

# 가구와 주택특성이 거주기간에 미치는 영향\*

## The Effects of Household and Housing Characteristics on the Duration of Residence

이경희 Lee Kyunghee\*\*, 민인식 Min Insik\*\*\*

### Abstract

Panel survey data basically provide useful information such as movement and status change of households because the same subject is repeatedly investigated by follow-up studies. However, making a history data containing the dynamics of household or individual requires the expertise and compactness in the data management. In this study, we describe the process of building residential history data based on Korean Labor & Income Panel Study(KLIPS) data. It is expected that this history data will be a useful asset for related researchers. In order to examine the completeness and applicability of developed data, we present the estimation results of the survival analysis model of residential duration influenced by the characteristics of household and residence. Empirical evidence from the shared-frailty survival model shows that age and education level of household head, household income, the number of children under high school and the non-metropolitan residence variable yield the positive effect of increasing the residential duration in a specific residence. Meanwhile, the effect of marital status and rental household is negatively associated with the duration. Step-by-step explanation for the building process of residential history data can be used as a guide for reference and practical assistance in creating similar types of historical data.

Keywords: The Duration of Residence, KLIPS, Residential History Data

### I. 서론

패널조사 데이터는 기본적으로 동일한 개체(Subject), 예를 들어 개인(Individual), 가구(Household), 사업체(Firm) 등의 대상에 대한 여러 시점에 걸친 반복, 추적 조사를 통해 구성된다. 따라서 일회성의 횡단면 조사 데이터에 비해 한계효과 추정 시 개체 고유의 특성(Group Heterogeneity)을 통제할 수 있다는 방법론적

이점을 가지고 있다. 그뿐만 아니라 이동(Transition), 변화(Change) 등과 같은 동적 변화(Dynamics)를 파악하는 데 유용한 정보를 제공한다. 그러나 패널데이터를 한 번이라도 사용해 본 경험이 있는 사람이라면 조사대상 개체의 동적 이력(Career or History) 정보를 분석용 데이터로 만드는 것이 생각보다 어려운 작업이라는 것을 알 수 있다. 경우에 따라서는 이러한 데이터 구축상의 어려움으로 인해 패널데이터 동적 변

\* 본 논문은 한국노동연구원의 KLI 패널 워킹페이퍼(2016년 1호)를 보완·발전시킨 것임.

\*\* 한국노동연구원 연구위원(제1저자) | Research Fellow, Korea Labor Institute | Primary Author | kheelee@kli.re.kr

\*\*\* 경희대학교 경제학과 교수(교신저자) | Prof., Dept. of Economics, Kyunghee Univ. | Corresponding Author | imin@khu.ac.kr

화를 고려하지 않고 횡단면적인 특성만을 분석내용으로 삼는 경우도 많다.

이러한 어려움을 고려하여 한국노동연구원(Korea Labor Institute: KLI)의 한국노동패널조사(Korean Labor and Income Panel Study: KLIPS)에서는 조사대상 개인의 직업력과 이력 정보를 정리하여 별도의 데이터 세트로 제공하고 있다. 본 논문에서는 직업력과 유사한 형태로서, KLIPS에 근거한 주거력(Residential History) 데이터를 구축하는 방법을 단계별로 설명하고 구축된 주거력 데이터의 활용 가능성에 대한 예시로서 거주기간에 대한 생존분석을 수행하고 그 결과를 해석한다.

개인이나 가구의 주거이동성(Residential Mobility)과 관련된 선행연구를 살펴보면 크게 두 가지 흐름을 발견할 수 있다. 하나는 주거환경학 또는 도시생태학적 관점에서 주거이동에 대한 결정 요인을 파악하고 개인이나 가구의 주거 조정 관련 의사결정 행동을 분석, 예측하고자 하는 흐름이다. 다른 하나는 주거이동에 의해 유발되는 긍정적 또는 부정적 결과(Outcome)를 이해하고자 하는 연구와 관련된다(Beck, Buttaro and Lennon 2016; Gambro and Joshi, 2016; Lupton 2016 등). 전자는 1980년대 이래 국내외 많은 연구자들이 꾸준히 연구해 온 주제인 반면, 후자는 1990년대에 시작되어 새롭게 관심을 받고 있는 분야이다(Lennon, Clark and Joshi 2016).

본 논문에서 제시하는 거주기간 생존분석은 전자와 관련된다. 본 논문과 같이 특정 주거지에서의 거주기간에 대한 결정 요인을 생존분석 방법론을 이용하여 분석한 연구로 Withers(1997), van der Vlist, Gorter, Nijkamp and Rietveld(2002), Decoster, Capéau and De Swert(2005), 최은선, 남진(2011), 이창효, 이승일(2012) 등이 있다.

기존 선행연구에서는 추적 가능한 주거력만을 추

출하여 거주기간을 분석하였지만, 본 논문에서는 주거력 구축에 대한 구체적인 알고리즘(Algorithm)을 제시하고 이에 근거한 새로운 주거력 데이터를 구축하고자 한다. 특히 KLIPS 1차(1998년)~18차(2015년) 데이터에서 조사대상 가구의 주거력 데이터를 구축하는 구체적인 과정을 제시한다. 이를 통해 주거이동 관련 연구자들이 유사한 주제를 다룰 때 도움이 되고자 한다. KLIPS는 조사대상 기간이 18년이기 때문에 대상 가구의 주거이동이 한 번만 발생하는 것이 아니라 두 번 이상 반복적으로 발생할 수도 있다. 선행연구의 생존분석은 반복 사건(Repeated Events)을 고려하지 않고 횡단면 생존분석을 시도하였다는 단점이 있다. 본 논문에서는 반복 사건 생존분석 방법론으로 Shared-Frailty모형을 설명한다. 또한 가구 및 주택특성에 초점을 두고 거주기간 또는 이사 위험이 어떻게 달라지는가를 살펴본다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 주거이동과 관련된 선행연구를 검토한다. 3장에서는 KLIPS에 근거한 주거력 데이터 구축과정을 설명하고 주요 변수에 대한 기초통계(가의 가구의 주거 이행 확률, 주택점유형태의 유지 및 전환 확률, 점유형태 및 주택유형별 평균 거주기간 등)를 계산하여 제시한다. 4장에서는 주거력 데이터를 활용하여 반복적으로 관찰할 수 있는 거주기간에 대한 생존분석모형을 추정하고 그 결과에 근거하여 점유형태와 가구주 나이에 따른 해저드함수를 비교한다. 특히 한 가구에서 주거이동(이사)이라는 사건 발생이 반복적으로 관찰되는 자료의 특성을 고려하여 Shared-Frailty 생존분석모형을 적용한다. 마지막 5장에서는 분석결과를 요약하고 연구의 시사점을 제시한다.

## II. 선행연구 검토

### 1. 이력 데이터 관련 선행연구

패널 서베이를 이용하여 이력 데이터를 구축하는 시도는 패널조사가 장기간 이루어짐으로써 자연스럽게 발생하는 부차적인 연구과제이다. Brüderl, Castiglioni, Ludwig and Proff et al.(2017)에서는 가구나 가구를 반복적으로 조사하는 패널자료를 통해 지난 조사시점과 비교하였을 때 특정 상황(주거 또는 직업 등)의 변화(Change)나 이행(Transition)을 관찰할 수 있는 사건사(Event History) 자료를 모을 수 있다는 장점이 있다고 강조한다.

이러한 사건사 데이터를 통해 특정 사건(Event)의 발생 시점 또는 지속 기간을 모델링할 수 있다. 그러나 기존 사건사 데이터는 회고적 질문에 의존하는 경우가 많기 때문에 회고 오류(Recall Mistakes)의 발생 가능성이 존재한다. 또한 회고적 질문에 대한 무응답이 상대적으로 많이 존재한다. 이러한 무응답과 응답 오류는 추정편의를 가져올 수 있는데, 패널 자료에 근거한 사건사 데이터 구축은 이러한 단점을 줄일 수 있다고 설명한다.

한국노동패널조사(KLIPS)에서는 패널조사에서 구축할 수 있는 이력 데이터로 직업력 데이터를 제공하고 있다. KLIPS의 직업력(Work History) 데이터는 1차 조사시점에서 과거에 가졌던 직업을 기술한 회고적 정보와 1차 조사 이후 조사된 개인의 모든 일자리에 관련된 정보가 담겨 있는 자료이다. 직업력 데이터는 개인이 최초로 노동시장에 진입한 이래 가졌던 모든 일자리의 변화를 살펴볼 수 있다는 장점이 있다.

한국조세재정연구원에서 조사, 발표하는 재정패널(National Survey of Tax and Benefit)에서는 부가조사

의 형태로 출산이력과 직업력 데이터를 연구자들에게 제공하고 있다. 6차연도 부가조사 형태로 조사된 출산이력은 2008~2011년 사이에 이루어진 가구원의 출산 경험을 파악하고 관련 정보를 수집하도록 한 조사이다. 출산에 관한 정보 및 출산 가구원의 근로 정보, 출산 아동의 보육시설 및 돌봄 상태 등을 조사하고 있다. 또한 2003년부터 조사시점까지 경험한 일자리에 대한 부가조사의 형태로 6차연도 조사시점에 직업력을 측정하였다. 직업은 1개월 이상 소득 활동을 한 경우 모두 응답하도록 하였으며 소득활동에는 무급 가족 종사자도 포함된다. KLIPS 직업력과 달리 재정패널의 이력 데이터는 부가조사 형태로 이루어졌다 는 점에서 차이가 있다.

민인식(2016a)의 연구에서는 재정패널의 1~8차 데이터를 이용하여 자동차 소유력과 구매력 데이터를 구축하는 방법을 제안하고 해당 데이터를 이용하여 자동차 보유와 변화에 관한 실증분석 결과를 제시하였다. 재정패널에서는 자동차 자산에 관한 정보를 매우 자세히 조사하였다. 자동차 브랜드, 모델, 구매가격, 구매연도, 제조연도, 배기량 등 관련 변수를 광범위하게 조사하고 있다. 같은 가구를 반복적으로 조사하는 재정패널 특성을 활용하면 가구의 자동차 보유가 어떤 식으로 변하고 있는지 관찰할 수 있다. 그러나 매년 서베이 조사답변을 기준으로 일관성 있는 자동차 보유 이력 데이터를 구축하기 위해서는 가구 보유 자동차가 변경되었다는 기준을 설정하고 이 기준에 따라 자동차 보유 변화를 측정할 필요가 있다. 이에 민인식(2016a)에서는 5가지 기준하에 자동차 구매력 데이터를 구축하고 자동차 보유기간에 대한 생존분석 결과를 제시하였다.

임미화, 임재만(2013)의 연구에서는 한국노동패널 1차연도(1998년)와 13차연도(2010년) 자료를 이용하여 주거력 데이터를 구축하였다. 두 시점 동안 특정

가구의 주거력유형 변화를 추적하여 이력 데이터를 구축한 후 주거력 변화와 자산변동 관계를 분석하였다. 노동패널 데이터에서 주거력을 구축하였다는 점에서 본 논문과 유사한 점이 있으나 노동패널의 단 2개 시점에서만 주거력을 구축하였다는 점에서 주거력 데이터로 활용하기에는 단점이 있다.

최은선, 남진(2011)과 이창효, 이승일(2012)의 연구에서도 노동패널 주거력 데이터를 구축하여 분석에 이용하였으나 다음과 같은 한계가 있다. 최은선, 남진(2011)에서는 KLIPS 데이터의 주거력을 충분히 사용하지 못하고 확실하게 추적 가능한 주거력만 선택하다 보니, 자가가구 509가구, 전세가구 642가구만을 분석대상으로 설정하였다. 이창효, 이승일(2012)에서는 본 논문과 달리 거주 시작연도 변수로 KLIPS에서 주어졌던 값을 그대로 사용하여 주거력 데이터가 부정확하거나 불완전할 수 있다.<sup>1)</sup>

최열, 김영민, 조승호(2010)의 연구에서는 부산시에 거주하는 만 40세 이상 성인 남녀를 대상으로 결혼 이후 최초 주거부터 현재까지의 주거력을 설문조사하였다. 결혼 이후 주거면적이 어떻게 변화하는지에 초점을 맞춘 분석이다. 양세화, 김묘정(2011)의 연구에서도 울산시의 특정 아파트 단지 거주자를 대상으로 주거력을 조사하였다. 자가가구의 81%, 임차가구의 91%가 이전 주거면적과 비교하였을 때 현재의 주거면적이 증가하였다고 응답하였고, 향후 이사에서도 주거면적을 늘릴 계획이라고 응답했다. 상기 두 연구에서 사용한 주거력 데이터는 특정 지역을 중심으로 한 소규모 회고적 서베이 데이터에 기초한다는 점이 본 논문에서 구축한 주거력 데이터와 차이가 있다.

## 2. 주거이동 관련 선행연구

개인이나 가구의 주거이동성의 주된 연구방향은 주거이동의 결정 요인을 파악하고 그러한 의사결정 행동을 분석, 예측하는 것이다. 특정 주거지에서 거주기간에 대한 결정 요인 분석은 주로 생존분석 방법론을 사용한다. Withers(1997)는 PSID(Panel Survey of Income Dynamics) 데이터와 이산 시간 로지스틱 생존 분석모형을 활용하여 주거이동성에 영향을 미치는 요인을 자가가구와 임차가구로 나누어 분석하였다. 자가가구의 경우 이사 위험은 시간과 무관하게 추정되어, 해저드함수가 거의 수평선(Horizontal Line)으로 나타났다. 또한 자가가구는 상대적으로 주거이동성이 매우 낮은 편이었으나, 임차가구는 초기 시점에서 이사 위험이 매우 높았다가 점차 감소하여 일정 기간 이후에는 거의 균일하게 유지되었다. 이사 위험을 결정하는 주요한 요인은 가구주의 나이보다는 해당 주택에 거주한 기간임을 보여준다. Withers(1997) 연구는 패널데이터를 사용하고 있다는 점에서 본 논문과 유사하다. 그러나 본 논문에서는 이산 시간(Discrete Time) 생존분석모형 대신 연속 시간(Continuous Time) 생존 분석모형을 사용하였다는 점에서 차이가 있다. 또한 Withers(1997)의 연구에서는 이사 사건이 반복적으로 발생한다는 점을 고려하지 않은 단점이 있다.

van der Vlist, Gorter, Nijkamp and Rietveld(2002)의 연구에서는 가구 단위의 주거이동성과 지역 주택시장 구조의 관계를 검증하였다. 주거이동성은 거주지역 주택시장의 수요·공급상황과 밀접한 관련이 있다. 해당 지역의 도시화 정도, 주택시장 규모, 정부의 규제 정책에 의해서 주거이동성의 차이가 어느 정도 설명

1) 이창효, 이승일(2012) 연구에서의 거주 시작연도 변수와 관련된 문제는 다음 장의 '주거력 데이터 구축' 부분에서 좀 더 자세히 설명하였음 (각주 5 참조).

된다. 그러나 지역 주택시장 구조뿐 아니라 여전히 가구특성에 의해 주거이동성이 설명되는 부분이 상당 부분 존재한다고 추정하였다. 본 논문에서는 가구 레벨의 미시데이터를 사용하는 반면, van der Vlist, Gorter, Nijkamp and Rietveld(2002)의 연구에서는 지역 레벨에서 주거이동 비율을 분석했다는 점에서 차이가 있다.

Decoster, Capéau and De Swerdt(2005)의 연구에서는 벨기에 가구패널데이터(Belgian Household Budget Survey)를 이용하여 가구 레벨에서 측정한 거주기간을 분석하였다. 반복 사건을 고려하지 않고 현 주거지에서의 거주기간에 대해 분석하였다. 첫 번째 단계에서는 비선형 최소자승법, 두 번째 단계에서는 제약된 최우추정방법을 이용하여 Duration모형을 추정한다. 지역변수, 소득, 교육수준, 그리고 이자율과 주택가격 변수들이 거주기간을 결정하는 주요한 변수라고 설명한다. 가구소득이 높을수록 주거이동성이 더 높고 이자율과 주택가격에 대한 예상(Expectation)이 높아질수록 주거이동을 늦추는 경향이 있다. 주택매매로 인해 발생하는 거래비용(취득세 포함)에 대한 정책변화가 주거이동에 미치는 효과 또한 시뮬레이션을 통해 제시하고 있는데, 거래비용 감소는 주거이동을 증가시키지만 이러한 정책 변화는 주택 매도자가 해당 주택판매가격을 높이는 효과로 인해 상쇄되는 결과로 나타난다.

최은선, 남진(2011)의 연구에서는 자가가구와 전세가구로 구분하여 거주기간에 미치는 영향 요인을 비교 분석하였다. KLIPS 데이터를 이용하였다는 점에서 본 논문과 유사하다. Cox 생존분석모형을 사용하여 분석한 결과 자가가구의 거주기간은 가구주 연령, 학력, 주택규모 변수가 유의하게 영향을 미치고 전세가구는 연령, 직업, 소득, 주택규모 변수가 거주기간과 양(+)의 관계로 나타남을 밝혔다. 특히 가구소득 변

수가 전세가구의 거주기간을 증가시킨다는 결과는 저소득층의 주거 불안정성을 의미한다고 예측할 수 있다.

이창효, 이승일(2012)의 연구는 가구구성원 변화로 인한 주거이동과 거주기간에 미치는 영향을 실증 분석하였다. KLIPS 데이터와 Cox 생존분석모형을 적용하였다. 가구구성원 변화 이후 기존 주택에서 거주하는 기간은 평균 2년 정도이고, 이는 수도권 평균 주택 거주기간인 6년의 33%에 불과하다. 주거이동 및 거주기간에 영향을 미치는 요인은 가구주 연령, 주택종류, 주택점유형태로 나타났고 가구주 연령이 높을수록, 단독주택에 거주하거나 자가가구인 경우 기존 주택에서 장기간 거주하는 것으로 확인된다. 그러나 가구주 학력, 가구원수, 가구소득은 유의하지 않았다. 특히 가구구성원이 증가하는 경우에는 고교생 이하 자녀가 있는 경우 거주기간이 길었고, 가구구성원이 감소하는 경우에는 고교생 이하 자녀가 있는 경우 오히려 거주기간이 짧은 것으로 나타났다. 가구구성원이 증가하는 경우와 감소하는 경우로 구분하여 거주기간에 영향을 미치는 요인에 차이가 있다는 분석결과를 얻었다는 점에서 연구의 의의가 존재한다.

이들 선행연구의 전체적인 결과는 가구주 연령, 학력, 직업, 성별, 인종, 통근거리, 가구소득, 가구유형, 자녀요인, 주택종류, 점유형태, 주택가격, 임대료, 거래세율 등의 가구 및 주택특성 그리고 주택시장 환경들이 거주기간과 유의한 관계가 있음을 시사한다. 가구주 연령이 높을수록, 여성이거나 소수 인종일수록, 고등학생 자녀가 있는 경우, 주택 종류가 단독주택인 경우, 자가인 경우, 주택 거래세율이 높을수록 특정 거주지에서 거주기간이 길어지는 양(+)의 효과를 보여주고 있다. 반면 가구주 학력이 높을수록, 주택가격이나 임대료가 높을수록 특정 주거지에서 주거이동

확률이 높아지는 효과를 가지는 것으로 추정된다. 가구소득, 주택규모, 가구주 직업의 효과는 점유형태—자가와 임차가구—에 따라 서로 다른 결과를 보이는 연구가 많았다. 전세가구인 경우에는 가구소득이 높거나 주택규모가 작은 경우, 가구주의 직업이 (상용직 대비) 자영업인 경우 특정 주거지에서 거주기간이 길어지는 경향을 보였다. 한편 자가가구인 경우에는 통근거리가 길어지거나 주택규모가 작아질수록 거주기간이 짧아지는 경향을 보였다.

주거이동과 관련된 선행연구 중에서 특정 주거지(주택)에서의 거주기간이 아닌 이주여부나 계획을 종속변수로 하여 로짓이나 프로빗모형을 이용한 선행연구들도 있다. Caldera Sánchez and Andrews(2011)는 OECD 국가들의 주거이동성 패턴을 노동시장 이동성과 관련지어 분석하였다. 가구와 국가 특성을 통제하였을 때 거래비용이 낮고 주택공급 수준이 높은 나라에서 주거이동성이 높고 주택모기지 접근이 쉬울수록 주거이동성이 높다는 결론을 얻었다. 문근식, 이현석(2016)의 연구에서는 주거이동 여부와 주거이동인 경우 지역 내, 외 이동으로 구분한 표본선택 이변량 프로빗모형을 설정하였다. 분석결과를 살펴보면 혼인 사건을 통해 지역 외 이동 확률이 높아지고 취학 자녀가 있는 경우에는 지역 내 이동 확률이 높아진다. 주거이동 확률이 높아지는 가구주 연령은 약 53세 전후로 추정되었고, 주거이동한 고령 가구는 지역 외 이동 확률이 높았다. 가구주와 배우자의 통근성에 따라 주거이동 및 지역 외 이동에 영향을 미치는 것으로 추정되었다.

본 논문에서는 이러한 주거이동 관련 선행연구들을 기초로 거주기간에 영향을 미칠 수 있는 변수들을

선택한다. 또한 거주기간 변수에서는 우측 절단(Right-Censored) 현상이 나타나기 때문에 생존분석모형을 사용하고 KLIPS 데이터에서 관찰되는 이사 사건의 반복성을 Shared-Frailty 형태로 고려하여 거주기간 모형을 분석한다.

### III. 주거력 데이터 및 기초통계

#### 1. 주거력 데이터 구축

본 절에서는 KLIPS 1차(1998년)~18차(2015년) 데이터에서 조사가구의 주거력을 추적하는 주거력 데이터를 구축하는 방법에 대해 구체적으로 설명한다. 패널조사의 각 시점(Wave)에서 가구의 주거 상태(거주지역, 점유형태, 주거면적, 주택유형 등)를 파악하기는 쉽다. 그러나 이사와 같이 주거변화를 판단하는 것은 연구자가 별도로 작업을 해야 한다.<sup>2)</sup> KLIPS 각 조사연도에는 이사여부 변수(h\_1401)가 있지만 이 변수만 가지고 각 가구의 이사여부를 판단하기는 쉽지 않다. 가령 특정 가구가 1~3차연도에 조사에 응답하였지만, 4차 조사에서는 제외되었다가 5차에 다시 패널에 복귀하였다고 가정하자. 만약 이 가구가 4차에 이사를 했다면 5차 조사시점에서 이 가구는 '이사하지 않음'으로 답하게 된다.<sup>3)</sup> 결과적으로 h\_1401 변수만 이용해서는 주거 변화를 판단하는 것이 불가능하다.

주거력 데이터 구축을 위해 각 가구의 거주주택에 '주택번호'를 부여함으로써 주거 변화를 추적하고자 한다. <Table 1>에서는 본 논문에서 주택변화를 판단하는 데 필요한 기준을 정리하여 제시한다.

2) 주거력 데이터 구축을 위한 Preliminary data를 위해 민인식(2016b)에서 제안한 Stata의 smart\_klips 명령어를 활용하였음.

3) 5차 조사 시점에서는 해당 연도에 이사여부를 물어보고 그에 대하여 답을 하기 때문임.

**Table 1** Steps to Assign New House ID

Step	Assignment Rule
1	A new house ID is assigned to the household when residential region gets changed between t and t-1 waves
2	If Step 1 is not satisfied, a new house ID is still given when residential ownership(own, jeonse, rent, other) gets changed
3	If Steps 1 and 2 are not satisfied, a new house ID can be assigned when housing type(apartment, single, townhouse/multi-family, other) gets changed
4	If Steps 1~3 are not satisfied, a new house ID still can be given when h_1401=1(an indicator of the move) is found in the KLIPS data
5	If Steps 1~4 are not satisfied, the house ID at time t is identical to the house ID at time t-1

<Table 1>의 기준에 따라 각 조사시점에서 거주 주택에 대해 번호를 부여할 수 있다. <Figure 1>에서는 주택번호(house\_nn)를 부여한 패널데이터 일부를 예시로 보여준다. hhid=1 가구는 조사기간 동안 4번의 주거력을 가지게 된다. 5차연도에 새로운 가구번호를

가지는 이유는 이사여부(h\_1401) 변수=1이기 때문이다. 7차연도에 새로운 가구번호를 가지는 이유는 주택점유형태가 바뀌었기 때문이며, 8차연도에 새로운 가구번호를 가지는 이유는 거주지역이 바뀌었기 때문이다. 그러나 8~15차연도에는 4번 주택에 계속 거주하였음을 알 수 있다.

실증분석의 핵심변수인 거주기간(Duration of Residence) 변수를 만들기 위해 해당 주택번호의 거주 시작연도(begin\_year)와 거주 종료연도(end\_year) 변수를 생성하여야 한다. 거주 시작연도는 KLIPS에서 h\_1415(거주 시작시기: 년)로 제공하고 있다. 그러나 같은 주택번호임에도 불구하고 h\_1415 변수값이 다른 경우가 꽤 존재한다.<sup>4)</sup> 또한 h\_1415 변수는 8차 Wave 이후에는 h\_1401=1(이사함) 경우에만 제공하고 있다. 따라서 KLIPS에서 제공하는 h\_1415 변수를 그대로 사용하여 begin\_year와 end\_year 변수를 만드는 것은 불가하다. 본 논문에서는 <Table 2>의 기준을 정해

**Figure 1** An Example of Panel Data with the Variable of House ID(house\_nn)

hhid	wave	house_nn	year	region	resid_type	hprice	h1401
1	1	1	1998	1	2		
1	2	2	1999	1	2		2
1	3	1	2000	1	2		2
1	5	2	2002	1	2		1
1	6	2	2003	1	2		2
1	7	3	2004	1	1	7000	2
1	8	4	2005	11	1	4500	1
1	9	4	2006	11	1	3000	2
1	10	4	2007	11	1	5000	2
1	11	4	2008	11	1	5000	2
1	12	4	2009	11	1	4000	2
1	13	4	2010	11	1	7000	2
1	14	4	2011	11	1	7000	2
1	15	4	2012	11	1	7000	2
1	16	4	2013	11	1	7000	2
1	17	4	2014	11	1	7000	2
1	18	4	2015	11	1	7000	2

Note:  : house\_nn=1 for hhid=1,  : house\_nn=2 for hhid=1,  : house\_nn=3 for hhid=1,  : house\_nn=4 for hhid\_1.

4) 기존 선행연구(이창효, 이승일 2012)에서는 거주 시작연도 변수를 KLIPS에서 주어진 값을 그대로 사용하였음. 따라서 합리적이지 않은 거주 시작연도를 걸러내지 못하는 문제가 있음. 또한 9차연도 이후에는 거주 시작연도가 주어지지 않기 때문에 해당 기간 데이터를 사용할 수 없는 문제가 발생함.

**Table 2\_ Construction of 'begin\_year' and 'end\_year' Variables**

Variable	Step	Assignment Rule
begin_year	1	The value of h__1415(beginning year of residence) in the first wave is assigned to the variable of begin_year. Thus the same begin_year appears for the same house ID
	2	When the begin_year is missing for a particular house ID, the first survey year is assigned to the begin_year of the house ID If h__1401(whether or not to move)=0 in that survey year A, (A-1) is assigned to the begin_year of the house ID
	3	If the begin_year allocated by Steps 1~2 is later than the survey year, the begin_year is replaced with the survey year
	4	We reasonably assume that the begin_year of house ID=1 should be earlier than the begin_year of house ID=2. If this assumption is not satisfied, the begin_year of house ID=2 is replaced with the survey year. Repeat the same assignment rule for all the other house ID's for a particular household
end_year	1	The end_year(last year of residence) of house ID=1 is filled with the begin_year of house ID=2. Construct the end_year of house ID=2,3,4 following the same rule
	2	Since we can not observe the end_year of the last house ID for a particular household, the last survey year is allocated to the end_year of that house ID

서 각 주택번호에 대해 begin\_year와 end\_year 변수를 생성한다.

<Table 2>의 기준에 따라 구축된 begin\_year와 end\_year 변수의 일부분을 <Figure 2>에서 제시한다. hhid=1의 경우 4번의 주거력을 가지고 있다. house\_nn=1은 1994년에 거주를 시작해서 2002년에 종료하였고 house\_nn=2는 2002년에 이사해서 2003년까지 거주하였다. house\_nn=3은 2003년에 다시 이사해서

2004년까지 거주하였다. 마지막 house\_nn=4는 2004년에 이사해서 마지막 조사시점인 2015년까지 계속 거주하고 있다. <Figure 2>와 같은 주거력 데이터를 모든 가구 id에 대해서 생성하면 본 논문의 실증분석에서 사용 가능한 주거력 데이터가 완성된다.<sup>5)</sup>

## 2. 주거력 기초통계 분석

연구의 관심 변수인 거주기간 변수는 아래 식과 같이 정의한다. 또한 house\_num 변수는 각 가구의 조사기간 동안 거주주택 건수로 정의한다. 가령 hhid=1 가구의 house\_num=4가 된다는 것을 <Figure 2>를 통하여 확인할 수 있다.

$$\text{duration} = (\text{end\_year} - \text{begin\_year}) + 1$$

$$\text{house\_num} = \text{각 가구의 거주주택 건수}$$

KLIPS 1~18차 패널데이터에서 구축된 주거력 데이터의 요약 통계량은 다음과 같다. 실증분석에 포함된 가구는 9,574가구이고, 각 가구는 평균 3.21개의 거주 이력을 가지고 있다.<sup>6)</sup> 최대 14번의 주거력을 가진 가구도 존재한다. 총 관측치(Observations)는 3만 560개이고 하나의 주택에서 평균 거주기간은 5.93년이다. 최소 1년~최대 60년 거주 이력 범위를 가진다. 그러나 각 가구의 마지막 주택번호에서는 end\_year값

**Figure 2\_ Residential History for hhid=1**

hhid	wave	house_nn	begin_year	end_year
1	1	1	1994	2002
1	5	2	2002	2003
1	7	3	2003	2004
1	8	4	2004	2015

5) 저자들이 구축한 주거력 데이터를 관련 분야 연구자들이 활용할 수 있도록 공개할 예정이다. KLIPS 19차(2016년) 데이터가 공개되는 2017년 9월경 주거력 데이터를 업데이트한 후 저자의 개인 블로그(blog.naver.com/housingdata)에 공개하고자 함.

6) duration 변수가 이상치(60년 초과)인 경우는 요약 통계량 계산에서 제외하였음.

을 마지막 조사시점으로 입력하였기 때문에 <Table 3>에서 제시한 평균 Duration은 하향 편향(Downward Bias)될 가능성이 있다는 점에 주의하여야 한다.

**Table 3\_ Summary Statistics for 'house\_num' and 'duration' Variables**

Variable	Sample Size	Mean	Median	Min	Max
house_num (unit: house)	9,574 house-holds	3.21	3	1	14
duration (unit: year)	30,560 obs.	5.93	3	1	60

<Table 4>에서는 점유형태와 주택유형에 따른 평균 거주기간을 계산하여 제시한다. 점유형태가 자가인 경우에 거주기간이 약 8.95년으로 가장 길다는 것을 알 수 있다. 자가 이외의 점유형태인 경우에는 평균 거주기간이 각각 3.9(전세), 3.8(월세), 4.2(기타)이다. 주택유형별로 평균 거주기간을 계산하면 단독주택인 경우가 8.88년으로 가장 길다. 아파트는 5.6년, 연립주택은 4.5년, 그리고 다세대 및 다가구는 3.9년으로 가장 짧다.

주거력 데이터를 이용하면 각 가구의 주거 변화를

**Table 4\_ Mean Residential Duration by Ownership/Housing Type**

Ownership				
Own	Jeonse	Rent	Other	
8.95 (12,206)	3.93 (10,487)	3.78 (5,337)	4.18 (2,496)	
Housing Type				
Single House	Apartment	Townhouse	Multi-Family House	Other
8.88 (8,108)	5.58 (11,181)	4.53 (3,424)	3.89 (5,556)	4.21 (2,229)

Note: The numbers in parentheses ( ) are the number of samples used in mean calculation.

관찰할 수 있으며, 이를 통해 이행행렬(Transition Matrix)을 계산할 수 있다. 거주지역에 대한 이행행렬은 <식 1>과 같이 정의한다.

$$\Pr(\text{region}_{it+1} = \text{경기} | \text{region}_{it} = \text{서울}) \quad \text{<식 1>}$$

위의 확률은 t시점에 서울에 거주하는 가구의 다음 거주지역이 경기도일 확률로 해석한다. <Table 5>에서는 거주지역을 서울, 인천, 경기로 한정하였을 때의 이행행렬을 보여준다. 이를 통해 거주지역의 계열 상관성(Serial Correlation)이 매우 높음을 알 수 있다. 서울지역에 거주한 가구는 다음 거주지 역시 서울을 선택할 확률이 87.6%이고, 경기→경기 확률은 90%, 인천→인천 확률 역시 90%이다. 한편 서울→경기로 이사할 확률은 상대적으로 높은 편(9.1%)임을 알 수 있다. 서울의 높은 주거비용이 주요한 원인으로 예상된다.

**Table 5\_ Transition Probability by Residential Region (unit: %)**

Region	Seoul(t+1)	Incheon(t+1)	Kyunggi(t+1)
Seoul(t)	87.6	0.7	9.1
Incheon(t)	2.1	89.9	5.5
Kyunggi(t)	4.9	1.2	90.3

주택점유형태에 따른 이행 확률은 <식 2>와 같이 정의할 수 있다. t시점에서 전세가가 다음 거주에서 자가로 변경될 확률을 의미한다.

$$\Pr(\text{owntype}_{it+1} = \text{자가} | \text{owntype}_{it} = \text{전세}) \quad \text{<식 2>}$$

<Table 6>에 따르면 자가→자가 확률은 67%이고, 전세→전세 확률은 54%이다. 월세 거주가가 다음 기에도 여전히 월세를 선택할 확률은 53%이다. 점유형태의 계열 상관성 역시 높다고 해석할 수 있다.

**Table 6** \_ Transition Probability by Ownership

Ownership	Own(t+1)	Jeonse(t+1)	Rent(t+1)
Own(t)	67.4	19.5	6.3
Jeonse(t)	27.9	54.2	13.3
Rent(t)	12.3	26.4	53.3

전세→자가 이행 확률은 28%이고 월세→자가 이행 확률은 12%에 불과하다. 전세 거주가구가 다음 거주에서는 자가가 될 확률이 월세 가구에 비해 2배 이상 더 높음을 확인할 수 있다.

#### IV. 거주기간에 대한 생존분석

##### 1. 계량방법론

거주기간에 대한 계량모형은 생존분석(Survival Analysis)을 선택한다. 종속변수인 거주기간(Duration) 변수는 항상 0보다 크고 정규분포(Normal Distribution)와는 다른 형태의 분포일 가능성이 크다. 또한 각 가구의 마지막 주거력은 마지막 거주시점을 관찰할 수 없기 때문에 우측 절단(Right-Censored) 현상이 나타난다. 주거력 데이터에서 드러나는 또 다른 특징은 특정 가구에 대한 사건 발생(이사 사건)이 반복적으로 관찰되는 것이다. 따라서 반복 사건(Repeated-Failure Events) 생존분석 접근법이 적용되어야 한다.

반복 사건, 즉 특정 가구에서 사건 발생이 두 번 이상 관찰될 수 있는 경우라면, 사건 발생까지 걸린 시간(Time to Event)은 같은 가구 id 내에서 서로 계열 상관을 가지게 된다. 반복 사건을 고려하는 생존분석 모형은 Shared-Fraily모형과 공분산 수정(Variance-Corrected) 접근법을 적용할 수 있다. 먼저 Shared-Fraily모형에서는 모형설정 과정에서 명확하게 특정 가구 내에서 발생한 사건 발생끼리 상관관계를 가정

한다. 특히 관찰되지 않는 Shared-Fraily( $\alpha_i$ )를 확률 효과(Random-Effects)로 포함하게 된다.  $\alpha_i$  확률변수에 대해서는 항상 양의 값을 갖는 특정한 분포함수를 가정한다. 둘째, 공분산 수정 접근법은 모형 추정에서는 반복 사건끼리의 의존성(Dependence)을 가정하지 않고 단지 추정량의 공분산 행렬(Covariance Matrix)에서 같은 가구 id 내의 상관관계를 고려하여 수정한 공분산 행렬을 제시하고, 이를 이용하여 표준오차와 계수 검정 결과를 제시한다. 본 논문에서는 Shared-Fraily 모형을 이용하여 거주기간 생존분석 결과를 제시한다.

Shared-Fraily모형은 준모수적(Semi-Parametric)인 접근법인 콕스회귀(Cox Regression)모형과 모수적(Parametric) 생존분석모형으로 나눌 수 있다. 먼저 Cox-Regression에서 가구  $i$ 의  $j$ 번째 주거력에 대해 해저드는 다음과 같이 정의한다.

$$h_{ij}(t) = h_0(t)\alpha_i \exp(x_{ij}\beta) \quad \text{<식 3>}$$

위의 <식 3>에서  $\alpha_i$ 는 가구 레벨 Shared-Fraily로 정의한다.  $h_0(t)$ 는 기준선 해저드함수(Baseline Hazard Function)이고 구체적인 형태를 가정하지 않는다. Fraily는 관찰되지 않으며 양의 값을 갖는 확률변수로 본다.  $v_i = \log\alpha_i$ 로 정의하면 <식 3>은 <식 4>와 같이 표현할 수 있다.

$$h_{ij}(t) = h_0(t)\exp(x_{ij}\beta + v_i) \quad \text{<식 4>}$$

<식 4>의  $v_i$ 는 패널선형회귀모형에서 확률효과(Random-Effects)와 유사하게 표현된다는 것을 알 수 있다. 모수적 생존분석모형에서도 가구-특정(Household-Specific) Shared-Fraily를 포함할 수 있다. 가구  $i$ 의 거주이력  $j$ 에서의 해저드함수는 다음과 같다.  $\alpha_i$ 를 통해 특정가구 내에서 거주기간은 양의 상관관계를 갖

게 된다.

$$h_{ij}(t | \alpha_i, x_{ij}) = \alpha_i h_{ij}(t | x_{ij})$$

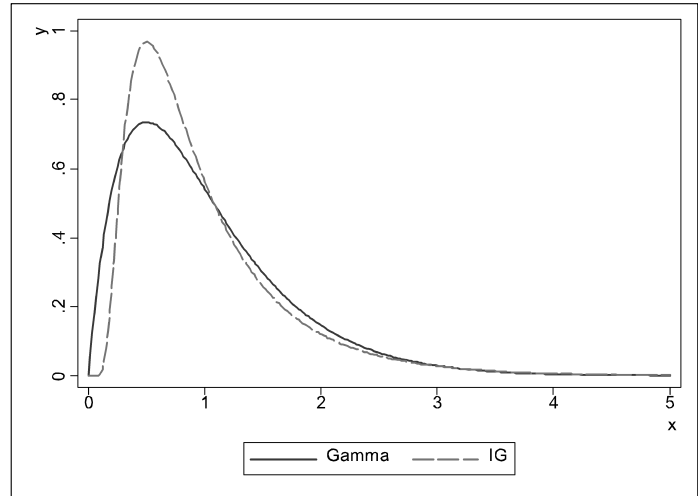
<식 5>

<식 5>에서  $h_{ij}(t | x_{ij})$ 는 모수적 해저드함수를 가정하게 된다. 모수적 분포함수로는 와이블 분포(Weibull Distribution) 또는 로그-로지스틱 분포(Log-Logistic Distribution)를 가정할 수 있다.  $\alpha_i$ 에 대한 분포함수는 평균이 1이고 분산이  $\theta$ 인 감마분포(Gamma Distribution), 또는 역가우시안 분포(Inverse Gaussian Distribution: IG Distribution)를 주로 선택한다. <Figure 3>에서는 평균=1, 분산=0.5인 감마분포와 역가우시안 분포의 확률밀도함수(Probability Density Function: PDF)를 비교한다. 두 분포 모두 0보다 큰 값을 갖고 오른쪽으로 꼬리가 긴 분포임을 알 수 있다. 관찰되지 않는 가구의 이질성(Household Heterogeneity)은 대부분 0~2 사이에 존재하고 거주 기간에 대해 극단적인 이질성을 갖는 소수의 가구가 계속 존재한다고 해석할 수 있다.

## 2. 실증분석 결과 및 해석

본 논문에서는 거주기간과 가구-주택특성의 관계를 분석하기 위해 Shared-Frailty를 포함한 모수적 생존분석모형을 추정하여 제시한다. 종속변수는 거주기간이고 설명변수는 가구의 인구통계학적 특성과 가구의 경제적 특성, 그리고 주택특성변수를 포함한다. <Table 7>에서는 종속변수와 설명변수를 정리하여 보여준다. 종속변수인 Duration 변수의 극단치를 제외하기 위해  $duration \leq 60$ 인 관측치만 분석에 포함하

Figure 3\_Density Plot: Gamma and Inverse Gaussian Distributions



였다. 추정에 사용된 표본은 8,914가구의 2만 6,504개 관측치(Obs)이다.

Shared-Frailty를 포함한 생존분석모형 추정결과는

Table 7\_Description of Variables

Variable Name	Description
Duration	Residential Duration (the duration for the right-censored house_id is computed from the last survey year)
Age	Age of Household Head
Sex	Gender of Household Head (1=male, 2=female)
Hnum	The Number of Children under High School Age
Edu	Education Level of Household Head (1=less than high-school diploma, 2=high-school diploma, 3=two-year college diploma and higher)
Married	Martial Status of Household Head (1=non-married(single/divorced/bereaved, 2=married(with spouse))
Region	Residential Region (1=metropolitan, 2=non-metropolitan)
Lincome	Log Household Income(2015 constant price)
Residtype	House Ownership (1=own, 2=jeonse, 3=rent and others)

<Table 8>에 제시되어 있다. 모수적 분포는 와이블 분포와 로그-로지스틱 분포로 가정하고, 각 모수적 분포하에서 Frailty 분포는 감마와 역가우시안 분포로 가정한다. 따라서 총 네 가지 모형 추정결과를 도출할 수 있다. 모수적 분포를 와이블 분포로 가정하는 경우에는 PH(Proportional Hazard) Metric과 AFT(Accelerated

Failure Time) Metric으로 추정결과를 제시할 수 있지만 로그-로지스틱 분포 가정에서는 AFT Metric 결과만 제시할 수 있기 때문에 <Table 8>에서는 비교를 위해 와이블 모형에서도 위험률(Hazard Ratio) 대신 추정계수(Estimated Coefficient)를 얻는 AFT Metric 결과를 제시한다.

**Table 8**\_Survival Analysis Results: Estimated Coefficients of Shared-Frailty Model

Variable	Parametric Distribution : Weibull Distribution		Parametric Distribution : Log-logistic Distribution	
	Frailty Distribution: Gamma	Frailty Distribution: IG	Frailty Distribution: Gamma	Frailty Distribution: IG
<i>age</i>	0.031 (0.0006)***	0.033 (0.0006)***	0.022 (0.0005)***	0.023 (0.0005)***
<i>lincome</i>	0.063 (0.007)***	0.059 (0.007)***	0.086 (0.007)***	0.085 (0.007)***
<i>hnum</i>	0.027 (0.008)***	0.023 (0.008)**	0.024 (0.007)***	0.022 (0.007)***
<i>2.sex</i> (female)	-0.005 (0.021)	-0.009 (0.021)	0.018 (0.019)	0.019 (0.018)
<i>2.married</i> (married)	-0.087 (0.019)***	-0.096 (0.020)***	-0.064 (0.018)***	-0.067 (0.018)***
<i>2.region</i> (non-metropolitan)	0.102 (0.014)***	0.118 (0.014)***	0.083 (0.013)***	0.092 (0.013)***
<i>2.residtype</i> (jeonse)	-0.649 (0.015)***	-0.671 (0.015)***	-0.473 (0.014)***	-0.485 (0.015)***
<i>3.residtype</i> (rent)	-0.651 (0.017)***	-0.671 (0.017)***	-0.484 (0.017)***	-0.497 (0.017)***
<i>2.edu</i> (high-school diploma)	0.151 (0.021)***	0.141 (0.021)***	0.226 (0.020)***	0.230 (0.021)***
<i>3.edu</i> (two-year college and higher)	0.227 (0.023)***	0.213 (0.024)***	0.328 (0.022)***	0.339 (0.023)***
constant	0.205 (0.075)***	0.156 (0.076)**	-0.104 (0.074)	-0.152 (0.075)**
$\log(p)$	0.442 (0.006)***	0.432 (0.006)***		
$\log(\theta)$	-0.545 (0.036)***	-0.334 (0.048)***		
$\log(\gamma)$			-0.762 (0.007)***	-0.759 (0.007)***
$\log(\theta)$			-0.944 (0.045)***	-0.888 (0.053)***
$\log L$	-28996	-29045	-28458	-28482
LR test $H_0 : \theta = 0$	2,005***	1,906***	1,037***	989***
<i>obs</i>	26,504	26,504	26,504	26,504

Note: \*\*\*, \*\*, \* indicate that the estimate is significant at 1%, 5% and 10% levels, respectively.

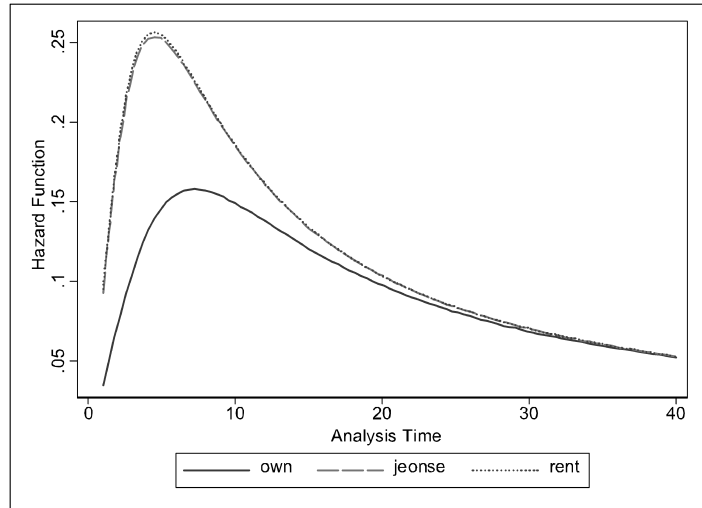
와이블 분포-Gamm Frailty를 가정한 생존분석 결과를 살펴보면  $\hat{p}=1.56$  이기 때문에 해저드함수가 증가하는 패턴임을 알 수 있다. 즉 다른 조건이 일정할 때 거주기간이 길어질수록 이사 사건 발생 리스크가 높아진다. 또한  $\alpha_i$ 의 분산 추정치  $\hat{\theta}=0.58$ 이고, 이 값은 0과 유의하게 다르다는 것을 알 수 있다. 각 가구의 거주기간에 관찰되지 않는 이질성이 통계적으로 유의하다고 해석한다. 로그-로지스틱 분포-Gamma Frailty를 가정한 생존분석 결과에서는  $\hat{\gamma}=0.467$ , 즉 1보다 작기 때문에 비단조적(Non-Monotone) 해저드함수를 가진다는 것을 예상할 수 있다.<sup>7)</sup> 거주기간이 일정 이상 길어지면 오히려 이사 발생 위험이 줄어든다. 분산 모수 추정치  $\hat{\theta}=0.389$ 이며, 역시 통계적으로 0과 유의하게 다르다. 거주기간에 영향을 미치는 가구 이질성이 유의하게 존재한다.<sup>8)</sup>

주요 설명변수와 거주기간의 관계를 살펴보면 대부분 예상과 일치함을 알 수 있다. 가구주 나이(Age)가 많을수록, 소득수준(Lincome)이 높을수록 거주기간이 길어진다. 또한 고등학생 이하 자녀수(Hnum)가 많을수록 거주기간이 길어진다고 추정된다. 초·중·고에 재학 중인 자녀가 있으면 전학 리스크가 발생하기에 빈번한 이사를 자제할 가능성이 있기 때문이다. 가구주의 성별은 통계적으로 유의하지 않은 반면, 가구주 혼인 여부는 음(-)으로 유의하다. 즉 기혼 유배우 가구에서 거주기간이

더 짧다. 광역시 거주가구의 경우 비광역시 거주가구에 비해 거주기간이 유의하게 짧다. 비광역시에 비해 광역시의 주거비용(주택가격 및 전월세 비용)이 더 높기 때문에 예상할 수 있는 결과이다. 점유형태를 살펴보면 자가 거주에 비해 전세나 월세 가구의 거주기간이 유의하게 짧은 것으로 추정된다. 가구주의 교육수준은 높을수록 거주기간이 유의하게 길어진다. Gamma-Frailty 결과에 대한 강건성 확인(Robustness Check)으로 IG-Frailty 분포를 가정한 결과 역시 <Table 8>에 포함되어 있다. 추정계수의 부호와 유의성은 Gamma-Frailty 경우와 거의 차이가 없다는 것을 확인할 수 있다.

<Figure 4, 5>에서는 관심 설명변수의 특정한 값에서 그린 해저드함수를 비교한다.<sup>9)</sup> 그래프 작성에 사용된 추정결과는 로그-로지스틱 분포와 Gamma-Frailty

Figure 4 \_ Estimated Hazard Function by Ownership

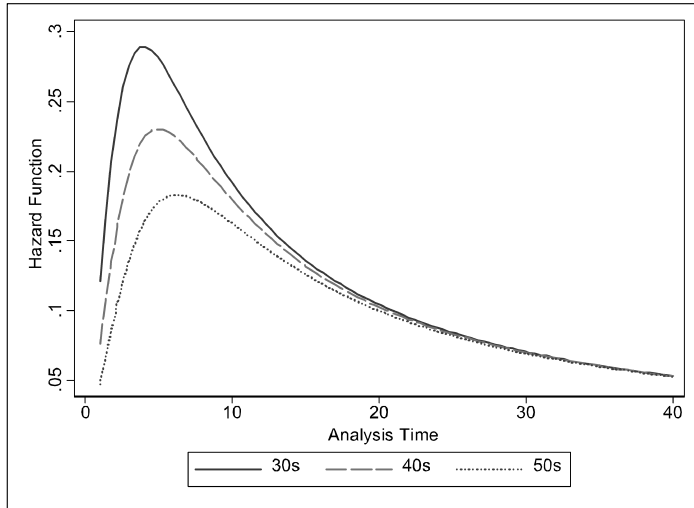


7) 로그-로지스틱 분포에서  $\gamma$ 값과 해저드함수 형태의 관계는 이경희, 민인식(2016)을 참고할 수 있음.

8) 귀무가설을 기각하기 때문에 반복 사건을 고려하지 않는 횡단면 생존분석 대신 반복 사건을 고려한 shared-frailty 생존분석이 더 적절한 모형임을 통계적으로 확인할 수 있음.

9) 관심 변수 이외의 변수는 평균에서 고정시킨 상태에서 그래프를 작성함.

Figure 5\_ Estimated Hazard Function by Household Head's Age



분포를 사용하였다.<sup>10)</sup> <Figure 4>에서는 점유형태(자가, 전세, 월세)에 따른 해저드함수 그래프를 보여 준다. <Table 8>에서  $\hat{\gamma} < 1$  이기 때문에 비단조적 해저드함수로 나타난다. 전세와 월세가구의 해저드함수는 거의 차이가 없으며 자가가구의 이사 발생 리스크는 전월세가구에 비해 현저하게 낮음을 확인할 수 있다. 전월세가구의 경우에는 Duration=4년 정도에 이사 발생 위험(Hazard)이 가장 높아지고 자가가구의 경우에는 Duration=6년 정도에 위험이 가장 높게 추정된다.

<Figure 5>에서는 가구주 연령에 따른 이사 발생 위험을 그래프로 제시하였다. 가구주 연령은 35세, 45세, 55세의 세 가지 경우로 고정한다. Age=35세일 때 사건 발생 리스크가 가장 높고 연령이 높아질수록 해저드가 낮아진다. 그러나 거주기간이 10년이 넘으면 이사 발생 위험은 가구주 연령과 무관해진다. 특정 주택에서 10년 이상 거주하면 가구주 나이와 무관하게 이사 위험이 매우 낮아진다는 것을 예상할 수 있다.

10) 해저드함수를 그리기 위한 모형은 로그우도함수값이 가장 큰 분포조합을 선택하였음.

## V. 요약 및 결론

서론에서 주지하였듯이 패널조사 데이터는 동일 대상을 여러 차수에 걸쳐 추적 조사한 내용을 담고 있기 때문에 조사대상의 이동(Transition)과 같은 동적 변화(Dynamics)를 파악하는 데 유용한 정보를 담고 있다. 본 논문에서는 KLIPS 데이터의 직업력과 유사하게 조사가구의 주거 변화를 추적한 주거력 데이터를 구축하는 방법을 소개하고 완성된 데이터를 연구자들이 활용할 수 있도록 기

여하고자 한다. 또한 구축된 데이터에 기초하여 각 주거지에서 거주기간에 대한 생존분석을 수행함으로써 그 활용가능성을 살펴보았다.

주거력 데이터는 가구의 주거변화에 대한 정보를 담고 있기 때문에 관련 연구자들에게 꼭 필요하지만 그 구축 과정이 그리 간단하지 않다. 특히 해당 순번 주택의 거주 시작연도 변수를 만드는 것이 가장 어려운 문제이다. 조사된 변수값과 필자들의 정한 기준을 적용하여 거주주택에 대한 시작연도와 종료연도 변수를 완성하였다. 이를 통해 주거력의 평균은 3.21번이고, 평균 거주기간은 5.93년임을 확인할 수 있다. 연구용 데이터를 구축하기 위해 주거력 데이터를 가구특성과 주택특성변수들과 병합(Merge)하였는데, 이렇게 병합된 데이터는 노동경제, 도시계획, 주거학, 부동산학, 교육, 아동복지 분야에서도 유용한 자료가 될 것으로 예상된다.

거주기간에 대한 생존분석모형을 통해 가구 및 주택특성변수들이 거주기간 또는 이사 위험에 어떠한

영향을 미치는지에 대해 살펴보았다. 반복 사건을 고려한 Shared-Frailty 모형 추정결과, 가구주의 나이와 교육수준, 가구소득, 고등학생 이하 자녀수, 비광역시 거주변수는 거주기간을 증가시키는 양(+)의 효과를 보인 반면, 기혼 유배우와 전월세 가구인 경우에는 거주기간과 부(-)적인 관계를 나타냈다. 가구주 성별은 모든 설명변수 중 유일하게 통계적 유의성을 확보하지 못했다. 가구주 교육수준과 성별 변수를 제외한 대부분의 변수들의 영향력은 선행연구와 부합한다. 로그-로지스틱 분포와 Gamma Frailty 분포를 사용한 모형에 근거하여 점유형태에 따른 해저드함수를 그려 보면 전세와 월세가구의 해저드함수는 거의 같은 패턴이고, 자가가구의 이사 발생 해저드는 전월세 가구에 비해 현저하게 낮다. 거주기간이 4~5년 정도일 때 이사 발생 해저드가 가장 높음을 확인할 수 있다. 가구주 연령에 따른 결과를 살펴보면 연령이 높아질수록 해저드가 낮아지는 형태로 나타나지만, 특정 주택에서 10년 이상 거주하면 가구주 연령과 무관하게 이사 발생 해저드가 매우 낮아진다.

본 논문에서는 주택과 가구특성이 거주기간에 미치는 영향을 실증적으로 분석하고자 하였다. 그러나 거주기간, 특히 자가 거주기간에 대해서는 부동산 관련 세제 역시 중요한 영향을 미칠 수 있다. 가령 1가구 1주택 양도세 면제 조건 또는 장기 보유에 따른 양도세 특별공제 제도는 자가 거주기간에 영향을 줄 수 있다. 이러한 부동산 관련 세제 변화가 거주기간에 유의한 영향을 미치는지 분석하지 못한 점은 본 논문의 한계점이라고 말할 수 있다. 다만 본 논문에서 제시한 KLIPS 주거력 구축과정에 대한 단계별 설명은 유사한 형태의 이력 데이터(혼인력 또는 출산력 등)를 만들 때 참고할 수 있고 실질적으로 도움이 되는 기초 가이드라인으로 활용될 수 있다는 점에서 본 논문의 또 다른 기여가 존재한다.

## 참고문헌 •••••

1. 문근식, 이현석. 2016. 표본선택 이변량 프로빗모형을 이용한 지역내의 주거이동 특성분석: 고령, 교육, 결혼, 직장 특성을 중심으로. *주거환경* 14권, 2호: 135-146.  
Mun Keunsik and Lee Hyunseok. 2016. An analysis for determinants of regional and local residential mobility in using bivariate probit model with sample selection. *Journal of The Residential Environment Institute of Korea* 14, no.2: 135-146.
2. 민인식. 2016a. 자동차 소유력과 구매력데이터 구축 및 활용. 2016년 재정패널 학술대회. 11월 25일, 서울: 전경련회관 컨퍼런스센터.  
Min Insik. 2016a. Building-up and application of car purchase history data. *Annual Conference 2016 of National Survey of Tax and Benefit*, November 25. Seoul: FKI Tower Conference Center.
3. \_\_\_\_\_. 2016b. Stata를 활용한 노동패널 실증분석. 2016년 한국노동패널조사 데이터 설명회 및 기초분석 방법론워크숍. 7월 8일, 서울: 중소기업중앙회 회의실.  
\_\_\_\_\_. 2016b. Statistical analysis using stata for KLIPS. *Annual Workshop 2016 of KLIPS Data Structure and Empirical Analysis*, July 8. Seoul: KBIZ Conference Hall.
4. 양세화, 김묘정. 2011. 자가거주가구와 임차가구의 주거이동 특성 비교. *한국가정관리학회지* 29권, 3호: 13-22.  
Yang Sehwa and Kim Myojung. 2011. A comparative study of residential mobility between home-owning households and renting households. *Journal of Korean Home Management Association* 29, no.3: 13-22.
5. 이경희, 민인식. 2016. 패널조사 응답지속성에 관한 연구: 한국노동패널조사를 중심으로. *조사연구* 17권, 3호: 1-24.  
Lee Kyunghee and Min Insik. 2016. A study on sample attrition in panel survey: Focusing on Korean Labor and Income Panel study. *Survey Research* 17, no.3: 1-24.
6. 이창효, 이승일. 2012. 가구구성원 변화에 따른 주거이동의 영향요인 분석: 수도권 거주가구의 주택 거주기간을 고려하여. *국토계획* 47권, 4호: 205-217.  
Yi Changhyo and Lee Seungil. 2012. Analyzing the factors on residential mobility according to the household member's change: In consideration of residential duration of the households in the Seoul metropolitan area. *Journal of Korea Planning Association* 47, no.4: 205-217.

7. 임미화, 임재만. 2013. 우리나라 가구의 주거력유형과 자산변동. *부동산연구* 23집, 3호: 79-97.  
Lim Mihwa and Lim Jaeman. 2013. An analysis of housing careers on Korean household and wealth variation. *Korea Real Estate Review* 23, no.3: 79-97.
8. 최열, 김영민, 조승호. 2010. 생애주기에 따른 주거이동 특성 분석. *대한토목학회논문집* 30권, 3D호: 313-321.  
Choi Yeol, Kim Youngmin and Cho Seongho. 2010. The analysis on the characteristics of residential mobility by life-cycle. *Journal of the Korean Society of Civil Engineers* 30, no.3D: 313-321.
9. 최은선, 남진. 2011. 자가가구와 전세가구의 거주기간에 미치는 영향 요인의 비교분석. *서울도시연구* 12권, 4호: 123-136.  
Choi Eunsun and Nam Jin. 2011. A study on comparing effect factors on the residential duration of homeowner and rental household. *Seoul Studies* 12, no.4: 123-136.
10. Beck, B., Buttaro, A. and Lennon, M. C. 2016. Home moves and child wellbeing in the first five years of life in the United States. *Longitudinal and Life Course Studies* 7, no.3: 240-264.
11. Bründerl, J., Castiglioni, L., Ludwig, V., Proff, K. and Schmiedeberg, C. 2017. Collecting event history data with a panel survey: Combining an electronic event history calendar and dependent interviewing. *Methods, Data, Analyses* 11, no.1: 45-66.
12. Caldera Sánchez, A. and Andrews, D. 2011. *To Move or Not to Move: What drives residential mobility rates in the OECD?* Paris: OECD Economics Department Working Paper no.846.
13. Decoster, A., Capéau, B. and De Swert, K. 2005. Residential Mobility in Belgium: A Duration Analysis. Leuven: Center for Economic Studies, Katholieke Universiteit Leuven.
14. Gambaro, L. and Joshi, H. 2016. Moving home in the early years: What happens to children in the UK?. *Longitudinal and Life Course Studies* 7, no.3: 265-287.
15. Lennon, M. C., Clark, W. and Joshi, H. 2016. Residential mobility and wellbeing: Exploring children's living situations and their implications for housing policy. *Longitudinal and Life Course Studies* 7, no.3: 197-200.
16. Lupton, R. 2016. Housing policies and their relationship to residential moves for families with young children. *Longitudinal and Life Course Studies* 7, no.3: 288-301.
17. van der Vlist, A. J., Gorter, C., Nijkamp, P. and Rietveld, P. 2002. Residential mobility and local housing-market differences. *Environment and Planning A* 34, no.7: 1147-1164.
18. Withers, S. D. 1997. Methodological considerations in the analysis of residential mobility: A test of duration, state dependence, and associated events. *Geographical Analysis* 29, no.4: 354-372.

- 
- 논문 접수일: 2017. 3. 23.
  - 심사 시작일: 2017. 4. 4.
  - 심사 완료일: 2017. 4. 13.

---

## 요약

주제어: 거주기간, KLIPS, 주거력 데이터

패널조사 데이터는 기본적으로 동일 대상을 여러 차례에 걸쳐 관찰하고 추적 조사한 내용을 토대로 하기 때문에 조사대상의 이동 및 상태 변화와 같은 유용한 정보를 제공한다. 그러나 이러한 이력 변화를 분석용 데이터로 만드는 일은 전문성과 치밀성을 요구한다. 본 논문에서는 한국노동패널조사(KLIPS) 데이터에 근거하여 조사가구의 주거력 데이터를 구축하는 과정을 설명한다. 이러한 이력 데이터는 관련 패널분석 연구자들에게 유용한 자료가 될 수 있을 것으로 예상된다. 주거력 데이터의 완성도 및 활용 가능성을 검토하기 위해 가구특성과 주택특성이 거주기간에 미

치는 영향을 생존분석모형 추정결과를 통해 제시한다. 반복 사건을 고려한 Shared-Frailty모형 추정결과, 가구의 나이와 교육수준, 가구소득, 고등학생 이하 자녀수, 비광역시 거주변수는 특정 주거력에서의 거주기간을 증가시키는 양(+)의 효과를 보인 반면, 기혼 유배우 또는 전월세 가구의 경우에는 거주기간과 부(-)적인 관계를 나타내었다. 주거력 데이터 구축 과정에 대한 단계별 설명은 유사한 형태의 이력 데이터를 만들 때 참고 및 실질적 도움이 되는 가이드라인으로 기여할 수 있다.

