

장기 인력수급 추이에 따른 소득세원의 변화

2003. 12

김 종 면 · 성 명 재

 한국조세연구원

* 상기 연구는 저자들의 개인적인 견해를 피력한 것이며 한국조세연구원의 공식적인 견해와 다를 수 있음을 밝혀두고자 함.

서 언

평균수명의 연장과 그에 따른 급속한 노령화의 진전, 출산율 저하 등으로 인해 우리나라는 인구·사회학적 구조가 매우 빠르게 변화하고 있다. 이는 중·장기적으로 경제활동참가연령인구의 구조적 변화를 초래함은 물론이고 노동공급 및 경제주체의 경제활동 양상 자체를 크게 변화시킬 것으로 예상된다. 실제로 중·장기적으로 향후 우리나라에서 일어날 것으로 예상되는 중요한 변화들은 인구구조 고령화와 깊이 관련되어 있다고 해도 과언이 아니다. 현재 우리나라에서는 고령화가 세계에서 유례를 찾아보기 힘들 정도로 급속히 진행되고 있는 만큼 이 방면의 연구활동도 매우 활발하게 진행되고 있다. 특히 최근들어 연금, 건강보험 등 사회보장 지출의 문제가 크게 부각되면서 경제학이나 경제정책에서도 고령화에 대한 관심이 고조되고 있다. 더 넓게는 사회보장 분야에 국한되지 않고 다양한 방면의 경제분석에서도 고령화와 관련된 연구가 이루어지고 있다.

그러나 고령화란, 달리 이야기해서, 인구의 연령분포가 변화하는 것을 의미함에도 불구하고, 현재까지는 연령분포의 변화를 경제변수들과 유기적으로 연계하여 구체적으로 분포의 미시적 변화효과를 분석하는 데에는 한계가 있었다. 이러한 한계는 물론 자료의 제약에도 이유가 있지만 보다 근본적으로는 제반 연구의 시발점인 소득의 연령별 분포를 예측할 수 있는 이론적 분석의 틀이 없었기 때문이라고 생각된다. 따라서 소득의 분포를 명시적으로 감안하여야 하는 연구주제에서도 평균소득 등 총량에 기반을 둔 분석 이상을 제

시하지 못하여 연구의 진척에 한계를 노정하여 왔다고 판단된다.

본 보고서는 이러한 한계를 극복하려는 시도로서 소득분포의 특성을 새로이 규명하고 이를 이용하여 소득분포를 전망할 수 있는 새로운 분석틀을 제시하고 있다. 따라서 본 보고서는 소득분포가 개입되는 대부분의 중·장기 전망에서 활용할 수 있는 결과를 제시하는 기초연구라고 간주할 수 있으며, 앞으로 소득분포나 고령화 관련 연구에서 한층 정치한 분석을 수행할 수 있는 기반을 마련하였다는 점에서 매우 의의가 크다고 하겠다.

본 연구는 상당히 방대한 주제들에 대해 세심하게 점검하였음에도 불구하고 방금 언급했듯이 기초연구로서 타 연구에서 쉽게 활용할 수 있도록 비교적 간략히 정리된 결과들을 제시하고 있다는 점이 돋보인다. 또한 가구소득의 분포가 자연대수 정규분포를 따른다는 사실을 우리나라 자료를 이용해서는 최초로 밝힌 점은 그 자체로서 학문적 의의가 크다고 생각되며 향후 다른 연구에서도 활용될 수 있는 가능성이 높다고 생각된다. 끝으로 소득분포의 특성을 자연대수 정규분포라고 규명하여 소득분포의 전망에 활용하는 접근법이나 지니계수를 소득의 평균과 분산으로부터 바로 추정할 수 있는 지니표면을 만들어 제시하는 부분에서 보인 창의성 역시 높이 평가하고 싶다.

본 보고서는 본 연구원의 김종면 박사와 성명재 박사가 공동으로 집필하였으며, 이정미 주임연구원, 이은경 주임연구원, 홍유남 연구조원, 윤혜순 연구조원이 자료수집과 원고정리에 수고를 아끼지 않았다. 저자들은 집필과정에서 지속적인 관심과 조언을 준 경희대 전택승 교수와 본 연구원의 김형준 박사에게 감사를 드린다. 또한 심사평과 세미나 논평을 통하여 보고서의 완성도 제고에 크게 도움을 준 본 연구원의 진병목 박사, 한국개발연구원의 최경수 박사, 성신여대 박기성 교수, 한국경제분석패널 전문가들의 기여에 감사하

며, 익명의 두 심사자에게도 사의를 밝힌다.

끝으로 본 보고서의 내용은 저자들의 개인적인 의견이며 본 연구원의 공식적인 견해가 아님을 밝혀둔다.

2003년 12월

韓國租稅研究院

院長 宋 大 熙

<요약 및 정책시사점>

1. 요약

1) 서론

우리나라는 선진국에서조차 찾아보기 어려울 정도로 인구구조가 매우 빠른 속도로 변화하고 있다. 범세계적으로 세계화·개방화·정보화 추세가 확산되면서 우리나라의 경제·사회구조도 크게 영향을 받고 있다. 더욱이 1997~1998년의 경제위기를 거치면서 기업 및 금융부문의 구조조정과 고용구조 등이 대폭 개편되었으며, 그 결과 소득분포 구조 또한 급속히 변화하였다. 이러한 추세적 변화는 세계 경제환경의 변화와 고령화의 진전 추세에 따라 앞으로도 당분간 지속될 것으로 전망된다.

우리 경제는 장차 고용구조 및 실업률 변동, 인구구조의 변화에 따른 노동공급과 수요의 불일치 등의 문제에 직면하게 될 것으로 예상된다. 더욱이 경제위기 이후의 사회안전망 확충과 복지사회로의 진입 등 본질적으로 우리의 재정을 팽창시킬 수 있는 요인이 점증하고 있다. 세입 측면에서는 인구의 노령화 등에 따라 장기적 인력수급 구조의 변화가 불가피해지고 있다. 각종 세원분포도 인력수급 구조의 변화에 크게 영향을 받을 것으로 예상된다. 따라서 중·장기적으로 변화된 제반 경제·사회적 환경 변화에 대응하여 세입·세출에 대한 중·장기적 예측이 더욱 중요성을 가지게 되었다.

그동안 우리나라에서는 각종의 재정수입·지출에 대한 연구가 많이 진행되었다. 이 중에서도 재정수입 부문과 관련해서는 소득세뿐만 아니라 소득에 기반하여 징수되는 각종 재정수입 관련 연구에서 소득총량을 기준으로 소요재정규모에 대한 추계가 이루어졌다. 이와 함께 재정지출 규모를 추계하는 경우에도 평균소득 등 총량변수를 사용하여 왔다. 그런데 재정수입 또는 지출은 기본적으로 총량지표에 의해 규모가 결정되기도 하지만, 총량 규모가 동일하다고 하더라도 각 구성원들의 상대소득분포에 의해 수입·지출 규모가 달라짐에 유의할 필요가 있다.

그런데 지금까지 우리나라에서 이루어진 관련 연구에서는 미시적 소득분포의 변화효과를 반영하지 못하고 단순히 집계변수(aggregate variables)를 통한 총량 위주의 분석에 그쳤다. 따라서 미시적 분포의 변화에 의한 효과 추정문제는 비껴나고 있다. 따라서 기존 연구에서는 소득분포의 변화에 의한 효과추정을 간과함으로써 연구목적과 수단 간에 괴리가 있었다고 할 수 있다. 본 연구는 이러한 소득분포 전망의 문제에 대해 유용한 해결책으로서 분포 자체를 전망할 수 있는 새로운 분석틀을 제시하고자 한다. 아울러 이러한 틀을 이용하여 분포 전망에 필요한 모수와 분포 특성 자체를 추정함으로써 중·장기 재정 관련 분야에 대한 기초연구로서 유용한 분석의 틀과 소득분포에 대한 추정결과 등을 제시하고자 한다.

2) 분석자료 및 추정기법

우리나라 소득분포의 전형적 특성(stylized fact)을 찾아내기 위하여 1982~2002년 사이의 도시가계자료를 이용하였다. 분석 기법으로는 가구소득 변수의 자연대수를 취한 자연대수 가구소

득을 연도별·가구주 연령별로 분류하여 분포함수를 비모수적인 회귀분석방법(nonparametric regression)으로 추정하였다. 이를 토대로 자연대수 소득의 정규분포 여부를 검정하였다. 검정결과 연도별·연령별 자연대수 소득분포가 정규분포를 따른다는 점을 실증적으로 밝혀내었으며, 그러한 관계가 시계열·연령별로 상당히 안정적인 관계임을 확인하였다. 소득분포의 안정적 변화패턴에 착안하여 중·장기적인 소득분포 예측방법을 제안하는 방법을 통해 소득세원의 변화추이 분석의 기초를 마련하였다.

3) 소득분포함수의 추정

소득분포의 구조 파악을 위해 본 연구에서는 다양한 검정방법을 동원하였다. (1) 자료상의 평균 및 분산을 통계학적 분포상의 평균 및 분포와 비교, (2) 경험분포와 이론적 확률분포와 비교, (3) 통계적 분포 검증을 통한 정규분포의 검정이라는 세 가지 방법을 통해 우리나라 도시가구의 소득이 연도별 또는 연도·연령별 소득분포에 있어 자연대수 정규분포와 유사한 수준을 넘어 실제로 자연대수 정규분포로 간주해도 무방하다는 사실을 실증적으로 확인하였다. 이를 통해 우리나라의 연령별 가구소득 분포는 평균과 분산이라는 2개의 모수로 분포 전체를 규정할 수 있다는 사실을 발견하였다. 이는 각종의 정책모의실험이나 경제모형 분석시에 모수적으로 주어진 소득분포가 자연대수 정규분포를 따른다는 점을 자연스럽게 전제할 수 있게 됨으로써 연구결과의 현실적 유용성이 크게 증대되었다고 할 수 있다.

우리는 흔히 1998년의 경제위기 이후에 소득분배구조가 급격히 악화되기 시작한 것으로 생각하는 경향이 있다. 즉, 1997년 연말에 외환위기가 도래하였고 1998년에 본격화되어 경제위기

상황이 가장 극심하였다고 본다. 그러나 본 연구에서 도시가구의 소득분포가 자연대수 정규분포를 따른다는 점을 검증하는 과정에서 나타난 결과에 의하면, 외환위기가 도래하기 이전에 이미 우리 경제의 소득분포 구조가 상당히 심각한 붕괴현상을 보이고 있었음을 확인할 수 있었다. 1997년의 자연대수 소득분포는 외환위기가 본격적으로 진행된 1998년과 같이 매우 심한 좌편향성을 보이고 있다. 이는 외환위기가 발발하여 경제붕괴가 시작된 1997년 말 이전에 이미 우리 경제는 보통의 불경기보다는 훨씬 심각한 구조적 문제의 표출이 상당히 진전되어 있었다는 사실을 시사하며, 보다 세밀한 추가적인 연구를 요구하는 결과라고 생각된다.

4) 연령별 상대소득분포의 추정

연령대별로 소득분포 구조를 검정한 결과 역시 각 연령별대 소득분포도 대체로 자연대수 정규분포를 따른다는 결론에 도달하였다. 또한 통합시계열 분석기법을 이용하여 연령별 자연대수 소득의 평균을 분석한 결과, 횡단면에서의 연령별 상대소득분포가 실질적으로 불변임을 보여주었다. 이를 통해 향후 연령별 소득분포 변화를 어느 정도 신뢰성을 갖고 예측할 수 있는 기반을 마련하였다.

5) 세대별 생애 연령-소득 곡선의 도출

횡단면 연령별 상대소득분포가 실질적으로 불변이라는 특성을 응용하여 세대별 생애 연령-소득곡선을 도출하였다. 1982~2002년의 21년간이라는 비교적 짧은 관측기간이라는 자료의 제약에

도 불구하고 40여년에 걸친 세대 생애 연령-소득 곡선을 추정할 수 있음을 보였다.

6) 지니표면을 이용한 자연대수 소득의 분산 추정

소득의 형평성을 측정하는 지니계수와 소득분포와의 관계를 구체적으로 규명하고 계량화하였다. 그 결과 지니계수는 지역적으로(locally) 자연대수 소득의 표준편차의 일차식으로 근사(approximation)시킬 수 있음을 보였다. 이러한 분석결과를 바탕으로 소득분포의 분산을 간접기법으로 추정하는 방법과 실제의 예를 제시하였다.

7) 소득세원 분포의 전망과 시사점

본 연구에서 찾아낸 소득분포의 특성을 사용하여 향후 소득분포의 변화를 추정하였다. 또한 이에 근거하여 향후 소득세의 추이를 예측해보는 사례를 보였다. 그 결과 향후 2032년까지 우리나라의 소득세 실효세부담률은 상승하는 추세를 보일 것으로 예측되었다. 다만 이 결과는 실효소득세부담률과 지니계수간의 상관관계만을 고려한 것일 뿐이다. 따라서 소득세제의 개편 여하에 따라서는 얼마든지 다른 결론에 도달할 수 있음에 유의할 필요가 있다.

2. 연구의 의의와 향후 보완과제

본 연구는 새로운 사실을 발굴하고 이를 이용한 새로운 방법론을 제시했다는 점에서 성격상 기초연구에 해당된다. 따라서 정

책에 직접적으로 적용할 수 있는 연구결과나 정책시사점을 제시하고 있지는 않다. 그러나 소득분포와 연관된 정책분야의 거의 모든 연구가 현실 연구에서 부딪혀왔던 가장 핵심적인 장애물 몇 가지를 해결할 수 있는 새로운 사실과 방법론을 발견·제시하였다는 점에서 본 연구의 의의와 학문적 기여도를 찾을 수 있다. 따라서 본 연구는 향후 소득분포나 소득분배와 연관된 대부분의 정책연구 및 학술연구에 상당히 큰 영향을 줄 것으로 생각된다. 즉, 본 연구의 결과는 기존 조세·재정 분석 중 상당 부분의 결과를 재검토할 필요성을 제기하고 있으며, 또한 향후 조세·재정 분석에서 본 연구에 기초하여 이전에는 분석이 불가능했던 새로운 시사점들이 도출될 수 있을 것으로 기대할 수 있다.

본 연구에서는 향후 소득분포를 전망하는 방법을 제시하고 있다. 따라서 이에 기초하여 세제개편이나 과세구간 재설정 등에 따른 장기효과를 분석하면서 세밀한 정책모의실험이 가능해진다. 또한 소득분포를 계량화하여 특성지음으로써, 공적부조 등 사회보장정책 기획에 있어서 수혜대상이나 금액 등 수혜기준 설정이 이전보다 훨씬 용이해지며, 이러한 정책에 따른 재정부담의 장기전망 또한 가능하게 해준다는 점에서 연구의 기여도·활용도 등을 논할 만하다.

이러한 연구가 더욱 빛을 발하기 위해서는 몇 가지 보완연구가 필요하다. 본 연구의 결과 소득분포는 자연대수 정규분포를 따르며 그러한 관계는 중·장기적으로 상당히 안정적인 모습을 보인다는 점이 밝혀졌다. 그러나 장래의 분포 예측에 있어 필요한 모수 중 분산의 경우에는 아직까지 결정요인에 대해 보다 정직한 분석이 요청된다. 또한 본 연구의 결과 중 소득분포의 중·장기적 안정성 여부에 대해서는 도시가계조사자료 이외의 다른 자료에 대해서도 동일한 결과를 얻을 수 있는지를 추가적으로

검정할 필요가 있다.

그 밖에 본 연구에서는 궁극적으로 중·장기적 소득세원의 분포 추정에 초점을 맞추고 있는데, 소득분포에 대해서는 상당한 정도의 정보를 제공해주고 있지만 과표의 총규모 및 상대분포와 세부담 분포에 대한 장기추세 분석에 있어서는 아직까지 만족할 만한 연구결과를 얻지 못하고 있다. 이는 중·장기적으로 소득세가 어떤 체계를 취하고 있을 것인지에 대해 사전적으로 아무런 정보가 주어지지 않기 때문이다. 뿐만 아니라 이를 예측하는 것도 현실적으로 불가능하기 때문에 과연 그러한 연구가 가능할 것인지에 대해 의문을 표시하지 않을 수 없음에도 유의할 필요가 있다.

그러나 이러한 연구에 대한 대안으로서, 가상적인 소득세제를 대상으로 어떤 세제가 어떤 세부담 분포를 초래할 것이며 어떠한 세부담 분포를 가져다주는 세제가 바람직한 것인지에 대해 평가해줄 연구방법을 제안할 수 있다. 이를 위해서는 기본적으로 주어진 정보, 즉 1982~2002년의 소득분포에 대한 정보를 기초로 향후의 소득분포와 각종 공제 관련 항목 등에 대한 예측(projection)을 통해 소득세제 정책시뮬레이션 연구를 수행할 수 있기를 기대한다. 물론 이러한 연구는 미시적 분포 특성을 추정 또는 검정하는 것보다 더 어려운 작업을 요구하는 만큼 미시분포의 변화추이에 발맞춘 개별관측치별 소득·소비분포에 대한 동시적 예측이 필요함에 유의하기 바란다.

목 차

I. 서론	25
1. 연구의 목적	25
2. 연령-소득 결합분포함수의 필요성	28
3. 분석방법	33
가. 개요	33
나. 소득분포함수의 추정	34
다. 정규분포의 검정	35
라. 소득분포의 특성에 따른 장기소득분포 구조의 추정	35
마. 장기소득분포의 변화가 장기소득세원 분포에 미치는 영향	37
II. 분석자료 및 추정기법	39
1. 분석자료	39
가. 분석자료의 종류	39
나. 도시가계조사자료의 특성	41
2. 커널 추정기법의 소개	47
III. 소득분포함수의 추정	50
1. 연도별 소득분포	50
가. 연도별 소득분포확률함수의 특성	50
나. 연도별 소득분포확률밀도함수의 추정	52
2. 연도별·연령별 자연대수 소득분포	70
3. 정규분포 여부의 검정	85

가. 검정통계량	85
나. 검정결과 I: 연도별 자료 분포	88
다. 검정결과 II: 연도별·연령별 자료 분포	89
4. 소결	94
가. 가구소득의 자연대수 정규분포 규명의 의의	94
나. 행렬형태의 표기	96
IV. 연령별 상대소득분포의 추정	98
1. 연령-소득 곡선의 도출	98
2. 회귀분석에 의한 세대 생애 연령-소득 곡선의 도출	106
V. 세대별 생애 연령-소득 곡선의 도출	119
1. 연도별 횡단면 연령-소득 곡선의 전망 도출	119
2. 세대별 생애 연령-소득 곡선의 도출	124
3. 개인소득과 대비한 가구소득의 특성	127
VI. 지니표면을 이용한 자연대수 소득의 분산 추정	131
1. 지니표면의 도출	131
2. 자연대수 소득분산의 추정	141
가. 분석방법	141
나. 지니계수 변화추이의 국제비교	142
다. 지니계수 변화율에 대한 예측	151
라. 자연대수 소득분산에 대한 예측결과	154
마. 소득 평균에 대한 가정	154
VII. 소득세원 분포의 전망 및 시사점	156
1. 소득세 실효세율과 소득분포의 상관관계	156
가. 소득세 실효세율과 소득 표준편차의 상관관계	156

나. 소득세 실효세율과 지니계수의 상관관계	161
2. 소득세 실효세부담률의 예측	162
VIII. 결론	165
참고문헌	168
부록 I. 근로자가구의 근로소득세 부담 분포	172
부록 II. 커널 추정기법의 소개	175
1. 비모수적 확률분포함수 추정량	175
2. 커널추정량의 점근적 특성 보장을 위한 가정	176
3. 커널추정량의 일치성과 점근적 정규분포	179
4. 커널추정량의 일치성과 점근적 정규분포	182
5. 커널추정량의 점근분포에 대한 일고	183
6. 층화무작위표본하에서의 커널추정량	184
부록 III. 연도-연령별 자연대수 가구소득 분포도	185

표 목 차

<표 I- 1> 특정 연도의 가구특성별 소득분포 표의 예시	29
<표 I- 2> 가구주 연령별 소득분포의 연도별 추이 (현황 및 전망)의 예시	31
<표 II- 1> 도시가계조사자료에 나타난 가구유형별 가구 수	44
<표 II- 2> 도시가계조사자료에 나타난 가구유형별 가구점유비	45
<표 II- 3> 가구유형별 평균소득	46
<표 II- 4> 가구유형별 가구주의 평균연령	47
<표 III- 1> 두 방법으로 추정한 도시가계가구의 자연대수소득 (x_t)의 평균과 분산	55
<표 III- 2> 1인당 경상GDP의 비율로 정의된 디플레이터의 도출	57
<표 III- 3> 1인당 경상GDP를 디플레이터로 사용한 x_t 의 평균과 분산의 비교	58
<표 III- 4> 연도별 정규분포 검정결과	89
<표 III- 5> 연도별·연령별 정규분포 검정결과 (유의수준=10%)	91
<표 III- 6> 연도별·연령별 정규분포 검정결과 (유의수준=5%)	92
<표 III- 7> 연도별·연령별 정규분포 검정결과 (유의수준=1%)	93
<표 IV- 1> 도시가계(1982~2002년)의 연령별 평균 자연대수 실질가구 소득의 분포	104

<표 IV- 2> 일방고정모형을 사용한 연령별 자연대수 소득의 회귀분석	114
<표 VI- 1> m의 변화에 따른 지니계수 추정식의 변화	140
<표 VI- 2> m>18 영역에서의 지니표면의 추정식	140
<표 VI- 3> 지니표면과 기존방법을 사용한 지니계수의 비교 ..	141
<표 VI- 4> 미국의 소득불평등도 계측결과	144
<표 VI- 5> 영국의 지니계수 추정결과	145
<표 VI- 6> 호주의 지니계수 추정결과	146
<표 VI- 7> 일본의 지니계수 추정결과	146
<표 VI- 8> 뉴질랜드의 지니계수(근로소득자)	147
<표 VI- 9> 캐나다의 지니계수	147
<표 VI-10> 싱가포르의 가계소득분배	147
<표 VI-11> 우리나라의 총소득 기준 지니계수 추정치	149
<표 VI-12> 지니계수 예측 시나리오에 대응한 자연대수소득 표준편차(분산) 추정결과	153
<표 VI-13> 자연대수 소득평균에 대한 가정	155
<표 VII- 1> 실효소득세율과 자연대수소득 표준편차간의 회귀분석 결과	160
<표 VII- 2> 실효소득세율과 지니계수간의 회귀분석 결과	161
<표 VII- 3> 소득세 실효세부담률 추정결과 I	163
<표 VII- 4> 소득세 실효세부담률 추정결과 II	164
<부표 I -1> 근로자가구의 근로소득세 부담액 분포 (무직가구 제외)	172
<부표 I -2> 근로자가구의 근로소득세 실효세부담률 분포 (무직가구 제외)	173
<부표 I -3> 근로자가구의 근로소득세 실효세부담 점유비 분포 (무직가구 제외)	174

그 립 목 차

[그림 III- 1] 막대그래프형 경험확률밀도함수와 이론상의 정규분포 확률밀도함수의 비교	60
[그림 III- 2] 연도별 추정경험분포와 정규분포의 비교 (1982~1984년)	63
[그림 III- 3] 1995년의 연령별 자연대수소득의 경험분포와 정규분포 확률밀도함수의 비교(25~36세)	71
[그림 III- 4] 중년기 소득의 좌편향 분포의 예	76
[그림 III- 5] 외환위기시(1997년) 연령별 자연대수 소득분포의 좌편향성(25~36세)	81
[그림 IV- 1] 3개 연도의 연령별 평균소득 수준의 비교	99
[그림 IV- 2] 세대 연령-소득 곡선의 도출 (1982년 당시 25세 세대의 예)	101
[그림 IV- 3] 도시가계의 경상소득 기준 세대 생애 연령- 소득 곡선	101
[그림 IV- 4] 도시가계의 1인당 실질소득 기준 세대 생애 연령- 소득 곡선	103
[그림 IV- 5] 도시가계의 연령별 평균 자연대수 가구소득 (경상성장률 디플레이터 적용)	105
[그림 IV- 6] 피용자보수와 영업잉여의 합	109
[그림 IV- 7] 총인구 추이(1960~2002년)	110
[그림 IV- 8] 각세별 연령분포(1982, 1992, 2002년)	111
[그림 IV- 9] 회귀분석으로 추정된 횡단면 연령별 상대소득 분포	116

[그림 IV-10] 횡단면 연령층간 상대소득 분포에 대한 연령별 총인구의 영향(1982년)	117
[그림 IV-11] 횡단면 연령층간 상대소득 분포에 대한 연령별 총인구의 영향(2002년)	118
[그림 V- 1] 총인구 추이(1981~2050년)	121
[그림 V- 2] 총인구 성장률(2002~2050년)	121
[그림 V- 3] 각세별 연령분포(1982, 2002, 2022, 2050년)	122
[그림 V- 4] 횡단면 연령층간 상대소득 분포에 대한 연령별 총인구의 영향(2002년)	123
[그림 V- 5] 횡단면 연령층간 상대소득 분포에 대한 연령별 총인구의 영향(2020년)	123
[그림 V- 6] 횡단면 연령층간 상대소득 분포에 대한 연령별 총인구의 영향(2050년)	124
[그림 V- 7] 세대별 생애 연령-자연대수소득 곡선 (명목성장-5% 가정)	126
[그림 V- 8] 세대별 생애 연령-자연대수소득 곡선 (명목성장 7% 가정)	127
[그림 VI- 1] 로렌츠 곡선 (Lorenz curve)의 예시	133
[그림 VI- 2] 지니표면: 자연대수가구소득의 평균 m 과 분산 s 에 따른 지니계수의 도출	136
[그림 VI- 3] 우리나라 지니계수의 변화추이	150
[그림 VI- 4] 지니계수 변화추이의 국제비교 I	150
[그림 VI- 5] 지니계수 변화추이의 국제비교 II	151
[부도 III- 1] 가구주 연령별 도시가구 소득분포함수의 추정 (1982년, 25~36세)	186
[부도 III- 2] 가구주 연령별 도시가구 소득분포함수의 추정 (1983년, 25~36세)	190
[부도 III- 3] 가구주 연령별 도시가구 소득분포함수의 추정 (1984년, 25~36세)	194

[부도 Ⅲ- 4] 가구주 연령별 도시가구 소득분포함수의 추정 (1985년, 25~36세)	198
[부도 Ⅲ- 5] 가구주 연령별 도시가구 소득분포함수의 추정 (1986년, 25~36세)	202
[부도 Ⅲ- 6] 가구주 연령별 도시가구 소득분포함수의 추정 (1987년, 25~36세)	206
[부도 Ⅲ- 7] 가구주 연령별 도시가구 소득분포함수의 추정 (1988년, 25~36세)	210
[부도 Ⅲ- 8] 가구주 연령별 도시가구 소득분포함수의 추정 (1989년, 25~36세)	214
[부도 Ⅲ- 9] 가구주 연령별 도시가구 소득분포함수의 추정 (1990년, 25~36세)	218
[부도 Ⅲ-10] 가구주 연령별 도시가구 소득분포함수의 추정 (1991년, 25~36세)	222
[부도 Ⅲ-11] 가구주 연령별 도시가구 소득분포함수의 추정 (1992년, 25~36세)	226
[부도 Ⅲ-12] 가구주 연령별 도시가구 소득분포함수의 추정 (1993년, 25~36세)	230
[부도 Ⅲ-13] 가구주 연령별 도시가구 소득분포함수의 추정 (1994년, 25~36세)	234
[부도 Ⅲ-14] 가구주 연령별 도시가구 소득분포함수의 추정 (1995년, 25~36세)	238
[부도 Ⅲ-15] 가구주 연령별 도시가구 소득분포함수의 추정 (1996년, 25~36세)	242
[부도 Ⅲ-16] 가구주 연령별 도시가구 소득분포함수의 추정 (1997년, 25~36세)	246
[부도 Ⅲ-17] 가구주 연령별 도시가구 소득분포함수의 추정 (1998년, 25~36세)	250

[부도 Ⅲ-18] 가구주 연령별 도시가구 소득분포함수의 추정 (1999년, 25~36세)	254
[부도 Ⅲ-19] 가구주 연령별 도시가구 소득분포함수의 추정 (2000년, 25~36세)	258
[부도 Ⅲ-20] 가구주 연령별 도시가구 소득분포함수의 추정 (2001년, 25~36세)	262
[부도 Ⅲ-21] 가구주 연령별 도시가구 소득분포함수의 추정 (2002년, 25~36세)	266

I. 서론

1. 연구의 목적

우리나라는 평균수명의 연장과 그에 따른 노령화의 급속한 진전, 출산율 저하 등으로 인해 인구·사회학적 구조가 매우 빠르게 변화하고 있다. 이는 중·장기적으로 경제활동참가연령인구의 구조적 변화를 초래함은 물론이고 노동공급 및 경제주체의 경제활동 양상 자체를 크게 변화시킨다. 따라서 다른 나라의 경험에서는 유례가 없을 정도로 급속하게 노령화가 진행되고 있는 우리나라 실정에 비추어 볼 때, 노령화와 연관된 경제연구가 활발히 진행되고 있는 것은 당연하다고 하겠다. 경제전반으로부터 공공정책으로 시야를 좁혀서 보아도 사회보장 지출 증가 등 노령화의 영향은 매우 크며 최근 에 점차 다양하고도 활발한 연구와 정책분석이 이루어지고 있다. 그러나 공공부문 지출의 이면에는 항상 수입·재원조달의 문제가 있다는 사실을 잊어서는 안 되며, 노령화와 관련되어 재정수입 측면에서의 연구는 상대적으로 미흡했다고 사료된다. 본 연구에서는 이러한 문제의식하에 세입, 그 중에서도 노령화와 밀접한 관련이 있는 소득세에 초점을 맞추고자 한다.

일반적으로 소득세의 실효부담은 크게 세법, 경제 전체의 총소득 수준, 소득분포의 세 가지 요인에 의해 결정된다. 이 가운데 외생적으로 주어지는 세법을 제외하면 총소득 수준과 소득분포는 소득세 실효세부담을 결정하는 데 가장 큰 요인이라고 할 수 있다. 그런 의미에서 경제 전체의 총소득과 소득분포를 소득세원이라고 지칭할

수 있다¹⁾. 이 중 소득세법·세제의 변화는 예측이라기보다는 시나리오 설정 등 가정을 한다는 성격이 강하므로 본 연구의 분석범위에서 제외하였다. 따라서 소득세원의 변화와 소득세제의 변화를 모두 감안하여야 도출할 수 있는 소득세수의 전망은 본 연구의 목적이 아님을 명시적으로 밝히는 바이며, 본 연구의 범위는 근본적으로 소득세원의 변화, 즉 소득분포의 변화를 분석하는 데 국한하였다²⁾.

성명재(2001)의 연구에 따르면 인구의 노령화와 이혼 증가와 같은 사회현상의 변화는 그 자체의 변화에서 그치지 않고 소득분배 구조에도 상당히 큰 영향을 미친다. 따라서 인구구조가 변화하면 세원분포에 영향을 미침으로써 단기적으로는 물론이고 중·장기적으로도 실효세부담을 변화시킨다.

다른 조건이 같을 경우, 만약 세법에서 각종 공제수준이나 세율구간 등을 모두 평균소득증가율과 동일하게 변화시킨다면 소득세의 실효세부담에는 아무런 영향을 미치지 않을 것이다. 그러나 현실에서는 시간이 경과함에 따라 소득계층별로 소득증가율이 상이할 뿐만 아니라 인구의 연령구조 등도 함께 변화한다. 그러므로 자연스럽게 소득의 상대분포도 변화하게 된다. 이는 곧 총소득과 각종 공제나 세율구간 등이 동일한 증가율로 변화하더라도 인구학적 특성

-
- 1) 다음 절에서도 언급하지만, 이는 사실 소득세뿐만 아니라 국민연금 및 건강보험 보험료 등 소득세와 유사하게 소득에 기반하여 징수되는 모든 공공부문 재정수입에 해당되는 이야기이다. 따라서 본 연구에서 소득세를 염두에 두고 전개되는 소득세원으로서의 소득분포에 관한 논의는 사실상 국민연금, 건강보험 등 보다 광범위한 사안 및 연구에 적용된다고 볼 수 있겠다.
 - 2) 제 VII장에서 소득세 실효부담률 전망을 제시함으로써 소득세수 추계가 제시되어 있다. 그러나 소득세제가 향후 어떻게 변할 것인가에 대해서 본격적인 분석이 수반되지 않았다는 사실에서도 알 수 있듯이, 제 VII장에서 제시된 소득세 실효부담률 전망은 본 보고서에서 보여준 바와 같이 소득세원의 변화를 전망하였을 때, 이를 응용하여 시도될 수 있는 분석, 즉 하나의 적용 사례로 제시하였다고 간주하는 것이 옳다고 본다.

이 변화함에 따라 소득분포가 달라지면 소득세 실효세부담도 함께 변화하게 됨을 시사한다.

인구의 노령화 등 인구학적 특성변화가 매우 빠른 속도로 진행되고 있는 현실을 직시할 때 그에 따른 세원분포의 변화를 추정하는 것은 매우 의미있는 일이다. 본 연구는 바로 이 점에 착안하여 인구 특성의 변화에 따른 총소득 및 소득분포, 즉 소득세원이 변화 가능성에 초점을 맞추어, 총소득 수준과 소득분포(세원 분포)의 변화를 추정함을 목적으로 한다. 그리고 실효세부담의 추정을 통해 시계열적 특성을 고려하여 소득분배 지수의 추정치를 기초로 총소득에 대비한 소득세 부담 비율, 즉 소득세 실효세부담률을 추정해 본다.

소득세원을 논함에 있어 우리는 흔히 근로자가구의 근로소득세 부담 문제만을 생각하기 쉽다. 우리 국민 중 상당수가 근로소득자이거나 가족으로 구성되어 있기 때문에 근로소득세는 일상생활로부터 떼려야 뗄 수 없는, 우리의 현실에 밀착된 세금이기 때문이다. 따라서 장기 소득세원의 추정 문제와 관련하여 근로소득세에 대응되는 소득세원에 대한 문제로 분석의 범위를 맞추고자 하는 견해가 있다.

그런데 본 연구의 제목에서 보듯이 중·장기적으로 우리나라의 인구구조가 노령화·저출산율 체제로 급속히 전환되고 있고 그러한 추세적 변화가 세원분포에 어떤 영향을 미치는지가 주된 관심의 대상이다. 그러한 관점에서 보면, 근로소득세의 세원분포 변화추이만을 추정하는 것은 인구구조 자체의 변화와 관련된 문제의 본질에서 상당히 벗어나 있음을 지적하지 않을 수 없다. 왜냐하면 기본적으로 인구구조학상 노령층은 대부분 무직가구나 소규모의 자영업을 영위하는 가구로 구성되어 있을 개연성이 높기 때문에 근로소득세에만 초점을 맞추게 되면 자연스럽게 노령화 등과 같은 인구구조의 변화에 의한 소득세원 분포의 변화효과를 제대로 추정하기 어렵게 되기 때문이다. 더욱이 이 경우 실업에 의한 무직가구까지 분석대상에서 배제되므로 실업률 변동 또는 경기변동에 따라 근로자가구,

보다 정확히는 취업근로자가구의 포괄범위가 가변화됨에 따라 분석 대상의 범위가 확정되지 않는다는 단점도 있음에 유의할 필요가 있다.

그러므로 이러한 관점에서 본 연구는 분석대상을 근로소득세에 국한시키지 않고 전체 소득세로 한다³⁾. 또한 근로소득, 또는 임금 소득에 관한 연구가 분석단위를 개인으로 설정하고 있음에 비해, 고령화와 연관된 제반 사회현상의 변화와 이의 정책적 시사점을 논함에 있어 개인보다는 가구 단위의 분석이 적절하다는 판단하에 분석의 기본단위를 개인이 아닌 가구에 두었음을 지적하고자 한다.

2. 연령-소득 결합분포함수의 필요성

본절에서는 소득의 연령간 분포를 전망과 관련하여 기존의 접근법이 갖는 문제점을 설명함으로써 본 연구에서 시도하는 소득분포의 특성 규명에 대한 논의의 단초를 찾고자 한다.

본 연구뿐만 아니라 우리 경제의 중장기 분석을 시도하는 대부분의 연구에는 소득의 연령별 분포에 대한 중장기 전망이 필요하다. 이와 같이 소득의 연령별 분포를 다루지 않을 수 없는 이유는 우리 경제가 중장기에 걸쳐 겪을 것으로 예상되는 변화 중 가장 중요한 요인인 인구 연령구조의 고령화 현상을 분석에 반영하기 위함이다. 이러한 맥락의 연구 주제에는 소득세의 중장기 분석, 국민연금이나 건강보험 등 소득세와 유사한 성격의 사회보험료 수입·재정수지에 대한 전망, 세대간 회계를 이용한 분석 등 조세와 사회보험 분야의 연구들이 포함되며, 그밖에도 고령화 문제와 관련된 주제라면 공히 소득의 연령별 분포 전망이라는 과제를 피해가기 어려울 것이다.

그런데 거시총량변수 위주의 비교적 단순한 분석방법을 택하지

3) 단, 근로소득세의 세부담 분포는 참고를 위해 부록에 제시하였다.

않는 한, 고령화 문제를 다루는 경제분야 연구는 대개의 경우 다음과 같은 접근법으로 이 과제를 다루게 된다. 즉 아래 <표 I-1>에서 보는 바와 같이 특정 연도의 인구 또는 가구에 대하여 연령과 기타 가구특성(예: 가구 구성원 수 등)으로 분류한 표를 각 연도별로 작성하게 된다. 이 표의 칸에는, 즉 각각의 연령층과 가구특성별로는, 평균소득수준과 그 연령·가구특성에 해당되는 인구를 기록하게 된다. 이때 향후 소득의 연령별 분포를 전망한다는 것은 이와 같은 표를 향후 전망이 필요한 연도에 대하여 작성하는 작업이 되겠다. 보다 구체적으로는, 표의 각 칸마다 평균소득수준과 인구가 어떻게 변할 것인지 예측하는 작업이 되겠다⁴⁾.

<표 I-1> 특정 연도의 가구특성별 소득분포 표의 예시

		가구특성 i (예: 가구 구성원 수)			
		i=1	i=2	... i=k ...	i=K
	25세	(N _{25,1} , P _{25,1})	(N _{25,2} , P _{25,2})	... (N _{25,k} , P _{25,k}) ...	(N _{25,K} , P _{25,K})
	26세	(N _{26,1} , P _{26,1})	(N _{26,2} , P _{26,2})	... (N _{26,k} , P _{26,k}) ...	(N _{26,K} , P _{26,K})
가 구 주 연 령	·	·	·	· ·	·
	·	·	·	· ·	·
	n세	(N _{n,1} , P _{n,1})	(N _{n,2} , P _{n,2})	... (N _{n,k} , P _{n,k}) ...	(N _{n,K} , P _{n,K})
	·	·	·	· ·	·
	·	·	·	· ·	·
	·	·	·	· ·	·
	65세	(N _{65,1} , P _{65,1})	(N _{65,2} , P _{65,2})	... (N _{65,k} , P _{65,k}) ...	(N _{65,K} , P _{65,K})

4) 고령화와 연관된 경제문제를 다루는 한 거의 예외 없이 이와 같은 분석틀을 사용하게 되는 이유는 기존의 이론모형 중 인구의 연령구조 변화와 경제이론을 유기적으로 통합한 모형이 없기 때문이다. 일반적으로 통용되는 정형화된 모형이 없는 마당에서, 현존하는 통계에 내재된 인구관련 정보, 특히 연령별 인구의 상대적 규모의 변화를 최대한 반영하는 방법은 이와 같이 가구주 연령 등 가구특성별로 계층을 세분화하여 각 계층별로 향후 추이에 대한 전망을 도출하는 접근법에 국한될 수밖에 없을 것이다.

이때 인구에 대한 예측은 분석모형에서 내생적으로(endogenously) 도출하지 못하고 통계청의 장래인구 추계자료를 대입하는 것이 일반적인 관례이다. 인구 등 통상적으로 사회적 변수(social variables)라고 칭하는 요인들과 소득 등의 경제변수(economic variables)가 시간의 경과에 따라 상호 결정요인으로 작용하면서 내생적으로 결정되는 모형이 사실 바람직하기는 하나, 흔히 이야기하는 자료의 제약울 굳이 거론하기 이전에 마땅히 응용할 수 있는 이론모형이 없다는 문제가 있기 때문이다. 따라서 인구 변동을 외생적으로 처리함을 크게 문제삼기는 어렵다고 볼 수 있다.

인구 예측과 마찬가지로 계층별 평균소득의 예측에서도 소득수준의 장래 변화는 사실상 외생적으로 결정되는 것으로 상정하는 것이 일반적이다. 소득의 전반적 수준의 변화나 경제성장률까지 분석모형에서 내생적으로 도출하는 경우, 연구의 범위가 너무 넓어지게 되기 때문이다. 그러므로 경제성장률 등 거시경제 전망은 외생변수로 상정하고, 이런 식으로 주어진 소득수준 변화의 영향이 서로 다른 연령 계층 간에 어떻게 작용하는지를 규명하는 것이 연구의 핵심이다. 이 같이 소득의 전반적 수준을 외생적으로 처리하는 것까지도 큰 무리는 없다고 사료되며, 본 연구에서 문제를 제기하고자 하는 것은 마지막 부분에서 소득의 연령간 분포의 변화를 전망하는 방법에 대해서이다.

더 자세한 설명을 위해 <표 I-2>를 보자. 서술의 편의를 위하여 <표 I-1>에서 두 가지로 나타냈던 가구특성을 가구주 연령 한 가지로 국한시키면, <표 I-2>에서의 각각의 열(column) 한 개가 <표 I-1>의 내용에 상응한다고 이해할 수 있다. 이때 과거의 자료를 이용하여 향후 소득의 연령별 분포를 도출하는 작업은, 현재까지 관측된 t 기 이전의 자료의 속성을 파악한 후, 이를 연장하여 $t+1$ 기 이후의 열을 작성함을 의미한다. 그런데 이때 크게 나누어 두 가지 방법 중에서 하나를 택하게 마련이다. 즉, 소득의 전반

적인 수준은 향후 예상되는 거시적인 경제성장률과 부합되도록 증가하는 것으로 전망하되, 각 관측 시점별로 연령별 소득의 상대적인 수준이 t 기 또는 t 기 이전 몇 년의 평균을 그대로 유지한다는 불변의 가정을 채택하는 방법이나, 또는 각 연령별로 <표 I-2> 상에서 횡으로 자료를 살펴보아 전반적인 소득상승 추세 이 외에도 증감 추세가 뚜렷하게 파악될 경우 이를 연장하여 $t+1$ 기 이후의 소득수준을 계산하는 방법이다. 이 두 가지 방법에 공통된 문제점 몇 가지를 살펴보면 다음과 같다.

<표 I-2> 가구주 연령별 소득분포의 연도별 추이(현황 및 전망)의 예시

		연도 t			
		t=82	t=83	... t=T ...	t=T
가 구 주 연 령	25세	(N _{25,1} , F _{25,1})	(N _{25,2} , F _{25,2})	... (N _{25,t} , F _{25,t}) ...	(N _{25,T} , F _{25,T})
	26세	(N _{26,1} , F _{26,1})	(N _{26,2} , F _{26,2})	... (N _{26,t} , F _{26,t}) ...	(N _{26,T} , F _{26,T})

	□세	(N _{□,1} , F _{□,1})	(N _{□,2} , F _{□,2})	... (N _{□,t} , F _{□,t}) ...	(N _{□,T} , F _{□,T})

	65세	(N _{65,1} , F _{65,1})	(N _{65,2} , F _{65,2})	... (N _{65,t} , F _{65,t}) ...	(N _{65,T} , F _{65,T})

첫째, 향후 소득의 연령별 분포를 불변으로 상정하든, 추세를 따라 증감한다고 상정하든, 어느 경우에도 상당히 강한 암묵적 가정이 들어 있음이 분명하다. 그러나 이러한 암묵적 가정이 정확히 무엇인지 알기 힘들 뿐더러, 그러한 암묵적 가정이 얼마나 타당한지 가늠하는 것은 더욱 어렵다는 문제가 있다. 둘째, 과거 추세를 연장하여 장래 전망을 도출하는 방법은 현재까지의 관측기간의 3분의 1

정도의 기간까지만 어느 정도 신뢰성이 있는 전망이 가능하며, 최대한으로 관측기간의 2분의 1에 해당하는 기간의 전망을 받아들이는 것이 관례이다. 현재 우리나라의 소득자료는 약 20년에 걸친 시계열 자료만 존재할 뿐이다. 이 정도의 자료로는 향후 10년 이상의 소득추이를 전망하는 것은 무리이다. 10년 정도의 분석으로도 인구고령화의 경제적 영향을 보기는 어렵다. 그러므로 통상 30~50년 정도의 기간에 걸친 분석이 필요하다. 따라서 단순히 과거 추세를 파악·연장하는 접근법으로 고령화에 대한 중장기 분석을 시도하는 데에는 한계가 있으며, 과거 추세 이 외에도 소득분포의 안정적 특성을 추가로 발굴하여 분석에 도입할 필요가 있다.

위의 두 가지 문제에 추가하여 또 한 가지 문제를 지적하자면, <표 I-1>이나 <표 I-2>와 같은 방법으로는, 비록 연령이나 기타 특성으로 세분화하기는 하였지만, 결국 분석의 산물이 평균소득이라는 하나의 계수이지 연령이나 기타 특성별로 소득분포 자체를 구한 것은 아니라는 점을 들 수 있다. 예를 들어 <표 I-1>이나 <표 I-2>와 같은 방법으로는 특정 시기에 특정 연령계층의 평균소득이 얼마인지, 또한 다른 연령계층에 비해 평균적으로 소득수준의 격차가 어떠한지는 논할 수 있으나, 특정계층 내에서 소득의 분포가 어떻게 될 것인지에 대한 정보가 없으므로 소득분배 같은 주제나 조세정책의 효과를 살펴보는 정책분석에서는 제약이 따를 수밖에 없다. 간단한 예를 들면 사회보장 정책분석에서는 특정 소득수준 이하의 가구가 얼마나 되는지 알 필요가 있으나, <표 I-1>이나 <표 I-2>에 있는 정보로는 답을 구할 수 없다. 이는 사실 소득의 연령별 분포와 관련된 연구뿐만 아니라, 분포보다는 분포의 평균 등 일부 특정 계수만을 사용하는 방법론을 따르는 연구의 공통된 한계라고 할 수 있다. 이 문제 역시 소득분포의 특성에 대한 보다 심층적인 분석을 요구한다고 할 수 있다.

3. 분석방법

가. 개 요

장기적으로 소득세원 분포를 예측하기 위해서는 개인들의 생애소득 흐름을 파악하여야 한다. 이를 위해서는 장기간을 대상으로 구축된 패널자료가 필수적이다. 그렇지만 전술한 바와 같이 국내에는 충분히 긴 기간을 대상으로 축적된 패널자료가 없는 실정이다. 다행히 통계청에서 수집하여 원시자료의 형태로 발표하는 도시가계조사자료의 경우에는 1982년도분부터 원시자료가 이용가능하다.

도시가계조사 원시자료는 지난 21년간이라는 비교적 긴 기간을 대상으로 축적되어 있다. 그렇지만 생애소득의 흐름을 빠짐없이 추적하기에는 충분하지 않다. 즉, 예를 들어 한 개인이 일생을 통해 본격적으로 소득을 획득하는 시기가 25세에서 65세에 이르는 41년 간이라고 하면 최소한 41년 이상의 기간을 대상으로 축적된 자료 필요하다. 그렇지만 현재 이용가능한 도시가계조사 원시자료는 1982~2002년의 21년간 자료뿐이다. 그러므로 이 자료를 이용하여 생애소득의 흐름을 추정하기 위해서는 각 연령대별로 소득의 연도별 흐름도를 추정하고 이를 연령대별로 중첩(overlapping)시키는 방법을 통해 간접적으로 생애소득의 흐름을 추정하여야 한다.

그런데 각 세대(generation)⁵⁾는 시계열적으로 서로 상이한 경제성장 경험(economic history)을 가지고 있다. 그러므로 예를 들어 현재 40세인 경제주체가 향후 5년간 경험하게 될 소득의 흐름이 현재 45세인 경제주체가 지난 5년간 경험하였던 소득의 흐름과 일치하지는 않는다. 뿐만 아니라 현재 45세인 경제주체가 지난 10~6년

5) 본 연구에서는 25~65세를 분석대상으로 하면서 한 세대는 동일 연령으로 구성된다고 가정한다. 그러므로 본 연구에서는 한 시점에서 모두 41개 세대가 존재한다. 분석의 편의상 25세 미만과 65세를 초과하는 자는 각각 25세 세대와 65세 세대로 분류하였다.

전에 경험하였던 소득의 흐름과 현재 40세인 경제주체가 지난 5년간 경험하였던 소득 흐름과도 다른 물론이다. 이는 각기 실질경제성장률과 노동 및 자본소득분배율이 서로 다를 뿐만 아니라 물가상승률도 차이가 나기 때문이다. 이러한 문제는 모두 자료의 시계열이 모든 세대를 통틀어 분석할 수 있을 만큼 오랜 기간을 대상으로 축적되어 있지 못하기 때문이다.

그러므로 차선책으로 본 연구에서는 일차적으로 각 연도별 자료로부터 평균적인 경제적 경험(economic history)의 차이에 의해 나타나는 소득흐름(income profile)의 차이요소를 제거한다. 이를 토대로 순수히 소득분포의 연령대별(또는 세대별) 변화패턴을 파악하고 이를 근거로 장래의 소득분포를 추정한다. 이것은 경제성장이 정제된 상황을 전제로 하는 만큼, 최종적으로는 목표연도와 기준연도 사이의 경상성장률 차이를 곱해줌으로써 최종적인 소득분포의 예측결과를 추정한다.

이러한 분석을 위한 중간과정에서는 소득분포 함수(확률밀도함수 또는 누적분포함수 등)의 추정, 정규분포 여부에 대한 가설검정, 경제성장이 정제된 상황하에서의 소득분포구조의 변화 추정, 장기적인 소득분포의 변화가 소득세원 분포에 미치는 영향에 대한 분석 등이 필요하다.

나. 소득분포함수의 추정

흔히 소득분포는 자연대수 정규분포(log-normal distribution)를 따른다고 한다. 즉, 소득에 자연대수를 취하면 소득분포가 정규분포를 따른다는 것이다. 우리나라에서 장기간을 대상으로 미시자료를 분석한 연구 중에서 자연대수를 취한 소득이 정규분포를 따른다는 것을 검정한 연구는 아직 없다. 만약 자연대수 소득분포가 정규분포, 즉 소득변수가 자연대수 정규분포를 따른다고 하면, 표본자료로

부터 평균과 분산의 두 가지 모수(parameter)만 추정함으로써 소득 분포에 대한 모든 정보를 얻을 수 있다.

그러므로 본 연구에서는 소득변수가 (자연대수) 정규분포를 따르는지를 검정하기 위해 일차적으로 비모수적인 방법을 사용하여 소득변수에 대한 분포함수(확률밀도함수; probability density function)를 추정한다.

다. 정규분포의 검정

비모수적 방법으로 확률밀도함수를 추정하면 자연대수 소득변수가 정규분포를 따르는지에 대한 통계적 검정이 필요하다. 가설검정을 위해서는 표본자료를 직접 이용하여 검정하는 방법과, 추정된 표본분포로부터 가상의 무작위표본을 추출하여 정규분포의 여부를 검정하는 두 가지 방법이 있다. 본 연구에서는 편의상 전자의 방법으로 정규분포의 여부를 검정한다.

라. 소득분포의 특성에 따른 장기소득분포 구조의 추정

다음 장에서의 검정결과를 미리 일부 소개한다면, 표본분포에 대한 정규분포 검정결과, 자연대수 소득변수의 표본분포는 대체로 정규분포를 따르는 것으로 검정되었다. 정규분포는 평균과 분산만으로 분포에 대한 모든 정보가 완전하게 규정되는 만큼, 자연대수 소득분포가 정규분포를 따른다는 사실은 장래에도 외부의 충격에 의해 소득분포(자연대수 소득분포)가 자연대수 정규분포(정규분포)에서 벗어나지 않는 한 평균소득과 분산의 두 가지 모수만을 추정함으로써 장기 소득분포 구조를 성공적으로 추정할 수 있음을 시사해준다.

자연대수 소득분포가 정규분포를 따른다고 할 때 표본으로부터 자연대수를 취한 표본자료로부터 평균과 분산을 추정하는 방법에는

크게 두 가지를 생각할 수 있다. 첫째는 표본자료로부터 평균과 분산공식을 이용하여 직접 계산하는 방법이고, 둘째 방법은 정규분포의 누적분포함수(cumulative distribution function)와 자연대수를 취한 소득변수 자료의 누적가중치의 차이를 최소화하여 주는 평균과 분산의 값을 최적화 과정을 통해 간접적으로 추정하는 방법이다.

위의 가점에서 설명하였듯이 일차적으로 경상성장률이 0%라는 가정하에서 각 연도의 원시자료로부터 자연대수를 취한 소득변수에 대해 이상의 방법을 이용하여 소득평균과 분산의 변화추이를 추정한다. 이를 토대로 평균과 분산의 변화패턴을 추출하여 미래의 평균소득과 분산을 예측함으로써 목표연도의 상대소득분포 구조를 추정한다. 이 단계에서 예측된 소득분포구조는 시간이 경과함에 따른 경제성장 효과를 제외하고 있기 때문에 최종적으로는 기준연도⁶⁾와 목표연도 사이의 1인당 경상GDP의 상대비를 곱해줌으로써 목표연도에 대한 장기소득분포 구조를 예측한다.

본 연구에서는 우선 자연대수 소득분포의 평균에 관하여 몇 가지 뚜렷한 안정적 특징을 찾아내어 미래 소득분포의 전망도출에 사용할 수 있음을 보여준다. 첫째, 횡단면에서 연령별 소득이(자연대수 소득의 평균을 기준으로 산정하였을 때) 과거 21년 동안 사실상 불변임을 보이고, 둘째, 이러한 횡단면 연령-소득 곡선이 전반적인 소득수준과 인구의 변화에 따라 변화하는 양상을 계량화하여 2050년까지의 연령별 소득분포 전망을 도출할 수 있음을 보인다. 이 과정에서 연구의 부산물로 우리나라의 경우에서 처음으로 세대별 생애 연령-소득 곡선(cohort life-cycle age-income profile)을 도출하여 제시한다.

본 연구가 제시하고자 하는 것이 궁극적으로는 소득분포의 몇 가지 계수가 아니라 분포 자체임을 상기하면, 자연대수 소득의 평균

6) 본 연구에서는 1995년을 기준연도로 한다.

뿐만 아니라 분산까지 제시하여야 할 것이다⁷⁾. 그러나 평균의 경우와는 달리 연령별 소득분포의 분산의 변화에 대해서는 과거 자료에서 뚜렷한 특징을 발견하기 어려웠으며, 정확한 분산의 전망은 본 연구에 포함시키기 힘든 제반 기타 요인에 대한 보다 광범위하고 심층적인 분석이 요구된다고 하겠다. 그럼에도 불구하고, 본 연구에서는 자연대수 소득분포와 지니계수(gini index)의 정확한 관계를 처음으로 밝히고 이를 계량화하여 제시하였으며, 동 관계를 이용하여 지니계수의 향후 추이에 관한 몇 가지 시나리오로부터 소득분포의 분산을 역산하는 방법으로 소득분산의 전망치를 제시한다.

마. 장기소득분포의 변화가 장기소득세원 분포에 미치는 영향

세제가 일정한 형태로 주어져 있다면, 일반적으로 평균소득이 증가할수록 소득세수가 증가한다. 그런데 세제가 일정하게 주어지고, 경제 전체의 총소득 규모가 변하지 않더라도 상대소득분포의 구조에 따라 해당 경제의 소득세 수입 규모는 달라질 수 있다. 총소득 또는 평균소득규모가 일정하더라도 소득분포의 분산이 커질수록 소득세수는 증가한다. 이는 소득세제가 소득공제제도와 누진세율체계 등의 누진과세체계를 지니고 있기 때문이다. 다시 말해 경제 전체의 총소득규모가 동일하다고 하더라도 소득이 일부의 계층에 집중되어 있다면 그렇지 않은 경제에 비해 높은 한계소득세율을 적용받는 납세자의 비율과 이들의 절대세부담이 커짐으로써 전체적인 소득세수가 증가하게 되기 때문이다.

소득세 부담을 결정하는 데 있어 가장 중요한 요소는 세제와 세정, 소득수준 및 소득분포 등이다. 특히 세정의 경우에는 소득포착률(또는 과표양성화율) 등과 관련하여 매우 중요한 요소이다.

7) 이미 언급하였듯이 소득이 자연대수 정규분포를 따른다는 사실을 보면 이상, 평균과 분산 2개의 계수를 밝히면 분포 전체를 알 수 있다.

그 밖에 장기적으로 소득세수에 영향을 미치는 외부적 요인으로 는 소득분배율, 인구증가율, 경제활동참가율, 실업률, 경제성장률, 물가상승률 등을 들 수 있다.

본 연구에서 분석·예측하고자 하는 것은 장기소득분포이다. 보다 궁극적으로는 장기적으로 소득분포 구조가 변화함에 따라 소득세의 세원에 어떤 영향을 미침으로써 소득세의 세수에 어떤 결과를 초래할 것인가에 대한 것이다. 그런데 다음 장에서 분석해본 결과 소득분포 구조는 나름대로 변화추세가 장기적인 안정성을 가지고 있어 추세분석을 통해 합리적으로 장기구조를 예측할 수 있다. 그러나 세수에 크게 영향을 미치는 세제나 세정의 경우에는 그렇지 않다. 세제·세정부문은 정책적 의지, 제반 과세환경, 재정수요, 과세기술, 납세의식, 조세경쟁 등과 같이 본질적으로 예측이 불가능하거나 분석이 용이하지 않아 불확실성이 매우 높다. 그러므로 현실적으로는 장래의 세제를 미리 예측하는 것이 매우 어려우며, 경우에 따라서는 바람직하지 않을 수도 있다. 따라서 본 연구에서는 그 부분에 대해서는 다루지 않기로 한다. 왜냐하면 소득세는 본질적으로 경제주체의 자원배분을 왜곡시키는 효과를 가지고 있기 때문에 장래의 소득세제에 대한 선부른 예측이 경제주체의 행태를 더욱 왜곡시킬 가능성이 있기 때문이다.

본 연구에서는 장래의 소득세제에 대한 구체적인 모습을 상정하기보다는 현행 소득세제의 기본틀 속에서 소득공제의 수준과 소득세율 과세구간의 범위가 물가변동 또는 경상소득증가율에 연동하여 조정된다는 것을 암묵적으로 가정하여 평균적인 소득자에 대한 실질소득세부담이 대체로 일정하게 유지되는 것을 전제로 하여 논의를 전개한다. 따라서 논의의 핵심은 소득분배율에 대한 전망과 평균소득의 증가추세 및 분산의 변화추이, 인구구조의 변화추이, 경제구조 변화에 따른 산업인력구조의 변화추이 등이 소득세의 세원에 미치는 영향을 분석한다.

II. 분석자료 및 추정기법

1. 분석자료

가. 분석자료의 종류

기본적으로 장기 인력수급추이 분석을 통한 소득세원 분포를 추정하기 위해서는 상당히 긴 기간을 대상으로 축적된 패널자료를 사용하여야 한다. 즉, 각 개인의 생애에 걸친 소득의 흐름을 일생 동안 추적·조사함으로써 현존하는 노동력 또는 미래에 경제활동에 참여하게 될 미래 세대의 노동력에 대한 소득흐름을 예측하기 위해서는 패널자료가 필수적이다.

우리나라에서는 1993~1998년을 대상으로 한 대우패널자료와 1998년부터 자료가 축적되고 있는 한국노동연구원의 한국노동패널 자료의 두 가지 패널자료가 있다. 그런데 불행히도 전자의 경우에는 1999년부터 자료가 단절되어 최신 정보를 알 수 없다. 후자의 경우에도 자료가 축적된 기간이 불과 3~4년에 불과할 정도로 매우 짧아 장기 인력수급 추이를 분석하는 데 있어서는 분명한 한계를 가지고 있다⁸⁾. 그러므로 현 시점에서는 패널자료를 이용하여 일생

8) 그 밖에 노동패널자료는 설문조사를 통한 원자료 구축단계에서 다음과 같은 편이가 있다고 생각하여 본 연구에서는 사용하지 않았다. 노동패널은 1년에 1차례 피조사대상자들을 대상으로 설문조사를 통해 소득세부항목 및 소비세부항목별로 설문조사를 통해 자료가 수집된다. 그런데 이러한 조사내용은 일정 기간을 대상으로 가계부를 작성하여 축적되거나, 또는 근로소득세 연말정산(익년 1월) 또는 종합소득세 신고시점(익년 5월) 등에 즈음하여 조사하여야만 피조사대상자들이 해당 설문항목에 대해 충분히 정확한 정보를 보유할 수 있기 때문에 전년

에 걸친 생애소득의 흐름을 파악하여 장기 인력수급 구조 및 소득 분포를 파악하는 것은 현실적으로 불가능하다고 할 수 있다.

그렇지만 비록 현 시점에서 장기인력수급 구조 또는 장기소득분포 구조를 분석하기 위해 필수적인 패널자료가 사실상 부재하기는 하지만 다행히도 통계청에서 매월 조사하고 분기별로 발표하는 도시가계조사자료를 이용하면 나름대로 이에 대응한 연구결과를 얻을 수 있다. 현재 도시가계조사자료의 경우에는 1982년부터 최근까지 원시자료를 이용할 수 있다. 따라서 이 자료를 활용하여 부족하나마

도 소득·소비에 대한 조사자료의 정확성을 기할 수 있다. 그러나 노동패널자료의 조사시점은 이들 시점과 차이가 크기 때문에 사실상 피조사대상자들이 객관적인 사실에 기초한 정보보다는 주관적이고 불확실한 기억에 의존하여 전년도 소득종류별·소비항목별로 소득·소비 지출 사항을 응답하게 된다. 이 경우 대부분의 사람들이 전년도의 총 소득과 소득종류별 소득금액에 대한 정보를 제대로 기억하지 못할 뿐만 아니라 소비항목의 경우에도 부정확한 기억력에 의존하여 자료를 작성하게 될 가능성이 높다. 그러므로 이러한 방법에 의해 구축된 소득·소비자료는 신뢰성에 대한 의문을 배제하기 어렵다.

바로 이러한 문제점 때문에 미국 상무부(U.S. Department of Commerce) 산하의 조사통계국(Bureau of the Census)을 비롯하여 대부분의 선진국에서는 소득·소비 관련 미시통계자료를 구축함에 있어 기억력에 의존한 설문조사를 행하지 않고 있음에 주목할 필요가 있다. 일례로 미국 상무부 산하의 조사통계국에서 매년 작성·발표하는 CPS (current population survey) 자료의 경우에는, 각 납세자들의 소득 관련 정보를 담고 있는 W2-form이 미국 국세청으로부터 개별납세자에게 통보되어 소득 관련 정보에 대한 내용을 충분히 숙지하고 있는 시점, 즉 소득세 자진신고납부 기간인 2~4월 사이에 조사를 행함으로써 정확성이 높은 소득 정보를 담은 미시 자료를 구축하고 있다. 더욱이 이 경우에는 소비항목에 대해서는 W2-form에 해당 정보를 담고 있지 않으므로 불가피하게 해당 피조사대상자들은 기억에 의존하여 응답하여야 하기 때문에 자료의 신뢰도를 확보할 수 없다는 점에서 소비 관련 항목은 조사대상에서 제외되고 있다. 바로 이러한 사실만 놓고 보더라도 시간이 상당히 지나간 이후의 시점에서 기억에 의존한 채로 수집되는 소득 및 소비금액, 그것도 총액이 아닌 세부항목별 금액에 대해 일정 수준 이상의 정확성을 가지고 피조사대상자들이 정확히 응답할 것으로 기대하는 것은 무리라고 생각된다. 본 연구에서는 이러한 이유로 상기자료보다는 도시가계조사자료를 사용하는 것이 타당하다고 판단하였다.

장기 인력분포 및 소득분포에 대한 정보를 추정할 수 있다.

도시가계조사자료는 매년 또는 정기적으로 표본의 일부를 새로운 표본가구로 교체함으로써 일정기간이 경과하게 되면 이전 기간의 표본가구와는 완전히 다른 새로운 가구로 표본이 구성된다. 그러한 의미에서 도시가계조사자료는 일부 패널자료로서의 특성을 가지고 있기는 하지만 현실적으로는 횡단면자료(cross section data)의 일종으로 보는 것이 적절하다.

본 연구에서는 1982~2002년의 21년간의 도시가계조사 원시자료를 이용하여 분석하였다.

나. 도시가계조사자료의 특성

1) 자료의 가공

통계청이 제공하는 도시가계조사자료는 월별로 조사·축적되며 분기를 단위로 자료가 제공된다. 따라서 연간자료를 생성하기 위해서는 각 분기별 자료를 동일가구별로 재분류하여 해당 항목을 일치시키고 합산한다. 그런데 분기단위의 원시자료에는 조사가 누락되어 특정 분기의 자료정보가 완전히 소실된 가구가 다수 포함되어 있다. 분기별 소득분포와 소비패턴이 가구별로는 상이한 분포를 지닌다고 하더라도 최소한 동일가구 내에서는 동일한 분포를 지닌다면 하나 또는 그 이상의 분기에서 자료가 모두 보고되지 않았다고 하더라도 보고된 분기의 자료만으로도 충분히 일치성을 지닌 소득 및 소비의 연간화 자료를 얻는 것이 가능하다.

그렇지만 이와 반대로 동일한 가구라고 하더라도 만약 분기별로 소득분포와 소비패턴이 계절에 따라 다르게 나타난다면 누락된 분기자료에 대해 추가적인 정보가 없는 한 해당 가구의 경우에는 연간자료에 대한 일치추정량을 얻는 것이 불가능하다.

도시가계조사자료는 기본적으로 계절성(seasonality)을 띠는 것으

로 알려져 있다. 그러므로 도시가계조사자료는 분기자료를 사용하든지 또는 자료를 연간화하여 사용하든지의 여부에 관계없이 계절성으로 인한 편의를 제거하는 것이 필수적이다. 만약 계절성을 제거하지 않는다면 분기별 자료 또는 연간자료를 토대로 추정된 소득불평등도가 계절성을 완전히 제거한 상태에서 나타나는 실제의 소득불평등도보다 과대추정하게 될 뿐만 아니라 추정결과의 일치성(consistency)도 위배되는 결과를 가져다준다. 계절편의를 제거하여야만 일치추정량을 구할 수 있다. 본 연구에서는 어떤 한 분기 또는 2개 이상의 분기에서 자료가 모두 누락된 가구를 분석대상자료에서 제외시켰다.

그런데 이와 같은 작업을 거쳐 계절성을 제거하고 자료를 연간화하게 되면 불가피하게 선택편의(selection bias)의 문제가 개입되는 것으로 알려져 있다. 그러므로 분기단위의 자료를 이용하여 연간자료를 획득함에 있어서는 계절성뿐만 아니라 선택편의까지도 함께 제거해주어야 한다.

본 연구에서는 계절성과 선택편의에 대한 검정은 물론이고, 두 가지 편의를 제거하기 위해 성명재(2002)에서 채택한 방법을 적용하였다. 그의 연구에서는 비모수적 최린추정법(nonparametric nearest neighbor estimation method)⁹⁾을 이용하여 상기의 두 가

9) 최린추정법이란 자료분포가 순탄(smooth)하고 연속미분가능(continuously differentiable)하다는 가정하에서, 추정하고자 하는 점(point) 근방의 자료에 높은 가중치를 주어 해당 점에서 추정하고자 하는 모수의 특성을 추정하는 것을 말한다. 성명재(2002)에서 사용한 최린추정법이란 계절성과 선택편의를 제거하는 과정에서 표본에서 제외되는 자료의 표본추출가중치를 이웃한 표본자료에 배분해주는 방법을 일컫는다. 이와 같이 해주는 이유는 자료분포가 순탄하고 연속미분가능하다면 어떤 점 부근의 자료는 해당 점과 분포특성상 공통점을 매우 많이 가지고 있으며 그 정도는 해당 점과의 거리가 가까울수록 더 커진다고 할 수 있기 때문에, 계절성이나 선택편의와 같은 교란요인(noise)을 제거하기 위해 불가피하게 표본에서 해당 자료를 제외시키더라도 전체적인 표본자료의 분포 특성을 유지하기 위해 해당 자료에 대한 표본추

지 편의를 제거하였다.

도시가계조사자료는 취업근로자가구에 대해서만 소득 관련 정보를 제공하며, 나머지 자영업자가구와 무직가구에 대해서는 소득 관련 정보를 제공해주지 않고 있다. 장기인력수급 구조 분석을 통해 소득세원 분포를 예측하기 위해서는 이들 가구에 대한 소득추정이 필수적이다. 본 연구에서는 성명재(2001)에서 수행한 방법, 즉 소득-소비간의 사후적 상관관계를 이용하여 가구유형에 관계없이 소비-소득의 상관관계가 동일하다는 가정에서 사후적인 의미에서의 소득 함수를 추정하여 비취업근로자가구, 즉 자영업자가구와 무직가구에 대한 소득 정보를 추정하였다¹⁰⁾.

2) 가구유형별 가구 분포

계절성과 선택편의를 제거한 후에 자료의 연간화를 통해 최종적으로 사용된 가구의 수는 1984년의 1,754가구에서 2000년의 4,547가구에 이른다.

전체적으로는 자영업자가구가 전체의 약 3분의 1 수준이며 나머지는 대부분 취업근로자가구로 구성되어 있으며 일부는 무직가구이다¹¹⁾.

도시가계조사자료는 기본적으로 층화무작위추출(stratified random sampling)을 통해 표본이 구성된다. 따라서 각 표본관측치는 서로

출가중치를 인근자료에 배분해주는 것을 말한다. 자세한 내용은 성명재(2002)를 참조하기 바람. 이와 관련된 통계학 이론적 내용은 비모수 또는 반(半)모수적 접근방법에 대해 설명하고 있는 일반적인 통계학 또는 계량경제학 교과서를 참조하기 바람.

10) 이에 대한 자세한 내용은 성명재(2001)를 참조하기 바람.

11) 본 연구에서는 분석의 편의상 무직가구를 광의의 근로자가구에 속하는 것으로 분류하였다. 일반적으로 통계청에서는 취업근로자가구만을 근로자가구로 분류하는데, 본 연구에서는 취업중인 근로자가구는 취업근로자가구로 별도로 호칭하며, 근로자가구라 하면 취업근로자가구와 무직가구를 통칭하기로 한다.

상이한 표본가중치가 부여된다. 그런데 여기에 추가하여 계절편의와 선택편의를 제거하는 과정에서 관측치별로 가중치가 다시 한 번 조정을 받는다. 그러므로 최종적으로 나타난 표본가중치를 기준으로 <표 II-2>와 같이 가구구성비를 구해보면, 표본가중치를 무시하고 절대도수(absolute frequency)를 토대로 작성된 <표 II-1>과 다소 차이를 보인다. 그렇지만 두 가지 기준에 의해 작성된 가구유형별 구성비의 절대치는 큰 차이를 보이지 않는다.

<표 II-1> 도시가계조사자료에 나타난 가구유형별 가구 수

(단위 : 가구)

	전체	근로자가구		자영업자
		전체	취업근로자	
1982	2,399	1,624	1,464	775
1983	2,448	1,674	1,493	774
1984	1,754	1,231	1,095	523
1985	1,937	1,385	1,258	552
1986	2,132	1,509	1,369	623
1987	2,095	1,522	1,386	573
1988	2,583	1,870	1,735	713
1989	3,388	2,332	2,143	1,056
1990	2,853	1,926	1,786	927
1991	2,950	1,987	1,840	963
1992	3,114	2,089	1,934	1,025
1993	3,456	2,414	2,230	1,042
1994	3,168	2,154	1,972	1,014
1995	3,609	2,464	2,242	1,145
1996	3,607	2,434	2,195	1,173
1997	3,618	2,459	2,197	1,159
1998	3,647	2,391	2,103	1,256
1999	3,581	2,390	2,012	1,191
2000	4,547	3,011	2,496	1,536
2001	3,629	2,384	1,982	1,245
2002	3,544	2,350	1,938	1,194

주: 상기 자료는 자료가공을 마친 이후의 연간소득자료를 기준으로 표본추출 가중치를 고려하지 않은 가구 수입.

<표 II-2> 도시가계조사자료에 나타난 가구유형별 가구점유비

(단위 : %)

	전체	근로자가구		자영업자
		전체	취업근로자	
1982	100	67.59	60.80	32.41
1983	100	68.46	60.96	31.54
1984	100	69.98	62.04	30.02
1985	100	70.86	64.39	29.14
1986	100	70.60	63.68	29.40
1987	100	73.23	66.77	26.77
1988	100	72.32	67.16	27.68
1989	100	69.38	64.42	30.62
1990	100	69.15	64.21	30.85
1991	100	68.67	63.56	31.33
1992	100	67.79	62.92	32.21
1993	100	69.68	64.79	30.32
1994	100	68.49	62.88	31.51
1995	100	67.72	62.13	32.28
1996	100	66.67	60.76	33.33
1997	100	66.87	60.42	33.13
1998	100	64.81	57.14	35.19
1999	100	66.17	55.97	33.83
2000	100	65.44	54.51	34.56
2001	100	64.57	54.36	35.43
2002	100	66.27	55.09	33.73

주: 상기 자료는 자료가공을 마친 이후의 연간소득자료를 기준으로 표본추출가중치를 고려하여 추정한 가구 비중임.

3) 평균소득과 평균연령 분포

가구당 평균경상총소득은 1982년 395만원에서 2002년에는 3,259만원으로 8.25배 수준으로 증가하였다. 이를 연간 기하평균증가율로 환산하면 11.1% 정도이다. 가구당 소득은 2002년 현재 자영업자가구가 평균 3,472만원으로 취업근로자가구의 3,287만원보다 185만원 정도 많다.

가구주의 평균연령은 1982년의 경우 38.4세였으나 점차 상승하는 추세를 보여 2002년에는 45.2세에 이른다. 이와 같이 도시가구의 가구주의 평균연령이 점진적으로 상승추세를 보이는 것은 인구의 노

령화에도 일부 관련이 있는 것으로 추측된다.

가구유형별로는 2002년 현재 자영업자가구가 45.7세이고 취업근로자가구는 42.4세로 자영업자가구의 연령이 근로자가구보다 약 3세 정도 더 높은 것을 알 수 있다. 그런데 무직가구를 포함하면 근로자가구의 평균연령은 45.2세로 상승하여 자영업자가구와 큰 차이를 보이지 않는다. 이는 곧 무직가구의 가구주 연령이 취업근로자가구나 자영업자가구에 비해 크게 높다는 것을 의미한다. 또한 이들 중 상당 부분이 노령의 은퇴가구로 구성되어 있음을 시사해준다.

<표 II-3> 가구유형별 평균소득

(단위 : 만원)

	전체	근로자가구		자영업자
		전체	취업근로자	
1982	395	385	386	417
1983	433	421	425	461
1984	482	465	466	522
1985	503	493	498	528
1986	574	559	566	607
1987	673	649	654	738
1988	805	775	780	883
1989	981	948	950	1,058
1990	1,166	1,127	1,137	1,251
1991	1,413	1,374	1,385	1,498
1992	1,644	1,611	1,631	1,715
1993	1,777	1,725	1,748	1,896
1994	2,021	1,979	2,015	2,114
1995	2,300	2,248	2,280	2,410
1996	2,575	2,537	2,583	2,649
1997	2,630	2,508	2,699	2,876
1998	2,304	2,167	2,386	2,555
1999	2,571	2,517	2,619	2,675
2000	2,902	2,798	2,883	3,097
2001	3,038	2,966	3,087	3,168
2002	3,259	3,151	3,287	3,472

주: 상기 자료는 자료가공을 마친 이후의 연간소득자료를 기준으로 가중평균한 값임.

<표 II-4> 가구유형별 가구주의 평균연령

(단위 : 세)

	전체	근로자		자영업자
		전체	취업근로자	
1982	38.4	37.5	36.6	40.4
1983	38.8	37.7	36.7	41.4
1984	37.9	36.6	36.2	41.1
1985	38.5	37.2	36.4	41.7
1986	38.8	37.5	36.9	41.9
1987	38.5	37.2	36.5	41.9
1988	38.9	37.8	37.2	41.8
1989	38.7	37.7	37.2	40.7
1990	39.2	38.4	37.9	40.8
1991	39.9	39.2	38.5	41.4
1992	40.2	39.7	39.0	41.4
1993	39.9	39.4	38.6	40.9
1994	41.2	41.0	39.9	41.6
1995	41.6	41.3	40.3	42.3
1996	42.5	42.1	40.9	43.3
1997	43.1	42.7	41.3	43.7
1998	42.3	42.2	40.6	42.5
1999	43.2	43.1	40.9	43.6
2000	44.2	44.1	42.0	44.4
2001	44.8	44.6	42.2	45.1
2002	45.4	45.2	42.4	45.7

주: 상기 자료는 자료가공을 마친 이후의 가구주 연령을 가중평균한 값임.

2. 커널 추정기법의 소개¹²⁾

어떤 연속미분가능한(continuously differentiable) 확률변수에 대한 확률밀도함수(probability density function)의 추정량(estimator)과 그에 대한 기대치를 각각 $f_N(x)$ (또는 $\hat{f}(x)$)와 $\tilde{f}(x)$ 라고 하자. 확률밀도함수를 추정하는 방법에는 여러 가지가 있는데, 본 연구에서는 커널(kernel)을 이용한 비모수적(nonparametric) 추정방법에

12) 커널추정기법에 대한 보다 자세한 이론적 해설은 부록을 참조하기 바란다.

의한 추정량을 사용하기로 한다. 어떤 점 x 에서의 확률밀도함수에 대한 커널추정량 $f_N(x)$ (또는 $\hat{f}(x)$)은 다음과 같이 정의할 수 있다.

$$f_N \equiv \hat{f}(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{X_i - x}{h}\right)$$

여기서 N 은 표본크기(sample size), $K(\cdot)$ 는 커널, h 는 x 의 근방 또는 주변(cell)의 폭(bandwidth) 또는 창넓이(window size)를 나타낸다. 커널추정량이 일치성을 가지기 위해서 h 는 표본크기가 무한대로 커질수록 0에 접근하는 특성을 가져야 한다. 그런데 만약 표본이 무작위추출법이 아닌 층화무작위추출법(stratified random sampling)으로 추출된 표본이라면, 표본자료 X_i 에 대한 표본추출 가중치를 w_i 라고 할 때 위의 식은 다음과 같이 수정된다.

$$f_N \equiv \hat{f}(x) = \frac{1}{h} \sum_{i=1}^N w_i K\left(\frac{X_i - x}{h}\right) \quad (1)$$

확률밀도함수에 대한 비모수적 추정방법의 하나로 커널추정량이 많이 사용된다. 위의 식 (1)과 같이 소득을 나타내는 정의구역(domain) 내에 있는 어떤 점 x 에서의 확률밀도함수 추정치를 정의할 수 있다. 기본적으로 식 (1)과 같이 확률밀도함수에 대한 추정치를 정의하는 배경에는 확률밀도함수가 순탄(smooth)하고 연속미분가능(continuously differentiable)하다는 정규성 조건(regularity condition)을 충족시켜준다는 가정하에서 어떤 점 x 의 근방(neighbor)에서는 x 와 유사한 분포특성을 가짐으로써 공통적인 특성을 많이 공유하고 있는 반면 x 에서 멀리 떨어질수록 x 의 분포특성과 다른 특성을 더 많이 가지게 됨으로써 동질성이 작아진다는 점에 착안하고 있다.

커널 $K(\cdot)$ 는 일종의 가중치와 같은 역할을 수행하는 것으로서

어떤 점 x 에 가까운 표본관측치에 대해서는 높은 가중치를 부여하고, x 에서 멀리 위치한 관측치에 대해서는 낮은 가중치를 부여하는 역할을 수행한다. 가중치를 부여함에 있어 h 는 가중치 부여강도를 결정해주는 요소이다. h 의 크기를 표본의 크기에 맞추어 적절히 조정해주면 커널추정량의 일치성과 점근적 정규분포 특성을 보장해줄 수 있다.

Ⅲ. 소득분포함수의 추정

1. 연도별 소득분포

가. 연도별 소득분포확률함수의 특성

연령·소득의 결합확률분포함수를 추정하는 작업은 우선 소득에 대한 한계확률분포함수, 즉 소득분포를 각 연도별로 추정하는 데서 시작하였다. 소득은 비대칭적 분포를 띠고 있다는 사실이 선형적으로 알려져 있으므로, 우선은 비대칭 확률분포함수 중 많이 연구되고 특성이 잘 알려진 함수와 비교하여 그 중 기존 소득분포와 가장 유사한 기존의 확률분포를 찾고, 차이가 나는 부분에 대한 세부조정을 통하여 소득분포함수를 근사(approximate)하려는 접근법이 원래 우리의 의도였다. 이에 따라 가장 먼저 도시가계연보 각 연도 자료에서 추출한 소득분포를 자연대수정규(log-normal)분포와 비교하였다. 그런데 의외로 소득분포는 근사할 필요도 없이 바로 자연대수정규분포라는 매우 강력한 결과를 얻었다. 따라서 소득분포함수 추정작업을 대부분 소득분포가 자연대수 정규분포임을 다각도로 검증하는 방향으로 진행되었다. 이하 본 장의 논의에서는 이 과정이 소개되어 있다.

대칭확률분포 중 정규분포(normal distribution)가 대표적으로 널리 통용되듯이 비대칭 확률분포함수 중에는 자연대수정규(log-normal)분포가 가장 널리 쓰인다. 이는 정규분포나 그 변형인 자연대수 정규분포의 특성이 잘 알려졌을 뿐만 아니라 분포함수의 특성을 이용한 여러 가지 추가적 결과의 도출이 용이하여 매우 유용하

기 때문이다. 참고로 비대칭확률분포함수 중 자연대수 정규분포 이외에도 많이 사용되는 분포로는 beta, gamma, weibull 분포 등이 있다.

본격적인 분석에 들어가기에 앞서, 정규분포와 자연대수 정규분포의 확률밀도함수와 양분포의 관계를 구체적으로 보기로 한다. 어떤 변수가 x 가 정규분포를 띠고 있고 평균값이 m , 분산이 s^2 이라고 할 때 일반적으로 이를 $x \sim N(m, s^2)$ 라고 표기하며, 이때 x 의 확률밀도함수 ϕ_N 는 다음과 같다.

$$\phi_N = \frac{\exp\left\{-\frac{1}{2}\left(\frac{x-m}{s}\right)^2\right\}}{s\sqrt{2\pi}}, \quad -\infty < x < \infty$$

이와 관련된 분포로 자연대수 정규분포를 생각할 수 있는 바, 평균 μ , 분산 σ^2 인 변수 y 의 자연대수(natural logarithm)를 취한 값이 정규분포를 하고 있으면 y 가 자연대수 정규분포를 하고 있다고 한다. 이를 $y \sim \lambda(\mu, \sigma^2)$ 라고 일반적으로 표기한다.

만약 자연대수를 취한 결과 변수를 x 라 지칭하고 그 평균과 분산이 각기 m, s^2 이라고 할 경우 $\ln y \sim N(m, s^2)$ 가 되며 원래의 변수 y 는 아래와 같은 자연대수확률 밀도함수를 갖는다.

$$\phi_\lambda(y) = \begin{cases} \frac{\exp\left\{-\frac{1}{2}\left(\frac{\ln y - m}{s}\right)^2\right\}}{ys\sqrt{2\pi}} & , y > 0 \\ 0 & , y \leq 0 \end{cases}$$

이때 (m, s^2) 와 (μ, σ^2) 에는 다음과 같은 관계가 성립하며, 두 개의 평균·분산 순서쌍 중 어느 하나만 알면 다른 순서쌍의 값을 계산할 수 있다.

(m, s^2) 를 알고 있을 경우 다음의 관계를 통해 평균과 분산을 추론할 수 있다.

$$\begin{aligned}\mu &= \exp\left\{\frac{2m+s^2}{2}\right\}, \\ \sigma^2 &= \exp(2m+2s^2) - \exp(2m+s^2)\end{aligned}$$

반대로 (μ, σ^2) 를 알면 (m, s^2) 는 아래와 같이 도출된다.

$$\begin{aligned}m &= \ln\left(\frac{\mu^2}{\sqrt{\sigma^2 + \mu^2}}\right), \\ s^2 &= \ln\left(\left(\frac{\sigma}{\mu}\right)^2 + 1\right)\end{aligned}$$

나. 연도별 소득분포확률밀도함수의 추정

본항에서는 우선 연도별 소득분포함수를 추정하고 그 특성에 대해 살펴본 후, 다시 연도·연령별 소득분포함수의 특성에 대해 살펴보기로 한다. 이하에서는 표기의 편의상 t기의 소득을 y_t 라 표기하고, y_t 의 자연대수(natural logarithm)를 $x_t \equiv \ln y_t$ 라고 쓰기로 한다.

y_t 가 자연대수 정규분포라는 사실을 3가지의 방법을 통하여 보이기로 한다. 첫째, x_t 의 평균과 분산을 도시가계자료에서 그대로 계산하는 경우와 x_t 가 정규분포라는 가정하에 추정한 값을 비교하는 방법이 있다. 둘째, x_t 의 경험분포(empirical distribution)가 이론적 확률분포와 얼마나 일치하는지 도표를 이용한 목측(eyeballing) 방법을 쓸 수 있다. 셋째, 통계적 검정(statistical testing)을 이용한 결과를 제시한다.

1) 평균과 분산의 비교

y_t 가 자연대수 정규분포이거나 이에 가까운 분포일 경우, 그 자연대수를 취한 변수 x_t 는 정규분포이거나 그에 매우 유사한 분포를 띠게 된다. 이 가설을 검정하는 방법 중 일차적으로 가장 손쉬운 방법의 하나로 평균과 분산을 이용하는 방법이 있다. 즉, 도시가계 자료에 수록된 y_t 로부터 자연대수를 취하여 x_t 를 얻은 후, x_t 의 평균과 분산을 계산하는 데 두 가지 방법을 이용할 수 있다. 첫 번째 방법은 분포를 모르고도 표본에서 바로 계산할 수 있는 통상적인 방법을 이용하여 평균과 분산을 구하는 방법이다. 두 번째는 x_t 의 표본이 정규분포에서 나왔다는 가정하에 미지의 평균과 분산을 간접적으로 추정하는 방법이다¹³⁾. 만일 x_t 가 실제로 정규분포라면, 두 가지 방법을 이용해 계산한 평균과 분산이 일치하거나 그 차이가 매우 작을 것이다.

<표 III-1>에는 이와 같이 두 가지 방법을 이용하여 계산한 x_t 의 평균과 분산이 비교되어 있다. 두 가지 계산 방법의 결과를 비교하기에 앞서, 단순 추정방법에 따른 평균과 분산의 값이 연도별로 변하는 추세를 한 번 살펴보기로 한다. <표 III-1>의 두 번째 열을 보면 x_t 의 평균이 1982년부터 2002년까지 꾸준히 증가함을 볼 수 있다. 다만 1997년과 1998년에는 연속하여 2년 동안 자연대수소득 x_t 의 평균이 감소한다. 이는 우리의 경상(가구)소득이 동기간에 꾸준히 증가하다가 1997년과 1998년에 외환위기를 맞아 경제와 소득수준이 마이너스 성장을 기록했던 것과 현상에 부합된다. x_t 의 증가폭을 보면 2002년의 평균값 17.17은 1982년의 평균값

13) 보다 구체적으로는, 표본이 정규분포에서 도출되었다는 가정하에 정규분포 누적분포함수(normal cumulative density function)와 자연대수를 취한 소득변수 자료의 누적가중치의 차이를 최소화하는 평균과 분산의 값을 최적화 과정을 통하여 간접적으로 추정하였다.

15.06보다 2.11, 즉 약 2배 늘었다. x_t 는 가구경상소득 y_t 의 자연대수 값이므로, x_t 가 1 정도 차이날 때 y_t 는 $e \approx 2.78$ 배 정도의 차이가 있다는 해석이 가능하다. 따라서 1982년부터 2002년까지 x_t 의 평균이 약 2배의 차이가 있다는 사실은 y_t 의 평균이 2.78²배, 즉 거의 8배 정도의 수준으로 증가했음을 의미한다¹⁴⁾.

x_t 의 분산의 연도별 추이를 살펴보면 평균의 추이와는 다른 양상을 보이고 있다. 즉, 1982년 이래 외환위기 직전인 1996년까지는 전반적으로 감소추세에 있어 소득분포가 개선되는 추세임을 알 수 있다. 그러나 1997년과 1998년에는 0.20 수준까지 낮아졌던 분산이 0.35와 0.44 수준으로 급격히 증가하여 소득분포에 심한 변동이 있었음을 알 수 있다. 1999년 이래 곧바로 다시 안정되기는 하였으나, 외환위기 직전보다는 높은 수준인 0.25 부근의 수치를 보이고 있어, 외환위기 당시보다는 소득분포가 안정성을 되찾기는 하였으나 위기 이전보다는 악화되었음을 알 수 있다.

다시 평균과 분산의 비교를 통한 x_t 의 정규분포 여부 검증으로 돌아가서, 우선 평균(m)을 비교해보면 1982년부터 2002년까지 매년 도마다 두 가지 방법에 의한 추정치가 거의 일치하고 있음을 알 수 있다. 따라서 적어도 이러한 검증방법으로서는 x_t 가 정규분포가 아니라는 가정을 기각하지 못한다. 여기서 눈여겨볼 것은 1997년과 1998년, 즉 외환위기 전후로 1년 정도의 기간에는 두 가지 방법으로 계산한 x_t 의 평균이 다른 연도에 비해 차이가 다소 커진다는 사실이다.

14) x_t 의 평균의 변화폭이 그대로 y_t 의 평균의 변화로 해석될 수는 없다. y_t 의 평균을 도출하려면 x_t 의 평균뿐만 아니라 x_t 의 분산도 이용해야 한다.

III. 소득분포함수의 추정 55

<표 III-1> 두 방법으로 추정된 도시가계가구의 자연대수소득(x_t)의 평균과 분산

연 도	표본에서 단순 추정		정규분포 가정하에 간접추정	
	평균 (m)	분산 (s ²)	평균 (m)	분산 (s ²)
1982	15.05654	0.26018	15.04247	0.24916
1983	15.15182	0.25549	15.14387	0.25041
1984	15.25906	0.25500	15.25275	0.24158
1985	15.30701	0.24006	15.29301	0.21644
1986	15.43969	0.24227	15.43347	0.22781
1987	15.59492	0.24886	15.58267	0.23852
1988	15.78104	0.23707	15.77491	0.21757
1989	15.98085	0.23019	15.97096	0.21177
1990	16.16728	0.20469	16.16080	0.19040
1991	16.36802	0.19178	16.36966	0.18604
1992	16.51950	0.19396	16.52163	0.17866
1993	16.59524	0.19689	16.59866	0.17543
1994	16.72486	0.19736	16.72922	0.18976
1995	16.85331	0.20054	16.86121	0.19010
1996	16.96280	0.20697	16.97223	0.19997
1997	16.93502	0.35006	16.99110	0.24710
1998	16.76566	0.44026	16.83627	0.32849
1999	16.94336	0.24185	16.95166	0.23340
2000	17.05265	0.26177	17.05913	0.25034
2001	17.10621	0.24542	17.10876	0.23816
2002	17.16859	0.26633	17.17773	0.25947

x_t 의 분산(s^2)의 경우에도 비슷한 현상을 볼 수 있다. 즉, 각 연도마다의 분산을 보면 단순추정치와 간접추정치 간에 서로 큰 차이가 나지 않는다. 다만 앞서 두 방법에 의한 평균치 간의 차이와 비교해 보았을 때 상대적으로 다소 차이가 크다는 느낌이 든다. 그러나 1997년과 1998년에는 두 방법으로 계산한 분산이 크게 차이가 난다. 앞서 평균의 경우보다도 두 방법의 차이가 크다는 점에 주목할 필요가 있다. 결론적으로 평균과 분산을 두 가지 방법으로 도출

하여 비교한 결과, 1997년과 1998년만을 제외하고 1982년부터 2002년 사이의 x_t 의 분포가 정규분포일 것이라는 가정을 기각할 수 없다. 즉, 도시가계자료에서 가구단위의 소득 y_t 가 자연대수 정규분포(log-normal distribution)의 형태를 띠고 있을 가능성이 높다는 사실을 확인할 수 있다. 1997년과 1998년도의 x_t 의 분포는 예외적으로 두 방법에 따른 계산결과가 차이가 커서 정규분포가 아닐 가능성을 강하게 시사한다고 해석할 수 있다.

2) 경험분포와 이론적 확률밀도함수의 목적 비교

위에서 우리가 행한 평균과 분산의 비교를 이용한 검증의 결과 x_t 의 분포가 정규분포라는 가설을 1997년과 1998년의 경우를 제외하고는 기각할 수 없다는 결과를 얻었다. 그러나 이러한 검증결과는 귀무가설의 기각이 모두 그렇듯이, x_t 가 정규분포일 것이라는 가능성을 강하게 시사할 뿐, x_t 를 정규분포로 보아도 무방하다는 의미는 아니다. 후자의 결론을 얻기 위해서는 이와 같이 x_t 의 평균이나 분산 등 일부 계수만 보는 것으로는 부족하며, 따라서 x_t 의 분포도를 직접 보면서 확인하는 절차가 중요하다.

가) 경상 디플레이터의 도출·적용

1982~2002년의 도시가계자료에서 가구소득의 분포도를 작성하기 앞서, 각 연도의 y_t 의 분포가 비슷한 수준에 있도록 디플레이터를 적용하여 조정한 후 자연대수를 취하여 x_t 를 새로이 구하였다. 이는 단순히 연도간 분포의 비교가 용이하도록 하기 위함이며, 각 연도별로 모든 가구의 소득 y_t 에 같은 디플레이터를 적용하기 때문에 우리가 검증하고자 하는 x_t 의 정규분포 여부에는 영향을 미치지 않는다. 이에 사용한 디플레이터는 보통 경상소득으로부터 실질소득을

III. 소득분포함수의 추정 57

계산할 때의 방식과 유사하다. 다만 y_i 가 가구 단위의 소득임을 감안하여 1인당 GDP의 비율을 사용하였으며, 각 연도의 소득분포가 비슷한 수준에 있도록 물가수준만을 감안하지 않고 경상소득의 비율을 사용하였다. 기준 연도는 1995년을 사용하였다. 그 이유는 과거보다는 되도록 최근 연도를 기준 연도로 설정하면서도, 외환위기를 겪는 과정에서 경제에 구조적 변화가 일어났을 가능성을 참작하여 외환위기 이후의 연도를 사용하는 것보다는 이전 연도를 사용하는 것이 타당하다고 생각했기 때문이다. 이러한 방식으로 작성하여 사용된 경상 디플레이터는 아래 <표 III-2>에 제시되어 있으며, 이를 적용하여 다시 구한 x_i 의 평균과 분산은 <표 III-3>에 제시되어 있다. 이 중 간접추정법을 통해 도출된 평균과 분산을 이하에서 검토하는 이론확률분포의 값으로 사용하였음을 밝혀둔다.

<표 III-2> 1인당 경상GDP의 비율로 정의된 디플레이터의 도출

항목	명목 GDP (a)	명목성장률	총인구 (b)	1인당 GDP (c=a/b)	경상 디플레이터 (c/837)
단위	억원	%	명	만원	1995년 기준
1981	473,826		38,723,248	122	0.146
1982	544,313	14.9	39,326,352	138	0.165
1983	638,575	17.3	39,910,403	160	0.191
1984	730,036	14.3	40,405,956	181	0.216
1985	813,123	11.4	40,805,744	199	0.238
1986	948,617	16.7	41,213,674	230	0.275
1987	1,111,977	17.2	41,621,690	267	0.319
1988	1,321,118	18.8	42,031,247	314	0.376
1989	1,481,970	12.2	42,449,038	349	0.417
1990	1,787,968	20.6	42,869,283	417	0.498
1991	2,165,109	21.1	43,295,704	500	0.598
1992	2,456,996	13.5	43,747,962	562	0.671
1993	2,774,965	12.9	44,194,628	628	0.750
1994	3,234,071	16.5	44,641,540	724	0.866
1995	3,773,498	16.7	45,092,991	837	1.000
1996	4,184,790	10.9	45,524,681	919	1.098

<표 III-2>의 계속

항목	명목 GDP (a)	명목성장률	총인구 (b)	1인당 GDP (c=a/b)	경상 디플레이터 (c/837)
1997	4,532,764	8.3	45,953,580	986	1.179
1998	4,443,665	-2.0	46,286,503	960	1.147
1999	4,827,442	8.6	46,616,677	1,036	1.237
2000	5,219,592	8.1	47,008,111	1,110	1.327
2001	5,515,575	5.7	47,342,828	1,165	1.392
2002	5,963,812	8.1	47,639,618	1,252	1.496

주: 경상디플레이터는 1인당 GDP의 비율로서 95년을 기준으로 하여 각 연도 비율을 계산한 것임.
자료: 한국은행, 경제통계, 통계청 통계DB.

<표 III-3> 1인당 경상GDP를 디플레이터로 사용한 x_t 의 평균과 분산의 비교

연도	표본에서 단순 추정		정규분포 가정하에 간접추정	
	평균 (m)	분산 (s ²)	평균 (m)	분산 (s ²)
1982	16.85594	0.26018	16.84187	0.24916
1983	16.80625	0.25549	16.79829	0.25041
1984	16.79198	0.25500	16.78566	0.24157
1985	16.74198	0.24006	16.72798	0.21644
1986	16.73049	0.24227	16.72427	0.22782
1987	16.73667	0.24886	16.72443	0.23852
1988	16.76025	0.23707	16.75412	0.21757
1989	16.85505	0.23019	16.84517	0.21177
1990	16.86363	0.20469	16.85715	0.19040
1991	16.88288	0.19178	16.88452	0.18604
1992	16.91829	0.19396	16.92041	0.17866
1993	16.88248	0.19689	16.88590	0.17543
1994	16.86906	0.19736	16.87342	0.18976
1995	16.85331	0.20054	16.86121	0.19010
1996	16.86887	0.20697	16.87830	0.19997
1997	16.77059	0.35006	16.82668	0.24710
1998	16.62831	0.44026	16.69892	0.32849
1999	16.73027	0.24185	16.73857	0.23339
2000	16.76982	0.26177	16.77631	0.25034
2001	16.77533	0.24542	16.77787	0.23816
2002	16.76582	0.26633	16.77496	0.25947

III. 소득분포함수의 추정 59

나) 확률밀도함수의 추정을 통한 정규분포 여부의 검정

본 절에서는 커널추정법을 통하여 표본의 자료로부터 추정한 경험확률밀도함수를 이론적인 정규분포 확률밀도함수와 비교하고자 한다. 그러나 그에 앞서 독자의 이해를 돕기 위해 커널추정법 대신 가장 단순한 형태인 막대그래프를 사용한 경험확률밀도함수에 대해 잠시 살펴보기로 한다.

아래의 [그림 III-1]에는 1995년의 소득자료로부터 도출된 두 개의 막대그래프(histogram)형 경험확률밀도함수가 제시되어 있다. 그림에서 종 모양의 곡선은 비교를 쉽게 하기 위하여 표본과 동일한 평균과 분산을 갖는 이론상의 정규분포 확률밀도함수를 그린 것이다. 막대그래프를 그리기 위해서는 자연대수 소득을 나타내는 수평축을 여러 개의 구간으로 나눈 다음, 표본에서 각 구간에 해당되는 관측치의 개수를 구한다. 본 연구에서 사용한 자료처럼 가중치가 부여된 경우에는 구간별로 관측된 소득자료의 가중치의 합을 사용한다. 이러한 관측 개수 또는 가중치의 합은 그래프에서 각 구간별 막대의 면적에 해당되므로 개수(가중치)를 각 구간의 길이로 나누면 막대의 높이를 구할 수 있다.

이때 구간들을 동일한 크기로 설정하는 것이 이상적이나, 분포의 꼬리로 갈수록 관측치가 적기 때문에 관측치가 없는 구간이 많이 발생하며, 간혹 관측치가 있는 구간에는 확률밀도가 지나치게 큰 값으로 추정되는 한계가 있다. 이를 피하기 위하여 분포의 중심에서는 구간을 비교적 짧게 설정하고 꼬리 쪽에서는 구간을 점차 크게 설정하는 것이 현실적이나, 표본의 특성에 맞춰 구간을 설정한다는 자의성이 있는 것이 단점이라 하겠다. [그림 III-2]의 경험확률밀도함수는 이와 같은 방법으로 구한 것이다. 이 중 위의 그림을 보면 대충 표본의 소득분포가 이론상의 정규분포와 유사하다는 사실을 알 수 있지만 그래프의 변화가 심하여 자세히 비교하기 힘들다. 그래서 위의 그림보다는 경험확률밀도함수의 순탄성(degree of

[그림 III-1] 막대그래프형 경험확률밀도함수와 이론상의 정규분포 확률밀도함수의 비교

smoothing)을 높이는 방법으로 구간을 더 크게 조정하여 [그림 III-2] 하단의 경험확률밀도함수를 도출하였다. 하단의 경험확률밀도함수를 보면 상단의 그래프보다는 개선되었으나, 역시 이 방법을 이용하여 이론상의 정규분포와 자세히 비교하는 데에는 한계가 있다. 특히 순탄성을 높임으로써 이론상의 확률밀도함수에 그래프가 더 근접하기는 하였으나, 그러기 위해서는 막대그래프의 구간이 넓어지면서 전반적으로 그래프가 정교하지 못하게 된다는 상충관계(trade-off)가 나타났다. 여러 가지 방식으로 구간의 크기를 다르게 조정하여도 [그림 III-2] 하단의 그래프보다 개선되지는 않았다.

본 연구에서는 이러한 막대그래프형 경험확률밀도함수의 한계를 피하기 위하여 커널(kernel)을 사용하여 순탄성(degree of smoothing)이 높은 경험확률밀도함수를 추정하여 사용하였다¹⁵⁾. 앞서 설명한 바와 같이 식 (12)를 기준으로 1982~2002년의 도시가계조사자료에 나타난 가구별 소득자료에 자연대수를 취한 자연대수 소득변수에 대하여 확률밀도함수를 추정한 결과는 [그림 III-2]와 같다. 커널기법을 이용하여 추정한 경험확률밀도함수를 각 연도마다 2개씩 볼 수 있으며, 비교의 편의를 위하여 이론상의 정규분포 확률밀도함수를 음영으로 표시하였다. 추정된 표본분포곡선은 ‘추정분포1’과 ‘추정분포2’의 두 가지이다. ‘추정분포1’은 표본분포 곡선의 모양이 삐죽삐죽한 반면 ‘추정분포2’의 곡선은 상당히 순탄하다(smooth). 이는 h 에 대한 초기값을 어떻게 주었느냐에 따라 서로 상이한 결과를 나타낸 것이다. ‘추정분포1’의 경우에는 해당 표본자료에 대해 폭(bandwidth) h 에 대해 $0.01(=1/100)$ 의 값을 준 경우이며 ‘추정분포2’의 경우에는 N 분의 300(단, N 은 표본크기)의 값을 준 것이다. 전

15) 커널기법의 근본원리는 막대그래프에서 구간을 조정하여 순탄성(smoothness)을 높이는 기법과 유사하나, 보다 세련된 기법을 사용함으로써 자의성을 상대적으로 많이 배제한다고 볼 수 있다. 커널을 이용한 추정의 자세한 방법은 본 보고서의 부록 II를 참조하기 바란다.

술한 바와 같이 커널추정량은 h 의 크기에 따라 다소 다른 모양을 나타낸다. 정확한 표현은 아니지만 이를 굳이 비유를 들어 설명한다면 ‘추정분포1’의 경우에는 저순탄화(undersmoothing) 작업을 수행한 반면 ‘추정분포2’는 고순탄화(oversmoothing) 작업을 수행하였다고 보면 이해가 용이하다¹⁶⁾.

그림을 보면 1997년과 1998년만 제외하고는 1982~2002년의 거의 모든 연도에서 경험확률밀도함수 2개 중 순탄성(smoothness)이 높은 함수의 모양이 종 모양의 정규분포와 매우 유사함을 알 수 있다. 1997년과 1998년의 경험확률밀도함수는 뚜렷한 좌편향(left-skewed) 현상을 보이고 있어, 정규분포와는 상당한 차이가 있다. 특히 눈여겨볼 만한 특징은 분포의 왼쪽 꼬리가 평균값에서 6 표준편차 떨어진 곳까지 뻗어 있다는 사실이다¹⁷⁾.

16) 그러나 이러한 표현이 정확한 것이 아님에 유의하기를 다시 한 번 강조하고자 한다.

17) 이에 더불어, 위의 <표 III-3>에서 볼 수 있듯이, 1997년과 1998년에 자연대수소득의 표준편차가 다른 연도에 비해 상당히 크다는 점까지 감안하면 -6 표준편차는 특히 매우 큰 폭의 소득변화에 해당된다.

III. 소득분포함수의 추정 63

[그림 III-2] 연도별 추정경험분포와 정규분포의 비교(1982~1984년)

- 주: 1. 회색 구역은 경험분포와 같은 평균, 분산의 정규분포 확률밀도함수임.
2. 실선으로 표시된 경험분포 2개는 커널기법으로 추정된 확률밀도함수임 (본문참조).
3. 수평축의 중심은 각 소표본의 평균에 맞추어져 있으며, 수직선의 간격은 각 소표본으로부터 계산한 표준편차의 값을 사용하였음.

[그림 III-2]의 계속(1985~1987년)

Ⅲ. 소득분포함수의 추정 65

[그림 Ⅲ-2]의 계속(1988~1990년)

[그림 III-2]의 계속(1991 ~ 1993년)

Ⅲ. 소득분포함수의 추정 67

[그림 Ⅲ-2]의 계속(1994~1996년)

[그림 III-2]의 계속(1997~1999년)

Ⅲ. 소득분포함수의 추정 69

[그림 Ⅲ-2]의 계속(2000~2002년)

2. 연도별·연령별 자연대수소득 분포

각 연도 내에서도 연령대별로 소득분배구조는 다른 양태를 보일 것으로 추정된다. 따라서 본 절에서는 각 연도의 자료를 연령대별로 25~65세의 41개 연령층으로 나누어 총 861개(=(41개/년)×21년)의 소표본(sub-sample)을 대상으로 자연대수 소득변수에 대한 확률밀도함수를 추정하였다. 본문에는 디플레이터 산출의 기준 연도인 1995년의 25~65세의 연령대에 대한 분포추정결과를 게재하였다 ([그림 III-3] 참조).

제II장에서 서술하였듯이 24세 이하와 66세 이상의 연령대는 편의상 각각 25세와 65세로 편입하였다. 그 이유는 첫째, 24세 이하와 66세 이상의 연령층의 경우에는 표본의 숫자가 매우 작아 독립적으로 분석하기 어려우며, 둘째, 이들 연령층은 경제활동연령층 가운데 중심적인 위치에 있지 않아 소득세원 분포를 추정함에 있어 큰 영향을 미치지 못하며, 셋째, 경제적 관점에서 기능이나 역할이 25세 또는 65세 연령층과 큰 차이를 보이지 않는다고 간주하여도 무리한 가정이 아니라고 판단되었기 때문이다.

그런데 각 연도별로 41개의 연령구간(또는 세대)으로 표본을 나누면 각 연령대별로 귀속되는 표본의 수가 평균적으로 수십개 정도에 불과한 경우가 대부분이다. 따라서 표본의 수가 너무 적으면 표본분포에 대한 추정결과의 표준편차가 너무 커 통계적인 추론이 쉽지 않다. 따라서 본 연구에서는 이러한 문제를 극복하기 위해 다음과 같은 비모수적인 접근방법을 통해 이 문제를 우회하였다. 다행히 연령 변수는 연속분포를 가지고 있으며 비슷한 연령대에서는 표본들이 서로 비슷한 특성을 공유하는 경향이 있다. 그러므로 연령 분포의 연속성(continuity)과 순탄성(smoothness)에 기초하여 ±3세의 범위 내에 있는 관측치에 대해서는 가중치를 가감하여 해당 연령대에 대한 표본분포를 추정하는 데 사용하였다. 주변 연령대에

Ⅲ. 소득분포함수의 추정 71

대한 가중치의 비율은, 해당 연령대의 관측치에 대한 가중비율을 1로 할 때 ± 1 세 관측치는 0.5, ± 2 세의 관측치는 0.25, ± 3 세의 관측치는 0.125, 그 밖의 관측치에 대해서는 0의 가중치를 부여하였다.

[그림 III-3] 1995년의 연령별 자연대수소득의 경험분포와 정규분포 확률밀도함수의 비교(25~36세)

- 주: 1. 회색 구역은 경험분포와 같은 평균, 분산의 정규분포 확률밀도함수임.
2. 실선으로 표시된 경험분포 2개는 커널기법으로 추정된 확률밀도함수임 (본문참조).
3. 수평축의 중심은 각 소표본의 평균에 맞추어져 있으며, 수직선의 간격은 각 소표본으로부터 계산한 표준편차의 값을 사용하였음.

[그림 III-3]의 계속(37~48세)

Ⅲ. 소득분포함수의 추정 73

[그림 Ⅲ-3]의 계속(49~60세)

[그림 III-3]의 계속(61 ~ 65세)

III. 소득분포함수의 추정 75

각 연도 내에서 각 연령대별 표본분포를 추정함에 있어서는 연도별 분포함수를 추정할 때와 마찬가지로 ‘추정분포1’과 ‘추정분포2’의 두 가지로 나누어 분석하였다. 이때 사용된 커널의 폭(bandwidth)은 각각 30분의 1과 N분의 300이다.

연령대별 소표본을 근거로 추정한 자연대수 소득변수에 대한 표본분포 역시 대체로 정규분포와 비슷한 것으로 추정되었다. 20대, 30대, 40대의 연령대에서는 표본분포가 단일정점분포(uni-modal distribution)를 보이는 반면 50대와 60대 가구의 상당 부분에서 미약하나마 평균소득 근처에서 이중정점분포(bi-modal distribution)를 보이는 것으로 나타났다. 대체로 50세 정도까지는 대부분의 사람들이 취업하여 상대적인 소득격차가 작은 편이지만 50세 내외를 기점으로 실직, 은퇴 등이 급격히 증가하여 동일연령대 내에서의 취업자와 비취업자간의 상대적인 소득격차가 벌어지면서 이중정점분포의 형태로 소득분배구조가 변하는 것으로 판단된다. 특히 이러한 경향은 50대 후반에 가장 크게 나타난다.

연령별 자연대수소득의 분포에서 나타나는 또 하나의 특징은 40대 중반 정도부터 50대 중반에 이르기까지 좌편향(left-skewed)된 소득분포가 상대적으로 많이 눈에 띈다는 점이다. 이러한 현상은 [그림 III-4]의 1995년 연령별 분포에서 볼 수 있으며, 아래의 [그림 III-4]에서 제시되었듯이 1982년의 40대 중반 이후, 1991년의 50대, 2002년의 40대 연령층에서도 나타나는 현상이다. 직관적으로는 이러한 현상이 40대 중반 이후 연장자에 대한 노동력(labor force)으로부터의 은퇴 또는 퇴출 압력에 따른 것일 가능성이 생각나지만, 다른 연도의 연령별 자연대수소득 분포에서는 이와 같은 우편향 분포가 발생하지 않으므로 단정하기는 어렵다¹⁸⁾. 또한 위에서 언급하

18) 직장에서 연장자에 대한 퇴출 압력은 경기불황기에 발생할 가능성이 높다. 이에 따라 중년기에 우편향 자연대수소득 분포가 발생하는 연도가 경제의 실질 GDP 성장률이 낮은 연도와 일치하는지 살펴보았으나 특별한 연관성을 발견하지 못하였다.

였듯이 연령별 소표본의 크기가 작아서 추정된 분포에 대한 신뢰도가 상대적으로 낮다는 문제가 있으며, 비록 추정된 경험확률밀도함수의 모양이 이론상의 분포와 차이가 있더라도 다음 절의 통계적 검증의 결과에서 볼 수 있듯이 정규분포라는 가설을 기각하지 못하는 경우가 대부분이라는 사실을 감안할 필요가 있다.

[그림 III-4] 중년기 소득의 좌편향 분포의 예
(1982년 40대 중후반 소득분포의 좌편향성)

- 주: 1. 회색 구역은 경험분포와 같은 평균, 분산의 정규분포 확률밀도함수임.
2. 실선으로 표시된 경험분포 2개는 커널기법으로 추정된 확률밀도함수임 (본문참조).
3. 수평축의 중심은 각 소표본의 평균에 맞추어져 있으며, 수직선의 간격은 각 소표본으로부터 계산한 표준편차의 값을 사용하였음.

Ⅲ. 소득분포함수의 추정 77

[그림 III-4]의 계속(1991년 50대 초중반 소득분포의 좌편향성)

[그림 III-4]의 계속(1995년 40대 중반 소득분포의 좌편향성)

Ⅲ. 소득분포함수의 추정 79

[그림 Ⅲ-4]의 계속(2002년 40대 중반 소득분포의 좌편향성)

앞의 [그림 III-3]의 연도별 자연대수소득 분포에서 보았듯이 외환위기 당시인 1997년과 1998년의 분포는 상당히 심한 좌편향성(left-skewness)을 보였다. 이제 당시의 소득분포를 연령별로 세분하여 검토하기로 한다. 1997년과 1998년의 연령별 자연대수소득 분포를 비교해본 결과 양 연도 간에 큰 차이가 발견되지 않으므로 본문에서는 1997년의 연령별 분포만 제시하였다.

[그림 III-5]에서 볼 수 있듯이 1997년에는 자연대수소득이 정규 분포를 하지 않고 있으며, 중년기뿐만 아니라 25세부터 65세까지 전 연령대에 걸쳐 좌편향적 분포를 하고 있으며, 분포의 왼쪽 꼬리가 평균값에서 5 표준편차 밑의 영역까지 뻗어 있다. 특히 1997년과 1998년 이외의 연도에서는 60대의 자연대수소득 분포가 작은 표본으로 인하여 불규칙적인 모양을 보이면서도 좌편향성을 보기는 힘들었으나, 외환위기 당시에는 60대 초반까지 좌편향적 분포를 보이고 있다. 이는 소득 상실 등 급격한 소득감소 현상이 전 경제에서 광범위하게 일어났음을 그대로 보여주는 것이다.

Ⅲ. 소득분포함수의 추정 81

[그림 III-5] 외환위기시(1997년) 연령별 자연대수소득분포의
좌편향성(25~36세)

- 주: 1. 회색 구역은 경험분포와 같은 평균, 분산의 정규분포 확률밀도함수임.
2. 실선으로 표시된 경험분포 2개는 커널기법으로 추정된 확률밀도함수임
(본문참조).
3. 수평축의 중심은 각 소표본의 평균에 맞추어져 있으며, 수직선의 간격
은 각 소표본으로부터 계산한 표준편차의 값을 사용하였음.
4. 1998년의 연령별 자연대수소득분포도 1997년과 유사한 좌편향적 특성을
보이고 있음.

[그림 III-5]의 계속(37~48세)

Ⅲ. 소득분포함수의 추정 83

[그림 Ⅲ-5]의 계속(49~60세)

[그림 III-5]의 계속(61 ~ 65세)

3. 정규분포 여부의 검정¹⁹⁾

가. 검정통계량

소득에 자연대수를 취한 값에 대한 확률밀도함수를 추정해본 결과 대부분의 경우에 얹어놓은 종 모양(bell-shape)의 형태를 띠고 있어 정규분포와 매우 유사한 것으로 보인다. 이에 본 절에서는 과연 도시가계조사상의 소득자료에 자연대수를 취한 변수가 정규분포를 따르는지를 검정해본다. 만약 검정결과 대체로 정규분포를 따르는 것으로 나타난다면 향후의 소득분포 구조를 예측함에 있어서 시사하는 바가 크다. 즉, 정규분포라 함은 평균과 분산만 추정한다면 분포에 관한 모든 정보를 다 알 수 있게 되기 때문에 장기인력수급 구조 추정을 통한 소득분포 구조를 예측함에 있어 매우 유용한 정보를 제공해줄 수 있기 때문이다.

표본크기가 N인 무작위 표본(random sample)이 정규분포를 따르는지의 여부는 다음과 같은 정규분포 검정(test for normality)을 통해 검정할 수 있다.

‘해당 표본이 정규분포를 따른다’는 귀무가설하에서 표본분포가 정규분포를 따르는지의 여부를 검정하기 위한 검정통계량은 Jarque and Bera(1987)의 방법²⁰⁾을 적용하였으며 검정통계량(test statistic)

19) 이에 대한 자세한 내용은 Greene(2000)을 참조하기 바란다.

20) 이 외에도 통계프로그램에서 일반적으로 쓰이는 정규분포 검정방법은 대략 다음과 같다. 표본의 크기가 3 이하일 때는 Shapiro-Wilk 통계치 W를 사용한다(Shapiro and Wilk 1965). W의 확률분포는 표본의 크기가 3 이하까지는 밝혀져 있으므로, 정확한 통계치 검증이 가능하다. 표본의 크기가 3보다 크나 6 이하일 때는 Royston(1982)에 따라 표본 크기 3에 해당하는 W의 값으로 변환하여 검증이 가능하며, 그보다 큰 표본은 일반적으로 표본 크기 2천 미만까지는 역시 Royston(1982)이 제안한 변환에 의하여 표본크기를 n이라고 하였을 때 통계량 $Z_n = ((1-Wn)^{\sqrt{n}} - \mu) / \sigma$ 를 이용한다. 이때 Z_n 은 표준정규분포를 하며, μ , σ 는 시뮬레이션을 통하여 n의 함수로서 결정되고 Z_n 의

은 다음과 같다²¹⁾.

값이 크면 정규분포 귀무가설이 기각된다. 표본의 크기가 2천 이상일 때는 일반적으로 알려진 Kolmogorov D 통계량을 이용하여 검증한다 (Stephens 1974).

본 연구에서 사용한 Jarque-Bera 기법은 위에 소개한 기법에 비하여 검증의 power가 약한 것으로 알려져 있다. 그럼에도 불구하고 동 기법은 계량경제 연구에서 가장 널리 사용되는 정규분포 검증기법이라 할 수 있다. 본 연구에서도 타 기법보다 Jarque-Bera 기법을 사용한 이유는, 정규분포 검증의 여러 기법들이 공통적으로 점근적 분포를 이용하며, 이때 검증 통계량의 식에 표본 크기 n 이 개입되나, 본 연구에서 사용한 표본에서는 가중치(sample weight)가 사용되어 n 의 개념이 명확하지 않다는 문제가 있기 때문이었다. 즉, 표본에서 제시된 가중치에 대한 해석을 여하히 하는가에 따라 검증 결과가 자의적으로 정해지는 문제가 있었다. 따라서 표본에 있는 가중치의 값들을 면밀히 검토하여 가중치 분포의 특성으로 미루어보아 표본작성 시의 본래 관측치 개수로 추정되는 n 을 역추정하여 검증과정에 대입하였다. 그러나 이러한 노력에도 불구하고 검증과정에서 사용한 n 의 값에 대하여 임의성을 완전히 배제할 수는 없는 일이므로 비교적 개념이 단순한 Jarque and Bera의 기법을 적용함으로써 검증결과의 해석에 있어서 추가적인 혼란을 피하고자 하였다.

- 21) skewness와 kurtosis를 기준으로 정규분포의 여부를 검정하는 데 있어서는 전체 표본 크기에 비해 특이관측치(outliers)가 불과 몇 개에 지나지 않더라도 종종 정규분포의 귀무가설이 기각되는 경우가 많다. 특히 일반적으로 표본의 크기가 상당히 클 경우에는 그러한 가능성이 크게 높아진다. 이는 식 (13)에서 보듯이 검정통계량을 구성하는 요소를 볼 때 표본의 크기가 클수록 검정통계치가 비례하여 커지는 것으로부터 쉽게 알 수 있다.

그런데 본 연구에서 사용하고 있는 도시가계조사자료의 경우에는 표본의 숫자가 최소 1,754개(1983년)에서 최고 4,547개(2000년)에 이를 정도로 표본의 크기가 크다. 따라서 실제의 모집단이 자연대수 정규분포(log-normal distribution)를 따르더라도 몇몇개의 특이관측치로 인해 귀무가설이 기각될 가능성이 높다.

그런데 도시가계조사자료는 기본적으로 층화무작위추출(stratified random sampling)로 표본이 구성될 뿐만 아니라 성명재(2002)에서 보듯이 계절성과 선택편의로 인해 표본조정을 통해 분포상의 편의(bias)를 교정해주는 과정에서 표본가중치가 관측치별로 일정하지 않다. 그러므로 실제의 표본크기(가구 수)와 가중치를 고려한 표본의 크기는 서로 일치하지 않는다. 본 연구에서는 정규분포를 검정함에 있어 이러한 문제를 극복하고자 다른 관측치에 비해 표본가중치가 낮은 관측치의 경우에는 표본의 수를 산정함에 있어 상대도수를 낮게 부여해주고 반대로 가중치가 높은 관측치의 경우에는 상대도수를 높게 부여해주는

$$L = N \left(\frac{\text{skewness}^2}{6} + \frac{(\text{kurtosis} - 3)^2}{24} \right) \quad (13)$$

그런데 skewness와 kurtosis는 일반적으로 각각 $\text{skewness} = E[(X - \mu)^3]$, $\text{kurtosis} = E[(X - \mu)^4]$ 으로 정의된다.

skewness는 어떤 분포가 대칭적인 모습을 보이는가를 살펴보는 것으로서 0의 값을 가지면 분포가 대칭적임을 의미하며 0에서 먼 값을 가질수록 분포가 한 쪽으로 치우쳐 있음을 나타낸다. 만약 skewness가 양의 값을 가지면 분포의 중심이 상대적으로 좌측에 분포하고 오른쪽으로는 분포의 꼬리가 길게 늘어져 있는 경우를 나타내며, 음의 값을 가지는 경우에는 반대의 경우를 나타낸다. kurtosis는 분포의 양쪽 꼬리 부분이 두터운지 아니면 얇은지를 저울질하는 것으로 분포의 꼬리 부분이 두터울수록 큰 값을 가지며, 반대로 분포의 꼬리 부분이 얇을수록 작은 값을 가진다. 따라서 skewness와 kurtosis를 이용하면 분포의 대칭성과 꼬리 부분의 두께를 기준으로 정규분포의 여부를 판가름하게 된다.

그런데 위와 같이 정규분포를 검정함에 있어서는 평균이나 분산의 크기에 따라 검정결과가 다르게 나타날 수 있기 때문에 그러한 문제를 불식시키기 위해 아래와 같이 skewness와 kurtosis를 정규화하여 검정통계량에 사용한다.

$$\text{skewness} = E \left[\frac{(X - \mu)^3}{\delta^3} \right] \quad (14)$$

$$\text{kurtosis} = E \left[\frac{(X - \mu)^4}{\delta^4} \right] \quad (15)$$

방법을 이용하여 층화무작위추출이 아닌 무작위추출(random sampling)에 의한 표본의 크기에 대응될 수 있는 표본의 크기 N을 재산정하여 검정통계치를 계산하였다.

식 (13)과 같이 검정통계량을 설정하면 ‘정규분포를 따른다’는 귀무가설(null hypothesis)하에서 기각역은, 유의수준이 10%, 5%, 1%일 때 각각 4.605, 5.991, 9.210 이상의 영역이다.

나. 검정결과 I: 연도별 자료 분포

식 (13)의 검정통계량을 기준으로 연도별·연령대별 자연대수 소득변수에 대한 정규분포의 검정결과는 각각 <표 III-4>과 <표 III-5>~<표 III-7>과 같다.

먼저 유의수준 10%를 기준으로 할 경우 21개의 연도별 분포 중에서 약 절반 정도인 10개 연도의 분포가 정규분포와 거리가 먼 것으로 검정되었다. 그러나 유의수준을 5% 또는 1%로 낮추면 ‘자연대수 소득변수가 정규분포를 따른다’는 귀무가설을 기각하는 경우가 각각 6개와 4개 연도로 급격히 줄어든다.

그런데 1997~1998년의 경제위기 기간은 매우 예외적인 기간으로 볼 수 있다. 그러므로 이 기간을 제외한다면 유의수준 5%와 1%에서 귀무가설을 기각하는 경우는 총 19개 연도 중에서 각각 4개와 2개 연도에 불과하다. 그러므로 일부의 예외적인 경우를 제외하고는 자연대수 소득변수는 정규분포를 따른다고 볼 수 있다. 더욱이 유의수준 1%에서 귀무가설을 기각하는 1993년과 1995년의 경우에도 [그림 III-2]에서 보듯이 추정분포가 정규분포와 매우 근사한 모습을 보이고 있다. 따라서 비록 정규분포의 귀무가설이 통계적으로 기각되어 정규분포와 다른 분포를 따르더라도 그 차이가 상당히 작아 정규분포로 근사시켜도 큰 무리가 없을 것으로 사료된다.

<표 Ⅲ-4> 연도별 정규분포 검정결과

	검정통계치	유의수준 10%	유의수준 5%	유의수준 1%
1982	4.469	0	0	0
1983	1.447	0	0	0
1984	0.631	0	0	0
1985	6.380	1	1	0
1986	1.258	0	0	0
1987	2.008	0	0	0
1988	4.054	0	0	0
1989	5.195	1	0	0
1990	6.182	1	1	0
1991	0.809	0	0	0
1992	5.067	1	0	0
1993	18.286	1	1	1
1994	3.234	0	0	0
1995	14.518	1	1	1
1996	5.939	1	0	0
1997	281.657	1	1	1
1998	182.699	1	1	1
1999	4.034	0	0	0
2000	3.690	0	0	0
2001	5.515	1	0	0
2002	3.533	0	0	0
계	-	10	6	4

주: 0은 '정규분포를 따른다'는 귀무가설을 기각하지 못하는 경우를 나타내며, 1은 주어진 유의수준하에서 귀무가설을 기각하는 경우임(귀무가설 기각 시 음영처리를 하였음).

다. 검정결과 Ⅱ: 연도별·연령별 자료 분포

연도별로 각 연령대(세대)의 자연대수 소득분포에 대한 정규분포 검정결과는 유의수준 10%, 5%, 1%에서 각각 <표 Ⅲ-5>~<표 Ⅲ-7>과 같다. 1개 연도당 총 41개 연령층이 존재하므로 21개년 자료는 총 861개의 소표본으로 세분된다. 861개의 소표본 가운데 유의수준 10%에서 '정규분포를 따른다'는 귀무가설이 기각된 경우는 64개, 유의수준 5%와 1%에서는 각각 44개와 19개이다. 이를 비율

로 환산하면 각각 7.4%, 5.1%, 2.2%이다. 더욱이 경제위기 기간인 1997~1998년은 매우 예외적인 경우이므로 경제위기의 2년을 분석 대상에서 제외하면 총 779개 소표본 가운데 귀무가설을 기각하는 경우는 유의수준별로 각각 19개, 13개, 3개로 급감하여 귀무가설 기각비율은 각각 2.4%, 1.7%, 0.4%로 크게 줄어든다.

유의수준은 말 뜻 그대로 귀무가설이 참(true)이라고 하더라도 표본자료를 이용하여 산출한 검정통계치가 기각역(rejection region)에 속하게 되어 귀무가설을 기각하게 되는 확률을 의미한다. 그런데 위에서 보듯이 경제위기 기간인 1997년과 1998년을 제외할 경우 각 소표본이 정규분포를 따른다는 귀무가설을 기각하는 비율이 유의수준을 크게 밑돈다. 이는 곧 비록 몇몇 소표본에서 정규분포를 따른다는 귀무가설이 기각되더라도 이는 주어진 유의수준하에서 충분히 기대할 수 있는 사건이라고 볼 수 있다. 더욱이 경제위기 기간인 1997~1998년을 포함한 경우에도 기각되는 소표본의 비율이 유의수준 1%인 경우를 제외하고는 유의수준보다 낮게 나온 것을 보더라도 각 연령대별 자연대수를 취한 소득분포가 정규분포를 따른다고 보아도 무방함을 의미한다고 할 수 있다. 즉, 소표본별 표본관측치의 수가 크지 않고, 연령대를 세분하였음에도 불구하고 이와 같이 귀무가설을 기각하는 비율이 낮게 나타난다는 점은, 현실적으로 각 연령대별 자연대수 소득변수의 분포가 대체로 정규분포를 따른다고 보아도 무방함을 시사해준다고 할 수 있다.

1997~1998년을 제외할 경우, 검정결과의 특징적인 점을 요약하자면, 20대 후반에서 30대 초반, 그리고 40대 일부의 연령층에서 정규분포의 귀무가설을 기각하는 빈도가 다른 연령층보다 상대적으로 높다는 점, 이와 반대로 50대 중반 이후의 연령층에서는 귀무가설을 기각하는 경우가 사실상 없다는 점, 1990년대 이후에는 20대와 30대의 연령층에서도 귀무가설을 기각하는 경우가 거의 없다는 점을 들 수 있다.

Ⅲ. 소득분포함수의 추정 91

그러나 전반적으로 모든 연령층의 분포에 대한 정규분포 검정결과를 종합해보면, 모든 연령층을 총괄한 연도별 분포에서뿐만 아니라 각 연령대별 분포도 사실상 정규분포를 따른다고 보아도 별 무리가 없을 것으로 사료된다.

<표 III-5> 연도별·연령별 정규분포 검정결과(유의수준=10%)

	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	계
25	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
26	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	3
27	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	4
28	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2
29	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	4
30	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	3
31	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	3
32	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	3
33	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	3
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
38	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	3
39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	2
44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
45	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	2
46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
48	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2
49	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2
50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
53	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
54	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	2
55	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
56	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
57	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
63	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
64	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
65	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
계	2	0	0	6	1	3	2	0	2	0	2	0	0	0	0	25	20	0	0	1	0	64

주: 0은 '정규분포를 따른다'는 귀무가설을 기각하지 못하는 경우를 나타내며, 1은 주어진 유의수준하에서 귀무가설을 기각하는 경우임(귀무가설 기각 시 음영처리를 하였음).

<표 III-6> 연도별·연령별 정규분포 검정결과(유의수준=5%)

	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	계
25	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
26	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
27	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
28	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
29	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2
30	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
32	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	3
33	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	3
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2
41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
45	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	2
46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2
49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
53	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
54	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	2
55	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
56	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
57	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
63	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
64	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
65	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
계	2	0	0	4	0	1	1	0	2	0	2	0	0	0	0	17	14	0	0	1	0	44

주: 0은 '정규분포를 따른다'는 귀무가설을 기각하지 못하는 경우를 나타내며, 1은 주어진 유의수준하에서 귀무가설을 기각하는 경우임(귀무가설 기각 시 음영처리를 하였음).

Ⅲ. 소득분포합수의 추정 93

<표 III-7> 연도별·연령별 정규분포 검정결과(유의수준=1%)

	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	계	
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
32	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	2
33	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	3
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	2
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	2
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	2
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	2
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
53	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
55	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
56	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
57	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
63	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
64	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
65	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
계	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0	9	7	0	0	0	0	0	19

주: 0은 '정규분포를 따른다'는 귀무가설을 기각하지 못하는 경우를 나타내며, 1은 주어진 유의수준하에서 귀무가설을 기각하는 경우임(귀무가설 기각 시 음영처리를 하였음).

4. 소 결

가. 가구소득의 자연대수 정규분포 규명의 의의

본 장에서는 1982년부터 2002년까지의 도시가계자료를 이용하여 가구 단위의 소득이 자연대수 정규분포를 따른다는 사실을 통계적으로 밝혔다. 보다 구체적으로는, 외환위기 당시인 1997년과 1998년만을 제외하고는 (가구주의) 연령을 구분하지 않은 각 연도의 전체 소득분포나, 25세부터 65세까지의 각 연령층의 소득분포나 모두 자연대수 정규분포로 간주해도 통계적으로 유의하다는 결과를 얻었다. 본 절에서는 이러한 발견이 함축하는 의미를 정리한다.

지금까지는 우리나라의 소득분포에 대해서는 경험적으로 관측된 자료를 근거로 과거 추이를 기술적(descriptive)으로 설명하는 데 그쳤다는 점에서 한계를 노정하였다고 할 수 있다. 즉, 소득분포의 성질을 어떠한 특성으로 함축적으로 나타내지 못하였다. 따라서 기존의 연구방법으로는 과거의 현상을 기술하고 설명하는 데 그쳤고, 경험적으로 안정적이라고 밝혀진 소득분포의 특성(stylized fact)을 이용하여 장래의 추이를 전망하기는 어려웠다. 이러한 상황에서 소득분포의 매우 중요한 특성으로서 자연대수 정규분포임을 밝혔다는 수확은 그 자체로서 매우 의미가 크다고 할 수 있으며, 경제전망이나 조세, 사회보장 등 분야에서 정책대안의 모의실험에 크게 기여할 수 있을 것으로 기대된다.

또한 소득이 자연대수 정규분포, 즉 자연대수소득이 정규분포임을 따른다는 점을 밝힘으로써, 정책운영에 중요한 요인으로 작용하는 소득분배의 형평성을 가늠하고 평가하는 새로운 시각이 제시되었다는 점 역시 성과로 꼽을 수 있다. 지금까지는 지니계수(gini index)로 측정된 소득분배의 변화 정도만을 생각할 수 있었으나, 본 연구의 결과로 통상적인 경기변화나 분배구조의 변화는 거의 예

외 없이 (자연대수)소득이 정규분포의 특성을 유지하면서 소득의 평균과 분산이 함께 변화하는 데 따른 것임을 알 수 있다.

반면, 외환위기 등과 같이 경제에 매우 드문 큰 충격이 가해졌을 때는 소득이 정규분포에서 이탈하면서 분배가 더욱 악화되는 것을 확인할 수 있었다. 즉, 통상적인 소득분배의 개선이나 악화는 소득의 분포가 자연대수 정규분포의 성질을 유지한다는 점에서 상대적으로 변화의 폭이 제약을 받으나, 경제가 극심한 충격을 받았을 때는 정규분포라는 제약을 벗어남으로써 보다 큰 폭의 소득분배 악화가 일어남을 볼 수 있었다.

본장에서 본 바와 같이 외환위기 당시인 1997년과 1998년의 자연대수소득 분포가 상당한 좌편향성(left-skewness)을 보였다는 사실의 발견은 나름대로의 의미를 갖는다. 외환위기의 발발이 1997년 말임을 상기할 때, 1997년 중 대부분의 기간은 비록 소득분포가 어느 정도 악화되었다고 해도 극심한 충격하에서 나타나는 좌편향적 분포보다는 대칭적 정규분포의 형태였을 것으로 생각하기 쉽다. 실제로 1997년 3/4분기까지의 거시통계에서는 경제상황이 이미 극도로 악화되었다는 징후는 찾아보기 힘들다. 이러한 경우, 본 연구에서 행한 방법으로 1997년의 자연대수소득 분포를 추정하면 정규분포를 도출하거나, 좌편향적 분포가 나타나더라도 편향성의 정도가 1998년에 비하여 매우 낮은 수준일 것이라고 기대할 수 있다.

그러나 본 연구에서는 1997년의 소득분포가 외환위기 발발 이후 경제에 대한 타격이 가장 심했던 1998년의 소득분포와 매우 유사한 수준으로 왜곡되었음을 발견하였다. 달리 이야기하자면, 외환위기가 터져서 경제붕괴가 시작된 1997년말 이전에 이미 우리 경제는 보통의 불경기보다는 훨씬 심각한 구조적 문제의 표출이 이미 상당히 진전되어 있었다는 해석이 가능하다. 이는 본 연구가 발굴한 중요한 결과의 하나로 생각할 수 있으며, 보다 세밀한 추가적인 연구를 요구하는 결과라고 생각된다.

III. 소득분포함수의 추정 97

대수소득 분포와 동일한 정보를 나타내고 있다.

본 장에서 각 연도마다 연령별 소득분포 역시 구하였는 바, 여러 개의 q^{ms}_t 를 취합하면 Q^{ms} 에 담겨 있는 정보와 동일하다. 이제 Q 에서 각 qt 간의 관계의 특성을 파악하면 Q 를 t 에 대해 일반화함으로써 $t=2003, 2004, 2005, \dots$ 의 장래에 대해서도 qt 의 예측치를 구할 수 있다. 즉 장래의 소득분포를 구체적으로 전망할 수 있게 된다. 본고에서는 다음 장에서 이 중 Q^m 만을 파악하기로 한다²²⁾.

22) Q^{ms}_a , 즉 자연대수 소득분포함수의 연령별 평균 이외에 연령별 분산 및 확률밀도까지 규명하면 연령-소득 결합분포함수를 추정할 수 있으나, 제I장에서 밝혔듯이 본 연구에서는 Q_m 을 직접 밝히고 Q_s 를 간접 추정하는 것으로 연구의 범위를 국한하였다.

IV. 연령별 상대소득분포의 추정

1. 연령-소득 곡선의 도출

본 장의 목적은 Q^m 의 속성을 파악함으로써 $t > 2002$ 인 장래의 연령별 소득수준을 전망할 수 있는 근거를 찾는 데 있다. 이를 위하여 우선 아래 [그림 IV-1]에서 1982, 1992, 2002년도의 연령별 평균소득을 그려보았다. 그림에서 보이는 3개의 곡선은 각 연도별 횡단면 연령-소득 곡선(cross-section age-income profile)이라는 해석이 가능하다. 세 개의 곡선이 곡률(curvature)이 다르지만 대체로 형태가 비슷하다는 사실로 미루어 보아, 각 연령에서 세 곡선의 수직구간 차이, 즉 연령별 평균소득의 차이는 대체로 실질성장률과 물가상승률, 즉 경제전반의 경성성장에 의한 것으로 짐작된다.

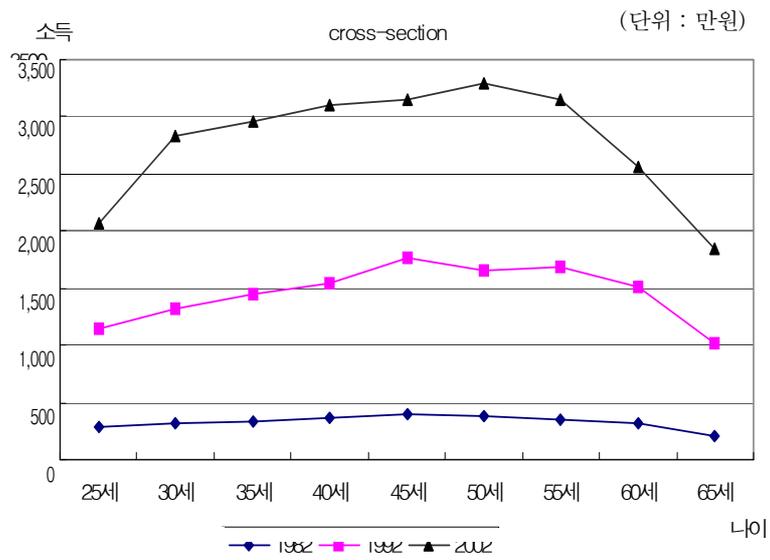
부연하자면, 소득분포는 시간이 경과함에 따라 물가상승과 실질소득이 성장함에 따라 직접적인 비교가 어렵다. 그러므로 본장에서는 1995년의 1인당 경성GDP를 기준으로 각 연도의 1인당 경성GDP에 대한 상대비만큼 해당연도의 모든 소득자료를 상수배로 할증 또는 할인하여, 경제성장이나 물가변동 등에 의한 소득변화요인을 제거하고 순수히 상대소득분포의 변화효과만을 비교하도록 하였다. 이로부터 상대소득분포의 변화추이에 대한 추세적 특징을 추출한다.

다시 말해서, [그림 IV-1]에서 보이는 3개 곡선의 차이는 세 가지 요인, 즉 경제전반의 실질성장, 물가상승, 그리고 연령간 순수상대소득분포의 변화에 기인한다고 볼 수 있다. 실질성장률과 물가상승은 경제 전반에 걸쳐 연령층에 구분없이 적용되는 변화인 반면, 연

IV. 연령별 상대소득분포의 추정 99

연령간 순수상대소득분포의 변화는 이러한 경상성장 효과를 배제한 후 연령간 상대소득 수준이 변했음을 의미한다. 그림에서는, 실질성장률과 물가상승 등 경상성장 효과를 배제한다면, 3개의 곡선이 대체로 같은 소득 수준에 있을 것이며, 이를 3개 연도의 연령간 순수상대소득분포 곡선이라고 해석할 수 있다. 이때 경상성장 효과를 배제한 3개 곡선의 곡률 등 모양에 차이가 있다면, 시간이 경과함에 따라 연령간 순수상대소득분포가 변했다는 것을 의미할 것이다. 본 장에서는 이와 반대로, 연령간 순수상대소득분포가 변하지 않음을 보이려고 한다.

[그림 IV-1] 3개 연도의 연령별 평균소득 수준의 비교



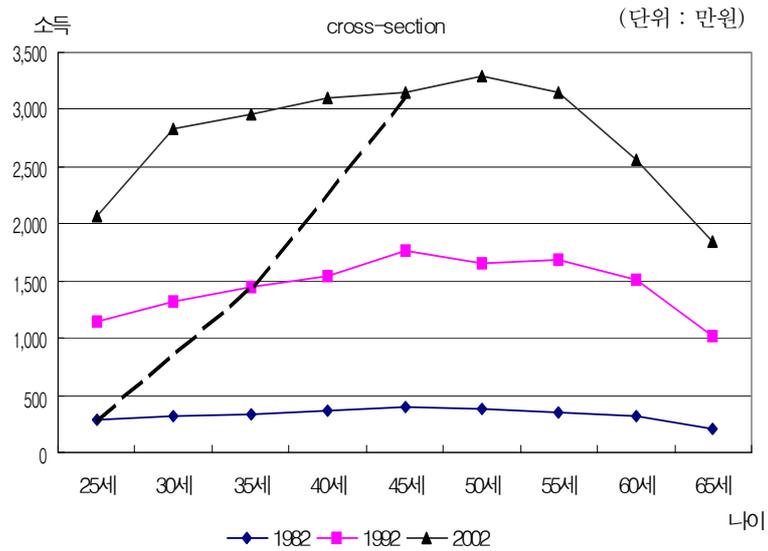
연령-소득 분포의 연도별 변화가 대부분 경상성장률로서 설명될 수 있을 것이라는 가설에 대해 보다 자세히 검증하기 위하여 횡단면이 아닌, 세대(cohort) 연령-소득 곡선을 그려보았다. 세대별 연

령-소득 곡선이란 특정 세대를 시간과 연령이 증가함에 따라 통시적으로(longitudinally) 추적하면서 소득 수준의 변화를 살피는 것이다. 예를 들어 [그림 IV-1]을 보면 1982년에 25세였던 세대는 1992년에 35세, 2002년에 45세가 되므로, 동 세대의 소득을 해당 연도의 횡단면 연령-소득 곡선에서 찾아 연결함으로써 점선으로 표시된 세대별 연령-소득 곡선을 구할 수 있다. 이때 세대를 구분(identify)하기 위하여 특정 연도와 당시의 세대 연령을 알아야 하며, 그림의 예에서는 1982년에 25세인 세대의 연령-소득 곡선이 도출되어 있다.

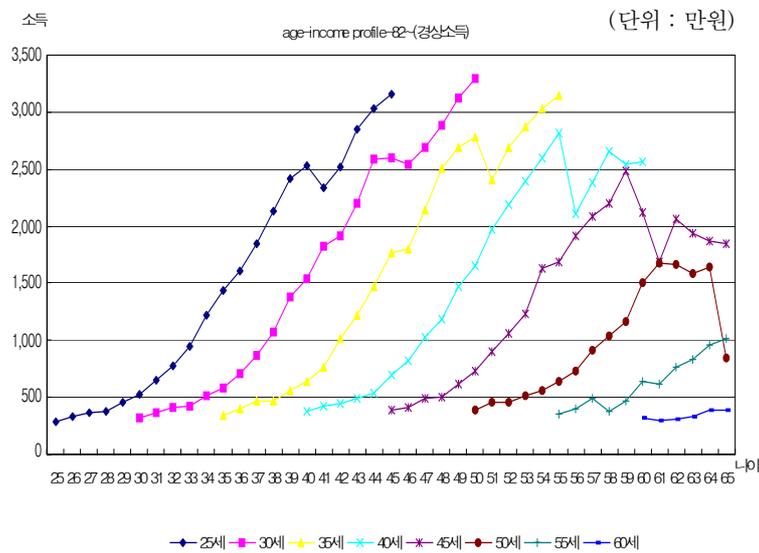
이러한 방법으로 1982~2002년의 도시기계자료에서 세대 연령-소득 곡선을 도출하여 [그림 IV-2]와 같이 그려보았다. 그림에서 여러 개의 연령-소득 곡선이 5세 간격으로 제시되어 있는바, 각 연령-소득 곡선의 맨 왼쪽 관측치는 1982년의 당해 연령 가구주의 평균소득이며, 각 연령-소득 곡선을 따라 우측으로 이동하면서 그 다음해의 도시기계자료에서 한 살 많은 연령층의 평균소득을 취하였다. 예를 들어 최좌측 선에서 최좌측 관측치는 1982년 25세 가구주의 평균소득이며, 같은 선상의 최우측 관측치는 2002년 46세 가구주의 평균소득이다. 도시기계자료가 1982년부터 2002년까지 21년치 밖에 없으므로 곡선들은 전 연령구간을 커버하지 못하고 있다. 그림에서 연령-소득 곡선이 겹쳐지지 않는 것은 방금 이야기하였듯이 주로 경상성장의 차이에 기인하는 것으로 짐작된다.

IV. 연령별 상대소득분포의 추정 101

[그림 IV-2] 세대 연령-소득 곡선의 도출(1982년 당시 25세 세대의 예)



[그림 IV-3] 도시가계의 경상소득 기준 세대 생애 연령-소득 곡선



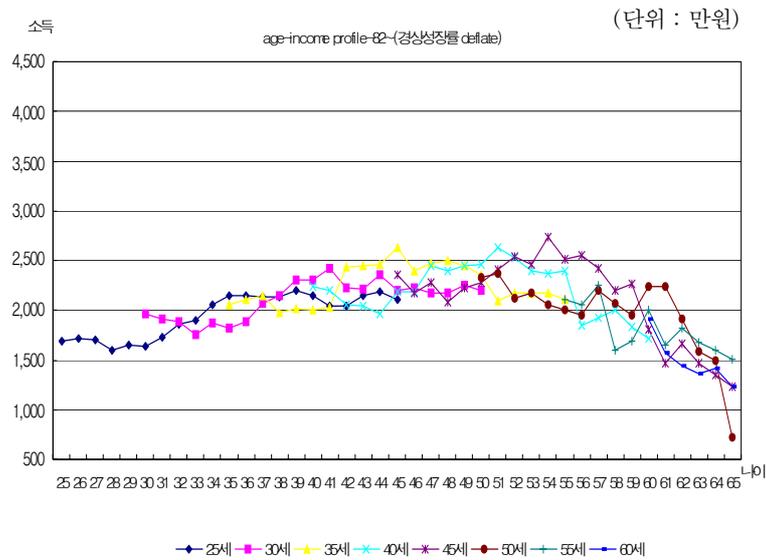
우리의 가설이 맞다면 위 [그림 IV-3]에서 사용한 평균경상소득 대신 각 소득 관측치를 경제의 경상성장률로 디플레이트한 후 다시 이의 연령별 평균을 구하여 사용하면, 여러 개의 연령-소득 곡선이 서로 겹쳐지면서 25~65세까지의 전 연령 구간에 대하여 한 개의 연령-소득 곡선이 구하여질 것이다. [그림 IV-4]에는 이러한 방식으로 앞서 제Ⅲ장에서 구한 1995년 기준 소득으로 하여 경상성장률로 디플레이트 시킨 소득을 사용하여 연령-소득 곡선을 다시 그렸다. 그 결과 여러 개의 연령-소득 곡선이 겹쳐지면서 대강 하나의 곡선의 모양이 드러났다. 이러한 추정결과로 미루어 짐작해 보건대 우리의 추측이 타당한 것으로 사료된다.

경상성장 효과를 배제한 순수 연령간 상대소득 분포, 즉 순수 연령-소득 곡선을 보다 정교하게 구하기 위하여 도시가계 각 연도의 소득자료에 경상성장 디플레이터를 적용한 후, 자연대수를 취하여 평균을 구하였다²³⁾. <표 IV-1>에서 보면 21개년별로 각 연령층의 평균자연대수소득을 구하여 보면, 표준편차가 상당히 작으므로 경상성장률로 디플레이트 시킨 연령별 평균자연대수소득이 연도에 따라 크게 변하지 않음을 알 수 있다. 따라서 21개 연도의 연령별 평균 자연대수소득의 평균을 다시 취하면 이를 q^m_{1995} 로 해석할 수 있다. [그림 IV-4]에서는 이렇게 구한 q^m_{1995} 를 다시 그림으로 제시하였다.

23) 위의 그림에서는 평균소득을 이용하여 연령-소득 곡선을 그렸으나, 본장에서 궁극적으로 구하려는 Q^m 은 자연대수소득의 평균이므로, 평균경상소득 대신 자연대수소득의 평균을 바로 구하였다.

IV. 연령별 상대소득분포의 추정 103

[그림 IV-4] 도시가계의 1인당 실질소득 기준 세대 생애 연령-소득 곡선



이러한 결과를 이용하여 1995년 이외의 q^m_t 는 다음과 같이 구할 수 있다. q^m_t 와 q^m_{1995} 의 차이는 경상성장률 디플레이터로 설명되므로, 다음이 성립한다²⁴⁾.

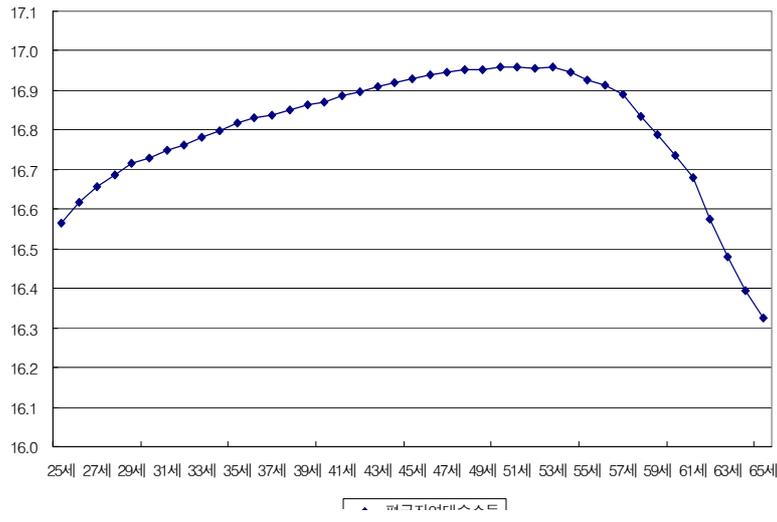
$$m_{t,a} = m_{1995,a} + \ln(\text{디플레이터}_t), \quad a = 25, 26, \dots, 65.$$

24) 자연대수를 취하기 전에 각 가구의 소득 y 를 해당연도의 디플레이터로 나누었으므로, 자연대수소득 x 의 평균은 디플레이터를 적용하지 않은 자연대수소득과 $\ln(\text{디플레이터})$ 만큼 차이가 나게 된다.

<표 IV-1> 도시가계(1982~2002년)의 연령별 평균 자연대수 실질가구소득의 분포

연령 (a)	m_{ta} 의 분포 (t = 1982, ... , 2002; 21개년)			
	평균	표준편차	최소값	최대값
25	16.5633391	0.0889748	16.3463867	16.6731995
26	16.6172057	0.0768850	16.4492268	16.7248682
27	16.6565797	0.0703560	16.5115383	16.7522315
28	16.6846892	0.0716395	16.5663847	16.7776541
29	16.7158757	0.0666811	16.6129048	16.7982411
30	16.7305352	0.0593853	16.6301040	16.8264187
31	16.7476801	0.0572150	16.6459605	16.8323471
32	16.7629453	0.0544487	16.6394152	16.8544241
33	16.7812291	0.0527312	16.6637091	16.8587787
34	16.7970238	0.0517958	16.6781815	16.8698479
35	16.8164260	0.0520657	16.7182085	16.8948262
36	16.8317434	0.0590097	16.7213942	16.9251779
37	16.8376304	0.0626434	16.7182470	16.9418841
38	16.8503882	0.0650736	16.7486584	16.9614256
39	16.8622268	0.0604326	16.7374185	16.9599614
40	16.8713381	0.0637852	16.7711629	16.9684237
41	16.8876936	0.0700971	16.7941117	16.9989609
42	16.8968173	0.0730168	16.8032234	17.0063633
43	16.9092726	0.0737210	16.8082775	17.0421214
44	16.9195759	0.0710888	16.8150807	17.0721699
45	16.9296796	0.0667307	16.8265996	17.0809184
46	16.9379951	0.0577250	16.8532929	17.0792410
47	16.9449883	0.0563107	16.8531432	17.0456392
48	16.9507237	0.0643992	16.8517836	17.0750581
49	16.9538169	0.0711869	16.8278364	17.0891556
50	16.9582967	0.0705772	16.8190060	17.0683684
51	16.9579516	0.0771820	16.8372303	17.0697668
52	16.9557076	0.0835814	16.8388267	17.0783266
53	16.9573896	0.0934222	16.7962282	17.1041053
54	16.9441133	0.0969989	16.7847358	17.1050934
55	16.9270691	0.0979495	16.7500762	17.0979628
56	16.9143853	0.0914893	16.7346524	17.0747459
57	16.8890145	0.1057449	16.7136202	17.0598939
58	16.8338451	0.1250704	16.6378975	17.0477074
59	16.7876665	0.1203852	16.6233398	17.0252384
60	16.7364138	0.1268398	16.5431035	16.9353489
61	16.6786270	0.1178656	16.4833248	16.9066265
62	16.5738424	0.1195459	16.2497873	16.7599128
63	16.4803468	0.1394753	16.0343358	16.6077141
64	16.3953809	0.2050374	15.7265201	16.6479930
65	16.3241834	0.2509650	15.4827223	16.6103561

[그림 IV-5] 도시가계의 연령별 평균 자연대수 가구소득
(경상성장률 디플레이터 적용)



이상의 논의에서 우리에게 필요한 q^m_t 의 속성을 파악하여 구체적인 값을 추정하였다. 그런데 다음 논의로 넘어가기에 앞서, 우리가 구한 q^m_{1995} 의 의미를 다시 보기로 하자. q^m_{1995} 는 정의상 일종의 횡단면 연령-소득 곡선임이 분명하나, 경상성장이 전혀 없는 경제에서는 통시적 세대 연령-소득 곡선(longitudinal cohort age-income profile)도 된다. 이러한 경제는 슈페터가 이야기한 정태적 단순 순환경제(circular flow economy)와 동일한 사고모의실험(thought experiment)이라고 해석할 수 있다.

그런데 본장에서 밝힌 바로는, 경상성장이 있는 현실에서도 연령간의 순수상대소득분포는 변하지 않으므로 q^m_{1995} 와 연령에 대해 불변인 일정 상수항의 차이만으로 q^m_t 를 손쉽게 구할 수 있다. 따라서 이러한 방식으로 아래의 Q^m 행렬을 구하면, 비록 41년간 관측된 패널자료가 없는 우리 현실에서도 25~65세의 전 근로생애

(working lifetime)에 걸쳐 세대 연령-소득 곡선(cohort age-income profile)을 구할 수 있다. 먼저 본장에서 제시한 방법으로 Q^m 을 구한 후, 동 행렬의 대각선 방향으로 m 의 값을 읽으면 세대 별 생애 연령-소득 곡선이 되는 것이다. 즉, t 연도에 가구주 연령이 i 인 가구의 평균 자연대수 소득을 y_{it} 라고 표기하였을 경우, t 연도의 횡단면 연령-소득 곡선은 $\{y_{it}\}_{i=25, \dots, 65}$ 라고 수열형태로 표시할 수 있다. 반면, t_0 연도에 출생한 세대의 생애 연령-소득 곡선은 $\{y_{it}\}_{i=25, \dots, 65; t=t_0+i}$ 라고 표기할 수 있다.

마지막으로 다음 절의 논의로 넘어가기 앞서, 위의 [표 IV-1]에서 볼 수 있는 또 하나의 중요한 결과를 지적하고자 한다. 각 연령 별로 1982~2002년의 21년간 (디플레이트된) 자연대수 가구소득의 평균에 대하여 다시 평균을 구하였을 때, 50세 이전의 연령층에서는 표준편차가 비교적 작아서 21년간 큰 변화가 없었음을 알 수 있다. 반면, 50세 이후의 연령층에서는 표준편차가 증가하며, 특히 경제활동 연령의 마지막 2년에서는 매우 큰 값을 보이고 있다. 즉, 장년을 넘어 노령 은퇴기로 접어드는 연령에서의 소득수준은 청·장년층의 소득에 비하여 상대적으로 불안정함을 알 수 있다²⁵⁾.

2. 회귀분석에 의한 세대 생애 연령-소득 곡선의 도출

위의 논의는 주로 직관에 의존하는 바, 그 결과로 도출된 연령-소득 곡선을 보다 정교한 분석에 활용하는 데에는 한계가 있을 것으로 생각되어 회귀분석을 시도하였다. 우선 도시가계 자료의 성질상 동일가구를 연도별로 추적하는 것이 불가능하기 때문에 가구단위의 자료가 있음에도 불구하고 미시적인 패널분석 기법은 제외할 수밖에 없었다. 이때 가장 적합한 회귀분석 기법은 시계열 자료 여

25) 이 결과를 지적해주신 박기성 교수님께 감사드린다.

러 개, 또는 횡단면 자료 여러 개의 시계열을 모은 자료의 분석에 사용되는 통합시계열(pooled time series, 또는 time series cross section) 분석기법이 되겠다. 이 경우 통상적으로 회귀식을 아래와 같이 쓸 수 있다.

$$y_{it} = \sum_{k=1}^K X_{itk} \beta_k + u_{it}, \quad i=1, \dots, N; \quad t=1, \dots, T$$

위의 식에서 하첨자 i 와 t 는 각기 횡단면에서의 계층과 시간(연도)을 나타내며, k 는 회귀분석에 쓰이는 독립변수들을 구분하는 하첨자이다. 본 연구에서는 횡단면에서의 계층구분을 가구주의 연령을 기준으로 하므로 하첨자 i 는 25, 26, ..., 65의 값을 가지며, 하첨자 t 는 분석기간이 1982~2002년임을 반영하여 82, 83, ..., 02로 표기한다. X 는 설명변수를 나타내며, β 는 X 의 계수를 나타낸다.

좌변의 종속변수로 본 연구에서는 각 연도 및 가구주 연령별 명목가구소득의 자연대수의 평균을 사용하였다. 평균소득이 아니라 가능하다면 개별 가구의 소득을 종속변수로 사용하는 것이 도시가계 자료에 내재된 정보를 보다 효율적으로 활용하는 방법이기도 하지만 위에서 언급하였듯이 개별 가구를 연도별로 추적하는 것이 불가능하다는 자료의 제약이 있다. 연도 및 연령별 평균소득을 도시가계자료에서 산출하면 21개 연도에서 각기 41개의 횡단면 연령층, 즉 861개의 관측치를 회귀분석에 사용할 수 있다.

설명변수로는 가구소득에 영향을 줄 수 있다고 선형적으로 판단되는 변수를 몇 개 선정하여 사용하였다. 이들 변수는 성질상 종속변수와 마찬가지로 미시자료가 아니라 거시시계열 통계변수를 몇 가지 채택하였으며, 주로 통계청과 한국은행 자료를 사용하였다.

우선 전반적인 소득수준의 변동이 가구소득에 주는 영향을 추출하기 위해서 총 요소소득의 자연대수를 설명변수로 채택하였

다26). 이때 중요소소득은 명목국내총생산(GDP)에 대해 연도의 요소소득비율을 곱하여 구할 수 있으며, 두 변수의 시계열은 한국은행 자료를 사용하였다. 국내총생산(GDP)이 아니라 중요소소득을 사용한 이유는 후자가 노동·자본·경영 등의 생산요소를 제공한 경제주체 중에서 노동을 제공한 대가로 가계에 분배되는 급여와 영업잉여를 나타내기 때문이다. 전자는 도시가계 자료에 포함된 가구들 중 임금근로자의 소득에 해당되며, 이 외에도 본 연구에 자영업

26) 피설명변수가 가구소득입에도 불구하고 설명변수로 일인당소득(중요소소득과 총인구)을 다시 사용하는 것은 일종의 순환논리라는 의문 제기가 가능하다. 이는 물론 인과관계가 상정된 통상적인 계량모형에서는 타당한 지적이다. 예를 들어 특정 연도에서 특정 연령 가구소득이 왜 다른 연령의 가구소득과 다른가, 또는 특정 연도·연령 가구의 소득이 다른 연도의 같은 연령 가구소득과 왜 다른가, 또한 얼마나 다른가를 밝히고자 하는 것이 회귀분석의 목적이라고 하자. 즉 그러한 차이를 설명하는 결정요인으로서 가구주나 가구 구성원의 학력, 결혼 연령, 자녀 관계 등을 설명변수로 쓰는 것이 타당하다면, 본문에서와 같이 일인당소득을 설명변수로 쓰는 것은 소득수준을 설명하기 위해서 소득수준을 설명변수로 사용하는 문제가 있다.

그러나 본문에서 제시한 회귀식의 취지는 이와 다르며, 인과관계야 어떻든 첫째, 각 연도의 연령-소득 곡선의 모양이 다르지 않고 높이만 다르다는 가설을 검증하는 것이며, 둘째, 그러한 검증을 하기 위해서 각 연도의 연령-소득 곡선의 높이의 차이를 어떠한 방법으로 최소화할 것인지를 회귀분석을 통하여 찾고자 하는 것이다. 두 번째 목적의 경우, 각 연도의 연령-소득 곡선의 차이는 우선 대부분 일인당소득의 차이로 설명될 것이라는 사실은 선형적으로 당연한 것이다. 이 경우에 연령-소득 곡선간의 다른 높이를 같은 수준이 되도록 조정함에 있어 일인당소득을 우선 쓰는 것이 가장 효과적임은 당연하며, 일인당소득을 쓰지 않고 다른 변수를 사용하는 것은 오히려 잘못되었다고 할 수 있다.

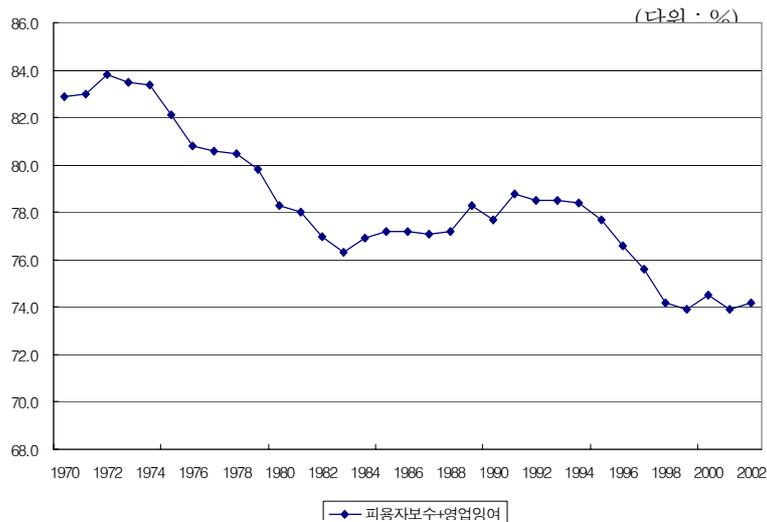
중요소소득과 총인구 등 단 2개의 변수로서 각 연도의 연령-소득 곡선 간의 높이의 차이가 효과적으로 조정되었다는 것은 회귀식의 R^2 가 매우 높다는 사실에서 알 수 있다. 아울러 회귀식의 R^2 가 매우 높다는 결과로부터 연령-소득 곡선을 같은 높이로 조정하였을 때 연령별 소득수준이 연도간에 거의 차이가 없다는 사실, 즉 연령-소득 곡선의 모양이 동일하다는 사실 역시 알 수 있다.

물론 이렇게 도출한 결과를 근거로 제V장에서와 같이 향후 전망을 시도할 때에는, 회귀결과가 지금까지 관측된 피설명변수와 설명변수 간의 매우 안정적인 상관관계를 나타내는 것으로 간주하여 회귀식을 이용한 예측치를 구하는 것이다. 즉 계량분석에서 통상적으로 하듯이 회귀식이 구조적 인과관계를 나타내는 것으로 해석하는 것은 아니다.

IV. 연령별 상대소득분포의 추정 109

자가 포함되고 그 소득은 영업잉여의 일부로 포함된다는 점을 반영한 것이다²⁷⁾. 또한 중요소소득비율이 분석기간에 걸쳐 비교적 안정적인 값을 유지한다면 국내총생산을 직접 사용한 경우와 거의 동일한 결과를 얻을 수 있을 것이다. 그러나 아래 [그림 IV-6]에서 볼 수 있듯이 중요소소득비율은 장기간에 걸쳐 다소 변동을 보이고 있다. 1990년대 후반 이후에는 75% 내외의 비교적 안정적인 추세를 보였고 그 이전에는 1980년대 초부터 1990년대 초 사이에 78% 내외의 추세를 보였다.

[그림 IV-6] 피용자보수와 영업잉여의 합

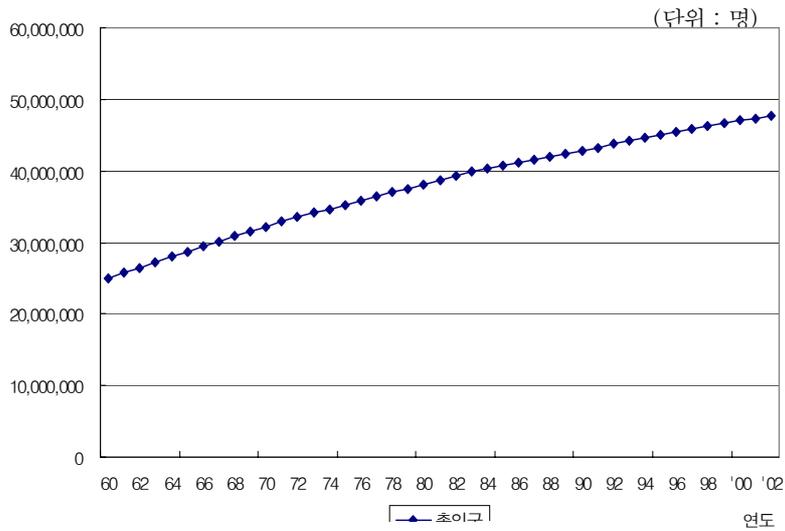


출처: 한국은행, 통계 Database

27) 국내총생산과 중요소소득비율의 시계열에 대하여 각각 자연대수를 취하여 별도의 설명변수로 사용해보기도 하였으나, 본고에서 제시된 결과와 크게 다르지는 않았다. 또한 중요소소득 대신 피용자보수를 사용한 결과도 역시 비슷하였다.

가구소득의 전반적인 수준에 영향을 줄 수 있는 또 하나의 변수로 총인구가 있으며, 회귀식에서는 총인구의 자연대수를 취하여 사용하였다. 즉, 국내총생산의 증가에는 1인당 생산의 증가와 함께 인구증가로 인한 효과가 있기 때문에 1인당 소득의 변화나 가구소득의 변화를 설명하기 위해서는 인구의 변동이 감안되어야 한다. 물론 가구 구성원의 수나 연령분포 등이 변하기 때문에 가구소득의 변화가 반드시 1인당 소득의 변화와 동일한 양상을 보이지는 않을 것이다. 아래 그림에서 볼 수 있는 바와 같이 총인구는 꾸준히 증가하여 왔으나 최근 성장세가 둔화되고 있는 점 이외에는 눈에 띄는 특징을 보이지 않고 있다.

[그림 IV-7] 총인구 추이(1960~2002년)



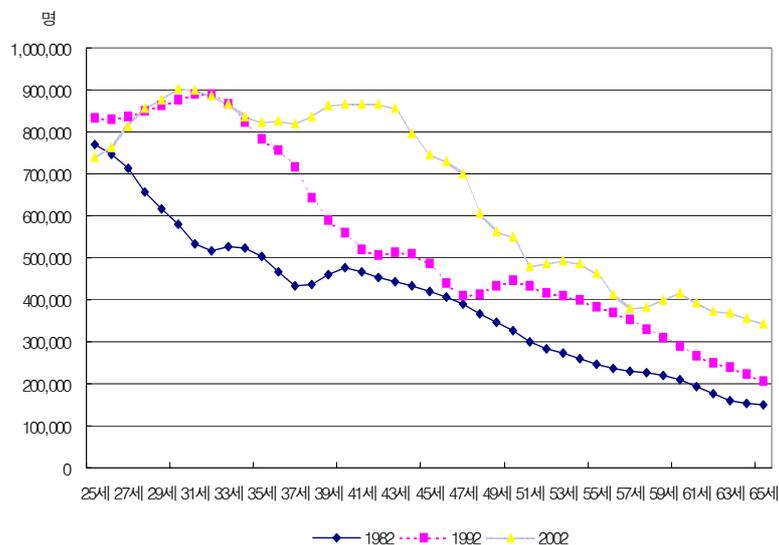
자료: 통계청, 통계 Database

마지막으로, 총인구와는 별도로 총인구 중 각 관측치의 가구주 연령과 동일한 연령의 인구를 설명변수로 채택하였다. 이는 특정 연령층의 인구가 총 인구에 비하여 상대적으로 많거나 적을 경우,

IV. 연령별 상대소득분포의 추정 111

타 연령층에 비하여 노동공급이 상대적으로 많거나 적게 됨으로써 연령층 간 급여의 상대수준에 영향을 줄 수 있을 것으로 상정했기 때문이다. 예를 들어 총인구 중 40세 인구가 상대적으로 많을 경우, 타 연령층에 비하여 노동공급이 크고 상대급여가 다소 낮을 것으로 기대할 수 있을 것이다. 즉, 선형적인 판단으로 이 변수는 상대적으로 영향이 작으면서 음의 값을 나타낼 것이라고 기대할 수 있다. 아래 그림에서 볼 수 있는 바와 같이 연령별 인구분포는 전반적으로 연령이 증가함에 따라 인구가 적어지는 우하향 곡선을 보이고 있으며, 시간의 경과에 따라 우측으로 이동하는 양상을 보이고 있다. 특히 수가 많은 젊은 연령층이 시간의 경과에 따라 30~40대의 경제활동 연령층에 진입하였으며, 노령층에 비하여 인구가 현격하게 많기 때문에 상대적인 연령별 노동공급과 급여에 영향을 줄 수 있는 가능성을 배제하기 어려울 것으로 보인다.

[그림 IV-8] 각세별 연령분포(1982, 1992, 2002년)



자료: 통계청, 통계 Database

이상과 같은 자료에 적용할 수 있는 통합시계열 회귀분석 기법은 여러 가지가 있으며, 오차항 u_{it} 의 속성에 대하여 어떠한 가정을 적용하는가에 따라 달라진다. 우선 오차항간의 자기상관 관계나 이분산(heteroskedasticity)을 상정할 수 있으나, 이 경우 다중공선성(multicollinearity) 때문에 회귀결과 도출이 불가능하거나 R^2 로 나타난 설명력이 낮아서 제외하였다²⁸⁾.

그 이외의 통합시계열분석 기법 중 일반적으로 사용되는 기법은 크게 일방효과모형(one-way effects model)과 양방효과모형(two-way effects model)으로 구분된다. 일방효과모형이란 오차항에 대하여 $u_{it} = v_i + \varepsilon_{it}$, 즉, 연도에 상관없이 횡단면 계층별로 고유의 효과 v_i 가 존재한다는 가정하에 회귀분석을 시도하는 방법이며, 이때 ε_{it} 에 대해서는 평균 0, 동분산 등 일반적인 가정을 한다. 양방효과모형은 일방효과모형과 유사하며, 다만 횡단면의 계층별 효과 v_i 에 추가하여 연도별 고유효과 e_t 가 존재한다는 가정하에 $u_{it} = v_i + e_t + \varepsilon_{it}$ 로 오차항의 속성을 정의한다.

일방 또는 양방효과모형은 v_i 와 e_t 의 속성을 i 나 t 에 대해서만 값을 달리하는 상수로 간주하는지, 아니면 확률변수(random variable)로

28) Parks(1967)의 기법은 오차항 u_{it} 에 대하여 $E(u_{it}^2) = \sigma_{it}^2$ (이분산), $E(u_{it}u_{jt}) = \sigma_{ij}$ (동시 상관), $u_{it} = \rho u_{i,t-1} + \varepsilon_{it}$ (자기상관)의 속성들을 상정한다. 그러나 동 기법에 의한 회귀분석은 다중공선성의 문제 때문에 결과도출이 되지 않았다.

Parks와는 달리 Da Silva(1975)는 $u_{it} = a_i + b_t + e_{it}$, $e_{it} = \alpha_0 \varepsilon_{it} + \alpha_1 \varepsilon_{i,t-1} + \dots + \alpha_n \varepsilon_{i,t-n}$ 이라는 이동평균 혼합 분산 오차항을 상정한다. 횡단면과 연도의 효과를 감안한 변수 a_i 와 b_t 를 채택하였으므로, 이는 본문에서 언급한 일방 또는 양방효과모형(one-way or two-way effects model)의 특수한 예로 간주할 수 있고, 다만 e_{it} 항이 연도 t 에 대해 서로 독립적이지 않고 이동평균 구조를 갖고 있다는 것이 특징이다. 동 기법을 적용하여 보았을 때 통계적 유의성을 확보한 결과를 얻을 수 있으나, 본문에서 제시한 일방고정효과(one-way fixed effects) 모형에 비하여 R^2 의 설명력에서 현격한 차이를 보였으며, 회귀분석 결과 역시 일방고정효과모형이 상대적으로 직관적인 해석이 용이하다는 장점이 있었다.

상정하는지에 따라 다시 고정효과모형(fixed effects model)과 확률 변수효과모형(random effects model)으로 구분되며, 각각의 경우에 따라 회귀분석 기법이 달라진다. 본 연구에서는 일방 및 양방 확률 변수효과모형(one-way and two-way random effects models)을 이용한 회귀분석을 우선 시도하였다. 그러나 설명변수에 대한 계수의 유의성이나 전체 회귀식의 R²값이 비교적 높은 수준으로 나타났음에도 불구하고, 확률변수모형 적용의 적합성을 판별하는 하우스만 검정(Hausman test)에서는 일반과 양방모형 모두 확률변수모형 적용이 적합하다는 귀무가설이 10%의 유의성 수준에서 기각되었다.

반면에 고정효과모형에서는 일방효과모형의 회귀분석 결과가 양호하여 아래에서 보다 상세히 살펴보기로 한다. 그러나 양방고정효과모형은 다중공선성의 문제 때문에 결과를 그대로 받아들이기에 무리가 있었다. 보다 구체적으로는, 종속변수인 연령·연도별 평균 자연대수 가구소득의 변화를 설명함에 있어서 연도별 고정효과의 영향을 추정한 후, 이에 추가하여 총요소소득의 변화가 설명할 수 있는 부분이 없었다²⁹⁾.

이상의 논의를 종합하자면, 통합시계열분석의 여러 가지 기법 중에 본 연구의 자료에 적용이 가능한 것은 일방고정효과모형이었으며, 동 모형의 회귀분석 결과가 해석도 가장 용이하였다. 일방고정효과모형은 결국 횡단면 계층별 고유효과를 더미변수로 추정하는 방법과 같으며, 앞의 회귀식을 다시 정리하여 아래와 같이 쓸 수 있다.

$$\begin{aligned} \text{평균자연대수가구소득}_{it} \\ &= a + \sum_{i=25}^{64} d_i \beta_i + \beta_{k1} \ln(\text{총요소소득}_t) + \beta_{k2} \ln(\text{총인구}_t) + \beta_{k3} \ln(i\text{세 총인구}_t) + e_{it}, \\ & \quad i=25, \dots, 65; \quad t=82, \dots, 02, \\ & \quad d_i=1 \text{ if age} = i \text{ (0 otherwise),} \\ & \quad e_{it} \sim i.i.d. \end{aligned}$$

29) 본 연구의 회귀분석에서 나타난 다중공선성의 문제는 설명변수인 총요소소득과 총인구가 연도별로는 변화를 보이나(variation across time), 각 연도별 횡단면 연령계층별로는 같은 값을(no variation in cross-section age groups) 사용하였기 때문이다. 이와 대조적으로, 설명변수로 사용된 연도·연령별 총인구는 우리가 시도한 여러 가지 통합회귀분석기법에서 항상 유의성을 나타내었다.

위의 식에서 d_{65} 는 제외더미(omitted dummy)로 처리되었으나, 회귀식에 상수항이 있기 때문에 추정된 α 의 값을 회귀식에서 제외된 β_{65} 로 해석할 수 있다. 이런 경우에 $\beta_i(i=25, \dots, 64)$ 는 제외된 β_{65} 와의 차이만을 추정하기 때문에 i 세 연령층에 해당되는 횡단면 고정효과의 크기는 $\alpha(=\beta_{65})+\beta_i(i=25, \dots, 64)$ 가 된다³⁰⁾. 위의 식에 따라 수행한 회귀분석의 결과는 아래 <표 IV-2>에 나와 있다³¹⁾.

<표 IV-2> 일방고정모형을 사용한 연령별 자연대수 소득의 회귀분석

횡단면 연령별 고정효과							
연령별 더미변수	추정계수	연령별 더미변수	추정계수	연령별 더미변수	추정계수	연령별 더미변수	추정계수
25세	0.322580	35세	0.569176	45세	0.647975	55세	0.619202
26세	0.376447	36세	0.574493	46세	0.656290	56세	0.606518
27세	0.415821	37세	0.580380	47세	0.663284	57세	0.581147
28세	0.443930	38세	0.593138	48세	0.669019	58세	0.525978
29세	0.475117	39세	0.604997	49세	0.672112	59세	0.479799
30세	0.483493	40세	0.602500	50세	0.664218	60세	0.428547
31세	0.500638	41세	0.618856	51세	0.663873	61세	0.354444
32세	0.515903	42세	0.627979	52세	0.661629	62세	0.249659
33세	0.534187	43세	0.640435	53세	0.663311	63세	0.156163
34세	0.549982	44세	0.650738	54세	0.650035	64세	0.071197
기타 설명변수							
상수항	102.111021	총요소 소득의 자연대수	1.438366	총인구의 자연대수	-6.208852	각연령층 총인구의 자연대수	-0.071573

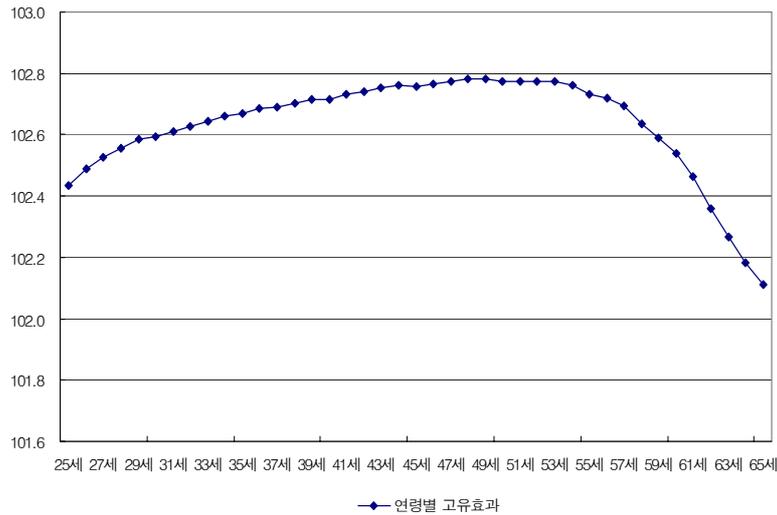
- 주: 1. 종속변수는 도시가계자료의 각 연도·가구주 연령별 자연대수가구소득의 평균
 $N = 861, R^2 = 0.9896$
 2. $\beta_i(i=25, \dots, 64)$, 즉 횡단면 계층별 고정효과가 없다는 귀무가설은 유의수준 0.0000 이하에서 기각($F = 88.8964$ (자유도 40, 817))
 3. 고정효과와 기타 설명변수의 추정계수는 모두 0.0001 이하의 유의수준을 보였으나, 다만 β_{64} 와 $\ln(\text{연령별 총인구})$ 만 계수의 유의수준이 각각 0.0029와 0.0223으로 추정

30) 실제로 회귀식에서 상수항을 제외하고 $\beta_i(i=25, \dots, 65)$ 를 추정하면 이와 동일한 결과를 얻는다.
 31) 회귀분석 결과에서 R^2 의 값이 매우 높으므로 단위근 검정(unit root test)의 필요성이 제기될 수도 있으나, 본 연구에서는 이를 시도하지 않았다. 본문의 회귀식은 인과관계 분석에 초점을 두었다기보다 횡단면 연령-소득 곡선의 정확한 모양을 계량적으로 추출하고, 그 모양이 사실상 불변이었음을 입증하는 데 목적이 있기 때문이다. 본문의 회귀식은 우리가 보이고자 하는 바를 입증하는 데 적절하다고 판단하였다. 달리 이야기하면, 회귀식에서 40여개의 많은 더미변수가 사용되는 바, 이들 더미변수들이 종속변수의 변이(variation)를 대부분 포착한다고 해석할 수 있다.

회귀식의 결과를 보면 종속변수인 연령별 평균자연대수가구소득을 결정함에 있어 연령층간의 횡단면 상대소득분포, 즉 횡단면 연령·소득 곡선의 모양을 결정하는 변수들이 있고, 이와 달리 연령·소득 곡선의 모양에는 영향을 주지 않으나 연령·소득 곡선의 전반적인 수준을 결정하는 요인들이 있다. 우선 횡단면 연령·소득 곡선의 모양은 주로 상수항과 연령 더미변수의 추정계수 값으로 결정된다.

아래 [그림 IV-9]에는 상수항과 더미변수의 값을 정리하여 제시하였으며, [그림 IV-5]에서 다른 방법으로 도출했던 연령별 상대소득분포 곡선과 동일한 모양임을 쉽게 알 수 있다. 그림에서 볼 수 있듯이, 횡단면 연령-소득 곡선의 특징은 가구주의 연령이 40대 중반에서 50대 중반에 이르는 구간에서 가구소득이 가장 높으며, 55세를 넘어서면서 가구소득이 급격히 하락한다는 점이다. 또한 앞의 <표 IV-1>과 [그림 IV-5]에서는 최고소득 연령이 53세였으나, 회귀분석의 결과에서는 가구주의 연령이 49세일 때 가구소득이 가장 높은 것으로 분석되었다. 그러나 앞에서 행한 분석의 결과에는 아래에서 논의할 총인구의 연령별 분포의 영향이 포함되었음을 상기하면, 실제로 두 방법으로 추정된 횡단면 연령별 상대소득분포는 동일하다고 판단된다.

[그림 IV-9] 회귀분석으로 추정된 횡단면 연령별 상대소득 분포



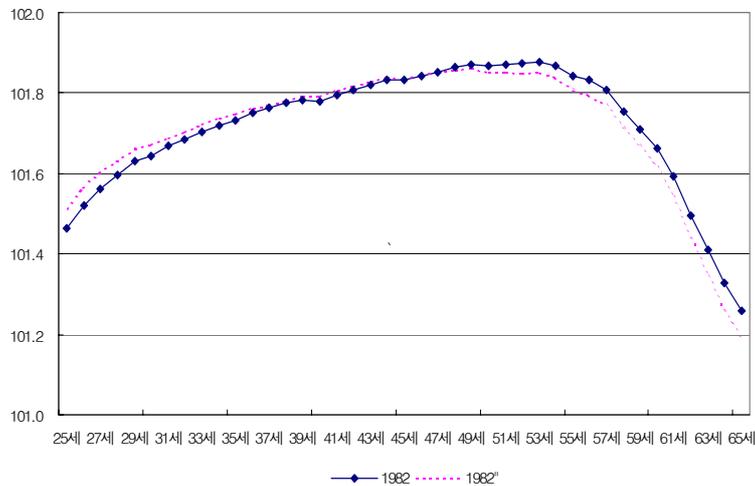
회귀분석 결과에서 상수항과 터미변수 이외에 횡단면에서의 상대소득 분포에 영향을 줄 수 있는 변수로는 각 연령의 총인구가 있다. 각 연령별 총인구는 두 가지 효과를 나타내는데, 앞의 [그림 IV-8]에서 보았듯이 1982년과 2002년 사이에 전반적인 인구증가로 인하여 횡단면 연령-소득 곡선을 하향 수직이동시키는 전반적인 효과가 있으며, 연령간 인구분포의 변화로 인하여 연령-소득 곡선의 모양에도 영향을 줄 수 있다. 아래 [그림 IV-10]과 [그림 IV-11]을 비교하면 알 수 있듯이 연령-소득 곡선의 높이에 거의 영향을 주지 않는 것으로 나타나 있으며, 동 곡선의 모양에도 미세한 조정만 나타나는 결과를 얻었다. 이는 연령별 총인구에 대한 추정계수가 매우 작은 음의 값(-0.0716)으로 추정된 결과에 따른 것이다. [그림 IV-10]과 [그림 IV-11]에서 가는 점선은 위의 [그림 IV-9]와 같은 모양을 하도록 각 연령층의 실제인구 대신 당해 연도의 연령별 인구의 평균치를 대입하여 도출하였다. 점선과 실선을 비교하면 연령

IV. 연령별 상대소득분포의 추정 117

별 횡단면 인구분포의 영향은, 앞에서 언급한 바와 같이 횡단면 최고소득 연령이 [그림 IV-9]의 49세에서 [그림 IV-5]와 동일하게 53세 부근으로 늦춰지는 것으로 나타남을 알 수 있다.

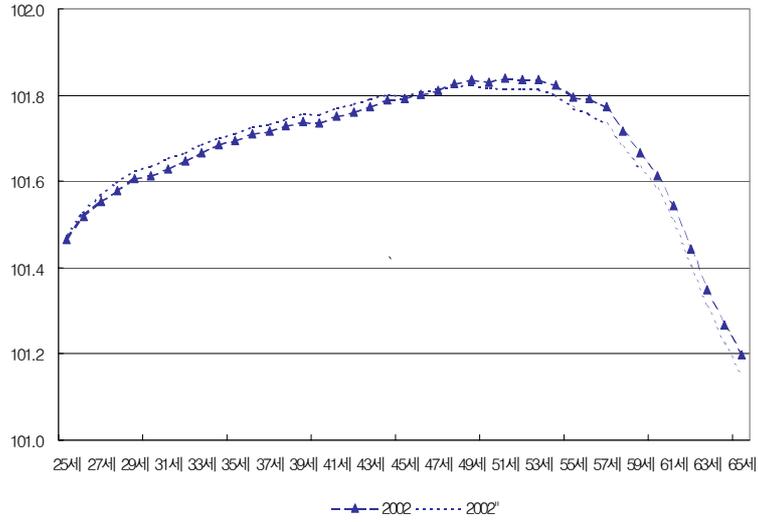
끝으로, 횡단면 연령-소득 곡선의 높이를 결정하는 변수로 총요소소득과 총인구가 있으며, 예상대로 $\ln(\text{총요소소득})$ 의 계수는 +1보다 다소 큰 값을 얻어서, 총요소소득의 변화가 가구소득 수준의 변화와 비슷한 비율로 일어나고 있음을 알 수 있다. 반면, $\ln(\text{총인구})$ 의 계수는 -6.21로 추정되어 예상대로 음의 값으로서 가구소득을 줄이는 효과를 보이고 있다. 그러나 그 절대값이 1과는 현격한 차이를 보이고 있다. 이러한 결과 중 일부는 총인구가 증가하면서 가구의 규모가 작아짐으로써 나타나는 효과라고 짐작되나, 정확한 경위는 본고에서 시도하는 것보다 세밀한 분석이 필요하다고 하겠다.

[그림 IV-10] 횡단면 연령층간 상대소득 분포에 대한 연령별 총인구의 영향(1982년)



주: 그림에서 가는 점선은 각 연령층의 실제인구 대신 당해 연도의 연령별 인구의 평균치를 대입하여 도출하였다. 즉, 연령별로 동일한 평균치를 사용하여 연령간 인구의 차이를 없앴으로써 위의 [그림 IV-9]와 같은 모양으로 돌아가면서도, 동시에 점선의 전반적인 높이가 연령별 인구의 효과가 반영된 실선과 같은 수준이 되도록 조정한 것이다.

[그림 IV-11] 횡단면 연령층간 상대소득 분포에 대한 연령별 총인구의 영향(2002년)



주: 그림에서 가는 점선은 각 연령층의 실제인구 대신 당해 연도의 연령별 인구의 평균치를 대입하여 도출하였다. 즉, 연령별로 동일한 평균치를 사용하여 연령간 인구의 차이를 없앴으로써 위의 [그림 IV-9]와 같은 모양으로 돌아가면서도, 동시에 점선의 전반적인 높이가 연령별 인구의 효과가 반영된 실선과 같은 수준이 되도록 조정한 것이다.

V. 세대별 생애 연령-소득 곡선의 도출

1. 연도별 횡단면 연령-소득 곡선의 전망 도출

본 장에서는 앞에서 얻은 회귀식을 이용하여 향후 횡단면 연령-소득 곡선을 전망하고, 다시 이를 이용하여 세대별 생애 연령-소득 곡선을 도출하고자 한다. 이러한 전망을 하기 위하여 우선 각 설명변수에 대한 향후 전망을 가정하여 회귀식에 대입하여야 한다. 그런데 본고에서 채택한 설명변수 전망은 세대 연령-소득 곡선 도출의 예시를 위하여 임의로 채택된 값을 사용하는 것이며, 실제 전망치로서 제시하는 것이 아님을 우선 밝히고자 한다.

우선 중요소소득의 향후 추이에 대한 가정이 필요하며, 이는 다시 중요소소득비율과 국내총생산의 향후 추이에 대한 가정을 필요로 한다. 중요소소득비율은 앞의 [그림 IV-6]에서 보았듯이 1990년대 이래 다소의 등락은 있지만 비교적 안정적인 추세를 보이고 있으므로 2002년도의 74.2%가 계속 유지된다고 가정하였다. 그러나 이 값은 향후 장기적으로 변할 수도 있다는 가능성이 배제될 수는 없음을 염두에 둘 필요는 있다.

국내총생산의 명목성장률은 우선 임의로 5% 수준을 향후 유지하는 것으로 설정하였다³²⁾. 앞의 회귀식에서 $\ln(\text{중요소소득} = \text{국내총}$

32) 국내총생산의 성장률을 일정 수준인 5%로 가정하는 것은 인구 증가율이 마이너스가 되는 2020년 초반 이후에 1인당 국내총생산의 증가율이 5% 이상이 되며 점차 증가한다고 암묵적으로 가정하는 것과 같다. 따라서 본문의 가정 대신 1인당 국내총생산이 일정 수준으로 증가한다는 가정을 대신 채택해보았으나, 결과에 사실상 영향이 없으므로 별도로 제시하지 않았다. 참고로, [그림 V-2]에서 볼 수 있듯이 인구

생산·요소소득비율)의 계수가 1.44로 추정되었으므로 5%의 명목 국내총생산 성장률은 자연대수 연령-소득 곡선을 매년 $0.072(=0.05 \cdot 1.44)$ 정도의 폭으로 상승시키는 효과가 있다.

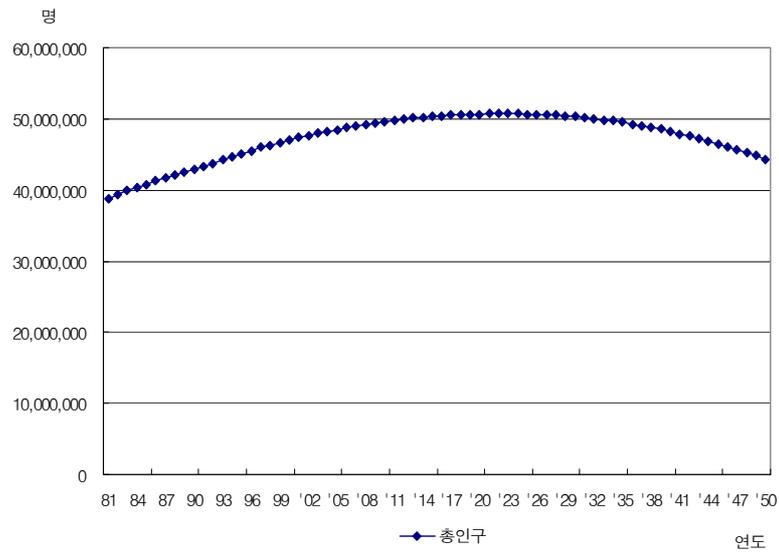
총인구 및 연령별 인구의 변화는 통계청의 장래인구 추계를 그대로 대입하였다. 총인구의 증가는 가구소득을 상당히 감소시키는 효과가 있음을 앞서서도 보았으나, 아래 [그림 V-1]에서 볼 수 있는 바와 같이 2022년 이후에 총인구가 감소할 것으로 예상되므로, 2022년 이후에는 추정 가구소득을 인상시키는 효과가 있을 것으로 보인다. 다만 회귀분석에서 추정된 총인구 변화의 영향은 총인구가 계속 증가하는 시기의 자료에 근거하는 것이므로, 총인구 변화가 감소세로 돌아선 이후에도 설명변수로서의 총인구의 영향의 크기와 방향이 예전과 같지 않을 수도 있다는 점을 부인할 수는 없다.

위의 논의를 종합하면, 2022년 이후에는 국내총생산(총요소소득)이나 총인구가 공히 연령-소득 곡선을 높이는 효과가 있다. 2022년 이전에는 두 요인의 효과가 반대로 작용하나, 총인구 성장률이 [그림 V-2]에서 볼 수 있듯이 0.7% 이하의 낮은 수준이므로 $\ln(\text{총인구})$ 의 추정계수가 비록 상대적으로 큰 음수이기는 하나, 총인구 증가의 효과가 대략 $-0.043(= -6.209 \cdot 0.007)$ 보다 작을 것이다. 따라서 명목국내총생산 성장률이 2%대 초반까지 하락하지 않는 한 국내총생산 증가의 영향을 상쇄하지는 못할 것으로 분석된다.

감소율이 가장 큰 2050년에도 -1%밖에 안되므로, 2000년 이후 2050년 이전의 기간 대부분에서 국내총생산과 1인당 국내총생산의 증가율의 차이는 미미할 수밖에 없다.

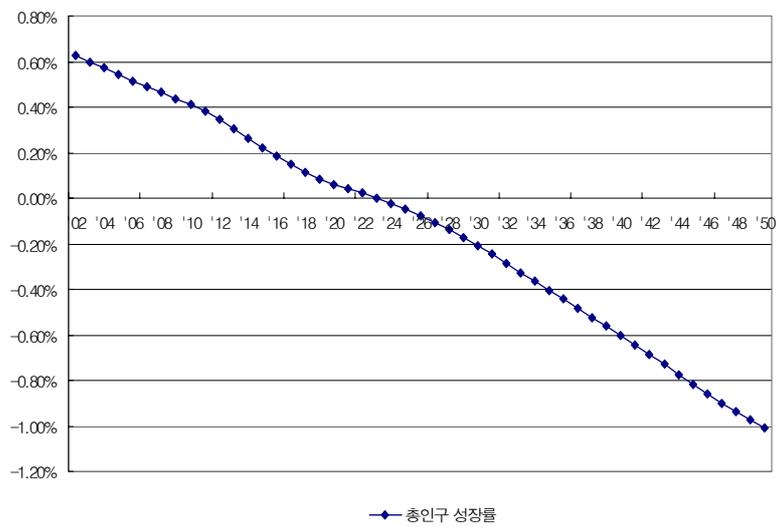
V. 세대별 생애 연령-소득 곡선의 도출 121

[그림 V-1] 총인구 추이(1981~2050년)



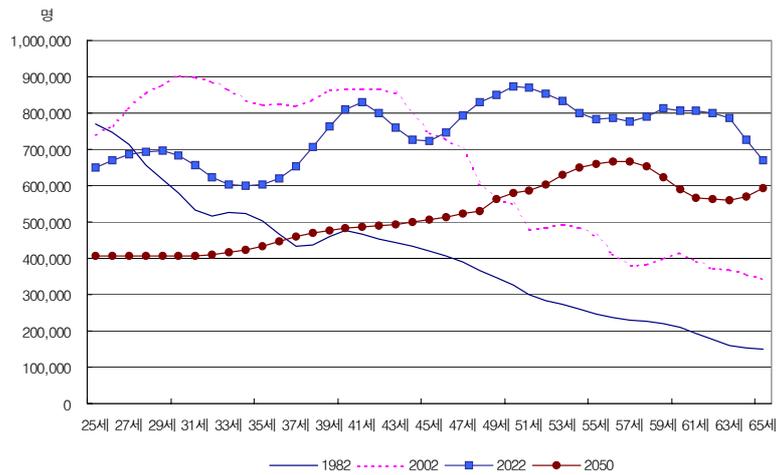
자료: 통계청, 통계 Database

[그림 V-2] 총인구 성장률(2002~2050년)



끝으로, 각 연령별 총인구의 영향을 감안하여야 한다. [그림 V-3]에서 볼 수 있듯이 1982년에서 2002년에 걸친 기간에는 연령별 인구분포가 우하향 곡선의 모양을 하고 있으나, 장래에는 인구구조의 고령화가 진전되면서 우상향 곡선으로 전향될 것으로 보인다. 따라서 앞의 회귀식에서 연령별 인구의 추정계수가 음으로 나왔으므로, 상대적으로 가구주 연령이 젊은 계층의 가구소득이 증가하고 장년 이후 계층의 소득은 감소할 것으로 예상된다. 다만 연령별 인구의 추정계수의 절대값이 작으므로, [그림 V-4]~[그림 V-6]에서 볼 수 있듯이 그 영향은 비교적 미미하다. [그림 V-4]를 보면 40대 중반 이전 가구주의 가구소득이 상대적으로 낮으며 50대 이후 계층은 소득이 증가된 효과가 있다. [그림 V-5]는 인구 성장이 감소세로 돌아서는 시기인 2020년에 연령별 인구가 연령간 상대 소득 분포에 미치는 영향이 거의 중립적임을 보인다. [그림 V-6]에서는 2050년에 이르러 상황이 2002년에 비해 반전되어 젊은 연령층 가구주의 가구소득이 높고 고령층의 가구소득이 낮아짐을 볼 수 있다.

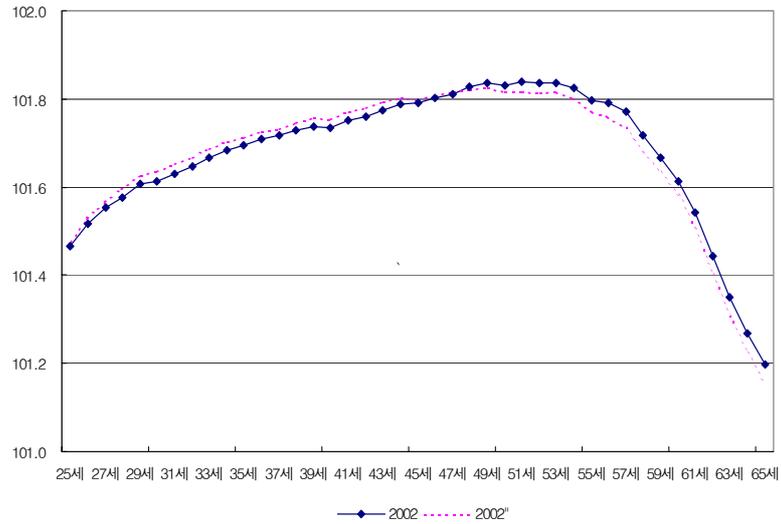
[그림 V-3] 각세별 연령분포(1982, 2002, 2022, 2050년)



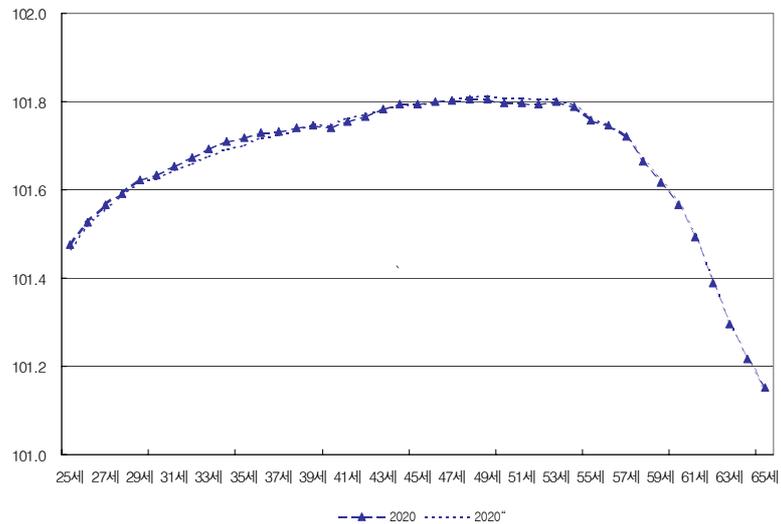
자료 : 통계청, 통계 Database

V. 세대별 생애 연령-소득 곡선의 도출 123

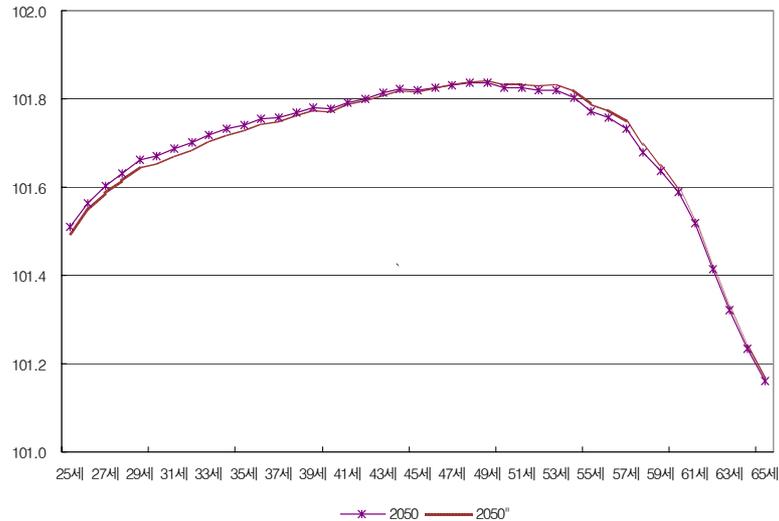
[그림 V-4] 횡단면 연령층간 상대소득 분포에 대한 연령별 총인구의 영향(2002년)



[그림 V-5] 횡단면 연령층간 상대소득 분포에 대한 연령별 총인구의 영향(2020년)



[그림 V-6] 횡단면 연령층간 상대소득 분포에 대한 연령별 총인구의 영향(2050년)



2. 세대별 생애 연령-소득 곡선의 도출

앞의 [그림 V-4]~[그림 V-6]의 곡선 3개는 해당 연도의 연령간 상대소득분포를 나타내고 있으며, 이에 총요소소득과 총인구 변화의 영향을 더하면 각 연도의 횡단면 연령-소득 곡선을 구할 수 있음을 설명하였다. 그러나 이렇게 구한 연령-소득 곡선은 [그림 V-4]~[그림 V-6]에서 제시한 곡선과 높이만 다를 뿐 모양이 동일하고, 향후 총요소소득과 총인구의 변화가 주는 영향 역시 이미 설명하였기 때문에 별도로 향후의 횡단면 연령-소득 곡선을 제시하지 않았다.

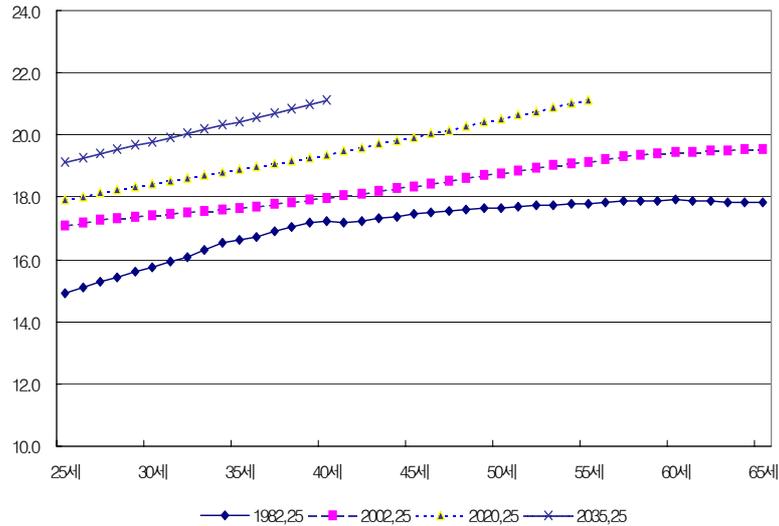
앞 절에서 설명한 방법으로 2002~2050년까지의 각 연도별 횡단면 연령-소득 곡선을 구하면, 앞 장의 [그림 IV-2]에서 예시한 방법을 적용하여 특정 자연인, 또는 특정 연령의 자연인의 집단인 세대(cohort)의 생애에 걸친 소득의 궤적을 구할 수 있다. 본고에서는

이를 세대 생애 연령-소득 곡선이라 지칭하기로 한다. 보다 구체적인 표기를 사용하자면, t 연도에 가구주 연령이 i 인 가구의 평균 자 연대수 소득을 y_{it} 라고 표기하였을 경우, t 연도의 횡단면 연령-소득 곡선은 $\{y_{it}\}_{i=25, \dots, 65}$ 라고 수열형태로 표시할 수 있다. 반면, t_0 연도에 출생한 세대의 생애 연령-소득 곡선은 $\{y_{it}\}_{i=25, \dots, 65; t=t_0+i}$ 라고 표기할 수 있다.

일단 각 연도별 횡단면 연령-소득 곡선을 구하면, 이로부터 여러 세대에 대하여 생애 연령-소득 곡선을 간단하게 구할 수 있다. 아래의 [그림 V-7]에는 향후 경상성장률을 5%로 가정하였을 때의 4개 세대별 생애 연령-소득 곡선이 제시되어 있다. 그림에서는 아래로부터 각각 1982, 2002, 2020, 2035년 현재 25세인 4개 세대의 생애 연령-소득 곡선이 제시되어 있으며, 제일 위쪽의 곡선 2개가 65세 이전에 단절된 것은 통계청의 장래인구 추계가 2050년까지만 제시되어 있어서 2050년 이후의 연령별 소득을 전망하지 못하였기 때문이다.

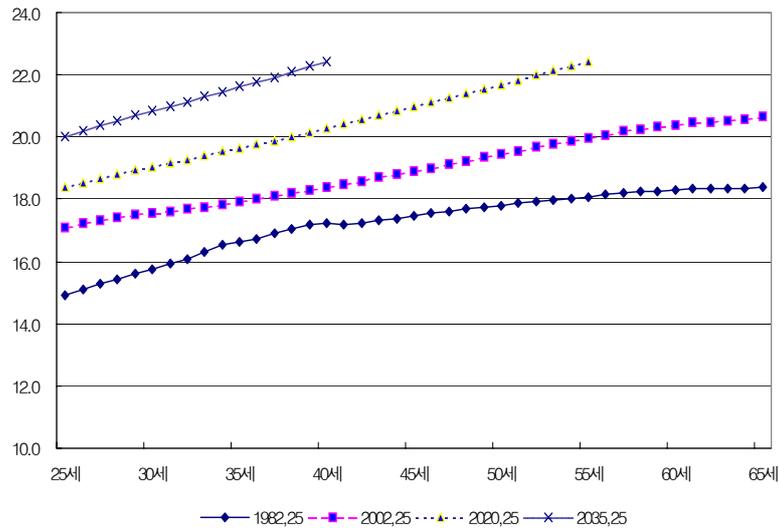
이와 같이 각 연도별 횡단면 연령-소득 곡선으로부터 도출된 세대 생애 연령-소득 곡선의 특징은 매년 횡단면 연령-소득 곡선이 상승함에 따라 횡단면 연령-소득 곡선에 비하여 최고소득 연령이 증가한다는 점이다. [그림 V-7]에서는 제일 아래의 곡선(1982년에 25세인 세대)을 보면 경제의 경상성장률이 앞으로 5%를 유지할 경우 전 생애 중 최고소득을 받는 연령이 60세가 된다. 이는 횡단면에서의 최고소득 연령인 50세 부근보다 약 10년 정도 늦어지는 결과이다.

[그림 V-7] 세대별 생애 연령-자연대수소득 곡선(명목성장-5% 가정)



이와 같이 생애 중 최고소득을 받는 연령이 횡단면에 비하여 늦어지는 현상은 물론 경상성장률이 높을수록 심해진다. [그림 V-8]에서는 경상성장률이 2002년 이후 7%를 유지한다는 가정하에 다시 4개 세대의 생애 연령-소득 곡선을 구하여 제시하였다. [그림 V-5]와 비교하면, 1982년에 25세인 세대의 생애 최고소득 연령이 61세로 1년 늦춰지는 효과가 있다. 또한, 2002년에 25세인 세대의 경우 5% 경상성장의 가정하에서는 역시 60세 부근 이후에 소득이 다소 감소하나, 7% 성장을 가정하면 은퇴연령인 65세까지 소득수준이 미미하나마 계속 증가하게 된다는 사실을 볼 수 있다.

[그림 V-8] 세대별 생애 연령-자연대수소득 곡선(명목성장 7% 가정)



3. 개인소득과 대비한 가구소득의 특성

다음 장의 자연대수 소득의 분산에 대한 논의로 넘어가기에 앞서, 본고에서 지금껏 밝힌 가구소득 분포의 특성에 대하여 개인소득 또는 개인의 임금과 대비하여 몇 가지 사항을 간략히 논하고자 한다. 우리는 제IV장에서 가구소득의 횡단면 연령-소득 곡선의 모양이 21년간 매우 안정적이었으며 이러한 안정성을 이용해 제V장에서는 세대별 생애 연령-소득 곡선을 도출할 수 있음을 보았다. 이때 가구소득의 횡단면 연령-소득 곡선(age-income profile)의 특징은 최고소득 연령이 50세 부근이었다. 이는 일반적으로 알려진 바와 같이 개인 근로자의 연령-임금 곡선(age-earnings profile)이 40대 초반에서 최고소득을 보이는 현상과 대조된다. 그러나 비록 가구의 주소득원인 가구주의 소득이 개인 연령-임금 곡선에 따라 40대 초반에 하락하기 시작한다고 해도, 가구주의 연령이 40대 초반에 이

르면 2세인 자녀가 경제활동 참여를 통해 소득을 창출하기 시작함으로써 가구주 소득의 하락을 당분간 상쇄할 수 있을 것이다. 이러한 요인들을 감안하였을 때, 가구의 연령-소득 곡선이 개인의 연령-소득 또는 연령-임금 곡선보다 낮은 연령에서 최고소득을 보인다는 사실은 오히려 당연한 결과라고 하겠다.

그런데 개인 연령-임금 곡선과 대비한 가구 연령-소득 곡선의 모양의 차이보다도 경제현상을 연구하는 저자들에게 더 흥미로운 사실은 가구 연령-곡선의 모양의 안정성이라는 특성이다. 즉, 21년이라는 장기간에 걸친 관측에도 불구하고 가구소득의 횡단면 연령-소득 구조가 변하지 않고 매우 안정적이라는 사실이다. 비록 본 연구에서 자세한 비교를 시도하지는 않았으나, 가구 소득분포는 개인을 관측단위로 한 소득분포보다 상당히 안정적인 특성이 있다고 판단된다.

또한 안정성을 보다 넓게 해석하면 제Ⅲ장에서 확인한 가구소득이 자연대수 정규분포라는 성질도 포함하여 생각할 수 있다. 1982년 이래 2002년까지 우리 경제가 경기변동, 소득수준, 산업구성, 연령구조, 제반 사회제도 등에서 많은 변화를 겪었음에도 불구하고 그 과정에서 가구소득의 분포는 총인구에서나 연령별 계층에서도 항상 자연대수 정규분포를 유지하였다. 거의 유일한 예외는 경제체제가 급격히 붕괴된 1997년과 1998년의 외환위기 때이며, 1998년의 소득분포가 크게 타격을 입어 매우 심한 좌편향성을 보였음에도 불구하고 그 다음 해인 1999년에는 바로 자연대수 정규분포로 복귀하였다는 사실은 자연대수 정규분포라는 특성이 매우 강하다는 점을 시사한다.

‘가구의 소득이 왜 자연대수 정규분포를 따르는가’하는 질문은 사람의 키가 왜 정규분포인가 묻는 것과 같이 답이 없는 질문이다. 본 연구에서도 이 질문에 대한 답을 구하기보다는 자연대수 정규분포라는 성질이 확인된 이상 그 성질을 이용하여 다른 유용한 결과

를 도출하는 작업에 전적으로 치중하였다. 그러나 여기서 잠시 지면을 할애하여, 비록 이론적 논리나 실증적 통계로부터 자연대수 정규분포를 도출할 수는 없어도, 가구소득의 분포가 매우 안정적이며 자연대수 정규분포를 따르는 이유에 대해 경제학적 통찰(insight)을 한 가지 음미해보는 것도 유용할 것이라고 사료된다.

저자들의 견해로는 가구소득 분포의 안정성을 이해하려면 기본 관측단위로 사용한 가구의 개념에서 답을 찾아야 할 것이다. 즉, 가구란 많은 의사결정이 이루어지는 내생적 조정 기구(endogenous adjustment mechanism)로 이해할 수 있다. 특히 가구 내에서 일어나는 내생적 의사결정도 많겠으나 가구형성 자체가 여러 경제·사회 변수들의 영향을 받아 내생적으로 결정된다는 사실은 매우 중요하다. 간단한 예를 들면, 어떠한 가구의 소득이 크게 증가하였을 때 그 가구의 일부 구성원들이 분가하여 새로운 가구를 형성할 수 있는 가능성이 커질 것이다. 만일 분가가 이루어진다면, 가구의 수가 1에서 2로 달라짐은 물론, 가구 규모, 가구 소득, 가구주 연령 등 가구유형의 구조에서 거의 모든 변수가 바로 영향을 받게 된다. 반대로 소득수준이 급감한 2개의 가구가 하나로 합치는 결과 역시 현실에서 흔히 관측되는 현상이며, 가구소득과 가구유형의 분포에 직접적인 영향이 있다.

내생적 조정이 이루어질 수 있는 여지가 많을수록 관측된 결과가 안정적이라는 사실은 경제학이 시사하는 가장 기본적인 원리의 하나이다. 예를 들어 수요와 공급곡선을 각기 따로 보면 큰 변화가 일어날 수 있는 여지가 많을 수 있다. 그러나 수요와 공급이 만나서 이루어지는 균형이라는 결과를 보면 그보다 상당히 안정적이다. 또 다른 비유를 하자면, 특정 산업부문이나 재화에 대해 부분균형(partial equilibrium)적 분석을 할 때 큰 변화가 예측되는 상황에서도, 시야를 체제 전체로 넓혀 일반균형(general equilibrium)하에서 종국적으로 나타나는 결과는 타 부문이나 재화와의 상호조정을 거

치면서 상당히 안정적인 것으로 나타나게 마련이다. 이는 왈라스(Walrasian) 경제체제에서 모든 것이 종국적으로는 상호대체적(substitutable)이라는 미시경제학의 기본원리를 나타내는 것으로 생각할 수 있다.

물론 개인에 귀속되는 의사결정과 경제변수들 역시 체제 내에서 이루어지는 내생적 결과들이다. 그러나 위에서 언급하였듯이 가구라는 단위에서는 개인의 경우에 추가하여 보다 많은 내생적 조정이 이루어질 수 있는 여지가 있다. 이렇게 볼 때, 가구를 단위로 하여 관측되는 소득분포가 개인의 소득분포에 비하여 안정적 특성이 있다는 사실은 그리 놀라운 것이라고 할 수 없다. 가구소득이 개인소득에 비하여 사실 그 의미가 모호하다는 단점에도 불구하고 본 연구에서 가구소득을 기본단위로 설정하여 사용한 것은 바로 개인소득에서는 찾기가 쉽지 않은, 안정적인 규칙성을 찾아 장기예측에 이용하려는 의도가 반영된 결과임을 밝힌다.

VI. 지니표면을 이용한 자연대수 소득의 분산 추정

1. 지니표면의 도출

제 I 장의 서론에서 밝혔듯이, 자연대수 소득의 평균의 경우에는 매우 안정적이고 규칙적인 특성을 쉽게 발견하여 미래 전망에 활용할 수 있음에 반하여, 자연대수 소득의 분산의 경우에는 전망에 활용할 수 있는 특징을 발견하기 어려웠다. 따라서 본 연구에서는 지니계수의 향후 추이에 대하여 몇 가지 시나리오를 제시하고 이로부터 역산으로 자연대수 소득의 분산을 간접추정하고자 한다. 이를 위하여 우선 본 절에서는 자연대수 소득의 분산과 지니계수 간에 1:1 대응관계가 있음을 보이고, 분산의 역산을 위하여 지니계수와의 관계를 계량화하여 제시하고자 한다.

어떠한 분포가 정규분포나 자연대수 정규분포임을 밝혔을 때 얻는 가장 유용한 결과 중 하나는 평균과 분산(또는 표준편차) 단 2개의 계수만으로 분포 전체를 알 수 있다는 사실이다. 따라서 가구소득이 자연대수 정규분포를 따른다는 것은 이미 언급하였듯이 단 2개의 계수만으로 소득분포를 알 수 있음을 의미한다. 그런데 소득분포 전체를 알면 소득분배의 형평성을 측정하는 지수로 많이 쓰이는 지니계수(gini index)를 산출할 수 있다는 당연한 사실을 상기하면, 가구소득이 자연대수 정규분포를 따른다는 것은 가구소득 또는 자연대수 가구소득의 평균과 분산 2개의 계수의 함수로서 바로 지니계수를 계산할 수 있다는 것을 의미한다.

이를 보다 자세히 설명하기 위해 소득을 y 라고 표기하고, y 의 평균과 분산을 각기 μ 와 σ^2 라고 하자. 이때 y 가 자연대수 정규분

포를 따른다는 사실을 $y \sim \lambda(\mu, \sigma^2)$ 라고 표기할 수 있다. 이때 y 의 자연대수를 취하여 구한 변수 x 는 정규분포를 따르며, x 의 평균과 분산을 각기 m 과 s^2 라고 할 때, 이를 $x \equiv \ln y \sim N(m, s^2)$ 라고 표기할 수 있다. 이때 x 의 확률밀도함수(probability density function)는 다음과 같다.

$$\phi_N = \frac{\exp\left\{-\frac{1}{2}\left(\frac{x-m}{s}\right)^2\right\}}{s\sqrt{2\pi}}, \quad -\infty < x < \infty$$

자연대수 정규분포를 따르는 y 의 확률밀도함수는 x 의 평균과 분산 m 과 s^2 를 이용하여 다음과 같이 적을 수 있다.

$$\phi_\lambda(y) = \begin{cases} \frac{\exp\left\{-\frac{1}{2}\left(\frac{\ln y - m}{s}\right)^2\right\}}{ys\sqrt{2\pi}}, & y > 0 \\ 0, & y \leq 0 \end{cases}$$

또한 (m, s^2) 나 (μ, σ^2) 중 어느 하나의 순서쌍을 알 경우 다른 하나의 순서쌍을 구할 수 있다. 즉, (m, s^2) 를 알 경우,

$$\begin{aligned} \mu &= \exp\left\{\frac{2m + s^2}{2}\right\}, \\ \sigma^2 &= \exp(2m + 2s^2) - \exp(2m + s^2) \end{aligned}$$

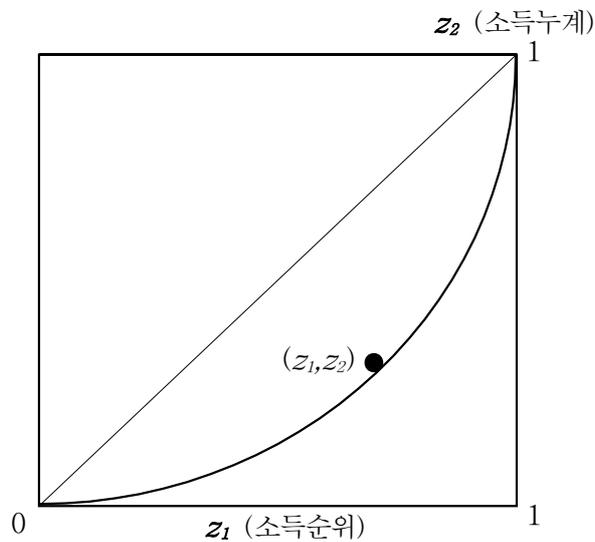
를 구할 수 있으며, 반대로 (μ, σ^2) 를 알면 다음과 같이 (m, s^2) 를 구할 수 있다.

$$m = \ln\left(\frac{\mu^2}{\sqrt{\sigma^2 + \mu^2}}\right),$$

$$s^2 = \ln\left(\left(\frac{\sigma}{\mu}\right)^2 + 1\right)$$

이제 위의 식들을 이용하여 지니계수를 산출하는 방법을 살펴보기로 하자. 우선 [그림 VI-1]을 보면 45도 선 아래의 호와 같은 형태의 곡선이 로렌츠 곡선(Lorenz curve)을 나타낸다. 이는 소득의 순위별로 전 가구를 횡축으로 나열하면서, 종축으로는 당해 소득순위까지의 누적총소득을 기록함으로써 구할 수 있다. 이때 양축의 상한은 1로 정의되도록 소득순위와 누적총소득의 단위를 조정한다.

[그림 VI-1] 로렌츠 곡선 (Lorenz curve)의 예시



이제 지니계수를 계산하기 위하여 로렌츠 곡선상의 임의의 점을 (z_1, z_2) 라고 표기하기로 하자. 이때 (z_1, z_2) 는 소득 a 를 ($0 < a < \infty$)

매개변수로 이용함으로써 다음과 같이 쓸 수 있다. 횡축의 값을 나타내는 $z_1(a)$ 는 $[0,1]$ 구간에서의 소득순위이므로, 다음아닌 소득의 누적분포함수이다. 따라서 소득 y 가 자연대수 정규분포일 경우 $[0,1]$ 구간에서 소득 a 에 해당하는 순위를 나타내는 $z_1(a)$ 는 다음 식과 같다.

$$z_1(a) \equiv \int_0^{\infty} \phi_{\lambda}(y) dy$$

또한 $z_2(a)$ 는 정의상 소득 a 까지의 누적총소득을 경제 내의 총 소득으로 나눈 값이므로, 이를 그대로 수식으로 표현하면 다음과 같다.

$$\begin{aligned} z_2(a) &\equiv \frac{\int_0^a y \phi_{\lambda}(y) dy}{\int_0^{\infty} y \phi_{\lambda}(y) dy} \\ &= \frac{\int_0^a y \phi_{\lambda}(y) dy}{\mu} \end{aligned}$$

이와 같이 $(z_1(a), z_2(a))$ 를 알면 로렌츠 곡선의 식을 암묵적으로 정의했다고 볼 수 있다. 이제 로렌츠 곡선 밑의 구역의 면적을 $g(m, s)$ 라고 표기하기로 하자. 지니계수는 45도 선 이하의 삼각형 중 45도 선과 로렌츠 곡선 사이 구역의 면적이 차지하는 비율로 정의된다. 그러므로 (m, s) 에 따라 주어지는 지니계수를 나타내는 함수인 지니표면(gini surface)을 $G(m, s) = 1 - 2g(m, s)$ 라고 쓸 수 있다. 우선 $g(m, s)$ 를 그 정의에 따라서 풀어쓰고, 그렇게 구한 식을 계산하면 다음과 같이 된다.

$$\begin{aligned}
g(m, s) &= \int_0^1 z_2 dz_1 = \int_0^\infty z_2(a) z_1'(a) da = \int_0^\infty z_2(a) \phi_\lambda(a) da \\
&= \int_0^\infty \frac{1}{\mu} \left[\int_0^a k \phi_\lambda(k) dk \right] \phi_\lambda(a) da
\end{aligned}$$

마지막 줄의 적분식은 더 이상 수식의 형태로는 풀 수 없으나, 구체적으로 특정한 (m, s) 의 값을 대입하면 그에 해당되는 $g(m, s)$ 의 값을 구할 수 있다. 이와 같은 방법으로 (m, s) 공간에서 아주 미세한 격자(grid)를 적용하여 약 10만개의 (m, s) 값에 대하여 지니계수를 구한 결과, 아래 [그림 VI-2]에 제시한 지니표면(gini surface)을 구할 수 있었다³³⁾. 그림에서는 3차원의 지니표면을 등고선을 사용하여 다시 2차원의 도형으로 나타내었다. 즉 그림에서 각각의 선은 같은 지니계수 값에 해당되는 (m, s) 조합의 궤적이며, 독자의 편의를 위하여 일정간격의 지니계수 값을 나타내는 동지니곡선(iso-gini curve)마다 해당 지니계수 값을 0.23부터 0.32까지 0.01 간격으로 표시하였다.

33) 이 지니표면을 구하기 위하여 횡축으로는 15부터 20까지 0.01 간격으로 501개의 m 값을 대입하였으며, 각각의 m 값에 대하여 종축으로는 0.001 간격으로 0.4부터 0.6까지 201개의 s 값을 대입하여 각각의 경우에 해당되는 지니계수값을 계산하였다. 이 과정에서 경희대의 전택승 교수님의 도움이 결정적이었음을 밝히며 감사를 표하고자 한다.

[그림 VI-2] 지니표면: 자연대수가구소득의 평균 m 과 분산 s 에
따른 지니계수의 도출

주: 실제 자료에서 도출한 각 연도의 m 과 s 값을 (<표 III-1> 참조) 지니표
면 위에 그림으로써 당해 연도의 지니계수를 읽을 수 있음. 지니표면을
사용한 상세한 지니계수 값은 아래 <표 VI-1>에 제시되어 있음.

VI. 지니표면을 이용한 자연대수 소득의 분산 추정 137

위의 그림에서 가장 두드러진 특징은 매우 높은 m 값의 영역을 제외하고는³⁴⁾ 동지니곡선이 수평선이라는 사실이다. 즉, 지니계수가 자연대수 가구소득의 평균 m 의 영향을 받지 않으며, 전적으로 표준편차 s 에 의해 결정된다는 사실을 볼 수 있다. 또한 동지니곡선의 간격이 일정함으로 보아, 지니계수를 자연대수 가구소득의 표준편차 s 의 1차 선형식으로 나타낼 수 있을 것으로 보인다. 실제로 m 의 값이 18 이하인 구역에 대하여 회귀식을 적용해 본 결과, 다음과 같이 지니계수를 도출하는 공식을 간단히 정리할 수 있다³⁵⁾. 이 식을 통하여 알 수 있는 사실은 소득분배의 형평성을 가늠하는 지니계수가 자연대수 소득분포의 분산(표준편차)에 의해 사실상 바로 결정된다는 것이다.

$$G(m, s) = 0.011217 + 0.529774 \times s, (R^2 = 0.999999) \quad (2)$$

역의 관계도 마찬가지로 설명이 가능하다. 따라서 위의 회귀방정식에 대하여 역회귀(reverse regression) 방정식을 설정하여 추정하면 다음의 결과를 얻을 수 있다.

34) 이러한 현상이 원래 지니표면의 속성 때문인지, 혹은 계산과정에서의 정확성(precision level)의 한계 때문인지는 명확하지 않음을 밝혀둔다.
 35) 자연대수 가구소득의 평균 m 을 회귀식에 포함시켜 보았으나, 예상대로 계수의 값이 0이거나 0과 매우 가까운 작은 값으로 추정되어 유의성을 확보하지 못하였다. 자연대수 가구소득의 표준편차 s 만을 사용한 회귀식은 3차식까지 시도하여 보았으며, 다소 의외로 계수의 값이 모두 0.0001 수준에서 유의한 회귀식 $G = -0.000395 + 0.567224s - 0.008252s^2 - 0.038620s^3$ 을 얻었다. 그러나 1차식에서 R^2 의 값이 이미 1로 나왔기 때문에 회귀식의 차수를 높임으로써 설명력이 향상되지 않았으며, 실제로 1차식과 3차식을 사용하여 계산한 지니계수 값에는 차이가 없었다. 또한 실제로 선형이라고 볼 수 있는 관계에서 3차식을 적용하여도 유의한 계수가 추정되는 현상은, 지니표면 도출과정에서 적분식을 계산할 때 컴퓨터의 정확성(precision level)의 한계에 기인했을 가능성을 배제할 수 없다.

$$s(m, g) = -0.021151 + 1.887522g, (R^2 = 0.999999) \quad (3)$$

위의 두 식 가운데 표준편차로 지니계수를 설명한 식 (2)의 회귀 방정식의 변형으로, 가구소득의 자연대수를 구하여 표준편차 s 를 도출하지 않고서도, 단지 자연대수를 취하지 않은 가구소득의 평균과 분산 (μ, σ^2)을 구하여 다음과 같이 간단하게 지니계수를 산정할 수 있다. 이때 경험적으로 실제로 관측되는 소득분포를 보면 평균 대 표준편차의 비율의 제곱이 작은 숫자이므로, 소득분포의 평균과 표준편차만 알면 아래와 같이 상당히 간단한 식을 이용하여 지니계수의 근사치를 계산할 수 있다. 아래의 식에서 $CV(y)$ 는 변수 y (소득)의 변이계수(coefficient of variation)로서, 분포의 표준편차를 평균으로 나눈 값으로 정의된다. 즉, 소득의 형평성을 측정하는 지니계수는 소득분포가 자연대수 정규분포일 경우, 결국 분포의 변이계수에 거의 정비례하여 결정됨을 볼 수 있다. 또한 (자연대수를 취하지 않은) 가구소득 분포의 평균과 표준편차인 μ - σ 평면에서 [그림 VI-2]와 같은 방법으로 지니표면을 그릴 경우, 동지니곡선은 $\frac{\sigma}{\mu}$ 의 값이 일정한 점들의 궤적이므로 원점을 지나는 직선 모양이 될 것임을 알 수 있다. 이때 원점을 지나는 직선이 시계 반대 방향으로 회전하면서 지니계수 값이 증가할 것이다.

$$\begin{aligned} G(m, s) \equiv \Gamma(\mu, \sigma) &= 0.011217 + 0.529774 \sqrt{\ln\left(\left(\frac{\sigma}{\mu}\right)^2 + 1\right)} \\ &\approx 0.011217 + 0.529774 \frac{\sigma}{\mu} \\ &\quad (\because \left(\frac{\sigma}{\mu}\right)^2 \text{의 값이 작으므로 } \ln\left(\left(\frac{\sigma}{\mu}\right)^2 + 1\right) \approx \left(\frac{\sigma}{\mu}\right)^2) \\ &\equiv 0.011217 + 0.529774 CV(y) \end{aligned}$$

VI. 지니표면을 이용한 자연대수 소득의 분산 추정 139

[그림 VI-2]로 다시 돌아가서, 자연대수 소득의 평균 m 이 높은 영역에서는 동지니곡선이 우하향으로 휘어짐을 볼 수 있다. 따라서 이 영역에서는 위에서 도출한 s 의 1차식으로서의 지니계수 추정식은 성립하지 않는다. 그러나 <표 VI-1>에서 볼 수 있듯이, 여기에서도 특정 m 값에 대해서는 지니계수를 s 의 1차식으로 구할 수 있다. 다만 동지니곡선 간의 간격이 m 의 값이 높아질수록 커진다는 사실을 s 의 계수에서 알 수 있다. 참고로, m 값이 18 이상인 영역에 국한해서 [그림 VI-2]의 지니표면 도출에 사용된 약 4만개의 m, s 지니값을 대입하여 지니계수를 m 과 s 의 다항식으로 추정한 결과, $R^2 = 0.9926$ 으로 <표 VI-2>에 제시된 식을 구할 수 있었다.

마지막으로, 본고에서 제시한 지니표면을 사용하여 산정한 지니계수와 통상적인 방법으로 계산한 지니계수를 비교함으로써 지니표면의 타당성을 검증할 수 있다. 간접적으로는 가구소득이 자연대수 정규분포를 따른다는 본 연구의 주장을 다시 한 번 우회적으로나마 검증할 수 있다. 아래의 <표 VI-3>에는 양 기법에 따른 지니계수의 시계열이 제시되어 있으며, 그 차이가 비교적 미미하다는 사실을 확인할 수 있다. 부연하자면, 지니표면을 이용하여 계산한 지니계수는 자연대수 가구소득이 정규분포라는 가정하에 구한 값이다. 따라서 정규분포에서 크게 벗어난 1997, 1998년의 외환위기 당시만 두 방법에 의한 지니계수의 값에 큰 차이가 보일 뿐이며, 그 외의 연도에서는 차이가 비교적 작다.

<표 VI-1> m 의 변화에 따른 지니계수 추정식의 변화

m	상수항	s 의 계수	R2
17.00	0.011217	0.529774	0.9999
17.25	0.011217	0.529774	0.9999
17.50	0.011217	0.529774	0.9999
17.75	0.011216	0.529775	0.9999
18.00	0.011209	0.529789	0.9999
18.25	0.011161	0.529896	0.9999
18.50	0.010869	0.530543	0.9999
18.75	0.009393	0.533846	0.9999
19.00	0.003279	0.547760	0.9999
19.25	-0.016877	0.594987	0.9992
19.50	-0.066491	0.718296	0.9980
19.75	-0.144171	0.943430	0.9992
20.00	-0.173080	1.163330	0.9995

주: 종속변수는 지니계수이며, 각 줄은 해당 m 에 국한하여 지니계수를 s 의 1차식으로 추정된 결과를 나타냄.

<표 VI-2> $m > 18$ 영역에서의 지니표면의 추정식

변수	추정계수
상수항	-371.856381
s	134.968517
s^2	-38.005982
s^3	21.856784
s^4	-0.002797
m	50.729354
m^2	-1.946444
m^3	0
m^4	0.000850
sm	-13.306363
s^2m	2.034259
s^3m	-1.167562
sm^2	0.326874
sm^3	0
s^2m^2	0

주: 종속변수: 지니계수, $R^2 = 0.9926$
 변수간의 다중공선성이 존재하여 각 계수의 추정치나 유의성에 의미를 부여할 수 없음.

<표 VI-3> 지니표면과 기존 방법을 사용한 지니계수의 비교

	지니계수	
	지니표면 사용	기존 방법
1982	0.281444	0.28678
1983	0.278997	0.28270
1984	0.278737	0.28109
1985	0.270785	0.27563
1986	0.271977	0.27368
1987	0.275500	0.27986
1988	0.269164	0.27082
1989	0.265392	0.26957
1990	0.250903	0.25236
1991	0.243216	0.24189
1992	0.244529	0.24147
1993	0.246288	0.24242
1994	0.246569	0.24293
1995	0.248460	0.24280
1996	0.252232	0.24685
1997	0.324663	0.28102
1998	0.362733	0.31121
1999	0.271749	0.26726
2000	0.282271	0.27988
2001	0.273667	0.27222
2002	0.272438	0.27997

주: 지니표면은 가구소득이 자연대수 정규분포를 따른다는 가정하에 지니계수를 도출한 것임. 따라서 동 가정이 위배된 1997, 1998년에는 기존 방법으로 도출한 지니계수와 비교적 큰 차이를 보임.

2. 자연대수 소득분산의 추정

가. 분석방법

소득분포가 자연대수 정규분포를 따른다고 하면 자연대수를 취한 소득의 평균과 분산, 그리고 상대소득분포를 나타내는 지니계수의 세 가지 요소는 다음과 같은 특징을 지닌다. 세 가지 요소 중 어느 하나가 임의의 값으로 주어졌다면 나머지 두 변수 사이에는 일대일

대응관계 또는 단조관계(monotonicity)가 성립한다. 그러므로 외생적인 요인을 감안하여 향후 특정 시점에서의 자연대수 소득의 평균을 추정하였다면, 주어진 자연대수 소득평균의 가정하에서 분산을 추정하든가(직접법) 또는 지니계수를 추정하고 지니계수 추정치에 대응되는 분산을 찾는 방법(간접법)을 이용하여 소득분포의 특성을 규정할 수 있다.

그런데 분산을 직접 추정하기에는 소득분포의 특성을 모두 파악하는 데 있어 필요한 정보가 충분하지 않다. 지니계수의 경우에도 그 값을 결정해주는 각종 요인이 충분히 알려져 있지 않기 때문에 지니계수를 직접 추정하는 것 또한 용이하지 않다. 다만 선진국에서의 지니계수 변화추이와 경제발전 단계 등을 감안하여 우회적으로 지니계수의 변화방향을 예측하는 방법을 생각할 수 있다. 이 방법은 논리적으로 기초가 탄탄하지는 않지만 과거의 추세와 선진국의 경험을 토대로 지니계수의 변화추이에 대한 시나리오를 작성·분석하는 데에는 도움이 된다. 따라서 본 절에서는 선진국과 우리나라의 소득분배 구조의 변화에 대한 경험자료를 근거로 우리나라 지니계수의 변화추이를 예상해보면서 향후 소득분포의 특성 중 하나인 분산을 추정해본다.

먼저 이를 위해 주요 선진국과 우리나라의 지니계수 변화추이를 비교해보고 이를 토대로 향후의 지니계수 변화추이를 예측해본다.

나. 지니계수 변화추이의 국제비교

선진국에서는 대부분 1970년대 말 또는 1980년대 초부터 지니계수가 상승추세를 보이기 시작하여 상대적인 소득분배의 격차가 확대되기 시작한 것으로 보인다.

미국의 경우에는 1960년대 말 이래 1970년대 말까지 지니계수가 0.4 수준에서 안정적인 모습을 보였다. 그러나 1980년대에 접어들면

VI. 지니표면을 이용한 자연대수 소득의 분산 추정 143

서부터 지니계수가 상승하기 시작하여 상대소득분포의 격차가 확대되기 시작한 것으로 추정된다. 이러한 추세는 최근까지도 지속되어 2002년 현재 지니계수는 0.462로 1970년대에 비해 약 0.06지니p 정도 높다.

영국의 경우에도 미국과 마찬가지로 1980년대부터 지니계수가 상승하기 시작한 것으로 보인다. 1970년대 말에는 지니계수가 0.43~0.45 수준에 머물렀으나 1980년대부터는 서서히 상승하기 시작하여 1980년대 말에는 0.5 수준으로 상승하였으며 1990년대에도 상승추세를 지속하여 2002년에는 0.53 수준이다. 1978~2002년 사이의 지니계수 상승폭은 0.10지니p에 이를 정도로 격차가 크다.

호주의 경우 지니계수의 변화를 논하기에는 자료의 축적기간이 짧아 추세적인 흐름을 파악하기 어렵다.

그렇지만 일본의 경우에는 지난 10여년간에 걸쳐 지니계수의 변화가 상당히 컸다. 시장소득을 기준으로 1981년 현재 일본의 지니계수는 0.3491이었는데 1996년에는 0.4412로 급상승하였다. 일본에서 소득분배 격차가 확대된 데에는 장기간의 경기침체 등이 근본적인 요인이 되었을 것으로 추정된다.

<표 VI-4> 미국의 소득불평등도 계측결과

	지니	Var (ln소득)	MLD (Mean log deviation)	타일지수	Atkinson 지수		
					e=0.25	e=0.50	e=0.75
1967	0.399	0.813	0.380	0.287	0.071	0.143	0.220
1968	0.388	0.779	0.356	0.273	0.067	0.135	0.208
1969	0.391	0.774	0.357	0.268	0.067	0.135	0.209
1970	0.394	0.805	0.370	0.271	0.068	0.138	0.214
1971	0.396	0.782	0.370	0.273	0.068	0.138	0.214
1972	0.401	0.790	0.370	0.279	0.070	0.140	0.216
1973	0.397	0.776	0.355	0.270	0.068	0.136	0.210
1974	0.395	0.740	0.352	0.267	0.067	0.134	0.207
1975	0.397	0.746	0.361	0.270	0.067	0.136	0.210
1976	0.398	0.756	0.361	0.271	0.068	0.137	0.211
1977	0.402	0.768	0.364	0.276	0.069	0.139	0.213
1978	0.402	0.784	0.363	0.275	0.069	0.139	0.213
1979	0.404	0.816	0.369	0.279	0.070	0.141	0.216
1980	0.403	0.792	0.375	0.274	0.069	0.140	0.216
1981	0.406	0.826	0.387	0.277	0.070	0.141	0.220
1982	0.412	0.864	0.401	0.287	0.072	0.146	0.226
1983	0.414	0.865	0.397	0.288	0.072	0.147	0.226
1984	0.415	0.859	0.391	0.290	0.073	0.147	0.225
1985	0.419	0.863	0.403	0.300	0.075	0.151	0.231
1986	0.425	0.888	0.416	0.310	0.077	0.155	0.237
1987	0.426	0.936	0.414	0.311	0.077	0.155	0.238
1988	0.427	0.899	0.401	0.314	0.078	0.155	0.236
1989	0.431	0.887	0.406	0.324	0.080	0.158	0.239
1990	0.428	0.888	0.402	0.317	0.078	0.156	0.236
1991	0.428	0.878	0.411	0.313	0.078	0.156	0.237
1992	0.434	0.925	0.416	0.323	0.080	0.160	0.242
1993	0.454	0.942	0.467	0.385	0.092	0.178	0.266
1994	0.456	0.976	0.471	0.387	0.092	0.180	0.268
1995	0.450	0.945	0.452	0.378	0.090	0.175	0.261
1996	0.455	0.954	0.464	0.389	0.093	0.179	0.266
1997	0.459	0.966	0.484	0.396	0.094	0.183	0.272
1998	0.456	0.973	0.488	0.389	0.093	0.081	0.271
1999	0.458	n.a.	0.476	0.386	0.092	0.180	0.268
2000	0.462	n.a.	0.490	0.404	0.096	0.185	0.275
2001	0.466	n.a.	0.515	0.414	0.098	0.189	0.282
2002	0.462	n.a.	0.514	0.398	0.095	0.186	0.279

주: 자료 체계의 불연속으로 1992년까지와 1993년부터의 추정결과는 일치성을 가지지 않음.

자료: U.S. Census Bureau, Money Income in the United States, 각 연도.

<표 VI-5> 영국의 지니계수 추정결과

	기준 소득(균등화 소득 기준)			
	시장소득 (A)	총소득 (B)	가처분소득 (C)	조세후 소득 (D)
1978	0.43	0.29	0.26	0.28
1979	0.44	0.30	0.27	0.29
1980	0.44	0.31	0.28	0.30
1981	0.46	0.31	0.28	0.31
1982	0.47	0.31	0.28	0.31
1983	0.48	0.32	0.28	0.31
1984	0.49	0.31	0.28	0.30
1985	0.49	0.32	0.29	0.32
1986	0.50	0.34	0.31	0.35
1987	0.51	0.36	0.33	0.36
1988	0.51	0.37	0.35	0.38
1989	0.50	0.36	0.34	0.37
1990	0.52	0.38	0.36	0.40
1991	0.51	0.37	0.35	0.39
1992	0.52	0.37	0.34	0.38
1993	0.53	0.38	0.35	0.38
1994~95	0.54	0.37	0.34	0.38
1995~96	0.52	0.36	0.33	0.37
1996~97	0.53	0.37	0.34	0.38
1997~98	0.53	0.37	0.34	0.38
1998~99	0.53	0.38	0.35	0.39
1999~2000	0.53	0.38	0.35	0.40
2000~2001	0.51	0.38	0.35	0.39
2001~2002	0.53	0.39	0.36	0.40

주: 1. 균등화된 소득은 가구원 수에 따라 가구별 실질소득을 조정된 것을 말함.

2. 조세후소득은 가처분소득에서 소비세 부담을 차감한 소득임.

자료: Harris(2000), Lakin(2002, 2003).

<표 VI-6> 호주의 지니계수 추정결과

	5분위 소득점유율						지니계수	
	1	2	3	4	5	계	총소득	가처분소득
1994~1995	3.6	9.3	15.2	24.0	47.9	100.0	0.443	-
1995~1996	3.8	9.1	15.0	23.7	48.3	100.0	0.437	-
1996~1997	3.8	9.4	15.2	24.0	47.5	100.0	0.444	-
1997~1998	3.8	9.0	15.0	23.9	48.3	100.0	0.446	-
1999~2000	3.8	9.0	15.0	23.8	48.5	100.0	0.448	0.396

자료: Australian Bureau of Statistics, Income Distribution, 1996~97, Australia, 1998.

——, Income Distribution, 1997~98, Australia, 1999.

——, Income Distribution, 1999~2000, Australia, 2001.

<표 VI-7> 일본의 지니계수 추정결과

	시장소득	재분배후 소득 (가처분소득)	세후소득 (시장소득-소득세)	사회보장후 재분배소득
1981	0.3491	0.3143	0.3301	0.3317
1984	0.3975	0.3426	0.3824	0.3584
1987	0.4049	0.3382	0.3879	0.3564
1990	0.4334	0.3643	0.4207	0.3791
1993	0.4394	0.3645	0.4255	0.3812
1996	0.4412	0.3606	0.4338	0.3721

주: 1. 시장소득(원소득) = 피용자보수 + 영업(사업)소득 + 농업소득 + 재산소득 + 가사노동소득 + 잡소득 + 사적보조(benefits privately received: 송금, 기업연금, 퇴직 관련 수혜금, 생명보험금 등)

2. 재분배후 소득 = 원소득 - 조세 - 사회보험료 + 사회보장수혜 + 의료비

3. 사회보장후 재분배소득(총소득에 해당) = 시장소득 - 사회보험료 + 사회보장수혜 + 의료비

4. 조세: 직접세 중 소득세, 주민세, 재산세(사업용 부동산 관련 재산세 제외), 자동차세(사업용 제외)

5. 사회보험료: 피용자보험료, 건강의료보험, 국민연금보험료

6. 사회보장수혜: 사회보장체계하에서의 연금수혜 등

7. 의료비: 의료비수혜 및 유사수혜의 현금가치

8. 변화율: 시장소득 기준 지니계수 대비 지니계수의 변화율

자료: "Survey on the Redistribution of Income(Fiscal 1996)" by the Research Section, Policy Planning and Evaluation Division, Minister's Secretariat, Ministry of Health and Welfare.

VI. 지니표면을 이용한 자연대수 소득의 분산 추정 147

뉴질랜드의 경우에는 시장소득(총소득)을 기준으로 할 때 1982~1996년 사이에 지니계수가 0.566에서 0.609(총소득 기준으로는 0.472→0.495)로 소폭 상승하였다. 상승폭만 고려한다면 일본·영국·미국에 비해 변화의 폭이 매우 작다. 그런데 뉴질랜드의 경우에는 이미 시장소득 단계에서 지니계수의 값이 0.6에 육박할 정도로 매우 높기 때문에 소득분배 격차가 더욱 확대되더라도 지니계수 단위로 표시된 소득분배 격차의 확대 정도는 상대적으로 작게 나타날 수밖에 없음에 유의할 필요가 있다.

캐나다의 경우에도 시장소득을 기준으로 할 때 지니계수는 1989년 0.396에서 1998년 0.437로 0.041지니p 상승하였다. 싱가포르의 경우에도 1990~2000년 사이에 지니계수가 0.436에서 0.481로 상승하였다.

<표 VI-8> 뉴질랜드의 지니계수(근로소득자)

15세 이상 국민 대상 지니계수	1982	1986	1991	1996
시장소득 기준	0.566	0.542	0.611	0.609
총소득 기준	0.472	0.437	0.474	0.495
가처분소득 기준	0.414	0.386	0.437	0.456

자료: Statistics New Zealand(1999)

<표 VI-9> 캐나다의 지니계수

	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
시장소득 기준 (A)	0.396	0.402	0.419	0.424	0.429	0.424	0.427	0.431	0.434	0.437
총소득 기준 (B)	0.329	0.330	0.337	0.336	0.339	0.334	0.341	0.347	0.353	0.357
세후소득 기준 (C)	0.292	0.291	0.297	0.295	0.298	0.292	0.297	0.305	0.312	0.315

주: 2인 이상 가구 기준

자료: Statistics Canada(2000).

<표 VI-10> 싱가포르의 가계소득분배

	1990	1995	1997	1998	1999	2000
지니계수 (전체)	0.436	0.443	0.444	0.446	0.467	0.481
(무소득자 가구 제외)	0.412	0.409	0.412	0.410	0.424	0.432

자료: Census of Population Office, Singapore Department of Statistics (2001).

이상에서 보듯이 호주를 제외한 나머지 선진국(미국·영국·일본·뉴질랜드·캐나다·싱가포르 등)에서는 모두 지난 10~20년 정도의 기간 동안 소득분배 격차를 나타내는 지니계수가 상승하는 모습을 보였다.

선진 각국에서 지니계수가 상승한 데에는 여러 가지 요인이 있겠지만 최근에는 정보화 사회의 급속한 진전에 따른 계층간 정보력 격차와 경제 패러다임의 변화에 따른 세대간·계층간 지식·경험의 차이 등이 확대되면서 나타난 현상인 것으로 추정된다. 그 밖에도 세계화·개방화의 진전에 따라 소득분배의 격차가 확대된 부분도 상당히 클 것으로 추정된다.

이 가운데 전자의 경우에는 새로운 정보화 사회가 전개되면서 이러한 환경변화에 적응한 소수가 상방으로 소득이 크게 증가한 데 따른 것인 반면, 후자의 경우에는 세계화·개방화에 따라 몰락한 부문에 종사하였던 사람들은 소득수준이 하방으로 이동하면서 양자간의 격차가 확대된 측면이 있다. 제반 환경적 요인에 대해 소득분배의 변화 양상은 상당히 다르게 나타나지만 결과적으로 지니계수를 상승시키는 방향으로 진전되었음에는 큰 차이가 없는 것으로 분석된다.

우리나라의 경우에는 1980년대 이래 1990년대 초까지 지니계수가 꾸준히 하락하여 소득분배 격차가 축소되었으며 1990년대 중반까지 0.24 정도의 수준에서 안정화되었다가, 1990년대 후반에 접어들면서부터 지니계수가 상승추세로 반전된 것으로 보인다. 이는 전 장에서 소득분포가 자연대수 정규분포의 틀에서 벗어난 것으로 분석된 1997~1998년의 경제위기 기간을 제외할 경우 1990년대 중반 이후의 지니계수의 궤적이 우상향하는 추세를 보이는 것으로부터 유추할 수 있다.

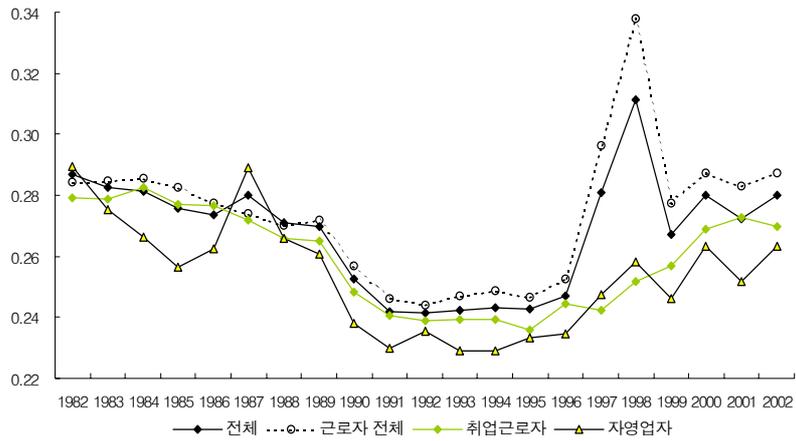
VI. 지니표면을 이용한 자연대수 소득의 분산 추정 149

<표 VI-11> 우리나라의 총소득 기준 지니계수 추정치

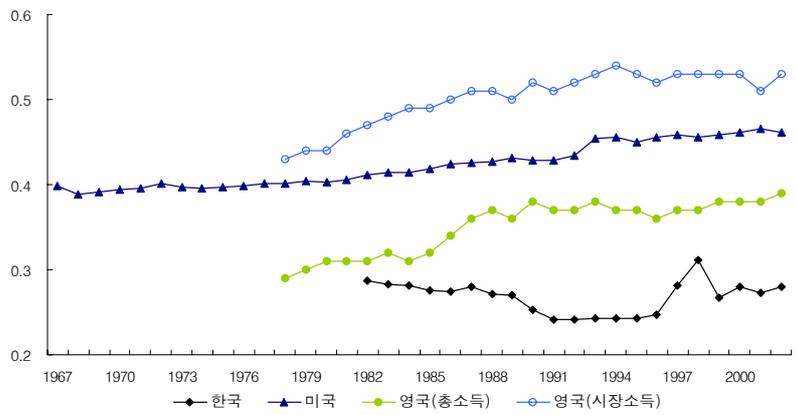
	전체	근로자가구		자영업자	실업률(%)	경제활동 참가율(%)	통계청 지니계수
		전체	취업근로자				
1982	0.28678	0.28444	0.27920	0.28937	4.4	58.6	-
1983	0.28270	0.28463	0.27879	0.27512	4.1	57.7	-
1984	0.28109	0.28539	0.28276	0.26641	3.8	55.8	-
1985	0.27563	0.28258	0.27683	0.25639	4.0	56.6	-
1986	0.27368	0.27732	0.27678	0.26253	3.8	57.1	-
1987	0.27986	0.27383	0.27181	0.28886	3.1	58.3	-
1988	0.27082	0.27034	0.26569	0.26583	2.5	58.5	-
1989	0.26957	0.27185	0.26503	0.26058	2.6	59.6	0.3039
1990	0.25236	0.25693	0.24819	0.23780	2.4	60.0	0.2948
1991	0.24189	0.24609	0.24042	0.22973	2.3	60.6	0.2869
1992	0.24147	0.24379	0.23900	0.23529	2.4	60.9	0.2836
1993	0.24242	0.24698	0.23922	0.22883	2.8	61.1	0.2817
1994	0.24293	0.24869	0.23947	0.22901	2.4	61.7	0.2845
1995	0.24280	0.24674	0.23579	0.23308	2.0	61.9	0.2837
1996	0.24685	0.25258	0.24448	0.23473	2.0	62.0	0.2907
1997	0.28102	0.29621	0.24209	0.24737	2.6	62.2	0.2830
1998	0.31121	0.33777	0.25163	0.25824	6.8	60.7	0.3157
1999	0.26726	0.27731	0.25700	0.24630	6.3	60.5	0.3204
2000	0.27988	0.28718	0.26878	0.26323	4.1	60.7	0.317
2001	0.27222	0.28288	0.27276	0.25156	3.7	60.8	0.319
2002	0.27997	0.28733	0.26991	0.26314	3.1	61.5	0.312

주: 통계청의 지니계수는 취업근로자가구만을 대상으로 한 것으로서, 각 분기별 소득분위의 소득을 분위별로 합산하여 연간소득분위를 산출하는 방법으로 추정된 것임.

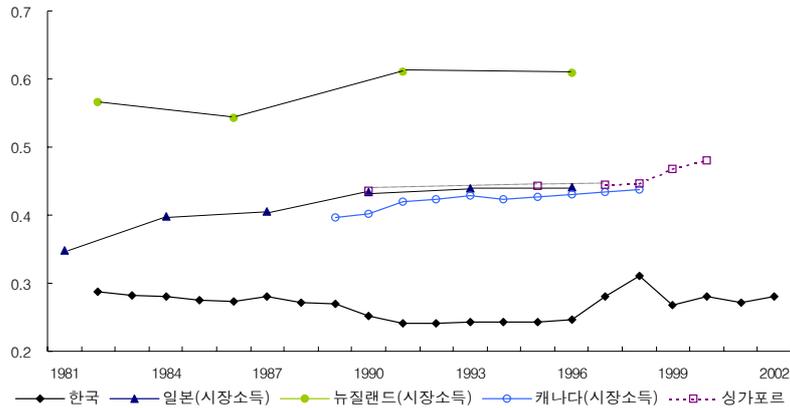
[그림 VI-3] 우리나라 지니계수의 변화추이



[그림 VI-4] 지니계수 변화추이의 국제비교 I



[그림 VI-5] 지니계수 변화추이의 국제비교 II



다. 지니계수 변화율에 대한 예측

앞에서 살펴보았듯이 상당수 선진국의 경우 이미 소득분배 격차의 확대 추세가 1980년대부터 시작되었다고 할 수 있다. 경제발전 단계상 우리나라가 선진국이 경험한 추세를 다소간 따라가고 있다는 점, 노령화, 이혼 급증, 정보력 격차 확대, 개방 가속화에 따른 경쟁의 격화 등이 모두 향후의 소득분배 격차를 확대시키는 요인으로 작용할 개연성이 높다는 점 등을 고려할 때 향후 우리나라의 소득분배 격차가 현재보다 축소될 것을 기대하기는 어려울 것으로 생각된다. 따라서 향후의 지니계수는 최소한 현재 수준보다 낮지는 않을 것으로 보인다. 오히려 제반 경제·사회적 여건의 변화추세를 감안하면 지니계수가 현재보다 확대될 여지가 큰 것으로 판단된다.

우리나라의 소득분배 구조가 현재 수준에서 안정화되는 것을 전제로 한다면 지니계수는 최근 수준(약 0.27~0.28) 정도를 유지할 것이다. 반면에 만약 소득분배 격차의 확대가 가속화된다면 그 속도는 선진국에서 경험하였던 추세적 변화속도 중 최대치를 상한으로 한다고 보아도 크게 무리하지 않을 것으로 생각된다. 현실적으

로는 최소치와 상한 사이의 정도에서 소득분배 격차 확대추세가 나타날 수도 있을 것으로 예상할 수 있다.

국가별로 기간의 장단에 따른 지니계수의 변화폭(상승폭)을 보면, 미국의 경우 20년간(1982~2002년) 지니계수가 0.05지니p, 영국은 같은 기간 동안 0.06지니p(1978~2002년 기간에는 0.10지니p), 일본은 15년간(1981~1996년) 0.09지니p, 뉴질랜드는 1982~1996년의 14년간 0.043지니p, 캐나다는 1989~1998년의 9년간 0.041지니p, 싱가포르는 1990~2000년의 10년간 0.045지니p씩 상승하였다. 이를 국가별로 산술평균해보면 지니계수 상승폭은 연간 0.0025~0.006지니p의 범위이다.

우리나라의 경우 지니계수가 상승하기 시작한 1995년부터 최근(2002년)까지 지니계수의 변화폭은 0.03717지니p, 연평균으로는 0.0053지니p 정도이다. 우리나라의 연평균 지니계수 상승폭은 상기 선진국에서의 변화폭 중 상위에 속한다고 할 수 있다.

국가별로 통계자료의 종류와 지니계수 산출방법이 상이하기 때문에 지니계수의 절대수준에 대한 직접 비교는 의미를 찾기 어려운 부분이 있는 것이 사실이다. 그렇지만 지니계수의 변화방향이나 변화율 등의 측면에서는 자료와 산출방법의 차이에도 불구하고 어느 정도 의미를 찾을 수 있을 것으로 사료된다. 따라서 위에서 살펴본 듯이 앞으로 우리나라의 지니계수가 상승하는 추세를 보일 것으로 가정한다면 선진국에서 최근 20여년간 지니계수가 상승하였던 추세 중 최대치를 상한으로 하여 향후 지니계수의 변화추이에 대한 시나리오를 작성할 수 있다.

먼저 2002년의 지니계수 0.28을 기준으로 할 때, 2022년을 예로 들면 지니계수의 하한치는 소득분배 구조의 안정화라는 가정하에 0.28 수준에 머물 것으로 가정한다. 이와 반대로 선진국에서의 지니계수 상승속도를 감안할 때 2022년의 지니계수는 0.38~0.40 정도를 상한으로 변화할 것으로 예상할 수 있다. 상기 선진국에서의 지

VI. 지니표면을 이용한 자연대수 소득의 분산 추정 153

니계수 상승폭의 중간 정도(연평균 0.003~0.004지니p)로 소득분배 격차가 확대된다면 지니계수는 평균적으로 0.34~0.36 정도에 이를 것으로 추정할 수 있다.

향후 지니계수의 예측치에 대해 <표 VI-12>에서 보는 바와 같이 시나리오를 작성할 수 있다.

<표 VI-12> 지니계수 예측 시나리오에 대응한 자연대수소득 표준편차(분산) 추정결과

	지니계수				자연대수 소득 표준편차			
	하한	중간 I	중간 II	상한	하한	중간 I	중간 II	상한
2002	0.27997	0.27997	0.27997	0.27997	0.50730	0.50730	0.50730	0.50730
2003	0.28	0.283	0.285	0.286	0.50736	0.51302	0.51679	0.51868
2004	0.28	0.286	0.290	0.292	0.50736	0.51868	0.52623	0.53001
2005	0.28	0.289	0.295	0.298	0.50736	0.52434	0.53567	0.54133
2006	0.28	0.292	0.300	0.304	0.50736	0.53001	0.54511	0.55266
2007	0.28	0.295	0.305	0.310	0.50736	0.53567	0.55454	0.56398
2008	0.28	0.298	0.310	0.316	0.50736	0.54133	0.56398	0.57531
2009	0.28	0.301	0.315	0.322	0.50736	0.54699	0.57342	0.58663
2010	0.28	0.304	0.320	0.328	0.50736	0.55266	0.58286	0.59796
2011	0.28	0.307	0.325	0.334	0.50736	0.55832	0.59229	0.60928
2012	0.28	0.310	0.330	0.340	0.50736	0.56398	0.60173	0.62061
2013	0.28	0.313	0.335	0.346	0.50736	0.56964	0.61117	0.63193
2014	0.28	0.316	0.340	0.352	0.50736	0.57531	0.62061	0.64326
2015	0.28	0.319	0.345	0.358	0.50736	0.58097	0.63004	0.65458
2016	0.28	0.322	0.350	0.364	0.50736	0.58663	0.63948	0.66591
2017	0.28	0.325	0.355	0.370	0.50736	0.59229	0.64892	0.67723
2018	0.28	0.328	0.360	0.376	0.50736	0.59796	0.65836	0.68856
2019	0.28	0.331	0.365	0.382	0.50736	0.60362	0.66779	0.69988
2020	0.28	0.334	0.370	0.388	0.50736	0.60928	0.67723	0.71121
2021	0.28	0.337	0.375	0.394	0.50736	0.61494	0.68667	0.72253
2022	0.28	0.340	0.380	0.400	0.50736	0.62061	0.69611	0.73386
2023	0.28	0.343	0.385	0.406	0.50736	0.62627	0.70554	0.74518
2024	0.28	0.346	0.390	0.412	0.50736	0.63193	0.71498	0.75651
2025	0.28	0.349	0.395	0.418	0.50736	0.63759	0.72442	0.76783
2026	0.28	0.352	0.400	0.424	0.50736	0.64326	0.73386	0.77916
2027	0.28	0.355	0.405	0.430	0.50736	0.64892	0.74330	0.79048
2028	0.28	0.358	0.410	0.436	0.50736	0.65458	0.75273	0.80181
2029	0.28	0.361	0.415	0.442	0.50736	0.66024	0.76217	0.81313
2030	0.28	0.364	0.420	0.448	0.50736	0.66591	0.77161	0.82446
2031	0.28	0.367	0.425	0.454	0.50736	0.67157	0.78105	0.83578
2032	0.28	0.370	0.430	0.460	0.50736	0.67723	0.79048	0.84711

주: 2002년의 자연대수소득의 표준편차는 직접계산법을 기준으로 0.51607임에 유의하기 바람. 위에서 2002년의 0.50730은 식 (3)을 기준으로 하였을 때 추정된 값임.

라. 자연대수 소득 분산에 대한 예측결과

지니계수의 값을 알 수 있다면 식 (3)의 역회귀방정식 추정결과를 바탕으로 자연대수 소득의 표준편차(분산)를 추정할 수 있다. 더욱이 분산의 값은 자연대수 소득의 평균에 의해서는 별로 영향을 받지 않으며 상대소득분포를 나타내는 지니계수와 상관관계가 높다. 그러므로 식 (3)을 기준으로 자연대수소득의 표준편차를 추정해보면 다음과 같다.

먼저 하한의 경우, 즉 지니계수가 0.28 수준을 그대로 유지하는 경우에는 자연대수 표준편차가 0.50736에서 일정할 것으로 예상된다. 만약 지니계수가 연간 0.03지니p씩, 0.05지니p씩, 또는 0.06지니p씩 상승한다고 하면 표준편차는 매년 각각 0.00566, 0.00944, 0.01325씩 함께 상승하는 것으로 추정되었다. 이에 따라 2032년 경우에는 표준편차가 각각 0.677, 0.790, 0.847 수준으로 상승할 것으로 예상되었다.

그런데 지난 20여년간의 소득분배구조 변화추이를 볼 때 그 정도 수준으로 높게까지 상승하지는 않을 가능성이 높을 것으로 추측된다. 이와 반대로 하한 시나리오의 경우에는 실현가능성이 낮으므로 실제로 향후의 소득세부담을 추정할 때에는 단순한 참고자료로만 활용하기로 한다.

마. 소득 평균에 대한 가정

다음 장에서는 소득분포에 대한 장기예측결과를 바탕으로 소득세의 실효세부담률이 어떠할지를 예측한다. 이때 설명변수로 사용되는 것은 소득분포의 특성을 규정하는 (자연대수)소득의 평균과 표준편차(또는 분산)이다. 소득평균에 대한 향후의 예측치는 외생적으로 결정된다고 가정한다. 경기변동이나 잠재성장률 추이 등을 감안하여

VI. 지니표면을 이용한 자연대수 소득의 분산 추정 155

야겠지만 본 연구에서는 명목소득의 연평균 증가율이 3%, 4%, 5%인 경우를 상정하여 각각 표준편차에 대한 세 가지 시나리오에 대응시켜 분석한다.

자연대수소득의 평균을 기준으로 하므로 명목소득증가율이 3%, 4%, 5%인 경우에는 각각의 자연대수 소득평균은 매년 각각 0.02956, 0.03922, 0.04879씩 증가한다.

<표 VI-13> 자연대수 소득평균에 대한 가정

	명목소득증가율 3%시	명목소득증가율 4%시	명목소득증가율 5%시
2002	17.16859	17.16859	17.16859
2003	17.19815	17.20781	17.21738
2004	17.22771	17.24703	17.26617
2005	17.25727	17.28625	17.31496
2006	17.28683	17.32547	17.36375
2007	17.31638	17.36469	17.41254
2008	17.34594	17.40391	17.46133
2009	17.37550	17.44313	17.51012
2010	17.40506	17.48236	17.55891
2011	17.43462	17.52158	17.60770
2012	17.46418	17.56080	17.65649
2013	17.49374	17.60002	17.70528
2014	17.52330	17.63924	17.75407
2015	17.55285	17.67846	17.80286
2016	17.58241	17.71768	17.85165
2017	17.61197	17.75690	17.90044
2018	17.64153	17.79612	17.94923
2019	17.67109	17.83534	17.99802
2020	17.70065	17.87456	18.04681
2021	17.73021	17.91378	18.09560
2022	17.75977	17.95300	18.14439
2023	17.78932	17.99222	18.19318
2024	17.81888	18.03145	18.24197
2025	17.84844	18.07067	18.29076
2026	17.87800	18.10989	18.33955
2027	17.90756	18.14911	18.38834
2028	17.93712	18.18833	18.43713
2029	17.96668	18.22755	18.48592
2030	17.99624	18.26677	18.53471
2031	18.02580	18.30599	18.58350
2032	18.05535	18.34521	18.63229

주: 연간명목소득증가율이 α (100 α %)라고 하면 자연대수소득의 평균은 매년 $\ln(1+\alpha)$ 씩 증가함.

VII. 소득세원 분포의 전망 및 시사점

1. 소득세 실효세율과 소득분포의 상관관계

가. 소득세 실효세율과 소득 표준편차의 상관관계

정치한 관계를 나타내지는 않지만 소득세 실효세부담률과 소득분포, 즉 소득의 평균과 표준편차 사이에는 통계적으로 일정한 상관관계가 존재한다고 할 수 있다. 주어진 소득세 체계하에서 경제 전체의 총소득이 일정할 경우 소득세의 실효세부담률은 소득분배상태가 완전균등할 경우에 가장 낮으며 세수규모도 가장 작다. 소득분배상태가 점차 불균등해질수록 실효세부담률이 상승하며 세수규모도 증가하기 시작한다. 궁극적으로 분배상태가 완전불균등하게 되면 소득세수가 최고조에 달하며 소득세의 평균실효세부담률도 극대화된다.

이와 달리 실업과 소득세 실효세부담률간에는 負(-)의 상관관계가 있다고 할 수 있다. 실업률이 상승하면 일반적으로 소득규모가 작아지므로 세수와 경제 전체의 평균실효세부담률이 하락함은 쉽게 이해가 된다. 물론 실업률이 하락하는 경우라면 경기도 좋지 않을 가능성이 높으므로 경기침체에 의한 실효세부담률 하락효과도 일조를 할 것으로 볼 수 있다. 물론 실업률이 하락하는 경우에는 이와 반대로 평균실효소득세부담률이 상승한다고 할 수 있다.

일반적으로 다른 조건이 동일하다면 소득이 증가할수록 소득세 실효세부담률 역시 상승한다. 이는 소득세가 기본적으로 누진세율 구조를 지니고 있기 때문이다.

다음의 <표 VII-1>은 1982~2002년 기간 동안 도시가계조사자료를 바탕으로 실효소득세부담률을 종속변수로 하고 실업률과 소득분배 상태를 나타내는 자연대수소득의 평균과 표준편차를 설명변수로 하여 회귀분석한 결과를 보여주고 있다. 다만 세법개정효과나 소득세의 과표양성화율 등과 같은 여타의 변수에 대해서는 이를 정량화 또는 정성화하여 설명변수로 전환하기 어려웠기 때문에 이를 대신하여 시간추세를 설명변수에 포함시켰다. 그리고 실효세부담 방정식의 구조조정 여부를 감안하여 1996년을 경계로 하는 수준더미를 포함시켜 분석하였다³⁶⁾.

시간추세를 설명변수에 포함시킨 이유를 간략히 설명하면 다음과 같다. 우리나라의 소득세법은 부정기적으로 개정이 이루어지므로 이를 직접 설명변수화하기는 어렵지만 장기적인 추세를 감안하면 대체로 4~6년 정도의 간격을 두고 실효세부담률이 크게 변화할 정도로 세법개정이 이루어져 왔다. 또한 최근에는 신용카드 소득공제와 영수증 복권제도 등을 통해 자영업자들에 대한 소득포착률이 상승하였다. 시간추세는 바로 이러한 추세적인 변화에 따른 효과를 요인에 구분없이 시계열적인 추세로 포착하는 역할을 담당하고 있다고 볼 수 있다.

회귀방정식의 형태는 수준(level) 자체를 선형화한 것과 각 변수

36) 물론 이 외에도 소득세의 실효세부담률을 설명해줄 수 있는 설명변수는 무수히 많을 것이다. 그러나 본 장에서 실효소득세부담률과 지니계수 등을 회귀분석을 통해 분석하는 이유는 분석 자체의 인과관계(causality) 분석 등에 있지 아니하고, 소득분배상태와 소득세 실효세부담률 사이의 상관관계를 고려하여, 향후 인구구조의 변화가 초래되었을 때 실험적으로 소득세의 실효세부담률에는 대체로 어떤 영향을 미칠 것인지에 대해 대강의 효과를 추정하기 위해서이다. 이는 전장에서 최근 21년간의 소득분포를 추정해 볼 때 경제위기의 예외적인 경우를 제외하고는 대체로 소득분포가 자연대수 정규분포를 따르는 것으로 추정되었으며 따라서 향후의 소득분포도 자연대수 정규분포를 따른다고 하면 지니계수 등으로 향후의 소득세 실효세부담률을 추정해볼 수 있을 것이기 때문이다.

(시간추세 제외)에 자연대수를 취하여 선형으로 모형화한 것의 두 가지를 상정하였다. 이 두 가지는 동시에 양립할 수 없는 모형이지만 분석의 편의상 모형의 설정오류(specification error)를 감안하지 않고 일단 양자 모두 타당한 것으로 받아들이기로 한다.

추정결과에 따르면 위에서 예측한 바와 같이 실효소득세부담률과 실업률 사이에는 負의 상관관계가 있는 것으로 추정되었다. LEVEL-2를 예로 들면, 실업률이 1%p 상승하면 소득세의 평균 실효세부담률은 0.20%p 정도 하락한다고 할 수 있다. 만약 LN-2의 모형이 맞다고 한다면 실업률이 1% 상승할 때 실효세부담률은 약 0.29% 하락한다고 할 수 있다³⁷⁾.

(자연대수) 소득의 표준편차와 실효세부담률 사이에는 위에서 예측한 바대로 正(+)의 상관관계가 있는 것으로 나타났다. LEVEL-2에 의하면 표준편차에 대한 계수 값이 1.828이므로 표준편차가 0.1p 상승하면 실효소득세부담률은 0.1828%p 상승한다고 할 수 있다. LN-1에 의하면, 자연대수소득의 표준편차의 계수는 0.494이므로 표준편차의 0.1p의 상승은 자연대수를 취한 실효소득세부담률을 0.0494p 상승시킨다.

그런데 자연대수 소득의 표준편차의 계수추정치에 대한 t-값은 대체로 절대수준이 1을 넘지 못하여 추정결과가 통계적으로 유의하지 않은 것으로 나타나고 있다. 이러한 이유는 분석대상 기간 동안 경제 위기 기간인 1997~1998년을 제외하고는 표준편차가 0.5 수준에서 변화의 폭이 매우 작다는 점을 지적할 수 있다. 만약 본 연구에서와 같이 어떤 설명변수의 변역이 매우 좁다면 비록 통계적으로 해당변수의 계수 값이 0이 아니라고 하더라도 대개의 경우 해당 변수의 계수 값은 통계적으로 유의하지 않은 것으로 나타나는 경우가 많다. 자연대수 소득의 표준편차의 경우가 바로 여기에 해당된다고 할 수 있다. 즉, 비록 계수추정치의 t-값이 작기는 하지만, 다른 여건이 동일하다면, 이론적으로 소득의 표준편차(분산)가 커질수록 실효세부담률이

37) 단위가 %p가 아님에 유의하기 바란다.

상승하는 것은 지극히 당연하다. 그러므로 본 연구에서는 비록 자연대수 소득의 표준편차에 대한 계수추정치의 t -값이 작더라도 표준편차를 설명변수에 계속 포함시켰음에 유의하기 바란다³⁸⁾.

시간추세에 대한 계수추정치는 모형에 관계없이 모두 양의 값을 가지는 것으로 추정되었다. 세법개정을 나타내는 변수를 명시적으로 표현하지 못하였기 때문에 세법개정효과도 시간추세에 반영되어 있다고 볼 수 있다. 물론 자영업자들의 소득포착률의 추세적 변화효과와 기타 소득세의 과세범위 확대 등에 의한 효과 등도 함께 포함되어 있다.

지난 20여년간의 실효소득세부담률 변화추이를 볼 때 상당한 정도의 부침은 있었지만 실효세부담률의 추세적 변화는 별로 크지 않았다고 할 수 있다. 1982~2002년 사이에 소득세 실효세부담률은 1983년에 최저인 2.59%, 1988년에 최고치인 4.2%를 기록하면서 큰 진폭을 그리기는 하였지만 1982년의 실효세부담률이 3.13%, 2002년에는 3.25%로 대동소이함을 볼 수 있다. 이는 소득증가 효과를 함께 고려할 경우, 그동안 수차례에 걸쳐 이루어져 왔던 세법개정을 통해 실질세부담의 증가가 완화되었음을 유추할 수 있다. 그러한 의미에서 볼 때 평균적으로 세법개정효과가 소득세 실효세부담률에 미친 효과는 별로 크지 않다고 하더라도 큰 무리는 없을 것으로 생각된다. 여기서 재미있는 것은 세법개정효과를 설명함에 있어 시간추세의 계수는 양수인 반면 자연대수 소득평균의 계수는 음수로 추정되었다는 점이다. 그동안의 세법개정은 매년 규칙적으로 이루어지지 않았다. 그리고 공제수준의 상향조정 등과 같은 세법개정은 명목소득이 일정수준 이상으로 증가하여 실질세부담 또는 세수가 일반적인 경상성장률을 크게 상회할 정도로 과다해지는 경우에 단행되는 경우가 많았다. 따라서 정확한 이유를 파악하기는 어렵지만

38) 이는 우리가 사전적으로 자연대수 소득의 표준편차가 소득세 실효세부담률 결정에 영향을 미친다는 것을 알고 있기 때문에 비로소 논리 전개가 가능한 것임에 유의할 필요가 있다.

소득변수는 주로 세법개정 등을 통한 세경감 효과를 포착하고 시간 추세는 세법개정효과를 제외한 나머지 요인에 의한 영향을 포착한 것으로 추측된다.

명확하게 파악하기는 어렵지만 경제위기 이후 신용카드 소득공제 제도 도입 및 신용카드 영수증 복권제도 등의 시행으로 자영업자들의 소득포착률이 상승하였다는 점이 여러 경로를 통해 밝혀지고 있다. 다른 조건이 동일하다면 소득포착률의 상승은 곧 소득세 실효세부담률의 상승을 가져다준다. 바로 이러한 것이 실효세부담률에 대한 시간추세의 계수가 양의 값을 가지게 된 하나의 원인으로 작용한 것으로 추정된다. 물론 1990년대 중반의 금융실명제 실시 등 꾸준한 소득세 과세범위의 확대추세도 시간추세의 계수 값이 양수를 가지게 하는 요인이 된 것으로 볼 수 있다.

<표 VII-1> 실효소득세율과 자연대수소득 표준편차간의 회귀분석 결과

	종속변수=소득			종속변수=자연대수 소득	
	LEVEL-1	LEVEL-2	LEVEL-3	LN-1	LN-2
상수	-350.84654 (-2.34349)	-275.79947 (-1.88936)	-238.65140 (-1.75832)	-106.92995 (-2.37131)	-72.29719 (-1.76253)
시간추세	0.18935 (2.36190)	0.14831 (1.90207)	0.13125 (1.78753)	0.05777 (2.39232)	0.03983 (1.79503)
D96	-0.50487 (-1.52614)	-0.49975 (-1.44967)	-	-1.15585 (-1.56405)	-
실업률	-	-0.20613 (-2.59792)	-	-	-
ln(실업률)	-0.93783 (-2.93417)	-	-0.94771 (-2.84970)	-0.28751 (-2.98644)	-0.29056 (-2.89092)
자연대수 소득 표준편차(s)	1.61561 (0.86806)	1.82774 (0.92171)	0.37291 (0.21411)	0.49439 (0.88192)	0.11080 (0.21049)
자연대수 소득 평균(m)	-1.39186 (-2.23023)	-1.01101 (-1.71365)	-1.14579 (-1.82612)	-0.41990 (-2.23378)	-0.34394 (-1.81379)
추정방법	OLS	OLS	OLS	OLS	OLS
R ²	0.44751	0.40026	0.36172	0.45899	0.37076
<조정-R ² >	<0.26335>	<0.20034>	<0.20216>	<0.27865>	<0.21345>
D.W.	1.93723	1.88793	1.73458	1.95638	1.75191
MSPE(%)	0.58501	0.62599	0.67935	0.42922	0.50035
MAPE(%)	5.61448	5.99009	6.24140	4.81392	5.38405

- 주: 1. 1982~2002년의 도시가계조사자료를 사용하여 추정한 결과임.
 2. () 안은 t-값임.
 3. 시간추세는 1982, ..., 2002임.
 4. D96은 수준 더미로 1996년부터 1의 값을 가지고 나머지 연도에는 0의 값을 가짐.
 5. MSPE(mean squared percentage error)와 MAPE(mean absolute percentage error)는 각각 전년도 대비 증감률을 기준으로 한 평균자승오차와 평균절대오차를 나타냄.

나. 소득세 실효세율과 지니계수의 상관관계

식 (2) 또는 식 (3)의 관계와 앞의 가節에서 살펴본 바를 놓고 볼 때 소득세 실효세부담률과 지니계수간의 상관관계도 표준편차를 기준으로 한 것과 유사할 것으로 예상된다. 아래의 <표 VII-2>에서는 참고로 실효소득세율과 지니계수간의 상관관계를 추정해보았다.

<표 VII-2> 실효소득세율과 지니계수간의 회귀분석 결과

	선형모형		자연대수 선형모형	
	LEVEL-G-1	LEVEL-G-2	LN-G-1	LN-G-2
상수	-182.37016 (-2.69333)	-91.99364 (-1.94729)	-52.52390 (-2.78411)	-29.87866 (-2.22587)
시간추세	0.09052 (2.73023)	0.04653 (1.99410)	0.02833 (2.88571)	0.01646 (2.37608)
D96	3.22145 (1.25417)	-0.59023 (-1.91507)	-1.98098 (-1.65039)	-0.20143 (-2.19205)
실업률	-0.17429 (-2.78058)	-0.19342 (-3.02376)	-0.05371 (-2.89591)	-0.06165 (-3.34129)
지니계수	23.43910 (3.00219)	12.56928 (2.34463)	1.85241 (3.10014)	1.11404 (2.63704)
지니계수×D96	-16.04919 (-1.52254)	-	-1.25090 (-1.47555)	-
추정방법	C-O	C-O	C-O	C-O
R ²	0.91563	0.88905	0.94073	0.92752
<조정-R ² >	<0.88750>	<0.86132>	<0.92097>	<0.90940>
D.W.	1.68268	1.73135	1.60683	1.65060
MSE(%)	0.58664	0.63977	0.42630	0.46476
MAD(%)	5.95624	6.45483	5.14083	0.556750

- 주: 1. 1982~2002년의 도시가계조사자료를 사용하여 추정한 결과임.
 2. () 안은 t-값임.
 3. 시간추세는 1982, ... , 2002임.
 4. D96은 수준 더미로 1996년부터 1의 값을 가지고 나머지 연도에는 0의 값을 가짐.
 5. 추정방법 중 C-O는 오차항의 1차 자기상관을 제거하는 방법 중 Cochrane -Orcutt 방법을 말함.
 6. MSE(mean squared error)와 MAD(mean absolute deviation)는 각각 전년도 대비 증감률을 기준으로 한 평균자승오차와 평균절대오차를 나타냄.
 7. C-O 기법으로 추정할 경우 R²와 조정-R²는 자기상관을 제거하기 위해 자기상관계수를 곱하여 차감한 회귀방정식상의 결정계수를 나타냄.

2. 소득세 실효세부담률의 예측

본 절에서는 <표 VII-1>의 다섯 가지 회귀분석 결과를 바탕으로 향후 30년간의 소득세 실효세부담률을 추정해보았다. 실효세율을 추정하기 위해서는 실업률, (자연대수)소득의 평균과 표준편차가 필요하다. 이 가운데 실업률은 외생적으로 3.5%에서 고정된다고 가정하였으며, 나머지는 제VI장의 <표 VI-13>과 <표 VI-14>에 제시된 바를 기준으로 하였다.

추정결과는 <표 VII-3>과 <표 VII-4>에 제시되었다. LN-1에 의한 것을 제외하고는 나머지 4개의 회귀분석 결과에 대응한 예측결과가 서로 대동소이하게 나타났다. 그렇지만 추세적으로는 시간이 경과함에 따라 소득세 실효세부담률이 계속 상승하는 것으로 예측되었다. 이는 선형회귀분석의 특성상 내분(interpolation)의 경우와 달리 일정 기간을 초과하여 외분(extrapolation)하는 경우에는 실제보다 과장되게 추정하는 경우가 많다는 점에 유의할 필요가 있다. 더욱이 본 연구에서 회귀분석시에 사용한 자료기간은 1982~2002년의 21년간에 불과하기 때문에 향후 30년을 예측하는 데에는 다소 무리가 있다. 그러나 향후 10년 또는 15~20년 이내의 기간에 대한 예측치는 나름대로 의미를 지닐 수 있는 수준인 것으로 사료된다.

VII. 소득세원 분포의 전망 및 시사점 163

<표 VII-3> 소득세 실효세부담률 추정결과 I

(단위 : %)

	LEVEL 1			LEVEL 2			LEVEL 3		
	중간-I	중간-II	상한	중간-I	중간-II	상한	중간-I	중간-II	상한
2003	3.63318	3.62583	3.61556	3.59442	3.59154	3.58532	3.39287	3.38320	3.37294
2004	3.79054	3.77584	3.75531	3.72319	3.71746	3.70501	3.49236	3.47304	3.45252
2005	3.94789	3.92586	3.89504	3.85196	3.84337	3.82469	3.59185	3.56287	3.53208
2006	4.10525	4.07587	4.03479	3.98075	3.96928	3.94438	3.69134	3.65270	3.61166
2007	4.26262	4.22586	4.17452	4.10953	4.09517	4.06405	3.79085	3.74253	3.69122
2008	4.41997	4.37588	4.31426	4.23830	4.22109	4.18374	3.89034	3.83236	3.77080
2009	4.57732	4.52589	4.45399	4.36707	4.34700	4.30342	3.98983	3.92219	3.85036
2010	4.73469	4.67589	4.59374	4.49585	4.47290	4.42311	4.08932	4.01202	3.92994
2011	4.89204	4.82588	4.73347	4.62462	4.59879	4.54278	4.18881	4.10184	4.00950
2012	5.04939	4.97590	4.87321	4.75339	4.72471	4.66247	4.28830	4.19168	4.08908
2013	5.20674	5.12591	5.01294	4.88216	4.85062	4.78214	4.38780	4.28151	4.16864
2014	5.36411	5.27592	5.15269	5.01095	4.97653	4.90183	4.48729	4.37134	4.24822
2015	5.52147	5.42592	5.29242	5.13973	5.10242	5.02151	4.58679	4.46117	4.32778
2016	5.67882	5.57593	5.43216	5.26850	5.22834	5.14120	4.68628	4.55100	4.40736
2017	5.83617	5.72594	5.57189	5.39727	5.35425	5.26087	4.78578	4.64083	4.48692
2018	5.99354	5.87596	5.71164	5.52606	5.48016	5.38056	4.88527	4.73067	4.56650
2019	6.15089	6.02595	5.85137	5.65483	5.60605	5.50024	4.98476	4.82050	4.64606
2020	6.30824	6.17597	5.99112	5.78360	5.73197	5.61993	5.08425	4.91033	4.72564
2021	6.46559	6.32598	6.13085	5.91237	5.85788	5.73960	5.18374	5.00016	4.80521
2022	6.62296	6.47599	6.27059	6.04115	5.98379	5.85929	5.28324	5.08999	4.88478
2023	6.78033	6.62599	6.41032	6.16993	6.10968	5.97896	5.38274	5.17982	4.96435
2024	6.93768	6.77599	6.55007	6.29870	6.23559	6.09865	5.48223	5.26964	5.04392
2025	7.09503	6.92600	6.68980	6.42747	6.36150	6.21833	5.58172	5.35947	5.12349
2026	7.25240	7.07601	6.82954	6.55626	6.48741	6.33802	5.68122	5.44931	5.20306
2027	7.40975	7.22602	6.96927	6.68503	6.61332	6.45769	5.78071	5.53914	5.28263
2028	7.56710	7.37602	7.10902	6.81380	6.73922	6.57738	5.88020	5.62897	5.36220
2029	7.72445	7.52603	7.24875	6.94257	6.86513	6.69706	5.97969	5.71880	5.44177
2030	7.88182	7.67605	7.38850	7.07136	6.99104	6.81675	6.07919	5.80863	5.52134
2031	8.03917	7.82606	7.52823	7.20013	7.11695	6.93642	6.17868	5.89847	5.60091
2032	8.19653	7.97605	7.66797	7.32891	7.24285	7.05611	6.27818	5.98829	5.68048

<표 VII-4> 소득세 실효세부담률 추정결과 II

(단위 : %)

	LN 1			LN 2		
	중간- I	중간- II	상한	중간- I	중간- II	상한
2003	1.34913	1.34617	1.34203	3.52545	3.51522	3.50441
2004	1.41568	1.40950	1.40083	3.63387	3.61282	3.59062
2005	1.48553	1.47580	1.46220	3.74563	3.71314	3.67896
2006	1.55882	1.54523	1.52627	3.86082	3.81623	3.76947
2007	1.63573	1.61791	1.59314	3.97957	3.92218	3.86220
2008	1.71643	1.69402	1.66295	4.10196	4.03108	3.95722
2009	1.80111	1.77371	1.73581	4.22811	4.14301	4.05457
2010	1.88997	1.85714	1.81187	4.35814	4.25802	4.15432
2011	1.98321	1.94449	1.89125	4.49217	4.37624	4.25652
2012	2.08106	2.03596	1.97412	4.63032	4.49775	4.36125
2013	2.18372	2.13173	2.06061	4.77272	4.62263	4.46854
2014	2.29147	2.23201	2.15090	4.91950	4.75098	4.57847
2015	2.40452	2.33700	2.24513	5.07081	4.88288	4.69111
2016	2.52315	2.44694	2.34351	5.22676	5.01846	4.80652
2017	2.64763	2.56204	2.44618	5.38750	5.15780	4.92477
2018	2.77826	2.68257	2.55337	5.55319	5.30100	5.04593
2019	2.91532	2.80874	2.66524	5.72398	5.44818	5.17006
2020	3.05915	2.94087	2.78202	5.90001	5.59945	5.29726
2021	3.21007	3.07921	2.90391	6.08146	5.75492	5.42758
2022	3.36845	3.22407	3.03115	6.26849	5.91470	5.56111
2023	3.53465	3.37571	3.16395	6.46129	6.07892	5.69792
2024	3.70903	3.53450	3.30259	6.66000	6.24768	5.83810
2025	3.89201	3.70077	3.44728	6.86482	6.42115	5.98172
2026	4.08404	3.87486	3.59833	7.07595	6.59943	6.12889
2027	4.28553	4.05713	3.75598	7.29356	6.78266	6.27967
2028	4.49695	4.24797	3.92056	7.51786	6.97098	6.43416
2029	4.71880	4.44780	4.09233	7.74906	7.16453	6.59245
2030	4.95163	4.65703	4.27164	7.98738	7.36345	6.75464
2031	5.19591	4.87610	4.45879	8.23302	7.56790	6.92081
2032	5.45227	5.10546	4.65416	8.48625	7.77801	7.09108

VIII. 결 론

본 연구는 개략적인 수준에서나마 최초로 미시분석에 기초한 소득분포를 추정하고 이를 바탕으로 소득세의 향후 장기 추이를 전망하였다는 점에서 고령화 문제에 대응하여 최근 활발하게 진행되는 중장기재정 연구에 기여하였다고 사료된다. 그러나 이러한 연구의 최종결과보다도 그 결과를 도출하는 과정에서 밝혀낸 새로운 사실들에 더 큰 의의를 부여할 수 있다고 판단된다. 즉, 소득세뿐만 아니라 소득에 기반하여 징수되는 재정수입이 관련된 제반 연구들이 공통적으로 부딪혀왔던 소득분포 전망의 문제에 대해 본 연구는 상당히 의미있고 유용한 해결책을 제시하였다고 사료된다. 분포의 몇 가지 계수가 아니라 분포 자체를 전망할 수 있는 새로운 분석틀을 제시하였으며, 이러한 틀을 이용하여 분포 전망에 필요한 계수들을 구체적으로 추정하여 제시하였다.

보다 구체적으로 연구의 의의를 정리하면, 뚜렷한 속성이 잘 파악되어 있지 않던 우리나라의 소득분포에 대하여 1982~2002년 사이의 도시가계자료를 이용하여 몇 가지의 정형화된 특성(stylized fact)을 새로이 발견하여 제시했다는 점을 꼽을 수 있다.

첫째, 우리나라 도시가구의 소득이 연도별 총소득에서나 연도·연령별 소득에서나 자연대수 정규분포와 유사한 수준을 넘어 실제로 자연대수 정규분포로 간주해도 무방하다는 사실을 보였다. 그러므로 단지 몇 개의 변수로써 소득분포 전체를 파악할 수 있으며, 정책모의실험이나 경제모형에 소득분포를 쉽게 대입하여 활용할 수 있다는 유용성이 있다. 또한 도시가구의 소득이 정규분포임을 보이는 과정에서 1997년 연말의 외환위기가 도래하기 이전에 우리 경제

의 소득분포에 이미 심각한 붕괴현상이 있었음을 보였다는 점도 연구의 성과로 꼽을 수 있겠다.

둘째, 자연대수를 취하였을 경우 횡단면에서의 연령별 상대소득 분포가 실질적으로 불변임을 보임으로써, 자료의 제약에도 불구하고 세대 생애 연령-소득 곡선을 추정할 수 있음을 보였다.

셋째, 소득의 형평성을 측정하는 지니계수와 소득분포와의 관계를 구체적으로 규명하고 계량화하였으며, 이를 이용하여 소득분포의 분산을 간접적으로 추정할 수 있음을 보였다.

본 연구의 한계에 대하여 언급하자면, 경험적으로 안정적인 규칙성을 찾아 정형화된 특성(stylized fact)으로 제시하기는 하였으나, 그러한 규칙성을 설명할 수 있는 인과관계에 대한 분석이 미흡하였다는 점을 지적할 수 있다. 또한 본고에서 제시한 횡단면 연령-소득 곡선 및 세대 생애 연령-소득 곡선이 평균자연대수소득의 형태로 되어 있으므로, 이를 다시 평균소득이나 소득분포로 환산하기 위해서는 자연대수 소득의 평균뿐만 아니라 분산 역시 추정하여 예측할 필요가 있다는 점도 지적할 수 있다.

본 연구에서는 소득분산 전망을 간접적인 방법으로밖에 다루지 못하였으나, 지니계수와 자연대수 정규분포 가구소득의 관계를 처음으로 규명했다는 성과를 대신 꼽을 수 있다.

마지막으로, 자연대수 소득의 평균과 분산에 더하여 연도별 가구 주 숫자의 연령별 분포의 특성까지 밝히면 연령-소득의 결합확률밀도함수를 제시할 수도 있었으나, 본고에서 이를 다루지 못한 점은 저자들에게 큰 아쉬움으로 남는다.

끝으로 본 연구에 이은 추후 연구 주제를 살펴봄으로써 본고를 마무리하고자 한다. 첫째, 방금 언급한 바와 같이 연령별 자연대수 소득분포의 분산까지 밝히면, 횡단면과 세대의 연령-소득 분포를 본고에서 제시한 바와 같이 평균소득을 기준으로 하여 밝힐 수 있을 뿐만 아니라, 각 소득분위별로 연령-소득의 양상을 추정하여 소

득분배의 변화에 대한 추가적인 시사점을 찾을 수 있다. 이는 특히 공적부조의 재정소요 파악 등 소득분위 등이 중요시되는 저소득층의 문제에 활용될 여지가 크다고 하겠다. 둘째, 분산과 가구주의 연령별 확률밀도까지 규명된 자연대수 가구소득의 분포를 앞으로 밝힌 후, 이를 소득세원의 향후 변화에 대한 예측분석에도 활용할 수 있다.

참고문헌

- 성명재, 『소비세에 대한 도·농별 소득계층별 세부담 및 역진도 실태 분석에 관한 연구』, 『재정논집』 제8집, 한국재정학회, 1993.
- _____, 『소비세 부담분포와 고세율 상품시장의 정상화에 관한 연구』, 연구보고서 93-02, 한국조세연구원, 1993.
- _____, 『외환위기 발생후 2년간의 소득·소비패턴 및 개인세부담의 변화 분석』, 연구보고서 00-02, 한국조세연구원, 2000.
- _____, 『소득분배 변화추이와 결정요인 분석: 도시가구를 중심으로』, 연구보고서 01-01, 한국조세연구원, 2001.
- _____, 『조세정책의 소득재분배 효과 분석에 관한 연구: 도시가계조사자료를 중심으로』, 연구보고서 02-01, 한국조세연구원, 2002.
- 성명재·전영준, 『경제위기 1년간 소득세·소비세 부담분포의 변화와 조세정책방향』, 연구보고서 99-03, 한국조세연구원, 1999.
- 전영준, 『인구구조 변동과 국민연금: 세대별 후생분석을 중심으로』, 『한국경제의 분석』, 제3권 제1호, 한국금융연구원, 1997.
- 전영준·김종면, 『사회보장정책의 경제적 효과 분석』, 연구보고서 01-07, 한국조세연구원, 2001.
- 전영준·성명재·노영훈·박종규, 『경제위기 1년의 조세정책 평가와 향후 정책방향』, 제33회 조세의 날 기념 심포지엄 발표자료, 한국조세연구원, 1999.
- Australian Bureau of Statistics, *Income Distribution 1996-97*, 1998.
- _____, *Income Distribution 1997-98*, 1999.

- _____, *Income Distribution 1999-2000*, 2001.
- Burniaux, Jean-Marc, Thai-Thanh Dang, Douglas Fore, and Michael Forster, "Income Distribution and Poverty in Selected OECD Countries," Economic Department Working Paper No. 189, ECO/WKP(98)2, 1998.
- Census of Population Office, "Household Income Growth and Distribution," Advance Data Release No. 7, Singapore Department of Statistics, 2001.
- Da Silva, J.G.C., "The Analysis of Cross-Sectional Time Series Data," Ph.D dissertation, Department of Statistics, North Carolina State University, 1975.
- Devroye, Robert J. and László Györfi, *Nonparametric Density Estimation*, John Wiley and Sons, 1985.
- Förster, Michael F. (Assisted by Michelle Pellizzari), "Trends and Driving Factors in Income Distribution and Poverty in the OECD Area," Labour Market and Social Policy-Occasional Papers No. 42, DEELSA/ELSA/WD (2000)3, OECD, 2000.
- Goldberger, Arthur S., *A Course in Econometrics*, Harvard University Press, 1991.
- Greene, William H., *Econometric Analysis*, Fourth Edition, Prentice Hall International, Inc., 2000.
- Härdle, W., *Applied Nonparametric Regression*, Econometric Society Monographs No. 19, Cambridge University Press, 1990.
- Harris, Tim, "The Effects of Taxes and Benefits on Household Income, 1998-99," Economic Trends No. 557, pp. 45-83, (UK) Office for National Statistics, April 2000.

- Jarque, C. M., and Bera, A. K., "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals," *Economic Letters*, 6, pp. 255-259, 1980.
- Jones Jr., Arthur F. and Daniel H. Weinberg, "The Changing Shape of the Nation's Income Distribution 1947-1998," Economics and Statistics Administration, the Census Bureau, U.S. Department of Commerce, June 2000.
- Lakin, Caroline, "The Effects of Taxes and Benefits on Household Income, 1999-2000," *Economic Trends* No. 569, pp. 35-79, (UK) Office for National Statistics, April 2001.
- _____, "The Effects of Taxes and Benefits on Household Income, 2000-2001," (UK) Office for National Statistics, 2002.
- _____, "The Effects of Taxes and Benefits on Household Income, 2001-2002," (UK) Office for National Statistics, 2003.
- Manski, Charles, F., *Analog Estimation Methods in Econometrics*, Chapman and Hall, 1988.
- Ministry of Health, Labour and Welfare, *Annual Report on Health and Welfare 1999*, Japan, 2000.
(http://www.mhlw.go.jp/search/mhlwe/wp/wp_hw/vol1/p1c2s2.html)
- Parks, R. W., "Efficient Estimation of a System of Regression Equations when Disturbances Are Both Serially and Contemporaneously Correlated," *Journal of American Statistical Association*, 62, pp. 500-509, 1967.
- Rao, B. L. S. Prakasa, *Nonparametric Functional Estimation*, Academic Press, 1983.

- Royston, J. P., "An Extension of Shapiro and Wilk's W Test for Normality to Large Samples," *Applied Statistics*, 131, pp. 115-124, 1982.
- Serfling, Robert J., *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, 1980.
- Shapiro, S. S. and Wilk, M. B., "An Analysis of Variance Test for Normality(Complete Samples)," *Biometrika*, 52, pp. 591-611, 1965.
- Silverman, B. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, 1986.
- Statistics Canada, *Income in Canada*, 1998.
- Statistics New Zealand, *New Zealand Now Incomes*, 1999.
- Stephens, M. A., "EDF Statistics for Goodness of Fit and Some Comparisons," *Journal of the American Statistical Association*, 69, pp. 730-737, 1974.
- Sung, Myung Jae, "Nonparametric Average Median Derivatives Estimation of Index Coefficients," Three Essays on Nonparametric Regression: Estimation of Index Coefficients and Preference Parameters, and Tests of Independence, Doctoral Dissertation, University of Wisconsin-Madison, 1992.
- US Census Bureau, *Money Income in the United States*, 각 연도.

부록 I. 근로자가구의 근로소득세 부담 분포

<부표 1-1> 근로자가구의 근로소득세 부담액 분포(무직가구 제외)

(단위 : 천원)

세부담액	1	2	3	4	5	6	7	8	9	10	평균
1982	1	5	13	23	28	43	60	96	146	418	78
1983	0	3	10	17	28	38	50	78	152	463	79
1984	1	8	14	26	36	46	71	101	199	550	98
1985	2	12	23	37	41	58	76	123	215	688	126
1986	4	18	34	41	62	78	110	175	293	759	155
1987	8	29	47	68	79	123	167	248	432	972	208
1988	20	53	83	114	171	206	291	395	663	1500	330
1989	10	50	72	119	146	224	298	425	661	1517	332
1990	21	56	87	151	160	248	314	446	722	1324	339
1991	19	58	83	110	161	208	301	458	666	1443	336
1992	33	87	124	184	262	355	503	747	1047	1999	530
1993	29	97	127	170	226	298	436	533	797	1820	438
1994	32	105	140	196	252	360	539	695	1090	2012	539
1995	39	129	172	242	371	504	735	956	1294	2444	679
1996	13	83	130	229	336	500	686	916	1406	2768	719
1997	0	31	103	178	287	397	633	923	1324	2782	684
1998	0	3	30	90	168	303	440	735	1181	2514	567
1999	1	6	43	96	177	268	425	703	1213	2441	565
2000	1	22	79	162	258	395	553	1000	1595	3032	687
2001	5	35	110	193	288	420	652	1188	2010	3859	912
2002	2	33	105	191	268	504	788	1236	2027	3798	925

<부표 1-2> 근로자가구의 근로소득세 실효세부담률 분포
(무직가구 제외)

(단위 : %)

실효세율	1	2	3	4	5	6	7	8	9	10	평균
1982	0.06	0.24	0.53	0.8	0.89	1.2	1.45	1.97	2.46	4.64	2.03
1983	0.02	0.16	0.38	0.56	0.79	0.95	1.11	1.47	2.35	4.67	1.85
1984	0.06	0.33	0.47	0.73	0.91	1.03	1.4	1.72	2.76	5.09	2.1
1985	0.12	0.42	0.71	1.01	1	1.26	1.46	2.04	2.93	6.1	2.54
1986	0.2	0.58	0.91	0.99	1.3	1.46	1.83	2.5	3.47	6.03	2.74
1987	0.32	0.82	1.12	1.4	1.43	2	2.38	3.02	4.38	6.66	3.19
1988	0.63	1.22	1.6	1.93	2.56	2.75	3.43	4.08	5.61	8.66	4.22
1989	0.26	0.91	1.13	1.63	1.81	2.47	2.91	3.6	4.63	7.17	3.49
1990	0.44	0.83	1.12	1.71	1.63	2.26	2.55	3.19	4.35	5.51	2.98
1991	0.31	0.71	0.86	1	1.32	1.53	1.99	2.65	3.32	5.14	2.42
1992	0.48	0.91	1.11	1.44	1.85	2.24	2.86	3.78	4.48	6.16	3.25
1993	0.39	0.94	1.04	1.23	1.46	1.75	2.29	2.49	3.2	5.08	2.51
1994	0.38	0.89	1.02	1.26	1.45	1.84	2.47	2.81	3.76	5.01	2.67
1995	0.42	0.96	1.09	1.36	1.86	2.27	2.94	3.37	3.91	5.46	2.98
1996	0.13	0.57	0.75	1.16	1.5	2	2.44	2.9	3.83	5.41	2.78
1997	0	0.24	0.6	0.88	1.25	1.54	2.17	2.78	3.38	5.11	2.54
1998	0	0.03	0.22	0.54	0.85	1.34	1.71	2.46	3.33	4.89	2.38
1999	0.01	0.04	0.26	0.5	0.81	1.09	1.53	2.2	3.2	4.5	2.16
2000	0.01	0.14	0.44	0.77	1.06	1.44	1.77	2.79	3.72	4.64	2.38
2001	0.04	0.21	0.57	0.86	1.14	1.47	2	3.14	4.47	5.88	2.96
2002	0.02	0.2	0.52	0.79	0.99	1.62	2.25	3.04	4.14	5.51	2.81

<부표 1-3> 근로자가구의 근로소득세 실효세부담 점유비
 분포(무직가구 제외)

(단위 : %)

세수비중	1	2	3	4	5	6	7	8	9	10	계
1982	0.12	0.69	1.88	3.02	3.39	5.26	7.32	12.68	17.87	47.77	100
1983	0.05	0.44	1.38	2.34	3.92	4.83	6.43	9.62	17.04	53.97	100
1984	0.11	1.02	1.57	2.6	3.36	4.68	6.4	10.08	19.4	50.77	100
1985	0.2	0.92	1.93	2.93	3.45	4.75	4.95	8.2	19	53.66	100
1986	0.27	1.31	2	2.92	4.36	5.18	6.58	9.79	17.51	50.09	100
1987	0.41	1.39	2.46	3.56	3.6	6.33	7.3	12.08	19.27	43.59	100
1988	0.64	1.71	2.48	3.78	5.44	6.64	8.04	11.18	18.94	41.16	100
1989	0.33	1.7	2.25	3.43	4.44	6.47	8.63	13.04	18.49	41.21	100
1990	0.68	1.69	2.68	4.8	4.95	7	9.13	11.62	19.14	38.3	100
1991	0.59	1.82	2.62	3.28	4.98	6.23	8.87	12.77	17.69	41.15	100
1992	0.63	1.71	2.33	3.52	4.87	6.58	8.93	14.5	20.08	36.85	100
1993	0.69	2.3	2.97	4.07	5.16	6.49	9.65	12.34	17.55	38.77	100
1994	0.61	1.89	2.75	3.48	4.68	6.58	9.81	13.31	20.11	36.79	100
1995	0.58	2	2.58	3.3	5.53	7.4	11.73	13.66	18.03	35.19	100
1996	0.19	1.2	1.76	3.11	4.19	6.76	9.33	13.32	21.17	38.97	100
1997	0	0.49	1.5	2.89	4.42	5.8	9.41	14.59	20.49	40.41	100
1998	0	0.05	0.58	1.63	3.22	5.41	7.74	13.33	21.98	46.05	100
1999	0.01	0.1	0.72	1.57	2.7	4.71	7.64	13.35	24.61	44.59	100
2000	0.01	0.33	1.12	2.48	3.81	5.61	8.19	13.79	24.7	39.97	100
2001	0.06	0.39	1.17	2.07	2.89	4.3	7.44	13.5	21.98	46.19	100
2002	0.02	0.35	1.08	2.13	2.7	5.39	8.04	14.04	25.09	41.17	100

부록 II. 커널 추정기법의 소개

1. 비모수적 확률분포함수 추정량

어떤 연속미분가능한(continuously differentiable) 확률변수에 대한 확률밀도함수(probability density function)의 추정량(estimator)과 그에 대한 기대치를 각각 $f_N(x)$ (또는 $\hat{f}(x)$)와 $\check{f}(x)$ 라고 하자. 확률밀도함수를 추정하는 방법에는 여러 가지가 있는데, 본 연구에서는 커널(kernel)을 이용한 비모수적(nonparametric) 추정방법에 의한 추정량을 사용하기로 한다. 어떤 점 x 에서의 확률밀도함수에 대한 커널추정량 $f_N(x)$ (또는 $\hat{f}(x)$)은 다음과 같이 정의할 수 있다.

$$f_N \equiv \hat{f}(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{X_i - x}{h}\right) \quad (1)$$

여기서 N 은 표본크기(sample size), $K(\cdot)$ 는 커널, h 는 x 의 근방 또는 주변(cell)의 폭(bandwidth) 또는 창넓이(window size)를 나타낸다. 커널추정량이 일치성을 가지기 위해서 h 는 표본크기가 무한대로 커질수록 0에 접근하는 특성을 가져야 한다.

확률밀도함수에 대한 비모수적 추정방법의 하나로 커널추정량이 많이 사용된다. 위의 식 (1)과 같이 소득을 나타내는 정의구역(domain) 내에 있는 어떤 점 x 에서의 확률밀도함수 추정치를 정의할 수 있다. 기본적으로 식 (1)과 같이 확률밀도함수에 대한 추정치를 정의하는 배경에는 확률밀도함수가 순탄(smooth)하고 연속미분가능(continuously differentiable)하다는 정규성 조건(regularity condition)을 충족시켜준다는 가정하에서 어떤 점 x 의 근방

(neighbor)에서는 x 와 유사한 분포특성을 가짐으로써 공통적인 특성을 많이 공유하고 있는 반면 x 에서 멀리 떨어질수록 x 의 분포특성과 다른 특성을 더 많이 가지게 됨으로써 동질성이 작아진다는 점에 착안하고 있다.

커널 $K(\cdot)$ 는 일종의 가중치와 같은 역할을 수행하는 것으로서 어떤 점 x 에 가까운 표본관측치에 대해서는 높은 가중치를 부여하고, x 에서 멀리 위치한 관측치에 대해서는 낮은 가중치를 부여하는 역할을 수행한다. 가중치를 부여함에 있어 h 는 가중치 부여강도를 결정해주는 요소이다. h 의 크기를 표본의 크기에 맞추어 적절히 조정해주면 커널추정량의 일치성과 점근적 정규분포 특성을 보장해줄 수 있다.

2. 커널추정량의 점근적 특성 보장을 위한 가정

식 (1)과 같이 연속확률밀도함수에 대한 추정량을 정의할 때 이 추정량이 일치성과 점근적 정규분포의 특성(asymptotic normality)을 가지기 위해서는 다음과 같은 가정이 필수적이다.

- 가정 I.** (1) N 이 무한대로 커질 때 h 는 0에 수렴($h \rightarrow 0$)
 (2) N 이 무한대로 커질 때 Nh 는 무한대로 발산($Nh \rightarrow \infty$)
 (3) N 이 무한대로 커질 때 Nh^3 는 0에 수렴($Nh^3 \rightarrow 0$)

만약 표본의 크기 N 이 고정되어 있다고 하면 h 가 0에 가까워질수록 식 (1)은 표본 자체의 상대도수에 가까워짐으로써 삐죽삐죽한 톱니 모양의 형태로 확률밀도함수의 함수형태가 추정되는 한편, h 가 무한대로 커질수록 확률밀도함수는 수평선 즉, 균등분포(uniform distribution)의 확률밀도함수와 유사해지게 된다. 그러므로 h 의 값이 어떤 특성을 가지느냐에 따라 확률밀도함수에 대한 추정결과는 크게 영향을 받는다.

가정 I 은 확률밀도함수에 대한 커널추정량이 일치성과 점근적 정규분포를 가지기 위해 필요한 h 의 점근적 특성을 가정하고 있다. h 의 크기는 표본의 크기에 따라 상대적인 가치를 지닌다. 즉, h 의 절대값이 일정하게 주어져 있다고 하면 N 이 작을 때에는 h 의 상대값이 작게 평가되지만 반대로 N 이 큰 값을 가지게 되면 h 의 상대값은 크게 평가할 수 있다.

h 는 추정하고자 하는 특정한 점 x 에 대한 근방(neighbor)의 폭(bandwidth)으로 볼 수 있는데, h 가 작을수록 x 의 근방에서는 점 x 와 동일한 분포특성을 많이 가지게 된다. 따라서 h 가 작을수록 커널추정치는 x 에서의 확률밀도함수의 값에 가까워진다. 그렇지만 이와 반대로 근방에서의 표본관측치가 너무 적다면 소표본특성(micronumerosity³⁹)으로 인해 추정오차가 크게 나타난다. 그러므로 근방에서의 표본 특성도 잘 반영하고 표본의 관측치도 충분히 큰 값을 얻기 위해서는 표본의 크기가 커질수록 h 는 작아지되 그 속도가 너무 빠르지 않게 함으로써 근방에 속하게 되는 근방표본치의 수도 전체 표본의 크기가 무한대로 갈수록 서서히 무한대로 커지도록 하여야 한다. 바로 가정 I 의 (1)과 (2)가 이러한 조건을 나타낸다. 그리고 가정 I 의 (3)은 점근적 정규분포를 추정함에 있어 점근편차(asymptotic bias)가 점근적으로 특이확률분포를 가지지 않고 0으로 흡수(degenerate)되어 버리도록 하는 조건이다.

- 가정 II.** (1) 소득을 나타내는 변량 $X \in \mathbb{X} \subset \mathbb{R}^1$ 은 양의 값을 가지는 연속확률변수이다.
- (2) 표본관측치 $X_i, i=1,2,\dots,N$ 은 모집단으로부터 독립적·동질적으로 추출(independently and identically drawn)되었다.

39) 이에 대한 자세한 내용은 Goldberger(1991)의 249쪽을 참조하기 바란다.

가정 III. (1) X 는 한정적(bounded)인 르베그밀도함수(Lebesgue density)를 가진다.

모든 $X \in \mathbb{X}$ 및 어떤 f^* 에 대해 $f(X) < f^* < \infty$

(2) $f: \mathbb{R}^1 \rightarrow \mathbb{R}^1$ 은 4번 연속미분가능하고 도함수(derivatives)는 한정적이고 다음의 립쉬츠 조건(Lipschitz condition)을 만족한다.

$$0 < f'(x) < \bar{f}' < \infty, X \in \mathbb{X}$$

$$|f(X+a) - f(X)| \leq \|a\|, \text{ 단 } \|\cdot\| \text{ 은 유클리드 놈(Euclidean norm)임.}$$

가정 II와 가정 III은 확률변수 X 에 대한 정규성을 보장해주는 조건이다. 이 가운데 특히 가정 III은 커널추정량의 일치성과 점근적 정규분포를 유도하는 과정에서 정규성을 보장하기 위해 필요한 가정이다. 다만 가정 III의 (2)에서는 f 를 4번 연속미분가능하다고 가정하였지만 2번 미분가능하더라도 충분하다. 미분가능횟수를 4회로 한 이유는 다음 항에서 커널추정량에 대한 기대치를 전개함에 있어 독자들의 이해를 돕기 위해 미분차수를 늘려주기 위함이다.

가정 IV. 커널 집합 $\mathbb{K}_{1,2}$ 는 측정가능·한정·실가함수(measurable bounded real-valued functions)인 $K: \mathbb{R}^1 \rightarrow \mathbb{R}^1$ 의 집합으로서 다음의 특성을 지닌다.

$$(1) |K(\cdot)| < K^0 < \infty$$

$$(2) \int K(u)du = 1$$

$$(3) \int uK(u)du = 0$$

$$(4) \int |u|^i \cdot |K(u)| du < \infty, i \geq 2$$

$$(5) K(-u) = K(u)$$

$$(6) \int K^2(u)du \leq \overline{K}^0 < \infty$$

가정 IV는 확률밀도함수를 추정함에 있어 가중치를 부여하는 역할을 하는 커널에 대한 가정으로, 대칭성(또는 Gaussian 특성)과 한정성 등을 보장함으로써 확률밀도함수 추정치가 발산하지 않고 성공적으로 본래의 분포에 수렴할 수 있는 조건을 부여하는 조건이다.

3. 커널추정량의 일치성과 점근적 정규분포

식 (1)의 확률밀도함수에 대한 커널추정량에 기대치를 취하여 전개하면 다음과 같다.

$$\begin{aligned}
 \tilde{f}(x) &\equiv E[f_N(x)] \\
 &= \frac{1}{Nh} \sum_{i=1}^N \int K\left(\frac{X_i - x}{h}\right) f(X_i) dX_i \\
 &= \frac{1}{h} \int K(u) f(x + hu) h du \\
 &= f(x) + hf^{(1)}(x) \int K(u) u du + \int K(u) \frac{f^{(2)}(x)(uh)^2}{2} du \\
 &\quad + \int K(u) \frac{f^{(3)}(x)(uh)^3}{6} du + \int K(u) \frac{f^{(4)}(x+h^*u)(uh)^4}{24} du \\
 &= f(x) + \int K(u) \frac{f^{(2)}(x)(uh)^2}{2} du + \int K(u) \frac{f^{(4)}(x+h^*u)(uh)^4}{24} du
 \end{aligned}$$

위의 식에서 세 번째 등호는 변수변환을 통해 성립한다. $(X_i - x)/h \equiv u$ 로 정의하면 $dX_i = du$ 의 관계가 나타난다. 그러므로 변수변환을 하면 세 번째 등호가 성립한다. 네 번째 등호는 연속미분가능한 함수에 대한 테일러 확장(Taylor expansion)을 통해 성립한다.

단, 여기서 함수에 상첨자로 붙어있는 것 가운데 () 안의 숫자는 미분차수를 나타낸다. 위의 전개식에서 세 번째 등호에서의 우변중 두 번째 항과 네 번째 항은 대칭성을 가지는 기함수의 특성상 적분값이 0이 된다. 세 번째 항과 다섯 번째 항은 h^2 와 h^4 를 제외할 경우 나머지 적분부분은 가정에 의해 (최대)값이 위로부터 한정

(bounded above)되어 있기 때문에 각각 $O(h^2)$ 와 $O(h^4)$ 이다. 그러므로

$$\tilde{f}(x) \equiv f(x) + \int K(u) \frac{f^{(2)}(x)(uh)^2}{2} du + O(h^4) \quad (2)$$

또는

$$\tilde{f}(x) \equiv f(x) + O(h^2) \quad (2-1)$$

으로 표현된다.

식 (1)과 같이 확률밀도함수에 대한 추정량을 정의하면 그에 따른 평균자승오차(mean squared error)는 다음과 같이 표현된다.

$$\begin{aligned} \text{MSE}(f_N) &= E[(f_N - f)^2] \\ &= E\{[(f_N - \tilde{f}) + (\tilde{f} - f)]^2\} \\ &= E[(f_N - \tilde{f})^2] + E[(f_N - \tilde{f})(\tilde{f} - f)] + E[(\tilde{f} - f)^2] \\ &= \text{Var}(f_N) + E[(\tilde{f} - f)^2] \end{aligned} \quad (3)$$

그런데 위의 식 (3) 가운데 마지막 등호의 우변 두 번째 항 중 () 안에 표현된 $(\tilde{f} - f)$ 는 점근적 편차(asymptotic bias)라고 한다. 여기에 식 (2-1)의 관계를 이용하면 점근적 편차는 $O(h^2)$ 의 점근적 특성을 가진다. 따라서 점근적 편차는 다음과 같이 점근적으로 무시할 수 있는 항이 된다.

$$E[(\tilde{f} - f)^2] = O(h^4) = o\left(\frac{1}{Nh}\right) \quad (4)$$

여기서 가정 I에 입각하여 $Nh^5 \rightarrow 0$ 이 되므로 마지막 등호는 $O(h^4) = O\left(\frac{Nh^5}{Nh}\right) = o\left(\frac{1}{Nh}\right)$ 의 관계가 성립한다. 그러므로 위의 평균자승편차(MSE)는 점근적으로 첫째 항과 일치하게 된다. 평균자승오차의 첫째 항 $\text{Var}(f_N)$ 은 다음과 같이 다시 쓸 수 있다.

$$\text{Var}(f_N) = E(f_N^2) - E^2(f_N) \quad (5)$$

$$\begin{aligned} E(f_N^2) &= \frac{1}{(Nh)^2} \sum_{i=1}^N \int K^2\left(\frac{X_i-x}{h}\right) f(X_i) dX_i \\ &\quad + (Nh)^2 \sum_{i=1}^N \sum_{j=1}^N \left[\int K\left(\frac{X_i-x}{h}\right) f(X_i) dX_i \right] \left[\int K\left(\frac{X_j-x}{h}\right) f(X_j) dX_j \right] \\ &= \frac{1}{(Nh)^2} \sum_{i=1}^N \int K^2\left(\frac{X_i-x}{h}\right) f(X_i) dX_i + \frac{N-1}{N} (E[f_N(x)])^2 \end{aligned} \quad (6)$$

이때 첫째 항과 둘째 항은 변수변환과 간단한 수학적 기교를 통해 각각 다음과 같이 표현될 수 있다.

$$\text{첫째 항} = \frac{1}{Nh} f(x) \int K^2(u) du + o(1) = \frac{1}{Nh} f(x) \int K^2(u) du + o\left(\frac{1}{Nh}\right) \quad (7)$$

$$\text{둘째 항} = \left(1 - \frac{1}{N}\right) E^2[f_N(x)] = E^2[f_N(x)] + O\left(\frac{1}{N}\right) = E^2(f_N) + o\left(\frac{1}{Nh}\right) \quad (8)$$

그러므로 식 (6)~식 (8)에 의해 분산, $\text{Var}(f_N)$,은 다음과 같다.

$$\begin{aligned} \text{Var}(f_N) &= E(f_N^2) - E^2(f_N) \\ &= \frac{1}{Nh} f(x) \int K^2(u) du + o\left(\frac{1}{Nh}\right) + E^2(f_N) + o\left(\frac{1}{Nh}\right) - E^2(f_N) \\ &= \frac{1}{Nh} f(x) \int K^2(u) du + o\left(\frac{1}{Nh}\right) \end{aligned} \quad (9)$$

식 (3)과 식 (9)에 따라 평균자승오차는 다음과 같다.

$$\begin{aligned} \text{MSE}(f_N) &= \text{Var}(f_N) + o\left(\frac{1}{Nh}\right) \\ &= \frac{1}{Nh} f(x) \int K^2(u) du + o\left(\frac{1}{Nh}\right) \end{aligned} \quad (10)$$

여기에서 중심극한정리(central limit theorem)에 따라 어떤 점 x 에서의 확률밀도함수에 대한 추정량은 점근적으로 정규분포를 가지게 된다.

4. 커널추정량의 일치성과 점근적 정규분포

식 (1)과 같이 어떤 점 x 에서의 확률밀도함수에 대한 점추정량 (point estimator)은 가정 I ~ 가정 IV의 조건을 충족할 때 다음과 같이 일치성과 점근적 정규분포를 가진다.

정리 I. 식 (1)의 $f_N(x)$ 는 $f(x)$ 에 대한 일치추정량이다.

(증명) 식 (2-1)에서 보듯이 $f_N(x) = \hat{f}(x) \rightarrow_p E[f_N(x)] = f(x) + O(h^2)$ 이다. 가정 I의 (1)에 따라 마지막 등호의 두 번째 항은 $o(1)$ 이다. 그러므로 대수의 법칙(law of large numbers)에 의해 $f_N(x) \rightarrow_p f(x)$ 이 된다. Q.E.D.

정리 II. 식 (1)의 커널추정량은 다음과 같은 점근적 정규분포를 가진다.

$$\sqrt{Nh}(f_N(x) - f(x)) \sim_A N\left(0, f(x) \int K^2(u) du\right) \quad (11)$$

(증명) 정리 I과 식 (9), 식 (10) 및 린드버그-레비 중심극한정리(Lindberg-Levy Central Limit Theorem)에 의해 식 (11)이 성립한다. Q.E.D.

위의 식 (11)은 어떤 점 x 에서 확률밀도함수에 대한 커널추정량이 점근적으로 정규분포를 가짐을 의미한다. 그런데 한 가지 주목할 점은 커널추정량 $f_N(x)$ 의 점근속도이다. 일반적으로 표본평균 \bar{X} 는 표본이 무한대로 커질수록 참값 μ 에 접근할 때 그 점근속도가 $1/\sqrt{N}$ 이다. 그렇지만 식 (1)에서 $f_N(x)$ 이 $f(x)$ 으로 수렴하는 데 있어서는 점근속도가 $1/\sqrt{Nh}$ 로 $1/\sqrt{N}$ 보다 느리다. 이는 전자와 같이 추정량이 모집단의 정의구역(domain) 전체를 대상으로 하는 전범위 추정량(global estimator)인 반면 본절에서 논의하고 있는

$f_N(x)$ 는 주로 점 x 의 근방을 대상으로 추정하는 지역추정량(local estimator)이라는 점에서 차이를 보이기 때문이다. 즉, 표본이 가지고 있는 정보를 많이 사용할수록 추정량이 더 많은 정보를 함유하게 되고 접근속도가 빨라지게 되어 중심극한정리에서 말하는 $1/\sqrt{N}$ 이 되는 반면 정보의 일부만을 사용하게 되면 그만큼 접근속도가 떨어지게 된다.

5. 커널추정량의 점근분포에 대한 일고

가정 I에 따르면 표본의 크기(N)가 무한대로 증가할 때 h 는 0으로 수렴한다. 그런데 Devroye and Györfi(1985)에 의하면 h 의 접근속도를 가정 I보다 완화시켜 0에 근접하는 속도를 다소 늦춘다면 점근적 정규분포에서의 분산 추정치가 다음과 같이 달라질 수 있음을 보여주고 있다.

$$E[(\hat{f}(x) - f(x))^2] \sim_A \frac{1}{Nh} f(x) \int K^2(u) du + \frac{1}{4} h^4 [f^{(2)}(x)]^2 \left(\int u^2 K(u) du \right)^2$$

$$\text{여기서 최적(optimal) } h \text{는 } \left[\frac{A}{N \int [f^{(2)}(z)]^2 dz} \right]^{1/5},$$

$$\text{단, } A = \frac{\int K^2(u) du}{\left(\int u^2 K(u) du \right)^2}$$

여기서 만약 f 가 정규분포의 확률밀도함수라고 하면

$$\int [f^{(2)}(z)]^2 dz = \frac{3}{8\sqrt{\pi}\sigma^5} \text{ 이 된다.}$$

6. 층화무작위표본하에서의 커널추정량

식 (1)은 모든 표본관측치의 가중치가 균등하게 배분되어 있는 무작위표본(random sample)에서 정의된 확률밀도함수에 대한 추정량이다. 그런데 본 연구와 같이 표본관측치별로 표본가중치가 서로 다른 층화무작위표본(stratified random sample)을 사용하는 경우에는 다음과 같이 추정량을 수정하여야 한다.

$$f_N(x) \equiv \hat{f}(x) = \frac{1}{h} \sum_{i=1}^N K\left(\frac{X_i - x}{h}\right) w_i \quad (12)$$

여기서 w_i 는 i 번째 표본관측치의 표본가중치를 나타낸다.

부록 Ⅲ. 연도-연령별 자연대수 가구소득 분포도

본 부록에는 1982년부터 2002년도까지의 기간 중 각 연도별로, 다시 25세부터 65세까지의 가구주 연령별로 추정된 자연대수 가구 소득의 분포도가 수록되어 있다. 본문 제Ⅲ장에서 설명한 바를 따라, 각 연도-연령 계층별로 smoothing의 정도를 달리한 2개의 추정경험분포가 제시되어 있으며, 이와 대조할 수 있도록 소표본과 동일한 평균과 분산의 정규분포 확률밀도함수가 음영으로 처리되어 제시되어 있다. 각각의 분포도에서 세로선의 간격은 당해 연도-연령 소표본에서 추정한 표준편차를 나타내며, 세로축의 높이는 확률 밀도 1에 맞추었음을 밝힌다.

[부도 III-1] 가구주 연령별 도시간구 소득분포함수의 추정
(1982년, 25~36세)

[부도 III-1]의 계속(37~48세)

[부도 III-1]의 계속(49~60세)

[부도 III-1]의 계속(61~65세)

[부도 III-2] 가구주 연령별 도시가구 소득분포함수의 추정
(1983년, 25~36세)

[부도 III-2]의 계속(37~48세)

[부도 III-2]의 계속(61 ~ 65세)

[부도 III-3] 가구주 연령별 도시가구 소득분포함수의 추정
(1984년, 25~36세)

[부도 III-3]의 계속(37~48세)

[부도 III-3]의 계속(61 ~ 65세)

[부도 III-4] 가구주 연령별 도시가구 소득분포함수의 추정
(1985년, 25~36세)

[부도 III-4]의 계속(37~48세)

[부도 III-4]의 계속(49~60세)

[부도 III-4]의 계속(61 ~ 65세)

[부도 III-5] 가구주 연령별 도시가구 소득분포함수의 추정
(1986년, 25~36세)

[부도 III-5]의 계속(37~48세)

[부도 III-5]의 계속(61~65세)

[부도 III-6] 가구주 연령별 도시가구 소득분포함수의 추정
(1987년, 25~36세)

[부도 III-6]의 계속(37~48세)

[부도 III-6]의 계속(49~60세)

[부도 III-6]의 계속(61~65세)

[부도 III-7] 가구주 연령별 도시간구 소득분포함수의 추정
(1988년, 25~36세)

[부도 III-7]의 계속(37~48세)

[부도 III-7]의 계속(49~60세)

[부도 III-7]의 계속(61~65세)

[부도 III-8] 가구주 연령별 도시간구 소득분포함수의 추정
(1989년, 25~36세)

[부도 III-8]의 계속(37~48세)

[부도 III-8]의 계속(61~65세)

[부도 III-9] 가구주 연령별 도시간구 소득분포함수의 추정
(1990년, 25~36세)

[부도 III-9]의 계속(37~48세)

[부도 III-9]의 계속(49~60세)

[부도 III-9]의 계속(61~65세)

[부도 III-10] 가구주 연령별 도시간구 소득분포함수의 추정
(1991년, 25~36세)

[부도 III-10]의 계속(37~48세)

[부도 III-10]의 계속(49~60세)

[부도 III-10]의 계속(61~65세)

[부도 III-11] 가구주 연령별 도시간구 소득분포함수의 추정
(1992년, 25~36세)

[부도 III-11]의 계속(37~48세)

[부도 III-11]의 계속(49~60세)

[부도 III-11]의 계속(61~65세)

[부도 III-12] 가구주 연령별 도시간구 소득분포함수의 추정
(1993년, 25~36세)

[부도 III-12]의 계속(37~48세)

[부도 III-12]의 계속(49~60세)

[부도 III-12]의 계속(61~65세)

[부도 III-13] 가구주 연령별 도시간구 소득분포함수의 추정
(1994년, 25~36세)

[부도 III-13]의 계속(37~48세)

[부도 III-13]의 계속(49~60세)

[부도 III-13]의 계속(61~65세)

[부도 III-14] 가구주 연령별 도시간구 소득분포함수의 추정
(1995년, 25~36세)

[부도 III-14]의 계속(37~48세)

[부도 III-14]의 계속(49~60세)

[부도 III-14]의 계속(61~65세)

[부도 III-15] 가구주 연령별 도시간구 소득분포함수의 추정
(1996년, 25~36세)

[부도 III-15]의 계속(37~48세)

[부도 III-15]의 계속(61~65세)

[부도 III-16] 가구주 연령별 도시간구 소득분포함수의 추정
(1997년, 25~36세)

[부도 III-16]의 계속(37~48세)

[부도 III-16]의 계속(49~60세)

[부도 III-16]의 계속(61~65세)

[부도 III-17] 가구주 연령별 도시간구 소득분포함수의 추정
(1998년, 25~36세)

[부도 III-17]의 계속(37~48세)

[부도 III-17]의 계속(49~60세)

[부도 III-17]의 계속(61~65세)

[부도 III-18] 가구주 연령별 도시간구 소득분포함수의 추정
(1999년, 25~36세)

[부도 III-18]의 계속(37~48세)

[부도 III-18]의 계속(61~65세)

[부도 III-19] 가구주 연령별 도시간구 소득분포함수의 추정
(2000년, 25~36세)

[부도 III-19]의 계속(37~48세)

[부도 III-19]의 계속(49~60세)

[부도 III-19]의 계속(61~65세)

[부도 III-20] 가구주 연령별 도시간구 소득분포함수의 추정
(2001년, 25~36세)

[부도 III-20]의 계속(37~48세)

[부도 III-20]의 계속(49~60세)

[부도 III-20]의 계속(61~65세)

[부도 III-21] 가구주 연령별 도시간구 소득분포함수의 추정
(2002년, 25~36세)

[부도 III-21]의 계속(37~48세)

[부도 III-21]의 계속(49~60세)

[부도 III-21]의 계속(61~65세)

장기인력수급추이에 따른 소득세원의 변화

金宗勉 · 成明宰

본 연구에서는 개략적인 수준에서나마 최초로 미시분석에 기초한 소득분포를 추정하고 이를 바탕으로 소득세의 향후 장기 추이를 전망하였다. 소득세뿐만 아니라 소득에 기반하여 징수되는 재정수입이 관련된 제반 연구에서는 물론, 재정지출 측면에서는 공적부조 등 사회보장 정책의 수혜계층 분석에 있어서 소득의 평균 등 총량이 같다고 하더라도 소득의 분포양상에 따라 결과가 다르다는 것은 당연한 사실이다. 그러나 지금까지 우리나라의 연구에서는 분포를 반영하지 못하고 총량위주의 분석만 이루어졌다는 점에서 연구목적과 수단 간에 괴리가 있었다고 할 수 있다. 본 연구는 이러한 소득분포 전망의 문제에 대해 유용한 해결책으로서 분포의 몇 가지 계수가 아니라 분포 자체를 전망할 수 있는 새로운 분석틀을 제시하였다. 또한, 이러한 틀을 이용하여 분포 전망에 필요한 계수들을 구체적으로 추정하여 제시하였다.

보다 구체적으로 연구의 의의를 정리하면, 뚜렷한 속성이 명백히 파악되어 있지 않던 우리나라의 소득분포에 대하여 1982~2002년 사이의 도시가계자료를 이용하여 몇 가지의 전형적 특성(stylized fact)을 새로이 발견하여 제시했다는 점을 꼽을 수 있다.

첫째, 우리나라 도시가구의 소득이 연도별 총소득에서나 연도·연령별 소득에서나 자연대수 정규분포와 유사한 수준을 넘어 실제로 자연대수 정규분포로 간주해도 무방하다는 사실을 보였다. 그러

므로 단지 2개의 모수로써 소득분포 전체를 파악할 수 있으며, 정책모의실험이나 경제모형에 소득분포를 쉽게 대입하여 활용할 수 있다는 유용성이 있다.

둘째, 도시가구의 소득이 정규분포임을 보이는 과정에서 1997년에 연말의 외환위기가 도래하기 이전에 우리 경제의 소득분포에 이미 심각한 붕괴현상이 있었음을 보였다. 본 연구에서 사용한 21개년의 소득자료 중 자연대수 정규분포가 아닌 것으로 밝혀진 것은 1993년과 1995년, 그리고 외환위기 당시인 1997년과 1998년에 국한되었다. 그런데 이 중 1993년과 1995년의 경우에는 비록 자연대수 정규분포를 따른다는 귀무가설이 기각되었으나 외관상 거의 자연대수 정규분포라고 하여도 무방하다. 다만 문제는 1997년과 1998년의 경우이다. 외환위기가 1997년 말에 발발하였고 실제로 1997년 3/4분기까지의 거시통계에서는 경제상황이 이미 극도로 악화되었다는 징후는 찾아보기 힘들다는 사실을 상기할 때, 1997년의 소득분포는 다른 해와 마찬가지로 자연대수 정규분포를 따르거나 그와 비슷한 모양을 할 것으로 생각하기 쉽다. 그러나 1997년의 자연대수 소득분포는 1998년과 같이 매우 심한 좌편향성을 보이고 있다. 이는 외환위기가 발발하여 경제붕괴가 시작된 1997년 말 이전에 이미 우리 경제는 보통의 불경기보다 훨씬 심각한 구조적 문제의 표출이 상당히 진전되어 있었다는 해석이 가능하며, 보다 세밀한 추가적인 연구를 요구하는 결과라고 생각된다. 즉, 1997년과 1998년에는 육안으로도 쉽게 구분될 정도로 소득분포가 예년의 경우에서 크게 벗어남으로써 경제위기가 소득분포에 미친 영향이 지대하였음을 시사해준다. 1999년에는 소득분포가 자연대수 정규분포로 회귀함으로써 일반적인 인식과 달리 소득분포가 상당히 안정적임을 시사하는 것도 특기할 만하다.

셋째, 자연대수를 취하였을 경우 횡단면에서의 연령별 상대소득분포가 실질적으로 불변임을 보임으로써, 향후 연령별 소득분포 변

화를 어느 정도 신뢰성을 갖고 예측할 수 있는 기반을 마련하였다. 본 연구에서는 이러한 특성을 응용한 예로서, 21년의 관측기간이라는 자료의 제약에도 불구하고 40여년에 걸친 세대 생애 연령-소득 곡선을 추정할 수 있음을 보였다.

넷째, 소득의 형평성을 측정하는 지니계수와 소득분포와의 관계를 구체적으로 규명하고 계량화하였으며, 이를 이용하여 소득분포의 분산을 간접적으로 추정할 수 있음을 보였다. 본 연구에서는 소득 분산 전망을 간접적인 방법으로밖에 다루지 못하였으나, 지니계수와 자연대수 정규분포 가구소득의 관계를 처음으로 규명했다는 성과를 대신 꼽을 수 있다.

마지막으로 본 연구에서 찾아낸 소득분포의 특성을 사용하여 향후 소득분포의 변화를 추정하고, 그에 근거하여 향후 소득세의 추이를 예측해보는 사례를 보임으로써 보고서를 마무리한다. 그 결과 향후 2032년까지 우리나라의 근로소득세 실효세율은 다소 상승하는 것으로 예측된다.

<Abstract>

Projecting Changes in the Income Tax Base Arising from Long-term Population Trends

John M. Kim and Myung Jae Sung

This report offers the first instance of income tax revenue projections based on properties of the income distribution which we extract from micro data analysis. It is obvious that the actual distribution of income does matter even when the mean or other aggregate indices of the distribution are the same. This applies both to studies concerned with fiscal revenues such as income tax, and to analyses of fiscal spending in the area of social security policy including social transfers. However, existing research in Korea has relied mainly on aggregate indices. This failure to incorporate distributional information points to a fundamental disagreement between methodology and the objective of income studies. Addressing this issue head on, we suggest a new analytical framework as a useful solution to the problem of projecting income distributions. Using this framework, we also produce specific estimates of parameters that can be used to forecast income distributions.

More specifically, the significance of this report is that it uncovers several stable characteristics, or stylized facts, of the household income distribution using annual data on urban

households in Korea from 1982 to 2002. Previously, the salient properties of income distributions in Korea had not been explicitly characterized.

First, we show that the distribution of urban household incomes in Korea can not only be approximated as, but practically are log-normal. We find that log-normality holds in both annual data for all households and also in each age-specific subsample for each year. Log-normality means that the entire income distribution can be represented with only two parameters, and that income distributions can be directly incorporated into policy simulations and analytical models with ease.

Second, in the course of showing that income distributions are log-normal, we find evidence that the Korean economy was already showing serious signs of a system collapse in 1997 well before the foreign exchange crisis erupted at the end of the year. Among the 21 years for which we examined annual income data, the significant exceptions to log-normal income distributions were found to be 1997 and 1998, which roughly coincide with the crisis period. A priori, one would expect the 1997 income distribution would be log-normal or close to log-normal, since the crisis erupted only at the end of the year and since virtually all post mortems of the crisis have been unable to uncover evidence of abnormalities in the economy's macro indices up to the 3rd quarter of that year. However, we show that the distribution of log-income for 1997 is marked left-skewed, similarly to that for 1998. This suggests that structural problems with far more serious consequences than a periodic recession had already progressed

to a considerable degree even before the system collapsed upon the crisis at the end of the year, an interpretation that calls for further detailed investigation.

Third, we establish a basis for projecting future changes in income distributions by showing that, expressed as log-income, relative income levels between age groups in cross section are more or less invariant with respect to time. As an example of an application that utilizes this property, we show that full cohort lifetime age-income profiles can be constructed from time series income data of durations much shorter than the typical life span.

Fourth, we give an explicit formulaic exposition of the relation between income distributions and the Gini index, which is often used to measure income inequality. We also show that this relation can be quantified with considerable precision, a result which we use to project the variance of income distributions indirectly. One limitation of this study is that we were not able to come up with a direct method of forecasting the variance of income distributions, but we believe our simple formula for determining Gini index values given any log-normal income distribution ought to compensate for this deficiency.

As a final exercise that concludes our report, we derive an actual forecast of future income distributions using the newly uncovered stylized facts of income distributions, based on which we attempt several projections of income tax revenues. Our projections up to 2032 show that the effective tax rate will be rising in the long run.

<著者略歷>

金 宗 勉

서울대학교 經濟學科 卒業
美國 Chicago大 經濟學 博士
現, 韓國租稅研究院 專門研究委員

成 明 宰

서울대학교 經濟學科 卒業
美國 Wisconsin-Madison大 經濟學 碩·博士
現, 韓國租稅研究院 研究委員

研究報告書 03-02

장기 인력수급 추이에 따른 소득세원의 변화

2003年 12月 15日 印刷
2002年 12月 22日 發行

著 者 金宗勉·成明宰
發行人 宋 大 熙
發行處 韓國租稅研究院

☐1318-7774 서울特別市 松坡區 可樂洞 79-6番地
電話: 2186-2114(代), 팩시밀리: 2186-2179

登 錄 1993年 7月 15日 第21-466號
組版 및 一 志 社
印 刷

© 韓國租稅研究院 2002

ISBN 89-8101-240-8

* 잘못 만들어진 책은 바꾸어 드립니다.

값 9,000원