Swiss Institute for Banking and Finance
Address: Unterer Graben 21, CH-9000 St. Gallen Switzerland.

Chapter 8

Performance Analysis

Reference: Elton, Gruber, Brown, and Goetzmann (2014) 25 and 26
More advanced material is denoted by a star (*). It is not required reading.

8.1 Performance Evaluation

8.1.1 The Idea behind Performance Evaluation

Traditional performance analysis tries to answer the following question: "should we include an asset in our portfolio, assuming that future returns will have the same distribution as in a historical sample." Since returns are random variables (although with different means, variances, etc) and investors are risk averse, this means that performance analysis will typically *not* rank the fund with the highest return (in a historical sample) first. Although that high return certainly was good for the old investors, it is more interesting to understand what kind of distribution of future returns this investment strategy might entail. In short, the high return will be compared with the risk of the strategy.

Most performance measures are based on mean-variance analysis, but the full MV portfolio choice problem is not solved. Instead, the performance measures can be seen as different approximations of the MV problem, where the issue is whether we should invest in fund p or in fund q. (We don't allow a mix of them.) Although the analysis is based on the MV model, it is not assumed that all assets (portfolios) obey CAPM's beta representation—or that the market portfolio must be the optimal portfolio for every investor. One motivation of this approach could be that the investor (who is doing the performance evaluation) is a MV investor, but that the market is influenced by non-MV investors.

Of course, the analysis is also based on the assumption that historical data are good forecasters of the future.

There are several popular performance measures, corresponding to different situations: is this an investment of your entire wealth, or just a small increment? However, all these measures are (increasing) functions of Jensen's alpha, the intercept in the CAPM regression

$$R^e_{it} = \alpha_i + b_i R^e_{mt} + \varepsilon_{it}, \text{ where} \tag{8.1}$$
$$\mathrm{E}\,\varepsilon_{it} = 0 \text{ and } \mathrm{Cov}(R^e_{mt}, \varepsilon_{it}) = 0.$$

Example 8.1 *(Statistics for example of performance evaluations) We have the following information about portfolios m (the market), p, and q*

	α	β	Std(ε)	μ^e	σ
m	0.00	1.00	0.00	10.00	18.00
p	1.00	0.90	14.00	10.00	21.41
q	5.00	1.30	3.00	18.00	23.59

Table 8.1: Basic facts about the market and two other portfolios, α, β, and Std(ε) are from CAPM regression: $R^e_{it} = \alpha + \beta R^e_{mt} + \varepsilon_{it}$

8.1.2 Sharpe Ratio and M^2: Evaluating the Overall Portfolio

Suppose we want to know if fund p is better than fund q to place *all* our savings in. (We don't allow a mix of them.) The answer is that p is better if it has a higher Sharpe ratio—defined as

$$SR_p = \mu^e_p / \sigma_p. \tag{8.2}$$

The reason is that MV behaviour (MV preferences or normally distributed returns) implies that we should maximize the Sharpe ratio (selecting the tangency portfolio). Intuitively, for a given volatility, we then get the highest expected return.

Example 8.2 *(Performance measure) From Example 8.1 we get the following performance measures*

A version of the Sharpe ratio, called M^2 (after some of the early proponents of the measure: Modigliani and Modigliani) is

$$M^2_p = \mu^e_{p*} - \mu^e_m \text{ (or } \mu_{p*} - \mu_m), \tag{8.3}$$

	SR	M^2	AR	Treynor	T^2
m	0.56	0.00		10.00	0.00
p	0.47	-1.59	0.07	11.11	1.11
q	0.76	3.73	1.67	13.85	3.85

Table 8.2: Performance Measures

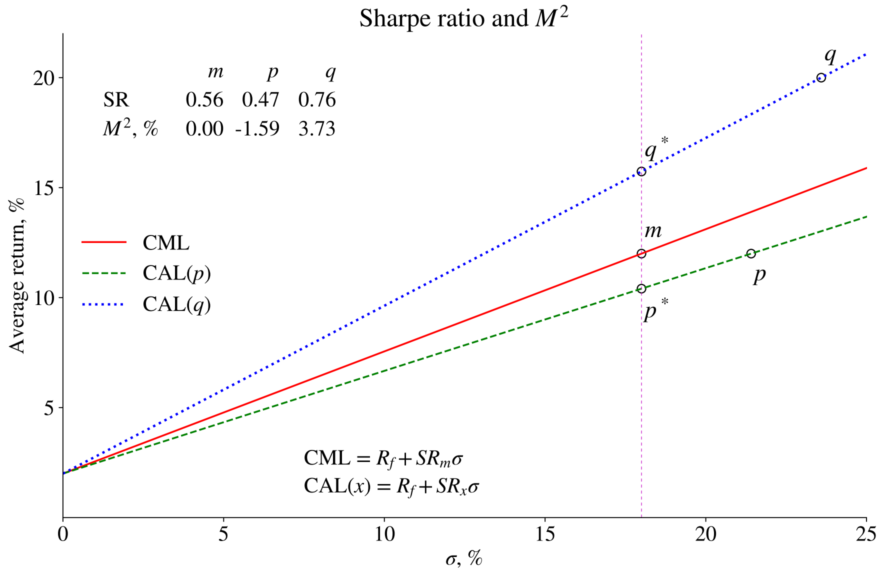

Figure 8.1: Sharpe ratio and M^2

where μ^e_{p*} is the expected return on a mix of portfolio p and the riskfree asset such that the volatility is the same as for the market return.

$$R_{p^*} = aR_p + (1-a)R_f, \text{ with } a = \sigma_m/\sigma_p. \tag{8.4}$$

This gives the mean and standard deviation of portfolio p^*

$$\mu^e_{p^*} = a\mu^e_p = \mu^e_p\sigma_m/\sigma_p \tag{8.5}$$

$$\sigma_{p^*} = a\sigma_p = \sigma_m. \tag{8.6}$$

The latter shows that R_{p^*} indeed has the same volatility as the market. See Example 8.2 and Figure 8.1 for an illustration.

M^2 has the advantage of being easily interpreted—it is just a comparison of two returns. It shows how much better (or worse) this asset is compared to the capital market line (which is the location of efficient portfolios provided the market is MV efficient).

However, it is just a scaling of the Sharpe ratio.

To see that, use (8.2) to write

$$\begin{aligned} M_p^2 &= SR_{p^*}\sigma_{p^*} - SR_m\sigma_m \\ &= \left(SR_p - SR_m\right)\sigma_m. \end{aligned} \tag{8.7}$$

The second line uses the facts that R_{p^*} has the same Sharpe ratio as R_p (see (8.5)–(8.6)) and that R_{p^*} has the same volatility as the market. Clearly, the portfolio with the highest Sharpe ratio has the highest M^2.

8.1.3 Appraisal Ratio: Which Portfolio to Combine with the Market Portfolio?

If the issue is "should I *add* fund p or fund q to my holding of the market portfolio?," then the appraisal ratio provides an answer. The appraisal ratio of fund p is

$$AR_p = \alpha_p / \operatorname{Std}(\varepsilon_{pt}), \tag{8.8}$$

where α_p is the intercept and $\operatorname{Std}(\varepsilon_{pt})$ is the volatility of the residual of a CAPM regression (8.1). (The residual is often called the tracking error.) A higher appraisal ratio is better.

If you think of $b_p R^e_{mt}$ as the benchmark return, then AR_p is the average extra return per unit of extra volatility (standard deviation). For instance, a ratio of 1.7 could be interpreted as a 1.7 USD profit per each dollar risked.

The motivation is that if we take the market portfolio and portfolio p to be the available assets, and then find the optimal (assuming MV preferences) combination of them, then the squared Sharpe ratio of the optimal portfolio (that is, the tangency portfolio) is

$$SR_c^2 = \left(\frac{\alpha_p}{\operatorname{Std}(\varepsilon_{pt})}\right)^2 + SR_m^2. \tag{8.9}$$

If the alpha is positive, a higher appraisal ratio gives a higher Sharpe ratio—which is the objective if we have MV preferences. See Example 8.2 for an illustration.

If the alpha is negative, and we rule out short sales, then (8.9) is less relevant. In this case, the optimal portfolio weight on an asset with a negative alpha is (very likely to be) zero—so those assets are uninteresting.

The *information ratio*

$$IR_p = \frac{\operatorname{E}(R_p - R_b)}{\operatorname{Std}(R_p - R_b)}, \tag{8.10}$$

where R_b is some benchmark return. The information ratio is similar to the appraisal ratio—although a bit more general. The denominator in 8.10 can be thought of as the tracking error relative to the benchmark—and the numerator as the average active return (the gain from actively deviating from the benchmark). Notice, however, that when the benchmark is $b_p R^e_{mt}$, then the information ratio is the same as the appraisal ratio. Instead, when R_f is the benchmark, then the information ratio equals the Sharpe ratio.

Proof. From the CAPM regression (8.1) we have

$$\operatorname{Cov}\begin{bmatrix} R^e_{it} \\ R^e_{mt} \end{bmatrix} = \begin{bmatrix} \beta_i^2\sigma_m^2 + \operatorname{Var}(\varepsilon_{it}) & \beta_i\sigma_m^2 \\ \beta_i\sigma_m^2 & \sigma_m^2 \end{bmatrix}, \text{ and } \begin{bmatrix} \mu^e_i \\ \mu^e_m \end{bmatrix} = \begin{bmatrix} \alpha_i + \beta_i\mu^e_m \\ \mu^e_m \end{bmatrix}.$$

Suppose we use this information to construct a mean-variance frontier for both R_{it} and R_{mt}, and we find the tangency portfolio, with excess return R^e_{ct}. We assume that there are no restrictions on the portfolio weights. Recall that the square of the Sharpe ratio of the tangency portfolio is $\mu^{e\prime}\Sigma^{-1}\mu^e$, where μ^e is the vector of expected excess returns and Σ is the covariance matrix. By using the covariance matrix and mean vector above, we get that the squared Sharpe ratio for the tangency portfolio (using both R_{it} and R_{mt}) is

$$\left(\frac{\mu^e_c}{\sigma_c}\right)^2 = \frac{\alpha_i^2}{\operatorname{Var}(\varepsilon_{it})} + \left(\frac{\mu^e_m}{\sigma_m}\right)^2.$$

■

8.1.4 Treynor's Ratio and T^2: Portfolio is a Small Part of the Overall Portfolio

Suppose instead that the issue is if we should add a *small* amount of fund p or fund q to an already well diversified portfolio (not the market portfolio). In this case, Treynor's ratio might be useful

$$TR_p = \mu^e_p/\beta_p. \tag{8.11}$$

A higher Treynor's ratio is better.

The TR measure can be rephrased in terms of expected returns—and could then be called the T^2 measure. Mix p and q with the riskfree rate to get the same β for both portfolios (here 1 to make it comparable with the market), the one with the highest Treynor's ratio has the highest expected return (T^2 measure). To show this consider the portfolio p^*

$$R_{p^*} = aR_p + (1-a)R_f, \text{ with } a = 1/\beta_p. \tag{8.12}$$

This gives the mean and the beta of portfolio p^*

$$\mu^e_{p^*} = a\mu^e_p = \mu^e_p/\beta_p \tag{8.13}$$

$$\beta_{p^*} = a\beta_p = 1, \tag{8.14}$$

so the beta is one. We then define the T^2 measure as

$$T^2_p = \mu^e_{p^*} - \mu^e_m = \mu^e_p/\beta_p - \mu^e_m, \tag{8.15}$$

so the ranking (of fund p and q, say) in terms of Treynor's ratio and the T^2 are the same. See Example 8.2 and Figure 8.2 for an illustration.

The basic intuition is that with a *diversified portfolio* and *small investment*, idiosyncratic risk doesn't matter, only systematic risk (β) does. Compare with the setting of the Appraisal Ratio, where we also have a well diversified portfolio (the market), but the investment could be large.

Example 8.3 *(Additional portfolio risk) We hold a well diversified portfolio (d) and buy a fraction 0.05 of asset i (financed by borrowing), so the return is $R = R_d + 0.05\left(R_i - R_f\right)$. Suppose $\sigma^2_d = \sigma^2_i = 1$ and that the correlation of d and i is 0.25. The variance of R is then*

$$\sigma^2_d + \delta^2\sigma^2_i + 2\delta\sigma_{id} = 1 + 0.05^2 + 2 \times 0.05 \times 0.25 = 1 + 0.0025 + 0.025,$$

so the importance of the covariance is 10 times larger than the importance of the variance of asset i.

Proof. (*Version 1: Based on the beta representation.) The derivation of the beta representation shows that for all assets $\mu^e_i = \mathrm{Cov}\left(R_i, R_m\right) A$, where A is some constant. Rearrange as $\mu^e_i/\beta_i = A\sigma^2_m$. A higher ratio than this is to be considered as a positive "abnormal" return and should prompt a higher investment. ■

Proof. (*Version 2: From first principles, kind of a proof...) Suppose we initially hold a well diversified portfolio (d) and we increase the position in asset i with the fraction δ by borrowing at the riskfree rate to get the return

$$R = R_d + \delta\left(R_i - R_f\right).$$

The incremental (compared to holding portfolio d) expected excess return is $\delta\mu^e_i$ and the incremental variance is $\delta^2\sigma^2_i + 2\delta\sigma_{id} \approx 2\delta\sigma_{id}$, since δ^2 is very small. (The variance of R

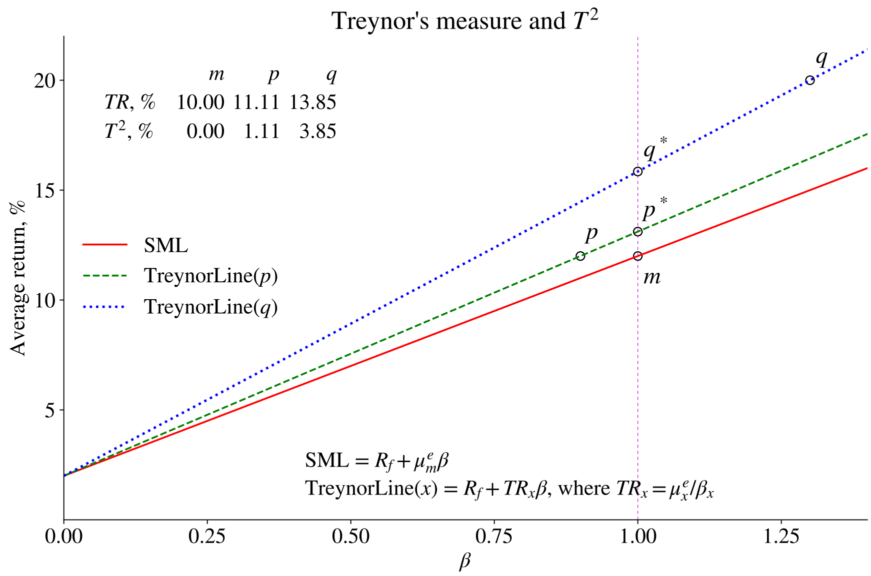

Figure 8.2: Treynor's ratio

is $\sigma_d^2 + \delta^2\sigma_i^2 + 2\delta\sigma_{id}$.) To a first-order approximation, the change (E $R_p - \text{Var}(R_p)k/2$) in utility is therefore $\delta\mu_i^e - k\delta\sigma_{id}$, so a high value of μ_i^e/σ_{id} will increase utility. This suggests μ_i^e/σ_{id} as a performance measure. However, if portfolio d is indeed well diversified, then $\sigma_{id} \approx \sigma_{im}$. We could therefore use μ_i^e/σ_{im} or (by multiplying by σ_{mm}), μ_i^e/β_i as a performance measure. ■

8.1.5 Relationships among the Various Performance Measures

The different measures can give different answers when comparing portfolios, but they all share one thing: they are increasing in Jensen's alpha. By using the expected values from the CAPM regression ($\mu_p^e = \alpha_p + \beta_p\mu_m^e$), simple rearrangements give

$$\begin{aligned} SR_p &= \frac{\alpha_p}{\sigma_p} + \text{Corr}(R_p, R_m)SR_m \\ AR_p &= \frac{\alpha_p}{\text{Std}(\varepsilon_{pt})} \\ TR_p &= \frac{\alpha_p}{\beta_p} + \mu_m^e. \end{aligned} \tag{8.16}$$

and M^2 is just a scaling of the Sharpe ratio. Notice that these expressions do not assume that CAPM is the right pricing model—we just use the definition of the intercept and slope in the CAPM regression.

Since Jensen's alpha is the driving force in all these measurements, it is often used as performance measure in itself. In a sense, we are then studying how "mispriced" a fund is—compared to what it should be according to CAPM. That is, the alpha measures the "abnormal" return.

Proof. (of (8.16)*) Taking expectations of the CAPM regression (8.1) gives $\mu_p^e = \alpha_p + \beta_p \mu_m^e$, where $\beta_p = \text{Cov}(R_p, R_m)/\sigma_m^2$. The Sharpe ratio is therefore

$$SR_p = \frac{\mu_p^e}{\sigma_p} = \frac{\alpha_p}{\sigma_p} + \frac{\beta_p}{\sigma_p}\mu_m^e,$$

which can be written as in (8.16) since

$$\frac{\beta_p}{\sigma_p}\mu_m^e = \frac{\text{Cov}(R_p, R_m)}{\sigma_m \sigma_p}\frac{\mu_m^e}{\sigma_m}.$$

The AR_p in (8.16) is just a definition. The TR_p measure can be written

$$TR_p = \frac{\mu_p^e}{\beta_p} = \frac{\alpha_p}{\beta_p} + \mu_m^e,$$

where the second equality uses the expression for μ_p^e from above. ∎

	α	SR	M^2	AR	Treynor	T^2
Market	0.00	0.35	0.00		6.39	0.00
Putnam	0.16	0.35	−0.06	0.04	6.59	0.20
Vanguard	2.36	0.54	3.38	0.61	10.38	3.99

Table 8.3: Performance Measures of Putnam Asset Allocation: Growth A and Vanguard Wellington, weekly data 1999:01-2020:12 (annualized figures)

8.1.6 Performance Measurement with More Sophisticated Benchmarks

Traditional performance tests typically rely on the alpha from a CAPM regression. The benchmark for the evaluation is then effectively a fixed portfolio consisting of assets that are correctly priced by the CAPM (obeys the beta representation). It often makes sense to use a more demanding benchmark. There are several popular alternatives.

If there are predictable movements in the market excess return, then it makes sense to add a "market timing" factor to the CAPM regression. For instance, Treynor and Mazuy (1966) argue that market timing is similar to having a beta that is linear in the market excess return

$$\beta_i = b_i + c_i R^e_{mt}. \tag{8.17}$$

Using in a traditional market model (CAPM) regression, $R^e_{it} = a_i + \beta_i R^e_{mt} + \varepsilon_{it}$, gives

$$R^e_{it} = a_i + b_i R^e_{mt} + c_i (R^e_{mt})^2 + \varepsilon_{it}, \tag{8.18}$$

where c captures the ability to "time" the market. That is, if the investor systematically gets out of the market (maybe investing in a riskfree asset) before low returns and vice versa, then the slope coefficient c is positive. The interpretation is not clear cut, however. If we still regard the market portfolio (or another fixed portfolio that obeys the beta representation) as the benchmark, then $a + c(R^e_{mt})^2$ should be counted as performance. In contrast, if we think that this sort of market timing is straightforward to implement, that is, if the benchmark is the market plus market timing, then only a should be counted as performance.

In other cases (especially when we think that CAPM gives systematic pricing errors), the performance is measured by the intercept of a multifactor model like the Fama-French model.

A recent way to merge the ideas of market timing and multi-factor models is to allow the coefficients to be time-varying. In practice, the coefficients in period t are only allowed to be linear (or affine) functions of some information variables in an earlier period, z_{t-1}. To illustrate this, suppose z_{t-1} is a single variable, so the time-varying (or "conditional") CAPM regression is

$$\begin{aligned} R^e_{it} &= (a_i + \gamma_i z_{t-1}) + (b_i + \delta_i z_{t-1}) R^e_{mt} + \varepsilon_{it} \\ &= \theta_{i1} + \theta_{i2} z_{t-1} + \theta_{i3} R^e_{mt} + \theta_{i4} z_{t-1} R^e_{mt} + \varepsilon_{it}. \end{aligned} \tag{8.19}$$

Similar to the market timing regression, there are two possible interpretations of the results: if we still regard the market portfolio as the benchmark, then the other three terms should be counted as performance. In contrast, if the benchmark is a dynamic strategy in the market portfolio (where z_{t-1} is allowed to affect the choice market portfolio/riskfree asset), then only the first two terms are performance. In either case, the performance is time-varying.

8.2 Holdings-Based Performance Measurement

As a complement to the purely return-based performance measurements discussed, it may also be of interest to study how the portfolio weights change (if that information is available). This highlights how the performance has been achieved.

Grinblatt and Titman's measure (in period t) is

$$GT_t = \sum\nolimits_{i=1}^{n} (w_{i,t-1} - w_{i,t-2}) R_{it}, \tag{8.20}$$

where $w_{i,t-1}$ is the weight on asset i in the portfolio chosen (at the end of) in period $t-1$ and $R_{i,t}$ is the return of that asset between (the end of) period $t-1$ and (end of) t. A positive value of GT_t indicates that the fund manager has moved into assets that turned out to give positive returns.

It is common to report a time-series average of GT_t, for instance over the sample $t = 1$ to T.

8.3 Performance Attribution

The performance of a fund depends on decisions taken on several levels. In order to get a better understanding of how the performance was generated, a performance attribution calculation can be very useful. It uses information on portfolio weights (for instance, in-house information) to decompose overall performance according to a number of criteria (typically related to different levels of decision making).

For instance, it could be to decompose the return (as a rough measure of the performance) into the effects of *(a)* allocation to asset classes (equities, bonds, bills); and *(b)* security choice within each asset class. Alternatively, for a pure equity portfolio, it could be the effects of *(a)* allocation to industries; and *(b)* security choice within each industry.

Consider portfolios p and b (for benchmark) from the same set of assets. Let n be the number of asset classes (or industries). Returns are

$$R_p = \sum_{i=1}^{n} w_i R_{pi} \text{ and } R_b = \sum_{i=1}^{n} v_i R_{bi}, \tag{8.21}$$

where w_i is the weight on asset class i (for instance, long T-bonds) in portfolio p, and v_i is the corresponding weight in the benchmark b. Analogously, R_{pi} is the return that the portfolio earns on asset class i, and R_{bi} is the return the benchmark earns. In practice, the benchmark returns are typically taken from well established indices.

Form the difference and rearrange ($(\pm w_i R_{bi})$) to get

$$R_p - R_b = \sum_{i=1}^{n} \left(w_i R_{pi} - v_i R_{bi} \right)$$
$$= \underbrace{\sum_{i=1}^{n} (w_i - v_i) R_{bi}}_{\text{allocation effect}} + \underbrace{\sum_{i=1}^{n} w_i \left(R_{pi} - R_{bi} \right)}_{\text{selection effect}}. \tag{8.22}$$

The first term is the *allocation effect* (that is, the importance of allocation across asset classes) and the second term is the *selection effect* (that is, the importance of selecting the individual securities within an asset class). In the first term, $(w_i - v_i) R_{bi}$ is the contribution from asset class (or industry) i. It uses the benchmark return for that asset class (as if you had invested in that index). Therefore the allocation effect simply measures the contribution from investing more/less in different asset class than the benchmark. If decisions on allocation to different asset classes are taken by senior management (or a board), then this is the contribution of that level. In the selection effect, $w_i \left(R_{pi} - R_{bi} \right)$ is the contribution of the security choice (within asset class i) since it measures the difference in returns (within that asset class) of the portfolio and the benchmark.

Remark 8.4 *(Alternative expression for the allocation effect*) The allocation effect is sometimes defined as* $\sum_{i=1}^{n} (w_i - v_i)(R_{bi} - R_b)$*, where* R_b *is the benchmark return. This is clearly the same as in (8.22) since* $\sum_{i=1}^{n} (w_i - v_i) R_b = R_b \sum_{i=1}^{n} (w_i - v_i) = 0$ *(as both sets of portfolio weights sum to unity).*

8.3.1 What Drives Differences in Performance across Funds?

Reference: Ibbotson and Kaplan (2000)

Plenty of research shows that the asset allocation (choice between markets or large market segments) is more important for mutual fund returns than the asset selection (choice of individual assets within a market segment). For other investors, including hedge funds, the leverage also plays a main role.

8.4 Style Analysis

Reference: Sharpe (1992)

Style analysis is a way to use econometric tools to find out the portfolio composition from a series of the returns, at least in broad terms.

The basic idea is to identify a number (5 to 10 perhaps) return indices that are expected to account for the brunt of the portfolio's returns, and then run a regression to find the portfolio "weights." It is essentially a multi-factor regression without any intercept and where the coefficients are constrained to sum to unity and to be positive

$$R^e_{pt} = \sum_{j=1}^{K} b_j R^e_{jt} + \varepsilon_{pt}, \text{ with} \tag{8.23}$$
$$\sum_{j=1}^{K} b_j = 1 \text{ and } b_j \geq 0 \text{ for all } j.$$

The coefficients are typically estimated by minimizing the sum of squared residuals. This is a nonlinear estimation problem, but there are very efficient methods for it (since it is a quadratic problem). Clearly, the restrictions could be changed to $U_j \leq b_j \leq L_j$, which could allow for short positions.

A pseudo-R^2 (the squared correlation of the fitted and actual values) is sometimes used to gauge how well the regression captures the returns of the portfolio. The residuals can be thought of as the effect of stock selection, or possibly changing portfolio weights more generally. One way to get a handle of the latter is to run the regression on a moving data sample. The time-varying weights are often compared with the returns on the indices to see if the weights were moved in the right direction.

See Figure 8.3 for an example.

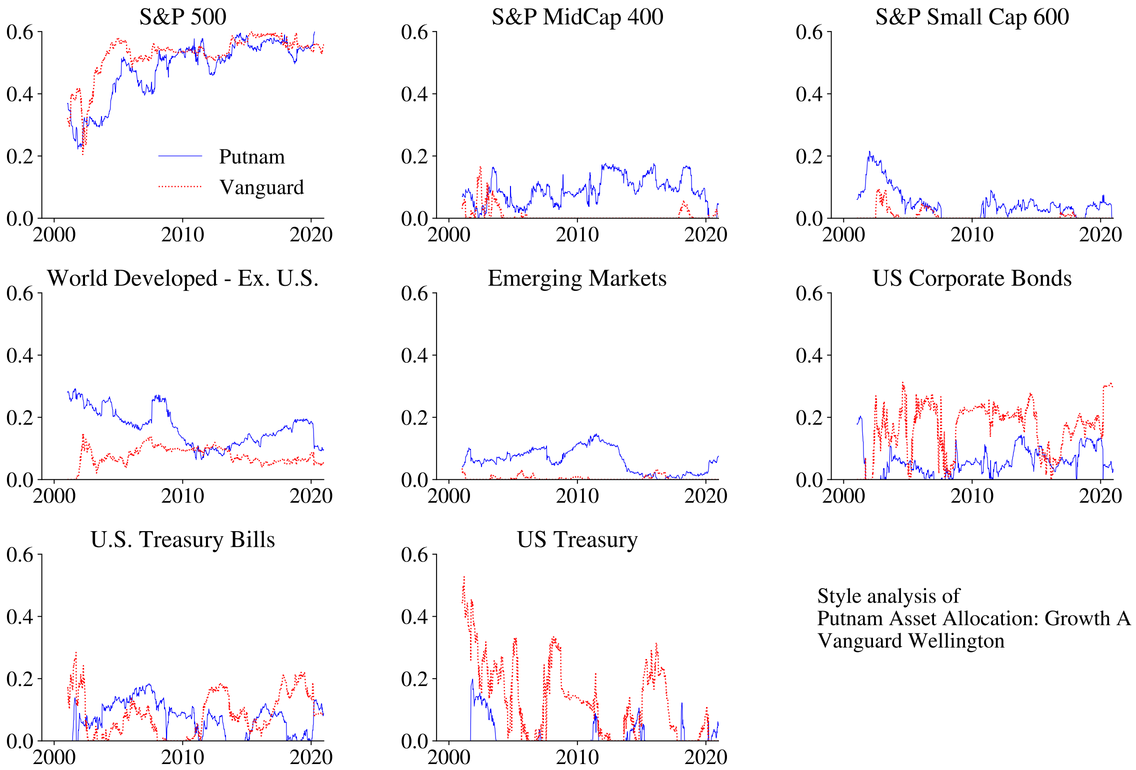

Figure 8.3: Example of style analysis, rolling data window