

검색어 통계 분석을 활용한 중도상환 관심도 측정

- 최근 빅데이터의 다양한 활용 사례 등장
- 검색어 통계 분석은 빅데이터를 활용하는 방법 가운데 직관적이고 사용이 용이하며, 관심도를 정량화할 수 있다는 장점이 있음
- 검색어 통계 분석을 통한 중도상환 관심도 측정치가 실제 중도상환과 관련이 있는지 검토하고, 실무 활용 가능성을 진단하고자 함

1 검토 배경

□ 최근 빅데이터의 다양한 활용 방안 주목

- 사용자에게 의해 누적된 데이터가 증가하고 데이터 처리 속도가 비약적으로 상승함에 따라 이를 활용하여 업무에 적용하는 사례 등장.
- 검색어 통계 서비스는 빅데이터 분석 가운데 직관적이고 사용이 용이하여 여러 분야에서 사용됨.
 - 구글은 감기에 대한 검색어를 분석하여 독감을 예측하는 시스템을 구축하였는데, 실제 독감 확산과 유사할 뿐만 아니라 오히려 2주가량 빠르게 예측.

□ 중도상환 관심도 측정

- 중도상환에 대한 관심도 측정을 통해 실제 중도상환이 설명될 수 있다면, 향후 중도상환율 예측 등의 분야에서 활용 가능.
- 본 보고서에서는 검색어 분석을 통해 중도상환에 대한 관심이 실제 중도상환을 설명하는지 확인하고, 검색어 분석의 실제 업무 활용 가능성을 진단해보고자 함.

-
- 작성자: 김형준 연구위원 (051-663-8178 / hkim@hf.go.kr)
 - 본고의 내용은 필자의 개인 의견으로 한국주택금융공사의 공식적인 견해와 다를 수 있습니다.
-

2 검색어 분석 서비스의 내용 및 종류

□ 검색어 분석 서비스란?

- 최근 포털사이트에서는 서비스 기간 동안 축적된 검색어 빅데이터를 바탕으로 검색어 분석 서비스를 제공.
- 특정 검색어에 대한 사람들의 ‘관심도’를 측정하기에 용이.
- 일반적으로 특정 검색어의 검색 횟수를 지수 형태로 표시.
 - 최대 검색 횟수를 가진 기간을 100으로 하여 나머지 기간의 검색 횟수 표시.
- 서비스를 제공하는 곳에 따라 보다 세분화된 분석 가능.

□ 구글 트렌드 (google.com/trends)

- 구글에서는 2012년 9월 Google Insights for Search 를 구글 트렌드로 통합, 이를 통해 현재까지 검색어 분석 서비스를 제공.
- 구글 검색에 축적된 데이터를 바탕으로 해당 검색 키워드의 검색 횟수를 지수 형태로 제공.
 - 주간 통계에서 의미있는 통계량이 나오지 않는 경우 월간 통계량 제공.
- 보다 세분화된 자료 검색 가능.
 - 검색 기간(2004년 이후), 지역, 카테고리, 검색경로 별로 세분화된 자료 검색 가능.
 - 관련 검색어 제공.

□ 네이버 트렌드 (trend.naver.com)

- 네이버는 2014년 12월부터 검색어 통계 서비스를 베타 형식으로 제공.
- 현재 네이버의 국내 검색점유율은 70% 이상이며, 해당 검색어에 대한 국내 관심도를 측정하기에 적합함.
- 검색 기간, 경로 별로 세분화된 자료 제공.
 - PC는 2007년 1월부터 현재까지, 모바일은 2010년 7월부터 현재까지 제공.
 - 검색 경로는 PC/모바일로 구분하여 조회 가능.

3 중도상환 관심도 측정을 위한 검색어 분석 활용

□ 안심전환대출 관심도 측정

- 지난 3월, 안심전환대출 출시 이후 공사 보증자리론의 중도상환이 급격하게 증가함. 그러나 모형 1과 같은 기존 설명 변수만으로는 이에 대해 충분히 설명하지 못함. (그림 1)

- 모형 1 (기존 설명 변수):

$$SMM_{i,t} = \beta_0 + \beta_1 X_{1,i,t} + \beta_2 X_{2,i,t} + \beta_3 X_{3,i,t} + \beta_4 X_{4,t} + \epsilon_{i,t}$$

$SMM_{i,t}$: i 풀의 t 기 중도상환율 (single monthly mortality)

$X_{1,i,t}$: i 풀의 t 기 평균 경과기간

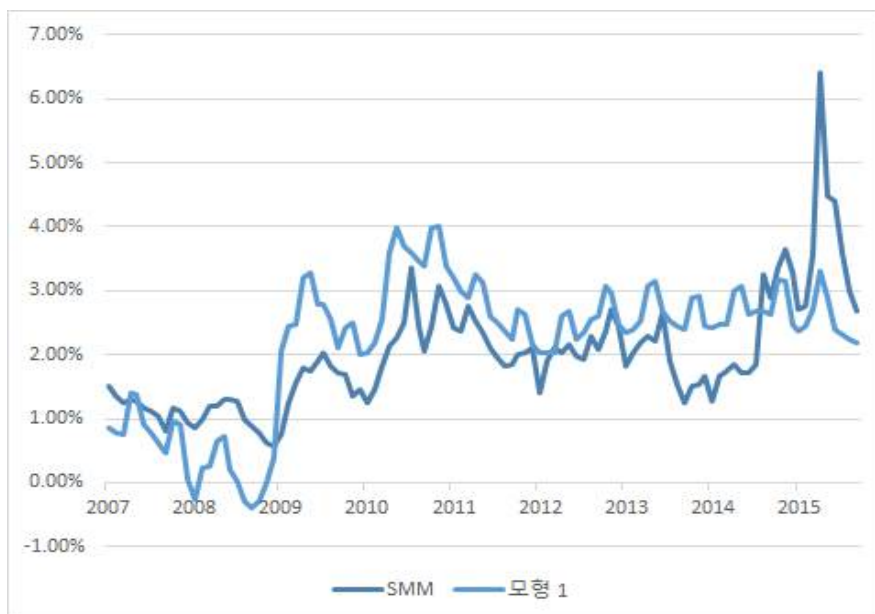
$X_{2,i,t}$: i 풀의 t 기 금리 스프레드 (= 계약금리 - 현재금리)

$X_{3,i,t}$: i 풀의 t 기 주택가격지수 변화율

$X_{4,t}$: t 기의 이사철 더미 (4, 5, 10, 11월 = 1, 나머지 = 0)

- 회귀분석 결과는 <표 1> 참조

[그림 1] 실제 조기상환과 모형 1의 설명



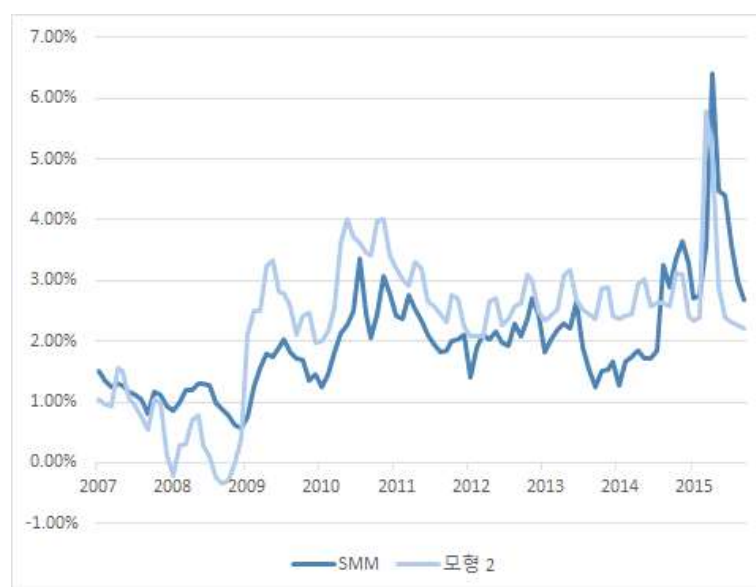
- 안심전환대출 출시에 따른 영향을 고려하기 위해 구글 트렌드의 검색
횟수 지표를 관심도에 대한 추정치로 사용하여, 모형 2와 같이 분석.
 - 보증자리론은 안심전환대출 대상이 아니므로 금리 차이만으로는 중도상환이 급증하는 현상을 설명하기에 부족함.
 - 안심전환대출 출시 및 홍보로 인한 중도상환에 대한 관심도 증가가 보증자리론의 중도상환으로 이어진 것으로 추정됨.
 - 일반적으로 정책 시행기간동안 더미변수를 할당하는 방법을 사용하지만, 더미변수는 시간에 따라 변화하는 관심도를 나타내기 어려움. 이를 보완하기 위하여 구글 트렌드의 검색 통계를 관심도에 대한 추정치로 사용.
- 모형 2 (안심전환대출 관심도 추가):

$$SMM_{i,t} = \beta_0 + \beta_1 X_{1,i,t} + \beta_2 X_{2,i,t} + \beta_3 X_{3,i,t} + \beta_4 X_{4,t} + \beta_5 X_{5,t} + \epsilon_{i,t}$$

$X_{5,t}$: t 기의 안심전환대출 관심도

 - 안심전환대출의 관심도는 구글 트렌드에서 “안심전환대출” 검색 빈도를 사용.
- 회귀분석 결과 모형 2는 기존 모형보다 중도상환에 대한 설명력이 높은 것으로 나타남.
 - <표 1>에서 모형 2의 Adjusted R-square는 0.4532 로 모형 1의 0.4244 보다 높음.

[그림 2] 실제 조기상환과 모형 2의 설명



□ 중도상환 관심도 측정

- 중도상환에 대한 관심도가 중도상환에 유의한 영향을 미치는지 확인하기 위하여 네이버 트렌드에서 나타나는 중도상환 관련 검색어의 검색 횟수를 지수화한 자료를 사용하여 모형 3과 같이 분석.
 - 중도상환 관련 검색어에는 ‘중도상환’, ‘조기상환’, ‘주택담보대출 갚아타기’를 선정. 일부 관련 검색어는 연관성 없는 항목을 함께 나열하므로 제외
 - 네이버 트렌드의 검색어 지수는 주간으로 제공되기 때문에 이를 월간 자료로 변환하는 작업이 필요. 본 보고서에서는 주간 검색 지수를 7로 나누어 이를 일간 검색 지수라 가정하고, 일간 검색 지수를 월간으로 합산하고 월별 날짜수를 보정하는 작업을 통해 월간 자료로 변환하여 사용.
 - 이를 그대로 사용하는 경우 변동성이 커 전반적인 관심도의 크기를 나타낸다고 보기 어려움. 6개월간의 이동평균을 사용하는 것이 가장 적합한 것으로 나타남.

- 모형 3 (중도상환 관심도 추가):

$$SMM_{i,t} = \beta_0 + \beta_1 X_{1,i,t} + \beta_2 X_{2,i,t} + \beta_3 X_{3,i,t} + \beta_4 X_{4,t} + \beta_5 X_{5,t} + (\beta_6 X_{6,t} + \beta_7 X_{7,t} + \beta_8 X_{8,t}) + \epsilon_{i,t}$$

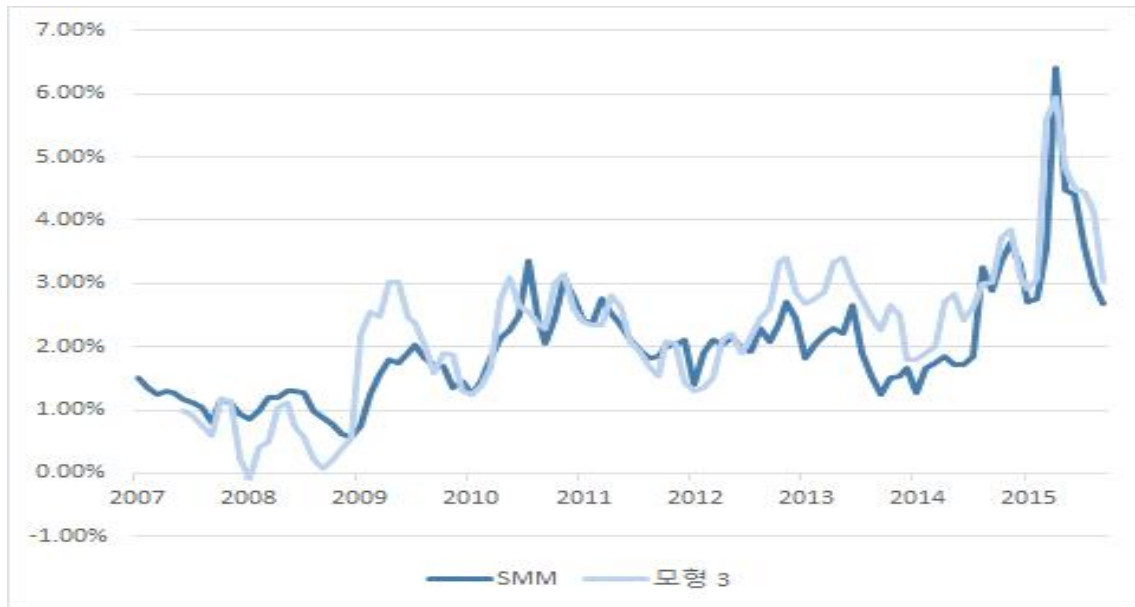
$X_{6,t}$: t 기의 네이버 트렌드 “중도상환” 지표 6개월 이동평균

$X_{7,t}$: t 기의 네이버 트렌드 “조기상환” 지표 6개월 이동평균

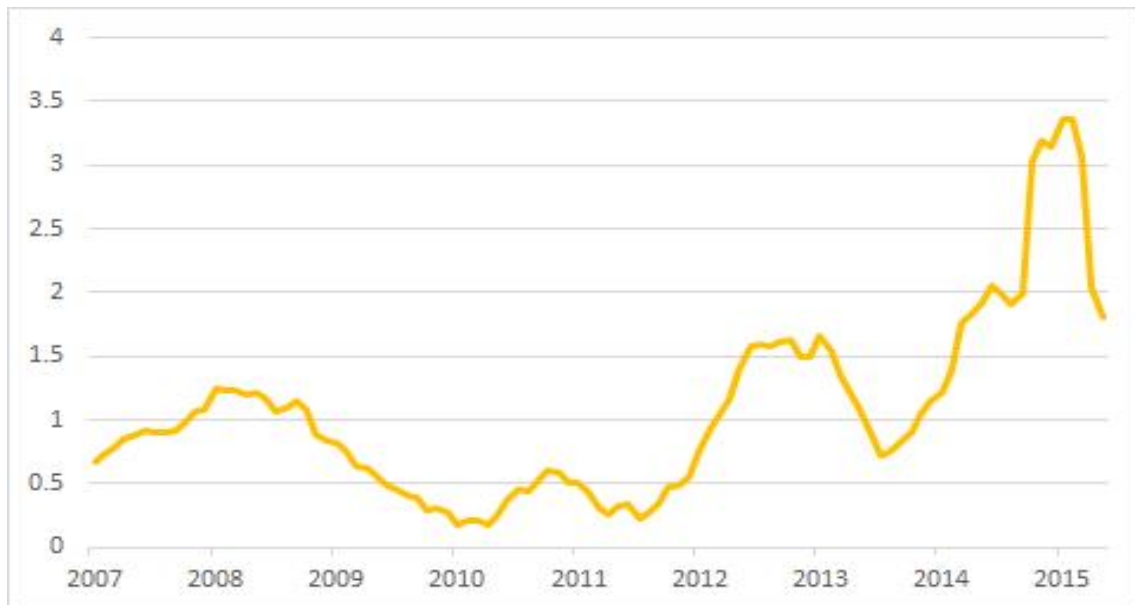
$X_{8,t}$: t 기의 네이버 트렌드 “주택담보대출 갚아타기” 지표 6개월 이동평균

- 분석결과 모형 3은 모형 1, 2 보다 중도상환에 대해 더욱 설명력이 높음.
 - <표 1>에서 모형 3의 Adjusted R-square 는 0.5257로 가장 높음.
 - 네이버 트렌드 “주택담보대출 갚아타기” 지표는 중도상환에 가장 크게 영향을 미침. “중도상환” 지표는 음의 값을 나타내는데, 이는 해당 검색어가 “중도상환 수수료” 등의 관련 검색어를 가짐으로써 오히려 중도상환을 감소시키는 영향을 미치는 것으로 해석됨. (2007년 이후 검색어 비중은 “주택담보대출 갚아타기” 64%, “중도상환” 29%, “조기상환” 7% 순으로 나타남)
- 낮은 금리, 주택가격상승 등으로 인한 영향이 이미 중도상환 관심도에 반영되어있을 수 있으나, 모든 변수가 유의하며 모형의 설명력이 높음.
 - 대출 연령을 제외한 나머지 변수의 부호와 유의성은 동일하게 유지됨

[그림 3] 실제 조기상환과 모형 3의 설명



[그림 4] 중도상환 관심도 변화*



※ 해당 지표는 <표 1> 모형 3의 결과를 바탕으로 중도상환, 조기상환, 주택담보대출 갈아타기 검색어 통계 지표를 선형 결합한 것임.

[표 1] 모형 1, 2, 3 의 회귀분석 결과

	모형 1		모형 2		모형 3	
	Coeff.	Std.	Coeff.	Std.	Coeff.	Std.
beta0	1.2035	0.0396	1.2898	0.0406	0.7198	0.1757
beta1	0.0280	0.0033	0.0173	0.0034		
beta2	1.3523	0.0461	1.4100	0.0468	1.3488	0.0259
beta3	-8.8343	0.7907	-6.4633	0.8088	-0.7010	0.2453
beta4	0.5147	0.0459	0.5131	0.0463	0.5548	0.0436
beta5			7.5181	0.3416	2.9226	0.3504
beta6					-1.1113	0.1913
beta7					1.6027	0.1835
beta8					8.1105	0.2839
R-sq	0.4247		0.4535		0.5261	
Adj. R-sq	0.4244		0.4532		0.5257	

※ 모든 추정치는 1% 수준에서 유의함

4 시사점

- 검색어 통계가 실제 중도상환에 영향을 미치는 것을 확인
 - 일반적인 거시경제지표 발표 주기가 분기/연도별임을 감안할 때, 매주 갱신되는 검색어 분석 지표 활용은 시장에 대한 보다 빠른 이해와 선제적 대응을 가능하게 하는 도구로서 가치가 있음.
- 업무분야에서의 다양한 활용 가능성 내재
 - 주택담보대출 및 유동화 분야
 - 중도상환 관심도 측정을 통한 중도상환 예측
 - 경제적 불안 요소에 대한 측정을 통한 연체/부도율 예측
 - 주택연금분야
 - 주택연금에 대한 직접적 관심도 측정을 통한 단기적 수요 예측
 - 고령화, 노인 복지 등에 관한 사회적 관심도 모니터링
- 향후 실질적인 업무 활용을 위해서는 보다 깊은 연구가 필요.
 - 검색기간 제한, 지역별 비교 불가, 검색어 선정의 자의성 등 검색어 통계가 지닌 한계점에 대한 이해가 필요