

Reese, Simon

Working Paper

Asymptotic Inference in the Lee-Carter Model for Modelling Mortality Rates

Working Paper, No. 2015:16

Provided in Cooperation with:

Department of Economics, School of Economics and Management, Lund University

Suggested Citation: Reese, Simon (2015) : Asymptotic Inference in the Lee-Carter Model for Modelling Mortality Rates, Working Paper, No. 2015:16, Lund University, School of Economics and Management, Department of Economics, Lund

This Version is available at:

<https://hdl.handle.net/10419/260154>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Working Paper 2015:16

Department of Economics
School of Economics and Management

Asymptotic Inference in the Lee-Carter Model for Modelling Mortality Rates

Simon Reese

May 2015



LUND
UNIVERSITY

Asymptotic inference in the Lee-Carter model for modelling mortality rates

Simon Reese*

May 26, 2015

Abstract

The most popular approach to modelling and forecasting mortality rates is the model of Lee and Carter (Modeling and Forecasting U. S. Mortality, *Journal of the American Statistical Association*, **87**, 659–671, 1992). The popularity of the model rests mainly on its good fit to the data, its theoretical properties being obscure. The present paper provides asymptotic results for the Lee-Carter model and illustrates its inherent weaknesses formally. Requirements on the underlying data are established and variance estimators are presented in order to allow hypothesis testing and the computation of confidence intervals.

JEL Classification: C33; C51; C53; J11.

Keywords: Lee-Carter model; mortality; common factor models; panel data.

1 Introduction

The Lee-Carter model (Lee and Carter, 1992) is nowadays one of the standard tools for forecasting mortality being used in a wide range of studies (for a non-exhaustive list, see Girosi and King, 2007). It has also sparked an entire strand in the literature that suggests extensions and methodological alternatives (see Plat, 2009, for a relatively recent review). Surprisingly, however, the basic theoretical properties of the Lee-Carter model have never been thoroughly investigated. This paper closes the gap by providing asymptotic results for the estimated parameters and by pointing out drawbacks in both the construction and the use of the model.

Consider the $X \times T$ matrix $\mathbf{M} = \{m_{x,t}\}$ of logarithmic mortality rates for age groups $x = 1, \dots, X$ and time periods $t = 1, \dots, T$. Following Lee and Carter (1992), it is defined as

$$\mathbf{M} = \boldsymbol{\alpha}' + \boldsymbol{\kappa}\boldsymbol{\beta}' + \mathbf{E}, \tag{1}$$

*Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden. Telephone: +46 46 222 79 11. Fax: +46 46 222 4613. E-mail address: simon.reese@nek.lu.se.

where the $X \times 1$ vectors $\alpha = [\alpha_1 \dots \alpha_X]'$ and $\beta = [\beta_1 \dots \beta_X]'$ are age-group specific intercepts and slope coefficients respectively and the $T \times 1$ vector $\kappa = [\kappa_1 \dots \kappa_T]'$ is an unobserved time trend. E is a matrix of deviations from the sum of common trend and individual intercepts. It is clear that this parameterization is a special case of the static factor model, which is characterized by the existence of a certain number of unobservable factors that may affect all entities in the sample differently. The time trend κ is the only latent factor that is assumed in (1). Its properties are not further specified by Lee and Carter, but the authors model the estimated trend ex-post as a Random Walk with Drift (RWD) when forecasting mortality rates. In order to avoid this inconsistency we treat the common trend directly as a RWD, formally defining it as

$$\kappa_t = \delta + \kappa_{t-1} + v_t. \quad (2)$$

This specification is not new to the literature. In fact, Girosi and King (2007) make the same assumption in order to show that the Lee-Carter model consisting of equations (1) and (2) is a special case of the RWD model. If the same equations are to be interpreted with regards to their implications for the factor model, it is insightful to decompose κ_t into its deterministic and stochastic components:

$$\kappa_t = t\delta + \eta_t \quad (3)$$

where $\eta_t = \eta_{t-1} + v_t$. Taking together equations (1) and (3) we obtain

$$m_{x,t} = \alpha_x + \beta_x t\delta + \beta_x \eta_t + e_{x,t} \quad (4)$$

which is nearly identical to the nonstationary common factor model with intercept and trend in Bai and Ng (2008, eq. 7.1). By estimating the sum of both the deterministic and the stochastic parts of the latent trend (3), Lee and Carter deviate from the customary approach in the PC literature. Furthermore, they use the identifying restriction $\sum_{t=1}^T \kappa_t = 0$ in order to estimate α_x with

$$\hat{\alpha}_x = T^{-1} \sum_{t=1}^T m_{x,t} \quad (5)$$

in a first step. However, from (4), it is obvious that the identifying restriction is necessarily violated since

$$T^{-1} \sum_{t=1}^T t\delta = \frac{T+1}{2}\delta \neq 0$$

which implies that the estimator of α is biased unless $\kappa_0 = -\frac{T+1}{2}\delta$. This result represents a weakness in the construction of the Lee-Carter model. Its consequences for the estimators of β and κ as well as the fitted values have not been established yet and need to be studied.

Estimators for the remaining parameters are obtained after subtracting the estimated individual intercepts. Let $\mathbf{M}^* = \mathbf{M} - \hat{\alpha}$ denote the individually demeaned mortality rates. An estimator of β which satisfies the additional identifying assumption $\sum_{x=1}^X \beta_x = 1$ is obtained using the singular value decomposition (SVD) $\mathbf{M}^* = \mathbf{U}\mathbf{\Sigma}\mathbf{V}'$, where \mathbf{U} (\mathbf{V}) is the $T \times T$ ($X \times X$) matrix containing the left (right) singular vectors of \mathbf{M}^* and $\mathbf{\Sigma}$ is a $T \times X$ matrix with the singular values $\sigma_i \forall i = 1, \dots, \min\{X, T\}$ in decreasing order on the diagonal and all other elements equal to zero. Lee and Carter set $\hat{\beta}_x = \mathbf{v}_1$ and $\hat{\kappa}_t = \sigma_1 \mathbf{u}_1$, where \mathbf{v}_1 (\mathbf{u}_1) is the right (left) singular vector corresponding to the largest singular value.¹ A problem with this estimation procedure is that the identifying restriction, which is a necessary normalisation for the SVD, is treated as an inherent characteristic of the data generating process. It is unclear what $\hat{\beta}$ estimates in the very likely case that the assumption is not satisfied. Furthermore, Lee and Carter treat their estimates as known parameters after having conducted the SVD, disregarding any estimation uncertainty. This entails the risk of drawing erroneous conclusions when comparing different parameters, since the difference might simply reflect estimation errors in small samples. Additionally, forecasting uncertainty is measured incorrectly since all estimated parameters are treated as known.

It is well known that the SVD used by Lee and Carter is directly related to the eigendecomposition since \mathbf{V} (\mathbf{U}) is the matrix of eigenvectors of $\mathbf{M}^* \mathbf{M}^*$ ($\mathbf{M}^* \mathbf{M}^*$). We can hence use results from the literature on factor models that rely on Principal Components (PC) for estimating a rotation of the latent factors. The asymptotic theory for PC-based estimation of static factor models is well-established (see Bai and Ng, 2008, for a survey) and can be used in order to derive the properties of the Lee-Carter model. Taking an asymptotic perspective enables us to provide results for measurement uncertainty and confidence intervals for a comparatively general class of data generating processes. If the Lee-Carter model were to be analyzed in a finite sample framework, prohibitively rigid assumptions on the stochastic properties of the model components would have to be made to yield tractable results. This is avoided in asymptotic analysis which provides its limit results as a good approximation for finite samples. The balance of the paper is as follows: Section 2 establishes the assumptions that need to be made and discusses the asymptotic distributions of the model estimates. In section 3, estimators for the asymptotic variances are presented whereas section 4 discusses the implications of estimation uncertainty for the confidence intervals of forecasts. Finally, section 5 concludes.

¹In this paper, we use the normalization $\sum_{x=1}^X \beta_x^2 = 1$ which is standard in most software packages. The order results are unaffected by this choice and the variance estimators in section are valid irrespective of the chosen normalisation.

2 Asymptotic theory for the model parameters

The model specification of Lee and Carter (1992) is fairly superficial, including solely the specification of expected value and variance of the model errors and an ex-post treatment of the latent trend as RWD. Additional assumptions have to be made in order to obtain tractable asymptotic distributions. In the following, let K denote a generic finite number. Furthermore, note that on several occasions the expression $X \rightarrow \infty$ is used, which, despite the marked increase in life expectancy during the 20th century, clearly appears awkward in the case of age groups. However, the expression is used to state a limit result and assumes only that life expectancy is not fixed at some upper level. The rate at which life expectancy (and thus age groups) increases is hence not restricted allows for an intuitively appealing very slow expansion rate relative to time periods.

Assumption 1 $v_t \sim i.i.d.(0, \sigma_v)$, $E[v^4] \leq K$, $T^{-3}\kappa'\kappa \in (0, K]$, and $\kappa_0 \leq K$.

Assumption 2

- (i) α_x is either deterministic s.t. $|\alpha_x| \leq K$, or stochastic s.t. $E[\alpha_x^4] \leq K$. In either case, $X^{-1}\alpha'\alpha \xrightarrow{p} \sigma_\alpha^2 > 0$ as $X \rightarrow \infty$.
- (ii) β_x is either deterministic s.t. $|\beta_x| \leq K$, or stochastic s.t. $E[\beta_x^4] \leq K$. In either case, $X^{-1}\beta'\beta \xrightarrow{p} \sigma_\beta^2 > 0$ as $X \rightarrow \infty$.

Assumption 3

- (i) $E[e_{x,t}] = 0$ and $E[e_{x,t}^8] \leq K$.
- (ii) $E[e_{x,t}e_{y,s}] = |\sigma_{xy,ts}| \leq \bar{\sigma}_{xy} \forall (t, s)$ and $|\sigma_{xy,ts}| < \tau_{ts} \forall (x, y)$ s.t. $\frac{1}{X} \sum_{x=1}^X \sum_{y=1}^X \bar{\sigma}_{xy} \leq K$, $\frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \tau_{ts} \leq K$ and $\frac{1}{XT} \sum_{x=1}^X \sum_{y=1}^X \sum_{t=1}^T \sum_{s=1}^T |\sigma_{xy,ts}| \leq K$.
- (iii) For every (t, s) , $E\left[|X^{-1/2} \sum_{x=1}^X (e_{x,s}e_{x,t} - E[e_{x,s}e_{x,t}])|^4\right] \leq K$.
- (iv) For each t , $X^{-1/2} \sum_{x=1}^X \beta_x e_{x,t} \xrightarrow{d} N(0, \Gamma_t)$, as $X \rightarrow \infty$ where $\Gamma_t = \lim_{X \rightarrow \infty} X^{-1} \sum_{x=1}^X \sum_{y=1}^X E[\beta_x \beta_y e_{x,t} e_{y,t}]$.
- (v) For each $s, t : s \neq t$, $\lim_{X \rightarrow \infty} X^{-1} \sum_{x=1}^X \sum_{y=1}^X E[|\beta_x \beta_y e_{x,t} e_{y,s}|] < \max\{\Gamma_t, \Gamma_s\}$
- (vi) For each x , $T^{-3/2} \sum_{t=1}^T \kappa_t e_{x,t} \xrightarrow{d} N(0, \Phi_x)$ as $T \rightarrow \infty$ where $\Phi_x = \lim_{T \rightarrow \infty} T^{-3} \sum_{t=1}^T \sum_{s=1}^T E[\kappa_t \kappa_s e_{x,t} e_{x,s}]$.

Assumption 4 $\{\beta_x\}$, $\{\kappa_t\}$ and $\{e_{x,t}\}$ are three mutually independent groups

Assumption 5 For all $t \leq T$, $x \leq X$, $\sum_{s=1}^T \tau_{ts} \leq K$ and $\sum_{x=1}^X |\bar{\sigma}_{xy}| \leq K$.

Assumption 6

(i) for each t ,

$$E \left[X^{-1/2} T^{-3/2} \sum_{s=1}^T (\mathbf{e}_s^* \mathbf{e}_t^* - E[\mathbf{e}_s^* \mathbf{e}_t^*]) \kappa_s^* \right]^2 \leq K$$

$$(ii) E \left[X^{-1/2} T^{-3/2} \sum_{t=1}^T \mathbf{e}_t^{*'} \boldsymbol{\beta} \kappa_t^* \right]^2 \leq K$$

These assumptions are standard in the PC literature and largely follow the ones in Bai and Ng (2008). It is important to note that Assumption 1 differs from the specification in the PC literature by requiring a normalisation of the squared sum by T^{-3} . This is necessary due to the dominating properties of the linear trend in the RWD. Assumption 2 is possibly more general than necessary by allowing for random age-group-specific coefficients on the latent trend. In fact, it appears plausible that $\boldsymbol{\beta}$ reflects the inherent characteristics of a specific age groups which would the impact of the latent trend a fixed parameter. Assumption 3 is especially interesting since it allows for considerably more generality in the model errors than assumed previously in the literature. In contrast to common claims (Brouns et al., 2005, Lee and Miller, 2001) heteroskedasticity can be allowed for and even weak serial and cross-sectional correlation in the model errors is permissible. For an in-depth discussion of all assumptions, the interested reader is referred to Bai and Ng (2008).

As alluded to previously, $\hat{\alpha}_x$ is biased since it estimates the sum of α_x and the mean of the latent trend. Since the average of κ is a linear function of the number of time periods, it is obvious that $\hat{\alpha}_x$ diverges at rate T . The following theorem states its asymptotic behaviour more precisely.

Theorem 1 *Under assumptions 1-5,*

$$\alpha_x - \hat{\alpha}_x \xrightarrow{d} \lim_{T \rightarrow \infty} \frac{T+1}{2} \beta_x \delta + \sqrt{T} \beta_x N(0, \sigma_v^2/3) + T^{-1/2} N(0, \bar{\sigma}_x^2)$$

where $\bar{\sigma}_x^2 = \text{Var}[T^{-1/2} \sum_{t=1}^T e_{x,t}]$

The second term on the right hand side implies that, in addition to a deterministic bias, the distribution of $\hat{\alpha}_x$ is affected by a stochastic term whose variance increases at rate T . Thus, generally $\hat{\alpha}_x$ does not deliver any insight about the numerical value of the age-specific intercept. However, comparisons between different intercept coefficients in the same sample are possible since the bias arising from inclusion of the mean latent trend is identical. A more problematic comparison is that of α across different samples (see e.g. Lee, 2000, or Janssen, 2013). Let $\hat{\alpha}_{x,j} = \alpha_{x,j} + \kappa_0 + \frac{T+1}{2} \beta_{x,j} \delta_j + T^{-1} \sum_{t=1}^T \eta_{t,j} + o_p(1)$ be the estimated baseline log mortality rate of age group x in sample $j = 1, 2$. The difference $\alpha_{x,1} - \alpha_{x,2}$ may not only be due

to the non-diminishing terms κ_0 and $^{-1} \sum_{t=1}^T \eta_{t,j}$, it is even affected by differences in the drift terms to a magnitude that increases with the number of time periods in the sample. This implies a possibility to systematically manipulate the comparison of intercepts in two samples by choosing the sample range adequately.

It is easy to see that the bias in the estimation of α_x implies that the estimator of κ , which is obtained from the difference of $m_{x,t}$ and α_x , is an estimator of the *demeaned* RWD, eliminating all information about the true trend levels. In this light, comparisons of estimated trends across samples, as e.g. done graphically in Tuljapurkar et al. (2000), appear pointless since they do not provide any insight about the relative position of different trends. Furthermore, κ_t and β_x can only be estimated up to a factor since estimation methods solely identify the product $\kappa_t \beta_x$ which is identical to $(\kappa_t h)(h^{-1} \beta_x)$ for any $h \neq 0$. In the Lee-Carter model, where only one latent factor is given, the SVD implies a fairly clear normalisation.

Proposition 1 *Under assumptions 1-6,*

$$h = X^{1/2} \sigma_\beta + O_p(X^{1/2} T^{-1/2}) + O_p(T^{-1}) \quad (6)$$

Obviously, this expression is nonbounded as the number of age groups tends to infinity, making it inconvenient for interpretation. Still, under the assumption that $X/T \rightarrow 0$ the multiplier converges to a multiple of the root mean square deviation of the slope coefficients from zero.

Given the definition of the normalisation h , it is possible to state the asymptotic distribution of the estimated factors.

Theorem 2 *Under Assumptions 1-6 and assuming that $X/T^4 \rightarrow 0$,*

$$(\hat{\kappa}_t - \kappa_t^* h) \xrightarrow{d} N(0, \Gamma_t \sigma_\beta^{-2})$$

for each t as $N, T \rightarrow \infty$ where $\kappa_t^ = \kappa_t - T^{-1} \sum_{t=1}^T \kappa_t$.*

Interestingly, $\hat{\kappa}_t$ does not converge, as it is subject to a certain persisting degree of estimation uncertainty irrespectively of the amount of information used. This observation is a direct consequence of the chosen normalisation in the SVD. The scaling entails a fast rate of convergence for $\hat{\beta}_x$ while keeping a certain margin of error around $\hat{\kappa}$. It is possible to use a different scaling which allows for the estimators of both β and κ to converge by adapting the estimation steps in the PC literature (see e.g. Bai and Ng, 2008). An important implication of Theorem 2 concerns studies that test for breaks in the latent trend of the Lee-Carter model (see e.g. Li, 2011). In order for the underlying statistical analysis to be valid, estimation uncertainty in κ must be taken into account.

Remark The admissible relative rate of convergence for the results to hold allows for all patterns that have been observed in the relative development of time and life expectancy. Both

the very slow increase in longevity until the industrial revolution and the steep rise of life expectancy thereafter, which implies an increase in the number of age groups involved the analysis of mortality rates, satisfy the requirement stated in theorem 1. It is worth noting that the presence of a drift term drives the requirements on X/T . Similar factor models that assume the latent trends to be a pure random walk (Bai, 2004) or to be stationary (Bai, 2003) require stricter assumptions.

In addition to the results on the latent factor, the asymptotic properties of its individual coefficients can be reported.

Theorem 3 *Under Assumptions 1-6,*

$$\sqrt{XT^3}(\hat{\beta}_x - \beta_x h^{-1}) \xrightarrow{d} \frac{12}{\delta^2 \sigma_\beta} N(0, \Phi_x)$$

for each x as $X, T \rightarrow \infty$.

In contrast to $\hat{\kappa}_t$, $\hat{\beta}_x$ converges fairly quickly to a multiple of the true parameters vector. But as already noticed by Lee (2000), the levels of this multiple have no meaning, a fact which is due to the inconvenient normalisation. Nevertheless, the established limiting results allow testing hypotheses on the relative magnitudes of slope coefficients at different ages.

The results presented for the latent trend and its slope coefficients naturally lead to a statement for the fitted values $\hat{m}_{x,t} = \hat{\alpha}_x + \hat{\kappa}_t \hat{\beta}_x$. Their asymptotic properties are given by Theorem 4 below.

Theorem 4 *Under Assumptions 1-6 and assuming that $X/T^4 \rightarrow 0$*

$$(X^{-1}v_1 + T^{-1}v_2)^{-1/2}(\hat{m}_{x,t} - E[m_{x,t}]) \xrightarrow{d} N(0, 1)$$

for each (x, t) as $X, T \rightarrow \infty$ where

$$v_1 = \left(\sigma_\beta^{-1} \beta_x\right)^2 \Gamma_t \sigma_\beta^{-2}$$

and

$$v_2 = \bar{\sigma}_x^2 + \left(\frac{12}{\delta} \left(\frac{t}{T} - \frac{1}{2}\right) \cdot\right)^2 \Phi_x$$

Interestingly, the fitted values of the model prove to be consistent despite the bad properties of $\hat{\alpha}_x$ and the persistent degree of estimation uncertainty around the estimated trend. The increasing margin of error in the estimated intercept is absorbed when summing it with the

demeaned fit $\hat{\kappa}_t \hat{\beta}_x$ whereas the multiplication of $\hat{\kappa}_t$ with its slope coefficients causes its estimation error to disappear asymptotically. The overall rate of convergence depends on the relative expansion rate and is $\min\{X^{-1/2}, T^{-1/2}\}$.

3 Variance estimators

The asymptotic variances in Theorems 1–4 are not known for the samples that one might work with and hence need to be estimated using a feasible procedure. For the latent trend the asymptotic variance

$$aVar[\hat{\kappa}_t] = \sigma_\beta^{-2} plim \sum_{x=1}^X X^{-1/2} \sigma_{x,t}^2 \beta_x^2$$

can be estimated by one of the three proposed estimators in Bai and Ng (2006). The most general of these, allowing for cross-section correlation and heteroskedasticity, is given by

$$\widehat{aVar}[\hat{\kappa}] = X^{-1} \sum_{x=1}^X \sum_{y=1}^X \hat{\beta}_x \hat{\beta}_y \sum_{t=1}^T \hat{e}_{x,t}^* \hat{e}_{y,t}^* \quad (7)$$

where $\hat{e}_{x,t}^* = m_{x,t}^* - \hat{\kappa}_t \hat{\beta}_x$. However, this generality requires imposing a stationarity assumption on the error terms which implies that the variance of the estimated trend is identical across time periods. This restriction can be lifted if one is willing to assume that the error term is cross-sectionally uncorrelated. In this case, it is possible to use the simpler term

$$\widehat{aVar}[\hat{\kappa}_t] = \sum_{y=1}^X (\hat{e}_{y,t}^*)^2 \hat{\beta}_y^2, \quad (8)$$

whereas additionally assuming homoskedasticity in the errors over both time and age groups allows using

$$\widehat{aVar}[\hat{\kappa}_t] = \hat{\sigma}_{e^*}^2 \sum_{y=1}^X \hat{\beta}_y^2 \quad (9)$$

where $\hat{\sigma}_{e^*}^2 = (XT - (X + T - 1))^{-1} \sum_{x=1}^X \sum_{t=1}^T (\hat{e}_{x,t}^*)^2$

Estimating the variance of $\hat{\beta}_x$ can be conducted at a similar level of generality as it was the case for the trend estimator. In order to allow for serial correlation, Bai (2003) suggests using a Newey-West estimator whose equivalent in the Lee-Carter model is given by

$$\widehat{aVar}[\sqrt{XT^3} \hat{\beta}_x] = XT^3 \left(\zeta_{0,x} + 2 \sum_{p=1}^r \left(1 - \frac{p}{r+1} \right) \zeta_{p,x} \right) \lambda_1^{-2}, \quad (10)$$

where $\zeta_{p,x} = T^{-1} \sum_{t=p+1}^T \hat{e}_{x,t}^* \hat{\kappa}_t \hat{e}_{x,t-p}^* \hat{\kappa}_{t-p}$ and $r = \lfloor 0.75T^{1/3} \rfloor$.

Concerning the fitted values, their estimated variance can be constructed from the variance estimators of the estimated latent trend and the slope coefficients since Theorem 4 implies that

$$\widehat{aVar}[\hat{m}_{x,t}] = \hat{\beta}_x^2 \widehat{Var}[\hat{\kappa}_t] + \hat{\kappa}_t^2 \widehat{Var}[\hat{\beta}_x] + T^{-1} \hat{\sigma}_x^2. \quad (11)$$

This expression requires a consistent estimator of $\bar{\sigma}_x^2$, which is given by

$$\hat{\sigma}_x^2 = T^{-1} \sum_{t=1}^T \hat{e}_{x,t}^2 + 2 \sum_{p=1}^r \left(1 - \frac{p}{r+1}\right) T^{-1} \sum_{t=p+1}^T \hat{e}_{x,t} \hat{e}_{x,t-p}. \quad (12)$$

4 Forecasting mortality rates

Predicting future values of the latent trend boils down to extrapolating the drift term from the last known observation. The drift term itself is unknown and usually² estimated as

$$\hat{\delta} = (T-1)^{-1} \sum_{t=2}^T \Delta \hat{\kappa}_t \quad (13)$$

where $\Delta \hat{\kappa}_t = \hat{\kappa}_t - \hat{\kappa}_{t-1}$. Its asymptotic properties can hence straightforwardly be derived from those of $\hat{\kappa}_t$.

Lemma 4.1 *Under assumptions 1-6 then as $X, T \rightarrow \infty$,*

$$\sqrt{\frac{T}{X}}(\hat{\delta} - \delta h) = A + O_p(X^{-1/2}) + O_p(T^{-3/2}).$$

where $A \xrightarrow{d} N(0, \sigma_\beta^2 \sigma_v^2)$.

Most interestingly, the variance of $\hat{\delta}$ increases at rate X/T which implies that tighter restrictions on the rate of convergence have to be imposed in order to ensure convergence. The fact that $\hat{\delta}$ estimates a multiple of δ is a consequence of the factor estimation which itself yields a scaled estimate of the latent trend. This also applies to the estimated random innovations

$$\begin{aligned} \hat{v} &= \Delta \hat{\kappa}_t - \hat{\delta} \\ &= v_t h + X^{-1/2} \Delta \mathbf{e}_t' \boldsymbol{\beta} \frac{\delta}{\sqrt{12}} + o_p(1) \end{aligned} \quad (14)$$

which, in addition to being scaled by the same factor as $\hat{\kappa}_t$ and $\hat{\delta}$, include a perturbation term that arises from the uncertainty involved in the estimation of $\hat{\kappa}$. The latter term is problematic

²see e.g. Girosi and King (2007, Ch.2)

because $\{\hat{v}_t, t = 1, \dots, T\}$ is used to estimate the variance of $v_{T+q}h$. The estimator

$$\hat{\sigma}_v^2 = (T-1)^{-1} \sum_{t=2}^T \hat{v}_t^2$$

includes the variance of the second term on (14) and hence introduces a positive bias in the estimated variance of future random innovations.

If one is interested in obtaining the distribution of forecasts of the latent trend, the above variance estimators are essential components for the estimation of the forecasting variance.

Theorem 5 Let $\hat{\kappa}_{T+q|T} = \hat{\kappa}_t + q\hat{\delta}$. Under assumptions 1-6 and assuming $X/T = O(1)$ and $v_t \sim i.i.d.n.(0, \sigma_v^2)$,

$$\frac{\hat{\kappa}_{T+q|T} - \kappa_{T+q}h}{\text{Var}[\hat{\kappa}_{T+q|T}]} \xrightarrow{d} N(0, 1),$$

where $\text{Var}[\hat{\kappa}_{T+q}] = a\text{Var}[\hat{\kappa}_T] + q^2 (\xi a\text{Var}[\sqrt{T/X}\hat{\delta}] + \sigma_v^2)$ and $\xi = \lim_{X, T \rightarrow \infty} X/T$

The additional assumption on the relative expansion rate of X and T arises from the properties of $\hat{\delta}$ and ensures that $\text{Var}[\hat{\kappa}_{T+q}]$ is finite. In applications, it is most often plausible to assume that $X/T \rightarrow 0$ so that the asymptotic variance can be simplified to $\text{Var}[\hat{\kappa}_{T+q}] = a\text{Var}[\hat{\kappa}_T] + q^2\sigma_v^2$. Still, even this simplified variance contains uncertainty due to the estimation of κ , a factor that is usually disregarded in practice. An estimator of the variance for the q -step forecast is given by taking the estimators of its components.

$$\widehat{\text{Var}}[\hat{\kappa}_{T+q}] = \widehat{\text{Var}}[\hat{\kappa}_T] + q^2 \left(X/T \widehat{\text{Var}}[\sqrt{T/X}\hat{\delta}] + \hat{\sigma}_v^2 \right) \quad (15)$$

where the term $X/T \widehat{\text{Var}}[\sqrt{T/X}\hat{\delta}]$ can be left out if $X/T \rightarrow 0$.

Interest concerning the distribution of forecasts in the Lee-Carter model is not restricted to the latent trend. Results concerning forecasts of age group-specific logged mortality rates are equally important.

Theorem 6 Let $\hat{m}_{x,T+q|T} = \hat{m}_{x,T} + q\hat{\delta}\hat{\beta}_x$. Under the assumptions of Theorem 5 and assuming $e_{x,t} \sim N(0, \sigma_x^2) \forall t$,

$$\frac{\hat{m}_{x,T+q|T} - \alpha_x - \beta_x \kappa_{T+q}}{\text{Var}[\hat{m}_{x,T+q|T}]} \xrightarrow{d} N(0, 1)$$

where $\text{Var}[\hat{m}_{x,T+q|T}] = a\text{Var}[\hat{m}_{x,T}] + q^2\beta_x^2(\xi a\text{Var}[\sqrt{T/X}\hat{\delta}] + \sigma_v^2) + \sigma_x^2$

Again, an upper bound for the admissible relative expansion rate has to be imposed in order to obtain a finite limiting variance for the forecasted logarithmic mortality rates. Additionally, the

asymptotic variance includes $\text{Var}[\hat{m}_{x,t}]$ and hence yet another source of estimation-related uncertainty that is usually neglected. Assuming the idiosyncratic components to be homoskedastic is actually not required to obtain the results of Theorem 6. However, it is necessary to get a feasible variance estimate. This estimate is given by

$$\widehat{\text{Var}}[\hat{m}_{x,T+q|T}] = \widehat{\text{aVar}}[\hat{m}_{x,T}] + q^2 \hat{\beta}_x^2 (X/Ta\widehat{\text{Var}}[\sqrt{T/X}\hat{\delta}] + \hat{\sigma}_v^2) + \hat{\sigma}_x^2. \quad (16)$$

5 Conclusion

The present paper considers the asymptotic properties of the Lee-Carter model for modelling and forecasting mortality rates. Taking an asymptotic perspective allows us to avoid a number of restrictive assumptions and to obtain results for a quite general class of DGPs which are good approximations in finite samples. We derive asymptotic distributions and provide variance estimators for the parameter estimators of the model. It is obvious that the superficial design of the Lee-Carter model leads to inconsistencies that impair the usefulness of the estimated parameters. Additionally, the trend estimate obtained from the SVD is not consistent, implying that, in contrast to common practice, estimation uncertainty should not be disregarded. Lastly, forecasted values from the model turn out to be affected by an inconvenient scaling choice in the SVD. This may result in additional forecasting uncertainty, depending on the relative rate of expansion.

References

- Bai, J., and S. Ng (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica* **70**, 191–221.
- Bai, J. (2003). Inferential Theory for Factor Models of Large Dimension, *Econometrica*, **71**, 135–171.
- Bai, J. (2004). Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data, *Journal of Econometrics* **122**, 137–183.
- Bai, J., and S. Ng (2008). Large Dimensional Factor Analysis. *Foundations and Trends in Econometrics* **3**, 89–163.
- Brouhns N., M. Denuit and I. van Keilegom (2005). Bootstrapping the Poisson log-bilinear model for mortality forecasting. *Scandinavian Actuarial Journal* **3**, 212–224.
- Davidson, J. (2000). *Econometric Theory*, Blackwell Publishing.
- Giroi, F. and G. King (2007). *Demographic Forecasting*, Princeton University Press.
- Hayashi, F. (2000). *Econometrics*, Princeton University Press.
- Janssen, F., L.J.G. van Wissen and A.E. Kunst (2013). Including the Smoking Epidemic in Internationally Coherent Mortality Projections. *Demography* **50**, 1341–1362.
- Lee, R.D. (2000). The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications. *North American Actuarial Journal* **4**, 80–91.
- Lee, R.D. and L. Carter (1992). Modeling and Forecasting U. S. Mortality. *Journal of the American Statistical Association* **87**, 659–671.
- Lee, R.D. and T. Miller (2001). Evaluation the performance of the Lee-Carter Method for Forecasting Mortality. *Demography* **38**, 537–549.
- Li, J.S., W. Chan and S. Cheung (2011). Structural Changes in the Lee-Carter Mortality Indexes. *North American Actuarial Journal* **15**, 13–31.
- Plat, R. (2009). On stochastic mortality modeling *Insurance: Mathematics and Economics* **45**, 393–404.
- Tuljapurkar, S., N. Li and C. Boe (2000), A universal pattern of mortality decline in the G7 countries. *Nature* **405**, 789–792.

A Notation and important equalities

We will conduct proofs of all Theorems based on the rescaled estimator

$$\tilde{\kappa} = \mathbf{u}_1.$$

Since \mathbf{u}_1 is an eigenvector of $\mathbf{M}^* \mathbf{M}^{*'}$, the equality

$$\mathbf{M}^* \mathbf{M}^{*'} \tilde{\kappa} = \tilde{\kappa} \lambda_1 \quad (17)$$

can be used to obtain

$$\tilde{\kappa} = \mathbf{M}^* \mathbf{M}^{*'} \tilde{\kappa} \lambda_1^{-1}$$

The above expression suggests that the rotation scalar \tilde{h} can be defined as $\tilde{h} = \boldsymbol{\beta}' \boldsymbol{\beta} \boldsymbol{\kappa}^{*'} \tilde{\kappa} \lambda_1^{-1}$. Using this definition, we obtain

$$\tilde{\kappa} - \boldsymbol{\kappa}^* \tilde{h} = (\boldsymbol{\kappa}^* \boldsymbol{\beta}' \mathbf{E}^{*'} \tilde{\kappa} + \mathbf{E}^* \boldsymbol{\beta} \boldsymbol{\kappa}^{*'} \tilde{\kappa} + \mathbf{E}^* \mathbf{E}^{*'} \tilde{\kappa}) \lambda_1^{-1},$$

and for each individual time period,

$$\begin{aligned} \tilde{\kappa}_t - \kappa_t^* \tilde{h} &= X^{-1} T^{-3} (\kappa_t^* \boldsymbol{\beta}' \mathbf{E}^{*'} \tilde{\kappa} + \mathbf{e}_t^{*'} \boldsymbol{\beta}' \boldsymbol{\kappa}^{*'} \tilde{\kappa} + \mathbf{e}_t^{*'} \mathbf{E}^{*'} \tilde{\kappa}) X T^3 \lambda_1^{-1} \\ &= (j_{1,t} + j_{2,t} + j_{3,t} + j_{4,t}) X T^3 \lambda_1^{-1}, \end{aligned} \quad (18)$$

where

$$j_{1,t} = X^{-1} T^{-3} \sum_{s=1}^T \kappa_t^* \boldsymbol{\beta}' \mathbf{e}_s^* \tilde{\kappa}_s \quad (19)$$

$$j_{2,t} = X^{-1} T^{-3} \sum_{s=1}^T \mathbf{e}_t^{*'} \boldsymbol{\beta} \boldsymbol{\kappa}_s^* \tilde{\kappa}_s \quad (20)$$

$$j_{3,t} = T^{-3} \sum_{s=1}^T E[\mathbf{e}_t^{*'} \mathbf{e}_s^*] \tilde{\kappa}_s \quad (21)$$

and

$$j_{4,t} = T^{-3} \sum_{s=1}^T (\mathbf{e}_t^{*'} \mathbf{e}_s^* / X - E[\mathbf{e}_t^{*'} \mathbf{e}_s^*]) \tilde{\kappa}_s \quad (22)$$

B Consistency of factor estimates

Lemma B.1 *Under assumption 1,*

$$T^{-3} \sum_{t=1}^T (\kappa_t^*)^2 = \frac{\delta^2}{12} + O_p(T^{-1})$$

Proof of Lemma B.1 Using 3, we obtain

$$\begin{aligned} T^{-3} \sum_{t=1}^T (\kappa_t^*)^2 &= T^{-3} \sum_{t=1}^T \left(t - \frac{T-1}{2} \right)^2 \delta^2 + T^{-3} \sum_{t=1}^T \eta_t^2 + T^{-3} \sum_{t=1}^T \left(t - \frac{T-1}{2} \right) \delta \eta_t \\ &= I + II + III \end{aligned}$$

where

$$I = T^{-3} \left(\frac{1}{12} T^3 + \frac{11}{12} T \right) \delta^2 = \frac{\delta^2}{12} + O(T^{-2})$$

by simple computation,

$$II \xrightarrow{d} T^{-2} \sigma_v^2 \int_0^1 \left(W(r) - \int_0^1 W(s) ds \right)^2 dr = O_p(T^{-2})$$

by Hayashi (2000, Proposition 9.2(b)), and

$$\begin{aligned} III &\leq \left(T^{-3} \sum_{t=1}^T \left(t - \frac{T-1}{2} \right)^2 \delta^2 \right)^{1/2} \left(T^{-3} \sum_{t=1}^T \eta_t^2 \right)^{1/2} \\ &= O_p(T^{-1}), \end{aligned}$$

which follows from the first two results. ■

Lemma B.2 *Under assumptions 1-5,*

$$X^{-1} T^{-3} \lambda_1 = \sigma_{\beta}^2 \delta^2 / 12 + O_p(T^{-1})$$

Proof of Lemma B.2 From (17) we obtain

$$X^{-1} T^{-3} \tilde{\mathbf{\kappa}} \mathbf{M}^* \mathbf{M}^{*'} \tilde{\mathbf{\kappa}} = X^{-1} T^{-3} \lambda_1$$

whose deviation from $\tilde{\kappa}'\kappa^*\beta'\beta\kappa^{*'}\tilde{\kappa}$ can be decomposed into

$$\begin{aligned} & X^{-1}T^{-3}(\tilde{\kappa}'\mathbf{M}^*\mathbf{M}^{*'}\tilde{\kappa} - \tilde{\kappa}'\kappa^*\beta'\beta\kappa^{*'}\tilde{\kappa}) \\ &= X^{-1}T^{-3}\tilde{\kappa}'\kappa^*\beta'\mathbf{E}'\tilde{\kappa} + X^{-1}T^{-3}\tilde{\kappa}'\mathbf{E}\beta\kappa^{*'}\tilde{\kappa} + X^{-1}T^{-3}\tilde{\kappa}'\mathbf{E}\mathbf{E}'\tilde{\kappa} \end{aligned} \quad (23)$$

Consider the last term on the RHS above.

$$\begin{aligned} X^{-1}T^{-3}\tilde{\kappa}'\mathbf{E}\mathbf{E}'\tilde{\kappa} &\leq X^{-1}T^{-3}\left(\sum_{t=1}^T\sum_{s=1}^T(\mathbf{e}_t\mathbf{e}_s)^2\right)^{1/2} \\ &= O_p(T^{-1}) \end{aligned}$$

The remaining two expressions in (23) are identical and of order

$$\begin{aligned} X^{-1}T^{-3}\tilde{\kappa}'\mathbf{E}\beta\kappa^{*'}\tilde{\kappa} &\leq X^{-1}T^{-3}\left(\left(\sum_{t=1}^T\tilde{\kappa}_t^2\right)\left(\sum_{t=1}^T(\mathbf{e}_t'\beta\kappa^{*'})^2\right)\right) \\ &\leq X^{-1}T^{-3}\left(\left(\sum_{t=1}^T\tilde{\kappa}_t^2\right)\left(\sum_{t=1}^T\left(\sum_{s=1}^T(\mathbf{e}_t'\beta\kappa_s^{*'})^2\right)\left(\sum_{s=1}^T\tilde{\kappa}_s^2\right)\right)\right)^{1/2} \\ &= X^{-1/2}T^{-1}\left(T^{-1}\sum_{t=1}^T(X^{-1/2}\mathbf{e}_t'\beta)^2T^{-3}\sum_{s=1}^T(\kappa_s^{*'})^2\right)^{1/2} \\ &= O_p(X^{-1/2}T^{-1}). \end{aligned} \quad (24)$$

These result show that asymptotically λ_1 is the largest eigenvalue of $\kappa^*\beta'\beta\kappa^{*'}$. Note however that the largest eigenvalues of $\kappa^*\beta'\beta\kappa^{*'}$ and $\kappa^{*'}\kappa^*\beta'\beta$ are identical. Using this equality as well as Assumption 2 and Lemma B.1 we obtain the required result. ■

Lemma B.3 *Under the assumptions of Lemma B.2,*

$$\frac{1}{T}\sum_{t=1}^T(\tilde{\kappa} - \kappa^*\tilde{h})^2 = O_p(X^{-1}T^{-3}) + O_p(T^{-6})$$

Proof of Lemma B.3 Considering (18), we can write

$$\begin{aligned} T^{-1}\sum_{t=1}^T(\tilde{\kappa}_t - \kappa_t^*\tilde{h})^2 &= T^{-1}\sum_{t=1}^T(j_{1,t} + j_{2,t} + j_{3,t} + j_{4,t})^2 X^2 T^6 \lambda_1^{-2} \\ &\leq 4X^2 T^6 \lambda_1^{-2} T^{-1}\sum_{t=1}^T(j_{1,t}^2 + j_{2,t}^2 + j_{3,t}^2 + j_{4,t}^2), \end{aligned} \quad (25)$$

whose individual components are

$$\begin{aligned}
T^{-1} \sum_{t=1}^T j_{1,t}^2 &= T^{-1} \sum_{t=1}^T \left(X^{-1} T^{-3} \kappa_t^* \sum_{s=1}^T \boldsymbol{\beta}' \mathbf{e}_s^* \tilde{\kappa}_s \right)^2 \\
&\leq X^{-1} T^{-3} \left(T^{-3} \sum_{t=1}^T (\kappa_t^*)^2 \right) \left(T^{-1} \sum_{s=1}^T (X^{-1/2} \boldsymbol{\beta}' \mathbf{e}_s^*)^2 \right) \left(\sum_{s=1}^T \tilde{\kappa}_s^2 \right) \\
&= O_p(X^{-1} T^{-3}),
\end{aligned} \tag{26}$$

$$\begin{aligned}
T^{-1} \sum_{t=1}^T j_{2,t}^2 &= T^{-1} \sum_{t=1}^T (\mathbf{e}_t^{*'} \boldsymbol{\beta})^2 \left(X^{-1} T^{-3} \sum_{s=1}^T \kappa_s^* \tilde{\kappa}_s \right)^2 \\
&\leq X^{-1} T^{-3} \left(T^{-1} \sum_{t=1}^T (X^{-1/2} \mathbf{e}_t^{*'} \boldsymbol{\beta})^2 \right) \left(T^{-3} \sum_{s=1}^T (\kappa_s^*)^2 \right) \left(\sum_{s=1}^T \tilde{\kappa}_s^2 \right) \\
&= O_p(X^{-1} T^{-3}).
\end{aligned} \tag{27}$$

and

$$\begin{aligned}
T^{-1} \sum_{t=1}^T j_{3,t}^2 &= T^{-1} \sum_{t=1}^T \left(T^{-3} \sum_{s=1}^T E[\mathbf{e}_t^{*'} \mathbf{e}_s^*] \tilde{\kappa}_s \right)^2 \\
&\leq T^{-6} \left(T^{-1} \sum_{t=1}^T \sum_{s=1}^T \tau_{t,s}^2 \right) \left(\sum_{s=1}^T \tilde{\kappa}_s^2 \right) \\
&= O_p(T^{-6}),
\end{aligned} \tag{28}$$

where the order of $\left(T^{-1} \sum_{t=1}^T \sum_{s=1}^T \tau_{t,s}^2 \right)$ is obtained via Lemma 1(i) in Bai and Ng (2002). Concerning $j_{4,t}$, let $\tilde{\gamma}(s, t) = \mathbf{e}_t^{*'} \mathbf{e}_s^* / X - E[\mathbf{e}_t^{*'} \mathbf{e}_s^*]$. Following the proof of Theorem 1 in Bai and Ng (2002) we obtain

$$\begin{aligned}
T^{-1} \sum_{t=1}^T j_{4,t}^2 &= T^{-7} \sum_{t=1}^T \left(\sum_{s=1}^T \tilde{\gamma}(s, t) \tilde{\kappa}_s \right)^2 \\
&= T^{-7} \sum_{t=1}^T \sum_{s=1}^T \sum_{u=1}^T \tilde{\gamma}(s, t) \tilde{\gamma}(u, t) \tilde{\kappa}_s \tilde{\kappa}_u \\
&\leq T^{-7} \left(\sum_{s=1}^T \sum_{u=1}^T \left[\sum_{t=1}^T \tilde{\gamma}(s, t) \tilde{\gamma}(u, t) \right]^2 \right)^{1/2} \left(\sum_{s=1}^T \sum_{u=1}^T [\tilde{\kappa}_s \tilde{\kappa}_u]^2 \right)^{1/2} \\
&\leq T^{-7} \left(\sum_{s=1}^T \sum_{u=1}^T \left[\sum_{t=1}^T \tilde{\gamma}(s, t) \tilde{\gamma}(u, t) \right]^2 \right)^{1/2} \sum_{s=1}^T \tilde{\kappa}_s^2.
\end{aligned} \tag{29}$$

Via the inequality

$$\begin{aligned} E \left[\left(\sum_{t=1}^T \tilde{\gamma}(s, t) \tilde{\gamma}(u, t) \right)^2 \right] &= E \left[\sum_{t=1}^T \tilde{\gamma}(s, t) \tilde{\gamma}(u, t) \tilde{\gamma}(s, v) \tilde{\gamma}(u, v) \right] \\ &\leq T^2 \max_{s, t} E \left[|\tilde{\gamma}(s, t)|^4 \right] \end{aligned} \quad (30)$$

and the fact that $E [|\tilde{\gamma}(s, t)|^4] \leq X^{-2}K$ by assumption 3(iii), equation (29) can be rewritten

$$\begin{aligned} T^{-1} \sum_{t=1}^T j_{4,t}^2 &\leq T^{-7} \left(\sum_{s=1}^T \sum_{u=1}^T O_p(X^{-2}T^2) \right)^{1/2} \\ &= O_p(X^{-1}T^{-5}). \end{aligned} \quad (31)$$

Collecting the results in (26)–(31) and noting that $X^{-1}T^{-3}\lambda_1 = O_p(1)$, as shown in Lemma B.2, we obtain the required result \blacksquare

Lemma B.4 *Under the assumptions of Lemma B.2,*

$$\tilde{h} = T^{-3/2} \frac{\sqrt{12}}{\delta} + O_p(X^{-1/2}T^{-5/2}) + O_p(T^{-4}) \quad (32)$$

Proof of Lemma B.4 We can use the decomposition

$$(\kappa^*)' \tilde{\kappa} = (\kappa^*)' (\tilde{\kappa} - \tilde{h} \kappa^*) + (\kappa^*)' \tilde{h} \kappa^*$$

so that

$$\tilde{h}(\kappa^*)' \tilde{\kappa} = (\tilde{h} \kappa^*)' (\tilde{\kappa} - \tilde{h} \kappa^*) + \tilde{h}^2 (\kappa^*)' \kappa^* \quad (33)$$

Now note that the LHS above can also be decomposed

$$\begin{aligned} \tilde{h}(\kappa^*)' \tilde{\kappa} &= (\tilde{h} \kappa^* - \tilde{\kappa})' \tilde{\kappa} + \tilde{\kappa}' \tilde{\kappa} \\ &= 1 + (\tilde{h} \kappa^* - \tilde{\kappa})' (\tilde{\kappa} - \kappa^* \tilde{h}) + (\tilde{h} \kappa^* - \tilde{\kappa})' \kappa^* \tilde{h} \end{aligned} \quad (34)$$

Setting equal (33) and (34), we obtain

$$\tilde{h}^2 = ((\kappa^*)' \kappa^*)^{-1} \left(1 + \sum_{t=1}^T (\tilde{\kappa}_t - \kappa_t^* \tilde{h})^2 \right) \quad (35)$$

after rearranging terms. From lemma B.1, it is known that $T^{-3}(\boldsymbol{\kappa}^*)'\boldsymbol{\kappa}^* \xrightarrow{d} \delta^2/12$, implying that

$$\tilde{h}^2 = T^{-3} \frac{12}{\delta^2} \left(1 + O_p(X^{-1/2}T^{-2}) + O_p(T^{-5}) \right). \quad (36)$$

which implies the required result. \blacksquare

Lemma B.5 *Under the assumptions of Lemma B.2,*

$$T^{-3/2}(\tilde{\boldsymbol{\kappa}}' - \tilde{h}\boldsymbol{\kappa}^{*'})\boldsymbol{\kappa}^* = O_p(X^{-1/2}T^{-3/2}) + O_p(T^{-3})$$

Proof of Lemma B.5. We can rewrite the expression as

$$T^{-3/2} \sum_{t=1}^T (\tilde{\kappa}_t - \tilde{h}\kappa_t^*)\kappa_t^* = \left(XT^3\lambda_1^{-1} \right) T^{-3/2} \sum_{t=1}^T (j_{1,t} + j_{2,t} + j_{3,t} + j_{4,t})\kappa_t^* \quad (37)$$

where

$$\begin{aligned} T^{-3/2} \sum_{t=1}^T j_{1,t}\kappa_t^* &= X^{-1}T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T \kappa_t^* \boldsymbol{\beta}' \mathbf{e}_s^* \tilde{\kappa}_s \kappa_t^* \\ &= X^{-1}T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T \kappa_t^* \boldsymbol{\beta}' \mathbf{e}_s^* \kappa_s^* \tilde{h} \kappa_t^* + X^{-1}T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T \kappa_t^* \boldsymbol{\beta}' \mathbf{e}_s^* (\tilde{\kappa}_s - \kappa_s^* \tilde{h}) \kappa_t^*. \end{aligned} \quad (38)$$

Using assumption 6(ii), we can determine the order of the first part of the above expression to be

$$\begin{aligned} X^{-1}T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T \kappa_t^* \boldsymbol{\beta}' \mathbf{e}_s^* \kappa_s^* \tilde{h} \kappa_t^* &= X^{-1/2} \left(T^{-3} \sum_{s=1}^T (\kappa_s^*)^2 \right) \left(X^{-1/2}T^{-3/2} \sum_{s=1}^T \boldsymbol{\beta}' \mathbf{e}_s^* \kappa_s^* \right) \tilde{h} \\ &= O_p(X^{-1/2}T^{-3/2}). \end{aligned} \quad (39)$$

The second part is given by

$$\begin{aligned} &X^{-1}T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T \kappa_t^* \boldsymbol{\beta}' \mathbf{e}_s^* (\tilde{\kappa}_s - \kappa_s^* \tilde{h}) \kappa_t^* \\ &\leq X^{-1/2}T^{-1/2} \left(T^{-3} \sum_{t=1}^T (\kappa_t^*)^2 \right) \left(T^{-1} \sum_{s=1}^T (X^{-1/2} \boldsymbol{\beta}' \mathbf{e}_s^*)^2 \right)^{1/2} \left(T^{-1} \sum_{s=1}^T (\tilde{\kappa}_s - \kappa_s^* \tilde{h})^2 \right)^{1/2} \\ &= O_p(X^{-1}T^{-2}) + O_p(X^{-1/2}T^{-7/2}). \end{aligned} \quad (40)$$

The term involving $j_{2,t}$ is solved similarly.

$$\begin{aligned}
T^{-3/2} \sum_{t=1}^T j_{2,t} \kappa_t^* &= X^{-1} T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T \mathbf{e}_t^{*'} \beta \kappa_s^* \tilde{\kappa}_s \kappa_t^* \\
&= X^{-1} T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T \mathbf{e}_t^{*'} \beta \kappa_s^* \tilde{h} \kappa_s^* \kappa_t^* + X^{-1} T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T \mathbf{e}_t^{*'} \beta \kappa_s^* (\tilde{\kappa}_s - \kappa_s^* \tilde{h}) \kappa_t^*
\end{aligned} \tag{41}$$

The first part of the summation above is

$$\begin{aligned}
X^{-1} T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T \mathbf{e}_t^{*'} \beta \kappa_s^* \tilde{h} \kappa_s^* \kappa_t^* &= X^{-1/2} \left(T^{-3} \sum_{s=1}^T (\kappa_s^*)^2 \right) \left(X^{-1/2} T^{-3/2} \sum_{t=1}^T \mathbf{e}_t^{*'} \beta \kappa_t^* \right) \tilde{h} \\
&= O_p(X^{-1/2} T^{-3/2})
\end{aligned} \tag{42}$$

under assumption 6(ii). The same assumption is used for the second part, giving

$$\begin{aligned}
&X^{-1} T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T \mathbf{e}_t^{*'} \beta \kappa_s^* (\tilde{\kappa}_s - \kappa_s^* \tilde{h}) \kappa_t^* \\
&\leq X^{-1/2} T^{-1} \left(T^{-1} \sum_{s=1}^T (\tilde{\kappa}_s - \kappa_s^* \tilde{h})^2 \right)^{1/2} \left(T^{-3} \sum_{s=1}^T (\kappa_s^*)^2 \left(X^{-1/2} T^{-3/2} \sum_{t=1}^T \mathbf{e}_t^{*'} \beta \kappa_t^* \right)^2 \right)^{1/2} \\
&= O_p(X^{-1} T^{-5/2}) + O_p(X^{-1/2} T^{-4})
\end{aligned} \tag{43}$$

For the term involving $j_{3,t}$,

$$\begin{aligned}
T^{-9/2} \sum_{t=1}^T j_{3,t} \kappa_t^* &= T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T E[\mathbf{e}_t^{*'} \mathbf{e}_s^{*'}] \tilde{\kappa}_s \kappa_t^* \\
&= T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T E[\mathbf{e}_t^{*'} \mathbf{e}_s^{*'}] \kappa_s^* \tilde{h} \kappa_t^* + T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T E[\mathbf{e}_t^{*'} \mathbf{e}_s^{*'}] (\tilde{\kappa}_s - \kappa_s^* \tilde{h}) \kappa_t^*
\end{aligned} \tag{44}$$

The second term above can be written

$$\begin{aligned}
&T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T E[\mathbf{e}_t^{*'} \mathbf{e}_s^{*'}] (\tilde{\kappa}_s - \kappa_s^* \tilde{h}) \kappa_t^* \\
&\leq T^{-2} \left(T^{-1} \sum_{t=1}^T \sum_{s=1}^T \tau_{ts}^2 \right)^{1/2} \left(T^{-1} \sum_{s=1}^T (\tilde{\kappa}_s - \kappa_s^* \tilde{h})^2 T^{-3} \sum_{t=1}^T (\kappa_t^*)^2 \right)^{1/2} \\
&= O_p(X^{-1/2} T^{-7/2}) + O_p(T^{-5}),
\end{aligned} \tag{45}$$

using Assumption 3(ii) to obtain the order result. The order of the first term is determined by

its nonzero expected value, whose order can be taken from assumption 3(v).

$$\begin{aligned}
E \left[T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T E[\mathbf{e}_t^{*'} \mathbf{e}_s^*] \kappa_s^* \tilde{h} \kappa_t^* \right] &= T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T E[\mathbf{e}_t^{*'} \mathbf{e}_s^*] E[\kappa_s^* \kappa_t^*] \tilde{h} \\
&\leq T^{-7/2} T^{-1} \sum_{t=1}^T \sum_{s=1}^T \tau_{st} T^2 \delta^2 \tilde{h} \\
&= O(T^{-3})
\end{aligned} \tag{46}$$

Finally, letting $\tilde{\gamma}(s, t) = \mathbf{e}_t^{*'} \mathbf{e}_s^* / X - E[\mathbf{e}_t^{*'} \mathbf{e}_s^*]$,

$$\begin{aligned}
T^{-9/2} \sum_{t=1}^T j_{4,t} \kappa_t^* &= T^{-6} \sum_{t=1}^T \sum_{s=1}^T \tilde{\gamma}(s, t) \tilde{\kappa}_s \kappa_t^* \\
&= T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T \tilde{\gamma}(s, t) \kappa_s^* \tilde{h} \kappa_t^* + T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T \tilde{\gamma}(s, t) (\tilde{\kappa}_s - \kappa_s^* \tilde{h}) \kappa_t^*,
\end{aligned} \tag{47}$$

whose second part is

$$\begin{aligned}
&T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T \tilde{\gamma}(s, t) (\tilde{\kappa}_s - \kappa_s^* \tilde{h}) \kappa_t^* \\
&\leq X^{-1/2} T^{-3/2} \left(T^{-2} \sum_{t=1}^T \sum_{s=1}^T X^{1/2} \tilde{\gamma}(s, t)^2 \right)^{1/2} \left(T^{-1} \sum_{s=1}^T (\tilde{\kappa}_s - \kappa_s^* \tilde{h})^2 T^{-3} \sum_{t=1}^T (\kappa_t^*)^2 \right)^{1/2} \\
&= O_p(X^{-1} T^{-3}) + O_p(X^{-1/2} T^{-9/2}).
\end{aligned} \tag{48}$$

The first part analyzed using assumption 6(i), yielding

$$\begin{aligned}
&T^{-9/2} \sum_{t=1}^T \sum_{s=1}^T \tilde{\gamma}(s, t) \kappa_s^* \tilde{h} \kappa_t^* \\
&\leq X^{-1/2} T^{-1} \left(T^{-1} \sum_{t=1}^T \left(X^{1/2} T^{-3/2} \sum_{s=1}^T \tilde{\gamma}(s, t) \kappa_s^* \right)^2 \right)^{1/2} \left(T^{-3} \sum_{t=1}^T (\kappa_t^*)^2 \right)^{1/2} \tilde{h} \\
&= O_p(X^{-1/2} T^{-5/2}).
\end{aligned} \tag{49}$$

Taking together all results, we obtain

$$T^{-3/2} (\tilde{\kappa}' - \tilde{h} \kappa^{*'}) \kappa^* = O_p(X^{-1/2} T^{-3/2}) + O_p(T^{-3}) \tag{50}$$

■

Lemma B.6 Under the assumptions of Lemma B.2,

$$(\tilde{\kappa}' - \tilde{h}\kappa^{*'})\tilde{\kappa}/T^{3/2} = O_p(X^{-1/2}T^{-3/2}) + O_p(T^{-3}) \quad (51)$$

Proof of Lemma B.6 Note that

$$(\tilde{\kappa}' - \tilde{h}\kappa^{*'})\tilde{\kappa}/T^{3/2} = (\tilde{\kappa}' - \tilde{h}\kappa^{*'}) (\tilde{\kappa} - \kappa^*\tilde{h})/T^{3/2} + (\tilde{\kappa}' - \tilde{h}\kappa^{*'})\kappa^*\tilde{h}/T^{3/2} \quad (52)$$

The lemma thus follows from Lemmas B.3 and B.5 . ■

Lemma B.7 Under the assumptions of Lemma B.2,

$$T^{-3/2}\kappa^{*'}\tilde{\kappa} = \delta/\sqrt{12} + O_p(T^{-1}) \quad (53)$$

Proof of Lemma B.7 Use the decomposition

$$T^{-3/2}\kappa^{*'}\tilde{\kappa} = T^{-3/2}\kappa^{*'}(\tilde{\kappa} - \tilde{h}\kappa^*) + (T^{-3}\kappa^{*'}\kappa^*)(T^{3/2}\tilde{h}) \quad (54)$$

Lemma B.5 establishes the order of the first term on the RHS. For the components of the second term, and Lemmas B.1 and B.4 can be used to obtain the result. ■

Proof of Proposition 1 Given that $\tilde{\kappa}_t$ is an estimator of $\kappa_t^*\tilde{h}$, the relation $\hat{\kappa}_t = \tilde{\kappa}_t\lambda_1^{1/2}$ implies that $\hat{\kappa}_t$ estimates $\kappa_t^*h = \kappa_t^*\tilde{h}\lambda_1^{1/2}$. Application of Lemmas B.2 and B.4 now yields the required result.

Proof of Theorem 1. Consider (4). Using Davidson (2000, 14.1.6) and White (2001, Th.5.20),

$$\hat{\alpha}_x \xrightarrow{d} \alpha_x + \lim_{T \rightarrow \infty} \beta_x \frac{T+1}{2} \delta + \sqrt{T}\beta_x N(0, \sigma_v^2/3) + \frac{1}{\sqrt{T}} N(0, \bar{\sigma}_x^2). \quad (55)$$

where $\bar{\sigma}_x^2 = \text{Var}[T^{-1/2} \sum_{t=1}^T e_{x,t}]$. ■

Proof of Theorem 2 Recall from (18) that the expansion

$$\sqrt{XT^3}(\tilde{\kappa}_t - \kappa_t^*\tilde{h}) = \sqrt{XT^3}(j_{1,t} + j_{2,t} + j_{3,t} + j_{4,t})XT^3\lambda_1^{-1}. \quad (56)$$

can be applied. Consider $\sqrt{X}j_{1,t}$, which we can be expanded to

$$\sqrt{XT^3}j_{1,t} = X^{-1/2}T^{-3/2} \sum_{s=1}^T \kappa_t^* \boldsymbol{\beta}' \mathbf{e}_s^* (\tilde{\kappa}_s - \kappa_s^* \tilde{h}) + X^{-1/2}T^{-3/2} \sum_{s=1}^T \kappa_t^* \boldsymbol{\beta}' \mathbf{e}_s^* \kappa_s^* \tilde{h} \quad (57)$$

For the first term on the RHS, we can apply Lemma B.3 to obtain

$$\begin{aligned} & X^{-1/2}T^{-3/2} \sum_{s=1}^T \kappa_t^* \boldsymbol{\beta}' \mathbf{e}_s^* (\tilde{\kappa}_s - \kappa_s^* \tilde{h}) \\ & \leq X^{-1/2}T^{-1} \left(\sum_{s=1}^T (\kappa_t^* \boldsymbol{\beta}' \mathbf{e}_s^*)^2 \right)^{1/2} \left(T^{-1} \sum_{s=1}^T (\tilde{\kappa}_s - \kappa_s^* \tilde{h})^2 \right)^{1/2} \\ & \leq T \left(T^{-2}(\kappa_t^*)^2 T^{-1} \sum_{s=1}^T (X^{-1/2} \boldsymbol{\beta}' \mathbf{e}_s^*)^2 \right)^{1/2} \left(O_p(X^{-1/2}T^{-3/2}) + O_p(T^{-3}) \right) \\ & = O_p(X^{-1/2}T^{-1/2}) + O_p(T^{-2}) \end{aligned} \quad (58)$$

The second term is

$$\begin{aligned} & X^{-1/2}T^{-3/2} \sum_{s=1}^T \kappa_t^* \boldsymbol{\beta}' \mathbf{e}_s^* \kappa_s^* \tilde{h} \\ & = T^1 \left(T^{-1} \kappa_t^* \right) \left(X^{-1/2}T^{-3/2} \sum_{s=1}^T \kappa_s^* \boldsymbol{\beta}' \mathbf{e}_s^* \right) \tilde{h} \end{aligned} \quad (59)$$

which is of order $O_p(T^{-1/2})$ by assumptions 3(iv) and 6(ii) and Lemma B.4. Taking together these two results, we obtain

$$\sqrt{XT^3}j_{1,t} = O_p(T^{-1/2}) \quad (60)$$

For the order of $j_{2,t}$ we use Lemma B.7 to obtain.

$$\begin{aligned} \sqrt{XT^3}j_{2,t} &= X^{-1/2}T^{-3/2} \sum_{s=1}^T \mathbf{e}_t^{*'} \boldsymbol{\beta} \kappa_s^* \tilde{\kappa}_s \\ &= X^{-1/2} \sum_{s=1}^T \mathbf{e}_t^{*'} \boldsymbol{\beta} \frac{\delta}{\sqrt{12}} + O_p(T^{-1}) \end{aligned} \quad (61)$$

Concerning $j_{3,t}$,

$$\sqrt{XT^3}j_{3,t} = \sqrt{X}T^{-3/2} \sum_{s=1}^T E[\mathbf{e}_t^{*'} \mathbf{e}_s^*] (\tilde{\kappa}_s - \kappa_s^* \tilde{h}) + \sqrt{X}T^{-3/2} \sum_{s=1}^T E[\mathbf{e}_t^{*'} \mathbf{e}_s^*] \kappa_s^* \tilde{h}, \quad (62)$$

where assumption 5 can be used to yield

$$\begin{aligned}
& \sqrt{XT}^{-3/2} \sum_{s=1}^T E[\mathbf{e}_t^{*'} \mathbf{e}_s^*] (\tilde{\kappa}_s - \kappa_s^* \tilde{h}) \\
& \leq \sqrt{XT}^{-1} \left(\sum_{s=1}^T |\tau_{ts}|^2 \right)^{1/2} \left(T^{-1} \sum_{s=1}^T (\tilde{\kappa}_s - \kappa_s^* \tilde{h})^2 \right)^{1/2} \\
& = \sqrt{XT}^{-1} O_p(1) \left(O_p(X^{-1/2} T^{-3/2}) + O_p(T^{-3}) \right) = O_p(T^{-5/2}) + O_p(\sqrt{XT}^{-4}). \tag{63}
\end{aligned}$$

Furthermore, note that $E[\sum_{s=1}^T E[\mathbf{e}_t^{*'} \mathbf{e}_s^*] \kappa_s^*] \leq \max_s E[|\kappa_s^*|] \sum_{s=1}^T |\tau_{ts}| \leq TK$ since the properties of the latent trend imply $E[|\kappa_t^*|] = O(T)$. Consequently,

$$\sqrt{XT}^{-3/2} \sum_{s=1}^T E[\mathbf{e}_t^{*'} \mathbf{e}_s^*] \kappa_s^* \tilde{h} = X^{-1} T^{-3/2} O_p(T) O_p(T^{-3/2}) = O_p(\sqrt{XT}^{-2}). \tag{64}$$

We can hence conclude that

$$\sqrt{XT}^{-3} j_{3,t} = \sqrt{XT}^{-3/2} \sum_{s=1}^T E[\mathbf{e}_t^{*'} \mathbf{e}_s^*] \kappa_s^* \tilde{h} + O_p(T^{-5/2}) + O_p(\sqrt{XT}^{-4}) = O_p(\sqrt{XT}^{-2}). \tag{65}$$

Finally,

$$\begin{aligned}
\sqrt{XT}^{-3} j_{4,t} &= \sqrt{XT}^{-3/2} \sum_{s=1}^T (\mathbf{e}_t^{*'} \mathbf{e}_s^* / X - E[\mathbf{e}_t^{*'} \mathbf{e}_s^*]) (\tilde{\kappa}_s - \kappa_s^* \tilde{h}) \\
&+ \sqrt{XT}^{-3/2} \sum_{s=1}^T (\mathbf{e}_t^{*'} \mathbf{e}_s^* / X - E[\mathbf{e}_t^{*'} \mathbf{e}_s^*]) \kappa_s^* \tilde{h}. \tag{66}
\end{aligned}$$

Here, assumption 3(iii) can be invoked to obtain

$$\begin{aligned}
& \sqrt{XT}^{-3/2} \sum_{s=1}^T (\mathbf{e}_t^{*'} \mathbf{e}_s^* / X - E[\mathbf{e}_t^{*'} \mathbf{e}_s^*]) (\tilde{\kappa}_s - \kappa_s^* \tilde{h}) \\
& \leq T^{-1/2} \left(T^{-1} \sum_{s=1}^T \left(X^{-1/2} \sum_{x=1}^X (e_{x,t}^* e_{x,s}^* - E[e_{x,t}^* e_{x,s}^*]) \right)^2 \right)^{1/2} \left(T^{-1} \sum_{s=1}^T (\tilde{\kappa}_s - \kappa_s^* \tilde{h})^2 \right)^{1/2} \\
& = T^{-1/2} \left(O_p(X^{-1/2} T^{-3/2}) + O_p(T^{-3}) \right) \\
& = O_p(X^{-1/2} T^{-2}) + O_p(T^{-7/2}) \tag{67}
\end{aligned}$$

and

$$\begin{aligned}
& \sqrt{X}T^{-3/2} \sum_{s=1}^T (\mathbf{e}_t^{*'} \mathbf{e}_s^* / X - E[\mathbf{e}_t^{*'} \mathbf{e}_s^*]) \kappa_s^* \tilde{h} \\
& \leq T^{1/2} \left(T^{-1} \sum_{s=1}^T (\mathbf{e}_t^{*'} \mathbf{e}_s^* / X^{1/2} - E[\mathbf{e}_t^{*'} \mathbf{e}_s^*])^2 \right)^{1/2} \left(T^{-3} \sum_{s=1}^T (\kappa_s^*)^2 \right)^{1/2} \tilde{h} \\
& = O_p(T^{-1}).
\end{aligned} \tag{68}$$

Collecting the order results of all components of (56), we obtain

$$\sqrt{XT}(\tilde{\kappa}_t - \kappa_t^* \tilde{h}) = \left(X^{-1/2} \mathbf{e}_t^{*'} \boldsymbol{\beta} \right) \frac{\delta}{\sqrt{12}} XT^3 \lambda_1^{-1} + O_p(T^{-1/2}) + O_p(\sqrt{X}T^{-2}) \xrightarrow{d} \sqrt{12} \delta \sigma_\beta N(0, \Gamma_t), \tag{69}$$

where the last result is obtained from using Assumption 2 and Lemmas B.2 and B.7. Pre-multiplication of the above with $(X^{-1}T^{-3}\lambda_1)^{1/2}$ yields the equivalent result for $\hat{\kappa}$, given the asymptotic results on λ_1 in Lemma B.2. ■

Proof of Theorem 3 Consider alternative estimator $\tilde{\boldsymbol{\beta}} = \mathbf{M}^{*'} \tilde{\boldsymbol{\kappa}}$. Recalling the factor structure of \mathbf{M}^* , we can write

$$\begin{aligned}
\tilde{\boldsymbol{\beta}}_x &= \beta_x \tilde{h}^{-1} \tilde{h} \boldsymbol{\kappa}^{*'} \tilde{\boldsymbol{\kappa}} + \mathbf{e}_x^{*'} \tilde{\boldsymbol{\kappa}} \\
&= \beta_x \tilde{h}^{-1} + \beta_x \tilde{h}^{-1} (\tilde{h} \boldsymbol{\kappa}^{*'} - \tilde{\boldsymbol{\kappa}}') \tilde{\boldsymbol{\kappa}} + \mathbf{e}_x^{*'} \boldsymbol{\kappa}^* \tilde{h} + \mathbf{e}_x^{*'} (\tilde{\boldsymbol{\kappa}} - \boldsymbol{\kappa}^* \tilde{h})
\end{aligned} \tag{70}$$

where the normalisation $\tilde{\boldsymbol{\kappa}}' \tilde{\boldsymbol{\kappa}} = 1$ is used in the last step. The order of the second term on the RHS can be inferred from Lemma B.6. Furthermore

$$\begin{aligned}
\mathbf{e}_x^{*'} (\tilde{\boldsymbol{\kappa}} - \boldsymbol{\kappa}^* \tilde{h}) &= \sum_{t=1}^T e_{x,t}^* (\tilde{\kappa}_t - \kappa_t^* \tilde{h}) \\
&= T \left(T^{-1} \sum_{t=1}^T (e_{x,t}^*)^2 \right)^{1/2} \left(T^{-1} \sum_{t=1}^T (\tilde{\kappa}_t - \kappa_t^* \tilde{h})^2 \right)^{1/2} \\
&= O_p(X^{-1/2} T^{-1/2}) + O_p(T^{-2})
\end{aligned} \tag{71}$$

and

$$\begin{aligned}
\mathbf{e}_x^{*'} \boldsymbol{\kappa}^* \tilde{h} &= \left(T^{-3/2} \sum_{t=1}^T e_{x,t}^* \kappa_t^* \right) (T^{3/2} \tilde{h}) \\
&\sim \frac{\sqrt{12}}{\delta} N(0, \Phi_x) + O_p(X^{-1/2} T^{-1}) + O_p(T^{-5/2})
\end{aligned} \tag{72}$$

as implied by Assumption 3(v) and Lemma B.4.

Using these order results, we can hence write

$$\begin{aligned}\tilde{\beta}_x - \beta_x \tilde{h}^{-1} &= \sum_{t=1}^T e_{x,t}^* \kappa_t^* \tilde{h} + O_p(X^{-1/2}) + O_p(T^{-3/2}) \\ &\sim \frac{\sqrt{12}}{\delta} N(0, \Phi_x) + O_p(X^{-1/2}) + O_p(T^{-1/2})\end{aligned}\quad (73)$$

The result for the estimator from the Lee-Carter model follows from the fact that $\hat{\beta} = \tilde{\beta} \lambda_1^{-1/2}$ (see e.g. Bai and Ng, 2008, section 3), yielding

$$\sqrt{XT^3}(\hat{\beta}_x - \beta_x(\tilde{h}\sqrt{\lambda_1})^{-1}) \sim \frac{12}{\delta^2 \sigma_\beta} N(0, \Phi_x) + O_p(X^{-1/2}) + O_p(T^{-1/2}) \quad (74)$$

■

Proof of Theorem 4 Note first that

$$\begin{aligned}\hat{m}_{x,t} - \alpha_x - \kappa_t \beta_x &= \hat{\alpha}_x + \hat{\kappa}_t \hat{\beta}_x - \kappa_t \beta_x \\ &= \alpha_x + T^{-1} \sum_{t=1}^T \kappa_t + T^{-1} \sum_{t=1}^T e_{x,t} + \hat{\kappa}_t \hat{\beta}_x - (\kappa_t^* + T^{-1} \sum_{t=1}^T \kappa_t) \beta_x \\ &= T^{-1} \sum_{t=1}^T e_{x,t} + \hat{\kappa}_t \hat{\beta}_x - \kappa_t^* \beta_x.\end{aligned}\quad (75)$$

Furthermore, we can substitute $\tilde{\kappa}_t \tilde{\beta}_x$ for $\hat{\kappa}_t \hat{\beta}_x$ since the two products are identical. Additionally expanding the term $(\tilde{\kappa}_t \tilde{\beta}_x - \kappa_t^* \beta_x)$, we obtain

$$\hat{m}_{x,t} - \alpha_x - \kappa_t \beta_x = T^{-1} \sum_{t=1}^T e_{x,t} + (\tilde{\kappa}_t - \kappa_t^* \tilde{h}) \tilde{h}^{-1} \beta_x + \tilde{\kappa}_t (\tilde{\beta}_x - \beta_x \tilde{h}^{-1}). \quad (76)$$

As argued in Theorem 1, $T^{-1/2} \sum_{t=1}^T e_{x,t} \xrightarrow{d} N(0, \bar{\sigma}_x^2)$. Next, using Theorem 2 the second part on the RHS can be written

$$\begin{aligned}(\tilde{\kappa}_t - \kappa_t^* \tilde{h}) \tilde{h}^{-1} \beta_x &= X^{-1/2} T^{-3/2} \left(X^{-1/2} \mathbf{e}_t^* \beta \frac{\delta}{\sqrt{12}} X T^3 \lambda_1^{-1} + O_p(T^{-1/2}) + O_p(\sqrt{X} T^{-2}) \right) \tilde{h}^{-1} \beta_x \\ &= X^{-1/2} \left(X^{-1/2} \mathbf{e}_t^* \beta \right) \sigma_\beta^{-2} \beta_x + O_p((XT)^{-1/2}) + O_p(T^{-2}),\end{aligned}\quad (77)$$

implying that

$$\sqrt{X}(\tilde{\kappa}_t - \kappa_t^* \tilde{h}) \tilde{h}^{-1} \beta_x \sim \sigma_\beta^{-1} \beta_x N(0, \Gamma_t \sigma_\beta^{-2}) + O_p(T^{-1/2}) + O_p(\sqrt{X} T^{-2}). \quad (78)$$

Concerning the third term, $\tilde{\kappa}_t$ can be expanded, yielding

$$\begin{aligned}\tilde{\kappa}_t(\tilde{\beta}_x - \beta_x \tilde{h}^{-1}) &= \kappa_t^* \tilde{h}(\hat{\beta}_x - \beta_x \tilde{h}^{-1}) + (\tilde{\kappa}_t - \kappa_t^* \tilde{h})(\tilde{\beta}_x - \beta_x \tilde{h}^{-1}) \\ &= T^{-1/2}(T^{-1}\kappa_t^*) \left(T^{-3/2}(\mathbf{e}_x^*)' \boldsymbol{\kappa}^* \right) T^3 \tilde{h}^2 + O_p(X^{-1/2}T^{-3/2}) + O_p(T^{-3}).\end{aligned}\tag{79}$$

From this result, we obtain

$$\begin{aligned}\sqrt{T}\tilde{\kappa}_t(\tilde{\beta}_x - \beta_x \tilde{h}^{-1}) &= \delta \left(\frac{t}{T} - \frac{1}{2} + O_p(T^{-1/2}) \right) \left(T^{-3/2}(\mathbf{e}_x^*)' \boldsymbol{\kappa}^* \frac{12}{\delta^2} + O_p(X^{-1/2}) + O_p(T^{-1/2}) \right) \\ &\sim \frac{12}{\delta} \left(\frac{t}{T} - \frac{1}{2} \right) N(0, \Phi_x)\end{aligned}\tag{80}$$

Collecting the above results and letting

$$v_1 = \left(\sigma_\beta^{-1} \beta_x \right)^2 \Gamma_t \sigma_\beta^{-2}$$

and

$$v_2 = \bar{\sigma}_x^2 + \left(\frac{12}{\delta} \left(\frac{t}{T} - \frac{1}{2} \right) \right)^2 \Phi_x$$

we can state that

$$(X^{-1}v_1 + T^{-1}v_2)^{-1/2}(\hat{m}_{x,t} - \kappa_t^* \beta_x) \xrightarrow{d} N(0, 1).\tag{81}$$

■

Proof of Lemma 4.1 By the definition of the estimator and by Theorem 2, we can write

$$\hat{\delta} = (T-1)^{-1} \sum_{t=2}^T (\delta h + v_t h + X^{-1/2} \Delta \mathbf{e}_t' \boldsymbol{\beta} \sigma_\beta^{-1} + O_p(T^{-1/2}) + O_p(\sqrt{XT}^{-2}))\tag{82}$$

As shown in the proof of Theorem 2, the $O_p(T^{-1/2})$ -term represents $\sqrt{XT^3}j_{1,t}$. Now note that

$$\begin{aligned}(T-1)^{-1} \sum_{t=2}^T \sqrt{XT^3} \Delta j_{1,t} &= X^{-1/2} T^{-3/2} (T-1)^{-1} \sum_{t=2}^T \Delta \kappa_t^* \sum_{s=1}^T \boldsymbol{\beta}' \mathbf{e}_s^* (\tilde{\kappa}_s - \kappa_s^* \tilde{h}) \\ &\quad + X^{-1/2} T^{-3/2} (T-1)^{-1} \sum_{t=2}^T \Delta \kappa_t^* \sum_{s=1}^T \boldsymbol{\beta}' \mathbf{e}_s^* \kappa_s^* \tilde{h}.\end{aligned}\tag{83}$$

The first term of this expression is given by

$$\begin{aligned}
& X^{-1/2}T^{-3/2}(T-1)^{-1} \sum_{t=2}^T \Delta \kappa_t^* \sum_{s=1}^T \boldsymbol{\beta}' \mathbf{e}_s^* (\tilde{\kappa}_s - \kappa_s^* \tilde{h}) \\
& \leq T^{-1/2}(\delta + (T-1)^{-1/2} \frac{1}{\sqrt{T-1}} \sum_{t=1}^T \nu_t) (T^{-1} \sum_{s=1}^T (X^{-1/2} \boldsymbol{\beta}' \mathbf{e}_s^*)^2)^{1/2} (T^{-1} \sum_{s=1}^T (\tilde{\kappa}_s - \kappa_s^* \tilde{h})^2)^{1/2} \\
& = T^{-1/2}(O_p(X^{-1/2}T^{-3/2}) + O_p(T^{-3})) \\
& = O_p(X^{-1/2}T^{-2}) + O_p(T^{-7/2})
\end{aligned} \tag{84}$$

similarly, for the first part we obtain

$$\begin{aligned}
& X^{-1/2}T^{-3/2}(T-1)^{-1} \sum_{t=2}^T \Delta \kappa_t^* \sum_{s=1}^T \boldsymbol{\beta}' \mathbf{e}_s^* \kappa_s^* \tilde{h} \\
& = (\delta + (T-1)^{-1/2} \frac{1}{\sqrt{T-1}} \sum_{t=1}^T \nu_t) (X^{-1/2}T^{-3/2} \sum_{s=1}^T \boldsymbol{\beta}' \mathbf{e}_s^* \kappa_s^*) \tilde{h} \\
& = O_p(T^{-3/2})
\end{aligned} \tag{85}$$

This result implies that we can write

$$\begin{aligned}
\sqrt{\frac{T-1}{X}}(\hat{\delta} - \delta h) &= X^{-1/2} \frac{1}{\sqrt{T-1}} \sum_{t=2}^T \nu_t h \\
&+ X^{-1/2} \frac{1}{\sqrt{X(T-1)}} \sum_{t=2}^T \Delta \mathbf{e}_t' \boldsymbol{\beta} \sigma_\beta^{-1} + O_p(X^{-1/2}T^{-1}) + O_p(T^{-3/2}) \\
&\sim N(0, \sigma_\beta^2 \sigma_v^2) + O_p(X^{-1/2}) + O_p(T^{-3/2}),
\end{aligned} \tag{86}$$

which completes the proof. ■

Proof of Theorem 5 We can write $\kappa_{T+q} = \kappa_T + q\delta + \sum_{p=1}^q \eta_{T+p}$. Hence,

$$\hat{\kappa}_{T+q|T} - \kappa_{T+q}^* h = (\hat{\kappa}_T - \kappa_T^* h) + q(\hat{\delta} - \delta h) + \sum_{p=1}^q \eta_{T+p} \tag{87}$$

From Theorem 2 and Lemma 4.1 it is known that the first two terms above are asymptotically normal with variances $\Gamma_t \sigma_\beta^{-2}$ and $X/T \sigma_v^2 \sigma_\beta^2 + T^{-1} \sigma_\beta^{-1} \Gamma_\Delta$ respectively. All covariances between individual terms of the expression above are equal to zero by assumption except for $\text{Cov}[\hat{\kappa}_T, \hat{\delta}]$. We can rewrite the latter

$$\text{Cov}[\hat{\kappa}_T, \hat{\delta}] = \sigma_\beta^{-2} E \left[X^{-1/2} (\mathbf{e}_T^*)' \boldsymbol{\beta} \frac{1}{T-1} \sum_{s=2}^T X^{-1/2} (\Delta \mathbf{e}_s)' \boldsymbol{\beta} \right] + O_p(\sqrt{X}T^{-2}) + O_p(XT^{-4}),$$

where the first part is

$$\begin{aligned}
& \sigma_{\beta}^{-2} E \left[X^{-1/2} (\mathbf{e}_T^*)' \beta \frac{1}{T-1} \sum_{s=2}^T X^{-1/2} (\Delta \mathbf{e}_s)' \beta \right] \\
&= \sigma_{\beta}^{-2} E \left[X^{-1/2} (\mathbf{e}_T^*)' \beta \frac{1}{T-1} (X^{-1/2} \mathbf{e}_T' \beta - X^{-1/2} \mathbf{e}_1' \beta) \right] \\
&= \sigma_{\beta}^{-2} E \left[\frac{1}{X(T-1)} \sum_{x=1}^X \sum_{y=1}^X e_{x,T}^* \beta_x e_{y,T} \beta_y \right] - \sigma_{\beta}^{-2} E \left[\frac{1}{X(T-1)} \sum_{x=1}^X \sum_{y=1}^X e_{x,1}^* \beta_x e_{y,T} \beta_y \right] \\
&= O(T^{-1}).
\end{aligned}$$

Thus, the covariance term can be disregarded if $X/T^4 \rightarrow 0$ and asymptotically the variance of the normally distributed term $\hat{\kappa}_{T+q|T} - \kappa_{T+q}^* h$ is given by $\text{Var}[\hat{\kappa}_{T+q|T}] = \text{aVar}[\kappa_T] + q^2 (\text{aVar}[\delta] + \sigma_{\beta}^2)$

Proof of Theorem 6 We can use the decomposition

$$\begin{aligned}
\hat{m}_{x,T+q|T} - m_{x,T+q} &= (\hat{m}_{x,T} - m_{x,T}) + q(\hat{\delta} - \delta h)h^{-1}\beta_x + q\hat{\delta}(\hat{\beta}_x - h^{-1}\beta_x) \\
&\quad - \sum_{p=1}^q \nu_{T+p}\beta_x - e_{x,T+q}
\end{aligned} \tag{88}$$

Theorem 4 establishes the limiting distribution of the first term on the RHS. Concerning the second term, note that

$$\begin{aligned}
(\hat{\delta} - \delta h)h^{-1}\beta_x &= X^{-1/2}(\hat{\delta} - \delta h)\sigma_{\beta}^{-1}\beta_x \\
&\sim (T-1)^{-1/2}\beta_x N(0, \beta_x^2 \sigma_v^2 + X^{-1}\sigma_{\beta}^{-2}\bar{\Gamma}_{\Delta}) + O_p(T^{-1})
\end{aligned} \tag{89}$$

by application of Lemma 4.1. Concerning the third term, note that

$$\begin{aligned}
\delta(\hat{\beta}_x - h^{-1}\beta_x) &= O_p \left(\sqrt{\frac{X}{T-1}} \right) O_p \left(\frac{1}{\sqrt{XT^3}} \right) \\
&= O_p(T^{-2})
\end{aligned} \tag{90}$$

which implies that this term can be disregarded. Using the assumptions on ν_t and $e_{x,t}$, we can hence conclude that

$$\hat{m}_{x,T+q|T} - m_{x,T+q} \sim N(0, \text{Var}(\hat{m}_{x,T+q|T})) + O_p(T^{-1}) \tag{91}$$

where $\text{Var}[\hat{m}_{x,T+q|T}] = \text{aVar}[\hat{m}_{x,T}] + q^2(\beta_x^2 \text{aVar}[\hat{\delta}] + \sigma_v^2) + \sigma_e^2$