8-10-2021

Statistical Inference for Mortality Models

Chen Ling

STATISTICAL INFERENCE FOR MORTALITY MODELS

BY

CHEN LING

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree

Of

Doctor of Philosophy

In the Robinson College of Business

Of

Georgia State University

GEORGIA STATE UNIVERSITY
ROBINSON COLLEGE OF BUSINESS
2021

ACCEPTANCE

This dissertation was prepared under the direction of the *Chen Ling* Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Business Administration in the J. Mack Robinson College of Business of Georgia State University.

Richard Phillips, Dean

DISSERTATION COMMITTEE

Liang Peng
Samuel Cox
Fangda Liu
Steve Lo
Stephen Shore
Ajay Subramanian

ABSTRACT

STATISTICAL INFERENCE FOR MORTALITY MODELS

BY

CHEN LING

JULY 19, 2021

Committee Chair: *Liang Peng*

Major Academic Unit: *Department of Risk Management and Insurance*

Underwriters of annuity products and administrators of defined-benefit pension plans with periodic payments are under financial obligation to their policyholders or participants until the death of the counterparty. Hence, the underwriters would be subject to longevity risk should the average lifespan of the entire population increase to an unforeseen level. Meanwhile, the fact that the effective federal funds rate is at its historic low level implies that the present value of life-contingent cash outflows for insurers is subject to the greatest amount of longevity risk. As a benchmark mortality model in the insurance industry is the Lee-Carter model, in this dissertation we summarize some flaws of model assumptions and the model's classical inference method. Based on the understanding of these flaws, we propose a modified Lee-Carter model, accompanied by a rigorous statistical inference with asymptotic results and satisfactory numerical and simulation results derived from a relatively small sample. Then we propose a bias-corrected estimator, which is consistent and asymptotically normally distributed regardless of the mortality index being a unit root or stationary AR(1) time series. We further extend the model to accommodate AR(2) process for the mortality index and apply it to a bivariate dataset of U.S. mortality rates. Finally, we conclude the dissertation by arguing that the proposed model is adequate and by suggesting some potential hedging practices based on that.

ACKNOWLEDGEMENTS

I cannot express how grateful I am to my Ph.D. advisor, Liang Peng, for his unwavering support, advisorship, and invaluable insight into my research area in the past five years. I would also like to extend my gratitude to my dissertation committee members for sharing their wonderful comments on my dissertation: Samuel Cox, Fangda Liu, Steve Lo, Stephen Shore, and Ajay Subramanian. I would also like to extend my deepest gratitude to my parents for continuous support during the past five years. This dissertation is made possible thanks to generous scholarships and fellowships from C.V. Starr, Leyton Hunter, Helen C. Leith, and the Center for the Economic Analysis of Risk (CEAR).

TABLE OF CONTENTS

PART 1

INTRODUCTION

Annuities and pension funds can be considered financial instruments to transfer people's wealth from when they are young and productive to post-retirement. Underwriters of annuity products and administrators of defined-benefit pension plans with periodic payments are under financial obligation to their policyholders until the death of the counterparty. Insurance underwriters employ actuaries to project the probability of future insured events (deaths, disabilities, and sicknesses) in order to price insurance products and ensure solvency for compliance purposes through adequate reserves. Mortality rate, or death rate, is defined as a measure of the number of deaths (either in general or due to a specific cause) in a particular population, relative to the size of that population, within each unit of time. To make a meaningful projection of future mortality rate, actuaries develop mortality models, a statistical or demographic model built for central death rates in the life table. Although it is not easy to predict the mortality of individuals, the mortality rate of an entire population or specific demography displays some patterns and evolves stably over time. Certain statistical or demographic methods might help us understand some patterns and trends regarding the evolvement, which helps determine future obligations of underwriters and administrators.

Underwriters of annuities are subject to multiple sources of risks. Because an underwriter makes periodic payments to policyholders after their retirement, the present value of future obligations is the discounted sum of those periodic payments made by the underwriter. Thus, one of the risks the underwriters are subject to is the interest rate risk, as the discounted value is sensitive to changing level of interest rate. However, in this dissertation, I am not discussing that risk. Instead, I am focusing on the longevity risk, which is defined as any potential risk originating from an increasing life expectancy of pensioners or annuity policyholders, which eventually translates to higher periodic payments and could translate to higher than expected cash outflows for pension administrators and annuity underwriters. Underwriters need to consider longevity risk before determining policy prices. In practice, actuaries make multiple assumptions to price an insurance product like an annuity, and some of those assumptions are mortality assumptions. While mortality improvement assumptions like scale AA are used to predict future mortality, a rigorous statistical model for mortality improvements is imperative. This dissertation will clearly explain how my proposed mortality model can be used to project future mortality improvements.

It is also worth mentioning that when the effective federal funds rate is at its historically low level, the present value of life-contingent cash outflows for insurers is subject to the greatest amount of longevity risk. It can be challenging for underwriters to manage the longevity risk in this low-interest-rate environment. Such risk can be managed through hedging practices, practices of transferring longevity risk to other financial institutions. A potential way to price such risk can be based on parametric mortality models (see Li et al., 2018).

The mortality model can be used to ensure solvency through adequate reserves. NAIC has issued guidelines that "require reserves for annuity contracts be based on the amount calculated using a projection of assets and liabilities". The projection of liabilities, of course, depends on a mortality model that projects future mortality improvements.

The study of mortality rates relates to annuities and other life insurance products according to insurance economics. Literature in this field attempt to explain for household's choice of annuities and life insurance products based on mortality rates and health conditions. Research in this area attempts to explain people's allocation of assets in annuities during or near retirement (see Milevsky, 1998). A basic assumption in this framework is to assume that people's utility function is homogeneous with respect to consumption at different points in time (see Ando & Modigliani, 1963). Based on this, people purchase private constant annuities to maximize the objective function, which is their expected utility function of consumption, weighted by probabilities of death (see Yagi & Nishigaki, 1993). Hence, the study of mortality facilitates our understanding of people's choice of wealth allocation in annuities.

There is a plethora of actuarial science literature dedicated to the projection of future mortality rates. It first dates back to the De Moivre model in 1725, which researchers believe to be the first model for mortality forecast. As summarized by Li et al. (2004), mortality forecasts are traditionally based on forecasters' subjective judgments, in light of historical data and expert opinions. The De Moivre model falls under this category of subjective judgments.

However, to obtain meaningful mortality projections, the modeling of mortality rates should be less based on subjective judgments (see Tabeau, 2001). Today's common practice is to calibrate and validate quantitative models based on data derived from the life table. Here we first review some basic facts of the life table. Suppose there are M ages (or age groups) and T years of observation, the life table of a given cohort or population is a panel data with the following variables: the number of who reached age x, or the survival function, $l(x,t)$, and the number of deaths at age x, $d(x,t)$, observed at the year t, $x = 1, \dots, M$, $t = 1, \dots, T$. The central death rate $m(x,t)$ is derived from that life table and is defined as

$$m(x,t) = \frac{d(x,t)}{L(x,t)} \tag{1.1}$$

where the average number of living at the age (or, age group) x is $L(x,t) = \int_0^1 l(x+u,t)du$ (see Pollard, 1975). Data of central death rate along with survival function, are generally available in the life table. Since insurers will generally use mortality tables for contract pricing and compliance purposes, a challenge in developing and using survival models is that survival probabilities are not constant over time (see Dickson et al., 2013). Hence, this calls for the development of modeling central death rates.

A benchmark model is Lee-Carter model Lee & Carter (1992), which proposes a two-step approach towards the modeling of central death rates. The first step is the decomposition of log central death rate $\log m(x,t)$ using singular value decomposition (SVD) subject to the identification constraints:

$$\log m(x,t) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}, \quad \sum_{x=1}^{M} \beta_x = 1, \quad \sum_{t=1}^{T} k_t = 0, \tag{1.2}$$

where $\varepsilon_{x,t}$'s are independent and identically distributed (i.i.d.) random errors with mean zero and finite variance, and the unobserved k_t's are called mortality index. Constraints regarding $\{\beta_x\}_{x=1}^{M}$ and $\{k_t\}_{t=1}^{T}$ are introduced in the model (1.2) in order to make sure model unknown parameters $\{\alpha_x\}_{x=1}^{M}$, $\{\beta_x\}_{x=1}^{M}$ and $\{k_t\}_{t=1}^{T}$ can be uniquely identified using the SVD.

The next step of Lee-Carter model is fitting and calibrating an ARIMA(p,d,q) time series based on the mortality index $\{k_t\}_{t=1}^{T}$:

$$\left(1 - \sum_{i=1}^{p} \phi_i B^i\right)(1-B)^d k_t = \mu + \left(1 + \sum_{i=1}^{q} \theta_i B^i\right) e_t, \tag{1.3}$$

where e_t's are i.i.d. white noises.

In conclusion, the original Lee-Carter mortality model is a two-step inference based on log central death rates $\log m(x,t)$ data and can be understood as a combination of singular value decomposition (SVD) subject to the identification constraints in (1.2) and the ARIMA(p,d,q) time series of k_t's in (1.3). Many papers in actuarial science have claimed that an application of this model and this two-step inference method to mortality data prefer a unit root time series model, i.e., $d = 1$ in (1.3). There is also the development of the R package 'demography', which implements the Lee-Carter model, which I will use as a benchmark to evaluate and compare the performance of our proposed estimators.

Since this seminal publication there is good literature dedicated to improving and modifying the Lee-Carter mortality model. Lee (2000) argued that instead of proceeding directly to modeling the parameter $\hat{k}_t$ as a time series process, the $\hat{k}_t$'s should be adjusted (taking $\hat{\alpha}_x$ and $\hat{\beta}_x$ as what is given in original Lee-Carter model) to reproduce the observed number of deaths, that is, the $\tilde{k}_t$'s solve $\sum_{x=1}^{M} D_{xt} = \sum_{x=1}^{M} E_{xt} \exp(\hat{\alpha}_x + \hat{\beta}\tilde{k}_t)$, where D_{xt} is the number of deaths of age x and observed in year t, and E_{xt} is the actual risk exposure. After that the $\tilde{k}_t$'s can be fitted into time series model. Brouhns et al. (2002) uses the same constraints

$\sum_{x=1}^{M} \beta_x = 1$, $\sum_{t=1}^{T} k_t = 0$ as in the Lee-Carter model but models the number of deaths data D_{xt}, $x = 1, \ldots, M$, $t = 1, \ldots, T$ as:

$$\begin{cases} D_{xt} = \text{Poisson}(E_{xt}\mu_x(t)) \\ \mu_x(t) = \exp(\alpha_x + \beta_x k_t) \end{cases} \tag{1.4}$$

and then maximize the following objective function as the first step:

$$\sum_{x,t} [D_{xt}(\alpha_x + \beta_x k_t) - E_{xt} \exp(\alpha_x + \beta_x k_t)] \tag{1.5}$$

Brouhs, Denuit and Vermunt (2002)'s second step is also fitting a time series to the mortality index, but to the $\hat{k}_t$'s that maximizes the above objective function. Li & Lee (2005) extended the Lee-Carter model to a group of population $m(x, t, i)$, where x, t, i denote the age group, observation time and the i-th population, respectively.

It is worth noting here the necessity of modeling multiple populations together. There is evidence of strong positive dependence between joint lives with real economic significance, so that modeling multiple populations together allows for reduced annuity valuation (see Frees et al., 1996). It makes much difference whether multiple populations live longer simultaneously or if one population lives longer while another one lives shorter. When the life expectations of multiple populations increase simultaneously, there is no way to hedge against such longevity risk just by selling annuities to policyholders from these populations. In this dissertation, I have a section dedicated to modeling the mortality of two populations.

There is a bunch of follow-up literature that continued discussion of this topic. Girosi & King (2007) cited more than a dozen of papers to confirm the broad implementation of the Lee-Carter model by policy analysts around the world. Cairns et al. (2011) compared six different stochastic mortality models:

$$\begin{aligned}
\log m(x,t) &= \beta_x^{(1)} + \beta_x^{(2)} k_t^{(2)} \\
\log m(x,t) &= \beta_x^{(1)} + \beta_x^{(2)} k_t^{(2)} + \beta_x^{(3)} \gamma_{t-x}^{(3)} \\
\log m(x,t) &= \beta_x^{(1)} + n_a^{-1} k_t^{(2)} + n_a^{-1} \gamma_{t-x}^{(3)} \\
\text{logit } q(t,x) &= k_t^{(1)} + k_t^{(2)}(x-\bar{x}) \\
\text{logit } q(t,x) &= k_t^{(1)} + k_t^{(2)}(x-\bar{x}) + k_t^{(3)}((x-\bar{x})^2 - \hat{\sigma}_x^2) + \gamma_{t-x}^{4} \\
\text{logit } q(t,x) &= k_t^{(1)} + k_t^{(2)}(x-\bar{x}) + \gamma_{t-x}^{(3)}(x_c - x)
\end{aligned} \tag{1.6}$$

where the functions $\beta_x^{(i)}$, $k_t^{(i)}$ and $\gamma_{t-x}^{(i)}$ are age, period and cohort effects, respectively; $\bar{x}$ is the mean age over the range of ages being used in the analysis; n_a is the number of ages. D'Amato et al. (2014) employed the Lee-Carter model to detect common longevity trends. The specification is:

$$\log m_{xt,i} = \alpha_{x,i} + \beta_{x,i} k_{t,i} + \varepsilon_{xt,i} \tag{1.7}$$

where i denotes the i-th population. Lin et al. (2014) employed the extended Lee-Carter model in Li & Lee (2005) to study the risk management of a defined benefit plan. Bisetti & Favero (2014) applied the Lee-Carter model to measure the impact of longevity risk on pension systems in Italy.

There is the wide application of bootstrap methods for quantifying uncertainty in mortality models. Bootstrap methods have been proposed for interval estimation and error projection purposes. Haberman & Renshaw (2009) proposed three different bootstrap methods to construct confidence intervals for interesting quantities based on the Lee-Carter framework and a generalized linear Poisson model. Li (2010) used parametric bootstrap. D'Amato et al. (2012) proposed sieve bootstrap method based on error terms in $\log m_{xt} = \alpha_x + \beta_x k_t + \varepsilon_{xt}$ where ε_{xt} follows from an AR(∞) model.

Aside from discrete time series models, continuous stochastic differential equation (SDE) is also used for modeling mortality data and the corresponding mortality index. Dahl (2004) selected an extended Cox-Ingersoll-Ross process; Biffis (2005) chose two different specifications for the intensity process; Schrager (2006) proposed an M-factor affine stochastic intensity; Luciano et al. (2008) modeled stochastic mortality for dependent lives.

Some recent literature focuses on hedging longevity risk originated from an increased life expectancy of a population. Milevsky & Promislow (2001) explored the topic of hedging against change in mortality rates using the (put) option, using both discrete and continuous time models for mortality. Cox & Lin (2007) considered natural hedging against mortality rates when insurance companies underwrite both life insurance and annuity products, which might help lower premiums than otherwise similar insurers (without such practice). Li & Hardy (2011) proposed the use by pension funds of a portfolio of q-Forward contracts to hedge against longevity risk.

It is also worth mentioning that there is abundant demography literature attempting to find subtle mortality trends within specific cohorts of a population. For example, Willets (2004) suggested that specific cohorts (people born in certain years) might experience mortality trends different from other cohorts, probably due to a combination of life habit factors or other health factors. Some more papers followed this discussion (see Richards et al., 2006). In this dissertation, we do not further pursue this topic. Instead, we are focusing on the widely applied Lee-Carter model Lee & Carter (1992).

We begin with some discussions of flaws and issues in the Lee-Carter paper, as we have argued in Liu et al. (2019b). An issue with the original Lee-Carter model is the model assumption on $\{k_t\}$ constraint in (1.2): $\sum_{k=1}^{T} k_t = 0$. Part of the Lee-Carter model is the time series model (1.3) characterizing mortality index $\{k_t\}$, which makes the series of $\{k_t\}$ random variables. The constraint that the sum of random k_t's equals to a constant number is unrealistic and too restrictive. For example, if one fits an AR(1) model to $\{k_t\}$, say $k_t = \mu + \phi k_{t-1} + e_t$, assuming the time series is stationary (i.e. $|\phi| < 1$), we must have $\frac{1}{T}\sum_{t=1}^{T} k_t \xrightarrow{p} \frac{\mu}{1-\phi}$ as $T \to \infty$. On the other hand, when $\mu \neq 0$ and $\phi = 1 + \frac{\gamma}{T}$ for some constant $\gamma \in \mathbb{R}$, we must have $\frac{k_T}{T} \xrightarrow{p} \mu \frac{1-e^{\gamma}}{-\gamma}$ as $T \to \infty$. (We can interpret $\frac{1-e^{\gamma}}{\gamma} = 1$ when $\phi = 1$, or $\gamma = 0$.) No matter which case of ϕ, these implications, combined with the constraint $\sum_{k=1}^{T} k_t = 0$, leads to $\mu = 0$ in (1.3), which is too constraining and does not fit real mortality data well. Therefore, a modified model constraint (rather than a constraint of the sum of mortality index) is more appropriate.

Another issue of the Lee-Carter model is that there are no asymptotic results for the derived estimates using singular value decomposition (SVD), so inference uncertainty cannot be quantified. A special case of the Lee-Carter was considered in Leng & Peng (2016) that all the β_x's are equal and subject to constraint $\sum_{x=1}^{M} \beta_x = 1$ (i.e., $\beta_1 = \cdots = \beta_M = \frac{1}{M}$), which implies that the two-step inference method in Lee & Carter (1992), as we have summarized above, does not produce inconsistent estimates, as long as the time series model of (1.3) is not exactly an ARIMA$(0, 1, 0)$ model.

It happens that some papers in actuarial science interpret the original Lee-Carter model in the wrong way. For example, by defining $m_0(x, t)$ as the true central death rates for age x in year t, Dowd et al. (2010), Cairns et al. (2011), Enchev et al. (2017) and some other papers interpreted Lee-Carter model as

$$\begin{gathered} \log m_0(x,t) = \alpha_x + \beta_x k_t \\ k_t = \mu + k_{t-1} + e_t, \quad \sum_{x=1}^{M} \beta_x = 1, \quad \sum_{t=1}^{T} k_t = 0 \end{gathered} \tag{1.8}$$

This interpretation (1.8) is confusing because it basically omits the unexplained error term ε_{xt} for mortality rate $\log m_0(x, t)$, so that mortality rate is random solely due to the randomness of k_t's. Another misinterpretation appeared in Li (2010) and Li et al. (2017b) that treated Lee-Carter model as $\log m(x, t) = \alpha_x + \beta_x k_t$ without the error term $\varepsilon_{x,t}$ in (1.2). This interpretation is problematic because it implies that $\log m(x_1, t)$ and $\log m(x_2, t)$ for different ages x_1 and x_2 are completely dependent as both are determined by the same random variable (mortality index) k_t. That is, central death rates are completely dependent across ages.

Based on these understandings, we conclude that the random error term $\varepsilon_{x,t}$ in (1.2) 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 error $\varepsilon_{x,t}$'s, the two-step inference procedure proposed by Lee & Carter (1992) may be inconsistent in the sense that the resulted estimators do not converge in probability to the true values as T

goes to infinity. More specifically, Leng & Peng (2016) considered a submodel of (1.2) with known β_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 a unit root AR(1) model exactly. So naturally, the research question to be answered is, can we make certain improvements to the original Lee-Carter method to obtain consistent estimators so that they approach true model parameters when there is enough number of years of observation ($T \to \infty$)? Even better, can we obtain model parameter estimates so close to true parameters that we can use them in developing hedging strategies against longevity risk?

As we have seen in literature, these questions are worth studying, but answering them is not easy. In this dissertation, I will take multiple steps in an attempt to answer this question. It first begins with a newly proposed model that changes model assumptions in a novel way, where these changes are based on the lessons that we learned. We have investigated the model (1.8) that the removal of error terms $\varepsilon_{x,t}$ from the Lee-Carter model implies that central death rates are completely dependent across ages. This dependency does not fit mortality data well, so we certainly want to keep error terms $\varepsilon_{x,t}$ in our model. We also studied the assumption $\sum_{k=1}^{T} k_t = 0$ in (1.2), which can lead to spurious results like $\mu = 0$ in (1.3). So we would like to get rid of this $\sum_{k=1}^{T} k_t = 0$ assumption. Nevertheless, to guarantee model identification, we have to introduce another model constraint regarding $\{\alpha_x\}$, which is detailed in the next chapter. Besides, since there are no asymptotic results for the derived estimates based on singular value decomposition, we need to propose a new set of estimators with asymptotic results. Our proposed new estimator utilizes the new model constraint regarding $\{\alpha_x\}$, and has some satisfactory asymptotic results under certain cases, depending on whether the $\{k_t\}$ series follows stationary or unit root process. So it becomes necessary to test whether the $\{k_t\}$ series has a unit root or near unit root. In response, we developed unit root tests for the proposed model based on Leng & Peng (2017). All of these proposed estimators and related results are put into Part 2, which refers to Liu et al.

(2019b).

Since the asymptotic results in Part 2 are contingent on whether or not mortality index $\{k_t\}$ follows unit root time series, an exciting question raises if one can estimate unknown model parameters asymptotically regardless of the property of $\{k_t\}$ (see Liu et al., 2019a). In Part 3 I refer to the paper Liu et al. (2019a) on details regarding the proposed bias-corrected estimation.

I further extend our proposed methods to accommodate two more cases. One of them is extending the modeling of mortality index $\{k_t\}$ to AR(2) process from AR(1) case. This case is addressed by my working paper and is detailed in Part 4. Since different asymptotic results are derived under different scenarios (i.e., k_t follow unit root AR(2) process), unit root test based on AR(2) time series is also supplied in Part 4. Finally, I have considered extending the mortality model and bias-corrected estimation to accommodate mortality data of two populations or two cohorts, based on the understanding of the necessity to model mortality of multiple populations together. The results are presented in the Part 5.

Last but not least, it is crucial to apply these models to actual mortality datasets because even with theoretical results regarding the estimators, we still need to empirically verify their performance, given a reasonable sample (number of years of observation, T, being not too big). So at the end of each part of this dissertation, I present numerical analysis and simulation study results. When estimates are close enough to the "real" parameters used to simulate mortality data in the simulation study, combined with confirmed asymptotic results, we can confidently conclude the performance of our proposed models and estimators. Finally, I conclude the dissertation with some discussions of results in Part 6. All R codes used for data analysis and simulation are provided at the end of this dissertation.

PART 2

STATISTICAL INFERENCE FOR LEE-CARTER MORTALITY MODEL AND CORRESPONDING FORECASTS

This Part is my published paper Liu et al. (2019b), but has been adapted to the format of dissertation.

Let $m(x,t)$ denote the observed central death rate for age (or age group) x in year t, where $x = 1, \ldots, M$ and $t = 1, \ldots, T$. To model the logarithms of the central death rates, Lee & Carter (1992) proposed the following simple linear regression model

$$\log m(x,t) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}, \quad \sum_{x=1}^{M} \beta_x = 1, \quad \sum_{t=1}^{T} k_t = 0, \tag{2.1}$$

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 $\{k_t\}_{t=1}^T$.

As an important task of modeling mortality rates is to forecast future mortality pattern so as to better hedge longevity risk, Lee & Carter (1992) further proposed to model the estimated mortality index by a simple time series model. In practice $\{k_t\}$ is often fitted to an ARIMA(p,d,q) model defined as

$$\left(1 - \sum_{i=1}^{p} \phi_i B^i\right) (1-B)^d k_t = \mu + \left(1 + \sum_{i=1}^{q} \theta_i B^i\right) e_t, \tag{2.2}$$

where e_t's are white noises.

In conclusion, the classic Lee-Carter mortality model proposed by Lee & Carter (1992) is a combination of (2.1) and (2.2), and a proposed two-step inference procedure is to first estimate parameters in (2.1) by the singular value decomposition method, and then to use the estimated k_t's to fit model (2.2). Many papers in actuarial science have claimed that an application of this model with its two-step inference procedure to mortality data prefers a unit root time series model, i.e., $d = 1$ in (2.2).

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 et al. (2002), Li & Lee (2005), Girosi & King (2007), Cairns et al. (2011), D'Amato et al. (2014), Lin et al. (2014), and Bisetti & Favero (2014).

Although the Lee-Carter model has become a benchmark in modeling mortality rates, there are some serious issues on its model assumptions and the proposed two-step inference procedure. First, since $\{k_t\}$ in (2.2) is random, the constraint $\sum_{t=1}^{T} k_t = 0$ in (2.1) becomes unrealistic and restrictive. For example, if one fits an AR(1) model to $\{k_t\}$, say $k_t = \mu + \phi k_{t-1} + e_t$, 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, when $\mu \neq 0$ and $\phi = 1 + \gamma/T$ for some constant $\gamma \in \mathbb{R}$, we have $k_T/T \xrightarrow{p} \mu \frac{1-e^{\gamma}}{-\gamma}$ as $T \to \infty$, where $\frac{1-e^{\gamma}}{-\gamma}$ is interpreted as 1 for $\gamma = 0$. That is, the constraint $\sum_{t=1}^{T} k_t = 0$ in (2.1) basically says μ in (2.2) must be zero. Hence, a modified model without any direct constraint on k_t's is more appropriate. A further difficulty in using the singular value decomposition method for model inference is that no asymptotic results are available for the derived estimators. When all β_x's are the same (i.e., $\beta_1 = \cdots = \beta_M = 1/M$), Leng & Peng (2016) proved that the proposed two-step inference procedure in Lee & Carter (1992) is inconsistent when the model (2.2) is not an ARIMA(0,1,0) model.

It also appears that some papers in actuarial science misunderstand the model. 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 et al. (2017) and others interpreted the Lee-Carter model as

$$\log m_0(x,t) = \alpha_x + \beta_x k_t, \quad k_t = \mu + k_{t-1} + e_t, \quad \sum_{x=1}^{M} \beta_x = 1, \quad \sum_{t=1}^{T} k_t = 0. \tag{2.3}$$

This is confusing because model (2.3) 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 et al. (2017b), 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 (2.1). This is 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 (2.1) 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 & Carter (1992) may be inconsistent in the sense that the resulted estimators do not converge in probability to the true values as T goes to infinity. More specifically, Leng & Peng (2016) considered a submodel of (2.1) with known β_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 & Peng (2017) proposed a way to test whether $\{k_t\}$ follows a unit root AR(2) model. Since Leng & Peng (2016) only considered a submodel of (2.1), it still remains open on whether the inference in Lee & Carter (1992) is consistent in estimating all unknown parameters and forecasting future mortality rates, and how to quantify the inference uncertainty even when $\{k_t\}$ does 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 error based on the Lee-Carter model and its two-step inference.

This Part of dissertation first modifies the classic Lee-Carter model without adding a constraint on k_t's for the sake of model identification. Second by focusing on fitting an AR(1) model to $\{k_t\}$ and assuming that the error sequence $\{(e_t, \varepsilon_{1,t}, \cdots, \varepsilon_{M,t})^\tau, t \ge 1\}$ is α-mixing, defined later, instead of independent random vectors, this Part proposes least squares estimators for the unknown quantities, provides a test for unit root, and derives the asymptotic distributions of the proposed estimators and unit root test when the mortality index $\{k_t\}$ follows a unit root or near unit root AR(1) process. Throughout A^τ denotes the transpose of the matrix or vector A. When the unit root hypothesis cannot be rejected, forecasting future mortality rates is provided too. We refer to section 2.1 for details. Section 2.2 presents a real data analysis and a simulation study. Some conclusions are summarized in section 2.3. All proofs are put in section 2.4.

2.1 Model, Estimation, Unit Root Test and Forecast

Model First we propose to replace (2.1) by

$$\log m(x,t) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}, \quad \sum_{x=1}^{M} \beta_x = 1, \quad \sum_{x=1}^{M} \alpha_x = 0, \tag{2.4}$$

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, for example, Chen & Cox (2009), Chen & Cummins (2010), Kwok et al. (2016), Biffis et al. (2017), Lin et al. (2017), Li et al. (2017a), Wong et al. (2017) and Zhu et al. (2017), this Part of dissertation considers a special case of (2.2):

$$k_t = \mu + \phi k_{t-1} + e_t, \tag{2.5}$$

where e_t's are white noises. Therefore the proposed modified Lee-Carter mortality model is a combination of (2.4) and (2.5), which does not impose any constraint on k_t's for model identification.

Estimation Next we propose a statistical inference for models (2.4) and (2.5). As we argue before, the two-step inference in Lee & Carter (1992) is hard to derive asymptotic results and may lead to inconsistent estimators. Therefore we need a method different from 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 $t = 1, \dots, T$. Then, by noting that $\sum_{x=1}^{M} \alpha_x = 0$ and $\sum_{x=1}^{M} \beta_x = 1$, we have

$$\hat{Z}_t = k_t + \eta_t \quad \text{for} \quad t = 1, \dots, T. \tag{2.6}$$

When $\{k_t\}$ is nonstationary such as unit root (i.e., $\phi = 1$ in (2.5)) or near unit root (i.e., $\phi = 1 + \gamma/T$ for some constant $\gamma \neq 0$ in (2.5)), k_t dominates η_t as t large enough, so $\hat{Z}_t$ behaves like k_t. This motivates us to minimize the following least squares

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

which leads to the least squares estimators for μ and ϕ as

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

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 estimators for α_x and β_x as

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

In order to derive the asymptotic properties of the above least squares estimators, we assume the following regularity conditions for the error sequence $\{(e_t, \varepsilon_{1,t}, \cdots, \varepsilon_{M,t})^\tau, t \geq 1\}$ in (2.4) and (2.5):

- C1) $E(e_t) = 0,\ E(\varepsilon_{x,t}) = 0$ for $t = 1, \ldots, T$ and $x = 1, \ldots, M$;

- C2) there exist $\beta > 2$ and $\delta > 0$ such that $\sup_t E|e_t|^{\beta+\delta} < \infty$ and $\sup_t E|\varepsilon_{x,t}|^{\beta+\delta} < \infty$ for $x = 1, \ldots, M$;

- C3) $\sigma_e^2 = \lim_{T\to\infty} E\{T^{-1}(\sum_{t=1}^T e_t)^2\} \in (0, \infty)$ and $\sigma_x^2 = \lim_{T\to\infty} E\{T^{-1}(\sum_{t=1}^T \varepsilon_{x,T})^2\} \in (0, \infty)$ for $x = 1, \ldots, M$;

- C4) the sequence $\{(e_t, \varepsilon_{1,t}, \cdots, \varepsilon_{M,t})^\tau\}$ is strong mixing with mixing coefficients

$$\alpha_m = \sup_{k\geq 1} \sup_{A\in\mathcal{F}_1^k, B\in\mathcal{F}_{k+m}^\infty} |P(A \cap B) - P(A)P(B)|$$

  such that $\sum_{m=1}^\infty \alpha_m^{1-2/\beta} < \infty$, where $\mathcal{F}_k^{k+m}$ denotes the σ-field generated by

$$\{(e_t, \varepsilon_{1,t}, \cdots, \varepsilon_{M,t})^\tau : k \leq t \leq k + m\}.$$

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

$$\left(\frac{\sum_{t=1}^{[Tr_1]} e_t}{\sigma_e\sqrt{T}}, \frac{\sum_{t=1}^{[Tr_2]} \varepsilon_{1,t}}{\sigma_1\sqrt{T}}, \cdots, \frac{\sum_{t=1}^{[Tr_{M+1}]} \varepsilon_{M,t}}{\sigma_M\sqrt{T}}\right)^\tau \xrightarrow{D} \boldsymbol{W}(\boldsymbol{r}) \text{ in the space } D([0, 1]^{M+1}), \quad (2.7)$$

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, and "$\xrightarrow{D}$" denotes the weak convergence of the associated probability measures. Throughout, we will use "$\xrightarrow{d}$" and "$\xrightarrow{p}$" to denote the convergence in distribution and in probability, respectively. We also use $W_i(r_i)$ to denote the ith marginal distribution of $\boldsymbol{W}(\boldsymbol{r})$ and define

$$W_*(t) = \sum_{i=2}^{M+1} \sigma_{i-1} W_i(t),\ f_\gamma(t) = \frac{1 - \exp(\gamma t)}{-\gamma},\ J_\gamma(t) = \int_0^t \exp\{(t - s)\gamma\}\, dW_1(s) \text{ for } t \in [0, 1],$$

$$\sigma^2 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^T E(e_t^2),\ c_0 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^T E\{\eta_{t-1}(e_t - e_{t-1} + \eta_t - \eta_{t-1})\},$$

$$c_x = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E\{(e_t+\eta_t)(\beta_x\eta_t - \varepsilon_{x,t})\} \text{ for } x = 1,\ldots,M.$$

Theorems below derive the asymptotic distributions of the proposed estimators when $\{k_t\}$ is near unit root, and show that the proposed estimators may be inconsistent when $\{k_t\}$ is stationary.

Theorem 2.1. *Assume (2.4) and (2.5) 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. Then the following convergences are true as $T \to \infty$.*

(i)

$$T^{3/2}(\hat{\phi} - \phi) \xrightarrow{d} \frac{\sigma_e}{\mu} \frac{\int_0^1 f_\gamma(s)\,dW_1(s) - W_1(1)\int_0^1 f_\gamma(s)\,ds}{\int_0^1 f_\gamma^2(s)\,ds - (\int_0^1 f_\gamma(s)\,ds)^2}.$$

(ii)

$$T^{1/2}(\hat{\mu} - \mu) \xrightarrow{d} \frac{\sigma_e W_1(1)\int_0^1 f_\gamma^2(s)\,ds - \sigma_e \int_0^1 f_\gamma(s)\,ds \int_0^1 f_\gamma(t)\,dW_1(t)}{\int_0^1 f_\gamma^2(s)\,ds - (\int_0^1 f_\gamma(s)\,ds)^2}.$$

(iii) *For $x = 1,\ldots,M$*

$$T^{1/2}(\hat{\alpha}_x - \alpha_x) \xrightarrow{d} \beta_x Y_* - Y_x,$$

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 $x = 1,\ldots,M$*

$$T^{3/2}(\hat{\beta}_x - \beta_x) \xrightarrow{d} \frac{\beta_x Z_* - Z_x}{\mu}$$

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

$$Z_x = \frac{\sigma_x W_{x+1}(1)\int_0^1 f_\gamma(s)\,ds - \sigma_x \int_0^1 f_\gamma(s)\,dW_{x+1}(s)}{\int_0^1 f_\gamma^2(s)\,ds - (\int_0^1 f_\gamma(s)\,ds)^2}.$$

Theorem 2.2. *Assume (2.4) and (2.5) hold with conditions C1)–C4), $\mu = 0$, $\phi = 1 + \gamma/T$ for some constant $\gamma \in \mathbb{R}$ and k_0 is a constant. Then the following convergences are true as $T \to \infty$.*

(i)

$$T(\hat{\phi} - \phi) \xrightarrow{d} \frac{\sigma_e^2\{\int_0^1 J_\gamma(t)\,dW_1(t) - W_1(1)\int_0^1 J_\gamma(t)\,dt\} + \frac{1}{2}(\sigma_e^2 - \sigma^2) + c_0}{\sigma_e^2\{\int_0^1 J_\gamma^2(t)\,dt - (\int_0^1 J_\gamma(t)\,dt)^2\}}.$$

(ii)

$$T^{1/2}(\hat{\mu} - \mu) \xrightarrow{d} \frac{\sigma_e^3 W_1(1)\int_0^1 J_\gamma^2(t)\,dt - \sigma_e\int_0^1 J_\gamma(t)\,dt\{\sigma_e^2\int_0^1 J_\gamma(t)\,dW_1(t) + \frac{1}{2}(\sigma_e^2 - \sigma^2) + c_0\}}{\sigma_e^2\{\int_0^1 J_\gamma^2(t)\,dt - (\int_0^1 J_\gamma(t)\,dt)^2\}}.$$

(iii) *For $x = 1, \ldots, M$,*

$$T^{1/2}(\hat{\alpha}_x - \alpha_x) \xrightarrow{d} \frac{\beta_x Y_* - Y_x + c_x\sigma_e\int_0^1 J_\gamma(t)\,dt}{\sigma_e^2\{\int_0^1 J_\gamma^2(t)\,dt - (\int_0^1 J_\gamma(t)\,dt)^2\}}$$

with

$$Y_* = \sigma_e^2\{\int_0^1 J_\gamma(t)\,dt\int_0^1 J_\gamma(t)\,dW_*(t) - W_*(1)\int_0^1 J_\gamma^2(t)\,dt\},$$

$$Y_x = \sigma_e^2\sigma_x\{\int_0^1 J_\gamma(t)\,dt\int_0^1 J_\gamma(t)\,dW_{x+1}(t) - W_{x+1}(1)\int_0^1 J_\gamma^2(t)\,dt\}.$$

(iv) *For $x = 1, \ldots, M$,*

$$T(\hat{\beta}_x - \beta_x) \xrightarrow{d} \frac{\beta_x Z_* - Z_x - c_x}{\sigma_e^2\{\int_0^1 J_\gamma^2(t)\,dt - (\int_0^1 J_\gamma(t)\,dt)^2\}}$$

with

$$Z_* = \sigma_e\{W_*(1)\int_0^1 J_\gamma(t)\,dt - \int_0^1 J_\gamma(t)\,dW_*(t)\},$$

$$Z_x = \sigma_e\sigma_x\{W_{x+1}(1)\int_0^1 J_\gamma(t)\,dt - \int_0^1 J_\gamma(t)\,dW_{x+1}(t)\}.$$

Theorem 2.3. *Assume (2.4) and (2.5) hold with conditions C1)–C4), $|\phi| < 1$ and ϕ is independent of T (i.e., $\{k_t\}$ is stationary). Further assume that the sequence $\{(e_t, \varepsilon_{1,t}, \cdots, \varepsilon_{M,t})^\tau\}$*

is strictly stationary. Then, for $x = 1, \ldots, M$, we have, as $T \to \infty$,

$$\hat{\phi} - \phi \xrightarrow{p} \frac{E(k_1e_2) + (1-\phi^2)E(k_1\eta_2) - \phi E(e_1\eta_1) - \phi E(\eta_1^2) + E(e_2\eta_1) + E(\eta_1\eta_2)}{\frac{E(e_1^2)}{1-\phi^2} + \frac{2\phi E(k_1e_2)}{1-\phi^2} + 2\phi E(k_1\eta_2) + 2E(e_1\eta_1) + E(\eta_1^2)},$$

$$\hat{\mu} - \mu \xrightarrow{p} -\frac{\mu}{1-\phi}\frac{E(k_1e_2) + (1-\phi^2)E(k_1\eta_2) - \phi E(e_1\eta_1) - \phi E(\eta_1^2) + E(e_2\eta_1) + E(\eta_1\eta_2)}{\frac{E(e_1^2)}{1-\phi^2} + \frac{2\phi E(k_1e_2)}{1-\phi^2} + 2\phi E(k_1\eta_2) + 2E(e_1\eta_1) + E(\eta_1^2)},$$

$$\hat{\alpha}_x - \alpha_x \xrightarrow{p} \frac{\mu}{1-\phi}\frac{\beta_x\{\phi E(k_1\eta_2) + E(\eta_1^2) + E(e_1\eta_1)\} - \phi E(k_1\varepsilon_{x,2}) - E(e_1\varepsilon_{x,1}) - E(\eta_1\varepsilon_{x,1})}{\frac{E(e_1^2)}{1-\phi^2} + \frac{2\phi E(k_1e_2)}{1-\phi^2} + 2\phi E(k_1\eta_2) + 2E(e_1\eta_1) + E(\eta_1^2)},$$

$$\hat{\beta}_x - \beta_x \xrightarrow{p} \frac{-\beta_x\{\phi E(k_1\eta_2) + E(\eta_1^2) + E(e_1\eta_1)\} + \phi E(k_1\varepsilon_{x,2}) + E(e_1\varepsilon_{x,1}) + E(\eta_1\varepsilon_{x,1})}{\frac{E(e_1^2)}{1-\phi^2} + \frac{2\phi E(k_1e_2)}{1-\phi^2} + 2\phi E(k_1\eta_2) + 2E(e_1\eta_1) + E(\eta_1^2)}.$$

Remark 2.1. *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 (2.4). When $\{k_t\}$ is near unit root, the asymptotic distributions of the proposed estimators are nonnormal when $\mu = 0$, and are normal with a faster rate of convergence for estimators $\hat{\phi}$ and $\hat{\beta}_x$ when $\mu \neq 0$. When $\{k_t\}$ is stationary and $\mu = 0$, the proposed estimators $\hat{\mu}$ and $\hat{\alpha}_x$ are consistent while $\hat{\phi}$ and $\hat{\beta}_x$ are inconsistent. When $\{k_t\}$ is stationary and $\mu \neq 0$, the proposed estimators are inconsistent.*

Remark 2.2. *The conditions C1)–C4) allow many weakly dependent time series such as finite order ARMA models under very general conditions on the underlying errors. Moreover the α-mixing condition can be replaced by other mixing conditions, for example, ψ-mixing and β-mixing, as long as the central limit theorem and law of large numbers can be employed.*

Remark 2.3. *If we do not add the assumption of strict stationarity in Theorem 2.3 above, the expectations in the right hand sides should be replaced by the corresponding limits of averages. For example, $E(k_1e_2)$ is replaced by $\lim_{T\to\infty} T^{-1}\sum_{t=2}^T k_{t-1}e_t$.*

Unit root test As many applications of the Lee-Carter model simply fit a unit root AR(1) model to the mortality index and Theorem 2.3 above shows that the proposed estimators may be inconsistent when $\{k_t\}$ is stationary, it becomes important/necessary to test $H_0 : \phi = 1$ in (2.5). Under H_0, we have $f_\gamma(s) = s$, so it follows from Theorem 2.1(i) above that

$$T^{3/2}(\hat{\phi}-\phi) \xrightarrow{d} \frac{12\sigma_e}{\mu}\{\frac{W_1(1)}{2} - \int_0^1 W_1(s)\,ds\} \sim N(0, \frac{12\sigma_e^2}{\mu^2}). \tag{2.8}$$

In order to employ the above limiting distribution to test $H_0 : \phi = 1$, one has to estimate σ_e^2. Unfortunately the estimators proposed by Phillips & Perron (1988) are not applicable due to the involved "measurement errors" η_t's and the possible dependence between η_t and e_t. But, when $\{e_t\}$ and $\{\varepsilon_{x,t}\}$ for $x = 1, \ldots, M$ are sequences of independent and identically distributed random variables and all sequences are independent, the estimator s_{Tl}^2 with $l = 1$ in Phillips & Perron (1988) can be employed. That is, for independent errors, a simple estimator for σ_e^2 is

$$\tilde{\sigma}_e^2 = \frac{1}{T-1}\sum_{t=2}^{T}\hat{e}_t^2 + \frac{2}{T-2}\sum_{t=3}^{T}\hat{e}_t\hat{e}_{t-1}. \tag{2.9}$$

As this dissertation deals with dependent errors, we employ the idea of block sample variance estimation in Carlstein (1986) and Politis & Romano (1993).

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

$$\hat{\sigma}_e^2 = L\left\{\frac{1}{T-L}\sum_{i=1}^{T-L}\hat{U}_i^2 - (\frac{1}{T-L}\sum_{j=1}^{T-L}\hat{U}_j)^2\right\}.$$

Theorem 2.4. *Suppose conditions in Theorem 2.1 hold. Further assume $L^{-1} + T^{-1}L \to 0$ as $T \to \infty$. Then $\hat{\sigma}_e^2 \xrightarrow{p} \sigma_e^2$ as $T \to \infty$.*

Based on (2.8) and Theorem 2.4, we reject the null hypothesis $H_0 : \phi = 1$ at level a if $\frac{\hat{\mu}^2 T^3(\hat{\phi}-1)^2}{12\hat{\sigma}_e^2} \geq \chi^2_{1,1-a}$, where $\chi^2_{1,1-a}$ denotes the $(1-a)$-th quantile of the chi-squared distribution with one degree of freedom.

Forecast When the null hypothesis $H_0 : \phi = 1$ is not rejected, we forecast the future mortality rates based on models (2.4) and (2.5) with $\phi = 1$ by

$$\log \widehat{m(x, T+d)} = \hat{\alpha}_x + \hat{\beta}_x\{\hat{Z}_T + d\hat{\mu}\} \quad \text{for} \quad d \geq 1.$$

Note that

$$\begin{aligned} & \log \widehat{m(x,T}+d) - \log m(x,T+d) \\ = \;& \hat{\alpha}_x + \hat{\beta}_x\{\hat{Z}_T + d\hat{\mu}\} - \alpha_x - \beta_x\{d\mu + k_T + \textstyle\sum_{s=1}^{d} e_{T+s}\} - \varepsilon_{x,T+d} \\ = \;& \hat{\alpha}_x - \alpha_x + d\{\hat{\beta}_x\hat{\mu} - \beta_x\mu\} + (\hat{\beta}_x - \beta_x)\hat{Z}_T + \beta_x\eta_T - \beta_x \textstyle\sum_{s=1}^{d} e_{T+s} - \varepsilon_{x,T+d} \\ = \;& \beta_x\eta_T - \beta_x \textstyle\sum_{s=1}^{d} e_{T+s} - \varepsilon_{x,T+d} + o_p(1). \end{aligned} \tag{2.10}$$

In order to quantify the uncertainties of the above forecasts, one has to estimate the distribution function

$$G_d(y) = P(\beta_x\eta_T - \beta_x \sum_{s=1}^{d} e_{T+s} - \varepsilon_{x,T+d} \le y).$$

Unfortunately it seems that $G_d(y)$ can not be estimated nonparametrically without imposing more conditions on $\varepsilon_{x,t}$'s. However, we have

$$\frac{1}{M}\sum_{x=1}^{M} \log \widehat{m(x,T}+d) - \frac{1}{M}\sum_{x=1}^{M} \log m(x,T+d) = \frac{1}{M}\{\eta_T - \sum_{s=1}^{d} e_{T+s} - \eta_{T+d}\} + o_p(1)$$

and the distribution function of

$$H_d(y) = P\Big(\frac{1}{M}(\eta_T - \sum_{s=1}^{d} e_{T+s} - \eta_{T+d}) \le y\Big)$$

can be estimated nonparametrically by

$$\hat{H}_d(y) = \frac{1}{T-d}\sum_{t=1}^{T-d} I(-\frac{1}{M}\sum_{s=1}^{d} \hat{e}_{t+s} \le y).$$

Hence, by defining

$$c_{l,a} = \sup\{y : \hat{H}_d(y) \le a/2\} \text{ and } c_{u,a} = \sup\{y : \hat{H}_d(y) \le 1 - a/2\},$$

an interval forecast for $\frac{1}{M}\sum_{x=1}^{M}\log m(x,T+d)$ with level a is obtained as

$$I_a = \left(\frac{1}{M}\sum_{x=1}^{M}\log \widehat{m(x,T}+d) - c_{u,a}, \quad \frac{1}{M}\sum_{x=1}^{M}\log \widehat{m(x,T}+d) - c_{l,a}\right).$$

Theorem 2.5. *Suppose conditions in Theorem 2.1 hold and $\phi = 1$. Then for any fixed integer $d \geq 1$ and $a \in (0,1)$,*

$$P(\frac{1}{M}\sum_{x=1}^{M}\log m(x,T+d) \in I_a) \to a \text{ as } T \to \infty.$$

2.2 Data Analysis and Simulation

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. This gives $M = 10$ and $T = 83$.

First, to implement the classic Lee-Carter model, we employ the statistical R package 'demography' to obtain estimates for α_x's, β_x's, k_t's, and then use the obtained estimates for k_t's to fit model (2.5) by using 'lm' in the statistical software R. We report the estimates for α_x's, β_x's, μ and ϕ in Tables 2.1, 2.2 and 2.3 for the female, male and combined mortality rates, respectively. As the asymptotic property for the estimate of ϕ 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 $H_0 : \phi = 1$ because of the employed two-step inference.

Second, we apply our proposed inference to fit models (2.4) and (2.5) to the female, male and combined mortality rates. Again we use 'lm' to obtain our proposed least squares

Table (2.1) Female mortality rates for ages 25 to 74. *Parameter estimates are obtained from fitting models (2.1) and (2.5) based on the two-step inference in Lee and Carter (1992).*

x	1	2	3	4	5	6	7	8	9	10
$\hat{\alpha}_x$	-7.011	-6.736	-6.377	-5.984	-5.572	-5.159	-4.770	-4.348	-3.929	-3.465
$\hat{\beta}_x$	0.135	0.128	0.119	0.106	0.095	0.090	0.083	0.080	0.081	0.083

	Estimate	Standard Error
$\hat{\mu}$	-0.157	0.022
$\hat{\phi}$	0.975	0.006

Table (2.2) Male mortality rates for ages 25 to 74. *Parameter estimates are obtained from fitting models (2.1) and (2.5) based on the two-step inference in Lee and Carter (1992).*

x	1	2	3	4	5	6	7	8	9	10
$\hat{\alpha}_x$	-6.262	-6.125	-5.844	-5.472	-5.048	-4.612	-4.204	-3.798	-3.415	-3.015
$\hat{\beta}_x$	0.088	0.094	0.106	0.109	0.108	0.108	0.103	0.099	0.096	0.090

	Estimate	Standard Error
$\hat{\mu}$	-0.118	0.024
$\hat{\phi}$	0.994	0.008

Table (2.3) Combined mortality rates for ages 25 to 74. *Parameter estimates are obtained from fitting models (2.1) and (2.5) based on the two-step inference in Lee and Carter (1992).*

x	1	2	3	4	5	6	7	8	9	10
$\hat{\alpha}_x$	-6.562	-6.381	-6.075	-5.697	-5.279	-4.853	-4.454	-4.045	-3.656	-3.237
$\hat{\beta}_x$	0.106	0.109	0.112	0.108	0.103	0.101	0.094	0.090	0.089	0.088

	Estimate	Standard Error
$\hat{\mu}$	-0.135	0.023
$\hat{\phi}$	0.983	0.007

estimates and report the estimates for α_x's, β_x's, μ and ϕ in Tables 2.4, 2.5 and 2.6 for the female, male and combined mortality rates, respectively. As before, the standard errors obtained from 'lm' is inaccurate since it ignores the involved η_t's and so one can not conclude whether $\phi = 1$ or not from these three tables. Although the estimate for ϕ obtained from the new method is similar to that obtained from the Lee-Carter method, estimates for μ are quite different for both methods since the new method does not assume $\sum_{t=1}^{T} k_t = 0$.

Third, we apply our proposed unit root test to the female, male and combined mortality

Table (2.4) Female mortality rates for ages 25 to 74. *Parameter estimates are obtained from fitting models (2.4) and (2.5) based on the proposed least squares estimation.*

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

	Estimate	Standard Error
$\hat{\mu}$	-1.389	0.290
$\hat{\phi}$	0.977	0.005

Table (2.5) Male mortality rates for ages 25 to 74. *Parameter estimates are obtained from fitting models (2.4) and (2.5) based on the proposed least squares estimation.*

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

	Estimate	Standard Error
$\hat{\mu}$	-0.441	0.399
$\hat{\phi}$	0.993	0.008

Table (2.6) Combined mortality rates for ages 25 to 74. *Parameter estimates are obtained from fitting models (2.4) and (2.5) based on the proposed least squares estimation.*

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

	Estimate	Standard Error
$\hat{\mu}$	-0.906	0.343
$\hat{\phi}$	0.985	0.007

rates, where we use $\hat{\sigma}_e^2$ with $L = 0.5\sqrt{T}, \sqrt{T}, 2\sqrt{T}$ and $\tilde{\sigma}_e^2$ (denoted by $L = *$) given in (2.9). Note that (2.8) 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 Pvalues are reported in Tables 2.7, 2.8 and 2.9 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 (2.7) Female mortality rates for ages 25 to 74. *Variance estimates, test statistics and Pvalues are reported for* $L = \lfloor \frac{1}{2}\sqrt{T} \rfloor, \lfloor \sqrt{T} \rfloor, \lfloor 2\sqrt{T} \rfloor$, *where* $'L = *'$ *denotes* $\tilde{\sigma}_e^2$.

L	$\hat{\sigma}_e^2$	Test statistic	Pvalue
$\lfloor \frac{1}{2}\sqrt{T} \rfloor$	0.042	75.938	2.927e-18
$\lfloor \sqrt{T} \rfloor$	0.052	61.378	4.709e-15
$\lfloor 2\sqrt{T} \rfloor$	0.038	85.207	2.687e-20
*	0.047	68.809	1.085e-16

Table (2.8) Male mortality rates for ages 25 to 74. *Variance estimates, test statistics and Pvalues are reported for* $L = \lfloor \frac{1}{2}\sqrt{T} \rfloor, \lfloor \sqrt{T} \rfloor, \lfloor 2\sqrt{T} \rfloor$, *where* $'L = *'$ *denotes* $\tilde{\sigma}_e^2$.

L	$\hat{\sigma}_e^2$	Test statistic	Pvalue
$\lfloor \frac{1}{2}\sqrt{T} \rfloor$	0.064	0.798	0.372
$\lfloor \sqrt{T} \rfloor$	0.073	0.692	0.405
$\lfloor 2\sqrt{T} \rfloor$	0.064	0.793	0.373
*	0.073	0.699	0.403

Table (2.9) Combined mortality rates for ages 25 to 74. *Variance estimates, test statistics and Pvalues are reported for* $L = \lfloor \frac{1}{2}\sqrt{T} \rfloor, \lfloor \sqrt{T} \rfloor, \lfloor 2\sqrt{T} \rfloor$, *where* $'L = *'$ *denotes* $\tilde{\sigma}_e^2$.

L	$\hat{\sigma}_e^2$	Test statistic	Pvalue
$\lfloor \frac{1}{2}\sqrt{T} \rfloor$	0.051	12.229	4.704e-4
$\lfloor \sqrt{T} \rfloor$	0.061	10.289	1.338e-3
$\lfloor 2\sqrt{T} \rfloor$	0.049	12.727	3.605e-4
*	0.058	10.789	1.021e-3

Finally, we examine the robustness of the above conclusion on the unit root hypothesis for the mortality index by rerunning the above unit root test for the male, female and combined population between 1 and 89 years old. Results are reported in Tables 2.10 – 2.12, which reach the same conclusion as that for the populations between 25 and 74 years old.

Table (2.10) Female mortality rates for ages 1 to 89. *Variance estimates, test statistics and Pvalues are reported for* $L = \lfloor \frac{1}{2}\sqrt{T} \rfloor, \lfloor \sqrt{T} \rfloor, \lfloor 2\sqrt{T} \rfloor$, *where* $'L = *'$ *denotes* $\tilde{\sigma}_e^2$.

L	$\hat{\sigma}_e^2$	Test statistic	Pvalue
$\lfloor \frac{1}{2}\sqrt{T} \rfloor$	0.165	60.684	6.703e-15
$\lfloor \sqrt{T} \rfloor$	0.197	50.802	1.021e-12
$\lfloor 2\sqrt{T} \rfloor$	0.169	59.118	1.485e-14
$*$	0.177	56.580	5.395e-14

Table (2.11) Male mortality rates for ages 1 to 89. *Variance estimates, test statistics and Pvalues are reported for* $L = \lfloor \frac{1}{2}\sqrt{T} \rfloor, \lfloor \sqrt{T} \rfloor, \lfloor 2\sqrt{T} \rfloor$, *where* $'L = *'$ *denotes* $\tilde{\sigma}_e^2$.

L	$\hat{\sigma}_e^2$	Test statistic	Pvalue
$\lfloor \frac{1}{2}\sqrt{T} \rfloor$	0.221	1.112	0.292
$\lfloor \sqrt{T} \rfloor$	0.244	1.005	0.316
$\lfloor 2\sqrt{T} \rfloor$	0.250	0.982	0.322
$*$	0.248	0.991	0.320

Table (2.12) Combined mortality rates for ages 1 to 89. *Variance estimates, test statistics and Pvalues are reported for* $L = \lfloor \frac{1}{2}\sqrt{T} \rfloor, \lfloor \sqrt{T} \rfloor, \lfloor 2\sqrt{T} \rfloor$, *where* $'L = *'$ *denotes* $\tilde{\sigma}_e^2$.

L	$\hat{\sigma}_e^2$	Test statistic	Pvalue
$\lfloor \frac{1}{2}\sqrt{T} \rfloor$	0.186	12.376	4.349e-4
$\lfloor \sqrt{T} \rfloor$	0.218	10.602	1.129e-3
$\lfloor 2\sqrt{T} \rfloor$	0.209	11.038	8.925e-4
$*$	0.202	11.406	7.322e-4

Simulation Study To examine the finite sample performance of the proposed estimators and unit root test, we consider models (2.4) and (2.5) with $M = 10$, α_x's, β_x's, μ being the estimates obtained from the female mortality rates in Table 2.4.

We assume $\varepsilon_{x,t}$'s are independent random variables with $N(0, \sigma_e^2/M)$, e_t's are independent random variables with $N(0, \sigma_e^2)$, and $\varepsilon_{x,t}$'s are independent of e_t's. We take σ_e^2 as $\tilde{\sigma}_e^2$ given in Table 2.7, i.e., the value with $L = *$. We draw $10,000$ random samples from models (2.4) and (2.5) with sample size $T = 80$ and 150, and consider $\phi = 1$.

First we compute the proposed estimators for α_x's, β_x's, μ and ϕ under the above settings and report the means and standard deviations of these estimators in Tables 2.13 and 2.14, 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 $\hat{\sigma}_e^2$ with $L = 0.5\sqrt{T}, \sqrt{T}, 2\sqrt{T}$, $\tilde{\sigma}_e^2$ denoted by $L = *$, and the true value σ_e^2 denoted by $L = **$ to compute the test statistic. Variance estimators and empirical sizes of the proposed unit root test are reported in the lower panel of Tables 2.13 and 2.14, which show that the size tends to be larger than the nominal level, the choice of L has an impact on the test, and the size becomes accurate as T is larger. We also find that the proposed test has a nontrivial power when $\phi = 1 - 2/T$, which is not reported here.

2.3 Conclusions

After articulating the issues on model assumptions, statistical inference, and existing misunderstandings of the classic Lee-Carter mortality model, this Part proposes a modified Lee-Carter model with no condition imposed on the unobserved mortality index for model identification. Further least squares estimators are proposed to estimate all unknown parameters, a unit root test is provided to test whether the mortality index follows a unit root AR(1) process, and the asymptotic distributions of the proposed estimators and unit root test are derived when the mortality index follows a unit root or near unit root process and errors satisfy some α-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 follows a unit root process. Forecasting future mortality rates is discussed too when the unit root hypothesis is not rejected. Some interesting future projects are i) to find unified methods for estimating parameters and forecasting future mortality rates regardless of whether the mortality index is stationary or near unit root or unit root, ii) to generalize the AR(1) model for the unobserved mortality index to an ARIMA(p,d,q) model, and iii) to show whether and how forecast errors can be quantified nonparametrically.

Table (2.13) $T = 80$. *The upper and middle panels report the means and standard errors in brackets for the new estimators based on models (2.4) and (2.5). The lower panel reports the mean and standard error in brackets for estimators for σ_e^2 and the size of the proposed unit root test, where $'L = *'$ and $'L = **'$ denote the test by using $\tilde{\sigma}_e^2$ and the true value σ_e^2, respectively.*

x	1	2	3	4	5	6	7	8	9	10
$\hat{\alpha}_x$	0.172	0.055	-0.022	-0.344	-0.474	-0.327	-0.337	-0.067	0.384	0.959
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
$\hat{\beta}_x$	0.135	0.127	0.119	0.106	0.096	0.091	0.083	0.080	0.081	0.083
	(2.308e-4)	(2.259e-4)	(2.272e-4)	(2.269e-4)	(2.274e-4)	(2.266e-4)	(2.258e-4)	(2.272e-4)	(2.271e-4)	(2.269e-4)

	μ	ϕ
True value	-1.389	1.000
Estimator	-1.393 (0.051)	1.000 (7.969e-4)

L	$\lfloor \frac{1}{2}\sqrt{T} \rfloor$	$\lfloor \sqrt{T} \rfloor$	$\lfloor 2\sqrt{T} \rfloor$	$*$	$**$
True value of σ_e^2	0.047	0.047	0.047	0.047	0.047
$\hat{\sigma}_e^2$	0.065 (0.014)	0.049 (0.016)	0.034 (0.018)	0.044 (0.024)	0.047
Unit root test size at 5%	3.91%	7.65%	14.87%	10.95%	6.70%
Unit root test size at 10%	7.79%	12.84%	21.60%	16.26%	12.10%

Table (2.14) $T = 150$. *The upper and middle panels report the means and standard errors in brackets for the new estimators based on models (2.4) and (2.5). The lower panel reports the mean and standard error in brackets for estimators for* σ_e^2 *and the size of the proposed unit root test, where* $'L = *'$ *and* $'L = **'$ *denote the test by using* $\tilde{\sigma}_e^2$ *and the true value* σ_e^2*, respectively.*

x	1	2	3	4	5	6	7	8	9	10
$\hat{\alpha}_x$	0.172	0.055	-0.022	-0.343	-0.474	-0.327	-0.337	-0.067	0.384	0.959
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
$\hat{\beta}_x$	0.135	0.127	0.119	0.106	0.096	0.091	0.083	0.080	0.081	0.083
	(8.873e-5)	(8.885e-5)	(8.818e-5)	(8.857e-5)	(8.819e-5)	(8.814e-5)	(8.891e-5)	(8.810e-5)	(8.829e-5)	(8.891e-5)

	μ	ϕ
True value	-1.389	1.000
Estimator	-1.391 (0.036)	1.000 (3.027e-4)

L	$\lfloor \frac{1}{2}\sqrt{T} \rfloor$	$\lfloor \sqrt{T} \rfloor$	$\lfloor 2\sqrt{T} \rfloor$	*	**
True value of σ_e^2	0.047	0.047	0.047	0.047	0.047
$\hat{\sigma}_e^2$	0.058 (0.011)	0.047 (0.014)	0.037 (0.017)	0.045 (0.018)	0.047
Unit root test size at 5%	3.93%	6.80%	12.22%	8.82%	5.70%
Unit root test size at 10%	8.14%	12.46%	19.02%	14.39%	11.08%

2.4 Proofs

Proof of Theorem 2.1. Note that

$$k_t = \mu + \phi k_{t-1} + e_t = \mu(\sum_{j=0}^{t-1} \phi^j) + \phi^t k_0 + \sum_{i=1}^{t} \phi^{t-i} e_i.$$

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

$$\begin{cases} T^{-2} \sum_{t=1}^{T} \tilde{k}_t^2 \xrightarrow{d} \sigma_e^2 \int_0^1 J_\gamma^2(s)\, ds, \\ T^{-3/2} \sum_{t=1}^{T} \tilde{k}_t \xrightarrow{d} \sigma_e \int_0^1 J_\gamma(s)\, ds, \\ T^{-1} \sum_{t=1}^{T} \tilde{k}_{t-1} e_t \xrightarrow{d} \sigma_e^2 \int_0^1 J_\gamma(s)\, dW_1(s) + \frac{1}{2}(\sigma_e^2 - \sigma^2). \end{cases} \tag{2.11}$$

As $T \to \infty$, it is easy to show that

$$\begin{cases} T^{-2} \sum_{t=2}^{T} \left(\sum_{j=0}^{t-2} \phi^j\right) \to \int_0^1 f_\gamma(s)\, ds, \\ T^{-3} \sum_{t=2}^{T} \left(\sum_{j=0}^{t-2} \phi^j\right)^2 \to \int_0^1 f_\gamma^2(s)\, ds, \\ T^{-3/2} \sum_{t=2}^{T} \left(\sum_{j=0}^{t-2} \phi^j\right) e_t \xrightarrow{d} \sigma_e \int_0^1 f_\gamma(s)\, dW_1(s), \\ T^{-5/2} \sum_{t=2}^{T} \left(\sum_{j=0}^{t-2} \phi^j\right)^2 e_t \xrightarrow{d} \sigma_e \int_0^1 f_\gamma^2(s)\, dW_1(s) \end{cases} \tag{2.12}$$

and

$$T^{-5/2} \sum_{t=2}^{T} (\sum_{j=0}^{t-2} \phi^j) \tilde{k}_{t-1} \xrightarrow{d} \sigma_e \int_0^1 f_\gamma(s) J_\gamma(s)\, ds. \tag{2.13}$$

It follows from (2.11), (2.12) and (2.13) that

$$\begin{aligned} \sum_{t=2}^{T} \hat{Z}_{t-1} &= \sum_{t=2}^{T} k_{t-1} + \sum_{t=2}^{T} \eta_{t-1} \\ &= \mu \sum_{t=2}^{T} \left(\sum_{j=0}^{t-2} \phi^j\right) + \sum_{t=2}^{T} \phi^{t-1} k_0 + \sum_{t=2}^{T} \tilde{k}_{t-1} + o_p(T) \\ &= \mu \sum_{t=2}^{T} \left(\sum_{j=0}^{t-2} \phi^j\right) + o_p(T^2) \end{aligned}$$

and

$$\begin{aligned}\sum_{t=2}^{T}\hat{Z}_{t-1}^2 &= \sum_{t=2}^{T}k_{t-1}^2+2\sum_{t=2}^{T}k_{t-1}\eta_{t-1}+\sum_{t=2}^{T}\eta_{t-1}^2\\ &= \sum_{t=2}^{T}\left(\mu\sum_{j=0}^{t-2}\phi^j+\phi^{t-1}k_0+\tilde{k}_{t-1}\right)^2\\ &\quad +2\sum_{t=2}^{T}\eta_{t-1}\left(\mu\sum_{j=0}^{t-2}\phi^j+\phi^{t-1}k_0+\tilde{k}_{t-1}\right)+\sum_{t=2}^{T}\eta_{t-1}^2\\ &= \mu^2\sum_{t=2}^{T}\left(\sum_{j=0}^{t-2}\phi^j\right)^2+o_p(T^3),\end{aligned}$$

implying that

$$\begin{cases} T^{-4}\left\{(T-1)\sum_{t=2}^{T}\hat{Z}_{t-1}^2-(\sum_{t=2}^{T}\hat{Z}_{t-1})^2\right\}\xrightarrow{p}\mu^2\left\{\int_0^1 f_\gamma^2(s)\,ds-(\int_0^1 f_\gamma(s)\,ds)^2\right\}, \\ T^{-4}\left\{T\sum_{t=1}^{T}\hat{Z}_{t}^2-(\sum_{t=1}^{T}\hat{Z}_{t})^2\right\}\xrightarrow{p}\mu^2\left\{\int_0^1 f_\gamma^2(s)\,ds-(\int_0^1 f_\gamma(s)\,ds)^2\right\}. \end{cases} \tag{2.14}$$

(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 conditions of Theorem 2.1, an application of the law of large numbers for an α-mixing sequence (see McLeish (1975)) and Hölder inequality implies that

$$\frac{\sum_{t=1}^{T}\eta_t}{T}=o_p(1),\ \frac{\sum_{t=1}^{T}\eta_t^2}{T}=O_p(1),\ \{\frac{\sum_{t=2}^{T}\eta_t\eta_{t-1}}{T}\}^2\le\{\frac{\sum_{t=2}^{T}\eta_t^2}{T}\}\{\frac{\sum_{t=2}^{T}\eta_{t-1}^2}{T}\}=O_p(1). \tag{2.15}$$

Note that

$$\begin{aligned} &\sum_{t=2}^{T}\tilde{k}_{t-1}(\eta_t-\phi\eta_{t-1})\\ =\ &\sum_{t=2}^{T}\tilde{k}_{t-1}\eta_t-\phi\sum_{t=1}^{T-1}\tilde{k}_t\eta_t\\ =\ &\sum_{t=2}^{T}\tilde{k}_{t-1}\eta_t-\phi\sum_{t=1}^{T-1}(\phi\tilde{k}_{t-1}+e_t)\eta_t\\ =\ &(1-\phi^2)\sum_{t=2}^{T-1}\tilde{k}_{t-1}\eta_t+\tilde{k}_{T-1}\eta_T-\phi^2\tilde{k}_0\eta_1-\phi\sum_{t=1}^{T-1}e_t\eta_t\\ =\ &O_p(T^{1/2})\end{aligned} \tag{2.16}$$

and

$$\sum_{t=2}^{T}(\sum_{j=0}^{t-2}\phi^j)(\eta_t-\phi\eta_{t-1})=O_p(T). \tag{2.17}$$

By (2.11), (2.15), (2.16) and (2.17), we have

$$
\begin{aligned}
& (T-1)\sum_{t=2}^T \hat{Z}_{t-1}(\hat{Z}_t - \phi\hat{Z}_{t-1}) \\
= \ & (T-1)\sum_{t=2}^T (k_{t-1}+\eta_{t-1})(\mu + e_t + \eta_t - \phi\eta_{t-1}) \\
= \ & (T-1)\sum_{t=2}^T (\mu\sum_{j=0}^{t-2}\phi^j + \phi^{t-1}k_0 + \tilde{k}_{t-1} + \eta_{t-1})(\mu + e_t + \eta_t - \phi\eta_{t-1}) \\
= \ & (T-1)\mu^2\sum_{t=2}^T \left(\sum_{j=0}^{t-2}\phi^j\right) + (T-1)\mu\sum_{t=2}^T \left(\sum_{j=0}^{t-2}\phi^j\right) e_t \\
& +(T-1)\mu\sum_{t=2}^T \tilde{k}_{t-1} + O_p(T^2)
\end{aligned}
$$

and

$$
\begin{aligned}
& \sum_{t=2}^T \hat{Z}_{t-1}\sum_{s=2}^T (\hat{Z}_s - \phi\hat{Z}_{s-1}) \\
= \ & \left(\mu\sum_{t=2}^T\sum_{j=0}^{t-2}\phi^j + k_0\sum_{t=2}^T\phi^{t-1} + \sum_{t=2}^T\tilde{k}_{t-1} + \sum_{t=2}^T\eta_{t-1}\right)\times \\
& \left(\mu(T-1) + \sum_{s=2}^T e_s + \sum_{s=2}^T\eta_s - \phi\sum_{s=2}^T\eta_{s-1}\right) \\
= \ & (T-1)\mu^2\sum_{t=2}^T\left(\sum_{j=0}^{t-2}\phi^j\right) + \mu\left(\sum_{t=2}^T\sum_{j=0}^{t-2}\phi^j\right)\sum_{s=2}^T e_s \\
& +(T-1)\mu\sum_{t=2}^T\tilde{k}_{t-1} + O_p(T^2),
\end{aligned}
$$

implying that

$$
\begin{aligned}
& (T-1)\sum_{t=2}^T \hat{Z}_{t-1}(\hat{Z}_t - \phi\hat{Z}_{t-1}) - \sum_{t=2}^T\hat{Z}_{t-1}\sum_{s=2}^T(\hat{Z}_s - \phi\hat{Z}_{s-1}) \\
= \ & (T-1)\mu\sum_{t=2}^T\left(\sum_{j=0}^{t-2}\phi^j\right)e_t - \mu\left(\sum_{t=2}^T\sum_{j=0}^{t-2}\phi^j\right)\sum_{s=2}^T e_s + O_p(T^2).
\end{aligned}
\tag{2.18}
$$

Hence, by (2.14) and (2.18),

$$
T^{3/2}(\hat{\phi}-\phi) \xrightarrow{d} \frac{\sigma_e}{\mu}\frac{\int_0^1 f_\gamma(s)\,dW_1(s) - W_1(1)\int_0^1 f_\gamma(s)\,ds}{\int_0^1 f_\gamma^2(s)\,ds - (\int_0^1 f_\gamma(s)\,ds)^2}.
$$

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

$$
\begin{aligned}
& \sum_{s=2}^T\hat{Z}_s\sum_{t=2}^T\hat{Z}_{t-1}^2 - \sum_{s=2}^T\hat{Z}_{s-1}\sum_{t=2}^T\hat{Z}_t\hat{Z}_{t-1} - (T-1)\mu\sum_{t=2}^T\hat{Z}_{t-1}^2 + \mu(\sum_{t=2}^T\hat{Z}_{t-1})^2 \\
= \ & \sum_{s=2}^T\hat{Z}_s\sum_{t=2}^T\hat{Z}_{t-1}^2 - \sum_{s=2}^T\hat{Z}_{s-1}\sum_{t=2}^T\hat{Z}_{t-1}^2 - \sum_{s=2}^T\hat{Z}_{s-1}\sum_{t=2}^T(\hat{Z}_t - \hat{Z}_{t-1})\hat{Z}_{t-1} \\
& -(T-1)\mu\sum_{t=2}^T\hat{Z}_{t-1}^2 + \mu(\sum_{t=2}^T\hat{Z}_{t-1})^2 \\
= \ & \sum_{t=2}^T\hat{Z}_{t-1}^2\sum_{s=2}^T(\hat{Z}_s - \hat{Z}_{s-1} - \mu) - \sum_{s=2}^T\hat{Z}_{s-1}\sum_{t=2}^T\hat{Z}_{t-1}(\hat{Z}_t - \hat{Z}_{t-1} - \mu) \\
= \ & I_1 - I_2.
\end{aligned}
$$

Using (2.11), (2.12), (2.13), (2.16) and (2.17), we have

$$
\begin{aligned}
I_1 &= \{\textstyle\sum_{t=2}^{T} \hat{Z}_{t-1}^2\}\{(\phi-1)\sum_{s=2}^{T} k_{s-1} + \sum_{s=2}^{T} e_s + \eta_T - \eta_0\} \\
&= \{\textstyle\sum_{t=2}^{T}(\mu\sum_{j=0}^{t-2}\phi^j + \phi^{t-1}k_0 + \tilde{k}_{t-1} + \eta_{t-1})^2\}\times \\
&\quad \{(\phi-1)(\mu\textstyle\sum_{t=2}^{T}\sum_{j=0}^{t-2}\phi^j + \sum_{t=2}^{T}\tilde{k}_{t-1} + k_0\sum_{t=2}^{T}\phi^{t-1}) + \sum_{t=2}^{T} e_t + \eta_T - \eta_0\} \\
&= \{\mu^2\textstyle\sum_{t=2}^{T}(\sum_{j=0}^{t-2}\phi^j)^2 + 2\mu\sum_{t=2}^{T}(\sum_{j=0}^{t-2}\phi^j)\tilde{k}_{t-1} + O_p(T^2)\}\times \\
&\quad \{(\phi-1)\mu\textstyle\sum_{t=2}^{T}(\sum_{j=0}^{t-2}\phi^j) + (\phi-1)\sum_{t=2}^{T}\tilde{k}_{t-1} + \sum_{t=2}^{T} e_t + O_p(1)\} \\
&= \mu^3(\phi-1)\textstyle\sum_{t=2}^{T}\left(\sum_{j=0}^{t-2}\phi^j\right)^2\sum_{s=2}^{T}\left(\sum_{j=0}^{s-2}\phi^j\right) \\
&\quad +\mu^2(\phi-1)\textstyle\sum_{t=2}^{T}\left(\sum_{j=0}^{t-2}\phi^j\right)^2\sum_{s=2}^{T}\tilde{k}_{s-1} \\
&\quad +2\mu^2(\phi-1)\textstyle\sum_{t=2}^{T}\left(\sum_{j=0}^{t-2}\phi^j\right)\sum_{s=2}^{T}\left(\sum_{j=0}^{s-2}\phi^j\right)\tilde{k}_{s-1} \\
&\quad +\mu^2\textstyle\sum_{t=2}^{T}\left(\sum_{j=0}^{t-2}\phi^j\right)^2\sum_{s=2}^{T} e_s + O_p(T^3)
\end{aligned}
$$

and

$$
\begin{aligned}
I_2 &= \{\textstyle\sum_{s=2}^{T}\hat{Z}_{s-1}\}\sum_{t=2}^{T}\{(k_{t-1}+\eta_{t-1})((\phi-1)k_{t-1} + e_t + \eta_t - \eta_{t-1})\} \\
&= \{\mu\textstyle\sum_{t=2}^{T}\sum_{j=0}^{t-2}\phi^j + \sum_{t=2}^{T}\tilde{k}_{t-1} + O_p(T)\}\sum_{t=2}^{T}\{(\mu\sum_{j=0}^{t-2}\phi^j + \phi^{t-1}k_0 + \tilde{k}_{t-1} + \eta_{t-1})\times \\
&\quad (\phi-1)\mu\textstyle\sum_{j=0}^{t-2}\phi^j + (\phi-1)\phi^{t-1}k_0 + (\phi-1)\tilde{k}_{t-1} + e_t + \eta_t - \eta_{t-1})\} \\
&= \{\mu\textstyle\sum_{t=2}^{T}\sum_{j=0}^{t-2}\phi^j + \sum_{t=2}^{T}\tilde{k}_{t-1} + O_p(T)\}\times \\
&\quad \{\mu^2(\phi-1)\textstyle\sum_{t=2}^{T}(\sum_{j=0}^{t-2}\phi^j)^2 + 2\mu(\phi-1)\sum_{t=2}^{T}(\sum_{j=0}^{t-2}\phi^j)\tilde{k}_{t-1} + \mu\sum_{t=2}^{T}(\sum_{j=0}^{t-2}\phi^j)e_t + O_p(T)\} \\
&= \mu^3(\phi-1)\textstyle\sum_{t=2}^{T}\left(\sum_{j=0}^{t-2}\phi^j\right)^2\sum_{s=2}^{T}\left(\sum_{j=0}^{s-2}\phi^j\right) \\
&\quad +\mu^2(\phi-1)\textstyle\sum_{t=2}^{T}\left(\sum_{j=0}^{t-2}\phi^j\right)^2\sum_{s=2}^{T}\tilde{k}_{s-1} \\
&\quad +2\mu^2(\phi-1)\textstyle\sum_{s=2}^{T}\left(\sum_{j=0}^{s-2}\phi^j\right)\sum_{t=2}^{T}\left(\sum_{j=0}^{t-2}\phi^j\right)\tilde{k}_{t-1} \\
&\quad +\mu^2\textstyle\sum_{s=2}^{T}\left(\sum_{j=0}^{s-2}\phi^j\right)\sum_{t=2}^{T}\left(\sum_{j=0}^{t-2}\phi^j\right)e_t + O_p(T^3),
\end{aligned}
$$

i.e.,

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

which implies that

$$
T^{1/2}(\hat{\mu}-\mu) \xrightarrow{d} \frac{\sigma_e W_1(1)\int_0^1 f_\gamma^2(s)\,ds - \sigma_e\int_0^1 f_\gamma(s)\,ds\int_0^1 f_\gamma(t)\,dW_1(t)}{\int_0^1 f_\gamma^2(s)\,ds - (\int_0^1 f_\gamma(s)\,ds)^2}
$$

by using (2.14).

(iii) Similar to the proof of (2.12), it follows from (2.7) that

$$\begin{cases} T^{-3/2}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\phi^j\right)\varepsilon_{x,t} \xrightarrow{d} \sigma_x\int_0^1 f_\gamma(s)\,dW_{x+1}(s), \\ T^{-3/2}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\phi^j\right)\eta_t \xrightarrow{d} \int_0^1 f_\gamma(s)\,dW_*(s). \end{cases} \tag{2.19}$$

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

$$\begin{aligned} & \sum_{t=1}^{T}\hat{Z}_t^2\sum_{s=1}^{T}\log m(x,s) - \sum_{t=1}^{T}\hat{Z}_t\sum_{s=1}^{T}\log m(x,s)\hat{Z}_s \\ & -\alpha_x T\sum_{t=1}^{T}\hat{Z}_t^2 + \alpha_x(\sum_{t=1}^{T}\hat{Z}_t)^2 \\ = & \sum_{t=1}^{T}\hat{Z}_t^2\sum_{s=1}^{T}\{\log m(x,s)-\alpha_x\} - \sum_{t=1}^{T}\hat{Z}_t\sum_{s=1}^{T}\hat{Z}_s\{\log m(x,s)-\alpha_x\} \\ = & \beta_x\left(\sum_{t=1}^{T}\hat{Z}_t^2\sum_{s=1}^{T}k_s - \sum_{t=1}^{T}\hat{Z}_t\sum_{s=1}^{T}k_s\hat{Z}_s\right) \\ & +\left(\sum_{t=1}^{T}\hat{Z}_t^2\sum_{s=1}^{T}\varepsilon_{x,s} - \sum_{t=1}^{T}\hat{Z}_t\sum_{s=1}^{T}\varepsilon_{x,s}\hat{Z}_s\right). \end{aligned} \tag{2.20}$$

From (2.11)-(2.13), (2.15), (2.16), (2.17) and the fact that $\sum_{t=1}^{T}\eta_t\tilde{k}_t = O_p(T)$, we have

$$\begin{aligned} & \sum_{t=1}^{T}\hat{Z}_t^2\sum_{s=1}^{T}k_s - \sum_{t=1}^{T}\hat{Z}_t\sum_{s=1}^{T}k_s\hat{Z}_s \\ = & \{\sum_{s=1}^{T}k_s\sum_{t=1}^{T}k_t\eta_t\} + \{\sum_{s=1}^{T}k_s\sum_{t=1}^{T}\eta_t^2\} - \{\sum_{s=1}^{T}k_s^2\sum_{t=1}^{T}\eta_t\} - \{\sum_{s=1}^{T}k_s\eta_s\sum_{t=1}^{T}\eta_t\} \\ = & \{\mu^2(\sum_{t=1}^{T}\sum_{j=0}^{t-1}\phi^j)(\sum_{t=1}^{T}\eta_t\sum_{j=0}^{t-1}\phi^j) + O_p(T^3)\} + \{O_p(T^3)\} \\ & -\{\mu^2\sum_{t=1}^{T}(\sum_{j=0}^{t-1}\phi^j)^2\sum_{t=1}^{T}\eta_t + O_p(T^3)\} - \{O_p(T^2)\} \\ = & \mu^2\{\sum_{s=1}^{T}(\sum_{j=0}^{s-1}\phi^j)\}\{\sum_{t=1}^{T}\eta_t(\sum_{j=0}^{t-1}\phi^j)\} - \mu^2\{\sum_{s=1}^{T}(\sum_{j=0}^{s-1}\phi^j)^2\}\{\sum_{t=1}^{T}\eta_t\} + O_p(T^3) \end{aligned} \tag{2.21}$$

and

$$\begin{aligned} & \sum_{t=1}^{T}\hat{Z}_t^2\sum_{s=1}^{T}\varepsilon_{x,s} - \sum_{t=1}^{T}\hat{Z}_t\sum_{s=1}^{T}\varepsilon_{x,s}\hat{Z}_s \\ = & \{\mu^2\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\phi^j\right)^2\sum_{s=1}^{T}\varepsilon_{x,s} + O_p(T^3)\} \\ & -\{\mu^2\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}\phi^j\right)\sum_{s=1}^{T}\left(\sum_{j=0}^{s-1}\phi^j\right)\varepsilon_{x,s} + O_p(T^3)\}. \end{aligned} \tag{2.22}$$

Therefore, it follows from (2.14), (2.19), (2.20)–(2.22) that

$$T^{1/2}(\hat{\alpha}_x - \alpha_x) \xrightarrow{d} \beta_x Y_* - Y_x.$$

(iv) As before, we can show that the numerator of $\hat{\beta}_x - \beta_x$ is

$$
\begin{aligned}
& T\sum_{s=1}^T \log m(x,s)\hat{Z}_s - \sum_{s=1}^T \log m(x,s)\sum_{t=1}^T \hat{Z}_t - \beta_x T\sum_{t=1}^T \hat{Z}_t^2 + \beta_x(\sum_{t=1}^T \hat{Z}_t)^2 \\
= \; & \beta_x T\left(\sum_{t=1}^T k_t\hat{Z}_t - \sum_{t=1}^T \hat{Z}_t^2\right) + \beta_x\left\{(\sum_{t=1}^T \hat{Z}_t)^2 - \sum_{s=1}^T k_s\sum_{t=1}^T \hat{Z}_t\right\} \\
& \quad + \left(T\sum_{s=1}^T \varepsilon_{x,s}\hat{Z}_s - \sum_{t=1}^T \hat{Z}_t\sum_{s=1}^T \varepsilon_{x,s}\right) \\
= \; & \beta_x\left(\sum_{t=1}^T \hat{Z}_t\sum_{s=1}^T \eta_s - T\sum_{t=1}^T \hat{Z}_t\eta_t\right) - \left(\sum_{t=1}^T \hat{Z}_t\sum_{s=1}^T \varepsilon_{x,s} - T\sum_{s=1}^T \varepsilon_{x,s}\hat{Z}_s\right) \\
= \; & \beta_x\left\{\mu\sum_{t=1}^T\left(\sum_{j=0}^{t-1}\phi^j\right)\sum_{s=1}^T \eta_s - T\mu\sum_{t=1}^T\left(\sum_{j=0}^{t-1}\phi^j\right)\eta_t + O_p(T^2)\right\} \\
& \quad - \left\{\mu\sum_{t=1}^T\left(\sum_{j=0}^{t-1}\phi^j\right)\sum_{s=1}^T \varepsilon_{x,s} - T\mu\sum_{t=1}^T\left(\sum_{j=0}^{t-1}\phi^j\right)\varepsilon_{x,t} + O_p(T^2)\right\}.
\end{aligned}
$$

Then it follows from (2.14) and (2.19) that

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

□

Proof of Theorem 2.2. When $\mu = 0$, we have $k_t = \phi k_{t-1} + e_t$ and it follows from Phillips (1987) that

$$
\begin{cases}
T^{-2}\sum_{t=1}^T k_t^2 \xrightarrow{d} \sigma_e^2\int_0^1 J_\gamma^2(t)\,dt, \\
T^{-3/2}\sum_{t=1}^T k_t \xrightarrow{d} \sigma_e\int_0^1 J_\gamma(t)\,dt, \\
T^{-1}\sum_{t=1}^T k_{t-1}e_t \xrightarrow{d} \sigma_e^2\int_0^1 J_\gamma(t)\,dW_1(t) + \frac{1}{2}(\sigma_e^2 - \sigma^2).
\end{cases}
\tag{2.23}
$$

Similar to the proof of Theorem 2.1, it is easy to show that

$$(T-1)\sum_{t=2}^T \hat{Z}_{t-1}^2 - (\sum_{t=2}^T \hat{Z}_{t-1})^2 = (T-1)\sum_{t=2}^T k_t^2 - (\sum_{t=2}^T k_t)^2 + o_p(T^3),$$

implying that

$$
\begin{cases}
T^{-3}\left\{(T-1)\sum_{t=2}^T \hat{Z}_{t-1}^2 - (\sum_{t=2}^T \hat{Z}_{t-1})^2\right\} \xrightarrow{d} \sigma_e^2\left\{\int_0^1 J_\gamma^2(t)\,dt - (\int_0^1 J_\gamma(t)\,dt)^2\right\}, \\
T^{-3}\left\{T\sum_{t=1}^T \hat{Z}_t^2 - (\sum_{t=1}^T \hat{Z}_t)^2\right\} \xrightarrow{d} \sigma_e^2\left\{\int_0^1 J_\gamma^2(t)\,dt - (\int_0^1 J_\gamma(t)\,dt)^2\right\}.
\end{cases}
\tag{2.24}
$$

(i) It follows from (2.23) that

$$\begin{aligned}\sum_{t=2}^T k_{t-1}(\eta_t - \phi\eta_{t-1}) &= \sum_{t=2}^T k_{t-1}\eta_t - \phi^2 \sum_{t=2}^T k_{t-2}\eta_{t-1} - \phi \sum_{t=2}^T e_{t-1}\eta_{t-1} \\ &= (1-\phi^2)\sum_{t=2}^{T-1} k_{t-1}\eta_t + k_{T-1}\eta_T - k_0\eta_1 - \phi\sum_{t=2}^T e_{t-1}\eta_{t-1} \\ &= -\sum_{t=2}^T e_{t-1}\eta_{t-1} + o_p(T),\end{aligned}$$

implying that

$$\begin{aligned} & (T-1)\sum_{t=2}^T \hat{Z}_{t-1}(\hat{Z}_t - \phi\hat{Z}_{t-1}) - \sum_{t=2}^T \hat{Z}_{t-1}\sum_{s=2}^T(\hat{Z}_s - \phi\hat{Z}_{s-1}) \\ = \; & (T-1)\sum_{t=2}^T k_{t-1}e_t + (T-1)\sum_{t=2}^T \eta_{t-1}(e_t - e_{t-1} + \eta_t - \eta_{t-1}) \\ & - \sum_{t=2}^T k_{t-1}\sum_{s=2}^T e_s + o_p(T^2). \end{aligned} \tag{2.25}$$

Under conditions of Theorem 2.2, by using the law of large numbers for an α-mixing sequence (see McLeish (1975)), we have

$$\frac{1}{T}\sum_{t=2}^T \eta_{t-1}(e_t - e_{t-1} + \eta_t - \eta_{t-1}) \xrightarrow{p} \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^T E\{\eta_{t-1}(e_t - e_{t-1} + \eta_t - \eta_{t-1})\}. \tag{2.26}$$

By (2.23), (2.25) and (2.26), we have

$$\begin{aligned} & T^{-2}\{(T-1)\sum_{t=2}^T \hat{Z}_{t-1}(\hat{Z}_t - \phi\hat{Z}_{t-1}) - \sum_{t=2}^T \hat{Z}_{t-1}\sum_{s=2}^T(\hat{Z}_s - \phi\hat{Z}_{s-1})\} \\ \xrightarrow{d} \; & \sigma_e^2\{\int_0^1 J_\gamma(t)\,dW_1(t) - W_1(1)\int_0^1 J_\gamma(t)\,dt\} + \tfrac{1}{2}(\sigma_e^2 - \sigma^2) + c_0. \end{aligned} \tag{2.27}$$

It follows from (2.24) and (2.27) that

$$T(\hat{\phi} - \phi) \xrightarrow{d} \frac{\sigma_e^2\{\int_0^1 J_\gamma(t)\,dW_1(t) - W_1(1)\int_0^1 J_\gamma(t)\,dt\} + \frac{1}{2}(\sigma_e^2 - \sigma^2) + c_0}{\sigma_e^2\{\int_0^1 J_\gamma^2(t)\,dt - (\int_0^1 J_\gamma(t)\,dt)^2\}}.$$

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

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

It follows from (2.23) that

$$
\begin{aligned}
& \sum_{t=2}^{T} \hat{Z}_{t-1}^2 \sum_{s=2}^{T} (\hat{Z}_s - \hat{Z}_{s-1}) \\
= \; & \sum_{t=2}^{T} (k_{t-1}^2 + 2k_{t-1}\eta_{t-1} + \eta_{t-1}^2)\{(\phi - 1) \sum_{s=2}^{T} k_{s-1} + \sum_{s=2}^{T} e_s + \eta_T - \eta_0\} \\
= \; & (\phi - 1) \sum_{t=2}^{T} k_{t-1}^2 \sum_{s=2}^{T} k_{s-1} + \sum_{t=2}^{T} k_{t-1}^2 \sum_{s=2}^{T} e_s + o_p(T^{5/2})
\end{aligned}
$$

and

$$
\begin{aligned}
& \sum_{s=2}^{T} \hat{Z}_{s-1} \sum_{t=2}^{T} \hat{Z}_{t-1}(\hat{Z}_t - \hat{Z}_{t-1}) \\
= \; & \{\sum_{s=2}^{T} k_{s-1} + \sum_{s=2}^{T} \eta_s\} \sum_{t=2}^{T} \{(k_{t-1} + \eta_{t-1})((\phi - 1)k_{t-1} + e_t + \eta_t - \eta_{t-1})\} \\
= \; & (\phi - 1) \sum_{t=2}^{T} k_{t-1}^2 \sum_{s=2}^{T} k_{s-1} + \sum_{s=2}^{T} k_{s-1} \sum_{t=2}^{T} k_{t-1}e_t \\
& + \sum_{s=2}^{T} k_{s-1} \sum_{t=2}^{T} \eta_{t-1}(e_t - e_{t-1} + \eta_t - \eta_{t-1}) + o_p(T^{5/2}),
\end{aligned}
$$

implying that

$$
\begin{aligned}
& \sum_{t=2}^{T} \hat{Z}_{t-1}^2 \sum_{s=2}^{T} (\hat{Z}_s - \hat{Z}_{s-1}) - \sum_{s=2}^{T} \hat{Z}_{s-1} \sum_{t=2}^{T} \hat{Z}_{t-1}(\hat{Z}_t - \hat{Z}_{t-1}) \\
= \; & \sum_{t=2}^{T} k_{t-1}^2 \sum_{s=2}^{T} e_s - \sum_{s=2}^{T} k_{s-1} \sum_{t=2}^{T} k_{t-1}e_t \\
& - \sum_{s=2}^{T} k_{s-1} \sum_{t=2}^{T} \eta_{t-1}(e_t - e_{t-1} + \eta_t - \eta_{t-1}) + o_p(T^{5/2}).
\end{aligned} \tag{2.28}
$$

Hence, by (2.23), (2.24) and (2.28),

$$
T^{1/2}(\hat{\mu} - \mu) \xrightarrow{d} \frac{\sigma_e^3 W_1(1) \int_0^1 J_\gamma^2(t)\,dt - \sigma_e \int_0^1 J_\gamma(t)\,dt\{\sigma_e^2 \int_0^1 J_\gamma(t)\,dW_1(t) + \frac{1}{2}(\sigma_e^2 - \sigma^2) + c_0\}}{\sigma_e^2\{\int_0^1 J_\gamma^2(t)\,dt - (\int_0^1 J_\gamma(t)\,dt)^2\}}.
$$

(iii) Similar to the proof of (2.23), it follows from (2.7) that

$$
\begin{cases}
T^{-1} \sum_{t=1}^{T} k_{t-1}\varepsilon_{x,t} \xrightarrow{d} \sigma_e \sigma_x \int_0^1 J_\gamma(t)\,dW_{x+1}(t), \\
T^{-1} \sum_{t=1}^{T} k_{t-1}\eta_t \xrightarrow{d} \sigma_e \int_0^1 J_\gamma(t)\,dW_*(t).
\end{cases} \tag{2.29}
$$

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

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

It follows from (2.23) and (2.29) that

$$\begin{aligned} & \sum_{t=1}^T \hat{Z}_t^2 \sum_{s=1}^T k_s - \sum_{t=1}^T \hat{Z}_t \sum_{s=1}^T k_s \hat{Z}_s \\ = & \sum_{t=1}^T k_t \eta_t \sum_{s=1}^T k_s + \sum_{t=1}^T \eta_t^2 \sum_{s=1}^T k_s - \sum_{t=1}^T \eta_t \sum_{s=1}^T k_s^2 - \sum_{t=1}^T \eta_t \sum_{s=1}^T k_s \eta_s \\ = & \sum_{t=1}^T k_t \eta_t \sum_{s=1}^T k_s + \sum_{t=1}^T \eta_t^2 \sum_{s=1}^T k_s - \sum_{t=1}^T \eta_t \sum_{s=1}^T k_s^2 + o_p(T^{5/2}) \end{aligned}$$

and

$$\begin{aligned} & \sum_{t=1}^T \hat{Z}_t^2 \sum_{s=1}^T \varepsilon_{x,s} - \sum_{t=1}^T \hat{Z}_t \sum_{s=1}^T \varepsilon_{x,s} \hat{Z}_s \\ = & \sum_{t=1}^T k_t^2 \sum_{s=1}^T \varepsilon_{x,s} - \sum_{t=1}^T k_t \sum_{s=1}^T k_s \varepsilon_{x,s} - \sum_{t=1}^T k_t \sum_{s=1}^T \eta_s \varepsilon_{x,s} + o_p(T^{5/2}), \end{aligned}$$

implying that

$$\begin{aligned} & \sum_{t=1}^T \hat{Z}_t^2 \sum_{s=1}^T \log m(x,s) - \sum_{t=1}^T \hat{Z}_t \sum_{s=1}^T \log m(x,s) \hat{Z}_s - \alpha_x T \sum_{t=1}^T \hat{Z}_t^2 + \alpha_x (\sum_{t=1}^T \hat{Z}_t)^2 \\ = & \beta_x \{ \sum_{t=1}^T k_t \eta_t \sum_{s=1}^T k_s + \sum_{t=1}^T \eta_t^2 \sum_{s=1}^T k_s - \sum_{t=1}^T \eta_t \sum_{s=1}^T k_s^2 \} \\ & + \sum_{t=1}^T k_t^2 \sum_{s=1}^T \varepsilon_{x,s} - \sum_{t=1}^T k_t \sum_{s=1}^T k_s \varepsilon_{x,s} - \sum_{t=1}^T k_t \sum_{s=1}^T \eta_s \varepsilon_{x,s} + o_p(T^{5/2}). \end{aligned} \tag{2.30}$$

By (2.29) and the law of large numbers for an α-mixing sequence McLeish (1975), we have

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T k_t \eta_t & = \phi \frac{1}{T} \sum_{t=1}^T k_{t-1} \eta_t + \frac{1}{T} \sum_{t=1}^T e_t \eta_t \\ & \xrightarrow{d} \int_0^1 J_\gamma(t)\, dW_*(t) + \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^T E(e_t \eta_t). \end{aligned} \tag{2.31}$$

Therefore it follows from (2.23), (2.24), (2.29)–(2.31) that

$$T^{1/2}(\hat{\alpha}_x - \alpha_x) \xrightarrow{d} \frac{\beta_x Y_* - Y_x + c_x \sigma_e \int_0^1 J_\gamma(t)\, dt}{\sigma_e^2 \{ \int_0^1 J_\gamma^2(t)\, dt - (\int_0^1 J_\gamma(t)\, dt)^2 \}}.$$

(iv) Note that the numerator of $\hat{\beta}_x - \beta_x$ can be written as

$$\begin{aligned} & T\sum_{s=1}^T \log m(x,s)\hat{Z}_s - \sum_{s=1}^T \log m(x,s)\sum_{t=1}^T \hat{Z}_t - \beta_x T\sum_{t=1}^T \hat{Z}_t^2 + \beta_x(\sum_{t=1}^T \hat{Z}_t)^2 \\ = & \beta_x\left(\sum_{t=1}^T \hat{Z}_t \sum_{s=1}^T \eta_s - T\sum_{t=1}^T \hat{Z}_t\eta_t\right) - \left(\sum_{t=1}^T \hat{Z}_t \sum_{s=1}^T \varepsilon_{x,s} - T\sum_{s=1}^T \varepsilon_{x,s}\hat{Z}_s\right) \\ = & \beta_x\left(\sum_{t=1}^T k_t \sum_{s=1}^T \eta_s - T\sum_{t=1}^T k_t\eta_t - T\sum_{t=1}^T \eta_t^2\right) \\ & - \left(\sum_{t=1}^T k_t \sum_{s=1}^T \varepsilon_{x,s} - T\sum_{s=1}^T k_s\varepsilon_{x,s} - T\sum_{s=1}^T \eta_s\varepsilon_{x,s}\right) + o_p(T^2). \end{aligned}$$

Then it follows from (2.23), (2.24), (2.29) and (2.31) that

$$T(\hat{\beta}_x - \beta_x) \xrightarrow{d} \frac{\beta_x Z_* - Z_x - c_x}{\sigma_e^2\{\int_0^1 J_\gamma^2(t)\,dt - (\int_0^1 J_\gamma(t)\,dt)^2\}}.$$

□

Proof of Theorem 2.3. If $|\phi| < 1$ and ϕ is independent of T, i.e., $\{k_t\}$ is stationary, it follows from the assumption of strict stationarity that, as $T \to \infty$,

$$\begin{cases} T^{-1}\sum_{t=2}^T k_t \xrightarrow{p} \frac{\mu}{1-\phi}, \\ T^{-1}\sum_{t=2}^T k_{t-1}e_t \xrightarrow{p} E(k_1e_2), \\ T^{-1}\sum_{t=2}^T k_t^2 \xrightarrow{p} \frac{\mu^2}{(1-\phi)^2} + \frac{E(e_1^2)}{1-\phi^2} + \frac{2\phi E(k_1e_2)}{1-\phi^2}. \end{cases} \tag{2.32}$$

By (2.6) and (2.32), we have

$$\begin{aligned} & T^{-2}\left\{(T-1)\sum_{t=2}^T \hat{Z}_{t-1}^2 - (\sum_{t=2}^T \hat{Z}_{t-1})^2\right\} \\ \xrightarrow{p} & \frac{E(e_1^2)}{1-\phi^2} + \frac{2\phi E(k_1e_2)}{1-\phi^2} + 2\phi E(k_1\eta_2) + 2E(e_1\eta_1) + E(\eta_1^2) \end{aligned} \tag{2.33}$$

and

$$T^{-2}\left\{T\sum_{t=1}^T \hat{Z}_t^2 - (\sum_{t=1}^T \hat{Z}_t)^2\right\} \xrightarrow{p} \frac{E(e_1^2)}{1-\phi^2} + \frac{2\phi E(k_1e_2)}{1-\phi^2} + 2\phi E(k_1\eta_2) + 2E(e_1\eta_1) + E(\eta_1^2). \tag{2.34}$$

Using the law of large numbers for an α-mixing sequence McLeish (1975) and noting the assumption of stationarity, we have

$$\begin{aligned} & T^{-2}\{(T-1)\sum_{t=2}^T \hat{Z}_t\hat{Z}_{t-1} - \sum_{s=2}^T \hat{Z}_s \sum_{t=2}^T \hat{Z}_{t-1}\} \\ \xrightarrow{p} \; & \tfrac{\phi}{1-\phi^2}E(e_1^2) + \tfrac{1+\phi^2}{1-\phi^2}E(k_1e_2) + (\phi^2+1)E(k_1\eta_2) + \phi E(e_1\eta_1) + E(e_2\eta_1) + E(\eta_1\eta_2) \end{aligned} \tag{2.35}$$

and

$$\begin{aligned} & T^{-2}\{\sum_{s=2}^T \hat{Z}_s \sum_{t=2}^T \hat{Z}_{t-1}^2 - \sum_{s=2}^T \hat{Z}_{s-1} \sum_{t=2}^T \hat{Z}_t\hat{Z}_{t-1}\} \\ \xrightarrow{p} \; & \tfrac{\mu}{1-\phi}\{\tfrac{E(e_1^2)}{1+\phi} - \tfrac{1-\phi}{1+\phi}E(k_1e_2) - (1-\phi)^2E(k_1\eta_2) \\ & +(2-\phi)E(e_1\eta_1) + E(\eta_1^2) - E(e_2\eta_1) - E(\eta_1\eta_2)\}. \end{aligned} \tag{2.36}$$

Similarly, for $x = 1, \ldots, M$, we have

$$\begin{aligned} & T^{-2}\{\sum_{s=1}^T \log m(x,s) \sum_{t=1}^T \hat{Z}_t^2 - \sum_{s=1}^T \log m(x,s)\hat{Z}_s \sum_{t=1}^T \hat{Z}_t\} \\ \xrightarrow{p} \; & \alpha_x\{\tfrac{E(e_1^2)}{1-\phi^2} + \tfrac{2\phi E(k_1e_2)}{1-\phi^2} + 2\phi E(k_1\eta_2) + 2E(e_1\eta_1) + E(\eta_1^2)\} \\ & + \tfrac{\mu\beta_x}{1-\phi}\{\phi E(k_1\eta_2) + E(e_1\eta_1) + E(\eta_1^2)\} - \tfrac{\mu}{1-\phi}\{\phi E(k_1\varepsilon_{x,2}) + E(e_1\varepsilon_{x,1}) + E(\eta_1\varepsilon_{x,1})\} \end{aligned} \tag{2.37}$$

and

$$\begin{aligned} & T^{-2}\{T\sum_{s=1}^T \log m(x,s)\hat{Z}_s - \sum_{s=1}^T \log m(x,s) \sum_{t=1}^T \hat{Z}_t\} \\ \xrightarrow{p} \; & \beta_x\{\tfrac{E(e_1^2)}{1-\phi^2} + \tfrac{2\phi E(k_1e_2)}{1-\phi^2} + \phi E(k_1\eta_2) + E(e_1\eta_1)\} + \phi E(k_1\varepsilon_{x,2}) + E(e_1\varepsilon_{x,1}) + E(\eta_1\varepsilon_{x,1}). \end{aligned} \tag{2.38}$$

Thus the theorem follows from (2.33)–(2.38). □

Proof of Theorem 2.4. Put $U_i = L^{-1}\sum_{j=1}^L e_{i+j}$ for $i = 1, \ldots, T-L$. Write

$$\begin{aligned} & \tfrac{L}{T-L}\sum_{i=1}^{T-L} \hat{U}_i^2 \\ = \; & \tfrac{L}{T-L}\sum_{i=1}^{T-L}(\hat{U}_i - U_i)^2 + \tfrac{L}{T-L}\sum_{i=1}^{T-L} U_i^2 + \tfrac{2L}{T-L}\sum_{i=1}^{T-L}(\hat{U}_i - U_i)U_i \\ = \; & I_1 + I_2 + I_3 \end{aligned} \tag{2.39}$$

and

$$\begin{aligned} & L\{\tfrac{1}{T-L}\textstyle\sum_{i=1}^{T-L}\hat{U}_i\}^2 \\ = \quad & L\{\tfrac{1}{T-L}\textstyle\sum_{i=1}^{T-L}(\hat{U}_i - U_i)\}^2 + L\{\tfrac{1}{T-L}\textstyle\sum_{i=1}^{T-L}U_i\}^2 + \tfrac{2L}{(T-L)^2}\textstyle\sum_{i=1}^{T-L}(\hat{U}_i - U_i)U_i \\ = \quad & I_4 + I_5 + I_6. \end{aligned} \tag{2.40}$$

It follows from (3.9) and (3.10) of Lahiri (2013) that

$$I_2 - I_5 \xrightarrow{p} \sigma_e^2 \text{ as } T \to \infty. \tag{2.41}$$

Similarly we have

$$\frac{L}{T-L}\sum_{i=1}^{T-L}\{\frac{1}{L}\sum_{j=1}^{L}\eta_{i+j}\}^2 = O_p(1). \tag{2.42}$$

Since

$$\begin{aligned} \hat{e}_t - e_t &= (\hat{Z}_t - \hat{\mu} - \hat{\phi}\hat{Z}_{t-1}) - (k_t - \mu - \phi k_{t-1}) \\ &= (\hat{Z}_t - k_t) - (\hat{\mu} - \mu) - (\hat{\phi}\hat{Z}_{t-1} - \phi\hat{Z}_{t-1} + \phi\hat{Z}_{t-1} - \phi k_{t-1}) \\ &= (\eta_t - \phi\eta_{t-1}) - (\hat{\mu} - \mu) - (\hat{\phi} - \phi)(k_{t-1} + \eta_{t-1}), \end{aligned}$$

we have

$$\hat{U}_i - U_i = \frac{1}{L}\sum_{j=1}^{L}(\eta_{i+j} - \phi\eta_{i+j-1}) - (\hat{\mu} - \mu) - (\hat{\phi} - \phi)\frac{1}{L}\sum_{j=1}^{L}(k_{i+j-1} + \eta_{i+j-1}). \tag{2.43}$$

Hence

$$
\begin{aligned}
I_1 &= \tfrac{L}{T-L}\sum_{i=1}^{T-L}\{\tfrac{1}{L}\sum_{j=1}^{L}(\eta_{i+j}-\phi\eta_{i+j-1})-(\hat{\mu}-\mu)-(\hat{\phi}-\phi)\tfrac{1}{L}\sum_{j=1}^{L}(k_{i+j-1}+\eta_{i+j-1})\}^2 \\
&\le \tfrac{3L}{T-L}\sum_{i=1}^{T-L}\{\tfrac{1}{L}\sum_{j=1}^{L}(\eta_{i+j}-\phi\eta_{i+j-1})\}^2+3L(\hat{\mu}-\mu)^2 \\
&\quad +(\hat{\phi}-\phi)^2\tfrac{3L}{T-L}\sum_{i=1}^{T-L}\{\tfrac{1}{L}\sum_{j=1}^{L}(k_{i+j-1}+\eta_{i+j-1})\}^2 \\
&\le \tfrac{3}{L(T-L)}\sum_{i=1}^{T-L}\{\eta_{i+L}-\phi\eta_i+(1-\phi)\sum_{j=1}^{L-1}\eta_{i+j}\}^2+3L(\hat{\mu}-\mu)^2 \\
&\quad +(\hat{\phi}-\phi)^2\tfrac{6L}{T-L}\sum_{i=1}^{T-L}\{\tfrac{1}{L}\sum_{j=1}^{L}k_{i+j-1}\}^2+(\hat{\phi}-\phi)^2\tfrac{6L}{T-L}\sum_{i=1}^{T-L}\{\tfrac{1}{L}\sum_{j=1}^{L}\eta_{i+j-1}\}^2 \\
&\le \tfrac{9}{L(T-L)}\sum_{i=1}^{T-L}\eta_{i+L}^2+\tfrac{9\phi^2}{L(T-L)}\sum_{i=1}^{T-L}\eta_i^2+\tfrac{9(1-\phi)^2}{L(T-L)}\sum_{i=1}^{T-L}\{\sum_{j=1}^{L-1}\eta_{i+j}\}^2+3L(\hat{\mu}-\mu)^2 \\
&\quad +(\hat{\phi}-\phi)^2\tfrac{6L}{T-L}\sum_{i=1}^{T-L}\{\tfrac{1}{L}\sum_{j=1}^{L}k_{i+j-1}\}^2+(\hat{\phi}-\phi)^2\tfrac{6L}{T-L}\sum_{i=1}^{T-L}\{\tfrac{1}{L}\sum_{j=1}^{L}\eta_{i+j-1}\}^2 \\
&= II_1+\cdots+II_6.
\end{aligned}
$$

Using (2.42), Theorem 2.1, $E\eta_i^2<\infty$, $1-\phi\to 0$ and $L\to\infty$ as $T\to\infty$, we have

$$II_i = o_p(1) \text{ for } i=1,2,3,4,6. \tag{2.44}$$

By noting that $\hat{\phi}-\phi = O_p(T^{-3/2})$ and $T^{-1}\max_{1\le t\le T}|k_t| = O_p(1)$, we have

$$II_5 = o_p(1). \tag{2.45}$$

Hence, it follows from (2.44) and (2.45) that

$$I_1 = o_p(1). \tag{2.46}$$

By Hölder inequality, we have

$$I_3 = O_p(\sqrt{|I_1 I_2|}) = o_p(1). \tag{2.47}$$

It follows from (2.43) and similar arguments in proving (2.46) that

$$I_4 = o_p(1). \tag{2.48}$$

Using Hölder inequality again, we have

$$I_6 = O_p(\sqrt{|I_4 I_5|}) = o_p(1). \tag{2.49}$$

Therefore the theorem follows from (2.41), (2.46)–(2.49). □

Proof of Theorem 2.5. By noting that

$$\begin{aligned}\textstyle\sum_{s=1}^d \hat{e}_{t+s} &= \textstyle\sum_{s=1}^d e_{t+s} + \eta_{t+d} - \eta_t \\ &\quad \textstyle -d(\hat{\mu}-\mu) - (\hat{\phi}-1)\sum_{s=1}^d k_{t+s-1} - (\hat{\phi}-1)\sum_{j=1}^d \eta_{t+s-1},\end{aligned}$$

$$\hat{\mu}-\mu = o_p(1), \quad \sup_{1\le t\le T}\{|(\hat{\phi}-1)\sum_{s=1}^d k_{t+s-1}| + |(\hat{\phi}-1)\sum_{s=1}^d \eta_{t+s-1}|\} = o_p(1),$$

we have

$$\hat{H}_d(y) = \frac{1}{T-d}\sum_{t=1}^{T-d} I\big(\frac{1}{M}(\eta_t - \sum_{s=1}^d e_{t+s} - \eta_{t+d}) \le y\big) + o_p(1) \text{ for any } y \in \mathbb{R}. \tag{2.50}$$

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

$$\frac{1}{T-d}\sum_{t=1}^{T-d} I\big(\frac{1}{M}(\eta_t - \sum_{s=1}^d e_{t+s} - \eta_{t+d}) \le y\big) \xrightarrow{p} H_d(y) \text{ for any } y \in \mathbb{R}. \tag{2.51}$$

Hence the theorem follows from (2.50) and (2.51). □

PART 3

BIAS CORRECTED INFERENCE FOR A MODIFIED LEE-CARTER MORTALITY MODEL

This Part is my published paper Liu et al. (2019a), but has been adapted to the format of dissertation.

The Lee-Carter model is a combination of the following two structures for modeling the central death rate $m(x,t)$ at age or age group $x = 1, \ldots, M$ and time $t = 1, \ldots, T$:

$$\log m(x,t) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}, \quad \sum_{x=1}^{M} \beta_x = 1, \quad \sum_{t=1}^{T} k_t = 0, \tag{3.1}$$

and

$$k_t = \mu + \rho k_{t-1} + e_t, \tag{3.2}$$

where $\{\varepsilon_{x,t}\}_{t=1}^{T}$ and $\{e_t\}_{t=1}^{T}$ are random errors with zero mean and finite variance, the unobserved $\{k_t\}$ is called the mortality index. A detailed assumption on the dependence of these random errors is given in the next section. Note that model (3.2) can be replaced by a more general time series model although researchers often claim that a unit root AR(1) model fits well to real mortality rates. Some recent applications of the above models (3.1) and (3.2) with $\rho = 1$ in actuarial science include Li et al. (2017b), Kwok et al. (2016), Enchev et al. (2017), Biffis et al. (2017), Lin et al. (2017), Wong et al. (2017), Zhu et al. (2017). Among these applications, a commonly employed statistical inference is the two-step procedure in Lee & Carter (1992), which first estimates α_x, β_x, k_t for $x = 1, \ldots, M$ and $t = 1, \ldots, T$ by the singular value decomposition method based on model (3.1) and then fits model (3.2) to the estimated k_t's. Unfortunately Leng & Peng (2016) showed that such an inference procedure may be inconsistent when the mortality index is not exactly an AR(1) unit root process.

So far many extensions and applications of this Lee-Carter model have appeared in the literature of actuarial science with an open statistical R package 'demography'. The two constraints in (3.1) ensure that the model is identifiable. Recently, to relax the very restrictive constraint $\sum_{t=1}^{T} k_t = 0$ on (3.2), which basically implies that $\mu = 0$, Liu et al. (2019b) proposed the following modified Lee-Carter model:

$$\log m(x,t) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}, \quad k_t = \mu + \phi k_{t-1} + e_t, \quad \sum_{x=1}^{M} \beta_x = 1, \quad \sum_{x=1}^{M} \alpha_x = 0, \tag{3.3}$$

where the condition $\sum_{x=1}^{M} \alpha_x = 0$ is not restrictive at all because the sum can be absorbed into μ via k_t. In order to estimate the unknown parameters and derive the asymptotic properties, Liu et al. (2019b) proposed the following inference procedure without using the singular value decomposition method under the setup that the sequence $\{(e_t, \varepsilon_{1,t}, \cdots, \varepsilon_{M,t})^\tau : t = 1, \ldots, T\}$ in (3.3) is an α-mixing sequence, where A^τ denotes the transpose of matrix or vector A.

Define $Z_t = \sum_{x=1}^{M} \log m(x,t)$ and $\eta_t = \sum_{x=1}^{M} \varepsilon_{x,t}$ for $t = 1, \ldots, T$. Then model (3.3) implies that $Z_t = k_t + \eta_t$ for $t = 1, \ldots, T$. When $\{k_t\}$ is a unit root or near unit root process, k_t dominates η_t for t large enough, which motivates Liu et al. (2019b) to estimate the unknown parameters by minimizing the following sums of squares

$$\sum_{t=2}^{T} (Z_t - \mu - \phi Z_{t-1})^2 \text{ and } \sum_{t=2}^{T} (\log m(x,t) - \alpha_x - \beta_x Z_t)^2$$

for $x = 1, \ldots, M$. That is, one solves the following score equations for $x = 1, \ldots, M$:

$$\begin{cases} \sum_{t=2}^{T} \{Z_t - \mu - \phi Z_{t-1}\} = 0, \; \sum_{t=2}^{T} \{Z_t - \mu - \phi Z_{t-1}\} Z_{t-1} = 0, \\ \sum_{t=2}^{T} \{\log m(x,t) - \alpha_x - \beta_x Z_t\} = 0, \; \sum_{t=2}^{T} \{\log m(x,t) - \alpha_x - \beta_x Z_t\} Z_t = 0. \end{cases} \tag{3.4}$$

However, when $\{k_t\}$ is a stationary sequence, the above least squares estimators are inconsistent, and the proposed unit root test in Liu et al. (2019b) does reject the unit root null hypothesis for some real mortality rates. This raises an interesting question on whether one could estimate these unknown parameters consistently regardless of whether $\{k_t\}$ is stationary or unit root or near unit root.

In this Part we propose a simple bias corrected estimation for the modified Lee-Carter model (3.3) and derive its asymptotic distribution regardless of the property of $\{k_t\}$ when errors are independent; see section 3.1 for details on the methodology and main asymptotic results. A simulation study and data analyses are given in section 3.2. Section 3.3 summarizes our contributions. All proofs are put into the Appendix.

3.1 Methodology and main asymptotic results

Throughout we use $\alpha_{x,0}, \beta_{x,0}, \mu_0, \phi_0$ to denote the true values of $\alpha_x, \beta_x, \mu, \phi$ in (3.3), respectively, hence $\sum_{x=1}^M \alpha_{x,0} = 0$ and $\sum_{x=1}^M \beta_{x,0} = 1$. Further we assume that

C1) $\{(e_t, \varepsilon_{1,t}, \cdots, \varepsilon_{M,t})^\tau : t = 1, \ldots, T\}$ is a sequence of independent and identically distributed random vectors with means zero and finite covariance matrix.

By noting that the inconsistency of the least squares estimators via solving (3.4) is due to the correlation between $Z_t - \mu_0 - \phi_0 Z_{t-1} = e_t + \eta_t - \phi_0 \eta_{t-1}$ and $Z_{t-1} = k_{t-1} + \eta_{t-1}$, we propose the simple bias corrected estimators via solving the following modified score equations:

$$\begin{cases} \sum_{t=3}^T \{Z_t - \mu - \phi Z_{t-1}\} = 0, \ \ \sum_{t=3}^T \{Z_t - \mu - \phi Z_{t-1}\} Z_{t-2} = 0, \\ \sum_{t=3}^T \{\log m(x,t) - \alpha_x - \beta_x Z_t\} = 0, \ \ \sum_{t=3}^T \{\log m(x,t) - \alpha_x - \beta_x Z_t\} Z_{t-1} = 0, \end{cases} \tag{3.5}$$

which give

$$\hat{\mu} = \frac{\sum_{s=3}^T Z_s \sum_{t=3}^T Z_{t-1} Z_{t-2} - \sum_{s=3}^T Z_{s-1} \sum_{t=3}^T Z_t Z_{t-2}}{(T-2)\sum_{t=3}^T Z_{t-1} Z_{t-2} - \sum_{s=3}^T Z_{s-1} \sum_{t=3}^T Z_{t-2}},$$

$$\hat{\phi} = \frac{(T-2)\sum_{t=3}^T Z_t Z_{t-2} - \sum_{s=3}^T Z_s \sum_{t=3}^T Z_{t-2}}{(T-2)\sum_{t=3}^T Z_{t-1} Z_{t-2} - \sum_{s=3}^T Z_{s-1} \sum_{t=3}^T Z_{t-2}},$$

$$\hat{\alpha}_x = \frac{\sum_{s=3}^T \log m(x,s) \sum_{t=3}^T Z_t Z_{t-1} - \sum_{s=3}^T \log m(x,s) Z_{s-1} \sum_{t=3}^T Z_t}{(T-2)\sum_{t=3}^T Z_t Z_{t-1} - \sum_{s=3}^T Z_s \sum_{t=3}^T Z_{t-1}},$$

$$\hat{\beta}_x = \frac{(T-2)\sum_{t=3}^T \log m(x,t) Z_{t-1} - \sum_{s=3}^T \log m(x,s) \sum_{t=3}^T Z_{t-1}}{(T-2)\sum_{t=3}^T Z_t Z_{t-1} - \sum_{s=3}^T Z_s \sum_{t=3}^T Z_{t-1}}$$

for $x = 1, \dots, M$. We remark that the estimator $\hat{\phi}$ is the same as the modified Yule-Walker estimator in Staudenmayer & Buonaccorsi (2005) for a time series model with measurement errors, and obviously we have $\sum_{x=1}^{M} \hat{\alpha}_x = 0$ and $\sum_{x=1}^{M} \hat{\beta}_x = 1$.

To present the asymptotic distribution of the proposed bias-corrected estimators, we need some notations. Put $\boldsymbol{\theta} = (\mu, \phi, \alpha_1, \beta_1, \cdots, \alpha_{M-1}, \beta_{M-1})^\tau$, $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\phi}, \hat{\alpha}_1, \hat{\beta}_1, \cdots, \hat{\alpha}_{M-1}, \hat{\beta}_{M-1})^\tau$ and let $\boldsymbol{\theta}_0 = (\mu_0, \phi_0, \alpha_{1,0}, \beta_{1,0}, \cdots, \alpha_{M-1,0}, \beta_{M-1,0})^\tau$ denote the true value of $\boldsymbol{\theta}$. Note that we exclude α_M and β_M in the above definitions due to the constraints $\sum_{x=1}^{M} \alpha_{x,0} = 0$ and $\sum_{x=1}^{M} \beta_{x,0} = 1$. For the stationary case, i.e., $|\phi_0| < 1$ independent of T, we define the symmetric matrix $\Sigma = (\sigma_{i,j})_{1 \le i,j \le 2M}$ with

$$\sigma_{1,1} = E(e_1 + (1-\phi_0)\eta_1)^2, \ \sigma_{1,2} = \frac{\mu_0}{1-\phi_0}\sigma_{1,1},$$

$$\sigma_{1,2x+1} = E\{(\varepsilon_{x,1} - \beta_{x,0}\eta_1)(e_1 + (1-\phi_0)\eta_1)\}, \ \sigma_{1,2x+2} = \frac{\mu_0}{1-\phi_0}\sigma_{1,2x+1},$$

$$\begin{aligned} \sigma_{2,2} \ &= \{E(e_1+\eta_1)^2 + \phi_0^2 E(\eta_1^2)\}\{\tfrac{\mu_0^2}{(1-\phi_0)^2} + \tfrac{E(e_1^2)}{1-\phi_0^2} + E(\eta_1^2) + 2E(e_1\eta_1)\} \\ &\quad -2\phi_0\{E(e_1\eta_1) + E(\eta_1^2)\}\{\tfrac{\mu_0^2}{(1-\phi_0)^2} + \tfrac{\phi_0 E(e_1^2)}{1-\phi_0^2} + \phi_0 E(e_1\eta_1)\}, \end{aligned}$$

$$\begin{aligned} \sigma_{2,2x+2} \ &= E\{e_1(\varepsilon_{x,1} - \beta_{x,0}\eta_1)\}\{\tfrac{\mu_0^2}{(1-\phi_0)^2} + \tfrac{\phi_0 E(e_1^2)}{1-\phi_0^2} + \phi_0 E(e_1\eta_1)\} \\ &\quad +E\{\eta_1(\varepsilon_{x,1} - \beta_{x,0}\eta_1)\}\{\tfrac{\mu_0^2}{1-\phi_0} - \phi_0 E(e_1\eta_1) - \phi_0 E(\eta_1^2)\}, \end{aligned}$$

$$\sigma_{2,2x+1} = \sigma_{1,2x+2}, \ \sigma_{2x+1,2y+1} = E\{(\varepsilon_{x,1} - \beta_{x,0}\eta_1)(\varepsilon_{y,1} - \beta_{y,0}\eta_1)\}, \ \sigma_{2x+1,2y+2} = \frac{\mu_0}{1-\phi_0}\sigma_{2x+1,2y+1},$$

$$\sigma_{2x+2,2y+2} = E\{(\varepsilon_{x,1} - \beta_{x,0}\eta_1)(\varepsilon_{y,1} - \beta_{y,0}\eta_1)\}\{\frac{\mu_0^2}{(1-\phi_0)^2} + \frac{E(e_1^2)}{1-\phi_0^2} + 2E(e_1\eta_1) + E(\eta_1^2)\}$$

for $1 \le x, y \le M-1$ and

$$\Gamma = diag(A_0, \cdots, A_{M-1}) \tag{3.6}$$

with

$$A_0 = \cdots = A_{M-1} = \begin{pmatrix} 1 & \frac{\mu_0}{1-\phi_0} \\ \frac{\mu_0}{1-\phi_0} & \frac{\mu_0^2}{(1-\phi_0)^2} + \frac{\phi_0 E(e_1^2)}{1-\phi_0^2} + \phi_0 E(e_1\eta_1) \end{pmatrix}.$$

For the nonstationary case, i.e., $\phi_0 = 1 + \rho/T$ for some $\rho \in \mathbb{R}$, we define the symmetric matrix $\tilde{\Sigma} = (\tilde{\sigma}_{i,j})_{1\le i,j\le 2M}$ with

$$\tilde{\sigma}_{1,1} = E(e_1^2),\ \tilde{\sigma}_{1,2} = E(e_1^2)\int_0^1 f_{\rho,\mu_0}(s)\,ds,$$

$$\tilde{\sigma}_{1,2x+1} = E\{e_1(\varepsilon_{x,1} - \beta_{x,0}\eta_1)\},\ \tilde{\sigma}_{1,2x+2} = E\{e_1(\varepsilon_{x,1} - \beta_{x,0}\eta_1)\}\int_0^1 f_{\rho,\mu_0}(s)\,ds,$$

$$\tilde{\sigma}_{2,2} = E(e_1^2)\int_0^1 f_{\rho,\mu_0}^2(s)\,ds,\ \tilde{\sigma}_{2,2x+1} = \tilde{\sigma}_{1,2x+2},$$

$$\tilde{\sigma}_{2,2x+2} = E\{e_1(\varepsilon_{x,1} - \beta_{x,0}\eta_1)\}\int_0^1 f_{\rho,\mu_0}^2(s)\,ds,\ \tilde{\sigma}_{2x+1,2y+1} = E\{(\varepsilon_{x,1} - \beta_{x,0}\eta_1)(\varepsilon_{y,1} - \beta_{y,0}\eta_1)\},$$

$$\tilde{\sigma}_{2x+1,2y+2} = E\{(\varepsilon_{x,1} - \beta_{x,0}\eta_1)(\varepsilon_{y,1} - \beta_{y,0}\eta_1)\}\int_0^1 f_{\rho,\mu_0}(s)\,ds,$$

$$\tilde{\sigma}_{2x+2,2y+2} = E\{(\varepsilon_{x,1} - \beta_{x,0}\eta_1)(\varepsilon_{y,1} - \beta_{y,0}\eta_1)\}\int_0^1 f_{\rho,\mu_0}^2(s)\,ds$$

for $1 \le x, y \le M-1$, where

$$f_{\rho,\mu_0}(s) = \begin{cases} \mu_0 \frac{e^{s\rho}-1}{\rho} & \text{if} \quad \rho \neq 0, \\ \mu_0 s & \text{if} \quad \rho = 0, \end{cases} \tag{3.7}$$

and

$$\tilde{\Gamma} = diag(\tilde{A}_0, \cdots, \tilde{A}_{M-1}) \tag{3.8}$$

with

$$\tilde{A}_0 = \cdots = \tilde{A}_{M-1} = \begin{pmatrix} 1 & \int_0^1 f_{\rho,\mu_0}(s)ds \\ \int_0^1 f_{\rho,\mu_0}(s)ds & \int_0^1 f_{\rho,\mu_0}^2(s)ds \end{pmatrix}.$$

Here we focus on the asymptotic result for the case of $\mu_0 \neq 0$ as real mortality rates are often in this situation. Results for the case of $\mu_0 = 0$ can be derived similarly, but with a different rate of convergence for $\hat{\phi}$ and $\hat{\beta}_x$. Throughout all limits in the theorems below are taken as $T \to \infty$ with a fixed M, and we use $\xrightarrow{d}$ and $\xrightarrow{p}$ to denote convergence in distribution and in probability, respectively.

Theorem 3.1. *Assume model (3.3) holds with C1) and $\mu_0 \neq 0$.*

i) When $|\phi_0| < 1$ independent of T (i.e., stationary case), we have

$$\sqrt{T}\Gamma\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\} \xrightarrow{d} N(0, \Sigma).$$

ii) When $\phi_0 = 1 + \rho/T$ for some constant $\rho \in \mathbb{R}$ (i.e., near unit root if $\rho \neq 0$ and unit root if $\rho = 0$), we have

$$D_T\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\} \xrightarrow{d} N(0, \widetilde{\Gamma}^{-1}\widetilde{\Sigma}\widetilde{\Gamma}^{-1}),$$

where D_T is a diagonal matrix with $T^{1/2}$ in the odd diagonal elements and $T^{3/2}$ in the even diagonal element.

Remark 3.1. *It is easy to check that Γ is singular only when $E(e_1^2) = (\phi_0^2 - 1)E(e_1\eta_1)$, $\tilde{\Gamma}$ is always nonsingular, Σ and $\tilde{\Sigma}$ are positive semidefinite. When e_t and $(\varepsilon_{1,t}, \cdots, \varepsilon_{M,t})^\tau$ are uncorrelated and the covariance matrix of $(\varepsilon_{1,t}, \cdots, \varepsilon_{M,t})^\tau$ is positive definite, it follows from Lemma 1 in the Appendix that $\Sigma = \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{12}^\tau & \Sigma_2 \end{pmatrix}$ and $\widetilde{\Sigma} = \begin{pmatrix} \widetilde{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \widetilde{\Sigma}_2 \end{pmatrix}$, where Σ_1 and $\widetilde{\Sigma}_1$ are positive definite 2×2 matrices, Σ_2 and $\widetilde{\Sigma}_2$ are positive definite $(2M-2) \times (2M-2)$ matrices.*

In order to employ the above theorem to construct a confidence region for $\boldsymbol{\theta}_0$ or a part of $\boldsymbol{\theta}_0$, or to test $H_0 : \phi_0 = 1$, one has to estimate Σ and $\widetilde{\Sigma}$ consistently, which can be done as follows.

Define $\tilde{Y}_{t,x}(\alpha_x, \beta_x) = \log m(x,t) - \alpha_x - \beta_x Z_t$ and $\hat{Y}_{t,x}(\alpha_x, \beta_x) = \{\log m(x,t) - \alpha_x - \beta_x Z_t\} Z_{t-1}$ for $x = 1, \ldots, M$ and $t = 3, \ldots, T$. Further put

$$Y_{t,1}(\mu, \phi) = Z_t - \mu - \phi Z_{t-1}, \quad Y_{t,2}(\mu, \phi) = \{Z_t - \mu - \phi Z_{t-1}\} Z_{t-2},$$

$$\begin{aligned} &\boldsymbol{W}_t(\boldsymbol{\theta}) \\ &= \big(Y_{t,1}(\mu,\phi), Y_{t,2}(\mu,\phi), \tilde{Y}_{t,1}(\alpha_1,\beta_1), \hat{Y}_{t,1}(\alpha_1,\beta_1), \cdots, \tilde{Y}_{t,M-1}(\alpha_{M-1},\beta_{M-1}), \hat{Y}_{t,M-1}(\alpha_{M-1},\beta_{M-1})\big)^\tau \end{aligned}$$

and

$$\widetilde{\boldsymbol{W}}_t(\boldsymbol{\theta}) = \big(Y_{t,1}(\mu,\phi), \tfrac{Y_{t,2}(\mu,\phi)}{T}, \tilde{Y}_{t,1}(\alpha_1,\beta_1), \tfrac{\hat{Y}_{t,1}(\alpha_1,\beta_1)}{T}, \cdots, \tilde{Y}_{t,M-1}(\alpha_{M-1},\beta_{M-1}), \tfrac{\hat{Y}_{t,M-1}(\alpha_{M-1},\beta_{M-1})}{T}\big)^\tau$$

for $t = 3, \cdots, T$. Then we propose to estimate Σ by

$$\widehat{\Sigma} = \frac{1}{T-2}\sum_{t=3}^{T} \boldsymbol{W}_t(\hat{\boldsymbol{\theta}})\boldsymbol{W}_t^\tau(\hat{\boldsymbol{\theta}}) + \frac{1}{T-3}\sum_{t=4}^{T} \boldsymbol{W}_t(\hat{\boldsymbol{\theta}})\boldsymbol{W}_{t-1}^\tau(\hat{\boldsymbol{\theta}}) + \frac{1}{T-3}\sum_{t=3}^{T-1} \boldsymbol{W}_t(\hat{\boldsymbol{\theta}})\boldsymbol{W}_{t+1}^\tau(\hat{\boldsymbol{\theta}}).$$

Estimating $\tilde{\Sigma}$ is done in the same as the previous one, $\widehat{\Sigma}$, replacing W_t by $\widetilde{W}_t$, which is denoted by $\hat{\widetilde{\Sigma}}$.

Theorem 3.2. *Assume model (3.3) holds with C1) and $\mu_0 \neq 0$.*

i) When $|\phi_0| < 1$ independent of T, we have $\widehat{\Sigma} \xrightarrow{p} \Sigma$ as $T \to \infty$.

ii) When $\phi_0 = 1 + \rho/T$ for some constant $\rho \in \mathbb{R}$, we have $\hat{\widetilde{\Sigma}} \xrightarrow{p} \widetilde{\Sigma}$ as $T \to \infty$.

Remark 3.2. *For testing $H_0 : \phi_0 = 1$, it follows from Theorem 3.1 that*

$$\big(\sqrt{T}(\hat{\mu} - \mu_0), T^{3/2}(\hat{\phi} - \phi_0)\big)^\tau \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \frac{\mu_0}{2} \\ \frac{\mu_0}{2} & \frac{\mu_0^2}{3} \end{pmatrix}^{-1} \begin{pmatrix} \tilde{\sigma}_{1,1} & \tilde{\sigma}_{1,2} \\ \tilde{\sigma}_{1,2} & \tilde{\sigma}_{2,2} \end{pmatrix} \begin{pmatrix} 1 & \frac{\mu_0}{2} \\ \frac{\mu_0}{2} & \frac{\mu_0^2}{3} \end{pmatrix}^{-1}\right),$$

which implies that

$$T^{3/2}(\hat{\phi} - \phi_0) \xrightarrow{d} N\big(0, \frac{144}{\mu_0^4}(\frac{\mu_0^2}{4}\tilde{\sigma}_{1,1} - \mu_0\tilde{\sigma}_{1,2} + \tilde{\sigma}_{2,2})\big),$$

where $\mu_0, \tilde{\sigma}_{i,j}$ can be estimated by $\hat{\mu}$ in Theorem 3.1 and $\hat{\tilde{\sigma}}_{i,j}$ in Theorem 3.2.

Remark 3.3. *We forecast r future mortality rates $\log m(x, T+1), \cdots, \log m(x, T+r)$ by*

$$\log \widehat{m(x, T+s)} = \hat{\alpha}_x + \hat{\beta}_x \hat{k}_{T+s} \quad \textit{for} \quad s = 1, \ldots, r,$$

where

$$\hat{k}_{T+1} = \hat{\mu} + \hat{\phi} Z_T,\ \hat{k}_{T+2} = \hat{\mu} + \hat{\phi}\hat{k}_{T+1},\ \cdots,\ \hat{k}_{T+r} = \hat{\mu} + \hat{\phi}\hat{k}_{T+r-1}.$$

Write

$$\begin{aligned}
& \log \widehat{m(x;T+1)} - \log m(x;T+1) \\
= \; & (\hat{\alpha}_x - \alpha_{x,0}) + (\hat{\beta}_x - \beta_{x,0})(\hat{\mu} + \hat{\phi} Z_T) + \beta_{x,0}(\hat{\mu} - \mu_0) + \beta_{x,0}(\hat{\phi} - \phi_0) Z_T \\
& + \beta_{x,0}\phi_0\eta_T - \beta_{x,0} e_{T+1} - \varepsilon_{x,T+1},
\end{aligned}$$

it becomes necessary to estimate the distribution function of $\beta_{x,0}\phi_0\eta_T - \beta_{x,0}e_{T+1} - \varepsilon_{x,T+1}$ *in order to quantify the forecast error. Unfortunately it remains unknown on how to estimate this distribution function nonparametrically, which is conjectured to be impossible. However, we have*

$$\sum_{x=1}^{M} \{\beta_{x,0}\phi_0\eta_T - \beta_{x,0}e_{T+1} - \varepsilon_{x,T+1}\} = \phi_0\eta_T - e_{T+1} - \eta_{T+1},$$

whose distribution function can be estimated nonparametrically by the empirical distribution of $\{Z_t - \hat{\mu} - \hat{\phi} Z_{t-1}\}_{t=1}^{T}$*. Therefore it is possible to quantify the uncertainty of the forecasts* $\sum_{x=1}^{M} \log \widehat{m(x;T+s)}$ *for* $s = 1, \ldots, r$*.*

3.2 Simulation study and data analysis

Data Analysis We compare the proposed bias corrected inference with the inference methods in Lee & Carter (1992) and in Part 2 for analyzing the U.S. mortality data obtained from the Human Mortality Database (HMD) [1]. We study the U.S. mortality data of population between 25 and 74 years old from 1933 to 2015, and we use the mortality data by 5-year age groups. Hence, $M = 10$ and $T = 83$.

We employ R package 'demography' to implement the classic Lee-Carter model. The estimates for α_x's, β_x's, μ's and ϕ's are reported in Tables 3.1–3.3 for the U.S. female, male and combined mortality rates, respectively. We observe that the proposed bias corrected estimate gives the smallest ϕ and largest $|\mu|$, and a clear difference in estimating α_x for these three methods.

Next we apply these three methods to forecast future logarithms of mortality rates

[1] http://www.mortality.org/cgi-bin/hmd-country.php?cntr=USA&level=1

Table (3.1) US female mortality rates.

	x	1	2	3	4	5	6	7	8	9	10
Lee and Carter (1992)	$\hat{\alpha}_x$	-7.011	-6.736	-6.377	-5.984	-5.572	-5.159	-4.770	-4.348	-3.929	-3.465
	$\hat{\beta}_x$	0.135	0.128	0.119	0.106	0.095	0.090	0.083	0.080	0.081	0.083
Part 2	$\hat{\alpha}_x$	0.172	0.055	-0.022	-0.344	-0.474	-0.327	-0.337	-0.067	0.384	0.959
	$\hat{\beta}_x$	0.135	0.127	0.119	0.106	0.096	0.091	0.083	0.080	0.081	0.083
Bias Corrected Inference	$\hat{\alpha}_x$	-0.025	-0.070	-0.085	-0.318	-0.453	-0.276	-0.261	-0.023	0.463	1.047
	$\hat{\beta}_x$	0.131	0.125	0.118	0.106	0.096	0.092	0.084	0.081	0.082	0.085

	$\hat{\mu}$	$\hat{\phi}$
Lee and Carter (1992)	-0.157	0.975
Part 2	-1.389	0.977
Bias Corrected Inference	-1.547	0.974

Table (3.2) US male mortality rates.

	x	1	2	3	4	5	6	7	8	9	10
Lee and Carter (1992)	$\hat{\alpha}_x$	-6.262	-6.125	-5.844	-5.472	-5.048	-4.612	-4.204	-3.798	-3.415	-3.015
	$\hat{\beta}_x$	0.088	0.094	0.106	0.109	0.108	0.108	0.103	0.099	0.096	0.090
Part 2	$\hat{\alpha}_x$	-2.068	-1.631	-0.789	-0.270	0.099	0.547	0.714	0.940	1.152	1.308
	$\hat{\beta}_x$	0.088	0.094	0.106	0.109	0.108	0.108	0.103	0.099	0.096	0.090
Bias Corrected Inference	$\hat{\alpha}_x$	-2.267	-1.838	-0.938	-0.305	0.105	0.623	0.856	1.063	1.300	1.400
	$\hat{\beta}_x$	0.084	0.090	0.103	0.108	0.108	0.109	0.106	0.102	0.099	0.092

	$\hat{\mu}$	$\hat{\phi}$
Lee and Carter (1992)	-0.118	0.994
Part 2	-0.441	0.993
Bias Corrected Inference	-0.662	0.989

Table (3.3) US combined mortality rates.

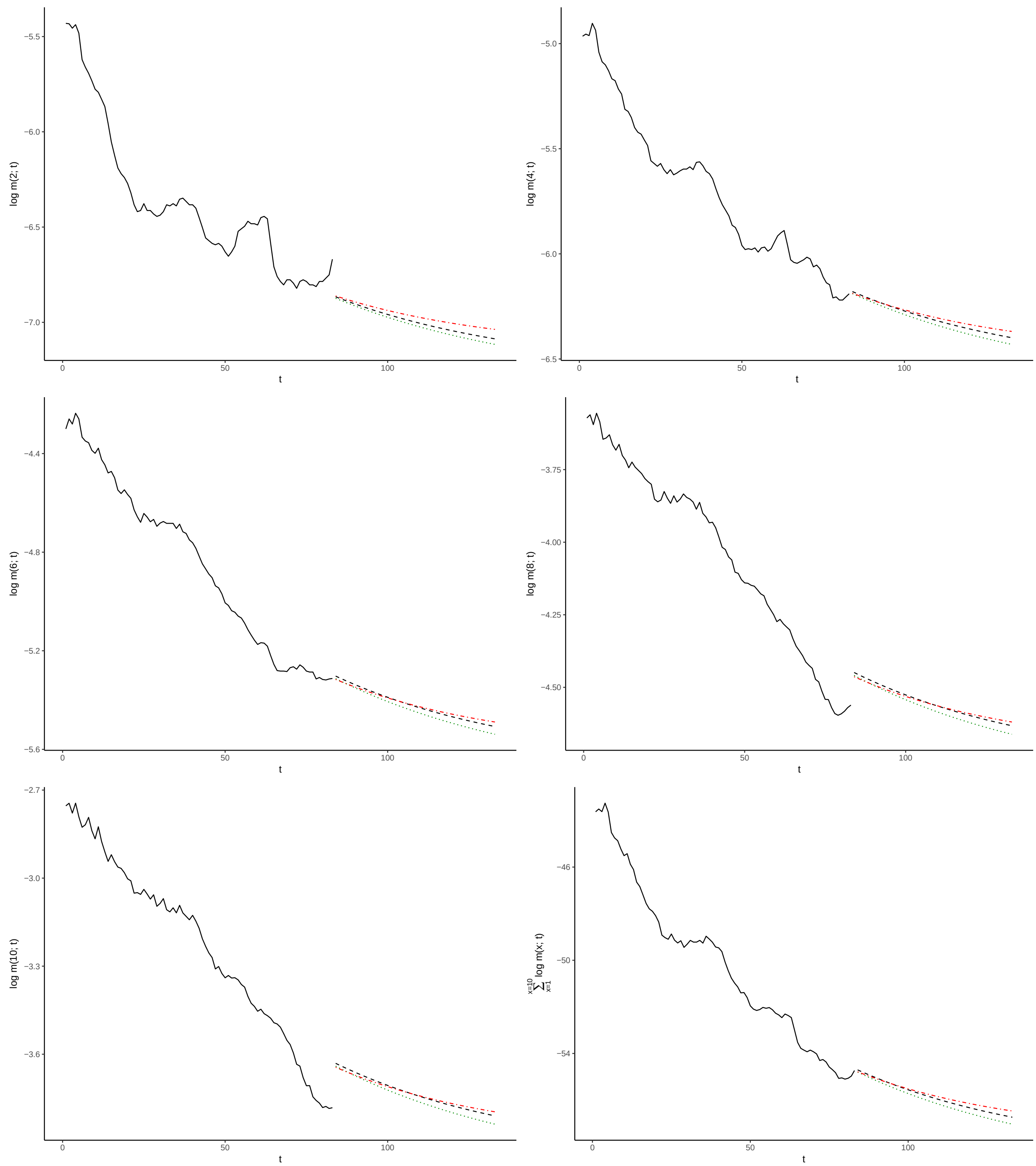
	x	1	2	3	4	5	6	7	8	9	10
Lee and Carter (1992)	$\hat{\alpha}_x$	-6.562	-6.381	-6.075	-5.697	-5.279	-4.853	-4.454	-4.045	-3.656	-3.237
	$\hat{\beta}_x$	0.106	0.109	0.112	0.108	0.103	0.101	0.094	0.090	0.089	0.088
Part 2	$\hat{\alpha}_x$	-1.264	-0.949	-0.452	-0.272	-0.105	0.219	0.293	0.509	0.815	1.205
	$\hat{\beta}_x$	0.105	0.108	0.112	0.108	0.103	0.101	0.094	0.091	0.089	0.088
Bias Corrected Inference	$\hat{\alpha}_x$	-1.495	-1.146	-0.576	-0.280	-0.083	0.304	0.430	0.613	0.939	1.294
	$\hat{\beta}_x$	0.101	0.104	0.110	0.108	0.103	0.103	0.097	0.093	0.091	0.090

	$\hat{\mu}$	$\hat{\phi}$
Lee and Carter (1992)	-0.135	0.983
Part 2	-0.906	0.985
Bias Corrected Inference	-1.095	0.981

for the next 50 years, and report the forecasts for the combined mortality rates in Figure 3.1. The bias corrected inference forecasts larger mortality rates than the method in Part 2. As Part 2 rejected the unit root null hypothesis for the combined mortality rates, only the forecasts based on the proposed bias corrected inference in Figure 3.1 are theoretically suitable.

Simulation Study This section investigates the finite sample performance of the proposed bias corrected estimators and compare them with the estimators in Part 2, which are inconsistent in the case of stationary mortality index.

Figure (3.1) True data and next 50 years' forecasts of $\log m(x,t)$ for US combined mortality rates. *The first five plots correspond to the age groups* $x = 2, 4, 6, 8, 10$ *and the last plot corresponds to the sum of* $\log m(x,t)$ *over all age groups* $x = 1, 2, \ldots, 10$*, where dashed line, dotted line and dashdotted line represent the Lee-Carter method, the method in Part 2 and the proposed bias corrected inference, respectively.*



We draw 10,000 random samples from model (3.3) with $M = 10$, α_x's, β_x's and μ being the estimates obtained from the female mortality rates based on the method in Part 2, which are given in the Table 3.1 of the Section 3.2. We further assume that the $\varepsilon_{x,t}$'s are independent random variables with normal distribution $N(0, \sigma_e^2/M)$ or $N(0, 5\sigma_e^2)$, e_t's are independent random variables with normal distribution $N(0, \sigma_e^2)$, while $\varepsilon_{x,t}$'s and e_t's are independent of each other. We take σ_e^2 as 0.047 ($= \tilde{\sigma}_e^2$) given in Table 2.7 in Part 2 which is the variance estimate of e_t, and consider sample size $T = 80$, 150, 300 and $\phi = 0.7$, 0.98. The simulation results for $T = 300$ are reported in Tables 3.4–3.6, which show that the bias corrected estimators have a larger standard error than the estimators in Part 2, but a smaller bias and a smaller mean squared error for the cases of $\phi = 0.7$ with $\varepsilon_{x,t} \sim N(0, \sigma_e^2/M)$ and $\phi = 0.98$ with $\varepsilon_{x,t} \sim N(0, 5\sigma_e^2)$, while both methods perform similar for the case of $\phi = 0.98$ with $\varepsilon_{x,t} \sim N(0, \sigma_e^2/M)$. Results in Tables 3.4 and 3.6 confirm that the method in Part 2 leads to inconsistent estimation when the mortality index is a stationary sequence. Results for $T = 80$ and 150 lead to similar conclusions, which are not reported here.

3.3 Conclusions

The Lee-Carter model in Lee & Carter (1992) has a restrictive constraint on the unobserved mortality index and suffers from possible inconsistent inference. Recently we proposed a modified Lee-Carter model without constraint on the mortality index and a consistent inference procedure when the mortality index is a near unit root or unit root AR(1) process with a nonzero intercept (as seen in Part 2). This section proposes a bias corrected inference, which is consistent with a normal limit regardless of whether the mortality index follows a stationary or near unit root or unit root AR(1) time series model with a nonzero intercept. This new inference is useful in forecasting future mortality rates as mortality index in real data may not be a unit root AR(1) process. Real data analysis does show that the bias corrected inference leads to larger forecasts of future mortality rates than the method in Part 2.

Appendix: Proofs of Theorems

For ease of notation, we use $Y_{t,j}, \tilde{Y}_{t,x}, \hat{Y}_{t,x}$ to denote $Y_{t,j}(\mu_0, \phi_0), \tilde{Y}_{t,x}(\alpha_{x,0}, \beta_{x,0}), \hat{Y}_{t,x}(\alpha_{x,0}, \beta_{x,0})$, respectively.

Lemma 3.1. *Suppose conditions in Theorem 3.1 hold.*

i) If $|\phi_0| < 1$ independent of T, then

$$\frac{1}{\sqrt{T}} \sum_{t=3}^{T} \boldsymbol{W}_t(\boldsymbol{\theta}_0) \xrightarrow{d} N(0, \Sigma) \quad \text{and} \quad \frac{1}{T} \sum_{t=3}^{T} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{W}_t(\boldsymbol{\theta}_0) \xrightarrow{p} \Gamma \quad \text{as} \quad T \to \infty.$$

Further if e_t is independent of $(\varepsilon_{1,t}, \cdots, \varepsilon_{M,t})^\tau$ and the covariance matrix of $(\varepsilon_{1,t}, \cdots, \varepsilon_{M,t})^\tau$ is positive definite, then we can write $\Sigma = \begin{pmatrix} \Sigma_1, & \Sigma_{12} \\ \Sigma_{12}^\tau & \Sigma_2 \end{pmatrix}$, where Σ_1 is a 2×2 positive definite matrix and Σ_2 is a $(2M-2) \times (2M-2)$ positive definite matrix.

ii) If $\phi_0 = 1 + \rho/T$ for some constant $\rho \in \mathbb{R}$, then

$$\frac{1}{\sqrt{T}} \sum_{t=3}^{T} \widetilde{\boldsymbol{W}}_t(\boldsymbol{\theta}_0) \xrightarrow{d} N(0, \widetilde{\Sigma}) \quad \text{and} \quad \frac{1}{T} \sum_{t=3}^{T} \frac{\partial}{\partial \boldsymbol{\theta}} \widetilde{\boldsymbol{W}}_t(\boldsymbol{\theta}_0) \xrightarrow{p} \widetilde{\Gamma} \quad \text{as} \quad T \to \infty.$$

Further if e_t is independent of $(\varepsilon_{1,t}, \cdots, \varepsilon_{M,t})^\tau$ and the covariance matrix of $(\varepsilon_{1,t}, \cdots, \varepsilon_{M,t})^\tau$ is positive definite, then we can write $\widetilde{\Sigma} = \begin{pmatrix} \widetilde{\Sigma}_1, & 0 \\ 0 & \widetilde{\Sigma}_2 \end{pmatrix}$, where $\widetilde{\Sigma}_1$ is a 2×2 positive definite matrix and $\widetilde{\Sigma}_2$ is a $(2M-2) \times (2M-2)$ positive definite matrix.

Proof. i) Note that for $x = 1, \ldots, M-1$,

$$\begin{cases} Y_{t,1} = e_t + \eta_t - \phi_0 \eta_{t-1}, \ Y_{t,2} = (e_t + \eta_t - \phi_0 \eta_{t-1})(k_{t-2} + \eta_{t-2}), \\ \tilde{Y}_{t,x} = \varepsilon_{x,t} - \beta_{x,0} \eta_t, \ \hat{Y}_{t,x} = (\varepsilon_{x,t} - \beta_{x,0} \eta_t)(k_{t-1} + \eta_{t-1}). \end{cases} \tag{3.9}$$

Define $Y_{t,1}^* = e_t + \eta_t - \phi_0 \eta_t$,

$$\begin{aligned} Y_{t,2}^* &= (e_t + \eta_t)(k_{t-2} + \eta_{t-2}) - \phi_0 \eta_t (k_{t-1} + \eta_{t-1}) \\ &= \{e_t + (1 - \phi_0)\eta_t\}(k_{t-2} + \eta_{t-2}) - \phi_0 \eta_t (k_{t-1} - k_{t-2} + \eta_{t-1} - \eta_{t-2}) \end{aligned}$$

and $\boldsymbol{W}_t^*(\boldsymbol{\theta}_0)$ as $\boldsymbol{W}_t(\boldsymbol{\theta}_0)$ with $Y_{t,1}$ and $Y_{t,2}$ replaced by $Y_{t,1}^*$ and $Y_{t,2}^*$, respectively. Let $\mathcal{F}_t$ denote the σ-field generated by $\{(e_s, \varepsilon_{1,s}, \cdots, \varepsilon_{M,s})^\tau : s \le t\}$. Then $\{\boldsymbol{W}_t^*(\boldsymbol{\theta}_0), \mathcal{F}_t\}_{t=1}^\infty$ is a martingale difference sequence. By the central limit theorem for a martingale difference sequence in Hall & Heyde (2014), $\frac{1}{\sqrt{T}}\sum_{t=3}^T \boldsymbol{W}_t^*(\boldsymbol{\theta}_0)$ converges in distribution to a multivariate normal distribution with zero means and covariance matrix $\Sigma = E\left\{\boldsymbol{W}_3^*(\boldsymbol{\theta}_0)\boldsymbol{W}_3^{*\tau}(\boldsymbol{\theta}_0)\right\}$. Since $\frac{1}{\sqrt{T}}\sum_{t=3}^T Y_{t,1} = \frac{1}{\sqrt{T}}\sum_{t=3}^T Y_{t,1}^* + o_p(1)$ and $\frac{1}{\sqrt{T}}\sum_{t=3}^T Y_{t,2} = \frac{1}{\sqrt{T}}\sum_{t=3}^T Y_{t,2}^* + o_p(1)$, we have

$$\frac{1}{\sqrt{T}}\sum_{t=3}^T \boldsymbol{W}_t(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{T}}\sum_{t=3}^T \boldsymbol{W}_t^*(\boldsymbol{\theta}_0) + o_p(1) \xrightarrow{d} N(0, \Sigma) \text{ as } T \to \infty.$$

In order to compute the covariance matrix $\Sigma = (\sigma_{i,j})_{1\le i,j\le 2M}$, by noting that

$$E(k_t) = \frac{\mu_0}{1-\phi_0},\ E(k_t^2) = \frac{\mu_0^2}{(1-\phi_0)^2} + \frac{E(e_1^2)}{1-\phi_0^2} \text{ and } E(k_t k_{t-1}) = \frac{\mu_0^2}{(1-\phi_0)^2} + \phi_0\frac{E(e_1^2)}{1-\phi_0^2},$$

it is straightforward to verify that for $1 \le x, y \le M-1$

$$\sigma_{1,1} = E(e_3 + (1-\phi_0)\eta_3)^2,\ \sigma_{1,2} = E\{(e_3 + (1-\phi_0)\eta_3\})^2\}E(k_1+\eta_1) = \frac{\mu_0}{1-\phi_0}\sigma_{1,1},$$

$$\sigma_{1,2x+1} = E\{(e_3 + (1-\phi_0)\eta_3)(\varepsilon_{x,3} - \beta_{x,0}\eta_3)\},$$

$$\sigma_{1,2x+2} = E\{(e_3 + (1-\phi_0)\eta_3)(\varepsilon_{x,3} - \beta_{x,0}\eta_3)(k_2+\eta_2)\} = \frac{\mu_0}{1-\phi_0}\sigma_{1,2x+1},$$

$$\begin{aligned}\sigma_{2,2} &= E\{(e_3+\eta_3)^2\}E\{(k_1+\eta_1)^2\} + \phi_0^2 E\{\eta_3^2\}E\{(k_2+\eta_2)^2\} \\ &\quad -2\phi_0 E\{\eta_3 e_3 + \eta_3^2\}E\{(k_1+\eta_1)(k_2+\eta_2)\} \\ &= \{E(e_1+\eta_1)^2 + \phi_0^2 E(\eta_1^2)\}\{\tfrac{\mu_0^2}{(1-\phi_0)^2} + \tfrac{E(e_1^2)}{1-\phi_0^2} + E(\eta_1^2) + 2E(e_1\eta_1)\} \\ &\quad -2\phi_0\{E(e_1\eta_1) + E(\eta_1^2)\}\{\tfrac{\mu_0^2}{(1-\phi_0)^2} + \tfrac{\phi_0 E(e_1^2)}{1-\phi_0^2} + \phi_0 E(e_1\eta_1)\},\end{aligned}$$

$$\sigma_{2,2x+1} = E\{(e_3 + (1-\phi_0)\eta_3)(\varepsilon_{x,3} - \beta_{x,0}\eta_3)\}E\{k_1+\eta_1\} = \sigma_{1,2x+2},$$

$$\begin{aligned}\sigma_{2,2x+2} &= E\{(e_3+\eta_3)(\varepsilon_{x,3}-\beta_{x,0}\eta_3)\}E\{(k_1+\eta_1)(k_2+\eta_2)\} \\ &\quad -\phi_0 E\{\eta_3(\varepsilon_{x,3}-\beta_{x,0}\eta_3)\}E(k_2+\eta_2)^2 \\ &= E\{e_1(\varepsilon_{x,1}-\beta_{x,0}\eta_1)\}\{\tfrac{\mu_0^2}{(1-\phi_0)^2}+\tfrac{\phi_0 E(e_1^2)}{1-\phi_0^2}+\phi_0 E(e_1\eta_1)\} \\ &\quad +E\{\eta_1(\varepsilon_{x,1}-\beta_{x,0}\eta_1)\}\{\tfrac{\mu_0^2}{1-\phi_0}-\phi_0 E(e_1\eta_1)-\phi_0 E(\eta_1^2)\},\end{aligned}$$

$$\sigma_{2x+1,2y+1} = E\{(\varepsilon_{x,3}-\beta_{x,0}\eta_3)(\varepsilon_{y,3}-\beta_{y,0}\eta_3)\},$$

$$\sigma_{2x+1,2y+2} = E\{(\varepsilon_{x,3}-\beta_{x,0}\eta_3)(\varepsilon_{y,3}-\beta_{y,0}\eta_3)\}E(k_2+\eta_2) = \tfrac{\mu_0}{1-\phi_0}\sigma_{2x+1,2y+1},$$

$$\sigma_{2x+2,2y+2} = E\{(\varepsilon_{x,3}-\beta_{x,0}\eta_3)(\varepsilon_{y,3}-\beta_{y,0}\eta_3)\}\{\tfrac{\mu_0^2}{(1-\phi_0)^2}+\tfrac{E(e_2^2)}{1-\phi_0^2}+2E(e_2\eta_2)+E(\eta_2^2)\}.$$

Note that

$$\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{W}_t(\boldsymbol{\theta}_0) = \begin{pmatrix} A_{t0} & 0 & \dots & 0 \\ 0 & A_{t1} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & A_{t(M-1)} \end{pmatrix},$$

where for $x = 1, \dots, M-1$

$$A_{t0} = -\begin{pmatrix} 1 & Z_{t-1} \\ Z_{t-2} & Z_{t-1}Z_{t-2} \end{pmatrix} \text{ and } A_{tx} = -\begin{pmatrix} 1 & Z_t \\ Z_{t-1} & Z_t Z_{t-1} \end{pmatrix}.$$

As $\{k_t\}$ is stationary, we have

$$\frac{1}{T}\sum_{t=3}^{T} k_t \xrightarrow{p} \frac{\mu_0}{1-\phi_0} \text{ and } \frac{1}{T}\sum_{t=3}^{T} k_t k_{t-1} \xrightarrow{p} \frac{\mu_0^2}{(1-\phi_0)^2} + \frac{\phi_0 E(e_1^2)}{1-\phi_0^2}$$

as $T \to \infty$ (see Lemma 1 in Leng & Peng (2016) too). Therefore we can show that

$$\frac{1}{T}\sum_{t=3}^{T} Z_{t-1} = \frac{1}{T}\sum_{t=3}^{T} k_{t-1} + \frac{1}{T}\sum_{t=3}^{T} \eta_{t-1} \xrightarrow{p} \frac{\mu_0}{1-\phi_0}$$

and

$$\begin{aligned}
&\tfrac{1}{T}\textstyle\sum_{t=3}^{T} Z_{t-1}Z_{t-2} \\
&= \tfrac{1}{T}\textstyle\sum_{t=3}^{T}(k_{t-1}+\eta_{t-1})(k_{t-2}+\eta_{t-2}) \\
&= \tfrac{1}{T}\textstyle\sum_{t=3}^{T} k_{t-1}k_{t-2} + \tfrac{1}{T}\sum_{t=3}^{T}\eta_{t-1}(k_{t-2}+\eta_{t-2}) + \tfrac{1}{T}\sum_{t=3}^{T}\eta_{t-2}(\mu_0+\phi_0 k_{t-2}+e_{t-1}) \\
&= \tfrac{1}{T}\textstyle\sum_{t=3}^{T} k_{t-1}k_{t-2} + \tfrac{1}{T}\sum_{t=3}^{T}\eta_{t-1}(k_{t-2}+\eta_{t-2}) + \mu_0(1+\phi_0)\tfrac{1}{T}\sum_{t=3}^{T}\eta_{t-2} \\
&\quad +\phi_0^2\tfrac{1}{T}\textstyle\sum_{t=3}^{T}\eta_{t-2}k_{t-3} + \tfrac{1}{T}\sum_{t=3}^{T} e_{t-1}\eta_{t-2} + \phi_0\tfrac{1}{T}\sum_{t=3}^{T} e_{t-2}\eta_{t-2} \\
&\xrightarrow{p} \tfrac{\mu_0^2}{(1-\phi_0)^2} + \tfrac{\phi_0 E(e_1^2)}{1-\phi_0^2} + \phi_0 E(e_1\eta_1),
\end{aligned}$$

since these three terms $\frac{1}{T}\sum_{t=3}^{T}\eta_{t-1}(k_{t-2}+\eta_{t-2})$, $\frac{1}{T}\sum_{t=3}^{T}\eta_{t-2}k_{t-3}$, $\frac{1}{T}\sum_{t=3}^{T}e_{t-1}\eta_{t-2}$ converge in probability to zero by the law of large numbers for a martingale difference sequence in Hall & Heyde (2014). That is,

$$\frac{1}{T}\sum_{t=3}^{T}\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{W}_t(\boldsymbol{\theta}_0) \xrightarrow{p} -\Gamma \text{ as } T\to\infty,$$

where Γ is defined in (3.6).

Since e_t is independent of $(\varepsilon_{1,t},\cdots,\varepsilon_{M,t})^\tau$, we have

$$\sigma_{1,1}\sigma_{2,2}-\sigma_{1,2}^2 = \{Ee_1^2+(1-\phi_0)^2E\eta_1^2\}\{\frac{(Ee_1^2)^2}{1-\phi_0^2}+2E(\eta_1^2)E(e_1^2)+(1+\phi_0^2)(E\eta_1^2)^2\}>0,$$

which implies that Σ_1 is positive definite.

Put $\tilde{\boldsymbol{Y}}_t=(\tilde{Y}_{t,1},\cdots,\tilde{Y}_{t,M-1})^\tau$, $\hat{\boldsymbol{Y}}_t=(\hat{Y}_{t,1},\cdots,\hat{Y}_{t,M-1})^\tau$ and $\boldsymbol{Y}_t=(\tilde{\boldsymbol{Y}}_t^\tau,\hat{\boldsymbol{Y}}_t^\tau)^\tau$. As the covariance matrix of $(\varepsilon_{1,t},\cdots,\varepsilon_{M,t})^\tau$ is positive definite, it is easy to see that the covariance matrix Ω of $\tilde{\boldsymbol{Y}}_t$ is positive definite, i.e., $|\Omega|\neq 0$. It is straightforward to compute the covariance matrix of $\boldsymbol{Y}_t$, which is $\begin{pmatrix}\Omega & \frac{\mu_0}{1-\phi_0}\Omega \\ \frac{\mu_0}{1-\phi_0}\Omega & (\frac{\mu_0^2}{(1-\phi_0)^2}+\frac{Ee_1^2}{1-\phi_0^2}+E\eta_1^2)\Omega\end{pmatrix}$ with determinant $|\Omega|^2(\frac{Ee_1^2}{1-\phi_0^2}+E\eta_1^2)^{M-1}>0$, i.e., the covariance matrix of $\boldsymbol{Y}$ is positive definite, which is equivalent to that Σ_2 is positive definite. Hence Lemma 3.1i) follows.

ii) When $\phi_0 = 1 + \rho/T$ for some constant $\rho \in \mathbb{R}$ and $\mu_0 \neq 0$, we have

$$k_t = \mu_0 \left(\sum_{j=0}^{t-1} \phi_0^j \right) + \phi_0^t k_0 + \sum_{i=1}^{t} \phi_0^{t-i} e_i,$$

and it follows from Chan & Wei (1987) that for $s \in [0,1]$

$$\sum_{i=1}^{[Ts]} \phi_0^{[Ts]-i} e_i = \phi_0^{[Ts]-T} \sum_{i=1}^{[Ts]} \phi_0^{T-i} e_i = O_p(\sqrt{T}),$$

which imply that, as $T \to \infty$

$$k_{[Ts]}/T \xrightarrow{p} f_{\rho,\mu_0}(s) \text{ for } s \in [0,1], \tag{3.10}$$

where $f_{\rho,\mu_0}(s)$ is defined in (3.7). By (3.9) and (3.10), we have

$$\frac{1}{\sqrt{T}} \sum_{t=3}^{T} \widetilde{\boldsymbol{W}}_t(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{T}} \sum_{t=3}^{T} \widetilde{\boldsymbol{U}}_t(\boldsymbol{\theta}_0) + o_p(1),$$

where $\widetilde{\boldsymbol{U}}_t(\boldsymbol{\theta}_0) = (V_{t,1}, V_{t,2}, \tilde{V}_{t,1}, \hat{V}_{t,1}, \ldots, \tilde{V}_{t,M-1}, \hat{V}_{t,M-1})^\tau$ and for $x = 1, \ldots, M-1$

$$\begin{cases} V_{t,1} = e_t + \eta_t - \eta_{t-1},\ V_{t,2} = (e_t + \eta_t - \eta_{t-1}) \frac{k_{t-2}}{T}, \\ \tilde{V}_{t,x} = \varepsilon_{x,t} - \beta_{x,0}\eta_t,\ \hat{V}_{t,x} = (\varepsilon_{x,t} - \beta_{x,0}\eta_t) \frac{k_{t-1}}{T}. \end{cases} \tag{3.11}$$

Similar to the proof for the stationary case, define $V_{t,1}^* = e_t$, $V_{t,2}^* = (e_t + \eta_t)\frac{k_{t-2}}{T} - \eta_t \frac{k_{t-1}}{T}$,

$$\widetilde{\boldsymbol{U}}_t^*(\boldsymbol{\theta}_0) = (V_{t,1}^*, V_{t,2}^*, \tilde{V}_{t,1}, \hat{V}_{t,1}, \ldots, \tilde{V}_{t,M-1}, \hat{V}_{t,M-1})^\tau,$$

and it follows from the central limit theorem for a martingale difference sequence in Hall & Heyde (2014) that $\frac{1}{\sqrt{T}} \sum_{t=3}^{T} \widetilde{\boldsymbol{U}}_t^*(\boldsymbol{\theta}_0) \xrightarrow{d} N(0, \tilde{\Sigma})$, where

$$\tilde{\Sigma} = (\tilde{\sigma}_{i,j})_{1 \le i,j \le 2M} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=3}^{T} E\left\{ \widetilde{\boldsymbol{U}}_t^*(\boldsymbol{\theta}_0) \widetilde{\boldsymbol{U}}_t^{*\tau}(\boldsymbol{\theta}_0) \right\}.$$

Since $\frac{1}{\sqrt{T}}\sum_{t=3}^{T}V_{t,1}=\frac{1}{\sqrt{T}}\sum_{t=3}^{T}V_{t,1}^{*}+o_p(1)$ and $\frac{1}{\sqrt{T}}\sum_{t=3}^{T}V_{t,2}=\frac{1}{\sqrt{T}}\sum_{t=3}^{T}V_{t,2}^{*}+o_p(1)$, we have

$$\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\widetilde{\boldsymbol{U}}_t(\boldsymbol{\theta}_0)=\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\widetilde{\boldsymbol{U}}_t^{*}(\boldsymbol{\theta}_0)+o_p(1)\xrightarrow{d}N(0,\tilde{\Sigma})\text{ as }T\to\infty.$$

It is straightforward to verify that, for $x,y=1,2,\ldots,M-1$, $\tilde{\sigma}_{1,1}=E(e_1^2)$,

$$\tilde{\sigma}_{1,2}=E(e_1^2+e_1\eta_1)\lim_{T\to\infty}\frac{1}{T}\sum_{t=3}^{T}\frac{k_{t-2}}{T}-E(e_1\eta_1)\lim_{T\to\infty}\frac{1}{T}\sum_{t=3}^{T}\frac{k_{t-1}}{T}=E(e_1^2)\int_0^1 f_{\rho,\mu_0}(s)\,ds,$$

$$\tilde{\sigma}_{1,2x+1}=E(e_1\varepsilon_{x,1})-\beta_{x,0}E(e_1\eta_1),$$

$$\tilde{\sigma}_{1,2x+2}=E(e_1\varepsilon_{x,1}-\beta_{x,0}e_1\eta_1)\lim_{T\to\infty}\frac{1}{T}\sum_{t=3}^{T}\frac{k_{t-1}}{T}=\{E(e_1\varepsilon_{x,1})-\beta_{x,0}E(e_1\eta_1)\}\int_0^1 f_{\rho,\mu_0}(s)\,ds,$$

$$\begin{aligned}\tilde{\sigma}_{2,2}\ &=E\{(e_1+\eta_1)^2\}\lim_{T\to\infty}\tfrac{1}{T}\textstyle\sum_{t=3}^{T}\frac{k_{t-2}^2}{T^2}+E(\eta_1^2)\lim_{T\to\infty}\frac{1}{T}\sum_{t=3}^{T}\frac{k_{t-1}^2}{T^2}\\&\quad-2E(e_1\eta_1+\eta_1^2)\lim_{T\to\infty}\tfrac{1}{T}\textstyle\sum_{t=3}^{T}\frac{k_{t-1}k_{t-2}}{T^2}\\&=E(e_1^2)\textstyle\int_0^1 f_{\rho,\mu_0}^2(s)\,ds,\end{aligned}$$

$$\begin{aligned}\tilde{\sigma}_{2,2x+1}\ &=E\{(e_1+\eta_1)(\varepsilon_{x,1}-\beta_{x,0}\eta_1)\}\lim_{T\to\infty}\tfrac{1}{T}\textstyle\sum_{t=3}^{T}\frac{k_{t-2}}{T}\\&\quad-E\{\eta_1(\varepsilon_{x,1}-\beta_{x,0}\eta_1)\}\lim_{T\to\infty}\tfrac{1}{T}\textstyle\sum_{t=3}^{T}\frac{k_{t-1}}{T}\\&=\{E(e_1\varepsilon_{x,1})-\beta_{x,0}E(e_1\eta_1)\}\textstyle\int_0^1 f_{\rho,\mu_0}(s)\,ds,\end{aligned}$$

$$\begin{aligned}\tilde{\sigma}_{2,2x+2}\ &=E\{(e_1+\eta_1)(\varepsilon_{x,1}+\beta_{x,0}\eta_1)\}\lim_{T\to\infty}\tfrac{1}{T}\textstyle\sum_{t=3}^{T}\frac{k_{t-1}k_{t-2}}{T^2}\\&\quad-E\{\eta_1(\varepsilon_{x,1}+\beta_{x,0}\eta_1)\}\lim_{T\to\infty}\tfrac{1}{T}\textstyle\sum_{t=3}^{T}\frac{k_{t-1}^2}{T^2}\\&=\{E(e_1\varepsilon_{x,1})-\beta_{x,0}E(e_1\eta_1)\}\textstyle\int_0^1 f_{\rho,\mu_0}^2(s)\,ds,\end{aligned}$$

$$\tilde{\sigma}_{2x+1,2y+1}=E\{(\varepsilon_{x,1}-\beta_{x,0}\eta_1)(\varepsilon_{y,1}-\beta_{y,0}\eta_1)\},$$

$$\tilde{\sigma}_{2x+1,2y+2}=E\{(\varepsilon_{x,1}-\beta_{x,0}\eta_1)(\varepsilon_{y,1}-\beta_{y,0}\eta_1)\}\int_0^1 f_{\rho,\mu_0}(s)\,ds,$$

$$\tilde{\sigma}_{2x+2,2y+2}=E\{(\varepsilon_{x,1}-\beta_{x,0}\eta_1)(\varepsilon_{y,1}-\beta_{y,0}\eta_1)\}\int_0^1 f_{\rho,\mu_0}^2(s)\,ds.$$

Note that

$$\frac{\partial}{\partial \boldsymbol{\theta}} \widetilde{\boldsymbol{W}}_t(\boldsymbol{\theta}_0) = \begin{pmatrix} \tilde{A}_{t0} & 0 & \dots & 0 \\ 0 & \tilde{A}_{t1} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \tilde{A}_{t(M-1)} \end{pmatrix},$$

where for $x = 1, \dots, M-1$

$$\tilde{A}_{t0} = -\begin{pmatrix} 1 & Z_{t-1} \\ \frac{1}{T} Z_{t-2} & \frac{1}{T} Z_{t-1} Z_{t-2} \end{pmatrix} \text{ and } \tilde{A}_{tx} = -\begin{pmatrix} 1 & Z_t \\ \frac{1}{T} Z_{t-1} & \frac{1}{T} Z_t Z_{t-1} \end{pmatrix}.$$

It follows from (3.10) that as $T \to \infty$

$$\frac{1}{T^2} \sum_{t=3}^{T} Z_{t-1} = \frac{1}{T^2} \sum_{t=3}^{T} k_{t-1} + o_p(1) \xrightarrow{p} \int_0^1 f_{\rho,\mu_0}(s)\, ds$$

and

$$\frac{1}{T^3} \sum_{t=3}^{T} Z_{t-1} Z_{t-2} = \frac{1}{T^3} \sum_{t=3}^{T} k_{t-1}^2 + o_p(1) \xrightarrow{p} \int_0^1 f_{\rho,\mu_0}^2(s)\, ds,$$

which imply that

$$\frac{1}{T} \{\sum_{t=3}^{T} \frac{\partial}{\partial \boldsymbol{\theta}} \widetilde{\boldsymbol{W}}_t(\boldsymbol{\theta}_0)\} \{\sqrt{T} D_T^{-1}\} \xrightarrow{p} -\tilde{\Gamma} \text{ as } T \to \infty,$$

where $\tilde{\Gamma}$ is defined in (3.8).

Since e_t is independent of $(\varepsilon_{1,t}, \cdots, \varepsilon_{M,t})^\tau$, we have $\tilde{\sigma}_{1,2x+1} = \tilde{\sigma}_{1,2x+2} = \tilde{\sigma}_{2,2x+1} = \tilde{\sigma}_{2,2x+2} = 0$ for $x = 1, \dots, M-1$ and

$$\tilde{\sigma}_{1,1} \tilde{\sigma}_{2,2} - \tilde{\sigma}_{1,2}^2 = (E e_1^2)^2 \{\int_0^1 f_{\rho,\mu_0}^2(s)\, ds - (\int_0^1 f_{\rho,\mu_0}(s)\, ds)^2\} > 0,$$

we can write $\widetilde{\Sigma} = \begin{pmatrix} \widetilde{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \widetilde{\Sigma}_2 \end{pmatrix}$ with $\widetilde{\Sigma}_1$ being positive definite.

Define $\tilde{\boldsymbol{V}}_t = (\tilde{V}_{t,1}, \cdots, \tilde{V}_{t,M-1})^\tau$, $\hat{\boldsymbol{V}}_t = (\hat{V}_{t,1}, \cdots, \hat{V}_{t,M-1})^\tau$ and $\boldsymbol{V}_t = (\tilde{\boldsymbol{V}}_t^\tau, \hat{\boldsymbol{V}}_t^\tau)^\tau$. As the covariance matrix of $(\varepsilon_{1,t},\cdots,\varepsilon_{M,t})^\tau$ is positive definite, it is easy to see that the covariance matrix $\tilde\Omega$ of $\tilde{\boldsymbol V}_t$ is positive definite, i.e., $|\tilde\Omega|\neq 0$. It is straightforward to compute the covariance matrix of $\boldsymbol V_t$, which is $\begin{pmatrix}\tilde\Omega & \tilde\Omega\int_0^1 f_{\rho,\mu_0}(s)\,ds\\ \tilde\Omega\int_0^1 f_{\rho,\mu_0}(s)\,ds & \tilde\Omega\int_0^1 f^2_{\rho,\mu_0}(s)\,ds\end{pmatrix}$ with determinant $|\tilde\Omega|^2\{\int_0^1 f^2_{\rho,\mu_0}(s)\,ds-(\int_0^1 f_{\rho,\mu_0}(s)\,ds)^2\}^{M-1}>0$, i.e., the covariance matrix of $\boldsymbol V_t$ is positive definite, which is equivalent to that $\widetilde\Sigma_2$ is positive definite. Hence Lemma 3.1ii) holds.

□

Proof of Theorem 3.1. It is easy to know that

$$0=\sum_{t=3}^T \boldsymbol W_t(\hat{\boldsymbol\theta})=\sum_{t=3}^T \boldsymbol W_t(\boldsymbol\theta_0)+\sum_{t=3}^T \frac{\partial}{\partial\boldsymbol\theta}\boldsymbol W_t(\boldsymbol\theta_0)(\hat{\boldsymbol\theta}-\boldsymbol\theta_0)$$

and

$$0=\sum_{t=3}^T \widetilde{\boldsymbol W}_t(\hat{\boldsymbol\theta})=\sum_{t=3}^T \widetilde{\boldsymbol W}_t(\boldsymbol\theta_0)+\sum_{t=3}^T \frac{\partial}{\partial\boldsymbol\theta}\widetilde{\boldsymbol W}_t(\boldsymbol\theta_0)(\hat{\boldsymbol\theta}-\boldsymbol\theta_0),$$

which imply the theorem by using Lemma 3.1. □

Proof of Theorem 3.2. The theorem can be proved straightforwardly by noting that

$$E\{\boldsymbol W_s(\boldsymbol\theta_0)\boldsymbol W_t^\tau(\boldsymbol\theta_0)\}=\mathbf 0 \text{ for } |s-t|\geq 2.$$

Hence we skip details. □

Table (3.4) Simulation results for $T = 300$, $\phi = 0.7$ and $\varepsilon_{x,t} \sim N(0, \sigma_e^2/M)$. Standard errors are given in brackets.

	x	1	2	3	4	5	6	7	8	9	10
Part 2	$\hat{\alpha}_x$	0.135	0.026	-0.042	-0.350	-0.469	-0.317	-0.319	-0.046	0.406	0.977
		(0.039)	(0.039)	(0.039)	(0.038)	(0.039)	(0.039)	(0.038)	(0.038)	(0.039)	(0.039)
	$\hat{\beta}_x$	0.127	0.121	0.115	0.104	0.096	0.093	0.087	0.085	0.086	0.087
		(8.374e-3)	(8.352e-3)	(8.390e-3)	(8.274e-3)	(8.337e-3)	(8.391e-3)	(8.256e-3)	(8.342e-3)	(8.375e-3)	(8.337e-3)
Bias Corrected Inference	$\hat{\alpha}_x$	0.173	0.056	-0.022	-0.344	-0.475	-0.326	-0.337	-0.067	0.385	0.957
		(0.087)	(0.087)	(0.087)	(0.087)	(0.087)	(0.088)	(0.088)	(0.087)	(0.088)	(0.087)
	$\hat{\beta}_x$	0.135	0.127	0.119	0.106	0.095	0.091	0.083	0.080	0.081	0.083
		(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)

		Estimate	Standard Error
Part 2	$\hat{\mu}$	-2.155	0.211
	$\hat{\phi}$	0.533	0.046
Bias Corrected Inference	$\hat{\mu}$	-1.425	0.323
	$\hat{\phi}$	0.692	0.070

Table (3.5) Simulation results for $T = 300$, $\phi = 0.98$ and $\varepsilon_{x,t} \sim N(0, \sigma_e^2/M)$. Standard errors are given in brackets.

	x	1	2	3	4	5	6	7	8	9	10
Part 2	$\hat{\alpha}_x$	0.171	0.055	-0.022	-0.344	-0.474	-0.327	-0.337	-0.067	0.385	0.959
		(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
	$\hat{\beta}_x$	0.135	0.127	0.119	0.106	0.096	0.091	0.083	0.080	0.081	0.083
		(2.299e-4)	(2.345e-4)	(2.332e-4)	(2.298e-4)	(2.309e-4)	(2.278e-4)	(2.309e-4)	(2.294e-4)	(2.339e-4)	(2.316e-4)
Bias Corrected Inference	$\hat{\alpha}_x$	0.172	0.055	-0.022	-0.344	-0.474	-0.327	-0.337	-0.067	0.384	0.959
		(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.014)	(0.015)	(0.015)	(0.015)	(0.015)
	$\hat{\beta}_x$	0.135	0.127	0.119	0.106	0.096	0.091	0.083	0.080	0.081	0.083
		(2.401e-4)	(2.426e-4)	(2.430e-4)	(2.409e-4)	(2.400e-4)	(2.383e-4)	(2.410e-4)	(2.406e-4)	(2.440e-4)	(2.420e-4)

		Estimate	Standard Error
Part 2	$\hat{\mu}$	-1.403	0.047
	$\hat{\phi}$	0.980	7.822e-4
Bias Corrected Inference	$\hat{\mu}$	-1.392	0.048
	$\hat{\phi}$	0.980	7.951e-4

Table (3.6) Simulation results for $T = 300$, $\phi = 0.98$ and $\varepsilon_{x,t} \sim N(0, 5\sigma_e^2)$. Standard errors are given in brackets.

	x	1	2	3	4	5	6	7	8	9	10
Part 2	$\hat{\alpha}_x$	0.089	-0.015	-0.068	-0.361	-0.466	-0.305	-0.294	-0.015	0.431	1.003
		(0.217)	(0.214)	(0.219)	(0.218)	(0.217)	(0.217)	(0.214)	(0.216)	(0.214)	(0.218)
	$\hat{\beta}_x$	0.133	0.126	0.118	0.105	0.096	0.091	0.084	0.081	0.082	0.084
		(3.600e-3)	(3.547e-3)	(3.620e-3)	(3.600e-3)	(3.556e-3)	(3.572e-3)	(3.530e-3)	(3.581e-3)	(3.534e-3)	(3.620e-3)
Bias Corrected Inference	$\hat{\alpha}_x$	0.175	0.056	-0.019	-0.347	-0.478	-0.328	-0.336	-0.064	0.383	0.958
		(0.237)	(0.235)	(0.237)	(0.237)	(0.237)	(0.235)	(0.234)	(0.235)	(0.235)	(0.237)
	$\hat{\beta}_x$	0.135	0.127	0.119	0.106	0.096	0.091	0.083	0.080	0.081	0.083
		(3.924e-3)	(3.893e-3)	(3.908e-3)	(3.921e-3)	(3.895e-3)	(3.875e-3)	(3.865e-3)	(3.897e-3)	(3.880e-3)	(3.932e-3)

		Estimate	Standard Error
Part 2	$\hat{\mu}$	-3.827	0.311
	$\hat{\phi}$	0.938	5.324e-3
Bias Corrected Inference	$\hat{\mu}$	-1.390	0.271
	$\hat{\phi}$	0.980	4.547e-3

PART 4

INFERENCE FOR THE LEE-CARTER MODEL WITH AN AR(2) PROCESS FOR UNOBSERVED MORTALITY INDEXES

This Part is my working paper which is currently under review, and has been adapted to the format of dissertation.

For pension funds, an important and challenging step to hedge longevity risk is to understand, model, and forecast mortality rates, given the rapid changes in environment and technology of our society. A benchmark mortality model is the so-called Lee-Carter model in Lee & Carter (1992), which models the central death rate $m(x,t)$ at age or age group $x = 1, \ldots, M$ and time $t = 1, \ldots, T$ by

$$\log m(x,t) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}, \quad \sum_{x=1}^{M} \beta_x = 1, \quad \sum_{t=1}^{T} k_t = 0, \tag{4.1}$$

and

$$k_t = \mu + \rho k_{t-1} + e_t, \tag{4.2}$$

where $\{\varepsilon_{x,t}\}_{t=1}^{T}$ and $\{e_t\}_{t=1}^{T}$ are random errors with zero means and finite variances, and the unobserved $\{k_t\}$ is called the mortality index. Note that the two constraints in (4.1) ensure that the model is identifiable. Although a general model for $\{k_t\}$ can be used, researchers in studying longevity risk often employ a unit root AR(1) model, i.e., $\rho = 1$ in (4.2); see, for example, Biffis et al. (2017), Enchev et al. (2017), Kwok et al. (2016), and Wong et al. (2017).

A widely used statistical inference in fitting models (4.1) and (4.2) is the two-step procedure in Lee & Carter (1992), which first estimates α_x and β_x for $x = 1, \ldots, M$ and k_t for $t = 1, \ldots, T$ by the singular value decomposition method, and then fits the time series model based on the estimated k_t's in the first step. Recently Leng & Peng (2016) pointed out that this two-step inference leads to an inconsistent inference when model (4.2) holds with $\rho \neq 1$ or $\{k_t\}$ is modeled by an AR(p) process with $p > 1$. As k_t's are unobservable, existing unit root tests in econometrics may not be applicable here. An application of the unit root test developed in Leng & Peng (2017) for models (4.1) and (4.2) shows that some mortality datasets reject the unit root hypothesis $H_0 : \rho = 1$. That is, simply using $\phi = 1$ is questionable. Moreover, as the stochastic structure of $\{k_t\}$ plays a crucial role in forecasting future mortality rates, the constraint on the random mortality index k_t becomes very restrictive, which indeed means $\mu = 0$ in (4.2) as pointed out in Part 2 of this dissertation.

To overcome these difficulties, Part 2 and Part 3 proposed unit root tests and consistent inferences for the following modified Lee-Carter model without imposing a constraint on the random mortality index:

$$\log m(x,t) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}, \quad k_t = \mu + \phi k_{t-1} + e_t, \quad \sum_{x=1}^{M} \beta_x = 1, \quad \sum_{x=1}^{M} \alpha_x = 0, \qquad (4.3)$$

where the condition $\sum_{x=1}^{M} \alpha_x = 0$ is not restrictive at all because the sum can be absorbed into μ via k_t. Now a natural question is whether an AR(p) model for $\{k_t\}$ is necessary to be adequate and accurate in forecasting future mortality rates. As k_t's are unobservable, such an extension becomes nontrivial at all. Focusing on an AR(2) process, in this Part we develop hypothesis tests and unified inferences regardless of $\{k_t\}$ being stationary or near unit root or unit root. An application to US mortality rates shows that using AR(1) model is suitable.

We organize this Part as follows. Section 4.1 presents the problems, methodologies, and main asymptotic results. A simulation study and data analysis are given in section 4.2. Some conclusions are summarized in section 4.3. All proofs are put into section 4.4.

4.1 Problems, methodologies and asymptotic results

As explained in the introduction, we consider the following modified Lee-Carter model:

$$\begin{cases} \log m(x,t) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}, \quad \sum_{x=1}^M \beta_x = 1, \quad \sum_{x=1}^M \alpha_x = 0, \\ k_t = \mu + \phi_1 k_{t-1} + \phi_2 k_{t-2} + e_t = \mu + \tilde{\phi}_1 k_{t-1} - \phi_2(k_{t-1} - k_{t-2}) + e_t, \end{cases} \tag{4.4}$$

where k_0 and k_{-1} are constants and $\tilde{\phi}_1 = \phi_1 + \phi_2$. Throughout assume that

- C) $\{\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{M,t})^\tau\}$ and $\{e_t\}$ are two independent sequences of independent and identically distributed random vectors with means zero and finite variances, where A^τ denotes the transpose of the matrix or vector A.

Define $Z_t = \sum_{x=1}^M \log m(x,t)$ and $\eta_t = \sum_{x=1}^M \varepsilon_{x,t}$ for $t = 1, \ldots, T$. Then it follows from (4.4) that

$$Z_t = k_t + \eta_t \text{ for } t = 1, \ldots, T. \tag{4.5}$$

When $\{k_t\}$ is a unit root process, Z_t and k_t have a similar size as t large enough. In this case, one may estimate $\mu, \tilde{\phi}_1, \phi_2$ by minimizing the least squares

$$\sum_{t=3}^T \{Z_t - \mu - \tilde{\phi}_1 Z_{t-1} + \phi_2(Z_{t-1} - Z_{t-2})\}^2,$$

which is equivalent to solving the score equations

$$\begin{cases} \sum_{t=3}^T \{Z_t - \mu - \tilde{\phi}_1 Z_{t-1} + \phi_2(Z_{t-1} - Z_{t-2})\} = 0, \\ \sum_{t=3}^T \{Z_t - \mu - \tilde{\phi}_1 Z_{t-1} + \phi_2(Z_{t-1} - Z_{t-2})\} Z_{t-1} = 0, \\ \sum_{t=3}^T \{Z_t - \mu - \tilde{\phi}_1 Z_{t-1} + \phi_2(Z_{t-1} - Z_{t-2})\}(Z_{t-1} - Z_{t-2}) = 0. \end{cases}$$

However, when $\{k_t\}$ is stationary, the above estimators are inconsistent due to the involved η_t in (4.5). Following the bias-correction idea in Part 3, one may apply backshift operator to Z_{t-1} and $Z_{t-1} - Z_{t-2}$ in the above scores. But it turns out that the choice of lags depends on whether $\phi_2 = 0$ or not in order to have a nondegenerate limit. This suggests to study the cases of $\phi_2 = 0$ and $\phi_2 \neq 0$ separately, and to develop a test for $\phi_2 = 0$.

First, consider the case of $\phi_2 = 0$. In this case, we propose the bias-corrected estimators $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta} = (\mu, \tilde{\phi}_1, \phi_2, \alpha_1, \beta_1, \ldots, \alpha_{M-1}, \beta_{M-1})^\tau$ by solving the score equations

$$\begin{cases} \sum_{t=4}^{T}\{Z_t - \mu - \tilde{\phi}_1 Z_{t-1} + \phi_2(Z_{t-1} - Z_{t-2})\} = 0, \\ \sum_{t=4}^{T}\{Z_t - \mu - \tilde{\phi}_1 Z_{t-1} + \phi_2(Z_{t-1} - Z_{t-2})\}Z_{t-2} = 0, \\ \sum_{t=4}^{T}\{Z_t - \mu - \tilde{\phi}_1 Z_{t-1} + \phi_2(Z_{t-1} - Z_{t-2})\}(Z_{t-2} - Z_{t-3}) = 0, \\ \sum_{t=4}^{T}\{\log m(x,t) - \alpha_x - \beta_x Z_t\} = 0, \quad \text{for } x = 1, \ldots, M-1, \\ \sum_{t=4}^{T}\{\log m(x,t) - \alpha_x - \beta_x Z_t\}Z_{t-1} = 0, \quad \text{for } x = 1, \ldots, M-1. \end{cases}$$

Note that α_M and β_M can be estimated by solving

$$\sum_{t=4}^{T}\{\log m(M,t) - \alpha_M - \beta_M Z_t\} = 0 \text{ and } \sum_{t=4}^{T}\{\log m(M,t) - \alpha_M - \beta_M Z_t\}Z_{t-1} = 0,$$

which is equivalent to estimators $-\sum_{x=1}^{M-1}\hat{\alpha}_x$ and $1 - \sum_{x=1}^{M-1}\hat{\beta}_x$. This is why we focus on $\boldsymbol{\theta}$ without α_M and β_M.

Let $\boldsymbol{\theta}_0 = (\mu_0, \tilde{\phi}_{1,0}, \phi_{2,0}, \alpha_{1,0}, \beta_{1,0}, \ldots, \alpha_{M-1,0}, \beta_{M-1,0})^\tau$ be the true value of $\boldsymbol{\theta}$ and denote $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\tilde{\phi}}_1, \hat{\phi}_2, \hat{\alpha}_1, \hat{\beta}_1, \ldots, \hat{\alpha}_{M-1}, \hat{\beta}_{M-1})^\tau$. The following theorem gives the asymptotic distribution of $\hat{\boldsymbol{\theta}}$, which is consistent regardless of the mortality index being unit root or stationary.

Theorem 4.1. *Suppose model (4.4) holds with C), $\phi_{2,0} = 0$, $\mu_0 \neq 0$.*

i) If $\tilde{\phi}_{1,0} = 1$, then

$$D_T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Sigma}(\boldsymbol{\Gamma}_1^{-1})^\tau),$$

where $\boldsymbol{\Sigma}$ and $\boldsymbol{\Gamma}_1$ are respectively defined by (4.11) and (4.12) below, and

$$D_T = diag(\sqrt{T}, T^{3/2}, \sqrt{T}, \sqrt{T}, T^{3/2}, ..., \sqrt{T}, T^{3/2}).$$

ii) If $|\tilde{\phi}_{1,0}| < 1$*, then*

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \widetilde{\boldsymbol{\Gamma}}_1^{-1} \widetilde{\boldsymbol{\Sigma}} (\widetilde{\boldsymbol{\Gamma}}_1^{-1})^\tau),$$

where $\widetilde{\boldsymbol{\Sigma}}$ *and* $\widetilde{\boldsymbol{\Gamma}}_1$ *are respectively defined by (4.14) and (4.15) below.*

To estimate $\boldsymbol{\Sigma}$, $\boldsymbol{\Gamma}_1$, $\widetilde{\boldsymbol{\Sigma}}$ and $\widetilde{\boldsymbol{\Gamma}}_1$, for $x = 1, \ldots, M-1$ and $t = 4, \ldots, T$, define $\tilde{Y}_{t,x}(\alpha_x, \beta_x) = \log m(x,t) - \alpha_x - \beta_x Z_t$, $\bar{Y}_{t,x}(\alpha_x, \beta_x) = \{\log m(x,t) - \alpha_x - \beta_x Z_t\} Z_{t-1}$,

$$\begin{cases} Y_{t,1}(\mu, \boldsymbol{\phi}) = Z_t - \mu - \tilde{\phi}_1 Z_{t-1} + \phi_2 (Z_{t-1} - Z_{t-2}), \\ Y_{t,2}(\mu, \boldsymbol{\phi}) = \{Z_t - \mu - \tilde{\phi}_1 Z_{t-1} + \phi_2 (Z_{t-1} - Z_{t-2})\} Z_{t-2}, \\ Y_{t,3}(\mu, \boldsymbol{\phi}) = \{Z_t - \mu - \tilde{\phi}_1 Z_{t-1} + \phi_2 (Z_{t-1} - Z_{t-2})\} (Z_{t-2} - Z_{t-3}), \end{cases} \tag{4.6}$$

$$\begin{aligned} \boldsymbol{W}_t(\boldsymbol{\theta}) &= \big(Y_{t,1}(\mu, \boldsymbol{\phi}), \tfrac{1}{T} Y_{t,2}(\mu, \boldsymbol{\phi}), Y_{t,3}(\mu, \boldsymbol{\phi}), \tilde{Y}_{t,1}(\alpha_1, \beta_1), \tfrac{1}{T} \bar{Y}_{t,1}(\alpha_1, \beta_1), \\ &\quad \ldots, \tilde{Y}_{t,M-1}(\alpha_{M-1}, \beta_{M-1}), \tfrac{1}{T} \bar{Y}_{t,M-1}(\alpha_{M-1}, \beta_{M-1})\big)^\tau \end{aligned}$$

and

$$\begin{aligned} \widetilde{\boldsymbol{W}}_t(\boldsymbol{\theta}) &= \big(Y_{t,1}(\mu, \boldsymbol{\phi}), Y_{t,2}(\mu, \boldsymbol{\phi}), Y_{t,3}(\mu, \boldsymbol{\phi}), \tilde{Y}_{t,1}(\alpha_1, \beta_1), \bar{Y}_{t,1}(\alpha_1, \beta_1), \\ &\quad \ldots, \tilde{Y}_{t,M-1}(\alpha_{M-1}, \beta_{M-1}), \bar{Y}_{t,M-1}(\alpha_{M-1}, \beta_{M-1})\big)^\tau. \end{aligned}$$

Then we propose to estimate $\boldsymbol{\Sigma}$, $\boldsymbol{\Gamma}_1$, $\widetilde{\boldsymbol{\Sigma}}$ and $\widetilde{\boldsymbol{\Gamma}}_1$ respectively by

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{T-3} \sum_{t=4}^{T} \boldsymbol{W}_t(\hat{\boldsymbol{\theta}}) \boldsymbol{W}_t^\tau(\hat{\boldsymbol{\theta}}) + \frac{1}{T-4} \sum_{t=5}^{T} \boldsymbol{W}_t(\hat{\boldsymbol{\theta}}) \boldsymbol{W}_{t-1}^\tau(\hat{\boldsymbol{\theta}}) + \frac{1}{T-4} \sum_{t=4}^{T-1} \boldsymbol{W}_t(\hat{\boldsymbol{\theta}}) \boldsymbol{W}_{t+1}^\tau(\hat{\boldsymbol{\theta}}),$$

$$\widehat{\boldsymbol{\Gamma}}_1 = \{\frac{1}{T-3} \sum_{t=4}^{T} \frac{\partial}{\partial \boldsymbol{\theta}^\tau} \boldsymbol{W}_t(\hat{\boldsymbol{\theta}})\} \{\sqrt{T} D_T^{-1}\},$$

$$\widehat{\widetilde{\boldsymbol{\Sigma}}} = \frac{1}{T-3} \sum_{t=4}^{T} \widetilde{\boldsymbol{W}}_t(\hat{\boldsymbol{\theta}}) \widetilde{\boldsymbol{W}}_t^\tau(\hat{\boldsymbol{\theta}}) + \frac{1}{T-4} \sum_{t=5}^{T} \widetilde{\boldsymbol{W}}_t(\hat{\boldsymbol{\theta}}) \widetilde{\boldsymbol{W}}_{t-1}^\tau(\hat{\boldsymbol{\theta}}) + \frac{1}{T-4} \sum_{t=4}^{T-1} \widetilde{\boldsymbol{W}}_t(\hat{\boldsymbol{\theta}}) \widetilde{\boldsymbol{W}}_{t+1}^\tau(\hat{\boldsymbol{\theta}}),$$

$$\widehat{\widehat{\boldsymbol{\Gamma}}}_1 = \frac{1}{T-3}\sum_{t=4}^{T}\frac{\partial}{\partial \boldsymbol{\theta}^\tau}\widetilde{\boldsymbol{W}}_t(\hat{\boldsymbol{\theta}}).$$

Theorem 4.2. *Under conditions of Theorem 4.1, as* $T \to \infty$*, we have* $\widehat{\boldsymbol{\Sigma}} \xrightarrow{p} \boldsymbol{\Sigma}$ *and* $\widehat{\boldsymbol{\Gamma}}_1 \xrightarrow{p} \boldsymbol{\Gamma}_1$ *when* $\tilde{\phi}_{1,0} = 1$*, and* $\widehat{\widetilde{\boldsymbol{\Sigma}}} \xrightarrow{p} \widetilde{\boldsymbol{\Sigma}}$ *and* $\widehat{\widetilde{\boldsymbol{\Gamma}}}_1 \xrightarrow{p} \widetilde{\boldsymbol{\Gamma}}_1$ *when* $|\tilde{\phi}_{1,0}| < 1$.

Remark 4.1. *For testing* $H_0 : \phi_2 = 0$*, we first test* $H_0 : \tilde{\phi}_1 = 1\,\&\,\phi_2 = 0$ *based on Theorem 4.1i). If this null hypothesis is rejected, we further test* $H_0 : \phi_2 = 0$ *under the assumption of* $|\tilde{\phi}_1| < 1$ *based on Theorem 4.1ii). More specifically, denote* $\widehat{\boldsymbol{\Gamma}}_1^{-1}\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\Gamma}}_1^{-1})^\tau = (\delta_{i,j})$, $\widehat{\widetilde{\boldsymbol{\Gamma}}}_1^{-1}\widehat{\widetilde{\boldsymbol{\Sigma}}}(\widehat{\widetilde{\boldsymbol{\Gamma}}}_1^{-1})^\tau = (\widetilde{\delta}_{i,j})$*, and test statistics*

$$\Delta_1 = \big(T^{3/2}(\hat{\tilde{\phi}}_1 - 1), T^{1/2}\hat{\phi}_2\big)\begin{pmatrix}\delta_{22} & \delta_{23}\\ \delta_{32} & \delta_{33}\end{pmatrix}^{-1}\begin{pmatrix}T^{3/2}(\hat{\tilde{\phi}}_1 - 1)\\ T^{1/2}\hat{\phi}_2\end{pmatrix}, \quad \Delta_2 = T\hat{\phi}_2^2/\tilde{\delta}_{33}.$$

Therefore, Δ_1 *has a chi-squared limit with two degrees of freedom under* $H_0 : \phi_1 = 1\&\phi_2 = 0$*, and* Δ_2 *has a chi-squared limit with one degree of freedom under* $H_0 : \phi_2 = 0$ *when* $|\phi_1| < 0$*.*

When the above two tests are rejected, we proceed to the case of $\phi_2 \neq 0$. In this case, the above estimators become biased due to the involvement of η_{t-2} in $\{Z_t - \mu - \tilde{\phi}_1 Z_{t-1} + \phi_2(Z_{t-1} - Z_{t-2})\}$, which suggests taking an extra lag by solving the following score equations

$$\begin{cases}\sum_{t=5}^{T}\{Z_t - \mu - \tilde{\phi}_1 Z_{t-1} + \phi_2(Z_{t-1} - Z_{t-2})\} = 0,\\ \sum_{t=5}^{T}\{Z_t - \mu - \tilde{\phi}_1 Z_{t-1} + \phi_2(Z_{t-1} - Z_{t-2})\}Z_{t-3} = 0,\\ \sum_{t=5}^{T}\{Z_t - \mu - \tilde{\phi}_1 Z_{t-1} + \phi_2(Z_{t-1} - Z_{t-2})\}(Z_{t-3} - Z_{t-4}) = 0,\\ \sum_{t=5}^{T}\{\log m(x,t) - \alpha_x - \beta_x Z_t\} = 0, \quad \text{for } x = 1, \cdots, M-1,\\ \sum_{t=5}^{T}\{\log m(x,t) - \alpha_x - \beta_x Z_t\}Z_{t-1} = 0, \quad \text{for } x = 1, \cdots, M-1.\end{cases}$$

Denote the resulted estimators for $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}^*$.

Theorem 4.3. *Suppose model (4.4) holds with C),* $\phi_{2,0} \neq 0$ *and* $\mu_0 \neq 0$*.*

i) If $\tilde{\phi}_{1,0} = 1 \,\&\, |\phi_{2,0}| < 1$, *then*

$$D_T(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}) \xrightarrow{d} N(0, \boldsymbol{\Gamma}_2^{-1}\boldsymbol{\Sigma}^*(\boldsymbol{\Gamma}_2^{-1})^\tau),$$

where $\boldsymbol{\Sigma}^*$ *and* $\boldsymbol{\Gamma}_2$ *are defined by (4.18) and (4.20) below, respectively.*

ii) If $\{k_t\}$ *is stationary, then*

$$\sqrt{T}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}) \xrightarrow{d} N(0, \widetilde{\boldsymbol{\Gamma}}_2^{-1}\widetilde{\boldsymbol{\Sigma}}^*(\widetilde{\boldsymbol{\Gamma}}_2^{-1})^\tau),$$

where $\widetilde{\boldsymbol{\Sigma}}^*$ *and* $\widetilde{\boldsymbol{\Gamma}}_2$ *are defined by (4.23) and (4.26) below, respectively.*

As before, $\boldsymbol{\Sigma}^*$, $\boldsymbol{\Gamma}_2$, $\widetilde{\boldsymbol{\Sigma}}^*$ and $\widetilde{\boldsymbol{\Gamma}}_2$ can be estimated as follows.

For $x = 1, \ldots, M-1$ and $t = 5, \ldots, T$, define $\tilde{Y}^*_{t,x}(\alpha_x, \beta_x) = \log m(x,t) - \alpha_x - \beta_x Z_t$, $\bar{Y}^*_{t,x}(\alpha_x, \beta_x) = \{\log m(x,t) - \alpha_x - \beta_x Z_t\} Z_{t-1}$,

$$\begin{cases} Y^*_{t,1}(\mu, \boldsymbol{\phi}) = Z_t - \mu - \tilde{\phi}_1 Z_{t-1} + \phi_2(Z_{t-1} - Z_{t-2}), \\ Y^*_{t,2}(\mu, \boldsymbol{\phi}) = \{Z_t - \mu - \tilde{\phi}_1 Z_{t-1} + \phi_2(Z_{t-1} - Z_{t-2})\} Z_{t-3}, \\ Y^*_{t,3}(\mu, \boldsymbol{\phi}) = \{Z_t - \mu - \tilde{\phi}_1 Z_{t-1} + \phi_2(Z_{t-1} - Z_{t-2})\}(Z_{t-3} - Z_{t-4}), \end{cases} \tag{4.7}$$

$$\begin{aligned} \boldsymbol{W}^*_t(\boldsymbol{\theta}) &= \big(Y^*_{t,1}(\mu, \boldsymbol{\phi}), \tfrac{1}{T} Y^*_{t,2}(\mu, \boldsymbol{\phi}), Y^*_{t,3}(\mu, \boldsymbol{\phi}), \tilde{Y}^*_{t,1}(\alpha_1, \beta_1), \tfrac{1}{T}\bar{Y}^*_{t,1}(\alpha_1, \beta_1), \\ &\quad \ldots, \tilde{Y}^*_{t,M-1}(\alpha_{M-1}, \beta_{M-1}), \tfrac{1}{T}\bar{Y}^*_{t,M-1}(\alpha_{M-1}, \beta_{M-1})\big)^\tau \end{aligned}$$

and

$$\begin{aligned} \widetilde{\boldsymbol{W}}^*_t(\boldsymbol{\theta}) &= \big(Y^*_{t,1}(\mu, \boldsymbol{\phi}), Y^*_{t,2}(\mu, \boldsymbol{\phi}), Y^*_{t,3}(\mu, \boldsymbol{\phi}), \tilde{Y}^*_{t,1}(\alpha_1, \beta_1), \bar{Y}^*_{t,1}(\alpha_1, \beta_1), \\ &\quad \ldots, \tilde{Y}^*_{t,M-1}(\alpha_{M-1}, \beta_{M-1}), \bar{Y}^*_{t,M-1}(\alpha_{M-1}, \beta_{M-1})\big)^\tau. \end{aligned}$$

Then we propose to estimate $\boldsymbol{\Sigma}^*$, $\boldsymbol{\Gamma}_2$, $\widetilde{\boldsymbol{\Sigma}^*}$ and $\widetilde{\boldsymbol{\Gamma}}_2$ respectively by

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}^*} &= \tfrac{1}{T-4}\textstyle\sum_{t=5}^T \boldsymbol{W}^*_t(\hat{\boldsymbol{\theta}}^*)\boldsymbol{W}^{*\tau}_t(\hat{\boldsymbol{\theta}}^*) + \tfrac{1}{T-5}\sum_{t=6}^T \boldsymbol{W}^*_t(\hat{\boldsymbol{\theta}}^*)\boldsymbol{W}^{*\tau}_{t-1}(\hat{\boldsymbol{\theta}}^*) + \tfrac{1}{T-6}\sum_{t=7}^T \boldsymbol{W}^*_t(\hat{\boldsymbol{\theta}}^*)\boldsymbol{W}^{*\tau}_{t-2}(\hat{\boldsymbol{\theta}}^*) \\ &\quad + \tfrac{1}{T-5}\textstyle\sum_{t=5}^{T-1} \boldsymbol{W}^*_t(\hat{\boldsymbol{\theta}}^*)\boldsymbol{W}^{*\tau}_{t+1}(\hat{\boldsymbol{\theta}}^*) + \tfrac{1}{T-6}\sum_{t=5}^{T-2} \boldsymbol{W}^*_t(\hat{\boldsymbol{\theta}}^*)\boldsymbol{W}^{*\tau}_{t+2}(\hat{\boldsymbol{\theta}}^*), \end{aligned}$$

$$\widehat{\mathbf{\Gamma}}_2 = \{\frac{1}{T-4}\sum_{t=5}^{T}\frac{\partial}{\partial\boldsymbol{\theta}^\tau}\boldsymbol{W}_t^*(\hat{\boldsymbol{\theta}}^*)\}\{\sqrt{T}D_T^{-1}\},$$

$$\begin{aligned}\widehat{\widetilde{\mathbf{\Sigma}}}^* &= \tfrac{1}{T-4}\textstyle\sum_{t=5}^{T}\widetilde{\boldsymbol{W}}_t^*(\hat{\boldsymbol{\theta}}^*)\widetilde{\boldsymbol{W}}_t^{*\tau}(\hat{\boldsymbol{\theta}}^*) + \tfrac{1}{T-5}\textstyle\sum_{t=6}^{T}\widetilde{\boldsymbol{W}}_t^*(\hat{\boldsymbol{\theta}}^*)\widetilde{\boldsymbol{W}}_{t-1}^{*\tau}(\hat{\boldsymbol{\theta}}^*) + \tfrac{1}{T-6}\textstyle\sum_{t=7}^{T}\widetilde{\boldsymbol{W}}_t^*(\hat{\boldsymbol{\theta}}^*)\widetilde{\boldsymbol{W}}_{t-2}^{*\tau}(\hat{\boldsymbol{\theta}}^*)\\ &\quad + \tfrac{1}{T-5}\textstyle\sum_{t=5}^{T-1}\widetilde{\boldsymbol{W}}_t^*(\hat{\boldsymbol{\theta}}^*)\widetilde{\boldsymbol{W}}_{t+1}^{*\tau}(\hat{\boldsymbol{\theta}}^*) + \tfrac{1}{T-6}\textstyle\sum_{t=5}^{T-2}\widetilde{\boldsymbol{W}}_t^*(\hat{\boldsymbol{\theta}}^*)\widetilde{\boldsymbol{W}}_{t+2}^{*\tau}(\hat{\boldsymbol{\theta}}^*),\end{aligned}$$

$$\widehat{\widetilde{\mathbf{\Gamma}}}_2 = \frac{1}{T-4}\sum_{t=5}^{T}\frac{\partial}{\partial\boldsymbol{\theta}^\tau}\widetilde{\boldsymbol{W}}_t^*(\hat{\boldsymbol{\theta}}^*).$$

Theorem 4.4. *Under conditions of Theorem 4.3, as $T\to\infty$, we have $\widehat{\mathbf{\Sigma}}^* \xrightarrow{p} \mathbf{\Sigma}^*$ and $\widehat{\mathbf{\Gamma}}_2 \xrightarrow{p} \mathbf{\Gamma}_2$ when $\tilde{\phi}_{1,0}=1$ & $|\phi_{2,0}|<1$, and $\widehat{\widetilde{\mathbf{\Sigma}}}^* \xrightarrow{p} \widetilde{\mathbf{\Sigma}}^*$ and $\widehat{\widetilde{\mathbf{\Gamma}}}_2 \xrightarrow{p} \widetilde{\mathbf{\Gamma}}_2$ when $\{k_t\}$ is stationary.*

Remark 4.2. *It is not surprising that the terms of $\sum_{t=5}^{T}(Z_{t-1}-Z_{t-2})(Z_{t-3}-Z_{t-4})$ in defining $\hat{\boldsymbol{\theta}}^*$ and $\sum_{t=5}^{T}(Z_{t-1}-Z_{t-2})(Z_{t-2}-Z_{t-3})$ in defining $\hat{\boldsymbol{\theta}}$ may become much smaller than $\sum_{t=5}^{T}(Z_{t-1}-Z_{t-2})^2$, which challenges the estimation of ϕ_2 when ϕ_2 is neither close to nor far away from zero. This is confirmed by the conducted simulation study below.*

4.2 Data analysis and simulation study

First we analyze the publicly available U.S. mortality data from Human Mortality Database (HMD) [1] by focusing on the U.S. population between 25 and 74 years old and using mortality data by 5-year age groups observed between the year 1933 and 2015. This results in $M = 10$ age groups and $T = 83$ years of observations. Estimates $\hat{\boldsymbol{\theta}}$ based on $\phi_2 = 0$ in Theorem 4.1 and $\hat{\boldsymbol{\theta}}^*$ based on $\phi_2 \neq 0$ in Theorem 4.3 for female, male and combined mortality rates are reported in Tables 4.1 and 4.2, respectively. These two tables show that the assumption of $\phi_2 = 0$ or $\phi_2 \neq 0$ has a noticeable difference on estimates for μ, ϕ_2, α_x, but not much difference for $\tilde{\phi}_1, \beta_x$. The computed test statistics for testing $H_0: \phi_1 = 1 \& \phi_2 = 0$ and $H_0: \phi_2 = 0$ with $|\phi_1| < 1$ are $\Delta_1 = 12.8001$ and $\Delta_2 = 0.2313$ for the

[1] http://www.mortality.org/cgi-bin/hmd-country.php?cntr=USA&level=1

combined mortality rates with the corresponding P-values 0.0017 and 0.6306, $\Delta_1 = 2.9704$ and $\Delta_2 = 0.1492$ for the male mortality rates with the corresponding P-values 0.2265 and 0.6993, and $\Delta_1 = 60.8755$ and $\Delta_2 = 0.5176$ for the female mortality rates with the corresponding P-values 6.0396×10^{-14} and 0.4719. These P-values suggest a stationary AR(1) model is sound for the female and combined mortality rates and a unit root AR(1) model is suitable for the male mortality rate, which is in line with the findings in Part 2 and supports the adequacy of using an AR(1) model instead of an AR(p) model. Below we conduct a simulation study to examine the challenges shown in developed theorems and Remark 4.1 that estimators for ϕ_2 and μ have a much slower rate of convergence in case of unit root mortality index.

Table (4.1) Estimates $\hat{\boldsymbol{\theta}}$ based on the assumption of $\phi_2 = 0$ in Theorem 4.1.

x	1	2	3	4	5	6	7	8	9	10
Combined α_x	-1.636	-1.255	-0.660	-0.290	-0.071	0.353	0.508	0.688	1.005	1.357
Combined β_x	0.098	0.102	0.108	0.108	0.104	0.104	0.099	0.094	0.093	0.091
Male α_x	-2.392	-1.955	-1.040	-0.334	0.109	0.670	0.940	1.152	1.382	1.468
Male β_x	0.081	0.087	0.101	0.108	0.108	0.110	0.107	0.103	0.100	0.094
Female α_x	-0.142	-0.137	-0.128	-0.305	-0.443	-0.247	-0.218	0.014	0.501	1.105
Female β_x	0.129	0.124	0.117	0.106	0.096	0.092	0.085	0.082	0.083	0.086

	μ	$\tilde{\phi}_1$	ϕ_1	ϕ_2
Combined	-0.848	0.985	1.302	-0.317
Male	-0.600	0.989	1.208	-0.218
Female	-0.693	0.988	1.616	-0.628

We draw 10,000 random samples with sample size $T = 300$ from model (4.4) with α_x, β_x, μ being the corresponding estimates for the male mortality rates given in Table 4.1, and independent normal random variables for $\varepsilon_{x,t}$ and e_t with the same standard deviation 0.01. We take $(\phi_1, \phi_2) = (\hat{\phi}_1 - 0.2, \hat{\phi}_2 + 0.2), (\hat{\phi}_1, \hat{\phi}_2), (\hat{\phi}_1 + 0.4, \hat{\phi}_2 - 0.4)$, where $\hat{\phi}_1 = 1.208$ and $\hat{\phi}_2 = -0.218$ are the corresponding estimates for the male mortality rates in Table 4.1. Therefore we investigate how the estimators in Theorems 4.1 and 4.3 are affected when ϕ_2 is close to zero, nonzero and far away from zero. The means and standard errors of estimates based on $\phi_2 = 0$ in Theorem 4.1 and $\phi_2 \neq 0$ in Theorem 4.3 are reported in Tables 4.3–4.5.

Table (4.2) Estimates $\hat{\boldsymbol{\theta}}^*$ based on the assumption of $\phi_2 \neq 0$ in Theorem 4.3.

x	1	2	3	4	5	6	7	8	9	10
Combined α_x	-1.789	-1.370	-0.751	-0.310	-0.054	0.407	0.594	0.772	1.072	1.430
Combined β_x	0.095	0.100	0.106	0.107	0.104	0.105	0.100	0.096	0.094	0.093
Male α_x	-2.523	-2.081	-1.149	-0.379	0.114	0.720	1.031	1.251	1.469	1.548
Male β_x	0.079	0.085	0.098	0.107	0.108	0.111	0.109	0.105	0.102	0.095
Female α_x	-0.274	-0.201	-0.176	-0.294	-0.423	-0.214	-0.170	0.054	0.532	1.165
Female β_x	0.127	0.123	0.116	0.107	0.096	0.093	0.086	0.082	0.084	0.087

	μ	$\tilde{\phi}_1$	ϕ_1	ϕ_2
Combined	-0.900	0.984	1.491	-0.507
Male	-0.575	0.989	1.679	-0.690
Female	-0.929	0.984	1.598	-0.615

These tables show that estimators for μ and ϕ_2 based on Theorem 4.1 are good when ϕ_2 is close to zero, but become worse when ϕ_2 is away from zero. The estimators for μ and ϕ_2 based on Theorem 4.3 perform well when ϕ_2 is far away from zero. This simulation study concludes that estimating ϕ_2 is challenging and the reason may be explained by Remark 4.1, i.e., using lags to correct the bias makes the inference inefficient. In conclusion, one may prefer an AR(1) model to an AR(p) model for the unobserved mortality index.

4.3 Conclusions

Researchers in actuarial science often fit a unit root AR(1) model for the unobserved mortality index. Recent developments show that the unit root hypothesis is rejected for some mortality datasets, the widely employed two-step inference in Lee & Carter (1992) leads to an inconsistent inference for a stationary AR(1) process, and a consistent inference is proposed for a modified Lee-Carter model with an AR(1) process for the unobserved mortality index. This paper investigates the possible advantages of using an AR(2) model instead of an AR(1) model by developing hypothesis tests and unified inferences regardless of the mortality index being stationary or unit root. Data analysis for the US mortality rates show that an AR(1) model is suitable, which is in line with Part 2.

4.4 Proofs

For ease of notation, we use $Y_{t,j}, \tilde{Y}_{t,x}, \bar{Y}_{t,x}, Y^*_{t,j}, \tilde{Y}^*_{t,x}, \bar{Y}^*_{t,x}$ to respectively denote $Y_{t,j}(\mu_0, \boldsymbol{\phi}_0)$, $\tilde{Y}_{t,x}(\alpha_{x,0}, \beta_{x,0})$, $\bar{Y}_{t,x}(\alpha_{x,0}, \beta_{x,0})$, $Y^*_{t,j}(\mu_0, \boldsymbol{\phi}_0), \tilde{Y}^*_{t,x}(\alpha_{x,0}, \beta_{x,0})$ and $\bar{Y}^*_{t,x}(\alpha_{x,0}, \beta_{x,0})$, which are defined in Section 4.1.

Lemma 4.1. *Suppose model (4.4) holds with C), $\phi_{2,0} = 0$, $\mu_0 \neq 0$.*

i) If $\tilde{\phi}_{1,0} = 1$, then

$$\frac{1}{\sqrt{T}} \sum_{t=4}^{T} \boldsymbol{W}_t(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}) \quad \text{and} \quad \{\frac{1}{T} \sum_{t=4}^{T} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{W}_t(\boldsymbol{\theta}_0)\}\{\sqrt{T} D_T^{-1}\} \xrightarrow{p} \boldsymbol{\Gamma}_1, \quad \text{as } T \to \infty,$$

where $\boldsymbol{W}_t$ is defined in Section 2, and $\boldsymbol{\Sigma}$ and $\boldsymbol{\Gamma}_1$ are given in (4.11) and (4.12) below respectively.

ii) If $|\tilde{\phi}_{1,0}| < 1$, then

$$\frac{1}{\sqrt{T}} \sum_{t=4}^{T} \widetilde{\boldsymbol{W}}_t(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}) \quad \text{and} \quad \frac{1}{T} \sum_{t=4}^{T} \frac{\partial}{\partial \boldsymbol{\theta}} \widetilde{\boldsymbol{W}}_t(\boldsymbol{\theta}_0) \xrightarrow{p} \widetilde{\boldsymbol{\Gamma}}_1, \quad \text{as } T \to \infty,$$

where $\widetilde{\boldsymbol{W}}_t$ is defined in Section 2, and $\widetilde{\boldsymbol{\Sigma}}$ and $\widetilde{\boldsymbol{\Gamma}}_1$ are given in (4.14) and (4.15) below respectively.

Proof. i) When $\tilde{\phi}_{1,0} = 1$, we have $k_t = \mu_0 + k_{t-1} + e_t$, which implies that, as $T \to \infty$

$$k_{[Ts]}/T \xrightarrow{p} \mu_0 s, \quad \text{for } s \in [0, 1]. \tag{4.8}$$

It is easy to check that

$$\begin{cases} Y_{t,1} = e_t + \eta_t - \eta_{t-1}, \\ Y_{t,2} = (e_t + \eta_t - \eta_{t-1})(k_{t-2} + \eta_{t-2}), \\ Y_{t,3} = (e_t + \eta_t - \eta_{t-1})(k_{t-2} - k_{t-3} + \eta_{t-2} - \eta_{t-3}), \\ \tilde{Y}_{t,x} = \varepsilon_{x,t} - \beta_{x,0}\eta_t, \ x = 1, 2, \ldots, M-1, \\ \bar{Y}_{t,x} = (\varepsilon_{x,t} - \beta_{x,0}\eta_t)(k_{t-1} + \eta_{t-1}), \ x = 1, 2, \ldots, M-1. \end{cases} \tag{4.9}$$

Then we have

$$\frac{1}{\sqrt{T}}\sum_{t=4}^{T}\boldsymbol{W}_t(\boldsymbol{\theta}_0)=\frac{1}{\sqrt{T}}\sum_{t=4}^{T}\boldsymbol{U}_t(\boldsymbol{\theta}_0)+o_p(1),$$

where $\boldsymbol{U}_t(\boldsymbol{\theta}_0)=(V_{t,1},V_{t,2},V_{t,3},\tilde{V}_{t,1},\bar{V}_{t,1},\ldots,\tilde{V}_{t,M-1},\bar{V}_{t,M-1})^\tau$ and for $x=1,\ldots,M-1$

$$\begin{cases} V_{t,1}=e_t,\ V_{t,2}=(e_t+\eta_t)\frac{k_{t-2}}{T}-\eta_t\frac{k_{t-1}}{T}, \\ V_{t,3}=(e_t+\eta_t)(\mu_0+e_{t-2}+\eta_{t-2}-\eta_{t-3})-\eta_t(\mu_0+e_{t-1}+\eta_{t-1}-\eta_{t-2}), \\ \tilde{V}_{t,x}=\varepsilon_{x,t}-\beta_{x,0}\eta_t,\ \bar{V}_{t,x}=(\varepsilon_{x,t}-\beta_{x,0}\eta_t)\frac{k_{t-1}}{T}. \end{cases} \tag{4.10}$$

Let $\mathcal{F}_t$ denote the σ-field generated by $\{(e_s,\varepsilon_{1,s},\ldots,\varepsilon_{M,s})^\tau : s\le t\}$. Then $\{\boldsymbol{U}_t(\boldsymbol{\theta}_0),\mathcal{F}_t\}_{t=1}^{\infty}$ is a martingale difference sequence. By the central limit theorem for a martingale difference sequence in Hall & Heyde (2014), $\frac{1}{\sqrt{T}}\sum_{t=4}^{T}\boldsymbol{U}_t(\boldsymbol{\theta}_0)$ converges in distribution to a multivariate normal distribution with zero means and covariance matrix $\boldsymbol{\Sigma}=(\sigma_{ij})_{i,j=1,2,\ldots,3+2(M-1)}$ satisfying

$$\frac{1}{T}\sum_{t=4}^{T}E\{\boldsymbol{U}_t(\boldsymbol{\theta}_0)\boldsymbol{U}_t^\tau(\boldsymbol{\theta}_0)|\mathcal{F}_{t-1}\}\xrightarrow{p}\boldsymbol{\Sigma},\quad \text{as}\quad T\to\infty, \tag{4.11}$$

i.e.,

$$\frac{1}{\sqrt{T}}\sum_{t=4}^{T}\boldsymbol{W}_t(\boldsymbol{\theta}_0)=\frac{1}{\sqrt{T}}\sum_{t=4}^{T}\boldsymbol{U}_t(\boldsymbol{\theta}_0)+o_p(1)\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{\Sigma}),\quad \text{as } T\to\infty.$$

By the assumption that $e_1,...,e_T,\eta_1,...,\eta_T$ are independent with mean zero, it is straightforward to verify that, for $x,y=1,2,\ldots,M-1$,

$$\sigma_{1,1}=E(e_1^2),\quad \sigma_{1,2}=E(e_1^2)\int_0^1\mu_0 s\,ds=\frac{\mu_0}{2}E(e_1^2),\quad \sigma_{1,3}=\mu_0E(e_1^2),\quad \sigma_{1,2x+2}=\sigma_{1,2y+3}=0,$$

$$\sigma_{2,2}=E(e_1^2)\int_0^1\mu_0^2s^2ds=\frac{\mu_0^2}{3}E(e_1^2),\quad \sigma_{2,3}=\mu_0E(e_1^2)\int_0^1\mu_0 s\,ds=\frac{\mu_0^2}{2}E(e_1^2),$$

$$\sigma_{2,2x+2}=\sigma_{2,2y+3}=0,\quad \sigma_{3,3}=E(e_1^2)\{\mu_0^2+E(e_1^2)+2E(\eta_1^2)\}+2E(\eta_1^2)\{E(e_1^2)+3E(\eta_1^2)\}.$$

Similarly we can show that for $x,y=1,2,\ldots,M-1$

$$\sigma_{3,2x+2}=\sigma_{3,2y+3}=0,\quad \sigma_{2x+2,2y+2}=E\{(\varepsilon_{x,1}-\beta_{x,0}\eta_1)(\varepsilon_{y,1}-\beta_{y,0}\eta_1)\},$$

$$\sigma_{2x+2,2y+3} = \frac{\mu_0}{2} E\{(\varepsilon_{x,1} - \beta_{x,0}\eta_1)(\varepsilon_{y,1} - \beta_{y,0}\eta_1)\},$$

$$\sigma_{2x+3,2y+3} = \frac{\mu_0^2}{3} E\{(\varepsilon_{x,1} - \beta_{x,0}\eta_1)(\varepsilon_{y,1} - \beta_{y,0}\eta_1)\}.$$

Further note that

$$\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{W}_t(\boldsymbol{\theta}_0) = \begin{pmatrix} A_{t0} & 0 & \dots & 0 \\ 0 & A_{t1} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & A_{t(M-1)} \end{pmatrix},$$

where

$$A_{t0} = -\begin{pmatrix} 1 & Z_{t-1} & -(Z_{t-1} - Z_{t-2}) \\ \frac{1}{T} Z_{t-2} & \frac{1}{T} Z_{t-1} Z_{t-2} & -\frac{1}{T}(Z_{t-1} - Z_{t-2}) Z_{t-2} \\ Z_{t-2} - Z_{t-3} & Z_{t-1}(Z_{t-2} - Z_{t-3}) & -(Z_{t-1} - Z_{t-2})(Z_{t-2} - Z_{t-3}) \end{pmatrix}$$

and

$$A_{tx} = -\begin{pmatrix} 1 & Z_t \\ \frac{1}{T} Z_{t-1} & \frac{1}{T} Z_t Z_{t-1} \end{pmatrix}, \ x = 1, \dots, M-1.$$

It is easy to show that as $T \to \infty$,

$$\frac{1}{T} \sum_{t=4}^{T} \frac{Z_t}{T} = \frac{1}{T} \sum_{t=4}^{T} \frac{k_t}{T} + o_p(1) \xrightarrow{p} \frac{\mu_0}{2},$$

$$\frac{1}{T} \sum_{t=4}^{T} (Z_{t-1} - Z_{t-2}) = \frac{1}{T} \sum_{t=4}^{T} (\mu_0 + e_{t-1} + \eta_{t-1} - \eta_{t-2}) \xrightarrow{p} \mu_0,$$

$$\frac{1}{T} \sum_{t=4}^{T} \frac{Z_t}{T} \frac{Z_{t-1}}{T} = \frac{1}{T} \sum_{t=4}^{T} (Z_{t-1}/T)^2 + o_p(1) \xrightarrow{p} \frac{\mu_0^2}{3},$$

$$\frac{1}{T} \sum_{t=4}^{T} (Z_{t-1} - Z_{t-2}) \frac{Z_{t-2}}{T} = \frac{1}{T} \sum_{t=4}^{T} (\mu_0 + e_{t-1} + \eta_{t-1} - \eta_{t-2}) \frac{k_{t-2}}{T} + o_p(1) \xrightarrow{p} \frac{\mu_0^2}{2}$$

and

$$\begin{aligned}
&\tfrac{1}{T}\textstyle\sum_{t=4}^{T}(Z_{t-1}-Z_{t-2})(Z_{t-2}-Z_{t-3})\\
&=\tfrac{1}{T}\textstyle\sum_{t=4}^{T}(\mu_0+e_{t-1}+\eta_{t-1}-\eta_{t-2})(\mu_0+e_{t-2}+\eta_{t-2}-\eta_{t-3})\\
&=\mu_0^2-\tfrac{1}{T}\textstyle\sum_{t=4}^{T}\eta_{t-2}^2+o_p(1)\\
&\xrightarrow{p}\mu_0^2-E(\eta_1^2),
\end{aligned}$$

implying that

$$\frac{1}{T}\{\sum_{t=4}^{T}\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{W}_t(\boldsymbol{\theta}_0)\}\{\sqrt{T}D_T^{-1}\}\xrightarrow{p}\boldsymbol{\Gamma}_1=:\begin{pmatrix} A_0 & 0 & \dots & 0\\ 0 & A_1 & \dots & 0\\ \dots & \dots & \dots & \dots\\ 0 & 0 & \dots & A_{M-1}\end{pmatrix} \tag{4.12}$$

with

$$A_0=-\begin{pmatrix} 1 & \frac{\mu_0}{2} & -\mu_0\\ \frac{\mu_0}{2} & \frac{\mu_0^2}{3} & -\frac{\mu_0^2}{2}\\ \mu_0 & \frac{\mu_0^2}{2} & -\mu_0^2+E(\eta_1^2)\end{pmatrix}\text{ and }A_x=-\begin{pmatrix} 1 & \frac{\mu_0}{2}\\ \frac{\mu_0}{2} & \frac{\mu_0^2}{3}\end{pmatrix}$$

for $x=1,\dots,M-1$. Hence Lemma 4.1i) holds.

ii) When $|\tilde{\phi}_{1,0}|<1$, it is easy to check that

$$\begin{cases} Y_{t,1}=e_t+\eta_t-\phi_{1,0}\eta_{t-1},\\ Y_{t,2}=(e_t+\eta_t-\phi_{1,0}\eta_{t-1})(k_{t-2}+\eta_{t-2}),\\ Y_{t,3}=(e_t+\eta_t-\phi_{1,0}\eta_{t-1})(k_{t-2}-k_{t-3}+\eta_{t-2}-\eta_{t-3}),\\ \tilde{Y}_{t,x}=\varepsilon_{x,t}-\beta_{x,0}\eta_t,\ x=1,2,\dots,M-1,\\ \bar{Y}_{t,x}=(\varepsilon_{x,t}-\beta_{x,0}\eta_t)(k_{t-1}+\eta_{t-1}),\ x=1,2,\dots,M-1.\end{cases} \tag{4.13}$$

Define

$$X_{t,1}=e_t+\eta_t-\phi_{1,0}\eta_t,\ X_{t,2}=(e_t+\eta_t)(k_{t-2}+\eta_{t-2})-\phi_{1,0}\eta_t(k_{t-1}+\eta_{t-1}),$$

$$X_{t,3}=(e_t+\eta_t)(k_{t-2}-k_{t-3}+\eta_{t-2}-\eta_{t-3})-\phi_{1,0}\eta_t(k_{t-1}-k_{t-2}+\eta_{t-1}-\eta_{t-2}),$$

and denote $\widetilde{\boldsymbol{U}}_t(\boldsymbol{\theta}_0)$ as $\widetilde{\boldsymbol{W}}_t(\boldsymbol{\theta}_0)$ with $Y_{t,1}$, $Y_{t,2}$, $Y_{t,3}$ replaced by $X_{t,1}$, $X_{t,2}$, $X_{t,3}$, respectively. Then $\{\widetilde{\boldsymbol{U}}_t(\boldsymbol{\theta}_0), \mathcal{F}_t\}_{t=1}^{\infty}$ is a martingale difference sequence. By the central limit theorem for a martingale difference sequence in Hall & Heyde (2014), $\frac{1}{\sqrt{T}}\sum_{t=4}^{T}\widetilde{\boldsymbol{U}}_t(\boldsymbol{\theta}_0)$ converges in distribution to a multivariate normal distribution with zero means and covariance matrix $\widetilde{\boldsymbol{\Sigma}} = (\tilde{\sigma}_{i,j})_{i,j=1,2,\ldots,3+2(M-1)}$ satisfying

$$\frac{1}{T}\sum_{t=4}^{T} E\{\widetilde{\boldsymbol{U}}_t(\boldsymbol{\theta}_0)\widetilde{\boldsymbol{U}}_t(\boldsymbol{\theta}_0)^{\tau}|\mathcal{F}_{t-1}\} \xrightarrow{p} \widetilde{\boldsymbol{\Sigma}}, \quad T \to \infty, \tag{4.14}$$

i.e.,

$$\frac{1}{\sqrt{T}}\sum_{t=4}^{T}\widetilde{\boldsymbol{W}}_t(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{T}}\sum_{t=4}^{T}\widetilde{\boldsymbol{U}}_t(\boldsymbol{\theta}_0) + o_p(1) \xrightarrow{d} N(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}}) \ \text{ as } T \to \infty.$$

To compute the covariance matrix, by noting that

$$E(k_t) = \frac{\mu_0}{1-\phi_{1,0}}, \quad E(k_t^2) = \frac{\mu_0^2}{(1-\phi_{1,0})^2} + \frac{E(e_1^2)}{1-\phi_{1,0}^2},$$

$$E(k_t k_{t-1}) = \frac{\mu_0^2}{(1-\phi_{1,0})^2} + \frac{\phi_{1,0}E(e_1^2)}{1-\phi_{1,0}^2} \text{ and } E(k_t k_{t-2}) = \frac{\mu_0^2}{(1-\phi_{1,0})^2} + \frac{\phi_{1,0}^2 E(e_1^2)}{1-\phi_{1,0}^2},$$

it is straightforward to verify that for $1 \le x, y \le M-1$

$$\tilde{\sigma}_{1,1} = E(e_1 + (1-\phi_{1,0})\eta_1)^2 = E(e_1^2) + (1-\phi_{1,0})^2 E(\eta_1^2),$$

$$\tilde{\sigma}_{1,2} = \{E(e_1^2) + (1-\phi_{1,0})^2 E(\eta_1^2)\}\frac{\mu_0}{1-\phi_{1,0}} = \frac{\mu_0}{1-\phi_{1,0}}\tilde{\sigma}_{1,1},$$

$$\tilde{\sigma}_{1,3} = 0, \quad \tilde{\sigma}_{1,2x+2} = (1-\phi_{1,0})E\{\eta_1(\varepsilon_{x,1} - \beta_{x,0}\eta_1)\},$$

$$\tilde{\sigma}_{1,2x+3} = (1-\phi_{1,0})E\{\eta_1(\varepsilon_{x,1} - \beta_{x,0}\eta_1)\}\frac{\mu_0}{1-\phi_{1,0}} = \frac{\mu_0}{1-\phi_{1,0}}\tilde{\sigma}_{1,2x+2},$$

$$
\begin{aligned}
\tilde{\sigma}_{2,2} &= E(e_t^2)E(k_{t-2}+\eta_{t-2})^2 + E(\eta_t^2)E(k_{t-2}+\eta_{t-2}-\phi_{1,0}k_{t-1}-\phi_{1,0}\eta_{t-1})^2 \\
&= \{E(e_1^2)+(1-\phi_{1,0})^2E(\eta_1^2)\}\{\tfrac{\mu_0^2}{(1-\phi_{1,0})^2}+\tfrac{E(e_1^2)}{1-\phi_{1,0}^2}\} \\
&\quad +E(\eta_1^2)\{(1+\tfrac{2\phi_{1,0}}{1+\phi_{1,0}})E(e_1^2)+(1+\phi_{1,0}^2)E(\eta_1^2)\},
\end{aligned}
$$

$$
\begin{aligned}
\tilde{\sigma}_{2,3} &= E(e_t+\eta_t)^2E(k_{t-2}^2-k_{t-2}k_{t-3}+\eta_{t-2}^2) \\
&\quad -\phi_{1,0}E(\eta_t^2)E(k_{t-1}k_{t-2}-k_{t-1}k_{t-3}) \\
&\quad -\phi_{1,0}E(\eta_t^2)E(k_{t-2}k_{t-1}-k_{t-2}^2-\eta_{t-2}^2) \\
&\quad +\phi_{1,0}^2E(\eta_t^2)E(k_{t-1}^2-k_{t-1}k_{t-2}+\eta_{t-1}^2) \\
&= \{E(e_1^2)+(1+\phi_{1,0})E(\eta_1^2)\}\tfrac{E(e_1^2)}{1+\phi_{1,0}}+\{E(e_1^2)+(1+\phi_{1,0}+\phi_{1,0}^2)E(\eta_1^2)\}E(\eta_1^2),
\end{aligned}
$$

$$
\tilde{\sigma}_{2,2x+2} = \frac{\mu_0}{1-\phi_{1,0}}\tilde{\sigma}_{1,2x+2} = \tilde{\sigma}_{1,2x+3},
$$

$$
\tilde{\sigma}_{2,2x+3} = E\{\eta_1(\varepsilon_{x,1}-\beta_{x,0}\eta_1)\}\{\frac{\mu_0^2}{1-\phi_{1,0}}-\phi_{1,0}E(\eta_1^2)\},
$$

$$
\begin{aligned}
\tilde{\sigma}_{3,3} &= E(e_t^2)E(k_{t-2}-k_{t-3}+\eta_{t-2}-\eta_{t-3})^2 \\
&\quad +E(\eta_t^2)E(k_{t-2}-k_{t-3}+\eta_{t-2}-\eta_{t-3})^2 \\
&\quad +\phi_{1,0}^2E(\eta_t^2)E(k_{t-1}-k_{t-2}+\eta_{t-1}-\eta_{t-2})^2 \\
&\quad -2\phi_{1,0}E(\eta_t^2)E\{(k_{t-2}-k_{t-3}+\eta_{t-2}-\eta_{t-3})(k_{t-1}-k_{t-2}+\eta_{t-1}-\eta_{t-2})\} \\
&= E(e_1^2)\{\tfrac{2E(e_1^2)}{1+\phi_{1,0}}+2E(\eta_1^2)\}+(1+\phi_{1,0}^2)E(\eta_1^2)\{\tfrac{2E(e_1^2)}{1+\phi_{1,0}}+2E(\eta_1^2)\} \\
&\quad -2\phi_{1,0}E(\eta_1^2)\{-\tfrac{1-\phi_{1,0}}{1+\phi_{1,0}}E(e_1^2)-E(\eta_1^2)\} \\
&= \tfrac{2}{1+\phi_{1,0}}\{E(e_1^2)\}^2+4E(e_1^2)E(\eta_1^2)+2(1+\phi_{1,0}+\phi_{1,0}^2)\{E(\eta_1^2)\}^2 \\
&= 2\tilde{\sigma}_{2,3},
\end{aligned}
$$

$$
\begin{aligned}
\tilde{\sigma}_{3,2x+3} &= E\{\eta_t(\varepsilon_{x,t} - \beta_{x,0}\eta_t)\}E(k_{t-1}k_{t-2} - k_{t-1}k_{t-3}) \\
&\quad -\phi_{1,0}E\{\eta_t(\varepsilon_{x,t} - \beta_{x,0}\eta_t)\}E(k_{t-1}^2 - k_{t-1}k_{t-2} + \eta_{t-1}^2) \\
&= -\phi_{1,0}E(\eta_1^2)E\{\eta_1(\varepsilon_{x,1} - \beta_{x,0}\eta_1)\},
\end{aligned}
$$

$$
\tilde{\sigma}_{3,2x+2} = 0, \quad \tilde{\sigma}_{2x+2,2y+2} = E\{(\varepsilon_{x,1} - \beta_{x,0}\eta_1)(\varepsilon_{y,1} - \beta_{y,0}\eta_1)\},
$$

$$
\tilde{\sigma}_{2x+2,2y+3} = E\{(\varepsilon_{x,1} - \beta_{x,0}\eta_1)(\varepsilon_{y,1} - \beta_{y,0}\eta_1)\}E(k_2 + \eta_2) = \frac{\mu_0}{1 - \phi_{1,0}}\tilde{\sigma}_{2x+2,2y+2},
$$

$$
\tilde{\sigma}_{2x+3,2y+3} = E\{(\varepsilon_{x,1} - \beta_{x,0}\eta_1)(\varepsilon_{y,1} - \beta_{y,0}\eta_1)\}\{\frac{\mu_0^2}{(1 - \phi_{1,0})^2} + \frac{E(e_1^2)}{1 - \phi_{1,0}^2} + E(\eta_1^2)\}.
$$

Note that

$$
\frac{\partial}{\partial \boldsymbol{\theta}}\widetilde{\boldsymbol{W}}_t(\boldsymbol{\theta}_0) = \begin{pmatrix} \tilde{A}_{t0} & 0 & \dots & 0 \\ 0 & \tilde{A}_{t1} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \tilde{A}_{t(M-1)} \end{pmatrix},
$$

where

$$
\tilde{A}_{t0} = -\begin{pmatrix} 1 & Z_{t-1} & -(Z_{t-1} - Z_{t-2}) \\ Z_{t-2} & Z_{t-1}Z_{t-2} & -(Z_{t-1} - Z_{t-2})Z_{t-2} \\ Z_{t-2} - Z_{t-3} & Z_{t-1}(Z_{t-2} - Z_{t-3}) & -(Z_{t-1} - Z_{t-2})(Z_{t-2} - Z_{t-3}) \end{pmatrix}
$$

and

$$
\tilde{A}_{tx} = -\begin{pmatrix} 1 & Z_t \\ Z_{t-1} & Z_tZ_{t-1} \end{pmatrix} \quad \text{for} \quad x = 1, \dots, M-1.
$$

It is easy to show that as $T \to \infty$

$$\frac{1}{T}\sum_{t=4}^{T} Z_t = \frac{1}{T}\sum_{t=4}^{T} k_t + o_p(1) \xrightarrow{p} \frac{\mu_0}{1-\phi_{1,0}}, \quad \frac{1}{T}\sum_{t=4}^{T}(Z_{t-1}-Z_{t-2}) \xrightarrow{p} 0,$$

$$\frac{1}{T}\sum_{t=4}^{T}(Z_{t-2}-Z_{t-3})Z_{t-1} = \frac{1}{T}\sum_{t=4}^{T} k_{t-1}(k_{t-2}-k_{t-3}) + o_p(1) \xrightarrow{p} \frac{\phi_{1,0}E(e_1^2)}{1+\phi_{1,0}},$$

$$\frac{1}{T}\sum_{t=4}^{T}(Z_{t-1}-Z_{t-2})Z_{t-2} \xrightarrow{p} -\frac{E(e_1^2)}{1+\phi_{1,0}} - E(\eta_1^2),$$

$$\frac{1}{T}\sum_{t=4}^{T} Z_t Z_{t-1} \xrightarrow{p} \frac{\mu_0^2}{(1-\phi_{1,0})^2} + \frac{\phi_{1,0}E(e_1^2)}{1-\phi_{1,0}^2}$$

and

$$\frac{1}{T}\sum_{t=4}^{T}(Z_{t-1}-Z_{t-2})(Z_{t-2}-Z_{t-3}) \xrightarrow{p} -\frac{(1-\phi_{1,0})E(e_1^2)}{1+\phi_{1,0}} - E(\eta_1^2),$$

implying that

$$\frac{1}{T}\sum_{t=4}^{T}\frac{\partial}{\partial\boldsymbol{\theta}}\widetilde{\boldsymbol{W}}_t(\boldsymbol{\theta}_0) \xrightarrow{p} \widetilde{\boldsymbol{\Gamma}}_1 =: \begin{pmatrix} \tilde{A}_0 & 0 & \dots & 0 \\ 0 & \tilde{A}_1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \tilde{A}_{M-1} \end{pmatrix}, \tag{4.15}$$

with

$$\tilde{A}_0 = -\begin{pmatrix} 1 & \frac{\mu_0}{1-\phi_{1,0}} & 0 \\ \frac{\mu_0}{1-\phi_{1,0}} & \frac{\mu_0^2}{(1-\phi_{1,0})^2} + \frac{\phi_{1,0}E(e_1^2)}{1-\phi_{1,0}^2} & \frac{E(e_1^2)}{1+\phi_{1,0}} + E(\eta_1^2) \\ 0 & \frac{\phi_{1,0}E(e_1^2)}{1+\phi_{1,0}} & \frac{(1-\phi_{1,0})E(e_1^2)}{1+\phi_{1,0}} + E(\eta_1^2) \end{pmatrix}$$

and

$$\tilde{A}_x = -\begin{pmatrix} 1 & \frac{\mu_0}{1-\phi_{1,0}} \\ \frac{\mu_0}{1-\phi_{1,0}} & \frac{\mu_0^2}{(1-\phi_{1,0})^2} + \frac{\phi_{1,0}E(e_1^2)}{1-\phi_{1,0}^2} \end{pmatrix} \quad \text{for} \quad x = 1, \dots, M-1. \tag{4.16}$$

Hence Lemma 4.1ii) holds. □

Lemma 4.2. *Suppose model (4.4) holds with C), $\phi_{2,0} \neq 0$ and $\mu_0 \neq 0$.*

i) If $\tilde{\phi}_{1,0} = 1$ & $|\phi_{2,0}| < 1$, then

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \boldsymbol{W}_t^*(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}^*) \quad and \quad \{\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{W}_t^*(\boldsymbol{\theta}_0)\}\{\sqrt{T}D_T^{-1}\} \xrightarrow{p} \boldsymbol{\Gamma}_2 \quad as \quad T \to \infty,$$

where $\boldsymbol{W}_t^*$ is defined in Section 2, and $\boldsymbol{\Sigma}^*$ and $\boldsymbol{\Gamma}_2$ are given in (4.18) and (4.20) below respectively.

ii) If $\{k_t\}$ is stationary, then

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \widetilde{\boldsymbol{W}}_t^*(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}^*) \quad and \quad \frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\boldsymbol{\theta}}\widetilde{\boldsymbol{W}}_t^*(\boldsymbol{\theta}_0) \xrightarrow{p} \widetilde{\boldsymbol{\Gamma}}_2 \quad as \quad T \to \infty,$$

where $\widetilde{\boldsymbol{W}}_t^*$ is defined in Section 2, and $\widetilde{\boldsymbol{\Sigma}}^*$ and $\widetilde{\boldsymbol{\Sigma}}_2$ are given in (4.23) and (4.26) below respectively.

Proof. i) When $\tilde{\phi}_{1,0} = \phi_{1,0} + \phi_{2,0} = 1$ and $|\phi_{2,0}| < 1$, we have $k_t - k_{t-1} = \mu_0 - \phi_{2,0}(k_{t-1} - k_{t-2}) + e_t$, which implies that

$$k_t - k_{t-1} = \frac{\mu_0}{1+\phi_{2,0}} + (k_0 - k_{-1} - \frac{\mu_0}{1+\phi_{2,0}})(-\phi_{2,0})^t + u_t,$$

where $u_t = \sum_{i=1}^{t}(-\phi_{2,0})^{t-i}e_i$ satisfies Assumption 2.1 of Phillips (1987). Note that

$$\begin{cases} Y_{t,1}^* = e_t + \eta_t - \phi_{1,0}\eta_{t-1} - \phi_{2,0}\eta_{t-2}, \\ Y_{t,2}^* = (e_t + \eta_t - \phi_{1,0}\eta_{t-1} - \phi_{2,0}\eta_{t-2})(k_{t-3} + \eta_{t-3}), \\ Y_{t,3}^* = (e_t + \eta_t - \phi_{1,0}\eta_{t-1} - \phi_{2,0}\eta_{t-2})(k_{t-3} - k_{t-4} + \eta_{t-3} - \eta_{t-4}), \\ \tilde{Y}_{t,x}^* = \varepsilon_{x,t} - \beta_{x,0}\eta_t,, \quad x = 1, 2, \ldots, M-1, \\ \bar{Y}_{t,x}^* = (\varepsilon_{x,t} - \beta_{x,0}\eta_t)(k_{t-1} + \eta_{t-1}), \quad x = 1, 2, \ldots, M-1. \end{cases} \tag{4.17}$$

Define $\boldsymbol{U}_t^*(\boldsymbol{\theta}_0) = (V_{t,1}^*, V_{t,2}^*, V_{t,3}^*, \tilde{V}_{t,1}^*, \bar{V}_{t,1}^*, \ldots, \tilde{V}_{t,M-1}^*, \bar{V}_{t,M-1}^*)^\tau$, where

$$V_{t,1}^* = e_t, \quad V_{t,2}^* = (e_t + \eta_t)\frac{k_{t-3}}{T} - \phi_{1,0}\eta_t\frac{k_{t-2}}{T} - \phi_{2,0}\eta_t\frac{k_{t-1}}{T},$$

$$V_{t,3}^* \ = (e_t + \eta_t)(k_{t-3} - k_{t-4} + \eta_{t-3} - \eta_{t-4}) - \phi_{1,0}\eta_t(k_{t-2} - k_{t-3} + \eta_{t-2} - \eta_{t-3})$$
$$-\phi_{2,0}\eta_t(k_{t-1} - k_{t-2} + \eta_{t-1} - \eta_{t-2}),$$

$$\tilde{V}_{t,x}^* = \varepsilon_{x,t} - \beta_{x,0}\eta_t \ \text{ and } \ \bar{V}_{t,x}^* = (\varepsilon_{x,t} - \beta_{x,0}\eta_t)\frac{k_{t-1}}{T} \ \text{ for } x = 1, 2, \ldots, M-1.$$

Then we have

$$\frac{1}{\sqrt{T}}\sum_{t=5}^{T}\boldsymbol{W}_t^*(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{T}}\sum_{t=5}^{T}\boldsymbol{U}_t^*(\boldsymbol{\theta}_0) + o_p(1) \quad \text{as} \quad T \to \infty.$$

Since $\{\boldsymbol{U}_t^*(\boldsymbol{\theta}_0), \mathcal{F}_t\}_{t=1}^{\infty}$ is a martingale difference sequence, by the central limit theorem for a martingale difference sequence in Hall & Heyde (2014), $\frac{1}{\sqrt{T}}\sum_{t=5}^{T}\boldsymbol{U}_t^*(\boldsymbol{\theta}_0)$ converges in distribution to a multivariate normal distribution with zero means and covariance matrix

$$\boldsymbol{\Sigma}^* = \lim_{T\to\infty}\frac{1}{T}\sum_{t=5}^{T}E\{\boldsymbol{U}_t^*(\boldsymbol{\theta}_0)\boldsymbol{U}_t^*(\boldsymbol{\theta}_0)^\tau|\mathcal{F}_{t-1}\} = (\sigma_{ij}^*)_{i,j=1,2,\ldots,3+2(M-1)}, \tag{4.18}$$

i.e.,

$$\frac{1}{\sqrt{T}}\sum_{t=5}^{T}\boldsymbol{W}_t^*(\boldsymbol{\theta}_0) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma}^*) \ \text{ as } T \to \infty.$$

To compute the covariance matrix, by noting that

$$k_t = k_0 + \frac{\mu_0}{1+\phi_{2,0}}t + (k_0 - k_{-1} - \frac{\mu_0}{1+\phi_{2,0}})\frac{(-\phi_{2,0})\{1-(-\phi_{2,0})^t\}}{1+\phi_{2,0}} + \sum_{i=1}^{t}u_i,$$

we have that for $s \in [0, 1]$

$$\frac{k_{[Ts]}}{T} = \frac{\mu_0}{1+\phi_{2,0}}s + o_p(1) \quad \text{as} \quad T \to \infty. Since \tag{4.19}$$

It follows from (4.17), (4.19) that

$$E(k_t - k_{t-1}) = \frac{\mu_0}{1+\phi_{2,0}}, \quad E(k_t - k_{t-1})^2 = \frac{\mu_0^2}{(1+\phi_{2,0})^2} + \frac{E(e_1^2)}{1-\phi_{2,0}^2},$$

$$E\{(k_t-k_{t-1})(k_{t-1}-k_{t-2})\} = \frac{\mu_0^2}{(1+\phi_{2,0})^2} - \frac{\phi_{2,0}E(e_1^2)}{1-\phi_{2,0}^2}$$

$$E\{(k_t-k_{t-1})(k_{t-2}-k_{t-3})\} = \frac{\mu_0^2}{(1+\phi_{1,0})^2} + \frac{\phi_{2,0}^2E(e_1^2)}{1-\phi_{2,0}^2}.$$

Using the above equations, it is straightforward to verify that for $x, y = 1, \ldots, M-1$

$$\sigma_{1,1}^* = E(e_1^2), \quad \sigma_{1,2}^* = \frac{\mu_0}{2(1+\phi_{2,0})}E(e_1^2), \quad \sigma_{1,3}^* = \frac{\mu_0}{1+\phi_{2,0}}E(e_1^2), \quad \sigma_{1,2x+2}^* = \sigma_{1,2y+3}^* = 0,$$

$$\sigma_{2,2}^* = E(e_1^2)\{\lim_{T\to\infty}\frac{1}{T}\sum_{t=5}^{T}\frac{k_{t-3}^2}{T^2}\} = \frac{\mu_0^2}{3(1+\phi_{2,0})^2}E(e_1^2),$$

$$\sigma_{2,3}^* = E(e_1^2)\{\lim_{T\to\infty}\sum_{t=5}^{T}\frac{k_{t-3}}{T}(k_{t-3}-k_{t-4})\} = \frac{\mu_0^2}{2(1+\phi_{2,0})^2}E(e_1^2), \quad \sigma_{2,2x+2}^* = \sigma_{2,2x+3}^* = 0,$$

$$\begin{aligned}
\sigma_{3,3}^* &= E(e_t^2)E(k_{t-3}-k_{t-4}+\eta_{t-3}-\eta_{t-4})^2 \\
&\quad +E(\eta_t^2)E(k_{t-3}-k_{t-4}+\eta_{t-3}-\eta_{t-4})^2 \\
&\quad +\phi_{1,0}^2E(\eta_t^2)E(k_{t-2}-k_{t-3}+\eta_{t-2}-\eta_{t-3})^2 \\
&\quad +\phi_{2,0}^2E(\eta_t^2)E(k_{t-1}-k_{t-2}+\eta_{t-1}-\eta_{t-2})^2 \\
&\quad -2\phi_{1,0}E(\eta_t^2)E\{(k_{t-3}-k_{t-4}+\eta_{t-3}-\eta_{t-4})(k_{t-2}-k_{t-3}+\eta_{t-2}-\eta_{t-3})\} \\
&\quad -2\phi_{2,0}E(\eta_t^2)E\{(k_{t-3}-k_{t-4}+\eta_{t-3}-\eta_{t-4})(k_{t-1}-k_{t-2}+\eta_{t-1}-\eta_{t-2})\} \\
&\quad +2\phi_{1,0}\phi_{2,0}E(\eta_t^2)E\{(k_{t-2}-k_{t-3}+\eta_{t-2}-\eta_{t-3})(k_{t-1}-k_{t-2}+\eta_{t-1}-\eta_{t-2})\} \\
&= \{E(e_1^2)+(1+\phi_{1,0}^2+\phi_{2,0}^2)E(\eta_1^2)\}\{\tfrac{\mu_0^2}{(1+\phi_{2,0})^2}+\tfrac{E(e_1^2)}{1-\phi_{2,0}^2}+2E(\eta_1^2)\} \\
&\quad +(-2\phi_{1,0}+2\phi_{1,0}\phi_{2,0})E(\eta_1^2)\{\tfrac{\mu_0^2}{(1+\phi_{2,0})^2}-\tfrac{\phi_{2,0}E(e_1^2)}{1-\phi_{2,0}^2}-E(\eta_1^2)\} \\
&\quad -2\phi_{2,0}E(\eta_1^2)\{\tfrac{\mu_0^2}{(1+\phi_{1,0})^2}+\tfrac{\phi_{2,0}^2E(e_1^2)}{1-\phi_{2,0}^2}\},
\end{aligned}$$

$$\sigma_{3,2x+2}^* = \sigma_{3,2y+3}^* = 0, \quad \sigma_{2x+2,2y+2}^* = E\{(\varepsilon_{x,1}-\beta_{x,0}\eta_1)(\varepsilon_{y,1}-\beta_{y,0}\eta_1)\}$$

$$\sigma_{2x+2,2y+3}^* = \frac{\mu_0}{2(1+\phi_{2,0})}E\{(\varepsilon_{x,1}-\beta_{x,0}\eta_1)(\varepsilon_{y,1}-\beta_{y,0}\eta_1)\},$$

$$\sigma^*_{2x+3,2y+3} = \frac{\mu_0^2}{3(1+\phi_{2,0})^2} E\{(\varepsilon_{x,1} - \beta_{x,0}\eta_1)(\varepsilon_{y,1} - \beta_{y,0}\eta_1)\}.$$

Further by noting that

$$\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{W}^*_t(\boldsymbol{\theta}_0) = \begin{pmatrix} A^*_{t0} & 0 & \dots & 0 \\ 0 & A^*_{t1} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & A^*_{t(M-1)} \end{pmatrix},$$

where

$$A^*_{t0} = -\begin{pmatrix} 1 & Z_{t-1} & -(Z_{t-1} - Z_{t-2}) \\ \frac{1}{T} Z_{t-3} & \frac{1}{T} Z_{t-1} Z_{t-3} & -\frac{1}{T}(\hat{Z}_{t-1} - Z_{t-2}) Z_{t-3} \\ Z_{t-3} - \hat{Z}_{t-4} & Z_{t-1}(Z_{t-3} - Z_{t-4}) & -(Z_{t-1} - Z_{t-2})(Z_{t-3} - Z_{t-4}) \end{pmatrix}$$

and

$$A^*_{tx} = -\begin{pmatrix} 1 & Z_t \\ \frac{1}{T} Z_{t-1} & \frac{1}{T} Z_t Z_{t-1} \end{pmatrix} \text{ for } x = 1, \dots, M-1,$$

we have

$$\frac{1}{T}\{\sum_{t=5}^{T} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{W}^*_t(\boldsymbol{\theta}_0)\}\{\sqrt{T} D_T^{-1}\} \xrightarrow{p} \boldsymbol{\Gamma}_2 =: \begin{pmatrix} A^*_0 & 0 & \dots & 0 \\ 0 & A^*_1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & A^*_{M-1} \end{pmatrix}, \tag{4.20}$$

with

$$A^*_0 = -\begin{pmatrix} 1 & \frac{\mu_0}{2(1+\phi_{2,0})} & -\frac{\mu_0}{1+\phi_{2,0}} \\ \frac{\mu_0}{2(1+\phi_{2,0})} & \frac{\mu_0^2}{3(1+\phi_{2,0})^2} & -\frac{\mu_0^2}{2(1+\phi_{2,0})^2} \\ \frac{\mu_0}{1+\phi_{2,0}} & \frac{\mu_0^2}{2(1+\phi_{2,0})^2} & -\frac{\mu_0^2}{(1+\phi_{2,0})^2} - \frac{\phi_{2,0}^2 E(e_1^2)}{1-\phi_{2,0}^2} \end{pmatrix}$$

and

$$A_x^* = -\begin{pmatrix} 1 & \frac{\mu_0}{2(1+\phi_{2,0})} \\ \frac{\mu_0}{2(1+\phi_{2,0})} & \frac{\mu_0^2}{3(1+\phi_{2,0})^2} \end{pmatrix} \quad \text{for} \quad x = 1, \ldots, M-1. \tag{4.21}$$

Hence Lemma 4.2i) holds.

ii) When the roots of $1 - \phi_{1,0}x - \phi_{2,0}x^2 = 0$ are outside of the unit circle, $\{k_t\}$ is a stationary sequence with $E(k_t) = \frac{\mu_0}{1-\phi_{1,0}-\phi_{2,0}}$. Note that

$$\begin{cases} Y_{t,1}^* = e_t + \eta_t - \phi_{1,0}\eta_{t-1} - \phi_{2,0}\eta_{t-2}, \\ Y_{t,2}^* = (e_t + \eta_t - \phi_{1,0}\eta_{t-1} - \phi_{2,0}\eta_{t-2})(k_{t-3} + \eta_{t-3}), \\ Y_{t,3}^* = (e_t + \eta_t - \phi_{1,0}\eta_{t-1} - \phi_{2,0}\eta_{t-2})(k_{t-3} - k_{t-4} + \eta_{t-3} - \eta_{t-4}), \\ \tilde{Y}_{t,x}^* = \varepsilon_{x,t} - \beta_{x,0}\eta_t, \quad x = 1, 2, \ldots, M-1, \\ \bar{Y}_{t,x}^* = (\varepsilon_{x,t} - \beta_{x,0}\eta_t)(k_{t-1} + \eta_{t-1}), \quad x = 1, 2, \ldots, M-1. \end{cases} \tag{4.22}$$

Define $X_{t,1}^* = e_t + (1 - \phi_{1,0} - \phi_{2,0})\eta_t$,

$$X_{t,2}^* = (e_t + \eta_t)(k_{t-3} + \eta_{t-3}) - \phi_{1,0}\eta_t(k_{t-2} + \eta_{t-2}) - \phi_{2,0}\eta_t(k_{t-1} + \eta_{t-1}),$$

$$\begin{aligned} X_{t,3}^* &= (e_t + \eta_t)(k_{t-3} - k_{t-4} + \eta_{t-3} - \eta_{t-4}) - \phi_{1,0}\eta_t(k_{t-2} - k_{t-3} + \eta_{t-2} - \eta_{t-3}) \\ &\quad -\phi_{2,0}\eta_t(k_{t-1} - k_{t-2} + \eta_{t-1} - \eta_{t-2}), \end{aligned}$$

and $\widetilde{\boldsymbol{U}}_t^*(\boldsymbol{\theta}_0)$ as $\widetilde{\boldsymbol{W}}_t^*(\boldsymbol{\theta}_0)$ with $Y_{t,1}^*$, $Y_{t,2}^*$, $Y_{t,3}^*$ replaced by $X_{t,1}^*$, $X_{t,2}^*$, $X_{t,3}^*$ respectively.

By the central limit theorem for a martingale difference sequence in Hall & Heyde (2014), $\frac{1}{\sqrt{T}}\sum_{t=5}^{T}\widetilde{\boldsymbol{U}}_t^*(\boldsymbol{\theta}_0)$ converges in distribution to a multivariate normal distribution with zero means and covariance matrix

$$\widetilde{\boldsymbol{\Sigma}}^* = \lim_{T\to\infty}\frac{1}{T}\sum_{t=5}^{T} E\{\widetilde{\boldsymbol{U}}_t^*(\boldsymbol{\theta}_0)\widetilde{\boldsymbol{U}}_t^*(\boldsymbol{\theta}_0)^\tau | \mathcal{F}_{t-1}\} = (\tilde{\sigma}_{ij}^*)_{i,j=1,2,\ldots,3+2(M-1)}, \tag{4.23}$$

i.e.,

$$\frac{1}{\sqrt{T}}\sum_{t=5}^{T}\widetilde{\boldsymbol{W}}_t^*(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{T}}\sum_{t=5}^{T}\widetilde{\boldsymbol{U}}_t^*(\boldsymbol{\theta}_0) + o_p(1) \xrightarrow{d} N(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}}^*) \ \text{ as } T \to \infty.$$

To compute the covariance matrix, by noting that $\{k_t\}$ is stationary, we have

$$\begin{cases} E(k_t^2) = \frac{\mu_0^2}{(1-\phi_{1,0}-\phi_{2,0})^2} + \frac{1-\phi_{2,0}}{1+\phi_{2,0}} \frac{E(e_1^2)}{(1-\phi_{2,0})^2-\phi_{1,0}^2} \\ E(k_t k_{t-1}) = \frac{\mu_0^2}{(1-\phi_{1,0}-\phi_{2,0})^2} + \frac{\phi_{1,0}}{1+\phi_{2,0}} \frac{E(e_1^2)}{(1-\phi_{2,0})^2-\phi_{1,0}^2} \\ E(k_t k_{t-2}) = \frac{\mu_0^2}{(1-\phi_{1,0}-\phi_{2,0})^2} + \frac{\phi_{1,0}^2+\phi_{2,0}-\phi_{2,0}^2}{1+\phi_{2,0}} \frac{E(e_1^2)}{(1-\phi_{2,0})^2-\phi_{1,0}^2} \\ E(k_t k_{t-3}) = \frac{\mu_0^2}{(1-\phi_{1,0}-\phi_{2,0})^2} + \phi_{1,0}\frac{\phi_{1,0}^2+2\phi_{2,0}-\phi_{2,0}^2}{1+\phi_{2,0}} \frac{E(e_1^2)}{(1-\phi_{2,0})^2-\phi_{1,0}^2}, \end{cases} \tag{4.24}$$

which imply that

$$\begin{cases} E\{k_t(k_t - k_{t-1})\} = \frac{E(e_1^2)}{(1+\phi_{2,0})(1-\phi_{2,0}+\phi_{1,0})} := \Delta, \\ E\{k_t(k_{t-1} - k_{t-2})\} = (\phi_{1,0} - \phi_{2,0})\Delta, \\ E\{k_t(k_{t-2} - k_{t-3})\} = (\phi_{2,0} + \phi_{1,0}^2 - \phi_{1,0}\phi_{2,0})\Delta, \\ E\{k_{t-1}(k_t - k_{t-1})\} = -\Delta, \\ Since E\{k_{t-2}(k_t - k_{t-1})\} = (\phi_{2,0} - \phi_{1,0})\Delta, \\ E\{(k_t - k_{t-1})^2\} = 2\Delta, \\ E\{(k_t - k_{t-1})(k_{t-1} - k_{t-2})\} = (\phi_{1,0} - \phi_{2,0} - 1)\Delta, \\ E\{(k_t - k_{t-1})(k_{t-2} - k_{t-3})\} = (2\phi_{2,0} - \phi_{1,0} + \phi_{1,0}^2 - \phi_{1,0}\phi_{2,0})\Delta. \end{cases} \tag{4.25}$$

It follows from (4.24) and (4.25) that, for $x = 1, \ldots, M-1$,

$$\tilde{\sigma}_{1,1}^* = E(e_1^2) + (1 - \phi_{1,0} - \phi_{2,0})^2 E(\eta_1^2),$$

$$\tilde{\sigma}_{1,2}^* = \frac{\mu_0}{1 - \phi_{1,0} - \phi_{2,0}}\{E(e_1^2) + (1 - \phi_{1,0} - \phi_{2,0})^2 E(\eta_1^2)\}, \quad \tilde{\sigma}_{1,3}^* = 0,$$

$$\tilde{\sigma}_{1,2x+2}^* = (1 - \phi_{1,0} - \phi_{2,0})E\{\eta_1(\varepsilon_{x,1} - \beta_{x,0}\eta_1)\}, \quad \tilde{\sigma}_{1,2x+3}^* = \mu_0 E\{\eta_1(\varepsilon_{x,1} - \beta_{x,0}\eta_1)\},$$

$$
\begin{aligned}
\tilde{\sigma}^*_{2,2} &= E(e_1^2)E(k_{t-3}+\eta_{t-3})^2 + E(\eta_1^2)E\{k_{t-3}+\eta_{t-3}-\phi_{1,0}(k_{t-2}+\eta_{t-2})-\phi_{2,0}(k_{t-1}+\eta_{t-1})\}^2 \\
&= E(e_1^2)\{\tfrac{\mu_0^2}{(1-\phi_{1,0}-\phi_{2,0})^2}+\tfrac{1-\phi_{2,0}}{1+\phi_{2,0}}\tfrac{E(e_1^2)}{(1-\phi_{2,0})^2-\phi_{1,0}^2}+E(\eta_1^2)\} \\
&\quad +E(\eta_1^2)\{\mu_0^2+E(e_1^2)+(1+\phi_{1,0}^2+\phi_{2,0}^2)E(\eta_1^2)\},
\end{aligned}
$$

$$
\begin{aligned}
\tilde{\sigma}^*_{2,3} &= \{E(e_1^2)+E(\eta_1^2)\}\{\Delta+E(\eta_1^2)\} - \phi_{1,0}E(\eta_1^2)\{-\Delta-E(\eta_1^2)\} - \phi_{2,0}E(\eta_1^2)(\phi_{2,0}-\phi_{1,0})\Delta \\
&\quad -\phi_{1,0}E(\eta_1^2)(\phi_{1,0}-\phi_{2,0})\Delta + \phi_{1,0}^2E(\eta_1^2)\{\Delta+E(\eta_1^2)\} + \phi_{1,0}\phi_{2,0}E(\eta_1^2)\{-\Delta-E(\eta_1^2)\} \\
&\quad -\phi_{2,0}E(\eta_1^2)(\phi_{2,0}+\phi_{1,0}^2-\phi_{1,0}\phi_{2,0})\Delta + \phi_{1,0}\phi_{2,0}E(\eta_1^2)(\phi_{1,0}-\phi_{2,0})\Delta \\
&\quad +\phi_{2,0}^2E(\eta_1^2)\{\Delta+E(\eta_1^2)\} \\
&= E(e_1^2)\Delta + 2E(e_1^2)E(\eta_1^2) + \{E(\eta_1^2)\}^2(1+\phi_{1,0}+\phi_{1,0}^2-\phi_{1,0}\phi_{2,0}+\phi_{2,0}^2),
\end{aligned}
$$

$$
\tilde{\sigma}^*_{2,2x+2} = \mu_0 E\{\eta_1(\varepsilon_{x,1}-\beta_{x,0}\eta_1)\}, \quad \sigma^*_{2,2x+3} = E\{\eta_1(\varepsilon_{x,1}-\beta_{x,0}\eta_1)\}\{\frac{\mu_0^2}{1-\phi_{1,0}-\phi_{2,0}}-\phi_{2,0}E(\eta_1^2)\},
$$

$$
\begin{aligned}
\tilde{\sigma}^*_{3,3} &= \{E(e_1^2)+(1+\phi_{1,0}^2+\phi_{2,0}^2)E(\eta_1^2)\}\{2\Delta+2E(\eta_1^2)\} \\
&\quad -2\phi_{1,0}E(\eta_1^2)\{(\phi_{1,0}-\phi_{2,0}-1)\Delta-E(\eta_1^2)\} \\
&\quad -2\phi_{2,0}E(\eta_1^2)(2\phi_{2,0}-\phi_{1,0}+\phi_{1,0}^2-\phi_{1,0}\phi_{2,0})\Delta \\
&\quad +2\phi_{1,0}\phi_{2,0}E(\eta_1^2)\{(\phi_{1,0}-\phi_{2,0}-1)\Delta-E(\eta_1^2)\} \\
&= 2E(e_1^2)\Delta + 4E(e_1^2)E(\eta_1^2) + 2\{E(\eta_1^2)\}^2(1+\phi_{1,0}+\phi_{1,0}^2-\phi_{1,0}\phi_{2,0}+\phi_{2,0}^2) \\
&= 2\tilde{\sigma}^*_{2,3},
\end{aligned}
$$

$$
\tilde{\sigma}^*_{3,2x+2} = 0, \quad \tilde{\sigma}^*_{3,2x+3} = -\phi_{2,0}E(\eta_1^2)E\{\eta_1(\varepsilon_{x,1}-\beta_{x,0}\eta_1)\},
$$

$$
\tilde{\sigma}^*_{2x+2,2y+2} = E\{(\varepsilon_{x,1}-\beta_{x,0}\eta_1)(\varepsilon_{y,1}-\beta_{y,0}\eta_1)\},
$$

$$\tilde{\sigma}^*_{2x+2,2y+3} = \frac{\mu_0}{1-\phi_{1,0}-\phi_{2,0}} E\{(\varepsilon_{x,1}-\beta_{x,0}\eta_1)(\varepsilon_{y,1}-\beta_{y,0}\eta_1)\},$$

$$\tilde{\sigma}^*_{2x+3,2y+3} = E\{(\varepsilon_{x,1}-\beta_{x,0}\eta_1)(\varepsilon_{y,1}-\beta_{y,0}\eta_1)\}\{\tfrac{\mu_0^2}{(1-\phi_{1,0}-\phi_{2,0})^2} + \tfrac{1-\phi_{2,0}}{1+\phi_{2,0}}\tfrac{E(e_1^2)}{(1-\phi_{2,0})^2-\phi_{1,0}^2} + E(\eta_1^2)\}.$$

Further by noting that

$$\frac{\partial}{\partial\boldsymbol{\theta}}\widetilde{\boldsymbol{W}}^*_t(\boldsymbol{\theta}_0) = \begin{pmatrix} \tilde{A}^*_{t0} & 0 & \dots & 0 \\ 0 & \tilde{A}^*_{t1} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \tilde{A}^*_{t(M-1)} \end{pmatrix},$$

where

$$\tilde{A}^*_{t0} = -\begin{pmatrix} 1 & Z_{t-1} & -(Z_{t-1}-Z_{t-2}) \\ Z_{t-3} & Z_{t-1}Z_{t-3} & -(Z_{t-1}-Z_{t-2})Z_{t-3} \\ Z_{t-3}-Z_{t-4} & Z_{t-1}(Z_{t-3}-Z_{t-4}) & -(Z_{t-1}-Z_{t-2})(Z_{t-3}-Z_{t-4}) \end{pmatrix},$$

and

$$\tilde{A}^*_{tx} = -\begin{pmatrix} 1 & Z_t \\ Z_{t-1} & Z_tZ_{t-1} \end{pmatrix} \text{ for } x = 1,\dots,M-1,$$

we can show that, as $T\to\infty$,

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\boldsymbol{\theta}}\widetilde{\boldsymbol{W}}^*_t(\boldsymbol{\theta}_0) \xrightarrow{p} \tilde{\boldsymbol{\Gamma}}_2 := \begin{pmatrix} \tilde{A}^*_0 & 0 & \dots & 0 \\ 0 & \tilde{A}^*_1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \tilde{A}^*_{M-1} \end{pmatrix}, \tag{4.26}$$

where

$$\tilde{A}^*_x = -\begin{pmatrix} 1 & \frac{\mu_0}{1-\phi_{1,0}-\phi_{2,0}} \\ \frac{\mu_0}{1-\phi_{1,0}-\phi_{2,0}} & \frac{\mu_0^2}{(1-\phi_{1,0}-\phi_{2,0})^2} + \frac{\phi_{1,0}}{1+\phi_{2,0}}\frac{E(e_1^2)}{(1-\phi_{2,0})^2-\phi_{1,0}^2} \end{pmatrix}, \quad x = 1,\dots,M-1,$$

and

$$\tilde{A}_0^* = -\begin{pmatrix} 1 & \frac{\mu_0}{1-\phi_{1,0}-\phi_{2,0}} & 0 \\ \frac{\mu_0}{1-\phi_{1,0}-\phi_{2,0}} & \frac{\mu_0^2}{(1-\phi_{1,0}-\phi_{2,0})^2} + \frac{(\phi_{1,0}^2+\phi_{2,0}-\phi_{2,0}^2)E(e_1^2)}{(1+\phi_{2,0})\{(1-\phi_{2,0})^2-\phi_{1,0}^2\}} & (\phi_{1,0}-\phi_{2,0})\Delta \\ 0 & (\phi_{2,0}+\phi_{1,0}^2-\phi_{1,0}\phi_{2,0})\Delta & (\phi_{1,0}-2\phi_{2,0}-\phi_{1,0}^2+\phi_{1,0}\phi_{2,0})\Delta \end{pmatrix}.$$

Hence Lemma 4.2i) follows from the above equations.

□

Proof of Theorem 4.1. Note that

$$0 = \sum_{t=4}^{T} \boldsymbol{W}_t(\hat{\boldsymbol{\theta}}) = \sum_{t=4}^{T} \boldsymbol{W}_t(\boldsymbol{\theta}_0) + \sum_{t=1}^{T} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{W}_t(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

and

$$0 = \sum_{t=4}^{T} \widetilde{\boldsymbol{W}}_t(\hat{\boldsymbol{\theta}}) = \sum_{t=4}^{T} \widetilde{\boldsymbol{W}}_t(\boldsymbol{\theta}_0) + \sum_{t=1}^{T} \frac{\partial}{\partial \boldsymbol{\theta}} \widetilde{\boldsymbol{W}}_t(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0).$$

Hence,

$$D_T\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\} = -\left\{\frac{1}{T}\sum_{t=4}^{T} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{W}_t(\boldsymbol{\theta}_0)(\sqrt{T} D_T^{-1})\right\}^{-1} \left\{\frac{1}{\sqrt{T}} \sum_{t=4}^{T} \boldsymbol{W}_t(\boldsymbol{\theta}_0)\right\}$$

when $\tilde{\phi}_{1,0} = 1$, and

$$\sqrt{T}\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\} = -\left\{\frac{1}{T}\sum_{t=4}^{T} \frac{\partial}{\partial \boldsymbol{\theta}} \widetilde{\boldsymbol{W}}_t(\boldsymbol{\theta}_0)\right\}^{-1} \left\{\frac{1}{\sqrt{T}} \sum_{t=4}^{T} \widetilde{\boldsymbol{W}}_t(\boldsymbol{\theta}_0)\right\}$$

when $|\tilde{\phi}_{1,0}| < 1$. Therefore the theorem follows from Lemma 4.1. □

Proof of Theorem 4.2. The theorem follows from the arguments in proving (4.12) and (4.15).

□

Proof of Theorem 4.3. Theorem 4.3 follows from Lemma 4.2 by using the same arguments in proving Theorem 4.1. □

Proof of Theorem 4.4. The theorem follows from the arguments in proving (4.20) and (4.26).

□

Table (4.3) Simulation results for $T = 300$, $\phi_1 = 1.208 - 0.2$, $\phi_2 = -0.218 + 0.2$, $sd(\varepsilon_{x,t}) = sd(e_t) = 0.01$.

	x	1	2	3	4	5	6	7	8	9	10
Assume $\phi_2 = 0$	$\hat{\alpha}_x$	-2.392	-1.955	-1.040	-0.334	0.109	0.670	0.940	1.152	1.382	1.468
		(1.685e-3)	(1.679e-3)	(1.687e-3)	(1.673e-3)	(1.666e-3)	(1.675e-3)	(1.691e-3)	(1.698e-3)	(1.687e-3)	(1.693e-3)
	$\hat{\beta}_x$	0.081	0.087	0.101	0.108	0.108	0.110	0.107	0.103	0.100	0.094
		(3.940e-5)	(3.935e-5)	(3.920e-5)	(3.907e-5)	(3.905e-5)	(3.906e-5)	(3.946e-5)	(3.941e-5)	(3.958e-5)	(3.937e-5)
Assume $\phi_2 \neq 0$	$\hat{\alpha}_x$	-2.392	-1.955	-1.040	-0.334	0.109	0.670	0.940	1.152	1.382	1.468
		(1.709e-3)	(1.699e-3)	(1.711e-3)	(1.695e-3)	(1.691e-3)	(1.697e-3)	(1.708e-3)	(1.722e-3)	(1.711e-3)	(1.717e-3)
	$\hat{\beta}_x$	0.081	0.087	0.101	0.108	0.108	0.110	0.107	0.103	0.100	0.094
		(3.990e-5)	(3.975e-5)	(3.967e-5)	(3.954e-5)	(3.961e-5)	(3.947e-5)	(3.980e-5)	(3.991e-5)	(4.005e-5)	(3.983e-5)

	μ	$\tilde{\phi}_1$	ϕ_1	ϕ_2
Assume $\phi_2 = 0$	-0.601	0.989	1.006	-0.016
	(0.093)	(1.635e-3)	(0.152)	(0.151)
Assume $\phi_2 \neq 0$	-0.932	0.984	0.464	0.520
	(15.866)	(0.279)	(26.005)	(25.726)

Table (4.4) Simulation results for $T = 300$, $\phi_1 = 1.208$, $\phi_2 = -0.218$, $sd(\varepsilon_{x,t}) = sd(e_t) = 0.01$.

	x	1	2	3	4	5	6	7	8	9	10
Assume $\phi_2 = 0$	$\hat{\alpha}_x$	-2.392	-1.955	-1.040	-0.334	0.109	0.670	0.940	1.152	1.381	1.468
		(1.833e-3)	(1.827e-3)	(1.835e-3)	(1.821e-3)	(1.811e-3)	(1.823e-3)	(1.840e-3)	(1.851e-3)	(1.837e-3)	(1.843e-3)
	$\hat{\beta}_x$	0.081	0.087	0.101	0.108	0.108	0.110	0.107	0.103	0.100	0.094
		(4.011e-5)	(4.005e-5)	(3.989e-5)	(3.978e-5)	(3.972e-5)	(3.977e-5)	(4.021e-5)	(4.021e-5)	(4.033e-5)	(4.012e-5)
Assume $\phi_2 \neq 0$	$\hat{\alpha}_x$	-2.392	-1.955	-1.040	-0.334	0.109	0.670	0.940	1.152	1.381	1.468
		(1.863e-3)	(1.853e-3)	(1.865e-3)	(1.850e-3)	(1.844e-3)	(1.850e-3)	(1.863e-3)	(1.881e-3)	(1.868e-3)	(1.874e-3)
	$\hat{\beta}_x$	0.081	0.087	0.101	0.108	0.108	0.110	0.107	0.103	0.100	0.094
		(4.070e-5)	(4.055e-5)	(4.047e-5)	(4.035e-5)	(4.041e-5)	(4.028e-5)	(4.063e-5)	(4.081e-5)	(4.089e-5)	(4.069e-5)

	μ	$\tilde{\phi}_1$	ϕ_1	ϕ_2
Assume $\phi_2 = 0$	-0.761	0.987	0.998	-0.011
	(0.119)	(2.096e-3)	(0.155)	(0.153)
Assume $\phi_2 \neq 0$	-0.955	0.983	0.742	0.241
	(92.313)	(1.622)	(119.762)	(118.139)

Table (4.5) Simulation results for $T = 300$, $\phi_1 = 1.208 + 0.4$, $\phi_2 = -0.218 - 0.4$, $sd(\varepsilon_{x,t}) = sd(e_t) = 0.01$.

	x	1	2	3	4	5	6	7	8	9	10
Assume $\phi_2 = 0$	$\hat{\alpha}_x$	-2.392 (2.525e-3)	-1.955 (2.532e-3)	-1.040 (2.531e-3)	-0.334 (2.520e-3)	0.109 (2.500e-3)	0.670 (2.524e-3)	0.940 (2.545e-3)	1.152 (2.573e-3)	1.382 (2.545e-3)	1.468 (2.553e-3)
	$\hat{\beta}_x$	0.081 (4.874e-5)	0.087 (4.882e-5)	0.101 (4.854e-5)	0.108 (4.845e-5)	0.108 (4.827e-5)	0.110 (4.854e-5)	0.107 (4.902e-5)	0.103 (4.937e-5)	0.100 (4.915e-5)	0.094 (4.900e-5)
Assume $\phi_2 \neq 0$	$\hat{\alpha}_x$	-2.392 (2.603e-3)	-1.955 (2.601e-3)	-1.040 (2.609e-3)	-0.334 (2.600e-3)	0.109 (2.587e-3)	0.670 (2.600e-3)	0.940 (2.611e-3)	1.152 (2.652e-3)	1.382 (2.622e-3)	1.468 (2.631e-3)
	$\hat{\beta}_x$	0.081 (5.015e-5)	0.087 (5.008e-5)	0.101 (4.994e-5)	0.108 (4.984e-5)	0.108 (4.990e-5)	0.110 (4.988e-5)	0.107 (5.020e-5)	0.103 (5.081e-5)	0.100 (5.051e-5)	0.094 (5.037e-5)

	μ	$\tilde{\phi}_1$	ϕ_1	ϕ_2
Assume $\phi_2 = 0$	-5.295 (594.785)	0.907 (10.419)	-1.365 (376.081)	2.272 (365.663)
Assume $\phi_2 \neq 0$	-0.528 (0.502)	0.991 (8.816e-3)	1.653 (0.312)	-0.662 (0.303)

PART 5

STATISTICAL INFERENCE FOR A MODIFIED TWO-POPULATION LEE-CARTER MORTALITY MODEL

This Part is my working paper which is currently under review, and has been adapted to the format of dissertation.

Lee-Carter mortality model in Lee & Carter (1992) has become a benchmark model in forecasting mortality, hedging longevity risk and pricing annuities. This model is a combination of the following two structures for modeling the central death rate $m(x,t)$ at age or age group $x = 1, \ldots, M$ and time $t = 1, \ldots, T$:

$$\log m(x,t) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}, \quad \sum_{x=1}^{M} \beta_x = 1, \quad \sum_{t=1}^{T} k_t = 0, \tag{5.1}$$

and

$$k_t = \mu + \rho k_{t-1} + e_t, \tag{5.2}$$

where $\{\varepsilon_{x,t}\}_{t=1}^{T}$ and $\{e_t\}_{t=1}^{T}$ are random errors with zero mean and finite variance, and the unobserved $\{k_t\}$ is called the mortality index. Although model (5.2) can be replaced by a more general time series model such as ARIMA model, many actuarial applications in the study of longevity risk simply assume $\rho = 1$ in (5.2); see Li et al. (2017b), Kwok et al. (2016), Enchev et al. (2017), Biffis et al. (2017), Lin et al. (2017), Wong et al. (2017), and Zhu et al. (2017).

A two-step statistical inference procedure in Lee & Carter (1992) is to first fit (5.1) by the singular value decomposition method and then fit (5.2) by the least squares estimation based on the estimated mortality index in the first step. This procedure has been widely applied in actuarial science without verifying the correctness. Until recently, Leng & Peng (2016) showed that such an inference procedure may be inconsistent when the mortality index is not exactly an AR(1) unit root process. As k_t's are unobservable, testing for unit root becomes nontrivial at all. An application of the unit root test developed in Leng & Peng (2017) shows that assuming unit root AR(1) model for $\{k_t\}$ is not suitable for some mortality data sets. Because fitting a time series model to $\{k_t\}$ is necessary for the purpose of forecasting mortality risk, the last constraint in (5.1) becomes extremely restrictive and basically implies that $\mu = 0$ in (5.2).

To solve the issues on inference inconsistency and restrictive constraint, in Part 2 we proposed the following modified Lee-Carter model:

$$\log m(x,t) = \alpha_x + \beta_x k_t + \varepsilon_{x,t}, \quad k_t = \mu + \phi k_{t-1} + e_t, \quad \sum_{x=1}^{M} \beta_x = 1, \quad \sum_{x=1}^{M} \alpha_x = 0. \qquad (5.3)$$

When $\{k_t\}$ is unit root or near unit root, least squares estimators are proposed and proved to be consistent with a normal limit, and a unit root test is provided too. However, these estimators become biased when $\{k_t\}$ is stationary. To unify the inference for (5.3), in Part 3 we proposed a bias-corrected inference procedure regardless of $\{k_t\}$ being stationary or near unit root or unit root.

When longevity risk transfers involve more than one population, it is important to develop multipopulation mortality models. Two-population stochastic mortality models have been studied and applied in recent years; see Li et al. (2015) and references therein. Many of these existing two-population mortality models are built upon the above Lee-Carter framework like Li & Lee (2005), Li & Hardy (2011), and Cairns et al. (2011). Hence statistical inference is similar to the two-step procedure in Lee & Carter (1992), which uses the singular value decomposition method to first estimate mortality index for each population and then infer the time series models. With no surprising, none of these papers concern the correctness and asymptotic properties of the employed statistical inference. Given the recent developments of statistical inference for Lee-Carter model in Leng & Peng (2016) and the proposed models in Parts 2 and 3, we conjecture that the widely employed statistical inference for two-population stochastic mortality models is problematic. After confirming this via a simulation study and locating the reason of inconsistency, we propose a unified inference regardless of the two mortality indexes being stationary or unit root.

More specifically, this Part considers a modified two-population Lee-Carter mortality model and the corresponding bias-corrected inference procedure, which is valid regardless of mortality indexes being stationary or unit root; see section 5.1 for details on methodologies and asymptotic results. A simulation study and data analyses are given in section 5.2. Section 5.3 summarizes our contributions. All proofs are put into the Appendix.

5.1 Methodologies and Asymptotic Results

Consider the bivariate mortality model

$$\log m^{(1)}(x,t) = \alpha_x^{(1)} + \beta_x^{(1)} k_t^{(1)} + \varepsilon_{x,t}^{(1)},\ \ \log m^{(2)}(x,t) = \alpha_x^{(2)} + \beta_x^{(2)} k_t^{(2)} + \varepsilon_{x,t}^{(2)}, \tag{5.4}$$

$$k_t^{(1)} = \mu^{(1)} + \phi^{(1)} k_{t-1}^{(1)} + e_t^{(1)},\ \ k_t^{(1)} - k_t^{(2)} = \mu^{(2)} + \phi^{(2)}(k_{t-1}^{(1)} - k_{t-1}^{(2)}) + e_t^{(2)}, \tag{5.5}$$

where $\{(\varepsilon_{x,t}^{(1)}, \varepsilon_{x,t}^{(2)})^\tau\}_{t=1}^T$ is a sequence of independent and identically distributed random vectors with zero means and finite variances for each x, $\{(e_t^{(1)}, e_t^{(2)})^\tau\}_{t=1}^T$ is a sequence of independent and identically distributed random vectors with zero means and finite variances, and A^τ denotes the transpose of the matrix or vector A. Due to the identification issue, like Lee-Carter model, one often assumes

$$\sum_{x=1}^M \beta_x^{(j)} = 1,\ \ \sum_{t=1}^T k_t^{(j)} = 0 \text{ for } j = 1, 2. \tag{5.6}$$

As argued in the introduction, the constraints $\sum_{t=1}^T k_t^{(j)} = 0$ for $j = 1, 2$ are very restrictive, and the two-step procedure with singular value decomposition method will lead to an inconsistent inference even under the often used assumptions of $\phi^{(1)} = 1$ and $|\phi^{(2)}| < 1$ in the study of longevity risk.

Like Part 2, we avoid adding constraints to the random mortality indexes by assuming that

$$\sum_{x=1}^{M} \alpha_x^{(i)} = 0, \quad \sum_{x=1}^{M} \beta_x^{(i)} = 1, \quad i = 1, 2, \tag{5.7}$$

where $\sum_{x=1}^{M} \alpha_x^{(i)} = 0$ is not restrictive at all as the sum can be absorbed into $k_t^{(i)}$. Define $\eta_t^{(i)} = \sum_{x=1}^{M} \epsilon_{x,t}^{(i)}$ and $Z_t^{(i)} = \sum_{x=1}^{M} \log m^{(i)}(x,t)$, then we have

$$Z_t^{(i)} = k_t^{(i)} + \eta_t^{(i)}, \quad i = 1, 2.$$

When both $\{k_t^{(1)}\}$ and $\{k_t^{(1)} - k_t^{(2)}\}$ are near unit root or unit root, as t large enough, $Z_t^{(1)}$ and $k_t^{(1)}$ are approximately the same, and $Z_t^{(1)} - Z_t^{(2)}$ and $k_t^{(1)} - k_t^{(2)}$ are approximately the same. Therefore, we could employ the least squares estimators via solving the following score functions

$$\begin{cases} \sum_{t=2}^{T}\{Z_t^{(1)} - \mu^{(1)} - \phi^{(1)} Z_{t-1}^{(1)}\} = 0 \\ \sum_{t=2}^{T}\{Z_t^{(1)} - \mu^{(1)} - \phi^{(1)} Z_{t-1}^{(1)}\} Z_{t-1}^{(1)} = 0 \\ \sum_{t=2}^{T}\{Z_t^{(1)} - Z_t^{(2)} - \mu^{(2)} - \phi^{(2)}(Z_{t-1}^{(1)} - Z_{t-1}^{(2)})\} = 0 \\ \sum_{t=2}^{T}\{Z_t^{(1)} - Z_t^{(2)} - \mu^{(2)} - \phi^{(2)}(Z_{t-1}^{(1)} - Z_{t-1}^{(2)})\}(Z_{t-1}^{(1)} - Z_{t-1}^{(2)}) = 0 \end{cases} \tag{5.8}$$

and

$$\begin{cases} \sum_{t=1}^{T}\{\log m^{(i)}(x,t) - \alpha_x^{(i)} - \beta_x^{(i)} Z_t^{(i)}\} = 0 \\ \sum_{t=1}^{T}\{\log m^{(i)}(x,t) - \alpha_x^{(i)} - \beta_x^{(i)} Z_t^{(i)}\} Z_t^{(i)} = 0 \end{cases} \tag{5.9}$$

for $i = 1, 2$ and $x = 1, 2, \ldots, M$. However, when either $\{k_t^{(1)}\}$ or $\{k_t^{(1)} - k_t^{(2)}\}$ is stationary, the above least squares estimators become biased as indicated by the simulation study below. By noting that the inconsistency of the least squares estimators via solving (5.8) is due to the correlation between $Z_t^{(1)} - \mu_0^{(1)} - \phi_0^{(1)} Z_{t-1}^{(1)} = e_t^{(1)} + \eta_t^{(1)} - \phi_0^{(1)} \eta_{t-1}^{(1)}$ and $Z_{t-1}^{(1)} = k_{t-1}^{(1)} + \eta_{t-1}^{(1)}$, we follow the idea of shifting a lag in Part 3 and propose the unified bias corrected estimators via solving the following modified score equations:

$$\begin{cases} \sum_{t=3}^{T}\{Z_t^{(1)} - \mu^{(1)} - \phi^{(1)} Z_{t-1}^{(1)}\} = 0 \\ \sum_{t=3}^{T}\{Z_t^{(1)} - \mu^{(1)} - \phi^{(1)} Z_{t-1}^{(1)}\} Z_{t-2}^{(1)} = 0 \\ \sum_{t=3}^{T}\{Z_t^{(1)} - Z_t^{(2)} - \mu^{(2)} - \phi^{(2)}(Z_{t-1}^{(1)} - Z_{t-1}^{(2)})\} = 0 \\ \sum_{t=3}^{T}\{Z_t^{(1)} - Z_t^{(2)} - \mu^{(2)} - \phi^{(2)}(Z_{t-1}^{(1)} - Z_{t-1}^{(2)})\}(Z_{t-2}^{(1)} - Z_{t-2}^{(2)}) = 0 \end{cases} \tag{5.10}$$

and

$$\begin{cases} \sum_{t=2}^{T}\{\log m^{(i)}(x,t) - \alpha_x^{(i)} - \beta_x^{(i)} Z_t^{(i)}\} = 0 \\ \sum_{t=2}^{T}\{\log m^{(i)}(x,t) - \alpha_x^{(i)} - \beta_x^{(i)} Z_t^{(i)}\} Z_{t-1}^{(i)} = 0 \end{cases} \tag{5.11}$$

for $i = 1, 2$ and $x = 1, 2, \ldots, M-1$.

To present the asymptotic distribution of the proposed bias-corrected estimators, we need some notations. Put

$$\boldsymbol{\theta} = (\mu^{(1)}, \phi^{(1)}, \mu^{(2)}, \phi^{(2)}, \alpha_1^{(1)}, \beta_1^{(1)}, \alpha_1^{(2)}, \beta_1^{(2)}, \cdots, \alpha_{M-1}^{(1)}, \beta_{M-1}^{(1)}, \alpha_{M-1}^{(2)}, \beta_{M-1}^{(2)})^\tau,$$

and let

$$\hat{\boldsymbol{\theta}} = (\hat{\mu}^{(1)}, \hat{\phi}^{(1)}, \hat{\mu}^{(2)}, \hat{\phi}^{(2)}, \hat{\alpha}_1^{(1)}, \hat{\beta}_1^{(1)}, \hat{\alpha}_1^{(2)}, \hat{\beta}_1^{(2)}, \cdots, \hat{\alpha}_{M-1}^{(1)}, \hat{\beta}_{M-1}^{(1)}, \hat{\alpha}_{M-1}^{(2)}, \hat{\beta}_{M-1}^{(2)})^\tau$$

and

$$\boldsymbol{\theta}_0 = (\mu_0^{(1)}, \phi_0^{(1)}, \mu_0^{(2)}, \phi_0^{(2)}, \alpha_{1,0}^{(1)}, \beta_{1,0}^{(1)}, \alpha_{1,0}^{(2)}, \beta_{1,0}^{(2)}, \cdots, \alpha_{M-1,0}^{(1)}, \beta_{M-1,0}^{(1)}, \alpha_{M-1,0}^{(2)}, \beta_{M-1,0}^{(2)})^\tau$$

denote the above bias-corrected estimators and the true value of $\boldsymbol{\theta}$, respectively. Note that we exclude $\alpha_M^{(i)}$ and $\beta_M^{(i)}$ in the above definitions due to the constraints $\sum_{x=1}^{M} \alpha_{x,0}^{(i)} = 0$ and

$\sum_{x=1}^{M} \beta_{x,0}^{(i)} = 1$ for $i = 1, 2$. Further define

$$\begin{cases} Y_{t,1}^{(1)}(\boldsymbol{\theta}_0) = Z_t^{(1)} - \mu_0^{(1)} - \phi_0^{(1)} Z_{t-1}^{(1)}, \\ Y_{t,2}^{(1)}(\boldsymbol{\theta}_0) = (Z_t^{(1)} - \mu_0^{(1)} - \phi_0^{(1)} Z_{t-1}^{(1)}) Z_{t-2}^{(1)}, \\ Y_{t,1}^{(2)}(\boldsymbol{\theta}_0) = Z_t^{(1)} - Z_t^{(2)} - \mu_0^{(2)} - \phi_0^{(2)} (Z_{t-1}^{(1)} - Z_{t-1}^{(2)}), \\ Y_{t,2}^{(2)}(\boldsymbol{\theta}_0) = \{Z_t^{(1)} - Z_t^{(2)} - \mu_0^{(2)} - \phi_0^{(2)} (Z_{t-1}^{(1)} - Z_{t-1}^{(2)})\} (Z_{t-2}^{(1)} - Z_{t-2}^{(2)}), \\ \tilde{Y}_{t,x}^{(i)}(\boldsymbol{\theta}_0) = \log m^{(i)}(x,t) - \alpha_{x,0}^{(i)} - \beta_{x,0}^{(i)} Z_t^{(i)}, \\ \hat{Y}_{t,x}^{(i)}(\boldsymbol{\theta}_0) = \{\log m^{(i)}(x,t) - \alpha_{x,0}^{(i)} - \beta_{x,0}^{(i)} Z_t^{(i)}\} Z_{t-1}^{(i)} \end{cases}$$

for $i = 1, 2$ and $x = 1, 2, \ldots, M-1$. Now we present the asymptotic results in four cases according to the values of $\phi_0^{(1)}$ and $\phi_0^{(2)}$. These theorems show that the proposed bias-corrected inference unifies all cases.

Theorem 5.1. *Suppose models (5.4) and (5.5) hold with (5.7). Further assume* $\mu_0^{(1)} \neq 0$, $\mu_0^{(2)} \neq 0$, $\phi_0^{(1)} = 1 + \frac{\rho_1}{T}$ *and* $\phi_0^{(2)} = 1 + \frac{\rho_2}{T}$ *for some* $\rho_1, \rho_2 \in \mathbb{R}$. *Then*

$$D_T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}_1^{-1} \boldsymbol{\Sigma}_1 \boldsymbol{\Gamma}_1^{-1}) \text{ as } T \to \infty,$$

where $\boldsymbol{\Sigma}_1$ *and* $\boldsymbol{\Gamma}_1$ *are respectively defined by (5.15) and (5.17) in the appendix, and*

$$D_T = diag(\sqrt{T}, T^{3/2}, \cdots, \sqrt{T}, T^{3/2}).$$

Theorem 5.2. *Suppose models (5.4) and (5.5) hold with (5.7). Further assume* $\mu_0^{(1)} \neq 0$, $\mu_0^{(2)} \neq 0$, $|\phi_0^{(1)}| < 1$ *and* $|\phi_0^{(2)}| < 1$. *Then*

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}_2^{-1} \boldsymbol{\Sigma}_2 \boldsymbol{\Gamma}_2^{-1}) \text{ as } T \to \infty,$$

where $\boldsymbol{\Sigma}_2$ *and* $\boldsymbol{\Gamma}_2$ *are respectively defined by (5.21) and (5.22) in the appendix.*

Theorem 5.3. *Suppose models (5.4) and (5.5) hold with (5.7). Further assume* $\mu_0^{(1)} \neq 0$,

$\mu_0^{(2)} \neq 0$, $\phi_0^{(1)} = 1 + \frac{\rho_1}{T}$ for some $\rho_1 \in \mathbb{R}$ and $|\phi_0^{(2)}| < 1$. Then

$$D_T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}_3^{-1}\boldsymbol{\Sigma}_3\boldsymbol{\Gamma}_3^{-1}) \text{ as } T \to \infty,$$

where $\boldsymbol{\Sigma}_3$ and $\boldsymbol{\Gamma}_3$ are respectively defined by (5.24) and (5.25) in the appendix, and

$$D_T = diag(\sqrt{T}, T^{3/2}, \sqrt{T}, \sqrt{T}, \sqrt{T}, T^{3/2}, \cdots, \sqrt{T}, T^{3/2}).$$

Theorem 5.4. *Suppose models (5.4) and (5.5) hold with (5.7). Further assume $\mu_0^{(1)} \neq 0$, $\mu_0^{(2)} \neq 0$, $|\phi_0^{(1)}| < 1$ and $\phi_0^{(2)} = 1 + \frac{\rho_2}{T}$ for some $\rho_2 \in \mathbb{R}$. Then*

$$D_T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}_4^{-1}\boldsymbol{\Sigma}_4\boldsymbol{\Gamma}_4^{-1}) \text{ as } T \to \infty,$$

where $\boldsymbol{\Sigma}_4$ and $\boldsymbol{\Gamma}_4$ are respectively defined by (5.26) and (5.27) in the appendix, and

$$D_T = diag(\sqrt{T}, \sqrt{T}, \sqrt{T}, T^{3/2}, \cdots, \sqrt{T}, \sqrt{T}, \sqrt{T}, T^{3/2}).$$

5.2 Data Analysis and Simulation Study

Data Analysis We investigate the US mortality data of male and female cohorts from year 1933 to year 2017, which are available from the Human Mortality Database (https://www.mortality.org).

Firstly we fit the models (5.4) and (5.6) by the singular value decomposition method in Lee & Carter (1992) and then fit model (5.5) by the least squares estimate. Secondly we fit models (5.4), (5.5) and (5.7) by the proposed least squares estimate without bias correction, i.e., solving equations (5.8) and (5.9). Thirdly we use the proposed bias-corrected estimate $\hat{\boldsymbol{\theta}}$ in Theorems 5.1–5.4 to fit models (5.4), (5.5) and (5.7).

Table 5.1 reports these estimates. As (5.6) and (5.7) are different, one would expect different estimates for $\mu^{(i)}, \alpha_x^{(i)}$ by the Lee-Carter inference and the proposed least squares estimation with or without bias correction. On the other hand, one could expect similar estimates for $\phi^{(i)}, \beta_x^{(i)}$ based on these three inferences if they are consistent. Table 5.1 shows that the estimate for $\phi^{(1)}$ based on the Lee-Carter inference is much closer to one (i.e., unit root) than the proposed least squares estimation, and the estimates for $\mu^{(i)}$ and $\phi^{(i)}$ are different for the proposed least squares estimation with or without bias correction. As estimates for $\phi^{(2)}$ based on these three inferences are 'significantly' smaller than one, it suggests that $\{k_t^{(1)} - k_t^{(2)}\}$ is a stationary sequence. Hence it would be cautious to use the Lee-Carter inference and the proposed least squares estimation without bias correction.

Finally we examine the effect of these three inferences on forecasting future mortality rates by carrying out an out-of-sample forecast of 50 more years. Figure 5.1 plots the historic mortality rates and the forecasts of $\log m^{(i)}(x,t)$ from the Lee-Carter inference (dashed line), the least squares estimation without bias correction (dotted line), and the least squares estimation with bias correction (dash-dotted line). It is clear that three inferences provide quite different forecasts. To confirm the inconsistency of the first two inferences, a simulation study is conducted below.

Figure (5.1) Forecasts of $\log m^{(i)}(x,t)$ along with historic values of US mortality data
Solid lines represent true historic values of mortality data; dashed lines represent forecasts by Lee-Carter model; dotted lines represent forecasts according to the proposed least squares estimators without bias correction, see (5.8) and (5.9); dash-dotted lines are forecasts by proposed bias corrected estimators, see (5.10) and (5.11)

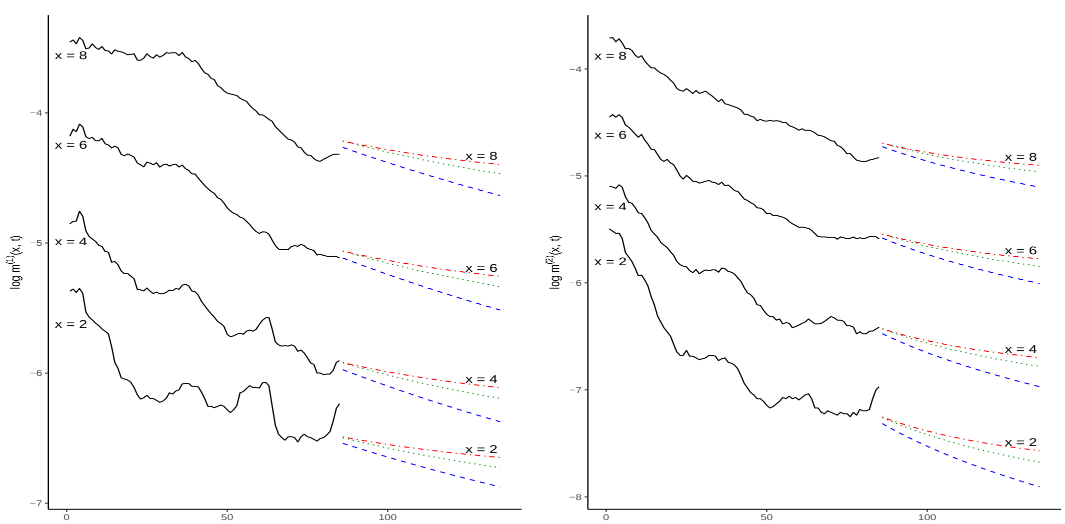


Table (5.1) Estimates of Bivariate Model Parameters Based on U.S. Cohorts

Lee-Carter Estimates of $\alpha_x^{(i)}$ and $\beta_x^{(i)}$

x	1	2	3	4	5	6	7	8	9	10
$\alpha_x^{(1)}$	-6.265	-6.128	-5.850	-5.482	-5.060	-4.624	-4.215	-3.810	-3.429	-3.028
$\beta_x^{(1)}$	0.085	0.091	0.104	0.109	0.108	0.109	0.104	0.101	0.098	0.093
$\alpha_x^{(2)}$	-7.018	-6.741	-6.385	-5.994	-5.582	-5.169	-4.780	-4.359	-3.941	-3.477
$\beta_x^{(2)}$	0.134	0.126	0.119	0.106	0.096	0.091	0.083	0.081	0.082	0.084

Lee-Carter Estimates of $\mu^{(i)}$ and $\phi^{(i)}$

$\mu^{(1)}$	$\phi^{(1)}$	$\mu^{(2)}$	$\phi^{(2)}$
-0.113	0.994	0.049	0.958

Estimates of $\alpha_x^{(i)}$ and $\beta_x^{(i)}$ by the Proposed Model (Without Bias Correction)

x	1	2	3	4	5	6	7	8	9	10
$\alpha_x^{(1)}$	-2.187	-1.755	-0.877	-0.276	0.129	0.569	0.741	0.994	1.245	1.415
$\beta_x^{(1)}$	0.085	0.091	0.104	0.109	0.108	0.108	0.103	0.100	0.098	0.093
$\alpha_x^{(2)}$	0.094	-0.030	-0.067	-0.344	-0.459	-0.317	-0.318	-0.020	0.444	1.017
$\beta_x^{(2)}$	0.133	0.126	0.118	0.106	0.096	0.091	0.083	0.081	0.082	0.084

Estimates of $\mu^{(i)}$ and $\phi^{(i)}$ by the Proposed Model (Without Bias Correction)

$\mu^{(1)}$	$\phi^{(1)}$	$\mu^{(2)}$	$\phi^{(2)}$
-0.622	0.989	0.274	0.958

Estimates of $\alpha_x^{(i)}$ and $\beta_x^{(i)}$ by the Proposed Model (With Bias Correction)

x	1	2	3	4	5	6	7	8	9	10
$\alpha_x^{(1)}$	-2.294	-1.867	-0.954	-0.289	0.139	0.610	0.809	1.053	1.324	1.469
$\beta_x^{(1)}$	0.083	0.089	0.102	0.108	0.109	0.109	0.105	0.101	0.099	0.094
$\alpha_x^{(2)}$	-0.012	-0.105	-0.106	-0.330	-0.447	-0.285	-0.278	7.739e-3	0.488	1.068
$\beta_x^{(2)}$	0.131	0.124	0.118	0.106	0.096	0.091	0.084	0.082	0.083	0.085

Estimates of $\mu^{(i)}$ and $\phi^{(i)}$ by the Proposed Model (With Bias Correction)

$\mu^{(1)}$	$\phi^{(1)}$	$\mu^{(2)}$	$\phi^{(2)}$
-0.856	0.985	0.285	0.956

Simulation Study As many papers in the study of longevity risk simply assume $\phi^{(1)} = 1$, this section uses simulated data to show that the Lee-Carter inference and the proposed least squares estimation without bias correction lead to an inconsistent inference when $\phi^{(1)} = 1$ and $|\phi^{(2)}| < 1$, but the proposed bias-corrected inference performs well.

We simulate 10,000 random samples with sample size $T = 80$, 150 and 300 from models (5.4), (5.5) and (5.7) with parameters being the proposed bias-corrected estimates in fitting the real dataset above except $\phi^{(1)} = 1$ and $\phi^{(2)} = 0.95$. We choose $\varepsilon_{x,t}$ and e_t to be independent and identically distributed random variables with $N(0, 0.1^2)$. We report the mean and standard deviation of estimates based on these three methods in Tables 5.2, 5.3, and 5.4. As argued before, we could only expect consistency for estimating $\phi^{(i)}$ and $\beta_x^{(i)}$ based on the Lee-Carter inference as constraints in (5.6) are different from those in (5.7). Results in Tables 5.2 to 5.4 show that i) the proposed least squares estimates without bias correction for $\mu^{(2)}$ and $\phi^{(2)}$ are inconsistent; ii) the Lee-Carter inference clearly gives an inconsistent estimation for $\phi^{(2)}$ when $T = 300$; iii) the standard errors for estimating $\alpha_x^{(i)}$ based on the Lee-Carter inference are much larger than those based on the proposed least squares estimation with or without bias correction; and iv) the bias-corrected inference performs quite well.

5.3 Conclusions

When longevity risk transfer involves more than one population, multipopulation mortality model is needed. Recently some two-population Lee-Carter related mortality models have been applied to the study of longevity risk and the employed statistical inference relies on the singular value decomposition method as the two-step inference procedure in Lee & Carter (1992). Surprisingly it seems no research on the study of inference consistence. Given the recent theoretical development for the statistical inference of Lee-Carter mortality model, this paper alerts the application of the Lee-Carter inference, confirms its inconsistence by a simulation study, and proposes a bias-corrected inference which is always consistent regardless of the mortality indexes being stationary or unit root, and performs well empirically.

Table (5.2) Estimates with standard deviations in brackets for $T = 80$

Lee-Carter Estimates of $\alpha_x^{(i)}$ and $\beta_x^{(i)}$

x	1	2	3	4	5	6	7	8	9	10
$\alpha_x^{(1)}$	-5.171	-4.955	-4.500	-4.049	-3.625	-3.178	-2.827	-2.466	-2.115	-1.786
	(0.044)	(0.047)	(0.053)	(0.057)	(0.057)	(0.057)	(0.055)	(0.053)	(0.052)	(0.049)
$\beta_x^{(1)}$	0.083	0.089	0.102	0.108	0.109	0.109	0.105	0.102	0.099	0.094
	(5.472e-4)	(5.351e-4)	(5.367e-4)	(5.353e-4)	(5.388e-4)	(5.274e-4)	(5.418e-4)	(5.379e-4)	(5.384e-4)	(5.337e-4)
$\alpha_x^{(2)}$	-5.131	-4.953	-4.693	-4.466	-4.198	-3.852	-3.565	-3.181	-2.746	-2.251
	(0.072)	(0.068)	(0.065)	(0.059)	(0.054)	(0.051)	(0.047)	(0.046)	(0.047)	(0.048)
$\beta_x^{(2)}$	0.131	0.124	0.118	0.106	0.096	0.091	0.084	0.082	0.083	0.085
	(5.060e-4)	(5.097e-4)	(5.089e-4)	(5.047e-4)	(5.037e-4)	(5.055e-4)	(5.030e-4)	(5.022e-4)	(5.068e-4)	(5.082e-4)

Lee-Carter Estimates of $\mu^{(i)}$ and $\phi^{(i)}$

$\mu^{(1)}$	$\phi^{(1)}$	$\mu^{(2)}$	$\phi^{(2)}$
-0.847	1.001	0.070	0.941
(0.012)	(6.724e-4)	(7.744e-3)	(0.013)

Estimates of $\alpha_x^{(i)}$ and $\beta_x^{(i)}$ by the Proposed Model (Without Bias Correction)

x	1	2	3	4	5	6	7	8	9	10
$\alpha_x^{(1)}$	-2.294	-1.867	-0.954	-0.289	0.138	0.610	0.808	1.053	1.324	1.470
	(0.022)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.022)	(0.022)	(0.021)
$\beta_x^{(1)}$	0.083	0.089	0.102	0.108	0.109	0.109	0.105	0.102	0.099	0.094
	(5.470e-4)	(5.350e-4)	(5.366e-4)	(5.352e-4)	(5.386e-4)	(5.272e-4)	(5.417e-4)	(5.378e-4)	(5.382e-4)	(5.335e-4)
$\alpha_x^{(2)}$	-0.012	-0.106	-0.106	-0.330	-0.447	-0.286	-0.278	8.205e-3	0.489	1.068
	(0.022)	(0.023)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.023)	(0.023)
$\beta_x^{(2)}$	0.131	0.124	0.118	0.106	0.096	0.091	0.084	0.082	0.083	0.085
	(5.059e-4)	(5.096e-4)	(5.088e-4)	(5.046e-4)	(5.036e-4)	(5.054e-4)	(5.029e-4)	(5.021e-4)	(5.067e-4)	(5.081e-4)

Estimates of $\mu^{(i)}$ and $\phi^{(i)}$ by the Proposed Model (Without Bias Correction)

$\mu^{(1)}$	$\phi^{(1)}$	$\mu^{(2)}$	$\phi^{(2)}$
-0.866	1.000	0.669	0.861
(0.029)	(7.541e-4)	(0.119)	(0.028)

Estimates of $\alpha_x^{(i)}$ and $\beta_x^{(i)}$ by the Proposed Model (With Bias Correction)

x	1	2	3	4	5	6	7	8	9	10
$\alpha_x^{(1)}$	-2.294	-1.867	-0.954	-0.289	0.138	0.610	0.808	1.053	1.324	1.470
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
$\beta_x^{(1)}$	0.083	0.089	0.102	0.108	0.109	0.109	0.105	0.102	0.099	0.094
	(5.567e-4)	(5.438e-4)	(5.462e-4)	(5.440e-4)	(5.484e-4)	(5.373e-4)	(5.520e-4)	(5.472e-4)	(5.479e-4)	(5.433e-4)
$\alpha_x^{(2)}$	-0.012	-0.106	-0.106	-0.330	-0.447	-0.286	-0.278	8.087e-3	0.488	1.068
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
$\beta_x^{(2)}$	0.131	0.124	0.118	0.106	0.096	0.091	0.084	0.082	0.083	0.085
	(5.167e-4)	(5.215e-4)	(5.194e-4)	(5.152e-4)	(5.140e-4)	(5.164e-4)	(5.131e-4)	(5.132e-4)	(5.176e-4)	(5.183e-4)

Estimates of $\mu^{(i)}$ and $\phi^{(i)}$ by the Proposed Model (With Bias Correction)

$\mu^{(1)}$	$\phi^{(1)}$	$\mu^{(2)}$	$\phi^{(2)}$
-0.857	1.000	0.293	0.948
(0.030)	(7.743e-4)	(0.106)	(0.024)

Table (5.3) Estimates with standard deviations in brackets for $T = 150$

Lee-Carter Estimates of $\alpha_x^{(i)}$ and $\beta_x^{(i)}$

x	1	2	3	4	5	6	7	8	9	10
$\alpha_x^{(1)}$	-7.658	-7.623	-7.565	-7.298	-6.878	-6.452	-5.969	-5.507	-5.088	-4.599
	(0.060)	(0.064)	(0.073)	(0.078)	(0.078)	(0.078)	(0.075)	(0.073)	(0.071)	(0.067)
$\beta_x^{(1)}$	0.083	0.089	0.102	0.108	0.109	0.109	0.105	0.102	0.099	0.094
	(2.094e-4)	(2.067e-4)	(2.082e-4)	(2.127e-4)	(2.082e-4)	(2.073e-4)	(2.084e-4)	(2.086e-4)	(2.076e-4)	(2.108e-4)
$\alpha_x^{(2)}$	-9.140	-8.751	-8.286	-7.706	-7.135	-6.645	-6.140	-5.679	-5.279	-4.850
	(0.096)	(0.091)	(0.086)	(0.078)	(0.071)	(0.067)	(0.062)	(0.060)	(0.061)	(0.063)
$\beta_x^{(2)}$	0.131	0.124	0.118	0.106	0.096	0.091	0.084	0.082	0.083	0.085
	(2.047e-4)	(2.043e-4)	(2.037e-4)	(2.035e-4)	(2.034e-4)	(2.028e-4)	(2.031e-4)	(2.065e-4)	(2.062e-4)	(2.033e-4)

Lee-Carter Estimates of $\mu^{(i)}$ and $\phi^{(i)}$

$\mu^{(1)}$	$\phi^{(1)}$	$\mu^{(2)}$	$\phi^{(2)}$
-0.851	1.000	0.041	0.937
(8.505e-3)	(2.434e-4)	(4.727e-3)	(0.011)

Estimates of $\alpha_x^{(i)}$ and $\beta_x^{(i)}$ by the Proposed Model (Without Bias Correction)

x	1	2	3	4	5	6	7	8	9	10
$\alpha_x^{(1)}$	-2.293	-1.867	-0.954	-0.289	0.138	0.610	0.809	1.053	1.324	1.470
	(0.016)	(0.015)	(0.016)	(0.016)	(0.016)	(0.016)	(0.015)	(0.015)	(0.015)	(0.016)
$\beta_x^{(1)}$	0.083	0.089	0.102	0.108	0.109	0.109	0.105	0.102	0.099	0.094
	(2.094e-4)	(2.066e-4)	(2.082e-4)	(2.127e-4)	(2.082e-4)	(2.073e-4)	(2.084e-4)	(2.086e-4)	(2.076e-4)	(2.108e-4)
$\alpha_x^{(2)}$	-0.012	-0.106	-0.106	-0.330	-0.447	-0.285	-0.278	7.924e-3	0.488	1.068
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
$\beta_x^{(2)}$	0.131	0.124	0.118	0.106	0.096	0.091	0.084	0.082	0.083	0.085
	(2.047e-4)	(2.043e-4)	(2.037e-4)	(2.035e-4)	(2.034e-4)	(2.028e-4)	(2.031e-4)	(2.065e-4)	(2.062e-4)	(2.033e-4)

Estimates of $\mu^{(i)}$ and $\phi^{(i)}$ by the Proposed Model (Without Bias Correction)

$\mu^{(1)}$	$\phi^{(1)}$	$\mu^{(2)}$	$\phi^{(2)}$
-0.862	1.000	0.836	0.839
(0.019)	(2.620e-4)	(0.127)	(0.026)

Estimates of $\alpha_x^{(i)}$ and $\beta_x^{(i)}$ by the Proposed Model (With Bias Correction)

x	1	2	3	4	5	6	7	8	9	10
$\alpha_x^{(1)}$	-2.294	-1.867	-0.954	-0.289	0.138	0.610	0.809	1.053	1.324	1.470
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
$\beta_x^{(1)}$	0.083	0.089	0.102	0.108	0.109	0.109	0.105	0.102	0.099	0.094
	(2.111e-4)	(2.095e-4)	(2.097e-4)	(2.148e-4)	(2.108e-4)	(2.093e-4)	(2.106e-4)	(2.114e-4)	(2.101e-4)	(2.124e-4)
$\alpha_x^{(2)}$	-0.012	-0.106	-0.106	-0.330	-0.447	-0.285	-0.278	7.866e-3	0.488	1.068
	(0.017)	(0.016)	(0.017)	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)	(0.017)	(0.016)
$\beta_x^{(2)}$	0.131	0.124	0.118	0.106	0.096	0.091	0.084	0.082	0.083	0.085
	(2.070e-4)	(2.066e-4)	(2.058e-4)	(2.063e-4)	(2.051e-4)	(2.056e-4)	(2.049e-4)	(2.086e-4)	(2.082e-4)	(2.051e-4)

Estimates of $\mu^{(i)}$ and $\phi^{(i)}$ by the Proposed Model (With Bias Correction)

$\mu^{(1)}$	$\phi^{(1)}$	$\mu^{(2)}$	$\phi^{(2)}$
-0.857	1.000	0.291	0.949
(0.019)	(2.667e-4)	(0.102)	(0.020)

Table (5.4) Estimates with standard deviations in brackets for $T = 300$

Lee-Carter Estimates of $\alpha_x^{(i)}$ and $\beta_x^{(i)}$

x	1	2	3	4	5	6	7	8	9	10
$\alpha_x^{(1)}$	-12.988 (0.083)	-13.342 (0.089)	-14.133 (0.102)	-14.262 (0.108)	-13.849 (0.108)	-13.468 (0.109)	-12.703 (0.105)	-12.026 (0.101)	-11.459 (0.099)	-10.629 (0.094)
$\beta_x^{(1)}$	0.083 (7.528e-5)	0.089 (7.372e-5)	0.102 (7.384e-5)	0.108 (7.320e-5)	0.109 (7.410e-5)	0.109 (7.276e-5)	0.105 (7.411e-5)	0.101 (7.325e-5)	0.099 (7.350e-5)	0.094 (7.409e-5)
$\alpha_x^{(2)}$	-17.609 (0.131)	-16.771 (0.124)	-15.874 (0.118)	-14.549 (0.106)	-13.340 (0.096)	-12.544 (0.092)	-11.578 (0.085)	-10.954 (0.082)	-10.630 (0.083)	-10.341 (0.085)
$\beta_x^{(2)}$	0.131 (7.339e-5)	0.124 (7.384e-5)	0.118 (7.349e-5)	0.106 (7.421e-5)	0.096 (7.418e-5)	0.091 (7.316e-5)	0.084 (7.378e-5)	0.082 (7.385e-5)	0.083 (7.402e-5)	0.085 (7.273e-5)

Lee-Carter Estimates of $\mu^{(i)}$ and $\phi^{(i)}$

$\mu^{(1)}$	$\phi^{(1)}$	$\mu^{(2)}$	$\phi^{(2)}$
-0.854 (5.882e-3)	1.000 (8.175e-5)	0.022 (2.812e-3)	0.927 (9.941e-3)

Estimates of $\alpha_x^{(i)}$ and $\beta_x^{(i)}$ by the Proposed Model (Without Bias Correction)

x	1	2	3	4	5	6	7	8	9	10
$\alpha_x^{(1)}$	-2.294 (0.011)	-1.867 (0.011)	-0.954 (0.011)	-0.289 (0.011)	0.138 (0.011)	0.610 (0.011)	0.809 (0.011)	1.053 (0.011)	1.324 (0.011)	1.470 (0.011)
$\beta_x^{(1)}$	0.083 (7.528e-5)	0.089 (7.372e-5)	0.102 (7.384e-5)	0.108 (7.320e-5)	0.109 (7.410e-5)	0.109 (7.275e-5)	0.105 (7.411e-5)	0.101 (7.325e-5)	0.099 (7.349e-5)	0.094 (7.409e-5)
$\alpha_x^{(2)}$	-0.012 (0.011)	-0.106 (0.011)	-0.106 (0.011)	-0.330 (0.011)	-0.447 (0.011)	-0.285 (0.011)	-0.278 (0.011)	7.702e-3 (0.011)	0.489 (0.011)	1.068 (0.011)
$\beta_x^{(2)}$	0.131 (7.338e-5)	0.124 (7.384e-5)	0.118 (7.349e-5)	0.106 (7.421e-5)	0.096 (7.418e-5)	0.091 (7.316e-5)	0.084 (7.377e-5)	0.082 (7.385e-5)	0.083 (7.402e-5)	0.085 (7.273e-5)

Estimates of $\mu^{(i)}$ and $\phi^{(i)}$ by the Proposed Model (Without Bias Correction)

$\mu^{(1)}$	$\phi^{(1)}$	$\mu^{(2)}$	$\phi^{(2)}$
-0.859 (0.012)	1.000 (8.489e-5)	1.175 (0.149)	0.783 (0.028)

Estimates of $\alpha_x^{(i)}$ and $\beta_x^{(i)}$ by the Proposed Model (With Bias Correction)

x	1	2	3	4	5	6	7	8	9	10
$\alpha_x^{(1)}$	-2.294 (0.011)	-1.867 (0.011)	-0.954 (0.011)	-0.289 (0.011)	0.138 (0.011)	0.610 (0.011)	0.809 (0.011)	1.053 (0.011)	1.324 (0.011)	1.470 (0.011)
$\beta_x^{(1)}$	0.083 (7.576e-5)	0.089 (7.411e-5)	0.102 (7.433e-5)	0.108 (7.367e-5)	0.109 (7.460e-5)	0.109 (7.318e-5)	0.105 (7.447e-5)	0.101 (7.366e-5)	0.099 (7.383e-5)	0.094 (7.436e-5)
$\alpha_x^{(2)}$	-0.012 (0.011)	-0.106 (0.011)	-0.106 (0.012)	-0.330 (0.011)	-0.447 (0.011)	-0.285 (0.011)	-0.278 (0.011)	7.649e-3 (0.011)	0.489 (0.011)	1.068 (0.011)
$\beta_x^{(2)}$	0.131 (7.376e-5)	0.124 (7.424e-5)	0.118 (7.400e-5)	0.106 (7.448e-5)	0.096 (7.461e-5)	0.091 (7.360e-5)	0.084 (7.410e-5)	0.082 (7.425e-5)	0.083 (7.449e-5)	0.085 (7.307e-5)

Estimates of $\mu^{(i)}$ and $\phi^{(i)}$ by the Proposed Model (With Bias Correction)

$\mu^{(1)}$	$\phi^{(1)}$	$\mu^{(2)}$	$\phi^{(2)}$
-0.856 (0.012)	1.000 (8.549e-5)	0.292 (0.116)	0.949 (0.022)

Appendix: Proofs of Theorems

We first introduce some lemmas before we prove the main results.

Lemma 5.1. *Suppose the conditions of Theorem 5.1 hold, then*

$$\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{W}_{1t}(\boldsymbol{\theta}_0)\xrightarrow{d}N(\mathbf{0},\boldsymbol{\Sigma}_1)\quad and\quad \frac{1}{T}\sum_{t=3}^{T}\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{W}_{1t}(\boldsymbol{\theta}_0)\xrightarrow{p}\boldsymbol{\Gamma}_1,\quad as\ T\to\infty,$$

where $\boldsymbol{\Sigma}_1$ *and* $\boldsymbol{\Gamma}_1$ *are defined in (5.15) and (5.17) respectively, and*

$$\boldsymbol{W}_{1t}(\boldsymbol{\theta}_0)=(Y_{t,1}^{(1)},\frac{Y_{t,2}^{(1)}}{T},Y_{t,1}^{(2)},\frac{Y_{t,2}^{(2)}}{T},\tilde{Y}_{t,1}^{(1)},\frac{\hat{Y}_{t,1}^{(1)}}{T},\tilde{Y}_{t,1}^{(2)},\frac{\hat{Y}_{t,1}^{(2)}}{T},\ldots,\tilde{Y}_{t,M-1}^{(1)},\frac{\hat{Y}_{t,M-1}^{(1)}}{T},\tilde{Y}_{t,M-1}^{(2)},\frac{\hat{Y}_{t,M-1}^{(2)}}{T})^{\tau}.$$

Proof. Since $\mu_0^{(1)}\neq 0$ and $\mu_0^{(2)}\neq 0$, we have

$$k_t^{(1)}=\mu_0^{(1)}\left(\sum_{j=0}^{t-1}(\phi_0^{(1)})^j\right)+(\phi_0^{(1)})^t k_0^{(1)}+\sum_{i=1}^{t}(\phi_0^{(1)})^{t-i}e_i^{(1)}.$$

It follows from Chan and Wei (1987) that for $s\in[0,1]$

$$\sum_{i=1}^{[Ts]}(\phi_0^{(1)})^{[Ts]-i}e_i^{(1)}=(\phi_0^{(1)})^{[Ts]-T}\sum_{i=1}^{[Ts]}(\phi_0^{(1)})^{T-i}e_i^{(1)}=O_p(\sqrt{T}),$$

which imply that, as $T\to\infty$

$$\frac{k_{[Ts]}^{(1)}}{T}\xrightarrow{p}f_{\rho_1,\mu_0^{(1)}}(s)\text{ for }s\in[0,1],\tag{5.12}$$

where

$$f_{\rho,\mu}(s)=\begin{cases}\mu\frac{e^{s\rho}-1}{\rho} & \text{if}\quad \rho\neq 0,\\ \mu s & \text{if}\quad \rho=0.\end{cases}\tag{5.13}$$

Similarly we have as $T\to\infty$

$$\frac{k_{[Ts]}^{(1)}-k_{[Ts]}^{(2)}}{T}\xrightarrow{p}f_{\rho_2,\mu_0^{(2)}}(s)\text{ for }s\in[0,1].\tag{5.14}$$

Note that

$$\begin{cases} Y_{t,1}^{(1)}(\boldsymbol{\theta}_0) = e_t^{(1)} + \eta_t^{(1)} - \phi_0^{(1)}\eta_{t-1}^{(1)}, \\ Y_{t,2}^{(1)}(\boldsymbol{\theta}_0) = (e_t^{(1)} + \eta_t^{(1)} - \phi_0^{(1)}\eta_{t-1}^{(1)})(k_{t-2}^{(1)} + \eta_{t-2}^{(1)}), \\ Y_{t,1}^{(2)}(\boldsymbol{\theta}_0) = e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)} - \phi_0^{(2)}(\eta_{t-1}^{(1)} - \eta_{t-1}^{(2)}), \\ Y_{t,2}^{(2)}(\boldsymbol{\theta}_0) = \{e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)} - \phi_0^{(2)}(\eta_{t-1}^{(1)} - \eta_{t-1}^{(2)})\}(k_{t-2}^{(1)} - k_{t-2}^{(2)} + \eta_{t-2}^{(1)} - \eta_{t-2}^{(2)}), \\ \tilde{Y}_{t,x}^{(i)}(\boldsymbol{\theta}_0) = \epsilon_{x,t}^{(i)} - \beta_{x,0}^{(i)}\eta_t^{(i)},\ i = 1, 2;\ x = 1, 2, \ldots, M-1, \\ \hat{Y}_{t,x}^{(i)}(\boldsymbol{\theta}_0) = (\epsilon_{x,t}^{(i)} - \beta_{x,0}^{(i)}\eta_t^{(i)})(k_{t-1}^{(i)} + \eta_{t-1}^{(i)}),\ i = 1, 2;\ x = 1, 2, \ldots, M-1. \end{cases}$$

Let $X_{t,1}^{(1)} = e_t^{(1)}$, $X_{t,2}^{(1)} = (e_t^{(1)} + \eta_t^{(1)})\frac{k_{t-2}^{(1)}}{T} - \eta_t^{(1)}\frac{k_{t-1}^{(1)}}{T}$, $X_{t,1}^{(2)} = e_t^{(2)}$,

$$X_{t,2}^{(2)} = (e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)})\frac{k_{t-2}^{(1)} - k_{t-2}^{(2)}}{T} - (\eta_t^{(1)} - \eta_t^{(2)})\frac{k_{t-1}^{(1)} - k_{t-1}^{(2)}}{T},$$

and define

$$\boldsymbol{U}_t(\boldsymbol{\theta}_0) = (X_{t,1}^{(1)}, X_{t,2}^{(1)}, X_{t,1}^{(2)}, X_{t,2}^{(2)}, \tilde{Y}_{t,1}^{(1)}, \frac{\hat{Y}_{t,1}^{(1)}}{T}, \tilde{Y}_{t,1}^{(2)}, \frac{\hat{Y}_{t,1}^{(2)}}{T}, \ldots, \tilde{Y}_{t,M-1}^{(1)}, \frac{\hat{Y}_{t,M-1}^{(1)}}{T}, \tilde{Y}_{t,M-1}^{(2)}, \frac{\hat{Y}_{t,M-1}^{(2)}}{T})^\tau.$$

Then we have

$$\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{W}_{1t}(\boldsymbol{\theta_0}) = \frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{U}_t(\boldsymbol{\theta_0}) + o_p(1), \quad T \to \infty.$$

Let $\mathcal{F}_t$ denote the σ-field generated by $\{(e_s^{(1)}, e_s^{(2)}, \varepsilon_{1,s}^{(1)}, \cdots, \varepsilon_{M,s}^{(1)}, \varepsilon_{1,s}^{(2)}, \cdots, \varepsilon_{M,s}^{(2)})^\tau : s \le t\}$, then $\{\boldsymbol{U}_t(\boldsymbol{\theta}_0), \mathcal{F}_t\}_{t=1}^{\infty}$ is a sequence of martingale difference. By the central limit theorem of martingale sequences in Hall & Heyde (2014), $\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{U}_t(\boldsymbol{\theta}_0)$ converges in distribution to a multivariate normal distribution with 0 means and covariance matrix $\boldsymbol{\Sigma}_1$ satisfying

$$\frac{1}{T}\sum_{t=3}^{T}E\{\boldsymbol{U}_t(\boldsymbol{\theta}_0)\boldsymbol{U}_t(\boldsymbol{\theta}_0)^\tau|\mathcal{F}_{t-1}\} \xrightarrow{p} \boldsymbol{\Sigma}_1 \stackrel{\text{def}}{=} (\sigma_{ij})_{4M\times 4M}, \tag{5.15}$$

i.e.,

$$\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{W}_{1t}(\boldsymbol{\theta_0}) = \frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{U}_t(\boldsymbol{\theta_0}) + o_p(1) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma}_1), \quad T \to \infty.$$

By the assumption that $\{(e_t^{(1)}, e_t^{(2)}, \varepsilon_{1,t}^{(1)}, \cdots, \varepsilon_{M,t}^{(1)}, \varepsilon_{1,t}^{(2)}, \cdots, \varepsilon_{M,t}^{(2)})^\tau\}_{t=1}^T$ is a sequence of independent random vectors with finite variance, it is straightforward to verify that

$$\sigma_{11} = E(e_t^{(1)})^2,\ \sigma_{12} = E(e_t^{(1)})^2 \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\, ds,\ \sigma_{13} = E(e_t^{(1)} e_t^{(2)}),$$

$$\sigma_{14} = E(e_t^{(1)} e_t^{(2)}) \int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\, ds,\ \sigma_{22} = E(e_t^{(1)})^2 \int_0^1 f^2_{\rho_1,\mu_0^{(1)}}(s)\, ds,$$

$$\sigma_{23} = E(e_t^{(1)} e_t^{(2)}) \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\, ds,\ \sigma_{24} = E(e_t^{(1)} e_t^{(2)}) \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s) f_{\rho_2,\mu_0^{(2)}}(s)\, ds,$$

$$\sigma_{33} = E(e_t^{(2)})^2,\ \sigma_{34} = E(e_t^{(2)})^2 \int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\, ds,\ \sigma_{44} = E(e_t^{(2)})^2 \int_0^1 f^2_{\rho_2,\mu_0^{(2)}}(s)\, ds.$$

Note that

$$\frac{k_{[Ts]}^{(2)}}{T} = \frac{k_{[Ts]}^{(1)}}{T} - \frac{k_{[Ts]}^{(1)} - k_{[Ts]}^{(2)}}{T} \quad \text{for } s \in [0,1], \tag{5.16}$$

then as $T \to \infty$, we have

$$\frac{1}{T} \sum_{t=2}^{T} \frac{k_t^{(2)}}{T} \xrightarrow{p} \int_0^1 \{f_{\rho_1,\mu_0^{(1)}}(s) - f_{\rho_2,\mu_0^{(2)}}(s)\}\, ds$$

and

$$\frac{1}{T} \sum_{t=2}^{T} (\frac{k_t^{(2)}}{T})^2 \xrightarrow{p} \int_0^1 \{f_{\rho_1,\mu_0^{(1)}}(s) - f_{\rho_2,\mu_0^{(2)}}(s)\}^2\, ds.$$

Similar to the proof above, for $x, y = 1, 2, \ldots, M-1$

$$\sigma_{1,4x+1} = E\{e_t^{(1)}(\epsilon_{x,t}^{(1)} - \beta_{x,0}^{(1)} \eta_t^{(1)})\}, \quad \sigma_{1,4x+2} = \sigma_{1,4x+1} \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\, ds,$$

$$\sigma_{1,4x+3} = E\{e_t^{(1)}(\epsilon_{x,t}^{(2)} - \beta_{x,0}^{(2)} \eta_t^{(2)})\},\ \sigma_{1,4x+4} = \sigma_{1,4x+3} \int_0^1 \{f_{\rho_1,\mu_0^{(1)}}(s) - f_{\rho_2,\mu_0^{(2)}}(s)\}\, ds,$$

$$\sigma_{2,4x+1} = \sigma_{1,4x+2},\ \sigma_{2,4x+2} = \sigma_{1,4x+1} \int_0^1 f^2_{\rho_1,\mu_0^{(1)}}(s)\, ds,\ \sigma_{2,4x+3} = \sigma_{1,4x+3} \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\, ds,$$

$$\sigma_{2,4x+4} = \sigma_{1,4x+3} \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s) \{f_{\rho_1,\mu_0^{(1)}}(s) - f_{\rho_2,\mu_0^{(2)}}(s)\}\, ds,\ \sigma_{3,4x+1} = E\{e_t^{(2)}(\epsilon_{x,t}^{(1)} - \beta_{x,0}^{(1)} \eta_t^{(1)})\},$$

$$\sigma_{3,4x+2} = \sigma_{3,4x+1} \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds,\ \sigma_{3,4x+3} = E\{e_t^{(2)}(\epsilon_{x,t}^{(2)} - \beta_{x,0}^{(2)}\eta_t^{(2)})\},$$

$$\sigma_{3,4x+4} = \sigma_{3,4x+3} \int_0^1 \{f_{\rho_1,\mu_0^{(1)}}(s) - f_{\rho_2,\mu_0^{(2)}}(s)\}\,ds,\ \sigma_{4,4x+1} = \sigma_{3,4x+1} \int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds,$$

$$\sigma_{4,4x+2} = \sigma_{3,4x+1} \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s) f_{\rho_2,\mu_0^{(2)}}(s)\,ds,\ \sigma_{4,4x+3} = \sigma_{3,4x+3} \int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds,$$

$$\sigma_{4,4x+4} = \sigma_{3,4x+3} \int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\{f_{\rho_1,\mu_0^{(1)}}(s) - f_{\rho_2,\mu_0^{(2)}}(s)\}\,ds,$$

$$\sigma_{4x+1,4y+1} = E\{(\epsilon_{x,t}^{(1)} - \beta_{x,0}^{(1)}\eta_t^{(1)})(\epsilon_{y,t}^{(1)} - \beta_{y,0}^{(1)}\eta_t^{(1)})\},\ \sigma_{4x+1,4y+3} = E\{(\epsilon_{x,t}^{(1)} - \beta_{x,0}^{(1)}\eta_t^{(1)})(\epsilon_{y,t}^{(2)} - \beta_{y,0}^{(2)}\eta_t^{(2)})\},$$

$$\sigma_{4x+1,4y+2} = \sigma_{4x+1,4y+1} \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds,\ \sigma_{4x+1,4y+4} = \sigma_{4x+1,4y+3} \int_0^1 \{f_{\rho_1,\mu_0^{(1)}}(s) - f_{\rho_2,\mu_0^{(2)}}(s)\}\,ds,$$

$$\sigma_{4x+2,4y+2} = \sigma_{4x+1,4y+1} \int_0^1 f^2_{\rho_1,\mu_0^{(1)}}(s)\,ds,\ \sigma_{4x+2,4y+3} = \sigma_{4x+1,4y+3} \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds,$$

$$\sigma_{4x+2,4y+4} = \sigma_{4x+1,4y+3} \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\{f_{\rho_1,\mu_0^{(1)}}(s) - f_{\rho_2,\mu_0^{(2)}}(s)\}\,ds,$$

$$\sigma_{4x+3,4y+3} = E\{(\epsilon_{x,t}^{(2)} - \beta_{x,0}^{(2)}\eta_t^{(2)})(\epsilon_{y,t}^{(2)} - \beta_{y,0}^{(2)}\eta_t^{(2)})\},$$

$$\sigma_{4x+3,4y+4} = \sigma_{4x+3,4y+3} \int_0^1 \{f_{\rho_1,\mu_0^{(1)}}(s) - f_{\rho_2,\mu_0^{(2)}}(s)\}\,ds,$$

$$\sigma_{4x+4,4y+4} = \sigma_{4x+3,4y+3} \int_0^1 \{f_{\rho_1,\mu_0^{(1)}}(s) - f_{\rho_2,\mu_0^{(2)}}(s)\}^2\,ds.$$

Moreover, note that

$$\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{W}_{1t}(\boldsymbol{\theta}_0) = \mathbf{diag}\left(A_{t0}^{(1)}, A_{t0}^{(2)}, A_{t1}^{(1)}, A_{t1}^{(2)}, \cdots, A_{t(M-1)}^{(1)}, A_{t(M-1)}^{(2)}\right)_{4(M-1)\times 4(M-1)},$$

where for $i = 1, 2$ and $x = 1, 2, \ldots, M-1$,

$$A_{t0}^{(1)} = -\begin{pmatrix} 1 & Z_{t-1}^{(1)} \\ \frac{Z_{t-2}^{(1)}}{T} & \frac{Z_{t-1}^{(1)} Z_{t-2}^{(1)}}{T} \end{pmatrix},\ A_{tx}^{(i)} = -\begin{pmatrix} 1 & Z_t^{(i)} \\ \frac{Z_{t-1}^{(i)}}{T} & \frac{Z_t^{(i)} Z_{t-1}^{(i)}}{T} \end{pmatrix},$$

and

$$A_{t0}^{(2)} = -\begin{pmatrix} 1 & Z_{t-1}^{(1)} - Z_{t-1}^{(2)} \\ \frac{1}{T}(Z_{t-2}^{(1)} - Z_{t-2}^{(2)}) & \frac{1}{T}(Z_{t-1}^{(1)} - Z_{t-1}^{(2)})(Z_{t-2}^{(1)} - Z_{t-2}^{(2)}) \end{pmatrix}.$$

It follows from (5.12), (5.14) and (5.16) that as $T \to \infty$

$$\frac{1}{T}\sum_{t=3}^{T} \frac{Z_{t-1}^{(1)}}{T} = \frac{1}{T}\sum_{t=3}^{T} \frac{k_{t-1}^{(1)}}{T} + \frac{1}{T^2}\sum_{t=3}^{T} \eta_{t-1}^{(1)} \xrightarrow{p} \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds,$$

$$\frac{1}{T}\sum_{t=3}^{T} \frac{Z_{t-1}^{(2)}}{T} = \frac{1}{T}\sum_{t=3}^{T} \frac{k_{t-1}^{(2)}}{T} + \frac{1}{T^2}\sum_{t=3}^{T} \eta_{t-1}^{(2)} \xrightarrow{p} \int_0^1 \{f_{\rho_1,\mu_0^{(1)}}(s) - f_{\rho_2,\mu_0^{(2)}}(s)\}\,ds,$$

$$\frac{1}{T^3}\sum_{t=3}^{T} Z_{t-1}^{(1)} Z_{t-2}^{(1)} = \frac{1}{T}\sum_{t=3}^{T} \frac{k_{t-1}^{(1)}}{T}\frac{k_{t-2}^{(1)}}{T} + o_p(1) \xrightarrow{p} \int_0^1 f^2_{\rho_1,\mu_0^{(1)}}(s)\,ds,$$

$$\frac{1}{T^3}\sum_{t=3}^{T} Z_{t-1}^{(2)} Z_{t-2}^{(2)} \xrightarrow{p} \int_0^1 \{f_{\rho_1,\mu_0^{(1)}}(s) - f_{\rho_2,\mu_0^{(2)}}(s)\}^2\,ds,\ \frac{1}{T^2}\sum_{t=3}^{T}(Z_{t-2}^{(1)} - Z_{t-2}^{(2)}) \xrightarrow{p} \int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds,$$

$$\frac{1}{T^3}\sum_{t=3}^{T}(Z_{t-2}^{(1)} - Z_{t-2}^{(2)})(Z_{t-1}^{(1)} - Z_{t-1}^{(2)}) \xrightarrow{p} \int_0^1 f^2_{\rho_2,\mu_0^{(2)}}(s)\,ds.$$

This implies that

$$\frac{1}{T}\sum_{t=3}^{T}\{\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{W}_{1t}(\boldsymbol{\theta}_0)\}\{\sqrt{T}D_T^{-1}\} \xrightarrow{p} \boldsymbol{\Gamma}_1 \text{ as } T \to \infty,$$

where

$$\boldsymbol{\Gamma}_1 = \mathbf{diag}\left(A_0^{(1)}, A_0^{(2)}, A_1^{(1)}, A_1^{(2)}, \cdots, A_{M-1}^{(1)}, A_{M-1}^{(2)}\right)_{4(M-1)\times 4(M-1)} \tag{5.17}$$

with

$$A_0^{(1)} = \cdots = A_{M-1}^{(1)} = -\begin{pmatrix} 1 & \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds \\ \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds & \int_0^1 f^2_{\rho_1,\mu_0^{(1)}}(s)\,ds \end{pmatrix},$$

$$A_0^{(2)} = -\begin{pmatrix} 1 & \int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds \\ \int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds & \int_0^1 f^2_{\rho_2,\mu_0^{(2)}}(s)\,ds \end{pmatrix}$$

and

$$A_1^{(2)} = \cdots = A_{M-1}^{(2)} = -\begin{pmatrix} 1 & \int_0^1 \{f_{\rho_1,\mu_0^{(1)}}(s) - f_{\rho_2,\mu_0^{(2)}}(s)\}\, ds \\ \int_0^1 \{f_{\rho_1,\mu_0^{(1)}}(s) - f_{\rho_2,\mu_0^{(2)}}(s)\}\, ds & \int_0^1 \{f_{\rho_1,\mu_0^{(1)}}(s) - f_{\rho_2,\mu_0^{(2)}}(s)\}^2\, ds \end{pmatrix}.$$

$\square$

Lemma 5.2. *Suppose the conditions of Theorem 5.2 hold, then*

$$\frac{1}{\sqrt{T}}\sum_{t=3}^{T} \boldsymbol{W}_{2t}(\boldsymbol{\theta}_0) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma}_2) \quad \text{and} \quad \frac{1}{T}\sum_{t=3}^{T} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{W}_{2t}(\boldsymbol{\theta}_0) \xrightarrow{p} \boldsymbol{\Gamma}_2, \quad \text{as } T \to \infty,$$

where $\boldsymbol{\Sigma}_2$ *and* $\boldsymbol{\Gamma}_2$ *are defined in (5.21) and (5.22) respectively, and*

$$\boldsymbol{W}_{2t}(\boldsymbol{\theta}_0) = (Y_{t,1}^{(1)}, Y_{t,2}^{(1)}, Y_{t,1}^{(2)}, Y_{t,2}^{(2)}, \tilde{Y}_{t,1}^{(1)}, \hat{Y}_{t,1}^{(1)}, \tilde{Y}_{t,1}^{(2)}, \hat{Y}_{t,1}^{(2)}, \cdots, \tilde{Y}_{t,M-1}^{(1)}, \hat{Y}_{t,M-1}^{(1)}, \tilde{Y}_{t,M-1}^{(2)}, \hat{Y}_{t,M-1}^{(2)})^\tau.$$

Proof. Since $\mu_0^{(1)} \neq 0$ and $\mu_0^{(2)} \neq 0$, we have

$$E(k_t^{(1)}) = \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}, \quad E(k_t^{(1)})^2 = (\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2 + \frac{E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2}. \tag{5.18}$$

and

$$E(k_t^{(1)} - k_t^{(2)}) = \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}}, \quad E(k_t^{(1)} - k_t^{(2)})^2 = (\frac{\mu_0^{(2)}}{1-\phi_0^{(2)}})^2 + \frac{E(e_t^{(2)})^2}{1-(\phi_0^{(2)})^2}. \tag{5.19}$$

Moreover, the stationarity of sequence $\{(k_t^{(1)}, k_t^{(1)} - k_t^{(2)})^\tau, t = 1, 2, \ldots, T\}$ implies that

$$E\{k_t^{(1)}(k_t^{(1)} - k_t^{(2)})\} = \frac{\mu_0^{(1)}\mu_0^{(2)}}{(1-\phi_0^{(1)})(1-\phi_0^{(2)})} + \frac{E(e_t^{(1)}e_t^{(2)})}{1-\phi_0^{(1)}\phi_0^{(2)}}. \tag{5.20}$$

Note that

$$\begin{cases} Y_{t,1}^{(1)}(\boldsymbol{\theta}_0) = e_t^{(1)} + \eta_t^{(1)} - \phi_0^{(1)}\eta_{t-1}^{(1)}, \\ Y_{t,2}^{(1)}(\boldsymbol{\theta}_0) = (e_t^{(1)} + \eta_t^{(1)} - \phi_0^{(1)}\eta_{t-1}^{(1)})(k_{t-2}^{(1)} + \eta_{t-2}^{(1)}), \\ Y_{t,1}^{(2)}(\boldsymbol{\theta}_0) = e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)} - \phi_0^{(2)}(\eta_{t-1}^{(1)} - \eta_{t-1}^{(2)}), \\ Y_{t,2}^{(2)}(\boldsymbol{\theta}_0) = \{e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)} - \phi_0^{(2)}(\eta_{t-1}^{(1)} - \eta_{t-1}^{(2)})\}(k_{t-2}^{(1)} - k_{t-2}^{(2)} + \eta_{t-2}^{(1)} - \eta_{t-2}^{(2)}), \\ \tilde{Y}_{t,x}^{(i)}(\boldsymbol{\theta}_0) = \epsilon_{x,t}^{(i)} - \beta_{x,0}^{(i)}\eta_t^{(i)}, \ i = 1, 2; \ x = 1, 2, \ldots, M-1, \\ \hat{Y}_{t,x}^{(i)}(\boldsymbol{\theta}_0) = (\epsilon_{x,t}^{(i)} - \beta_{x,0}^{(i)}\eta_t^{(i)})(k_{t-1}^{(i)} + \eta_{t-1}^{(i)}), \ i = 1, 2; \ x = 1, 2, \ldots, M-1. \end{cases}$$

Let $X_{t,1}^{(1)} = e_t^{(1)} + (1 - \phi_0^{(1)})\eta_t^{(1)}$, $X_{t,1}^{(2)} = e_t^{(2)} + (1 - \phi_0^{(2)})(\eta_t^{(1)} - \eta_t^{(2)})$,

$$X_{t,2}^{(1)} = (e_t^{(1)} + \eta_t^{(1)})(k_{t-2}^{(1)} + \eta_{t-2}^{(1)}) - \phi_0^{(1)}\eta_t^{(1)}(k_{t-1}^{(1)} + \eta_{t-1}^{(1)}),$$

$$X_{t,2}^{(2)} = (e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)})(k_{t-2}^{(1)} - k_{t-2}^{(2)} + \eta_{t-2}^{(1)} - \eta_{t-2}^{(2)}) - \phi_0^{(2)}(\eta_t^{(1)} - \eta_t^{(2)})(k_{t-1}^{(1)} - k_{t-1}^{(2)} + \eta_{t-1}^{(1)} - \eta_{t-1}^{(2)}),$$

and define

$$\boldsymbol{U}_t(\boldsymbol{\theta}_0) = (X_{t,1}^{(1)}, X_{t,2}^{(1)}, X_{t,1}^{(2)}, X_{t,2}^{(2)}, \tilde{Y}_{t,1}^{(1)}, \hat{Y}_{t,1}^{(1)}, \tilde{Y}_{t,1}^{(2)}, \hat{Y}_{t,1}^{(2)}, \cdots, \tilde{Y}_{t,M-1}^{(1)}, \hat{Y}_{t,M-1}^{(1)}, \tilde{Y}_{t,M-1}^{(2)}, \hat{Y}_{t,M-1}^{(2)})^\tau.$$

Then we have

$$\frac{1}{\sqrt{T}} \sum_{t=3}^{T} \boldsymbol{W}_{2t}(\boldsymbol{\theta_0}) = \frac{1}{\sqrt{T}} \sum_{t=3}^{T} \boldsymbol{U}_t(\boldsymbol{\theta_0}) + o_p(1), \quad T \to \infty.$$

Let $\mathcal{F}_t$ denote the σ-field generated by $\{(e_s^{(1)}, e_s^{(2)}, \varepsilon_{1,s}^{(1)}, \cdots, \varepsilon_{M,s}^{(1)}, \varepsilon_{1,s}^{(2)}, \cdots, \varepsilon_{M,s}^{(2)})^\tau : s \le t\}$, then $\{\boldsymbol{U}_t(\boldsymbol{\theta}_0), \mathcal{F}_t\}_{t=1}^{\infty}$ is a sequence of martingale difference. By the central limit theorem of martingale sequences in Hall & Heyde (2014), $\frac{1}{\sqrt{T}} \sum_{t=3}^{T} \boldsymbol{U}_t(\boldsymbol{\theta}_0)$ converges in distribution to a multivariate normal distribution with 0 means and covariance matrix $\boldsymbol{\Sigma}_2$ satisfying

$$\frac{1}{T} \sum_{t=3}^{T} E\{\boldsymbol{U}_t(\boldsymbol{\theta}_0)\boldsymbol{U}_t(\boldsymbol{\theta}_0)^\tau | \mathcal{F}_{t-1}\} \xrightarrow{p} \boldsymbol{\Sigma}_2 \stackrel{\text{def}}{=} (\sigma_{ij})_{4M \times 4M}, \tag{5.21}$$

i.e.,

$$\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{W}_{2t}(\boldsymbol{\theta_0}) = \frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{U}_t(\boldsymbol{\theta_0}) + o_p(1) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_2), \quad T \to \infty.$$

By the assumption that $\{(e_t^{(1)}, e_t^{(2)}, \varepsilon_{1,t}^{(1)}, \cdots, \varepsilon_{M,t}^{(1)}, \varepsilon_{1,t}^{(2)}, \cdots, \varepsilon_{M,t}^{(2)})^\tau\}_{t=1}^T$ is a sequence of independent random vectors with finite variance, it is straightforward to verify that

$$\sigma_{11} = E\{e_t^{(1)} + (1-\phi_0^{(1)})\eta_t^{(1)}\}^2, \quad \sigma_{12} = \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\sigma_{11},$$

$$\sigma_{13} = E\{[e_t^{(1)} + (1-\phi_0^{(1)})\eta_t^{(1)}][e_t^{(2)} + (1-\phi_0^{(2)})(\eta_t^{(1)} - \eta_t^{(2)})]\},\ \sigma_{14} = \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}}\sigma_{13},$$

$$\begin{aligned}\sigma_{22} &= \{E(e_t^{(1)} + \eta_t^{(1)})^2 + (\phi_0^{(1)})^2 E(\eta_t^{(1)})^2\}\{(\tfrac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2 + \tfrac{E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2} + 2E(e_t^{(1)}\eta_t^{(1)}) + E(\eta_t^{(1)})^2\}\\ &\quad -2\phi_0^{(1)} E\{\eta_t^{(1)}(e_t^{(1)} + \eta_t^{(1)})\}\{(\tfrac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2 + \tfrac{\phi_0^{(1)}E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2} + \phi_0^{(1)}E(e_t^{(1)}\eta_t^{(1)})\},\end{aligned}$$

$$\sigma_{23} = \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}}\sigma_{13},\ \sigma_{33} = E\{e_t^{(2)} + (1+\phi_0^{(2)})(\eta_t^{(1)} - \eta_t^{(2)}\}^2,\ \sigma_{34} = \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}}\sigma_{33},$$

$$\begin{aligned}\sigma_{24} &= E\{(e_t^{(1)} + \eta_t^{(1)})(e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)})\}\{\tfrac{\mu_0^{(1)}\mu_0^{(2)}}{(1-\phi_0^{(1)})(1-\phi_0^{(2)})} + \tfrac{E(e_t^{(1)}e_t^{(2)})}{1-\phi_0^{(1)}\phi_0^{(2)}} + E[(e_t^{(1)} + \eta_t^{(1)})(\eta_t^{(1)} - \eta_t^{(2)})]\\ &\quad +E(e_t^{(2)}\eta_t^{(1)})\} - \phi_0^{(2)}E\{(e_t^{(1)} + \eta_t^{(1)})(\eta_t^{(1)} - \eta_t^{(2)})\}\{\tfrac{\mu_0^{(1)}\mu_0^{(1)}}{(1-\phi_0^{(1)})(1-\phi_0^{(2)})} + \tfrac{\phi_0^{(2)}E(e_t^{(1)}e_t^{(2)})}{1-\phi_0^{(1)}\phi_0^{(2)}} + \phi_0^{(2)}E(e_t^{(2)}\eta_t^{(1)})\}\\ &\quad -\phi_0^{(1)}E\{\eta_t^{(1)}(e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)})\}\{\tfrac{\mu_0^{(1)}\mu_0^{(1)}}{(1-\phi_0^{(1)})(1-\phi_0^{(2)})} + \tfrac{E(e_t^{(1)}e_t^{(2)})}{1-\phi_0^{(1)}\phi_0^{(2)}} + \phi_0^{(1)}E[e_t^{(1)}(\eta_t^{(1)} - \eta_t^{(2)})]\}\\ &\quad +\phi_0^{(1)}\phi_0^{(2)}E\{\eta_t^{(1)}(\eta_t^{(1)} - \eta_t^{(2)})\}\{\tfrac{\mu_0^{(1)}\mu_0^{(2)}}{(1-\phi_0^{(1)})(1-\phi_0^{(2)})} + \tfrac{E(e_t^{(1)}e_t^{(2)})}{1-\phi_0^{(1)}\phi_0^{(2)}} + E[(e_t^{(1)} + \eta_t^{(1)})(\eta_t^{(1)} - \eta_t^{(2)})] + E(e_t^{(2)}\eta_t^{(1)})\},\end{aligned}$$

$$\begin{aligned}\sigma_{44} &= \{E(e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)})^2 + (\phi_0^{(2)})^2E(\eta_t^{(1)} - \eta_t^{(2)})^2\}\{(\tfrac{\mu_0^{(2)}}{1-\phi_0^{(2)}})^2 + \tfrac{E(e_t^{(2)})^2}{1-(\phi_0^{(2)})^2} + 2E[e_t^{(2)}(\eta_t^{(1)} - \eta_t^{(2)})]\\ &\quad +E(\eta_t^{(1)} - \eta_t^{(2)})^2\} - 2\phi_0^{(2)}E\{(e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)})(\eta_t^{(1)} - \eta_t^{(2)})\}\{(\tfrac{\mu_0^{(2)}}{1-\phi_0^{(2)}})^2 + \tfrac{\phi_0^{(2)}E(e_t^{(2)})^2}{1-(\phi_0^{(2)})^2}\\ &\quad +\phi_0^{(2)}E[e_t^{(2)}(\eta_t^{(1)} - \eta_t^{(2)})]\}.\end{aligned}$$

Note that

$$E(k_t^{(2)}) = E(k_t^{(1)}) - E(k_t^{(1)} - k_t^{(2)}) = \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}} - \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}}.$$

Then for $x = 1, 2, \ldots, M-1$, we have

$$\sigma_{1,4x+1} = E\{[e_t^{(1)}+(1-\phi_0^{(1)})\eta_t^{(1)}](\epsilon_{x,t}^{(1)}-\beta_{x,0}^{(1)}\eta_t^{(1)})\},\ \sigma_{1,4x+3} = E\{[e_t^{(1)}+(1-\phi_0^{(1)})\eta_t^{(1)}](\epsilon_{x,t}^{(2)}-\beta_{x,0}^{(2)}\eta_t^{(2)})\},$$

$$\sigma_{1,4x+2} = \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\sigma_{1,4x+1}, \quad \sigma_{1,4x+4} = \{\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}} - \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}}\}\sigma_{1,4x+3},$$

$$\sigma_{2,4x+1} = \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\sigma_{1,4x+1}, \quad \sigma_{2,4x+3} = \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\sigma_{1,4x+3},$$

$$\begin{aligned}\sigma_{2,4x+2} \quad &= E\{(e_t^{(1)}+\eta_t^{(1)})(\epsilon_{x,t}^{(1)}-\beta_{x,0}^{(1)}\eta_t^{(1)})\}\{(\tfrac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2 + \tfrac{\phi_0^{(1)}E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2} + \phi_0^{(1)}E(e_t^{(1)}\eta_t^{(1)})\}\\ &-\phi_0^{(1)}E\{\eta_t^{(1)}(\epsilon_{x,t}^{(1)}-\beta_{x,0}^{(1)}\eta_t^{(1)})\}\{(\tfrac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2 + \tfrac{E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2} + 2E(e_t^{(1)}\eta_t^{(1)}) + E(\eta_t^{(1)})^2\},\end{aligned}$$

$$\begin{aligned}\sigma_{2,4x+4} \quad &= E\{(e_t^{(1)}+\eta_t^{(1)})(\epsilon_{x,t}^{(2)}-\beta_{x,0}^{(2)}\eta_t^{(2)})\}\{(\tfrac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2 - \tfrac{\mu_0^{(1)}\mu_0^{(2)}}{(1-\phi_0^{(1)})(1-\phi_0^{(2)})} + \tfrac{\phi_0^{(1)}E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2}\\ &-\tfrac{\phi_0^{(2)}E(e_t^{(1)}e_t^{(2)})}{1-\phi_0^{(1)}\phi_0^{(2)}} + \phi_0^{(1)}E(e_t^{(1)}\eta_t^{(1)}) - \phi_0^{(1)}E(e_t^{(2)}\eta_t^{(1)})\}\\ &-\phi_0^{(1)}E\{\eta_t^{(1)}(\epsilon_{x,t}^{(2)}-\beta_{x,0}^{(2)}\eta_t^{(2)})\}\{(\tfrac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2 + \tfrac{E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2} - \tfrac{\mu_0^{(1)}\mu_0^{(2)}}{(1-\phi_0^{(1)})(1-\phi_0^{(2)})}\\ &-\tfrac{E(e_t^{(1)}e_t^{(2)})}{1-\phi_0^{(1)}\phi_0^{(2)}} + E(e_t^{(1)}\eta_t^{(2)}) + E[\eta_t^{(1)}(e_t^{(1)}-e_t^{(2)})] + E(\eta_t^{(1)}\eta_t^{(2)})\},\end{aligned}$$

$$\sigma_{3,4x+1} = E\{[e_t^{(2)}+(1-\phi_0^{(2)})(\eta_t^{(1)}-\eta_t^{(2)})](\epsilon_{x,t}^{(1)}-\beta_{x,0}^{(1)}\eta_t^{(1)})\}, \quad \sigma_{3,4x+2} = \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\sigma_{3,4x+1},$$

$$\sigma_{3,4x+3} = E\{[e_t^{(2)}+(1-\phi_0^{(2)})(\eta_t^{(1)}-\eta_t^{(2)})](\epsilon_{x,t}^{(2)}-\beta_{x,0}^{(2)}\eta_t^{(2)})\},\ \sigma_{3,4x+4} = \{\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}-\frac{\mu_0^{(2)}}{1-\phi_0^{(2)}}\}\sigma_{4,4x+3},$$

$$\sigma_{4,4x+1} = \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}}\sigma_{3,4x+1}, \quad \sigma_{4,4x+3} = \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}}\sigma_{3,4x+3},$$

$$\begin{aligned}\sigma_{4,4x+2} \quad &= E\{(e_t^{(2)}+\eta_t^{(1)}-\eta_t^{(2)})(\epsilon_{x,t}^{(1)}-\beta_{x,0}^{(1)}\eta_t^{(1)})\}\{\tfrac{\mu_0^{(1)}\mu_0^{(2)}}{(1-\phi_0^{(1)})(1-\phi_0^{(2)})} + \tfrac{\phi_0^{(1)}E(e_t^{(1)}e_t^{(2)})}{1-\phi_0^{(1)}\phi_0^{(2)}} + \phi_0^{(1)}E[e_t^{(1)}(\eta_t^{(1)}-\eta_t^{(2)})]\}\\ &-\phi_0^{(2)}E\{(\eta_t^{(1)}-\eta_t^{(2)})(\epsilon_{x,t}^{(1)}-\beta_{x,0}^{(1)}\eta_t^{(1)})\}\{\tfrac{\mu_0^{(1)}\mu_0^{(2)}}{(1-\phi_0^{(1)})(1-\phi_0^{(2)})} + \tfrac{E(e_t^{(1)}e_t^{(2)})}{1-\phi_0^{(1)}\phi_0^{(2)}} + E[e_t^{(1)}(\eta_t^{(1)}-\eta_t^{(2)})]\\ &+E(e_t^{(1)}\eta_t^{(1)}) + E[\eta_t^{(1)}(\eta_t^{(1)}-\eta_t^{(2)})]\}.\end{aligned}$$

$$\begin{aligned}\sigma_{4,4x+4} &= E\{(e_t^{(2)}+\eta_t^{(1)}-\eta_t^{(2)})(\epsilon_{x,t}^{(2)}-\beta_{x,0}^{(2)}\eta_t^{(2)})\}\{\tfrac{\mu_0^{(1)}\mu_0^{(2)}}{(1-\phi_0^{(1)})(1-\phi_0^{(2)})}+\tfrac{\phi_0^{(1)}E(e_t^{(1)}e_t^{(2)})}{1-\phi_0^{(1)}\phi_0^{(2)}}-(\tfrac{\mu_0^{(2)}}{1-\phi_0^{(2)}})^2\\ &+\tfrac{\phi_0^{(2)}E(e_t^{(2)})^2}{1-(\phi_0^{(2)})^2}+E[(\phi_0^{(1)}e_t^{(1)}-\phi_0^{(2)}e_t^{(2)})(\eta_t^{(1)}-\eta_t^{(2)})]\}\\ &-\phi_0^{(2)}E\{(\eta_t^{(1)}-\eta_t^{(2)})(\epsilon_{x,t}^{(2)}-\beta_{x,0}^{(2)}\eta_t^{(1)})\}\{\tfrac{\mu_0^{(1)}\mu_0^{(2)}}{(1-\phi_0^{(1)})(1-\phi_0^{(2)})}+\tfrac{E(e_t^{(1)}e_t^{(2)})}{1-\phi_0^{(1)}\phi_0^{(2)}}-(\tfrac{\mu_0^{(2)}}{1-\phi_0^{(2)}})^2\\ &-\tfrac{E(e_t^{(2)})^2}{1-(\phi_0^{(2)})^2}+E(e_t^{(2)}\eta_t^{(2)})+E[(e_t^{(1)}-e_t^{(2)})(\eta_t^{(1)}-\eta_t^{(2)})]+E[\eta_t^{(2)}(\eta_t^{(1)}-\eta_t^{(2)})]\}.\end{aligned}$$

Similarly, for for $x, y = 1, 2, \ldots, M-1$, we can show that

$$\sigma_{4x+1,4y+1} = E\{(\epsilon_{x,t}^{(1)}-\beta_{x,0}^{(1)}\eta_t^{(1)})(\epsilon_{y,t}^{(1)}-\beta_{y,0}^{(1)}\eta_t^{(1)})\},\ \sigma_{4x+1,4y+3} = E\{(\epsilon_{x,t}^{(1)}-\beta_{x,0}^{(1)}\eta_t^{(1)})(\epsilon_{y,t}^{(2)}-\beta_{y,0}^{(2)}\eta_t^{(2)})\},$$

$$\sigma_{4x+1,4y+2} = \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\sigma_{4x+1,4y+1},\ \sigma_{4x+1,4y+4} = (\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}-\frac{\mu_0^{(2)}}{1-\phi_0^{(2)}})\sigma_{4x+1,4y+3},$$

$$\sigma_{4x+2,4y+2} = \sigma_{4x+1,4y+1}\{(\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2+\frac{E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2}+E(\eta_t^{(1)})^2+2E(e_t^{(1)}\eta_t^{(1)})\},$$

$$\sigma_{4x+2,4y+3} = \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\sigma_{4x+1,4y+3},\ \sigma_{4x+3,4y+3} = E\{(\epsilon_{x,t}^{(2)}-\beta_{x,0}^{(2)}\eta_t^{(2)})(\epsilon_{y,t}^{(2)}-\beta_{y,0}^{(2)}\eta_t^{(2)})\},$$

$$\begin{aligned}\sigma_{4x+2,4y+4} &= \sigma_{4x+1,4y+3}\{(\tfrac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2+\tfrac{E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2}-\tfrac{\mu_0^{(1)}\mu_0^{(2)}}{(1-\phi_0^{(1)})(1-\phi_0^{(2)})}-\tfrac{E(e_t^{(1)}e_t^{(2)})}{1-\phi_0^{(1)}\phi_0^{(2)}}\\ &+E(e_t^{(1)}\eta_t^{(2)})+E(e_t^{(1)}\eta_t^{(1)})-E(e_t^{(2)}\eta_t^{(1)})+E(\eta_t^{(1)}\eta_t^{(2)})\},\end{aligned}$$

$$\sigma_{4x+3,4y+4} = \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}}\sigma_{4x+3,4y+3},$$

$$\begin{aligned}\sigma_{4x+4,4y+4} &= \sigma_{4x+3,4y+3}\{(\tfrac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2+\tfrac{E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2}+(\tfrac{\mu_0^{(2)}}{1-\phi_0^{(2)}})^2+\tfrac{E(e_t^{(2)})^2}{1-(\phi_0^{(2)})^2}-\tfrac{2\mu_0^{(1)}\mu_0^{(2)}}{(1-\phi_0^{(1)})(1-\phi_0^{(2)})}\\ &-\tfrac{2E(e_t^{(1)}e_t^{(2)})}{1-\phi_0^{(1)}\phi_0^{(2)}}+2E(e_t^{(1)}\eta_t^{(2)})-2E(e_t^{(2)}\eta_t^{(2)})+E(\eta_t^{(2)})^2\}.\end{aligned}$$

Furthermore, note that

$$\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{W}_{2t}(\boldsymbol{\theta}_0) = \mathbf{diag}\left(A_{t0}^{(1)}, A_{t0}^{(2)}, A_{t1}^{(1)}, A_{t1}^{(2)}, \cdots, A_{t(M-1)}^{(1)}, A_{t(M-1)}^{(2)}\right)_{4(M-1)\times 4(M-1)},$$

where for $i = 1, 2$ and $x = 1, 2, \ldots, M-1$,

$$A_{t0}^{(1)} = -\begin{pmatrix} 1 & Z_{t-1}^{(1)} \\ Z_{t-2}^{(1)} & Z_{t-1}^{(1)}Z_{t-2}^{(1)} \end{pmatrix},\ A_{tx}^{(i)} = -\begin{pmatrix} 1 & Z_t^{(i)} \\ Z_{t-1}^{(i)} & Z_t^{(i)}Z_{t-1}^{(i)} \end{pmatrix},$$

and

$$A_{t0}^{(2)} = -\begin{pmatrix} 1 & Z_{t-1}^{(1)} - Z_{t-1}^{(2)} \\ Z_{t-2}^{(1)} - Z_{t-2}^{(2)} & (Z_{t-1}^{(1)} - Z_{t-1}^{(2)})(Z_{t-2}^{(1)} - Z_{t-2}^{(2)}) \end{pmatrix}.$$

It follows from (5.18), (5.19) and (5.20) that as $T \to \infty$

$$\frac{1}{T}\sum_{t=3}^{T} Z_t^{(1)} = \frac{1}{T}\sum_{t=3}^{T} k_t^{(1)} + \frac{1}{T}\sum_{t=3}^{T} \eta_t^{(1)} \xrightarrow{p} \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}},$$

$$\frac{1}{T}\sum_{t=3}^{T} Z_t^{(2)} = \frac{1}{T}\sum_{t=3}^{T} k_t^{(2)} + \frac{1}{T}\sum_{t=3}^{T} \eta_t^{(2)} \xrightarrow{p} \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}} - \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}},$$

$$\begin{aligned} \tfrac{1}{T}\textstyle\sum_{t=3}^{T} Z_t^{(1)} Z_{t-1}^{(1)} &= \tfrac{1}{T}\textstyle\sum_{t=3}^{T} k_t^{(1)} k_{t-1}^{(1)} + \tfrac{1}{T}\textstyle\sum_{t=3}^{T} k_t^{(1)} \eta_{t-1}^{(1)} + o_p(1) \\ &\xrightarrow{p} (\tfrac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2 + \tfrac{\phi_0^{(1)} E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2} + \phi_0^{(1)} E(e_t^{(1)} \eta_t^{(1)}), \end{aligned}$$

$$\begin{aligned} \tfrac{1}{T}\textstyle\sum_{t=3}^{T} Z_t^{(2)} Z_{t-1}^{(2)} &= \tfrac{1}{T}\textstyle\sum_{t=3}^{T} k_t^{(2)} k_{t-1}^{(2)} + \tfrac{1}{T}\textstyle\sum_{t=3}^{T} k_t^{(2)} \eta_{t-1}^{(2)} + o_p(1) \\ &\xrightarrow{p} E\{k_t^{(1)} k_{t-1}^{(1)}\} - E\{k_t^{(1)}(k_{t-1}^{(1)} - k_{t-1}^{(2)})\} - E\{k_{t-1}^{(1)}(k_t^{(1)} - k_t^{(2)})\} \\ &\quad + E\{(k_t^{(1)} - k_t^{(2)})(k_{t-1}^{(1)} - k_{t-1}^{(2)})\} + E\{k_t^{(2)} \eta_{t-1}^{(2)}\} \\ &= (\tfrac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2 + \tfrac{\phi_0^{(1)} E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2} - \tfrac{2\mu_0^{(1)}\mu_0^{(2)}}{(1-\phi_0^{(1)})(1-\phi_0^{(2)})} - \tfrac{(\phi_0^{(1)}+\phi_0^{(2)})E(e_t^{(1)} e_t^{(2)})}{1-\phi_0^{(1)}\phi_0^{(2)}} \\ &\quad + (\tfrac{\mu_0^{(2)}}{1-\phi_0^{(2)}})^2 + \tfrac{\phi_0^{(2)} E(e_t^{(2)})^2}{1-(\phi_0^{(2)})^2} + \phi_0^{(1)} E(e_t^{(1)} \eta_t^{(2)}) - \phi_0^{(2)} E(e_t^{(2)} \eta_t^{(2)}), \end{aligned}$$

$$\frac{1}{T}\sum_{t=3}^{T} (Z_{t-1}^{(1)} - Z_{t-1}^{(2)}) = \frac{1}{T}\sum_{t=3}^{T} (k_{t-1}^{(1)} - k_{t-1}^{(2)}) + o_p(1) \xrightarrow{p} \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}},$$

$$\frac{1}{T}\sum_{t=3}^{T} (Z_{t-1}^{(1)} - Z_{t-1}^{(2)})(Z_{t-2}^{(1)} - Z_{t-2}^{(2)}) \xrightarrow{p} (\frac{\mu_0^{(2)}}{1-\phi_0^{(2)}})^2 + \frac{\phi_0^{(2)} E(e_t^{(2)})^2}{1-(\phi_0^{(2)})^2} + \phi_0^{(2)} E\{e_t^{(2)}(\eta_t^{(1)} - \eta_t^{(2)})\}.$$

This implies that

$$\frac{1}{T}\sum_{t=3}^{T} \{\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{W}_{2t}(\boldsymbol{\theta}_0)\} \xrightarrow{p} \boldsymbol{\Gamma}_2 \text{ as } T \to \infty,$$

where

$$\boldsymbol{\Gamma}_2 = \mathbf{diag}\left(A_0^{(1)}, A_0^{(2)}, A_1^{(1)}, A_1^{(2)}, \cdots, A_{M-1}^{(1)}, A_{M-1}^{(2)}\right)_{4(M-1)\times 4(M-1)} \tag{5.22}$$

with

$$A_0^{(1)} = \cdots = A_{M-1}^{(1)} = -\begin{pmatrix} 1 & \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}} \\ \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}} & (\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2 + \frac{\phi_0^{(1)}E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2} + \phi_0^{(1)}E(e_t^{(1)}\eta_t^{(1)}) \end{pmatrix},$$

$$A_0^{(2)} = -\begin{pmatrix} 1 & (\frac{\mu_0^{(2)}}{1-\phi_0^{(2)}})^2 \\ (\frac{\mu_0^{(2)}}{1-\phi_0^{(2)}})^2 & (\frac{\mu_0^{(2)}}{1-\phi_0^{(2)}})^2 + \frac{\phi_0^{(2)}E(e_t^{(2)})^2}{1-(\phi_0^{(2)})^2} + \phi_0^{(2)}E\{e_t^{(2)}(\eta_t^{(1)} - \eta_t^{(2)})\} \end{pmatrix}$$

and

$$A_1^{(2)} = \cdots = A_{M-1}^{(2)} = -\begin{pmatrix} 1 & \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}} - \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}} \\ \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}} - \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}} & a \end{pmatrix},$$

where

$$\begin{aligned} a &= (\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2 + \frac{\phi_0^{(1)}E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2} - \frac{2\mu_0^{(1)}\mu_0^{(2)}}{(1-\phi_0^{(1)})(1-\phi_0^{(2)})} - \frac{(\phi_0^{(1)}+\phi_0^{(2)})E(e_t^{(1)}e_t^{(2)})}{1-\phi_0^{(1)}\phi_0^{(2)}} \\ &\quad +(\frac{\mu_0^{(2)}}{1-\phi_0^{(2)}})^2 + \frac{\phi_0^{(2)}E(e_t^{(2)})^2}{1-(\phi_0^{(2)})^2} + \phi_0^{(1)}E(e_t^{(1)}\eta_t^{(2)}) - \phi_0^{(2)}E(e_t^{(2)}\eta_t^{(2)}). \end{aligned}$$

□

Lemma 5.3. *Suppose the conditions of Theorem 5.3 hold, then*

$$\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{W}_{3t}(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_3) \quad \text{and} \quad \frac{1}{T}\sum_{t=3}^{T}\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{W}_{3t}(\boldsymbol{\theta}_0) \xrightarrow{p} \boldsymbol{\Gamma}_3, \quad \text{as } T \to \infty,$$

where $\boldsymbol{\Sigma}_3$ *and* $\boldsymbol{\Gamma}_3$ *are defined in (5.24) and (5.25) respectively, and*

$$\boldsymbol{W}_{3t}(\boldsymbol{\theta}_0) = (Y_{t,1}^{(1)}, \frac{Y_{t,2}^{(1)}}{T}, Y_{t,1}^{(2)}, Y_{t,2}^{(2)}, \tilde{Y}_{t,1}^{(1)}, \frac{\hat{Y}_{t,1}^{(1)}}{T}, \tilde{Y}_{t,1}^{(2)}, \frac{\hat{Y}_{t,1}^{(2)}}{T}, \cdots, \tilde{Y}_{t,M-1}^{(1)}, \frac{\hat{Y}_{t,M-1}^{(1)}}{T}, \tilde{Y}_{t,M-1}^{(2)}, \frac{\hat{Y}_{t,M-1}^{(2)}}{T})^\tau.$$

Proof. In this case, $\{k_t^{(1)}\}_{t=1}^T$ is unit root or near unit root, while $\{k_t^{(1)} - k_t^{(2)}\}_{t=1}^T$ is stationary. Assume $\mu_0^{(1)} \neq 0$ and $\mu_0^{(2)} \neq 0$, then (5.12) and (5.19) hold, which implies that

$$\frac{k_{[Ts]}^{(2)}}{T} = \frac{k_{[Ts]}^{(1)}}{T} - \frac{k_{[Ts]}^{(1)} - k_{[Ts]}^{(2)}}{T} \xrightarrow{p} f_{\rho_1,\mu_0^{(1)}}(s), \tag{5.23}$$

for $s \in [0,1]$ as $T \to \infty$.

Note that

$$\begin{cases} Y_{t,1}^{(1)}(\boldsymbol{\theta}_0) = e_t^{(1)} + \eta_t^{(1)} - \phi_0^{(1)}\eta_{t-1}^{(1)}, \\ Y_{t,2}^{(1)}(\boldsymbol{\theta}_0) = (e_t^{(1)} + \eta_t^{(1)} - \phi_0^{(1)}\eta_{t-1}^{(1)})(k_{t-2}^{(1)} + \eta_{t-2}^{(1)}), \\ Y_{t,1}^{(2)}(\boldsymbol{\theta}_0) = e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)} - \phi_0^{(2)}(\eta_{t-1}^{(1)} - \eta_{t-1}^{(2)}), \\ Y_{t,2}^{(2)}(\boldsymbol{\theta}_0) = \{e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)} - \phi_0^{(2)}(\eta_{t-1}^{(1)} - \eta_{t-1}^{(2)})\}(k_{t-2}^{(1)} - k_{t-2}^{(2)} + \eta_{t-2}^{(1)} - \eta_{t-2}^{(2)}), \\ \tilde{Y}_{t,x}^{(i)}(\boldsymbol{\theta}_0) = \epsilon_{x,t}^{(i)} - \beta_{x,0}^{(i)}\eta_t^{(i)},\ i = 1,2;\ x = 1,2,\ldots,M-1, \\ \hat{Y}_{t,x}^{(i)}(\boldsymbol{\theta}_0) = (\epsilon_{x,t}^{(i)} - \beta_{x,0}^{(i)}\eta_t^{(i)})(k_{t-1}^{(i)} + \eta_{t-1}^{(i)}),\ i = 1,2;\ x = 1,2,\ldots,M-1. \end{cases}$$

Let $X_{t,1}^{(1)} = e_t^{(1)}$, $X_{t,2}^{(1)} = (e_t^{(1)} + \eta_t^{(1)})\frac{k_{t-2}^{(1)}}{T} - \eta_t^{(1)}\frac{k_{t-1}^{(1)}}{T}$, $X_{t,1}^{(2)} = e_t^{(2)} + (1-\phi_0^{(2)})(\eta_t^{(1)} - \eta_t^{(2)})$,

$X_{t,2}^{(2)} = (e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)})(k_{t-2}^{(1)} - k_{t-2}^{(2)} + \eta_{t-2}^{(1)} - \eta_{t-2}^{(2)}) - \phi_0^{(2)}(\eta_t^{(1)} - \eta_t^{(2)})(k_{t-1}^{(1)} - k_{t-1}^{(2)} + \eta_{t-1}^{(1)} - \eta_{t-1}^{(2)})$,

and define

$$\boldsymbol{U}_t(\boldsymbol{\theta}_0) = (X_{t,1}^{(1)}, X_{t,2}^{(1)}, X_{t,1}^{(2)}, X_{t,2}^{(2)}, \tilde{Y}_{t,1}^{(1)}, \frac{\hat{Y}_{t,1}^{(1)}}{T}, \tilde{Y}_{t,1}^{(2)}, \frac{\hat{Y}_{t,1}^{(2)}}{T}, \cdots, \tilde{Y}_{t,M-1}^{(1)}, \frac{\hat{Y}_{t,M-1}^{(1)}}{T}, \tilde{Y}_{t,M-1}^{(2)}, \frac{\hat{Y}_{t,M-1}^{(2)}}{T})^\tau.$$

Then we have

$$\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{W}_{3t}(\boldsymbol{\theta_0}) = \frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{U}_t(\boldsymbol{\theta_0}) + o_p(1), \quad T \to \infty.$$

Let $\mathcal{F}_t$ denote the σ-field generated by $\{(e_s^{(1)}, e_s^{(2)}, \varepsilon_{1,s}^{(1)}, \cdots, \varepsilon_{M,s}^{(1)}, \varepsilon_{1,s}^{(2)}, \cdots, \varepsilon_{M,s}^{(2)})^\tau : s \le t\}$, then $\{\boldsymbol{U}_t(\boldsymbol{\theta}_0), \mathcal{F}_t\}_{t=1}^{\infty}$ is a sequence of martingale difference. By the central limit theorem of martingale sequences in Hall & Heyde (2014), $\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{U}_t(\boldsymbol{\theta}_0)$ converges in distribution to a multivariate normal distribution with 0 means and covariance matrix $\boldsymbol{\Sigma}_3$ satisfying

$$\frac{1}{T}\sum_{t=3}^{T}E\{\boldsymbol{U}_t(\boldsymbol{\theta}_0)\boldsymbol{U}_t(\boldsymbol{\theta}_0)^\tau|\mathcal{F}_{t-1}\} \xrightarrow{p} \boldsymbol{\Sigma}_3 \overset{\text{def}}{=} (\sigma_{ij})_{4M\times 4M}, \tag{5.24}$$

i.e.,

$$\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{W}_{3t}(\boldsymbol{\theta_0}) = \frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{U}_t(\boldsymbol{\theta_0}) + o_p(1) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma}_3), \quad T \to \infty.$$

By the assumption that $\{(e_t^{(1)}, e_t^{(2)}, \varepsilon_{1,t}^{(1)}, \cdots, \varepsilon_{M,t}^{(1)}, \varepsilon_{1,t}^{(2)}, \cdots, \varepsilon_{M,t}^{(2)})^\tau\}_{t=1}^T$ is a sequence of independent random vectors with finite variance, it is straightforward to verify that

$$\sigma_{11} = E(e_t^{(1)})^2, \quad \sigma_{12} = E(e_t^{(1)})^2 \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds, \quad \sigma_{13} = E(e_t^{(1)}e_t^{(2)}) + (1-\phi_0^{(2)})E\{e_t^{(1)}(\eta_t^{(1)} - \eta_t^{(2)})\},$$

$$\sigma_{14} = \{E(e_t^{(1)}e_t^{(2)}) + (1-\phi_0^{(2)})E[e_t^{(1)}(\eta_t^{(1)} - \eta_t^{(2)})]\}\frac{\mu_0^{(2)}}{1-\phi_0^{(2)}} = \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}}\sigma_{13},$$

$$\sigma_{22} = E(e_t^{(1)})^2 \int_0^1 f^2_{\rho_1,\mu_0^{(1)}}(s)\,ds, \quad \sigma_{23} = \sigma_{13}\int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds,\ \sigma_{24} = \sigma_{13}\frac{\mu_0^{(2)}}{1-\phi_0^{(2)}}\int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds,$$

$$\sigma_{33} = E\{e_t^{(2)} + (1-\phi_0^{(2)})(\eta_t^{(1)} - \eta_t^{(2)})\}^2, \quad \sigma_{34} = \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}}\sigma_{33},$$

$$\begin{aligned}\sigma_{44} &= \{E(e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)})^2 + (\phi_0^{(2)})^2E(\eta_t^{(1)} - \eta_t^{(2)})^2\}\{(\tfrac{\mu_0^{(2)}}{1-\phi_0^{(2)}})^2 + \tfrac{E(e_t^{(2)})^2}{1-(\phi_0^{(2)})^2} + 2E[e_t^{(2)}(\eta_t^{(1)} - \eta_t^{(2)})] \\ &\quad + E(\eta_t^{(1)} - \eta_t^{(2)})^2\} - 2\phi_0^{(2)}E\{(e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)})(\eta_t^{(1)} - \eta_t^{(2)})\}\{(\tfrac{\mu_0^{(2)}}{1-\phi_0^{(2)}})^2 + \tfrac{\phi_0^{(2)}E(e_t^{(2)})^2}{1-(\phi_0^{(2)})^2} \\ &\quad + \phi_0^{(2)}E[e_t^{(2)}(\eta_t^{(1)} - \eta_t^{(2)})]\}.\end{aligned}$$

Similarly, for $x = 1, 2, \ldots, M-1$, we have

$$\sigma_{1,4x+1} = E\{e_t^{(1)}(\epsilon_{x,t}^{(1)} - \beta_{x,0}^{(1)}\eta_t^{(1)})\},\ \sigma_{1,4x+3} = E\{e_t^{(1)}(\epsilon_{x,t}^{(2)} - \beta_{x,0}^{(2)}\eta_t^{(2)})\},$$

$$\sigma_{1,4x+2} = \sigma_{2,4x+1} = \sigma_{1,4x+1}\int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds, \quad \sigma_{1,4x+4} = \sigma_{2,4x+3} = \sigma_{1,4x+3}\int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds,$$

$$\sigma_{2,4x+2} = \sigma_{1,4x+1}\int_0^1 f^2_{\rho_1,\mu_0^{(1)}}(s)\,ds, \quad \sigma_{2,4x+4} = \sigma_{1,4x+3}\int_0^1 f^2_{\rho_1,\mu_0^{(1)}}(s)\,ds,$$

$$\sigma_{3,4x+1} = E\{[e_t^{(2)} + (1-\phi_0^{(2)})(\eta_t^{(1)} - \eta_t^{(2)})](\epsilon_{x,t}^{(1)} - \beta_{x,0}^{(1)}\eta_t^{(1)})\}, \quad \sigma_{3,4x+2} = \sigma_{3,4x+1}\int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds,$$

$$\sigma_{3,4x+3} = E\{[e_t^{(2)} + (1-\phi_0^{(2)})(\eta_t^{(1)} - \eta_t^{(2)})](\epsilon_{x,t}^{(2)} - \beta_{x,0}^{(2)}\eta_t^{(2)})\},\ \sigma_{3,4x+4} = \sigma_{3,4x+3}\int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds,$$

$$\sigma_{4,4x+1} = \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}}\sigma_{3,4x+1}, \quad \sigma_{4,4x+3} = \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}}\sigma_{3,4x+3},$$

$$\sigma_{4,4x+2} = \sigma_{3,4x+1}\frac{\mu_0^{(2)}}{1-\phi_0^{(2)}}\int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds, \quad \sigma_{4,4x+4} = \sigma_{3,4x+3}\frac{\mu_0^{(2)}}{1-\phi_0^{(2)}}\int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds,$$

And for $x, y = 1, 2, \ldots, M-1$, we have

$$\sigma_{4x+1,4y+1} = E\{(\epsilon_{x,t}^{(1)} - \beta_{x,0}^{(1)}\eta_t^{(1)})(\epsilon_{y,t}^{(1)} - \beta_{y,0}^{(1)}\eta_t^{(1)})\},\ \sigma_{4x+1,4y+2} = \sigma_{4x+1,4y+1}\int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds,$$

$$\sigma_{4x+1,4y+3} = E\{(\epsilon_{x,t}^{(1)} - \beta_{x,0}^{(1)}\eta_t^{(1)})(\epsilon_{y,t}^{(2)} - \beta_{y,0}^{(2)}\eta_t^{(2)})\},\ \sigma_{4x+1,4y+4} = \sigma_{4x+1,4y+3}\int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds, ,$$

$$\sigma_{4x+2,4y+2} = E\{(\epsilon_{x,t}^{(1)} - \beta_{x,0}^{(1)}\eta_t^{(1)})(\epsilon_{y,t}^{(1)} - \beta_{y,0}^{(1)}\eta_t^{(1)})\}\int_0^1 f^2_{\rho_1,\mu_0^{(1)}}(s)\,ds,$$

$$\sigma_{4x+2,4y+3} = \sigma_{4x+1,4y+3}\int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds, \quad \sigma_{4x+2,4y+4} = \sigma_{4x+1,4y+3}\int_0^1 f^2_{\rho_1,\mu_0^{(1)}}(s)\,ds,$$

$$\sigma_{4x+3,4y+3} = E\{(\epsilon_{x,t}^{(2)} - \beta_{x,0}^{(2)}\eta_t^{(2)})(\epsilon_{y,t}^{(2)} - \beta_{y,0}^{(2)}\eta_t^{(2)})\},\ \sigma_{4x+3,4y+4} = \sigma_{4x+3,4y+3}\int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds,$$

$$\sigma_{4x+4,4y+4} = \sigma_{4x+3,4y+3}\int_0^1 f^2_{\rho_1,\mu_0^{(1)}}(s)\,ds.$$

Furthermore, note that

$$\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{W}_{3t}(\boldsymbol{\theta}_0) = \mathbf{diag}\left(A_{t0}^{(1)}, A_{t0}^{(2)}, A_{t1}^{(1)}, A_{t1}^{(2)}, \cdots, A_{t(M-1)}^{(1)}, A_{t(M-1)}^{(2)}\right)_{4(M-1)\times 4(M-1)},$$

where for $i = 1, 2$ and $x = 1, 2, \ldots, M-1$,

$$A_{t0}^{(1)} = -\begin{pmatrix} 1 & Z_{t-1}^{(1)} \\ \frac{Z_{t-2}^{(1)}}{T} & \frac{Z_{t-1}^{(1)}Z_{t-2}^{(1)}}{T} \end{pmatrix},\ A_{tx}^{(i)} = -\begin{pmatrix} 1 & Z_t^{(i)} \\ \frac{Z_{t-1}^{(i)}}{T} & \frac{Z_t^{(i)}Z_{t-1}^{(i)}}{T} \end{pmatrix},$$

and

$$A_{t0}^{(2)} = -\begin{pmatrix} 1 & Z_{t-1}^{(1)} - Z_{t-1}^{(2)} \\ Z_{t-2}^{(1)} - Z_{t-2}^{(2)} & (Z_{t-1}^{(1)} - Z_{t-1}^{(2)})(Z_{t-2}^{(1)} - Z_{t-2}^{(2)}) \end{pmatrix}.$$

It follows from (5.12), (5.19) and (5.23) that as $T \to \infty$

$$\frac{1}{T}\sum_{t=3}^{T}\frac{Z_t^{(1)}}{T} = \frac{1}{T}\sum_{t=3}^{T}\frac{k_t^{(1)}}{T} + \frac{1}{T^2}\sum_{t=3}^{T}\eta_t^{(1)} \xrightarrow{p} \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds,$$

$$\frac{1}{T}\sum_{t=3}^{T}\frac{Z_t^{(2)}}{T} = \frac{1}{T}\sum_{t=3}^{T}\frac{k_t^{(2)}}{T} + \frac{1}{T^2}\sum_{t=3}^{T}\eta_t^{(2)} \xrightarrow{p} \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds,$$

$$\frac{1}{T}\sum_{t=3}^{T}\frac{Z_t^{(1)}}{T}\frac{Z_{t-1}^{(1)}}{T} = \frac{1}{T}\sum_{t=3}^{T}\frac{k_t^{(1)}}{T}\frac{k_{t-1}^{(1)}}{T} + o_p(1) \xrightarrow{p} \int_0^1 f^2_{\rho_1,\mu_0^{(1)}}(s)\,ds,$$

$$\frac{1}{T}\sum_{t=3}^{T}\frac{Z_t^{(2)}}{T}\frac{Z_{t-1}^{(2)}}{T} = \frac{1}{T}\sum_{t=3}^{T}\frac{k_t^{(2)}}{T}\frac{k_{t-1}^{(2)}}{T} + o_p(1) \xrightarrow{p} \int_0^1 f^2_{\rho_1,\mu_0^{(1)}}(s)\,ds,$$

$$\frac{1}{T}\sum_{t=3}^{T}(Z_{t-1}^{(1)} - Z_{t-1}^{(2)}) = \frac{1}{T}\sum_{t=3}^{T}(k_{t-1}^{(1)} - k_{t-1}^{(2)}) + o_p(1) \xrightarrow{p} \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}},$$

$$\frac{1}{T}\sum_{t=3}^{T}(Z_{t-1}^{(1)} - Z_{t-1}^{(2)})(Z_{t-2}^{(1)} - Z_{t-2}^{(2)}) \xrightarrow{p} (\frac{\mu_0^{(2)}}{1-\phi_0^{(2)}})^2 + \frac{\phi_0^{(2)}E(e_t^{(2)})^2}{1-(\phi_0^{(2)})^2} + \phi_0^{(2)}E\{e_t^{(2)}(\eta_t^{(1)} - \eta_t^{(2)})\}.$$

This implies that

$$\frac{1}{T}\sum_{t=3}^{T}\{\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{W}_{3t}(\boldsymbol{\theta}_0)\} \xrightarrow{p} \boldsymbol{\Gamma}_3 \text{ as } T \to \infty,$$

where

$$\boldsymbol{\Gamma}_3 = \mathbf{diag}\left(A_0^{(1)}, A_0^{(2)}, A_1^{(1)}, A_1^{(2)}, \cdots, A_{M-1}^{(1)}, A_{M-1}^{(2)}\right)_{4(M-1)\times 4(M-1)} \tag{5.25}$$

with

$$A_0^{(1)} = A_1^{(1)} = \cdots = A_{M-1}^{(1)} = A_1^{(2)} = \cdots = A_{M-1}^{(2)} = -\begin{pmatrix} 1 & \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds \\ \int_0^1 f_{\rho_1,\mu_0^{(1)}}(s)\,ds & \int_0^1 f^2_{\rho_1,\mu_0^{(1)}}(s)\,ds \end{pmatrix}$$

and

$$A_0^{(2)} = -\begin{pmatrix} 1 & \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}} \\ \frac{\mu_0^{(2)}}{1-\phi_0^{(2)}} & (\frac{\mu_0^{(2)}}{1-\phi_0^{(2)}})^2 + \frac{\phi_0^{(2)}E(e_t^{(2)})^2}{1-(\phi_0^{(2)})^2} + \phi_0^{(2)}E\{e_t^{(2)}(\eta_t^{(1)} - \eta_t^{(2)})\} \end{pmatrix}.$$

□

Lemma 5.4. *Suppose the conditions of Theorem 5.4 hold, then*

$$\frac{1}{\sqrt{T}}\sum_{t=3}^{T} \boldsymbol{W}_{4t}(\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_4) \quad \textit{and} \quad \frac{1}{T}\sum_{t=3}^{T} \frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{W}_{4t}(\boldsymbol{\theta}_0) \xrightarrow{p} \boldsymbol{\Gamma}_4, \quad \textit{as } T \to \infty,$$

where $\boldsymbol{\Sigma}_4$ *and* $\boldsymbol{\Gamma}_4$ *are defined in (5.26) and (5.27) respectively, and*

$$\boldsymbol{W}_{4t}(\boldsymbol{\theta}_0) = (Y_{t,1}^{(1)}, Y_{t,2}^{(1)}, Y_{t,1}^{(2)}, \frac{Y_{t,2}^{(2)}}{T}, \tilde{Y}_{t,1}^{(1)}, \hat{Y}_{t,1}^{(1)}, \tilde{Y}_{t,1}^{(2)}, \frac{\hat{Y}_{t,1}^{(2)}}{T}, \cdots, \tilde{Y}_{t,M-1}^{(1)}, \hat{Y}_{t,M-1}^{(1)}, \tilde{Y}_{t,M-1}^{(2)}, \frac{\hat{Y}_{t,M-1}^{(2)}}{T})^{\tau}.$$

Proof. In this case, we have

$$E(k_t^{(1)}) = \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}, \quad E(k_t^{(1)})^2 = (\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2 + \frac{E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2}.$$

and

$$\frac{k_{[Ts]}^{(1)} - k_{[Ts]}^{(2)}}{T} \xrightarrow{p} f_{\rho_2,\mu_0^{(2)}}(s) \text{ for } s \in [0,1]$$

with $f_{\rho_2,\mu_0^{(2)}}(s)$ defined in (5.13). This implies that

$$\frac{k_{[Ts]}^{(2)}}{T} = \frac{k_{[Ts]}^{(1)}}{T} - \frac{k_{[Ts]}^{(1)} - k_{[Ts]}^{(2)}}{T} \xrightarrow{p} -f_{\rho_2,\mu_0^{(2)}}(s) \text{ for } s \in [0,1].$$

Note that

$$\begin{cases} Y_{t,1}^{(1)}(\boldsymbol{\theta}_0) = e_t^{(1)} + \eta_t^{(1)} - \phi_0^{(1)}\eta_{t-1}^{(1)}, \\ Y_{t,2}^{(1)}(\boldsymbol{\theta}_0) = (e_t^{(1)} + \eta_t^{(1)} - \phi_0^{(1)}\eta_{t-1}^{(1)})(k_{t-2}^{(1)} + \eta_{t-2}^{(1)}), \\ Y_{t,1}^{(2)}(\boldsymbol{\theta}_0) = e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)} - \phi_0^{(2)}(\eta_{t-1}^{(1)} - \eta_{t-1}^{(2)}), \\ Y_{t,2}^{(2)}(\boldsymbol{\theta}_0) = \{e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)} - \phi_0^{(2)}(\eta_{t-1}^{(1)} - \eta_{t-1}^{(2)})\}(k_{t-2}^{(1)} - k_{t-2}^{(2)} + \eta_{t-2}^{(1)} - \eta_{t-2}^{(2)}), \\ \tilde{Y}_{t,x}^{(i)}(\boldsymbol{\theta}_0) = \epsilon_{x,t}^{(i)} - \beta_{x,0}^{(i)}\eta_t^{(i)},\ i = 1, 2;\ x = 1, 2, \ldots, M-1, \\ \hat{Y}_{t,x}^{(i)}(\boldsymbol{\theta}_0) = (\epsilon_{x,t}^{(i)} - \beta_{x,0}^{(i)}\eta_t^{(i)})(k_{t-1}^{(i)} + \eta_{t-1}^{(i)}),\ i = 1, 2;\ x = 1, 2, \ldots, M-1. \end{cases}$$

Let $X_{t,1}^{(1)} = e_t^{(1)} + (1 - \phi_0^{(1)})\eta_t^{(1)}$, $X_{t,1}^{(2)} = e_t^{(2)}$,

$$X_{t,2}^{(1)} = (e_t^{(1)} + \eta_t^{(1)})(k_{t-2}^{(1)} + \eta_{t-2}^{(1)}) - \phi_0^{(1)}\eta_t^{(1)}(k_{t-1}^{(1)} + \eta_{t-1}^{(1)}),$$

$$X_{t,2}^{(2)} = (e_t^{(2)} + \eta_t^{(1)} - \eta_t^{(2)})\frac{k_{t-2}^{(1)} - k_{t-2}^{(2)}}{T} - (\eta_t^{(1)} - \eta_t^{(2)})\frac{k_{t-1}^{(1)} - k_{t-1}^{(2)}}{T},$$

and define

$$\boldsymbol{U}_t(\boldsymbol{\theta}_0) = (X_{t,1}^{(1)}, X_{t,2}^{(1)}, X_{t,1}^{(2)}, X_{t,2}^{(2)}, \tilde{Y}_{t,1}^{(1)}, \hat{Y}_{t,1}^{(1)}, \tilde{Y}_{t,1}^{(2)}, \frac{\hat{Y}_{t,1}^{(2)}}{T}, \cdots, \tilde{Y}_{t,M-1}^{(1)}, \hat{Y}_{t,M-1}^{(1)}, \tilde{Y}_{t,M-1}^{(2)}, \frac{\hat{Y}_{t,M-1}^{(2)}}{T})^\tau.$$

Then we have

$$\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{W}_{4t}(\boldsymbol{\theta_0}) = \frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{U}_t(\boldsymbol{\theta_0}) + o_p(1), \quad T \to \infty.$$

Let $\mathcal{F}_t$ denote the σ-field generated by $\{(e_s^{(1)}, e_s^{(2)}, \varepsilon_{1,s}^{(1)}, \cdots, \varepsilon_{M,s}^{(1)}, \varepsilon_{1,s}^{(2)}, \cdots, \varepsilon_{M,s}^{(2)})^\tau : s \le t\}$, then $\{\boldsymbol{U}_t(\boldsymbol{\theta}_0), \mathcal{F}_t\}_{t=1}^{\infty}$ is a sequence of martingale difference. By the central limit theorem of martingale sequences in Hall & Heyde (2014), $\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{U}_t(\boldsymbol{\theta}_0)$ converges in distribution to a multivariate normal distribution with 0 means and covariance matrix $\boldsymbol{\Sigma}_4$ satisfying

$$\frac{1}{T}\sum_{t=3}^{T}E\{\boldsymbol{U}_t(\boldsymbol{\theta}_0)\boldsymbol{U}_t(\boldsymbol{\theta}_0)^\tau|\mathcal{F}_{t-1}\} \xrightarrow{p} \boldsymbol{\Sigma}_4 \stackrel{\text{def}}{=} (\sigma_{ij})_{4M\times 4M}, \tag{5.26}$$

i.e.,

$$\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{W}_{4t}(\boldsymbol{\theta_0})=\frac{1}{\sqrt{T}}\sum_{t=3}^{T}\boldsymbol{U}_t(\boldsymbol{\theta_0})+o_p(1)\stackrel{d}{\to}N(\mathbf{0},\boldsymbol{\Sigma}_4),\quad n\to\infty.$$

By the assumption that $\{(e_t^{(1)},e_t^{(2)},\varepsilon_{1,t}^{(1)},\cdots,\varepsilon_{M,t}^{(1)},\varepsilon_{1,t}^{(2)},\cdots,\varepsilon_{M,t}^{(2)})^\tau\}_{t=1}^T$ is a sequence of independent random vectors with finite variance, it is straightforward to verify that

$$\sigma_{11}=E\{e_t^{(1)}+(1-\phi_0^{(1)})\eta_t^{(1)}\}^2,\quad \sigma_{12}=\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\sigma_{11},$$

$$\sigma_{13}=E(e_t^{(1)}e_t^{(2)})+(1-\phi_0^{(1)})E(e_t^{(2)}\eta_t^{(1)}),\quad \sigma_{14}=\sigma_{13}\int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds,$$

$$\begin{aligned}\sigma_{22}\ &=\{E(e_t^{(1)}+\eta_t^{(1)})^2+(\phi_0^{(1)})^2E(\eta_t^{(1)})^2\}\{(\tfrac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2+\tfrac{E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2}+2E(e_t^{(1)}\eta_t^{(1)})+E(\eta_t^{(1)})^2\}\\ &\quad-2\phi_0^{(1)}E\{\eta_t^{(1)}(e_t^{(1)}+\eta_t^{(1)})\}\{(\tfrac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2+\tfrac{\phi_0^{(1)}E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2}+\phi_0^{(1)}E(e_t^{(1)}\eta_t^{(1)})\},\end{aligned}$$

$$\sigma_{23}=\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\sigma_{13},\quad \sigma_{24}=\sigma_{13}\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds,$$

$$\sigma_{33}=E(e_t^{(2)})^2,\quad \sigma_{34}=E(e_t^{(2)})^2\int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds,\quad \sigma_{44}=E(e_t^{(2)})^2\int_0^1 f^2_{\rho_2,\mu_0^{(2)}}(s)\,ds.$$

Similarly, for $x=1,2,\dots,M-1$, we have

$$\sigma_{1,4x+1}=E\{[e_t^{(1)}+(1-\phi_0^{(1)})\eta_t^{(1)}](\epsilon_{x,t}^{(1)}-\beta_{x,0}^{(1)}\eta_t^{(1)})\},\ \sigma_{1,4x+3}=E\{[e_t^{(1)}+(1-\phi_0^{(1)})\eta_t^{(1)}](\epsilon_{x,t}^{(2)}-\beta_{x,0}^{(2)}\eta_t^{(2)})\},$$

$$\sigma_{1,4x+2}=\sigma_{2,4x+1}=\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\sigma_{1,4x+1},\quad \sigma_{1,4x+4}=-\sigma_{1,4x+3}\int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds,$$

$$\begin{aligned}\sigma_{2,4x+2}\ &=E\{(e_t^{(1)}+\eta_t^{(1)})(\epsilon_{x,t}^{(1)}-\beta_{x,0}^{(1)}\eta_t^{(1)})\}\{(\tfrac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2+\tfrac{\phi_0^{(1)}E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2}+\phi_0^{(1)}E(e_t^{(1)}\eta_t^{(1)})\}\\ &\quad-\phi_0^{(1)}E\{\eta_t^{(1)}(\epsilon_{x,t}^{(1)}-\beta_{x,0}^{(1)}\eta_t^{(1)})\}\{(\tfrac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2+\tfrac{E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2}+2E(e_t^{(1)}\eta_t^{(1)})+E(\eta_t^{(1)})^2\},\end{aligned}$$

$$\sigma_{2,4x+3}=\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\sigma_{1,4x+3},\quad \sigma_{2,4x+4}=-\sigma_{1,4x+3}\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds,$$

$$\sigma_{3,4x+1}=E\{e_t^{(2)}(\epsilon_{x,t}^{(1)}-\beta_{x,0}^{(1)}\eta_t^{(1)})\},\quad \sigma_{3,4x+2}=\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\sigma_{3,4x+1},$$

$$\sigma_{3,4x+3} = E\{e_t^{(2)}(\epsilon_{x,t}^{(2)} - \beta_{x,0}^{(2)}\eta_t^{(2)})\},\ \sigma_{3,4x+4} = -\sigma_{3,4x+3}\int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds,$$

$$\sigma_{4,4x+1} = \sigma_{3,4x+1}\int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds, \quad \sigma_{4,4x+3} = \sigma_{3,4x+3}\int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds,$$

$$\sigma_{4,4x+2} = \sigma_{3,4x+1}\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds, \quad \sigma_{4,4x+4} = -\sigma_{3,4x+3}\int_0^1 f^2_{\rho_2,\mu_0^{(2)}}(s)\,ds.$$

And for $x, y = 1, 2, \ldots, M-1$, we have

$$\sigma_{4x+1,4y+1} = E\{(\epsilon_{x,t}^{(1)} - \beta_{x,0}^{(1)}\eta_t^{(1)})(\epsilon_{y,t}^{(1)} - \beta_{y,0}^{(1)}\eta_t^{(1)})\}, \quad \sigma_{4x+1,4y+2} = \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\sigma_{4x+1,4y+1},$$

$$\sigma_{4x+1,4y+3} = E\{(\epsilon_{x,t}^{(1)} - \beta_{x,0}^{(1)}\eta_t^{(1)})(\epsilon_{y,t}^{(2)} - \beta_{y,0}^{(2)}\eta_t^{(2)})\}, \quad \sigma_{4x+1,4y+4} = -\sigma_{4x+1,4y+3}\int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds,$$

$$\sigma_{4x+2,4y+2} = \sigma_{4x+1,4y+1}\{(\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2 + \frac{E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2} + E(\eta_t^{(1)})^2\},$$

$$\sigma_{4x+2,4y+3} = \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\sigma_{4x+1,4y+3}, \quad \sigma_{4x+2,4y+4} = -\sigma_{4x+1,4y+3}\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}}\int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds,$$

$$\sigma_{4x+3,4y+3} = E\{(\epsilon_{x,t}^{(2)} - \beta_{x,0}^{(2)}\eta_t^{(2)})(\epsilon_{y,t}^{(2)} - \beta_{y,0}^{(2)}\eta_t^{(2)})\},\ \sigma_{4x+3,4y+4} = -\sigma_{4x+3,4y+3}\int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds,$$

$$\sigma_{4x+4,4y+4} = \sigma_{4x+3,4y+3}\int_0^1 f^2_{\rho_2,\mu_0^{(2)}}(s)\,ds.$$

Furthermore, note that

$$\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{W}_{4t}(\boldsymbol{\theta}_0) = \mathbf{diag}\left(A_{t0}^{(1)}, A_{t0}^{(2)}, A_{t1}^{(1)}, A_{t1}^{(2)}, \cdots, A_{t(M-1)}^{(1)}, A_{t(M-1)}^{(2)}\right)_{4(M-1)\times 4(M-1)},$$

where

$$A_{t0}^{(1)} = -\begin{pmatrix} 1 & Z_{t-1}^{(1)} \\ Z_{t-2}^{(1)} & Z_{t-1}^{(1)}Z_{t-2}^{(1)} \end{pmatrix},\ A_{t1}^{(1)} = \cdots = A_{t(M-1)}^{(1)} = -\begin{pmatrix} 1 & Z_t^{(1)} \\ Z_{t-1}^{(1)} & Z_t^{(1)}Z_{t-1}^{(1)} \end{pmatrix},$$

$$A_{t1}^{(2)} = \cdots = A_{t(M-1)}^{(2)} = -\begin{pmatrix} 1 & Z_t^{(2)} \\ \frac{1}{T}Z_{t-1}^{(2)} & \frac{1}{T}Z_t^{(2)}Z_{t-1}^{(2)} \end{pmatrix},$$

and

$$A_{t0}^{(2)} = -\begin{pmatrix} 1 & Z_{t-1}^{(1)} - Z_{t-1}^{(2)} \\ \frac{1}{T}(Z_{t-2}^{(1)} - Z_{t-2}^{(2)}) & \frac{1}{T}(Z_{t-1}^{(1)} - Z_{t-1}^{(2)})(Z_{t-2}^{(1)} - Z_{t-2}^{(2)}) \end{pmatrix}.$$

It follows from (5.18), (5.19) and (5.20) that as $T \to \infty$

$$\frac{1}{T}\sum_{t=3}^{T} Z_t^{(1)} = \frac{1}{T}\sum_{t=3}^{T} k_t^{(1)} + \frac{1}{T}\sum_{t=3}^{T} \eta_t^{(1)} \xrightarrow{p} \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}},$$

$$\frac{1}{T}\sum_{t=3}^{T} \frac{Z_t^{(2)}}{T} = \frac{1}{T}\sum_{t=3}^{T} \frac{k_t^{(2)}}{T} + \frac{1}{T^2}\sum_{t=3}^{T} \eta_t^{(2)} \xrightarrow{p} -\int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds,$$

$$\begin{aligned} \tfrac{1}{T}\textstyle\sum_{t=3}^{T} Z_t^{(1)} Z_{t-1}^{(1)} &= \tfrac{1}{T}\textstyle\sum_{t=3}^{T} k_t^{(1)} k_{t-1}^{(1)} + \tfrac{1}{T}\textstyle\sum_{t=3}^{T} k_t^{(1)} \eta_{t-1}^{(1)} + o_p(1) \\ &\xrightarrow{p} (\tfrac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2 + \tfrac{\phi_0^{(1)} E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2} + \phi_0^{(1)} E(e_t^{(1)}\eta_t^{(1)}), \end{aligned}$$

$$\frac{1}{T}\sum_{t=3}^{T} \frac{Z_t^{(2)}}{T}\frac{Z_{t-1}^{(2)}}{T} = \frac{1}{T}\sum_{t=3}^{T} \frac{k_t^{(2)}}{T}\frac{k_{t-1}^{(2)}}{T} + o_p(1) \xrightarrow{p} \int_0^1 f^2_{\rho_2,\mu_0^{(2)}}(s)\,ds,$$

$$\frac{1}{T}\sum_{t=3}^{T} \frac{Z_{t-1}^{(1)} - Z_{t-1}^{(2)}}{T} = \frac{1}{T}\sum_{t=3}^{T} \frac{k_{t-1}^{(1)} - k_{t-1}^{(2)}}{T} + o_p(1) \xrightarrow{p} \int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds,$$

$$\frac{1}{T}\sum_{t=3}^{T} \frac{Z_{t-1}^{(1)} - Z_{t-1}^{(2)}}{T}\frac{Z_{t-2}^{(1)} - Z_{t-2}^{(2)}}{T} \xrightarrow{p} \int_0^1 f^2_{\rho_2,\mu_0^{(2)}}(s)\,ds.$$

This implies that

$$\frac{1}{T}\sum_{t=3}^{T}\{\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{W}_{4t}(\boldsymbol{\theta}_0)\} \xrightarrow{p} \boldsymbol{\Gamma}_4 \text{ as } T \to \infty,$$

where

$$\boldsymbol{\Gamma}_4 = \mathbf{diag}\left(A_0^{(1)}, A_0^{(2)}, A_1^{(1)}, A_1^{(2)}, \cdots, A_{M-1}^{(1)}, A_{M-1}^{(2)}\right)_{4(M-1)\times 4(M-1)} \tag{5.27}$$

with

$$A_0^{(1)} = \cdots = A_{M-1}^{(1)} = -\begin{pmatrix} 1 & \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}} \\ \frac{\mu_0^{(1)}}{1-\phi_0^{(1)}} & (\frac{\mu_0^{(1)}}{1-\phi_0^{(1)}})^2 + \frac{\phi_0^{(1)} E(e_t^{(1)})^2}{1-(\phi_0^{(1)})^2} + \phi_0^{(1)} E(e_t^{(1)}\eta_t^{(1)}) \end{pmatrix},$$

$$A_0^{(2)} = -\begin{pmatrix} 1 & \int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds \\ \int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds & \int_0^1 f^2_{\rho_2,\mu_0^{(2)}}(s)\,ds \end{pmatrix}$$

and

$$A_1^{(2)} = \cdots = A_{M-1}^{(2)} = -\begin{pmatrix} 1 & -\int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds \\ -\int_0^1 f_{\rho_2,\mu_0^{(2)}}(s)\,ds & \int_0^1 f^2_{\rho_2,\mu_0^{(2)}}(s)\,ds \end{pmatrix}.$$

□

PART 6

CONCLUSION

In this dissertation, I review the classical Lee & Carter (1992) paper and some recent literature with a mortality model adapted from the classic paper. Once again, how mortality rates are modeled has numerous implications in annuity and pension fund industry practices. Practitioners can rely on parametric mortality models for pricing, longevity risk management, and compliance purposes. Besides, by clarifying some details of Lee-Carter alike mortality models, we are also contributing to the theoretical literature.

Let me summarize some major results of this dissertation now. In Part 2 our proposed model reflects our understanding that imposing a constraint regarding mortality index k_t yields unreasonable implications. By replacing the constraint regarding k_t with a new constraint regarding model parameters α_x, we are able to derive our estimation method using $\hat{Z}_t$ to approach k_t, which is contingent on $\sum_{x=1}^{M} \alpha_x = 0$ and some very general regularity conditions for the error sequence in the model. Based on the regularity conditions in Part 2, we have different asymptotic results under different cases regarding AR(1) time series model parameters (can be inconsistent estimator under some cases). We present a unit root test for mortality data, given that the presence of unit root decides asymptotic results. According to our analysis of U.S. mortality rate data, the proposed test rejects the unit root hypothesis for male and combined mortality rates but fails to reject the unit root hypothesis for the male mortality rates. Hence, this calls for unified methods for estimating parameters and forecasting mortality rates regardless of whether the mortality index is stationary or near unit root or unit root.

Part 3 of this dissertation provides an answer to this question. In Part 3, we point out that inconsistency of estimators proposed in Part 2 is due to the correlation of two terms in the score equations which are used to derive them. As a result, one way to mitigate this issue is to modify the score equations by taking an additional lag. The new score equations eventually enable us to derive new bias-corrected estimators consistent with a normal limit regardless of whether the mortality index follows a stationary or unit root AR(1) time series model with a nonzero intercept. The accompanying simulation study confirmed this by showing bias-corrected estimators display more negligible bias and smaller mean squared error.

In Part 4 of this dissertation, I study whether an AR(1) model, as appeared in the dissertation, is adequate for characterizing mortality index in the available mortality rate dataset among the AR(p) family of models. This question is not as trivial as model selection for a time series because the mortality index is unobservable. In this Part, I focus on an AR(2) process and develop hypothesis tests whether the AR(2) parameter $\phi_2 = 0$ is based on a consistent estimator. When applied to U.S. mortality data, results of tests show that AR(1) model is adequate for female, male, and combined datasets. Meanwhile, the stationary AR(1) model is sound for female and combined mortality rates; the unit root AR(1) model is suitable for male mortality rates.

Last but not least, I apply the bias-corrected inference method to a bivariate setting in Part 5. Since most life insurers and annuity underwriters have policyholders from multiple populations, they need to understand the risk of a portfolio of insurance policies targeting multiple populations. Our model in Part 5 can be used for such risk modeling. To develop a two population model, I apply the bias-corrected inference method developed in 3 to the mortality index $\{k_t^{(1)}\}$ and the difference of indices $\{k_t^{(1)} - k_t^{(2)}\}$. Some major asymptotic results follow naturally from results in Part 3.

Implications of our model on longevity hedging can be argued as follows: in Li et al. (2018), hedging strategies are presented that use derivatives to hedge against time-t values of longevity deltas of (unpaid) annuity liability, which is a function of individual mortality model parameters. Although I use a modified Lee-Carter mortality model in this dissertation, the corresponding time-t value of longevity deltas can be derived similarly. My future research would be to develop delta hedging strategies based on the proposed models in this dissertation.

This dissertation makes contributions to several strands of literature. First, I continue the conversation regarding inference pitfalls of some recent theoretical literature related to the Lee-Carter mortality model. In my dissertation, I focus on proposing a tweaked model with modified parameter constraints and corresponding statistical inference. I was able to show asymptotic results for proposed estimators. Some of the results in this dissertation are based on Leng & Peng (2016) and Leng & Peng (2017).

Second, contributions are made to time series literature because this dissertation provides an application of time series models and α-mixing process. All asymptotic results and the unit root test included in this dissertation are contingent on regularity conditions as mentioned in the dissertation. These regularity conditions should be very general and can accommodate most real mortality datasets as long as the unexplained error terms are not too big.

Last but not least, we are making contributions to actuarial industry practice by proposing this alternative mortality model. In addition to asymptotic results which imply that the estimators converge to actual model parameters when the number of observations (in the mortality dataset, the number of observed years, T) is big enough, our real data analysis combined with simulation study have shown that even for reasonably small datasets (e.g., the U.S. mortality dataset), our bias-corrected inference can produce estimates that approach actual model parameters very well. As we have mentioned before, a suitable model parameter estimator can be very helpful to industry practices like longevity hedging.

One shortcoming of methodology in this dissertation is that the number of model parameters is relatively big (twice the number of age groups plus the number of AR time series model parameters), so this might cause problems for developing a practical delta hedging strategy. Some other literature uses models with very few model parameters, usually only two. This is an issue that is worth further investigation.

PART 7

R CODE

In this part I attach all R code for data analysis and simulation study in this dissertation.

For single population (univariate) datasets (in Parts 2, 3 and 4):

```
# required R packages:
# dplyr, stringr, demography, CADFtest, ggplot2, latex2exp, snowfall, rlecuyer

## DATA IMPORT

# USA data, 5 x 1, male, female and total data
# http://www.mortality.org/cgi-bin/hmd/country.php?cntr=USA&level=1

# load raw data files
# to download data other than 1x1, do not use the 'demography' package functions

DA_lifetbl_both <- read.table('bltper_5x1.txt', skip = 2, header = TRUE,
    stringsAsFactors = FALSE)
DA_lifetbl_male <- read.table('mltper_5x1.txt', skip = 2, header = TRUE,
    stringsAsFactors = FALSE)
DA_lifetbl_female <- read.table('fltper_5x1.txt', skip = 2, header = TRUE,
    stringsAsFactors = FALSE)

## DATA CLEANUP

# replace 110+ with 110, change data format to numeric
library(dplyr)
library(stringr)

DA_lifetbl_both <- mutate(DA_lifetbl_both,
    Age0 = as.numeric(str_extract(DA_lifetbl_both$Age, '[0-9]+')))
DA_lifetbl_male <- mutate(DA_lifetbl_male,
    Age0 = as.numeric(str_extract(DA_lifetbl_male$Age, '[0-9]+')))
DA_lifetbl_female <- mutate(DA_lifetbl_female,
    Age0 = as.numeric(str_extract(DA_lifetbl_female$Age, '[0-9]+')))
```

```
# focus on ages of 25 - 74
DA_lifetbl_both <-
    DA_lifetbl_both[DA_lifetbl_both$Age0 >= 25 & DA_lifetbl_both$Age0 < 75,]
DA_lifetbl_male <-
    DA_lifetbl_male[DA_lifetbl_male$Age0 >= 25 & DA_lifetbl_male$Age0 < 75,]
DA_lifetbl_female <-
    DA_lifetbl_female[DA_lifetbl_female$Age0 >= 25 & DA_lifetbl_female$Age0 < 75,]

# DA_n_agegroups: number of age groups (M in paper)
DA_n_agegroups <- length(unique(DA_lifetbl_both$Age0)) # DA_n_agegroups = 10

# DA_n_periods: number of years observed (T in paper)
DA_n_periods <- length(unique(DA_lifetbl_both$Year)) # DA_n_periods = 83

## IMPLEMENTATION OF THE ORITINAL LEE-CARTER MODEL

LC_func <- function(LC_n_agegroups, LC_n_periods,
    LC_lifetbl_mx, LC_agelist, LC_yearlist) {

    library(demography)

    # wrap up available data for input into 'lca' function
    LC_wrapped <- demogdata(matrix(data = LC_lifetbl_mx, nrow = LC_n_agegroups,
        ncol = LC_n_periods, byrow = FALSE),
        pop = matrix(0, LC_n_agegroups, LC_n_periods),
        ages = LC_agelist, years = LC_yearlist,
        type = 'mortality', label = 'USA_5x1', name = 'lee-carter', lambda = 0)

    # apply 'lca' function in 'demography' package
    LC_LCA.fitting <- lca(LC_wrapped, adjust = 'none')

    # ax
    LC_ax <- LC_LCA.fitting$ax
    names(LC_ax) <- NULL

    # bx
    LC_bx <- LC_LCA.fitting$bx
    names(LC_bx) <- NULL

    # kt
    LC_kt <- array(LC_LCA.fitting$kt)

    # below: fitting k_t with (5) in Draft #4
```

```
    LC_kt1 <- LC_kt[-1]
    LC_kt0 <- LC_kt[-LC_n_periods]
    LC_kt.fitting <- lm(LC_kt1 ~ LC_kt0)

    # report format: (estimate, standard error)
    LC_mu <- LC_kt.fitting$coefficients[1]
    names(LC_mu) <- NULL

    LC_mu_se <- coef(summary(LC_kt.fitting))[, 'Std. Error'][1]
    names(LC_mu_se) <- NULL

    LC_phi <- LC_kt.fitting$coefficients[2]
    names(LC_phi) <- NULL

    LC_phi_se <- coef(summary(LC_kt.fitting))[, 'Std. Error'][2]
    names(LC_phi_se) <- NULL

    # Augmented Dickey-Fuller test for unit root
    library(CADFtest)
    LC_kt.test <- CADFtest(LC_kt ~ 1, type = 'drift', lags = 1)
    LC_kt.test.p.value <- LC_kt.test$p.value

    return(list(LC_ax, LC_bx, LC_kt,
        LC_mu, LC_mu_se, LC_phi, LC_phi_se, LC_kt.test.p.value))

}

DA_LC_both <- LC_func(LC_n_agegroups = DA_n_agegroups,
    LC_n_periods = DA_n_periods,
    LC_lifetbl_mx = DA_lifetbl_both$mx,
    LC_agelist = unique(DA_lifetbl_both$Age0),
    LC_yearlist = unique(DA_lifetbl_both$Year))
DA_LC_male <- LC_func(LC_n_agegroups = DA_n_agegroups,
    LC_n_periods = DA_n_periods,
    LC_lifetbl_mx = DA_lifetbl_male$mx,
    LC_agelist = unique(DA_lifetbl_male$Age0),
    LC_yearlist = unique(DA_lifetbl_male$Year))
DA_LC_female <- LC_func(LC_n_agegroups = DA_n_agegroups,
    LC_n_periods = DA_n_periods,
    LC_lifetbl_mx = DA_lifetbl_female$mx,
    LC_agelist = unique(DA_lifetbl_female$Age0),
    LC_yearlist = unique(DA_lifetbl_female$Year))
```

```
## IMPLEMENTATION OF THE ORIGINAL AR(1) INFERENCE METHOD

OM_func <- function(OM_n_agegroups, OM_n_periods,
    OM_lifetbl_mx, OM_test.L) {

    # generate table of central mortality rate
    OM_logM <- matrix(data = OM_lifetbl_mx, nrow = OM_n_agegroups,
        ncol = OM_n_periods, byrow = FALSE)

    # dimension of mortality rate data:
    # number of years observed (T, OM_n_periods) X
    # number of age groups (M, OM_n_agegroups)
    OM_logM <- t(log(OM_logM))

    # compute \hat{Z}_t
    OM_Zhat_t <- apply(OM_logM, 1, sum)
    OM_Zhat_t1 <- OM_Zhat_t[-1]
    OM_Zhat_t0 <- OM_Zhat_t[-OM_n_periods]

    OM_Zhat_t.fitting <- lm(OM_Zhat_t1 ~ OM_Zhat_t0)

    # report format: (estimate, standard error)
    OM_mu <- OM_Zhat_t.fitting$coefficients[1]
    names(OM_mu) <- NULL

    OM_mu_se <- coef(summary(OM_Zhat_t.fitting))[, 'Std. Error'][1]
    names(OM_mu_se) <- NULL

    OM_phi <- OM_Zhat_t.fitting$coefficients[2]
    names(OM_phi) <- NULL

    OM_phi_se <- coef(summary(OM_Zhat_t.fitting))[, 'Std. Error'][2]
    names(OM_phi_se) <- NULL

    # define function: OM_estimate.alphabeta.x
    # estimate an element of \alpha_x and \beta_x
    OM_estimate.alphabeta.x <- function(arg1, arg2) {
        # arg1: a column of OM_logM
        # arg2: array, OM_Zhat_t
        argsize <- length(arg1) # dimension of function input parameter
        # the two input arguments are assumed to have the same length, or
        # error will be reported
```

```
    # compute estimates
    a.x <- (sum(arg1) * sum(arg2 ^ 2) -
        sum(arg1 * arg2) * sum(arg2)) /
        (argsize * sum(arg2 ^ 2) - (sum(arg2)) ^ 2)
    b.x <- (argsize * sum(arg1 * arg2) -
        sum(arg1) * sum(arg2)) /
        (argsize * sum(arg2 ^ 2) - (sum(arg2)) ^ 2)

    # return a pair of estimates of \alpha_x and \beta_x
    return(c(a.x, b.x))
} # end function OM_estimate.alphabeta.x

# \alpha_x and \beta_x
OM_axbx <- apply(OM_logM, 2, OM_estimate.alphabeta.x, OM_Zhat_t)

# adjust \mu values for the test
OM_mu.adjusted <- OM_mu - OM_Zhat_t[1] + OM_phi * OM_Zhat_t[1]

# compute \hat{e}_t (t = 2 ... T)
OM_ehat_t <- OM_Zhat_t1 - OM_mu - OM_phi * OM_Zhat_t0
OM_ehat_t1 <- OM_ehat_t[-1]
OM_ehat_t0 <- OM_ehat_t[-(OM_n_periods - 1)]

# compute \hat{U}_i (i = 2 ... T-L+1)
# OM_test.L: the value of L for the Chi-square test
OM_Uhat_i <- apply(embed(OM_ehat_t, OM_test.L), 1, sum) / OM_test.L

# \hat{U}_i (i = 1 ... T-L+1)
OM_Uhat_i <- c(sum(OM_ehat_t[2:OM_test.L]) / OM_test.L, OM_Uhat_i)

# compute \hat{\sigma}_e^2
OM_sigma.hat_e.sq <- OM_test.L * ((sum(OM_Uhat_i ^ 2)) /
    (OM_n_periods - OM_test.L + 1) -
    ((sum(OM_Uhat_i)) / (OM_n_periods - OM_test.L + 1)) ^ 2)

# alternative \hat{\sigma}_e^2
OM_sigma.hat_e.sq_alt <- sum(OM_ehat_t ^ 2) / OM_n_periods +
    2 / OM_n_periods * sum(OM_ehat_t1 * OM_ehat_t0)

# unit-root test: Chi-square test statistic and p-value
OM_unitrootteststatistic <- OM_mu.adjusted ^ 2 * OM_n_periods ^ 3 *
    (OM_phi - 1) ^ 2 / (12 * OM_sigma.hat_e.sq)
```

```
    OM_unitrootteststatistic_alt <- OM_mu.adjusted ^ 2 * OM_n_periods ^ 3 *
        (OM_phi - 1) ^ 2 / (12 * OM_sigma.hat_e.sq_alt)

    OM_unitroottest_p.value <- pchisq(OM_unitrootteststatistic, df = 1,
        lower.tail = FALSE)

    OM_unitroottest_alt_p.value <- pchisq(OM_unitrootteststatistic_alt,
        df = 1, lower.tail = FALSE)

    return(list(OM_axbx[1,], OM_axbx[2,],
        OM_mu, OM_mu_se, OM_phi, OM_phi_se,
        OM_sigma.hat_e.sq, OM_sigma.hat_e.sq_alt,
        OM_unitrootteststatistic, OM_unitrootteststatistic_alt,
        OM_unitroottest_p.value, OM_unitroottest_alt_p.value))

}

# choose the value of L (L = 5, 10, 15) for the Chi-square test
# (during implementation of our method)
DA_OM_test.L <- floor(0.5 * sqrt(DA_n_periods))

DA_OM_both <- OM_func(OM_n_agegroups = DA_n_agegroups,
    OM_n_periods = DA_n_periods,
    OM_lifetbl_mx = DA_lifetbl_both$mx,
    OM_test.L = DA_OM_test.L)
DA_OM_male <- OM_func(OM_n_agegroups = DA_n_agegroups,
    OM_n_periods = DA_n_periods,
    OM_lifetbl_mx = DA_lifetbl_male$mx,
    OM_test.L = DA_OM_test.L)
DA_OM_female <- OM_func(OM_n_agegroups = DA_n_agegroups,
    OM_n_periods = DA_n_periods,
    OM_lifetbl_mx = DA_lifetbl_female$mx,
    OM_test.L = DA_OM_test.L)

## IMPLEMENTATION OF THE BIAS CORRECTED AR(1) INFERENCE METHOD

BC_func <- function(BC_n_agegroups, BC_n_periods, BC_lifetbl_mx) {

    # generate table of central mortality rate
    BC_logM <- matrix(data = BC_lifetbl_mx, nrow = BC_n_agegroups,
        ncol = BC_n_periods, byrow = FALSE)

    # dimension of mortality rate data:
```

```
# number of years observed (T, BC_n_periods) X
# number of age groups (M, BC_n_agegroups)
BC_logM <- t(log(BC_logM))

# compute \hat{Z}_t
BC_Zhat_t <- apply(BC_logM, 1, sum)

# compute sums of (cross products of) \hat{Z}_t
BC_Zhat_t_no.first.1 <- BC_Zhat_t[-1]
BC_Zhat_t_no.first.2 <- BC_Zhat_t[-c(1, 2)]
BC_Zhat_t_no.last.1 <- BC_Zhat_t[-BC_n_periods]
BC_Zhat_t_no.last.2 <- BC_Zhat_t[-c(BC_n_periods - 1, BC_n_periods)]
BC_Zhat_t_no.first.1.last.1 <- BC_Zhat_t[-c(1, BC_n_periods)]

# sum of \hat{Z}_t
BC_Zhat_t_sum_t <- sum(BC_Zhat_t_no.first.2)

# sum of \hat{Z}_{t-1}
BC_Zhat_t_sum_t.1 <- sum(BC_Zhat_t_no.first.1.last.1)

# sum of \hat{Z}_{t-2}
BC_Zhat_t_sum_t.2 <- sum(BC_Zhat_t_no.last.2)

# sum of \hat{Z}_t\hat{Z}_{t-2}
BC_Zhat_t_sum_t_t.2 <- sum(BC_Zhat_t_no.first.2 * BC_Zhat_t_no.last.2)

# sum of \hat{Z}_{t-1}\hat{Z}_{t-2}
BC_Zhat_t_sum_t.1_t.2 <-
    sum(BC_Zhat_t_no.last.2 * BC_Zhat_t_no.first.1.last.1)

# estimator of \mu
BC_muhat <- (BC_Zhat_t_sum_t * BC_Zhat_t_sum_t.1_t.2 -
    BC_Zhat_t_sum_t.1 * BC_Zhat_t_sum_t_t.2) /
    ((BC_n_periods - 2) * BC_Zhat_t_sum_t.1_t.2 -
    BC_Zhat_t_sum_t.1 * BC_Zhat_t_sum_t.2)

# estimator of \phi
BC_phihat <- ((BC_n_periods - 2) * BC_Zhat_t_sum_t_t.2 -
    BC_Zhat_t_sum_t * BC_Zhat_t_sum_t.2) /
    ((BC_n_periods - 2) * BC_Zhat_t_sum_t.1_t.2 -
    BC_Zhat_t_sum_t.1 * BC_Zhat_t_sum_t.2)

# define function: BC_estimate.alphabeta.x
```

```
    # estimate an element of \alpha_x and \beta_x
    BC_estimate.alphabeta.x <- function(arg1, arg2) {
        # arg1: a column of BC_logM
        # arg2: array, BC_Zhat_t
        argsize <- length(arg1) # dimension of function input parameter
        # the two input arguments are assumed to have the same length, or
        # error will be reported

        arg1_no.first.1 <- arg1[-1]
        arg1_no.first.2 <- arg1[-c(1, 2)]
        arg1_no.last.1 <- arg1[-argsize]
        arg2_no.first.1 <- arg2[-1]
        arg2_no.first.2 <- arg2[-c(1, 2)]
        arg2_no.last.1 <- arg2[-argsize]
        arg2_no.first.1.last.1 <- arg2[-c(1, argsize)]

        # compute estimates
        a.x <- (sum(arg1_no.first.2) * sum(arg2_no.first.2 * arg2_no.first.1.last.1) -
            sum(arg2_no.first.2) * sum(arg2_no.first.1.last.1 * arg1_no.first.2)) /
            ((argsize - 2) * sum(arg2_no.first.2 * arg2_no.first.1.last.1) -
            sum(arg2_no.first.2) * sum(arg2_no.first.1.last.1))
        b.x <- ((argsize - 2) * sum(arg2_no.first.1.last.1 * arg1_no.first.2) -
            sum(arg1_no.first.2) * sum(arg2_no.first.1.last.1)) /
            ((argsize - 2) * sum(arg2_no.first.2 * arg2_no.first.1.last.1) -
            sum(arg2_no.first.2) * sum(arg2_no.first.1.last.1))

        # return a pair of estimates of \alpha_x and \beta_x
        return(c(a.x, b.x))
    } # end function BC_estimate.alphabeta.x

    # \alpha_x and \beta_x
    BC_axbx <- apply(BC_logM, 2, BC_estimate.alphabeta.x, BC_Zhat_t)

    return(list(BC_axbx[1, ], BC_axbx[2, ], BC_muhat, BC_phihat))

}

DA_BC_both <- BC_func(BC_n_agegroups = DA_n_agegroups,
    BC_n_periods = DA_n_periods,
    BC_lifetbl_mx = DA_lifetbl_female$mx)
DA_BC_male <- BC_func(BC_n_agegroups = DA_n_agegroups,
    BC_n_periods = DA_n_periods,
    BC_lifetbl_mx = DA_lifetbl_male$mx)
```

```
DA_BC_female <- BC_func(BC_n_agegroups = DA_n_agegroups,
    BC_n_periods = DA_n_periods,
    BC_lifetbl_mx = DA_lifetbl_female$mx)

## IMPLEMENTATION OF THE AR(2) INFERENCE METHOD

A2_func <- function(A2_n_agegroups, A2_n_periods, A2_lifetbl_mx) {

    # generate table of central mortality rate
    A2_logM <- matrix(data = A2_lifetbl_mx, nrow = A2_n_agegroups,
        ncol = A2_n_periods, byrow = FALSE)

    # dimension of mortality rate data:
    # number of years observed (T, A2_n_periods) X
    # number of age groups (M, A2_n_agegroups)
    A2_logM <- t(log(A2_logM))

    # compute \hat{Z}_t
    A2_Zhat_t <- apply(A2_logM, 1, sum)

    A2_Zhat_t_no.first.1 <- A2_Zhat_t[-1]
    A2_Zhat_t_no.last.1 <- A2_Zhat_t[-A2_n_periods]
    A2_Zhat_t_no.first.2 <- A2_Zhat_t[-c(1, 2)]
    A2_Zhat_t_no.first.1.last.1 <- A2_Zhat_t[-c(1, A2_n_periods)]
    A2_Zhat_t_no.last.2 <- A2_Zhat_t[-c(A2_n_periods - 1, A2_n_periods)]
    A2_Zhat_t_no.first.3 <- A2_Zhat_t[-c(1, 2, 3)]
    A2_Zhat_t_no.first.2.last.1 <- A2_Zhat_t[-c(1, 2, A2_n_periods)]
    A2_Zhat_t_no.first.1.last.2 <-
        A2_Zhat_t[-c(1, A2_n_periods - 1, A2_n_periods)]
    A2_Zhat_t_no.last.3 <-
        A2_Zhat_t[-c(A2_n_periods - 2, A2_n_periods - 1, A2_n_periods)]
    A2_Zhat_t_no.first.4 <- A2_Zhat_t[-c(1, 2, 3, 4)]
    A2_Zhat_t_no.first.3.last.1 <- A2_Zhat_t[-c(1, 2, 3, A2_n_periods)]
    A2_Zhat_t_no.first.2.last.2 <-
        A2_Zhat_t[-c(1, 2, A2_n_periods - 1, A2_n_periods)]
    A2_Zhat_t_no.first.1.last.3 <-
        A2_Zhat_t[-c(1, A2_n_periods - 2, A2_n_periods - 1, A2_n_periods)]
    A2_Zhat_t_no.last.4 <-
        A2_Zhat_t[-c(A2_n_periods - 3, A2_n_periods - 2,
        A2_n_periods - 1, A2_n_periods)]

    # AR(2) TIME SERIES: WHEN \phi_2 = 0
```

```
# the 3 x 3 equation system for \mu, \tilde{\phi}_1 and \phi_2
A2_E1_a11 <- A2_n_periods - 3
A2_E1_a12 <- sum(A2_Zhat_t_no.first.2.last.1)
A2_E1_a13 <- A2_Zhat_t[2] - A2_Zhat_t[A2_n_periods - 1]
A2_E1_b1 <- sum(A2_Zhat_t_no.first.3)
A2_E1_a21 <- sum(A2_Zhat_t_no.first.1.last.2)
A2_E1_a22 <- sum(A2_Zhat_t_no.first.1.last.2 *
    A2_Zhat_t_no.first.2.last.1)
A2_E1_a23 <- sum(A2_Zhat_t_no.first.1.last.2 ^ 2) - A2_E1_a22
A2_E1_b2 <- sum(A2_Zhat_t_no.first.3 * A2_Zhat_t_no.first.1.last.2)
A2_E1_a31 <- A2_Zhat_t[A2_n_periods - 2] - A2_Zhat_t[1]
A2_E1_a32 <- A2_E1_a22 -
    sum(A2_Zhat_t_no.first.2.last.1 * A2_Zhat_t_no.last.3)
A2_E1_a33 <- sum(A2_Zhat_t_no.first.1.last.2 ^ 2) -
    sum(A2_Zhat_t_no.first.1.last.2 * A2_Zhat_t_no.last.3) -
    A2_E1_a32
A2_E1_b3 <- A2_E1_b2 -
    sum(A2_Zhat_t_no.first.3 * A2_Zhat_t_no.last.3)

# solving for \mu, \tilde{\phi}_1 and \phi_2
A2_E1_Amatrix <- matrix(
    c(A2_E1_a11, A2_E1_a12, A2_E1_a13, A2_E1_a21, A2_E1_a22,
    A2_E1_a23, A2_E1_a31, A2_E1_a32, A2_E1_a33), 3, 3, byrow = TRUE)
A2_E1_Bvector <- c(A2_E1_b1, A2_E1_b2, A2_E1_b3)
A2_E1_solvector <- solve(A2_E1_Amatrix) %*% A2_E1_Bvector

A2_mu <- A2_E1_solvector[1]
A2_phi1tilde <- A2_E1_solvector[2]
A2_phi2 <- A2_E1_solvector[3]
A2_phi1 <- A2_phi1tilde - A2_phi2

# storage space for \alpha_x and \beta_x where x is age group
A2_axbx <- NULL

# the 2 x 2 equation system for \alpha_x and \beta_x where x is age group
for (i in 1:A2_n_agegroups) {

    # central mortality rate of the specific age group
    A2_logM_x <- A2_logM[, i]

    # compute various sums of log m(x,t) for the specific age group
    A2_logM_x_no.first.3 <- A2_logM_x[-c(1, 2, 3)]
```

```
    # the 2 x 2 equation system for \alpha_x and \beta_x for the specific age group
    A2_E2x_a11 <- A2_n_periods - 3
    A2_E2x_a12 <- sum(A2_Zhat_t_no.first.3)
    A2_E2x_b1 <- sum(A2_logM_x_no.first.3)
    A2_E2x_a21 <- sum(A2_Zhat_t_no.first.2.last.1)
    A2_E2x_a22 <- sum(A2_Zhat_t_no.first.3 * A2_Zhat_t_no.first.2.last.1)
    A2_E2x_b2 <- sum(A2_Zhat_t_no.first.2.last.1 * A2_logM_x_no.first.3)

    # solving for \alpha_x and \beta_x for the specific age group
    A2_E2x_Amatrix <- matrix(
        c(A2_E2x_a11, A2_E2x_a12, A2_E2x_a21, A2_E2x_a22),
        2, 2, byrow = TRUE)
    A2_E2x_Bvector <- c(A2_E2x_b1, A2_E2x_b2)
    A2_E2x_solvector <- solve(A2_E2x_Amatrix) %*% A2_E2x_Bvector
    A2_axbx <- cbind(A2_axbx, A2_E2x_solvector)

} # end the 2 x 2 equation system for \alpha_x and \beta_x

# define \tilde{Y}_{t,x} and \hat{Y}_{t,x} matrices (BASED ON ESTIMATES)
# (number of years observed T - 3, A2_n_periods - 3, 4 ... T) X
# (number of age groups M - 1, A2_n_agegroups - 1, 1 ... M - 1)
A2_Ytilde_est.matrix <-
    A2_logM[4:A2_n_periods, 1:(A2_n_agegroups - 1)] -
    matrix(rep(1, A2_n_periods - 3), ncol = 1) %*%
    matrix(A2_axbx[1, 1:(A2_n_agegroups - 1)], nrow = 1) -
    matrix(A2_Zhat_t_no.first.3, ncol = 1) %*%
    matrix(A2_axbx[2, 1:(A2_n_agegroups - 1)], nrow = 1)
A2_Yhat_est.matrix <-
    diag(A2_Zhat_t_no.first.2.last.1) %*% A2_Ytilde_est.matrix

# define Y_{t,1}, Y_{t,2} and Y_{t,3} vectors (BASED ON ESTIMATES)
# (number of years observed T - 3, A2_n_periods - 3, 4 ... T)
A2_Yt1_est.vector <- A2_Zhat_t_no.first.3 - A2_mu -
    A2_phi1tilde * A2_Zhat_t_no.first.2.last.1 +
    A2_phi2 *
    (A2_Zhat_t_no.first.2.last.1 - A2_Zhat_t_no.first.1.last.2)
A2_Yt2_est.vector <- A2_Yt1_est.vector * A2_Zhat_t_no.first.1.last.2
A2_Yt3_est.vector <- A2_Yt2_est.vector -
    A2_Yt1_est.vector * A2_Zhat_t_no.last.3

# define W_t and \tilde{W}_t matrices (BASED ON ESTIMATES)
# (number of years observed T - 3, A2_n_periods - 3, 4 ... T) X
# (2 x number of age groups M + 1, 2 x A2_n_agegroups + 1)
```

```
A2_Wt_est.matrix <- cbind(A2_Yt1_est.vector,
    A2_Yt2_est.vector / A2_n_periods, A2_Yt3_est.vector)
A2_Wt_tilde_est.matrix <- cbind(A2_Yt1_est.vector,
    A2_Yt2_est.vector, A2_Yt3_est.vector)

for (i in 1:(A2_n_agegroups - 1)) {
    A2_Wt_est.matrix <- cbind(A2_Wt_est.matrix,
        A2_Ytilde_est.matrix[, i],
        A2_Yhat_est.matrix[, i] / A2_n_periods)
    A2_Wt_tilde_est.matrix <- cbind(A2_Wt_tilde_est.matrix,
        A2_Ytilde_est.matrix[, i], A2_Yhat_est.matrix[, i])
}

dimnames(A2_Wt_est.matrix) <- NULL
dimnames(A2_Wt_tilde_est.matrix) <- NULL

# estimate \Sigma and \tilde{\Sigma} matrices (BASED ON ESTIMATES)
# (2 x number of age groups M + 1, 2 x A2_n_agegroups + 1) X
# (2 x number of age groups M + 1, 2 x A2_n_agegroups + 1)
A2_Sigma_est.matrix <- matrix(0, nrow = 2 * A2_n_agegroups + 1,
    ncol = 2 * A2_n_agegroups + 1)
A2_Sigma_tilde_est.matrix <- matrix(0,
    nrow = 2 * A2_n_agegroups + 1, ncol = 2 * A2_n_agegroups + 1)

for (j in 1:(A2_n_periods - 3)) {
    A2_Sigma_est.matrix <- A2_Sigma_est.matrix +
        matrix(A2_Wt_est.matrix[j, ], ncol = 1) %*%
        matrix(A2_Wt_est.matrix[j, ], nrow = 1) / (A2_n_periods - 3)
    A2_Sigma_tilde_est.matrix <- A2_Sigma_tilde_est.matrix +
        matrix(A2_Wt_tilde_est.matrix[j, ], ncol = 1) %*%
        matrix(A2_Wt_tilde_est.matrix[j, ], nrow = 1) /
        (A2_n_periods - 3)
}
for (j in 2:(A2_n_periods - 3)) {
    A2_Sigma_est.matrix <- A2_Sigma_est.matrix +
        matrix(A2_Wt_est.matrix[j, ], ncol = 1) %*%
        matrix(A2_Wt_est.matrix[j - 1, ], nrow = 1) /
        (A2_n_periods - 4)
    A2_Sigma_tilde_est.matrix <- A2_Sigma_tilde_est.matrix +
        matrix(A2_Wt_tilde_est.matrix[j, ], ncol = 1) %*%
        matrix(A2_Wt_tilde_est.matrix[j - 1, ], nrow = 1) /
        (A2_n_periods - 4)
}
```

```
for (j in 1:(A2_n_periods - 4)) {
    A2_Sigma_est.matrix <- A2_Sigma_est.matrix +
        matrix(A2_Wt_est.matrix[j, ], ncol = 1) %*%
        matrix(A2_Wt_est.matrix[j + 1, ], nrow = 1) /
        (A2_n_periods - 4)
    A2_Sigma_tilde_est.matrix <- A2_Sigma_tilde_est.matrix +
        matrix(A2_Wt_tilde_est.matrix[j, ], ncol = 1) %*%
        matrix(A2_Wt_tilde_est.matrix[j + 1, ], nrow = 1) /
        (A2_n_periods - 4)
}

# diagonal matrix D_T
# (2 x number of age groups M + 1, 2 x A2_n_agegroups + 1) X
# (2 x number of age groups M + 1, 2 x A2_n_agegroups + 1)
A2_DT_diag <- c(sqrt(A2_n_periods), A2_n_periods ^ (3 / 2),
    sqrt(A2_n_periods))
for (i in 1:(A2_n_agegroups - 1)) {
    A2_DT_diag <-
        c(A2_DT_diag, sqrt(A2_n_periods), A2_n_periods ^ (3 / 2))
}
A2_DT_matrix <- diag(A2_DT_diag)

# estimate \Gamma_1 and \tilde{\Gamma}_1 matrices (BASED ON ESTIMATES)
# (2 x number of age groups M + 1, 2 x A2_n_agegroups + 1) X
# (2 x number of age groups M + 1, 2 x A2_n_agegroups + 1)
A2_Gamma1_est.matrix <- matrix(0,
    nrow = 2 * A2_n_agegroups + 1, ncol = 2 * A2_n_agegroups + 1)
A2_Gamma1_tilde_est.matrix <- matrix(0,
    nrow = 2 * A2_n_agegroups + 1, ncol = 2 * A2_n_agegroups + 1)

for (i in 4:A2_n_periods) { # sum of matrix differentiations

    A2_Gamma1_t <- matrix(0,
        nrow = 2 * A2_n_agegroups + 1, ncol = 2 * A2_n_agegroups + 1)
    A2_Gamma1_tilde_t <- matrix(0,
        nrow = 2 * A2_n_agegroups + 1, ncol = 2 * A2_n_agegroups + 1)

    # matrix differentiation by blocks
    # top left 3 X 3 block
    A2_Gamma1_t[1, 1] <- -1
    A2_Gamma1_t[1, 2] <- -A2_Zhat_t[i - 1]
    A2_Gamma1_t[1, 3] <- A2_Zhat_t[i - 1] - A2_Zhat_t[i - 2]
    A2_Gamma1_t[2, 1] <- -A2_Zhat_t[i - 2] / A2_n_periods
```

```
A2_Gamma1_t[2, 2] <- -A2_Zhat_t[i - 1] * A2_Zhat_t[i - 2] / A2_n_periods
A2_Gamma1_t[2, 3] <- A2_Zhat_t[i - 2] *
    (A2_Zhat_t[i - 1] - A2_Zhat_t[i - 2]) / A2_n_periods
A2_Gamma1_t[3, 1] <- A2_Zhat_t[i - 3] - A2_Zhat_t[i - 2]
A2_Gamma1_t[3, 2] <- A2_Zhat_t[i - 1] *
    (A2_Zhat_t[i - 3] - A2_Zhat_t[i - 2])
A2_Gamma1_t[3, 3] <- (A2_Zhat_t[i - 1] - A2_Zhat_t[i - 2]) *
    (A2_Zhat_t[i - 2] - A2_Zhat_t[i - 3])

A2_Gamma1_tilde_t[1, 1] <- -1
A2_Gamma1_tilde_t[1, 2] <- -A2_Zhat_t[i - 1]
A2_Gamma1_tilde_t[1, 3] <- A2_Zhat_t[i - 1] - A2_Zhat_t[i - 2]
A2_Gamma1_tilde_t[2, 1] <- -A2_Zhat_t[i - 2]
A2_Gamma1_tilde_t[2, 2] <- -A2_Zhat_t[i - 1] * A2_Zhat_t[i - 2]
A2_Gamma1_tilde_t[2, 3] <- A2_Zhat_t[i - 2] *
    (A2_Zhat_t[i - 1] - A2_Zhat_t[i - 2])
A2_Gamma1_tilde_t[3, 1] <- A2_Zhat_t[i - 3] - A2_Zhat_t[i - 2]
A2_Gamma1_tilde_t[3, 2] <- A2_Zhat_t[i - 1] *
    (A2_Zhat_t[i - 3] - A2_Zhat_t[i - 2])
A2_Gamma1_tilde_t[3, 3] <- (A2_Zhat_t[i - 1] - A2_Zhat_t[i - 2]) *
    (A2_Zhat_t[i - 2] - A2_Zhat_t[i - 3])

# top right 3 X { (A2_n_agegroups - 1) x 2 } block
# all zero

# bottom left { (A2_n_agegroups - 1) x 2 } X 3 block
# all zero

# bottom right { (A2_n_agegroups - 1) x 2 } X { (A2_n_agegroups - 1) x 2 } block
# 2 x 2 block diagonal matrices
for (j in 1:(A2_n_agegroups - 1)) {
    A2_Gamma1_t[j * 2 + 2, j * 2 + 2] <- -1
    A2_Gamma1_t[j * 2 + 2, j * 2 + 3] <- -A2_Zhat_t[i]
    A2_Gamma1_t[j * 2 + 3, j * 2 + 2] <-
        -A2_Zhat_t[i - 1] / A2_n_periods
    A2_Gamma1_t[j * 2 + 3, j * 2 + 3] <-
        -A2_Zhat_t[i] * A2_Zhat_t[i - 1] / A2_n_periods

    A2_Gamma1_tilde_t[j * 2 + 2, j * 2 + 2] <- -1
    A2_Gamma1_tilde_t[j * 2 + 2, j * 2 + 3] <- -A2_Zhat_t[i]
    A2_Gamma1_tilde_t[j * 2 + 3, j * 2 + 2] <- -A2_Zhat_t[i - 1]
    A2_Gamma1_tilde_t[j * 2 + 3, j * 2 + 3] <-
        -A2_Zhat_t[i] * A2_Zhat_t[i - 1]
```

```
    }

    A2_Gamma1_est.matrix <- A2_Gamma1_est.matrix + A2_Gamma1_t
    A2_Gamma1_tilde_est.matrix <-
        A2_Gamma1_tilde_est.matrix + A2_Gamma1_tilde_t

} # end sum of matrix differentiations

A2_Gamma1_est.matrix <-
    (A2_Gamma1_est.matrix / (A2_n_periods - 3)) %*%
    (sqrt(A2_n_periods) * solve(A2_DT_matrix))
A2_Gamma1_tilde_est.matrix <-
    A2_Gamma1_tilde_est.matrix / (A2_n_periods - 3)

# Theorem 1 asymptotic variance matrix
A2_Theorem1_avar_est.matrix <- solve(A2_Gamma1_est.matrix) %*%
    A2_Sigma_est.matrix %*% t(solve(A2_Gamma1_est.matrix))
A2_Theorem1_avar_tilde_est.matrix <-
    solve(A2_Gamma1_tilde_est.matrix) %*% A2_Sigma_tilde_est.matrix %*%
    t(solve(A2_Gamma1_tilde_est.matrix))

# Theorem 1 hypothesis tests

# testing H_0: \tilde{\phi}_1 = 1 & \phi_2 = 0 based on Theorem 1i
A2_Theorem1i_chisq2 <- as.numeric(
    matrix(c(A2_DT_diag[2] * (A2_phi1tilde - 1),
    A2_DT_diag[3] * A2_phi2), nrow = 1) %*%
    solve(A2_Theorem1_avar_est.matrix[2:3, 2:3]) %*%
    matrix(c(A2_DT_diag[2] * (A2_phi1tilde - 1), A2_DT_diag[3] * A2_phi2),
    ncol = 1))

# testing H_0: \phi_2 = 0 based on Theorem 1ii
A2_Theorem1ii_chisq1 <- (A2_DT_diag[3] * A2_phi2) ^ 2 /
    A2_Theorem1_avar_tilde_est.matrix[3, 3]

# AR(2) TIME SERIES: WHEN \phi_2 != 0

# the 3 x 3 equation system for \mu^*, \tilde{\phi}_1^* and \phi_2^*
A2_E3_a11 <- A2_n_periods - 4
A2_E3_a12 <- sum(A2_Zhat_t_no.first.3.last.1)
A2_E3_a13 <- A2_Zhat_t[3] - A2_Zhat_t[A2_n_periods - 1]
A2_E3_b1 <- sum(A2_Zhat_t_no.first.4)
A2_E3_a21 <- sum(A2_Zhat_t_no.first.1.last.3)
```

```
A2_E3_a22 <- sum(A2_Zhat_t_no.first.3.last.1 *
    A2_Zhat_t_no.first.1.last.3)
A2_E3_a23 <- sum(A2_Zhat_t_no.first.1.last.3 *
    (A2_Zhat_t_no.first.2.last.2 - A2_Zhat_t_no.first.3.last.1))
A2_E3_b2 <- sum(A2_Zhat_t_no.first.4 * A2_Zhat_t_no.first.1.last.3)
A2_E3_a31 <- A2_Zhat_t[A2_n_periods - 3] - A2_Zhat_t[1]
A2_E3_a32 <- sum(A2_Zhat_t_no.first.3.last.1 *
    (A2_Zhat_t_no.first.1.last.3 - A2_Zhat_t_no.last.4))
A2_E3_a33 <-
    sum((A2_Zhat_t_no.first.2.last.2 - A2_Zhat_t_no.first.3.last.1) *
    (A2_Zhat_t_no.first.1.last.3 - A2_Zhat_t_no.last.4))
A2_E3_b3 <- sum(A2_Zhat_t_no.first.4 *
    (A2_Zhat_t_no.first.1.last.3 - A2_Zhat_t_no.last.4))

# solving for \mu^*, \tilde{\phi}_1^* and \phi_2^*
A2_E3_Amatrix <- matrix(
    c(A2_E3_a11, A2_E3_a12, A2_E3_a13, A2_E3_a21, A2_E3_a22,
    A2_E3_a23, A2_E3_a31, A2_E3_a32, A2_E3_a33), 3, 3, byrow = TRUE)
A2_E3_Bvector <- c(A2_E3_b1, A2_E3_b2, A2_E3_b3)
A2_E3_solvector <- solve(A2_E3_Amatrix) %*% A2_E3_Bvector

A2_mu.s <- A2_E3_solvector[1]
A2_phi1tilde.s <- A2_E3_solvector[2]
A2_phi2.s <- A2_E3_solvector[3]
A2_phi1.s <- A2_phi1tilde.s - A2_phi2.s

# storage space for \alpha_x^* and \beta_x^* where x is age group
A2_axbx.s <- NULL

# the 2 x 2 equation system for \alpha_x^* and \beta_x^* where x is age group
for (i in 1:A2_n_agegroups) {

    # central mortality rate of the specific age group
    A2_logM_x <- A2_logM[, i]

    # compute various sums of log m(x,t) for the specific age group
    A2_logM_x_no.first.4 <- A2_logM_x[-c(1, 2, 3, 4)]

    # the 2 x 2 equation system for \alpha_x^* and \beta_x^* for the specific age group
    A2_E4x_a11 <- A2_n_periods - 4
    A2_E4x_a12 <- sum(A2_Zhat_t_no.first.4)
    A2_E4x_b1 <- sum(A2_logM_x_no.first.4)
    A2_E4x_a21 <- sum(A2_Zhat_t_no.first.3.last.1)
```

```
    A2_E4x_a22 <- sum(A2_Zhat_t_no.first.4 * A2_Zhat_t_no.first.3.last.1)
    A2_E4x_b2 <- sum(A2_Zhat_t_no.first.3.last.1 * A2_logM_x_no.first.4)

    # solving for \alpha_x^* and \beta_x^* for the specific age group
    A2_E4x_Amatrix <- matrix(
        c(A2_E4x_a11, A2_E4x_a12, A2_E4x_a21, A2_E4x_a22),
        2, 2, byrow = TRUE)
    A2_E4x_Bvector <- c(A2_E4x_b1, A2_E4x_b2)
    A2_E4x_solvector <- solve(A2_E4x_Amatrix) %*% A2_E4x_Bvector
    A2_axbx.s <- cbind(A2_axbx.s, A2_E4x_solvector)

} # end the 2 x 2 equation system for \alpha_x^* and \beta_x^*

# define \tilde{Y}_{t,x}^* and \hat{Y}_{t,x}^* matrices (BASED ON ESTIMATES)
# (number of years observed T - 4, A2_n_periods - 4, 5 ... T) X
# (number of age groups M - 1, A2_n_agegroups - 1, 1 ... M - 1)
A2_Ytilde.s_est.matrix <-
    A2_logM[5:A2_n_periods, 1:(A2_n_agegroups - 1)] -
    matrix(rep(1, A2_n_periods - 4), ncol = 1) %*%
    matrix(A2_axbx.s[1, 1:(A2_n_agegroups - 1)], nrow = 1) -
    matrix(A2_Zhat_t_no.first.4, ncol = 1) %*%
    matrix(A2_axbx.s[2, 1:(A2_n_agegroups - 1)], nrow = 1)
A2_Yhat.s_est.matrix <-
    diag(A2_Zhat_t_no.first.3.last.1) %*% A2_Ytilde.s_est.matrix

# define Y_{t,1}^*, Y_{t,2}^* and Y_{t,3}^* vectors (BASED ON ESTIMATES)
# (number of years observed T - 4, A2_n_periods - 4, 5 ... T)
A2_Yt1.s_est.vector <- A2_Zhat_t_no.first.4 - A2_mu.s -
    A2_phi1tilde.s * A2_Zhat_t_no.first.3.last.1 +
    A2_phi2.s *
    (A2_Zhat_t_no.first.3.last.1 - A2_Zhat_t_no.first.2.last.2)
A2_Yt2.s_est.vector <- A2_Yt1.s_est.vector * A2_Zhat_t_no.first.1.last.3
A2_Yt3.s_est.vector <- A2_Yt2.s_est.vector -
    A2_Yt1.s_est.vector * A2_Zhat_t_no.last.4

# define W_t^* and \tilde{W}_t^* matrices (BASED ON ESTIMATES)
# (number of years observed T - 4, A2_n_periods - 4, 5 ... T) X
# (2 x number of age groups M + 1, 2 x A2_n_agegroups + 1)
A2_Wt.s_est.matrix <- cbind(A2_Yt1.s_est.vector,
    A2_Yt2.s_est.vector / A2_n_periods, A2_Yt3.s_est.vector)
A2_Wt.s_tilde_est.matrix <- cbind(A2_Yt1.s_est.vector,
    A2_Yt2.s_est.vector, A2_Yt3.s_est.vector)
```

```
for (i in 1:(A2_n_agegroups - 1)) {
    A2_Wt.s_est.matrix <- cbind(A2_Wt.s_est.matrix,
        A2_Ytilde.s_est.matrix[, i],
        A2_Yhat.s_est.matrix[, i] / A2_n_periods)
    A2_Wt.s_tilde_est.matrix <- cbind(A2_Wt.s_tilde_est.matrix,
        A2_Ytilde.s_est.matrix[, i], A2_Yhat.s_est.matrix[, i])
}

dimnames(A2_Wt.s_est.matrix) <- NULL
dimnames(A2_Wt.s_tilde_est.matrix) <- NULL

# estimate \Sigma^* and \tilde{\Sigma}^* matrices (BASED ON ESTIMATES)
# (2 x number of age groups M + 1, 2 x A2_n_agegroups + 1) X
# (2 x number of age groups M + 1, 2 x A2_n_agegroups + 1)
A2_Sigma.s_est.matrix <- matrix(0, nrow = 2 * A2_n_agegroups + 1,
    ncol = 2 * A2_n_agegroups + 1)
A2_Sigma.s_tilde_est.matrix <- matrix(0,
    nrow = 2 * A2_n_agegroups + 1, ncol = 2 * A2_n_agegroups + 1)

for (j in 1:(A2_n_periods - 4)) {
    A2_Sigma.s_est.matrix <- A2_Sigma.s_est.matrix +
        matrix(A2_Wt.s_est.matrix[j, ], ncol = 1) %*%
        matrix(A2_Wt.s_est.matrix[j, ], nrow = 1) / (A2_n_periods - 4)
    A2_Sigma.s_tilde_est.matrix <- A2_Sigma.s_tilde_est.matrix +
        matrix(A2_Wt.s_tilde_est.matrix[j, ], ncol = 1) %*%
        matrix(A2_Wt.s_tilde_est.matrix[j, ], nrow = 1) /
        (A2_n_periods - 4)
}
for (j in 2:(A2_n_periods - 4)) {
    A2_Sigma.s_est.matrix <- A2_Sigma.s_est.matrix +
        matrix(A2_Wt.s_est.matrix[j, ], ncol = 1) %*%
        matrix(A2_Wt.s_est.matrix[j - 1, ], nrow = 1) /
        (A2_n_periods - 5)
    A2_Sigma.s_tilde_est.matrix <- A2_Sigma.s_tilde_est.matrix +
        matrix(A2_Wt.s_tilde_est.matrix[j, ], ncol = 1) %*%
        matrix(A2_Wt.s_tilde_est.matrix[j - 1, ], nrow = 1) /
        (A2_n_periods - 5)
}
for (j in 3:(A2_n_periods - 4)) {
    A2_Sigma.s_est.matrix <- A2_Sigma.s_est.matrix +
        matrix(A2_Wt.s_est.matrix[j, ], ncol = 1) %*%
        matrix(A2_Wt.s_est.matrix[j - 2, ], nrow = 1) /
        (A2_n_periods - 6)
```

```
    A2_Sigma.s_tilde_est.matrix <- A2_Sigma.s_tilde_est.matrix +
        matrix(A2_Wt.s_tilde_est.matrix[j, ], ncol = 1) %*%
        matrix(A2_Wt.s_tilde_est.matrix[j - 2, ], nrow = 1) /
        (A2_n_periods - 6)
}
for (j in 1:(A2_n_periods - 5)) {
    A2_Sigma.s_est.matrix <- A2_Sigma.s_est.matrix +
        matrix(A2_Wt.s_est.matrix[j, ], ncol = 1) %*%
        matrix(A2_Wt.s_est.matrix[j + 1, ], nrow = 1) /
        (A2_n_periods - 5)
    A2_Sigma.s_tilde_est.matrix <- A2_Sigma.s_tilde_est.matrix +
        matrix(A2_Wt.s_tilde_est.matrix[j, ], ncol = 1) %*%
        matrix(A2_Wt.s_tilde_est.matrix[j + 1, ], nrow = 1) /
        (A2_n_periods - 5)
}
for (j in 1:(A2_n_periods - 6)) {
    A2_Sigma.s_est.matrix <- A2_Sigma.s_est.matrix +
        matrix(A2_Wt.s_est.matrix[j, ], ncol = 1) %*%
        matrix(A2_Wt.s_est.matrix[j + 2, ], nrow = 1) /
        (A2_n_periods - 6)
    A2_Sigma.s_tilde_est.matrix <- A2_Sigma.s_tilde_est.matrix +
        matrix(A2_Wt.s_tilde_est.matrix[j, ], ncol = 1) %*%
        matrix(A2_Wt.s_tilde_est.matrix[j + 2, ], nrow = 1) /
        (A2_n_periods - 6)
}

# estimate \Gamma_2 and \tilde{\Gamma}_2 matrices (BASED ON ESTIMATES)
# (2 x number of age groups M + 1, 2 x A2_n_agegroups + 1) X
# (2 x number of age groups M + 1, 2 x A2_n_agegroups + 1)
A2_Gamma2_est.matrix <- matrix(0,
    nrow = 2 * A2_n_agegroups + 1, ncol = 2 * A2_n_agegroups + 1)
A2_Gamma2_tilde_est.matrix <- matrix(0,
    nrow = 2 * A2_n_agegroups + 1, ncol = 2 * A2_n_agegroups + 1)

for (i in 5:A2_n_periods) { # sum of matrix differentiations

    A2_Gamma2_t <- matrix(0,
        nrow = 2 * A2_n_agegroups + 1, ncol = 2 * A2_n_agegroups + 1)
    A2_Gamma2_tilde_t <- matrix(0,
        nrow = 2 * A2_n_agegroups + 1, ncol = 2 * A2_n_agegroups + 1)

    # matrix differentiation by blocks
    # top left 3 X 3 block
```

```
A2_Gamma2_t[1, 1] <- -1
A2_Gamma2_t[1, 2] <- -A2_Zhat_t[i - 1]
A2_Gamma2_t[1, 3] <- A2_Zhat_t[i - 1] - A2_Zhat_t[i - 2]
A2_Gamma2_t[2, 1] <- -A2_Zhat_t[i - 3] / A2_n_periods
A2_Gamma2_t[2, 2] <- -A2_Zhat_t[i - 1] * A2_Zhat_t[i - 3] / A2_n_periods
A2_Gamma2_t[2, 3] <- A2_Zhat_t[i - 3] *
    (A2_Zhat_t[i - 1] - A2_Zhat_t[i - 2]) / A2_n_periods
A2_Gamma2_t[3, 1] <- A2_Zhat_t[i - 4] - A2_Zhat_t[i - 3]
A2_Gamma2_t[3, 2] <- A2_Zhat_t[i - 1] *
    (A2_Zhat_t[i - 4] - A2_Zhat_t[i - 3])
A2_Gamma2_t[3, 3] <- (A2_Zhat_t[i - 1] - A2_Zhat_t[i - 2]) *
    (A2_Zhat_t[i - 3] - A2_Zhat_t[i - 4])

A2_Gamma2_tilde_t[1, 1] <- -1
A2_Gamma2_tilde_t[1, 2] <- -A2_Zhat_t[i - 1]
A2_Gamma2_tilde_t[1, 3] <- A2_Zhat_t[i - 1] - A2_Zhat_t[i - 2]
A2_Gamma2_tilde_t[2, 1] <- -A2_Zhat_t[i - 3]
A2_Gamma2_tilde_t[2, 2] <- -A2_Zhat_t[i - 1] * A2_Zhat_t[i - 3]
A2_Gamma2_tilde_t[2, 3] <- A2_Zhat_t[i - 3] *
    (A2_Zhat_t[i - 1] - A2_Zhat_t[i - 2])
A2_Gamma2_tilde_t[3, 1] <- A2_Zhat_t[i - 4] - A2_Zhat_t[i - 3]
A2_Gamma2_tilde_t[3, 2] <- A2_Zhat_t[i - 1] *
    (A2_Zhat_t[i - 4] - A2_Zhat_t[i - 3])
A2_Gamma2_tilde_t[3, 3] <- (A2_Zhat_t[i - 1] - A2_Zhat_t[i - 2]) *
    (A2_Zhat_t[i - 3] - A2_Zhat_t[i - 4])

# top right 3 X { (A2_n_agegroups - 1) x 2 } block
# all zero

# bottom left { (A2_n_agegroups - 1) x 2 } X 3 block
# all zero

# bottom right { (A2_n_agegroups - 1) x 2 } X { (A2_n_agegroups - 1) x 2 } block
# 2 x 2 block diagonal matrices
for (j in 1:(A2_n_agegroups - 1)) {
    A2_Gamma2_t[j * 2 + 2, j * 2 + 2] <- -1
    A2_Gamma2_t[j * 2 + 2, j * 2 + 3] <- -A2_Zhat_t[i]
    A2_Gamma2_t[j * 2 + 3, j * 2 + 2] <-
        -A2_Zhat_t[i - 1] / A2_n_periods
    A2_Gamma2_t[j * 2 + 3, j * 2 + 3] <-
        -A2_Zhat_t[i] * A2_Zhat_t[i - 1] / A2_n_periods

    A2_Gamma2_tilde_t[j * 2 + 2, j * 2 + 2] <- -1
```

```
            A2_Gamma2_tilde_t[j * 2 + 2, j * 2 + 3] <- -A2_Zhat_t[i]
            A2_Gamma2_tilde_t[j * 2 + 3, j * 2 + 2] <- -A2_Zhat_t[i - 1]
            A2_Gamma2_tilde_t[j * 2 + 3, j * 2 + 3] <-
                -A2_Zhat_t[i] * A2_Zhat_t[i - 1]
        }

        A2_Gamma2_est.matrix <- A2_Gamma2_est.matrix + A2_Gamma2_t
        A2_Gamma2_tilde_est.matrix <-
            A2_Gamma2_tilde_est.matrix + A2_Gamma2_tilde_t

    } # end sum of matrix differentiations

    A2_Gamma2_est.matrix <- (A2_Gamma2_est.matrix / (A2_n_periods - 4)) %*%
        (sqrt(A2_n_periods) * solve(A2_DT_matrix))
    A2_Gamma2_tilde_est.matrix <-
        A2_Gamma2_tilde_est.matrix / (A2_n_periods - 4)

    # Theorem 3 asymptotic variance matrix
    A2_Theorem3_avar_est.matrix <- solve(A2_Gamma2_est.matrix) %*%
        A2_Sigma.s_est.matrix %*% t(solve(A2_Gamma2_est.matrix))
    A2_Theorem3_avar_tilde_est.matrix <-
        solve(A2_Gamma2_tilde_est.matrix) %*% A2_Sigma.s_tilde_est.matrix %*%
        t(solve(A2_Gamma2_tilde_est.matrix))

    # Theorem 3 hypothesis tests

    # no tests based on Theorem 3

    return(list(A2_axbx[1, ], A2_axbx[2, ],
        A2_mu, A2_phi1tilde, A2_phi1, A2_phi2,
        A2_Theorem1i_chisq2, A2_Theorem1ii_chisq1,
        A2_axbx.s[1, ], A2_axbx.s[2, ],
        A2_mu.s, A2_phi1tilde.s, A2_phi1.s, A2_phi2.s,
        A2_Yt1.s_est.vector))

}

DA_A2_both <- A2_func(A2_n_agegroups = DA_n_agegroups,
    A2_n_periods = DA_n_periods,
    A2_lifetbl_mx = DA_lifetbl_both$mx)
DA_A2_male <- A2_func(A2_n_agegroups = DA_n_agegroups,
    A2_n_periods = DA_n_periods,
    A2_lifetbl_mx = DA_lifetbl_male$mx)
```

```
DA_A2_female <- A2_func(A2_n_agegroups = DA_n_agegroups,
    A2_n_periods = DA_n_periods,
    A2_lifetbl_mx = DA_lifetbl_female$mx)

## CODE FOR DATA SIMULATION IN PAPER

SM_func <- function(SM_loopindex, SM_n_agegroups, SM_n_periods,
    SM_ax, SM_bx, SM_mu, SM_phi1, SM_phi2,
    SM_sd_e_t, SM_sd_epsilon, SM_agelist, SM_OM_test.L){
    # begin simulation loop

    # data generating process
    # below: data generating process is based on input parameters after the
    #        simulation loop
    #        check input parameters

    # k_t in simulation (a new k_t series for each simulation loop)
    # assume k_{-1} = k_0 = 0 and by induction

    SM_k_t <- SM_mu + rnorm(1, mean = 0, sd = 1) * SM_sd_e_t
    SM_k_t <- c(SM_k_t,
        SM_mu + SM_phi1 * SM_k_t + rnorm(1, mean = 0, sd = 1) * SM_sd_e_t)
    for (i in 3:SM_n_periods) {
        SM_k_t <- c(SM_k_t,
            SM_mu + SM_phi1 * SM_k_t[i - 1] + SM_phi2 * SM_k_t[i - 2] +
            rnorm(1, mean = 0, sd = 1) * SM_sd_e_t)
    }
    names(SM_k_t) <- NULL

    # mortality rate data 'logM' in simulation
    # the dimension of SM_logM is: SM_n_periods X SM_n_agegroups (M)
    SM_logM <- t(cbind(SM_ax, SM_bx) %*% as.matrix(rbind(1, SM_k_t)) +
        matrix(rnorm(SM_n_agegroups * SM_n_periods, mean = 0, sd = 1) * SM_sd_epsilon,
        nrow = SM_n_agegroups))

    # recover the original mx array
    # this is reverse process of deriving logM from life table's 'mx' variable
    SM_mx <- array(exp(t(SM_logM)))

    # implementation of the original lee-carter model
    # include LC_func here ...

    SM_LC <- LC_func(LC_n_agegroups = SM_n_agegroups,
```

```
        LC_n_periods = SM_n_periods,
        LC_lifetbl_mx = SM_mx,
        LC_agelist = SM_agelist,
        LC_yearlist = 1:SM_n_periods)

    # implementation of the original ar(1) inference method
    # include OM_func here ...

    SM_OM <- OM_func(OM_n_agegroups = SM_n_agegroups,
        OM_n_periods = SM_n_periods,
        OM_lifetbl_mx = SM_mx,
        OM_test.L = SM_OM_test.L)

    # implementation of the bias corrected ar(1) inference method
    # include BC_func here ...

    SM_BC <- BC_func(BC_n_agegroups = SM_n_agegroups,
        BC_n_periods = SM_n_periods,
        BC_lifetbl_mx = SM_mx)

    # implementation of the ar(2) inference method
    # include A2_func here ...

    SM_A2 <- A2_func(A2_n_agegroups = SM_n_agegroups,
        A2_n_periods = SM_n_periods,
        A2_lifetbl_mx = SM_mx)

    return(list(SM_LC[[1]], SM_LC[[2]], SM_LC[[3]], SM_LC[[4]],
        SM_LC[[5]], SM_LC[[6]], SM_LC[[7]], SM_LC[[8]],
        SM_OM[[1]], SM_OM[[2]], SM_OM[[3]], SM_OM[[4]],
        SM_OM[[5]], SM_OM[[6]], SM_OM[[7]], SM_OM[[8]],
        SM_OM[[9]], SM_OM[[10]], SM_OM[[11]], SM_OM[[12]],
        SM_BC[[1]], SM_BC[[2]], SM_BC[[3]], SM_BC[[4]],
        SM_A2[[1]], SM_A2[[2]], SM_A2[[3]], SM_A2[[4]],
        SM_A2[[5]], SM_A2[[6]], SM_A2[[7]], SM_A2[[8]],
        SM_A2[[9]], SM_A2[[10]], SM_A2[[11]], SM_A2[[12]],
        SM_A2[[13]], SM_A2[[14]]))

} # end simulation loop

# number of simulation loops
SM_loopindex <- 1:10000
```

```
# number of simulated age groups
SM_n_agegroups <- DA_n_agegroups

# number of simulated periods (80, 150, 500, 1000)
SM_n_periods <- 300

# ar(2) model parameters for simulation
SM_ax <- DA_A2_male[[1]]
SM_bx <- DA_A2_male[[2]]
SM_mu <- DA_A2_male[[3]]
SM_phi1 <- DA_A2_male[[5]]+0.4
SM_phi2 <- DA_A2_male[[6]]-0.4
SM_sd_e_t <- 0.01#sd(DA_A2_male[[15]]) / sqrt(DA_n_agegroups)
SM_sd_epsilon <-SM_sd_e_t# sd(DA_A2_male[[15]]) / sqrt(DA_n_agegroups)
SM_agelist <- unique(DA_lifetbl_both$Age0)
SM_OM_test.L <- floor(2 * sqrt(SM_n_periods))

# prepare parallel clusters for simulation
library(snowfall)
library(rlecuyer)
set.seed(12345)
sfInit(parallel = FALSE)
#sfInit(parallel = TRUE, cpus = 8, type = 'SOCK')
#i <- sfClusterSetupRNGstream(12345)

# execute simulation loops
SM_result <- sfLapply(SM_loopindex, SM_func, SM_n_agegroups, SM_n_periods,
    SM_ax, SM_bx, SM_mu, SM_phi1, SM_phi2, SM_sd_e_t, SM_sd_epsilon,
    SM_agelist, SM_OM_test.L)
sfStop()

# end
```

For two-population (bivariate) datasets (in Part 5):

```
## DATA IMPORT

# USA data, 5 x 1, male and female cohort data
# http://www.mortality.org/cgi-bin/hmd/country.php?cntr=USA&level=1

# load raw data files
# to download data other than 1x1, do not use the 'demography' package functions
```

```
DA_lifetbl_male <- read.table('mltper_5x1.txt',
    skip = 2, header = TRUE, stringsAsFactors = FALSE)
DA_lifetbl_female <- read.table('fltper_5x1.txt',
    skip = 2, header = TRUE, stringsAsFactors = FALSE)

## DATA CLEANUP

# same as data cleanup code in previous section ...

## IMPLEMENTATION OF THE ORIGINAL LEE-CARTER MODEL

# same as lee-carter model implementation in previous section...

DA_LC_male <- LC_func(LC_n_agegroups = DA_n_agegroups,
    LC_n_periods = DA_n_periods,
    LC_lifetbl_mx = DA_lifetbl_male$mx,
    LC_agelist = unique(DA_lifetbl_male$Age0),
    LC_yearlist = unique(DA_lifetbl_male$Year))
DA_LC_female <- LC_func(LC_n_agegroups = DA_n_agegroups,
    LC_n_periods = DA_n_periods,
    LC_lifetbl_mx = DA_lifetbl_female$mx,
    LC_agelist = unique(DA_lifetbl_female$Age0),
    LC_yearlist = unique(DA_lifetbl_female$Year))

LC_diff_func <- function(diff_n_agegroups, diff_n_periods,
    diff_var1, diff_var2) {

    # difference of k_t's of the two cohorts
    diff_kt <- diff_var1[[3]] - diff_var2[[3]]

    # below: fitting k_t with (5) in Draft #4
    diff_kt1 <- diff_kt[-1]
    diff_kt0 <- diff_kt[-diff_n_periods]
    diff_kt.fitting <- lm(diff_kt1 ~ diff_kt0)

    # report format: (estimate, standard error)
    diff_mu <- diff_kt.fitting$coefficients[1]
    names(diff_mu) <- NULL

    diff_mu_se <- coef(summary(diff_kt.fitting))[, 'Std. Error'][1]
    names(diff_mu_se) <- NULL
```

```
    diff_phi <- diff_kt.fitting$coefficients[2]
    names(diff_phi) <- NULL

    diff_phi_se <- coef(summary(diff_kt.fitting))[, 'Std. Error'][2]
    names(diff_phi_se) <- NULL

    return(list(rep(NA, diff_n_agegroups), rep(NA, diff_n_agegroups),
        diff_kt, diff_mu, diff_mu_se, diff_phi, diff_phi_se, NA))

}

DA_LC_diff <- LC_diff_func(diff_n_agegroups = DA_n_agegroups,
    diff_n_periods = DA_n_periods, diff_var1 = DA_LC_male,
    diff_var2 = DA_LC_female)

## IMPLEMENTATION OF THE BIVARIATE LEE-CARTER MORTALITY MODEL

BI_func <- function(BI_n_agegroups, BI_n_periods,
    BI_lifet1_mx, BI_lifet2_mx) {

    # generate two tables of central mortality rate
    BI_logM.1 <- matrix(data = BI_lifet1_mx, nrow = BI_n_agegroups,
        ncol = BI_n_periods, byrow = FALSE)
    BI_logM.2 <- matrix(data = BI_lifet2_mx, nrow = BI_n_agegroups,
        ncol = BI_n_periods, byrow = FALSE)

    # dimension of mortality rate data:
    # number of years observed (T, BI_n_periods) X
    # number of age groups (M, BI_n_agegroups)
    BI_logM.1 <- t(log(BI_logM.1))
    BI_logM.2 <- t(log(BI_logM.2))

    # compute \hat{Z}_t
    BI_Zhat_t.1 <- apply(BI_logM.1, 1, sum)
    BI_Zhat_t.2 <- apply(BI_logM.2, 1, sum)
    BI_Zhat_diff <- BI_Zhat_t.1 - BI_Zhat_t.2

    BI_Zhat_t.1_no.first.1 <- BI_Zhat_t.1[-1]
    BI_Zhat_t.2_no.first.1 <- BI_Zhat_t.2[-1]
    BI_Zhat_diff_no.first.1 <-
        BI_Zhat_t.1_no.first.1 - BI_Zhat_t.2_no.first.1

    BI_Zhat_t.1_no.last.1 <- BI_Zhat_t.1[-BI_n_periods]
```

```
BI_Zhat_t.2_no.last.1 <- BI_Zhat_t.2[-BI_n_periods]
BI_Zhat_diff_no.last.1 <-
    BI_Zhat_t.1_no.last.1 - BI_Zhat_t.2_no.last.1

BI_Zhat_t.1_no.first.2 <- BI_Zhat_t.1[-c(1, 2)]
BI_Zhat_t.2_no.first.2 <- BI_Zhat_t.2[-c(1, 2)]
BI_Zhat_diff_no.first.2 <-
    BI_Zhat_t.1_no.first.2 - BI_Zhat_t.2_no.first.2

BI_Zhat_t.1_no.first.1.last.1 <-
    BI_Zhat_t.1[-c(1, BI_n_periods)]
BI_Zhat_t.2_no.first.1.last.1 <-
    BI_Zhat_t.2[-c(1, BI_n_periods)]
BI_Zhat_diff_no.first.1.last.1 <-
    BI_Zhat_t.1_no.first.1.last.1 - BI_Zhat_t.2_no.first.1.last.1

BI_Zhat_t.1_no.last.2 <-
    BI_Zhat_t.1[-c(BI_n_periods - 1, BI_n_periods)]
BI_Zhat_t.2_no.last.2 <-
    BI_Zhat_t.2[-c(BI_n_periods - 1, BI_n_periods)]
BI_Zhat_diff_no.last.2 <-
    BI_Zhat_t.1_no.last.2 - BI_Zhat_t.2_no.last.2

# WHEN WITHOUT BIAS CORRECTION

# the 2 x 2 equation system for \mu(1) and \phi(1)
BI_E1_a11 <- BI_n_periods - 1
BI_E1_a12 <- sum(BI_Zhat_t.1_no.last.1)
BI_E1_b1 <- sum(BI_Zhat_t.1_no.first.1)
BI_E1_a21 <- sum(BI_Zhat_t.1_no.last.1)
BI_E1_a22 <- sum(BI_Zhat_t.1_no.last.1 ^ 2)
BI_E1_b2 <-
    sum(BI_Zhat_t.1_no.first.1 * BI_Zhat_t.1_no.last.1)

# solving for \mu(1) and \phi(1)
BI_E1_Amatrix <-
    matrix(c(BI_E1_a11, BI_E1_a12, BI_E1_a21, BI_E1_a22),
    2, 2, byrow = TRUE)
BI_E1_Bvector <- c(BI_E1_b1, BI_E1_b2)
BI_E1_solvector <- solve(BI_E1_Amatrix) %*% BI_E1_Bvector

BI_mu.1 <- BI_E1_solvector[1]
BI_phi.1 <- BI_E1_solvector[2]
```

```
# the 2 x 2 equation system for \mu(2) and \phi(2)
BI_E2_a11 <- BI_n_periods - 1
BI_E2_a12 <- sum(BI_Zhat_diff_no.last.1)
BI_E2_b1 <- sum(BI_Zhat_diff_no.first.1)
BI_E2_a21 <- sum(BI_Zhat_diff_no.last.1)
BI_E2_a22 <- sum(BI_Zhat_diff_no.last.1 ^ 2)
BI_E2_b2 <-
    sum(BI_Zhat_diff_no.first.1 * BI_Zhat_diff_no.last.1)

# solving for \mu(2) and \phi(2)
BI_E2_Amatrix <-
    matrix(c(BI_E2_a11, BI_E2_a12, BI_E2_a21, BI_E2_a22),
    2, 2, byrow = TRUE)
BI_E2_Bvector <- c(BI_E2_b1, BI_E2_b2)
BI_E2_solvector <- solve(BI_E2_Amatrix) %*% BI_E2_Bvector

BI_mu.2 <- BI_E2_solvector[1]
BI_phi.2 <- BI_E2_solvector[2]

# storage space for \alpha_x and \beta_x where x is age group
BI_axbx <- NULL

# the equation systems for \alpha_x(i) and \beta_x(i)
# where x is age group, i = 1, 2
for (i in 1:BI_n_agegroups) {

    # central mortality rate of the specific age group
    BI_logM_x.1 <- BI_logM.1[, i]
    BI_logM_x.2 <- BI_logM.2[, i]

    # the 2 x 2 equation system for \alpha_x(1) and \beta_x(1)
    # for the specific age group
    BI_E3x_a11 <- BI_n_periods
    BI_E3x_a12 <- sum(BI_Zhat_t.1)
    BI_E3x_b1 <- sum(BI_logM_x.1)
    BI_E3x_a21 <- sum(BI_Zhat_t.1)
    BI_E3x_a22 <- sum(BI_Zhat_t.1 ^ 2)
    BI_E3x_b2 <- sum(BI_logM_x.1 * BI_Zhat_t.1)

    # solving for \alpha_x(1) and \beta_x(1)
    # for the specific age group
    BI_E3x_Amatrix <-
```

```
        matrix(c(BI_E3x_a11, BI_E3x_a12, BI_E3x_a21, BI_E3x_a22),
        2, 2, byrow = TRUE)
    BI_E3x_Bvector <- c(BI_E3x_b1, BI_E3x_b2)
    BI_E3x_solvector <- solve(BI_E3x_Amatrix) %*% BI_E3x_Bvector

    # the 2 x 2 equation system for \alpha_x(2) and \beta_x(2)
    # for the specific age group
    BI_E4x_a11 <- BI_n_periods
    BI_E4x_a12 <- sum(BI_Zhat_t.2)
    BI_E4x_b1 <- sum(BI_logM_x.2)
    BI_E4x_a21 <- sum(BI_Zhat_t.2)
    BI_E4x_a22 <- sum(BI_Zhat_t.2 ^ 2)
    BI_E4x_b2 <- sum(BI_logM_x.2 * BI_Zhat_t.2)

    # solving for \alpha_x(2) and \beta_x(2)
    # for the specific age group
    BI_E4x_Amatrix <-
        matrix(c(BI_E4x_a11, BI_E4x_a12, BI_E4x_a21, BI_E4x_a22),
        2, 2, byrow = TRUE)
    BI_E4x_Bvector <- c(BI_E4x_b1, BI_E4x_b2)
    BI_E4x_solvector <- solve(BI_E4x_Amatrix) %*% BI_E4x_Bvector

    # c( \alpha_x(1), \beta_x(1), \alpha_x(2), \beta_x(2) )
    BI_axbx <- cbind(BI_axbx, c(BI_E3x_solvector[1], BI_E3x_solvector[2],
        BI_E4x_solvector[1], BI_E4x_solvector[2]))

} # end the equation systems for \alpha_x(i) and \beta_x(i)

# WHEN WITH BIAS CORRECTION

# the 2 x 2 equation system for \mu(1) and \phi(1)
BI_E5_a11 <- BI_n_periods - 2
BI_E5_a12 <- sum(BI_Zhat_t.1_no.first.1.last.1)
BI_E5_b1 <- sum(BI_Zhat_t.1_no.first.2)
BI_E5_a21 <- sum(BI_Zhat_t.1_no.last.2)
BI_E5_a22 <-
    sum(BI_Zhat_t.1_no.first.1.last.1 * BI_Zhat_t.1_no.last.2)
BI_E5_b2 <-
    sum(BI_Zhat_t.1_no.first.2 * BI_Zhat_t.1_no.last.2)

# solving for \mu(1) and \phi(1)
BI_E5_Amatrix <-
    matrix(c(BI_E5_a11, BI_E5_a12, BI_E5_a21, BI_E5_a22),
```

```
    2, 2, byrow = TRUE)
BI_E5_Bvector <- c(BI_E5_b1, BI_E5_b2)
BI_E5_solvector <- solve(BI_E5_Amatrix) %*% BI_E5_Bvector

BI_mu.1.s <- BI_E5_solvector[1]
BI_phi.1.s <- BI_E5_solvector[2]

# the 2 x 2 equation system for \mu(2) and \phi(2)
BI_E6_a11 <- BI_n_periods - 2
BI_E6_a12 <- sum(BI_Zhat_diff_no.first.1.last.1)
BI_E6_b1 <- sum(BI_Zhat_diff_no.first.2)
BI_E6_a21 <- sum(BI_Zhat_diff_no.last.2)
BI_E6_a22 <-
    sum(BI_Zhat_diff_no.first.1.last.1 * BI_Zhat_diff_no.last.2)
BI_E6_b2 <-
    sum(BI_Zhat_diff_no.first.2 * BI_Zhat_diff_no.last.2)

# solving for \mu(2) and \phi(2)
BI_E6_Amatrix <-
    matrix(c(BI_E6_a11, BI_E6_a12, BI_E6_a21, BI_E6_a22),
    2, 2, byrow = TRUE)
BI_E6_Bvector <- c(BI_E6_b1, BI_E6_b2)
BI_E6_solvector <- solve(BI_E6_Amatrix) %*% BI_E6_Bvector

BI_mu.2.s <- BI_E6_solvector[1]
BI_phi.2.s <- BI_E6_solvector[2]

# storage space for \alpha_x and \beta_x where x is age group
BI_axbx.s <- NULL

# the equation systems for \alpha_x(i) and \beta_x(i)
# where x is age group, i = 1, 2
for (i in 1:BI_n_agegroups) {

    # central mortality rate of the specific age group
    BI_logM_x.1 <- BI_logM.1[, i]
    BI_logM_x.2 <- BI_logM.2[, i]

    # compute various sums of log m(x,t)
    # for the specific age group
    BI_logM_x.1_no.first.1 <- BI_logM_x.1[-1]
    BI_logM_x.2_no.first.1 <- BI_logM_x.2[-1]
```

```
    # the 2 x 2 equation system for \alpha_x(1) and \beta_x(1)
    # for the specific age group
    BI_E7x_a11 <- BI_n_periods - 1
    BI_E7x_a12 <- sum(BI_Zhat_t.1_no.first.1)
    BI_E7x_b1 <- sum(BI_logM_x.1_no.first.1)
    BI_E7x_a21 <- sum(BI_Zhat_t.1_no.last.1)
    BI_E7x_a22 <-
        sum(BI_Zhat_t.1_no.first.1 * BI_Zhat_t.1_no.last.1)
    BI_E7x_b2 <-
        sum(BI_logM_x.1_no.first.1 * BI_Zhat_t.1_no.last.1)

    # solving for \alpha_x(1) and \beta_x(1)
    # for the specific age group
    BI_E7x_Amatrix <-
        matrix(c(BI_E7x_a11, BI_E7x_a12, BI_E7x_a21, BI_E7x_a22),
        2, 2, byrow = TRUE)
    BI_E7x_Bvector <- c(BI_E7x_b1, BI_E7x_b2)
    BI_E7x_solvector <- solve(BI_E7x_Amatrix) %*% BI_E7x_Bvector

    # the 2 x 2 equation system for \alpha_x(2) and \beta_x(2)
    # for the specific age group
    BI_E8x_a11 <- BI_n_periods - 1
    BI_E8x_a12 <- sum(BI_Zhat_t.2_no.first.1)
    BI_E8x_b1 <- sum(BI_logM_x.2_no.first.1)
    BI_E8x_a21 <- sum(BI_Zhat_t.2_no.last.1)
    BI_E8x_a22 <- sum(BI_Zhat_t.2_no.first.1 * BI_Zhat_t.2_no.last.1)
    BI_E8x_b2 <- sum(BI_logM_x.2_no.first.1 * BI_Zhat_t.2_no.last.1)

    # solving for \alpha_x(2) and \beta_x(2)
    # for the specific age group
    BI_E8x_Amatrix <-
        matrix(c(BI_E8x_a11, BI_E8x_a12, BI_E8x_a21, BI_E8x_a22),
        2, 2, byrow = TRUE)
    BI_E8x_Bvector <- c(BI_E8x_b1, BI_E8x_b2)
    BI_E8x_solvector <- solve(BI_E8x_Amatrix) %*% BI_E8x_Bvector

    # c( \alpha_x(1), \beta_x(1), \alpha_x(2), \beta_x(2) )
    BI_axbx.s <- cbind(BI_axbx.s,
        c(BI_E7x_solvector[1], BI_E7x_solvector[2],
        BI_E8x_solvector[1], BI_E8x_solvector[2]))

} # end the equation systems for \alpha_x(i) and \beta_x(i)
```

```
    return(list(BI_mu.1, BI_phi.1, BI_mu.2, BI_phi.2, BI_axbx,
        BI_mu.1.s, BI_phi.1.s, BI_mu.2.s, BI_phi.2.s, BI_axbx.s,
        BI_Zhat_t.1, BI_Zhat_t.2, BI_Zhat_diff))

}

j <- BI_func(BI_n_agegroups = DA_n_agegroups,
    BI_n_periods = DA_n_periods,
    BI_lifet1_mx = DA_lifetbl_male$mx,
    BI_lifet2_mx = DA_lifetbl_female$mx)
DA_BI_noncorrect <- j[c(1:5, 11:13)]
DA_BI_biascorrected <- j[6:13]

## DATA SIMULATION

SM_func <- function(SM_loopindex, SM_n_agegroups, SM_n_periods,
    SM_ax.1, SM_ax.2, SM_bx.1, SM_bx.2,
    SM_mu.1, SM_mu.2, SM_phi.1, SM_phi.2,
    SM_sd_e_t.1, SM_sd_e_t.2, SM_sd_epsilon, SM_agelist) {
    # begin simulation loop

    # data generating process
    # below: data generating process is based on input parameters
    #        after the simulation loop
    #        check input parameters

    # k_t in simulation (a new k_t series for each simulation loop)
    # assume k_0 = 0 and by induction

    SM_k_t.1 <- SM_mu.1 + rnorm(1, mean = 0, sd = 1) * SM_sd_e_t.1
    SM_k_t.diff <- SM_mu.2 + rnorm(1, mean = 0, sd = 1) * SM_sd_e_t.2
    for (i in 2:SM_n_periods) {
        SM_k_t.1 <-
            c(SM_k_t.1, SM_mu.1 + SM_phi.1 * SM_k_t.1[i - 1] +
            rnorm(1, mean = 0, sd = 1) * SM_sd_e_t.1)
        SM_k_t.diff <-
            c(SM_k_t.diff, SM_mu.2 + SM_phi.2 * SM_k_t.diff[i - 1] +
            rnorm(1, mean = 0, sd = 1) * SM_sd_e_t.2)
    }
    names(SM_k_t.1) <- NULL
    names(SM_k_t.diff) <- NULL

    SM_k_t.2 <- SM_k_t.1 - SM_k_t.diff
```

```
# mortality rate data 'logM' in simulation
# the dimension of SM_logM is: SM_n_periods X SM_n_agegroups (M)
SM_logM.1 <-
    t(cbind(SM_ax.1, SM_bx.1) %*% as.matrix(rbind(1, SM_k_t.1)) +
    matrix(rnorm(SM_n_agegroups * SM_n_periods, mean = 0, sd = 1) *
    SM_sd_epsilon, nrow = SM_n_agegroups))
SM_logM.2 <-
    t(cbind(SM_ax.2, SM_bx.2) %*% as.matrix(rbind(1, SM_k_t.2)) +
    matrix(rnorm(SM_n_agegroups * SM_n_periods, mean = 0, sd = 1) *
    SM_sd_epsilon, nrow = SM_n_agegroups))

# recover the original mx array
# this is reverse process of deriving logM from
# life table's 'mx' variable
SM_mx.1 <- array(exp(t(SM_logM.1)))
SM_mx.2 <- array(exp(t(SM_logM.2)))

# implementation of the original lee-carter model
# include LC_func here ...

SM_LC.1 <- LC_func(LC_n_agegroups = SM_n_agegroups,
    LC_n_periods = SM_n_periods,
    LC_lifetbl_mx = SM_mx.1,
    LC_agelist = SM_agelist,
    LC_yearlist = 1:SM_n_periods)
SM_LC.2 <- LC_func(LC_n_agegroups = SM_n_agegroups,
    LC_n_periods = SM_n_periods,
    LC_lifetbl_mx = SM_mx.2,
    LC_agelist = SM_agelist,
    LC_yearlist = 1:SM_n_periods)

# include LC_diff_func here ...

SM_LC.diff <- LC_diff_func(diff_n_agegroups = SM_n_agegroups,
    diff_n_periods = SM_n_periods, diff_var1 = SM_LC.1,
    diff_var2 = SM_LC.2)

# implementation of the bivariate lee-carter mortality model
# include BI_func here ...

j <- SM_BI_func(BI_n_agegroups = SM_n_agegroups,
    BI_n_periods = SM_n_periods,
```

```
        BI_lifet1_mx = SM_mx.1,
        BI_lifet2_mx = SM_mx.2)
    SM_BI_noncorrect <- j[c(1:5, 11:13)]
    SM_BI_biascorrected <- j[6:13]

    return(list(SM_LC.1, SM_LC.2, SM_LC.diff,
        SM_BI_noncorrect, SM_BI_biascorrected))

} # end simulation loop

# number of simulation loops
SM_loopindex <- 1:10000

# number of simulated age groups
SM_n_agegroups <- DA_n_agegroups

# number of simulated periods (80, 150, 500, 1000)
SM_n_periods <- 150

# bivariate model parameters for simulation
SM_ax.1 <- DA_BI_biascorrected[[5]][1, ]
SM_ax.2 <- DA_BI_biascorrected[[5]][3, ]
SM_bx.1 <- DA_BI_biascorrected[[5]][2, ]
SM_bx.2 <- DA_BI_biascorrected[[5]][4, ]
SM_mu.1 <- DA_BI_biascorrected[[1]]
SM_mu.2 <- DA_BI_biascorrected[[3]]
SM_phi.1 <- 1
SM_phi.2 <- 0.95
SM_sd_e_t.1 <- 0.1
SM_sd_e_t.2 <- 0.1
SM_sd_epsilon <- 0.1
SM_agelist <- unique(DA_lifetbl_male$Age0)

# prepare parallel clusters for simulation
library(snowfall)
library(rlecuyer)
set.seed(123)
#sfInit(parallel = FALSE)
sfInit(parallel = TRUE, cpus = 8, type = 'SOCK')
i <- sfClusterSetupRNGstream(123)

# execute simulation loops
SM_result <-
```

```
    sfLapply(SM_loopindex, SM_func, SM_n_agegroups, SM_n_periods,
    SM_ax.1, SM_ax.2, SM_bx.1, SM_bx.2,
    SM_mu.1, SM_mu.2, SM_phi.1, SM_phi.2,
    SM_sd_e_t.1, SM_sd_e_t.2, SM_sd_epsilon, SM_agelist)
sfStop()

# end
```

Bibliography

Ando, A., & Modigliani, F. (1963). The "life cycle" hypothesis of saving: Aggregate implications and tests. *The American economic review*, *53*(1), 55–84.

Biffis, E. (2005). Affine processes for dynamic mortality and actuarial valuations. *Insurance: mathematics and economics*, *37*(3), 443–468.

Biffis, E., Lin, Y., & Milidonis, A. (2017). The cross-section of asia-pacific mortality dynamics: Implications for longevity risk sharing. *Journal of Risk and Insurance*, *84*(S1), 515–532.

Bisetti, E., & Favero, C. A. (2014). Measuring the impact of longevity risk on pension systems: The case of italy. *North American Actuarial Journal*, *18*(1), 87–103.

Brouhns, N., Denuit, M., & Vermunt, J. K. (2002). A poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and economics*, *31*(3), 373–393.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., & Khalaf-Allah, M. (2011). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, *48*(3), 355–367.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *The annals of statistics*, *14*(3), 1171–1179.

Chan, N. H., & Wei, C.-Z. (1987). Asymptotic inference for nearly nonstationary ar (1) processes. *The Annals of Statistics*, *15*(3), 1050–1063.

Chen, H., & Cox, S. H. (2009). Modeling mortality with jumps: Applications to mortality securitization. *Journal of Risk and Insurance*, *76*(3), 727–751.

Chen, H., & Cummins, J. D. (2010). Longevity bond premiums: The extreme value approach and risk cubic pricing. *Insurance: Mathematics and Economics*, *46*(1), 150–161.

Cox, S. H., & Lin, Y. (2007). Natural hedging of life and annuity mortality risks. *North American Actuarial Journal*, *11*(3), 1–15.

Dahl, M. (2004). Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts. *Insurance: mathematics and economics*, *35*(1), 113–136.

Dickson, D. C., Hardy, M., Hardy, M. R., & Waters, H. R. (2013). *Actuarial mathematics for life contingent risks*. Cambridge University Press.

Dowd, K., Cairns, A. J., Blake, D., Coughlan, G. D., Epstein, D., & Khalaf-Allah, M. (2010). Evaluating the goodness of fit of stochastic mortality models. *Insurance: Mathematics and Economics*, *47*(3), 255–265.

D'Amato, V., Haberman, S., Piscopo, G., Russolillo, M., & Trapani, L. (2014). Detecting common longevity trends by a multiple population approach. *North American Actuarial Journal*, *18*(1), 139–149.

D'Amato, V., Haberman, S., & Russolillo, M. (2012). The stratified sampling bootstrap for measuring the uncertainty in mortality forecasts. *Methodology and Computing in Applied Probability*, *14*(1), 135–148.

Enchev, V., Kleinow, T., & Cairns, A. J. (2017). Multi-population mortality models: fitting, forecasting and comparisons. *Scandinavian Actuarial Journal*, *2017*(4), 319–342.

Frees, E. W., Carriere, J., & Valdez, E. (1996). Annuity valuation with dependent mortality. *Journal of risk and insurance*, (pp. 229–261).

Girosi, F., & King, G. (2007). Understanding the lee-carter mortality forecasting method. *Gking. Harvard. Edu.*

Haberman, S., & Renshaw, A. (2009). On age-period-cohort parametric mortality rate projections. *Insurance: Mathematics and Economics*, *45*(2), 255–270.

Hall, P., & Heyde, C. C. (2014). *Martingale limit theory and its application*. Academic press.

Kwok, K. Y., Chiu, M. C., & Wong, H. Y. (2016). Demand for longevity securities under relative performance concerns: Stochastic differential games with cointegration. *Insurance: Mathematics and Economics*, *71*, 353–366.

Lahiri, S. N. (2013). *Resampling methods for dependent data*. Springer Science & Business Media.

Lee, R. (2000). The lee-carter method for forecasting mortality, with various extensions and applications. *North American actuarial journal*, *4*(1), 80–91.

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting us mortality. *Journal of the American statistical association*, *87*(419), 659–671.

Leng, X., & Peng, L. (2016). Inference pitfalls in lee–carter model for forecasting mortality. *Insurance: Mathematics and Economics*, *70*, 58–65.

Leng, X., & Peng, L. (2017). Testing for a unit root in lee–carter mortality model. *ASTIN Bulletin: The Journal of the IAA*, *47*(3), 715–735.

Li, H., De Waegenaere, A., & Melenberg, B. (2017a). Robust mean–variance hedging of longevity risk. *Journal of Risk and Insurance*, *84*(S1), 459–475.

Li, J. S., Chan, W., & Zhou, R. (2017b). Semicoherent multipopulation mortality modeling: the impact on longevity risk securitization. *Journal of Risk and Insurance*, *84*(3), 1025–1065.

Li, J. S.-H. (2010). Pricing longevity risk with the parametric bootstrap: A maximum entropy approach. *Insurance: Mathematics and Economics*, *47*(2), 176–186.

Li, J. S.-H., & Hardy, M. R. (2011). Measuring basis risk in longevity hedges. *North American Actuarial Journal*, *15*(2), 177–200.

Li, J. S.-H., Li, K. K. J., & Balasooriya, U. (2018). Constructing out-of-the-money longevity hedges using parametric mortality indexes. *North American Actuarial Journal*.

Li, J. S.-H., Zhou, R., & Hardy, M. (2015). A step-by-step guide to building two-population stochastic mortality models. *Insurance: Mathematics and Economics*, *63*, 121–134.

Li, N., & Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the lee-carter method. *Demography*, *42*(3), 575–594.

Li, N., Lee, R., & Tuljapurkar, S. (2004). Using the lee–carter method to forecast mortality for populations with limited data. *International Statistical Review*, *72*(1), 19–36.

Lin, Y., Shi, T., & Arik, A. (2017). Pricing buy-ins and buy-outs. *Journal of Risk and Insurance*, *84*(S1), 367–392.

Lin, Y., Tan, K. S., Tian, R., & Yu, J. (2014). Downside risk management of a defined benefit plan considering longevity basis risk. *North American Actuarial Journal*, *18*(1), 68–86.

Liu, Q., Ling, C., Li, D., & Peng, L. (2019a). Bias-corrected inference for a modified lee–carter mortality model. *ASTIN Bulletin: The Journal of the IAA*, *49*(2), 433–455.

Liu, Q., Ling, C., & Peng, L. (2019b). Statistical inference for lee-carter mortality model and corresponding forecasts. *North American Actuarial Journal*, (pp. 1–29).

Luciano, E., Spreeuw, J., & Vigna, E. (2008). Modelling stochastic mortality for dependent lives. *Insurance: Mathematics and Economics*, *43*(2), 234–244.

McLeish, D. L. (1975). A maximal inequality and dependent strong laws. *The Annals of probability*, *3*(5), 829–839.

Milevsky, M. A. (1998). Optimal asset allocation towards the end of the life cycle: to annuitize or not to annuitize? *Journal of Risk and Insurance*, (pp. 401–426).

Milevsky, M. A., & Promislow, S. D. (2001). Mortality derivatives and the option to annuitise. *Insurance: Mathematics and Economics*, *29*(3), 299–318.

Phillips, P. C. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika*, *74*(3), 535–547.

Phillips, P. C., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, *75*(2), 335–346.

Politis, D. N., & Romano, J. P. (1993). On the sample variance of linear statistics derived from mixing sequences. *Stochastic Processes and their Applications*, *45*(1), 155–167.

Pollard, J. H. (1975). *Mathematical models for the growth of human populations*. Cambridge University Press Cambridge.

Richards, S. J., Kirkby, J., & Currie, I. D. (2006). The importance of year of birth in two-dimensional mortality data. *British Actuarial Journal*, *12*(1), 5–38.

Schrager, D. F. (2006). Affine stochastic mortality. *Insurance: mathematics and economics*, *38*(1), 81–97.

Staudenmayer, J., & Buonaccorsi, J. P. (2005). Measurement error in linear autoregressive models. *Journal of the American Statistical Association*, *100*(471), 841–852.

Tabeau, E. (2001). *A review of demographic forecasting models for mortality*, (pp. 1–32). Springer.

Willets, R. (2004). The cohort effect: insights and explanations. *British Actuarial Journal*, *10*(4), 833–877.

Wong, T. W., Chiu, M. C., & Wong, H. Y. (2017). Managing mortality risk with longevity bonds when mortality rates are cointegrated. *Journal of Risk and Insurance*, *84*(3), 987–1023.

Yagi, T., & Nishigaki, Y. (1993). The inefficiency of private constant annuities. *Journal of Risk and Insurance*, (pp. 385–412).

Zhu, W., Tan, K. S., & Wang, C. (2017). Modeling multicountry longevity risk with mortality dependence: A lévy subordinated hierarchical archimedean copulas approach. *Journal of Risk and Insurance*, *84*(S1), 477–493.