Biometrika (2011), **98**, 4, *pp.* 1007–1014
doi: 10.1093/biomet/asr044
Advance Access publication 13 October 2011

Generalized linear time series regression

BY ENNO MAMMEN

Department of Economics, University of Mannheim, 68131 Mannheim, Germany
emammen@rumms.uni-mannheim.de

JENS PERCH NIELSEN

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

AND BERND FITZENBERGER

Department of Applied Econometrics, Albert-Ludwigs-University, 79085 Freiburg, Germany
bernd.fitzenberger@vwl.uni-freiburg.de

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 were 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, quasilikelihood estimation in a generalized linear time series model.

Some key words: Cross-section; Generalized linear time series model; Latent time series; Linear model.

1. INTRODUCTION

Often time series data do not directly lend themselves to a classical analysis, because the series of interest is unobserved. In econometrics, one finds recent extensions of classical panel data models where calendar effects are modelled as genuine time series to allow future forecasts. See, for example, the recent study of Linton et al. (2009), which gives a review of panel data in this particular context, adds a latent time series to a standard nonparametric regression problem, and shows that under certain assumptions one can analyse the estimated latent time series as if it had been fully observed from the beginning. In this paper, we consider a broad class of parametric models with latent time series and we 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 Lee & Carter (1992) and Carter & Lee (1992), which 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 that under certain regularity assumptions this procedure is valid also when the latent time series is incorporated into the model from the outset. While Lee & Carter (1992) and Carter & Lee (1992) only consider calendar and age effects on mortality, recent studies also model cohort effects; see, for example, Cairns et al. (2009) and our labour market example. This adds two important issues: identification of the model and forecasting. 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, 2008b) address these two problems. In this paper, we consider one important additive version of the age-period-cohort model and we show that one can estimate the calendar effect and forecast it as if it were fully known from the beginning (Fitzenberger et al., 2001; Fitzenberger & Wunderlich, 2002). For some related nonparametric models, see Park et al. (2009) and Linton et al. (2009). While the full model formulation approach has for some time been used in demographics and actuarial science (Lee & Miller, 2001; Renshaw & Haberman, 2003a, 2003b; Wong-Fupuy & Haberman, 2004; Li & Chan, 2005), it has only recently found its way into econometrics and empirical finance (Fengler et al., 2007; Park et al., 2009).

2. ESTIMATION IN A LINEAR TIME SERIES MODEL

In this section, we consider a linear time series cross-section model

$$Y_{it} = X_{it}^{\mathrm{T}}\beta + (Z_{it}^{\mathrm{T}}\theta)(R_t^{\mathrm{T}}\gamma + \eta_t) + \varepsilon_{it} \quad (t = 1, \ldots, T;\ i = 1, \ldots, I). \tag{1}$$

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 R_t are deterministic covariates to model the time trend. The process η_t is a common unobserved latent time series. We assume that the first elements of Z_{it} and θ are equal to unity. This linear model is related to the one factor models of Bai & Ng (2006) and Bai (2009).

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 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 were known. We put $\mu_t = R_t^{\mathrm{T}}\gamma + \eta_t$, $\nu_t = \theta\mu_t$ and we consider the following estimator of β and fit of μ_t,

$$\begin{aligned}
(\hat\beta, \hat\nu_t) &= \arg\min_{\beta,\nu_t} \sum_{i=1,\ldots,I;\ t=1,\ldots,T} (Y_{it} - X_{it}^{\mathrm{T}}\beta - Z_{it}^{\mathrm{T}}\nu_t)^2,\\
(\hat\theta, \hat\mu_t) &= \arg\min_{\theta,\mu_t} \sum_{i=1,\ldots,I;\ t=1,\ldots,T} \{Z_{it}^{\mathrm{T}}(\hat\nu_t - \theta\mu_t)\}^2,
\end{aligned} \tag{2}$$

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

$$\hat\gamma = \arg\min_{\gamma} \sum_{t=1}^{\mathrm{T}} (\hat\mu_t - R_t^{\mathrm{T}}\gamma)^2. \tag{3}$$

Fits of the time series model for η_t are based on the estimation of its autocovariances

$$\hat\rho_h = T^{-1} \sum_{t=1}^{\mathrm{T}-h} \hat\eta_t \hat\eta_{t+h} \ \ (h \geqslant 0), \tag{4}$$

with $\hat\eta_t = \hat\mu_t - R_t\hat\gamma$. We assume that:

Assumption 1. The variables ε_{it} have conditional mean zero $E(\varepsilon_{it} \mid \mathcal{X}) = 0$ and they fulfil the conditional dependence condition: $|E(\varepsilon_{it}\varepsilon_{js} \mid \mathcal{X})| \leqslant \Delta(|i-j|, |s-t|)$, $|E(\eta_u\eta_v\varepsilon_{it}\varepsilon_{js} \mid \mathcal{X})| \leqslant \Delta(|i-j|, |s-t|)$ $(i, j = 1, \ldots, I; u, v, s, t = 1, \ldots, T)$ for a function Δ with $\sum_{i,s=0}^{\infty} \Delta(i, s) < \infty$. Here the conditioning variable $\mathcal{X}$ equals $\{X_{it}, Z_{it} : i = 1, \ldots, I;\ t = 1, \ldots, T\}$.

Assumption 2. The process η_t is mean zero and the terms $T^{-1/2}\sum_{t=1}^{\mathrm{T}-h} R_t\eta_{t+h}$, $T^{-1/2}\sum_{t=1}^{\mathrm{T}-h} R_{t+h}\eta_t$ $(h \geqslant 0)$ and $T^{-1}\sum_{t=1}^{\mathrm{T}} \eta_t^2$ and $(T^{-1}\sum_{t=1}^{\mathrm{T}} \eta_t^2)^{-1}$ are absolutely bounded in probability.

Assumption 3. It holds that $I, T \to \infty$ and $T = o(I^2)$.

Assumption 4. The matrix $T^{-1}I^{-1}\sum_{i=1,\ldots,I;\ t=1,\ldots,T}\tilde{X}_{it}\tilde{X}_{it}^{\mathrm{T}}$ converges in probability to a matrix Γ that has full rank. Here, $\tilde{X}_{it}^{\mathrm{T}}$ is the vector $X_{it}^{\mathrm{T}} - Z_{it}^{\mathrm{T}}A_t^{-1}I^{-1}\sum_{j=1}^{I} Z_{jt}X_{jt}^{\mathrm{T}}$ and A_t is the matrix $I^{-1}\sum_{j=1}^{I} Z_{jt}Z_{jt}^{\mathrm{T}}$. The norms of the elements of matrices $I^{-1}\sum_{i=1}^{I} Z_{it}X_{it}^{\mathrm{T}}$ and $I^{-1}\sum_{i=1}^{I} R_t X_{it}$ $(t = 1, \ldots, T)$ are uniformly bounded in probability. The operator norm of A_t and A_t^{-1} is uniformly stochastically bounded.

Assumption 5. The covariables R_t may depend on T and $T^{-1}\sum_{t=1}^{\mathrm{T}} R_t R_t^{\mathrm{T}}$ converges to a full rank matrix Ψ.

Assumption 1 allows dependencies between the errors that are local in time and after a suitable reordering of the individuals also local in i. In applications the reordering may correspond to closeness of the individuals in age, region or other characteristics. Assumption 5 can be achieved by normalizing of R_t such that $T^{-1}\sum_{t=1}^{\mathrm{T}} R_t R_t^{\mathrm{T}}$ equals the identity matrix. Assumptions 2 and 4 hold, for example, under appropriate stationarity and ergodicity assumptions on X_{it} and Z_t.

We compare $\hat{\gamma}$ and $\hat{\rho}_h$ with the theoretical oracle estimators

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

Our main result is that the difference between $\hat{\gamma}$ and $\tilde{\gamma}$ and the difference between $\hat{\rho}_h$ and $\tilde{\rho}_h$ 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 were known. These oracle properties are stated in the following theorem.

THEOREM 1. *Under Assumptions* 1–5, $\hat{\gamma} = \tilde{\gamma} + o_P(T^{-1/2})$ *and* $\hat{\rho}_h = \tilde{\rho}_h + o_P(T^{-1/2})$, *for* $h \geqslant 0$.

The basic argument in the proof of this theorem is to show that $\hat{\mu}_t - \mu_t$ 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 stated in Theorem 1. Our model contains as a special case $Z_{it}^{\mathrm{T}}\theta \equiv 1$. 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 $\hat{\theta} - \theta$.

Example 1. We use German labour market data, from a 2% random sample of employees subject to social security, grouped by year and age for the time period 1980–2004. Fitzenberger et al. (2001) and Fitzenberger & Wunderlich (2002) used earlier versions of these data. The data also involve information about receipt of unemployment benefits. We analyse annual observations of log median real wages, deflated by the consumer price index, for full-time working males with a completed vocational training degree who work full-time. We measure the age-year group specific unemployment rate by the share of benefit recipients. We model log wages as a function of age, cohort and year effects as well as the unemployment rate by $V_{it} = \alpha_i + \xi_t + \kappa_{t-i} + U_{it}\beta + e_{it}$ 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 $Y_{it} = V_{i+1,t+2} - V_{i,t+1} - V_{i+1,t+1} + V_{it}$, $X_{it} = U_{i+1,t+2} - U_{i,t+1} - U_{i+1,t+1} + U_{it}$, $\varepsilon_{it} = e_{i+1,t+2} - e_{i,t+1} - e_{i+1,t+1} + e_{it}$ and $\mu_t = \xi_{t+2} - 2\xi_{t+1} + \xi_t$. For these second differences, the parameters α_i and κ_{t-i} cancel and

$$Y_{it} = X_{it}^{\mathrm{T}}\beta + \mu_t + \varepsilon_{it}. \tag{5}$$

This is an example of model (1). 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, respectively, in order to estimate the age or the cohort process.

Table 1. *Ratios of mean squared differences for the oracle estimator and sample estimator in scenarios (i)–(iv). Monte Carlo standard deviation is in parentheses and the calculation is based on the delta method*

Ratios	(i)	(ii)	(iii)	(iv)
Oracle-to-sample(ϕ_1)	0·995 (0·008)	0·996 (0·004)	0·997 (0·011)	0·885 (0·048)
Oracle-minus-sample(ϕ_1)	0·015 (0·040)	0·004 (0·050)	0·038 (0·040)	0·800 (0·050)
Oracle-to-sample(ϕ_2)	0·996 (0·007)	0·993 (0·003)	0·981 (0·010)	0·588 (0·041)
Oracle-minus-sample(ϕ_2)	0·010 (0·050)	0·002 (0·040)	0·021 (0·040)	0·654 (0·040)

Our proposal is to first treat μ_t as a parameter and to estimate μ_t and β in equation (5) by least squares; we estimate heteroscedasticity robust standard errors. The coefficient of the unemployment rate β is estimated as 0·097 (0·042), here and henceforth standard errors in parentheses. This significantly positive effect of unemployment likely reflects an inverse labour demand relationship (Card & Lemieux, 2001). Our estimates suggest that μ_t is fairly precisely estimated, based on standard errors that are estimated, here and in the following, as if μ_t were known; our theory says that the difference is asymptotically negligible. Next, we investigate the time series process governing μ_t. The estimated first and second order autocorrelations are $-0{\cdot}002$ (0·209) and $-0{\cdot}476$ (0·209), respectively. A Dickey–Fuller test suggests that μ_t is not nonstationary. Estimating an autoregressive process of order 2, i.e., $\mu_t = \phi_1\mu_{t-1} + \phi_2\mu_{t-2} + e_t$, yields $\hat{\phi}_1 = 0{\cdot}0089$ (0·202) and $\hat{\phi}_2 = -0{\cdot}498$ (0·202). We find no autocorrelation in the error term e_t. Further detailed analysis suggests that the cumulated process of μ_t is stationary.

To illustrate the usefulness of our theory, we simulate data based on an idealized version of our sample data. We hold the unemployment rates fixed and we replicate their values when we increase the sample size. We use the model estimates for β, ϕ_1 and ϕ_2 to simulate data. We draw the error terms from a normal distribution with zero mean and a variance corresponding to the estimated sample variance. We simulate 1000 random samples for each scenario. We focus on the estimation of ϕ_1 and ϕ_2. Table 1 reports the ratio of the mean squared error between the oracle estimators, which assumes knowledge of the true simulated μ_t, and the estimators based on the estimated μ_t, which we denote by oracle-to-sample(ϕ), where $\phi \in \{\phi_1, \phi_2\}$, and the ratio between the mean squared difference between the sample estimator and the oracle estimator divided by the mean squared error of the sample estimator, which we denote by oracle-minus-sample(ϕ). These ratios are defined as: oracle-to-sample $(\phi) = \sum_{s=1}^{1000}(\tilde{\phi}_s - \hat{\phi})^2 / \sum_{s=1}^{1000}(\hat{\phi}_s - \hat{\phi})^2$ and oracle-minus-sample $(\phi) = \sum_{s=1}^{1000}(\tilde{\phi}_s - \hat{\phi}_s)^2 / \sum_{s=1}^{1000}(\hat{\phi}_s - \hat{\phi})^2$, where s denotes the simulated sample, $\hat{\phi}$ is the parameter used to simulate the data, $\hat{\phi}_s$ is the sample estimator in the sth simulated sample and $\tilde{\phi}_s$ is the oracle estimator. We consider four scenarios: (i) the same sample size as in the data application with $I = 34$ and $T = 23$, (ii) $T = 100$ and $I = 100$, (iii) $I = 15$ and $T = 23$ using only age up to 40, (iv) as in (i) but increasing the standard deviation of the error term in model (1) by a factor 10. For scenario (i), the mean squared error of the oracle estimator for ϕ_1 are ϕ_2 are 0·995 times and 0·996 times, respectively, as large as the mean squared error for the sample estimator, referring to the oracle-to-sample ratios reported in Table 1, i.e., for the original sample type the result in Theorem 1 is strongly supported. Furthermore, the oracle-minus-sample ratio is only 1·5 and 1·0% showing the difference between the two estimators is small. Considering scenarios (ii) and (iii), the oracle-to-sample ratio is nondecreasing in the sample size, subject to the Monte Carlo simulation error, and the oracle-minus-sample ratio falls with the sample size. Considering scenario (iv), the oracle-to-sample ratio falls considerably with a larger variance of the error term. These results show that the theoretical considerations are useful for the sample sizes considered here. Further simulation results show that also for estimating the first order autocorrelation of the cumulated process of μ_t we obtain an oracle-to-sample ratio close to one for scenario (i), even though this is beyond the scope of our theory.

Example 2. Our general model (1), with covariates Z_{it} and parameter θ, can be used in a labour/health study, where health outcomes depend upon macroconditions, i.e., γ_t and η_t, interacting with workplace conditions Z_{it}. When there is a lot of overtime in a cyclical boom, stressful jobs may result in worse health outcomes. In contrast, when recessionary periods increase the fear of unemployment, this may make health worse. Portrait et al. (2010) provides an example, where the cohort process could be modelled as a time series, as in Example 1.

3. Generalized linear time series model

In this section, we introduce our generalization through the link function G:

$$Y_{it} = G\{X_{it}^{\mathrm{T}}\beta + (Z_{it}^{\mathrm{T}}\theta)(R_t^{\mathrm{T}}\gamma + \eta_t)\} + \varepsilon_{it} \quad (t = 1, \ldots, T;\ i = 1, \ldots, I). \tag{6}$$

The function G is a known link function. Again as in § 2, we observe the response Y_{it} and the random covariables X_{it} and Z_{it}. As above, the vectors R_t are deterministic covariates to model the time trend and we put $\mu_t = R_t^{\mathrm{T}}\gamma + \eta_t$ and $\nu_t = \theta\mu_t$.

For the theoretical discussion, we assume that for a weighting function w the estimator $\hat{m}_{it} = X_{it}^{\mathrm{T}}\hat{\beta} + Z_{it}^{\mathrm{T}}\hat{\nu}_t$ is defined by Assumption 6.

Assumption 6. The estimators are approximate solutions of the score equations

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

Examples for estimators that fulfil Assumption 6 are quasilikelihood estimators in generalized linear time series models. For a positive function V the quasilikelihood function is defined as $Q(\tau; y) = \int_\tau^y (s-y)V(s)^{-1}\,ds$ where τ is the expectation of Y, i.e., in our case $\tau = G(X^{\mathrm{T}}\beta + Z^{\mathrm{T}}\nu)$. The quasilikelihood estimator satisfies the two equations in Assumption 6 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 quasilikelihood estimation.

Assumption 7. The functions G and w are twice differentiable and have a bounded second derivative. The weight function w has bounded support.

Assumption 8. It holds that

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

We conjecture that Assumption 7 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 beyond the scope of this paper. In applications we would propose to use the quasilikelihood estimator that corresponds to a weighting function with unbounded support. Assumption 8 is motivated by Assumption 3. For each parameter ν_t one has I observations. This suggests a rate of order $O_p(I^{-1/2})$ which is equal to $o_P(T^{-1/4})$ according to Assumption 3. A formal mathematical theory under which technical Assumption 8 holds is also beyond the scope of this paper.

As in the last section, the time series μ_t and the parameter θ is estimated as in (2). The trend parameter γ is estimated by least squares (3). Again, we consider fits of time series models for η_t that are based on the estimation of its autocovariances $\hat{\rho}_h$ for $h \geqslant 0$, see (4). We compare $\hat{\gamma}$ and $\hat{\rho}_h$ 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 (6).

Theorem 2. *Under Assumptions* 1–8, *it holds that* $\hat{\gamma} = \widetilde{\gamma} + o_P(T^{-1/2})$ *and* $\hat{\rho}_h = \widetilde{\rho}_h + o_P(T^{-1/2})$ $(h \geqslant 0)$.

Example 3. Using (6) one can combine the traditional chain-ladder approach with a well-defined time series analysis of the calendar effect. Consider the case where Y_{it} is the number of claims in an insurance portfolio and where $X_{it}^{\mathrm{T}}\beta$ 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. 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 times the market value of the company. This illustrates the importance of improving the econometric methodology for this problem. Our model above for the first time gives 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 recent extension of the chain-ladder model allowing for calendar effects see Kuang et al. (2008a, 2008b) that derive the nontrivial rules of identification and forecasting in this context.

Example 4. Similar to Example 1, Fitzenberger & Wunderlich (2004) investigate age, time and cohort effects in labour force participation by females. This analysis could be implemented by estimating a generalized linear time series model using a probit or logit link function. Estimating the time series of labour force participation of females helps to analyse the contributions in the pay-as-you-go social security system or the need for child care.

4. Generalized time series regression

In this section, we briefly consider our final and most general model:

$$Y_{it} = G\left\{h_\beta\left(X_{it}\right) + g_\theta\left(Z_{it}\right)\left(R_t^{\mathrm{T}}\gamma + \eta_t\right)\right\} + \varepsilon_{it}. \tag{7}$$

The model is as in (6) but with the extension of our introduction of two parametric families of functionals h_β and g_θ. 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 (6).

The model of Lee & Carter (1992) and Carter & Lee (1992) is a special case of (7). Our more general formulation allows the applied statistician to modify Lee and Carter's original model. 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 modifications of the Lee–Carter structure that are also contained in our model framework. For recent applications to the financial construction of survivor linked bonds, see Blake et al. (2006) and Dowd et al. (2006).

Acknowledgement

We thank two referees, the associate editor, and the editor for very useful comments. The first and the last author acknowledge financial support provided by the Deutsche Forschungsgemeinschaft through the research group Statistical Regularization and Qualitative Constraints. We are grateful to Dirk Antonczyk and Stefanie Schäfer for excellent research assistance in compiling the data used in Example 1.

Appendix

Proof of Theorem 1. First

$$\hat{\beta} - \beta = \left(T^{-1}I^{-1}\sum_{i=1,\ldots,I,t=1,\ldots,T}\widetilde{X}_{it}\widetilde{X}_{it}^{\mathrm{T}}\right)^{-1} T^{-1}I^{-1}\sum_{i=1,\ldots,I,t=1,\ldots,T}\widetilde{X}_{it}\varepsilon_{it}, \tag{A1}$$

$$\hat{\nu}_t - \nu_t = I^{-1}A_t^{-1}\sum_{i=1}^{I} Z_{it}\varepsilon_{it} - I^{-1}A_t^{-1}\sum_{i=1}^{I} Z_{it}X_{it}^{\mathrm{T}}(\hat{\beta}-\beta). \tag{A2}$$

It can be checked that $E(\|\hat\beta-\beta\|^2 \mid \mathcal{X}) = O_P(T^{-1}I^{-1})$ and thus $\|\hat\beta-\beta\| = O_P(T^{-1/2}I^{-1/2})$. Because of Assumptions 4 and 5 this implies that

$$\sup_{t=1,\ldots,T}\left\|\hat\nu_t-\nu_t-I^{-1}A_t^{-1}\sum_{i=1}^{I} Z_{it}\varepsilon_{it}\right\| = O_P(I^{-1/2}T^{-1/2}), \tag{A3}$$

$$T^{-1/2}I^{-1}\sum_{i=1,\ldots,I,t=1,\ldots,T} R_t X_{it}^{\mathrm{T}}(\hat\beta-\beta) = O_P(I^{-1/2}).$$

From (2) we get that $I\sum_{t=1,\ldots,T}(\hat\nu_{t,1}-\hat\mu_t)^2 \leqslant \sum_{i=1,\ldots,I,t=1,\ldots,T}\{Z_{it}^{\mathrm{T}}(\hat\nu_t-\hat\theta\hat\mu_t)\}^2 \leqslant \sum_{i=1,\ldots,I,t=1,\ldots,T}\{Z_{it}^{\mathrm{T}}(\hat\nu_t-\nu_t)\}^2$. This bound and (A3) can be used to show that $\hat\theta-\theta = O_P(I^{-1/2})$ and $T^{-1}\sum_{t=1}^{\mathrm{T}}(\hat\mu_t-\mu_t)^2 = O_P(I^{-1})$. With these expansions, one can approximate the score function that is the derivatives of the left-hand side of (2) with respect to θ and μ_t. After some algebra one gets that $\hat\theta-\theta = O_P(T^{-1/2}+I^{-1})$ where the rate I^{-1} is a quadratic approximation error coming from the above bounds of order $O_P(I^{-1/2})$ for $\hat\theta-\theta$ and $\hat\mu_t-\mu_t$. This bound and the linearized score equation can be used to show that

$$\hat\mu_t-\mu_t = T^{-1/2}I^{-1}\sum_{i=1}^{I} w_{it}\varepsilon_{it} + O_P(\delta_t) \tag{A4}$$

with $w_{it} = (\theta^{\mathrm{T}}Z_{it})/(\theta^{\mathrm{T}}A_t\theta)$ and $\delta_t = (T^{-1/2}I^{-1/2}+I^{-1})(1+\|\hat\nu_t-\nu_t\|+\|\mu_t\|)$. For $\hat\gamma = \widetilde\gamma + o_P(T^{-1/2})$ we have to show

$$T^{-1/2}\sum_{t=1}^{\mathrm{T}} R_t(\hat\mu_t-\mu_t) = o_P(1). \tag{A5}$$

For the proof of this claim note first that, because of (A4), $T^{-1/2}\sum_{t=1}^{\mathrm{T}} R_t(\hat\mu_t-\mu_t) = T^{-1/2}I^{-1}\sum_{i=1,\ldots,I,t=1,\ldots,T} R_t w_{it}\varepsilon_{it} + O_P(I^{-1/2})$. Because of (A1), (A4) and (A5) the right-hand side of this equation is of order $O_P(I^{-1/2})$. This shows (A5). Thus, $\hat\gamma = \widetilde\gamma + o_P(T^{-1/2})$ is shown.

We now show $\hat\rho_h = \widetilde\rho_h + o_P(T^{-1/2})$. We have to show for $h \geqslant 0$ that $T^{1/2}(\hat\rho_h-\widetilde\rho_h) = T^{-1/2}\sum_{t=1}^{\mathrm{T}-h}\{(\hat\mu_t-R_t^{\mathrm{T}}\hat\gamma)(\hat\mu_{t+h}-R_{t+h}^{\mathrm{T}}\hat\gamma)-\eta_t\eta_{t+h}\} = o_P(1)$. For this claim we will show that

$$T^{-1/2}\sum_{t=1}^{\mathrm{T}-h}\eta_{t+h}(\hat\mu_t-R_t^{\mathrm{T}}\hat\gamma-\eta_t) = O_P(T^{-1/2}+I^{-1/2}), \tag{A6}$$

$$T^{-1/2}\sum_{t=1}^{\mathrm{T}-h}\eta_t(\hat\mu_{t+h}-R_{t+h}^{\mathrm{T}}\hat\gamma-\eta_{t+h}) = O_P(T^{-1/2}+I^{-1/2}), \tag{A7}$$

$$T^{-1/2}\sum_{t=1}^{\mathrm{T}-h}(\hat\mu_{t+h}-R_{t+h}^{\mathrm{T}}\hat\gamma-\eta_{t+h})(\hat\mu_t-R_t^{\mathrm{T}}\hat\gamma-\eta_t) = O_P(T^{-1/2}+I^{-1/2}). \tag{A8}$$

Because of (A4), we have that uniformly for $t=1,\ldots,T$, $\hat\mu_t-R_t^{\mathrm{T}}\hat\gamma-\eta_t = I^{-1}\sum_{i=1}^{I} w_{it}\varepsilon_{it} - R_t^{\mathrm{T}}(\hat\gamma-\gamma)+O_P(\delta_t)$. Using this expansion, it can be easily seen that (A6) and (A7) follow from Assumption 3 and $\hat\gamma-\gamma = O_P(T^{-1/2})$, $T^{-1/2}I^{-1}\sum_{t=1}^{\mathrm{T}-h}\sum_{i=1}^{I}\eta_t w_{it}\varepsilon_{it+h} = O_P(I^{-1/2}T^{-1/2})$ and $T^{-1/2}I^{-1}\sum_{t=1}^{\mathrm{T}-h}\sum_{i=1}^{I}\eta_{t+h}w_{it}\varepsilon_{it} = O_P(I^{-1/2}T^{-1/2})$. The latter expansions follow from Assumption 1.

For $\hat\rho_h = \widetilde\rho_h + o_P(T^{-1/2})$ it remains to check (A8). This can be done by showing that $T^{-1/2}\sum_{t=1}^{\mathrm{T}-h}R_t^{\mathrm{T}}(\hat\gamma-\gamma)R_{t+h}^{\mathrm{T}}(\hat\gamma-\gamma) = O_P(T^{-1/2})$, $T^{-1/2}I^{-1}\sum_{t=1}^{\mathrm{T}-h}\sum_{i=1}^{I}\varepsilon_{it}R_{t+h}^{\mathrm{T}}(\hat\gamma-\gamma) = O_P(I^{-1/2}T^{-1/2})$, $T^{-1/2}I^{-1}\sum_{t=1}^{\mathrm{T}-h}\sum_{i=1}^{I}\varepsilon_{it+h}R_t^{\mathrm{T}}(\hat\gamma-\gamma) = O_P(I^{-1/2}T^{-1/2})$ and $T^{-1/2}I^{-2}\sum_{t=1}^{\mathrm{T}-h}\sum_{i=1}^{I}\sum_{j=1}^{I}\varepsilon_{it+h}\varepsilon_{jt} = O_P(I^{-1})$. These expansions can be shown by using Assumptions 1 and 5. □

Proof of Theorem 2. By expanding the score functions in Assumption 6 we get stochastic expansions of $\hat\beta-\beta$ and $\hat\nu_t-\nu_t$ where the first terms are weighted modifications of the right-hand side of (A1) or (A2), respectively. These expansions are of order $o_P(T^{-1/2})$. To get these expansions, one applies that $\|\hat\beta-\beta\|^2$

and $\|\hat{\nu}_t - \nu_t\|^2$ are of order $o_P(T^{-1/2})$, see Assumption 8. The further proof of Theorem 2 can be carried out by similar arguments as in the proof of Theorem 1. □

REFERENCES

BAI, J. (2009). Panel data models with interactive fixed effects. *Econometrica* **77**, 1229–79.

BAI, J. & NG, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica* **74**, 1133–50.

BLAKE, D., CAIRNS, A. J. G., DOWD, K. & MACMINN, R. (2006). Longevity bonds: financial engineering, valuation and hedging. *J. Risk Insur.* **73**, 647–72.

BROUHNS, N., DENUIT, M. & VERMUNT, J. K. (2002). A Poisson log-bilinear approach to the construction of projected lifetables. *Insur. Math. Econ.* **31**, 373–93.

CAIRNS, A. J. G., BLAKE, D., DOWD, K., COUGHLAN, G. D., EPSTEIN, D., ONG, A. & BALEVICH, I. (2009). A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *N. Am. Actuar. 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. *Quart. J. Econ.* **116**, 705–46.

CARTER, L. R. & LEE, R. D. (1992). Modelling and forecasting U.S. sex differentials in mortality. *Int. J. Forecasting* **8**, 393–411.

DOWD, K., BLAKE, D., CAIRNS, A. J. G. & DAWSON, P. (2006). Survivor swaps. *J. Risk Insur.* **73**, 1–17.

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

FITZENBERGER, B. & WUNDERLICH, G. (2004). The changing life cycle pattern in female employment: a comparison of Germany and the UK. *Scottish J. Polit. Econ.* **51**, 302–28.

FITZENBERGER, B. & WUNDERLICH, G. (2002). Gender wage differences in West Germany: a cohort analysis. *German Econ. Rev.* **3**, 379–414.

FITZENBERGER, B., HUJER, R., MCCURDY, T. E. & SCHNABEL, R. (2001). Testing for uniform wage trends in West Germany: a cohort analysis using quantile regressions for censored data. *Empirical Econ.* **26**, 41–86.

KUANG, D., NIELSEN, B. & NIELSEN, J. P. (2008a). Forecasting with the age-period-cohort model and the extended chain-ladder model. *Biometrika* **95**, 987–91.

KUANG, D., NIELSEN, B. & NIELSEN, J. P. (2008b). Identification of the age-period-cohort model and the extended chain-ladder model. *Biometrika* **95**, 979–86.

LEE, R. D. & CARTER, L. R. (1992). Modelling and forecasting U.S. mortality. *J. Am. Statist. Assoc.* **87**, 659–71.

LEE, R. D. & MILLER, T. (2001). Evaluating the performance of the Lee–Carter method for forecasting mortality. *Demography* **8**, 537–49.

LI, S. & CHAN, W. (2005). Outlier analysis and mortality forecasting: the United Kingdom and Scandinavian countries. *Scand. Actuar. J.* **3**, 187–211.

LINTON, O., NIELSEN, J. P. & NIELSEN, S. F. (2009). Nonparametric regression with a latent time series. *Economet. J.* **12**, 187–207.

PARK, B., MAMMEN, E., HÄRDLE, W. & BORAK, S. (2009). Time series modelling with semiparametric factor dynamics. *J. Am. Statist. Assoc.* **104**, 284–98.

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**, 617–42.

RENSHAW, A. E. & HABERMAN, S. (2003a). On the forecasting of mortality reduction factors. *Insur. Math. Econ.* **2**, 379–401.

RENSHAW, A. E. & HABERMAN, S. (2003b). On the forecasting of mortality reduction factors. *Appl. Statist.* **2**, 119–37.

WONG-FUPUY, C. & HABERMAN, S. (2004). Projection mortality trends: recent developments in the United Kingdom and the United States. *N. Am. Actuar. J.* **8**, 56–83.

[*Received December* 2008. *Revised June* 2011]