Fama French Factor Model

모두를 위한 Factor Series I

David Geonsu Kim, Kalvin Kiyoon Jang

모윅 cOMPANY

**[0] Abstract**

CAPM 모형의 확립 이후로 시장에 존재하는 리스크 요소를 팩터로 정의하며 이를 활용하여 포트폴리오의 수익률을 설명하고, 궁극적으로 초과수익률을 거두기 위한 시도가 지금까지 계속되었습니다. 그 중에서 가장 유명한 연구는 Fama and French (1992, 1993)의 연구로, CAPM의 시장수익률 팩터에 더해, 소형주와 가치주가 갖는 리스크 팩터를 추가하여 포트폴리오의 초과수익률을 설명하기 위한 시도였죠. 이는 팩터의 대표적인 연구로 일컬어지며 현재까지도 팩터 모델링의 대명사로 자리매김하고 있습니다. Fama and French는 이어서 2015년 논문을 통해 투자와 수익성에 대한 리스크 팩터를 추가하여 5팩터 모델을 추가적으로 발표하였고 기존 3팩터로 설명되지 못한 초과수익률에 대해 보완한 연구를 내놓았습니다.

앞으로 저희(회사명)가 발간할 리포트 시리즈는 팩터 모델링을 보다 많은 대중에게 알려서 이해를 돕고 직접 replication할 수 있도록 API와 데이터를 제공하는 데 도움을 줄 것을 목적으로 합니다. 뿐만 아니라 궁극적으로는 저희만의 팩터를 독자적인 모델을 통해 찾고 이를 대쉬보드를 이용해 모든 사람들이 자유롭게 볼 수 있도록 할 것입니다. 이를 위해서는 우선적으로 Fama-French 5팩터 모델에 대한 이해가 선행되어야 할 것이며 우리의 첫 리포트인 이번 리포트가 팩터 모델에 대한 넓고 깊은 이해를 가져다줄 것으로 기대하고 있습니다. 리포트는 다음과 같이 구성됩니다. 1장은 Fama-French 팩터 모델의 논문에 나온 내용과 팩터를 나누는 기준, data에 대해 명확히 밝히고 해당 논문이 갖는 의의를 요약하는 부분입니다. 2장에서는 저희가 한국 회사들의 시장, 재무 데이터를 통해 replication한 결과와 그 논리를 소개합니다. 3장은 이번 리포트를 발간하면서 발행자인 David와 Kalvin의 논의를 바탕으로 2장에 나왔던 논리를 보충하고, 팩터 모델에 대한 이해를 심화시킬 수 있도록 David와 Kalvin의 짧은 논평을 실었습니다. 4장에서는 마지막으로 우리의 API와 코드에 대한 설명을 제공하며 독자들이 replication하기 편할 수 있도록 했습니다.

한국 시장에 대한 replication 결과는 다음과 같이 요약될 수 있습니다. 1) FF 5F와 같이 모든 팩터가 통계적으로 유의하게 포트폴리오의 초과수익률을 설명할 수 있었으며 이런 비중은 5x5 포트폴리오 중 ~~로 갈수록 높아졌다. 2) 모델의 설명력인 R^2가 Fama and French의 논문보다는 낮은 경향을 가지는 것으로 나타났다. 3) ~~~~~

이번 리포트의 의의는 팩터 모델링의 근간이 되는 연구인 Fama-French 5팩터 모델을 한국 시장에서 replication하고 이에 대한 과정과 결과를 모든 사람이 보면서 따라할 수 있도록 공개했다는 것에 있습니다. 지금껏 여러 기관 및 개인이 해당 논문을 replication한 시도는 있어왔지만, Factor만을 공개하거나 팩터를 정의하는데 필요한 기준을 자의적으로 변경하여 각기 다른 기준을 갖고 팩터를 정의하는 문제점이 있었습니다. 이런 부분 때문에 개인이 실습으로 Fama-French 5팩터 모델을 구현하기에도 어려움이 따랐을 거라 생각합니다. 저희의 이번 리포트는 이런 문제점을 해결하기 위해 replication하는 코드를 공개해 모두가 동일한 데이터와 기준을 갖고 팩터를 정의하고 모델링할 수 있도록 하는데 집중했습니다. 저희의 이런 시도가 팩터 모델링의 대중적인 이해를 넓히는 데 긍정적인 영향을 주기를 바랍니다.

**[1] Paper review**

**(1) 모델의 설정**

Fama and French (2015) 연구는 회귀식을 팩터 회귀식을 이용해서 초과수익률에 대한 팩터의 설명력을 찾는 연구였습니다. 데이터는 1963년 7월부터 2013년 12월까지의 606개월간의 NYSE, AMEX, NASDAQ 데이터였으며 팩터는 시장 초과수익률 팩터에 SMB, HML, RMW, CMA를 더한 5개로 설정했습니다. 아래의 표1은 Fama and French의 연구에서 팩터를 어떻게 정의했는지에 대한 설명입니다. 시장초과수익률을 나타내는 Rm-Rf 팩터를 제외한 나머지 4개의 팩터는 모두 특정 기준에 대해 상위 포트폴리오와 하위 포트폴리오를 구성한 후, 그 둘의 수익률 차이로 나타낼 수 있습니다. Rm-Rf 팩터는 CAPM에서와 마찬가지로 시장 초과수익률로 이루어진 팩터입니다.

표 1. Fama-French 5팩터 모델의 2x3 sorts 팩터

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 팩터명 | 팩터 기준 | Masking 결정 세부 식 | Masking 시점 | 포트폴리오 breakpoint |
| Rm-Rf | Market return | 시장 초과수익률 | - | - |
| SMB | Size (크기) | 시가총액 | 6월 말 | NYSE의 중앙값 |
| HML | B/M (평가수준) | 자본총계 / 시가총액 | 12월 말 | NYSE의 [30%, 70%] |
| RMW | OP (수익성) | (영업손익 - 금융비용) / 총자산 | 12월 말 | NYSE의 [30%, 70%] |
| CMA | Inv (투자패턴) | 당기순이익 / 총자산 | t-1기 12월 말  / t-2기 12월 말 | NYSE의 [30%, 70%] |

사실 Fama and French (2015) 논문에서는 표1에서 제시하고 있는 포트폴리오 breakpoint이외에도 두 가지의 조합을 더 고려했습니다. 표1에서 보이는 팩터는 2x3 sorts 기준의 팩터이고, 표1 이외에 2x2 sorts 기준 팩터와 2x2x2x2 sorts 기준 팩터라는 두 가지의 팩터 세트가 논문에는 추가적으로 수록되어있습니다. 2등분을 할 때 중앙값으로 한 이유, 3등분을 할 때 상하위 30%를 설정해서 두 포트폴리오의 수익률 차를 팩터로 정의한 이유에 대해서는 저자도 이유를 밝히고 있지는 않습니다. 다만 해당 breakpoint에서 가장 모델의 성과가 두드러지게 나왔기 때문에 기준을 저렇게 세웠을 것으로 추측할 수는 있습니다.

그리고 팩터와 혼동하면 안 되는 것이 바로 회귀식에서 종속변수에 해당하는 포트폴리오 수익률입니다. 종속변수에 들어가는 포트폴리오도 팩터들과 동일한 기준을 통해 만들어지지만, 차이점은 우선 시가총액을 통해 Size를 2등분 또는 5등분으로 분류한 후, 분류된 해당 포트폴리오 안에서 추가적인 기준을 이용해 세분화한다는 점입니다. 25 Size-B/M 포트폴리오의 경우, 시가총액을 기준으로 5등분 포트폴리오를 만든 후, 각 포트폴리오별로 B/M기준을 통해 다시 한 번 5등분 해서 총 25개의 포트폴리오를 만든 것입니다. 32 Size-B/M-Inv의 경우는 시가총액을 기준으로 2분할 포트폴리오를 만든 후, 각 포트폴리오 안에서 B/M과 Inv를 기준으로 4등분해서 32개의 포트폴리오를 만드는 것을 말합니다. 아 그리고 종속변수에 들어가는 포트폴리오는 시가총액에 비례해서 비율을 정하게 됩니다.

표 2. Fama-French 5팩터 모델의 경우의 수

|  |  |
| --- | --- |
| 6세트의 포트폴리오 구성 | 25 Size-B/M, 25 Size-OP, 25 Size-Inv,  32 Size-B/M-OP, 32 Size-B/M-Inv, 32 Size-OP-Inv |
| 3세트의 팩터 구성 | 2x3 sorts on Size and B/M, or Size and OP, or Size and Inv  2x2 sorts on Size and B/M, or Size and OP, or Size and Inv  2x2x2x2 sorts on Size, B/M, OP, Inv |
| 7개의 회귀모형 | 3개의 Three-factor models (Rm-Rf + SMB + HML / RMW / CMA)  3개의 Four-factor models (HML/RMW/CMA 중 하나씩만 뺀 회귀식)  1개의 Five-factor models |

정리하자면 Fama and French (2015) 논문은 6세트의 포트폴리오 구성을 종속변수에 대해 3세트의 팩터 구성과 7개의 회귀모형을 고려한 회귀분석을 진행한 연구라고 할 수 있습니다.

**(2) 모델의 결과**

3세트의 팩터 구성에서는 2x2나 2x3이나 2x2x2x2나 비슷한 성과를 보였고 three factor에서는 2x3을 사용했기 때문에, 그리고 2x2x2x2의 경우에는 결과가 robust하지 않기 때문에 2x3을 추천한다고 Fama and French는 밝히고 있습니다. 회귀모형 선택에서는 HML의 영향이 다른 팩터에 의해 대부분 설명되기 때문에, HML을 제외한 4팩터 모형이나, HML팩터를 다른 4개의 팩터 공간에 직교시킨 HMLO팩터를 대신 사용한 5팩터 모형을 추천한다고 말합니다. 6세트의 포트폴리오 구성 중 4세트에서, RMW와 CMA의 팩터계수가 음수인 소형주가 가장 좋지 못한 성과를 보였으며, 이는 적은 수익성에도 불구하고 많은 투자를 하는 소형주는 심각한 성과를 보인다는 의미라고 논문에서는 해석하고 있습니다.

표3과 표4, 표5는 Fama and French(2015)의 결과를 옮겨놓은 표입니다. 저자가 결론에서 추천한 2x3 sorts 5팩터 모델에 대한 결과값이며, 종속변수는 표3의 경우 Size-B/M으로 구분한 25포트폴리오, 표4의 경우 Size-OP로 구분한 25포트폴리오, 표5의 경우 Size-Inv로 구분한 25포트폴리오입니다. 논문에서도 이 세 가지의 결과를 중점으로 결과를 보여주었기 때문에 많은 조합 중 세 가지 결과를 선택해서 저희 리포트에 실었습니다.

표 3. Size-B/M 기준 25포트폴리오에 대한 회귀분석 결과

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | |
| B/M | Low | 2 | 3 | 4 | High | Low | 2 | 3 | 4 | High |
| Five-Factor coefficients: Rm-Rf, SMB, HMLO, RMW, CMA | | | | | | | | | | |
|  |  |  | a |  |  |  |  | t(a) |  |  |
| Small | -0.29 | 0.11 | 0.01 | 0.12 | 0.12 | -3.31 | 1.61 | 0.17 | 2.12 | 1.99 |
| 2 | -0.11 | -0.10 | 0.05 | -0.00 | -0.04 | -1.73 | -1.88 | 0.95 | -0.04 | -0.64 |
| 3 | 0.02 | -0.01 | -0.07 | -0.02 | 0.05 | 0.40 | -0.10 | -1.06 | -0.25 | 0.60 |
| 4 | 0.18 | -0.23 | -0.13 | 0.05 | -0.09 | 2.73 | -3.29 | -1.81 | 0.73 | -1.09 |
| Big | 0.12 | -0.11 | -0.10 | -0.15 | -0.09 | 2.50 | -1.82 | -1.39 | -2.33 | -0.93 |
|  |  |  | h |  |  |  |  | t(h) |  |  |
| Small | -0.43 | -0.14 | 0.10 | 0.27 | 0.52 | -10.11 | -4.38 | 3.90 | 10.12 | 17.55 |
| 2 | -0.46 | -0.01 | 0.29 | 0.43 | 0.69 | -15.22 | -0.45 | 11.77 | 16.78 | 24.44 |
| 3 | -0.43 | 0.12 | 0.37 | 0.52 | 0.67 | -14.70 | 3.71 | 12.28 | 17.07 | 18.75 |
| 4 | -0.46 | 0.09 | 0.38 | 0.52 | 0.80 | -15.18 | 2.76 | 11.03 | 15.88 | 20.26 |
| Big | -0.31 | 0.03 | 0.26 | 0.62 | 0.85 | -14.12 | 1.09 | 7.54 | 21.05 | 18.74 |
|  |  |  | r |  |  |  |  | t(r) |  |  |
| Small | -0.58 | -0.34 | 0.01 | 0.11 | 0.12 | -13.26 | -10.56 | 0.31 | 3.89 | 3.95 |
| 2 | -0.21 | 0.13 | 0.27 | 0.26 | 0.21 | -6.75 | 4.89 | 10.35 | 9.86 | 7.04 |
| 3 | -0.21 | 0.22 | 0.33 | 0.28 | 0.33 | -6.99 | 6.77 | 10.36 | 8.98 | 8.88 |
| 4 | -0.19 | 0.27 | 0.28 | 0.14 | 0.25 | -6.06 | 7.75 | 7.99 | 4.16 | 6.14 |
| Big | 0.13 | 0.25 | 0.07 | 0.23 | 0.02 | 5.64 | 8.79 | 2.07 | 7.62 | 0.49 |
|  |  |  | c |  |  |  |  | t(c) |  |  |
| Small | -0.57 | -0.12 | 0.19 | 0.39 | 0.62 | -12.27 | -3.46 | 6.59 | 13.15 | 19.10 |
| 2 | -0.59 | 0.06 | 0.31 | 0.55 | 0.72 | -17.76 | 1.94 | 11.27 | 19.39 | 22.92 |
| 3 | -0.67 | 0.13 | 0.42 | 0.64 | 0.778 | -20.59 | 3.64 | 12.52 | 18.97 | 19.62 |
| 4 | -0.51 | 0.31 | 0.51 | 0.60 | 0.79 | -15.11 | 8.33 | 13.35 | 16.41 | 18.03 |
| Big | -0.39 | 0.26 | 0.41 | 0.66 | 0.73 | -16.08 | 8.38 | 10.80 | 19.88 | 14.54 |

표3에서는 Size-B/M 기준으로 분류한 25개의 포트폴리오에 대한 팩터별 계수 및 알파값과 그에 따른 t통계량이 나와있습니다. 팩터가 설명하지 못하는 초과수익률을 의미하는 a를 주목했을 때, B/M기준으로 낮은 포트폴리오일수록 a값의 t통계량의 절대값이 높은 경향을 볼 수 있습니다. 또한 h, r, c값을 볼 때, B/M기준으로 낮은 포트폴리오일수록 계수들이 음의 값을 띄는 경향인 짙으며, t통계량 상으로도 매우 유의한 결과를 나타냅니다. 반면, 포트폴리오 간에 Size별로는 큰 차이는 보이지 않았습니다.

표 4. Size-OP 기준 25포트폴리오에 대한 회귀분석 결과

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | |
| B/M | Low | 2 | 3 | 4 | High | Low | 2 | 3 | 4 | High |
| Five-Factor coefficients: Rm-Rf, SMB, HMLO, RMW, CMA | | | | | | | | | | |
|  |  |  | a |  |  |  |  | t(a) |  |  |
| Small |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |
| Big |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Small |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |
| Big |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Small |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |
| Big |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Small |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |
| Big |  |  |  |  |  |  |  |  |  |  |

표 5. Size-Inv 기준 25포트폴리오에 대한 회귀분석 결과

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | |
| B/M | Low | 2 | 3 | 4 | High | Low | 2 | 3 | 4 | High |
| Five-Factor coefficients: Rm-Rf, SMB, HMLO, RMW, CMA | | | | | | | | | | |
|  |  |  | a |  |  |  |  | t(a) |  |  |
| Small |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |
| Big |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Small |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |
| Big |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Small |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |
| Big |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Small |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |
| Big |  |  |  |  |  |  |  |  |  |  |

**[2] Experiment**

저희는 위에서 살펴본 Fama and French (2015)의 방법론과 결론의 한국 주식시장에서도 적용되는지 알아보기 위해 한국 데이터로 재구성하였습니다. 데이터는 2009년 10월부터 2021년 6월까지, 141개월 기간에 대한 KOSPI, KOSDAQ 데이터입니다. 기업의 가격과 시가총액데이터는 KRX데이터를 이용했으며, 재무데이터는 Sirius\_API[[1]](#footnote-1)를 활용하여 가져왔습니다. 팩터를 계산하기 위한 포트폴리오 masking은 표1과 동일한 형태를 취했으며, NYSE대신 KOSPI로 대체하여 breakpoint를 설정했습니다. 또한 시장수익률은 KOSPI와 KOSDAQ의 시총가중 수익률을 택하였고, 무위험이자율은 CD91일물 수익률로 계산하였습니다. 표6은 저희가 구현한 모형에 대한 결과값이며

표 6. 한국데이터로 구현한 Fama-French 5팩터 모형 결과값

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | |
| B/M | Low | 2 | 3 | 4 | High | Low | 2 | 3 | 4 | High |
| Five-Factor coefficients: Rm-Rf, SMB, HMLO, RMW, CMA | | | | | | | | | | |
|  |  |  | a |  |  |  |  | t(a) |  |  |
| Small |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |
| Big |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Small |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |
| Big |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Small |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |
| Big |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Small |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |
| Big |  |  |  |  |  |  |  |  |  |  |

**[3] Talk**

(3) Talk. transcript처럼 대화체로 쓰되, 위에서 쓰기 애매한 내용들을 쓰자. 생동감있게? 친근하게? 적으면 어떨까? —> AQR에 cliff’s perspective 아티클이 구어체로 적혀있긴 하다. 편하게 잘 읽힘.

🡪 오픈 디스커션 느낌. 자유롭게. 서평 형식으로 한 문단씩.

David:

Kalvin:

David:

Kalvin:

David:

Kalvin:

팩터가 갖는 의미. 팩터 계수가 민감도 이외의 해석이 있지 않을까.

R^2가 높게 나오는 이유가 뭘까.

이걸로 수익을 낼 거면 어떻게 해야할까

[4] Reference

1. Sirius API는 저희가 만든 API로, 가공된 재무데이터를 손쉽게 불러올 수 있는 API입니다. 저희 홈페이지의 Sirius API Guideline을 참고해주시기 바랍니다. [↑](#footnote-ref-1)