

본 자료의 내용은 한국은행의 공식견해가 아니라  
집필자 개인의 견해라는 점을 밝힙니다. 따라서  
본 자료의 내용을 보도하거나 인용할 경우에는  
집필자명을 반드시 명시하여 주시기 바랍니다.

# 금융자산시장 쏠림현상이 금융안정에 미치는 영향 : 이론적·실증적 분석 프레임워크

국민대학교 경제학과

김현학

# 금융자산시장 쏠림현상이 금융안정에 미치는 영향: 이론적·실증적 분석 프레임워크

김 현 학

## 요약

본 연구는 한국 금융시장의 고질적 취약 요인으로 지목되어 온 '쏠림현상'이 금융안정(FSI)에 미치는 동태적 영향을 이론적/실증적 프레임워크를 통해 분석하였다. 쏠림현상은 투자자의 사적 정보를 무시한 군집행동과 금융기관의 자금 쏠림으로 구분되며, 이는 정보 폭포 및 평판 관리와 비대칭적 보상구조로 인한 대리인 문제와 같은 미시적 동인에 의해 촉발되어 시스템 리스크를 직접적으로 증폭시키는 도미노 효과를 야기한다. 본 연구는 Chang et al. (2000)의 CSAD 방법론을 적용한 군집행동 지표(HERD)와 함께, BERT 기반 텍스트 감성분석 및 검색어 급등도를 결합하여 군집행동심리지수(HSI)를 구축하였으며, 확장된 VAR 모형과 기계학습 모형(랜덤포레스트, LSTM)을 통해 그 선행성을 검증하였다. 실증 분석 결과, 정형데이터에 기반한 군집심리의 충격은 금융안정에 미치는 영향이 제한적으로 나타났지만, 텍스트 자료 기반으로 작성한 HSI는 시장의 '공포 지수'(VIX)와 강한 역(-)의 상관관계를 보이는 '탐욕 지수'로서, 주요 주가지수 급락 이벤트에 뚜렷하게 선행하는 조기경보 신호 역할을 수행했다. 특히 위기 예측을 위한 기계학습 모형에서는 HSI가 전통적인 신용스프레드와 더불어 중요한 예측 변수로 식별되었는데, 이는 금융안정 예측에서 시장 심리의 동태적 패턴 학습에 HSI가 상당한 영향력이 있음을 시사한다. 반면, 통화정책(기준금리) 충격은 군집심리나 신용쏠림 구조에 유의한 영향을 미치지 못하는 한계를 보였다. 따라서 본 연구는 HSI를 활용한 선제적 '쏠림 경보체계' 구축, 통화정책과 거시건전성 정책(LTV/DSR 등)의 유기적 연계, 그리고 금융기관의 클로백(clawback) 도입 등 보상구조 개선을 통한 미시적 행동 편향 완화가 쏠림 리스크를 효과적으로 관리하는 데 필수적임을 정책적으로 제시한다.

---

\* 교수, 국민대학교 경제학과, hyunhak.kim@kookmin.ac.kr

# I. 서론

## 1. 한국 금융시장에서의 '쏟림현상' 정의 및 맥락

금융시장의 안정성은 국가 경제의 지속 가능한 성장을 위한 핵심 전제조건이다. 그러나 금융시장은 내재적으로 변동성을 가지며, 때로는 특정 자산이나 부문으로 자금이 급격하게 집중되는 현상으로 인해 시스템 전체의 안정성이 위협받기도 한다. 한국 금융시장에서는 이러한 현상을 '쏟림현상'이라 지칭하며, 이는 금융 시스템의 고질적인 취약 요인으로 빈번하게 거론되어 왔다(강경훈, 2006). 본 보고서는 한국 금융자산시장에서 관찰되는 쏟림현상이 금융안정에 미치는 영향을 분석하기 위한 포괄적인 이론적·실증적 프레임워크를 제시하는 것을 목적으로 한다.

쏟림현상은 단일한 개념이 아니며, 그 발생 주체와 시장에 따라 구분하여 이해할 필요가 있다. 본 연구에서는 쏟림현상을 다음과 같이 두 가지 유형으로 구분하여 정의한다(강경훈, 2006; Bikhchandani & Sharma, 2000).

1. **투자자 수준의 군집행동(Investor-level Herd Behavior):** 이는 자본시장에서 개인 또는 기관 투자자들이 자신의 사적 정보나 분석을 무시하고, 다른 시장 참여자들의 투자 결정을 모방하는 경향을 의미한다(Banerjee, 1992; Bikhchandani, Hirshleifer, & Welch, 1992; Hirshleifer & Teoh, 2003). 이러한 행동은 상호 간의 명시적인 조정 없이도 유사한 투자 패턴을 만들어내며, 주식, 펀드 등 특정 금융상품에 대한 수요를 급격히 증가시키거나 감소시킨다.
2. **금융기관 수준의 자금 쏟림(Institution-level Concentration of Funds):** 이는 은행, 신용카드사 등 금융기관들이 특정 산업 부문, 금융상품, 또는 소위 '테마(theme)'로 경쟁적으로 자금을 공급하는 현상을 지칭한다(강경훈, 2006; 지동현, 2010). 이는 개별 자산에 대한 투자를 넘어 특정 사업 전략을 집단적으로 추종하는 행태로, 한국 금융계에서 '패거리 금융문화'라는 비판을 받을 정도로 뚜렷하게 나타나는 특징이다(강경훈, 2006).

이러한 쏟림현상의 파괴적인 잠재력은 한국의 여러 금융위기 사례를 통해 명확히 드러났다. 이러한 사례들은 쏟림현상이 단순히 특정 자산 시장에 국한되지 않고, 금

용 시스템 전반에 걸쳐 구조적인 취약성을 드러내는 반복적인 패턴임을 보여준다. 표 1은 1999년 이후 우리나라의 주요 자산 쓸림현상의 사례를 정리한 것이다.

표 1 쓸림현상의 역사적 사례

| 사례/테마       | 기간        | 주체        | 주요 동인                                  | 핵심 지표 (정점)   | 주요 결과  |
|-------------|-----------|-----------|--|--|--|
| 바이코리아 펀드 열풍 | 1999-2000 | 개인/기관 투자자 | IT 버블, 주식 시장에 대한 낙관적 전망                | 주식형 펀드 수탁고 급증  | IT 버블 붕괴 후 펀드 환매 사태, 주가 급락 (강경훈, 2006)                   |
| 신용카드 사태     | 2001-2003 | 신용카드사, 은행 | 정부의 카드 사용 장려 정책, 카드사 간 과도한 외형 경쟁       | 신용카드사 연체율 (14.1%, 2003년) (강경훈, 2006)                 | 대규모 신용불량자 발생, 카드사 부실화, 시스템 유동성 위기 (강경훈, 2006; 지동현, 2010) |
| 중소기업 대출 쓸림  | 2002-2003 | 은행        | 대기업 대출 수요 감소, 가계부실 우려에 따른 대체 자금 운용처 모색 | 중소기업 대출 연평균 28% 급증 (강경훈, 2006)                       | 부실 심사, 연체율 상승 후 급격한 대출 회수(역쓸림) (강경훈, 2006; 지동현, 2010)    |
| 주택담보대출 쓸림   | 2005-2006 | 은행        | 저금리 기조, 부동산 가격 상승 기대, 은행 간 금리 인하 경쟁    | 주택담보대출 잔액 2배 증가 (2002년 88조 → 2004년 169조) (강경훈, 2006) | 가계부채 급증, 금융감독당국의 LTV/DTI 규제 강화 초래 (강경훈, 2006)            |

## 2. 문헌 연구: 군집행동의 미시적 기초

쓸림현상의 원인을 이해하기 위해서는 그 미시적 동인을 설명하는 다양한 이론적 접근법을 검토할 필요가 있다. 본 절에서는 정보 기반, 평판 및 보상 기반, 그리고 행동주의적 관점에서 군집행동의 발생 메커니즘을 분석한 주요 선행 연구들을 고찰한다.

### 2.1. 정보 기반 군집행동과 정보 폭포 (Information-Based Herding and Informational Cascades)

군집행동에 대한 초기 이론 연구는 합리적인 경제 주체들이 불완전한 정보 하에서 어떻게 타인의 행동을 모방하게 되는지에 초점을 맞췄다. Banerjee (1992)와 Bikhchandani, Hirshleifer, and Welch (1992)의 연구는 '정보 폭포(informational cascade)'라는 개념을 통해 이를 설명한다. 이들의 모델에 따르면, 개인들은 순차적으로 의사결정을 내리면서 앞선 사람들의 행동을 관찰한다. 특정 시점이 되면, 앞선 사람들의 행동에서 추론된 정보가 자신이 가진 사적 정보보다 더 유용하다고 판단하여 자신의 정보를 무시하고 대중의 행동을 따르는 것이 합리적인 선택이 된다 (Banerjee, 1992; Bikhchandani et al., 1992; Welch, 1992). 이렇게 형성된 정보 폭포는 소수의 초기 정보에 의해 전체 집단의 행동이 결정되게 만들어, 집단적 의사결정이 매우 취약하고 새로운 정보에 의해 쉽게 역전될 수 있는 특성을 갖게 한다 (Bikhchandani et al., 1992; Welch, 1992).

금융시장에서의 군집행동은 가격 메커니즘의 존재로 인해 더욱 복잡한 양상을 띤다. Avery and Zemsky (1998)는 가격이 정보를 효율적으로 반영할 경우 단순한 정보 폭포는 발생하기 어렵다고 주장했다. 그러나 이들은 '사건 불확실성(event uncertainty)', 즉 정보 이벤트의 발생 여부 자체가 불확실한 다차원적 불확실성 하에서는 군집행동이 발생할 수 있음을 보였다. Cipriani and Guarino (2014)는 이러한 이론을 바탕으로 실제 거래 데이터를 통해 군집행동을 추정할 수 있는 구조적 모형을 개발하여, 이론과 실증 분석의 간극을 좁혔다(Cipriani & Guarino, 2014). 이들의 연구는 금융시장에서의 군집행동이 비합리성의 발현이 아니라, 불완전한 정보 환경 하에서 이루어지는 합리적 추론의 결과일 수 있음을 시사한다.

## 2.2. 평판 및 보상 기반 군집행동 (Agency Problems)

정보적 요인 외에, 대리인 문제(agency problems) 역시 군집행동의 중요한 동인으로 작용한다. 특히 금융투자업 종사자들의 의사결정은 평판과 보상구조에 큰 영향을 받는다.

Scharfstein and Stein (1990)은 투자 매니저들이 자신의 능력에 대한 시장의 평가, 즉 평판에 대한 우려 때문에 군집행동을 보일 수 있다고 주장했다. 만약 독자적인 판단으로 투자에 실패할 경우, 이는 매니저의 능력 부족으로 평가되어 경력에 치명적인 타격을 줄 수 있다. 반면, 다수를 따라 투자했다가 실패할 경우 "모두가 틀렸다"는 변명이 가능해 실패에 대한 비난을 분산시킬 수 있다. 이러한 '숫자 속의 안전(safety in numbers)' 추구는 개인적으로는 합리적일 수 있으나, 사회적으로는 비효율적인 자원 배분과 상호 연계된 위험(correlated risk-taking)을 초래한다(Scharfstein & Stein, 1990; Rajan, 2006).

비대칭적 보상구조는 이러한 경향을 더욱 심화시킨다. 금융투자업 종사자들의 보수는 단기적인 성공에 대해서는 막대한 보너스를 지급하지만, 실패에 대해서는 책임이 제한적인 '내가 이기면 내 돈, 네가 지면 네 돈(Heads I win, Tails you lose)' 구조를 갖는 경우가 많다(강경훈 & 여은정, 2010). 이러한 구조는 과도한 위험 추구와 함께, 단기적으로 높은 수익을 낼 것으로 기대되는 인기 자산으로의 쏠림을 강력하게 유인한다. 2008년 글로벌 금융위기(GFC) 이전, 수많은 금융기관들이 복잡한 구조의 주택저당증권(MBS)과 부채담보부증권(CDO)에 집중적으로 투자했던 현상은 이러한 보상구조가 낳은 대표적인 쏠림 사례로 지적된다(Rajan, 2006; 강경훈 & 여은정, 2010). 이에 대한 규제적 대응으로 미국의 도드-프랭크법(Dodd-Frank Act)은 경영진의 성과급을 재무제표 재작성 시 환수할 수 있도록 하는 '클로백(clawback)' 조항을 도입했다(Acharya & Richardson, 2012).

## 2.3. 행동주의적 및 심리적 동인 (Behavioral and Psychological Drivers)

합리적 선택 이론의 틀을 넘어 행동경제학과 사회심리학은 군집행동의 비합리적 측면을 조명한다. 투자자들은 과신(overconfidence), 가용성 휴리스틱(availability heuristic) 등 인지 편향에 의해 체계적인 판단 오류를 범할 수 있다. 또한, 시장 전

반의 투자심리(investor sentiment)나 공포와 탐욕 같은 감정적 요인들이 확산되면서, 투자자들이 펀더멘털 분석을 무시하고 군중의 흐름에 휩쓸리게 될 수 있다(Barberis et al., 1998; Shiller, 2015).

특히 '불확실성'은 모든 이론적 관점에서 군집행동을 촉발하는 핵심적인 촉매제 역할을 한다. 금융위기나 팬데믹과 같이 불확실성이 극도로 높아지는 시기에는, 개인들이 자신의 분석에 대한 확신을 잃고 타인의 행동에서 단서를 찾으려는 경향이 강해진다. 이러한 심리적 기제는 군집행동을 증폭시키고 시장의 변동성을 키우는 악순환을 낳는다(Schmitt & Westerhoff, 2017; Shiller, 1990).

#### **2.4. 거시 금융 연계: 군집행동에서 시스템 리스크로**

미시적 수준에서의 군집행동은 거시 금융 안정성에 중대한 영향을 미친다. 군집행동은 본질적으로 투자 전략의 다양성을 감소시켜, 한 방향으로의 대규모 주문 불균형을 야기한다. 이는 직접적으로 자산 가격의 변동성을 증폭시키고, 가격을 펀더멘털 가치로부터 괴리시켜 자산 버블을 형성하는 주요 원인이 된다(Lux, 1995; Shiller, 2015). 다수 경제 주체들의 상호 연관된 행동은 시장을 불안정하게 만들고 금융 시스템의 취약성을 높인다(Cont & Bouchaud, 2000; Ionescu, 2012).

특히 금융기관 수준의 자금 쓸림이 발생할 경우, 시스템 리스크는 더욱 직접적인 위협이 된다. 예를 들어, 모든 은행이 부동산 담보대출에 집중적으로 자금을 공급하면, 금융 시스템 전체가 부동산 시장이라는 단일 위험 요인에 과도하게 노출된다. 이 상황에서 부동산 시장에 부정적인 충격이 발생하면, 대출 부실이 동시다발적으로 발생하여 개별 은행의 문제를 넘어 금융 시스템 전체의 연쇄적인 부실로 이어지는 '도미노 효과'를 촉발할 수 있다. 이는 2008년 글로벌 금융위기의 핵심 전개 과정이었으며, 서브프라임 모기지 시장의 부실이 MBS와 CDO라는 매개를 통해 전 세계 금융 시스템으로 확산된 바 있다(Rajan, 2006). 따라서 쓸림현상은 단순한 시장 이상 현상이 아니라, 시스템 리스크를 내재적으로 증폭시키는 핵심 요인으로 간주되어야 한다(Ionescu, 2012; Ooi, 2025).

## II. 쓸림현상의 실증적 측정

이론적 논의를 바탕으로, 본 장에서는 쓸림현상을 계량적으로 측정하는 방법론들을 검토하고, 국내외 시장에서 축적된 실증적 증거들을 종합한다. 이는 III장에서 제시될 분석 프레임워크의 방법론적 타당성을 뒷받침하는 근거가 된다.

### 1. 군집행동의 탐지 및 계량화 방법론

#### 1.1. 수익률 분산도 모형 (Return Dispersion Models)

수익률 분산도 모형은 자산 수익률이 시장 평균 수익률로부터 얼마나 흩어져 있는지를 측정하여 군집행동을 탐지한다. 합리적 자산가격결정모형에 따르면, 시장이 큰 폭으로 움직일 때 개별 자산들은 시장 요인에 대한 민감도(베타)가 각기 다르기 때문에 수익률의 분산도가 증가해야 한다. 반면, 군집행동은 투자자들이 개별적 신념을 억제하고 시장의 컨센서스를 따르는 현상이므로, 이 기간 동안 수익률 분산도는 오히려 감소하거나 비례적으로 덜 증가하는 경향을 보인다(Christie & Huang, 1995; Chang, Cheng, & Khorana, 2000).

- CSSD 모형 (Cross-Sectional Standard Deviation): Christie and Huang (1995)이 제안한 선구적인 방법으로, 개별 주식 수익률의 횡단면 표준편차(CSSD)를 계산하고, 이를 시장이 급등락한 날을 나타내는 더미 변수에 회귀분석한다. 이 더미 변수의 계수가 통계적으로 유의한 음(-)의 값을 가질 경우 군집행동의 증거로 해석한다(Christie & Huang, 1995).
- CSAD 모형 (Cross-Sectional Absolute Deviation): Chang, Cheng, and Khorana (2000)가 제안한 CSAD 모형은 이상치(outlier)에 덜 민감한 절대 편차를 사용하며, 현재 가장 널리 사용되는 방법론이다. 이 모형의 핵심은 수익률 분산도와 시장 수익률 간의 '비선형적' 관계를 검증하는 데 있다. 회귀식에 시장 수익률의 제곱항( $R_{m,t}^2$ )을 포함시켜 다음과 같이 분석한다:

$$CSAD_t = \gamma_0 + \gamma_1 |R_{m,t}| + \gamma_2 R_{m,t}^2 \varepsilon_t \quad (1)$$

여기서  $\gamma_2$  계수가 통계적으로 유의한 음(-)의 값을 가질 경우( $\gamma_2 < 0$ ), 이는 시장의 움직임이 클 때 분산도의 증가세가 둔화되거나 오히려 감소함을 의미하며, 군집 행동의 강력한 증거로 간주된다(Chang et al., 2000; Chiang & Zheng, 2010).

## 1.2. 최근 연구

- 구조적 모형 (Structural Models): Cipriani and Guarino (2014)의 연구와 같이, 단순한 통계적 검증을 넘어 군집행동 이론 모형의 내재적 파라미터들(예: 정보 이벤트 발생 확률, 사적 정보의 정확성 등)을 고빈도 거래 데이터를 이용해 직접 추정하는 접근법이다. 이 방법은 진정한 의미의 정보 기반 군집행동과, 단순히 공통된 정보에 반응하여 나타나는 외견상의 군집(spurious clustering)을 구분할 수 있게 해준다(Cipriani & Guarino, 2014).

Cipriani와 Guarino(2014)는 정보 기반 군집행동을 설명하기 위해 '정보 이벤트 불확실성(information-event uncertainty)'이라는 개념을 도입했다. 구조적 접근법은 단순히 거래가 물리는 현상을 넘어, 정보 비대칭 환경에서 합리적 행위자들이 왜 자신의 정보를 무시하고 타인을 따르는지를 이론적 틀 안에서 설명하고 그 확률을 정량적으로 측정할 수 있게 해준다는 점에서 큰 의의가 있다.

- 행위자 기반 모형 (Agent-Based Models, ABM): ABM은 다양한 특성을 가진 행위자들(예: 기술적 분석가, 기본적 분석가, 시장조성자)의 상호작용을 컴퓨터 시뮬레이션을 통해 구현하여, 군집행동을 포함한 거시적 시장 현상(예: 변동성 군집, 버블)이 어떻게 미시적 행동으로부터 발생하는지를 탐구한다. 복잡한 동학을 분석하는 데 강력한 도구로서 금융시장의 변동성 군집, 팻테일(fat-tail) 수익률 분포 등 복잡한 동태적 현상을 설명하는데 강점을 갖지만, 모형의 보정(calibration)과 정량적 추정이 어렵다는 한계가 있다(Kim & Kim, 2014; Lux, 1998; Wray & Bishop, 2016). ABM은 특정한 단일 모형보다는 다양한 아이디어를 구현하는 시뮬레이션 프레임워크에 가깝다. 행위자 기반 모형의 군집행동을 모델링하는 주요 방식으로는 물리학의 상호작용 입자 시스템에서 아이디어를 얻은 상호작용을 통한 의견 전파모형(Opinion Dynamics via Interaction), 시장의 불확실성(주로 변동성으로 측정)이 높아질 때, 행위자들이 자신의 분석보다 다른 사람들의 행동을 더 의존하게 되어 행동의 이질성이 감소하고 동조화가 강화되는 현상을 모형화

한 불확실성에 따른 행동 동조화 모형 (Behavioral Synchronization under Uncertainty) 등이 있다. 이처럼 행위자 기반 모형은 행위자들의 제한된 합리성과 상호작용 규칙을 통해, 실제 금융시장에서 관찰되는 복잡한 거시 현상들이 어떻게 미시적 수준에서 창발하는지를 탐구하는 강력한 도구를 제공할 수 있다. 표 2는 앞서 설명한 4가지 방식의 군집 효과 측정방법을 정리한 것이다.

표 2 군집 효과 측정 방법들

| 모형            | 주요 제안자                    | 기본 원리                              | 핵심 방정식/검증   | 데이터 요구사항           | 장점/단점   |
|---------------|---------------------------|------------------------------------|---|--------------------|---|
| <b>CSSD</b>   | Christie & Huang (1995)   | 군집행동 발생 시 수익률 분산도 감소               | 극단적 시장 움직임 더미 변수의 음(-)의 유의성                       | 일별 주식 수익률          | 선구적 모형, 개념이 직관적<br>이상치에 민감                            |
| <b>CSAD</b>   | Chang et al. (2000)       | 군집행동 발생 시 수익률 분산도와 시장 수익률 간 비선형 관계 | $CSAD_t$ 공식에서 $\gamma_2 < 0$ 검증                   | 일별 주식 수익률          | 이상치에 강건, 널리 사용되는 표준 방법론<br>진정한 군집행동과 외견상 군집을 구분하기 어려움 |
| <b>구조적 모형</b> | Cipriani & Guarino (2014) | 이론 모형의 파라미터를 직접 추정하여 군집행동 확률 계산    | 최대우도추정법(MLE)을 통한 파라미터 ( $\alpha, \mu, \tau$ 등) 추정 | 고빈도 거래 데이터 (틱 데이터) | 이론적 기반이 명확, 군집행동의 원인 추론 가능<br>데이터 요구 수준이 높고 모형이 복잡함   |

|                                     |   |   |                              |   |   |
|-------------------------------------|---|---|------------------------------|---|---|
| <b>행위자<br/>기반 모<br/>형<br/>(ABM)</b> | Lux (1995),<br>Cont &<br>Bouchaud<br>(2000),<br>Schmitt &<br>Westerhoff<br>(2017) | 제한된 합리<br>성을 가진 이<br>질적 행위자<br>들의 상호작<br>용을 시뮬레<br>이션하여 거<br>시적 시장 현<br>상을 분석 | 특정 방정<br>식보다는<br>시뮬레이션<br>기반 | 이론적 시뮬레<br>이션이 주 목<br>적이거나, 실제<br>시장 데이터<br>(가격, 거래량<br>등)를 '양식화<br>된 사실'과 비<br>교하여 모형<br>보정 및 검증<br>에 사용 | 복잡한 동학 및<br>창발 현상 분석<br>에 강력, 행위자<br>이질성 포착 용<br>이<br>모형 보정 및<br>정량적 추정이<br>어려움, 결과가<br>초기 가정에 민<br>감 |
|-------------------------------------|---|---|------------------------------|---|---|

## 2. 금융 시장에서의 증거

글로벌 연구들은 군집행동이 항시적으로 나타나는 현상이 아니라, 특정 시장 상황에 따라 그 강도가 달라진다는 점을 일관되게 보여준다. 특히 2008년 글로벌 금융위기나 COVID-19 팬데믹과 같은 시장 스트레스가 심하고 변동성이 높은 기간에 군집행동이 현저하게 강화되는 경향이 있다(Bouri et al., 2021; Chiang & Zheng, 2010; Hwang & Salmon, 2004).

군집행동은 선진 시장보다 신흥 및 프론티어 시장에서 더 뚜렷하게 나타나는 경향이 있다. 이는 정보 비대칭성이 크고, 시장 투명성이 낮으며, 상대적으로 경험이 적은 투자자 비중이 높은 신흥 시장의 구조적 특성에 기인하는 것으로 해석된다(Chang et al., 2000; Chiang & Zheng, 2010). 군집행동의 강도는 시장 상승기(강세장)보다 하락기(약세장)에 더 강하게 나타나는 비대칭적 특성을 보이는 경우가 많다(Chang et al., 2000; Chiang & Zheng, 2010).

한국 시장의 쏠림현상은 이승호(2016)에 잘 정리되어 있다. 해당 보고서에는 자산 쏠림 현상을 해당자산으로 자금이 이동하는 것으로 보고 크게 네 가지 유인으로 인해 자본쏠림현상이 발생한다고 보고 있다.

먼저 은행에서 투자금융으로의 이동으로, 전통적인 은행 예금(저축성 자금)에서 증권사 및 자산운용사가 취급하는 금융투자상품(투자성 자금)으로 자금이 이동하는 거대한 흐름이 관찰된다. 은행의 수신 비중은 감소한 반면, 증권사와 자산운용사의 비

중은 크게 증가했다. 두 번째로, 자금이동의 이동의 동인으로, 지속적인 저금리 기조와 인구 고령화이고 이로 인해 구조적 변인이 생겼음을 지적한다. 저금리는 은행 예금의 매력을 감소시키는 반면, 고령화는 노후 대비를 위해 더 높은 기대수익률을 추구하게 만들어 투자성 상품에 대한 수요를 촉진하는 방향이다. 세 번째로, 인기상품에 대한 집중이다. 투자성 상품 카테고리 내에서도 자금 쓸림 현상이 뚜렷하다. 2000년대 중반에는 주식형 펀드에 자금이 몰렸으나, 최근에는 채권형 펀드, 혼합형 펀드, 그리고 특히 파생결합증권(ELS, DLS)과 같은 특정 상품으로 자금이 집중되는 패턴을 보인다. 이는 '그때그때 뜨는' 상품으로 자금이 몰리는 전형적인 '주제의 쓸림' 현상이다. 마지막으로 유인으로 불확실성을 꼽으며, 수익률 추구 경향에도 불구하고, 글로벌 불확실성이 높아지는 시기에는 위험회피 성향이 강화되어 투자성 상품에서 자금이 유출되고 안전자산인 예금성 상품으로 회귀하는 '안전자산 선호' 현상이 나타난다. 이러한 동학은 자금 흐름에 밀고 당기는 변동성을 야기한다

이러한 분석은 한국의 쓸림현상을 온전히 이해하기 위해서는 국제적으로 통용되는 자본시장 균집행동 측정 방법론만으로는 불충분함을 시사한다. CSAD와 같은 고빈도 수익률 기반의 측정 지표는 주식시장에서의 투자자 균집행동을 포착하는 데 유용하지만, 은행의 특정 부문 대출 집중과 같은 저빈도(월별 또는 분기별) 대차대조표 데이터에 나타나는 금융기관 수준의 자금 쓸림을 측정하지 못하는 방법론적 한계가 존재한다. 따라서 한국의 금융안정성을 종합적으로 분석하기 위해서는 자본시장 균집행동과 신용시장 자금 쓸림을 모두 포괄할 수 있는 이원적인 분석 체계가 필수적이다.

### III. 금융안정 영향 분석을 위한 프레임워크

#### 1. 데이터 및 변수 구성

본 분석은 고빈도 시장 데이터와 저빈도 거시경제 및 금융기관 데이터를 통합적으로 활용한다. 분석에 필요한 변수와 데이터 출처는 다음과 같다.

표 3 변수 리스트

| 변수명 (기호)                | 설명   | 빈도 | 단위         | 데이터 출처 (시계열 코드)                        |
|-------------------------|--|----|------------|--|
| 금융안정지수 (FSI)            | 금융 시스템의 단기적 안정 상황을 종합적으로 나타내는 지수. 은행, 금융시장, 대외, 실물경제 등 6개 부문 20개 지표로 구성. | 월별 | 지수 (0~100) | 한국은행 경제 통계시스템 (ECOS)                   |
| 자본시장 균집 행동 ( $HERD_t$ ) | CSAD 모형을 통해 추정된 시변동적 균집행동 계수의 월별 평균값.                                    | 월별 | 계수 값       | KOSPI 200 구성 종목 일별 수익률 데이터 (FnGuide 등) |
| 신용시장 쏠림 ( $CONC_t$ )    | 총 은행 대출 대비 주택담보 대출의 비중. 특정 부문으로의 자금 집중도를 측정.                             | 월별 | %          | 한국은행 경제 통계시스템 (ECOS)                   |
| 기준금리 ( $PolicyRate_t$ ) | 한국은행의 통화정책 방향을 나타내는 정책금리.  | 월별 | %          | 한국은행 경제 통계시스템 (ECOS)                   |
| VKOSPI                  | KOSPI 200 지수 옵션 가격   | 월별 | 지수         | 한국거래소                                  |

|                                      |  |            |            |  |
|--------------------------------------|--|------------|------------|--|
|                                      | 에 내재된 시장 변동성 지수. 국내 시장의 위험 인식을 반영.             |            |            | (KRX)                                      |
| <b>VIX</b>                           | S&P 500 지수 옵션 가격에 내재된 변동성 지수. 글로벌 위험회피 심리를 대변. | 월별         | 지수         | CBOE, Federal Reserve Economic Data (FRED) |
| <b>LTV/DTI 규제 (REG<sub>t</sub>)</b>  | 주택담보대출 관련 LTV/DTI 규제의 강도를 나타내는 더미 변수 또는 지수.    | 월별         | 더미/지수      | 정책 발표 자료 기반 자체 구축                          |
| <b>실질 GDP 성장률 (GDP<sub>gt</sub>)</b> | 국내 실물경제 상황을 나타내는 지표.                           | 분기 (월별 보간) | 전년 동기 대비 % | 한국은행 경제 통계시스템 (ECOS)                       |
| <b>소비자물가지수 (CPI<sub>t</sub>)</b>     | 인플레이션 압력을 측정하는 지표.                             | 월별         | 전년 동월 대비 % | 통계청  |

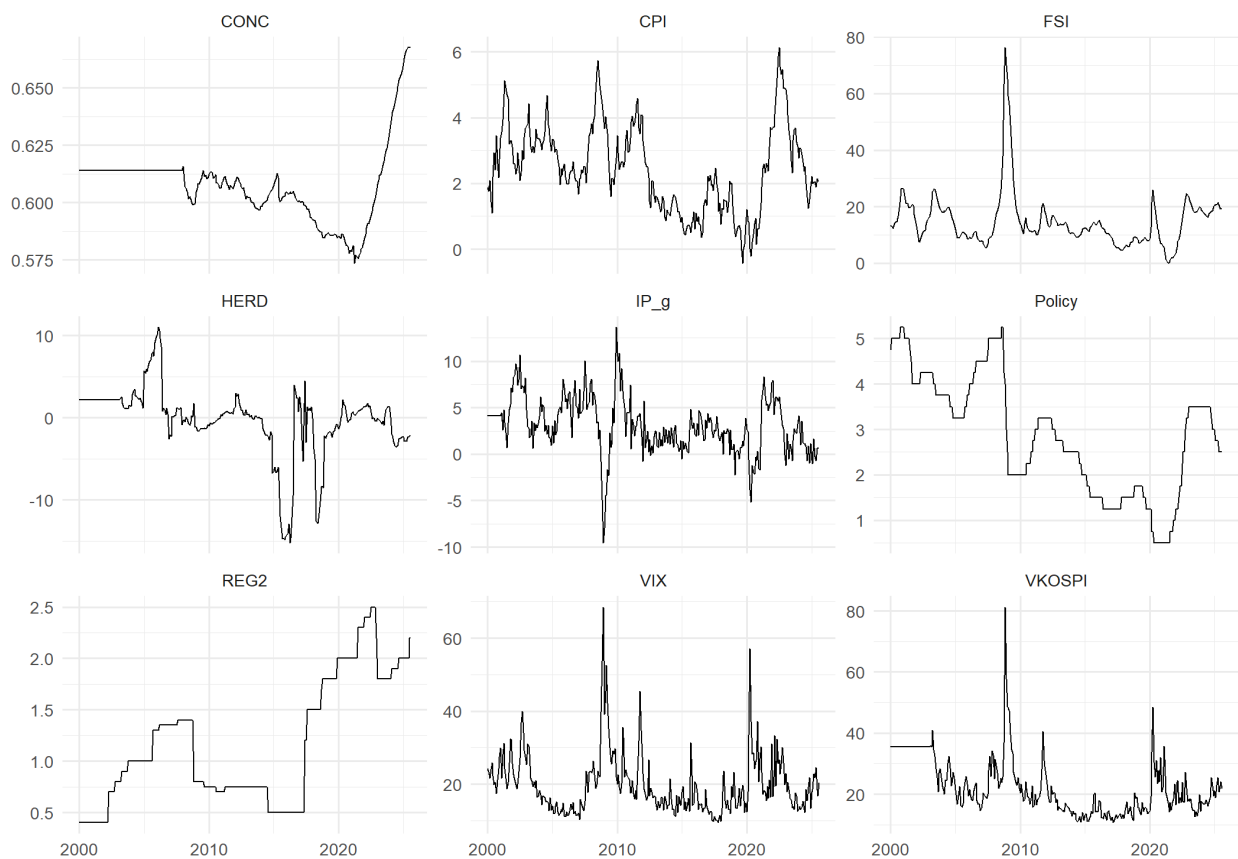
먼저 종속변수는 금융안정을 나타내는 것으로, 주 지표는 한국은행의 금융안정지수(FSI)를 사용한다. FSI는 금융 시스템의 전반적인 스트레스 수준을 종합적으로 보여 주어 정책적 함의가 크다. 분석의 강건성을 위해 시장 변동성을 나타내는 VKOSPI와 국가 신용위험을 나타내는 CDS 프리미엄을 보조 지표로 활용할 수 있다.

한편, 쏠림현상을 나타낼 핵심 독립변수로는 자본시장 균집행동 지수와 신용시장 쏠림지수가 있다. 자본시장 균집행동 (HERD<sub>t</sub>)은 Chang et al. (2000)의 CSAD 방법론을 적용하여 측정한다. KOSPI 200 구성 종목의 일별 수익률 데이터를 사용하여 일별 CSAD<sub>t</sub>를 계산한 후, 이동창(rolling-window) 회귀분석 또는 칼만 필터(Kalman filter)를 이용해 시변동(time-varying) 계수  $\gamma_{2,t}$ 를 추정한다(Hwang & Salmon, 2004). 최종적으로 이 일별 계수들을 월별 평균 등으로 집계하여 월별 시계열 변수 HERD<sub>t</sub>를 생성한다. 신용시장 쏠림 (CONC<sub>t</sub>)은 금융기관 수준의 자금 쏠림을 포착하기 위해, 한국은행 ECOS에서 제공하는 부문별 대출 데이터를 활용한다. 예를 들어, 총 은행 대출 대비

주택담보대출 비중을 계산하여, 이 비율의 급격하고 지속적인 상승을 쏠림 현상의 발생으로 정의한다.

통화정책의 기초는 한국은행 기준금리와 M2 증가율로 측정한다(Gong & Dai, 2017; Krokida et al., 2020). 한국의 주요 거시건전성 정책 수단인 LTV·DTI 규제 강도 변화는 시점별 더미 변수나 규제 수준을 지수화하여 모형에 포함시킨다. 이 외에 실물경제 상황(GDP 성장률, CPI)과 글로벌 요인(VIX, 미 연준 금리)을 통제변수로 포함하여 분석의 정확성을 높인다.

그림 1. 주요 변수 시계열



## 2. 분석 모형 및 추정 전략

변수들 간의 동태적 상호관계를 분석하기 위해, 본 연구는 벡터 자기회귀(VAR) 모형을 핵심 분석틀로 사용한다. VAR 모형은 특정 변수들 간의 인과관계에 대한 강한 이론적 제약을 가하지 않고, 데이터 자체의 동학을 통해 변수들 간의 상호 영향을 분석하는 데 적합하다(Loisel, Pommeret, & Portier, 2012).

### 2.1. 모형 설정 및 추정 절차

기준 VAR 모형은 다음과 같이 설정할 수 있다:

$$Y_t = c + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + \varepsilon_t \quad (2)$$

여기서  $Y_t$  는 내생변수들의 벡터이며, 예를 들어 다음과 같이 구성될 수 있다:

$$Y_t = \{HERD_t, CONC_t, PolicyRate_t, Reg_t, GDPg_t, CPI_t\}$$

이러한 구성은 신용시장 쏠림, 자본시장 군집행동, 금융안정, 통화정책, 그리고 국내외 변동성 간의 동태적 상호작용을 동시에 분석할 수 있게 한다.

분석에 사용될 모든 시계열 변수에 대해 ADF(Augmented Dickey-Fuller), PP(Phillips-Perron) 검정 등을 실시하여 안정성(stationarity)을 확인하여, 불안정적인 변수는 차분하여 안정적 시계열로 변환하였으며, 단위근 검정의 결과에 따라, *HERD*, *Conc*, *Policy*, *CPI*, *Reg2* 변수 모두 적절히 차분하였다. 또한 AIC를 이용하여 VAR 모형의 최적 시차(p)를 결정하였다.

이와 같이 전처리한 시계열 변수들을 대상으로 변수들 간의 통계적 선행 관계를 파악하기 위해 그랜저 인과관계(Granger Causality) 검정을 실시하고, VAR모형 추정을 통해 충격반응함수를 계산한다. 이와 같은 분석을 통해 예를 들어, ' $HERD_t$ 가  $FSI_t$ 를 그랜저 인과하는가?' 또는 ' $PolicyRate_t$ 가  $CONC_t$ 를 그랜저 인과하는가?'와 군집행동 변수( $HERD_t$ )에 충격이 발생했을 때 금융안정지수( $FSI_t$ )가 어떻게 반응하는지를 시각적으로 분석하여 그 영향의 크기와 지속성을 파악한다. 마지막으로 각 변수의 예측

오차 분산 중에서 다른 변수들의 충격이 차지하는 비중을 시차별로 계산하였다. 이를 통해 금융안정 변동의 주요 원인이 무엇인지, 즉 군집행동이나 신용 쓸림이 다른 요인들에 비해 얼마나 중요한지를 정량적으로 평가하고자 한다.

## 2.2. 모형 분석 결과

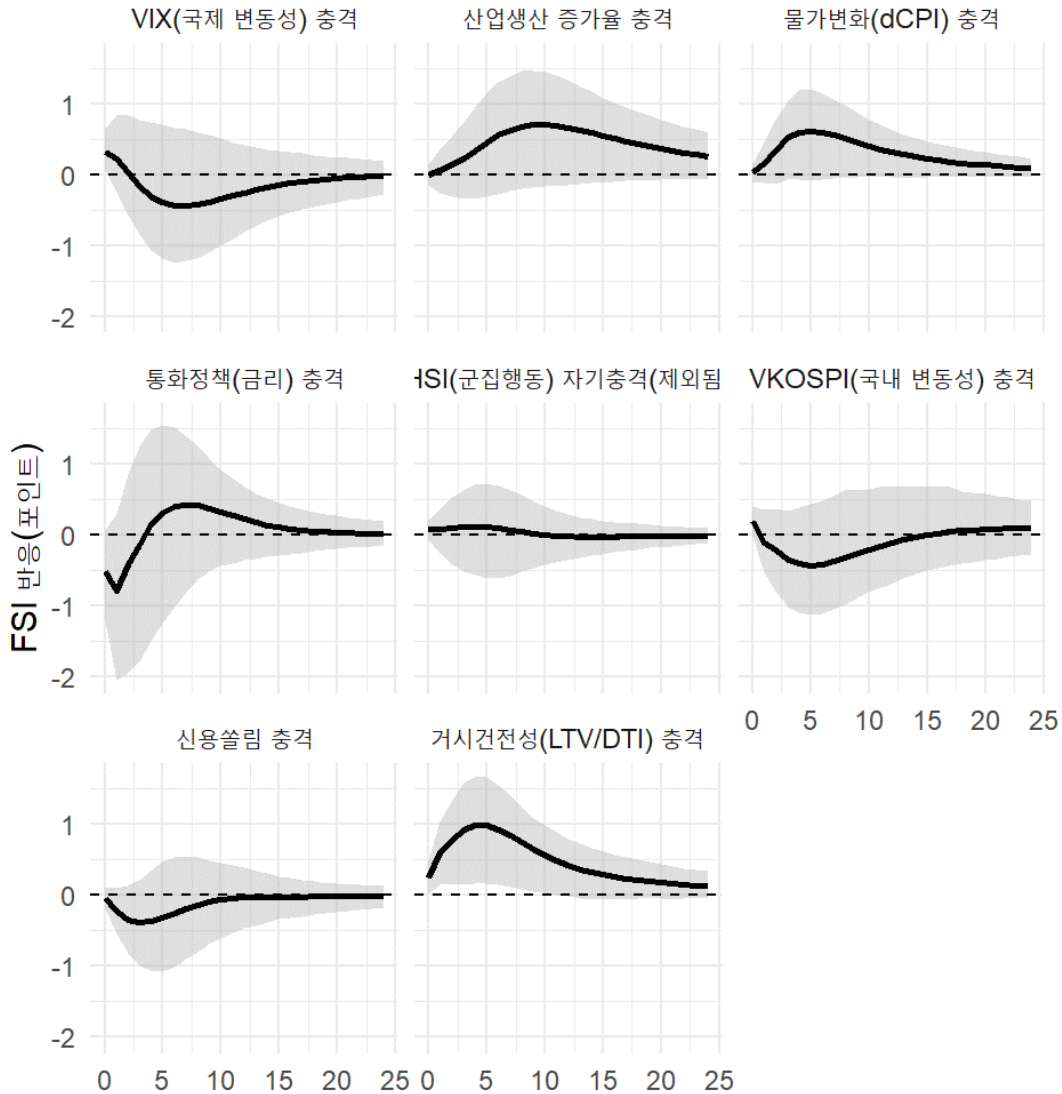
앞장에서 논의한 바와 같이 정보비대칭과 사회적 학습, 평판 위험 및 보수체계 등이 결합될 경우 투자자들은 합리적 근거보다 남들의 행동을 따라가는 군집행동을 보일 수 있고, 이러한 쓸림현상은 자산시장 거품과 금융시스템 불안정으로 이어질 수 있다 (Banerjee, 1992; Bikhchandani et al., 1992; Scharfstein & Stein, 1990). 이를 실증적으로 확인하기 위해 HSI(Herd Sentiment Index)와 거시변수들을 포함한 VAR 모형을 추정하고 군집행동의 발생 메커니즘과 거시정책 효과를 분석하였다. 주요 변수로는 본 연구에서 구축한 군집심리 지수(HSI), 한국은행 금융안정보고서의 금융불안지수(FSI)(시스템 리스크 지표), 기준금리(통화정책), 및 거시건전성 지수(LTV·DSR 규제 등 정책변화를 반영) 등이 포함되었고, 산업생산증가율, 글로벌 VIX 등 통제변수도 고려되었다. VAR 모형의 안정성 요건을 충족한 상태에서 충격반응분석과 Granger 인과관계 검정을 실시하였다.

본 연구에서는 군집행동과 신용쓸림이 금융시스템 안정성에 미치는 동태적 영향을 분석하기 위해 확장형 VAR 모형을 추정하였다. 모형에는 금융안정 상황을 나타내는 금융불안지수(FSI), 투자자들의 군집심리를 계량화한 군집행동 심리 스코어(Herd Sentiment Index, HSI), 국내 은행권의 신용쓸림 지표(부문별 대출 집중도 지수), 중앙은행의 통화정책을 대표하는 기준금리(콜금리)와 금융당국의 거시건전성 정책강도 지수(예: LTV·DSR 등 여신규제 종합지수), 그리고 글로벌 요인으로 VIX 지수(미국 주식시장의 변동성 지표)를 포함하였다. 분석 표본 기간은 HSI 및 신용쓸림 지표가 확보되는 2003년 이후 월별 자료로 하였으며, 최적 시차는 AIC 기준 3 시차로 설정하였다. VAR 모형의 안정성 조건을 만족하였고 자기상관 등 진단검정에서도 양호한 것으로 나타났다.

먼저 <그림 2>에서 FSI에 대한 충격반응을 살펴보면, 거의 대부분의 변수에서 급작스런 FSI에 대한 반응은 유의미하지 않게 나온다. FSI 자체가 위험도를 나타내는 지수이지 직접적으로 거시 변수에 영향을 미치는 변수가 아니기 때문에 충격반응이

크지 않을 수 있다. 다만 거시건전성과 통화정책의 경우 어느 정도 영향력이 있는 것으로 나타나는데, 이는 FSI 를 통해 금융시장의 위험성에 대한 경고가 나타나는 경우, 정책적으로 선제 조치를 취한 결과라고도 볼 수 있다.

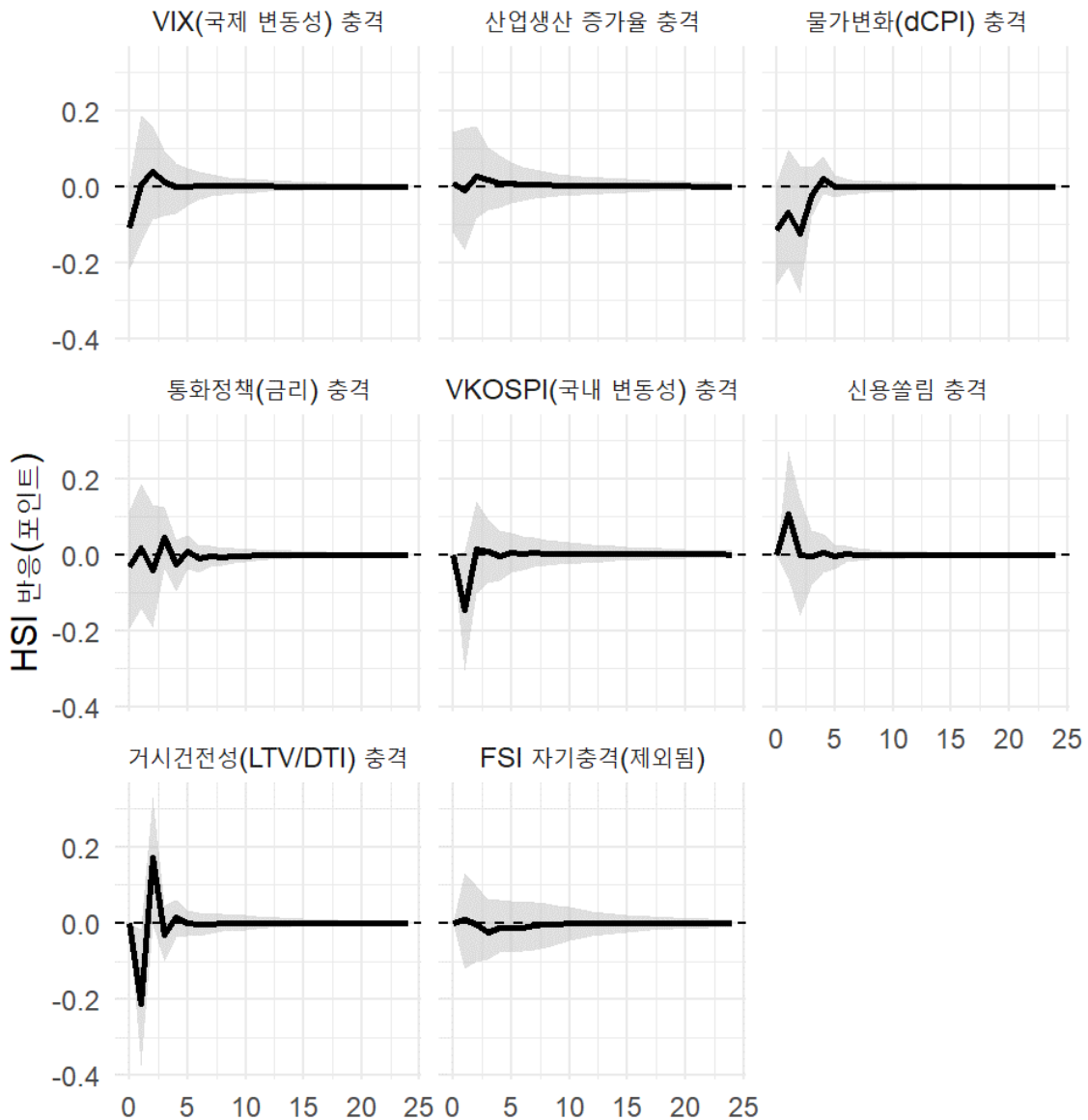
그림 2. FSI 충격에 대한 반응함수



다음으로 군집심리 충격에 대한 반응(그림 3)을 살펴보면, 투자자들의 군집행동이 갑작스럽게 강화되었을 때 금융시장과 실물경제에 미치는 영향을 가늠할 수 있다. d HERD 지표의 급등 충격은 무엇보다 금융안정성의 악화로 직결되지는 않는데, FSI 가 유의미하게 충격을 받지 않는 것으로 나타나, 기존의 연구들과는 다른 모습을 보인다. 일반적인 군집심리지표가 상승하게 되면 시장에서 투자자들이 위험자산에 일방적으로 몰리거나 회피하면서 불균형이 축적되고 금융 불안으로 이어질 수 있다는

동학을 예측할 수 있는데, FSI로 측정되는 금융불안을 직접적으로 유발하는 것은 아닌 것으로 나타났다. 다만, 군집행동은 금융안정의 측면보다는 변동성 충격으로 조금 더 잘 나타난 모습을 보이고 있다. VKOSPI에는 군집심리의 결집이 영향을 받는 것으로 나타나는데, 초과수익을 쫓아 모두가 한 방향으로 매매할 경우 가격 변동성이 커지고 시장 과열 및 급변 동향이 나타난다고 할 수 있다. 글로벌 VIX도 비슷한 방향으로 움직이는데, 비단 우리나라에서만 군집현상이 발견되는 것이 아니라면 동일한 동학으로도 설명할 수 있다. 이와 같은 결론을 얻기 위해서는 글로벌 시장의 관점에서 군집 현상에 대한 분석이 필요하다.

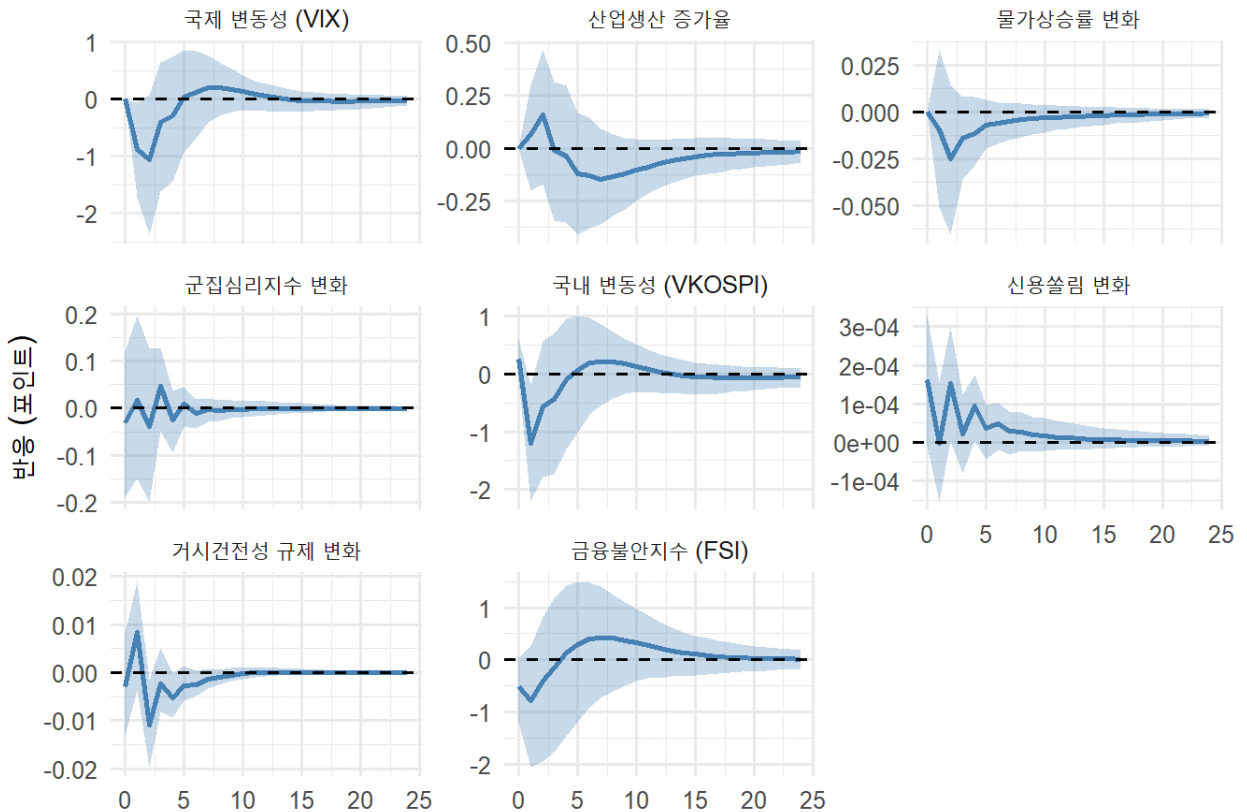
**그림 3. HERD의 충격에 대한 반응함수**



실물경제 부문에 대해서는 군집심리 충격으로 인한 직접 효과가 크지 않았다. 산업생산의 경우 초기 몇 분기 동안 소폭 상승과 하락을 반복하다가 통계적으로 유의미하지 않은 미약한 감소세를 보였고, 물가도 미세한 하락 압력이 나타났으나 그 효과는 제한적이었다. 이는 투자심리 요인만으로는 단기에 실물경제 경로에 큰 영향을 주지 못하며, 다만 금융시장의 불안정이 지속되고 이를 통해 자산가격 조정이나 신용경로를 통해서만 간접적으로 실물에 파급될 수 있음을 시사한다. 끝으로, 군집심리 충격에 대해 거시건전성 지표인 dCONC와 dREG2 어느 쪽도 유의한 반응을 보이지 않았다. 이는 시장 심리 변화에 정책당국이 즉각적으로 대응하기 어렵거나 또는 당시 거시건전성 정책이 별도로 조정되지 않은 결과로 볼 수 있다.

본 모형의 정합성을 확인하기 위해 그림 4에서 통화정책 긴축 충격이 발생할 경우 주요 변수들의 반응 경로를 확인할 수 있다. 국제 금융시장의 변동성(VIX)과 국내 주가시장의 변동성(VKOSPI)은 충격 직후 일시적으로 하락하였다. 특히 VKOSPI는 1분기 시점에 통계적으로 유의하게 큰 폭의 하락을 보여주며, 예상치 못한 통화긴축이 발표될 때 시장의 과열 심리가 순간적으로 진정되는 효과를 시사한다. 그러나 이러한 안정 효과는 지속되지 못하고 2분기 이후에는 오히려 변동성이 상승 전환하여 중기적으로 금융시장 변동성이 커질 가능성을 내포한다. 실제로 금융불안지수(FSI)의 경우 통화정책 충격에 대해 초기(0~2분기)에 하락 반응을 보이다가 이후 상승세로 방향을 바꾸어 5~7분기경 정점을 형성한 후 서서히 기초 수준으로 복귀하는 역U자형 경로를 나타냈다. 이는 통화긴축이 초기에 물가안정 기대 제고 등으로 시장안정을 가져올 수 있으나, 시간 경과에 따라 경영여건 악화와 위험자산 가격 하락 등의 영향으로 오히려 금융불안이 높아질 수 있음을 보여준다. 실물부문의 경우 통화긴축 충격 이후 산업생산 증가율(IP\_g)은 1~2분기 동안 다소 상승하는 price puzzle 현상이 관찰되지만 통계적으로 유의하지는 않았으며, 3분기 이후부터는 음(-)의 반응으로 전환되어 산업생산이 점차 둔화되는 추이를 보였다. 이는 통화정책의 실물경기 하방압력이 시차를 두고 나타나는 전형적인 패턴으로 해석된다. 물가상승률의 증감률(dCPI)은 통화긴축 충격에 대체로 지속적인 하락 반응을 보여, 결과적으로 긴축적 통화정책이 예측대로 물가를 낮추는 방향으로 작용함을 확인할 수 있다. 전반적으로 통화 긴축이 금융 시장에 미치는 단기적 진정 효과와 중기적 금융 불안정 증가(역U자형 경로), 그리고 실물 경기의 시차를 둔 하방 압력 등 대부분의 경로가 일반적인 거시경제 및 금융 안정 이론에 부합한다.

그림 4. 통화정책 충격에 대한 반응함수



다만, 'Price Puzzle' 현상은 거시경제 모형에서 자주 관찰되지만 이론적으로는 예외적인 현상으로, 이 부분에 대한 추가적인 분석 또는 모형의 진단은 필요할 수 있다.

한편 군집심리지수의 변화(dHERD)는 통화정책 충격에 대해 뚜렷한 방향성을 보이지 않고 미미한 진폭의 진동에 그쳐, 금리 변화만으로는 투자자들의 군집행동을 유의하게 변화시키기 어려움을 시사한다. 아울러 신용대출(dCONC)이나 거시건전성 규제 지표(dREG2)는 통화충격에 거의 반응하지 않는 것으로 나타나, 기준금리 변동이 단기적으로 금융시장의 구조적 대출이나 거시건전성 정책기조를 직접적으로 변화시키지는 않음을 알 수 있다. 종합하면, 통화정책 충격은 자산시장 변동성과 금융불안에 단기적인 진정 효과 후 중기적인 불안정 증가 효과를 함께 야기하며, 실물경제에는 지연된 경기둔화 및 물가하락 압력으로 나타나는 반면, 군집행동이나 신용분포 구조에는 유의한 영향이 미미한 것으로 해석된다. 이는 전통적 통화정책 수단이 투자자 심리나 자금대출에는 직접적 영향을 주기 어렵다는 한계를 보여준다.

이상의 동태적 관계에 대한 그레인저 인과 검정 결과(표 3)는 주요 변수들 간 영향 방향을 통계적으로 확인해준다. 분석 결과, 군집심리 지표(HSI)의 변화가 향후 금

용불안지수(FSI)에 미치는 영향은 통계적으로 유의하지 않은 것으로 나타났다. 이는 그림 2의 충격반응함수에서 군집심리(HSI) 충격에 대해 FSI가 통계적으로 유의미한 반응을 보이지 않는 결과와 일치한다. 즉, 본 모형 내에서는 군집행동의 심화 자체가 직접적으로 즉각적인 금융불안을 유발한다는 증거는 발견되지 않았다.

표 3 그랜저 인과성 검정 결과

| 원인변수 (X)        | 결과변수 (Y)        | F 통계량 | p-값   |
|-----------------|-----------------|-------|-------|
| 통화정책 (기준금리)     | 군집행동 (HSI)      | 0.18  | 0.835 |
| 통화정책 (기준금리)     | 금융불안 (FSI)      | 9.40  | 0.000 |
| 거시건전성 정책 (Reg2) | 군집행동 (HSI)      | 5.68  | 0.004 |
| 군집행동 (HSI)      | 금융불안 (FSI)      | 0.04  | 0.958 |
| 금융불안 (FSI)      | 거시건전성 정책 (Reg2) | 19.09 | 0.000 |
| 글로벌 변동성 (VIX)   | 금융불안 (FSI)      | 2.93  | 0.055 |

반면, 정책 변수들은 주요 경제 변수들에 대해 뚜렷한 인과성을 보였다. 우선 통화정책(기준금리)은 금융불안(FSI)에 대해 1% 유의수준에서 강한 그레인저 인과관계(p=0.000)를 보였으며, 그림 4에서 확인한 바와 같이 금리 충격은 시차를 두고 금융불안지수의 변동을 유발하는 것으로 나타났다. 또한, 거시건전성 정책(Reg2)은 군집행동(HSI)에 대해 유의한 인과관계(p=0.004)를 보여, LTV·DTI와 같은 규제 정책의 변화가 투자자들의 쓸림 심리에 유의미한 영향을 미칠 수 있음을 시사한다.

종합하면, 본 분석 결과는 투자자의 군집행동이 독자적으로 금융불안을 야기하기 보다는, 통화정책과 거시건전성 정책의 기조 변화가 금융시장 안정과 투자자 심리에 더 결정적인 동인으로 작용함을 보여준다. 특히 거시건전성 정책이 군집심리에 유의한 영향을 미친다는 점은, 시장 과열 시 금리 정책뿐만 아니라 대출 규제 등 미시적 정책 수단을 적극 활용하는 것이 투자자 심리 안정에 효과적일 수 있음을 시사한다. 따라서 향후 금융안정을 위해서는 군집심리에 대한 지속적인 모니터링과 더불어, 이러한 정책 수단들이 시장 심리와 안정성에 미치는 파급효과를 정교하게 고려한 정책 운용이 필수적이다.

### 2.3. 정책적 시사점

이상의 VAR기반 분석 결과를 토대로 금융안정 정책 방향에 대한 몇 가지 시사점을 도출할 수 있다. 첫째, 금융시장 군집행동에 대한 조기경보체계를 구축하여 선제 대응할 필요가 있다. 본 연구에서 군집심리지표(dHERD)가 금융안정에 미치는 영향이 크지는 않지만, 향후 불확실성을 선행할 수 있음은 확인이 되었다. 이에 따라 군집심리 측정이 금융안정 측면에서 정례적으로 모니터링 할 지표임이 확인되었다. 그러나 군집심리 지표는 결과론적인 지표이기 때문에 이를 선제적으로 확인할 수 있는, 예를 들어 텍스트마이닝 등을 활용한 투자심리 지표(HSI 등)를 개발의 필요성이 강조된다. 이와 같은 지표가 일정 임계치를 초과할 경우 “쏟림 경보”를 발령한 뒤 경고 메시지의 시장 전파나 일부 부문의 거래·신용 규제를 단계적으로 시행하는 방안을 고려할 수 있다. 이를 통해 군중심리가 과열되는 국면에서 사전에 위험 신호를 포착하고 완화 조치를 취함으로써, 군집행동의 부작용을 줄일 수 있을 것이다.

둘째, 통화정책과 거시건전성정책의 유기적 연계가 중요하다. 군집행동이 심화된 시장에서는 금리정책만으로는 이러한 비이성적 폭주를 억제하기에 한계가 있다. 실제 분석에서도 금리충격이 군집심리에 유의한 영향을 주지 못했으며, 오히려 군집행동이 금융불안을 야기하는 것으로 나타났다. 따라서 자산시장의 쏟림과 금융불안 징후가 포착될 경우 LTV·DSR 등 거시건전성 규제수단을 신속하게 발동하여 특정 부문의 과열을 진정시키는 등 통화정책과 보완적으로 운용할 필요가 있다. 실제 Shiller(2000) 등의 연구도 과열된 시장에서는 금리정책만으로 대응하기 어렵다는 점을 강조하고 있어, 통화정책과 거시건전성정책의 정책 공조는 금융불안정 위험을 관리하는 핵심 축이라 할 수 있다.

마지막으로, 정형 데이터에 기반한 쏟림 현상의 측정은 선제적인 금융 안정에 대한 조기 경보로서 선제적인 역할이 제한적이기 때문에 군집행동을 소셜미디어, 뉴스 등 비정형 데이터에서 탐지되는 집단심리 변화에서 측정할 필요성이 강조된다. 와 밀접한데, 이러한 정보를 활용하면 보다 신속한 이상징후 포착이 가능하다. 이에 따라 본 연구에서는 보다 선제적인 금융 쏟림 현상을 측정하기 위해 뉴스미디어에서 나타나는 심리적인 쏟림현상을 측정하기 위해 군집행동 심리지수를 작성하고자 한다. 이를 통해 투자심리 변동에 기인한 금융불안정 리스크를 선제적으로 관리할 수 있을 것으로 판단된다.

## IV. 텍스트 마이닝을 통한 군집행동 심리지수 작성

금융자산시장 쓸림현상의 조기 경보를 위해, 본 장에서는 군집행동 심리 스코어(Herd Sentiment Index, HSI)를 구축하고 이를 활용한 위기 예측 모형을 제시한다. HSI는 투자자들의 군집행동에 따른 시장 심리를 정량화한 지표로서, 뉴스와 소셜미디어의 감성 및 투자자 관심도 변화를 결합하여 측정한다. 이후 HSI 등 텍스트 마이닝 지표를 독립변수로 활용하여 위기 발생 확률을 예측하는 로지스틱 회귀, 랜덤포레스트, LSTM 등의 기계학습 모형을 정의하고 비교한다. 마지막으로 각 모형의 위기 예측 성능을 ROC 곡선, AUC, 정밀도, 재현율, F1 점수와 같은 평가 지표를 통해 살펴본다.

### 1 HSI 구축 개요 및 시각화

본 연구에서는 뉴스·포털 텍스트 및 검색량 데이터를 활용해 군집행동 심리 스코어(Herd Sentiment Index, HSI)를 도출하고, 이를 통해 쓸림현상의 조기경보 신호를 파악한다. 기존의 단순 감성분석(예: 긍정어 수 - 부정어 수 비율)은 문맥 해석력이 부족하므로, 보다 정교한 문맥 기반 감성분석을 위해 BERT(Bidirectional Encoder Representations from Transformers) 모델을 활용한다.

#### (1) 데이터 준비 및 전처리

국내 주요 경제신문 및 포털뉴스(예: 네이버 뉴스), 경제전문 블로그/SNS 게시물 중 주식, 부동산, 암호자산 관련 키워드("폭등", "급락", "투자열풍" 등)가 포함된 기사·게시글을 수집할 수 있지만, 본 연구에서는 모집단이라고 할 수 있는 기사의 수가 비교적 일정하고 표본기간내에 누락없이 제공되고 있는 기간통신사인 연합뉴스의 기사만을 대상으로 한다. 기사들은 모두 네이버에서 제공되는 네이버뉴스를 스크래핑한 데이터이다. 연합뉴스는 다만 통신사이기 때문에 동일한 기사가 여러 번 올라오는 경우가 있어서 기사 제목과 본문 길이 기준으로 동일하게 일치하는 데이터를 제거하였다. 수집된 기사는 HTML 태그, 공백, 특수문자 제거 등 텍스트 정규화 및 클리닝을 실시한다. 이와 같은 작업을 통해 얻은 텍스트 원본을 한국어 형태소 분석기를 이용하거나 BERT의 WordPiece 기반 토큰라이저를 활용하여 텍스트를 토큰화한다. 즉, 뉴스 데이터 중 아래의 특정 키워드들이 본문 또는 제목에 존재하는 기사만

추출하여 이에 대해 감성분석을 적용하였다. 기사 검색의 키워드는 '주식', '부동산', '암호자산', '비트코인', '투자', '펀드', '증시', '코스피', '코스닥', '아파트', '매매', '전세', '가상화폐', '암호화폐'들이다.

## (2) BERT 기반 감성분석

BERT 모델은 문맥을 양방향으로 인지하므로, 기사 단위 감성 점수를 보다 정밀하게 산출할 수 있다. 한국어 BERT(KoBERT 등)에 금융 관련 감성 레이블링된 코퍼스를 추가 학습(fine-tuning)하여, 각 문서  $d$ 의 감성 점수  $SentBERT(d)$ 를  $-1$ 에서  $+1$  사이의 연속값으로 도출한다.

$$SentBERT \in [-1, +1]$$

값이  $+1$ 에 가까울수록 낙관적,  $-1$ 에 가까울수록 비관적 투자심리를 의미한다. 문단별 점수를 계산한 뒤 가중평균하여 기사 단위 점수를 산출할 수도 있다. 감성분석 모델은 KR-FinBERT-SC<sup>i</sup>를 사용하였고, 라벨은 Positive/Negative/Neutral 의 3 분류로 실시하였다. Kss 를 사용하여 문장을 분리하여 문장별로 sentiment 를 계산하는 방식을 적용하였다<sup>ii</sup>. 지수 작성 방식은 base polarity 방식으로 (긍정-부정)/(긍정+부정)으로 계산하여 긍정/부정 극성이 얼마나 다른 빈도로 나타났는지에 대해서만 나타내는 지표와 normalized polarity 방식으로 (긍정-부정)/(긍정+부정+중립)으로 중립을 넣어 전반적으로 normalized 되는 방식 두 가지가 있는데, 본 연구에서는 normalized polarity 로 지수를 작성하였다.

## (3) 검색어 급등 지표(Search Spike)

포털 검색량 데이터를 활용하여 특정 자산 관련 키워드의 관심 급등도를 측정한다. 시점  $t$ 에서 자산  $a$ 에 대한 검색량을  $G_{a,t}$ 라 하고, 직전 4주간 평균 검색량을  $\bar{G}_{a,(t-4:t-1)}$ 라 하면, 검색어 급등 지표는 다음과 같이 정의된다. 값이 클수록 대중 관심이 급증했음을 의미하며, 이는 슐림 리스크 상승 가능성을 시사한다.

$$SearchSpike_{a,t} = \frac{G_{a,t}}{\frac{1}{4} \sum_{j=1}^4 G_{a,t-j}}$$

<sup>i</sup> <https://huggingface.co/snunlp/KR-FinBert-SC>

<sup>ii</sup> <https://github.com/likejazz/korean-sentence-splitter>  
<https://github.com/hyunwoongko/kss>

#### (4) 군집행동 심리 스코어(HSI)

BERT 기반 감성 점수와 검색어 급등 지표를 결합하여 HSI 를 정의한다.

$$HSI_t = \alpha_1 \cdot \overline{SentBERT}_t + \alpha_2 \cdot \overline{SearchSpike}_t$$

여기서  $\overline{SentBERT}_t$ 는 시점  $t$ 의 전체 기사 평균 감성 점수,  $\overline{SearchSpike}_t$ 는 주요 키워드의 평균 검색 급등 지표이며,  $\alpha_i$ 는 학습을 통해 추정된 가중치를 보통 사용하게 되는데, 본 연구에서는 임의로  $\alpha_1 = 0.7$ 을 부과하였다. 검색어 급등 지표에서 발생하는 노이즈들을 일정 정도 조정하기 위함이다. 이렇게 작성된 HSI 가 일정 임계치를 초과하면 자산시장 과열 국면으로 분류하고 쉐림 경보를 발령한다.

## 2. 구체화

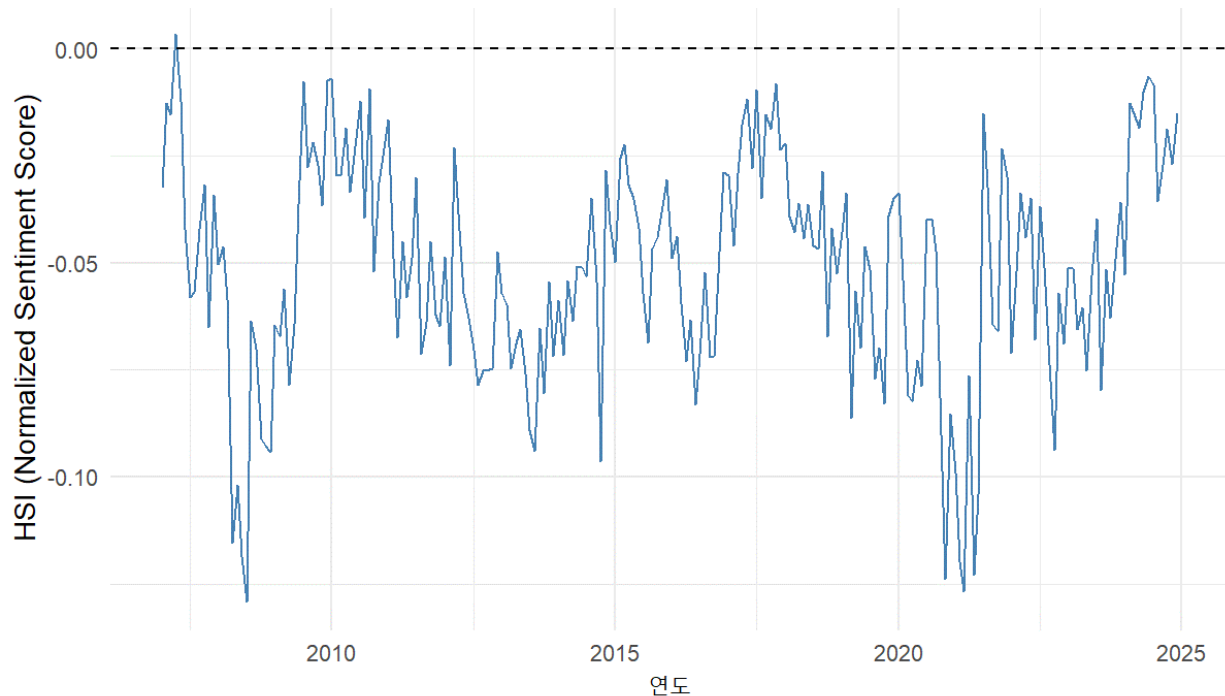
HSI(Herd Sentiment Index)는 텍스트 데이터 감성분석 결과와 검색어 급등 지표 등 여러 투자 심리 척도를 결합하여 산출한 종합 조기경보 지표이다. 구체적으로, 특정 시점  $t$ 에서 뉴스 기사 및 SNS 상의 평균 감성 점수와 자산 관련 검색량 급등도를 가중합하여 계산한다. 이를 수식으로 표현하면 다음과 같다:

$$HSI_t = \alpha_1 \cdot \overline{SentBERT}_t + \alpha_2 \cdot \overline{SearchSpike}_t \quad (2)$$

여기서  $\overline{SentBERT}_t$ 는 시점  $t$ 에 수집된 모든 관련 뉴스 및 소셜미디어 게시물에 대한 BERT 기반 평균 감성 점수를 나타낸다. BERT 모델은 문맥을 고려한 강력한 자연어 처리 모형으로, 금융 텍스트 감성분석에 활용할 경우 사람의 언어 뉘앙스를 잘 포착하는 성능을 보인다(Devlin et al., 2019). 실제로 금융 뉴스에 사전 학습된 변형 모델인 FinBERT 등을 통해 투자자 심리를 계량화하는 연구들이 보고되고 있다. 한편  $\overline{SearchSpike}_t$ 는 여러 주요 자산 관련 검색어들의 급등도를 나타내며, 특정 기간 동안 온라인 검색량이 과거 평균 대비 얼마나 급증했는지를 측정한다. 예를 들어  $G$ 를 시점  $t$ 의 자산  $a$ 에 대한 검색량(정규화된 점수),  $\bar{G}_{a,t-4:t-1}$ 을 직전 4주간 해당 자산 검색량의 평균이라고 하면, 검색어 급등 지표는 대략  $\frac{G_{a,t}}{\bar{G}_{a,t-4:t-1}}$ 와 유사한 형태로 정의된다. 검색량 급등 지표가 큰 양수값을 보이면 해당 자산에 대한 대중 관심이 폭증한 것으로 해석할 수 있으며, 이는 곧 자산시장 쉐림 리스크의 상승 가능성을 시사한다(Oust & Eidjord, 2020). 수식 (2)에서  $\alpha_1, \alpha_2$  각 구성 요소의 가중치로, 과거 데이터 학습을 통해 최적화된다. 예를 들어 과거 여러 자산 버블 및 붕괴 사례(버블

정점 시기 및 급락 시점)에 대해 HSI를 산출한 뒤, 과열 국면을 가장 잘 포착하도록  $\alpha_i$  값을 정할 수 있다. 본 연구에서는  $\alpha_1, \alpha_2$ 의 값을 각 요소를 10~90% 비율로 그리드 검색을 통해 과열 국면을 가장 잘 포착하는 것으로 보이는 7:3의 비율로 최종적으로 HSI를 도출하였다. 한편, 구축된 HSI가 일정 임계치를 초과하면 해당 시기를 자산시장 과열 국면으로 분류하여 “쏟림 경보”를 발령한다. 이 임계치는 과거 버블 사례를 이용한 학습 데이터에서 HSI의 분포를 분석하여 결정할 수 있다. HSI 시계열을 추적하면 시장심리가 중립적 단계에서 과열로 진행되는 변곡점을 포착할 수 있으며, 다른 지표(예: 자금흐름 지표, 신용용자잔고 등)와의 상관관계를 통해 HSI 상승이 실제 자산가격 급등락에 선행하는지 검정할 수 있다.

그림 5. 추정된 HSI 지수



HSI는 군집행동에 따른 투자심리를 실시간 모니터링하기 위한 지표로서, 텍스트 마이닝 기법 중 감성분석에는 BERT 기반 모델을 활용하여 투자 관련 뉴스의 긍정·부정 분위기를 계량화하고, 검색어 급등도를 통해 시장의 관심 쏠림 현상을 포착함으로써 산출된다. 향후 HSI가 급등하여 임계치를 넘길 경우 이는 시장 과열의 조기 경보로 활용될 수 있으며, 금융당국은 이에 대한 선제적 대응을 고려할 수 있다. HSI는

다음 장에서 다루는 위기에측 모형의 핵심 독립변수로 활용되어, 향후 6개월 내 버블 붕괴 가능성을 정량적으로 예측하는 데 기여하게 된다.

## V. 기계학습 예측모형을 통한 위기 발생 확률 추정

### 1. 데이터 및 모형

#### 1.1. 데이터

본 연구에서는 2010년부터 최근까지의 월별 시계열 데이터를 활용하여 자산시장 위기 발생 확률을 예측한다. 분석 대상 기간에는 국내외 주요 금융시장 국면(예: 국지적 버블 형성 및 붕괴)이 포함되어 있으며, 이를 토대로 종속변수를 위기 발생 여부의 이진변수로 구성한다. 구체적으로 과거 자산가격의 버블 형성과 급락 시점을 식별하여, 향후 6개월 내 해당 자산가격이 급락하면 종속변수를 1로 라벨링하고 그렇지 않으면 0으로 설정하였다. 이러한 위기 정의는 Phillips 외(2015) 등의 선행연구를 참고한 것으로, 자산가격 붕괴의 조기경보를 위한 기준을 제공한다. 한편 독립변수로는 투자자 군집행동과 거시환경을 설명할 수 있는 다양한 지표들을 포함하였다.

본 연구에서 구축한 군집심리 지수인 HSI (Herd Sentiment Index)는 뉴스 기사 및 인터넷 검색량에 대한 빅데이터 텍스트마이닝을 통해 산출한 지표로서 시장의 과열 심리를 반영한다. 또한 중앙은행에서 발표하는 금융불안지수(FSI)도 포함하였는데, FSI는 주식·채권·외환시장 변동성, 신용스프레드, 은행건전성 지표 등 다수의 금융변수를 종합해 금융시스템 전반의 스트레스 수준을 나타내는 지수이다. 이 밖에도 신용경색 위험을 보여주는 신용스프레드(예: 회사채 대비 국고채 금리차), 통화정책 기조를 반영하는 단기 기준금리(금리), 시중 유동성 지표(M2 통화량) 등을 주요 거시변수로 포함하였다.

#### 1.2. 모형

위기 발생 확률을 예측하기 위해 세 가지 기계학습 모형을 도입하였다.

(i) 로지스틱 회귀 모형은 통계학에서 널리 활용되는 이진분류 모델로, 사건 발생 확률을 로그 오즈(log-odds)의 선형함수로 표현한다. 본 연구에서는 위기 발생 여부를 확률적으로 예측하기 위한 기본 모형으로 로지스틱 회귀를 활용하였다. 로지스틱 회귀는 설명변수와 위기 발생 간의 선형 관계 및 방향성을 해석하기 쉽고, 기준모형(baseline)으로서 복잡한 기계학습 기법과의 성능 비교를 위한 출발점이 된다. 로지스틱 회귀는 위기 발생 확률을 다음과 같이 추정한다.

$$P(Y_{t+6} = 1|X_t) = \frac{1}{1 + \exp(-(\beta_0 + \sum_{i=1}^k \beta_i X_{i,t}))}$$

여기서  $X_{i,t}$ 는 시점  $t$ 의 독립변수이다.

(ii) 랜덤포레스트 모형은 Breiman(2001)이 제안한 대표적인 앙상블 학습 기법으로서, 다수의 결정트리를 결합해 예측력을 높인 모델이다. 개별 의사결정나무들을 무작위로 샘플링된 데이터셋에 대해 구축하고 그 예측을 평균내는 배깅(Bagging) 기법을 사용함으로써 과적합을 방지하고 안정적인 예측 성능을 얻는다. 또한 트리 생성 시 변수 일부만 무작위 선택하여 분기하는 특징이 있어 변수 중요도 측정과 비선형·복잡한 관계 포착에 강점이 있다. 랜덤포레스트를 도입한 이유는 위기 예측에 영향을 미치는 요인들 간 비선형 상호작용까지 모형에 반영하고자 함이며, 이를 통해 전통적 회귀모형으로는 식별하기 어려운 잠재적 위험 신호를 포착할 수 있을 것으로 기대되기 때문이다. 아울러 랜덤포레스트는 변수 중요도를 제공하므로 어떤 변수가 위기 예측에 특히 중요한지도 분석할 수 있어 해석에 도움을 준다. 랜덤포레스트는 다음과 같이 다수의 결정트리  $f_b(X)$ 를 앙상블하여 예측한다.

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B f_b(X)$$

여기서  $B$ 는 트리 개수이다. 각 트리는 부트스트랩 샘플과 랜덤 변수 선택으로 학습된다.

(iii) LSTM 모형은 순환신경망(RNN)의 한 종류인 Long Short-Term Memory 네트워크로, Hochreiter 와 Schmidhuber(1997)가 제안한 이후 시계열 패턴 학습에 널리 활용되고 있다. LSTM 은 내부에 게이트 구조를 두어 긴 시계열 데이터에서도 장기 의존성(long-term dependency)을 학습할 수 있는 것이 특징이다. 본 연구에서는 위기 발생 징후가 시간의 흐름에 따라 누적되거나 변화하는 양상을 포착하기 위해 LS

TM을 예측모형으로 채택하였다. 즉, 전기간에 걸친 시계열 정보와 변수들의 동태적 상호작용을 모델에 내재화함으로써, 단순 통계모형을 뛰어넘는 비선형 동학적 패턴을 학습시키고자 한 것이다. LSTM은 다음과 같이 시계열 입력  $x_t$ 와 은닉상태  $h_{t-1}$ 를 받아 장기 의존성을 학습한다. 기본 수식은 다음과 같다.

$$\begin{aligned}
 i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i), \\
 f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f), \\
 o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o), \\
 \tilde{c}_t &= \tanh(W_c x_t + U_c h_{t-1} + b_c), \\
 c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \\
 h_t &= o_t \odot \tanh(c_t),
 \end{aligned}$$

이를 통해 과거  $p$ 개월치 시계열로부터 향후 6개월 내 위기 발생 확률을 추정한다.

### 1.3. 추정방법과 예측 정확도 평가

예측모형의 훈련과 추정은 시계열 교차검증 기법과 각 모형별 고유의 학습 알고리즘을 통해 이루어졌다. 먼저 데이터를 훈련셋과 검증셋으로 시계열 분할하여 과거 구간으로 모형을 학습하고 이후 기간에 대한 예측력을 검증하였다. 이는 미래 정보를 사용하지 않도록 시간 순서를 보존한 교차검증으로, 모형의 실제 조기경보 성능을 평가하기 위한 것이다. 로지스틱 회귀모형의 모수 추정에는 최대우도추정법(MLE)을 사용하여, 주어진 데이터에서 위기 발생 여부에 대한 우도(likelihood)를 최대화하는 모형 계수를 추정했다.

한편 랜덤포레스트는 부트스트래핑으로 생성된 다수의 의사결정나무들을 기반으로 학습되는데, 각 트리의 예측오차를 선택하지 않은 자료(OOB, out-of-bag)에 대해 측정함으로써 모형의 복잡도(트리 개수, 최대 깊이 등)와 튜닝 파라미터를 최적화하였다. 이러한 OOB 오류평가는 별도의 검증셋 없이도 내부적으로 모형의 일반화 성능을 추정하게 해주며, Breiman(2001)에서도 랜덤포레스트의 성능을 평가하고 하이퍼파라미터를 선택하는 데 OOB 오류를 활용함을 보여준다.

LSTM과 같은 심층신경망 모형은 역전파(backpropagation) 알고리즘을 통해 학습되는데, 시계열 데이터의 문맥을 고려한 BPTT (Backpropagation Through Time) 방식으로 모형 매개변수(가중치와 바이어스)를 반복적으로 갱신하였다. 즉, 예측오차에

대한 기울기를 과거 시점까지 전파하여 가중치를 조정함으로써, LSTM 이 훈련 데이터의 패턴을 점진적으로 학습하도록 하였다. 세 모형 모두 학습 과정에서 과적합을 방지하기 위해 정규화나 조기종료(early stopping) 등의 기법을 병행하였으며, 최종적으로 검증 데이터에 대한 예측 정확도가 가장 높은 모형 설정을 선택하였다.

모형의 성능 평가는 위기 탐지 능력과 오경보 최소화라는 조기경보 체계의 목적에 비추어 다양한 지표를 사용하였다. 첫째, 위기 발생 여부를 종합적으로 분류하는 능력은 ROC 곡선(Receiver Operating Characteristic Curve) 및 그 곡선하면적(AUC)을 통해 평가한다. ROC 곡선은 위기 판정을 위한 민감도(재현율)와 특이도 간의 상충관계를 시각화한 것으로, AUC 값이 1 에 가까울수록 모형의 분류 능력이 우수함을 의미한다. 둘째, 위기 예측의 정확도를 세부적으로 살펴보기 위해 정밀도와 재현율을 활용한다. 정밀도는 모형이 위기 경보를 발령한 사례 중 실제로 위기가 발생한 비율을 가리키며, 경보 정확도를 측정한다. 반면 재현율은 실제 발생한 위기 중 모형이 사전에 경보를 올린 비율로서, 놓치지 않고 얼마나 위기를 감지하는지를 나타낸다. 일반적으로 정밀도와 재현율은 반비례적 관계에 있으므로 한쪽 지표만 높이는 데 한계가 있다. 따라서 두 지표의 조화평균인 F1 점수를 함께 고려하여 모형의 전반적인 분류 성능을 평가한다. F1 점수는 정밀도와 재현율 사이의 균형을 수치 하나로 표현함으로써, 위기 예측 모형의 종합적 정확도를 나타내는 데 유용한데, 이는 정밀도와 재현율의 조화평균으로 정의된다. 다음은 정확도, 재현율, F1 의 공식을 정리한 것이다.

$$Precision = \frac{TP}{TP+FP} , Recall = \frac{TP}{TP+FN} , F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision+Recall}$$

이를 통해 모형의 조기경보 정확도를 다각도로 검증한다.

셋째, 예측확률과 실제 관측값 간의 오차 크기를 직접 측정하기 위해 평균제곱예측오차(MSFE)도 산출하였다. MSFE 는 모형이 산출한 위기 발생 확률과 실제 결과(0 또는 1) 사이의 차이에 대한 제곱오차의 평균으로, 예측값의 칼리브레이션 정확도를 보여주는 지표이다. 값이 작을수록 모형이 위기 발생 확률을 실제 발생 여부에 가깝게 예측하고 있음을 의미한다. 끝으로, 이러한 다양한 성능지표를 도입한 것은 단순 정확도 하나만으로는 금융위기와 같은 희소 이벤트의 예측력을 충분히 평가하기 어렵기 때문이다. 예를 들어 위기의 발생 빈도가 낮은 상황에서 정확도 지표는 대부분의 기간을 '비위기'로 예측해도 높게 나올 수 있다. 따라서 본 연구는 정밀도, 재현율 처럼 오경보와 놓친 위기 간 트레이드오프를 구체적으로 측정하는 지표와, ROC-AUC

처럼 임계치에 무관하게 모형 성능을 평가하는 지표를 함께 고려하였다. 이러한 평가 방법을 통해 각 예측모형의 조기경보 체계로서의 유용성을 다각도로 검증하고자 한다.

## 2. HSI의 활용: 조기경보 신호 및 상관분석

앞서 구축한 HSI 지표를 활용하여, 시장 과열 여부에 대한 조기경보 신호를 포착하고 기타 금융변수들과의 관계를 분석할 수 있다. 우선 HSI가 일정 임계치를 초과하면 해당 시점을 자산시장 과열(bubble) 국면으로 분류하며, 이는 잠재적 버블 붕괴 위험이 높아졌음을 시사한다. 이러한 임계치는 과거 자산시장 버블 사례의 데이터로부터 최적화할 수 있는데, HSI 수준이 과거 버블 국면에서 보였던 값 이상일 때 경보를 발령하도록 설정하는 것이다. 예를 들어 과거 닷컴버블, 신용카드 사태, 암호자산 급등기 등의 사례에서 HSI 지표를 산출한 뒤, 버블 정점 이전 기간의 HSI 분포를 분석하여 경보 기준치를 결정할 수 있다.

또한 HSI의 선행성을 평가하기 위해, HSI와 자산시장 실물 지표들과의 상관관계를 검정한다. 예를 들어 HSI 상승이 주식시장 가격 변동성 확대나 신용스프레드 상승 등에 선행하는지 Granger 인과관계 검정 등을 통해 확인할 수 있다. 자금흐름(펀드 유입액 등)이나 신용융자잔고와 HSI 간의 동조성을 분석하면, 군집행동 심리가 실제 자금쏠림 현상과 어떻게 연관되는지 파악할 수 있다. 이러한 분석을 통해 HSI가 실제 시장위기(자산 가격 급락)에 얼마나 앞서 신호를 주는지 검증할 수 있으며, HSI와 기타 지표들을 조합한 복합적인 조기경보 체계도 고안해볼 수 있다.

결과적으로 HSI는 단순히 하나의 심리지표에 그치지 않고, 기존의 거시·미시 금융 지표들과 함께 종합적으로 고려함으로써 자산시장 과열 탐지 및 위기 예방적 정책 대응에 유용한 인사이트를 제공한다. 다음 절에서는 HSI를 포함한 다양한 변수를 활용하여 향후 위기 발생 확률을 예측하는 기계학습 모형들을 구축하고 성능을 평가한다.

### 2.1. HSI vs 코스피 및 코스닥 종합지수

HSI와 한국의 대표 주가지수인 KOSPI의 시계열을 비교한다. 그림 6은 그래프는 HSI와 두 주가지수 간에 매우 뚜렷한 정(+)의 상관관계, 즉 강한 경기순응성(pro-cyc

lality)을 보여준다. 주식시장이 상승하는 강세장 국면에서는 HSI 가 높은 수준을 유지하며 함께 상승하고, 시장이 하락하는 약세장 국면에서는 HSI 역시 급격히 하락하는 패턴이 일관되게 관찰된다. 특히, HSI 의 정점과 저점은 주가지수의 주요 변곡점과 거의 동행하는 모습을 보여, 두 변수 간의 동조성이 매우 높음을 시사한다. 이 강력한 상관관계는 단순한 동행 현상을 넘어, 투자자 심리와 시장 수익률 간의 인과관계에 대한 근본적인 질문을 제기한다. 즉, 낙관적인 투자 심리(높은 HSI)가 주가 상승을 견인하는가, 아니면 주가 상승이 투자자들을 낙관적으로 만드는가 하는 '닭과 달걀'의 문제이다.

**그림 6. HSI 와 KOSPI**

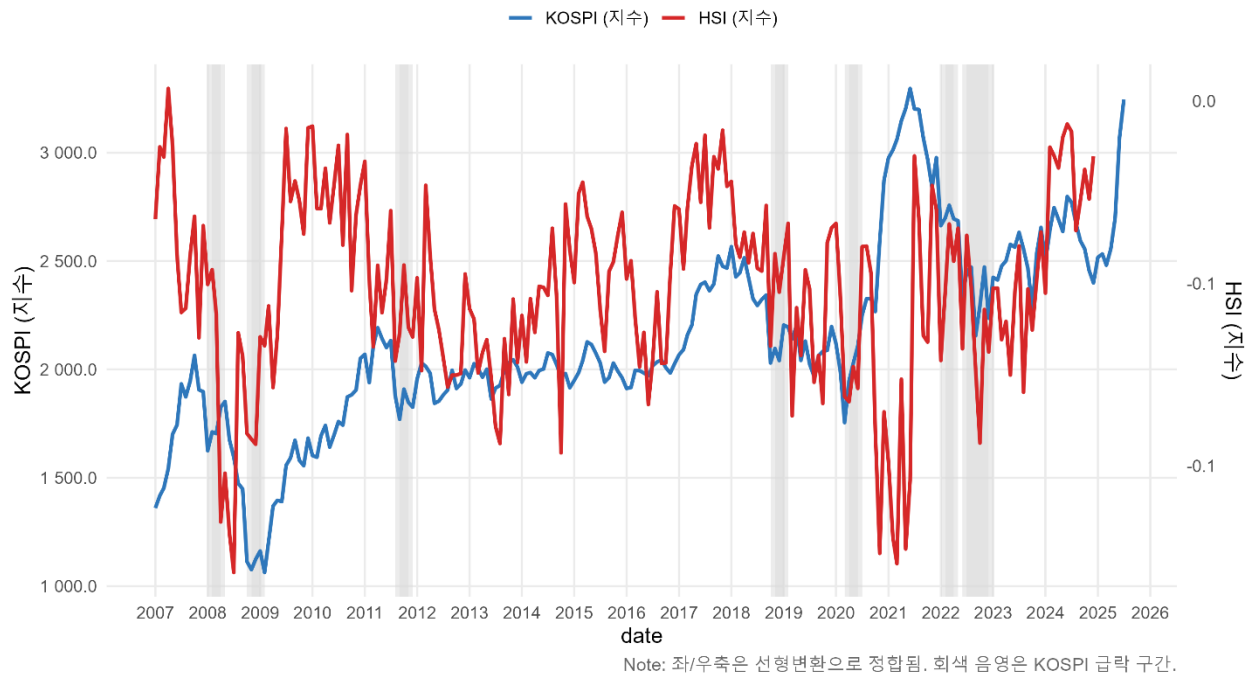


그림 6의 시각적 증거만으로는 이 인과관계를 명확히 규명할 수 없지만, 두 변수 간의 긴밀한 동조성은 강력한 피드백 루프(feedback loop)의 존재를 암시한다. 이 피드백 루프는 다음과 같은 과정으로 작동할 수 있다. 먼저, 긍정적인 외부 충격(예: 예상보다 좋은 경제지표 발표, 기술 혁신 등)이 발생하면 주가가 상승한다. 이 초기 가격 상승은 투자자들의 심리를 개선시켜 HSI 를 끌어올린다. 이렇게 높아진 투자 심리는 더 많은 투자자들을 시장으로 유인하고, 기존 투자자들의 위험 선호 성향을 높여

추가적인 매수세를 유발한다. 이 매수세는 다시 주가를 상승시키고, 상승한 주가는 다시 HSI 를 높이는 자기강화적(self-reinforcing) 순환을 만들어낸다.

이러한 메커니즘은 DeLong et al. (1990)이 제시한 '소음 거래자(noise trader)' 모델의 추세 추종(positive-feedback trading) 전략과 일치하며, Shiller (2015)가 '자연 발생적 폰지 과정(naturally occurring Ponzi processes)'이라고 묘사한 버블 형성의 핵심 동력이다. 즉, HSI 와 주가지수 간의 동행성은 자산시장 버블이 형성되고 증폭되는 엔진을 시각적으로 보여주는 증거로 해석될 수 있다

이러한 강한 상관관계는 HSI 가 시장의 전반적인 '분위기(mood)'나 '열기(temperature)'를 효과적으로 포착하는 지표임을 확인시켜 준다.<sup>2</sup> 정책 당국의 관점에서 볼 때, 지속적으로 높은 수준을 유지하는 HSI 는 단순히 양호한 경제 펀더멘털의 반영이 아니라, 시장이 심리적 과열 상태에 진입하여 잠재적인 심리 반전 충격에 취약해지고 있음을 시사하는 경고 신호로 해석될 수 있다.

다만 팬데믹 기간 동안 주가지수가 급등하는 시기에 HSI 의 경우 오히려 낮아지는 경향을 보이는데, HSI 가 신문기사를 바탕으로 하는 지수이므로 당시에는 투자와 관련된 기사보다는 팬데믹에 대한 기사 비중이 높다보니 HSI 가 투자 심리를 충분히 반영하지 못할 수도 있었음을 시사하는 것으로 보인다. 이는 지수 산정에서 보다 다양한 텍스트 자료를 도입함으로써 극복할 수 있을 것으로 보인다.

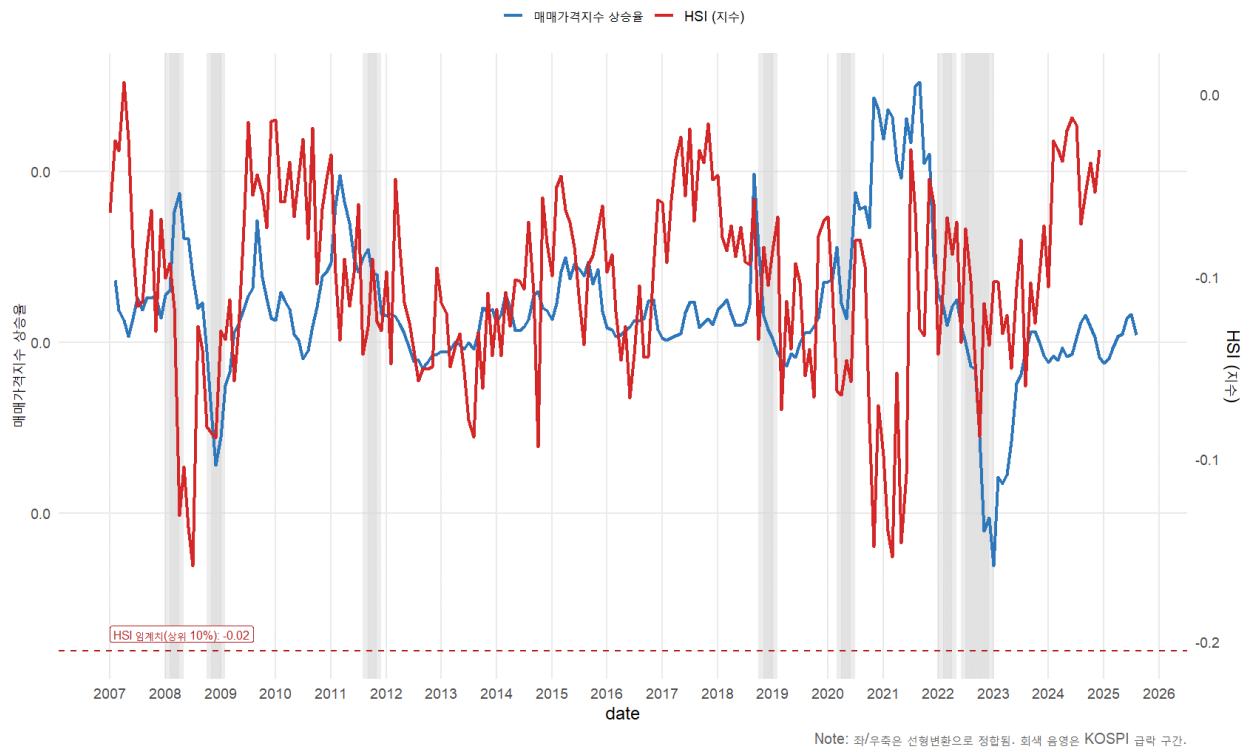
## 2.2. HSI vs 아파트 매매가격지수

그림 7 은 군집행동 심리 지수(HSI)와 주택매매가격의 전기 대비 상승률을 비교하였다. HSI 가 높은 수준을 유지하며 투자 심리가 과열되는 시기에는 주택 가격의 상승 속도 역시 빨라지는 경향이 나타나고, 반대로 HSI 가 급락하며 시장 심리가 위축될 때 주택 가격 상승률은 둔화되거나 하락세로 전환된다. 이는 주식 시장을 중심으로 포착된 투자 심리가 부동산이라는 실물 자산 시장으로 전이되는 '심리 전이' 현상을 명확히 보여주는 증거다.

특히 주목할 점은 HSI 가 주택매매가격 상승률에 대해 선행하는 경향을 보인다는 점이다. 예를 들어, 2021 년 초중반에 HSI 가 먼저 하락하기 시작하자, 주택가격 상승률은 정점에 이르렀고 HSI 가 저점에 도달하니 이후 투자 심리가 하락하면서 주택가격 상승률은 하락하게 되었다. 이와 더불어 HSI 가 다시 상승하는 경향을 보이는데,

이는 HSI 가 투자 과열뿐만 아니라 매도를 하는 시점을 포착하기도 한다는 것을 의미한다. 이러한 시차는 HSI 가 포착하는 광범위한 '위험 선호' 분위기와 주가 상승에 따른 '부의 효과'가 시차를 두고 상대적으로 느리게 움직이는 부동산 시장의 매수세를 자극하는 경로를 시사한다.

**그림 7. HSI 와 아파트매매가격지수 상승률**



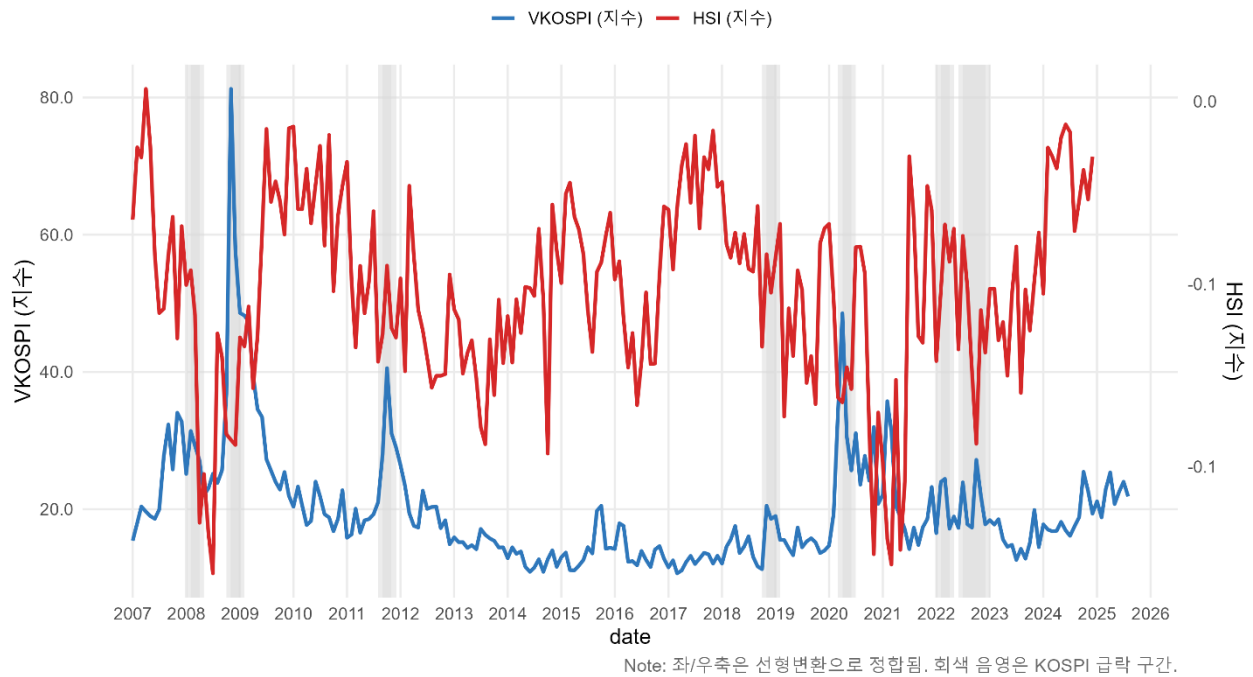
결론적으로, 주택매매가격 상승률과 HSI 의 비교 분석은 HSI 가 단순히 주식 시장에 국한된 지표가 아니라, 자산 시장 전반의 과열 심리를 측정하는 강력한 선행 지표임을 다시 한번 확인시켜 준다. 특히 부동산 시장의 가격 변동 '속도'를 예측하는데 중요한 단서를 제공함으로써, 거시 건전성 정책 수립에 있어 매우 유용한 조기 경보 신호로 활용될 수 있는 잠재력을 보여준다.

### 2.3. HSI vs 주가변동성 (VIX 유사 지표)

그림 8 은 HSI 와 시장의 기대 변동성을 나타내는 vix 유사 지수(이하 vix) 간의 관계를 보여준다. 두 지표 사이에는 매우 명확하고 강력한 역(-)의 상관관계가 존재한

다. HSI 가 높은 수준(낙관 또는 탐욕)에 있을 때, VIX 는 낮은 수준에서 안정적인 모습을 보인다. 반대로, HSI 가 급락할 때(비관 또는 공포) VIX 는 폭발적으로 급등한다. VIX 는 시장 참여자들의 공포 심리를 측정하는 지표로 널리 알려져 '공포 지수(Fear Index)'라는 별칭을 가지고 있다. 이 그래프는 HSI 가 VIX 의 거울 이미지, 즉 '탐욕 지수(Greed Index)'로서 기능함을 명확히 보여준다.

**그림 8. HSI 와 VKOSPI**



이 관계는 행동 재무학의 핵심적인 동학을 반영한다. 낮은 변동성이 지속되는 평온한 시기는 투자자들을 안심시키고 점진적으로 위험에 둔감하게 만든다. 이러한 환경은 과신과 낙관론이 자라나기 좋은 토양이 되며, 이는 HSI 의 상승으로 이어진다. 이것이 바로 '폭풍 전의 고요' 상태이다. 특히 주목할 점은 이 관계의 비대칭성이다. HSI 는 정점에서부터 비교적 완만하게 하락하는 경향을 보이는 반면, VIX 의 급등은 매우 갑작스럽고 폭력적이다. 이는 "강세장은 계단을 오르지만, 약세장은 창문 밖으로 뛰어내린다"는 시장의 오랜 격언을 반영한다.

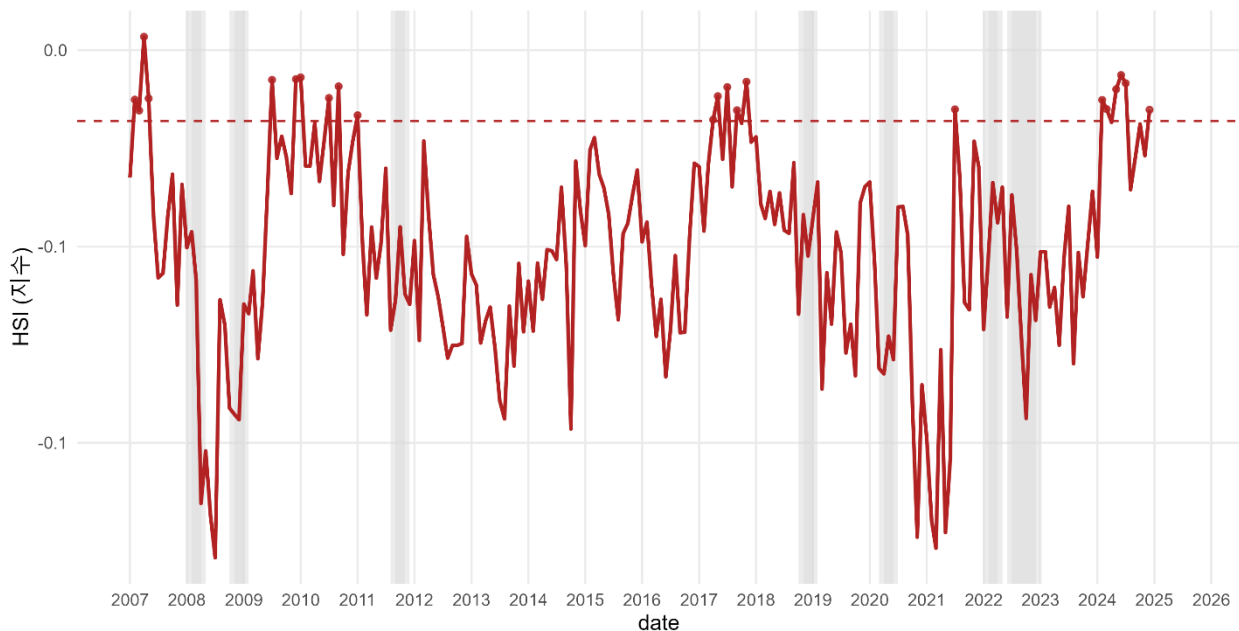
이러한 동학은 심리와 변동성이 서로를 강화하는 피드백 루프를 형성한다.<sup>11</sup> 낮은 변동성은 높은 심리를 조장하고, 이는 레버리지 축적과 같은 불균형을 야기할 수 있다. 이후 특정 충격이 발생하면 심리가 붕괴되고, 이는 다시 변동성을 극단적인 수준으로 끌어올리는 과정으로 이어진다. VIX 가 현재 시장의 '공포' 수준을 측정하는

데 탁월한 지표라면, HSI 는 그 공포에 선행하는 '안일함(complacency)' 또는 '과열'의 정도를 측정하는 데 더 유용할 수 있다. 즉, HSI 는 낮은 변동성 국면에서 축적되는 잠재적 위험을 지속적으로 측정하는 보완적인 도구를 제공한다. 이는 위험 관리 및 금융안정 모니터링 관점에서 HSI 가 가지는 중요한 가치이다

#### 2.4. HSI 임계치 초과 및 주가지수 급락 이벤트 비교

그림 9는 HSI가 사전에 정의된 임계치를 상회하는 구간을 표시하고, 이를 실제 주가지수 급락 이벤트와 비교한다. 시각적으로, HSI가 극단적인 과열 국면에 진입했던 시기들이 이후의 주요 시장 조정 또는 붕괴에 선행하는 경향이 뚜렷하게 나타난다.

그림 9. HSI 임계치 초과시점과 주가지수 급락 비교



주: 회색음영은 KOSPI 급락구간으로 KOSPI 월수익률이 -10%이하인 구간에서 3개월간을 나타낸 것이고, 붉은 점은 HSI 임계치(상위 10%)초과 시점을 나타낸 것이다.

이 그림은 다음장에 제시된 복잡한 기계학습 모형의 예측 결과를 직관적으로 검증하는 역할을 한다. 특히 LSTM 모형은 시계열 데이터의 장기 의존성과 패턴을 학습하는 데 특화된 딥러닝 모델로, HSI 를 위기 발생의 가장 중요한 단일 예측 변수로

식별했다. 이 그래프는 그 이유를 시각적으로 설명해준다. 즉, 시장 붕괴 전에는 미디어와 대중의 관심이 폭증하며 극단적인 흥분 상태(매우 높은 HSI)가 나타나고, 이로 인해 시장이 펀더멘털을 초과하여 과도하게 확장된 후 급격한 조정을 겪는 패턴이 반복적으로 관찰된다.

이는 버블의 생성과 붕괴 동학의 전형적인 모습이다. HSI 는 '비이성적 과열' 국면을 포착하고, 이후의 시장 급락은 그 필연적인 결과인 '붕괴(bust)' 국면에 해당한다.

이 그래프는 HSI 가 조기 경보 지표로서 상당한 잠재력을 가지고 있음을 가장 직접적으로 보여준다. 물론 완벽한 예측 변수는 아니지만, HSI 가 임계치를 상회하는 것은 시장 위험이 펀더멘털 요인에서 벗어나 투기적인 군집행동에 의해 주도되고 있다는 강력한 신호로 해석되어야 한다. 이는 정책 당국과 규제 기관에게 시장의 과열을 식히기 위한 경기 대응적 거시건전성 조치의 시행을 고려해야 할 시점임을 알리는 중요한 경고등 역할을 할 수 있다.

### 3. 위기 발생 확률 예측을 위한 기계학습 모형의 실증분석

이 절에서는 자산시장 위기(버블 붕괴)의 발생 여부를 사전에 예측하기 위한 세 가지 기계학습 모형을 정의한다. 로지스틱 회귀는 선형 결합에 기반한 확률 예측 모형이고, 랜덤포레스트는 다수의 결정나무를 앙상블하여 비선형 관계도 포착할 수 있는 모형이며, LSTM 은 시계열 데이터의 패턴을 학습하는 딥러닝 순환신경망 모형이다. 각 모형의 수학적 구조와 특징을 간략히 살펴보고, 본 연구에의 적용 방법을 정의한다. 한편 위기 여부에 대한 종속변수  $Y_t$ 는 과거 자산가격 버블 및 급락 발생 여부를 이진값(0/1)으로 라벨링하여 구성한다. 구체적으로 "시점  $t$ 로부터 6 개월 내에 자산가격의 급격한 폭락 발생"을  $Y_t = 1$ 로 정의하고, 그렇지 않으면  $Y_t = 0$ 으로 둔다(Phillips et al., 2015). 독립변수  $X_t$ 에는 앞서 구축한 HSI 를 비롯하여 자금시장 쏠림 관련 지표들(예: 자금흐름 등)과 거시변수들(예: 금리, 신용스프레드, 유동성(M2) 등)을 포함한다. 이제 세 가지 예측 모형을 차례로 정의한다.

#### 3.1. 위기 발생확률 모형의 예측 결과

군집행동(herd behavior)은 투자자들이 자신의 정보를 무시하고 다수의 행동을 추종함으로써 발생하는 현상으로, 정보전염(informational cascade)이나 평판 동기 등

다양한 요인에 의해 촉발된다(Banerjee, 1992; Scharfstein & Stein, 1990). 이러한 쓸림현상은 한 번 시작되면 쉽게 멈추지 않고 시장과열과 버블 붕괴로까지 이어져 금융불안정을 증폭시킬 수 있다. 실제로 Spyrou(2013)는 군집행동이 시장의 극단적 변동성과 위기의 한 원인임을 여러 선행연구를 통해 지적하면서, 이러한 군집심리의 쓸림을 조기에 감지하여 경고하는 시스템의 필요성을 강조하였다. 본 절에서는 이러한 맥락에서 조기경보 예측모형을 구축하고 성능을 비교하였다. 실증분석에는 로지스틱 회귀, 랜덤포레스트, LSTM 의 세 가지 예측모형을 활용하여 자산시장 위기 발생 여부를 사전에 예측하였다. 각 모형의 예측 성능은 <표 4>에 요약되어 있으며, 성능 지표로 평균제곱예측오차(MSFE)를 사용하였다. 그 결과 로지스틱 회귀의 MSFE 는 0.257, LSTM 은 0.326 으로 나타난 반면, 랜덤포레스트는 0.135 로 가장 낮은 오차를 보여주었다.

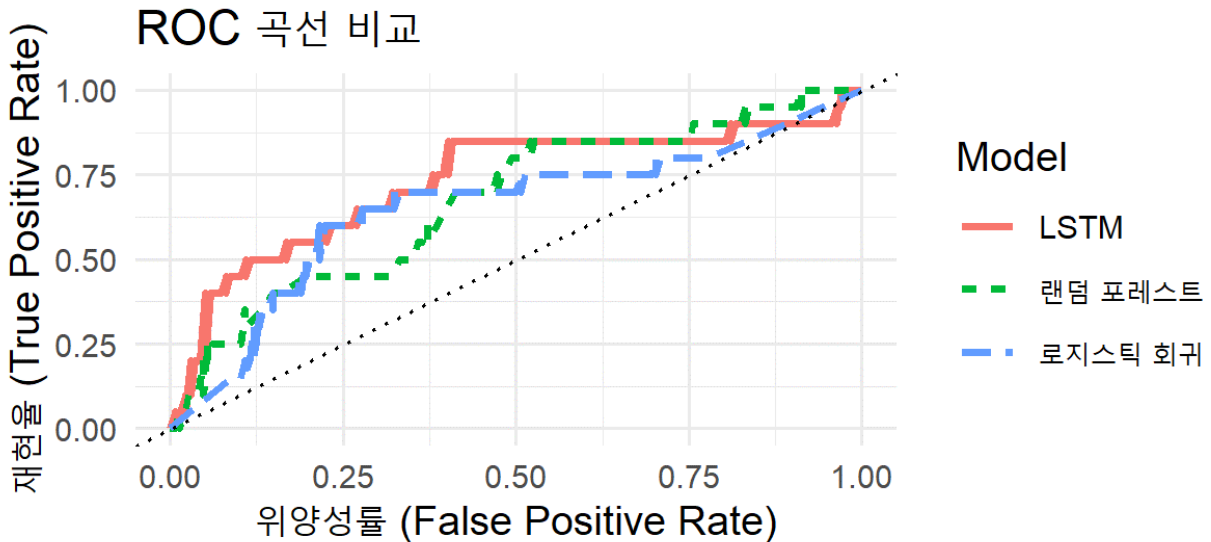
**표 4. 위기 발생확률 모형들의 MSFE**

| <b>Model</b>         | <b>MSFE</b> |
|----------------------|-------------|
| <b>Random Forest</b> | 0.135       |
| <b>Logistic Reg.</b> | 0.257       |
| <b>LSTM</b>          | 0.326       |

로지스틱 회귀(Logistic Regression), 랜덤포레스트(Random Forest), LSTM 모형을 동일한 데이터에 적용한 결과, 각 모형의 성능지표에서 상이한 특징이 관찰되었다. 정량적 예측 정확도를 나타내는 MSFE(평균제곱예측오차) 기준으로는 랜덤포레스트가 0.135 로 가장 우수하였고, 로지스틱 회귀(0.257)와 LSTM(0.326)은 그보다 큰 오차를 보여 정량적 예측력이 다소 떨어지는 것으로 나타났다. 반면 그림 10 을 보면 분류 성능을 나타내는 AUC(ROC 곡선 아래 면적)의 경우 LSTM 모형의 값이 가장 높게 나타나 위기 발생 여부를 식별해내는 능력에서 LSTM 이 다른 모형을 앞서는 것으로 평가된다. 이는 랜덤포레스트가 예측값 자체를 정확히 맞추는 데는 강점을 보인 반면, LSTM 은 위기 사례 대 비위기 사례를 구분하는 순위화 능력에서 뛰어난 의미를 가진다. 다시 말해 한 모형의 정량적 정확도가 높다고 해서 분류 능력도 항상 우수한 것은 아니며, 본 연구 결과에서도 MSFE 와 AUC 지표에서 모형별 우열이 엇갈려 나타났다. 이러한 결과는 정량적 예측성과와 분류 능력이 상호 보완적인 측면이

있음을 시사하며, 조기경보 모델을 평가할 때 단일 지표에만 의존하지 말고 여러 성능지표를 종합적으로 고려해야 함을 보여준다.

그림 10. 세 모형들의 ROC 비교



다만 모형 해석력 측면에서 보면, 로지스틱 회귀의 장점도 분명하다. 로지스틱 회귀는 추정된 계수의 부호와 크기를 통해 각 독립변수가 위기 발생 확률에 미치는 영향을 직접적으로 파악할 수 있어 해석이 용이하다.

조기경보 체계(Early Warning System) 구축 관점에서 몇 가지 정책적 시사점을 도출할 수 있다. 첫째, 조기경보 모델을 평가할 때는 정량 예측 정확도와 분류 능력 간 균형에 주목해야 한다. 위기 발생 가능성을 정확한 수치로 예측하는 것도 중요하지만, 궁극적으로 정책당국에 필요한 것은 위기 징후를 놓치지 않고 포착하는 능력이다. 랜덤포레스트의 낮은 MSFE는 향후 위험 수준을 정량적으로 정밀하게 추계하는데 유용하며, 이를 통해 경보 단계별 대응조치를 세분화하는 등 정교한 사전대응이 가능해진다. 반면 LSTM의 높은 AUC는 실제 위기가 발생할 사례를 조기에 식별해내는 데 유리하여, 위기 신호를 조금이라도 앞서 포착하고 경보를 발령할 민감도를 높이는 데 기여할 수 있다. 따라서 정책당국은 한 가지 지표나 모형에만 의존하기보다, 정량적 지표와 분류 지표를 모두 고려한 다각도의 성능 평가를 통해 모형을 선택해야 한다. 예컨대, 위험도 지수 예측에는 랜덤포레스트의 결과를 참고하되, 경보 발령

여부 판단에는 LSTM의 신호를 동시에 활용하는 식의 통합적 접근도 고려할 수 있다. 이는 상호 보완적인 모델 특성을 활용하여 오경보(false alarm)와 경보누락(missed alarm)을 동시에 줄이는 전략이 될 것이다.

둘째, 조기경보 모형의 해석가능성과 수용성도 정책적으로 중요하다. Jarmulska(2020)이 지적한대로, 로지스틱 회귀처럼 모형 구조가 투명한 경우에는 경보 발령의 근거를 이해관계자들에게 쉽게 설명할 수 있고, 변수별 기여도를 공개함으로써 정책 대응의 정당성을 확보하기 쉽다. 반대로 LSTM이나 랜덤포레스트는 예측력이 높더라도 결과를 설명하기 어려운 블랙박스 모델로 인식될 수 있으므로, 금융당국이 이를 채택할 때는 결과 해석을 보완하는 장치가 필요하다. 예를 들어 랜덤포레스트의 경우 변수중요도 지표나 SHAP 가치 등을 제시하여 어떤 요인이 경보에 영향을 미쳤는지 밝혀줄 수 있고, LSTM의 경우에도 시계열 모형 특성을 고려한 주의해야 할 지표의 추이를 함께 제공하는 방안을 생각해볼 수 있다. 이러한 해석 보완 장치는 조기경보에 대한 신뢰도 제고와 오류 발생 시 책임 소재 규명에도 기여하여, 정책당국이 기계학습 기반 경보체계를 안정적으로 운용하는 데 도움을 줄 것이다.

셋째, 모형 간 특성에 따른 활용 전략이 요구된다. 본 분석에서 세 모형은 각기 장단점을 보였으므로, 조기경보 체계에 단일 모형을 독점적으로 사용하는 것은 바람직하지 않을 수 있다. 대신 복수의 모형을 병행 활용하여 서로 보완토록 하면 경보체계의 견고성을 높일 수 있다. 예를 들어 1차적으로는 해석이 용이한 로지스틱 회귀나 랜덤포레스트로 기본 경보 신호를 생성하고, 동시에 LSTM으로 보다 복잡한 패턴을 감지하여 추가적인 경보를 보조적으로 운용하는 방안을 생각해볼 수 있다. 이렇게 하면 한 모형의 약점을 다른 모형이 보완하면서 전체적인 예측 성능과 신뢰도를 높일 수 있다. 실제로 최근 연구들에서도 전통 통계모형과 기계학습모형을 결합한 하이브리드 조기경보체계를 제안하는 등, 여러 기법의 장점을 동시에 활용하려는 노력이 나타나고 있다. 궁극적으로 조기경보 시스템의 목적은 완벽한 예측이 아니라 위험징후에 대한 사전 경고이므로, 다소의 오차가 있더라도 미리 경고를 발하여 대비할 수 있게 하는 것이 중요하다. 다양한 모형의 결과를 종합적으로 고려하는 접근은 이러한 사전 경고 기능을 최대화하는 데 유리하며, 정책당국이 위기 대응에 필요한 “골든타임”을 확보하는 데 기여할 것이다.

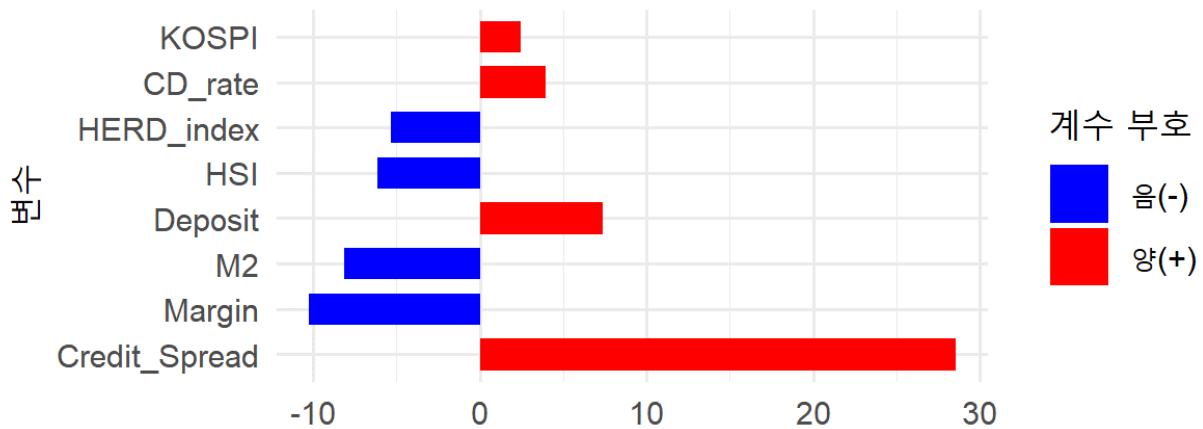
### 3.2. 위기 발생확률모형의 해석

앞서 5.3.1 절의 예측 성능 비교 결과를 통해 로지스틱 회귀, 랜덤 포레스트, LSTM 등 세 모형이 위기 발생 여부를 예측하는 방식에 차이가 있음을 확인하였다. 그러나 이러한 예측 결과만으로는 각 모형이 어떤 변수에 민감하게 반응하고, 어떤 요인을 주요 위험 신호로 간주하는지까지는 충분히 파악하기 어렵다. 이에 따라 본 절에서는 세 모형 각각에 대해 변수 중요도 분석을 수행하고자 한다. 이를 통해 자산시장 위기 예측 과정에서 각 모형이 핵심적으로 활용하는 변수가 무엇인지 밝히고, 이러한 변수 중요도의 차이를 모형 간에 비교·분석하고자 한다.

#### (1) 로지스틱 회귀 분석 결과

전통적인 선형 모델인 로지스틱 회귀 분석에서는 Credit\_Spread(신용스프레드)가 위기 발생 확률을 설명하는 가장 강력한 변수로 나타났다. 이는 기업의 신용 위험 증가가 금융 시장의 불안정성을 직접적으로 반영한다는 금융 이론과 일치하는 결과이다. 그 뒤를 이어 CD\_rate(CD 금리)와 Margin(신용융자잔고)이 유의미한 변수로 식별되었으며, 이는 거시경제 환경의 금리 수준과 시장 내 레버리지(leverage)가 위기 예측의 주요 요인임을 시사한다. 반면, 군집행동 지수인 HERD\_index와 HSI의 영향력은 상대적으로 미미하게 나타나, 선형 모델에서는 이들 변수가 가진 복잡한 관계를 효과적으로 포착하지 못하는 것으로 해석된다.

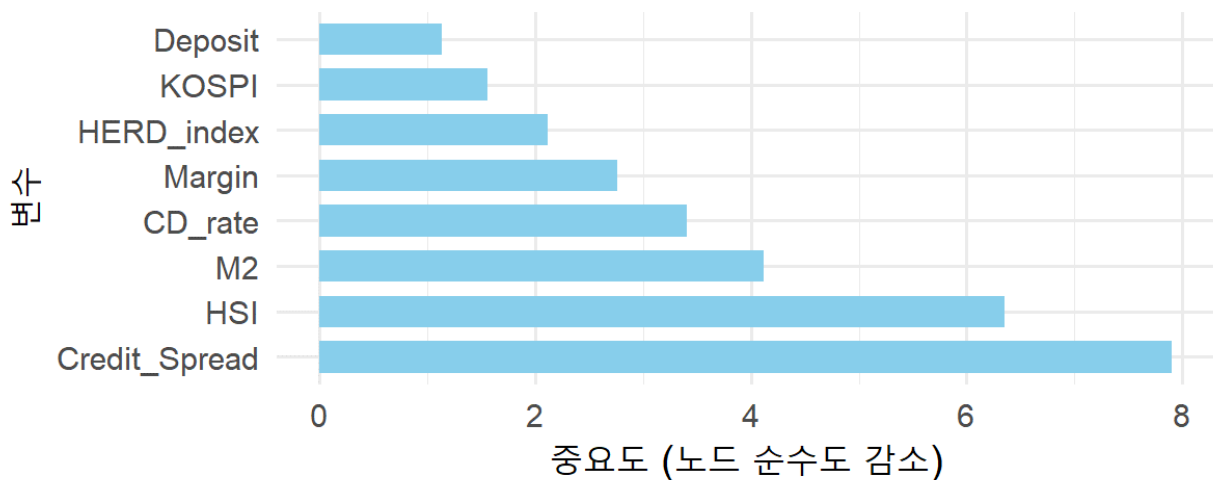
그림 11. 로지스틱 회귀 모형의 계수 중요도



## (2) 랜덤 포레스트 분석 결과

비선형 및 변수 간 상호작용 포착에 강점이 있는 랜덤 포레스트 모델에서도 Credit\_Spread 는 여전히 가장 높은 중요도를 기록하며, 해당 변수의 안정적인 예측력을 재확인시켜 주었다. 하지만 로지스틱 회귀 분석과 가장 두드러지는 차이점은 BERT 기반의 군집행동 심리 스코어인 HSI 가 두 번째로 중요한 변수로 부상한 점이다. 이는 시장 참여자들의 심리 상태가 위기 발생과 직접적인 선형 관계는 아닐지라도, 복합적인 비선형 관계 속에서 매우 높은 설명력을 가짐을 의미한다. 또한, M2(통화량) 변수의 중요도 역시 로지스틱 회귀 모델에 비해 유의미하게 상승하여, 시장 유동성의 비선형적 영향력을 시사했다.

그림 12. 랜덤 포레스트 변수 중요도



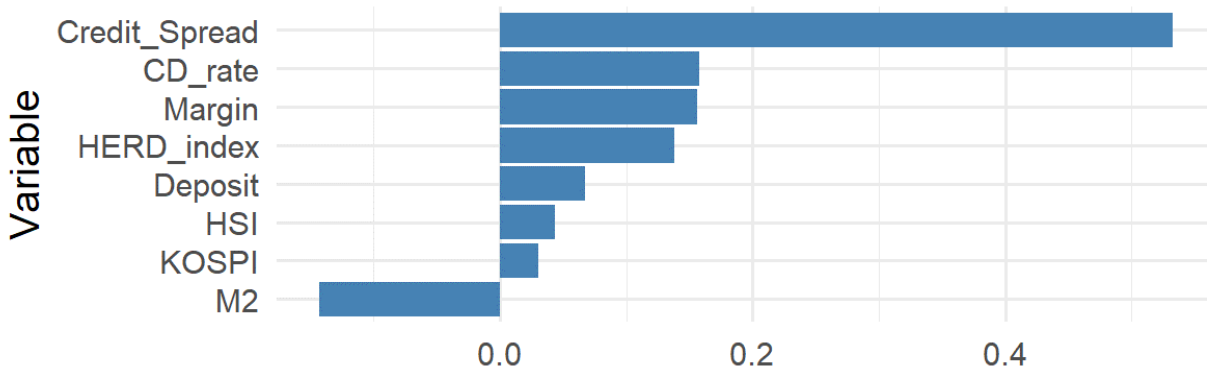
## (3) LSTM 변수 중요도 분석 결과

딥러닝 모형인 LSTM 은 랜덤포레스트와 달리 내재된 변수 중요도 계산 기능을 갖고 있지 않다. 이는 신경망의 학습 구조가 복잡하고, 파라미터가 직접적인 해석을 제공하지 않기 때문이다. 따라서 LSTM 에서 변수별 중요도를 추정하려면 외부적인 접근 방법이 필요하다. 본 연구에서는 이러한 목적을 위해 Permutation Importance 방식에 기반한 평가 절차를 활용하였다. 이 방법은 각 변수의 값을 무작위로 섞은(permuted) 후, 해당 변형이 예측 성능에 미치는 영향을 정량화함으로써 변수 중요도를 간접적으로 추정하는 방식이다.

구체적인 절차는 다음과 같다. 먼저, 학습된 LSTM 모형에 대해 테스트셋을 입력하여 예측값을 산출하고, 이를 기반으로 기준 예측 오차(baseline MSFE)를 계산한다. 이후 각 독립변수에 대해 개별적으로 순열(permutation) 처리를 수행하는데, 이는 특정 변수의 입력값을 샘플 순서에 무관하게 무작위로 섞는 방식이다. 이때 다른 변수들은 그대로 유지하고 하나의 변수만 무작위로 교란시킨다. 이렇게 변형된 입력 데이터셋을 동일한 LSTM 모델에 입력하여 다시 예측값을 산출하고, 그에 따른 변형 후 MSFE 를 계산한다. 해당 변수의 값이 예측에 중요했다면, 값을 섞음으로 인해 예측력이 저하되고 MSFE 가 증가하게 된다.

변수별 중요도는 “(변형 후 MSFE - 기준 MSFE) / 기준 MSFE”로 정의되며, 이는 해당 변수를 무작위화했을 때 예측 성능이 얼마나 감소하는지를 비율로 나타낸 값이다. 이 점수를 통해 각 변수의 상대적 중요도를 정량적으로 비교할 수 있으며, 값이 클수록 해당 변수가 예측 정확도에 큰 기여를 한다는 해석이 가능하다. 마지막으로 이러한 중요도 점수는 내림차순으로 정렬되어 막대 그래프 등의 시각화 결과로 제시되며, 예측 모형이 어떤 변수를 핵심적으로 활용했는지를 설명하는 데 활용된다. Permutation Importance 는 모형의 구조나 학습 알고리즘에 의존하지 않고 예측 결과만을 기반으로 평가하므로, 모형-불가지론적(model-agnostic)이며 LSTM 뿐 아니라 다양한 예측모형에 적용 가능하다.

**그림 13. LSTM 모형의 Permutation Importance**



Permutation Importance 를 기반으로 LSTM 모델에서는 Credit\_Spread(신용스프레드)가 가장 중요한 예측 변수로 나타났다. 이는 금융 위기를 예측하는 데 있어, 기업의 신용 위험 확대와 같은 전통적인 금융 스트레스 지표가 여전히 가장 결정적인 선행 정보를 포함하고 있음을 시사한다. Credit\_Spread 의 뒤를 이어 CD\_rate(CD 금리)와 Margin(신용융자잔고) 변수가 높은 중요도를 보여, 금리 수준과 시장 내 레버리

지 요인이 위기 예측에 중요한 역할을 함을 뒷받침했다. 반면, HSI는 상대적으로 낮은 중요도를 기록했는데, 이는 LSTM 모델이 텍스트 기반의 심리 지표보다는 정량적 금융 변수들의 시계열적 패턴과 변동성 학습에 더 민감하게 반응했기 때문으로 해석된다.

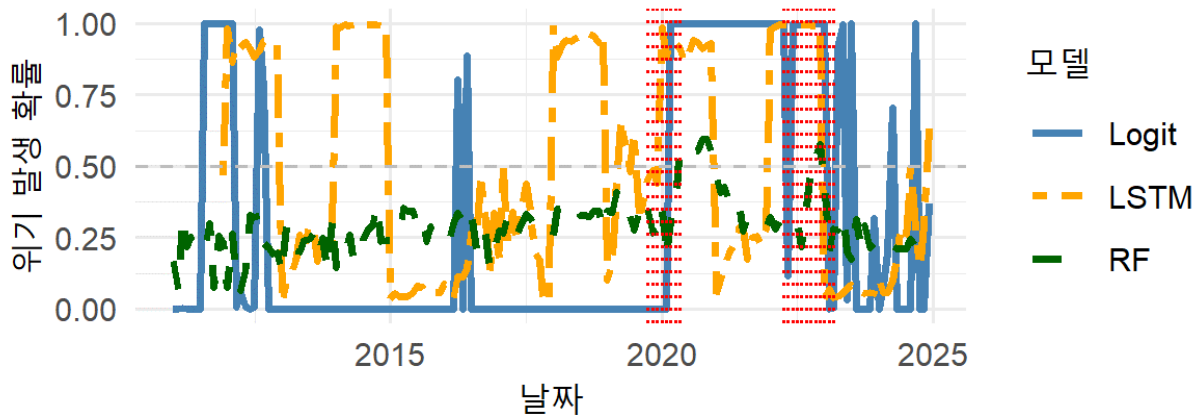
세 가지 예측 모형(로지스틱 회귀, 랜덤포레스트, LSTM)의 변수 중요도 분석을 종합한 결과, 신용스프레드(Credit Spread)가 모든 모형에서 공통적으로 위기 예측에 가장 핵심적인 역할을 하는 것으로 나타났다. 로지스틱 회귀에서는 신용스프레드의 확대가 향후 6개월 내 자산가격 급락 가능성을 유의하게 증가시키는 가장 강력한 양(+)의 변수로 확인되었으며, 반면 HSI와 HERD\_index는 음(-)의 계수를 보여 위기 확률과 역의 관계를 나타내거나 그 영향이 제한적이었다. 한편 랜덤포레스트 모형의 변수 중요도에서는 신용스프레드가 1위를 차지하였고, HSI가 그 뒤를 이어 두 번째로 높은 비중을 차지했다. 이는 비선형 모형에서 시장심리가 일정 임계치를 넘을 경우 HSI가 위기 예측에 중요한 정보를 제공함을 시사한다. LSTM 딥러닝 모형 역시 신용스프레드를 압도적인 1위 예측 인자로 활용하였으나, HSI의 중요도는 상대적으로 낮게 평가하여 정량적 금융 지표의 시계열 추세를 더 중시하는 경향을 보였다. 이처럼 세 모형은 신용스프레드라는 공통된 핵심 지표를 강조하면서도, 로지스틱 회귀는 선형적 위험 증가를, 랜덤포레스트는 심리지표(HSI)와의 비선형적 상호작용을, LSTM은 금융 지표의 동태적 패턴을 포착한다는 점에서 상호 보완적인 특성을 보였다.

이상의 결과는 자산시장 조기경보체계 및 감독정책 측면에서 중요한 시사점을 제공한다. 우선, 세 모형에서 공통적으로 가장 중요하게 지목된 신용스프레드(Credit Spread)를 평시에 최우선으로 면밀히 모니터링하는 것이 필요하다. 회사채-국채 금리 차로 측정되는 신용스프레드의 급격한 확대는 금융시스템 내 스트레스 누적을 나타내는 가장 확실한 경고 신호이기 때문이다. 또한, 랜덤포레스트 모형에서 입증된 HSI의 유용성을 고려하여, 뉴스·검색 데이터 기반의 군집심리 지표를 보조적인 모니터링 지표로 활용할 필요가 있다. 따라서 중앙은행 및 금융당국은 신용스프레드 확대를 위기 임박의 핵심 징후로 삼되, HSI가 비정상적으로 높아지는 현상을 시장 과열의 선행 신호로 해석하여 "쏟림경보" 발령 등의 선제 조치를 단계적으로 고려해야 한다. 나아가 각 모형이 포착하는 위험의 성격이 다르므로, 단일 지표에 의존하기보다 신용지표와 심리지표를 종합적으로 판단하는 다각적인 대응 체계를 구축함으로써 자산시장 불안정을 조기에 억제하고 잠재적 위험을 완화할 수 있을 것으로 기대된다.

#### (4) 예측 확률 비교

그림 14 는 앞서 언급한 세 모형들의 위기 예측확률을 2011 년 후에 실제 적용해 본 것이다. 빨간 점선은 실제 위기를 나타내는 것으로 앞서 정의한 대로 최근 6 개월 이내의 자산 가격의 폭락한 경우를 나타낸 것이다. 세 모형 모두 실제 위기 발생 시점(2020 년, 2022 년경 등)을 전후하여 위기 예측 확률을 1.00 에 가깝게 높이는 등, 위기 예측에 대한 우수한 성능을 보였다. RF 모형(녹색 점선)은 위기가 아닌 평상시에는 예측 확률을 0 에 가깝게 낮게 유지하며, 실제 위기 시점에만 확실히 높은 확률을 보이는 경향이 있다. 이는 RF 가 오작동(False Positive) 가능성이 가장 낮아 정량적 예측 정확도가 높음을 시사한다. Logit(파란색 실선)과 LSTM(주황색 파선) 모형은 실제 위기 시점에는 높은 확률을 보이지만, 위기가 아닌 시점에도 예측 확률이 0.5 이상으로 치솟거나 오작동하는 경우가 랜덤 포레스트보다 빈번하게 관찰된다.

그림 14. 위기 예측모형의 예측 확률 비교



주: 빨간 점선은 실제 위기 발생 시점을 나타냄

세 모형은 각기 다른 장단점과 예측 패턴을 보이므로, 단일 모형에 의존하는 것은 바람직하지 않다. 랜덤 포레스트는 오차가 가장 낮고(MSFE 0.135) 평상시 위험도를 낮게 유지하여, 향후 위험 수준을 정량적으로 정밀하게 추계하고 경보 단계별 대응 조치를 세분화하는 데 유용하다. Logit 이나 LSTM 은 평상시 오작동 가능성은 높지만, 위기 상황에 대한 경고를 민감하게 울리는 데 그 의미를 둘 수 있다. 특히 LSTM 은 위기 사례를 식별해내는 분류 능력(AUC)이 가장 높게 나타났다. 따라서 정책당국은 랜덤 포레스트의 결과를 주된 위험도 지표로 참고하되, LSTM 의 신호를 보조적으로 활용하여 경보 발령의 민감도를 높이는 통합적 접근을 고려해야 한다.

조기 경보 시스템의 궁극적인 목적은 위험 징후에 대한 사전 경고를 통해 대비할 시간(골든 타임)을 확보하는 것이다. Logit 이나 LSTM 처럼 오작동 가능성이 있더라도 위기 상황에 대한 경고를 신속히 하는 것을 정책적으로 추구한다면, 오경보를 감수 하고서라도 경보 누락을 최소화하는 전략을 택해야 한다. 다만 오경보가 지나치게 잦으면 시장의 경보 신뢰도가 저하될 수 있으므로, 경보 발령 시에는 모형 해석의 투명성을 높여 이해관계자들에게 근거를 명확히 설명하는 보완 장치가 필수적이다. 한편 랜덤 포레스트 및 LSTM 과 같은 비선형 모형에서 군집심리 지표(HSI)가 전통적인 신용스프레드에 준하는 주요 예측 변수로 식별되었으므로, 빅데이터 기반의 심리 지표 모니터링은 금융안정 정책에 있어 새로운 보완적 접근을 제공한다. 이는 전통적인 거시 지표가 포착하지 못하는 시장 내부의 동조화 및 과열 현상을 조기에 파악 하는 데 기여한다

## VI. 결론

본 연구는 한국 금융자산시장에서 고질적인 취약 요인으로 지목되어 온 ‘쏟림현상’이 금융안정 및 시스템 리스크에 미치는 동태적 영향을 분석하고, 특히 빅데이터 기반의 심리 지표를 활용한 조기경보체계 구축 가능성을 탐색하는 데 그 목적이 있다.

쏟림현상은 투자자의 군집행동 (개인/기관 투자자가 사적 정보를 무시하고 타인의 결정을 모방)과 금융기관의 자금 쏟림 (특정 부문으로 경쟁적으로 자금 공급)이라는 두 가지 유형으로 구분하여 정의되었으며, 바이코리아 펀드 열풍, 신용카드 사태, 주택담보대출 쏟림 등 한국의 역사적 금융위기 사례를 통해 그 파괴적인 잠재력이 확인되었다. 이러한 쏟림현상의 미시적 동인으로는 불완전한 정보 하의 합리적 모방을 설명하는 정보 폭포(Informational Cascades) 이론, 평판 및 비대칭적 보상구조로 인한 대리인 문제, 그리고 과신이나 가용성 휴리스틱 같은 행동주의적 편향 등이 작용함이 문헌을 통해 확인되었다. 특히 금융기관 수준의 자금 쏟림은 시스템 리스크를 직접적으로 증폭시켜, 특정 부문 충격이 시스템 전반으로 확산되는 도미노 효과를 초래할 수 있다.

본 연구는 쏟림현상을 종합적으로 분석하기 위해 자본시장 군집행동(HERD)과 신용시장 쏟림(CONC)을 포괄하는 이원적 분석 체계를 수립하고, 벡터 자기회귀(VAR) 모형과 기계학습 예측 모형을 활용하여 다음과 같은 핵심 결과를 도출하였다. VAR

모형의 충격반응 분석 결과, 투자자들의 군집심리가 갑작스럽게 강화되는 충격은 직접적으로 금융불안지수(FSI)의 반응을 유발하지는 않는다. 다만 군집행동의 심화가 변동성이 높아져 추후 금융불안을 유발할 원인이 될 수 있음을 보였다. 통화정책(기준금리) 긴축 충격은 자산시장 변동성(VKOSPI)에 단기적인 진정 효과를 보였으나, 중기적으로는 오히려 금융불안(FSI)을 증가시키는 역 U 자형 경로를 나타냈다. 무엇보다 통화정책 충격이 군집심리 지수(dHERD)나 신용쏠림(dCONC/dREG2)에는 유의한 영향을 미치지 못하는 것으로 나타나, 전통적 금리 정책 수단이 투자자 심리나 자금쏠림 구조를 직접적으로 변화시키는 데는 한계가 있음을 시사한다.

BERT 기반 감성분석과 검색어 급등 지표를 결합하여 구축한 군집행동 심리 스코어(HSI)는 시장 과열 심리를 효과적으로 포착하는 조기경보 지표로서의 잠재력을 확인했다. HSI는 KOSPI 및 아파트 매매가격지수 상승률과 강력한 정(+)의 동조성을 보이며, 주가지수의 주요 변곡점과 거의 동행하는 자기강화적 순환(self-reinforcing loop)을 시각적으로 보여주었다. 또한, HSI는 시장 변동성 지수(VIX 유사 지표)와 강력한 역(-)의 상관관계를 보여, '탐욕 지수'로서 평온한 국면에서 축적되는 잠재적 위험(안일함)을 측정하는 데 유용함을 입증했다. HSI가 임계치(상위 10%)를 초과했던 시점들은 이후의 주요 주가지수 급락 이벤트에 선행하는 경향이 뚜렷하게 나타났다.

향후 6개월 내 자산가격 급락(위기)을 예측하는 기계학습 모형 비교 결과, 랜덤포레스트 모형이 MSFE(0.135) 기준 가장 우수한 정량적 예측 정확도를 보였으며, LSTM 모형은 AUC 기준 가장 뛰어난 분류 성능을 나타내, 모형 간 상호 보완성이 필요함을 시사했다. 변수 중요도 분석 결과, 비선형 모형(랜덤포레스트, LSTM)에서는 뉴스/검색 데이터 기반의 HSI가 전통적인 신용스프레드(Credit Spread)를 능가하지는 못하지만, 특히 랜덤포레스트 모형에서는 상당히 중요한 예측 변수로 식별되었다. 이는 금융 위기 예측에서 정적인 지표보다 시간의 흐름에 따른 시장 심리의 동태적 패턴이 더 결정적인 선행 정보를 포함하고 있음을 강력하게 시사한다.

본 연구 결과는 금융자산시장 쏠림현상으로 인한 금융불안정 리스크를 선제적으로 관리하기 위한 네 가지 정책 방향을 제시한다. 첫째, 군집행동 조기경보체계의 구축 및 정례 모니터링 강화가 필수적이다. 이번 연구에서 새롭게 시도한 HSI가 금융불안을 선행함을 확인한 만큼, 정책 당국은 빅데이터 및 인공지능 활용 역량을 강화하여 HSI를 정례 모니터링 지표에 편입해야 한다. HSI가 임계치를 초과할 경우 "쏠림

경보"를 발령하고 단계적인 완화 조치를 시행하여, 군중심리가 과열되는 국면에서 선제적으로 위험 신호를 포착할 필요가 있다. 둘째, 통화정책과 거시건전성정책의 유기적인 연계 운용이 중요하다. 금리 정책만으로는 군집심리의 비이성적 폭주를 억제하기 어렵기 때문에, 자산시장의 쓸림 및 금융불안 징후가 포착될 경우 LTV/DSR 등 거시건전성 규제 수단을 신속하게 발동하여 통화정책과 보완적으로 운용해야 한다. 셋째, 금융회사 보상구조 개선을 통한 미시적 행동 편향 완화 노력도 병행되어야 한다. 단기 실적 위주의 비대칭적 보상체계는 운용역의 단기적 군집 투자를 부추기므로, 성과보수를 장기 성과와 연계하고 클로백(clawback) 제도를 도입하여 과도한 위험 추구 유인을 억제할 필요가 있다. 넷째, 투자자 보호와 정보 공시 투명성을 제고하여 군집행동의 폐해를 완화해야 한다. 불완전 판매에 대한 검사를 강화하고, 투자자들이 포모(FOMO) 심리에 흔들리지 않도록 행동재무학적 편향에 대한 투자자 교육을 강화해야 한다.

결론적으로, 투자자 군집행동은 금융 불안정을 유발하는 핵심 동력이며, 전통적인 정책 수단만으로는 이를 통제하는 데 한계가 있다. 따라서 HSI와 같은 선행 심리 지표를 활용한 선제적 조기경보 시스템 구축과 함께, 거시적/미시적 정책 보완 장치들을 통합적으로 운용하는 다각적 접근이 필요하다. 이러한 전략은 시장 쓸림현상을 억제하고 거시경제와 금융안정의 양대 목표를 보다 효율적으로 달성하는 데 기여할 것이다.

## 참고문헌

- 강경훈(2006). 우리나라 금융시장의 쏠림현상, 금융조사보고서, 2006-08, 한국금융연구원
- 강경훈·여은정 (2011), 금융투자업 종사자 보수체계의 비대칭성과 투자 쏠림현상, 재무연구 24(2):455-481
- 이승호. (2016). 국내 금융권 자금흐름의 특징 및 변동요인 분석, 연구보고서 16-06, 한국자본시장연구원
- 지동현. (2010). 금융회사의 대출쏠림 억제방안. 주간금융브리프, 19(34), 3-6.
- Banerjee, A. V. (1992). A simple model of herd behavior. *Quarterly Journal of Economics*, 107(3), 797–817.
- Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3), 307–343.
- Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5), 992–1026.
- Bikhchandani, S., & Sharma, S. (2000). Herd behavior in financial markets (IMF Working Paper No. 00/48). International Monetary Fund.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Chang, E. C., Cheng, J. W., & Khorana, A. (2000). An examination of herd behavior in equity markets: An international perspective. *Journal of Banking & Finance*, 24(10), 1651–1679.
- Christie, W. G., & Huang, R. D. (1995). Following the pied piper: Do individual returns herd around the market? *Financial Analysts Journal*, 51(4), 31–37.
- Cipriani, M., & Guarino, A. (2014). Estimating a structural model of herd behavior in financial markets. *American Economic Review*, 104(1), 224–251.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019* (pp. 4171–4186).
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.

- Hwang, S., & Salmon, M. (2004). Market stress and herding. *Journal of Empirical Finance*, 11(4), 585–616.
- Jarmulska, B. (2020). Random forest versus logit models: which offers better early warning of fiscal stress?, Working Paper No. 2408, European Central Bank.
- Loisel, O., Pommeret, A., & Portier, F. (2012). Monetary policy and the herd behavior of asset prices. *Journal of Economic Dynamics and Control*, 36(8), 1133–1148.
- Oust, A., & Eidjord, O. M. (2020). Can Google search data be used as a housing bubble indicator? *International Real Estate Review*, 23(2), 267–308.
- Phillips, P. C. B., Shi, S., & Yu, J. (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review*, 56(4), 1043–1078.
- Rajan, R. G. (2006). Has finance made the world riskier? *European Financial Management*, 12(4), 499–533.
- Scharfstein, D. S., & Stein, J. C. (1990). Herd behavior and investment. *American Economic Review*, 80(3), 465–479.
- Shiller, R. J. (2015). *Irrational exuberance* (3rd ed.). Princeton University Press.
- Spyrou, S. (2013). Herding in financial markets: A review of the literature. *Review of Behavioral Finance*, 5(2), 175–194.
- Tran, N. T. (2024). A systematic review of the monetary policy and herd behavior. *Journal of Management, Economics, and Industrial Organization*, 8(1), 39–78.