

국제 금융 시장 지수 선물 하방 리스크의 정보 전이 네트워크 개발 및 예측에서의 응용

최인수¹⁾, 김우창²⁾

주식 시장은 경제학 및 금융의 중요한 연구 분야 중 하나로 광범위하게 연구됐다. 특히 주가와 주가의 리스크에 대한 데이터를 기반으로 증시를 분석하고 예측하는 연구는 금융공학 분야 내에서 가장 활발히 다루어지고 있는 주제이며 구체적으로 주가 및 리스크를 바탕으로 한 시장 전반에 대한 예측은 개인이나 기관투자자들이 최적의 투자전략을 수립하기 위한 중요한 요소 중 하나이다. 본 연구에서는 리스크 중 하방 리스크에 집중한 연구들을 진행하였으며, 하방 리스크의 지표로 최대 예상 손실액(VaR, value-at-risk)과 조건부 최대 예상 손실액(CVaR, Conditional value-at-risk)의 이론적 상한으로 알려진 엔트로피 최대 예상 손실액(EVaR, entropic value-at-risk)을 활용하였다. 이러한 EVaR의 일별 변화율에 대한 경험적 분포가 정규성을 만족하지 않는 데이터에 대해서 분포의 형태에 강건한 정보 불확실성 감소 측면에서의 인과관계 측정 지표인 전이 엔트로피(transfer entropy) 개념을 활용하여 국제 금융 시장 지수 선물의 EVaR에 대한 정보 전이를 바탕으로 한 인과관계 네트워크를 구성하였다. 구체적으로는 전이 엔트로피 측정 시 발생할 수 있는 유한크기 효과(finite size effect)를 조정하는 데 있어서 효과적인 지표인 효율적 전이 엔트로피를 활용하였다. 그리고 구성된 네트워크를 활용하여 금융 시장의 하방 리스크에 대한 정보 전이를 분석하였으며, 또한 정보 전이의 존재 여부를 바탕으로 국제 금융 시장 지수 선물별 하방 리스크 예측 시 이러한 네트워크의 위상(topology)을 바탕으로 구성된 군집을 활용하면 더 적은 수의 데이터로 전체 데이터 세트를 사용할 때와 같은 유사한 예측력, 혹은 더욱 나은 예측력을 갖게 되는지 확인하고자 하였다. 이때 예측에는 의사결정나무 모형 기반의 그래디언트 부스팅(gradient boosting) 기법을 사용한 세 모형을 소프트 보팅(soft voting) 방식으로 활용하였다.

핵심 주제어: 전이 엔트로피, 엔트로피 최대 예상 손실액(EVaR), 네트워크 모형, 하방 리스크 증감 예측, 기계 학습

JEL 분류기준: F3

1. 서론

주식 시장은 경제학 및 금융의 중요한 연구 분야 중 하나로 광범위하게 연구됐다. 특히 주가와 수익률 데이터를 기반으로 한 증시를 분석하고 예측하는 연구는 금융 분야 내에서 가장 활발히 다루어지고 있는 주제이다. 특히, 구체적으로는 주가 및 시장 전반에 대한 예측은 투자자들이 최적의 투자전략을 수립하기 위한 필수 과제 중 하나이다. 이러한 주식 시장의 예측은 시계열과 관련된 가장 중요하고 도전적인 문제 중 하나로 평가받는다(Chen et al., 2017). Fama (1970)에 의해 주창되었으며 Fama (1991)에 의해 수정된 개념인 무작위 경로를 따르기 때문에 예측할 수 없다는 효율적 시장 가설(efficient market hypothesis)이 확립되었음에도 예측을 통한 수익 발생 모형에 관한 연구는 지속해서 이루어지고 있다(Weng et al., 2017). 실제로 Fama (1991), Malkiel

* 본 논문의 일부 선행 연구 및 방법론 서술은 2022년 6월 발간 예정인 대한산업공학회지 48(3)의 출판 예정 논문 <정보 전이 네트워크를 활용한 국제 금융 시장 지수 선물의 분석 및 국내 시장 지수 선물 등락 예측> 중 제1 저자가 서술한 부분을 일부 개고하여 서술하였음

1) 제1 저자, 한국과학기술원 산업 및 시스템공학과 박사과정

2) 교신저자, 한국과학기술원 산업 및 시스템공학과 교수, 주소: 대전광역시 유성구 대학로 291(한국과학기술원) E2-2 #3107, e-Mail: wkim@kaist.ac.kr, 전화: 042-350-3129

(2003), Atsalakis and Valavanis (2009), Kumar et al. (2016)의 연구 결과에서 요약된 것과 같이 이러한 금융 시장이 효율적이지 못하다는 연구 역시 꾸준히 이루어지고 있으며 시간이 지남에 따라 수익을 창출할 수 있는 예측 모형은 시장이 효율적이지 못함을 의미하는 강력한 증거를 나타낼 뿐만 아니라 금융 포트폴리오 운영에 있어서 매우 중요한 요소이기에 상기한 것과 같이 그 중요성이 역설되고 있다(Henrique et al., 2019).

이러한 금융 시장의 분석 및 예측에 있어 상관 분석은 두 변수의 유사성 정도를 측정하기 위한 간단하고 좋은 지표이다. 상관관계 기반 시계열 분석에 관한 여러 연구가 상관관계를 바탕으로 금융 시장의 특성에 대해 밝혀낸 바 있다(Plerou et al., 2002; Kim et al., 2011; Kumar and Deo, 2012). 또한, 금융 시장이 복잡계 네트워크로 상정되며 이에 관한 연구가 이루어지면서 Pearson 상관계수 등과 같은 선형 상관관계 외에도 모든 금융 시장의 움직임이 동조화(synchronization)되는 것이 아니라 일정한 시간에 걸쳐서 이루어짐을 확인하게 되면서 금융 시장에 대해서 선형적인 상관관계 외에도 비선형적인 의존관계(dependency)나 인과관계(causality)와 같이 시차를 둔 형태 등 다른 방법으로 정량화 및 시각화하려는 시도 역시 발생하기 시작했다. 이에 연구자들은 이러한 흐름에서 금융 시장의 비선형적 의존성과 인과관계 지표에 관해서도 관심을 두고 연구하기 시작했다.

이러한 흐름에서 시계열 사이의 인과관계를 정의하기 위해 Granger (1969)에 의해 Granger 인과관계가 도입되었으며 이는 시계열의 선형적 인과관계를 측정하는 지표로 널리 사용되기 시작하였다. 이러한 Granger 인과관계는 일반적으로 벡터 자기 회귀모형(VAR, vector autoregression model)을 전제로 하며 이러한 벡터 자기 회귀모형은 데이터의 정규성을 바탕으로 해야 한다는 조건이 있다. 하지만 여러 연구자는 가격 데이터나 수익률 데이터는 일반적으로 정규성을 만족하지 않으며 금융 데이터가 실제 경험적 분포가 정규 분포가 아닌 다른 분포, 구체적으로는 Pareto 분포(Pareto distribution), t - 분포(t - distribution), Cauchy 분포(Cauchy distribution), Laplace 분포(Laplace distribution) 등에 가깝다는 연구 결과를 제시한 바 있다 (Mandelbrot and Taylor, 1967; Bouchaud, 2001; Linden, 2001; Pierou et al., 2001; Cowan and Sergeant, 2001; Young et al., 2006; Young, 2008; Almonte et al., 2011; Beitinger et al., 2015; Lausberg et al., 2020).

이에 기존의 연구들은 주로 중심극한정리에 기반하여 데이터의 정규성을 전제하고 Granger 인과관계를 측정하였다. 하지만 이러한 정규성이 전제되어야 하는 Granger 인과관계의 문제에 대해서 여러 대안이 제시되었으며 일부 연구들은 시스템 내에서 통계 물리학의 개념을 활용하기 시작했다. 경제학 및 금융 분야 역시 경제 또는 금융 시스템에서 금융 시장 간의 관계를 정량화하기 위해 복잡계 시스템 및 정보 이론과 같은 통계 물리학의 개념을 활용하여 측정하려는 여러 연구가 이루어졌다(Mantegna and Stanley, 1999; Noh, 2000; Bonanno et al. 2003; Zunino et al. 2008; Chi et al. 2010; Bekiros et al., 2017).

이러한 흐름에서 Schreiber (2000)가 전이 엔트로피(TE) 개념을 제시하였다. 전이 엔트로피는 Shannon (1948)이 제시한 정보 엔트로피로부터 발생하며, 특정 변수에서 변수로의 정보 불확실성의 감소를 측정하는 비모수적 척도이다. 전이 엔트로피는 시스템 내의 비대칭 상호작용을 포착하고 주행 및 반응 요소를 효율적으로 구분할 수 있는 능력 때문에 신경 과학 및 사회 과학 분야의 연구에 활용됐다. 예를 들어, Vicente et al. (2011)은 신경과학 분야에서 전이 엔트로피가 비선형 상호작용에 대한 효율적 연결을 탐지하는 능력을 향상할 수 있음을 확인하였고, Faes et al. (2013)은 전이 엔트로피를 활용한 심혈관계 및 신경학적 시계열의 분석이 매우 효과적이며, 이를 통해 전이 엔트로피를 활용한 생리학적 메커니즘 조사를 위한 프레임워크를 제시하였다. Kim et al. (2016)은 전이 엔트로피를 이용하여 통계적 가정 없이 복잡계 사회적 네트워크 내에서 광역 수준의 정보 전이를 기반으로 한 인과관계 분석이 가능함을 확인하였다.

이러한 흐름에서 예측과 인과관계는 매우 밀접한 관련이 있으며, 상기한 것과 같은 인과관계 네트워크 기반의 연구가 이루어짐에 따라 본 연구에서는 금융 시장 지수 선물의 인과관계 네트워크를 구성하고 여기에 금융 시장 지수 선물의 예측을 접목하고자 하였다. 이러한 방법을 통해 본 실험에서는 금융 시장 지수 선물의 하방 리스크에 대한 예측력 향상의 여부와 더불어 특정 금융 지수 선물과 밀접한 더 적은 수의 데이터만을 이용하여 더 많은 데이터를 이용한 것보다 더 나은 효과를 낼 수 있는지를 확인하고자 하였다. 이러한 금융 시장 네트워크를 예측에 활용한 예는 Kwon et al. (2005)이 개별 주식 단위에서 교차-상관계수를 이용한 네트워크를 구성한 뒤 유전 알고리즘 및 순환 신경망을 활용한 예측을 통해 매수 및 매도 전략을 구성한 연구와 Liu et al. (2019)이 금융 지수 선물의 부실성을 금융 시장 네트워크와 그래디언트 부스팅 기계 학습 모형을 이용하여 예측한 연구 등이 있다.

본 연구에서는 이러한 흐름에서 전이 엔트로피 개념을 활용하여 주요 금융 시장 지수 선물들의 정보 흐름을 바탕으로 한 인과관계 네트워크를 구성하고 구성된 네트워크의 연결 여부를 바탕으로 국제 금융 시장 지수 선물의 하방 리스크 증감에 대한 예측력이 개선될 수 있는지 확인하는 것을 목표로 하였다.

본 연구의 구성은 다음과 같다. 2장에서는 데이터에 관해 설명하며 3장에서는 연구 방법론에 관해 설명한다. 4장에서는 3장의 연구 방법론을 바탕으로 한 금융 시장 지수 선물의 하방 리스크에 대한 네트워크를 구성한다. 5장에서는 4장에서 구성된 네트워크를 바탕으로 하여 국내 금융 시장의 예측 결과에 대해서 분석하였다. 6장에서는 연구의 결론, 한계점 및 의의에 관해 설명한다.

2. 데이터 및 연구 방법론

2.1 데이터 설명 및 수집

본 연구에서는 2002년 1월 1일부터 2021년 12월 31일까지의 기간을 연구를 위한 데이터 수집의 기간으로 설정하였다. 이 중 금융 시장 지수 선물 간의 수익률의 흐름에 대한 네트워크의 도식화를 위한 실험 기간 및 기계 학습의 학습을 위한 실험 기간은 2002년 1월 1일부터 2017년 12월 31일이었다. KOSPI의 학습된 모형에 대한 테스트 기간은 2018년 1월 1일부터 2021년 12월 31일로 설정하였다. 실험에 사용된 국제 금융 시장 지수 선물은 <Table 1>과 같이 2020년 기준 국제 금융 시장의 시총 및 영향도를 고려하여 총 39개로, 모두 지수 선물 데이터를 활용하여 데이터열을 생성하였다. 또한 시각화의 편의성을 위하여 각 국가 및 금융 시장 지수 선물은 좌측의 열과 같이 티커(ticker)로 표현하였다. 본 연구에서는 연구에서 활용되는 국가별로 최대 시차 차이가 24시간 이상이 날 때는, 일별 데이터의 일자가 달라짐에 따라 본 연구에서 정의하고자 하는 인과관계의 시차에 대한 전제가 필요하나 본 연구에서 사용된 금융 시장 간의 최대 시차 차이는 24시간보다 작으므로 인과관계의 계산에는 유의미한 차이가 없을 것으로 판단하고 실험을 진행하였다.

Table 1. Selected global market indices

Ticker	Financial index
AEX	네덜란드 금융 시장 지수 선물
AS51	호주 금융 시장 지수 선물
ATX	오스트리아 금융 시장 지수 선물
BEL20	벨기에 금융 시장 지수 선물
CAC	프랑스 금융 시장 지수 선물
DAX	독일 금융 시장 지수 선물
FBMKLCI	말레이시아 금융 시장 지수 선물
FSSTI	싱가포르 금융 시장 지수 선물
FTSEMIB	이탈리아 금융 시장 지수 선물
HSI	홍콩 금융 시장 지수 선물
IBEX	스페인 금융 시장 지수 선물
IBOV	브라질 금융 시장 지수 선물
INDEXCF	러시아 금융 시장 지수 선물
INDU	미국 다우존스 산업평균지수 선물
JCI	인도네시아 금융 시장 지수 선물
KOSDAQ	대한민국 코스닥 지수 선물
KOSPI	대한민국 코스피 지수 선물
MEXBOL	멕시코 금융 시장 지수 선물
MID	미국 S&P Mid Cap 지수 선물
N100	유럽 Euronext 100 지수 선물
NDX	미국 NASDAQ 100 지수 선물
NIFTY	인도 금융 시장 지수 선물
NKY	일본 금융 시장 지수 선물
OMXS	스웨덴 금융 시장 지수 선물
OSEAX	노르웨이 금융 시장 지수 선물

PCOMP	필리핀금융 시장 지수 선물
SASEIDX	사우디아라비아 지수 선물
SENSEX	인도 금융 시장 지수 선물
SET50	태국 금융 시장 지수 선물
SHCOMP	중국 금융 시장 지수 선물
SMI	스위스 금융 시장 지수 선물
SPTSX	캐나다 금융 시장 지수 선물
SPX	미국 S&P 500 지수 선물
SX5E	유럽 EuroStoxx50 지수 선물
TOP40	남아프리카공화국 금융 시장 지수 선물
TWSE	대만 금융 시장 지수 선물
UKX	영국 금융 시장 지수 선물
VIX	미국 변동성 지수 선물
W5000	미국 월서5000 지수 선물

3. 연구 방법론

3.1 엔트로피 최대 예상 손실액 (EVaR, Entropic Value-at-Risk)

금융 수학과 확률적 최적화에서 위험 척도의 개념은 무작위 결과 또는 위험 위치에 관련된 위험을 정량화하는 데 사용된다. 지금까지 많은 하방 리스크에 대한 지표가 제안되었으며 각각 특정 특성을 가지고 있다. 엔트로피 최대 예상 손실액(Entropic value at risk, EVaR)은 Ahmadi-Javid (2011)이 도입한 일관성 있는 위험 척도이며, 특정 α 값 하에서 체르노프 부등식(Chernoff inequality)에 의해 최대 예상 손실액(VaR, value-at-risk)과 조건부 최대 예상 손실액(CVaR, conditional value-at-risk)의 상한임이 수학적으로 증명되어 있다. EVaR은 또한 상대 엔트로피의 개념을 사용하여 표현될 수 있다는 점에서 엔트로피 최대 예상 손실액으로 불리며, EVaR은 CVaR의 일부 계산 비효율성을 해결하기 위해 개발되었다. 이때 EVaR은 90일의 역사적 데이터를 사용하여 일별 이동창으로 측정하였다. 측도 공간 (Ω, F, P) 에 대해서 확률변수 X 와 Borel 가측 함수 $X: \Omega \rightarrow \mathbf{R}$ 의 집합 L_{M^+} 에 대해 모든 $z \geq 0$ 에 대해 적률생성함수 $M_X(z)$ 를 정의한다면, $1 - \alpha$ 신뢰 수준에서 $X \in L_{M^+}$ 인 $EVaR$ 은 다음과 같이 정의되며 이때, 확률변수 $X \in L_{M^+}$ 는 포트폴리오에서의 손해(losses)를 사용하게 된다.

$$EVaR_{1-\alpha}(X) := \inf_{z > 0} \left\{ z^{-1} \ln \left(\frac{M_X(z)}{\alpha} \right) \right\}$$

본 연구에서는 $\alpha=0.01$, 즉 신뢰 수준 99% 하에서의 EVaR을 계산하고자 하였으며, 본 연구의 방향성이 하기할 전이 엔트로피를 사용하는 이유가 데이터 분포의 성질에 대해 자유로운 분석을 시도하고자 하였으므로 경험적 방법(historical method)으로 이러한 EVaR을 계산하였다.

3.2 전이 엔트로피

변형이 없는 기본적인 형태의 그래인저 인과관계를 산출하기 위해서는 정규성, 정상성, 선형성과 같은 데이터의 통계적 성질에 대한 가정이 전제된다. 하지만 2장의 결과에서 확인할 수 있듯 본 연구에서 사용된 데이터 세트는 전부 정규성을 전혀 만족하지 못하였으며, 이는 금융 데이터의 변화량 데이터가 일반적으로 정규성을 만족하지 못한다는 기존의 연구 결과를 따른다고 볼 수 있다. 이러한 결과에 따라 본 연구에서는 위에 언급된 가정 없이 사용할 수 있는 방법론을 활용하고자 하였고, 그 결과 정보 이론을 기반으로 하여 인과관계를 측정하는 Schreiber (2000)이 제시한 전이 엔트로피(transfer entropy) 개념을 활용하였다. 전이 엔트로피는 Shannon

엔트로피를 기반으로 두 변수 사이의 정보 흐름을 확인하기 위한 비모수 측정 방법론이다. 전이 엔트로피는 인과관계를 정보의 흐름에 따른 해소된 불확실성을 측정한다는 아이디어를 바탕으로 시스템 내 인과관계를 정량화하고 선행 변수(source)로부터 종속 변수(target)에 대한 인과관계를 크기를 측정한다. 이러한 전이 엔트로피는 정보 또는 물리학 분야를 넘어서 신경 과학, 전기 공학, 화학 공학 분야에서도 인과관계를 측정하는 지표로써 사용되고 있다. 엔트로피 이론은 데이터의 성질과 무관하게 모형을 활용할 수 있으며, 비선형적 인과관계 역시 측정할 수 있다는 장점이 있으며 이러한 장점은 엔트로피를 기반으로 한 인과관계 지표인 전이 엔트로피 역시 유효하다.

이러한 전이 엔트로피를 계산하기 위해서는 우선 정보 엔트로피의 정의가 필요하다. 본 연구에서는 Shannon (1948)이 제시한 정보 엔트로피 개념을 활용하였으며, Shannon의 정보 엔트로피 $H(X)$ 는 확률공간 Ω 에서 확률질량함수 $p(x)$ 를 갖는 확률변수 X 에 대해서 식 (1) 과 같이 정의된다. 이러한 엔트로피를 바탕으로 전이 엔트로피는 두 시계열 변수 X_t 와 Y_t 에 대해 식 (2) 와 같이 정의된다. 이때 식 (2) 의 k 와 l 은 각 시계열 X_t 와 Y_t 에서 전이 엔트로피의 계산에 활용되는 시차(lag)를 의미하며, $x_t^{(k)} = \{x_t, x_{t-1}, \dots, x_{t-k+1}\}$, $y_t^{(l)} = \{y_t, y_{t-1}, \dots, y_{t-l+1}\}$ 인 시계열을 의미한다.

$$H(X) = - \sum_{x \in \Omega} p(x) \log_2 p(x) \quad (1)$$

$$\begin{aligned} TE_{Y \rightarrow X}^{(k,l)}(t) &= H(X_{t+1} | X_t, \dots, X_{t-k+1}) - H(X_{t+1} | X_t, \dots, X_{t-k+1}, Y_t, \dots, Y_{t-l+1}) \\ &= \sum_i p(x_{t+1}, x_t^{(k)}, y_t^{(l)}) \log_2 p(x_{t+1} | x_t^{(k)}, y_t^{(l)}) - \sum_i p(x_{t+1}, x_t^{(k)}, y_t^{(l)}) \log_2 p(x_{t+1} | x_t^{(k)}) \\ &= \sum_i p(x_{t+1}, x_t^{(k)}, y_t^{(l)}) \log_2 \frac{p(x_{t+1} | x_t^{(k)}, y_t^{(l)})}{p(x_{t+1} | x_t^{(k)})} \end{aligned} \quad (2)$$

본 연구에서는 두 시계열의 시차 k 와 l 에 대해서 $k = l = 1$ 인 경우를 가정하였다. 이는 가장 짧은 시차 형태로 이러한 시차 형태를 사용하는 것은 약형 효율적 시장가설(efficient market hypothesis), 즉 무작위 행보 이론(random walk behavior)을 가정하기 때문이다(Bekiros et al., 2017). 약형 효율적 시장가설은 가격 움직임, 물량, 수익 데이터가 주가에 영향을 미치지 않고 향후 방향을 예측하는 데 사용될 수 없다는 주장이며 이러한 형태의 가정을 하면 가장 짧은 시차를 통해 TE를 계산하게 되는 것이 가장 적합한 시차로 해석될 수 있다. 이렇게 되면 전이 엔트로피는 다음과 같이 표현될 수 있다.

$$\begin{aligned} TE_{Y \rightarrow X}^{(1,1)}(t) &= \sum_i p(x_{t+1}, x_t, y_t) \log_2 \frac{p(x_{t+1} | x_t, y_t)}{p(x_{t+1} | x_t)} \\ &= \sum_i p(x_{t+1}, x_t, y_t) \log_2 \frac{p(x_{t+1}, x_t, y_t) p(x_t)}{p(x_{t+1}, x_t) p(x_t, y_t)} \end{aligned} \quad (3)$$

이러한 $k = l = 1$ 시차 하에서 본 연구에서는 엔트로피를 기반으로 한 인과관계 측정 지표인 전이 엔트로피(transfer entropy, TE)의 발전된 형태로 효율적 전이 엔트로피(effective transfer entropy, ETE)를 활용하고자 하였다. 이러한 효율적 전이 엔트로피는 유한 시계열에서 발생할 수 있는 유한크기효과(finite size effect)를 제거한 형태이다. 이는 Marschinski and Kantz (2002)에 의해 처음 제시되었으며 여러 가지 형태가 있으나 본 연구에서는 Boba et al. (2015)에서 제시한 두 시계열을 모두 섞어 계산된 식 (4) 과 같은 Z값과 같은 형태의 효율적 전이 엔트로피값을 사용하기로 하였으며, 총 1,000회의 시뮬레이션을 시행하였다. 식 (4) 의 $\overline{TE}_{shuffle}$ 은 두 시계열을 모두 섞어 1,000회 반복된 전이 엔트로피의 평균값 $\sigma_{shuffle}$ 은 두 시계열을 모두 섞어 1,000회 반복된 계산된 전이 엔트로피의 평균값에 대한 표준편차를 의미한다.

$$ETE = \frac{TE - \overline{TE}_{shuffle}}{\sigma_{shuffle}} \quad (4)$$

전이 엔트로피는 금융 분야에서 주로 2000년대 후반부터 금융자산과 금융 시장의 인과관계를 확인하는 데 사용됐다. 2000년대에 Marschinski and Kantz (2002)는 독일 DAX Xetra Stock Index (DAX) 와 미국 Dow Jones Industrial Average Index 사이의 인과관계를 보고하였으며, Kwon and Yang (2008)은 정보 이전의 방향을 보여 준 결과, 시장 지수 선물이 미국 증시의 개인 주식에 영향을 미친다는 것을 발견했다. 2010년대 이후에는 Dimpfl and Peter(2013)는 신용위험의 가격 책정에 대한 회사채 시장에 대한 신용부도스왑 시장의 인과관계와 VIX와 iTraxx Europe의 시장위험과 신용위험 사이의 동적 관계를 위기 이전, 위기 이후 기간의 관점에서 분석했다. Sensoy et al. (2014) 은 효율적 전이 엔트로피를 사용하여 몇몇 신흥국가의 환율과 주가 간 정보 흐름의 강도와 방향을 조사했다. 전이 엔트로피를 기반으로, Bekiros et al. (2017) 는 미국 주식 및 상품 시장에서 네트워크 역학을 전이 엔트로피를 기반으로 산출하였다. 그리고 Lim et al. (2017)은 전이 엔트로피를 기반으로 신용부도스왑 시장과 미국의 주식 시장 간의 정보 흐름을 구조를 분석하였다. 최근에 Jang et al. (2019)는 TE를 사용하여 비트코인, 금, S&P 500 지수 선물, 미국 달러 간의 인과관계를 연구하였으며 Yue et al. (2020; 2020) 은 중국 금융 상품 간의 수익률의 인과관계에 대해 TE를 이용하여 분석하고, 또한 미국과 중국의 금융 시장에서의 정보 흐름에 대해 비교 분석했다. 이러한 선행 연구를 바탕으로 본 연구에서는 전이 엔트로피가 본 연구의 목적인 금융 시장 지수 선물의 하방 리스크에 대한 네트워크의 구축과 이를 바탕으로 한 KOSPI의 예측에 있어 적합하다고 판단하였고, 이를 인과관계의 지표로 활용하였다.

이때, 전이 엔트로피는 두 시계열 데이터의 확률분포를 이산화한 후 산출이 가능한데 본 연구에서는 가장 보편적으로 사용되는 시계열 데이터 이산화 방법론인 동일한 간격으로 정의된 히스토그램을 활용하였으며, 이때 히스토그램 구간의 수는 주어진 데이터에 대해 히스토그램 기반으로 결합 엔트로피를 계산할 때 수치 해석적으로 최적의 구간 수를 제시한 Hacine-Gharbi and Ravier (2018)의 연구를 참조하였다.

3.3 네트워크 분석

3.3.1 네트워크 밀도(Network Density)

네트워크의 밀도는 네트워크 노드들 사이에서 연결된 정도를 의미한다. D_i 를 금융 시장 지수 선물 i 의 금융 시장 지수 선물의 하방 리스크에 대한 네트워크의 밀도로 정의할 때, 금융 시장 지수 선물의 하방 리스크에 대한 네트워크 내의 주식의 수가 n_i 개라 할 때, 발생 가능한 $n_i(n_i - 1)$ 개의 인과관계 중 통계적으로 유의미한 인과관계의 개수 k_i 의 비율로 정의한다. 이를 요약하면 식 (5) 과 같다.

$$D_i = \frac{k_i}{n_i(n_i - 1)} \quad (5)$$

3.3.2 연결 중심성(Degree Centrality)

연결 중심성은 전이 엔트로피에 의해 인과관계를 갖는 주식의 개수를 집계하는 것으로 내향 연결 정도 중심성과 외향 연결 정도 중심성으로 구분된다.

$$N_j^I: (S \rightarrow j) = \frac{1}{N-1} \sum_{i \neq j} (i \rightarrow j) \quad (6)$$

$$N_j^O: (j \rightarrow S) = \frac{1}{N-1} \sum_{i \neq j} (j \rightarrow i) \quad (7)$$

식 (6)의 N_j^I 는 내향 연결 중심성(in-degree centrality)으로 구성된 금융 시장 지수 선물의 하방 리스크에 대한 네트워크 S 내에서 효율적 전이 엔트로피에 의해 영향력을 받는 다른 금융 시장 지수 선물의 개수를 의미하며 식 (7)의 N_j^O 는 외향 연결 중심성(out-degree centrality)으로 금융 시장 지수 선물의 하방 리스크에 대한 네트워크 S 내에서 효율적 전이 엔트로피에 의해 영향력을 행사하고 있는 다른 금융 시장 지수 선물의 개수를 의미한다. 이때 $i \neq j$ 이다.

3.3.3 페이지랭크(PageRank)

중심성은 네트워크 분석에 있어서 가장 많이 사용되는 지표로서, 노드가 가지는 영향력을 측정하기 위해 중심에 위치하는 정도를 측정한 지표이다. 이 중 Brin et al. (1999)이 고안한 PageRank는 자사 검색엔진에서 웹페이지의 순위를 측정하기 위해 고안한 중심성 측정 알고리즘으로, 네트워크상의 경로들을 고려하여 노드의 중심성을 결정하는 고유벡터 기반 위세 중심성 지표 중 가장 유명한 방법이다. 특히, 유사한 측정 방법인 고유벡터 중심성(eigenvector centrality)과 Katz 중심성(Katz centrality) 등의 지표의 단점을 보완한 방법으로 알려져 있다. PageRank는 유향 그래프(directed graph)의 연결 분석 방법으로서, 각 노드의 영향 그리고 현재 및 미래의 정보 흐름의 세 요소를 모두 측정할 수 있다.

PageRank를 계산하는 방법은 다음과 같다. 네트워크에 N 개의 노드가 있을 때, 이 네트워크는 $\mathbf{A} \in \mathbb{R}^{N \times N}$ 라는 인접행렬(adjacency matrix)로 나타낼 수 있다. 이 인접행렬을 통해 각 노드의 PageRank를 원소로 하는 열벡터 $\mathbf{r} \in \mathbb{R}^N$ 를 아래와 같이 수학적으로 계산할 수 있다.

$$\mathbf{r} = (1 - \alpha)(\mathbf{I} - \alpha \mathbf{A}^T \mathbf{D}^{-1})^{-1} \mathbf{1} \quad (8)$$

식 (8)에서 $\mathbf{I} \in \mathbb{R}^{N \times N}$ 는 단위행렬이며, $\mathbf{1} \in \mathbb{R}^N$ 는 모든 원소가 1인 열벡터이고, $\mathbf{D} \in \mathbb{R}^{N \times N}$ 는 $\max(K_{out}^i, 1)$ 를 원소로 하는 대각행렬이다. 이때 K_{out}^i 는 노드 i 에서 뺄어나가는 유향 선분의 개수를 의미한다. α 는 0에서 1 사이의 값을 가지는 감쇠치(damping factor)로, 본 연구에서는 PageRank를 고안한 연구에서 사용된 통상적인 값인 0.85로 정했다(Brin et al, 1999). 위의 식을 이용하면, 거듭제곱 반복법(power method)을 통해 값으로 PageRank의 값을 도출해낼 수 있게 된다. PageRank의 값이 클수록 그 노드의 중심성이 강하다는 뜻이기 때문에 본 연구에서 구성할 금융 시장 네트워크에서 해당 금융 지수 선물의 하방 리스크가 갖는 영향력이 강하다는 의미로 해석할 수 있다.

3.3.4 군집 분석

네트워크의 군집을 생성하기 위하여 다음과 같은 Dugué and Perez (2015)의 유향 네트워크에 대한 Louvain 알고리즘에 관한 연구를 활용하였다. 해당 연구에서는 유향 네트워크의 군집화를 위하여 다음과 같은 네트워크의 모듈성(modularity)을 최대화하는 방식으로 활용하였다. 이때 모듈성(modularity)은 네트워크 내에서 상대적으로 밀접한 관계를 맺고 있는 하위 집단을 찾기 위해 주로 사용되는 척도로, 모듈성이 크다는 것은 집단 내 정점(금융 시장 지수 선물)들 사이의 관계가 집단 간 금융들 사이의 관계에 비해 강한 것을 의미한다. 다시 말해, 하위 그룹이 명확하게 분류될수록 모듈성의 값이 큼을 의미한다.

$$Q = \frac{1}{w} \sum_{i,j} (A_{ij} - \gamma \frac{d_i^+ d_j^-}{w}) \delta_{c_i, c_j} \quad (9)$$

A 는 인접행렬, c_i 는 노드 i 의 군집, d_i^+ 와 d_i^- 는 외향 및 내향 가중치를 의미하며 $w = \mathbf{1}^T A \mathbf{1}$ 이며 δ 는 Kronecker의 델타이다. 마지막으로, γ 는 해상도 모수(resolution parameter)이다.

4. 실험 결과

4.1 기초통계량

Table 2. Descriptive Statistics of Global Indices' EVaR

Financial Index	Average	Standard Deviation	Minimum	Maximum	Q1	Median	Q3	Skewness	Kurtosis	W	p-value (W)	JB	p-value (JB)	DF	p-value (DF)
ATX	0.0388	0.0198	0.0137	0.0975	0.0262	0.0317	0.0478	1.5680	1.9470	0.8175	0.0000** *	4143.4145	0.0000** *	-2.8002	0.0582*
BEL20	0.0338	0.0185	0.0116	0.0981	0.0223	0.0276	0.0403	2.0825	4.5972	0.7638	0.0000** *	11699.2099	0.0000** *	-2.6755	0.0784*
IBOV	0.0464	0.0211	0.0211	0.1140	0.0327	0.0371	0.0506	1.8849	3.1008	0.7595	0.0000** *	7244.8270	0.0000** *	-3.1223	0.0249**
SPTSX	0.0281	0.0192	0.0117	0.0898	0.0182	0.0215	0.0278	2.2509	3.9746	0.6390	0.0000** *	10965.6460	0.0000** *	-2.6277	0.0874*
DAX	0.0369	0.0169	0.0140	0.0846	0.0246	0.0289	0.0476	1.1129	0.7088	0.8730	0.0000** *	1659.6170	0.0000** *	-2.7397	0.0674*
IBEX	0.0400	0.0210	0.0136	0.0973	0.0241	0.0353	0.0451	1.2921	0.9991	0.8521	0.0000** *	2334.8133	0.0000** *	-3.0204	0.0330**
N100	0.0344	0.0167	0.0132	0.0826	0.0233	0.0276	0.0433	1.2891	1.3167	0.8635	0.0000** *	2548.7157	0.0000** *	-2.7260	0.0697*
SXSE	0.0370	0.0163	0.0136	0.0756	0.0241	0.0314	0.0486	0.7026	-0.4545	0.9205	0.0000** *	663.8427	0.0000** *	-2.6755	0.0784*
CAC	0.0382	0.0190	0.0143	0.0852	0.0243	0.0302	0.0474	0.9880	0.0294	0.8818	0.0000** *	1188.2465	0.0000** *	-2.6777	0.0780*
UKX	0.0305	0.0157	0.0111	0.0752	0.0207	0.0250	0.0360	1.5987	1.8548	0.7986	0.0000** *	4155.2297	0.0000** *	-2.4603	0.1254
FTSEMIB	0.0423	0.0238	0.0163	0.1183	0.0262	0.0370	0.0511	1.7223	2.8249	0.8081	0.0000** *	6034.2429	0.0000** *	-3.0343	0.0318**
MEXBOL	0.0312	0.0106	0.0155	0.0568	0.0221	0.0283	0.0398	0.5279	-0.5547	0.9329	0.0000** *	432.9327	0.0000** *	-2.8011	0.0581*
AS51	0.0268	0.0150	0.0096	0.0697	0.0175	0.0215	0.0307	1.6452	1.8964	0.7826	0.0000** *	4386.1330	0.0000** *	-2.4246	0.1349
FBMKLCI	0.0270	0.0222	0.0066	0.1014	0.0149	0.0186	0.0310	2.1396	3.8234	0.6794	0.0000** *	10012.4346	0.0000** *	-2.6237	0.0882*
FSSTI	0.0282	0.0127	0.0109	0.0635	0.0184	0.0250	0.0330	1.2014	0.9168	0.8814	0.0000** *	2011.5161	0.0000** *	-2.4261	0.1345
HSI	0.0350	0.0171	0.0158	0.0930	0.0242	0.0289	0.0415	1.9049	3.6121	0.7827	0.0000** *	8380.0378	0.0000** *	-2.4960	0.1164
JCI	0.0390	0.0163	0.0100	0.0802	0.0265	0.0337	0.0491	0.8573	-0.0471	0.9178	0.0000** *	895.1825	0.0000** *	-2.9424	0.0406**
NIFTY	0.0388	0.0225	0.0119	0.0899	0.0231	0.0298	0.0471	1.2343	0.2429	0.8028	0.0000** *	1871.9926	0.0000** *	-2.5141	0.1121
NKY	0.0403	0.0159	0.0155	0.0885	0.0305	0.0351	0.0448	1.5167	1.9944	0.8398	0.0000** *	4007.7432	0.0000** *	-3.1400	0.0237**
PCOMP	0.0371	0.0198	0.0154	0.0937	0.0238	0.0297	0.0449	1.7171	2.1192	0.7615	0.0000** *	4952.2887	0.0000** *	-2.8998	0.0454**
SASEIDX	0.0449	0.0190	0.0132	0.0848	0.0282	0.0429	0.0601	0.4912	-0.7930	0.9412	0.0000** *	485.1801	0.0000** *	-2.3078	0.1695
SENSEX	0.0373	0.0213	0.0113	0.0911	0.0232	0.0286	0.0476	1.2352	0.4169	0.8255	0.0000** *	1909.5021	0.0000** *	-2.6269	0.0876*
SET50	0.0407	0.0241	0.0114	0.1079	0.0258	0.0348	0.0416	1.5478	1.4380	0.7801	0.0000** *	3543.4381	0.0000** *	-2.9623	0.0386**
SHCOMP	0.0403	0.0155	0.0156	0.0734	0.0277	0.0377	0.0517	0.4813	-0.8618	0.9404	0.0000** *	508.1430	0.0000** *	-1.5804	0.4934
TOP40	0.0324	0.0132	0.0141	0.0711	0.0253	0.0284	0.0341	1.6563	2.0464	0.7838	0.0000** *	4610.3043	0.0000** *	-2.9070	0.0445**
TWSE	0.0330	0.0102	0.0108	0.0527	0.0234	0.0324	0.0418	-0.0639	-1.1368	0.9555	0.0000** *	398.3386	0.0000** *	-2.6676	0.0798*
INDEXCF	0.0512	0.0286	0.0153	0.1501	0.0308	0.0490	0.0614	1.7935	4.1560	0.8164	0.0000** *	9162.6587	0.0000** *	-2.3742	0.1492
AEX	0.0351	0.0176	0.0115	0.0760	0.0226	0.0273	0.0425	1.0929	0.2189	0.8576	0.0000** *	1467.9882	0.0000** *	-2.5517	0.1034
OSEAX	0.0352	0.0164	0.0136	0.0810	0.0236	0.0326	0.0432	1.2672	1.1550	0.8648	0.0000** *	2359.2416	0.0000** *	-2.1471	0.2260
SMI	0.0314	0.0151	0.0124	0.0653	0.0219	0.0237	0.0390	0.9188	-0.3819	0.8615	0.0000** *	1071.9775	0.0000** *	-2.6609	0.0810*

OMXS	0.0354	0.0152	0.0127	0.0721	0.0225	0.0318	0.0476	0.8077	-0.3576	0.9032	0.0000** *	833.0575	0.0000** *	-2.8345	0.0535**
INDU	0.0304	0.0187	0.0119	0.0917	0.0172	0.0257	0.0356	1.9035	3.3317	0.7647	0.0000** *	7781.8830	0.0000** *	-2.6453	0.0840*
MID	0.0343	0.0214	0.0149	0.0994	0.0231	0.0270	0.0340	1.9364	2.7233	0.6974	0.0000** *	6816.3574	0.0000** *	-2.6236	0.0882*
SPX	0.0310	0.0185	0.0127	0.0849	0.0186	0.0269	0.0339	1.7527	2.2974	0.7617	0.0000** *	5341.9011	0.0000** *	-2.4543	0.1270
W5000	0.0312	0.0188	0.0129	0.0868	0.0192	0.0268	0.0318	1.7782	2.3628	0.7561	0.0000** *	5543.9132	0.0000** *	-2.4780	0.1209
NDX	0.0366	0.0182	0.0178	0.0857	0.0233	0.0312	0.0410	1.5194	1.3187	0.7884	0.0000** *	3337.3305	0.0000** *	-2.7750	0.0619*
VIX	0.1650	0.0389	0.0774	0.2296	0.1266	0.1694	0.1969	-0.4030	-0.9781	0.9382	0.0000** *	488.9859	0.0000** *	-3.2387	0.0179**
KOSPI	0.0356	0.0183	0.0134	0.0828	0.0223	0.0301	0.0470	1.1106	0.5030	0.8722	0.0000** *	1577.7501	0.0000** *	-2.8233	0.0550*
KOSDAQ	0.0496	0.0178	0.0182	0.0836	0.0356	0.0447	0.0621	0.4460	-0.8502	0.9395	0.0000** *	462.2144	0.0000** *	-2.9026	0.0450**

2.1절에서 언급된 데이터의 EVaR에 대한 기초 통계량 값은 위와 같다. Table 2에서 *, **, ***은 각각 $\alpha = 0.1$, $\alpha = 0.05$, $\alpha = 0.01$ 수준에서 해당 검정통계량이 통계적으로 유의미함을 의미한다. 그 결과, 정규성은 본 논문에서 전이 엔트로피를 활용하게 된 중요한 이유 중 하나이므로 결과의 강건성을 위하여 Jarque-Bera 검정과 Shapiro-Wilk 검정의 두 가지 정규성 검정 시행하였으며, 그 결과 39개의 모든 데이터열에 대해서 두 검정 모두에서 정규성을 만족하는 데이터열은 없음을 Jarque-Bera 검정의 검정 통계량 JB와 Shapiro-Wilk 검정의 검정 통계량 W를 통해 확인할 수 있다. 또한 정상성(stationarity)을 측정하기 위하여 Augmented Dickey-Fuller 검정을 실시하였다. 이러한 통계 검정 결과는 본 연구에서 후술할 내용과 같이 Granger 인과관계와 비교해 통계적 성질에 대한 전제로부터 자유로운 전이 엔트로피 개념을 이용하여 금융 시장 네트워크를 구축하는 데에 타당한 통계적 근거로 활용될 수 있다.

4.2 네트워크 구성 결과 및 해석

본 연구에서는 ETE로 상정된 Z의 값을 통해 산출된 p-값에 대해 유의 수준 $\alpha = 0.1, 0.05, 0.01$ 에 대해서 위와 같은 세 가지 네트워크를 구성하였다. 네트워크 전체를 기준으로 한 주요 지표는 다음과 같다. 이때, 39개의 노드에 대해서 가능한 총 연결의 개수는 39×38 개, 즉 1482개이다. 또한, 후술할 예측 실험을 위해 데이터 유출(data leakage)을 방지하고자 위의 그래프는 2002년 1월 1일부터 2017년 12월 31일까지의 16개년 치의 데이터에 대한 국제 금융 시장 지수 선물의 하방 리스크에 대한 네트워크이며 각 유의 수준에 대해 네트워크 수준에서의 주요 지표는 다음과 같다.

Table 3. Network-level network measures of global financial market index EVaR network

Attribute	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
Total Connected Nodes	39	38	24
Total Connections	143	78	27
Network Density	9.65%	5.26%	1.82%
Maximum Degree Centrality	12	7	3
Minimum Degree Centrality	3	1	1
Average Degree Centrality	7.33	4.00	1.80

총 가능한 연결 수의 50%를 기준으로 할 때, 모든 네트워크는 매우 성긴(sparse) 네트워크로 분류할 수 있다. 이는 계산 기간이 길어질수록 유의미성이 매우 높은 정보 불확실성 기반의 인과관계가 남게 되는 ETE의 특성이 반영된 것으로 해석할 수 있다. 그리고 $\alpha = 0.1$ 유의 수준에서만 네트워크에서 제외된 노드, 즉 국제 금융 시장 지수 선물 중 인과 관계적으로 고립된 시장 지수 선물이 없음을 확인할 수 있다.

이에, 본 연구에서는 $\alpha = 0.1$ 을 기준으로 한 네트워크를 활용하여 후술할 실험을 진행하였다. $\alpha = 0.05$ 수준에서 구성된 국제 금융 시장 지수 선물의 하방 리스크 네트워크의 주요 네트워크 지표에 대해서 노드 수준의 분석 결과는 다음과 같다.

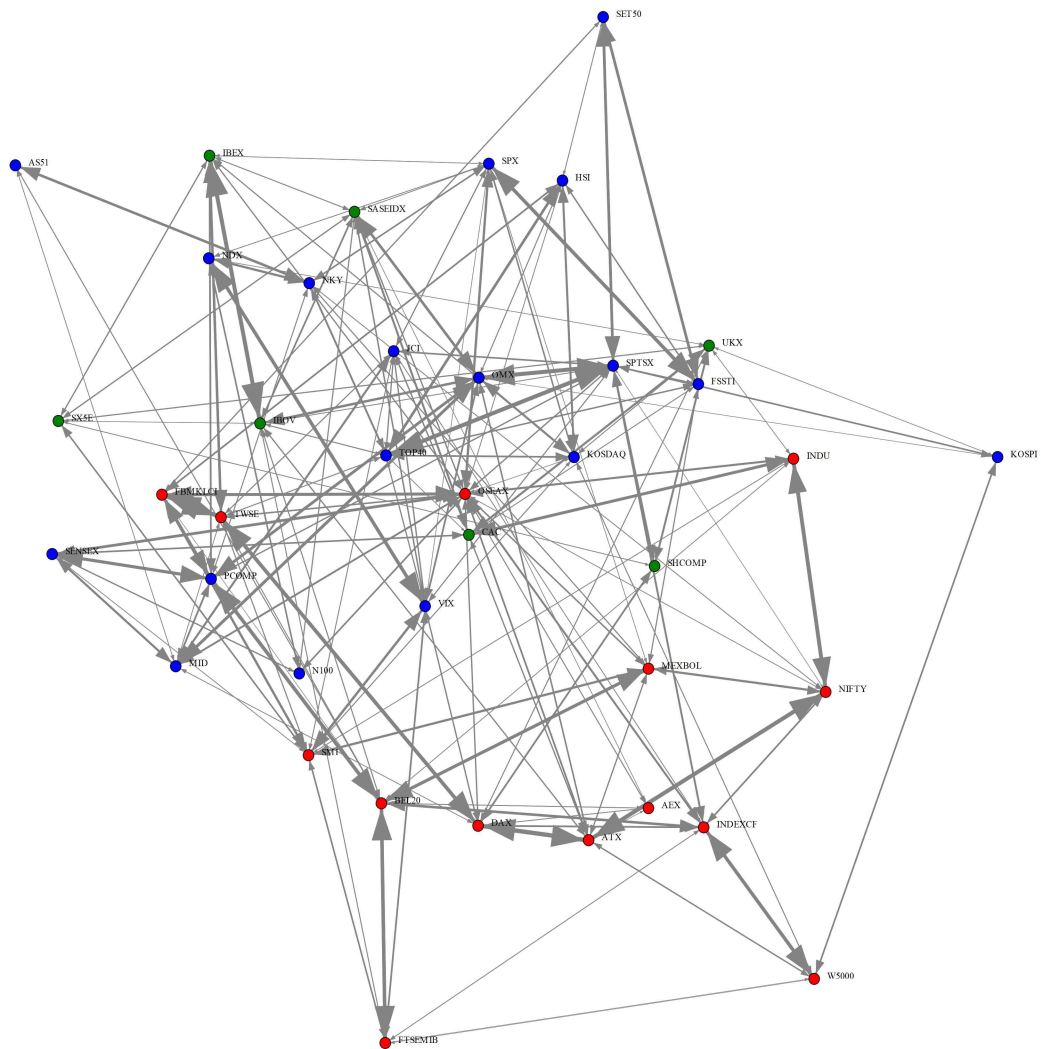


Figure 1. Global financial market indices' EVaR network

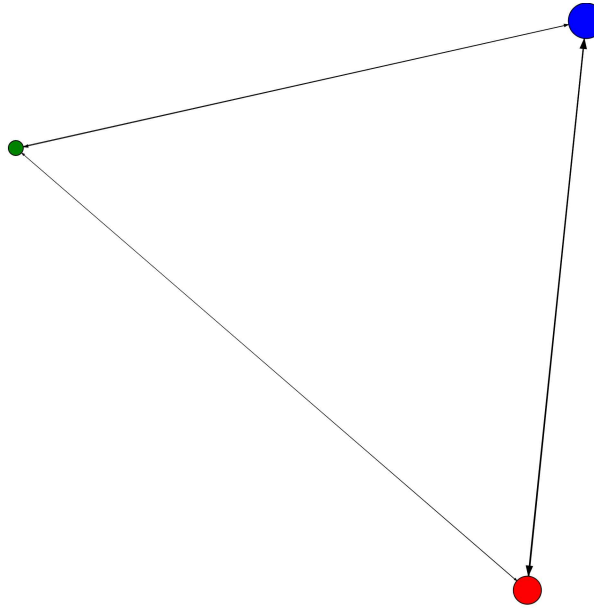


Figure 2. Clustering aggregation of information flow

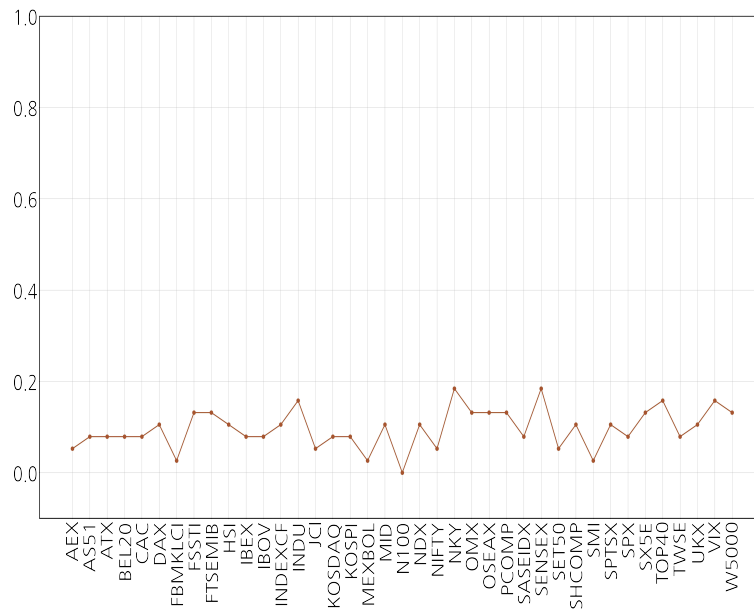


Figure 3. In-Degree of global financial market index EVaR network

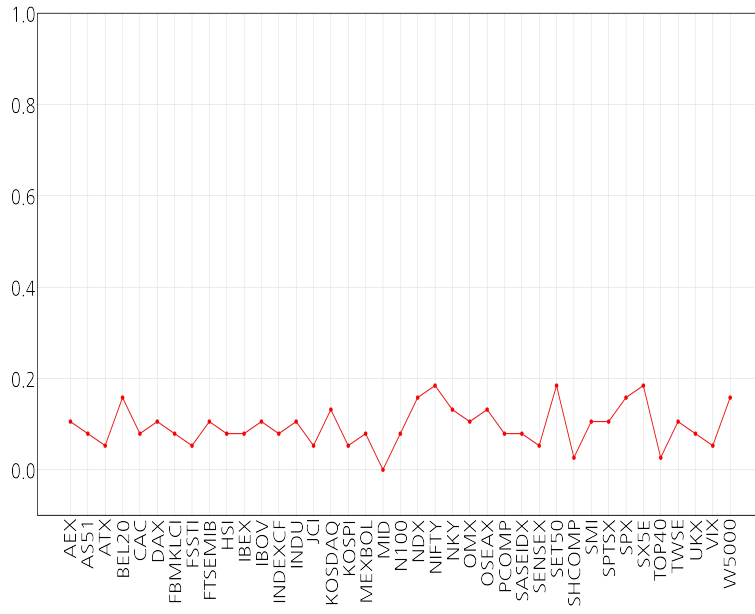


Figure 4. Out-degree of global financial market index EVaR network

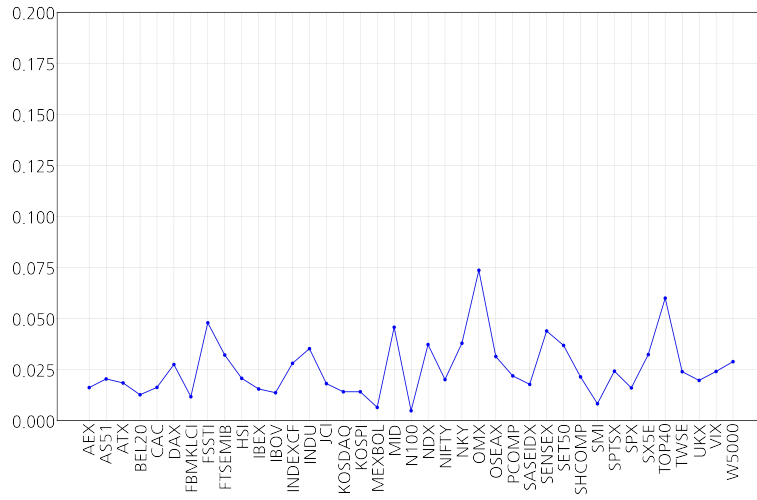


Figure 5. PageRank of global financial market index EVaR network

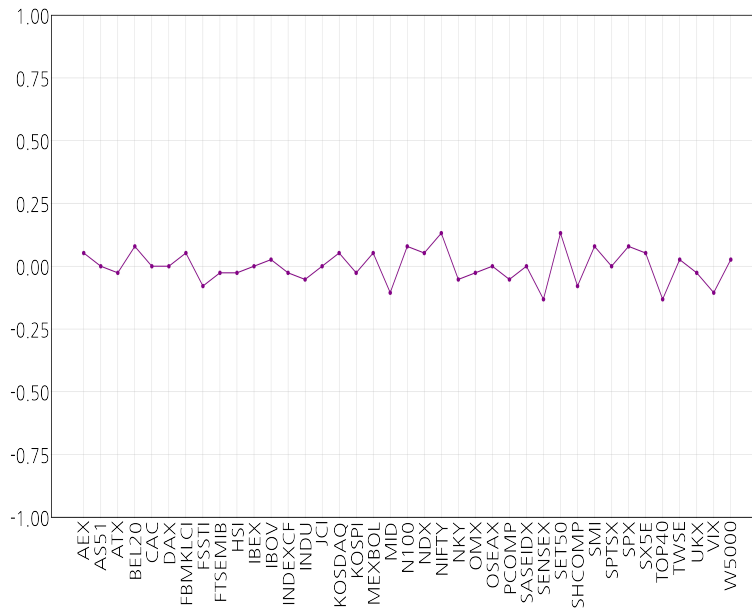


Figure 6. Net Degree (Out-Degree - In-Degree)
of global financial market index EVaR network

Table 4. Key index futures by centrality measures of EVaR network

Rank	Index	In-Degree	Index	Out-Degree	Index	PageRank	INDEX	Net-Degree
1	NKY	0.1842	SX5E	0.1842	OMXS	0.0736	SHCOMP	0.0526
2	SENSEX	0.1842	NIFTY	0.1842	TOP40	0.0600	FBMKLCI	0.0526
3	INDU	0.1579	SET50	0.1842	FSSTI	0.0479	FSSTI	0.0526
4	TOP40	0.1579	W5000	0.1579	MID	0.0458	FTSEMIB	0.0526
5	VIX	0.1579	NDX	0.1579	SENSEX	0.0439	DAX	0.0263
6	FSSTI	0.1316	BEL20	0.1579	NKY	0.0379	AEX	0.0263
7	FTSEMIB	0.1316	SPX	0.1579	NDX	0.0373	JCI	0.0263
8	OMXS	0.1316	NKY	0.1316	SET50	0.0369	NIFTY	0.0263
9	OSEAX	0.1316	OSEAX	0.1316	INDU	0.0353	SET50	0.0263
10	PCOMP	0.1316	KOSDAQ	0.1316	SX5E	0.0324	N100	0.0263

Figures 3-6과 Table 4에 요약된 4개 노드 수준의 네트워크 지표 결과에서는 다음과 같은 점을 주목할 수 있다. 우선 내향 연결 중심성의 경우 일본의 금융 시장 지수인 NIKKEI 225 (NKY)의 지수 선물과 인도의 금융 시장 지수인 BSE SENSEX (SENSEX)의 지수 선물이 하방 리스크 네트워크 내에서 가장 큰 내향 연결 중심성을 가짐을 확인했다. 또한 외향 연결 중심성의 경우 SX5E (유럽 EuroStoxx 50 지수), NIFTY (인도 Nifty 50 지수), SET50 (태국 SET 50 지수)의 지수 선물의 하방 리스크가 전체 지표에서 가장 큰 외향 연결 중심성을 가짐을 확인했다. 이는 정보 불확실성 감소 측면에서 상기한 세 개의 금융 시장 지수의 하방 리스크가 타 국가 금융 시장 지수의 하방 리스크에 인과관계를 가짐을 의미한다. 또한 그리고 Figure 6에서 외향 연결 중심성으로부터 내향 연결 중심성의 값을 뺀 값을 확인했을 때, 중국의 상해종합주가지수(SHCOMP)의 지수 선물의 값이 양의 값으로 가장 큰 값을 가짐에 따라 타 시장으로의 외향적 영향력이 상대적으로 가장 크다는 것을 확인할 수 있다. 동일하게 말레이시아 시장 지수 선물 (FBMKLCI), 싱가포르 STI 30 지수 선물 (FSSTI), 이탈리아 FTSE MIB 지수 선물 (FTSEMIB) 등 역시 가장 큰 외향 연결 영향력을 지닌다고 볼 수 있다. 하방 리스크 네트워크 관점에서

불 경우 비교적 동남아시아 시장 등 신흥 시장의 강세가 보이는 점을 주목할 수 있다. 단, 이는 선형적인 분석이 아니며 하방 리스크의 움직임에 대한 것이기 때문에 수익률 등 선형 금융 시장 네트워크 분석의 경우와 다른 결과가 도출될 수 있다.

4.3 네트워크 군집의 구성

3.2절에서 언급한 것과 같이 Dugué and Perez (2015)의 유향 네트워크에 대한 Louvain 알고리즘을 활용하여 다음과 같이 3개로 나누어진 네트워크 군집 결과를 Figure 1에서 확인할 수 있다.

구체적인 군집 분석 결과는 Table 5와 같다.

Table 5. Clustering results

Country	Financial Index	# of Cluster
AEX	네덜란드 금융 시장 지수 선물	2
AS51	호주 금융 시장 지수 선물	1
ATX	오스트리아 금융 시장 지수 선물	2
BEL20	벨기에 금융 시장 지수 선물	2
CAC	프랑스 금융 시장 지수 선물	3
DAX	독일 금융 시장 지수 선물	2
FBMKLCI	말레이시아 금융 시장 지수 선물	2
FSSTI	싱가포르 금융 시장 지수 선물	1
FTSEMIB	이탈리아 금융 시장 지수 선물	2
HSI	홍콩 금융 시장 지수 선물	1
IBEX	스페인 금융 시장 지수 선물	3
IBOV	브라질 금융 시장 지수 선물	3
INDEXCF	러시아 금융 시장 지수 선물	2
INDU	미국 다우존스 산업평균지수 선물	2
JCI	인도네시아 금융 시장 지수 선물	1
KOSDAQ	대한민국 코스닥 지수 선물	1
KOSPI	대한민국 코스피 지수 선물	1
MEXBOL	멕시코 금융 시장 지수 선물	2
MID	미국 S&P Mid Cap 지수 선물	1
N100	유럽 Euronext 100 지수 선물	1
NDX	미국 NASDAQ 100 지수 선물	1
NIFTY	인도 금융 시장 지수 선물	2
NKY	일본 금융 시장 지수 선물	1
OMXS	스웨덴 금융 시장 지수 선물	1
OSEAX	노르웨이 금융 시장 지수 선물	2
PCOMP	필리핀 금융 시장 지수 선물	1
SASEIDX	사우디아라비아 지수 선물	3
SENSEX	인도 금융 시장 지수 선물	1
SET50	태국 금융 시장 지수 선물	1
SHCOMP	중국 금융 시장 지수 선물	3
SMI	스위스 금융 시장 지수 선물	2
SPTSX	캐나다 금융 시장 지수 선물	1
SPX	미국 S&P 500 지수 선물	1
SX5E	유럽 EuroStoxx50 지수 선물	3
TOP40	남아프리카공화국 금융 시장 지수 선물	1
TWSE	대만 금융 시장 지수 선물	2
UKX	영국 금융 시장 지수 선물	3

VIX	미국 변동성 지수 선물	1
W5000	미국 월서5000 지수 선물	2

4.3 네트워크 군집을 이용한 국제 시장 금융 지수 선물의 하방 리스크 증감 예측 실험

4.3.1 실험 설명

본 연구에서는 기계 학습 모형을 활용하여 4.2절에서 생성된 군집을 이용하여 하방 리스크 증감 예측을 실시하였다. 실험 실시를 위해 다음과 같은 실험을 구상하였다. 이때 타깃 데이터는 왜곡할 수 있어 벤치마크 실험을 제외하고 예측에 직접적으로 포함하지 않았다.

- Experiment 1: 특정 타깃을 제외한 실험 내 전체 국제 시장 지수 선물 데이터열을 이용한 예측
- Experiment 2: 군집 1에 속한 국제 시장 지수 선물 데이터열을 이용한 예측
- Experiment 3: 군집 2에 속한 국제 시장 지수 선물 데이터열을 이용한 예측
- Experiment 4: 군집 3에 속한 국제 시장 지수 선물 데이터열을 이용한 예측
- Experiment 5: 군집 2와 3에 속한 국제 시장 지수 선물 데이터열을 이용한 예측

모형의 성능을 측정하는 실험 기간은 2.1절에서 상기한 것과 같이 2018년 1월 1일부터 2021년 12월 31일까지이며, 2002년 1월 1일부터 2017년 12월 31일까지의 데이터를 학습 데이터로 활용하였다.

학습을 위해 포함되는 데이터의 형태는 국제 시장 지수 선물의 일일 수익률의 형태로 포함이 되었으며, 하방 리스크 증감 분류를 위하여 타깃은 각각의 지수 선물에서 하방 리스크가 전일보다 증가하거나 변함이 없으면 1, 감소하면 0으로 변환하여 이진 분류 문제(binary classification problem)로 정의하였다. 그리고 실제 테스트 단계에서는 2018년 1월 1일부터 2021년 12월 31일 사이의 영업일 사이의 임의의 시점 t 에 대해서 $t-1$ 시점의 금융 시장 지수 선물 데이터만을 이용하여 예측을 시행하였다.

이러한 하방 리스크의 등락을 예측하기 위한 기계 학습 모형에는 그라디언트 부스팅 알고리즘 기반 모형인 XGBoost(Chen and Guestrin, 2016), Light gradient boosting machine(Ke et al., 2017, 이하 LightGBM), 그리고 CatBoost 알고리즘(Prokhorenkova et al., 2018)을 사용하였다. XGBoost, LightGBM, 그리고 CatBoost 모형은 본 연구에서 활용되는 정형 데이터에 대해서 매우 좋은 분류 성능을 보이는 트리 기반의 모형들로 알려져 있으며 이러한 점에 착안하여 본 연구에서는 해당 모형들을 사용하였다. 또한, 초모수(hyperparameter) 최적화를 위하여 Python의 Optuna (Yakiba et al., 2019) 패키지를 활용하였다.

4.3.2 실험 결과

실험 결과를 주요 성과 지표를 통해 요약하면 다음과 같다. 실험은 모형별로 무작위 난수 발생 시드를 변경하여 Optuna로 20회 학습을 진행한 후 해당 초모수들을 사용하여 각 100회를 진행하였으며 각 성과 평가 지표의 평균값 및 표준편차를 표기하였다. 또한 선정된 세 모형을 동일한 비율로 사용한 간접 투표 분류기(soft voting classifier) 결과도 실험을 평균적인 성능을 확인하고자 포함하였다. 이때 학습 성과를 평가하기 위한 지표로는 정확도(accuracy)를 사용하였다.

Table 5. Experiment results (average accuracy of 100 time simulations of 39 targets' EVaR difference)

ML Algorithm	Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 5
XGBoost	54.00%	53.18%	45.24%	43.94%	46.95%
LightGBM	53.94%	53.52%	45.04%	44.15%	46.75%
CatBoost	54.35%	53.59%	45.11%	44.22%	46.95%
SoftVotingClassifier (with equal-weights)	54.83%	54.96%	45.59%	44.90%	47.64%

그 결과, 전체 데이터열을 사용한 Experiment 1이 가장 좋은 결과를 보이는 것을 확인하였다. 단, 세 가지 모형을 동일 비중으로 조합한 간접 투표 분류기의 경우 Experiment 2가 더 좋은 결과를 보임을 확인했다. 또한 단독으로 사용한 벤치마크의 경우 일종의 동전 던지기(coin tossing) 문제에 가깝거나 이보다 더 낮은 예측률을 보임에 따라 EVaR의 경우 약형 효율적 시장가설(weak form of the efficient market hypothesis)로 해석할 수 있는 결과를 얻어낼 수 있음을 확인하였다. 즉, 무작위 행보(random walk)에 가까운 움직임을 보였다고 해석할 수 있다. 군집2와 군집3을 이용한 Experiment 3, 4의 결과가 가장 좋지 않은 것을 확인했으며 군집 1을 제외한 두 군집을 합쳐 실험한 Experiment 5의 경우에도 Experiment 1 및 Experiment 2보다 좋은 정확도를 얻지 못했다.

결론적으로, 모든 데이터열을 사용하는 것보다 정보 흐름 기반의 네트워크 기반의 국제 금융 시장 지수 선물의 하방 리스크 네트워크 내에서 이를 예측할 때 정확도를 기준으로 대부분 좋은 결과가 도출되는 것을 확인할 수 있었다. 이는 하방 리스크 증감에 대해서 더욱 적은 데이터열로 동등하거나 그 이상의 예측 정확도를 얻을 수 있음을 의미한다. 또한 특정 금융 시장 지수 선물의 하방 리스크가 속한 군집 외의 국가의 금융 시장 지수 선물 하방 리스크로 예측하는 어떠한 실험 결과(Experiment 3, Experiment 4, Experiment 5)보다 타깃이 되는 국제 금융 시장 지수 선물이 속하는 군집으로 예측하는 것이 정확도를 기준으로 모든 기계 학습 모형에서 좋은 예측 결과가 나오는 것을 확인할 수 있었다.

5. 결론

주식 시장은 경제학 및 금융의 중요한 연구 분야 중 하나로 광범위하게 연구됐다. 특히 주가와 수익률 데이터를 기반으로 증시를 분석하고 예측하는 연구는 금융 분야 내에서 가장 활발히 다루어지고 있는 주제이며 구체적으로 주가 및 시장 전반에 대한 리스크, 특히 하방 리스크에 대한 예측은 투자자들이 최적의 투자전략을 수립하기 위한 중요한 요소 중 하나이다.

본 연구에서는 전이 엔트로피 개념을 활용하여 주요 금융 시장 지수 선물의 하방 리스크에 대한 정보 흐름을 바탕으로 한 인과관계 네트워크를 구성하였다. 그리고 구성된 네트워크를 활용하여 금융 시장을 분석하였으며, 또한 정보 흐름의 존재 여부를 바탕으로 국제 금융 시장 지수 선물의 하방 리스크 증감에 대한 예측력이 개선될 수 있는지 확인하고자 하였다.

이를 위하여 정보 불확실성의 감소량을 측정하는 전이 엔트로피를 인과관계의 측정 지표로 상정하였다. 그리고 전이 엔트로피 측정 시 발생할 수 있는 유한크기 효과(finite size effect)를 조정하는 데 있어서 효과적인 지표인 효율적 전이 엔트로피를 활용하여 정보 흐름 네트워크를 구성하였으며 이를 이용하여 금융 지수 선물 간의 인과관계를 분석하고 국제 금융 시장 지수 선물의 하방 리스크 증감 예측에 활용하였다. 그 결과, 금융 시장 지수 선물을 효율적 전이 엔트로피를 이용한 인과관계 네트워크를 활용하여 금융 시장의 복잡계 네트워크 분석이 가능함을 확인하였고, 구성된 네트워크를 활용하여 국내 금융 시장 하방 리스크 증감 예측에 있어 더 적은 데이터열을 활용하여 개선된 예측 결과를 확인할 수 있었다.

참고문헌

- Ahmadi-Javid, Amir (2011). An information-theoretic approach to constructing coherent risk measures. St. Petersburg, Russia: Proceedings of IEEE International Symposium on Information Theory. pp. 2125–2127. doi:10.1109/ISIT.2011.6033932.
- Ahmadi-Javid, Amir (2012). "Entropic value-at-risk: A new coherent risk measure". Journal of Optimization Theory and Applications. 155 (3): 1105–1123. doi:10.1007/s10957-011-9968-2.
- Ahmadi-Javid, Amir (2012). "Addendum to: Entropic Value-at-Risk: A New Coherent Risk Measure". Journal of Optimization Theory and Applications. 155 (3): 1124–1128. doi:10.1007/s10957-012-0014-9.
- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019, July). Optuna: A next-generation hyperparameter optimization framework. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 2623–2631).
- Alareeni, B. A., & Hamdan, A. (2020). ESG impact on performance of US S&P 500-listed firms. Corporate Governance: The International Journal of Business in Society.
- Altomonte, C., Barba Navaretti, G., Di Mauro, F., and Ottaviano, G. (2011). Assessing competitiveness: how firm-level data can help (No. 2011/16). Bruegel Policy Contribution.

- Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques—Part II: Soft computing methods. *Expert systems with applications*, 36(3), 5932-5941.
- Baitinger, E., Kutsarov, I., Maier, T., Storr, M., and Wan, T. (2015). A holistic approach to diversification management: The diversification delta strategy applied to non-normal return distributions. *Credit and Capital Markets*, 48(1), 89–119.
- Bekiros, S., Nguyen, D. K., Junior, L. S., & Uddin, G. S. (2017). Information diffusion, cluster formation and entropy-based network dynamics in equity and commodity markets. *European Journal of Operational Research*, 256(3), 945-961.
- Boba, P., Bollmann, D., Schoepe, D., Wester, N., Wiesel, J., & Hamacher, K. (2015). Efficient computation and statistical assessment of transfer entropy. *Frontiers in Physics*, 3, 10.
- Bonanno, G., Caldarelli, G., Lillo, F., & Mantegna, R. N. (2003). Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E*, 68(4), 046130.
- Bouchaud, J. P. (2001). Power laws in economics and finance: some ideas from physics. *Quantitative finance*, 1(1), 105.
- Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
- Chen, H., Xiao, K., Sun, J., & Wu, S. (2017). A double-layer neural network framework for high-frequency forecasting. *ACM Transactions on Management Information Systems (TMIS)*, 7(4), 1-17.
- Chi, K. T., Liu, J., & Lau, F. C. (2010). A network perspective of the stock market. *Journal of Empirical Finance*, 17(4), 659-667.
- Cowan, A. R., and Sergeant, A. M. (2001). Interacting biases, non-normal return distributions and the performance of tests for long-horizon event studies. *Journal of banking and finance*, 25(4), 741–765.
- Dimpfl, T., and Peter, F. J. (2013). Using Transfer Entropy to Measure Information Flows between Financial Markets. *Studies in Nonlinear Dynamics and Econometrics*, 17(1), 85–102.
- Dugué, N., & Perez, A. (2015). Directed Louvain: maximizing modularity in directed networks (Doctoral dissertation, Université d'Orléans).
- Faes, L., Nollo, G., & Porta, A. (2013). Compensated transfer entropy as a tool for reliably estimating information transfer in physiological time series. *Entropy*, 15(1), 198-219.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2), 383-417.
- Fama, E. F. (1991). Efficient capital markets: II. *The journal of finance*, 46(5), 1575-1617.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
- Gokmenoglu, K. K., & Fazlollahi, N. (2015). The interactions among gold, oil, and stock market: Evidence from S&P500. *Procedia Economics and Finance*, 25, 478-488.
- Graham, M., Kiviaho, J., & Nikkinen, J. (2013). Short-term and long-term dependencies of the S&P 500 index and commodity prices. *Quantitative Finance*, 13(4), 583-592.
- Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424-438.
- Hacine-Gharbi, A. and Ravier, P.(2018), A binning formula of bi-histogram for joint entropy estimation using mean square error minimization, *Pattern Recognition Letters*, 101(1): 21-28.
- Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2018). Stock price prediction using support vector regression on daily and up to the minute prices. *The Journal of finance and data science*, 4(3), 183-201.
- Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, 124, 226-251.
- Jabeur, S. B., Mefteh-Wali, S., & Viviani, J. L. (2021). Forecasting gold price with the XGBoost algorithm and SHAP interaction values. *Annals of Operations Research*, 1-21.
- Jang, S., Yi, E., Kim, W. C., and Ahn, K. (2019). Information Flow between Bitcoin and Other Investment Assets. *Entropy*, 21(11), 1116.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1-2), 307-319.
- Kim, M., Newth, D., & Christen, P. (2016). Macro-level information transfer in social media: Reflections of crowd phenomena. *Neurocomputing*, 172, 84-99.
- Kumar, D., Meghwani, S. S., & Thakur, M. (2016). Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets. *Journal of Computational Science*, 17, 1-13.
- Kumar, S., & Deo, N. (2012). Correlation and network analysis of global financial indices. *Physical Review E*, 86(2), 026101.
- Kwon, O., and Yang, J. S. (2008). Information Flow between Composite Stock Index and Individual Stocks. *Physica A: Statistical Mechanics and Its Applications*, 387(12), 2851–2856.
- Kwon, Y. K., Choi, S. S., & Moon, B. R. (2005, June). Stock prediction based on financial correlation. In *Proceedings of the 7th annual conference on Genetic and evolutionary computation* (pp. 2061-2066).
- Lausberg, C., Lee, S., Müller, M., Oertel, C., and Schultheiß, T. (2020). Risk measures for direct real estate investments with

- non-normal or unknown return distributions. *Zeitschrift für Immobilienökonomie*, 6(1), 3–27.
- Lim, K., Kim, S., and Kim, S. Y. (2017). Information Transfer Across Intra/Inter-Structure of CDS and Stock Markets. *Physica A: Statistical Mechanics and Its Applications*, 486, 118–126.
- Linden, M. (2001). A model for stock return distribution. *International Journal of Finance and Economics*, 6(2), 159–169.
- Liu, J., Wu, C., & Li, Y. (2019). Improving financial distress prediction using financial network-based information and GA-based gradient boosting method. *Computational Economics*, 53(2), 851–872.
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- Lundberg, S. M., & Lee, S. I. (2017). Consistent feature attribution for tree ensembles. arXiv preprint arXiv:1706.06060.
- Lundberg, S. M., Erion, G. G., & Lee, S. I. (2018). Consistent individualized feature attribution for tree ensembles. arXiv preprint arXiv:1802.03888.
- Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of economic perspectives*, 17(1), 59–82.
- Mandelbrot, B., and Taylor, H. M. (1967). On the distribution of stock price differences. *Operations research*, 15(6), 1057–1062.
- Mantegna, R. N., & Stanley, H. E. (1999). *Introduction to econophysics: correlations and complexity in finance*. Cambridge university press.
- Marschinski, R., and Kantz, H. (2002). Analyzing the Information Flow between Financial Time Series. *European Physical Journal B – Condensed Matter and Complex Systems*, 30(2), 275–281.
- Noh, J. D. (2000). Model for correlations in stock markets. *Physical Review E*, 61(5):5981.
- Oatley, T., Winecoff, W. K., Pennock, A., & Danzman, S. B. (2013). The political economy of global finance: A network model. *Perspectives on Politics*, 11(1), 133–153.
- Plerou, V., Gopikrishnan, P., Gabaix, X., Amaral, L. A. N., and Stanley, H. E. (2001). Price fluctuations, market activity and trading volume. *Quantitative Finance*, 1(2), 262.
- Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. A. N., Guhr, T., & Stanley, H. E. (2002). Random matrix approach to cross correlations in financial data. *Physical Review E*, 65(6), 066126.
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulina, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems*, 31.
- Sandoval, L. (2014). Structure of a Global Network of Financial Companies Based on Transfer Entropy. *Entropy*, 16(8), 4443–4482.
- Sansa, N. A. (2020). The Impact of the COVID-19 on the Financial Markets: Evidence from China and USA. *Electronic Research Journal of Social Sciences and Humanities*, 2.
- Sensoy, A., Sobaci, C., Sensoy, S., and Alali, F. (2014). Effective Transfer Entropy Approach to Information Flow between Exchange Rates and Stock Markets. *Chaos, Solitons and Fractals*, 68, 180–185.
- Schreiber, T. (2000). Measuring information transfer. *Physical review letters*, 85(2):461.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423.
- Shapley, L. S. (1953). A value for n-person games, *Contributions to the Theory of Games*, 2, 307–317.
- Shynkevich, Y., McGinnity, T. M., Coleman, S. A., Belatreche, A., & Li, Y. (2017). Forecasting price movements using technical indicators: Investigating the impact of varying input window length. *Neurocomputing*, 264, 71–88.
- So, M. K., Chu, A. M., & Chan, T. W. (2021). Impacts of the COVID-19 pandemic on financial market connectedness. *Finance Research Letters*, 38, 101864.
- Sun, X., Liu, M., & Sima, Z. (2020). A novel cryptocurrency price trend forecasting model based on LightGBM. *Finance Research Letters*, 32, 101084.
- Turchenko, V., Beraldi, P., De Simone, F., & Grandinetti, L. (2011, September). Short-term stock price prediction using MLP in moving simulation mode. In *Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems (Vol. 2, pp. 666–671)*. IEEE.
- Vicente, R., Wibral, M., Lindner, M., & Pipa, G. (2011). Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *Journal of computational neuroscience*, 30(1), 45–67.
- Weng, B., Ahmed, M. A., & Megahed, F. M. (2017). Stock market one-day ahead movement prediction using disparate data sources. *Expert Systems with Applications*, 79, 153–163.
- World Bank (2021). Market capitalization of listed domestic companies (current US\$) | Data. data.worldbank.org. Retrieved 2021-09-20.
- Young, M. S. (2008). Revisiting non-normal real estate return distributions by property type in the US. *The Journal of Real Estate Finance and Economics*, 36(2), 233–248.
- Young, M. S., Lee, S. L., and Devaney, S. P. (2006). Non-normal real estate return distributions by property type in the UK. *Journal of Property Research*, 23(2), 109–133.
- Yue, P., Cai, Q., Yan, W., and Zhou, W. X. (2020). Information Flow Networks of Chinese Stock Market Sectors. *IEEE Access*, 8, 13066–13077.
- Yue, P., Fan, Y., Batten, J. A., and Zhou, W. X. (2020). Information Transfer between Stock Market Sectors: A Comparison between the USA and China. *Entropy*, 22(2), 194.

- Zhang, D., Hu, M., & Ji, Q. (2020). Financial markets under the global pandemic of COVID-19. *Finance research letters*, 36, 101528.
- Zunino, L., Tabak, B. M., Figliola, A., Pérez, D. G., Garavaglia, M., & Rosso, O. A. (2008). A multifractal approach for stock market inefficiency. *Physica A: Statistical Mechanics and its Applications*, 387(26), 6558-6566.