

# 우리나라 소득 구조의 동태적 분석

김재영 · 이병찬

본 논문에서는 한국노동패널의 1998~2008년 11년간 5,000가구를 대상으로 조사한 패널자료를 사용하여 우리나라 소득구조의 동태적 변이과정을 분석하였다. 주로 소득의 구성과 이동성에 초점을 맞추어 분석을 진행하였는데, 위수회귀, 전이행렬, 그리고 경계자기회귀모형 등을 이용하였다. 소득의 구성과 관련해서는 금융소득과 부동산소득의 특수성에 중점을 두고 분석하였고, 이동성에 관해서는 그 정도와 계층별 차이에 중점을 두고 분석하였다. 분석결과 우리나라의 소득구조 동학에 대해 다음과 같은 사실들을 발견하였다. (1) 임금의 분포보다 금융 및 부동산 소득의 분포가 더 불평등하며 부동산 소득이 그 정도가 더 심하다. (2) 고소득층은 중산층에 비해 임금소득이 아니라 금융 및 부동산 소득을 많이 올리고, 최상위층일수록 부동산 소득이 중요해진다. (3) 근로소득의 분배는 지속적으로 악화되고 있다. (4) 연간 소득이동성이 크에도 불구하고 고소득을 올리는 데에는 지속성이 있다. (5) 중산층 및 저소득층이 경기침체에 더 취약하다.

주제어: 소득구성, 동태적 변이

## 1. 서론

자본주의 경제에서 소득불평등은 매우 중요한 주제이며 그동안 많은 학자들이 관심을 가지고 연구해왔다.<sup>(1)</sup> 그 중 소득불평등의 측정 문제와 관련하여서는 지니계수와 엡킨슨지수 등이 개발되어있다. 그러나 이것들은 기본적으로 소득구조의 '정태적'인 측면에 초점을 맞추어 특정 시점의 소득 구조가 가진 특성을 파악하는데 사용되는 도구이며, 그 동태적인 변화를 분석하는 데는 적절하지 않다. 이들 지수로는 소득구조의 변화가 “부익부 빈익빈”의 양상을 갖는지, “부익빈 빈익부”의 양상을 갖는지 구분할 수 없는 것이다. 소득구조의 '동태적' 변화양상은 정태적 소득불평등문제 못지않게 최근 전 세계적인 관심사가 되고 있고, 본고에서는 지니계수로는 포착할 수

(1) 그간의 연구에 대한 소개서로 李俊求(1989) 및 Campano, F. and D. Salvatore(2006)를 들 수 있다.

없는 동학에 초점을 맞추려한다.

전통적으로 소득의 동태적 변화는 사회학자들이 주로 연구하여 왔다. 사회학자들은 ‘사회이동(social mobility)’이라는 용어로 그 동태적 변화를 규정하며 주로 세대 간 직업위신의 변화를 연구하였는데, 주로 부 혹은 계급의 대물림현상에 관심을 기울였다[양춘(2000), 홍두승·구해근(2001)]. 그러나 계량경제학적 기법을 사용하면 소득구조의 동태적 변화에 대한 보다 체계적이고 정량적인 분석을 할 수 있다. 특히, 이 문제와 관련한 방대한 패널자료가 축적되어 있어 계량경제학적 분석이 적합한 상황이다.

본 논문에서는 한국의 소득자료를 이용하여 소득이동(income mobility)의 양상 및 그 특징적인 현상들을 밝히려한다. 특히 소득의 구성<sup>(2)</sup>과 이동성에 분석의 초점을 맞추고 있는데, 이를 위해 Koenker and Hallock(2001)의 위수회귀(quantile regression) 방법과 Tsay(1989)의 경계자기회귀모형(threshold AR, TAR)을 사용할 것이다.

본고의 구성은 다음과 같다. 2절에서는 소득이동을 연구하는 선행연구와 그 방법론들을 살펴볼 것이다. 3절에서는 본고에서 사용한 자료에 대해 설명하고, 4절에서는 여러 소득 원천들의 구성을 살펴보는데 이를 위해 위수회귀기법을 사용할 것이다. 5절에서는 마르코프 전이행렬(Markov transition matrix)과 경계자기회귀모형을 사용하여 소득의 이동성을 분석할 것이다.

## 2. 선행연구

소득구조의 동학을 마르코프 연쇄(Markov chain)와 전이행렬 개념을 이용하여 분석한 대표적인 연구로 Atkinson(1981), Fields and Ok(1999), Markandya(1982, 1984), Shorrocks(1976, 1978a) 등이 있다. Shorrocks(1976)는 영국의 자료를 이용하여 소득이동이 마르코프과정을 따르는 지를 검정하였고, 2차 마르코프 과정으로 소득이동을 적절히 모형화 할 수 있다고 결론 내렸다.

소득이 마르코프 과정을 따를 경우 적절한 정규성가정하에서 그 분포가 극한분포(stationary, limiting distribution)로 수렴하고, 따라서 장기적으로 기회의 평등이 실현된다고 볼 수 있기에 이 결론이 의미가 있다. 그러나 Shorrocks가 사용한 표본은

---

(2) 근로소득, 금융소득, 부동산소득 등 다양한 소득 요인들이 어떠한 비율로 구성되어 있는지를 의미한다.

1963, 1966, 1970년 세 해에 걸친 동년배 영국남성 800명의 소득 자료였다. 즉 두 번의 소득이동자료만 가지고서 분석을 진행한 것이다. 더욱이 1년에 한 계층만 이동할 수 있다는 제한적인 가정 하에 분석하였기에, 이 결론을 최근의 풍부한 자료를 이용하여 다시 도출해 볼 필요가 있다.

또한 소득이동의 사회적 비용, 혹은 이득을 측정하는 연구들도 진행되었는데, 주로 전이행렬의 특성에 의존하는 경우가 많았다. 예를 들면 Atkinson(1981)은 두 전이행렬 사이에 특정 관계가 성립하면 한 행렬이 다른 행렬보다 더 많은 소득이동을 의미한다고 정의하고, 이 경우 그 사회후생함수 값이 더 커진다고 주장하였다. 한편 Markandya(1982)는 전이행렬의 관점에서 더 많은 사회이동이 더 낮은 사회후생을 야기하는 조건을 보여주기도 하였다. Markandya(1984)는 마르코프 과정이 극한분포를 가진다는 점에 착안하여, 그 극한분포 하에서의 사회후생들을 비교함으로써 소득이동의 후생비용을 측정하는 방법을 제안하기도 하였다.

전이행렬을 이용한 분석은 무엇보다도 단순한 것이 특징이다. 이는 장점이자 단점인데, 단순하고 한 눈에 파악할 수 있기에 직관적인 이해를 돕는 것은 장점이지만, 모형에 구조를 주기가 힘들고, 전이행렬을 구성하는 과정에서 많은 정보들을 잃기에 세밀한 분석이 힘든 것은 단점이다. 본고에서는 직관적인 이해를 돕기 위해 4장에서 전이행렬을 사용함과 동시에, 분위 간 이동 정보뿐 아니라, 소득 자체의 크기까지 활용하기 위해 경계자기회귀모형도 사용할 것이다. 경계자기회귀모형을 사용하면 소득 계층별 소득의 변화과정을 분석할 수 있을 뿐 아니라 소득계층간의 전이 과정도 분석할 수 있다.

1990년대 이후 진행된 연구들은 주로 패널자료를 이용하였다. Jarvis and Jenkins (1998)는 영국의 자료를 이용하였는데, 기본적으로 전이행렬을 이용한 분석법을 채택하였으나 나이와 성별에 따라 소집단을 구성하여 집단별로 다른 결론을 이끌어낸 것이 특징이다. Corak and Heisz(1999)의 경우 캐나다의 세대 간 소득이동성을 분석하기 위해 회귀분석 및 전이행렬을 사용하였는데, 부모가 최하위층일 때 자녀 소득의 이동성이 가장 크며, 부모의 로그 소득과 자녀의 로그 소득이 역-V자 관계를 가짐을 보였다. Geweke and Keane(2000)의 경우 미국자료를 사용하였는데, 연령, 본인과 부모의 교육수준, 그리고 결혼 여부 같은 변수들로 로그 소득을 추정하는 회귀분석을 수행하였다.

이처럼 패널자료를 활용하면 다양하고 더 구체적인 분석을 할 수 있다는 장점이

있다. 따라서 한국의 패널자료를 분석하여 한국의 상황을 구체적으로 이해한다면 그 결과를 정책적인 목적으로도 상당히 유용하게 사용할 수 있을 것이다. 다만 기존의 연구에서 사용한 회귀모형들로는 계층에 따른 비선형성을 설명할 수 없다는 한계가 있다. 그러나 소득 동학의 경우 소득 층위에 따라 상당한 이질성이 예상되고, 따라서 특정한 경계를 기점으로 다른 구조를 갖는 모형을 사용하는 것이 좋을 것이다. Tsay (1989)는 바로 이러한 특성을 가지는 시계열 모형인 경계자기회귀모형(Threshold Auto-regression model)을 제안하고 분석하였는데, 본고에서는 이를 사용하려 한다. 다만 이 모형이 패널자료를 분석하기 위해 개발된 모형이 아니기 때문에, 패널자료에 맞도록 모형을 적절히 수정하여 사용할 것이다.<sup>(3)</sup>

### 3. 자료

본고에서 사용한 소득자료는 모두 한국노동패널(KLIPS: Korea Labor and Income Panel Study)에서 얻은 것이다. 한국노동패널은 1998년에 5,000가구를 대상으로 조사를 시작하여 2008년까지 11년간 추적 조사한 자료인데, 이 중 11년간 지속적으로 조사에 응한 2,607가구의 자료만을 추려내었다. 패널은 소득을 근로소득, 금융소득, 부동산소득, 사회보험, 이전소득, 그리고 기타소득의 6가지로 나누어 수록하고 있는데, 사회보험, 이전소득, 그리고 기타소득의 영향력이 미비하다고 판단하여, 총소득( $y_{i,t}$ ), 근로소득( $l_{i,t}$ ), 금융소득( $f_{i,t}$ ), 부동산소득( $r_{i,t}$ )의 네 변수를 구성하였다. 따라서 총 114,708개의 변수( $11 \times 2607 \times 4$ )를 사용하였다.

소득자료 이외에 소비자물가지수, 실질경제성장률, 그리고 경기종합지수를 사용하였는데, 이는 모두 국가통계포털(KOSIS)에서 구하였다. 소비자물가지수(2005=100)는 소득자료를 물가조정하는데 사용하였다. 실질경제성장률은 소득에서 성장 및 경기변동의 효과를 제거하는데 사용하였다. 공통의 성장추세를 제거함으로써 변수들의 안정성(stationarity)을 확보함과 동시에 소득의 상대적 이동에 더 초점을 맞출 수 있다고 생각했기 때문이다. Bernasconi and Dardanoni(2004)에 의하면 소득이동은 “구

(3) 물론 현재 개발 중인 패널 시계열 기법 중, Cermeno and Grier(2001)나 Pakel *et al.*(2009)의 패널 GARCH 모형을 이용하여 비선형성을 분석하는 것도 가능할 것이다. 다만 경계자기회귀모형의 ‘경계’ 개념이 직관적으로 더 매력적이기 때문에 본고에서는 경계자기회귀모형을 사용하였다.

조이동(structural mobility)”과 “교환이동(exchange mobility)”으로 나눌 수 있는데, 구조이동은 소득의 한계분포가 변하는 것을 의미하고, 교환이동은 한계분포 내에서 위치, 소득의 순위가 변하는 것을 의미한다. 그리고 구조이동은 다시 두 가지로 나눌 수 있는데, 평균소득이 일정한 상태에서 상대소득분포가 변하는 것이 하나고 평균소득 자체가 변화하는 것, 즉 경제 성장이 다른 하나이다. 본고에서는 성장의 효과를 제거하는 대신 상대소득분포의 변화와 교환이동에 집중하려한다.

마지막으로 경기종합지수는 5장에서 불황 더미변수를 만드는데 사용하였다. 경기동행지수에서 성장을 제거한 “동행지수 순환변동치”를 사용하였는데 12개월 중 7개월 이상 지수가 100 미만인 해는 불황으로 간주하였다. 이 기준을 따를 경우 1998, 1999, 2001, 2005년이 불황이고, 남은 7개년은 호황이다.

#### 4. 소득의 구성 - 금융 및 부동산소득의 역할

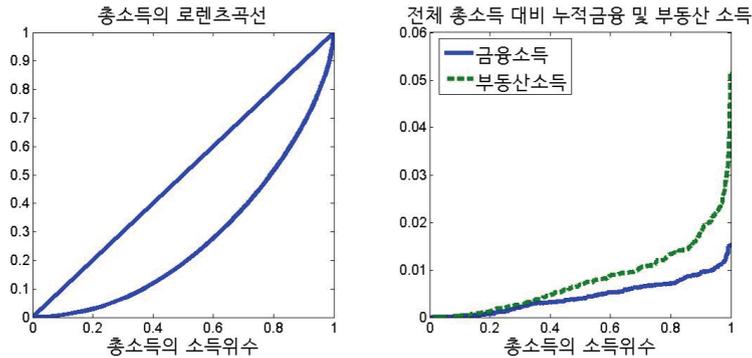
이 절에서는 금융소득과 부동산소득이 계층별로 어떻게 분포하고 이것이 소득구조의 동학에 어떤 영향을 미치는지를 분석할 것이다. 이를 통해 고소득층을 고소득층이게 하는 것이 금융, 부동산 등 투자소득임을 밝힐 것이다.

우선 직관적인 이해를 도모하기 위해 2008년 한 해의 자료만 살펴보자. 전체 2,607가구 중 근로소득이 있는 가구는 2,119가구(81%)인데 반해, 금융소득이 있는 가구는 296가구(11%), 부동산소득이 있는 가구는 332가구(13%)뿐이다. 금융소득과 부동산소득이 소수에게 집중되어있음을 짐작할 수 있는 대목이다. 둘 중 하나라도 있는 가구는 546가구(21%)뿐이고, 둘 다 있는 가구는 82가구(3%)뿐이다. 금융과 부동산이 대표적인 자산 축적 및 투자수단임을 감안할 때 이는 매우 놀라운 수치이다.

〈표 1〉은 2008년 한 해 동안 총소득을 기준으로 할 때 상위 가계들이 각 항목별 소득을 얼마나 점유했는지를 나타낸 표이다. 상위 5%와 금융소득이 만나는 지점의

〈표 1〉 총소득 기준 고소득층의 각 항목별 점유비율

2008	근로소득	금융소득	부동산소득	금융+부동산
상위 1%	4.36%	11.39%	38.36%	32.20%
5%	17.62%	31.38%	57.40%	51.46%
10%	29.86%	38.52%	64.58%	58.62%
30%	65.56%	59.19%	81.40%	76.32%



〈그림 1〉 금융소득 및 부동산소득이 총소득에서 차지하는 비율

31.38%는 총소득 기준 상위 5%의 가계가 전체 금융소득의 31.38%를 얻었다는 의미이다. 마지막 열은 전체 가계의 금융소득과 부동산소득의 합에 대한 고소득층의 금융소득과 부동산소득의 합을 나타낸다. 표를 보면 금융소득과 부동산소득이 고소득층에 상당한 정도로 집중되어있으며, 특히 부동산소득이 그 정도가 더 심함을 확인할 수 있다. 일반적으로 부동산 투자를 위해 필요한 억 단위의 자본금이 중산층 이하에게는 상당한 부담이기에, 부동산투자과 그 과실이 상위층에 집중되는 것이라 짐작할 수 있다. 금융소득은 상대적으로 적은 자본으로도 투자할 수 있기에 상위 30%에서 근로소득과의 누적점유율이 역전되는 것도 확인할 수 있다.

〈그림 1〉은 2008년 자료를 가지고 그린 그림이다. 좌측 그림은 총소득의 로렌츠곡선을 그린 것이고, 우측 그림은 총소득이 가장 낮은 사람부터 순서대로 누적금융소득과 누적부동산소득이 전체 총 소득에서 차지하는 비율을 도시한 것이다. 예를 들어 총소득으로 따진 소득의 위수(quantile)가 약 0.7일 때 총소득대비 누적부동산소득의 비율이 0.01인데, 이는 총소득기준 소득 하위 70% 이내인 사람들의 부동산소득을 모두 합한 것이 경제 전체의 소득 중 1%를 차지한다는 뜻이다.

〈그림 1〉을 보면 고소득층이 금융과 특히 부동산소득의 대부분을 차지하고 있음을 확인할 수 있다. 예를 들어 총소득기준 하위 80%가 전체 총 소득 중 대략 2% 정도를 금융소득과 부동산소득으로 얻고 있는데 반해, 상위 20%가 총 소득의 4.5% 이상을 금융 및 부동산 소득으로 얻는 것이다. 또한 부동산소득은 대략 상위 5%, 금융소득은 30% 정도에 상당히 밀집되어 있는데, 결국 부동산소득이 금융소득보다 더 최상위층에 집중되어 있는 것이다. 결국 〈표 1〉에서와 마찬가지로 고소득층을 특징짓

〈표 2〉 연도별, 소득별 지니계수

	근로소득	금융소득	부동산소득	총소득
1998	0.449	0.964	0.972	0.438
1999	0.439	0.953	0.974	0.425
2000	0.432	0.980	0.968	0.423
2001	0.463	0.965	0.971	0.439
2002	0.456	0.967	0.976	0.449
2003	0.457	0.967	0.970	0.445
2004	0.478	0.973	0.973	0.535
2005	0.484	0.956	0.968	0.449
2006	0.488	0.958	0.961	0.507
2007	0.500	0.957	0.967	0.505
2008	0.508	0.960	0.966	0.457

는 것이 금융 및 부동산소득이며, 이 두 소득은 상당히 불평등한 분포를 가지고 있음을 확인할 수 있다.

〈표 2〉는 연도별, 소득별 지니계수를 계산한 것으로, Deaton(1997, p. 139)이 제시한 다음의 식을 이용하였다.

$$(4.1) \quad \gamma = \frac{N+1}{N-1} - \frac{2}{N(N-1)\mu} \sum_{i=1}^N \rho_i x_i$$

이 식의  $\gamma$ ,  $x_i$ ,  $\mu$ ,  $N$ 은 각각 순서대로 지니계수, 소득, 평균소득, 가구 수이고,  $\rho_i$ 는 소득의 순위로 소득이 가장 높은 가계가 1이고 가장 낮은 가계가  $N$ 이다.

〈표 2〉를 보면 우선 금융소득과 부동산소득이 극단적으로 불평등하며 그 정도가 시간이 흐름에 따라 그다지 변하지 않음을 확인할 수 있다. 이와는 반대로 근로소득은 시간이 지날수록 점점 불평등해지는 특징을 보인다. 한국의 소득분배구조가 지난 10여 년간 지속적으로 악화되는 데 가장 큰 영향을 준 것이 근로소득분배의 악화일 것이라 짐작할 수 있는 대목이다. 총소득의 경우 사회보험이나 정부이전소득 등의 영향으로 부침을 보이나 그래도 악화되는 추세를 보인다.<sup>(4)</sup>

(4) 〈표 2〉의 값들이 통계청에서 제시하는 수치(0.3 내외)보다 높은데, 패널에서 표본을 선택하는 과정에서, 혹은 11년 동안 지속적으로 응답한 가구만 추리는 과정에서 약간의 편향이 발생한 것이 아닌지 의심된다. 또 하나 흥미로운 것은 2001년을 제외하면 불황이었던 해(1998, 1999, 2005)에 지니계수가 작아졌다는 것이다. Janvry and Sadoulet(1996)는 라틴 아메리카의

이상의 논의를 통해 고소득층을 고소득층이게 하는 것이 금융소득과 부동산소득임을 확인하였다. 이제 위수회귀(quantile regression)기법을 이용하여 이를 분석할 것인데, 그전에 위수회귀기법에 대해 간단히 소개하겠다.<sup>(5)</sup>

종속변수  $y$ 는 설명변수  $X$ 와 오차항  $\varepsilon$ 에 의해 결정된다고 하자.

$$(4.2) \quad y = g(X, \varepsilon).$$

이때  $\alpha$ -위수(quantile)에 대하여 위수회귀를 하여 얻은  $\hat{\beta}$ 은 식 (4.3)이 나타내는 의미를 갖는다.

$$(4.3) \quad F_{y|X}^{-1}(\alpha) = X\hat{\beta}.$$

다시 말해  $X\hat{\beta}$ 는  $y|X$ 의  $\alpha$ -위수에 대한 추정치이다. 전통적인 회귀분석 모형인 “ $y = X\beta + \varepsilon$ ”에서는  $X$ 값에 따라서  $y|X$  분포의 평균만이 변하고 분포자체는 그대로이지만, 현실적으로 이 가정이 충족되리라 기대하기는 어렵다. 다행히 이런 경우에도 위수회귀기법은 적용될 수 있다는 것이 이 방법의 장점이다. 실제로  $\hat{\beta}$ 를 계산할 때는 선형 계획법(linear programming)을 이용한다.

여기서는 종속변수로 총소득( $y_{i,t}$ )을, 독립변수로 상수항, 근로소득( $l_{i,t}$ ), 금융소득( $f_{i,t}$ ), 그리고 부동산소득( $r_{i,t}$ )을 3기전 자료까지 사용하였다.

$$(4.4) \quad y_{i,t+1} = g(1, l_{i,t}, l_{i,t-1}, l_{i,t-2}, f_{i,t}, f_{i,t-1}, f_{i,t-2}, r_{i,t}, r_{i,t-1}, r_{i,t-2}, \varepsilon)$$

따라서 전체 자료가 11기간으로 구성되어 있는데 설명변수로 과거 3기의 자료가

---

경우 불황을 겪으며 지니계수가 커졌다고 보고하였다. 5절에서 추정한 불황더미의 계수를 보면 저소득층이 고소득층보다 불황에 큰 영향을 받기 때문에 이 결과는 매우 흥미롭다. 한 가지 가능한 설명은 지니계수가 양 극단보다는 중산층의 변화에 민감하게 반응한다는 사실을 이용하는 것이다[李俊求(1989)]. 5절에서 추정한 결과를 보면 전기에 고소득층이었다가 중산층이 된 사람들이 저소득층이었다가 중산층이 된 사람들보다 불황에 민감하게 반응하고, 이것이 양극단(저소득층과 고소득층)의 반(anti) 평등적인 변화보다 더 큰 친(pro) 평등적인 변화로 계산되었다고 해석할 수 있다. 아무래도 고소득층에서 중산층으로 이동한 사람들이 저소득층에서 중산층으로 이동한 사람들보다 부유하기 때문이다.

(5) 이하의 내용은 Koenker and Hallock(2001)을 참조하였다.

〈표 3〉 위수회귀 결과

	1	$l_{i,t}$	$l_{i,t-1}$	$l_{i,t-2}$	$f_{i,t}$	$f_{i,t-1}$	$f_{i,t-2}$	$r_{i,t}$	$r_{i,t-1}$	$r_{i,t-2}$
5%	16.3	0.28	0.07	0.06	0.02	0.01	0.01	0.00	0.01	0.00
15%	106.4	0.46	0.11	0.07	0.01	0.00	0.00	0.01	0.01	0.00
25%	161.6	0.54	0.13	0.08	0.01	0.00	0.00	0.02	0.02	0.01
35%	206.4	0.58	0.15	0.09	0.01	0.02	0.06	0.03	0.03	0.03
45%	261.1	0.61	0.17	0.09	0.18	0.07	0.13	0.05	0.04	0.06
55%	322.4	0.62	0.18	0.11	0.19	0.21	0.19	0.10	0.09	0.07
65%	406.3	0.64	0.19	0.13	0.26	0.19	0.21	0.17	0.11	0.10
75%	531.6	0.63	0.22	0.16	0.43	0.32	0.28	0.32	0.10	0.18
85%	744.6	0.66	0.24	0.18	0.42	0.63	0.38	0.60	0.13	0.18
95%	1328.5	0.80	0.26	0.27	0.71	0.94	0.30	1.03	0.18	0.15

\*색칠은 계수가 0.1 이하임을 의미한다.

쓰이므로, 각 가구별로 8개의 식이 나오고, 2,607가구가 있으므로 총 20,856개의 식을 이용하여  $\hat{\beta}$ 을 구하였다. 그 결과는 〈표 3〉과 같다.

〈표 3〉의 세로축은  $\alpha$ -위수를, 가로축은 각각 해당변수의 추정된 계수 값을 의미한다. 우선 눈에 띄는 것은 표의 우측상단이 전부 0에 가깝다는 것이다. 5~35% 위수를 추정하는 데에는 금융소득과 부동산소득은 거의 무시되고 있다. 또한 35~85% 위수를 추정하는데 근로소득은 큰 차이를 만들지 못하는 것도 확인할 수 있다. 결국 투자소득을 얼마나 올리느냐 혹은 올리지 못하느냐에 의해 현재상황하에서 다음 기에 (현재에 비해 상대적으로) 고소득층이 되는지, 혹은 저소득층이 되는지가 거의 결정되는 것으로 보인다. 특히 (상대적으로) 저소득층이 될 경우 투자소득이 없기 때문에 얼마만큼의 노동소득을 거두느냐가 다음 기 소득에 대한 거의 유일한 결정요소이고, 중산층이 될 경우 금융소득이, 그리고 고소득층이 될 경우 금융과 부동산소득이 중요한 결정요소로 작용하는 것도 볼 수 있다.

또한 5%와 15%, 85%와 95% 간 계수의 급격한 변화도 흥미로운 부분이다. 이는 최하위와 최상위 5%의 경우 전체 사회에 비해 상당히 이질적인 동태적 특성을 가지고 있음을 시사한다.

〈표 4〉 1998~1999, 1998~2008년의 총소득 분위별 전이행렬

	1998~1999					1998~2008				
	1	2	3	4	5	1	2	3	4	5
1	64%	22%	9%	4%	3%	50%	27%	13%	6%	5%
2	23%	46%	24%	12%	6%	23%	27%	22%	16%	12%
3	7%	20%	37%	25%	10%	14%	24%	22%	25%	14%
4	4%	7%	20%	34%	21%	9%	13%	25%	28%	23%
5	3%	4%	11%	25%	61%	5%	8%	17%	25%	45%

〈표 5〉 1998~1999, 1998~2008년의 근로소득 분위별 전이행렬

	1998~1999					1998~2008				
	1	2	3	4	5	1	2	3	4	5
1	63%	23%	9%	4%	3%	50%	27%	13%	6%	6%
2	23%	46%	24%	12%	6%	23%	27%	22%	16%	12%
3	7%	20%	37%	25%	10%	14%	24%	22%	25%	14%
4	4%	7%	20%	34%	21%	8%	14%	25%	28%	24%
5	3%	4%	10%	25%	61%	5%	8%	18%	26%	44%

## 5. 소득의 이동성

### 5.1. 전이행렬을 이용한 분석

이 절에서는 소득의 이동성, 혹은 변동성을 중점적으로 분석한다. 이를 통해 각 가계의 소득이동 정도가 경제 전체의 성장에 따른 이동 폭에 비해 상당히 크다는 것을 밝히고, 또한 계층별로 그 정도가 상이함을 보일 것이다.

여기서도 모형을 통해 분석하기에 앞서 표와 그림으로 접근해보자. 다음의 전이행렬들이 직관적 이해에 도움이 될 것이다.

〈표 4〉는 98~99, 98~08의 두 기간에 해당하는 총소득( $y_{i,t}$ )의 전이행렬이다. 각 행렬의  $(i, j)$ -원소는  $j$ -분위에서 출발하여  $i$ -분위가 된 비율을 의미한다. 소득의 크기 자체가 아니라 분위를 기준으로 삼았기 때문에 위의 표는 교환이동만을 반영하고 있다. 98~99 전이행렬을 보면 특히 중산층에서 매해 교환이동이 상당한 정도로 발생하지만, 동시에 양극단의 경우 어느 정도 지속성(persistency)도 있음을 알 수 있다.

〈표 5〉는 근로소득( $l_{i,t}$ )의 전이행렬인데, 〈표 4〉와 거의 유사한 양태를 보인다.

〈표 6〉 1998~1999, 1998~2008년의 금융소득 분위별 전이행렬

	1998~1999			1998~2008		
	n = 472			n = 494		
	1	2	3	1	2	3
1	0%	66%	56%	0%	76%	76%
2	59%	13%	10%	40%	6%	7%
3	41%	21%	33%	60%	18%	17%

〈표 7〉 1998~1999, 1998~2008년의 부동산소득 분위별 전이행렬

	1998~1999			1998~2008		
	n = 283			n = 429		
	1	2	3	1	2	3
1	10%	58%	40%	15%	70%	60%
2	54%	28%	16%	42%	19%	14%
3	37%	14%	44%	43%	11%	26%

〈표 8〉 1998~1999, 1998~2008년의 금융소득과 부동산소득의 유무로 구성된 전이행렬

	금융소득				부동산소득			
	1998~1999		1998~2008		1998~1999		1998~2008	
	無	有	無	有	無	有	無	有
無	91%	61%	90%	76%	95%	47%	89%	57%
有	9%	39%	10%	24%	5%	53%	11%	43%

〈표 6〉과 〈표 7〉은 각각 금융소득( $f_{i,t}$ )과 부동산소득( $r_{i,t}$ )의 전이행렬이다. 두 번째 행의 'n'은 두 해중 한 해라도 양의 소득을 얻은 가구의 수인데, 예를 들어 〈표 6〉의 'n = 472'는 98, 99년 중 한 해라도 금융소득이 있었던 가구가 472가구라는 뜻이다. 총소득이나 근로소득과는 다르게 표본이 적기 때문에 3분위로 나누어서 전이행렬을 계산하였다. 〈표 6〉과 〈표 7〉에서 두 가지 특징을 발견할 수 있는데, 금융소득이나 부동산소득을 올리는 가구 자체가 상당히 적다는 것이 하나이고, 이동성이 근로소득에 비해 훨씬 크다는 것이 다른 하나이다.

각 소득의 유무만 가지고 전이행렬을 만들 수도 있는데, 〈표 8〉을 보면 애초에 금융소득과 부동산소득의 보유여부 자체가 상당히 변동성이 큰 성질임을 알 수 있다. 그럼에도 불구하고 소득이 없었던 사람보다는 있었던 사람이 다시 소득이 있을 확률

이 훨씬 크다는 점에서, 이 두 소득이 고소득층의 지속성에 상당히 기여하고 있다고 짐작할 수 있다. 그리고 4절의 논의에 미루어 짐작할 수 있듯이 이러한 효과는 금융 소득보다 부동산소득에서 더 크게 나타난다.

위의 전이행렬들을 해석한 결과를 정리하면 다음 두 가지 추측이 가능하다. 하나는 소득의 이동성이 상당히 크다는 것이고, 다른 하나는 그럼에도 불구하고 금융과 부동산소득의 동학 때문에 이동성이 계층에 따라 다르게 나타날 가능성이 존재한다는 것이다.

## 5.2. 경계자기회귀모형을 이용한 분석

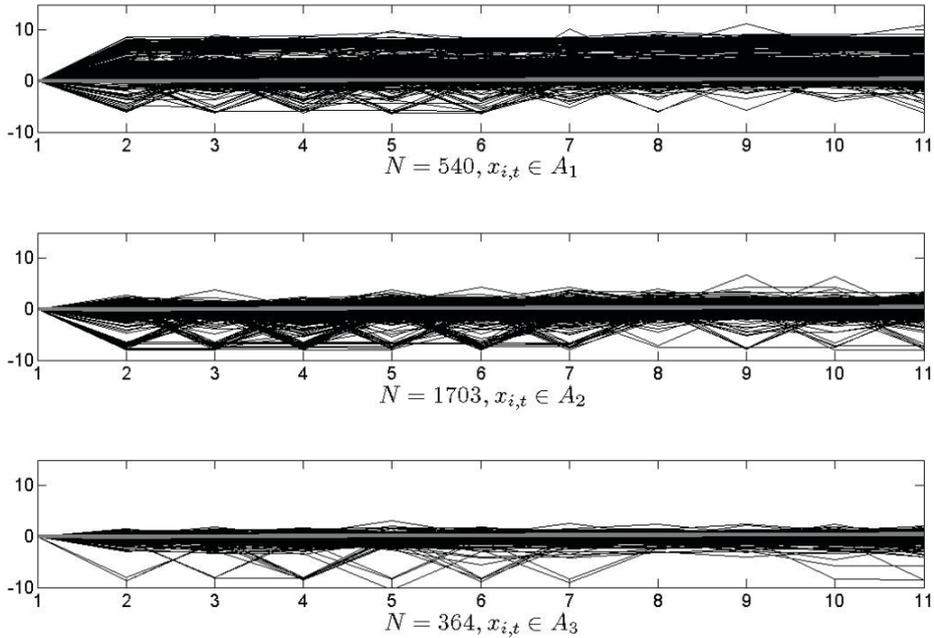
### 5.2.1. 자료

이제 경계자기회귀모형(TAR)을 구성하여 앞서 전이행렬을 통해 분석한 내용들을 더 자세하게 살펴보자. 상황을 간단하게 하기 위하여, 여기서는 총소득( $y_{i,t}$ )변수만을 사용한다. 성장률을 변수로 사용하기 위하여 log를 취하는데, 6종류의 소득이 모두 0이라고 보고한 경우 총소득을 강제로 1로 수정하였다. 1년간 총 소득이 정말로 0원 이라기보다는, 현실적으로 그 만큼 수입이 적었다는 의미로 파악하는 것이 옳으리라 생각되기에, 이 정도의 조작은 무방하리라 생각된다. 만약 이러한 조작을 하지 않으려면 11년의 소득자료 중 하나라도 0이 있는 가구는 표본에서 제외해야 하는데, 이럴 경우 정보의 손실이 상당하고, 실직상황을 포함한 저소득층의 동학을 올바르게 파악하지 못하는 등의 문제가 생길 수 있다. 따라서 자의적이지만 이와 같은 수정이 불가피하다.

$\log y_{i,t}$ 를 변수로 사용하는 것은 두 가지 이유 때문이다. 첫째는, 차분을 취할 경우 성장률 변수가 되어 안정성(stationarity)을 보장할 수 있다는 것이고,<sup>(6)</sup> 둘째는 경제를 구성하는 모든 가계의 소득이 같은 비율로 성장할 경우 소득불평등 측면에서 아무런 차이가 발생하지 않는 것이다. 이처럼 소득의 상대적인 구조만 문제시하는 것, 즉도의 0차 동차성은 소득불평등측도의 주요 공리 중 하나이다.

이제 새로운 표기와 변수를 도입하자.  $F_{i,t}$ 를  $t$ 기의 경제전체의 (누적)소득분포함수라고 하자.  $x_{i,t}$ 를  $i$ 가계의  $t$ 기의 위수(quantile)로 정의한다.

(6) 평균을 0으로 맞추기 위하여 이후 경제전체의 성장률도 뺀 것이다.



〈그림 2〉 1998년 계층을 기준으로 한 가계별 성장경로 그림

$$(5.1) \quad x_{i,t} \equiv F_t^{-1}(y_{i,t})$$

적절한  $0 < \gamma_L < \gamma_H < 1$ 에 대해,  $x_{i,t} \leq \gamma_L$ 이면 저소득층,  $\gamma_L < x_{i,t} < \gamma_H$ 이면 중산층,  $x_{i,t} \geq \gamma_H$ 이면 고소득층이라 하자.  $A_1 \equiv [0, \gamma_L]$ ,  $A_2 \equiv [\gamma_L, \gamma_H]$ ,  $A_3 \equiv [\gamma_H, 1]$ 로 각각 정의한다.

〈그림 2〉는  $\gamma_L = 0.22$ ,  $\gamma_H = 0.86$ 일 때  $x_{i,1} \in A_q$ 인 가구  $i$ -의 성장경로를  $q$ -번째 그림에 도시한 것이다.<sup>(7)</sup> 성장경로는  $(\log y_{i,t} - \log y_{i,1})_t$ 로 정의한다. 즉 1998년의 소득대비 얼마나 성장했는지를 파악하는 것이다. 첫 번째 그림의  $N = 540$ 은 1998년에 저소득층이었던 가계의 수가 540임을 의미하고, 그림의 개별 선 하나 하나는  $x$ -축의 시간을 따라서 각각 한 가계의 성장경로를 도시한 것이다. 즉 이 그림에는 540개 가계의 11년간 성장경로가 중첩되어있다. 중간 회색실선은 경제전체의 성장률, 즉 실질 GDP 성장률을 따르는 성장경로를 나타낸다. 로그차분 변수를 사용하므로 세로 값이 1 커지는 것은 소득이 대략 2.7배로 늘어남을 의미한다.

〈그림 2〉를 보면 소득이동성이 실질 GDP 성장이 거의 무시될 정도로 큼을 확인할

(7) 이후 모형을 추정할 결과가 0.22, 0.86이기 때문에 미리 그 값을 사용하여 그림을 그렸다.

수 있으며, 그 정도가 고소득층이 될수록 작아지는 것도 확인할 수 있다. 다시 말해 개별 검은 선분들이 상당히 위아래로 폭 넓게 움직이는데, 이는 매년 소득이 크게 늘었다가 크게 줄어드는 등 그 이동성이 상당히 큼을 의미한다. 또한 개별 가계에 대해 그 변동 폭이 매우 크기에 개별 가계에 대해서는 경제 전체의 성장경로인 회색실선이 큰 의미를 갖지 못하는 것이다. 그리고 첫 해 저소득층이었던 가계들의 성장경로가 중산층 이상의 그것보다 더 위아래로 폭 넓게 움직이고, 중산층이었던 가계들의 그것 또한 고소득층의 그것보다 더 폭넓게 움직임을 확인할 수 있다. 즉 전체적으로 개별 가계들의 소득이동성이 크기는 하지만 고소득층의 경우 그 값이 상대적으로 작고 이는 고소득층에 속한 사람들은 상당히 안정적으로 그들의 지위를 유지함을 나타낸다.

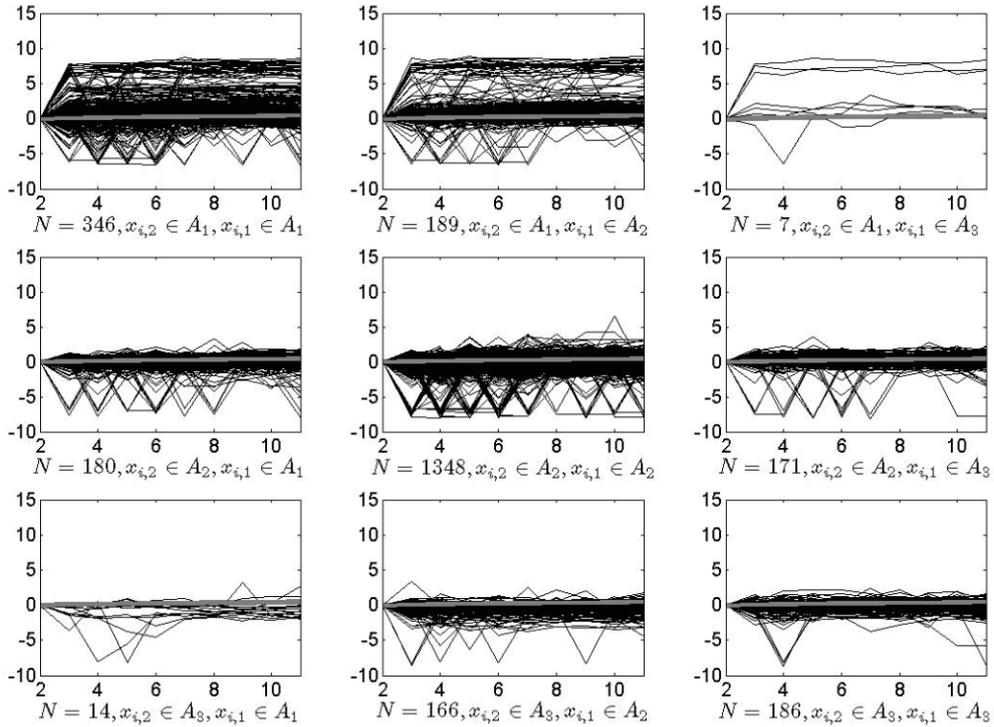
또한 그림에서 V자나 L자와 같은 형태로 바닥을 치는듯한 움직임은 당해 총소득이 0임을 의미한다.<sup>(8)</sup> 즉 실직을 한 것인데, 저소득층이나 중산층에 비해 고소득층의 경우 이러한 가계의 수가 훨씬 적고 그나마도 대부분 V자 형태로 바로 다음 해에 소득을 올리는 것을, 혹은 직장을 구하는 것을 확인할 수 있다. 즉 고소득층의 동학이 훨씬 안정적이다.

〈그림 3〉은 〈그림 2〉와 같은 것을 1998년과 1999년 두 기의 소득을 기준으로 그린 것이다. 다시 말해  $p$ 행  $q$ 열의 그림은  $x_{i,1} \in A_q, x_{i,2} \in A_p$ 인 가구  $i$ -들의 성장경로를 도시한 것으로, 이 경우 기준은  $\log y_{i,2}$ 이다. 다시 말해 1행 3열의 그림은 1998년에 고소득층이었지만 1999년에는 저소득층이 된 가계들의 소득이 향후 어떻게 변해갔는지를 도시해놓은 것이다. 이 그림에서 예컨대  $t=4$ 일 때 그 값이 2라면, 이는 해당 가계의 2001년 소득은  $e^2 y_{i,2}$ 임을 의미한다.<sup>(9)</sup>  $N=7$ 은 그러한 가계가 전체 표본 중 7가구에 불과함을 의미한다.

그림의 V자나 U자 형태의 움직임은 실직했다가 복직하는 과정을 나타내는데, 고소득층은 실직이 적으며, 실직 후에도 1년 만에 바로 복직하지만, 중산층이나 저소득층은 그렇지 않다. 즉 중산층이나 저소득층에서는 U자 혹은 L자 모양의 움직임이 관측된다. 또한 대부분의 성장경로가 W자 형태로 움직이는데, 전년도에 성장했으면 다음해에는 침체되는 패턴을 보인다. 이동성은 역시 고소득층, 특히 두 기 연속 고소득

(8) 앞서 서술하였지만 총소득을 0으로 보고한 경우 log 변수를 분석에 사용하기 위하여 그 값을 1로 바꾸었다. 따라서 log를 씌운 값은 0이다.

(9)  $\log y_{i,4} - \log y_{i,2} = 2 \Rightarrow y_{i,4} = e^2 y_{i,2}$ .



(그림 3) 1998년과 1999년도 계층을 기준으로 한 가계별 성장경로 그림

층이었던 경우가 제일 작으며, 저소득층으로 갈수록 커진다. 즉 그림의 우하단에서 좌상단으로 올수록 그 이동성이 크다. 또한 1행의 그림들에서는 급격하게 뛰어오르는 움직임이 관찰되는데, 이는 실직 때문에 저소득층이 되었다가 1년 후 복직하는 사람들을 의미한다. 크게 보면 5 정도로 뛰어오르는 집단과 10 정도로 뛰어오르는 두 집단으로 나뉘는데, 중산층으로 복귀하는 가구들과 고소득층으로 복귀하는 가구들을 나타내는 것이라고 해석할 수 있다. 이러한 움직임들, 즉 실직상황에 빠진 사람들의 움직임까지 잡아내려면 경계(threshold)를 늘리는 것도 좋은 방법이 될 것이다. 그러나 본고에서는 표본 수에 대한 우려 때문에 이를 행하지는 못하였다.

5.2.2. 모형

자료를 살펴보았으니 이제 모형을 세워 이를 분석하자.  $g_t$ 를 실질 GDP 성장률이라 할 때,  $z_{i,t}$ 를 다음과 같이 정의한다.

$$(5.2) \quad z_{i,t} \equiv \log y_{i,t} - \log y_{i,t-1} - g_t$$

여기서  $z_{i,t}$ 는  $i$ -가구의  $t$ 기 성장률에서 경제 전체의 성장률을 뺀 값이다. 즉 소득에 대한 공통의 충격을 뺀 부분이라고 생각할 수 있다. 총 11기간의 자료가 있으므로, 한 가구당 10개의  $z_{i,t}$ 값을 계산할 수 있고, 따라서 26,070개( $2,607 \times 10$ )의 자료를 사용할 수 있다.

이제 다음과 같은 모형을 생각한다.

$$(5.3) \quad z_{i,t+1} = \sum_{p,q=1,2,3} I_{p,q}(x_{i,t}, x_{i,t-1}) \\ \times (\beta_{0,p,q} + \beta_{1,p,q} D_{i,t} + \beta_{2,p,q} z_{i,t} + \beta_{3,p,q} z_{i,t-1} + \varepsilon_{i,t,p,q}),$$

$$I_{p,q}(x_{i,t}, x_{i,t-1}) = \begin{cases} 1, & x_{i,t} \in A_p, x_{i,t-1} \in A_p, \\ 0, & o.w. \end{cases}$$

단,  $A_1 \equiv [0, \gamma_L], A_2 \equiv (\gamma_L, \gamma_H), A_3 \equiv [\gamma_H, 1]$

$$D_{i,t} = \begin{cases} 1, & t = 1, 2, 4, 8 \\ 0, & o.w. \end{cases},$$

$$\varepsilon_{i,t,p,q} \sim N(0, \sigma_{p,q}^2).$$

모형을 순서대로 살펴보자. 기본적으로 위의 모형은 전기( $t$ )와 그 전기( $t-1$ )에 경제의 평균 성장치보다 얼마나 소득이 증가했는지, 즉 ‘초과’성장치(경제의 평균 성장치를 ‘초과’하여 성장하는 부분)의 두 과거 값( $z_{i,t}, z_{i,t-1}$ )에 의해 이번기의 ‘초과’성장치( $z_{i,t+1}$ )가 결정되는 모형이다. 다만 그 관계가 전기와 그 전기에 어떤 계층에 있었느냐에 의해서도 영향을 받는다. 이를 위해  $I_{p,q}$ 를 구성하는데, 이 변수는 전기에  $x_{i,t}$ 가  $A_p$ 의 원소이고, 그 전기에  $x_{i,t-1}$ 이  $A_q$ 의 원소였을 때에만 1이 되는 값이다. 즉 전기에 중산층, 그 전기에 고소득층이었다면  $I_{2,3}$ 은 1이고, 다른  $(p, q)$ 에 대해  $I_{p,q} = 0$ 이 된다. 이때  $A_p$ 를 결정하는  $\gamma_L$ 과  $\gamma_H$ 도 모수화하여 모형에 반영하였다. 저소득층과 중산층을 구분 짓는 모수가  $\gamma_L$ , 중산층과 고소득층을 구분 짓는 모수가  $\gamma_H$ 이므로, 결국 어느 지점에서 이동성의 동학이 질적인 변화를 보이든지도 자료에서 추정하겠다는 뜻이다. 이러한 접근을 취하는 이유는 저소득층과 중산층, 그리고 고소득층을 구분하는 통일된 기준이 없기 때문이다. 연구자에 따라, 목적에 따라 다른 기준들이 사용되고 있기 때문에, 아예 이 값도 모수화하여 자료를 가장 잘 설명하는 값으로 추정하는 편이 낫

다고 생각하였다.

$D_{i,t}$ 는 불황더미이다. 3절에서 언급했다시피 1998, 1999, 2001, 그리고 2005년의 4개년이 불황이고, 이때  $D_{i,t}$ 는 1, 그렇지 않은 경우 0이다.  $\beta_{1,p,q}$ 를 추정하면 불황이 계층별로 상이한 영향을 미치는지 여부를 분석할 수 있다.

패널자료를 다룰 때는 오차항의 분산에 큰 관심을 기울여야 한다. 위의 모형에서  $\varepsilon$ 은  $i, t$ 에 대해서는 *iid*이고,  $p, q$ 에 대해서는 독립이라 가정하는데, 오차항이 위와 같은 모형을 따르도록 하기 위하여 집단 공통의 충격들을 전부 제거하는 방법을 사용하였다. 예를 들어  $z$ 변수를 구성할 때는 경제전체의 성장 부분을 외생적으로 제거하고,  $\beta_{0,p,q}$ 를 포함하여 같은 계층 구성원끼리 공유하는 부분 또한 제거하였다. 또한 자기상관부분은 AR(2) 형태의 모형을 가정하여 해결하였다. 즉 오차항의 독립성에 위배될만한 부분들은 전부 오차항 밖으로 빼내어 제거하였다.<sup>(10)</sup> 이러한 방법으로 위와 같은 오차항의 구조를 가정하는 이유는, *iid* 정규분포 표본을 가정할 경우 그 추정과 검정이 상당히 편해지기 때문이다.

일반적으로 TAR 모형을 쓰는 경우 SETAR(self-exciting TAR)를 많이 쓴다. 즉 경계(threshold)를 구성할 때,  $z$ 의 과거 값을 기준으로 삼는 것이다. 그러나 본고의 맥락에서는 성장률 자체( $z$ )보다는 저소득층, 혹은 고소득층의 지위가 어떠한 식으로 유지되는지가 더 본질적인 물음이다. 결국 소득구조를 연구하면서 궁극적으로 관심을 갖는 변수는 성장률을 통해 계산된 실제 소득이지 성장률이 아닌 것이다. 따라서 성장률보다는  $x_{i,t}$ 를 경계를 구성하는 변수로 사용하는 편이 더 타당하고 해석하기도 좋다고 생각하여 위와 같은 경계구조를 상정하였다.

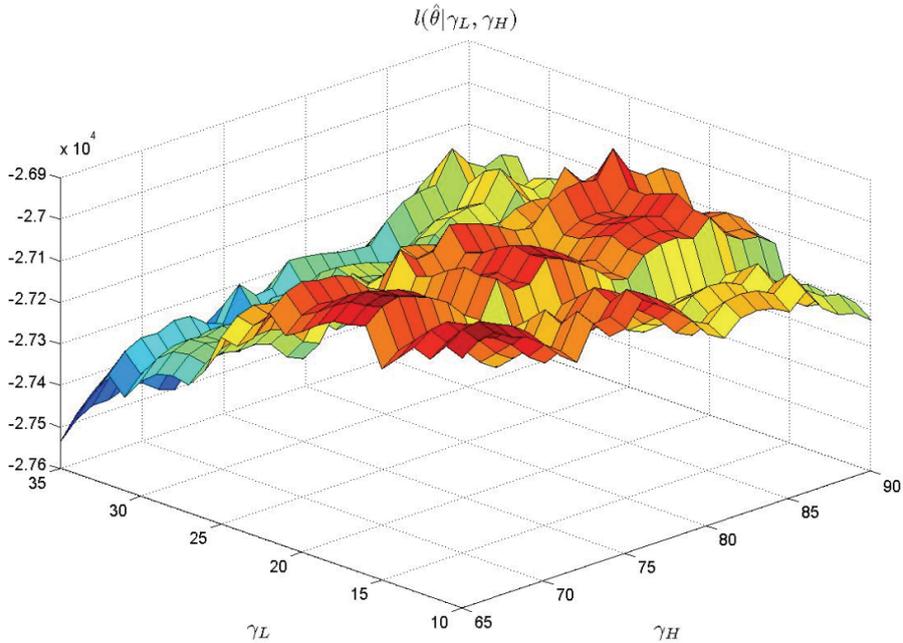
### 5.2.3. 추정 방법

이제 추정의 문제를 생각하자. 위 모형의 모수는 다음과 같다.

$$(5.4) \quad \theta \equiv (\gamma, \beta, \Sigma) \in \mathbb{R}^2 \times \mathbb{R}^{4 \times 3 \times 3} \times \mathbb{R}^{3 \times 3},$$

$$\text{단, } \gamma \equiv (\gamma_L, \gamma_H), \beta \equiv (\beta_{j,p,q})_{j=0,1,2,3}, \Sigma \equiv (\sigma_{p,q}^2)_{p,q=1,2,3}.$$

(10) 소득분배구조와 경제성장의 관계를 규명하는 것 또한 매우 중요한 작업인데, 소득구조가 경제성장에 직접적인 영향을 준다면 위와 같은 방법이 문제를 일으킬 소지가 있다. 그러나 필자의 능력으로 소득구조와 경제성장의 상호작용을 규명하는 것은 아직 불가능하기에, 본고에서는 아예 성장부분을 빼버렸다. 위에서 다루는 작은 모형만으로 경제성장과 분배를 한꺼번에 다루기가 버거웠기 때문이다. 그러나 이 문제는 분명 중요한 문제이고, 규명되어야 할 문제이다. 양자의 상호작용을 밝혀낸다면 성장과 분배 중 무엇을 택할지에 대한 오랜 갈등에 대해 경제학적인 해결책을 제시할 수 있기 때문이다.



〈그림 4〉  $\log L(\hat{\theta} | \gamma)$

위 모수의 최우추정(MLE)은 두 단계로 행해진다. 우선 각  $\gamma$ 에 대하여  $L(\theta | \gamma)$ 를 최대화하는  $\hat{\beta}, \hat{\Sigma} | \gamma$ 를 찾는다. 이렇게  $\gamma$ 가 주어져 있을 때 (조건부)가능도함수의 최대값들을 구하는 것이 첫 번째 단계이고, 이 중 그 값이 가장 커지는  $\gamma$ 와 그때의  $\hat{\beta}, \hat{\Sigma} | \gamma$ 를 찾는 것이 추정의 두 번째 단계이다. 위의 두 단계 접근을 이용하면  $\hat{\beta}$ 을 구할 수 있는데, 두 번째 단계는 격자탐색(grid search)법을 이용하면 되므로, 문제는 첫 번째 단계에서  $\hat{\beta}, \hat{\Sigma} | \gamma$ 를 구하는 일이 된다. 이 문제에 대해서는 다음과 같이 접근한다. 우선  $\gamma_L$ 과  $\gamma_H$ 에 특정 값을 부여하면 위의 모형에서  $I_{p,q}(x_{i,t}, x_{i,t-1})$ 에 해당하는 부분을 0 아니면 1로 확정할 수 있고, 이 경우 (조건부)가능도비함수가 다음과 같이 구성된다.

$$(5.5) \quad L(\theta | \gamma, Z) = \prod_{p,q} \prod_{I_{p,q}=1} \phi \left( \frac{z_{i,t+1} - \beta_{0,p,q} - \beta_{1,p,q} D_{i,t} - \beta_{2,p,q} z_{i,t} - \beta_{3,p,q} z_{i,t-1}}{\sigma_{p,q}} \right),$$

단  $\phi$ 는 표준정규분포의 확률밀도함수,  $Z \equiv (z_{i,t})_{i,t}$

이제  $\hat{\beta}, \hat{\Sigma} | \gamma$ 를 찾는 문제는, 각  $p, q$ 에 대하여

$$\prod_{I_{p,q}=1} \phi \left( \frac{z_{i,t+1} - \beta_{0,p,q} - \beta_{1,p,q} D_{i,t} - \beta_{2,p,q} z_{i,t} - \beta_{3,p,q} z_{i,t-1}}{\sigma_{p,q}} \right)$$

를 최대화하는 모수들을 찾는 문제로 바뀐다.  $\varepsilon$ 이  $i, t$ 에 대해서는 *iid*이고,  $p, q$ 에 대해서는 독립이라 가정했기 때문에 이러한 접근이 가능하다. 이때 주어진  $p, q$ 에 대해 오류항이 *iid* 정규분포임을 이용하면  $\beta_{0,p,q}, \beta_{1,p,q}, \beta_{2,p,q}, \beta_{3,p,q} | \gamma$ 를 최소잔차추정법 (Ordinary Least Square)으로 쉽게 구할 수 있다. 또한 잔차를 이용하면  $\sigma_{p,q}^2 | \gamma$ 도 쉽게 계산할 수 있다. 이러한 방법을 통해 모형을 추정한 결과는 다음과 같다.

우선  $\hat{\gamma}_L = 0.22, \hat{\gamma}_H = 0.88$ 이다.  $\gamma$ 를 추정할 때  $\gamma_L$ 에는 10%부터 1%간격으로 35%까지 격자를 주었고,  $\gamma_H$ 에는 65%부터 1% 간격으로 90%까지 격자를 주었는데 각 경우에 최대화된 (조건부)로그 가능도함수값들은 <그림 4>에 도시되어있다.

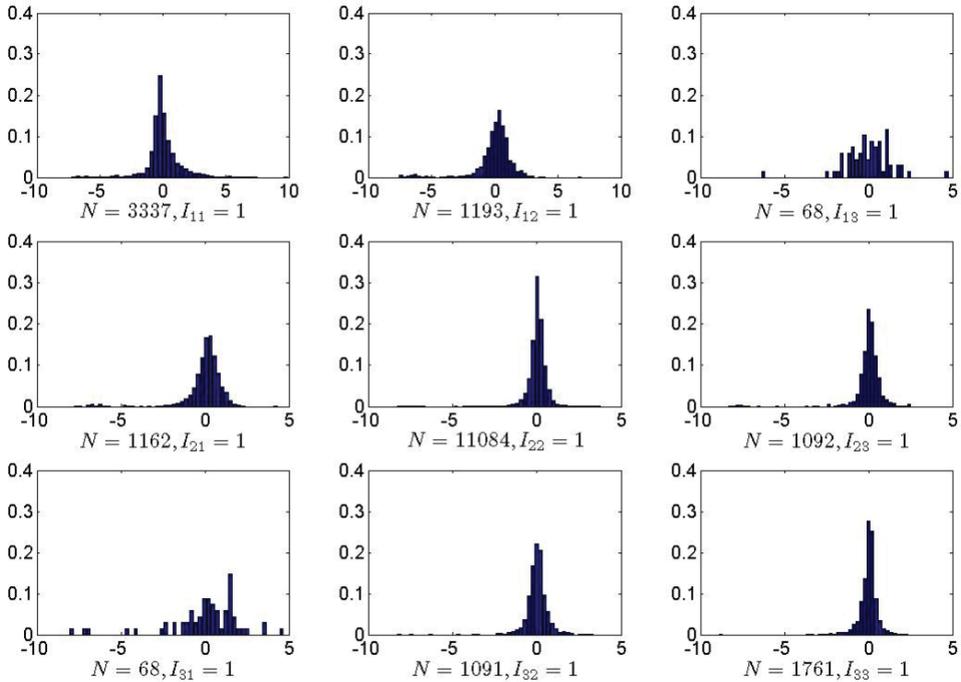
5.2.4. 추정 결과

<표 9>에  $\hat{\beta}, \hat{\Sigma}$  값들과 표본 수를 정리해놓았는데, 이때 표본 수는 “ $I_{p,q}(x_{i,t}, x_{i,t-1}) = 1$ ”를 만족하는 순서쌍  $(i, t)$ 의 개수를 의미한다. 최대화된 로그가능도 함수 값은 -26,933.9이다.

변수들의 기본단위가 로그차분임을 감안할 때 오차항의 분산에 대한 추정 값들은 앞서 살펴본 것처럼 큰 이동성을 반영한다. 그리고 그 정도가 계층에 따라 다른 것도 쉽게 확인할 수 있는데, 눈에 띄는 사실은  $\hat{\sigma}_{3,3}^2$  값이 가장 작다는 것이다. 이는 두기 연속 고소득층이었던 가구들이 가장 안정적으로 그 위치를 유지함을 의미한다. 또한 극단적인 변화를 경험한데다가 그 표본수가 매우 작아 안정적인 추론이 힘든  $\hat{\sigma}_{3,1}^2$ 를

<표 9>  $\hat{\beta}, \hat{\Sigma}$

$p$	$q$	표본 수	$\hat{\beta}_{0,p,q}$	$\hat{\beta}_{1,p,q}$	$\hat{\beta}_{2,p,q}$	$\hat{\beta}_{3,p,q}$	$\hat{\sigma}_{p,q}^2$
1	1	3,336	0.278	-0.149	-0.553	-0.272	2.128
1	2	1,165	-0.421	-0.205	-0.843	-0.070	2.412
1	3	65	-0.621	0.644	-0.779	-0.435	2.009
2	1	1,137	-0.145	-0.111	-0.155	-0.086	1.525
2	2	11,108	-0.038	-0.072	-0.334	-0.034	0.541
2	3	1,095	0.010	-0.157	-0.072	-0.090	1.020
3	1	66	-1.024	0.056	-0.400	-0.262	5.010
3	2	1,094	0.155	-0.004	-0.928	-0.062	0.536
3	3	1,760	-0.123	-0.058	-0.539	-0.178	0.302

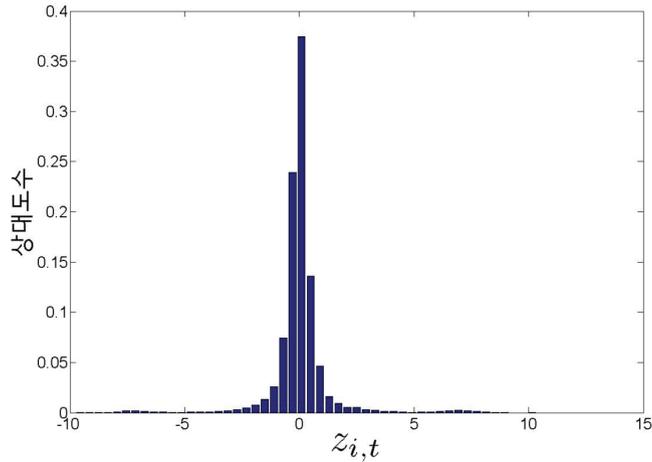


〈그림 5〉 잔차 $_{p,q}$ 의 히스토그램

제외하면, 고소득층이 가장 그 이동성이 적고, 계층이 내려가면서 그 이동성이 점점 커지는 것을 확인할 수 있다.

자기회귀(Auto-regression) 계수들인  $\widehat{\beta}_{2,p,q}, \widehat{\beta}_{3,p,q}$  이 음수라는 사실은 〈그림 2〉, 〈그림 3〉에서 살펴본 W자형태의 움직임을 반영하는 것이라 생각된다. 즉 기본적으로 “평균으로의 회귀”현상이 발생하는 것으로 보이는데, 계층구조와 상관없이 초과 성장한 가구는 다음 기에는 과소성장하고 그 반대의 경우도 성립한다는 의미로 해석할 수 있다.

$\widehat{\beta}_{1,p,q}$ 은 불황 더미변수의 계수이다. 표본 수가 적어 신뢰성이 떨어지는 값들인  $\widehat{\beta}_{1,1,3}, \widehat{\beta}_{1,3,1}$ 을 제외하면 그 값이 모두 음수인데, 흥미로운 것은 계층이 내려갈수록 그 영향이 커진다는 점이다. 고소득층은 불황이라고 해서 특별한 반응을 보이지 않지만, 저소득층으로 내려갈수록 민감하게 반응하는 것은 흥미로운 현상이다. 각주 6에서 살펴보았듯이 지니계수만 보면 불황기에 소득구조가 개선되는 것처럼 보이지만, 위의 추정결과는 이것이 그렇게 단순히 생각할 수 있는 문제가 아님을 시사하고 있다. 또한  $\widehat{\beta}_{1,p,q} = 0$ 을 가정하고 추정한 최대화된 로그가능도 함수 값(-28,400)을 이용하면

〈그림 6〉  $z_{i,t}$ 들의 히스토그램

가능도비검정(Likelihood ratio test)을 할 수 있는데, 이 경우 검정통계량은 2,932.3이고,  $p$ -값은 0이다. 즉  $D_{i,t}$ 는 유효한 설명변수이다.

마지막으로 잔차를 살펴보자. 〈그림 5〉는 위 모형을 추정하고 남은 잔차들의 히스토그램이다. 모든 경우 분포가 상당한 정도로 중앙에 집중되어 있으나 그 꼬리도 상당히 두터워 극단적인 값들이 꽤 발생한다. 대부분의 경우 조건부 평균 근처에서  $z_{i,t+1}$ 값이 형성되나 극단적인 값이 형성될 확률도 꽤 되고, 그 침도가 매우 큰 분포이다. 이런 모양은 흔히 사용하는 정규분포,  $t$ -분포, 로지스틱분포 등에서는 볼 수 없는데, 실제로 위의 세 분포함수를 가정한 최우추정은 그 결과가 그다지 히스토그램과 상응하지 않는다. 분석의 편의를 위해서 정규분포가정을 하였지만, 〈그림 5〉는 그것이 틀렸을 수도 있음을 시사한다.

이러한 현상이 발생하는 가장 근본적인 원인은  $z_{i,t}$ 들이 비슷한 분포를 갖기 때문인데, 〈그림 6〉과 〈표 10〉에 이것이 잘 나타나 있다.<sup>(11)</sup>

(11) 보다 흥미로운 것은 교환이동도 비슷한 분포를 따른다는 사실이다. 교환이동은 주어진 분포 하에서 소득 순위가 변하는 이동이다. 즉  $t$ 기와  $t+1$ 기 사이의 교환이동은  $(x_{i,t+1} - x_{i,t})$ 로 파악할 수 있다. 즉 하위 10%에서 이듬해 하위 40%가 된 가계의 경우 1년간 30%만큼의 교환이동을 겪었다고 할 수 있다. 자료에 나타나는 교환이동의 그림을 그려보면 그 형태가 〈그림 5〉, 〈그림 6〉과 흡사하고, 따라서 흔히 사용되는 분포들로는 근사하기가 힘들다. 따라서 소득이동을 다루는데 적절하며 정규분포처럼 쉽지는 않더라도 다루기 쉬운 분포함수를 개발할 필요가 있어 보인다.

〈표 10〉  $z_{it}$ 의 기초 통계량

평균	분산	왜도	첨도
0.0250	1.6610	0.3604	21.7095

## 6. 결론

지금까지 한국의 노동, 금융, 부동산 소득 자료를 이용하여 소득의 정태적, 동태적 특성을 분석하였다. 위수회귀, 전이행렬, 그리고 경계자기회귀모형 등의 모형과 기법을 사용하였고 다음의 결과를 얻었다.

소득의 구성과 관련하여서는 금융소득과 부동산소득을 주의 깊게 살펴보았는데, 지니계수와 점유비율에 대한 분석 결과 이들 소득이 임금소득에 비해서 상당히 불평등하게 분배되어있으며, 부동산소득이 그 정도가 더 심함을 확인할 수 있었다. 또한 위수회귀 분석 결과 고소득층을 고소득층에게 하는 것은 부동산과 금융소득이며 중산층에서 고소득층으로 넘어갈수록 부동산소득의 중요성이 더욱 커짐을 확인할 수 있었다. 바꿔 말하면 저소득층은 이러한 소득을 올리지 못하고 오직 임금소득만 받는다는 뜻이다. 금융자산이든 부동산이든 본질적으로 투자할 수 있는 위험자산이라는 측면에서는 동일하기 때문에 이러한 결과는 상당히 흥미롭다. 부동산자산의 특수성에 의한 정보제약이나 투자를 위한 자본조달비용의 문제 때문에 부동산을 통해 얻을 수 있는 수익이 고소득층에게 집중되고 있는 것이 아닌지 의심스럽다. 또한 많은 사람들이 우려하듯이 본고에서 고려한 기간 동안에는 임금소득의 불평등성이 지속적으로 증가하였음을 확인하였다. 다만 2008년 이후의 자료는 분석하지 못하였기 때문에 전세계적인 금융위기 이후에 생긴 변화는 확인할 수 없었던 점이 아쉽다.

경계자기회귀모형을 이용하여 계층별 이동성의 경향과 그 크기를 분석하였는데, 자료를 통해 살펴보았듯이 연간 소득이동성은 상당히 크지만, 그럼에도 불구하고 고소득을 올리는 데에는 지속성이 있음을 보았다. 즉 고소득층은 이동성의 정도에 대해서 다른 계층과 질적으로 다른 동학을 가지고 있는 것이다. 이는 실직에 대해서도 마찬가지인데, 고소득층은 실업의 빈도가 작고, 실업하더라도 다음해에 다시 직장을 얻지만, 중산층 이하의 경우 그렇지 않았다. 또한 전기와 그 전기 성장률을 기준으로 집단을 나누었을 때, 각 집단 내에서는 평균으로의 회귀가 관측되는데, 다시 말해 평

균보다 더 많이 소득이 증가하였을 경우 이듬해 지속적으로 증가하는 것이 아니라 줄어드는 것이다. 이러한 현상의 원인 중 일시적으로 거둔 부동산 및 금융자산소득의 부침이 차지하는 비중이 얼마나 되는지는 추후 확인해보아야 할 것이다. 또한 하위 22%와 88% 지점이 경곶값으로 추정되었는데, 저소득층과 고소득층의 비율이 비대칭적인 것도 흥미로운 사실이다. 마지막으로 중산층과 저소득층이 불황에 더 민감하게 반응함을, 다시 말해 더 많은 피해를 봄을 확인할 수 있었다. 이 결과는 거시경제적 경기안정화정책이 소득불평등을 완화하는 효과를 가질 수도 있음을 시사하기에 흥미롭다.

본고에서는 한국노동패널자료를 이용하여 한국의 소득구조에 대한 몇 가지 사실들을 살펴보았다. 앞서 위수회귀와 경제자기모형을 분석할 때 사용한 설명변수로서 주로 과거 값만을 사용했는데, 이는 상당한 단순화라고 할 수 있을 것이다. 설명력이 있는 다른 변수들을 찾고 해석하는 과정은 앞으로의 과제로 남겨둔다. 경제자기모형에서 서로 독립인 정규분포오차를 가정한 것 또한 잔차항에 대한 분석을 통해서 한계가 있을 수 있음이 드러났기에, 다루기 쉬우면서 현실과 합치하는 분포를 개발하는 것도 중요한 과제이다. 마지막으로 경험연구를 통해 알아낸 사실들이 발생하는 내적인 이유를 밝혀줄 수 있는 경제학 모형을 구성하고 풀어내는 일이 중요한데, 그래야만 정책함의가 담긴 주장들을 설득력있게 제시할 수 있을 것이기 때문이다.

교신저자, 서울대학교 경제학부 교수

151-746 서울특별시 관악구 대학동 관악로1

전화: (02)880-6390

E-mail: jykim017@snu.ac.kr

서울대학교 경제학부 석사과정

151-746 서울특별시 관악구 대학동 관악로1

E-mail: bc1105@snu.ac.kr

## 참고문헌

- 국가통계포털(KOSIS), [www.kosis.kr](http://www.kosis.kr)
- 한국노동패널(KLIPS), [www.kli.re.kr](http://www.kli.re.kr)
- 양춘(2000): 『한국사회: 계층구조와 동학』, 서울, 고려대학교 출판부.
- 李俊求(1989): 『所得分配의 理論과 現實』, 서울, 茶山出版社.
- 홍두승, 구해근(2001): 『사회계층·계급론』, 제2판, 서울, 다산출판사.
- Atkinson, A.B.(1981): “The measurement of economic mobility,” in Atkinson, A.B. (Ed.), *Essays in Honor of Jan Pen*, Reprinted in *Social Justice and Public Policy*, Brighton, Wheatshef Books Ltd.
- Bernasconi, M., and V. Dardanoni(2004): “An Experimental Analysis of Social Mobility Comparisons,” in Cowell, F.(Ed.), *Inequality, welfare and income distribution: Experimental approaches*, Greenwich, Connecticut, JAI Press, 55-84.
- Campano, F., and D. Salvatore(2006): *Income distribution*, New York, Oxford University Press.
- Cermeño, R., and K.B. Grier(2001): “Modeling GARCH processes in panel data: Theory, simulations and examples,” unpublished working paper, <http://www.ou.edu/cas/econ/research/2001-3.pdf>.
- Corak, M., and A. Heisz(1999): “The Intergenerational Earnings and Income Mobility of Canadian Men: Evidence from Longitudinal Income Tax,” *The Journal of Human Resources*, **34**, 3, 504-533.
- Deaton, A.(1997): *Analysis of Household Surveys*, Baltimore MD, Johns Hopkins University Press.
- Fields, G.S., and E.A. Ok(1999): “The Measurement of Income Mobility: An Introduction to the Literature,” in Silber, J.(Ed.), *Handbook of income inequality measurement*, Boston, Kluwer academic publishers, 557-598.
- Geweke, J., and M. Keane(2000): “An empirical analysis of earnings dynamics among men in the PSID: 1968-1989,” *Journal of Econometrics*, **96**, 293-356.
- Janvry, A., and E. Sadoulet(1996): “Growth, Inequality, and Poverty in Latin America: A Casual Analysis, 1970-94,” Working paper, 784.

- Jarvis, S., and S.P. Jenkins(1998): “How Much Income Mobility Is There In Britain?” *The Economic Journal*, **108**, 428-443.
- King, M.A.(1983): “An Index of Inequality with Applications to Horizontal Equity and Social Mobility,” NBER Working Paper, **1468**.
- Koenker, R., and K.F. Hallock(2001): “Quantile Regression,” *Journal of Economic Perspectives*, **15**, 4, 143-156.
- Markandya, A.(1982): “Intergenerational Exchange Mobility and Economic Welfare,” *European Economic Review*, **17**, 307-324.
- Markandya, A.(1984): “The Welfare Measurement of Changes in Economic Mobility,” *Economica*, **51**, **204**, 457-471.
- Pakel, C., Shephard, N., and K. Sheppard(2009): “Nuisance parameters, composite likelihoods and a panel of GARCH models,” *Economics Series Working Papers* **458**, University of Oxford, Department of Economics.
- Shorrocks, A.F.(1976): “Income Mobility and the Markov Assumption,” *The Economic Journal*, **86**, 566-578.
- Shorrocks, A.F.(1978a): “The Measurement of Mobility,” *Econometrica*, **46**, **5**, 1013-1024.
- Shorrocks, A.F.(1978b): “Income Inequality and Income Mobility,” *Journal of Economic Theory*, **19**, 376-393.
- Tsay, R.S.(1989): “Testing and Modeling Threshold Autoregressive Processes,” *Journal of the American Statistical Association*, **84**, **405**, 231-240.

**Abstract**

## Dynamic Analysis of Income Composition and Transition of Korea

Jae Young Kim and ByungChan Lee

We analyze dynamic pattern of income composition and transition for Korean panel data from 1998 to 2008. We adopt the quantile regression, the transition matrix method and the threshold autoregression model for the analysis. For the analysis of income composition we focus incomes from financial assets and real properties. For the analysis of income transition we focus on the transition between income groups. Our results show some interesting features of income composition and transition of Korea in the period.

**Keywords:** Income composition, Transition, Dynamic pattern