

2요인 선형 확률적 사망률 모형

엄영호*, 김계홍**

2009년 5월 22일

[요약문]

한 국가의 전체 사망률은 횡단면적으로 매시점마다 연령에 따라 지수함수형태로 증가하고 시계열적으로 모든 연령에서 하락하는 추세를 나타낸다. 이 논문은 사망률의 횡단면 및 시계열 추세 특성을 각각 보험수리분야의 Gompertz 모형과 재무분야의 이자율 기간구조모형으로 널리 활용되는 상태변수의 선형모형(Affine model)을 결합한 새로운 사망률 모형을 제시하고 있다. 연속시간 선형모형을 사망률에 적용시킬 경우 생존함수에 대한 닫힌 해(Closed-form solution)를 구할 수 있는 장점이 있다. 본 논문은 연속시간 선형모형을 사망률 적합에 사용한 기존 연구중에서 최초로 횡단면 및 시계열 자료 전체를 추정된 결과를 제시하고 있다. 추정결과는 기존 사망률 연구에서 선형확산 모형만을 이용한 방식의 설명력이 한계가 있으므로 횡단면 추이를 설명하는 함수를 포함시키는 것이 필수적이란 사실을 시사한다. 주요 분석대상은 한국과 미국의 남성사망률이며 이중 미국의 경우에 대하여 사망률 적합 및 예측 결과를 기존 사망률 모형중 횡단면 및 시계열 자료 전체를 이용하는 Lee&Carter (1992) 모형 및 Cairns, Blake&Dowd(2006) 모형과 비교하여 본 논문 모형의 성과가 우수함을 보이고 있다.

* 연세대학교 경영학과 교수, yeom@base.yonsei.ac.kr

** 연세대학교 경영학과 박사과정, 삼성금융연구소 수석연구원

1. 서론

사망률에 대한 올바른 추정은 사망률에 연동된 보험 및 금융상품의 가치평가와 리스크 분석의 정확도 제고를 위한 필수적 요소이다. 전통적인 보험상품에 대한 보험수리적 가치평가방법은 연령별 횡단면 사망률 추정치를 특정 연령의 미래 사망률로 이용하는 방식이다. 그러나 실제 사망률은 연령별과 시점별로 차이가 존재하므로 사망률 추정시 두 가지 요소를 동시에 고려하는 것이 중요하다. 사망률 자료가 갖는 특징적인 현상은 횡단면적으로 연령이 높아짐에 따라 사망률이 지수함수형태로 증가하는 점과 시간 경과에 따라 동일 연령의 사망률이 낮아지는 점이다. 전통적인 보험수리적 방법은 보험사의 경험 자료로부터 0세부터 100세까지 남성과 여성의 사망률을 추정된 결과인 경험생명표를 기반으로 보험상품의 가치를 계산하므로 실제 사망률 시계열에 존재하는 추세하락 특성을 반영하지 않거나 또는 별도의 단순한 가정을 통하여 이를 반영하는 단점을 지니고 있다. 그리고, 최근 몇 년 사이 전통적인 보험상품 이외에 사망률에 연계된 파생보험상품¹⁾이 나타나고 있다. 경험생명표는 보험사의 인수효과(underwriting effect)로 인하여 사망률 추정을 위한 표본에 편의(bias)가 있고, 외부에서 관찰하기 어려우며 또한 시계열 특징을 파악하기 위해 필요한 자료수가 확보되지 않는 단점이 있으므로 보험파생상품의 원리금 결정지표로 사용되기 어려운 것이 현실이다. 이러한 이유로 보험파생상품들은 보다 사망률 추정의 객관성을 확보하기 위하여 이자지급이나 상환액의 크기가 전체 인구중 특정연령군의 실제 사망자 수에 연동되어 결정되는 구조를 갖고 있다. 이 경우 사망률 추정 및 예측은 국민생명표를 이용하여 분석하는 것이 가능하다.

본 논문은 국민생명표상의 사망률 자료에 대한 추정 및 예측모형으로 재무분야의 이자율 기간구조모형을 채용하여 각 연령대별 사망확률 시계열에 적용시키는 방법을 다루고 있다. 기존 연구결과에서 선형 기간구조모형(Affine Term Structure Model)을 사망률 분석에 이용한 사례는 대부분 특정 시점에서 각 연령별 사망확률 또는 특정 연령에 대한 cohort 자료를 추정된 점이 한계로 지적될 수 있는데, 이 논문은 하나의 모형으로 모든 시간과 연령대의 사망률을 설명하는 방법을 제시하고 있다. 본 논문이 기존결과를 개선시킨 부분은 이자율 기간구조모형 외에 횡단면 사망률 추이를 설명하는 함수를 고려하여 모형의 적합도를 증진시킨 점과 전체 횡단면 및 시계열 사망률 자료를 추정된 점이다. 주요 분석모형은 횡단면 사망률 추이를 설명하는 Gompertz 함수와 시계열 추세를 설명하는 Vasicek(1977) 모형을 기본으로 잠재적 추세 요인변수가 각각 1개와 2개일 경우에 대하여 모수추정 결과를 비교하고 성과가 우수한 2요인의 경우에 대하여 기존 모형과 추정 및 예측성과를 비교하고 있다.

이후로 2장에서 사망률 기간구조와 관련된 기본 개념을 설명하고, 기존 사망률 설명모형중 연속시간 확산(continuous-time diffusion) 모형을 적용한 사례에 대하여 간략히 살펴보고, 3장에서 본 논문이 제시하는 모형을 설명한 후, 4장에서 추정 및 예측성과를 검토한다. 마지막으로 5장에서 결론을 언급한다.

1) 2003년 Swiss Re와 2004년 EIB/BNP Paribas의 Mortality bond를 고려할 수 있다.(Blake 등(2006) 참조)

2. 기존연구

2.1 생존함수와 사망률 intensity

생존함수 $S(t, x, T)$ 는 t 시점에 연령이 x 인 사람이 향후 $T-t$ 기간 동안 살아있을 확률을 의미하며 다음과 같이 정의된다.

$$S(t, x, T) = P(\tau_{t,x} > T-t) = 1 - F(t, x, T) = \int_{T-t}^{\infty} f(t, x, t+u) du$$

여기서,

- $\tau_{t,x}$: t 시점 x 연령의 사망 정지시간(stopping time) 확률변수
- F, f : 사망정지시간의 누적분포함수와 확률밀도함수

위 식을 경과시간에 대하여 미분하면 정지시간(stopping time) $\tau_{t,x}$ 의 확률밀도함수 $f(t, x, t+s)$ 는 다음과 같이 표현된다.

$$f(t, x, t+s) = - \frac{dS(t, x, t+s)}{ds}$$

$t+s$ 시점 $x+s$ 연령의 사망 가능성과 관련된 위태율(hazard rate) 또는 강도(intensity) 함수 $\mu(t+s, x+s)$ 는 다음과 같이 표현된다.

$$\mu(t+s, x+s) ds = \frac{f(t, x, t+s)}{S(t, x, t+s)} ds = -d \log S(t, x, t+s)$$

이 경우, 누적 강도는 다음과 같다.

$$M(t, x, t+s) = \int_0^s \mu(t+u, x+u) du = -\log S(t, x, t+s)$$

따라서 생존함수와 강도간 다음 관계가 성립한다.

$$S(t, x, T) = \exp(-M(t, x, T)) = \exp\left(-\int_0^{T-t} \mu(t+u, x+u) du\right)$$

여기서 강도가 확률과정일 경우 생존함수식은 다음과 같이 변형된다.

$$S(t, x, T) = P[\tau_{t,x} > T-t | F_t] = E\left[\exp\left(-\int_0^{T-t} \mu(t+u, x+u) du\right) \middle| F_t\right]$$

단, F_t 는 정보집합족(filtration)을 의미한다.

위 식의 기댓값은 강도가 선형확산과정(affine diffusion process)을 따를 경우 다음과 같은 지수선형 함수(exponential affine function) 형태로 표현될 수 있다는 사실이 알려져 있다.

$$E\left[\exp\left(-\int_0^{T-t} \mu(t+u, x+u) du\right) \middle| F_t\right] = \exp[A(T-t) + B(T-t)\mu(t, x)]$$

2.2 관련연구 고찰

여기서는 재무이론의 동태적 이자율 기간구조 모형(Dynamic Term Structure Model : DTSM)을 이용하여 확률적 사망률 모형을 구축한 사례를 중심으로 설명한다. 사망률 추정 및 예측을 위해 확산모형(Diffusion Model)을 적용한 기존 연구는 크게 사망률 중심과 생존함수 중심의 두가지 방식으로 나눌 수 있다. 첫 번째로 사망률중심 방식은 확산모형을 단순히 시간 또는 연령에 따른 사망률 추이를 나타내는 행태 방정식(behavioral equation)으로 사용하는 형태이다. 이 경우 확산모형을 통하여 예측한 값이 바로 미래 사망률이며, 생존함수는 예측된 사망률로부터 생존확률을 구하고 이를 축차적으로 곱하여 생성시킴으로 연속시간 이자율 기간구조모형의 순간(instantaneous) 이자율에 해당하는 순간 사망률 개념을 고려할 필요가 없다. 이와 관련된 대표적인 연구는 Milevsky & Promislow(2001), Schrager(2006), Biffis & Denuit(2005) 그리고 Biffis, Denuit & Devolder(2006) 등을 고려할 수 있다. 두 번째로 생존함수중심 방식은 순간 사망률을 고려하는 형태이다. 이 경우 확산과정은 순간사망률의 동학을 표현하며 이를 적분하여 생존함수가 생성된다. 이러한 방식과 관련된 대표적인 연구는 Dahl(2004) 과 Biffis (2005) 등이 있다.

먼저, Milevsky와 Promislow(2001)는 연금상품에 내재된 사망옵션을 평가하기 위하여 재무모형중 위험체권의 가치평가모형을 적용하였다. 즉, 사망과 관련된 미래 위태율(hazard rate)에 대하여 확률적 모형을 구축하고 이를 바탕으로 보험상품의 가치를 평가하는 체계를 구축하였다. 특히, 0시점에 연령이 x 인 사람의 위태율 동학을 다음과 같이 Gompertz 함수 형태에서 지수부분에 위치한 1개의 상태변수가 평균역전현상을 나타내는 브라운운동을 따르는 것으로 가정하였다.

$$\mu(0+t, x+t) = \mu(0, x) \exp(gt + \sigma Y_t),$$

$$dY_t = -bY_t dt + dW_t^h$$

위 모형은 비록 연속형 확률변수 형태로 위태율을 설명하고 있지만 앞에서 언급한 바와 같이 별도로 생존함수를 고려하지 않고 각 연도별 위태율을 해당년도의 사망확률로 고려하였다. 여기서 위태율의 동학은 이자율 기간구조 모형중 Black, Derman & Toy(1990) 모형과 유사한 형태이며 위 모형을 이용한 연금상품의 가치평가는 이항수행도(binomial tree)를 이용하여 미래 각 년도별 사망률을 생성시켜서 이용하였다. 특히, 이 논문은 70세 연금가입자의 가입이후 년별 사망확률을 분석하였다.

Dahl(2004)은 특정 연령(x)의 위태율을 다음과 같은 복수 리스크 요인을 갖는 확산모형으로 표현하였다.

$$d\mu(t, x) = \alpha(t, x, \mu(t, x))dt + [\sigma(t, x, \mu(t, x))]^T dW_t$$

이와 같은 확산모형에서 표류항(drift term) α 와 확산항(diffusion term) σ 가 특정 조건을 만족하는 경우, 연령이 $x+t$ 인 사람이 t 시점에서 T 시점까지 생존할 확률 즉, 생존함수 G 는 다음과 같이 표현되는 점에 착안하여 선형 확률적 사망률 모형(Affine Stochastic Mortality Model : ASMM) 체계를 구축하였다.

$$G(t, x, \mu(t, x), T) = E^P[e^{-\int_t^T \mu(u, x) du} | \mu(t, x)] = e^{A(t, x, T) - B(t, x, T)\mu(t, x)}$$

$$\text{단, } G(t, x, \mu_{[x+t]}, T) = S(t, x, T)$$

Biffis(2005)는 초단기 위태율을 두 부분으로 나누고 이를 바탕으로 생존함수를 보험계리적 생존확률(${}_{T-t}p_{x+t}$)과 상태변수에 영향을 받는 무작위항의 두 부분으로 구분하여 다음과 같이 표현하였다.

$$\mu(s) = \mu_m(s) + \Lambda(Y_s)$$

$$P(\tau > T | F_t) = E(e^{-\int_t^T (\mu_m(s) + \Lambda(Y_s)) ds} | F_t)$$

$$= e^{-\int_t^T \mu_m(s) ds} E(e^{-\int_t^T \Lambda(Y_s) ds} | F_t)$$

$$= {}_{T-t}p_{x+t} e^{A(t, T) + B(t, T)Y_t}$$

$$\text{단, } \mu(s) = \mu(0 + s, x + s)$$

상태변수는 다음과 같이 점프를 고려한 Vasicek 모형을 채택하였다.

$$dY_t = k(\bar{Y} - Y_t)dt + \sigma dW_t - dJ_t$$

지금까지 언급한 논문들은 모수들에 대하여 엄밀한 통계적 추정이 아닌 휴리스틱 접근을 통하여 계산하고 있으므로 실제적인 모형의 설명력을 판단하는데 다소 무리가 있는 것이 사실이다. 또한 위태율 또는 생존함수 모형에 대한 수치적 분석결과가 특정시점의 연령대별 사망률 또는 특정 연령의 시점별 사망률 추이에 한정되어 있으므로 사망률의 시계열 추세 반영부분에서 문제의 소지가 있다. 비록 Dahl(2004)과 Biffis(2005)이 제시한 모형은 각 시점별 및 연령별 사망률을 동적인 관점에서 설명하는 형태로 설정되어 있지만 실제 추정결과를 보여주지 않기 때문에 모형의 성과에 대한 직접적인 파악이 곤란한 점이 한계로 지적될 수 있다.

Schrager(2006)는 상태공간모형(State-Space Model)을 채용하여, 각 시점별로 전체 연령의 사망률을 설명하는 함수를 관찰방정식(measurement equation)으로 설정하고 관찰함수내 각 항의 계수가 선형확산 모형을 따르는 형태로 전이방정식(transition equation)을 구축하였다. 논문에서 고려한 관찰방정식은 다음의 두 가지 모형이다.

관찰방정식

Thiele 모형

$$\mu(t, x, t+1) = Y_{t,1} \exp(-\tau_1 x) + Y_{t,2} \exp(-\tau_2 (x - \eta)^2) + Y_{t,3} \exp(\tau_3 x)$$

여기서,

$$\mu(t, x, t+1) : t \text{ 시점 } x \text{ 연령의 1년 사망률}$$

Makeham 모형

$$\mu(t, x, t+1) = Y_{t,1} + Y_{t,2} \exp(x)$$

전이방정식

$$dY(t) = \kappa(\theta - Y(t))dt + \Sigma dW_t$$

Schrager(2006)는 위 모형을 칼만필터링(Kalman Filtering)을 적용하여 추정한 결과를 제시하였다. 이 모형은 단 한번의 추정으로 모든 연령에 대한 사망률 추이 분석이 가능하다는 장점을 지니고 있지만, 생존함수 생성방식이 아닌 사망률 생성방식을 채택하고 있으므로 특정 연령의 생존함수 계산을 위해 연령의 증가에 따른 미래 사망률 전체를 생성시켜야 하는 단점을 가지고 있다.

Biffis & Denuit(2005) 와 Biffis, Denuit & Devolder(2006) 는 대표적인 사망률 예측모형인 Lee & Carter(1992) 모형에서 사망률 추세지수를 확산과정으로 변환시키고 독특한 형태의 측도변환(measure change)을 적용하여 실제 사망률과 위험중립 사망률 모형의 구축을 시도하였다. 이들이 고려한 모형은 다음과 같다.

실제 사망률 :

$$\mu(t, x) = \exp(\alpha(x+t) + \beta(x+t) \cdot \kappa_t)$$

실제 사망추세지수 :

$$d\kappa_t = \delta(t, \kappa_{t-1})dt + \sigma(t, \kappa_t)dW_t$$

위 사망률 모형은 국민생명표상의 사망률로 부터 추정된다. 저자들은 국민생명표와 보험상품 가격결정에 사용하는 경험생명표를 구분하여 전자를 실제확률측도, 후자를 위험중립측도로 간주하였다. 측도변환을 위한 Radon-Nikodym 도함수는 다음과 같이 정의하였다.

$$\xi_t = \frac{d\tilde{P}}{dP}\Big|_{F_t} = 1 - \int_0^t \xi_s - \eta_s dW_s + \sum_{i=1}^n \int_0^t \xi_s - \phi_s^i dM_s^i$$

측도변환은 2단계로 이루어지는데, 먼저 위식 우변의 두 번째 항에 의해 실제 추세지수는 위험중립측도에서 다음과 같이 표현된다.

$$d\kappa_t = (\delta - \eta\sigma)dt + \sigma d\tilde{W}_t$$

보험상품 가격결정에 사용되는 경험생명표는 국민생명표와 변동성의 차이가 존재하므로 경험생명표에서 추정된 추세지수를 다음과 같이 표현한다.

$$d\tilde{\kappa}_t = \tilde{\delta} dt + \tilde{\sigma} d\tilde{W}_t$$

실제 추세지수와 위험중립 추세지수간의 변동성 차이를 다음과 같이 보정하므로써 2단계 측도변환이 이

루어진다.

$$\tilde{\kappa}_t = \frac{\sigma^{\tilde{\kappa}}}{\sigma} \kappa_t$$

이 경우 η 는 다음과 같이 계산된다.

$$\eta = \frac{\delta}{\sigma} - \frac{\tilde{\delta}^{\tilde{\kappa}}}{\sigma^{\tilde{\kappa}}}$$

또한, 위험중립측도하의 사망률은 다음과 같다.

$$\tilde{\mu}_t^x = \exp(\tilde{\alpha}(x+t) + \tilde{\beta}(x+t) \cdot \tilde{\kappa}_t)$$

이 모형은 실제 확률측도와 위험중립확률측도 구분을 시도했다는 의미가 있을 수 있지만, 측도변환시 일반적인 표류항 변환 외에 변동성을 조정한 부분에서 논란의 소지가 있을 수 있다. 부연하면, 두 번째 측도변환은 임의로 변동폭을 조정한 결과일 뿐이고 실제확률측도와 위험중립확률측도가 각각 다른 확률분포를 따르기 때문에 동등성(equivalence) 조건이 보장되지 않는 한계점이 존재한다.

3. 선형 확률적 사망률 모형

3.1 복수요인을 갖는 확률적 사망률 모형

$\mu(t, x, t+1)$ 을 t 년도에서 $t+1$ 년도까지 1년간 x 연령의 사망률(즉, t 년도의 q_x)로 정의하자. 특정시점에서 연령대별 사망률 곡선은 다음과 같이 Gompertz 모형으로 추정할 경우 적합도가 높게 나타나는 것으로 알려져 있다.

$$\mu(t, x, t+1) \sim \frac{\alpha}{\beta} e^{\beta x}$$

이 함수를 이용하여 $t+s$ 시점 $x+s$ 연령의 순간사망률을 다음과 같이 정의한다.

$$\mu(t+s, x+s) = \alpha e^{\beta(x+s)} + \iota^T Y(t+s) \quad (1)$$

위 식에서 우변의 첫 번째 항은 횡단면적으로 연령에 따른 순간사망률의 수준을 결정하고, 두 번째 항은 요인벡터로서 시간경과에 따른 순간사망률의 변화를 나타낸다. 그리고 ι 는 모든 원소가 1인 합벡터(sum vector)를 의미한다. 상태변수(Y)는 다음의 다변량 Vasicek 모형을 고려한다.

$$dY(t+s) = \kappa(\theta - Y(t+s))ds + \Sigma dW_{t+s} \quad (2)$$

여기서,

- κ : 평균회귀 속도를 나타내는 주대각행렬,
- θ : 안정상태 수렴값 벡터,
- dW_{t+s} : 브라운운동 벡터

생존함수중심 방식을 적용한 기존 문헌들은 일반적으로 모든 연령과 시점을 고려하는 형태로 모형이 설정되어 있지만 실제 추정결과를 제시하지 않은 점이 한계점으로 지적될 수 있다. 본 연구에서 제시하는 모형은 생존함수 중심 방식하에 단 한번 추정하여 전체연령의 생존함수에 대한 표시가 가능하다. 제시한 모형체계에서 t 년도에 x 연령인 사람이 미래 τ 년 동안 생존할 확률을 도출한 결과는 다음과 같다.

$$\begin{aligned} S(t, x, t+\tau) &= P(\tau_{t,x} > \tau) \quad (3) \\ &= E \left[\exp \left(- \int_0^\tau \mu(t+s, x+s) ds \right) \middle| F_t \right] \\ &= E \left(e^{-\int_0^\tau (\alpha e^{\beta(x+s)} + \iota^T Y(t+s)) ds} \middle| F_t \right) \\ &= e^{-\int_0^\tau (\alpha e^{\beta(x+s)}) ds} E \left(e^{-\int_0^\tau (\iota^T Y_{t+s}) ds} \middle| F_t \right) \\ &= \exp \left(\frac{\alpha}{\beta} e^{\beta x} (1 - e^{-\beta \tau}) \right) \exp(-A(\tau) - B(\tau)^T Y_t) \end{aligned}$$

여기서,

$$B(\tau) = (I - \exp(-\kappa^T \tau)) (\kappa^T)^{-1} \iota,$$

$$A(\tau) = (\iota \tau - B(\tau))^T \theta - \frac{1}{2} [\iota^T \kappa^{-1} \Xi(\tau) (\kappa^T)^{-1} \iota] \tau,$$

$$\Xi_{ij}(\tau) = \tilde{\Sigma}_{ij} \left[1 - \frac{1 - \exp(-\kappa_i \tau)}{\kappa_i \tau} - \frac{1 - \exp(-\kappa_j \tau)}{\kappa_j \tau} + \frac{1 - \exp(-(\kappa_i + \kappa_j) \tau)}{(\kappa_i + \kappa_j) \tau} \right],$$

$$\tilde{\Sigma} = \Sigma \Sigma^T,$$

단, $\exp(-\kappa^T \tau)$ 는 i 번째 주대각원소가 $e^{-\kappa_i \tau}$ 인 주대각행렬

사망률과 관련된 통계자료는 각 국가마다 매년 연령별 또는 연령군별로 1년간 사망확률을 추정한 결과를 발표하고 있다. 위 모형은 실제 연령별 1년 사망률 자료로부터 계산한 생존확률을 모형의 생존확률 함수식과 일치시키므로써 관련 모수들을 추정하게 된다. 이를 위하여 먼저 x 연령의 1년 사망확률과 생존확률간의 관계를 다음과 같이 설정한다.

$$e^{-\mu(t, x, t+1)} = S(t, x, t+1) \quad (4)$$

보험계리분야에서 1년 생존확률을 $1 - \mu(t, x, t+1)$ 과 같이 계산함에도 불구하고 위와 같이 설정한 이유는 사망률이 충분히 작은 경우 두 식의 계산결과가 유의미한 차이를 보이지 않고 또한 모수추정시 위 식과 같은 형태가 보다 편리하기 때문이다.²⁾ 위 식에 대수변환(log transformation)을 취하고 정리하면 다음과 같이 표현된다.

$$\begin{aligned} \mu(t, x, t+1) &= -\ln S(t, x, t+1) \quad (5) \\ &= -\frac{\alpha}{\beta} e^{\beta x} (1 - e^{-\beta}) + A(1) + B(1)^T Y_t \end{aligned}$$

위 식에서 Gompertz 함수는 전체 표본구간에 걸쳐 사망률의 횡단면 추이를 단 하나의 함수로 나타내고 사망률의 시간에 따른 변화가 이 함수에서 벗어나는 형태를 나타내는 것을 의미한다.

t 년도 x 연령의 $\tau(>2)$ 년동안 생존확률을 실제 자료로 표현하는 방법은 다음 두 가지를 고려할 수 있다.

$$S(t, x, t+\tau) = \begin{cases} \exp \left(- \sum_{i=0}^{\tau-1} \mu(t, x+i, t+1) \right) & \text{(I)} \\ \exp \left(- \sum_{i=0}^{\tau-1} \mu(t+i, x+i, t+i+1) \right) & \text{(II)} \end{cases} \quad (6)$$

첫 번째 방식은 현재 보험사에서 보험상품 가격결정시 사용하는 경험생명표에 의한 생존확률 계산방식 으로서 필요한 경우 여기에 확률의 변동 추세(trend)를 고려한다. 두 번째는 이론적으로 타당한 방식이

2) 만일 식(4)의 좌변에서 지수함수를 쓰지 않는 경우, 사망확률 표현식에서 이중지수함수가 나타나기 때문에 수치적 미분을 이용한 모수 추정시 최적값으로 수렴하는 초기치 설정이 쉽지 않을 수 있다.

지만 실제 보험가입자의 연령별 사망률 시계열자료의 입수 및 분석이 쉽지 않은 단점이 있다. 본 논문에서는 방식(I)에 의하여 각 년도마다 횡단면 연령별 사망률로 생존확률을 계산하고 이 형태가 시간에 따라 변동하는 추이를 상태변수로 표현하는 방식을 채용하고 있다. 이 경우 방식(II)의 생존확률은 Vasicek 모형으로부터 예측한 미래 상태변수를 대입하여 구하고자 하는 생존기간까지 매 1년 단위의 생존함수를 계산하여 다음과 같이 구할 수 있다.

$$S(t, x, t+1; Y_t)S(t+1, x+1, t+2; \hat{Y}_{t+1|t}) \cdots S(t+\tau-1, x+\tau-1, t+\tau; \hat{Y}_{t+\tau-1|t})$$

4장에서는 요인이 1개와 2개인 경우에 대하여 실제 추정결과를 보여주고 있는데 각 경우에 대하여 순간 사망률과 요인 동학은 다음과 같다.

1요인 모형

$$\begin{aligned} \mu(t+s, x+s) &= \alpha e^{\beta(s+x)} + Y_1(t+s) \\ dY(t+s) &= \kappa_1(\theta_1 - Y(t+s))ds + \sigma_1 dW_{t+s} \end{aligned}$$

이 경우 생존함수는 다음과 같이 표현된다.

$$\begin{aligned} S(t, x, t+\tau) &= \exp\left(\frac{\alpha}{\beta} e^{\beta x} (1 - e^{-\beta\tau})\right) e^{-A_1(\tau) - B_1(\tau) Y_t} \\ B_1(\tau) &= \frac{(1 - e^{-\kappa_1\tau})}{\kappa_1} \\ A_1(\tau) &= \left(\theta - \frac{\sigma_1^2}{2\kappa_1^2}\right) \left(\tau - \frac{(1 - e^{-\kappa_1\tau})}{\kappa_1}\right) + \frac{\sigma_1^2}{4\kappa_1^3} (1 - e^{-\kappa_1\tau})^2 \end{aligned}$$

2요인 모형

$$\begin{aligned} \mu(t+s, x+s) &= \alpha e^{\beta(s+x)} + \iota^T Y(t+s) \\ dY(t+s) &= \kappa(\theta - Y(t+s))ds + \Sigma dW_{t+s} \end{aligned}$$

$$\text{여기서, } \kappa = \text{diag}(\kappa_1, \kappa_2), \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

$$\iota = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$

이 경우 생존함수는 다음과 같다.

$$\begin{aligned} S(t, x, t+\tau) &= \exp\left(\frac{\alpha}{\beta} e^{\beta x} (1 - e^{-\beta\tau})\right) e^{-A_2(\tau) - B_2(\tau)^T Y_t} \\ B_2(\tau) &= (I - \exp(-\kappa^T \tau)) (\kappa^T)^{-1} \iota, \\ A_2(\tau) &= (\iota^T \tau - B(\tau))^T \theta + \frac{1}{2} [\iota^T \kappa^{-1} \Xi(\tau) (\kappa^T)^{-1} \iota] \tau \end{aligned}$$

앞에서 언급한 바와 같이 이 모형은 모든 연령대별 사망률 시계열을 단 한번 추정하여 각 연령대별 생존함수를 생성시키는 것이 가능하며, 모수들은 동적 기간구조 추정 문제와 관련된 기존 방법론을 적용할 수 있다. 다음에 모형의 모수 추정방법을 설명한다.

3.2 복수요인 확률적 사망률 모형 추정방법

제시한 모형은 사망률의 횡단면 특성 즉, 연령이 증가할수록 사망률이 지수적(exponential)으로 증가하는 특성을 나타내는 1개 함수를 중심으로 각 연도별 및 연령별 사망률과 이 함수와의 편차가 Vasicek 모형을 따르는 1개 또는 2개 요인으로 표시하는 형태이다. 이 모형의 모수들은 준최우추정법을 적용하여 추정이 가능한데, 이와 관련된 자세한 추정방법은 다음과 같다. 먼저, 관찰된 사망률이 관찰되지 않은 잠재적 요인에 의해서 시계열적 상관을 갖게 되므로 우선적으로 각 년도마다 미관찰된 요인을 추출하는 과정이 필요하다. Pearson&Sun (1994)은 이차율 기간구조 추정문제에서 관찰된 이차율로부터 잠재적 요인을 추출하여 조건부 확률밀도함수를 구축하는 방법을 제시하였는데, 이 방법은 우도함수를 요인의 시계열적 전이확률밀도함수와 사망률 예측오차의 2부분으로 나누어 고려하는 방법이다. 여기서 첫 번째 부분의 요인계산은 각 시점별로 요인수와 동일한 수의 자료를 이용하여 예측오차가 없다는 가정하에 사망률과 요인간 연립방정식을 설정하고 이를 풀어서 요인을 계산하는 것이고, 두 번째는 앞에서 계산된 요인을 동일시점의 다른 자료에 적용하여 사망률에 대한 예측오차를 생성시키는 방식이다. 이와 같은 방식에 의해 대수변환(log transformation)을 취한 우도함수는 다음과 같은 형태로 나타난다.

$$l(\phi, D) = \sum_t \left(-\ln|H_1| + \ln p(Y_t | Y_{t-1}, \phi) + \sum_\tau \ln(Z(t, \tau) | Y_t, D, \phi) \right) \quad (7)$$

(7)식 우변의 세 번째 항에서 $Z(t, \tau)$ 는 다음과 같이 정의된다.

$$Z(t, \tau) = H_0(\tau) + H_1(\tau) Y(t) \quad (8)$$

이 식을 다시 풀어쓰면 다음과 같다.

$$\sum_{i=0}^{\tau-1} \mu(t, x+i, t+1) = -\frac{\alpha}{\beta} e^{\beta x} (1 - e^{\beta \tau}) + A(\tau) + B(\tau)^T Y_t$$

대수변환 우도함수의 우변항은 다음과 같이 정의된다.

$$H_1 = \frac{\partial Z}{\partial Y}$$

$$\ln p(Y_t | Y_{t-1}, \phi) = -\frac{1}{2} (Y_t - E_{t-1}(Y_t))^T \Sigma_Y^{-1} (Y_t - E_{t-1}(Y_t)) - \frac{1}{2} \ln |\Sigma_Y|$$

$$\ln p(Z_t | Y_{t-1}, D, \phi) = -\frac{1}{2} \epsilon_t^T \Sigma_\epsilon^{-1} \epsilon_t - \frac{1}{2} \ln |\Sigma_\epsilon|$$

$$\epsilon_t = \sum_{i=0}^{\tau-1} \mu(t, x+i, t+1) + \frac{\alpha}{\beta} e^{\beta x} (1 - e^{\beta \tau}) - A(\tau) - B(\tau)^T Y_t$$

4. 추정결과 및 활용

4.1 자료

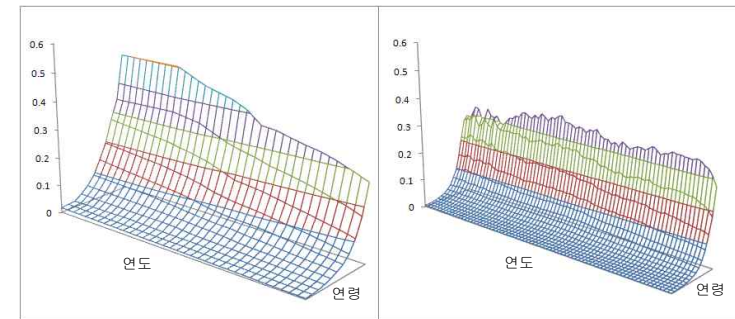
모형에 대한 실증분석을 위해 사용한 자료는 한국과 미국의 연령군별 남성 사망률 시계열로써 한국 자료는 통계청에서 제공하는 1970년부터 2005년까지 5세군별 남성 사망률 자료를 사용하였고, 미국은 Human Mortality Database³⁾에서 제공하는 1933년부터 2004년까지 5세군별 남성 사망률 자료를 사용하였다.

한국의 경우 25세군부터 75세군까지 총 11개 군을 고려하였고, 미국의 경우는 95세 연령군까지 총 15개 군을 고려하였다. 사용된 연령군은 25~29세(a25), 30~34세(a30), 35~39세(a35), 40~44세(a40), 45~49세(a45), 50~54세(a50), 55~59세(a55), 60~64세(a60), 65~69세(a65), 70~74세(a70), 75~79세(a75), 80~84세(a80), 85~89세(a85), 90~94세(a90), 95~99세(a95) 군이다. 다음 쪽의 표는 각 국가별 사망률의 기초 통계량을 나타낸다. 기초통계량에서 나타나는 특징적인 현상은 신생아군에서 10세군까지 사망률이 감소하다가 그 이후 연령이 높아질수록 다시 증가하는 것과 시간에 따른 변동성이 사망률 수준에 비해

한다는 사실이다. 한국은 모든 연령에서 미국보다 높게 나타나며, 75세군에 대하여 표준편차를 평균으로 나눈 변동계수(Coefficient of Variation)는 한국과 미국이 각각 18.9%와 17.8%로 나타난다. 사망률의 최소값은 가장 최근 사망률 추정치로 한국의 2005년과 미국의 2004년 자료이다. 한국과 미국의 시간에 따른 사망률 추이를 나타낸 [그림 1]은 양국에서 사망률이 전반적으로 하락하고 있지만, 한국의 사망률 개선속도가 보다 빠르게 나타나는 것을 알 수 있다.

[표 1] 한국과 미국 사망률의 기초 통계량

구분	한국				미국			
	평균	표준 편차	최대값	최소값	평균	표준 편차	최대값	최소값
a25	0.0102	0.0040	0.0167	0.0037	0.0021	0.0007	0.0040	0.0013
a30	0.0125	0.0044	0.0194	0.0049	0.0024	0.0008	0.0047	0.0014
a35	0.0179	0.0053	0.0247	0.0076	0.0033	0.0011	0.0066	0.0019
a40	0.0286	0.0086	0.0393	0.0128	0.0047	0.0015	0.0086	0.0029
a45	0.0436	0.0129	0.0607	0.0209	0.0072	0.0021	0.0118	0.0044
a50	0.0635	0.0195	0.0918	0.0311	0.0111	0.0031	0.0167	0.0063
a55	0.0927	0.0293	0.1414	0.0441	0.0166	0.0041	0.0227	0.0092
a60	0.1364	0.0402	0.1958	0.0682	0.0250	0.0054	0.0326	0.0141
a65	0.2002	0.0537	0.2760	0.1083	0.0365	0.0073	0.0482	0.0213
a70	0.2847	0.0593	0.3505	0.1714	0.0541	0.0099	0.0703	0.0327
a75	0.4126	0.0780	0.5030	0.2740	0.0804	0.0143	0.1070	0.0517
a80					0.1209	0.0191	0.1568	0.0820
a85					0.1809	0.0255	0.2321	0.1277
a90					0.2567	0.0248	0.3092	0.1809
a95					0.3353	0.0224	0.3784	0.2766



[그림 1] 한국(좌)과 미국(우) 사망률 추이

3) www.mortality.org 참조

4.2 사망률 모형 추정결과

횡단면 시계열자료 전체에 대한 실증분석은 Gompertz함수를 포함하고 요인을 각각 1개와 2개를 고려한 모형과 Vasicek 요인만 2개를 고려한 모형을 추정하여 그 결과를 비교하였다. 여기서 요인을 결정하기 위한 기준 사망률은 1요인의 경우 45세군을 사용하였고, 2요인의 경우 추가요인을 한국의 경우 65세군 그리고 미국의 경우 85세군을 사용하였다. 실제 자료가 군별로 존재하기 때문에 실제 생존함수는 각 군의 연령에 대하여 해당군의 대표 사망률을 동일하게 적용하여 생성하였고 각 연령군의 생존만기는 해당군에 속한 최고 연령의 사망만기, 즉 30세군의 경우 35로 간주하였다.

[표 2]와 [표 3]의 추정결과는 Vasicek 모형만을 고려한 경우보다 Gompertz 함수를 포함시킨 모형이 추정성평가 우월한 것으로 나타났다. 그리고 2요인 추정결과에서 나타나는 특이한 사실은 두 요인들이 평균회귀속도(κ)가 매우 작아서 결과적으로 요인 동학이 음의 표류향(drift)을 갖는 무작위보행(random walk) 과정과 유사한 형태로 나타난 점이다. 이 점을 보다 자세히 보기 위하여 요인들을 추출한 결과를 [그림 2]에 제시하였다. 30세군 사망률을 이용하여 추출한 1요인의 경우 사망률이 시간에 따라 점차적으로 하락하는 추세를 나타내고 있다. 30세군과 50세군 사망률을 이용하여 추출한 2요인의 경우는 첫 번째와 두 번째 요인이 각각 두 연령대의 사망률 상승 및 하락 추세를 나타내고 있는데 결과적으로 하락효과가 상대적으로 크게 나타나는 것으로 볼 수 있다. 기술적으로 사망률과 요인은 다음과 같은 관계를 갖는다.

$$Z(t, \tau) = \sum_{i=0}^{\tau-1} \mu(t, x+i, t+1) \propto B(\tau)^T Y_t$$

우변항의 $B(\tau)$ 를 45세군과 65세군에 대해서 계산한 결과는 다음과 같은데 식에서 50와 70은 각각 45세군과 65세군에 속한 최고연령인 49세와 69세의 사망만기를 의미한다.

$$B = \begin{bmatrix} B_1(50) & B_2(50) \\ B_1(70) & B_2(70) \end{bmatrix} = \begin{bmatrix} 31.99 & 27.89 \\ 51.99 & 41.75 \end{bmatrix}$$

이 경우 첫 번째 요인(Y_{1t})과 두 번째 요인(Y_{2t})은 각각 B_1 과 B_2 벡터방향으로 움직이며 그 상관관계 $\left(\frac{B_1 \cdot B_2}{\sqrt{\|B_1\| \|B_2\|}} \right)$ 는 0.9993으로 나타나므로 두 요인은 거의 유사하게 움직임을 알 수 있다. 이제 위 식을 요인에 대하여 정리하면 다음과 같다.

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} \propto \begin{bmatrix} B_2(70)Z(t, 50) - B_2(50)Z(t, 70) \\ B_1(50)Z(t, 70) - B_1(70)Z(t, 50) \end{bmatrix}$$

B 의 행렬식(determinant)이 123.69이므로 Y_t 의 부호는 위 식에 의해서 결정된다. 50세군 사망률에 대한 30세군 사망률의 비중이 1987년 21.5%를 정점으로 지속적으로 감소하고 있는데 이것은 사망률이 전 연령에 걸쳐 하락하며 동시에 기울기가 감소하고 있음 또는 상대적으로 높은 연령대에서 사망률의 개선 효과가 크게 나타나는 것을 의미한다. 따라서 30세군과 50세군의 누적 사망확률값의 비율 $\left(\frac{Z(t, 35)}{Z(t, 55)} \right)$ 이 $\frac{B_2(35)}{B_2(55)}$ 에 비하여 작으므로 첫 번째 요인은 음수로 증가하는 추세를 나타내고 두 번째 요인은 양수로 감소하는 추세를 나타낸다. 30세군 사망률을 이용하여 추출한 1요인의 경우 사망률이 시간에 따라 점차적으로 하락하는 추세를 나타내고 있으며, 30세와 50세군 사망률을 이용하여 추출한 2요인의 경우는 각각 상승과 하락추세를 나타내는 것을 살펴볼 수 있다. 일반적으로 주성분분석을 적용하거나 또는 이차원 기간구조 모형중의 하나인 Nelson&Siegel(1987) 모형과 같이 시계열의 형태적 특성을 고려하는 특정함수를 적용한 경우 각 요인들이 전반적인 시계열의 수준이나 기울기(보다 정확히 장단기 스프레드) 등의 움직임을 나타내는 것으로 해석할 수 있지만, 본 논문에서와 같은 선형모형의 경우 요인들이 1차 연립방정식으로 표현되기 때문에 개별 요인 고유의 성격을 특징짓는 것이 쉽지 않은 단점이 있다.

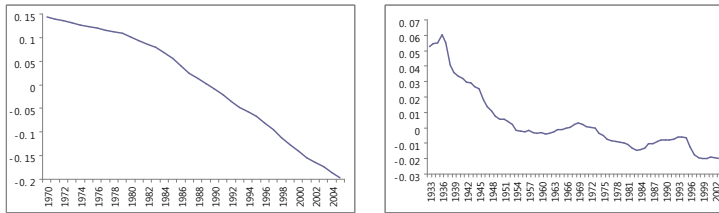
[표 2] 한국 사망률 추정결과(전체 시계열 대상)

변수	Gompertz+1요인		Gompertz+2요인		2요인	
	추정치	표준오차	추정치	표준오차	추정치	표준오차
α	0.000933	0.000053	0.001266	0.000217		
β	0.0529102	0.001084	0.074733	0.002111		
κ_1	0.113408	0.001300	5.10E-08	6.63E-09	1.70E-08	1.64E-08
θ_1	0.007690	0.001735	0.005775	0.001126	0.003334	0.001388
σ_1	0.017198	0.005700	0.050803	6.131096	0.055985	6.975357
κ_2			0.025603	1.158733	0.017100	5.941001
θ_2			0.000002	0.000001	0.000002	1.79E-07
σ_2			0.000815	0.000103	0.005693	0.000016
$\sigma_{e,25}$	0.011978	0.005029	0.006324	0.000676	0.025163	0.001939
$\sigma_{e,30}$	0.021794	0.002617	0.003559	0.000361	0.009636	0.000816
$\sigma_{e,35}$	0.013074	0.001575	0.001630	0.000197	0.004433	0.000717
$\sigma_{e,40}$	0.003668	0.000452	0.001164	0.000145	0.003613	0.000457
$\sigma_{e,50}$	0.000507	0.000062	0.001434	0.000173	0.005046	0.000592
$\sigma_{e,55}$	0.001182	0.000506	0.001124	0.000136	0.001733	0.000673
$\sigma_{e,60}$	0.008437	0.001018	0.001853	0.000210	0.006516	0.000970
$\sigma_{e,65}$	0.013330	0.001595				
$\sigma_{e,70}$	0.017760	0.002124	0.003815	0.000459	0.025562	0.003632
$\sigma_{e,75}$	0.023203	0.002784	0.005135	0.000435	0.016631	0.005900
logL	1530.5		1519.5		1417.1	

[표 3] 미국 사망률 추정결과(전체 시계열 대상)

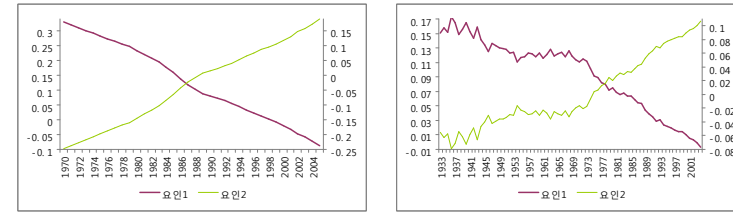
변수	Gompertz+1요인		Gompertz+2요인		2요인	
	추정치	표준오차	추정치	표준오차	추정치	표준오차
α	0.000194	0.000005	0.000328	0.000006		
β	0.077581	0.000338	0.072080	0.000172		
κ_1	0.697099	0.012886	0.000001	0.000001	0.000038	3.847
θ_1	1.603236	0.000063	0.005445	0.000120	0.006007	0.000171
σ_1	0.013030	0.000122	0.069805	3.396119	111.803528	17.632033
κ_2			0.025605	1.890387	111.845630	16.375696
θ_2			0.000004	0.475	0.000068	0.000011
σ_2			0.000398	0.000035	0.001166	0.000121
$\sigma_{\epsilon,25}$	0.004746	0.000407	0.001713	0.000144	0.004536	0.000446
$\sigma_{\epsilon,30}$	0.002078	0.000181	0.001073	0.000090	0.002363	0.000238
$\sigma_{\epsilon,35}$	0.001036	0.000093	0.000503	0.000042	0.000862	0.000091
$\sigma_{\epsilon,40}$	0.000553	0.000041	0.000250	0.000028	0.000452	0.000040
$\sigma_{\epsilon,50}$	0.000396	0.000022	0.000728	0.000061	0.000741	0.000056
$\sigma_{\epsilon,55}$	0.001084	0.000091	0.000966	0.000081	0.001228	0.000115
$\sigma_{\epsilon,60}$	0.001612	0.000135	0.001128	0.000094	0.001909	0.000180
$\sigma_{\epsilon,65}$	0.002228	0.000186	0.001174	0.000096	0.002594	0.000235
$\sigma_{\epsilon,70}$	0.002970	0.000248	0.001069	0.000090	0.002901	0.000255
$\sigma_{\epsilon,75}$	0.003981	0.000332	0.000882	0.000085	0.002264	0.000197
$\sigma_{\epsilon,80}$	0.005227	0.000434	0.001151	0.000096	0.004939	0.000420
$\sigma_{\epsilon,85}$	0.006809	0.000564				
$\sigma_{\epsilon,90}$	0.008018	0.000693	0.001938	0.000177	0.012865	0.001091
$\sigma_{\epsilon,95}$	0.007567	0.000594	0.002061	0.000177	0.023354	0.001976
$\log L$	4626.1		6188.7		5188.7	

Gompertz+1요인

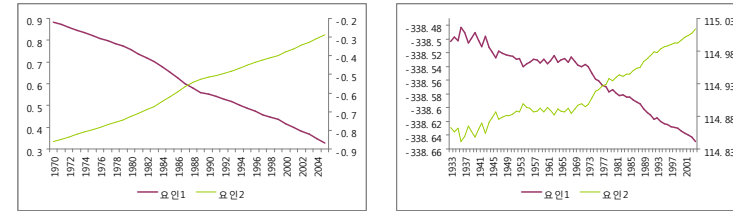


[그림 2] 한국(좌)과 미국(우)의 요인 추출 결과(계속)

Gompertz+2요인



2요인



[그림 2] 한국(좌)과 미국(우)의 요인 추출 결과

4.3 모형의 예측성과 비교

기존연구에서 연속시간으로 횡단면 및 시계열 사망률 자료를 동시에 추정할 모형이 없었으므로 본 논문에서는 기존에 발표된 이산시간 모형중 Cairns, Blake&Dowd(2006)(이하 CBD)와 Lee-Carter(1992)(이하 LC) 모형을 선택하여 미국 남성 사망률을 추정할 결과를 비교하여 설명한다. 우선 첫 번째로 CBD모형은 특정 연령(예를 들어 60세) 이상의 횡단면 및 시계열 사망률 자료를 설명하는 모형으로서, 각 시점에서 횡단면 연령별 사망률은 로지스틱함수를 이용하고 로지스틱 함수의 추정계수를 ARIMA모형으로 설정하고 있다. 자세한 모형식은 다음과 같다.

$$q_{t,x} = \frac{e^{A_{1t} + A_{2t}x}}{1 + e^{A_{1t} + A_{2t}x}}, \quad (9)$$

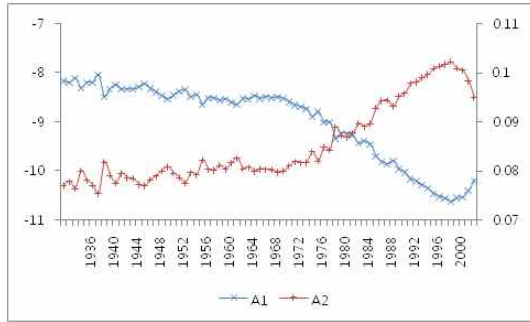
$$A_t = A_{t-1} + \mu + \eta_t \quad (10)$$

여기서, $A_t = [A_{1t} \ A_{2t}]'$, $\mu = [\mu_1 \ \mu_2]'$, $\eta \sim N(0, H)$

위 사망률 식에 대수변환을 취하고 정리하면 다음과 같이 선형함수형태가 도출된다.

$$\ln\left(\frac{q_{t,x}}{1-q_{t,x}}\right) = A_{1t} + A_{2t}x \quad (9')$$

CBD모형의 추정은 2단계로 이루어지는데 먼저 1단계에서 각 연도별 횡단면 사망률 자료에 대하여 (9') 식을 추정하여 추정계수(A_t) 시계열을 도출하고, 2단계에서 이 계수 시계열로부터 μ 와 H 를 추정한다. 본 연구에서는 60세군부터 95세군까지의 사망률을 추정하였으며 다음 [그림 3]은 1단계 계수 시계열 추정결과를 나타낸다.

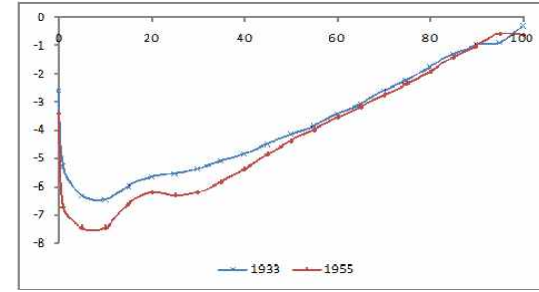


[그림 3] CBD 모형 로지스틱 함수의 계수 시계열 추이

2단계로 (10)식 추정결과를 다음과 같다.

$$\mu = \begin{bmatrix} -0.028451 \\ 0.000251 \end{bmatrix}, H = \begin{bmatrix} 0.012456 & -0.000185 \\ -0.000185 & 2.85E-6 \end{bmatrix}$$

한편, 횡단면 사망률 추이는 영유아기에 높게 나타나는 특성이 있으므로 [그림 4]에서와 같이 전체 연령에 대한 로지스틱 변환 추이가 비선형으로 나타나게 된다. 그러므로 CBD 모형은 로지스틱 변환후 비선형성이 크지 않은 일부 연령대에서만 설명이 가능하다.



[그림 4] 미국 사망률 로지스틱 변환 결과 : 1933년과 1955년

두 번째로 LC모형은 모든 연령별 그리고 시계열 사망률 자료를 대상으로 모형이 구축된다. 이 모형은 전체 사망률 자료를 특이값 분해(Singular Value Decomposition : SVD)를 적용하여 보다 적은 모수로 사망률의 횡단면 및 시계열 행태를 설명한다. 자세한 모형식은 다음과 같다.

$$\log q(t,x) = A_x + B_x K_t \quad (11)$$

$$K_t \sim ARIMA(p,d,q) \quad (12)$$

여기서, A_x, B_x : x 연령대 고유의 사망계수, K_t : 대수 사망률 추세 지수

위 모형의 추정도 2단계로 이루어지는데 먼저 $\sum B_x = 1$ 과 $\sum K_t = 0$ 의 두 조건을 가정한 후, 대수 변환 시킨 사망률을 각 연령군에 대하여 시계열 평균을 구한 값을 A_x 로 고려하고 각 사망률에서 평균을 차감한 값을 대상으로 특이값분해(Singular Value Decomposition)를 적용하여 각각 특이값과 특이벡터를 B_x 와 K_t 로 결정하는데 여기서 K_t 는 A_x 와 B_x 를 고정시킨 후, 각 연도별 사망자수($D(t)$)와 연령별 인구수($N(x,t)$)를 고려하여 다음 식으로 다시 계산하게 된다.

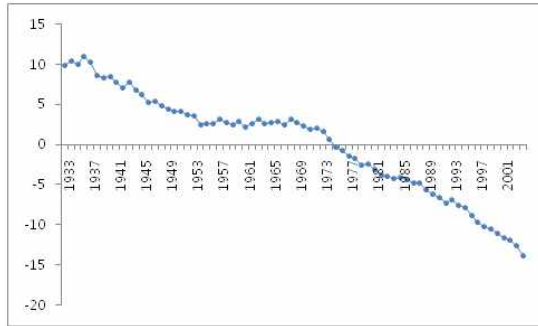
$$D(t) = \sum_x [N(x,t)e^{A_x + B_x K(t)}] \quad (13)$$

그리고 2단계에서 K_t 에 대하여 별도로 ARIMA모형을 적용하여 추정한다. 다음 표는 A_x 와 B_x 추정결과를 나타낸다.

[표 4] LC모형 A_x 와 B_x 추정결과

연령군	A_x	B_x	연령군	A_x	B_x
a0	-3.8115	0.1136	a50	-4.5449	0.0465
a1	-6.8832	0.1209	a55	-4.1318	0.0412
a5	-7.6306	0.1086	a60	-3.7170	0.0358
a10	-7.5567	0.0870	a65	-3.3324	0.0325
a15	-6.5563	0.0386	a70	-2.9358	0.0301
a20	-6.2089	0.0403	a75	-2.5379	0.0292
a25	-6.2187	0.0422	a80	-2.1262	0.0259
a30	-6.0728	0.0432	a85	-1.7198	0.0227
a35	-5.7772	0.0459	a90	-1.3647	0.0143
a40	-5.4013	0.0466	a95	-1.0951	-0.0044
a45	-4.9806	0.0465			

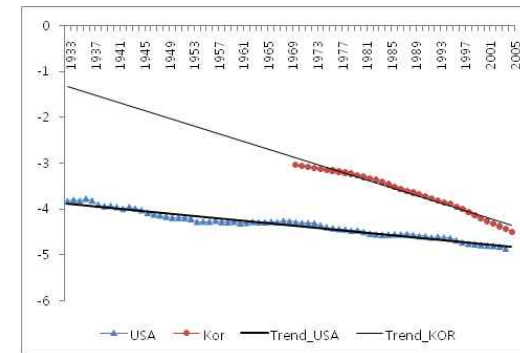
[그림 5]는 추정된 K_t 추이를 나타낸다.



[그림 5] LC모형 사망률 추세지수(K_t) 추이

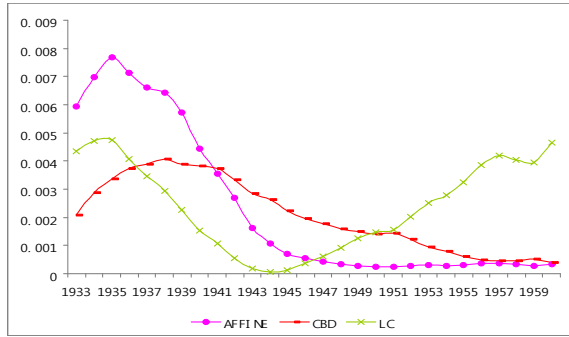
사망률 추세를 모형으로 설명할 경우 두가지 주의할 점이 존재한다. [그림 6]은 미국과 한국의 대수 변환 사망률의 연도별 평균의 추이와 선형추세선을 표시한 결과이다. 미국의 사망률 추세는 선형추세선을 중심으로 교차점이 여러번 나타나고 있지만, 한국의 사망률 추세는 선형추세선과의 교차점이 단 2번으로 지속적으로 감소하는 형태를 나타내고 있다. 이 그림에서 유추할 수 있는 첫 번째 주의점은 추세를 선형으로 나타낼 경우 대수변환된 사망률의 궁극적 예측(eventual forecasts)이 음의 무한대로 발산, 즉 사망률이 0으로 수렴하므로 생명이 무한하다는 의미를 갖게 되는 점이다. 현실적으로 사망률 예측이 미래 30년을 넘어서는 경우가 거의 없다 하더라도 선형추세를 통하여 예측할 경우 예측치의 적정성에 대한 판단이 필요하다. 두 번째 주의점은 한국의 사망률 추세가 지속적으로 감소하는(또는 아래로 오펜) 형태를 나타내는 것에 대한 처리문제이다. 지수적 감소추세는 지속될 수 없는 것이므로 이 문제는

시계열 자료수가 충분하지 못한 것에 기인한 것으로 볼 수 있다. 본 논문에서 채택한 Vasicek 모형은 수렴하는 성질을 지니고 있는데 특히 사망률 시계열의 경우 추정치가 아래로 불룩한 형태로 하락하면서 수렴하게 된다. 이와 같은 전형적인 유한표본편의(finite sample bias) 문제로 인하여 한국의 사망률에 대한 추정결과는 선형추세를 갖는 다른 2개 모형에 비하여 본 논문이 제시하는 모형(이후 Affine 모형)의 오차가 크게 나타난다. 이와같은 지수적으로 감소하는 추세 문제는 한국의 남성 뿐만 아니라 여성 사망률 시계열에서 동일하게 나타나는 현상이다. 이와 관련하여 Eom&Kim(2009)은 연도별 사망률 추세 부분을 이산형 로지스틱 쇠퇴(discrete-time logistic decay) 모형을 적용하여 추정 및 예측 성과를 개선한 결과를 제시하고 있으므로 이 논문에서는 다루지 않기로 한다. 향후 자료가 충분히 쌓인 경우 Affine 모형을 사용하는 것이 가능하지만, 자료가 부족하더라도 사망률에 연동된 수명채권의 가치평가를 위해서는 이 모형을 사용하는 것이 의미가 있을 것으로 기대된다.



[그림 6] 한국과 미국 사망률 시계열 추이

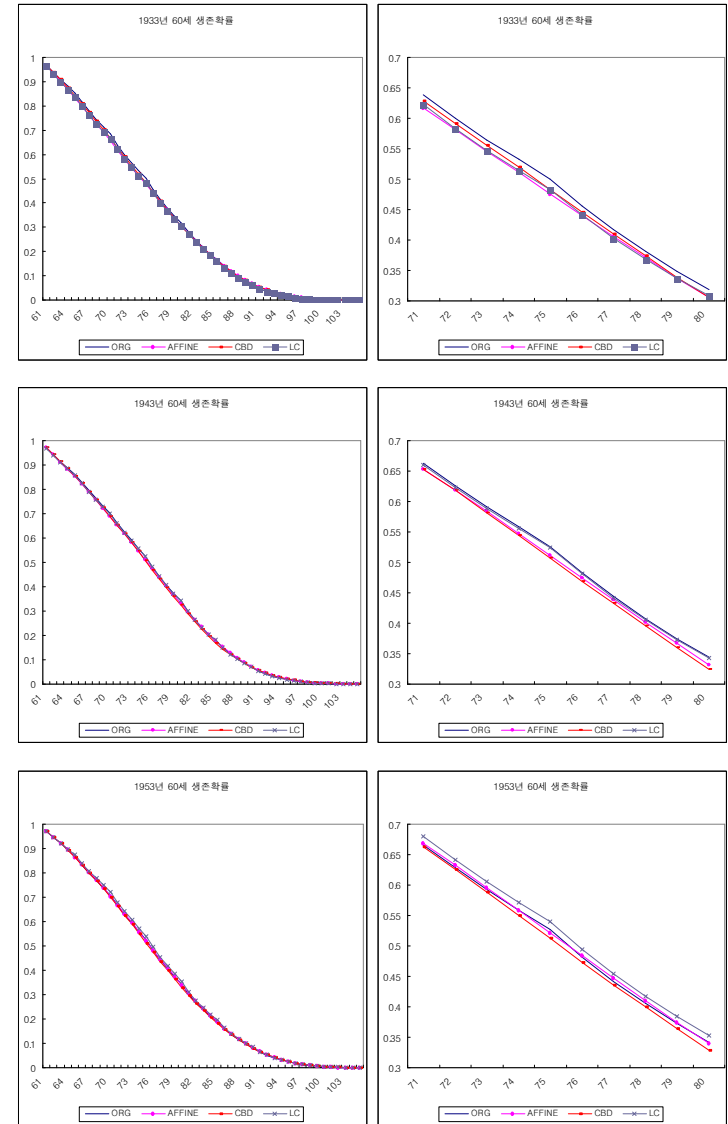
이후의 적합도 및 예측성과는 미국의 경우로 한정하여 분석한 결과를 비교한다. [그림 7]은 1933년부터 1960년까지 매년초 60세 남성의 해당시점 이후 100세까지의 생존함수를 3개 모형으로 추정한 오차의 제곱합을 표시한 결과이다. 이와 같은 방식으로 추정성과를 비교하는 이유는 3개 모형의 성격이 다소 차이가 있기 때문이다. 사망률 추정모형은 사망률의 횡단면 및 시계열 추이에서 나타나는 특징적 현상을 설명하는 것이 중요하다. 3개 모형중 Affine 모형과 LC모형은 횡단면 연령별 사망률 추이 및 연도별 사망률 추세를 동시에 고려하므로 추정오차에 횡단면 제약과 시계열 제약이 반영된 반면, CBD모형은 1단계 추정에서 각 연도마다 횡단면 연령별 사망률 추이를 하나의 모형으로 추정하기 때문에 추정오차가 횡단면 제약만을 반영하게 된다. 따라서 CBD 모형의 오차는 시계열 제약이 반영되지 않기 때문에 다른 2개 모형의 오차에 비하여 개념적으로 작게 나타날 가능성이 있다. 모형의 적합력 비교시 이와같은 차이점을 인식한 상태에서 그 의미를 분석할 필요가 있다.



[그림 7] 추정모형의 오차제곱합 비교

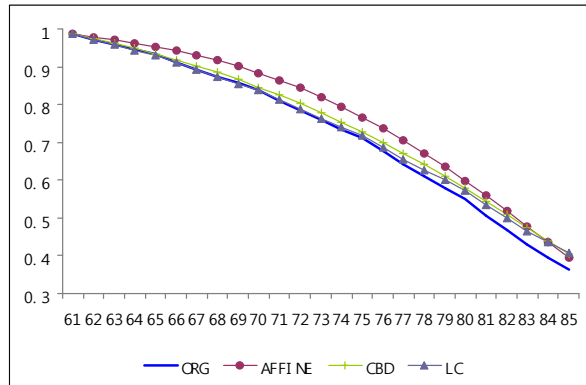
미국 남성의 사망률은 1933년 이후로 약 10여년간 전반적인 사망률 수준이 등락을 반복하면서 3개 모형의 적합력이 상대적으로 낮게 나타나고 있지만 그 이후로는 안정적으로 하락하는 추이를 나타내기 때문에 Affine 모형과 CBD 모형의 적합오차가 작게 나타나고 있다. 특히 1946년 이후로는 Affine 모형의 오차가 비교모형중 가장 작게 나타나고 있다. 여기서 LC 모형은 1944년까지 오차가 감소하다가 그 이후로 증가하는 형태를 나타내고 있는데 이 것은 각 시점에서 연령별 사망률 추이의 기울기(그리고 곡률)이 변동하기 때문인 것으로 판단된다. 즉, LC 모형은 SVD의 특성상 각 연령별 민감도(B_x)가 고정되어 있지만 Affine 모형은 2개의 추세지수를 이용하고 CBD 모형은 직접 회귀계수를 변동시키므로써 이러한 사망률 시계열의 특징적 현상에 대한 설명이 가능한 것이다.

다음 [그림 8]은 각각 1933년, 1943년 그리고 1953년 60세 cohort의 추정 생존확률과 실제생존확률 추이를 나타낸다. 왼쪽 그림은 60세 기준으로 105세까지의 생존확률을 나타낸 것이고, 오른쪽 그림은 차이를 보다 자세히 보기 위하여 바로 위 그림에서 70세부터 80세 구간의 생존확률을 별도로 표시한 결과이다.



[그림 8] 모형별 생존함수 추정값 비교

모형의 예측성과 비교를 위하여 60세 남성의 미래 25년동안, 즉 85세까지의 연별 생존확률을 예측하였다. 다음의 [그림 9]는 3개 모형으로 예측한 생존확률과 2004년 횡단면 실제 사망률로 계산한 생존확률을 나타낸다. 그림에서 2004년 실제 사망률에 의한 생존확률에 비하여 모형에 의한 생존확률이 전체적으로 높은 것은 3개 모형이 시계열 추세를 반영하므로써 시간경과에 따라 사망률 개선효과가 나타나는 것으로 볼 수 있다. 특히 모형간 사망률 개선효과의 경우 Affine 모형이 상대적으로 빠르게 나타나는 반면 CBD 모형과 LC 모형은 다소 천천히 나타나는 점이다. CBD 모형과 LC 모형의 추세는 음(-)의 표류향(drift)을 갖는 무작위보행(random walk) 과정으로 표시되므로 추정기간의 평균적인 하락추세를 반영하여 특정구간의 개선효과가 상대적으로 과소 또는 과대평가될 개연성이 존재하며 궁극적으로 사망률 예측치가 0으로 수렴(즉, 이 경우 생존확률은 0보다 큰 값으로 수렴)하게 된다. 반면에 Affine 모형의 경우 추세부분이 AR(1) 과정으로 표시되므로 단기적으로 (감소) 추세가 반영되지만 충분한 시간이 흐른후 사망률이 특정 확률로 수렴하게 되므로 생존확률이 0으로 수렴하게 된다. 이부분은 그림에서 생존확률의 상대적 개선효과가 84세 이후 역전되는 것으로 나타난다.



[그림 9] 모형별 생존함수 예측값 비교

4.4 수명채권 가격결정 사례

여기서는 2004년 유럽투자은행(European Investment Bank : EIB)과 BNP Paribas 수명채권과 같이 매년도마다 특정연령의 생존률만큼 이자를 지급하는 형태의 수명채권에 대한 가치평가 결과를 제시한다. EIB/BNP 수명채권은 2002년말 영국의 잉글랜드와 웨일즈 지역 65세 남성전체를 모집단으로 하여 미래 25년간 이 집단의 생존률에 5천만 파운드를 곱한 값을 지급하는 조건으로 발행되었다. 이 채권의 가격은 £540M이었으며, 이자지급은 2004년말부터 시작되었는데 자세한 이자지급 공식은 다음과 같다.

$$\begin{aligned}
 &2004년 : S(2002,65,2003) \times \pounds 50M \\
 &2005년 : S(2002,65,2004) \times \pounds 50M \\
 &\quad \vdots \\
 &2028년 : S(2002,65,2027) \times \pounds 50M
 \end{aligned}$$

다음의 [표 5]는 예측된 생존확률을 이용하여 60세 남성의 생존확률에 따라 원리금을 지급하는 수명채권(Longevity bond)의 가격을 계산한 결과로써, 각 연령 생존확률에 대한 할인율은 3%, 5% 그리고 2005년 1월 3일 달러화 스왑 기간구조를 이용하였다.⁴⁾ 이와 같은 방식으로 수명채권 뿐만 아니라 수명스왑(Longevity swap)에서 변동부(floating leg) 현금흐름을 설계하는 것이 가능하다. 어느 기간구조를 이용하든지에 상관없이 Affine 모형에 의한 채권가격이 가장 높게 나타남을 알 수 있다.

[표 5] 모형별 수명채권 가격비교

	ORG	Affine	CBD	LC
3% Flat	13.2972	13.9843	13.5572	13.4191
5% Flat	11.0278	11.5513	11.2317	11.1295
2005-01-03	13.2287	13.9107	13.5169	13.3699

기존 모형에서 사망률 추세에 대한 설명을 위해 표류향이 있는 무작위보행과정을 적용한 것은 사망률이 0으로 수렴하는 한계가 있지만, Affine 모형에서와 같이 0이 아닌 특정 사망률로 수렴하는 것은 상식적으로 의미가 있을 수 있다. 사망률 하한 예측치의 정확성 여부는 이 논문의 범위를 벗어나는 것으로 차후에 자료가 충분히 쌓일 경우 실증분석의 영역에서 해석이 가능할 것으로 기대된다.

4) 이자율과 사망률은 독립으로 가정하였다.

5. 결론

이 논문은 재무분야의 이자율 기간구조 모형중 Vasicek 모형과 전통적인 보험수리분야의 Gompertz 모형을 결합하여 사망률 자료를 추정하고 예측하는 모형을 제시하였다. 제시한 모형은 다음 3가지 측면에서 의미가 있을 것으로 기대된다. 우선 첫 번째로 이 논문은 연속시간 이자율 기간구조 모형을 사망률 적합에 사용한 연구중에서 최초로 횡단면 및 시계열 자료 전체를 추정할 결과를 제시하고 있다는 점이다. 전체 사망률에 대한 설명이 가능한 이유는 순간 사망률을 시간 뿐만 아니라 연령의 함수로 표현하므로써 모형의 설명력을 증진시킨 것에 기인하는 것으로 볼 수 있다. 두 번째는 기존 논문들이 이자율 기간구조 모형만을 이용한 것은 한계가 있으므로 본 논문에서와 같이 횡단면 추이를 설명하는 함수를 포함시키는 것이 필수적이라는 사실을 보인 점이다. 대부분의 연속시간 이자율 기간구조 모형은 횡단면 추이에서 곡률이 사망률의 추이와 부호가 반대로 나타나는 경향이 있다. 따라서 연령에 따라 사망률이 지속적으로 증가하는 추이를 나타내는 함수를 고려하고 평균역전현상을 갖고 안정적으로 수렴하는 모형을 사용하는 것이 설명력 증진에 보다 중요한 요소임을 보인 것이 의미가 있을 것으로 기대된다. 마지막으로 세 번째는 개념적인 문제로서 사망률의 추세를 Vasicek 모형으로 설명하므로써 사망률이 0보다 큰 어떤 작은 값으로 수렴하게 된다는 사실이다. 물론 사망률 하한의 의미는 본문에서 언급한 바와 같이 추후에 자료가 더 쌓일 경우 보다 다양한 해석이 가능할 것으로 기대된다. 본 논문의 한계점으로는 사망률 횡단면 및 시계열 추이 적합을 위하여 이미 알려진 모형 위주로 결합하고 있다는 점이다. 추후에 사망률의 특징적 현상을 반영하는 보다 적합한 확산과정에 대한 모색이 여전히 필요하며 제시한 모형의 성과가 모형 설명력의 우열을 판단하는 근거가 되기를 기대해 본다.

참고문헌

- [1] Babel, B., Bomsdorf, E. & Schmidt, R., Forecasting German mortality using panel data procedures, *Journal of Population Economics*, vol. 21, 541-555, 2008
- [2] Bauer, D., Kling, A. & Russ, J., A Universal Pricing Framework for Guaranteed Minimum Benefits in Variable Annuities, wp, 2006
- [3] Biffis, E., Affine processes for dynamic mortality and actuarial valuations, *Insurance: Mathematics and Economics* 37: 443-468, 2005
- [4] Biffis, E & Denuit, M, Lee-Carter goes risk-neutral: An application to the Italian annuity market, *Giornale dell'Istituto Italiano degli Attuari*, Vol: LXIX, 33 - 53, 2006
- [5] Biffis, E, Denuit, M. & Devolder, P, Stochastic Mortality Under Measure Changes, wp, 2005
- [6] Blake, D., Cairns, A.J.G., and Dowd, K. (2006) Living with mortality: longevity bonds and other mortality-linked securities, *British Actuarial Journal* 12: 153-197
- [7] Bongaarts, J., Long-Range Trends in Adult Mortality: Models and Projection Methods, wp, 2004
- [8] Cairns, A.J.G., Blake, D. & Dowd, K., Pricing death: Frameworks for the valuation and securitisation of mortality risk, wp, 2005
- [9] Cairns, A.J.G., Blake, D., and Dowd, K., A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration, *Journal of Risk and Insurance* 73, 687-718, 2006
- [10] Cox, S.H. & Lin, Y., Natural hedging of lif and annuity mortality risks, In Proceedings of the 14th International AFIR Colloquium, Boston, 483-507, 2004
- [11] Cox, S.H. & Lin, Y., Securitization of Mortality Risks in Life Annuities, *Journal of Risk and Insurance*, vol. 72, No. 2, 227-252, 2005
- [12] Dahl, M., Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts, *Insurance: Mathematics and Economics* 35, 113-136, 2004
- [13] Eom, Y. H & Kim, G. H., A Stochastic Mortality Model with Lower Bound : Lee-Carter(1992) Extension, wp, 2009
- [14] Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org
- [15] Lee, R.D., The Lee-Carter method for forecasting mortality, its various extensions and applications, *North American Actuarial Journal* 4: 80-93, 2000
- [16] Lee, R.D. & Carter, L.R., Modelling and forecasting U.S. mortality, *Journal of the American Statistical Association* 87: 659-675, 1992
- [17] Luciano, E. & Vigna, E., Non mean reverting affine processes for stochastic mortality, wp, 2006

- [18] Milevsky, M. & Promislow, S.D., Mortality derivatives and the option to annuitise, *Insurance: Mathematics and Economics* 29: 299-318, 2001
- [19] Nelson, C. R. & A. F. Siegel, "Parsimonious Modeling of Yield Curves," *Journal of Business*, 60, 1987, 473-489.
- [20] Pearson, Neil D & Sun, Tong-Sheng, Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model," *Journal of Finance*, vol. 49(4), pp. 1279-1304, 1994
- [21] Renshaw, A.E. & Haberman, S., On the forecasting of mortality reductions factors, *Insurance: Mathematics and Economics* 32: 379-401, 2003
- [22] Richards, S. & Gavin J., Financial aspects of longevity risk, wp, 2004
- [23] Schrage, D., Affine stochastic mortality, *Insurance: Mathematics and Economics* 38: pp. 81-97, 2006
- [24] Vasicek, O., An Equilibrium Characterization of the Term Structure, *Journal of Financial Economics*, 5, 177-188, 1977