Statistical Inference For Lee-Carter Mortality Model And Corresponding Forecasts

QING LIU\(^1\), CHEN LING\(^2\) AND LIANG PENG\(^2\)

Abstract. Although the Lee-Carter model has become a benchmark in modeling mortality rates and forecasting mortality risk, there exist some serious issues on its inference and interpretation in the literature of actuarial science. After pointing out these pitfalls and misunderstandings, this paper proposes a modified Lee-Carter model, provides a sound statistical inference and derives the asymptotic distributions of the proposed estimators and unit root test when the mortality index is nearly integrated and errors in the model satisfy some mixing conditions. After a unit root hypothesis is not rejected, future mortality forecasts can be obtained via the proposed inference. An application of the proposed unit root test to US mortality rates rejects the unit root hypothesis for the female and combined mortality rates, but does not reject the unit root hypothesis for the male mortality rates.

Key words and phrases: Asymptotic distribution, Forecast, Lee-Carter model, Mixing sequence, Mortality rates

1 Introduction

As of December 1, 2017, Google scholar gave 156,000 results after searching the key words "Lee Carter mortality model", which clearly shows the importance and wide applications of the Lee-Carter mortality model. Indeed the Lee-Carter model has become a benchmark in modeling mortality rates, forecasting mortality risk and hedging longevity risk.

Let \(m(x, t)\) denote the observed central death rate for age (or age group) \(x\) in year \(t\), where

\(^1\)School of Statistics and Research Center of Applied Statistics, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330013, P.R.China

\(^2\)Department of Risk Management and Insurance, Georgia State University, Atlanta, GA 30303, USA
\( x = 1, \cdots, M \) and \( t = 1, \cdots, T \). To model the logarithms of the central death rates, Lee and Carter (1992) proposed the following simple linear regression model

\[
\log m(x, t) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}, \quad \sum_{x=1}^{M} \beta_x = 1, \quad \sum_{t=1}^{T} k_t = 0, \tag{1}
\]

where \( \varepsilon_{x,t} \)'s are random errors with mean zero and finite variance, and the unobserved \( k_t \)'s are called mortality index. Note that the above two constraints ensure that the model is identifiable.

Since \( k_t \)'s are unobservable, the so-called singular value decomposition method is employed to estimate the unknown quantities, \( \{\alpha_x\}_{x=1}^{M}, \{\beta_x\}_{x=1}^{M} \) and \( \{k_t\}_{t=1}^{T} \), which makes the derivation of the asymptotic distributions of these estimators impossible.

As an important task of modeling mortality rates is to forecast future mortality pattern so as to better forecast mortality risk and hedge longevity risk, Lee and Carter (1992) further proposed to model the estimated mortality index by a simple time series model. As a matter of fact, Lee and Carter (1992) assumed that \( \{k_t\} \) follows from an ARIMA\((p,d,q)\) model defined as

\[
\left(1 - \sum_{i=1}^{p} \phi_i B^i\right) \left(1 - B\right)^d k_t = \mu + \left(1 + \sum_{i=1}^{q} \theta_i B^i\right) \epsilon_t, \tag{2}
\]

where \( \epsilon_t \)'s are white noises.

In conclusion, the classic Lee-Carter mortality model proposed by Lee and Carter (1992) is a combination of (1) and (2), and a proposed two-step inference procedure by them is to first estimate parameters in (1) by the singular value decomposition method and then to use the estimated \( k_t \)'s to fit model (2). Many papers in actuarial science have claimed that an application of this model and this two-step inference procedure to mortality data prefers a unit root time series model, i.e., \( d = 1 \) in (2).

Since this seminal publication, many extensions and applications have appeared in the literature of actuarial science with an open statistical R package (‘demography’), where a key step in forecasting future mortality rates is to fit a time series model to the unobserved mortality index. Some references are Brouhns, Denuit and Vermunt (2002), Li and Lee (2005), Girosi and
Although the Lee-Carter model has become a benchmark in modeling mortality rates, there are some serious issues on its model assumptions and proposed two-step inference procedure. First, since \( \{k_t\} \) in (2) is random, the constraint \( \sum_{t=1}^{T} k_t = 0 \) in (1) becomes unrealistic and restrictive. For example, if one fits an AR(1) model to \( \{k_t\} \), say \( k_t = \mu + \phi k_{t-1} + \varepsilon_t \), then we have \( T^{-1} \sum_{t=1}^{T} k_t \xrightarrow{p} \mu/(1 - \phi) \) as \( T \to \infty \) when \(|\phi| < 1\) independent of \( T \). On the other hand, we have \( k_T/T \xrightarrow{p} \mu \{ \lim_{x \to 1} \frac{1-e^x}{e^x-x} \} \) as \( T \to \infty \) when \( \mu \neq 0 \) and \( \phi = 1 + \gamma/T \) for some constant \( \gamma \in \mathbb{R} \). That is, the constraint \( \sum_{t=1}^{T} k_t = 0 \) in (1) basically says \( \mu \) in (2) must be zero. Hence, a modified model without any direct constraint on \( k_t \)'s is more appropriate. Another difficulty in using the singular value decomposition method in Lee and Carter (1992) is that no asymptotic results are available for the derived estimators. That is, there is no way to quantify the inference uncertainty. When all \( \beta_x \) are the same (i.e., \( \beta_1 = \cdots = \beta_M = 1/M \)), Leng and Peng (2016) proved that the proposed two-step inference procedure in Lee and Carter (1992) is inconsistent when the model (2) is not an exact ARIMA(0,1,0) model.

It also happens that some papers in actuarial science understand the model in a wrong way. For example, by defining \( m_0(x,t) \) as the true central death rate for age \( x \) in year \( t \), Dowd et al. (2010), Cairns et al (2011), Enchev, Kleinow and Cairns (2017) and others interpreted the Lee-Carter model as

\[
\log m_0(x,t) = \alpha_x + \beta_x k_t, \quad k_t = \mu + \varepsilon_t, \quad \sum_{x=1}^{M} \beta_x = 1, \quad \sum_{t=1}^{T} k_t = 0. \tag{3}
\]

Obviously this is quite confusing because model (3) basically says the true mortality rate \( m_0(x,t) \) is random due to the randomness of \( k_t \)'s. Another misinterpretation appears in Li (2010) and Li, Chan and Zhou (2015), where the Lee-Carter model is treated as \( \log m(x,t) = \alpha_x + \beta_x k_t \) without the random error \( \varepsilon_{x,t} \) in (1). This is quite problematic because it simply says that \( \log m(x,t) \) and \( \log m(y,t) \) are completely dependent as both are determined by the same random variable.
$k_t$. That is, central death rates are completely dependent across ages.

In summary, the random error term $\varepsilon_{x,t}$ in (1) is necessary in order to avoid the unrealistic implication that the central death rates are completely dependent across ages. Due to the presence of these random errors $\varepsilon_{x,t}$'s, the two-step inference procedure proposed by Lee and Carter (1992) may be inconsistent in the sense that the resulted estimators do not converge to the true values when $T$ goes to infinity. More specifically, Leng and Peng (2016) considered a submodel of (1) with known $\beta_x$'s (i.e., $\beta_1 = \cdots = \beta_M = \frac{1}{M}$) and showed that the two-step inference procedure is inconsistent in identifying the true dynamics of the mortality index when $k_t$’s follow an ARIMA(p,0,q) or ARIMA(p,1,q) model with $p + q > 0$, but it is consistent when $k_t$’s follow an ARIMA(0,1,0) model exactly (i.e., a unit root AR(1) model). Further Leng and Peng (2017) proposed a way to test whether $k_t$’s follow a unit root AR(p) model and showed that a blind application of the R package 'demography' leads to a different conclusion. Since Leng and Peng (2016) only considered a submodel of (1), it still remains open on whether the inference in Lee and Carter (1992) is consistent in estimating all unknown parameters and forecasting future mortality rates even when $k_t$’s do follow a unit root AR(1) process. It also remains unknown whether the bootstrap method in D’Amato et al. (2012) are consistent in quantifying the forecasting accuracy of longevity risk based on the Lee-Carter model.

This paper first modifies the classic Lee-Carter model without adding a constraint on $k_t$’s for the sake of model’s identification. Second by focusing on fitting an AR(1) model to $\{k_t\}$ and assuming that errors are $\alpha$-mixing instead of independent, this paper proposes estimators for unknown quantities and a test for unit root, and further derives the asymptotic distributions of the proposed estimators and the unit root test when the mortality index $\{k_t\}$ follows from a unit root or near unit root AR(1) process. When the unit root hypothesis can not be rejected, forecasting future mortality rates is provided. We refer to Section 2 for details. Section 3 presents a real data analysis and a simulation study. Some conclusions are summarized in Section 4. All
proofs are put in Section 5.

2 Model, Estimation, Unit Root Test and Forecast

2.1 Model

First we propose to replace (1) by

$$\log m(x, t) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}, \quad \sum_{x=1}^{M} \beta_x = 1, \quad \sum_{x=1}^{M} \alpha_x = 0,$$

(4)

where $\varepsilon_{x,t}$’s are random errors with zero mean and finite variance for each $x$. It is clear that we do not directly impose a constraint on the unobserved random mortality index $k_t$ to ensure that the proposed model is identifiable. We also remark that the assumption of $\sum_{x=1}^{M} \alpha_x = 0$ is not restrictive at all as we can simply move the sum to $k_t$ if $\sum_{x=1}^{M} \alpha_x \neq 0$.

As literature argues that real datasets often prefer a unit root AR(1) model and some applications of the Lee-Carter model simply assume a unit root AR(1) model such as Chen and Cox (2009), Chen and Cummins (2010), Kwok, Chiu and Wong (2016), Biffis, Lin and Milidonis (2017), Lin, Shi and Arik (2017), Li, De Waegenaeve and Melenberg (2017), Wong, Chiu and Wong (2017) and Zhu, Tan and Wang (2017), we consider the following special case of (2):

$$k_t = \mu + \phi k_{t-1} + \epsilon_t,$$

(5)

where $\epsilon_t$’s are white noises. Therefore the proposed modified Lee-Carter mortality model is a combination of (4) and (5), which does not impose any constraint on $k_t$’s for model’s identification.

2.2 Estimation

Next we consider a statistical inference for models (4) and (5). As we argue before, the two-step inference in Lee and Carter (1992) is hard to derive asymptotic results and may lead to
inconsistent estimators. Therefore we need a different statistical inference without using the
singular value decomposition method.

Put $\hat{Z}_t = \sum_{x=1}^{M} \log m(x, t)$ and $\eta_t = \sum_{x=1}^{M} \varepsilon_{x,t}$ for $t = 1, \cdots, T$. Then, by noting that
$\sum_{x=1}^{M} \alpha_x = 0$ and $\sum_{x=1}^{M} \beta_x = 1$, we have

$$\hat{Z}_t = k_t + \eta_t \text{ for } t = 1, \cdots, T. \quad (6)$$

When $\{k_t\}$ is nonstationary such as unit root (i.e., $\phi = 1$ in (5)) or near unit root (i.e., $\phi = 1 + \gamma/T$
for some constant $\gamma \neq 0$ in (5)), $k_t$ dominates $\eta_t$ as $t$ large enough, and so $\hat{Z}_t$ behaves like $k_t$ in
this case. This motivates us to minimize the following least squares

$$\sum_{t=2}^{T} \left( \hat{Z}_t - \mu - \phi \hat{Z}_{t-1} \right)^2,$$

and obtain the least squares estimator for $\mu$ and $\phi$ as

$$\hat{\mu} = \frac{\sum_{s=2}^{T} \hat{Z}_s \sum_{t=2}^{T} \hat{Z}_{t-1} - \sum_{s=2}^{T} \hat{Z}_{s-1} \sum_{t=2}^{T} \hat{Z}_t \hat{Z}_{t-1}}{(T-1) \sum_{t=2}^{T} \hat{Z}_{t-1}^2 - (\sum_{t=2}^{T} \hat{Z}_{t-1})^2},$$

$$\hat{\phi} = \frac{(T-1) \sum_{s=2}^{T} \hat{Z}_t \hat{Z}_{t-1} - \sum_{s=2}^{T} \hat{Z}_s \sum_{t=2}^{T} \hat{Z}_{t-1}}{(T-1) \sum_{t=2}^{T} \hat{Z}_{t-1}^2 - (\sum_{t=2}^{T} \hat{Z}_{t-1})^2}.$$

Similarly, by minimizing the following least squares

$$\sum_{t=1}^{T} \left( \log m(x, t) - \alpha_x - \beta_x \hat{Z}_t \right)^2,$$

we obtain the least squares estimator for $\alpha_x$ and $\beta_x$ as

$$\hat{\alpha}_x = \frac{\sum_{s=1}^{T} \log m(x, s) \sum_{t=1}^{T} \hat{Z}_t - \sum_{s=1}^{T} \log m(x, s) \hat{Z}_s \sum_{t=1}^{T} \hat{Z}_t}{T \sum_{t=1}^{T} \hat{Z}_t^2 - (\sum_{t=1}^{T} \hat{Z}_t)^2},$$

$$\hat{\beta}_x = \frac{T \sum_{s=1}^{T} \log m(x, s) \hat{Z}_s - \sum_{s=1}^{T} \log m(x, s) \sum_{t=1}^{T} \hat{Z}_t}{T \sum_{t=1}^{T} \hat{Z}_t^2 - (\sum_{t=1}^{T} \hat{Z}_t)^2}.$$

In order to derive the asymptotic properties of the above proposed estimators, we assume
the following regularity conditions for the stationary errors in (4) and (5).

- C1) $E(e_t) = 0$, $E(\varepsilon_{x,t}) = 0$ for $t = 1, \cdots, T$ and $x = 1, \cdots, M$;
- C2) there exist $\beta > 2$ and $\delta > 0$ such that $\sup_t E|e_t|^\beta + \delta < \infty$, $\sup_t E|\varepsilon_{x,t}|^{\beta + \delta} < \infty$ for
  $x = 1, \cdots, M$;
• C3) $\sigma^2 = \lim_{T \to \infty} E\{T^{-1}(\sum_{t=1}^{T} e_t)^2\} \in (0, \infty)$ and $\sigma^2 = \lim_{T \to \infty} E\{T^{-1}(\sum_{t=1}^{T} \varepsilon_{x,T})^2\} \in (0, \infty)$ for $x = 1, \ldots, M$;

• C4) sequence $\{e_t, \varepsilon_{1,t}, \ldots, \varepsilon_{M,t}\}^T$ is strong mixing with mixing coefficients $\alpha_m = \sup_{k \geq 1} \sup_{A \in F^k, B \in F_{k+m}^\infty} |P(A \cap B) - P(A)P(B)|$

such that $\sum_{m=1}^{\infty} \alpha_m^{1-2/\beta} < \infty$, where $F^k$ denotes the $\sigma$-field generated by $\{e_t, \varepsilon_{1,t}, \ldots, \varepsilon_{M,t}\}^T : k \leq t \leq k+m \}$ and $A^T$ denotes the transpose of the matrix or vector $A$.

Under the above regularity conditions, it is known that for $r = (r_1, \ldots, r_{M+1})^T \in [0, 1]^{M+1}$ there is a $(M + 1)$-dimensional Gaussian Process $\{W(r)\}$ such that

$$\left(\frac{\sum_{t=1}^{[Tr_1]} e_t}{\sigma_e \sqrt{T}}, \frac{\sum_{t=1}^{[Tr_2]} \varepsilon_{1,t}}{\sigma_1 \sqrt{T}}, \ldots, \frac{\sum_{t=1}^{[Tr_{M+1}]} \varepsilon_{M,t}}{\sigma_M \sqrt{T}}\right)^T \overset{D}{\to} W(r)$$ in the space $D([0, 1]^{M+1})$, (7)

where $[x]$ is the floor function, $D([0, 1]^{M+1})$ denotes the space of real-valued functions on $[0, 1]^{M+1}$ that are right continuous and have finite left limits, "$\overset{D}{\to}$" denotes the weak convergence of the associated probability measures. Throughout we will use "$\overset{d}{\to}$" and "$\overset{p}{\to}$" to denote the convergence in distribution and in probability, respectively. We also use $W_i(r_i)$ to denote the marginal distribution of $W(r)$ for the $i$th variable and define

$$W_i(s) = \sum_{i=2}^{M+1} \sigma_{i-1} W_i(s) \text{ for } s \in [0, 1].$$

As $\mu$ in (5) is often nonzero in real mortality rates, the following theorem provides the asymptotic theory for the proposed estimators when (5) holds with $\mu \neq 0$ and $\phi = 1 + \gamma/T$ for some constant $\gamma \in \mathbb{R}$. Similar results can be derived with a different rate of convergence for $\hat{\phi}$ when $\mu = 0$ and $\phi = 1 + \gamma/T$.

Theorem 1. Assume (4) and (5) hold with conditions C1)–C4), $\mu \neq 0$, $\phi = 1 + \gamma/T$ for some constant $\gamma \in \mathbb{R}$ and $k_0$ is a constant. Define $f_\gamma(s) = (1 - e^{\gamma s})/(-\gamma)$ for $s \in [0, 1]$. Then the following convergences are true as $T \to \infty$.  

7
(i) \[ T^{3/2}(\hat{\phi} - \phi) \overset{d}{\rightarrow} \frac{\sigma_e}{\mu} \int_0^1 f_\gamma(s) dW_1(s) - W_1(1) \int_0^1 f_\gamma(s) ds \] 

(ii) \[ T^{1/2}(\hat{\mu} - \mu) \overset{d}{\rightarrow} \frac{\sigma_e W_1(1) \int_0^1 f_\gamma(t) dW_1(t) - \sigma_e \int_0^1 f_\gamma(s) ds \int_0^1 f_\gamma(t) dW_1(t)}{\int_0^1 f_\gamma^2(s) ds - (\int_0^1 f_\gamma(s) ds)^2} \] 

(iii) For \( x = 1, 2, \ldots, M \) 
\[ T^{1/2}(\hat{\alpha}_x - \alpha_x) \overset{d}{\rightarrow} \beta_x Y_* - Y_x, \] 
where 
\[ Y_* = \frac{\int_0^1 f_\gamma(s) ds \int_0^1 f_\gamma(t) dW_*(t) - W_*(1) \int_0^1 f_\gamma^2(s) ds}{\int_0^1 f_\gamma^2(s) ds - (\int_0^1 f_\gamma(s) ds)^2} \] 
and 
\[ Y_x = \frac{\sigma_x \int_0^1 f_\gamma(s) ds \int_0^1 f_\gamma(t) dW_{x+1}(t) - \sigma_x W_{x+1}(1) \int_0^1 f_\gamma^2(s) ds}{\int_0^1 f_\gamma^2(s) ds - (\int_0^1 f_\gamma(s) ds)^2} \]

(iv) For \( x = 1, 2, \ldots, M \) 
\[ T^{3/2}(\hat{\beta}_x - \beta_x) \overset{d}{\rightarrow} \frac{\beta_x Z_* - Z_x}{\mu} \] 
where 
\[ Z_* = \frac{W_*(1) \int_0^1 f_\gamma(s) ds - \int_0^1 f_\gamma(s) dW_*(s)}{\int_0^1 f_\gamma^2(s) ds - (\int_0^1 f_\gamma(s) ds)^2} \] 
and 
\[ Z_x = \frac{\sigma_x W_{x+1}(1) \int_0^1 f_\gamma(s) ds - \sigma_x \int_0^1 f_\gamma(s) dW_{x+1}(s)}{\int_0^1 f_\gamma^2(s) ds - (\int_0^1 f_\gamma(s) ds)^2} \]

Remark 1. It is easy to check that \( \sum_{x=1}^M \hat{\alpha}_x = 0 \) and \( \sum_{x=1}^M \hat{\beta}_x = 1 \), which satisfy the constraints on \( \{\alpha_x\} \) and \( \{\beta_x\} \) given in (4). Also it is not difficult to see from the proof that the proposed estimators are inconsistent when \( \{k_t\} \) is a stationary AR(1) process, i.e., \( |\phi| < 1 \) independent of \( T \).

2.3 Unit root test

As many applications simply fit a unit root AR(1) model to the mortality index, it becomes important to test \( H_0 : \phi = 1 \) in (5). Under \( H_0 \), we have \( f_\gamma(s) = s \) and so it follows from Theorem
1(i) that
\[ T^{3/2}(\hat{\phi} - \phi) \overset{d}{\to} \frac{12\sigma_e}{\mu} \left\{ \frac{W_1(1)}{2} - \int_0^1 W_1(s) \, ds \right\} \sim N(0, \frac{12\sigma_e^2}{\mu^2}). \quad (8) \]

In order to employ the above limit to test \( H_0 : \phi = 1 \), one has to estimate \( \sigma_e^2 \). Unfortunately the estimators proposed by Phillips and Perron (1988) are not applicable due to the involved "measurement errors" \( \eta_t \)'s. Here we employ the idea of block sample variance estimation in Carlstein (1986) and Politis and Romano (1993).

For integer \( L \), define \( \hat{e}_t = \hat{Z}_t - \hat{\mu} - \hat{\phi} \hat{Z}_{t-1} \) for \( t = 2, \cdots, T \), and \( \hat{U}_i = \sum_{j=1}^L \hat{e}_{i+j} \) for \( i = 1, \cdots, T - L \). Then for estimating \( \sigma_e^2 = \lim_{T \to \infty} E(\frac{\sum_{t=2}^T e_t^2}{T})^2 \), we consider
\[ \hat{\sigma}_e^2 = L \left\{ \frac{1}{T - L} \sum_{i=1}^{T-L} \hat{U}_i^2 - \left( \frac{1}{T - L} \sum_{j=1}^{T-L} \hat{U}_j \right)^2 \right\}. \]

**Theorem 2.** Suppose conditions in Theorem 1 hold. Further assume \( L^{-1} + T^{-1}L \to 0 \) as \( T \to \infty \). Then \( \hat{\sigma}_e^2 \overset{p}{\to} \sigma_e^2 \) as \( T \to \infty \).

Based on (8) and Theorem 2, we reject the null hypothesis \( H_0 : \phi = 1 \) at level \( a \) if \( \frac{T^{3/2}(\hat{\phi} - 1)^2}{12\hat{\sigma}_e^2} \geq \chi^2_{1,1-a} \), where \( \chi^2_{1,1-a} \) denotes the \((1 - a)\)-th quantile of the chi-squared distribution with one degree of freedom.

**Remark 2.** When \( \{e_t\} \) and \( \{\varepsilon_{xt}\} \) for \( x = 1, \cdots, M \) are sequences of independent and identically distributed random variables and all sequences are independent, a different estimator for \( \sigma_e^2 \) without a tuning parameter is
\[ \tilde{\sigma}_e^2 = \frac{1}{T - 1} \sum_{t=2}^T \hat{e}_t^2 + \frac{2}{T - 2} \sum_{t=3}^T \hat{e}_t \hat{e}_{t-1}. \]

### 2.4 Forecast

When the null hypothesis \( H_0 : \phi = 1 \) is not rejected, we could forecast the future \( d \) mortality rates based on models (4) and (5) with \( \phi = 1 \), which are
\[ \log \tilde{m}(x, T + r) = \hat{\alpha}_x + \hat{\beta}_x \{ \hat{Z}_T + r\hat{\mu} \} \quad \text{for} \quad r = 1, 2, \cdots, d. \]
Note that
\[
\log m(x, T + r) - \log m(x, T + r) \\
= \hat{\alpha}_x + \hat{\beta}_x \{ \hat{Z}_T + r \hat{\mu} \} - \alpha_x - \beta_x \{ r \mu + k_T + \sum_{s=1}^r \epsilon_{T+s} \} - \varepsilon_{x,T+r} \\
= \hat{\alpha}_x - \alpha_x + r \{ \hat{\beta}_x \hat{\mu} - \beta_x \mu \} + (\hat{\beta}_x - \beta_x) \hat{Z}_T + \beta_x \eta_T - \beta_x \sum_{s=1}^r e_{T+s} - \varepsilon_{x,T+r} \\
= \beta_x \eta_T - \beta_x \sum_{s=1}^r e_{T+s} - \varepsilon_{x,T+r} + o_p(1). 
\]
(9)

In order to quantify the uncertainties of the above forecasts, one has to estimate the distribution function
\[ G_r(y) = P(\beta_x \eta_T - \beta_x \sum_{s=1}^r e_{T+s} - \varepsilon_{x,T+r} \leq y). \]

Unfortunately it seems that \( G_r(y) \) can not be estimated nonparametrically without imposing more conditions on \( \varepsilon_{x,t} \)'s. However, we have
\[
\frac{1}{M} \sum_{x=1}^M \log m(x, T + r) - \frac{1}{M} \sum_{x=1}^M \log m(x, T + r) = \frac{1}{M} \{ \eta_T - \sum_{s=1}^r e_{T+s} - \eta_{T+r} \} + o_p(1)
\]
and the distribution function of
\[ H_r(y) = P(\frac{1}{M}(\eta_T - \sum_{s=1}^r e_{T+s} + \eta_{T+r}) \leq y) \]
can be estimated nonparametrically by
\[ \hat{H}_r(y) = \frac{1}{T-r} \sum_{t=1}^{T-r} I(\frac{1}{M} \sum_{s=1}^r \hat{\epsilon}_{t+s} \leq y). \]

Define
\[ c_{l,a} = \sup\{ y : \hat{H}_r(y) \leq a/2 \} \quad \text{and} \quad c_{u,a} = \sup\{ y : \hat{H}_r(y) \leq 1 - a/2 \}. \]

Then an interval forecast for \( \frac{1}{M} \sum_{x=1}^M \log m(x, T + r) \) with level \( a \) is
\[ I_a = \left( \frac{1}{M} \sum_{x=1}^M \log m(x, T + r) - c_{u,a}, \quad \frac{1}{M} \sum_{x=1}^M \log m(x, T + r) - c_{l,a} \right). \]

**Theorem 3.** Suppose conditions in Theorem 1 hold and \( \phi = 1 \). Then for any fixed integer \( r \geq 1 \) and \( a \in (0,1) \),
\[ P\left( \frac{1}{M} \sum_{x=1}^M \log m(x, T + r) \in I_a \right) \to a \quad \text{as} \quad T \to \infty. \]
3 Data Analysis and Simulation

3.1 Data Analysis

To illustrate how the proposed model and inference can be applied to mortality data and how the new method differs from the classic Lee-Carter model, we employ the mortality data from the Human Mortality Database (HMD) (see http://www.mortality.org/cgi-bin/hmd/country.php?cntr=USA&level=1). To gain a robust conclusion, we study the central death rates of U.S. female, male and combined population between 25 and 74 years old from year 1933 to year 2015, and use the mortality data by 5-year age groups. Hence we have $M = 10$ and $T = 83$.

First, to implement the classic Lee-Carter model, we employ the statistical R package ‘demography’ to obtain estimates for $\alpha_x$’s, $\beta_x$’s, $k_t$’s, and then use the obtained estimates for $k_t$’s to fit model (5) by using ‘lm’ in the statistical software R. We report the estimates for $\alpha_x$’s, $\beta_x$’s, $\mu$ and $\phi$ in Tables 1, 2 and 3 for the female, male and combined mortality rates, respectively. As the asymptotic property for the estimate of $\phi$ is unknown, one can not simply use the standard errors obtained from ‘lm’ to conclude whether $\phi = 1$ or not. Also one can not employ the commonly employed unit root tests based on estimates of $k_t$’s to test $H_0 : \phi = 1$.

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_x$</td>
<td>0.135</td>
<td>0.128</td>
<td>0.119</td>
<td>0.106</td>
<td>0.095</td>
<td>0.090</td>
<td>0.083</td>
<td>0.080</td>
<td>0.081</td>
<td>0.083</td>
</tr>
</tbody>
</table>

Table 1: Female mortality rates. Parameter estimates are obtained from fitting models (1) and (5) based on the two-step inference in Lee and Carter (1992).
Next we apply our proposed inference to fit models (4) and (5) to the female, male and combined mortality rates. Again we use ‘lm’ to obtain our proposed least squares estimates and report the estimates for $\alpha_x$’s, $\beta_x$’s, $\mu$ and $\phi$ in Tables 4, 5 and 6 for the female, male and combined mortality rates, respectively. As before, the standard errors obtained from ‘lm’ is inaccurate since it ignores the involved $\eta_t$’s and so one can not conclude whether $\phi = 1$ or not.
from these three tables. Although the estimate for $\phi$ obtained from the new method is similar to that obtained from the Lee-Carter method, estimates for $\mu$ are quite different for both methods since the new method does not assume $\sum_{t=1}^{T} k_t = 0$.

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}_x$</td>
<td>0.172</td>
<td>0.055</td>
<td>-0.022</td>
<td>-0.344</td>
<td>-0.474</td>
<td>-0.327</td>
<td>-0.337</td>
<td>-0.067</td>
<td>0.384</td>
<td>0.959</td>
</tr>
<tr>
<td>$\hat{\beta}_x$</td>
<td>0.135</td>
<td>0.127</td>
<td>0.119</td>
<td>0.106</td>
<td>0.096</td>
<td>0.091</td>
<td>0.083</td>
<td>0.080</td>
<td>0.081</td>
<td>0.083</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\mu}$</td>
<td>-1.389</td>
</tr>
<tr>
<td>$\hat{\phi}$</td>
<td>0.977</td>
</tr>
</tbody>
</table>

Table 4: Female mortality rates. *Parameter estimates are obtained from fitting models (4) and (5) based on the proposed least squares estimation.*

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}_x$</td>
<td>-2.068</td>
<td>-1.631</td>
<td>-0.789</td>
<td>-0.270</td>
<td>0.099</td>
<td>0.547</td>
<td>0.714</td>
<td>0.940</td>
<td>1.152</td>
<td>1.308</td>
</tr>
<tr>
<td>$\hat{\beta}_x$</td>
<td>0.088</td>
<td>0.094</td>
<td>0.106</td>
<td>0.109</td>
<td>0.108</td>
<td>0.108</td>
<td>0.103</td>
<td>0.099</td>
<td>0.096</td>
<td>0.090</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\mu}$</td>
<td>-0.441</td>
</tr>
<tr>
<td>$\hat{\phi}$</td>
<td>0.993</td>
</tr>
</tbody>
</table>

Table 5: Male mortality rates. *Parameter estimates are obtained from fitting models (4) and (5) based on the proposed least squares estimation.*
Table 6: Combined mortality rates. Parameter estimates are obtained from fitting models (4) and (5) based on the proposed least squares estimation.

Finally we apply our proposed unit root test to the female, male and combined mortality rates, where we use $\tilde{\sigma}_e^2$ with $L = 0.5\sqrt{T}$, $\sqrt{T}$, $2\sqrt{T}$ and $\tilde{\sigma}_e^2$ (denoted by $L = \ast$) given in Remark 2. Note that (8) requires $k_0/T \to 0$ as $T \to \infty$. Since $|\hat{Z}_1|/T$ is around 0.5 which is far larger than zero, the limiting distribution of the proposed unit root test under the unit root null hypothesis will be away from a chi-squared distribution for the given $T = 83$. Therefore we apply the proposed unit root test to $\{\hat{Z}_t - \hat{Z}_1\}_{t=1}^T$. The obtained variance estimates, test statistics and P-values are reported in Tables 7, 8 and 9 for the female, male and combined mortality rates, respectively. As we see, these quantities are quite robust to the choice of $L$. Moreover, the proposed test rejects the unit root hypothesis for the female and combined mortality rates, but fails to reject the unit root hypothesis for the male mortality rates.

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\alpha}_x$</td>
<td>-1.264</td>
<td>-0.949</td>
<td>-0.452</td>
<td>-0.272</td>
<td>-0.105</td>
<td>0.219</td>
<td>0.293</td>
<td>0.509</td>
<td>0.815</td>
<td>1.205</td>
</tr>
<tr>
<td>$\hat{\beta}_x$</td>
<td>0.105</td>
<td>0.108</td>
<td>0.112</td>
<td>0.108</td>
<td>0.103</td>
<td>0.101</td>
<td>0.094</td>
<td>0.091</td>
<td>0.089</td>
<td>0.088</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\mu}$</td>
<td>-0.906</td>
<td>0.343</td>
</tr>
<tr>
<td>$\hat{\phi}$</td>
<td>0.985</td>
<td>0.007</td>
</tr>
<tr>
<td>$L$</td>
<td>$\hat{\sigma}_e^2$</td>
<td>Test statistic</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>$\left[\frac{1}{2}\sqrt{T}\right]$</td>
<td>0.042</td>
<td>75.938</td>
</tr>
<tr>
<td>$\left[\sqrt{T}\right]$</td>
<td>0.052</td>
<td>61.378</td>
</tr>
<tr>
<td>$\left[2\sqrt{T}\right]$</td>
<td>0.038</td>
<td>85.207</td>
</tr>
<tr>
<td>*</td>
<td>0.047</td>
<td>68.809</td>
</tr>
</tbody>
</table>

Table 7: Female mortality rates. Variance estimates, test statistics and Pvalues are reported for $L = \left[\frac{1}{2}\sqrt{T}\right], \left[\sqrt{T}\right], \left[2\sqrt{T}\right]$, where $'L = *'$ denotes $\tilde{\sigma}_e^2$.

<table>
<thead>
<tr>
<th>$L$</th>
<th>$\hat{\sigma}_e^2$</th>
<th>Test statistic</th>
<th>Pvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\left[\frac{1}{2}\sqrt{T}\right]$</td>
<td>0.064</td>
<td>0.798</td>
<td>0.372</td>
</tr>
<tr>
<td>$\left[\sqrt{T}\right]$</td>
<td>0.073</td>
<td>0.692</td>
<td>0.405</td>
</tr>
<tr>
<td>$\left[2\sqrt{T}\right]$</td>
<td>0.064</td>
<td>0.793</td>
<td>0.373</td>
</tr>
<tr>
<td>*</td>
<td>0.073</td>
<td>0.699</td>
<td>0.403</td>
</tr>
</tbody>
</table>

Table 8: Male mortality rates. Variance estimates, test statistics and Pvalues are reported for $L = \left[\frac{1}{2}\sqrt{T}\right], \left[\sqrt{T}\right], \left[2\sqrt{T}\right]$, where $'L = *'$ denotes $\tilde{\sigma}_e^2$.

<table>
<thead>
<tr>
<th>$L$</th>
<th>$\hat{\sigma}_e^2$</th>
<th>Test statistic</th>
<th>Pvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\left[\frac{1}{2}\sqrt{T}\right]$</td>
<td>0.051</td>
<td>12.229</td>
<td>4.704e-4</td>
</tr>
<tr>
<td>$\left[\sqrt{T}\right]$</td>
<td>0.061</td>
<td>10.289</td>
<td>1.338e-3</td>
</tr>
<tr>
<td>$\left[2\sqrt{T}\right]$</td>
<td>0.049</td>
<td>12.727</td>
<td>3.605e-4</td>
</tr>
<tr>
<td>*</td>
<td>0.058</td>
<td>10.789</td>
<td>1.021e-3</td>
</tr>
</tbody>
</table>

Table 9: Combined mortality rates. Variance estimates, test statistics and Pvalues are reported for $L = \left[\frac{1}{2}\sqrt{T}\right], \left[\sqrt{T}\right], \left[2\sqrt{T}\right]$, where $'L = *'$ denotes $\tilde{\sigma}_e^2$. 
3.2 Simulation Study

To examine the finite sample performance of the proposed estimator and unit root test, we consider models (4) and (5) with \( M = 10 \), \( \alpha_x \)'s, \( \beta_x \)'s, \( \mu \) being the estimates obtained from the female mortality rates in Table 4.

We assume \( \varepsilon_{x,t} \)'s are independent random variables with \( N(0, \sigma^2_e/M) \), \( e_t \)'s are independent random variables with \( N(0, \sigma^2_e) \) and \( \varepsilon_{x,t} \)'s are independent of \( e_t \)'s. We take \( \sigma^2_e \) as \( \tilde{\sigma}^2_e \) given in Table 7, i.e., the value with \( L = \ast \). We draw 10,000 random samples from models (4) and (5) with sample size \( T = 80 \) and 150, and consider \( \phi = 1 \).

First we compute the proposed estimators for \( \alpha_x \)'s, \( \beta_x \)'s, \( \mu \) and \( \phi \) under the above settings and report the means and standard deviations of these estimators in Tables 10 and 11, which show that estimators for \( \alpha_x, \beta_x, \mu, \phi \) are accurate.

Second we investigate the size of the proposed unit root test under the above settings. We use \( \hat{\sigma}^2_e \) with \( L = 0.5\sqrt{T}, \sqrt{T}, 2\sqrt{T} \), \( \tilde{\sigma}^2_e \) denoted by \( L = \ast \) and the true value \( \sigma^2_e \) denoted by \( L = \ast\ast \) to compute the test statistics. Estimators and empirical sizes are reported in the lower panel of Tables 10 and 11, which show that the size tends to be larger than the nominal level, and the choice of \( L \) has an impact on the test. The size becomes accurate as \( T \) is larger. We also find that the proposed test has a good power when \( \phi = 1 - 2/T \), which we do not report here.
Table 10: $T = 80$. The upper and middle panels report the means and standard errors in brackets for the new estimators based on models (4) and (5). The lower panel reports the mean and standard error in brackets for estimators for $\sigma_e^2$ and the size of the proposed unit root test, where 'L = *' and 'L = **' denote the test by using $\bar{\sigma}_e^2$ and the true value $\sigma_e^2$, respectively.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{\alpha}_e$</td>
<td>0.172</td>
<td>0.055</td>
<td>-0.022</td>
<td>-0.344</td>
<td>-0.474</td>
<td>-0.327</td>
<td>-0.337</td>
<td>-0.067</td>
<td>0.384</td>
<td>0.959</td>
</tr>
<tr>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td></td>
</tr>
<tr>
<td>$\tilde{\beta}_e$</td>
<td>0.135</td>
<td>0.127</td>
<td>0.119</td>
<td>0.106</td>
<td>0.096</td>
<td>0.091</td>
<td>0.083</td>
<td>0.080</td>
<td>0.081</td>
<td>0.083</td>
</tr>
<tr>
<td>(2.308e-4)</td>
<td>(2.259e-4)</td>
<td>(2.272e-4)</td>
<td>(2.269e-4)</td>
<td>(2.274e-4)</td>
<td>(2.266e-4)</td>
<td>(2.258e-4)</td>
<td>(2.272e-4)</td>
<td>(2.271e-4)</td>
<td>(2.269e-4)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>$\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True value</td>
<td>-1.389</td>
</tr>
<tr>
<td>Estimator</td>
<td>-1.393 (0.051)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$L$</th>
<th>$[\frac{1}{\sqrt{T}}]$</th>
<th>$[\sqrt{T}]$</th>
<th>$[2\sqrt{T}]$</th>
<th>*</th>
<th>**</th>
</tr>
</thead>
<tbody>
<tr>
<td>True value of $\sigma_e^2$</td>
<td>0.047</td>
<td>0.047</td>
<td>0.047</td>
<td>0.047</td>
<td>0.047</td>
</tr>
<tr>
<td>$\bar{\sigma}_e^2$</td>
<td>0.065 (0.014)</td>
<td>0.049 (0.016)</td>
<td>0.034 (0.018)</td>
<td>0.044 (0.024)</td>
<td>0.047</td>
</tr>
<tr>
<td>Unit root test size at 5%</td>
<td>3.91%</td>
<td>7.65%</td>
<td>14.87%</td>
<td>10.95%</td>
<td>6.70%</td>
</tr>
<tr>
<td>Unit root test size at 10%</td>
<td>7.79%</td>
<td>12.84%</td>
<td>21.60%</td>
<td>16.26%</td>
<td>12.10%</td>
</tr>
</tbody>
</table>
Table 11: \( T = 150. \) The upper and middle panels report the means and standard errors in brackets for the new estimators based on models (4) and (5). The lower panel reports the mean and standard error in brackets for estimators for \( \sigma_e^2 \) and the size of the proposed unit root test, where \( 'L = \ast' \) and \( 'L = \ast\ast' \) denote the test by using \( \hat{\sigma}_e^2 \) and the true value \( \sigma_e^2 \), respectively.

4 Conclusions

After articulating the issues with model assumptions, inference and existing misunderstandings of the classic Lee-Carter mortality model, this paper proposes a modified Lee-Carter model with no condition imposed on the unobserved mortality index for model’s identification. Further a least squares estimator is proposed to estimate all unknown parameters, a unit root test is derived to test whether the mortality index follows from a unit root AR(1) process, and the asymptotic distributions of the proposed estimator and unit root test are derived when the mortality index follows from a unit root or near unit root process and errors satisfy some \( \alpha \)-mixing conditions. An application of the proposed unit root test to US mortality rates rejects the unit root hypothesis for the female and combined mortality rates, but fails to reject the unit root hypothesis for the male mortality rates. This finding does contradict the common argument...
in the literature of actuarial science that mortality index follow a unit root process. Forecasting future mortality rates is discussed too.

5 Proofs

Proof of Theorem 1. Note that

\[ k_t = \mu + \phi k_{t-1} + e_t = \mu \sum_{j=0}^{t-1} \phi^j + \phi^t k_0 + \sum_{j=0}^{t-1} \phi^{t-j} e_i. \]

Put \( \tilde{k}_t = \sum_{i=0}^{t} \phi^{t-i} e_i \), we have \( \tilde{k}_t = \phi \tilde{k}_{t-1} + e_t \), and then it follows from Phillips (1987) that

\[
\begin{align*}
T^{-2} \sum_{t=1}^{T} \tilde{k}_t^2 & \overset{d}{\to} \sigma^2 \int_0^1 J_\gamma^2(s) \, ds, \\
T^{-3/2} \sum_{t=1}^{T} \tilde{k}_t & \overset{d}{\to} \sigma \int_0^1 J_\gamma(s) \, ds, \\
T^{-1} \sum_{t=1}^{T} \tilde{k}_{t-1} e_t & \overset{d}{\to} \sigma^2 \int_0^1 J_\gamma(s) \, dW_1(s) + \frac{1}{2} (\sigma^2 - \sigma^2),
\end{align*}
\]

where \( J_\gamma(t) = \int_{t-1}^{t} e^{(t-s)\gamma} \, dW_1(s) \) and \( \sigma^2 = \lim_{n \to \infty} \frac{1}{T} \sum_{t=1}^{T} E(e_t^2) \). As \( T \to \infty \), it is easy to show that

\[
\begin{align*}
T^{-2} \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right) & \overset{d}{\to} \sigma^2 \int_0^1 J_\gamma(s) \, ds, \\
T^{-3} \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right)^2 & \overset{d}{\to} \int_0^1 f_\gamma^2(s) \, ds, \\
T^{-3/2} \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right) e_t & \overset{d}{\to} \sigma \int_0^1 f_\gamma(s) \, dW_1(s), \\
T^{-5/2} \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right)^2 e_t & \overset{d}{\to} \sigma \int_0^1 f_\gamma^2(s) \, dW_1(s),
\end{align*}
\]

and

\[
T^{-5/2} \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right) \tilde{k}_{t-1} e_t \overset{d}{\to} \sigma \int_0^1 f_\gamma(s) J_\gamma(s) \, ds.
\]

It follows from (10), (11) and (12) that

\[
\sum_{t=2}^{T} \tilde{k}_{t-1} = \sum_{t=2}^{T} k_{t-1} + \sum_{t=2}^{T} \eta_{t-1} \\
= \mu \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right) + \sum_{t=2}^{T} \phi^{t-1} k_0 + \sum_{t=2}^{T} \tilde{k}_{t-1} + o_p(T) \\
= \mu \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right) + o_p(T^2)
\]
and
\[ \sum_{t=2}^{T} \hat{Z}_{t-1}^2 = \sum_{t=2}^{T} k_{t-1}^2 + 2 \sum_{t=2}^{T} k_{t-1} \eta_{t-1} + \sum_{t=2}^{T} \eta_{t-1}^2 \]
\[ = \sum_{t=2}^{T} \left( \mu \sum_{j=0}^{t-2} \phi^j + \phi^{t-1}k_0 + \tilde{k}_{t-1} \right) \]
\[ + 2 \sum_{t=2}^{T} \eta_{t-1} \left( \mu \sum_{j=0}^{t-2} \phi^j + \phi^{t-1}k_0 + \tilde{k}_{t-1} \right) + \sum_{t=2}^{T} \eta_{t-1}^2 \]
\[ = \mu^2 \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right)^2 + o_p(T^3), \]
which imply that
\[
\begin{cases}
    T^{-4} \left\{ (T-1) \sum_{t=2}^{T} \hat{Z}_{t-1}^2 - (\sum_{t=2}^{T} \hat{Z}_{t-1})^2 \right\} \xrightarrow{p} \mu^2 \left\{ \int_0^T f_0^2(s) \, ds - (\int_0^T f_1(s) \, ds)^2 \right\}, \\
    T^{-4} \left\{ T \sum_{t=1}^{T} \hat{Z}_{t}^2 - (\sum_{t=1}^{T} \hat{Z}_{t})^2 \right\} \xrightarrow{p} \mu^2 \left\{ \int_0^T f_0^2(s) \, ds - (\int_0^T f_1(s) \, ds)^2 \right\}.
\end{cases} \tag{13}
\]

(i) Consider the numerator of \( \hat{\phi} - \phi \), which is
\[
(T - 1) \sum_{t=2}^{T} \hat{Z}_{t-1} (\hat{Z}_t - \phi \hat{Z}_{t-1}) - \sum_{t=2}^{T} \hat{Z}_{t-1} \sum_{s=2}^{T} (\hat{Z}_s - \phi \hat{Z}_{s-1}).
\]

Under the conditions of Theorem 1, an application of the law of large numbers for an \( \alpha \)-mixing sequence and Hölder inequality implies that
\[
\frac{\sum_{t=1}^{T} \eta_t}{T} = o_p(1), \quad \frac{\sum_{t=1}^{T} \eta_t^2}{T} = O_p(1), \quad \left\{ \frac{\sum_{t=2}^{T} \eta_{t-1}^2}{T} \right\} = O_p(1). \tag{14}
\]

Note that
\[
\sum_{t=2}^{T} \tilde{k}_{t-1} (\eta_t - \phi \eta_{t-1})
\]
\[ = \sum_{t=2}^{T} \tilde{k}_{t-1} \eta_t - \phi \sum_{t=1}^{T-1} \tilde{k}_t \eta_t \]
\[ = \sum_{t=2}^{T} \tilde{k}_{t-1} \eta_t - \phi \sum_{t=1}^{T-1} (\phi \tilde{k}_{t-1} + \epsilon_t) \eta_t \]
\[ = (1 - \phi^2) \sum_{t=2}^{T-1} \tilde{k}_{t-1} \eta_t + \tilde{k}_{T-1} \eta_T - \phi^2 \tilde{k}_0 \eta_1 - \phi \sum_{t=1}^{T-1} \epsilon_t \eta_t \]
\[ = O_p(T^{1/2}) \]
and
\[
\sum_{t=2}^{T} \sum_{j=0}^{t-2} \phi^j (\eta_t - \phi \eta_{t-1}) = O_p(T). \tag{16}
\]
By (10), (14), (15) and (16), we have

\[(T - 1) \sum_{t=2}^{T} \hat{Z}_{t-1} (\hat{Z}_t - \phi \hat{Z}_{t-1}) \]
\[= (T - 1) \sum_{t=2}^{T} (k_t + \eta_t - \phi \eta_{t-1}) \]
\[= (T - 1) \sum_{t=2}^{T} (\mu \sum_{j=0}^{t-2} \phi^j + \phi^{j-1} k_0 + \hat{k}_{t-1} + \eta_{t-1})(\mu + \epsilon_t + \eta_t - \phi \eta_{t-1}) \]
\[= (T - 1) \mu^2 \sum_{t=2}^{T} (\sum_{j=0}^{t-2} \phi^j) + (T - 1) \mu \sum_{t=2}^{T} (\sum_{j=0}^{t-2} \phi^j) \epsilon_t \]
\[+ (T - 1) \mu \sum_{t=2}^{T} \hat{k}_{t-1} + O_p(T^2) \]

and

\[\sum_{t=2}^{T} \hat{Z}_{t-1} \sum_{s=2}^{T} (\hat{Z}_s - \phi \hat{Z}_{s-1}) \]
\[= \left( \mu \sum_{t=2}^{T} \sum_{j=0}^{t-2} \phi^j + k_0 \sum_{t=2}^{T} \phi^{t-1} + \sum_{t=2}^{T} \hat{k}_{t-1} + \sum_{t=2}^{T} \eta_{t-1} \right) \times \]
\[\left( \mu(T - 1) + \sum_{s=2}^{T} \epsilon_s + \sum_{s=2}^{T} \eta_s - \phi \sum_{s=2}^{T} \eta_{s-1} \right) \]
\[= (T - 1) \mu^2 \sum_{t=2}^{T} (\sum_{j=0}^{t-2} \phi^j) + \mu \left( \sum_{t=2}^{T} \sum_{j=0}^{t-2} \phi^j \right) \sum_{s=2}^{T} \epsilon_s \]
\[+ (T - 1) \mu \sum_{t=2}^{T} \hat{k}_{t-1} + O_p(T^2), \]

which imply that

\[(T - 1) \sum_{t=2}^{T} \hat{Z}_{t-1} (\hat{Z}_t - \phi \hat{Z}_{t-1}) - \sum_{t=2}^{T} \hat{Z}_{t-1} \sum_{s=2}^{T} (\hat{Z}_s - \phi \hat{Z}_{s-1}) \]
\[\overset{d}{\rightarrow} \frac{\sigma_\epsilon}{\mu} \int_0^1 \frac{1}{f_\gamma(s)} dW_1(s) - W_1(1) \int_0^1 f_\gamma(s) ds \]
\[\int_0^1 f_\gamma^2(s) ds - (\int_0^1 f_\gamma(s) ds)^2]. \]

(ii) Write the numerator of \( \hat{\mu} - \mu \) as

\[\sum_{s=2}^{T} \hat{Z}_s \sum_{t=2}^{T} \hat{Z}^2_{t-1} - \sum_{s=2}^{T} \hat{Z}_{s-1} \sum_{t=2}^{T} \hat{Z}_t \hat{Z}_{t-1} - (T - 1) \mu \sum_{t=2}^{T} \hat{Z}^2_{t-1} + \mu (\sum_{t=2}^{T} \hat{Z}_{t-1})^2 \]
\[\overset{d}{=} \sum_{t=2}^{T} \hat{Z}_s \sum_{t=2}^{T} \hat{Z}^2_{t-1} - \sum_{s=2}^{T} \hat{Z}_{s-1} \sum_{t=2}^{T} \hat{Z}^2_{t-1} - \sum_{s=2}^{T} \hat{Z}_{s-1} \sum_{t=2}^{T} (\hat{Z}_t - \hat{Z}_{t-1}) \hat{Z}_{t-1} \]
\[\overset{d}{=}\]
\[\overset{d}{=}
\]
\[\overset{d}{=}
\]
\[\overset{d}{=}
\]
\[\overset{d}{=} \sum_{t=2}^{T} \hat{Z}^2_{t-1} + \mu (\sum_{t=2}^{T} \hat{Z}_{t-1})^2 \]
\[\overset{d}{=}
\]
\[\overset{d}{=}
\]
\[\overset{d}{=}
\]
\[\overset{d}{=} \sum_{t=2}^{T} \hat{Z}^2_{t-1} + \sum_{t=2}^{T} (\hat{Z}_t - \hat{Z}_{t-1} - \mu) \]
\[\overset{d}{=}
\]
\[\overset{d}{=}
\]
\[\overset{d}{=}
\]
\[\overset{d}{=} I_1 - I_2. \]

21
Using (10), (11), (12), (15) and (16), we have

\[
I_1 = \{ \sum_{t=2}^{T} \hat{Z}_{t-1} \} \{(\phi - 1) \sum_{s=2}^{T} k_{s-1} + \sum_{s=2}^{T} e_s + \eta_T - \eta_0 \}
\]

\[
= \{ \sum_{t=2}^{T} (\mu \sum_{j=0}^{t-2} \phi^j + \phi^{t-1} k_0 + \hat{k}_{t-1} + \eta_{t-1})^2 \} \times \\
\{(\phi - 1)(\mu \sum_{j=0}^{T} \sum_{t=2}^{T} \phi^j + \sum_{t=2}^{T} \hat{k}_{t-1} + k_0 + \sum_{t=2}^{T} \phi^{t-1}) + \sum_{t=2}^{T} e_t + \eta_T - \eta_0 \}
\]

\[
= \{ \mu^2 \sum_{t=2}^{T} (\sum_{j=0}^{t-2} \phi^j)^2 + 2\mu \sum_{t=2}^{T} (\sum_{j=0}^{t-2} \phi^j) \hat{k}_{t-1} + O_p(T^2) \} \times \\
\{(\phi - 1)(\mu \sum_{j=0}^{T} \sum_{t=2}^{T} \phi^j) + (\phi - 1) \sum_{t=2}^{T} \hat{k}_{t-1} + \sum_{t=2}^{T} e_t + O_p(1) \}
\]

\[
= \mu^3(\phi - 1) \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right)^2 \sum_{s=2}^{T} \left( \sum_{j=0}^{s-2} \phi^j \right) \\
+ \mu^2(\phi - 1) \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right)^2 \sum_{s=2}^{T} \hat{k}_{s-1} \\
+ 2\mu^2(\phi - 1) \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right) \sum_{s=2}^{T} \left( \sum_{j=0}^{s-2} \phi^j \right) \hat{k}_{s-1} \\
+ \mu^2 \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right)^2 \sum_{s=2}^{T} e_s + O_p(T^3)
\]

and

\[
I_2 = \{ \sum_{s=2}^{T} \hat{Z}_{s-1} \} \sum_{t=2}^{T} \{(k_{t-1} + \eta_{t-1})((\phi - 1)k_{t-1} + e_t + \eta_t - \eta_{t-1}) \}
\]

\[
= \{ \mu \sum_{t=2}^{T} \sum_{j=0}^{t-2} \phi^j + \sum_{t=2}^{T} \hat{k}_{t-1} + O_p(T) \} \sum_{t=2}^{T} \{(\mu \sum_{j=0}^{t-2} \phi^j + \phi^{t-1} k_0 + \hat{k}_{t-1} + \eta_{t-1}) \times \\
(\phi - 1)\mu \sum_{j=0}^{t-2} \phi^j + (\phi - 1)\phi^{t-1} k_0 + (\phi - 1)\hat{k}_{t-1} + e_t + \eta_t - \eta_{t-1} \}
\]

\[
= \{ \mu \sum_{t=2}^{T} \sum_{j=0}^{t-2} \phi^j + \sum_{t=2}^{T} \hat{k}_{t-1} + O_p(T) \} \times \\
\{ \mu^2(\phi - 1) \sum_{s=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right)^2 + 2\mu(\phi - 1) \sum_{s=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right) \hat{k}_{t-1} + \mu \sum_{s=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right) e_t + O_p(T) \}
\]

\[
= \mu^3(\phi - 1) \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right)^2 \sum_{s=2}^{T} \left( \sum_{j=0}^{s-2} \phi^j \right) \\
+ \mu^2(\phi - 1) \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right)^2 \sum_{s=2}^{T} \hat{k}_{s-1} \\
+ 2\mu^2(\phi - 1) \sum_{s=2}^{T} \left( \sum_{j=0}^{s-2} \phi^j \right) \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right) \hat{k}_{t-1} \\
+ \mu^2 \sum_{s=2}^{T} \left( \sum_{j=0}^{s-2} \phi^j \right) \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right) e_t + O_p(T^3),
\]

i.e.,

\[
I_1 - I_2 = \mu^2 \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right)^2 \sum_{s=2}^{T} e_s - \mu^2 \sum_{s=2}^{T} \left( \sum_{j=0}^{s-2} \phi^j \right) \sum_{t=2}^{T} \left( \sum_{j=0}^{t-2} \phi^j \right) e_t + O_p(T^3),
\]

which implies that

\[
T^{1/2}(\hat{\mu} - \mu) \xrightarrow{d} \frac{\sigma_W W_1(1) \int_0^1 f_r^2(s) \, ds - \sigma_\phi \int_0^1 f_\gamma(s) \, ds \int_0^1 f_r(t) \, dW_1(t)}{\int_0^1 f_\gamma^2(s) \, ds - (\int_0^1 f_\gamma(s) \, ds)^2}
\]

22
by using (13).

(iii) Similar to (11), it follows from (7) that

\[
\begin{align*}
T^{-3/2} & \sum_{t=1}^{T} \left( \sum_{j=0}^{t-1} \phi^j \right) \varepsilon_{x,t} \xrightarrow{d} \sigma_x \int_0^1 f_\gamma(s) \, dW_{x+1}(s), \\
T^{-3/2} & \sum_{t=1}^{T} \left( \sum_{j=0}^{t-1} \phi^j \right) \eta_t \xrightarrow{d} \int_0^1 f_\gamma(s) \, dW_s(s).
\end{align*}
\]

Write the numerator of \( \hat{\alpha}_x - \alpha_x \) as

\[
\begin{align*}
\sum_{t=1}^{T} \hat{Z}_t^2 \sum_{s=1}^{T} \log m(x,s) & - \sum_{t=1}^{T} \hat{Z}_t \sum_{s=1}^{T} \log m(x,s) \hat{Z}_s \\
- \alpha_x \sum_{t=1}^{T} \hat{Z}_t^2 + \alpha_x \sum_{t=1}^{T} \hat{Z}_t 2 \\
= & \sum_{t=1}^{T} \hat{Z}_t^2 \sum_{s=1}^{T} \{ \log m(x,s) - \alpha_x \} - \sum_{t=1}^{T} \hat{Z}_t \sum_{s=1}^{T} \hat{Z}_s \{ \log m(x,s) - \alpha_x \} \\
= & \beta_x \left( \sum_{t=1}^{T} \hat{Z}_t^2 \sum_{s=1}^{T} k_s - \sum_{t=1}^{T} \hat{Z}_t \sum_{s=1}^{T} k_s \hat{Z}_s \right) \\
& + \left( \sum_{t=1}^{T} \hat{Z}_t^2 \sum_{s=1}^{T} \varepsilon_{x,s} - \sum_{t=1}^{T} \hat{Z}_t \sum_{s=1}^{T} \varepsilon_{x,s} \hat{Z}_s \right).
\end{align*}
\]

From (10)-(12), (14), (15), (16) and the fact that \( \sum_{t=1}^{T} \eta_t k_t = O_p(T) \), we have

\[
\begin{align*}
\sum_{t=1}^{T} \hat{Z}_t^2 \sum_{s=1}^{T} k_s & - \sum_{t=1}^{T} \hat{Z}_t \sum_{s=1}^{T} k_s \hat{Z}_s \\
= & \{ \sum_{s=1}^{T} k_s \sum_{t=1}^{T} k_t \eta_t \} + \{ \sum_{s=1}^{T} k_s \sum_{t=1}^{T} k_t \eta_t \} - \{ \sum_{s=1}^{T} k_s \sum_{t=1}^{T} k_t \eta_t \} \\
= & \{ \mu^2 (\sum_{t=1}^{T} \sum_{j=0}^{t-1} \phi^j) (\sum_{t=1}^{T} \eta_t \sum_{j=0}^{t-1} \phi^j) + O_p(T^3) \} \} + \{ \mu^2 (\sum_{t=1}^{T} \sum_{j=0}^{t-1} \phi^j) (\sum_{t=1}^{T} \eta_t \sum_{j=0}^{t-1} \phi^j) + O_p(T^3) \} \\
- \{ \mu^2 (\sum_{t=1}^{T} \sum_{j=0}^{t-1} \phi^j) (\sum_{t=1}^{T} \eta_t \sum_{j=0}^{t-1} \phi^j) + O_p(T^3) \} - \{ \mu^2 (\sum_{t=1}^{T} \sum_{j=0}^{t-1} \phi^j) (\sum_{t=1}^{T} \eta_t \sum_{j=0}^{t-1} \phi^j) + O_p(T^3) \} \\
= & \mu^2 \{ \sum_{t=1}^{T} (\sum_{j=0}^{t-1} \phi^j) \}\{ \sum_{t=1}^{T} \eta_t (\sum_{j=0}^{t-1} \phi^j) \} - \mu^2 \{ \sum_{t=1}^{T} (\sum_{j=0}^{t-1} \phi^j) \} \{ \sum_{t=1}^{T} \eta_t \} + O_p(T^3)
\end{align*}
\]

and

\[
\begin{align*}
\sum_{t=1}^{T} \hat{Z}_t^2 \sum_{s=1}^{T} \varepsilon_{x,s} & - \sum_{t=1}^{T} \hat{Z}_t \sum_{s=1}^{T} \varepsilon_{x,s} \hat{Z}_s \\
= & \{ \mu^2 \sum_{t=1}^{T} \left( \sum_{j=0}^{t-1} \phi^j \right)^2 s_{x,s} + O_p(T^3) \} \\
- \{ \mu^2 \sum_{t=1}^{T} \left( \sum_{j=0}^{t-1} \phi^j \right) \} \sum_{s=1}^{T} \left( \sum_{j=0}^{t-1} \phi^j \right) \varepsilon_{x,s} + O_p(T^3) \}.
\end{align*}
\]

Therefore, it follows from (13), (18), (19)–(21) that

\[
T^{1/2} (\hat{\alpha}_x - \alpha_x) \xrightarrow{d} \beta_x Y_s - Y_x.
\]
(iv) As before, we can show that the numerator of $\hat{\beta}_x - \beta_x$ is

\[
T \sum_{s=1}^{T} \log m(x, s) \hat{Z}_s - \sum_{s=1}^{T} \log m(x, s) \sum_{t=1}^{T} \hat{Z}_{lt} - \beta_x T \sum_{t=1}^{T} \hat{Z}_{lt}^2 + \beta_x \sum_{s=1}^{T} \hat{Z}_s^2
\]

\[
= \beta_x T \left( \sum_{t=1}^{T} k_t \hat{Z}_t - \sum_{t=1}^{T} \hat{Z}_{lt}^2 \right) + \beta_x \left\{ (\sum_{t=1}^{T} \hat{Z}_t)^2 - \sum_{s=1}^{T} k_s \sum_{t=1}^{T} \hat{Z}_t \right\}
\]

\[
+ \left( T \sum_{s=1}^{T} \hat{Z}_s \sum_{t=1}^{T} \hat{Z}_{lt} - \sum_{t=1}^{T} \hat{Z}_{lt} \sum_{s=1}^{T} \hat{Z}_s \right)
\]

\[
= \beta_x \left\{ \mu \sum_{t=1}^{T} \left( \sum_{j=0}^{t-1} \phi^j \right) \sum_{s=1}^{T} \eta_s - T \mu \sum_{t=1}^{T} \left( \sum_{j=0}^{t-1} \phi^j \right) \eta_t + O_p(T^2) \right\}
\]

Then it follows from (13) and (18) that

\[
T^{3/2}(\hat{\beta}_x - \beta_x) \xrightarrow{d} \frac{\beta_x Z_x - Z_x}{\mu}.
\]

\[\square\]

**Proof of Theorem 2.** Put $U_i = L^{-1} \sum_{j=1}^{L} e_{i+j}$ for $i = 1, \ldots, T - L$. Write

\[
\frac{L}{T-L} \sum_{i=1}^{T-L} \hat{U}_i^2
\]

\[
= \frac{L}{T-L} \sum_{i=1}^{T-L} (\hat{U}_i - U_i)^2 + \frac{L}{T-L} \sum_{i=1}^{T-L} U_i^2 + \frac{2L}{T-L} \sum_{i=1}^{T-L} (\hat{U}_i - U_i) U_i
\]

\[
= I_1 + I_2 + I_3
\]

and

\[
\left( \frac{1}{T-L} \sum_{i=1}^{T-L} \hat{U}_i \right)^2
\]

\[
= \left( \frac{1}{T-L} \sum_{i=1}^{T-L} (\hat{U}_i - U_i) \right)^2 + \left( \frac{1}{T-L} \sum_{i=1}^{T-L} U_i \right)^2 + \frac{2L}{(T-L)^2} \sum_{i=1}^{T-L} (\hat{U}_i - U_i) U_i
\]

\[
= I_4 + I_5 + I_6.
\]

It follows from (3.9) and (3.10) of Lahiri (2003) that

\[
I_2 - I_5 \xrightarrow{p} \sigma_e^2 \text{ as } T \to \infty.
\]

Similarly we have

\[
\frac{L}{T-L} \sum_{i=1}^{T-L} \left( \frac{1}{L} \sum_{j=1}^{L} e_{i+j} \right)^2 = O_p(1).
\]

24
Since
\[ \tilde{e}_t - e_t = (\hat{Z}_t - \hat{\mu} - \hat{\phi} \hat{Z}_{t-1}) - (k_t - \mu - \phi k_{t-1}) \]
\[ = (\hat{Z}_t - k_t) - (\hat{\mu} - \mu) - (\hat{\phi} \hat{Z}_{t-1} - \phi \hat{Z}_{t-1} + \phi \hat{Z}_{t-1} - \phi k_{t-1}) \]
\[ = (\eta_t - \phi \eta_{t-1}) - (\hat{\mu} - \mu) - (\hat{\phi} - \phi)(k_{t-1} + \eta_{t-1}), \]
we have
\[ \hat{U}_i - U_i = \frac{1}{L} \sum_{j=1}^{L} (\eta_{i+j} - \phi \eta_{i+j-1}) - (\hat{\mu} - \mu) - (\hat{\phi} - \phi) \frac{1}{L} \sum_{j=1}^{L} (k_{i+j-1} + \eta_{i+j-1}). \quad (26) \]

Hence
\[ I_1 = \frac{L}{T - L} \sum_{i=1}^{T-L} \left\{ \frac{1}{L} \sum_{j=1}^{L} (\eta_{i+j} - \phi \eta_{i+j-1}) - (\hat{\mu} - \mu) - (\hat{\phi} - \phi) \frac{1}{L} \sum_{j=1}^{L} (k_{i+j-1} + \eta_{i+j-1}) \right\}^2 \]
\[ \leq \frac{3L}{T - L} \sum_{i=1}^{T-L} \left\{ \frac{1}{L} \sum_{j=1}^{L} (\eta_{i+j} - \phi \eta_{i+j-1}) \right\}^2 + 3L(\hat{\mu} - \mu)^2 \]
\[ + (\hat{\phi} - \phi)^2 \frac{3L}{T - L} \sum_{i=1}^{T-L} \left\{ \frac{1}{L} \sum_{j=1}^{L} (k_{i+j-1} + \eta_{i+j-1}) \right\}^2 \]
\[ \leq \frac{3}{T - L} \sum_{i=1}^{T-L} \eta_{i+L}^2 + \frac{9\phi^2}{T - L} \sum_{i=1}^{T-L} \eta_{i}^2 + \frac{9(1-\phi)}{T - L} \sum_{i=1}^{T-L} \left( \sum_{j=1}^{L} \eta_{i+j} \right)^2 + 3L(\hat{\mu} - \mu)^2 \]
\[ + (\hat{\phi} - \phi)^2 \frac{6L}{T - L} \sum_{i=1}^{T-L} \left\{ \frac{1}{L} \sum_{j=1}^{L} (k_{i+j-1}) \right\}^2 + (\hat{\phi} - \phi)^2 \frac{6L}{T - L} \sum_{i=1}^{T-L} \left\{ \frac{1}{L} \sum_{j=1}^{L} (\eta_{i+j}) \right\}^2 \]
\[ = I_1 + \cdots + I_6. \]

Using (25), Theorem 1, $E\eta_i^2 < \infty$, $1 - \phi \to 0$ and $L \to \infty$ as $n \to \infty$, we have
\[ II_i = o_p(1) \text{ for } i = 1, 2, 3, 4, 6. \quad (27) \]

By noting that $\hat{\phi} - \phi = O_p(T^{-3/2})$ and $T^{-1} \max_{1 \leq t \leq T} |k_t| = O_p(1)$, we have
\[ II_5 = o_p(1). \quad (28) \]

Hence, it follows from (27) and (28) that
\[ I_1 = o_p(1). \quad (29) \]

By H"older inequality, we have
\[ I_3 = O_p(\sqrt{|I_1|}) = o_p(1). \quad (30) \]
It follows from (26) and similar arguments in proving (29) that

\[ I_4 = o_p(1). \quad (31) \]

Using Hölder inequality again, we have

\[ I_6 = O_p(\sqrt{|I_4I_5|}) = o_p(1). \quad (32) \]

Therefore the theorem follows from (24), (29)–(32).

\[ \square \]

**Proof of Theorem 3.** By noting that

\[
\sum_{s=1}^{r} \hat{e}_{t+s} = \sum_{s=1}^{r} e_{t+s} + \eta_{t+r} - \eta_t \quad -r(\hat{\mu} - \mu) - (\hat{\phi} - 1) \sum_{s=1}^{r} k_{t+s-1} - (\hat{\phi} - 1) \sum_{j=1}^{r} \eta_{t+s-1},
\]

\[
\hat{\mu} - \mu = o_p(1), \quad \sup_{1 \leq t \leq T} \left\{ |(\hat{\phi} - 1) \sum_{s=1}^{r} k_{t+s-1}| + |(\hat{\phi} - 1) \sum_{s=1}^{r} \eta_{t+s-1}| \right\} = o_p(1),
\]

we have

\[
\hat{H}_r(y) = \frac{1}{T-r} \sum_{t=1}^{T-r} I\left( \frac{1}{M}(\eta_t - \sum_{s=1}^{r} e_{t+s} - \eta_{t+r}) \leq y \right) + o_p(1) \text{ for any } y \in \mathbb{R}. \quad (33)
\]

Since \( \{\eta_t - \sum_{s=1}^{r} e_{t+s} - \eta_{t+r}\} \) is strong mixing and satisfies condition C4, we have

\[
\frac{1}{T-r} \sum_{t=1}^{T-r} I\left( \frac{1}{M}(\eta_t - \sum_{s=1}^{r} e_{t+s} - \eta_{t+r}) \leq y \right) \overset{P}{\to} H_r(y) \text{ for any } y \in \mathbb{R}. \quad (34)
\]

Hence the theorem follows from (33) and (34).

\[ \square \]

**References**


