EFFICIENT ESTIMATION OF FACTOR MODELS WITH
TIME AND CROSS-SECTIONAL DEPENDENCE

Alexander Heinemann

Master Thesis CEMFI No. 1501

July 2015

CEMFI
Casado del Alisal 5; 28014 Madrid
Tel. (34) 914 290 551. Fax (34) 914 291 056
Internet: www.cemfi.es

This paper is a revised version of the Master's Thesis presented in partial fulfilment of the 2013-2015 Master in
Economics and Finance at the Centro de Estudios Monetarios y Financieros (CEMFI). I would like to thank my
advisor, Dante Amengual, for his invaluable guidance and support during all stages of the project. In addition, I am
grateful to Enrique Sentana and Manuel Arellano for their constructive comments and econometric advice. I
appreciate the valuable conversations and fruitful discussions with Josep Pijoan-Mas, Pedro Mira, Luca Repetto and
Guillermo Tellechea. The usual disclaimer applies.
EFFICIENT ESTIMATION OF FACTOR MODELS WITH TIME AND CROSS-SECTIONAL DEPENDENCE

Abstract

This paper studies the efficient estimation of large-dimension factor models with time and simultaneously cross-sectional dependence assuming multiplicative separability of the covariance matrix. The asymptotic distribution of the estimator of the factor and factor-loading space under factor stationarity is derived and compared to the one of the principal component estimator. The paper also considers the case when factors exhibit a unit root. We provide feasible estimators and show in simulation that they are more efficient than the principal component estimator infinite samples. In application, the estimation procedure is employed to estimate the Lee-Carter model and life expectancy is forecasted.

Alexander Heinemann
Maastricht University
a.heinemann@alumni.maastrichtuniversity.nl
1 Introduction

Factor Models have become an important tool in the present economics and finance literature. Setting large data sets in relation to relatively few latent factors, one reduces the problem’s dimension while being able to capture a large amount of data variability. Factors are regarded as latent economic variables that capture the co-movement of a large number of variables. For instance, in modeling mortality rates for different age groups over time, Lee and Carter (1992) propose a model with a single factor representing the time variation in mortality rates. A distinction is made between static and dynamic factor models. Albeit its name, the static factor model also allows for factor dynamics, however they are not explicitly modeled. On the other hand, the pre-fix dynamic is also used for models in which the factors are considered to be static, but the errors follow dynamic processes (i.e. Breitung and Tenhofen (2011)). Regarding the estimation of the static factor model, the workhorse in the literature is the method of principal components proposed by Chamberlain and Rothschild (1982). Its asymptotic distribution of factors and loadings is established by Bai (2003) within the framework of large cross-sectional ($N$) and time-series ($T$) dimensions. He also studies the asymptotic distribution of the factor and loading space in large-dimension factor models with non-stationary dynamic factors (Bai, 2004). In both publications, he considers the case of time and cross-correlation in the idiosyncratic error, though he does not explicitly capture the dependence structure. In general, although there exist well-established estimating procedures for static and dynamic factor models, efficiency considerations have not received much attention in the literature. Exceptions are Breitung and Tenhofen (2011) and Choi (2012). Breitung and Tenhofen take a GLS approach to account for autocorrelation and heteroskedasticity in the idiosyncratic error in a static factor model with dynamic errors. In contrast, Choi proposes an estimator allowing for cross-sectional dependence in static factor models.

This paper studies the efficient estimation of large-dimension factor models with time and simultaneously cross-sectional dependence allowing for heteroskedasticity while assuming multiplicative separability of the covariance matrix. Estimators for the model’s factors and factor loadings are proposed, which have a (conditional) maximum likelihood interpretation under normality. In the spirit of Choi (2012),
we build on the results of Bai (2003), to derive the asymptotic distribution of the unfeasible estimators of the factor and factor-loading space assuming factor stationarity. The feasible counterparts are shown to be asymptotically equivalent under mild conditions. The paper also considers the case when factors exhibit a unit root. Given this specification, the asymptotic distribution of the estimators of the factor and factor loading space is obtained adopting the approach of Bai (2004). The efficiency gain over the principal component estimator in finite sample is illustrated in a Monte Carlo simulation. In addition, the proposed estimation procedure is applied to the Lee-Carter model and a natural explanation is provided for autocorrelation and cross-sectional dependence.

The structure of the paper is as follows: Section 2 introduces the model in more detail and stresses on the error dependence structure. Section 3 discusses the estimation procedure when the dependence structure is known and proposes feasible estimators. In Section 4 the estimators’ asymptotic distribution and their relative efficiency is studied. Section 5 reports a detailed Monte-Carlo evaluation of the estimators. In section 6, we revisit the Lee-Carter model. Section 7 summarizes and discusses interesting avenues for further research.

2 Model

We consider a factor model for a panel of $i = 1, \ldots, N$ groups and $t = 1, \ldots, T$ periods written as

$$X_{i,t} = \lambda_i' F_t + e_{i,t}$$  \hspace{1cm} (1)

where $F_t$ is a vector of latent factors of length $r$ and $\lambda_i$ is a vector of factor coefficients of length $r$. Alternatively, one can express the model in compact matrix notation:

$$X = FN' + e$$  \hspace{1cm} (2)

where $X$ is a $T \times N$ matrix with elements $X_{i,t}$, $F = (F_1, \ldots, F_T)'$ is a matrix of dimension $T \times r$, $\Lambda = (\lambda_1, \ldots, \lambda_N)'$ is an $N \times r$ matrix of factor coefficients and $e$ denotes a $T \times N$ matrix with idiosyncratic error elements $e_{i,t}$. Throughout the paper, $r$ is assumed to be known and we denote the norm of matrix $A$ by $\|A\| = [tr(\Lambda A')]^{1/2}$. 

4
The matrices $F$ and $\Lambda$ are unknown and it is the objective to estimate the matrices efficiently under time and simultaneously cross-sectional dependence.

**Assumption A: Time and Cross-sectional Dependence**

The idiosyncratic component $e_{i,t}$ has a stationary heterogeneous autoregressive form satisfying:

(i) $e_{i,t} = \phi_i(L) \theta_t(B) \sigma_i \varepsilon_{i,t}$

(ii) $\phi_i(L) = \phi(L)$ for all $i$ and $\theta_t(B) = \theta(B)$ for all $t$

where $L$ and $B$ are the lag-operators with respect to $t$ and $i$, $\phi_i(L)$ and $\theta_t(B)$ are invertible lag polynomials and $\varepsilon_{i,t}$ has unit variance.

The covariance structure defined in Assumption A (i) can account for autocorrelation through $\phi_i(L)$, cross-correlation through $\theta_t(B)$ and heteroskedasticity through $\sigma_i$; it nests the model considered by Breitung and Tenhofen (2011), that is a model of heteroskedasticity and autocorrelation only when $\theta_t(B) = 1$. Assumption A (ii) is an auxiliary assumption: it restricts the correlation structure to be the same across years as well as across units, which seems to be reasonable assumption in the application of mortality rates. It reduces the parameter dimension and simplifies the derivation of the asymptotic distribution. Combining Assumptions A (i) and (ii) results in the following Kronecker product structure:

$$\text{Cov}(\text{vec}(e)) = \Theta' \Theta \otimes \Phi \Phi'$$  

where $\Phi$ and $\Theta$ are two matrices of dimensions $T \times T$ and $N \times N$ with bounded eigenvalues and $\Phi \Phi'$ and $\Theta' \Theta$ being positive definite. Generally, we have:

$$\Phi \Phi' = \begin{pmatrix} 1 & \phi_1 & \cdots & \phi_{T-1} \\ \phi_1 & 1 & \cdots & \phi_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{T-1} & \phi_{T-2} & \cdots & 1 \end{pmatrix}, \quad \Theta' \Theta = \Sigma \otimes \begin{pmatrix} 1 & \theta_1 & \cdots & \theta_{N-1} \\ \theta_1 & 1 & \cdots & \theta_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{N-1} & \theta_{N-2} & \cdots & 1 \end{pmatrix} \Sigma$$  

where $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_N)$. If $\Phi = I_T$, the model reduces to Choi’s model. In contrast, if $\Theta = I_N$, then it reduces to a model with autocorrelation only.
Assumption B: Independence
\( \{F_t\}_{t=1}^T, \{e_{i,t}\}_{i,t=1}^{N,T} \) and \( \{\lambda_i\}_{i=1}^N \) are mutually independent.

Assumption B limits \( F_t, e_{i,t} \) and \( \lambda_i \) to be mutually independent groups across \( i \) and \( t \), while dependence within a group is allowed. This assumption is slightly stronger than the corresponding one in Bai (2003) and is assumed for clearness. It can be relaxed at a cost of notation. However, the assumption of strict exogeneity:

\[
E(e_{i,t}|F_1, \ldots, F_T, \lambda_1, \ldots, \lambda_N) = 0
\]

for all \((i,t)\), which is implied by Assumption B, must be maintained. It is needed for consistency of the GLS estimator proposed in section 3. Align with Hansen and Hodrick (1980), in the absence of the strict exogeneity assumption, the GLS transformation generally induces dependence between the transformed factors/factor-loadings and the transformed error, which produces generally inconsistent estimators.

Remark: An alternative covariance structure that accounts for cross- and time correlation is based on a Kronecker sum structure, e.g. \( \text{Cov}(\text{vec}(e)) = A \oplus B \). An important special case of this type is the two-way error component model given by \( e_{i,t} = \alpha_i + \delta_t + u_{i,t} \) that is often used in Panel data models. A special feature of the Kronecker sum covariance structure is that it prohibits correlation among different units at different times, e.g. \( \text{Cov}(e_{i,t}, e_{j,s}) = 0 \) for all \( i \neq j \) and \( t \neq s \). This assumption may be reasonable in certain applications such as a Panel of GDP time series of different countries, in which two different countries at distinct times are uncorrelated. However, this assumption may be too restrictive for geographical models, where the \( T \) and \( N \) dimensions correspond to the Cartesian axes. The covariance structure proposed in Assumption A implies a Kronecker product structure, e.g. \( \text{Cov}(\text{vec}(e)) = A \otimes B \). In contrast, it can accommodate for correlation among different units at different times, which comes at a structural cost since the number of parameters is the same. Employing one or the other covariance structure depends on the economic application. Having a mortality model as application in mind, where the \( T \) and \( N \) dimensions correspond to time and age respectively, we subsequently assume a Kronecker product structure to account for the fact people aged \( i \) in year \( t \) were aged \( i - 1 \) in year \( t - 1 \).
3 GLS Estimation Method

In the spirit of the General Method of Moments (GMM), we consider the estimation of $F_t, t = 1, \ldots, T$ and $\lambda_i, i = 1, \ldots, N$ based on the $E(e_{i,t}) = 0$. More precisely, the estimators $\hat{F}$ and $\hat{\Lambda}$ minimize the following criterion function:

$$\hat{F}, \hat{\Lambda} = \arg\min_{F, \Lambda} \text{vec}(X - FA')' A_{NT} \text{vec}(X - FA')$$

where $A_{NT}$ is an $NT \times NT$, possibly random, non-negative definite weight matrix. When the weighting matrix is the identity matrix, the estimators $\hat{F}$ and $\hat{\Lambda}$ minimize the (unweighted) total sum of squares and hence coincide with the principal components (PCA) estimators. If the idiosyncratic errors are heteroskedastic, cross- or autocorrelated, the PCA estimator is not efficient. For the heteroskedastic and cross-correlation case, Choi (2012) suggests a GLS-type estimator that minimize the weighted sum of squares. Generalizing his approach to account for temporal dependence at the same time, the weighting matrix is taken to be the inverse of $Cov(\text{vec}(e))$ subsequently. Hence, $\hat{F}$ and $\hat{\Lambda}$ minimize:

$$\text{vec}(X - FA')' [Cov(\text{vec}(e))]^{-1} \text{vec}(X - FA') = \text{vec}(\varepsilon)' \text{vec}(\varepsilon) = tr(\varepsilon'\varepsilon)$$

where $\varepsilon$ is a $T \times N$ matrix with elements $\varepsilon_{i,t}$.

3.1 Unfeasible GLS

For the time being, suppose that the matrices $\Phi$ and $\Theta$ are known. Then using $\varepsilon = \Phi^{-1} e \Theta^{-1}$ one can rewrite (2) as:

$$\Phi^{-1}X\Theta^{-1} = \Phi^{-1}FA'\Theta^{-1} + \varepsilon.$$  \hspace{1cm} (7)

Defining $Y = \Phi^{-1}X\Theta^{-1}, G = \Phi^{-1}F$ and $\Gamma' = \Lambda'\Theta^{-1}$, we can express (7) by

$$Y = GT' + \varepsilon.$$  \hspace{1cm} (8)
Therefore, the criterion function, that \( \hat{F} \) and \( \hat{\Lambda} \) minimize, is:

\[
tr \left( (Y - G\Gamma')'(Y - G\Gamma') \right). \tag{9}
\]

By definition of \( \Gamma \) and \( G \), there exists a one-to-one mapping between \((G, \Gamma)\) and \((F, \Lambda)\). Therefore, minimizing (9) with respect to \( \Lambda \) is equivalent to minimizing with respect to \( \Gamma \). We proceed as follows: first, we find \( \hat{G} \) and \( \hat{\Gamma} \) that minimize (9) and then we recover \( \hat{F} \) and \( \hat{\Lambda} \) using the one-to-one mapping. To obtain \( \hat{G} \) and \( \hat{\Gamma} \), we follow Choi (2012) by minimizing first with respect to \( \Gamma \) (as a function of \( G \)) and thereafter minimizing with respect to \( G \). The first-order condition with respect to \( \Gamma \) leads to \( \hat{\Gamma} (G) = Y'G(G'G)^{-1} \), which has a simple OLS regression form. Plugging this formula into (9), we obtain:

\[
tr \left( Y'(I - G(G'G)^{-1}G')Y \right). \tag{10}
\]

Minimizing (10) with respect to \( G \) is equivalent to

\[
\max_G \quad tr \left( (G'G)^{-1}G'YY'G \right). \tag{11}
\]

Standardizing \( \frac{1}{T}G'G = I_r \), \( \hat{G} \) is given by the first principal components of the matrix \( \frac{1}{NT}YY' \) and \( \hat{\Gamma} \) is equal to \( \frac{1}{T}Y'\hat{G} \). Using the one-to-one mapping between \((F, \Lambda)\) and \((G, \Gamma)\), the (unfeasible) GLS estimators of \( F \) and \( \Lambda \) are:

\[
\hat{F} = \Phi \hat{G} \tag{12a}
\]
\[
\hat{\Lambda} = \Theta' \hat{\Gamma}. \tag{12b}
\]

Under normality the estimators of \( F \) and \( \Lambda \) coincide with the conditional maximum likelihood estimator (CMLE), that maximize the following log-likelihood:

\[
\mathcal{L} = -\frac{TN}{2}ln(2\pi) - \frac{1}{2}tr \left( (Y - G\Gamma')(Y - G\Gamma') \right). \tag{13}
\]

This can be easily verified by comparing equations (9) with (13). The CMLEs need to be distinguished from the unconditional maximum likelihood estimator (MLE) because the MLE requires the knowledge of the probabilistic structure of \( \{F_t\}_{t=1}^T \) and
\{\lambda_i\}_{i=1}^N \text{ (Choi, 2012). Under non-normality, the estimator is commonly interpreted as conditional pseudo maximum likelihood estimator. We close this subsection with a remark on the (conditional) maximum likelihood estimator when Assumption A (ii) fails.}

**Remark:** Dropping Assumption A (ii), the first order conditions of the criterion function are:

\[
\frac{\partial}{\partial \lambda_i}: \quad \frac{1}{\sigma_i^2} \sum_t \varepsilon_{i,t} \phi_i^{-1}(L) F_t = 0 \quad (14a)
\]

\[
\frac{\partial}{\partial F_t}: \quad \sum_i \frac{1}{\sigma_i^2} \varepsilon_{i,t} \theta_i^{-1}(B) \lambda_i = 0 \quad (14b)
\]

Following Breitung and Tenhofen (2011) closely, one can show that the GLS estimator can be obtained by running the following least-squares regressions until convergence:

\[
\left( \phi_i^{-1}(L) X_{i,t} \right) = \left( \phi_i^{-1}(L) \tilde{F}_t \right) \lambda_i + \theta_i(B) \sigma_i \varepsilon_{i,t} \quad (t = \ldots, T) \quad (15a)
\]

\[
\left( \theta_i^{-1}(B) \frac{1}{\sigma_i} X_{i,t} \right) = \left( \theta_i^{-1}(B) \frac{1}{\sigma_i} \tilde{\lambda}_i \right) F_t + \phi_i(L) \varepsilon_{i,t} \quad (i = \ldots, N) \quad (15b)
\]

where \{\tilde{F}_t\}_{t=1}^T and \{\tilde{\lambda}_i\}_{i=1}^N are the estimates obtained in the previous step. The PCA-estimates \{\tilde{F}_t\}_{t=1}^T and \{\tilde{\lambda}_i\}_{i=1}^N can be used for initialization. To reduce the computational burden, Breitung and Tenhofen propose to run a single iteration that generates a 2-step estimator.

### 3.2 Feasible GLS

In the previous subsection, it was assumed that \Phi and \Theta are known. In practice, they are unknown such that one needs to obtain estimates. However, \Phi and \Theta are generally not identified, although \Phi\Phi' and \Theta'\Theta are (up to scale), which can be seen in equation (3). Without loss of generality, one can assume that \Phi and \Theta are the Cholesky factorization matrices of \Phi\Phi' and \Theta'\Theta respectively. This has to do with the fact that one needs only any \Phi and \Theta matrices such that \varepsilon = \Phi^{-1} e \Theta^{-1}, has
mutually uncorrelated elements with unit variance. Since every Hermitian positive-definite matrix has a unique Cholesky decomposition, one can focus on estimating \( \Phi \Phi' \) and \( \Theta' \Theta \) rather than \( \Phi \) and \( \Theta \). Because \( \Phi \Phi' \) and \( \Theta' \Theta \) are identified only up to scale, one matrix can be normalized without loss of generality.

To estimate \( \Phi \Phi' \) and \( \Theta' \Theta \) consistently, a specific dependence structure is imposed, that depends on a finite set of parameters. For instance, one could specify an AR(1) structure for \( \Phi \Phi' \) and a heterogeneous MA(1) structure for \( \Theta' \Theta \):

\[
\Phi \Phi' = \begin{pmatrix}
1 & \phi & \cdots & \phi^{T-1} \\
\phi & 1 & \cdots & \phi^{T-2} \\
\vdots & \vdots & \ddots & \vdots \\
\phi^{T-1} & \phi^{T-2} & \cdots & 1
\end{pmatrix},
\quad
\Theta' \Theta = \begin{pmatrix}
1 & \theta & \cdots & 0 \\
\theta & 1 & \cdots & \vdots \\
\vdots & \vdots & \ddots & \theta \\
0 & \cdots & \theta & 1
\end{pmatrix} \Sigma.
\]

(16)

where \( \Sigma = diag(\sigma_1, \ldots, \sigma_N) \). Equation (3) specifies the moment conditions of \( \Phi \Phi' \) and \( \Theta' \Theta \). Vectorizing the expression and eliminating redundant elements (due to symmetry) one obtains a vector of \( \frac{TN(TN+1)}{2} \) distinct moment conditions:

\[
E \left( vech \left[ vec(e) vec(e)' - \Theta' \Theta \otimes \Phi \Phi' \right] \right).
\]

(17)

Using the General Method of Moments while replacing \( e \) by an consistent estimate, one obtains estimates for the underlying parameters of \( \Theta' \Theta \) and \( \Phi \Phi' \), that are consistent under regulatory conditions. A natural choice for the sample counterpart of \( e \) is \( \tilde{e} = X - \tilde{F} \tilde{\Lambda}' \), where \( \tilde{F} \) and \( \tilde{\Lambda} \) are the first \( r \) principal components and factor loadings of \( \frac{1}{N} XX' \). Consistency of \( \tilde{e} \) is established in Stock and Watson (2002) under fairly weak assumptions. Hence, the resulting feasible GLS estimators can be obtained by applying the following steps:

1. Estimate \( X = F \Lambda' + e \) with PCA yielding consistent \( \tilde{F} \) and \( \tilde{\Lambda} \)
2. Use \( \tilde{e} = X - \tilde{F} \tilde{\Lambda}' \) and GMM to obtain estimates for \( \Theta' \Theta \) and \( \Phi \Phi' \)
3. Apply Cholesky to obtain \( \tilde{\Theta} \) and \( \tilde{\Phi} \)
4. Apply PCA on \( \tilde{Y} = \tilde{\Phi}^{-1} X \tilde{\Theta}^{-1} \) to obtain \( \tilde{G} \) and \( \tilde{\Gamma} \)
5. Recover \( \hat{F} \) and \( \hat{\Lambda} \) using \( \hat{F} = \tilde{\Phi} \hat{G} \) and \( \hat{\Lambda} = \tilde{\Theta}' \hat{\Gamma} \).

In reference to Breitung and Tenhofen (2011), one can consider iterative estimators, that repeat steps two to five until convergence. Although further iterations do not change the estimator’s asymptotic distribution, they are likely to improve performance in small samples.

4 Asymptotics of the GLS Estimators

In this section we study the asymptotic distributions of the factor and factor loadings estimators as well as their vector product under the assumption of factor stationarity. Moreover, we evaluate their asymptotic efficiency relative to their corresponding PCA estimators.

4.1 Asymptotic Distribution Under Factor Stationarity

To derive the asymptotic distribution of the estimators of the factor and factor loading space when the factors are stationary, we follow Bai (2003). First, we consider unfeasible estimators when \( \Phi \) and \( \Theta \) are known and thereafter turn to feasible estimators. One needs to assume that both, the cross-sectional (\( N \)) and time series (\( T \)) dimension, grow large. In particular we assume the following rates:

**Assumption C:** 
Rates
\( N, T \to \infty \) jointly such that \( \frac{\sqrt{N}}{T} \to 0 \) and \( \frac{\sqrt{T}}{N} \to 0 \).

The first part states, that \( N \) and \( T \) simultaneously grow large, whereas the second and the third expression limit \((N,T)\) to increase along particular trajectories, called pathwise limits. In addition, the following regulatory conditions are set for the factors and the factor loadings:

**Assumption D:** 
Factors and Factor Loadings

(i) \( E \left\| \sum_{s=1}^{T} \left[ \Phi^{-1} \right]_{t,s} F_s \right\|^4 \leq M < \infty \) for all \( t \) and \( \frac{1}{T} F' (\Phi \Phi')^{-1} F \overset{P}{\to} \Sigma_{F*} \) for some non-random \( r \times r \) positive definite matrix \( \Sigma_{F*} \).
\( \sum_{j=1}^{N} \left[ \Theta^{-1} \right]_{j,i} \lambda_j \leq M < \infty \) for all \( i \) and \( \frac{1}{N} \Lambda' (\Theta' \Theta)^{-1} \Lambda \Rightarrow \Sigma_{\Lambda^*} \) for some non-random \( r \times r \) positive definite matrix \( \Sigma_{\Lambda^*} \).

Assumption D is standard in the literature. It puts restrictions on the fourth moments and ensures the limiting matrices to be non-random positive definite. Moreover, one needs two simple central limit theorems to hold:

**Assumption E: Central Limit Theorem**

(i) for each \( t \) as \( N \to \infty \) we assume \( \frac{1}{\sqrt{N}} \Lambda' \Theta^{-1} \varepsilon' \Phi_t \stackrel{d}{\to} N(0, \Omega_t) \)

(ii) for each \( i \) as \( T \to \infty \) we assume \( \frac{1}{\sqrt{T}} \Phi_t' \varepsilon \Theta_i \stackrel{d}{\to} N(0, \Psi_i) \)

with \( \Omega_t = \lim_{N \to \infty} \left( \frac{1}{N} \Lambda' \Theta^{-1} \varepsilon' \Phi_t \varepsilon \Theta^{-1} \Lambda \right) \), \( \Psi_i = \operatorname{plim}_{T \to \infty} \left( \frac{1}{T} \Phi_t' \varepsilon \Theta_i \varepsilon \Phi_t' \varepsilon \right) \), \( \Phi_t = (\Phi_t, \ldots, \Phi_T)' \) and \( \Theta_i = (\Theta_1, \ldots, \Theta_N)' \).

Assumption E is needed to establish asymptotic normality of the parameter estimators \( F_t \) and \( \lambda_i \) and is satisfied by various mixing processes (Bai, 2003). Last, one needs a unique limit of \( \frac{1}{T} \hat{F}_t' (\Phi \Phi' )^{-1} F_t' \), in order for the limiting distributions of \( \hat{F}_t \) and \( \hat{\lambda}_i \) to be identified.

**Assumption F: Distinct Eigenvalues**

The eigenvalues of the \( r \times r \) matrix \( \Sigma_{\Lambda^*} \Sigma_{F^*} \) are distinct, where \( \Sigma_{\Lambda^*} \) and \( \Sigma_{F^*} \) are defined in Assumption D.

Based on the previous assumptions, Theorem 1 states the asymptotic distributions of the factor estimator \( \hat{F}_t \), loading estimator \( \hat{\lambda}_i \) and the common component estimator \( \hat{C}_{i,t} = \hat{F}_t' \hat{\lambda}_i \).

**Theorem 1: Asymptotic Distribution of the Unfeasible GLS Estimator**

Suppose that Assumptions A-F hold. Then,

(i) \( \sqrt{N} \left( J^{-1} \hat{F}_t - F_t \right) \stackrel{d}{\to} N \left( 0, W_{\hat{F}_t} \right) \) with \( W_{\hat{F}_t} = \Sigma_{\Lambda^*}^{-1} \Omega_t \Sigma_{\Lambda^*}^{-1} \)

(ii) \( \sqrt{T} \left( J \hat{\lambda}_i - \lambda_i \right) \stackrel{d}{\to} N \left( 0, W_{\hat{\lambda}_i} \right) \) with \( W_{\hat{\lambda}_i} = \Sigma_{F^*}^{-1} \Psi_i \Sigma_{F^*}^{-1} \)
where $J$ is a rotation matrix, $\Sigma_{F_*}$ and $\Sigma_{\Lambda_*}$ are defined in Assumption D and $\Omega_t$ and $\Psi_i$ are defined in Assumption E. Moreover, defining $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$, then under Assumptions A-E:

(iii) $\delta_{NT} \left( \hat{C}_{i,t} - C_{i,t} \right) \overset{d}{\rightarrow} N \left( 0, W_{\hat{C}_{i,t}} \right)$ with $W_{\hat{C}_{i,t}} = \lim_{N,T \to \infty} \delta_{NT}^2 \left[ \frac{\lambda_i' W_i \hat{F}_t}{N} + \frac{F_i' \hat{W}_1 F_i}{T} \right]$.

It is well known that the estimators of factors and factor loadings are only identified up to rotation since $F_t' \lambda_i$ is observationally equivalent to $(F_t J^{-1})(J \lambda_i)$. Following Bai and Ng (2002), the factors and factor loadings in Theorem 1 are rotated using

$$J = \frac{\Lambda' (\Theta' \Theta)^{-1} \Lambda F' (\Phi \Phi')^{-1} \hat{F}}{N \left( \Phi_t' \Phi_t \right) \left( \Theta_i' \Theta_i \right) \left( F_t' (\Phi \Phi')^{-1} F_t \right)}$$

where $V_{r,NT}$ is a diagonal matrix consisting of the first $r$ eigenvalues of $\frac{1}{NT} Y Y'$ in decreasing order. Having established the asymptotic distributions for the unfeasible GLS estimators under factor stationarity, we turn now to feasible estimators. The unknown matrices $\Phi$ and $\Theta$ need to be replaced by consistent estimators.

**Assumption G: Consistent Estimation of $\Phi$ and $\Theta$**

$\tilde{\Theta} \overset{p}{\rightarrow} \Theta$ and $\tilde{\Phi} \overset{p}{\rightarrow} \Phi$ as $N, T \to \infty$.

The extension of Theorem 1 to feasible estimators under the additional assumption G is straight-forward and is stated in Theorem 2:

**Theorem 2: Asymptotic Distributions of the Feasible GLS Estimators**

If Assumptions A-G hold, then the feasible GLS estimators have the same asymptotic distributions as their unfeasible counterpart in Theorem 1.

### 4.2 Asymptotic Efficiency

In this subsection, we study the relative asymptotic efficiency assuming that the covariance structure is correctly specified. Under Assumptions A and B, $\Omega_t$ and $\Psi_i$ reduce to $\Phi_t' \Phi_t$ and $\Sigma_{\Lambda_*}$ and $\Theta_i' \Theta_i$ and $\Sigma_{F_*}$ respectively such that:

(i) $W_{\hat{F}_t} = \lim_{N,T \to \infty} N (\Phi_t' \Phi_t) \left( \Lambda' (\Theta' \Theta)^{-1} \Lambda \right)^{-1}$

(ii) $W_{\hat{\lambda}_i} = \lim_{N,T \to \infty} T (\Theta_i' \Theta_i) \left( F_t' (\Phi \Phi')^{-1} F_t \right)^{-1}$
For the sake of comparison, we reformulate Theorem 1-3 of Bai (2003) in order to compare directly the asymptotic variances of the PCA estimators with those of the GLS estimators. The respective asymptotic variances are:

(i) \( W_{\tilde{\lambda}_i} = \lim_{N,T \to \infty} \delta_{NT}^2 \left[ (\Phi_{\lambda}' \Phi_{\lambda}) \lambda_i (\Lambda' (\Theta' \Theta)^{-1} \Lambda)^{-1} \lambda_i 
+ (\Theta' \Theta) F_t' (F' (\Phi' \Phi')^{-1} F) F_t \right] \).

(ii) \( W_{\tilde{F}_t} = \lim_{N,T \to \infty} N (\Phi_{\lambda}' \Phi_{\lambda}) \lambda_i (\Lambda' (\Theta' \Theta)^{-1} \Lambda)(\Lambda' \Lambda)^{-1} \).

The following theorem is similar to Theorem 2 of Breitung and Tenhofen (2011) and shows that the GLS estimators are generally more efficient than the PCA estimators if the temporal and cross-sectional dependence structure of the error \( e_{it} \) is correctly specified.

**Theorem 3: Asymptotic Efficiency of the unfeasible GLS Estimator**

If Assumptions A-F hold, then the GLS estimators of \( \lambda_i \) and \( F_t \) are asymptotically more efficient than the respective PCA estimators. The difference of the respective asymptotic covariance matrices is positive semi-definite.

An analog theorem can be stated for the feasible estimator using Assumption G in addition. For \( r = 1 \), we define the relative efficiency by the ratio of the asymptotic variance of the GLS estimator and the asymptotic variance of the PCA estimator. Figures 1a, 1b and 1c plot the relative efficiency of the estimators of \( F_t \) and \( \lambda_i \) and \( C_{i,t} \) as a function of the time and cross-correlation parameters in a 3-D plot under AR(1) specification.

---

\(^1\)Originally, Bai’s factors and loadings are based on the rotation: \( H = \frac{\Lambda' \Lambda}{N} \frac{E' E}{T} V_N^{-1} \) with \( V_N \) being a diagonal matrix consisting of the first \( r \) eigenvalues of \( \frac{1}{NT} XX' \).
Figure 1a shows that the ratio of the asymptotic variances of the GLS estimator and the PCA estimator of $\lambda_i$ does not depend on $\theta$. In contrast, it strongly varies with the values of $\phi$. For values of $\phi$ close to 0, the asymptotic variances of the GLS and the PCA estimator are approximately the same. The larger $\phi$ grows in absolute terms, the smaller the asymptotic variance of the GLS estimator compared to the asymptotic variance of the PCA estimator. Figure 1b is the analog of Figure 1a. The ratio of the asymptotic variances of the GLS estimator and the PCA estimator of $F_t$ does not depend on $\phi$, but strongly varies with $\theta$. Figure 1c plots the ratio of the asymptotic variances of the GLS estimator and the PCA estimator of $C_{i,t}$. The asymptotic variance of the PCA estimator explodes compared to asymptotic variance of the GLS estimator when $\theta$ and $\phi$ approach extreme values.

4.3 Asymptotic Distribution Under Factor Non-stationarity

Next, we consider the asymptotic distribution of the estimators when $F_t$ is a vector of integrated processes, i.e. $F_t \sim I(1)$. It must be noted that the unit root formulation of $F_t$ violates Assumption D and hence Theorem 1 does not apply. To derive the asymptotic distributions of the GLS estimators$^2$, one needs to reformulate Assumptions B-F in the spirit of Bai (2004) while assuming no cointegration among the common stochastic trends$^3$. The modified assumptions are stated in the appendix, that are needed for the following theorem:

**Theorem 4: Asymptotic Distribution of the Unfeasible GLS Estimator**

Suppose that Assumptions A,B'-F' hold. Then,

(i) $\sqrt{N} \left( J^{-1} \hat{F}_t - F_t \right) \xrightarrow{d} N \left( 0, \Sigma_{A*}^{-1} \Omega_t \Sigma_{A*}^{-1} \right)$

(ii) $T \left( J \hat{\lambda}_i - \lambda_i \right) \xrightarrow{d} \Sigma_{F_o}^{-1} \left( \int B_u dB_{\varepsilon(i)}^t \right)$

where $J$ is the rotation matrix given in equation (18), $\Sigma_{A*}$ is defined in Assumption D, $\Sigma_{F_o}$ is the distributional limit of $\frac{1}{T^2} F'(\Phi\Phi')^{-1} F$ and $\Omega_t$ and $\int B_u dB_{\varepsilon(i)}^t$ are defined in Assumption E'.

$^2$we standardize by $\frac{1}{T^2} F'(\Phi\Phi')^{-1} \hat{F} = I_r$ instead.

$^3$The no cointegration property is preserved in the GLS transformation.
The proof of Theorem 4 is given in the Appendix. In the spirit of section 4, one can also derive the asymptotic distribution of the feasible GLS estimator using Assumption G as well as the distribution of the common component under factor non-stationarity. The asymptotic comparison with the PCA estimator can be done in a similar manner as presented in subsection 4.2. Having obtained these extensions for the stationary case, the derivations for the non-stationary case are straightforward and hence omitted.

5 Monte Carlo Analysis

This section reports simulation results that illustrate the estimation efficiency of the feasible GLS estimators and the estimators of the method of principal components. Since the GLS estimates are based on the standardization \( \frac{1}{T} \tilde{F}' \left( \tilde{\Phi} \tilde{\Phi}' \right)^{-1} \tilde{F} = I_r \) while the method of principal components uses the standardization \( \frac{1}{T} \tilde{F}' \tilde{F} = I_r \), the estimators cannot be directly compared. Therefore, we focus on the estimation of the common component \( C_{i,t} = F_t' \lambda_i \) for the sake of comparison. More precisely, the comparison is based on the empirical mean squared error (MSE) of the estimated common components, suggested by Choi (2012), of the feasible GLS, \( \hat{C}_{i,t} = \hat{F}_t' \hat{\lambda}_i \), and of the method of principal components, \( \tilde{C}_{i,t} = \tilde{F}_t' \tilde{\lambda}_i \). Moreover, we compare the results to Breitung and Tenhofen (2011)’s 2-step GLS estimator, referred as B&T estimator subsequently, that can account for time correlation, but not for cross-correlation.

We use \( S = 10,000 \) iterations to calculate the empirical MSEs. We only report the statistic for \( t = [T/2] \) and \( i = [N/2] \), where \([a]\) is the integer closest to \(a\); other values of \((t,i)\) give similar results and are thus omitted. We report empirical MSEs, rather than sample variances because biases may also matter in finite samples in estimating the common components.

We study the estimators’ performance along different dimensions. In particular, we explore different sample sizes, factor stationarity and non-stationarity as well as alternative covariance structures. The following subsection describes the data generation in detail.

\(^4\)To be precise, it can account for time correlation and heteroskedasticity.
5.1 Simulation Procedure

For the simulation analysis, the following sample sizes are considered: \( N = 50, 100, 200 \) and \( T = 50, 100, 200 \). For each \((N, T)\) combination, three cases are examined:

(i) **factor stationarity** \((r = 1)\): \( F_t \overset{iid}{\sim} N(0, 1) \)

(ii) **factor non-stationarity** \((r = 1)\): \( F_t = F_{t-1} + \eta_t, \; \eta_t \overset{iid}{\sim} N(0, 1) \)

(iii) **factor stationarity and non-stationarity** \((r = 2)\): \( F_{t,1} \overset{iid}{\sim} N(0, 1) \) and \( F_{t,2} = F_{t-1,2} + \eta_t, \; \eta_t \overset{iid}{\sim} N(0, 1) \)

In all three cases the factor loading is chosen to follow the law of standard normal distribution, \( \lambda_t \overset{iid}{\sim} N(0, 1) \). Each case is combined with three covariance structures:\(^5\)

(i) **Moving Average Process** (MA(1)): In this specification, the covariance of \( vec(c) \) has a Kronecker product form with \( \Theta'\Theta \) and \( \Phi\Phi' \) modeled by MA(1)’s. The MA(1) covariance structure is modeled in the following way:

\[
\Phi\Phi' = \begin{pmatrix}
1 & \phi & \cdots & 0 \\
\phi & 1 & \cdots & \vdots \\
\vdots & \vdots & \ddots & \phi \\
0 & \cdots & \phi & 1
\end{pmatrix}, \quad \Theta'\Theta = \begin{pmatrix}
1 & \theta & \cdots & 0 \\
\theta & 1 & \cdots & \vdots \\
\vdots & \vdots & \ddots & \theta \\
0 & \cdots & \theta & 1
\end{pmatrix}
\]

with parameters set to \( \phi = 0.3 \) and \( \theta = 0.3 \). Figure 2a illustrates the MA(1) covariance matrix type by plotting the matrix \( \Phi\Phi' \) in a 3-D graph.

(ii) **Autoregressive Process** (AR(1)): This specification is similar two the first one, however \( \Theta'\Theta \) and \( \Phi\Phi' \) are modeled by AR(1)’s instead. In particular, we use the following:

\[
\Phi\Phi' = \begin{pmatrix}
1 & \phi & \cdots & \phi^{T-1} \\
\phi & 1 & \cdots & \phi^{T-2} \\
\vdots & \vdots & \ddots & \vdots \\
\phi^{T-1} & \phi^{T-2} & \cdots & 1
\end{pmatrix}, \quad \Theta'\Theta = \begin{pmatrix}
1 & \theta & \cdots & \theta^{N-1} \\
\theta & 1 & \cdots & \theta^{N-2} \\
\vdots & \vdots & \ddots & \vdots \\
\theta^{N-1} & \theta^{N-2} & \cdots & 1
\end{pmatrix}
\]

\(^5\)We abstract from heteroskedasticity.
where parameters are set to $\phi = 0.5$ and $\theta = 0.5$. Figure 2b illustrates the $AR(1)$ covariance matrix type by plotting the matrix $\Phi\Phi'$ in a 3-D graph.

(iii) Misspecification ($MS$): Last, we consider the case where the covariance of $vec(e)$ has a Kronecker product structure, but $\Theta'\Theta$ and $\Phi\Phi'$ are both complicated matrices. They are the average of four positive definite MATLAB® gallery test matrices ($lehmer$, $minij$, $gcdmat$, $prolate$) that are standardized such that each $e_{i,t}$ has unit variance. Figure 2c illustrates the unconventional form of $\Phi\Phi'$. If the structure of $\Theta'\Theta$ and $\Phi\Phi'$ were known, the transformed GLS model could be correctly specified. However, the structure of $\Theta'\Theta$ and $\Phi\Phi'$ is unknown, though can be reasonably approximated by an $AR(1)$ covariance structure. Using an $AR(1)$ approximation, the GLS estimator is based on a misspecified model.

The simulations are carried out for the Gaussian distribution$^6$. Under Gaussian modeling the error matrix $e$ follows a matrix normal distribution with probability density function:

$$f(e|\Theta, \Phi) = (2\pi)^{-NT/2}|\Theta'\Theta|^{-T/2}|\Phi\Phi'|^{-N/2}exp\left[-\frac{1}{2}tr\left((\Theta'\Theta)^{-1}e'(\Phi\Phi')^{-1}e\right)\right].$$

A realization from this distribution can be drawn by simulating $e = \Phi \varepsilon \Theta$, where $\varepsilon$ is an $T \times N$ matrix with independent elements, that are drawn from a standard normal distribution.

5.2 Results

This subsection reports the simulation results. We start with examining the validity of the approximate normal distribution of the feasible GLS estimators in finite samples.

Result 1: Approximate Normality

The asymptotic approximation works well for the feasible GLS estimators when sample size is modest.

---

$^6$We also ran simulations with $e$ drawn from a matrix-t distribution with 10 degrees of freedom. The results are similar to the Gaussian case.
Figure 3 plots the non-parametrically\(^7\) estimated densities of the estimated common components with AR(1)-specification. This suggests that the asymptotic approximation works well for the GLS estimators when sample size is modest \((N = 50, T = 50)\). With regards to Theorem 3, we find the density of the GLS-estimated common component to be less disperse. The estimated variance for the PCA-estimated common component is 0.044, whereas the estimated variance for the GLS-estimated common component is 0.024 (see Table 1)\(^8\). Thus, using the GLS approach one has a precision gain of \(1 - \frac{0.024}{0.044} = 45.45\%\). This illustration suggests large efficiency gains of the feasible GLS over the method of principal component. This hypothesis is strongly confirmed by the results in Table 1. The table summarizes the efficiency of the GLS, B&T and the PCA estimator by reporting the empirical MSE of the estimated common component scaled by 100.

**Result 2: Efficiency Gains**
Under time and cross-sectional correlation, the GLS method performs best, whereas the method of principal components performs worst in terms of MSE efficiency.

This result can be established by comparing Panel A, B and C in Table 1. It comes at no surprise that the B&T estimator performs worse than the GLS estimator, but better than the PCA estimator since it can account for time, but not for cross-correlation. The third simulation result is in correspondence with the sample size.

**Result 3: Sample Size**
The MSE of the estimated common component of all three estimators tends to decrease as the sample size increases.

This result is coherent with intuition: the larger the sample, the more precise the estimators and thus the smaller the MSE.

**Result 4: Stationarity**
All three estimation methods perform better under factor stationarity.

\(^{7}\)Nadaraya-Watson kernel regression
\(^{8}\)The estimated variances coincide with their empirical MSE suggesting that potential bias are negligible.
To establish simulation result four, one needs to compare column 3 with 6, 4 with 7 and 5 with 8 of Table 1 for each estimator. The empirical MSE of the GLS-estimated common component is fairly smaller under factor stationarity than under non-stationarity. This pattern can also be identified for the PCA estimator and the B&T estimator.

Summarizing the main simulation results, we find that the GLS estimator outperforms the PCA and the B&T estimator for sample sizes up to $(N, T) = (200, 200)$ when time correlation and cross correlation is present in the errors. It operates best under factor stationarity. Efficiency gains differ greatly across specifications depending on the extend of correlation present in the errors.

6 Application: Lee-Carter Revisited

6.1 The Lee-Carter Model

Lee and Carter (1992) proposed a stochastic mortality model that has become the “leading statistical model of mortality in the demographic literature” (Deaton and Paxson, 2001) and has been employed by the US Census Bureau and the United Nations (Girosi and King, 2007). They apply the method of principal component to estimate a model of the US central death rate. The central death rate of age group $i$ in calendar year $t$, denoted by $m_{i,t}$, plays a key role in actuarial science as $\exp(-m_{i,t})$ is the estimate of the conditional probability of surviving (Brouhns et al., 2005). One can easily use the concept of conditional probability of surviving to derive important actuarial variables such as expected remaining lifetime. Therefore, researchers typically focus on modeling and forecasting the central death rate. Figure 4a plots the central death rate for the Netherlands in one-year intervals between 1925 and 2009 ($T = 85$) covering the age groups from 0 to 94 ($N = 95$). Lee and Carter propose a model that sets the log central death rate in a linear relationship to the parameters:

$$log(m_{i,t}) = a_i + b_i k_t + \xi_{i,t} \quad (19)$$

where $a_i$ and $b_i$ are age-specific parameters. It is important to note that $k_t$ is not a regressor, but also a parameter frequently interpreted as the time-varying mortality
factor. Whereas $exp(a_i)$ represents the general shape of the force of mortality schedule, $b_i$ captures the age-group-specific sensitivity to the time-varying mortality factor $k_t$.

In a first step, Lee and Carter estimate $a_i$ by the time-series mean of age group $i$: $\hat{a}_i = \frac{1}{T} \sum_t \ln(m_{i,t})$. Removing the time-series means from the data, they perform a PCA on the demeaned data in order to extract a single factor with corresponding loading, which are the estimates of $\{k_t\}_{t=1}^T$ and $\{b_i\}_{i=1}^N$ respectively. To be aligned with the standard notation of the PCA literature, we rewrite the model in matrix form as in equation (2), where $X$ denotes a $T \times N$ matrix of observed time-series demeaned log central death rate, $F = (k_1, \ldots, k_T)'$ represents the latent factor of dimension $T \times 1$, $\Lambda = (b_1, \ldots, b_N)'$ is the $N \times 1$ vector of factor loadings and $e$ is an $T \times N$ matrix of idiosyncratic errors. Lee and Carter set the number of factors $r$ to be 1. Other researchers considered the possibility of $r > 1$ such as Renshaw and Haberman (2005), who assume $r = 2$. Rather than assuming ex-ante the number of factors, we follow the path of inferring $r$ from the data. To estimate consistently the number of factors for the Dutch data set, we employ a formal statistical procedure proposed by Bai and Ng (2002, 2004). Three different criterion functions, stated in equation (12) of Bai (2004), are utilized all suggesting $r = 1$; it corresponds to a single factor model that is in line with Lee and Carter. Applying Lee and Carter’s approach to the Dutch data set, one gets estimates for the factor and loading that are represented by green dashed lines in Figure 6. Figure 6a illustrates the estimated age-group-specific sensitivity to the time-varying mortality factor. It points out that over the period from 1925 to 2009, the largest gains in terms of mortality reduction occurred at infant ages, that are linked to improvements in nutrition, birth weight and advances in pediatric medicine (Mohangoo et al., 2013). Figure 6b suggests that the general time-schedule of mortality rates have decreased over time. There is a striking peak in the estimated time-schedule of mortality rates that is clearly related to World War II. For the sake of illustration, the following timeline presents the key events of 1940 to 1945 and plots the corresponding mortality factor estimates in greater detail. The World War II period has been well-documented for the Netherlands by De Jong (1972): the Netherlands were invaded by Nazi Germany in May 1940. After the bombing of Rotterdam, that demanded more than 800 civilian causalities, the Netherlands officially surrendered. Resistance against the Nazi’s occupation and
the imposed forced labor, but above all the tragic deportation of the Jewish Dutch increased the number of victims. The number of deaths peaked in the so-called *hunger winter* of 1944 – 1945, after Nazi Germany cut off all food shipments to the western provinces, which led to severe malnutrition and wide-spread starvation. The deliberation occurred in May 1945 with Canadian forces entering the Netherlands from the east.

Except for the war years, the mortality factor seems to follow a stochastic trend. To test for a unit-root in the common factor in the Dutch data, we follow Bai and Ng (2004) by employing an augmented Dickey-Fuller test on the estimated factor. With a p-value of 5.81%, we accept the null hypothesis of factor non-stationarity against the alternative of factor (trend-) stationarity.

### 6.2 Heteroskedasticity, Cross- and Time Correlation

Several authors have commented on non-random patterns in the residuals such as time and cross-correlation after having fit the Lee-Carter model to different data sets (see Renshaw and Haberman (2005); Koissi et al. (2006); Dowd et al. (2010)). Figure 4b illustrates these findings by plotting the Lee-Carter residuals for the Netherlands. The *waving-behavior* in the Lee-Carter residuals suggests that the errors exhibit some form of dependence. A formal test of cross- and time-correlation strongly rejects the null hypothesis of random errors. For instance one can apply a Ljung-Box test for autocorrelation for each cross-section and each time series. Doing so, the null hypothesis of independence is 82-out-of-85 times rejected at 5% for the time series and 90-out-of-95 times rejected at 5% for the cross-sections.

The question arises why time and cross-correlation is persistent in the model’s residuals. Some researchers have argued in favor for additional factors, but since the first factor usually explains a large fraction of the data’s variation (95.73% for Dutch data), the benefit of adding factors is minor. Furthermore, adding more factors seems only to mitigate the issue slightly. This becomes evident in the Dutch data set: by including two more factors \((r = 3)\), one is able to explain additional 2.32% of the data’s variation, however, the Ljung-Box test still rejects null hypothesis of independence in 69 out of 85 cases at 5% for the time series and in 46 out of 95 cases at 5% for the cross-sections.
A possible explanation for time and cross-correlation is based on the nature of the data. The Human Mortality Database (2015) provides a detailed methods protocol on how central death rates are computed using raw data from Census and other sources. The central death rate for age group $i$ in year $t$, $m_{i,t}$, is calculated as the ratio of the estimated death counts $D_{i,t}$ and some estimated population reference value $E_{i,t}$.

$$m_{i,t} \overset{\text{def.}}{=} \frac{D_{i,t}}{E_{i,t}}$$

This estimated population reference value $E_{i,t}$ accounts for the number of people that were exposed to mortality risk at age $i$ in year $t$ and is therefore called exposure-to-risk. One source of correlation is the measurement of the exposure-to-risk $E_{i,t}$. It is calculated in two steps. In a first step a so-called intercensal survival method is applied to obtain yearly January 1st population estimates for each age group from Census data. Since the exposure-to-risk for age group $i$ in year $t$ refers to the period $[t, t + 1)$, the mean of the January 1st population estimates of the year $t$ and $t + 1$ is taken. Doing this calculation recursively, this procedure induces time correlation in the data since the January 1st population estimates of year $t + 1$ are used to estimate the exposure-to-risk for the years $t$ and $t + 1$.

Moreover, Census data is collected only every 5 or 10 years. Therefore the intercensal survival method calculates January 1st population estimates by following cohorts of the previous Census date and then subtracts the yearly death counts. Doing this recursively, one obtains some population estimates for the next Census date, that usually do not coincide with the Census raw data. The difference is mainly due to migration flows. In order to equalize the population estimates at the Census date to those of the Census raw data, migration flows are equally split over the intercensal period. This procedure further induces correlation in the data.

Another convincing argument for dependence are cohort effects, that have been well-documented in the literature (Renshaw and Haberman, 2005). They result from the fact that one follows individuals over time. For instance, a person that is aged 20 in year 2000 is likely to be observed in year 2001 at age 21. The presence of such cohort effects imply correlation among different units at different times, that

---

9...plus an adjustment term that accounts for the timing of deaths, which is typically 0 except for infant ages
cannot be captured by models with a Kronecker sum covariance structure (see discussion of Assumption A). This favors the Kronecker product covariance structure, which can account for such type of spatial correlation. To model the dependency structure and to estimate $F$ and $\Lambda$, we choose a feasible GLS approach as described in section 3.2. We estimate $\Phi$ and $\Theta$ using the parametric structure described in (4) with $\phi_t = \phi^i$ and $\theta_i = \theta^i$. The implied covariance structure is more flexible than the $AR(1)$ structure described in Section 6.1 since it additionally allows for cross-sectional heteroskedasticity. Its parameter set $(\phi, \theta, \sigma_1, \ldots, \sigma_N)$ increases as $N$ increases, however consistency of the GMM estimates is established since the time dimension $T$ grows large as well. Accounting for heteroskedasticity, the sample partial autocorrelation functions suggest that the adjusted residuals exhibit an $AR(1)$ covariance structure. This can be verified by checking the sample partial autocorrelation functions (PACFs) of each time and cross-sectional series. For illustration, Figure 5 plots the sample PACF of the heteroskedasticity-adjusted residual series of age group $i = 0$. It indicates a time-series correlation structure with parameter $\phi$ of approximately 0.80. In fact, we estimate $\hat{\phi} = 0.83$ and $\hat{\theta} = 0.59$. For the sample size considered in the application ($N = 95$, $T = 85$), simulations show that the GLS estimator based on a heterogeneous $AR(1)$ structure outperforms the PCA estimator even under misspecification, where $\Phi\Phi'$ and $\Theta'\Theta$ have an MS structure illustrated in Figure 2c. In particular under unit-root specification and normality, the empirical MSE of the PCA-estimated common component is 0.185 whereas the empirical MSE of the GLS-estimated common component is 0.141.

The GLS estimates of the mortality factor $F_t$ and the factor loading $\lambda_i$ are plotted in solid blue in Figure 6. They are rotated (scaled) in order to compare them with the PCA estimates. For the case of the Netherlands, the GLS decomposition explains 95.23% of the data’s variation, which is 0.5 percentage points smaller than the respective PCA value. By construction, the PCA decomposition attains the highest value of explained sum of squares (ESS). The discrepancy of the GLS and PCA based ESS suggests that the PCA estimates slightly overfit the model. It can be verified that the PCA estimates follow the general pattern of the GLS estimates. However, considerable differences occur at young ages in Figure 6a, where the GLS estimate of $\lambda_i$ is sizable smaller than the corresponding PCA estimate. Regarding the estimates of the time-varying mortality factor for the period of World War II, the PCA estimate.
deflects stronger upward than the GLS estimate suggesting an milder war effect than presumed by using the PCA method.

6.3 Forecasting Life Expectancy

Having used a GLS approach to fit the Lee-Carter mortality model to the Dutch data set, it is natural to consider a forecasting exercise in order to compute life expectancy. Life expectancy is defined as the average number of years lived by a group of individuals of same age. For extinct cohorts it is completely determined and can simply be calculated by averaging the ages of death over all cohort members. However, for cohorts with some survivors it can only be forecasted using information based on past mortality experience. As Lee and Carter (1992) suggest, we adjust the life tables for the evolution of mortality by projecting the mortality factor of the past into the future. The expected remaining lifetime of a person aged \( i \) in year \( t \), denoted by \( E_{i,t} \), can be directly calculated from the central death rates. Under the assumption that the force of mortality is constant within a year/age cell, Brouhns et al. (2005) derive the following formula for the life expectancy:

\[
E_{i,t} = \frac{1 - p_{i,t}}{m_{i,t}} + \sum_{k>0} \left( \prod_{j=1}^{k-1} p_{i+j,t+j} \right) \frac{1 - p_{i+k,t+k}}{m_{i+k,t+k}}
\]

(20)

where \( p_{i,t} = \exp(-m_{i,t}) \) denotes the probability of an individual aged \( i \) to survive to year \( t + 1 \), conditional on having survived to year \( t \). To project life tables one needs to forecast the central death rate. Lee and Carter (1992) propose to fit an ARIMA model to the time-varying mortality factor. Doing so for the Dutch data set, we exclude the war year observations (1940 – 1945) from the sample, which is the usual approach in the literature. Using the Box-Jenkins procedure, we find an ARIMA(1,1,0) to fit best to both, the GLS-estimated and the PCA-estimated, mortality factor. The parameters estimates and their estimated standard deviations can be found in Table 2. The drift parameter \( d \) is key and the ARIMA estimates based

\[10\] This is different from the standard period age life expectancy measure, which assumes that the current age-specific mortality rates will continue in the future. The stationarity assumption seems not to be plausible in regard to Figure 6b.

\[11\] Due to data availability, we assume that live ends at age 95, i.e. \( p_{95,t} = 0 \) for all \( t \). This assumption generates a downward-bias of \( e_{i,t} \), however the bias is negligible.
on the GLS- and the PCA-estimated factors are very close. Difference occur in the estimated auto-regressive parameter $\alpha$, which is statistically significant. Last, the standard deviation $\sigma$ of the error component is slightly larger in the ARIMA model fitted to the GLS-estimated factor. Its difference, however, is not statistically significant at 5%. Instead of relying on analytical solutions for point forecasts, we follow a bootstrap method similar to the one proposed by Brouhns et al. (2005) in order to take into account sampling errors in the parameters. For the life expectancy forecasts based on the GLS estimates we proceed as follows:

For $s = 1, \ldots S$

1. draw values for $(F_{t-1}^s, F_T^s)$ from the asymptotic distribution of $(\hat{F}_{T-1}, \hat{F}_T)$
2. draw values $(d^s, \alpha^s, \sigma^s)$ from the asymptotic distribution of $(\hat{d}, \hat{\alpha}, \hat{\sigma})$
3. draw a path $\{F_{t}^s\}_{t=T+1}^{T+h}$ by resampling from the ARIMA errors
4. draw a realization $\{\lambda_i^s\}_{i=1}^N$ from the asymptotic distribution of $\{\lambda_i\}_{i=1}^N$
5. calculate $X_{i,T+\tau} = F_{T+\tau}^s \lambda_i^s + e_{i,T+\tau}^s$, where $e_{i,T+\tau}^s$ is calculated by resampling from $\hat{\epsilon}$ and obtaining $e^s = \tilde{\Phi}\hat{\epsilon}\tilde{\Theta}$
6. obtain $\{m_{i,t}^s\}_{i,t=1}^{N,T+h}$ and calculate $F_{i,t}^s$

Finally, take the sample average across simulations, where we choose $S = 5,000$.

In a similar fashion, we obtain life expectancy forecasts based on the PCA estimates, where we resample from $\hat{\epsilon}$ directly instead through $\hat{\epsilon}$. The results for a forecast horizon until year 2060 are summarized in Table 3. The table can be read as follows: an individual at age 30 in the year 2030 has an expected remaining lifetime of 54.43 years according to the PCA based forecast, whereas only 53.40 years according to the GLS based forecast. Comparing the different forecasts, it is striking that the PCA based forecasts are generally larger up to a difference of 1.41 years. Next, we have a closer look at life expectancy of an unborn. Figure 8 plots the evolution of the

\[12\text{In the application, resampling from } \hat{\epsilon} \text{ directly, without accounting for the dependence in the error, does not alter the qualitative results.}\]
life expectancy forecasts of an unborn over time. It must be noted that both life expectancy point forecasts exhibit a sharp decrease in the year 1945. This drop can be interpreted as follows: people born in the *hunger winter* had a lower life expectancy due to malnutrition and the risk of starvation. However, the GLS based point forecast for the year 1945 is approximately 3.5 years higher than the corresponding PCA based point forecast. Prior to the year 1945 all PCA point forecasts are lower than the GLS based point forecast of the life expectancy of an unborn. With increasing years, the PCA point forecasts exceed the GLS point forecasts. In the year 2015, this difference amounts to approximately 1.2 years increasing further up to 1.4 years by the year 2060. Having chosen a bootstrap approach, it is straight-forward to obtain prediction intervals for the life expectancy of an unborn using the sample quantiles across simulations. It must be noted that the PCA based sample quantiles are not reliable for interval prediction since the correlation structure of the errors has not been taken into account. In Figure 8 we have plotted the GLS based 5% and 95% sample quantiles. For the postwar period, we cannot reject the validity of the PCA forecast within the time period up to year 2060. However, for the pre-war period, the validity of the PCA forecast can be rejected at 5% (one-sided test).

In a sensitivity analysis, we explore the driving factors leading to the results presented in Table 3 and Figure 8. It turns out that the results are robust to a large differences in the estimated values for $\alpha$ and $\sigma$. The results, however, are sensitive to differences in the estimated ARIMA drift $d$. The difference in the third digit results in the divergence of the point forecasts in Figure 8. This emphasizes the importance of precise estimation of the time-varying mortality factor $F_t$ by using the GLS procedure instead of PCA; Precise estimation is key in order to accurately estimate the drift component in the second step.

### 6.3.1 Gender Gap

It is a well known fact that mortality and life expectancy vary by gender; in most countries, men live shorter lives than women. To explore the gender gap for the Netherlands we repeat the estimation and forecasting exercise using gender-specific central death rates. The data has been provided by the HMD and the same period and age groups are selected as in the gender-neutral case (years: 1925 – 2009; ages:
Figure 9 plots the estimated gender-specific factor and factor loadings based on the GLS and the PCA method, where cross- and time correlation is described in (4) with $\phi_t = \phi^i$ and $\theta_i = \theta^i$.

The factor and factor loading estimates are rotated for the sake of comparison. Figure 9a plots the estimates of the gender-specific mortality factors. It can be observed that not only the gender-specific GLS estimates are very similar to their respective PCA estimates, but also that the estimates for male and female tend to move together. A departure of this co-movement occurs for the World War II period, in which the male GLS and PCA based factor estimates peak significantly higher. The difference seems plausible in regard to the fighting against the Nazis that took place in 1944/45 on Dutch territory (e.g. battle of Arnhem) with support by Dutch men on the Allied side.

Figure 9b plots the estimates of the gender-specific factor loading. Again, one observes that the gender-specific GLS estimates are rather similar to their respective PCA estimates. An exception occurs at infant ages for the male estimates indicating that the disparity in Figure 6a is driven by men. Moreover, one finds generally lower GLS and PCA based estimates for male compared to female. This evidence suggests that women achieved larger gains in terms of mortality reduction. This is align with the literature that attributes the strong mortality improvements to the fact that the economic status of women has risen relative to men over the considered period, which in return benefited women’s health (Gorman and Read, 2007).

Using the bootstrap method, that has been previously described, we forecasts life expectancy of an unborn for the year 2015, where cross- and time correlation is modeled by an heterogeneous $AR(1)$ structure. The results are presented in Table 4.

The forecast values confirm the conjecture that women outlive men in the Netherlands. The gender disparity is quantified to be 6.13 years for the PCA based forecasts, whereas it is slightly smaller for the GLS based forecasts (5.85 years). Using the bootstrap standard errors we find the gender gap to be statistically significant at 1%. Interestingly, the gender-specific forecast values based on PCA and GLS differ by less than in the gender-neutral case (see Table 3). The question why this gender discrepancy arises has only been partially answered and is the focus of ongoing research. According to Gorman and Read (2007), there are socioeconomic, biological and behavioral explanations. From a behavioral standpoint, it appears that men
consume more (and more frequently) alcohol and tobacco, whereas "adult women under the age of 65 report more doctor visits" (Gorman and Read, 2007). Smoking is not only a key determinant of the gender gap but is also considered to be the driving factor of the gap’s shrinking size since men have been reducing smoking faster than women. This trend is also evident in the PCA and GLS based life expectancy forecasts. By 2060, the gender gap for an unborn reduces to 5.93 years according to the PCA based point forecasts and 5.77 years according to the GLS based point forecasts. Next, we turn to comparing life expectancy across countries.

6.3.2 Cross-Country Comparison

To compare gender-neutral life expectancy of the Netherlands with its neighbors, we select the BeNeLux states: (Belgium, Netherlands, Luxembourg) as well as Germany. For Belgium, HMD data is available until the year 2012, therefore we select the years from 1925 until 2012. For Luxembourg, the full HMD data is selected, which covers the period starting in 1960 and ending in 2009. Last, for Germany only data is available since the reunification of East and West Germany in 1990. Due to the small sample size, we use data for its western part instead, that is available for the period 1956–2011.

Life expectancy forecasts based on PCA and GLS for an unborn in the year 2015 for each country are obtained by employing the bootstrap method. Figure 10 summarizes the results. For the Netherlands, Luxembourg and West Germany, the PCA based life expectancy point forecast exceeds the corresponding GLS based forecast. The discrepancy is largest for the Netherlands with a difference of 85.46 – 84.20 = 1.26 years. In contrast, in the case of Belgium the GLS based life expectancy point forecast exceeds the corresponding PCA value by 83.20 – 81.93 = 1.27 years. Moreover, the countries’ ranking is the same whether ranked by the PCA based point forecasts or by the GLS based point forecast: for an unborn in the year 2015, life expectancy appears to be highest in Luxembourg and lowest in Belgium. In between, the western region of Germany ranks second and the Netherlands third. Linking the results to income, the results verify the cross-sectional relationship between life expectancy and

\[13\] Luxembourg is a small country with approximately 500,000 citizens (in 2013). For specific years and age combination, it appears that the central death rate is exactly zero such that Lee and Carter’s log-transformation is not possible. In that case we interpolate between years.
GDP per capita, that was first described by Preston (1975): it states that individuals born in wealthier countries, on average, can expect to live longer compared to those born in poor countries. Luxembourg, with a GDP per capita of $110,697 in 2013 (World Bank, 2015), has clearly the highest life expectancy according to the GLS and PCA based forecasts. On the other edge, Belgium with an GDP per capita of $46,978 in 2013 (World Bank, 2015), has the lowest life expectancy. The Netherlands with an average income of $50,793 in 2013 (World Bank, 2015) rank between Luxembourg and Belgium. Although Germany (full country) has only a GDP per captia of $46,269 in 2013 (World Bank, 2015), the western part is considered to be more economically developed than its eastern counterpart suggesting a significant higher GDP per capita for West Germany\textsuperscript{14}. To sum up, life expectancy forecasts vary widely across the BeNeLux states and West Germany. Their PCA and GLS based rankings coincide and are positively related to the level of economic development.

7 Conclusion

This paper studies the estimation of large-dimensional factor models with time and simultaneously cross-sectional dependence while allowing for heteroskedasticity. Under the assumption of multiplicative separability of the covariance matrix, we employ a GLS type approach to efficiently estimate the factor and factor loading space. We propose unfeasible GLS estimators of the factor and factor loading space and provide feasible counterparts, that are easy to implement. Under factor stationarity, the asymptotic distributions of the unfeasible GLS estimators are derived when both, N and T, grow large. Since the estimation of the covariance parameters does not affect the limiting distribution of the GLS estimators, the feasible ones are asymptotically as efficient as their unfeasible counterparts. In comparison with the method of principal components, we prove theoretically the relative efficiency of the GLS estimators over their corresponding PCA estimators under correct covariance specification. Moreover, a unit root specification is considered and the asymptotic distributions of the GLS estimators of the factor and factor loading space are derived. A thorough Monte Carlo analysis supports the asymptotic results. The simulation study shows substan-

\textsuperscript{14}According to the German development bank KfW, the former East German states have a per capita income that is 84% of the West German states (Matthews, 2014).
tial gains of the GLS estimators over the the method of principal components (PCA) in terms of precision in finite samples. Moreover, the proposed GLS estimator also outperforms Breitung and Tenhofen (2011)’s GLS estimator, which cannot account for cross-correlation. Results are robust against a wide range of misspecifications suggesting to favor the GLS approach in practice.

In application, we revisit the Lee-Carter model that uses the method of principal components to estimate a single latent factor with corresponding loading in a data set of log central death rates. For a Dutch data set, time and cross-correlation is shown to be present in the Lee-Carter residuals. An intuitive explanation for dependence is provided that is based on how central death rates are typically measured. In addition, cohort effects generate correlation among different units at different times, which suggest the multiplicative separability of the covariance structure. The GLS type estimation procedure is applied to the log central death rates to gain estimation precision on the mortality factor. Fitting an ARIMA model to the GLS-estimated factors, we find a lower drift estimate compared to using the PCA-estimated factors. However, they are not statistically different when taking the parameter uncertainty into account. Employing a bootstrap procedure that accounts for parameter uncertainty, we show that the Dutch life expectancy point forecasts based on GLS are substantially lower than the ones based on PCA. By the year 2060, the difference amounts to 1.41 years suggesting that the longevity problem in the Netherlands is less severe than usually pronounced. Differentiating by gender, we forecast that females born in the year 2015 live, on average, 5.85 years longer than men and that the gender gap narrows slightly by 2060. Comparing the BeNeLux states and Germany, we find substantial life expectancy differences that are coherent with Preston’s relationship between life expectancy and income.

Several methodological issues remain. One issue is to combine the ARIMA specification directly in the estimation stage. The implied dynamic factor model with latent ARIMA processes and time and simultaneously cross-sectional dependent errors overcomes the two-stage estimation procedure employed in estimating the Lee-Carter model. In that respect a spectral EM algorithm seems to be promising for estimation (see Fiorentini et al., 2014). Moreover, from a theoretical standpoint it is worthwhile to derive the asymptotic distribution of the GLS estimator when Assumption A (ii) is relaxed.
A Appendix: Proofs

A.1 Proof of Theorem 1

The proof of Theorem 1 consists of two parts. First, we obtain the asymptotic distribution of \( \hat{G}_t, \hat{\gamma}_i \) and \( \hat{G}_t', \hat{\gamma}_i \), where \( \Gamma = (\gamma_1, \ldots, \gamma_N)' \). In a second step, we derive the asymptotic distribution of \( \hat{F}_t, \hat{\lambda}_i \) and \( \hat{C}_{i,t} \) using the one-to-one mapping between \((G, \Gamma)\) and \((F, \Lambda)\).

Lemma 1: Asymptotic Distribution of the Unfeasible GLS Estimator

Suppose that Assumptions A-F hold. Then,

(i) \[ \sqrt{N} \left( J^{-1} \hat{G}_t - G_t \right) = J^{-1} V_{sNT}^{-1} \frac{\hat{G}_t G_t'}{\sqrt{N}} + o_p(1) \xrightarrow{d} N \left( 0, W_{\hat{G}_t} \right) \]

(ii) \[ \sqrt{T} (J \hat{\gamma}_i - \gamma_i) = J V_{sNT}^{-1} \frac{G_t G_t'}{T} \frac{\Gamma \varepsilon_i}{\sqrt{T}} + o_p(1) \xrightarrow{d} N \left( 0, W_{\hat{\gamma}_i} \right) \]

where \( J \) and \( V_{sNT} \) are defined in equation (18). Moreover, defining \( \delta_{NT} = \min \{ \sqrt{N}, \sqrt{T} \} \), then under Assumptions A-E:

(iii) \[ \delta_{NT} \left( \hat{G}_t' \hat{\gamma}_i - G_t' \gamma_i \right) = \delta_{NT} \left( \frac{\Gamma' \varepsilon_t}{\sqrt{N}} \right) \left( \frac{\Gamma' \varepsilon_t}{\sqrt{N}} \right)^{-1} \frac{\Gamma \varepsilon_t}{\sqrt{N}} \]

\[ + \delta_{NT} \left( \frac{G_t' G_t'}{T} \right) \left( \frac{G_t' G_t'}{T} \right)^{-1} \frac{G_t' \varepsilon_i}{\sqrt{T}} + O_p \left( \frac{1}{\delta_{NT}} \right) \xrightarrow{d} N \left( 0, W_{\hat{G}_t' \hat{\gamma}_i} \right) \]

Proof of Lemma 1: Assumptions A-E/F imply the application of Theorem 1, 2 and 3 of Bai (2003) where \( G \) is Bai’s \( F \), \( \Gamma \) is Bai’s \( \Lambda \) and \( \varepsilon \) is Bai’s \( e \).

Using Lemma 1, we can establish the asymptotic distribution of the unfeasible estimators of \( F_t \) and \( \lambda_i \) as stated in Theorem 1 noting that \( F_t = \sum_{s=1}^{T} \Phi_{t,s} G_s \) and \( \lambda_i = \sum_{j=1}^{N} \Theta_{j,i} \gamma_i \).

Proof of Theorem 1: As \( \sqrt{N} (J^{-1} \hat{G}_t - G_t) \) is asymptotically normally distributed under Assumption A-F by Lemma 1 (i), the weighted sum

\[ \sum_{s=1}^{T} \Phi_{t,s} \sqrt{N} (J^{-1} \hat{G}_s - G_s) = \sqrt{N} (J^{-1} \hat{F}_t - F_t) \]
is also asymptotically normally distributed. The asymptotic variance is:

\[
W_{\hat{F}_t} = \lim_{N,T \to \infty} J^{-1} V_{sNT}^{-1} \frac{\hat{G}' G}{T} \left[ \frac{\Gamma' \varepsilon' \Phi_t \varepsilon' \Gamma}{N} \right] \frac{G' \hat{G}}{T} V_{sNT}^{-1} J^{-1}
\]

\[
= \Sigma_{A*}^{-1} \Omega_t \Sigma_{A*}^{-1}
\]

Similarly, as \(\sqrt{T}(J \hat{\gamma}_i - \gamma_i)\) is asymptotically normally distributed under Assumption A-F by Lemma 1 (ii), the weighted sum

\[
\sum_{j=1}^{N} \Theta_{j,i} \sqrt{T}(J \hat{\gamma}_j - \gamma_j) = \sqrt{T}(J \hat{\lambda}_i - \lambda_i)
\]

is also asymptotically normally distributed. The asymptotic variance is:

\[
W_{\hat{\lambda}_i} = \lim_{N,T \to \infty} J^{-1} V_{sNT}^{-1} \frac{\hat{G}' G}{T} \frac{\Gamma' \varepsilon' \Theta_i \varepsilon' \Gamma}{T} \frac{\Gamma' \hat{G}}{T} \frac{G' \hat{G}}{T} V_{sNT}^{-1} J^{-1}
\]

\[
= \Sigma_{F*}^{-1} \Psi_i \Sigma_{F*}^{-1}
\]

Last, as \(\delta_{NT} \left( \hat{G}' \hat{\gamma}_i - G' \gamma_i \right)\) is asymptotically normally distributed under Assumption A-E by Lemma 1 (iii), the weighted sum

\[
\sum_{s=1}^{T} \sum_{j=1}^{N} \Phi_{t,s} \Theta_{j,i} \delta_{NT} \left( \hat{G}' \hat{\gamma}_j - G' \gamma_j \right) = \delta_{NT} \left( \hat{C}_{i,t} - C_{i,t} \right)
\]

\[
= \frac{\delta_{NT}}{\sqrt{N}} \lambda_i' \left[ \frac{\Gamma' \varepsilon' \Phi_t}{N} \right]^{-1} \frac{G' \varepsilon' \Theta_i}{\sqrt{T}} + \frac{\delta_{NT}}{\sqrt{T}} \frac{F_t' \left( G' \Gamma \right)^{-1} \frac{G' \varepsilon' \Theta_i}{\sqrt{T}}}{\sqrt{T}} + O_p \left( \frac{1}{\delta_{NT}} \right)
\]

is also asymptotically normally distributed. We consider the asymptotic variance by parts. The asymptotic variance of the first term is given by:

\[
\lim_{N,T \to \infty} \delta_{NT}^2 \frac{\delta_{NT}^2}{N} \lambda_i' \left[ \frac{\Gamma' \varepsilon' \Phi_t \varepsilon' \Gamma}{N} \right] \left( \frac{\Gamma' \varepsilon' \Phi_t \varepsilon' \Gamma}{N} \right)^{-1} \lambda_i
\]

\[
= \lim_{N,T \to \infty} \delta_{NT}^2 \frac{\delta_{NT}^2}{N} \lambda_i' \Sigma_{A*}^{-1} \Omega_t \Sigma_{A*}^{-1} \lambda_i
\]
Similarly, the asymptotic variance of the second term is given by:

\[
\lim_{N,T \to \infty} \frac{\delta^2_{NT}}{T} F_t \left( \frac{G' G}{T} \right)^{-1} \left[ \frac{G' \varepsilon \Theta_i \Theta'_j \varepsilon' G}{T} \right] \left( \frac{G' G}{T} \right)^{-1} F_t
\]

\[
= \lim_{N,T \to \infty} \frac{\delta^2_{NT}}{T} F_t' \Sigma_{F*} \Psi_i \Sigma_{F*}^{-1} F_t
\]

Last, we show that the covariance between both terms vanishes:

\[
\lim_{N,T \to \infty} \frac{\delta^2_{NT}}{T \sqrt{T} \sqrt{N}} \lambda_i' \left( \frac{\Gamma' \Gamma}{N} \right)^{-1} \left[ \sum_{s=1}^{T} \sum_{j=1}^{N} \Phi_{t,s} \Theta_{j,i} \left( \frac{\Gamma' \varepsilon_s \varepsilon_j G}{\sqrt{N} \sqrt{T}} \right) \right] \left( \frac{G' G}{T} \right)^{-1} F_t
\]

\[
= \lim_{N,T \to \infty} \frac{1}{\max\{N, T\}} \left( \lambda_i' \left( \frac{\Gamma' \Gamma}{N} \right)^{-1} \lambda_i \right) \left( F_t' \left( \frac{G' G}{T} \right)^{-1} F_t \right) = 0
\]

\[\square\]

### A.2 Proof of Theorem 2

The proof of Theorem 2 consists of two parts. Lemma 2 states an intermediate result needed for the applicability of Lemma 1:

**Lemma 2:** Under Assumptions A-G, \( \tilde{\Phi}^{-1} X \tilde{\Theta}^{-1} \) follows the same asymptotic distribution as \( \Phi^{-1} X \Theta^{-1} \).

**Proof of Lemma 2:**

\[
\tilde{\Phi}^{-1} X \tilde{\Theta}^{-1} = G \Gamma' + \varepsilon + (\tilde{\Phi}^{-1} - \Phi^{-1}) F \Lambda' (\tilde{\Theta}^{-1} - \Theta^{-1})
\]

\[
+ (\tilde{\Phi}^{-1} - \Phi^{-1}) \varepsilon (\tilde{\Theta}^{-1} - \Theta^{-1})
\]

where the last two terms vanish in probability under Assumption G as \( N, T \to \infty \).

This establishes the applicability of Lemma 1 when \( \Phi \) and \( \Theta \) are replaced by consistent estimator. It must be noted that the rotation is changed to

\[
\hat{J} = \frac{1}{NT} \left( \Lambda' (\hat{\Theta}' \hat{\Theta})^{-1} \Lambda \right) \left( F' (\hat{\Phi} \hat{\Phi}')^{-1} F \right) \hat{V}_{NT}^{-1}
\]
where $\hat{V}_{NT}$ is a diagonal matrix consisting of the first $r$ eigenvalues of
\[
\frac{1}{NT}\Phi^{-1}X\left(\hat{\Theta}'\hat{\Theta}\right)^{-1}X\Phi'^{-1}
\]
in decreasing order. In a second step, we prove the analog of Theorem 1:

**Proof of Theorem 2:** As $\sqrt{N}(\hat{J}^{-1}\hat{G}_t - G_t)$ is asymptotically normally distributed under Assumption A-G, the weighted sum can be written as:
\[
\sum_{s=1}^{T} \hat{\Phi}_{t,s}\sqrt{N}(\hat{J}^{-1}\hat{G}_s - G_s) = \sqrt{N}(\hat{J}^{-1}\hat{F}_t - F_t) + \sqrt{N}\sum_{s=1}^{T} \left(\hat{\Phi}_{t,s} - \Phi_{t,s}\right) (\hat{J}^{-1}\hat{G}_s - G_s).
\]

Since $\hat{\Phi} \xrightarrow{p} \Phi$, the second term vanishes asymptotically and hence, the feasible estimator of $F_t$ follows the same asymptotic distribution as the unfeasible estimator of $F_t$ under Assumptions A-G. Similarly, as $\sqrt{T}(\hat{J}\hat{\lambda}_i - \lambda_i)$ is asymptotically normally distributed under Assumption A-G, the weighted sum can be written as:
\[
\sum_{j=1}^{N} \tilde{\Theta}_{j,i}\sqrt{T}(\hat{J}\hat{\gamma}_j - \gamma_j) = \sqrt{T}(\hat{J}\hat{\lambda}_i - \lambda_i) + \sqrt{T}\sum_{j=1}^{N} \left(\tilde{\Theta}_{j,i} - \Theta_{j,i}\right) (\hat{J}\hat{\gamma}_j - \gamma_j).
\]

Since $\tilde{\Theta} \xrightarrow{p} \Theta$, the second term vanishes asymptotically and hence, the feasible estimator of $\lambda_i$ follows the same asymptotic distribution as the unfeasible estimator of $\lambda_i$ under Assumptions A-G. Last, as $\delta_{NT}\left(\hat{G}'_t\hat{\gamma}_i - G'_i\gamma_i\right)$ is asymptotically normally distributed under Assumption A-E by Lemma 1, the weighted sum can be written as:
\[
\sum_{s=1}^{T} \sum_{j=1}^{N} \hat{\Phi}_{t,s}\tilde{\Theta}_{j,i}\delta_{NT}\left(\hat{G}'_s\hat{\gamma}_j - G'_s\gamma_j\right)
\]
\[
= \delta_{NT}\left(\hat{C}_{i,t} - C_{i,t}\right) + \delta_{NT}\sum_{s=1}^{T} \sum_{j=1}^{N} \left(\hat{\Phi}_{t,s}\tilde{\Theta}_{j,i} - \Phi_{t,s}\Theta_{j,i}\right) \left(\hat{G}'_s\hat{\gamma}_j - G'_s\gamma_j\right).
\]

Since $\hat{\Phi} \xrightarrow{p} \Phi$ and $\tilde{\Theta} \xrightarrow{p} \Theta$, the second term vanishes asymptotically and hence, the feasible estimator of $C_{i,t}$ follows the same asymptotic distribution as the unfeasible
estimator of $C_{i,t}$ under Assumptions A-E and G.

\[ \square \]

A.3 Proof of Theorem 3

The subsequent proof of follows the argument of Breitung and Tenhofen (2011) closely. First, we compare the asymptotic covariance matrices of the GLS estimator and the PCA estimator of $F_t$. Consider

\[
\text{var} \left( H^{-1} \tilde{F}_t \right) = \text{var} \left( J^{-1} \hat{F}_t \right) + \text{var} \left( H^{-1} \tilde{F}_t - J^{-1} \hat{F}_t \right) + \text{cov} \left( J^{-1} \hat{F}_t, H^{-1} \tilde{F}_t - J^{-1} \hat{F}_t \right) + \text{cov} \left( H^{-1} \tilde{F}_t - J^{-1} \hat{F}_t, J^{-1} \hat{F}_t \right).
\]

It follows that $J^{-1} \hat{F}_t$ is asymptotically more efficient than $H^{-1} \tilde{F}_t$ if

\[
N \text{cov} \left( J^{-1} \hat{F}_t, H^{-1} \tilde{F}_t - J^{-1} \hat{F}_t \right) \rightarrow 0.
\]

Since

\[
\lim_{N,T \rightarrow \infty} N E \left[ \left( J^{-1} \hat{F}_t - F_t \right) \left( H^{-1} \tilde{F}_t - F_t \right) \right] = \lim_{N,T \rightarrow \infty} N \left( \Lambda' (\Theta' \Theta)^{-1} \Lambda \right)^{-1} \Lambda' (\Theta' \Theta)^{-1} E(e_t e'_t) \Lambda \left( \Lambda' \Lambda \right)^{-1}
\]

\[
= \lim_{N,T \rightarrow \infty} N \left( \Phi_t' \Phi_t \right) \left( \Lambda' (\Theta' \Theta)^{-1} \Lambda \right)^{-1}
\]

\[
= \lim_{N,T \rightarrow \infty} N E \left[ \left( J^{-1} \hat{F}_t - F_t \right) \left( J^{-1} \hat{F}_t - F_t \right) \right]
\]

it follows that the difference of the asymptotic covariance matrices of the GLS- and the PCA estimator of $F_t$ is positive semidefinite. Similarly, we compare the asymptotic covariance matrices of the GLS estimator and the PCA estimator of $\lambda_i$. Consider

\[
\lim_{N,T \rightarrow \infty} T E \left[ \left( J \hat{\lambda}_i - \lambda_i \right) \left( H \tilde{\lambda}_i - \lambda_i \right) \right] = \lim_{N,T \rightarrow \infty} T \left( \Theta'_i \Theta_i \right) \left( F' (\Phi \Phi')^{-1} F \right)^{-1} F' (\Phi \Phi')^{-1} \left( \Phi \Phi' \right) F \left( F' F \right)^{-1}
\]

\[
= \lim_{N,T \rightarrow \infty} T \left( \Theta'_i \Theta_i \right) \left( F' (\Phi \Phi')^{-1} F \right)^{-1}
\]
Hence, the GLS estimator of $\lambda_i$ is asymptotically more efficient than the corresponding PCA estimator.

\[ \Box \]

### A.4 Assumptions and Proof of Theorem 4

We define $u_t = G_t - G_{t-1}$, that is a $r \times 1$ vector of zero-mean I(0) processes (not necessarily i.i.d.). We proceed to enlist the modified assumptions:

**Assumption B’**: Independence

\[ \{u_t\}_{t=1}^T, \{\varepsilon_{i,t}\}_{i,t=1}^{N,T} \text{ and } \{\lambda_i\}_{i=1}^N \text{ are mutually independent.} \]

**Assumption C’**: Rates

\[ N, T \to \infty \text{ jointly such that } \frac{N}{T^3} \to 0. \]

**Assumption D’**: Common Stochastic Trends

1. \( E\|u_t\|^{4+\delta} \leq M \) for some \( \delta > 0 \) and for all \( t \leq T \)

2. As \( T \to \infty \), \( \frac{1}{T^7} F'(\Phi'\Phi)^{-1}F \overset{d}{\to} \int B_u B_u' \), where \( B_u \) is a vector of Brownian motions with covariance matrix \( \lim_{T \to \infty} \left( \frac{1}{T} \sum_{s=1}^T \sum_{t=1}^T E(u_t u'_s) \right) \) is positive definite.

3. \( \lim \inf_{T \to \infty} \left( \log(\log(T)) \frac{1}{T^7} F'(\Phi'\Phi)^{-1}F \right) = D \), where \( D \) is a non-random positive definite matrix.

4. \( E \left\| \sum_{s=1}^T [\Phi^{-1}]_{0,s} F_s \right\|^4 \leq M < \infty \)

**Assumption E’**: Limiting Distribution

1. for each \( t \), as \( N \to \infty \), we assume \( \frac{1}{\sqrt{N}} \Lambda'_t \Theta^{-1} \varepsilon'_t \Phi_t \overset{d}{\to} N(0, \Omega_t) \)

2. for each \( i \), as \( T \to \infty \), we assume \( \frac{1}{T} F'(\Phi^{-1})' \varepsilon \Theta_i \overset{d}{\to} \int B_u dB^{(i)}_\varepsilon \)

where \( \Omega_t \) is defined in Assumption E, \( B_u \) is the vector of Brownian motions defined in Assumption D’ (ii), \( B^{(i)}_\varepsilon \) is a scalar Brownian motion process and \( B_u \) and \( B^{(i)}_\varepsilon \) are independent.
Assumption F': distinct eigenvalues

The eigenvalues of the \( r \times r \) random matrix \( \Sigma_{\Lambda^*}^{-1/2} \int B_u B_u' \Sigma_{\Lambda^*}^{1/2} \) are distinct almost surely.

Lemma 4 states the asymptotic distribution of the factor and factor loadings of the transformed model defined in (8) (formulated in the \( (G, \Gamma) \)-space) when \( G_t \) is of order \( I(1) \).

**Lemma 4: Asymptotic Distribution of \( \hat{\mathbf{G}}_t \) and \( \hat{\gamma}_i \)**

Suppose that Assumptions A, B'-F' hold. Then,

(i) \( \sqrt{N} \left( J^{-1} \hat{\mathbf{G}}_t - \mathbf{G}_t \right) = \left( \frac{\Gamma' \Gamma}{N} \right)^{-1} \frac{\Gamma' \varepsilon_t}{\sqrt{N}} + o_p(1) \xrightarrow{d} N \left( 0, W_{\hat{\mathbf{G}}_t} \right) \)

(ii) \( T \left( J \hat{\gamma}_i - \gamma_i \right) = \left( \frac{G'G}{T^2} \right)^{-1} \frac{G' \varepsilon_i}{T} + o_p(1) \)

where the limiting distribution of (ii) is conditionally normal.

**Proof of Lemma 4:** Assumptions A, B'-F' imply the application of Corollary 1 and 2 of Bai (2003) where \( (G, \Gamma, \varepsilon) \) is Bai's \( (F, \Lambda, e) \).

Using Lemma 4, we can establish the asymptotic distribution of the unfeasible estimators of \( F_t \) and \( \lambda_i \) as stated in Theorem 1 noting that \( F_t = \sum_{s=1}^{T} \Phi_{t,s} G_s \) and \( \lambda_i = \sum_{j=1}^{N} \Theta_{j,i} \gamma_i \).

**Proof of Theorem 4:** As \( \sqrt{N}(J^{-1} \hat{\mathbf{G}}_t - \mathbf{G}_t) \) is asymptotically normally distributed under Assumption A, B'-F' by Lemma 4 (i), the weighted sum

\[
\sum_{s=1}^{T} \Phi_{t,s} \sqrt{N}(J^{-1} \hat{\mathbf{G}}_s - \mathbf{G}_s) = \sqrt{N}(J^{-1} \hat{\mathbf{F}}_t - \mathbf{F}_t)
\]

is also asymptotically normally distributed. The asymptotic variance is:

\[
\lim_{N,T \to \infty} \left( \frac{\Gamma' \Gamma}{N} \right)^{-1} \left[ \frac{\Gamma' \varepsilon' \Phi_t \varepsilon \Gamma}{N} \right] \left( \frac{\Gamma' \Gamma}{N} \right)^{-1} = \Sigma_{\Lambda^*}^{-1} \Omega_t \Sigma_{\Lambda^*}^{-1}
\]

where \( \Sigma_{\Lambda^*} \) is the distributional limit of \( \frac{1}{N} \Lambda' \left( \Theta' \Theta \right)^{-1} \Lambda \) and \( \Omega_t \) is defined in Assumption E. Furthermore, since the limiting distribution of \( T(J \hat{\gamma}_i - \gamma_i) \) is conditional
normal under Assumption A,B'-F' by Lemma 4 (ii), the weighted sum
\[ \sum_{j=1}^{N} \Theta_{j,i} T(J\hat{\gamma}_j - \gamma_j) = T(J\hat{\lambda}_i - \lambda_i) \]
follows also a conditional normal distribution in the limit. We have:
\[ T(J\hat{\lambda}_i - \lambda_i) = \left( \frac{G'G}{T^2} \right)^{-1} \frac{G'\epsilon\Theta_i}{T} + o_p(1) \xrightarrow{d} \Sigma_F \left( \int B_u dB_u^{(i)} \right) \]
where \( \Sigma_F \) is the distributional limit of \( \frac{1}{T^2} F' (\Phi\Phi')^{-1} F \).

\[ \square \]

References


Bai, J. and S. Ng (2002). "Determining the Number of Factors in Approximate Factor Models". Econometrica 70(1), 191–221.


Chamberlain, G. and M. Rothschild (1982, October). "Arbitrage, Factor Structure,


Girosi, F. and G. King (2007). ”Understanding the Lee-Carter Mortality Forecasting Method”.


Mohangoo, A., C. Hukkelhoven, P. Achterberg, P. Elferink-Stinkens, A. Ravelli, G. Rijninks-van Driel, P. Tamminga, A. Waelput, K. van der Pal-de Bruin, and


### Tables

<table>
<thead>
<tr>
<th></th>
<th>Stationary</th>
<th></th>
<th>Non-stationary</th>
<th></th>
<th>Stationary &amp; Non-stationary</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r = 1</td>
<td></td>
<td>r = 1</td>
<td></td>
<td>r = 2</td>
<td></td>
</tr>
<tr>
<td>N T</td>
<td>AR(1) MA(1) MS</td>
<td>AR(1) MA(1) MS</td>
<td>AR(1) MA(1) MS</td>
<td>AR(1) MA(1) MS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 50</td>
<td>2.43 3.24 1.62</td>
<td>6.03 4.61 14.49</td>
<td>11.38 8.80 25.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 100</td>
<td>1.87 2.48 1.11</td>
<td>3.84 3.13 13.62</td>
<td>7.55 6.39 22.72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 200</td>
<td>1.49 1.98 0.83</td>
<td>2.58 2.34 13.1</td>
<td>5.04 4.67 21.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 50</td>
<td>1.78 2.38 1.14</td>
<td>5.45 3.78 14.98</td>
<td>10.52 7.53 25.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 100</td>
<td>1.20 1.61 0.64</td>
<td>3.24 2.36 12.56</td>
<td>6.34 4.64 22.68</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 200</td>
<td>0.93 1.24 0.50</td>
<td>2.05 1.62 13.49</td>
<td>3.90 3.13 22.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200 50</td>
<td>1.49 1.99 0.80</td>
<td>5.23 3.43 14.13</td>
<td>9.71 6.53 25.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200 100</td>
<td>0.89 1.18 0.49</td>
<td>2.97 1.94 13.86</td>
<td>5.58 3.76 22.41</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200 200</td>
<td>0.60 0.80 0.32</td>
<td>1.77 1.21 13.06</td>
<td>3.34 2.36 20.99</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel A: GLS

|                  |  |  |  |  |
|------------------| | | | |
| 50 50            | 3.60 3.85 10.24 | 7.04 5.06 16.52 | 13.09 9.57 27.27 |
| 50 100           | 2.86 3.00 7.37 | 4.63 3.53 15.02 | 9.24 7.21 24.11 |
| 50 200           | 2.46 2.48 5.64 | 3.43 2.80 14.08 | 6.52 5.41 22.95 |
| 100 50           | 2.38 2.71 6.22 | 6.03 3.99 16.16 | 11.59 7.98 25.88 |
| 100 100          | 1.68 1.86 3.71 | 3.65 2.53 13.11 | 7.24 5.06 23.2 |
| 100 200          | 1.38 1.48 3.16 | 2.48 1.84 13.93 | 4.62 3.48 22.35 |
| 200 50           | 1.81 2.23 3.97 | 5.58 3.55 15.14 | 10.32 6.79 25.71 |
| 200 100          | 1.14 1.31 2.72 | 3.22 2.05 14.23 | 6.01 3.96 22.58 |
| 200 200          | 0.81 0.90 1.33 | 1.96 1.31 13.24 | 3.69 2.52 21.15 |

Panel B: B&T

|                  |  |  |  |  |
|------------------| | | | |
| 50 50            | 4.36 4.07 13.05 | 7.36 5.10 21.44 | 13.86 9.64 32.97 |
| 50 100           | 3.29 3.18 9.52 | 4.75 3.55 18.52 | 9.65 7.25 28.52 |
| 50 200           | 2.68 2.57 6.94 | 3.48 2.80 16.26 | 6.64 5.42 25.94 |
| 100 50           | 3.20 3.03 9.08 | 6.38 4.02 21.22 | 12.45 8.07 31.69 |
| 100 100          | 2.09 2.02 5.29 | 3.82 2.55 16.46 | 7.68 5.11 27.49 |
| 100 200          | 1.62 1.58 4.03 | 2.52 1.84 16.35 | 4.75 3.49 25.37 |
| 200 50           | 2.65 2.54 6.16 | 5.94 3.59 20.10 | 11.25 6.86 31.58 |
| 200 100          | 1.51 1.45 3.98 | 3.32 2.05 17.76 | 6.36 3.99 26.94 |
| 200 200          | 1.00 0.98 1.99 | 2.01 1.31 15.54 | 3.80 2.53 23.99 |

Panel C: PCA

Table 1: Efficiency comparison for the estimation of common components under Gaussian specification. The table reports empirical MSEs scaled by 100.
<table>
<thead>
<tr>
<th>parameter</th>
<th>dependent variable</th>
<th>$\hat{F}_t$</th>
<th>$\tilde{F}_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>$-0.031$</td>
<td>$(0.006)$</td>
<td>$-0.032$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>$-0.451$</td>
<td>$(0.061)$</td>
<td>$-0.261$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>$0.074$</td>
<td>$(0.004)$</td>
<td>$0.059$</td>
</tr>
</tbody>
</table>

Table 2: ARIMA estimates based on $\hat{F}_t$ and $\tilde{F}_t$. Standard errors in brackets.

<table>
<thead>
<tr>
<th>Time</th>
<th>Age</th>
<th>2010</th>
<th>2020</th>
<th>2030</th>
<th>2040</th>
<th>2050</th>
<th>2060</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: GLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>83.86</td>
<td>84.52</td>
<td>85.14</td>
<td>85.72</td>
<td>86.26</td>
<td>86.77</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>73.22</td>
<td>73.92</td>
<td>74.57</td>
<td>75.18</td>
<td>75.75</td>
<td>76.29</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>62.49</td>
<td>63.29</td>
<td>63.98</td>
<td>64.62</td>
<td>65.22</td>
<td>65.78</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>51.82</td>
<td>52.63</td>
<td>53.40</td>
<td>54.06</td>
<td>54.69</td>
<td>55.28</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>41.07</td>
<td>42.03</td>
<td>42.79</td>
<td>43.52</td>
<td>44.17</td>
<td>44.78</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>30.50</td>
<td>31.46</td>
<td>32.33</td>
<td>33.04</td>
<td>33.73</td>
<td>34.34</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>20.51</td>
<td>21.31</td>
<td>22.12</td>
<td>22.88</td>
<td>23.52</td>
<td>24.14</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>11.55</td>
<td>12.19</td>
<td>12.82</td>
<td>13.45</td>
<td>14.05</td>
<td>14.59</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>4.63</td>
<td>5.07</td>
<td>5.45</td>
<td>5.82</td>
<td>6.21</td>
<td>6.59</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>1.28</td>
<td>1.38</td>
<td>1.46</td>
<td>1.55</td>
<td>1.63</td>
<td>1.73</td>
<td></td>
</tr>
<tr>
<td>Panel B: PCA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>85.06</td>
<td>85.83</td>
<td>86.51</td>
<td>87.13</td>
<td>87.68</td>
<td>88.18</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>74.26</td>
<td>75.11</td>
<td>75.87</td>
<td>76.55</td>
<td>77.15</td>
<td>77.70</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>63.36</td>
<td>64.34</td>
<td>65.17</td>
<td>65.92</td>
<td>66.58</td>
<td>67.18</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>52.49</td>
<td>53.50</td>
<td>54.43</td>
<td>55.25</td>
<td>55.98</td>
<td>56.63</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>41.55</td>
<td>42.70</td>
<td>43.66</td>
<td>44.56</td>
<td>45.34</td>
<td>46.05</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>30.79</td>
<td>31.95</td>
<td>33.00</td>
<td>33.90</td>
<td>34.74</td>
<td>35.49</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>20.60</td>
<td>21.60</td>
<td>22.61</td>
<td>23.54</td>
<td>24.35</td>
<td>25.12</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>11.39</td>
<td>12.30</td>
<td>13.10</td>
<td>13.91</td>
<td>14.66</td>
<td>15.34</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>4.50</td>
<td>5.03</td>
<td>5.56</td>
<td>6.05</td>
<td>6.56</td>
<td>7.04</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>1.26</td>
<td>1.36</td>
<td>1.47</td>
<td>1.58</td>
<td>1.70</td>
<td>1.82</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Expected remaining lifetime forecasts based on GLS and PCA for different years and different ages.
<table>
<thead>
<tr>
<th></th>
<th>PCA</th>
<th>GLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>86.67</td>
<td>86.64</td>
</tr>
<tr>
<td>Male</td>
<td>80.54</td>
<td>80.79</td>
</tr>
</tbody>
</table>

Δ  6.13  5.85

Table 4: Life expectancy forecasts of an unborn in the year 2015.
Figures

(a) Relative efficiency of the estimators of the factor loading $\lambda_i$

(b) Relative efficiency of the estimators of the factor $F_t$

(c) Relative efficiency of the estimators of the common component $C_{i,t}$

Figure 1: Relative efficiency, defined by the ratio of the asymptotic variance of the GLS estimator and the asymptotic variance of the PCA estimator. AR(1) specification for $\Theta'\Theta$ and $\Phi\Phi'$ with parameters $\theta$ and $\phi$ respectively.
Figure 2: Three alternative covariance structures used for simulation.

(a) MA(1)

(b) AR(1)

(c) MS

Figure 3: Common component density under factor stationarity, normality and AR(1) specification with $\phi = 0.5$ and $\theta = 0.5$. Sample size: $N = 50$, $T = 50$. The density is non-parametrically estimated based on $S = 10,000$ observations.
Figure 4: Lee-Carter data and residuals for the Netherlands

(a) Central death rate

(b) Lee-Carter residuals

Figure 5: Sample partial autocorrelation function of the heteroskedasticity-adjusted residual series of age group $i = 0$. 
(a) Estimated age-group-specific factor loading $\lambda_i$ aka $b_i$

(b) Estimated time-varying mortality factor $F_t$ aka $k_t$

Figure 6: The PCA and GLS method applied to the Dutch central death rates.

Figure 7: Timeline of the World War II events with the corresponding mortality factor estimates.
Figure 8: GLS and PCA based life expectancy point forecasts of an unborn over time with 5% and 95% GLS quantiles.

Figure 9: The PCA and GLS method applied to the gender-specific Dutch central death rates.
Figure 10: GLS and PCA based life expectancy point forecasts of an unborn in year 2015 for Belgium, the Netherlands, Luxembourg and West Germany.
**MASTER'S THESIS CEMFI**

0801 *Paula Inés Papp*: “Bank lending in developing countries: The effects of foreign banks”.

0802 *Liliana Bara*: “Money demand and adoption of financial technologies: An analysis with household data”.

0803 *J. David Fernández Fernández*: “Elección de cartera de los hogares españoles: El papel de la vivienda y los costes de participación”.

0804 *Máximo Ferrando Ortí*: “Expropriation risk and corporate debt pricing: the case of leveraged buyouts”.

0805 *Roberto Ramos*: “Do IMF Programmes stabilize the Economy?”.  

0806 *Francisco Javier Montenegro*: “Distorsiones de Basilea II en un contexto multifactorial”.

0807 *Clara Ruiz Prada*: “Do we really want to know? Private incentives and the social value of information”.

0808 *Jose Antonio Espin*: “The “bird in the hand” is not a fallacy: A model of dividends based on hidden savings”.

0809 *Víctor Capdevila Cascante*: “On the relationship between risk and expected return in the Spanish stock market”.

0810 *Lola Morales*: “Mean-variance efficiency tests with conditioning information: A comparison”.

0811 *Cristina Soria Ruiz-Ogarrio*: “La elasticidad micro y macro de la oferta laboral familiar: Evidencia para España”.

0812 *Carla Zambrano Barbery*: “Determinants for out-migration of foreign-born in Spain”.

0813 *Álvaro de Santos Moreno*: “Stock lending, short selling and market returns: The Spanish market”.

0814 *Olivia Peraita*: “Assessing the impact of macroeconomic cycles on losses of CDO tranches”.

0815 *Iván A. Kataryniuk Di Costanzo*: “A behavioral explanation for the IPO puzzles”.

1001 *Oriol Carreras*: “Banks in a dynamic general equilibrium model”.

1002 *Santiago Pereda-Fernández*: “Quantile regression discontinuity: Estimating the effect of class size on scholastic achievement”.

1003 *Ruxandra Ciupagea*: “Competition and “blinders”: A duopoly model of information provision”.

1004 *Rebeca Anguren*: “Credit cycles: Evidence based on a non-linear model for developed countries”.

1005 *Alba Diz*: “The dynamics of body fat and wages”.

1101 *Daniela Scidá*: “The dynamics of trust: Adjustment in individual trust levels to changes in social environment”.

1102 *Catalina Campillo*: “Female labor force participation and mortgage debt”.

1103 *Florina Raluca Silaghi*: “Immigration and peer effects: Evidence from primary education in Spain”.

1104 *Jan-Christoph Bietenbeck*: “Teaching practices and student achievement: Evidence from TIMSS”.
Andrés Gago: “Reciprocity: Is it outcomes or intentions? A laboratory experiment”.

Rocío Madera Holgado: “Dual labor markets and productivity”.

Lucas Gortazar: “Broadcasting rights in football leagues and TV competition”.

Mª Elena Álvarez Corral: “Gender differences in labor market performance: Evidence from Spanish notaries”.

José Alonso Olmedo: “A political economy approach to banking regulation”.

Joaquín García-Cabo Herrero: “Unemployment and productivity over the business cycle: Evidence from OECD countries”.

Luis Díez Catalán: “Collective bargaining and unemployment during the Great Recession: Evidence from Spain”.

Cecilia Dassatti Camors: “Macroprudential and monetary policy: Loan-level evidence from reserve requirements”.

Juan Carvajal: “Choosing to invest in human capital through adult education”.

Haritz Garro: “Desire to win and public information in majority rule elections”.

Álvaro Martín Herrero: “Credit and liquidity risk in sovereign bonds”.

José Carreño: “Housing bubbles, doubts and learning”.

Ester Núñez de Miguel: “Excuse me, do you speak English? An international evaluation”.

Alexander Heinemann: “Efficient estimation of factor models with time and cross-sectional dependence”.