Return Prediction for Dynamic Trading Strategies

Campbell R. Harvey
Duke University and NBER
1. Introduction

- The most important components in any trading strategy are the forecasts of the asset returns.

- That is, mistakes in the expected returns inputs will have a much larger impact that mistakes in the risk inputs.

- Hence, it is critical to get the best possible forecasting methodology for the asset returns.
Two types of prediction:

Time Series--


Cross sectional--

The cross-section of returns is predicted for next period based on the cross-section of attributes in the previous period. Exercise predicts winners and losers. Examples: Ferson and Harvey (1994).
1. Introduction

Goals of Presentation:

1. Demonstrate contribution to the cross-sectional prediction models.

2. Review models of time-series prediction.

2. Cross-sectional predictability

Example:

Suppose we wanted to forecast the returns on 2000 NYSE stocks next quarter.

- The predictor variable might include the price-book value ratio of each stock.

Use a linear regression model – but now we use only two quarters of data.

\[ \text{Returns}_{93Q4} = c_0 + c_1 (P/B)_{93Q3} + \text{error} \]

- Note that the predictor variable is lagged.

- We can use a host of additional variables such as
  1. Other accounting information.
  2. Sector dummy variables.
  3. Past performance measures (to capture reversals).
  5. Public insider trading information....
2. Cross-sectional predictability

Out-of-sample forecasts can then be formed:

⇒ With data for \( P/E \) for 2000 firms in 93Q4, we can forecast the 94Q1 returns using the previously estimated coefficients, \( c_0, c_1 \).

- In practice, up to 50 fundamental variables are used in the exercise.

- Also, in practice, the \( c \) coefficients may change through time.

- To implement these models, it is crucial to somehow forecast the "\( c \)" coefficients.

- Previous research provides little guidance as to how to forecast the \( c \) coefficients. Many ad hoc procedures are popular in practice.
Enter Fama and French (1992).

They run cross-sectional regressions with both the beta (implied by asset pricing theory) and fundamental attributes.

- They find "no significant relation" between beta and the cross-section of expected returns when fundamental attributes are also considered in the regression.

General interpretation:

CAPM dead?
2. Cross-sectional predictability

Here’s our contribution:

Asset pricing theory is very explicit:

*The cross-section of expected returns is related to the securities’ betas and the (common) economy-wide risk premiums.*

- Why not let the risk loadings be a function of the firm attributes.

- This approach may lead to improved cross-sectional prediction models.
2. Cross-sectional predictability

Suppose we are considering a one-factor CAPM world where the factor is the market return.

Suppose asset $i$’s beta with respect to the market is predicted by the price to book ratio ($P/B$).

Let’s write the model:

$$\beta_{i,t+1} = b_{0i} + b_{1i}(P/B)_{i,t}.$$  

Notice that the lagged price to book is revealing information about next period’s beta.

$\Rightarrow$ This is called a “conditional beta” model.
What does the theory say about the time-series of returns:

The usual model for asset $i$’s excess returns is

$$r_{i,t+1} = \alpha_i + \beta_i r_{m,t+1} + \text{error}.$$  

Let’s plug in our model for beta:

$$r_{i,t+1} = \alpha_i + [b_0 + b_1(P/B)_{it}] r_{m,t+1} + \text{error}$$

$$r_{i,t+1} = \alpha_i + b_0 r_{m,t+1} + b_1[(P/B)_{it}] \times [r_{m,t+1}] + \text{error}$$
2. Cross-sectional predictability

This is similar to the standard market model equation — however, there are a number of differences.

1. The beta is allowed to change through time as a function of the price to book ratio.

2. An extra term enters the equation — the product of the market return and the price to book ratio.

3. The $b_1$ coefficient tells us the response of the firm’s beta to changes in the book to price ratio.

- Note, that the relation between risk and price to book is assumed constant.
2. Cross-sectional predictability

Implications:

What we have learned from the time-series regressions can be applied to cross-sectional prediction.

Let's return to the CAPM example with price to book as an attribute. Consider the time-series regression:

\[ r_{i,t+1} = \text{intercept} + b_1 r_{m,t+1} \times \left( \frac{P}{B} \right)_{it} + \text{error} \]

Like \( c_1 \)

Important difference:

- \( c_1 = b_1 r_{m,t+1} \) is asset specific!

- This is why it is difficult to forecast the cross-sectional coefficient \( c_1 \). In addition, arbitrary rescaling of attributes can mess up properties.
2. Cross-sectional predictability

Proposed solution:

In cross-sectional prediction, regress returns on:

\[ b_{1i}(P/B)_{it} \]

We get \( b_{1i} \) from asset-specific time-series regressions.

\[ \Rightarrow b_{1i}(P/B)_{it} \text{ is the "optimally filtered attribute"} \]

\[ \Rightarrow \text{Filtered attribute should improve cross-sectional prediction.} \]
2. Cross-sectional predictability

Specification of the Beta Function:

Important to get the risk function properly specified.

Specifications can be explored with GMM system. Suppose we wanted to test whether the squared price-to-book ratio enters the beta function.

\[ \beta_{i,t+1} = b_{0i} + b_{1i} P/B_{i,t} + b_{2i} P/B_{i,t} \]

The system can be estimated:

\[ u_t = f - Z \delta \]
\[ \eta_t = u_t' u_t / \beta_i - u_t r_{it} \]

where \( f \) represents a set of factor returns.

The hypothesis \( b_{2i} = 0 \) can be tested – without imposing an asset pricing model.
The goal is to forecast the change in price: $P_t$ to $P_{t+1}$.

To take the stand that the return is unpredictable implies:

1. No ability to forecast the economy and how it influences a particular security’s cash flows.

2. No ability to forecast the economy’s impact on the risk of the security.

3. No ability to forecast economy-wide rewards for risk.

Most now agree that at least one of these items is predictable.

⇒ The key is translating this potentially small degree of predictability into successful trading strategies.
3. Time-series predictability

A Simple Example:

We will assume a linear regression model is being considered by a U.S. fund manager that chooses to allocate into two portfolios: the S&P 500 and money market instruments (T-bill).

We will examine three strategies:

1. Unconditional asset allocation with average risk aversion.

2. Unconditional asset allocation with higher than average risk aversion.

3. Conditional asset allocation with average risk aversion.

Each of these strategies will produce a different allocation. They will be tracked over the January 1970–September 1991 period.

This is a ‘simple’ example because only two asset classes are considered.
3. Time-series predictability

**Strategy 1:** (Buy-hold)

This strategy implies a buy and hold equities portfolio. Unconditionally, the average equity return is much higher than the average money market return. Hence, a manager with average risk aversion will hold 100% equities.

**Strategy 2:** (90/10)

This strategy also implies a buy and hold portfolio. However, to lower the risk of the portfolio, the manager holds a combination of money market and equities. For our example, we will assume a 90% equity and 10% money market composition.

**Strategy 3:** (Conditional)

This strategy will likely produce a portfolio that switches among the two asset classes depending upon the forecasts of the equity returns.
If the forecasted equity returns are always above the money market rate, the Strategy 3 will be identical to Strategy 1.

*It is in this sense that the Unconditional Asset Allocation is a special case of the Conditional Asset Allocation. The strategies will be identical when it is impossible to accurately forecast the equity returns.*
The Details of Strategy 3:

A linear regression is used to forecast the S&P 500 return.

The regression equation is:

$$SPRET_t = \alpha_0 + \alpha_1 \text{BILL}_{t-1} + \alpha_2 \text{SPDIV}_{t-1} + \alpha_3 \text{3-1BILL}_{t-1} + \alpha_4 \text{Baa-Aaa}_{t-1} + \varepsilon_t$$

where

1BILL = Yield on one month U.S. Treasury Bill,
SPDIV = Annual dividend yield on S&P 500 stock index,
3-1BILL = Return spread on 3 and 1 month U.S. T-bills,
Baa-Aaa = Yield spread on U.S. Baa and Aaa rated bonds,
\varepsilon = Regression error (unexpected part of the return).

Notice that this is a forecasting equation. The conditioning information (1BILL, SPDIV, 3-1BILL and Baa-Aaa) are available at time $t - 1$.

These variables are used to forecast the next period returns for time $t$. 
3. Time-series predictability

Implementing Strategy 3:

If this equation is estimated with monthly data over the 1947:2–1991:9 period, the $R^2$ is 6.9%.

The R-square measures the precision of our predictions.

- An $R^2=100\%$ means our predictions are perfect.
- An $R^2=0\%$ implies our predictions are equal to $\alpha_0$ – which is just the average equity return.

An $R^2$ of zero, implies that the unconditional strategy (Strategy 1 – buy and hold) is the best.

Hence, the worst case scenario for our forecasting model implies the unconditional asset allocation strategy.
Implementing Strategy 3:

Although the $R^2$ is low, it is best to evaluate our forecasting model using dollars rather than statistical measures.

It is important to produce out-of-sample forecasts.

METHOD

1. Estimate coefficients using returns data through the end of December 1969 (and information variables through November 1969).

2. Apply coefficients to the information variables known at the end of December 1969 and calculate a forecast of the stock return for January 1970. Trade based on the forecasted return.


4. Use new coefficient and information variables known at the end of January 1970 to form the forecasted return for February 1970.

5. Repeat this procedure every month.
3. Time-series predictability

The Trades:

⇒ If forecasted equity return is greater than money market rate, enter the equity market with 100% allocation.

⇒ If forecasted equity return is less than the money market rate but greater than zero, then enter the money market with a 100% allocation.

⇒ If forecasted equity return is less than zero, then short equities with -100% weight. Assume that 50% non-interest bearing margin is required.

Evaluate position at the end of each month.
3. Time-series predictability

Evaluation:

<table>
<thead>
<tr>
<th>Performance Measure</th>
<th>Strategy 1 Buy-Hold</th>
<th>Strategy 2 90/10</th>
<th>Strategy 3 Conditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total profit</td>
<td>233.87%</td>
<td>226.40%</td>
<td>346.58%</td>
</tr>
<tr>
<td>Monthly profit</td>
<td>0.90%</td>
<td>0.86%</td>
<td>1.33%</td>
</tr>
<tr>
<td>Monthly std. dev.</td>
<td>4.68%</td>
<td>4.21%</td>
<td>4.10%</td>
</tr>
<tr>
<td>Annual profit</td>
<td>10.76%</td>
<td>10.41%</td>
<td>15.94%</td>
</tr>
<tr>
<td>Annual std. dev.</td>
<td>16.22%</td>
<td>14.59%</td>
<td>14.21%</td>
</tr>
</tbody>
</table>
3. Time-series predictability

- Conditional strategy greatly enhances returns and lowers overall risk.

- The benefits are impressive given the fairly low $R^2$.

- Return enhancement would not be eliminated by reasonable transactions costs.

- Potential for even higher returns if filter is applied to forecast. I.e., if forecasted equity return is trivially above the T-bill rate do not enter the equity market.

- My empirical work clearly documents predictability in U.S. and International stock and bond markets and, especially, in new emerging stock markets.
4. Nonlinear predictability

Most of the work has focused on linear prediction.

Nonlinearities and changing projection coefficients often proxied by rolling window estimation.

Some new techniques:

1. Nonparametric density estimation.
2. Neural Nets/Genetic Algorithms
3. Entropy encoding (Data Compression).
4. Nonlinear predictability

In the computer science literature, entropy-based encoding is used for the problem of data compression.

There are distinct links between the data compression task and a returns forecasting exercise.

- Both are designed to learn from patterns.

- It will turn out that the recent neural nets and genetic algorithms are special cases of the entropy encoding.

Let’s explore some of the basic concepts.
Entropy is a measure of a series’ predictability.

\[ \text{Entropy} = - \sum p_y \log(p_y) \]

where \( p \) is the probability and we sum over all the events.

Example:

1. Two-headed coin flip. Perfectly predictable. \( \log(1) = 0 \) so Entropy=0.

2. N-sided dice. Impossible to predict. \( -\log(1/N) \to \infty \) as \( N \to \infty \).
4. Nonlinear predictability

How to get the probabilities:

1. Sample 150 data points

2. Sort data (lowest to highest).

3. Draw equal area rectangles around each two adjacent points, $y_1, y_2$.

4. $\sum A_i = 1$ by construction. Each box is equal area. Hence, $A_i = 1/150$ for each $i$.

5. Area of rectangles is: $1/N = P_y(y_1 - y_2)$.

6. Solve for $P_y$. Do this for each rectangle. Calculate entropy.
4. Nonlinear predictability

Approaches

- Single training period, 0-1 encoding (rise/fall).
- Multiple training periods, 0-1 encoding.
- Uniform quantization (rounded % returns) – Single tree.
- Uniform quantization – Multiple trees.
- Scalar Quantization – Single trees.
- Scalar Quantization – Multiple trees.
- Vector Quantization – Single trees.
- Vector Quantization – Multiple trees.
4. Nonlinear predictability

Performance tracking with:

Compression ratio:

Ratio is the average internal path length (root to leaf) divided by the number of leaves.

⇒ When ratio equals one, no predictability.

⇒ Linked asymptotically to the entropy measure.
4. Nonlinear predictability

0-1 Tree (binary)

Start with string: \{1 0 1 1 0 0 0 1 1 1 0 1\}
Previous two strategies are not ideally suited for capturing the magnitude of the expected returns.

Scalar quantization

1. Sample some of the data to create rectangles as before.

2. Value of box will be the centroid \((y_1 - y_2)/2\).

3. Assign all data to boxes.

4. Build n-ary tree.

5. Forecasts will calculate expected values using the probabilities \(p_y\) which can be calculated from [1].
4. Nonlinear predictability

Scalar strategy only considers past values of series. Vector strategy takes conditioning information into account.

Vector quantization (one conditioning variable)

1. Sample some of the data to create returns-interest rate scatter.

2. Using Vornoi diagram create pillars.

3. Proceed as before.
4. Nonlinear predictability

Preliminary results are very encouraging.

Predictability in many daily stock markets.

• More predictability in emerging markets. This is consistent with the results in Harvey (1994) using linear models.
Predicting Global Industry Returns
U.S. $ Returns

Data through 1991:09
Predicting Developed Market Returns
U.S. $ Returns

R-square

Country
Aus Aut Bel Can Den Fra Ger HK Ita Jap Net Nor S/M Spa Swe Swi UK US

Data through 1991:09
Predictability of Equity Returns

R-square

Country

Common Information

Local Information

U.S. $ Returns, Full Sample