

# Machine Learning vs. Economic Restrictions: Evidence from Stock Return Predictability[1]

이종우  
FRE Lab

'24.09.06

1. Purpose
2. Methodology: GKX [4], CPZ [2], IPCA [7], CA[4], KNS[8]
3. Experiment and Result
4. Conclusion

- Anomaly들은 distressed stock(파산에 가까운 기업의 주식)에 대부분 포진하며 short position을 잡았을 때 이익이 대부분 발생한다는 연구들이 다수 존재
- Anomaly : 특정한 asset들은 기대 수익이 리스크에 대한 보상이 아님에도 발생 가능
- 하지만 anomaly를 파악하는 것은 최근 market liquidity와 arbitrage에 의해 매우 어려움
- 각각의 anomaly들을 통합하는 능력이 우수한 기계학습 방법론들은 mispricing을 파악할 수 있게 해주며 이에 기계학습 방법들을 주식시장에 적용하는 다양한 연구들이 진행되어 옴
- 본 논문은 방법론들의 성능 비교가 아닌, 기계학습 기반 투자 방법이 적절한 제약이 있는 투자환경에서도 여전히 잘 적용되는지와 어떠한 강점을 보이는지를 확인하고자 함

## 2. 방법론(1)

### 1. GKX[3]

- Three hidden layer FCNN, ReLU function, Batch Norm 그리고 학습에 있어 Lasso penalty 사용
- $E_t(r_{i,t+1}) = g(z_{i,t}) = (x_t \otimes c_{i,t})' \theta$ , 이 때  $c_{i,t}$ 는 각 주식의 변수(OHLCV 등)  $x_t$ 는 거시경제 변수

### 2. CPZ[2]

- Stock Discount Factor(SDF)를 no-arbitrage pricing theory를 인공지능망에 적용함으로써 추정
- Feedforward Network : SDF의 General functional form을 학습
- LSTM : SDF의 time variation을 거시경제적 상태의 함수로써 학습
- GAN : Largest possible pricing error를 minimize하는 SDF가 추산가능한 SDF에 가장 근사하다는 연구[5]에 근거하여 Asset pricing modeler는 best asset pricing model을 선택하고 adversary는 worst 모델에서의 moment condition을 선택하는 minmax 문제로 수식화
- 이는 가장 mispriced된 portfolio를 찾고 asset pricing model를 수정하는 것으로 해석이 가능
- 추정한 SDF를 통해 자산의 expected return을 계산하고 portfolio를 구성

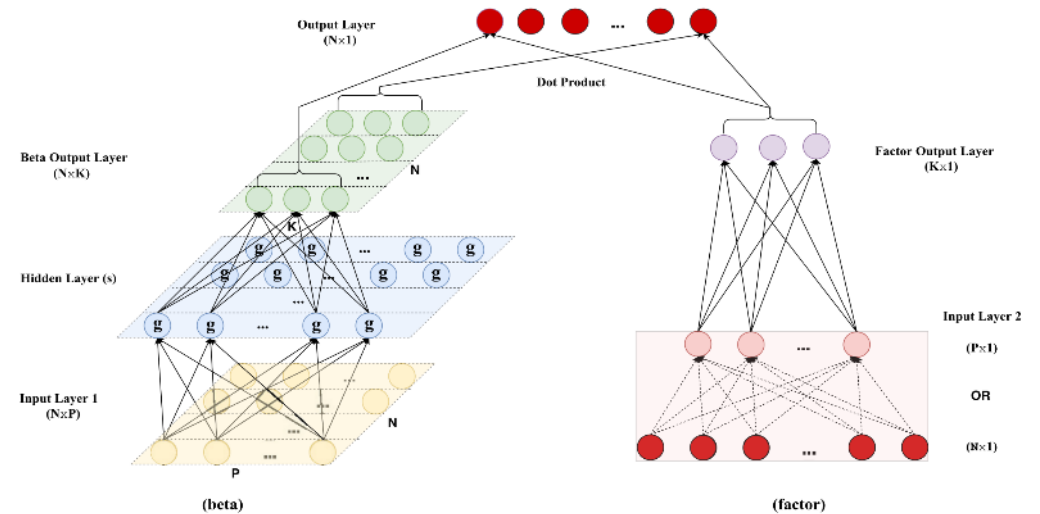
## 2. 방법론(1)

### 3. Instrumented Principal Component Analysis(IPCA)[7]

- Characteristic과 기대수익률의 관계가 latent risk factor에 의한 보상에 의해 유도되는 경우 이에 대응되는 latent factor를 식별, 반대의 경우 out risk에 의한 보상임을 인지하고 Anomaly intercept에 할당하는 모델

### 4. Conditional Autoencoder[4]

- Nonlinear factor model을 가정, Firm Characteristics를 포함하도록 하는 conditional autoencoder를 통해 latent factor들을 학습



Conditional Autoencoder model[4]

### 5. KNS[8]

- Equity portfolio를 사용하여 SDF를 추정
- Ridge regression with 3-fold cross validation을 통해 Hanssen-Jagannathan distance를 최소화하는 방향으로 학습

- Hanssen-Jagannathan distance : Distance between the SDF of asset pricing model and a set of correct SDFs and defined as  $\delta = \min_{m \in \mathcal{M}} \|m - y\|$  [6]

### 3. 실험(1) Data

---

- GKX, IPCA, CA Data : NYSE, AMEX 그리고 NASDAQ daily, monthly stock data, Quarterly and Monthly financial statement data, 74 industry section, macroeconomic predictors 사용
- CPZ Data : US stock data with 46 firm characteristics, 178 macroeconomic predictors
- KNS는 GKX, IPCA, CA를 통해 예측한 return을 통해 포트폴리오를 구성

### 3. 실험(2) 진행방식

---

1. Full sample을 사용하여 학습
  2. 상대적으로 cheap-to-trade stock sample로 학습(주식의 가격이 싼 것이 아닌 손해(비용)이 클 가능성이 낮은)
  3. 위의 sample에서 economic restrictions를 적용한 subsample들로 학습
    - ① Exclude microcaps(시가총액이 \$50M~300M인 기업)
    - ② Exclude nonrated firms(신용 평가가 없는 기업)
    - ③ Exclude financially distressed firms(신용 평가가 하락한 기업)
- 세 개의 실험을 진행
    - ① Economic restrictions를 적용했을 때 return predictability의 변화를 측정
    - ② 시간(또는 거시 경제 상황)에 따른 return predictability의 변화를 측정
    - ③ Full sample에서 machine learning method들이 어떠한 특성들을 추출하는지를 확인



## 4. 결과(1)

---

### 실험 1

- GKX, CPZ, CA는 economic restriction이 적용되었을 때 portfolio에서 상당한 성능하락을 보임
- Linear model인 IPCA는 Full sample에서는 GKX, CPZ, CA보다 성능이 낮으나 restriction 적용 sample에서 더 강건한 성능을 보임
- 네 방법론 모두 traditional 방법론(Fama-French Factor model, CAPM)보다 모든 sample에서 좋은 성능을 보임
- 또한 downside risk를 줄이고 extreme crash로부터 투자자를 보호할 수 있으며 portfolio rebalancing에서 high turnover를 보임
- KNS를 통해 SDF-implied tangency portfolio를 구성한 결과
  1. Traditional 방법론에 비해 높은 Sharpe ratio를 보임
  2. Economic restriction 적용 시 성능 하락
  3. Extreme long-short position을 취하는 결과를 보임

## 4. 결과(2)

---

### 실험 2

- GKX, CPZ는 high investor sentiment, high market volatility & low market liquidity 시장에서 더 좋은 portfolio 성능을 보임
- IPCA, CA와 같은 Beta pricing model은 trading profit의 low time series variation을 보였으며 low limits-to-arbitrage market에서는 다른 모델들을 능가함
- 모든 시장에서 economic restriction들을 적용하면 모든 모델의 성능이 하락함

### 실험 3

- 기계학습에 경제학의 empirical 또는 stylized fact들을 적용하면 mispriced stock들을 파악 가능
- Corporate investment가 많은 주식과 Idiosyncratic volatility를 가진 주식에 long position을 잡는 결과를 보임
  - Idiosyncratic volatility : 전체 시장의 volatility 로는 설명이 어려운 자산의 volatility
- 산업군 내 포트폴리오 전략이 산업군 간 포트폴리오 전략보다 더 좋은 성능을 보임

## 5. 결론

---

- 기계학습 기반 모델들은 market volatility가 높고 market liquidity가 낮은 시장에서 더 좋은 성능을 보임
- 기계학습 기반 SDF-tangency 포트폴리오에서 high turnover를 보이고 extreme long-short position을 취하는 결과를 보임
- Economic restriction들 추가하였을 때 성능이 꽤 하락하는 경향을 보였지만 더 적은 downside risk와 long position에서 이익을 만들어내는 결과를 확인함
- 경제학의 empirical facts를 통해 기계학습은 mispriced 주식들을 훌륭히 구별할 수 있음을 보임

- 1) Avramov, D., Cheng, S., & Metzker, L. (2023). Machine learning vs. economic restrictions: Evidence from stock return predictability. **Management Science**, 69(5), 2587-2619.
- 2) Chen, L., Pelger, M., & Zhu, J. (2024). Deep learning in asset pricing. **Management Science**, 70(2), 714-750.
- 3) Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. **The Review of Financial Studies**, 33(5), 2223-2273.
- 4) Gu, S., Kelly, B., & Xiu, D. (2021). Autoencoder asset pricing models. **Journal of Econometrics**, 222(1), 429-450.
- 5) Hansen, L. P., & Jagannathan, R. (1997). Assessing specification errors in stochastic discount factor models. **The Journal of Finance**, 52(2), 557-590.
- 6) Kan, R., & Zhou, G. (2004). Hansen-Jagannathan distance: Geometry and exact distribution. School of Economics and Finance, University of Hong Kong.
- 7) Kelly, B. T., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. **Journal of Financial Economics**, 134(3), 501-524.
- 8) Kozak, S., Nagel, S., & Santosh, S. (2020). Shrinking the cross-section. **Journal of Financial Economics**, 135(2), 271-292.