DFG-SNF Research Group FOR916

Statistical Regularization and Qualitative Constraints

Enno Mammen   Jens Perch Nielsen   Bernd Fitzenberger

Generalised Linear Time Series Regression

Preprint FOR916 10-19

Updated version

Preprint-Series of the Research Group FOR916

FOR916/10-19   **2010**

Biometrika (2010), **xx**, x, *pp*. 1–10
© 2007 Biometrika Trust
Printed in Great Britain

GENERALISED LINEAR TIME SERIES REGRESSION

BY ENNO MAMMEN
Department of Economics, University of Mannheim
L7, 3-5, 68131 Mannheim, Germany
E mail: emammen@rumms.uni-mannheim.de

JENS PERCH NIELSEN
Cass Business School, City University, London
106 Bunhill Row, London EC1Y 8TZ, United Kingdom
E mail: Jens.Nielsen.1@city.ac.uk

BERND FITZENBERGER
Dept. of Applied Econometrics, Albert-Ludwigs-University
PO Box, 79085 Freiburg, Germany
E mail: bernd.fitzenberger@vwl.uni-freiburg.de

SUBMITTED TO BIOMETRIKA

SUMMARY

We consider a cross section model that contains an individual component, a deterministic time trend and an unobserved latent common time series component. We show the following oracle property. The parameters of the latent time series and the parameters of the deterministic time trend can be estimated with the same asymptotic accuracy as if the parameters of the individual component would be known. We consider this model in two settings: least squares fits of linear specifications of the individual component and the parameters of the deterministic time trend and, more generally, quasi-likelihood estimation in a GLM model.

Some key words: cross sections; linear models; latent time series models; Generalised Linear Models.

1. INTRODUCTION.

Often time series data do not directly lend itself to a classical time series analysis, because the time series of interest is unobserved. In econometrics one finds recent extensions of classical panel data models where the calendar effects are not just in-sample parameters, but they are modeled as genuine time series to allow future forecasts. See for example the recent study of Linton et al. (2009) that also gives a short review of panel data in this particular context. Linton et al. (2009) adds a latent time series to a standard nonparametric regression problem and show that under certain assumptions one can analyse the estimated latent time series in exactly the same way as if it had been fully observed from the beginning. In this paper we will consider a

We thank two anonymous referees, the associate editor, and the editor for very useful comments. We are grateful to Dirk Antonczyk and Stefanie Schäfer for excellent research assistance in compiling the data used in example 1. The research of this paper was supported by the German Science Foundation (DFG) in the framework of the DFG-SNF Research Group FOR916: Statitical Regularization and Qualitative Constraints. Inference, Algortihms, Asymptotics and Applications.

broad class of parametric models with latent time series and we will show the analogous oracle property. One motivation for our analysis comes from an econometric labour market application with an additional identifiability issue arising in demographical age-period-cohort models. In demography the principle of a latent time series has long been used in mortality estimation and prediction, in particular since the appearance of the two important papers Lee & Carter (1992); Carter & Lee (1992) that first consider the in-sample mortality model in two steps. After the in-sample parameters have been estimated, the estimated calendar effect is redefined into a time series and is analysed as such. In this paper we prove for the first time that under certain regularity assumptions this procedure is indeed valid also when the latent time series is incorporated into the model from the outset. While Lee & Carter (1992); Carter & Lee (1992) only considered calendar and age effects on the mortality, recent mortality studies are also modeling cohort effects, see for example Cairns et al. (2009) and compare also our labour market example. This adds two important issues: identification of the model and designing forecast procedures. The identification issue arises because the calendar time is a simple addition of cohort year and age. The forecasting issue arises from the fact that the forecast can be poorly specified even when in-sample parameters are fully identified. Kuang et al. (2008a,b) did consider these two problems. In this paper we consider one important version of the age-period-cohort model and show that - after proper identification rules - one can estimate the calendar effect and forecast it as if the calendar effect had been fully known from the beginning. For some related nonparametric models, see Park et al. (2009) and Linton et al. (2009). While the full model formulation approach has been used in many practical papers for some time in demographics and actuarial science, see e.g. Lee & Miller (2001), Renshaw & Haberman (2003a,b), Wong-Fupuy & Haberman (2004) and Li & Chan (2005), this approach has only recently found its way into econometrics and empirical finance, see Park et al. (2009) and Fengler et al. (2007). We develop our model in three steps. In Section 2, we consider a linear time series cross section model

$$Y_{it} = X_{it}^T\beta + (Z_{it}^T\theta)(\psi_t^T\gamma + \eta_t) + \varepsilon_{it}, \tag{1}$$

with an obvious notation that will be specified in Section 2. In Section 3 we introduce our generalisation to the framework of generalised linear models by introducing the link function G :

$$Y_{it} = G\left\{X_{it}^T\beta + (Z_{it}^T\theta)(\psi_t^T\gamma + \eta_t)\right\} + \varepsilon_{it}, \tag{2}$$

notational details follow in Section 3. Our leading example of this model is reserving in non-life insurance. The oldest and most celebrated actuarial reserving model is the so called Chain-Ladder model. Only recently the stochastic properties of this model has been investigated, see England & Verrall (2002) for a review. However, while the non-life actuarial literature is concerned about modeling the calendar time effect, there is no real solid theory around this, see Zehnwirth (1994) for an early attempt. Our model in Section 3 combines the traditional Chain-Ladder approach with a well defined time series analysis of the calendar effect. In Section 4, we consider our final and most general model:

$$Y_{it} = G\left\{h_\beta\left(X_{it}\right) + g_\theta\left(Z_{it}\right)\left(\psi_t^T\gamma + \eta_t\right)\right\} + \varepsilon_{it}, \tag{3}$$

where h_β and g_θ are parametric families of functionals. This last model has the Lee-Carter model as a special case and it can for example also serve to modify the applications of the models (1) and (2).

2. ESTIMATION IN LINEAR MODEL TIME SERIES.

We consider the following cross section model

$$Y_{it} = X_{it}^T \beta + (Z_{it}^T \theta)(\psi_t^T \gamma + \eta_t) + \varepsilon_{it} \tag{4}$$

for $1 \leq t \leq T, 1 \leq i \leq I$. For simplicity of notation, we assume that the upper limit I of individuals i does not depend on time t. We observe the response Y_{it} and the random covariables X_{it} and Z_{it}. The vectors ψ_t are deterministic covariates to model the time trend. The process η_t is a common unobserved latent time series. We consider estimation of the time trend parameter γ and parametric fits for the time series structure of the process η_t. The parameter β is the regression parameter. We suppose that the first element of Z_{it} and θ is equal to one.

EXAMPLE 1. We use German labour market data grouped by year and age for the time period 1980 to 2004 analogous to Fitzenberger et al. (2001); Fitzenberger & Wunderlich (2002).[1] We analyze annual observations of log median real wages (deflated by the consumer price index) for medium–skilled male workers who work full–time.[2] We measure the age–year group specific unemployment rate as the share of benefit recipients among all observations. We model log wages as a function of separable age, cohort (year of birth), and year effects as well as the the unemployment rate:

$$V_{it} = \alpha_i + \xi_t + \kappa_{t-i} + U_{it}\beta + e_{it}$$

with V_{it} log median wages, U_{it} unemployment rate, i age, t time, $t-i$ cohort. Here, α_i, κ_{t-i} and β are parameters, ξ_t is a time series. The aim is inference on the dynamics of ξ_t. Note the identification problem arising from the linear relationship between age, cohort, and year. Therefore, we focus on the second differences:

$$\begin{aligned} Y_{it} &= V_{i+1,t+2} - V_{i,t+1} - V_{i+1,t+1} + V_{it}, \\ X_{it} &= U_{i+1,t+2} - U_{i,t+1} - U_{i+1,t+1} + U_{it}, \\ \varepsilon_{it} &= e_{i+1,t+2} - e_{i,t+1} - e_{i+1,t+1} + e_{it}, \\ \mu_t &= \xi_{t+2} - 2\xi_{t+1} + \xi_t. \end{aligned}$$

For these second differences, the parameters α_i and κ_{t-i} cancel out and it holds that[3]

$$Y_{it} = X_{it}^T \beta + \mu_t + \varepsilon_{it}.$$

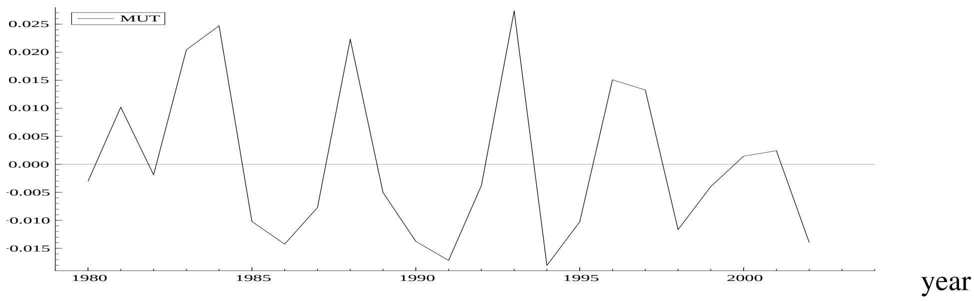
This is an example of model (4) in the paper. Our proposal is to treat in a first step μ_t as a parameter and to estimate μ_t and β by least squares. We estimate the first and second order autocorrelations of μ_t to be 0.007 and -0.493, respectively. Further investigations via augmented Dickey Fuller tests suggest that the cumulated process of μ_t is trend stationary whereas the cumulated process of the latter is not stationary. This suggests that the original process ξ_t is I(1). The coefficient of the unemployment rate β is estimated to be .097 with a heteroscedasticity robust standard error of .042. This significantly positive effect of unemployment suggests an inverse labor demand relationship across age groups, see Card & Lemieux (2001). Finally, a regression of μ_t on the first lag of the second difference of the log gross domestic product in West Germany plus an intercept yields a significantly positive coefficient of .342 suggesting

[1] The data stem from the IABS (IAB employment subsample). The IABS involves a 2% random sample of employees who participate in the German Social Security System. The data involve information about spells of benefit receipt which we use as proxy for unemployment. Although the data set starts in 1975, before 1980 the information on benefit receipt is incomplete.

[2] Medium–skilled workers have a completed vocational training (apprenticeship) degree and no university degree, see Fitzenberger & Wunderlich (2002). This skill group comprises the majority of all workers.

[3] Analogously, we could take the second order differences $Y_{it} = V_{i+2,t+1} - V_{i+1,t} - V_{i+1,t+1} + V_{it}$ to eliminate ξ_t and ρ_{t-i} (or $Y_{it} = V_{i,t+1} - V_{i,t} - V_{i+1,t+1} + V_{i+1,t}$ to eliminate ξ_t and α_i) in order to estimate the age (or the cohort) process.

Fig. 1. Estimates for μ_t by year t



strongly procyclical behavior with a lag of one year (we find no significant effect of the second difference in the same year). □

EXAMPLE 2. Our general model (4) (with covariates Z_{it} and parameter θ) can be used in a labor/health study, where health outcomes depend upon "macroconditions" (i.e. γ_t and η_t) interacting with workplace conditions Z_{it}. For instance, when workers do a lot of overtime in a cyclical boom, then stressful jobs (jobs requiring a lot of flexibility of the worker) may result in worse health outcomes. Differently, if recessionary periods increase the fear of unemployment for certain types of workers, this may make their health worse. See also Portrait et al. (2010) for a related empirical analysis with a long cohort dimension, for which the cohort process could be modelled as a time series (see footnote 3). □

For our model (4), we propose to estimate β and γ by least squares. Our main result is to show the following oracle property. Asymptotically the least squares estimator of γ and estimators of the time series parameters of η_t work as well as if the nuisance parameter β would be known. We put

$$\mu_t = \psi_t^T \gamma + \eta_t, \qquad \nu_t = \theta \mu_t$$

and we consider the following estimator of β and fit of μ_t

$$(\widehat{\beta}, \widehat{\nu}_t) = \arg\min_{\beta, \nu_t} \sum_{1 \le i \le I, 1 \le t \le T} (Y_{it} - X_{it}^T \beta - Z_{it}^T \nu_t)^2,$$
$$(\widehat{\theta}, \widehat{\mu}_t) = \arg\min_{\theta, \mu_t} \sum_{1 \le i \le I, 1 \le t \le T} [Z_{it}^T (\widehat{\nu}_t - \theta \mu_t)]^2, \tag{5}$$

where the second argmin runs over vectors θ with first element equal to one. The time trend parameter γ is estimated by least squares

$$\widehat{\gamma} = \arg\min_{\gamma} \sum_{1 \le t \le T} (\widehat{\mu}_t - \psi_t^T \gamma)^2. \tag{6}$$

Fits of the time series model for η_t are based on the estimation of its autocovariances for $h \ge 0$

$$\widehat{\rho}_h = T^{-1} \sum_{t=1}^{T-h} \widehat{\eta}_t \widehat{\eta}_{t+h} \tag{7}$$

with $\widehat{\eta}_t = \widehat{\mu}_t - \psi_t \widehat{\gamma}$. We assume that

(A1) The variables ε_{it} have conditional mean zero $E[\varepsilon_{it}|\mathcal{X}] = 0$ and they fulfill the following conditional dependence condition: $|E[\varepsilon_{it}\varepsilon_{js}|\mathcal{X}| \leq \Delta(|i-j|, |s-t|)$, $|E[\eta_u\eta_v\varepsilon_{it}\varepsilon_{js}|\mathcal{X}| \leq \Delta(|i-j|, |s-t|)$ $(1 \leq i, j \leq I, 1 \leq u, v, s, t \leq T)$ for a function Δ with $\sum_{i,s\geq 0} \Delta(i,s) < \infty$. Here the conditioning variable $\mathcal{X}$ is equal to $(X_{it}, Z_{it} : 1 \leq i \leq I, 1 \leq t \leq T)$.

(A2) The process (η_t) is mean zero and the terms $T^{-1/2}\sum_{t=1}^{T-h} \psi_t\eta_{t+h}$, $T^{-1/2}\sum_{t=1}^{T-h} \psi_{t+h}\eta_t$ (with $h \geq 0$) and $T^{-1}\sum_{t=1}^{T} \eta_t^2$ and $(T^{-1}\sum_{t=1}^{T} \eta_t^2)^{-1}$are (absolutely) bounded in probability.

(A3) It holds that $I, T \to \infty$ and $T = o(I^2)$.

(A4) The matrix $T^{-1}I^{-1}\sum_{1\leq i\leq I, 1\leq t\leq T} \widetilde{X}_{it}\widetilde{X}_{it}^T$ converges in probability to a matrix Γ that has full rank. Here, $\widetilde{X}_{it}^T$ is the vector $X_{it}^T - Z_{it}^T A_t^{-1} I^{-1}\sum_{j=1}^{I} Z_{jt}X_{jt}^T$ and A_t is the matrix $I^{-1}\sum_{j=1}^{I} Z_{jt}Z_{jt}^T$. The norm of the elements of matrices $I^{-1}\sum_{1\leq i\leq I} Z_{it}X_{it}^T$ for $1 \leq t \leq T$ and $I^{-1}\sum_{1\leq i\leq I} \psi_t X_{it}$ are uniformly bounded in probability. The operator norm of A_t and A_t^{-1} is uniformly stochastically bounded.

(A5) The covariables ψ_t may depend on T and it holds that $T^{-1}\sum_{1\leq t\leq T} \psi_t\psi_t^T$ converges to a matrix Ψ that has full rank.

We compare $\widehat{\gamma}$ and $\widehat{\rho}_h$ with the following theoretical oracle estimators:

$$\widetilde{\gamma} = \left(\sum_{t=1}^{T} \psi_t\psi_t^T\right)^{-1} \sum_{t=1}^{T} \psi_t\mu_t, \qquad \widetilde{\rho}_h = T^{-1}\sum_{t=1}^{T-h} \eta_t\eta_{t+h}.$$

Our main result is that the difference between $\widehat{\gamma}$ and $\widetilde{\gamma}$ and the difference between $\widehat{\rho}_h$ and $\widetilde{\rho}_h$ is asymptotically negligible. This means that asymptotically the least squares estimator of γ and the estimators of the time series parameters of η_t work as well as if the nuisance parameter β would be known. These oracle properties are stated in the following theorem.

THEOREM 1. *Under assumptions (A1) - (A5), it holds for* $h \geq 0$

$$\widehat{\gamma} = \widetilde{\gamma} + o_P(T^{-1/2}), \qquad (8)$$

$$\widehat{\rho}_h = \widetilde{\rho}_h + o_P(T^{-1/2}). \qquad (9)$$

The basic argument in the proof of this theorem is to show that $\widehat{\mu}_t - \mu_t$ is asymptotically equivalent to a weighted average of the error variables ε_{it}. In the calculation of the least squares estimator of γ and of the empirical autocovariances this term is averaged such that it does not contribute to any first order differences for these estimators and one gets the asymptotic equivalences (8)-(9). Our model contains as a special case that $Z_{it}^T\theta \equiv 1$. Then the theorem holds without the assumption $T = o(I^2)$ and it is only required that $I, T \to \infty$. In the general model the additional requirement $T = o(I^2)$ is needed to control the rate of $\widehat{\theta} - \theta$.

3. ESTIMATION IN GLM TIME SERIES.

We now come to the discussion of estimation in GLM models with latent time series components. Now we assume that

$$Y_{it} = G[X_{it}^T\beta + (Z_{it}^T\theta)(\psi_t^T\gamma + \eta_t)] + \varepsilon_{it} \qquad (10)$$

for $1 \le t \le T, 1 \le i \le I$. The function G is a known link function. Again, we observe the response Y_{it} and the random covariables X_{it} and Z_{it}. As above, the vectors ψ_t are deterministic covariates to model the time trend and we put $\mu_t = \psi_t^T \gamma + \eta_t$ and $\nu_t = \theta \mu_t$.

EXAMPLE 3. Consider the case where Y_{it} is the number of claims in an insurance portfolio and where $X_{it}^T \beta$ is the sum of two functions. The first function depends on the underwriting year i. The second function depends on the development period $t - i$, i.e. the time it takes for a claim to develop to the point t where the claim is reported to the insurance company. If the sum of these two functions is modeled as a spline function with a finite number of knots the spline parameters are given by β. The development period can often be several years, in personal accidents up to 20 years, and this is of course a challenge to the insurance industry to prudently account for this econometric assessment. Without the calendar effects this model exactly amounts to the celebrated Chain-Ladder model when G is the exponential link function. For many companies the value of such outstanding liabilities is several time the market value of the company. This illustrates the importance to keep improving the econometric methodology of this problem. Our model above gives for the first time a way to assess the Chain-Ladder type of regression estimates along with consistently and well defined time series effect that can be analysed as a standard time series. For a review on Chain-Ladder type models and their extensions, see England & Verrall (2002). For a recent extension of the chain-ladder model allowing for calendar effects see Kuang et al. (2008a,b) that derive the nontrivial rules of identification and forecasting in this context. □

EXAMPLE 4. Similar to the analysis of wages in example 1, Fitzenberger & Wunderlich (2004) investigate age, time, and cohort effects in labor force participation by females in both West Germany and the United Kingdom. This analysis could be implemented by estimating a GLM time series model using a probit or logit link function. It is important to estimate the time series dynamics in labor force participation of females (apart from cohort effects) in order to analyze the contributions in the pay-as-you-go social security system or the need for child care. □

For the theoretic discussion of GLM Time Series we need the following additional assumptions:

(A6) It holds that

$$\sup_{1 \le t \le T} \|\widehat{\nu}_t - \nu_t\| = o_P(T^{-1/4}),$$
$$\|\widehat{\beta} - \beta\| = o_P(T^{-1/4}).$$

(A7) The estimators are approximate solutions of the following score equations

$$\sup_{1 \le t \le T} \left| I^{-1} \sum_{1 \le i \le I} [Y_{it} - G(\widehat{m}_{it})] w(\widehat{m}_{it}) \right| = o_P(T^{-1/2}),$$
$$I^{-1} T^{-1} \sum_{1 \le i \le I, 1 \le t \le T} [Y_{it} - G(\widehat{m}_{it})] w(\widehat{m}_{it}) X_{it} = o_P(T^{-1/2}).$$

Here, $\widehat{m}_{it} = X_{it}^T \widehat{\beta} + Z_{it}^T \widehat{\nu}_t$ and w is a weighting function.

Examples for estimators that fulfill (A7) are quasi-likelihood estimators in generalised linear models (GLM). For a positive function V the quasi-likelihood [QL] function is defined as $Q(\tau; y) = \int_\tau^y (s - y) V(s)^{-1} \, ds$ where τ is the (conditional) expectation of Y, i.e. in our case

$\tau = G\{X^T\beta + Z^T\nu\}$. The quasi-likelihood estimator fulfills the two equations in (A7) with $w(u) = G'(u)/V[G(u)]$. In particular, the equations hold with the right hand sides replaced by zero. In the next assumption we assume that w has bounded support. This simplifies the asymptotic discussion but allows only truncated versions of quasi-likelihood estimation.

(A8) The functions G and w are twice differentiable and have a bounded second derivative. The weight function w has a bounded support.

We conjecture that (A8) could be weakened to allow a sequence of weight functions with increasing support or even to allow weight functions with unbounded support. But in both cases, one would need rather technical tail conditions. These theoretical discussions are out of the scope of this paper. In applications we would propose to use the quasi-likelihood estimator that corresponds to a weighting function with unbounded support.

As in the last section, the time series μ_t and the parameter θ is estimated as in (5). The trend parameter γ is estimated by least squares (6). Again, we consider fits of time series models for η_t that are based on the estimation of its autocovariances $\widehat{\rho}_h$ for $h \geq 0$, see (7). We compare $\widehat{\gamma}$ and $\widehat{\rho}_h$ with their theoretical oracle estimators $\widetilde{\gamma}$ and $\widetilde{\rho}_h$ that are defined as in the last section. We now state an oracle property for the GLM time series model.

THEOREM 2. *Under assumptions (A1)—(A8), it holds that* $\widehat{\gamma} = \widetilde{\gamma} + o_P(T^{-1/2})$ *and* $\widehat{\rho}_h = \widetilde{\rho}_h + o_P(T^{-1/2})$ *(for* $h \geq 0$*).*

4. GENERALISED TIME SERIES REGRESSION

We now come to a short discussion of a generalised time series model:

$$Y_{it} = G\left\{h_\beta\left(X_{it}\right) + g_\theta\left(Z_{it}\right)\left(\psi_t^T\gamma + \eta_t\right)\right\} + \varepsilon_{it}.$$

The model is as in Section 2 but with the extension of our introduction of two parametric families h_β and g_θ.

EXAMPLE 5. The celebrated model of Lee & Carter (1992) and Carter & Lee (1992) is a special case of this model. Notice that our more general model formulation leave the applied statistician with the opportunity to challenge Lee and Carters original model with various modifications of it. For some recent literature estimating the Lee-Carter parameters based on poisson regression, see Brouhns et al. (2002) and see Cairns et al. (2009) for recent modifications of the Lee-Carter structure that is also contained in our model framework. For recent applications of this type of models to the financial construction of survivor linked bonds, see Blake et al. (2006) and Dowd et al. (2006). None of these papers contain a full model structure, where the time series structure is incorporated into the model framework from the beginning. The formulation of this full model structure is therefore a contribution of this paper. □

5. PROOFS.

Proof of Theorem 1. First note that

$$\widehat{\beta} - \beta = \left(T^{-1}I^{-1} \sum_{1\leq i\leq I, 1\leq t\leq T} \widetilde{X}_{it}\widetilde{X}_{it}^T\right)^{-1} T^{-1}I^{-1} \sum_{1\leq i\leq I, 1\leq t\leq T} \widetilde{X}_{it}\varepsilon_{it}, \tag{11}$$

$$\widehat{\nu}_t - \nu_t = I^{-1}A_t^{-1} \sum_{1\leq i\leq I} Z_{it}\varepsilon_{it} - I^{-1}A_t^{-1} \sum_{1\leq i\leq I} Z_{it}X_{it}^T(\widehat{\beta} - \beta). \tag{12}$$

It can be easily checked that $E[\|\widehat{\beta}-\beta\|^2|\mathcal{X}] = O_P(T^{-1}I^{-1})$ and thus $\|\widehat{\beta}-\beta\| = O_P(T^{-1/2}I^{-1/2})$. Because of (A4) and (A5) this implies that

$$\sup_{1\le t\le T} |\widehat{\nu}_t - \nu_t - I^{-1}A_t^{-1}\sum_{1\le i\le I} Z_{it}\varepsilon_{it}| = O_P(I^{-1/2}T^{-1/2}), \qquad (13)$$

$$T^{-1/2}I^{-1}\sum_{1\le i\le I, 1\le t\le T} \psi_t X_{it}^T(\widehat{\beta}-\beta) = O_P(I^{-1/2}).$$

From the definition (5) we get that $I\sum_{1\le t\le T}(\widehat{\nu}_{t,1}-\widehat{\mu}_t)^2 \le \sum_{1\le i\le I,1\le t\le T}[Z_{it}^T(\widehat{\nu}_t - \widehat{\theta}\widehat{\mu}_t)]^2 \le \sum_{1\le i\le I,1\le t\le T}[Z_{it}^T(\widehat{\nu}_t-\nu_t)]^2$. This bound and (13) can be used to show that $\widehat{\theta}-\theta = O_P(I^{-1/2})$ and $T^{-1}\sum_{1\le t\le T}(\widehat{\mu}_t-\mu_t)^2 = O_P(I^{-1})$. With these expansions one can approximate the score function that is given by the derivatives of the left hand side of (5) with respect to θ and μ_t. After some algebra one gets that $\widehat{\theta}-\theta = O_P(T^{-1/2}+I^{-1})$ where the rate I^{-1} is a quadratic approximation error coming from the above bounds of order $O_P(I^{-1/2})$ for $\widehat{\theta}-\theta$ and $\widehat{\mu}_t-\mu_t$. This bound and the linearized score equation can be used to show that

$$\widehat{\mu}_t - \mu_t = T^{-1/2}I^{-1}\sum_{1\le i\le I} w_{it}\varepsilon_{it} + O_P(\delta_t) \qquad (14)$$

with $w_{it} = (\theta^T Z_{it})/(\theta^T A_t\theta)$ and $\delta_t = (T^{-1/2}I^{-1/2}+I^{-1})(1+\|\widehat{\nu}_t-\nu_t\|+\|\mu_t\|)$. For claim (8) we have to show

$$T^{-1/2}\sum_{t=1}^{T}\psi_t(\widehat{\mu}_t-\mu_t) = o_P(1). \qquad (15)$$

For the proof of this claim note first that, because of (14), $T^{-1/2}\sum_{t=1}^{T}\psi_t(\widehat{\mu}_t-\mu_t) = T^{-1/2}I^{-1}\sum_{1\le i\le I,1\le t\le T}\psi_t w_{it}\varepsilon_{it} + O_P(I^{-1/2})$. Because of (A1), (A4) and (A5) the right hand side of this equation is of order $O_P(I^{-1/2})$. This shows (15). Thus, (8) is shown.

We now show (9). We have to show for $h \ge 0$ that

$$T^{1/2}(\widehat{\rho}_h - \widetilde{\rho}_h) = T^{-1/2}\sum_{t=1}^{T-h}([\widehat{\mu}_t - \psi_t^T\widehat{\gamma}][\widehat{\mu}_{t+h}-\psi_{t+h}^T\widehat{\gamma}] - \eta_t\eta_{t+h}) = o_P(1).$$

For this claim we will show

$$T^{-1/2}\sum_{t=1}^{T-h}\eta_{t+h}(\widehat{\mu}_t-\psi_t^T\widehat{\gamma}-\eta_t) = O_P(T^{-1/2}+I^{-1/2}), \qquad (16)$$

$$T^{-1/2}\sum_{t=1}^{T-h}\eta_t(\widehat{\mu}_{t+h}-\psi_{t+h}^T\widehat{\gamma}-\eta_{t+h}) = O_P(T^{-1/2}+I^{-1/2}), \qquad (17)$$

$$T^{-1/2}\sum_{t=1}^{T-h}(\widehat{\mu}_{t+h}-\psi_{t+h}^T\widehat{\gamma}-\eta_{t+h})(\widehat{\mu}_t-\psi_t^T\widehat{\gamma}-\eta_t) = O_P(T^{-1/2}+I^{-1/2}). \qquad (18)$$

Now, because of (14), we have that uniformly for $1\le t\le T$ the following expansion holds: $\widehat{\mu}_t - \psi_t^T\widehat{\gamma} - \eta_t = I^{-1}\sum_{1\le i\le I} w_{it}\varepsilon_{it} - \psi_t^T(\widehat{\gamma}-\gamma) + O_P(\delta_t)$. Using this expansion, it can be easily seen that (16) and (17) follow from (A3) and $\widehat{\gamma}-\gamma = O_P(T^{-1/2})$, $T^{-1/2}I^{-1}\sum_{t=1}^{T-h}\sum_{i=1}^{I}\eta_t w_{it}\varepsilon_{it+h} = O_P(I^{-1/2}T^{-1/2})$ and

$T^{-1/2}I^{-1}\sum_{t=1}^{T-h}\sum_{i=1}^{I}\eta_{t+h}w_{it}\varepsilon_{it} = O_P(I^{-1/2}T^{-1/2})$. Note that the latter expansions follow from (A1).

For (9) it remains to check (18). This can be done by showing $T^{-1/2}\sum_{t=1}^{T-h}\psi_t^T(\widehat{\gamma}-\gamma)\psi_{t+h}^T(\widehat{\gamma}-\gamma) = O_P(T^{-1/2})$, $T^{-1/2}I^{-1}\sum_{t=1}^{T-h}\sum_{i=1}^{I}\varepsilon_{it}\psi_{t+h}^T(\widehat{\gamma}-\gamma) = O_P(I^{-1/2}T^{-1/2})$, $T^{-1/2}I^{-1}\sum_{t=1}^{T-h}\sum_{i=1}^{I}\varepsilon_{it+h}\psi_t^T(\widehat{\gamma}-\gamma) = O_P(I^{-1/2}T^{-1/2})$, and $T^{-1/2}I^{-2}\sum_{t=1}^{T-h}\sum_{i=1}^{I}\sum_{j=1}^{I}\varepsilon_{it+h}\varepsilon_{jt} = O_P(I^{-1})$. These expansions can be shown by using (A1) and (A5). This completes the proof of (9). □

Proof of Theorem 2. By expanding the score functions in (A7) we get stochastic expansions of $\widehat{\beta}-\beta$ and $\widehat{\nu}_t-\nu_t$ where the first terms are weighted modifications of the right hand side of (11) or (12), respectively. These expansions are of order $o_P(T^{-1/2})$. To get these expansions, one applies that $\|\widehat{\beta}-\beta\|^2$ and $|\widehat{\nu}_t-\nu_t|^2$ are of order $o_P(T^{-1/2})$, see (A6). The further proof of Theorem 2 can be carried out by similar arguments as in the proof of Theorem 1. □

REFERENCES

BLAKE, D., CAIRNS, A., DOWD, K. & MACMINN, R. (2006). Longevity bonds: Financial engineering, valuation and hedging. *Journal of Risk and Insurance* **73**, 647–672.

BROUHNS, N., DENUIT, M. & VERMUNT, J. (2002). A poisson log-bilinear approach to the construction of projected lifetables. *Insurance: Mathematics and Economics* **31**, 373–393.

CAIRNS, A., BLAKE, D., DOWD, K., COUGHLAN, G., EPSTEIN, D., ONG, A. & BALEVICH, I. (2009). A quantitative comparison of stochastic mortality models using data from england and wales and the united states. *North American Actuarial J.* **13**, 1–35.

CARD, D. & LEMIEUX, T. (2001). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *Quarterly Journal of Economics* **116**, 705–746.

CARTER, L. & LEE, R. (1992). Modelling and forecasting u.s sex differentials in mortality. *International Journal of Forecasting* **8**, 393 – 411.

DOWD, K., BLAKE, D., CAIRNS, A. & DAWSON, P. (2006). Survivor swaps. *Journal of Risk and Insurance* **73**, 1–17.

ENGLAND, P. & VERRALL, R. (2002). Stochastic claims reserving in general insuranced. *British Actuarial Journal* **8**, 443–518.

FENGLER, M., HÄRDLE, W. & MAMMEN, E. (2007). A semiparametric factor model for implied volatility surface dynamics. *J. Financial Econometrics* **5**, 189–218.

FITZENBERGER, B., HUJER, R., MACURDY, T. & SCHNABEL, R. (2001). "Testing for Uniform Wage Trends in West Germany: A Cohort Analysis Using Quantile Regressions for Censored Data". *Empirical Economics* **26(1)**, 41–86.

FITZENBERGER, B. & WUNDERLICH, G. (2002). "Gender Wage Differences in West Germany: A Cohort Analysis". *German Economic Review* **3(4)**, 379–414.

FITZENBERGER, B. & WUNDERLICH, G. (2004). "The Changing Life Cycle Pattern in Female Employment: A Comparison of Germany and the UK". *Scottish Journal of Political Economy* **51(3)**, 302–328.

KUANG, D., NIELSEN, B. & NIELSEN, J. (2008a). Forecasting with the age-period-cohort model and the extended chain-ladder model. *Biometrika* **95**, 987–991.

KUANG, D., NIELSEN, B. & NIELSEN, J. (2008b). Identification of the age-period-cohort model and the extended chain-ladder model. *Biometrika* **95**, 979–986.

LEE, R. & CARTER, L. (1992). Modelling and forecasting u.s mortality. *Journal of the American Statistical Association* **87**, 659–671.

LEE, R. & MILLER, T. (2001). Evaluating the performance of the lee-carter method for forecasting mortality. *Demography* **8**, 537–549.

LI, S. & CHAN, W. (2005). Outlier analysis and mortality forecasting: The united kingdom and scandinavian countries. *Scandinavian Actuarial Journal* **3**, 187–211.

LINTON, O., NIELSEN, J. & NIELSEN, S. (2009). Nonparametric regression with a latent time series. *Econometrics Theory (in print)* .

PARK, B., MAMMEN, E., HRDLE, W. & BORAK, S. (2009). Time series modelling with semiparametric factor dynamics. *J. Amer. Statist. Assoc. (in print)* .

PORTRAIT, F., ALESSIE, R. & DEEG, D. (2010). Do early life and contemporaneous macroconditions explain health at older ages? an application to functional limitations of dutch older individuals. *J. Popul. Econ.* **23**, 617642.

RENSHAW, A. & HABERMAN, S. (2003a). On the forecasting of mortality reduction factors. *Insurance: Mathematics and Economics* **2**, 379–401.

RENSHAW, A. & HABERMAN, S. (2003b). On the forecasting of mortality reduction factors. *Applied Statistics* **2**, 119–137.

WONG-FUPUY, C. & HABERMAN, S. (2004). Projection mortality trends: Recent developments in the united kingdom and the united states. *North American Actuarial Journal* **8**, 56–83.

ZEHNWIRTH, B. (1994). Probabilistic development factor models with applications to loss reserve variability, prediction intervals, and risk-based capital. *Casualty Actuarial Society Forum* , 447–605.