응용통계연구 (2010)
23(5), 909–921
DOI: 10.5351/KJAS.2010.23.5.909

BLS 무응답 보정법을 이용한 대체법과 이월대체법에 관한 연구

이상은[1] · 신기일[2]

[1]경기대학교 응용정보통계학과, [2]한국외국어대학교 통계학과

(2010년 5월 접수, 2010년 7월 채택)

요 약

패널 자료에서 무응답이 발생한 경우에는 횡시점회귀대체법(cross-wave regression imputation) 등과 같은 대체법을 이용하여 무응답 문제를 해결한다. 최근 표본 틀(sampling frame) 자료를 이용하여 무응답 가중치 보정을 하는 BLS 무응답 보정법은 패널 자료에도 적용 가능한 방법으로 알려져있다. 본 논문에서는 패널자료에서 BLS 무응답 보정법을 이용한 대체법을 연구하였으며 자료가 경향이 있는 비정상시계열(nonstationary process with drift)을 따른 다는 조건하에서 BLS 무응답 보정법과 횡시점회귀대체법의 하나인 이월대체법(carry-over imputation)과의 이론적 관계를 살펴보았다. 모의실험을 통하여 이론적인 결과를 확인하였으며, 2007년 매월노동통계 자료를 이용하여 두 방법의 우수성을 비교하였다.

주요용어: BLS 방법, 무응답 보정, 무응답 대체, 횡시점회귀대체, 이월대체.

1. 서론

추출된 표본이 모집단을 잘 대표하는 것은 매우 중요하다. 그러나 표본 설계와는 달리 실사 과정에서 많은 무응답이 발생할 수 있다. 특히 패널자료를 이용한 분석에서는 무응답에서 발생하는 결측치(missing value) 처리 방법이 총계 추정의 정확도에 중요한 역할을 하고 있기 때문에 결측치 대체에 관한 연구가 활발히 진행되고 있다 (Rubin, 1987; Little과 Rubin, 2002).

무응답 문제를 해결하기 위한 방법으로 결측치에 대체값을 대체하는 결측치 대체법과 가중치를 보정하는 방법이 있다 (신민웅과 이상은, 2001; Särndal 등, 1992). 패널 자료의 결측치 대체법으로 쉬우면서도 효과적인 방법이 횡시점회귀대체법이다. 이 방법은 전 시점 자료를 독립변수로, 현 시점 자료를 종속변수로 하는 단순회귀 모형을 이용하여 대체하는 회귀대체법이다 (김동욱과 노영화, 2003; Lepkowski, 1989). 반면 직접 결측치를 대체하지 않고 가중치를 변화시켜 무응답 문제를 해결하는 방법 중 하나가 BLS 무응답 보정법이다. 이 방법은 미국 노동통계국(Bureau of labor statistics)이 사업체 조사 시 발생한 무응답 문제를 해결하기 위해 만든 것으로 패널자료 분석에 결측 대체법의 하나로 사용될 수 있다. 최근 김석과 신기일 (2009)은 무응답이 있는 패널 자료의 총계를 추정할 경우에 BLS 보정법의 우수성을 확인하였다. 그러나 총계추정에 있어서 BLS 무응답 보정법과 횡시점회귀대체법 등 패널 자료 분석

이 논문은 2008년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2008-313-C00141).
[2]교신저자: (449-791) 경기도 용인시 모현면 왕산리 산 89, 한국외국어대학교 통계학과, 교수.
E-mail: keyshin@hufs.ac.kr

에서 사용하는 결측치 대체법과의 관계에 관한 연구는 이루어지지 않았다. 따라서 BLS 무응답 보정법을 이용한 대체법과 횡시점회귀대체법과 같은 패널자료 대체법을 연구하고 그 우수성을 비교하는 것은 매우 중요하다.

본 연구에서는 2절에서 횡시점회귀대체법과 이 방법의 특별한 경우인 이월대체법(carry-over imputation), 그리고 BLS 무응답 보정법을 이용한 결측치 대체법을 이론적으로 살펴보았다. 자료가 경향이 있는 비정상시계열(non-stationary process with drift) 또는 이의 특별한 경우인 일반 비정상시계열(non-stationary process)로 생성될 경우에 BLS 무응답 보정법에서 얻어진 결측치 대체법과 이월대체법이 근사적으로 같은 결과를 주는 것을 확인하였다. 얻어진 결과를 뒷받침하기 위해 3절에서 모의실험을 실시하였으며 노동부의 2007년 매월노동통계자료 중 3월 자료와 5월 자료를 패널 자료와 같은 구조가 되도록 만든 후 각 방법의 우수성을 비교하였다.이를 위하여 본 논문에서는 결측 메카니즘으로 MCAR(missing completely ar random)을 가정하여 분석하였다. 끝으로 4절에 토의 및 결론이 있다.

2. BLS 무응답 보정법과 횡시점회귀대체법(Cross-Wave Regression Imputation)

미국 노동통계국(Bureau of Labor Statistics)의 BLS 무응답 가중치 보정법은 사업체 조사에서 표본틀의 자료와 실사에서 얻어진 자료 사이에 상관관계가 높은 경우 사용하는 방법이다. BLS 방법의 무응답 가중치 보정법은 패널 자료에서 완전한 $t-1$ 시점의 자료가 얻어진 경우 t 시점의 무응답 가중치 보정에 사용가능하며 결과 또한 매우 우수한 것으로 나타났다 (김석과 신기일, 2009). 전반적인 BLS 방법에 관한 자세한 내용은 이상은 (2008)을 참조하기 바란다. 본 논문에서는 먼저 BLS 무응답 보정법을 결측치 대체법으로 전환하는 방법을 연구하였다. 다음으로 패널 자료에서처럼 연속되는 두 시점 자료가 존재할 경우 사용할 수 있는 방법인 횡시점회귀대체법과 이의 특별한 경우인 이월대체법에 관하여 연구하였다 (Lepkowski, 1989; Tremblay, 1994). 그리고 전환된 BLS 결측치 대체법을 횡시점회귀대체법, 이월대체법과 비교하였다.

일반적으로 실제 자료 분석에서는 모집단을 몇 개의 층으로 나누어 분석하는 층화 추출법이 사용된다. 이때 나누어진 부차 모집단(subpopulation) 별로 무응답 보정과 대체가 이루어지기 때문에 본 논문에서는 하나의 층만을 고려하여 연구하였다. 또한 t 시점과 $t-1$ 시점에서 자료가 얻어졌으며 $t-1$ 시점의 자료는 완전한 자료라고 가정하였다. $t-1$ 시점의 자료가 완전한 자료라는 가정은 첫 시점 자료는 일반적인 횡시점 대체법으로 완전한 자료를 만들 수 있기 때문에 전혀 무리가 없는 가정이다.

본 논문에서 사용할 자료의 구성은 다음과 같다. 먼저 t 시점의 자료는 다음과 같이 두 부분 $(y_{i,t}^O, y_{i,t}^m)$으로 나누어진다. 여기서 O는 관측(observed)을 의미하고 m은 결측(missing)을 의미하며 i는 i번째 관측값을 의미한다. n개의 자료 중에서 r개가 관측되었으며 따라서 $n-r$개가 결측되었다. $y_{i,t}^m$은 결측되었기 때문에 대체값 $\hat{y}_{i,t}^m$을 구하여 대체한다. $t-1$ 시점의 자료도 두 부분, $(y_{i,t-1}^O, y_{i,t-1}^m)$으로 나누어진다. $t-1$ 시점의 자료는 완전한 자료이므로 $(y_{i,t-1}^O, y_{i,t-1}^m)$는 모두 얻어진 값이다. $y_{i,t-1}^O$는 r개로 구성되었으며 $y_{i,t-1}^m$은 $n-r$개로 구성된다. 즉

$y_{i,t-1}^O$: $t-1$ 시점에서 조사된 값이며 t 시점의 값도 얻어진 값,
$y_{i,t}^O$: t 시점에서 조사된 값,
$y_{i,t-1}^m$: $t-1$ 시점에서 조사된 값이며 t 시점의 값은 결측된 값,
$y_{i,t}^m$: t 시점에서 결측된 값,
$\hat{y}_{i,t}^m$: t 시점에서 대체값.

2.1. 횡시점회귀대체(Cross Wave Regression Imputation)

횡시점회귀대체는 무응답이 있는 t번째 시점에서 $y_{i,t}^O$를 종속변수로 두고 응답이 있는 가장 가까운 과거 $t-1$번째 시점의 응답값 $y_{i,t-1}^O$을 독립변수로 하는 단순 회귀모형을 사용한다. 즉 무응답이 있는 조사단위를 제외한 후 회귀모형을 추정한다. 모형 $y_{i,t}^O = \beta_0 + \beta_1 y_{i,t-1}^O + \epsilon_i$를 이용하여 모수의 추정값 $\hat{\beta}_0$, $\hat{\beta}_1$을 구한 후 t번째 시점의 i번째 대체값 $\hat{y}_{i,t}^m = \hat{\beta}_0 + \hat{\beta}_1 y_{i,t-1}^m$을 무응답 $y_{i,t}^m$에 대체한다. 연속적으로 $t-1$번째 시점과 t번째 시점에 무응답이 있는 경우 먼저 $t-1$번째 시점의 응답값, $y_{i,t-1}^O$을 종속변수로 두고 $t-2$번째 시점의 응답값, $y_{i,t-2}^O$을 독립변수로 하여 회귀모형을 추정하여 대체한다. 같은 방법으로 t번째 시점의 무응답도 추정하여 대체한다. 다음으로 횡시점회귀대체법의 간단한 방법인 이월대체(carry-over imputation)에 관한 설명은 다음과 같다. 먼저 이월대체는 횡시점회귀대체법에서 구한 모형 $\hat{y}_{i,t}^m = \hat{\beta}_0 + \hat{\beta}_1 y_{i,t-1}^m$에 $\hat{\beta}_0 = 0$과 $\hat{\beta}_1 = 1$을 대입하여 대체하는 방법이다. 즉 무응답 $y_{i,t}^m$에 전 시점의 응답값 $y_{i,t-1}^m$을 대체하는 방법이다. 또한 연속 시점에서 무응답이 있는 경우는 응답이 있는 가장 가까운 과거 시점의 응답값으로 대체한다. 만약 두 시점의 평균이 다를 경우에는 이월대체법의 확장인 평균이 다른 이월대체법을 사용할 수 있다. 이 경우 횡시점회귀대체법에서 $\hat{\beta}_1 = 1$만을 적용하면 된다.

2.2. BLS 무응답 보정(Nonresponse Adjustment; NRA)

BLS 무응답 보정법은 가중치를 보정하는 방법이다. 먼저 기본 가중치가 동일하다고 가정하고 기본 가중치를 w라 표시하자. 이는 일반적으로 층별 가중치는 다르나 층내 가중치는 동일하고 본 연구에서는 한 개 층만을 다루기 때문에 가중치가 동일하다는 가정은 무리가 없다.

따라서 무응답이 존재하지 않는다고 가정하면 총계 추정치는 다음과 같이 구해진다.

$$\hat{t}^B = \sum_{i=1}^{n} w y_{i,t}. \tag{2.1}$$

이제 t 시점에 결측치가 있어 $y_{i,t}^O$만 얻어졌다고 하자. 그러면 결측치의 영향을 줄이기 위해 가중치를 보정해야 한다. 이때 사용할 BLS 보정인자(adjustment factor)는 다음과 같이 구해진다.

$$\text{NRA factor} = f^{NRA} = \frac{\sum_{i=1}^{n} y_{i,t-1}}{\sum_{i=1}^{r} y_{i,t-1}^O}. \tag{2.2}$$

따라서 무응답 보정 가중치는 $w^A = w \times f^{NRA}$가 되고 총계 추정치, $\hat{t}^B$는 다음과 같이 구해진다.

$$\hat{t}^B = \sum_{i=1}^{r} w^A y_{i,t}^O = \sum_{i=1}^{r} \left(w \times f^{NRA} \right) y_{i,t}^O = w \sum_{i=1}^{r} \left(\frac{\sum_{i=1}^{n} y_{i,t-1}}{\sum_{i=1}^{r} y_{i,t-1}^O} \right) y_{i,t}^O = w \left(\frac{\sum_{i=1}^{n} y_{i,t-1}}{\sum_{i=1}^{r} y_{i,t-1}^O} \right) \sum_{i=1}^{r} y_{i,t}^O. \tag{2.3}$$

2.3. 평균이 다른 이월대체법과 BLS 보정법을 이용한 총계 추정량

완전한 $t-1$ 시점의 자료가 얻어졌고 결측치가 있는 t 시점 자료의 총계를 추정한다고 하자. $t-1$ 시점 자료의 평균과 t 시점 자료의 평균을 각각 μ_t와 μ_{t-1}이라 하자. 또한 t시점의 n개 자료 중에서 r개의 자료가 얻어졌고 나머지 $n-r$개의 자료에서 결측이 발생하였다고 가정하자. 또한 결측 메카니즘은 MCAR을 따른다고 가정하자. 전 절에서도 언급하였듯이 본 연구에서는 횡시점회귀대체법의 특수한 경

우인 이월대체법에 관하여 연구하였다. 따라서 이월대체법이 적용되는 모형으로 자료가 생성되었다고 가정하자. 그러나 각 시점별로 평균이 같다고 가정하기 보다는 평균이 다른 모형이 일반적이므로 이월대체법의 확장모형이 분석에 사용되었다. 즉 자료는 다음의 모형에서 생성되었다.

모형 $$y_{i,t} - \mu_t = y_{i,t-1} - \mu_{t-1} + a_{i,t}, \quad a_{i,t} \sim iid(0, \sigma_a^2). \tag{2.4}$$

가정 1: $\mathrm{Var}(y_{i,t-1}) = \sigma_{t-1}^2 < \infty$, $\mathrm{Var}(y_{i,t}) = \sigma_{t-1}^2 + \sigma_a^2$.

가정 2: $E(y_{i,t}) = \mu_t$, $E(y_{i,t-1}) = \mu_{t-1}$.

전술한 바와 같이 $y_{i,t-1}^O : t-1$ 시점에서 조사된 값, $y_{i,t}^O : t$ 시점에서 조사된 값, $y_{i,t-1}^m : t-1$ 시점에서 조사된 값 그리고 $y_{i,t}^m : t$ 시점에서 결측된 값이다. 모형식 (2.4)는 경향이 있는 확률보행과정(random work process with drift)에 해당된다. 일반적으로 두 시점(wave)의 평균은 다르며 분산 또한 다를 것이다. 그러나 시계열 모형의 하나인 경향이 있는 확률보행과정을 자료 생성 모형으로 사용할 경우 $t-1$ 시점의 분산에 백색잡음과정(white noise process)의 분산이 더해진 값으로 t 시점의 분산이 구해져 항상 분산이 증가하는 결과가 나오게 된다. 물론 이 가정은 현실에서 잘 맞지 않을 수도 있다.

2.3.1. 평균이 다른 이월대체법을 이용한 총계 추정량 2.1절에서 설명한 기존의 이월대체법은 두 시점의 평균이 같은 경우에 해당하는 방법이다. 즉 자료가 확률보행과정인 다음의 모형을 따를 때 타당한 방법이다.

$$y_{i,t} = y_{i,t-1} + a_{i,t}, \quad a_{i,t} \sim iid(0, \sigma_a^2). \tag{2.5}$$

그러나 본 연구에서는 경향이 있는 확률보행과정을 자료 생성 모형으로 고려하였기 때문에 평균이 다른 이월대체법의 총계 추정량을 사용하게 된다. 이제 모형 (2.4)에 의한 대체값은

$$\hat{y}_{i,t}^m = \left(\bar{y}_t^O - \bar{y}_{t-1}^O\right) + y_{i,t-1}^m \tag{2.6}$$

이고 $\bar{y}_t^O = 1/r \sum_{i=1}^r y_{i,t}^O$, $\bar{y}_{t-1}^O = 1/r \sum_{i=1}^r y_{i,t-1}^O$이므로 $\hat{t}^c$는 다음과 같이 얻어진다.

$$\hat{t}^c = w\left(\sum_{i=1}^r y_{i,t}^O + (n-r)\left(\bar{y}_t^O - \bar{y}_{t-1}^O\right) + \sum_{i=r+1}^n y_{i,t-1}^m\right). \tag{2.7}$$

표현을 간단히 하기 위해 $S_1 = \sum_{i=1}^r y_{i,t}^O$, $S_2 = \sum_{i=1}^r y_{i,t-1}^O$, $S_3 = \sum_{i=r+1}^n y_{i,t-1}^m$이라 하자. 그러면

$$\hat{t}^c = w\left(S_1 + (n-r)\left(\bar{y}_t^O - \bar{y}_{t-1}^O\right) + S_3\right) = w\left(S_1 + \frac{(n-r)}{r}(S_1 - S_2) + S_3\right) \tag{2.8}$$

가 된다.

2.3.2. BLS 무응답 가중치 보정법을 이용한 총계 추정량 식 (2.3)의 결과를 이용하면 BLS 무응답 보정법의 총계 추정량은 다음과 같다

$$\hat{t}^B = w\left(\frac{\sum_{i=1}^n y_{i,t-1}}{\sum_{i=1}^r y_{i,t-1}^O}\right)\sum_{i=1}^r y_{i,t}^O = w\left(\frac{\sum_{i=1}^r y_{i,t-1}^O + \sum_{i=r+1}^n y_{i,t-1}^m}{\sum_{i=1}^r y_{i,t-1}^O}\right)\sum_{i=1}^r y_{i,t}^O$$

$$= w\sum_{i=1}^{r} y_{i,t}^{O} + w\left(\frac{\sum_{i=r+1}^{n} y_{i,t-1}^{m}}{\sum_{i=1}^{r} y_{i,t-1}^{O}}\right)\sum_{i=1}^{r} y_{i,t}^{O} = w\sum_{i=1}^{r} y_{i,t}^{O} + w\left(\frac{\sum_{i=1}^{r} y_{i,t}^{O}}{\sum_{i=1}^{r} y_{i,t-1}^{O}}\right)\sum_{i=r+1}^{n} y_{i,t-1}^{m}$$

$$= w\sum_{i=1}^{r} y_{i,t}^{O} + w\frac{\bar{y}_t^{O}}{\bar{y}_{t-1}^{O}} \times \sum_{i=1}^{n-r} y_{i,t-1}^{m}. \tag{2.9}$$

따라서 BLS 무응답 보정법은 결측치의 대체값으로

$$\hat{y}_{i,t}^{m} = \frac{\bar{y}_t^{O}}{\bar{y}_{t-1}^{O}} \times y_{i,t-1}^{m} \tag{2.10}$$

을 사용한 결과가 되므로 BLS 결측치 대체법은 비추정(ratio estimation) 대체법과 같은 형태의 대체값으로 결측치를 대체하는 방법으로 정의할 수 있다. 이제 식 (2.10)을 (2.8)과 같은 형태로 만들면 다음과 같다.

$$\hat{t}^{B} = w\left(S_1 + \frac{S_1}{S_2}S_3\right). \tag{2.11}$$

2.4. 총계 추정량의 특징

이 절에서는 이월대체법과 BLS 무응답 보정법에서 얻어진 총계 추정량의 특징을 살펴보았다. 즉 모형 (2.4) 하에서 얻어진 $\hat{t}^{B}$의 기댓값과 분산을 구하였으며 $\hat{t}^{c}$의 기댓값을 유도하였다. 또한 $\hat{t}^{B}$와 $\hat{t}^{c}$의 관계도 구하였다. 이를 위하여 먼저 모형 (2.4)와 이에 주어진 가정 하에서 S_1, S_2, S_3에 관한 결과를 살펴보면 다음과 같다.

$$E(S_1) = r\mu_t, \qquad E(S_2) = r\mu_{t-1}, \qquad E(S_3) = (n-r)\mu_{t-1}, \tag{2.12}$$

$$\mathrm{Var}(S_1) = r(\sigma_{t-1}^2 + \sigma_a^2), \qquad \mathrm{Var}(S_2) = r\sigma_{t-1}^2, \qquad \mathrm{Var}(S_3) = (n-r)\sigma_{t-1}^2, \tag{2.13}$$

$$E\left(\frac{1}{S_2}\right) = E\left(\frac{1}{\sum_{i=1}^{r} y_{i,t-1}^{O}}\right) \approx \frac{1}{r}\frac{1}{\mu_{t-1}}\left(1 + \frac{1}{r}\frac{\sigma_{t-1}^2}{\mu_{t-1}^2}\right). \tag{2.14}$$

식 (2.14)는 $E(y) = \mu$라 할 때 정석오와 신기일 (2008)에서 얻어진

$$E\left(y^{-k}\right) \approx \mu^{-k}\left(1 + \frac{1}{2}k(k+1)\frac{\sigma_{t-1}^2}{\mu^2}\right)$$

를 이용한 결과이다. 정리 2.1과 2.2는 모형 (2.4)와 가정 1, 2하에서 $\hat{t}^{c} = w(S_1 + (n-r)(\bar{y}_t^{O} - \bar{y}_{t-1}^{O}) + S_3)$의 기댓값과 분산에 관한 결과이다.

정리 2.1 모형 (2.4)와 가정 1, 2하에서 이월대체법으로 구해진 $\hat{t}^{c} = w(S_1 + (n-r)(\bar{y}_t^{O} - \bar{y}_{t-1}^{O}) + S_3)$는 불편 추정량이다. 즉 $E(\hat{t}^{c}) = wn\mu_t$이다.

증명: 먼저 $E(S_1) = r\mu_t$, $E(S_2) = r\mu_{t-1}$, $E(S_3) = (n-r)\mu_{t-1}$이고 $E(\bar{y}_t^{O}) = 1/rE(S_1) = \mu_t$, $E(\bar{y}_{t-1}^{O}) = 1/rE(S_2) = \mu_{t-1}$이므로 $E(\hat{t}^{c}) = wn\mu_t$가 되어 불편 추정량이다. □

정리 2.2 모형 (2.4)와 가정 1, 2하에서 이월대체법으로 구해진 $\hat{t}^{c} = w(S_1 + (n-r)(\bar{y}_t^{O} - \bar{y}_{t-1}^{O}) + S_3)$의 분산은 다음과 같다.

$$\mathrm{Var}\left(\hat{t}^{c}\right) = w^2\left\{n\left(\sigma_{t-1}^2 + \sigma_a^2\right) + \frac{n}{r}(n-r)\sigma_a^2\right\}.$$

증명: (S_1, S_2)와 S_3는 독립이므로, 식 (2.7) $\mathrm{Var}(\hat{t}^c) = \mathrm{Var}(n/rS_1 - (n-r)/rS_2 + S_3)$은 다음과 같다.

$$\frac{n^2}{r^2}\mathrm{Var}(S_1) + \frac{(n-r)^2}{r^2}\mathrm{Var}(S_2) + \mathrm{Var}(S_3) - \frac{2n(n-r)}{r^2}\mathrm{Cov}(S_1, S_2),$$

여기서 $\mathrm{Cov}(S_1, S_2) = \sum_i \mathrm{Cov}(y_{i,t}^O,\, y_{i,t-1}^O) = r\sigma_{t-1}^2$이므로 이를 대입하면 다음의 결과를 얻는다.

$$\mathrm{Var}\left(\hat{t}^c\right) = w^2 n\sigma_{t-1}^2 + w^2\frac{n^2}{r}\sigma_a^2 = w^2 n\left(\sigma_{t-1}^2 + \sigma_a^2\right) + w^2\frac{n}{r}(n-r)\sigma_a^2.$$

□

정리 2.1과 2.2를 살펴보면 경향이 있는 확률보행과정 하에서 평균이 다른 이월대체법은 불편 추정량이지만 분산은 결측치가 없는 경우에 비해 과대 추정되는 것을 확인 할 수 있다. 만약 $\mu_{t-1} = \mu_t$인 경우를 가정한다면 즉 일반 확률보행과정 하에서는 $\hat{t}^c = w(S_1 + S_3)$가 되고 이 경우 $\hat{t}^c$는 정리 2.1에 의해 불편 추정량이 된다. 그러나 $\mathrm{Var}(\hat{t}^c) = w^2\{r(\sigma_{t-1}^2 + \sigma_a^2) + (n-r)\sigma_{t-1}^2\} = w^2(n\sigma_{t-1}^2 + r\sigma_a^2)$이므로 결측치가 없는 경우에 비해 분산이 과소 추정되고 있음을 알 수 있다. 다음으로 정리 2.3은 모형 (2.4)와 가정 1, 2하에서 $\hat{t}^B$의 기댓값에 관한 내용이다.

정리 2.3 모형 (2.4)와 가정 1, 2하에서 BLS 무응답 보정법의 총계 추정량, $\hat{t}^B = w(S_1 + S_1/S_2\, S_3)$의 기댓값은 다음과 같다.

$$E\left(\hat{t}^B\right) \approx wn\mu_t + w(n-r)\left(\mu_t - \mu_{t-1}\right)\frac{\sigma_{t-1}^2}{r\mu_{t-1}^2}.$$

증명: 먼저 (S_1, S_2)와 S_3는 독립이므로 식 (2.13)에 의해

$$\begin{aligned} E\left(\hat{t}^B\right) &= wE\left(S_1 + \frac{S_1}{S_2}S_3\right) = wE(S_1) + wE\left(\frac{S_1}{S_2}\right)E(S_3) \\ &= rw\mu_t + (n-r)w\mu_{t-1}E\left(\frac{S_1}{S_2}\right) \end{aligned} \tag{2.15}$$

이 된다. 여기서 $E(S_1/S_2) = E[(\sum_{i=1}^r(\mu_t - \mu_{t-1}) + S_2 + \sum_{i=1}^r a_{i,t}){S_2}^{-1}] = 1 + r(\mu_t - \mu_{t-1})E(1/S_2) + E((\sum_{i=1}^r a_{i,t})S_2^{-1})$고 $\sum_{i=1}^r a_{i,t}$와 S_2는 독립이므로 $E((\sum_{i=1}^r a_{i,t})S_2^{-1}) = E(\sum_{i=1}^r a_{i,t})E(1/S_2)$이 된다. 따라서 식 (2.14) 결과를 대입하면 $E(S_1/S_2) \approx 1 + (\mu_t - \mu_{t-1})/\mu_{t-1}(1 + 1/r\,\sigma_{t-1}^2/\mu_{t-1}^2)$이 성립한다. 이 결과를 식 (2.15)에 대입하면 다음의 결과를 얻는다.

$$\begin{aligned} E\left(\hat{t}^B\right) &\approx wr\mu_t + w(n-r)\mu_{t-1}\left(1 + \frac{(\mu_t - \mu_{t-1})}{\mu_{t-1}}\left(1 + \frac{\sigma_{t-1}^2}{r\mu_{t-1}^2}\right)\right) \\ &= wr\mu_t + w(n-r)\mu_t + w(n-r)(\mu_t - \mu_{t-1})\frac{\sigma_{t-1}^2}{r\mu_{t-1}^2} \\ &= wr\mu_t + w(n-r)(\mu_t - \mu_{t-1})\frac{\sigma_{t-1}^2}{r\mu_{t-1}^2}. \end{aligned}$$

□

따라서 모형 (2.4) 하에서 $E(\hat{t}^B)$는 불편 추정량이 아니다. 그러나 일반적인 자료에서는 $\sigma_{t-1}^2/(r\mu_{t-1}^2) \approx 0$이 되므로 이 조건하에서는 근사적으로 불편 추정량이 된다. 물론 평균이 동일한 경우에는 불편 추정량이 된다. 다음의 정리 2.4는 $\hat{t}^c$와 $\hat{t}^B$의 관계에 관한 내용이다.

표 **3.1.** $n = 500$인 경우의 MSE

결측비율	μ_{t-1}	θ	방법	$\sigma^2_{t-1} = 9$			$\sigma^2_{t-1} = 25$		
				$\sigma^2_a = 1$	$\sigma^2_a = 1.5$	$\sigma^2_a = 2$	$\sigma^2_a = 1$	$\sigma^2_a = 1.5$	$\sigma^2_a = 2$
5%	100	0	t^B	26.39	39.55	53.22	26.41	37.97	53.97
			t^c	26.39	39.55	53.22	26.40	37.96	53.97
		10	t^B	29.16	40.05	53.68	32.06	44.67	65.04
			t^c	26.24	38.15	51.48	26.07	37.71	55.94
	200	0	t^B	24.65	39.81	50.81	23.72	38.64	52.79
			t^c	24.65	39.81	50.81	23.72	38.65	52.79
		10	t^B	27.37	38.97	55.81	27.71	37.95	56.56
			t^c	26.61	38.21	56.04	26.02	36.73	54.88
20%	100	0	t^B	136.77	179.88	248.34	133.36	186.28	248.06
			t^c	136.77	179.87	248.32	133.36	186.28	248.10
		10	t^B	135.61	208.96	257.25	157.04	214.77	312.86
			t^c	123.26	198.23	244.68	122.96	178.75	264.55
	200	0	t^B	128.15	193.33	241.37	124.21	187.33	260.74
			t^c	128.15	193.32	241.40	124.23	187.33	260.71
		10	t^B	125.44	189.06	254.74	129.34	188.23	238.08
			t^c	123.11	186.45	250.68	122.04	179.42	231.26

정리 2.4 모형 (2.4)와 가정 1, 2하에서, $\bar{y}^m_{t-1}/\bar{y}^O_{t-1} \approx 10$이면 $\hat{t}^c \approx \hat{t}^B$가 된다.

증명: $S_1 + S_1/S_2\, S_3 = S_1 + S_3 + (S_1 - S_2)/S_2\, S_3$이다. 여기서 $(S_1 - S_2)S_2\, S_3 = (\bar{y}^O_t - \bar{y}^O_{t-1})(n - r)\bar{y}^m_{t-1}/\bar{y}^O_{t-1}$이 되므로 $\bar{y}^m_{t-1}/\bar{y}^O_{t-1} \approx 1$인 조건하에서는 $(S_1 - S_2)/S_2\, S_3 \approx (n - r)(\bar{y}^O_t - \bar{y}^O_{t-1})$이 된다. 따라서

$$\hat{t}^B = w\left(S_1 + \frac{S_1}{S_2}S_3\right) \approx w\left(S_1 + (n - r)\left(\bar{y}^O_t - \bar{y}^O_{t-1}\right) + S_3\right) = \hat{t}^c.$$

□

이상을 종합해 보면 두 모집단의 평균값에 상관없이 $\hat{t}^c$는 불편 추정량이나 $\hat{t}^B$는 평균이 같은 경우에만 불편 추정량이 된다. 또한 $\hat{t}^c$의 분산은 평균값에 따라 과대 추정량 또는 과소 추정량이 되는 것을 확인할 수 있으며 $\bar{y}^m_{t-1}/\bar{y}^O_{t-1} \approx 1$인 경우 두 추정량은 근사적으로 같은 추정량이 된다.

3. 모의실험 및 자료 분석

3.1. 모의실험

2절에서 얻어진 결과를 확인하기 위해 간단한 모의실험을 실시하였다. 먼저 $t-1$ 시점의 자료 n개는 정규분포를 이용하여 생성하였다. 층의 크기에 따른 영향을 살펴보기 위해 자료 수 $n = 500, 1000$을 사용하였다. 또한 $\mu_{t-1} = 100, 200$, $\sigma^2_{t-1} = 9, 25$을 사용하였다. 다음으로 t 시점의 자료를 얻기 위해 모형 (2.4), $y_{i,t} - \mu_t = y_{i,t-1} - \mu_{t-1} + a_{i,t}$을 이용하였다. 여기서 $\mu_t = \mu_{t-1} + \theta$에서 $\theta = 0, 10$을 이용하였으며 백색잡음과정 $a_{i,t} \sim N(0, \sigma^2_a)$에서 σ^2_a의 영향을 보기 위해 $\sigma^2_a = 1, 1.5, 2$를 사용하였다. 또한 결측비율을 5%와 20%로 하였으며 반복은 1,000번을 실시하였다. 비교를 위해 사용된 통계량은 MSE와 편향(Bias)이며 MSE 결과는 표 3.1과 3.2에 나와 있다. 모의실험 결과 $\theta = 0$인 경우 두 추정 결과가 일치

표 **3.2.** $n = 1000$인 경우의 MSE

결측비율	μ_{t-1}	θ	방법	$\sigma^2_{t-1}=9$			$\sigma^2_{t-1}=25$		
				$\sigma^2_a=1$	$\sigma^2_a=1.5$	$\sigma^2_a=2$	$\sigma^2_a=1$	$\sigma^2_a=1.5$	$\sigma^2_a=2$
5%	100	0	$t^B=t^c$	53.48	85.86	108.83	53.49	83.13	107.59
		10	t^B	59.45	80.81	106.83	64.83	85.29	124.63
			t^c	54.43	77.37	103.62	50.01	73.51	106.04
	200	0	$t^B=t^c$	50.07	79.07	102.50	53.54	78.80	109.61
		10	t^B	53.07	87.67	98.85	56.61	81.91	113.69
			t^c	51.71	86.12	98.03	53.94	78.27	111.87
20%	100	0	$t^B=t^c$	250.10	418.56	485.05	260.85	398.13	547.91
		10	t^B	298.69	356.33	536.24	284.85	431.20	598.78
			t^c	267.84	327.86	513.02	222.13	368.04	548.13
	200	0	$t^B=t^c$	243.80	375.72	496.87	265.50	394.41	507.16
		10	t^B	256.02	378.03	515.80	276.20	386.89	528.29
			t^c	249.13	375.65	512.44	261.76	371.40	507.86

표 **3.3.** $n = 500$인 경우의 Bias

결측비율	μ_{t-1}	θ	방법	$\sigma^2_{t-1}=9$			$\sigma^2_{t-1}=25$		
				$\sigma^2_a=1$	$\sigma^2_a=1.5$	$\sigma^2_a=2$	$\sigma^2_a=1$	$\sigma^2_a=1.5$	$\sigma^2_a=2$
5%	100	0	$t^B=t^c$	−0.25	0.28	−0.38	−0.16	−0.14	0.02
		10	t^B	−0.06	−0.34	−0.07	0.15	0.03	0.28
			t^c	−0.03	−0.32	−0.09	0.17	0.11	0.18
	200	0	$t^B=t^c$	0.42	0.36	0.17	0.14	0.40	0.03
		10	t^B	0.04	−0.14	−0.05	0.09	−0.09	0.17
			t^c	0.01	−0.14	−0.04	0.10	−0.12	0.18
20%	100	0	$t^B=t^c$	1.07	−0.22	0.49	0.71	0.10	0.15
		10	t^B	−0.11	−1.19	0.25	0.47	0.60	−0.67
			t^c	−0.00	−1.27	0.43	0.37	0.46	−0.49
	200	0	$t^B=t^c$	−0.02	0.31	0.13	−0.18	0.55	0.41
		10	t^B	−0.07	−0.01	0.12	−0.00	−0.30	0.24
			t^c	−0.06	0.02	0.12	−0.04	−0.30	0.26

하여 표 3.2에는 이 결과를 생략하였다. 편향 결과는 표 3.3과 3.4에 나와 있으며 같은 이유로 $\theta = 0$인 결과는 생략하였다. 또한 응답자 자료와 간단한 무응답 비율 보정 가중치를 적용한 결과는 이미 김석과 신기일 (2009)에 나와 있기 때문에 본 모의실험에서는 이를 생략하였다.

표 3.1과 3.2의 MSE 결과를 살펴보면 모든 결과에서 t^c가 우수한 것을 확인 할 수 있다. 이는 자료가 경향이 있는 확률보행과정을 따르도록 생성하였기 때문으로, 예상된 결과이다. 여기서 두 시점간의 평균이 같은 경우, 즉 일반 이월대체법을 사용한 경우에는 모든 표에서 같은 결과를 주고 있음을 확인 할 수 있다. 다음으로 표 3.1 결과에서 MSE의 비율을 살펴보면 어떤 요인이 결과에 많은 영향을 주는 지 파악할 수 있다. 먼저 MSE의 결과에 가장 많은 영향을 주는 것은 $\theta = \mu_t - \mu_{t-1}$이다. 즉 두 시점간의 평균 차이가 클수록 두 방법의 MSE 결과에 크게 영향을 주고 있다는 것을 알 수 있다. 다음으로 MSE 결과 차이에 영향을 주는 것은 μ_{t-1}의 크기이다. $\mu_{t-1} = 100$인 경우에 비해 $\mu_{t-1} = 200$인 경우 결과의 차이가 감소하는 경향을 보이고 있다. 다음으로 σ^2_{t-1}이 영향을 주고 있으며 또한 σ^2_a이 영향을 주고 있다. 이러한 결과는 정리 2.3과 2.4의 내용과 일치하는 것이며 표 3.2에서도 같은 내용을 확인 할 수 있다. 편향 결과인 표 3.3과 3.4를 살펴보면 두 추정량 모두 매우 작은 편향을 갖고 있음을 확인 할 수 있

표 **3.4.** $n = 1000$인 경우의 Bias

결측비율	μ_{t-1}	θ	방법	$\sigma^2_{t-1} = 9$			$\sigma^2_{t-1} = 25$		
				$\sigma^2_a = 1$	$\sigma^2_a = 1.5$	$\sigma^2_a = 2$	$\sigma^2_a = 1$	$\sigma^2_a = 1.5$	$\sigma^2_a = 2$
5%	100	0	$t^B = t^c$	−0.71	−0.12	0.07	0.27	−0.22	−0.12
		10	t^B	−0.23	−0.00	−0.16	0.02	−0.35	−0.09
			t^c	−0.17	−0.03	−0.15	0.01	−0.40	−0.04
	200	0	$t^B = t^c$	0.11	−0.23	−0.28	0.10	−0.45	−0.50
		10	t^B	0.16	0.27	−0.22	−0.17	0.18	0.32
			t^c	0.12	0.22	−0.18	−0.14	0.18	0.21
20%	100	0	$t^B = t^c$	−0.03	−0.88	0.64	−0.66	0.62	−0.68
		10	t^B	−1.27	0.45	0.24	−0.18	−0.85	0.53
			t^c	−1.03	0.60	0.48	−0.11	−0.65	0.45
	200	0	$t^B = t^c$	−0.29	0.63	0.09	0.70	0.50	−0.29
		10	t^B	−0.83	−0.62	0.03	−0.90	0.17	1.45
			t^c	−0.73	−0.55	0.09	−0.76	0.29	1.38

표 **3.5.** 층별 표준편차, 상관계수 및 기울기

층	모집단 수	표준편차		상관계수	기울기
		3월	5월		
C1	1251	3842.62	3536.63	0.894	0.82
C2	983	81727.47	79936.04	0.943	0.92
C3	397	1749711.58	1521347.64	0.945	0.822

으며 편향을 기준으로 두 방법의 우수성을 평가하기는 어렵다.

3.2. 자료 분석

실제 자료 분석에서는 세 가지 방법이 사용되었다. 먼저 횡시점회귀대체법이 사용되었으며 전술한 바와 같이 $t-1$ 시점의 자료를 독립변수로, t 시점의 자료를 종속변수로 하는 단순회귀모형을 사용하여 결측치를 대체하였다. 이 방법에서 얻어진 총계 추정량을 $\hat{t}^R$이라 표시하였다. 다음으로 횡시점회귀대체법의 특수한 경우인 평균이 다른 이월대체법이 사용되었다. 이 경우의 총계 추정량은 $\hat{t}^c$이다. 세 번째는 BLS 방법에서 얻어진 총계 추정량으로 $\hat{t}^B$이다. 자료 분석에 사용된 자료는 2007년 3월과 5월의 매월노동통계 자료이다. 이 자료는 약 7,000개로 이루어졌으며 규모별 임금 총액을 구하는 것이 목적이다. 일반적으로 월 평균 임금을 구하기 위해서는 규모별 임금 총액이 구해져야 하기 때문에 규모별 임금 총액은 매우 중요한 통계이다. 이제 3월 자료를 $t-1$ 시점 자료인 $y_{i,t-1}$로 5월 자료를 t 시점 자료인 $y_{i,t}$로 정하였다. 모집단은 7개 규모로 나누어져 있으며 각 규모가 하나의 층으로 구분 되어있다. 본 연구에서는 이 중에서 다음의 3개 층을 이용하였다.

C1: 5–9인, C2: 100–299인, C3: 500인 이상.

각 층에서 5%, 10% 그리고 20% 자료를 결측 처리하였다. 이제 각 층별 3월 자료와 5월 자료의 표준편차, 상관계수 그리고 기울기를 구한 결과는 표 3.5와 같다.

표 3.5에서 표준편차를 살펴보면 규모가 큰 사업체 일수록 커지는 것을 알 수 있다. 각 층의 표준편차를 보면 3월에 비해 5월의 표준편차가 작아져 경향이 있는 확률보행과정에 주어진 가정과 일치하지 않는다. 이 자료는 3월과 5월에 모두 같은 층에 포함된 자료만을 사용하기 때문에 시점이 증가해도 표준

표 **3.6.** 매월 노동통계 자료를 이용한 비교 통계량 결과(C1층)

비교통계량	추정량	결측비율		
		5%	10%	20%
MSE	t^B	170885377.57	399088289.52	870866196.74
	t^c	185831418.78	432416854.39	943545220.01
	t^R	156679174.70	366218446.41	799677209.41
Bias	t^B	−351.988	−324.889	32.427
	t^c	−234.474	−484.567	225.127
	t^R	−611.676	158.494	−248.604
RMSE	t^B	.000009095	.000021241	.000046351
	t^c	.000009891	.000023015	.000050219
	t^R	.000008339	.000019492	.000042562
ARE	t^B	.002375250	.003646774	.005450662
	t^c	.002481509	.003772360	.005681530
	t^R	.002278001	.003523706	.005201345

편차가 크게 될 가능성은 많지가 않다. 각 층별로 얻어진 상관계수는 상대적으로 높은 편이며 기울기를 살펴보면 모두 “1”보다 작게 나타났기 때문에 이월대체법을 사용하기에는 무리 일 수 있다. 결과적으로 이월대체법 보다는 횡시점회귀대체법과 BLS 대체법이 우수한 결과를 줄 것으로 예상된다.

각 층별 결과를 비교하기 위해 사용된 비교 통계량은 MSE, 편향(Bias)에 추가하여 상대평균제곱오차(RMSE) 그리고 평균절대상대오차(ARE)를 사용하였다. 여기서 상대평균제곱오차(RMSE)와 평균절대상대오차(ARE)의 정의는 다음과 같다.

$$\text{RMSE} = \frac{1}{1000}\sum_{k}^{1000}\left(\frac{t_k - \hat{t}_k}{t_k}\right)^2,$$

$$\text{ARE} = \frac{1}{1000}\sum_{k}^{1000}\left|\frac{t_k - \hat{t}_k}{t_k}\right|,$$

여기서 t_k는 모집단 총계이고, $\hat{t}_k$는 k번째 반복에서 구해진 총계 추정치이다. 다음의 표 3.6에서 표 3.8에 세 가지 대체법에 관한 비교 결과를 나타내었다.

표 3.6의 결과를 살펴보면 횡시점회귀대체 결과가 우수한 것을 확인 할 수 있다. 또한 표 3.7과 표 3.8의 결과를 살펴보면 횡시점회귀대체 추정량 t^R과 BLS 추정량인 t^B의 결과가 매우 유사하며 이월대체 추정량인 t^c 보다는 우수한 결과를 확인할 수 있다. 이러한 결과는 예상된 결과이며 따라서 기울기가 “1”보다 작은 경우에는 이월대체법 보다는 다른 두 방법을 사용하는 것이 타당하다고 하겠다.

4. 토의 및 결론

총계 추정에 있어서 BLS 무응답 보정법은 가중치를 보정하여 무응답의 영향력을 줄이는 방법이고 횡시점회귀대체법 또는 이월대체법은 무응답으로 인해 발생한 결측치에 값을 대체하여 무응답의 영향을 줄이는 방법이다. 따라서 두 방법은 목적은 같지만 근본적으로 다른 개념을 갖고 있다. 본 논문에서는 BLS 무응답 보정법이 비추정 대체법 형태의 대체법과 같음을 보였다. 이 결과를 이용하여 경향이 있는 확률 모형을 가정했을 때 평균이 다른 이월대체법과 근사적으로 같은 결과를 주는 것을 확인 하였다. 모의실험 결과 두 시점의 평균이 같은 경우 이월대체법과 BLS 무응답 보정법의 결과는 완전히 일치하는 것을 확인하였다. 자료 분석 결과를 살펴보면 기울기가 “1”보다 작고 평균이 다른 경우 평균이 다른 이

표 **3.7.** 매월 노동통계 자료를 이용한 비교 통계량 결과(C2층)

비교통계량	추정량	결측비율		
		5%	10%	20%
MSE	t^B	38600321811	73146033024	174398527373
	t^c	39364707299	74447539934	177355165865
	t^R	38245284435	73308396850	174096690369
Bias	t^B	1080.02	253.044	−3927.16
	t^c	919.75	467.276	−2956.82
	t^R	1848.22	71.044	−5525.16
RMSE	t^B	.000005747	.000010890	.000025965
	t^c	.000005861	.000011084	.000026405
	t^R	.000005694	.000010914	.000025920
ARE	t^B	.001858141	.002601703	.004093636
	t^c	.001876488	.002623541	.004132832
	t^R	.001852782	.002607792	.004070534

표 **3.8.** 매월 노동통계 자료를 이용한 비교 통계량 결과(C3층)

비교통계량	추정량	결측비율		
		5%	10%	20%
MSE	t^B	5.4596E12	1.1676E13	2.6289E13
	t^c	6.7360E12	1.4694E13	3.4084E13
	t^R	5.6632E12	1.1839E13	2.6149E13
Bias	t^B	80551.48	−123007.72	−140214.30
	t^c	116508.50	−139116.42	−100380.11
	t^R	69643.74	−86929.48	−134539.79
RMSE	t^B	.000049790	.000106487	.000239747
	t^c	.000061431	.000134007	.000310834
	t^R	.000051647	.000107971	.000238471
ARE	t^B	.005226097	.008089779	.012344
	t^c	.006150947	.009300599	.014279
	t^R	.005201810	.008033131	.012204

월대체법을 사용하기 보다는 횡시점회귀대체법 또는 BLS 대체법을 사용하는 것이 더 좋은 결과를 주고 있음을 확인 할 수 있었다.

이론적으로는 각 시점에서 얻어진 패널 자료는 시계열 자료이므로 시계열 자료 분석을 사용하는 것이 타당하다. 시계열 자료는 평균이 일정하고 분산이 일정한 정상시계열과 시점에 따라 평균 또는 분산이 달라지는 비정상시계열이 있다. 비정상시계열인 경우 평균이 같은 비정상시계열 또는 경향이 있는 비정상시계열로 나누어진다. 그러나 일반적인 자료에서는 평균과 분산이 다르고 또한 비정상 시계열인 경우 t 시점의 분산이 $t-1$ 시점에 비해 작을 수 있기 때문에 시계열 모형을 고집하는 것은 문제가 될 수 있다. 따라서 패널 자료를 시계열 자료로 보느냐 아니면 회귀분석 자료로 보느냐는 매우 중요한 문제가 될 수 있다.

참고문헌

김동욱, 노영화 (2003). 대체방법별 GEE추정량 비교, <응용통계연구>, **16**, 407–426.

김석, 신기일 (2009). 상관관계와 표본 크기에 따른 BLS 무응답 보정의 효율성 비교, <응용통계연구>, **22**, 1301–1313.

신민웅, 이상은 (2001). <표본조사를 위한 표본설계>, 교우사.

이상은 (2008). 표본조사에 따른 추정방법비교: 가중치조정기법을 중심으로, <응용통계연구>, **21**, 413–427.

정석오, 신기일 (2008). 평균제곱상대오차에 기반한 비모수적 예측, 한국통계학회논문집, **15**, 255–264.

Lepkowski, J. M. (1989). *Treatment of Wave Nonresponse in Panel Survey*, in Panel Survey(D. Kasprzyk, ed.), John Wiley & Sons, 348–374.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, John Wiley & Sons, New Jersey.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons, New York.

Särndal, C. E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*, Springer.

Tremblay, A. (1994). *Longitudinal imputation of SIPP food stamp benefit, The survey of income and program participation*, Working paper No. 208, U.S. Department of Commerce Bureau of the Census, URL: http://www.census.gov/sipp/

A Comparison of BLS Non-Response Adjustment and Cross-Wave Regression Imputation Methods

Sang Eun Lee[1] · Key-Il Shin[2]

[1]Department of Applied Statistics, Kyonggi University
[2]Department of Statistics, Hankuk University of Foreign Studies

(Received May 2010; accepted July 2010)

Abstract

Cross-wave regression imputation and carry-over imputation method are generally used in the analysis of panel data with missing values. Recently it is known that the BLS non-response adjust method has good statistical properties. In this paper we show that the BLS method can be considered as an imputation method with a similar formula of a ratio-estimator. In addition, we show that the carry-over imputation and BLS imputation are approximately the same under the assumption that data follow a non-stationary process with drift. Small simulation studies and real data analysis are performed. For the real data analysis, a monthly labor statistic (2007) is used.

Keywords: Panel data, missing value, cross-wave regression imputation, carry-over imputation, BLS non-response adjust method.

This work was supported by the Korea Research Foundation Grant funded by the Korean Government(MOEHRD, Basic Research Promotion Fund)(KRF-2008-313-C00141).

[2]Corresponding author: Professor, Department of statistics, Hankuk University of Foreign Studies, Yonginsi, Kyonggy 449-791, Korea. E-mail: keyshin@hufs.ac.kr