Recent Empirical Studies on Return Predictability

WANG, Yanchu
Shanghai University of Finance and Economics

October 2024

Motivation

- Forecasting stock returns (risk premium) is of great interests to both academics and practitioners.
- Economic theory indicates that expected returns should be time-varying and predictable.
- However, empirical findings fail to provide such evidence, especially for out-of-sample benefits.
- Welch and Goyal (2008) conclude that numerous economic variables with in-sample predictability fail to deliver out-of-sample gains.
- The failure may be due to:
  1. Model uncertainty and instability (Rapach, Strauss, and Zhou, 2010).
  2. Regime uncertainty and parameter uncertainty (Zhu and Zhu, 2013).
  3. State variable uncertainty (Binsbergen and Koijen, 2010).□

Rapach, Strauss, and Zhou (2010)

- Previous findings.
- Combination methods.
- Forecast evaluation.
- Empirical Results.
- Macroeconomic links.
- Conclusion

Previous Findings.

- Return predictability is a long-history topic in finance.
- A set of variables are considered to be helpful in forecasting returns:
  1. Firm-level valuation ratios: such as dividend-price, earning price, book-to-market etc.
  2. Economic activity indicators: such as interest rate, inflation rate, term and default spread etc.
  3. Other market data: consumption-wealth ratio, stock market volatility etc.
- Most existing studies focus on in-sample tests and conclude that there is significant evidence of return predictability.
- But it remains controversial, see the review paper of Spiegel (2008).

Combination Forecasts

- Consider two predictive variables: dividend yield and term spread.
- Previous literature shows that both variables can predict stock returns.
- But each variable alone may capture different components of economic conditions, and thus may give a false signal during certain periods.
- If these two variables are correlated, then a combination of them should be less volatile and provide better performance.
- We may extend the idea to numerous factors.

Combination Forecasts

- The combination forecasts are linked to macroeconomics.
- Fama and French (1989) and Cochrane (1999, 2007) argue that heightened risk aversion during economic recession require a higher risk premium.
- The combination forecasts reach local maxima very near NBER-dated business-cycle troughs.
- The forecasting power becomes stronger when economy is bad.
- The instability of individual predictive factors is related to instability in real economy.

Predictive Regression Model

- A standard predictive regression model:

$$r_{t+1} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{t+1},$$

  where r_{t+1} is the return on a stock market index in excess of the risk-free rate, and $x_{i,t}$ is a predictive variable.
- The out-of-sample forecasts are generated using a recursive estimation window.
- The full sample T is divided into a in-sample portion with m observations, and out-of-sample portion with q observations. $T = m + q$.

Predictive Regression Model

- The first forecast is given by
$$\widehat{r}_{i,m+1} = \widehat{\alpha}_{i,m} + \widehat{\beta}_{i,m} x_{i,m},$$
where $\widehat{\alpha}_{i,m}$ and $\widehat{\beta}_{i,m}$ are the OLS estimates of α_i and β_i generated by regressing $\{r_t\}_{t=2}^{m}$ on a constant and $\{x_{i,t}\}_{t=1}^{m-1}$.
- The next forecast is given by
$$\widehat{r}_{i,m+2} = \widehat{\alpha}_{i,m+1} + \widehat{\beta}_{i,m+1} x_{i,m+1},$$
where $\widehat{\alpha}_{i,m+1}$ and $\widehat{\beta}_{i,m+1}$ are the OLS estimates of α_i and β_i generated by regressing $\{r_t\}_{t=2}^{m+1}$ on a constant and $\{x_{i,t}\}_{t=1}^{m}$.
- Continue the procedure, we finally generate q forecasts based on $x_{i,t}$.
- We generate return forecasts using 15 individual factors (with details later).
- The historical average benchmark: $\overline{r}_{t+1} = \sum_{j=1}^{t} r_j$.
- If $x_{i,t}$ contains useful information, then $\widehat{r}_{t+1}$ should perform better than $\overline{r}_{t+1}$.

Forecast Combinations

- Bates and Granger (1969) point out that combination of individual factors can outperform individual factors.
- Forecast combination has recently received attention in macroeconomic forecasting literature (Stock and Watson, 1999, 2003, 2004).
- The combination forecasts of r_{t+1} is given by

$$\widehat{r}_{c,t+1} = \sum_{i=1}^{N} \omega_{i,t} \widehat{r}_{i,t+1},$$

  where $N = 15$, and $\{\omega_{i,t}\}_{i=1}^{N}$ are the *ex ante* combining weights formed at time t.
- The first q_0 observations from the out-of-sample period are used as the initial holdout period for estimating $\omega_{i,t}$.
- Thus we have $q - q_0$ combining forecasts in total.

Forecast Combinations

- How to estimate $\omega_{i,t}$?
- The mean combination forecast: $\omega_{i,t} = 1/N$.
- In practice, a trimmed mean combination forecast:

$$\omega_{i,t} = 0,$$

with the smallest and largest values, and

$$\omega_{i,t} = 1/(N-2),$$

with the remaining individual forecast.

Forecast Combinations

- The discount mean square prediction error (DMSPE) combining forecast based on Stock and Watson (2004):

$$\omega_{i,t} = \phi_{i,t}^{-1} / \sum_{j=1}^{N} \phi_{j,t}^{-1},$$

where

$$\phi_{i,t} = \sum_{s=m}^{t-1} \theta^{t-1-s} (r_{s+1} - \widehat{r}_{i,s+1})^2.$$

- The DMSPE method assigns greater weights to individual forecasts with lower MSPE values.
- When $\theta < 1$, greater weight is attached to the recent forecast accuracy.
- $\theta = 1.0$ or 0.9 in empirical analysis.

Forecast Evaluation

- Out-of-sample R^2_{OS} based on Campbell and Thompson (2008)

$$R^2_{OS} = 1 - \frac{\sum_{k=q_0+1}^{q} (r_{m+k} - \widehat{r}_{m+k})^2}{\sum_{k=q_0+1}^{q} (r_{m+k} - \overline{r}_{m+k})^2}.$$

- R^2_{OS} compares the specified forecast performance with the historical average forecasts.
- If $R^2_{OS} > 0$, then $\widehat{r}_{t+1}$ is better than $\overline{r}_{t+1}$.

Forecast Evaluation

- Even if $R^2_{OS} > 0$, we don't know its significance.
- In order to test that, we follow Clark and West (2007) and first define

$$f_{t+1} = (r_{t+1} - \overline{r}_{t+1})^2 - [(r_{t+1} - \widehat{\overline{r}}_{t+1})^2 - (\overline{r}_{t+1} - \widehat{\overline{r}}_{t+1})^2].$$

- We the regress $\{f_{s+1}\}_{s=m+q_0}^{T-1}$ on a constant and calculate the t-statistic.
- This is called MSPE-adjusted statistic and provide a one-sided (upper-tail) test.

Forecast Evaluation

- Even if R^2_{OS} is significantly positive, its values are usually small for predictive regression models.
- This raises the issue of economic significance.
- Let's consider the utility gain for a mean-variance investor who uses predictive returns to construct his optimal portfolio between stocks and risk-free bonds.
- It can be proved that the optimal weight for stocks is

$$w_t = (\frac{1}{\gamma})(\frac{r_{t+1}}{\sigma_{t+1}}).$$

- To be realistic, let w_t lie between 0% and 150%.

Forecast Evaluation

- If the investor use historical average forecast, then at time t, he need to decide the weight applied for $t+1$

$$w_{0,t} = (\frac{1}{\gamma})(\frac{\bar{r}_{t+1}}{\widehat{\sigma}^2_{t+1}}),$$

where $\widehat{\sigma}^2_{t+1}$ is the rolling-window estimate of the variance of stock returns.

- Over the out-of-sample period, the realized average utility level is

$$\widehat{v}_0 = \widehat{\mu}_0 - 0.5\gamma\widehat{\sigma}^2_0,$$

where $\widehat{\mu}_0$ and $\widehat{\sigma}^2_0$ are the sample mean and variance of the portfolio constructed using the weights over the out-of-sample period.

Forecast Evaluation

- We then compute the utility gain with individual factor or combined factor, j :
$$w_{j,t} = (\frac{1}{\gamma})(\frac{\widehat{r}_{t+1}}{\widehat{\sigma}^2_{t+1}}).$$
- The utility gain is
$$\widehat{v}_j = \widehat{\mu}_j - 0.5\gamma\widehat{\sigma}^2_j.$$
- The certainty equivalent return (CER) is the difference between $\widehat{v}_j$ and $\widehat{v}_0$.
- We choose $\gamma = 3$.

Data

- Following Welch and Goyal (2008), the paper considers 15 variable (quarterly) for 1947:1 - 2005:4:
  1. Dividend-price ratio (log): D/P
  2. Dividend yield (log): D/Y
  3. Earnings-price ratio (log): E/P
  4. Dividend-payout ratio (log): D/E
  5. Stock variance: sum of squared daily returns on the S&P 500 index
  6. Book-to-market ratio: B/M
  7. Net equity expansion: twelve-month sum of net issues to total end-of-year market cap
  8. Treasury bill rate

Data

- 9. Long-term yield
  10. Long-term return
  11. Term spread: difference between long-term yield and the Treasury bill rate
  12. Default yield spread: difference between BAA- and AAA- corporate bond yield
  13. Default return spread: difference between long-term corporate bond and long-term government bond returns
  14. Inflation
  15. Investment-to-capital ratio
- Consider three different out-of-sample periods: 1965:1-2005:4, 1976:1-2005:4, and 2000:1-2005:4.

Forecasting Results

- Cumulative square prediction error for individual factors vs. historical average benchmark (Figure 1).
- Cumulative square prediction error for combining forecasts vs. historical average benchmark (Figure 2).
- Detailed results (Table 1).

Robustness Checks (Selected)

- Factor stabilization:
  1. Correlation between individual factors (Table 3)
  2. Combining forecasts reduce forecast variability (Figure 3).
  3. "Kitchen sink" Model

$$r_{t+1} = \alpha + \beta_1 x_{1,t} + \dots + \beta_N x_{N,t} + \varepsilon_{t+1}.$$

  It performs much worse than the combination method!
- It can be shown that the mean combination is a restricted forecast from a multiple regression model.

Links to Real Economy

- Equity premium forecasts and NBER-dated business-cycle phases (Figure 6).
- Correlation of equity premium forecasts and GDP growth (Table 4).
- Forecasting gains during "good" and "bad" growth periods (Table 5).

Conclusions

- Forecast combining method provides convincing evidence of out-of-sample predictability.
- It is successful since it is a compromise:
  1. Individual forecasts are too volatile.
  2. The historical average is too smooth.
- The result shows that existing asset pricing models relying on one or a few state variables will have difficulty in accurately tracking the expected return over time.