Beta vs. Characteristics: A Practitioner's Perspective

Daehwan Kim

Department of Economics, Konkuk University, Seoul, South Korea

daehwan.kim@post.harvard.edu

January 14, 2013

Abstract

The first part of the paper demonstrates the similarity between the beta model (a.k.a. covariance model) and the characteristics model in the ability to reduce portfolio risk. When the two models are applied to the S&P 500 stocks in the period between 1980 and 2010, one model appears better than the other for some sub-periods, but the pattern is reversed for other sub-periods. Also, the difference, when it exists, is not economically significant. In the second part of the paper, the similarity between the two models is traced back to the correlation between beta and characteristics and also the correlation between factor and the characteristics price. A Monte Carlo simulation suggests that the difference between the two models is unlikely to increase substantially even if the underlying correlation structure changes unfavorably.

Keywords

Beta model, Characteristics model, Portfolio risk, Beta-characteristics correlation, Factor-characteristics price correlation

JEL Classification

G11, G12, G17

1. Introduction

You are about to create a portfolio of stocks. For the purpose of risk control, you want to use a model of stock returns. You do not have many choice alternatives. Most likely, you will consider only two models as serious alternatives: the beta model (a.k.a covariance model[1]) and the characteristics model. To make an informed choice in this situation, what are the things that you need to know? Probably you will ask yourself the following questions:

- Which model performed better historically?
- What is the loss when you choose a 'wrong' model?
- What influences the relative performances of the models?
- What is the worst that can happen when you choose a 'wrong' model?

This paper attempts to answer these questions. The main findings of the paper can be summarized as follows: No one model dominates the other in the historical performance. In some periods, the beta model produced a lower-risk portfolio whereas, in other periods, the characteristics model did better. Was the difference significant? Statistically, "sometimes," but economically "not much." The magnitude of the loss, measured in terms of the portfolio volatility, was small. The worse model increased the annualized standard deviation by 0.5% or so. In terms of the tracking error, the loss was even smaller. Given the small magnitude of the loss, one is led to the question of why these two models are so similar. Two things are at work here: Beta and characteristics are correlated, and also factor and the characteristics price are correlated. Finally, if these correlations disappear, will the loss from choosing a wrong model increase substantially? My analysis suggests that the loss is not going to increase much. There is more room for loss reduction than for loss increase from the current level. The rest of the paper elaborates on these findings.

There is a large body of literature that compares the beta model and the characteristics model in terms of their ability to estimate mean returns. This literature might be called the 'beta vs. characteristics debate.'[2] Daniel and Titman (1997) have started the debate, arguing in favor of characteristics. David, Fama, and French (2000) have presented a counter-argument based on new data. In response, Daniel, Titman, and Wei (2001) have provided evidence from Japan in favor of characteristics. Many other authors, including Pastor and Stambaugh (2000), Brennan, Chordia, and Subrahmanyam (1998), and Lewellen (1999) have contributed to this literature.

Less attention has been paid to the models' ability to estimate and control risk. The first major

---

[1] I use the term 'beta model' rather than 'covariance model' as the latter can be awkward in an expression such as 'the covariance forecast of the covariance model.'

[2] It is better known as 'covariance vs. characteristics debate.' For the reason explained in the previous footnote, I use the term 'beta' instead of 'covariance.'

contributor here is Connor (1995). He emphasized the fact that many practitioners use these models to control risk, rather than to estimate mean returns. His analysis, however, is limited in that he has examined only the in-sample fits of the models. The paper by Chan, Karceksi, and Lakonishok (1999) comes close to the current paper in many respects. Chan et al. have examined the economic consequences in terms of the realized variances of the global minimum variance portfolios and the realized tracking errors of the global minimum tracking error portfolios. The main difference between the current paper and that by Chan et al. is that Chan et al. have not considered the characteristics model as conventionally used.[3] Thus, the analysis by Chan et al. does not answer the question of how the characteristics model fares compared to the beta model. In addition, Chan et al. have not examined the sources of the similarity between the two models, and the range of the potential loss of the model choice.

For the first part of the paper, I consider two beta models: the one factor model in the spirit of the Capital Asset Pricing Model (which I denote as BM1) and the three-factor model of Fama and French (BM3). I also consider the “zero-factor” model (BM0), where the sample variance-covariance matrix of returns is used as the variance-covariance matrix estimate. For the characteristics model, I consider a two-stage model (CM2), where the first stage is identical to the CAPM and, in the second stage, the residual from the first stage is regressed on firm characteristics. The market capitalization and the book-to-market ratio are the characteristics that I use. Later analyses focus on the two of the four models: the three-factor beta model (BM3) and the two-stage characteristics model (CM2).

Data for S&P 500 stocks, from 1975 to 2010, are used in this study. First, I consider an investor at the end of 1980, who estimates the models using the past 5-year data (1976-1980). Based on the model estimates, the investor creates portfolios. I track the performance of these portfolios for the next 5 years (1981-1985). Then I move the time window 5 years forward. Now is the end of 1985. The investor estimates the models using the past 5-year data (1981-1985), and creates portfolios based on the estimates. I track the performance of these portfolios for the next 5 years (1986-1990). And I repeat the same analysis every 5 years. In total, there are 6 sets of analyses, corresponding to 6 portfolio formation years (1980, 1985, 1990, 1995, 2000, and 2005).

The 5-year holding period is longer than the holding period of other studies.[4] There are two reasons for this choice. First, if the holding period is short, then realized variances and covariances need to be

[3] Instead, they have considered a characteristics-matching approach.

[4] For example, Chan, Karceksi, and Lakonisho (1999) consider one-year and three-year holding periods.

calculated from short time-series, which introduces a lot of noise into the analysis.[5] Second, to improve the comparison of the characteristics model and the beta model, I assume that characteristics remain relevant for the entire estimation period (5 years). Thus, it does not seem excessive to assume that characteristics remain relevant for the next 5 years as well. The same can be said about the beta.

A somewhat non-standard feature in my analysis is that I use only the initial characteristics in the estimation of the characteristics model. That is, I take the characteristics from the beginning of the estimation period and do not update them. The reason I do this is to put the characteristics model and the beta model in the equal footing. In the estimation of the beta model, the beta is assumed to be constant throughout the estimation period. Thus, updating the characteristics would be advantageous to the characteristics model. More details on the models and the estimation are explained in Section 3.

I carry out four sets of post-estimation analyses. First, I examine the accuracy of the variance and covariance forecasts from each model (Section 5). Each model generates a variance forecast for each stock, and a covariance forecast for each pair of stocks. I compare these forecasts to the realized values, and assess the forecast accuracy using the root mean squared error (RMSE) and Pearson's and Kendall's correlation coefficients. Then the forecast accuracies of alternative models are compared. It turns out that the alternative models have comparable forecast accuracies. While this analysis is interesting in its own right, it is limited in that the economic consequence is not considered.

The second set of analyses (Section 6) focuses on the economic consequences. I compare the realized variances of the global minimum variance portfolios and the realized tracking errors of the global minimum tracking error portfolios. The word 'global' in the name indicates that these portfolios have the smallest variances (tracking errors) among the efficient portfolios. Thus, mean estimates are not used in the construction of these portfolios. The difficulty of estimating means is well known; to avoid potential problems arising from mean estimates, I only consider these portfolios. This is also the approach taken by Chan, Karceksi, and Lakonishok (1999).

In terms of realized variances, the beta model turns out to do somewhat better than the characteristics model. The beta model has lower variances than the characteristics model in four out of six periods; however, the difference is significant only in one period. In terms of the tracking errors, the beta model does better than the characteristics model in five out of six periods, two of them having

---

[5] Chan, Karceksi, and Lakonisho (1999) have noted this after comparing the estimates from the one-year horizon and the three-year horizon. See page 949 of their article.

significant differences. There is, however, an opposite case as well: the characteristics model is significantly better than the beta model in one period. The magnitude of difference is relatively small. The annualized standard deviation increases only by 0.5% when the characteristics model is used; the tracking errors change even less. Thus, it would be fair to say that the difference between the two models is not economically significant.

In the third set of analyses (Section 7), I identify the parameters that affect the relative performances of the models, and estimate those parameters. It has been noted by many authors that the correlation between beta and characteristics makes the two models similar. I develop this idea further, and show how the multivariate correlation between beta and characteristics influences the similarity between the variance-covariance matrix estimates of the alternative models. I also look at a less noted aspect: the correlation between factor and the characteristics price matters as well. Here again, I show how the multivariate correlation between factor and the characteristics price affects the variance-covariance matrix estimates. Using 2001-2005 data, I estimate the relevant parameters, and show that substantial correlations do exist in the data.

I interpret the increase in realized variances or tracking errors as the 'loss' from choosing a 'wrong' model. The final analysis (Section 8) is a Monte Carlo experiment designed to check the sensitivity of the 'loss' to the variations in those parameters highlighted above. I vary the data generating process such that the correlation between beta and characteristics is at its maximum and also at its minimum. I also adjust the correlation between factor and the characteristics price. The goal is to show the range of the loss that an investor may encounter as the correlation structure changes. It turns out that there is not much room for loss increase. That is, the current loss estimate is close to the maximum value.

The bottom line of the analysis is that, at least for practitioners, which model to choose is probably not as important as it seems. Surveying the documents of commercial software (Menchero, Orr, and Wang, 2011; Northfield, 2012a, 2012b; Asprouli and Lamb, 2002) one gets the sense that the characteristics model is more popular among practitioners, whereas the opposite is probably true among academics. The disparity is somewhat understandable in light of my findings. As far as practitioners' use of these models is concerned, the two models are probably not very different.

The rest of the paper is organized as follows. Section 2 includes literature review. The models and the data are described in Sections 3 and 4, respectively. Sections 5 through 8 provide the main analyses of the paper. Section 9 concludes.

2. Literature

I first review the literature related to the “beta vs. characteristics debate,”[6] and comment on the differences in perspective from the current study. This literature focuses on the ability of the models to estimate expected returns, whereas the primary interest of the current paper is the ability of the models to estimate and control risk. Those papers directly relevant to risk control are reviewed next.

The “beta vs. characteristics debate” literature can be divided into three strands in terms of methodologies and perspectives. The first strand includes the seminal paper by Daniel and Titman (1997) and the paper by Davis, Fama and French (2000). Their primary methodology is triple sort, and characteristics are allowed to influence expected returns, but not volatility. The second strand includes Brennan, Chordia, and Subrahmanyam (1998) and Lewellen (1999). In these studies, the residuals from a beta model are regressed on characteristics, allowing characteristics to influence both volatility and expected returns. The last strand includes the studies following Fama-MacBeth tradition. Raw returns (i.e. not residuals) are regressed on characteristics, and characteristics replace the role of beta completely. I review each of these three groups of studies, in turn.

Daniel and Titman (1997) propose the view that “characteristics matter more,” and Daniel, Titman, and Wei (2001) elaborate on this further. These studies adopt the ‘triple sort’ methodology; i.e. stocks are first sorted by size and the book-to-market ratio (BM), and then again by beta. When 9 size-BM sorted groups are created, one may create 3 beta-sorted groups within each size-BM group. By taking a long position in the top beta group and a short position in the bottom beta group, it is possible to create a ‘characteristics-balanced’ portfolio within each size-BM group. Then the zero-returns of the characteristics-balanced portfolios are the evidence for the view that characteristics matter more. Alternatively, when the returns of the characteristics-balanced portfolios are regressed on the SMB and the HML factors, the intercept estimate can be considered the average return of the ‘beta-balanced’ portfolio. The non-zero returns of the beta-balanced portfolios indicate the supremacy of characteristics. Davis, Fama and French (2000) repeat this analysis, and show that Daniel and Titman’s results are peculiar to the 1973-1993 sample, and the opposite conclusion can be drawn for the period before and after that. In response, Daniel, Titman, and Wei (2001) present the evidence from the Japanese stock market in favor of the characteristics view.[7]

---

[6] See footnote 2.

[7] Hou, Karolyi, and Ko (2011) present further evidence from international stock markets. Their evidence

Pastor and Stambaugh (2000) apply a Bayesian perspective on the triple sort methodology. They interpret the alternative models as the alternative priors on expected returns: the prior corresponding to the beta model has the zero-intercept restriction, whereas the prior corresponding to the characteristics model requires zero expected returns on the characteristics-matched portfolios. They then analyze the investment choices of the investors with the alternative priors, and the 'utility loss' of the investors when they are forced to take someone else' portfolios. Their conclusion is that, when real-life constraints are considered, the difference between the models is not significant.

Daniel and Titman (1997) and other authors have noted that whether the beta model can be distinguished from the characteristics model depends on the correlation between beta and characteristics. If the correlation is perfect, distinguishing between the two models becomes impossible. While these authors have not reported the correlation between beta and characteristics, the beta of the characteristics-balanced portfolio is indicative of this correlation. In case of the HML-beta, it is highly significant with the t statistic of up to 14. I extend this line of thoughts further in this paper. I examine the consequence of the correlation between beta and characteristics. Daniel and Titman (1997) and other authors have also noted that the difference between the two models is minor when factor returns are small.[8] In later analyses, I extend this idea as well. I show that the correlation between factor returns (such as the SMB returns) and the characteristics price affects the relative performances of the models.

From a practitioner's perspective, one limitation of these studies is that the characteristics model, in the sense that a practitioner uses, has not been considered seriously. In a practitioner's view, characteristics influence return volatility as well as expected returns. Not so in the studies reviewed above. In the framework suggested by Daniel, Titman, and Wei (2001), characteristics affect expected returns, but not the variances of returns. The variances of returns solely depend on beta. This idea can be illustrated by the following system of equations:

$$\begin{aligned} R_{it} &= \mathrm{E}[R_{it}] + \beta_i f_t + \varepsilon_{it} \\ \mathrm{E}[R_{it}] &= \alpha + \delta x_i + \lambda \beta_i \end{aligned} \tag{1}$$

Above, x and β are vectors of characteristics and beta, respectively, and δ and λ are the characteristics price and the factor price, respectively. Note that there are two sources of time

---

supports the beta view as far as the cash flow-to-price (C/P) factor is concerned. For the other factors, the evidence is in favor of the characteristics view.

[8] Daniel and Titman (1997) noted that the SMB factor returns make the power of their test small.

variations: f_t and $\varepsilon_{\mathrm{it}}$. x and β can be allowed to vary over time, but this is not relevant for now. A practitioner would like to allow the characteristics price δ to vary over time, just as factor realization f_t varies over time. Such possibility is not considered in the above studies.

In the studies by Brennan, Chordia, and Subrahmanyam (1998) and by Lewellen (1999), characteristics are allowed to influence return volatility though that aspect is not the focus of these studies. In these studies, the residuals from the beta model are regressed on characteristics. If characteristics matter, the regression coefficients would be significant. If characteristics do not matter, the coefficients would be insignificant. Brennan, Chordia, and Subrahmanyam (1998) focus on the cross-sectional relationships, and find that characteristics matter. Lewellen (1999) focuses on the time-series relationship, and reaches at the opposite conclusion.

Characteristics have a more prominent role in the tradition of the Fama-Macbeth methodology. In the original Fama-Macbeth article (1973), the only characteristics variable is the specific risk (i.e. residual volatility). In Fama and French (1992), however, both ME and BM, together with beta, are included as regressors in cross-sectional regressions. In Fama and French (1997), ME and BM are included as regressors in time-series regressions. Fama and French (2008) adopt characteristics-only cross-sectional regression. These studies, however, maintain the view that beta is more important. Characteristics are used, not because they affect returns, but because they are good proxies for beta.

The use of firm characteristics for the purpose of risk control can be traced back to Rosenberg (1974). His paper has been influential among the practitioners, partly because Rosenberg has developed software (BARRA) based on his idea, and his software became very popular among practitioners. Other software such as Northfield and BITA Plus follows the similar ideas.[9] Connor (1995) follows this tradition. He compares alternative models in terms of the models' ability to explain volatility. One limitation is that his study focuses on the in-sample fit of alternative models.

While not examining the characteristics model, Chan, Karceski, and Lakonishok (1999) investigate the same question that Connor has examined: how alternative models perform in terms of controlling risk. Their innovation is to focus on the economic consequence. They compare models in terms of the realized variances of the global minimum variance portfolios and the realized tracking errors of the

9 See Menchero, Orr, and Wang (2011), Northfield (2012a), Asprouli and Lamb (2002) for an overview of the BARRA, Northfield, and BITA Plus models, respectively. Northfield (2012b) shows how the beta model is implemented by this type of software.

global minimum tracking error portfolios. The current paper complements Chan, Karceski, and Lakonishok (1999) by including the characteristics model and also expanding the analysis in several directions.

3. Models

The goal of this paper is to compare the beta model and the characteristics model. I consider two versions of the beta model—the one factor model in the spirit of CAPM (BM1) and the three factor model of Fama and French (BM3). The characteristics model is a two-stage model (CM2) where the first stage is the CAPM and the second stage includes two characteristics variables. I also consider a zero-factor model (BM0), whose only parameters are the mean vector and the variance-covariance matrix. So, in total, four models are considered in this paper.

Certainly, there are other interesting models that I do not consider here. For the purpose of risk control, some shrinkage-based models have been extensively discussed. (See, for example, Ledoit and Wolf (2003).) I do not consider these models here as the main focus is the relative merit of the beta model and the characteristics model.

The primary use of a model in this study is to produce an estimator for the variance-covariance matrix of excess returns Σ_r and the covariance matrix between excess returns and the excess market returns $\Sigma_{r,M}$. Below I specify the estimators for Σ_r and $\Sigma_{r,M}$ for each model, after describing the model itself.

(i) The Zero-Factor Model (BM0)

The zero-factor model can be written as the following:

$$\begin{pmatrix} r_t \\ f_{1,t} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu_r \\ \mu_{f,1} \end{pmatrix}, \begin{pmatrix} \Sigma_r & \Sigma_{r,M}' \\ \Sigma_{r,M} & \Sigma_M \end{pmatrix} \right) \tag{2}$$

where $f_{1,t}$ is the excess market return, i.e. the market return in excess of the riskfree rate. No factor structure has been specified above, thus the name "zero-factor" model. $f_{1,t}$ is not a factor as it does not impose any structure on Σ_r. $f_{1,t}$ is included above only for the purpose of specifying the covariance matrix $\Sigma_{r,M}$. As the IID property is assumed, the sample variance-covariance matrix $\hat{\Sigma}_r^0$

is the consistent estimator of Σ_r, and the sample covariance matrix $\hat{\Sigma}^0_{r,M}$ is the consistent estimator of $\Sigma_{r,M}$.[10] (The consequence of the singularity of $\hat{\Sigma}^0_r$ will be discussed later.)

Note that all the other models described below can be considered special cases of the zero-factor model, given the IID property of those models. Thus, $\hat{\Sigma}^0_r$ is a consistent estimator of Σ_r for those models as well, though it may not be the most efficient estimator.

(ii) The Beta Model (BM1 and BM3)

A beta model can be written as the following:

$$r_t = \alpha + \beta f_t + \varepsilon_t, \qquad f_t \sim N(\mu_f, \Sigma_f), \qquad \varepsilon_t \sim N(0, D_\varepsilon) \tag{3}$$

In the above equation, r_t is an N-by-1 vector of month-t excess stock returns, N being the number of stocks, α is an N-by-1 vector of intercepts, β is an N-by-K matrix of betas, K being the number of factors, f_t is a K-by-1 vector of month-t factor realizations, and ε_t is an N-by-1 vector of errors. I assume that D_ε is a diagonal matrix, i.e. there is no correlation between errors of different stocks. In the one-factor beta model, f_t includes only the excess market return. In the three-factor beta model, f_t includes HML and SMB in addition to the excess market return.

For simplicity, I adopt the standard regression model assumptions such as the independence between f_t and ε_t, and the IID of f_t and ε_t. Thus, α, β, and D_ε can be consistently estimated by the ordinary least square method, and μ_f and Σ_f can be consistently estimated by sample mean and sample variance-covariance matrix. Let us denote those estimates by $\hat{\alpha}$, $\hat{\beta}$, $\hat{D}_\varepsilon$, $\hat{\mu}_f$, and $\hat{\Sigma}_f$. Then the variance-covariance matrix of returns can be estimated as:

$$\hat{\Sigma}^{BM}_r = \hat{\beta}\hat{\Sigma}_f\hat{\beta}' + \hat{D}_\varepsilon \tag{4}$$

Note that the diagonal elements of $\hat{\Sigma}_r$ correspond to those of the sample variance-covariance matrix $\hat{\Sigma}^0_r$. Only off-diagonal elements of $\hat{\Sigma}^B$ are different.

Let us now consider the estimator for $\Sigma_{r,M}$. By partitioning vectors $\hat{\beta}' = [\hat{\beta}_1, \hat{\beta}_2{}']$ and $f_t = [f_{1,t}, f_{2,t}{}']$ and considering the regression of $f_{2,t}$ on $f_{1,t}$, i.e. $f_{2,t} = \hat{a} + \hat{b} f_{1,t} + \hat{e}$,

[10] I consider the consistency rather than the unbiasedness, as it is much simpler to determine the consistency in later analyses. Normality assumption is not needed in first analyses though I do use this assumption in a later Monte Carlo experiment.

$$
\begin{aligned}
r_t &= \hat{\alpha} + \hat{\beta}_1 f_{1,t} + \hat{\beta}_2 f_{2,t} + \hat{\varepsilon}_t \\
&= \hat{\alpha} + \hat{\beta}_2 a + (\hat{\beta}_1 + \hat{\beta}_2 \hat{b}) f_{1,t} + \hat{\beta}_2 \hat{e} + \hat{\varepsilon}_t
\end{aligned} \tag{5}
$$

The last line shows that, in the regression of r_t on $f_{1,t}$ and the constant, the slope coefficient would be $\hat{\beta}_1 + \hat{\beta}_2 \hat{b}$. Thus, the estimator for the covariance between r_t on $f_{1,t}$ is

$$
\hat{\Sigma}_{r,M}^{BM} = \hat{\Sigma}_{f,11} (\hat{\beta}_1 + \hat{\beta}_2 \hat{b})(\hat{\beta}_1 + \hat{\beta}_2 \hat{b})' \tag{6}
$$

Note that this estimator is in fact identical to the sample covariance matrix between the individual excess returns and the market excess return, i.e. $\hat{\Sigma}_{r,M}^{0}$.[11]

(iii) The Characteristics Model (CM2)

For the characteristics model, I consider a two-stage characteristics model: In the first stage, the excess stock return is related to the excess market return, and the residual return is identified. In the second stage, the residual return is explained by characteristics. An alternative formulation ("single-stage characteristics model") would have related the excess stock return directly to the characteristics. I choose the former mainly because of the need to generate the minimum tracking error portfolio in a later analysis.[12] Chan, Karceski, and Lakonishok (1999) emphasize that the tracking error minimization produces a more interesting model comparison. For the tracking error minimization, one needs to include the benchmark into the model in some way. It is rather difficult to incorporate the benchmark into a single-stage characteristics model, and most practitioners would adopt a two-stage characteristics model for the purpose of the tracking error minimization.[13]

The first stage of the characteristics model can be written as:

$$
r_t = \alpha_M + \beta_M f_{1,t} + \tilde{r}_t \tag{7}
$$

r_t is the vector of excess stock returns, $f_{1,t}$ is the excess market return (scalar), and $\tilde{r}_t$ is the vector of residual returns. I denote the excess market return as $f_{1,t}$ to emphasize the fact that it is identical to the first element of f_t in the beta model. α_M and β_M are the constant and the slope coefficients,

[11] In the regression of y on x and the constant, i.e. $y = \hat{a} + \hat{b}x + \hat{e}$, $\hat{b}^2 \sigma_x^2$ is identical to the sample covariance between x and y.

[12] In an earlier stage of the research, I examined a longer list of models including a single-stage characteristics model. The performance of the single-stage characteristics model is comparable to that of the two-stage characteristics model.

[13] Rosenberg's (1974) article, on which BARRA model is based, suggests that the characteristics model be applied to the residual from the market return regression. Northfield makes this explicit.

respectively. The second stage of the characteristics model can be written as:

$$\tilde{r}_t = \iota\delta_{1,t} + x\delta_{2,t} + \eta_t = X\Delta_t + \eta_t, \qquad \Delta_t \sim \mathrm{N}(0, \Sigma_\Delta), \qquad \eta_t \sim \mathrm{N}(0, D_\eta) \tag{8}$$

In the above equation, ι is an N-by-1 vector of ones, $\delta_{1,t}$ is the intercept term (scalar), x is an N-by-K matrix of characteristics, $\delta_{2,t}$ is a K-by-1 vector of slope coefficients, and η_t is an N-by-1 vector of error terms. $\Delta_t{}' = (\delta_{1,t}, \delta_{2,t}{}')$ is called the characteristics price. Note that this equation implies cross-sectional regressions. By estimating this equation for different t's, we get time-series of estimates. The time-series mean of Δ_t is zero since $\tilde{r}_t$ is the error term in the time-series regression and has zero time-series means. As before, I assume that D_η is diagonal. The characteristics model includes two characteristics (K=2): BM (the book-to-market ratio) and ME (the market capitalization).

Once again, I adopt the assumptions that allow the OLS estimation of the parameters. This means independence between $f_{1,t}$ and $\tilde{r}_t$ in the first-stage equation; also the independence between x and η_t in the second-stage equation. Note that the first-stage equation is to be estimated from time-series regressions, whereas the second stage equation is to be estimated from cross-sectional regressions, as mentioned earlier. μ_Δ, Σ_Δ and Σ_η can be consistently estimated by the sample counterparts from the estimated Δ_t, and the second-stage residuals η_t. (This may not be efficient, but good enough for our purpose.) Let us denote the resulting estimates by $\hat{\Delta}_t$, $\hat{\mu}_\Delta$, $\hat{\Sigma}_\Delta$, and $\hat{\Sigma}_\eta$. Then the variance-covariance matrix of returns can be estimated as:

$$\hat{\Sigma}_r^{CM} = \hat{\Sigma}_{f,11}\hat{\beta}_M\hat{\beta}_M' + X\hat{\Sigma}_\Delta X' + \hat{\Sigma}_\eta \tag{9}$$

The estimator for $\Sigma_{r,M}$ can be obtained from the first-stage regression:

$$\hat{\Sigma}_{r,M}^{CM} = \hat{\Sigma}_{f,11}\hat{\beta}_M \tag{10}$$

Note that this estimator is identical to the sample covariance matrix, i.e. $\hat{\Sigma}_{r,M}^0$.[14] That is, there is no difference among the models in how $\Sigma_{r,M}$ is to be estimated.

Note that characteristics vector x is held constant for the whole estimation period. This is certainly not the conventional way of estimating a characteristics model. This study adopts this assumption to make the characteristics model more comparable to the beta model. In the beta model, the beta is held constant for the whole estimation period. Allowing the characteristics to vary over time might give an undue advantage to the characteristics model. So I fix the value of the characteristics vector and do not update it. It is fixed at the beginning-of-the-estimation-period value; choosing a later value would

[14] See footnote 11.

amount to explaining the return with future characteristics.

4. Data

Data of S&P 500 stocks, for the period from 1976 to 2010, are used in this study. Portfolios are created at the end of year *t*, where *t* is one of 1980, 1985, 1990, 1995, 2000, and 2005. For each year *t*, I identify the stocks belonging to S&P 500 as of the end of year *t*. Then for each stock, I collect monthly returns for the 10-year period around year *t*, i.e. from year *t-4* to year *t+5*. I also collect characteristics as of the beginning of year *t-4*. As mentioned earlier, only the beginning of the period characteristics are used in the estimation. Two characteristics variables are included: book-to-market ratio (BM) and the market capitalization (ME). The numerator of BM is the book value from the last fiscal quarter ending prior to the beginning of year t-4. I also collect Fama-French factor returns for the entire period.[15]

Table 1 shows summary statistics for each portfolio creation year (i.e. for each 'cohort' of stocks). For some periods, the mean of individual stock returns is somewhat higher than the mean of MRF plus the riskfree rate. This corresponds to the size effect as shown by the value of SMB.[16] The size effects vary quite a bit across the cohorts, whereas the value effect is rather stable. Relatively large values of ME and BM for the 2005 cohort reflect the high valuation at the end of year 2000.

[Table 1 about here.]

5. Accuracy of Variance and Covariance Forecasts

In this section, I examine the accuracy of variance and covariance forecasts of the alternative models. Following the convention, forecast accuracy is measured in terms of the root mean squared errors between forecasts and realized values. I also consider correlations between forecasts and realized

[15] Downloaded from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[16] For a single period, the mean of individual stock returns is an equal-weighted average, whereas the market return is a value-weighted average. Thus, the small firm effect makes the mean of individual stock returns higher than the market return.

values as additional measures of accuracy.

Note that, given the IID property, predicting the future is equivalent to estimating the population parameters. Thus, the *estimator* of the variance-covariance matrix $\hat{\Sigma}_r$ would be the best *predictor* of the future variance-covariance matrix. So, the elements of $\hat{\Sigma}_r$ are compared to the realized values, i.e. the sample variances and covariances of the future months. $\hat{\Sigma}_r$ is obtained out of the data between year *t-4* and year *t*, where *t* is the year of portfolio creation. The sample variances and covariances are based on the data between year *t+1* and year *t+5*.

Table 2 shows the root mean squared error and two correlation measures (Pearson's and Kendall's) for variance forecasts.[17] Table 3 shows the same statistics for covariance forecasts. Statistics are presented for each cohort. Bootstrap standard errors are presented as well.[18] Note that the two beta models and the zero-factor model produce the identical variance forecasts (i.e. the diagonal elements of $\hat{\Sigma}_r$ are identical.) Thus, in Table 2, only one set of statistics is reported for these models.

[Table 2 about here.]

[Table 3 about here.]

Forecast accuracy varies from cohort to cohort. In case of variance forecasts, the root mean squared error is less than 50% for the first three cohorts, but is more than 100% for the last two cohorts.[19] The correlation between the forecasts and the realized values is particularly low for the 2005 cohort, possibly reflecting the impact of the global financial crisis. A similar pattern is noticeable in the case of covariance forecasts. Forecasts are more accurate for earlier cohorts, and less accurate for later cohorts. As noted by Chan, Karceski, and Lakonishok (1999), variance forecasts are somewhat more accurate than covariance forecasts. The correlation between the forecasts and the realized values is

---

[17] Kendall's correlation measures the rank correlation. It is the answer to the following question: Given a pair of observations, what is the probability that the ordering of this pair by the first variable is identical to the ordering of the pair by the second variable?

[18] Alternatively, standard errors can be calculated by assuming a bivariate normal distribution. However, bootstrap standard errors are known to be more robust.

[19] Note that the root mean squared error is in the unit of monthly variance. Thus, the root mean squared error of 50% is, without the percentage, 0.005. This is comparable to the mean absolute deviation reported by Chan, Karceski, and Lakonishok (1999).

mostly above 0.5 in the case of variance forecasts, but always below 0.5 in the case of covariance forecasts.

From Table 2, it is clear that the beta model and the characteristics model produce comparable variance forecasts. There is no discernable difference between the forecasts of the two models. In the case of covariance forecasts, the difference is somewhat more noticeable. For some cohorts (e.g. cohort 1990), the beta model appears to produce better forecasts, but for others (e.g. cohort 1985 and cohort 2000), the characteristics model appears to produce better forecasts. While the analysis of this section is indicative, it would be more meaningful to consider the economic consequences. More rigorous statistical analysis is deferred to the next section where economic consequences are considered.

6. Economic Consequences

In this section, I examine the economic significance of the difference among the models. I construct the global minimum variance portfolios using the alternative models, and compare their "out-of-sample" realized variances. I also construct the global minimum tracking error portfolios, and compare their out-of-sample realized tracking errors. There is a possibility that realized variances or tracking errors are similar even though the parameter estimates are different. In such a case, one could conclude that the difference among the models is not significant economically.

The word 'global' in the name indicates that it has the smallest variance (tracking error) among all the minimum variance (tracking error) portfolios. The use of the global minimum variance (tracking error) portfolios instead of some other mean-variance (mean-tracking error) efficient portfolios can be justified by two reasons. First, estimating and predicting mean returns are quite difficult; whether the models considered here are useful for such purposes is questionable. Second, I would like to focus on the risk control aspect of the models. In this respect, I am following the study design of Chan, Karceski, and Lakonishok (1999). As I do not consider any other minimum variance (tracking error) portfolios in this paper, I drop the word 'global' when there is no possibility of confusion.

Given an estimator $\hat{\Sigma}_r$, the minimum variance portfolio is the solution to the following problem:

$$\min w'\hat{\Sigma}_r w \quad \text{s.t.} \quad w'\iota = 1 \tag{11}$$

Solving the problem, the following weight vector is obtained:

$$w^{MV} = \hat{\Sigma}_r^{-1}\iota/(\iota'\hat{\Sigma}_r^{-1}\iota) \tag{12}$$

Tracking error is the difference between the portfolio return and the benchmark. When the market is the benchmark, the variance of the tracking error can be estimated as $\widehat{\text{var}}(r_p - r_M) = w'\hat{\Sigma}_r w - 2w'\hat{\Sigma}_{r,M} + \hat{\sigma}_M^2$, where $\hat{\Sigma}_{r,M}$ is the covariance between the return vector and the market return. The last term does not depend on the portfolio. Thus, the portfolio that minimizes the variance of the tracking error can be found by solving:

$$\min w'\hat{\Sigma}_r w - 2w'\hat{\Sigma}_{r,M} \quad \text{s.t.} \quad w'\iota = 1 \tag{13}$$

This portfolio might be called the minimum variance-of-tracking-error portfolio. Following the convention among practitioners, however, I call this the minimum tracking error portfolio. The solution is the following:

$$w^{\text{TE}} = \left(1 - \iota'\hat{\Sigma}_r\hat{\Sigma}_{r,M}\right)\frac{\hat{\Sigma}_r^{-1}\iota}{\iota'\hat{\Sigma}_r^{-1}\iota} + \hat{\Sigma}_r\hat{\Sigma}_{r,M} \tag{14}$$

It can be interpreted as the weighted average of the minimum variance portfolio (w^{MV}) and the hedging portfolio ($\hat{\Sigma}_r\hat{\Sigma}_{r,M}/\iota'\hat{\Sigma}_r\hat{\Sigma}_{r,M}$). In the actual implementation, I impose the no-short sale constraint and also the maximum weight of 0.05 for both the minimum variance problem and the minimum tracking error problem. This is in accordance with the common industry practice.[20] In any case, the maximum weight is binding only infrequently.[21]

I do not create portfolios for the zero-factor model (BM0). For the zero-factor model, $\hat{\Sigma}_r$ is singular (since N is greater than T in our estimation.) When $\hat{\Sigma}_r$ is singular, the above minimization problems do not have a solution. There are three responses to the singularity problem. One response is to replace $\hat{\Sigma}_r^{-1}$ in the solution with the Moore-Penrose generalized inverse $\hat{\Sigma}_r^{+}$, as Ledoit and Wolf (2003) did[22]. This approach does not achieve the global minimum. The second response is to solve the

---

[20] Allowing the short-sale makes the return calculation somewhat challenging, as there is a possibility that the return is less than -100%.

[21] Chan, Karceski, and Lakonishok (1999) impose the maximum weight of 0.02, which I believe is unnecessarily stringent. It is also binding more frequently.

[22] When $\hat{\Sigma}_r = R'R/(N-1)$, $\hat{\Sigma}_r^{+1} = (N-1)R'(RR')^{-1}(RR')^{-1}R$. See Theorem 1.3.2 of Campbell and Meyer (1979). The resulting solution can be interpreted as the minimal least square solution of the first order condition, i.e. it has the smallest vector-norm $\|\;\;\|$ among all the least square solutions, where a least square solution minimizes $\|\hat{\Sigma}_r w - \lambda\iota\|$ subject to $\iota' w = 1$. See Theorem 2.1.1 of Campbell and Meyer (1979). A related approach, based on a Krylov method, is discussed by Bajeux-Besnainou, Bandara, and Bura (2012). Ipsen and Meyer (1998) and Calvetti, Lewis, and Reichel (2000) show that Krylov methods lead to generalized-

problem after imposing some boundary conditions. This is the approach of Chan, Karceski, and Lakonishok (1999). In this approach, the solution is obtained at the boundary. The third response is not to proceed further, which is what I do in this paper.

For each portfolio, I calculate the 'out-of-sample' monthly portfolio returns, from which realized variances and tracking errors are determined. A realized tracking error is measured as the variance of the portfolio return in excess of the market return. In the current set-up, there is no reason to expect that the portfolio has a high Sharpe ratio, as no mean estimates have been used. Nonetheless, one might be curious whether the portfolio with a higher volatility has a compensating return. So I calculate Sharpe ratios as well. To assess the significance of the difference in the realized variances and tracking errors, I carry out the Pitman-Morgan test.[23] For the Sharpe ratios, I carry out Jobson-Korkie (1981) test as corrected by Memmel (2003).[24]

Table 4 reports the performances of the minimum variance portfolios. The most interesting question is whether the realized variances differ substantially across models. Statistically, the answer to this question is 'sometimes.' In 8 out of 12 t tests, the differences in the realized variances are statistically significant. However, many of these test results are contradictory. The BM1 portfolio has a significantly lower variance than the CM2 portfolio for the 2005 cohort, but the opposite is true for the 1990, 1995, and 2000 cohorts. The BM3 portfolio has a significantly lower variance than the CM2 portfolio for the 1990 cohort, but the opposite is true for the 1985, 2000, and 2005 cohorts. More importantly, the difference in realized variances is not substantial in the economic sense. In the case of the 2005 cohort, the variances of the BM3 and CM2 portfolios (10.87% vs. 11.23% per month) are translated to the annualized standard deviations of 11.40% and 10.77%.[25] Thus, the difference is less than 1% of the annualized standard deviation. The difference is greater for the 2000 cohort, but it is still less than 2% of the annualized standard deviation (11.22% vs. 9.46%). Table 4 also reports the predicted variances and the realized Sharpe ratios of the minimum variance portfolios. The predicted

---

inverse-based solutions.

[23] Pitman (1939) and Morgan (1939) proposed to test the equality of the variances of two variables, say, $\sigma_X = \sigma_Y$ by testing the significance of the correlation between the sum and the difference, i.e. $\rho_{X+Y,X-Y} = 0$. A significant $\rho_{X+Y,X-Y}$ indicates a significant difference between σ_X and σ_X.

[24] Under the null hypothesis that two Sharpe ratios are the same, i.e. $\mu_X/\sigma_X = \mu_Y/\sigma_Y$, the following z-statistic has the standard normal distribution asymptotically: $z = (\hat{\mu}_x/\hat{\sigma}_X - \hat{\mu}_Y/\hat{\sigma}_Y)/\sqrt{V}$, where $V = (1/T)[2 - 2\hat{\rho} + (1/2)(\hat{\mu}_X^2/\hat{\sigma}_X^2 + \hat{\mu}_Y^2/\hat{\sigma}_Y^2 - 2\hat{\rho}^2\hat{\mu}_X\hat{\mu}_Y/(\hat{\sigma}_X\hat{\sigma}_Y))]$.

[25] $\text{Annualized standard deviation} = \sqrt{12}\sqrt{\text{monthly variance}}$

variances are extremely poor predictors of the realized variances. Thus, comparing these predictors is not meaningful. Sharpe ratios are comparable among models, and the differences are not significant.

[Table 4 about here.]

[Table 5 about here.]

Table 5 reports the performance of the minimum tracking error portfolios. Compared to Table 4, the differences among the models are even smaller. The tracking error of the BM1 portfolio is not significantly different from that of the CM2 portfolio. In the case of the BM3 portfolio vs. the CM2 portfolio, the difference is significant in three instances. In one of three instances, the CM2 portfolio has a lower tracking error, whereas, in the rest, the BM3 portfolio has a lower tracking error. The conclusion that I made from Table 4 is still valid: the difference between the models is not economically significant. The predicted tracking error is again a poor predictor of the realized tracking error, and the difference in Sharpe ratios is also minor.

7. Sources of Similarity

The analysis so far suggests that the beta model and the characteristics model are similar to each other in terms of the ability to control portfolio risk. Given the finding, a question arises naturally: what are the sources of this similarity? In this section, I trace the sources of the similarity to the correlation between beta and characteristics, and also to the correlation between factor and the characteristics price. I estimate these correlations from the data as well.

The relevance of the correlation between beta and characteristics has been noted by several authors, including Daniel and Titman (1997) and Davis, Fama, and French (2000). They have pointed out that the high correlation between beta and characteristics makes it difficult to determine which one is driving the return process. My analysis clarifies and extends this idea. My analysis shows that the correlation between beta and characteristics matters, but only if the true model is the beta model. If the true model is the characteristics model, it is the correlation between factor and the characteristics price that is relevant.

In the following, I take the approach of assuming a particular data generating process (DGP) and analyzing the consequence of choosing a wrong model. First, I assume that the BM3 model is the true DGP, and examine whether the CM2 model leads to an accurate estimate. Then I assume that the CM2 model is the true DGP, and examine whether the BM3 model leads to an accurate estimate. I do not consider the BM1 model any more, as this model does not add any insights to the analysis. Both the minimum variance portfolio and the minimum tracking error portfolio depend on the estimator of Σ_{r}. So below I ask when the estimator of Σ_{r} based on a wrong model is a good estimator of Σ_{r}. As all the models produce the identical estimators for $\Sigma_{\mathrm{r,M}}$, I do not consider those estimators.

First, suppose that the BM3 model is the true DGP. Then the true variance-covariance matrix of returns is

$$\Sigma_r = \beta \Sigma_f \beta' + D_\varepsilon \tag{15}$$

The estimator based on the CM2 model is

$$\hat{\Sigma}_r^{CM} = \hat{\beta}_M \Sigma_{f,11} \hat{\beta}_M' + X \hat{\Sigma}_\Delta X' + \hat{D}_\eta \tag{16}$$

The question for the moment is: under what circumstance is $\hat{\Sigma}_r^{CM}$ a good estimator of Σ_r?

In the appendix, I show that the diagonal part of $\hat{\Sigma}_r^{CM} - \Sigma_r$ converges in probability to

$$D\{2\psi\Sigma_e\psi' - \psi\Sigma_e{\beta_2}' - \beta_2\Sigma_e\psi' + 2\xi D_\varepsilon \xi' - \xi D_\varepsilon - D_\varepsilon \xi'\} \tag{17}$$

whereas the off-diagonal part of $\hat{\Sigma}_r^{CM} - \Sigma_r$ converges in probability to

$$O\{\psi\Sigma_e\psi' - \psi\Sigma_e{\beta_2}' - \beta_2\Sigma_e\psi' + \xi D_\varepsilon \xi' - \xi D_\varepsilon - D_\varepsilon \xi'\} \tag{18}$$

Above, $D\{\ \}$ indicates the diagonal part, and $O\{\ \}$ indicates the off-diagonal part. Σ_e is the variance-covariance matrix of the error terms in the regression of $f_{2,t}$ on $f_{1,t}$, i.e. $f_{2,t} = a + bf_{1,t} + e$. That is, it is the variation in $f_{2,t}$ that is not related to the variation in $f_{1,t}$. What is more important are the terms ψ and ξ. ψ captures the lack of the correlation between beta β_2 and characteristics X. ξ is also related to the correlation between beta and characteristics, but in an indirect manner.

ξ and ψ have regression-based interpretations as well. Each column of ξ can be considered as the error in the regression of each column of the identity matrix on the columns of X.[26] Similarly, each column of ψ can be considered the error in the regression of each column of β_2 on the columns of X, i.e. $\beta_{2,i} = \pi_1 + \pi_2 x_i + \psi_i$.[27] Thus, ψ would be small when β_2 is well 'explained' by the columns of X.

[26] In matrix notation, $\xi = X(X'X)^{-1}X'$.

[27] In matrix notation, $\psi = \beta_2 - X(X'X)^{-1}X'\beta_2$.

Beta and characteristics may be perfectly correlated, i.e. the variation in β_2 can be completely 'explained' by the variation in *X*, in which case each element of ψ is zero. ξ cannot be zero, however. Given that there are *N* columns in the identity matrix and only two columns in *X*, it is not possible to 'explain' all the columns of the identity matrix with the columns of *X* (when *N* is greater than 2.)

If both ψ and ξ were zero, then each element of $\hat{\Sigma}_r^{CM} - \Sigma_r$ would converge to zero, i.e. the CM2 model would produce a consistent estimator of the variance-covariance matrix of returns, and there would be no difference in the risk of the portfolios based on the CM2 model and the BM3 model asymptotically. (In reality, ξ cannot be, and thus, $\hat{\Sigma}_r^{CM} - \Sigma_r$ cannot converge to zero.) Even if ψ and ξ are nonzero, as long as they are small, the difference between the CM2 model and the BM3 model becomes small as well. Summarizing this discussion, the following statement can be made:

> Statement 1: *When the beta model is the true data generating process, the characteristics model produces a more accurate estimate if the correlation between beta and characteristics is higher.*

I now assume that the CM2 model is the true DGP. Then the true variance-covariance matrix of returns is

$$\Sigma_r = \beta_M \Sigma_{f,11} \beta_M' + X \Sigma_\Delta X' + D_\eta \tag{19}$$

The estimator based on the BM3 model is

$$\hat{\Sigma}_r^{BM} = \hat{\beta} \hat{\Sigma}_f \hat{\beta}' + \hat{D}_\varepsilon \tag{20}$$

This time I ask the question of when $\hat{\Sigma}_r^{BM}$ is a good estimator of Σ_r.

In the appendix, I show that the diagonal elements of $\hat{\Sigma}_\mathrm{r}^\mathrm{BM} - \Sigma_\mathrm{r}$ converge in probability to zero, and that the off-diagonal part of $\hat{\Sigma}_\mathrm{r}^\mathrm{BM} - \Sigma_\mathrm{r}$ converges in probability to $X\Sigma_\nu X'$. Σ_ν is the variance-covariance matrix of the error terms in the regression of Δ_t on $f_{2,t}$, i.e. $\Delta_t = \phi_1 + \phi_2 f_{2,t} + \nu_t$. Thus, Σ_ν indicates the lack of correlation between the characteristics price Δ_t and factor realization $f_{2,t}$. Given that there are 3 elements in Δ_t and two elements in $f_{2,t}$, Σ_ν cannot be zero. However, it can be quite close to zero.

If Σ_ν were zero, $\hat{\Sigma}_\mathrm{r}^\mathrm{BM} - \Sigma_\mathrm{r}$ would converge in probability to zero, and $\hat{\Sigma}_\mathrm{r}^\mathrm{BM}$ would be a consistent estimator. This is not possible as Σ_ν cannot be zero. However, one can still say that $\hat{\Sigma}_\mathrm{r}^\mathrm{BM}$ is a good estimator when Σ_ν is small. Thus, the following statement can be made.

Statement 2: *When the characteristics model is the true data generating process, the beta model produces a more accurate estimate if the correlation between the characteristics price and factor realization is higher.*

I estimate ψ, ξ, and Σ_ν from the 2005 cohort data. Table 7 reports the results of the regression associated with ψ, i.e. $\beta_{2,i} = \pi_1 + \pi_2 x_i + \psi_i$. Both elements of x_i are highly significant, though R squared of the regression is only around 10%. Whether this level of correlation is meaningful or not will be discussed in the next section in a Monte Carlo simulation context. Table 8 reports the summary statistics for the diagonal and the off-diagonal elements of ξ. ξ has large diagonal elements and small off-diagonal elements, i.e. it is similar to the identity matrix. Note that when ξ is close to the identity matrix, the variance estimates are very accurate, i.e. all the terms involving ξ in (17) is zero. Table 9 reports the results of the regression associated with Σ_ν, i.e. $\Delta_t = \phi_1 + \phi_2 f_{2,t} + \nu_t$. All the elements in the first row of ϕ_2 (the coefficients of SMB) are significant, whereas the two of the three elements in the second row (the coefficients of HML) are significant. The R squared is between 14% and 63%, making Σ_ν significantly smaller than the variance-covariance matrix of Δ_t.[28]

[Table 7 about here.]

[Table 8 about here.]

[Table 9 about here.]

8. A Sensitivity Analysis

In the previous section, I identified the two sources of the similarity between the beta model and the characteristics model: the correlation between beta and characteristics, as reflected in ψ, and the correlation between factor and the characteristics price, as reflected in Σ_ν. I presented the estimates of ψ and Σ_ν as well. For an investor who is facing a choice between the beta model and the characteristics model, one remaining issue would be: What if the true values of ψ and Σ_ν are quite different from the estimates from the previous section? Would the loss from choosing a wrong

[28] Note that the ratio of the variance of ν_t to the variance of Δ_t equals 1 minus R squared.

model—the increase in variances and tracking errors—be significantly larger? In this section, I examine the effect of the variation in ψ and Σ_ν on the loss.

To examine the effect of the variation in ψ and Σ_ν, I carry out a Monte Carlo simulation. I choose the extreme values of ψ and Σ_ν, generate pseudo data sets, and evaluate the performance of the alternative models. To keep the analysis manageable, the other parameters of the data generating process are not varied.

The analysis of this section can be compared to the study by Pastor and Stambaugh (2000). Adopting a Bayesian framework, Pastor and Stambaugh (2000) calculate the posterior distribution for each model, and evaluate the reduction in the expected utility from making a choice based on a wrong model. Let $P_1, \cdots, P_n$ be posteriors, corresponding to n models. Let $a_1, \cdots, a_n$ be the choices (or actions), each of which is optimal under each of the alternative models. Given n posteriors and n choices, there is an n-by-n matrix of consequences, represented by $\{C_{ij}\}$, where i is an index of posteriors and j is an index of choices. Pastor and Stambaugh (2000) focus on how the utility from each choice varies across posteriors, i.e., how large $\max C_{\cdot j} - \min C_{\cdot j}$ is for each j.[29]

I adopt the perspective of Pastor and Stambaugh (2000), but implement the analysis in a non-Bayesian framework. That is, instead of calculating 'posterior distributions,' I consider alternative 'data generating processes.' I estimate alternative DGPs corresponding to alternative models, and examine the consequences of alternative choices under these DGPs. I prefer to work with DGPs rather than with posteriors, primarily because the former is simpler than the latter. To calculate posteriors, one needs to come up with priors, which is not an easy task. To avoid the difficulty of specifying priors, one might work with non-informative priors. However, specifying non-informative priors is not a trivial task either. With DGPs, there is no need to specify priors and subsequent analysis is simpler. However, working with DGPs rather than with posteriors has one obvious drawback: Simulating out of a DGP tends to underestimate the true magnitude of uncertainty. One could limit this problem by allowing variations in the parameter values of DGPs, i.e. by carrying out multiple sets of simulations.

---

[29] In a different application, Chamberlain (2000) suggested to apply the minimax principle, i.e. the utility from choosing a_j is determined by $\min_i C_{ij}$. The analysis by McCulloch and Rossi (1990) can be understood in the same framework. They examined the consequence of each choice under the posterior that is most favorable to the choice, i.e. $C_{11}, \cdots, C_{nn}$. Kendall and Stambaugh's (1996) analysis is different from other analysis mentioned here, in that they evaluate alternative models under single posterior.

I elaborate on this point below.

In Pastor and Stambaugh (2000), given posterior P_{i}, the predictive distribution of returns is determined, and the distribution of the consequence of choosing action a_{j} is determined. In my approach, a DGP includes the distribution of future returns. Given an estimate of the parameters of the DGP, say, θ_{i}, the distribution of the consequence of choosing action a_{j} is determined. The predictive distribution in Pastor and Stambaugh (2000) and the DGP in my analysis are similar to each other in that they summarize what we know about the future.

The main difference between the predictive distribution and the DGP is whether the parameter uncertainty is allowed. A predictive distribution allows the parameter uncertainty through the posterior distribution, which is a distribution of θ. A DGP does not allow the parameter uncertainty as it is based on a point estimate of θ. Thus, simulation analysis based on a DGP may underestimate the true uncertainty regarding the future. To compensate for this tendency of understating the magnitude of uncertainty, I consider multiple values of θ for each model. That is, I vary the values of θ from the point estimates that were determined from the data. While a single DGP underestimates the parameter uncertainty, multiple DGPs may provide more accurate view of reality.

Given the high dimension of θ, it is not practical to vary every element of θ. I focus on those elements of θ that critically affects the consequence of choices, as identified in the previous section. I consider two sets of DGPs, one corresponding to the beta model and the other corresponding to the characteristics model. In the first set, I vary the correlation between β_2 and X, and in the second set, I vary the correlation between $f_{2,t}$ and Δ_t. The analysis in Section 7 shows that, if the beta model is the true DGP, the relative performances of the two models critically depend on the correlation between β_2 and X. If the characteristics model is the true DGP, the relative performances of the two models depend on the correlation between $f_{2,t}$ and Δ_t.

When the beta model is true, the DGP can be written as:

$$r_t = \alpha + \beta f_t + \varepsilon_t, \qquad f_t \sim N\left(\mu_f, \Sigma_f\right), \qquad \varepsilon_t \sim N(0, D_\varepsilon) \tag{21}$$

Let us denote the i-th row of β as $\beta_{\mathrm{i}}' = \left(\beta_{1,i}, \beta_{2,i}{}'\right)$. $\beta_{2,i}$ is generated from the following:

$$\beta_{2,i} = \pi_1 + \pi_2 x_i + \psi_i, \quad x_i \sim N(\mu_x, \Sigma_x), \quad \psi_i \sim N\left(0, \Sigma_\psi\right) \tag{22}$$

I do not model $\beta_{1,i}$, as it does not affect our analysis. The same is true for α. Thus, the entire DGP is described by the following parameters: $\left(\alpha, \beta_1, \mu_f, \Sigma_f, D_\varepsilon, \pi_1, \pi_2, \mu_x, \Sigma_x, \Sigma_\psi\right)$. t runs from 1 to T+H

(T=H=60), and i runs from 1 to N (N=467)[30]. f_t, ε_t, x_i and ψ_i are assumed to be pairwise independent.

I leave the values of $\alpha, \beta_1, \mu_f, \Sigma_f, D_\varepsilon, \mu_x$ and Σ_x at their estimates, and modify the values of π_1, π_2, and Σ_ψ. Three possible values of π and Σ_ψ are selected: (BM3a) as estimated from the data; (BM3b) $\beta_{2,i}$ is maximally correlated with x_i, i.e. Σ_ψ is zero; (BM3c) $\beta_{2,i}$ is minimally correlated with x_i, i.e. π_2 is zero. In all cases, I maintain the unconditional mean and variance of $\beta_{2,i}$. That is, I keep $\mu_{\beta,2} = \pi_1 + \pi_2 \hat{\mu}_x$ and $\Sigma_{\beta,22} = \pi_2 \hat{\Sigma}_x \pi_2' + \Sigma_\psi$ at their estimates $\hat{\mu}_{\beta,2}$ and $\hat{\Sigma}_{\beta,22}$. When $\beta_{2,i}$ is maximally correlated with x_i, we have $\pi_2 \Sigma_x \pi_2' = \hat{\Sigma}_{\beta,22}$; thus, one possibility is to choose $\pi_2 = \hat{\Sigma}_{\beta,22}^{1/2} \left(\hat{\Sigma}_x^{1/2}\right)^{-1}$ where $\hat{\Sigma}_{\beta,22}^{1/2}$ and $\hat{\Sigma}_x^{1/2}$ are the lower triangular matrices obtained from the Cholesky decomposition.[31] When $\beta_{2,i}$ is minimally correlated with x_i, we have $\Sigma_\psi = \hat{\Sigma}_{\beta,22}$.

When the characteristics model is true, the DGP can be written as:

$$r_t = \alpha_M + \beta_M f_{1,t} + \tilde{r}_t \tag{23}$$

where the residual return is determined from the characteristics equation:

$$\tilde{r}_t = X\Delta_t + \eta_t, \quad x_i \sim N(\mu_x, \Sigma_x), \quad \eta_t \sim N\left(0, D_\eta\right) \tag{24}$$

I assume that Δ_t is generated from the following:

$$\Delta_t = \phi_1 + \phi_2 f_{2,t} + \nu_t, \quad f_t \sim N\left(\mu_f, \Sigma_f\right), \quad \nu_t \sim N(0, \Sigma_\nu) \tag{25}$$

The entire DGP is represented by the following parameters: $\left(\alpha_M, \beta_M, D_\eta, \phi_1, \phi_2, \Sigma_\nu, \mu_f, \Sigma_f, \mu_x, \Sigma_x\right)$. t runs from 1 to T+H (T=H=60), and i runs from 1 to N (N=467). η_t, ν_t, f_t, x_i are assumed to be pairwise independent.

I leave the values of $\alpha_M, \beta_M, D_\eta, \mu_f, \Sigma_f, \mu_x$ and Σ_x at their estimates, and modify the values of ϕ_1, ϕ_2, and Σ_ν. Three possible values of ϕ_1, ϕ_2, and Σ_ν are selected: (CM2a) as estimated from the data; (CM2b) Δ_t is maximally correlated with $f_{2,t}$, i.e. Σ_ν is as close to zero as possible; (CM2c) Δ_t is minimally correlated with $f_{2,t}$, i.e. ϕ_2 is zero. In all cases, I maintain the unconditional mean and variance of Δ_t. Thus, $\phi_1 + \phi_2 \hat{\mu}_f = \hat{\mu}_\Delta$ and $\phi_2 \hat{\Sigma}_{f,22} \phi_2{}' + \Sigma_\nu = \hat{\Sigma}_\Delta$. Note that it is not possible to have $\Sigma_\nu = 0$. $\phi_2 \hat{\Sigma}_{f,22} \phi_2{}'$ has the rank of 2, whereas $\hat{\Sigma}_\Delta$ has the rank of 3. Thus, these two cannot be identical. Instead, I look for the value of ϕ_2 that minimizes the sum of the squared elements of

[30] 467 is the number of stocks in cohort 2005, after excluding stocks with missing values.

[31] Thus, $\hat{\Sigma}_{\beta,22} = \hat{\Sigma}_{\beta,22}^{1/2} \hat{\Sigma}_{\beta,22}^{1/2}{}'$ and $\hat{\Sigma}_X = \hat{\Sigma}_x^{1/2} \hat{\Sigma}_x^{1/2}{}'$.

Σ_ν. That is, Δ_t is maximally correlated with $f_{2,t}$, given the restriction, when ϕ_2 is the solution to the minimization problem: $\min\left\|\hat{\Sigma}_\Delta - \phi_2\hat{\Sigma}_{f,22}\phi_2{}'\right\|$. Here, $\|\quad\|$ is Frobenius norm, i.e. the sum of the square of each element. This problem has a standard solution, which is based on the singular value decomposition of $\hat{\Sigma}_\Delta$ (Eckart and Young, 1936.) Given that $\hat{\Sigma}_\Delta$ is a symmetric nonsingular matrix, one may state the solution in terms of the eigenvalue decomposition.[32] Let the eigenvalue decomposition of $\hat{\Sigma}_\Delta$ be PDP'. I eliminate the smallest eigenvalue (in absolute term) and the eigenvector associated with it. Let us denote the result as $P^*D^*P^{*\prime}$, which has the rank of 2. Then ϕ_2 should satisfy $P^*D^*P^{*\prime} = \phi_2\hat{\Sigma}_{f,22}\phi_2{}'$. I select ϕ_2 as $P^*(D^*)^{\frac{1}{2}}\left(\hat{\Sigma}_{f,22}^{1/2}\right)^{-1}$, where $\hat{\Sigma}_{f,22}^{1/2}$ is the lower-triangular matrix obtained from the Cholesky decomposition. (The solution is not unique.)

In sum, I consider six DGPs: the baseline beta model (BM3a), the beta model with maximal correlation (BM3b), and the beta model with minimal correlation (BM3c); the baseline characteristics model (CM2a), the characteristics model with maximal correlation (CM2b), and the characteristics model with minimal correlation (CM2c). From each DGP, I simulate 500 samples, and compare the realized variances and tracking-errors of the beta model-based and the characteristics model-based portfolios.

Table 9 shows the performance of the minimum variance portfolios under alternative DGPs. Note that, compared to the real-data performance reported in Table 4, both predicted and realized variances are smaller, and the Sharpe ratios are higher. This is due to the under-estimation of the uncertainty as mentioned earlier. In any case, my primary interest is in how the numbers change across DGPs, rather than the absolute sizes of the numbers.

[Table 9 about here.]

[Table 10 about here.]

The panel (C) of the table shows the difference in the realized variances of the two models. All the differences are statistically significant. This is to be expected given the relatively large simulation size. When the baseline beta model (BM3a) is the DGP, the difference between the two models is 3.06%. When the correlation between beta and characteristics is maximal (BM3b), the difference becomes very small (0.3%) as expected. Also, when the correlation between beta and characteristics is zero

[32] For a nonsingular matrix, the singular value decomposition is identical to the eigenvalue decomposition.

(BM3c), the difference becomes larger (4.19%). This is also as expected. What is interesting is that the effect of making the correlation zero is not so large. There is no so much change from the level based on the current estimate. It suggests that the difference between the beta model and the characteristics model have a lot of room to decrease, but not so much room to increase.

When the characteristics model is the DGP (CM2a, CM2b, CM2c), the differences in the realized variances are smaller. Recall from the previous section that, when the characteristics model is the DGP, the variance estimators of the beta model are consistent. Thus, the difference between the models arises mostly from covariance estimates, and it is to be expected that the difference is not large. When the correlation between factor and the characteristics price is large (CM2b), the difference declines; when the correlation is zero (CM2c), the difference increases. This time, the real-data based estimates (CM2a) appears closer to the maximal correlation case (CM2b) rather than the zero correlation case (CM2c).The overall message, however, is that the difference between the two models does not increase dramatically.

Table 10 presents the performance of the minimum tracking-error portfolios. The reported realized tracking errors are around 1% ~ 2%, which are comparable to the tracking errors reported in Table 5. When the beta model is the DGP, the difference in the realized tracking errors can increase noticeably, from about 0.05% to about 0.94%. Note that these are monthly variances of the excess returns. 2% monthly variance is translated into 4.9% annualized standard deviation. 1% monthly variance is translated into 3.5% annualized standard deviation. Thus, in terms of annualized standard deviation, the two models produce the tracking errors of 4.9% vs. 3.5%. One may say that it is a substantial difference. When the characteristics model is the DGP, however, the difference changes marginally, from about 0.15% to about 0.40%.

The findings of this section can be summarized in terms of the loss from choosing a wrong model. If the true data generating process is the beta model and the investor chooses the characteristics model, the loss to the investor is estimated to be about 3% extra variance per month. If the correlation between beta and characteristics becomes much smaller than it is now, this loss may increase up to about 4% extra variance per month. On the other hand, if the correlation becomes large, the loss may disappear completely. In terms of the tracking error, the loss varies from none to 1% extra variance per month. If the true data generating process is the characteristics model, the loss to the investor who selects the beta model is relatively small.

9. Conclusion

For the purpose of the risk control, does it really matter whether one uses the beta model or the characteristics model? The analysis in this paper suggests that it does not matter so much. Historically, both models have comparable 'track records.' In some periods, the beta model produced better portfolios, whereas in other periods, the characteristics model did better. The difference between the two models depends on the correlation between beta and characteristics and also the correlation between factor and the characteristics price. If these correlations change unfavorably, the cost of choosing a wrong model may increase, but not very far from what we have experienced historically. The only exception is when the true model is the beta model and the investor creates a tracking error portfolio. Even in this case, however, considering that our analysis shows the worst possible scenario, the cost of choosing a wrong model is unlikely to be substantial in most situations.

While academics rarely use the characteristics model for the purpose of risk control, many practitioners do. This may have more to do with the intuitive appeal and the ease of use than with the belief in the superiority of this model. Nonetheless, if the characteristics model were significantly inferior to the beta model, this preference for the characteristics model would have been a puzzling phenomenon. My analysis provides an assurance that the practitioners' preference for the characteristics model is not an irrational phenomenon.

Appendix

In this appendix, I prove two claims made in Section 7: (i) If the BM3 model is the true DGP, then the diagonal part of $\hat{\Sigma}_{\mathrm{r}}^{\mathrm{CM}} - \Sigma_{\mathrm{r}}$ converges in probability to

$$\mathrm{D}\{2\psi\Sigma_e\psi' - \psi\Sigma_e{\beta_2}' - \beta_2\Sigma_e\psi' + 2\xi\mathrm{D}_\varepsilon\xi' - \xi\mathrm{D}_\varepsilon - \mathrm{D}_\varepsilon\xi'\} \tag{26}$$

whereas the off-diagonal part of $\hat{\Sigma}_{\mathrm{r}}^{\mathrm{CM}} - \Sigma_{\mathrm{r}}$ converges in probability to

$$\mathrm{O}\{\psi\Sigma_e\psi' - \psi\Sigma_e{\beta_2}' - \beta_2\Sigma_e\psi' + \xi\mathrm{D}_\varepsilon\xi' - \xi\mathrm{D}_\varepsilon - \mathrm{D}_\varepsilon\xi'\} \tag{27}$$

(ii) If the CM2 is the true DGP, then the diagonal elements of $\hat{\Sigma}_{\mathrm{r}}^{\mathrm{BM}} - \Sigma_{\mathrm{r}}$ converge in probability to zero, and that the off-diagonal part of $\hat{\Sigma}_{\mathrm{r}}^{\mathrm{BM}} - \Sigma_{\mathrm{r}}$ converges in probability to $X\Sigma_\nu X'$.

Proof of (i):

Suppose that the BM3 model is the true DGP. Then the true variance-covariance matrix of the return is

$$\Sigma_r = \beta\Sigma_f\beta' + D_\varepsilon \tag{28}$$

Σ_r can be further broken down in the following way. Let us partition β and f_t such that $\beta' = [\beta_1, \beta_2']$ and $f_t = [f_{1,t}, f_{2,t}']$. Consider the regression of $f_{2,t}$ on $f_{1,t}$, i.e. $f_{2,t} = a + bf_{1,t} + e$, where $a = \mu_{f,2} - \Sigma_{f,21}\Sigma_{f,11}^{-1}\mu_{f,1}$ and $b = \Sigma_{f,21}\Sigma_{f,11}^{-1}$. The return can be expressed as the following:

$$\begin{aligned} r_t &= \alpha + \beta_1 f_{1,t} + \beta_2 f_{2,t} + \varepsilon_t \\ &= \alpha + \beta_2 a + (\beta_1 + \beta_2 b) f_{1,t} + \beta_2 e + \varepsilon_t \end{aligned} \tag{29}$$

Based on the above expression, Σ_r can be written as the following:

$$\Sigma_r = (\beta_1 + \beta_2 b)\Sigma_{f,11}(\beta_1 + \beta_2 b)' + \beta_2 \Sigma_e \beta' + D_\varepsilon \tag{30}$$

where $\Sigma_e = \Sigma_{f,2} - b\Sigma_{f,11}b'$.

The estimator based on the CM2 model is

$$\hat{\Sigma}_r^{CM} = \hat{\beta}_M \Sigma_{f,11}\hat{\beta}_M' + X\hat{\Sigma}_\Delta X' + \hat{D}_\eta \tag{31}$$

Comparing $\hat{\Sigma}_r^{CM}$ to Σ_r, the first part of $\hat{\Sigma}_r^{CM}$, i.e. $\hat{\beta}_M\Sigma_{f,11}\hat{\beta}_M'$ converges to the first part of Σ_r, i.e. $(\beta_1 + \beta_2 b)\Sigma_{f,11}(\beta_1 + \beta_2 b)'$, to be shown shortly. Then we only need to compare the remaining parts, i.e. $+X\hat{\Sigma}_\Delta X' + \hat{D}_\eta$ to $\beta_2\Sigma_e\beta' + D_\varepsilon$.

The convergence of $\hat{\beta}_M\Sigma_{f,11}\hat{\beta}_M'$ to $(\beta_1 + \beta_2 b)\Sigma_{f,11}(\beta_1 + \beta_2 b)'$ can be established in the following way. Given the IID property of the BM3 model,

$$\hat{\beta}_M \to cov(r_t, f_{1,t})\Sigma_{f,11}^{-1} = \beta_1 + \beta_2 b \tag{32}$$

Thus,

$$\hat{\beta}_M \Sigma_{f,11}\hat{\beta}_M' \to (\beta_1 + \beta_2 b)\Sigma_{f,11}(\beta_1 + \beta_2 b)' \tag{33}$$

I now compare $X\hat{\Sigma}_\Delta X' + \hat{D}_\eta$ to $\beta_2\Sigma_e\beta' + D_\varepsilon$. The OLS estimator of Δ_t is $\hat{\Delta}_t = (X'X)^{-1}X'\tilde{r}_t$ and its variance-covariance matrix estimator is

$$\hat{\Sigma}_\Delta = (X'X)^{-1}X'\widehat{var}(\tilde{r}_t)X(X'X)^{-1} \tag{34}$$

Thus,

$$X\hat{\Sigma}_\Delta X' = X(X'X)^{-1}X'\widehat{var}(\tilde{r}_t)X(X'X)^{-1}X' \tag{35}$$

From the fact that $\hat{\eta}_t = \tilde{r}_t - X\hat{\Delta}_t$,

$$\begin{aligned} \hat{D}_\eta &= D\{\widehat{var}(\tilde{r}_t) + X\hat{\Sigma}_\Delta X' - \widehat{cov}(\tilde{r}_t, \hat{\Delta}_t{}')X' - X\widehat{cov}(\hat{\Delta}_t, \tilde{r}_t{}')\} \\ &= D\{\widehat{var}(\tilde{r}_t) + X(X'X)^{-1}X'\widehat{var}(\tilde{r}_t)X(X'X)^{-1}X'' \\ &\quad -\widehat{var}(\tilde{r}_t)X(X'X)^{-1}X' - X(X'X)^{-1}X'\widehat{var}(\tilde{r}_t)\} \end{aligned} \tag{36}$$

Note that $\hat{\eta}_t$ is correlated with $\hat{\Delta}_t$, as $\hat{\Delta}_t$ is obtained from cross-sectional regressions, not from time-series regressions. Thus, the above formula cannot be reduced further. As T increases, $\widehat{\mathrm{var}}(\tilde{r}_t)$ estimates the population parameter consistently, i.e.

$$\widehat{\mathrm{var}}(\tilde{r}_t) \to var(\tilde{r}_t) = \beta_2 \Sigma_e \beta_2' + D_\varepsilon \tag{37}$$

Consider the diagonal elements of $X\hat{\Sigma}_\Delta X' + \hat{D}_\eta$. We would like to check whether the probability limit of $D\left(X\hat{\Sigma}_\Delta X' + \hat{D}_\eta\right)$ is close to $D(\beta_2 \Sigma_e \beta_2' + D_\varepsilon)$. From the formula above, one can see that the probability limit of $D\left(X\hat{\Sigma}_\Delta X' + \hat{D}_\eta\right)$ is identical to $D(\beta_2 \Sigma_e \beta_2' + D_\varepsilon)$ if $X(X'X)^{-1}X'$ were an identity matrix. Of course, $X(X'X)^{-1}X'$ cannot be an identity matrix (the rank of the former is at most 3, whereas the rank of the latter is N,) so the probability limit of $D\left(X\hat{\Sigma}_\Delta X' + \hat{D}_\eta\right)$ cannot be identical to $D(\beta_2 \Sigma_e \beta_2' + D_\varepsilon)$. I define ξ as the difference between $X(X'X)^{-1}X'$ and the identity matrix *I*, i.e. $I = X(X'X)^{-1}X' + \xi$. We can also see from the formula above that the difference between the probability limit of $D\left(X\hat{\Sigma}_\Delta X' + \hat{D}_\eta\right)$ and $D(\beta_2 \Sigma_e \beta_2' + D_\varepsilon)$ is substantially reduced if $X(X'X)^{-1}X'\beta_2$ is close to β_2. Let ψ be the difference between β_2 and $X(X'X)^{-1}X'\beta_2$, i.e. $\beta_2 = X(X'X)^{-1}X'\beta_2 + \psi$. Then

$$D\left(X\hat{\Sigma}_\Delta X' + \hat{D}_\eta\right) \to D\begin{Bmatrix} \beta_2 \Sigma_e \beta_2{}' - \psi \Sigma_e \beta_2{}' - \beta_2 \Sigma_e \psi' + 2\psi \Sigma_e \psi' \\ +D_\varepsilon - \xi D_\varepsilon - D_\varepsilon \xi' + 2\xi D_\varepsilon \xi' \end{Bmatrix} \tag{38}$$

I now consider the off-diagonal elements of $X\hat{\Sigma}_\Delta X' + \hat{D}_\eta$. I would like to check whether the probability limit of $O\left(X\hat{\Sigma}_\Delta X'\right)$ is close to $O(\beta_2 \Sigma_e \beta_2')$. As we have seen above,

$$O\left(X\hat{\Sigma}_\Delta X'\right) \to O\begin{Bmatrix} \beta_2 \Sigma_e \beta_2{}' - \psi \Sigma_e \beta_2{}' - \beta_2 \Sigma_e \psi' + \psi \Sigma_e \psi' \\ -\xi D_\varepsilon - D_\varepsilon \xi' + \xi D_\varepsilon \xi' \end{Bmatrix} \tag{39}$$

This completes the proof.

Proof of (ii):

Suppose that the characteristics model is the true DGP. Then the true variance-covariance matrix of returns is

$$\Sigma_r = \beta_M \Sigma_{f,11} \beta_M' + X \Sigma_\Delta X' + D_\eta \tag{40}$$

The estimator based on the BM3 model is

$$\hat{\Sigma}_r^{BM} = \hat{\beta} \hat{\Sigma}_f \hat{\beta}' + \hat{D}_\varepsilon \tag{41}$$

As we noted before, the variance estimates of the beta model are identical to the sample variances.

That is, $\mathrm{D}(\hat{\Sigma}_{\mathrm{r}}^{\mathrm{BM}}) = D(\hat{\Sigma}_{\mathrm{r}})$. Thus, the diagonal elements of $\hat{\Sigma}_{\mathrm{r}}^{\mathrm{BM}}$ consistently estimates the diagonal elements of Σ_{r}.

For the off-diagonal elements, I start from the "partitioning" of $\hat{\Sigma}_{\mathrm{r}}^{\mathrm{BM}}$:

$$\hat{\Sigma}_r^{BM} = (\hat{\beta}_1 + \hat{\beta}_2\hat{b})\hat{\Sigma}_{f,11}(\hat{\beta}_1 + \hat{\beta}_2\hat{b})' + \hat{\beta}_2\hat{\Sigma}_e\hat{\beta}' + \hat{D}_\varepsilon \tag{42}$$

where $\hat{\mathrm{b}} = \hat{\Sigma}_{f,21}\hat{\Sigma}_{f,11}^{-1}$ and $\hat{\Sigma}_e = \hat{\Sigma}_{\mathrm{f},2} - \hat{b}\hat{\Sigma}_{f,11}\hat{b}'$. The first part of $\hat{\Sigma}_{\mathrm{r}}^{\mathrm{BM}}$ converges to the first part of Σ_{r} as $\hat{\beta}_1 + \hat{\beta}_2\hat{b} \to \beta_{\mathrm{M}}$ and $\hat{\Sigma}_{\mathrm{f},11} \to \Sigma_{\mathrm{f},11}$. Recall from the "partial regression formula" that $\hat{\beta}_2$ can be interpreted as the slope in the regression of $\tilde{\mathrm{r}}_{\mathrm{t}}$ on $\mathrm{f}_{2,\mathrm{t}}$ and the constant, and equivalently the regression of $\tilde{\mathrm{r}}_{\mathrm{t}}$ on e_{t}. (e_{t} is the error in the regression of $\mathrm{f}_{2,\mathrm{t}}$ on $\mathrm{f}_{1,\mathrm{t}}$ and the constant. See the formula above (29).) Thus,

$$\begin{aligned}\hat{\beta}_2 &\to cov(\tilde{r}_t, f_{2,t}{}')\Sigma_{f,22}^{-1} = Xcov(\Delta_t, f_{2,t}{}')\Sigma_{f,22}^{-1} \\ &= cov(\tilde{r}_t, e_t{}')\Sigma_e^{-1} = Xcov(\Delta_t, e_t{}')\Sigma_e^{-1}\end{aligned} \tag{43}$$

and

$$\begin{aligned}\hat{\beta}_2\hat{\Sigma}_e\hat{\beta}' &\to Xcov(\Delta_t, f_{2,t}{}')\Sigma_{f,22}^{-1}cov(e_t, \Delta_t{}')X' \\ &= Xcov(\Delta_t, f_{2,t}{}')\Sigma_{f,22}^{-1}cov(f_{2,t}, \Delta_t{}')X'\end{aligned} \tag{44}$$

Above, I assume that η_{t} is independent from $f_{2,t}$. I also use the property that Δ_t is independent from $f_{1,t}$ (which comes from the fact that $\tilde{\mathrm{r}}_{\mathrm{t}}$ is the residual.)

Let $\phi_1 + \phi_2 f_{2,t}$ be the best linear predictor of Δ_t. That is, $\Delta_t = \phi_1 + \phi f_{2,t} + \nu_t$, where ν_t has the property of the regression error term. Then

$$O(\hat{\beta}_2\hat{\Sigma}_e\hat{\beta}') \to X\Sigma_\Delta X' - X\Sigma_\nu X' \tag{45}$$

This completes the proof.

References

Asprouli, M, and D. Lamb (2002), *BITA Plus Multi Market Equity Risk Model*, London: Cor Risk Solutions.

Bajeux-Besnainou, Isabelle, Wachindra Bandara, and Efstathia Bura (2012). "A Krylov Subspace Approach to Large Portfolio Optimization," *Journal of Economic Dynamics and Control*, 36(11), 1688-1699.

Brennan, Michael J., Tarun Chordia, and Avanidhar Subrahmanyam (1998), "Alternative Factor

Specifications, Security Characteristics, and the Cross-Section of Expected Stock Returns," *Journal of Financial Economics*, 49, 345-373.

Calvetti, D., B. Lewis, and L. Reichel (2000), "GMRES-type Methods for Inconsistent Systems," *Linear Algebra and its Applications*, 316, 157-169.

Campbell, S. L., and C. D. Meyer (1979), *Generalized Inverse of Linear Transformations*, London: Pitman.

Chamberlain, Gary (2000), "Econometric Applications of the Maximin Expected Utility," *Journal of Applied Econometrics*, 15, 625-644.

Chan, Louis K. C., Jason Karceski, and Josef Lakonishok (1999), "On Portfolio Optimization," *Review of Financial Studies*, 12(5), 937-974.

Connor, Gregory (1995), "The Three Types of Factor Models: A Comparison of Their Explanatory Power," *Financial Analysts Journal*, 51(6), 42-46.

Daniel, Kent, and Sheridan Titman (1997), "Evidence on the Characteristics of Cross-sectional Variation in Stock Returns," *Journal of Finance*, 52(1), 1-33.

Daniel, Kent, Sheridan Titman, and K. C. John Wei (2001), "Explaining the cross-section of stock returns in Japan: Factors or Characteristics?" *Journal of Finance*, 56(2), 743-766.

Davis, James, Eugene F. Fama, and Kenneth F. French (2000), "Characteristics, covariances, and average returns: 1929-1997," *Journal of Finance*, 55(1), 389-406.

Eckart, Carl, and Gale Young (1936), "The Approximation of One Matrix with Another of Lower One," *Psychometrika*, 1(3), 211-218.

Fama, Eugene F., and Kenneth R. French (1992), "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47(2), 427-465.

-------- (1997), "Industry Costs of Equity," *Journal of Financial Economics*, 43, 153-193.

-------- (2008), "Dissecting Anomalies," *Journal of Finance*, 63(4), 1653-1678

Fama, Eugene F., and James D. MacBeth (1973), "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 1973, 607-636.

Hou, Kewei, G. Andrew Karolyi, and Bong-Chan Kho (2011), "What Factors Drive Global Stock Returns?" *The Review of Financial Studies*, 24(8), 2527-2574.

Ipsen, Ilse C., and Carl D. Meyer (1998), "The Idea behind Krylov Methods," *The American Mathematical Monthly*, 105(10), 889-899.

Jobson, J. D. and Bob M. Korkie (1981), "Performance hypothesis testing with the Sharpe and Treynor measures," *Journal of Finance*, 36, 889-908.

Kendall, Shmuel, and Robert F. Stambaugh (1996), "On the Predictability of Stock Returns: An Asset Allocation Perspective," *Journal of Finance*, 51(2), 385-424.

Ledoit, Olivier, and Michael Wolf (2003), “Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection,” *Journal of Empirical Finance*, 10, 603-621.

Lewellen, Jonathan (1999), “The Time-series Relations among Expected Return, Risk, and Book-to-Market,” *Journal of Financial Economics*, 54, 1999, 5-43

McCulloch, Robert, and Peter E. Rossi (1990), “Posterior, Predictive, and Utility-Based Approaches to Testing the Arbitrage Pricing Theory,” *Journal of Financial Economics*, 28(1), 7-38.

Memmel, Christoph (2003), “Performance Hypothesis Testing with the Sharpe Ratio,” *Finance Letters*, 1, 21-23.

Menchero, Jose, D. J. Orr, and Jun Wang (2011), *The Barra US Equity Model (USE4) Methodology Notes*, New York: MSCI Inc.

Morgan, W. A. (1939), “A Test of the Significance of the Difference Between the Two Variances in a Sample From a Normal Bivariate Population,” *Biometrika*, 31(1), 13-19.

Northfield (2012a), *U.S. Fundamental Equity Risk Model*, Boston: Northfield Information Services, Inc.

-------- (2012b), *U.S. Macroeconomic Equity Risk Model*, Boston: Northfield Information Services, Inc.

Pastor, Lubos, and Robert F. Stambaugh (2000), “Comparing Asset Pricing Models: an Investment Perspective,” *Journal of Financial Economics*, 56(3), 335-381.

Pitman, E. J. G. (1939), “A Note on Normal Correlation,” *Biometrika*, 31(1), 9-12.

Rosenberg, Barr (1974), “Extra-Market Components of Covariance among Security Prices,” *Journal of Financial and Quantitative Analysis*, 9(2), 263-274.

Tables

Table 1. Summary Statistics

Statistics are reported for each cohort of stocks, i.e. the S&P 500 index constituents as of the end of year t (where t is one of 1980, 1985, 1990, 1995, 2000, and 2005.) Monthly returns for the 10-year period around year t, from year t-4 to year t+5, have been collected. ME is the market capitalization in billion dollars as of the beginning of year t-4. BM is the book-to-market ratio; the book value is from the last fiscal quarter ending before the beginning of year t-4, and the market value is identical to ME. Stocks with missing ME and BM are excluded from the analysis. MRF (the market return in excess of the riskfree rate), SMB (small minus big), HML (high minus low), and RF (riskfree rate) are monthly Fama-French factor returns for the 10-year period.

	N	Mean	SD	Min	Max		N	Mean	SD	Min	Max
		(A) Cohort 1980						(D) Cohort 1995			
Return	43324	1.47	8.85	-60.00	152.29	Return	51233	1.51	9.40	-75.92	120.00
ME	375	1.45	3.54	0.01	36.24	ME	448	4.74	7.97	0.06	64.53
BM	375	1.04	0.73	0.05	4.39	BM	448	0.73	0.63	0.01	8.34
MRF	120	0.58	4.37	-13.23	12.13	MRF	120	0.99	3.96	-16.21	10.30
SMB	120	0.75	2.47	-9.91	6.98	SMB	120	0.04	4.13	-16.62	22.06
HML	120	0.59	2.80	-8.52	8.59	HML	120	0.29	3.69	-12.87	12.39
RF	120	0.72	0.25	0.34	1.35	RF	120	0.39	0.09	0.21	0.56
		(B) Cohort 1985						(E) Cohort 2000			
Return	44087	1.31	9.37	-88.34	118.52	Return	50548	1.43	11.20	-98.04	209.62
ME	380	2.19	4.46	0.01	39.61	ME	433	8.99	14.90	0.04	120.26
BM	380	0.89	0.64	0.01	6.13	BM	433	0.39	0.25	0.02	2.84
MRF	120	0.40	4.80	-23.14	12.43	MRF	120	0.56	4.67	-16.21	8.18
SMB	120	-0.16	2.31	-8.41	6.15	SMB	120	0.30	4.38	-16.62	22.06
HML	120	0.57	2.61	-5.95	7.56	HML	120	0.46	3.98	-12.87	13.88
RF	120	0.69	0.22	0.29	1.35	RF	120	0.30	0.15	0.06	0.56
		(C) Cohort 1990						(F) Cohort 2005			
Return	52879	1.29	8.93	-56.41	109.02	Return	53317	0.85	11.58	-98.66	396.33
ME	445	3.20	6.22	0.05	95.61	ME	467	22.85	46.16	0.77	475.00
BM	445	0.61	0.36	0.01	3.79	BM	467	0.37	0.31	0.01	2.68
MRF	120	0.73	4.27	-23.14	12.43	MRF	120	0.20	4.95	-18.55	11.04
SMB	120	-0.17	2.39	-8.41	8.43	SMB	120	0.58	2.76	-6.53	6.98
HML	120	0.23	2.27	-5.95	6.50	HML	120	0.40	3.07	-9.93	13.88
RF	120	0.45	0.14	0.21	0.79	RF	120	0.18	0.15	0.00	0.54

Table 2. Accuracy of Variance Forecasts

Variance forecasts are compared to the realized variances for each cohort of stocks, i.e. the S&P 500 index constituents as of the end of year t (where t is one of 1980, 1985, 1990, 1995, 2000, and 2005.) Variance forecasts are the estimates obtained from the 5-year period prior to the end of year t, i.e. from year t-4 to year t. Realized variances are the sample variances from the 5-year period after the end of year t, i.e. from year t+1 to year t+5. See the text for the description of the models. RMSE is the root mean squared error. Rho is the Pearson correlation, and tau is Kendall's rank correlation. Figures inside parenthesis are bootstrap standard errors.

Model:	BM0/BM1/BM3	CM2
	(A) Cohort 1980	
N	375	375
RMSE	45.43	45.19
	(6.05)	(6.18)
Rho	0.57	0.56
	(0.05)	(0.06)
Tau	0.53	0.53
	(0.04)	(0.04)
	(B) Cohort 1985	
N	380	380
RMSE	43.54	43.29
	(2.40)	(2.34)
Rho	0.60	0.60
	(0.04)	(0.04)
Tau	0.47	0.46
	(0.04)	(0.04)
	(C) Cohort 1990	
N	445	445
RMSE	48.65	48.31
	(2.14)	(2.31)
Rho	0.67	0.67
	(0.03)	(0.03)
Tau	0.53	0.53
	(0.03)	(0.03)

Model:	BM0/BM1/BM3	CM2
	(D) Cohort 1995	
N	448	448
RMSE	89.57	89.86
	(7.59)	(7.31)
Rho	0.62	0.63
	(0.05)	(0.05)
Tau	0.49	0.48
	(0.04)	(0.04)
	(E) Cohort 2000	
N	433	433
RMSE	122.31	123.75
	(21.65)	(21.88)
Rho	0.61	0.61
	(0.08)	(0.08)
Tau	0.48	0.49
	(0.03)	(0.03)
	(F) Cohort 2005	
N	467	467
RMSE	302.76	302.10
	(53.44)	(52.11)
Rho	0.03	0.03
	(0.04)	(0.04)
Tau	0.24	0.23
	(0.03)	(0.03)

Table 3. Accuracy of Covariance Forecasts

Covariance forecasts are compared to the realized variances for each cohort of stocks, i.e. the S&P 500 index constituents as of the end of year t (where t is one of 1980, 1985, 1990, 1995, 2000, and 2005.) Covariance forecasts are the estimates obtained from the 5-year period prior to the end of year t, i.e. from year t-4 to year t. Realized covariances are the sample variances from the 5-year period after the end of year t, i.e. from year t+1 to year t+5. See the text for the description of the models. RMSE is the root mean squared error. Rho is the Pearson correlation, and tau is Kendall's rank correlation. Figures inside parenthesis are bootstrap standard errors.

	Model: BM0	BM1	BM3	CM2
	(A) Cohort 1980			
N	70125	70125	70125	70125
RMSE	15.65	14.20	14.60	14.64
	(0.08)	(0.07)	(0.07)	(0.07)
Rho	0.44	0.43	0.42	0.43
	(0.00)	(0.00)	(0.00)	(0.00)
Tau	0.33	0.32	0.32	0.32
	(0.00)	(0.00)	(0.00)	(0.00)
	(B) Cohort 1985			
N	72010	72010	72010	72010
RMSE	22.78	22.77	22.72	22.06
	(0.07)	(0.07)	(0.07)	(0.07)
Rho	0.47	0.45	0.46	0.46
	(0.00)	(0.00)	(0.00)	(0.00)
Tau	0.34	0.33	0.34	0.34
	(0.00)	(0.00)	(0.00)	(0.00)
	(C) Cohort 1990			
N	98790	98790	98790	98790
RMSE	32.13	30.69	30.88	31.19
	(0.06)	(0.05)	(0.06)	(0.05)
Rho	0.39	0.36	0.38	0.36
	(0.00)	(0.00)	(0.00)	(0.00)
Tau	0.27	0.26	0.27	0.26
	(0.00)	(0.00)	(0.00)	(0.00)

	Model: BM0	BM1	BM3	CM2
	(D) Cohort 1995			
N	100128	100128	100128	100128
RMSE	23.20	23.29	23.13	23.07
	(0.12)	(0.13)	(0.13)	(0.13)
Rho	0.24	0.17	0.21	0.18
	(0.01)	(0.00)	(0.00)	(0.01)
Tau	0.16	0.13	0.15	0.14
	(0.00)	(0.00)	(0.00)	(0.00)
	(E) Cohort 2000			
N	93528	93528	93528	93528
RMSE	31.82	26.89	28.72	26.06
	(0.21)	(0.24)	(0.21)	(0.24)
Rho	0.43	0.57	0.47	0.57
	(0.01)	(0.00)	(0.01)	(0.00)
Tau	0.18	0.30	0.19	0.31
	(0.00)	(0.00)	(0.00)	(0.00)
	(F) Cohort 2005			
N	108811	108811	108811	108811
RMSE	57.49	56.19	56.01	55.42
	(1.08)	(1.10)	(1.16)	(1.13)
Rho	0.21	0.21	0.22	0.22
	(0.01)	(0.01)	(0.01)	(0.01)
Tau	0.24	0.25	0.25	0.26
	(0.00)	(0.00)	(0.00)	(0.00)

Table 4. Minimum Variance Portfolios

Out of each cohort, i.e. the S&P 500 index constituents as of the end of year t (where t is one of 1980, 1985, 1990, 1995, 2000, 2005), the minimum variance portfolios are created and the predicted variances (Pred Var) are estimated based on three models – BM1, BM3, and CM2. Models are estimated from the 5-year period between year t-4 and year t. See the text for the description of these models. The monthly returns of these portfolios are calculated for the 5-year period from year t+1 to year t+5. Realized variances (Real Var) and the realized Sharpe ratios (Real SR) are calculated out of these returns. For Real Var, the figures inside the square brackets are the t statistics of Morgan-Pitman test for the null hypothesis that BM1 and BM3 variances are not different from CM2 variances. For Real SR, the figures inside the square brackets are the Jobson-Korkie z-statistic for the null hypothesis that BM1 and BM3 Sharpe ratios are not different from CM2 Sharpe ratios. * and ** indicate the significance at the 90% and 95% levels, respectively.

Model:	BM1	BM3	CM2
(A) Cohort 1980			
Pred Var	2.98	3.65	4.27
Real Var	10.43	10.05	10.29
	[0.71]	[0.59]	
Real SR	0.34	0.33	0.33
	[1.11]	[0.17]	
(B) Cohort 1985			
Pred Var	3.67	4.08	4.59
Real Var	19.99	23.20	20.33
	[0.47]	[3.08]	
		**	
Real SR	0.19	0.18	0.21
	[1.16]	[1.32]	
(C) Cohort 1990			
Pred Var	6.02	8.73	6.80
Real Var	10.00	7.47	9.59
	[3.06]	[2.93]	
	**	**	
Real SR	0.21	0.16	0.21
	[0.59]	[1.20]	

Model:	BM1	BM3	CM2
(D) Cohort 1995			
Pred Var	1.69	2.03	2.01
Real Var	14.79	12.92	13.44
	[3.00]	[0.47]	
	**		
Real SR	0.13	0.10	0.14
	[0.59]	[0.93]	
(E) Cohort 2000			
Pred Var	0.85	2.98	5.92
Real Var	10.26	10.51	7.45
	[2.94]	[2.93]	
	**	**	
Real SR	0.16	0.14	0.08
	[1.39]	[1.02]	
(F) Cohort 2005			
Pred Var	0.73	1.30	1.91
Real Var	9.22	10.85	9.68
	[1.65]	[2.06]	
	*	**	
Real SR	0.07	0.07	0.07
	[0.17]	[0.14]	

Table 5. Minimum Tracking Error Portfolios

Out of each cohort, i.e. the S&P 500 index constituents as of the end of year t (where t is one of 1980, 1985, 1990, 1995, 2000, 2005), the minimum tracking error portfolios are created and the predicted tracking errors (Pred TE) are estimated based on three models – BM1, BM3, and CM2. Models are estimated from the 5-year period between year t-4 and year t. See the text for the description of these models. The tracking error is measured as the variance of the portfolio return in excess of the benchmark. The monthly returns of these portfolios are calculated for the 5-year period from year t+1 to year t+5. Realized tracking errors (Real TE) and the realized Sharpe ratios (Real SR) are calculated out of these returns. For Real TE, the figures inside the square brackets are the t statistics of Morgan-Pitman test for the null hypothesis that BM1 and BM3 tracking errors are not different from CM2 tracking errors. For Real SR, the figures inside the square brackets are the Jobson-Korkie z-statistic for the null hypothesis that BM1 and BM3 Sharpe ratios are not different from CM2 Sharpe ratios. * and ** indicate the significance at the 90% and 95% levels, respectively.

	Model:		
	BM1	BM3	CM2
	(A) Cohort 1980		
Pred TE	0.10	0.11	1.22
Real TE	1.14	0.86	1.22
	[0.52]	[2.73]	
		**	
Real SR	0.18	0.16	0.17
	[0.65]	[0.83]	
	(B) Cohort 1985		
Pred TE	0.11	0.12	0.57
Real TE	0.55	0.53	0.46
	[0.94]	[0.75]	
Real SR	0.10	0.10	0.11
	[1.33]	[1.23]	
	(C) Cohort 1990		
Pred TE	0.09	0.09	0.37
Real TE	0.26	0.28	0.31
	[1.36]	[0.53]	
Real SR	0.36	0.36	0.36
	[0.26]	[0.37]	

	Model:		
	BM1	BM3	CM2
	(D) Cohort 1995		
Pred TE	0.08	0.08	0.35
Real TE	6.18	5.18	6.09
	[0.47]	[4.66]	
		**	
Real SR	0.23	0.24	0.24
	[1.42]	[0.69]	
	(E) Cohort 2000		
Pred TE	0.28	0.39	4.02
Real TE	1.52	1.16	1.13
	[1.34]	[0.10]	
Real SR	0.14	0.10	0.07
	[2.12]	[1.08]	
	**		
	(F) Cohort 2005		
Pred TE	0.11	0.13	1.15
Real TE	0.63	0.71	0.58
	[0.59]	[1.94]	
		*	
Real SR	0.04	0.04	0.03
	[0.97]	[0.53]	

Table 6. Regression of Beta on Characteristics

For the 2005 cohort, i.e. the S&P 500 index constituents as of the end of year 2005, the BM3 model is estimated from the 5-year period between year 2001 and year 2005. Beta 2 (SMB) and beta 3 (HML) are the coefficients on SMB and HML in the BM3 model, respectively. See the text for the further description of the BM3 model. ME is the market capitalization as of the beginning of year 2001. BM is the book-to-market ratio, where the book value is as of the last fiscal quarter ending prior to the beginning of year 2001 and the market value is identical to ME. Stocks with missing BM and ME are excluded. The regression reported below are the cross-sectional regression (N=467). The figures inside the round brackets are the standard errors, and the figures inside the square brackets are t statistics. * and ** indicate the significance at the 90% and 95% levels, respectively.

	Dependent Variable:	
	Beta 2 (SMB)	Beta 3 (HML)
RHS Variables:		
Intercept	0.2036	0.0694
	(0.0426)	(0.0616)
	[4.78]	[1.13]
	**	
ME	-0.0037	-0.0020
	(0.0005)	(0.0008)
	[-6.71]	[-2.48]
	**	**
BM	0.1991	0.6549
	(0.0812)	(0.1174)
	[2.45]	[5.58]
	**	**
R sq.	0.12	0.09

Table 7. Matrix ξ

ξ is a 467-by-467 square matrix. Columns of ξ are the residual in the regression of columns of the identity matrix I on the columns of matrix X, i.e. $\xi = I - X(X'X)^{-1}X'$. X is a 467-by-2 matrix; the first and the second columns are ME and BM of the 2005 cohort, i.e. the S&P 500 index constituents as of the end of year 2005. ME is the market capitalization as of the beginning of year 2001. BM is the book-to-market ratio, where the book value is as of the last fiscal quarter ending prior to the beginning of year 2001 and the market value is identical to ME. Stocks with missing BM and ME are excluded.

	Mean	SD	Min	Max
Diagonal Elements	0.9936	0.0146	0.7880	0.9979
Off-diagonal Elements	-0.0021	0.0029	-0.1320	0.0175

Table 8. Regression of the Characteristics Prices on the Factor Realizations

For the 2005 cohort, i.e. the S&P 500 index constituents as of the end of year 2005, the CM2 model is estimated from the 5-year period between year 2001 and year 2005. Delta 1, delta 2 (ME), and delta 3 (BM) are the characteristics prices, i.e. the intercept and coefficients on ME and BM in the CM2 model. See the text for the further description of the CM2 model. SMB and HML are monthly Fama-French factor returns. The regression reported below are a time-series regression (from January 2001 to December 2005; T=60). The figures inside the round brackets are the standard errors, and the figures inside the square brackets are t statistics. * and ** indicate the significance at the 90% and 95% levels, respectively.

	Dependent Variable:		
	Delta 1	Delta 2 (ME)	Delta 3 (BM)
RHS Variables:			
Intercept	-0.2237	0.0046	-0.6425
	(0.2114)	(0.0012)	(0.3659)
	[-1.06]	[3.65]	[-1.76]
		**	*
SMB	0.1969	-0.0036	0.2556
	(0.0666)	(0.0004)	(0.1153)
	[2.96]	[-9.24]	[2.22]
	**	**	**
HML	0.0817	-0.0021	0.5512
	(0.0642)	(0.0004)	(0.1111)
	[1.27]	[-5.41]	[4.96]
		**	**
R sq.	0.14	0.63	0.31

Table 9. Sensitivity Analysis – Minimum Variance Portfolios

For the 2005 cohort, i.e. the S&P 500 index constituents as of the end of year 2005, the BM3 model and the CM2 model are estimated from the 5-year period between year 2001 and year 2005. BM3a is a data generating process based on this estimate of the BM3 model. In BM3b, the correlation between factor and characteristics is assumed to be maximal, whereas, in BM3c, the correlation is assumed to be zero. CM2a is based on the estimate of the CM2 model. In CM2b, the correlation between the characteristics price and factor is assumed to be maximal, where, in CM2c, the correlation is assumed to be zero. Out of each data generating process, a 10-year pseudo-random sample is generated; the BM3 model and the CM2 model are estimated from the first 5-year data; the minimum variance portfolios are created, and the predicted variances (Pred Var) are estimated; the realized variances (Real Var) and the Sharpe ratios (Real SR) are calculated from the last 5-year data. The entire simulation is repeated 500 times. * and ** indicate the significance at the 90% and 95% levels, respectively.

		DGP: BM3a	BM3b	BM3c	CM2a	CM2b	CM2c
				(A) Model: BM3			
Pred Var	mean	0.8510	0.8473	0.8453	0.6424	0.5563	0.5907
	s.d.	0.1364	0.1424	0.1258	0.1290	0.1151	0.1176
Real Var	mean	2.7628	2.6832	2.7555	2.5818	2.5483	2.8870
	s.d.	0.8982	0.8084	0.7800	0.5939	0.5087	0.6310
Real SR	mean	0.6407	0.6615	0.6406	0.6357	0.6262	0.5811
	s.d.	0.1414	0.1531	0.1466	0.1537	0.1454	0.1383
				(B) Model: CM2			
Pred Var	mean	1.0454	1.1019	1.4145	1.7816	1.8222	1.8912
	s.d.	0.1878	0.1456	0.2560	0.2564	0.2588	0.2640
Real Var	mean	5.8241	2.3804	6.9414	2.3807	2.3742	2.5388
	s.d.	1.7534	0.6489	1.7545	0.5122	0.4662	0.5131
Real SR	mean	0.5172	0.6671	0.5362	0.6531	0.6411	0.6159
	s.d.	0.1385	0.1513	0.1392	0.1510	0.1496	0.1435
				(C) Test of Difference			
Real Var	mean	-3.0613	0.3028	-4.1859	0.2011	0.1741	0.3482
	s.e.	(0.0592)	(0.0180)	(0.0665)	(0.0112)	(0.0104)	(0.0152)
	t stat	[-51.72]	[16.87]	[-62.90]	[17.98]	[16.82]	[22.93]
		**	**	**	**	**	**

Table 10. Sensitivity Analysis – Minimum Tracking Error Portfolios

For the 2005 cohort, i.e. the S&P 500 index constituents as of the end of year 2005, the BM3 model and the CM2 model are estimated from the 5-year period between year 2001 and year 2005. BM3a is a data generating process based on this estimate of the BM3 model. In BM3b, the correlation between factor and characteristics is assumed to be maximal, whereas, in BM3c, the correlation is assumed to be zero. CM2a is based on the estimate of the CM2 model. In CM2b, the correlation between the characteristics price and factor is assumed to be maximal, where, in CM2c, the correlation is assumed to be zero. Out of each data generating process, a 10-year pseudo-random sample is generated; the BM3 model and the CM2 model are estimated from the first 5-year data; the minimum tracking-error portfolios are created, and the predicted tracking error (Pred TE) are estimated; the realized tracking errors (Real TE) and the Sharpe ratios (Real SR) are calculated from the last 5-year data. The tracking error is measured as the variance of the portfolio return in excess of the benchmark. The entire simulation is repeated 500 times. * and ** indicate the significance at the 90% and 95% levels, respectively.

		DGP: BM3a	BM3b	BM3c	CM2a	CM2b	CM2c
				(A) Model: BM3			
Pred TE	mean	0.1082	0.1083	0.1086	0.1603	0.1526	0.1155
	s.d.	0.0047	0.0049	0.0048	0.0205	0.0186	0.0081
Real TE	mean	0.9470	0.9135	0.9415	1.8844	1.8346	2.0890
	s.d.	0.5248	0.4696	0.5481	0.5771	0.6169	0.5359
Real SR	mean	0.1829	0.1921	0.1945	0.2176	0.2246	0.2126
	s.d.	0.1149	0.1187	0.1249	0.1204	0.1153	0.1252
				(B) Model: CM2			
Pred TE	mean	0.4904	0.2904	0.9866	1.3439	1.3806	1.3709
	s.d.	0.0927	0.0355	0.2235	0.2290	0.2322	0.2295
Real TE	mean	0.9978	0.8010	1.8792	1.7388	1.6789	1.6899
	s.d.	0.4742	0.4494	0.7509	0.5461	0.5608	0.4943
Real SR	mean	0.1880	0.1795	0.2529	0.2173	0.2216	0.2140
	s.d.	0.1150	0.1194	0.1247	0.1196	0.1167	0.1240
				(C) Test of Difference			
Real TE	mean	-0.0508	0.1125	-0.9377	0.1456	0.1557	0.3991
	s.e.	(0.0118)	(0.0050)	(0.0176)	(0.0104)	(0.0086)	(0.0140)
	t stat	[-4.30] **	[22.29] **	[-53.15] **	[14.04] **	[18.10] **	[28.45] **