

—  
**정책모기지  
DB 분석 및 활용**

# 04

송완영, 민병철, 백인걸, 김형준  
한국주택금융공사 주택금융연구원 연구위원

## CONTENTS

149	I 서론
151	II 공사 DB의 현황 및 활용방안
157	III DB자료를 이용한 연체율 분석
178	IV 활용방안: 주택금융정보시스템 구축
181	V 결론 및 시사점
183	참고문헌

## 요약

### ● 연구배경

- ▶ 공사 주택담보대출 DB(DB)는 관련연구 수행에 한계
  - 화면 및 전산 요청을 통한 간접 접근 방식으로 운영되어 제한된 자료 접근
  - 지속적인 상시 관찰이 용이하지 않아 자료 분석 기반 연구(data-driven research) 수행에 어려움
- ▶ 이에 연구용 주택담보대출DB 설계 및 구축을 통해 주택담보대출 관련 연구 적시성 제고

### ● DB 설계 목표

- ▶ 상시 접근과 자유로운 조회가 가능한 연구목적 주택담보대출DB
  - 원자료(raw data)에 가까운 자료 제공으로 다양한 활용 가능성 유지 및 외부 공동 연구가 가능하도록 필요 시 비식별 조치 적용 솔루션 포함

### ● 생존분석 적용 차주 연체율 분석

- ▶ 생존분석 방법론과 공사 DB자료를 활용하여 연체율을 분석한 결과, 대출목적 및 담보의 소재지에 따라 유의한 연체율의 차이를 확인

### ● 주택금융정보시스템 구축

- ▶ 연구용 DB를 중심으로 외부 관련기관 자료와 정보연계를 통한 주택금융정보시스템을 구축하고, 정책지표 발굴 등 후속 연구수행을 통해 주택금융시장 모니터링과 정책지원 기능 강화

## I 서론

- 데이터 수집, 저장, 분석 기술의 비약적인 발달로 4차 산업혁명이 진행됨에 따라 데이터 기반 의사결정에 대한 관심이 급증
  - ▶ 데이터 수집이 용이하지 않았던 과거에는 직관에 의존한 결정이 많이 이루어졌는데, 직관을 통한 결정이란 개인의 축적된 경험이라는 정형·비정형 데이터를 이용한 의사결정방법이라 볼 수 있음
  - ▶ 데이터 축적이 충분히 이루어지지 않았거나 최신 데이터 수집이 어려울 경우, 오랜 경험에 기반한 직관적 결정이 더 나올 수 있음
  - ▶ 하지만 현대사회에서는 데이터 수집 및 저장 기술의 발달로 이러한 어려움이 없어 데이터의 적극적인 활용이 대세
- 공사는 업무와 관련된 풍부한 공사내부자료를 DB화하여 보유하고 있으나, 이를 다양한 의사결정과정에 적극적으로 활용하고 있다고 보기는 다소 어려움
  - ▶ 주택정보를 포함한 주택담보대출 미시정보 DB 미 구축으로 주택시장과의 연계분석 제한적
  - ▶ 자금흐름이 아닌 잔액중심 집계로 인해 주택담보대출시장 현황 및 환경 변화 분석 제한적
  - ▶ 차주 대상이 한정적인 장기고정금리대출로, 전체 주택담보대출 특성으로 일반화에 한계
- 이 보고서에서는 공사 DB의 현황을 분석하고, 이를 적극적으로 활용하기 위한 연구용 DB의 설계방향 및 이를 활용한 주택금융정보시스템 구축방안을 제시함
  - ▶ 주택금융 Needs 다양화, 주택 특수성, 주택시장 변동성 확대, 주택담보대출시장 확대 추세를 고려한 종합적 분석 필요

\* 본고의 내용은 필자의 개인 의견으로 한국주택금융공사의 공식적인 견해와 다를 수 있습니다.

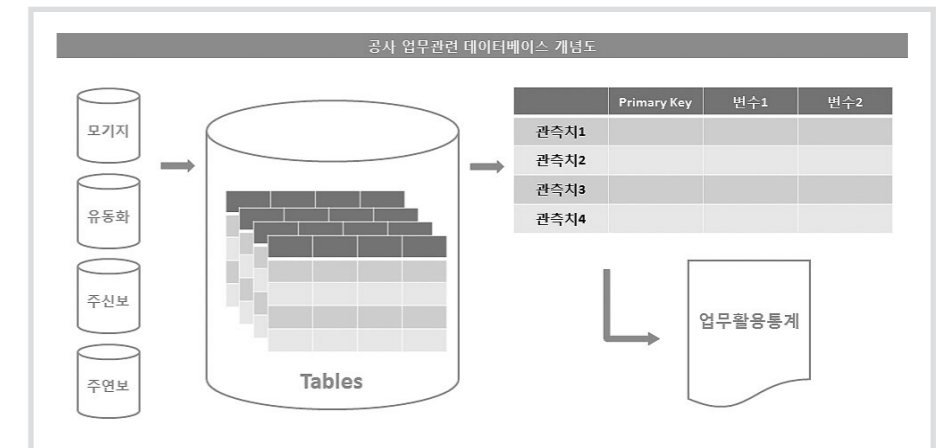
- ▶ 차주 대출특성, 가구특성, 상환방식, 담보특성, 담보물의 가격변화 정보를 통합 축적하여 특성별 및 그룹별 분석을 통하여 다양한 정책지표 개발과 여신기준 세분화에 활용
- ▶ 장기적으로 주택시장 변동성 전이에 대한 사전적 대응체계 구축과 상시(실시간) 모니터링 시스템을 통하여 주택금융시장의 안정성 제고 및 주택금융지표 제공

## II 공사 DB의 현황 및 활용방안

### 1. 현황 및 연구용 DB의 필요성

- 공사의 업무관련 DB는 자료의 수집 및 관리를 위한 설계로, 자료 활용을 위해서는 별도의 생성과정 필요 ([그림 4-1])
  - ▶ 모기지, 유동화, 주신보, 주연보 계정으로 구분
  - ▶ 각 계정은 해당 업무와 관련된 여러 테이블을 보유하고 있으며, 각 테이블은 여러 칼럼들로 구성
    - 예: 주연보 계정은 보증접수원장·담보 물건 정보 등의 테이블로 구성되어 있으며, 보증접수원장 테이블에는 접수일자·배우자유무 등의 칼럼들이 있음

[그림 4-1] 공사 업무관련 DB 개념도



※ 자료 : Google, 네이버, 한국주택금융공사

- 현재 공사의 DB는 업무에서 발생하는 다양한 정보를 수집·관리하기 위하여 수많은 테이블을 가지고 있으며, 자료의 요청이 있을 때마다 이를 기본 키(primary key)로 연결하는 구조
  - ▶ 업무용 DB는 자료의 효율적인 축적 및 관리를 목적으로 하므로 그 수가 방대하고 즉각적 활용이 어려움 ([표 4-1])

[표 4-1] 공사 업무관련 DB의 대략적 규모

DB	테이블 수(개)	총 칼럼 수(개)
모기지	300+	8000+
유동화	800+	19000+
주신보	600+	13000+
주연보	300+	7000+

▶ 개별 연구마다 맞춤형 자료를 따로 생성하므로 자료 정의에서 생성까지 시간이 소요되며, 전산 업무가 과중되는 결과 초래

● 공사는 풍부한 공사내부자료와 전문연구인력을 보유하고 있으나, 자료 접근성이 낮아 이를 충분히 활용하고 있지 못함

▶ 연구자료 확보를 위해서는 연구계획수립 및 연구계획서 제출, 필요자료 요건정의, 사업부서 및 전산부서 협조 등의 단계 필요

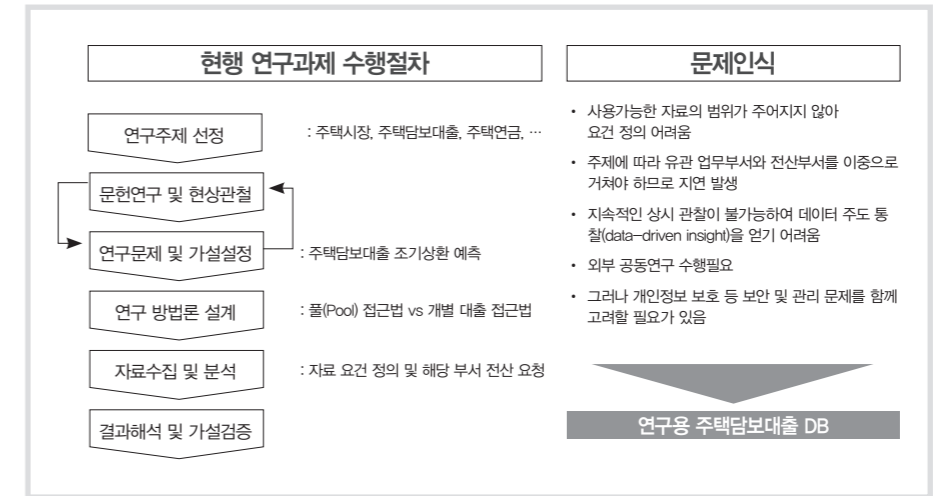
▶ 연구계획수립 이후 자료 확보까지 1~2주 가량의 시간이 소요되어 연구시간 확보에 어려움

▶ 연구위원들은 공사 DB에서 보유하고 있는 자료의 범위와 내용을 파악하지 못하여 연구에 한계

▶ 先 계획, 後 분석의 현행 연구과정으로는 자료분석에서 연구 아이디어를 얻는 자료 분석 기반 연구 불가능

● 이러한 자료 접근의 제한성은 연구를 일회성으로 그치게 하며, 지속적인 상시 관찰을 통하여 숨은 트렌드나 의미를 파악하는 연구 수행을 불가능하게 하므로 기존 DB를 활용한 연구용 주택담보대출DB가 필요함

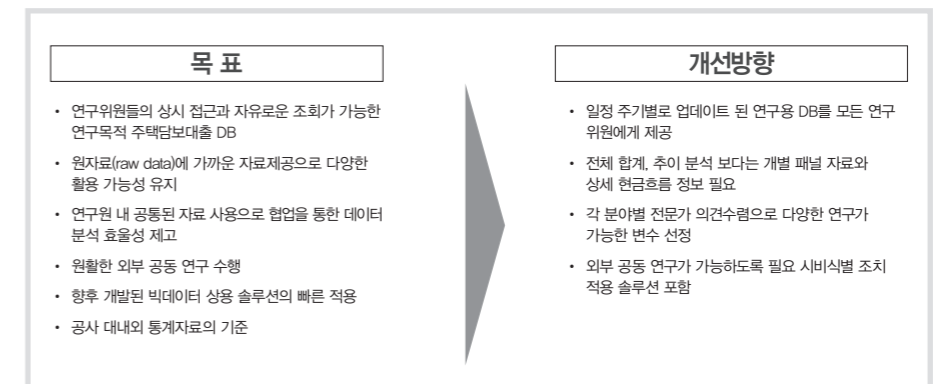
[그림 4-2] 연구용 주택담보대출DB의 필요성



## 2. 연구용 주택담보대출 DB의 설계

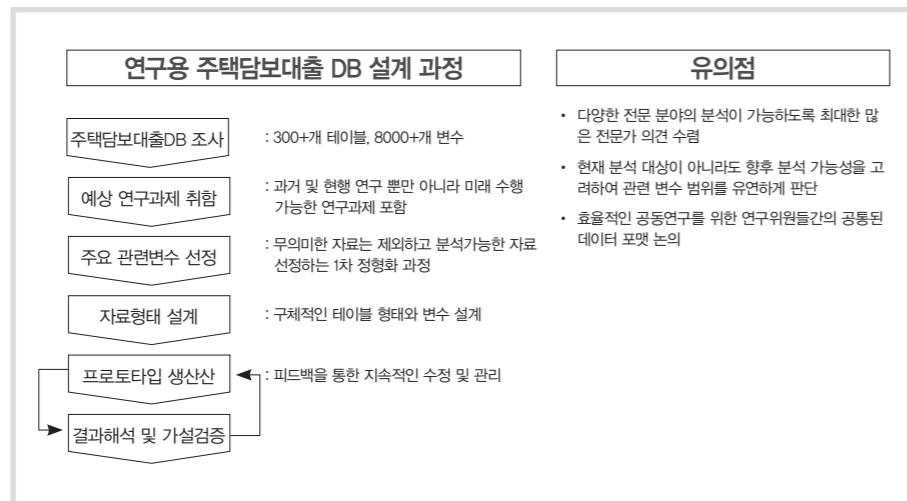
● 연구 중심 DB는 기존의 관리 및 업무활용 DB와 다른 접근이 필요하므로, 명확한 목표를 설정하고 이를 토대로 개선방향과 필요요건을 설계

[그림 4-3] 연구용 주택담보대출DB의 목표



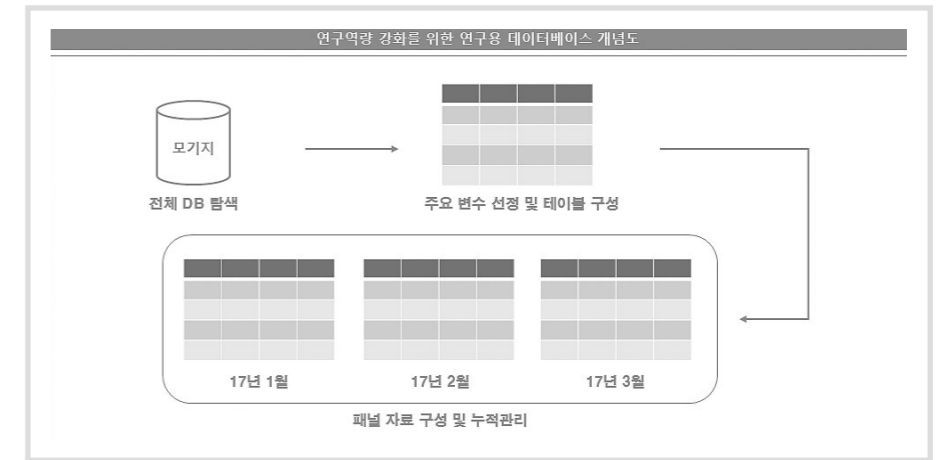
- 참여하는 구성원의 의견을 수렴하고, 장래 확장성을 감안하여 다양한 관련변수 선정
- 장기적으로 주택금융정보시스템 구축 작업과의 연계를 염두에 두고 이와 관련된 자료 선정
  - ▶ 개별적으로 가지는 정보는 적더라도, 결합을 통해 시너지 창출 가능한 자료들을 함께 포함

[그림 4-4] 연구용 주택담보대출 DB의 설계 과정



- 각 업무용 DB 계정으로 부터 필요한 자료를 추출하여 연구용 원장 자료를 생성, 이를 매월 누적 관리([그림 4-5])
  - ▶ 직관적인 자료 이해를 위하여 각 업무분야 별로 연구에 필요한 자료를 하나의 테이블로 요약
  - ▶ 연구위원과 업무부서의 의견을 반영하여 연구 및 현황파악에 필요한 자료를 선정
  - ▶ 동일한 형태의 자료를 월별로 누적 관리하는 이유는 담보물의 교체 등과 같은 시계열 변화를 함께 살펴보기 위함

[그림 4-5] 연구용 주택담보대출 DB 개념도

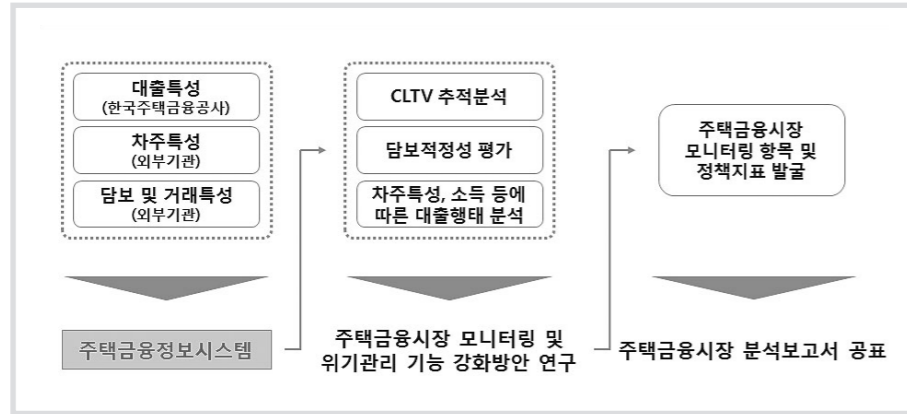


- 데이터 축적 과정에서 발생하는 원본 데이터 소실, 오기, 불완전 데이터에 대하여 일관된 데이터 정화(data cleansing) 작업 실시
  - ▶ 개별적으로 실시하던 데이터 정화작업을 표준화하여 원장 누적 단계에서 실시함으로써 동일한 자료를 사용한 연구 가능

### 3. 연구용 주택담보대출DB의 의의

- 주택금융연구의 허브 역할
  - ▶ 연구용 주택담보대출DB를 중심으로 외부 관련 기관과의 정보연계를 통해 주택금융정보시스템을 구축하고, 데이터 기반 주택금융시장 모니터링과 정책지원 기능 강화

[그림 4-6] 주택금융정보시스템 구축 및 활용방안 예시

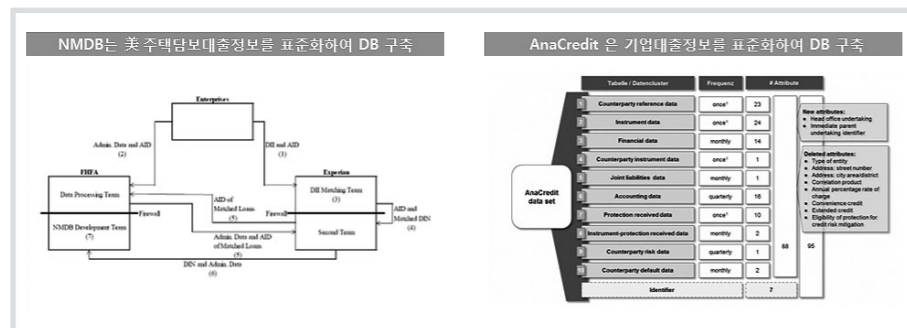


● 주택담보대출DB의 표준

▶ 데이터에 기반한 시장 선도적 연구와 적극적 대외 홍보를 통해, 주택담보대출 연구 분야의 표준 DB 위상을 구축하고 타 금융기관의 자발적 참여 독려

- 미국은 연방주택금융청(FHFA)과 소비자금융감독원(CFPB)이 공동으로 참여하여 주택담보대출 마이크로 DB인 NMDB(National Mortgage Database) 구축
- 유럽은 ECB 주도로 EU지역 은행들의 기업 대출 정보를 표준화하여 마이크로 DB인 AnaCredit(Analytical Credit Database) 구축

[그림 4-7] 주택금융정보시스템 해외사례



## DB 자료를 이용한 연체율 분석<sup>1)</sup>

- 이 장에서는 본격적인 DB구축의 사전 작업으로 DB 예시를 생성한 후 공사 주담대의 연체율에 영향을 미치는 요소들에 대한 분석을 수행함

### 1. 주택담보대출 분석의 특징

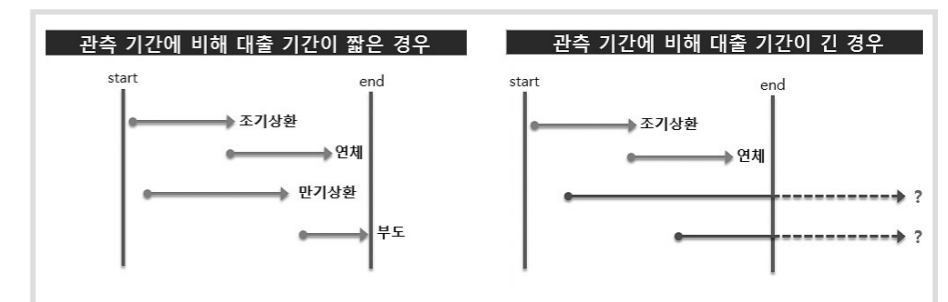
● 주택담보대출의 특징

- ▶ 대출기간이 통상 10~30년으로 주택(부동산)을 담보로 하는 장기 대출
- ▶ 신용대출에 비해 상환능력 및 담보능력이 우월하고 대손의 비율이 낮은 특징이 있음

● 주택담보대출 분석의 특징

- ▶ 장기 대출이기 때문에 많은 경우 대출 만기 이전에 관측이 종료되어 사건의 발생여부가 불확실한 자료가 발생하는데, 이를 '중도 절단된 자료 (censored data)'라 함
- ▶ 이러한 중도 절단된 자료의 경우, 신용위험 분석에서 일반적으로 사용되는 로짓(logit) 모형을 이용한 '로지스틱 회귀분석 (Logistic regression)' 등은 추정 편의(bias) 발생가능

[그림 4-8] 주택담보대출의 대출 종료 특징



1) 본 장의 분석은 국토연구원과의 공동연구로 진행되었음

● 주택담보대출의 연체율 분석에 일반적인 방법론을 적용하였을 때 생기는 문제점

- ▶ 일반적인 방법론은 중도 절단된 데이터에 대한 고려가 없기 때문에 실제로 연체가 발생하지 않은 차주와 연체가 발생하기 전 관측이 종료된 차주를 구분하지 못함
  - 상환능력이 우수하여 5년 동안 연체를 하지 않은 차주와, 상환능력은 떨어지나 대출 개시 후 6개월여 밖에 되지 않아 아직 연체를 하지 않은 차주를 동일하게 취급함
- ▶ 최근 급증하고 있는 주택담보대출 공급량 추이를 감안하면, 이러한 분석의 한계는 시의성 있는 분석을 저해하며 분석 결과를 활용한 주택담보대출 관련 의사결정 과정을 왜곡시킬 우려가 있음
  - [표 4-2] 는 각 연도에 공급된 대출들이 평균적으로 첫 연체를 경험하기까지 걸린 시간을 나타냄<sup>2)</sup>

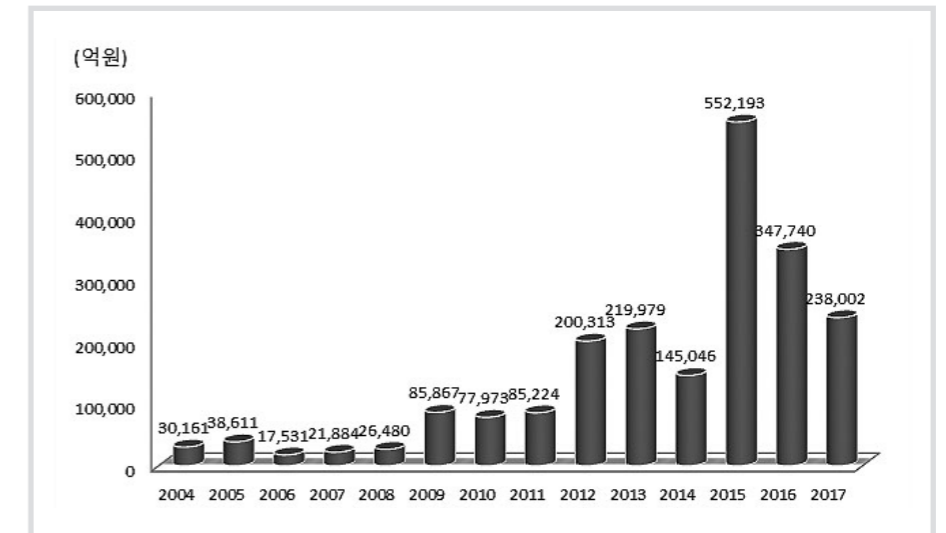
[표 4-2] 연체까지 걸린 평균시간

(단위:개월)	30일 연체	60일 연체	90일 연체
전체	21.87	24.79	26.55
2004	26.45	32.57	36.51
2005	23.51	26.11	27.03
2006	25.40	24.27	29.00
2007	29.84	35.41	39.57
2008	25.55	26.65	30.78
2009	29.93	32.00	36.33
2010	29.64	30.03	33.41
2011	25.15	28.29	28.36
2012	19.20	21.76	23.50
2013	20.08	22.35	21.73
2014	15.02	15.33	18.42
2015	13.16	14.35	15.91
2016	8.65	9.80	10.38
2017	6.00		

2) 분석 대상은 본 연구에서 사용한 표본이며, 표본추출과정은 본문 Ⅲ.3.1에 있음

- 2011년까지는 대략 25개월~30개월 사이의 값을 보이지만 이후부터는 평균값이 크게 떨어지는데, 이는 주택담보대출의 특성변화를 나타낼 수도 있지만 관측 종료로 인해 연체까지 긴 시간이 걸리는 차주들의 연체를 관측하지 못하여 생긴 오차의 영향이 존재함
- 예를 들어, 관측종료 시점이 2017년 1월 1일인 경우, 이후에 발생한 연체는 관측되지 못하여 2016년에 공급된 대출들의 연체까지 걸리는 기간은 12개월을 넘지 못하는 것으로 기록됨
- 2011년 이전의 연체까지 걸리는 평균 기간이 그대로 유지되었다고 가정할 때, 2015년 이후에 공급된 대출들 중 연체가 발생할 가능성이 높은 대출들도 관측종료로 인해 연체와 무관한 차주로 분류됨
- 공사의 MBS 공급이 최근 급증하고 있어([그림 4-9]) 분석 자료의 상당부분이 2015년 이후의 데이터임을 감안하면, 일반적으로 사용하는 로지스틱 회귀분석은 현실과 괴리된 결과를 발생시킬 위험이 있음

[그림 4-9] 연도별 MBS 발행실적



주 : 2017년은 8월까지의 실적

- ▶ 대출기간이 길어 관측기간 동안 연체상태에서 정상상환 상태로 돌아갔다가 다시 연체를 하는 경우가 존재하여 분석이 까다로움
  - 특정시점에 연체상태에 있는 차주들을 연체차주로 분류할 경우, 과거에 연체를 하였으나 현재는 정상상환 상태에 있는 차주들은 연체차주가 아닌 것으로 분류됨. 이 경우, 10년 동안 연체를 하지 않은 차주와 여러 차례 연체를 한 차주가 무차별한 것으로 간주될 위험
  - 위 문제를 해결하기 위해 특정 기간 내 연체를 경험한 차주들을 연체차주로 분류하더라도, 연체까지 걸리는 시간을 고려하기 힘들다는 문제가 존재함. 예를 들어, 대출을 받은 지 3개월 만에 연체를 한 차주와 3년 만에 연체를 한 차주의 위험도를 구분하지 못함

## 2. 연체율 분석의 목적 및 차별성

- (목적) 차주 신용도, 대출의 특성(위험도) 및 외부 변수가 연체율에 미치는 영향을 분석
- (특징) 중도 절단된 자료들의 비중이 높은 분석 대상의 특성을 고려하여 생존분석 방법론을 사용
  - ▶ 생존분석은 사건의 발생 여부뿐만 아니라 사건의 발생까지 걸리는 시간 또한 고려하는 분석방식
  - ▶ 대출 시작 1년 만에 관측이 종료된 대출의 경우, 연체와 무관한 대출로 간주하는 것이 아니라 '첫 1년 이내에는 연체를 하지 않았다'고 분류함으로써 관측 종료로 인해 발생할 수 있는 분석의 한계들을 극복
  - ▶ 첫 연체까지 걸리는 시간이 분석대상이므로 관측 시점에 따라 이전에 연체를 경험한 차주를 연체경험이 없는 차주와 구분하지 못할 수 있는 문제에서 자유로움
  - ▶ 연체까지 걸리는 시간을 고려함으로써 대출을 받은 지 3개월 만에 연체를 한 차주와 3년 만에 연체를 한 차주를 구분할 수 있음

- (차별성) 공사 DB 자료의 활용과 시계열 변수들의 포함으로 시간에 따라 연체율에 영향을 미치는 요소들의 변화를 반영
  - ▶ 조기상환으로 원금의 일부를 상환한 차주의 경우 연체위험이 줄어들었으나, 조기상환 정보가 없을 시 이러한 차주의 위험도 변화를 반영할 수 없음
    - 공사 DB의 조기상환 정보 활용으로 이러한 변화 반영
  - ▶ 차주의 담보 소재지 및 담보 형태와 감정원의 월별 가격지수를 결합함으로써 담보의 가치 변화를 정밀하게 반영 가능

## 3. 분석자료 및 기초통계

### 3-1) 분석대상 및 표본추출과정

- 2004년 6월부터 2017년 7월 사이에 공사가 유동화한 주택담보대출 중 차주 특성을 기준으로 10%를 샘플링하여 사용
  - ▶ 대상상품: t-보금자리론, t플러스 보급자리론, u-보금자리론, e-보금자리론, 적격대출, MBS-SWAP
  - ▶ 소수의 변동금리 대출들은 표본에서 제외하고 고정금리, 준고정금리 및 고정변동 상품만 분석
    - 고정금리형 상품과 변동금리형 상품의 연체율은 서로 다른 특징을 가질 것으로 추정되는데, 대부분이 고정금리형 상품인 분석자료 특성상 본 연구에서는 고정금리형 상품에 집중
  - ▶ (샘플링 방법) 공사가 2004년부터 2017년 7월까지 유동화한 주택 담보대출 개별 제출 자료를 연도별, 연령대별, 소득별, 지역별로 14\*9\*10\*16 개의 부분집합([표 4-3])으로 분류한 후, 비례층화랜덤추출법(비례층화표본추출법)을 사용하여 10%의 표본자료를 추출



[표 4-3] 표본추출 기준

기준	구간	설명
연도	14개	2004, ..., 2017
연령대	9개	10대, 20대, ..., 80대, 90대 이상
연소득	10개	1000만원 미만, 1000만원~2000만원, ..., 9000만원 이상
지역	16개	특별시, 광역시, 도 단위

- ▶ 데이터 전처리 과정 후 최종 169,302개 차주에 대한 정보사용
- ▶ 전체 모집단과 데이터 전처리 과정 전 표본의 기초통계량을 비교하면 전체적으로 큰 오차는 발생하지 않는 것으로 보임([표 4-4])
  - 실제 분석은 데이터 전처리 과정을 통해 이상치들을 제거한 표본을 사용하였으나, 추출한 표본에 존재하는 이상치 비율과 모집단에 존재하는 이상치 비율이 크게 다르지 않다고 본다면 데이터 전처리 과정 전 표본의 기초통계량과 모집단의 기초통계량을 비교하는 것이 적절
  - 샘플링에 고려되지 않은 변수들을 이용하여 모집단과 표본의 유사도를 검증

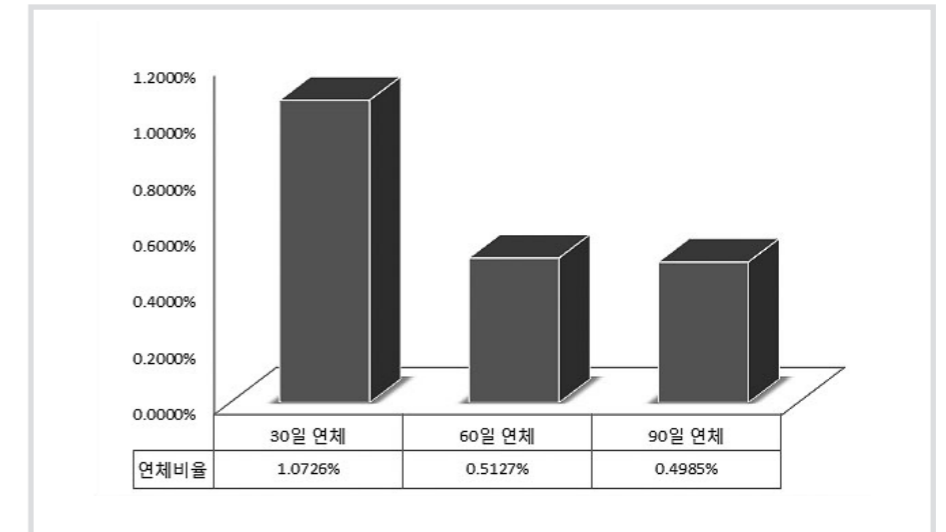
[표 4-4] 모집단과 표본의 특성 비교

	모집단	표본	차이
대출금액	98,320,884.58	98,533,333.25	-0.22%
LTV	57.57	57.56	0.03%
DTI	33.16	33.28	-0.36%
신용등급	2.90	2.90	0.07%
담보가치	283,379,901.91	283,964,382.88	-0.21%

### 3-2) 분석 대상의 기초통계량

- 이 절에서는 분석 자료의 기초통계를 정리하였음. 연체를 경험한 비율에 대하여는 [그림 4-10]에 나타내었고, 설명변수로 사용한 자료들은 자료의 특성에 따라 [표 4-5]와 [표 4-6]에 나누어 기초통계량을 나타냄
- ▶ 전체 169,302명의 차주 중 30일 기준 연체비율은 1.0726%(1,816명), 60일 기준으로는 0.5127%(868명), 90일 기준으로는 0.4985%(844명)가 연체를 경험함
  - 참고로 일반적인 연체율은 (현재 연체중인 금액)/(전체 대출 잔액) 으로 정의되어 연체를 해소한 차주는 포함하지 않고 계산됨
  - 본 연구에서는 한 번이라도 연체를 경험한 차주가 분석 대상이 되므로 이 값은 일반적인 연체율과는 다소 차이가 있을 수 있음

[그림 4-10] 연체비율



- ▶ 담보물로 설정된 주택가격은 평균 2억 2,454만원이고 대출금액은 평균 9,920만원, 분석기간 중 원금 대비 대출잔액의 비율은 평균 0.9인 것으로 나타남
- ▶ 차주의 평균 연령은 43.02세이며 대출 당시 차주의 소득대비 원리금 상환비율인 DTI는 평균 30.7%, 적용된 가산금리는 1.36%임

[표 4-5] 연속형 변수 기초통계량

기준	mean	std.dev	min	median	max
주택가격(만원)	22,454	12,985	1,500	19,750	90,000
대출금액(만원)	9,920	6,790	100	8,500	50,000
대출 잔액비율	0.90	0.16	0.00	0.96	1.00
연령	43.02	10.21	19	41	90
DTI	30.70	19.24	0.00	26.28	109.89

- ▶ 2004년부터 2017년 중 주택담보대출은 2015년에 54,529건(32.21%)으로 가장 많이 시행되었으며, 2016년 29,749건(17.57%), 2013년 20,124건(11.89%) 순임
- ▶ 대출의 목적은 구입이 82,673건(48.83%)으로 가장 많고 보전이 53,122건(31.38%), 상환은 33,163건(19.59%)임
- ▶ 대출금리는 고정금리가 144,286건(85.22%)으로 대부분을 차지했으며 준고정금리는 24,629건(14.55%), 고정변동이 387건(0.23%)임
- ▶ 차주의 신용등급은 1등급이 46,813명(27.65%), 2등급 35,301명(20.85%), 3등급 29,297명(17.3%) 순으로 신용등급이 낮을수록 비중이 줄어들어
- ▶ 주택유형은 아파트가 152,642건(90.16%)으로 압도적으로 높은 비율을 차지하고 다세대빌라는 11,695건(6.91%)임
- ▶ 전체 169,302개의 대출 중 105,799건(62.49%)은 원리금균등방식으로, 40,011건(23.63%)은 원금균등방식으로 상환이 이루어짐
- ▶ 연소득이 4천만원 이상인 차주는 58,509명(34.56%), 2천만원 미만은 56,294명(33.25%), 2천만원 이상 3천만원 미만은 27,993명(16.53%), 3천만원 이상 4천만원 미만은 26,506명(15.66%)으로 나타남
- ▶ 대출이 발생한 지역은 경기도가 53,981건(31.88%), 서울 26,117건(15.43%), 인천 14,857건(8.78%)으로 수도권지역이 전체의 56.09%를 차지함

[표 4-6] 범주형 변수 기초통계량

대출연도				
2004	2005	2006	2007	2008
3,733	5,123	1,654	3,394	4,109
2009	2010	2011	2012	2013
5,436	5,309	5,707	17,190	20,124
2014	2015	2016	2017	계
12,206	54,529	29,749	1,039	169,302
대출목적				
보전	구입	상환	결측	계
53,122	82,673	33,163	344	169,302
대출금리				
고정변동	고정	준고정	계	
387	144,286	24,629	169,302	
신용등급				
CB1	CB2	CB3	CB4	
46,813	35,301	29,297	21,153	
CB5	CB6	CB7	CB8	
16,272	11,264	6,140	1,826	
CB9	CB0			계
588	648			169,302
주택유형				
아파트	다세대빌라	단독주택	복합주택	
152,642	11,695	2,167	121	
상가주택		연립주택	계	
16		2,661	169,302	
대출상환방식				
원리금균등	원금균등	원금균등, 일시납	원리금균등, 일시납	
105,799	40,011	2,574	14,426	
일시납	체증식상환	체증식상환, 일시납	계	
1	2,671	3,820	169,302	
소득				
2천만원 미만	2천만원 이상 3천만원 미만	3천만원 이상 4천만원 미만	4천만원 이상	계
56,294	27,993	26,506	58,509	169,302

지역				
서울	강원	경기	경남	
100,541	3,089	53,981	11,691	
경북	광주	대구	대전	
4,671	6,103	7,689	4,900	
부산	울산	인천	전남	
12,742	4,662	14,857	2,671	
전북	제주	충남	충북	계
4,509	785	6,404	4,431	169,302
비수도권		수도권		계
74,347		94,955		169,302

## 4. 분석모형

### 4-1) 생존분석 방법론 개요

- 생존분석(Survival Analysis) (Cox, 1984)

- ▶ 어떠한 사건(event)이 발생하기까지 걸리는 시간에 대해 분석하여 생존함수(survival function) 또는 생존곡선(survival curve)을 추정하는 통계기법

- 새로운 치료방법이나 신약이 생존에 미치는 효과, 질병/기계 오작동 발생 등

- ▶ Deng, Quigley and Van Order(2000)에서는 다양한 방식으로 대출이 종료될 수 있는 주택담보대출의 특성을 고려하여 조기상환위험과 부도위험을 함께 보는 경쟁위험모형(Competing Risk Model) 제안

- ▶ 중도절단이 발생하는 대출기간이 긴 주택담보대출을 살펴보기에 적합

- 중도절단(Censoring)

- ▶ 사건의 발생여부에 대한 불확실한 자료를 ‘중도 절단된 자료’(censored data)라 함

- 연구 종료까지 사건이 발생하지 않거나, 연구 중간에 추적이 되지 않는 경우 등

- ▶ 중도절단에는 우측 중도절단(right censoring), 좌측 중도절단(left censoring), 구간 중도절단(interval censoring) 3가지 종류가 존재함

- 우측 중도절단 : 연구 종료 후에 사건이 발생할 수 있는 경우, 연구 중간에 추적이 되지 않는 경우
- 좌측 중도절단 : 이미 사건이 발생한 상태로 연구 시작한 경우
- 구간 중도절단 : 일정 주기로 관측을 해 사건 발생 시점을 정확히 알 수 없고 특정 기간 사이라고만 아는 경우

▶ 주택담보대출은 대부분 장기대출로, 관측 기간에 비해 대출 기간이 긴 특징이 있어 우측 중도절단이 있는 경우가 다수

- 생존함수(Survival Function)

▶  $t$ 는 시간을 나타내는 확률변수(random variable),  $T$ 는 사건이 발생한 시점

- $f(t)$  :  $t$ 의 확률밀도함수 (the probability density function for  $T$ )

- $F(t)$  :  $t$ 의 누적분포함수 (the cumulative distribution function for  $T$ )  
즉,  $F(t) = Prob(T \leq t)$

▶ 생존함수  $S(t)$ 는  $t$ 시점 이후에 생존할 확률을 의미

$$S(t) = Prob(T > t) = 1 - F(t)$$

- Hazard Function(위험함수)

▶ 위험함수  $h(t)$ 는  $t$ 시점까지는 살아있는데  $t$ 시점 직후 사망할 확률을 의미

$$h(t) = \frac{f(t)}{S(t)} = \lim_{\Delta t \rightarrow 0} \frac{Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

▶ 누적위험함수(cumulative hazard function)는  $t$ 시점 내에 사망할 확률을 의미하며 다음과 같이 정의됨

$$H(t) = \int_0^t h(u) du$$

● 생존함수와 누적위험함수의 관계

$$\begin{aligned} \triangleright H(t) &= \int_0^t h(u)du = \int_0^t \frac{f(u)}{1-F(u)} du = [-\ln(1-F(u))]_0^t \\ &= -\ln\{S(t)\} \end{aligned}$$

$$\triangleright S(t) = \exp[-H(t)]$$

$$\triangleright F(t) = 1 - \exp[-H(t)]$$

$$\triangleright f(t) = h(t)\exp[-H(t)]$$

● Cox의 비례위험회귀모형(Proportional Hazard Model)

▶ 생존분석 방법 중 가장 많이 쓰이는 모형

▶ 모든 설명변수가 0일 때 사건이 발생할 확률을 기본적 위험  $h_0(t)$  (baseline hazard)라 하며, 이를 이용하여 위험함수  $h(t)$ 를 다음과 같이 표현함

$$h(t) = h_0(t) \times \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})$$

▶ 기본적 위험에 대한 가정을 필요로 하지 않는 준모수적 (semi-parametric) 방법

- 모형의 특성상 추정 과정에서 기본적 위험이 약분되어 사라지기 때문에 이에 대한 추정이 불필요

▶ 이자율, 주택가격 변화 등의 시계열 자료 적용 가능

▶  $\exp(\beta_j)$ 를 위험비(Hazard Ratio, HR)라 하고, 1을 기준으로 해석을 달리 함

-  $HR = 1$  ( $\beta_j = 0$ ): j번째 설명변수는 사건의 발생에 영향을 미치지 않음

-  $HR > 1$  ( $\beta_j > 0$ ): j번째 설명변수는 사건의 발생 위험을 증가시킴

-  $HR < 1$  ( $\beta_j < 0$ ): j번째 설명변수는 사건의 발생 위험을 감소시킴

$$-\frac{h_0(t)\exp\left(\sum_{k=1}^j \beta_k x_k + \beta_j(x_j + 1)\right)}{h_0(t)\exp\left(\sum_{k=1}^{j-1} \beta_k x_k + \beta_j x_j\right)} = \exp(\beta_j) \text{이므로 위험비는 } j\text{번째 변수를}$$

1만큼 증가시켰을 때 위험함수가 증가하는 비율을 나타냄.

즉,  $\exp(\beta_j) = 1.04$  라는 것은 j번째 변수를 1만큼 증가시켰을 때 위험함수는 4%만큼 증가한다는 것을 의미함

▶ 위험비는 상수 값이므로, 시간 t에 관계없이 사건이 발생할 위험률은 일정함

4-2) 분석 방법론

● (분석모형) 생존분석 Survival Analysis 방법론 중 일반적으로 가장 많이 사용되는 Cox의 비례위험회귀모형모형을 적용

▶ 30일 연체, 60일 연체, 90일 연체로 나누어 각 차주의 연체 확률에 영향을 미치는 요인들을 분석

- 생존 분석 특성상, 최초 연체 후 정상적으로 상환하다가 다시 연체를 할 경우, 두 번째 연체는 분석 대상이 아니게 되는데, 이러한 특징은 분석방식의 한계

● (분석과정) 연체율에 가장 큰 영향을 미칠 것으로 예상되는 기본적인 변수들로 기본모형을 구축한 후, 분석 목적에 따라 변수 추가

▶ 범주형 변수(categorical variable)가 많은 자료의 특성상, 모든 변수들을 한 번에 포함시킬 경우 추정해야 할 변수가 많아 추정의 정확도가 떨어질 위험이 있음

- 신용등급, 연령, 소득은 양적 변수(quantitative variable)이긴 하지만, 각 값들의 간격이 연체율에 일정한 영향을 미친다고 보기 어려워 더미 변수로 사용

▶ (기본 모형) 가장 기본적인 형태로, 차주의 신용도, 대출의 위험도, 그리고 외부요인으로 연체율을 설명

- 차주의 신용도: 신용등급

- 대출의 위험도: DTI, 대출원금, 담보가치, 잔여원금비율

- 외부요인: 주택가격변화, 실업률, 물가지수

▶ (신용등급의 적정성 모형) 신용등급이 차주의 신용도를 충분히 반영하고 있는지 확인하기 위하여 기본 모형에 소득수준과 연령대에 대한 변수 추가

▶ (기타 특성 모형) 기본 모형에 기타 연체율에 영향을 미칠 것으로 예상되는 요인들(대출 목적, 상환방식, 수도권 여부, 가산금리 수준) 추가

▶ (지역별 차이 모형) 지역별 차이를 세분화하여 살펴보기 위하여 기본모형에 광역시·도 단위의 지역 변수 추가

[표 4-7] 변수 정의

변수명	설명
차주의 신용도 관련 변수	CB1 ~CB9 차주의 신용등급 더미변수(1~9등급). 금융거래 부재로 신용 등급을 매길 수 없는 0등급이 base case.
	연령1 ~연령3 차주의 연령대에 대한 더미변수. 30세 미만을 base case로 하여 30-40, 40-50, 50+로 구분.
	소득1 ~소득3 차주의 소득구간에 대한 더미변수. 2천만원 미만을 base case로 하여 2천만원-3천만원, 3천만원-4천만원, 4천만원+로 구분.
대출 특성 관련 변수	잔여원금 비율 잔여 원금/대출 원금*100. 대출 후 1개월 단위로 존재하는 시계열 변수. (단위: %)
	대출원금 대출원금. (단위: 1,000만원)
	담보가치 대출시점 담보의 평가가치. (단위: 1,000만원)
	DTI 대출시점 차주의 전체 DTI. (공사 추정치)
	구입목적, 상환목적 구입목적 관련 더미변수. 보존 목적이 base case.
	상환_ 상환방식 관련 더미변수. 원리금균등이 base case로 하여 원금균등, 원금균등.일시납, 원리금균등.일시납, 일시납, 체증식상환, 체증식상환.일시납으로 구분.
외부 요인 관련 변수	주택가격 변화 차주의 담보 주택 형태에 따라 단독주택, 아파트, 연립다세대 월별 가격 지수 사용, 담보의 소재지에 따른 지역별 지수 사용. 각 지수 들은 CPI를 이용하여 실질 변수로 변환 후 로그 차분하여 대출 시점 대비 변화량 형태로 사용. (단위: %, 출처: 갑정원)
	실업률 실업률. CPI를 이용하여 실질 변수로 변환 후 로그 차분하여 대출 시점 대비 변화량 형태로 사용. (단위: %, 출처: 한국은행)
	물가지수 물가지수. (단위: %, 출처: 한국은행)
	수도권 담보의 소재지 관련 더미 변수. 비수도권을 base case로 하여 수도권(서울, 경기, 인천)을 구분.
	강원, 경기 ... 담보의 소재지 관련 더미 변수. 서울을 base case로 하여 각 지역을 나타냄
	가산금리 차주의 대출 금리 - 대출 당시 국고채 3년 금리. (단위: %, 국고채 자료 출처: 한국은행)

## 5. 분석결과

- 대출원금 대비 담보가치, DTI, 신용등급, 그리고 잔여원금 비율 모두 이론적으로 예측되는 방향으로 유의한 영향력을 나타냄
  - ▶ 대출원금, 잔여원금 비율, DTI, 그리고 신용등급은 높을수록 연체가능성은 증가하며 담보가치는 클수록 연체위험이 감소함
- 대출 시점 대비 물가상승은 연체가능성을 증가시키지만 담보가격의 변화나 실업률의 변화는 연체율에 영향을 미치지 못함
  - ▶ 물가가 상승하면 가계의 기본적인 생활비가 높아져 대출 상환에 사용할 수 있는 가용자원이 감소하므로 연체위험이 높아질 수 있음
  - ▶ 일반적으로 담보가치의 상승은 부의 효과(wealth effect)와 향후 추가적인 주택가격 상승에 대한 기대를 불러 일으켜 연체율을 낮추는 부(-)의 관계를 가짐
  - ▶ 하지만 분석 대상기간 동안 한국의 주택가격은 꾸준히 올랐으므로, 차주 입장에서는 주택 구입 시점 대비 일시적인 가격 하락이 있었다더라도 향후 주택가격 상승에 대한 예상이 흔들리지 않았다면 오히려 가격 회복까지 버티겠다고 결정할 수 있음
  - ▶ 이러한 경우 담보가치의 변화와 연체율 간의 관계는 중립적이거나 심지어 정(+)의 관계를 나타낼 수도 있으므로, 앞선 부(-)의 관계를 나타내는 표본들과 뒤섞여 유의하지 않은 것처럼 보일 가능성이 있음. 이러한 비대칭적인 영향에 대하여는 향후 추가적인 연구가 필요

[표 4-8] 기본모형 추정 결과

변수명	연체 30		연체 60		연체 90	
	coef	exp(coef)	coef	exp(coef)	coef	exp(coef)
CB1	-2.7638	0.0631 ***	-2.7350	0.0649 ***	-2.0602	0.1274 ***
CB2	-1.9280	0.1454 ***	-2.1993	0.1109 ***	-1.8023	0.1649 ***
CB3	-1.4985	0.2235 ***	-1.5108	0.2207 ***	-1.0730	0.3420 **
CB4	-0.9222	0.3976 ***	-0.9049	0.4046 **	-0.4104	0.6634
CB5	-0.5762	0.5620 **	-0.5527	0.5754	-0.1761	0.8385
CB6	0.0137	1.0138	-0.0593	0.9424	0.3569	1.4289
CB7	0.5695	1.7673 **	0.6099	1.8402	0.9617	2.6163 *
CB8	0.8806	2.4124 ***	1.0053	2.7328 **	1.4518	4.2707 ***
CB9	1.1323	3.1027 ***	1.1309	3.0984 **	1.5585	4.7518 ***
DTI	0.0066	1.0066 ***	0.0060	1.0060 ***	0.0051	1.0051 **
대출원금	0.0923	1.0966 ***	0.1378	1.1478 ***	0.1440	1.1549 ***
담보가치	-0.0485	0.9527 ***	-0.0715	0.9310 ***	-0.0797	0.9234 ***
잔여원금 비율	0.0206	1.0208 ***	0.0194	1.0196 ***	0.0269	1.0272 ***
주택가격 변화	-0.0012	0.9988	-0.0020	0.9980	-0.0038	0.9962
실업률	0.0044	1.0044	0.0005	1.0005	0.0025	1.0025
물가지수	0.0477	1.0488 ***	0.0440	1.0450 **	0.0734	1.0761 ***
log-lik	-19,242.84		-8,999.953		-8,707.342	
obs	4,638,985		4,664,977		4,667,941	

주: \*\*\*, \*\*, \*는 각각 0.1%, 1%, 5% 유의수준을 의미함

● 신용등급은 연체율에 영향을 미칠 수 있는 차주의 개인적 특성을 충분히 잘 반영하고 있는 것으로 보임

- ▶ 60일 연체 및 90일 연체의 경우, 차주의 연령, 소득수준, 성별 모두 신용등급에 반영되어 있는 것 이상의 정보를 주지 못함
- ▶ 소득과 연령의 경우 30일 연체에 유의한 영향을 미치는데, 이는 신용등급이 소득을 고려한 장기적인 상환능력은 적절히 예측하지만 일시적인 유동성 부족으로 인한 30일 연체를 예측하는 데는 한계를 보이는 것일 가능성이 있음

▶ 금융거래 부재로 신용등급 산출이 되지 않는 차주의 경우 6등급과 유사한 연체율을 보이는 것으로 추정

[표 4-9] 신용등급의 적정성 모형 추정 결과

변수명	연체 30		연체 60		연체 90	
	coef	exp(coef)	coef	exp(coef)	coef	exp(coef)
CB1	-2.7378	0.0647 ***	-2.7097	0.0666 ***	-2.0511	0.1286 ***
CB2	-1.9050	0.1488 ***	-2.1728	0.1139 ***	-1.7864	0.1676 ***
CB3	-1.4763	0.2285 ***	-1.4882	0.2258 ***	-1.0555	0.3480 **
CB4	-0.9033	0.4052 ***	-0.8853	0.4126 **	-0.3933	0.6749
CB5	-0.5553	0.5739 *	-0.5345	0.5860	-0.1609	0.8514
CB6	0.0371	1.0378	-0.0406	0.9602	0.3713	1.4495
CB7	0.5884	1.8010 **	0.6272	1.8724 *	0.9782	2.6598 *
CB8	0.8989	2.4570 ***	1.0217	2.7779 **	1.4689	4.3446 ***
CB9	1.1432	3.1368 ***	1.1554	3.1754 ***	1.5839	4.8740 ***
연령1	-0.1557	0.8558	-0.2233	0.7999	-0.1287	0.8792
연령2	-0.1511	0.8597	-0.1387	0.8704	-0.0243	0.9760
연령3	-0.3861	0.6797 ***	-0.2463	0.7817	-0.1307	0.8775
소득1	-0.1078	0.8979	0.0555	1.0571	0.0290	1.0294
소득2	-0.2055	0.8143 **	-0.1342	0.8744	-0.2286	0.7957
소득3	-0.1994	0.8192 **	-0.0199	0.9803	0.1126	1.1192
DTI	0.0050	1.0050 ***	0.0056	1.0056 **	0.0055	1.0055 **
대출원금	0.0933	1.0978 ***	0.1375	1.1474 ***	0.1424	1.1530 ***
담보가치	-0.0445	0.9565 ***	-0.0701	0.9323 ***	-0.0804	0.9228 ***
잔여원금 비율	0.0205	1.0207 ***	0.0194	1.0196 ***	0.0269	1.0272 ***
주택가격 변화	-0.0009	0.9991	-0.0020	0.9980	-0.0039	0.9961
실업률	0.0043	1.0043	0.0005	1.0005	0.0024	1.0024
물가지수	0.0511	1.0524 ***	0.0442	1.0452 *	0.0717	1.0743 ***
log-lik	-19,229.33		-8,996.116		-8,701.203	
obs	4,635,921		4,664,977		4,667,941	

주: \*\*\*, \*\*, \*는 각각 0.1%, 1%, 5% 유의수준을 의미함

- 상환 목적으로 받은 대출의 경우 구입 목적 혹은 보존 목적의 대출보다 뚜렷하게 낮은 연체율을 보임
  - ▶ 다른 모든 조건이 같을 때 30일, 60일, 90일 연체 모두에서 상환 목적의 대출은 연체 가능성이 기타 대출 대비 절반 수준으로 분석되는데, 이는 상환을 통해 이전에 비해 이자 부담이 낮은 대출로 옮겨갔기 때문으로 보임
- 상환 방식은 30일 연체 및 60일 연체에는 영향을 미치지만 90일 연체에는 영향을 주지 못하는 것으로 보임
  - ▶ 체증식 상환방식을 택하였거나 일부 일시납 상환의 조건이 있는 대출들의 연체확률이 상대적으로 높은 것으로 드러남
    - 일부 일시납 상환의 조건은 차주의 일상적인 상환부담을 경감시키는 조건이므로, 이러한 조건이 연체에 영향을 미쳤다가보다 연체 위험이 높은 차주들이 이러한 조건을 선택했을 가능성이 있어 추가적인 분석 필요
  - ▶ 일시납 상환 대출의 경우 표본 중 1건밖에 되지 않아 결과에서 제외
- 수도권과 비수도권에 공급된 대출은 뚜렷한 차이를 나타낸다고 보기 어려워 지역별로 세분화하여 분석할 필요

[표 4-10] 기타 특성 모형 추정 결과

변수명	연체 30			연체 60			연체 90		
	coef	exp(coef)		coef	exp(coef)		coef	exp(coef)	
CB1	-2.6520	0.0705	***	-2.6050	0.0739	***	-1.9256	0.1458	***
CB2	-1.8120	0.1634	***	-2.1030	0.1221	***	-1.6616	0.1898	***
CB3	-1.3930	0.2483	***	-1.4000	0.2465	***	-0.9992	0.3682	*
CB4	-0.8549	0.4253	***	-0.8320	0.4352	**	-0.4275	0.6521	
CB5	-0.5549	0.5741	*	-0.5429	0.5811		-0.1971	0.8211	
CB6	-0.0094	0.9906		-0.1155	0.8909		0.2862	1.3314	
CB7	0.5262	1.6920	*	0.5615	1.7530		0.8971	2.4525	*
CB8	0.8308	2.2950	***	0.9226	2.5160	**	1.3551	3.8770	***
CB9	1.0540	2.8680	***	1.0630	2.8960	**	1.4739	4.3662	***
DTI	0.0075	1.0080	***	0.0067	1.0070	***	0.0059	1.0060	**
대출원금	0.0577	1.0590	***	0.1113	1.1180	***	0.1261	1.1344	***
담보가치	-0.0305	0.9699	***	-0.0584	0.9433	***	-0.0649	0.9371	***
잔여원금 비율	0.0136	1.0140	***	0.0137	1.0140	***	0.0214	1.0216	***
구입목적	0.0104	1.0100		-0.0700	0.9324		-0.1829	0.8329	*
상환목적	-0.7592	0.4680	***	-0.4591	0.6318	*	-0.4651	0.6280	*
상환_원금 균등	-0.0574	0.9442		-0.0817	0.9216		-0.0883	0.9154	
상환_원금 균등,일시납	0.3721	1.4510	*	0.2217	1.2480		-0.1974	0.8209	
상환_원리금 균등,일시납	0.5002	1.6490	***	0.0948	1.0990		0.1146	1.1214	
상환_일시납	-4.2650	0.0141		-4.5480	0.0106		-3.8056	0.0222	
상환_체증식 상환	0.5899	1.8040	***	0.6046	1.8300	**	0.2142	1.2389	
상환_체증식 상환,일시납	0.8898	2.4350	***	0.6492	1.9140	***	0.2375	1.2681	
가산금리	0.2723	1.3130	***	0.4204	1.5230	***	0.4025	1.4955	***
주택가격 변화	0.0007	1.0010		-0.0007	0.9993		-0.0027	0.9973	
실업률	0.0054	1.0050	*	-0.0015	0.9985		-0.0019	0.9981	
물가지수	0.0688	1.0710	***	0.0500	1.0510	**	0.0772	1.0802	***
수도권	0.0102	1.0100		0.1829	1.2010	*	-0.0149	0.9852	
log-lik	-17,820.16			-8,274.981			-7,812.734		
obs	4,133,100			4,158,437			4,161,434		

주: \*\*\*, \*\*, \*는 각각 0.1%, 1%, 5% 유의수준을 의미함

- 지역별로 세분화하여 연체율을 분석한 경우, 수도권 내 경기도와 인천의 연체확률이 서울보다 높은 것으로 나타남

- ▶ 경기도와 인천은 30일, 60일, 90일 연체 모두에서 서울보다 높은 연체 확률을 나타냄
- ▶ 기타 지역에서는 일관된 경향성을 발견하기 어려워, 수도권과 비수도권의 차이보다 수도권 내 연체율의 차이가 유의미한 것으로 보임

[표 4-11] 지역별 차이 모형 추정 결과

변수명	연체 30		연체 60		연체 90				
	coef	exp(coef)	coef	exp(coef)	coef	exp(coef)			
CB1	-2.7389	0.0646	***	-2.7030	0.0670	***	-2.0523	0.1284	***
CB2	-1.8987	0.1498	***	-2.1627	0.1150	***	-1.7942	0.1663	***
CB3	-1.4719	0.2295	***	-1.4770	0.2283	***	-1.0640	0.3451	**
CB4	-0.8959	0.4082	***	-0.8747	0.4170	**	-0.4053	0.6668	
CB5	-0.5536	0.5749	*	-0.5275	0.5901	.	-0.1703	0.8434	
CB6	0.0306	1.0311		-0.0372	0.9635		0.3631	1.4377	
CB7	0.5895	1.8031	**	0.6338	1.8848	*	0.9652	2.6253	*
CB8	0.8985	2.4560	***	1.0197	2.7724	**	1.4478	4.2535	***
CB9	1.1525	3.1661	***	1.1620	3.1964	***	1.5643	4.7795	***
DTI	0.0066	1.0066	***	0.0059	1.0059	***	0.0055	1.0055	**
대출원금	0.0902	1.0944	***	0.1352	1.1447	***	0.1407	1.1511	***
담보가치	-0.0489	0.9523	***	-0.0722	0.9304	***	-0.0750	0.9277	***
잔여원금 비율	0.0198	1.0200	***	0.0184	1.0186	***	0.0265	1.0269	***
주택가격 변화	-0.0013	0.9987		-0.0020	0.9980		-0.0037	0.9963	
실업률	0.0044	1.0044		0.0007	1.0007		0.0026	1.0026	
물가지수	0.0478	1.0489	***	0.0426	1.0436	*	0.0679	1.0703	***
강원	0.1764	1.1929		0.2079	1.2310		0.2713	1.3117	
경기	0.2074	1.2304	*	0.3002	1.3501	*	0.3647	1.4400	**
경남	0.0223	1.0225		0.0723	1.0750		0.4559	1.5776	*

변수명	연체 30		연체 60		연체 90				
	coef	exp(coef)	coef	exp(coef)	coef	exp(coef)			
경북	-0.0038	0.9962		0.1598	1.1733		0.5647	1.7589	*
광주	0.1049	1.1106		0.0785	1.0817		0.2671	1.3062	
대구	-0.3784	0.6850	*	-0.2461	0.7818		-0.0870	0.9167	
대전	-0.3817	0.6827	*	-0.7404	0.4769	*	-0.1382	0.8710	
부산	0.2456	1.2784	*	0.0601	1.0620		0.1250	1.1332	
울산	-0.2819	0.7544		-0.2985	0.7419		0.2546	1.2900	
인천	0.2208	1.2471	*	0.3435	1.4099	*	0.4825	1.6201	**
전남	0.1907	1.2101		0.2428	1.2748		0.6165	1.8523	*
전북	-0.3556	0.7008		-0.1692	0.8443		0.4146	1.5138	
제주	0.6650	1.9445	*	0.7353	2.0862		0.4443	1.5593	
충남	-0.0700	0.9324		-0.0925	0.9116		0.1675	1.1823	
충북	0.0787	1.0818		-0.0465	0.9546		0.4608	1.5853	*
log-lik	-19,214.73		-8,980.821		-8,694.298				
obs	4,638,985		4,664,977		4,667,941				

주: \*\*\*, \*\*, \*는 각각 0.1%, 1%, 5% 유의수준을 의미함



# IV

## 활용방안: 주택금융정보시스템 구축

### 1. 배경 및 목적

- 가계부채 리스크 확대
  - ▶ 최근 가계부채 증가속도에 대하여 우려가 커지는 상황에서 가계부채의 50%를 초과하는 주택담보대출이 빠른 속도로 증가함
  - ▶ 가계부채 총량에 대한 리스크 분석 및 차주별, 금융업권별 등 종합적 관점에서의 리스크 분석 필요
- 기존 공사 주택담보대출 DB 한계
  - ▶ 인구사회구조 및 주택시장의 변화에 따라 주택담보대출에 대한 정교한 분석이 요구되나, 기 구축된 자료로 다변화하는 주택금융수요와 위험 요인 변화를 판단하기에 역부족
  - ▶ 공사는 공사취급상품의 특성정보를 정기적으로 분석하고 있으나, 대상이 한정적인 장기고정금리대출로, 전체 주택담보대출에서 차지하는 비율이 10%~15% 수준에 불과해 일반화 할 수 없는 한계
  - ▶ 그간 공사는 주택시장의 변동성 확대에 대비, 주택 및 주택금융시장 분석지표 개발, 리스크관리 방안 등 데이터 분석에 기반한 연구와 지표개발을 시도하였으나, 제한된 시장 한계에 따른 대표성 문제 지적
- Micro DB 구축 필요성 부각
  - ▶ 가계부채의 체계적 관리를 위하여 2015년부터 한국은행, 금융감독원, 금융위원회는 각각 가계부채 미시자료(micro-data)를 별도의 방법으로 구축
  - ▶ 한국은행은 NICE 신용정보에 구축된 DB를 활용하며, 금융위원회는 한국신용정보원을 통하여 전 금융권 신용정보 집중

▶ 미국 및 EU 등 주요 선진국에서는 2008년 금융위기 이후 금융시장 안정성 모니터링과 정책판단 기초자료 활용 목적으로 마이크로 DB를 보완 및 신규로 구축하는 추세

### 2. 주택금융정보시스템 구축

- 주택담보대출 데이터와 차주 가계신용 데이터가 결합된 광범위한 패널 데이터 구축
  - ▶ 주택금융시장의 안정성 제고를 위해 주택금융 증장기 대책 마련 및 맞춤형 정책 추진 지원
  - ▶ 대출특성, 주택특성 및 차주별 특성이 결합된 마이크로데이터 구축
  - ▶ 주택시장과 금융시장을 연계한 입체적 분석을 통한 금융정책 지표 발굴, 시장 모니터링 기능 강화 및 정책판단의 근거로 활용
  - ▶ 주택금융시장에 대한 체계적 모니터링 수단 마련을 통한 통합 리스크 관리 체계 구축
- 공사 연구용 주택담보대출 DB를 중심으로 외부 관련기관 자료와 정보 연계를 통해 주택금융정보시스템을 구축하고, 빅데이터 기반 주택금융 시장 모니터링과 정책지원 기능 강화
  - ▶ 주택시장과 금융환경의 연계분석을 통하여 다양한 정보를 분석하고, 이를 통하여 정부, 금융기관, 금융소비자 간의 정보비대칭을 해소하여 주택금융시장의 투명성 및 안정성 제고

### 3. 주택금융정보시스템 활용 및 연구 예시

- 서민주거 안정 목적 정책금융 효과 분석
  - ▶ 공적모기지 이용 차주에 대해 대출실행 이후 과정에 대한 조사를 통하여 정책금융의 효과 실증 분석
    - 대출 이후 추가대출(고정·변동) 여부, 카드소비 및 소득 추이 분석
    - 대출 이후 2차 주택 구입 또는 이동 여부 파악(실거래가 매칭)
    - 정책모기지 공급이 주거 및 소득 수준 개선에 미친 직·간접적 영향 분석
- 정책모기지 선택 요소 분석
  - ▶ 효율적 공적모기지 공급전략 수립 제고
  - ▶ 고정금리 공사모기지 대출자와 유사한 특성을 가진 표본 집단에 대해 어떤 요인이 공적·일반모기지 선택에 주요한지 분석
- 금융정책지표 개발
  - ▶ 담보목적물 가격정보, 차주 소득정보 등을 연계하여 주기적으로 CLTV(Current Lone to Value), CDTI(Current Debts to Income) 산정
    - 담보능력의 적정성, 채무상환능력의 적정성 여부 판단
    - 정보 축적 후 연체율, 부도율 등과의 연계 정보를 활용하여 적정 LTV, 적정 DTI 등 보다 객관적인 정책 판단기준 산정 가능
  - ▶ 담보목적물의 가치변동 측정을 위한 주택가격지수 개발
    - 일정시점마다 업데이트되는 주택 가격정보를 이용하여 담보목적물의 가치변동 측정을 목적으로 한 주택가격지수를 개발
    - 주택시장의 변화와 변동성 예측을 통하여 주택시장 및 주택금융시장의 조기 경보지표로 활용

## V 결론 및 시사점

- 데이터 분석 기술이 발달함에 따라 기존 방법으로는 어렵거나 불가능 하던 영역을 개척하여 새로운 서비스를 제공할 수 있게 되었으며 이에 데이터 분석의 잠재력과 가능성이 재조명 되고 있음
  - ▶ 본 연구는 이러한 추세를 공사 주택담보대출 자료에 적용해 보고자 하는 장기 프로젝트의 일환으로 시작되었음
- 공사 주담대 DB는 자료의 안전한 관리와 효율적인 업무를 위해 구축되었으나 연구목적에 있어 한계 존재
  - ▶ 시스템과 전산 요청을 통한 간접 접근 방식으로 운영되어 제한된 자료 접근 한계가 존재하여 관련 연구를 일회성으로 그치게 함
  - ▶ 지속적인 상시 관찰이 용이하지 않아 숨은 트렌드나 의미를 파악하는 연구 수행에 어려움이 있음
- 연구용 주택담보대출DB 설계와 구축을 통해 주담대 관련 연구 적시성 및 시의성을 제고
  - ▶ 설계된 Layout을 기반으로 주택담보대출의 연체율에 미치는 요소에 대한 분석 실시
- 생존분석 방법론을 이용하여 일반적인 주택담보대출 분석에서 발생할 수 있는 문제점들을 일부 극복하였으며, 주택금융공사 DB자료와 시계열 자료의 활용으로 연체율에 영향을 미치는 변수들을 좀 더 다양하고 정밀하게 분석함
- 2004년 6월부터 2017년 7월 사이에 공급된 주택담보대출 중 169,302건의 표본에 대하여 약 460만개의 시계열 관측치를 사용함

● 주요 결과는 다음과 같음

- ▶ LTV, DTI, 그리고 잔여원금 비율은 낮을수록, 신용등급은 높을수록 연체율이 낮아짐
- ▶ 대출 시점 대비 물가상승은 연체가능성을 증가시키지만 담보가격의 변화나 실업률의 변화는 영향을 미치지 못하는 것으로 나타남
- ▶ 상환 목적으로 받은 대출의 경우 구입 목적 혹은 보존 목적의 대출보다 뚜렷하게 낮은 연체율을 보임
- ▶ 수도권과 비수도권의 차이보다 수도권 내의 연체율 차이가 유의미한 것으로 분석됨

● 본 연구에서 수행된 작업을 기초로 향후 주택담보대출 데이터와 차주가계신용 데이터가 결합된 광범위한 패널 데이터 구축 예정

- ▶ 공사 연구용 주택담보대출 DB를 중심으로 외부 관련기관 자료와 정보연계를 통해 주택금융정보시스템 구축
- ▶ 정책지표 발굴 등 후속 연구수행을 통해 주택금융시장 모니터링과 지원 기능 강화

참고문헌

1. 김유정·문영기, "주택가격변동과 주택담보대출 연체율의 동태적 분석", 한국주거환경학회, 제9권 제2호, 2011, pp. 43-57.
2. 김정렬, "가계부채 규모와 연체율 결정요인에 관한 연구", 한일경상학회, 제66호, 2014, pp. 25-36.
3. 심종원·정의철·정현정, "주택담보대출 연체율 결정 요인에 관한 연구", 부동산학 연구, 제15집 제2호, 2009, pp. 81-96.
4. 정호성, "차주별 패널자료를 이용한 주택담보대출의 연체요인에 대한 연구", BOK 경제연구 제2017-6, 한국은행
5. Cox, David Roxbee, and David Oakes, "Analysis of survival data.", Vol. 21. CRC Press, 1984.
6. Demyanyk, Yuliya, and Otto Van Hemert, "Understanding the subprime mortgage crisis.", The Review of Financial Studies 24.6: 1848-1880, 2009.
7. Deng, Yongheng, et al. "Mortgage default and low downpayment loans: the costs of public subsidy.", Regional science and urban economics 26.3: 263-285, 1996.
8. Gerardi, Kristopher, et al. "Can't pay or won't pay? Unemployment, negative equity, and strategic default.", No. w21630. National Bureau of Economic Research, 2015.
9. IMF, "Article IV Consultation—Press Release; Staff Report; And Statement By the Executive Director For The Republic Of Korea.", IMF Country Report, No. 16/276, 2016.
10. Kim, H., Cho, H., & Ryu, D., "Characteristics of Mortgage Terminations: an Analysis of a Loan-Level Dataset.", The Journal of Real Estate Finance and Economics, 1-30, 2017.
11. Kim, Yong Sun, and Kwang Kyu Lim, "A Study on Mortgage Loan Borrowing Behavior of Korean Households: A Micro Perspective.", 2013.
12. Mian, Atif, and Amir Sufi. "The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis.", The Quarterly Journal of Economics 124.4: 1449-1496, 2009.
13. Tian, Chao Yue, Roberto G. Quercia, and Sarah Riley. "Unemployment as an adverse trigger event for mortgage default.", The Journal of Real Estate Finance and Economics 52.1: 28-49, 2016.