確率的 支配 關係 檢定法을 利用한 經濟分析[1]

朴 炳 建 · 黃 潤 宰

확률적 지배 관계란, 어떻게 두 확률 변수를 비교하고 위계를 정할 수 있는지에 대하여 기준을 제시해주는 이론이다. 초기 확률적 지배 관계의 개념은 이론적인 측면에서의 접근이 이루어졌으나, 확률적 지배 관계에 대한 계량 경제학적 분석은 McFadden(1989)의 연구 이후 최근까지 많은 연구들이 이루어지고 있다. 확률 분포함수와 효용함수에 대한 가정은 확률 변수들을 비교하는 데에 있어서 출발점이 된다. 확률적 지배 관계는 분포 함수와 효용함수에 대해 매우 약한 가정만을 필요로 한다는 특징을 가진다. 예를 들어 후생 경제학에서는 불평등과 빈곤에 대한 연구에서 확률적 지배 관계를 이용하여 소득의 분포를 비교한다. 또한, 재무 경제학에서는 다양한 투자 대상의 불확실성을 비교함에 있어서 확률적 지배 관계가 중요한 개념으로 사용된다. 본 논문의 목적은 확률적 지배 관계의 검정법에 대한 최근의 계량경제학의 문헌을 소개하고 경제학에의 적용 사례를 통해 검정법의 유용성을 살펴보는 데 있다. 구체적으로 확률적 지배 관계의 정의와 이론적 측면을 살펴보고, 확률적 지배 관계에 대한 유한 차원과 무한 차원의 계량적 검정 방법에 대하여 설명한다. 또한, 확률적 지배 관계를 응용한 다양한 개념과 계량적 검정 방법들을 소개하며 확률적 지배 관계를 이용한 실증적 분석의 적용 사례를 제시한다.

1. 序 論

실증적 분석을 중요시하는 경향은 최근 경제학 제 분야에서 일어나고 있는 현상이다. 이론적 접근으로 분석되었던 많은 대상들에 대한 자료의 축적과 컴퓨터의 발전으로 인한 연산 속도의 발전으로 이론과 실증의 統合的 完結性에 대한 추구는 경제학의 흐름이 되고 있다. 본 논문에서 다룰 확률적 지배 관계(stochastic dominance)에 대한 통계적 검정 방법 또한 이러한 큰 흐름의 하나로 볼 수 있다.

확률적 지배 관계란, 어떻게 두 확률 변수를 비교하고 위계를 정할 수 있는지에 대하여 기준을 제시해주는 이론이다. 초기 확률적 지배 관계의 개념은 이론적인 측면에서의 접근이 이루어졌다. Lehmann(1955)이 최초로 정의를 하였고, Rothschild and Stiglitz(1970) 등이 확률적 지배 관계의 개념을 경제학에 도입하는 데 큰 기여를 한 대표적 연구들이다.

(1) 본 연구는 제원연구재단의 지원으로 이루어진 것임.

확률적 지배 관계에 대한 계량 경제학적 분석 방법이 제시된 것은 그 이후의 일이다. McFadden(1989)에서 간단한 상황에서의 검정 방법이 개발된 이후 최근까지 많은 연구들이 지속적으로 이루어지고 있다.

확률 분포 함수와 효용함수에 대한 가정은 확률 변수들을 비교하는 데에 있어서 출발점이 된다. 예컨대, 두 확률 변수가 모두 정규 분포를 따르고 분산이 동일하다면 기대값을 비교함으로써 두 확률 변수의 비교는 충분하다. 혹은 경제적 주체는 위험 중립적이라는 가정을 한다면 역시 확률 변수의 기대값만으로 확률 변수의 우열이 결정된다. 그러나 이러한 경우 가정이 매우 한정적이므로 우리가 분석하고 싶은 많은 경우에 적용할 수 없다. 하지만, 확률적 지배 관계는 분포 함수와 효용함수에 대해 매우 약한 가정만을 필요로 한다. 이러한 장점 덕분에 최근 경제학의 많은 분야에서 널리 쓰인다. 후생 경제학에서는 불평등과 빈곤에 대한 연구에서 확률적 지배 관계를 이용하여 소득의 분포를 비교한다. 또한, 재무 경제학에서는 다양한 투자 대상의 불확실성을 비교함에 있어서 확률적 지배 관계가 중요한 개념으로 사용된다. 본 논문의 목적은 확률적 지배 관계의 검정법에 대한 최근의 계량경제학의 문헌을 소개하고 경제학에의 적용 사례를 통해 검정법의 유용성을 살펴보는 데 있다.

본 논문은 다음과 같은 구조로 이루어져 있다. 2장에서는 확률적 지배 관계의 정의와 이론적 측면을 다룰 것이다. 3장에서는 확률적 지배 관계에 대한 유한 차원과 무한 차원의 계량적 검정 방법에 대하여 설명할 것이다. 이를 바탕으로 4장에서는 확률적 지배 관계를 응용한 다양한 개념과 계량적 검정 방법들을 소개하고자 한다. 마지막으로 5장에서는 확률적 지배 관계를 이용한 실증적 분석의 적용 사례를 제시하며, 6장의 결론으로 본 논문 마무리할 것이다.

2. 確率的 支配 關係의 定義 및 概念

확률적 지배 관계(stochastic dominance)는 확률 변수의 누적 분포 함수(cumulative distribution function) 혹은 그것의 적분으로 표현되는 함수를 통해 정의된다. 비교의 대상이 될 두 개의 집단 1과 2가 있다고 하자. 집단 1, 2로부터 추출된 확률 변수를 각각 X_1과 X_2라고 하자. 두 확률 변수는 실수 구간 $[0, \overline{X}]$ 사이의 값을 갖고, 확률 변수의 누적 분포 함수는 각각 $F_k(x), k = 1, 2$를 따른다고 가정한다.

이를 이용하여 1차 확률적 지배 관계는 다음과 같이 정의된다.

(定義: 1차 확률적 지배 관계) 다음 세 명제는 동치이다.

(1) X_1이 X_2를 1차 확률적으로 지배한다.

(2) $[0, \overline{X}]$의 모든 점 x에 대하여 다음 부등식이 만족되고, 일부의 점에 대하여 강부등식이 성립된다.

$$F_1(x) \leq F_2(x)$$

(3) 임의의 단조 증가하는 함수 $u: [0, \overline{X}] \rightarrow \mathbb{R}$에 대하여 다음 부등식이 만족되고, 일부 단조증가 함수에 대하여 강부등식이 성립된다.

$$E[u(X_1)] \geq E[u(X_2)]$$

1차 확률적 지배 관계는 X_1의 실현된 값이 임의의 점 x보다 작을 확률이 X_2의 그것보다 작은 경우에 성립된다. 그리고 이것의 동치로서 세 번째 명제는 효용함수가 단조 증가한다는 가정하에서 期待效用(expected utility)이 한쪽이 크다는 경제학적 의미를 지닌다. 다시 말하여 합리적인 경제 주체가 단조 증가하는 효용함수를 가지고 있고, 기대효용을 극대화하는 선택을 한다면, 1차 확률적으로 지배당하는 선택을 하지 않을 것이라는 의미이다. 또한, 이러한 접근 방법은 효용함수가 단조 증가한다는 매우 약한 가정만 필요하다는 점에서 매우 강인한(robust) 판단 기준이 된다.

비슷한 방식으로 2차 확률적 지배 관계는 다음과 같이 정의한다.

(定義: 2차 확률적 지배 관계) 다음 세 명제는 동치이다.

(1) X_1이 X_2를 2차 확률적으로 지배한다.

(2) $[0, \overline{X}]$의 모든 점 x에 대하여 다음 부등식이 만족되고, 일부의 점에 대하여 강부등식이 성립된다.

$$\int_{-\infty}^{x} F_1(t)dt \leq \int_{-\infty}^{x} F_2(t)dt$$

(3) 임의의 단조 증가하고 오목한(concave) 함수 $u: [0, \overline{X}] \rightarrow \mathbb{R}$에 대하여 다음 부등식이 만족되고, 일부 단조증가 오목 함수에 대하여 강부등식이 성립된다.

$$E[u(X_1)] \geq E[u(X_2)]$$

두 번째 명제에서 보듯이 2차 확률적 지배 관계의 정의는 누적 분포 함수의 積分을 이용하여 정의된다. 두 번째 명제로부터 직관적인 의미를 찾기는 힘들지만, 그와 동치인 세 번째 명제는 경제학적으로 매우 중요한 의미를 담고 있다. 단조 증가하고 오목한 효용함수는 경제학에서 von Neumann-Morgenstern 효용함수로 불리는 가장 널리 쓰이는 형태의 효용함수이다. 2차 확률적 지배 관계는 임의의 von Neumann-Morgenstern 효용함수에 대한 기대효용의 대소 관계를 의미하며, 더 나아가 필요충분조건을 이루어 이론적으로 중요한 개념이다. 또한, von Neumann-Morgenstern 효용함수의 가정과 기대효용 가설 하에서의 경제적인 주체의 선택을 분석할 때 2차 확률적 지배 관계를 이용하여 대상의 분포 함수만으로 결론을 도출하는 것이 가능해지므로 실증적으로도 중요하다.

보다 일반적인 s차 확률적 지배 관계도 비슷한 방식으로 정의된다. 표기의 편의를 위하여 다음과 같이 s차 누적 분포 함수를 정의하자.

(定義) $k = 1, 2$에 대하여 확률 변수 X_k의 s차 누적 분포 함수 $D_k^s(x)$는 다음과 같이 정의된다.

$$D_k^s(x) = F_k(x), \ s = 1 \text{인 경우}$$

$$D_k^s(x) = \int_0^x D_k^{s-1}(t)dt, \ s \geq 2 \text{인 경우}$$

이를 이용하여 일반적인 s차 확률적 지배 관계를 정의하면 다음과 같다.

(定義: **s차 확률적 지배 관계**) $[0, \overline{X}]$의 모든 점 x에 대하여 다음 부등식이 만족되고, 일부의 점에 대하여 강부등식이 성립될 때, X_1이 X_2를 s차 확률적으로 지배한다고 정의한다.

$$D_1^s(x) \leq D_2^s(x)$$

s차 확률적 지배 관계는 s가 증가할수록 보다 많은 제약을 가진 효용함수에 대한 기대효용의 대소관계로 해석될 수 있다.

3. 確率的 支配 關係의 統計的 檢定 方法

본 장에서 앞으로 다룰 내용은 각 집단으로부터의 자료 $\{X_{ki} : i = 1, \ldots, N\}$이 관측되었을 때, 이를 바탕으로 두 집단 사이에 확률적 지배 관계가 존재하는지 통계적으로 검정하는 문제를 다룰 것이다. 그러나 확률적 지배 관계에 대한 정의는 누적 분포 함수 간의 부등식의 형태로 이루어져 있기 때문에 모든 점에서의 비교를 토대로 한다. 이로 인하여 일반적으로 함수 간의 등식으로 이루어진 가설을 검정하는 것보다 문제가 복잡하다. 이 문제를 해결하기 위한 연구는 크게 두 갈래로 구분할 수 있다. 우선 대부분의 연구들은 제한된 대립가설에서만 적용되는, 실질적인 검정력이 유한 차원인 검정 방법들에 관련된 것이다. Anderson(1996), Davidson and Duclos(2000) 등이 이러한 계열의 대표적인 연구이다. 또 다른 방식은 Kolmogorov-Smirnov type 또는 Cramer-von Mises type 통계량을 이용하여 일반적인 대립가설에 대하여 검정을 행하는 방식이 있다. 이러한 연구로는 McFadden(1989), Klecan, McFadden, and McFadden(1991), Barrett and Donald(2003), Linton, Maasoumi, and Whang(2005) 등이 있다.

앞으로 소개하는 확률적 지배 관계의 검정 방법에 대하여 특별한 언급이 없다면 귀무가설은 다음과 같이 설정한다.

$$H_0\text{: } X_1\text{는 } X_2\text{를 } s\text{차 확률적으로 지배한다.}$$

그리고 대부분의 검정 방법은 미지의 누적 분포 함수에 대한 추정을 필요로 한다. 이에 대하여 經驗的 分布 函數(empirical distribution function)을 바탕으로 한 누적 분포 함수의 표본 상대값(sample analogue)은 계산이 쉬우며 좋은 성질을 지니는 추정치이기에 앞으로 많이 사용될 것이다. 표기 편의를 위해 여기서 정의하려 한다.

$$\hat{D}_k^s(x) = \frac{1}{N}\sum_{i=1}^{N} 1(X_{ki} \le x), \quad s = 1\text{인 경우}$$

$$\hat{D}_k^s(x) = \int_0^x \hat{D}_k^{s-1}(t)dt, \quad s = 2, 3, \ldots\text{인 경우}$$

참고로 부분 적분을 이용하면 다음과 같이 쉽게 계산이 가능하다.

$$\hat{D}_k^s(x) = \frac{1}{N(s-1)!}\sum_{i=1}^{N}(x - X_{ki})^{s-1}1(X_{ki} \le x)$$

또한, 각 연구마다 가정하는 바가 상이하기 때문에 본 장에서는 논의의 편의를 위하여 다음 가정을 유지할 것이다. 그리고 4장에서 가정이 일반화된 경우에 대하여 논할 것이다.

(가정 1) $F_1(\cdot)$와 $F_2(\cdot)$는 동일한 정의역 $[0, \overline{X}]$을 가지며 그 위 모든 점에서 연속이다.
(가정 2) $k = 1, 2$에 대하여 $F_k(\cdot)$는 르벡 측도(Lebesgue measure)에 대한 확률 밀도 함수(probability density function)가 존재한다.
(가정 3) $k = 1, 2$에 대하여 $\{X_{ki} : i = 1, \ldots, N\}$는 독립적으로 누적 분포 함수 $F_k(x)$에서 추출된 표본이다.
(가정 4) X_1과 X_2는 서로 확률적으로 독립이다.

3.1. 有限 次元의 檢定 方法

Anderson(1996)과 Davidson and Duclos(2000)의 검정 방법은 임의로 정해진 유한한 점에서 대상을 비교하여 확률적 지배 관계에 대한 검정을 한다. Anderson(1996)은 임의의 몇 개의 점에서의 누적 분포 함수를 계산하여 이를 바탕으로 多項 檢定(multinomial test)를 시행한다. 반면 Davidson and Duclos(2000)는 누적 확률 분포의 교차하는 점을 추정하여 그것을 검정 통계량으로 삼는다. 이러한 방법은 유한 차원의 정보만을 사용하기 때문에 귀무가설 하에서의 분포를 구하기 쉬워 매우 실용적이라는 장점이 있다. 하지만 임의로 정해진 점을 사용하기 때문에 검정의 非一貫性(inconsistency)이 야기될 수도 있다.

3.1.1. Anderson 檢定 方法

Anderson의 방법은 Pearson의 분포 가정에 대한 適合度(goodness of fit) 검정에 기반을 두고 있다. $[0, \overline{X}]$에 임의의 분할(partition) $0 = t_0 < t_1 < \ldots < t_M = \overline{X}$이 주어졌다고 하자. 그러면 정의역은 M개의 구간 $\{(t_{m-1}, t_m] : m = 1, \ldots, M\}$으로 나뉘는데, $\{X_{ki} : i = 1, \ldots, N\}$에서 각 구간에 속한 개수를 $N_k^1, N_k^2, \ldots, N_k^M$이라 정의하자. 그러면 이것은 $p_{km} = \int_{t_{m-1}}^{t_m} dF_k$에 대하여 $P_k = (p_k^1, \ldots, p_k^M)'$의 모수(parameter)를 가지는 多項 分布(multinomial distribution)을 따른다.

P_k를 표본 평균 $\hat{P}_k = 1/N\,(N_k^1, N_k^2, \ldots, N_k^M)$으로 추정하고 이를 이용하여 Pearson의 다항 검정의 원리를 적용하면 쉽게 다음 사실을 알 수 있다.

$$\sqrt{N}(\hat{P}_1 - \hat{P}_2) \xrightarrow{d} N(P_1 - P_2,\ \Omega_1 + \Omega_2)$$

$$\text{단, } \Omega_k = \begin{pmatrix} p_k^1(1-p_k^1) & -p_k^1 p_k^2 & \cdots & -p_k^1 p_k^M \\ -p_k^2 p_k^1 & p_k^2(1-p_k^2) & & -p_k^2 p_k^M \\ \vdots & & \ddots & \vdots \\ -p_k^M p_k^1 & -p_k^M p_k^2 & \cdots & p_k^M(1-p_k^M) \end{pmatrix}$$

우선 1차 확률적 지배 관계 검정을 위해 다음 행렬을 생각해보자.

$$I_1 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & & 0 \\ \vdots & & \ddots & \\ 1 & 1 & & 1 \end{pmatrix}$$

$I_1 P_1 = \left(P_k^1,\ P_k^1 + P_k^2,\ \ldots,\ \sum_{m=1}^{M} p_k^m \right)'$이므로 다시 표현하면 $(F_k(t_1), F_k(t_2), \ldots, F_k(t_M))'$으로 쓸 수 있다. 귀무가설 하에서는 모든 t_m에 대하여 $F_1(t_m) - F_2(t_m) \le 0$이므로 앞서 정의한 행렬을 이용하여 $I_1(P_1 - P_2) \le 0$(2)와 같이 쓸 수 있다. 그러면 검정 통계량으로 $T^1 = \sqrt{N}\, I_1(\hat{P}_1 - \hat{P}_2)$을 생각해 볼 수 있다. 이 통계량의 귀무가설 하에서 극한 분포는 다음과 같다.

$$T^1 \xrightarrow{d} N(0,\ I_1(\Omega_1 + \Omega_2)I_1')$$

여기서 대립가설을 $I_1(P_1 - P_2) > 0$로 설정하고 쉽게 정규 분포를 이용한 검정을 할 수 있다. 그러나 여기서 주의할 점은 $I_1(P_1 - P_2)$의 부호가 판정되지 않는 대립가설에 대해서는 검정력(power)을 상실한다는 점이다. 예를 들면 두 누적 분포 함수가 교차하는 경우에는 $I_1(P_1 - P_2)$의 일부 원소는 양수이고 다른 원소는 음수가 되어 대립가설에 속하지 않는다.

고차 확률적 지배 관계 검정을 위해서는 새로운 행렬을 정의해야 한다.

---

(2) 임의의 벡터 v에 대해 $v \ge 0$는 v의 모든 원소가 0보다 크다는 것을 의미한다.

$$I_2 = \begin{pmatrix} t_1 - t_0 & 0 & 0 & \cdots & 0 \\ t_2 - t_0 & t_2 - t_1 & 0 & \cdots & 0 \\ t_2 - t_0 & t_3 - t_1 & t_3 - t_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_2 - t_0 & t_3 - t_1 & t_4 - t_2 & \cdots & t_M - t_{M-1} \end{pmatrix}$$

그러면 s차 확률적 지배 관계 검정의 귀무가설은 $I_2^{s-1} I_1 (P_1 - P_2) \le 0$으로 표현되므로 앞서와 비슷한 방식으로 쉽게 검정을 할 수 있다. 물론 대립가설이 제한적이라는 문제도 동일하다.

3.1.2. Davidson-Duclos 檢定 方法

Davidson-Duclos 검정 방법은 두 집단의 분포 함수가 交叉하는 지점을 추정하고 그것을 바탕으로 검정하는 것을 핵심으로 한다. 이제 z_s라는 점을 다음과 같이 정의하자. 만일 $x \in (0, z_s)$인 모든 x에 대해서는 $D_1^s(x) > D_2^s(x)$이 성립되고 z_s에 대해서는 $D_1^s(z_s) = D_2^s(z_s)$를 만족하는 점이 존재한다면 그 점으로 z_s를 정의하자. 만일 모든 점에서 $D_1^s(x) > D_2^s(x)$인 경우에는 $z_s = \overline{X}$로 정의하자. 그리고 모든 점에서 $D_1^s(x) < D_2^s(x)$라면 $z_s = 0$이라고 하자.

이번에는 귀무가설을 확률적 지배 관계가 존재하지 않는다는 것으로 생각하자. 즉 두 집단의 누적 분포 함수가 적어도 한 점 이상 교차하므로 z_s가 0 또는 $\overline{X}$가 아닌 곳에서 정의되는 경우이다.

경험적 분포 함수를 이용하여 $D_k^s(x)$에 대한 추정치 $\hat{D}_k^s(x)$를 구하고 z를 다음과 같이 추정하자.

$$\hat{z} = \inf_{x \in [0, \overline{X}]} \{x : \hat{D}_1^s(x) = \hat{D}_2^s(x)\}$$

이것을 바로 검정 통계량으로 사용할 수 있는데, 검정 통계량의 극한 분포는 다음을 따른다.

$$\sqrt{N}(\hat{z} - z) \xrightarrow{d} N(0, \ \Omega)$$

그리고 분산은 다음과 같이 정의된다.

$$\Omega = \frac{Var[(z - X_{1i})^{s-1}1(z \geq X_{1i}) - (z - X_{2i})^{s-1}1(z \geq X_{2i})]}{(s-1)!(D_1^{s-1}(z) - D_2^{s-1}(z))^2}$$

이 검정 방법은 검정 통계량의 계산과 기각역의 계산이 간편하여 실용적이라는 장점이 있다. 하지만, 표본으로 추정한 누적 분포 함수가 교차하지 않아 $\hat{z}_s$가 0 또는 $\overline{X}$일 때와 실제 누적 분포 함수가 일정 구간에서 일치하거나 여러 번 교차하는 등 매우 가까울 때 검정이 잘 작동하지 않는다는 문제점이 있다.

3.2. Kolmogorov-Smirnov type 檢定 統計量

McFadden(1989)에서는 상한(supremum)을 이용한 확률적 지배 관계 검정이 Kolmogorov-Smirnov type(이하 KS type) 검정 통계량의 성질을 지닌다는 것을 알아냈다. 그 이후 Klecan *et al.*(1991), Barrett and Donald(2003), Linton *et al.*(2005)를 비롯한 많은 연구들이 KS type 검정 통계량을 기반으로 하고 있다. 이러한 검정 통계량은 모든 점에서 분포를 비교하는 것이기 때문에, 모든 대립가설에 대하여 一致性(consistency)을 지니는 검정이 가능하다. 즉, 대립가설은 집단 1이 집단 2를 확률적으로 지배하지 않는 모든 경우를 포함한다.

일반적인 KS type 검정 통계량은 다음과 같은 汎函數(functional)에 대한 검정을 기반으로 한다.

$$d = \sup_{x \in [0, \overline{X}]} [D_1^s(x) - D_2^s(x)]$$

주어진 범함수를 이용하면 귀무가설은 $d \leq 0$으로 표현될 수 있다. 그리고 대립가설은 모두 $d > 0$인 경우에 포함된다. 앞서 정의한 경험적 누적 분포 함수 $\hat{D}_k^s(x)$를 이용하여 다음과 같은 검정 통계량을 생각할 수 있다.

$$T = \sqrt{N} \sup_{x \in [0, \overline{X}]} [\hat{D}_1^s(x) - \hat{D}_2^s(x)]$$

이를 이용하여 적절한 상수 c에 대하여 $T \geq c$인 경우 귀무가설을 기각하는 검정을 구성할 수 있다. 그런데 KS type의 통계량은 그 대표본 극한(asymptotic) 분포가 일반적으로 잘 알려진 분포를 따르지 않거나 복잡한 경우가 많아 기각역을 구하는 것이 매우 힘들다. 그래서 주로 사용되는 방법은 bootstrap 방법이다. 이 방법은 근래 급속도로 발전하는 컴퓨터의 계산 능력을 이용하여 매우 간단한 방법으로 기각역을 구할 수 있게 하였다.

또한, 표본의 개수가 적은 경우에는 오히려 대표본 극한 분포보다 정확한 검정 規模(size)를 보이는 것으로 알려져 있다. 여기서는 Barrett and Donald(2003)에서 제시한 표준적인 bootstrap 방법과 Linton *et al.*(2005)에서 제시한 부분 추출 방법(subsampling)을 소개하려 한다.

Barrett and Donald(2003)는 방법을 다음과 같이 제시하였다. 우선 집단 k에 대하여 관측된 자료 $\{X_{ki} : i = 1, \ldots, N\}$로부터 독립적인 복원 추출을 통해 새로운 표본 $\{X_{ki}^{*}\}$을 뽑는다. 이를 이용하여 앞에서와 같은 방법으로 $\hat{D}_k^{s*}$를 구할 수 있다. 그리고 다음과 같은 값을 생각하자.

$$T^{*} = \sqrt{N}\, \sup_{x \in [0, \overline{X}]} [(\hat{D}_1^{s*}(x) - \hat{D}_2^{s*}(x)) - (\hat{D}_1^{s}(x) - \hat{D}_2^{s}(x))]$$

우리는 귀무가설 하에서의 분포를 구하는 것이므로 위와 같은 再調整(recentering) 과정이 필요하다. 만일 이러한 과정을 B번 반복한다면, 총 B개의 새로 추출된 검정 통계량 $\{T_b^{*} : b = 1, \ldots, B\}$을 얻는다. B가 충분히 크다면 재표본(resample)을 바탕으로 계산된 검정 통계량의 분포는 T의 귀무가설 하에서의 극한 분포에 수렴할 것이다. 크기가 α인 검정의 기각역은 다음과 같다.

$$c^{*}(\alpha) = \inf\left\{w : \frac{1}{B}\sum_{b=1}^{B} 1(T_b^{*} \le w) \le 1 - \alpha\right\}$$

한편, Linton, Maasoumi, and Whang(2005)에서는 부분 추출 방법을 이용한 기각역 설정 방법을 제시하였다. 우선 집단 k에 대한 표본 $\{X_{ki} : i = 1, \ldots, N\}$에 대하여 정해진 부분 표본의 크기 J만큼의 부분 표본 $\{X_{ki} : i = j, \ldots, j + J - 1\}$을 $j = 1, \ldots, N - J + 1$에 대하여 추출한다. 즉 $N - J + 1$개의 부분 표본을 재추출한다. j번째 표본을 이용하여 계산한 검정 통계량을 T_j^{**}라고 하자. 위와 마찬가지로 크기 α의 기각역은

$$c^{**}(\alpha) = \inf\left\{w : \frac{1}{N - J + 1}\sum_{j=1}^{N-J+1} 1(T_j^{**} \le w) \le 1 - \alpha\right\}$$

Linton *et al.*(2005)의 부분 추출 방법은 두 집단의 분포 함수가 부분적으로 일치할 때 일반적인 bootstrap 방법보다 높은 검정력을 지닌다. 그러나 이 경우에 부분 표본의 크기 J에 따라 결과가 크게 달라진다. Politis and Romano(1994)는 분산(variance)을 최소화하는 방법을 이용하여 부분 표본의 크기를 결정하는 방법을 제시하였다. 그러나 이 방법은 이

론적인 근거가 충분하지 않은 임시방편의 성격이 강하다. 그래서 Linton, Maasoumi, and Whang(2005)에서는 다양한 부분 표본 크기에 대하여 검정을 행한 다음 모든 결과를 보고하는 방식 등을 제시한다.

4. 確率的 支配 關係 槪念의 擴張 및 應用

본 장에서는 지금까지 알아본 확률적 지배 관계의 개념이 어떻게 확장될 수 있으며 어떻게 적용될 수 있는지 다양한 경우에 대하여 알아보고자 한다.

4.1. 確率的 最大性(Stochastic Maximality)

지금까지의 분석은 비교 대상 집단을 2개로 가정하고 있었다. 만일 비교 대상을 유한한 임의의 K로 확장해도 비슷한 이론이 정립될 수 있다. 우리가 비교하고 싶은 K개의 확률 변수를 $\{X_1, X_2, \ldots, X_k\}$로 나타내자. Klecan, McFadden, and McFadden(1991)에서는 $\{X_1, X_2, \ldots, X_k\}$ 중에서 어떤 확률변수도 다른 확률변수에 대해 s차 확률적 지배 관계를 갖지 않는다면 우리는 그 집합이 s차 확률적 최대성(stochastic maximality)을 지녔다고 정의한다.

확률적 최대성 검정은 예를 들어 다수의 펀드 간의 비교에 적용할 수 있다. 만일 시장이 효율적이라면 확률적으로 지배되는 펀드는 사라질 것이므로 시장에 존재하는 펀드 사이에는 확률적 최대성이 존재할 것이라는 귀무가설을 통계적으로 검정하여 효율 시장 가설이 성립하는지 알아볼 수 있는 것이다. 이에 관련하여 다음 식을 생각해보자.

$$d = \min_{k \neq l} \sup_{x \in [0, \overline{X}]} [D_k^s(x) - D_l^s(x)]$$

확률적 최대성의 귀무가설 하에서는 $d > 0$일 것이다. 이제 검정 통계량을 다음과 같이 설정하면 bootstrap과 부분 추출 방법도 3장에서 제시한 것과 동일하게 적용된다〔Linton, Maasoumi, and Whang(2005)〕.

$$T = \sqrt{N} \min_{k \neq l} \sup_{x \in [0, \overline{X}]} [\hat{D}_k^s(x) - \hat{D}_l^s(x)]$$

4.2. 多次元 確率 變數의 比較

경제학에서, 특히 후생을 비교함에 있어서, 여러 가지 후생 지표를 동시에 비교하는 것이 일반적이다. 예를 들어, 일반적으로 효용함수가 소비와 여가 시간에 의존하는 함수라

고 가정할 때, 두 국가의 평균적인 후생을 비교한다면 일인당 소비와 평균 여가 시간을 동시에 비교해야 할 것이다. 이러한 경우 1차원에서의 확률적 지배 관계와는 다른 정의가 필요하다. 왜냐하면, 1차원 변수에서 확률적 지배 관계는 각각의 限界 分布 函數(marginal distribution function)에 기반을 둔 것이지만 다차원 변수로 확장될 경우 結合 分布 函數(joint distribution function)을 생각해야 하기 때문이다. 이러한 경우에 관하여 McCaig and Yatchew(2007)에서는 확률적 지배 관계의 다차원으로의 확장에 대한 이론을 제시하였다.

간단히 두 개의 집단의 2차원 확률 벡터를 생각해보자. 즉 두 집단 $k = 1, 2$에 대해 집단 k의 두 확률 변수, (X_k^1, X_k^2)를 가지고 있고, 그 종합 분포 함수를 $F_k(x^1, x^2)$라고 하자. 그리고 두 집단은 동일한 효용함수 $U(x^1, x^2)$를 가지고 있다고, 그 효용함수가 다음과 같은 조건을 만족한다고 가정하자.

$$U \in \{U : U_1 \geq 0, U_2 \geq 0, U_{12} \leq 0\}$$

이는 X^1과 X^2가 증가함에 따라 효용은 증가하며, 두 변수 간에는 상호 대체 관계가 있는 효용함수의 형태를 의미한다. 평균적인 후생의 격차는 부분 적분을 이용하여 다음과 같이 정리할 수 있다.

$$\begin{aligned}
&E[U(X_1^1,\ X_1^2) - U(X_2^1,\ X_2^2)] \\
&= \int_0^{\bar{X}^2} \int_0^{\bar{X}^1} U_{12}(x_1,\ x_2)[F_1(x_1,\ x_2) - F_2(x_1,\ x_2)]dx_1 dx \\
&- \int_0^{\bar{X}^1} U_1(x_1,\ \bar{X}_2)[F_1(x_1,\ \bar{X}_2) - F_2(x_1,\ \bar{X}_2)]dx_1 \\
&- \int_0^{\bar{X}^2} U_2(\bar{X}_1,\ x_2)[F_1(\bar{X}_1,\ x_2) - F_2(\bar{X}_1,\ x_2)]dx_2
\end{aligned}$$

단, 여기서 U_i는 $[\partial U(x^1, x^2)]/\partial x_i$, U_{ij}는 $[\partial^2 U(x^1, x^2)]/\partial x_j \partial x_i$를 의미한다. 만일 모든 점에서 $F_1(x^1, x^2) \leq F_2(x^1, x^2)$가 만족되면, 집단 1의 기대효용이 집단 2에 비해 높으므로 집단 1이 집단 2를 1차 확률적으로 지배한다고 할 수 있다.

혹은 효용함수에 대한 가정을 $U \in \{U : U_1 \geq 0, U_2 \geq 0, U_{12} \geq 0\}$으로 바꾸어도 같은 결과를 얻을 수 있다. 이 경우에는 두 변수가 상호 보완적인 관계에 있다. 앞에서와 마찬가지로 모든 점에서 $F_1(x^1, x^2) \leq F_2(x^1, x^2)$가 만족된다면,

$$E[U(X_1^1,\ X_1^2) - U(X_2^1,\ X_2^2)]$$
$$= -\int_0^{\bar{X}^2}\int_0^{\bar{X}^1} U_{12}(x_1,\ x_2)[F_1(x_1,\ x_2) - F_2(x_1,\ x_2)]dx_1dx_2$$
$$-\int_0^{\bar{X}^1} U_1(x_1,\ 0)[F_1(x_1,\ 0) - F_2(x_1,\ 0)]dx_1$$
$$-\int_0^{\bar{X}^2} U_2(0,\ x_2)[F_1(0,\ x_2) - F_2(0,\ x_2)]dx_2 \geq 0$$

즉, 앞에서와 다른 형태의 효용함수에 대해서도 1차 확률적 지배 관계가 동일한 조건 하에서 성립된다.

이제 일반적으로 M차원을 생각해보자. 이 경우 s차 누적 분포 함수는 다음과 같이 정의된다.

$$D_k^s(x^1,\ x^2,\ \ldots,\ x^M) = \int_0^{x^1}\int_0^{x^2}\cdots\int_0^{x^M} D_k^{s-1}(t^1,\ t^2,\ \ldots,\ t^M)dt^M \ldots\ dt^2dt^1$$

그리고 s차 확률적 지배 관계의 정의는 다음과 같다. 임의의 M차원 벡터 X에 대하여 $D_1^s(X) \leq D_2^s(X)$이 만족되면 집단 1이 집단 2에 대하여 s차 확률적으로 지배한다고 한다. 앞에서와 비슷하게 다음 범함수를 생각해보자.

$$d = \sup_x [D_1^s(x) - D_2^s(x)]$$

이를 바탕으로 귀무가설 하에서 $d < 0$을 검정하기 위한 검정 통계량은 다음과 같다.

$$T = \sqrt{N}\ \sup_x [\hat{D}_1^s(x) - \hat{D}_2^s(x)]$$

다차원의 확률 변수 간의 확률적 지배 관계를 생각할 때, 우리는 일부 차원의 변수만의 확률적 지배 관계를 검정해야 하는 경우도 생각해 볼 수 있다. 1차 확률적 지배 관계에서는 전체 변수의 확률적 지배 관계가 일부 변수에 대한 확률적 지배 관계를 의미하지만, $s \geq 2$인 경우에는 별개의 문제가 된다.

예컨대, 집단 1과 집단 2의 임금과 여가 시간을 알고 있을 때, 집단 1의 임금이 집단 2의 임금을 2차 확률적으로 지배하고, 집단 1의 여가 시간도 집단 2의 여가 시간을 2차 확률적으로 지배할 수 있다. 그러나 이것은 집단 1의 임금과 여가 시간이 집단 2의 임금

과 여가 시간에 대하여 2차원에서의 2차 확률적 지배 관계를 누린다는 것에 대하여 필요조건도 아니고 충분조건도 아니다.

혹은 임금에 대해서는 집단 1이 집단 2를 확률적으로 지배하지만, 여가 시간은 집단 2가 집단 1을 확률적으로 지배하는지 검정하는 것에 관심이 있을 수도 있다. 보다 일반적으로 귀무가설을 표현하면 $M_1 + M_2 = M$인 M_1과 M_2에 대하여

H_0: 첫 M_1 차원의 벡터에 대해서 집단 1이 집단 2를 s차 확률적으로 지배하고, 마지막 M_2 차원의 벡터에 대해서는 집단 2가 집단 1을 s차 확률적으로 지배한다.

$X = (X^{M_1}, X^{M_2})$라고 쓰자. 그리고 한계 분포 함수를 $F_{k,M_1}(X) = F_k(X^{M_1}, \overline{X}^{M_2})$와 $F_{k,M_2}(X) = F_k(\overline{X}^{M_1}, X^{M_2})$로 나타내자. 그리고 각각의 한계 분포 함수를 반복 적분하여 $D^s_{k,M_1}(X^{M_1})$과 $D^s_{k,M_2}(X^{M_2})$를 구할 수 있다. 이를 바탕으로 다음을 정의하자.

$$\tilde{d} = \sup_{x^{M_1}} [D^s_{1,M_1}(x^{M_1}) - D^s_{2,M_1}(x^{M_1})] + \sup_{x^{M_2}} [D^s_{2,M_2}(x^{M_2}) - D^s_{1,M_2}(x^{M_2})]$$

귀무가설 하에서는 윗 식이 음수가 될 것이다. 이를 검정하기 위한 검정 통계량은 다음과 같다.

$$\tilde{T} = \sqrt{N} \sup_{x^{M_1}} [\hat{D}^s_{1,M_1}(x^{M_1}) - \hat{D}^s_{2,M_1}(x^{M_1})] + \sqrt{N} \sup_{x^{M_2}} [\hat{D}^s_{2,M_2}(x^{M_2}) - \hat{D}^s_{1,M_2}(x^{M_2})]$$

또 다른 응용으로 위에서 제시한 두 가지의 귀무가설을 함께 검정할 수 있다. 이 경우 T와 $\tilde{T}$를 더하여 검정 통계량으로 삼으면 다음 귀무가설을 검정할 수 있다.

H_0: 첫 M_1 차원의 벡터에 대해서 집단 1이 집단 2를 s차 확률적으로 지배하고, 마지막 M_2 차원의 벡터에 대해서도 집단 1이 집단 2를 s차 확률적으로 지배한다. 그리고 동시에 M차원 벡터에 대해서는 집단 1이 집단 2를 s차 확률적으로 지배한다.

4.3. 相互 依存的인 確率 變數의 比較

지금까지 우리는 독립적인 확률 변수에 대해서만 고려하였다. 그러나 일반적으로는 우리는 두 가지의 확률적 의존성을 고려해야 한다. 첫째는 집단 사이의 확률적 의존성이다. 만일 우리가 납세 이전과 이후의 소득에 대한 자료를 조사하여 그 분포를 비교할 때,

각 개인에 대하여 납세 전후의 소득은 상관관계가 아주 높을 것이다. Linton, Maasoumi, and Whang(2005)에 의하면 다행히도 이러한 상호 의존성은 특별한 가정을 필요로 하지 않으며 큰 문제를 일으키지 않는다. 검정 통계량도 그대로 사용 가능하다. 단 점근적 분포를 계산할 때 분산을 계산할 때 상호 의존성으로 인해 共分散(covariance)이 0이 아니므로 이를 고려해야 하는 부분이 생긴다. 그리고 bootstrap 방법을 사용할 때 주의를 해야 한다. 이러한 문제는 $X_i = (X_{1i}, X_{2i})$으로 정의하고 표본을 재추출할 때 $k = 1, 2$의 자료를 한 쌍으로 생각하여 재표본을 추출하면 해결된다.

또 다른 확률적 의존성은 각 집단 내에서 일어나는 확률적 의존성이다. 특히 시계열(time series) 자료를 분석할 때 주로 발생한다. 자산 1과 자산 2의 일별 수익률 자료를 이용하여 비교한다면, 추세(momentum)로 인하여 수익률 사이의 시계열적 상관관계가 존재할 것을 추측할 수 있다. 하지만, 이러한 상호 의존성이 적절한 조건을 만족한다면 위에서 사용한 검정 통계량을 그대로 사용할 수 있음이 Klecan, McFadden, and McFadden (1990)과 Linton, Maasoumi, and Whang(2005)에서 증명되었다. 그 조건은 다음과 같다.

(가정 3′) 각 $k = 1, 2$에 대하여 $\{X_{ki} : i = 1, \ldots, N\}$는 강정상성(strict stationarity)을 지니고 있으며 α 혼합 과정을 따르고 강혼합계수(strong mixing coefficient)는 $\alpha(m) = O(m^{-A})$를 만족한다. 단 $A > 1$인 양수이다.

이 조건이 만족된다면 검정 통계량은 잘 알려진 극한 분포를 가진다. 그러나 주의할 점은 bootstrap 방법을 사용할 경우, 상호독립적으로 표본 재추출을 하면 자료가 지닌 상호 의존성이 재추출된 표본에서 나타나지 않으므로 문제가 발생한다. 이 경우 자료의 생성 과정에 대한 유한 차원 혹은 무한 차원의 모수를 포함하는 모형 가정을 바탕으로 한 표본 재추출을 하거나 상호 의존성 구조를 살릴 수 있는 block bootstrap 등의 방법을 사용해야 한다.

4.4. 回歸 殘差(Regression Residual)을 利用한 比較

Linton, Massoumi, and Whang(2005)에서는 비교의 대상이 되는 확률 변수를 모형 추정 결과 얻어지는 잔차로 생각하는 방식을 제안하였다. 우리는 많은 경우에 대상을 비교할 때 특정한 조건을 가정하거나 외부적인 영향을 제거하는 것을 필요로 할 때가 많다. 예컨대 두 집단의 소득을 비교한다면, 각 집단에 속한 개인의 교육수준, 나이 등의 요소로부터의 영향을 제거한 후 비교하는 것이 의미를 지닐 것이다. 이러한 일반화는 특히 정책 결정에 있어서 매우 유용한 분석 방식이 된다. 이러한 상황에서 우리가 비교하고자

하는 확률 변수 $\{X_{ki} : i = 1, \ldots, N\}$는 관측이 불가능하다. 대신, 관측 가능한 변수 $\{Y_{ki}, Z_{ki} : i = 1, \ldots, N\}$와 알려지지 않은 모수 μ_k, θ_k에 의존한다. 변수들과 사이에는 구체적으로 다음과 같은 일반적인 선형 모형의 가정이 만족된다.

$$Y_{ki} = \mu_k + Z_{ki}\theta_k + \varepsilon_{ki}$$
$$E(\varepsilon_{ki} \mid Z_{ki}) = 0$$

즉, $\{X_{ki} : i = 1, \ldots, N\}$는 위의 선형 회귀분석 모형에서 절편(intercept)이 포함되어 있는 攪亂項(disturbance term)으로 생각할 수 있다. 그리고 θ_k를 추정하는 문제에 관해서는 다양한 방법이 존재한다. 적절히 θ_k를 $\hat{\theta}_k$로 추정하였다면 $Y_{ki} - Z_{ki}\hat{\theta}_k$를 $\hat{X}_{ki}$로 정의하고 X_{ki}의 비교에 대신 이용할 수 있을 것이다.

혹은 선형 모형의 가정이 적절하지 않다면 보다 일반적인 함수의 형태를 가정할 수도 있다. 알려지지 않은 유한차원의 모수 θ_k뿐 아니라 알려지지 않은 함수 $\tau_k(\cdot)$에도 X_{ki}가 의존한다고 가정하자. 예를 들어 $X_{ki} = Y_{ki} - Z^1_{ki}\theta_k - \tau_k(Z^2_{ki})$와 같은 부분 선형 모형(Partially linear model)이나 $X_{ki} = Y_{ki} - \tau_k(Z_{ki}\theta_k)$와 같은 단일 지수 모형(Single index model)을 생각해볼 수 있다.[(3)]

$\{\hat{X}_{ki} : i = 1, \ldots, N\}$를 구했다면 이것을 $\{X_{ki} : i = 1, \ldots, N\}$과 같이 생각하고 검정 통계량을 구할 수 있다. 즉, $\hat{F}_k(x) = \frac{1}{N}\sum_{i=1}^{N} 1(\hat{X}_{ki} \le x)$을 바탕으로 경험적 누적 분포 함수를 구하고 이를 이용하여 다음과 같은 검정 통계량을 계산하면 된다.

$$T = \sqrt{N}\,\sup_{x \in [0, \overline{X}]} [\hat{D}^s_1(x) - \hat{D}^s_2(x)]$$

그리고 bootstrap 방법을 사용할 때는 모수 μ_k, θ_k, τ_k의 추정 과정을 반영해야 한다. 간단히 선형 모형에 대한 일반적인 bootstrap 과정을 소개하면 다음과 같다.

(1) $\{\hat{X}_{ki} : i = 1, \ldots, N\}$로부터 재추출을 통해 새로운 표본 $\{\hat{X}^*_{ki} : i = 1, \ldots, N\}$를 뽑는다.

(2) 평균이 0인 성질을 만족하는 잔차 $\hat{\varepsilon}^*_{ki} = \hat{X}^*_{ki} - \frac{1}{N}\left(\sum_{j=1}^{N} \hat{X}^*_{kj}\right)$을 정의한다.

---

(3) 비모수 및 준모수모형을 이용한 엄밀한 확률적 지배 관계 검정법은 Linton, Song, and Whang (2010)을 참조하라.

(3) $\hat{Y}^{*}_{ki} = \hat{\mu}_k + Z_{ki}\hat{\theta}_k + \hat{\varepsilon}^{*}_{ki}$와 같이 종속 변수를 복원한다.

(4) $\{\hat{Y}^{*}_{ki}, Z_{ki} : i = 1, \ldots, N\}$를 이용하여 $\hat{\mu}^{*}_{k}, \hat{\theta}^{*}_{k}$을 추정한다.

(5) $\hat{X}^{**}_{ki} = \hat{Y}^{*}_{ki} - Z_{ki}\hat{\theta}^{*}_{k}$으로 정의하면 이는 모형 추정 과정을 반영하여 추출된 표본이다.

(6) $\{\hat{X}^{**}_{ki} : i = 1, \ldots, N\}$을 이용하여 bootstrap 검정 통계량을 계산한다.

(7) 위의 과정을 충분히 많이 반복한다.

4.5. 確率的 支配 效率性(Stochastic Dominance Efficiency)

금융 시장의 資産構成(portfolio)은 확률적 우월성의 중요한 연구 대상이다. 금융 시장에서 투자자가 보유한 자산 구성이 과연 시장에서 가장 효율적인 자산 구성인지 아닌지 검정하는 것은 투자자의 최적행동을 분석함에 있어서 매우 중요한 역할을 한다. 그러나 이러한 경우 두 개의 집단이 주어지는 것이 아니고 유한 개의 서로 다른 자산의 선형 구성을 모두 비교해야 하는 문제에 직면하게 된다. 즉, $\{X_t = (X_{1t}, \ldots, X_{Kt})' : t = 1, \ldots, T\}$을 K 개의 자산 수익률에 대한 T기간 동안의 관측치라고 하고, $\{Y_t : t = 1, \ldots, N\}$를 우리가 비교하고 싶어하는 기준 자산(Benchmark Asset)의 수익률이라고 하자. $\sum_{k=1}^{K} \lambda_k = 1$와 $\lambda_k \geq 0, \forall k = 1, \ldots, K$을 만족시키는 벡터 공간 Λ와 $\lambda = (\lambda_1, \ldots, \lambda_k)'$를 생각하자. 그러면 k번째 자산에 λ_k의 가중치를 주는 자산구성의 수익률을 $X_t'\lambda$라고 나타낼 수 있다. 만일 자산 Y가 임의의 $\lambda \in \Lambda$가 주어지더라도 $X_t'\lambda$에 대해 s차 확률적으로 지배되지 않을 때, Y가 s차 확률적 지배 효율성을 지닌다고 정의한다.

전통적인 금융 경제학의 평균-분산의 2차원의 정보만을 이용한 소위 평균-분산 효율적 경계역 접근 방법(Mean-Variance Efficient Frontier approach)과 비교하였을 때, 확률적 지배 효율성을 이용한 접근 방법은 투자자의 효용함수 혹은 자산 수익률 분포 함수에 대한 가정을 매우 一般化시키기 때문에 잘못된 가정으로부터 잘못된 결론이 도출될 우려가 적은 보다 강인한(robust) 접근 방법이다.

다음과 같은 귀무가설을 설정하자.

H_0: Y가 X의 선형 결합으로 구성된 어떠한 자산구성에 대해서도 s차 확률적으로 지배되지 않는다. 즉, Y는 X에 대해 s차 확률적 지배 효율성을 가진다.

이 귀무가설을 검정하기 위해 다음 식을 생각해보자.

$$d = \max_{\lambda \in \Lambda} \min_{x \in X} [D^{s}_{Y}(x) - D^{s}_{\lambda}(x)]$$

여기서 $D^s_Y(\cdot)$와 $D^s_\lambda(\cdot)$는 확률적 지배 관계의 정의에서와 마찬가지로 각각 Y와 $X_t'\lambda$의 분포 함수의 반복 적분으로 정의된 함수이다. 만일 귀무가설이 옳다면 d는 0보다 작거나 같을 것이다. 그러나 이를 검정의 기초 함수로 채택하기에는 한 가지 문제가 있다. 그것은 대립가설 하에서도 일부 경우에 $d=0$이 될 수 있다는 것이다. 예를 들어 1차 확률적 지배 효율성을 검정하고자 할 때, 특정한 λ에 대해 $X'\lambda$와 Y의 분포 함수가 일부분에서 일치할 수 있다. 이러한 경우 대립가설에서 $d=0$의 값을 갖게 되어 검정이 힘들어진다. Linton, Post, and Whang(2005)에서는 이러한 문제를 해결하기 위해 분포 함수가 일치하는 구간 B를 고려한다.

$$d_B = \max_{\lambda \in \Lambda} \min_{x \in B} [D^s_Y(x) - D^s_\lambda(x)]$$

d_B는 대립가설 하에서는 항상 양수이다. 그러므로 이를 바탕으로 검정 통계량을 정의할 수 있다. $\hat{B}$를 각각 B에 대한 추정치라 하였을 때, 검정 통계량은 다음과 같다.

$$\hat{d}_{\hat{B}} = \max_{\lambda \in \Lambda} \min_{x \in \hat{B}} \sqrt{N}\, [\hat{D}^s_Y(x) - \hat{D}^s_\lambda(x)]$$

위의 검정 통계량의 기각역을 구하기 위해서는 bootstrap이나 부분 추출 방법을 사용할 수 있다.

4.6. 確率的 單調性(Stochastic Monotonicity)

확률적 우월성은 유한개의 집단 간의 비교에 적용 가능한 이론이다. 그러나 유한개의 집단이 아닌, 연속적인 분포를 이루는 특성들 사이에서의 비교를 생각할 필요성이 있다. 예를 들어, 부모가 모두 직장 생활을 하는 경우와 그렇지 않은 경우에 자녀들의 학업 성취도를 비교하는 문제는 2개의 집단을 비교하는 문제이지만, 부모의 소득에 따라 자녀들의 학업 성취도가 어떻게 변하는지 비교하는 것은 “所得”이라는 연속적인 특성에 근거한 비교가 된다. 이러한 경우 기존에 제시된 확률적 우월성의 개념을 적용하기 힘들고 확률적 단조성이라는 새로운 개념의 도입이 필요하다. Lee, Linton, and Whang(2009)에서는 이러한 경우에 적용 가능한 확률적 우월성의 개념을 연속적인 경우로 확장하여 확률적 단조성이라는 개념을 제시한다. X_t의 변화에 따라 Y_t의 분포를 비교하는 경우를 상정하자. 즉, 위의 예시에서 X_t는 부모의 소득, Y_t는 자녀들의 학업 성취도에 해당된다. 엄밀한 정의를 위해 몇 가지 정의가 필요하다. $F(Y|X)$를 X에 대한 Y의 조건부 확률 분포 함수라고 하자. 그리고 이 조건부 분포함수로부터 $D^s(Y|X)$를 반복 적분으로 정의할 수 있다.

그러면 다음과 같이 s차 확률적 단조성이 정의된다.

$$D^s(y\,|\,x) \le D^s(y\,|\,x'),\ \forall x \ge x',\ \forall y \in \mathbb{R}$$

즉 x가 커질수록 그에 대한 Y의 조건부 분포는 s차 확률적 지배 관계를 누리는 쪽으로 변한다는 의미이다.

X와 Y의 두 변수의 관계를 분석하기 위한 대표적인 방법으로 회귀분석이 있다. 그러나 회귀분석은 설명변수 X에 대한 종속변수 Y의 조건부 기대값을 구하여 X와의 관계를 보는 것이다. 반면 확률적 단조성은 분포의 변화 전체를 분석하는 방법이므로 보다 강한 결론을 이끌어낼 수 있다. 그러나 이에 대한 통계적 검정 방법은 이론적으로 매우 어려운 문제이다. 여기서는 간단한 1차 확률적 단조성에 대한 검정 방법을 소개하고자 한다. 우리가 검정하고자 하는 귀무가설은

H_0: X에 대한 Y의 조건부 확률 분포는 1차 확률적 단조성을 지닌다.

이다. 이를 검정하기 위해 다음의 U-통계량을 고려하자.

$$\hat{U}_N(y,\ x) = \frac{2}{N(N-1)} \sum_{1 \le i < j \le N} [1(Y_i \le y) - 1(Y_j \le y)]\mathrm{sgn}(X_i - X_j)K_h(X_i - x)K_h(X_j - x)$$

여기서 $K_h(\cdot) = h^{-1}K(\cdot/h)$, $\mathrm{sgn}(x) = 1(x>0) - 1(x<0)$, 그리고 h는 bandwidth를 의미한다. 적절한 조건하에서 $N \to \infty$에 따라 우리는 다음이 성립함을 보일 수 있다.

$$\begin{aligned} &h^{-1}E\hat{U}_N(y,\ x) \\ &\to F_x(y\,|\,x)\left(\iint |u_1 - u_2|\, K(u_1)K(u_2)du_1du_2\right)[f_X(x)]^2 \end{aligned}$$

여기서 $F_x(y\,|\,x)$는 $F_{Y|X}(y\,|\,x)$의 x에 대한 편미분 값이다. 따라서 귀무가설 하에서는 모든 (y, x)에 대하여 $F_x(y\,|\,x) \le 0$이나, 대립가설 하에서는 $F_x(y\,|\,x) > 0$이 되는 (y, x)가 존재하므로 다음의 검정 통계량을 고려할 수 있다.

$$S_N = \sup_{(y,x)} \frac{\hat{U}_N(y,\ x)}{c_N(x)}$$

여기서 $c_N(x) = \hat{\sigma}_N(x) / \sqrt{N}$ 을 의미하며

$$\hat{\sigma}_N^2(x) = \frac{4}{N(N-1)(N-2)} \sum_{1 \le i \ne j \ne k \le N} \operatorname{sgn}(X_i - X_j) \operatorname{sgn}(X_i - X_k)$$
$$\times K_h(X_j - x) K_h(X_k - x) [K_h(X_i - x)]^2$$

을 나타낸다.

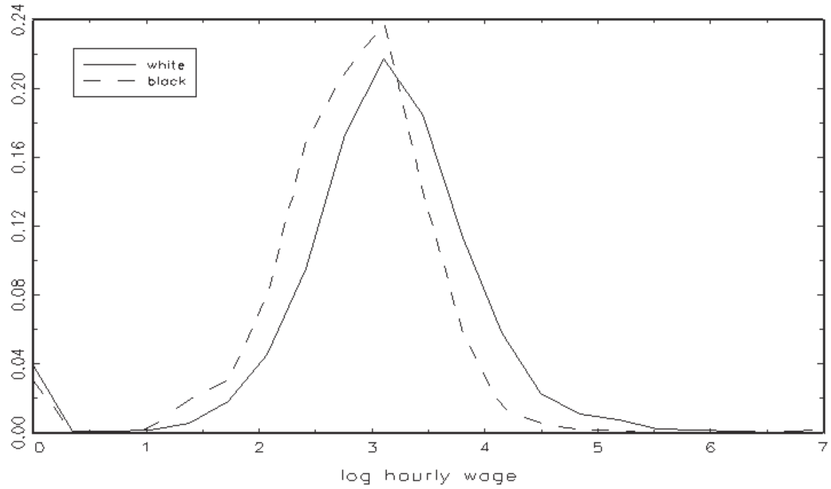
위의 통계량은 Ghosal, Sen, and van der Vaart(2000)의 회귀함수(즉, 조건부 평균함수)의 단조성을 검정하기 위해 고려한 Rank Correlation 통계량의 idea를 조건부 분포함수의 경우로 확장한 것이다. Lee, Linton, and Whang(2009)은 적절한 조건하에서 통계량 S_N의 점근적 분포는 극한치 분포(Extreme Value distribution)가 됨을 보였다. 즉,

$$\lim_{N \to \infty} \Pr[4\beta_N (S_N - \beta_N) < x] = \exp(-e^{-x})$$

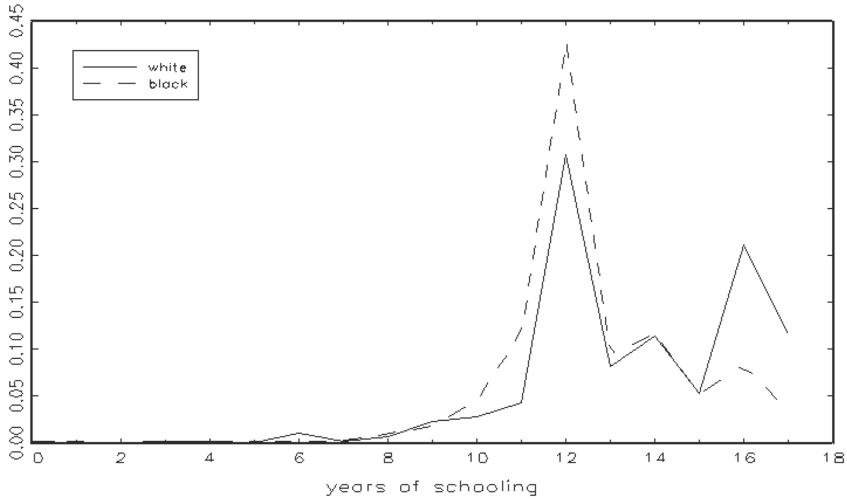
여기서, β_N는 $N \to \infty$에 따라 적절한 속도로 무한대로 발산하는 수열이다. 후자의 논문은 이 검정법을 Y가 아들의 소득이고 X가 부모의 소득인 경우 세대 간 소득의 이전문제에 적용하여 경험적 유용성을 입증하였다. 그러나 Lee, Linton, and Whang(2009)의 검정법은 극한치 분포에의 수렴의 속도가 매우 느린 것으로 알려져 소규모 표본에서 사용되기에는 한계가 있다.

5. 實證分析 事例: 人種 間 賃金 隔差 分析

본 장에서는 Park(2008)에서 행해진 흑인과 백인 사이의 임금 격차를 검정하는 실증 분석을 통해 확률적 지배 관계가 실증적으로 어떻게 사용되는지 알아보고자 한다. 보다 구체적으로, 교육 수준이 인종 간의 임금 격차에 미치는 영향을 제거하였을 때 임금의 분포를 비교함으로써 인종 간 불평등을 해소하기 위한 정부의 교육 정책의 효과를 살펴보고자 한다. 본 주제에 관련된 기존의 연구는 대부분 선형 회귀분석의 모형을 이용하여 인종에 대한 더미 변수(dummy variable)에 대한 추정 계수를 비교하는 방법을 사용한다. 그러나 이러한 분석 방법은 분포의 한 측면만을 바라보는 것이므로 확률적 지배 우월성을 사용하여 보다 강인한(robust) 분석이 가능하다. 또한, 교육이 임금에 미치는 영향을 추정하기 위해 비모수적 부분 선형 모형(Partial linear model)을 사용하였다. 본 실증 연구에 사용된 자료는 미국의 PSID(Panel Study of Income Dynamics)로부터 구한 2005년 횡단면 자



〈그림 1〉 黑人과 白人의 賃金 分布



〈그림 2〉 黑人과 白人의 敎育 水準 分布

료이다. 분석에 사용된 자료는 직업을 가진 2,950명의 백인과 964명의 흑인에 대한 임금, 교육수준, 성별, 나이, 근로시간, 거주지역, 혼인 여부, 노동조합 가입 여부, 종사 산업, 종사 업무의 자료를 사용하였다.

〈그림 1〉은 흑인과 백인의 시간당 임금에 자연로그를 취하여 분포를 그린 것이다.[(4)] 전반적으로 백인의 임금 분포가 흑인보다 오른쪽으로 치우친 것을 확인할 수 있다. 즉,

백인의 임금 분포가 흑인의 임금 분포에 대하여 확률적 지배 관계를 가진다는 것을 추측할 수 있으며, 이것이 통계적으로 유의한지 검정하고자 한다.

〈그림 2〉는 흑인과 백인의 교육 수준의 분포를 나타낸다. 교육 수준은 몇 년의 교육을 이수하였는지를 기준으로 하였다. 그림에서 보다시피 대학 교육을 마친 비율이 백인에게서 월등히 높음을 볼 수 있다. 백인의 높은 교육 수준이 흑인에 비해 높은 임금을 받는 주된 원인의 하나라고 일반적으로 생각된다. 다음과 같은 방법으로 확률적 지배 관계의 관점에서 교육 수준이 인종 간 임금 격차의 주된 원인이라는 주장의 타당성을 검정해보고자 한다. 우선, 임금을 교육 수준과 다른 설명변수에 대하여 회귀분석을 하고, 회귀 결과를 토대로 교육 수준이 일정한 수준으로 통제되었을 때 과연 확률적 지배 관계가 관측되는지를 볼 것이다. 만일 확률적 지배 관계가 관측된다면 교육 수준 이외의 다른 요인이 남아서 인종 간의 격차를 계속 유발하고 있음을 알 수 있다.

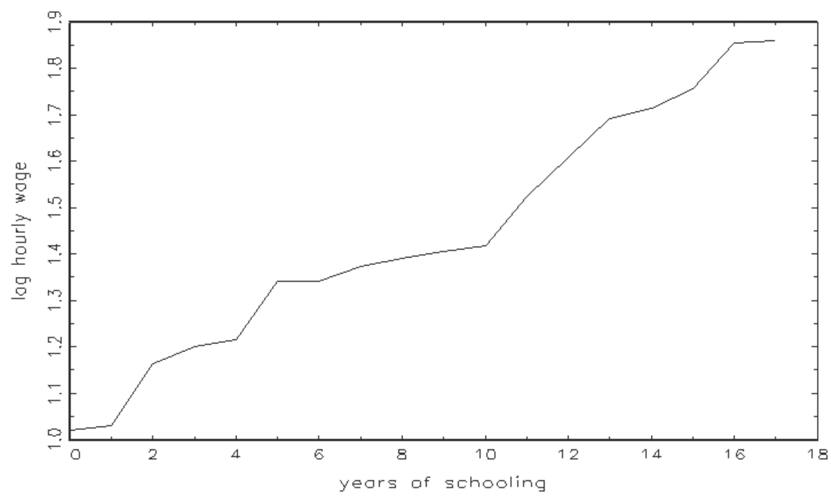
이제 $k=b, w$는 각각 흑인과 백인을 나타낸다고 하자. 그리고 y_k, d_k, s_k, z_k는 각각 자연로그를 취한 시간당 임금, 흑인과 백인을 나타내는 더미 변수, 교육 수준, 기타 설명 변수를 나타낸다고 하자. 그리고 이 변수들의 관계를 다음과 같이 설정하자.

$$y_k = d_k\gamma + \tau(s_k) + z_k\beta + x_k$$

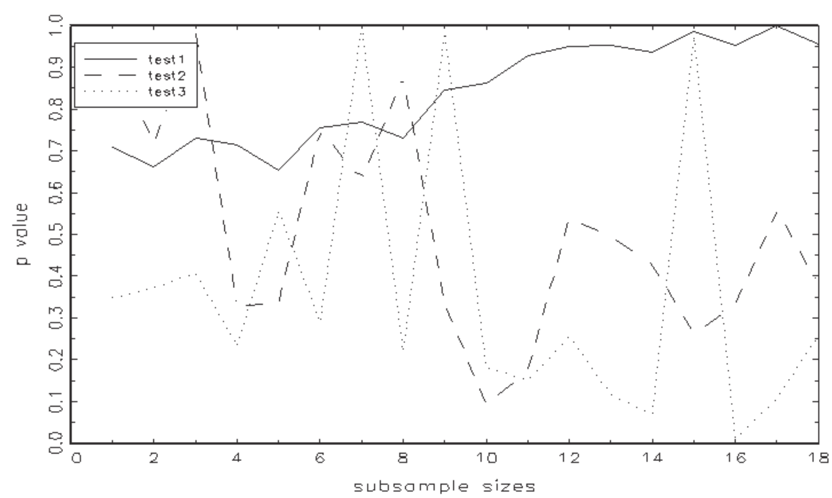
γ, β는 미지의 벡터이고, τ는 미지의 함수이다. x_k는 이 관계식에서 다른 설명변수로 설명되지 않는 오차 항이다. Robinson(1988)의 부분 선형 모형 추정 방법을 사용하여 τ를 추정한 결과가 〈그림 3〉에 나타나 있다. 구체적으로 Nadaraya-Watson 추정 방법에 Silverman의 bandwidth를 이용했으며 Chernozhukov *et al.*(2007)의 재배열을 통한 단조함수 추정 방법을 사용하였다.

이 추정치를 바탕으로 우리는 세 가지의 변수에 대하여 1차 확률적 지배 관계와 2차 확률적 지배 관계의 관점에서 비교할 것이다. 첫 번째로 비교할 변수는 아무런 통제가 되지 않은 자연로그를 취한 시간당 임금이다. 이를 통해서 우선 흑인과 백인의 임금에 확률적 지배 관계가 실제로 존재하는지 알아볼 수 있다. 둘째는 첫 번째 변수에 교육 수준을 제외한 다른 설명변수를 통제한 $y_k - z_k\hat{\beta}$이고, 마지막으로 교육수준까지 통제된 $y_k - \hat{\tau}(s_k) - z_k\hat{\beta}$을 비교할 것이다. 편의를 위해 $y_k^2 = y_k - z_k\hat{\beta}$, $y_k^3 = y_k - \hat{\tau}(s_k) - z_k\hat{\beta}$으로 표기하자. y_k^2의 비교는 교육 수준의 영향을 알아보기 위한 대조군에 해당되며, y_k^3의 결과와 비교하

(4) 임금이 0인 경우 정의되지 않기 때문에 모든 자료에 1을 더하여 로그를 취하였다.



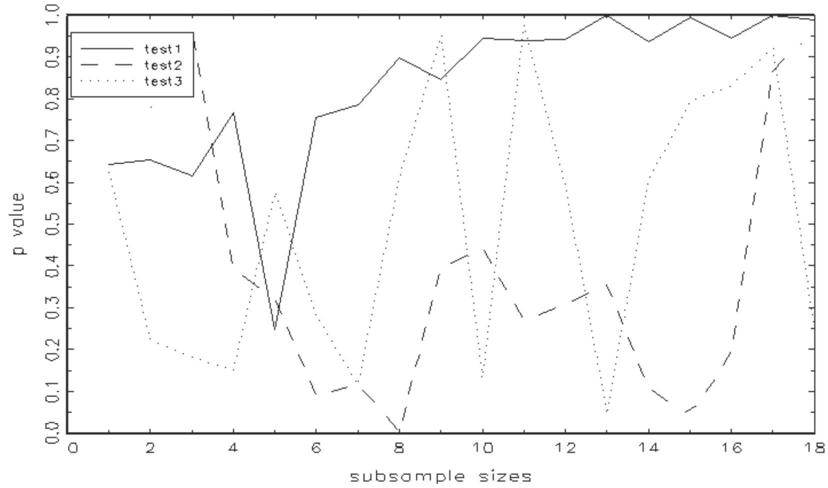
〈그림 3〉 敎育 水準에 대한 賃金 函數 推定



〈그림 4〉 1次 確率的 支配 關係 檢定 結果

여 교육이 임금 불평등에 미치는 영향을 살펴볼 것이다.

각 변수에 대하여 1차와 2차 확률적 지배 관계의 존재를 귀무가설로 설정한 후 총 6개의 검정을 시행하였다. 귀무가설 하에서의 분포는 Linton, Maasoumi, and Whang(2005)에서 제시한 部分 抽出 方法(subsampling method)을 이용하여 구하였다. 두 집단의 부분 표본의 크기의 비율은 Linton(2004)에서 제시한 방법을 사용하였고, 부분 표본의 크기를 바



〈그림 5〉 2次 確率的 支配 關係 檢定 結果

꾸어 18번의 p값을 구하였다. 〈그림 4〉와 〈그림 5〉는 18개의 p값을 그래프로 나타낸 것이다.

결과를 살펴보면 1차와 2차 모두 y_k는 전반적으로 높은 p값을 보인다. 즉, 흑인과 백인 사이의 임금에는 1차와 2차 확률적 지배 관계가 모두 존재한다고 생각할 수 있다. 그리고 y_k^2와 y_k^3의 결과를 보면 추정된 변수를 사용했기 때문에 결과가 매우 불안정하지만 두 p값 사이의 차이가 크지 않은 것으로 보인다. 즉, 교육 수준을 통제하거나 그렇지 않거나 결과가 크게 달라지지 않은 것을 보아 교육 수준 이외의 다른 요인들이 인종 간의 임금 차이에 이미 존재함을 추측할 수 있다.

6. 結 論

지금까지 우리는 확률적 지배 관계의 개념과 응용에 대하여 구체적으로 살펴보았다. 확률적 지배 관계는 최소한의 가정만을 바탕으로 경제학적으로 유의미한 확률 변수 간의 비교를 가능하게 해준다. 이에 대한 계량 경제학적 검정 방법 중, Anderson 검정 방법, Davidson-Duclos 검정 방법 등은 계산이 간편하지만 일부 대립가설에서 검정력이 떨어지며, KS type 통계량을 사용한 검정 방법들은 모든 대립가설에 대하여 일치성을 보장하지만 극한 분포를 도출하기가 쉽지 않은 단점이 있다. 또한 시계열 확률 변수, 다차원 확률 변수, 회귀 잔차 등에 사용되는 방법도 개발되어 있으며, 확률적 지배 효율성, 확률적 단

조성 등 특수한 상황에 보다 적합한 방법들도 존재한다. 그러므로 분석 대상과 목적에 따라 적절히 사용해야 할 것이다.

Yale 大學校 經濟學科 博士課程
서울大學校 經濟學部 BK21 事業團
Department of Economics
Yale University
New Haven, CT 06520
E-mail: byounggun.park@yale.edu

서울大學校 經濟學部 敎授
151-746 서울특별시 관악구 관악로 599
전화: (02)880-6362
팩스: (02)886-4231
E-mail: whang@snu.ac.kr

參 考 文 獻

Anderson, G.(1996): "Nonparametric Tests of Stochastic Dominance in Income Distribution," *Econometrica*, **64**, **5**, 1183-1193.

Barrett, G. F., and S. G. Donald(2003): "Consistent Tests for Stochastic Dominance," *Econometrica*, **71**, **1**, 71-104.

Chernozhukov, V., I. Fernandez-Val, and A. Galichon(2007): "Improving Estimates of Monotone Functions by Rearrangement," working paper.

Davidson, R., and J.-Y. Duclos(2000): "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality," *Econometrica*, **68**, **6**, 1435-1464.

Ghosal, S., A. Sen, and A. W. van der Vaart(2000): "Testing Monotonicity of Regression," *Annals of Statistics*, **28**, 1054-1082.

Klecan, L., R. McFadden, and D. McFadden(1991): "A Robust Test for Stochastic Dominance," working paper.

Lee, S., O. Linton, and Y.-J. Whang(2009): "Testing for Stochastic Monotonicity," *Econometrica*, **77**, **2**, 585-602.

Lehmann, E. L.(1955): "Ordered Families of Distributions," *Annals of Mathematical Statistics*, **26**, 399-419.

Linton, O.(2004): "Nonparametric Inference for Unbalanced Time Series Data," *Econometric Theory*, **21**, 143-157.

Linton, O., E. Maasoumi, and Y.-J. Whang(2005) : "Consistent Testing for Stochastic Dominance under General Sampling Schemes," *Review of Economic Studies*, **72**, 735-765.

Linton, O., T. Post, and Y.-J. Whang(2005): "Testing for Stochastic Dominance Efficiency," working paper.

Linton, O., K. Song, and Y.-J. Whang(2010): "An Improved Bootstrap Test of Stochastic Dominance," *Journal of Econometrics*, **154**, 186-202.

McCaig, B., and A. Yatchew(2007): "International Welfare Comparisons and Nonparametric Testing of Multivariate Stochastic Dominance," *Journal of Applied Econometrics*, **22**, 951-969.

McFadden, D.(1989): "Testing for Stochastic Dominance," in T. B. Fomby, and T. K. Seo(eds.), *Studies in the Economics of Uncertainty: In Honor of Josef Hadar*, Springer-Verlag.

Park, B. G.(2008): "Testing for Stochastic Dominance with an Application to Racial Wage Inequality," Seoul National University.

Politis, D. N., and J. P. Romano(1994): *Subsampling*, Springer-Verlag.

Robinson, P. M.(1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, **56**, **4**, 931-954.

Rothschild, M., and J. E. Stiglitz(1970): "Increasing Risk: I. A Definition," *Journal of Economic Theory*, **2**, 225-243.