PUC-Rio - Certificação Digital Nº 0912808/CA

3 Forecasting the cross-sectional variation in stock returns

3.1 Introduction

Several studies present models which explain the variability in stock returns. By grouping related stocks in portfolios and considering long-term statistics, some models can explain up to 90% of returns variability. If such variability exists and is also related to expected returns, it should be possible to check its existence assuming a forecast environment (without look-ahead bias) and evaluating statistics over shorter horizons. Under this environment, a model which can forecast stock return variability could be plugged red into a portfolio optimization framework in order to generate economic value.

Several studies present models which explain the variability in stock returns. By grouping related stocks in portfolios and considering long-term statistics, some models can explain up to 90% of returns variability. If such variability exists and is also related to expected returns, it should be possible to check its existence assuming a forecast environment (without look-ahead bias) and evaluating statistics over shorter horizons. Under this environment, a model which can forecast stock return variability could be plugged into a portfolio optimization framework in order to generate economic value.

Our work can be divided in three parts, all related to the forecast of the cross-sectional expected return variation in stocks. The first part evaluates the forecasting performance of standard, skeptical and naive models using alternative statistics. The second part is an attempt to find better models according to the alternative statistics proposed in the first part. The third part evaluates all expected return models in the previous parts from the portfolio manager point of view by plugging the model into a Markowitz optimization framework. Better models should generate better return/risk ratios if the covariance model is the same. In the second part, we search for better models using two different approaches. The first approach defines a model set containing several models estimated by the Fama-Macbeth and searches for better model combinations according to each of the short-run statistics. The second approach defines a single model containing several factors and estimates using the LASSO technique (Tibshirani (1996)). Besides, we use a set of portfolios beyond the traditional portfolios sorted by size and book-to-market. The set of portfolios include: 100 portfolios sorted on size and book-to-market, 48 portfolios sorted by industry, 25 portfolios sorted by size and 1-year momentum and 10 portfolios sorted on earning/price.

The benchmark models are the CAPM (Sharpe (1964)) and the Fama-French 3-factor model (Fama & French (1993a)). Others works on the cross-sectional variation of stock expected returns include: Jagganathan & Wang (1996), Lettau & Ludvigson (2001) and Petkova (2006). All these articles have in common the fact they consider fitted and realized returns averaged across the entire sample when checking the model capacity to explain cross-sectional variation. Instead, we use monthly and average monthly return over one year to check forecasting power (not explanatory). We are not aware of other works using short-horizon statistics or using LASSO in the context of asset returns.

Our results suggest the CAPM and Fama-French do not perform better than naive models (for example, the historical mean) if we consider return statistics over short horizons or return/risk ratios of optimized portfolio assuming each expected return model. This part can be seen as an expansion of the work Welch & Goyal (2008) to evaluate models to explain the cross-sectional variation of stock expected returns. Finally, combination of models and LASSO estimation may provide better results, as shown by a relevant positive economic value from both approaches. A Sharpe ratio maximizer investor would be willing to spend more than 200 basis points to exchange the S&P500 by the portfolios generated by a modified Markowitz optimization using these models.

The paper is organized as follows. Section 3.2 formulates the problem, that is, the variability of expected returns. Section 3.3 describes the statistics we use and evaluate benchmark, naive and skeptical models according to them. Section 3.4 attempts to search for better models on a linear multi-factor environment. In 3.4.1 we define a set containing several linear models and attempt to find good combinations according to short-run statistics. In 3.4.2 we use LASSO estimation on a linear model with 24 factors. Section 3.5 presents the experiment to evaluate models according to the economic value of the Markowitz optimized portfolio obtained by plugging the model. Section 3.6 concludes.

3.2 Problem Formulation

Let $r_{i,t+1}$ be the holding period return of an asset (or portfolio) i from period t to $t+1$. We define the return in excess of the market $r^*_{i,t+1}$ as:

PUC-Rio - Certificação Digital Nº 0912808/CA

$$r^*_{i,t+1} = r_{i,t+1} - r_{b,t+1} \tag{3-1}$$

where $r_{b,t+1}$ is the benchmark return. In this paper, we use the S&P500 index as benchmark.

The variability we want to forecast is the expected return in excess of the market from period t to $t+1$, conditional on information known at t, for a set of portfolios:

$$\mathbb{E}_t[r^*_{i,t+1}] = \mathbb{E}_t[r_{i,t+1} - r_{b,t+1}] \tag{3-2}$$

Therefore, we evaluate models which attempt to explain $\mathbb{E}_t[r^*_{i,t+1}]$ by forecasting return for each asset i, without attempting to forecast the benchmark expected return.

3.3 Short-horizon statistics

In order to evaluate the forecasting power of several stock return forecast models, we use 183 portfolios from Kenneth French's website:

1. Group 1: 100 portfolios sorted on size (market equity) and book-to-market (ratio of book equity to market equity).

2. Group 2: 25 portfolios sorted on size (market equity) and momentum (prior 2-12 returns). The monthly size breakpoints are the NYSE market equity quintiles. The monthly prior (2-12) return breakpoints are NYSE quintiles. The portfolios constructed each month include NYSE, AMEX, and NASDAQ stocks with prior return data.

3. Group 3: 48 portfolios sorted on industry.

4. Group 4: 10 portfolios sorted on earnings/price ratio. Portfolios are formed on E/P at the end of each June using NYSE breakpoints.

By using multiple groups of portfolios, we attempt to expand our analysis beyond traditional sorted by only one or two characteristics (mainly to explain size and value effects). Lewellen et al. (2010) describe various concerns which are most severe when a couple of factors explain nearly all of the variation in expected returns. A desirable model should be able to successfully forecast variations in expected returns among different portfolios when there is some variation, and, do not forecast spurious variations when there are none. So we include portfolios sorted by industry and earning/price ratio, which are groups with lesser spread in expected returns.

The returns are from January 1961 to December 2008. Table 3.1 presents the cross-sectional standard deviation of average monthly returns σ_{avgret}, average cross-sectional standard deviation $\bar{\sigma}$ and the total variability explained

PUC-Rio - Certificação Digital Nº 0912808/CA

Table 3.1: Cross-sectional statistics

This table presents cross-sectional statistics for the 4 portfolios groups and the entire set of portfolios. The statistics are the cross-sectional standard deviation of average monthly returns $\sigma_a vgret$, average cross-sectional standard deviation $\bar{\sigma}$ and the total variability explained by the first 5 principal components (PC1-PC5).

Group	σ_{avgret}	$\bar{\sigma}$	PC1	PC2	PC3	PC4	PC5
Group 1	0.003	0.030	77.0	3.6	2.9	1.0	0.7
Group 2	0.004	0.024	81.2	7.1	4.6	1.7	1.0
Group 3	0.002	0.042	56.4	6.5	4.2	3.7	2.6
Group 4	0.002	0.017	85.6	6.0	2.7	1.3	1.0
All groups	0.003	0.033	72.2	3.1	2.4	2.1	1.3

by the first 5 principal components (PC1-PC5) of the portfolio returns. The values of σ_{avgret} and $\bar{\sigma}$ show the variability of monthly returns is far greater than the variability of average monthly returns over the period. As expected, the returns of Group 3 are less explained by the first 5 principal components.

We propose three statistics to evaluate the forecasting power of the cross-sectional variation in stock returns. They ressemble the cross-section R^2 used since Fama & French (1993b), with two differences. First, the returns are from the average of distinct horizons (monthly, annual, entire sample). Second, the weight of the each portfolio's forecasting error is related to the market value of the portfolio.

Let S_m be the first statistic as follows:

$$S_{m,t} \equiv 1 - \frac{\sum_{i=1}^{N} \omega_{i,t} e_{i,t}^2}{\sum_{i=1}^{N} \omega_{i,t}(r_{i,t+1}^* - \bar{r}_{i,t+1}^*)^2} \tag{3-3}$$

where i represents a index of portfolios, $\omega_{i,t}$ is the portfolio's market value, $e_{i,t}$ is the difference between forecast and realized returns, $r_{i,t+1}^*$ is the realized return in excess of the market and $\bar{r}_{i,t+1}^*$ is the cross-sectional average return in excess of the market.

The next statistic, S_y, considers average monthly returns over each year:

$$S_{y,t} \equiv 1 - \frac{\sum_{i=1}^{N} \omega_{i,t}(e_{i,t+1}^y)^2}{\sum_{i=1}^{N} \omega_{i,t}(r_{i,t+1}^y - \bar{r}_{i,t+1}^y)^2} \tag{3-4}$$

where i represents a index of portfolios,$\omega_{i,t}$ is the portfolio's market value, $e_{i,t}^y$ is the difference between average forecast and average realized returns over a year, $r_{i,t+1}^y$ is the monthly excess return averaged over one year and $\bar{r}_{i,t+1}^y$ is the cross-sectional average of $r_{i,t+1}^y$.

Finally, the statistic S_s consider average monthly returns over the entire sample:

$$S_s \equiv 1 - \frac{\sum_{i=1}^{N} \omega_{i,t}(e_{i,t+1}^s)^2}{\sum_{i=1}^{N} \omega_{i,t}(r_{i,t+1}^s - \bar{r}_{i,t+1}^s)^2} \tag{3-5}$$

where i represents a index of portfolios, t and $t+1$ represent periods, $\omega_{i,t}$ is a measure of the portfolio's market value, $e_{i,t}^s$ is the difference between average

PUC-Rio - Certificação Digital Nº 0912808/CA

forecast and average realized returns over the sample, $r^s_{i,t+1}$ is the monthly excess return averaged over the sample and $\bar{r}^s_{i,t+1}$ is the cross-sectional average of $r^s_{i,t+1}$. S_s is more related to the cross-sectional R^2 used to measure the explanatory power of models which attempt to explain cross-section variation in returns.

At first, we use these statistics to compare the performance of benchmark, naive and skeptical return models. These models attempt to model the expected return in excess of the market $\mathbb{E}_t[r^*_{i,t+1}]$ and they are defined as follows:

1. Model 1: $\mathbb{E}_t[r^*_{i,t+1}] = 0$ for every portfolio i, "no variation"
2. Model 2: $\mathbb{E}_t[r^*_{i,t+1}] = c_i\mathbb{E}_t[r_{m,t+1} - r_{f,t}]$, "CAPM"
3. Model 3: $\mathbb{E}_t[r^*_{i,t+1}] = c_{1,i}\mathbb{E}_t[r_{m,t+1} - r_{f,t}] + c_{2,i}\mathbb{E}_t[SMB_{t+1}] + c_{3,i}\mathbb{E}_t[HML_{t+1}]$, "FF 3-factor"
4. Model 4: $\mathbb{E}_t[r^*_{i,t+1}] = a_i$, "intercept"
5. Model 5: $\mathbb{E}_t[r^*_{i,t+1}] = a_i + b_i r^*_{i,t}$, "intercept + lag"

The "no variation" model simply states there is not expected variation in portfolio returns from t to $t+1$ conditional on all information available at t. The two following models use factors and are based on Sharpe (1964) CAPM and the 3-factor model proposed in Fama & French (1996). The last two models go in the opposite direction. Instead of risk premia for factors times factor loadings, they allow intercepts ("intercept" model) and intercepts plus lag times a coefficient.

The "CAPM" and "FF 3-factor" models are estimated by the Fama-Macbeth procedure (Fama & Macbeth (1973)). This procedure involves two steps and the objective is to estimate both factor loadings and factor risk premia. Let the expected return in excess of the market be defined as follows:

$$\mathbb{E}_t[r^*_{i,t+1}] = \alpha_i + \beta_i'\bar{\delta}_t \tag{3-6}$$

where β_i is a vector containing factor loadings and $\bar{\delta}_t$ is the risk premia for each factor.

In the first step, $\bar{\delta}_t$ is replaced by the realized return of the factor-mimicking portfolio, so α_i and β_i are estimated for each portfolio i by OLS from:

$$r^*_{i,t+1} = \alpha_i + \beta_i' R_{X,t+1} + e_{i,t} \tag{3-7}$$

where $R_{X,t}$ is the return of the factor-mimicking portfolios.

The second step involves the estimation of the risk premia for each factor and uses $\hat{\alpha}_i$ and $\hat{\beta}_i$ estimated during the first step. Given $\hat{\alpha}_i$ and $\hat{\beta}_i$, we estimate

PUC-Rio - Certificação Digital Nº 0912808/CA

Table 3.2: Results - growing window of estimation

This table presents average values of $\bar{S}_m$, $\bar{S}_y$ and the S_s statistic for the "no variation", "CAPM", "FF 3-factor", "intercept" and "intercept + lag" models. The value in parenthesis is the standard-deviation of the statistic over the sample. The estimation procedure uses a growing window of estimation.

Model	$\bar{S}_m \times 100$	$\bar{S}_y \times 100$	$S_s \times 100$
"no variation"	13.13 (15.37)	20.07 (17.63)	14.05
"CAPM"	12.80 (15.54)	18.61 (18.85)	13.78
"FF 3-factor"	12.64 (16.37)	17.80 (25.11)	30.70
"intercept"	13.05 (16.34)	19.38 (24.85)	77.34
"intercept + lag"	12.65 (18.42)	30.73 (24.85)	80.51

the risk premia for each factor at period t by OLS from the cross-section of portfolios:

$$r^*_{i,t+1} - \hat{\alpha}_i = \delta'_t \hat{\beta}_i + u_{i,t} \tag{3-8}$$

Finally, the risk premia $\bar{\delta}_{t+1}$ is estimated as the average of the estimated risk premia from the last, say, w periods.

In the "intercept" and "intercept + lag" models, a_i and b_i are estimated by OLS. Hence, $\hat{a}_i$ is the sample average return in excess of the market in the "intercept" model.

Table 3.2 presents the statistics for the models defined before, using a growing window to estimate coefficients and risk premia associated to each factor. The difference in S_m is small, although the 'no cross-sectional variation model' performs better. The statistic S_y presents a greater variation, favoring the no-cross-section variance model and the model with intercept plus one lag. The models which allow intercepts provide a greater explanation of the average variation in returns, as shown in the last column. The result for the "FF 3-Factor" model for S_s is lesser than displayed in previous papers. This occurs because the set of portfolios does not include only portfolios sorted by size and book-to-market and the estimation of coefficients and risk premia uses only information available until each period.

Given evidence of time-varying risk premia and factor exposures, we re-estimate coefficients and risk premia for each factor by using a rolling window of 10 years in both parts of the estimation procedure. Table 3.3 shows the results. This reduce the S_m for the models with intercept and has little effect in the other two models. There is small average decrease in S_y and an overall increase in S_s.

These results suggest that if we remove look-ahead bias to estimate coefficients and risk premia to each factor and include portfolios sorted by others characteristics, the "FF 3-factor" performs much worse than the naive models which replace factors by intercepts and lags. Besides, the statistics over shorter horizons show a negligible difference in performance between models, with some advantage to "no variation" and "intercept + lag" models.

PUC-Rio - Certificação Digital Nº 0912808/CA

Table 3.3: Results - rolling window of estimation

This table presents average values of $\bar{S}_m$, $\bar{S}_y$ and the S_s statistic for the "no variation", "CAPM", "FF 3-factor", "intercept" and "intercept + lag" models. The value in parenthesis is the standard-deviation of the statistic over the sample. The estimation procedure uses a rolling window of estimation (10 years).

Model	$\bar{S}_m \times 100$	$\bar{S}_y \times 100$	$S_s \times 100$
"no variation"	13.13 (15.37)	20.07 (17.63)	14.05
"CAPM"	12.96 (15.39)	19.76 (17.55)	20.08
"FF 3-factor"	12.51 (16.40)	16.62 (25.47)	40.54
"intercept"	12.08 (17.01)	16.01 (26.56)	93.60
"intercept + lag"	11.24 (19.50)	25.23 (25.24)	93.40

3.4 Search for a better model

Results from Section 3.3 suggest the benchmark models are not better than the skeptical model to explain cross-sectional variation in stock returns. This raises the following question: can a single linear or combination of linear models achieve better results?

In order to answer to this question, we attempt two model combination approaches. The first approach (Section 3.4.1) selects models from a large set according to the performance of the combination. The set contains only models estimated by the Fama-Macbeth procedure. The second approach (Section 3.4.2) estimates a single model containing several factors by using the LASSO Method (e.g. Hastie et al. (2008)).

3.4.1 Combination of models estimated by Fama-Macbeth procedure

The following equation describe the general functional form of the returns in excess of the market for portfolio i:

$$\mathbb{E}_t[r^*_{i,t+1}] = \alpha_i + \beta_i \delta_{M,t+1} + c_i \delta_{X,t+1} + d_i r^*_{i,t} \tag{3-9}$$

where $\delta_{M,t+1}$ is the market risk premia and $\delta_{X,t+1}$ is the risk premia associated to another factor, from the list:

1. Small-minus-big and high-minus-low portfolios;
2. Term spread;
3. Variation in term spread;
4. Default spread;
5. Variation in default spread;
6. Aggregated earnings/price ratio;

PUC-Rio - Certificação Digital Nº 0912808/CA

7. Variation in aggregated earnings/price ratio;
8. Variation in the risk-free rate;
9. Variation in unemployment;
10. Inflation (accumulated, one year).

We obtain term spread, default spread, inflation, aggregated earnings/price ratio and risk-free rate from Amit Goyal's website. Small-minus-big and high-minus-low portfolios returns are also from Kenneth French's website[1]

In addition, the set of models is defined by combining the following assumptions:

1. $\alpha_i = 0$ vs. α_i free;
2. $\beta_i = 0$ vs. β_i free;
3. $c_i = 0$ vs. c_i free;
4. $d_i = 0$ vs. d_i free;
5. Risk premia for each factor estimated from last 1,3,6,12 or 120 months.

It is noteworthy that this functional form and assumptions include the "no variation", "CAPM", "intercept" and "intercept + lag" models evaluated in Section 3.3.

All coefficients are estimated by the Fama-Macbeth procedure. The first stage uses a rolling windows of 120 months, while the second stage uses average risk premia from last 1,3,6,12 or 120 months.

We impose an additional constraint: for each group of portfolios, the weighted sum of expected returns in excess of the market equal zero, that is: $\sum \omega_{i,t}\mathbb{E}_t[r^*_{i,t+1}] = 0$ for each group of portfolios. In order to impose this constraint, we shift all the forecast excess returns from a given group of portfolios for a same value.

Given the set of 473 models defined before, we attempt to find the best equal-weighted combination of n models according to $\bar{S}_m$ or $\bar{S}_y$. As n increases, the number of possible combinations increases exponentially. Hence, we restrict the combinations we explore by the following procedure:

1. Find the best model (according to the statistic) - this is the best model combination ($M1$) for $n = 1$;

[1] Available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

PUC-Rio - Certificação Digital Nº 0912808/CA

Table 3.4: Results - best equal-weighted combination of 7 models

This table presents average values of S_m and S_y and the S_s statistic for the "no variation" and "FF 3-factor" models, as well as the best equal-weighted combinations of 7 models considering S_m and S_y . The value in parenthesis is the standard-deviation of the statistic over the sample. The estimation procedure uses a rolling window of estimation (10 years).

Model	$\bar{S}_m \times 100$	$\bar{S}_y \times 100$	$S_s \times 100$
Model "No variation"	13.13 (15.37)	20.07 (17.63)	14.05
Model "FF 3-factor"	12.51 (16.40)	16.62 (25.47)	40.54
Best combination for $\bar{S}_m$	13.71 (16.71)	30.33 (18.09)	52.40
Best combination for $\bar{S}_y$	2.31 (39.86)	44.71 (19.17)	60.99

Table 3.5: Results - best equal-weighted combination of 7 models, out-of-sample procedure

This table presents average values of S_m and S_y and the S_s statistic for the "no variation" and "FF 3-factor" models, as well as the best equal-weighted combinations of 7 models considering S_m and S_y.The best model combination from the first half of the sample is used in the second half, and vice-versa. The value in parenthesis is the standard-deviation of the statistic over the sample. The estimation procedure uses a rolling window of estimation (10 years).

Model	$\bar{S}_m \times 100$	$\bar{S}_y \times 100$	$S_s \times 100$
Model "No variation"	13.13 (15.37)	20.07 (17.63)	14.05
Model "FF 3-factor"	12.51 (16.40)	16.62 (25.47)	40.54
Best combination for $\bar{S}_m$	13.20 (16.59)	28.76 (17.43)	42.71
Best combination for $\bar{S}_y$	2.48 (41.40)	43.98 (19.33)	62.81

2. Find the best equal-weight combination of two models containing $M1$. The two portfolios are the best combination ($M2$) for $n = 2$;

3. Repeat the procedure using $M2$ to select the best combination ($M3$) for $n = 3$, and so on.

Table 3.4 presents the results for the best 7-model combinations according to S_m and S_y. The best combination according to S_m has better performance than both benchmark models. The best model combination according to S_y performs much better according to S_y and S_s statistics, but poor performance when considering S_m.

Although the estimated coefficients and risk premia use only information available in the beginning of each period, the approach to select and combine models uses the performance over the entire sample. Therefore, even if the estimation procedure does not suffer from look-ahead bias, the model selection approach does. Besides, given the large set of available models, we cannot rule out data-mining bias. We consider this approach to be robust to both issues. As an argument, we evaluate a modified version of this approach. The first step is to split the sample in two, from 1971 to 1989 and from 1990 to 2008. Then, the selected models from the first part are used in the second part, and vice-versa. Table 3.5 presents the results for the best 7-model combinations using this adjustment.

The model combination according to S_m still performs little better than the "no variation". However, this variation is negligible if the volatility of this

PUC-Rio - Certificação Digital Nº 0912808/CA

PUC-Rio - Certificação Digital Nº 0912808/CA

measure is taken in account. Finally, the model combination according to S_y performs ahead of others (except S_m) and almost as well as the selected model using the entire sample.

3.4.2 Combination of models estimated by LASSO

The LASSO is a shrinkage method which restricts the estimated coefficients by $\sum_{i=1}^{N}|\beta_i| \leq t$, for some t.

Equivalently, the LASSO can be seen modeled as a Bayesian estimation. Let $\{\beta_j\}_{j=0}^{p}$ be a set of coefficients to be estimated from the equation $y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + \epsilon_i$ where i identifies each sample element. The sample contains N elements. The parameters are estimated from the criterion:

$$\tilde{\beta} = \arg\min_{\beta}\{\sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\} \tag{3-10}$$

The sum $\sum_{j=1}^{p}|\beta_j|$ can be seen as the log-prior density for $\{\beta_j\}_{j=1}^{p}$. Hence, λ controls the relative weight on the prior. Instead of defining a specific value for λ, we estimate the coefficients by using 6 different values of λ and combine the forecasts.

We consider a multi-factor model with a set of 24 lagged variables (in the sense that x_t is known at period t), given by:

$$\mathbb{E}_t[r^*_{i,t+1}] = \alpha_i + \beta_i x_t \tag{3-11}$$

where β_i is a vector of 24 coefficients and x_t is a vector of 24 lagged factors. This model is different from the models used in Section 3.4.1. Those models use concurrent factors and estimate the risk premia associated to each factor, in a 2-step procedure. Therefore, the expected return is the combination of estimated factor coefficients and risk premia. Instead, this model uses lagged variables as factors and the expected return is the combination of estimated factors coefficients and current factor value.

The factors we use are: earnings price ratio , variation in earnings price ratio (difference between consecutive months), variation in earnings price ratio (difference between 3 months), variation in earnings price ratio (difference between 12 months), variation in earnings (difference between consecutive months). variation in earnings (difference between 3 months), variation in earnings (difference between 12 months), book-to-market ratio, market excess (1 month interval), market excess return (3 months), market excess return (12 months), risk-free rate, variation in risk-free rate (1 month interval), inflation, term spread, term spread variation (consecutive months), default spread, default spread variation (consecutive months), small-minus-

Table 3.6: Results - Performance of models estimated by Lasso
This table presents average values of S_m and S_y and the S_s statistic for the "no variation" and "FF 3-factor" models, as well as for models estimated by the Lasso technique.

Model	$\bar{S}_m \times 100$	$\bar{S}_y \times 100$	$S_s \times 100$
Model "No variation"	13.13 (15.37)	20.07 (17.63)	14.05
Model "FF 3-factor"	12.51 (16.40)	16.62 (25.47)	40.54
$\lambda = 0.04$	13.05 (16.34)	19.38 (24.85)	77.34
$\lambda = 0.02$	13.05 (16.34)	19.39 (24.85)	77.34
$\lambda = 0.01$	13.11 (16.39)	19.60 (24.96)	77.94
$\lambda = 0.005$	13.07 (17.17)	21.96 (25.48)	78.33
$\lambda = 0.0025$	12.03 (19.61)	26.54 (26.52)	81.62
$\lambda = 0.00125$	9.48 (24.39)	30.18 (28.52)	81.46
Equal-weighted combination	13.31 (17.46)	26.01 (24.92)	80.84

big accumulated returns (1 month interval), small-minus-big accumulated returns (3 month interval), small-minus-big accumulated returns (12 month interval), high-minus-low accumulated returns (1 month interval), high-minus-low accumulated returns (3 months interval), high-minus-low accumulated returns (12 months interval), vix (1-month lag). All factors are used with 1-month lag.

The coefficients of Equation 3-11 are estimated by the Lasso technique using values of λ in the set $\{0.00125, 0.0025, 0.005, 0.01, 0.02, 0.04\}$. Table 3.6 presents the statistics for the models estimated by each λ and the equal-weighted combination of all models. The results give multiple insights:

1. The changes in performance for distinct values of λ suggest there is a trade-off between good performance on short-horizon and long-horizon statistics;

2. The combined forecast seems to perform better than forecasts from single models;

3. The combined forecast perform better than the benchmark models in all statistics.

Finally, we can compare the performance of the equal-weighted combination of lasso models to the out-of-sample best combinations according to $\bar{S}_m$ or $\bar{S}_y$. Compared to the $\bar{S}_m$ combination, the lasso combination presents equivalent values of $\bar{S}_m$ and $\bar{S}_y$ and greater S_s (80 versus 42). Compared to the S_y combination, the lasso combination shows lesser $\bar{S}_y$ and greater $\bar{S}_m$ and S_s.

PUC-Rio - Certificação Digital Nº 0912808/CA

PUC-Rio - Certificação Digital Nº 0912808/CA

3.5 Economic Value of each model

As an alternative way to classify the models, we attach each portfolio into a portfolio optimization framework. In resume, we optimize the portfolio of a Sharpe Ratio maximizer investor on a investable set consisting of all the 183 portfolios considered in the previous sections as available assets. We follow the optimization procedure described in Chapter 1, with two differences. First, the same covariance model is use in for all models. This ensures the benefits come from the expected return model. Second, the leverage is set to the same value at all periods.

The expected conditional covariance model is the same from the "CAPM" model in Chapter 1 - a single factor and the residual covariance between distinct assets is assumed to be 0. The leverage is adjusted so the sum of short positions equals 100% of the investor's wealth at any period. For each asset, the expected return is defined as the forecast expected return in excess of the market by a simple assumption of the market expected return: the average market return up to this period.

By following a procedure similar to Fleming et al. (2001), we estimate the economic value of the Markowitz optimized portfolios for a Sharpe Ratio maximizer investor.

For a given portfolio p, we take the average and standard deviation of monthly returns over each year i: $\bar{r}_i^p$ and std_i^p. The economic value Δ_p of portfolio p is given by:

$$\sum_{i=1}^{n} \frac{\bar{r}_i^p - \Delta_p}{std_i^p} = \sum_{i=1}^{n} \frac{\bar{r}_i^b}{std_i^b} \tag{3-12}$$

where n, $\bar{r}_i^b$, and std_i^b are the number of years and the average and standard deviation of the benchmark monthly returns, respectively. The benchmark is the portfolio optimized considering the "No variation" described in Section 3.3.

The value of Δ_p represents the amount a Sharpe Ratio maximizer investor is willing to pay to be indifferent between portfolio p and the benchmark.

Table 3.7 shows the economic value for several portfolios optimized using the Markowitz optimization. Compared to the "no variation" model, all models have positive economic value, except the "FF 3-factor" model. The naive models display a good performance, but are beat by the models we find using both the model combination and LASSO approaches. The best combination for $\bar{S}_y$ does not perform well, what is another suggestion one should look for short horizon statistics when evaluating models.

Table 3.7: The economic value of Markowitz optimized portfolios

Portfolio	Sharpe-ratio (basis points)
Model "CAPM"	50
Model "FF 3-factor"	-38
"intercept"	154
"intercept + lag"	196
Best combination for $\bar{S}_m$	229
Best combination for $\bar{S}_y$	26
LASSO (equal-weighted combination)	386

3.6 Conclusions

This work attempted to compare standard, naive and skeptical by short horizon statistics and model performance when plugged to a Markowitz optimization framework. Naive portfolios perform better than traditional models like the CAPM or the Fama-French 3 factor model according both criteria. Besides, we considered two distinct approaches to select better models. The first approach defined a model set containing several models estimated by the Fama-Macbeth and searched for better model combinations according to each of the short-run statistics. The second part defined a single model containing several factors and estimated it using the LASSO technique. Both approaches generate models which perform better according to the Markowitz experiment. These models also perform slightly better according to the short horizon statistics we propose. The first part can be seen as an expansion of the work Welch & Goyal (2008) to evaluate models that explain the cross-sectional variation of stock expected returns.

Further improvements to the current analysis may include: adding factors uncorrelated to the ones we use, obtaining theoretical results about the relation between short horizon statistics performance and economic value; expanding the portfolio set.

PUC-Rio - Certificação Digital Nº 0912808/CA