Goldman Sachs
January 1998

Cross-Sectional Prediction for Dynamic Trading Strategies

Campbell R. Harvey
Duke University and NBER
1. Introduction

There are three streams to my research:

- Implications of time-series predictability of asset returns (lower frequency) for asset pricing and asset allocation

- Cross-sectional prediction (bottom-up strategies).

- High frequency forecasting.
Time-series prediction:


- Simple OLS models estimated on monthly data.
- Nonparametric density estimation.
- Regime switching models with disturbances which follow mixtures of normals.
- Generalization of GARCH methods to conditional skewness and kurtosis.
3. Lower frequency cross-sectional prediction

Predict which assets will outperform and underperform with purely quantitative methods.

- Popular business these days: BNP/Cooper Neff Advisors supposedly trade 5% of the Big Board volume.

- Popular quantitative products being brought to the market that allow many investors to access quantitative models (e.g. FACTSET’s ALPHATESTING).

⇒ My research deals with some of the most pressing problems in implementing these models.
4. The details of 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.

\[ \text{Returns}_{97Q4} = c_0 + c_1 (P/B)_{97Q3} + \text{error} \]

- Note that the predictor variable is lagged.

- We can use a host of additional variables such as
  1. Other accounting information.
  2. Past performance measures (momentum).
  3. Volume, volatility, correlation, ....
  4. Public insider trading information....
4. The details of cross-sectional predictability

Out-of-sample forecasts can then be formed:

⇒ With data for \( P/E \) for 2000 firms in 97Q4, we can forecast the 98Q1 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.

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?
Ferson and I have proposed the following 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.
4. The details of 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.

⇒ This is called a “conditional beta” model.
4. The details of cross-sectional predictability

What does the theory say about the time-series of returns:

The usual model for asset i’s excess returns is

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

Let’s plug in our model for beta:

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

\[ r_{i,t+1} = \alpha_i + b_0 r_{m,t+1} + b_1 [(P/B)_{it}] \times [r_{m,t+1}] + \text{error} \]
4. The details of 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.
4. The details of 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:

\[ r_{i,t+1} = \text{intercept} + b_{1i}r_{m,t+1} \times (P/B)_{it} + \text{error} \]

Like \( c_1 \)

Important difference:

- \( c_1 = b_{1i}r_{m,t+1} \) 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.
4. 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.

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

\[ \Rightarrow \text{Filtered attribute should improve cross-sectional prediction.} \]
4. The details of 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.

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

The system can be estimated:

\[
\begin{align*}
\mathbf{u}_t &= \mathbf{f}_t - \mathbf{Z}_{t-1} \delta \\
\eta_t &= \mathbf{u}_t' \mathbf{u}_t \beta_{i,t} - \mathbf{u}_t r_{it}
\end{align*}
\]

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

The hypothesis \( b_{2i} = 0 \) can be tested – without imposing an asset pricing model.
5. Implementing cross-sectional predictability

Recent 1997 working paper with Ferson: "Conditioning Variables and the Cross-Section of Stock Returns"

- We re-examine the Fama and French model.

- When we allow for time-series conditioning information, we are able to reject their model for 25 of the 25 portfolios that they use.

- We are the first to integrate time-series prediction and cross-sectional prediction.
Basic idea of the Ferson-Harvey paper:

- Use time-series methods to predict asset returns.
- Use out-of-sample time-series forecasts, as "additional factor" in cross-sectional prediction model.
- This new factor is very powerful. It has the ability to "knock-out" a number of standard accounting factors.
- The Fama-French type of model is strongly rejected.
5. Implementing cross-sectional predictability

Implications:

- We have only tested this model on U.S. data with monthly frequency.

- Our preliminary analysis suggests that our time-series measure is similar to a momentum like factor – but it is specified in a much different way that the usual approach (e.g. Carhart).

- Our model might work at higher frequency (daily, weekly). But it has not been tested.

- Need to be aware of the challenge that the infrequent trading problem poses.
6. 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).
6. 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.
6. Nonlinear predictability

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. \(-\log(1/N) \to \infty \text{ as } N \to \infty\).
6. 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.
6. 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.
- Glodjo-Harvey Proprietary Quantization Approaches.
6. 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.
6. Nonlinear predictability

0-1 Tree (binary)

Start with string: \{1 0 1 1 0 0 0 1 1 1 0 1\}
6. Nonlinear predictability

Implementation:

- Track compression ratio. If it drops below a prespecified threshold then retrain. This is the case of multiple trees.

- In forecasting, if you end up at leaf, adjust the prefix by dropping the first observation and tracing new path.

- Straightforward to adjust to n-ary tree. Round returns to percentage points (30-ary tree).
6. Nonlinear predictability

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 \(((x_1 - x_2)/2)\).

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].
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. Proceede as before.
6. Nonlinear predictability

Results for high frequency futures very encouraging.

Has not been applied to equity markets.

- Most likely to work better in non-U.S. markets.