Factor Investing: Hierarchical Ensemble Learning

Guanhao Feng[1] Jingyu He [2]

[1]City University of Hong Kong

[2]Chicago Booth

R/Finance 2019

Research Questions and Methods

Cross-section return prediction

- I Hierarchical prior

Factor investing and predictability

- II Posterior ensemble learning

Dynamic portfolio choice

- III Predictive system

Motivation I: Hierarchical prior

- A long history in finance research about predicting the index, see Goyel and Welch (2008).
- A recent literature in machine learning prediction for individual stocks, see Gu, Kelly and Xiu (2018).
  - Time Series Model: $R_{\mathbf{i},t+1} = f_{\mathbf{i}}(X_t)$
  - Pooled Model: $R_{\mathbf{i},t+1} = F(X_t)$

Motivation I: Hierarchical prior

- A long history in finance research about predicting the index, see Goyel and Welch (2008).
- A recent literature in machine learning prediction for individual stocks, see Gu, Kelly and Xiu (2018).
  - Time Series Model: $R_{i,t+1} = f_i(X_t)$
  - Pooled Model: $R_{i,t+1} = F(X_t)$

We want to explore the model heterogeneity with a large sample size.

- Hierarchical Model: $f_i(\cdot) \sim F(\cdot)$

Motivation II: Posterior ensemble learning

- Ensemble learning: $\tilde{f}_{\mathbf{i}}(X_t) = \frac{1}{K}\sum_{k=1}^{K} f_{\mathbf{i}}^{(k)}(X_t)$
- This modern machine learning, to reduce prediction error, has a Bayesian history.
  - Parameter uncertainty: Posterior Prediction
  - Model uncertainty: Bayesian Model Averaging

Motivation II: Posterior ensemble learning

- Ensemble learning: $\tilde{f}_{\mathbf{i}}(X_t) = \frac{1}{K}\sum_{k=1}^{K} f_{\mathbf{i}}^{(k)}(X_t)$
- This modern machine learning, to reduce prediction error, has a Bayesian history.
  - Parameter uncertainty: Posterior Prediction
  - Model uncertainty: Bayesian Model Averaging

We mainly focus on reducing the parameter uncertainty by posterior ensemble learning.

Motivation III: Predictive system

- The predictive model of macro predictors is time-varying.
  long-term yield, inflation, stock market variance, ...
- Alpha and beta are predictable on fundamental characteristics.
  dividend yield, accrual, gross profit, ...
- Most machine learning prediction only provides the predictive returns, while we still need the predictive covariance matrix.
  This is why Bayesian modeling is powerful for asset allocation.

Motivation III: Predictive system

- The predictive model of macro predictors is time-varying.
  long-term yield, inflation, stock market variance, ...
- Alpha and beta are predictable on fundamental characteristics.
  dividend yield, accrual, gross profit, ...
- Most machine learning prediction only provides the predictive returns, while we still need the predictive covariance matrix.
  This is why Bayesian modeling is powerful for asset allocation.

We provide a model of seemingly unrelated regressions.

Our solution: Hierarchical Ensemble Learning (HEL)

Following Polson and Tew (1999), we adopt a hierarchical structure in an SUR framework for the cross-sectional dependence among assets.

- A conditional predictive regression on macro predictors.
- Alpha and beta are driven by firm characteristics.
- Explore the joint predictability by a hierarchical framework.
- Reduce prediction error by the posterior ensemble learning.
- An SUR structure for both predictive returns and covariance matrix.

The R Package

BayesPortfolio

Factor Investing: Hierarchical Ensemble Learning.

The package will be posted on this website before end of May 19, 2019. Please check out the package later for update. Thanks!

jingyuhe.com/bayesportfolio

Hierarchical Ensemble Learning

Model Setup I

For each asset i and time period t, return of asset i at time period $t+1$ is modeled as

$$r_{i,t+1} = \alpha_{\mathbf{i},t} + x_t^\intercal \beta_{\mathbf{i},t} + \epsilon_{i,t+1}, \tag{1}$$

where x_t is the vector for Q macro predictors. Residual vector

$(\epsilon_{i,t+1}, \cdots, \epsilon_{N,t+1})^\intercal$ contains shocks to all asset returns and are assumed to follow a multivariate normal distribution $N(0, \Sigma)$ with full covariance matrix.

Model Setup II

Coefficients $\alpha_{i,t}$ and $\beta_{i,t}$ are assumed to be time-varying, driven by the asset characteristics as follows

$$\alpha_{\mathbf{i},t} = \eta_i^a + z_{\mathbf{i},t}^{\intercal}\theta_i^a, \quad (2)$$

$$\beta_{\mathbf{i},t} = \eta_i^b + \theta_i^b z_{\mathbf{i},t}, \quad (3)$$

where θ_i^b is a matrix coefficient of size $Q \times P$ and $z_{\mathbf{i},t}$ is the vector for P portfolio characteristics.

If we plug the time-varying coefficients into equation (1), we obtain an unconditional predictive regression on z_t, x_t, and their interactions $z_t \otimes x_t$:

$$r_{i,t+1} = \eta_i^a + z_{i,t}^{\intercal}\theta_i^a + x_t^{\intercal}\eta_i^b + \left(x_t \otimes z_{\mathbf{i},t}\right)^{\intercal}\theta_i^b + \epsilon_{i,t+1}. \quad (4)$$

Model Setup III

We simplify notations of our main equation (4) as

$$r_{i,t+1} \quad = \quad f_{i,t}^{\top} b_i + \epsilon_{i,t}, \tag{5}$$

where $f_{i,t} = [1, z_{i,t}, x_t, (x_t \otimes z_{i,t})]$, $\otimes$ denotes Kronecker product. b_i is a vector of all coefficients $b_i = [\eta_i^{\alpha}, \theta_i^{\alpha}, \eta_i^{b}, \theta_i^{b}]$.

For each asset i, stack equations of different time period t as

$$r_i \quad = \quad f_i^{\top} b_i + \epsilon_i, \tag{6}$$

where $r_i = (r_{i,2}, \cdots, r_{i,T+1})^{\top}$, $\epsilon_i = (\epsilon_{i,1}, \cdots, \epsilon_{i,T})^{\top}$, and f_i is a matrix with T rows.

Seemingly Unrelated Regressions

The SUR setup is organized by assets. Stacking all equations asset by asset, we have

$$R = FB + E, \tag{7}$$

where

$$R = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix}, \quad F = \begin{bmatrix} f_1 & 0 & 0 & 0 \\ 0 & f_2 & 0 & 0 \\ 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & f_N \end{bmatrix}, \quad B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix}, \quad E = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}.$$

Here, R is a $NT \times 1$ vector of stacked vector of firm returns, F is a $NT \times NK$ block diagonal matrix, B is an $NK \times 1$ vector, and E is $NT \times 1$ stacked vector of residuals.

Hierarchical Prior Structure

b_i is independent and identical draw from the following multivariate hierarchical prior

$$b_i \sim N(\bar{b}, \Delta_b), \quad \bar{b} \sim N(0, \Delta_{\bar{b}}), \quad \Delta_b \sim IW(\nu_b, V_b) \tag{8}$$

The likelihood function is multivariate normal:

$$l(E \mid B, \Omega) \propto |\Omega|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(R - FB)^\intercal \Omega^{-1}(R - FB)\right\}. \tag{9}$$

The joint posterior can be expressed as

$$p(B, \Omega, \bar{b}, \Delta_b \mid R, F) \propto l(E \mid B, \Omega)p(\Omega)p(B \mid \bar{b}, \Delta_b)p(\bar{b})p(\Delta_b). \tag{10}$$

Markov chain Monte Carlo Scheme I

(1) Update B through a multivariate normal distribution:

$$B \mid \bar{b}, \Delta_b, \Omega, R, F \sim N\Big(b^*, (F^\intercal \Omega^{-1} F + I_N \otimes \Delta_b^{-1})^{-1}\Big), \tag{11}$$

where $b^* = (F^\intercal \Omega^{-1} F + I_N \otimes \Delta_b^{-1})^{-1} \times (F^\intercal \Omega^{-1} R + (I_N \otimes \Delta_b^{-1})(\iota_N \otimes \bar{b}))$, and ι_N denotes an $N \times 1$ vector of ones.

(2) Update $\bar{b}$ through a multivariate normal distribution:

$$\bar{b} \mid B, \Delta_b \sim N\left(\frac{1}{N}(\iota_N \otimes I_K)^\intercal B, \frac{1}{N}\Delta_b\right), \tag{12}$$

where ι_N denotes an $N \times 1$ vector of ones and I_K is $K \times K$ dimensional identity matrix.

Markov chain Monte Carlo Scheme II

(3) Update Δ_b through an inverse-Wishart distribution:

$$\Delta_b \mid B, \bar{b} \sim IW\left(\nu_b + N, ((D - \bar{b} \otimes \mathbf{1}_N^\intercal)(D - \bar{b} \otimes \mathbf{1}_N^\intercal)^\intercal + V_b))^{-1}\right). \qquad (13)$$

(4) Update Σ through an inverse-Wishart distribution:

$$\Sigma \mid B, R, F \sim IW(\nu_\Sigma + T, V_\Sigma + \tilde{E}^\intercal \tilde{E}), \qquad (14)$$

where $\tilde{E} = [\hat{\epsilon}_1, \cdots, \hat{\epsilon}_N]$ is a $T \times N$ matrix of residuals, $\hat{\epsilon}_i = r_i - f_i^\intercal b_i$.

Note that $\Omega = \Sigma \otimes I_N$. Then we finish updating Ω.

Ensemble Forecast and Portfolio Optimization

The Bayesian forecast is the ensemble forecast:

$$\widehat{r}_{i,t+1}^{(j)} = f_{i,t} b_i^{(j)} \tag{15}$$

$$\frac{1}{J}\sum_{j=1}^{J} \widehat{r}_{i,t+1}^{(j)} = f_{i,t} \frac{1}{J}\sum_{j=1}^{J} b_i^{(j)}. \tag{16}$$

The portfolio is built to maximize the mean-variance utility function:

$$U(W) = \exp\left\{\mathbf{E}(R_{p,t+1}) - \frac{\gamma}{2}\mathrm{Var}(R_{p,t+1})\right\}, \tag{17}$$

where $R_{p,t+1} = W^\intercal R_{t+1}$ is the future portfolio return and γ is the coefficient for risk aversion. We simply restrict the short selling and require $\sum W_i = 1$ and $W_i \geq 0$.

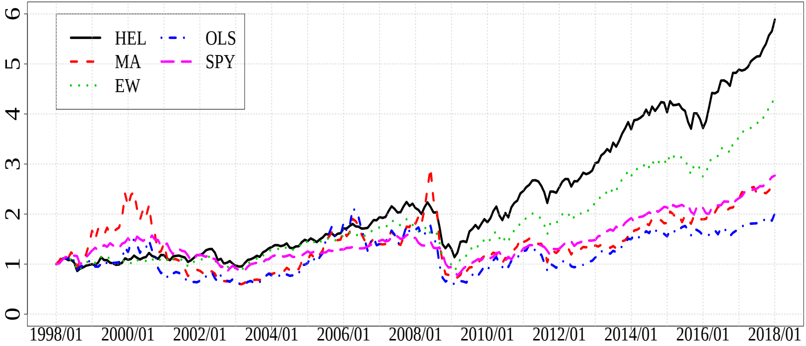
Empirical Findings

Data

We study the U.S. equity market from 1978 Jan to 2018 Jan.

- To compare portfolio optimization methods for different numbers of assets, three different versions of classifications are evaluated: 10, 30, and 49 industries.
- We use five macroeconomic predictors: treasury bill, long-term yield, inflation, stock market variance, and lag excess market return.
- The chosen firm characteristics include all main categories, such as the book-to-market ratio, earning-price ratio, investment growth, return on equity, inventory, accrual, dividend yield, gross profit, capitalization ratio, and asset turnover.

Cumulative Performance for 49 Industries

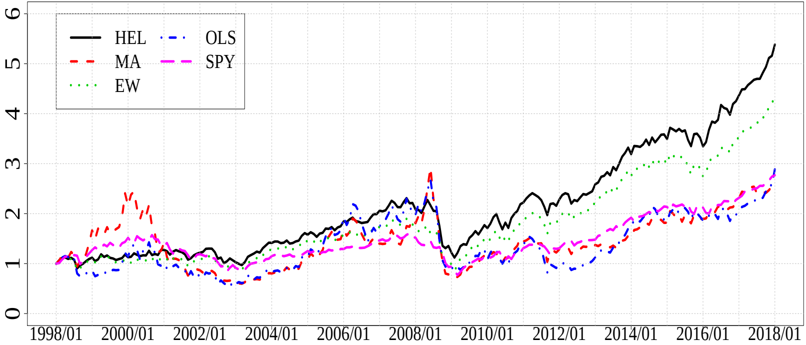


These figures provide the cumulative performance of the dynamic monthly updated portfolios by a 120-month rolling window. We plot the cumulative returns of multiple methods for 49 industry portfolios, including hierarchical ensemble learning (HEL), regression (OLS), moving average (MA), and the equally weighted portfolios (EW). We also add S&P 500 (SPY) as a benchmark for the passive investment.

Predictor Evaluation

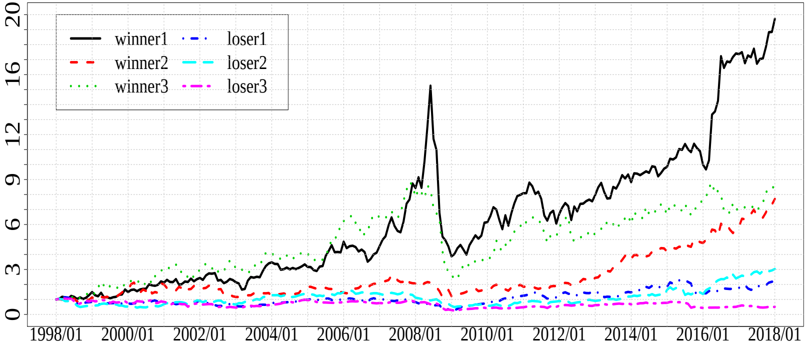
- To evaluate the usefulness of one fundamental characteristic, we need to examine if the corresponding θ_i^a and θ_i^b are jointly zero.
- In the unconditional formulation, we need to test whether six (five interactions plus one intercept) parameters are jointly zero for the characteristic and all its interaction with macro predictors.
- We find significant macro predictors (long-term yield, inflation, and stock market variance) as well as asset characteristics (dividend yield, accrual, and gross profit).

Cumulative Performance for Different Industries: Selected Model



These figures provide the cumulative performance of the dynamic monthly updated portfolios by a 120-month rolling window, for the selected model.

Cumulative Performance for Sorted Industries: Selected Model



We plot the cumulative returns of the first three winning industries and first three losing industries predicted by the model on a monthly rebalancing. The industry winners and losers are sorted on the predictive returns of our model.

Final Takeaway

- Our method is built for the stock-predictor-characteristics dynamic.
- We incorporate the hierarchical ensemble design for average signals.

- We show positive performance for the dynamic portfolio strategy.
- We find useful predictors (long-term yield, inflation, stock market variance) and portfolio fundamentals (dividend yield, accrual, and gross profit).

The R Package

BayesPortfolio

Factor Investing: Hierarchical Ensemble Learning.

The package will be posted on this website before end of May 19, 2019. Please check out the package later for update. Thanks!

jingyuhe.com/bayesportfolio