Asset Management: Advanced Investments

Investment Strategy Development

Alexandre Ziegler

University of Zurich

Fall 2020

Slides for 28th Sept and 5th Oct

Overview

Introduction

- In this first part of the course we will consider the most common methods for investment strategy development and how to deal with the main issues that arise in the development process.

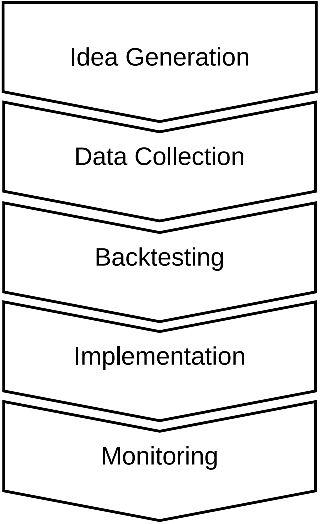
- These aspects will be useful to keep in mind when we consider common investment strategies later in the semester.

- We will cover a lot of material, but many of the topics have been covered in other finance and statistics classes.
    - If you have prior knowledge, these notes consolidate the most important things in one place.
    - If you don't, our goal is to make you aware of the main issues. You can always come back to these notes and gather additional literature when you need to.

- Strategy development can become very tedious if you try to do everything precisely. These notes therefore also provide shortcuts to allow you to assess whether a particular aspect is important or not when analyzing a given strategy.

Overview of the Strategy Development Process

Goal: Obtain the highest possible return for a given level of risk.

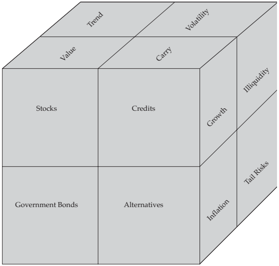


Overview

Idea Generation: Sources of Returns

- Finding ideas is the hard part of strategy development.
- Conceptually one can view returns as coming from three sources:

1. **Asset classes** (equities, government bonds, credit, real estate, commodities...)
2. **Strategy styles** (value, carry, trend/ momentum, volatility).
3. Underlying **risk factors** driving portfolio returns (growth, inflation, illiquidity, and tail risks).



Source: Ilmanen (2012).

- When devising an investment strategy, you are effectively taking exposure to asset classes, strategy styles, and risk factors.
- Put differently, before devising a strategy, you should think about **where your returns are coming from** (e.g. a risk premium, timing ability, a market anomaly) and **what kinds of risks you are willing to bear**.

Idea Generation: Sources of Returns

As an example, here are the sources of returns of some common strategies:

- **Constant allocation strategies**:
  - Strategies with a constant allocation to asset classes earn the risk premia on these asset classes over time.
  - The allocation depends on the (unconditional) risk premium per unit of risk, the level of risk, and the timing of losses.

- **Timing strategies**:
  - Strategies that rotate among asset classes try to be invested in a particular asset class when its risk premium per unit of risk is high.
  - The size of the allocation depends on conditional risk premia and risk estimates that are based on indicators / predictors.

- **Selection strategies**:
  - Selection strategies usually take long and short positions in assets within an asset class. They typically do not target the risk premium of the asset class but the returns provided by strategy styles.
  - The size of the allocation depends on the strategy's return distribution and how it relates to that of other investments.

Idea Generation: Starting Points and Desirable Properties

- Here are a few common starting points that you can use to find ideas:
  - Newspapers, scientific papers, blogs;
  - Study of successful investors;
  - Observation of market behavior;
  - Analysis of central bank/government policies;
  - Analysis of regulations.
- Ideally your strategy should be:
  - Simple;
  - Robust;
  - Easy to implement;
  - Exposed to risks that are easy to understand.

Idea Generation: Sanity Check

Once you have a candidate strategy, a sanity check is useful:

- Do you have an edge? What is it?
- What data do you need to analyze the idea? Can you get it?
- Are instruments available to implement the idea?
- How large are the transaction costs for these instruments?
- How often would you need to monitor the strategy and trade? Is that feasible?
- Would you be able to withstand drawdowns?
- Who is on the other side of your trades?

Overview

Data Collection

- To avoid going back and forth, think through what analyses you want to perform to test a proposed strategy **before** downloading the data.
- The first decision to make is what database(s) to use:
  - **General data sources**: Bloomberg, Datastream, Yahoo Finance
  - **Specialized sources**: CRSP, Compustat, Optionmetrics, IBES
- It is usually a good idea to first download data for a few securities and review it to make sure it is what you need. Once you are comfortable with what you get, download (slightly) more data than strictly needed to avoid having to do it multiple times.
- When selecting and collecting data you should:
  - Check how **dividends and interest** are accounted for and futures contracts rolled over (this will be discussed in detail as part of return computations);
  - Beware of **biases** (the main issues and solutions will be discussed as part of the principles of strategy development).

Data Collection: Understanding and Checking Data

It is essential to understand and check your data before using it:

- **Understanding your data**:
  - When is the data available (is the reference date also the publication date)?
  - What exactly is in a given field?
  - Is the field raw or calculated?
  - If calculated, how? Is it based on a model, and if so what are the assumptions?
  - Are prices adjusted for splits, dividends, rollovers? Do bond prices include accrued interest or not?
  - Are prices trade prices, quotes, or model prices? From an exchange or OTC?
  - What do open and close prices really represent?
- **Checking your data**:
  - Are there outliers or implausible values? How can they be cleaned up?
  - How are missing values coded and how frequent are they? How will you deal with them?

Overview

Computing Returns: Ordinary & Log Returns

- Since the goal is to obtain the highest possible return for a given level of risk, one needs to know how to measure returns.
- One distinguishes ordinary returns R_t and log returns r_t.
- Consider an asset n whose price at time t is $P_{n,t}$ and that pays interest or dividends (net of any taxes to simplify the notation) $D_{n,t}$.
- The **ordinary return** on the asset is computed as:

$$R_{n,t} = \frac{P_{n,t} + D_{n,t}}{P_{n,t-1}} - 1 \ . \qquad (1)$$

  Intuition: You paid $P_{n,t-1}$ for the asset at time $t-1$. At time t you have $P_{n,t} + D_{n,t}$. The gain per unit of invested capital is $R_{n,t}$.
- **Log returns** are given by:

$$r_{n,t} = \ln\left(\frac{P_{n,t} + D_{n,t}}{P_{n,t-1}}\right) = \ln(1 + R_{n,t}) \ . \qquad (2)$$

Computing Returns: Overview

Computing Cumulative Returns

- Suppose you invest an amount V_0 in an asset or portfolio at time 0. Each period t you earn a return of R_t or r_t.
- After T periods, you have:

$$V_T = V_0 \prod_{t=1}^{T} (1 + R_t) \tag{3}$$

$$= V_0 \exp\left(\sum_{t=1}^{T} r_t\right) . \tag{4}$$

- Consider **cumulative returns** $R_{0,T}$ and $r_{0,T}$. Log returns are time-additive but ordinary returns are not:

$$R_{0,T} = \frac{V_T}{V_0} - 1 = \prod_{t=1}^{T} (1 + R_t) - 1 \neq \sum_{t=1}^{T} R_t \ , \tag{5}$$

$$r_{0,T} = \ln\left(\frac{V_T}{V_0}\right) = \sum_{t=1}^{T} r_t \ . \tag{6}$$

Computing Average Returns

- For **ordinary** returns, the **average return per period** is the **geometric** mean:

$$\overline{R}_{0,T} = \left(\frac{V_T}{V_0}\right)^{\frac{1}{T}} - 1 = \left(\prod_{t=1}^{T}(1+R_t)\right)^{\frac{1}{T}} - 1 \ . \tag{7}$$

- For **log** returns, the **average return per period** is the **arithmetic** mean:

$$\overline{r}_{0,T} = \frac{1}{T}\ln\left(\frac{V_T}{V_0}\right) = \frac{1}{T}\sum_{t=1}^{T} r_t \ . \tag{8}$$

- The arithmetic mean of ordinary returns over multiple periods doesn't tell you if you are making money or not.

Computing Returns: Overview

Computing Returns: Assessing Profitability

Here are a few important things to keep in mind when assessing the profitability of investing in an asset or portfolio:

- **For realized returns:**
  - *Ordinary returns*:
    - If your arithmetic average return is zero you lost money (unless the return was zero each period).
    - If your geometric average return is zero you are flat.
  - *Log returns*:
    - If your (arithmetic) average log return is zero you are flat.
- **For expected returns (assuming random and uncorrelated returns):**
  - *Ordinary returns*: If your expected one-period return is zero then:
    - Over a single period on average you are flat.
    - Over multiple periods on average you are flat.
  - *Log returns*: If your expected one-period log return is zero then:
    - Over a single period on average you make money.
    - Over multiple periods on average you make money.

Computing Returns: Assessing Profitability

Summarizing:

Return that is zero	Profit over one period	Profit over multiple periods
Arithmetic average realized one-period return	0	–
Geometric average realized one-period return	0	0
Arithmetic avg. realized one-period log return	0	0
Expected one-period return	0	0
Expected one-period log return	+	+

Computing Returns: Overview

Computing Returns: Time Scaling

- What is the relation between average returns at different frequencies?
- For example, suppose you have a time series of monthly returns. What is the average annual return $\overline{R}_{0,T}^{\text{annual}}$?
- Let m be the number of periods in a year. The **average annual return** solves

$$\left(1+\overline{R}_{0,T}^{\text{annual}}\right)^{\frac{T}{m}} = \prod_{t=1}^{T}(1+R_t) \ , \tag{9}$$

yielding

$$\overline{R}_{0,T}^{\text{annual}} = \left(\prod_{t=1}^{T}(1+R_t)\right)^{\frac{m}{T}} - 1 \ . \tag{10}$$

- The **average annual log return** is

$$\overline{r}_{0,T}^{\text{annual}} = \frac{m}{T}\sum_{t=1}^{T} r_t \ . \tag{11}$$

Computing Returns: Overview

Computing Portfolio Returns

- Generally you will be interested in the **overall return on your portfolio**.
- Suppose we have N assets, $n = 1, \ldots, N$. Let $w_{n,t}$ denote the portfolio weights, i.e. the fraction of wealth invested in asset n at time t.
- The ordinary return and log return of the portfolio in period $t+1$ are:

$$R_{p,t+1} = \sum_{n=1}^{N} w_{n,t} R_{n,t+1} = \mathbf{w}_t' \mathbf{R}_{t+1} \ , \tag{12}$$

$$r_{p,t+1} = \ln(1 + R_{p,t+1}) \ . \tag{13}$$

- Note that ordinary returns aggregate in the cross-section, whereas log returns do not. We saw earlier that log returns aggregate over time but ordinary returns do not. Hence,
  - Whenever you want to compute the **one-period** return of a **portfolio** of assets, it is easier to work with **ordinary returns**.
  - Whenever you want to compute the total or average return over **multiple periods**, it is easier to work with **log returns**.

Computing Portfolio Returns: Riskless Asset Position

- Sometimes people include the riskless asset in the N assets, sometimes they treat it as asset 0.
- Both approaches work fine as long as one is consistent; the formulas are just slightly different:
  1. If you choose to **include the riskless asset in the N assets**, then portfolio returns are given by (12), i.e. one has

$$R_{p,t+1} = \mathbf{w}_t' \mathbf{R}_{t+1} \ . \tag{14}$$

     If all the assets are funded, the portfolio weights will sum to one, $\mathbf{w}_t' \mathbf{1} = 1 \ \forall t$.
  2. If you choose to **let the riskless asset be asset 0**, portfolio returns are given by

$$R_{p,t+1} = \mathbf{w}_t' \mathbf{R}_{t+1} + (1 - \mathbf{w}_t' \mathbf{1}) R_{f,t+1} \ , \tag{15}$$

     where $R_{f,t+1}$ denotes the riskless return on an investment made at t and maturing at $t+1$ and $w_{0,t} = (1 - \mathbf{w}_t' \mathbf{1})$ is the fraction of wealth held in the riskless asset (with negative values corresponding to borrowing/leverage).
- It is important to keep track of the size of the position in the riskless asset.

Computing Returns: Overview

Computing Returns on Long-Short Portfolios

- Many investment strategies involve **long-short portfolios**. For example, a momentum strategy will buy assets that have performed well in the past and short assets that performed poorly in the past.
- To understand how to compute returns on such portfolios, remember what happens when you short a security:
  1. You borrow the security from someone that has it.
  2. You sell the security on the market.
  3. Your account is credited with the proceeds from the short sale.
  4. You have a liability to return the security to the lender and to pay him a lending fee as well as the interest and dividends from the security.
  5. At some point in the future, you buy back the security and return it to the lender.

Computing Returns on Long-Short Portfolios

- Suppose that you fund your account with 100 in cash and then short stocks for 50 and buy stocks for 80. Your account would then look as follows:

Assets		**Liabilities**	
Stocks held long	80	Stocks held short	50
Cash	70	Account equity	100
Total	150	Total	150

- The total return on the portfolio (i.e. the return on the account equity) will have the following components:
  + Return on the stocks held long;
  + Interest on the cash;
  − Return on the stocks held short;
  − Stock lending fees for the stocks held short (hard to estimate).

Computing Returns on Long-Short Portfolios

- Formally, letting $\mathbf{w}_t$ denote the percentage position in the different assets (positive for longs, negative for shorts) and letting the riskless asset be asset 0, the return on the portfolio is given by

$$R_{p,t+1} = \mathbf{w}_t'\mathbf{R}_{t+1} + (1 - \mathbf{w}_t'\mathbf{1})R_{f,t+1} + \sum_{n=1}^{N} \min[w_{n,t}, 0] L_{n,t+1} \ , \qquad (16)$$

  where $L_{n,t+1}$ is the percentage stock lending fee from t to $t+1$.

- Thus, one can apply the usual portfolio return computation formula up to an **adjustment for stock lending fees**.
- People often neglect these fees because they are hard to estimate precisely, in which case the usual portfolio return computation formula can be applied without changes.
- A better approach is to subtract an estimate of the average lending fee $\overline{L}$ times the overall fraction of the portfolio held short from the returns, i.e.

$$R_{p,t+1} = \mathbf{w}_t'\mathbf{R}_{t+1} + (1 - \mathbf{w}_t'\mathbf{1})R_{f,t+1} + \overline{L} \sum_{n=1}^{N} \min[w_{n,t}, 0] \ . \qquad (17)$$

Returns on Zero-Cost Long-Short Portfolios

- A widespread alternative approach is to keep track of the long and short legs separately and require the weights of each to sum to one. In this case one thinks of the portfolio as a **zero-cost long-short portfolio**, in which the proceeds from the shorts are used to finance the purchase of the longs.
- Letting $\mathbf{w}_L$ denote the portfolio held long and $\mathbf{w}_S$ that held short, one obtains the returns on each leg as:

$$R^L_{p,t+1} = \mathbf{w}'_L \mathbf{R}_{t+1} \ , \tag{18}$$

$$R^S_{p,t+1} = \mathbf{w}'_S (\mathbf{R}_{t+1} + \mathbf{L}_{t+1}) \ . \tag{19}$$

- The return on the zero-cost long-short portfolio is the difference in returns between the two legs:

$$R_{p,t+1} = R^L_{p,t+1} - R^S_{p,t+1} = \mathbf{w}'_L \mathbf{R}_{t+1} - \mathbf{w}'_S (\mathbf{R}_{t+1} + \mathbf{L}_{t+1}) \ . \tag{20}$$

- Note that since we have not included the return on the riskless asset, $R_{p,t+1} = R^L_{p,t+1} - R^S_{p,t+1}$ represents an **excess return**. The raw return would be $R_{p,t+1} = R^L_{p,t+1} - R^S_{p,t+1} + R_{f,t+1}$.

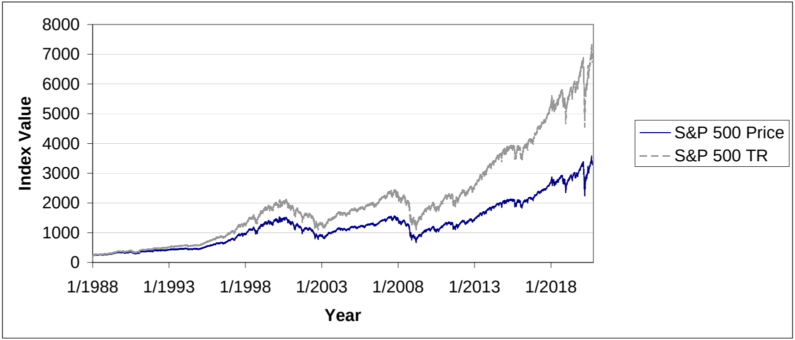
Computing Returns: Overview

Accounting for Dividends and Interest

- Recall that the return on an asset n whose price at time t is $P_{n,t}$ and that pays interest or dividends (net of any taxes) $D_{n,t}$ is given by

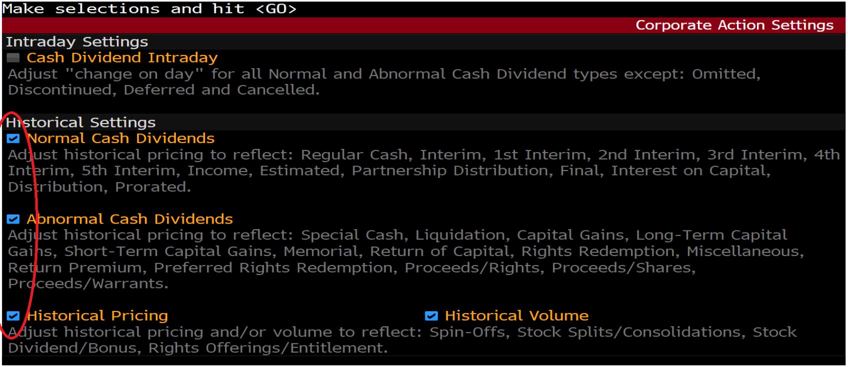
$$R_{n,t} = \frac{P_{n,t} + D_{n,t}}{P_{n,t-1}} - 1 \ . \tag{21}$$

- When computing returns, you have to make sure that you account for dividends and interest. As the chart below shows, their effect can be large.



Accounting for Dividends and Interest

- If you use Bloomberg to download your data, the **DPDF** function allows making the necessary adjustments automatically.



- These settings also affect downloads via Excel (unless you override them).
- CRSP also provides returns that are adjusted for splits and dividends, as well as an adjustment factor for prices.

Accounting for Dividends and Interest: Withholding Taxes

- An additional adjustment to portfolio returns might be needed if you are subject to withholding taxes on dividends or interest that are not creditable/refundable in full.
- The issue can be important if your portfolio has short positions because
  - On **long** positions, you receive dividends **net of withholding tax**;
  - On **short** positions, you have to pay the **full dividends** to the lender.
- Suppose that your long positions pay out 200 in gross dividends subject to 30% withholding and that your short positions have dividends of 100. Suppose there are no capital gains or losses. Your net return then looks as follows:

Dividend income	200
Dividends paid on shorts	−100
Withholding tax	−60
Profit after tax	40

- A useful **shortcut** to assess the impact of withholding taxes is to subtract the product of the dividend yield on long positions and the withholding tax rate from your returns.

Accounting for Dividends and Interest: Withholding Taxes

- Cross-border investments create an additional complexity. Provided there is a tax treaty between the two countries involved, part of the foreign withholding tax can be refunded, and part can be credited against the domestic tax. However, that credit is generally subject to a limitation.
- Assume the dividends from the previous example and suppose that your capital gains are 900 and that the tax rate on your profits is 20%. With a credit limitation, your after-tax profit will be only 780 instead of 800 with full crediting:

Case	Full crediting	Credit limitation
Dividend income	200	200
Dividends paid on shorts	−100	−100
Capital gains	900	900
Taxable income	1'000	1'000
Foreign withholding tax	−60	−60
Domestic income taxes (net of credit)	−140	−160
Profit after tax	800	780

Accounting for Dividends and Interest: Withholding Taxes

- Formally, letting V_t denote the value of the portfolio before any taxes, D_t the (aggregate) dividends paid on long positions, τ_d the domestic tax rate and τ_f the nonrefundable part of the foreign withholding tax rate, one has:

Withholding tax on long dividends	$\tau_f D_t$
Tax credit	$-\min(\tau_f, \tau_d) D_t$
Domestic income tax	$\tau_d (V_t - V_{t-1})$
Total tax cost	$\tau_d (V_t - V_{t-1}) + \max(\tau_f - \tau_d, 0) D_t$

- Expressed in terms of returns, the tax is:

$$\frac{\tau_d (V_t - V_{t-1}) + \max(\tau_f - \tau_d, 0) D_t}{V_{t-1}} = \tau_d R_{p,t} + \max(\tau_f - \tau_d, 0) \frac{D_t}{V_{t-1}} \quad (22)$$

and the **after-tax return** is:

$$R_{p,t}^{\text{net}} = R_{p,t}(1 - \tau_d) - \max(\tau_f - \tau_d, 0) \frac{D_t}{V_{t-1}} \ . \quad (23)$$

- Thus, the **drag on returns from taxation** is the sum of:
  1. the product of the percentage return before any taxes and the domestic income tax rate; and
  2. the product of the dividend yield on long positions and the part of the withholding tax rate that is neither creditable nor refundable.

Computing Returns: Overview

Portfolios with Futures Contracts

- Does the presence of **futures contracts** in a portfolio affect the return computation formula?
- A futures is an **unfunded position**. Any change in the futures price is settled in cash at the end of each trading day. Hence, the value of the contract is zero, and we can't really compute a percentage return on the investment.
- Let $R_{p,t+1}$ be the portfolio return *excluding* futures (i.e. from the N other assets), F_t the futures price and N_F the portfolio's futures exposure (number of contracts times multiplier). The change in the total value of the portfolio V_t is:

$$V_{t+1} - V_t = V_t R_{p,t+1} + N_{F,t}(F_{t+1} - F_t) \ . \tag{24}$$

- The percentage return on the portfolio *with* futures, $R^F_{p,t+1}$, is:

$$R^F_{p,t+1} = \frac{V_{t+1}}{V_t} - 1 = R_{p,t+1} + \frac{N_{F,t}F_t}{V_t}\frac{F_{t+1} - F_t}{F_t} = R_{p,t+1} + w_{F,t}R_{F,t+1} \ . \tag{25}$$

- Hence, the return on a portfolio comprising futures can be computed as usual. The difference is that **the weights need not sum to one**.

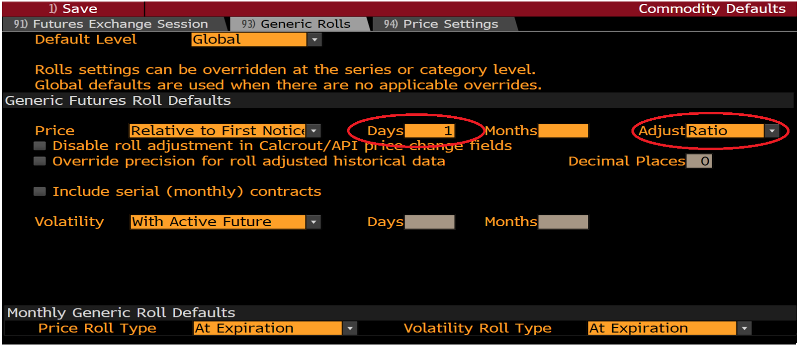
Portfolios with Futures Contracts: Rollovers

- Some care must be taken when a futures position is rolled over. The above statement on how to compute returns remains true, but one must make sure that one is using a **given** futures contract and **not mixing contracts with different maturities**.
- Consider for example the E-Mini S&P 500 futures shown below.
  - Through September 17, you would use the September contract to compute returns.
  - Starting September 18 at the latest, you would use the December contract.

Date	ESU0	Return U0 (%)	ESZ0	Return Z0 (%)
14/09/2020	3382.50		3372.25	
15/09/2020	3405.25	**0.6726**	3395.00	0.6746
16/09/2020	3389.50	−**0.4625**	3379.50	−0.4566
17/09/2020	3361.50	−**0.8261**	3351.00	−0.8433
18/09/2020	3353.60	−0.2350	3316.25	−**1.0370**
21/09/2020	(expired)	n/a	3275.00	−**1.2439**

Portfolios with Futures Contracts: Rollovers

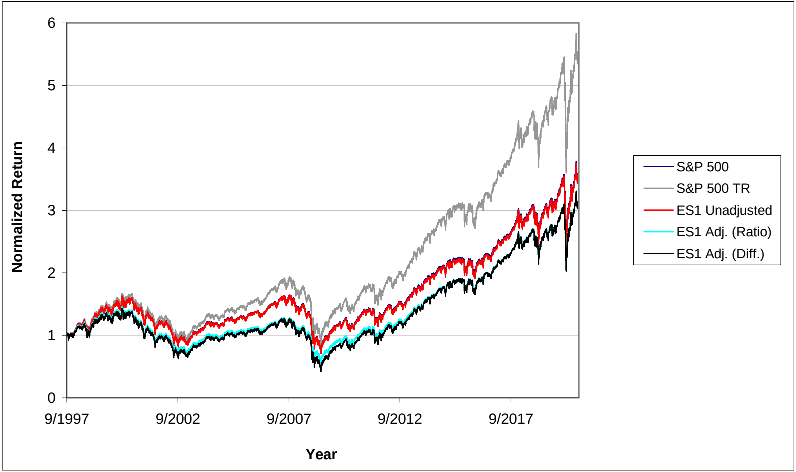
- One way to do this is to have your code figure out which contract to use.
- An easier way is to use the generic futures in Bloomberg. In our example, ES1 would be the front-month contract rolled over automatically. However, you must make sure to **adjust the prices for the rollover** using the **GFUT** function.



- The rollover adjustment is needed because futures prices differ across maturities. The price spread between two maturities reflects the difference between the riskless interest rate and the dividend yield.
- For example, for the rollover from ESU0 to ESZ0 (buy Z0, sell U0):

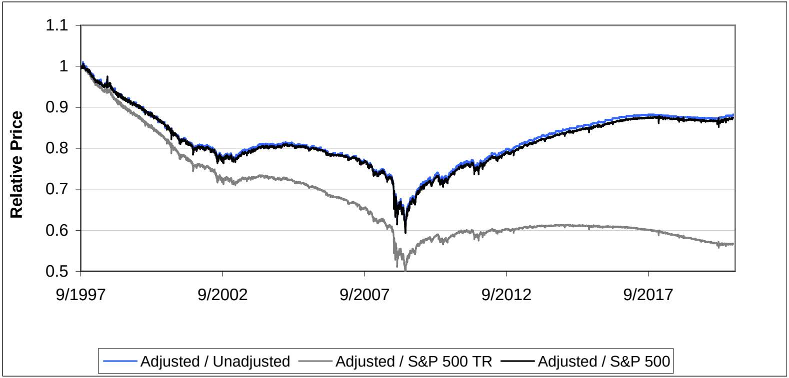
Portfolios with Futures Contracts: Rollovers

- Failure to perform the rollover adjustment will cause large errors in returns:



Portfolios with Futures Contracts: Rollovers

- It is important to note that futures returns will generally differ from the returns on both price and total return indices.



- Absent market frictions, **futures returns equal index returns including dividends minus the riskless rate**.

Proof:

- Suppose that the value of the underlying **price index** follows:

$$dS_t = (\mu_t - q)S_t dt + \sigma_t S_t dB_t \ , \tag{26}$$

  where q is the dividend yield and dB_t a Brownian motion increment. Note that the drift μ_t and volatility σ_t can be time-varying and stochastic.

- Letting r denote the riskless rate, the price of a futures maturing at T is:

$$F_t = S_t e^{(r-q)(T-t)} \ . \tag{27}$$

- Using the Ito formula, the return on the futures contract is:

$$dF_t = dS_t e^{(r-q)(T-t)} - S_t(r-q)e^{(r-q)(T-t)}dt = \frac{dS_t}{S_t}F_t - (r-q)F_t dt \ .$$

- This can be rewritten as:

$$\frac{dF_t}{F_t} = \frac{dS_t}{S_t} - (r-q)dt = (\mu_t - r)dt + \sigma_t dB_t \ . \tag{28}$$

- For a **total return index** one has $dS_t = \mu_t S_t dt + \sigma_t S_t dB_t$ and $F_t = S_t e^{r(T-t)}$; the expression for the futures return is also given by (28).

Overview

Performance Measurement: Basic Idea

- The idea of performance measurement is to relate the portfolio's realized return to its risk. For a given level of risk, the higher the realized return, the better the portfolio's performance.
- There are numerous performance measures. They differ in the kind of risk they take into account (total, systematic, or idiosyncratic risk) and in the way they control for it.
- Performance measures are usually based on **excess returns**, i.e. the return on the portfolio in excess of that on the riskless asset, $R_{p,t} - R_{f,t}$.

Performance Measurement: Overview

Computing Excess Returns

- Some care should be taken when computing excess returns. Indeed, it is not always appropriate to subtract the riskless rate.
- Whether this should be done depends on the extent to which the portfolio must be funded. Put differently, it is the answer to the question:

"If I didn't have cash and wanted to trade this portfolio, what fraction of the portfolio value would I need to borrow?"

- Here are a few examples based on that rule:

Type of portfolio	Subtract R_f?
Long-only portfolio of funded positions (stocks, bonds, ETFs, options)	Yes
Zero-cost long-short portfolio	No
Futures or forward contracts, long or short	No

- Intermediate cases are not typical but possible. For example, if the portfolio buys stocks for 100%, shorts bonds for 40% and buys futures for 30%, one should subtract 60% of the riskless rate from the returns.

Performance Measurement: Overview

Sharpe Ratio

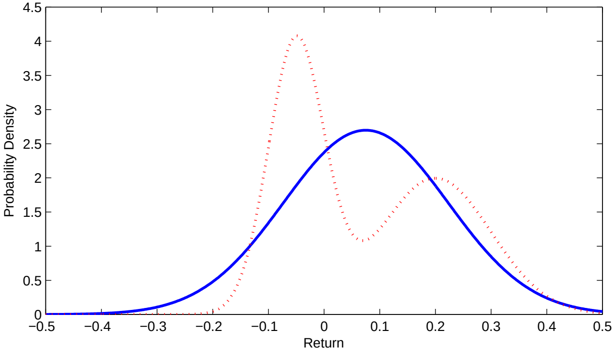
- A key performance measure is the **Sharpe ratio**. It gives the average return achieved by a portfolio in excess of the risk-free rate, $R_p - R_f$, per unit of total portfolio risk σ_p. Formally,

$$SR_p = \frac{R_p - R_f}{\sigma_p} \ . \tag{29}$$

- The higher this ratio, the better portfolio performance and the more desirable is holding a particular portfolio.
- The Sharpe ratio is an appropriate measure of performance for an investor concerned with the **full risk** of the portfolio, i.e. for an investor evaluating a portfolio that represents the majority of his investment.

Sharpe Ratio

- The Sharpe ratio has the advantage that its computation does not require a benchmark portfolio.
- It has two drawbacks:
  1. It does not distinguish whether the returns are generated by taking systematic or idiosyncratic risk.
  2. It only considers the first two moments of the return distribution. Both distributions shown below have mean 7.5% and standard deviation 15%.



Performance Measurement: Overview

Alpha

- Another important performance measure is **alpha**. A portfolio's alpha is the excess return on the portfolio over the return that would have been achieved by investing in a benchmark portfolio that has the **same systematic risk**.
- A portfolio's alpha can be estimated and its significance assessed by regressing the portfolio's excess return $R_{p,t} - R_{f,t}$ on the excess returns of the systematic risk factors $R_{j,t}$:

$$R_{p,t} - R_{f,t} = \alpha_p + \sum_{j=1}^{J} \beta_j R_{j,t} + \varepsilon_{p,t} \ . \tag{30}$$

- For alpha to be meaningful, the $R_{j,t}$s must be **excess returns on (portfolios of) traded assets**.

Alpha: Benchmark Models

Alpha is generally computed using one or several standard benchmark models:

- The **Market Model**:

$$R_{p,t} - R_{f,t} = \alpha_p + \beta_p(R_{M,t} - R_{f,t}) + \varepsilon_{p,t} \ . \tag{31}$$

- The **Fama-French 3-Factor Model**:

$$\begin{aligned} R_{p,t} - R_{f,t} =& \alpha_p + \beta_{M,p}(R_{M,t} - R_{f,t}) + \beta_{SMB,p}(R_{S,t} - R_{B,t}) \\ &+ \beta_{HML,p}(R_{H,t} - R_{L,t}) + \varepsilon_{p,t} \ . \end{aligned} \tag{32}$$

- The **Fama-French-Carhart 4-Factor Model**:

$$\begin{aligned} R_{p,t} - R_{f,t} =& \alpha_p + \beta_{M,p}(R_{M,t} - R_{f,t}) + \beta_{SMB,p}(R_{S,t} - R_{B,t}) \\ &+ \beta_{HML,p}(R_{H,t} - R_{L,t}) + \beta_{UMD,p}(R_{U,t} - R_{D,t}) + \varepsilon_{p,t} \ . \end{aligned} \tag{33}$$

- The **Fung-Hsieh Model**:

$$\begin{aligned} R_{p,t} - R_{f,t} =& \alpha_p + \beta_{M,p}(R_{M,t} - R_{f,t}) + \beta_{SMB,p}(R_{S,t} - R_{B,t}) \\ &+ \beta_{Term,p}(R_{10yTry,t} - R_{f,t}) + \beta_{Default,p}(R_{BAA,t} - R_{10yTry,t}) \\ &+ \beta_{FX,p}(R_{FXStraddle,t} - R_{f,t}) + \beta_{Cdty,p}(R_{CdtyStraddle,t} - R_{f,t}) \\ &+ \beta_{Bond,p}(R_{BondStraddle,t} - R_{f,t}) + \varepsilon_{p,t} \ . \end{aligned} \tag{34}$$

Performance Measurement: Overview

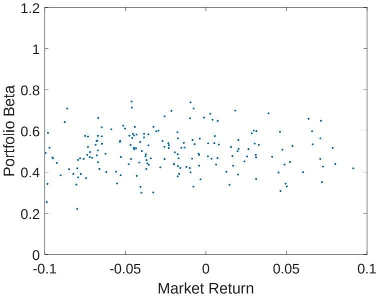
Measuring Timing Ability

- We saw earlier that timing involves increasing the exposure to an asset class when its risk premium per unit of risk is high, and reducing it when it is low.
- There are two ways to assess whether one has timing ability:
  1. plotting beta against market return; and
  2. comparing portfolio return and market return.
- We now discuss these two approaches with the case of rotating between stocks and bonds in mind. The same reasoning holds for other asset classes, interpreting "market return" as "return on the asset class" and β as "exposure to the asset class".

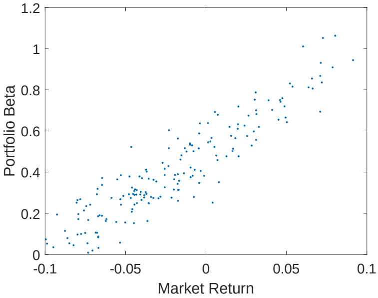
1. Plotting Beta against Market Return

- If the strategy is successful at timing the market, there should be a positive relation between the β of the portfolio and the market return (or the difference between the market return and the bond return).

No timing ability

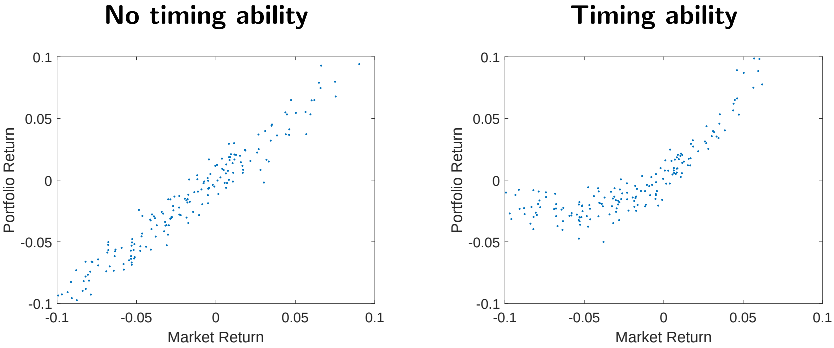


Timing ability



2. Comparing Portfolio Return and Market Return

- If the strategy is successful at timing the market, the portfolio's exposure to fluctuations in the market is greater when the market goes up than when it goes down. This yields a convex relation between the market return and the return on the portfolio.



Measuring Timing Ability

- Thus, one should be able to detect timing ability by testing for nonlinearity in the relation between the market return and the return on the portfolio.
- The **Treynor/Mazuy procedure** is to fit a quadratic curve to the performance data, i.e. run the (linear) regression

$$R_{p,t} - R_{f,t} = \alpha_p + \beta_p \left(R_{M,t} - R_{f,t}\right) + \gamma_p \left(R_{M,t} - R_{f,t}\right)^2 + \varepsilon_{p,t} \ . \qquad (35)$$

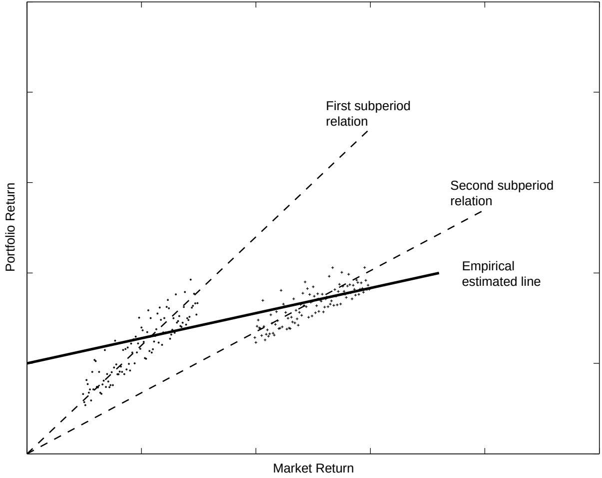
  There is evidence of timing ability if the coefficient γ_p is positive (and statistically significant).
- The **Merton/Henriksson procedure** is to recognize that the portfolio return in the timing case is similar to the payoff diagram of a long market plus call on the market position and to run the regression

$$R_{p,t} - R_{f,t} = \alpha_p + \beta_p \left(R_{M,t} - R_{f,t}\right) + \gamma_p \max(R_{M,t} - R_{f,t}, 0) + \varepsilon_{p,t} \ . \qquad (36)$$

  Again, there is evidence of timing ability if γ_p is positive.
- Unfortunately, the power of these tests is low.

Measuring Timing Ability

- Note that α_p and β_p will be distorted if the strategy involves timing but timing is not modeled explicitly when trying to measure performance.



Performance Measurement: Overview

Performance Measurement: Summary

Summarizing, when assessing a portfolio or strategy's performance, you should typically do the following:

- Plot an equity line.
- Generate a histogram of portfolio returns.
- Compute (at least) the following statistics on the returns:
  - Average
  - Standard Deviation
  - Sharpe Ratio
  - Skewness
  - Kurtosis
  - Maximum
  - Minimum
  - Factor Exposures
  - Alpha
  - Autocorrelation

Overview

Principles of Strategy Development

When developing a strategy, it is essential to keep the following in mind:

1. Always worry about statistical significance:
   - If something is not significant you actually don't have anything.
   - When assessing significance, beware of nonspherical residuals, overlapping returns, and persistent predictors.
2. Beware of overfitting / data mining:
   - If you make numerous trials you'll end up getting something that looks significant even though there is nothing.
   - If you develop a timing model that recommends few switches you might be overfitting the data. (This is generally due to persistent predictors.)
3. Beware of biases:
   - Look-ahead bias: Make sure that your strategy *only uses data that was available at the time the investment decision had to be made*.
   - Selection bias. Is your sample a random/representative draw?
   - Backfilling bias. Was data backfilled in the database you are using?
4. Account for transaction costs.
5. Implement what you tested.

Principles of Strategy Development: Overview

Statistical Significance

- Unless an effect is statistically significant, it should not be used as a basis for an investment strategy.
- We shall discuss the four main issues around statistical significance, namely:
  1. How statistical significance can be assessed when returns or regression residuals are not normally distributed or not independent and identically distributed (IID).
  2. How to deal with overlapping returns when assessing significance.
  3. What to do when the predictors used in the analysis exhibit persistence.
  4. The impact of multiple testing on statistical significance.
- We shall begin with (1). Items (2) and (3) are best dealt with in the context of timing models and will be considered later. Item (4) ("Overfitting") is dealt with in the next section.

Significance: Nonnormal and Nonspherical Residuals

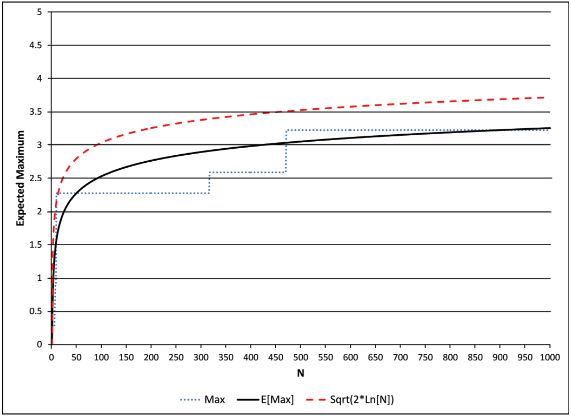
- The returns on many assets are **not normally distributed**. For example:
  - Stock returns over short periods are known to have fat tails.
  - Option returns exhibit large skewness.

  This will show up as **nonnormality of the residuals** in return regressions.
- There is evidence that **volatility is time-varying** so asset returns are not IID.
- What impact does this have on regression analysis?
  - **Nonnormal Residuals**: By the Gauss-Markov theorem, the OLS estimator is the best linear unbiased estimator (BLUE) if the errors are IID. **Normality is not required**.
  - **Non-IID Residuals**: When the residuals are not IID (correlated or heteroskedastic):
    - The **OLS estimator is still unbiased** (but inefficient; the GLS estimator is BLUE in this case; in effect GLS transforms the errors to make them uncorrelated and homoskedastic).
    - However, the **OLS standard errors are incorrect**. To assess the sampling variance of the estimates, you should use **robust standard errors**.

Principles of Strategy Development: Overview

Avoiding Overfitting

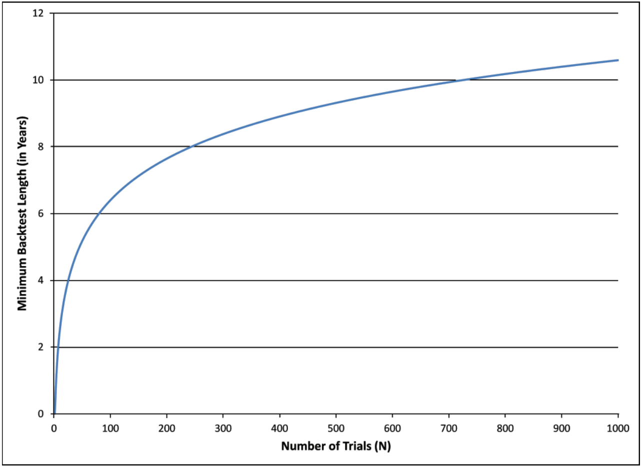
- When searching for strategies, people often try multiple possibilities on the same dataset. The "best" strategy in-sample can have a large Sharpe ratio even if there is actually nothing in the data. The chart below shows the expectation of the maximum of a set of N normal random numbers.



Source: Bailey et al. (2014).

Avoiding Overfitting

- The larger the number of trials, the longer the backtest length required to avoid overfitting.



Source: Bailey et al. (2014).

Avoiding Overfitting

The following approaches can be used to avoid overfitting:

- *Keep track of the number of trials*.
- *Keep the number of trials low*.
- *Keep the strategy relatively simple*: If your strategy is a simple "if-else", you are less likely to be overfitting than if it involves many layers of nested ifs.
- *Run out-of-sample tests*: Split the data in two pieces. One is used to develop the strategy, the other to perform an out-of-sample test of its performance.
- *Investigate the robustness of the strategy to changes in parameters*: If performance deteriorates sharply when you run a monthly model on the second trading day of the month instead of the first one, or when you change the holding period from say one calendar month to 20 trading days, your returns are likely to be due to luck. (An obvious exception is if your strategy involves calendar anomalies.)

Principles of Strategy Development: Overview

Avoiding Biases: Overview

- Data biases can cause one to mistakenly think that one found a profitable investment strategy.
- It is important to be especially careful because these biases can arise in several places:
  - In the way that assets or data items that are considered in the analysis are collected or selected.
  - In the way that the data are analyzed.
- Hence, special care must be taken when collecting, selecting, and analyzing data.
- We now describe the most common data biases and how to avoid them.

Avoiding Biases: Look-ahead Bias

- **Look-ahead bias** arises if a strategy makes use of information that was not actually available at the time the investment decision was made.
- For concreteness, suppose that you try to predict the returns on an asset using a regression

$$R_{t+1} = \alpha + \beta x_t + \varepsilon_{t+1} \; . \tag{37}$$

- To avoid look-ahead bias, you have to:
  1. Make sure that the value of x_t was really known at time t. Note that it is not sufficient that x_t relates to period t. Typical problem cases are the following:
     - *Accounting data* gets published with a lag of several months.
     - *Some financial market data* (e.g. open interest, mutual fund NAV) gets published with a lag of one trading day.
     - *Time zones*. US markets close after Asian markets. So the US return for the day is not known when Japan closes, and you can't use it to decide whether to invest in Japanese stocks at the close of that day.
     - *Data revisions*.
  2. Make sure you estimate the regression coefficients using only data through period t. If you estimate them using data through the end of your sample *they actually include information about the future*.

Avoiding Biases: Selection Bias

- **Selection bias** arises when the sample is not a random draw from the underlying population, causing assets with certain characteristics to be over- or under-represented in the sample.
- For example, suppose you backtest a strategy using hedge fund data. Since reporting to hedge fund databases is voluntary, funds with poor performance are less likely to be included in your sample.
- *Fix*: If possible, use the entire population instead of a sample.

Avoiding Biases: Survivorship Bias

- **Survivorship bias** is a form of selection bias that arises when the sample used in the analysis only includes assets that are still traded at the end of the sample period.
- Examples:
  - You backtest a strategy using the current members of a stock market index. In doing so you leave out firms that left the index in previous years.
  - You backtest a strategy using bonds that are traded at the end of your sample period. In doing so you leave out the ones that matured and the ones that defaulted.
  - You backtest a strategy using commodities for which futures contracts currently exist. In doing so you leave out contracts that were delisted.
- *Fix*: In general survivorship bias can be avoided by:
  - Using the population of all assets in a given category as universe, or
  - Using the constituents of an index at a time preceding each investment decision as universe.

Avoiding Biases: Backfilling Bias

- **Backfilling / instant history bias** arises if historical data is backfilled into a database when an asset is added to it.
- For example, when hedge fund databases add a fund, they include returns before the inclusion date.
- *Fix*: Provided that the database contains information on the effective inclusion date, drop data for the assets before their inclusion date.

Principles of Strategy Development: Overview

Accounting for Transaction Costs

- Assessing the impact of transaction costs on strategy returns requires an estimate of these costs and of the strategy's turnover.
- The case of *proportional transaction costs* is the easiest to deal with. Let c denote the trading cost per currency unit (including commissions, the bid-ask spread, and any transaction taxes).
- That cost could differ by asset but we shall assume that it is the same for all assets considered in the strategy. The extension to the general case is straightforward.
- A rough estimate of the effect of transaction costs on the return per period is just the product of c and the strategy's average turnover. For example, if transaction costs are $c = 0.2\%$ and the strategy's turnover is 250% per year, transaction costs will reduce returns by about 0.5% per year.

Accounting for Transaction Costs

- For enhanced precision, one can subtract the product of c and the turnover at the end of each period from the return in that period.
- Letting $\mathbf{w}$ denote the vector of portfolio weights and $|...|$ component-wise absolute value, total turnover in period t can be approximated as

$$z_t = |\mathbf{w}_t - \mathbf{w}_{t-1}|'\mathbf{1} \ . \tag{38}$$

- This approximation does not account for the change in portfolio weights induced by differences in returns across assets. To include this effect, compute the portfolio weights $\hat{\mathbf{w}}_t$ at the end of the period, right before trading takes place, using

$$\hat{w}_{n,t} = w_{n,t-1}\frac{1 + R_{n,t}}{1 + R_{p,t}} \ . \tag{39}$$

Total turnover is then

$$z_t = |\mathbf{w}_t - \hat{\mathbf{w}}_t|'\mathbf{1} \ . \tag{40}$$

- This will be slightly off due to dividend and interest payments, but typically close enough.

Overview

Risk Premium Estimation

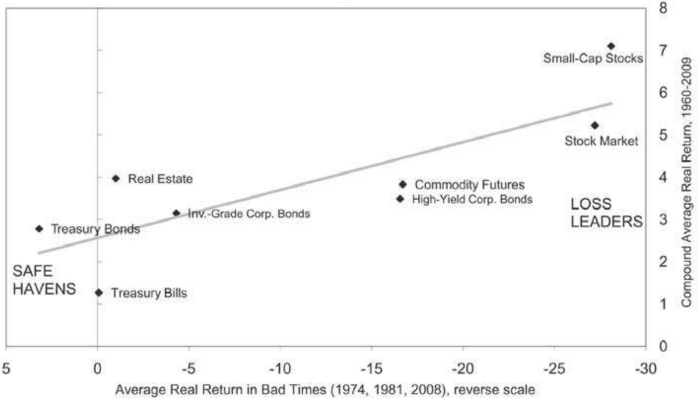
- The simplest strategies are those targeting a constant allocation to a number of asset classes.
- The key decisions to make for constant proportion strategies are:
  - Which asset classes should be included and in what proportions?
  - How often should the portfolio be rebalanced towards the target weights?
- As mentioned above, these strategies earn the risk premia on the asset classes over time, and the size of the allocation depends on the (unconditional) risk premium per unit of risk, the level of risk, and the timing of losses.

Risk Premium Estimation

- **Unconditional** risk premia on the different asset classes can be estimated as the difference between the average return on the risky asset and the average return on the riskless asset.
- For the different asset classes, the following can be used:
  - *Equity risk premium*: average total return on stocks;
  - *Bond risk premium*: average total return on long-term debt;
  - *Credit risk premium*: average total return on risky debt;
  - *Real estate risk premium*: average total return on real estate;
  - *Commodity risk premium*: average total return on commodities.
- Since average return estimates are noisy, one must use **long sample periods**.
- A word of caution: There is evidence that risk premia are **not constant** over time. Estimating **conditional** risk premia is much harder than obtaining unconditional estimates. This is what timing models typically try to do.

Risk Premium Estimation

- For histories of 50 years, Ilmanen (2011) provides the following estimates:



- Note that one can also estimate return premia within an asset class using average returns. For example, one can estimate the size premium as the average return of a portfolio that is long small firms and short large firms.

Risk Premium Estimation: Fama-MacBeth Method

- The **Fama-MacBeth method** is another method to estimate risk premia, which will also prove to be useful when we consider selection models.

- Suppose there are N assets and that excess returns on asset n are given by:

$$R_{n,t} - R_{f,t} = a_n + \boldsymbol{\beta}_n' \mathbf{f}_t + \varepsilon_{n,t} \ , \qquad t = 1, \ldots, T, \quad n \text{ given}, \tag{41}$$

  where the $\mathbf{f}$s are factors driving returns and a_n is an asset-specific constant (not an alpha!).

- The method involves the following steps:
  1. Run n time series regressions, one for each asset, to obtain the $\boldsymbol{\beta}_n$s.
  2. For each time period, run a cross-sectional regression of excess returns on the $\boldsymbol{\beta}$s to estimate the returns on the factors during a given period:

$$R_{n,t} - R_{f,t} = \gamma_t + \boldsymbol{\beta}_n' \boldsymbol{\lambda}_t + \alpha_{n,t} \ , \qquad n = 1, \ldots, N \ , \quad t \text{ given}. \tag{42}$$

  3. Obtain the estimates of the α_ns and the factor risk premia $\boldsymbol{\lambda}$ as their time series averages:

$$\hat{\boldsymbol{\lambda}} = \frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{\lambda}}_t \ , \qquad \hat{\alpha}_n = \frac{1}{T} \sum_{t=1}^{T} \hat{\alpha}_{n,t} \text{ for each asset } n. \tag{43}$$

Risk Premium Estimation: Fama-MacBeth Method

- To assess the significance of the risk premia and the intercepts, compute the covariance matrices of the estimates as

$$\text{cov}(\hat{\boldsymbol{\lambda}}) = \frac{1}{T^2} \sum_{t=1}^{T} (\hat{\boldsymbol{\lambda}}_t - \hat{\boldsymbol{\lambda}})(\hat{\boldsymbol{\lambda}}_t - \hat{\boldsymbol{\lambda}})' \ , \tag{44}$$

$$\text{cov}(\hat{\boldsymbol{\alpha}}) = \frac{1}{T^2} \sum_{t=1}^{T} (\hat{\boldsymbol{\alpha}}_t - \hat{\boldsymbol{\alpha}})(\hat{\boldsymbol{\alpha}}_t - \hat{\boldsymbol{\alpha}})' \ . \tag{45}$$

- Note that we are using the variation of the cross-sectional regression coefficients over time to assess the significance of their average.
- The resulting standard errors are not perfect, as they *do not account for the fact that the $\boldsymbol{\beta}$s have been estimated*. One can account for this by running a single cross-sectional regression for the sample average excess return and using (messy) formulas for the standard errors. In that case one would run:

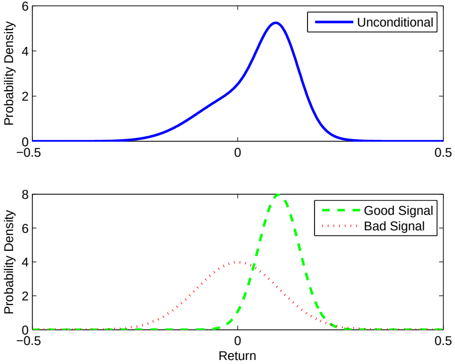
$$\mathbb{E}(R_{n,t} - R_{f,t}) = \gamma + \boldsymbol{\beta}_n' \boldsymbol{\lambda} + \alpha_n \ , \qquad n = 1, \ldots, N. \tag{46}$$

Overview

Timing Models: Basic Idea

- Timing models attempt to enhance returns by varying the exposure to asset classes or subsets of assets within such classes over time. Examples:
  - Market timing models switch between stocks, bonds and cash.
  - Sector rotation models switch between cyclical and defensive sectors.

- The key to developing a timing model is to find indicators for which the assets' **conditional** return distribution differs from the **unconditional** one.



Timing Models: Example

- To see how finding such indicators allows improving portfolio performance, let us solve the investor's portfolio problem for the above distribution both with and without knowledge of the signal.
- In the distribution depicted on the previous page, good and bad market conditions are equally likely, and the returns conditional on the signal are normally distributed with the following parameters:
  - When observing a good signal, $\mu_1 = 0.1$ and $\sigma_1 = 0.05$.
  - When observing a bad signal, $\mu_2 = 0$ and $\sigma_2 = 0.1$.
- Suppose that the riskless rate is 2% and that the investor has mean-variance preferences with a risk aversion $a = 10$.

Timing Models: Example

- **Without** knowledge of the signal, the investor must base his portfolio decision on the **unconditional** distribution. Letting $p = 1/2$ denote the probability of good market conditions, the average and standard deviation of returns are given by:

$$\mu = p\mu_1 + (1-p)\mu_2 = 0.05 \ , \tag{47}$$

$$\sigma = \sqrt{p(\sigma_1^2 + (\mu_1 - \mu)^2) + (1-p)(\sigma_2^2 + (\mu_2 - \mu)^2)} = 0.0935 \ . \tag{48}$$

- Hence, the Sharpe ratio from investing in the asset with constant exposure is

$$SR_{\text{No Timing}} = \frac{\mu - R}{\sigma} = \frac{0.05 - 0.02}{0.0935} = 0.3207 \ . \tag{49}$$

- The investor's optimal investment in the risky asset is given by

$$w_{NT} = \frac{1}{a}\frac{\mu - R}{\sigma^2} = \frac{1}{10}\frac{0.05 - 0.02}{0.0935^2} = 0.3429 \ , \tag{50}$$

  and his portfolio has expected return $\mu_p = w_{NT}\mu + (1 - w_{NT})R = 0.0303$ and standard deviation $\sigma_p = |w_{NT}|\sigma = 0.0321$.

Timing Models: Example

- **With** access to the signal, the investor selects his portfolio depending on the **conditional** distribution corresponding to the signal's value:

$$\text{Good signal:} \quad w_1 = \frac{1}{a}\frac{\mu_1 - R}{\sigma_1^2} = \frac{1}{10}\frac{0.1 - 0.02}{0.05^2} = 3.20 \ . \tag{51}$$

$$\text{Bad signal:} \quad w_2 = \frac{1}{a}\frac{\mu_2 - R}{\sigma_2^2} = \frac{1}{10}\frac{0 - 0.02}{0.1^2} = -0.20 \ . \tag{52}$$

- Hence, the portfolio's conditional expected return and standard deviation are

$$\mu_{p,1} = w_1\mu_1 + (1 - w_1)R = 0.2760 \ , \quad \sigma_{p,1} = |w_1|\sigma_1 = 0.16 \ , \tag{53}$$

$$\mu_{p,2} = w_2\mu_2 + (1 - w_2)R = 0.0240 \ , \quad \sigma_{p,2} = |w_2|\sigma_2 = 0.02 \ . \tag{54}$$

- Applying the earlier formula for mixture distributions yields

$$\mu_p = p\mu_{p,1} + (1 - p)\mu_{p,2} = 0.15 \ , \tag{55}$$

$$\sigma_p = \sqrt{p(\sigma_{p,1}^2 + (\mu_{p,1} - \mu_p)^2) + (1 - p)(\sigma_{p,2}^2 + (\mu_{p,2} - \mu_p)^2)} = 0.1699 \ .$$

- Hence, the Sharpe ratio from a timed investment in the asset is

$$SR_{\text{Timing}} = \frac{\mu_p - R}{\sigma_p} = \frac{0.15 - 0.02}{0.1699} = 0.7650 \ . \tag{56}$$

Timing Models: Overview

Predictability

- As the example illustrates, timing models rely on predictability in asset returns.
- Predictability studies **patterns in the return distribution over time**.
- The goal is to identify a set of **variables that forecast the future return distribution** in order to identify periods in which risky asset returns are attractive and periods in which they are not.
- Most of the literature focuses on predicting expected returns, but one can also attempt to predict higher moments of the return distribution.
- In academia researchers have focused mainly on in-sample predictability, where the parameters are estimated over the entire sample.
- Working with **out-of-sample predictions** is critical for trading applications.

Predictability: Approaches

Let x_t denote the predictive variable being considered and $R_{t,t+h}$ the future return, where h is the forecasting horizon. (When there is no risk of confusion we shall just write R_{t+h} instead of $R_{t,t+h}$.)
To assess whether x_t has predictive ability, the following methods can be used:

1. **Split the sample** based on the values of x_t (e.g. positive versus negative values, values above or below the sample median,...). Compute the summary statistics of R_{t+h} for the subsamples and run a test for identical distributions.

2. Estimate the **predictive regressions**

$$R_{t+h} = \alpha_h + \beta_h x_t + \varepsilon_{t+h} \ , \tag{57}$$

$$R^2_{t+h} = \alpha_h + \beta_h x_t + \varepsilon_{t+h} \ , \tag{58}$$

   and assess the significance of the betas. The first regression tells you something about expected returns, the second about the variance. You can also investigate higher moments if you see some nonnormality.

3. Run **predictive regressions using dummy variables** $D_{x_t > \tilde{x}_t}$ instead of x_t:

$$R_{t+h} = \alpha_h + \beta_h D_{x_t > \tilde{x}_t} + \varepsilon_{t+h} \ , \tag{59}$$

$$R^2_{t+h} = \alpha_h + \beta_h D_{x_t > \tilde{x}_t} + \varepsilon_{t+h} \ . \tag{60}$$

Predictability: Approaches

- If the critical values you use when splitting the sample in subsamples (point (1) above) are based on the entire sample, you are working in-sample. The same holds if the regression coefficients (points (2) and (3)) are estimated over the entire sample.

- If the critical values and the regression coefficients are estimated with rolling or expanding windows, then you are working out-of-sample. Of course, the forecast should be based on information available within the window you used for the estimation (no look-ahead bias).

Timing Models: Overview

Common Predictors

The finance literature has extensively investigated predictive variables across asset classes. Here is a list of the most common ones:

- Valuation ratios:
  - Dividend/Price ratio (D/P).
  - Earnings/Price ratio (E/P and E10/P).
  - Book-to-Market ratio (B/M).
  - Consumption/Wealth ratio (C/W or CAY).
- Interest rate variables:
  - T-bill rate and T-bill rate minus its 12-month moving average.
  - Term spread, i.e. long term yield minus short term rate (TS).
  - Credit spread, e.g. BBB yield minus AAA yield (CS).
- Volatility variables:
  - Realized volatility.
  - Implied volatility, e.g. VIX.
- Macroeconomic variables:
  - GDP Growth (GDP).
  - Inflation (π).
- News sentiment.

Common Predictors: Intuition

- D/P, E/P, and B/M are ratios of accounting-based fundamental variables and market prices. They measure a discrepancy between fundamentals and market expectations / valuation.
- T-bill, TS, and CS reflect monetary policy and market expectations about future monetary policy and the business cycle. For instance, TS reflects business cycle movements and anticipated changes in monetary policy, while CS reflects the market's assessment of future default risk.
- VIX reflects perceived risk.
- GDP and inflation represent macroeconomic fundamentals of the economy.
- News sentiment reflects market participants' feeling about recent events.

Timing Models: Overview

Assessing the Value of Predictability

- Using the information provided by predictive regressions can yield large improvements in portfolio returns even if the regression R^2s are low.
- To see why, consider the following example, analyzed in Campbell and Thompson (2008). Suppose that the risky asset's excess return is given by

$$R^e_{t+1} = \mu + x_t + \varepsilon_{t+1} \ , \tag{61}$$

  where
  - μ is the unconditional average excess return,
  - x_t is the predictor variable assumed to have zero mean and constant variance σ_x^2, and
  - ε_{t+1} is a random shock with mean zero and constant variance σ_ε^2.
- As before, the investor has mean-variance preferences with risk aversion a.

- **If the investor does not observe** x_t, his optimal risky asset investment is

$$w_{NT} = \frac{1}{a} \frac{\mu}{\sigma_x^2 + \sigma_\varepsilon^2} \ . \tag{62}$$

- His portfolio's average excess return is

$$\mu_{p,NT} - R_f = \frac{1}{a} \frac{\mu^2}{\sigma_x^2 + \sigma_\varepsilon^2} = \frac{1}{a} S^2 \ , \tag{63}$$

where S is the unconditional Sharpe ratio of the risky asset.

Assessing the Value of Predictability

- **If the investor observes** x_t, he selects a portfolio weight

$$w_T = \frac{1}{a}\frac{\mu + x_t}{\sigma_\varepsilon^2} \ . \tag{64}$$

Note that the denominator is now σ_ε^2 because the variation in x_t does not contribute to risk.

- The average excess return is now

$$\mu_{p,T} - R_f = \frac{1}{a}\frac{\mu^2 + \sigma_x^2}{\sigma_\varepsilon^2} = \frac{1}{a}\frac{S^2 + R^2}{1 - R^2} \ , \tag{65}$$

where the second equality follows from the fact that the R^2 from the regression of excess returns on the predictor x_t is

$$R^2 = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\varepsilon^2} \ . \tag{66}$$

Assessing the Value of Predictability

- The improvement in the portfolio's expected excess return is

$$\mu_{p,T} - \mu_{p,NT} = \frac{1}{a}\frac{R^2}{1-R^2}\left(1+S^2\right) \ . \tag{67}$$

If R^2 and S^2 are both small (as is typically the case for short time intervals), the value will be close to R^2/a.

- The proportional improvement in the portfolio's expected excess return is

$$\frac{\mu_{p,T} - \mu_{p,NT}}{\mu_{p,NT} - R_f} = \frac{R^2}{1-R^2}\frac{1+S^2}{S^2} \ . \tag{68}$$

Again, if R^2 and S^2 are both small (as is typically the case for short time intervals), the value will be close to R^2/S^2.

Assessing the Value of Predictability

Campbell and Thompson (2008) provide the following numerical examples that reveal that the effect can be quite large:

- **For monthly data:**
  - The monthly Sharpe ratio of US stocks is $S = 0.108$, so $S^2 = 0.012$.
  - The monthly out-of-sample R^2 for the smoothed earnings-price ratio is 0.43%.
  - Hence, the investor can increase average monthly portfolio returns by $0.43/1.2 = 36\%$.
  - This translates to 5.2% per year for unit risk aversion and 1.7% per year for a risk aversion of 3.
- **For annual data:**
  - The annual Sharpe ratio is $S = 0.374$, yielding $S^2 = 0.118$.
  - The annual out-of-sample R^2 is 7.9%.
  - The relative improvement in returns is $7.9/11.8 = 67\%$.
  - The absolute increase in return is 9.6% per year for unit risk aversion and 3.2% per year for a risk aversion of 3.

Timing Models: Overview

Practical Considerations

There are a few key decisions to make when developing a timing model:

1. *How long is one period?* One day, one month, one year?
   - How fast-moving and volatile are the target assets?
   - Compute turnover. Checking the signal every day does not mean that you trade every day.
2. *What is the forecasting horizon h?*
3. *Do you use the signal to decide on the entire portfolio each period or only on a fraction?*
   - How accurate and fast-moving is the signal?
   - Is turnover a concern?
4. *Should you use the raw signal or a transformed version?*
   - Dummies will miss the magnitude, but are more robust to outliers.
5. *Assuming you find something, how do you want to implement the model?*
   - Binary long / short or long / flat using thresholds;
   - Ternary long / flat / short using thresholds;
   - Scaled long / short based on signal strength.

Timing Models: Overview

Statistical Issues with Predictive Regressions

- Recall that predictive regressions are usually of the form

$$R_{t+h} = \alpha_h + \beta_h x_t + \varepsilon_{t+h} \ , \tag{69}$$

  where
  - R_{t+h} denotes the return to be forecast (one can also use continuously compounded returns r_{t+h});
  - x_t is a stationary variable that might predict returns (e.g. lagged returns, financial variables, asset fundamentals); and
  - h denotes the time lag which could be days, months, quarters or years.

- The null hypothesis of no predictability is $H_0 : \beta_h = 0$.

- Predictive regressions raise a number of statistical issues:
  1. Standard errors must be adjusted to account for the presence of **overlapping observations**.
  2. Using **persistent predictors** which (weakly) predict short-term returns implies that predictability adds up over longer horizons.
  3. Coefficient estimates will be **biased** when moves in asset prices affect the predictive variable.

Timing Models: Overview

Overlapping Observations

- In many settings, **predictability works better at long horizons**. In these cases one typically estimates predictive regressions with a forecasting horizon $h > 1$.
- One approach would be to estimate the regression including one observation every h periods. This works but means throwing away a large part of the data.
- It is as such preferable to estimate the regression including all the data. However, doing so will cause the error terms to be serially correlated.
- The standard errors need to be adjusted to account for this autocorrelation. The most common way to do so is to use Newey-West (1987) standard errors.
- If the forecasting horizon is h periods, there are $h - 1$ overlapping observations and you should include $h - 1$ leads and lags in the Newey-West standard error estimator.

Overlapping Observations: Statistical Analysis

- Consider predicting a two-period return. Let us work with log returns for simplicity.
- Let $r_{t,t+2}$ denote the return from period t to period $t+2$ and $r_{t+1,t+3}$ be the return from period $t+1$ to period $t+3$:

$$r_{t,t+2} = r_{t+1} + r_{t+2} = \alpha + \beta x_t + \varepsilon_{t+2} \ , \quad (70)$$

$$r_{t+1,t+3} = r_{t+2} + r_{t+3} = \alpha + \beta x_{t+1} + \varepsilon_{t+3} \ . \quad (71)$$

- Suppose that $\alpha = \beta = 0$. Then

$$\varepsilon_{t+2} = r_{t+1} + r_{t+2} \ , \quad (72)$$

$$\varepsilon_{t+3} = r_{t+2} + r_{t+3} \ . \quad (73)$$

- The return for period 2, r_{t+2}, shows up in both error terms. This causes serial correlation in the error terms. The OLS estimator is no longer efficient.
- As claimed above, if your forecasting horizon is h periods, you have $h-1$ overlapping observations.

Timing Models: Overview

Persistent Predictors

- A common feature of many if not most predictors documented in the literature is that they are **persistent**. This is in particular the case for the following:
  - Short-term interest rates;
  - Term spreads;
  - Credit spreads;
  - Dividend yields.
- Novy-Marx (2014) shows that their significance is likely to be spurious by finding a host of other predictors such as political variables, the weather, sunspot activity, and the conjunction of planets.
- To avoid spurious inference, use differences of the explanatory variables or remove their cyclical component.

Persistent Predictors: Examples

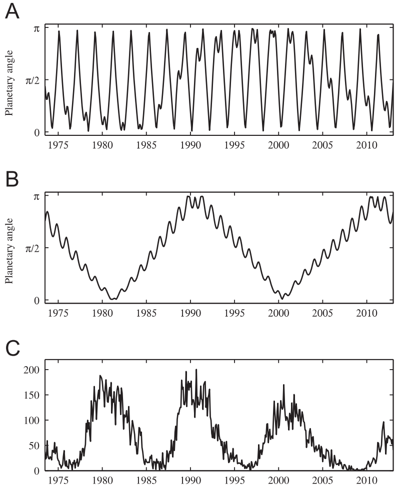


Fig. 4. Celestial phenomena. This figure shows the levels of the celestial predictive variables. Panel A shows the aspect of Mars and Saturn (i.e., the angle between the planets and an Earth observer). Panel B shows the aspect of Jupiter and Saturn. Panel C shows the number of sunspots observed each month. The data cover July 1973 to December 2012.

Strategy	average monthly anomaly excess returns ($\hat{m}$)	sensitivity of anomaly returns to the predictive variable (b)
Panel A: Angle between Mars and Saturn as predictive variable		
Market	0.51	0.53
	[2.37]	[2.29]
Return on assets	0.69	−0.52
	[3.07]	[−2.17]
Earnings-to-price	1.11	−0.55
	[5.10]	[−2.37]
Failure probability	0.87	−0.76
	[2.53]	[−2.06]
Ohlson's O-score	0.30	−0.52
	[1.45]	[−2.38]
Idiosyncratic volatility	0.33	−0.85
	[0.96]	[−2.28]
Panel B: Angle between Jupiter and Saturn as predictive variable		
Size	0.24	−0.55
	[1.12]	[−2.34]
Net issuance	0.73	−0.45
	[5.43]	[−3.05]
Asset growth	0.57	−0.48
	[3.25]	[−2.48]
Betting against beta	0.97	−0.39
	[6.03]	[−2.18]
Panel C: Observed number of sunspots as predictive variable		
UMD	0.66	0.01
	[3.14]	[2.25]
Earnings momentum (PEAD)	0.70	0.01
	[4.27]	[2.18]
Cyclic moving average of observed number of sunspots as predictive variable		
Size	0.24	−0.04
	[1.12]	[−2.09]
Ohlson's O-score	0.30	0.05
	[1.45]	[3.11]
Seasonality	0.87	0.04
	[4.56]	[2.70]

Source: Novy-Marx (2014).

Persistent Predictors: Statistical Analysis

- Let us show why predictability tends to be stronger at longer horizons.

- Suppose we forecast one-period returns with a variable x which follows an AR(1) process. For simplicity, assume that the intercepts are zero. The setup is:

$$r_{t+1} = \beta x_t + \varepsilon_{t+1} \ , \tag{74}$$

$$x_{t+1} = \rho x_t + \eta_{t+1} \ . \tag{75}$$

- The two-period return is

$$\begin{aligned} r_{t+1} + r_{t+2} &= (\beta x_t + \varepsilon_{t+1}) + (\beta x_{t+1} + \varepsilon_{t+2}) \\ &= \beta x_t + \beta(\rho x_t + \eta_{t+1}) + \varepsilon_{t+1} + \varepsilon_{t+2} \\ &= \beta(1+\rho)x_t + (\beta \eta_{t+1} + \varepsilon_{t+1} + \varepsilon_{t+2}) \ . \end{aligned} \tag{76}$$

Persistent Predictors: Statistical Analysis

- Similarly, the three-period return is

$$r_{t+1} + r_{t+2} + r_{t+3} = \beta(1 + \rho + \rho^2)x_t + \left(\beta(\rho\eta_{t+1} + \eta_{t+2}) + \varepsilon_{t+1} + \varepsilon_{t+2} + \varepsilon_{t+3}\right) . \qquad (77)$$

- Hence, if ρ is positive and sufficiently high, the slope coefficient builds up with the forecasting horizon. One can also show that the R^2 similarly increases with horizon.

- Thus, if short-term returns are slightly predictable by a persistent/ slow-moving variable, that predictability adds up over longer horizons.

- The increased forecasting power shows up as increasing slope coefficients and R^2s.

- Put differently, long-horizon predictability arises mechanically from persistent regressors, and findings of predictability at different horizons are not separate facts.

Timing Models: Overview

Stambaugh Bias

- Stambaugh (1999) considered the following common forecasting setup:

$$r_{t+1} = \alpha + \beta x_t + \varepsilon_{t+1} \ , \quad (78)$$

$$x_{t+1} = \theta + \rho x_t + \eta_{t+1} \ . \quad (79)$$

- Recall the OLS assumption: $\mathbb{E}[\varepsilon|x] = 0$.
- This assumption will be violated if the predictor x depends on the asset's price.
- For example, suppose that x is the D/P ratio. A positive shock to prices in period $t+1$ raises r_{t+1} meaning $\varepsilon_{t+1} > 0$. This also lowers D/P at time $t+1$, i.e. one has $\eta_{t+1} < 0$.
- A similar effect would arise when predicting bond returns using interest rates or yield spreads, or currency returns using the forward premium/discount.
- A negative correlation between ε_{t+1} and η_{t+1} violates the OLS assumption of strict exogeneity. This causes the OLS estimator to be biased.

Stambaugh Bias

- Stambaugh showed that the bias of the OLS estimator is related to the bias in $\hat{\rho}$, the sample first-order autocorrelation of x_t:

$$\mathbb{E}[\hat{\beta} - \beta] = \frac{\sigma_{\varepsilon\eta}}{\sigma_\eta^2} \mathbb{E}[\hat{\rho} - \rho] \ . \tag{80}$$

- Kendall (1954) shows that under normality,

$$\mathbb{E}[\hat{\rho} - \rho] \approx -\frac{(1+3\rho)}{T} < 0 \ , \tag{81}$$

where T is the sample size. Combining,

$$\mathbb{E}[\hat{\beta} - \beta] \approx -\frac{\sigma_{\varepsilon\eta}}{\sigma_\eta^2} \frac{(1+3\rho)}{T} \ . \tag{82}$$

- Hence, a negative covariance $\sigma_{\varepsilon\eta} < 0$ implies that $\hat{\beta}$ is upward biased and the bias increases in ρ. Also, $\hat{\beta}$ is positively skewed and has higher variance and kurtosis than the normal sampling distribution of the OLS estimator.
- Correcting for these issues will weaken evidence of return predictability.

Timing Models: Overview

Out-of-sample Prediction

- In-sample predictive regressions are subject to look-ahead bias because the coefficient estimates depend on future returns.
- To eliminate this look-ahead bias, one performs **out-of-sample predictions** in which one forecasts returns at time $t+h$ using **information through time t only**.
- Usually the out-of-sample forecast of a specific predictive regression model is compared to the prediction of some benchmark model. Academic papers often use the average historical return as benchmark.
- Procedure:
  1. Stand at time t, look back m periods and run a predictive regression on this sample (i.e. using x_{t-m} through x_{t-h} and $r_{t-m,t-m+h}$ through $r_{t-h,t}$).
  2. Using the regression coefficients just estimated, compute the expected value of the future return $\hat{r}_{t+h}$.
  3. Repeat steps 1 and 2, expanding or rolling the estimation window by one period until the most recent observation is used to forecast the future return.

Out-of-sample Prediction

- This procedure will generate a time-series of out-of-sample return predictions $\hat{r}_{t+h}$. For each period t, you can compute the squared forecast error

$$(r_{t+h} - \hat{r}_{t+h})^2 \ , \tag{83}$$

where r_{t+h} is the realized/observed return.

- Let $\hat{r}_{M,t+h}$ be the return forecast from the predictive regression model and $\hat{r}_{B,t+h}$ the forecast based on the benchmark model.

- Compute Mean Squared Forecast Errors (MSFE) for both models:

$$MSFE_M = \mathbb{E}\left[(r_{t+h} - \hat{r}_{M,t+h})^2\right] \ , \tag{84}$$

$$MSFE_B = \mathbb{E}\left[(r_{t+h} - \hat{r}_{B,t+h})^2\right] \ . \tag{85}$$

- One often reports the difference in root MSFEs, which is given by

$$\Delta RMSFE = \sqrt{MSFE_B} - \sqrt{MSFE_M} \ . \tag{86}$$

Out-of-sample Prediction

- The out-of-sample R^2 is

$$R^2_{OOS} = 1 - \frac{MSFE_M}{MSFE_B} \; . \tag{87}$$

- A positive R^2_{OOS} means that the MSFE of the predictive regression model is smaller than the MSFE of the benchmark model.

- To test the null hypothesis $H_0 : MSFE_B \leq MSFE_M$ or $H_0 : R^2_{OOS} \leq 0$, Clark and West (2007) proposed the following test statistic: the Newey-West t-statistic (with $h-1$ lags) from regressing $\tilde{d}_{M,t}$ on a constant, with

$$\tilde{d}_{M,t} = \hat{u}^2_{B,t} - \left[\hat{u}^2_{M,t} - \left(\hat{r}_{B,t} - \hat{r}_{M,t}\right)^2\right] \; , \tag{88}$$

where

$$\hat{u}_{B,t} = r_t - \hat{r}_{B,t} \; , \tag{89}$$

$$\hat{u}_{M,t} = r_t - \hat{r}_{M,t} \; . \tag{90}$$

Overview

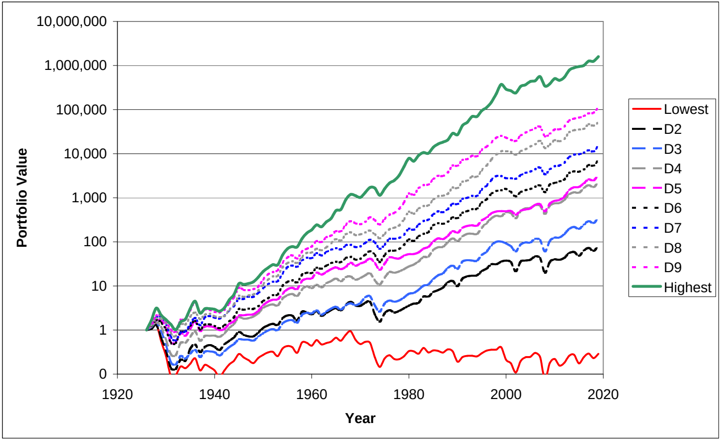
Selection: Basic Idea

- Selection models attempt to generate returns by **selecting assets from a cross-section based on some criteria**. Usually the assets are in the same asset class but this is not required. Examples:
  - **Carry**: Select currencies based on the level of interest rates.
  - **Value**: Select stocks based on their book-to-market ratio.
  - **Momentum**: Select stocks based on their past returns.

- Usually one proceeds as follows:
  1. **Sort the assets into groups** based on the value of the variable of interest. The number of groups is chosen to strike a balance between stronger effects at the extremes (which speaks for many groups) and the wish to reduce exposure to idiosyncratic risk (which requires many assets per group). Terciles, quintiles and deciles are most commonly used. One can also set the number of assets per group directly.
  2. **Build portfolios of the assets in each group.** The portfolios can be equally weighted, value-weighted, or weighted based on the strength of the signal (value or rank) or their risk.
  3. **Investigate the return properties of the portfolios.**

Selection: Example

- The chart below shows the value over time of value-weighted decile portfolios of US stocks built monthly based on return momentum (computed as the return from 12 months to 1 month prior to the portfolio formation date):



Source: Kenneth French Data Library.

Selection Models: Overview

Selection: Multivariate Case

- The approach we described so far is appropriate if a single variable is believed to affect expected returns.
- In many applications, one will want to incorporate the impact of several variables. Three approaches can be used in this case:
    1. Bivariate or three-way sorts;
    2. Scoring;
    3. Regression.
- We now consider them in detail.

Selection: Bivariate or Three-Way Sorts

- As the name says, this approach involves sorting over several variables.
- For example, a two-way sort into terciles along the dimensions size and book/market would look as follows:

		Book/Market		
		Value	Medium	Growth
	Small			
Size	Medium			
	Large			

- There are actually two approaches:
  1. **Independent sort:** Sort on the different variables independently. The resulting portfolios will be the intersection of portfolios built on the individual variables. The number of assets in each portfolio will differ significantly if the two variables are related.
  2. **Dependent sort:** Sort based on a first variable. Within each quantile of the first variable, sort based on a second variable. The advantage of this approach is that the number of assets in each cell is similar.

Selection: Scoring

- This approach involves attributing scores to each asset for each variable considered and then aggregating them into an overall score for the asset.
- The steps are:
  1. For each variable considered, attribute a score to each asset based on its quantile, for example:
     - The number of the quantile;
     - $+1$ for the top quantile, 0 for the middle quantiles, -1 for the bottom quantile.

     Make sure that a higher score means better returns for all the variables.
  2. Compute an overall score for the asset as the sum or average of the scores for the different variables. Weighting the variables is possible but beware of overfitting.

Asset	Individual scores			Overall
	Size	Value	Momentum	score
Firm X	1	0	1	2
Firm Y	-1	1	0	0
Firm Z	0	-1	-1	-2

  3. Construct the portfolio based on the overall score.

Selection: Regression

- This approach uses regressions to estimate the expected returns on the different assets.
- The steps are:
  1. Use the Fama-MacBeth method or cross-sectional regression to estimate the exposure of the different assets to the factors and the factor risk premia. Repeating what we saw in Section 7:

$$R_{n,t} - R_{f,t} = a_n + \boldsymbol{\beta}_n' \mathbf{f}_t + \varepsilon_{n,t} \ , \qquad t = 1, \ldots, T, \quad n \text{ given}; \tag{91}$$

$$R_{n,t} - R_{f,t} = \gamma_t + \boldsymbol{\beta}_n' \boldsymbol{\lambda}_t + \alpha_{n,t} \ , \qquad n = 1, \ldots, N, \quad t \text{ given}; \tag{92}$$

$$\hat{\boldsymbol{\lambda}} = \frac{1}{T} \sum_{t=1}^{T} \hat{\boldsymbol{\lambda}}_t \ , \qquad \hat{\alpha}_n = \frac{1}{T} \sum_{t=1}^{T} \hat{\alpha}_{n,t} \text{ for each asset } n. \tag{93}$$

  2. Compute the expected excess return on each asset based on the factor exposures and the factor risk premia:

$$\mathbb{E}(R_{n,t} - R_{f,t}) = \gamma + \boldsymbol{\beta}_n' \hat{\boldsymbol{\lambda}} \ , \qquad n = 1, \ldots, N. \tag{94}$$

- This is similar to scoring except that the weights of the variables differ depending on the cross-sectional distribution of the factor exposures and the factor risk premia. Note that the result does *not* depend on how the factors are scaled so you don't have to worry about that.

Selection Models: Overview

Selection: Implementation Options

Selection models can be implemented in three ways:

1. **Long only**. In this case, the return each period is equal to the return on the portfolio being held long.

2. **Long-short**. In this case, the return each period is equal to (see Section 4):
   - \+ The return on the portfolio being held long;
   - \+ The riskless rate (omitted for zero-cost long-short portfolios);
   - − The return on the portfolio being held short;
   - − The lending fees for the portfolio being held short.

3. **Long with a hedge for market risk**. In this case, if the hedge is implemented using futures, the return each period is equal to:
   - \+ The return on the portfolio being held long;
   - − The hedge ratio times the return on the futures.

Overview

Combining Strategies

- Individual strategies won't work all the time, so it is advisable to manage your assets using **multiple strategies**.
- You are used to combining individual assets to form a portfolio. Based on the expected return and risk of the individual assets, you can assess the expected return and risk of a portfolio.
- The same actually applies when combining strategies. **Risk and return combine in the same way, as if each strategy were an individual asset.**
- You could use portfolio optimization to allocate funds to the different strategies. Your overall portfolio would then be the weighted average of the positions implied by the individual strategies.
- **Be cautious about running portfolio optimization** on your strategies:
  - The returns on the individual strategies are likely to be non-normal.
  - Even if you beware of overfitting, you will have been experimenting to get good returns in each strategy. Since standard portfolio optimizers react strongly to small differences in expected returns, you are likely to get extreme allocations.

Combining Strategies: Optimization vs. Heuristics

- Instead of running portfolio optimization to allocate funds to strategies, it is therefore useful to have heuristics.
- The advantage of heuristics is that they tend to be more robust (less overfitting). But using them might cost in terms of performance.
- Determining the alpha and the factor exposures when combining strategies is trivial. **Alpha and the factor exposures for the overall portfolio will be the weighted average of their values for the individual strategies.**
- Determining the Sharpe ratio of a strategy combination is slightly more involved. We now consider a few typical combination approaches and investigate their Sharpe ratios. This will also allow us to understand when adding strategies is likely to be beneficial for overall performance.

Portfolios constructed using	
Optimization	**Heuristics**
1. Tangency portfolio	3. Risk parity portfolio
2. Risky asset only global MVP	4. Equally weighted portfolio

Combining Strategies: Tangency Portfolio

- Suppose you were to run **mean-variance portfolio optimization**. Let $\boldsymbol{\mu}$ denote the vector of expected returns and $\boldsymbol{\Sigma}$ the variance-covariance matrix of returns. For a risk aversion a, the **optimal portfolio** is

$$\mathbf{w} = \frac{1}{a}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - R_f\mathbf{1}) \ . \tag{95}$$

- What Sharpe ratio will you achieve? The optimal portfolio's expected return and variance are given by

$$\mu_p = R_f + \mathbf{w}'(\boldsymbol{\mu} - R_f\mathbf{1}) = R_f + \frac{1}{a}(\boldsymbol{\mu} - R_f\mathbf{1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - R_f\mathbf{1}) \ , \tag{96}$$

$$\sigma_p^2 = \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} = \frac{1}{a^2}(\boldsymbol{\mu} - R_f\mathbf{1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - R_f\mathbf{1}) \ . \tag{97}$$

- Hence, the best Sharpe ratio that you can achieve is:

$$\overline{SR}_p = \frac{\mu_p - R_f}{\sigma_p} = \sqrt{(\boldsymbol{\mu} - R_f\mathbf{1})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - R_f\mathbf{1})} \ . \tag{98}$$

Combining Strategies: Risky Asset Only Global MVP

- Suppose you do not trust the expected return estimates. Then a candidate portfolio would be the **risky asset only global minimum variance portfolio**

$$\mathbf{w} = \frac{\mathbf{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}} \ . \tag{99}$$

- What Sharpe ratio would that give you? We have

$$\mu_p = \mathbf{w}'\boldsymbol{\mu} = \frac{\mathbf{1}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}} \ , \tag{100}$$

$$\sigma_p^2 = \mathbf{w}'\mathbf{\Sigma}\mathbf{w} = \frac{1}{\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}} \ . \tag{101}$$

- In this case your Sharpe ratio would be:

$$SR_p = \frac{\mu_p - R_f}{\sigma_p} = \frac{\frac{\mathbf{1}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}} - R_f}{\sqrt{\frac{1}{\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}}}} = \frac{\mathbf{1}'\mathbf{\Sigma}^{-1}(\boldsymbol{\mu} - R_f\mathbf{1})}{\sqrt{\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}}} \ . \tag{102}$$

Combining Strategies: Risk Parity Portfolio

- A **risk parity portfolio** weights assets according to the inverse of their volatility,

$$w_i = \frac{1/\sigma_i}{\sum_{n=1}^{N} 1/\sigma_n} \,. \tag{103}$$

- Letting ρ_{mn} denote the pairwise correlation between assets m and n, the expected return and variance of returns of the portfolio are

$$\mu_p = \mathbf{w}'\boldsymbol{\mu} = \frac{\sum_{n=1}^{N} \frac{\mu_n}{\sigma_n}}{\sum_{n=1}^{N} \frac{1}{\sigma_n}} \,, \tag{104}$$

$$\sigma_p^2 = \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} = \frac{N + \sum_{n=1}^{N} \sum_{m \neq n} \rho_{mn}}{\left(\sum_{n=1}^{N} \frac{1}{\sigma_n}\right)^2} = \frac{N + N(N-1)\overline{\rho}}{\left(\sum_{n=1}^{N} \frac{1}{\sigma_n}\right)^2} \,, \quad \overline{\rho} \equiv \frac{\sum_{n=1}^{N} \sum_{m \neq n} \rho_{mn}}{N(N-1)}$$

- The resulting Sharpe ratio is:

$$SR_p = \frac{\mu_p - R_f}{\sigma_p} = \frac{\sum_{n=1}^{N} \frac{\mu_n}{\sigma_n} - R_f \sum_{n=1}^{N} \frac{1}{\sigma_n}}{\sqrt{N + N(N-1)\overline{\rho}}} = \sqrt{\frac{N}{1 + (N-1)\overline{\rho}}} \frac{\sum_{n=1}^{N} \frac{\mu_n - R_f}{\sigma_n}}{N} \,.$$

- For an **equally weighted portfolio**, $\mathbf{w} = \frac{1}{N}\mathbf{1}$, expected return and variance are given by

$$\mu_p = \mathbf{w}'\boldsymbol{\mu} = \frac{\sum_{n=1}^{N} \mu_n}{N} \equiv \overline{\mu} \, , \tag{105}$$

$$\sigma_p^2 = \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w} = \frac{1}{N^2}\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1} = \frac{\overline{\sigma^2}}{N} + \frac{N-1}{N}\overline{\sigma} \, , \quad \overline{\sigma^2} \equiv \frac{\sum_{n=1}^{N} \sigma_n^2}{N}, \quad \overline{\sigma} \equiv \frac{\sum_{n=1}^{N} \sum_{m \neq n} \sigma_{mn}}{N(N-1)}$$

- The portfolio's Sharpe ratio is:

$$SR_p = \frac{\mu_p - R_f}{\sigma_p} = \frac{\overline{\mu} - R_f}{\sqrt{\frac{1}{N}\overline{\sigma^2} + \frac{N-1}{N}\overline{\sigma}}}$$

$$\approx \frac{\overline{\mu} - R_f}{\sqrt{\frac{1}{N}\overline{\sigma^2} + \frac{N-1}{N}\overline{\rho}\overline{\sigma^2}}} = \sqrt{\frac{N}{1 + (N-1)\overline{\rho}}} \frac{\overline{\mu} - R_f}{\sqrt{\overline{\sigma^2}}} \, . \tag{106}$$

Combining Strategies: Summary

- Using these expressions, one can compute the portfolios and the Sharpe ratios resulting from the different approaches.
  - If the Sharpe ratios **do not differ much** across approaches, it is advisable to go for an **equally weighted** portfolio.
  - If they do, **tilting the portfolio** towards the weights recommended by the tangency portfolio can be justified, but it is in general not advisable to go all the way.
  - If the tangency portfolio recommends **shorting a strategy**, one should probably **drop it**. This will typically happen if it is highly correlated with another one but has lower average excess return per unit of risk.
- In addition to the overall portfolio's Sharpe ratio, one should take a look at its **skewness** and its **worst return**. A portfolio with a lower Sharpe ratio may be preferred if it allows achieving a higher (or less negative) overall skewness or worst return.

Overview

Implementation: Overview

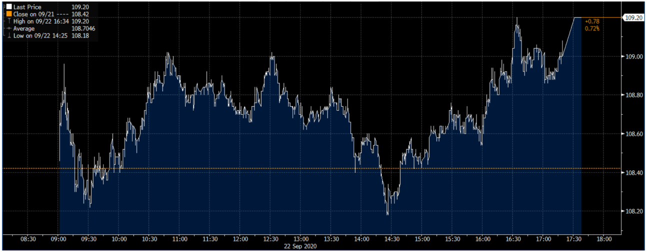
- There are often many ways to implement the same basic strategy. For example, if you have a market timing model, you could implement it by buying individual stocks, index ETFs, or index futures.
- Aspects to consider when determining the best implementation include:
  - **Access**: Is the instrument available to you?
  - **Convenience**: Number of transactions required, available trading hours, currencies in which the instruments are traded, OTC vs. exchange traded.
  - **Additional risks**: Are your positions subject to counterparty risk?
  - **Liquidity and transaction costs**: Commissions, spreads, transaction taxes.
  - **Robustness**: Is the instrument always available or is there a risk you can't get it? For example, if you need to short an asset, can you find it and what is the recall risk?
  - **Taxes**: Often tax rates on capital gains and dividends differ. For example, in many countries dividends are subject to withholding tax but capital gains are tax free for nonresidents. So an implementation with few dividends is better.
  - **Time of trades**: At the open, close, during the session?

Implementation: Overview

Implementation: Trading at the Open or at the Close

- For strategies with low turnover, the exact time of the transactions should not be critical. However, for high-turnover strategies, it is important to be careful about when you trade.
- Many exchanges have special rules for trading at the open or at the close. Prices at these times are also quite volatile.

Intraday price chart for Nestlé



Implementation: Overview

Implementation: Leverage Constraints

- If your strategy uses leverage, it is important to make sure that you can actually obtain it and **account for its cost** in your backtest.
- **Portfolio returns** (15) **account for the cost of leverage**. This expression can be extended to the case where the borrowing rate exceeds the lending rate.
- The **amount of leverage** that you can get is driven by regulations, the lending standards of your bank or broker, and the type of instrument:
  - For stock/bond positions, you can usually get 2:1 total (long plus short).
  - For futures positions, the achievable leverage is driven by the margin requirements of individual instruments and is usually quite large.
  - For long option positions, it is generally hard to get leverage.
  - For short option positions, margin requirements may exceed the contract's value, especially for OTM options.
  - For OTC derivatives, each bank or broker will have its rules.
- The easiest way to get leverage without hitting constraints is to implement some of your strategies using futures.

Implementation: Leverage Constraints

- As an example, consider the following situation:
  - You have 2 strategies. The first involves a long/short portfolio of stocks and the other trades interest rate futures.
  - Your bank allows you to have 100% of the account value (net of futures margin requirements) long and another 100% short in stocks.
  - The interest rate strategy only requires 40% margin.
- If you allocate 50% of the account value to each strategy, you actually only use up 50% + 50% * 40% = 70% of the account value. This means that you can trade your models at a maximum of 1/0.7 = 142.86%:
  - 50% * 1.4286 = 71.43% long/short in stocks uses 71.43% of the account value.
  - 50% * 1.4286 = 71.43% in futures requires 28.57% of the account value.
- Generally you will not want to go that high because this would require positions to be liquidated as soon as you experience a drawdown. But you might consider trading your strategies at 110% or 120%.

Implementation: Overview

Implementation: Issues Arising with Short Positions

- Particular issues arise if your strategy involves **short positions** in securities.
- In order to short a security and hold the position overnight, you need to **borrow** it.
  - Usually banks and brokers have inventory that they lend out against a fee. If many people are trying to short at the same time, they may be unable to find the security or the lending fee might be large.
  - Whenever you have borrowed a security, the lender can request it to be returned. If this happens, your bank or broker will try to borrow it from someone else, but if they cannot you will get a notice that they need to recall it, which will force you to close the short position on short notice.
- An additional risk with strategies involving short positions is that regulators might impose **short selling bans**. This generally occurs when a security drops sharply in value over a short period. Existing short positions are usually grandfathered, but new ones cannot be opened.

Implementation: Issues Arising with Short Positions

- Because of these problems, when analyzing a long-short strategy, it is useful to
  - investigate the **profitability of the long and short legs separately**; and
  - consider **alternatives to shorting** securities that you need to borrow.
- Conceptually, the short leg of a strategy can improve performance for two reasons:
  - Selection of securities which subsequently drop in value; and
  - Reduction of the strategy's market risk.

Implementation: Issues Arising with Short Positions

An approach that works well in practice is the following:

1. Compute the performance statistics of
    - the full long-short strategy;
    - a strategy only involving the long leg;
    - a strategy only involving the short leg; and
    - a strategy involving the long leg and a short position in index futures. Often the size of the index futures position is chosen based on the long leg's beta but this is not required
2. If the strategy involving only the long leg has the best performance, forget about shorting.
3. If the performance when hedging with index futures is close to that of the full long-short strategy, it is better to use the strategy with the index futures.

Implementation: Issues Arising with Short Positions

(continued)

4. If the performance using the index futures is significantly worse than that of the full long-short strategy, the full long-short strategy is preferred, provided that its implementation is realistic:
   - Try to get a sense of whether you are likely to encounter problems to borrow the securities or to face short selling bans.
   - If this is a concern you can consider synthesizing shorting in different ways:
     - By using option positions (long put, short call). The concern is liquidity and transaction costs.
     - By selling the index futures and buying all the index members which you do not want to short. This works but only on a limited scale because it ties up a lot of capital.
     - By shorting ETFs or buying short ETFs in which the stocks you want to short have a large weight. This raises its own issues which we consider in the next section.

Implementation: Overview

Implementation: Use of Levered and Inverse ETFs

- Levered and inverse ETFs are useful but should be used with care.
- The chart below shows the returns on three ETFs, IYF (1X long the Dow Jones U.S. Financial Sector Index), UYG (Ultra Financials, 2X long) and SKF (Ultrashort Financials, 2X short).

- Over the two-year period considered, buying IYF would have resulted in a 50% loss, but buying SKF would not have resulted in a 100% gain. Rather, it would also have resulted in a loss.

Implementation: Use of Levered and Inverse ETFs

- In the figure, both the levered long and the levered short ETFs are worth less than the underlying index (proxied by IYF) at the end of the period.
- Does this mean that something is wrong with levered ETFs and that it is possible to earn profits by simply shorting them?
- The answer is no. These ETFs actually do what they say, but they are **rebalanced dynamically on a daily basis**.
- For example, at the end of a day where the **underlying index price went up**:
  - the **levered ETF** will **increase its position** in the index;
  - the **short ETF** will **reduce its (short) position** in the index.

Implementation: Use of Levered and Inverse ETFs

- Hence, levered and inverse ETFs will tend to
  - do well in trending markets; and
  - do poorly in markets that exhibit reversals.
- The table below reports the expected gross return from a one-year investment in each type of ETF (for $\mu = 5\%$, $\sigma = 20\%$) depending on daily return autocorrelation ρ.

Autocorrelation	Long 1X	Long 2X	Long 3X	Short 1X	Short 2X	Short 3X
$\rho = -0.25$	1.0423	1.0692	1.0797	0.9443	0.8777	0.8029
$\rho = 0$	1.0510	1.1045	1.1607	0.9515	0.9053	0.8615
$\rho = 0.25$	1.0647	1.1637	1.3055	0.9643	0.9547	0.9703

Implementation: Use of Levered and Inverse ETFs

- Hence, if you hold a levered or short ETF to proxy for leverage or shorting as part of a model that has a holding period exceeding one day, you need to **adjust the size of the ETF position over time to undo the effect of the dynamic rebalancing inside the ETF**. (You obviously don't need to rebalance if you backtested using the ETF.)
- Note that this problem arises with levered and inverse ETFs. Long and short positions in single long ETFs and futures do not require rebalancing before the end of the model's holding period.

Overview

Summary: Sources of Returns, Types of Strategies

- The goal when developing a strategy is to obtain the **highest possible return for a given level of risk**.
- Before devising a strategy, think about **where your returns are coming from** (e.g. a risk premium, timing ability, a market inefficiency) and **what kinds of risks you are willing to bear**.
- The main types of strategies are:
  - **Constant allocation strategies** earning unconditional risk premia.
  - **Timing strategies** earning conditional risk premia.
  - **Selection strategies** earning the returns provided by strategy styles.

Summary: Data Collection

- It is important to understand and check the data before using it. Common issues are:
  - Publication dates that are after the reference date.
  - Outliers, data errors, missing values.
  - Biases in the selection of data (look-ahead bias, selection bias, backfilling bias, survivorship bias).
  - Prices that are not adjusted for splits, dividends, and interest.
  - Futures prices that are not adjusted for rollovers.

Summary: Returns, Performance Measurement

- When computing the **one-period return of a portfolio** of assets, it is easier to use **ordinary** returns.
- When **aggregating returns over time**, it is easier to use **log** returns.
- The return on a portfolio comprising **futures** contracts can be computed as usual, but the **weights need not sum to one**.
- To assess a portfolio's **performance**:
  - Plot an equity line;
  - Generate a histogram of portfolio returns;
  - Compute (at least) mean, standard deviation, Sharpe ratio, skewness, kurtosis, maximum, minimum, factor exposures, alpha, and autocorrelation.
- One can assess **timing ability** by investigating the **convexity** of the relation between market returns and portfolio returns.

Summary: Principles of Strategy Development

- The key **principles of strategy development** are:
  1. Always worry about statistical significance.
  2. Beware of overfitting / data mining.
  3. Beware of biases.
  4. Account for transaction costs.
  5. Implement what you tested.
- The factors impeding the correct assessment of statistical significance can be addressed as follows:
  - **Nonspherical residuals** can be addressed by using **robust standard errors**.
  - **Overlapping returns** can be addressed by using **Newey-West standard errors**.
  - Problems associated with **persistent predictors** can be addressed by differencing or removing the cyclical component.
- **Overfitting** can be addressed by running few trials, using long sample periods, and running out-of-sample tests.
- One can assess the impact of **transaction costs** on strategy returns by tracking the strategy's **turnover**.

Summary: Risk Premia, Timing Models, Selection Models

- **Unconditional risk premia** can be estimated as the difference between the average return on the risky asset and the average return on the riskless asset. Since average return estimates are noisy, one must use long periods.
- When developing a **timing model**, one seeks indicators for which the assets' **conditional distribution** differs from the unconditional one. This can be assessed in a number of ways:
  1. By looking at subsamples constructed based on the indicators' value.
  2. By running predictive regressions using the indicators as explanatory variables.
  3. By running predictive regressions using dummy variables constructed based on the indicator values as explanatory variables.
- **Selection models** attempt to generate returns by selecting assets from a cross-section based on some criteria. The most common methods are:
  1. Sorts (univariate or multivariate);
  2. Scoring;
  3. Regression.

Summary: Combining Strategies

- The risk of overfitting is large when combining strategies.
- Instead of portfolios based on optimization such as the tangency portfolio or the global minimum variance portfolio, one will often choose to use allocations based on heuristics (equally weighted portfolios or risk parity portfolios).
- One should also consider the skewness and the worst return of the overall portfolio.

Summary: Implementation

- Factors that must be considered when selecting the best way to **implement** a strategy include
  - access,
  - convenience,
  - additional risks such as counterparty risk,
  - liquidity/transaction costs,
  - robustness, and
  - taxes.
- In addition, one should:
  - Be careful when **trading at the open or at the close**.
  - Analyze the amount of **leverage** that is desired and possible without being exposed to liquidation risk.
  - Investigate **alternatives to shorting** securities that must be borrowed.
  - **Rebalance** positions on a regular basis when using **levered or inverse ETFs**.

References I

- Bailey, David H., Jonathan M. Borwein, Marcos Lopez de Prado, and Qiji Jim Zhu (2014): Pseudomathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-sample Performance. *Notices of the American Mathematical Society* 61(5), 458–471.
- Campbell, John Y. and Samuel B. Thompson (2008): Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?, *Review of Financial Studies* 21(4), 1509–1531.
- Clark, Todd E. and Kenneth D. West (2007): Approximately Normal Tests for Equal Predictive Accuracy in Nested Models, *Journal of Econometrics* 138, 291–311.
- Ilmanen, Antti (2011): *Expected Returns: An Investor's Guide to Harvesting Market Rewards*. Chichester, U.K.: John Wiley & Sons.
- Ilmanen, Antti (2012): *Expected Returns on Major Asset Classes*. Chichester, U.K.: John Wiley & Sons.

References II

- Kendall, M. G. (1954): Note on Bias in the Estimation of Autocorrelation, *Biometrika* 41, 403–404.
- Newey, Whitney K. and Kenneth D. West (1987): A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703–708.
- Novy-Marx, Robert (2014): Predicting Anomaly Performance with Politics, the Weather, Global Warming, Sunspots, and the Stars, *Journal of Financial Economics* 112, 137–146.
- Petersen, Mitchell A. (2009): Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches, *Review of Financial Studies* 22(1), 435–480.
- Stambaugh, Robert F. (1999): Predictive Regressions, *Journal of Financial Economics* 54, 375–421.