

제4장

경제활동인구조사의 가중값 조정

정미옥 · 김서영 · 정동명 · 정구현

제1절 서론

1. 연구배경

통계청에서는 국민의 경제활동(취업, 실업, 노동력 등) 특성을 파악할 목적으로 약 32,000 표본가구에 대해 매월 경제활동인구조사를 실시하고 있다. 경제활동인구조사는 2005년부터 모집단 변화의 반영도를 제고하고 응답자의 부담을 경감하기 위해 1/36 연동표본제(rotation sampling system)를 채택하고 있다. 이 연동표본제에 의해 표본으로 선정된 가구는 36개월 동안 조사에 참여하고, 이후 표본에서 완전히 제외된다. 그리고 매월 전체표본의 1/36인 약 900가구가 교체되는 방식으로 수행되며 이에 따라 중복되는 표본의 비율은 전월대비 35/36, 전년 동월대비 24/36이다. 해당 월의 표본은 최초 조사참여시점이 서로 다른 36개의 부표본(sub-sample)으로 나뉘어 지는데 이 표본그룹을 연동그룹(rotation group)이라 한다.

선행연구들에 따르면 패널조사에서는 같은 표본을 반복적으로 조사하기 때문에 측정오차(measurement error)의 일종인 조사 참여기간에 따른 패널편향(panel bias)이 발생한다고 알려져 있다(Lessler, 1992). 마찬가지로 연동표본제도 표본이 순차적으로 교체되는 반복조사 형태이기 때

문에 편향에 대한 문제가 발생한다. 이를 연동그룹편향(rotation group bias) 혹은 조사 참여기간 효과(time-in-survey effect)라고 부른다(이하 ‘연동그룹편향’). 이러한 편향문제가 발생하게 되는 원인은 명확하게 밝혀 지진 않았지만, 반복된 접촉에 따른 응답자와 조사자 간의 상황변화, 연동그룹 간 조사환경의 차이 등이 주요인으로 추측되고 있다.

미국, 캐나다, 호주 등 많은 나라의 노동력조사에서도 다양한 형태의 연동표본제를 채택하고 있으며, 자국의 연동표본 효과를 반영할 수 있는 고유한 추정방법을 개발하여 사용하고 있다. 예를 들면, 미국의 AK복합추정량과 호주의 BLUE-B1복합추정량은 중복된 표본의 과거정보와 연동그룹편향을 고려하고 있는 대표적인 추정방법이다. 반면 캐나다의 MR2복합추정량은 중복된 표본의 과거 조사 자료를 이용하여 추정의 효율을 높이고는 있으나 연동그룹편향은 반영하지 못하고 있다. 이러한 MR2추정량의 경우, 연동그룹편향이 일단 발생하게 되면 시간이 지남에 따라 이 편향이 누적되면서 추정값(estimates)에 심각한 오류를 범할 수 있다고 Fuller(1999)는 지적하였다. 이처럼 연동표본제에 대한 추정방법 연구에서는 중복되는 표본의 정보를 효과적으로 활용하는 방법과 연동그룹편향을 제거하는 문제가 추정값의 신뢰성 측면에서 주요 관심사항이라 하겠다.

2. 연구목적 및 필요성

경제활동인구조사는 2005년 1월부터 연동표본제를 적용하여 2007년 12월 현재 36번째 연동그룹에 대한 조사를 시작하였다. 즉, 연동의 한 주기(cycle)가 지남에 따라 2007년 12월에 비로소 36개의 연동그룹이 모두 생성되었으며 2005년 1월 처음 조사를 시작했던 연동그룹이 36개월 동안의 조사를 마치고 교체되었다. 현 시점에서 연동표본 효과를 충분히 반영할 수 있는 추정량을 개발하기에는 연동그룹들에 대한 정보가 아직은 부족하다. 그러나 한 주기의 연동그룹이 모두 생성되었다는 점은 연동표본 효과에 대해 탐색할 수 있는 기회를 제공한다. 뿐만 아니라 조사현장에서의 경험이나 연동표본 조사결과 분석을 통해서도 연동그룹들 간의 편향 가능성이 인지되어 온 만큼 연동표본효과를 파악하는 것은 중요하다.

연동표본 추정량과 관련하여 임경은(2007)은 AK복합추정량의 적용 가능성을 검토한바 있다. 그러나 AK복합추정량은 연동그룹편향 측면에서 신규그룹과 나머지 그룹 간 차이만을 고려한 방법으로서 우리의 현실을 반영할 수 있는 A, K 값의 결정에 대한 추가 연구가 필요할 것으로 보인다.

2008년 3월 현시점에서 연동표본효과를 정확하게 반영할 수 있는 복합추정량을 단기에 개발하기에는 크게 두 가지 측면에서 한계가 있다. 첫째, 기존의 다양한 복합추정량에 대해 경제활동인구조사에의 적용 가능성을 검토하는 기초연구가 부족하다. 둘째, 연동주기가 이제 한 주기가 완료됨에 따라 연동그룹 효과를 파악하기에는 이용 가능한 자료가 상당히 부족하다. 이에 본 연구에서는 복합추정량 적용에 앞서 연동그룹편향의 발생 여부를 확인하고 이를 제거할 수 있는 방법을 찾아내는 데 초점을 두고자 한다. 이는 경제활동인구조사의 연동그룹편향 조정을 위해 필요한 단계이며 향후 복합추정량 개발을 위한 사전 연구로서의 의미를 갖는다.

3. 연구내용 및 범위

앞에서 언급한 바와 같이 연동표본조사에 적합한 추정량 개발에 앞서 연동그룹별 편향을 찾는 방법으로 본 연구의 범위를 제한하고자 한다. 우선 전월에 이어 해당 월에도 조사가 지속되고 있는 표본그룹과 해당 월에 교체되어 조사를 처음 시작한 표본그룹들 간 응답수준에 차이가 있는지 여부와 이러한 현상이 추정값에 미치는 영향을 파악해야 한다. 특히 우리나라의 조사환경은 주거문화의 특성상 가구의 전입·전출이 빈번하게 발생하는 편으로 잦은 표본교체가 불가피하다. 경제활동인구조사에서는 표본가구의 이주로 인하여 매월 교체되는 표본가구의 수가 연동표본제에 의해 매월 교체되는 가구의 수와 비슷한 수준이다. 따라서 연동그룹편향을 고려할 때 연동교체와 전입교체 모두를 반영하도록 하겠다. 연동그룹편향은 해당 월에 처음 표본에 진입한 그룹이 조사가 진행되는 처음 몇 개월 동안은 꾸준히 나머지 그룹과 특성 추정값에 차이를 보이다가 일정 시점이 지나면서 다소 안정되어가는 경향을 보이고 있다.

본 연구에서는 경제활동인구조사 자료의 실증분석을 통해 표본교체가 추정값에 어떠한 영향을 미치는지 탐색하고, 이를 가중값(weight)을 이용하여 조정하는 방안에 대해 제시하고자 한다. 제2절에서는 연동그룹편향에 대한 개요와 각 국의 추정방법과 특징에 대해 소개한다. 그리고 제3절에서는 경제활동인구조사의 현행 추정방법을 살펴보고, 표본교체 현황을 분석하여 연동교체와 전입교체에 의한 신규표본그룹이 추정값에 반영되는 정도를 줄여줄 수 있는 가중값 조정방법을 제안한다.

제2절 연동그룹편향 조정을 위한 추정방법

1. 연동그룹편향

패널조사는 동일한 표본을 일정한 시간 간격을 두고 반복적으로 측정하여 자료를 얻는 방법으로 같은 표본에 대해 조사가 반복적으로 수행되는 특성으로 인하여 패널편향이 발생한다고 알려져 있다. 응답자가 조사에 반복적으로 참여한 횟수가 응답수준에 영향을 미칠 수 있다는 연구가 Kemsley(1961), Turner(1961), Waksberg and Neter(1965) 등에 의해 수행되었고, Hansen(1955)은 노동력조사에서 이러한 문제가 특히 실업률 추정값에 영향을 미친다고 언급하였다.

패널조사 형태의 일종인 연동표본제에서도 마찬가지로 반복조사에 의한 측정오류가 발생한다. 미국의 상시인구조사(CPS: Current Population Survey)에서의 반복 인터뷰와 관련한 편향 문제는 많은 연구에서 지속적으로 언급되어 왔는데, 특히 Bailar(1975)는 이러한 CPS에서의 패널편향을 연동그룹편향으로 정의하고 연동그룹편향이 추정값에 미치는 영향에 대해 연구하였다. 연동그룹편향은 반복된 인터뷰에서 서로 다른 조사 참여기간에 의한 응답 수준의 차이로 정의된다. 이러한 현상은 명확하지는 않지만, 반복된 접촉에 의한 응답자와 조사자의 상황변화, 연동그룹간의 조사환경 차이 등에 의해 발생될 수 있다. 또한 호주의 노동력조사(LFS: Labour Force Survey)에서는 이러한 편향을 조사 참여기간 효과라고도 부르며, 이 효과가 추정값에 미치는 영향력을 줄여주기 위해 B1이라는 계

수를 과거 자료에 근거하여 찾아내었다.

연동그룹편향은 연동그룹지수(RGI : Rotation Group Index)에 의해 파악될 수 있다. 이 지수는 주어진 연동그룹에서 관심 특성을 갖는 사람 수를 모든 연동그룹에 대해 평균한 값으로 나누고 100을 곱하여 다음과 같이 계산한다.

$$RGI_r = \frac{\text{r번째 연동그룹에서 관심특성을 갖는 사람의 수}}{\text{모든 연동그룹에서 관심특성을 갖는 사람 수의 평균}} \times 100$$

여기서 $r = 1, \dots, R$ 로 연동그룹번호를 나타낸다. 어떤 연동그룹에 대한 지수가 100이 나왔다면 해당 연동그룹의 추정값이 전체 추정값과 같음을, 100보다 크다면 해당 연동그룹에서 전체보다 크게 추정하고 있음을 의미한다.

Bailar(1975)에 따르면, 미국 CPS의 1968~1969년도의 자료를 대상으로 연동그룹지수를 구했을 때 실업자 수에 대한 연동그룹지수가 최초로 표본으로 들어온 첫 번째 연동그룹에서 120으로 높게 나타났다. 이는 처음 표본으로 들어온 연동그룹에서의 실업자 수가 전체 연동그룹 평균에 비해 약 20%정도 크게 추정된다는 것을 의미한다. 연동그룹별 추정값이 전체에 비해 편향되게 나타나는 이러한 차이를 연동그룹편향으로 설명할 수 있다

연동그룹편향은 가법모형(additive model)과 승법모형(multiplicative model)을 이용하여 측정할 수 있다. 여기서 Y_t 는 추정해야할 t 월의 모수(population parameter)이고, y_t 는 t 월의 비추정량이며, y_{tr} 은 t 월의 r th 연동그룹의 비추정량이라 가정하자.

먼저 연동그룹편향이 다음과 같은 가법형태라고 가정한다면 비추정량(ratio estimator)의 기댓값은 다음과 같다.

$$E(y_{tr}) = \frac{1}{R} Y_t + a_r$$

$$E(y_t) = Y_t + \sum_{r=1}^R a_r$$

$$\cdot a_r = E(y_{tr}) - \frac{1}{R} Y_t : r\text{번째 연동그룹편향}$$

$$\cdot y_t = \sum_{r=1}^R y_{tr}$$

이 경우 어느 한 시점에서의 비추정량 y_t 는 연동그룹편향의 총합이 0이 아닌 경우 연동그룹편향에 의해 편향될 수 있으며 시점 간 수준의 변화 (month-to-month changes)에 대해서는 비편향이다.

만약 연동그룹편향이 승법성격을 지니고 있다면 위와 같은 가법모형을 적용하는 것은 적합하지 않으며 그 대안으로 가장 단순한 형태인 승법모형(multiplicative model)을 Solon(1986)이 다음과 같이 제안하였다.

$$E(y_{tr}) = (1 + b_r) \left(\frac{1}{R} Y_t \right)$$

$$E(y_t) = (1 + \bar{b}) Y_t$$

$$\cdot b_r = \frac{R \cdot E(y_{tr}) - Y_t}{Y_t} : r \text{ 번째 연동그룹편향}$$

$$\cdot \bar{b} = \frac{1}{R} \sum_{r=1}^R b_r$$

승법모형을 가정하였을 때 비추정량 y_t 는 연동그룹편향에 의해 어느 시점의 수준 뿐 아니라 수준의 변화에도 편향을 가져올 수 있다는 것을 이론과 경험적으로 밝혔다.

2. 미국 CPS의 추정방법

미국의 CPS는 미국 전체 및 각 주(state)에 대한 노동력 실태 파악을 목적으로 매월 824개의 PSU(Primary Sampling Unit)로부터 약 72,000가구를 대상으로 실시하는 노동력 통계조사이다. 이 조사의 수행은 가구 조사에 대한 경험이 풍부한 센서스국(Census Bureau)에서, 조사결과와 처리와 분석은 노동통계국(Bureau of Labor Statistics)에서 하고 있다. 또한 CPS에서는 1953년부터 4-8-4 연동표본제를 도입하였는데, 표본가구가 연속되는 4개월 동안 조사에 참여하다가 이후 8개월 동안은 조사를 중단하고 그 다음 4개월 동안 다시 조사에 참여한 후 교체되는 방식으로 수행된다. 이 연동표본제에서는 8개의 연동그룹 중 6개 그룹이 전월에 이어서 조사를 받게 되기 때문에 중복표본비율이 전월대비 75%, 전년 동월대비 50%이다.

CPS에서의 추정과정은 크게 세 단계의 가중값 조정 과정을 거친다. 첫째, 설계가중값으로 단순불편추정값을 구하고 무응답에 대한 조정을 수행한다. 둘째, 이미 알려진 모집단의 성별, 연령별, 인종별, 민족성별¹⁾ 분포 등에 대해 추정값을 일치시키는 비추정단계를 4단계에 걸쳐 실시한다. 셋째, 복합추정단계로 AK복합추정방법을 사용하여 전 단계의 추정값을 개선시킨다. 이때 총계추정을 위한 AK복합추정량은 다음과 같다.

$$Y_t^{AK} = (1 - K) \hat{Y}_t + K(Y_{t-1}^{AK} + \Delta_t) + A\hat{\beta}_t$$

- $\hat{Y}_t = \sum_{i=1}^8 x_{t,i}$
- $\Delta_t = \frac{4}{3} \sum_{i \in S} (x_{t,i} - x_{t-1,i-1})$
- $\hat{\beta}_t = \sum_{i \notin S} x_{t,i} - \frac{1}{3} \sum_{i \in S} x_{t,i}$
- $S = \{2,3,4,6,7,8\}$: 전월에 이어 연속 조사하는 그룹(지속표본)
- $K = 0.4$ (실업자), 0.7 (취업자)
- $A = 0.3$ (실업자), 0.4 (취업자)

AK복합추정량은 해당 월의 비추정값(\hat{Y}_t)과 전월의 AK복합추정값(Y_{t-1}^{AK}) 그리고 지속표본 그룹의 변동 분(Δ_t)이 K 가중값에 의해 가중 평균되고, 신규표본과 지속표본 그룹 간 추정값의 차이가 A 가중값에 의해 반영된다. 즉, 위의 식에서 A 값에 의해 결정되는 마지막 항($\hat{\beta}_t$)이 연동그룹편향에 대한 부분으로서 조사를 처음 참여하거나 중단했다가 다시 참여하는 연동그룹의 실업자 추정값이 다른 그룹에 비해 높게 나타나는 연동그룹편향을 조정하기 위해 추가되었다. 따라서 AK복합추정량은 추정량 자체에 신규그룹과 나머지그룹들 간의 연동그룹편향을 조정할 수 있는 방법이라 할 수 있다.

3. 캐나다 LFS의 추정방법

캐나다의 LFS는 다단계 층화 표본설계에 의해 선택된 54,000가구에

1) 여기서 민족성이란 라틴아메리카(Hispanic) 사람인지 아닌지를 의미한다.

대해 매월 조사를 수행하고 있으며 연동표본제를 실시하고 있다. 캐나다의 연동표본제는 1/6교체 시스템으로 표본으로 선정된 가구는 연속되는 6개월 동안 조사에 참여한 후 교체되는 방식으로 중복표본비율은 전월대비 5/6이다. LFS의 추정 시스템은 2000년도 이전까지는 설계가 중값을 회귀추정방법에 의해 성별, 연령별 지역별로 조정해주는 방식이었다. 이러한 시스템은 당월 자료만을 이용하여 추정하는 방법으로 중복되는 표본의 정보를 탐색하여 추정값의 정도를 향상시키는 방법은 아니었다. 캐나다 통계청에서는 미국에서 사용하는 K복합추정량과 AK복합추정량에 대해 25년 동안 여러 번에 걸쳐 검토하였으나 기존 추정 프로그램에 복합추정방법을 적용하는 것에 매우 큰 압력을 받아 채택하지 않았다. 결국 기존에 있던 일반화회귀추정량(GR : Generalized Regression estimator)의 프로그램을 이용할 수 있는 MR (Modified Regression)² 복합추정량을 Singh(1997)이 제안하여 2001년부터 사용하고 있다.

MR2 복합추정방법은 성별, 연령과 같은 모집단 정보(x)와 아래와 같이 정의되는 제어변수(z)를 이용하여 일반화회귀추정방법을 적용하는 방법으로서 MR2 추정량은 다음과 같다.

$$\hat{y}_t^M = \hat{y}_t^H + ((X_t, Z_t) - (\hat{x}_t^H, \hat{z}_t^H))\beta_t^M$$

$$\cdot \beta_t^M = \left(\sum_i w_{ti}^\pi (x_{ti}, z_{ti})' (x_{ti}, z_{ti}) \right)^{-1} \sum_i w_{ti}^\pi (x_{ti}, z_{ti})' y_{ti}$$

$$\cdot z_{ti} = \begin{cases} \theta_t^{-1} (y_{t-1,i} - y_{ti}) + y_{ti} & \text{if } i \in A_t \\ y_{ti} & \text{if } i \in B_t \end{cases}$$

- A_t : 전월에 이어 연속 조사하는 그룹(지속표본)
- B_t : 처음 조사되는 그룹
- θ_t : t시점에서 중복되는 표본의 비율(5/6)

여기서 제어변수는 지속표본그룹의 전월 추정값을 이용하여 해당 월의 추정값을 개선시키기 위한 변수로 이 변수를 정의하는 것이 바로 MR2 추정방법의 핵심이라고 할 수 있다. MR2 추정량은 지속표본그룹의 추정값에 대한 변동분을 고려했기 때문에, 만약 지속표본과 신규표본 간

성향이 다를 경우 편향이 발생할 수 있는 단점이 있다. 또한 이러한 편향이 일단 발생하게 되면 시간이 지남에 따라 계속 누적되게 되어 편향이 커지게 되는 문제점(drift problem)이 있다. MR2 복합추정량은 연동그룹편향에 대한 조정부분이 반영되지 않은 추정방법으로 이러한 문제점이 지적되고 있다.

4. 호주 LFS의 추정방법

호주에서는 매월 약 30,000의 가구에 대해 LFS를 수행하고 있고, 연동표본제를 실시하고 있다. 호주의 연동표본제는 1/8교체 방식으로 표본 가구는 8개월 동안 조사에 참여한 후 교체된다. 매월 표본의 1/8이 교체되므로 전월대비 중복표본의 비율은 7/8이다. 호주에서는 추정을 위해 당월자료만을 이용하는 일반화회귀추정량을 2007년 이전까지 사용해왔다. 그러나 수년 동안 호주 통계청에서는 이를 대체할 수 있는 추정량을 찾기 위해 다양한 추정방법들에 대한 연구를 수행하였다. Bell(2001)은 AK 복합추정량, MR 복합추정량, 최량선형불편추정량(BLUE : Best Linear Unbiased Estimator)을 LFS자료에 적용·비교해봄으로써 호주자료에 적합한 추정량을 찾으려고 노력했다. 이 연구결과를 근거로 호주에서는 표본오차(sampling error)가 낮은 BLUE-B1추정량을 2007년부터 적용하고 있다.

BLUE-B1 추정량은 1단계에서 연동그룹 수준에서의 BLUE추정량을 구하고 2단계에서는 모집단 보조정보를 이용하여 아래와 같이 일반화회귀추정방법을 적용하고 있다.

$$\begin{aligned} \hat{y}_t^{B1} &= \hat{y}_t^{BH} + (X_t - \hat{x}_t^{BH}) \hat{\beta} \\ \cdot \hat{\beta} &= \left(\sum_{s=t-l}^t \sum_i w_{si}^\pi x_{si}^\# x_{si}^\# \right)^{-1} \sum_{s=t-l}^t \sum_i w_{si}^\pi x_{si}^\# y_{si}^\# \\ \cdot \hat{y}_t^{BH} &= \sum_{s=t-l}^t \sum_i w_{si}^\pi y_{si}^\# \\ \cdot y_{ti}^\# &= a_{tR(t,i)} y_{ti} \end{aligned}$$

여기서 $a_{tR(t,i)}$ 를 B1계수(multiplier)라 부르며, 이 B1계수는 7개월 동안에 걸친 표본의 응답이 해당 월 추정값에 미치는 영향력의 정도를 결정하는 인자이다. B1계수는 연동그룹별로, 해당 월과의 시차 별로 산정되는 값으로 과거 LFS 자료 연동그룹들의 상관행렬구조 모형으로부터 최적의 값을 찾아내었으며, 5년 마다 새로운 값을 찾아낼 예정이다. 따라서 호주의 BLUE-B1추정량은 과거 7개월 자료의 정보와 연동그룹별로 추정값에 미치는 영향력을 조정하는 B1계수를 이용하여 추정량 자체로 연동그룹편향을 조정할 수 있는 방법이라 할 수 있다.

다음의 <표 4-1>은 지금까지 검토한 추정량들의 주요 특성을 요약한 것이다. MR2추정량과 일반화회귀추정량, 비추정량은 연동그룹편향에 대한 조정 부분을 고려하지 않는다. 반면, AK는 전월에 조사되지 않고 금월에 표본으로 포함된 신규그룹에 대한 편향을 조정하고 있다. 또한 BLUE-B1은 신규그룹 뿐 아니라 모든 연동그룹에 대한 편향을 조정해주는 방식을 취하고 있다. 각 국에서는 자국의 연동표본 효과를 최대한 반영해 줄 수 있는 복합추정방법을 개발하여 이용하고 있다. 우리도 향후에는 기존의 방법론에 대한 종합적인 검토와 더불어 연동표본제만의 특성을 반영할 수 있는 복합추정량 개발이 요구된다.

<표 4-1> 각 국의 추정량 특성 비교

내용	GR, ratio	AK	MR2	BLUE-B1
모집단정보 반영	○	○	○	○
월간 변동 반영	×	○	○	○
연동그룹편향조정	×	○	×	○
복잡성	단순함	덜 복잡함	복잡함	복잡함
이용자료	해당 월 자료	해당 월 자료와 이전 월의 가중된 자료	해당 월 자료와 이전 월의 가중된 자료	해당 월 자료와 이전 6개월의 가중된 자료

제3절 경제활동인구조사에서의 연동그룹편향 조정

1. 현행 추정방법

경제활동인구조사의 추정방법은 추계인구를 보조정보로 활용하는 비추정방법(ratio estimation)이다. 설계가중값과 추계인구에 의한 사후층화 가중값의 곱으로 구해지는 응답한 가구원의 가중값을 이용하여 추정하고 있다.

가. 설계가중값(design weight)

설계가중값은 표본설계로부터 직접 구해지는데, 전국에서 표본조사구가 표본으로 추출될 확률의 역수와 각 표본조사구내 가구조사 완료율의 역수를 곱하여 다음과 같이 작성된다. 여기서 조사구내 표본가구 수를 20으로 정했기 때문에 설계가중값은 h지역 내에서는 모든 가구가 동일한 가중값을 갖게 되는 자체가중이 된다.

$$w_{hij}^0 = \frac{S_h}{n_h S_{hi}} \times \frac{M_{hi}}{m_{hi}} = \frac{S_h}{n_h \times 20}$$

- $h = 1, \dots, 25$: 지역 층
- $i = 1, \dots, n_h$: 표본조사구
- $j = 1, \dots, m_{hi}$: 표본가구
- S : 설계당시 가구 수

나. 사후층화가중값(post-stratified weight)

비추정방법은 모집단의 인구총계(demographic totals)와 같은 보조정보를 이용하여 표본으로부터 얻은 추정값을 벤치마킹(benchmarking)하기 위해 사용되며 이를 사후층화 조정이라고도 한다. 경제활동인구조사에서는 인구주택총조사 자료를 보정하여 공식인구로서 사용되고 있는 성별, 연령별 추계인구를 모집단 정보로 사용하였다. 다만 군인 및 재소자 등 현실적으로 가구에서 조사가 불가능한 인구는 추계인구에서 제외시킨다.

$$w_{hijk}^1 = \frac{X_{h,sa}}{\hat{X}_{h,sa}}$$

- $k = 1, \dots, l_{hij}$: 표본가구원
- $X_{h,sa}$: 성별, 연령별 15세 이상 모집단 추계인구 수
- $\hat{X}_{h,sa}$: 표본에서 조사된 성별, 연령별 15세 이상 인구 수

다. 총계(total)추정

경제활동인구조사에서 취업, 실업과 같은 각 특성에 대한 총계 추정 방법은 표본추출 당시의 추출률과 모집단의 보조정보인 추계인구를 활용한 비추정방법이다. 따라서 최종 가중값은 앞에서 살펴보았던 두 가중값의 곱으로 산출되며, 총계 추정량은 다음과 같다.

$$w_{hijk} = w_{hijk}^0 \times w_{hijk}^1$$

$$\hat{Y} = \sum_h \hat{Y}_h$$

$$\hat{Y}_h = \sum_i \sum_j \sum_k w_{hijk} \cdot y_{hijk}$$

또한 총계추정에 대한 분산 추정량은 차분산(successive difference)방법에 의해 다음과 같이 계산된다.

$$Var(\hat{Y}) = \sum_h Var(\hat{Y}_h)$$

$$Var(\hat{Y}_h) = \frac{(1-f_h)n_h}{2(n_h-1)} \sum_i Z_{hi}^2$$

$$\cdot Z_{hi} = dY_{hi} - R_h \cdot dX_{hi}$$

$$\cdot dY_{hi} = Y_{hi} - Y_{hi+1}, \quad dX_{hi} = X_{hi} - X_{hi+1}$$

$$\cdot Y_{hi} = \sum_j \sum_k w_{hijk} \cdot y_{hijk}, \quad X_{hi} = \sum_j \sum_k w_{hijk}$$

$$\cdot R_h = \sum_i \frac{Y_{hi}}{X_{hi}}$$

2. 표본교체 현황 및 특성

경제활동인구조사의 연동표본모형은 1/36교체로서 연동표본제로 인해 매월 교체되는 가구 수는 캐나다의 1/6교체, 호주의 1/8교체에 비하면 적은 편이다. 반면 한 연동그룹이 표본으로 선정되어 조사를 완료한 후 교체되기까지는 36개월이 걸리므로 교체주기는 긴 편이다. 그리고 경제활동인구조사의 표본교체는 연동표본제에 의한 연동교체와 표본가구의 전·출입으로 인해 불가피하게 발생하는 전입교체가 있다. <표 4-2>는 2005년에 발생한 연동교체와 전입교체자의 수를 정리한 것이다. 매월 연동교체자는 약 1,900명, 전입교체자는 약 1,500~2,300명 정도가 발생하고 있다. 즉, 전입에 의한 교체가 연동교체 만큼 매월 발생하고 있으며 이들이 전체표본에서 차지하는 비율은 약 5% 정도이다. 이러한 현상은 2006년과 2007년에서도 유사하게 나타난다. 이에 본 연구는 경제활동인구조사에서의 연동그룹편향을 탐색하기 위해 연동교체와 전입교체를 모두 포함시켜 ‘전입교체’라 정의하고, 이 전입회차그룹에 대한 응답수준의 차이를 확인한다.

<표 4-2> 월별 연동교체 및 전입자 현황

(단위: 명)

년월	지속자	연동교체자	전입교체자	총 교체전입자	전입비율
2005.02.	68,620	1,816	1,570	3,386	4.70%
2005.03.	67,778	1,880	1,858	3,738	5.23%
2005.04.	67,904	1,902	1,568	3,470	4.86%
2005.05.	67,107	1,961	2,054	4,015	5.65%
2005.06.	67,241	1,814	1,851	3,665	5.17%
2005.07.	67,355	1,953	1,880	3,833	5.38%
2005.08.	67,135	1,958	1,693	3,651	5.16%
2005.09.	66,698	1,975	2,323	4,298	6.05%
2005.10.	67,330	1,937	1,778	3,715	5.23%
2005.11.	67,479	1,888	1,752	3,640	5.12%
2005.12.	67,526	1,982	1,774	3,756	5.27%

처음 표본에 포함되어 조사를 시작하는 표본그룹과 조사를 지속해오던 그룹은 서로 다른 응답성향을 보일 수 있다. 이를 확인하기 위해 신규그룹이 최초 조사에서 응답한 내용이 조사 참여기간이 늘어남에 따라 어떻게 변화해 가는지를 살펴보고자 한다. <표 4-3>은 2005년 1월부터 2006년 12월의 자료를 이용하여 조사참여 기간에 따른 응답상태의 전이 과정을 분석한 결과이다. <표 4-3>을 보면, 해당 월에 고용상태를 비경제활동(이하 '비경')으로 응답하였을 때 그 다음 달에도 비경이라 응답한 경우가 해당 월 최초 진입그룹에서 86.2%로 전체평균 89.5%에 비해 다소 낮게 나타났다. 반대로 최초 응답이 비경이었으나 그 다음 회차에서 취업과 실업으로의 상태전이가 일어난 경우는 최초진입그룹에서는 각각 6.1%와 1.4%로 전체평균은 각각 4.2%와 0.8%로 나타났다. 즉, 최초진입그룹이 전체평균에 비해 각각 약 2%p, 0.6%p 높은 편이고, 3회차 이상에서도 같은 현상을 보인다. 이는 조사가 반복적으로 진행되면서 응답자의 고용상태가 바뀌는 현상이 신규그룹에서 전체에 비해 높게 나타나는 것을 뜻한다.

<표 4-3> 조사 참여기간에 따른 최초 응답상태의 전이패턴

구분	전이 상태	2회차	3회차	4회차	5회차	6회차
해당 월 최초진입그룹	비경→비경	86.2%	77.5%	68.2%	67.6%	64.0%
	비경→취업	6.1%	9.3%	10.2%	11.1%	12.1%
	비경→실업	1.4%	1.5%	1.5%	1.6%	1.3%
전체그룹	비경→비경	89.5%	82.0%	74.8%	71.0%	66.7%
	비경→취업	4.2%	6.5%	7.8%	8.4%	8.7%
	비경→실업	0.8%	1.0%	1.0%	1.0%	0.9%
구분	전이 상태	2회차	3회차	4회차	5회차	6회차
해당 월 최초진입그룹	취업→취업	85.9%	80.1%	72.2%	72.9%	70.1%
	취업→실업	0.7%	1.1%	1.1%	1.4%	1.3%
	취업→비경	2.7%	4.2%	4.7%	5.1%	5.6%
전체그룹	취업→취업	91.7%	85.6%	79.2%	75.8%	71.6%
	취업→실업	0.6%	0.9%	1.0%	1.1%	1.1%
	취업→비경	2.6%	3.8%	4.5%	4.8%	5.1%

구분	전이 상태	2회차	3회차	4회차	5회차	6회차
해당 월 최초진입그룹	실업→실업	55.5%	33.8%	23.5%	18.3%	14.7%
	실업→취업	26.7%	37.3%	36.3%	38.9%	40.2%
	실업→비경	9.7%	13.8%	15.4%	16.0%	14.5%
전체그룹	실업→실업	57.1%	38.2%	26.8%	21.1%	17.1%
	실업→취업	24.4%	32.6%	35.2%	36.4%	36.1%
	실업→비경	11.6%	16.1%	18.0%	18.8%	18.7%

이처럼 최초 진입한 그룹에서의 상태전이 비율이 전체그룹과 차이가 나는 이유를 주어진 자료만으로 명확히 파악할 수는 없다. 그러나 다음과 같은 몇 가지 요인들이 작용했을 것이라 생각할 수 있다. 우선 표본 가구원이 처음으로 조사에 참여한 시점에는 조사원과의 신뢰관계가 아직 형성되지 않은 상태이다. 그래서 이 시기에는 사생활 노출을 우려한 응답자가 거짓되거나 부실한 응답을 할 가능성이 높다. 예컨대, 고용상태를 ‘취업’으로 했을 경우 설문 분량이 늘어나게 되는 것을 귀찮게 생각하여 거짓응답을 할 수 있으며, 자신의 직업을 밝히고 싶지 않은 경우 정확한 응답을 회피할 수도 있다. 또한 조사초기에는 조사용어나 개념이 익숙하지 않기 때문에 조사과정에서 혼동이 있었을 수도 있다.

최초진입그룹과 전체그룹과의 응답패턴의 차이를 미루어 보았을 때 경제활동인구조사에서도 조사 참여기간에 따라 응답수준 차이가 발생할 가능성이 높다. 따라서 연동그룹편향에 대한 탐색이 필요하며, 추정 과정에 이러한 연동그룹효과를 반영해 줄 필요가 있다.

3. 가중값 조정방법

이 장에서는 표본교체로 인한 편향을 파악하고, 이를 통해 진입초기 표본의 불안정한 응답패턴이 추정값에 미치는 영향력을 줄일 수 있는 방법을 제안하고자 한다. 그러나 추정방법을 고려함에 있어 다음과 같은 이유로 인하여 연구의 범위가 제한되었다. 첫째, 2005년 1월에 1번 연동그룹 조사를 시작으로 2007년 12월에 36번 연동그룹이 처음 조사를 시작하기 때문에 36개 모든 연동그룹들에 대한 정보가 분석하기에 충분하지

않다. 둘째, 현재 경제활동인구조사에서는 마이크로데이터(micro-data) 이
 용자들이 직접 추정값을 계산해볼 수 있도록 가구원별 가중값을 포함하
 여 자료를 제공하고 있다. 따라서 기존의 제공방법을 유지하기 위해 가
 중값을 이용하여 조정할 수 있는 방법으로 추정방법의 검토범위를 제한
 하였다. 셋째, 추정값의 정도를 높이면서도 실무에 적용하기 편리한 방
 법을 제안하고자 했다. 즉, 계산과정이 간편하고, 기존 추정시스템의 변
 경을 최소화하는 방법을 우선시하였다. 이와 같은 검토기준을 고려하여
 다음과 같은 추정방법을 제안한다.

가. 진입회차그룹별 응답수준의 차이

해당 가구가 표본으로 진입교체되어 얼마동안 조사를 지속해왔는지
 를 전월자료와 연계시켜 비교함으로써 진입회차를 계산하였다. 그리고
 진입회차에 따라 전체표본을 진입회차그룹이라는 36개의 그룹으로 구
 성하였다. 즉, 해당 월에 연동교체나 전입교체로 인하여 표본에 처음
 들어온 그룹은 진입 1회차 그룹이 되고, 전월에 처음 진입하여 해당 월
 에도 존재하는 표본은 진입 2회차 그룹이 된다. 이렇게 구성된 진입회
 차그룹별로 응답수준에 차이가 있는지를 살펴보았다. 2절에서 소개되
 었던 연동그룹지수에 연동그룹 대신 진입회차그룹을 적용시키면 아래
 와 같은 진입회차그룹지수를 구할 수 있다. 이 지수값이 100이면 해당
 진입회차그룹의 추정값이 전체그룹의 추정값과 같다는 것을 의미한다.

$$I_r = \frac{r\text{-th 진입회차그룹에 대한 특성 추정값}}{\text{모든 진입회차그룹들의 특성 추정값에 대한 평균}} \times 100$$

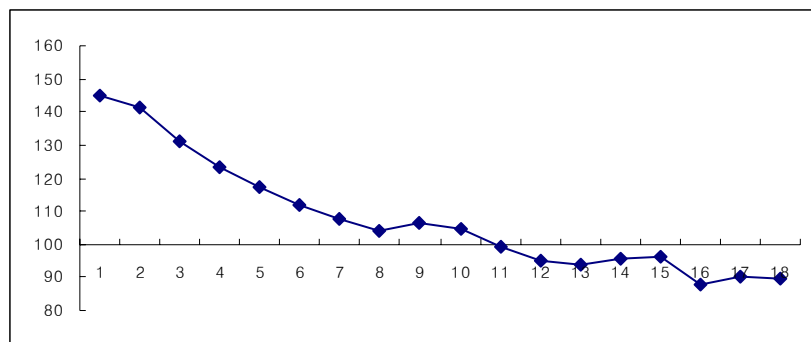
여기서 $r = 1, \dots, 36$ 은 진입회차이다. 진입회차에 의해 구성된 표본그
 룬 간에는 규모가 서로 다르기 때문에 직접적인 사람 수의 비교는 어렵
 다. 따라서 진입회차별 응답수준의 비교를 위한 특성 추정값으로 총계
 가 아닌 실업률이나 고용률 등의 비율에 대한 추정값을 이용하였다.

경제활동인구조사의 2005년 6월부터 2007년 12월까지의 31개월 자
 료를 이용하여 진입회차그룹지수를 구해보면 <표 4-4>와 같다. 진입 1
 회차 그룹의 실업률지수는 145.1로 표본으로 최초 진입한 그룹의 실업

를 추정값이 전체 평균보다 약 45% 높게 나타났다. 또한 진입회차가 늘어남에 따라 지수 값이 점차 줄어들고 있다. 반대로 고용률은 진입 1회차 그룹의 고용률 지수가 96.2로 전체 평균보다 낮게 나타나다가 회차가 증가할 수록 100에 가까워지고 있음을 알 수 있다. 이처럼 추정값은 진입회차그룹에 따라 차이를 보인다. 이러한 진입회차그룹별 추정값들은 서로 다른 기댓값을 갖기 때문에 진입회차편향을 조정해 줄 필요가 있다. 진입회차편향은 실업률 추정값에 큰 영향을 미친다고 알려져 있다(Hansen, 1955). 경제활동인구조사에서도 다른 추정값에 비하여 실업률이 진입회차별로 큰 차이를 보이고 있기 때문에 본 연구는 가중값 조정을 위해 실업률을 기준으로 연구를 진행하였다.

〈표 4-4〉 진입회차그룹별 각 특성 추정값에 대한 진입회차그룹지수(I_r)

진입회차(r)	1	2	3	4	5	6	7	8	9	...
고용률 지수	96.2	92.9	94.6	95.5	96.3	97.3	97.9	98.3	98.4	...
실업률 지수	145.1	141.3	131.3	123.1	117.2	111.6	107.4	104.3	106.4	...
비경비율 지수	109.3	107.1	106.0	105.1	103.8	103.0	102.6	102.3	102.1	...
그룹 수	31	31	31	31	31	31	30	29	28	...



[그림 4-1] 진입회차 변화에 따른 실업률 진입회차그룹지수

〈표 4-4〉에서와 같이 진입회차가 늘어남에 따라 진입회차그룹지수 값이 100에 가까워지면서 해당 그룹의 추정값이 전체평균에 가까워지는 경향이 있다. 이는 진입교체초기그룹의 응답수준이 조사가 반복적으로

진행되면서 전체평균으로 가까워지며 점차 안정화 된다는 것을 의미한다. [그림 4-1]은 진입회차 변화에 따른 실업률지수를 나타낸다. 마찬가지로 실업률 지수 값은 6회차까지 줄어들다가 7회차 이후엔 그 감소폭이 줄어들며 100근처의 값을 갖게 된다. 그러므로 7회차 이상 조사에 참여한 표본그룹은 어느 정도 안정적인 응답경향을 보이고 있는 그룹이라 할 수 있다. 따라서 표본이 교체되어 안정적으로 응답하게 되는 조사회차를 정함으로써 계산상 편의를 도모하고자 한다.

이에 따라 안정되기 이전의 진입회차 1~6회차 그룹과 안정된 이후의 7회차 이상 그룹과 같이 총 7개의 그룹으로 재구성하고, 안정되기 이전과 이후의 진입회차 그룹별 추정값 반영도를 다르게 주었다. 2005년 6월부터 2008년 3월까지의 34개월 자료를 이용하여 총 7개 진입회차그룹별로 실업률에 대한 진입회차그룹지수를 다시 계산한 결과가 <표 4-5>와 같다. <표 4-4>에서 36개 그룹에 대해 살펴보았던 결과와 유사하게 1회차 실업률 지수가 147로 높게 나타났으며 점차 줄어들다 5~7회차에 이르게 되면 점차 안정되어가는 추세를 보인다.

<표 4-5> 7개 진입회차그룹별 실업률 추정값에 대한 진입회차그룹지수(I_r)

진입회차(r)	1	2	3	4	5	6	7+
실업률 지수	147	132	124	119	113	109	90

나. 진입회차그룹별 가중값 산정

본 절에서는 안정되기 이전의 진입초기 표본에 의한 추정값의 편향을 줄여줄 수 있는 방법으로 다음과 같은 가중값 조정방법을 제안한다. 이 방법은 안정되기 이전의 진입회차그룹에는 가중값을 줄여주고, 안정된 이후의 진입회차그룹에는 가중값을 다소 높여주어 추정값에 미치는 영향력을 조정해주는 가법모형의 형태를 사용한다. 이때 w_r^* 는 진입회차가중값으로서 각 진입회차그룹별 추정값이 전체 추정값에 반영되는 정도를 결정하는 계수이다.

$$w_r^* = \begin{cases} 1 - \alpha_r, & \text{if } r \notin S \\ 1 + \beta, & \text{if } r \in S \end{cases}$$

$$\begin{aligned} & \cdot S = \{7, 8, \dots, 36\} \\ & \cdot \alpha_r = \frac{I_r - 100}{100}, \quad \beta = \frac{\sum_{r \notin S} \alpha_r w_{hijk}}{\sum_{r \in S} w_{hijk}} \end{aligned}$$

진입초기그룹의 진입회차가중값은 해당 진입회차그룹 추정값과 전체평균 추정값과의 차이의 크기가 반영되어 있는 진입회차지수를 이용하여 산정하였다. 그리고 안정된 그룹의 진입회차가중값은 해당 월의 15세 이상 인구를 유지하면서 안정되지 않은 그룹에서 줄여준 만큼의 영향력을 안정된 그룹에서 높여줄 수 있도록 매월 산정된다. 이와 같은 방식으로 계산된 진입회차가중값이 <표 4-6>과 <표 4-7>이며 진입초기 그룹의 가중값은 1보다 작고, 안정그룹은 1보다 큰 가중값을 갖는다. 진입회차가중값은 경제활동인구조사의 33개월 자료에 근거하여 경험적으로(empirical) 구해진 값이므로 최신자료의 경향을 파악하기 위해 연동주기인 3년마다 다시 계산할 필요가 있다.

<표 4-6> 진입초기그룹에 대한 진입회차가중값

진입회차	1	2	3	4	5	6	S(7+)
w_r^*	0.53	0.68	0.76	0.81	0.87	0.91	매월 산정됨 (1+β)

<표 4-7> 안정그룹에 대한 월별 진입회차가중값

년월	1+β	년월	1+β
2006. 01	1.10	2007. 01	1.11
2006. 02	1.09	2007. 02	1.10
2006. 03	1.10	2007. 03	1.10
2006. 04	1.10	2007. 04	1.12
2006. 05	1.10	2007. 05	1.12
2006. 06	1.10	2007. 06	1.11
2006. 07	1.10	2007. 07	1.11
2006. 08	1.10	2007. 08	1.10
2006. 09	1.10	2007. 09	1.10
2006. 10	1.10	2007. 10	1.11
2006. 11	1.11	2007. 11	1.11
2006. 12	1.10	2007. 12	1.12

다. 진입회차가중값 조정방안

본 연구에서 제안한 추정과정을 정리하면 다음과 같다. 최종 가중값은 표본설계로부터 구한 설계가중값과 진입회차별 가중값 그리고 추계인구에 대한 사후층화가중값의 곱으로 산출된다. 총계 추정량과 분산 추정량은 아래의 식을 이용할 수 있다. 분산추정 방법으로는 추출방법이나 무응답 여부에 관계없이 일반적으로 적용 가능한 잭나이프(jackknife) 방법을 이용하였다.

$$\begin{aligned} \hat{Y} &= \sum_h \hat{Y}_h \\ \cdot \hat{Y}_h &= \sum_r w_r^* \hat{Y}_{hr} = (1 - \alpha_r) \sum_{r \notin S} \hat{Y}_{hr} + (1 + \beta) \sum_{r \in S} \hat{Y}_{hr} \\ \cdot \hat{Y}_{hr} &= \sum_i \sum_j \sum_k w_{hijk} y_{hijk} \\ \cdot w_{hijk} &= w_{hijk}^0 \times w_{hijk}^1 \end{aligned}$$

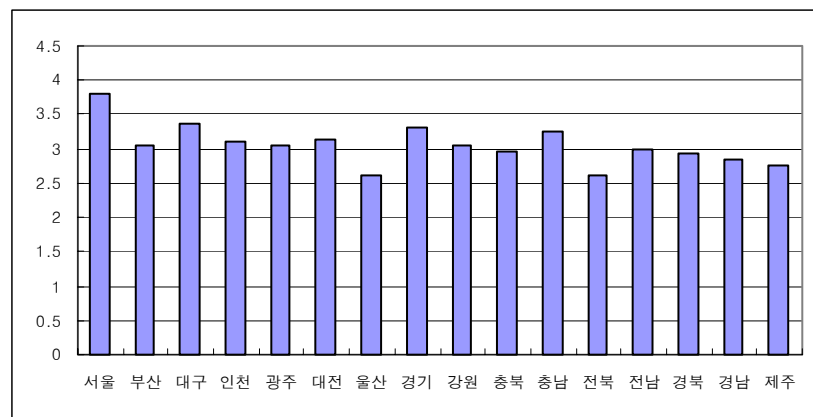
$$\begin{aligned} var(\hat{\theta}) &= \sum_{h=1}^H \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} (\hat{\theta}_{(hi)} - \hat{\theta}_{(h\cdot)})^2 \\ \cdot \hat{\theta} &= X \cdot \left(\frac{\hat{Y}}{\hat{X}} \right) \\ \cdot X &: \text{추계인구(benchmarking 대상)} \\ \cdot \hat{Y} &: \text{가중된 특성별 조사인구} \\ \cdot \hat{X} &: \text{가중된 15세이상 조사인구} \\ \cdot \hat{\theta}_{(h\cdot)} &= \sum_i \frac{\hat{\theta}_{(hi)}}{n_h} \end{aligned}$$

4. 가중값 조정결과

가. 시도별 가중값을 이용한 추정결과

현행 경제활동인구조사는 전국과 시도별로 추정값을 공표하고 있다. 따라서 시도별로 연동그룹편향이 어떻게 나타나고 있으며, 전국의 경향과는 얼마나 다른지 파악하기 위해서 전국 실업률을 근거로 산정된 진

입회차가중값을 이용하기에 앞서 다음과 같은 내용을 검토하였다. 2005년 6월부터 2007년 12월 자료로 분석한 [그림 4-2]의 시도별 전입자 현황을 보면 평균 전입자 비율이 시도별로 약간 차이가 있다. 매월 전입으로 교체되는 표본의 비율이 서울이 약 3.8%, 경기도가 3.3%로 수도권에서의 전입교체 비율이 높으며 울산이 2.6%로 가장 낮다. 이처럼 매월 발생하는 전입자 규모가 시도별로 다소 차이가 있고, 지역별로 응답자들의 성향이나 조사방법 등이 각기 다르기 때문에 시도별 전입회차편향이 다르게 나타날 가능성이 있다.



[그림 4-2] 시도별 전입자 현황

이를 확인하고자 시도별로 실업률에 대한 전입회차지수를 <표 4-8>과 같이 시산해보았다. 모든 시도에서 1회차 실업률 지수가 100보다 큰 값으로 전입 1회차 그룹의 실업률이 전체보다 높게 추정되고 있으며 조사 횟수가 늘어남에 따라 점차 100에 가까운 값으로 줄어들면서 전국 경향과 유사하게 추정되고 있다. 서울의 경우 1회차와 2회차 실업률 지수가 약 120으로 전입회차에 따른 추정값의 차이가 다른 지역에 비해서 상대적으로 적다. 또한 강원, 충남, 전남, 경북 지역에서는 1회차 실업률 지수가 200보다 크게 나오는 등 지수 값이 시도에 따라 매우 큰 차이를 보이고 있다.

<표 4-8> 시도별 실업률 진입회차지수 시산 결과

진입회차 지역	1	2	3	4	5	6	7+
서울	121	122	114	111	111	102	94
부산	131	132	125	117	105	113	92
대구	151	141	139	118	111	119	89
인천	116	115	124	121	120	124	93
광주	133	125	116	117	110	106	93
대전	178	137	141	137	133	110	84
울산	172	143	120	125	134	137	87
경기	144	134	126	113	107	106	90
강원	213	222	204	168	160	171	75
충북	159	149	151	132	127	146	89
충남	217	178	147	143	113	115	79
전북	148	127	148	141	133	133	93
전남	208	171	162	170	179	144	74
경북	201	187	144	154	141	127	79
경남	140	123	135	141	158	136	90
제주	174	172	174	172	166	146	86

특정지역의 진입 1회차 그룹에서 전체 추정값보다 100% 이상 큰 값으로 추정하게 되는 원인은 다음과 같이 추측해 볼 수 있다. 즉, 해당 지역의 경우 표본의 규모가 크지 않고 이를 다시 진입회차에 따라 7개의 그룹으로 나누었기 때문에 진입회차그룹내 표본 규모는 매우 작아지게 된다. 그리고 그 표본 중에서 실업자가 차지하는 비중은 매우 적기 때문에 해당 지역의 진입회차그룹내 실업률은 그 그룹의 표본규모에 매우 의존하게 된다. 실제로 3~4회차 그룹의 실업자 수가 0명인 지역의 시도별 진입회차별로 구해진 실업률 자체를 신뢰하기가 곤란하며 이를 근거로 하여 구해진 진입회차지수에 오류의 가능성이 제기될 수 있다.

시도별 실업률 진입회차지수를 바탕으로 시도별 가중값을 산정하여 2007년 12월 자료에 대한 시도별 추정값을 구해 보았다. 단, 시도별 가중값을 산정할 때 지수의 값이 200을 넘으면 가중값이 음수가 나오기 때문에 이를 방지하고자 해당 지역의 진입회차그룹을 추정값 산정 시 제외시켰다. <표 4-9>는 진입회차가중값을 적용하지 않고 기존의 비추정방법을 이용한 시도별 추정값 (A), 전국 실업률을 기준으로 산정한 진

입회차가중값을 적용한 추정값 (B), 그리고 시도별 실업률을 기준으로 시도별 진입회차가중값을 적용한 추정값 (C)를 나타낸다. 이를 비교해 보면, (C)의 추정값이 (B)보다 (A)로부터 조정 폭이 더 커진다. 특히 충남은 시도별 가중값을 적용했을 때 고용률이 약 1.1%p 정도의 차이가 나고, 강원도의 실업률은 약 0.7%p의 차이를 보이는 등 전국 가중값으로 조정한 추정결과 보다 월별, 지역별로 편차가 매우 크게 나타난다.

앞에서 언급한 바와 같이 시도별 진입회차지수는 다소 불안정한 측면이 있고, 이를 근거로 추정된 결과 역시 변동 폭이 매우 크게 나타나고 있다. 이런 이유에서 본 연구는 전국 실업률을 바탕으로 산출한 가중값을 이용하여 추정하고자 한다. 이는 시도별 실업률 지수가 전국에서 구한 지수와 유사하게 진입초기에 높게 나타나다가 점차 안정되어 가는 추세를 보이기 때문에 전국 가중값을 이용하여도 이러한 경향은 반영될 수 있으며 시도 간 불안정성은 줄일 수 있을 것으로 기대되기 때문이다.

〈표 4-9〉 2007년 12월 시도별 추정결과

(단위: %)

지역	진입회차가중값 미조정 (A)			전국 진입회차가중값 조정 (B)			시도 진입회차가중값 조정 (C)		
	고용	실업	비경	고용	실업	비경	고용	실업	비경
서울	59.8	4.1	37.7	59.9	4.1	37.5	59.9	4.1	37.6
부산	55.4	3.9	42.4	55.4	3.7	42.4	55.5	3.8	42.4
대구	56.8	2.7	41.6	56.9	2.6	41.6	56.9	2.6	41.6
인천	59.7	4.2	37.7	59.9	4.2	37.5	59.9	4.2	37.5
광주	56.2	3.7	41.6	56.5	3.6	41.4	56.4	3.6	41.5
대전	57.2	3.7	40.7	57.7	3.4	40.3	57.9	3.2	40.1
울산	58.7	2.6	39.7	59.0	2.4	39.5	59.1	2.3	39.5
경기	59.9	2.9	38.3	60.0	2.8	38.3	60.0	2.8	38.3
강원	54.4	2.0	44.6	54.4	1.8	44.6	54.7	1.3	44.6
충북	55.8	2.5	42.8	55.7	2.3	42.9	55.7	2.3	43.0
충남	61.3	1.8	37.6	61.8	1.6	37.1	62.5	1.5	36.6
전북	57.8	2.6	40.7	58.0	2.5	40.5	58.1	2.5	40.4
전남	63.6	1.8	35.2	63.9	1.7	35.0	64.3	1.7	34.6
경북	60.9	1.7	38.1	61.1	1.5	38.0	61.5	1.3	37.7
경남	59.7	2.4	38.9	60.3	2.3	38.3	60.3	2.3	38.3
제주	69.2	2.0	29.4	69.5	1.8	29.2	69.8	1.7	29.0

나. 전국 가중값을 이용한 추정결과

<표 4-10>과 <표 4-11>은 진입회차가중값을 적용하지 않은 기존의 비추정방법과 본 연구에서 제안하는 진입회차가중값을 적용한 방법에 대한 추정결과이다. 진입회차가중값을 조정하였을 때 취업자의 수가 조정하지 않았을 때 보다 약 10,000~130,000명이 증가하였으며 실업자는 약 5,000~33,000명 정도가 줄어들었다. 고용률은 0.02~0.34%p, 실업률 측면에서는 0.04~0.15%p 정도가 진입회차가중값에 의하여 조정되었다. 진입초기그룹에서 실업률이 높게 나타났던 편향을 줄여줌으로써 기존의 추정값보다 약 0.1%p만큼 낮아진 값으로 추정되었다.

진입회차가중값 조정 여부에 대한 추정값의 분산을 추정값의 상대표준오차(CV : Coefficient of Variation)를 이용해서 비교해 보았다. <표 4-12>를 살펴보면, 두 방법의 추정값에 대한 CV값은 차이를 보이지 않는다. 또한 2007년 월별, 시도별 추정값을 2006년 동월과 비교해 보았다. [그림 4-3]은 두 방법에 대한 전년 동월 대비 증감률을 나타낸 것으로 두 방법의 증감률이 거의 비슷하게 나타났다. 이때 증감률은 아래와 같이 계산되었다.

$$\text{증감률}(\%) = \frac{\text{금년 동월 실업률} - \text{전년 동월 실업률}}{\text{전년 동월 실업률}}$$

본 연구에서 제안한 진입회차가중값을 조정하였을 때의 실업률에 대한 추정결과는 상대표준오차측면에서 조정하지 않았을 때와 큰 차이가 없고, 시도별 전년 동월 대비도 미조정시와 유사한 경향을 보이는 것으로 나타났다. 따라서 진입회차가중값 조정 방법은 진입회차에 의한 편향을 조정해줌으로써 추정값의 정도를 높이는 데 기여할 수 있을 것으로 보인다.

〈표 4-10〉 진입회차기증값 미조정 시 추정결과

(단위: 천명, %)

년월	15세이상 인구	경제활동 인구	취업자	실업자		비경제활 동인구	
				고용률	실업률		
2006. 01	38,546	23,340	22,471	58.3	869	3.87	15,206
2006. 02	38,592	23,365	22,412	58.1	953	4.25	15,227
2006. 03	38,640	23,769	22,848	59.1	921	4.03	14,871
2006. 04	38,684	24,088	23,242	60.1	846	3.64	14,597
2006. 05	38,730	24,267	23,484	60.6	783	3.33	14,463
2006. 06	38,776	24,320	23,501	60.6	819	3.49	14,455
2006. 07	38,814	24,270	23,447	60.4	823	3.51	14,544
2006. 08	38,835	23,972	23,164	59.6	807	3.49	14,864
2006. 09	38,854	24,096	23,330	60.0	766	3.28	14,758
2006. 10	38,874	24,253	23,463	60.4	789	3.36	14,621
2006. 11	38,892	24,225	23,458	60.3	767	3.27	14,667
2006. 12	38,905	23,773	22,989	59.1	784	3.41	15,132
2007. 01	38,950	23,580	22,729	58.4	851	3.74	15,370
2007. 02	38,996	23,536	22,674	58.1	861	3.80	15,461
2007. 03	39,044	23,960	23,121	59.2	840	3.63	15,084
2007. 04	39,091	24,337	23,520	60.2	817	3.47	14,755
2007. 05	39,137	24,537	23,758	60.7	779	3.28	14,600
2007. 06	39,178	24,593	23,816	60.8	777	3.26	14,585
2007. 07	39,205	24,545	23,750	60.6	795	3.35	14,660
2007. 08	39,231	24,214	23,458	59.8	756	3.22	15,018
2007. 09	39,259	24,341	23,622	60.2	719	3.04	14,918
2007. 10	39,290	24,482	23,750	60.4	733	3.09	14,808
2007. 11	39,316	24,471	23,739	60.4	732	3.09	14,845
2007. 12	39,342	23,993	23,257	59.1	736	3.16	15,349

〈표 4-11〉 진입회차가중값 조정 시 추정결과

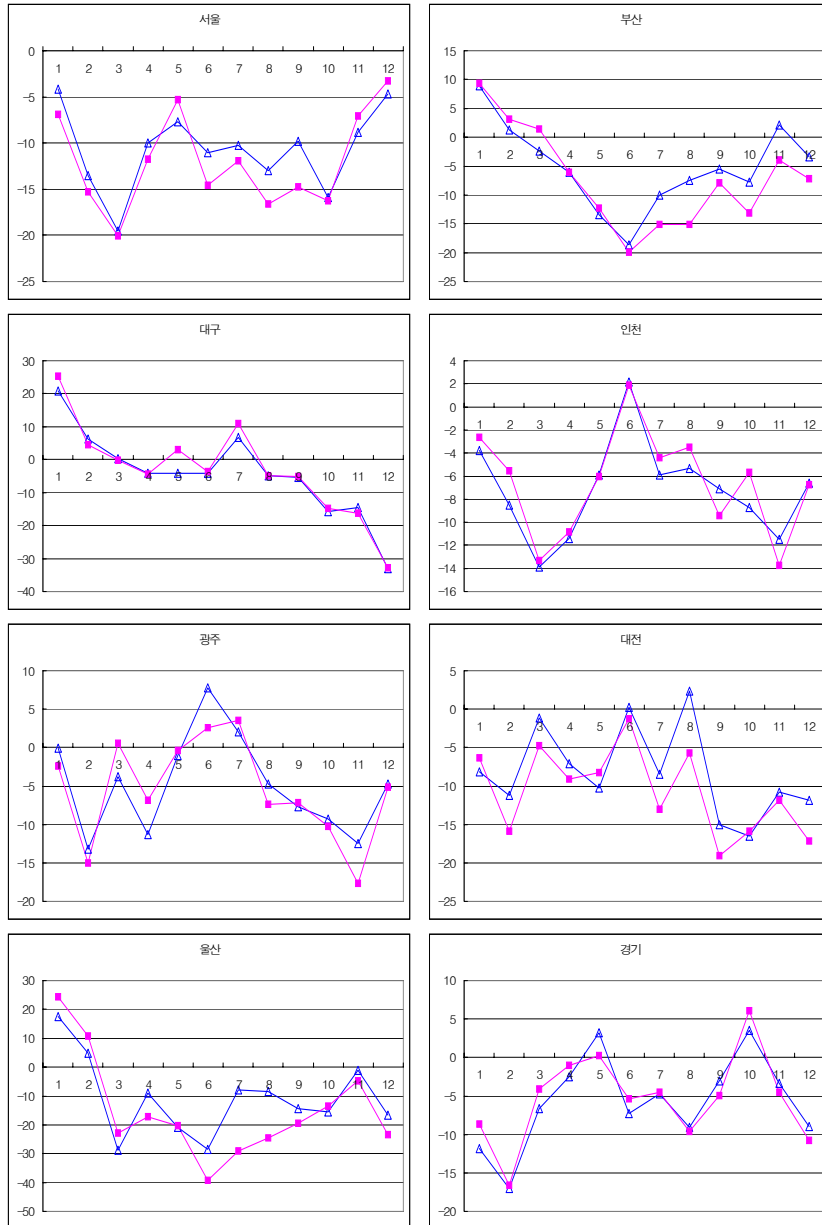
(단위: 천명, %)

년월	15세이상 인구	경제활동 인구	취업자	실업자		비경제활 동인구	
				고용률	실업률		
2006. 01	38,546	23,407	22,556	58.5	851	3.77	15,139
2006. 02	38,592	23,430	22,503	58.3	928	4.12	15,162
2006. 03	38,640	23,846	22,943	59.4	903	3.94	14,794
2006. 04	38,684	24,169	23,334	60.3	835	3.58	14,515
2006. 05	38,730	24,325	23,553	60.8	772	3.28	14,405
2006. 06	38,776	24,368	23,556	60.7	813	3.45	14,407
2006. 07	38,814	24,326	23,518	60.6	808	3.44	14,488
2006. 08	38,835	24,040	23,240	59.8	800	3.44	14,795
2006. 09	38,854	24,115	23,357	60.1	758	3.25	14,738
2006. 10	38,874	24,272	23,502	60.5	771	3.28	14,602
2006. 11	38,892	24,249	23,493	60.4	756	3.22	14,643
2006. 12	38,905	23,845	23,070	59.3	775	3.36	15,060
2007. 01	38,950	23,657	22,824	58.6	833	3.65	15,294
2007. 02	38,996	23,616	22,780	58.4	836	3.67	15,380
2007. 03	39,044	24,089	23,255	59.6	834	3.59	14,955
2007. 04	39,091	24,403	23,601	60.4	801	3.40	14,688
2007. 05	39,137	24,572	23,802	60.8	769	3.23	14,565
2007. 06	39,178	24,599	23,842	60.9	757	3.18	14,579
2007. 07	39,205	24,572	23,805	60.7	767	3.22	14,633
2007. 08	39,231	24,252	23,527	60.0	725	3.08	14,979
2007. 09	39,259	24,365	23,680	60.3	685	2.89	14,894
2007. 10	39,290	24,476	23,759	60.5	717	3.02	14,814
2007. 11	39,316	24,516	23,798	60.5	718	3.02	14,800
2007. 12	39,342	24,054	23,334	59.3	720	3.08	15,288

〈표 4-12〉 진입회차가중값 조정 여부에 따른 CV 추정결과

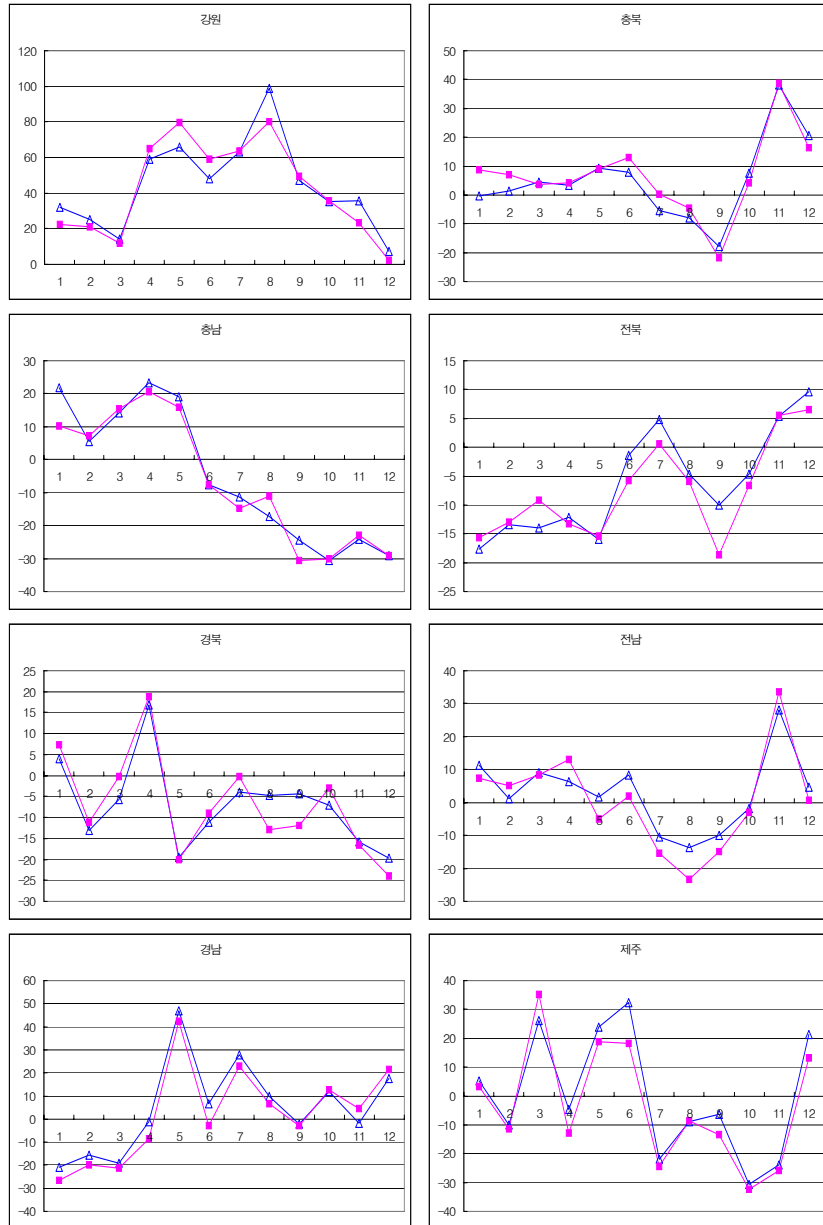
년월	진입회차가중값 미조정 시 CV				진입회차가중값 조정 시 CV			
	경활 인구	취업자	실업자	비경 인구	경활 인구	취업자	실업자	비경 인구
2006. 01	0.43	0.44	3.35	0.65	0.43	0.44	3.41	0.66
2006. 02	0.42	0.43	3.27	0.64	0.42	0.43	3.30	0.65
2006. 03	0.41	0.42	3.30	0.65	0.41	0.42	3.35	0.65
2006. 04	0.41	0.42	3.53	0.68	0.41	0.42	3.60	0.68
2006. 05	0.40	0.41	3.40	0.67	0.40	0.41	3.44	0.67
2006. 06	0.40	0.42	3.36	0.67	0.40	0.42	3.39	0.68
2006. 07	0.39	0.41	3.32	0.66	0.40	0.41	3.39	0.66
2006. 08	0.41	0.42	3.42	0.66	0.40	0.42	3.50	0.66
2006. 09	0.40	0.41	3.52	0.65	0.40	0.41	3.51	0.65
2006. 10	0.40	0.42	3.33	0.67	0.40	0.41	3.39	0.67
2006. 11	0.40	0.41	3.48	0.66	0.40	0.41	3.55	0.67
2006. 12	0.41	0.42	3.45	0.64	0.41	0.41	3.50	0.64
2007. 01	0.41	0.43	3.43	0.63	0.41	0.42	3.43	0.63
2007. 02	0.40	0.42	3.51	0.61	0.40	0.41	3.42	0.62
2007. 03	0.41	0.42	3.58	0.65	0.40	0.41	3.67	0.64
2007. 04	0.42	0.43	3.41	0.69	0.40	0.42	3.47	0.67
2007. 05	0.41	0.42	3.49	0.69	0.40	0.41	3.54	0.68
2007. 06	0.41	0.41	3.53	0.68	0.40	0.41	3.59	0.68
2007. 07	0.40	0.41	3.94	0.67	0.41	0.41	3.70	0.68
2007. 08	0.42	0.42	3.92	0.68	0.42	0.42	3.87	0.67
2007. 09	0.44	0.44	4.11	0.72	0.44	0.44	3.99	0.71
2007. 10	0.42	0.42	3.77	0.70	0.43	0.43	3.85	0.70
2007. 11	0.45	0.45	3.89	0.73	0.45	0.45	4.03	0.74
2007. 12	0.46	0.47	3.83	0.72	0.45	0.46	3.96	0.71

△ 진입회차가종값 미조정 ■ 진입회차가종값 조정



[그림 4-3] 시도별 전년동월(2006년) 대비 실업률 증감률(2007년)

△ 진입회차가종값 미조정 ■ 진입회차가종값 조정



[그림 4-3] 시도별 전년동월(2006년) 대비 실업률 증감률(2007년)(계속)

제4절 결론

본 연구는 기존의 고정표본제에서 사용하던 추정량을 복합추정량으로 변경을 검토하기 전 단계로서 현재까지 연동 1주기가 진행된 연동표본제의 특성을 탐색하였다. 경제활동인구조사에서의 연동표본교체와 전·출입 표본가구의 교체가 추정값에 어떠한 영향을 미치는지 확인하기 위해 2005년 1월부터 2007년 12월까지의 3년간 자료를 이용하여 분석해보았다. 전체 표본의 약 5%정도가 매월 교체되고 있으며, 진입교체로 인한 표본들 간 조사 참여기간의 차이가 응답수준에 영향을 미치는 것으로 나타났다. 진입회차가 낮은 그룹일수록 실업률 추정값이 높게 나타나고 있어서 전체 추정값이 편향될 가능성이 있었다. 따라서 전체 추정값을 안정적으로 구하기 위하여 진입회차그룹별로 추정값에 미치는 영향력의 정도를 다르게 책정해주는 방법을 활용하였다. 즉, 전체표본을 진입회차가 높아 어느 정도 응답형태가 안정적인 그룹과 진입회차가 낮은 초기그룹으로 진입회차에 따라 나누고 안정그룹은 추정값에 반영되는 정도를 높여주고, 진입초기그룹에는 반영도를 낮춰줌으로써 진입회차편향을 가중값을 통해 조정할 수 있는 방안을 제시하였다.

본 연구에서는 과거에 조사된 자료를 분석함으로써 연동그룹편향을 파악하고 그 결과를 반영할 수 있는 방법을 모색하였다. 그러나 진입초기 그룹의 응답성향 및 측정과정에서 나타나는 오류 등을 포착하기 위해서는 좀 더 이론적이고 구체적인 방법에 의한 연구를 진행할 필요가 있다. 또한 경제활동인구조사는 반복조사의 형태이므로 지속되고 있는 표본그룹의 월간 변동이나 전월 추정값과 같은 과거의 정보를 활용하여 추정값을 개선시키는 복합추정방법에 대한 지속적인 연구가 필요할 것으로 보인다. 본 연구에서 제안한 방법은 연동그룹의 영향을 가중값으로 반영하는 방법이기 때문에 순수한 연동그룹 효과를 포착하는 데는 한계가 있을 수 있다. 따라서 기존의 비추정방법을 연동표본제에 적합한 방법으로 변경하기 위해서는 호주의 사례와 같이 여러 추정기법들을 비교·검토함으로써 우리의 조사환경에 맞는 방법을 찾아내는 노력을 기울여야 한다. 우리나라의 경우 연동표본의 교체규모가 외국사례에 비

해 크지 않고, 연동주기가 36개월로 매우 길다는 점 등을 고려하여 향후 경제활동인구조사에 적합한 추정기법 개발이 필요하며, 본 연구가 이를 위한 주요한 발판이 될 수 있을 것으로 기대한다.

참고 문헌

- 임경은(2007), “표본조사에서의 가중값 조정방안 : 추정방법 중심으로”, 통계개발원, 「국가통계발전을 위한 통계기법의 개선」, pp.7-53.
- 통계청(2008), “경제활동인구조사 지침서”.
- _____ (2007), “가구부문 표본개편 결과”.
- Australian Bureau of Statistics(2007), “Forthcoming Changes to Labour Force Statistics”, Information Paper.
- Bailar, B.A.(1975), “The Effects of Rotation Group Bias on Estimates from Panel Surveys”, *Journal of the American Statistical Association*, Vol. 70, No. 349, pp.23-30.
- Bell, P.A.(2001), “Comparison of Alternative labour Force Survey Estimators”, *Survey Methodology*, Vol. 27, No. 1, pp.53-63.
- Fuller, W.A.(1999), “Canadian Regression Composite Estimation”, Unpublished manuscript.
- Fuller, W.A. and Rao, J.N.K.(2001), “A Regression Composite Estimator with Application to the Canadian Labour Force Survey”, *Survey Methodology*, Vol. 27, No. 1, pp.45-51.
- Gambino, Jack, Kennedy, Brian and Singh, Mangala P.(2001), “Regression Composite Estimation for the Canadian Labour Force Survey: Evaluation and Implementation”, *Survey Methodology*, Vol. 27, No. 1, pp.65-74.
- Lessler, J.T. and Kalsbeek, W.D.(1992), *Nonsampling Error in Surveys*, John Wiley & Sons, Inc.
- Singh, A.C., Kennedy, B. and Wu, S.(2001), “Regression composite estimation for the Canadian Labour Force Survey with a rotating panel design”, *Survey Methodology*, Vol. 27, No. 1, pp.33-44.
- Singh, A.C., Kennedy, B., Wu, S. and Brisebois, F.(1997), “Composite

Estimation for the Canadian Labour Force Survey”, *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp.300-305.

Solon, G.(1986), “Effects of Rotation Group Bias on Estimation of Unemployment”, *Journal of Business & Economic Statistics*, Vol. 4, No. 1, pp.105-109.

Kemsley, W.F.F.(1961), “The household Expenditure Enquiry of the Ministry of Labor”, *Applied Statistics*, Vol. 10, No. 3, pp.117-135.

Hansen, M.H. *et. al.*(1955), “The Redesign of the Census Current Population Survey”, *Journal of the American Statistical Association*, Vol. 50, No. 271, pp.701-719.

Statistics Canada(2007), “Guide to the Labour Force Survey”.

Turner, P.(1961), “Inter-Week Variations in Expenditure Recorded During a Two-Week Survey of Family Expenditure”, *Applied Statistics*, Vol. 10, No. 3, pp.136-146.

U.S. Bureau of Labour Statistics and U.S. Census Bureau(2006), “Design and Methodology: Current Population Survey”.

Waksberg, J. and Neter, J.(1965), “*Response Errors in Collection of Expenditures Data by Household Interviews: An Experimental Study*”, Technical Paper, No. 11, U.S. Bureau of the Census.