inverse covariance clustering의 국면을 활용한 동적 포트폴리오

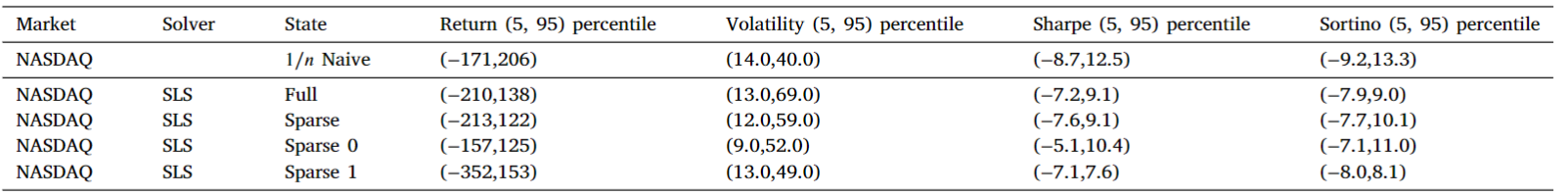
하나증권 퀀트 챌린지

FICC Quant modeling 부문 윤원식

본 프로젝트는 하나증권에서 주최한 FICC QUANT 부문의 프로젝트를 위한 자료입니다.

**참고 문헌**

참고한 문헌은 Yuanrong Wang, Tomaso Aste의 ‘Dynamic portfolio optimization with inverse covariance clustering’를 참고하였습니다. 해당 논문은 기존 마코프 국면이나 여러 머신러닝 군집화 방법의 한계점을 개선하여 inverse covariance clustering 방안을 활용해 자산 비중을 시간에 따라 유동적으로 변화시킬 수 있는 방법론을 제시합니다. 해당 논문에 의하면 ICC 방식의 국면모델은 짧은 시계열로 활용이 가능해 급격한 손실에 빠르게 대응할 수 있을 뿐만 아니라 국면의 안정성 측면에서도 우수해 VAR, Sharpe ratio, PML 등 다양한 지표 측면에서 우수한 성과를 보였습니다.

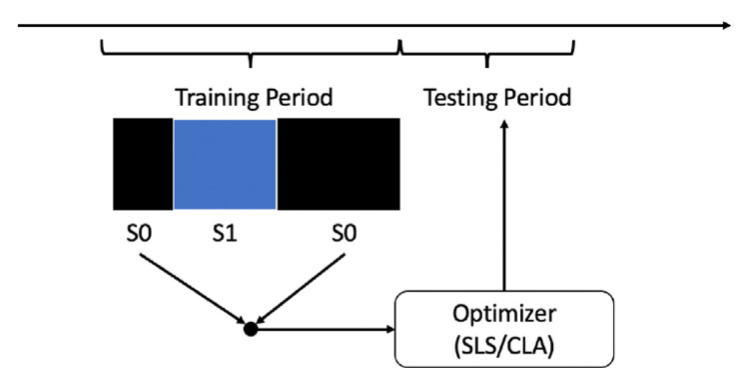


<그림 1: 논문 속 성과 NASDAQ 성과 평가 결과>

따라서, 본 프로젝트는 해당 논문의 방법론을 인용하여 모델을 구현하였습니다.

해당 논문의 Asset allocation 방법론은 다음과 같이 크게 3단계로 나누었습니다.

먼저, 주어진 여러 거시적 금융위험 지표와 각종 index지표들이 담겨있는 train dataset에서 inverse covariance clustering(ICC) 방법을 활용해 time series별로 군집화를 진행합니다. 논문에서 나온 방식처럼 훈련기간 마지막 최근 20일을 기준으로 가장 우세한 그룹을 향후 지속될 것이라 가정을 하고 진행을 합니다. 가장 우세한 최근의 국면과 동일한 군집을 따로 추출하여 각 자산별 기대 수익률과 변동성 및 공분산을 추정하여 mean-variance optimization을 진행합니다. 해당 과정을 통해 도출된 비중을 다음 한달간 수익을 관찰합니다.



<그림2 : 논문 속 모델링의 예시>

해당 논문의 장점으로는 기존 히든 마르코프를 이용한 국면모델이 해결하지 못한 차원문제를 해결할 수 있습니다. 여러 금융지표와 가격데이터를 혼합하여 국면을 측정하면 자연히 차원이 늘어날 수밖에 없는데, 기존 마코프 모형에서는 차원의 저주(curse of dimensionality)에서 벗어날 수 없었습니다. 또한, Gaussian Mixture나 KNN을 활용한 군집화 역시 제시된 바가 있지만, 이는 시간에 흐름에 따라 의존하는 구조적 특성을 파악하는데 어려움이 있습니다. 이러한 한계점을 해결하기위해 본 논문은 기존의 센서 신호데이터에 사용되는 ICC method를 활용해 극복하고자 하였습니다. ICC method는 시계열간의 네트워크 구조의 변화를 감지하며 짧은 시계열의 subsuquence의 likelihood와 sparse precision matrix를 기반으로 국면의 빠른 변화와 국면 모수 추정의 강건함을 보이는 결과가 나왔습니다. 이에 따라, 본 논문 방법론을 활용하여 고정 금리형 ETF 및 금 ETF를 이용한 최적 자산배분 전략을 개발해보았습니다. 위의 논문은 주식 시장의 주식들을 이용하여 자산배분을 실행하였지만, 해당 전략은 채권과 금, 달러만을 이용하여 자산배분하였습니다. 달러의 경우, DXY에 투자하는 것이 아닌 어느 곳에 투자하지 않고 환전된 달러로 가지고 있다고 생각하였습니다. 금융위기 시장에서 늘 채권보다도 달러가 강세인 점을 생각하여 해당 전략을 취하였습니다.

또한, 장기 단기 금리, 모든 채권들의 데이터를 활용해 국면을 측정하면 더욱 정확한 자산배분이 가능하겠지만, 해당 프로젝트에서 모든 데이터를 모수를 추정하기에 충분한 데이터기간을 구하기 힘든 점이 있었습니다. 본 프로젝트에서의 가장 큰 제약은 ‘데이터’ 입니다. 해당 ETF 혹은 벤치마크의 데이터가 충분히 많으면 국면에 따라 포트폴리오 비중을 설정하기 쉽지만, 모든 데이터가 존재하는 시점은 조사한 바로는 최대 2010년도입니다. 2010년~2014년까지로 국면 분석을 하는 것은 지나치게 좁은 범위의 시계열을 가지고 국면을 분석하기에 의미있는 비중 추정이 어렵다고 판단하였습니다. 따라서, 본 프로젝트에서는 시계열의 길이가 긴 대표적인 자산군(ex. 금, 하이일드 채권, 국채 ETC)의 index 데이터와 금융 지표들을 가져와 국면을 추정하였습니다. 그리고 국면에 따라 어느 자산군이 우세한지 확인하는 과정을 거치고 대표적인 자산군 지수들을 활용하여 risk-parity 포트폴리오를 만들게끔 하였습니다. risk-parity를 적용한 이유는 기존 논문에서 적용한 MVO 포트폴리오는 추정할 모수가 많기 때문에 상당히 불안정한 특성을 가지고 있다. 뿐만 아니라 해당 전략은 국채의 가격 상승을 통해 적극적인 차익거래를 하는 전략이 아닌 고정금리를 꾸준히 받으며 채권 시장의 급격한 하락을 방어하며 운용하는 포트폴리오입니다. 따라서 샤프지수자체를 초점 맞추기보다는 변동성 측면에서 접근하는 리스크패리티가 가장 적합하다고 생각하였습니다. 이후 비중을 산출하면 산출한 자산군 비중에서 단기 장기채권의 비중을 리스크 패리티 전략으로 동일 위험으로 자산을 배분하게끔 하였습니다.

전체적인 방법론을 요약하면 다음과 같습니다.

1. 우선, 큰 자산군 6개를 다음과 같이 나눕니다.

하이일드(회사채 포함), 금, 국채, 달러, TIPS

1. 이후 자산군에 따라 대표적인 지표들과 각종 금융 지표들을 활용하여 자산의 국면을 ICC-PO method를 활용하여 국면을 추정합니다.
2. 추정한 국면 중 가장 최근 10주 동안 가장 많이 관찰된 국면을 현재의 국면이라고 판단합니다. (논문에서는 20일로 가정)
3. 이후 추정한 최근 국면과 같은 국면이었던 과거의 데이터들을 merge하여 risk-parity 포트폴리오 최적화를 하여 자산군마다의 비중을 산출합니다.
4. 이후 자산군 속에서 장기,단기,중기 채권들의 비중을 risk-parity 전략으로 동일 위험 기여도로 자산을 배분하게끔 설정해줍니다.
5. 매월 월간 리밸런싱을 위의 1~5단계를 반복 진행해줍니다.

**활용 데이터**

우선 데이터는 14년까지 밖에 없어 모수를 추정하기 위해서 아래의 데이터들을 블룸버그 터미널을 통해 수집하였습니다. 국면에 활용한 데이터로는 다음과 같습니다.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **TICKER** | USGG3M  Index | NFCIADJ  Index | LUACTRUU  Index | DXY Curncy | XAU BGN  Curncy |
| **Data**  **Description** | 3month corporate bond | Chicago FED Adjusted Financial Condition | corporate\_  bond | dollar  index | Gold  index |
| **TICKER** | CSI BARC  Index | I14031US  Index | LUACOAS  Index | TIPS  index | US0003M  Index |
| **Data**  **Description** | Highyield spread | Highyield data | Investment Grade | TIPS | US treasury\_bond |
| **TICKER** | VIX index |  |  |  |  |
| **Data**  **Description** | VIX |  |  |  |  |

<표1 : 국면데이터를 구하기 위해 사용된 데이터>

제시된 데이터중 활용하지 않은 데이터로는 원달러 환율과 LG98OAS Index 입니다. 원달러 환율의 경우는 ETF 자산들 대부분은 외국 자산임에 비해 원달러 환율은 국내의 위험도 포함된 지표이기에 적절하지 않다고 판단하였습니다. 이에 따라 원달러 환율대신에 DXY index를 편입하여 미국 시장의 위험을 위주로 국면을 판단해주게끔 하였습니다. 하이일드 스프레드인 LG98OAS Index는 블룸버그에서 찾을 수 없는 티커로 나와 해당 데이터를 활용하지 않고 CSI BARC Index를 활용하였습니다. 위의 데이터 모두 주간으로 데이터를 변환해주었고 데이터의 트렌드를 제거해주기 위해 논문에서와 같이 차분을 하여 진행하였습니다. 추가적으로 VIX 지표를 사용했습니다. 이는 시장의 급격한 하락을 가장 잘 반영하는 지표이기에 추가하였습니다.

추가적으로 주어진 ETF 데이터 역시 변동성을 추정하기엔 데이터 수가 부족하여 과거 데이터를 블룸버그 터미널을 통해 가져왔습니다.

1. ice u.s. treasury 20+ year bond index(TLT)
2. ICE BofA 7-10 Year US Corporate Index Total Return Index Value(IEF)
3. ICE US Treasury 1-3 Year Index(SHY)
4. ICE Short US Treasury Securities Index(SHV)
5. ICE BofA 15+ Year US Inflation-Linked Treasury Index(LTPZ)  x
6. Bloomberg U.S. Treasury Inflation-Protected Securities (TIPS) 0-5 Year Index (VTIP)
7. Bloomberg U.S. 10+ Year Corporate Bond Index.  (VCLT)
8. Bloomberg U.S. 5-10 Year Corporate Bond Index (VCIT)
9. Bloomberg U.S. 1-5 Year Corporate Bond Index (VCSH)
10. ICE BofA US Treasury Bill Index (GSY)
11. Markit iBoxx® USD Liquid High Yield Index(HYG)
12. ICE BofA US High Yield TR USD (SHYG)
13. Gold(IAU)
14. ICE U.S. Treasury Core Bond Index(GOVT)
15. Bloomberg U.S. Treasury Inflation Protected Securities (TIPS) Index
16. Markit iBoxx® USD Liquid Investment Grade Index (LQD)

**파라미터 설정**

해당 전략의 파라미터는 총 4가지가 있습니다.

1. **국면의 개수**

먼저 Dynamic portfolio optimization with inverse covariance clustering에 따르면, 해당 군집화 방법에서 국면의 개수는 자유롭게 정해도 되지만, 적절한 국면을 선택하려면 BIC값을 기준으로 선택해야 한다고 합니다. 이에 따라 데이터가 모두 존재하는 2002년부터 2013년까지 국면을 2개부터 5개까지 나누어본 결과의 BIC는 다음과 같습니다.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 국면개수 | 2\_regime | 3\_regime | 4\_regime | 5\_regime |
| BIC | 33740 | 18804 | 36651 | 44584 |

국면의 개수가 3개에서 가장 작은 것을 알 수 있었으며, 이는 파라미터를 조정해도 동일한 결과를 얻었습니다. 따라서 본 프로젝트는 국면의 개수를 3개로 한정하여 진행을 하였습니다.

1. **Train dataset의 기간**

데이터가 모두 존재하는 2000년 12월부터 시행하였습니다. 이는 위의 데이터들이 모두 주간 데이터로 존재하는 날짜입니다. 이전 날짜에서는 월간데이터로 존재하는 경우가 대부분이라 2000년 12월부터 국면을 추정하는 데이터로 넣어주었습니다.

1. **비중 변경의 주기**

비중 변경의 주기는 논문에서와 동일하게 월간 리밸런싱을 하였습니다.

1. **국면을 판단한 기준**

국면을 판단하는 기준은 논문에서와 동일하게 최근 n일간 가장 많이 관찰된 국면을 앞으로 다음 한달간 유지된 국면이라고 가정하였습니다. 논문에서는 20일로 추정을 하였지만, 해당 프로젝트에서는 5주로 추정을 하였습니다.

1. **포트폴리오 최소 비중**

각 자산군별 비중을 최소 5% 최대 50%까지로 제약조건을 걸어주었습니다. 이는 국면의 추정이 불안정할 경우 MVO 최적화 비중이 한쪽 자산군으로 쏠릴 수 있기 때문에 이를 방지하기 위해 설정해주었습니다.

따라서, 위와 같은 이유로 parameter을 선정하여 모델의 성과를 측정해보았습니다. 그외 ICC의 파라미터 (iteration,beta)는 논문의 값과 비슷한 각각 0.11,300으로 설정하였습니다.

**코드 세부 내용**

다음으로는 코드의 과정을 설명 드리겠습니다. 모두 class화를 하지 않고 데이터 정제와 자산군 속 세부 비중설정은 주피터 노트북으로 과정을 담았습니다. 이는 데이터 전처리 과정이 복잡하여 이해가 안되실 수 있는 점을 감안하여 파트를 총 3개로 나누어 코딩을 진행하였습니다.

전체적인 코드 실행 과정은 다음과 같습니다.

1. Data\_preprocessing\_step1

해당 주피터노트북 파일은 국면모델에 넣어줄 인풋 데이터를 정제해주는 파일입니다. 아웃풋 값으로는 all\_ETF\_data과 all\_index\_data 데이터를 내보냅니다. all\_ETF\_data는 비중 최적화를 해주기 위해 필요한 ETF 과거 데이터입니다. all\_index\_data는 국면 모델에 필요한 금융위험지표와 기타 인덱스 지표들이 담겨있습니다.

1. ficc\_quant/TICC/ ICC\_regime\_step2.py

해당 파일은 Data\_preprocessing\_step1에서 나온 데이터들을 활용하여 국면을 날짜마다 예측하고, 국면을 활용해 MVO 최적화를 해주는 코드입니다. 날짜별 자산군 투자 비중이 담긴 데이터프레임이 output 값으로 산출되며, 이를 활용하여 자산 군집 내에서 step3코드에서 risk parity 전략을 실행될 수 있게끔 해줍니다.

1. weight\_riskparity\_step3

해당 파일은 3번째 단계로 step2에서 나온 비중 값을 가지고 자산군 내에서 risk parity 전략을 시행하여 최종 비중을 산출되게끔 해줍니다. Output 파일은 ‘last\_portfolio\_weight.xlsx’파일이며, 날짜별 ETF의 비중이 담겨있습니다. 또한, 백테스팅을 하기 위한 ‘all\_ETF\_data.xlsx’ 파일 역시 내보냅니다. 해당 파일은 백테스팅에 필요한 ETF의 과거 데이터를 변화율 형태로 변환해줍니다.

1. Backtesting\_step4

해당 파일은 최종적으로 나온 비중을 가지고 백테스팅을 하는 코드입니다. 백테스팅의 비교 벤치마크는 동일비중 투자 포트폴리오입니다.

최종 결과