Asset Management I

Week 5

Daniele Bianchi[1]

whitesphd.com

[1]School of Economics and Finance
Queen Mary, University of London

Summary

This week we introduce basic concepts of regression trees and random forest. Based on advanced research we are going to start to look into returns predictability within the context of investment decision making.

Contents

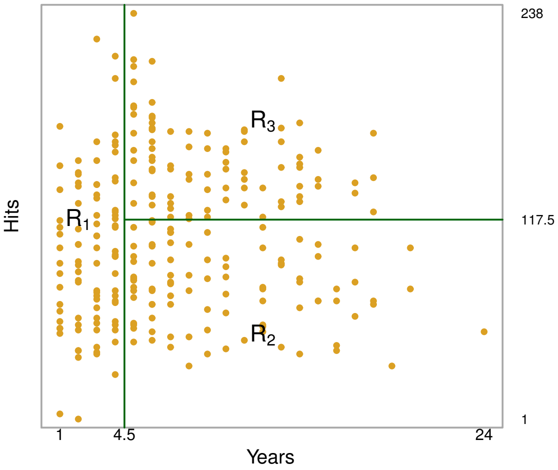
Recap

Recap

- Regularization
  - Introducing additional constraints to get more sensible solution
  - Avoid overfitting
- Solve a common problem: How can we make progress with there are many possible predictors of an outcome variable?
- This lecture:
  - Introduction to Trees
  - Calibration
  - Introduction to Asset Management
- Note: I have removed additional cross validation methods for now. It will be covered later in the course, should there be time.

Classification Trees

The Basics of a Tree



Source: "An introduction to Statistical Learning", James et al. (2013).

The Basics of a Tree

- Top split: Years < 4.5.
- Next split: Number of baseball hits $<$ 117.5.
- Three regions, and corresponding estimates of salary.
- R_1, R_2, and R_3 are called *terminal nodes* or leaves of the tree.
- Split rules are called *internal nodes*.

Classification Trees

- The basic idea of trees is to partition the space of X into subspaces, and then to estimate the classification outcome for new units as the most commonly occurring class in the training set with values in the same subspace.
- The partitioning is sequential - called "recursive binary splitting," one covariate at a time.
- Guidance for making these splits is provided by some criterion.

Classification Trees: Splitting

- One way to split is to compute the classification error rate as the fraction of training observations in each region that does not belong to the most commonly occurring class, i.e., $1 - \max_k \hat{p}_{mk}$ where $\hat{p}_{mk}$ is the proportion of training observations in the mth region belonging to the kth class.
- Another (preferable) one is the Gini index: $G = \Sigma_k \hat{p}_{mk}(1 - \hat{p}_{mk})$. Clearly, this will be low when all $\hat{p}_{mk}$ observations are close to 0 or 1, i.e., nodes are "pure".
- A third is called cross-entropy: $D = -\Sigma_k \hat{p}_{mk} \log \hat{p}_{mk}$, which has similar properties to the Gini index.

A Tree Cookbook - I

- Pick any X variable X_p. Pick a threshold t and consider splitting the data on this threshold, i.e., depending on whether

$$X_{i,p} \leq t \text{ versus } X_{i,p} > t$$

- Let the estimator then be:

$$f_{p,t}(x) = \left\{ \begin{array}{l} \max_k \hat{p}_k(X_{i,p} \leq t) \text{ for } x_p \leq t \\ \max_k \hat{p}_k(X_{i,p} > t) \text{ for } x_p > t \end{array} \right\}$$

- Then compute the criterion function G, and find both the covariate p^* and the threshold t^* that solves:

$$\arg\min_{p,t} G(f_{p,t}(x))$$

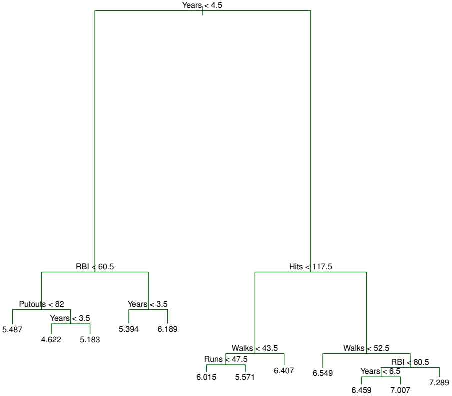
A Tree Cookbook - II

- Partition the covariate space into two subspaces based on whether $X_{i,p^*} \leq t^*$ or not.
- Repeat, splitting each subspace in the way that leads to the biggest improvement in the objective function.
- Keep splitting the subspaces to minimize the objective function, optionally, with a penalty λ for the number of splits (also called leaves).
    - That is, with penalty, objective function is now $Q = G(f_{p,t}(x)) + \lambda(\#$ of leaves).

Pruning the Tree

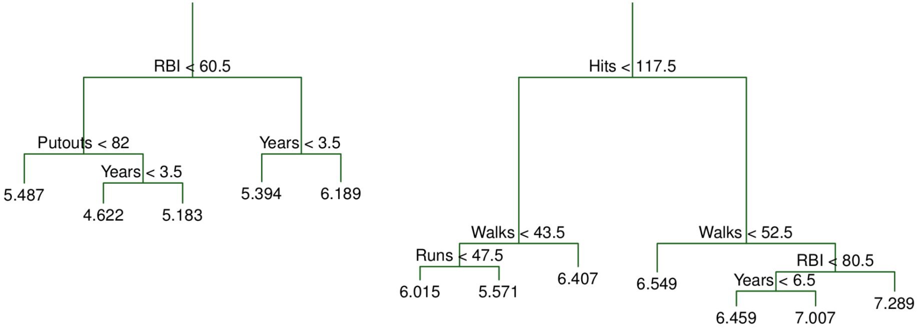
- A frequently used approach is to grow a big tree by using a deliberately small value of the penalty term, or simply growing the tree till the leaves have a preset small number of observations.
- Then go back and prune branches or leaves that do not collectively improve the objective function sufficiently.

A Big Tree

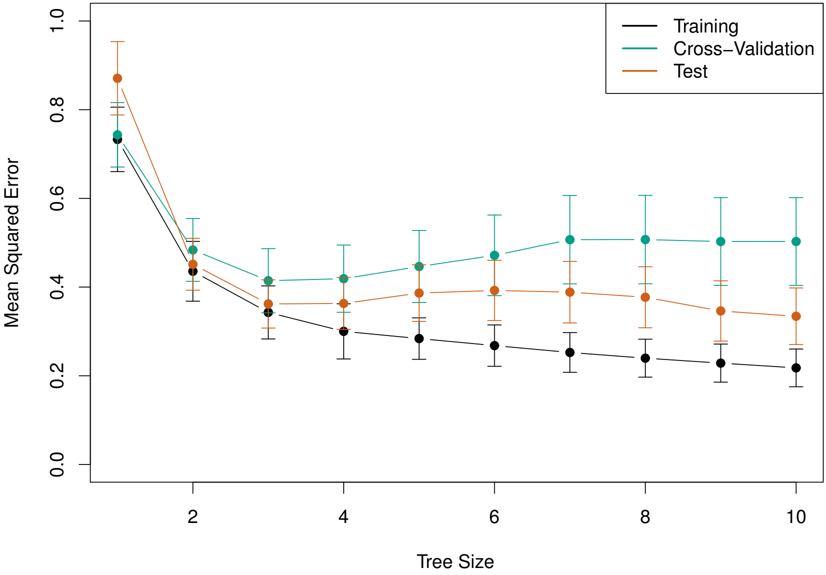


Source: "An introduction to Statistical Learning", James et al. (2013).

A Big Tree


RBI < 60.5
Putouts < 82
Years < 3.5
Years < 3.5
5.487
4.622
5.183
5.394
6.189
Hits < 117.5
Walks < 43.5
Walks < 52.5
Runs < 47.5
RBI < 80.5
Years < 6.5
6.015
5.571
6.407
6.549
6.459
7.007
7.289

How to pick the optimal number of leaves?



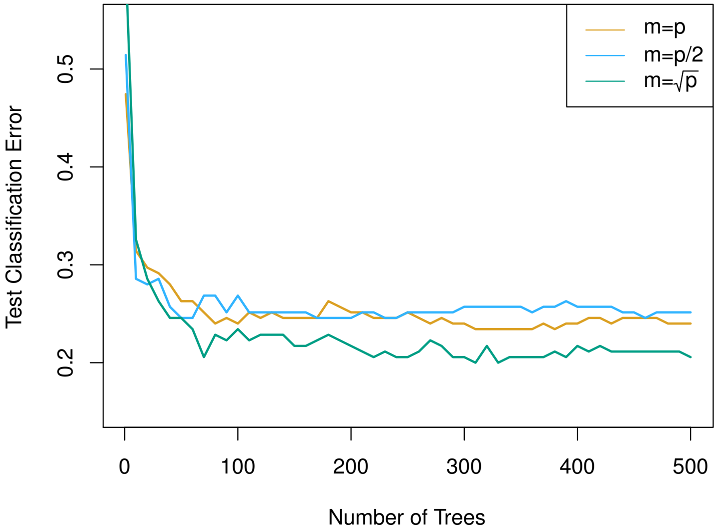
Two types of Trees

- Regression Tree: The outcome for the defined group is defined as the mean of the outcome for all of the observations.
- Classification Tree: Similar, but it is used to predict a qualitative response or / categories.
- Key points:
  1. Trees are very easy to explain, displayed graphically.
  2. Handle both quantitative and qualitative predictors.
  3. Prediction accuracy low. So, need to do something to improve it.

Random Forests

- Intuition: If one tree is a problem, we grow many trees.
- And from growing the “forest”, we average the outcome to improve accuracy.
- Steps:
  1. You randomly select a fraction of your training data.
  2. You fit a tree on this “bootstrapped” sample, and store it.
  3. Importantly, for every split in the tree, you restrict the split candidates to a random sample of predictors in your data.

Gains from Random Forest



Calibration

Calibration

- Classification algorithms are useful, but for many applications, we need a probability that an observation belongs to a particular class label, rather than the predicted class label alone.
- Many algorithms we have studied cannot give you the probabilities that an outcome will be of a predicted class label.
- Example:
    1. Knowing if stock i is going to have positive or negative return next day (Hard classification).
    2. What is the probability that stock i is going to have positive or negative return next day (Probabilistic classification).
- Both useful. (1) For momentum based investing, and (2) for hedging purposes.

"Hard" and Probabilistic Classification

- Many algorithms in their basic form aren't set up to give you probabilities. They are "hard" classifiers rather than probabilistic classifiers:
    - Tree-based methods.
    - Support vector machines.
- However, others are naturally probabilistic:
    - Logistic regressions.
    - Naive Bayes.
    - More generally, in econometrics, such probabilistic classifiers are broadly studied in the area of discrete choice.

Probabilistic to Categories

- It's easy to transform a probabilistic classifier into a "hard" classifier.
- The simplest way (for a binary "hard" classification) is to set a cutoff level, and simply classify observations above and below the cutoff into the two classes.
- Generally speaking, since probabilistic classifiers optimize conditional probabilities on the training set, this cutoff is optimally 50%.
- More generally, you can find the optimal decision rule (in the training dataset) as the one that minimizes the risk, i.e., the expected loss function (or in most applications, the empirical risk, which is the average of the loss function on the training dataset.

Introduction to Asset Management

- Predictability.
    - Definitions and preliminaries.
    - Non-return forecasting variables.
- The out-of-sample predictability debate.
    - Goyal and Welch (2008)
- New approaches with big data. (Next week)

Market Efficiency

- Fama (1970) defines a market as efficient if "prices fully reflect all available information". In practice this means:

  $$R_{i,t+1} = \Theta_{it} + U_{i,t+1},$$

  where Θ_{it} is the rationally expected return on asset i, and $U_{i,t+1}$ has zero expectation with respect to the information set at t.

- This will only have content if we can restrict Θ_{it} using an economic model.

- Thus market efficiency is not testable except in combination with a model of expected returns.
  - This is called the joint hypothesis problem.

Time-Series Efficiency

- Fix i, model returns over t.
- Economic model for Θ_{it} is a time series model. The simplest such model is $\Theta_{it} = \Theta$, a constant.
- Equilibrium models with time-varying expected returns also possible.
- We will concentrate in this lecture on explaining (and more importantly, forecasting) aggregate stock index behaviour.

Defining Time-Series Efficiency

- Even with a model of expected returns, we need to specify what is in the information set used to form expectations of $U_{i,t+1}$ at time t.
- Fama defines 3 forms of the efficient market hypothesis, corresponding to what is included in the information set used to forecast $U_{i,t+1}$:
    - Weak form. Past returns.
    - Semi-strong form. Past publicly available information, e.g. stock splits, trading volume, dividends, earnings (also returns).
    - Strong form. Any past information, even if it is only available to insiders.

Examples

Michael Jensen (1978): "There is no other proposition in economics which has more solid evidence supporting it than the Efficient Markets Hypothesis".

Examples

Robert Shiller (1984): "Returns on speculative assets are nearly unforecastable; this fact is the basis of the most important argument in the oral tradition against a role for mass psychology in speculative markets. One form of this argument claims that because real returns are nearly unforecastable, the real price of stocks is close to the intrinsic value, that is, the present value with constant discount rate of optimally forecasted future real dividends. This argument... is one of the most remarkable errors in the history of economic thought".

Debate continues to this day...

- The debate continues to this day. Debates currently center around:
  - Whether predictability exists at all, in and out of sample.
  - If predictability exists, whether the proximate cause is frictions or irrationality.

Short and Long-Run Predictability

- Short-term return predictability is easy to detect if it is present, and hard to explain using a risk-based asset pricing model.
    - Can have modest effects on prices, and disappear quickly if arbitrageurs discover it.
    - Can also disappear if transactions costs decline making arbitrage cheaper (e.g., decimalization).
- Long-term return predictability can have large effects on prices; harder to detect without a very long time series.
    - Can be explained by a more sophisticated model of risk and return.

Alternative Time Series Hypotheses

- To devise meaningful time-series tests and interpret the results, it is useful to consider alternative hypotheses:
- Market prices contaminated by short-term noise, generating short-run reversals.
  - Noise could be caused by illiquidity (bid-ask bounce or other issues).
- Market reacts sluggishly to information, generating short-run predictability based on past returns or information releases.
- Market prices deviate substantially from efficient levels; long-lasting deviations are hard to arbitrage.
  - Generates long-run reversal and predictability based on price levels, and is difficult to detect in the short run.

Tests of Autocorrelation in Stock returns

- Do past returns predict future returns (weak-form market efficiency)?
  - One way to check is to inspect the autocorrelations of stock returns.

- There has been a significant amount of literature devoted to this topic, which includes (but is not limited to):
  - Box-Pierce statistics.
  - Variance ratio statistics.
  - Inference on long horizon and short-horizon return autocorrelations and asymptotic inference.
  - Cross-correlations and cross-autocorrelations (lead-lag effects).
- We won't spend too much time on this, but present analogous results on time-series momentum which employ regression analysis.
  - We can then simply apply big data techniques to these regression problems.

Time-Series Momentum

- Moskowitz, Ooi, and Pedersen (2012) investigate time-series regressions for a large set of assets, of the form:

$$r_t^s/\sigma_{t-1}^s = \alpha + \beta_h r_{t-h}^s/\sigma_{t-h-1}^s + \varepsilon_t^s,$$

  - where σ_{t-1}^s is the square-root of ex-ante volatility, which is estimated as:
$$\sigma_t^2 = 261 * \sum_{i=0}^{\infty}(1-\delta)\delta^i(r_{t-1-i} - \bar{r}_t)^2$$
  - the weights $(1-\delta)\delta^i$ add up to 1, and $\bar{r}_t$ is the exponentially weighted moving average return.
  - δ chosen such that $\sum_{i=0}^{\infty}(1-\delta)\delta^i i = \frac{\delta}{1-\delta} = 60$ days.

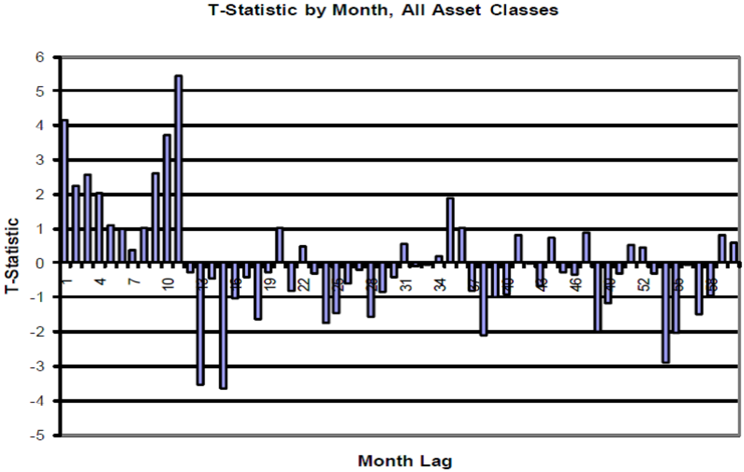
Time-Series Momentum

- These authors also estimate:

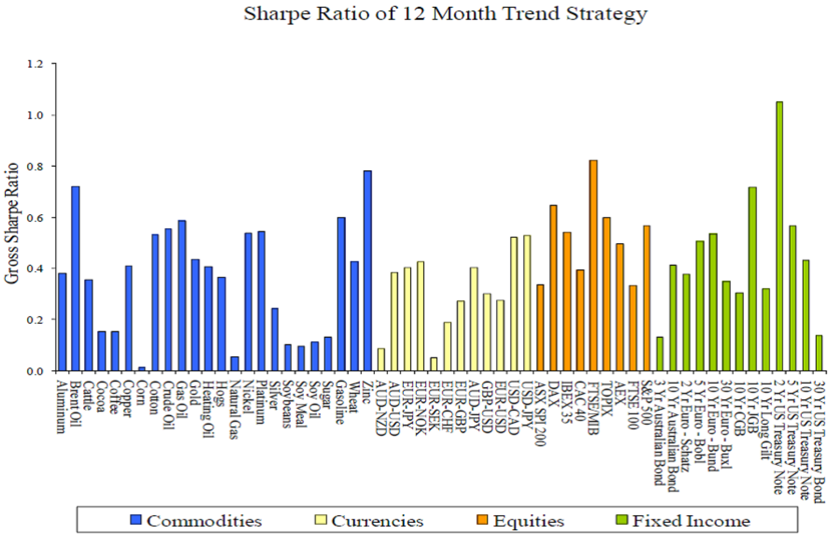
$$r_t^s/\sigma_{t-1}^s = \alpha + \beta_h sign(r_{t-h}^s) + \varepsilon_t^s.$$

- We can easily apply techniques such as regression trees to this analysis to check for nonlinearities in the forecasting relationship.

Time-Series Momentum



Time-Series Momentum



Semi-Strong Form Efficiency

- Authors have also tried many different forecasting variables for returns.
- As mentioned, debate in the literature about whether predictability actually exists, i.e., is predictability statistically and economically meaningful?
- We adopt a purely statistical approach, for now.
- A fast summary of the evidence for non-return predictability follows.

A Fast Summary of Forecasting Variables

- Nominal Interest Rates and Inflation:
  - the short-term interest rate.
  - the long-term bond yield.
  - the term spread between long- and short-term Treasury yields.
  - the default spread between corporate and Treasury bond yields.
  - the lagged rate of inflation.

A Fast Summary of Forecasting Variables

- Valuation Ratios.
- Each of these ratios has some accounting measure of corporate value in the numerator, and market value in the denominator:
  - the price dividend ratio.
  - the earnings price ratio.
  - the smoothed price earnings ratio: proposed by Campbell and Shiller (1988a, 1998b) is the ratio of current price to a 10-year moving average of earnings.
  - the book to market ratio.

D/P Ratio and Stock Prices

Table 20.1. *OLS regressions of percent excess returns (value weighted NYSE − treasury bill rate) and real dividend growth on the percent VW dividend/price ratio*

Horizon k	$R_{t\to t+k} = a + b(D_t/P_t)$			$D_{t+k}/D_t = a + b(D_t/P_t)$		
(years)	b	$\sigma(b)$	R^2	b	$\sigma(b)$	R^2
1	5.3	(2.0)	0.15	2.0	(1.1)	0.06
2	10	(3.1)	0.23	2.5	(2.1)	0.06
3	15	(4.0)	0.37	2.4	(2.1)	0.06
5	33	(5.8)	0.60	4.7	(2.4)	0.12

$R_{t\to t+k}$ indicates the k-year return. Standard errors in parentheses use GMM to correct for heteroskedasticity and serial correlation. Sample 1947–1996.

A Fast Summary of Forecasting Variables

- Other Forecasting Variables:
  - the equity share of new issues proposed by Baker and Wurgler (2000).
  - net equity issuance $NTIS$, which represents the net issuing activity (IPOs, SEOs, stock repurchases, less dividends) of firms as a percentage of their market capitalization.
  - the consumption-wealth ratio of Lettau and Ludvigson (2001).
  - the aggregate portfolio illiquidity of hedge funds, proposed by Kruttli, Patton, and Ramadorai (2014).

Kruttli, Patton, and Ramadorai (2013)

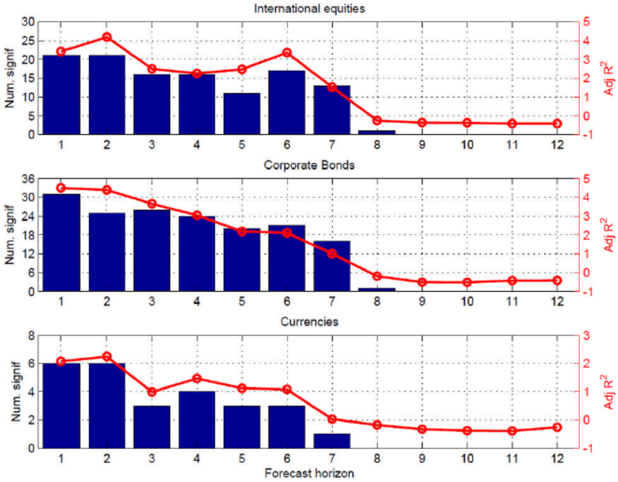


Figure 2 A: This figure shows the average adjusted R^2 and number of significant coefficients of the hedge fund illiquidity measure for different forecast horizons (in months). The results are for single predictor in-sample regressions, i.e. the only predictor is the hedge fund illiquidity measure.

Why Do We Care?

- Return predictability is evidence of market inefficiency; prices take long but temporary swings away from fundamental value.
- Predictability is evidence of time-varying equilibrium expected returns (note this is an excess return prediction).
    - Fama and French (1989) suggest that expected returns vary over business cycles, (i.e., takes a higher risk premium to induce stock holdings at the bottom of a recession).
    - Mechanically, when expected returns go up, prices go down.
    - We see low prices, followed by higher returns required by the market.
    - Changes in expected returns could be driven by changing investment opportunity sets or changing prices for risk.

Why Do We Care?

- A huge normative literature has sprung up around this, that purports to tell people how to invest.
    - If returns are predictable, especially out of sample, it helps us think about portfolio choice in the real world.
    - This is where we are headed next.
- An important debate in the literature (and in practice) about whether we can indeed find any variables that help us predict aggregate market returns.
- Our goal is to see if some of the big data techniques that we have discussed can help.

The Predictability Debate

Examples

Goyal and Welch (2008): "Our paper has systematically investigated the empirical real-world out-of-sample performance of plain linear regressions to predict the equity premium. We find that none of the popular variables has worked – and not only post-1990... Our profession has yet to find a variable that has had meaningful robust empirical equity premium forecasting power, at least from the perspective of a real-world investor."

Examples

Campbell and Thompson (2008): "In this note we show that forecasting variables with significant forecasting power insample generally have a better out-of-sample performance than a forecast based on the historical average return, once sensible restrictions are imposed on the signs of coefficients and return forecasts. The out-of-sample predictive power is small, but we find that it is economically meaningful."

Goyal and Welch (2008): The Method

- The Goyal and Welch procedure is to run a regression:

$$r_{t+1} = a + bx_t + \varepsilon_{t+1}$$

from time $t = 1$ to time $t = \tau$.

- Then use

$$\mathrm{E}_\tau[r_{\tau+1}] = a + bx_\tau$$

as a forecast of the next period return.

- This defines an out-of-sample forecast error of the forecasting model as:

$$f_{\tau+1} = r_{\tau+1} - \mathrm{E}_\tau[r_{\tau+1}]$$

Goyal and Welch (2008): The Method

- We can define the 'historical mean return' forecast error as:

$$m_{\tau+1} = r_{\tau+1} - \frac{1}{\tau}\sum_{t=1}^{\tau} r_t$$

  Note that this is just forecasting the next period return using the sample average.

- Define a Mean Squared Error (MSE) of each model as:

$$\frac{SSE(R)}{T} = \frac{1}{T}\sum_{\tau=start}^{end} \left(f_{\tau+1}\right)^2 \quad \text{and} \quad \frac{SSE(M)}{T} = \frac{1}{T}\sum_{\tau=start}^{end} \left(m_{\tau+1}\right)^2$$

The Method

- Next, compute the difference between the (square roots of) these mean-squared errors:

$$\Delta RMSE = \sqrt{SSE(M)/T} - \sqrt{SSE(R)/T}$$

- Clearly, if the forecasting model is better, it will have smaller forecast errors, i.e., $\sqrt{SSE(M)/T} > \sqrt{SSE(R)/T}$, and thus $\Delta RMSE > 0$.
- In the results shown below, $\Delta RMSE\ D + 20$ begins the annual forecasting exercise twenty years after data first become available, and $\Delta RMSE\ 1965$ begins the annual forecasting exercise in 1965.

Many Predictors Insignificant Insample

Panel A: Insignificant in-sample predictors

	Variable	Full Data: Data	IS $\overline{R}^2$	OOS ΔRMSE D+20	OOS ΔRMSE 1965–
d/p	Dividend Price Ratio	1872–2004	0.47	–0.1092	–0.0908
d/y	Dividend Yield	1872–2004	0.89	–0.0971	–0.3162
e/p	Earning Price Ratio	1872–2004	1.00	–0.0886	0.0899
d/e	Dividend Payout Ratio	1872–2004	–0.75	–0.3140	–0.1846
svar	Stock Variance	1885–2004	–0.76	–2.3405	0.0104
ntis	Net Equity Expansion	1927–2004	–0.03	–0.0352	–0.2303
tbl	T-Bill Rate	1920–2004	0.57	–0.1083	–0.1318
lty	Long Term Yield	1919–2004	–0.53	–0.4638	–0.7499
ltr	Long Term Return	1926–2004	1.00	–0.7696	–1.2016
tms	Term Spread	1920–2004	0.30	–0.0488	–0.0008
dfy	Default Yield Spread	1919–2004	–1.20	–0.1376	–0.1249
dfr	Default Return Spread	1926–2004	0.38	–0.0330	–0.0194
infl	Inflation	1919–2004	–0.98	–0.1939	–0.0714

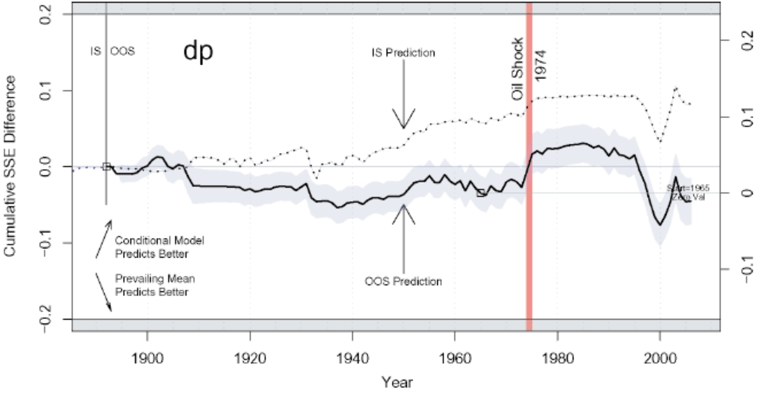
Desirable Predictor Characteristics?

A well-specified signal would inspire confidence in a potential investor if it had:

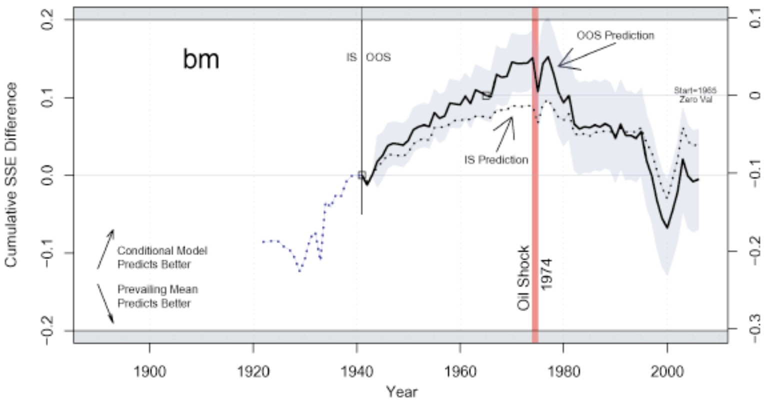
1. both significant IS and OOS performance;
2. an irregular by upward drift;
3. an upward drift not just in one short or unusual sample period;
4. an upward drift that remains positive over the most recent decades.

Bad News for D/P

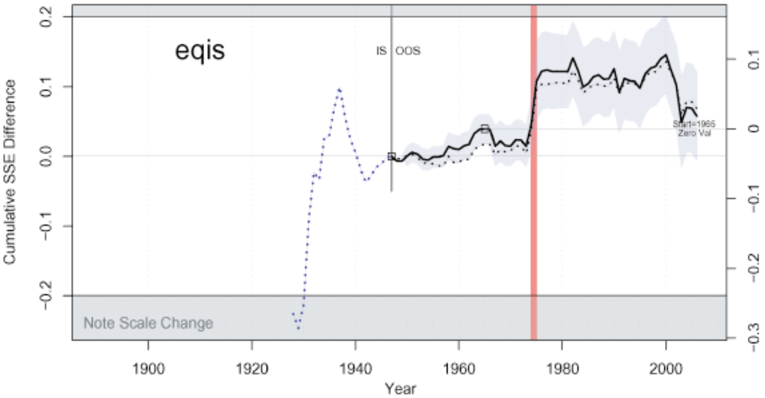
Blue Band shows statistical significance at the 95% level.



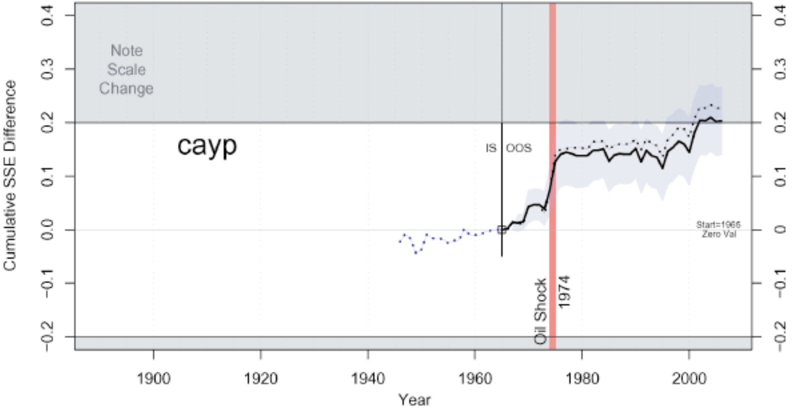
Bad News for B/M



Better OOS Prediction: The Equity Share



Better OOS Prediction: CAY



Next week

- Predicting returns with big data techniques.
- Current literature.
- Introduction to your assignment.

References

1. Cochrane, John H. Asset pricing: Revised edition. Princeton university press, 2009.
2. Welch, Ivo, and Amit Goyal. "A comprehensive look at the empirical performance of equity premium prediction." The Review of Financial Studies 21.4 (2008): 1455-1508.