1. Return Predictability

Contents

1 Learning objective:

- **Reading list:** Fama and French (1989), Goyal and Welch (2008), Campbell and Thomson (2008), Lecture notes
- Knowledge and understanding of return predictability in financial markets, its role in asset pricing, and implications for the dynamics of asset prices
- Skills to discuss and implement in-sample and out-of-sample return predictability methods, evaluate the empirical findings, and reflect on their implications for the dynamics of asset prices

2 Suggested presentation 10 min

1. Empirical

Part I
Short version

3 Return predictability

3.1 Introduction

1. Question of whether risk premia as excess returns on financial assets can be predicted
2. Naturally interesting ⇒ better investment performance + answer theoretical asset pricing questions
3. Contradictory authors ⇒ 1. Document predictability + 2. Prediction is spurious
   (a) Predictability time varying in itself

3.2 Base equation

Regression form:

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

If $\beta \neq 0 \Rightarrow$ predictability
Seems simple, but not always easy.
Troublesome predictors ⇒ econometrical problems, e.g. persistence as we shall see.

3.3 Campbell-Shiller representation?

3.4 Stambaugh bias

Two-equation system

$$\begin{aligned} r_{t+1} &= \alpha + \beta x_t + \varepsilon_{t+1} \\ x_{t+1} &= \lambda + \rho x_t + \varepsilon_{t+1} \end{aligned}$$

$\hat{\rho}$ is downward bias in a finite sample.
This biases $\hat{\beta}$ as:

$$\mathbb{E}\left[\hat{\beta} - \beta\right] \approx -\underbrace{\gamma}_{\frac{\sigma_{\varepsilon v}}{\sigma_\varepsilon}} \underbrace{\left(\frac{1+3\rho}{T}\right)}_{\mathbb{E}[\hat{\rho}-\rho]}$$

Errors negatively correlated -> $\gamma < 0$, $\downarrow \mathbb{E}[\hat{\rho} - \rho] = \uparrow \mathbb{E}\left[\hat{\beta} - \beta\right]$
Errors positively correlated -> $\gamma > 0$, $\uparrow \mathbb{E}[\hat{\rho} - \rho] = \downarrow \mathbb{E}\left[\hat{\beta} - \beta\right]$
Bias increases in γ and ρ, decreases in T

3.5 Solutions

1. Bootstrapping
   (a) Evaluate the β in a distribution suffering from same bias
   (b) Going into detail with this
2. Lewellens defense
   (a) Realizes max bias at $\rho = 1$
   (b) Adjust coefficients for bias
   (c) Get same coefficients, but much lower bias
   (d) Staumbaugh bias conditioned on estimated persistence and true persistence:

$$\mathbb{E}\left[\hat{\beta} - \beta | \hat{\rho}, \rho\right] = \gamma \left[\hat{\rho} - \rho\right]$$

$$\hat{\beta}_{Adj.} = \hat{\beta} - \gamma \left[\hat{\rho} - 1\right]$$

$$Var\left[\hat{\beta}_{Adj.}\right] = \frac{\sigma_\eta^2}{(T\sigma_x^2)}$$

3. Cochrane's defense
   (a) If returns are <u>not</u> predictable, dividend growth <u>must</u> be predictable. But it's not.
   (b) System:

$$\begin{aligned} \Delta d_{t+1} &= \alpha_d + \beta_d \left(d_t - p_t\right) + \varepsilon_{d,t+1} \\ r_{t+1} &= \alpha_r + \beta_r \left(d_t - p_t\right) + \varepsilon_{r,t+1} \\ d_{t+1} - p_{t+1} &= \lambda + \rho \left(d_t - p_t\right) + \varepsilon_{dp,t+1} \end{aligned}$$

   Parameter link:

$$\beta_r = 1 + \beta_d - \phi\rho$$

   If e.g. $\phi = 0.97$ and we know that $\rho \leq 1$, then

$$\beta_r > \beta_d + 0.03$$

4. Campbell & Thompson
   (a) Restrict the predictive formula
   (b) Both the excess return and slope
   (c) Find much higher OOS R^2

3.5.1 Bootstrapping

Have $\mathcal{H}_0 : \beta = \beta_0$

1. Estimate parameters by OLS
2. Generate T random numbers of errors from multivariate normal distribution
3. Generate random initial value of x_0 from mean and variance of predictor
4. Use estimated coefficietns to obtain T observations of $x_{1,\ldots T}$
5. Use the estimated constant with β_0 and the generated predictor values and errors to get T estimated returns
6. Estimate predictive regression from return formula on simulated data to obtain first $\tilde{\beta}^{(1)}$

7. Repeat to get more $\tilde{\beta}^{(1,\dots,M)}$

Can now compute the correct critical values and p-values.
Nonparametric -> Draw (with replacement) residuals and predictor values.
P values under null hypothesis:

$$\mathbb{P}\left[\tilde{\beta} > \hat{\beta}\right] = \frac{1}{M}\sum_{i=1}^{M} 1\left\{\tilde{\beta}^{(i)} > \hat{\beta}\right\}$$

$$\text{Bias}\left(\hat{\beta}\right) = \frac{1}{M}\sum_{i=1}^{M}\tilde{\beta}^{(i)} - \beta_0$$

3.6 Out of sample predictability

1. Split in-sample and out-of-sample
   (a) Rolling or expanding
2. Benchmark: Historical average
   (a) Gu et al: Artifcially "easy" due to noise
3. $MSFE = \frac{1}{T-R}\sum_{i=R+1}^{T}(r_i - \hat{r}_i)^2$
4. $R^2_{OOS} = 1 - \frac{MSFE_x}{MSFE_{HA}} = 1 - \frac{\sum_{i=R+1}^{T}(r_i-\hat{r}_i)^2}{\sum_{i=R+1}^{T}(r_i-\bar{r}_i)^2}$
   (a) $R^2_{OOS} > 0$ -> predictive regression has lower avg. MSFE than benchmark
5. Test for $\mathcal{H}_0 : R^2_{OOS} \leq 0$ vs $\mathcal{H}_1 : R^2_{OOS} > 0$
   (a) Diebold-mariano
      i. Loss differentials, d_i
         A. $d_i = (r_i - \bar{r}_i)^2 - (r_i - \hat{r}_i)^2$
      ii. Run regression to test $\mathcal{H}_0 : \mathbb{E}[d_i] \leq 0$
         A. $d_i = \theta + \varepsilon_i$
      iii. Standard t-test with Newey West standard errors to evaluate $\theta \leq 0$
   (b) Clark and West
      i. Diebold-Mariano severely undersized for nested models due to an additional parameter being zero under the null, leading to underperfroamnce
      ii. New MSFE-adjustment for errors:
         A. $f_i = (r_i - \bar{r}_i)^2 - \left[(r_i - \hat{r}_i)^2 - (\bar{r}_i - \hat{r}_i)^2\right]$
      iii. Run regression to test $\mathcal{H}_0 : \mathbb{E}[f_i] \leq 0$
         A. $f_i = \theta + \epsilon_i$
      iv. Standard t-test with NW std. errors to evaluate $\theta \leq 0$
         A. Test has approximate standard normal asymptotics for nested models and good small sample properties

Part II

Long version

4 Introduction

5 Return predictability

5.1 Introduction

Return predictability is concerned with the question of whether risk premia, in the shape of excess returns on financial assets, are predictable as the risk free rate may vary over time. It is a naturally interesting topic as predictability would yield benefits in better asset pricing models, investment performance and generally answering theoretical asset pricing questions.
There are both papers and authors that document predictability while others argue that predictability is purely spurious:

- Documenting predictability:
  - Stocks: Campbell and Shiller (1988), Fama and French (1989)
  - Bonds: Fama and Bliss (1987), Campbell and Shiller (1991)
- Argue that predictability is spurious and performs poorly out-of-sample:
  - Stocks: Stambaugh (1986, 1999), Nelson and Kim (1993), Goyal and Welch (2008)
  - Bonds: Thornton and Valente (2012), Sarno et al. (2016)

This may in part be explained by predictability itself varying over time, as the papers did not appear at exactly the same time.

5.2 Base equation

On regression form, return predictability is typically studied using the multivariate predictive regression model:

$$r_{t+1} = \alpha + \boldsymbol{x'}_t \boldsymbol{\beta} + \varepsilon_{t+1}$$

If excess returns are predictable, then expected excess returns vary over time as a function of x_t

$$\mathbb{E}_t \left[r_{t+1} \right] = \widehat{\alpha} + \boldsymbol{x}_t' \widehat{\boldsymbol{\beta}}$$

where we say that if $\boldsymbol{\beta} \neq 0$, we can speak of predictability, and thus this is what we want to test on.
Many predictors happen to be highly persistent, though, giving rise to some annoying econometric challenges and small sample bias, which we'll get into.

- Excess returns contains a predictable an unpredictable part $\Rightarrow$ it can be hard to determine the unpredictable part.
- How to deal with model uncertainty and instability?
- We do not know "The Model" or the data generating process (DGP) for returns

5.3 Campbell-Shiller loglinear approximation

Start from the definition of returns, taking log and rewriting *(further explanation of this derivation is available in the Return Predictability note)*

$$\begin{aligned} R_{t+1} &= \frac{P_{t+1} + D_{t+1}}{P_t} \\ r_{t+1} &= \ln \left(P_{t+1} + D_{t+1} \right) - \ln \left(P_t \right) \\ &= p_{t+1} - p_t + \ln \left(1 + \exp \left(d_{t+1} - p_{t+1} \right) \right) \end{aligned}$$

Where the last term is a nonlinear function of the log dividend price. Taking a first-order Taylor approximation yields the approximation:

$$r_{t+1} \approx \kappa + \phi p_{t+1} + (1-\phi)\, d_{t+1} - p_t$$

with

$$\phi = \frac{1}{1+\exp\left(\overline{d-p}\right)}$$
$$\kappa = -\ln\phi - (1-\phi)\ln\left(1/\phi - 1\right)$$

Solving the approxmiation forward (inserting for the future price) and isolating for p_t, using the transversaility condition $\lim_{j\to\infty} \phi^j p_{t+j} = 0$ will yield:

$$p_t = \frac{\kappa}{1-\phi} + \mathbb{E}_t\left[\sum_{j=0}^{\infty} \phi^j \left[(1-\phi)\, d_{t+1+j} - r_{t+1+j}\right]\right]$$

Where we see that if prices vary, those variations can come from changing expectations about future dividends, changing expectations about future returns, or both.
Can also be stated in terms of the log-dividend price ratio and iterating forward, yielding:

$$d_t - p_t = \frac{-\kappa}{1-\phi} + \mathbb{E}_t\left[\sum_{j=0}^{\infty} \phi^j \left[-\Delta d_{t+1+j} + r_{t+1+j}\right]\right]$$

This equation is central to the predictability debate because it implies that the log dividend- price must predict either future expected dividends, future expected returns, or a combination of the two. It also tells us that the log dividend-price ratio should predict returns with a positive coefficient. This comes from Campbell and Shiller (1988).

5.4 Stambaugh's finite sample bias

One econometric issue when testing predictability is however as found by Stambaugh that OLS estimates are subject to a small sample bias.
There might be a correlation between the errors and the predictor variable which can cause a bias.
Consider a system where predictor x_t follows AR(1) process

$$r_{t+1} = \alpha + \beta x_t + \varepsilon_{t+1}$$
$$x_{t+1} = \lambda + \rho x_t + \nu_{t+1}$$

Note that the error terms ϵ_{t+1} and ν_{t+1} are white white noise errors that are contemporaneously correlated with covariance $\sigma_{\epsilon\nu}$.

The OLS estimate $\hat{\rho}$ is downward-biased in a finite sample (Kendall, 1954).

$$\mathbb{E}\left[\hat{\rho} - \rho\right] \approx \left(\frac{1+3\rho}{T}\right)$$

The bias in $\hat{\beta}$ is then given by (Stambaugh (1986, 1999))

$$\mathbb{E}\left[\hat{\beta} - \beta\right] = \underbrace{\frac{\sigma_{\varepsilon\nu}}{\sigma_{\varepsilon}^2}}_{\gamma}\, \mathbb{E}\left[\hat{\rho} - \rho\right]$$

(SKIP BELOW)
Can now interpret γ as a regression coefficent

$$\varepsilon_{t+1} = \gamma \nu_{t+1} + \eta_{t+1}$$

The bias in $\hat{\beta}$ can be approximated as

$$\mathbb{E}\left[\hat{\beta}-\beta\right] \approx -\frac{\sigma_{\varepsilon\nu}}{\sigma_{\varepsilon}^2}\left(\frac{1+3\rho}{T}\right) = -\gamma\left(\frac{1+3\rho}{T}\right)$$

If the errors are negatively correlated, then $\gamma < 0$ and the downward bias in $\hat{\rho}$ produces an upward bias in $\hat{\beta}$. The bias will increase in γ and ρ, but decrease in sample size T - thus being a small sample bias.

5.5 Parametric bootstrap

In spite of the small sample bias, we can conduct valid inference using techniques such as bootstrapping, where we evaluate the biased coefficient in an empirical distribution suffering the same bias, meaning that inference will be valid.

- In other words we generate an empirical distribution of beta estimates under the null hypothesis - which here would naturally be $\beta = 0$ i.e. no predictability.
- This empirical distribution will be shifted by the bias and thus will allow us to do valid inference.
- It can be done using a parametric or residual based bootstrap
  - The parametric bootstrap makes a distributional assumption about the error terms which are re-sampled.
- Residual based bootstrap approach:
  - Do not rely on distributional assumptions.
  - Samples error terms from actual data by drawing T values of $(\epsilon_{t+1}, \nu_{t+1})$ with replacement from the set of obtained residuals.
  - Obtaining initial value of x_t by randomly drawing from the observed predictor variable.
  - The two steps above replaces step 2 and 3 in the steps outlined below.

Steps:

1. Estimate parameters by OLS
2. Generate T random numbers of errors from multivariate normal distribution
3. Generate a random initial value of the predictor of interest, based on the unconditional mean and variance of the predictor variable.
4. Use estimated coefficients with the generated errors to obtain T observations of the predictor.
5. Use the estimated constant with the hypothesized value β_0 with the generated values of the predictor and return errors to obtain T observations of estimated returns
6. Estimate the predictive regression from the return formula on the simulated data to obtain $\tilde{\beta}^{(1)}$
7. Repeat steps 2-6 M times to obtain $\tilde{\beta}^{(1)}, \tilde{\beta}^{(2)}, ..., \tilde{\beta}^{(M)}$

Can then compute the (upper) one-sided p-value under the null hypothesis as

$$\mathbb{P}\left[\tilde{\beta} > \hat{\beta}\right] = \frac{1}{M}\sum_{i=1}^{M} 1\left\{\tilde{\beta}^{(i)} > \hat{\beta}\right\}$$

or compute the correct critical value from the relevant percentiles of the distribution.

This approach automatically deals wtih bias since the mean of the empirical distribution moves with the size of the bias.

To estimate the bias we can simply calculate the mean deviation of simulated slope coefficients and the estimate β,

$$\text{Bias}\left(\hat{\beta}\right) = \frac{1}{M}\sum_{i=1}^{M}\tilde{\beta}^{(i)} - \beta_0$$

5.6 Lewellen's (2004) defense

Lewellen states that while predictive regressions are subject to small-sample biases, the correction used by prior studies can substantially understate forecasting power. He therefore uses a different test to show that dividend yield predicts market returns. Incorporating information about the sample autocorrelation of DY helps produce more powerful tests of predictability. Incorporating this information into empirical tests has two effects: (1) the slope estimate is often larger than the standard bias-adjusted estimate; and (2) the variance of the estimate is much lower. In combination, the two effects can substantially raise the power of empirical tests.

Conditioning the Stambaugh bias on the estimated persistence $\hat{\rho}$and the true persistence ρ

$$\mathbb{E}\left[\hat{\beta} - \beta | \hat{\rho}, \rho\right] = \gamma \left[\hat{\rho} - \rho\right]$$

we don't know ρ, but we know that $\rho = 1$ is where the maximum bias occurs as $d_t - p_t$ is not explosive. This gives us

$$\hat{\beta}_{Adj.} = \hat{\beta} - \gamma \left[\hat{\rho} - 1\right]$$

variance equal to

$$Var\left[\hat{\beta}_{Adj.}\right] = \frac{\sigma_\eta^2}{(T\sigma_x^2)}$$

These bias-adjusted coefficients are found to be similar to unadjusted coefficents, but with much lower variance.

5.7 Cochrane's (2008) defense

Defends return predictability by directing attention to the inability of the log dividend-price to predict dividend growth. The idea is that if price dividend cannot explain dividend growth, then it must explain returns. From the abstract:
If returns are not predictable, dividend growth must be predictable, to generate the observed variation in divided yields. I find that the absence of dividend growth predictability gives stronger evidence than does the presence of return predictability.

System:

$$\begin{aligned}
\Delta d_{t+1} &= \alpha_d + \beta_d (d_t - p_t) + \varepsilon_{d,t+1} \\
r_{t+1} &= \alpha_r + \beta_r (d_t - p_t) + \varepsilon_{r,t+1} \\
d_{t+1} - p_{t+1} &= \lambda + \rho (d_t - p_t) + \varepsilon_{dp,t+1}
\end{aligned}$$

Parameters linked as

$$\beta_r = 1 + \beta_d - \phi\rho$$

If $\phi = 0.97$ and we know that $\rho \leq 1$, then

$$\beta_r > \beta_d + 0.03$$

In the above system, Cochrane finds that $\beta_d \approx 0$, providing indirect evidence that $\beta_r > 0$, meaning that stock returns are predictable.

5.8 Long-horizon predictability

NOTE: Nævn dette i starten af præsentation, når 1-period return-predictability regression introduceres.
It may be relevant to forecast excess returns over a longer horizon.

- Strong connection between short- and long-horizon predictability.
- If we do not have predictability on the short horizon, then we will not have predictability on long horizon either.
- However, we can test the two independently.

- It is important to note that long horizon predictive regressions introduces the problem of overlapping samples which causes serial correlated errors ($\varepsilon_{t\to t+h} \sim MA(h-1)$)
  - Newey-West is a poor solution $\to$ used Hodrick (1992) or bootstrapping.
  - Bootstrapping is ideal as it deals with both serial correlation and small-sample-bias simultaneously.

When we want to forecast excess returns over a longer horizon, we consider the equation

$$r_{t\to t+h} = \alpha_h + \beta_h x_t + \varepsilon_{t\to t+h}$$

with h-period log excess return (risk premia) computed as:

$$\begin{aligned} r_{t\to t+h} = \ln(1 + R_{t\to t+h}) &= \sum_{i=1}^{h} \ln(1 + R_{t+i}) \\ &= \sum_{i=1}^{h} r_{t+i} \end{aligned}$$

Often end up using overlapping returns in studies of long-horizon predictability, due to data limitations. Now the errors are distributed differently, and the OLS standard errors will no longer be valid.

Boostrapping long-horizon regressions
Replace step 6 with:
6a. Use T simulated one-period returns (or risk premia) r^*_{t+1} to build multi-period returns (or risk premia) $r^*_{t\to t+h}$.
6b. Estimate multi-period model on the simulated data to obtain $\tilde{\beta}^1_h$
Requiers no estimate of standard error and thus automatically deals with the overlapping data problem, while automatically and simultaneously accounting for small-sample bias.

5.9 Out of sample predictability

We now move on to test whether investors can use real time predictive variables to make investment decisions. The sample is split into two parts; one in-sample part on which we base the forecast on and an out-of-sample part which we will have to predict. The main object of interest for evaluating the forecast is generally the forecast error which we compare to error of other benchmarks.

Split the sample to test returns in out-of-sample. Can be done either **rolling** or **expanding**.

- Rolling: an advantage if parameters vary over time
- Expanding: best for constant parameters, uses more information $\to$ has higher strength

The choice of split between in-sample and out-of-sample is important.

- Consider whether parameters are constant
- Long enough in-sample period to have high strength
- Long enough out-of-sample period to evaluate the forecasting ability

Natural benchmark: Historical average *(likewise estimated rolling or expanding)*

$$\bar{r}_{t+1} = \frac{1}{t}\sum_{i=1}^{t} r_i$$

Mean squared forecast error (MSFE):

$$MSFE = \frac{1}{T-R}\sum_{i=R+1}^{T} (r_i - \hat{r}_i)^2$$

R^2_{OOS} (proportional reduction in MSFE for the predictive regression relative to HA):

$$R^2_{OOS} = 1 - \frac{MSFE_x}{MSFE_{HA}} = 1 - \frac{\sum_{i=R+1}^{T} (r_i - \hat{r}_i)^2}{\sum_{i=R+1}^{T} (r_i - \overline{r}_i)^2}$$

$R^2_{OOS} > 0$ means that the predictive regression has lower average MSFE than HA, implying that the predictor contains relevant relevant information not contained in the HA. Therefore we'd like to test whether this is the case as:

$$\mathcal{H}_0 : R^2_{OOS} \leq 0$$
$$\mathcal{H}_1 : R^2_{OOS} > 0$$

Can be tested in two ways:

5.9.1 Diebold-Mariano

Construct series of loss differentials, d_i
$d_i = (r_i - \overline{r}_i)^2 - (r_i - \hat{r}_i)^2$
Then test $\mathcal{H}_0 : \mathbb{E}[d_i] \leq 0$ by running the regression

$$d_i = \theta + \varepsilon_i$$

and perform standard t-test on θ using Newey West standard errors to evaluate the null $\theta \leq 0$.
The test has standard asymptotic distribution for non-nested models, but is severely **undersized for nested models**.

5.9.2 Clark and West

MSFE-adjusted test for nested models. Construct:

$$f_i = (r_i - \overline{r}_i)^2 - \left[(r_i - \hat{r}_i)^2 - (\overline{r}_i - \hat{r}_i)^2\right]$$

With the last term adjusting for the fact that we'd expect the predictive regression to underperform as there is an additional parameter that is zero under the null.

Then test: $\mathcal{H}_0 : \mathbb{E}[f_i] \leq 0$ by running the regression of $f_i = \theta + \epsilon_i$ and perform standard t-test using NW std. errs. to evaluate the null $\theta \leq 0$. This test has approximate standard normal asymptotics for nested models and good small sample properties.

5.9.3 Goyal Welch graphical device (CDSFE)

Goyal Welch graphical device: cumulative difference in squared forecast errors (CDSFE)

$$CDSFE_t = \sum_{i=R+1}^{t} (r_i - \overline{r}_i)^2 - \sum_{i=R+1}^{t} (r_i - \hat{r}_i)^2$$

Informative about the timing of the value of predictor information

- **Positive slope:** Predictive models outperforms the benchmark in terms of MSFE
- **Negative slope:** Predictive models underperforms the benchmark in terms of MSFE

Goyal and Welch (2008) further finds limited evidence of in-sample predictability and virtually no out-of-sample predictability over time and argue that what limited evidence we might see is fully attributable to the 1974 oil crisis - the graphical device is thus sensitive in times of economic unrest.

5.10 Campbell and Thompson (2008)

Defends out of sample predictability by showing that by imposing simple constraints improves the out-of-sample performance. E.g. expected excess returns (risk premia) should be non-negative while slope coefficients should align with theory.

$$\hat{r}_{t+1} = \max\left\{0, \hat{\alpha} + \max\left\{0, \hat{\beta}\right\} x_t\right\}$$

Where we can take one or the other of these restrictions, or do them jointly as above. This is shown to generate significant improvements in R^2_{OOS}

5.11 Additional Notes

There are several drawbacks using only a single predictor.

- Hard to select predictors á priori
- $1/N$ strategy is not necessarily the best, i.e. use different weighting schemes
- Individual predictors may by particularly unstable over time

Alternatives include forecast combination or machine learning techniques for dimension reduction, shrinkage or variable selection.
It is possible to compute the utility gains of predictability by imposing a utility function on the investor.
Finally it is worth noting that the mixed empirical evidence for predictability of asset returns might be due to predictability itself varying over time. There is ex post empirical evidence supporting time-varying predictability of returns in different asset classes.

5.12 Goyal and Welch (2008)

- comprehensively reexamines the performance of variables that have been suggested by the academic literature to be good predictors of the equity premium.
- find that by and large, these models have predicted poorly both in-sample (IS) and out-of-sample (OOS) for 30 years
- these models seem unstable, as diagnosed by their out-of-sample predictions and other statistics
- these models would not have helped an investor with access only to available information to profitably time the market.