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

Simon Reese*

March 30, 2015

PRELIMINARY DRAFT

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 establishes the assumptions required for these to hold. Variance estimators are presented in order to allow hypothesis testing and the computation of confidence intervals.

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-exhausitve 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.

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 $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \dots & \alpha_X \end{bmatrix}'$ and $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 & \dots & \beta_X \end{bmatrix}'$ are age-group specific intercepts and slope coefficients respectively and $\boldsymbol{\kappa} = \begin{bmatrix} \kappa_1 & \dots & \kappa_T \end{bmatrix}'$ is an unobserved time trend. $\mathbf{E}$ 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 $\boldsymbol{\kappa}$ is the only latent factor that is assumed in (1). Its properties are not further specified by Lee and Carter, but the authors treat the estimated trend as a Random Walk with Drift (RWD) when forecasting mortality rates. We take this practice as a reason to assume that the common trend itself is a RWD, formally defining it as

$$\kappa_t = \delta + \kappa_{t-1} + \nu_t. \tag{2}$$

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:

$$\kappa_t = t\delta + \eta_t \tag{3}$$

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

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

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

$$\hat{\alpha} = T^{-1} \sum_{t=1}^{T} m_{x,t} \tag{5}$$

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

$$T^{-1} \sum_{t=1}^{T} t\delta = \frac{T+1}{2}\delta \neq 0 \tag{6}$$

which implies that the estimator of α is biased. The consequences of this result for the estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ 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 $\mathbf{M}^* = \mathbf{M} - \hat{\boldsymbol{\alpha}}$ denote the individually demeaned mortality rates. An estimator of $\boldsymbol{\beta}$ which satisfies the additional identifying assumption $\sum_{x=1}^{X} \beta_x = 1$ is obtained using the singular value decomposition (SVD) $\mathbf{M}^* = \mathbf{U\Sigma V}'$, where $\mathbf{U}$ ($\mathbf{V}$) is the $T \times T$ ($X \times X$) matrix containing the left (right) singular vectors of $\mathbf{M}^*$ and $\boldsymbol{\Sigma}$ is a $T \times X$ matrix with the singular values $\sigma_i \; \forall i = 1, \ldots, \min\{X, T\}$ 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 assumption, which is a necessary normalisation for the SVD, is treated as an inherent characteristic of the data generating process. It is unclear what $\hat{\boldsymbol{\beta}}$ 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 estimated parameters are treated as known.

It is well known that the SVD used in the Lee-Carter model is directly related to the eigendecomposition since $\mathbf{V}$ ($\mathbf{U}$) is the matrix of eigenvectors of $\mathbf{M}^{*\prime}\mathbf{M}^*$ ($\mathbf{M}^{*\prime}\mathbf{M}^*$). 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. The balance of the paper is as follows: Section 2 establishes the assumptions that need to be made and reports the asymptotic distributions of the model estimates. In section 3, estimators for the asymptotic variance are presented and the implication of estimation uncertainty for the confidence intervals is discussed. Finally, section 4 concludes.

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 in addition to the definition of the model's log-linear structure. Additional assumptions have to be made in order to ensure some desirable properties of the estimators. In the following, let K denote a generic finite number. Furthermore, note that on several occasions the expression $X \to \infty$ 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

[1] In this paper, we use the normalization $\sum_{x=1}^{X} \beta_x^2 = 1$, which is in line with common practice.

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 *$\nu_t \sim i.i.d(0, \sigma_\nu)$, $E[\nu^4] \leq K$, $T^{-3}\boldsymbol{\kappa}'\boldsymbol{\kappa} \in (0, K]$, and $\kappa_0 \leq K$.*

Assumption 2 *β_x is either deterministic s.t. $|\beta_x| \leq K$, or stochastic s.t. $E\left[\beta_x^4\right] \leq K$. In either case, $X^{-1}\boldsymbol{\beta}'\boldsymbol{\beta} \xrightarrow{p} \sigma_{\boldsymbol{\beta}}^2 > 0$ as $X \to \infty$.*

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* $X \to \infty$ *where*

$\Gamma_t = \lim_{X\to\infty} X^{-1}\sum_{x=1}^{X}\sum_{y=1}^{X} E[\beta_x\beta_y e_{x,t}e_{y,t}]$.

(v) *For each* x, $T^{-3/2}\sum_{t=1}^{T}\kappa_t e_{x,t} \xrightarrow{d} N(0, \Phi_x)$ *as* $T \to \infty$ *where*

$\Phi_x = \lim_{T\to\infty} T^{-3}\sum_{t=1}^{T}\sum_{s=1}^{T} E[\kappa_t\kappa_s e_{x,s}e_{x,t}]$.

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

Assumption 5 *For all $t \leq T$, $x \leq X$, $\sum_s^T \tau_{ts} \leq K$ 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) $E\left[X^{-1/2}T^{-3/2}\sum_{t=1}^{T}\mathbf{e}_t^{*\prime}\boldsymbol{\beta}\kappa_t^*\right]^2 \leq K$

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, 2005, Lee and Miller, 2001) heteroskedasticity can be allowed for and even weak serial 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 $\boldsymbol{\kappa}$ 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,*

$$\alpha_x - \hat{\alpha}_x \xrightarrow{d} \lim_{T\to\infty} \frac{T+1}{2}\beta_x\delta + \sqrt{T}\beta_x N(0,\sigma_v^2/3) + T^{-1/2}N(0,\overline{\sigma}_x^2) \tag{7}$$

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

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. Concerning the comparisons of coefficients, Theorem 1 has different purpose-specific implications. Testing for the equality of two or more intercepts is unproblematic. Their estimates are affected by exactly the same bias term and the only relevant source of estimation error is the third term on the RHS of (7). This term converges to zero at rate $\sqrt{T}$ and a variance estimator can be calculated from the model residuals. However, there are cases where intercept coefficients are compared for the same age group in different samples, such as mortality rates in different countries and/or from different causes. The mortality rates in these different samples can not necessarily be assumed to be driven by the same latent trend which means that the bias term in the estimated coefficients differs. Comparing those estimators thus does not yield any reliable indication about the relation of the true parameter values.

It is easy to see that the bias in the estimation of α_x implies that the estimator of $\boldsymbol{\kappa}$, which is obtained from the difference of $m_{x,t}$ and α_x, is an estimator of the *demeaned* RWD. A further point concerning the estimation of the latent trend is that only the product $\kappa_t\beta_x$. Still this product is identical to $(\kappa_t\hat{h})(\hat{h}^{-1}\beta_x)$ for any $\hat{h} \neq 0$, implying that the factor and its coefficients can only be estimated up to a factor. In the Lee-Carter model, where only one latent factor is given, the SVD chooses a normalisation which has a fairly clear expression. Its asymptotic limit is given in the following.

Proposition 1 *Under assumptions 1-6,*

$$\hat{h} = X^{1/2}\sigma_\beta + O_p(X^{1/2}T^{-1/2}) + O_p(T^{-1}) \tag{8}$$

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

Given the definition of the normalisation $\hat{h}$, it is possible to state the asymptotic distribution of the estimated factors.

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

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

where $\kappa_t^* = \kappa_t - T^{-1}\sum_{t=1}^{T} kappa_t$.

This theorem adds an important insight into the literature showing that $\hat{\kappa}_t$ is subject to a certain persisting degree of estimation uncertainty irrespectively of the amount of information used. This bears an important consequence for the interpretation of the latent trend since the usual large sample argumentation that motivates neglecting the estimation error does not apply here.

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

Remark The persistence of estimation uncertainty in $\hat{\kappa}_t$ is a direct consequence of the chosen normalisation in the SVD. This 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 $\boldsymbol{\beta}$ and $\boldsymbol{\kappa}$ to converge by adapting the estimation steps in the PC literature (see e.g. Bai and Ng, 2008).

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,*

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

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

In contrast to $\hat{\kappa}_t$, $\hat{\beta}_x$ converges fairly quickly to a multiple of the true parameters vector. Additionally, the multiplier $\hat{h}$ allows for a rather cumbersome interpretation of $\hat{\beta}_x$ as an estimator of $\sqrt{X}$ times the slope coefficient for one age group normalized by the root mean squared deviation of all slope coefficients from zero.

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

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

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

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

$$v_1 = \left(\frac{\delta^2}{12}\beta_x\right)^2 \Gamma_t \tag{9}$$

and

$$v_2 = \left(\frac{12}{\delta}\left(\frac{t}{T} - \frac{1}{2}\right)\right)^2 \Phi_x \tag{10}$$

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 either $X^{-1/2}$ or $T^{-1/2}$.

3 Variance estimators

The asymptotic variances in Theorems 1–3 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

$$aVar[\hat{\kappa}_t] = \frac{\delta^4 \sigma_\beta^2}{12^2} plim \sum_{x=1}^{X} X^{-1/2} \sigma_{x,t}^2 \boldsymbol{\beta}_x^2 \tag{11}$$

can be estimated as

$$\widehat{aVar}[\hat{\kappa}_t] = \sum_{y=1}^{X} (\hat{e}_{y,t}^*)^2 \hat{\beta}_y^2, \tag{12}$$

where $\hat{e}^*_{x,t} = m^*_{x,t} - \hat{\kappa}_t \hat{\beta}_x$. Similarly, an estimator of the variance of $\hat{\beta}_x$ is given by

$$\widehat{aVar}[\hat{\beta}_x] = \sum_{s=1}^{T} (\hat{e}^*_{s,x})^2 \hat{\kappa}_s^2 \lambda_1^{-2}. \tag{13}$$

finally, the dispersion of the fitted values around their expected value can be estimated by

$$\widehat{aVar}[\hat{m}_{x,t}] = \sum_{y=1}^{X} (\hat{\mathbf{e}}^*_t)^2 \hat{\beta}_y^2 \hat{\beta}_x^2 + \sum_{s=1}^{T} (\hat{e}^*_{s,x})^2 \hat{\kappa}_s^2 \lambda_1^{-2} \hat{\kappa}_t^2 \tag{14}$$

4 Forecasting mortality rates

Still to be written...

5 Conclusion

The present paper considers the asymptotic properties of the Lee-Carter model for modelling and forecasting mortality rates. We have used results from the closely related literature on Principal Components estimation of static factor models to establish the asymptotic distributions of the model estimators. It has been shown that the Lee-Carter model works under a quite general set of assumptions about the true data generating process and that the fitted values are consistent. However, several problems with individual estimators have been highlighted: The estimated intercepts are excessively biased and cannot be interpreted in any useful way, although relative comparisons of different intercepts in the same sample are pssible. Interpretation of the slope coefficients is possible but cumbersome, due to a multiplicator that depends on the sample size. Finally, the trend estimate does not converge and thus needs to be combined with a measure of estimation uncertainty. This latter problem carries on to forecasts of mortality rates with the Lee-Carter model, a fact that is until now disregarded in the applied literature.

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.

Davidson, J. (2000). *Econometric Theory*, Blackwell Publishing.

Hayashi, F. (2000). *Econometrics*, Princeton University Press.

Lee, R.D. and L. Carter (1992). Modeling and Forecasting U. S. Mortality. *Journal of the American Statistical Association* **87**, 659–671

A Properties of the demeaned Random Walk with Drift

It is helpful to decomposed the demeaned RWD into its deterministic and stochastic components, as done in (3):

$$\kappa_t^* = \delta t - \frac{T-1}{2} + \eta_t^* \tag{15}$$

where $\eta_t^* = \eta_t - T^{-1}\sum_{s=1}^{T}\eta_s$ and $\eta_t = \eta_{t-1} + \nu_t$. It is obvious that the first part of this sum is of order $O(T)$. For the second part, note that

$$\begin{aligned}
\eta_t^* &= \sum_{s=1}^{t}\nu_s - T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{t}\nu_s \\
&\xrightarrow{d} T^{1/2}\sigma\left(W(r) - \int_0^1 W(r)dr\right) \\
&= O_p(T^{1/2})
\end{aligned} \tag{16}$$

where the second last step is due to Hamilton (1994, equations 17.3.7 and 17.3.16). Concerning Hamilton, 1994, equation 17.3.16, note that

$$T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{t}\nu_s = T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{t}\nu_{s-1} + T^{-1}\sum_{t=1}^{T}\nu_t, \tag{17}$$

where $\nu_0 = 0$. The second term on the RHS converges to $T^{-1/2}\sigma W(1)$ which goes to zero at rate $T^{-1/2}$. we can hence conclude that the entire expression at the beginnin of this paragraph is $O_p(T)$.

Taking the sum over κ_t^* changes the asymptotic order. For the deterministic part, we obtain

$$\sum_{t=1}^{T}(\delta t - \frac{T-1}{2}) = \frac{1}{2}(T^2 - T)(\delta - 1) = O(T^2) \tag{18}$$

whereas the stochastic part is

$$\sum_{t=1}^{T}\eta_t^* = \sum_{t=1}^{T}\sum_{s=1}^{t}\nu_s - T\,T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{t}\nu_s = 0. \tag{19}$$

The cumulative sum of a RWD is hence of order $O(T^2)$. Additionally

$$T^{-2}\sum_{t=1}^{T}\kappa_t^* \to \frac{\delta-1}{2}. \tag{20}$$

Furthermore,

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

where

$$I=\left(\frac{1}{12}T^3+\frac{11}{12}T\right)\delta^2=O(T^3) \tag{21}$$

by simple computation,

$$II \xrightarrow{d} T\sigma_\nu^2\int_0^1\left(W(r)-\int_0^1 W(s)ds\right)^2 dr = O_p(T) \tag{22}$$

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

$$\begin{aligned}III &\leq \left(\sum_{t=1}^{T}\left(t-\frac{T-1}{2}\right)^2\delta^2\right)^{1/2}\left(\sum_{t=1}^{T}\eta_t^2\right)^{1/2} \\ &= O_p(T^2),\end{aligned} \tag{23}$$

which follows from the first two results. We can hence conclude that $\sum_{t=1}^{T}(\kappa_t^*)^2$ is $O_p(T^3)$. Moreover,

$$T^{-3}\sum_{t=1}^{T}(\kappa_t^*)^2=\frac{\delta^2}{12}+O_p(T^{-1}). \tag{24}$$

B Notation and important equalities

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

$$\tilde{\boldsymbol{\kappa}}=\mathbf{u}_1. \tag{25}$$

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

$$\mathbf{M}^*\mathbf{M}^{*\prime}\tilde{\boldsymbol{\kappa}}=\tilde{\boldsymbol{\kappa}}\lambda_1 \tag{26}$$

can be used to obtain

$$\tilde{\boldsymbol{\kappa}} = \mathbf{M}^*\mathbf{M}^{*\prime}\tilde{\boldsymbol{\kappa}}\lambda_1^{-1} \tag{27}$$

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

$$\tilde{\boldsymbol{\kappa}} - \boldsymbol{\kappa}^* h = (\boldsymbol{\kappa}^*\boldsymbol{\beta}'\mathbf{E}^{*\prime}\tilde{\boldsymbol{\kappa}} + \mathbf{E}^*\boldsymbol{\beta}\boldsymbol{\kappa}^{*\prime}\tilde{\boldsymbol{\kappa}} + \mathbf{E}^*\mathbf{E}^{*\prime}\tilde{\boldsymbol{\kappa}})\lambda_1^{-1}, \tag{28}$$

and for each individual time period,

$$\begin{aligned}\tilde{\kappa}_t - \kappa_t^* h &= X^{-1}T^{-3}(\kappa_t^*\boldsymbol{\beta}\mathbf{E}^{*\prime}\tilde{\boldsymbol{\kappa}} + \mathbf{e}_t^{*\prime}\boldsymbol{\beta}'\boldsymbol{\kappa}^{*\prime}\tilde{\boldsymbol{\kappa}} + \mathbf{e}_t^{*\prime}\mathbf{E}^{*\prime}\tilde{\boldsymbol{\kappa}})XT^3\lambda_1^{-1} \\ &= (j_{1,t} + j_{2,t} + j_{3,t} + j_{4,t})XT^3\lambda_1^{-1},\end{aligned} \tag{29}$$

where

$$j_{1,t} = X^{-1}T^{-3}\sum_{s=1}^{T}\kappa_t^*\boldsymbol{\beta}'\mathbf{e}_s^*\tilde{\kappa}_s \tag{30}$$

$$j_{2,t} = X^{-1}T^{-3}\sum_{s=1}^{T}\mathbf{e}_t^{*\prime}\boldsymbol{\beta}\boldsymbol{\kappa}_s^*\tilde{\kappa}_s \tag{31}$$

$$j_{3,t} = T^{-3}\sum_{s=1}^{T}E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^*]\tilde{\kappa}_s \tag{32}$$

and

$$j_{4,t} = T^{-3}\sum_{s=1}^{T}(\mathbf{e}_t^{*\prime}\mathbf{e}_s^*/X - E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^*])\tilde{\kappa}_s) \tag{33}$$

C Consistency of factor estimates

We begin with two Lemmas that are used intensely in the following proofs.

Lemma C.1

$$X^{-1}T^{-3}\lambda_1 = \sigma_\beta^2\delta^2/12 + O_p(T^{-1}) \tag{34}$$

Proof of Lemma C.1 From (35) we obtain

$$X^{-1}T^{-3}\tilde{\boldsymbol{\kappa}}\mathbf{M}^*\mathbf{M}^{*\prime}\tilde{\boldsymbol{\kappa}} = X^{-1}T^{-3}\lambda_1 \tag{35}$$

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

$$\begin{aligned}&X^{-1}T^{-3}(\tilde{\boldsymbol{\kappa}}'\mathbf{M}^*\mathbf{M}^{*\prime}\tilde{\boldsymbol{\kappa}} - \tilde{\boldsymbol{\kappa}}'\boldsymbol{\kappa}^*\boldsymbol{\beta}'\boldsymbol{\beta}\boldsymbol{\kappa}^{*\prime}\tilde{\boldsymbol{\kappa}})\\ &= X^{-1}T^{-3}\tilde{\boldsymbol{\kappa}}'\boldsymbol{\kappa}^*\boldsymbol{\beta}'\mathbf{E}'\tilde{\boldsymbol{\kappa}} + X^{-1}T^{-3}\tilde{\boldsymbol{\kappa}}'\mathbf{E}\boldsymbol{\beta}\boldsymbol{\kappa}^{*\prime}\tilde{\boldsymbol{\kappa}} + X^{-1}T^{-3}\tilde{\boldsymbol{\kappa}}'\mathbf{E}\mathbf{E}'\tilde{\boldsymbol{\kappa}}\end{aligned} \tag{36}$$

Consider the last term on the RHS above.

$$\begin{aligned}X^{-1}T^{-3}\tilde{\boldsymbol{\kappa}}'\mathbf{E}\mathbf{E}'\tilde{\boldsymbol{\kappa}} &\leq X^{-1}T^{-3}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}(\mathbf{e}_t\mathbf{e}_s)^2\right)^{1/2}\\ &= O_p(T^{-1})\end{aligned} \tag{37}$$

The remaining two expressions in (36) are identical and of order

$$\begin{aligned}X^{-1}T^{-3}\tilde{\boldsymbol{\kappa}}'\mathbf{E}\boldsymbol{\beta}\boldsymbol{\kappa}^{*\prime}\tilde{\boldsymbol{\kappa}} &\leq X^{-1}T^{-3}\left(\left(\sum_{t=1}^{T}\tilde{\kappa}_t^2\right)\left(\sum_{t=1}^{T}(\mathbf{e}_t'\boldsymbol{\beta}\boldsymbol{\kappa}^{*\prime}\tilde{\boldsymbol{\kappa}})^2\right)\right)\\ &\leq X^{-1}T^{-3}\left(\left(\sum_{t=1}^{T}\tilde{\kappa}_t^2\right)\left(\sum_{t=1}^{T}\left(\sum_{s=1}^{T}(\mathbf{e}_t'\boldsymbol{\beta}\kappa_s^{*\prime})^2\right)\left(\sum_{s=1}^{T}\tilde{\kappa}_s^2\right)\right)\right)^{1/2}\\ &= X^{-1/2}T^{-1}\left(T^{-1}\sum_{t=1}^{T}(X^{-1/2}\mathbf{e}_t'\boldsymbol{\beta})^2T^{-3}\sum_{s=1}^{T}(\kappa_s^{*\prime})^2\right)^{1/2}\\ &= O_p(X^{-1/2}T^{-1}).\end{aligned} \tag{38}$$

These result show that asymptotically λ_1 is the largest eigenvalue of $\boldsymbol{\kappa}^*\boldsymbol{\beta}'\boldsymbol{\beta}\boldsymbol{\kappa}^{*\prime}$. Note however that the largest eigenvalues of $\boldsymbol{\kappa}^*\boldsymbol{\beta}'\boldsymbol{\beta}\boldsymbol{\kappa}^{*\prime}$ and $\boldsymbol{\kappa}^{*\prime}\boldsymbol{\kappa}^*\boldsymbol{\beta}'\boldsymbol{\beta}$ are identical. Using this equality as well as Assumption 2 and the results established in appendix A we obtain the required result. ■

Lemma C.2 *Under assumptions* `yet to be written`*,*

$$\frac{1}{T}\sum_{t=1}^{T}(\tilde{\boldsymbol{\kappa}} - \boldsymbol{\kappa}^*h)^2 = O_p(X^{-1}T^{-3}) + O_p(T^{-6}) \tag{39}$$

Proof of Lemma C.2 Considering (29), we can write

$$\begin{aligned}T^{-1}\sum_{t=1}^{T}(\tilde{\kappa}_t - \kappa_t h)^2 &= T^{-1}\sum_{t=1}^{T}(j_{1,t} + j_{2,t} + j_{3,t} + j_{4,t})^2 X^2 T^6 \lambda_1^{-2}\\ &\leq 4X^2T^6\lambda_1^{-2}T^{-1}\sum_{t=1}^{T}(j_{1,t}^2 + j_{2,t}^2 + j_{3,t}^2 + j_{4,t}^2),\end{aligned} \tag{40}$$

whose individual components are

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T} j_{1,t}^{2} &= T^{-1}\sum_{t=1}^{T}\left(X^{-1}T^{-3}\kappa_t^{*}\sum_{s=1}^{T}\boldsymbol{\beta}'\mathbf{e}_s^{*}\tilde{\kappa}_s\right)^{2} \\
&\leq X^{-1}T^{-3}\left(T^{-3}\sum_{t=1}^{T}(\kappa_t^{*})^{2}\right)\left(T^{-1}\sum_{s=1}^{T}(X^{-1/2}\boldsymbol{\beta}'\mathbf{e}_s^{*})^{2}\right)\left(\sum_{s=1}^{T}\tilde{\kappa}_s^{2}\right) \\
&= O_p(X^{-1}T^{-3}), \qquad (41)
\end{aligned}
$$

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T} j_{2,t}^{2} &= T^{-1}\sum_{t=1}^{T}(\mathbf{e}_t^{*\prime}\boldsymbol{\beta})^{2}\left(X^{-1}T^{-3}\sum_{s=1}^{T}\kappa_s^{*}\tilde{\kappa}_s\right)^{2} \\
&\leq X^{-1}T^{-3}\left(T^{-1}\sum_{t=1}^{T}(X^{-1/2}\mathbf{e}_t^{*\prime}\boldsymbol{\beta})^{2}\right)\left(T^{-3}\sum_{s=1}^{T}(\kappa_s^{*})^{2}\right)\left(\sum_{s=1}^{T}\tilde{\kappa}_s^{2}\right) \\
&= O_p(X^{-1}T^{-3}). \qquad (42)
\end{aligned}
$$

and

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T} j_{3,t}^{2} &= T^{-1}\sum_{t=1}^{T}\left(T^{-3}\sum_{s=1}^{T}E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^{*}]\tilde{\kappa}_s\right)^{2} \\
&\leq T^{-6}\left(T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_{t,s}^{2}\right)\left(\sum_{s=1}^{T}\tilde{\kappa}_s^{2}\right) \\
&= O_p(T^{-6}), \qquad (43)
\end{aligned}
$$

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 $\tilde{\gamma}(s,t) = \mathbf{e}_t^{*\prime}\mathbf{e}_s^{*}/X - E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^{*}]$. Following the proof of Theorem 1 in Bai and Ng (2002) we obtain

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T} j_{4,t}^{2} &= T^{-7}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\tilde{\gamma}(s,t)\tilde{\kappa}_s\right)^{2} \\
&= T^{-7}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{u=1}^{T}\tilde{\gamma}(s,t)\tilde{\gamma}(u,t)\tilde{\kappa}_s\tilde{\kappa}_u \\
&\leq T^{-7}\left(\sum_{s=1}^{T}\sum_{u=1}^{T}\left[\sum_{t=1}^{T}\tilde{\gamma}(s,t)\tilde{\gamma}(u,t)\right]^{2}\right)^{1/2}\left(\sum_{s=1}^{T}\sum_{u=1}^{T}[\tilde{\kappa}_s\tilde{\kappa}_u]^{2}\right)^{1/2} \\
&\leq T^{-7}\left(\sum_{s=1}^{T}\sum_{u=1}^{T}\left[\sum_{t=1}^{T}\tilde{\gamma}(s,t)\tilde{\gamma}(u,t)\right]^{2}\right)^{1/2}\sum_{s=1}^{T}\tilde{\kappa}_s^{2}. \qquad (44)
\end{aligned}
$$

Via the inequality

$$E\left[\left(\sum_{t=1}^{T}\tilde{\gamma}(s,t)\tilde{\gamma}(u,t)\right)^2\right] = E\left[\sum_{t=1}^{T}\tilde{\gamma}(s,t)\tilde{\gamma}(u,t)\tilde{\gamma}(s,v)\tilde{\gamma}(u,v)\right]$$
$$\leq T^2 \max_{s,t} E\left[|\tilde{\gamma}(s,t)|^4\right] \tag{45}$$

and the fact that $E\left[|\tilde{\gamma}(s,t)|^4\right] \leq X^{-2}K$ by assumption 3(iii), equation (44) can be rewritten

$$T^{-1}\sum_{t=1}^{T} j_{4,t}^2 \leq T^{-7}\left(\sum_{s=1}^{T}\sum_{u=1}^{T} O_p(X^{-2}T^2)\right)^{1/2}$$
$$= O_p(X^{-1}T^{-5}). \tag{46}$$

Collecting the results in (41)–(46) and noting that $X^{-1}T^{-3}\lambda_1 = O_p(1)$, as shown in Lemma C.1, we obtain the required result ■

Lemma C.3 *Under the assumptions of Lemma C.4,*

$$h = T^{-3/2}\frac{\sqrt{12}}{\delta} + O_p(X^{-1/2}T^{-5/2}) + O_p(T^{-4}) \tag{47}$$

Proof of Lemma C.3 We can use the decomposition

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

so that

$$h(\boldsymbol{\kappa}^*)'\tilde{\boldsymbol{\kappa}} = (h\boldsymbol{\kappa}^*)'(\tilde{\boldsymbol{\kappa}} - h\boldsymbol{\kappa}^*) + h^2(\boldsymbol{\kappa}^*)'\boldsymbol{\kappa}^* \tag{48}$$

Now note that the LHS above can also be decomposed

$$h(\boldsymbol{\kappa}^*)'\tilde{\boldsymbol{\kappa}} = (h\boldsymbol{\kappa}^* - \tilde{\boldsymbol{\kappa}})'\tilde{\boldsymbol{\kappa}} + \tilde{\boldsymbol{\kappa}}'\tilde{\boldsymbol{\kappa}}$$
$$= 1 + (h\boldsymbol{\kappa}^* - \tilde{\boldsymbol{\kappa}})'(\tilde{\boldsymbol{\kappa}} - \boldsymbol{\kappa}^*h) + (h\boldsymbol{\kappa}^* - \tilde{\boldsymbol{\kappa}})'\boldsymbol{\kappa}^*h \tag{49}$$

Setting equal (48) and (49), we obtain

$$h^2 = \left((\boldsymbol{\kappa}^*)'\boldsymbol{\kappa}^*\right)^{-1}\left(1 + \sum_{t=1}^{T}(\tilde{\kappa}_t - \kappa_t^* h)^2\right) \tag{50}$$

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

$$h^2 = T^{-3}\frac{12}{\delta^2}\left(1 + O_p(X^{-1/2}T^{-2}) + O_p(T^{-5})\right). \tag{51}$$

which implies the required result. ■

Lemma C.4 *Under Assumptions 1-5,*

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

Proof of Lemma C.4. We can rewrite the expression as

$$T^{-3/2}\sum_{t=1}^{T}(\tilde{\kappa}_t - h\kappa_t^{*\prime})\kappa_t^* = \left(XT^3\lambda_1^{-1}\right)T^{-3/2}\sum_{t=1}^{T}(j_{1,t} + j_{2,t} + j_{3,t} + j_{4,t})\kappa_t^* \tag{52}$$

where

$$\begin{aligned} T^{-3/2}\sum_{t=1}^{T} j_{1,t}\kappa_t^* &= X^{-1}T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\kappa_t^*\boldsymbol{\beta}'\mathbf{e}_s^*\tilde{\kappa}_s\kappa_t^* \\ &= X^{-1}T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\kappa_t^*\boldsymbol{\beta}'\mathbf{e}_s^*\kappa_s^* h\kappa_t^* + X^{-1}T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\kappa_t^*\boldsymbol{\beta}'\mathbf{e}_s^*(\tilde{\kappa}_s - \kappa_s^* h)\kappa_t^*. \end{aligned} \tag{53}$$

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

$$\begin{aligned} X^{-1}T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\kappa_t^*\boldsymbol{\beta}'\mathbf{e}_s^*\kappa_s^* h\kappa_t^* &= X^{-1/2}\left(T^{-3}\sum_{s=1}^{T}(\kappa_t^*)^2\right)\left(X^{-1/2}T^{-3/2}\sum_{s=1}^{T}\boldsymbol{\beta}'\mathbf{e}_s^*\kappa_s^*\right)h \\ &= O_p(X^{-1/2}T^{-3/2}). \end{aligned} \tag{54}$$

The second part is given by

$$\begin{aligned} &X^{-1}T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\kappa_t^*\boldsymbol{\beta}'\mathbf{e}_s^*(\tilde{\kappa}_s - \kappa_s^* h)\kappa_t^* \\ &\leq X^{-1/2}T^{-1/2}\left(T^{-3}\sum_{t=1}^{T}(\kappa_t^*)^2\right)\left(T^{-1}\sum_{s=1}^{T}(X^{-1/2}\boldsymbol{\beta}'\mathbf{e}_s^*)^2\right)^{1/2}\left(T^{-1}\sum_{s=1}^{T}(\tilde{\kappa}_s - \kappa_s^* h)^2\right)^{1/2} \\ &= O_p(X^{-1}T^{-2}) + O_p(X^{-1/2}T^{-7/2}). \end{aligned} \tag{55}$$

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

$$T^{-3/2}\sum_{t=1}^{T} j_{2,t}\kappa_t^* = X^{-1}T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{e}_t^{*\prime}\boldsymbol{\beta}\kappa_s^*\tilde{\kappa}_s\kappa_t^*$$
$$= X^{-1}T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{e}_t^{*\prime}\boldsymbol{\beta}\kappa_s^* h\kappa_s^*\kappa_t^* + X^{-1}T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{e}_t^{*\prime}\boldsymbol{\beta}\kappa_s^*(\tilde{\kappa}_s - \kappa_s^* h)\kappa_t^* \tag{56}$$

The first part of the summation above is

$$X^{-1}T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{e}_t^{*\prime}\boldsymbol{\beta}\kappa_s^* h\kappa_s^*\kappa_t^* = X^{-1/2}\left(T^{-3}\sum_{s=1}^{T}(\kappa_s^*)^2\right)\left(X^{-1/2}T^{-3/2}\sum_{t=1}^{T}\mathbf{e}_t^{*\prime}\boldsymbol{\beta}\kappa_t^*\right)h$$
$$= O_p(X^{-1/2}T^{-3/2}) \tag{57}$$

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

$$X^{-1}T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{e}_t^{*\prime}\boldsymbol{\beta}\kappa_s^*(\tilde{\kappa}_s - \kappa_s^* h)\kappa_t^*$$
$$\leq X^{-1/2}T^{-1}\left(T^{-1}\sum_{s=1}^{T}(\tilde{\kappa}_s - \kappa_s^* h)^2\right)^{1/2}\left(T^{-3}\sum_{s=1}^{T}(\kappa_s^*)^2\left(X^{-1/2}T^{-3/2}\sum_{t=1}^{T}\mathbf{e}_t^{*\prime}\boldsymbol{\beta}\kappa_t^*\right)^2\right)^{1/2}$$
$$= O_p(X^{-1}T^{-5/2}) + O_p(X^{-1/2}T^{-4}) \tag{58}$$

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

$$T^{-9/2}\sum_{t=1}^{T} j_{3,t}\kappa_t^* = T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^{*\prime}]\tilde{\kappa}_s\kappa_t^*$$
$$= T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^{*\prime}]\kappa_s^* h\kappa_t^* + T^{-9/2}\sum_{t=1}^{T}\sum_{t=1}^{T}E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^{*\prime}](\tilde{\kappa}_s - \kappa_s^* h)\kappa_t^* \tag{59}$$

The second term above can be written

$$T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^{*\prime}](\tilde{\kappa}_s - \kappa_s^* h)\kappa_t^*$$
$$\leq T^{-2}\left(T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_{ts}^2\right)^{1/2}\left(T^{-1}\sum_{s=1}^{T}(\tilde{\kappa}_s - \kappa_s^* h)^2 T^{-3}\sum_{t=1}^{T}(\kappa_t^*)^2\right)^{1/2}$$
$$= O_p(X^{-1/2}T^{-7/2}) + O_p(T^{-5}), \tag{60}$$

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).

$$E\left[T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^*]\kappa_s^* h\kappa_t^*\right] = T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^*]E[\kappa_s^*\kappa_t^*]h$$
$$\leq T^{-7/2}T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\tau_{st}T^2\delta^2 h$$
$$= O(T^{-3}) \tag{61}$$

Finally, letting $\tilde{\gamma}(s,t) = \mathbf{e}_t^{*\prime}\mathbf{e}_s^*/X - E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^*]$,

$$T^{-9/2}\sum_{t=1}^{T} j_{4,t}\kappa_t^* = T^{-6}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{\gamma}(s,t)\tilde{\kappa}_s\kappa_t^*$$
$$= T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{\gamma}(s,t)\kappa_s^* h\kappa_t^* + T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{\gamma}(s,t)(\tilde{\kappa}_s - \kappa_s^* h)\kappa_t^*, \tag{62}$$

whose second part is

$$T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{\gamma}(s,t)(\tilde{\kappa}_s - \kappa_s^* h)\kappa_t^*$$
$$\leq X^{-1/2}T^{-3/2}\left(T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T}X^{1/2}\tilde{\gamma}(s,t)^2\right)^{1/2}\left(T^{-1}\sum_{s=1}^{T}(\tilde{\kappa}_s - \kappa_s^* h)^2 T^{-3}\sum_{t=1}^{T}(\kappa_t^*)^2\right)^{1/2}$$
$$= O_p(X^{-1}T^{-3}) + O_p(X^{-1/2}T^{-9/2}). \tag{63}$$

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

$$T^{-9/2}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{\gamma}(s,t)\kappa_s^* h\kappa_t^*$$
$$\leq X^{-1/2}T^{-1}\left(T^{-1}\sum_{t=1}^{T}\left(X^{1/2}T^{-3/2}\sum_{s=1}^{T}\tilde{\gamma}(s,t)\kappa_s^*\right)^2\right)^{1/2}\left(T^{-3}\sum_{t=1}^{T}(\kappa_t^*)^2\right)^{1/2} h$$
$$= O_p(X^{-1/2}T^{-5/2}). \tag{64}$$

Taking together all results, we obtain

$$T^{-3/2}(\tilde{\boldsymbol{\kappa}}' - h\boldsymbol{\kappa}^{*\prime})\boldsymbol{\kappa}^* = O_p(X^{-1/2}T^{-3/2}) + O_p(T^{-3}) \tag{65}$$

■

Lemma C.5 *Under the assumptions of Lemma C.4,*

$$(\tilde{\boldsymbol{\kappa}}' - h\boldsymbol{\kappa}^{*\prime})\tilde{\boldsymbol{\kappa}}/T^{3/2} = O_p(X^{-1/2}T^{-3/2}) + O_p(T^{-3}) \tag{66}$$

Proof of Lemma C.5 Note that

$$(\tilde{\boldsymbol{\kappa}}' - h\boldsymbol{\kappa}^{*\prime})\tilde{\boldsymbol{\kappa}}/T^{3/2} = (\tilde{\boldsymbol{\kappa}}' - h\boldsymbol{\kappa}^{*\prime})(\tilde{\boldsymbol{\kappa}} - \boldsymbol{\kappa}^{*}h)/T^{3/2} + (\tilde{\boldsymbol{\kappa}}' - h\boldsymbol{\kappa}^{*\prime})\boldsymbol{\kappa}^{*}h/T^{3/2} \tag{67}$$

The lemma thus follows from Lemmas C.2 and C.4 . ■

Lemma C.6

$$T^{-3/2}\boldsymbol{\kappa}^{*\prime}\tilde{\boldsymbol{\kappa}} = \delta/\sqrt{12} + O_p(T^{-1}) \tag{68}$$

Proof of Lemma C.6 Use the decomposition

$$T^{-3/2}\boldsymbol{\kappa}^{*\prime}\tilde{\boldsymbol{\kappa}} = T^{-3/2}\boldsymbol{\kappa}^{*\prime}(\tilde{\boldsymbol{\kappa}} - h\boldsymbol{\kappa}^{*}) + (T^{-3}\boldsymbol{\kappa}^{*\prime}\boldsymbol{\kappa}^{*})(T^{3/2}h) \tag{69}$$

Lemma C.4 establishes the order of the first term on the RHS. For the components of the second term, (24) and Lemma (C.3) can be used to obtain the result. ■

Proof of Proposition 1 Given that $\tilde{\kappa}_t$ is an estimator of $\kappa_t^* h$, the relation $\hat{\kappa}_t = \tilde{\kappa}_t\lambda_1^{1/2}$ implies that $\hat{\kappa}_t$ estimates $\kappa_t^*\hat{h} = \kappa_t^* h\lambda_1^{1/2}$. Appliation of Lemmas C.1 and C.3 now yie lds the required result.

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

$$\hat{\alpha}_x \xrightarrow{d} \alpha_x + \lim_{T\to}\beta_x\frac{T+1}{2}\delta + \sqrt{T}\beta_x N(0,\sigma_v^2/3) + \frac{1}{\sqrt{T}}N(0,\overline{\sigma}_x^2). \tag{70}$$

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

Proof of Theorem 2 Recall from (29) that the expansion

$$\sqrt{XT^3}(\tilde{\kappa}_t - \kappa_t^* h) = \sqrt{XT^3}(j_{1,t} + j_{2,t} + j_{3,t} + j_{4,t})XT^3\lambda_1^{-1}. \tag{71}$$

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

$$\sqrt{XT^3}j_{1,t} = X^{-1/2}T^{-3/2}\sum_{s=1}^{T}\kappa_t^*\boldsymbol{\beta}'\mathbf{e}_s^*(\tilde{\kappa}_s - \kappa_s^* h) + X^{-1/2}T^{-3/2}\sum_{s=1}^{T}\kappa_t^*\boldsymbol{\beta}'\mathbf{e}_s^*\kappa_s^* h \tag{72}$$

For the first term on the RHS, we can apply Lemma C.2 to obtain

$$\begin{aligned}
&X^{-1/2}T^{-3/2}\sum_{s=1}^{T}\kappa_t^*\boldsymbol{\beta}'\mathbf{e}_s^*(\tilde{\kappa}_s-\kappa_s^*h) \\
&\leq X^{-1/2}T^{-1}\left(\sum_{s=1}^{T}(\kappa_t^*\boldsymbol{\beta}'\mathbf{e}_s^*)^2\right)^{1/2}\left(T^{-1}\sum_{s=1}^{T}(\tilde{\kappa}_s-\kappa_s^*h)^2\right)^{1/2} \\
&\leq T\left(T^{-2}(\kappa_t^*)^2\,T^{-1}\sum_{s=1}^{T}(X^{-1/2}\boldsymbol{\beta}'\mathbf{e}_s^*)^2\right)^{1/2}\Big(O_p(X^{-1/2}T^{-3/2})+O_p(T^{-3})\Big) \\
&= O_p(X^{-1/2}T^{-1/2})+O_p(T^{-2}) \qquad (73)
\end{aligned}$$

The second term is

$$\begin{aligned}
&X^{-1/2}T^{-3/2}\sum_{s=1}^{T}\kappa_t^*\boldsymbol{\beta}'\mathbf{e}_s^*\kappa_s^*h \\
&= T^1\left(T^{-1}\kappa_t^*\right)\left(X^{-1/2}T^{-3/2}\sum_{s=1}^{T}\kappa_s^*\boldsymbol{\beta}'\mathbf{e}_s^*\right)h \qquad (74)
\end{aligned}$$

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

$$\sqrt{XT^3}j_{1,t}=O_p(T^{-1/2}) \qquad (75)$$

For the order of $j_{2,t}$ we use Lemma C.6 to obtain.

$$\begin{aligned}
\sqrt{XT^3}j_{2,t}&=X^{-1/2}T^{-3/2}\sum_{s=1}^{T}\mathbf{e}_t^{*\prime}\boldsymbol{\beta}\kappa_s^*\tilde{\kappa}_s \\
&=X^{-1/2}\sum_{s=1}^{T}\mathbf{e}_t^{*\prime}\boldsymbol{\beta}\frac{\delta}{\sqrt{12}}+O_p(T^{-1}) \qquad (76)
\end{aligned}$$

Concerning $j_{3,t}$,

$$\sqrt{XT^3}j_{3,t}=\sqrt{X}T^{-3/2}\sum_{s=1}^{T}E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^*](\tilde{\kappa}_s-\kappa_s^*h)+\sqrt{X}T^{-3/2}\sum_{s=1}^{T}E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^*]\kappa_s^*h, \qquad (77)$$

where assumption 5 can be used to yield

$$\begin{aligned}
&\sqrt{X}T^{-3/2}\sum_{s=1}^{T}E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^*](\tilde{\kappa}_s-\kappa_s^*h)\\
&\leq \sqrt{X}T^{-1}\left(\sum_{s=1}^{T}|\tau_{ts}|^2\right)^{1/2}\left(T^{-1}\sum_{s=1}^{T}(\tilde{\kappa}_s-\kappa_s^*h)^2\right)^{1/2}\\
&=\sqrt{X}T^{-1}O_p(1)\left(O_p(X^{-1/2}T^{-3/2})+O_p(T^{-3})\right)=O_p(T^{-5/2})+O_p(\sqrt{X}T^{-4}). \qquad (78)
\end{aligned}$$

Furthermore, note that $E[|\sum_{s=1}^{T}E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^*]\kappa_s^*|]\leq \max_s E[|\kappa_s^*|]\sum_{s=1}^{T}|\tau_{ts}|\leq TK$ since the properties of the latent trend imply $E[||\kappa_t^*||]=O(T)$. Consequently,

$$\sqrt{X}T^{-3/2}\sum_{s=1}^{T}E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^*]\kappa_s^*h=X^{-1}T^{-3/2}O_p(T)O_p(T^{-3/2})=O_p(\sqrt{X}T^{-2}). \qquad (79)$$

We can hence conclude that

$$\sqrt{XT^3}j_{3,t}=\sqrt{X}T^{-3/2}\sum_{s=1}^{T}E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^*]\kappa_s^*h+O_p(T^{-5/2})+O_p(\sqrt{X}T^{-4})=O_p(\sqrt{X}T^{-2}). \qquad (80)$$

Finally,

$$\begin{aligned}
\sqrt{XT^3}j_{4,t}&=\sqrt{X}T^{-3/2}\sum_{s=1}^{T}(\mathbf{e}_t^{*\prime}\mathbf{e}_s^*/X-E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^*])(\tilde{\kappa}_s-\kappa_s^*h)\\
&\quad+\sqrt{X}T^{-3/2}\sum_{s=1}^{T}(\mathbf{e}_t^{*\prime}\mathbf{e}_s^*/X-E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^*])\kappa_s^*h. \qquad (81)
\end{aligned}$$

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

$$\begin{aligned}
&\sqrt{X}T^{-3/2}\sum_{s=1}^{T}(\mathbf{e}_t^{*\prime}\mathbf{e}_s^*/X-E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^*])(\tilde{\kappa}_s-\kappa_s^*h)\\
&\leq T^{-1/2}\left(T^{-1}\sum_{s=1}^{T}\left(X^{-1/2}\sum_{x=1}^{X}(e_{x,t}^*e_{x,s}^*-E[e_{x,t}^*e_{x,s}^*])\right)^2\right)^{1/2}\left(T^{-1}\sum_{s=1}^{T}(\tilde{\kappa}_s-\kappa_s^*h)^2\right)^{1/2}\\
&=T^{-1/2}\left(O_p(X^{-1/2}T^{-3/2})+O_p(T^{-3})\right)\\
&=O_p(X^{-1/2}T^{-2})+O_p(T^{-7/2}) \qquad (82)
\end{aligned}$$

and

$$
\begin{aligned}
&\sqrt{X}T^{-3/2}\sum_{s=1}^{T}(\mathbf{e}_t^{*\prime}\mathbf{e}_s^*/X - E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^*])\kappa_s^* h \\
&\leq T^{1/2}\left(T^{-1}\sum_{s=1}^{T}(\mathbf{e}_t^{*\prime}\mathbf{e}_s^*/X^{1/2} - E[\mathbf{e}_t^{*\prime}\mathbf{e}_s^*])^2\right)^{1/2}\left(T^{-3}\sum_{s=1}^{T}(\kappa_s^*)^2\right)^{1/2} h \\
&= O_p(T^{-1}).
\end{aligned} \tag{83}
$$

Collecting the order results of all components of (71), we obtain

$$
\sqrt{XT}(\tilde{\kappa}_t - \kappa_t^* h) = \left(X^{-1/2}\mathbf{e}_t^{*\prime}\boldsymbol{\beta}\right)\frac{\delta}{\sqrt{12}}XT^3\lambda_1^{-1} + O_p(T^{-1/2}) + O_p(\sqrt{X}T^{-2}) \quad \xrightarrow{d} \sqrt{12}\delta\sigma_\beta N(0,\Gamma_t), \tag{84}
$$

where the last result is obtained from using Assumption 2 and Lemmas C.1 and C.6. Premultiplication of the above with $(X^{-1}T^{-3}\lambda_1)^{1/2}$ yields the equivalent result for $\hat{\boldsymbol{\kappa}}$, given the asymptotic results on λ_1 in Lemma C.1. ■

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

$$
\begin{aligned}
\tilde{\beta}_x &= \beta_x h^{-1} h\boldsymbol{\kappa}^{*\prime}\tilde{\boldsymbol{\kappa}} + \mathbf{e}_x^{*\prime}\tilde{\kappa} \\
&= \beta_x h^{-1} + \beta_x h^{-1}(h\boldsymbol{\kappa}^{*\prime} - \tilde{\boldsymbol{\kappa}}')\tilde{\boldsymbol{\kappa}} + \mathbf{e}_x^{*\prime}\boldsymbol{\kappa}^* h + \mathbf{e}_x^{*\prime}(\tilde{\boldsymbol{\kappa}} - \boldsymbol{\kappa}^* h)
\end{aligned} \tag{85}
$$

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 C.5. Furthermore

$$
\begin{aligned}
\mathbf{e}_x^{*\prime}(\tilde{\boldsymbol{\kappa}} - \boldsymbol{\kappa}^* h) &= \sum_{t=1}^{T} e_{x,t}^*(\tilde{\kappa}_t - \kappa_t^* h) \\
&= T\left(T^{-1}\sum_{t=1}^{T}(e_{x,t}^*)^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}(\tilde{\kappa}_t - \kappa_t^* h)^2\right)^{1/2} \\
&= O_p(X^{-1/2}T^{-1/2}) + O_p(T^{-2})
\end{aligned} \tag{86}
$$

and

$$
\begin{aligned}
\mathbf{e}_x^{*\prime}\boldsymbol{\kappa}^* h &= \left(T^{-3/2}\sum_{t=1}^{T} e_{x,t}^*\kappa_t^*\right)\left(T^{3/2}h\right) \\
&= \frac{\sqrt{12}}{\delta} \sim N(0,\Phi_x) + O_p(X^{-1/2}T^{-1}) + O_p(T^{-5/2})
\end{aligned} \tag{87}
$$

as implied by Assumption 3(v) and Lemma C.3.

Using these order results, we can hence write

$$\begin{aligned}\tilde{\beta}_x - \beta_x h^{-1} &= \sum_{t=1}^{T} e^*_{x,t}\kappa^*_t h + O_p(X^{-1/2}) + O_p(T^{-3/2}) \\ &\sim \frac{\sqrt{12}}{\delta} N(0,\Phi_x) + O_p(X^{-1/2}) + O_p(T^{-1/2}) \end{aligned} \tag{88}$$

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

$$\sqrt{XT^3}(\hat{\beta}_x - \beta_x(h\sqrt{\lambda_1})^{-1}) \sim \frac{12}{\delta^2\sigma_\beta} N(0,\Phi_x) + O_p(X^{-1/2}) + O_p(T^{-1/2}) \tag{89}$$

■

Proof of Theorem 4 Given that $\hat{\boldsymbol{\kappa}}\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\kappa}}\tilde{\boldsymbol{\beta}}$, the properties of the common component in the Lee-Carter model can be analysed using the alternative estimators $\tilde{\boldsymbol{\kappa}}$ and $\tilde{\boldsymbol{\beta}}$ The deviations of the model fit from the true mortality rates can be written

$$\hat{m}^*_{x,t} - \kappa^*_t\beta_x = (\tilde{\kappa}_t - \kappa^*_t h)h^{-1}\beta_x + \tilde{\kappa}_t(\tilde{\beta}_x - \beta_x h^{-1}). \tag{90}$$

Using Theorem 2 the first part on the RHS can be written

$$\begin{aligned}(\tilde{\kappa}_t - \kappa^*_t h)h^{-1}\beta_x &= X^{-1/2}T^{-3/2}\left(X^{-1/2}\mathbf{e}^{*\prime}_t\boldsymbol{\beta}\right)\left(T^{-3}\frac{\delta}{\sqrt{12}} + O_p(T^{-1/2}) + O_p(\sqrt{X}T^{-2})\right)h^{-1}\beta_x \\ &= X^{-1/2}\left(X^{-1/2}\mathbf{e}^{*\prime}_t\boldsymbol{\beta}\right)\frac{\delta}{\sqrt{12}}(T^{3/2}h)^{-1}\beta_x, \end{aligned} \tag{91}$$

implying that

$$\sqrt{X}(\tilde{\kappa}_t - \kappa^*_t h)h^{-1}\beta_x \sim \frac{\delta^2}{12}\beta_x N(0,\Gamma_t) + O_p(T^{-1/2}) + O_p(\sqrt{X}T^{-2}). \tag{92}$$

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

$$\begin{aligned}\tilde{\kappa}_t(\tilde{\beta}_x - \beta_x h^{-1}) &= \kappa^*_t h(\hat{\beta}_x - \beta_x h^{-1}) + (\tilde{\kappa}_t - \kappa^*_t h)(\tilde{\beta}_x - \beta_x h^{-1}) \\ &= T^{-1/2}(T^{-1}\kappa^*_t)\left(T^{-3/2}(\mathbf{e}^*_x)'\boldsymbol{\kappa}^*\right)T^3h^2 + O_p(X^{-1/2}T^{-3/2}) + O_p(T^{-3}). \end{aligned} \tag{93}$$

From this result, we obtain

$$
\begin{aligned}
T^{1/2}\tilde{\kappa}_t(\tilde{\beta}_x - \beta_x h^{-1}) &= \delta\left(\frac{t}{T} - \frac{1}{2} + O_p(T^{-1/2})\right)\left(T^{-3/2}(\mathbf{e}_x^*)'\boldsymbol{\kappa}^*\frac{12}{\delta^2} + O_p(X^{-1/2} + O_p(T^{-1/2}))\right) \\
&\sim \frac{12}{\delta}\left(\frac{t}{T} - \frac{1}{2}\right)N(0,\Phi_x) \qquad (94)
\end{aligned}
$$

Collecting the above results and letting

$$
v_1 = \left(\frac{\delta^2}{12}\beta_x\right)^2\Gamma_t \qquad (95)
$$

and

$$
v_2 = \left(\frac{12}{\delta}\left(\frac{t}{T} - \frac{1}{2}\right)\right)^2\Phi_x \qquad (96)
$$

we can state that

$$
(X^{-1}v_1 + T^{-1}v_2)^{-1/2}(\hat{m}_{x,t} - \kappa_t^*\beta_x) \xrightarrow{d} N(0,1) \qquad (97)
$$

■.