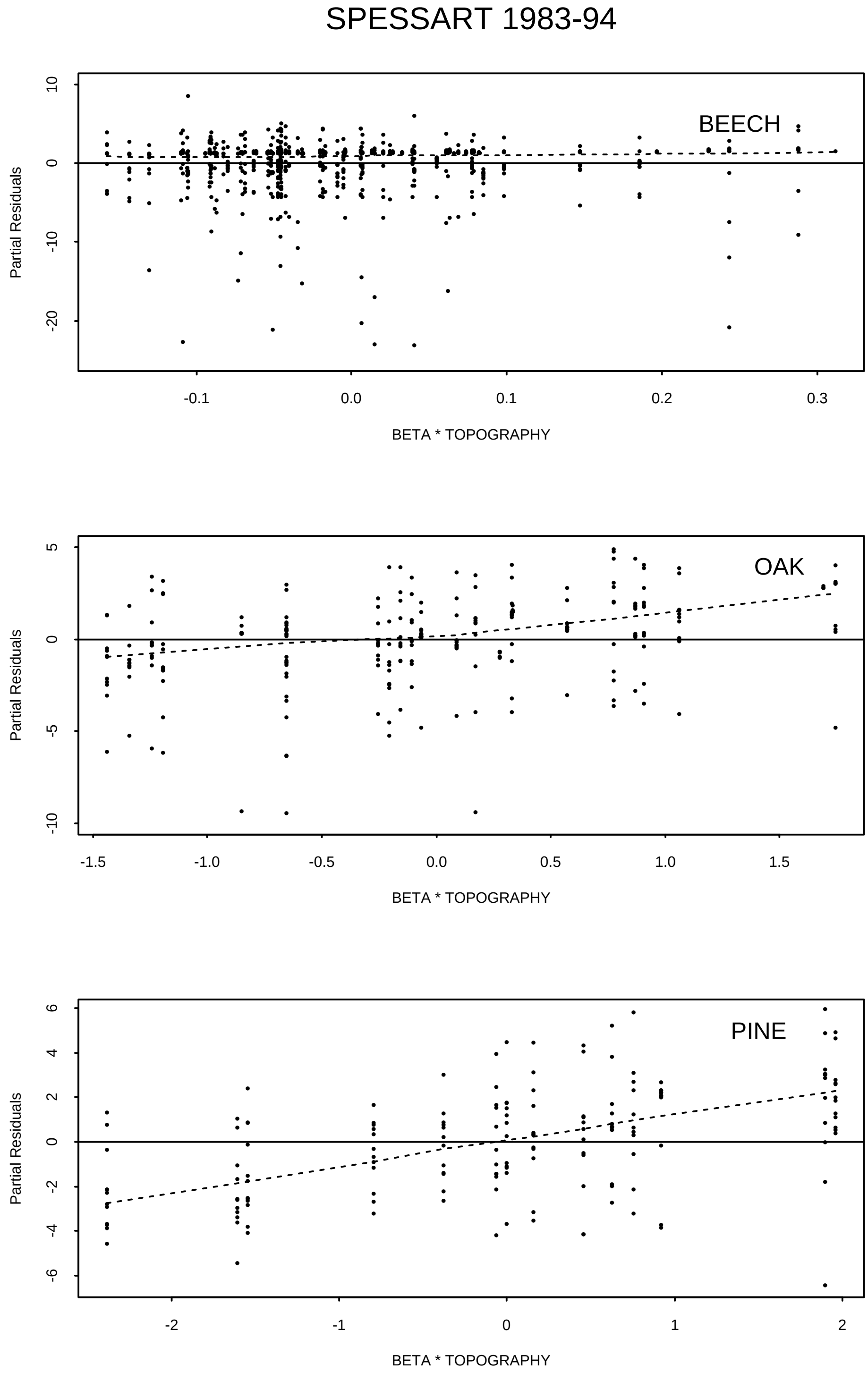

Figure 1: Partial residuals for the regressor set topography plotted over the regressor term beta∗topography, for each of the three tree species. A smoothing curve was fitted to the scatterplot.

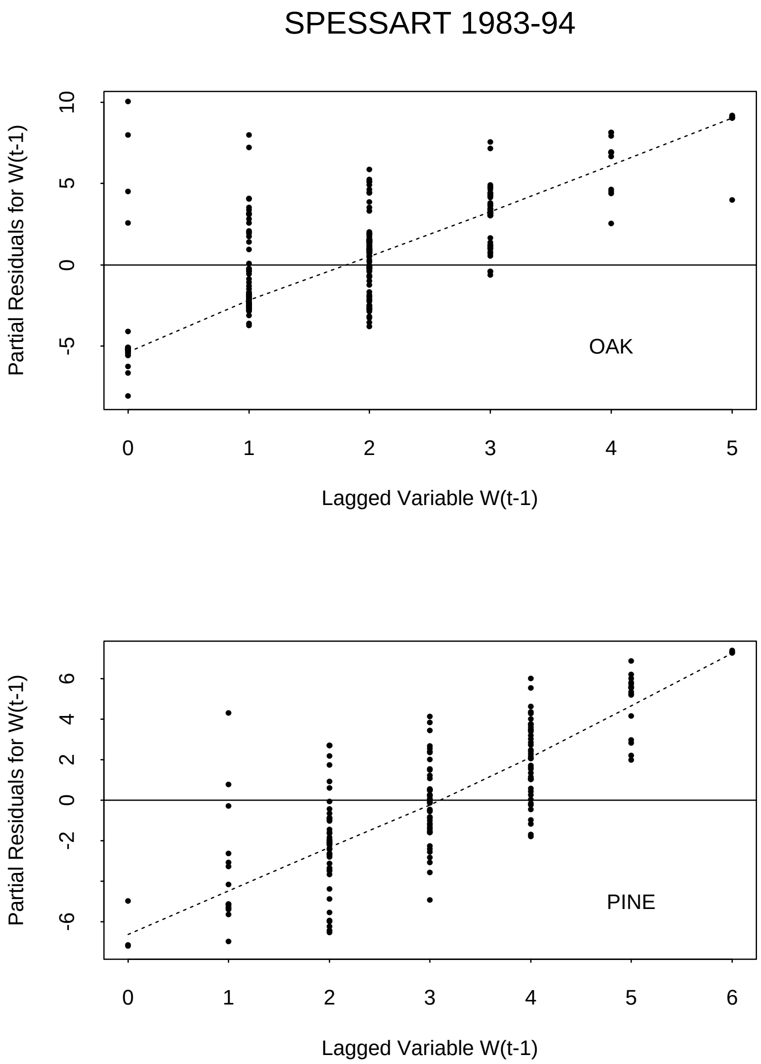

Figure 2: Partial residuals for the regressor W_{t-1}, the lagged (ordinally scaled) damage category, plotted over the values of W_{t-1}, for oak and pine trees. A smoothing curve was fitted to the scatterplot.

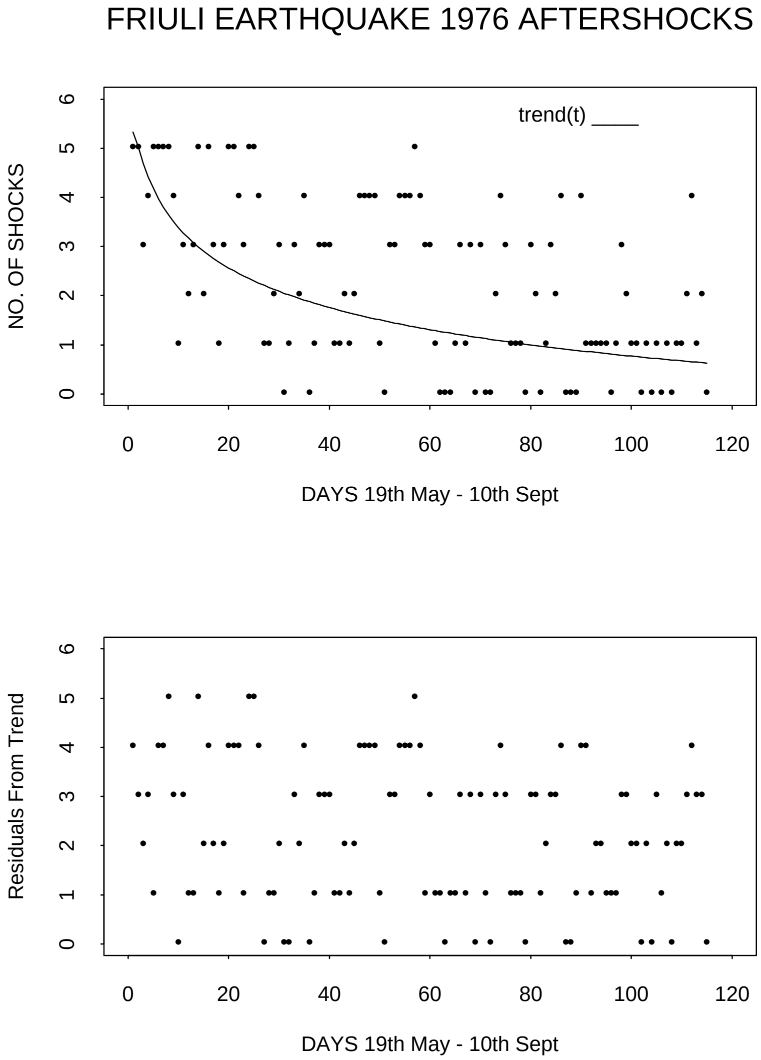

Figure 3: a) (top) Number Y_t of shocks per day plotted over the aftershock period of 115 consecutive days, together with a trend function. b) (bottom) Ordinally scaled partial residual variable Y_t^* from trend, plotted over the 115 consecutive days.

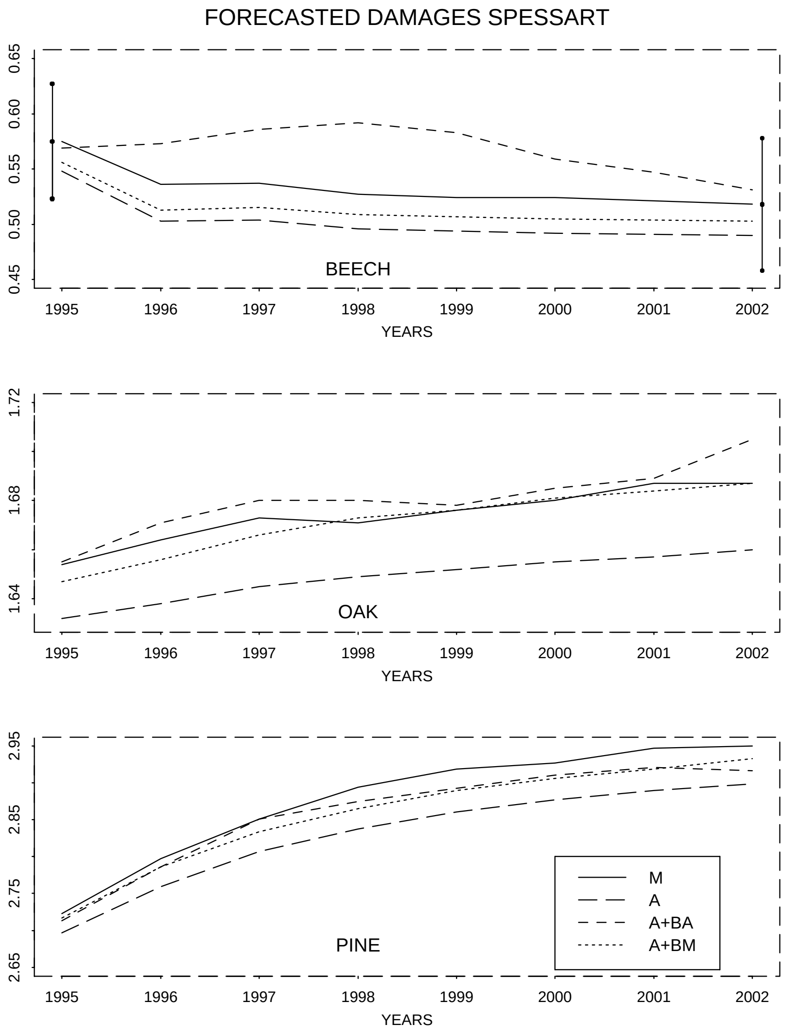

Figure 4: Forecasted forest damages for the years 1995 - 2002, for each of the three tree species. The approximation $\check{m}_T(l)$ $[A]$ is plotted, together with the corrections $\check{m}_T(l) + \hat{B}_T^{(A)}(l)$ and $\check{m}_T(l) + \hat{B}_T^{(M)}(l)$ $[A + BA$ and $A + BM]$ and with the Monte-Carlo solution $\overline{m}_T(l)$ $[M]$. At the beginning and the end of the beech curves, a confidence interval $M \pm s$ is indicated by vertical bars; in the case of the oak and pine curves these bars would overlap the whole plot area ($s \approx 0.15$ and $s \approx 0.30$, resp.) and are therefore omitted.