

# KLIPS에 근거한 주거력 데이터 구축 및 활용

이 경 희\* · 민 인 식\*\*

패널조사 데이터는 기본적으로 동일 대상을 한 번 이상의 여러 차수에 걸쳐 관찰하고 추적 조사한 내용을 토대로 하기 때문에 조사대상의 이동, 상태변화와 같은 유용한 정보를 제공하지만, 이러한 변화이력을 분석용으로 데이터화하는 것은 상당한 전문성과 치밀함을 요구한다. 이에 한국노동패널조사(KLIPS)는 이미 직업력을 따로 구축하여 별도 데이터세트로 제공하고 있으며 이는 다양한 실증분석 연구에 활용되고 있다. 본 연구에서는, 이러한 직업력과 유사한 형태로, KLIPS에 근거한 조사가구의 주거력 데이터를 새로 구축하여 연구자들에게 제공함으로써 조사의 활용도를 더욱 높이는데 기여하고자 한다. 가구의 주거력을 기준으로 만든 주거력 데이터를 long-type으로 변환한 후, 가구특성 데이터와 병합된 주거패널데이터와 병합함으로써 만들어지는, 가구특성과 주거관련 변수를 포함한 주거력 데이터는 노동경제뿐만 아니라 도시계획, 주거학, 부동산, 교육, 아동복지 등의 분야에서도 유용한 자료가 될 것이다. 한편, 본 연구에서는 복잡다단한 구축과정을 거쳐 만들어진 최종 주거력 데이터의 완성도 및 활용가능성을 검토하는 차원에서, 기본적인 거주기간 분석을 수행하였다. 반복사건을 고려한 frailty 생존분석 모형을 추정한 결과, 가구주의 나이와 교육수준, 가구소득, 고등학교 이하 자녀 수, 비광역시 거주 변수는 특정 주거이력에서의 거주기간을 증가시키는 양(+)의 효과를 보인 반면, 기혼유배우 또는 전·월세 가구의 경우에는 거주기간과 부(-)적인 관계를 나타냈다. 또한 로그로지스틱 분포- frailty Gamma 분포 모형의 추정결과에 근거하여 가구주 연령에 따른 해저드 함수를 그려보면, 예상대로 연령이 높아질수록 해저드가 낮아지는 형태로 나타나는데, 특정 주택에서의 거주기간이 10년 이상 오래 지속되면 가구주 연령과 상관없이 이사발생 해저드가 낮아진다. 저자들은 본 연구를 통해 제공되는 KLIPS 주거력 데이터가 다양한 분야의 연구에서 활용되기를 기대한다. 그리고 구축과정에 대한 단계별 설명 자료는 유사한 형태의 이력 데이터 - 예를 들어 혼인력 - 를 만들 때 참고 가능한, 실질적으로 도움이 되는 기초 가이드라인으로서의 활용가능성이 높을 것으로 본다.

## 1. 서 론 및 선행연구

패널조사 데이터는 기본적으로 동일한 개체, 예를 들어 개인이나 가구, 사업체 등을 대상으로 여러 번에 걸친 반복·추적 조사를 통해 구성된다. 따라서 일회성의 횡단면조사 데이터에 비해 여러 가지 효과 추정 시 개체고유의 특성을 통제할 수 있는 방법론적 이점을 가지고 있을 뿐만 아니라, 특히, 이동, 변화 등과 같은 동학(dynamics)을 파악하는데 유용한 정보를 제공한다.

---

\* 한국노동연구원 연구위원 (kheele@kli.re.kr)

\*\* 경희대학교 경제학과 교수 (imin@khu.ac.kr)

그러나 패널데이터를 한 번이라도 사용해본 경험이 있는 사람이라면 조사대상 개체의 동적 이력 (career or history) 정보를 데이터화하는 것이 생각보다 어려운 작업이라는 것을 알 것이다. 경우에 따라서는 이러한 데이터 구축 상의 문제 때문에 분석 자체를 하지 못하는 경우도 있을 수 있다. 때문에 한국노동연구원(Korea Labor Institute: 이하 KLI)의 한국노동패널조사(Korean Labor and Income Panel Study: 이하 KLIPS)와 같은 패널데이터에서는 조사대상 개인의 직업력과 같은 이력 정보를 정리하여 별도의 데이터로 제공해준다. 본 연구에서는, 직업력과 유사한 형태로, KLIPS에 근거한 조사가구의 주거력 데이터를 구축하는 방법을 단계별로 상세히 설명하고, 조사의 활용도를 높이기 위해 최종산출물인 주거력 데이터를 연구자들에게 제공하고자 한다. 또한 구축된 주거력 데이터의 활용가능성에 대한 예시로서 거주기간에 대한 생존분석을 수행하고 그 결과를 제시한다.

개인이나 가구의 주거이동성(residential mobility)과 관련된 선행연구들을 살펴보면 크게 두 가지 흐름이 발견된다. 하나는 주거환경학 또는 도시생태학적 관점에서 주거이동에 대한 결정요인을 파악하고 개인이나 가구의 주거조정 관련 의사결정행동을 분석·예측하고자 하는 연구의 흐름이다. 다른 하나는 주거이동에 의해 유발되는, 긍정적 또는 부정적 결과(outcome)를 이해하고자 하는 연구와 관련된다(Beck, Buttaro & Lennon, 2016; Gambaro & Joshi, 2016; Lupton, 2016 등). 전자는 1980년대 이래 국내외의 많은 연구자들에 의해 꾸준히 연구되어 온 주제인 반면, 후자는 1990년대에 시작되어 새로운 관심을 받고 있는 분야이다(Lennon, Clark & Joshi, 2016). 본 연구에서 제시하고자 하는, KLIPS의 주거력 데이터를 활용한 거주기간 분석은 전자와 관련된다. 본 연구와 같이 특정 주거지에서의 거주기간에 대한 결정요인을 생존분석 방법론을 이용하여 분석한 연구들로는 Withers(1997), van der Vlist, Gorter, Nijkamp & Rietveld(2002), Decoster, Capéau & De Swertdt(2005), 최은선·남진(2011), 이창호·이승일(2012) 등이 있다.<sup>1)</sup> 이들 선행연구 결과에 따르면, 가구주 연령·학력·직업·성별·인종, 통근거리, 가구소득, 가구유형, 자녀관련 요인, 주택종류, 주택규모, 점유형태, 주택가격/임대료, 지역, 주택거래세율, 이자율 등의 가구 및 주택 특성, 환경, 정책 요인들이 거주기간과 유의한 관계를 보이는 것으로 나타났다. 대체적으로, 가구주 연령이 높을수록, 여성이거나 소수인종인 경우, 고등학생 이하 자녀가 있는 경우, 주택종류가 아파트 대비 비아파트인 경우, 점유형태가 임차 대비 자가인 경우, 그리고 주택거래세율이 높을수록, 특정 주거지에서의 거주기간이 길어지는(이주확률이 낮아지는) 양(+)의 효과를 보였다. 반면, 가구주 학력이 높을수록, 주택가격이나 임대료가 높을수록, 특정 주거지에서의 거주기간이 짧아지는(주거이동 확률이 높아지는) 음(-)의 효과를 가지는 것으로 추정되었다. 가구주 직업, 가구소득, 주택규모, 통근거리 등의 영향은 점유형태(자가 vs. 임차가구)에 따라 상이한 결과를 보이는 경우가 많았다.<sup>2)</sup> 본 연구에서는 가구 및 주택 관련 특성에 초점을 두고 거주기간 또는 이사위험이 어떻게 달라지는지에

1) 주거이동과 관련된 이론적 배경 및 선행연구들에 대한 자세한 사항은 조성진(2014), 문근식·이현석(2016) 등 참고. 특정 주거지(주택)에서의 거주기간이 아닌 이주여부/계획을 종속변수로 하여 로짓(Logit)이나 프로빗(Probit) 모형 등을 이용하여 분석한 연구들도 많이 있다(Sánchez & Andrews, 2011; 최열·김형준, 2012; 조성진, 2014; 문근식·이현석, 2016 등).

2) 전세가구의 경우에는, 가구소득이 높거나 주택규모가 작은 경우, 가구주의 직업이 (상용직 대비) 자영업인 경우, 특정 주거지에서의 거주기간이 길어지는 경향을 보였다. 한편, 자가가구의 경우에는, 통근거리가 길거나 주택규모가 작을수록, 거주기간이 짧아지는(주거이동 확률이 높아지는) 양상을 보였다.

대해 살펴보고자 한다.

본 연구는 다음과 같이 구성되어 있다. 이어지는 II장에서는 KLIPS에 근거한 주거패널데이터 구축과정을 설명하고 주요 변수에 대한 기초 통계 - 임의의 가구의  $t$  시점 이사확률, (이행확률을 이용한)  $t$  시점에 이사하지 않은 가구가 1년 후 이사할 확률, 자가 및 전월세 비율 추이, 주택점유 형태의 유지 및 전환 확률 등 - 를 보여준다. 그리고 이러한 주거패널데이터를 KLIPS의 다른 조사변수들과 병합하여 주거관련 연구용 데이터를 만드는 과정을 설명하고, 1인 가구의 점유형태와 주거비 비율을 예로 들어 가구특성과 주거 변수의 관계를 살펴본다. III장에서는 가구의 주거이력을 기준으로 주거력 데이터를 구축하는 과정과 여기에 포함된 8개 변수에 대해 설명하고, 주택순번과 해당순번 주택에서의 거주기간 변수에 대한 기초통계량, 이사시기별 평균 거주기간 그래프를 제시한다. II장에서와 마찬가지로 가구특성과 병합된 주거력 데이터를 만드는 과정을 단계별로 설명하고 이를 이용하여 점유형태/주택유형별 평균 거주기간을 비교분석한다. IV장에서는 주거력 데이터를 활용하여 각 주거이력에서의 거주기간에 대한 생존분석 모형을 추정하고, 그 결과에 근거하여 점유형태와 가구주 나이에 따른 헤저드 함수를 비교한다. 특히 한 가구에 대해 이사라는 사건 발생이 반복적으로 관찰되는 자료의 특성(반복사건)을 고려하여 frailty 생존분석 모형을 적용한다. 마지막으로 V장에서는 분석결과를 요약하고 시사점을 제시한다.

## II. 주거패널데이터

### 1. 주거패널데이터 구축

본 장에서는 KLIPS에 근거한 주거관련 데이터 중 첫 번째인 패널 long-type데이터 (이하 주거패널데이터) 구축과정과 해당 변수에 대해 설명한다. 주거패널데이터는 조사가구의 주거관련 변수를 조사 차수(wave)별로 모아서 패널 long-type 형태로 만들어진 것이다. 주거패널데이터를 구축함으로써 주거관련 변수들이 패널데이터 분석에 손쉽게 활용될 수 있다. 관심 있는 KLIPS 내의 다른 변수들과 주거패널데이터가 쉽게 연결(merge)될 수 있는 장점이 있다.

<표 1>에서는 주거패널데이터의 변수이름과 변수 정의에 대해서 정리한다. 저자들이 만든 데이터에는 11개의 변수가 포함되어 있다. KLIPS 원 데이터에서 그대로 가져온 변수도 있지만 특정 변수들은 재코딩 작업을 거쳐서 완성하였다.<sup>3)</sup> 가령 *hsize* 변수는 자가인 경우에는 전체평수 변수(*h\_1410*), 자가 이외의 경우에는 실사용평수 면적(*h\_1411*)을 대입하여 생성하였다. *hsize* 변수는 미터법 사용의 정부정책을 따라 평수를 모두 제곱미터( $m^2$ )로 환산한 값이다. 또한 *hbegin* 변수 역시 KLIPS 원 데이터에는 거주시작 시점을 년과 월로 나눈 변수를 제공하고 있지만 본 연구에서 만든 데이터에서는 year-month를 묶어 하나의 변수로 만들어서 제공한다.

---

3) 주거패널데이터 구축 시 Stata 명령어 *klips\_use*와 *klips\_add*를 이용하였다(민인식, 2016).

<표 1> 주거패널데이터의 변수

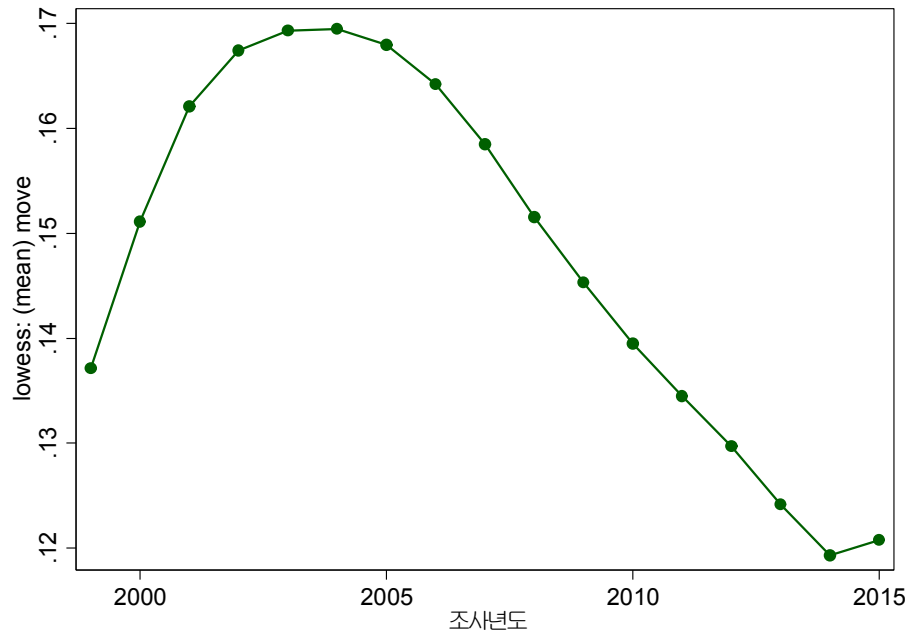
변수이름	변수정의	변수설명
<i>hhid</i>	가구id	가구 id
<i>wave</i>	조사 wave	조사 wave (1차~ 18차)
<i>year</i>	조사년도	조사년도 (1998년~2015년)
<i>move</i>	이사여부	해당 wave에 이사여부 (1=이사 0=비이사)
<i>residtype</i>	점유형태	주택의 점유형태 (1=자가 2=전세 3=월세 4=기타)
<i>housetype</i>	주택유형	주택유형 (1=단독주택 2=아파트 3=연립 4=다세대·다가구 5=상가 및 기타)
<i>hsize</i>	주거면적	주거면적: 자가인 경우 전체 면적, 자가 이외 경우 실사용면적
<i>hprice</i>	주택시가	주택의 시가(자가 이외의 경우는 결측치)
<i>deposit</i>	보증금	임차보증금(자가인 경우에는 결측치)
<i>monthlyrent</i>	월세	월세(점유형태가 월세 또는 기타인 경우에만 관찰됨)
<i>hbegin</i>	거주시작 년월	현재 주택 거주시작 년월

## 2. 기초 통계분석

주거와 관련된 변수 중 핵심적인 변수는 이사여부이다. 주거패널데이터에서 매년도 이사여부를 *move* 변수를 통해 관찰할 수 있다. *move* 변수는 최초 조사년도인 wave 1(1998년)에는 조사에 포함되지 않았다.<sup>4)</sup> 2차 wave인 1999년부터 이사여부 변수가 조사에 포함되었다. 1999년~2015년까지 매 조사년도별로 *move* 변수의 평균을 계산하면 임의의 가구의  $t$  시점 이사확률로 해석할 수 있다. [그림 1]에서  $\Pr(\text{move}_t = 1)$  을 그래프로 보여준다. 시간의 흐름에 따른 패턴을 이해하기 위해 lowess smoothing 값을 이용하여 그래프를 작성하였다. 1999년~2005년까지 꾸준히 이사확률이 증가하지만 2006년부터는 이사확률이 꾸준히 감소한다. 2006년이 주택거래 폭증과 부동산 시장 호황이었던 사실과 무관하지 않은 것으로 보인다. 2015년 현재 임의 가구의 이사확률은 12.1%이다. 즉 100가구 중 12가구가 2015년에 이사했다고 해석할 수 있다.

4) 따라서 *move* 변수의 1998년 값은 모두 결측치이다.

[그림 1]  $t$  시점의 이사확률



이행확률(transition probability)을 계산하여  $t$  시점에 이사하지 않은 가구가 1년 후 이사할 확률을 계산해보자.  $t$  시점과  $t+1$  시점의 이행확률은 (식 1)과 같이 정의한다. <표 2>의 이행확률 결과에 따르면  $t$  시점에 이사하지 않은 가구가 그 이듬해에 이사할 확률은 10.6%이고 그 이듬해에도 여전히 이사하지 않을 확률은 89.3%이다.  $t$  시점에 이사진 가구도 그 이듬해에도 이사할 확률이 17.9%이고 이사하지 않고 그대로 머물 확률은 82.0%이다. 이행확률 계산은 범주형 변수인  $move$ 의 serial correlation을 계산하는 것과 같다. 이사를 하지 않은 가구가 다음 기에도 이사를 하지 않을 확률이 89.3%로 매우 높다. 그리고 현 기에 이사를 했던 가구가 다음 기에도 이사를 할 확률은 18% 정도이다.

$$\Pr(move_{t+1}=1 | move_t=0) : \text{이행확률} \quad (\text{식 1})$$

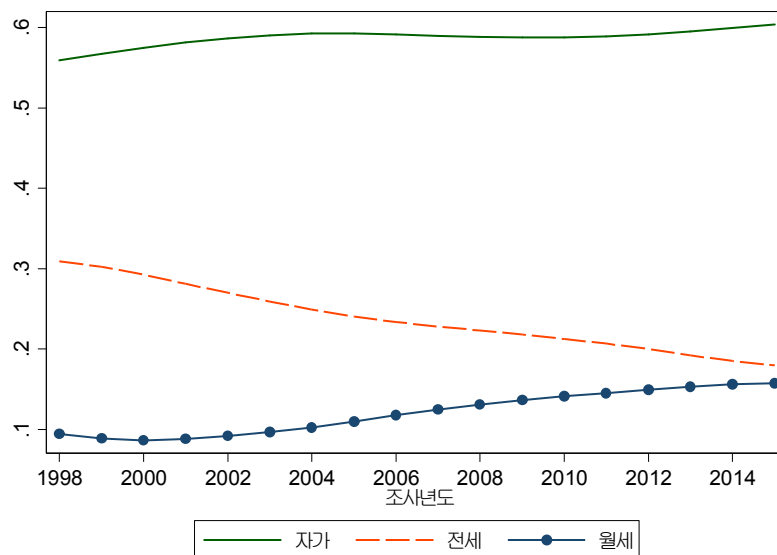
<표 2>  $move$  변수의 이행확률

(단위: %)

		$t+1$ 시점	
		$move_{t+1}=0$	$move_{t+1}=1$
$t$ 시점	$move_t=0$	89.3	10.6
	$move_t=1$	82.0	17.9

*residtype* 변수는 주택의 점유형태를 의미하는 변수이다. *residtype* 변수를 이용하여 년도별 자가점유 비율을 계산할 수 있다. 매년도 조사된 가구에서 *residtype* = 1(자가), *residtype* = 2(전세), *residtype* = 3(월세) 비율을 계산한 후 lowess smoothing한 결과를 [그림 2]에서 제시한다. 자가점유 비율은 1998년 55%에서 2015년 60%까지 꾸준히 상승하는 패턴이다. 최근 주택 임대차 시장에서 전세가 감소하고 월세가 증가하는 현상은 KLIPS에서도 관찰된다. 전세비율은 1998년 30%에서 2015년 17.8%까지 하락하고 있는 반면 월세비율은 1998년 9.4%에서 2015년 15.6%까지 꾸준히 상승하고 있다.

[그림 2] 자가 및 전월세 점유비율



*residtype* 변수를 이용하여 이행행렬을 계산해보자. 점유형태가 범주형 변수이므로 serial correlation을 통해 계산하고자 한다.  $t$  시점에서 자가주택에 거주하는 가구는  $t+1$  시점에서도 자가주택에 거주할 확률이 매우 높을 것으로 예상가능하다. 한편  $t$  시점의 전·월세 거주가구에 대해서는 1년 후인  $t+1$  시점에 자가소유로 전환할 확률이 어느 정도인지 확인할 수 있다. <표 3>의 결과에 따르면 주택점유의 계열상관은 매우 높다는 것을 알 수 있다.<sup>5)</sup> 자가 가구가 다음 기에도 자가 가구일 확률은 95.4%이고 전세가구의 계열상관은 79.9%, 월세가구의 계열상관은 80.5%이다. 전·월세 가구가 주택소유로 전환될 이행확률은 다음과 같이 계산된다. 아래 결과를 보면, 월세 가구에 비해 전세 가구가 다음 기에 주택소유로 전환될 확률이 2배 이상 더 높다는 것을 알 수 있다.

$$\Pr(residtype_{t+1} = 1 | residtype_t = 2) = 11.7\%$$

5) 점유형태가 “상가 및 기타”인 가구의 이행확률은 제시하지 않는다.

$$\Pr(residtype_{t+1}=1|residtype_t=3) = 4.9\%$$

<표 3> 주택점유형태의 이행행렬

(단위: %)

		<i>t</i> +1 시점		
		<i>residtype</i> <sub><i>t</i>+1</sub> = 1	<i>residtype</i> <sub><i>t</i>+1</sub> = 2	<i>residtype</i> <sub><i>t</i>+1</sub> = 3
<i>t</i> 시점	<i>residtype</i> <sub><i>t</i></sub> = 1	95.4	2.76	0.85
	<i>residtype</i> <sub><i>t</i></sub> = 2	11.7	79.9	6.5
	<i>residtype</i> <sub><i>t</i></sub> = 3	4.9	11.3	80.5

### 3. 주거패널데이터 활용

본 소절에서는 앞서 구축된 주거패널데이터를 KLIPS의 다른 조사변수와 병합하여 주거관련 연구용 데이터를 만드는 과정에 대해 설명한다. 주거관련 변수와 다른 가구특성 변수를 병합하여 하나의 데이터 세트를 만들기 위해서는 민인식(2016)에서 제안한 klips\_use와 klips\_add 명령어를 활용하는 것이 편리하다.

제 1 단계: klips\_use 명령어를 이용하여 가구특성 변수를 포함한 데이터 세트를 만든다. 파일이름은 klips\_final.dta로 저장된다. wd(c:\klips) 옵션에서는 KLIPS 1차~18차 데이터가 저장되어 있는 폴더이름을 지정한다.

```
klips_use 거주지역 가구구성별 가구주나이 자녀수0세6세 자녀수7세12세 ///
자녀수13세15세 가구원수 가구주혼인 가구주교육 총소득 부채총액 ///
고등학교이하자녀수 ///
, wave(01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18) wd(c:\klips)
```

제 2 단계: 제 1 단계에서 만든 klips\_final.dta 파일과 주거패널데이터를 병합한다. 주거패널데이터는 klips\_주거패널데이터.dta 라고 가정하자. 아래 Stata 코드와 같이 merge 명령어를 이용하여 두 데이터 파일을 병합하여 새로운 파일 klips\_주거패널데이터1.dta로 저장하면 된다.

```
cd c:\klips
use klips_final, clear
merge 1:1 hhid wave using klips_주거패널데이터 , nogen
save klips_주거패널데이터1 , replace
```

<표 4>에서는 위 과정을 통해 완성된 klips\_주거패널데이터1.dta에 포함된 변수 및 관측치(obs.)에 대해서 간단하게 정리한다. 완성된 예제 파일에는 9,574가구의 99,017 관측치(obs.)가 포함되어 있으며, <표 1>에서 제시한 주거관련 변수 8개와 저자들이 지정한 가구특성 변수 12개가 포함되어 있다.

<표 4> klips\_주거패널데이터1.dta 파일

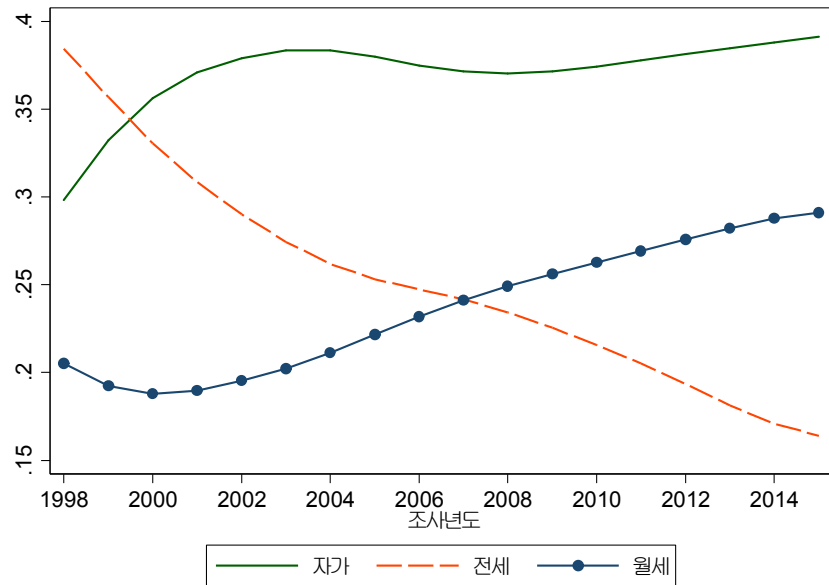
변수	id 및 wave 변수: 3개 가구특성 변수: 12개 주거관련 변수: 8개
가구 수	9,574 가구
observations	99,017 개

<표 4>에서 구축된 연구용 데이터를 이용하여 간단하게 가구특성과 주거 변수가 어떤 관련이 있는지 살펴볼 수 있다. 통계청 2015년 인구총조사 결과에 따르면 2015년 1인 가구비율이 27%에 이른다. 그러나 KLIPS 데이터에서 2015년 1인 가구비율은 이보다 낮은 23%로 나타난다.<sup>6)</sup> 이러한 차이를 염두에 두면서, 1인 가구의 비중이 늘어나는 사회적 현상에 비추어 이들 가구의 주거관련 특성을 살펴보고자 한다. 먼저 1인 가구의 자가, 전월세 점유비율을 그래프로 보여주고자 한다. [그림 3]에서는 세 가지 타입의 점유형태 비율을 lowess smoothing한 결과를 제시한다. [그림 3]을 보면, 1인 가구의 자가 비율은 1998년 29%에서 2015년 38%로 전반적인 상승세를 보인다. 그러나 2인 이상 가구의 자가 비율이 2015년 68%인 것과 비교하면 30%p이상 낮은 수준에 불과하다. 주목할 점은 1인 가구의 경우에도 이미 2007년에 월세 비율이 전세 비율을 초과하기 시작하였다는 점이다. 2015년 기준 1인 가구의 전세 비율은 16%, 월세 비율은 28%로, 월세 거주비율이 12%p 더 높다.

6) KLIPS에서 1인 가구를 정확히 식별(identify)하기 위해서는 가구원수 변수 뿐 아니라 가구원의 동거여부까지 고려해서 1인 가구를 정의해야 할 것으로 판단된다.



[그림 3] 1인 가구의 점유형태



1인 가구의 소비 중에서 가장 큰 비중을 차지하는 것은 주거비일 것이다. 이는 [그림 3]에서 1인 가구의 임대차 가구비율이 2015년 기준 62%<sup>7)</sup>가 넘는 것을 보더라도 어느 정도 예상할 수 있다. 이를 좀 더 구체적으로 살펴보기 위해, [그림 4]에서는 1인 가구의 주거비 부담비율을 계산하여 시계열 그래프로 보여준다. 자가 가구를 제외한 전월세 가구에 대해서만 주거비를 계산한다. 전세 가구와 월세 가구에 대해 월 주거비는 다음과 같이 계산한다. 이자율은 연도별 예금은행 가계대출 금리를 선택하여 계산에 사용한다.<sup>8)</sup> 주거비 비율은 1인 가구의 월 주거비와 월 소득의 비율로 계산한다.<sup>9)</sup> [그림 4]의 월 주거비 비율을 살펴보면 2002년 23%에서 최고점을 찍고 그 이후 꾸준히 감소한다. 2015년에는 월 소득의 17%를 주거비로 사용한다. 주거비 비율이 감소하는 패턴은 최근 이자율 하락을 반영하는 것으로 보인다. 최근 시점으로 올수록 1인 가구의 주거비 자체는 높아졌을 것이다. 그러나 소득대비 주거비 비율은 낮아지는 것으로 보인다.

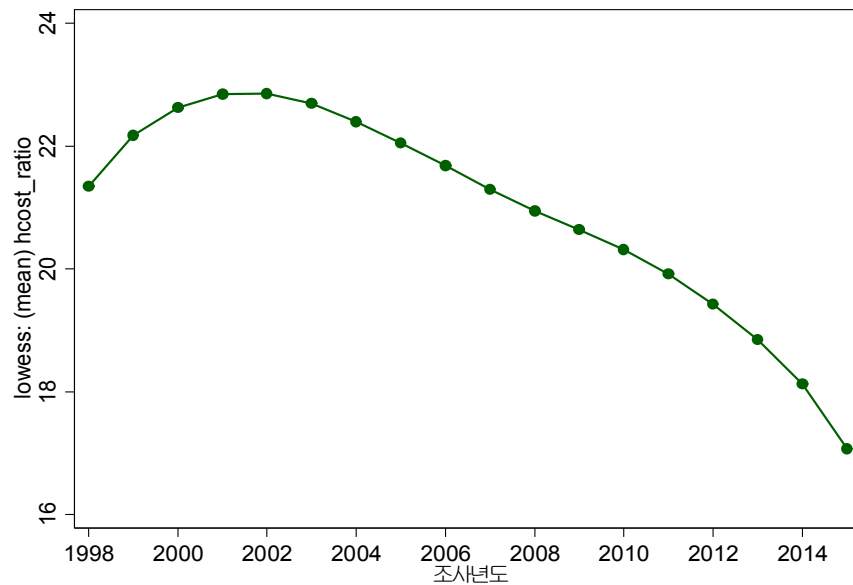
점유형태	월 주거비 계산
전세가구	(전세보증금 × 이자율) ÷ 12
월세가구	{(월세보증금 × 이자율) ÷ 12} + 월세

7) 임대차 가구비율 = 100% - 자가 비율 38% = 62%. 한편 전월세 비율은 16% + 28% = 44%.

8) 한국은행 경제통계시스템에서 연도별 이자율 변수를 가져와서 사용한다.

9) 월 주거비 비율이 100%가 넘는 경우는 이상치(outlier)로 간주하고 제외하였다.

[그림 4] 1인 가구의 주거비 비율



### III. 주거력 데이터

#### 1. 주거력 데이터 구축

본 장에서는 가구의 주거이력을 기준으로 만든 데이터를 설명한다. II장의 주거패널데이터는 각 조사시점에서 어떤 주거상태인지 파악할 수는 있지만 가구의 주거변화에 대해서는 파악하기 어려운 데이터 구조이다. KLIPS에서 직업력 데이터를 새로 구축하여 제공하듯이 저자들은 주거력 데이터를 완성하여 연구자들에게 제공하고자 한다. 주거력 데이터에서 포함하고자 하는 변수와 데이터 구조는 <표 5>와 [그림 5]와 같다.

<표 5>에서는 주거력 데이터에 포함된 8개 변수에 대해 설명하고 있다. 가장 핵심적인 변수는 *housebegin*와 *houseend* 변수이다. 어떤 가구의 특정 주택에서 몇 년간 거주했는지를 계산하기 위해서는 필수적인 변수들이다. 두 변수를 이용하여 거주기간(*duration*) 변수를 생성할 수 있다. [그림 5]에서 변수의 구체적인 내용을 확인할 수 있다. hhid=1번 가구는 KLIPS 1차~18차 조사를 통해 총 4번의 주거이력이 있다는 것을 알 수 있다. house\_num=1에서는 1994년~2000년(7년) 동안 거주하였고 house\_num=2에서는 2002년~2002년(1년) 동안 거주하였다. 2001년에는 어떤 주거이력이 있는지 확인할 수 없었다. house\_num=3에서는 2003년 1년의 거주이력을 확인할 수 있다. 마지막으로 house\_num=4의 경우 2004년~현재 2015년(마지막 조사년도)까지 12년째 거주하고 있다는 것을 알 수 있다. house\_num=4에서 마지막 거주년도를 알 수 없기 때문에 censored=1을 갖게 된다.

house\_bwave와 house\_ewave 변수를 통해 해당 거주이력을 KLIPS 원 데이터와 병합할 수 있다.

<표 5> 주거력 데이터 변수

변수이름	변수정의	변수설명
<i>hhid</i>	가구 id	가구 id
<i>houseenum</i>	주택순서	거주 주택의 순번
<i>housebegin</i>	거주 시작년도	해당 순번 주택의 거주 시작년도
<i>houseend</i>	마지막 거주년도	해당 순번 주택의 마지막 거주년도
<i>housebwave</i>	거주시작 wave	거주 시작년도에 매칭되는 wave
<i>houseewave</i>	거주이전 wave	마지막 거주년도에 매칭되는 wave
<i>censored</i>	우측절단 여부	마지막 거주년도가 관찰되지 않으면 1 마지막 거주년도가 관찰되면 0
<i>duration</i>	거주기간	해당순번 주택에서 거주기간 (마지막 거주년도 - 거주시작년도 +1)로 계산

[그림 5] 주거력 데이터 구조 예시

	hhid	house_num	house_begin	house_end	house_bwave	house_ewave	censored	duration	
1	1	1	1994	2000	1	3	0	7	
2	1	2	2002	2002	5	5	0	1	
3	1	3	2003	2003	6	6	0	1	
4	1	4	2004	2015	7	18	1	12	
5	2	1	1993	1998	1	1	0	6	
6	2	2	1999	2000	2	3	0	2	
7	2	3	2001	2015	4	18	1	15	

주거력 데이터 구축에서 어려운 부분은 “거주시작 년도(*housebegin*)” 변수를 만드는 것이다. KLIPS에서 h\_1415변수가 “해당 주택의 거주시작 년도”를 조사하고 있다. 그러나 이 변수에는 오류가 상당 부분 존재한다. 가령 같은 주택에 거주함에도 불구하고 wave 1에서 답한 h\_1415 변수 값과 wave 2에서 답한 h\_1415 변수 값이 서로 다른 경우가 있다. 또한 wave 1과 wave 2에 서로 다른 주택에 거주함에도 불구하고 wave 1의 거주시작년도가 wave 2의 거주시작년도보다 나중인 경우도 존재한다. 또한 8차 년도 이후 조사에서는 “이사여부”=0(이사하지 않음)이라고 답한 경우에는 h\_1415 변수값을 입력하지 않고 결측치로 되어 있는 것에 주의해야 한다. 이외에도 주거력 데이터를 만들기에 수많은 걸림돌이 있다는 것을 발견하였다. 따라서 KLIPS 원 데이터에 기반하여 저자들의 개인적인 판단(<표 6> 참고)을 적용하여 새로운 주거력 데이터를 만드는 방향을 선택한다. <표 6>의 과정 이외에도 주거력 데이터를 완성하기 위해 저자들의 자의적 판단이 사소하게 적

용된다.<sup>10)</sup> 최종적으로 완성된 데이터 파일은 **klips\_주거력데이터.dta**로 저장하였다. 해당 데이터 파일은 9,344가구의 총 30,516개 obs.로 구성되며 변수의 개수는 8개이다.

<표 6> *housebegin* 변수 구축 과정

단계	과정 설명
제 1 단계	<i>hchange</i> 변수를 새로 만든다. 이 변수는 <i>t</i> 시점의 주거상태를 ( <i>t-s</i> )시점의 주거상태와 비교하여 주택이 바뀌었으면 <i>hchange</i> = 1이고 바뀌지 않았다면 <i>hchange</i> = 0이 된다. <i>hchange</i> 는 KLIPS에서 조사하는 “이사여부” 변수와 정확히 일치하지는 않는다.
제 2 단계	<i>hchange</i> = 1이 되는 경우 1) 거주지역이 바뀐 경우 2) 주택 점유유형이 바뀐 경우 3) 주택 종류가 바뀐 경우 4) “이사여부” 변수에서 “이사”로 답한 경우 위 4가지 경우 중 한 가지라도 해당하면 <i>hchange</i> = 1이 된다.
제 3 단계	<i>hchange</i> = 1인데 <i>h1415</i> 가 결측치로 되어 있으면 <i>h1415</i> 값은 해당 조사년도에 이사한 것으로 간주한다.
제 4 단계	<i>hchange</i> = 1인데 <i>t</i> 시점의 <i>h1415</i> 값이 ( <i>t-s</i> )시점의 <i>h1415</i> 보다 작으면 해당 조사년도에 이사한 것으로 간주한다.
제 5 단계	<i>hchange</i> = 0인데 <i>t</i> 시점의 <i>h1415</i> 값과 ( <i>t-s</i> )시점의 <i>h1415</i> 값이 서로 다르면 <i>t</i> 시점의 <i>h1415</i> 값을 ( <i>t-s</i> )시점의 <i>h1415</i> 값으로 대신한다.
제 6 단계	wave 8 이후의 조사에서 <i>hchange</i> = 0인 경우, <i>h1415</i> 이 결측이면 ( <i>t-s</i> )시점의 <i>h1415</i> 값으로 대신한다.

## 2. 기초 통계분석

본 소절에서는 주거력 데이터를 이용하여 가구들의 주거이력에 관한 기초통계량을 정리하고자 한다. 먼저 *housenum* 와 *duration* 변수의 기초통계량은 <표 7>에 제시한다. 평균 3.26번 주거이력이 관찰되며 최소 주거이력은 1번 그리고 주거이력이 가장 많이 관찰되는 가구는 13번이다. 주의할 점은 어떤 가구는 1998년 첫 번째 조사 이전에 이미 주거이력이 존재할 수 있다. 그러나 1998년 이전 주거이력은 알 수 없기 때문에 평균 3.25는 하향편향(downward bias)될 가능성이 있다. *duration* 변수의 평균은 5.12년이다.<sup>11)</sup> 즉 KLIPS에서 조사된 가구는 특정 주택에 이사하면 평균 5.12년 거주하는 것을 알 수 있다. 그러나 *duration* 변수는 우측절단(right-censored)인 경우가 있기 때문에 평균 값 역시 downward bias될 가능성이 크다.

10) 자세한 구축과정에 대한 설명은 교신저자(민인식)에게 문의할 수 있다.

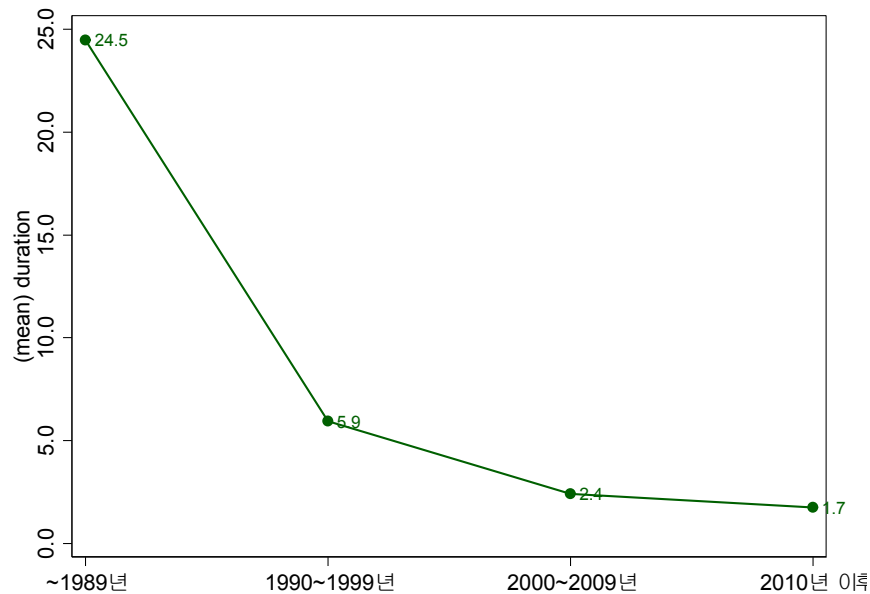
11) *duration* 변수값이 60년 초과인 경우는 이상치로 제외하고 평균을 계산하였다.

<표 7> *housenum* 과 *duration* 변수

변수	표본 수	평균	표준편차	중앙값	최소값	최대값
<i>housenum</i>	9,344 가구	3.26	2.27	3	1	13
<i>duration</i>	30,285 obs	5.12	7.46	3	1	60

[그림 6]에서는 이사시기별 평균 거주기간을 그래프로 보여준다.<sup>12)</sup> 각 가구의 주거이력은 아직 완료되지 않는 경우는 우측절단으로 나타난다. 따라서 [그림 6]의 평균을 계산할 때는 우측절단(right-censored) 주거이력은 제외하고 계산한다. 주거시작 년도를 4개 구간, 즉 주거시작연도가 1) 1989년 이전, 2) 1990년~1999년, 3) 2000년~2009년, 4) 2010년 이후에 이사한 경우로 나누었다. 과거 시점에 시작한 주거이력의 경우 평균 거주기간이 24.5년으로 매우 길다. 그러나 1990년대에 시작한 주거이력은 평균 거주기간이 5.9년으로 대폭 짧아진다. 2000년대에 시작한 주거이력은 2.4년으로 더 짧아지는 것을 알 수 있다. [그림 6]을 통해 최근 주택 매매가격 및 전세가격 상승 등으로 인해 평균 거주기간이 점차 짧아지고 있다는 것을 예상할 수 있다.

[그림 6] 이사 시기별 평균 거주기간



12) <표 7>과 마찬가지로 *duration* > 60인 표본은 제외하고 계산하였다.

### 3. 주거력 데이터의 활용

주거력 데이터는 8개 변수 외에 점유형태, 주택가격 등 주거관련 변수 및 가구특성 변수들을 포함하고 있지 않다. 따라서 가구특성과 주거이력을 연결하여 분석하기 위해서는 데이터를 병합할 필요가 있다. 민인식(2016)의 klips\_use와 klips\_add 명령어를 이용하면 간단하게 병합된 데이터 세트를 새로 만들 수 있다.

제 1 단계: klips\_use 명령어를 사용하여 가구특성 변수를 klips\_final.dta 데이터로 만들고 II장에서 제시한 주거패널데이터를 병합한 후 klips\_주거패널데이터1.dta 파일로 저장한다.

```
klips_use 거주지역 가구구성별 가구주나이 자녀수0세6세 자녀수7세12세 ///
자녀수13세15세 가구원수 가구주혼인 가구주교육 총소득 부채총액 ///
고등학교이하자녀수 ///
, wave(01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18) wd(c:\klips)

merge 1:1 hhid wave using klips_주거패널데이터 , nogen
save klips_주거패널데이터1 , replace
```

제 2 단계: klips\_주거력데이터.dta을 long-type으로 바꾼 후 1단계의 klips\_주거패널데이터1.dta와 병합한다.

```
use klips_주거력데이터, clear
ren (house_begin house_end) (house_year1 house_year2)
ren (house_bwave house_ewave) (house_wave1 house_wave2)

reshape long house_year house_wave , i(hhid house_num) j(nn)
ren house_wave wave
order hhid house_num wave
merge m:1 hhid wave using klips_주거패널데이터1
keep if _m==3
drop _m
order hhid house_num wave
drop move hbegin
save klips_주거력데이터1, replace
```

완성된 데이터의 일부분을 [그림 7]에서 보여준다. *nn* 변수가 각 주거이력의 *start*와 *end*를 나타낸다. 가령 *nn*=1 이면 시작시점, *nn*=2 이면 마지막시점을 의미한다. 즉 *hhid*=1번 가구의 첫 번째 주거이력은 1994년에 시작해서 2000년까지 거주하였다. 그리고 거주기간은 7년이고 점유형태는 전세임을 알 수 있다. 두 번째 주거이력은 2002년 1년 동안 전세로 거주하였다. 세 번째 주거이력은 2003년~2003년 1년 간 전세로 거주하였다. 네 번째 주거이력은 2004년~2015년까지 12년 동안 자가로 계속 거주하고 있다.

[그림 7] klips\_주거력데이터1.dta 파일

	hhid	house_num	wave	nn	house_year	censored	duration	resid_type
1	1	1	1	1	1994	0	7	2
2	1	1	3	2	2000	0	7	2
3	1	2	5	1	2002	0	1	2
4	1	2	5	2	2002	0	1	2
5	1	3	6	2	2003	0	1	2
6	1	3	6	1	2003	0	1	2
7	1	4	7	1	2004	1	12	1
8	1	4	18	2	2015	1	12	1
9	2	1	1	1	1993	0	6	2
10	2	1	1	2	1998	0	6	2
11	2	2	2	1	1999	0	2	2
12	2	2	3	2	2000	0	2	1
13	2	3	4	1	2001	1	15	2
14	2	3	18	2	2015	1	15	2

가구특성과 병합된 주거력 데이터를 이용하면 점유형태별/주택유형별 평균 거주기간을 쉽게 계산할 수 있다. *duration* ≤ 60인 경우에만 평균 계산에 사용하였으며 우측절단(right-censored) 경우도 포함하였기 때문에 평균은 downward bias될 가능성이 있다.

<표 8> 점유형태/주택유형별 평균 거주기간

점유형태	자가	전세	월세	기타	
	8.28년 (11,950)	2.94년 (9,620)	2.94년 (5,161)	3.16년 (2,354)	
주택유형	단독주택	아파트	연립	다세대/다가구	상가 및 기타
	8.40년 (7,567)	4.61년 (11,220)	3.75년 (3,177)	3.06년 (5,030)	3.42년 (2,067)

주: ( )안의 숫자는 평균 계산에 사용된 주거이력 관측치 수(observations)이다.

## IV. 거주기간에 대한 생존분석

### 1. 계량방법론

각 주거이력에서의 거주기간에 대한 계량모형 분석은 생존분석(survival analysis) 접근법을 선택한다. 종속변수인 거주기간(duration) 변수는 항상 0보다 크고 정규분포(normal distribution)와는 다른 형태의 분포를 따를 가능성이 크다. 또한 각 가구의 마지막 주거이력은 마지막 거주시점을 관찰할 수 없기 때문에 우측절단(right-censored) 현상이 나타난다. 본 연구의 데이터에서 드러난 또 다른 특징은 하나의 가구에 대해 사건발생(이사)이 반복적으로 관찰되는 것이다. 따라서 반복사건(repeated-failure event) 생존분석 접근법이 적용되어야 할 것으로 판단된다.

반복사건(repeated-failure event)은 특정 가구에서 사건발생이 2번 이상 관찰될 수 있는 경우이며 사건발생까지 걸린 시간(time to event)은 가구 id 내에서 서로 상관관계를 가지게 된다. 반복사건을 고려하는 생존분석 모형으로는 frailty 모형과 variance-corrected 접근법을 선택할 수 있다. 먼저 frailty 모형에서는 모형 설정에서 명확하게 failure times 간 상관관계를 가정한다. 특히 관찰되지 않는 shared-frailty( $\alpha_i$ )를 확률효과(random-effects term)로 포함한다.  $\alpha_i$  확률변수에 대해서는 항상 양의 값을 갖는 특정한 분포함수를 가정한다. 둘째, variance-corrected 접근법은 모형설정에서는 failure times 간 의존성을 가정하지는 않고 단지 추정량의 공분산 행렬(covariance matrix)에서 같은 가구 id 내의 상관관계를 고려하여 재조정된 공분산 행렬을 계산한다. 본 연구에서는 frailty 모형을 이용하여 거주기간 생존분석 결과를 보여주고자 한다.

frailty 모형은 준모수적 접근법인 cox-regression과 모수적 생존분석 모형으로 추정할 수 있다. 먼저 cox-regression에서 가구  $i$ 의  $j$  번째 주거이력에 대한 해저드(hazard)는 다음과 같이 정의한다.

$$h_{ij}(t) = h_0(t)\alpha_i \exp(x_{ij}\beta) \quad (\text{식 2})$$

위 식에서  $\alpha_i$ 는 가구레벨 frailty로 정의한다.  $h_0(t)$ 는 기준선 해저드함수(baseline hazard function)이고 구체적인 형태를 가정하지 않는다. frailty는 관찰되지 않으며 양의 값을 갖는 확률변수로 본다.  $v_i = \log \alpha_i$ 로 정의하면 (식 2)는 (식 3)과 같이 표현할 수 있다.

$$h_{ij}(t) = h_0(t)\exp(x_{ij}\beta + v_i) \quad (\text{식 3})$$

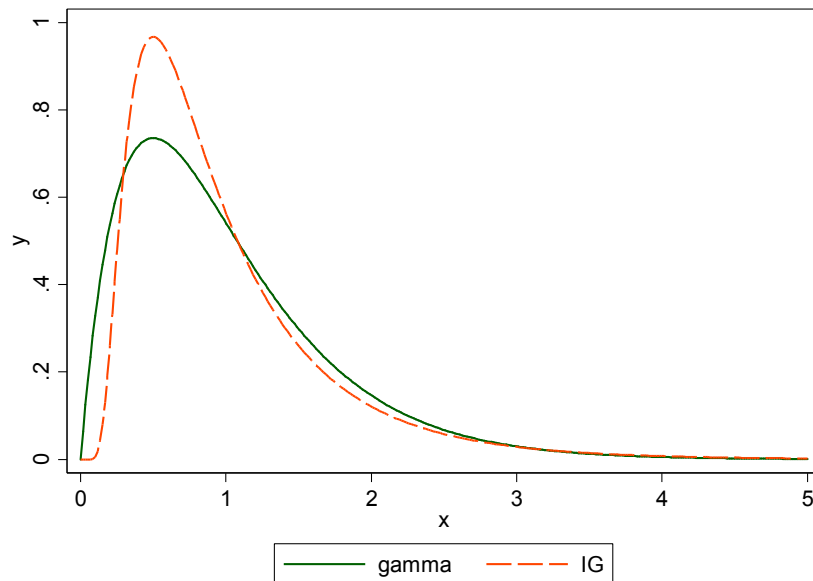
(식 3)에서  $v_i$ 는 패널 선형회귀모형의 확률효과(random effects)와 유사하게 표현됨을 알 수 있다. 모수적 생존분석 모형에서도 가구에 특정한(household-specific) frailty를 고려할 수 있다. 가구  $i$ 의 거주이력  $j$ 에서의 해저드(hazard) 함수는 다음과 같이 가정한다.  $\alpha_i$ 를 통해 특정 가구 내에서 거주기간은 서로 양의 상관관계를 가지게 된다.



$$h_{ij}(t | \alpha_i, x_{ij}) = \alpha_i h_{ij}(t | x_{ij}) \quad (\text{식 4})$$

위 식에서  $h_{ij}(t|x_{ij})$  는 모수적 해저드 함수(hazard function)를 가정한다. 모수적 분포함수로는 와이불(Weibull) 분포 또는 로그로지스틱(log-logistic) 분포를 가정할 수 있다.  $\alpha_i$ 에 대한 분포함수는 평균이 1이고 분산이  $\theta$ 인 감마분포(Gamma distribution) 또는 역가우시언 분포(Inverse-Gaussian distribution: IG distribution)를 선택할 수 있다. [그림 8]에서는  $mean=1$  이고  $variance=0.5$ 인 감마 분포와 역가우시언 분포의 확률밀도함수(probability density function: pdf)를 비교한다. 두 분포 모두 0보다 큰 값을 갖는 분포이고 오른쪽으로 꼬리가 긴 분포임을 알 수 있다. 관찰되지 않는 가구 이질성(household heterogeneity)은 대부분 0~2 사이에 존재하고 거주기간에 대해 극단적인 이질성을 갖는 소수의 가구가 존재한다고 해석할 수 있다.

[그림 8] 감마와 역가우시언 분포의 확률밀도함수(pdf)



## 2. 실증분석 결과 및 해석

본 연구의 실증분석에서는 frailty를 포함한 모수적 생존분석 모형을 추정한다. 종속변수인 거주기간에 대한 설명변수는 가구의 인구통계학적 특성과 가구의 경제적 요인 변수를 포함한다. <표 9>에서는 종속변수와 설명변수의 이름과 변수 정의를 정리하여 보여준다. 종속변수인 *duration* 변수의 극단치를 제외하기 위해  $duration \leq 60$ 인 경우만 추정에 사용한다. 분석대상 표본은 9,172가구의 29,096 obs.이다.

&lt;표 9&gt; 변수 설명

변수 속성	변수 이름	변수 설명
종속변수	<i>duration</i>	거주기간 우측절단(right-censored)인 경우에는 2015년까지 거주했다고 가정함
설명변수	<i>age</i>	가구주 나이
	<i>sex</i>	가구주 성별 (1=남자, 2=여자)
	<i>hnum</i>	고등학교 이하 자녀 수
	<i>edu</i>	가구주 교육수준 (1=고졸미만, 2=고졸 또는 대학중퇴, 3=전문대졸 이상)
	<i>married</i>	가구주 혼인상태 (1=미혼/기혼무배우 2=기혼유배우)
	<i>region</i>	거주지역 (1=광역시 2=비광역시)
	<i>lincome</i>	2015년 실질소득으로 환산한 가구 실질소득의 로그값
	<i>residtype</i>	주택점유형태 (1=자가 2=전세 3=월세 및 기타)

frailty를 고려한 생존분석 모형 추정결과는 <표 10>에 제시되어 있다. 모수적 분포는 와이블(Weibull)과 로그로지스틱(log-logistic) 분포, 각 생존분석 모형에서 frailty의 분포는 감마와 역가우시언 분포로 가정하여 총 4가지 모형 추정결과를 제시한다. 오차항( $\alpha_i$ ) 분포가 와이블인 경우에는 hazard metric과 AFT(Accelerated Failure Time) metric으로 추정결과를 제시할 수 있지만 로그로지스틱인 경우에는 AFT metric만 가능하기 때문에, 결과 비교를 위해 와이블 모형에서도 위험률(hazard ratio) 대신 추정계수(estimated coefficients)를 제시한다.

모수적 분포함수로써 와이블(Weibull) 분포를 가정한 생존분석 결과에서 먼저  $\hat{p} = 1.29$  (frailty: Gamma)이기 때문에 increasing hazard가 추정됨을 알 수 있다. 즉 다른 조건이 일정할 때 거주기간이 길어질수록 이사사건 발생 리스크가 높아진다. 또한  $\alpha_i$ 의 분산 추정치는  $\hat{\theta} = 0.49$  (frailty: Gamma)로 0과 유의하게 다르다는 것을 알 수 있다. 각 가구의 거주기간에 관찰되지 않는 이질성이 통계적으로 존재한다. 로그로지스틱(log-logistic) 분포를 가정한 생존분석 결과에서는  $\hat{\gamma} = 0.564$  (frailty: Gamma), 즉 1보다 작기 때문에 비단조적(non-monotone) hazard임을 알 수 있다.<sup>13)</sup> 거주기간이 어느 이상 길어지면 오히려 이사발생 위험이 줄어들게 된다. 분산 모수는  $\hat{\theta} = 0.324$  (frailty: Gamma)로 추정되며 통계적으로 0과 유의하게 다르다. 따라서 거주기간에 영향을 미치는 각 가구의 이질성이 유의하게 존재한다.

13) 로그로지스틱 분포에서  $\gamma$  값과 해저드 함수와의 관계는 이경희·민인식(2016) 참고.

&lt;표 10&gt; 생존분석 모형 추정결과

	모수적 분포함수 : Weibull distribution		모수적 분포함수 : Log-logistic distribution	
	frailty 분포: Gamma	frailty 분포: IG	frailty 분포: Gamma	frailty 분포: IG
<i>age</i>	0.032 (0.0007)***	0.034 (0.0007)***	0.024 (0.0006)***	0.025 (0.0007)***
<i>lincome</i>	0.056 (0.009)***	0.053 (0.009)***	0.067 (0.009)***	0.066 (0.009)***
<i>hnum</i>	0.027 (0.009)***	0.021 (0.009)**	0.028 (0.009)***	0.025 (0.009)***
<i>2.sex</i> (여자)	-0.003 (0.025)	-0.010 (0.025)	-0.037 (0.022)	-0.010 (0.025)
<i>2.married</i> (기혼 유배우)	-0.074 (0.023)***	-0.083 (0.023)***	-0.037 (0.022)*	-0.039 (0.022)*
<i>2.region</i> (비광역시)	0.111 (0.016)***	0.124 (0.017)***	0.090 (0.015)***	0.098 (0.015)***
<i>2.residtype</i> (전세)	-0.850 (0.017)***	-0.869 (0.017)***	-0.681 (0.017)***	-0.692 (0.017)***
<i>3.residtype</i> (월세)	-0.833 (0.020)***	-0.850 (0.020)***	-0.692 (0.020)***	-0.704 (0.020)***
<i>2.edu</i> (고졸)	0.152 (0.024)***	0.144 (0.024)***	0.237 (0.023)***	0.242 (0.024)***
<i>3.edu</i> (전문대졸 이상)	0.219 (0.026)***	0.206 (0.027)***	0.332 (0.026)***	0.341 (0.027)***
상수항	0.085 (0.086)	0.063 (0.087)	-0.198 (0.086)	-0.231 (0.087)
$\log(p)$	0.261 (0.006)***	0.256 (0.006)***		
$\log(\theta)$	-0.699 (0.039)***	-0.521 (0.049)***		
$\log(\gamma)$			-0.572 (0.006)***	-0.573 (0.007)***
$\log(\theta)$			-1.126 (0.050)***	-1.066 (0.056)***
$\log L$	-32585	-32614	-32125	-32132
LR 검정 $H_0: \theta = 0$	1,683***	1,625***	807***	791***
<i>obs</i>	26,657	26,657	26,657	26,657

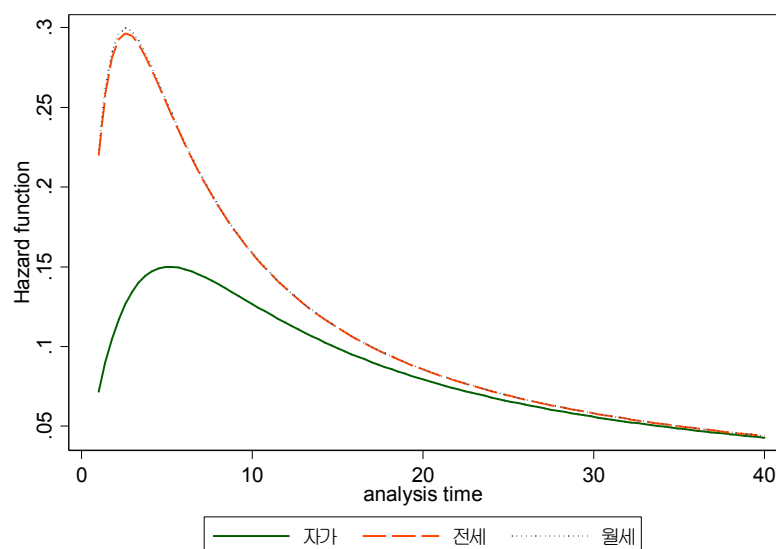
주: \*\*\*, \*\*, \*은 각각 1%, 5% 그리고 10% 유의수준에서 유의함을 의미한다.

설명변수와 거주기간과의 관계를 살펴보면 대부분 예상과 일치한다. 가구주의 나이(*age*)가 많을수록, 소득수준이 높을수록 특정 주거이력에서 거주기간이 길어진다. 또한 고등학교 이하 자녀의 수가 많을수록 거주기간이 길어지는 것으로 나타났다. 초·중·고등학교 학생이 있다면 전학 리스크가 발생하기 때문에 빈번한 이사를 자제할 가능성이 크기 때문인 것으로 보인다. 가구주의 성별은 통계적으로 유의하지 않은 반면, 가구주 혼인여부 변수는 음(-)으로 유의하다. 즉 기혼유배우 가구의 경우 거주기간이 더 짧다.<sup>14)</sup> 한편, 광역시 거주 가구의 경우 비광역시 거주 가구에 비해 거주기

간이 유의하게 짧다. 비광역시에 비해 광역시의 전월세 가격이 높기 때문에 예상되는 추정결과이다. 점유형태를 살펴보면 자가 거주에 비해 전세/월세 거주가구의 거주기간이 유의하게 짧게 추정된다. 가구주의 교육수준이 높아질수록 거주기간은 유의하게 길어진다.

[그림 9]와 [그림 10]에서는 설명변수의 특정한 값에서 그린 해저드 함수(hazard function)를 비교한다.<sup>15)</sup> 그래프 작성을 위한 추정결과는 로그로지스틱 분포- frailty Gamma 분포를 사용한 모형을 선택하였다. [그림 9]에서는 점유형태(자가/전세/월세 가구)에 따른 해저드 함수 그래프를 보여준다. 추정결과에서  $\hat{\gamma} < 1$  이기 때문에 비단조적 해저드 함수로 표현된다. 전세와 월세가구의 해저드 함수는 거의 같으며, 자가 가구의 이사발생 리스크는 전월세 가구에 비해 현저하게 작다는 것을 알 수 있다. 전월세 가구의 경우에는 duration=4년 정도에 이사발생 위험률(hazard rate)이 가장 높아지고, 자가 가구의 경우에는 duration=5년 정도에 위험률이 가장 높게 추정된다.

[그림 9] 점유형태에 따른 hazard function

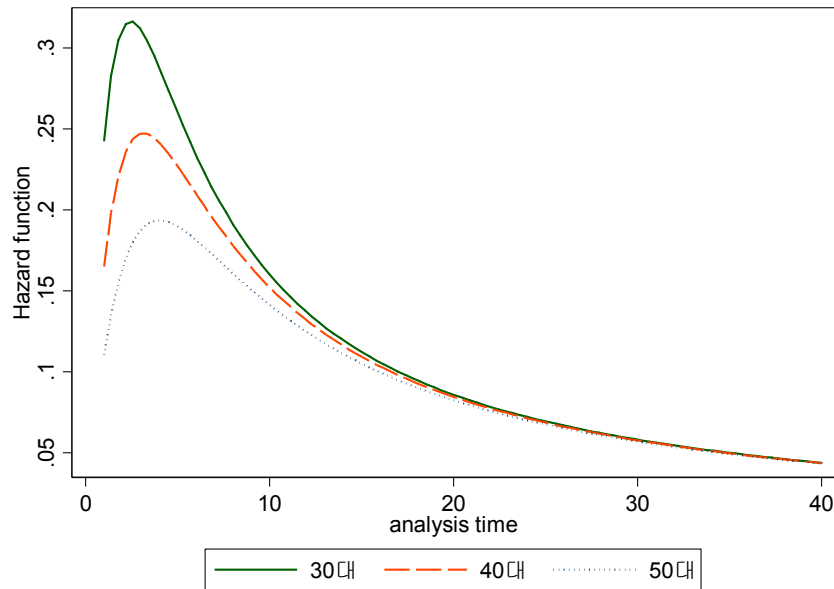


[그림 10]에서는 가구주 연령에 따른 이사사건 발생 위험(hazard)을 그래프로 보여준다. 가구주 연령은 35세, 45세, 55세의 3가지 경우로 고정한다.  $age = 35$ 일 때 사건발생 리스크가 가장 높고 연령이 높아질수록 해저드가 낮아진다. 그러나 거주기간이 10년이 넘으면 이사발생 위험이 가구주 연령과는 무관한 것으로 보인다. 특정 주택에서 10년 이상 오래 거주하면 가구주 나이와 무관하게 이사발생 리스크가 낮아진다.

14) 로그로지스틱 분포 모형에서는 5% 유의수준에서는 유의하지 않게 추정된다.

15) 관심변수 이외의 변수는 평균에서 고정시킨 상태에서 그래프를 작성한다.

[그림 10] 가구주 나이에 따른 hazard function



## V. 결론 및 시사점

주지하듯이 패널조사 데이터는 기본적으로 동일 대상을 한 번 이상의 여러 차수에 걸쳐 추적 조사한 내용을 담고 있기 때문에 조사대상의 이동, 상태변화 등과 같은 다이내믹스를 파악하는데 유용한 정보를 제공한다. 한국노동패널조사(KLIPS)는 특히 데이터화하기 번거롭고 어려운 직업력을 따로 구축하여 제공하고 있어 다양한 실증분석 연구에 활용되고 있다. 본 연구에서는, 이러한 직업력과 유사한 형태로, KLIPS에 근거한 조사가구의 주거력 데이터를 구축하여 연구자들에게 제공함으로써 조사의 활용도를 더욱 높이는데 기여하고자 한다. 또한 구축된 주거력 데이터에 근거하여 각 주거지에서의 거주기간에 대한 생존분석을 수행함으로써 그 활용가능성을 살펴보고자 한다.

우선, 본 연구는 주거력 데이터를 만들기에 앞서 주거패널데이터를 만드는 것으로부터 시작한다. 여기에서 주거패널데이터란 조사가구의 주거관련 변수들을 조사 차수별로 모아 패널 long-type 형태로 구축한 데이터(11개 변수 포함)를 말하는데, 이 중에서 가장 핵심적인 변수는 이사여부 변수라고 할 수 있다. 이와 관련하여 이사여부 변수의 조사연도별 평균값, 즉 임의의 가구의  $t$  시점 이사확률을 살펴보면, 1999년 13.8%를 시작으로 2005년까지는 꾸준히 증가하다가 2006년부터는 감소세로 돌아서 2015년 기준 이사확률은 12.1%까지 낮아지는 모습을 보인다.

이렇게 구축된 주거패널데이터는 다른 가구특성 데이터와 병합하여 가구특성과 주거변수 사이의 관계를 분석하는데 활용될 수 있다. 본 연구에서는 급속한 증가세를 보이는 1인 가구의 주거관련 특성, 구체적으로 점유형태별(자가 및 전월세) 비율과 월 주거비 비율의 추이를 살펴보았다. 전자

와 관련하여 주목할 점은 1인 가구의 경우에도 2007년에 이미 월세 비율이 전세 비율을 초과하기 시작하였다는 점이며, 후자, 즉 월 주거비 비율 측면에서는 2002년 23%를 기점으로 꾸준한 감소세를 보이고 있는데 이는 최근 이자율 하락의 영향일 가능성이 크다는 점이다.

한편, 주거력 데이터는 주거패널데이터와는 달리 가구의 주거변화에 대한 정보를 담고 있기 때문에 더욱 유용할 수 있지만 데이터 구축에 더 많은 어려움이 따른다. 특히 해당 순번 주택의 거주 시작년도 변수를 만드는 것이 가장 어려운데, 본 연구에서는 단계별로 판단기준을 적용(<표 6> 참조)하여 이 변수를 생성하였다. 주거력 데이터에 포함된 8개 변수(<표 5> 참조) 중에서 주거이력 수와 거주기간의 평균을 살펴보면 각각 3.25번과 5.12년으로 나타났는데 두 값 모두 하향편향의 가능성이 있음에 주의해야 한다. 또한 이사시기별 구간으로 나누어 평균 거주기간을 계산해보면 최근으로 올수록 평균 거주기간이 짧아지는 양상을 보임을 알 수 있다.

분석용 데이터 구축과 관련된 마지막 단계는, 이러한 주거력 데이터를 long-type으로 변환한 후, 가구특성 데이터와 병합된 주거패널데이터와 병합하는 것이다. 이렇게 만들어진 가구특성과 주거 관련 변수를 포함한 주거력 데이터는 노동경제뿐만 아니라 도시계획, 주거학, 부동산, 교육, 아동복지 등의 분야에서도 유용한 자료가 될 것이다.

이와 관련하여 본 연구에서는 최종 주거력 데이터를 활용하여 기본적인 거주기간 분석을 수행하였다. 구체적으로 개인이나 가구의 주거이동성(residential mobility)과 관련된 선행연구들을 참고하여, 가구 및 주택 관련 특성들이 거주기간 또는 이사위험에 어떠한 영향을 미치는지에 대해 살펴보았다. 반복사건을 고려한 frailty 생존분석 모형 추정 결과, 가구주의 나이와 교육수준, 가구소득, 고등학교 이하 자녀 수, (광역시 대비) 비광역시 거주 변수는 특정 주거이력에서의 거주기간을 증가시키는 양(+)의 효과를 보인 반면, 기혼유배우 가구, 전·월세 가구의 경우에는 거주기간과 부(-)적인 관계를 나타냈다. 가구주 성별은 모든 설명변수 중 유일하게 통계적 유의성을 확보하지 못했다. 가구주 교육수준과 성별을 제외한 대부분의 변수들의 영향력은 기존 연구결과와 부합한다. 한편, 로그로지스틱 분포-frailty Gamma 분포 모형의 추정결과에 근거하여 점유형태에 따른 해저드 함수를 그려보면, 전세와 월세가구의 해저드 함수는 거의 같고, 자가 가구의 이사발생 해저드(위험)는 전월세 가구에 비해 현저하게 작으며, 거주기간이 4~5년 정도일 때 해저드가 가장 높다는 것을 알 수 있다. 가구주 연령에 따른 결과는 연령이 높아질수록 해저드가 낮아지는 형태로 나타나는데, 특정 주택에서의 거주기간이 10년 이상 오래 지속되면 가구주 연령과 상관없이 이사발생 해저드가 낮아진다.

저자들은 본 연구를 통해 제공되는 KLIPS 주거력 데이터가 다양한 분야의 연구에서 활용되기를 기대한다. 또한 구축과정에 대한 단계별 설명 자료는 유사한 형태의 이력 데이터(예, 혼인력)를 만들 때 참고 가능한, 실질적으로 도움이 되는 기초 가이드라인으로서의 활용가능성이 높을 것으로 본다.

## 참 고 문 헌

- 문근식·이현석(2016), 「표본선택이변량 프로빗모형을 이용한 지역내외 주거이동특성분석 - 연령·교육·결혼·직장특성을 중심으로」, 『주거환경』, 제14권 2호, pp. 135-146.
- 민인식(2016), 「Stata를 활용한 노동패널 실증분석」, 2016년 한국노동패널조사 데이터 설명회 자료집.
- 이경희·민인식(2016), 「패널조사 응답지속성에 관한 연구」, 『조사연구』, 제17권 3호, pp. 1-24.
- 이창호·이승일(2012), 「가구 구성원 변화에 따른 주거이동의 영향 요인 분석 - 수도권 거주가구의 주택 거주기간을 고려하여」, 『국토계획』, 제47권 4호, pp. 205-217.
- 조성진(2014), 「잠재주택수요 추정과 주거이동 결정요인 연구」, 건국대학교 부동산학과 박사학위논문.
- 최열·김형준(2012), 「수도권 및 비수도권의 주거이동 결정요인 비교 분석」, 『국토계획』, 제47권 4호, pp. 219-231.
- 최은선·남진(2011), 「자가가구와 전세가구의 거주기간에 미치는 영향 요인의 비교분석」, 『서울도시연구』, 제12권 4호, pp. 123-136.
- Beck, Brenden, Anthony Buttarro, and Mary C. Lennon(2016), “Home Moves and Child Wellbeing in the First Five Years of Life in the United States”, *Longitudinal and Life Course Studies*, 7(3), pp. 240-264.
- Decoster, André, Bart Capéau, and Kris De Swerdt(2005), “Residential Mobility in Belgium: A Duration Analysis”, CES, Katholieke Universiteit Leuven.
- Gambaro, Ludovica and Heather Joshi(2016), “Moving Home in the Early Years: What Happens to Children in the UK?”, *Longitudinal and Life Course Studies*, 7(3), pp. 265-287.
- Lennon, Mary C., William A.V. Clark, and Heather Joshi(2016), “GUEST EDITORIAL: Residential Mobility and Wellbeing: Exploring Children’s Living Situations and Their Implications for Housing Policy”, *Longitudinal and Life Course Studies*, 7(3), pp. 197-200.
- Lupton, Ruth(2016), “Housing Policies and Their Relationship to Residential Moves for Families with Young Children”, *Longitudinal and Life Course Studies*, 7(3), pp. 288-301.
- Sánchez, Aida C. and Dan Andrews(2011), “TO MOVE OR NOT TO MOVE: What Drives Residential Mobility Rates in the OECD?”, Economics Department Working Papers No.846.
- van der Vlist, Arno, Cees Gorter, Peter Nijkamp, and Piet Rietveld(2002), “Residential Mobility and Local Housing-market Differences”, *Environment and Planning A*, 34(7), pp. 1147-1164.
- Withers, Suzanne D.(1997), “Methodological Considerations in the Analysis of Residential Mobility: A Test of Duration, State Dependence, and Associated Events”, *Geographical Analysis*, 29(4), pp. 354-372.