APPROXIMATING FACTOR MODELS ON MULTI-POPULATION MORTALITY MODELING

Richard MacMinn
Ko-Lun Kung
Weiyu Kuo
Jason Chenghsien Tsai

Outline

- Motivations
- Literatures
- Approximating factor models
- Results

Motivations for Multi-Population Modeling

- Global convergence of mortality trends
  - Individual mortality forecasts were likely to be divergent in the long-run
- Multi-population longevity product
  - Vita series
- Basis risk
  - Using the mortality-linked products that have different underlying populations
- Managing multi-national insurance company
  - Indeed, almost all insurers cover multi-populations.

Motivations for Approximating Factor Modeling

- Although the factor model has proven its success in mortality modeling such as Lee and Carter (1992), some challenges occur to multi-population modeling
  - The number of ages is often greater than the number of observations, which violates the assumption of factor model analysis
  - The i. i. d. and diagonality assumptions are almost always violated.
- We introduce approximating factor models to the modeling on the mortality rates of multiple populations.

Literatures

- Joint-k Model
- Cointegrated model
- One-common-factor model
  - Li and Lee (2005)
- Augmented common factor model
  - Li and Hardy (2011)
- Model for mortality indices
  - Lin et al. (2013)
  - Chen et al. (2014)

Multi-Population Mortality Rate Curve Modeling

- Augmented Common Model (ACM) (i: country, x: age, and t: time)

$$m_i(x,t) = a_i(x) + B(x)K(t) + b_i(x)k_i(t) + \epsilon_i(x,t),$$

- The common factor is extracted from weighted average of population mortality matrices

$$M_{N\times T} = \sum_{i=1}^{I} w_i \left[\; M_{N\times T}^{(i)} \;\right]$$

The First, Easy Generalization

- We treat each mortality rate as a risk factor. We stack mortality rate matrices of all *i* countries

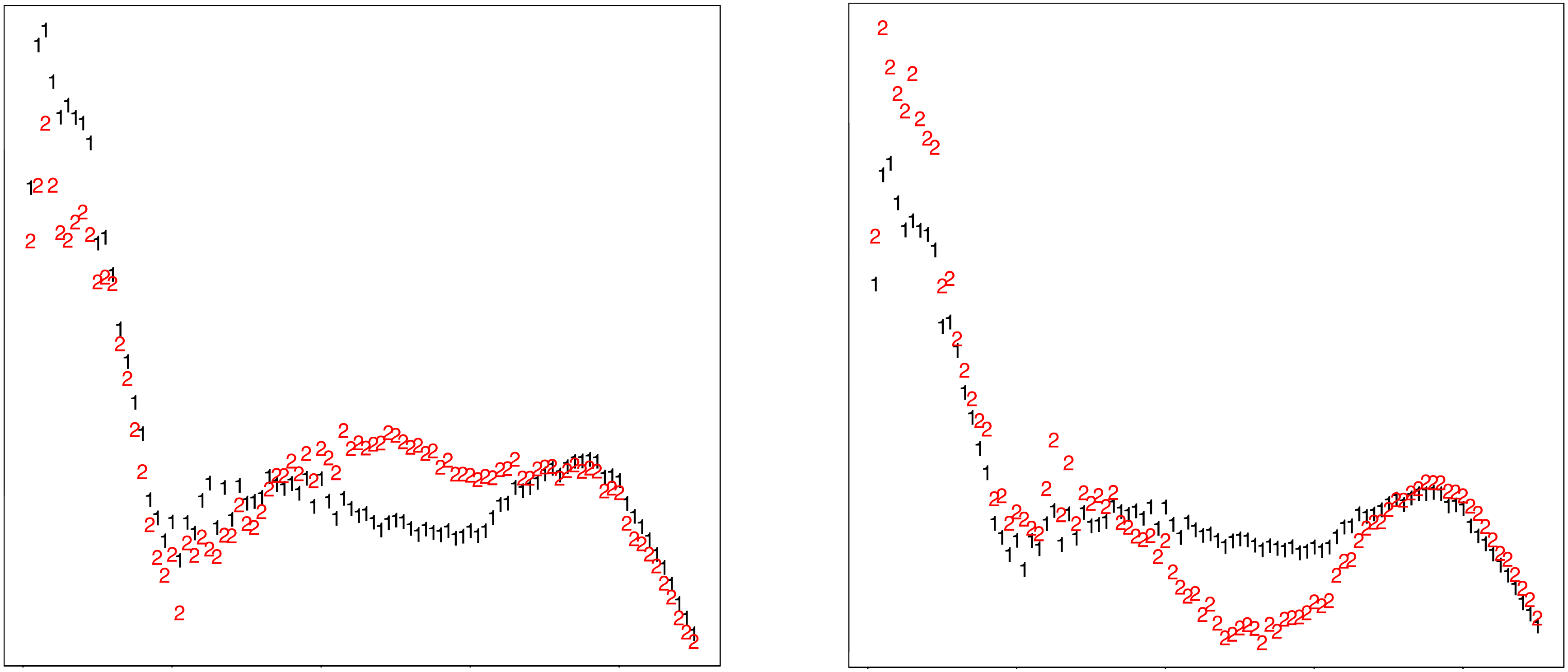
$$M_{(N_1+N_2+\cdots+N_I)\times T} = \begin{bmatrix} M^{(1)}_{N_1\times T} \\ M^{(2)}_{N_2\times T} \\ \vdots \\ M^{(I)}_{N_I\times T} \end{bmatrix}$$

- Then estimate the factor model with PCA. Now all population can have its own age sensitivity to factor shocks

$$m_i(x,t) = a_i(x) + B_i(x)K(t) + b_i(x)k_i(t) + \epsilon_i(x,t).$$

An illustration on age sensitivity estimates

- Averaging data resulted in under/over-estimation in age 0 – 10 and 40 - 60 range between population



Covariances of mortality rates

- Decomposing the covariance matrix

$$\mathrm{cov}(M) = B\mathrm{var}(K)B^\top + b\mathrm{var}(k)b^\top + \mathrm{cov}(\epsilon\epsilon^\top)$$

- Among multiple populations,
  - ACM implies mortality rate from any distinct populations is always perfectly correlated since (a) independent factors do not contribute to covariance between populations and (b) loading are the same across populations
  - Our framework relaxes such assumption, i.e., mortality rates can be positively or negatively correlated

Conventional PCA

- The model

$$M = \bar{M} + BF + \epsilon$$

- Spectral decomposition to sample covariance to estimate the loading B

$$\hat{\Sigma} = \hat{\text{cov}}(M) = BB^\top, \quad B_{\cdot j} = \sqrt{\lambda_j} e_j$$

where (λ_j, e_j) is the jth eigenvalue and eigenvector

- Then estimate factor F via regression

$$\hat{F} = (B^\top B)^{-1} B^\top (M - \bar{M})$$

Asymptotic PCA

- Conner and Korajczyk (1986) and Tsay (2013)
- To extract factors,
  - instead of doing spectral decomposition to covariance of original mortality matrix, which is dimension N x N
  - we do spectral decomposition to the covariance of transposed mortality matrix, which is dimension T x T
- Intuition:
  - We know the model has at most min(N, T) principal components. The extra N – T eigenvalues are hence equal to 0. To find at most T eigenvalues we therefore just need a T x T covariance matrix.

Asymptotic PCA

- Advantages:
  - The percentage of variation explained is greater than or equal to traditional PCA (such as Lee-Carter)
  - Incorporate cross-sectional heteroskadasticity

Asymptotic PCA (APCA)

- ☐ GLS Estimation

1. Compute the cross-product $\hat{\Omega}_T$ where X_n is the age-*n* log mortality rates (1 x T vector)

$$\widehat{\Omega_T} = \frac{1}{N-1} \sum_{n=1}^{N} (X_n - \bar{X}_n)(X_n - \bar{X}_n)^\top.$$

2. Obtain first *m* eigenvectors
3. For each age *n*, regress the eigenvectors to (demeaned) data and retrieve residual variance $\hat{\sigma}_n^2$
4. Scale $X_{T\times N}$ as $X\hat{D}^{-1/2}$, where $\hat{D} = \mathrm{diag}(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_N^2)$
5. Compute the cross-product of $X\hat{D}^{-1/2}$
6. Obtain first *m* eigenvectors from cross-product of $X\hat{D}^{-1/2}$ as factor estimates

Heteroskedasticity Factor Analysis (HFA)

- Jones (2001)
  - Use maximum likelihood to run the GLS estimation and find the residual variances in dimension T
  - Incorporate time series heteroskedasticity

1. Compute $\hat{\Omega}_T$
2. Set $\hat{D}_0$ as the initial guess of diagonal residual covariance D
3. Find first *m* eigenvectors for $\hat{D}_0^{-1/2}\hat{\Omega}_T\hat{D}_0^{-1/2}$
4. Let V and Λ be the matrix of first *m* eigenvectors and eigenvalues. Compute factor estimate $\hat{F} = \hat{D}_0^{-1/2}V(\Lambda - I)^{-1/2}$
5. Update estimate of D by $\hat{D} = \hat{\Omega}_T - \hat{F}^\top \hat{F}$
6. Iterate step 3 – 5 till convergence criterion is achieved

POET
Principal Orthogonal ComplEment Thresholding

- Fan, Liao and Mincheva (2011)
  - Estimate the off-diagonal correlation to capture higher-order variation
  - Estimate the factors and loadings with standard PCA
- The estimated residual covariance matrix

$$\hat{\Sigma}_\epsilon = \mathrm{cov}\widehat{(}\epsilon\epsilon^\top) = (\hat{r}_{ij})_{N \times N}$$

only takes value when entry exceeds certain threshold to avoid spurious estimate:

$$\hat{\Sigma}_\epsilon^\mathcal{T} = (\hat{r}_{ij}^\mathcal{T})_{N \times N}, \quad \hat{r}_{ij}^\mathcal{T} = \hat{r}_{ij} I(|\hat{r}_{ij}| \geq \tau_{ij})$$

- The thresholding can be "hard" (i.e. all or nothing) or "soft" (i.e. tuned down estimate)

Model Comparisons

- Data: UK, US, Spain, Italy and France 45 – 100 male and female populations, 1933 – 2007
- We compared the log likelihood and BIC of 9 models: ACM, PCA, APCA, HFA, POET; Independent LC, CBD, Plat (2009), and Colin and O'Hare (2012) (M10)
  - Plat (2009)

$$\log m_i(x,t) = a_i(x) + k_i^{(1)}(t) + k_i^{(2)}(t)(\bar{x}-x) + k_i^{(3)}(t)(\bar{x}-x)^+ + \epsilon_i(x,t)$$

  - M10

$$\begin{aligned}\log m_i(x,t) = a_i(x) + k_i^{(1)}(t) + k_i^{(2)}(t)(\bar{x}-x) \\ + k_i^{(3)}(t)\Big((\bar{x}-x)^+ + \big((\bar{x}-x)^+\big)^2\Big) + \epsilon_i(x,t)\end{aligned}$$

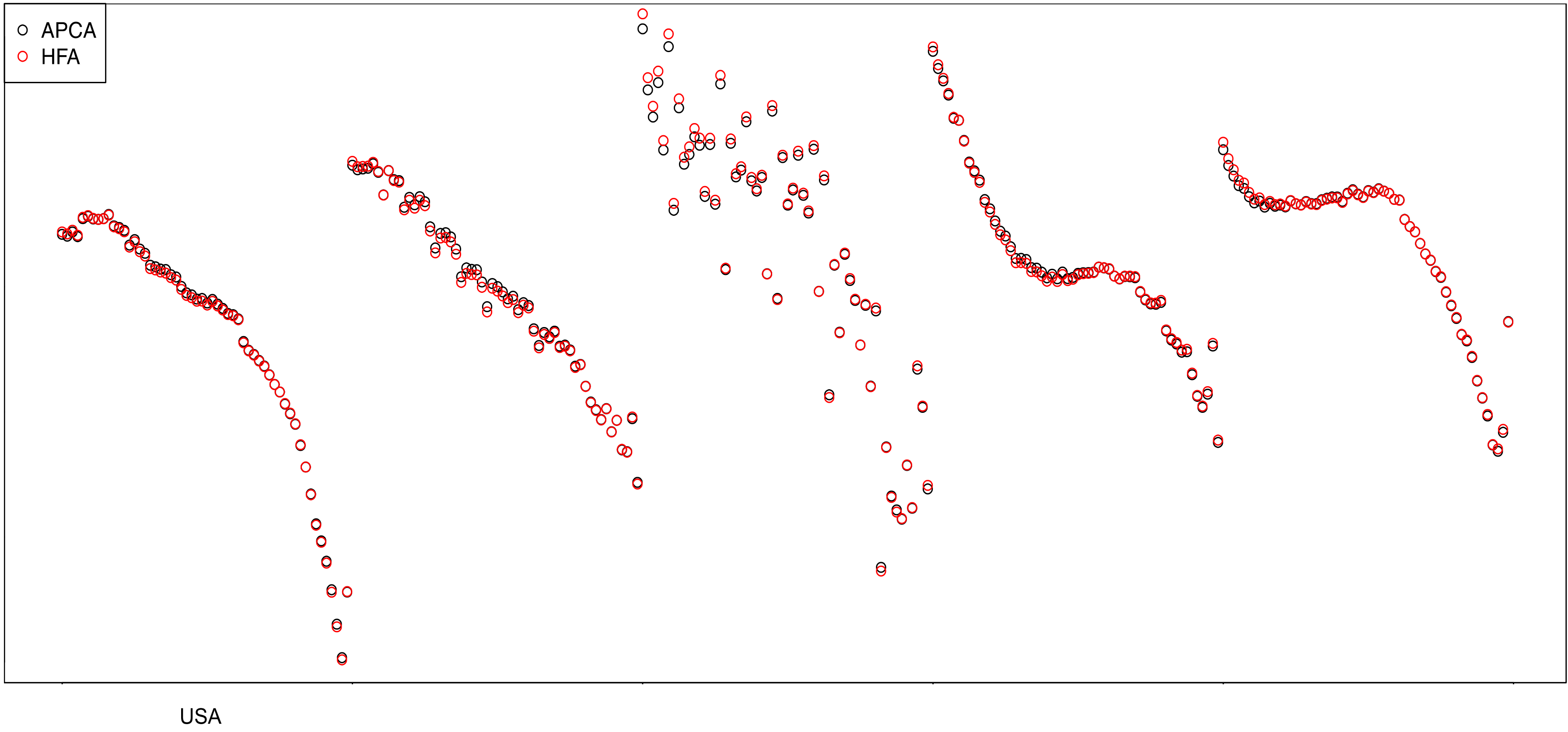
Time-Series Modeling

- ☐ We follow the literatures in choosing time-series models for common factors.

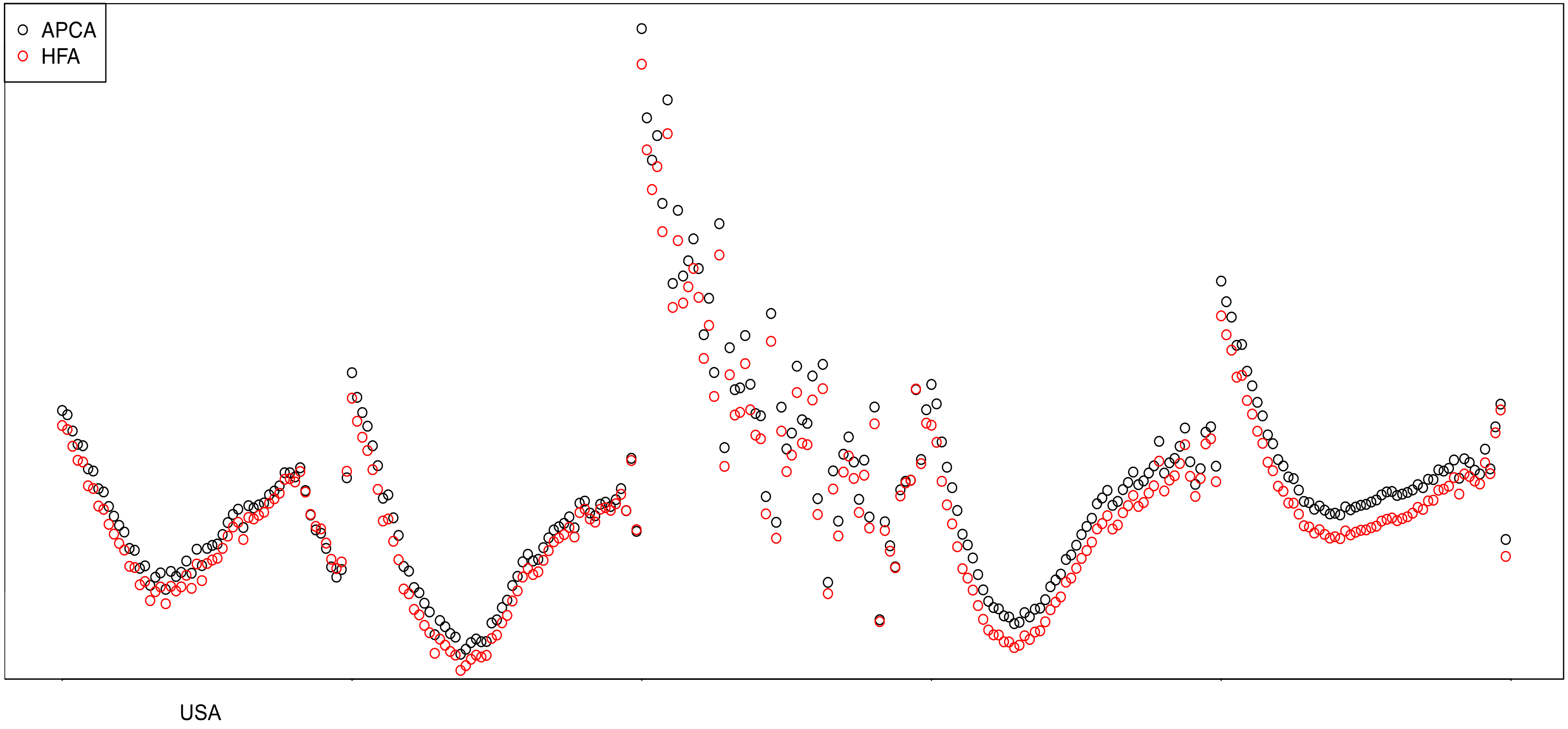
Model	**ACM**	**PCA, APCA, HFA and POET**	**Independent LC, CBD, Plat and M10**
Time series model for first trend factor	ARIMA(0,1,0)	ARIMA(0,1,0)	ARIMA(0,1,0)
Time series model for second factor and beyond*	auto.arima by BIC	auto.arima by BIC	AR(1)

* If factor is an I(1) process then we choose ARIMA(0,1,0)

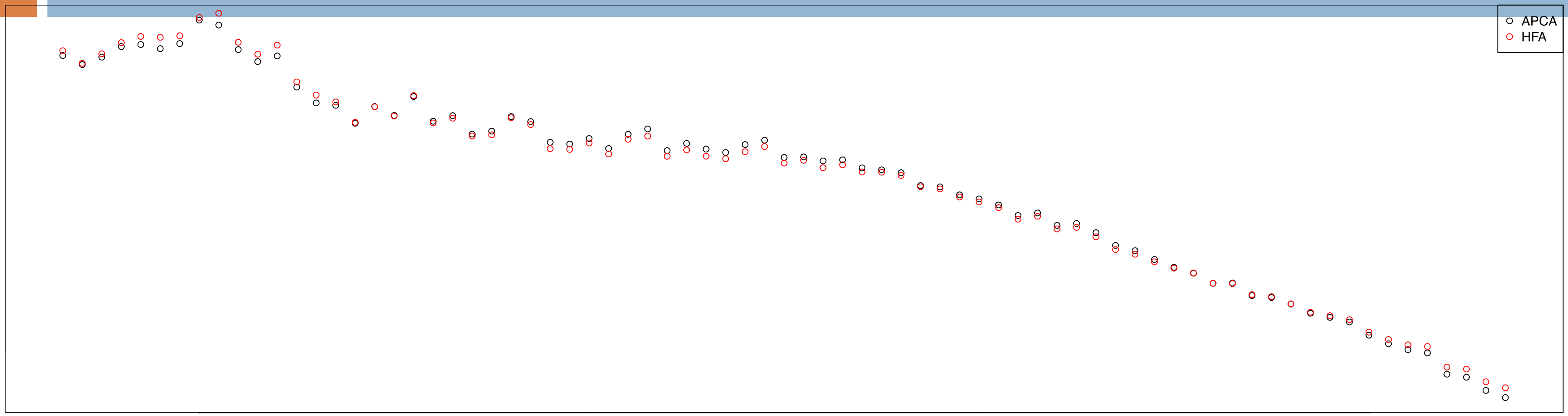
Loadings to Factor 1

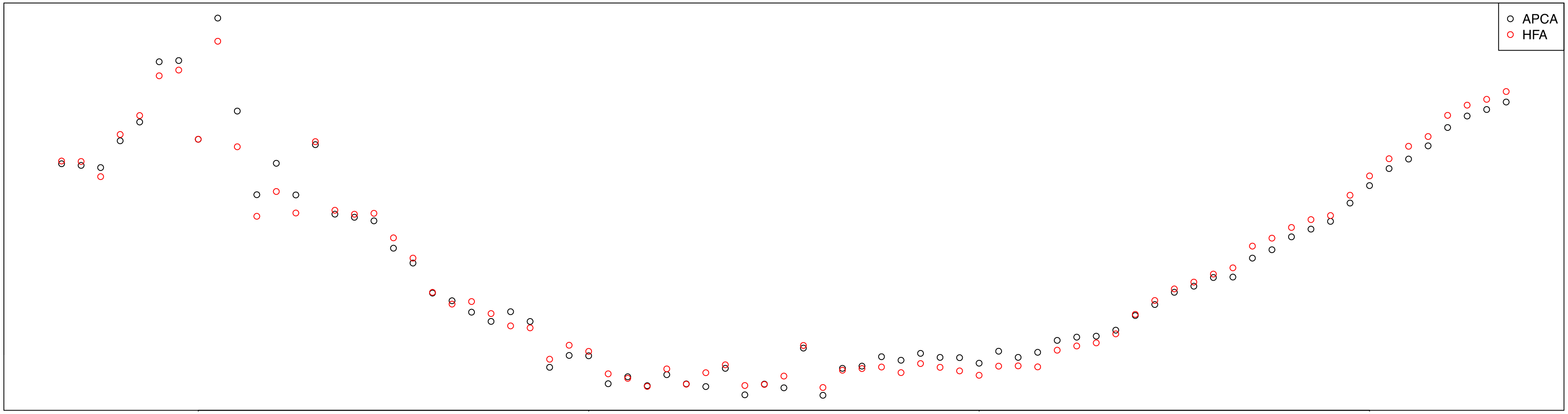


Loadings to Factor 2



Factors 1 and 2





Goodness of Fit (Male)

45-100

	ACM	PCA	APCA	HFA	POET	LC	CBD	Plat	M10
Log Likelihood	25928	25930	30678	30516	32237	25476	1222	28232	28579
BIC	-44103	-44805	-51523	-50254	-45884	-44533	5071	-45219	-45911

55-100

	ACM	PCA	APCA	HFA	POET	LC	CBD	Plat	M10
Log Likelihood	22665	21595	26313	26089	27229	22534	343	24217	23633
BIC	-38316	-37249	-44451	-42587	-41951	-39264	6680	-37410	-36241

65-100

	ACM	PCA	APCA	HFA	POET	LC	CBD	Plat	M10
Log Likelihood	17390	16222	20740	20309	21101	17080	-702	18885	18493
BIC	-28512	-27602	-34937	-32221	-33481	-28977	8584	-27024	-26240

Goodness of Fit (Female)

45-100

	ACM	PCA	APCA	HFA	POET	LC	CBD	Plat	M10
Log Likelihood	29599	29909	32291	31934	33705	29258	13520	31931	28911
BIC	-51446	-52762	-54749	-53089	-50362	-52096	-19527	-52616	-46576

55-100

	ACM	PCA	APCA	HFA	POET	LC	CBD	Plat	M10
Log Likelihood	25603	24893	27231	26980	28397	25499	12783	28160	26710
BIC	-44192	-43845	-46286	-44370	-41986	-45193	-18200	-45296	-42396

65-100

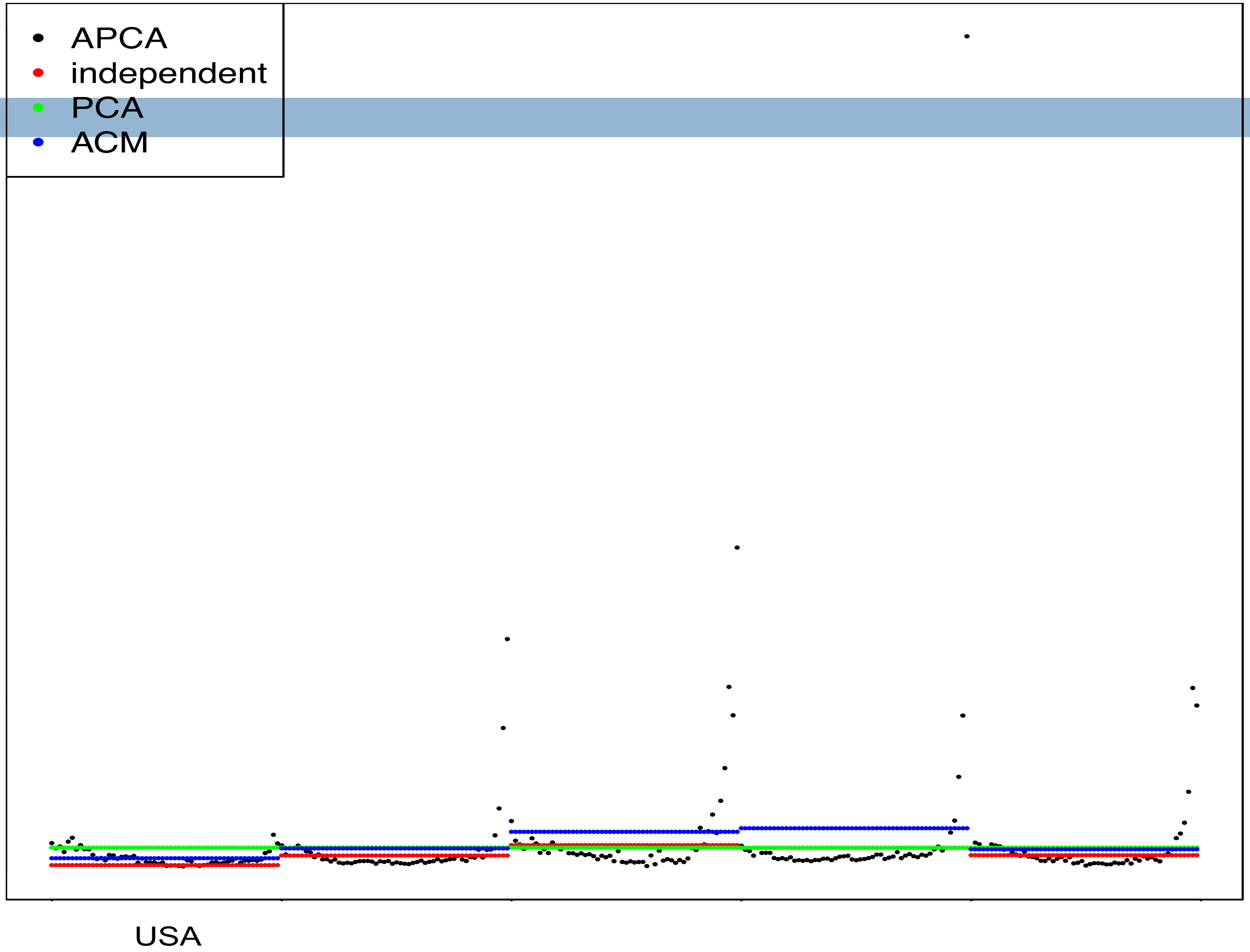
	ACM	PCA	APCA	HFA	POET	LC	CBD	Plat	M10
Log Likelihood	20395	19511	21662	21566	22452	20455	10330	22163	22283
BIC	-34523	-34182	-36781	-34734	-34814	-35727	-13480	-33579	-33819

- The results are robust for other sampling periods as well.

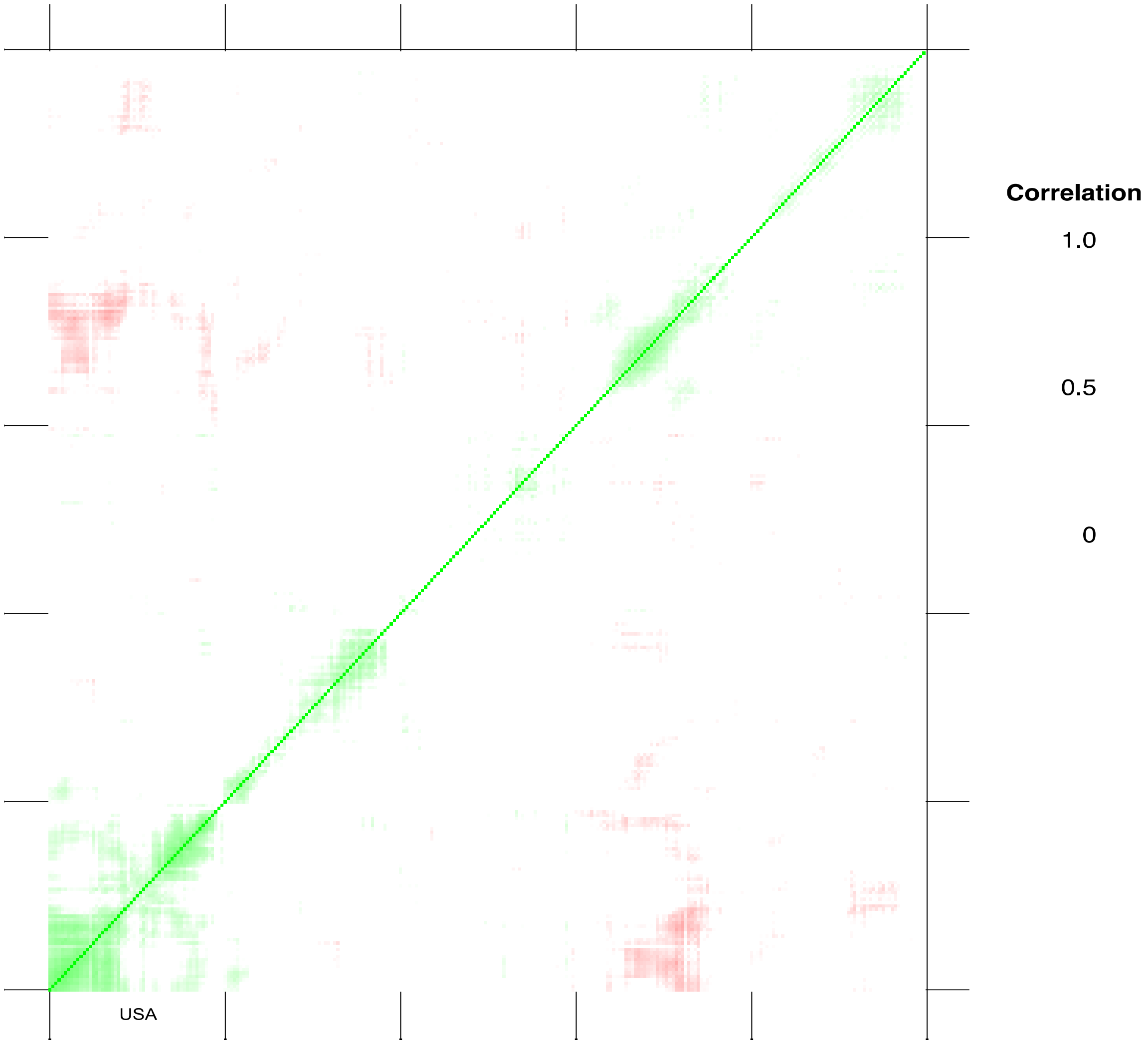
Remarks

- What are the impacts of different covariance structures?
  - Variances
    - Homoscedasticity – overshoot variances
  - Correlations
    - Assuming perfect correlation tends to overestimate risk
    - Assuming no correlations across populations contradicts the empirical observation that mortality across different countries move together

Residual variance comparison



POET Residual Correlations



Forecasting Accuracies

- In-sample period 1933 – 1997; Out-of-sample 1998 - 2007

45-100

	ACM	PCA	APCA	HFA	POET	Indep.	Plat	M10
RMSE	0.137	0.123	0.119	0.117	0.123	0.141	0.192	0.189
MAPE(%)	4.197	3.395	3.317	3.275	3.395	3.964	6.252	5.955

55-100

	ACM	PCA	APCA	HFA	POET	Indep.	Plat	M10
RMSE	0.116	0.129	0.124	0.119	0.129	0.129	0.142	0.157
MAPE(%)	4.050	3.859	3.722	3.558	3.859	4.154	4.435	4.948

65-100

	ACM	PCA	APCA	HFA	POET	Indep.	Plat	M10
RMSE	0.117	0.137	0.126	0.113	0.137	0.131	0.119	0.140
MAPE(%)	4.546	4.595	4.297	3.742	4.595	4.626	4.272	4.919

Conclusions

- The approximating factor models is theoretically sound on multi-population mortality modeling
  - It provide consistent estimate of model in N > T setting
  - It relaxes the assumption of no correlation between residuals and perfect correlation between mortality rates of different population.
- Our results show that they are empirically appealing in fitting and forecasting as well.
- The approximation factor model can also serves as a generalized Lee-Carter model in individual populations.