

R을 이용한 분위수회귀 분석 : 경제외적 요인이 기대수명에 미치는 영향

김명중* · 박범조**

요약

1978년 Koenker와 Bassett에 의해 소개된 분위수회귀 추정량은 이상점(outliers)이나 오차항 분포에 민감하게 반응하지 않는 로버스트성을 갖고 있으며, 회귀모형의 종속변수에 대한 기대치나 중앙치를 추정하는 기존의 방법들과는 다르게 다양한 분위수값을 추정하여 종속변수분포의 특성을 알 수 있게 해주는 추가적인 장점을 갖는다. 이런 장점으로 인해 분위수회귀 이론과 추정 알고리즘에 대한 연구가 경제학 분야에서 활발하게 진행되고 있다. 따라서 본 연구는 박범조(2003)의 논문을 기초로 분위수회귀추정의 기본 개념 소개와 함께 분위수회귀 분석에 가장 적합한 소프트웨어 중 하나인 R 소프트웨어의 'quantreg'(Koenker, 2013) 패키지를 이용하여 경제외적 요인들이 국가별 기대수명에 미치는 영향을 실증 분석해 봄으로써 응용경제 분야의 전문가들에게 분위수회귀모형을 이용한 실증분석의 이해와 활용성을 증진시키고자 한다.

핵심주제어 : 분위수회귀 분석, R 소프트웨어, 'quantreg' 패키지, 분위수회귀 추정량의 점근적 특성, 경제외적 요인과 기대수명

I. 서론

분위수회귀(quantile regression) 분석은 1978년 Koenker와 Bassett에 의해 소개된 이후 종속변수의 조건부 분포에 대한 자세한 정보를 제공하고, 이상점이나 오차항 분포에 민감하게

* 단국대학교 상경대학 경제학과 대학원생, E-mail : mjkim@dankook.ac.kr

** 교신저자, 단국대학교 상경대학 경제학과 교수, E-mail : bjpark@dankook.ac.kr

〈논문 투고일〉 2013. 9. 2

〈논문 수정일〉 2013. 9. 22

〈게재 확정일〉 2013. 9. 24

반응하지 않는 로버스트성(robustness)을 가지며, 단조변형(monotone transformation)에 대해 등변성(equivariance)을 갖는 장점으로 인해 계량경제학 및 통계학은 물론 응용학문 분야에서 매우 중요한 연구 주제로 지속적인 발전을 해오고 있다(Powell, 1986; Chaudhuri, 1991; Jurečková and Prochazka, 1994; Koenker and Park, 1996; Koenker and Zhao, 1996; Koenker and Xiao, 2002; Koenker, 2005; Komunjer, 2002; Chakraborty, 2003; Yin and Cai, 2005; Chernozhukov and Hansen, 2008; Katoa, Galvao Jr. and Montes-Rojas, 2012; Galvao Jr., Lamarchez and Lima, 2013). 특히 최근 종속변수의 조건부 분포에 대한 극단(extrimal) 혹은 꼬리(tail) 분위수에 대한 관심이 집중되면서 정규분포에 기초한 분위수회귀 추론 방법보다 극단치(extreme value) 이론을 적용하여 자료의 중요한 특성을 도출하고 있다(Engle and Manganelli, 2004; Chernozhukov and Fernández-Val, 2011).

또한 최근 컴퓨터 연산 속도의 급속한 향상과 다양한 계산 알고리즘의 발전은 분위수회귀 추정과 추론을 이용한 실증적 연구와 활용성에 대한 관심을 확산시키고 있으며 종속변수의 조건부 분포의 집중화 경향만을 계산하던 과거의 연구에서는 도출할 수 없었던 다양하고 유용한 실증분석 결과를 얻고 있다. 초기의 실증분석은 주로 노동경제학 분야에 집중된 경향이 있었다. 예를 들어 노동시장의 임금구조 분석, 실업지속기간에 미치는 영향 분석, 학교 질이나 학급의 크기가 학생의 교육성과에 미치는 영향 분석 등에 분위수회귀 모형이 아주 유용하게 적용되었다(Chamberlain, 1994; Buchinsky, 1994; Park, 1997; Koenker and Biliias, 2001, Levin 2001). 하지만 최근에는 기업 창업규모의 조건부 분포(Mata and Machado, 1996), R&D투자와 판매량의 관계(Nahm, 2001), 정보기술수준과 기업성과(박범조, 2002), 그리고 한국 제조기업의 생산성과 수출 활동(김민중, 2013) 등과 같은 다양한 분야의 실증연구에 활용되고 있을 뿐만 아니라 금융 및 재무관리 분야의 분석에도 폭넓게 활용되고 있다. 예를 들어 VaR(value at risk) 모형에 적용(Taylor, 1999; Engle and Manganelli, 2004; Chernozhukov and Umantsev, 2001; Lee and Saltoğlu, 2002), 포트폴리오(portfolio) 투자 스타일 분석(Basset and Chen, 2001), EMS체제하의 외환 수익률의 비대칭 변동성(volatility) 연구(Park, 2002), 주식시장 수익률을 분석하기 위한 CAPM(capital asset pricing model)모형에 적용(Barnes and Hughes, 2003), 외환거래량, 환율변동성 그리고 GARCH 효과의 관계에 대한 분석(박범조, Park, 2007), 충격정보 확률변동성 모형을 이용한 외환시장의 변동성과 거래량의 관계 분석(박범조, 2008), G7 주식시장을 중심으로 살펴본 국제주식시장의 공분산 위험 동학(dynamics) 분석(Li and Yen, 2011) 등의 연구에 분위수회귀 분석이 적용되었다.

본 연구는 분위수회귀 분석을 위한 분위수회귀 모형과 추정 방법에 대해 설명하고 실증분석을 위해 통계분석을 위한 범용성과 전문성을 갖춘 R 소프트웨어¹⁾(R Core Team, 2012;

<http://www.r-project.org>와 분위수회귀 분석을 위한 패키지 ‘quantreg’(Koenker, 2013)를 활용한다. 인구구조의 고령화 자체가 국가 경제에 미치는 영향은 매우 심각하기 때문에 국가별 기대수명에 영향을 미치는 요인에 대한 실증적 연구는 중요하다. 따라서 본 연구는 국가별 기대수명에 대한 분위수회귀 모형을 설정하고 추정해봄으로써 유치원·초·중·고등 교육이나 흡연, 비만, HIV 출현율과 같은 경제외적 요인들과 기대수명의 관계를 면밀하게 분석해보고자 한다. 한편 이런 실증분석은 응용계량경제학자와 경제의 다양한 분야의 전문가들에게 R 소프트웨어를 이용한 분위수회귀 분석의 이해와 활용성을 향상시키는데 도움을 줄 수 있을 것으로 기대한다. 이 논문은 다음과 같이 구성된다. 제2절은 분위수, 표본분위수, 회귀분위수의 개념을 소개하고 선형 분위수회귀추정량의 점근적 정규성에 대해 간단히 설명한다. 제3절은 R 소프트웨어를 이용하여 국가별 기대수명의 분위수회귀모형을 추정하고 분석한다. 마지막 절에서는 실증분석 결과의 요약과 함께 결론을 간략하게 언급한다.

II. 분위수회귀 모형²⁾

1. 분위수

분위수회귀 모형을 쉽게 이해할 수 있도록 분위수의 개념을 살펴보자. 만일 어떤 사람이 거주하고 있는 도시에서 그 사람보다 소득이 낮은 사람의 비율이 τ 이고 소득이 높은 사람의 비율이 $(1-\tau)$ 라면 이 사람의 소득은 도시민들 소득에 대한 자료에서 τ^{th} 분위수(τ^{th} quantile)에 위치하게 된다. 예를 들어, 이 사람의 소득이 중앙값(median)인 0.5^{th} 분위수에 위치하면 (즉, 중앙값 = 0.5^{th} 분위수) 도시민들 중에 절반은 이 사람보다 소득이 높고 나머지 절반은 이 사람보다 소득이 낮음을 의미하게 된다. 또한 사분위수(quartiles)는 도시민들의 소득에 대한 자료를 넷으로 나누어 각 그룹에 속하는 관측치의 비율이 $1/4$ 로 동일하게 위치할 것이다.

이에 대한 개념을 수식으로 나타내기 위해 어떤 연속확률변수 X 가 있으며, 이 확률변수는

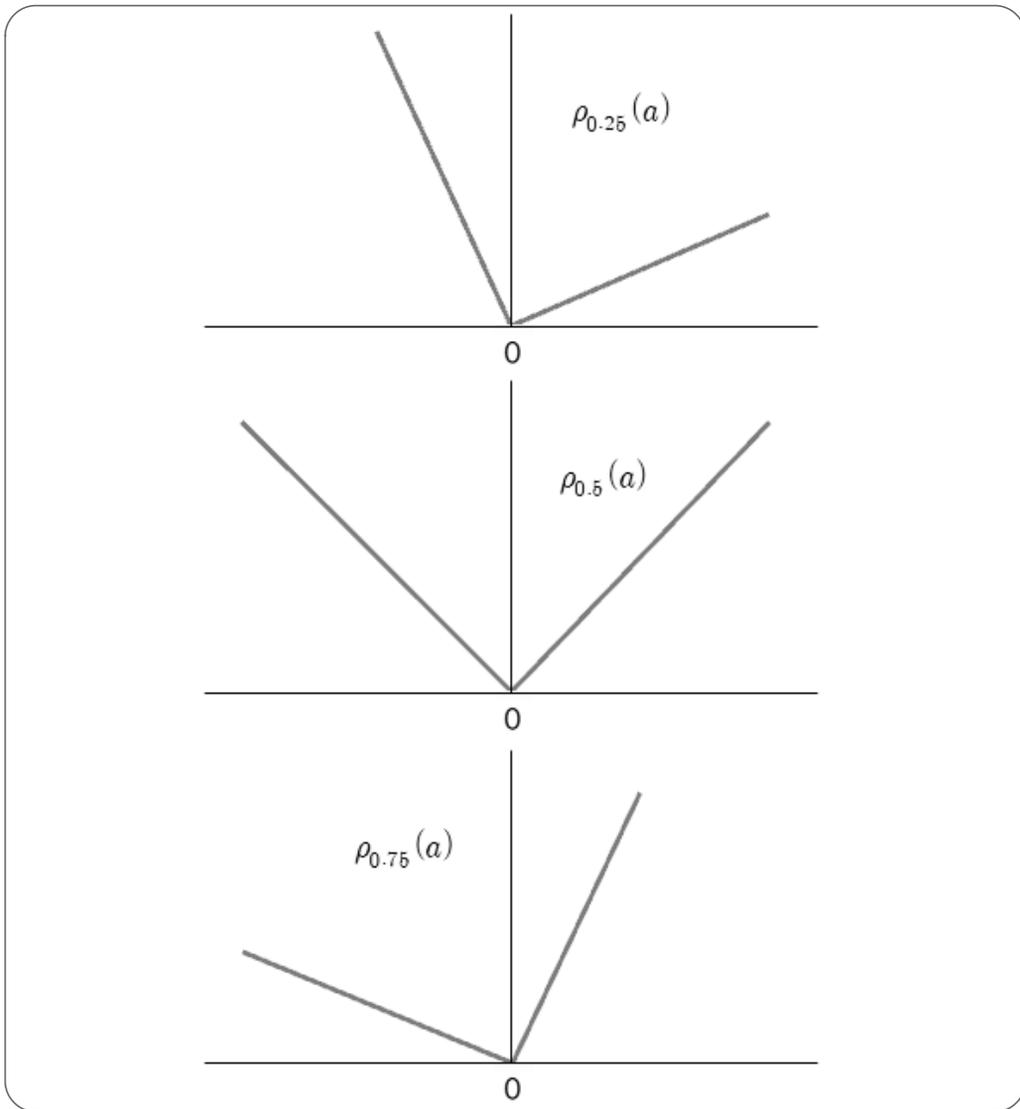
- 1) R 소프트웨어는 뉴질랜드 Auckland 대학의 Ross Ihaka와 Robert Gentleman에 의해 통계계산과 그래프 분석을 위해 개발된 프로그래밍 언어로 효율적인 자료 처리와 저장, 행렬 및 배열 형태의 연산, 자료분석을 위한 유용한 도구 모임의 제공, 그래프 분석을 위한 편리성 등의 장점을 갖는다. 특히 R 소프트웨어는 벨 연구소(Bell Lab.)의 Becker, Chambers, and Wilks(1988)에 의해 개발되어 통계학 분야에서 광범위하게 사용되어 왔던 S 언어와 유사한 언어구조를 갖기 때문에 통계분석과 관련된 매우 다양한 패키지들이 제공되어 그 활용성이 매우 높다(박범조, 2012).
- 2) 이 절의 내용은 박범조(2003, 2013)에서 인용 및 수정하였다.

다음과 같은 연속확률분포함수를 갖는다고 가정하자.

$$F(x) = P(X \leq x) \tag{2.1}$$

여기서 F 는 연속확률분포함수, P 는 확률함수, 그리고 x 는 X 가 취하는 특정한 값이다. 0과 1사이에 속하는 어떤 τ 에 대해 다음 등식을 만족하는 x 를 X 의 τ^{th} 분위수라고 정의한다.

[그림 1] 분위수회귀 $\rho_{\tau}(a)$ 함수



$$F^{-1}(\tau) = \inf \{x : F(x) \geq \tau\} \quad (2.2)$$

만일 $\tau = 0.5$ 일 경우 $F^{-1}(0.5)$ 는 0.5th 분위수인 중앙값을 의미한다. 분위수는 다음과 같이 표현되는 손실함수(loss function)로부터 구해질 수 있다.

$$\rho_\tau(a) = a(\tau - I(a < 0)) \quad (2.3)$$

여기서 $\tau \in (0, 1)$ 이며 $I(\cdot)$ 는 표시함수(indicator function)로 $a < 0$ 이면 1 아니면 0이 된다.

2. 표본분위수

τ th 표본분위수의 점추정치는 기대손실함수를 최소화하는 \hat{x} 가 되며 기대손실함수는 다음과 같다.

$$E[\rho_\tau(X - \hat{x})] = (\tau - 1) \int_{-\infty}^{\hat{x}} (x - \hat{x}) dF(x) + \tau \int_{\hat{x}}^{\infty} (x - \hat{x}) dF(x) \quad (2.4)$$

연속확률분포함수 F 가 단조함수(monotonic function)이므로 기대손실함수를 최소화하는 $\{x : F(x) = \tau\}$ 의 원소가 존재한다. 현실적으로 모집단자료 대신 표본자료를 이용하여 분위수를 추정하는 경우 연속확률분포함수를 모르기 때문에 다음과 같은 경험적 분포함수(empirical distribution function) F_n 를 대신 사용하여 추정하게 된다.

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x) \quad (2.5)$$

이 때 경험적 분포함수를 이용한 다음과 같은 기대손실함수를 최소화하는 \hat{x} 를 구할 수 있다.

$$\int \rho_\tau(x - \hat{x}) dF_n(x) = \frac{1}{n} \sum_{i=1}^n \rho_\tau(x - \hat{x}) = \min! \quad (2.6)$$

따라서 \hat{x} 는 τ th 표본분위수가 된다.

3. 분위수회귀

Koenker and Bassett(1978)에 의해 제안된 것처럼 이 표본분위수의 개념을 회귀식에 적용하기 위해 다음과 같이 단순한 위치모형(location model)을 가정한다.

$$y_i = \beta + \varepsilon_i \tag{2.7}$$

여기서 y_i 는 연속확률분포함수 F 를 갖는 i.i.d.(independent identically distributed) 확률변수이다. 이 때 표본분위수는 다음 식을 최소화하는 해(solution)가 된다.

$$\min_{\beta \in R} \sum_{i=1}^n \rho_{\tau}(y_i - \beta) \tag{2.8}$$

이 표본추정량의 개념을 회귀추정에 적용하기 위해 다음과 같은 선형회귀 모형을 가정한다.

$$y_i = x_i' \beta + \varepsilon_i \tag{2.9}$$

여기서 ε_i 는 0에 대해 대칭인 연속 확률분포함수 F 를 갖는 i.i.d. 확률변수이며 β 는 모수 벡터이다. 참고로 선형회귀모형에 대한 최소제곱추정량에 대해 알아보자. x 가 주어질 때 y 의 조건부 평균을 $\mu(x, \beta) = x' \beta$ 로 정의한다면 일반적인 최소제곱추정치 $\hat{\beta}$ 는 다음의 최소화 문제를 만족하는 해를 구함으로써 얻어진다.

$$\min_{\beta \in R^p} \sum_{i=1}^n (y_i - x_i' \beta)^2 \tag{2.10}$$

이제 유사한 방법으로 앞에서 구한 표본분위수의 개념을 선형회귀모형에 확장하여 분위수 회귀추정량을 도출할 수 있다. 보다 구체적으로 설명하면, x 가 주어질 때 y 의 τ^{th} 조건부 분위수 함수를

$$Q_y(\tau | x) = x' \beta_{\tau} \tag{2.11}$$

로 정의할 수 있으며, 분위수회귀추정치 $\hat{\beta}_\tau$ 는 표본분위수의 최소화 식의 스칼라 β 대신 $x'\beta$ 로 대체한 다음의 최소화 문제를 만족하는 해로부터 구해진다.

$$\min_{\beta \in R^p} \sum_{i=1}^n \rho_\tau(y_i - x_i' \beta) \quad (2.12)$$

특별히 $\tau = 0.5$ 인 경우 앞의 최소화 문제는 다음과 같아지며,

$$\min_{\beta \in R^p} \sum_{i=1}^n |y_i - x_i' \beta| \quad (2.13)$$

이 최소화 문제를 만족하는 분위수회귀추정치는 일반적으로 최소절대치(least absolute value) 혹은 l_1 -추정치라고 불린다.

Koenker and Bassett(1978)은 선형 분위수회귀추정량의 일치성과 점근적 정규성에 대해 증명을 하였으며, 후에 i.i.d. 오차항을 갖는 선형모형에 대한 경험적 분위수회귀 함수의 특성에 대해서도 설명하였다. 즉, 정규성 조건을 만족한다면 $\tau \in (0, 1)$ 에 대해 분위수회귀추정량 $\hat{\beta}_\tau$ 은 다음과 같이 점근적 정규성을 만족한다.

$$\sqrt{n}(\hat{\beta}_\tau - \beta_\tau) \rightarrow_d N(0, \tau(1-\tau)H^{-1}\Omega_x H^{-1}) \quad (2.14)$$

여기서 $H = E[f_y(X'\beta)XX']$

$$X = x_1, x_2, \dots, x_n$$

$$\Omega_x = E(XX').$$

III. R을 이용한 국가별 기대수명의 분위수회귀모형 분석

본 연구에서는 R 소프트웨어를 이용하여 실증분석을 수행하기 위해 인간의 기대수명(life expectancy at birth) 변화를 설명하는 회귀모형 설정하고, 2009년을 기준으로 전 세계 60개 국가에서 수집한 횡단면 자료(cross-section data)를 사용한다. 구체적인 회귀모형은 다음과

같으며, 기대수명 자료의 특성을 고려하고 다른 설명변수와의 관계를 면밀하게 분석하기 위해 기존 연구들(예, Gradstein and Kaganovich, 2004; Shaw et al., 2005)과는 다르게 분위수 회귀 분석을 수행한다.

$$Life_i = \beta_0 + \beta_1 White_i + \beta_2 Yellow_i + \beta_3 GDP.pc_i + \beta_4 Snt_i + \beta_5 S.Pre_i + \beta_6 S.Pri_i + \beta_7 S.Sec_i + \beta_8 S.Ter_i + \beta_9 Smk_i + \beta_{10} Obs_i + \beta_{11} Alc_i + \beta_{12} CO2_i + \beta_{13} HIV_i + \varepsilon_i, \quad i = 1, \dots, 60$$

여기서,

- A. $Life_i$: 기대수명(life expectancy at birth, year)³⁾
- B. $White_i$: 백인종이면 1의 값을 갖는 더미변수(dummy variable).
- C. $Yellow_i$: 황인종이면의 1의 값을 갖는 더미변수.⁴⁾
- D. $GDP.pc_i$: 1인당 실질 GDP(GDP per capita, constant 2005 US\$)
- E. Snt_i : 향상된 위생시설 확충 정도(improvement sanitation facilities, % of population with access)
- F. $S.Pre_i$: 유치원 재학 비율(School enrollment, preprimary, % of gross)
- G. $S.Pri_i$: 초등학교 재학 비율(School enrollment, primary, % of gross)
- H. $S.Sec_i$: 중등학교 재학 비율(School enrollment, secondary, % of gross)
- I. $S.Ter_i$: 고등학교 재학 비율(School enrollment, tertiary, % of gross)⁵⁾
- J. Smk_i : 흡연율(current smoking of any tobacco product, age-standardized⁶⁾ rate)
- K. Ob_i : 비만인구 비율(prevalence of obesity, BMI⁷⁾ ≥ 30 , age-standardized estimate)

3) 기대수명(life expectancy)이란 자살이나 각종 사고로 인한 사망률이 일정한 수준으로 유지된다고 가정했을 때, 주어진 연령에서 어떤 사람이 앞으로 얼마나 더 살 수 있을지를 통계적으로 계산한 평균 생존 연수를 말한다. 그리고 출생 시점(0세)을 기준으로 측정된 기대수명을 기대수명(life expectancy at birth)이라고 한다.

4) 개별 국가의 인종 분류는 대표인종으로 하였으며, 인종이 혼합되어 있는 경우 다수(50% 이상)를 차지하는 인종을 대표인종으로 하였다.

5) 유치원, 초등학교, 중등학교, 고등학교 재학 비율은 개별 국가에서 공식적인 재학 연령의 전체 인구 중 현재 해당 학교에 재학 중인 인구의 비율에 100을 곱한 값으로 정의된다.

6) 연령표준화(age-standardized)는 국가 혹은 지역별로 편차가 큰 연령분포를 동일하게 하여 연령구조가 가지는 효과를 제거하고 대상이 되는 어떤 특성의 절대적인 비교를 가능하게 한다.

7) 체질량지수(Body Mass Index)는 키가 t 미터, 체중이 w 킬로그램일 때 w/t^2 으로 정의되며, 일반적

2	ARM	73,64554	1	0	2913,601	90.1	31,74654	
3	AUS	81,54390	1	0	42403,596	100.0	80,88884	
4	AUT	80,08293	1	0	45859,426	100.0	96,10451	
5	BRB	76,41302	0	0	12885,783	100.0	113,18190	
	S.Pri	S.Sec	S.Ter	Smk	Obs	Alc	CO2	HIV
1	117,5184	88,53210	71,23081	27	29.7	9.35	4,365367	0.4
2	104,1306	92,95095	50,04239	27	24.0	13.66	1,513424	0.2
3	104,4562	129,22019	75,91419	21	26.8	10.21	18,375410	0.2
4	100,1030	99,62421	60,20413	46	20.9	12.40	7,449048	0.3
5	114,0570	101,20360	71,56275	7	34.7	6.42	5,638384	0.9

R에 불러와 출력된 자료에서 변수 이름은 각각 $Life=Life_i$, $White=White_i$, $Yellow=Yellow_i$, $GDP.pc=GDP.pc_i$, $Snt=Snt_i$, $S.Pre=S.Pre_i$, $S.Pri=S.Pri_i$, $S.Sec=S.Sec_i$, $S.Ter=S.Ter_i$, $Smk=Smk_i$, $Obs=Obs_i$, $Alc=Alc_i$, $CO2=CO2_i$, $HIV=HIV_i$ 를 나타내며, Console 2에는 summary() 함수를 이용하여 개별 변수들의 기본 통계량(최솟값, 1사분위수, 중앙값, 평균, 3사분위수, 최댓값)을 살펴보는 방법이 소개되어 있다.

[Console 2]

```

> summary(Life)

```

Code	Life	White	Yellow
ARG : 1	Min. : 50.54	Min. : 0.0000	Min. : 0.0000
ARM : 1	1st Qu. : 68.98	1st Qu. : 0.0000	1st Qu. : 0.0000
AUS : 1	Median : 73.73	Median : 1.0000	Median : 0.0000
AUT : 1	Mean : 72.50	Mean : 0.5167	Mean : 0.3333
BEL : 1	3rd Qu. : 78.81	3rd Qu. : 1.0000	3rd Qu. : 1.0000
BFA : 1	Max. : 82.93	Max. : 1.0000	Max. : 1.0000
(Other) : 54			
GDP.pc	Snt	S.Pre	
Min. : 343.4	Min. : 9.50	Min. : 2.818	
1st Qu. : 2414.2	1st Qu. : 70.58	1st Qu. : 50.134	
Median : 7635.2	Median : 95.35	Median : 77.580	
Mean : 16274.6	Mean : 80.93	Mean : 69.888	
3rd Qu. : 24546.3	3rd Qu. : 100.00	3rd Qu. : 96.150	
Max. : 78457.4	Max. : 100.00	Max. : 126.301	

S,Pri		S,Sec		S,Ter		Smk	
Min.	: 62.01	Min.	: 11.79	Min.	: 1,422	Min.	: 4.00
1st Qu.	: 98.92	1st Qu.	: 75.83	1st Qu.	: 27,931	1st Qu.	: 17.50
Median	: 103.07	Median	: 91.79	Median	: 51,152	Median	: 26.00
Mean	: 103.04	Mean	: 85.71	Mean	: 46,878	Mean	: 24.07
3rd Qu.	: 108.67	3rd Qu.	: 101.60	3rd Qu.	: 63,988	3rd Qu.	: 30.25
Max.	: 131.81	Max.	: 129.22	Max.	: 91,593	Max.	: 46.00
Obs		Alc		CO2		HIV	
Min.	: 1.10	Min.	: 0.320	Min.	: 0.04516	Min.	: 0.10
1st Qu.	: 13.38	1st Qu.	: 4.915	1st Qu.	: 1.45388	1st Qu.	: 0.10
Median	: 21.80	Median	: 8.755	Median	: 4.85957	Median	: 0.30
Mean	: 19.53	Mean	: 9.057	Mean	: 5.04705	Mean	: 0.64
3rd Qu.	: 26.30	3rd Qu.	: 12.635	3rd Qu.	: 7.46103	3rd Qu.	: 0.80
Max.	: 34.70	Max.	: 23.010	Max.	: 18.37541	Max.	: 6.20

주 : Min., 1st Qu. Median, Mean, 3rd Qu. Max.는 각각 최솟값, 1사분위수, 중앙값, 평균, 3사분위수, 최댓값을 나타내며, 분석 결과는 Table 1에 정리되어 있다.

종속변수인 기대수명의 기초통계량을 간략히 살펴보면 2009년도에 60개 국가를 대상으로 수집한 자료에서 최솟값은 50.54세, 최댓값은 82.93세, 평균 기대수명은 72.5세로 나타났으며, 1사분위수가 68.98세인데 비해 3사분위수는 78.81세로 나타난 것으로 보아, 종속변수는 왼쪽 긴 꼬리 분포를 보이고 있음을 대략적으로 알 수 있다.

Console 3은 R에서 분위수회귀 분석을 수행하고 결과를 반환하는 방법을 보여준다. (1)은 분석을 위한 명령문, (2)는 분석 결과를 출력하는 명령문이며, (3)은 개별 변수에 대한 계수 추정치의 변화를 보여주는 그래프를 생성하는 명령문이다. 분석에 사용할 분위수는 0.1, 0.25, 0.5, 0.75, 0.9로 하였다.

[Console 3]

```

> taus<-c(0.1, 0.25, 0.5, 0.75, 0.9)
> Q.Life.res<-rq(Life~White+Yellow+GDP.pc+Snt+S.Pre+S.Pri+
+ S.Sec+S.Ter+Smk+Obs+Alc+CO2+HIV, tau=taus, data=Life) (1)
> summary(Q.Life.res, se="nid") (2)

Call: rq(formula=Life~White+Yellow+GDP.pc+Snt+S.Pre+S.Pri+
S.Sec+S.Ter+Smk+Obs+Alc+CO2+HIV, tau=taus, data=Life)

tau: [1] 0.1

```

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	45,16861	0,91885	49,15802	0,00000 ***
White	6,45282	0,48033	13,43409	0,00000 ***
Yellow	5,53036	0,45670	12,10935	0,00000 ***
GDP.pc	0,00010	0,00001	8,23303	0,00000 ***
Snt	0,11806	0,01161	10,16704	0,00000 ***
S.Pre	0,08543	0,00202	42,24790	0,00000 ***
S.Pri	0,06542	0,00840	7,78884	0,00000 ***
S.Sec	-0,00300	0,01962	-0,15305	0,87903
S.Ter	-0,01216	0,00654	-1,85842	0,06951 .
Smk	-0,03453	0,01670	-2,06759	0,04433 *
Obs	0,12409	0,01204	10,30239	0,00000 ***
Alc	-0,25456	0,03684	-6,91066	0,00000 ***
CO2	-0,32481	0,03100	-10,47674	0,00000 ***
HIV	-1,84572	0,14435	-12,78613	0,00000 ***

Call: rq(formula=Life~White+Yellow+GDP.pc+Snt+S.Pre+S.Pri+S.Sec+S.Ter+Smk+Obs+Alc+CO2+HIV, tau=taus, data=Life)

tau: [1] 0,25

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	50,04387	6,00618	8,33207	0,00000 ***
White	4,79518	2,75506	1,74050	0,08846 .
Yellow	4,27352	2,70492	1,57991	0,12098
GDP.pc	0,00013	0,00003	4,18788	0,00013 ***
Snt	0,11388	0,03852	2,95652	0,00490 **
S.Pre	0,07734	0,02045	3,78205	0,00045 ***
S.Pri	0,04164	0,04290	0,97055	0,33685
S.Sec	-0,00358	0,04116	-0,08709	0,93098
S.Ter	-0,01445	0,03936	-0,36701	0,71529
Smk	-0,05311	0,04809	-1,10445	0,27514
Obs	0,11576	0,04853	2,38531	0,02124 *
Alc	-0,18220	0,07855	-2,31952	0,02486 *
CO2	-0,25659	0,13962	-1,83770	0,07257 .
HIV	-2,26777	1,23323	-1,83888	0,07239 .

Call: rq(formula=Life~White+Yellow+GDP.pc+Snt+S.Pre+S.Pri+S.Sec+S.Ter+Smk+Obs+Alc+CO2+HIV, tau=taus, data=Life)

tau: [1] 0,5

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	51,72549	5,46873	9,45841	0,00000 ***
White	4,40703	2,61394	1,68597	0,09857 .
Yellow	4,89720	2,45654	1,99353	0,05215 .
GDP.pc	0,00012	0,00002	5,07650	0,00001 ***
Snt	0,08934	0,04511	1,98044	0,05365 .
S.Pre	0,08158	0,02100	3,88413	0,00033 ***
S.Pri	0,02445	0,04945	0,49442	0,62336
S.Sec	-0,00248	0,04585	-0,05417	0,95704
S.Ter	0,03487	0,03278	1,06378	0,29298
Smk	-0,02364	0,05444	-0,43417	0,66619
Obs	0,08422	0,07462	1,12865	0,26490
Alc	-0,25935	0,10022	-2,58784	0,01288 *
CO2	-0,09480	0,16447	-0,57640	0,56716
HIV	-0,73215	1,15974	-0,63131	0,53096

Call: rq(formula=Life~White+Yellow+GDP.pc+Snt+S.Pre+S.Pri+S.Sec+S.Ter+Smk+Obs+Alc+CO2+HIV, tau=taus, data=Life)

tau: [1] 0,75

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	56,07466	5,17938	10,82652	0,00000 ***
White	5,90378	3,11090	1,89777	0,06401 .
Yellow	5,61239	2,63635	2,12885	0,03865 *
GDP.pc	0,00010	0,00004	2,66933	0,01047 *
Snt	0,08446	0,04533	1,86337	0,06880 .
S.Pre	0,04279	0,01772	2,41458	0,01979 *
S.Pri	0,01284	0,04481	0,28657	0,77573
S.Sec	0,04894	0,04276	1,14455	0,25831
S.Ter	0,00776	0,03838	0,20210	0,84073
Smk	-0,13812	0,06762	-2,04246	0,04686 *
Obs	0,06138	0,07017	0,87476	0,38625
Alc	-0,08566	0,09234	-0,92769	0,35841
CO2	-0,13557	0,11933	-1,13613	0,26179
HIV	-1,18577	0,60593	-1,95693	0,05644 .

Call: rq(formula=Life~White+Yellow+GDP.pc+Snt+S.Pre+S.Pri+S.Sec+S.Ter+Smk+Obs+Alc+CO2+HIV, tau=taus, data=Life)

```

tau: [1] 0.9
Coefficients:

```

	Value	Std. Error	t value	Pr(> t)
(Intercept)	57.63783	6.92427	8.32403	0.00000 ***
White	3.16932	3.68929	0.85906	0.39476
Yellow	4.52090	2.56925	1.75962	0.08512 .
GDP.pc	0.00014	0.00004	3.31337	0.00180 **
Snt	0.11295	0.05529	2.04275	0.04683 *
S.Pre	0.04857	0.02242	2.16625	0.03551 *
S.Pri	-0.00919	0.08592	-0.10693	0.91531
S.Sec	0.04471	0.06068	0.73692	0.46491
S.Ter	0.04403	0.04990	0.88231	0.38220
Smk	-0.14598	0.05953	-2.45215	0.01806 *
Obs	0.00151	0.11043	0.01368	0.98914
Alc	-0.00867	0.17918	-0.04837	0.96163
CO2	-0.28282	0.29877	-0.94660	0.34879
HIV	-1.18649	0.54331	-2.18381	0.03411 *

```

>plot(summary(Q.Life.res,se="nid"))
(3)

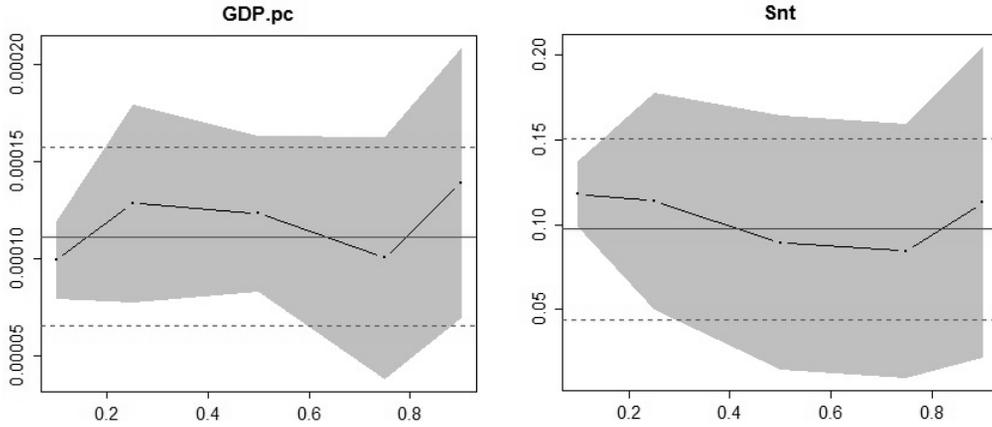
```

주 : '***' : 0.1%, '**' : 1%, '*' : 5%, '.' : 10% 유의수준에서 각각 유의함을 나타낸다. 분석 결과는 Table 2에 정리되어 있다.

명령문 (3)의 실행으로 생성된 그래프는 하나의 창(window) 안에 모두 그려지는데, 설명의 편의를 위해 그래프를 개별로 추출하여 설명하도록 한다. 분위수회귀 추정 결과를 살펴보면 다음과 같다.

본 연구에 사용된 회귀모형에서 경제적 요인이라고 판단할 수 있는 1인당 실질 GDP (GDP.pc)와 향상된 위생시설 확충 정도(Snt) 변수의 계수추정치 변화는 각각 [10% : 0.00010 → 25% : 0.00013 → 50% : 0.00012 → 75% : 0.00010 → 90% : 0.00014], [10% : 0.11806 → 25% : 0.11388 → 50% : 0.08934 → 75% : 0.08446 → 90% : 0.11295]로, 모든 분위수에서 두 변수는 기대수명 향상에 정(+)의 영향을 주지만, 기대수명이 점차 높아질수록 수명향상에 대한 영향력이 지속적으로 감소하다가 소폭 상승하는 경향을 보여주고 있다. 이는 인간의 기대수명 향상에 경제적인 요인이 여전히 중요한 변수이지만, 경제력 증가에 의한 기대수명 향상의 효과는 국가별로 차이가 존재하며, 기대수명이 높은 국가들 보다는 주로 기대수명이 낮은 국가들에서 그 영향력이 보다 큰 것으로 나타났다. 두 변수에 대한 계수 추정치 변화 그래프는 [Figure 1]에 나타나 있다.

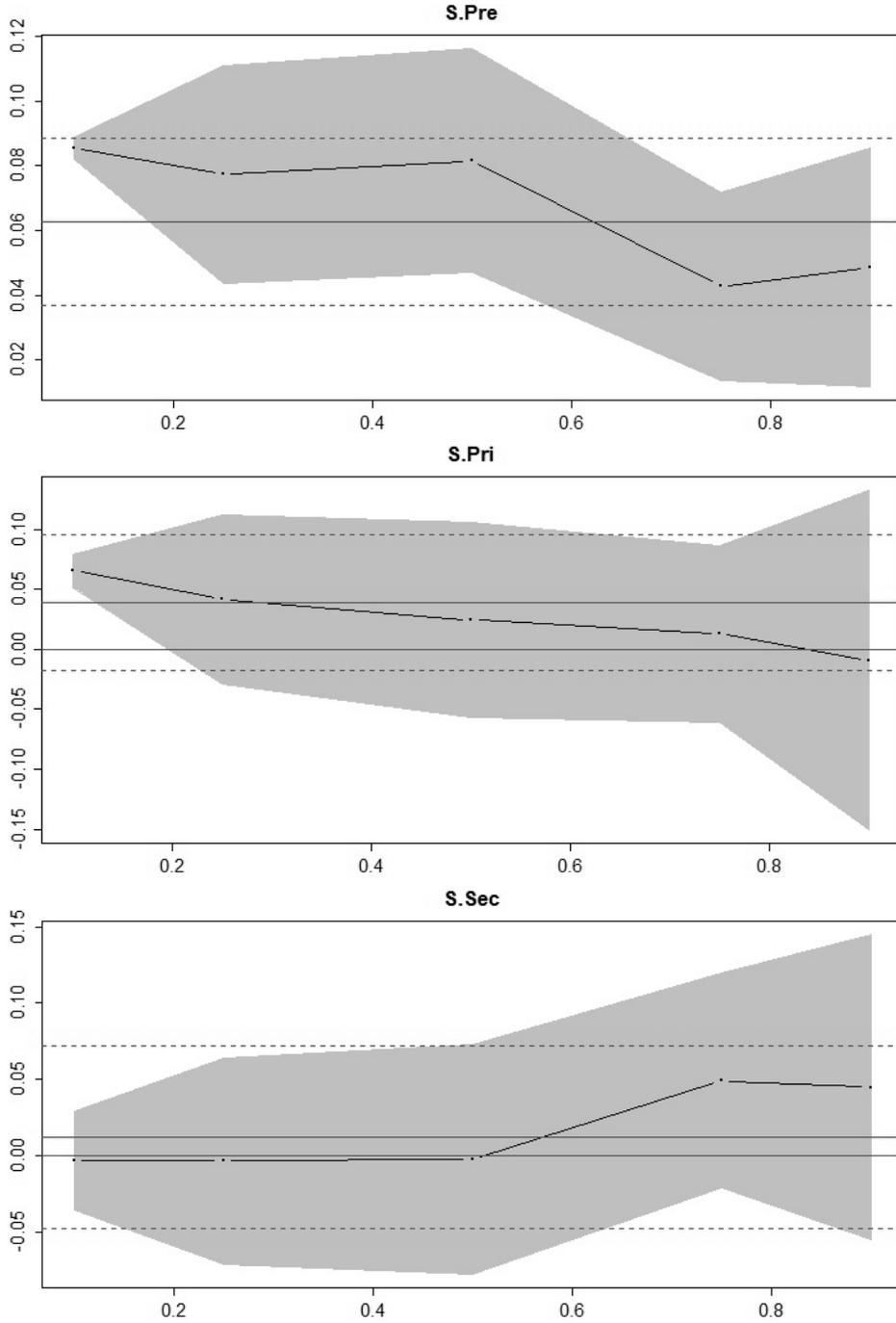
[Figure 1] 1인당 실질 GDP(GDP.pc)와 향상된 위생시설 확충 정도(Snt) 변수의 계수 추정치 그래프

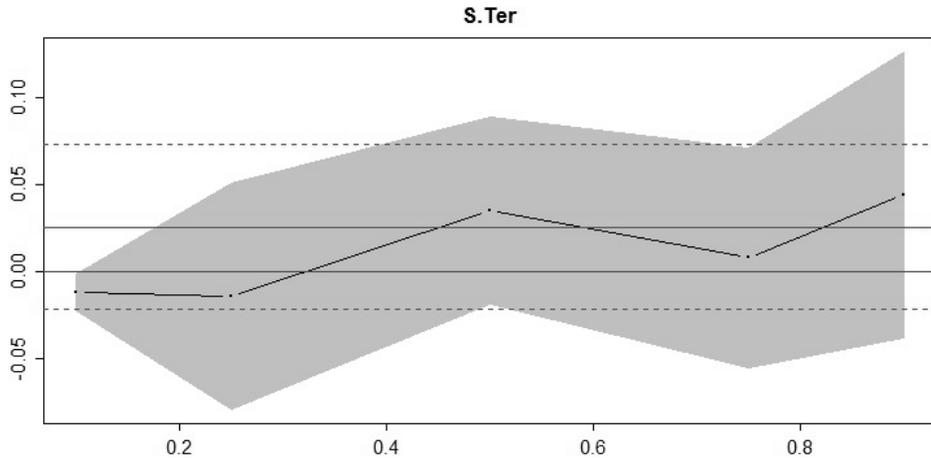


※ $\tau=0.2 \sim \tau=0.8$ 구간에서 공통적으로 기대수명 향상에 대한 두 변수의 영향력이 지속적으로 감소하고 있음을 살펴볼 수 있다. 참고로 회색 음영 처리된 부분은 분위수회귀 분석 결과에 대한 95% 신뢰구간이며, 수평으로 그은 실선과 점선은 각각 최소제곱추정(OLS) 결과와 95% 신뢰구간을 나타낸다.

경제외적 요인 중에서 교육에 대한 영향력을 나타내 주는 유치원·초·중·고등학교 재학 비율(S.Pre, S.Pri, S.Sec, S.Ter) 변수의 계수추정치를 살펴보면 전반적으로 교육은 인간의 기대수명 향상에 긍정적인 영향을 주고 있음을 알 수 있다. 하지만 유치원과 초등학교 재학 비율 변수의 계수 추정치 변화는 각각 [10% : 0.08543 → 25% : 0.07734 → 50% : 0.08158 → 75% : 0.04279 → 90% : 0.04857], [10% : 0.06542 → 25% : 0.04164 → 50% : 0.02445 → 75% : 0.01284 → 90% : -0.00919]로, 유치원 교육과 초등 교육을 받는 인구의 증가는 전반적으로 기대수명을 향상시키지만, 기대수명이 높은 국가들에 비해 상대적으로 기대수명이 낮은 국가들에서 기대수명을 더 크게 향상시키는 것으로 나타났다. 반면, 중등학교와 고등학교 재학 비율 변수의 계수 추정치 변화를 각각 살펴보면 [10% : -0.00300 → 25% : -0.00358 → 50% : -0.00248 → 75% : 0.04894 → 90% : 0.04471], [10% : -0.01216 → 25% : -0.01445 → 50% : 0.03487 → 75% : 0.00776 → 90% : 0.04403]로, 중등 교육과 고등 교육의 경우 유치원 교육이나 초등 교육과 달리 기대수명이 낮은 국가들에서는 수명 향상에 부정적인 영향을 주지만, 기대수명이 길어질수록 수명 향상에 긍정적인 영향을 주는 것으로 나타났다.

[Figure 2] 유치원 재학 비율(S.Pre) · 초등학교 재학 비율(S.Pri) · 중학교 재학 비율(S.Sec) · 고등학교 재학 비율(S.Ter) 변수의 계수 추정치 그래프





※ 기대수명이 늘어날수록 유치원이나 초등 교육 보다는 중등 교육 및 고등 교육이 기대수명 향상에 상대적으로 큰 영향을 줄 수 있다.

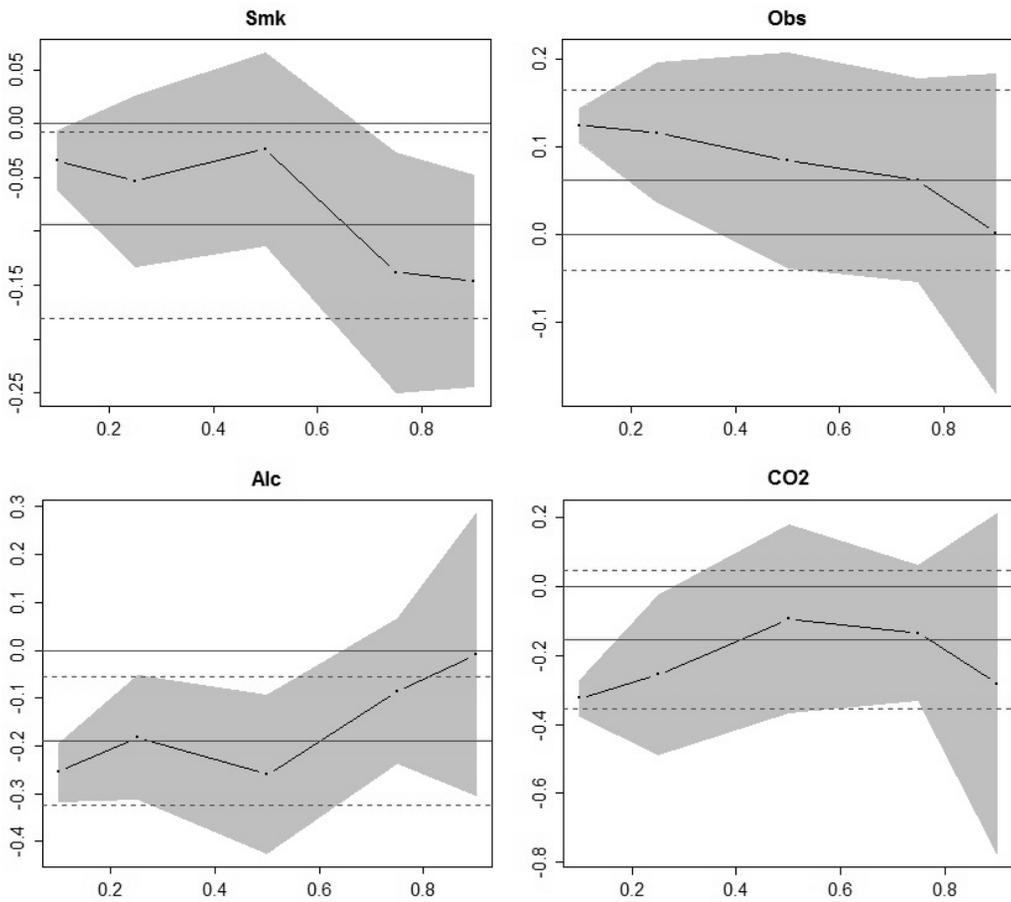
종합하면 기대수명이 낮은 국가들에서는 유치원이나 초등 교육이 기대수명 증가에 큰 정(+)의 영향을 주는데 반해, 기대수명이 점차 늘어나면서 유치원이나 초등 교육보다는 중등 교육이나 고등 교육이 기대수명 증가에 상대적으로 더 큰 정(+)의 영향을 줄 수 있다. 교육 관련 변수들에 대한 계수 추정치 변화 그래프는 [Figure 2]에 나타나 있다.

흡연이나 비만, 음주, 대기오염과 같은 비경제적 요인들은 기대수명에 부정적인 영향을 주는 것으로 나타났다. 흡연율(Smk) 변수의 계수추정치 변화는 [10% : -0.03453 → 25% : -0.05311 → 50% : -0.02364 → 75% : -0.13812 → 90% : -0.14598]로, 모든 분위수에서 흡연은 기대수명을 감소시키며, 인간의 기대수명이 점차 길어지면서 흡연량 증가에 의한 기대수명 감소 영향이 점점 더 커짐을 알 수 있다.

비만인구 비율(Obs) 변수의 계수추정치 변화는 [10% : 0.12409 → 25% : 0.11576 → 50% : 0.08422 → 75% : 0.06138 → 90% : 0.00151]로, 비만인구 비율의 증가는 전반적으로 기대수명 향상에 정(+)의 영향을 주는 것으로 나타났다. 이는 비만이 각종 성인병의 원인으로 건강을 위협하는 요인이기는 하지만 평균적으로 영양 상태가 좋아지는 경향을 반영하기 때문으로 해석된다. 그러나 [Figure 3]에서처럼 기대수명이 길어질수록 그 영향력이 지속적으로 감소하는 것으로 볼 때, 기대수명이 낮은 국가에서는 비만이 건강을 악화시켜 기대수명을 감소시키는 영향보다 영양상태의 호전으로 인한 기대수명 증가 효과가 크지만, 기대수명이 높은 국가에서는 상대적으로 비만이 건강을 악화시키는 효과가 커지면서 기대수명 증가에 대한 영향력이 감소하는 것으로 해석할 수 있다.

1인당 주류 소비량(Alc) 변수의 계수추정치 변화는 [10% : -0.25456 → 25% : -0.18220 → 50% : -0.25935 → 75% : -0.08566 → 90% : -0.00867]로, 음주가 건강상태를 악화시켜 기대수명을 감소시킬 것이라는 일반적인 통념에 부합하는 결과를 보여주었다. 모든 분위수에서 주류 소비 증가는 인간의 기대수명을 감소시킨다. 특히 주류 소비량의 증가는 기대수명이 높은 국가에서보다 낮은 국가에서 상대적으로 기대수명 감소에 더 큰 영향을 미친다.

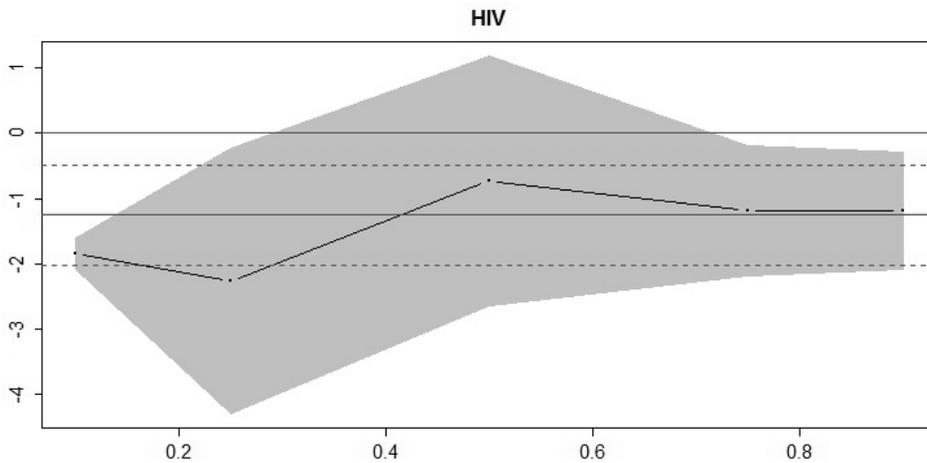
[Figure 3] 흡연율(Smk), 비만인구 비율(Obs), 1인당 주류소비량(Alc), CO2 배출량(CO2) 변수의 계수 추정치 그래프



※ 흡연이나 비만, 음주, 대기오염과 같은 비경제적 요인들은 전반적으로 기대수명에 부정적인 영향을 주며, 흡연율과 비만인구 비율은 기대수명이 높은 국가에서, 1인당 주류소비량과 CO2 배출량은 기대수명이 낮은 국가에서 상대적으로 더 큰 영향을 주는 것으로 나타났다.

대기오염 정도를 반영하는 CO2 배출량(CO2)¹⁰⁾의 계수추정치 변화는 [10% : -0.32481 → 25% : -0.25659 → 50% : -0.09480 → 75% : -0.13557 → 90% : -0.28282]로, CO2 배출량의 증가 역시 모든 분위수에서 인간의 기대수명을 감소시키며, [Figure 3]의 계수추정치 변화 그래프에도 잘 나타나 있듯이 기대수명이 낮은 국가에서 상대적으로 기대수명 감소에 더 큰 영향을 미침을 알 수 있다.

[Figure 4] HIV 출현율(HIV) 변수의 계수추정치 그래프



※ HIV 출현율 증가는 전체적으로 기대수명을 감소시키는 영향을 주며, 기대수명이 낮은 저개발 국가들에서 더 큰 부(-)의 영향을 주는 것으로 나타났다.

후천성 면역결핍 증후군(AIDS)을 일으키는 원인 바이러스인 HIV(human immunodeficiency virus) 출현율 변수의 계수 추정치는 [10% : -1.84572 → 25% : -2.26777 → 50% : -0.73215 → 75% : -1.18577 → 90% : -1.18649]로, HIV 출현율의 증가는 모든 분위수에서 기대수명을 감소시키며, [Figure 4]에 잘 나타나 있듯이, 기대수명이 낮은 국가에서 더 큰 부(-)의 영향을 주는 것으로 나타났다. 이는 일반적으로 HIV 감염자가 기대수명이 높은 선진국보다 기대수명

10) 대기 중에는 약 0.03%의 이산화탄소(carbon dioxide)가 존재하며, 이중 대부분은 가열된 해양이나 화산폭발에 의해 방출되지만, 본 연구에 사용된 자료는 화석 연료(fossil fuels)의 연소나 시멘트 제조, 화력발전, 연료 소비 등 산업화에 의해 국가적으로 발생하는 이산화탄소의 총량을 측정하는 것이다. 지협적으로 발생한 0.1%~1%의 이산화탄소에 노출되면 나른함과 두통을 느끼며, 8%~10%의 이산화탄소에 노출되면 질식에 의한 사망에도 이를 수 있다. 2009년 말 미국 환경보호청(EPA)에서도 이산화탄소(CO2)가 건강을 위협한다며 규제의 필요성을 역설한 바 있다. [출처 : WDI, Considine, G. D. et al.(2005)]

이 낮은 후진국에 더 많으며¹¹⁾, 의료기술이 발달한 선진국에서는 약물로 오랜 기간 동안 생명유지가 가능하지만, 상대적으로 후진국에서는 HIV 감염자가 의료 혜택을 받기 어렵기 때문으로 해석된다.

〈Table 1〉 종속변수(Life)와 독립변수의 기초 통계량

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Life	50.54	68.98	73.73	72.50	78.81	82.93
White	0.0000	0.0000	1.0000	0.5167	1.0000	1.0000
Yellow	0.0000	0.0000	0.0000	0.3333	1.0000	1.0000
GDP_pc	343.4	2414.2	7635.2	16274.6	24546.3	78457.4
Snt	9.50	70.58	95.35	80.93	100.00	100.00
S.Pre	2.818	50.134	77.580	69.888	96.150	126.301
S.Pri	62.01	98.92	103.07	103.04	108.67	131.81
S.Sec	11.79	75.83	91.79	85.71	101.60	129.22
S.Ter	1.422	27.931	51.152	46.878	63.988	91.593
Smk	4.00	17.50	26.00	24.07	30.25	46.00
Obs	1.10	13.38	21.80	19.53	26.30	34.70
Alc	0.320	4.915	8.755	9.057	12.635	23.010
CO2	0.04516	1.45388	4.85957	5.04705	7.46103	18.3754
HIV	0.10	0.10	0.30	0.64	0.80	6.20

※ Min., 1st Qu. Median, Mean, 3rd Qu. Max.는 각각 최솟값, 1사분위수, 중앙값, 평균, 3사분위수, 최댓값을 나타낸다.

〈Table 2〉 2009년 60개 국가에 대한 횡단면자료의 분위수회귀 분석 결과

τ	0.1	0.25	0.5	0.75	0.9
(Intercept)	45.1686 (0.0000) ***	50.0439 (0.0000) ***	51.7255 (0.0000) ***	56.0747 (0.0000) ***	57.6378 (0.0000) ***

11) 본 연구에 사용된 60개 국가의 횡단면자료에서 기대수명 하위 30%(70세 이하)에 해당하는 국가들의 1인당 실질 GDP 평균은 약 1,549.04\$이며, 상위 30%(78세 이상)에 해당하는 국가들의 평균은 약 39,880.43\$이다. 또한 동일한 기준에서 HIV 출현율은 기대수명 하위 30% 국가에서 평균 1.31%, 기대수명 상위 30% 국가에서 평균 0.3%로 나타났다.

τ	0.1	0.25	0.5	0.75	0.9
White	6.4528 (0.0000) ***	4.7952 (0.0885) ·	4.4070 (0.0986) ·	5.9038 (0.0640) ·	3.1693 (0.3948)
Yellow	5.5304 (0.0000) ***	4.2735 (0.1210)	4.8972 (0.0522) ·	5.6124 (0.0387) *	4.5209 (0.0851) ·
GDP.pc	0.0001 (0.0000) ***	0.0001 (0.0001) ***	0.0001 (0.0000) ***	0.0001 (0.0105) *	0.0001 (0.0018) **
Snt	0.1181 (0.0000) ***	0.1139 (0.0049) **	0.0893 (0.0537) ·	0.0845 (0.0688) ·	0.1130 (0.0468) *
S.Pre	0.0854 (0.0000) ***	0.0773 (0.0005) ***	0.0816 (0.0003) ***	0.0428 (0.0198) *	0.0486 (0.0355) *
S.Pri	0.0654 (0.0000) ***	0.0416 (0.3369)	0.0245 (0.6234)	0.0128 (0.7757)	-0.0092 (0.9153)
S.Sec	-0.0030 (0.8790)	-0.0036 (0.9310)	-0.0025 (0.9570)	0.0489 (0.2583)	0.0447 (0.4649)
S.Ter	-0.0122 (0.0695) ·	-0.0145 (0.7153)	0.0349 (0.2930)	0.0078 (0.8407)	0.0440 (0.3822)
Smk	-0.0345 (0.0443) *	-0.0531 (0.2751)	-0.0236 (0.6662)	-0.1381 (0.0469) *	-0.1460 (0.0181) *
Obs	0.1241 (0.0000) ***	0.1158 (0.0212) *	0.0842 (0.2649)	0.0614 (0.3863)	0.0015 (0.9891)
Alc	-0.2546 (0.0000) ***	-0.1822 (0.0249) *	-0.2594 (0.0129) *	-0.0857 (0.3584)	-0.0087 (0.9616)
CO2	-0.3248 (0.0000) ***	-0.2566 (0.0726) ·	-0.0948 (0.5672)	-0.1356 (0.2618)	-0.2828 (0.3488)
HIV	-1.8457 (0.0000) ***	-2.2678 (0.0724) ·	-0.7322 (0.5310)	-1.1858 (0.0564) ·	-1.1865 (0.0341) *

※ '***' : 0.1%, '**' : 1%, '*' : 5%, '·' : 10% 유의수준에서 각각 유의함을 나타내며, 괄호 안의 숫자는 P -value이다.

V. 결 론

본 연구에서는 Koenker and Bassett(1978)이 선형 회귀모형에 소개한 분위수 회귀추정법 (method of quantile regression estimation)을 소개하고, 분위수회귀 분석에 가장 적합한 소프트웨어 중 하나인 R 소프트웨어의 'quantreg'(Koenker, 2013) 패키지를 이용하여 경제외적 요인이 기대수명에 미치는 영향을 설명하는 회귀모형을 2009년 60개 국가의 자료를 이용하

여 추정하고 실증적으로 분석하였다. 따라서 본 연구의 목적은 분위수회귀 분석과 관련된 독창적인 계량기법이나 실증분석을 통해 새로운 현상을 발견하기보다 응용 계량경제학자 및 전문가들에게 분위수회귀 분석에 필요한 기존의 계량기법을 소개하고 R 소프트웨어의 활용성을 높이는 데 있다.

고령화 사회를 대비하여 기대수명에 영향을 미치는 경제적·경제외적 요인에 대한 실증분석을 수행하였으며, 그 결과에 의하면 1인당 실질 GDP나 향상된 위생시설 확충 정도와 같은 경제적인 요인들은 여전히 기대수명 향상에 유의미한 정(+)의 영향을 주지만, 기대수명이 길어짐에 따라 경제적인 요인들의 영향력은 점차 감소하며 오히려 비경제적인 요인들, 예컨대 중·고등 교육이나 흡연, 비만과 같은 요인들이 기대수명 연장 및 축소에 상대적으로 더 큰 영향을 주는 것으로 나타났다. 또한 1인당 주류소비량이나 대기오염(CO₂ 배출량), HIV 출현율과 같은 요인들은 기대수명이 낮은 후진국가에서 기대수명 향상을 크게 저해하는 요인으로 밝혀졌다.

본 연구에서는 가장 일반적인 선형회귀모형에 대한 분위수회귀 추정법과 실증분석에 대해 다루고 있지만 향후 연구에서는 GARCH 모형을 포함한 시계열 모형, 패널 모형, 해저드 모형(hazard model), 다변량 모형 등과 같은 다양한 계량모형에 대한 분위수회귀 추정법의 이론적 소개와 함께 실증분석을 수행하기 위한 프로그래밍 기법에 대해 추가적인 설명이 요구된다.

참고문헌

- 김민중(2013), Productivity Performance and Exporting Activity of Korean Manufacturing Firms: Quantile Regression Approach, 『경제분석』, 제19권 제2호, 97-120.
- 박범조(2002), 정보기술이 기업의 사업성과에 미치는 효과 분석: 위수회귀 접근법, 『계량경제학보』, 13권 2호, 61-93.
- 박범조(2003), 분위수 회귀접근법, 『계량경제학보』, 제14권 제4호, 93-122.
- 박범조(2008), 외환시장의 변동성과 거래량의 관계 분석: 충격정보 확률변동성 모형 이용, 『경제분석』, 제14권 제4호, 47-84.
- 박범조(2012), R을 이용한 패널자료분석 : OECD 국가의 자동차 휘발유 소비량 분석 적용, 『산업연구』, 제16권 제1호, 83-100.
- 박범조(2013), 응용 계량경제학 : R 활용, 시그마프레스.
- Becker, R., Chambers, J. and Wilks, A.(1988), The New S Language, London: Chapman & Hall.
- Buchinsky, M.(1994), Changes in US Wage Structure 1963-87: An Application of Quantile Regression, *Econometrica*, 62, 405-458.
- Chakraborty, B.(2003), On Multivariate Quantile Regression, *Journal of Statistical Planning and Inference*, 110, 109-132.
- Chamberlain, G.(1994), Quantile Regression, Censoring and the Structure of Wages, in *Advances in Econometrics*, C. Sims, ed., New York: Elsevier, 171-209.
- Chaudhuri, P.(1991), Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation, 19(2), *The Annals of Statistics*, 760-777.
- Chernozhukov, V. and Fernández-Val, I.(2011), Inference for Extremal Conditional Quantile Models, with an Application to Market and Birthweight Risks, *Review of Economic Studies*, 78(2), 559-589.
- Chernozhukov, V. and Hansen, C.(2008), Instrumental Variable Quantile Regression: A Robust Inference Approach,” *Journal of Econometrics*, 142(1), 379-398.
- Chernozhukov, V. and Umantsev, L.(2001), Conditional Value-at-Risk: Aspects of Modeling and Estimation, *Empirical Economics*, 26, 271-292.
- Considine, G. D.(2005), Van Nostrand's encyclopedia of chemistry, 5th edition, Wiley

Press.

- Engle, R. and Manganelli, S.,(1999), CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, *Journal of Business and Economic Statistics*, 22, 367-381.
- Galvao, A., Lamarche, C. and Lima, L.(2013), Estimation of Censored Quantile Regression for Panel Data with Fixed Effects. Forthcoming, *Journal of the American Statistical Association*.
- Gradstein, M. and Kaganovich, M.(2004), Aging population and education finance. *Journal of Public Economics*, 88, 2469-2485.
- Katoa, K., Galvao Jr., A. and Montes-Rojas, G.(2012), Asymptotics for Panel Quantile Regression Models with Individual Effects, *Journal of Econometrics*, 170(1), 76-91.
- Jurečková, J. and B. Prochazka(1994), "Regression Quantiles and Trimmed Least Squares Estimators in the Nonlinear Regression Model," *Journal of Nonparametric Statistics*, 3, 201-222.
- Koenker, R.(2005), *Quantile Regression*, Cambridge University Press.
- Koenker, R.(2013), Package 'quantreg', Version 5.04, URL <http://cran.r-project.org/package=quantreg>.
- Koenker, R. and Bassett, G.(1978), Regression Quantiles, *Econometrica*, 46, 33-50.
- Koenker, R. and Biliias, Y.(2001), Quantile Regression for Duration Data: A Reappraisal of the Pennsylvania Reemployment Bonus Experiments, *Empirical Economics*, 26, 199-220.
- Koenker, R. and Park, B. J.(1996), An Interior Point Algorithm for Nonlinear Quantile Regression, *Journal of Econometrics*, 71, 265-283.
- Koenker, R. and Xiao, Z.(2002), Quantile Autoregression, Unit Roots and Asymmetric Interest Rate Dynamics, *UIUC*, Manuscript.
- Koenker, R. W. and Zhao, Q.(1996), "Conditional Quantile Estimation and Inference for ARCH Models," *Econometric Theory*, 12, 793-813.
- Lee, T. H. and Saltoğlu, B.(2002), Assessing the Risk Forecasts for Japanese Stock Market, *Japan and the World Economy*, 14, 63-85.
- Levin, J.(2001), For Whom the Reductions Count: Quantile Regression for Duration Data: A Quantile Regression Analysis of Class Size on Scholastic Achievement, *Empirical Economics*, 26, 221-246.

- Li, M. and Yen, S.(2011), Re-examining Covariance Risk Dynamics in International Stock Markets Using Quantile Regression Analysis, *Acta Oeconomica*, 61(1), 33-59.
- Mata, J. and Machado, J.(1996), Firm start-up size: A Conditional Quantile Approach, *European Economic Review*, 40, 1305-1323.
- Nahm, J. W.(2001), Nonparametric Quantile Regression Analysis of R&D-Sales Relationship for Korean Firms, *Empirical Economics*, 26, 259-270.
- Park, B. J.(1997), The Effect of Education on the Duration of Unemployment, *Journal of Economic Theory and Econometrics*, 3(1), 139-162.
- Powell, J. L.(1986), Censored Regression Quantiles, *Journal of Econometrics*, 32, 143-155.
- R Core Team(2012), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Shaw, J., Horrace, W. and Vogel, R.(2005), The Determinants of Life Expectancy: An Analysis of the OECD Health Data, *Southern Economic Journal*, 71(4), 768-783.
- Taylor, J.(1999), A Quantile Regression Approach to Estimating the Distribution of Mutiperiod Returns, *Journal of Derivatives*, 7(1), 64-78.
- Yin, G. and Cai, J.(2005), Quantile Regression Models with Multivariate Failure Time Data, *Biometrics*, 61(1), 151-161.

Quantile Regression Approach Using R Statistical Software: Analyzing the Effect of Non-economic Factors on Life Expectancy at Birth

Kim, Myung-Joong* · Park, Beum-Jo**

ABSTRACT

This paper concisely explains quantile regression models, introduced by Koenker and Bassett(1978). The quantile regression approach has recently received a lot of attention because it contains robustness that has not responded sensitively to influences from outliers or error term distribution and offers complete view of the effect of covariates on the location, scale, and shape of the conditional distribution of the response variable.

Due to these advantages, studies on quantile regression theories and estimation algorithm are progressing actively in economics. Thus, the main purpose of this paper is to introduce 'quantreg'(Koenker, 2013) package in R software that should be one of the appropriate statistical packages for the quantile regression analysis. This paper also applies it to an empirical study on the effect of non-economic factors on life expectancy at birth. This application might help experts in empirical economics to improve their understanding of the quantile regression analysis.

Key Words : Quantile regression analysis, R software, 'quantreg' package, Asymptotic properties of quantile regression estimator, Non-economic factors and life expectancy at birth.

* Graduate Student, Dept. of Economics, Dankook University, Jukjeon-dong, Suji-gu, Yongin-si, Gyeonggi-do, 448-701, Korea. E-mail: mjkim@dankook.ac.kr

** Professor, Dept. of Economics, Dankook University, Jukjeon-dong, Suji-gu, Yongin-si, Gyeonggi-do, 448-701, Korea. E-mail: bjpark@dankook.ac.kr