

GARCH-LSTM 결합모형을 이용한

펀드 위험등급 예측

김아람, 박민호

승실대학교 IT 융합학과

Fund Risk Rating Prediction

Using GARCH-LSTM

Kim A Ram, Park Min ho

Soongsil Univ

요 약

본 논문은 LSTM 모델에 조건부 이분산성을 적용한 금융 시계열 분석 모형인 GARCH(1,1) 모형을 결합한 GARCH(1,1)-LSTM 통합모형으로 차기 결산일 시점의 펀드 변동성을 예측하였고, 예측된 변동성을 SVM 분류기에 넣어 펀드의 예상 위험등급을 도출했다. 선제적 펀드 데이터 예측을 통해 펀드 판매사가 능동적으로 불완전 판매 리스크를 대응할 수 있도록 큰 도움을 주고 나아가 펀드 산업에 대한 금융소비자의 신뢰도 향상에 기여하고자 한다.

I. 서론

급변하는 금융시장 속 다양하고 복잡한 금융상품 출시로 새로운 투자기회가 확대됐지만 금융 소비자와 금융기관의 정보격차 확대에 따른 잠재적 위험도 커졌다. 2019년 라임 사태와 2020년 옵티머스 사태 등 연이은 대형 금융범죄는 금융 소비자에게 막대한 피해를 줬고 금융소비자보호 강화와 금융업자에 대한 감독이 확대되는 계기가 됐다.

21년 3월 금융소비자보호법 시행 이후 금융상품 판매업자는 금융상품 불완전 판매가 발생하지 않도록 설명 의무를 충실히 준수해야 한다. 일환으로 펀드의 정보 변경 누락을 방지하고자 펀드평가사로부터 수시공시대행, 위험등급 모니터링 시스템을 도입하여 불완전 판매 리스크에 대응하고 있다.

하지만 자산운용회사의 공시 정보는 사후적이라는 한계 있다. 또한 증권신고의 효력이 발생하는 당일 공시로 지정하고 있기 때문에 판매시점과 시간 격차가 생길 수 있다. 이에 펀드 판매업체에서 선제적으로 공시 변경 여부를 예측하고자 하는 수요가 확대되고 있다. 변동성 지표는 위험관리의 핵심 변수이기 때문에 사전적으로 파악하는 것이 중요하다. 본 연구에서는 펀드 주간 수익률을 이

용하여 결산일 도래 시점의 변동성을 추정하고 추정한 변동성으로부터 펀드 위험등급을 사전 예측하고자 한다.

II. 이론적 배경

펀드 위험등급은 금융소비자들이 펀드에 가입할 때 참고해야 하는 중요한 변동성 지표다. 2016년 1월 기존 5단계 이상에서 6단계 이상으로 투자위험등급이 세분화되었고, 설정 후 3년이 경과한 펀드에 대해서는 수익률 변동성을 기준으로 위험등급을 산정하고 있다. 금융감독원 가이드라인에 따르면 변동성은 매년 결산시점에 측정하며, 해당 결산일 기준 이전 3년(156주)간 펀드의 주간 수익률의 연환산 표준편차와 등급 기준표상의 표준편차 상한치를 비교하여 등급을 산출한다.

<표-1> 위험등급 기준표

등급	1(고위험)	2	3	4	5	6(저위험)
표준편차	25%초과	25%이하	15%이하	10%이하	5%이하	0.5%이하

주) (연환산 표준편차 기준 예시, 금감원 가이드)

III. 연구방법

3.1 연구자료

목표 변수인 펀드의 위험등급은 자산운용회사가 투자설명서 상에 명시한 위험등급이며, 수집 기간은 펀드 위험등급이 5단계에서 6단계 체계로 바뀐 2016년 6월 24일부터 2022년 12월 31일까지다. 주간수익률과 결산관련 데이터는 펀드평가사인 한국펀드평가의 자료를 이용했다. 대상 펀드는 2022년 12월 31일 기준 운용중인 국내주식형 운용펀드이며 펀드 수는 4,025개이다.

3.2 연구모형

연구는 두 단계로 진행한다. 변동성 예측 단계에서는 시계열 분석에서 주목받고 있는 LSTM 모형과 LSTM 모형에 이분산성 시계열 모형인 GARCH(1,1)을 결합한 GARCH-LSTM 모형으로 변동성을 예측한 후 RMSE(평균제곱근 오차)로 두 모형의 예측 정확도를 비교하였다. 위험등급 예측 단계에서는 SVM 알고리즘을 사용했다.

3.2.1. LSTM

LSTM 모형은 기존의 RNN이 출력과 먼 위치에 있는 정보를 기억할 수 없다는 단점을 보완한 모형으로 역전과 기법을 통해 장단기 기억을 학습한다. 본 연구에서는 금융 시계열 자료를 효과적으로 학습하기 위해 LSTM 모형을 이용하여 펀드의 변동성을 예측하였다.

3.2.2. GARCH

고전적 회귀 분석 모형은 서로 다른 관측 시점의 오차항들의 분포는 동일한 분산을 갖는다는 전제를 갖고 있다. 하지만 실제 금융 시계열은 변동성 자체도 시간에 흐름에 따라 바뀌는 조건부 이분산성을 갖는다. 조건부 이분산성 관련 연구에서 널리 사용되는 모형은 Bollerslev (1986)가 제안한 GARCH 모형으로, 수익률의 변동성이 과거의 변동성에 의존하여 결정되도록 모형화하였다. 본 문에서는 펀드의 주간 수익률을 단일 변량 변수로 GARCH(1,1) 모형에 넣어 T+1 시점의 변동성을 예측하였다.

3.2.3. GARCH - LSTM

LSTM 모형에 GARCH(1,1) 모형을 결합하여 금융 시계열의 특성을 반영하고자 한다. GARCH - LSTM 모형에서는 주 단위의 1개월 표준편차와 GARCH(1,1)모형에서 추정된 모수를 LSTM 입력 변수로 이용했다. 출력 변수는 한 달 후의 1개월 표준편차이다. 본 연구에서는 Python을 통해 모델링 했으며, 하이퍼 파라미터인 epoch은 200으로, batch size는 32으로 지정했다.

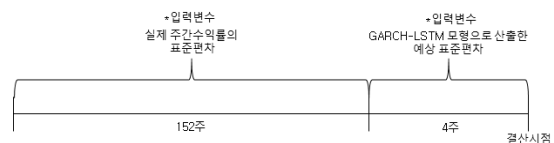
156주간의 변동성 예측 모형으로 입력변수의 최소 ROW는 156개이다. 훈련,테스트,검증 셋으로 나누어 교차 검증 하기에는 데이터의 수가 적기

때문에 데이터를 K개로 분할하고, K개의 모델을 만들어 K-1개 분할에서 훈련하고 나머지 분할에서 평가하는 K-fold 교차 검증을 이용해 모델을 평가했다. LSTM 모형과 GARCH - LSTM 모형의 예측력 비교를 위한 손실함수로는 평균 제곱근 오차(RMSE)를 이용하였다. RMSE는 오차의 제곱을 평균으로 나눈 값의 제곱근이며 RMSE가 0에 가까울수록 정확도가 높다고 할 수 있다.

3.2.4. SVM

위험등급 예측 모형으로 분류를 위해 사용되는 대표적 머신러닝 알고리즘 SVM을 사용했다. 입력 변수는 결산일 3년 전 시점으로부터의 152주 주간수익률의 연환산 표준편차와 GARCH-LSTM 모형으로 예측한 4주 연환산 표준편차이다. 목표 변수는 1~6 단계의 펀드 위험등급이다. 본 연구에서는 Python을 통해 모델링 했으며, 가장 많이 사용되는 RBF kernel을 사용했고 하이퍼 파라미터인 gamma와 c는 디폴트 값을 이용했다. 검증은 변동성 예측 모형 검증과 같은 K-fold 교차 검증을 사용했다.

<그림-1> SVM 모형 입력 변수 구성



IV. 결론

변동성 예측 모형으로는 LSTM 단일 모형보다 GARCH - LSTM 결합모형의 손실 값이 적었다. 특히 코로나 팬데믹 같은 고변동성 구간에서 LSTM 보다 GARCH - LSTM의 변동성 예측이 정확했다. 위험등급 예측 단계의 SVM 모형에서는 예측 정확도 88.34%로 나왔다. 본 연구는 사전적 의미가 있는 위험등급 예측 모형을 제안 하였으며 위 제시된 내용에 나아가 LSTM 및 SVM의 시뮬레이션 구성 확대 및 입력변수 추가를 통해 예측 정확도를 높일 계획이다.

참고문헌

[1] Jun Han Deok, "A Study on the Main Contents and Improvement Tasks of Financial Consumer Protection Act", 2021

[2] Bollerslev, T.(1986). Generalized autoregressive conditional heteroscedasticity, Journal of Econometrics, 31, 307-327

[3] S. Y. Kim and Y. H. Lee, "Comparison of a Class of Nonlinear Time Series models (GARCH, IGARCH, EGARCH)", 2006

[4] Juyeon park and In Kwon Yeo, "Prediction Value Estimation in Transformed GARCH Models", 2009

[5] SeRin Park and Changryong Baek, "On multivariate GARCH model selection based on risk management", 2014

[6] Kim Kyoung Jin, "Analysis of the Predicted Volatility of Imported Meat Price based on GARCH and LSTM"