

2022 경제학부 학술제 보고서

시계열 통계모델과 머신러닝 모델의 변수 유의성 판단 비교 및 시사점

Team Deepconomics

B782040 정성운

C035006 강서연



Abstract

경제학에서 중요한 이슈 중 하나는 엄밀한 가정과 증명을 통해 우리 사회에서 상호영향력을 주고받는 변수들 간의 관계를 파악하고, 이들을 기반으로 의사결정에 있어 이론적 근거의 역할을 하는 것이다. 이러한 절차를 위해, 경제통계학 분야에선 오래 전부터 다양한 문제를 해결하기 위한 방법들이 고안되었다. 그러나 동시에 머신러닝 분야 또한 데이터 패턴 파악을 통한 예측의 '부산물'로 변수의 중요도를 파악할 수 있다. 본 팀은 경제학에서 전통적으로 사용되어 왔던 시계열 모델과 머신러닝 모델을 비교하여 각 모델이 판단한 변수의 유의성과 예측력을 비교하고, 이로부터 우리 경제학이 한 발자국 나아갈 수 있는 방법을 고안한다.

1. 개요

컴퓨터 하드웨어 산업의 눈부신 발전과 더불어 막강한 계산력을 요구해 이론상으로만 존재했던 머신러닝 모델들이 세상에 나오고 있다. 바둑을 두는 알파고부터 시작해 최근 '그림을 그리는 AI'로 대중을 놀라게 했던 Stable-Diffusion 모델들처럼 이제 더 이상 머신러닝 기술은 공상과학 소설의 꿈 같은 얘기가 아닌 어느새 우리 현실 속에 자리잡고 있다.

이러한 머신러닝의 영향력은 경제학 영역에서도 발휘되고 있다. 특정 변수에 대한 요인분석과 이를 통한 미래 예측은 경제학이 지금까지 풀어왔던 문제였고, 이는 머신러닝이 특화되어 있는 분야이기 때문이다. 그러나 기존에 많이 활용되어 왔던 경제학의 예측 모델들 또한 같은 문제를 해결하기 위해 고안된 모델들로, 두 분야의 모델은 같은 목적을 가진 다른 방법론이라고 할 수 있다.

본 팀은 날이 갈수록 국제사회에서 중요해지고 있는 외국인 투자금액이라는 변수에 대해 영향을 미치는 변수들을 교과서와 논문에 기반해 선정하고, 이들을 각각 전통 통계모형과 머신러닝 모델로 학습하여 각 모델이 중요하게 여긴 변수를 비교해보고자 한다.

2. 데이터셋 구축

데이터 구축 단계는 분석에 있어 가장 기초적이면서 중요한 단계라고 할 수 있다. 본 팀은 교과서와 논문 등에서 언급되는 외국인 투자액에 영향을 미치는 변수들을 파악하고 동시에 논문과 교과서에는 언급되지 않았지만 본 팀이 생각하기에 영향을 미칠 것으로 예상되는 변수들도 추가적으로 넣어주었다.

다만 데이터를 수집할 때, 머신러닝 모델의 특성상 많은 데이터가 필요했기에, 적은 수의 데이터만 확보 가능한 변수는 수집 대상에서 제외하였다. 데이터 수집 기준엔 동시에 본 팀은 일단위와 월단위, 연단위가 있었지만, 학습 속도와 확보되는 데이터 수를 고려하여 가장 적합하다고 판단한 월단위 데이터를 수집하기로 결정하였다.

데이터를 주로 수집은 주로 KOSIS 국가통계포털¹과 e-나라지표²를 이용하였으며, 그 외 기업에서 제공하는 자료들은 해당 기업 홈페이지를 방문하여 확보하였다.

최종적으로 구축된 Tabular Data는 2004년 8월부터 2022년 9월까지 총 218건의 데이터와 종속변수를 포함하여 27개의 Feature로 이루어졌다.

선정된 독립변수들과 제외된 변수의 목록은 아래와 같다.

| | |
|--------|------------------------------------------------------------------------------------------------|
| 독립변수 | MSCI지수, FTSE지수, 외환보유액, 외평채 스프레드, 국고채, 회사채 CD금리, 콜금리, 기준금리, 미연준금리, AAI 투자심리지수, VIX |
| 종속변수 | 외국인 증권투자 현황 |
| 제외된 변수 | CDS 프리미엄, CRS금리, 물가 상승률, 북한 도발 여부 자국 주요 기업의 긍정적 뉴스 여부(혹은 부정적 뉴스 여부) 국가 신용 등급, 가산금리투자심리지수 |

Table 1 선정된 변수와 제외 변수 목록, AAI 투자심리지수와 VIX지수는 본 팀이 추가적으로 유의미하다고 판단하여 선정하였다. 제외된 변수들은 유의미할 수 있으나 데이터 수 부족이나 수집에 있어 비효율적이라고 판단되어 제외되었다.
외평채 스프레드와 국고채 같은 경우 기간에 따라 2개 이상의 변수로 표현되었다. Table 2를 참고할 것.

¹ KOSIS국가통계포털 : <https://kosis.kr/index/index.do>

² e-나라지표 : <https://www.index.go.kr/main.do?cate=1>

최종적으로 구축된 데이터셋의 예시는 아래와 같다.

| Date | 외국인 보유 금액 | 외환보유 액 | FTSE | MSCI | 외원채 _2013 | 외원채 _2014_04 | 외원채 _2014_09 | 외원채 _2015 | 외원채 _2016 | 외원채 _2019 | - | 국고채_3 년 | 국고채_5 년 | 국고채_10 년 | 회사채_3 년 | CD | 금금 리 | 기준금 리 | 연준금 리 | AAII | VIX |
|-------------|--------------|-----------|--------|--------|--------------|-----------------|-----------------|--------------|--------------|--------------|---|------------|------------|-------------|------------|------|---------|----------|----------|---------|--------|
| 200408 월 | 165.6 | 1705.0 | 103.86 | 237.37 | 69.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | - | 3.82 | 4.01 | 4.33 | 4.38 | 3.70 | 3.60 | 3.50 | 1.429 | 2.975 | 15.290 |
| 200409 월 | 173.1 | 1745.0 | 107.45 | 245.87 | 75.0 | 0.0 | 97.0 | 0.0 | 0.0 | 0.0 | - | 3.59 | 3.75 | 4.14 | 4.10 | 3.52 | 3.52 | 3.50 | 1.605 | 15.940 | 13.340 |
| 200410 월 | 171.3 | 1784.0 | 107.08 | 243.79 | 56.0 | 0.0 | 73.0 | 0.0 | 0.0 | 0.0 | - | 3.54 | 3.67 | 4.05 | 4.01 | 3.52 | 3.52 | 3.50 | 1.760 | 19.500 | 16.270 |
| 200411 월 | 179.1 | 1926.0 | 111.54 | 252.75 | 62.0 | 0.0 | 85.0 | 0.0 | 0.0 | 0.0 | - | 3.42 | 3.53 | 3.96 | 3.86 | 3.42 | 3.34 | 3.25 | 1.932 | 34.900 | 13.240 |
| 200412 월 | 179.7 | 1991.0 | 113.59 | 256.40 | 66.0 | 0.0 | 82.0 | 0.0 | 0.0 | 0.0 | - | 3.28 | 3.39 | 3.85 | 3.73 | 3.39 | 3.27 | 3.25 | 2.156 | 34.600 | 13.290 |
| - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 202205 월 | 695.3 | 4477.0 | 392.18 | 814.54 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | - | 3.02 | 3.23 | 3.30 | 3.78 | 1.77 | 1.51 | 1.75 | 0.758 | -27.175 | 26.190 |
| 202206 월 | 593.0 | 4383.0 | 340.45 | 706.72 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | - | 3.48 | 3.65 | 3.64 | 4.23 | 1.99 | 1.75 | 1.75 | 1.187 | -26.960 | 28.709 |
| 202207 월 | 629.9 | 4386.0 | 359.64 | 748.12 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | - | 3.24 | 3.30 | 3.30 | 4.13 | 2.41 | 1.97 | 2.25 | 1.654 | -19.500 | 21.330 |
| 202208 월 | 629.9 | 4364.0 | 358.50 | 744.58 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | - | 3.25 | 3.30 | 3.32 | 4.23 | 2.79 | 2.26 | 2.50 | 2.330 | -7.825 | 25.870 |
| 202209 월 | 593.0 | 4168.0 | 312.36 | 649.07 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | - | 3.90 | 3.94 | 3.90 | 4.90 | 3.01 | 2.53 | 2.50 | 2.580 | -33.520 | 31.620 |

Table 2 구축된 데이터셋의 Data Frame, 일부 변수는 원활한 Display를 위해 ...으로 생략되었다.

3. 모델 선택

본 팀의 목표는 머신러닝 모델과 전통적으로 사용되어 왔던 시계열 모형의 변수 중요도를 파악하는 것이다. 때문에 다음과 같은 제한 사항을 충족하는 모델들을 선정할 필요가 있었다.

첫째, 결과에 대한 설명 가능해야 할 것. 모델이 판단하는 변수의 중요도를 확인할 수 없는 모델은 우리의 목적에 부합하지 않았으므로, 모델 후보에서 제외되었다. 이 조건에서 제외된 대표적인 방법론이 딥러닝(Deep Learning)으로, 딥러닝의 학습 방법상 정답값과 예측값의 Loss를 최소화 시키는 방향으로 신경망 내부의 Weight들을 조절해나가는 방식으로 이 Weight들만 보고서 직관적인 해석을 얻기는 매우 까다롭다. 물론 Computer Vision 영역에서 특히 주목 받고 있는 XAI(Explainable AI)는 이미지 관련 작업에 한해 직관적으로 이해 가능한 설명을 제공해주지만 XAI 또한 본 데이터셋에 부합하지 않으므로 활용할 수 없었다.

둘째, 적은 양의 데이터로도 충분히 학습가능할 것. 확보된 데이터는 2004년 8월부터 2022년 9월까지 총 218건의 데이터로 이루어져있다. 데이터는 많으면 많을수록 좋지만 확보된 데이터가 적은 만큼, 적은 데이터의 수로도 원활한 학습이 가능한 모델이어야 했다.

셋째, 단변량이 아닌 다변량의 모델이어야 할 것. 본 팀의 목적은 각 분야의 모델이 여러 개의 변수 중 어떤 변수를 중요하게 봤는지 확인하는 것이다. 그에 따라 하나의 변수

를 보고 판단하는 모델이 아닌 여러 변수의 영향력을 고려할 수 있는 모델을 선택 했어야 했다.

위와 같은 조건 하에 본 팀은 전통적 통계 모형으로 VAR을, 머신러닝 모델에선 XGBoost를 선택하였다. 두 알고리즘 모두 각 분야에서 대표적인 알고리즘으로, 현업에서도 주어진 조건 하에 적합하다고 판단될 시, 널리 사용되고 있다.

VAR(Vector Auto Regression)은 예측할 변수의 과거값뿐만 아니라 예측할 변수와 의존성이 있는 다른 변수들까지 고려하는 모델이다. VAR을 선택한 이유는 본 팀이 구축한 데이터셋이 일변량 시계열이 아닌 다변량 시계열 데이터로 이루어져 있으며, 동시에 목적 또한 변수들 사이의 의존성과 상관관계를 파악해야 했기 때문이다. 때문에 다변량 시계열 모델 중 가장 기본이 되는 VAR을 선택하였다. 추가적으로, 정상성을 충족해야 한다는 조건이 마침 활용될 머신러닝 모델인 XGBoost의 조건과 일치하여 동등한 조건 하의 비교가 가능했다.

XGBoost(Extream Gradient Boosting)는 Boosting기법의 대표적인 알고리즘으로 기존의 Gradient Boost 기법에서 병렬 학습이 가능하도록 구현했다. CART 기반의 Regression 문제와 Classification 문제에서 뛰어난 성적을 보이며 Tabular Data를 다루는 업무에서 주로 사용되는 모델이다. 다만 Tree 기반의 모델이기 때문에 데이터의 주어진 범위 내에서의 예측 능력은 뛰어나지만, 데이터가 가지고 있는 범위 외의 예측을 해야하는 경우 예측 성능이 떨어진다는 단점이 있다. 때문에 시계열 데이터를 XGBoost를 통해 예측하는 경우, 정상성을 띄도록 데이터를 가공해주어야 한다는 주의점이 있다.

위의 두 모델 모두 200여개 정도 되는 적은 데이터에 대해 훌륭한 예측력과 결과에 대한 Input의 해석이 가능한 모델로 본 팀의 목적에 적절히 부합하다고 판단하였다. 동시에 두 모델은 분석을 함에 있어 동일한 데이터 전처리를 요구하여 같은 조건 하의 예측력 비교와 변수 선택 차이를 알아보는 작업에 있어 적합할 것으로 기대했다.

4. Data EDA

수집이 완료된 데이터들을 목적에 맞게 시각화하여 데이터셋 자체가 내포하고 있는 의미를 찾아내는 과정인 **EDA**(Exploratory Data Analysis)를 선행적으로 진행하였다. 시간에 따른 변수의 변화나 변수간 상관 관계를 시각화 함으로써, 모델링을 하기 앞서 해야할 전처리에 대한 방향성을 잡기 위함이다.

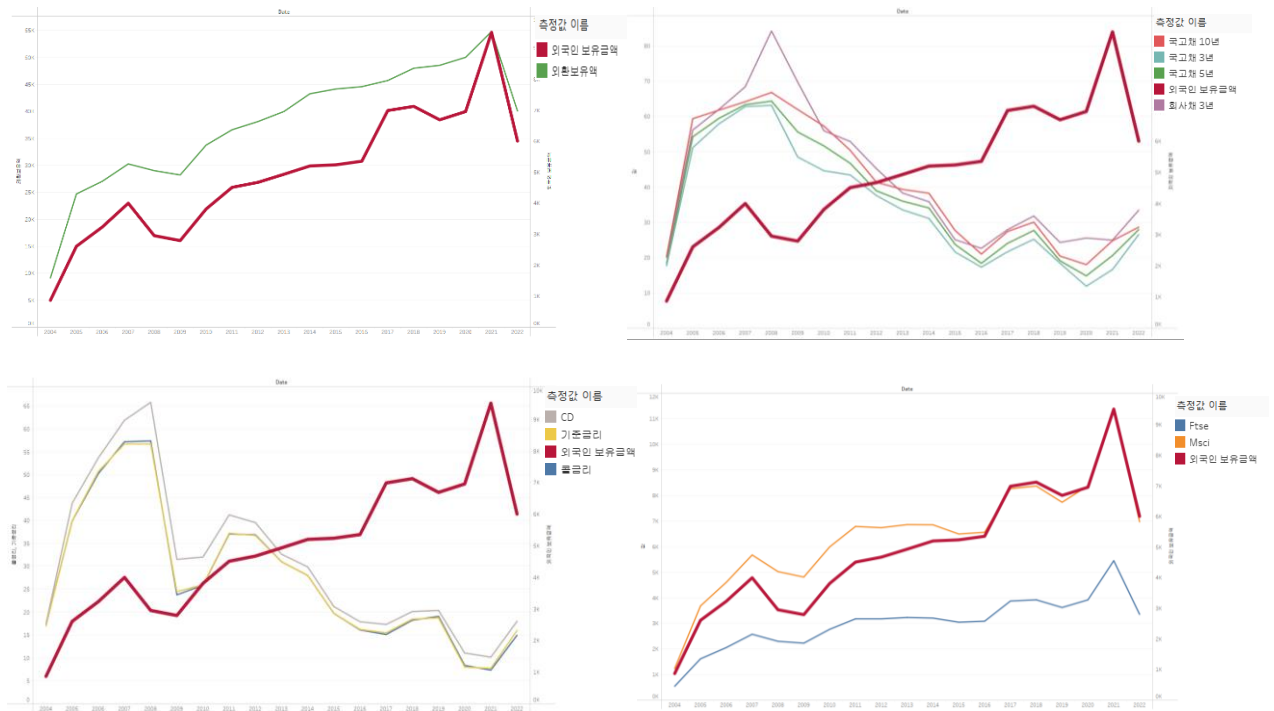


Figure 1 예측 변수인 외국인보유금액(빨간선)과 그 외 변수들의 추세

(왼쪽 위) 예측 변수와 외환보유액, (오른쪽 위) 예측 변수와 국고채, 회사채 관련 변수,

(왼쪽 아래) 예측 변수와 금리관련 변수, (오른쪽 아래) 예측 변수와 투자지수 관련 변수

Figure 1은 예측 대상이 되는 외국인 투자금액과 독립변수들로 넣어준 그 외 변수들의 추세선을 함께 시각화한 자료이다. 대부분의 변수들 모두 일정한 추세를 가지고 있으며 이는 본 팀이 구축한 데이터셋이 non-Stationary 하다는 것을 의미한다.

물론 시각적으로도 정상성이 충족되지 않음을 확인할 수 있었지만, 추가적으로 정상성을 통계적으로 검정하기 위해 ADF Test(Augmented Dickey-Fuller Test)와 KPSS(Kwiatkowski-Phillips-Schmidt-Shin)를 이용해 교차검정을 실시하였다. ADF Test의 경우 p-value가 0.05 미만인 경우에 정상성을 충족한다고 판단하며, KPSS 검정은 p-value가 0.05보다 커야 정상성을 충족한다고 판단한다. 아무런 전처리를 하지 않은 상태의 Raw Data가 가지는 검정 상태를 표로 나타낸 결과는 아래 표와 같다.

| 변수명 | KPSS | ADF | 정상성 여부 | 변수명 | KPSS | ADF | 정상성 여부 |
|-------------|-------|-------|-------------|----------|-------|-------|-------------|
| 외국인_보유금액 | 0.01 | 0.566 | FALSE | 외평채_2028 | 0.01 | 0.243 | FALSE |
| 외환보유액 | 0.01 | 0.569 | FALSE | 외평채_2029 | 0.01 | 0.25 | FALSE |
| FTSE | 0.01 | 0.297 | FALSE | 외평채_2044 | 0.01 | 0.797 | FALSE |
| MSCI | 0.01 | 0.27 | FALSE | 외평채_2048 | 0.01 | 0.841 | FALSE |
| 외평채_2013 | 0.01 | 0.171 | FALSE | 국고채_3년 | 0.01 | 0.477 | FALSE |
| 외평채_2014_04 | 0.045 | 0.051 | FALSE | 국고채_5년 | 0.01 | 0.517 | FALSE |
| 외평채_2014_09 | 0.01 | 0.137 | FALSE | 국고채_10년 | 0.01 | 0.565 | FALSE |
| 외평채_2015 | 0.02 | 0.103 | FALSE | 회사채_3년 | 0.01 | 0.483 | FALSE |
| 외평채_2016 | 0.01 | 0.525 | FALSE | CD | 0.01 | 0.326 | FALSE |
| 외평채_2019 | 0.061 | 0.006 | TRUE | 콜금리 | 0.01 | 0.346 | FALSE |
| 외평채_2023 | 0.01 | 0.279 | FALSE | 기준금리 | 0.01 | 0.349 | FALSE |
| 외평채_2024 | 0.01 | 0.514 | FALSE | 연준금리 | 0.011 | 0.042 | FALSE |
| 외평채_2027 | 0.01 | 0.518 | FALSE | AAII | 0.1 | 0 | TRUE |
| | | | | VIX | 0.1 | 0 | TRUE |

Table 3 Raw Data의 정상성 검정 결과. ADF와 KPSS를 교차적으로 이용하여 검정하였다. 둘 중 하나라도 만족하지 못하는 경우엔 정상성을 충족하지 못하는 것으로 판단하였다.

아무런 전처리를 하지 않은 Raw Data의 경우, 2019년물 외평채와 AAII, VIX 지수를 제외하고는 모두 정상성을 충족시키지 못했다. 본 팀은 정상성을 충족시켜주기 위해 가장 기본적으로 떠올릴 수 있는 방법인 차분(Differencing)과 최근 통계학에서 널리 쓰이고 있는 Box-cox 변환 계열을 고려하였다. 이들에 대한 대처는 추후 나오는 전처리 파트를 참고하길 바란다.

본 팀의 목적은 각 모델이 판단하는 변수의 영향력 즉, 상관관계를 확인하는 것이기 때문에 다중공선성의 여부를 사전에 미리 파악할 필요가 있었다. 데이터셋에 존재하는 다중공선성을 가장 직관적으로 시각화 시킬 수 있는 방법은 Correlation Matrix를

Heatmap을 통해 이용하는 것이다. 아래 그림은 Heatmap을 통해 Correlation Matrix를 시각화한 그림이다.

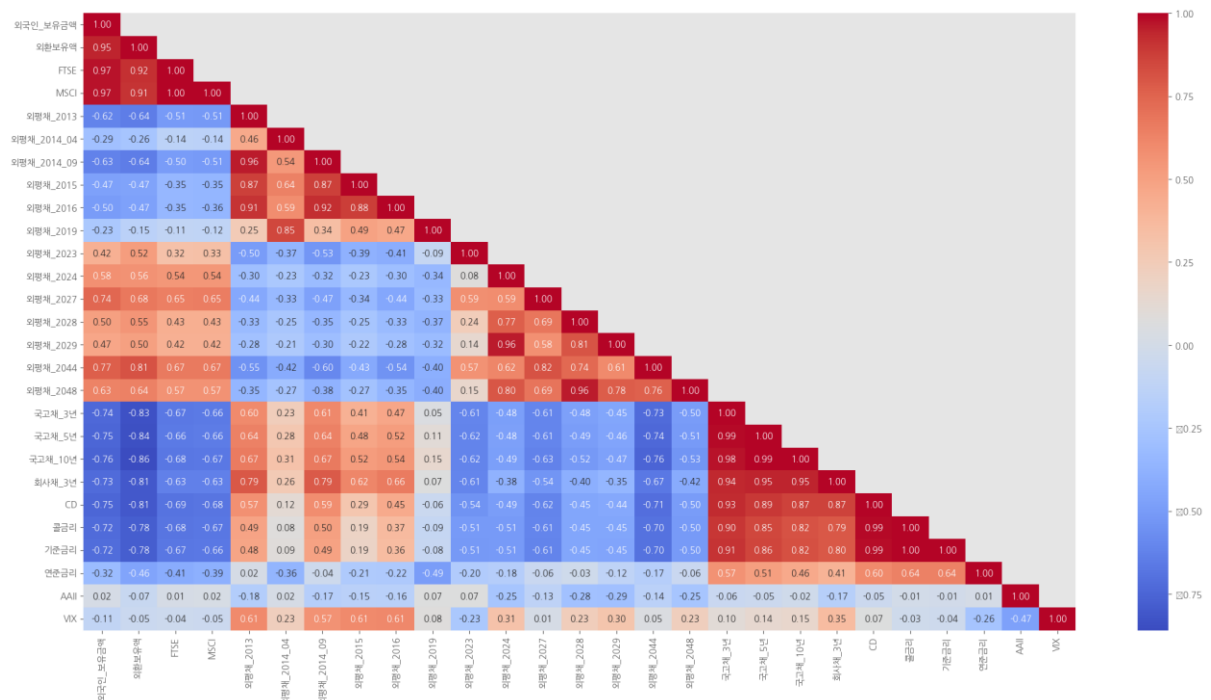


Figure 2 Correlation Matrix를 Heatmap으로 시각화한 그림

변수 간의 상관관계를 파란색부터 빨간색으로 표현하였으며, 양 극단의 색깔로 갈수록 (색이 진해질수록) 두 변수 간의 상관관계가 높다는 것을 의미한다. VIX와 AAI 변수를 제외하고는 대부분의 변수가 다른 변수와 상관관계가 존재한다는 것을 알 수 있었다. 이는 외평채나 국고채와 같이 같은 종류의 변수임에도 만기연도에 따라 분리해둔 것이 원인으로 작용한 것으로 추정된다.

다중공선성 역시 시각화를 통해서도 충분히 확인할 수 있지만, 통계적 검정을 통해 확인할 필요가 있다고 판단하였다. 다중공선성이 높은 변수를 탐지하는 방법인 VIF(Variance Inflation Factor, 분산팽창요인)을 활용하였다. VIF 수치는 1부터 무한대의 값을 가지며, 5를 넘기면 주의, 10을 넘기면 다중공선성이 존재한다고 판단한다. VIF를 이용하여 Raw Data의 변수들의 다중공선성을 측정한 결과는 아래와 같다.

| 변수명 | VIF | 변수명 | VIF |
|----------|----------|-------------|-------------|
| FTSE | 50811.62 | 외평채_2029 | 54.29 |
| MSCI | 43556.42 | 외평채_2014_09 | 46.47 |
| 국고채_5년 | 15621.93 | 외평채_2013 | 40.94 |
| 국고채_3년 | 6252.25 | 외평채_2016 | 31.26 |
| 국고채_10년 | 4787.25 | VIX | 23.33 |
| 콜금리 | 2831.92 | 연준금리 | 21.52 |
| 외국인 보유금액 | 2473.36 | 외평채_2015 | 18.08 |
| 기준금리 | 1895.49 | 외평채_2044 | 17.33 |
| CD | 1260.44 | 외평채_2019 | 15.33 |
| 외환보유액 | 547.05 | 외평채_2027 | 15.12 |
| 회사채_3년 | 447.78 | 외평채_2014_04 | 15.12 |
| 외평채_2048 | 92.83 | 외평채_2023 | 10.81 |
| 외평채_2028 | 91.60 | AAII | 2.09 |
| 외평채_2024 | 62.70 | | |

Table 4 5 Raw Data의 VIF 측정 결과

측정 결과, AAII 변수를 제외하고는 모두 VIF 수치가 10 이상으로 다중공선성이 존재하는 것으로 나타났다. 본래 다중공선성을 극복하는 방법에는 VIF 수치가 높은 변수부터 차례대로 제거하는 방법이 있으나, 본 팀은 과도하게 많은 다중공선성을 극복함과 동시에 변수들이 가지고 있는 정보량을 최대한 보존하기 위해 PCA(Principle Component Analysis, 주성분분석)를 활용하기로 결정하였다.

EDA 과정을 통해 알게된 사실을 정리하면, 구축된 데이터셋의 변수들은 대부분 non-stationary한 상태를 띄고 있으며 변수간 다중공선성 또한 매우 높게 나타났다. 본 팀은 이를 극복하기 위해 차분, Box-cox 변환, PCA를 이용하기로 결정하였다.

5. 전처리

EDA 파트에서 확인했던 바와 같이, 본 팀이 구축한 데이터셋은 사용하고자 한 모델인 VAR과 XGBoost에 적합하지 않았다. 동시에 변수의 중요도 파악에 가장 중요한 다중공선성 제거가 필수적이기 때문에 이를 종합적으로 고려하여 PCA와 차분, box-cox 변환을 시도했다.

가장 먼저 차원축소 기법 중에 하나인 PCA를 적용하였다. PCA는 데이터들의 분산을 가장 잘 설명하는 축(Axis)을 찾아내는 방법으로 쉽게 비유하면 피자맛을 느끼기 위해 피자를 먹는 대신 피자빵을 먹는 것과 비슷하다고 할 수 있다. PCA를 사용하는 목적은 변수의 데이터 수에 비해 너무 많아 차원의 저주³에 빠졌을 때 이를 극복하기 위함이나 다중공선성을 극복하기 위해 사용된다. 그러나 원래의 변수를 그대로 사용하는 것이 아니기 때문에 원변수가 가지고 있는 정보의 일부손실은 감안해야한다. 해당 기법을 사용하는 목적은 다중공선성을 제거하기 위함이지만 동시에 정보손실을 최소화시키기 위해 EDA에서 실시했던 Correlation Matrix의 시각화 결과를 바탕으로 압축 군을 형성하였다. 그렇게 형성한 차원축소 대상은 아래와 같다.

| 압축군 | 변수 |
|----------|---------------------------------|
| 투자지수 관련 | 외환보유액, FTSE, MSCI |
| 과거외평채 관련 | 2022년 기준 과거시점의 외평채 스프레드 |
| 미래외평채 관련 | 2022년 기준 미래시점의 외평채 스프레드 |
| 국고채 관련 | 국고채 3년, 국고채 5년, 국고채 10년, 회사채 3년 |
| 금리관련 | CD, 콜금리, 기준금리, 미연준금리 |

Table 6 PCA 압축군과 재료 변수 목록

변수를 압축할 때의 축의 수는 사용자가 지정해야 하는 parameter로, 원 변수들을 몇 개의 축으로 압축을 시킬 것인지를 정할 수 있다. 본 팀은 변수들의 분산을 최소 90% 설명할 수 있는 축의 수를 지정해주었으며 압축된 축의 설명력은 투자지수 관련 변수는 축 1개로 95.8% , 과거 외평채 관련 변수는 축 2개로 93.8%, 미래 외평채 관련 변수는 축 3개로 93.5%, 국고채 관련 변수는 축 1개로 97.4%, 금리 관련 변수 또한 축 1개로 99.3% 이다. 차원축소에 활용하지 않은 연준금리, AAIL, VIX 변수는 다른 변수와 비교했을

³ 차원의 저주 : 차원(Feature)이 증가하면서 개별 차원 내의 학습할 데이터의 수가 적어지게 되어 발생하는 현상. 데이터들 간의 거리가 희소(Sparse)해진다는 특징 때문에 거리 기반 모델이나 데이터의 수가 중요한 모델의 학습 성능이 저하된다는 특징이 있다.

때 타변수와의 비교적 낮은 상관관계가 보여 차원축소 대상에서 제외시켰다.

차원축소를 진행한 후에 압축된 변수들에 대해 위에서 진행했던 KPSS-ADF 교차검정을 재 실시 하였다.

| 변수명 | KPSS | ADF | 정상성 여부 | 변수명 | KPSS | ADF | 정상성 여부 |
|-------------|-------|-------|-------------|-------------|-------|-------|-------------|
| 투자지수 | 0.01 | 0.728 | FALSE | 미래 외평채 1 | 0.01 | 0.566 | FALSE |
| 국고채관련 | 0.01 | 0.735 | FALSE | 미래 외평채 2 | 0.039 | 0.446 | FALSE |
| 금리관련 | 0.01 | 0.469 | FALSE | 미래 외평채 3 | 0.012 | 0.000 | TRUE |
| 과거 외평채 1 | 0.01 | 0.293 | FALSE | AAII | 0.1 | 0.000 | TRUE |
| 과거 외평채 2 | 0.061 | 0.036 | TRUE | VIX | 0.1 | 0.000 | TRUE |
| 연준금리 | 0.01 | 0.038 | FALSE | | | | |

Table 7 차원축소 이후의 KPSS-ADF 검정 결과

여전히 대부분의 변수에 대해 아직 non-Stationary한 상태에 있음을 알 수 있다. Non-Stationary한 변수들을 Stationary한 상태로 만들어주기 위해 두 가지 방법을 고안하였다. 하나는 가장 기본적으로 사용되는 차분(Differencing)이고, 다른 하나는 거듭곱 변환(Power Transformation)의 일종인 Box-cox 변환법이다.

거듭곱 변환의 일종인 Box-cox 변환법은 모든 값이 양수여야 한다는 제약조건이 있어, 음수 값이 포함되어 있는 PCA로 압축 데이터셋엔 적합하지 않았다. 때문에 Box-cox의 일반화된 기법인 Yeo-Johnson을 이용하였으나 단독으로 사용했을 때와 차분과 함께 적용했을 때 모두 모든 변수에 대해 정상성이 확보되지 않았다.

차분은 lag=1 부터 lag=12까지 적용시켰고, 그 중 가장 정보 손실이 적으면서 모든 변수에 대해 정상성을 충족한 lag=1을 적용하기로 결정하였다.

최종적으로 전처리된 데이터에 대한 KPSS-ADF 검정 결과와 VIF 계수, 모든 변수에 대해 그린 plot은 아래와 같다.

| 변수명 | KPSS | ADF | 정상성 여부 | 변수명 | KPSS | ADF | 정상성 여부 |
|-------------|-------|-------|--------|-------------|------|-------|--------|
| 투자지수 | 0.1 | 0.000 | TRUE | 미래 외평채 1 | 0.1 | 0.000 | TRUE |
| 국고채관련 | 0.1 | 0.004 | TRUE | 미래 외평채 2 | 0.1 | 0.000 | TRUE |
| 금리관련 | 0.1 | 0.000 | TRUE | 미래 외평채 3 | 0.1 | 0.000 | TRUE |
| 과거 외평채 1 | 0.1 | 0.000 | TRUE | AAll | 0.1 | 0.000 | TRUE |
| 과거 외평채 2 | 0.063 | 0.034 | TRUE | VIX | 0.1 | 0.000 | TRUE |
| 연준금리 | 0.1 | 0.005 | TRUE | | | | |

Table 8 PCA와 차분 이후 KPSS-ADF 검정 결과

| 변수명 | VIF | 변수명 | VIF |
|-------------|-------|-------------|-------|
| VIX | 1.84 | 과거 외평채 1 | 1.257 |
| 금리관련 | 1.67 | 국고채 관련 | 1.21 |
| 연준금리 | 1.63 | 미래 외평채 1 | 1.20 |
| 과거 외평채 2 | 1.38 | 미래 외평채 2 | 1.15 |
| AAll | 1.33 | 미래 외평채 3 | 1.09 |
| 투자지수 | 1.258 | | |

Table 9 PCA 이후 VIF 측정 결과

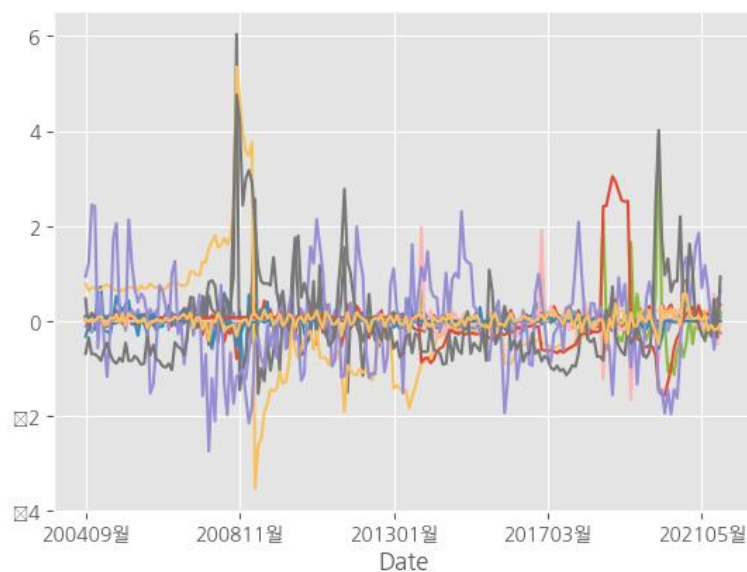


Figure 3 PCA와 차분 이후의 모든 변수의 Plot

4. 모델링

본 팀의 목적이 모델이 내놓은 변수 유의성 차이 파악인 만큼, 동일한 데이터로 학습하고 예측할 필요가 있었다. 데이터에 대한 전처리는 VAR과 XGBoost 모두 위의 전처리 파트의 과정을 따랐으며, 학습 데이터는 두 모델 모두 동일하게 2004년 9월부터 2021년 11월까지의 데이터로, 테스트 데이터는 2021년 12월부터 2022년 9월까지의 데이터를 사용했다.

먼저 VAR 모델에 대한 모델링부터 실시했다. Python의 VAR 모델은 파라미터로서 지연 차수인 p 의 값을 지정해줄 수 있다. 이때 p 는 몇기 이전의 값을 참고할 것인지를 정하는 값이며, 대개 모델의 적합도를 판단하는 AIC의 값이 제일 낮은 값⁴을 선택한다. 파이썬의 $p=1$ 인 경우부터 순차적으로 확인해본 결과, $p=14$ 일 때 AIC 점수와 모델의 실질적인 예측에 있어 가장 좋은 성능을 보였다. 추가적으로 $p=15$ 인 지점부터 다시 AIC 점수가 반등하는 것이 관찰되어, $p=14$ 로 학습된 모델을 계수추정과 변수 예측에 최종적으로 활용하였다. 학습 결과는 본 보고서의 부록을 참고하길 바란다.

단순히 모델이 OLS로 추정했을 때 나온 변수의 중요도뿐만 아니라, 충격반응함수를 이용하여 특정 변수가 target 변수에 대해 어떤 방향의 영향을 미치는지를 검정했다. 모델의 학습 결과와 충격반응함수의 결과를 함께 고려했을 때, VAR 모델이 판단한 유의미한 변수를 판단하였다. Impuls의 영향력은 $|0.5|$ 를 기준으로 하여 해당 수치보다 영향력이 없는 변수는 제외하였다.

추가적으로, 변수의 상관관계뿐만 아니라 선후관계 또한 본 팀의 과제에 있어 중요한 문제이기 때문에 선후관계에 해단 검증은 Granger Causality를 이용한 Matrix를 python으로 구현하여 확인하였다. Granger Causality 선후관계까지 고려한 종속변수에 유의미했던 변수는 아래와 같다. (상위 3개만 표시)

| 변수 | VAR+IMPULSE | GRANGER |
|--------|-------------|----------|
| 미래 외평채 | 0.04** | 0.001*** |
| 금리관련 | 0.045** | 0.062* |
| AAII | 0.068** | 0.06* |

Table 10 VAR 모델로 판단한 종합적 변수 중요도 선정 결과

⁴ AIC의 값이 낮을수록 주어진 데이터셋에 적합하다고 판단한다.

본 데이터셋으로 학습된 VAR모델로 테스트셋의 기간동안을 예측하게 하였다. 예측에 대한 시각화 자료는 아래와 같다. 본 모델의 예측력 척도는 Mean Squared Error를 이용하였으며, VAR 모델의 테스트 데이터셋에 대한 MSE는 0.161 이었다.

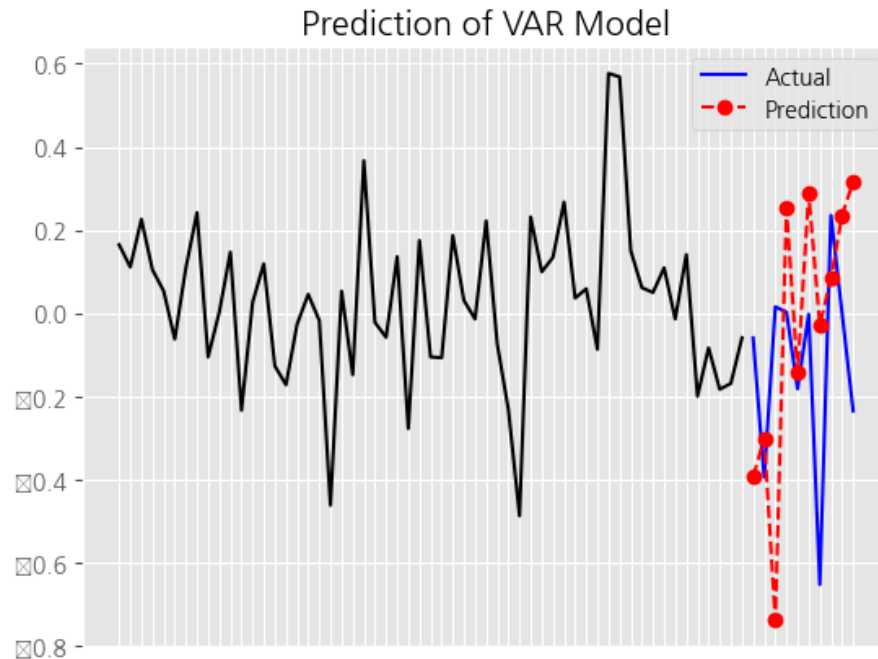


Figure 4 VAR 모델의 변수에 대한 예측 시각화, 검정색으로 표시한 plot은 train data셋으로, 예측에 반영하지 않았다.

다음은 XGBoost를 이용한 모델링이다. XGBoost는 본래 Tree Model의 일종으로 CART 알고리즘 기반이기에 Classification Task에 많이 사용되어왔지만, 최근엔 이를 Regression 영역에서도 활용하여 Kaggle과 같은 여러 머신러닝 경진대회에서 정형 데이터 task 에서 좋은 성적을 내고 있다. XGBoost를 비롯한 CART(Classification And Regression Tree) 알고리즘 기반의 트리모델들은 Regression 문제를 풀 때, node를 쪼개는 지표 index를 Entropy나 Gini 계수가 아닌 실제값과 예측값의 오차를 의미하는 Error를 기준으로 사용한다. 이런 기준으로 Data Point들을 쪼개다 보면 쪼개 영역을 기준으로 생성된 leaf node들에 Data Point들이 들어가게 되는데, 이 Data Point의 평균값을 최종 예측값으로 이용한다. 요약하면, 분기를 쪼갤 때 Error를 이용하며, leaf node에 속한 Data Point들의 평균값을 최종 예측값으로 반환한다고 할 수 있다.

XGBoost를 이용한 모델링은 VAR에 비해 조정해줄 파라미터의 수가 적다. 사실상 조절해줄 수 있는 parameter는 VAR에 비해 많으나, default로 설정되어 있는 parameter 만으

로도 충분히 좋은 예측력을 얻어낼 수 있어 hyper parameter 조정은 하지 않았다. 실제로 hyper parameter는 모델의 성능을 최적으로 내야할 때 사용되며, parameter 조정으로 얻을 수 있는 모델 성능 향상은 데이터 가공에 비해 크지 않은 편이다. 따라서 일반적으로 머신러닝 모델의 성능 향상을 기대할 땐 데이터를 적절한 방향으로 재가공하지만 본 팀의 목적은 변수의 중요도를 파악하는 것이 우선이었으며, 위에서 언급한 바와 같이 모델 자체에 default로 설정되어 있는 parameter 만으로도 충분히 좋은 예측력과 학습결과를 얻을 수 있었다. XGBoost 모델의 예측에 대한 시각화 자료는 아래와 같다. 본 모델의 MSE는 0.020 이었다. (VAR's MSE : 0.161)

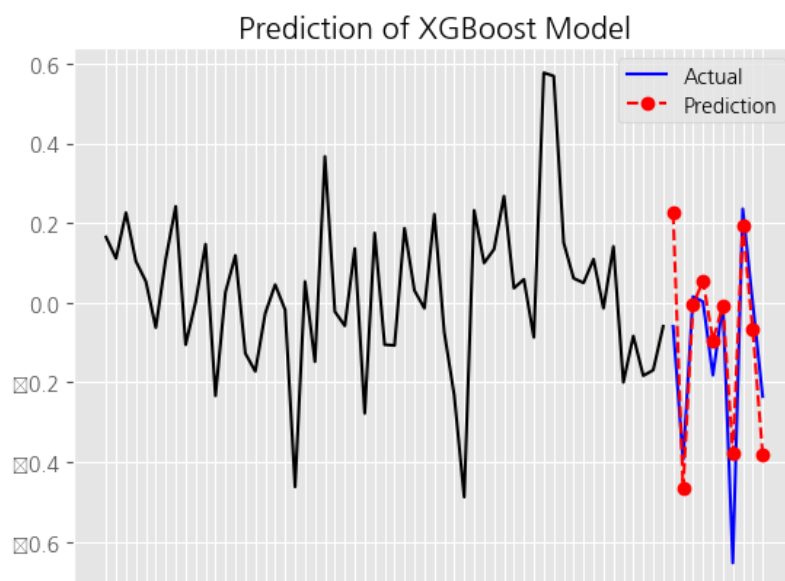


Figure 5 XGBoost 모델의 예측 시각화

XGBoost 와 같은 CART 알고리즘 기반 모델의 장점은 트리의 분기를 쪼갤 때, 어떤 변수를 중요하게 보며 쪼갬는지를 통해 변수의 중요도를 파악할 수 있다는 것이다. 이를 Feature Importance 기법이라고 하며, 다만 주의할 것은 이는 말 그대로 모델이 분기를 쪼갤 때 중요하게 사용했던 변수를 말하는 것이지, 그 변수가 통계적으로 중요하다든가, 모델의 성능을 결정짓는 중요 변수라는 뜻은 아니라는 것이다. 그러나 모델이 학습할 때 중요하다고 판단되었던 변수라는 점에서 본 팀의 변수 중요도 파악 목적엔 적합하다고 판단하였다. 그러나 위에서 언급한 한계점을 극복하기 위해, 좀 더 엄밀한 의미에서의 변수 중요도를 파악할 수 있는 Permutation Importance 기법을 적용하였다. 해당 기법은 변수 내부의 값들을 임의로 섞어(shuffle) 모델의 예측력이 하락하는 정도로 변수의 중요도를 파악하는 기법이다. 두 기법을 이용해 파악한 변수 중요도는 아래와 같다.

| 순위 | FEATURE IMPORTANCE | PERMUTATION IMPORTANCE |
|----|--------------------|------------------------|
| 1 | 투자지수 | 투자지수 |
| 2 | 국고채 관련 | AAII 심리지수 |
| 3 | 금리관련 | 연준금리 |
| 4 | 과거외평채_1 | 과거_외평채_1 |

Table 11 머신러닝 모델의 변수 중요도 검사 결과

5. 결론 및 시사점

실험 결과, 기존의 전통 통계학 모델은 VAR의 학습 결과와 충격반응함수, Granger Causality를 모두 종합적으로 고려하여 **미래 외평채, 금리관련, AAI**와 같은 변수가 유의미하다고 판단했으며 머신러닝 모델은 Feature Importance와 Permutation Importance를 고려했을 때 **투자지수와 과거외평채**를 중요한 변수로 판단했다고 결론 내릴 수 있다.

모델의 예측력을 잠시 떠나, 본 팀은 위와 같은 결과로 두 모델이 같은 데이터셋을 학습했음에도 불구하고 변수 중요도 선정에 있어 다른 결과를 내놓은 것에 대해 생각해 보았다. VAR 모델은 예측하는 변수와 더불어 다른 변수들의 “과거”를 함께 고려한다. 본인 뿐만 아니라 본인에게 영향을 줄 수 있는 다양한 변수들의 과거까지 고려하여 회귀계수를 추정하고 학습된 정보를 바탕으로 미래를 예측한다. 다만 VAR 모델 결과식을 단독으로 사용하기엔 변수들 간의 관계를 직관적으로 이해하기 어렵기 때문에 충격반응함수나 Granger Causality, 본 과정에는 포함되지 않았지만 Variance Decomposition과 같은 통계적 기법들을 추가적으로 이용하여 최종 결론에 이른다. 즉, 결론에 이르는 과정이 머신러닝에 비해 상대적으로 엄밀하게 증명된 과정을 거치게 된다.

그러나 머신러닝은 그 내부적인 추정 절차만 봤을 때, 모델이 받은 데이터의 패턴을 수많은 연산을 통해 파악하는 것일 뿐 그 과정에서 엔지니어링적인 연산 외에는 그 어떤 추가적인 통계적 검증을 요구하지 않는다. Feature Importance나 Permutation Importance와 같은 변수 중요도 추정 방법 또한 예측을 최대한 잘 하려는 과정에서 알게된 사실을 말하는 것이지, 정말 엄밀하게는 그 변수가 통계적으로 유의하다는 것을 의미하지는 않는다. 그러나 머신러닝은 그 목적에 부합하게, 통계적 기법에 비해 상대적으로 매우 높은 예측력을 자랑한다. 데이터의 패턴을 수많은 연산을 통해 파악하고 이렇게 파악된 패턴을 통해 주어진 Feature에 대한 Prediction이 주가 되는 분야에선 매우 유용하게 활용될 수 있음은 경험적으로, 이론적으로도 많은 연구와 증명이 이루어져왔다. 그러나 위에서

말한 바와 같이, 변수 간의 상관관계를 추정하거나 영향력을 엄밀하게 추정하는 일에 있어서는 머신러닝 엔지니어들 사이에서도 회의적인 입장이 대부분이다.

이것이 머신러닝이 경제 외의 도메인에서는 뛰어난 예측력을 바탕으로 많이 활용되고 있지만 경제학계에서의 머신러닝 모델의 활용은 아직까지 어려운 이유라고 생각한다. 데이터셋이 충분하며, 주어진 Feature에 대한 예측이 주된 업무라면, 머신러닝 혹은 딥러닝을 적극적으로 고려할만 하다. 그러나, 경제학은 엄밀하게 세워진 논리와 가정들로 우리 사회에서 상호작용을 주고받는 경제 변수간의 관계를 파악하는 학문이기도 하다. 경제통계학 기반의 변수추정은 논리와 타당성이 증명된 기법으로 변수의 영향력을 파악함으로써 특정 정책에 반영할 만큼 설득력을 확보할 수 있지만, 머신러닝의 “예측의 부산물”로 나온 변수 설명은 예측에 대한 설명은 가능하지만, 그것이 곧 변수 간 상관관계를 설명한다고 할 수는 없기에 아직까지 해당 분야의 설득력은 미흡한 상태라고 할 수 있다.

두 분야 모두 각자의 강점을 지니고 있다. 각 분야의 강점만을 활용하고자 한다면 한계에 금방 부딪히겠지만, 두 분야의 강점을 적절히 앙상블(Ensemble)시킬 수 있는 방안을 찾을 수만 있다면 경제학 발전에 큰 기여를 할 수 있을 것으로 기대한다. 예를 들어, 통계학의 변수 추정과 머신러닝의 뛰어난 예측력을 더한다면, 데이터셋 구축에 들어가는 비용을 줄일 수 있음과 동시에 예측력까지 확보하게 되어 정책 의사결정에 드는 절차적 비용을 크게 절감시킬 수 있을 것이다.

한 분야만을 고집하는 것이 아닌, 다양하게 활용될 수 있는 기술을 적절히 도입하여 도메인에 있어 새로운 방향을 찾을 수 있는 더 다양한 기회와 세상을 기대한다.

참고자료

- [1] 최아름, 구지현 (2017). 한국의 외국인직접투자에 영향을 주는 요인에 관한 연구 - 문화적 차원과 부패인식지수를 중심으로-. *Journal of Digital Convergence*
- [2] 금융보안원 (2017). 국내외 금융권 머신러닝 도입 현황.
- [3] 정재현 (2019). 머신러닝을 활용한 정책설계 : 출산 결정요인을 중심으로.
- [4] 김수현 (2020). Macroeconomic and Financial Market Analyses and Predictions through Deep Learning. *BOK경제연구*
- [5] 조병현 (2020). 미국인들의 투자 심리와 변화의 가능성. *Yuanta Research*

Appendix

Results for equation Target

| Results for equation Target | | | | | L6.Target | | | | |
|-----------------------------|-------------|------------|--------|-------|--------------|-----------|----------|--------|-------|
| | coefficient | std. error | t-stat | prob | | -0.928861 | 0.854411 | -1.087 | 0.277 |
| const | -0.049506 | 0.035568 | -1.392 | 0.164 | L7.투자지수 | 0.055287 | 0.529657 | 0.104 | 0.917 |
| L1.투자지수 | -0.594086 | 0.416546 | -1.426 | 0.154 | L7.국고채관련 | -0.156175 | 0.147148 | -1.061 | 0.289 |
| L1.국고채관련 | -0.180917 | 0.122058 | -1.482 | 0.138 | L7.금리관련 | -0.113183 | 0.208555 | -0.543 | 0.587 |
| L1.금리관련 | 0.028761 | 0.179977 | 0.160 | 0.873 | L7.과거_외평채_1 | 0.076558 | 0.056266 | 1.361 | 0.174 |
| L1.과거_외평채_1 | 0.001714 | 0.049604 | 0.035 | 0.972 | L7.과거_외평채_2 | -0.011709 | 0.078774 | -0.149 | 0.882 |
| L1.과거_외평채_2 | -0.023068 | 0.061763 | -0.373 | 0.709 | L7.미래_외평채_1 | -0.057228 | 0.057684 | -0.992 | 0.321 |
| L1.미래_외평채_1 | -0.008526 | 0.074285 | -0.115 | 0.909 | L7.미래_외평채_2 | -0.028363 | 0.048572 | -0.584 | 0.559 |
| L1.미래_외평채_2 | -0.050937 | 0.071072 | -0.717 | 0.474 | L7.미래_외평채_3 | -0.037060 | 0.083013 | -0.446 | 0.655 |
| L1.미래_외평채_3 | -0.086754 | 0.080081 | -1.083 | 0.279 | L7.연준금리 | -0.023221 | 0.446496 | -0.052 | 0.959 |
| L1.연준금리 | -0.118204 | 0.382807 | -0.309 | 0.757 | L7.AAII | -0.028401 | 0.035733 | -0.795 | 0.427 |
| L1.AAII | -0.020275 | 0.030443 | -0.666 | 0.505 | L7.VIX | -0.031172 | 0.052443 | -0.594 | 0.552 |
| L1.VIX | 0.005256 | 0.053702 | 0.098 | 0.922 | L7.Target | 0.091015 | 0.809233 | 0.112 | 0.910 |
| L1.Target | 0.930425 | 0.724659 | 1.284 | 0.199 | L8.투자지수 | 0.054617 | 0.513421 | 0.106 | 0.915 |
| L2.투자지수 | -0.343668 | 0.486508 | -0.706 | 0.480 | L8.국고채관련 | -0.129373 | 0.151980 | -0.851 | 0.395 |
| L2.국고채관련 | 0.160816 | 0.121668 | 1.322 | 0.186 | L8.금리관련 | 0.320889 | 0.249255 | 1.287 | 0.198 |
| L2.금리관련 | -0.336088 | 0.198484 | -1.693 | 0.090 | L8.과거_외평채_1 | -0.040095 | 0.049104 | -0.817 | 0.414 |
| L2.과거_외평채_1 | -0.002468 | 0.052966 | -0.047 | 0.963 | L8.과거_외평채_2 | -0.063081 | 0.078157 | -0.807 | 0.420 |
| L2.과거_외평채_2 | -0.006166 | 0.086047 | -1.001 | 0.317 | L8.미래_외평채_1 | 0.119315 | 0.067160 | 1.777 | 0.076 |
| L2.과거_외평채_1 | 0.019816 | 0.068736 | 0.288 | 0.773 | L8.미래_외평채_2 | 0.010114 | 0.048256 | 0.210 | 0.834 |
| L2.미래_외평채_2 | 0.006698 | 0.063917 | 1.356 | 0.175 | L8.미래_외평채_3 | 0.019300 | 0.079493 | 0.243 | 0.808 |
| L2.미래_외평채_3 | 0.259431 | 0.123110 | 2.107 | 0.035 | L8.연준금리 | -0.129244 | 0.394980 | -0.327 | 0.744 |
| L2.연준금리 | -0.556080 | 0.349303 | -1.592 | 0.111 | L8.AAII | 0.062624 | 0.034342 | 1.824 | 0.068 |
| L2.AAII | 0.015214 | 0.027242 | 0.558 | 0.577 | L8.VIX | 0.017282 | 0.045461 | 0.380 | 0.704 |
| L2.VIX | 0.047214 | 0.059785 | 0.790 | 0.430 | L8.Target | -0.030884 | 0.799062 | -0.039 | 0.969 |
| L2.Target | 0.717871 | 0.764734 | 0.939 | 0.348 | L9.투자지수 | 0.590612 | 0.442853 | 1.334 | 0.182 |
| L3.투자지수 | 0.813952 | 0.551927 | 1.475 | 0.140 | L9.국고채관련 | 0.028818 | 0.114366 | 0.252 | 0.801 |
| L3.국고채관련 | 0.068305 | 0.117478 | 0.581 | 0.561 | L9.금리관련 | -0.191481 | 0.236224 | -0.811 | 0.418 |
| L3.금리관련 | -0.259913 | 0.181401 | -1.433 | 0.152 | L9.과거_외평채_1 | 0.018326 | 0.044934 | 0.408 | 0.683 |
| L3.과거_외평채_1 | -0.022665 | 0.049949 | -0.454 | 0.650 | L9.과거_외평채_2 | 0.118633 | 0.055706 | 2.130 | 0.033 |
| L3.과거_외평채_2 | 0.116468 | 0.087722 | 1.328 | 0.184 | L9.미래_외평채_1 | 0.090513 | 0.066854 | 1.354 | 0.176 |
| L3.과거_외평채_1 | 0.018520 | 0.064300 | 0.288 | 0.773 | L9.미래_외평채_2 | 0.134709 | 0.050201 | 2.683 | 0.007 |
| L3.미래_외평채_2 | 0.022488 | 0.064009 | 0.351 | 0.725 | L9.미래_외평채_3 | 0.093551 | 0.072974 | 1.282 | 0.200 |
| L3.미래_외평채_3 | -0.227755 | 0.122548 | -1.858 | 0.063 | L9.연준금리 | -0.134474 | 0.501503 | -0.268 | 0.789 |
| L3.연준금리 | 0.757998 | 0.390318 | 1.942 | 0.052 | L9.AAII | -0.019992 | 0.036980 | -0.541 | 0.589 |
| L3.AAII | 0.008127 | 0.028561 | 0.285 | 0.776 | L10.과거_외평채_1 | 0.025999 | 0.044497 | 0.584 | 0.559 |
| L3.VIX | 0.024155 | 0.059594 | 0.405 | 0.685 | L10.과거_외평채_2 | -0.084857 | 0.057480 | -1.476 | 0.140 |
| L3.Target | -0.968561 | 0.817249 | -1.185 | 0.236 | L10.미래_외평채_1 | -0.079948 | 0.081461 | -0.981 | 0.326 |
| L4.투자지수 | 0.049115 | 0.548797 | 0.089 | 0.929 | L10.미래_외평채_2 | -0.022090 | 0.070792 | -0.312 | 0.755 |
| L4.국고채관련 | -0.088489 | 0.113584 | -0.779 | 0.436 | L10.미래_외평채_3 | -0.197165 | 0.095903 | -2.056 | 0.040 |
| L4.금리관련 | -0.070362 | 0.193244 | -0.364 | 0.716 | L10.연준금리 | 0.224921 | 0.462074 | 0.487 | 0.626 |
| L4.과거_외평채_1 | 0.021394 | 0.054056 | 0.396 | 0.692 | L10.AAII | -0.000692 | 0.031819 | -0.022 | 0.983 |
| L4.과거_외평채_2 | -0.004237 | 0.071941 | -0.059 | 0.953 | L10.VIX | -0.007754 | 0.048145 | -0.161 | 0.872 |
| L4.미래_외평채_1 | 0.121310 | 0.064069 | 1.893 | 0.058 | L10.Target | -1.411316 | 0.738789 | -1.910 | 0.056 |
| L4.미래_외평채_2 | 0.081396 | 0.060516 | 1.345 | 0.179 | L11.투자지수 | 0.107260 | 0.598247 | 0.179 | 0.858 |
| L4.미래_외평채_3 | 0.051742 | 0.129954 | 0.398 | 0.691 | L11.국고채관련 | 0.044548 | 0.107844 | 0.413 | 0.680 |
| L4.연준금리 | 0.540262 | 0.477272 | 1.132 | 0.258 | L11.금리관련 | -0.001323 | 0.202100 | -0.007 | 0.995 |
| L4.AAII | -0.022697 | 0.027089 | -0.838 | 0.402 | L11.과거_외평채_1 | 0.033226 | 0.049456 | 0.672 | 0.502 |
| L4.VIX | 0.049457 | 0.058647 | 0.843 | 0.399 | L11.과거_외평채_2 | -0.021703 | 0.055209 | -0.393 | 0.694 |
| L4.Target | 0.071232 | 0.813115 | 0.088 | 0.930 | L11.미래_외평채_1 | -0.052424 | 0.081483 | -0.643 | 0.520 |
| L5.투자지수 | 0.155323 | 0.527956 | 0.294 | 0.769 | L11.미래_외평채_2 | 0.028656 | 0.058930 | 0.486 | 0.627 |
| L5.국고채관련 | 0.016639 | 0.110363 | 0.151 | 0.880 | L11.미래_외평채_3 | 0.315373 | 0.115965 | 2.720 | 0.007 |
| L5.금리관련 | -0.198469 | 0.169628 | -1.170 | 0.242 | L11.연준금리 | -0.012205 | 0.478582 | -0.026 | 0.980 |
| L5.과거_외평채_1 | 0.056496 | 0.050620 | 1.116 | 0.264 | L11.AAII | -0.014403 | 0.023609 | -0.610 | 0.542 |
| L5.과거_외평채_2 | -0.014130 | 0.074710 | -0.180 | 0.850 | L11.VIX | 0.062390 | 0.048869 | 1.277 | 0.202 |
| L5.미래_외평채_1 | -0.021571 | 0.071551 | -0.301 | 0.763 | L11.Target | 0.284346 | 0.895790 | 0.317 | 0.751 |
| L5.미래_외평채_2 | 0.056297 | 0.066292 | 0.849 | 0.396 | L12.투자지수 | 0.983880 | 0.553010 | 1.779 | 0.075 |
| L5.미래_외평채_3 | 0.017330 | 0.122605 | 0.141 | 0.888 | L12.국고채관련 | -0.005620 | 0.120159 | -0.047 | 0.963 |
| L5.연준금리 | 0.180339 | 0.404890 | 0.445 | 0.656 | L12.금리관련 | -0.386744 | 0.192557 | -2.008 | 0.045 |
| L5.AAII | 0.006435 | 0.027471 | 0.234 | 0.815 | L12.과거_외평채_1 | 0.020648 | 0.049788 | 0.415 | 0.678 |
| L5.VIX | -0.038808 | 0.045767 | -0.848 | 0.396 | L12.과거_외평채_2 | -0.004692 | 0.050796 | -0.092 | 0.926 |
| L5.Target | -0.035842 | 0.812519 | -0.044 | 0.965 | L12.미래_외평채_1 | 0.048650 | 0.086681 | 0.561 | 0.575 |
| L6.투자지수 | 0.657292 | 0.554799 | 1.185 | 0.236 | L12.미래_외평채_2 | 0.019591 | 0.066296 | 0.296 | 0.768 |
| L6.국고채관련 | 0.049101 | 0.124103 | 0.396 | 0.692 | L12.미래_외평채_3 | -0.138901 | 0.137010 | -1.014 | 0.311 |
| L6.금리관련 | 0.059969 | 0.101200 | 0.314 | 0.754 | L12.연준금리 | 1.201316 | 0.512036 | 2.346 | 0.019 |
| L6.과거_외평채_1 | 0.071839 | 0.053825 | 1.335 | 0.182 | L12.AAII | -0.012219 | 0.023914 | -0.511 | 0.609 |
| L6.과거_외평채_2 | -0.010802 | 0.070342 | -0.154 | 0.878 | L12.VIX | -0.052434 | 0.052200 | -1.004 | 0.315 |
| L6.미래_외평채_1 | -0.032439 | 0.061420 | -0.528 | 0.597 | L12.Target | -1.508811 | 0.836880 | -1.803 | 0.071 |
| L6.미래_외평채_2 | 0.031111 | 0.052913 | 0.588 | 0.557 | L13.투자지수 | 0.331825 | 0.552586 | 0.600 | 0.548 |
| | | | | | L13.국고채관련 | -0.065919 | 0.106523 | -0.619 | 0.536 |

Figure 6 Python Statsmodels의 VAR 모델로 추정된 학습 결과 (p=14)

| | 투자지수_x | 국고채관련_x | 금리관련_x | 과거_외형채_1_x | 과거_외형채_2_x | 미래_외형채_1_x | 미래_외형채_2_x | 미래_외형채_3_x | 연준금리_x | AAII_x | VIX_x | Target_x |
|------------|--------|---------|--------|------------|------------|------------|------------|------------|--------|--------|--------|----------|
| 투자지수_y | 1.0000 | 0.2958 | 0.0123 | 0.3532 | 0.2158 | 0.0044 | 0.1031 | 0.0195 | 0.2568 | 0.0992 | 0.0012 | 0.0125 |
| 국고채관련_y | 0.0000 | 1.0000 | 0.0018 | 0.0000 | 0.0000 | 0.5245 | 0.2609 | 0.4876 | 0.0022 | 0.0362 | 0.0103 | 0.0014 |
| 금리관련_y | 0.0003 | 0.0024 | 1.0000 | 0.0000 | 0.0016 | 0.5467 | 0.7605 | 0.3236 | 0.0000 | 0.0047 | 0.0000 | 0.0054 |
| 과거_외형채_1_y | 0.1635 | 0.0001 | 0.0029 | 1.0000 | 0.0000 | 0.9538 | 0.8437 | 0.8726 | 0.0000 | 0.0157 | 0.0093 | 0.3717 |
| 과거_외형채_2_y | 0.0022 | 0.0000 | 0.0000 | 0.0000 | 1.0000 | 0.9978 | 0.9092 | 0.8217 | 0.0000 | 0.1470 | 0.0002 | 0.2001 |
| 미래_외형채_1_y | 0.0069 | 0.5447 | 0.1126 | 0.9561 | 0.9285 | 1.0000 | 0.0005 | 0.0018 | 0.0021 | 0.0568 | 0.0001 | 0.0012 |
| 미래_외형채_2_y | 0.1496 | 0.2298 | 0.7634 | 0.8833 | 0.5179 | 0.0984 | 1.0000 | 0.0107 | 0.7969 | 0.0765 | 0.5582 | 0.1350 |
| 미래_외형채_3_y | 0.0539 | 0.7516 | 0.5454 | 0.9371 | 0.6283 | 0.0002 | 0.0258 | 1.0000 | 0.4250 | 0.2817 | 0.3220 | 0.0074 |
| 연준금리_y | 0.0169 | 0.0197 | 0.0012 | 0.0704 | 0.1891 | 0.2769 | 0.0780 | 0.0000 | 1.0000 | 0.0109 | 0.0000 | 0.0037 |
| AAII_y | 0.0332 | 0.2257 | 0.0154 | 0.4763 | 0.1222 | 0.2901 | 0.1015 | 0.3534 | 0.0112 | 1.0000 | 0.0476 | 0.0176 |
| VIX_y | 0.2593 | 0.1833 | 0.0586 | 0.0688 | 0.0122 | 0.0406 | 0.2807 | 0.0030 | 0.0000 | 0.4167 | 1.0000 | 0.3458 |
| Target_y | 0.0817 | 0.2577 | 0.0620 | 0.6886 | 0.1517 | 0.0006 | 0.1036 | 0.0010 | 0.0590 | 0.0621 | 0.0146 | 1.0000 |

Figure 7 Granger Causality Matrix

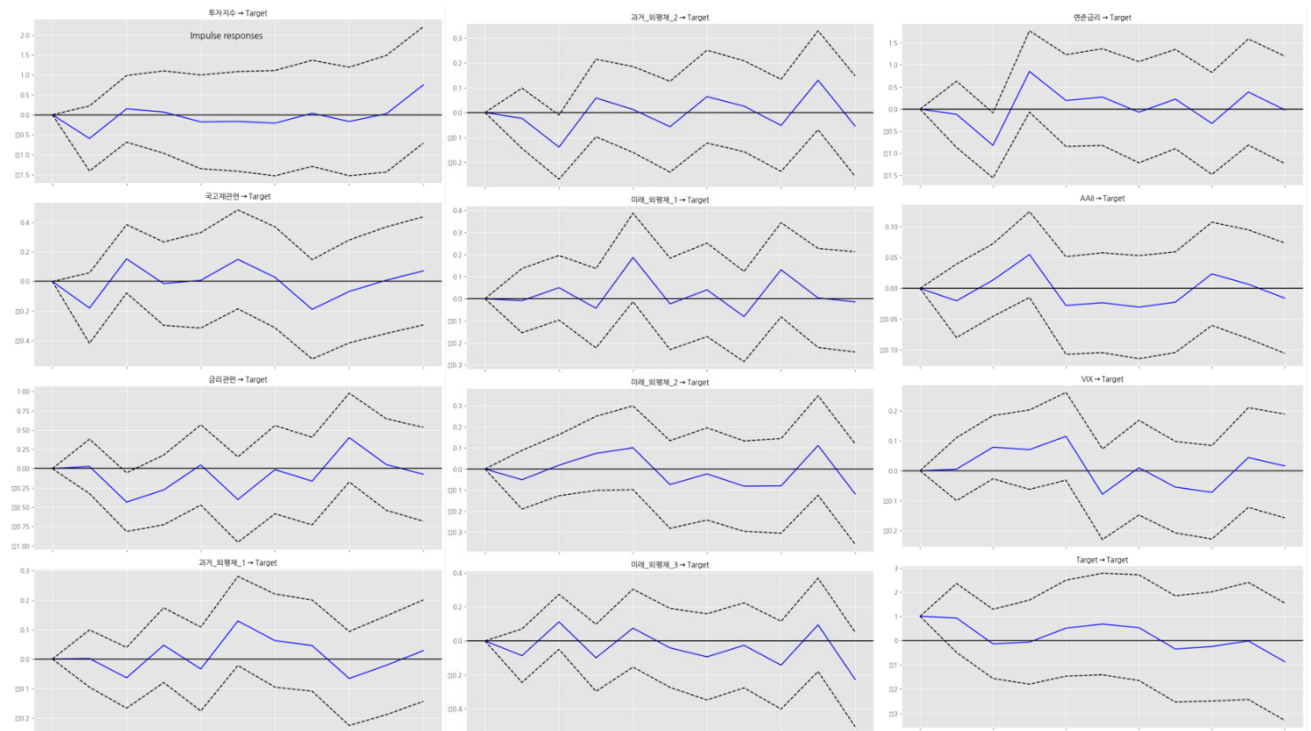


Figure 8 VAR 모델의 Impulse Response Function, impulse의 영향력이 |0.5| 이상인 변수들이 유의미한 반응을 주었다고 판단하여 변수 선정에 있어 참고하였다. Yticks의 범위가 변수별로 상이하므로 그래프 모양만을 본 것이 아닌 수치 위주로 판단하였다.

Information

Contact

paragonyun@naver.com

※본 보고서에 활용된 코드는 아래 깃허브에서 확인하실 수 있습니다.

<https://github.com/paragonyun/Deepconomics>