Empirical Asset Pricing

Francesco Franzoni

Swiss Finance Institute - USI Lugano

The Efficient Market Hypothesis

Lecture Outline

1. Definition of Market Efficiency
2. Efficient Capital Markets II (Fama, JF, 1991)
3. Efficiently Inefficient (Berk and Green, JPE, 2004; Garleanu and Pedersen, JF, 2018)
4. Predicting Returns with Text Data (Ke, Kelly, and Xiu, 2019)
5. Market efficiency and learning (Martin and Nagel, 2019)

Relevant readings:

- Campbell, Lo, and MacKinlay: chapter 1
- Fama, E. F., 1991, "Efficient Capital Markets II", Journal of Finance
- Garleanu and Pedersen, 2018, "Efficiently Inefficient Markets for Assets and Asset Management", Journal of Finance
- Ke, Kelly, and Xiu, 2019, "Predicting Returns with Text Data", Working Paper
- Martin and Nagel, "Market Efficiency in the Age of Big Data", NBER Working Paper 26586

1. Market Efficiency

Historical Perspective

- Origins of the Efficient Market Hypothesis (EMH) date back to Bachelier (1900): efficiency of Paris Bourse

- Samuelson (1965): in an informationally efficient market price changes should be unpredictable

- Fama (1970): a market is efficient if prices fully reflect all available information

- Malkiel (1992): "...A capital market is said to be efficient if it *fully and correctly reflects all relevant information* in determining security prices. Formally, the market is said to be efficient with respect to some information set...if *security prices would be unaffected by revealing that information to all partecipants*. Moreover, efficiency with respect to an information set ...implies that it is *impossible to make economic profits* by trading on the basis of [that information]..."

Three Ideas

1. Prices fully reflect all the available information: suggestive but empirically useless

2. Prices would not move if the available info was revealed to market participants (because they have already exploited it): thought experiment, still useless

3. Abnormal profits should not be possibly made by trading on the information: truly operational concept

Strategies for testing efficiency

- The third idea suggested the empirical strategies for testing market efficiency

1. Look at profits generated by professional market participants. If they achieve superior returns (after adjusting for risk) then markets are not efficient. (Ex: mutual fund managers)
   - Problem: you do not observe information they use
   - They may be missing some relevant information
   - Conclude for efficiency, but really is lack of skill
   - Or, based on Berk and Green's (2004) theory, there is skill, but decreasing returns to scale make abnormal profits equal to zero
2. Look at hypothetical trading strategies based on explicitly specified information set. Do they earn superior returns?
   - One needs to specify the information set
   - Also, need to specify model for risk
   - Trading costs?

Taxonomy of Information Sets

- To implement the second approach, one needs to specify the information set
- Different info sets imply different forms of efficiency that can be classified as follows (Roberts, 1967) :

1. **Weak-form Efficiency**: The information set includes only the history of the prices or returns themselves
2. **Semistrong-form Efficiency**: The information set includes all information known to all market participants (publicly available information)
3. **Strong-form Efficiency**: The information set includes all information known to any market participant (private information)

Impossibility of Perfect Efficiency

- In a fundamental work, Grossman and Stiglitz (1980) argue that, if there are costs of gathering information, the market cannot be perfectly efficient

- Consider an investment firm. It has to pay millions to set up a research division

- It would not do it, if it could not make profits from trading on its research

- Then, there must be some profit from collecting information and trading on it: some degree of inefficiency

- Otherwise, nobody would pay the cost of collecting the information and the market would break down

Abnormal Returns

- A crucial point in the tests to define what superior, or abnormal, returns are
- We can define them as follows:

$$\text{Abnormal Return} = \text{Realized Return-Normal Return}$$

- What about Normal Returns?
- They are the reward for the risk of the investment
- Need to specify a model for Normal, or Expected, Returns (e.g. CAPM, APT, etc.)
- Then, abnormal returns can be obtained as

$$AR_{t+1} = R_{t+1} - E^M(R_{t+1})$$

- The superscript M denotes the fact that the expected return depends on the model for risk

- The EMH can be expressed as

$$H_0 : E\left(AR_{t+1} | I_t\right) = 0$$

- If the abnormal return is predictable using the info in I_t, then the hypothesis of market efficiency is rejected

Innovations in Finance

- It was typical in the past to assume constant normal returns for an asset
- For example: the Random-Walk Hypothesis

$$RW : p_{t+1} = k + p_t + \varepsilon_{t+1}$$

  where k is a constant reward for risk

- On daily data $k \approx 0$ and the assumption does not harm
- On longer horizons, however, risk premia can vary in a predictable way. E.g.: require high expected return in recession, low in expansion
- Hence, recent equilibrium models allow for time-varying expected (=normal) returns

Joint-Hypothesis Problem (JHP)

- Fama (1971)
- Indeed, a test of EMH contains a joint hypothesis:
  1. That markets are efficient
  2. That you are choosing the right model for risk
- This implies that market efficiency can never be rejected (but see later qualifications...)
- The whole debate between: 'Rational Finance' and 'Behavioral Finance' can be framed in terms of the JHP
- Also, perfect efficiency is unrealistic given frictions in the market, and costs of gathering information
- The question is whether deviations from EMH exceed reasonable transaction costs

2. Efficient Capital Markets?

Background

- In 1970, Fama, Efficient Capital Markets I: EMH holds unambiguosly
- In 1991: first negative results for CAPM and return predictability by Fama and French
- Need to rephrase initial (hardcore 'rational') position:
  - Not true rejection of EMH, but expressions of JHP
- Likely that we need to rethink models for risk
- JHP makes empirical conclusions ambiguous but not irrelevant: we have learnt a lot about properties of asset prices

New Taxonomy of Tests of EMH

1. Tests for Predictability:
   (a) Time-Series
   (b) Cross-Sectional
   - Cross-sectional tests are one of the main topics of this class
   - Here the JHP is very strong
2. Event Studies ($=$ tests of semi-strong efficiency)
   - Compute returns after the release of public information and see if returns are different from zero after event
   - Because of short window, computing risk adjustment is not crucial
   - Hence, this is as close as you can get to a pure test of EMH
3. Tests for Private Information ($=$ tests of strong efficiency)

(a) Tests of mutual/hedge fund performance

(b) Tests for insider trading

- Here, you have JHP
- But you would not expect to find efficiency because of Grossman and Stiglitz

Quick Review of Results

1. Tests of Predictability

a) Times-Series Predictability

- Want to predict returns over time, typically on an index

$$E\left(AR_{t+1}|I_t\right) = 0 \quad t = 1...T$$

where I_t is either past returns or other public information

Results:

- At short horizons (daily, weekly, monthly):
  - Almost no predictability: RW hypothesis works for the market
  - Some predictability for small stocks, due to infrequent trading (positive cross-correlation, Lo and MacKinley, 1990, RFS) and bid-ask bounce (negative autocorrelation)
  - It would not survive after reasonable transaction costs
  - Cannot be considered evidence of irrational pricing (either overreaction or underreaction)

- At long horizons (multi-year):
  - Negative autocorrelations over 3-5 years (Fama and French, 1988, JPE)

$$\begin{aligned} R_{t+3} &= a + bR_t + \varepsilon_{t+3} \\ \hat{b} &\simeq -0.3 \text{ autocorr.} \end{aligned}$$

  - Two possible explanations:

    1. Irrational Pricing: mean-reverting sentiment (Summers, 1986, JF)

$$p_t = p^* + s_t$$

where p^* is fundamental value and s_t is mean-reverting sentiment

* Notice: this explanation can accomodate lack of short horizon predictability if temporary component of prices (sentiment) moves slowly

$$s_t = \rho s_{t-1} + u_t$$

Let ρ be close to one

* On short horizon, $p_t \simeq p_{t-1}$ is almost constant
* On long horizon, $E_t\left(p_{t+k}\right) = p^* + \rho^k s_t$, $\rho^k \simeq 0$, you have mean reversion

2. Rational Pricing: time-varying expected returns because of time-varying risk aversion

  * E.g.: positive shock to expected returns $\longrightarrow$ the price drops (negative returns today) $\longrightarrow$ higher returns in the future (because higher exp. return): negative autocorrelation

  * Also this explanation is consistent with lack of short horizon predictability if exp. ret. is slowly mean-reverting

– Long-horizon returns on index are predictable using valuation ratios: D/P, E/P, etc.

  * By Gordon's Dividend Discount Model:

$$P = \frac{D}{R - G}$$

  So, there is positive relationship between R and D/P: risk based explanation
  Of course, you need expected returns to be time-varying, otherwise D/P would be constant

  * But sentiment story could apply instead: optimism $\longrightarrow P \uparrow \longrightarrow$ D/P$\downarrow$, but then bubble bursts and $R < 0$: positive correlation of D/P and returns

– Clear manifestations of the JHP

b) Cross-Sectional Predictability

- In this case you look at

$$E\left(AR^i_{t+1}|I_t\right) = 0 \quad i = 1...N$$

- That is: are there stock characteristics (size, B/M, past returns, volatility, profitability, etc.) that allow us to predict different returns on different assets?

- Suppose you use CAPM for $E^M\left(R_{t+1}\right)$

$$E^M\left(R^i_{t+1}\right) = R_f + \beta_i\left(E\left(R^m_{t+1}\right) - R_f\right)$$

- Then, in the time-series regression

$$R^i_{t+1} = \alpha_i + R_f + \beta_i\left(R^m_{t+1} - R_f\right) + \varepsilon^i_{t+1}$$

  the EMH amounts to

$$H_0 : \alpha_i = 0 \quad i = 1...N$$

  which is indistinguishable from a test of CAPM (see later)

- Joint Hypothesis Problem

The Bid-Ask Bounce (CLM, 3.2.1)

- Roll's (1984) Model
- It accounts for the impact of bid-ask spread on time-series properties of returns
- Relevant at short frequencies (daily or shorter)
- Let P_t^* be the fundamental value of the security and P_t be the observed market price

$$P_t = P_t^* + I_t \frac{s}{2}$$

where s is the bid-ask spread and I_t is an indicator variable denoting whether the transaction took place at the ask ($I_t = 1$) or at the bid ($I_t = -1$)

$$I_t \quad IID \begin{cases} +1 \text{ with probability } 1/2 \\ -1 \text{ with probabilty } 1/2 \end{cases}$$

$E\left(I_t\right) = 0$ and $Var\left(I_t\right) = 1$

- Assume that there are no changes in the fundamental value of the security: $\Delta P_t^* = 0$

- Then, the process for price changes (returns) becomes

$$\Delta P_t = (I_t - I_{t-1})\,\frac{s}{2}$$

- Under the assumption of IID I_t we can compute

$$\begin{aligned}
Var\,(\Delta P_t) &= \frac{s^2}{2} \\
Cov\,(\Delta P_{t-1}, \Delta P_t) &= -\frac{s^2}{4} \\
Cov\,(\Delta P_{t-k}, \Delta P_t) &= 0, \quad k > 1 \\
Corr\,(\Delta P_{t-1}, \Delta P_t) &= -\frac{1}{2}
\end{aligned}$$

- Despite the fact that fundamental value is fixed, returns exhibit volatility and negative serial correlation, as a result of the *bid-ask bounce*

- Intuition: returns are either zero or the opposite of the return in the prior period

2. Event Studies

- Fama, Fisher, Jensen, and Roll, 1969, International Economic Review
- Event studies look at average returns after release of public information for a few days (short-run event studies) or for up to a year (long-run event studies)
- For short-run studies, given average daily market return is 0.04%, the risk adjustment does not matter much
- Short-run studies are the cleanest test of market efficiency
- Long-run event studies suffer from JHP and statistical problems (cross-sectional correlation of return)
- Results at the time of Fama's paper: information impounded at the time of release
- Later results: *Post-Earnings Announcement Drift* (Bernard and Thomas, 1990)
  - Positive earnings surprises trigger positive price drift and vice versa

- Underreaction to information
- In recent times, PEAD has decreased in magnitude (increased efficiency)
- Difficult to give rational explanation

3. Tests for Private Information

a) Insider Trading

- Prices start to rise a few days before positive announcement and vice versa
- But not all the way to the post-announcement level
- Profits from insider trading
- Evidence against strong efficiency

b) Professional Portfolio Managers

- Question: do they generate abnormal returns using the information they gather?
- Approach: test $H_0 : \alpha_i = 0$ in

$$\begin{aligned} E\left(R^i_{t+1}\right) - R_f &= \alpha_i + \beta_i^1 \left(E\left(R^m_{t+1}\right) - R_f\right) \\ &\quad + \beta_i^2 F_2 + \beta_i^3 F_3 + \ldots \end{aligned}$$

  where R^i_{t+1} is return on the fund

- If $\alpha > 0$ the manager has skill and reject strong efficiency
- JHP: what is the right model for risk?

Results:

- Jensen (1968): average return net of fees is 1% below benchmark
  - Adding back the fees $\simeq 0$
  - Concludes: no private information or skill
- Ippolito (1989): $+0.83\%$ above benchmark

- But this evidence disappears with multifactor model (adding size and B/M)
- Carhart (1997): zero outperformance when accounting for momentum returns
- No skill or private information: just exploiting existing anomalies (public information)
- Results on mutual fund lack of outperformance spurred the passive mutual fund industry
- More recent results: some persistence in outperformance exists for some ‘star’ managers (Kosowski, Timmerman, Wermers, and White, JF, 2006)
- Other recent results suggest that some fund managers appear to have skill:
  - Kacperczyk, Sialm, and Zheng (2005): industry concentration
  - Kacperczyk, Sialm, and Zheng (2006): return gap
  - Cremers and Petajisto (2009): active share
  - Amihud and Goyenko (2013): R-squared
  - Puckett and Yan (2011): interim trading skill using transaction data

3. Efficiently Inefficient

Efficient Market for Asset Managers

- Berk and Green (2004, JPE) build a neoclassical model for the choice of active managers by investors
- The model features investors learning about manager skill from past performance and decreasing returns to fund-scale
- Manager's skill in scarce supply, while investors are in large supply
- Hence, asset managers have monopoly power over the fees that they set
- In equilibrium, managers generate before-fee alphas
- But alphas are zero after fees as managers extract all the rents
- Other results: due to learning about manager skill, flows rationally chase past performance
- But flows are not predictive of future performance exactly because alphas are zero after fees in expectation
- In this model, there is *efficient allocation to asset managers*, in the sense that performance is equalized across managers after fees

Friction in the Search for Asset Managers

- Garleanu and Pedersen (2018, JF) overlay a model for the choice of asset managers by investors to a Grossman and Stiglitz framework

- The model features a search cost for investors (think about due diligence), besides the cost for gathering information about assets

- As a result of the search cost, investors must be indifferent between investing with an informed asset manager and investing in an uninformed way (i.e. via a passive fund)

- Therefore, in equilibrium, the informed managers' after-fee performance is positive (and covers the search cost of the marginal investor)

- This model features an 'efficient level of inefficiency'

- As in Grossman and Stiglitz, assset markets cannot be perfectly efficient, otherwise there would be no active manager/investors

- Plus, unlike Berk and Green, you get that some asset managers outperform after fees
- This model makes predictions that are consistent with a number of stylized facts:
  1. Fama's prediction (Market Efficiency) is that managers underperform by the amount of the fees. Instead, Garleanu and Pedersen's prediction is that some (skilled) managers generate outperformance after fee. The recent literature cited above confirms that the best managers outperform consistently (e.g. Kosowski et al. 2006). This evidence also contradicts the Berk and Green prediction
  2. Investors that are more likely to bear the search cost (i.e. do the due diligence) are more likely to invest in outperforming managers. Consistent with evidence in Evans and Falenbrach (2012) that mutual funds with institutional share classes outperform other mutual funds
  3. Related, mutual funds that only service institutional investors outperform mutual funds that service retail clients who are less likely to engage in a search (Gerakos, Linnainmaa, Morse 2016)
  4. Large investors have better performance than smaller investors (Gerakos, Linnainmaa, Morse 2016), consistent with a fixed search cost and the better ability to bear this cost by larger investors

5. Anomalies are more likely to arise in securities that are covered by managers for which search costs are higher (e.g. private equity, convertible bonds, etc.)
6. Anomalies are larger in markets that are more costly to study (e.g. equity more than bonds)
7. Fees are higher for assets whose managers are more costly to search (e.g. hedge funds vs. mutual funds), because there is less entry and more mispricing
8. Fees are higher for managers investing in more mispriced assets, because investors obtain higher returns. Note that some friction (high costs of either information or search) must prevent the entry of new managers and the mispricing from disappearing

4. Predicting Returns with Text and Machine Learning

Predicting Returns with Text Data (Ke, Kelly, Xiu, 2019)

- Test whether returns can be predicted with sentiment from news articles
- If yes: Violation of semi-strong form of EMH
- Possible channels
  - Limits to arbitrage (see later class)
  - Rationally limited attention
- Develop a text mining approach to predict returns based on *supervised learning*
- Different from previous text-based approaches, because it is specifically targeted to return prediction
- Three steps:
  1. Isolate a list of sentiment words

2. Assign sentiment weights to these words

3. Aggregate terms into article-level score used within a trading strategy

- Multiple advantages:
  - Specifically adapted to the context at hand; it does not rely on pre-existing dictionaries
  - Very transparent supervised learning approach: minimum computing power and white box
  - Provides properties of estimators under mild assumptions (we do not focus on this part here, but it is a nice feature)

Previous approaches

- Tetlock (2007) applies the Harvard-IV psychosocial dictionary to articles from the WSJ to predict sentiment in index returns

- Loughran and McDonald (2011) create a new dictionary specifically designed for finance. They use it to classify 10Ks and other financial communications and show that the sentiment score correlates with returns

- These papers do not carry out a supervised selection of sentiment words

- Jegadeesh and Wu (2013) is precursor that also does supervised estimation of sentiment words

1. Screen sentiment words

- Supervised learning
- Use a **training sample** to screen for sentiment-charged words
- Attach to each article y a 'label', that is, the sign of the associated stock return
- For each word, compute the frequency at which it appears with positive returns

$$f_j = \frac{\#\text{articles including word } j \text{ AND having } sign(y) = 1}{\#\text{ articles including word } j} \tag{1}$$

- Set thresholds α_+, α_- for frequencies above/below which words are deemed positive/negative
- Also, set minimum threshold κ for the count of articles in which word j appears for statistical reliability (the denominator in equation (1))

2. Learn sentiment topics

- Again, supervised learning
- S is the list of sentiment-charged words
- The *topics* are two probability distributions $O = [O_+, O_-]$ on the list S that give the probability of each word in the maximally positive and negative topic, respectively
- A word j is positive if the j^{th} entry in $O_+ - O_-$ is positive, and vice versa
- Let $\tilde{d}_{i,[S]}$ denote the vector of sentiment-charged word frequencies for article i
- Then, the statistical model is assumed to be

$$E\left(\tilde{d}_{i,[S]}\right) = p_i O_+ + (1 - p_i) O_- \tag{2}$$

  where p_i is the **sentiment score** of article i
- How to estimate O?

- Following equation (2), from a regression of $\tilde{d}_{i,[S]}$ on p_i. Neither variable is observed. We need proxies for them

- From the first step, we obtain $\hat{S}$, which is an estimate of S

- Again within the training sample, we obtain $\hat{p}_i$

$$\hat{p}_i = \frac{\text{rank of } y_i \text{ in } \{y_l\}_{l=1}^{n}}{n}$$

  i.e., the standardized return rank of article i

- Finally regress $\tilde{d}_{i,[\hat{S}]}$ on $\hat{p}_i$ to obtain $\hat{O}$. Note that these are indeed $\left|\hat{S}\right|$ (numerosity of $\hat{S}$) regressions with n observations (number of articles)

3. Scoring news articles

- The preceeding steps estimate $\hat{S}$ and $\hat{O}$. Next, they need to produce a sentiment score p_i for the articles that are **not** in the training sample

- They estimate p_i by maximum likelihood, based on the model in equation (2)
  - For each article, estimate p_i using $\left|\hat{S}\right|$ observations on $\tilde{d}_{i,[\hat{S}]}$ and $\hat{O}$

- They add a penalty to the likelihood function to reduce the noise coming from a limited number of observations: $\lambda \log \left(p_i \left(1 - p_i\right)\right)$, which shrinks the estimate of p_i towards 1/2

Implementation

- Data: *Dow Jones Newswires Machine Text Feed and Archive.* 1989-2017. +10M articles involving only one firm
- 15 year rolling window:
  - first 10 years: training sample
  - next 5 years: validation sample
  - out-of-sample prediction: next one year
- Details
  - Estimate models in training sample corresponding to grid of parameters
  - Choose parameter constellation in validation sample minimizing loss function

$$\ell^1 - \text{norm} = \sum_{i=1}^{n} \left| p_i - \frac{\text{rank of } y_i \text{ in } \{y_l\}_{l=1}^{n}}{n} \right|$$

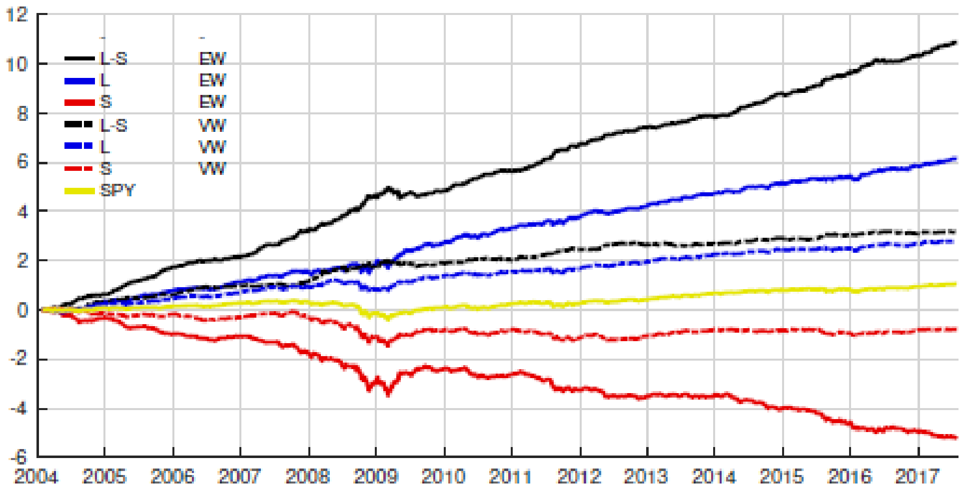
    for all articles n in the validation sample

– Using the chosen parameters, obtain p_i for each article in out-of-sample period

– Repeat 14 times

Trading Strategy

- Go long/short in 50 articles with highest/lowest p_i
- Open position at market open, close at market close



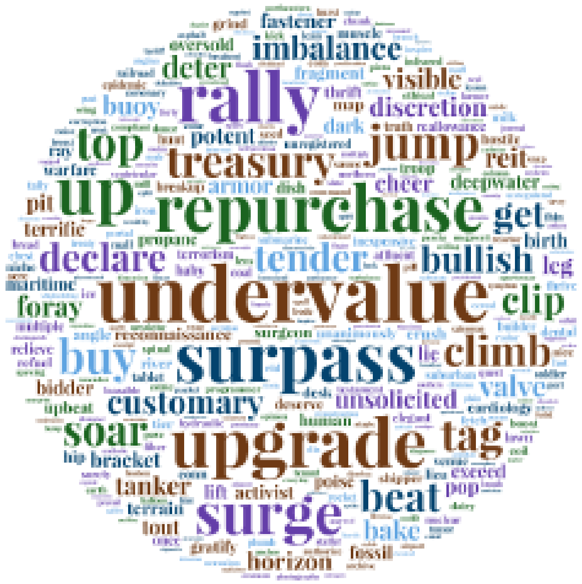
Note: This figure compares the out-of-sample cumulative log returns of portfolios sorted on sentiment scores. The black, blue, and red colors represent the long-short (L-S), long (L), and short (S) portfolios, respectively. The solid and dashed lines represent equal-weighted (EW) and value-weighted (VW) portfolios, respectively. The yellow solid line is the S&P 500 return (SPY).

- Sharpe Ratio of EW L-S is 4.29 (annualized); SR of VW L-S = 1.33
- Turnover 94% daily, suggesting some stocks are held for longer than a day

Most Impactful Words

Negative Words

Positive Words



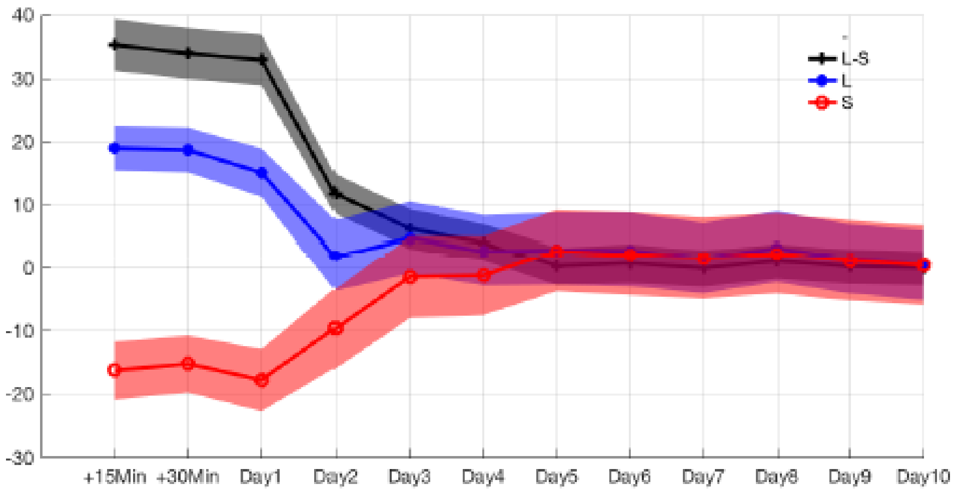
Lead-Lag Relations

- Trade on day -1 relative to news. SR = 5.88
- Meaning: news are anticipated, either because of previous news, or private information, or reverse causality (i.e. a large return is realized and the article comes out about it on the next day)

Formation	Sharpe Ratio	Turnover	Average Return	FF3 α	FF3 R^2	FF5 α	FF5 R^2	FF5+MOM α	FF5+MOM R^2
				Day −1					
L-S	5.88	94.5%	45	45	0.1%	44	0.5%	44	0.6%
L	2.30	95.9%	20	20	0.8%	21	1.1%	21	1.1%
S	2.08	93.2%	25	24	0.5%	24	1.2%	24	1.2%
				Day 0					
L-S	10.78	94.6%	93	93	0.4%	93	0.5%	92	0.8%
L	5.34	96.0%	50	48	7.0%	49	7.8%	49	8.1%
S	3.56	93.3%	43	45	6.0%	44	7.0%	43	7.5%
				Day +1					
L-S	4.29	94.6%	33	33	1.8%	32	3.0%	32	4.3%
L	2.12	95.8%	19	16	40.0%	16	40.3%	17	41.1%
S	1.21	93.4%	14	17	33.2%	16	34.2%	16	36.3%
				Day −1 to +1					
L-S	12.38	94.6%	170	170	1.0%	169	2.3%	169	2.8%
L	5.67	95.9%	89	86	22.3%	86	23.2%	87	24.1%
S	3.83	93.3%	81	85	16.7%	82	18.7%	82	20.1%

Note: The table repeats the analysis of Table 2 for the equal-weighted long-short (L-S) portfolios plotted in Figure 7, as well as their long (L) and short (S) legs. Sharpe ratios are annualized, while returns and alphas are in basis points per day.

- After 5 days news contents become useless



Note: This figure compares average one-day holding period returns to the news sentiment trading strategy as a function of when the trade is initiated. We consider intra-day high frequency trading that takes place either 15 or 30 minutes after the article's time stamp and is held for one day (denoted +15min and +30min, respectively), and daily open-to-open returns initiated from one to 10 days following the announcement. We report equal-weighted portfolio average returns (in basis points per day) in excess of an equal-weighted version of the S&P 500 index, with 95% confidence intervals given by the shaded regions. We consider the long-short (L-S) portfolio as well as the long (L) and short (S) legs separately.

5. Market Efficiency and Learning

In-sample vs. out-of-sample predictability

- Investors may need to learn about the asset pricing model parameters over time from the observed data

- This fact can generate in-sample predictability, but no out-of-sample predictability

- This situation is more likely when the number of predictors J is large relative to the number of assets N (or the number of periods T)

- Because it takes more data to really understand the relevance of a predicting variable and the learning process is slower

Interpretation of Market Efficiency

- Let us leave aside the Joint Hypothesis Problem. Consider a market with risk neutral investors
- What is the implication of the sentence "In an efficient market, prices fully reflect all the available information"?
- It depends on the model

1. Rational Expectations model: investors know all the relevant parameters of the cash-flow prediction model
   - Implication: Returns are not predictable both in-sample and out-of-sample
   - Logic: Rational investors use all the available information in the correct model. The econometrician does not know more than the investors
   - Out-of-sample testing is not recommended because of lower power of tests, given that fewer observations are used for prediction (Cochrane 2008; Campbell and Thompson 2008)

2. Learning model: Bayesian investors learn about the relevant parameters using information up to time t

   - Implication: Returns are predictable in sample, but not out-of-sample

   - Logic: Investors use information up to time t, whereas the econometrician uses information up to time $T > t$. The econometrician knows more

   - But this does not mean that investors were not doing the best they could given the available information

   - Importance of out-of-sample tests, not just because of data mining and p-hacking, but also because of false predictability resulting from learning

   - I.e., the econometrician should try to predict $t+1$ returns using information up to time t

A simple example

- Based on Lewellen and Shanken (2002)
- A stock pays dividend according to a process with unknown mean $\bar{d}$

$$d_t = \bar{d} + \varepsilon_t$$

- Investors need to estimate $\bar{d}$ and have a diffuse prior (they basically know nothing)
- In this case, the expecation of $\bar{d}$ is the mean of observed dividends: $\bar{d}_t = \sum_{i=1}^{t} \frac{d_i}{t}$
- Hence, when d_t is large investors update upwards their estimate of $\bar{d}$ and the price of the stock p_t rises
- However, a large d_t could just be due to noise (i.e. a large realization of ε_t)
- In this case, next periods' prices will have to be corrected downwards

- In this economy, an econometrician regressing returns on dividend yields finds a negative coefficient

- Alternatively, investors may have a very strong prior on low dividends

- In this case, a large realization of d_t due to a large $\bar{d}$ is accompanied by investors' underreaction

- Next period posteriors and prices will slowly adjust upwards

- In this economy, an econometrician observes returns continuation (i.e. time-series momentum)

- Eventually, investors learn about $\bar{d}$ and predictability should disappear

Martin and Nagel (2019)

- With the advent of big data, there are many variables X that can help predict cash flows
- Potentially, the number J of these variables is larger than the number of stocks N
- When these new variables emerge, investors do not have enough information (i.e. stocks) to estimate with large precision the cross-sectional implications of these predictors
- Investors need time to learn about the predictive ability of these variables for the cash-flow process
- Thus, in-sample, cross-sectional predictability can emerge
- However, out-of-sample predictability is not present

Conclusion

- In-sample evidence of predictability could be due to

  1. Market inefficiency resulting from irrationality or limited rationality (behavioral finance) and limits to arbitrage
  2. Wrong model for risk (rational finance à la Fama and French)
  3. Parameter uncertainty and bayesian learning

- Importance of out-of-sample testing to establish 3 vs. 1 and 2