**※ 애널리스트 보고서 제목의 정보력 검증: 텍스트 어조를 중심으로**

|  |  |
| --- | --- |
| **요약** | ● 애널리스트 보고서 제목에서 추출한 텍스트 어조의 정보력 검증  ● Data: 2009년 ~ 2018년 한국에서 발행된 애널리스트 보고서 자료  ● Purpose: 정보변수와 발표일 주변 5일 간의 누적초과수익률과의 관계를 살펴봄  ● Result:  - 추천의견 변경, 목표주가 변경, 텍스트 어조에 대해 유의미한 시장반응을 발견  - 특히 제목 텍스트의 어조(tone)는 다른 정보 변수들을 통제한 후에도 유의미한 시장 반응 발견  - 부정적 어조의 정보력이 더 큼↑  - 소규모 기업, 애널리스트 수가 ↓인 기업 = 정보효과 ↑  - 어조의 정보력을 증권사 및 애널리스트의 특성에 따라 구분했을 때 증권사의 특성은 유의미한 반면, 애널리스트 특성은 유의미한 결과 X |
| **주제어** | 애널리스트 보고서, 추천의견, 목표주가, 이익예측치, 텍스트 분석, 어조 |
| **자료** | ● 애널리스트 보고서: FnGuide의 DataGuide를 통해서 추출  - 유가증권과 코스닥 시장 포함  - 애널리스트 자료를 1년과 분기별 예측기준으로 제공 -> 1년 예측기준 자료 사용  - 2009 ~2018년 총 74,548 개의 관측치  ● 보고서 어조를 긍정어와 부정어 목록을 사용하여 측정  ● 주가 자료: FnGuide의 DataGuide를 통해서 추출  - 주가 자료를 통해 보고서의 정보력을 검증  ● 기업의 특성변수: FnGuide의 DataGuide를 통해서 추출  - 통제변수로 사용  - ex) 기업의 시가총액, 자기자본의 장부가치, 산업분류… |
| **변수** | ● 애널리스트 보고서에서 정보력을 가지고 있는 변수로 4가지 고려  1) 추천의견  ① 매도 = 1  ② 비중축소 = 2  ③ 중립 = 3  ④ 매수 = 4  ⑤ 적극매수 = 5  \*\* 추천 투자등급은 증권사마다 상이하지만, 이원홈, 최수미, 김동순, 엄승섭의 분류방식을 그대로 사용  \*\* 추천변경() = 추천의견값 – 이전보고서의 추천의견값  ex) 2 = 추천의견이 2단계 ↑, -2 = 추천의견이 2단계 ↓  2) 목표주가의 변화율 ( =  3) 이익예측치 (  4) 텍스트 분석을 통한 어조 변수(  - Tone = 긍정어 177개, 부정어 180개의 목록을 토대로, 애널리스트 보고서 제목의 (긍정어수 합계 – 부정어수 합계)  ● 누적초과수익률을 통해 애널리스트 보고서 정보에 대한 시장반응을 측정  - CAR(-2,2) = 발표일자를 포함한 5일 동안의 누적초과수익률(CAR, cumulative abnormal return)  - 초과수익률(AR, abnormal return) = 개별 기업의 수익률 – 당일의 시장수익률  ● 요약: 애널리스트 보고서의 정보변수 4개와 CAR(-2,2)의 관계를 통해 시장반응 검증 🡪 어떤 정보를 투자자들이 가장 유의미하게 인식하는지 파악 |
| **요약통계량** | ● 애널리스트의 추천의견이 낙관적인 쪽으로 치우쳐저 있음  - 매도 = 17건(0.02%), 비중축소 = 104건(0.14%), 매수 = 53,132(71.27%)  - 미국과 비교했을 때, 한국 애널리스트의 매도의견이 더욱 희소    - 추천의견등급↑ 🡺 목표주가 변화율 ↑, 이익예측치 변화율 ↑, 어조도 긍정적으로 ↑  - TONE 변수 = 애널리스트 보고서 제목의 (긍정어 수 합계 – 부정어 수 합계)  - TONE의 최솟값 = -3 : 긍정어는 없고 부정어만 3개 나오는 경우가 대부분  : 긍정어와 부정어 수가 함께 나오기 보다는 한쪽으로 쏠려있는 경향 有  - TONE 최댓값 = 5  - TONE의 분포: (TONE = 1)은 70.54%, (TONE = 0)은 17.67%, (TONE = -1) 은 7.54%  ● 변수들 간의 상관관계 |
| **실증분석**  **결과** | **1. 보고서 발표일 전후의 수익률**  \*\* AR(abnormal return, 초과수익률) = (개별 기업의 수익률 - 해당일의 시장  수익률)  ● 추천의견  통계적으로 유의X  - 추천의견 하향(DOWN\_GR) 발표일의 AR(0)은 평균 -0.847%, 추천의견  상향(UP\_GR) 발표일은 평균 0.379%의 초과수익률  - CAR(-6, -1)은 하향발표일 전에 -0.273%, 상향발표일 전에 0.772%로 수익률이 발표일 이전에 이미 하향하거나 상향하는 흐름  ● 목표주가  - 목표주가 하향(ΔTPRC(-)) 발표일은 평균 -0.184%, 상향(ΔTPRC(+)) 발표일은  평균 0.327%의 초과수익률 -> 둘 다 통계적으로 유의  - CAR(0, 5) = 목표주가 하향 후 6일 동안의 누적수익률은 -0.302%이며, 상향은 0.409%로 조금 더 큼 -> 목표주가 발표의 경우 하향의 영향력은 약하고 상향의 영향이 더 큼  ● 이익예측치  - 이익예측치 하향(ΔEPS(-)) 발표일은 평균 0.022%로 예측과 반대이지만 통계적  유의성은 X  - 상향(ΔEPS(+)) 발표일은 평균 0.245%의 초과수익률을 가지며 통계적으로 유의미  - 이익예측치 하향의 누적초과수익률을 살펴보면, CAR(-2, 2)에서도 양(+)인 초과수익률이 CAR(0, 5)에서야 음(-)의 수익률로 전환 VS. 상향의 누적초과수익률은 계속해서 양(+)의 값  - 전반적으로 EPS 예측치 발표가 목표주가 정보보다 영향력이 작으며, 상대적으로 상향의 영향력이 하향보다 더 큼  ● 보고서의 어조  - 보고서의 어조가 부정적인 발표일(NEG\_TONE)은 평균 -0.210%,  긍정적인 발표일(POSG\_TONE)은 평균 0.262%의 초과수익률  - 주변일인 전날 AR(-1)과 다음날 AR(1)도 발표일과 동일한 방향  - CAR(-6,-1) : 누적초과수익률은 부정적인 발표일 전에 -0.212%,  긍정적인 발표일 전에 0.701%로 수익률이 발표일 이전에 이미 하향하거나  상향하는 흐름  - CAR(0,5) : 부정적 보고서 이후 수익률이 하락하고,  긍정적 보고서 이후 수익률은 상승하며 통계적으로도 유의미  - CAR(0,10) : 하향의 경우는 5일이 지나면 하락세가 약해지지만, 상향의 경우는 5일 이후에도 상승세가 지속 --> 긍정적 발표일의 영향이 부정적 발표일보다 더 큼  ● 결론    ⅰ) 추천의견 변경의 경우 : 상향, 하향 다 의미있는 영향 & 하향의 발표일 효과가 큼  ⅱ) 목표주가, EPS예측치의 경우  : 상향의 효과가 하향의 효과보다 더 뚜렷  : EPS예측치의 경우는 하향의 효과가 거의 X  ⅲ) 어조 변수의 유용성을 발견  : 긍정적인 어조의 발표일 경우의 수익률의 추세는 추천의견 상향의 경우와 유사 : 부정적인 어조의 발표일은 목표주가 하향의 경우와 비슷  **2. 애널리스 보고서 정보의 주가예측력**    ● 종속변수: CAR(-2, 2) = 발표일 전후 2일을 포함한 5일의 누적초과수익률  ● 설명변수:  - ΔRECOMM은 추천변경 변수  - ΔTPRC은 목표주가의 변화율  - ΔEPS은 이익예측치 변화율  - TONE은 애널리스트 보고서의 어조   = ( 보고서 제목의 긍정어수 합계 - 부정어 수의 합계)  ● 통제변수  - Nanalyst는 해당 연도에 해당 기업에 대해 보고서를 발표한 애널리스트의 수  - 기업규모(Size)는 연말의 주가에 상장주식수를 곱한 값에 로그를 취한 값  - B/M은 자기자본의 장부가치를 연말의 시가총액으로 나눈 값  ● 회귀식에는 표기하지 않았지만, 산업효과와 연도효과를 통제하기 위해 산업 및 연도 더미변수를 사용  ● 결과표  <표4>    - 모형 (3)의 ΔEPS만이 통계적 유의성 X --> 회귀계수: 0.230, t-value: 1.319  - 모형 (5)에서 4개의 변수를 모두 사용하였을 때, ΔRECOMM와 ΔTPRC, TONE은 여전히 통계적으로 유의 VS. ΔEPS는 유의미 X  - 모형 (5)의 관측치는 23,445개=전체 표본의 1/2 <-- ΔEPS 관측 표본 작아서  - 모형 (6)에서는 ΔEPS 변수를 제외  : ΔRECOMM, ΔTPRC와 TONE 3개 변수 모두 통계적으로 유의미하며,  특히 TONE 변수의 유의성이 강하게 발견  변수들이 내포하는 정보의 상황에 따라 세분한 새로운 변수를 사용하여 분석    ● ΔRECOMM (추천변경)  - ΔRECOMM>0일 때 --> 상향(upgrade) UP\_GR = 1  ΔRECOMM<0일 때 --> UP\_GR = 0  - ΔRECOMM<0일 때 --> 하향(downgrade) DOWN\_GR = 1  나머지일 때 --> DOWN\_GR = 0  ● TONE  - TONE>0일 때 --> POS\_TONE = 1  나머지일 때 --> POS\_TONE = 0  - TONE<0일 때 --> NEG\_TONE= 1  나머지일 때 --> NEG\_TONE= 0  ● 통제 변수(애널리스트 수(Nanalyst)와 기업규모(Size), B/M) 그대로 사용  ● ΔEPS 변수는 제외  ● 결과표  <표5>    통계적으로 유의 X  - 모형 (1)에서 UP\_GR와 DOWN\_GR는 각각 회귀계수 0.594(t-값 = 3.18), -1.852 (t-값 = -10.88)로 모두 통계적으로 유의  - 추천변경 하향의 경우 애널리스트 보고서 발표로 수익률이 -1.85% 정도  하락하며, 상향의 경우는 0.59% 정도 상승  - 애널리스트 하향 추천의견이 상향 의견보다 더 큰 정보력을 가지고 있음  - 모형 (2)에서 통제변수들을 추가한 경우도 결과는 비슷  - 모형 (3)에서 POS\_TONE과 NEG\_TONE은 각각 회귀계수 0.137(t-값 = 1.50), -0.531 (t-값 = -4.61)  - TONE 변수를 긍정과 부정으로 구분하였을 때, 이 역시 부정의 경우 하락하는 수익률이 더 큼  - 모형 (5)는 추천변경, 어조 변수, 목표주가 변수를 사용  - 모형 (6)은 추천변경, 어조 변수, 목표주가 변수, 통제변수  - 모형 (5)와 (6)의 추천변경 상향(UP\_GR)과 긍정적 어조(POS\_TONE) 변수는  양(+)의 회귀계수를 가지기는 하지만 통계적 유의성 X  - 반면 추천변경 하향(DOWN\_GR)와 부정적 어조(NEG\_TONE) 변수는 음(-)의 회귀계수를 가지며 통계적으로도 유의미  - 애널리스트 보고서의 내용이 부정적일 때가 긍정적일 때보다 시장에 미치는 영향력이 큰 비대칭성이 존재함을 보여준다  **3. 텍스트 어조에 대한 추가분석**  ● 지금까지는 애널리스트 보고서의 텍스트 어조가 추천의견 변경이나 목표  주가를 통제한 후에도 나름대로의 정보력을 가지고 있음을 확인  ● 여기서는 텍스트 어조의 정보효과를 분석  1) 기업에 따라서 보고서 텍스트의 정보효과는 다를 수 있으므로 '기업특성변수'를 사용하여 애널리스트 보고서 텍스트 어조의 정보효과 차이 검증  - 기업특성 변수: 기업규모 (size = log(기업의 연말 주가 X 상장주식수)), 애널리스트 수 (기업에 대해 해당 연도 동안 보고서를 발행한 애널리스트의 수)  - 검증을 위해 TONE 변수와 기업특성변수와의 교차항 추가하여 회귀분석  <표6>    통계적으로 유의 X  - 모형 (1)에는 POS\_TONE×Size와 NEG\_TONE×Size 변수 추가  : 모형 (1)에서 POS\_TONE×Size의 회귀계수는 -0.100(t-값 = -1.86)로 통계적으로 유의미  : 긍정적 어조와 수익률 상승과의 관계가 대규모 기업에서 약해짐  : NEG\_TONE×Size의 회귀계수는 0.123(t-값 = 2.68)로 통계적으로 유의미  : 부정적 어조와 수익률 하락의 관계가 대규모 기업에서 더 약함  - 모형(1)을 통해 대규모 기업보다는 소규모 기업에서 애널리스트 보고서의 정보효과가 더 강하게 나타남을 보여줌  - 모형 (2)는 텍스트 어조 TONE 변수와 애널리스트 수(Nanalyst)와의 교차항을 추가한 결과  - NEG\_TONE×Nanalyst의 회귀계수는 0.034(t-값 = 4.27)로 부정적인 어조인 경우만 통계적으로 유의미  - 요약: 어조와 수익률 간의 관계가 규모가 작고 애널리스트 수가 적은 기업에서 더 강함 --> 애널리스트의 수가 정보생성과 연관되어 정보 비대칭의 측정치가 됨. 이런 측면에서 애널리스트 수 󰀶 -> 정보가치 󰀵  2) 증권사의 규모가 애널리스트 보고서 정보의 질에 영향을 줄 수 있으므로 연도별로 해당 증권사의 보고서를 발표한 대상기업의 수와 애널리스트 수 사용  - (연도별로 해당 증권사의 보고서를 발표한 대상기업의 수)가 상위 50%   --> N\_Firm = 1 , 아니면 N\_Firm = 0  - (연도별로 해당 증권사의 보고서를 발표한 애널리스트의 수)가 상위 50%   --> N\_Anal = 1 , 아니면 N\_Anal = 0  - TONE 변수와 N\_Firm와의 교차항, TONE 변수와 N\_Anal와의 교차항 추가하여 회귀분석  <표7>    통계적으로 유의 X  통계적으로 유의 X  통계적으로 유의 X  - 두 모형 모두 긍정적인 어조의 교차항이 통계적으로 유의미  - POS\_TONE의 회귀계수의 유의성 사라짐  - 긍정적 어조의 영향력의 대부분이 규모가 큰 증권사에서 나옴  3) 애널리스트의 특성에 따른 영향을 고려  - 애널리스트 특성을 반영한 더미변수(Dummy)와 TONE 변수의 교차항을 추가하여 회귀분석  <표8>  통계적으로 유의 X    통계적으로 유의 X  - 모형(1)은  애널리스트가 다루는 기업의 수가 하위 50% --> Dummy = 1  나머지 --> Dummy = 0  - 모형(2)은  표본기간 동안의 보고서를 제출한 경력 연수가 상위 50% --> Dummy = 1  나머지 --> Dummy = 0  - 모형(3)은  증권사를 이직한 경험이 있는 애널리스트 --> Dummy = 1  나머지 --> Dummy = 0  - 모형(4)은  베스트 애널리스트로 선정된 경험이 있으면 --> Dummy = 1  없으면 --> Dummy = 0  \*\* 베스트 애널리스트 여보는 한국금융투자협회와 한국거래소가 후원하고 조선일보와 에프앤가이드가 공동으로 선정하는 애널리스트 자료 사용  - 모형 (1)에서 POS\_TONE×Dummy의 회귀계수만 -0.255(t-값 = -1.769)로  통계적으로 유의미  - 애널리스트의 특성을 고려하는 것이 유의미한 정보력을 보여주지는 못한다  - 앞에서 증권사의 특성은 긍정적 어조에서 유의미한 역할을 한 것과 대조됨  ● 애널리스트들의 텍스트 정보생성에 있어 증권사 수준에서는 의미있는 영향력을 주지만, 애널리스트 개개인의 특성들까지 반영되기는 어려움  **4. 강건성 검증**  ● 수익률에 대한 강건성 검증  1) 누적초과수익률(CAR) 대신 보유초과수익률(BHAR, buy-and-hold abnormal return)을 사용  <표9\_패널 A>    - 보유초과수익률(BHAR)  = 해당 종목의 5일 동안의 보유수익률 - 시장수익률을 사용한 5일 동안의 보유수익률  - CAR를 사용한 결과와 크게 다르지 않음  2) 초과수익률을 구할 때 사용하는 벤치마크로 본문에서는 시장수익률을 사용하였지만, 그 대안으로 DGTW(Daniel, Grinblatt, Titman, and Wermers, 1997) 벤치마크 수익률을 사용  <표9\_패널 B>    - 전반적으로 시장수익률을 벤치마크로 구한 CAR를 사용한 분석 결과와 크게 다르지 않음  3) CAR의 기간을 변경  <표9\_패널 C>    - 발표일 이후 6일 동안의 누적초과수익률, CAR(0, 5)을 검증  - 패널 C의 결과는 본문의 결과와 비슷  ● 다음 강건성 검증은 다양한 하위표본을 사용하여 실증분석 결과를 재검토  1) 전체 표본기간을 하위기간으로 나눔  <표10\_패널A>    코스피 지수가 하락한 연도  통계적으로 유의 X  - 2009년부터 2018년까지 기간을 시장 상황에 따라 시장수익률이 양(+)인 연도와 음(-)인 연도로 구분  - 2011년, 2014년, 2018년 = 코스피 지수가 하락한 연도  나머지 = 모두 코스피 지수가 상승한 기간  - <표 5>의 모형 (4)= TONE(POS,NEG)변수와 통제변수를 시행한 결과  - 전반적으로 NEG\_TONE이 POS\_TONE보다 더 유의미한 결과  - NEG\_TONE의 회귀계수는 시장수익률이 음(-)인 기간이 양(+)인 기간보다 더 큼  - 시장상황이 좋은 기간에는 POS\_TONE의 영향이 커지고, 시장상황이 좋은 않은 기간에는 NEG\_TONE의 영향력이 더 커지는 경향 (?)  2) 유가증권시장과 코스닥시장으로 구분  <표10\_패널B>    통계적으로 유의 X  - POS\_TONE은 유가증권시장에서만 유의미  - NEG\_TONE은 두 시장 모두에서 유의미  - NEG\_TONE의 회귀계수는 유가증권시장이 -0.600, 코스닥시장이 -0.636으로  큰 차이 X  - 관측수에 있어서는 유가증권시장이 35,998개로 코스닥시장 8,454개  보다 3배 이상 많음  3) 기업규모에 따라 하위표본으로 나눔  <표10\_패널C>    - 유가증권시장과 코스닥시장으로 구분한 후에 그 안에서 기업규모의 평균이 상위 50% 주식 = BIG  하위 50% 주식 = SMALL  - POS\_TONE = 유가증권시장의 대형주(BIG)에서만 유의미  - NEG\_TONE = 모든 그룹에서 유의미한 값  - 소형주(SMALL)의 NEG\_TONE 회귀계수가 대형주(BIG)보다 큼 |
| **결론** | ● 본 연구는 애널리스트들이 제공하는 다양한 정보들의 상호관계와 정보력에 대한 종합적 분석이 주요 목적  ● 텍스트 정보가 주는 정보력도 추가하여 검증  - 애널리스트 보고서의 제목을 추출하여 해당 텍스트가 가지는 어조(tone) 변수를 구성  ● FnGuide가 제공하는 DataGuide에서 추출한 2009년부터 2018년까지 약  5만 개의 애널리스트 보고서 자료를 사용  ● 애널리스트 보고서의 정보 변수들을 추천의견, 목표주가, 이익예측치, 텍스트 분석을 통한 어조(tone) 4가지 범주로 분류, 이들 변수들과 애널리스트 보고서 발표일 이후 5일 간의 누적초과수익률과의 관계를 회귀분석을 통해 검증  ● 주요 결과  1) 애널리스트 보고서 변수들 중에서 추천의견 변경, 목표주가 변화율과 어조 등 3개 범주의 변수들에 시장이 유의미하게 반응하는 것을 발견, 이익예측치 변화율 변수에 대해서는 유의미 X  - 특히 텍스트 어조 변수는 다른 정보 변수들을 통제한 후에도 유의미한 정보력  - 어조 변수를 긍정과 부정으로 구분하였을 때 부정적 어조의 정보 효과가 강하였다  2) 텍스트 어조의 정보효과는 기업의 특성에 따라 차이  - 대규모 기업보다는 소규모 기업에서, 애널리스트 수가 많은 기업보다는 애널리스트 수가 적은 기업에서 어조의 정보 효과가 더 강함  3) 텍스트 어조의 정보력을 증권사 및 애널리스트의 특성에 따라 구분  - 규모가 큰 증권사가 긍정적 어조에서 유의미한 역할을 하여 증권사의  특성은 유의미  - 애널리스트의 특성을 고려하는 것이 유의미한 정보력 X |

**질문**

● '2. 애널리스 보고서 정보의 주가예측력' 부분에서 TONE 변수의 유의성이 강하게 발견되었다는 것의 근거가 명확히 이해되지 않음

● AR = 수익률 = 개별 기업의 수익률 - 시장수익률

여기서, 시장수익률은 사업분야별(ex. 은행, 전기전자, ...) 시장주익률을 의미하는 것인지?

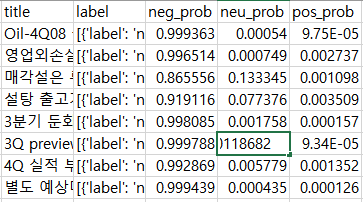
● 통계적으로 유의하다고 결정하는 t-value값의 기준은 무엇일까?

● TONE 변수를 (긍정적 단어의 개수 - 부정적 단어의 개수)로 하고,

- TONE>0일 때 --> POS\_TONE = 1, 나머지일 때 --> POS\_TONE = 0

- TONE<0일 때 --> NEG\_TONE= 1, 나머지일 때 --> NEG\_TONE= 0

- TONE 변수를 생성하는 더 좋은 방법은 없을까?   
 예를 들면, 아래와 같은 TONE의 확률값을 어떻게 이용하면 좋을지??



● 선형회귀분석 말고 비선형 머신러닝 방법을 도입해야하지 않을까?

● 한 단어만 있는 경우는 삭제해야 할까?



