

Asset Pricing Models: Implications for Expected Returns and Portfolio Selection

A. Craig MacKinlay

University of Pennsylvania and NBER

Ľuboš Pástor

University of Chicago

When a risk factor is missing from an asset pricing model, the resulting mispricing is embedded within the residual covariance matrix. Exploiting this phenomenon leads to expected return estimates that are more stable and precise than estimates delivered by standard methods. Portfolio selection can also be improved. At an extreme, optimal portfolio weights are proportional to expected returns when no factors are observable. We find that such portfolios perform well in simulations and in out-of-sample comparisons.

Linear factor-based asset pricing models have solid theoretical foundations in finance following the work of Ross (1976). However, recent empirical evidence questions the usefulness of these models, since they do not seem to fully explain the cross-section of average returns.¹ This article evaluates the usefulness of factor-based models from a different perspective. We assume that asset returns have an exact factor structure, but one of the factors is unobserved. Under this assumption, we develop implications for estimation of expected returns and for portfolio selection. The usefulness of imposing an exact factor structure is assessed by comparing the expected return estimates and portfolios obtained by invoking these implications to their counterparts delivered by standard methods.

When a risk factor is missing from an asset pricing model, the resulting mispricing is related to the covariance matrix. In the absence of such a relation, mispriced assets could be collected to form asymptotic arbitrage opportunities. In this article the relation between the mispricing and the residual

We thank Bruce Grundy, Ravi Jagannathan, Jay Shanken, Rob Stambaugh, an anonymous referee, and the participants at the 1999 NBER Summer Institute, Yale School of Management Asset Pricing Conference, and at the finance seminars at Boston College, Carnegie Mellon University, Columbia University, Copenhagen Business School, University of Alberta, University of Chicago, University of Frankfurt, Vanderbilt University, Washington State University, and Wharton for helpful comments. Research support from the Geewax-Terker Research Program of the Rodney White Center for Financial Research (MacKinlay), the Center for Research in Security Prices, and the Dimensional Fund Advisors (Pástor) is gratefully acknowledged. Address correspondence to Craig MacKinlay, The Wharton School, 3620 Locust Walk, Philadelphia, PA 19104-6367, or e-mail: acmack@wharton.upenn.edu.

¹ See, for example, Lakonishok, Shleifer, and Vishny (1994), MacKinlay (1995), and Daniel and Titman (1997).

covariance matrix is imposed as a restriction in the estimation of the mispricing component of expected return and in portfolio selection.

It is a common practice in finance to estimate the expected return and its mispricing component separately for each asset. In this article the mispricing vector is estimated jointly for a number of assets. The returns on the other assets are related to an asset's mispricing through the link of the mispricing vector to a common residual covariance matrix. Since covariances can be estimated more precisely than means [see Merton (1980)], the restricted mean estimators that incorporate the link could be quite precise. Simulations and empirical analysis confirm that this is indeed the case. In simulations, the expected return estimates that incorporate the link are more precise than both sample means and the James–Stein shrinkage estimators of the mean, unless the sample size is large. With any misspecification, the unbiased sample means become more precise in sufficiently large samples. In empirical analysis, the restricted expected return estimates are economically reasonable and more stable over time than both historical averages and grand means of asset returns. Comparisons of the restricted estimators with their alternative counterparts reveal the usefulness of imposing the factor structure.

The link between the means and covariances has striking implications for portfolio selection. A strong form of the link when no factors are observable leads to a simplification in which the identity matrix can be used in place of the covariance matrix when calculating the tangency portfolio weights. In simulations, the identity covariance matrix is compared to the sample covariance matrix as well as to the true covariance matrix that is used to generate returns. These comparisons are conducted based on the true Sharpe ratios of the optimal portfolios constructed using the various covariance matrixes. Without exception, the identity matrix outperforms not only the sample covariance matrix, but also the true covariance matrix, regardless of what estimator of the mean is used. This result appears to be robust to violations of the underlying assumptions. Using the identity matrix effectively links the estimator of the expected return to the covariance matrix, whereas using the true covariance matrix does not. The gains from using the identity matrix are especially large when a relatively imprecise estimator of expected excess return is used, such as the usual sample mean. Portfolios constructed using the identity covariance matrix also perform very well in out-of-sample empirical analysis.

The article proceeds as follows. Section 1 presents the factor structure used in the analysis, including the link between the mispricing and the residual covariance matrix. Section 2 discusses the application of the “link-based” approach to estimation of expected returns and to portfolio selection. Section 3 contains some simulation results illustrating the usefulness of imposing the link. Section 4 uses actual stock market data to compare the link-based portfolio selection strategy with some traditional strategies. Section 5 concludes.

1. Linear Factor-Based Asset Pricing Models

Consider a linear factor model for the returns on N given assets. Let z_t denote the vector of returns on these N assets in period t , with mean μ and covariance matrix V . Throughout the article, the term “return” on a given asset denotes the asset’s return in excess of the riskless rate, unless the asset is a zero-investment position. For a set of K factor portfolios, the following linear relation between the asset returns and the factor portfolio returns is assumed:

$$z_t = \alpha + Bz_{Kt} + \epsilon_t, \quad (1)$$

$$E[\epsilon_t] = 0, \quad E[\epsilon_t \epsilon_t'] = \Sigma, \quad \text{cov}[z_{Kt}, \epsilon_t] = 0. \quad (2)$$

B is the $N \times K$ matrix of factor sensitivities, z_{Kt} is the K -vector of factor portfolio returns in period t , α is the N -vector of mispricings, and ϵ_t is the N -vector of disturbances. Assume that the factor portfolios are not linear combinations of the N assets, so that Σ is full rank.

An exact K -factor pricing relation implies that each element of the mispricing vector α equals zero. If exact pricing does not hold due to a missing factor, then α is nonzero and embedded in the residual covariance matrix Σ . The relation between α and Σ can be developed using the *optimal orthogonal portfolio* discussed in MacKinlay (1995). This is the unique portfolio constructed from the population of assets which is optimal since it can be combined with the factor portfolios to form the tangency portfolio and is orthogonal to the factor portfolios.² The optimal orthogonal portfolio is denoted by h . The usefulness of portfolio h comes from the fact that when it is added to the K -factor model in Equation (1), the intercept α vanishes and the factor sensitivity matrix B is not altered. The optimality property leads to the intercept vanishing, and the orthogonality condition leads to B being unchanged. Since the mispricing vanishes when h is added to the K factors, h can be thought of as an omitted factor in a linear asset pricing model. Adding the omitted factor,

$$z_t = Bz_{Kt} + \beta_h z_{ht} + u_t, \quad (3)$$

$$E[u_t] = 0, \quad E[u_t u_t'] = \Phi, \quad (4)$$

$$\text{cov}[z_{Kt}, u_t] = 0, \quad \text{cov}[z_{ht}, u_t] = 0, \quad (5)$$

$$E[z_{ht}] = \mu_h, \quad E[(z_{ht} - \mu_h)^2] = \sigma_h^2, \quad (6)$$

where β_h is the N -vector of assets’ sensitivities to the omitted factor h and z_{ht} denotes the return on portfolio h at time t . The link between α and Σ results

² It is assumed that the K factor portfolios cannot be combined to form the tangency portfolio or the global minimum variance portfolio. See Roll (1980) for general properties of orthogonal portfolios. See Hansen and Jagannathan (1997) for a related approach.

from comparing Equations (1) and (3). Taking the unconditional expectations in both equations,

$$\alpha = \beta_h \mu_h, \quad (7)$$

and by equating the variance of ϵ_t with the variance of $\beta_h z_{ht} + u_t$,

$$\begin{aligned} \Sigma &= \beta_h \beta_h' \sigma_h^2 + \Phi \\ &= \alpha \alpha' \frac{1}{s_h^2} + \Phi, \end{aligned} \quad (8)$$

where s_h denotes $\frac{\mu_h}{\sigma_h}$, the Sharpe ratio of portfolio h . Equation (8) links the model mispricing α to the residual covariance matrix Σ , given an assumption about the structure of Φ . For example, the derivation of APT in Ross (1976) assumes that Φ is diagonal. In general, if Σ is unrelated to α , asymptotic arbitrage opportunities exist. MacKinlay (1995) presents examples of such opportunities and illustrates that, in the absence of the link, unrealistic risk-return trade-off positions can be created with even a small number of assets.

2. Expected Returns and Portfolio Selection

This section extends the analysis of Section 1 to develop implications for estimation of expected returns and for portfolio selection. The environments with one factor and with multiple factors are analyzed in separate subsections.

2.1 One factor

A statistical model in which asset returns are generated by one factor has often been employed in finance since the introduction of the diagonal model by Sharpe (1963) and its application by Treynor and Black (1973). In this section, asset returns are assumed to be generated by a one-factor model, but the factor h is assumed to be unobserved.³ This model corresponds to that presented in Equations (3)–(6) for $K = 0$:

$$z_t = \beta_h z_{ht} + u_t, \quad (9)$$

$$E[u_t] = 0, \quad E[u_t u_t'] = \Phi, \quad \text{cov}[z_{ht}, u_t] = 0. \quad (10)$$

Since z_{ht} is unobserved, the model used in the analysis corresponds to that presented in Equations (1) and (2) for $K = 0$:

$$z_t = \alpha + \epsilon_t, \quad (11)$$

$$E[\epsilon_t] = 0, \quad E[\epsilon_t \epsilon_t'] = \Sigma. \quad (12)$$

³ One might think of the true market portfolio as the unobserved factor. In that case, this framework addresses some of the concerns in Roll (1977). However, our assumptions about the structure of Φ are not required to obtain the CAPM.

Following the argument from the end of the previous section, Equations (9)–(12) imply $\Sigma = \alpha\alpha'/s_h^2 + \Phi$. Thus the mean α appears in the covariance matrix of returns. This observation is used in the estimation of expected returns and in portfolio selection.

The usual arbitrage pricing theory assumption in a one-factor model would be that the elements of u_t are sufficiently uncorrelated so that, as the number of assets in a portfolio goes to infinity, the residual variance of the portfolio goes to zero. Such an assumption does not put sufficient structure on Φ for the purposes of this article. This article makes the stronger assumption that Φ is diagonal and proportional to the identity matrix:

$$\Phi = \sigma^2 I. \quad (13)$$

This assumption in combination with the relation in Equation (8) is referred to as a “strong form” of the link between α and Σ . While the assumption may appear strong, this article demonstrates that the payoff from the rigid structure delivered by this assumption typically dominates the cost due to possible misspecification.

A reasonable alternative way of specifying Φ would be to make it diagonal but not necessarily proportional to the identity matrix, such as in Sharpe’s diagonal model. Such an approach would account for potential differences across the N assets in the residual variances from the model that includes the unobserved factor. This could be useful particularly if the N assets are fundamentally different from each other, such as if bond portfolios are mixed with stock portfolios. However, this approach would come at the expense of having to estimate additional $(N - 1)$ parameters when the model is taken to the data. Simulation results not reported here have shown that the benefits of relaxing the assumption of Equation (13) when the N assets belong to the same broad asset class (for example, when all the assets are individual stocks or stock portfolios) are small, if any. Since our subsequent analysis uses assets from the same class, the assumption that Φ is proportional to the identity matrix is maintained throughout the article.

2.1.1 Estimation of expected return. For estimation purposes, the disturbances ϵ_t in Equation (11) are assumed to be normally distributed. Although returns on many assets appear to depart from normality to some extent, this assumption allows convenient estimation of the parameters restricted by the link using the maximum likelihood procedures. The likelihood function is proportional to

$$L(\alpha, s_h^2, \sigma^2 | z_1, \dots, z_T) \propto \left| \frac{\alpha\alpha'}{s_h^2} + \sigma^2 I \right|^{-T/2} \times \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (z_t - \alpha)' \left(\frac{\alpha\alpha'}{s_h^2} + \sigma^2 I \right)^{-1} (z_t - \alpha) \right\}. \quad (14)$$

The maximum likelihood estimators of α , s_h^2 , and σ^2 are obtained by maximizing the joint log-likelihood function with s_h^2 and σ^2 constrained to be positive. Since closed-form solutions for the maximum likelihood estimates are not available, a constrained quasi-Newton numerical procedure is adopted.⁴

Separately, the model is also estimated conditional on a given value of s_h . From a Bayesian perspective, conditioning on s_h is an extreme case of having prior information about the maximum Sharpe ratio. The corresponding results will therefore be informative about the value of having such prior information. To some extent, the value of s_h can be viewed as the degree of confidence in the K -factor pricing model. A dogmatic belief in the model, which rules out mispricing, corresponds to $s_h \rightarrow 0$. The other extreme, $s_h \rightarrow \infty$, can be viewed as complete skepticism about the model, combined with the belief that the maximum Sharpe ratio in the population of assets can be arbitrarily large. A different way of incorporating a degree of belief in an asset pricing model is proposed by Pástor and Stambaugh (1999) in estimation of expected returns, and by Pástor (2000) in portfolio selection.

In simulations and in empirical analysis, expected return estimates and portfolios obtained from the model restricted by the link $\Sigma = \alpha\alpha'/s_h^2 + \sigma^2 I$ are compared to alternative estimates and portfolios that do not impose the restriction. Such a comparison facilitates judgment of the gains from incorporating the factor structure.

2.1.2 Portfolio selection. In a mean-variance framework with a riskless asset, the optimal portfolio of the risky assets, the “tangency portfolio,” is the portfolio with the maximum Sharpe ratio among all portfolios of the N assets. Let x_{N^*} denote the N vector of the tangency portfolio weights. A standard result from mean-variance mathematics is that

$$x_{N^*} = (\iota' V^{-1} \mu)^{-1} V^{-1} \mu, \quad (15)$$

where ι is a conforming vector of ones. Recall from Equations (11) and (12) that, in the model with no observed factors, the mean μ of asset returns z_t is equal to α and the covariance matrix V is equal to Σ . Hence,

$$x_{N^*} = (\iota' \Sigma^{-1} \alpha)^{-1} \Sigma^{-1} \alpha. \quad (16)$$

In simulations and in empirical analysis, the optimal weights are computed as in Equation (16), using different estimators of α and Σ . Given the structure of Σ from Equation (8), the inverse of Σ equals [see Morrison (1990, p. 69)]

$$\Sigma^{-1} = \left(\frac{\alpha\alpha'}{s_h^2} + \Phi \right)^{-1} = \Phi^{-1} - \frac{1}{s_h^2 + \alpha'\Phi^{-1}\alpha} \Phi^{-1} \alpha\alpha' \Phi^{-1}. \quad (17)$$

⁴ The actual estimation uses Matlab’s constrained minimization routine *constr*, which uses a sequential quadratic programming method. For a detailed exposition of this method, see Matlab’s Optimization Toolbox manual.

Substituting the inverse of Σ into Equation (16), the weight vector simplifies to

$$x_{N^*} = (\iota' \Phi^{-1} \alpha)^{-1} \Phi^{-1} \alpha. \quad (18)$$

It is interesting that since the mean α is embedded in the covariance matrix Σ , the residual covariance matrix Φ replaces the total covariance matrix in computing the tangency portfolio weights. When the link between α and Σ is imposed in the strong form $\Phi = \sigma^2 \mathbf{I}$, the restricted tangency portfolio weight vector is

$$x_{N^*} = (\iota' \alpha)^{-1} \alpha. \quad (19)$$

This simple result is striking. Imposing one-factor structure with an unobserved factor implies that the tangency portfolio weights are proportional to the vector of mean returns. In other words, imposing a strong form of the link is equivalent to using an identity matrix as a covariance matrix in portfolio selection.

There are two possible sources of gains from using the restricted estimator of the tangency portfolio weight vector relative to an unrestricted estimator. First, the restricted estimator of the mean is likely to be more precise due to including the information in the covariances. Second, the link between α and Σ leads to the simplified expression in Equation (19). One need not use the restricted estimator of α to realize the gains from this second source. Any estimator of the mean can be embedded in the covariance matrix, leading to the simplification in Equation (19). Hence plugging in the identity matrix for the covariance matrix in Equation (16) links together the mean and the covariance matrix for any estimator of α . The importance of both sources of gains is illustrated in simulations in Section 3.

2.2 Multiple factors

In the multifactor case, an exact $(K + 1)$ -factor pricing relation with K observed factors is examined. The analysis is presented for the two-factor case ($K = 1$), but the extension for $K > 1$ is straightforward. Let p denote the observed factor portfolio, and z_{pt} denote its return at time t . The two-factor linear model is

$$z_t = \alpha + \beta z_{pt} + \epsilon_t \quad (20)$$

$$E[\epsilon_t] = 0, \quad E[\epsilon_t \epsilon_t'] = \Sigma = \alpha \alpha' / s_h^2 + \Phi. \quad (21)$$

Recall that the second factor can be thought of as the optimal orthogonal portfolio h that annihilates α upon its addition to the model.

2.2.1 Estimation of expected return. The expected return for the N assets is

$$\mu = \alpha + \beta\mu_p. \quad (22)$$

The expected return has two components, the mispricing vector with respect to the K -factor model, α , and the factor-related component, $\beta\mu_p$. The link between the mispricing and the residual covariance matrix can be used in estimating α . As in Section 2.1.1, by imposing a rigid structure on Φ as in Equation (13), maximum likelihood can be employed to estimate the model parameters conditional on the returns on the factor portfolio p . The maximum likelihood estimate of α is an estimate of the mispricing component of the expected return. This estimate can be combined with an estimate of the factor-related component to obtain an estimate of the expected return.

2.2.2 Portfolio selection. Suppose that the set of investable assets consists of N assets and $K = 1$ traded factor portfolios. The objective of mean-variance portfolio selection is to find the tangency portfolio of the $N + 1$ assets. The process of choosing the tangency portfolio can be broken down into two stages, as pointed out by Treynor and Black (1973) and Gibbons, Ross, and Shanken (1989), among others. In the first stage, the so-called active portfolio of the N assets is constructed. In the second stage, this active portfolio is combined with the factor portfolio to form the tangency portfolio.⁵ In subsequent analysis, we employ a modified version of this active portfolio, which is uncorrelated with the factor portfolio. The modified active portfolio can be viewed as the optimal orthogonal portfolio in the universe of the $N + 1$ assets. The weights in the active portfolio will be shown to depend on the mispricing and not on the mean or variance of the observed factor. The optimal mix of the active and factor portfolios depends on the factor portfolio's mean and variance.

Let x_{N+} denote the $(N + 1)$ vector of weights in the tangency portfolio, whose first N elements represent the weights in the N assets and the $(N + 1)^{st}$ element represents the weight in the factor portfolio. It is shown in the appendix that

$$x_{N+} = c_1 \begin{bmatrix} \Sigma^{-1}\alpha \\ -\beta'\Sigma^{-1}\alpha + \frac{\mu_p}{\sigma_p^2} \end{bmatrix}, \quad (23)$$

where c_1 is a normalizing constant such that $\iota'x_{N+} = 1$. When the structure $\Sigma = \alpha\alpha'/s_h^2 + \Phi$ from Section 1 is incorporated and substituted into

⁵ When more than one traded factor portfolio is available, the second stage splits into two steps. One step is choosing the optimal portfolio of the factor portfolios and the second step is choosing the mix of that optimal portfolio with the active portfolio.

Equation (23), the tangency portfolio weights can be written as

$$x_{N^+} = c_1 \begin{bmatrix} c_2 \Phi^{-1} \alpha \\ -c_2 \beta' \Phi^{-1} \alpha \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ c_1 \frac{\mu_p}{\sigma_p^2} \end{bmatrix}, \quad (24)$$

where $c_2 = 1/(1 + \alpha' \Phi^{-1} \alpha / s_h^2)$.⁶ Consider an active portfolio a , whose weights are given by

$$x_a = c_3 \begin{bmatrix} \Phi^{-1} \alpha \\ -\beta' \Phi^{-1} \alpha \end{bmatrix}, \quad (25)$$

where c_3 is a normalizing constant such that $1'x_a = 1$. From Equation (24), the tangency portfolio of the $(N + 1)$ assets can be formed by combining portfolio a and portfolio p . (Recall that portfolio p is the $(N + 1)^{st}$ asset.) Using the fact that portfolio a is uncorrelated with portfolio p (since its beta equals zero, i.e., $[\beta' \ 1]x_a = 0$), the tangency portfolio is formed with the weights $\frac{\mu_a/\sigma_a^2}{\mu_a/\sigma_a^2 + \mu_p/\sigma_p^2}$ in a and $\frac{\mu_p/\sigma_p^2}{\mu_a/\sigma_a^2 + \mu_p/\sigma_p^2}$ in p . Furthermore, because a and p are uncorrelated, the squared Sharpe ratio of the tangency portfolio is the sum of the squared Sharpe ratios of a and p .

The active portfolio in Equation (25) does not yet reflect the restrictions imposed by exact factor pricing, since Φ is unrestricted. Restricting Φ to be proportional to the identity matrix simplifies the active portfolio weights:

$$x_{a_r} = c_4 \begin{bmatrix} \alpha \\ -\beta' \alpha \end{bmatrix}, \quad (26)$$

where c_4 is a normalizing constant such that $1'x_{a_r} = 1$. The weights on the N assets are simply proportional to the mispricing vector. The weight on the factor portfolio makes the active portfolio zero-beta with respect to the factor.

As mentioned earlier, the modified active portfolio a can be viewed as the optimal orthogonal portfolio. This portfolio can be added to the K observed factors and cause the mispricing to disappear. As shown in Equation (26), a 's weights in the N assets with restricted Φ are proportional to the mispricing vector. Combining the two observations implies that empirically constructed factor portfolios whose weights in the N assets are approximately proportional to α should do a good job in bringing the model mispricing close to zero. An example of such empirically constructed factors are the Fama–French SMB and HML factor portfolios described earlier. Both of

⁶ Inverting Σ and postmultiplying the result by α gives $\Sigma^{-1}\alpha = \left[\frac{1}{1 + \alpha' \Phi^{-1} \alpha / s_h^2} \right] \Phi^{-1} \alpha = c_2 \Phi^{-1} \alpha$, where c_2 denotes the fraction in square brackets.

these portfolios essentially assume long positions in stocks with positive α estimates and short positions in stocks with negative α estimates (where α s are estimated from the market model), so when added as factors to the market portfolio, they bring the mispricing closer to zero than the CAPM.

The previously noted two possible sources of gains from using the exact factor pricing restrictions for portfolio selection are again present. First, the restricted estimator of the mispricing vector α could be more precise. Second, the link between α and Σ leads to the simplified expression in Equation (26).

3. Simulation Results

This section presents a simulation experiment designed to evaluate the potential gains from linking the means to the covariances. In a simulated environment, the true mean and covariance matrix are known, which is helpful in assessing the usefulness of their alternative estimators. The analysis with one unobserved factor is applied in two environments. In the first environment, returns are generated by a single factor, as assumed in the estimation. This environment provides a measure of the maximum gains from imposing the $\alpha - \Sigma$ link. In the second environment, returns are generated by two factors, but the estimation is again conducted assuming only one unobserved factor. This environment illustrates the benefits of using the restricted framework under a misspecified model.

3.1 Environment I: One-factor return-generating model

Returns on $N = 10$ assets are generated from a one-factor model:

$$z_t = \beta_h z_{ht} + u_t, u_t \sim N(0, \sigma^2 I_N), z_{ht} \sim N(\mu_h, \sigma_h^2), \quad t = 1, \dots, T. \quad (27)$$

The parameters of the simulated environment are specified to correspond approximately to monthly returns. The elements of β_h are evenly spread between 0.5 and 1.5, and $\sigma = (20/\sqrt{12})\%$. The moments of the factor portfolio returns are set equal to $\mu_h = (8/12)\%$ and $\sigma_h = (16/\sqrt{12})\%$, so the Sharpe ratio of the factor portfolio h is $s_h = 0.1443$ (or 0.5 annualized). One thousand samples are generated for three values of T : 60, 120, and 360. The estimation results are reported in Table 1 for expected returns and in Table 2 for portfolio selection.

3.1.1 Expected return. Four estimators of the expected return vector α are calculated for each asset. The first estimator is the (unrestricted) sample average return, denoted by a subscript u . The second estimator is the James–Stein estimator, subscripted by JS . Following Jorion (1986), the James–Stein estimator of the expected return vector is computed as

$$\hat{\alpha}_{JS} = \hat{w} \hat{\alpha}_{GMV} + (1 - \hat{w}) \hat{\alpha}, \quad (28)$$

Table 1
Simulation results for expected returns in a model with one unobserved factor

	$T = 60$	$T = 120$	$T = 360$
$\overline{\hat{\alpha}_u - \alpha}$	0.0018	-0.0223	0.0112
$\left\{ \overline{(\hat{\alpha}_u - \alpha)^2} \right\}^{\frac{1}{2}}$	0.9776	0.7006	0.3983
$\overline{\hat{\alpha}_{JS} - \alpha}$	-0.1893	-0.1913	-0.1536
$\left\{ \overline{(\hat{\alpha}_{JS} - \alpha)^2} \right\}^{\frac{1}{2}}$	0.7064	0.5245	0.3451
$\overline{\hat{\alpha}_{r1} - \alpha}$	0.0828	0.0147	0.0179
$\left\{ \overline{(\hat{\alpha}_{r1} - \alpha)^2} \right\}^{\frac{1}{2}}$	0.4436	0.3732	0.2598
$\overline{\hat{\alpha}_{r2} - \alpha}$	-0.0044	-0.0004	-0.0006
$\left\{ \overline{(\hat{\alpha}_{r2} - \alpha)^2} \right\}^{\frac{1}{2}}$	0.1308	0.0921	0.0529

One thousand different samples of length T of returns on $N = 10$ assets are generated from a one-factor model

$$z_t = \beta_h z_{ht} + u_t, \quad u_t \sim N(0, \sigma^2 I_N), \quad z_{ht} \sim N(\mu_h, \sigma_h^2), \quad t = 1, \dots, T,$$

where the elements of β_h are spread evenly between 0.5 and 1.5, $\sigma = 20\%$, $\mu_h = 8\%$, and $\sigma_h = 16\%$, all annualized. The true expected asset returns are denoted by $\alpha = \beta_h \mu_h$, and $s_h \equiv \mu_h / \sigma_h$ denotes the factor's Sharpe ratio.

$\hat{\alpha}_s - \alpha$ denotes the grand mean across simulations and across N assets of the deviations of the expected return estimates from the true expected returns. $\left\{ \overline{(\hat{\alpha}_s - \alpha)^2} \right\}^{\frac{1}{2}}$ denotes the square root of the grand mean across simulations and across N assets of the squared deviations of the expected return estimates from the true expected returns. The subscript s takes on values from the set ($u, JS, r1, r2$), where u stands for the unrestricted estimate, JS stands for the James–Stein estimate, $r1$ stands for the restricted estimate when the true value of s_h is unknown, and $r2$ stands for the restricted estimate when s_h is known. All values are reported in percentage points per month.

where $\hat{\alpha}$ is the vector of sample means of the assets' returns, $\hat{\alpha}_{GMV}$ is the average return on the global minimum variance portfolio,

$$\hat{\alpha}_{GMV} = \frac{\iota' \hat{V}^{-1} \hat{\alpha}}{\iota' \hat{V}^{-1} \iota}, \quad (29)$$

and

$$\hat{w} = \min \left(1, \frac{(N-2)/T}{(\hat{\alpha} - \hat{\alpha}_{GMV} \iota)' \hat{V}^{-1} (\hat{\alpha} - \hat{\alpha}_{GMV} \iota)} \right), \quad (30)$$

where $\hat{V}$ is the sample covariance matrix of returns. In repeated samples, this shrinkage estimator has lower risk than the sample mean.⁷ The third estimator is the restricted estimator $r1$, obtained by maximizing the likelihood function in Equation (14). The fourth estimator is obtained by maximizing the likelihood conditional on the true value $s_h = 0.1443$, and denoted by $r2$.⁸

⁷ For more details, see the references in Jorion (1986). In an earlier version of the article, we reported results for a James–Stein estimator that shrinks the sample means $\hat{\alpha}$ to their grand mean $\iota' \hat{\alpha} / \iota'$. Those results are similar to the results reported here.

⁸ In some samples, the sample mean of the factor returns is negative, although the population mean is positive. In such cases, the restricted mean return estimates are predominantly negative across assets, and the negative mean estimates of the high-mean assets (those with high factor loadings) are larger in absolute value than those of the low-mean assets. To handle this problem without using unobserved information, the signs of the restricted estimates are reversed when their sum across assets is negative (i.e., when the expected return on the estimated global minimum variance portfolio is negative). The sign reversal yields more sensible expected

For each estimator, measures of bias and precision are presented. The bias is calculated as the estimate's average deviation from the true expected return. The precision is calculated as the root mean square error (RMSE), the square root of the average squared deviation. All averages are grand means, taken across the 1,000 samples and across the $N = 10$ assets.

Table 1 reports the results for expected returns, all on a monthly basis. The average biases are close to zero, except for the James–Stein estimator, whose bias is -0.19% with $T = 60$ and $T = 120$. The mild bias is not surprising, since all expected returns are forced to be the same, while they are not the same in the population. The RMSE varies substantially across the four estimators, and the variation is most dramatic for $T = 60$. The RMSE for the unrestricted estimate is 0.98% . The James–Stein estimator reduces the RMSE to 0.71% , and the first restricted estimator to 0.44% . The RMSE is further reduced to 0.13% when the value of s_h is known. The reason behind this reduction is that s_h^2 is difficult to estimate precisely, since it enters the likelihood function in Equation (14) only by dividing $\alpha\alpha'$, which is also being estimated. Estimation error in s_h^2 can result in multiplicative estimation error in the estimates of α . The reduction in RMSE becomes less pronounced as T increases, since sample mean is an unbiased estimator, unlike the restricted estimators. However, some reduction is still present for $T = 360$: the RMSE of the sample mean is 0.40% , whereas that of the second restricted estimator is only 0.05% .

In an earlier version of the article, the simulation exercise was also repeated with $N = 30$ assets, leading to the same conclusions. Using the information in the covariances increases the precision of the expected return estimates, especially for smaller values of T . As T increases, the gains from restricting the parameters are smaller due to the unbiasedness of the unrestricted estimator. Additional information about the Sharpe ratio of the unobserved factor is very helpful in increasing the precision of the expected return estimates: for $T = 60$, the mean square error is on average reduced by a factor of about 50.

3.1.2 Portfolio selection. Portfolio selection is examined next, using the same set of simulated samples. Given the estimates $\tilde{\mu}$ for the expected return vector and $\tilde{V}$ for the covariance matrix, the vector of tangency portfolio weights can be calculated as

$$\tilde{x} = (\iota' \tilde{V}^{-1} \tilde{\mu})^{-1} \tilde{V}^{-1} \tilde{\mu}. \quad (31)$$

In a simulated environment, it is natural to compare portfolios based on their true Sharpe ratios. Let μ and V denote the true mean and covariance matrix,

return estimates. The occurrence of the simulated samples with the sign reversal is low, at most 9% with $T = 60$, so the impact of the sign reversal on the results reported in Tables 1 and 3 is low. There is no impact at all on the portfolio selection results in Tables 2 and 4, since reversing the signs of all mean returns has no effect on the tangency portfolio weights.

respectively. The true expected excess return of portfolio $\tilde{x}$ is $\mu(\tilde{x}) = \tilde{x}'\mu$, and the true standard deviation is $\sigma(\tilde{x}) = [\tilde{x}'V\tilde{x}]^{\frac{1}{2}}$. The true Sharpe ratio is calculated as

$$\tilde{r} = \frac{\mu(\tilde{x})}{\sigma(\tilde{x})}. \quad (32)$$

Five different values are considered for the expected return—the true value, the unrestricted estimate u (the sample mean), the James–Stein estimate JS , the restricted estimate $r1$, and the restricted estimate $r2$. Five different values are used for the covariance matrix—the true value, the unrestricted estimate u (the sample covariance matrix), the restricted estimate $r1$, the restricted estimate $r2$, and the identity matrix. Recall that using the identity matrix corresponds to imposing the mean–covariance link. For restricted estimates of the mean, using the identity matrix is equivalent to using the restricted covariance matrix. For the true mean, using the identity is the same as using the true covariance matrix, since the true covariance matrix is chosen such that the simplification of Equation (18) into Equation (19) applies.

For each combination of the mean vector and the covariance matrix, Table 2 reports the average (monthly) true Sharpe ratios across the 1,000 simulations. Below each average, the standard deviation of the 1,000 ratios is reported. Panel A contains the results for $T = 60$. The value of 0.1351 in the top left corner of the panel is the true maximum Sharpe ratio for the 10 assets, calculated using the true mean and covariance matrix. With the true mean, the choice of the covariance matrix has a fairly small effect on the tangency portfolio's true Sharpe ratio. However, when the mean is estimated, the choice of the covariance matrix is much more important. In general, using the sample mean ($\hat{\alpha}_u$) results in poorly behaved portfolios, with average Sharpe ratios of about 0.04. However, when the identity matrix is used as the covariance matrix, the average Sharpe ratio for the sample mean is much higher, 0.1130. This is a striking result—with the mean return estimated as the sample average, using the identity matrix is far better than using the true covariance matrix. This finding is also observed for the other expected return estimates. The James–Stein estimator outperforms the sample mean for every covariance matrix, and the highest Sharpe ratio (of 0.1269) is obtained using the identity matrix. The restricted estimates both outperform the James–Stein estimator for every covariance matrix, and the Sharpe ratio (of 0.1349) for the identity matrix is very close to the true value.

The striking result that the identity matrix outperforms not only the sample covariance matrix but also the true covariance matrix is consistent with the importance of sample interaction between the means and covariances. If the mean α is replaced by an arbitrary estimator $\hat{\alpha}$, it is important to embed $\hat{\alpha}$ in the estimator of the covariance matrix and use $\hat{\Sigma} = \hat{\alpha}\hat{\alpha}'/s_h^2 + \hat{\sigma}^2I$. When $\hat{\Sigma}$ is inverted as in Equation (17) and postmultiplied by $\hat{\alpha}$, the resulting tangency portfolio weights are proportional to $\hat{\alpha}$ [see Equation (19)]. In other words,

Table 2
Simulation results for true Sharpe ratios of tangency portfolios in a model with one unobserved factor

	Σ	$\hat{\Sigma}_u$	$\hat{\Sigma}_{r1}$	$\hat{\Sigma}_{r2}$	I_N
Panel A: $T = 60$					
α	0.1351	0.1243	0.1284	0.1284	0.1351
	0.0000	0.0048	0.0024	0.0025	0.0000
$\hat{\alpha}_u$	0.0422	0.0398	0.0439	0.0396	0.1130
	0.0383	0.0369	0.0390	0.0406	0.0347
$\hat{\alpha}_{JS}$	0.0767	0.0699	0.0808	0.0781	0.1269
	0.0367	0.0349	0.0351	0.0383	0.0219
$\hat{\alpha}_{r1}$	0.1236	0.1341	0.1349	—	0.1349
	0.0052	0.0011	0.0001	—	0.0001
$\hat{\alpha}_{r2}$	0.1234	0.1342	—	0.1349	0.1349
	0.0053	0.0010	—	0.0001	0.0001
Panel B: $T = 360$					
α	0.1351	0.1334	0.1336	0.1336	0.1351
	0.0000	0.0008	0.0006	0.0007	0.0000
$\hat{\alpha}_u$	0.0845	0.0832	0.0849	0.0839	0.1312
	0.0289	0.0292	0.0294	0.0296	0.0139
$\hat{\alpha}_{JS}$	0.0997	0.0980	0.1000	0.0996	0.1337
	0.0219	0.0230	0.0228	0.0211	0.0043
$\hat{\alpha}_{r1}$	0.1331	0.1350	0.1350	—	0.1350
	0.0009	0.0001	0.0000	—	0.0000
$\hat{\alpha}_{r2}$	0.1330	0.1350	—	0.1350	0.1350
	0.0010	0.0001	—	0.0000	0.0000

One thousand different samples of length T of returns on $N = 10$ assets are generated from a one-factor model

$$z_t = \beta_h z_{ht} + u_t, u_t \sim N(0, \sigma^2 I_N), z_{ht} \sim N(\mu_h, \sigma_h^2), t = 1, \dots, T,$$

where the elements of β_h are spread evenly between 0.5 and 1.5, $\sigma = 20\%$, $\mu_h = 8\%$, and $\sigma_h = 16\%$, all annualized. The true mean and covariance matrix of the asset returns are denoted by $\alpha = \beta_h \mu_h$ and $\Sigma = \beta_h \beta_h' \sigma_h^2 + \sigma^2 I_N$, respectively, their estimates are denoted by $\hat{\alpha}_s$ and $\hat{\Sigma}_s$, and $s_h \equiv \mu_h / \sigma_h$ denotes the factor's Sharpe ratio. The subscript s takes on values from the set $(u, JS, r1, r2)$, where u stands for the unrestricted estimate, JS stands for the James–Stein estimate, $r1$ stands for the restricted estimate when the true value of s_h is unknown, and $r2$ stands for the restricted estimate when s_h is known. The true Sharpe ratios corresponding to $\hat{\alpha}_i$ and $\hat{\Sigma}_j$ are computed as

$$\frac{\hat{w}_{ij}' \alpha}{(\hat{w}_{ij}' \Sigma \hat{w}_{ij})^{1/2}},$$

where the estimated optimal weight vector is $\hat{w}_{ij} = (t' \hat{\Sigma}_j^{-1} \hat{\alpha}_i)^{-1} \hat{\Sigma}_j^{-1} \hat{\alpha}_i$. The values reported below the true monthly Sharpe ratios are their standard deviations across simulations.

embedding a mean estimator (restricted or not) in the covariance matrix is equivalent to using an identity covariance matrix in computing the tangency portfolio. Using the true covariance matrix $\Sigma = \alpha \alpha' / s_h^2 + \sigma^2 I$ does not link the mean estimator with the covariance matrix, which reduces the true Sharpe ratio, as shown in Table 2.

Note that the results for both restricted estimators are almost identical. This contrasts with the different performance of $r1$ and $r2$ in the estimation of expected return. The reason is that dividing a vector of expected returns by a scalar has no effect on tangency portfolio weights. As mentioned earlier, s_h^2 enters the likelihood function only by dividing $\alpha \alpha'$. The estimation error in s_h^2 divides all elements of α by approximately the same scalar, which has a very small impact on the optimal weights. Since α enters the likelihood also without being divided by s_h^2 , the results for $r1$ and $r2$ could potentially differ, but the differences are very small throughout the article. Hence the knowledge of the Sharpe ratio of the optimal orthogonal portfolio is not

very useful in portfolio selection without observed factors. Moreover, since portfolio selection does not seem to depend on s_h^2 , the difficulty in estimating s_h^2 has no effect on the interesting portfolio implications of the α - Σ link.

Panel B of Table 2 contains portfolio selection results for $T = 360$. The conclusions are similar to those from Panel A, except that the unrestricted estimate shows some improvement. This is not unexpected, since the sample mean is unbiased and its variance declines as T increases. The James–Stein estimator again outperforms the unrestricted estimator, and the highest Sharpe ratios are again achieved by the restricted estimators. As in Panel A, all expected return estimators perform best with the identity matrix, even better than with the true covariance matrix.

In an earlier version of the article, the simulation exercise has also been repeated with $N = 30$ and $N = 80$ assets. With the larger values of N , the true Sharpe ratios for the unrestricted mean estimate are even lower, and the benefits from using the identity matrix are higher than for $N = 10$. Otherwise the larger values of N lead to the same conclusion. There appears to be a substantial payoff in portfolio selection from embedding the means in the covariance matrix, especially when the mean is estimated poorly.

3.2 Environment II: Two-factor return-generating model

The purpose of analyzing the second environment is to investigate the robustness of the one-factor analysis in a misspecified model. The estimation is conducted assuming one underlying factor, although the returns on $N = 10$ assets are simulated from a two-factor model:

$$z_t = \alpha + \beta z_{pt} + \epsilon_t, \epsilon_t \sim N(0, \Sigma), z_{pt} \sim N(\mu_p, \sigma_p^2), t = 1, \dots, T. \quad (33)$$

Recall that the second factor can be thought of as the optimal orthogonal portfolio h that causes α to disappear upon its addition to the model. Since $\Sigma = \alpha\alpha'/s_h^2 + \sigma^2 I$, the mean and covariance matrix of returns are

$$\mu = \alpha + \beta\mu_p \quad \text{and} \quad V = \frac{\alpha\alpha'}{s_h^2} + \beta\beta'\sigma_p^2 + \sigma^2 I. \quad (34)$$

The parameters of the simulation are specified to correspond approximately to monthly returns. The elements of α are evenly spread between $-(5/12)\%$ and $(5/12)\%$, creating an annualized spread of 10%. The elements of β are evenly spread between 0.5 and 1.5. The elements of α and β are randomly paired and the pairs are held fixed across the simulations. The moments of the factor portfolio p are set to $\mu_p = (8/12)\%$ and $\sigma_p = (16/\sqrt{12})\%$, so that the annualized Sharpe ratio of portfolio p is 0.5. Also, $s_h = 0.0953$ (or 0.33 annualized) and $\sigma = (20/\sqrt{12})\%$. The Sharpe ratios of the factor h underlying α and of portfolio p imply that the annualized maximum Sharpe ratio for this simulated environment is 0.6.⁹ One thousand samples are generated for three

⁹ This value is calculated as $\sqrt{12} \times \sqrt{(0.0953)^2 + (\frac{0.5}{\sqrt{12}})^2}$.

Table 3
Simulation results for expected returns in a model with two unobserved factors

	<i>T</i> = 60	<i>T</i> = 120	<i>T</i> = 360
$\overline{\hat{\mu}_u - \mu}$	0.0075	-0.0076	0.0113
$\{(\hat{\mu}_u - \mu)^2\}^{\frac{1}{2}}$	1.0564	0.7336	0.4247
$\overline{\hat{\mu}_{JS} - \mu}$	-0.1530	-0.1534	-0.1204
$\{(\hat{\mu}_{JS} - \mu)^2\}^{\frac{1}{2}}$	0.7276	0.5432	0.3661
$\overline{\hat{\mu}_{r1} - \mu}$	0.0559	0.0031	-0.0129
$\{(\hat{\mu}_{r1} - \mu)^2\}^{\frac{1}{2}}$	0.5353	0.4698	0.3848
$\overline{\hat{\mu}_{r2} - \mu}$	0.1125	0.1147	0.1146
$\{(\hat{\mu}_{r2} - \mu)^2\}^{\frac{1}{2}}$	0.3736	0.3490	0.3333

One thousand different samples of length *T* of returns on *N* = 10 assets are generated from what is effectively a two-factor model

$$z_t = \alpha + \beta z_{pt} + \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma = \alpha\alpha' / s_h^2 + \sigma^2 I_N), \quad z_{pt} \sim N(\mu_p, \sigma_p^2), \quad t = 1, \dots, T,$$

where the elements of α are spread evenly between -5% and 5%, elements of β are spread evenly between 0.5 and 1.5 and randomly assigned to the elements of α , $\sigma = 20\%$, $\mu_p = 8\%$, and $\sigma_p = 16\%$, all annualized. Above, s_h is the Sharpe ratio of the optimal orthogonal portfolio, which is set equal to 0.33 per year. The values of s_h and $s_p = \mu_p / \sigma_p$ together imply that the Sharpe ratio of the optimal mean variance efficient portfolio equals 0.6 per year. The true expected asset returns are denoted by $\mu = \alpha + \beta \mu_p$.

$\hat{\mu}_s - \mu$ denotes the grand mean across simulations and across *N* assets of the deviations of the expected return estimates from the true expected returns. $\{(\hat{\mu}_s - \mu)^2\}^{\frac{1}{2}}$ denotes the square root of the grand mean across simulations and across *N* assets of the squared deviations of the expected return estimates from the true expected returns. The subscript *s* takes on values from the set (*u*, *JS*, *r1*, *r2*), where *u* stands for the unrestricted estimate, *JS* stands for the James–Stein estimate, *r1* stands for the restricted estimate when the true value of s_h is unknown, and *r2* stands for the restricted estimate when s_h is fixed at 0.6 per year. All values are reported in percentage points per month.

values of *T*: 60, 120, and 360. The estimation results are reported in Table 3 for expected returns and in Table 4 for portfolio selection.

3.2.1 Expected return. Four estimators of the expected return vector μ are considered: the unrestricted estimator *u*, the James–Stein estimator *JS*, the restricted estimator *r1* that incorrectly assumes a one-factor structure, and the restricted estimator *r2* conditional on $s_h = 0.1732$ (or 0.6 annualized) that incorrectly assumes a one-factor structure. Conditioning on that particular value of s_h in the computation of *r2* assumes that the Sharpe ratio of the unobserved factor is equal to the true maximum Sharpe ratio in the two-factor environment.

Table 3 contains the biases and precisions of the alternative expected return estimators, analogous to Table 1. The average biases are again fairly close to zero, no more in absolute value than 0.15% for the James–Stein estimator.¹⁰ The RMSE varies substantially across the estimators. For *T* = 60, the RMSE for the unrestricted estimator is 1.06%, whereas it is only 0.73% for the

¹⁰ An inspection of the biases for each asset reveals that the biases of the restricted estimators can sometimes be substantial. Although the biases are typically small, for several assets, the bias of *r2* exceeds 0.50%. Such a finding is not unexpected, since the model is misspecified. Despite the bias, each asset's RMSE for the restricted estimator is lower than for the unrestricted estimator. Also, the biases of the James–Stein estimators are typically larger than for the restricted estimators.

Table 4

Simulation results for true sharpe ratios of tangency portfolios in a model with two unobserved factors

	V	$\hat{V}_u$	$\hat{V}_{r1}$	$\hat{V}_{r2}$	I_N
Panel A: $T = 60$					
μ	0.1557	0.1433	0.1424	0.1418	0.1472
	0.0000	0.0057	0.0059	0.0061	0.0000
$\hat{\mu}_u$	0.0506	0.0461	0.0652	0.0629	0.1094
	0.0481	0.0484	0.0571	0.0559	0.0486
$\hat{\mu}_{JS}$	0.0845	0.0771	0.0971	0.0990	0.1245
	0.0437	0.0430	0.0533	0.0501	0.0326
$\hat{\mu}_{r1}$	0.1177	0.1286	0.1282	—	0.1282
	0.0182	0.0087	0.0082	—	0.0082
$\hat{\mu}_{r2}$	0.1177	0.1284	—	0.1280	0.1280
	0.0183	0.0088	—	0.0083	0.0083
Panel B: $T = 360$					
μ	0.1557	0.1538	0.1400	0.1401	0.1472
	0.0000	0.0009	0.0026	0.0025	0.0000
$\hat{\mu}_u$	0.1054	0.1037	0.1177	0.1166	0.1413
	0.0326	0.0337	0.0259	0.0261	0.0170
$\hat{\mu}_{JS}$	0.1158	0.1136	0.1293	0.1289	0.1389
	0.0253	0.0277	0.0231	0.0204	0.0139
$\hat{\mu}_{r1}$	0.1275	0.1297	0.1285	—	0.1285
	0.0075	0.0033	0.0032	—	0.0032
$\hat{\mu}_{r2}$	0.1275	0.1297	—	0.1285	0.1285
	0.0075	0.0032	—	0.0031	0.0031

One thousand different samples of length T of returns on $N = 10$ assets are generated from what is effectively a two-factor model

$$z_t = \alpha + \beta z_{pt} + \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma = \alpha\alpha' / s_h^2 + \sigma^2 I_N), \quad z_{pt} \sim N(\mu_p, \sigma_p^2), \quad t = 1, \dots, T,$$

where the elements of α are spread evenly between -5% and 5% , elements of β are spread evenly between 0.5 and 1.5 and randomly assigned to the elements of α , $\sigma = 20\%$, $\mu_p = 8\%$, and $\sigma_p = 16\%$, all annualized. Above, s_h is the Sharpe ratio of the optimal orthogonal portfolio, which is set equal to 0.33 per year. The values of s_h and $s_p = \mu_p / \sigma_p$ together imply that the Sharpe ratio of the optimal mean–variance efficient portfolio equals 0.6 per year. The true mean and covariance matrix of the asset returns are denoted by $\mu = \alpha + \beta\mu_p$ and $V = \beta\beta'\sigma_p^2 + \Sigma$, respectively, and their estimates are denoted by $\hat{\mu}_s$ and $\hat{V}_s$. The subscript s takes on values from the set $(u, JS, r1, r2)$, where u stands for the unrestricted estimate, JS stands for the James–Stein estimate, $r1$ stands for the restricted estimate when the true value of s_h is unknown, and $r2$ stands for the restricted estimate when s_h is fixed at 0.6 per year. The true Sharpe ratios corresponding to $\hat{\mu}_i$ and $\hat{V}_j$ are computed as

$$\frac{\hat{w}'_{ij}\mu}{(\hat{w}'_{ij}V\hat{w}_{ij})^{1/2}},$$

where the estimated optimal weight vector is $\hat{w}_{ij} = (\hat{V}_j^{-1}\hat{\mu}_i)^{-1}\hat{V}_j^{-1}\hat{\mu}_i$. The values reported below the true monthly Sharpe ratios are their standard deviations across simulations.

James–Stein estimator, 0.54% for the restricted estimator $r1$, and 0.37% for the restricted estimator $r2$. As in Table 1, the reduction in RMSE across estimates is less pronounced for larger values of T . For $T = 360$, the RMSE decreases from 0.42% for the unrestricted estimate to 0.33% for the second restricted estimate, and JS has a lower RMSE than $r1$.

In an earlier version of the article, the analysis was repeated with $N = 30$ assets, leading to the same conclusions. Despite the misspecification, using the information in the covariances increases the precision of expected return estimates, especially for smaller values of T . As T grows, the gains from imposing the restrictions disappear since the restricted estimators remain biased, unlike the sample means. Additional information about the maximum Sharpe ratio for the whole environment reduces the RMSE, but the

reduction is small as the bias sometimes increases when that information is incorporated in a misspecified model.

3.2.2 Portfolio selection. Portfolio selection is examined next, using the same set of simulated samples. Table 4 reports the simulation results in the same format as Table 2. Panel A contains the results for $T = 60$. The true maximum Sharpe ratio is 0.1557. In this second environment, using the identity matrix with the true mean is no longer the same as using the true covariance matrix. The true covariance matrix now has a two-factor structure [see Equation (34)], and the simplifications of Equations (18) and (19) occur only with a one-factor structure. Using the sample mean always leads to the lowest Sharpe ratios. This unrestricted estimate performs worst with the sample covariance matrix and the true covariance matrix, and best with the identity matrix. This result deserves emphasis for two reasons. First, even in a misspecified model, imposing the link implied by the one-factor model leads to substantially higher Sharpe ratios. Second, as in the first environment, using the true covariance matrix with the unrestricted mean estimate works poorly. Using the identity covariance matrix, which effectively links the covariance matrix to the sample mean, leads to substantial improvement.

We also observe that shrinking the sample means increases the true Sharpe ratios. The identity matrix is also the best choice for the covariance matrix with the James–Stein estimator of the mean. Since the restrictions are now misspecified, the restricted estimators of the mean, r_1 and r_2 , no longer lead to average Sharpe ratios close to the true maximum. However, their Sharpe ratios still uniformly exceed those obtained using the sample means, and even the James–Stein estimators. The Sharpe ratios are the lowest when the true covariance matrix is used.

Panel B of Table 4 contains portfolio selection results for $T = 360$. Using the identity matrix, the average Sharpe ratio is now higher for the unrestricted estimate than for the restricted estimates. This is not surprising. As mentioned earlier, the unrestricted estimate is unbiased and converges to the true mean as T increases. In all cases, the identity matrix continues to be a reasonable choice for the covariance matrix, despite model misspecification in this two-factor environment. This result highlights the importance of imposing a relation between the mean and the covariance matrix in portfolio selection.

3.3 Summary of simulation evidence

A number of interesting results emerge from the simulations. First, recall that using the identity covariance matrix imposes a strong link between the mean and the covariance matrix in a one-factor model with an unobserved factor. In all scenarios with estimated means, the true Sharpe ratios are higher when the identity matrix is used than when the true covariance matrix is used, even in the presence of misspecification. This powerful result supports an

important economic role of factor models. Second, portfolio selection can be improved when the estimate of the mean incorporates the information in the covariances. This improvement is present in small and medium samples even when the underlying assumptions are violated. In large samples, with any misspecification, the improvement disappears as the variance reduction is more than offset by the bias in the restricted estimates.

The environment analyzed in this subsection illustrates the robustness of the results to model misspecification. Further analysis reveals that the usefulness of linking the mean and the covariance matrix is also robust to other forms of misspecification. The simulations in both environments have been repeated with less structure imposed on the true return-generating covariance matrix of u_t .¹¹ The results in both environments, not reported here to conserve space, are similar to those reported in Tables 1–4. In an earlier version of the article, the simulations in both environments have also been repeated for $N = 30$ and $N = 80$, leading to the same conclusions. Across the simulations, the restricted estimates of the mean are more precise than the unrestricted estimates, and the identity matrix beats the true covariance matrix in portfolio selection for all estimates of the mean. In fact, for the larger values of N , the gains from using the identity covariance matrix are even larger.

4. Empirical Illustrations

This section provides an illustration on actual data of the role that the economic restrictions of exact factor pricing can play in estimation of expected returns and in portfolio selection. Empirical results are presented for two universes of assets: 10 size-sorted portfolios and 30 individual stocks. For each universe, two scenarios are considered—one in which no factors are observed, and the other in which one factor is observed.

4.1 Decile portfolios

The size decile portfolios are value-weighted portfolios of NYSE, AMEX, and NASDAQ stocks sorted by market capitalization, with an equal number of NYSE stocks in each decile. Excess decile returns are obtained from the CRSP for January 1926–December 1996 using the one-month Treasury bill return as the risk-free rate.

A monthly series of expected return estimates and optimal portfolio weights is computed using a “rolling sample approach” for windows of $T = 60$ and $T = 120$ months. The first sample period for the 120-month window is January 1926–December 1935. The decile returns from that period are used

¹¹ Let V_u denote a matrix whose all diagonal elements are equal to σ^2 and all off-diagonal elements are equal to $\sigma^2/3$. In environment I, the true covariance matrix of u_t was set equal to V_u (as opposed to $\sigma^2 I$). In environment II, the true covariance matrix of ϵ_t was set equal to $\alpha\alpha'/s_h^2 + V_u$ (as opposed to $\alpha\alpha'/s_h^2 + \sigma^2 I$).

to estimate expected returns and calculate portfolio weights, and the return on the resulting portfolio is recorded for January 1936. The 120-month window is then rolled forward so that January 1926 is dropped and January 1936 is added to the sample, and the portfolio return is recorded for February 1936. This process is continued until a time series of observations through December 1996 is built. The same approach is used for the 60-month window, except that the January 1931–December 1935 period is used for the initial sample.

Three expected return estimates are generated for each decile: the sample mean, the grand mean of the 10 deciles, and the restricted estimate r_2 . The restricted estimate is computed assuming one unobservable factor for $s_h = 0.1732$ (or 0.6 annualized) and $s_h = 0.2887$ (or 1.0 annualized). Only the restricted estimates conditional on s_h are presented. As previously noted, s_h^2 is difficult to estimate. Since expected return is scaled by s_h^2 , treating s_h^2 as a free parameter leads to occasional jumps in expected return estimates in response to changes in the estimate of s_h^2 . Conditioning on s_h eliminates this problem.

Figures 1 and 2 present the expected return estimates for $T = 60$ for deciles 1 (smallest firms), 4, 7, and 10 (largest firms). Figure 1 contains the results for $s_h = 0.1732$. The sample mean and grand mean estimates are highly variable. As would be expected, the returns on small stocks are more variable than the returns on large stocks. For all deciles, the unrestricted estimates are negative in the early 1940s and in the mid-1970s. The restricted estimates are much smoother. The expected return generally varies between 1 and 2% per month for decile 1, and between 0.5 and 1% per month for decile 10. In general, the estimates are economically reasonable. The tendency for the unrestricted estimates to move together with the restricted estimates appears weak. Using the information in covariances leads to expected return estimates that are quite different from the unrestricted estimates. Also note that the restricted estimates are positively correlated across the deciles, since they are linked to the same covariance matrix.

Figure 2 contains the results for $s_h = 0.2887$. The higher Sharpe ratio for the unobservable factor portfolio implies higher restricted estimates of the mean. These restricted estimates in fact seem to be too high. For example, the expected return for deciles 1 and 4 is often near or above 2% per month. If these expected returns seem too high, then $s_h = 0.2887$ is probably too high, and the true annualized maximum Sharpe ratio is smaller than 1.0. Comparing the restricted and unrestricted estimates leads to the same conclusions as in Figure 1. The restricted estimates are much less variable and do not appear to move together with the unrestricted estimates.

Table 5 presents out-of-sample monthly Sharpe ratios for eight investment strategies, based on the 1936–1996 period as well as on two and three subperiods of equal length. Panel A presents the Sharpe ratios for three passive strategies: holding the value-weighted market index, the equal-weighted

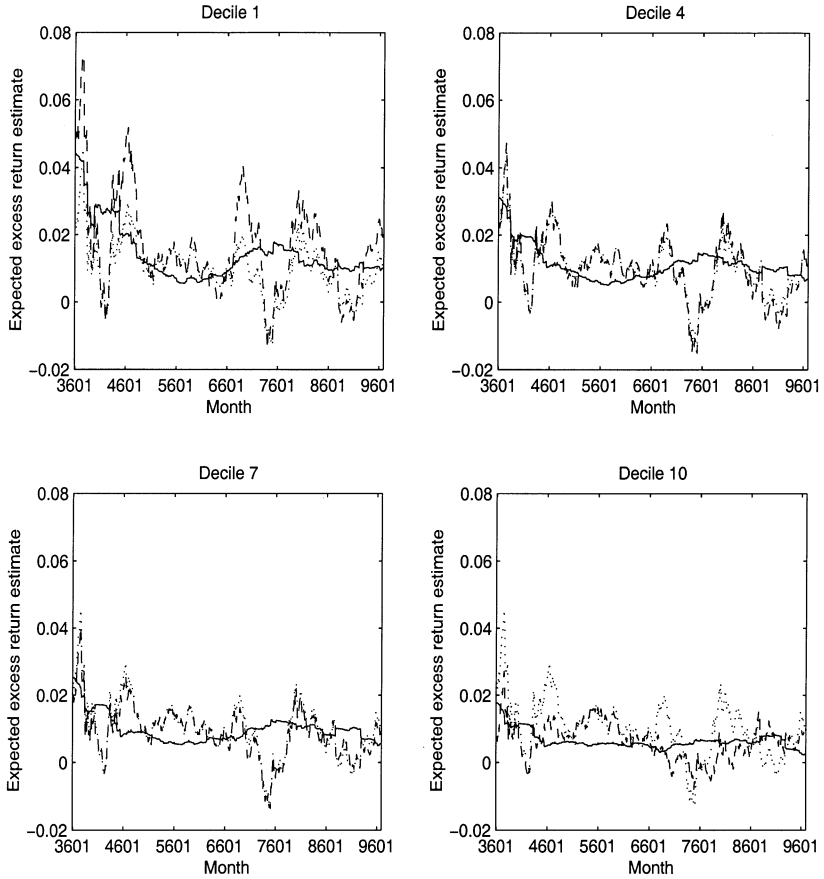


Figure 1

Expected returns for size decile portfolios, $T = 60$, $s_h = 0.1732$

Estimates of expected monthly excess returns for size decile portfolios in January 1936–December 1996 using three different techniques. Decile 1 represents the smallest stocks and decile 10 the largest stocks. The value of s_h corresponds to an annualized population maximum Sharpe ratio of 0.6. The *solid* line corresponds to the *restricted* estimates that link the means to the covariances, the *dashed* line to the *unrestricted* estimates and the *dotted* line represents *grand means* across all 10 deciles.

market index, and an equal-weighted portfolio of the 10 deciles. The Sharpe ratios are similar across the three passive strategies. They range from 0.0784 to 0.2056, with the lowest values in the periods that include the down-market of 1973–1974. Panel B presents the results for portfolios constructed using the unrestricted estimators of the mean and covariance matrix. The unrestricted estimators perform very poorly.¹² The Sharpe ratios are low and often

¹² The fact that using the unrestricted estimators to compute optimal portfolio weights often results in poorly behaved portfolios has been documented and discussed in a number of studies, including Dickinson (1974),

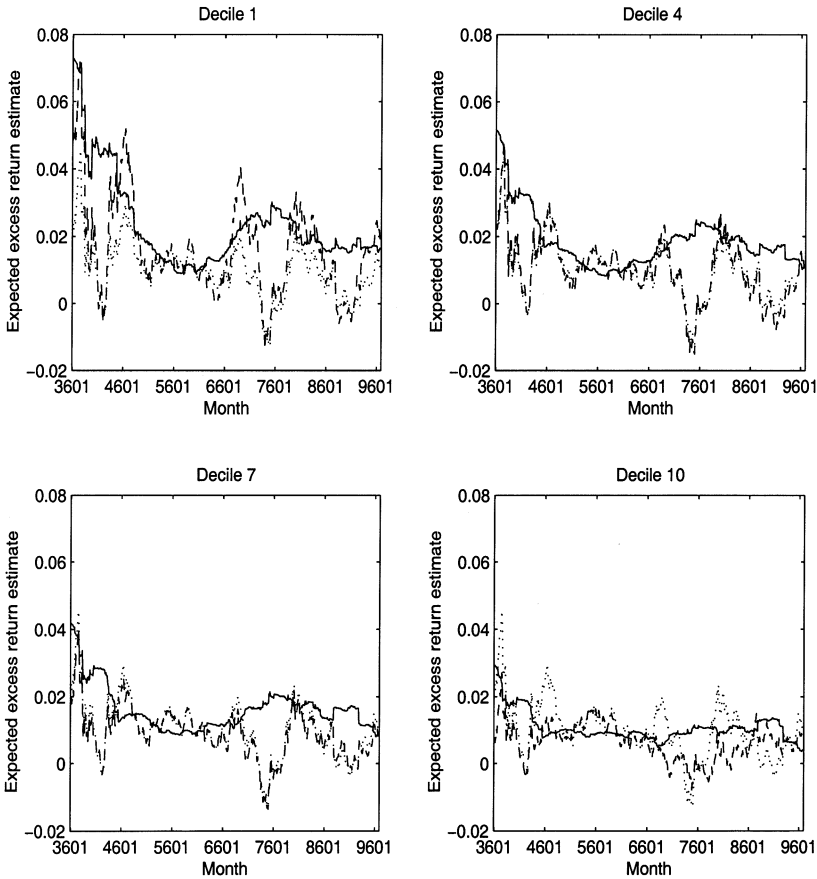


Figure 2
Expected returns for size decile portfolios, $T = 60$, $s_h = 0.2887$
Estimates of expected monthly excess returns for size decile portfolios in January 1936–December 1996 using three different techniques. Decile 1 represents the smallest stocks and decile 10 the largest stocks. The value of s_h corresponds to an annualized population maximum Sharpe ratio of 1.0. The *solid* line corresponds to the *restricted* estimates that link the means to the covariances, the *dashed* line to the *unrestricted* estimates and the *dotted* line represents *grand means* across all 10 deciles.

negative; for the overall period, they are equal to 0.0555 and -0.0356 for estimation windows of 60 and 120 months, respectively.¹³

Panel C presents the results for the unrestricted mean estimator combined with the identity matrix. Recall that using the identity covariance matrix is consistent with imposing the link between the means and covariances.

Jobson and Korkie (1980), Michaud (1989), Best and Grauer (1991), Jorion (1991), Black and Litterman (1992), Green and Hollifield (1992), and Chopra and Ziemba (1993).

¹³ One is cautioned not to attach much significance to the magnitudes of the negative Sharpe ratios. If the mean return is negative, the magnitude of such a ratio increases as the standard deviation decreases, so that the negative ratio decreases as risk decreases.

Table 5
Out-of-sample Sharpe ratios for $N = 10$ size decile portfolios

	1936–1996	1936–1966	1966–1996	1936–1956	1956–1976	1976–1996
A. Passive strategies						
VW Market	0.1486	0.1951	0.1016	0.2056	0.0784	0.1523
EW Market	0.1583	0.1936	0.1209	0.1977	0.1007	0.1743
EW 10	0.1532	0.1917	0.1134	0.1909	0.1036	0.1618
B. Sample mean and covariance matrix						
$T = 60$	0.0555	0.0006	0.0800	−0.0018	0.0764	0.0715
$T = 120$	−0.0356	0.0774	−0.0737	0.0918	0.0368	−0.0914
C. Sample mean with an identity matrix						
$T = 60$	0.1728	0.1919	0.1601	0.1942	0.1258	0.1981
$T = 120$	−0.0349	0.1823	−0.0519	0.1821	−0.0283	−0.0623
D. James–Stein mean with a sample covariance matrix						
$T = 60$	0.0584	−0.0053	0.0842	−0.0219	0.0787	0.0729
$T = 120$	−0.0390	0.0548	−0.0638	0.0587	0.0382	−0.0790
E. James–Stein mean with an identity matrix						
$T = 60$	0.0665	0.1909	0.0108	0.1911	0.1180	−0.0047
$T = 120$	0.1443	0.2082	0.0909	0.2136	0.0873	0.1296
F. Restricted mean and covariance matrix, unobserved factor						
$T = 60$	0.1533	0.1892	0.1156	0.1884	0.1062	0.1631
$T = 120$	0.1532	0.1895	0.1150	0.1889	0.1049	0.1632
G. Restricted mean and covariance matrix, observed factor						
$T = 60, s_h = 0.0289$	0.1253	0.1493	0.0966	0.1528	0.0623	0.1566
$T = 60, s_h = 0.0577$	−0.0332	−0.0484	0.1111	−0.0597	0.0648	0.1674
$T = 120, s_h = 0.0289$	0.1423	0.1839	0.0926	0.1931	0.0859	0.1332
$T = 120, s_h = 0.0577$	0.1281	0.1542	0.0980	0.1694	0.0984	0.1225
H. Modified restricted estimates, observed factor						
$T = 60, s_h = 0.0289$	0.1278	0.1540	0.0969	0.1575	0.0620	0.1588
$T = 60, s_h = 0.0577$	0.1306	0.1448	0.1150	0.1481	0.0691	0.1789
$T = 120, s_h = 0.0289$	0.1427	0.1838	0.0938	0.1925	0.0862	0.1358
$T = 120, s_h = 0.0577$	0.1379	0.1651	0.1058	0.1706	0.1013	0.1398

Out-of-sample monthly Sharpe ratios are computed for several portfolio strategies. Among the passive strategies in panel A, the “VW market” strategy invests in a value-weighted index of all NYSE + AMEX + NASDAQ stocks, the “EW market” strategy invests in an equally weighted index of those stocks, and the “EW 10” strategy invests equal dollar amounts in $N = 10$ size decile portfolios every month. Panels B–H contain the results for strategies that form mean variance efficient portfolios of the decile portfolios using the stated estimates of the mean and covariance matrix of the deciles’ returns. Every month, T most recent returns are used to estimate the return moments and the portfolios are rebalanced on a monthly basis. In panel F, the restricted strategies produce identical results for any value of s_h , the Sharpe ratio of the optimal orthogonal portfolio. Panel H corresponds to the strategy in panel G, except that if the restricted estimates imply extreme weights (i.e., if at least one weight in the portfolio exceeds 50% in absolute value), the weights in the decile portfolios are scaled down such that the largest absolute weight equals 50% and the remaining funds are invested in the VW market defined in panel A.

In all but one case, the Sharpe ratios in panel C are higher than in panel B, supporting the usefulness of imposing the link. Using a 60-month window, the Sharpe ratios in panel C are quite high and in most cases exceed those of the passive portfolios. The Sharpe ratio for the overall period is 0.1728, higher than 0.1583 for the best performing passive portfolio, the equal-weighted market index. However, using a 120-month window, the results are mixed, with negative Sharpe ratios in some periods. Clearly the unrestricted mean estimator can lead to rather unstable optimal weights, even when the covariance matrix is an identity.

Panels D and E present the results for the James–Stein estimator JS of the mean. With the sample covariance matrix, JS does not seem to offer

any improvement over the unrestricted estimator in this sample. With the identity matrix, JS does worse than the sample mean for $T = 60$, but much better than the sample mean for $T = 120$. The identity matrix with JS performs much better than the sample covariance matrix. This confirms our simulation findings that the choice of the covariance matrix is important in portfolio selection and that imposing the mean-variance link by using the identity covariance matrix is useful.

Panel F of Table 5 reports the results for the restricted estimators of the mean and covariance matrix with no observed factors. Recall that using the restricted covariance matrix is equivalent to simply using the identity matrix. As discussed earlier, the portfolio weights are not sensitive to s_h . The restricted portfolio does well, with Sharpe ratios that are always positive and quite high. Comparison with the low Sharpe ratios for the unrestricted estimates in panel B clearly reveals the usefulness of imposing the restrictions. The Sharpe ratios in panel F are always higher than the corresponding ratios in panel B. Moreover, comparison with panels C, D, and E reveals that using the restricted mean in general provides more reasonable portfolios than using either the sample mean or the James–Stein mean, even with the identity covariance matrix.

The performance of the active portfolios constructed using the restricted estimators (panel F) can also be compared to the performance of passive portfolios (panel A). Note that, in the overall period, the active portfolios outperform the VW market. However, the best ex post performance is delivered by holding the EW market. It is important to realize that, under certain circumstances, an EW portfolio can be interpreted as a portfolio that imposes the mean–variance link. Consider the more general specification of the link, in which Φ is restricted to be diagonal but not necessarily proportional to the identity. Recall from Equation (18) that the weights in the optimal portfolio are proportional to $\Phi^{-1}\alpha$. If the expected returns α are positively related and approximately proportional to the residual variances (the diagonal elements of Φ), then the optimal portfolio is approximately equally weighted.¹⁴

The restricted estimation approach is next applied with one observable factor portfolio, the value-weighted market index. The estimates of the mispricing α relative to the one-factor model are presented in Figures 3 and 4, for $T = 60$. The estimates are plotted for $s_h = 0.0289$ and $s_h = 0.0577$, which correspond to the annualized Sharpe ratios for the optimal orthogonal portfolio of 0.1 and 0.2, respectively. For all deciles, the restricted mispricing estimates are much more stable than the unrestricted estimates. However, the restricted estimates “jump” when they cross zero. Conditioning on a nonzero value of s_h implies the existence of mispricing, so all elements of α cannot have values very close to zero. The estimated α for decile 10 is close

¹⁴ This is theoretically plausible. For example, Merton (1987) proposes a model of capital market equilibrium with incomplete information in which expected returns are positively related to residual variances.

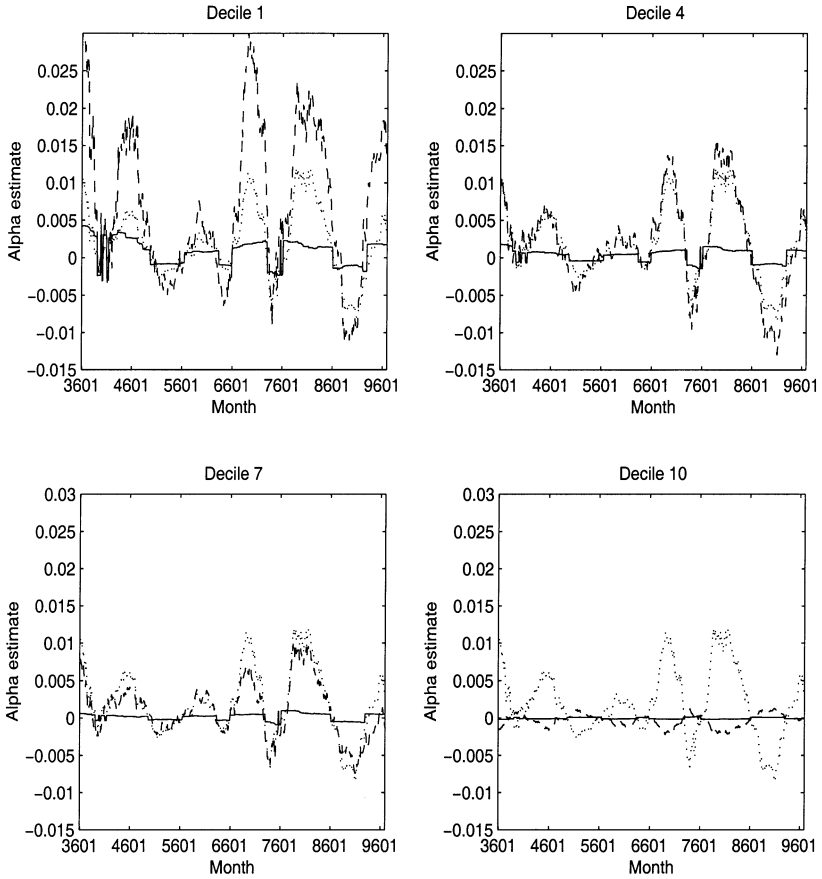


Figure 3

Mispricing for size decile portfolios, $T = 60$, $s_h = 0.0289$

Estimates of monthly mispricing for size decile portfolios in January 1936–December 1996 using three different techniques. Decile 1 represents the smallest stocks and decile 10 the largest stocks. The value of s_h corresponds to the annualized Sharpe ratio of the optimal orthogonal portfolio of 0.1. The *solid* line corresponds to the *restricted* estimates that link the means to the covariances, the *dashed* line to the usual *unrestricted* case and the *dotted* line represents *grand means* across all 10 deciles.

to zero, which could be expected since this large firm decile dominates the value-weighted market index.

In portfolio selection, the value-weighted market is not included among the investable assets, since it is approximately a linear combination of the decile portfolios, and is very highly correlated with the decile containing the largest stocks. Therefore the optimal portfolio computed here is not the orthogonal active portfolio, but rather the tangency portfolio of the 10 assets:

$$x_{[10]} = \frac{1}{c_r} \hat{V}_r^{-1} \hat{\mu}_r, \quad (35)$$

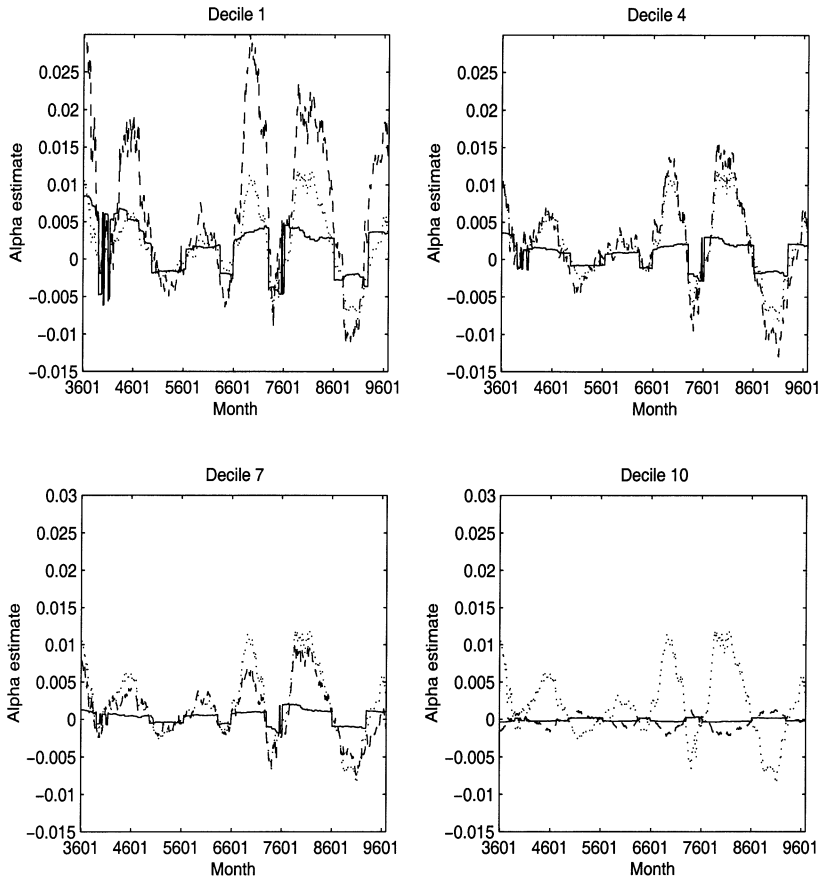


Figure 4
Mispricing for size decile portfolios, $T = 60$, $s_h = 0.0577$
Estimates of monthly mispricing for size decile portfolios in January 1936–December 1996 using three different techniques. Decile 1 represents the smallest stocks and decile 10 the largest stocks. The value of s_h corresponds to the annualized Sharpe ratio of the optimal orthogonal portfolio of 0.2. The *solid* line corresponds to the *restricted* estimates that link the means to the covariances, the *dashed* line to the usual *unrestricted* case and the *dotted* line represents *grand means* across all 10 deciles.

where $\hat{\mu}_r = \hat{\alpha}_r + \hat{\beta}_r \hat{\mu}_p$, $\hat{V}_r = \hat{\alpha}_r \hat{\alpha}_r' / s_h^2 + \hat{\beta}_r \hat{\beta}_r' \hat{\sigma}_p^2 + \hat{\sigma}_r^2 I$, and $c_r = \iota' \hat{V}_r^{-1} \hat{\mu}_r$. The restricted estimates are computed assuming that there is one observable and one unobservable factor. In addition, estimates of the mean $\hat{\mu}_p$ and standard deviation $\hat{\sigma}_p$ of the factor returns are needed. These parameters are estimated by their sample values for the CRSP value-weighted market index from January 1926 through the last month in the estimation window.

Panel G of Table 5 reports the results for the restricted estimators of the mean and covariance matrix with one observed factor. Since the portfolio weights are now sensitive to s_h , the results are reported for $s_h = 0.0289$ as

well as for $s_h = 0.0577$. The Sharpe ratios are much higher than for the unrestricted estimates in Panel B, except for $T = 60$ and $s_h = 0.0577$ in the periods that include the 1936–1956 period. Comparing the Sharpe ratios from Panels F and G leads to the somewhat counterintuitive result that one is better off not using the value-weighted market as the factor portfolio when constructing the optimal portfolio weights. This surprising result is consistent with the importance of the link, given the empirical fact that stock returns have one dominant factor. In Panel F, the dominant factor is missing, so there are substantial benefits from imposing the link. In Panel G, where the important factor has been largely accounted for and only relatively unimportant factors are omitted, the benefits of imposing the link are reduced. Moreover, the value-weighted market may not be a perfect proxy for the single dominant factor. The usefulness of observing the returns on the value-weighted market is further compromised given the imprecision of the estimates of its mean $\hat{\mu}_p$ and variance $\hat{\sigma}_p^2$.

Recall that s_h is the Sharpe ratio of the missing (orthogonal) factor. Hence a large value for s_h implies an important missing factor and could potentially lead to extreme portfolio weights. Technically c_r in Equation (35) can be close to zero, which results in extreme weights. In panel G, this problem is present for $s_h = 0.0577$. Between 1936 and 1996, the absolute value of the maximum weight exceeds 100% in 146 months for $T = 60$ and in 101 months for $T = 120$. For $s_h = 0.0289$, the absolute value of the maximum weight never exceeds 100%.

To investigate the impact of extreme weights, the results of panel G are replicated with one modification. The portfolio weights are adjusted so that no long or short position in any single decile exceeds 50%. In each month for which the largest absolute weight exceeds 50%, the portfolio is modified such that the weight is reduced to 50% with the same sign. The magnitudes of the other decile weights are reduced proportionally. The value-weighted market is added to the portfolio as an extra asset so that the weights sum to one.¹⁵

The results for this modified approach are reported in Panel H. The Sharpe ratios with the modification are generally close to those without the modification, with a few exceptions. The very low Sharpe ratios from panel G for $T = 60$ and $s_h = 0.0577$ have increased substantially. The negative ratios for the periods that include the 1936–1956 period are now positive and in line with the passive indexes. Clearly the results for this period in panel G are driven by some months with extreme weights.

4.2 Individual Stocks

A universe of 30 individual stocks is selected using the recent composition of the Dow Jones Industrial Index. In order to be included in the sample, a stock

¹⁵ The simple ad-hoc modification proposed here is only one of many ways of imposing portfolio constraints. For example, a reasonable alternative would be to impose a restriction on the sum of squared weights.

Table 6
Out-of-sample Sharpe ratios for $N = 30$ Dow Jones stocks

	1957–1996	1957–1976	1977–1996	1957–1970	1970–1983	1983–1996
A. Passive strategies						
VW Market	0.1180	0.0863	0.1489	0.1140	0.0730	0.1766
EW Market	0.1369	0.1100	0.1682	0.1568	0.1505	0.1022
EW 30	0.1401	0.1425	0.1380	0.1619	0.0807	0.1874
B. Sample mean and covariance matrix						
$T = 60$	−0.0281	−0.0299	−0.0290	−0.0863	0.0372	−0.0694
$T = 120$	−0.0196	−0.0806	0.1049	0.1549	−0.1029	0.2068
C. Sample mean with an identity matrix						
$T = 24$	0.1300	0.1000	0.1583	0.1290	0.0826	0.1883
$T = 60$	−0.0521	0.0050	−0.0758	0.1327	−0.1046	0.2068
$T = 120$	0.0511	0.0721	0.0698	0.1537	0.0840	0.2336
D. James–Stein mean with a sample covariance matrix						
$T = 60$	−0.0480	−0.0412	−0.0811	−0.0910	0.0082	−0.0665
$T = 120$	0.0744	0.0008	0.1565	0.1990	−0.0848	0.2378
E. James–Stein mean with an identity matrix						
$T = 24$	0.1401	0.1425	0.1380	0.1619	0.0807	0.1874
$T = 60$	0.0146	−0.0143	0.1444	0.2351	−0.0642	0.2052
$T = 120$	0.1025	0.0613	0.1495	0.1538	−0.0080	0.2127
F. Restricted mean and covariance matrix, unobserved factor						
$T = 24$	0.1254	0.1367	0.1156	0.1559	0.0587	0.1701
$T = 60$	0.1173	0.1288	0.1076	0.1362	0.0744	0.1480
$T = 120$	0.1247	0.1351	0.1159	0.1471	0.0824	0.1525
G. Restricted mean and covariance matrix, observed factor						
$T = 24, s_h = 0.0289$	0.1136	0.1087	0.1181	0.1432	0.0575	0.1507
$T = 24, s_h = 0.0577$	0.1041	0.1259	0.0850	0.1630	0.0364	0.1235
$T = 60, s_h = 0.0289$	0.0994	0.0543	0.1458	0.0855	0.0253	0.2039
$T = 60, s_h = 0.0577$	0.0776	0.0264	0.1321	0.0595	−0.0192	0.2190
$T = 120, s_h = 0.0289$	0.1065	0.0592	0.1546	0.0932	0.0303	0.2126
$T = 120, s_h = 0.0577$	0.0892	0.0325	0.1492	0.0683	−0.0087	0.2341

Out-of-sample monthly Sharpe ratios are computed for several portfolio strategies. Among the passive strategies in panel A, the “VW market” strategy invests in a value-weighted index of all NYSE + AMEX + ZNASDAQ stocks, the “EW market” strategy invests in an equally weighted index of those stocks, and the “EW 30” strategy invests equal dollar amounts in $N = 30$ Dow Jones stocks every month. Panels B–G contain the results for strategies that form mean variance efficient portfolios of the Dow Jones stocks using the stated estimates of the mean and covariance matrix of the stocks’ returns. Every month, T most recent returns are used to estimate the return moments and the portfolios are rebalanced on a monthly basis. In panel F, the restricted strategies produce identical results for any value of s_h , the Sharpe ratio of the optimal orthogonal portfolio. Panel G contains analogous results with an observed factor, which is considered investable.

is required to be continuously listed on an exchange in the past 50 years.¹⁶ Among the stocks that currently constitute the Dow Jones 30 index, 22 satisfy this requirement. The additional eight stocks that satisfy the requirement are selected from earlier Dow Jones lists.

Table 6 presents the portfolio selection results for the individual stocks. The analysis is analogous to that for the decile portfolios, with two exceptions. First, when the link between the mean and covariance matrix is imposed, the results for $T = 24$ are added in order to illustrate the approach for $T < N$.

¹⁶ This criterion provides a reasonably long period for out-of-sample comparisons. The effect of survivorship bias is unlikely to be important, since comparisons are conducted between strategies that combine the same set of stocks.

One advantage of the restricted approach to portfolio selection is that there are no restrictions on the number of assets N . In contrast, using the unrestricted approach, the number of assets must be smaller than the number of time-series observations T , so that the sample estimate of the covariance matrix is nonsingular. Second, in one-factor analysis, the factor portfolio is now included as an investable asset, since it is no longer a linear combination of the N assets. The optimal portfolio is constructed using the result in Equation (24) with the restriction $\Phi = \sigma^2 I$.

The conclusions for the individual stocks are broadly similar to those for the decile portfolios. The Sharpe ratios obtained using the unrestricted estimates (panel B) are low and often negative. Using the identity matrix with the unrestricted mean (panel C) leads to some improvement. Using the identity matrix with the James–Stein (panel E) and the restricted (panel F) mean estimates is even better. The restricted estimates with an observed factor (panel G) perform well, except for a few cases with a large s_h , for the same reason as in the previous subsection. As in Table 5, one is better off not using the value-weighted market as an observed factor portfolio. The only strategy that yields positive and reasonable Sharpe ratios in all subperiods and for all values of T (including $T = 24$) is the one that uses the restricted estimates with no observed factors (panel F). Recall that this strategy uses the identity covariance matrix with the restricted estimate of the mean, which incorporates the information from the covariances.

While both empirical examples are only illustrative, their results are promising. The restricted expected return estimates are economically reasonable and much less variable than historical averages. The portfolio selection results show that imposing the link between the means and covariances dominates using the estimates that do not invoke implications of an exact factor structure.

5. Conclusion

The debate on the usefulness of factor-based pricing models has centered primarily on their ability to explain the cross section of expected returns. Some recent evidence has been unfavorable toward these models. However, it is unlikely that any economic model will explain all phenomena that can be culled from the data. This article adopts a more positive approach to evaluating the usefulness of factor-based models. By imposing certain implications of these models in estimation of expected returns and in portfolio selection, it is shown that these models can play an important role in finance despite their apparent imperfections.

The central theme of the article is that, given a factor structure with an unobserved factor, the vector of expected returns is directly linked to the covariance matrix of returns. Hence the covariances provide information that

can be useful in estimating the means. Simulations and empirical illustrations show that this is indeed the case. Substantially higher precision can be obtained by employing an expected return estimator that incorporates the link. However, imposing the link is not costless. Since an important assumption must be made about the covariance structure, the proposed estimators of expected returns are likely to be biased. With a long and stable time series of returns, it is possible for the noisier but unbiased unrestricted estimates to be more precise. Nevertheless, the assumption of stability of the return distribution over very long horizons is not necessarily appropriate, so the role of the bias should not be overstated.

The link between expected returns and covariances is also important in portfolio selection. Imposing the link is especially useful when the factors are difficult to identify precisely and/or are unstable over time. With no observed factors, a strong form of the link leads to a simplification in which an identity matrix is used as a covariance matrix when calculating the tangency portfolio weights.¹⁷ In all simulated cases with estimated expected returns, the true Sharpe ratio of the tangency portfolio constructed using the identity matrix is higher than using the true covariance matrix. This result holds even when the underlying assumptions are violated. The driving force behind the result is that using the identity matrix effectively links the expected returns to the covariances, whereas using the true covariance matrix does not. The difference between using the true covariance matrix and the identity is especially large when the usual sample mean is used to estimate the expected return.

This article is also related to the literature on the construction of factors for linear asset pricing models. In empirical research, factors constructed using two very different techniques have been widely used. One factor-construction technique extracts common factors from the sample covariance matrix of returns using factor analysis or principal components [e.g., Roll and Ross (1980), Connor and Korajczyk (1986)]. Another factor-building technique assigns assets to several groups based on certain characteristics so that sample mean returns differ across the groups. The difference in average returns across the groups can serve as a factor in a multifactor pricing relation. This second approach is used to build the Fama and French (1993) SMB and HML factors.¹⁸ The two factor-building techniques lead to different factors since they use different information to construct the factors. The Connor–Korajczyk factors are obtained from the information in sample variances and covariances, whereas the Fama–French factors are obtained from the information in sample means. The approach used in this article can be interpreted

¹⁷ If the investable assets are sufficiently different from each other in terms of their residual variances, such as when assets from different asset classes are mixed together, a weaker form of the link that replaces the identity matrix with a more general diagonal matrix can be imposed.

¹⁸ The SMB factor is the return of a portfolio of low market value of equity firms minus the return of a portfolio of high market value of equity firms. The HML factor is the return of a portfolio of firms with high book value to market value of equity ratios minus the return of a portfolio of firms with low book value to market value of equity ratios.

as a mixture of the two techniques. Since the mean return is explicitly linked to the covariance matrix of returns, factors are built using the sample information about *both* the means and the covariances.

The important issue of estimation risk pioneered by Bawa, Brown, and Klein (1979) is not addressed, for tractability reasons. When computing the optimal portfolios, the estimates of the return moments are treated as true parameters. Estimation risk can be incorporated in a Bayesian framework. With a noninformative prior, the Bayesian tangency portfolio weights correspond to the unrestricted estimator of the weight vector in Equation (16), which alleviates the concerns about estimation risk when unrestricted estimators are used.¹⁹ Using restricted estimators of the return moments also has a Bayesian interpretation, albeit less formal. The restricted optimal weights are likely to be close to the weights obtained in a Bayesian framework when the only prior information is that there is a perfect link between model mispricing and the residual covariance matrix. Conditioning on a certain value for the maximum Sharpe ratio is similar in nature to specifying a dogmatic prior belief that the Sharpe ratio is known. Pursuing a Bayesian approach that allows the data to update a less dogmatic prior about the link is beyond the scope of this study, but is one direction for future research.

This article provides economic motivation for using a simply structured covariance matrix, such as an identity matrix, in portfolio selection. The motivation assumes that expected asset returns are driven solely by the assets' covariances with an unobserved factor. Nevertheless, an investor may believe that the assets' risk premia are not entirely risk-based, as suggested by MacKinlay (1995). In that case, a simply structured covariance matrix may serve as a benchmark around which a Bayesian investor can center his prior beliefs about the covariance matrix. The tightness of the prior would reflect the degree of belief that the mispricing is risk based. This economic motivation complements the statistical motivation of Ledoit (1997) for shrinking the sample covariance matrix toward the identity matrix.

Several other issues warrant further study. The robustness of the simulation result that the identity matrix beats the true covariance matrix in portfolio selection is important. This result has strong implications for the value of efforts that seek to develop refined estimators of the covariance matrix for portfolio selection. The framework proposed here provides a natural way of allocating among many assets without imposing artificial constraints. The gains from imposing the link when the number of assets exceeds the number of time-series observations deserve further investigation. Finally, empirical analysis using different sets of assets would be informative.

¹⁹ The correspondence between the Bayesian weights and the unrestricted weights follows because the mean of the Bayesian predictive density is equal to the maximum likelihood estimate of the mean, and the predictive covariance matrix is a scalar multiple of the sample covariance matrix.

While there is substantial empirical evidence questioning the validity of factor-based asset pricing models, our results suggest that casting the models aside could be premature. Even if the pricing relation is not exact, assuming that mispricing is due to risk can serve as a powerful tool for estimation of expected return and for portfolio selection.

Appendix

The appendix derives the formula in Equation (23) for the vector of optimal weights in the tangency portfolio. From basic mean-variance analysis, this $(N + 1) \times 1$ vector equals

$$x_{N+} = (\iota' V^{*-1} \mu^*)^{-1} V^{*-1} \mu^*, \quad (\text{A1})$$

where μ^* is the $(N + 1)$ vector of expected excess returns with the factor portfolio as the $(N + 1)^{\text{st}}$ element and V^* is the $(N + 1) \times (N + 1)$ excess return covariance matrix:

$$V^* = \begin{bmatrix} V & \beta \sigma_p^2 \\ \beta' \sigma_p^2 & \sigma_p^2 \end{bmatrix} = \begin{bmatrix} \beta \beta' \sigma_p^2 + \Sigma & \beta \sigma_p^2 \\ \beta' \sigma_p^2 & \sigma_p^2 \end{bmatrix}, \quad (\text{A2})$$

where $V = \beta \beta' \sigma_p^2 + \Sigma$ has been substituted.

V^* can be analytically inverted using the formula for a partitioned inverse [see Morrison (1990, p. 69)]:

$$V^{*-1} \equiv \begin{bmatrix} \Sigma^{-1} & -\Sigma^{-1} \beta \\ -\beta' \Sigma^{-1} & \frac{1}{\sigma_p^2} + \beta' \Sigma^{-1} \beta \end{bmatrix}. \quad (\text{A3})$$

Using Equation (A3), $\mu^* = [\mu' \ \mu_p]$, and substituting $\alpha = \mu - \beta \mu_p$ from Equation (22), we obtain

$$\begin{aligned} V^{*-1} \mu^* &= \begin{bmatrix} \Sigma^{-1} \mu - \Sigma^{-1} \beta \mu_p \\ -\beta' \Sigma^{-1} \mu + \beta' \Sigma^{-1} \beta \mu_p + \frac{\mu_p}{\sigma_p^2} \end{bmatrix} \\ &= \begin{bmatrix} \Sigma^{-1} \alpha \\ -\beta' \Sigma^{-1} \alpha + \frac{\mu_p}{\sigma_p^2} \end{bmatrix}. \end{aligned} \quad (\text{A4})$$

Returning to Equation (A1), the portfolio weight vector is

$$x_{N+} = c_1 \begin{bmatrix} \Sigma^{-1} \alpha \\ -\beta' \Sigma^{-1} \alpha + \frac{\mu_p}{\sigma_p^2} \end{bmatrix}, \quad (\text{A5})$$

where c_1 is a normalizing constant such that $\iota' x_{N+} = 1$.

References

- Bawa, V. S., S. J. Brown, and R. W. Klein, 1979, *Estimation Risk and Optimal Portfolio Choice*, North Holland, Amsterdam.
- Best, M. J., and R. R. Grauer, 1991, "On the Sensitivity of Mean-Variance Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results," *Review of Financial Studies*, 4, 315-342.

- Black, F., and R. Litterman, 1992, "Global Portfolio Optimization," *Financial Analysts Journal*, September–October 1992.
- Chopra, V. K., and W. T. Ziemba, 1993, "The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice," *Journal of Portfolio Management*, Winter 1993, 6–11.
- Connor, G., and R. A. Korajczyk, 1986, "Performance Measurement Within the Arbitrage Pricing Theory: A New Framework for Analysis," *Journal of Financial Economics* 15, 373–394.
- Daniel, K., and S. Titman, 1997, "Evidence on the Characteristics of Cross-Sectional Variation in Stock Returns," *Journal of Finance*, 52, 133.
- Dickinson, J. P., 1974, "The Reliability of Estimation Procedures in Portfolio Analysis," *Journal of Financial and Quantitative Analysis*, June, 447–462.
- Fama, E. F., and K. R. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics* 33, 3–56.
- Gibbons, M., S. A. Ross, and J. Shanken, 1989, "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57, 1121–1152.
- Green, R. C., and B. Hollifield, 1992, "When Will Mean-Variance Efficient Portfolios Be Well Diversified?" *Journal of Finance*, 47, 1785–1809.
- Hansen, L. P., and R. Jagannathan, 1997, "Assessing Specification Errors in Stochastic Discount Factor Models," *Journal of Finance*, 52, 557–590.
- Jobson, J. D., and B. Korkie, 1980, "Estimation for Markowitz Efficient Portfolios," *Journal of American Statistical Association*, 75, 544–554.
- Jorion, P., 1986, "Bayes-Stein Estimation for Portfolio Analysis," *Journal of Financial and Quantitative Analysis*, 21, 279–292.
- Jorion, P., 1991, "Bayesian and CAPM Estimators of the Means: Implications for Portfolio Selection," *Journal of Banking and Finance*, 15, 717–727.
- Lakonishok, J., A. Shleifer, and R. W. Vishny, 1994, "Contrarian Investment, Extrapolation, and Risk," *Journal of Finance*, 49, 1541–1578.
- Ledoit, O., 1997, "Improved Estimation of the Covariance Matrix of Stock Returns, with an Application to Portfolio Selection," Finance Working Paper no 6-97, UCLA.
- MacKinlay, A. C., 1995, "Multifactor Models do not Explain Deviations from the CAPM," *Journal of Financial Economics*, 38, 3–28.
- Merton, R. C., 1980, "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361.
- Merton, R. C., 1987, "A Simple Model of Capital Market Equilibrium with Incomplete Information," *Journal of Finance*, 42, 483–510.
- Michaud, R. O., 1989, "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?" *Financial Analysts Journal*, 45(1), 31–42.
- Morrison, 1990, *Multivariate Statistical Methods*, 3rd ed., McGraw-Hill, New York.
- Pástor, Ľ., 2000, "Portfolio Selection and Asset Pricing Models," *Journal of Finance*, 55, 179–223.
- Pástor, Ľ., and R. F. Stambaugh, 1999, "Costs of Equity Capital and Model Mispricing," *Journal of Finance*, 54, 67–121.
- Roll, R., 1977, "A Critique of the Asset Pricing Theory's Tests: Part I," *Journal of Financial Economics*, 4, 129–176.
- Roll, R., 1980, "Orthogonal Portfolios," *Journal of Financial and Quantitative Analysis*, 15, 1005–1023.

Roll, R., and S. A. Ross, 1980, "An Empirical Investigation of the Arbitrage Pricing Theory," *Journal of Finance*, 35, 1073–1103.

Ross, S. A., 1976, "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341–360.

Sharpe, W. F., 1963, "A Simplified Model for Portfolio Analysis," *Management Science*, 9, 277–293.

Treynor, J. L., and F. Black, 1973, "How to Use Security Analysis to Improve Portfolio Selection," *Journal of Business*, 46, 65–86.