Two sided analysis of variance with a latent time series

Lars Hougaard Hansen
University of Copenhagen and Codan forsikring
Bent Nielsen
Nuffield College, University of Oxford
Jens Perch Nielsen
Codan forsikring

18 October 2004

Abstract: Many real life regression problems exhibit some kind of calender time dependency and it is often of interest to predict the behavior of the regression function along this calender time direction. This can be formulated as a regression model with an added latent time series and the task is to be able to analyse this series. In this paper we engage this through a two step procedure, firstly we treat the time dependent elements as parameters and estimate them in the two-sided analysis of variance setup, secondly we use the estimated time series as predictor of the latent time series. An application to risk theory is discussed.

Key Words: regression, time series, risk theory.

1 Introduction

We start by defining the standard linear regression model with common slope and different intercept in each group, where the groups correspond to changing calender years. This model is well studied and estimation can be carried out by Ordinary Least Squares. If we reformulate the model such that the intercept term in each year is stochastic, we can use the intercept estimates from the linear model as predictions for the latent time series. We wish to draw inference about the parameters in the model for the time series, and the theory of the estimator for the first model gives us a starting point for this task.

The obvious way to predict the latent time series would be to use conditional expectations in some form but these are often difficult to calculate under general assumptions and, more importantly, they require a fixed model for the time series. Our method enables separating the time series analysis from the prediction of the series, and gives the opportunity to wait deciding on the distribution of the time series until one has the predicted series.

Suppose an unbalanced two-sided array of data is observed with a dependent variable Y_{ti} and a q-dimensional covariate X_{ti} so $t = 1, \ldots T$ is a time index while the number of individuals $i = 1, \ldots, n_t$ vary over time. A standard two-sided analysis of variance model can be formulated as

$$Y_{ti} = \alpha_t + X_{ti}' \gamma + e_{ti},$$

where the error terms, e_{ti}, are independent over individuals, i, and time t, and identically distributed with mean zero and variance η^2. The least squares estimators for α_t and γ are then found by minimising the sum of squared deviations

$$\sum_{t=1}^{T}\sum_{i=1}^{n_t}\left\{\left(Y_{ti}-\overline{Y}_t\right)-\left(X_{ti}-\overline{X}_t\right)'\gamma\right\}^2+\sum_{t=1}^{T}n_t\left(\overline{Y}_t-\overline{X}_t'\gamma-\alpha_t\right)^2,$$

where $\overline{Y}_t$ and $\overline{X}_t$ represent averages over individuals with a common time index t. This gives the estimators

$$\hat{\gamma}=\left\{\sum_{t=1}^{T}\sum_{i=1}^{n_t}\left(X_{ti}-\overline{X}_t\right)\left(X_{ti}-\overline{X}_t\right)'\right\}^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n_t}\left(X_{ti}-\overline{X}_t\right)Y_{ti},\qquad \hat{\alpha}_t=\overline{Y}_t-\overline{X}_t'\hat{\gamma}.$$

The total number of observations is denoted $n=\sum_{t=1}^{T}n_t$.

We will now formulate a latent time series regression model for the same array of data. This is formulated as

$$Y_{ti}=\mu_t+X_{ti}'\beta+\varepsilon_{ti}, \tag{1}$$

where $\mu_1,\ldots,\mu_T$ is the latent time series. As in the two-sided analysis of variance model we estimate the regression coefficient β and the latent time series μ_t by

$$\hat{\beta}=\left\{\sum_{t=1}^{T}\sum_{i=1}^{n_t}\left(X_{ti}-\overline{X}_t\right)\left(X_{ti}-\overline{X}_t\right)'\right\}^{-1}\sum_{t=1}^{T}\sum_{i=1}^{n_t}\left(X_{ti}-\overline{X}_t\right)Y_{ti},\qquad \hat{\mu}_t=\overline{Y}_t-\overline{X}_t'\hat{\beta}.$$

Imagining a situation where T is large and n_t even larger we will show three types of results: (i) The regression function can be estimated and analysed in the same way as if the time component had been deterministic. (ii) The latent time series can be estimated very accurately. (iii) Oracle efficiency: the estimated time series can be analysed as the time series itself, had it be given by an oracle.

2 Analysis of the latent time series regression model

The three types of results for the latent time series regression model are now discussed in detail.

To formulate the conditions precisely, let $\|\cdot\|$ be the spectral norm so $\|A\|^2$ equals the maximal eigenvalue of $A'A$ for a matrix A and reduces to $A'A$ for a vector A.

2.1 The regression estimator

The first result shows that the regression estimator $\hat{\beta}$ is asymptotically normal in similar way to what would arise in a two-sided analysis of variance model. The result is formulated for large sample length of the time series, $T\to\infty$, in terms of a Liapounov Central Limit Theorem.

The essence of the conditions is that the innovations ε_{ti}, given the regressors, have a standardised distribution, while the regressors X_{ti} are allowed to vary over both indices as long as the information is spread across the individuals and over time. Importantly, the limiting distribution of $\hat{\beta}$ does not depend on neither the nature of the latent time series nor is the number of individuals at each time point, n_t, required to increase with T.

Theorem 1 *Suppose that for some sequence a_T depending on T, it holds that*

(i) *the variables satisfy*

> (a) *the arrays $(\varepsilon_{ti}, X_{ti},$ for $i = 1, \ldots, n_t)$ are independent over t,*
>
> (b) *the pairs $(\varepsilon_{ti}, X_{ti})$ are independent over i,*

(ii) *the innovations satisfy, for some $k < \infty$,*
(a) $\mathsf{E}(\varepsilon_{ti}|X_{ti}) = 0$, $\quad (b)$ $\mathsf{Var}(\varepsilon_{ti}|X_{ti}) = \sigma^2$, $\quad (c)$ $\max_{t,i} \mathsf{E}(\|\varepsilon_{ti}\|^4 |X_{ti}) < k$.

(iii) *the regressors satisfy, for a positive definite matrix Σ,*

> (a) $a_T^{-2} \mathsf{E} \sum_{t=1}^T \sum_{i=1}^{n_t} \left(X_{ti} - \overline{X}_t\right) \left(X_{ti} - \overline{X}_t\right)' \rightarrow \Sigma$,
>
> (b) $a_T^{-2} \sum_{t=1}^T \sum_{i=1}^{n_t} \left(X_{ti} - \overline{X}_t\right) \left(X_{ti} - \overline{X}_t\right)' \xrightarrow{\mathsf{P}} \Sigma$,
>
> (c) $a_T^{-4} \mathsf{E} \sum_{t=1}^T n_t \sum_{i=1}^{n_t} \left\|X_{ti} - \overline{X}_t\right\|^4 \rightarrow 0$.

Then, for $T \rightarrow \infty$,

$$a_T \left(\hat{\beta} - \beta\right) \xrightarrow{\mathsf{P}} \mathsf{N}\left(0, \sigma^2 \Sigma^{-1}\right).$$

2.2 Prediction of the time series

We will seek to apply time series analysis to the estimated time series $\hat{\mu}_1, \ldots, \hat{\mu}_T$ with a view to predicting future values of μ_t. This will be possible when the estimated time series

$$\hat{\mu}_t = \overline{Y}_t - \overline{X}_t' \hat{\beta} = \mu_t + \overline{X}_t' \left(\beta - \hat{\beta}\right) + \overline{\varepsilon}_t \tag{2}$$

is close to latent time series μ_t uniformly over time. This issue is investigated in Theorem 2 below. Noting that $\hat{\beta} - \beta$ is invariant to changes in the value of β the predicted time series $\hat{\mu}_t$ inherits this property.

In addition to the assumptions of Theorem 1 the Theorem 2 concerning $\hat{\mu}_t$ requires that the number of individuals n_t at each point in time growths faster than T.

Theorem 2 *Suppose the assumptions of Theorem 1 are satisfied, and that,*

(iv) $\max_t(n_t^{-1}) = \mathrm{o}(T^{-1})$,

(v) $a_T^{-2} \mathsf{E} \sum_{t=1}^T n_t^{-1} \sum_{i=1}^{n_t} \|X_{ti}\|^2 \rightarrow 0$.

Then, for $T \to \infty$,

$$\sum_{t=1}^{T} (\hat{\mu}_t - \mu_t)^2 = \mathsf{o}_{\mathsf{P}}(1).$$

2.3 Analysis of the predicted time series

As an example of a time series model for μ_t consider the first order autoregression

$$\mu_t = \rho\mu_{t-1} + \xi_t \qquad (t = 2, \ldots, T) \tag{3}$$

conditional on μ_1, where the innovations ξ_t constitute a martingale difference sequence with variance ω^2. Had the latent time series μ_t been observed statistical analysis would typically be based on the standardised least squares estimator

$$\mathsf{t} = \frac{\sum_{t=2}^{T} \xi_t \mu_{t-1}}{\left(\sum_{t=2}^{T} \mu_{t-1}^2\right)^{1/2}}. \tag{4}$$

The asymptotic properties of this statistic are well known: when $|\rho| < 1$ it is normal distributed due to a Central Limit Theorem argument for martingale differences, see Hall and Heyde (1980, p.172), when $|\rho| = 1$ it has a non-standard distribution that can be represented using Brownian motions, see Dickey-Fuller (1979), whereas when $|\rho| > 1$ and the innovations are independent and normal it is asymptotically normal distributed, see Anderson (1959).

The parameters ρ, ω^2 can be estimated by least squares estimators based on the estimated series $\hat{\mu}_t$ given as

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \hat{\mu}_t \hat{\mu}_{t-1}}{\sum_{t=2}^{T} \hat{\mu}_{t-1}^2}, \qquad \hat{\omega}^2 = \frac{1}{T-1}\left\{\sum_{t=2}^{T} \hat{\mu}_t^2 - \frac{\left(\sum_{t=2}^{T} \hat{\mu}_t \hat{\mu}_{t-1}\right)^2}{\sum_{t=2}^{T} \hat{\mu}_{t-1}^2}\right\}.$$

The following result shows that for very weak assumptions to the innovation process the estimators $\hat{\rho}, \hat{\omega}^2$ will have properties similar to what could have been achieved had the latent time series been given to us by an oracle. In particular the series $\hat{\mu}_t$ can be analysed well regardless of whether the latent time series is stationary, or non-stable.

Theorem 3 *Suppose, as stated in Theorem* 2, *that*

$$\sum_{t=1}^{T} (\hat{\mu}_t - \mu_t)^2 = \mathsf{o}_{\mathsf{P}}(1).$$

Suppose, further, that (ξ_t) is a martingale difference sequence with respect to a filtration $(\mathcal{F}_t)$ satisfying $\mathsf{E}(\xi_t^2|\mathcal{F}_{t-1}) = \omega^2$ and $\max_t \mathsf{E}(|\xi_t|^3|\mathcal{F}_{t-1}) < \infty$. Then, for any $\rho \in \mathbf{R}$ and

$T \to \infty$ *it holds*

$$\begin{aligned}
\textstyle\sum_{t=2}^{T} \hat{\mu}_{t-1}^2 &= \textstyle\sum_{t=2}^{T} \mu_{t-1}^2 \left\{1 + \mathrm{o}_{\mathsf{P}}(1)\right\}, \\
\hat{\mathsf{t}} = \left(\textstyle\sum_{t=2}^{T} \hat{\mu}_{t-1}^2\right)^{1/2} (\hat{\rho} - \rho) &= \frac{\sum_{t=2}^{T} \xi_t \mu_{t-1}}{\left(\sum_{t=2}^{T} \mu_{t-1}^2\right)^{1/2}} \left\{1 + \mathrm{o}_{\mathsf{P}}(1)\right\} + \mathrm{o}_{\mathsf{P}}(1), \\
\hat{\omega}^2 &= \frac{1}{T} \textstyle\sum_{t=2}^{T} \xi_t^2 + \mathrm{o}_{\mathsf{P}}(1).
\end{aligned}$$

3 Risk Theory

The model presented above has an application to Risk Theory. In the standard setting of Risk Theory, see Bühlmann (1970), Beard, Pentikainen and Pesonen (1984) or Norberg (1990), the focus concentrates on the investigation of the claim process,

$$C_\tau = \sum_{i=1}^{N_\tau} Z_i, \qquad (\tau \geq 0),$$

where N_τ is a counting process representing the number of claims occurred in a time interval $[0; \tau]$ and Z_i is the size of the ith claim. In practice, it is of interest to study the growth of claim sizes in order to predict future liabilities. It is therefore natural to include a time dependency in the model for Z_i to describe this pattern, and this is where the latent time series regression model proves its worth. Grouping the claims in years, gives us the opportunity to use the setup from the previous section,

$$C_\tau^* = \sum_{t=1}^{\tau} \sum_{i=1}^{n_t} Y_{ti}, \qquad (\tau \in \mathbf{N}_0),$$

where n_t is the number of claims in year t and, Y_{ti} is the size of the ith claim in the tth year, following the model defined by (1) and (3).

A slight modification has to be made to Theorems 1 and 2 when using the latent time series regression model in Risk Theory, to allow for stochastic behavior of n_t. The following holds.

Theorem 4 *Suppose that the Assumptions* (i, a), (iii), (v) *hold, while Assumptions* (i, b), (ii) *are modified to hold conditionally on* n_t, *and* (iv) *is replaced by*

(iv') $\max_t \mathsf{E}(n_t^{-1} | n_t > 0) = \mathrm{o}(T^{-1})$.

Then, for $T \to \infty$,

$$\sqrt{a_T}\left(\hat{\beta} - \beta\right) \xrightarrow{\mathsf{P}} \mathsf{N}\left(0, \sigma^2 \Sigma\right) \qquad \textit{and} \qquad \sum_{t=1}^{T} \left(\hat{\mu}_t - \mu_t\right)^2 = \mathrm{o}_{\mathsf{P}}(1).$$

The assumptions of Theorem 4 can be shown to be valid in many specific situations. In many Risk Theory models it is assumed that the number of individuals is $\mathsf{Poisson}(\lambda_t)$-distributed. Assuming that $\max \lambda_t^{-1} = \mathrm{o}\left(T^{-1}\right)$ it is clear that Assumption (iv') is satisfied. To see this, use Jensen's inequality,

$$\mathsf{E}\left(n_t^{-1}|n_t>0\right) \leq \left\{\mathsf{E}\left(n_t|n_t>0\right)\right\}^{-1} = \frac{1-\exp\left(-\lambda_t\right)}{\lambda_t} \leq \frac{1}{\lambda_t} = \mathrm{o}\left(T^{-1}\right).$$

4 Illustration

The asymptotic theory is now illustrated by a simulation, where the covariates X_{ti} are chosen to be trending over time. We will assume that $T=20$ periods are considered and the numbers of individuals, n_t, are independently $\mathsf{Poisson}\left(\lambda\right)$-distributed where $\lambda = 100$, so it is not unreasonable to assume that $\lambda^{-1} = \mathrm{o}\left(T^{-1}\right)$. The regressors X_{ti} are assumed to be $\Gamma\left(\delta_t, 1\right)$-distributed where the shape parameter $\delta_t = 1 + t/10$ growths linearly over time and the regression coefficient is chosen to be $\beta = 0$ due to the invariance with respect to β. The conditional distribution of the errors ε_{ti} given regressors is assumed to be $\mathsf{N}(0, 3^2)$, while the innovations ξ_t are $\mathsf{N}(0, 1^2)$ and thus have much smaller variation than the errors, ε_{ti}. Finally, assume an autoregressive coefficient of unity, $\rho = 1$.

In this situation the Assumptions $(i), (ii), (iv')$ of Theorem 4 are trivially met. Choosing $a_T^2 = \lambda T^2$ and $T^{-2}\sum_{t=1}^{T}\delta_t \to \Sigma = 1/20$ some tedious calculations presented in the Appendix show that Assumptions (iii) and (v) are met.

Table 1. *Simulation results*

$\lambda-$value	10	20	40	80	160
$\widehat{\mathsf{E}}\sum_{t=1}^{T}\left(\hat{\mu}_t - \mu_t\right)^2$	22·61	10·54	5·16	2·55	1·26
$\widehat{\mathsf{E}}\sum_{t=1}^{T}\overline{\varepsilon}_t^2$	20·44	9·50	4·62	2·28	1·14
$\widehat{\mathsf{E}}(\widehat{\mathsf{t}}-\mathsf{t})^2$	3·97	1·20	0·362	0·127	0·0498

Table 1 reports the results of the simulation study. For each of five different values of the Poisson parameter λ the expectation of three statistics were simulated using 5000 repetitions. For the smallest value, $\lambda = 10$, it happened 6 times that n_t took the value 0 which was then replaced by 1.

The first row of Table 1 investigates the convergence $\sum_{t=1}^{T}\left(\hat{\mu}_t - \mu_t\right)^2 = \mathrm{o}_{\mathsf{P}}(1)$ stated in Theorem 4. The expectation of this term appear to be of order λ^{-1}. Even for a modest value of λ it is small compared to the expected sum of squared innovations of the latent time series, which is $\mathsf{E}\sum_{t=1}^{T}\xi_t^2 = 20$. The proof of the convergence result uses the decomposition $\hat{\mu}_t - \mu_t = \overline{X}_t'(\beta - \hat{\beta}) + \overline{\varepsilon}_t$, see (2). While both terms vanish a comparison of the first two rows of Table 1 indicates that the term $\overline{\varepsilon}_t$ is dominating in accordance with the proof.

In the third row of Table 1 the convergence of Theorem 3 is investigated. The mean square error $\widehat{\mathsf{E}}(\widehat{\mathsf{t}}-\mathsf{t})^2$ is simulated and reported in Table 1. This statistic also appears to be vanishing fast. Once again, it is relatively small compared to the expectation $\mathsf{E}\mathsf{t}^2$, in that the limiting distribution of t^2 has expectation of about 1.142, see Nielsen (1997). When λ is larger than about 40 and thus somewhat larger than T this mean square error is therefore of relatively minor importance.

5 Discussion

We have introduced a latent time series regression model and extended it to a specific risk theoretical setting. While the concept of a latent time series is new in risk theory, it has some history in the area of mortality estimation and in panel data analysis.

It was introduced in mortality estimation in a widely cited paper by Lee and Carter (1992). First, they estimate how the mortality depends of three non-parametrically specified functions, two depending on age and one depending on calendar time. This, their first step, is rather similar to our first step estimating our two components from a standard analyses of variance. Afterwards, they analyse their non-parametric function depending on calendar time as a simple autoregressive function much like we investigate our calendar effect as a time series. The major contribution of this paper is to consider the latent time series to be part of the original model formulation and to formulate standard theorems of mathematical statistics to investigate the properties of such a model. One interesting future line of research could be to treat the important mortality model of Lee and Carter in a similar way. Another interesting future direction of research could be to investigate the possibility of including such a latent time series in risk theoretic models of outstanding insurance liabilities based on aggregated data, see England and Verrall (2000) for a review of such models.

The proposed model is also related to some recent latent time series models for balanced panel data, so n_t is constant over t. Forni, Hallin, Lippi and Reichlin (2000) consider a model of the form

$$Y_{ti} = \gamma_i(L)\,u_t + \varepsilon_{ti},$$

so $\gamma_i(L)\,u_t$ is a moving average process where the lag polynomials $\gamma_i(L)$ have coefficients depending on i while the innovations u_t are common for all individuals. They propose an estimator for the latent component $\gamma_i(L)\,u_t$. In the simple situation where $\gamma_i(L)=1$ this is done in a similar way to our formulation of $\hat{\mu}_t$ by estimating u_t by $\overline{Y}_t$. Pesaran (2003) looks at the model

$$Y_{ti} = \alpha_i + \beta_i Y_{t-1,i} + \gamma_i f_t + \varepsilon_{ti}.$$

and analyses the regression function $\alpha_i + \beta_i Y_{t-1,i}$ with a view to testing the joint unit root hypothesis $\beta_i = 1$ for all i. This is done in a similar way to our model in that the latent time series $\gamma_i f_t$ is replaced by $c_i + d_i \overline{Y}_{t-1}$ where c_i and d_i are auxiliary parameters. The results presented in this paper could possibly inspire the formulation of theoretical properties of these more complicated models.

6 Acknowledgements

Discussion with D.R. Cox is gratefully acknowledged. All simulations has been performed using R 1.7.1.

A Proofs

Proof of Theorem 1. Consider the triangular array $y_{Tt} = a_T^{-1} \sum_{i=1}^{n_t} \left(X_{ti} - \overline{X}_t\right) \varepsilon_{ti}$ and show that the Liapounov conditions of Davidson (1994, p. 373) are satisfied:
(1) By Assumption (i, a) then y_{Tt} is independent over t.
(2) By Assumptions (i, b), (ii, a) then $\mathsf{E}\left(y_{Tt}\right) = 0$.
(3) By Assumptions (i, b), (ii, b), (iii, a) then $s_T^2 = \mathsf{Var}\left(y_{Tt}\right) \to \sigma^2 \Sigma$.
(4) Note first, that with $z_{ti} = X_{ti} - \overline{X}_t$ it holds

$$a_T^4 \mathsf{E}\left\|y_{T,t}\right\|^4 = \mathsf{E}\left\|\sum_{i=1}^{n_t} z_{ti}\varepsilon_{ti}\right\|^4 = \mathsf{E}\left(\sum_{i,j} \varepsilon_{ti}\varepsilon_{tj} z_{tj}' z_{ti}\right)^2 .$$

Conditioning on the regressors and n_t and using Assumption (ii) this is bounded by

$$a_T^4 \mathsf{E}\left\|y_{T,t}\right\|^4 \leq k \sum_{i=1}^{n_t} \left\|z_{ti}\right\|^4 + 2\sigma^2 \sum_{i \neq j} \left\|z_{ti}\right\|^2 \left\|z_{tj}\right\|^2 .$$

By Jensen's inequality this in turn is bounded, for some $K < \infty$, by

$$a_T^4 \mathsf{E}\left\|y_{T,t}\right\|^4 \leq K n_t \sum_{i=1}^{n_t} \left\|z_{ti}\right\|^4 .$$

Thus the Liapounov condition $\sum_{t=1}^T \mathsf{E}\left\|y_{T,t}\right\|^4 \to 0$ is met under Assumption (iii, c). ∎

Proof of Theorem 2. Due to the expression (2) and Markov's inequality it suffices to argue that $\sum_{t=1}^T \overline{\varepsilon}_t^2$ and $(\beta - \hat{\beta})' \sum_{t=1}^T \overline{X}_t \overline{X}_t' (\beta - \hat{\beta})$ vanish.

To show that $\sum_{t=1}^T \overline{\varepsilon}_t^2$ vanish use Chebychev's inequality to see that

$$\mathsf{P}\left(\sum_{t=1}^T \overline{\varepsilon}_t^2 > \delta\right) \leq \delta^{-1} \mathsf{E} \sum_{t=1}^T \overline{\varepsilon}_t^2 = \delta^{-1} \sum_{t=1}^T \mathsf{E}\overline{\varepsilon}_t^2 .$$

It therefore suffices to show $\mathsf{E}\overline{\varepsilon}_t^2 = \mathrm{o}\left(T^{-1}\right)$. Using first Assumption (i, b) and then the bound to n_t in Assumption (iv) it holds

$$\mathsf{E}\overline{\varepsilon}_t^2 = \mathsf{E}\sigma^2 / n_t = \mathrm{o}\left(T^{-1}\right).$$

It follows from Theorem 1 that $(\hat{\beta} - \beta) = \mathrm{O}_{\mathsf{P}}\left(a_T^{-1}\right)$ so it suffices to show that $a_T^{-2} \sum_{t=1}^T \overline{X}_t \overline{X}_t'$ vanishes. Using first Chebychev's inequality, then twice the triangle inequality, and Jensen's inequality

$$\begin{aligned} \delta \mathsf{P}\left(\left\|\sum_{t=1}^T \overline{X}_t \overline{X}_t'\right\| > \delta\right) &\leq \mathsf{E}\left\|\sum_{t=1}^T \overline{X}_t \overline{X}_t'\right\| \leq \mathsf{E} \sum_{t=1}^T \left\|\overline{X}_t \overline{X}_t'\right\| = \sum_{t=1}^T \mathsf{E}\left\|\overline{X}_t\right\|^2 \\ &\leq \sum_{t=1}^T \mathsf{E}\left(\frac{1}{n_t} \sum_{i=1}^{n_t} \left\|X_{t,i}\right\|\right)^2 \leq \sum_{t=1}^T \mathsf{E} \frac{1}{n_t} \sum_{i=1}^{n_t} \left\|X_{t,i}\right\|^2 . \end{aligned}$$

This is of order $\mathrm{o}(a_T^2)$ as desired due to Assumption (v). ■

Proof of Theorem 3. The key to this result is to show

$$\left(\sum_{t=2}^{T} \mu_t^2\right)^{-1} = \mathrm{O}_{\mathsf{P}}(T^{-1}) \qquad \text{and} \qquad \sum_{t=2}^{T} \xi_t^2 = \mathrm{O}_{\mathsf{P}}(T). \tag{5}$$

For the stated assumptions to ξ this holds for any $\rho \in \mathbf{R}$ according to Lai and Wei (1983, Theorems 1 and 3).

It is first shown that $\sum_{t=2}^T \hat{\mu}_{t-1}^2 = \sum_{t=2}^T \mu_{t-1}^2 \{1 + \mathrm{o}_{\mathsf{P}}(1)\}$. Let $d_t = \hat{\mu}_t - \mu_t$ and use Cauchy-Schwarz's inequality to see

$$\sum_{t=2}^T \hat{\mu}_{t-1}^2 \leq \sum_{t=2}^T \mu_{t-1}^2 + 2\left(\sum_{t=2}^T \mu_{t-1}^2 \sum_{t=2}^T d_{t-1}^2\right)^{1/2} + \sum_{t=2}^T d_{t-1}^2$$

Theorem 2 and the first property in (5) give the desired result.

Turning to the first expression of interest the numerator satisfies

$$\sum_{t=2}^T \left(\hat{\mu}_t - \rho\hat{\mu}_{t-1}\right)\hat{\mu}_{t-1} = \sum_{t=2}^T \left(\xi_t + d_t - \rho d_{t-1}\right)\left(\mu_{t-1} + d_{t-1}\right) = \sum_{t=2}^T \xi_t \mu_{t-1} + R,$$

where R is some remainder term. By the triangle inequality and Cauchy-Schwarz's inequality R is found to be of the order

$$\left(\sum_{t=2}^T \xi_t^2 \sum_{t=2}^T d_{t-1}^2\right)^{1/2} + (1+|\rho|)\left(\sum_{t=2}^T \mu_{t-1}^2 \sum_{t=1}^T d_t^2\right)^{1/2} + (1+|\rho|)\sum_{t=1}^T d_t^2.$$

Theorem 2 and (5) then show $R = \mathrm{o}_{\mathsf{P}}(\sum_{t=1}^T \mu_t^2)^{1/2}$. Together with the result for denominator shown above this implies the desired result.

The estimator $\hat{\omega}^2$ satisfies

$$\begin{aligned}(T-1)\hat{\omega}^2 &= \sum_{t=2}^T \left(\hat{\mu}_t - \rho\hat{\mu}_{t-1}\right)^2 - \frac{\left\{\sum_{t=2}^T \left(\hat{\mu}_t - \rho\hat{\mu}_{t-1}\right)\hat{\mu}_{t-1}\right\}^2}{\sum_{t=2}^T \hat{\mu}_{t-1}^2} \\ &= \sum_{t=2}^T \left(\hat{\mu}_t - \rho\hat{\mu}_{t-1}\right)^2 - \left(\hat{\rho} - \rho\right)^2 \sum_{t=2}^T \hat{\mu}_{t-1}^2.\end{aligned}$$

The first term equals $\sum_{t=2}^T \left(\xi_t + d_t - \rho d_{t-1}\right)^2$ which in turn equals $\sum_{t=2}^T \xi_t^2 + \mathrm{o}_{\mathsf{P}}(T)$ by an argument as that for the numerator of the first expression. The second term is asymptotically equivalent to the statistic (4) due to the first part of this Theorem, which in turn is of order $\mathrm{o}_{\mathsf{P}}(T)$ according to Nielsen (2001, Lemma A3). ■

Proof of Theorem 4. The proofs of Theorems 1, 2 can be adapted:

Asymptotic distribution of $\hat{\beta}$.

(a) It suffices to argue that the bound to $\mathsf{E}\overline{\varepsilon}_t^2$ holds. By iterated expectations

$$\mathsf{E}\overline{\varepsilon}_t^2 = \mathsf{E}\left(\overline{\varepsilon}_t^2 | n_t > 0\right)\mathsf{P}\left(n_t > 0\right) \leq \mathsf{E}\left(\overline{\varepsilon}_t^2 | n_t > 0\right) = \mathsf{E}\left\{\mathsf{E}\left(\overline{\varepsilon}_t^2 | n_t > 0, n_t\right) | n_t > 0\right\}.$$

This equals $\sigma^2 \mathsf{E}\left\{n_t^{-1}|n_t > 0\right\}$ due to Assumptions $(i')\,,(ii',b)$. This is in turn bounded by $T^{-(1+a)}$ according to assumption (iv').

(b) This argument can be done in the same way as (a) by conditioning on n_t.

(c) Same argument as in the proof of Theorem 2.

(d) The argument concerning the bound to $\mathsf{E}\overline{X}_t^2$ has to be adapted along the lines of (a). ■

Finally it is argued that the Assumptions (iii) and (v) are met in the illustration of Section 4. First note that

$$\mathsf{E}n_t = \lambda, \qquad \mathsf{E}n_t^2 = \lambda^2 + \lambda,$$

$$\mathsf{E}\left(X_{ti}|n_t\right) = \delta_t, \qquad \mathsf{E}\left(X_{ti}^2|n_t\right) = \delta_t^2 + \delta_t,$$

$$\mathsf{E}\left\{\left.(X_{ti} - \delta_t)^2\right| n_t\right\} = \delta_t, \qquad \kappa_t = \mathsf{E}\left\{\left.(X_{ti} - \delta_t)^4\right| n_t\right\} = 3\delta_t^2 + 6\delta_t.$$

Noting that $n_t\left(X_{ti} - \overline{X}_t\right) = (n_t - 1)\left(X_{ti} - \delta_t\right) - \sum_{k\neq i}\left(X_{tk} - \delta_t\right)$ it is seen that

$$\begin{aligned}
\mathsf{E}\left\{\left.\left(X_{ti} - \overline{X}_t\right)^2\right| n_t\right\} &= \delta_t\left\{1 + \mathrm{O}\left(n_t^{-1}\right)\right\}, \\
\mathsf{E}\left\{\left.\left(X_{ti} - \overline{X}_t\right)^4\right| n_t\right\} &= \kappa_t\left\{1 + \mathrm{O}\left(n_t^{-1}\right)\right\} + \delta_t^2 \mathrm{O}\left(n_t^{-1}\right),
\end{aligned}$$

while using that $\left(X_{ti} - \overline{X}_t\right) = \left(X_{ti} - \delta_t\right) - \left(\overline{X}_t - \delta_t\right)$ it is seen that, for $i \neq j$,

$$\mathsf{E}\left\{\left.\left(X_{ti} - \overline{X}_t\right)^2\left(X_{tj} - \overline{X}_t\right)^2\right| n_t\right\} = \delta_t^2\left\{1 + \mathrm{O}\left(n_t^{-1}\right)\right\} + \kappa_t \mathrm{O}\left(n_t^{-2}\right).$$

Turning to the Assumptions it then follows for (iii,a) that

$$\begin{aligned}
& a_T^{-2}\sum_{t=1}^{T}\mathsf{E}\sum_{i=1}^{n_t}\mathsf{E}\left\{\left.\left(X_{ti} - \overline{X}_t\right)^2\right| n_t\right\} \\
= \; & \frac{1}{\lambda T^2}\sum_{t=1}^{T}\delta_t \mathsf{E}n_t\left\{1 + \mathrm{o}(1)\right\} = \frac{1}{T^2}\sum_{t=1}^{T}\delta_t\left\{1 + \mathrm{o}(1)\right\} \to \Sigma.
\end{aligned}$$

Next, for (iii,b) it suffices to prove that

$$\mathsf{E}\left[a_T^{-2}\sum_{t=1}^{T}\left\{\sum_{i=1}^{n_t}\left(X_{ti} - \overline{X}_t\right)^2 - \lambda\delta_t\right\}\right]^2 = I_1 + I_2 \to 0,$$

where

$$\begin{aligned}
I_1 &= \frac{1}{\lambda^2 T^4}\sum_{t=1}^{T}\mathsf{E}\left\{\sum_{i=1}^{n_t}\left(X_{ti} - \overline{X}_t\right)^2 - \lambda\delta_t\right\}^2 = \frac{1}{\lambda^2 T^4}\sum_{t=1}^{T}\kappa\lambda\left\{1 + \mathrm{o}(1)\right\} \to 0, \\
I_2 &= \frac{1}{\lambda^2 T^4}\sum_{s\neq t}\mathsf{E}\left\{\sum_{i=1}^{n_t}\left(X_{ti} - \overline{X}_t\right)^2 - \lambda\delta_t\right\}\mathsf{E}\left\{\sum_{i=1}^{n_s}\left(X_{si} - \overline{X}_s\right)^2 - \lambda\delta_s\right\} = 0.
\end{aligned}$$

For (iii, c) and (iv) it holds

$$a_T^{-4}\sum_{t=1}^{T}\mathsf{E}n_t\sum_{i=1}^{n_t}\mathsf{E}\left\{\left(X_{ti}-\overline{X}_t\right)^4\middle| n_t\right\} = \frac{1}{\lambda^2 T^4}\sum_{t=1}^{T}\kappa\lambda^2\left\{1+\mathrm{o}\left(1\right)\right\}\to 0,$$

$$a_T^{-2}\sum_{t=1}^{T}\mathsf{E}n_t^{-1}\sum_{i=1}^{n_t}\mathsf{E}\left(X_{ti}^2\middle| n_t\right) = \frac{1}{\lambda T^2}\sum_{t=1}^{T}\delta_t^2\left\{1+\mathrm{o}\left(1\right)\right\}\to 0.$$

B References

Anderson, T.W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* 30, 676-687.

Beard, R.E., Pentikainen, T., Pesonen, E. (1984) *Risk Theory*, Chapmann and Hall, London, 3rd edition

Bühlmann, H. (1970) *Mathematical methods in risk theory*, Springer-Verlag, Heidelberg.

Davidson, J. (1994) *Stochastic limit theory*, Oxford University Press.

Dickey, D.A. and Fuller, W.A. (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427-431.

England, P. and Verrall, R.J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal* 8, 443-544.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2002) The generalized dynamic-factor model: identification and estimation. *Review of Economics and Statistics* 82, 540-554.

Hall, P., and Heyde, C.C. (1980). *Martingale Limit Theory and Its Applications.* Academic Press, San Diego.

Lai, T.L. and Wei, C.Z. (1983b) Asymptotic properties of general autoregressive models and strong consistency of least-squares estimates of their parameters. *Journal of Multivariate Analysis* 13, 1-23.

Lee, R. and Carter, L. (1992) Modelling and forecasting U.S. mortality. *Journal of the American Statistical Association* 87, 659-671.

Nielsen, B. (1997) Bartlett correction of the unit root test in autoregressive models. *Biometrika* 84, 500-504.

Nielsen, B. (2001) The asymptotic distribution of unit root tests of unstable autoregressive processes. *Econometrica* 69, 211-219.

Norberg, R. (1990) Risk theory and its statistics environment. *Statistics* 21, 273-299.

Pesaran, M.H. (2003) A simple panel unit root test in the presence of cross section dependence. Cambridge University DAE Working Paper 0346.

Sundt, B. (1993) *An introduction to non-life insurance mathematics*, Veröf. des Inst. für Versich., Universität Mannheim, 28, Karlsruhe, 3rd edition