Kellogg School of Management
FINC 585-3: Asset Pricing
Spring 2024

Prof: Torben G. Andersen & Zhengyang Jiang
Jose Antunes Neto

Assignment 5

Problem 1.

For this question, we are going to use the dataset `industry_daily.csv` available on CANVAS. This dataset contains 38 Industry Porfolio daily returns avaiable on Kenneth French‘s website. Along with this, we will also use the Fama-French 5 + Momentum factors under the `ffdaily.csv` file.

A) For a range of $k = 1, \dots, 10$, calculate the k^{th} principal components of these industry excess returns (use the risk-free data available on `ffdaily.csv`). Remember to standardize the returns to have zero mean and unity variance before extracting the principal components. Plot the total explained variance ratio for each of these models. Using this measure, how many factors would you choose? Assuming normality of the log-returns, calculate both the AIC and BIC for each model as well. How many factors would you choose based on these criteria? Compare the results.

B) Pick up the model with 8 factors. Using rotation, how do the principal components correlate with the FF factors?

C) The Canonical Correlation Analysis is a methodology to capture the association between two sets of variables. Given data $X_{N_1 \times T}$ and $Y_{N_2 \times T}$ and a given number of components c, we look to find the $c \leq \min\{N_1, N_2\}$ linear combinations of X and Y with the greatest correlation, i.e.:

$$A, B = \underset{A,B}{\operatorname{argmax}} \quad \operatorname{corr}(A'X, B'Y) \tag{1}$$

and such that the linear combinations are uncorrelated with each other,

$$\mathbb{C}\big(A'X, A_j'X\big) = \mathbb{C}\big(B'Y, B_j'Y\big) = 0 \qquad \forall j = 1, \dots, c-1$$

The transformed variables $A'X$ and $B'Y$ are called the canonical variables, and their correlation is the canonical correlation (or score).

Use the Canonical Correlation Analysis to identify the relationship between the FF factors and the principal components. Using 1 to 5 CCA components, calculate the canonical correlation between these sets. How does it change with the number of components?

D) Plot the loadings of the PCA for each industry. Give a brief explanation of the results.

Solution.

A) After calculating the excess returns for the $N = 38$ industry portfolios, the remaining dataset contains $T = 15265$ daily returns. Using $k = 1, \dots, 10$ principal components, I calculate the total explained variance ratio for each of the models. This is shown in Figure 1. As we see, the variance ratio increases monotonically with the number of factors, something that is explained by the nature of the PCA.

Assuming that the idiosyncratic returns (a.k.a. residuals) are normally distributed, independent over time and with a fixed covariance matrix between industries, the log-likelihood function is given by

$$\log \mathcal{L}\left[(\varepsilon_{j,t})_{j=1,\dots,N;t=1,\dots,T}\right] = -\frac{NT}{2}\log 2\pi - \frac{T}{2}\log\det(\Sigma) - \frac{1}{2}\sum_{t=1}^{T}\varepsilon_t'\Sigma^{-1}\varepsilon_t,$$

Estimating Σ from the time series of residuals I can calculate both the Akaike and the Schwarz Information Criteria (AIC and BIC, respectively) as:

$$AIC(k) = 2k - 2\log\mathcal{L}$$
$$BIC(k) = k\log NT - 2\log\mathcal{L}$$

These are ploted along the explained variance in Figure 1. As it is usual with these measures, they tend to indicate a way more parsimonious model, with maximum criteria being achieved at only 1 factor for all measures.

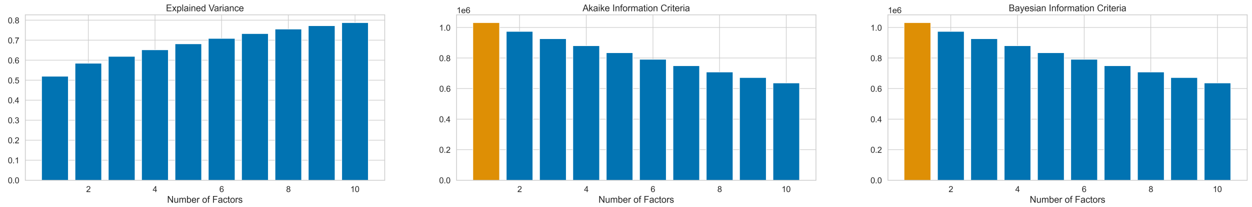
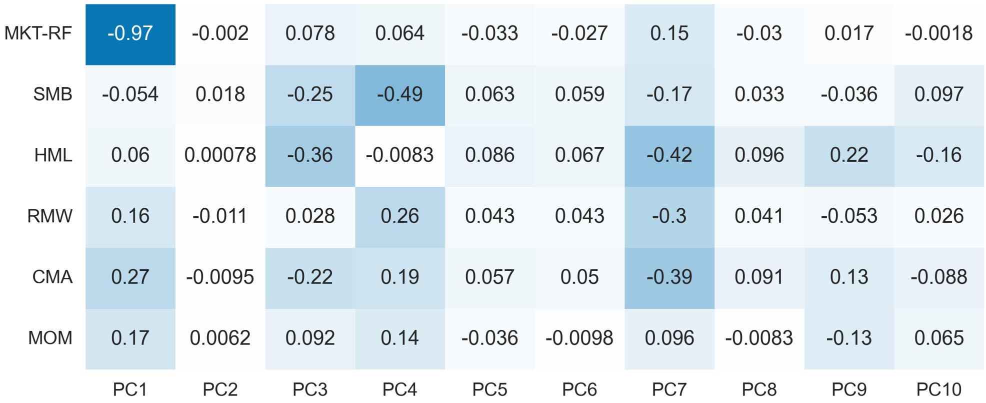


Figure 1: Explained Variance, AIC and BIC as a function of the number of factors

B) Using the model with 10 factors, I calculate the correlation matrix between the principal components and the Fama-French factors. This is shown in Figure 2. As the figure shows, the first component is highly correlated with the market factor, meaning that the market portfolio is responsible for most of the variation in the industry portfolios. This gets clearer when we look at Figure 3 which shows the time series of the Market and PC1. The two series very much like each other, indicating that indeed Market is the most important factor. On the other components, PC3 looks to correlate with HML, SMB and CMA. PC4 has high correlation with the Small minus Big (SMB) factor. Finally, the 7th component strongly correlates with the value factor (HML). The other components do not display significant correlation with any of the FF factors.

C) Using $c = 1, \dots 5$, I calculate the canonical correlation between the PC and the FF factors. The CCA score, or canonical correlation, is shown in Figure 4. As we see, the highest canonical correlation is achived with 3 components.



	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
MKT-RF	-0.97	-0.002	0.078	0.064	-0.033	-0.027	0.15	-0.03	0.017	-0.0018
SMB	-0.054	0.018	-0.25	-0.49	0.063	0.059	-0.17	0.033	-0.036	0.097
HML	0.06	0.00078	-0.36	-0.0083	0.086	0.067	-0.42	0.096	0.22	-0.16
RMW	0.16	-0.011	0.028	0.26	0.043	0.043	-0.3	0.041	-0.053	0.026
CMA	0.27	-0.0095	-0.22	0.19	0.057	0.05	-0.39	0.091	0.13	-0.088
MOM	0.17	0.0062	0.092	0.14	-0.036	-0.0098	0.096	-0.0083	-0.13	0.065

Figure 2: Correlation between Principal Components and FF

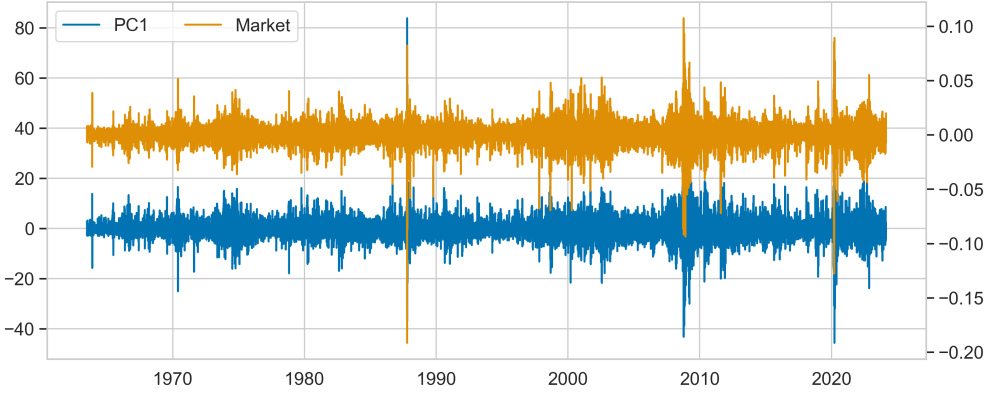


Figure 3: PC1 and Market Portfolio

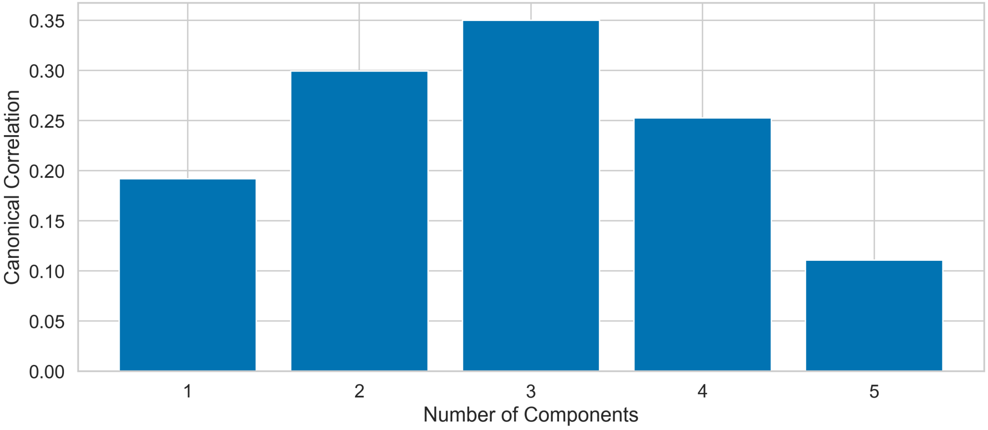


Figure 4: Canonical Correlation between PCA and FF

D) Finally, the PCA loadings are shown in Figure 5. Again, the market portfolio has decent loadings in all of the 38 industry portfolios. We can see that some principal components capture specific industries. For example, PC8 has higher loadings only on Steam and Water industries, PC2 on Agricultural, Garbage and Government. Factors 3, 4, 9 and 10 look to capture some of the remaining variance left by the market portfolio.

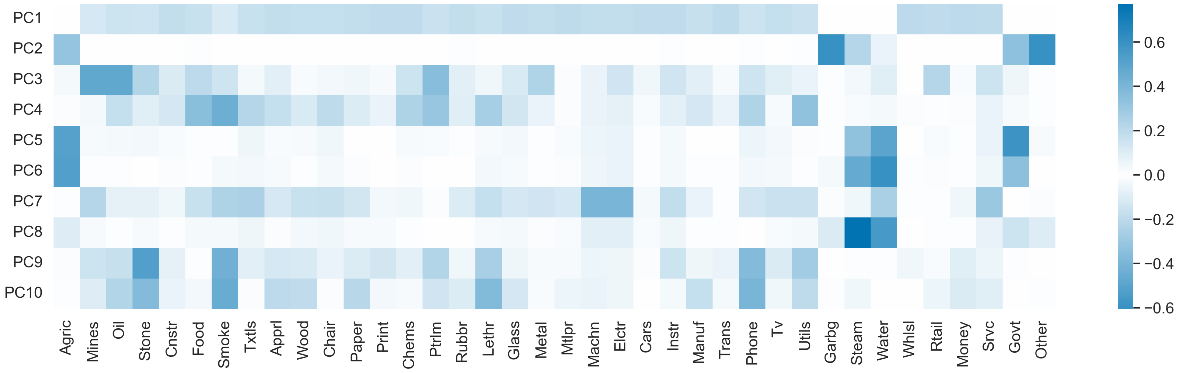


Figure 5: Factor Loadings on the PCA

□

Problem 2.

Connor and Korajczyk (1986) develop a method to measure the performance of a mutual fund portfolio. Using a large number of funds, they show that the Treynor and Black (1973) Appraisal Ratio completely captures investors ranking over funds. A simplified version of the model is described as follows:

Let fund's j excess return at time t be denoted as $r_{j,t}$ for $j = 1, \dots, N$ and the riskless asset return be $r_{0,t}$. These returns are generated by a finite set of latent factors $\mathbf{f_t} = \left(f_t^{(1)}, \dots, f_t^{(K)}\right)$ where K is known. The return of fund j is then given by:

$$r_{j,t} = \alpha_j + \sum_{k=1}^{K} \beta_j^{(k)} \left(\gamma^{(k)} + f_t^{(k)}\right) + \varepsilon_{j,t} \tag{2}$$

with $\mathbb{E}(\mathbf{f_t}) = 0$, $\mathbb{E}(\mathbf{f_t}\mathbf{f_t}') = I_K$ and $\mathbb{E}(\varepsilon_{j,t}) = 0$, $\mathbb{V}(\varepsilon_{j,t}) = \sigma_j^2$. We also assume that $\varepsilon_{\cdot,t}$ is uncorrelated between funds and mean-independent on the set of factors. The Appraisal Ratio is defined as $t_j = \alpha_j/\sigma_j$, and according to Theorem 4, this ratio is sufficient to rank funds[1].

Let $\mathbf{R}$ be the $N \times T$ matrix of fund returns. The authors show that equation (2) can be estimated by the following algorithm.

1) Compute a $k \times T$ principal component matrix of $\frac{1}{n}\mathbf{R}'\mathbf{R}$, denote it as $\widehat{\mathbf{f}}$;

2) Run a time series regression

$$r_j = \widehat{\alpha}_j + \sum_{k=1}^{K} \widehat{\beta}_j^{(k)} \left(\widehat{\gamma}^{(k)} + \widehat{f}^{(k)}\right) + \widehat{\varepsilon}_j \tag{3}$$

3) Calculate $\widehat{t_j} = \widehat{\alpha}_j/\widehat{\sigma}_j$.

4) For a large N and T, this estimator is consistent and asymptotically normal:

$$\operatorname*{plim}_{N\to\infty} T^{1/2}\left(\widehat{t_j} - t_j\right) \xrightarrow{d} N\left(0, 1 + \sum_{k=1}^{K} \gamma_k^2\right) \qquad \text{as } T \to \infty \tag{4}$$

In this exercise, we will apply this model to a dataset of mutual funds. The dataset `fund_returns.csv` available on CANVAS contains a time series of monthly log-returns for a large set of mutual funds from 2000 to 2023 obtained from CRSP.[2]. We will also use the data on the Fama-French factors and risk-free rate available on the `ffdaily.csv` file.

A) For k from 1 to 6, calculate the principal components of the funds' excess returns. Plot the explained variance ratio for each of these components. How do these PC compare with the Fama-French factors?

[1] i.e, an investor would strictly prefer to switch from investing in fund j to fund l if, and only if, $t_j > t_l$

[2] This data is available on WRDS under the `MONTHLY_RETURNS` dataset on the `CRSP` library.

B) Connor and Korajczyk (1993) provide a methodology to choose the optimal number of factors K in an approximate factor model. Their test is based on the idea that additional (non-informative) factors will not increase the explained variance of the model. Their algorithm is described as the following iteration:

1) For a given number of factors k, estimate the model with k and $k+1$ factors. Let the residuals be $\widehat{\varepsilon}_{j,t}$ and $\widehat{\varepsilon}^*_{j,t}$, respectively;

2) Calculate adjusted squared residuals

$$\begin{aligned} \widehat{\sigma}_{j,t} &= \widehat{\varepsilon}_{j,t} / \left[1 - (k+1)/T - k/N\right] \\ \widehat{\sigma}^*_{j,t} &= \widehat{\varepsilon}^*_{j,t} / \left[1 - (k+2)/T - k/N\right] \end{aligned} \tag{5}$$

3) Calculate $\widehat{\mathbf{\Delta}}$ by subtracting the cross-sectional means of $\widehat{\sigma}_{j,t}$ in odd periods from the cross-sectional means of $\widehat{\sigma}^*_{j,t+1}$ in even periods:

$$\widehat{\Delta}_s = \mu_{2s-1} - \mu^*_{2s} \qquad s = 1, \ldots, \lfloor T/2 \rfloor \tag{6}$$

where $\mu_t = N^{-1} \sum_{j=1}^{N} \widehat{\sigma}^2_{j,t}$ and $\mu^*_t = N^{-1} \sum_{j=1}^{N} \left(\widehat{\sigma}^*_{j,t}\right)^2$. Under the null, as $N \to \infty$

$$\sqrt{N} \widehat{\Delta} \xrightarrow{d} N\left(0, \Gamma\right) \tag{7}$$

4) Using the time series of $\widehat{\Delta}$, calculate the mean and covariance matrix $\widehat{\Gamma}$ and perform a one-sided zero mean test:

$$\mathcal{H}_0 : \mathbb{E}\left[\widehat{\Delta}\right] \leq 0 \qquad \mathcal{H}_1 : \mathbb{E}\left[\widehat{\Delta}\right] > 0$$

5) If the null is rejected, start over with $k+1$ factors. Otherwise, set $K = k$ as the optimal number of factors.

Using their test, find the optimal K. Use this number for the rest of the exercise.

C) Estimate (3) for each fund and calculate their Appraisal Ratio. Plot the distribution of this measure. According to this, how many funds do really outperform the market? Use the distribution on (4) to test the significance of their performance.

D) Equation (2) assumes that the factor loadings $\beta_j^{(k)}$ are constant over time. Split the sample in half. Redo C) for each subsample and compare the results. How many funds do outperform in both periods? Interpret.

Solution.

A) After combining the industry portfolios dataset with the FF factors, the remaining dataset contains 288 monthly excess returns for 4674 mutual funds. Calculating the principal components for $k = 1, \ldots, 6$, Figure 6 plots the explained variance ratio

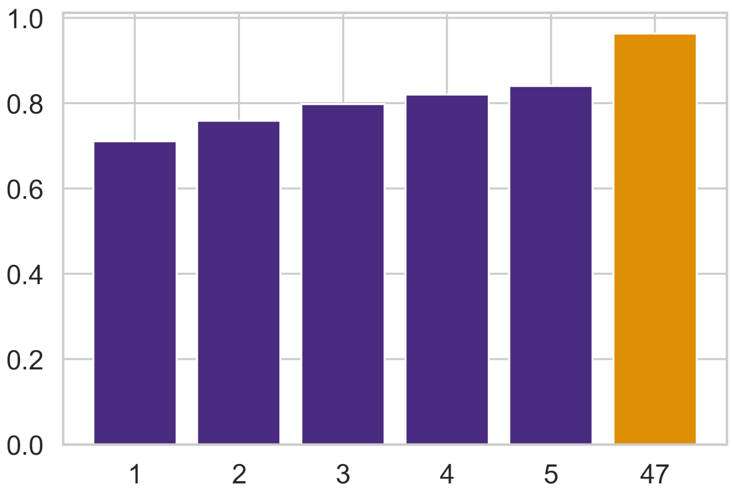


Figure 6: Explained Variance Ratio for the set of k components

for each of these portfolios. From this figure, we see that the amount of variance explained by the first factors is quite low compared to the previous exercise. The correlation between the 6 principal components and the FF factors is depicted in Figure 7. Comparing it with the previous exercise, the first principal component still mostly captures the market portfolio. However, we see a greater influence of the momentum factors in each of the components. Overall the correlation is lower than before, with maximum values of arounf 0.20.

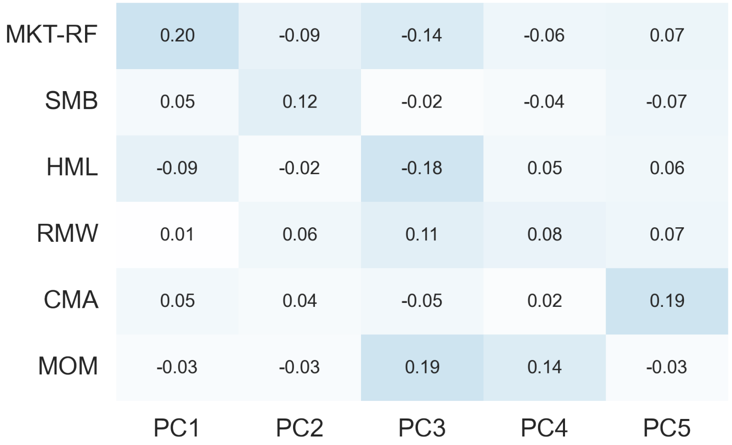


Figure 7: Correlation between Principal Components and FF Factors

B) I use the Connor and Korajczyk (1993) to find the optimal number of factors. Using the entire sample, a total of 47 factors is required to explain a large portion of the variance of the 4674 mutual funds. This is higher than the number of factors found in the industry portfolios, but it is expected given the larger amount of mutual funds and the plethora of different strategies they might follow. Figure 6 shows the explained variance ratio for the first 6 components and the optimal number. We can see that there is a significant increase in the explained variance ratio compared to the model with 6 components.

C) Using the optimal number of factors, I estimate the Appraisal Ratio for each fund in the sample and calculate the critical values of (4). The distribution is shown in Figure 8. We see that there is a significant number of funds outperforming the market in the overall sample. A total of 1158 funds have a positive and significant appraisal ratio when using the asymptotic distribution.

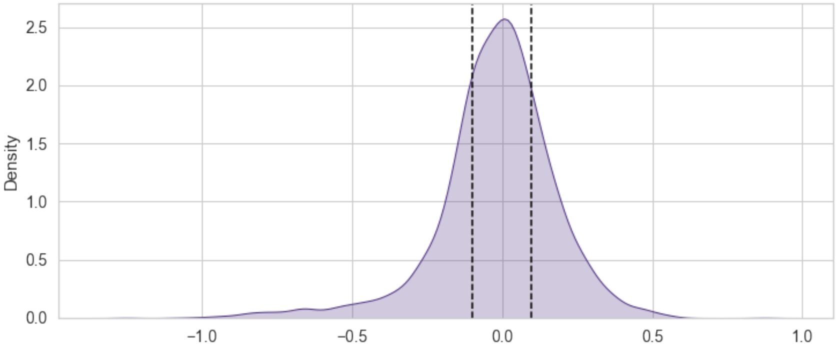


Figure 8: Distribution of Appraisal Ratio

D) Using data for over 20 years of returns, it is hard to expect that the sensitivity of funds to each factor would have stayed the same. Different macroeconomic conditions, for example, can affect the strategies and performances of each fund. To account for this, I estimate the same model using 2 halves of the sample. The first subsample contains data from 2000 to 2011, while the second spans from 2012 to 2023. I run the model for each subsample and calculate the optimal value of factors. Unlike in the overall sample, the subsamples indicated an optimal number of 1 and 12 factors respectively, smaller than the full sample. Figure 9 shows the distribution of the Appraisal Ratio for both subsamples. As we see, these ratios tend to be higher in the first subsample, compared to the second. A total of 1638 ($\approx$ 35%s) of funds outperformed in the first half and 729 ($\approx$ 15%) in the second half and only a 333 ($\approx$ 7%) of funds did well in both samples. This exhibits the difficulty in maintaining a good strategy throughout different market conditions.

□

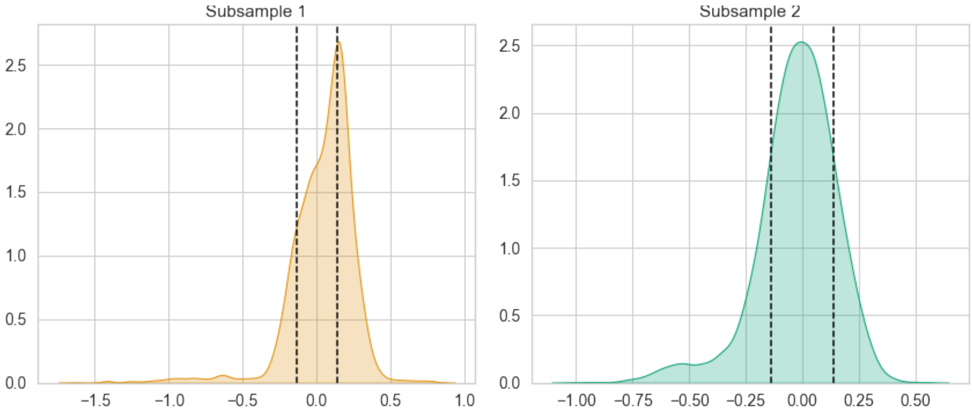


Figure 9: Distribution of Appraisal Ratio for Subsamples

References

G. Connor and R. A. Korajczyk. Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of financial economics*, 15(3):373–394, 1986.

G. Connor and R. A. Korajczyk. A test for the number of factors in an approximate factor model. *the Journal of Finance*, 48(4):1263–1291, 1993.

J. L. Treynor and F. Black. How to use security analysis to improve portfolio selection. *The journal of business*, 46(1):66–86, 1973.