

家口所得의 市道別 隔差 및 要因 分析*

盧 應 源**

논문 초록 15개 광역시·도별 가구소득 자료가 통계청의 『1996 가구소비실태조사 보고서』로 가용하게 됨에 따라, 본 연구는 복잡한 표본설계로 추출된 표본자료(표본수=24,290가구)에 입각하여 설계-일치 회귀계수의 분산을 추정하는 복제분산 추정법을 적용함으로써 광역시·도의 가구소득함수를 추정하고 지역간 평균 가구소득의 격차와 요인을 분석한다. 지역별 평균 가구소득은 최고 서울의 2,721만원에서 최하 경북의 2,110만원(서울의 72.5%)까지 분포하지만, 각 지역 가구소득의 서울대비 격차(350만~860만원)의 약 1/2 이상(단, 인천은 26%)은 가구당 취업자수, 가구주의 교육수준, 금융자산 보유액 등 가구특성들의 차이에 기인한다. 그 나머지 130만~280만원은 동일한 가구특성을 보유한 가구로서 단지 거주지역이 다르다는 이유로 인해 나타나는 순수한 지역 특성효과이며, 그것은 지역별 3차산업 비율과 지역내 총생산 점유율로 잘 설명된다.

핵심주제어: 지역별 가구소득, 소득격차, 복제분산 추정법

경제학문헌목록 주제분류: D31

* 본고는 1998년 9월 1일 제4회 통계의 날 행사의 기념 세미나 및 9월 8일 충남대학교 경상대학 부설 경영경제연구소 세미나에서 발표한 논문을 전가구를 대상으로 확장하고 몇 가지 오류를 수정한 것이며, 1998년 11월 28일 한국계량경제학회 정기학술대회에서 발표된 것이다. 가구조사 원자료와 함께 그 체계와 내용을 친절하게 안내해주신 통계청 사회통계과의 우사임 사무관, 그리고 유익한 논평을 해주신 분들께, 특히 이정우, 배진한, 김성태 교수, 그리고 익명의 심사위원들께 감사드리며, 본고의 의견과 오류는 전적으로 필자의 책임임을 밝혀둔다.

** 충남대학교 경제학과 교수(nhoew@hanbat.chungnam.ac.kr)

I. 서론

본 연구는 통계청의 『1996 가구소비실태조사 보고서』 자료를 이용하여, 우리나라 15개 시·도(6대 도시와 9개 도) 별 비농가 소득함수를 추정하고, 지역별 가구 평균소득 격차의 요인을 실증적으로 분석하고자 한다.

그동안 우리나라는 전체적으로 높은 경제성장을 달성하면서도 지역간 불균등 발전에 따라 지역갈등 문제가 대두되고 있다. 그러나 지역간 경제력 또는 소득 격차에 관한 실증적 연구는 자료의 미비 등으로 1990년대에 들어서야 김대환(1992), 김성태 외(1991) 및 이정우(1997) 등에 의해 시작되었다. 이들은 대부분 지역내 총생산 및 1인당 지역내 총생산의 시계열 자료에 입각하여 수도권과 영남권의 소득이 충청-호남권에 비해 상대적 격차는 점차 감소해 왔으나 그 절대적 수준이 높았음을 보임으로써 지역간 소득격차가 지역갈등의 한 요인임을 밝힌 바 있다. 그런데 이들 연구는 1인당 국내총생산 자료에 입각하여 소득격차를 확인하였으나 그 요인분석으로 나아가지는 못하였다. 단, 김성태 외(1991)는 1970~1987년 자료에 입각하여 광공업 생산함수를 대도시권(서울, 부산), 대도시 인접권(경기, 경북, 경남) 및 비도시권(충청, 호남, 제주)의 3지역별로 추정함으로써 지역별 격차를 요소(노동과 자본) 증가율, 노동 수익률 탄력성 및 기술진보율의 차이로 설명하였으나, 이러한 상이한 특성들이 유의적으로 다른가에 관한 검증은 생략하였다.

지역별 1인당 지역내총생산은 지역별 1인당 경제력을 나타내기는 하지만, 생산 측면에서 집계된 소득개념이므로 지역주민들의 경제적 복지수준과는 상당한 거리가 있으며 생활실태 분석에는 가구소득이 더 적절하다.¹⁾ 가구소득 분배에 관한 연구는 일찍이 도시가계조사와 농가경제조사 자료를 이용한 주학중(1979), 반성환(1979) 등의 연구가 있으나 주로 전국의 인적 분배, 직종별(근로자-사업자) 분배 및 농가소득 분배가 중심이었고,²⁾ 지역별로는 都農간의 소득분배를 주로 다루었다. 사실, 본 연구는 지역별로 지역내 가구소득의 인적 분배문제가 아니라, 평균 가구소득의 지역간 격차를 분석하고자 한다. 이 분야에서도 1985년에 5,107개 표본가

1) 국민소득통계의 1인당 소득과 『1996 가구소비실태조사 보고서』의 1인당 소득 간의 차이에 관해서는 강석훈(1998)을 참조하라.

2) 최근까지의 한국의 인적 소득분배 및 지역경제력 격차에 관한 문헌에 관해서는 각각 이정우(1997a)와 이정우(1997b)를 참조하라.

구를 독자적으로 추출·조사한 권순원 외(1992)의 연구에서는 지역을 시부와 군부로 나누어 지역간 소득격차를 계산하였을 뿐이며, 1993년도 대우경제연구소의 한국가구패널조사 자료(4,547가구)를 이용한 이은우(1995)는 지역별 가구소득 함수를 추정하였으나 전국을 서울, 기타 도시 및 농촌으로만 구분하였다.

그리하여 본 연구와 같이 행정구역상의 대도시와 도를 구분하여 광역 지방단체에 입각한 지역별 가구소득의 격차와 그 요인을 분석한 지역별 연구는 지금까지 기초조사자료가 없었기 때문에 수행될 수 없는 실정이었다.³⁾ 단, 『1996 가구소비실태조사 보고서』는 전국에서 2만 4,290개에 달하는 표본가구들을 추출·조사하였으나 농가는 제외하였다.

본 연구는 통계적 및 계량경제학적 추정방법론의 관점에서도 한국에서는 새로운 시도를 하였다. 즉, 『1996 가구소비실태조사 보고서』 자료 등과 같이 복잡한 표본설계에 입각하여 수집된 자료를 이용하여 회귀 분석할 경우, ‘표본설계 가중 추정량’이 근사적 불편성(approximate unbiasedness)과 일치성을 갖는 추정량임이 밝혀져 있다. 그런데, 이 추정량의 표준오차는 널리 사용되는 SAS, SPSS 등 상용 통계 프로그램으로는 올바르게 계산되지 않으며, SAS의 가중(‘이분산 교정용 가중치’) 최소자승 추정법으로 구한 표준오차는 과소 추정되므로 잘못된 통계적 추론을 낳을 수도 있음이 Carlson(1988), Brogan(1998) 등에 강조되어 있다. 이 표준오차는 이 목적에 맞게 개발된 분산추정용 프로그램을 사용하여야 제대로 계산될 수 있다. 그런데, 이 프로그램들을 사용하기 위해서는 표본설계에 관한 자세한 정보(특히, 표본조사구를 절반씩 양분한 半分標本(half-sample) 식별코드 및 복제 가중치 자료)가 필요한데, 그 정보가 『1996 가구소비실태조사 보고서』의 공표된 원자료 목록에 수록되어 있지 않다. 최근 들어 경제·사회 등 각 분야에서 대규모 표본조사를 수행하는 우리나라의 정부 및 민간기관들이 점차 그 원자료를 공표하고 있으나, 이러한 필수 자료는 대부분 아직도 공개하지 않고 있어 그 자료를 이용한 표본설계 가중 추정량이 계산되지 못하고 있으며 대부분의 연구자들도 그 추정법을 잘 알지 못하는 실정이다. 따라서, 본고에서는 다소 자의적이거나 표본조사 가구들을 양분하여

3) 통계청은 1991년 10-12월에 제1차 가구소비실태조사를 실시한 바 있으나, 그 자료가 널리 알려지지 않았고, 1994년부터 대우경제연구소의 〈한국가구패널조사〉 자료가 발표되고 있으나 표본 가구수가 약 4,500개로서 15개 지역별 연구를 하기에는 신뢰도에 문제가 있을 것으로 생각된다.

WesVarPC라는 분산추정 프로그램을 이용하여 회귀계수와 그 표준오차를 계산하였다. 이러한 방법론의 중요성 때문에 제Ⅱ절에서 추정방법을 설명하였다.

제Ⅲ절에서는 소득분포의 통상적 특성상 지역별 가중평균 소득, 대수 가구소득의 평균, 중위소득, 가구원 1인당 소득 등 여러가지 가구소득의 대표치들을 추정하였다. 1996년도 가중평균 소득은 서울가구가 2,911만원으로서 가장 높을 뿐만 아니라 다른 지역에 비해 월등히 높고, 경북이 그 72.5%인 2,110만원으로서 가장 낮다. 그러나 이러한 격차가 통계적으로 유의한가를 검정한 결과, 9개 도의 가구소득들은 서울가구소득과 유의적으로 다르지만 5대 도시의 소득들은 서울의 그것과 유의적으로 다르지 않았다.

제Ⅳ절에서는 각 지역 가변수와 함께 여러가지 가구특성 변수들을 도입하여 가구소득함수를 추정하였다. 가구특성들을 통제할 때, 즉 유사한 가구특성을 가진 가구들간에 비교할 때, 충남·대전·경남을 제외한 각 지역의 소득은 서울가구소득보다 유의적으로 더 낮음이 발견되었다. 다음으로, 추정된 가구소득함수에 기초하여 지역별 가구소득의 격차를 가구특성 요인들과 순수한 지역특성 요인들에 기인하는 부분으로 분해하였다. 끝으로, 지역 가변수로 상징되는 순수한 지역효과를 발휘하는 구체적 경제변량을 포착하기 위해 지역의 산업구조 및 성장 잠재력을 나타내는 지역특성 변수들을 추가하여 회귀분석을 시도하여, 지역별 3차산업 비율과 지역내총생산 점유율이 클수록 지역특성 효과가 유리하게 작용함을 발견하였다. 제Ⅴ절에서는 본 연구결과를 요약하고 지역개발 정책에 관한 정책적 함의를 도출하고 아울러 그 한계를 지적하였다.

Ⅱ. 표본설계와 추정방법

『1996 가구소비실태조사 보고서』의 표본은 전국의 총가구(모집단) 중에서 무작위로 추출된 단순 무작위표본(simple random sample)이 아니라, 표본조사의 효율성과 지역별 정보 획득이라는 목표에 적합하도록 2단계 層化系統抽出法(stratified systematic sampling)이라는 복잡한 표본설계(complex sampling design) 과정을 거쳐 추출된 것이다. 이렇게 추출된 표본의 관측치들은 ‘동일·독립적 분포’(i. i. d.)라는 특성을 갖지 않으므로, 이것을 가정한 전통적 통계적 추론법을 적용하면 분산

추정치에 큰 편이성이 발생한다. 올바른 분산 추정법을 소개하기 전에, 그것에 필요한 표본설계의 주요 내용을 먼저 살펴보기로 하자.

1. 표본 설계⁴⁾

『1996 가구소비실태조사 보고서』는 전국의 가구를 목표틀(target frame)로 하지만, 1990년 인구주택총조사의 보통조사구와 섬조사구를 행정구역별로 정리한 후 추출간격 10으로 계통추출한 ‘인구주택총조사 10% 표본조사구’ 중에서 섬지역과 시설단위 조사구를 제외한 18,524개 보통조사구를 표본틀(sampling frame)로 한다. 행정구역(및 지역)별 정보를 획득하고 표본의 대표성을 높이기 위해 1992년 7월 1일 현재의 행정구역에 따라 이 표본틀을 6대 도시(서울, 부산, 대구, 인천, 광주, 대전) 및 9개 각 도의 시부 및 군부(18개) 등 총 24개의 집단으로 層化하여, 각 層(stratum)별로 아래의 2단계 표본추출법에 따라 표본가구를 추출한다.

제1단계로, 특정 분류지표에 따라 배열된 조사구 명부로부터 모집단(목표틀)의 ‘크기의 測度’(1990년 인구주택총조사 결과의 가구수를 10으로 나누어 반올림한 수)에 비례하는 확률로 1차 추출단위(PSU)인 조사구를 계통추출한다. 여기서 조사구數는 비용을 고려하여 약 3만 가구(조사구 1,150개)로 책정된 전국 표본가구數를 失業者數의 상대표준오차와 目標精度를 기준으로 지역별로 할당하여 결정된다. 제2단계로, 제1단계에서 추출된 표본조사구에 한해 거처당 가구수가 균등하도록(10개씩) 거처들을 구역화하고 크기의 측도와 같은 수의 조사구역(segments)으로 분할한 다음, 서로 인접한 3개의 조사구역을 무작위로 추출하여, 그 안의 적격 가구들을 모두 조사한다(단, 조사에 불응 또는 미응답 가구는 전국 평균 약 30%인데, 이들에 대해서는 인접 가구의 조사자료로 대체함).

이와 같이 추출된 표본의 층별 가구수, 모집단(목표틀) 가구수 및 표본가구의 全標本加重值(full sample weight)는 <표 1>에 수록되어 있다. 여기서, 가구 가중치가 표본의 확률구조를 나타내지만, 본고에서 사용할 컴퓨터 프로그램에서는 전표본 가중치를 필요로 하므로 그것을 계산해 두었다.

4) 본 항의 내용은 통계청(1998) 각 권 제1장의 조사개요와 통계청(1997)을 참조하였다.

〈표 1〉 표본설계: 층별 가구수(모집단 및 표본) 및 가구 가중치

A. 행정 구역	B. 층	C. 표본 가구수	D. 가구 가중치	E. 가구 가중치 소계	F. 모집단 가구수	G. 모집단 가구 구성비	H. 표본 추출률 (%)	I. 가구 전표본 가중치
서울	서울	3,111	9.78E-05	0.30417	3,453,731	30.42	0.090	1,110.2
부산	부산	2,539	3.75E-05	0.09527	1,081,720	9.53	0.235	426.0
대구	대구	1,712	3.14E-05	0.05369	609,668	5.37	0.281	356.1
인천	인천	2,144	2.55E-05	0.05471	621,184	5.47	0.345	289.7
광주	광주	1,710	1.61E-05	0.02748	312,041	2.75	0.548	182.5
대전	대전	1,669	1.48E-05	0.02476	281,085	2.48	0.594	168.4
경기	경기 시부	1,915	6.38E-05	0.12209	1,386,239	12.21	0.138	723.9
	경기 군부	816	5.32E-05	0.04339	492,719	4.34	0.166	603.8
강원	강원 시부	782	2.41E-05	0.01881	213,593	1.88	0.366	273.1
	강원 군부	442	2.48E-05	0.01095	124,343	1.10	0.355	281.3
충북	충북 시부	776	1.96E-05	0.01519	172,468	1.52	0.450	222.3
	충북 군부	354	2.05E-05	0.00726	82,428	0.73	0.429	232.8
충남	충남 시부	636	1.45E-05	0.00921	104,575	0.92	0.608	164.4
	충남 군부	386	3.70E-05	0.01427	162,014	1.43	0.238	419.7
전북	전북 시부	751	3.05E-05	0.02287	259,704	2.29	0.289	345.8
	전북 군부	201	3.25E-05	0.00653	74,188	0.65	0.271	369.1
전남	전남 시부	609	2.51E-05	0.01531	173,804	1.53	0.350	285.4
	전남 군부	326	4.13E-05	0.01347	152,979	1.35	0.213	469.3
경북	경북 시부	698	3.65E-05	0.02551	289,604	2.55	0.241	414.9
	경북 군부	454	4.44E-05	0.02014	228,737	2.01	0.198	503.8
경남	경남 시부	1,223	5.17E-05	0.06325	718,178	6.33	0.170	587.2
	경남 군부	446	5.25E-05	0.02343	266,015	2.34	0.168	596.4
제주	제주 시부	413	1.47E-05	0.00606	68,819	0.61	0.600	166.6
	제주 군부	177	1.23E-05	0.00217	24,688	0.22	0.717	139.5
총계		24,290		1.00000	11,354,524	100	0.214	

층별 표본추출률(H열) = $100 \times \text{층의 표본가구수} / \text{층의 모집단 가구수}$,

개별 가구 가중치(D열) = $\text{층의 가구 가중치 소계} / \text{층의 표본가구수}$

층별 가구 가중치 소계(E열) = $\text{해당 층의 표본추출률의 역수} / \text{모집단 총가구수} (=11,354,540)$

가구 全標本加重値(I열) = $\text{개별 가구 가중치} \times \text{모집단 총가구수}$

= $\text{표본추출률의 역수} / \text{층의 표본가구수}$

= $\text{각 표본가구가 대표하는 모집단 가구수}$

2. 회귀계수 및 분산 추정법

복잡한 서베이 자료(survey data)에 입각한 통계적 추론에는 표본추출 이론가들의 ‘표본설계 접근법’(design-based approach)과 고전적 통계적 추론 이론가의 ‘모형(model-based) 접근법’이 오랫동안 대립적으로 발전하여 왔는데, 양자의 극단적 형태는 모집단에 대한 성격규정 및 통계적 추론대상에 대한 상이한 입장에 입각하여 상이한 통계적 추론방식을 전개한다. 그러나 1950년대 후반 이후 양자를 결합한 접근법이 발전되고 있다. 이들에 대한 자세한 설명은 아래에 인용된 참고문헌들에 맡기고, 여기서는 간략히 설명하기로 한다.⁵⁾

전국 N 개 가구들로 구성된 모집단 $U = \{U_1, U_2, \dots, U_N\}$ 에서 크기 n 의 표본 $s = \{U_{i_1}, U_{i_2}, \dots, U_{i_n}\}$ 을 추출하여 자료

$$(Y_i, X_{1i}, X_{2i}, \dots, X_{mi}), \quad i = i_1, i_2, \dots, i_n$$

을 측정오차 없이 관측한다고 하자. 표기의 편리상, 표본을 i 의 하첨자만 사용하여 $s = \{1, 2, \dots, n\}$ 으로 표기하기로 하자. 여기서, Y_i 는 i 가구의 연간 소득, X_{ji} 는 i 가구의 가구 특성 변량 X_j 의 값(예, 취업자수, 가구주 연령, 금융자산 등)이다. 표본설계는 U 의 모든 부분집합들의 집합에 대한 확률측도(probability measure) $p(\cdot)$ 으로 정의되며, 확률변수 S 가 s 값을 가질 확률을 $p(s)$, i 가구가 추출될 확

5) 경제학과의 교육과정에 통계학(및 계량경제학)이 들어 있으나 다른 사회과학분야 교육과정에 들어 있는 ‘조사방법론’을 거의 다루지 않기 때문에, 본 절은 일반 독자에게 다소 생경할지 모른다. 즉, 경제학에서는 주로 ‘단순임의 추출법(srs)’에 기초한 추정법만 다루며 실제 표본 조사에 널리 사용되는 다른 표본추출법들은 거의 다루지 않는다. 그런데, 표본추출이론 분야의 대표적 교재인 Cochran(1977)의 Sampling Technique에서도 모집단의 평균, 총계 및 비율 등을 추정하는 단순 통계량만 다룰 뿐 회귀분석이나 분산분석 등 보다 복잡한 분석적 통계량은 다루지 않는다. 최근에 서베이 원자료가 점차 많이 제공되고 있으므로, 대학원 경제학 교육에서 본절의 분석법을 다루어야 할 것이다.

Cassel and Saerndal(1977)은 표본설계접근법과 모형접근법에 관한 연구들을 엄밀하게 집약하여 정리하고 있으나 기술통계량의 추론문제에 국한되어 있으며, Nathan(1988)은 다소 덜 엄밀하기는 하지만 회귀분석, 분산분석 등 다변량 분석에서의 추론문제를 다루고 있다. 이하의 서술은 이 논의를 간략히 정리한 것이며, 그것은 줄고(1998)의 부록에 좀더 자세히 설명되어 있으나, 더 자세한 내용은 참고문헌을 보기 바란다. 또한 필자의 홈페이지 서베이자료 통계분석법(<http://business.chungnam.ac.kr/~nhoew/hc/hamain.htm>)란에 구체적 계산방법과 참고자료를 게시해 두었다.

를 $\pi_i \equiv \sum_{s=1}^N p(s)$ 로 표기하자. 그리고, 모집단의 k 번째 ($k=1, 2, \dots, N$) 단위가 표본으로 추출되면 1, 아니면 0인 확률변수 q_k 를 정의하면, $E[q_k] = \pi_k$ 이다. 모집단 전체 가구의 자료를 $n \times (m+1)$ 차원 행렬 (Y, X)로 표기하면, 표본설계 접근법에서는 모집단 회귀계수

$$B = (X'X)^{-1}X'Y \quad (1)$$

를 파라미터라 간주하며, 그 불편 추정량은

$$\hat{B} = (X'X)^{-1}X'\Pi^{-1}QY,$$

$$\text{단, } Q = \text{diag}(q_1, \dots, q_N), \quad \Pi = \text{diag}(\pi_1, \dots, \pi_N) = E[Q]$$

로 주어진다. 여기서 불편성은 표본설계-불편성(design-unbiasedness) 또는 p -불편성 개념으로서 기대치는 표본 확률측도 $p(\cdot)$ 에 입각하여 계산된다:

$$E_p[\hat{B}] = \sum_s p(s) \hat{B} = B.$$

그런데, 모집단 전체의 자료 $X'X$ 은 통상 관찰되지 않으므로 그 불편 추정량 $X'\Pi^{-1}QX$ 를 대용하여

$$\hat{\hat{B}} = (X'\Pi^{-1}QX)^{-1}X'\Pi^{-1}QY = (X_s'W_sX_s)^{-1}X_s'W_sY_s \quad (2)$$

를 사용할 수밖에 없다. 여기서 하첨자 s 는 표본으로 추출된 것만을 포함하는 부분 행렬(예컨대, Y_s 는 표본 s 에 추출된 가구들만의 소득벡터임)을 나타내며, $W_s = \text{diag}(\pi_i^{-1}; i \in s)$ 이다. 식(2)는 표본설계 가중(design-weighted) 추정량 또는 허비츠-톰슨(Horvitz-Thomson) 추정량이라 불리는데, B 에 대해 소량의 편이를 보유하므로 '근사적 불편' 추정량임이 밝혀져 있다. 그리고 \hat{B} 의 공분산 행렬과 그 불편 추정량은 각각 아래로 주어진다(Joerup and Rennermalm, 1976):

$$\begin{aligned} V[\hat{B}] &= (X'X)^{-1}X'\Pi^{-1}V[QY]\Pi^{-1}X(X'X)^{-1} \\ \hat{V}[\hat{B}] &= (X'X)^{-1}X'Q\Psi QX(X'X)^{-1} \end{aligned} \quad (3)$$

단, $V[QY] = QY$ 의 공분산 행렬, $\Psi = \{\Psi_{ij}\}$,

$$\Psi_{ij} = \begin{cases} (1-\pi_i)\pi_i^{-2}Y_i^2, & j=i \\ (\pi_i^{-1}\pi_j^{-1}-\pi_{ij}^{-1})Y_iY_j, & j \neq i \end{cases}, \quad \pi_{ij} = \sum_{s \ni i, j} p(s) = E[q_i q_j].$$

그런데, 식(3)에서 보듯 각 \hat{B}_k 의 분자와 분모는 모두 확률변수를 이루므로, 모든 \hat{B}_k 들에 대해 불편 분산 추정량을 도출하는 것은 불가능하며, 아래에서 소개될

별도의 분산 추정기법을 사용하여 계산한다.

한편, 초모형 접근법은 Y 를 고정된 값이 아니라 가설적인 초모집단(super-population)의 확률과정 또는 모형분포(model distribution: ξ)를 따르는 확률변수 벡터 $Y = (Y_1, Y_2, \dots, Y_N)'$ 이 실현된 것으로 간주하며, 이 모형분포의 파라미터를 추정하고자 한다. 예컨대, Y 가 동일-독립적 분포(i.i.d.)를 이룬다는 통상적인 계량경제학의 초모집단 모형은

$$Y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n) \quad (4)$$

으로 표현되며, 이 분포의 파라미터인 β 와 σ^2 이 추론대상이다. Nathan(1988)에 따르면, 실제로 선형 모형(4)가 타당하게 성립하고 표본설계가 모든 X 를 포함한다면 고전적 회귀분석법의 통상최소자승 추정량(OLSE)과 그 분산 추정량

$$\hat{\beta} = (X_s' X_s)^{-1} X_s' Y_s \text{ 및 } \hat{V}[\hat{\beta}] = \hat{\sigma}^2 (X_s' X_s)^{-1}$$

도 β 의 효율적인 불편 추정량이다. 그런데, 여기서 불편성은 모형-불편성(model-unbiasedness) 또는 ξ -불편성 개념으로서 기대치는 ξ 분포, 즉 식(4)의 분포에 입각하여 정의된다. 그러나, 이 추정량은 설계-불편성을 갖지 못한다.

표본설계·모형 결합 접근법은 식(1)의 B 가 확률변수로서 식(4)의 β 의 실현치라 간주하며 통계량의 분포는 표본추출 확률 $p(s)$ 와 모형분포의 결합분포를 따른다. Nathan(1998)에 따르면⁶⁾ 복잡한 표본설계에 식(3)과 같은 '동분산 초모집단 모형'을 결합한 경우, 식(2)의 표본설계 가중 추정량

$$\hat{B}_{HT} = (X_s' W_s X_s)^{-1} X_s' W_s Y_s \quad (2)$$

는 상술한 바와 같이 소량의 편이를 갖지만 설계일치(design-consistent) 추정량이며, 그 분산-공분산 추정치는

$$\hat{V}[\hat{B}_{HT}] = \hat{\sigma}^2 (X_s' W_s X_s)^{-1} X_s' W_s^2 X_s (X_s' W_s X_s)^{-1} \quad (5)$$

으로 주어진다. \hat{B}_{HT} 는 모형-불편이며, $\hat{\beta}$ 가 보유하지 않는 설계-일치성도 보유한다. 따라서, 본고에서는 허비츠-툼슨 추정량을 추정하기로 한다.

그런데, SAS, SPSS 등의 통상적인 회귀분석 프로그램으로 식(2)와 같은 회귀계

6) Nathan(1988)은 이 결합접근법을 순수한 설계접근법이라 부르며 전개하고 있지만, "그 내부에 '동분산 모형'(6')을 전제하여야 B에 대한 관심이 정당화될 수 있다"(p. 255)는 말로 시작함으로써 사실상 결합접근법임을 시사하고 있다.

수는 계산할 수 있지만, 그 분산 추정량 (3)과 (5)는 계산할 수 없다. 그것을 계산해 주는 컴퓨터 프로그램은 여러가지가 있으나, 그 계산법은 크게 複製法(method of replication)과 테일러 전개법으로 대별되며,⁷⁾ 여러 프로그램들의 성능 및 특성은 Lepkowski와 Bowles(1996) 및 Carlson(1998)에 잘 소개되어 있다. 본고에서는 복제법을 적용한 Westat사의 WesVarPC 프로그램을 사용하였다.⁸⁾

복제법은 전체 표본으로부터 표본의 설계내용을 잘 반영하도록 부분 표본(복제 표본이라 함) G 개를 추출하여 각 복제표본에서 추정치를 계산하고, 그것들과 全標本 추정치간의 편차를 이용하여 계산된다. 즉, 전체 표본자료에 입각한 파라미터 θ 의 추정치를 $\hat{\theta}$, G 개의 각 복제 표본에서 구한 θ 에 대한 추정치를 $\hat{\theta}_k$ ($k = 1, 2, \dots, G$)로 표기하면, $\hat{\theta}$ 의 복제 분산 추정치는

$$\hat{V}(\hat{\theta}) = c \sum_{k=1}^G (\hat{\theta}_k - \hat{\theta})^2$$

으로 계산된다. 여기서 c 는 복제방식에 따라 결정되는 상수로서, 본고에서 적용한 Fay 방식에서는 $c = \frac{1}{G(1-K)^2}$ 이며 $K=0.3$ 으로 설정하였다(Brick et al., 1997, p. 145). 복제표본은, 각 층에서 오직 2개의 1차 추출단위를 추출한 경우를 가정하여, 각 층별로 그 중 1개씩 추출하여 ‘半分표本’을 구성하되, 전표본가중치 계산에 적용된 것과 동일한 방식으로 이 복제표본에 가중치(복제 가중치)를 새로 부여함으로써 창출되며, H 개로 층화된 표본일 경우 표본정보의 상실이 없는 최소의 복제표본數(G)는 H 이상인 4의 최소 배수임이 밝혀져 있다. 『1996 가구소비실태조사 보고서』에서는 $H=24$ 이므로 $G=24$ 이다.⁹⁾

7) 복제방식에는 Balanced Repeated Replication(BRR), Jack-knife Repeated Replication(JRR) I 및 II, Fay법 등이 있다. Kish and Frankel(1974)은 이들 및 테일러 전개법의 장단점을 시뮬레이션을 통해 비교·검토하고 있으며, Wolter(1985)의 『분산추정법 입문』에는 보다 최근까지의 이론적 업적들이 요약 및 정리되어 있다.

8) SAS사는 장치 SAS Ver. 7에 테일러 전개법을 이용한 분산계산 프로그램(SAS/STAT package 내 PROC SurveySelect, SurveyMeans, 및 SurveyReg)을 추가할 예정이며, SPSS(Ver 8.0 for Windows, 97.12.30)는 1998년 6월 22일부터 WesVar PC를 개량한 WesVar Complex Samples 3.0을 별도로 판매하고 있다.

9) 복제표본들은 Hadamard 행렬을 이용하여 만드는데, 그 구체적 방법은 Wolter(1985) 부록 A에 자세히 설명되어 있다. 그런데 WesVarPC 프로그램에서는 반분표본 코드, 층코드 및 복제방식만 지정하면, 복제표본의 작성 등 나머지 계산과정은 자동으로 수행되어 회귀계수와 그 분산 추정치가 모두 계산된다.

복제분산을 계산하기 위해서는 먼저 상술한 반분 표본을 만들어야 한다. 『1996 가구소비실태조사 보고서』에서는 계통추출법을 적용하되 각 추출구간에서 1개의 확률적 추출번호(random start number)를 선정하여 각 구간마다 1개씩 추출하였으므로, 반분 표본을 만들 수 있도록 설계되어 있지 않다. 따라서, 현 단계에서는 조사구들을 강제로 양분하는 방법밖에 없으며,¹⁰⁾ 간단한 자의적(ad hoc) 방법은 각 층의 조사구들을 2개의 확률적 집단으로 양분하는 것이다(Wolter, 1995, pp. 131~132). 반분 표본은 원칙적으로 1차 추출단위인 ‘표본조사구들’을 2개의 확률집단으로 구분하여야 하지만, 조사구 식별번호가 통계청이 공표한 자료에 수록되어 있지 않으므로 부득이 2차 추출단위인 표본가구들을 직접 두 그룹으로 나누되, 다소 자의적이지만 조사구 순으로 배열된 표본가구들의 처음 절반을 제1집단으로 하고 그 나머지를 제2집단으로 구분하였다.¹¹⁾

Ⅲ. 지역별 가구소득 추정

1. 지역별 대표적 가구소득 추정

『1996 가구소비실태조사 보고서』의 조사단위는 가구(household)로서 주거와 생계를 같이하는 사람들의 집단으로 정의되며, 가구원 수에 따라 1인 가구(전가구의 12.9%)와 2인 이상 가구로 구분된다. 한 지역내에서도 가구들의 소득은 아주 상이

10) 최강식(1997)은 노동부의 <직종별 임금조사> 원자료를 이용한 연구에서 기업규모별 추출률의 역수 배만큼 각 표본을 복제하여 삽입한 다음, 모집단의 기업규모별 비율로 임의추출표본을 재추출하고, 중복 추출된 표본은 제거하되 가중치를 재조정하는 방식으로 단순 무작위 추출 표본을 재작성함으로써 복잡한 표본설계 자료의 문제를 해결하였다. 그러나, 이 방법은 상당한 자료의 손실을 초래할 것으로 생각된다. 또한 분석하려는 모집단의 속성(특성) 여하에 따라 단순무작위표본이 언제나 그 특성 추정에 최적인 것도 아니다.

11) 이 방식도 여전히 문제점을 안고 있으므로, 장차 통계청에서 조사구 식별번호를 제공하면 본 연구에서의 회귀분석을 다시 해볼 필요가 있다. 왜냐하면 회귀계수의 표준오차는 반분표본 설정방식에 따라 상당히 다른 값으로 산출되기 때문이다. 줄고(1998)에서는 SAS 프로그램 설정 착오로 인해 표본가구들을 가구주의 교육수준(학교 급별 내림차순 및 졸업, 중퇴, 진학의 순)에 따라 정돈한 후 전후반으로 나누어 반분표본을 구성하였는데, 이 경우 무응답 가구의 중복 등의 이유로 인해 표준오차가 너무 작게(단순 무작위 표본으로 간주하고 계산한 것보다 더 작게) 추정되었다.

하며 하나의 분포를 이루고 있다. 이 소득분포의 대표적 소득으로서 (가중)평균소득과 중위소득이 널리 사용된다.

〈표 2〉에 1996년도 지역별 가구소득의 여러가지 대표치가 추정되어 있다. 제A열은 각 지역의 가구소득 가중¹²⁾ 평균치이다. 전국 가구의 가중평균 소득은 2,392만원인데, 6대 도시와 9개 도의 15개 지역 중에서 서울의 평균 가구소득이 2,721만원으로서 가장 높고, 경북이 1,880만원으로서 가장 낮다. 서울 가구소득을 기준으로 15개 지역을 〈표 3〉의 B열과 같이 상위권, 중위권 및 하위권의 세 집단으로 구분해볼 수 있다. 여기서 두드러진 특징은, 서울가구의 평균소득이 다른 지역보다 월등히 높아 서울소득의 95~100%에 속하는 지역이 없으며, 서울 가구소득의 80% 미만인 하위권은 道지역으로만 구성되어 있다는 점이다.

〈표 2〉 지역별 대표적 가구소득

(금액단위: 천원)

지역	A. 평균소득			B. 중위소득			C. 대수소득		D. 가구원 1인당		E. 1인당 GRP	
		지수	오차		지수	오차	평균	환산 소득	소득	지수	(96년)	지수
전국	23,930	87.9	752	21,198	88.4	764	9,8792	19,520	8,011	88.9	8,571	100.5
서울	27,213	100.0	2,267	23,989	100.0	2,768	10,0413	22,956	9,014	100.0	8,526	100.0
부산	24,917	91.6	2,495	20,196	84.2	1,313	9,8580	19,111	8,062	89.4	6,863	80.5
대구	22,940	84.3	1,412	20,391	85.0	1,278	9,8631	19,207	7,515	83.4	6,652	78.0
인천	22,877	84.1	134	20,443	85.2	390	9,8740	19,418	7,238	80.3	8,357	98.0
광주	23,218	85.3	2,603	20,540	85.6	2,339	9,8423	18,813	7,609	84.4	6,953	81.5
대전	22,911	84.2	3,651	19,798	82.5	4,161	9,8551	19,056	7,728	85.7	6,819	80.0
경기	22,419	82.4	10	20,399	85.0	256	9,8550	19,053	7,584	84.1	8,460	99.2
강원	20,795	76.4	2,101	17,987	75.0	1,354	9,7050	16,399	7,375	81.8	7,748	90.9
충북	20,936	76.9	751	18,448	76.9	268	9,7252	16,734	6,971	77.3	9,568	112.2
충남	22,801	83.8	185	20,519	85.5	748	9,7789	17,657	7,831	86.9	9,659	113.3
전북	21,985	80.8	1,784	19,189	80.0	1,109	9,7803	17,681	7,137	79.2	7,815	91.7
전남	19,820	72.8	1,901	17,200	71.7	2,291	9,5746	14,394	7,102	78.8	9,574	112.3
경북	18,819	69.2	1,296	17,987	75.0	847	9,5810	14,487	7,139	79.2	9,366	109.9
경남	23,095	84.9	271	21,594	90.0	829	9,8485	18,929	7,917	87.8	11,751	137.8
제주	21,638	79.5	394	18,684	77.9	459	9,6829	16,041	7,229	80.2	7,499	88.0

(오차는 표준오차임)

- 12) 6대 도시는 각 가구가 동일한 가중치를 가지므로 가중평균과 단순평균이 일치하지만, 각 도의 경우 시부 가구와 군부 가구의 가중치가 다르므로 가중평균이 적용된다.

〈표 3〉 가구소득 수준별 지역 구분

구분	A. 서울 소득의	B. 평균소득 기준	C. 동일 가구특성하의 평균소득 기준	D. 역대수 환산 소득 기준	E. 중위소득 기준
상위권	95~100%	(서울)	(서울, 충남, 대전, 경남)	(서울)	(서울)
	90~95%	부산	대구, 광주, 부산, 경기		경남
중위권	85~90%	광주	제주, 전남, 전북, 강원, 인천, 충북, 경북		광주, 충남, 대구, 인천, 경기
	80~85%	경남, 대구, 대전, 인 천, 충남, 경기, 전북		인천, 대구, 부산, 대 전, 경기, 경남, 광주	부산, 대전, 전북
하위권	75~80%	제주, 충북, 강원		전북, 충남	제주, 충북, 강원, 경북
	70~75%	전남		충북, 강원	전남
	60~70%	경북		제주, 경북, 전남	

주: C열은 〈표 4〉에서 통계적으로 유의적인 결과에 기초하여 지역을 A년의 격차 구분에 따라 분류한 것임.

중위소득은 서베이 표본에 입각한 소득의 대표치로서 널리 사용되는데, 〈표 2〉의 B열에 계산되어 있다. 전국의 중위소득은 2,120만원이며 중위소득이 가장 높은 지역 역시 서울로서 2,399만원이고 가장 낮은 곳은 전남으로서 1,720만원(서울의 71.7%)이다. 어느 지역에서나 중위소득은 가중평균 소득보다 낮지만, 중위소득의 지역간 범위는 평균소득의 그것보다 약간 작다. 참고로, 미국 통계청(1997)이 조사한 과세전 화폐소득 기준 중위소득의 州別 분포를 보면 최저 수준인 웨스트 버지니아州的 중위소득이 가장 높은 알래스카州的 50%에 불과하여, 우리나라보다 훨씬 불균등하게 분포하고 있음을 알 수 있다.¹³⁾

소득(개인 소득이든 가구 소득이든)은 일반적으로 우측으로 긴 꼬리를 가진 형태로 분포되어 있으므로 중위소득은 평균소득보다 낮다. 소득의 이러한 분포특성 때문에 정규분포에 기초한 통계적 추정 및 검증법을 적용하기 위해 통상 대수(log) 소득을 사용한다. 이 글에서도 회귀분석시에 대수소득을 사용할 것이므로, 제C열에

13) 미국 통계청(1997)이 상주인구조사(CPS)로 전국 5만여 가구를 서베이(46,800가구는 인터뷰에 응하고 3,200가구는 불응함)하여 1996년 화폐소득을 조사한 결과임. 미국 통계청은 이 표본이 본래 미국 전체 추정을 목적으로 설계되어 있어서 주변 추정치는 그 신빙도가 떨어짐을 경고하면서, 그 오차를 줄이기 위해 2개년간 이동평균을 사용하였음을 명시하고 있으며, 특히 이동평균의 '차이' 자료를 보는 데 만족할 것을 권고함.

대수 소득의 지역별 평균을 계산해 두었다. 그리고, 그것의 실제 크기를 알 수 있도록 그것에 역대수를 취해 통상 소득으로 환산한 금액도 수록해 두었는데, 이 환산소득은 가구소득의 평균과는 상당히 다르지만 지역별 중위소득과는 대체로 유사한 수준임을 알 수 있다.¹⁴⁾ 그러나, <표 3>의 D난에서 보는 바와 같이, 환산소득 기준으로 지역을 상·중·하위권으로 구분하면 B난의 평균소득 기준의 구분과 유사하다. 참고로, D열은 가구원 1인당 지역별 평균소득으로서 최저인 충북이 서울의 78.3%이며 가구당 평균소득보다 지역간 격차가 적다. 이것은 통계청이 발표한 E열의 1996년도 1인당 (경상)지역내 총생산과 비교하기 위해 수록하였다. 물론, (지역내)총생산과 가구소득의 개념이 다르므로 직접 비교 가능한 것은 아니나, 지역별 차이(지수) 및 순위는 가구소득과 판이함을 알 수 있다.

그런데 이상의 지역별 대표적 소득은 모두 점추정치이며, <표 2>에 그 표준오차가 계산되어 있다. 이것은 WesVarPC 프로그램으로 계산한 복제분산 추정치의 제곱근인 표준오차인데, 평균소득의 표준오차는 경기의 1만원에서 대전의 365만원까지 걸쳐 있다.¹⁵⁾ 따라서, 15개 지역별 평균 가구소득에는 상당한 차이가 있지만, 그 표준오차도 상당히 크므로 지역간 평균소득의 차이가 유의적으로 존재한다고 볼 수 있는가를 통계적으로 검증할 필요가 있다. 이를 위해, j 가구 연간소득(단위: 천원) y_j 의 대수값을 지역 가변수(dummy variable) D_r (j 가구가 r 지역이면 1, 아니면 0임)를 회귀변수로 하는 모형

$$\ln y_j = \beta_0 + \sum_{r=2}^{15} \beta^r D_r + \epsilon_j \quad (6)$$

14) 가구소득이 0인 가구의 대수치는 정의되지 않으므로, 24,290개의 표본가구 중 양의 가구소득을 보유한 24,274가구에 대해 계산한 것이며, IV절의 회귀분석에서도 마찬가지이다.

15) 이 표준오차의 계산방식에는 여러가지가 있는데(Wolter, 1985, Ch. 7 참조), 이 표의 수치는 복제법으로 계산된 것이다. 참고로, 통계청이 표본가구 중 2인 이상 가구들만을 대상으로 계차법(successive difference method)으로 표준오차를 계산한 것이 있는데, 본 연구에서도 동일 표본에 대해 복제법으로 표준오차를 계산한 결과, 경기도의 경우 3.7만원으로서 통계청이 구한 값 65.3만원의 0.057배이고 대전의 경우 369.4만원으로서 통계청의 95.9만원의 3.85배이다. 두 계산법이 상이하므로 양자가 정확하게 일치하기를 기대할 수는 없으나, 이 정도의 차이는 어느 쪽의 계산방법에 문제가 있음을 시사한다. 이 차이가 반분표본 설정방식에 기인하는가는 장차 검토되어야 할 과제 중의 하나이다. 익명의 심사위원도 경기도 가구소득의 표준오차가 다른 지역과는 비교할 수 없을 만큼 작은 1만원이라는 결과에 대해 강한 의문을 제기하였다. 그런데, 경기도 중위소득의 표준오차는 전국에서 제일 작기는 하지만 충북의 그것과 유사한 크기이며 평균소득의 표준오차처럼 극적인 차이를 보이지 않는다.

를 WesVarPC 프로그램으로 추정하였는데, 그 결과가 <표 4>이다. 여기서, 서울 지역($r=1$)을 기준으로 설정하였으므로, 서울지역 가변수 D_1 은 생략되며, $r=2, \dots, 15$ 는 5개 광역시와 9개 도를 나타낸다. 먼저, 15개 시·도간 가구소득 회귀 분석에서는 부산, 대구, 인천, 광주, 대전 등 5대 도시의 가구소득은 서울가구의 소득과 유의적으로 다르지 않으며 오직 각 도의 가구소득들만 서울가구의 소득에 비해 유의적으로 작고, 특히 전남의 경우 서울 가구에 비해 46.7%나 더 낮은 것으로 추정되었다(<표 2>에서 보듯, 대수 소득 평균의 역대수 환산소득은 37.3% 낮음). 그런데, 이 모형은 그 결정계수가 3.7%에 불과하므로 모형의 설명력이 극히 낮다. 아래 제IV절에서 가구소득에 영향을 미치는 가구의 인적 및 물적 자원변수들을 도입함으로써 지역별 가구소득의 차이를 분석할 것이다.

<표 4> 지역 가변수에 대한 회귀분석

종속변수	대수소득		결정계수 0.0368		
회귀변수	회귀계수	표준오차	회귀변수	회귀계수	표준오차
상수항	10.0413	0.0839**			
부 산	-0.1833	0.1025	충 북	-0.3161	0.0906**
대 구	-0.1783	0.1053	충 남	-0.2624	0.0845**
인 천	-0.1674	0.0840	전 북	-0.2611	0.1140*
광 주	-0.1990	0.1400	전 남	-0.4667	0.1339**
대 전	-0.1862	0.1811	경 북	-0.4603	0.1303**
경 기	-0.1864	0.0847*	경 남	-0.1929	0.0923*
강 원	-0.3363	0.1172**	제 주	-0.3584	0.0955**

주: **(*)는 1% (5%) 유의수준에서 유의적임을 나타냄.

IV. 지역별 가구당 연간소득의 결정요인 분석

1. 가구소득함수

가구당 소득이 어떤 요인들에 의해 결정되는가, 특히 지역별로 어떤 차이가 있는가를 분석하기로 하자. 특정 지역 개별 가구의 소득결정 요인들은 크게 가구의 소득창출적 특성(간단히, 가구특성)과 지역의 소득창출적 특성(간단히, 지역특성)의 2

가지로 구분할 수 있다. 구체적 지역특성 변수들은 아래 4항에서 도입하기로 하고, 우선 (6)에서처럼 그것들 전체를 대변하는 변수로서 지역 가변수(D_r)를 도입하고 j 가구의 i 가구특성 변수 x_{ij} ($i=1,2,\dots,m$)를 도입한 '가구특성 모형'

$$\ln y_j = \beta_0 + \sum_{i=1}^m \beta_i x_{ij} + \sum_{r=2}^{15} \beta_r D_r + \varepsilon_j \quad (7)$$

를 추정하기로 하자.

이제, 가구소득에 영향을 미치는 가구특성들을 살펴보자. 가구소득은 가구원들이 번 근로소득, 사업소득, 부업소득, 재산소득 및 이전소득으로 세분되어 조사되어 있다. 각 소득의 전국 가구 평균 구성비는 각각 59.7%, 0.5%, 29.7%, 5.1%, 5.0%로서 근로소득과 사업소득의 합계가 가구소득의 90%을 차지하나 지역별로 상당한 차이가 있으며, 각 부문소득을 전혀 벌지 않는 가구의 비율에도 차이가 많다.

본고에서는 근로소득의 결정요인으로서 가구당 취업자수, 가구주의 성별, 연령, 교육수준, 종사상의 지위 및 배우자의 연령 변수를 도입하였다. 교육수준은 초등학교 졸업 이하, 중학교 졸업, 고등학교 졸업, 전문대졸, 대학졸, 대학원졸의 6개 등급, 종사상의 지위는 봉급자, 노무자, 자영자, 경영자, 자유업자, 무직 등 분류불능의 6개 등급으로 나누어 가변수를 도입하되 그 구성비가 가장 높은 고졸자와 노무자를 각각 기준으로 하였다.¹⁶⁾ 『1996 가구소비실태조사 보고서』에서 사업소득 관련 자료는 상술한 일반적 가구특성 이외에 사업소득 수취주체별(가구주, 배우자 및 기타) 소득금액, 가구주가 사업자인 경우 사업자 종류가 조사되어 있다. 따라서 본 연구에서는 아주 미흡하지만 사업소득의 결정요인으로서 사업자 종류를 자영자, 경영자 및 자유업자로 구분하여 가구주의 종사상의 지위 가변수 형태로 도입하였다.

재산소득은 이자소득, 배당소득, 부동산 임대소득, 기타 재산소득(저작권 및 특허 사용료 등)으로 구성되는데, 이들을 낳는 자산으로서 본 가구조사에서 조사된 것은 1996년 11월 30일 현재 가구의 각종 금융자산 총액('충저축 보유액'이라는 명칭으로 발표됨) 뿐이다. 소유 부동산 가치는 조사되어 있지 않으나, 주택에 관해 각 가구

16) 대부분의 가변수 분류는 통계청(1998)과 일치하지만, 교육수준 분류에서는 대학 2년 이상 재학중이거나 중퇴한 경우는 전문대 졸업자로 분류하고 2년 미만 재학 및 중퇴자는 고졸자로 분류하여, 이들을 모두 전문대 졸업자로 분류한 통계청의 분류와 차이가 있다. 그리고, 가구주의 성별 구분 가변수는 가구주가 여성이면 1, 남성이면 0으로 하였다.

의 주택 소유 여부, 현행 주거의 상태(입주형태, 월세 또는 월세평가액 및 규모)가 조사되어 있다. 본고에서는 재산소득의 설명변수로서 금융자산 보유액(단위: 천원)을 도입하고, 부동산 임대소득의 대리변수로서 불완전하나마 자가주택의 소유 여부(전국의 주택소유 가구 비율은 58.2%로 추정됨) 가변수를 도입하였다.¹⁷⁾ 위에서 언급한 각 가구특성 변수들의 지역별 평균치는 부록에 수록해 두었다.

2. 가구특성 모형의 추정

가구특성 모형(7)의 설계일치 추정치는 <표 5>에 수록되어 있다. 이 회귀모형의 결정계수는 61%로서 횡단면 분석결과인 점을 감안하면 상당히 높은 편이며, 대전, 충남, 경남 이외의 지역 가변수들과 자유업자 이외의 가구특성 변수들이 모두 5% 유의수준에서 유의적이다.

이 추정결과에 따르면, 다른 조건이 동일할 때 취업자수가 1명 증가하면 가구소득은 전국 평균적으로 26.7% 상승한다. 가구주 연령이 증가함에 따라 39.3세까지는 가구소득이 상승하지만 그 이후로는 하락하는데, 배우자의 연령의 경우 70.3세까지 가구소득이 상승한다.¹⁸⁾ 그리고, 가구주가 여성인 가구는 통상 어머니나 할머니가 손자녀와 함께 사는 母子 가구인데, 가구주가 남성인 가구에 비해 가구소득이 11.5% 더 낮다. 다른 조건이 동일할 때 가구주의 학력에 따라 가구소득은 꾸준히 상승하는데, 고교졸업자 가구를 기준으로 할 때 초등학교 이하 학력 가구주의 가구소득은 27% 낮고 중졸자 가구는 11% 낮으며, 전문대졸 가구는 2.2%, 대졸자 가구는 13%, 그리고 대학원 졸업자 가구는 23% 더 높다. 가구주의 종사상의 지위 측면에서는 노무자 가구를 기준으로 할 때, 봉급자 가구의 가구소득은 13% 높고, 사업자 가구 중의 자영자 가구는 12%, 경영자 가구는 37% 더 높다. 그러나, 사업자

17) 소유 주택에 거주하지 않는 가구의 주택 임대료는 재산소득에 합산되어 있으나, 자가 거주 가구의 임대료 평가액은 별도로 조사되어 있으나 재산소득에 합산되어 있지 않다. 따라서 후자를 가구소득에 합산한 후, 주택의 임대료 및 귀속 임대료를 설명변수로 도입할 수도 있으나, 임대료는 지역별로 편차가 많으며 지역의 고유한 특성을 반영한다고 볼 수 있으므로 지역특성을 포착하는 지역 가변수와 부분적으로 중복되므로 도입하지 않았다.

18) 배우자 연령을 x 로 표기하고 상하첨자를 생략할 때, $\frac{d \ln y}{dx} = \left(\frac{dy}{y} \right) \left(\frac{1}{dx} \right) = 0.008980 - 2 \times 6.33861 \times 10^{-5} x \geq 0$ 을 풀면 $x \leq 70.3$ 이다. 익명의 심사위원이 지적한 바와 같이, 이 결과는 상식적으로 잘 이해되지 않으며 장차 연구할 과제로 남겨 둔다.

〈표 5〉 가구특성 모형(7)의 추정 결과

종속변수	대수소득		결정계수	0.6110	
회귀변수	회귀계수	표준오차	회귀변수	회귀계수	표준오차
상수항	8.4331	0.127**			
부 산	-0.0968	0.023**	배우자 연령	0.0090	0.001**
대 구	-0.0622	0.024*	연령제곱	-6.39E-05	0.000**
인 천	-0.1240	0.022**	여성 가구주	-0.1154	0.011*
광 주	-0.0753	0.027**	가구주 교육	(고졸 기준)	
대 전	-0.0969	0.069	초등교 미만	-0.2693	0.014*
경 기	-0.0989	0.026**	중졸	-0.1137	0.010*
강 원	-0.1239	0.037**	전문대졸	0.0218	0.008**
충 북	-0.1372	0.027**	대학졸	0.1304	0.016*
충 남	-0.0658	0.041	대학원졸	0.2300	0.035*
전 북	-0.1182	0.025**	가구주 지위	(노무자 기준)	
전 남	-0.1132	0.044*	봉급자	0.1306	0.012*
경 북	-0.1498	0.052**	자영자	0.1231	0.011*
경 남	-0.0650	0.035	경영자	0.3696	0.035*
제 주	-0.1130	0.033**	자유업자	0.1639	0.051
취업자수	0.2674	0.009**	무직 및 기타	-0.0465	0.022*
가구주 연령	0.0429	0.007**	금융자산	5.68E-06	0.000**
연령제곱	-5.45E-04	0.000**	주택소유	0.1847	0.009**

주: **(*)는 1%(5%) 유의수준에서 유의적임을 나타냄.

가구 중 자유업자의 소득이 노무자 가구소득과 동일하다는 가설은 5% 유의수준에서 기각되지 않는다(유의확률이 5.1%이므로, 간신히 기각되지 않음). 다음으로, 재산관련 변수로서 자신의 주택을 소유한 가구의 소득은 무주택 가구에 비해 18.5% 더 많다. 그리고, 전국 평균 가구소득에서 계산하면 금융자산이 100만원 증가할수록 가구소득은 13.6만원 증가하는 것으로 나타났는데, 참고로 1996년 12월 평균 회사채 수익률(장외시장 3년물)은 12.57%였다.

상술한 바와 같이 가구특성 요인들에 따라 지역별 가구소득이 유의적으로 달라지는데, 제반 가구특성들의 지역별 차이에 기인하는 소득차이는 ‘가구특성 부존효과’로 볼 수 있다. 그런데, 가구특성들이 모두 동일한 가구들이더라도 그 가구가 위치한 지역에 따라 여전히 가구소득이 다를 수도 있는데, 이러한 ‘순수한 지역효과’는 본 모형의 지역 가변수들에 의해 포착되며 그것들은 대전·충남·경남을 제외하면

모두 5% 유의수준에서 유의적이다. 물론, 가구 평균소득이 가장 높은 서울을 기준 지역으로 설정하였기 때문에, 각 지역 가변수의 회귀계수는 모두 음으로 추정되었다. 따라서 이것은 각 지역의 서울지역에 대한 상대적 지역적 불리성을 측정한다.

〈표 5〉의 추정 결과는, 동일한 가구특성을 가진 가구들을 비교할 때 예컨대 경북 가구소득은 전국에서 가장 낮아 서울 가구소득에 비해 15% 낮음을 의미한다. 그리고, 경북을 비롯하여 충북, 인천, 강원, 전북, 전남, 제주 가구소득은 서울 가구소득에 비해 10% 이상 더 낮다. 그런데, 대전, 충남, 경남 등 3지역의 지역 가변수들은 그 계수가 비록 음이지만 5% 유의수준에서 유의적이지 않다. 이것은 이 지역의 가구소득이 서울 가구소득과 통계적으로 유의적인 차이가 없음을 의미한다. 즉, 가구특성들이 동일하다면 이 지역들에 거주하는 가구의 소득은 서울가구의 소득과 유의적으로 다르지 않다. 이러한 지역효과에 의한 지역별 소득격차의 비율에 따라 지역들을 분류한 것이 〈표 3〉의 C열에 정리되어 있는데, B열의 평균소득 전체의 격차에서 가구특성 부존효과를 차감한 것이기 때문에 B열보다 격차 비율의 폭이 상당히 적게 나타난다.

3. 지역간 소득격차의 요인분해

모형(7)을 추정한 〈표 5〉에서 가구특성 변수(18개) 및 지역 가변수(14개)의 회귀계수는, 통계적으로 유의적인 경우, 각 변수가 개별적으로 1단위씩 증가하거나 각 가변수에 해당하는 가구일 때, 통계적으로 유의적인 가구소득의 평균적인 증가율을 말해 준다. 이것은 가구소득이 그 가구의 가구특성 부존량들과 거주지역에 따라 달라짐을 의미한다. 따라서, 이 결과를 이용하면 〈표 2〉에 제시된 지역별 가구 평균소득 격차를 가구특성 요인(18개 가구특성 변수들 전체의 결합 효과)과 지역특성 요인(지역 가변수로 포착됨)의 2가지로 크게 분해할 수 있다. 그리고, 가구특성 요인은 여러 개로 세분된 등급변수들을 하나로 묶어서 8개의 가구특성 종류별 세부요인으로 분해할 수 있다.

이를 위해, 〈표 2〉에서 지역별 평균소득을 계산할 때 적용한 것처럼, 각 지역에서 가구특성의 평균치(\bar{x}_i)를 보유한 가상적인 가구를 그 지역의 '평균가구'라 정의하자. 그러면, 모형(7)의 추정결과를 이용한 각 지역 평균가구의 가구소득 추정치는

$$\ln y^1 \equiv \beta_0 + \sum_{i=1} \hat{\beta}_i \bar{x}_i^1 \quad (\text{서울: } r=1), \quad (8a)$$

$$\ln y^r \equiv \beta_0 + \sum_{i=1} \hat{\beta}_i \bar{x}_i^r + \beta^r D_r \quad (\text{서울 이외: } r=2, \dots, 15) \quad (8b)$$

로 주어진다. 이것은 대수값이므로, 알기 쉽게 역대수를 취해 구한 환산소득을 <표 6>의 A행(추정 가구소득)에 표시해 두었다. A의 a행은 <표 5>에 수록된 모든 설명 변수들을 식(8) 우변에 포함시켜 계산한 것이고, b행은 5% 유의수준에서 유의적인 설명변수들(비유의적인 대전, 충남, 경남의 지역 가변수와 자유업자 가변수 제외)만을 포함시켜 계산한 것이다. 19) 20)

(8)로부터 각 지역의 서울대비 대수소득 격차는

$$\ln y^r - \ln y^1 = \sum_{i=1} \hat{\beta}_i (\bar{x}_i^r - \bar{x}_i^1) + \beta^r \quad (9)$$

으로 표시되며, <표 6>의 B행에 계산되어 있다(서울 평균가구 추정소득은 a행의 경우 22,956천원, b행의 경우 22,935천원임). 식(9) 우변의 각 항은 아래에서 설명하는 바와 같이 각각 r 지역 평균가구의 '가구특성 부존량 효과'와 '순수한 지역특성 효과'로 해석될 수 있다.

먼저, 식(9) 우변의 첫 항은 " r 지역의 추정소득식 (8b)에서 지역효과를 나타내는 β^r 항을 무시한 경우(기준 지역인 서울의 지역 가변수는 0으로 설정되었기 때문에 서울과 동일한 지역특성을 보유한다고 가정하는 것과 동일함)의 소득"

19) 이 추정소득(대수치)을 <표 2>의 지역별 대수소득 평균치와 비교하기 위해 전자의 후자에 대한 비율을 계산해 보면, b행 방식의 대전, 충남, 경남 경우에만 101%이고 그 외 모두 100%로서 양자가 매우 근사하다.

20) 초고에서는 b행 방식으로만 계산하였으나, 익명의 심사위원(들)은 그것이 오류이며 a행 방식처럼 모든 설명변수를 포함시키는 것이 타당하다고 지적하였다. 그런데, 회귀모형(7)의 추정 결과(<표 5>) 중, 예천대, 충남 지역 가변수의 "회귀계수가 통계적으로 유의적이 아니라는 것"은 "다른 조건들(각종 가구 특성들)이 동일할 때 충남지역 가구소득이 서울가구 소득과 유의적으로 다르지 않음"을 의미한다. 이 '통계적 결론'을 절대 금액 단위로 바꾸어 표현하고자 한 것이 <표 6>인데, 거기서 충남 평균가구의 서울 대비 소득격차 총액 -412만원은 "다른 조건들(각종 가구 특성들)이 다르기 때문에 발생한 것"이며 "단지 거주지역이 서울이 아니라 충남이기 때문에 발생한 것은 아니라는 것"으로 해석된다. 바로 이러한 관점에서 계산한 것이 b행이다. 이에 반해, a행 계산은 <표 6>에서 충남 평균가구의 서울 대비 소득격차 총액 -530만원 중 -133만원이란 "다른 조건들(각종 가구 특성들)이 동일할 때 거주지역이 서울이 아닌 충남이기 때문에 발생한 것"으로 해석함을 의미한다. 이것은 상술한 '통계적 결론'과 상치되는 것으로 생각된다. 두 경우를 모두 계산해 두었으므로, 독자들은 관점에 따라 타당하다고 생각되는 바를 취할 수 있을 것이며, 이 점에 관한 다른 공개적 논증 및 논평은 언제나 환영하는 바이다.

$$\widehat{\ln y}_0^r \equiv \widehat{\beta}_0 + \sum_{i=1} \widehat{\beta}_i \bar{x}_i^r$$

에서 비교기준이 되는 서울 평균가구의 소득 $\widehat{\ln y}^1$ 를 뺀 차이와 같다:

$$\widehat{\ln y}_0^r - \widehat{\ln y}^1 = \sum_{i=1} \widehat{\beta}_i (\bar{x}_i^r - \bar{x}_i^1). \quad (10)$$

따라서, 이것은 r 지역 평균가구의 각종 가구특성 부존량이 서울 평균가구의 그것과 다르기 때문에 발생하는 소득 차이를 나타내며, ‘ r 지역의 가구특성 부존효과’라 부를 수 있다. 이것은 <표 6>의 E행에 계산되어 있고, 전체 소득 격차 (9)에 대한

<표 6> 지역별 가구특성 부존효과와 순수 지역효과

(단위: 1,000원)

	부산	대구	인천	광주	대전	경기	강원
A. 추정 가구소득 a	19,081	19,197	19,410	18,813	19,056	19,053	16,399
b	19,048	19,181	19,405	18,804	20,965	19,034	16,375
B. 서울대비 격차 a	-3,874	-3,759	-3,546	-4,143	-3,900	-3,903	-6,556
b	-3,887	-3,755	-3,530	-4,131	-1,971	-3,901	-6,560
C. 지역특성효과 a	-2,028	-1,308	-2,621	-1,567	-2,030	-2,070	-2,415
b	-2,026	-1,307	-2,620	-1,566	0	-2,068	-2,412
D. 구성비(%)	52	35	74	38	52(0)	53	37
E. 가구특성효과 a	-1,846	-2,451	-925	-2,576	-1,870	-1,832	-4,142
b	-1,861	-2,448	-910	-2,565	-1,971	-1,833	-4,148
F. 구성비(%)	48	65	26	62	48(100)	47	63
	충북	충남	전북	전남	경북	경남	제주
A. 추정 가구소득 a	16,734	17,657	17,681	14,394	14,474	18,940	16,041
b	16,714	18,814	17,638	14,368	14,468	20,191	16,032
B. 서울대비 격차 a	-6,222	-5,299	-5,275	-8,561	-8,482	-4,016	-6,915
b	-6,221	-4,121	-5,297	-8,567	-8,468	-2,744	-6,904
C. 지역특성효과 a	-2,700	-1,329	-2,388	-2,076	-2,755	-1,357	-2,180
b	-2,697	0	-2,384	-2,074	-2,753	0	-2,178
D. 구성비(%)	43	25(0)	45	24	32	34(0)	32
E. 가구특성효과 a	-3,522	-3,970	-2,886	-6,485	-5,727	-2,659	-4,735
b	-3,524	-4,121	-2,913	-6,493	-5,715	-2,744	-4,726
F. 구성비(%)	57	75(100)	55	76	68	66(100)	68

주: a=‘모든 회귀변수를 포함’한 경우이고, b=‘비유의적 변수 제외’한 경우이며, D와 F행에서는 두 방식의 구성비가 다른 지역에 대해서만 b방식의 값을 괄호 안에 표시함.

구성비가 F행에 나와 있다.²¹⁾

다음으로, r 지역의 순수한 지역효과는 “ r 지역의 평균가구가 기준지역인 서울의 평균가구와 동일한 가구특성 부존량(\bar{x}_i^1)을 보유한다면 갖게 될 소득”

$$\ln y_c^r \equiv \hat{\beta}_0 + \sum_{i=1}^n \hat{\beta}_i \bar{x}_i^1 + \hat{\beta}^r, \quad r = 2, \dots, 15$$

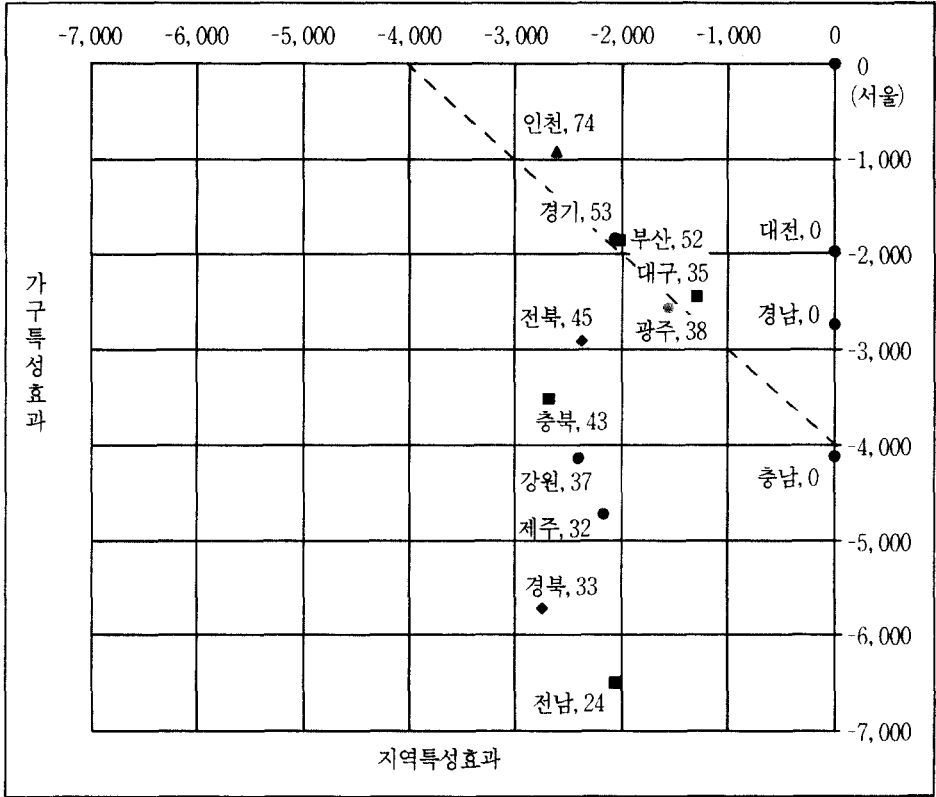
에서 기준이 되는 서울가구의 소득을 뺀 차이 $\ln y_c^r - \ln y^1 = \hat{\beta}^r$ 로 측정할 수 있는데, 이것은 식(9) 우변의 둘째 항과 일치한다. 이것은 <표 6>의 C행에 계산되어 있고, 전체 소득 격차 (9)에 대한 구성비가 E행에 나와 있다.

<표 6>에는 각 지역별 평균가구의 서울가구 대비 소득격차를 이상의 두 효과별로 나누어 그 절대치와 구성비가 a와 b의 양 경우로 나누어 함께 표시되어 있어 여간 복잡하지 않다. <표 6>의 b방식 결과를 두 효과의 절대적 크기를 나타내는 평면에 알기 쉽게 도시한 것이 <표 7>이다. 이 표에서 소득격차 계산의 기준이 되는 서울 평균가구는 원점인 우상 모서리에 위치한다. 그리고, 각 지역의 서울대비 총소득격차의 절대금액은 等總隔差線(예, 표에서 우하향하는 빗금친 선은 -400만원 등총격차 선임)이 원점에서 떨어진 거리로 측정되는데 그것은 지역별로 약 200만~860만원의 범위에 걸쳐 있다. 그 중 지역특성 효과(횡축)의 절대 금액은 0~275만원의 범위에 걸쳐 있는데, 각 지역은 그 크기에 따라 3개 집단으로 구분된다: ① 서울에 비해 불리한 지역특성 효과가 없는 대전·충남·경남, ② 지역특성 효과가 130만~160만원인 대구와 광주, ③ 지역특성 효과가 200만~275만원인 그 외의 9개 시·도 지역.

이에 비해, 가구특성 효과의 절대 금액은 인천의 91만원에서 전남의 649만원까지 보다 폭넓게 분포하고 있다. 그런데 서울대비 소득격차 총액이 530만원 이상인 전북, 충북, 강원, 제주, 경북, 전남에서는 모두 지역특성효과가 200만~275만원이며, 그 나머지는 가구특성 부존량 부족에 기인한다. 이들 중, 소득격차가 650만원 이상인 강원, 제주, 경북, 전남 지역에서는 가구특성 부존량 효과가 63% 이상(지역특성 효과가 37% 이하)이다. 따라서 서울대비 소득격차가 큰 주요인은 지역 가구들의 가구특성 부존량이 부족한 것이라고 말할 수 있다.

21) 식(9) 우변은 대수단위로 측정한 두 효과의 합계이므로, 역대수로 전환할 경우 두 효과 중 어느 것을 먼저 계산하느냐에 따라 그 값이 크게 달라진다. 따라서, 대수단위로 측정한 두 효과의 구성비를 총격차인 제B행에 곱하여 E행과 F행의 절대 금액을 계산하였다. 또한, 대부분의 지역에서 a와 b경우의 구성비가 거의 동일하므로 하나로 표시하였으며, 크게 다른 3지역만 괄호 안에 b경우의 구성비를 표시해 두었다.

〈표 7〉 가구특성효과 및 지역효과별 지역 분포



주: b방식 기준. 좌표축 단위: 천원, 지역명 우측의 숫자는 지역특성효과와 구성비(%).

〈표 8〉은 가구특성 부존효과를 다시 가구특성 변수별로 세분하여 계산한 것이다.

이것은 식 (10) 을 x_i 변수의 기여비율 $\frac{\hat{\beta}_i(\bar{x}_i^r - \bar{x}_i^1)}{\sum_i \hat{\beta}_i(\bar{x}_i^r - \bar{x}_i^1)}$ 에 따라 분해하여 계산한

것인데, 가구주 교육변수처럼 여러 개의 가변수들로 구성된 것들은 이 비율들을 더하여 1개의 변수로 통합하였다.

가구특성 중, 지역에 따라 100만원 이상의 소득차이를 설명하는 비교적 영향력이 큰 가구특성은 취업자수, 가구주의 연령 및 교육수준이다. 전남과 경북의 가구당 평균 취업자수는 1.16~1.17명으로서 서울의 1,565명에 비해 현저히 낮기 때문에 취업자수 부족에 따른 소득격차가 크다. 그리고, 강원과 전남의 가구주 평균 연령은 47~48세로서 서울의 41.5세보다 높고 또한 가구소득이 최고조에 이르는 39.3

세 이상이므로 가구주 연령이 고령이어서 소득격차가 큰 편이다. 경기와 경남을 제외한 도 지역들에서는 가구주 교육수준이 서울에 비해 낮은 데 따른 소득격차가 100만원 이상이다. 가구 특성 중 자가주택 소유자 비율은 다른 대부분의 가구특성과는 달리 경기(48.2%)와 제주(42.9%)를 제외하고는 서울(50.7%)보다 타지역에서 오히려 높으므로 가구특성 부존량 ‘부족’ 효과가 음의 값이 아닌 양의 값으로 나타난다.

금융자산 부족효과는 지역별로 27만~83만원의 범위에 걸쳐 있는데, 제주만 2.5만원으로 아주 작다. 이것은 가구당 평균 금융자산 보유액이 서울의 2,053만원에 비해 제주는 2,030만원으로서 비교적 높으며, 다른 지역들에서는 1,250만~1,754만원 수준이기 때문이다.

〈표 8〉 가구특성 부존량 효과의 세분

(단위: 천원)

	부산	대구	인천	광주	대전	대전*	경기	강원	충북
취업자수	-31	-416	-265	-725	-681	-650	-537	-946	-786
가구주 연령	-212	-219	85	-530	-299	-285	140	-1,224	-859
배우자 연령	95	-89	70	-180	81	78	-3	76	-17
여성가구주	-84	-60	117	-138	-52	-50	61	-93	-103
가구주 교육	-838	-715	-561	-427	-431	-411	-449	-1,502	-1,293
가구주 지위	-369	-324	-366	-272	-302	-277	-339	-525	-543
금융자산	-544	-658	-517	-353	-567	-541	-607	-565	-383
주택소유	122	33	527	60	280	268	-99	630	460
	충남	충남*	전북	전남	경북	경남	경남*	제주	
취업자수	-854	-829	-558	-1,495	-1,445	-701	-680	-779	
가구주 연령	-981	-952	-664	-1,627	-1,194	-194	-188	-1,387	
배우자 연령	-169	-164	96	-358	-375	-175	-169	-500	
여성가구주	-136	-132	-130	-194	-123	-7	-7	-246	
가구주 교육	-1,622	-1,574	-1,176	-1,708	-1,438	-769	-746	-1,032	
가구주 지위	-452	-409	-377	-686	-725	-563	-544	-479	
금융자산	-453	-440	-590	-833	-714	-554	-538	-25	
주택소유	546	530	487	407	300	220	213	-277	

주: b방식 기준이며, *표시된 것은 a방식임.

4. 지역 특성 효과의 요인

위에서 지역특성을 단지 지역 가변수로 포착하는 모형 (7) 을 추정한 결과, 일부 지역의 경우 소득격차의 50% 이상을 지역 가변수가 설명하고 있음을 확인하였다. 그렇다면, 이 지역 가변수에 의해 통계적으로 포착된 지역특성 효과는 구체적으로 어떤 경제적 변량들에 의해 발생하는가?

이에 관한 기존의 연구는 아직 없는 것 같다. 그러나, 몇 가지 일반적 요인들은 생각해 볼 수 있다. 우선, 지역별 임금률이나 이자율 등 요소가격을 설명 요인으로 삼을 수 있겠지만, 노동시장이나 자본시장이 지역별로 분리되어 있지 않다고 보면 이것들은 적절한 요인이 될 것 같지 않다. 따라서, 가구소득을 증대시키는 지역의 고유한 경제적 특성은 그 지역의 경제적 기반 내지 성장 잠재력을 나타내는 구조적 및 장기적 경제변수들 중에서 발견할 수 있을 것으로 기대된다. 본 연구에서는 일차적 시도로서 1996년도 지역내총생산 자료에 입각한 산업구조 변수들(지역별 광공업 비율, 3차산업 비율, 3차산업의 시·도별 구성비), 지역 시장의 크기 내지 성장 잠재력 변수(지역별 지역내총생산 점유율, 지역내총생산 성장률, 인구밀도), 지역 노동시장 기본 변수(노동참가율, 실업률) 등을 시도하였다.

r 지역의 이러한 지역특성 변수들을 $x_k^r (k=1, 2, \dots, g)$ 로 표기하면, 가구특성 및 지역특성 변수들을 도입한 가구소득 추정 모형은

$$\ln y_j = \beta_0 + \sum_{i=1}^m \beta_i x_{ij} + \sum_{k=1}^g \beta_k x_k^r + \varepsilon_j \quad (11)$$

와 같이 설정할 수 있으며, 추정 결과는 <표 9>에 수록되어 있다. 그런데, 상술한 지역특성 변수들을 모두 도입하여 식(11)을 추정하면, 추가된 지역특성 변수들은 그 어느 것도 통계적 유의성을 갖지 못하였으므로 그 결과는 수록하지 않았다.

추가된 지역특성들이 모두 유의성을 갖는 경우는 <표 9>에 수록된 4가지 모형뿐이며, 또한 가구특성변수들은 모두 유의적이다. 모형(가)는 지역내총생산(GRP)의 점유율과 지역별 2차산업 비율(%: 각 지역에서 전산업을 100%로 하는 2차산업의 구성비)이 지역특성을 잘 반영함을 보여주는데, 전자가 1%포인트 증가할수록 그 지역 가구소득이 0.4%씩 증가함을 의미한다. 그런데, 후자가 1%포인트 증가함에 따라 가구소득이 0.1%씩 하락한다는 것은 다소 의외의 결과로 생각되지만, 이 점은 모형(다)와 (라)에서도 약간 다른 방식으로 입증된다. 즉, 모형(다)는 지역별

3차산업 비율(%)이 1%포인트 증가함에 따라 그 지역 가구소득이 0.2%씩 상승함을 말해 주며, 모형(라)는 3차산업의 시·도별 구성비(%: 전국 3차산업을 100%로 하는 시·도별 구성비)가 각각 1%포인트 증가함에 따라 그 지역 가구소득이 0.4% 상승함을 말해 준다. 모형(나)는 지역내총생산 점유율만 도입할 경우 그 효과가 더욱 크게 나타남을 보여준다. 그런데, 이 모형들의 결정계수는 지역 가변수만을 사용한 모형(7)보다 오히려 약간 낮다.

〈표 9〉 지역특성 모형의 추정

모형	(가)	(나)	(다)	(라)
결정계수(%)	60.99	60.91	60.87	60.99
상수항	8.343**	8.285**	8.224**	8.300**
취업자수	0.268**	0.270**	0.269**	0.268**
가구주 연령	0.043**	0.044**	0.043**	0.043**
연령제곱	-0.001**	-0.001**	-0.001**	-0.001**
배우자 연령	0.009**	0.009**	0.009**	0.009**
연령제곱	0.000**	0.000**	0.000**	0.000**
여성 가구주	-0.115**	-0.113**	-0.117**	-0.114**
초등교 미만	-0.272**	-0.273**	-0.275**	-0.272**
중졸	-0.114**	-0.114**	-0.115**	-0.114**
전문대졸	0.023**	0.024**	0.023**	0.023**
대학졸	0.132**	0.135**	0.134**	0.133**
대학원졸	0.233**	0.239**	0.235**	0.234**
봉급자	0.130**	0.131**	0.131**	0.131**
자영자	0.123**	0.126**	0.122**	0.124**
경영자	0.372**	0.378**	0.376**	0.373**
자유업자	0.162**	0.164**	0.159**	0.164**
무직 및 기타	-0.046*	-0.044*	-0.044*	-0.045**
금융자산	0.000**	0.000**	0.000**	0.000**
주택소유	0.184**	0.182**	0.182**	0.184**
지역별 GRP 점유율	0.004**	0.005**		
지역별 2차산업 비율	-0.001*			
지역별 3차산업 비율			0.002**	
3차산업 시·도별 구성비				0.004**

주: **(*)는 1%(5%) 유의수준에서 유의적임을 나타냄.

V. 결 론

이 글에서는 1996년도 가구소비실태조사 자료에 의거하여 전국의 15개 시·도별 가구당 평균 소득을 추정하고 그 격차의 요인을 분석하였다. 주요 가구특성들을 도입하여 가구소득함수를 추정한 결과, 각 지역들의 가구소득이 서울가구 소득보다 작은 차액(350만~860만원)의 약 1/2이상(단, 인천은 26%)이 가구특성들의 부존량 차이로 인한 것이며, 그 나머지만 동일한 가구특성을 보유한 가구로서 단지 거주지역이 다르기 때문에 나타나는 순수한 지역적 소득격차로 나타났다. 또한, 지역별로 상이하기는 하지만 대체로 대규모 소득격차를 유발하는 가구특성 변수는 가구당 취업자수, 가구주의 교육수준 및 연령이었다.

순수한 지역특성에 기인하는 가구소득 격차가 절대 금액상으로는 지역에 따라 130만~280만원이지만, 소득격차 총액에 대한 구성비는 지역에 따라 대체로 0~53%를 점하며 인천의 경우 예외적으로 74%나 차지한다. 이 지역특성 효과는 지역 내총생산 점유율이 높고 지역별 2차산업 비율이 낮은(3차산업 비율이 높은) 지역일 수록 높으며 가구소득을 상승시키는 것으로 나타났다.

이상의 분석은 중앙 및 지방 정부의 지역 균형발전 정책에 간접적으로 관련된 몇 가지 시사점을 제시해 준다. 첫째로, 각 시·도별로 가구소득 격차의 주요 요인이 가구특성 효과와 지역특성 효과 측면에서 상이하므로 각 지역에 적합한 대책을 수립하고 시행할 필요가 있다. 예컨대, 인천의 경우 타 지역과는 달리 소득격차 총액은 전국에서 가장 작지만, 그것의 74%가 순수한 지역특성 효과로 밝혀졌으므로 불리한 지역특성의 개선에 주력하고, 충남·경남·대전의 경우 소득격차 전부가 가구특성 부존량의 부족에 기인하므로 가구특성을 확충할 필요가 있다. 둘째로, 소득격차가 큰 대부분의 지역에서는 가구특성 효과도 큰데, 이 지역들에서는 가구특성별로 취약점을 보완할 필요가 있다. 예컨대, 전남과 경북의 경우 가구당 취업자수가 서울 등 타 지역에 비해 현저하게 낮고 이로 인한 소득격차가 상당히 크므로, 그 원인을 조사하여 적절한 대책을 마련한다거나, 경기와 경남을 제외한 도지역에서는 가구주 교육수준이 낮은 것이 소득격차의 중요 요인이므로 장차 도민의 고등교육기관 진학을 촉진하고 또한 고등교육 이수자의 도내 정착을 유도하는 방안을 강구할 수 있을 것이다. 셋째로, 순수 지역특성효과는 3차산업 비중이 높을수록 크게 작용하므로, 3차산업 육성에 주력하는 것이 간접적으로 당해 지역 가구소득을 높이는

방법이다.

그런데, 본 연구는 몇 가지 한계를 지니고 있다. 우선, 1개년의 횡단면 조사자료에 입각한 정태분석이기 때문에, 이것을 기초로 도출된 동태적 소득증대(경제성장)에 관한 정책적 함의는 그만큼 제한적일 수밖에 없으며 장차 자료의 누적을 통해 더욱 연구되어야 할 것이다. 특히, 순수한 지역효과를 발생시키는 지역경제 기반에 관해서는 장차 이론적 연구와 더불어 실증적 연구가 진행되어야 할 것이다. 둘째, 가구소득이 의미를 갖는 것은 가구의 생활수준과 직결되어 있기 때문인데, 가구의 궁극적 물질적 복지와 이를 위한 정책적 대책을 개발하기 위해서는 가구 가치분소득의 격차를 탐구할 필요가 있다. 더 나아가, 지역별 가구복지를 더욱 엄밀하게 파악하기 위해서는 지역의 자연적 문화적 환경 등의 비금전적 효용(또는 비효용) 가치도 어느 정도 고려하여야 할 것이다. 사실상, 주민들의 거주이전의 자유가 보장되어 있는 사회에서 사람들이 합리적으로 거주 지역을 선택하는 티부(Tibou) 효과가 작용한다면, 이동이 완료된 균형상태에서는 본 연구에서 계산된 불리한 금전적인 순수한 지역효과가 그 지역의 유리한 비금전적 이득에 의해 완전히 상쇄된다고 볼 수 있을 것이다. 셋째, 본 연구의 회귀분석에서 회귀변수의 유의성 판정은 그 표준오차에 의해 좌우되며, 표준오차는 제II절에서 언급한 바와 같이 반분표본 설정방식에 따라 크게 달라진다. 장차 표본설계 단계에서부터 반분표본을 설정하여 정확한 표본오차를 계산하여야 할 것이다.

■ 參考文獻

1. 강석훈, "서베이 데이터와 집계 데이터의 비교 연구: 가구소비실태조사와 국민계정을 중심으로", 한국계량경제학회, 1998년도 학술발표대회 논문집 제1권, pp. 319~335.
2. 김대환, "불균등 산업화와 지역격차", 『경제민주화의 길』, 비봉출판사, 1992.
3. 권순원·고일동·김관영·김선웅, 『분배불균등의 실태와 주요정책과제』, 한국개발연구원, 1992.
4. 노응원, "우리나라 지역별 가족가구 소득의 격차 및 요인분석", 1998, 통계청 세미나(98.9.1) 발표문. 통계청, 『지역별 소득분배 및 생활실태 분석』(1998. 10)의 제II장에 수록됨. 홈페이지 <http://business.chungnam.ac.kr/~nhoew/research/reg-dif-fy-sem.exe>에서 내려받을 수 있음.
5. 반성환, "농가소득의 결정요인과 소득분배", 주학중 편, 『한국의 소득분배와 결정요인(상)』, 한국개발연구원, 1979, pp. 111~179.

6. 이은우, “도농간 소득격차의 실태와 원인”, 『경제발전연구』, 1995년 12월(창간호), pp. 249~267.
7. 이정우, 『소득분배론』, 비봉출판사, 개정판, 1997a
8. ———, “한국의 지역적 경제력 격차: 장기추세와 한일비교”, 『경제발전연구』, 제3권, 1997b, pp. 291~312.
9. 주학중, “계층별 소득분배의 추계와 변동요인”, 주학중 편, 『한국의 소득분배와 결정요인(상)』, 한국개발연구원, 1979, pp. 73~107.
10. 최강식, 『기술진보와 노동시장의 변화』, 한국노동연구원, 1997.
11. 통계청, “표본개편 결과보고서(가구부문)”, 1997. 12.
12. 통계청, 『1996 가구소비실태조사보고서』, 제1-4권, 1998.
13. Brick, J. M. et al., *A User's Guide to WesVarPC*, Ver. 2.1, WestVar, Inc., 1997.
14. Brogan, “Pitfalls of using standard statistical software packages for sample survey data” (http://www.fas.harvard.edu/~stats/survey-soft/donna_brogan.html), P. Armitage and T. Colton (eds.), *Encyclopedia of Biostatistics*, John Wiley, 1998에 수록 예정.
15. Carlson, B. L., “Software for Statistical Analysis of Sample Survey Data” (http://www.fas.harvard.edu/~stats/survey-soft/blc_oeb.html), P. Armitage and T. Colton (eds.), *Encyclopedia of Biostatistics*, John Wiley, 1998에 수록 예정.
16. Cassel, C. M., C. E. Saerndal, and J. H. Wretman, *Foundation of Inference in Survey Sampling*, John Wiley & Sons, Inc., 1977.
17. Cochran, W. G., *Sampling Techniques*, John Wiley & Sons, Inc., 3rd ed., 1977.
18. Joerup, H. and B. Rennermalm, “Regression Analysis in Samples from Finite Population”, *Scandinavian Journal of Statistics*, 3, 1976, pp. 33~36.
19. Kish, L. and M. R. Frankel, “Inference from complex samples”, *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol. 36, No. 1, 1974, pp. 1~37.
20. Lepkowski, J. and J. Bowles, “Sampling Error Software for Personal Computers”, *The Survey Statistician*, No. 35, December 1996, pp. 10~17 (<http://www.fas.harvard.edu/~stats/survey-soft/iass.html>).
21. Nathan, G., “Inference based on Data from Complex Sample Designs”, in P. R. Krishnaiah and C. R. Rao (eds.), *Handbook of Statistics*, Vol. 6, Elsevier Science Publishers B. V. (1988), pp. 247~266.
22. U. S. Bureau of the Census, Current Population Reports, pp. 60~197, *Money Income in the United States: 1996 (with Separate Data on Valuation of Noncash Benefits)*, U. S. Government Printing Office, Washington, DC, 1997.
23. Wolter, K. M., *Introduction to Variance Estimation*, Springer-Verlag, 1985.

〈부록〉 가구특성 변수의 지역별 평균치

	서울	부산	대구	인천	광주	대전	경기	강원
취업자수	1,465	1,459	1,391	1,418	1,335	1,349	1,369	1,283
가구주 연령	41.5	44.0	42.6	41.2	42.8	42.2	39.9	47.0
연령제공	1,862	2,084	1,969	1,835	2,016	1,943	1,729	2,411
배우자 연령	31.0	32.3	30.8	31.2	30.0	32.0	30.3	33.0
연령제공	1,310	1,427	1,346	1,290	1,307	1,397	1,220	1,528
여성 가구주	0.193	0.228	0.218	0.145	0.251	0.214	0.168	0.235
가구주 교육	고졸							
초등교 미만	0.110	0.182	0.157	0.137	0.167	0.168	0.133	0.311
중졸	0.139	0.174	0.189	0.152	0.143	0.157	0.158	0.175
전문대졸	0.085	0.067	0.082	0.081	0.077	0.091	0.079	0.046
대학졸	0.207	0.122	0.128	0.125	0.188	0.174	0.146	0.095
대학원졸	0.037	0.015	0.013	0.006	0.028	0.046	0.015	0.022
가구주 지위	노무자							
봉급자	0.293	0.211	0.193	0.222	0.265	0.256	0.263	0.216
자영자	0.260	0.264	0.269	0.251	0.262	0.251	0.219	0.222
경영자	0.034	0.015	0.028	0.012	0.015	0.015	0.015	0.012
자유업자	0.005	0.011	0.005	0.001	0.003	0.009	0.006	0.009
무직 및 기타	0.092	0.102	0.117	0.063	0.150	0.108	0.090	0.181
금융자산	20,533	15,957	15,015	16,220	17,543	15,979	15,420	15,422
주택소유가변수	0.507	0.539	0.516	0.642	0.523	0.576	0.482	0.682
	충북	충남	전북	전남	경북	경남	제주	전국
취업자수	1,315	1,311	1,361	1,160	1,171	1,343	1,314	1,386
가구주 연령	44.8	45.4	45.0	47.7	44.8	42.2	44.3	42.4
연령제공	2,208	2,255	2,203	2,516	2,243	1,940	2,217	1,961
배우자 연령	31.6	30.5	32.7	28.9	28.2	29.5	26.8	30.8
연령제공	1,405	1,375	1,477	1,327	1,241	1,224	1,133	1,313
여성 가구주	0.238	0.250	0.249	0.285	0.251	0.196	0.304	0.205
가구주교육	고졸							
초등교 미만	0.295	0.319	0.268	0.358	0.297	0.171	0.257	0.170
중졸	0.140	0.168	0.161	0.158	0.158	0.168	0.128	0.156
전문대졸	0.069	0.062	0.064	0.067	0.061	0.078	0.086	0.077
대학졸	0.109	0.103	0.140	0.073	0.073	0.124	0.126	0.151
대학원졸	0.024	0.017	0.019	0.009	0.003	0.015	0.016	0.022
가구주지위	노무자							
봉급자	0.217	0.202	0.236	0.175	0.165	0.232	0.234	0.246
자영자	0.199	0.288	0.247	0.227	0.209	0.177	0.247	0.241
경영자	0.013	0.006	0.016	0.004	0.003	0.017	0.004	0.020
자유업자	0.007	0.014	0.015	0.011	0.003	0.006	0.004	0.007
무직 및 기타	0.144	0.151	0.152	0.239	0.200	0.123	0.182	0.114
금융자산	17,099	16,696	15,377	12,520	13,692	15,997	20,303	17,115
주택소유가변수	0.634	0.649	0.638	0.628	0.596	0.563	0.429	0.543