Time Series Predictability and Stock Returns

Svetlana Bryzgalova

Pasargad Advanced Research Initiative

Big Data Economics

Tehran, Iran

September 5, 2016

A bit of history

What is predictability and why it matters?

A stock return r_{t+1} is predictable by a factor f_t if

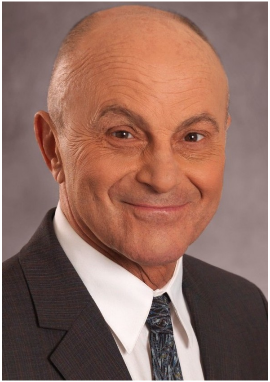
$$\mathbb{E}[r_{t+1}|I_t \backslash f_t] \neq \mathbb{E}[r_{t+1}|I_t]$$

.

Predictability of stock market returns is important for both researchers and practitioners:

- optimal asset allocation, portfolio choice
- welfare implications
- understanding the source of risks in the economy
- developing suitable structural models

Efficient Market Hypothesis



Eugene Fama
Nobel Prize in Economics, 2013

Weak form of market efficiency states that market prices already reflect all past publicly available information, hence it should not **systematically** bring additional **riskless/risk-adjusted** profits.

Up until 80s any sort of asset prices predictability (both fundamental and technical models) above transaction costs level was considered a violation of the EMH.

Basic test for market efficiency:

$$r_{t+1} = \alpha_0 + \rho f_t + \epsilon_{t+1}$$

$H_0: \ \rho = 0$

EMH strikes back

A simple test for predictability

Regression of returns on lagged returns (annual data, 1927-2012): $r_{t+1} = a + br_t + \epsilon_{t+1}$

Asset class	b	$t(b)$	R^2
Stock	0.04	0.33	0.002
T-bill	0.91	19.5	0.83
Excess stock return	0.04	0.39	0.00

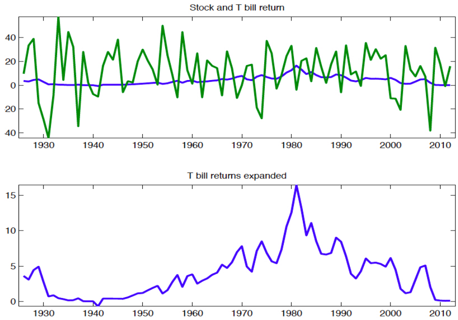


Figure 1 : Historical returns on stocks and T-bills. Source.

Market predictability

Let's use price-dividend ratio as a predictor:

Horizon k	b	$t(b)$	R^2
1 year	3.8	2.6	0.09
5 years	20.6	3.4	0.28

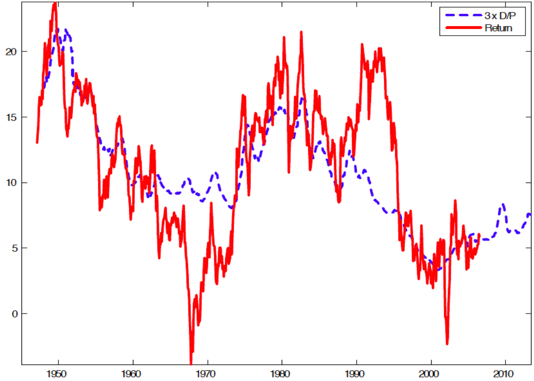


Figure 2 : 7year market returns and price-dividend ratio. Source.

Markets are predictable

Robert Shiller
Nobel Prize in Economics, 2013

Long-term returns are predictable by fundamental characteristics.

Shiller (1984): low price-dividend ratio today leads to future high returns tomorrow.

Ample empirical evidence suggesting other fundamental ratios also predict returns.

Overall advocates behavioural models in asset pricing that can explain bubbles, predictability, over-reaction, extrapolation of past performance, herding...

Occam's razor

- Let's try to use a single approach to thinking about decision-making in many different contents: a simple assumption can get you a long way forward!
- 'Models should be as simple as possible, but not simpler': overreaction and underreaction work differently depending on the context, application etc.
- Key question: whether a particular violation of rationality has an aggregate implication? Does the beauty contest **cause** market inefficiency on aggregate?
- So far no direct empirical link from micro inefficiencies to macro observations.
- Ample empirical evidence that when all the major **risks** are considered, in many applications economy behaves **as if** people were rational.
- Regardless of the rules, plenty decision-making is done by sophisticated decision-makers that faced with strict regulations, will try to go around them and adapt (Lucas critique)
- Any policy should be internally consistent.
- ABM and traditional economics approach are very similar: they both assume exogenous shocks and a certain structure of the economy, the only difference is the **propagation** mechanism for the shocks.

Understanding the null hypothesis

Early evidence for predictability of the market returns was interpreted as the violation of EMH.

$$r_{t+1} = \alpha_0 + \rho f_t + \epsilon_{t+1}$$

Inference on ρ, however, is done under a joint hypothesis:

1. market efficiency
2. linearity of the predictive model
3. constant expected returns
4. additional technical assumptions

Understanding the null hypothesis

Early evidence for predictability of the market returns was interpreted as the violation of EMH.

$$r_{t+1} = \alpha_0 + \rho f_t + \epsilon_{t+1}$$

Inference on ρ, however, is done under a joint hypothesis:

1. market efficiency
2. linearity of the predictive model
3. constant expected returns
4. additional technical assumptions

It is now wildly accepted that **expected returns are time-varying**:

- time-varying risk aversion
- long-run consumption risk
- time-varying opportunities for risk sharing (e.g. the impact of the housing collateral)

A simple model for stock returns

Let's start with the definition for gross stock returns:

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t}$$

Log-linearizing this identity:

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} = \frac{\left(1 + \frac{P_{t+1}}{D_{t+1}}\right) \frac{D_{t+1}}{D_t}}{\frac{P_t}{D_t}}$$

$$r_{t+1} = \log\left(1 + e^{pd_{t+1}}\right) + \Delta d_{t+1} - pd_t$$

$$r_{t+1} \approx \log\left(1 + e^{pd}\right) + \frac{e^{pd}}{1 + e^{pd}} \left(pd_{t+1} - pd\right) + \Delta d_{t+1} - pd_t$$

$$r_{t+1} \approx \rho(p_{t+1} - d_{t+1}) + \Delta d_{t+1} - (p_t - d_t)$$

where $\rho = \frac{1}{1+\frac{D}{P}} \approx 0.96$ (for annual data, $P/D \approx 20$).

Present value relationship

Start with the return identity:

$$r_{t+1} \approx \rho(p_{t+1} - d_{t+1}) + \Delta d_{t+1} - (p_t - d_t)$$

Solve forward to express the present value relationship:

$$pd_t \approx \rho \times pd_{t+1} + \Delta d_{t+1} - r_{t+1}$$

$$pd_t \approx \sum_{j=1}^{k} \rho^{j-1} \Delta d_{t+j} - \sum_{j=1}^{k} \rho^{j-1} r_{t+j} + \rho^k \left(pd_{t+k}\right)$$

when $\rho^k \left(pd_{t+k}\right) \to 0$ (transversality condition, no rational bubbles)

$$pd_t \approx \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} - \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$$

Ex ante,

$$pd_t \approx \mathbb{E}_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} - \mathbb{E}_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$$

Campbell-Shiller decomposition

$$pd_t \approx \mathbb{E}_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} - \mathbb{E}_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$$

$$dp_t \approx \bar{dp} + \mathbb{E}_t \sum_{j=1}^{\infty} \rho^{j-1} \left[(r_{t+j} - \bar{r}) - (\Delta d_{t+j} - \bar{d}) \right]$$

Conditional on the fact that expected returns and dividend growth are both stationary, if there is a deviation in the price-dividend ratio from its long-run mean, it should forecast

- future returns,
- future dividend growth,
- or both

Early attempts to assess predictability estimated:

$$r_{t+1} - \bar{r} = \kappa_r (dp_t - \bar{dp}) + \epsilon^r_{t+1}$$
$$\Delta d_{t+1} - \bar{d} = \kappa_d (dp_t - \bar{dp}) + \epsilon^d_{t+1}$$
$$dp_{t+1} - \bar{dp} = \phi_d (dp_t - \bar{dp}) + \epsilon^{dp}_{t+1}$$

Indeed, empirical evidence seems to support $\kappa_r > 0$.

Big data - big choice

Last decades saw a tremendous increase in data availability, computational power and statistics development:

- Many markets, many asset classes
- Powerful software allows to easily run thousands of regressions in minutes
- High frequency data
- Technical analysis
- New datasets
- Linear/nonlinear models, various inference procedures lead to the question of model selection and its validation...

The search for predictability continues...

Price-dividend ratio was not the only statistically and economically significant predictor:

- earnings-price ratio: Lamont (1998)
- consumption-wealth ratio: Lettau and Ludvigson (2001)
- labour income-to-consumption: Santos and Veronesi (2006)
- cross-sectional price of risk: Polk, Thompson, and Vuolteenaho (2006)
- housing collateral ratio: Lustig and S. Van Nieuwerburgh (2005)
- short term interest rates: Fama and Schwert (1977)
- credit spreads: Keim and Stambaugh (1986)
- the term structure slope: Campbell (1987)
- stock volatility: French, Schwert and Stambaugh (1987)
- equity share of new issuance: Baker and Wurgler (2000)
- aggregate short interest: Lamont and Stein (2004)
- investor sentiment: Baker and Wurgler (2006)

Furthermore, long-term returns are much more predictable:

$$\sum_{j=1}^{H} r_{t+1} - r = \kappa_r^H (dp_t - \bar{dp}) + \epsilon_{t,t+H}^r$$

with $\kappa_r^H > \kappa_r$.

Prediction zoo

Abundant empirical predictability not only for the market, but also for the performance of particular popular trading strategies/portfolios.

Novy-Marx (2014) identifies a series of new important predictors for the market and a set of popular trading strategies:

Prediction zoo

Abundant empirical predictability not only for the market, but also for the performance of particular popular trading strategies/portfolios.

Novy-Marx (2014) identifies a series of new important predictors for the market and a set of popular trading strategies:

- the party of the sitting president
- the weather
- global warming
- the El Ni-ño phenomenon
- sunspot activity
- the conjunctions of the planets

Not only these results are statistically significant, quite often there is a plausible story, explaining the mechanism!

Stock market and the weather

Kamstra, Kramer, and Levi (2003): higher returns in colder months

- Higher levels of depression in Autumn (Seasonal Affective Disorder, SAD)
- Depression could lower risk appetite of the investors, and lead to lower prices
- Therefore, yields in winter become higher
- In spring, the mechanism is reversed

Cai and Wei (2006): "lower temperature can lead to aggression... [which] could result in more risk-taking... We therefore expect lower temperature to be related to higher stock returns."

But...

Stock market and the weather

Kamstra, Kramer, and Levi (2003): higher returns in colder months

- Higher levels of depression in Autumn (Seasonal Affective Disorder, SAD)
- Depression could lower risk appetite of the investors, and lead to lower prices
- Therefore, yields in winter become higher
- In spring, the mechanism is reversed

Cai and Wei (2006): "lower temperature can lead to aggression... [which] could result in more risk-taking... We therefore expect lower temperature to be related to higher stock returns."

But...

Hirshleifer and Shumway (2003): "psychological evidence and casual intuition predict that sunny weather is associated with upbeat mood", "sunshine is strongly significantly [positively] correlated with stock returns."

Stock and the weather

Novy-Marx (2014):
Cold weather in Manhatten predicts not just market returns, but also many trading strategies: market, but also for small cap strategies, value strategies, and strategies based on long run reversals, asset growth, and asset turnover .

Hot weather in Manhatten leads to abnormally high performance of many earnings related anomalies, including those based on return-on-assets, earnings-to-price, gross margins, financial strength, etc.

Stock and the weather

Novy-Marx (2014):
Cold weather in Manhatten predicts not just market returns, but also many trading strategies: market, but also for small cap strategies, value strategies, and strategies based on long run reversals, asset growth, and asset turnover .

Hot weather in Manhatten leads to abnormally high performance of many earnings related anomalies, including those based on return-on-assets, earnings-to-price, gross margins, financial strength, etc.

Problem:

Stock and the weather

Novy-Marx (2014):
Cold weather in Manhatten predicts not just market returns, but also many trading strategies: market, but also for small cap strategies, value strategies, and strategies based on long run reversals, asset growth, and asset turnover .

Hot weather in Manhatten leads to abnormally high performance of many earnings related anomalies, including those based on return-on-assets, earnings-to-price, gross margins, financial strength, etc.

Problem:

not only weather from NYC predicts traders's mood and behaviour, but also that in Bozeman, Montana, or Hawaii.

Celestial powers

Yu, Zheng, and Zhu (2006): "...since psychological studies associate **full moon phases** with depressed mood, this study hypothesizes that stocks are valued less and thus returns are lower during full moon periods."

Contradicts Kamstra, Kramer and Levis (2003) who show that investor depression leads to an *increase* in expected returns.

Novy-Marx (2004): focus on celestial angles and sunspots

- "The aspects of Mercury and Venus with the outer planets appear particularly important for the performance of anomalies, predicting the returns of the market, and strategies based on market cap, book-to-market, momentum, gross profitability..."
- "... Mars disproportionately influences our animal instincts, especially aggression, which is strongly associated with risk-taking."
- "High levels of solar activity seem to inhibit investors capacity to process information, reducing the rate at which news gets incorporated into prices. This increases the profitability of strategies that exploit slow adjustments of prices to fundamentals."
- "... both Black Monday (October, 1987) and the start of the great recession (2007) came at minimums in the solar cycle, times of negligible sunspot activity"

What's next in prediction?

Models and inference

Lars P. Hansen
Nobel Prize in Economics, 2013

"I view the work I've done related to statistics and economics as roughly speaking, how to do something without having to do everything. So economic models – how any model by definition isn't right.

When someone just says, 'Oh, your model is wrong.' That's not much of an insight. What you want to know is, is wrong in important ways or wrong in ways that are less relevant? And you want to know what does the data really say about the model?"

Concerns for spurious predictors

There are many concerns arising around simple regressions:

- In-sample vs out-of-sample performance
- Persistency and cyclicality of regressors
- Small sample (not just the number of time series observations!)
- Instability in the linear relationship, structural breaks
- Model selection
- P-hacking, multiple testing

Concerns for spurious predictors

There are many concerns arising around simple regressions:

- In-sample vs out-of-sample performance
- Persistency and cyclicality of regressors
- Small sample (not just the number of time series observations!)
- Instability in the linear relationship, structural breaks
- Model selection
- P-hacking, multiple testing

Two types of solutions:

- imposing a structural model on the reduced-form specification
- Proper statistical tools

Persistence and Stambaugh bias

Most financial ratios are very persistent, and it could substantially bias the results of the predictive regressions:

$$r_t = \beta x_{t-1} + \epsilon_t^r$$
$$r_t = \beta x_{t-1} + \epsilon_t^x$$

where $\epsilon_t := [\epsilon_t^r, \epsilon_t^x]$ is a martingale difference sequence innovations with

$$\mathbb{E}_{F_{t-1}}[\epsilon_t \epsilon_t'] = \begin{bmatrix} \Sigma_r & \Sigma_{rx} \\ \Sigma_{rx} & \Sigma_x \end{bmatrix}$$

Centered OLS coefficient estimator:

$$\hat{\beta} - \beta = \frac{\sum_{t=1}^n x_{t-1}\epsilon_t^r}{\sum_{t=1}^n (x_{t-1}^2)} = \frac{\sum_{t=1}^n x_{t-1}\epsilon_t^{r,x}}{\sum_{t=1}^n (x_{t-1}^2)} + \left(\frac{\Sigma_{r,x}}{\Sigma_x}\right) \frac{\sum_{t=1}^n x_{t-1}\epsilon_t^x}{\sum_{t=1}^n (x_{t-1}^2)}$$
$$= \frac{\sum_{t=1}^n x_{t-1}\epsilon_t^{r,x}}{\sum_{t=1}^n (x_{t-1}^2)} + \left(\frac{\Sigma_{r,x}}{\Sigma_x}\right)(\hat{\rho} - \rho)$$

where $\epsilon_t^{r,x} = \epsilon_t^r - \frac{\Sigma_{r,x}}{\Sigma_x}\epsilon_t^x$ and $\hat{\rho} = \left(\sum_{t=1}^n x_{t-1}^2\right)^{-1} \sum_{t=1}^n x_{t-1}x_t$.

Under normality and with a stationary regressor ($\rho < 1$), Stambaugh (1999) showed that

$$\mathbb{E}\left[\hat{\beta} - \beta\right] = -\frac{\Sigma_{r,x}}{\Sigma_x}\left(\frac{1+3\rho}{n}\right)$$

Persistence and Stambaugh bias

For highly persistent variables, the asymptotics is non-standard and Stambaugh bias is no longer a finite sample problem, that can be easily corrected.

Campbell and Yogo (2006) and Jansson and Moreira (2006): local-to-unity modelling tool, $\rho = 1 + \frac{c}{n}$.

$$n(\hat{\beta} - \beta) = \frac{\frac{1}{n}\sum_{t=1}^{n} x_{t-1}\epsilon_t^r}{\frac{1}{n^2}\sum_{t=1}^{n}(x_{t-1}^2)} \Rightarrow \frac{\int J_x^c(\tau)dB(\tau)}{\int J_x^c(\tau)^2 d\tau}$$

where B is Brownian motion and $J_x^c(\tau)$ is a linear diffusion.

Valkanov (2003): long-horizon regressions induce further complications.

Main solutions:

- Bonferroni method: Cavanagh et al. (1995) , Campbell and Yogo (2006)
- Conditional likelihood with sufficient statistics: Jansson and Moreira (2006)
- Control function approach: Elliott (2011)
- IVX method: Phillips and Lee (2012).

Phillips and Lee (2013) extend the IVX approach to allow for general degrees of persistence (mildly integrated, integrated, explosive, etc) and long horizon regressions.

Novy-Marx (2014) weather factors

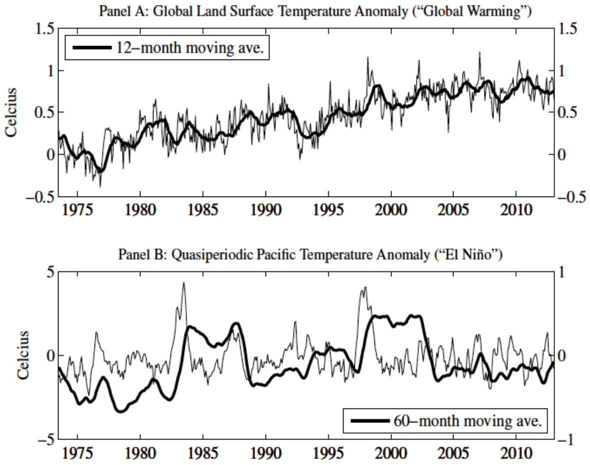


Fig. 3. Other Climatic Variables. This figure shows the levels of two non-seasonal climatic variables. Panel A shows the global average land temperature relative to the 1951-1980 base period. Panel B shows the deviations from average surface temperatures of the tropical Eastern Pacific Ocean. The data cover July 1973 to December 2012.

Imposing a structure

Niewerbourg and Koijen (2007)

$$\Delta d_{t+1} - \bar{d} = z_t + \epsilon^d_{t+1}, \quad z_{t+1} = \phi z_t + \epsilon^z_{t+1}$$
$$r_{t+1} - \bar{r} = x_t + \epsilon^r_{t+1}, \quad x_{t+1} = \phi x_t + \epsilon^x_{t+1}$$

Fundamental accounting identity implies that

$$\epsilon^r_{t+1} = \frac{-\rho}{1-\rho\phi}\epsilon^x_{t+1} + \frac{\rho}{1-\rho\phi}\epsilon^z_{t+1} + \epsilon^D_{t+1}, \qquad dp_t - \bar{dp} = \frac{x_t - z_t}{1-\rho\phi}$$

Three fundamental shocks:

- innovation in expected dividends, ϵ^z_{t+1}
- innovation in expected returns, ϵ^x_{t+1}.
- innovation in unexpected dividends, ϵ^d_{t+1}.

Since $\rho, \phi > 0$ and $\rho\phi < 1$, a positive shock to expected returns leads to a negative contemporaneous return, and a shock to expected or unexpected dividend growth induces a positive contemporaneous return.

Assume all three shocks are serially and cross-sectionally uncorrelated, except for

- $cov(\epsilon^x_{t+1}, \epsilon^z_{t+1}) = \chi$
- $cov(\epsilon^d_{t+1}, \epsilon^z_{t+1}) = \lambda$

Contaminated predictor problem

- Returns are predicted by x_t
- However, when we run a regression of returns on PD-ratio, the predictor is noisy: pd_t varies due to x_t and z_t
- This creates an error-in-variables problem and biased parameter estimates

$$\kappa_r = \frac{cov(r_{t+1}, dp_t)}{var(dp_t)} = \frac{(1-\rho\phi)(\sigma^2_{\epsilon^x_t} - \chi)}{\sigma^2_{\epsilon^x_t} + \sigma^2_{\epsilon^z_t} - 2\chi}$$

If growth rates are constant ($\chi = \sigma^2_{\epsilon^z_t} = 0$), then the dividend-price ratio is a perfect predictor of returns and $\kappa_r^* = 1 - \rho\phi$.

$$\kappa_r^* - \kappa_r = \frac{(1-\rho\phi)(\sigma^2_{\epsilon^z_t} - \chi)}{\sigma^2_{\epsilon^x_t} + \sigma^2_{\epsilon^z_t} - 2\chi}$$

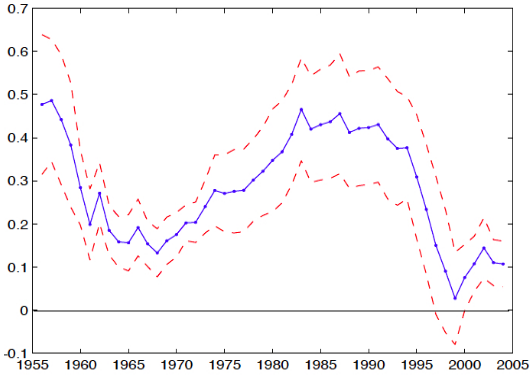
If there is a sufficiently high correlation between expected dividend growth and expected returns, the bias could be upward.

Parameter instability

Niewerbourg and Koijen (2007): consider a standard predictive regression:

$$r_{t+1} - \bar{r} = \kappa_r(dp_t - \bar{dp}) + \epsilon^r_{t+1}$$

Focus on rolling window regression (30 years): fit varies from 0 to 0.3, $\hat{\kappa}_r$ from 0 to 0.5.



In-sample vs out-of-sample, parameter instability

Goyal and Welsh (2009):

- most predictive results are no longer significant in-sample
- they were never significant out-of-sample
- a lot of the old significance results originated during the years of the Oil Shock (1973-75)
- Model selection does not improve their performance

Most technical predictors found in the earlier papers are useless.

Multiple testing: why is it a problem?

- You have a set of $n = 20$ predictors and you assess their performance one by one, using a t-test at $\alpha = 5\%$ significance level.
- For simplicity, assume tests are independent
- Under the null of no predictability, what are the chances that at least one variable will turn out significant?

$$Prob = 1 - (1 - 0.05)^{20} \approx 64\%$$

- Simple Bonferroni correction: set the test level to α/n. In this example,

$$Prob = 1 - (1 - 0.05/20)^{20} \approx 0.0488\%$$

- Depending on the correlation structure between the tests, Bonferroni correction could lead to very conservative tests, producing high rates of false negatives.
- Romano and Wolf (2010) : resampling and subsampling approach with the focus on generalized family-wise error rate.

Is there any hope for predictability left?

Proper statistical inference eliminates a huge number of spurious predictors. Is there anything left?

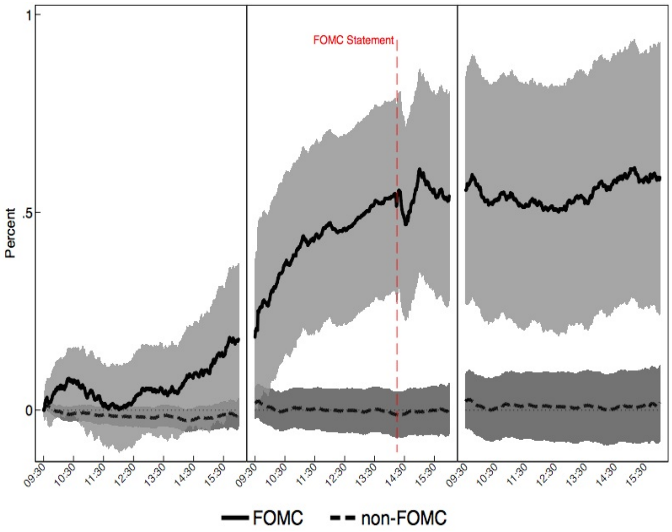
Three examples:

- fundamental analysis: FOMC announcement drift
- technical analysis (LASSO): sparse short-lived predictors in the cross-section of stocks
- technical analysis (Random Forest): non-monotonic predictive content of the past stocks prices

Lucca and Moench (2015)

- Federal Open Market Committee meets on average 8 times a year and decide on the monetary policy (fed funds rate, forward guidance, etc)
- The dates and times of the announcements are known in advance, since 1994 they summarise the outcomes at 2.15pm
- Since 1994, the S&P index has on average increased 49 basis points in the 24 hours before scheduled FOMC announcements and did not revert.
- For comparison, during any other 24 hour window, S&P 500 increased only by 0.5 basis points.
- As a result, about 80% of the market excess return during the last 20 years was due to FOMC announcement drift
- A simple 24 hour buy and hold strategy before the announcement yields Sharpe ratio of 1.14.
- The finding is unlikely to be due to data snooping and small sample effects.
- FOMC announcement drift is not present in bonds
- S&P 500 does not exhibit similar drift before major macroeconomic announcements (nonfarm payroll employment, weekly initial claims for unemployment insurance, and CPI)

Cumulated S&P500 excess returns



Many stocks, many predictors

Imagine, I want to check whether current stock returns of Apple and all the other stocks can be predicted by recent returns on some other stocks.

- U.S.: 5,000 on exchanges
- 10,000 on OTC markets
- world-wide: about 63,000

The nature of predictability is believed to be sparse and short-lived. What to do?

LASSO

- Least Absolute Shrinkage and Selection Operator (Tibshirani, 1996) is a tool for simultaneous variable selection and parameter estimation.
- Consider a linear model:

$$r_t = x'_{t-1}\beta + \epsilon_t$$

  where $t = 1..T$, x_{t-1} is a $p \times 1$ vector of predictors.
- When $p > T$, OLS is no longer applicable.
- LASSO enforces sparcity and recovers influential signals at the same time:

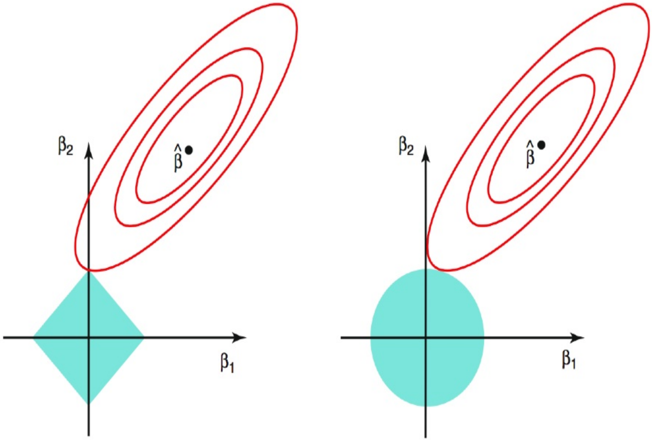
$$\hat{\beta}_{L1} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{T} \left(r_t - x'_{t-1}\beta\right)^2 + \lambda_T \sum_{j=1}^{p} |\beta_j|$$

- Note, the sparcity is enforced through L1-regularization, not shrinkage in general
- A solution to ridge regression won't have the same property:

$$\hat{\beta}_{L2} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{t=1}^{T} \left(r_t - x'_{t-1}\beta\right)^2 + \lambda_T \sum_{j=1}^{p} |\beta_j|^2$$

Graphical interpretation

Contours of the error and constraint functions for lasso and ridge regressions:



Sparse predictors in the stock exchange

Chinco, Clark-Joseph and Ye (2016): predictability of one-minute returns:

$$\hat{\beta}_{ols} = \arg \min_{\beta \in \mathbb{R}^{1+3n}} \left\{ \frac{1}{2 \times 30} \sum_{\tau=t-29}^{t} \left(r_{i,t} - \beta_0 - \sum_{l=1}^{3} \beta_{i,l} r_{j,t-l} \right)^2 + \lambda \sum_{l=1}^{3} |\beta_{i,l}| \right\}$$

$$\hat{\beta}_{L1} = \arg \min_{\beta \in \mathbb{R}^{1+3n}} \left\{ \frac{1}{2 \times 30} \sum_{\tau=t-29}^{t} \left(r_{i,t} - \beta_0 - \sum_{j=1}^{n} \sum_{l=1}^{3} \beta_{j,l} r_{j,t-l} \right)^2 + \lambda \sum_{j=1}^{n} \sum_{l=1}^{3} |\beta_{j,l}| \right\}$$

Focus on NYSE-listed stocks ($n \approx 2000$)and the out-of-sample performance of

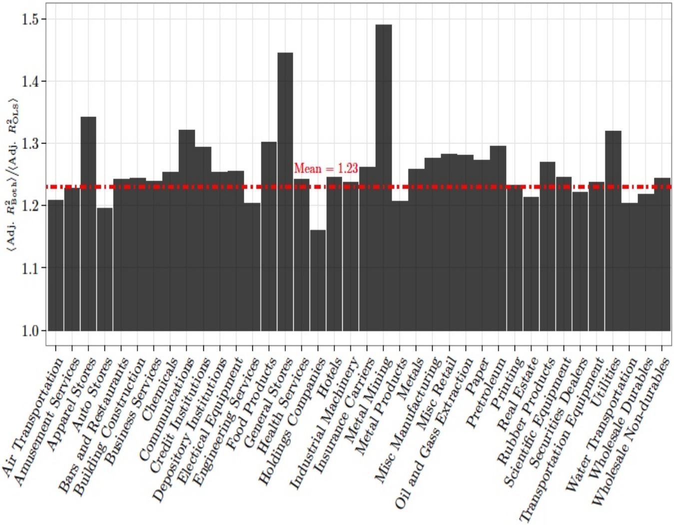
- OLS estimator, using the stock's own lags
- OLS + lag selection
- LASSO on its own and other stocks' lags

LASSO leads to substantial gains in forecasting abilities:

- On average, LASSO selects 11 predictors (own and other stocks' lags)
- 23% growth in the quality of out-of-sample forecasting
- signals vary over time and stocks, could be related to industry and supply chain
- trading strategies delivers 0.4% per month (net of costs)

Heterogeneity across industries

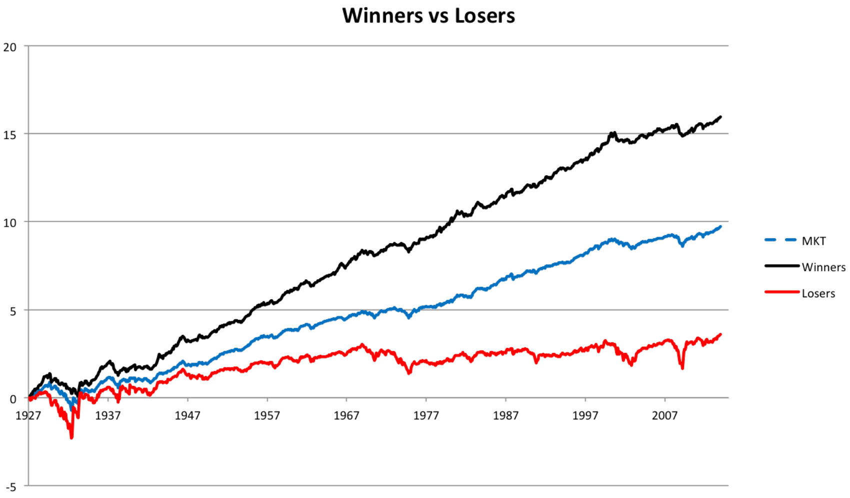
LASSO improvement in the out-of-sample R^2-adjusted across different industries:



Random Forest

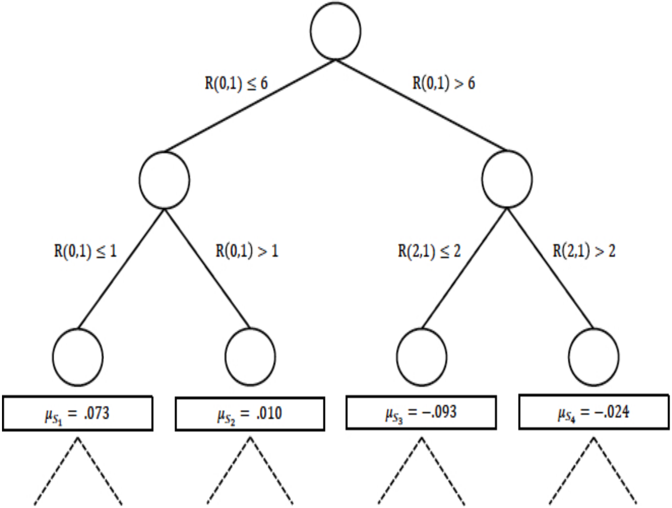
- Which information in the past returns is most relevant for today's forecast?
- Jagadeesh and Titman (1993) found the evidence of momentum: stocks that performed over the last 12 months (excluding the very last one), tend to continue do well.
- Momentum strategy: sort all the stocks by their past performance, buy past winners and sell losers.
- Moritz and Zimmerman (2016) question whether the impact is monotonic.
- Tree-based consitional portfolio sorts: based on the attribute $x_t \geq x^*$, leading to the highest differences in returns.
- Focus on low frequency: monthly returns
- Main idea: minimize within-group variation
- 20 firm characteristics with pairwise and three-way interactions lead to 16700 potential variables for sorting.
- Model averaging across randomly selected subsets of variables

Standard momentum: winners vs. losers (WML)



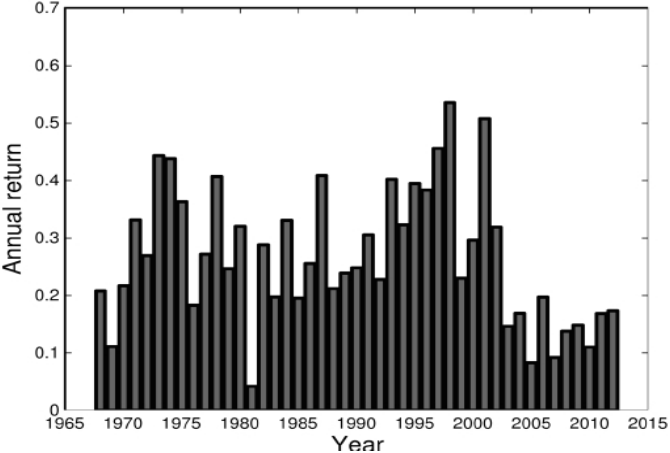
An example of a tree

All securities are sorted into two groups based on the threshold in the attribute that minimize within-group variation of returns.



Strategy payoff

The trading strategy based on buying highest decile, and selling the lowest, one, delivers high risk-adjusted returns.



Introduction to cross-sectional prediction

- Tree-based conditional sorts evaluated out-of-sample returns predictions.
- Key concept: we did not try to predict the time series of returns!

Introduction to cross-sectional prediction

- Tree-based conditional sorts evaluated out-of-sample returns predictions.
- Key concept: we did not try to predict the time series of returns!
- Instead, we predicted which stocks will have higher returns that others
- This is a different dimension of predictability: cross-sectional!
- Most of the current predictability literature in empirical asset pricing focuses on this aspect: which stocks will have higher returns?

Famous cross-sectional predictors

Value vs Growth:

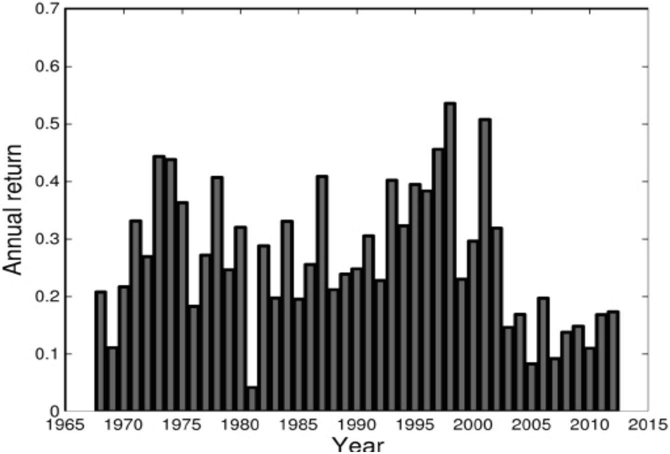
- stocks with low book-to-market ratio (growth companies) tend to have lower returns compared with those with low B/M ratio (value companies) $\Rightarrow$ **Value premium**
- HML strategy: sort stocks based on their BM ratio, form a portfolio by buying **high** B/M stocks and selling **low** B/M stocks.

Big vs Small:

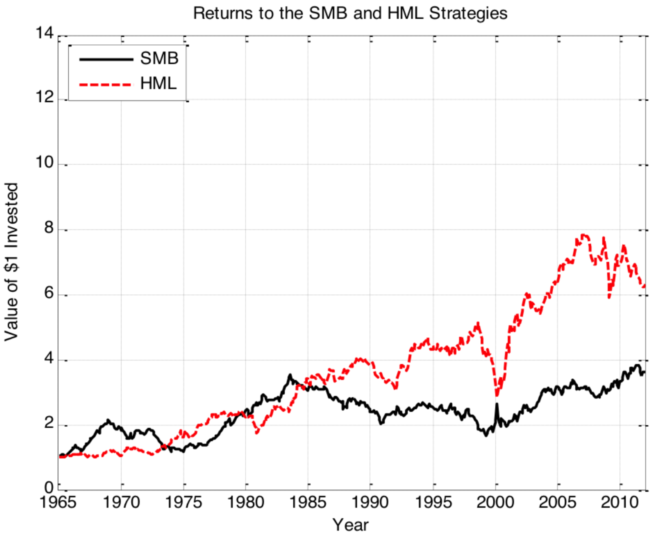
- stock with small market capitalisation (small caps) tend to have higher returns than the companies with large market capitalisation (large caps) $\Rightarrow$ **Size premium**
- SMB strategy: sort stocks based on their size, form a portfolio by buying **small** cap stocks and selling **big** cap stocks.

Strategy payoff

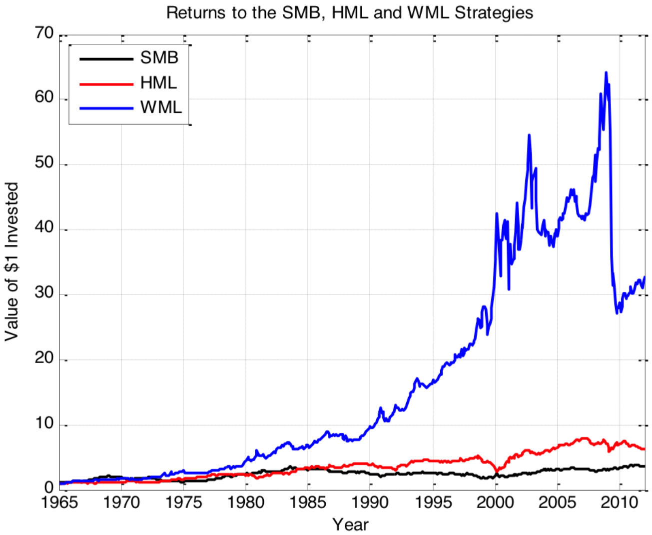
The trading strategy based on buying highest decile, and selling the lowest, one, delivers high risk-adjusted returns.



Long-short strategies



Long-short strategies



Free lunch?

With thousands of stocks and thousands of characteristics, what other strategies could we use?

Do these returns indicate *'good'* companies to invest?

Cross-sectional predictability vs Efficient Market Hypothesis?

Free lunch?

With thousands of stocks and thousands of characteristics, what other strategies could we use?

Do these returns indicate *'good'* companies to invest?

Cross-sectional predictability vs Efficient Market Hypothesis?

Next lecture!