Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns

Michael W. Brandt
Fuqua School of Business, Duke University, and NBER

Pedro Santa-Clara
UCLA Anderson School, Universidade Nova de Lisboa, and NBER

Rossen Valkanov
Rady School, UCSD

We propose a novel approach to optimizing portfolios with large numbers of assets. We model directly the portfolio weight in each asset as a function of the asset's characteristics. The coefficients of this function are found by optimizing the investor's average utility of the portfolio's return over the sample period. Our approach is computationally simple and easily modified and extended to capture the effect of transaction costs, for example, produces sensible portfolio weights, and offers robust performance in and out of sample. In contrast, the traditional approach of first modeling the joint distribution of returns and then solving for the corresponding optimal portfolio weights is not only difficult to implement for a large number of assets but also yields notoriously noisy and unstable results. We present an empirical implementation for the universe of all stocks in the CRSP–Compustat data set, exploiting the size, value, and momentum anomalies. (*JEL* G11, G12)

Stock characteristics, such as the firm's market capitalization, book-to-market ratio, or lagged return, are related to the stock's expected return, variance, and covariance with other stocks.[1] However, exploiting this fact in portfolio

---

We thank an anonymous referee, Michael Brennan, Kent Daniel, Rob Engle, Larry Harris, Joel Hasbrouck (the editor), Ravi Jagannathan, Ken Kroner, Bruce Lehmann, Francis Longstaff, Anthony Lynch, Richard Roll, Avanidhar Subrahmanyam, Alan Timmerman, Raman Uppal, Halbert White, and seminar participants at Columbia, CUNY Baruch, Emory, HEC Lausanne, INSEAD, LBS, LSE, NYU, Rochester, Tulane, UCSD, UNC, Universidade Nova de Lisboa, Universidade Católica Portuguesa, University College Dublin, University of Michigan, Tilburg, USC, Vienna, the Chicago Quantitative Alliance, the Innovations in Financial Econometrics Conference in Honor of the 2003 Nobel at NYU, the Conference in Honor of Jan Mossin at the Norwegian School of Economics, the Gutmann Public Lecture, the Los Angeles Quantitative Investment Association, Barclays Global Investors, Mellon Capital, Nomura, the Society of Quantitative Analysts, and Wells Capital Management for their comments and suggestions. Send correspondence to Michael W. Brandt, Fuqua School of Business, Duke University, 1 Towerview Drive, Durham, NC 27708; telephone: (919) 660-1948; fax: (919) 882-9157. E-mail: mbrandt@duke.edu.

[1] Fama and French (1996) find that these three characteristics robustly describe the cross-section of expected returns. Chan, Karceski, and Lakonishok (1998) show that these characteristics are also related to the variances and covariances of returns.

doi:10.1093/rfs/hhp003 Advance Access publication February 13, 2009

management has been, up to now, extremely difficult. The traditional mean–variance approach of Markowitz (1952) requires modeling the expected returns, variances, and covariances of all stocks as functions of their characteristics. This is not only a formidable econometric problem given the large number of moments involved and the need to ensure the positive definiteness of the covariance matrix, but the results of the procedure are also notoriously noisy and unstable (e.g., Michaud 1989). In practice, the Markowitz approach is therefore implemented along with a number of different fixes, including shrinkage of the estimates, imposing a factor structure on the covariance matrix, estimation of expected returns from an asset pricing model, or constraining the portfolio weights.[2] While these fixes generally improve the properties of the optimized portfolio, they require substantial resources, such as the tools developed by BARRA, Northfield, and other companies. As a result, formal portfolio optimization based on firm characteristics is seldom implemented by asset managers (with the notable exception of quant managers, who are a small part of the profession), even though it has the potential to provide large benefits to investors.[3]

We propose a simple new approach to equity portfolio optimization based on firm characteristics. We parameterize the portfolio weight of each stock as a function of the firm's characteristics and estimate the coefficients of the portfolio policy by maximizing the average utility the investor would have obtained by implementing the policy over the historical sample period.

Our approach has a number of conceptual advantages. First, we avoid completely the auxiliary, yet very difficult, step of modeling the joint distribution of returns and characteristics and instead focus directly on the object of interest—the portfolio weights. Second, parameterizing the portfolio policy leads to a tremendous reduction in dimensionality. For a problem with N stocks, the traditional Markowitz approach requires modeling N first and $(N^2 + N)/2$ second moments of returns. With preferences other than the simplistic quadratic utility, the traditional approach involves a practically unmanageable number of higher moments for even a relatively small number of stocks (e.g., 100 stocks have over 300,000 third moments). In contrast, our approach involves modeling only N portfolio weights regardless of the investor's preferences and the joint distribution of asset returns. Because of this reduction in dimensionality, our approach escapes the common statistical problems of imprecise coefficient estimates and overfitting, while allowing us to solve very large-scale problems with arbitrary preferences. Third, but related, our approach captures implicitly the relation between the characteristics and expected returns, variances, covariances, and even higher-order moments of returns since they affect the

[2] See Jobson and Korkie (1980, 1981); Frost and Savarino (1986, 1988); Jorion (1986); Black and Litterman (1992); Chan, Karceski, and Lakonishok (1999); Pastor (2000); Pastor and Stambaugh (2000, 2002); Jagannathan and Ma (2002); and Ledoit and Wolf (2003, 2004). Brandt (2004) surveys the literature.

[3] See, for instance, Chan, Karceski, and Lakonishok (1999) and Jagannathan and Ma (2002).

distribution of the optimized portfolio's returns, and therefore the investor's expected utility. Fourth, by framing the portfolio optimization as a statistical estimation problem with an expected utility objective function (a "maximum expected utility" estimator as opposed to the usual least-squares or maximum-likelihood estimators), we can easily test individual and joint hypotheses about the optimal portfolio weights.

From a practical perspective, our approach is simple to implement and produces robust results in and out of sample. It is also easily modified and extended. We discuss a number of possible extensions, including the use of different objective functions, the use of different parameterizations of the portfolio policy to accommodate short-sale constraints, and conditioning the portfolio policy on macroeconomic predictors. Perhaps most interestingly, from a practical perspective, we show how our approach can be extended to capture the effect of transaction costs.

Our paper is related to a recent literature on drawing inferences about optimal portfolio weights without explicitly modeling the underlying return distribution. Brandt (1999) and Aït-Sahalia and Brandt (2001) model the optimal allocations to stocks, bonds, and cash as nonparametric functions of variables that predict returns. Nigmatullin (2003) extends their nonparametric approach to incorporate parameter and model uncertainty in a Bayesian setting. More closely related to our paper are Brandt and Santa-Clara (2006), who study a market-timing problem involving stocks, bonds, and cash by modeling the optimal portfolio weights as functions of the predictors. Specifically, they model the weight in each asset class as a separate function (with coefficients that are specific to the asset class) of a common set of macroeconomic variables. Their approach is relevant for problems involving a few assets that have fundamentally different characteristics, such as the allocation of capital across different asset classes. In contrast, our paper models the weight invested in each asset as the same function (with common coefficients) of asset-specific variables. This is the relevant problem when choosing among a large number of essentially similar assets, such as the universe of stocks.

We use our approach to optimize a portfolio of all the stocks in the CRSP-Compustat data set from 1974 to 2002, using as characteristics the market capitalization, book-to-market ratio, and lagged one-year return of each firm. The investor is assumed to have constant relative risk aversion (CRRA) preferences. Our empirical results document the importance of the firm characteristics for explaining deviations of the optimal portfolio weights from observed market capitalization weights. Relative to market cap weights, the optimal portfolio with and without short-sale constraints allocates considerably more wealth to stocks of small firms, firms with high book-to-market ratios (value firms), and firms with high lagged returns (winners). With a relative risk aversion of five, the certainty-equivalent gain from investing in the optimal portfolio relative to holding the market is an annualized 11.1% in sample and 5.4% out of sample. The benefits are even greater when we allow the coefficients of the portfolio policy to depend on the slope of the yield curve. We present results for long-only portfolio policies and find that the constraint has significant costs for the investor. We examine the impact of increasing the level of risk aversion on the portfolio policy and find essentially that size and momentum become less appealing while value retains its importance. Finally, we incorporate transaction costs. We show that, with a simple policy function that features a no-trade boundary, the portfolio turnover is reduced by up to 50% with only marginal deterioration in performance.

The remainder of the paper proceeds as follows. We describe the basic idea and various extensions of our approach in Section 1. The empirical application is presented in Section 2. We conclude in Section 3.

1. Methodology

1.1 Basic idea

Suppose that at each date t, there is a large number, N_t, of stocks in the investable universe.[4] Each stock i has a return of $r_{i,t+1}$ from date t to $t+1$ and is associated with a vector of firm characteristics $x_{i,t}$ observed at date t. For example, the characteristics could be the market capitalization of the stock, the book-to-market ratio of the stock, and the lagged twelve-month return on the stock. The investor's problem is to choose the portfolio weights $w_{i,t}$ to maximize the conditional expected utility of the portfolio's return $r_{p,t+1}$,

$$\max_{\{w_{i,t}\}_{i=1}^{N_t}} \mathrm{E}_t[u(r_{p,t+1})] = \mathrm{E}_t\left[u\left(\sum_{i=1}^{N_t} w_{i,t} r_{i,t+1}\right)\right]. \tag{1}$$

We parameterize the optimal portfolio weights as a function of the stocks' characteristics,

$$w_{i,t} = f(x_{i,t}; \theta). \tag{2}$$

In a large part of the paper, we concentrate on the following simple linear specification for the portfolio weight function:

$$w_{i,t} = \bar{w}_{i,t} + \frac{1}{N_t}\theta^\top \hat{x}_{i,t}, \tag{3}$$

where $\bar{w}_{i,t}$ is the weight of stock i at date t in a benchmark portfolio, such as the value-weighted market portfolio, θ is a vector of coefficients to be estimated, and $\hat{x}_{i,t}$ are the characteristics of stock i, standardized cross-sectionally to have zero mean and unit standard deviation across all stocks at date t. Note that, rather than estimating one weight for each stock at each point in time, we

[4] Our method automatically accommodates the realistic case of a varying number of stocks through time. This is not trivially done in the traditional approach, as discussed by Stambaugh (1997).

estimate weights as a single function of characteristics that applies to all stocks over time—a portfolio policy.

This particular parameterization captures the idea of active portfolio management relative to a performance benchmark. The intercept is the weight of the stock in the benchmark portfolio, and the term $\theta^\top \hat{x}_{i,t}$ represents the deviations of the optimal portfolio weight from this benchmark. The characteristics are standardized for two reasons. First, the cross-sectional distribution of the standardized $\hat{x}_{i,t}$ is stationary through time, while that of the raw $x_{i,t}$ may be nonstationary. Second, the standardization implies that the cross-sectional average of $\theta^\top \hat{x}_{i,t}$ is zero, which means that the deviations of the optimal portfolio weights from the benchmark weights sum to zero, and therefore that the optimal portfolio weights always sum to one. Finally, the term $1/N_t$ is a normalization that allows the portfolio weight function to be applied to an arbitrary and time-varying number of stocks. Without this normalization, doubling the number of stocks without otherwise changing the cross-sectional distribution of the characteristics results in twice as aggressive allocations, even though the investment opportunities are fundamentally unchanged.

There are a number of alternative ways to normalize the firm characteristics. One alternative is to subtract the mean characteristic of the industry (at a given level of aggregation) rather than the mean of the universe. In this way, the standardized characteristics measure deviations from the industry, which may clean out systematic operational or financial differences across industries. Asness, Porter, and Steven (2001) stress the importance of industry normalizations. Besides the impact of purifying the signal for expected returns, using industry-normalized characteristics is likely to reduce the risk of the portfolio since there will be lower net exposure to industries. Another alternative is to run a cross-sectional regression each period of each given characteristic on other firm variables (possibly including industry dummies), and take the residuals of that regression as inputs to the portfolio policy. These residuals are the component of the characteristic that is orthogonal to the regression's explanatory variables and will therefore remove all commonality in the characteristics due to those variables.

The most important aspect of our parameterization is that the coefficients θ are constant across assets and through time. Constant coefficients across assets implies that the portfolio weight in each stock depends only on the stock's characteristics and not on the stock's historic returns. Two stocks that are close to each other in characteristics associated with expected returns and risk should have similar weights in the portfolio even if their sample returns are very different. The implicit assumption is that the characteristics fully capture all aspects of the joint distribution of returns that are relevant for forming optimal portfolios. Constant coefficients through time means that the coefficients that maximize the investor's conditional expected utility at a given date are the same for all dates, and therefore also maximize the investor's unconditional expected utility.

These two facts imply that we can rewrite the conditional optimization with respect to the portfolio weights $w_{i,t}$ in Equation (1) as the following unconditional optimization with respect to the coefficients θ:

$$\max_{\theta} \ \mathrm{E}[u(r_{p,t+1})] = \mathrm{E}\left[u\left(\sum_{i=1}^{N_t} f(x_{i,t};\theta) r_{i,t+1}\right)\right]. \tag{4}$$

We can then estimate the coefficients θ by maximizing the corresponding sample analog

$$\max_{\theta} \ \frac{1}{T}\sum_{t=0}^{T-1} u(r_{p,t+1}) = \frac{1}{T}\sum_{t=0}^{T-1} u\left(\sum_{i=1}^{N_t} f(x_{i,t};\theta) r_{i,t+1}\right), \tag{5}$$

for some prespecified utility function (e.g., quadratic or CRRA). In the linear policy case (3), the optimization problem is

$$\max_{\theta} \frac{1}{T}\sum_{t=0}^{T-1} u\left(\sum_{i=1}^{N_t}\left(\bar{w}_{i,t} + \frac{1}{N_t}\theta^{\top}\hat{x}_{i,t}\right) r_{i,t+1}\right). \tag{6}$$

Four observations about our approach are worth making at this point. First, optimizing a portfolio of a very large number of stocks is extremely simple. Given the relatively low dimensionality of the parameter vector, it is computationally trivial to optimize the portfolio with nonlinear optimization methods.[5] The computational burden of our approach only grows with the number of characteristics entering the portfolio policy, not with the number of assets in the portfolio. Second, the formulation is numerically robust. We optimize the entire portfolio by choosing only a few parameters θ. This parsimony reduces the risk of in-sample overfitting since the coefficients will only deviate from zero if the respective characteristics offer an interesting combination of return and risk consistently across stocks and through time. For the same reason, the optimized portfolio weights tend not to take extreme values.

Third, the linear policy (3) conveniently nests the long–short portfolios construction of Fama and French (1993) or its extension in Carhart (1997). To see how this is the case, assume that the portfolio policy in Equation (2) is parameterized in a linear manner as in (3). Let the benchmark weights be the market capitalization weights and the characteristics be defined as 1 if the stock is in a top quantile, −1 if it is in the bottom quantile, and zero for intermediate quantiles of market capitalization (*me*), book to market ratio (*btm*), and past

[5] For most common utility functions and given the linearity of the portfolio policy (3) in the coefficients θ, it is easy to derive analytically the gradient and the Hessian of the optimization problem.

return (*mom*). Then, the portfolio return is

$$\begin{aligned} r_{p,t+1} &= r_{m,t+1} + \theta_{me} \sum_{i=1}^{N_t} \left(\frac{1}{Q_t} me_{i,t} \right) r_{i,t+1} + \cdots \theta_{btm} \sum_{i=1}^{N_t} \left(\frac{1}{Q_t} btm_{i,t} \right) r_{i,t+1} \\ &\quad + \theta_{mom} \sum_{i=1}^{N_t} \left(\frac{1}{Q_t} mom_{i,t} \right) r_{i,t+1} \\ &= r_{m,t+1} + \theta_{me} r_{smb,t+1} + \theta_{btm} r_{hml,t+1} + \theta_{mom} r_{wml,t+1}, \end{aligned} \tag{7}$$

where $r_{smb,t+1}$, $r_{hml,t+1}$, and $r_{wml,t+1}$ are the returns to "small-minus-big," "high-minus-low," and "winners-minus-losers" portfolios, and Q_t is the number of firms in the quantile. Under this interpretation, the theta coefficients are the weights put on each of the factor portfolios. To find the weight of the portfolio in each individual stock, we still need to multiply the coefficients θ by the respective characteristics.

While our approach nests the problem of optimally investing in factor-mimicking long–short portfolios, the reverse is only true when the portfolio policy is linear and unconstrained. In the more general and practically relevant case of constrained portfolio weights, such as the long-only specification discussed in Section 1.3.2, the optimal portfolio can no longer be seen as a choice among long–short factor portfolios. The reason is that with long–short factor portfolios, the overall portfolio constraints cannot be imposed on a stock-by-stock basis. Similarly, the portfolio policy proposed in Section 1.4 to deal with transaction costs is nonlinear and recursive. That policy also cannot be implemented by a static choice of long–short factor portfolios.

Fourth, the optimization takes into account the relation between the characteristics and expected returns, variances, covariances, and even higher-order moments of returns, to the extent that they affect the distribution of the optimized portfolio's returns, and therefore the investor's expected utility. In the optimization, the degree of cross-sectional predictability of each component of the joint return distribution is intuitively weighted by its impact on the overall expected utility of the investor.

To better understand this point, we can approximate the expected utility of the investor with a Taylor series expansion around the portfolio's expected return $\mathrm{E}[r_{p,t+1}]$,

$$\begin{aligned} \mathrm{E}[u(r_{p,t+1})] &\approx u(\mathrm{E}[r_{p,t+1}]) + \frac{1}{2} u''(\mathrm{E}[r_{p,t+1}]) \mathrm{E}[(r_{p,t+1} - \mathrm{E}[r_{p,t+1}])^2] \\ &\quad + \frac{1}{6} u'''(\mathrm{E}[r_{p,t+1}]) \mathrm{E}[(r_{p,t+1} - \mathrm{E}[r_{p,t+1}])^3] + \cdots . \end{aligned} \tag{8}$$

This expansion shows that, in general, the investor cares about all the moments of the distribution of the portfolio return.[6] Since the portfolio return is given by

$$r_{p,t+1} = \sum_{i=1}^{N_t} f(x_{i,t};\theta) r_{i,t+1}, \tag{9}$$

the moments of its distribution depend implicitly on the joint distribution of the returns and characteristics of all firms. The coefficients θ affect the distribution of the portfolio's return by changing the weights given to the returns of the individual firms in the overall portfolio.

To perform a comparable portfolio optimization using the traditional Markowitz approach requires modeling the means, variances, and covariances of all the stocks as functions of their characteristics. This entails estimating for each date t a large number of N_t conditional expected returns and $(N_t^2 + N_t)/2$ conditional variances and covariances. Besides the fact that the number of these moments grows quickly with the number of stocks, making robust estimation a real problem, it is extremely challenging to estimate the covariance matrix as a function of stock characteristics in a way that guarantees its positive definiteness. Furthermore, extending the traditional approach beyond first and second moments, when the investor's utility function is not quadratic, is practically impossible because it requires modeling not only the conditional skewness and kurtosis of each stock but also the numerous high-order cross-moments.

Finally, when the benchmark is the value-weighted market, m, the return of the linear portfolio policy (3) can be written as

$$r_{p,t+1} = \sum_{i=1}^{N_t} \bar{w}_{i,t} r_{i,t+1} + \sum_{i=1}^{N_t} \left(\frac{1}{N_t} \theta^\top \hat{x}_{i,t} \right) r_{i,t+1} = r_{m,t+1} + r_{h,t+1}, \tag{10}$$

where h is a long–short hedge fund with weights $\theta^\top \hat{x}_{i,t}/N_t$ that add up to zero. Therefore, problem (8) can be reinterpreted as the problem of a hedge fund that optimizes its portfolio to maximize the utility of an investor who already holds the market (i.e., the market is a background risk for the investor).

1.2 Statistical inference

By formulating the portfolio problem as a statistical estimation problem, we can easily obtain standard errors for the coefficients of the weight function. The "maximum expected utility" estimate $\hat{\theta}$, defined by the optimization problem

[6] This is especially important in dealing with assets with distributions that significantly depart from normality, such as options and credit-sensitive securities. Santa-Clara and Saretto (2006) provide an application of our approach to option portfolios.

(5) with the linear portfolio policy (3), satisfies the first-order conditions[7]

$$\frac{1}{T}\sum_{t=0}^{T-1} h(r_{t+1}, x_t; \theta) \equiv \frac{1}{T}\sum_{t=0}^{T-1} u'(r_{p,t+1}) \left(\frac{1}{N_t} \hat{x}_t^\top r_{t+1} \right) = 0, \tag{11}$$

and can, therefore, be interpreted as a method of moments estimator. From Hansen (1982), the asymptotic covariance matrix of this estimator is

$$\Sigma_\theta \equiv \mathrm{AsyVar}[\hat{\theta}] = \frac{1}{T}[G^\top V^{-1} G]^{-1}, \tag{12}$$

where

$$G \equiv \frac{1}{T}\sum_{t=0}^{T-1} \frac{\partial h(r_{t+1}, x_t; \theta)}{\partial \theta} = \frac{1}{T}\sum_{t=0}^{T-1} u''(r_{p,t+1}) \left(\frac{1}{N_t} \hat{x}_t^\top r_{t+1} \right) \left(\frac{1}{N_t} \hat{x}_t^\top r_{t+1} \right)^\top \tag{13}$$

and V is a consistent estimator of the covariance matrix of $h(r, x; \theta)$.

Assuming marginal utilities are uncorrelated, which is true by construction when the portfolio policy is correctly specified and the optimization is unconstrained, we can consistently estimate V by

$$\frac{1}{T}\sum_{t=0}^{T-1} h(r_{t+1}, x_t; \hat{\theta}) h(r_{t+1}, x_t; \hat{\theta})^\top. \tag{14}$$

If we want to allow for the possibility of a misspecified portfolio policy (e.g., for the purpose of specification testing discussed below), or if constraints are imposed, we may instead use an autocorrelation-adjusted estimator of V (e.g., Newey and West 1987).

Alternatively, the covariance matrix of coefficients $\hat{\Sigma}_\theta$ can be estimated by bootstrap. For that, we simply generate a large number of samples of returns and characteristics by randomly drawing monthly observations from the original data set (with replacement).[8] For each of these bootstrapped samples, we estimate the coefficients of the optimal portfolio policy and compute the covariance matrix of the coefficients across all the bootstrapped samples. This approach has the advantage of not relying on asymptotic results, and takes into account the potentially nonnormal features of the data. We use bootstrapped standard errors in the empirical analysis below.

The resulting estimate of the covariance matrix of the coefficients $\hat{\Sigma}_\theta$ can be used to test individual and joint hypotheses about the elements of θ. These tests address the economic question of whether a given characteristic is related to the

[7] With more general portfolio policies, we also need to differentiate $f(x_{i,t}; \theta)$ with respect to θ.

[8] We also experimented with block bootstrapping techniques that maintain the time-series dependence of the data (e.g., Politis and Romano 1994). The resulting inferences are qualitatively the same.

moments of returns in such a way that the investor finds it optimal to deviate from the benchmark portfolio weights according to the realization of the characteristic for each stock. It is important to recognize that this is not equivalent to testing whether a characteristic is cross-sectionally related to the conditional moments of stock returns for at least two reasons. First, the benchmark portfolio weights may already reflect an exposure to the characteristics, and it may not be optimal to change that exposure. Second, a given characteristic may be correlated with first and second moments in an offsetting way, such that the conditionally optimal portfolio weights are independent of the characteristic.

The interpretation of our approach as a method of moments estimator suggests a way of testing the functional specification of the portfolio policy. In going from Equation (1) to Equation (4), we assume that the functional form of the portfolio policy is correct, to replace $w_{i,t}$ with a function of $x_{i,t}$, and that the coefficients are constant through time, to condition down the expectation. If either assumption is incorrect, the marginal utilities in Equation (11) will be correlated with variables in the investor's information set at date t, which may include missing characteristics or variables that are correlated with the variation in the coefficients. We can therefore perform specification tests for the portfolio policy using the standard overidentifying restrictions test of Hansen (1982).

Finally, note that the method of moments interpretation does not necessarily render our approach frequentist and therefore unable to accommodate finite-sample uncertainty about the parameters and model specification. Nigmatullin (2003) shows how to interpret first-order conditions similar to Equation (11) from a Bayesian perspective using the idea of an empirical likelihood function and explains how to incorporate parameter and model uncertainty. While his application deals with the nonparametric approach of Aït-Sahalia and Brandt (2001), the general idea applies directly to our approach.

1.3 Refinements and extensions

Besides its effectiveness and simplicity, an important strength of our approach is that the basic idea is easily refined and extended to suit specific applications. We now discuss some of the possible refinements and extensions to illustrate the flexibility of our approach.

1.3.1 Objective functions. The most important ingredient of any portfolio choice problem is the investor's objective function. In contrast to the traditional Markowitz approach, our specification of the portfolio choice problem can accommodate any choice of objective function. The only implicit assumption is that the conditional expected utility maximization problem (1) be well specified with a unique solution. Besides the standard HARA preferences (which nest constant relative risk aversion, constant absolute risk aversion, log, and quadratic utility), our approach can also be applied to behaviorally motivated utility functions, such as loss aversion, ambiguity aversion, or disappointment aversion, as well as practitioner-oriented objective functions, including maximizing the Sharpe or information ratios, beating or tracking a benchmark, controlling drawdowns, or maintaining a certain value at risk (VaR).[9]

In most of the empirical applications, we use standard CRRA preferences over wealth,

$$u(r_{p,t+1}) = \frac{(1 + r_{p,t+1})^{1-\gamma}}{1-\gamma}. \tag{15}$$

The advantage of CRRA utility is that it incorporates preferences toward higher-order moments in a parsimonious manner. In addition, the utility function is twice continuously differentiable, which allows us to use more efficient numerical optimization algorithms that make use of the analytic gradient and Hessian of the objective function. We also offer results for the minimum variance and maximum Sharpe ratio portfolios.

1.3.2 Portfolio weight constraints. By far the most common departure from the basic portfolio choice problem (1) in practice is to impose constraints on the optimal portfolio weights. In our approach, these constraints have to be imposed through the parameterization of the portfolio policy. For example, consider the case of the no-short-sale constraint in long-only equity portfolios. The simplest way to impose this constraint through the portfolio policy is to truncate the portfolio weights in Equation (3) at zero. Unfortunately, in doing so, the optimal portfolio weights no longer sum to one (setting the negative weights to zero results in a sum of weights greater than one). We, therefore, need to renormalize the portfolio weights as follows:

$$w_{i,t}^{+} = \frac{\max[0, w_{i,t}]}{\sum_{j=1}^{N_t} \max[0, w_{j,t}]}. \tag{16}$$

Besides guaranteeing positivity of the portfolio weights, this specification is also an example of a nonlinear parameterization of the portfolio weight function (2).

One computational problem with this specification of the portfolio policy function is its nondifferentiability at $w_{i,t} = 0$. In order to compute the standard errors of the estimated θ from first-order conditions analogous to Equation (11), we require first-order derivatives. One way to overcome this problem in practice is to approximate the function $\max[0, y]$ between two close points $y = 0$ and $y = \alpha > 0$ with either a third- or a fifth-order polynomial with smooth first- or first- and second-order derivatives at the end points, respectively. Using bootstrapped standard errors is an obvious approach to avoid the problem.

[9] Benartzi and Thaler (1995); Aït-Sahalia and Brandt (2001); Ang, Bekaert, and Liu (2005); and Gomes (2005), among others, examine the role of behaviorally motivated preference in portfolio choice. Practitioner-oriented objective functions are considered, for example, by Roy (1952); Grossman and Vila (1989); Browne (1999); Basak and Shapiro (2001); Tepla (2001); and Alexander and Baptista (2002).

1.3.3 Nonlinearities and interactions. Although we explicitly specified the portfolio policy (3) as a linear function of the characteristics, the linearity assumption is actually innocuous because the characteristics $x_{i,t}$ can always contain nonlinear transformations of a more basic set of characteristics $y_{i,t}$. This means that the linear portfolio weights can be interpreted as a more general portfolio policy function $w_{i,t} = \bar{w}_{i,t} + g(y_{i,t}; \theta)$ for any $g(\cdot;\cdot)$ that can be spanned by a polynomial expansion in the more basic state variables $y_{i,t}$. Our approach therefore accommodates very general departures of the optimal portfolio weights from the benchmark weights.

Cross-products of the characteristics are an interesting form of nonlinearity because they have the potential to capture interactions between the characteristics. For instance, there is considerable evidence in the literature that the momentum effect is concentrated in the group of growth (low book-to-market) firms (e.g., Daniel and Titman 1999). Our approach can capture this empirical regularity by including the product of the book-to-market ratio and the one-year lagged return as an additional characteristic.

In practice, we need to choose a finite set of characteristics as well as possible nonlinear transformations and interactions of these characteristics to include in the portfolio policy specification. This variable selection for modeling portfolio weights is no different from variable selection for modeling expected returns with regressions. The characteristics and their transformations can be chosen on the basis of individual t tests and joint F tests computed using the covariance matrix of the coefficient estimates, or on the basis of out-of-sample performance.

1.3.4 Time-varying coefficients. The critical assumption required for conditioning down the expectation to rewrite the conditional problem (1) as the unconditional problem (4) is that the coefficients of the portfolio policy are constant through time. While this is a convenient assumption, there is no obvious economic reason for the relation between firm characteristics and the joint distribution of returns to be time invariant. In fact, there is substantial evidence that economic variables related to the business cycle forecast aggregate stock and bond returns.[10] Moreover, the cross-section of expected returns appears to be time varying as a function of the same predictors (e.g., Cooper, Gulen, and Vassalou 2000).

To accommodate possible time variation in the coefficients of the portfolio policy, we can explicitly model the coefficients as functions of the business cycle variables. Given a vector of predictors observable at date t, denoted by

[10] For example, Keim and Stambaugh (1986); Campbell and Shiller (1988); Fama and French (1988, 1989); Fama (1990); Campbell (1991); and Hodrick (1992) report evidence that the stock market returns can be forecasted by the dividend–price ratio, the short-term interest rate, the term spread, and the credit spread.

z_t, we can extend the portfolio policy (3) as

$$w_{i,t} = \bar{w}_{i,t} + \frac{1}{N_t}\theta^\top(z_t \otimes x_{i,t}), \tag{17}$$

where $\otimes$ denotes the Kronecker product of two vectors. In this form, the impact of the characteristics on the portfolio weight varies with the realization of the predictors z_t.

1.3.5 Shrinkage. Shrinkage estimation is an effective technique for reducing the effect of estimation error and in-sample fitting in portfolio optimization. In shrinkage estimation, "shrunk" estimates are constructed as a convex combination of sample estimates and shrinkage targets. The shrinkage targets are either of statistical nature, such as the grand mean of all estimates, or are generated by the predictions of a theoretical model. The efficacy of shrinkage to a statistical target in portfolio choice problems is demonstrated by Jobson and Korkie (1981); Jorion (1986); Frost and Savarino (1988); and DeMiguel, Garlappi, and Uppal (2007), among others. Theoretically motivated shrinkage targets are advocated by Black and Litterman (1992); Kandel and Stambaugh (1996); Pastor (2000); and Pastor and Stambaugh (2000).

Shrinkage estimation is traditionally applied to the parameters of the return-generating process. The idea of downweighting the information contained in a single set of data in favor of an ex ante reasonable benchmark is, however, equally applicable to our method. Recall that the portfolio weight parameterization (3) can be interpreted as a data-driven tilt away from holding the benchmark portfolio. If this benchmark portfolio is an ex ante efficient portfolio according to some theoretical model, such as the market portfolio for the CAPM, it is natural to consider shrinking the parameterized portfolio weights toward these benchmark weights. This is mechanically accomplished by simply reducing the absolute magnitudes of the θ coefficients relative to their in-sample estimates. The extent of shrinkage depends, as with all shrinkage estimators, on the potential magnitude of the estimation error in θ, as well as on the strength of the investor's belief in the theoretical model.

1.4 Transaction costs

In this section, we show how to optimize portfolio policies taking into account transaction costs. For a given policy such as (3), the turnover of each period is the sum of all the absolute changes in portfolio weights from one period to the next, net of the changes that occur mechanically due to the relative returns of different assets in the portfolio. Ignoring these mechanical changes in weights for now, we define turnover as

$$T_t = \sum_{i=1}^{N_t} |w_{i,t} - w_{i,t-1}|. \tag{18}$$

Therefore, the return to the portfolio net of trading costs is

$$r_{pt+1} = \sum_{i=1}^{N_t} w_i r_{i,t+1} - c_{i,t}|w_{i,t} - w_{i,t-1}|, \tag{19}$$

where $c_{i,t}$ reflects the proportional transaction cost for stock i at time t. These transaction costs may be estimated directly from market liquidity measures or may be modeled as a function of the stocks' characteristics, such as their market capitalization. Note that we should use estimates of one-way trading costs to input in the equation above since our measure of turnover already includes both the buys and sells (positive and negative changes in weights). We can then find the optimal values of the coefficients by optimizing the average utility of the returns net of trading costs.

The linear functional form of policy (3) is clearly not optimal in the presence of transaction costs. Magill and Constantinides (1976); Taksar, Klass, and Assaf (1988); and Davis and Norman (1990) study the optimal portfolio choice between a risky and a riskless asset in the presence of proportional trading costs.[11] They show that the optimal policy is characterized by a boundary around the target weight for the risky asset. When the current weight is within this boundary, it is optimal not to trade. When the current weight is outside the boundary, however, it is optimal to trade to the boundary, but not to the target. This result is intuitive since when the weight is close to the target, there is only a second-order small gain from rebalancing to the target but a first-order cost from trading. Leland (2000) studies the optimal portfolio problem with multiple risky assets and proportional transaction costs. He finds again that the optimal policy has a no-trade zone with partial adjustment of the portfolio weights to the border when the current holdings are outside the no-trade zone.

Motivated by this theoretical literature, we propose the following functional form of the portfolio weights in the presence of transaction costs, which also illustrates how easy it is in our approach to deal with nonlinear and recursive portfolio policies. Start with an initial portfolio, given by our previous optimal policy,

$$w_{i,0} = \bar{w}_{i,0} + \theta^\top x_{i,0}. \tag{20}$$

Then, each period, define a "target" portfolio that is given by the same policy,

$$w_{i,t}^t = \bar{w}_{i,t} + \theta^\top x_{i,t}. \tag{21}$$

Before trading at time t, the portfolio is the same as the portfolio at time $t-1$ with the weights changed by the returns from $t-1$ to t. Call this the "hold"

[11] See also Dixit (1991); Dumas (1991); Shreve and Soner (1994); and Akian, Menaldi, and Sulem (1996).

portfolio,

$$w_{i,t}^{h} = w_{i,t-1}\frac{1 + r_{i,t}}{1 + r_{p,t}}. \tag{22}$$

If, on one hand, the hold portfolio is sufficiently close to the target portfolio, it is better not to trade. We define the distance between the portfolios as a sum of squares. It follows that

$$w_{i,t} = w_{i,t}^{h} \quad \text{if} \quad \frac{1}{N_t}\sum_{i=1}^{N_t}\left(w_{i,t}^{t} - w_{i,t}^{h}\right)^2 <= k^2. \tag{23}$$

In this way, the no-trade region is a hypersphere of radius k around the target portfolio weights. This is not necessarily the shape of the optimal trade region, and we propose it only as a simple approximation.

If, on the other hand, the hold portfolio is sufficiently far from the target, the investor should trade to the frontier of the no-trade region. In that case, the new portfolio is a weighted average of the hold portfolio and the target portfolio,

$$w_{i,t} = \alpha_t w_{i,t}^{h} + (1-\alpha_t)w_{i,t}^{t}, \quad \text{if} \quad \frac{1}{N_t}\sum_{i=1}^{N_t}\left(w_{i,t}^{t} - w_{i,t}^{h}\right)^2 > k^2. \tag{24}$$

We can pick α_t such that the new portfolio w_t is exactly at the boundary to capture the intuition that the investor should trade to the boundary when outside of the no-trade region,

$$\begin{aligned}\frac{1}{N_t}\sum_{i=1}^{N_t}\left(w_{i,t}^{t} - w_{i,t}\right)^2 &= \frac{1}{N_t}\sum_{i=1}^{N_t}\left(w_{i,t}^{t} - \alpha_t w_{i,t}^{h} - (1-\alpha_t)w_{i,t}^{t}\right)^2 \\ &= \alpha_t^2\frac{1}{N_t}\sum_{i=1}^{N_t}\left(w_{i,t}^{t} - w_{i,t}^{h}\right)^2. \end{aligned} \tag{25}$$

Setting this equal to k^2 and solving for α, we obtain

$$\alpha_t = \frac{k\sqrt{N_t}}{\left(\sum_{i=1}^{N_t}\left(w_{i,t}^{t} - w_{i,t}^{h}\right)^2\right)^{1/2}}. \tag{26}$$

It is worth reiterating that the functional form of the portfolio policy described above is, as in the base case without transaction costs, only an approximation of the theoretically optimal, but unfortunately unknown, functional form. The quality of this approximation is inherently application specific. However, one of the strengths of our approach is the ease with which different portfolio policy functions can be implemented and compared.

2. Empirical Application

To illustrate the simplicity, the flexibility, and, most importantly, the effectiveness of our approach, we present an empirical application involving the universe of all listed stocks in the United States from January 1964 to December 2002. We first describe the data and then present results for the base case and various extensions, both in and out of sample. Unless otherwise stated, we assume an investor with CRRA preference and a relative risk aversion of five. In the application, the investor is restricted to only invest in stocks. We do not include the risk-free asset in the investment opportunity set. The reason is that the first-order effect of allowing investments in the risk-free asset is to vary the leverage of the portfolio, which only corresponds to a change in the scale of the stock portfolio weights and is not interesting per se.

2.1 Data

We use monthly firm-level returns from CRSP as well as firm-level characteristics obtained from the CRSP–Compustat merged data set, from January 1964 to December 2002. For each firm in this data set, we construct the following variables at the end of each fiscal year: the log of the firm's market equity (*me*), defined as the log of the price per share times the number of shares outstanding, and the firm's log book-to-market ratio (*btm*), defined as the log of one plus book equity (total assets minus liabilities, plus balance sheet deferred taxes and investment tax credits, minus preferred stock value) divided by market equity.[12] We use the standard timing convention of leaving at least a six-month lag between the fiscal year-end characteristics and the monthly returns, to ensure that the information from the annual reports would have been publicly available at the time of the investment decision. From the CRSP database, we record for each firm the lagged one-year return (*mom*) defined as the compounded return between months $t-13$ and $t-2$. Similar definitions of the three characteristics are commonly used in the literature (e.g., Fama and French 1996). The Appendix provides further details about the firm-level data, including the exact definitions of the components of each variable. We use size, book-to-market, and momentum as conditioning characteristics in the portfolio optimization since we want to compare our results with previous studies and these characteristics are the most widely used in the literature.

The number of firms in our sample is generally trending upward, with an average annual growth rate of 4.2%. The average number of firms throughout our sample is 3680, with the fewest firms in February 1964 (1033 firms) and the most firms in November 1997 (6356 firms).

Figure 1 describes the three firm characteristics. The first column plots the cross-sectional means of the (nonstandardized) characteristics at each month in our sample. The second column shows the corresponding cross-sectional

[12] Taking logs makes the cross-sectional distribution of *me* and *btm* more symmetric and reduces the effect of outliers.

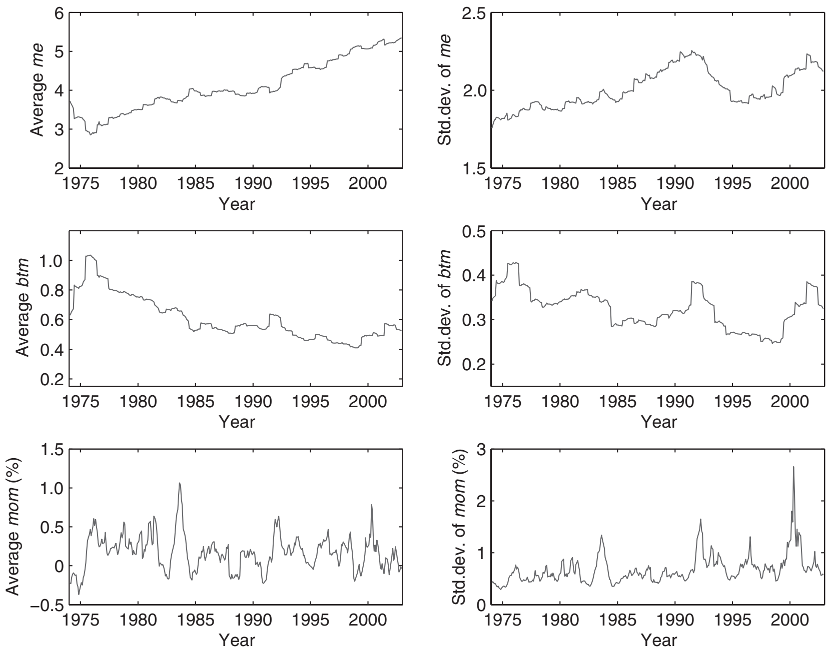

Figure 1
Summary statistics of characteristics
The figure displays cross-sectional means and standard deviations of the firm characteristics *me*, *btm*, and *mom* in every month from January 1974 to December 2002. For each month and firm, the characteristics are *me*, defined as the log of market equity, *btm*, defined as the log of one plus the ratio of book equity divided by market equity, and *mom*, defined as the lagged twelve-month return. The reported means and standard deviations are computed across firms at each point in time.

standard deviations. Recall that the characteristics enter the portfolio policy function in a standardized form. The plots in Figure 1 can be used to translate given values of the standardized characteristics at a particular date in the sample into the original characteristics at the same date.

We use the one-month Treasury bill rate as the risk-free rate. In an extension of our basic approach, we model the coefficients of the portfolio policy as functions of the term spread (*tsp*), defined as the difference in the yields to maturity of the ten-year Treasury note and the one-year Treasury bill. Monthly interest rate data is obtained from the DRI database for the same sample period as the stock data.

2.2 Base case

Table 1 presents the results for the base case, in which the over- or underweighting of each stock, relative to the value-weighted market portfolio, depends on the firm's market capitalization, book-to-market ratio, and lagged one-year return, using the policy function in Equation (3). The table is divided into four sections describing separately the (i) parameter estimates and standard

Table 1
Simple linear portfolio policy

Variable	VW	EW	In Sample PPP	Out of Sample PPP
θ_{me}	–	–	−1.451	−1.124
Std. err.	–	–	(0.548)	(0.709)
θ_{btm}	–	–	3.606	3.611
Std. err.	–	–	(0.921)	(1.110)
θ_{mom}	–	–	1.772	3.057
Std. err.	–	–	(0.743)	(0.914)
LRT p-value	–	–	0.000	0.005
$\lvert w_i \rvert \times 100$	0.023	0.023	0.083	0.133
max $w_i \times 100$	3.678	0.023	3.485	4.391
min $w_i \times 100$	0.000	0.023	−0.216	−0.386
$\sum w_i I(w_i < 0)$	0.000	0.000	−1.279	−1.447
$\sum I(w_i \leq 0)/N_t$	0.000	0.000	0.472	0.472
$\sum \lvert w_{i,t} - w_{i,t-1} \rvert$	0.097	0.142	0.990	1.341
CE	0.064	0.069	0.175	0.118
$\bar{r}$	0.139	0.180	0.262	0.262
$\sigma(r)$	0.169	0.205	0.188	0.223
SR	0.438	0.564	1.048	0.941
α	–	–	0.174	0.177
β	–	–	0.311	0.411
$\sigma(\epsilon)$	–	–	0.181	0.214
IR	–	–	0.960	0.829
me	2.118	−0.504	−0.337	−0.029
btm	−0.418	0.607	3.553	3.355
mom	0.016	0.479	1.623	2.924

This table shows estimates of the portfolio policy with three characteristics: size (me), book-to-market ratio (btm), and momentum (mom), as specified in Equation (3) and optimized for a power utility function with a relative risk aversion of five. We use data from the merged CRSP–Compustat database from January 1964 to December 2002. In the "out-of-sample" results, we use data until December 1973 to estimate the coefficients of the portfolio policy and then form out-of-sample monthly portfolios using those coefficients in the next year. Every subsequent year, we reestimate the portfolio policy by enlarging the sample. All statistics are reported for the period January 1974 to December 2002. The columns labeled "VW," "EW," and "PPP" display statistics of the market-capitalization-weighted portfolio, the equally weighted portfolio, and the optimal parametric portfolio policy, respectively. The first set of rows shows the estimated coefficients of the portfolio policy with bootstrapped standard errors in parentheses. The bootstrapped p-value of the Wald test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. The "out-of-sample" results display time-series averages of coefficients, standard errors, and p-values. The second set of rows shows statistics of the portfolio weights averaged across time. These statistics include the average absolute portfolio weight, the average minimum and maximum portfolio weights, the average sum of negative weights in the portfolio, the average fraction of negative weights in the portfolio, and the turnover in the portfolio. The third set of rows displays average portfolio return statistics: certainty-equivalent return, average return, standard deviation, and Sharpe ratio of returns; the alpha, beta, and volatility of idiosyncratic shocks of a market model regression; and the information ratio. The final set of rows displays the average normalized characteristics of the portfolio. The average risk-free rate in the sample is 0.061 (annualized).

errors, (ii) distribution of the portfolio weights, (iii) properties of the optimized portfolio returns, and (iv) average characteristics of the portfolio. This format is the same for all tables in the paper. The sample goes from January 1974 to December 2002 since we lose the first ten years of data to estimate the initial portfolio for the out-of-sample experiments.

The first few rows of Table 1 present the estimated coefficients of the portfolio policy along with their standard errors estimated from 1000 bootstrapped samples.[13] In the third column, the deviations of the optimal weights from the benchmark weights decrease with the firm's market capitalization (size) and increase with both the firm's book-to-market ratio (value) and its lagged one-year return (momentum). The signs of the estimates are consistent with the literature. The investor overweights small firms, value firms, and past winners and underweights large firms, growth firms, and past losers. Since the characteristics are standardized cross-sectionally, the magnitudes of the coefficients can be compared to each other. Quantitatively, a high book-to-market ratio leads to the largest overweighting of a stock. All three coefficients are highly significant. We also test whether all three coefficients are jointly equal to zero using a Wald test, and the bootstrapped p-value of this test is reported in the row labeled "Wald p-value."[14]

The next few rows of Table 1 describe the weights of the optimized portfolio (in the second column) and compare them to the weights of the market portfolio (in the first column) and the equal-weighted portfolio (in the second column). The average absolute weight of the optimal portfolio is about four times that of the market (0.08% versus 0.02%). Not surprisingly, the active portfolio takes larger positions; however, these positions are not extreme. The average (over time) maximum and minimum weights of the optimal portfolio are 3.49% and −0.22%, respectively, while the corresponding extremes for the market portfolio are 3.68% and 0.00%. The average sum of negative weights in the optimal portfolio is −128%, which implies that the sum of long positions is on average 228%. Finally, the average fraction of negative weights (shorted stocks) in the optimal portfolio is 0.47. Overall, the optimal portfolio does not reflect unreasonably extreme bets on individual stocks and could well be implemented by a combination of an index fund that reflects the market and a long–short equity hedge fund. Finally, one might suspect that the optimal portfolio policy requires unreasonably large trading activity. Fortunately, this is not the case. The average turnover (measured using Equation (18) as the sum of one-way trades) of the optimized portfolio is 99% per year, as compared to an average turnover of 9.7% per year for the market portfolio (due to new listings, delistings, equity issues, etc.) and 14.2% per year for the equal-weighted portfolio. This further shows that the optimal portfolio is eminently implementable and that the returns are unlikely to be affected much by trading costs. Of course, the low turnover is a result of using persistent variables. Using variables that changed more through time would undoubtedly result in higher turnover.

The following rows of Table 1 characterize the performance of the optimal portfolio relative to the market and the equal-weighted portfolios. For ease of interpretation, all measures are annualized. The optimal portfolio has a volatility slightly larger than that of the market portfolio but lower than the

[13] We use bootstrapped standard errors since they produce slightly more conservative tests (larger standard errors) than using estimates of the asymptotic covariance matrix in Equation (12).

[14] When the bootstrapped p-value from the Wald test is less than 0.001, we report it as 0.000.

equal-weighted portfolio (18.8%, 16.9%, and 20.5%, respectively). The optimal portfolio policy has a much higher average return of 26.2% as opposed to 13.9% for the market and 18.0% for the equal-weighted portfolio. This translates into a Sharpe ratio that is more than twice the Sharpe ratio of the market or the equal-weighted portfolio. The certainty equivalent captures the impact of the entire distribution of returns according to the risk preferences of the investor and is therefore the measure that best summarizes performance. The optimal portfolio policy offers a certainty-equivalent gain of roughly 11% relative to the market or the equal-weighted portfolios. We can use a regression of the excess returns of the active portfolio on the excess return of the market to evaluate the active portfolio's alpha, market beta, and residual risk, and then use these statistics to compute the portfolio's information ratio. The alpha of the portfolio is over 17% with a low market beta of only 0.31.[15] Dividing the alpha by the residual volatility of 18.1% produces an information ratio of 0.96. Finally, a word of caution. We should point out that it is not very surprising that the optimal portfolio outperforms the market because we are optimizing in sample and have chosen characteristics that are known to be associated with substantial risk-adjusted returns.

We can decompose the optimal portfolio returns into the market return and the return on a long–short equity hedge fund along the lines of Equation (10). The average return of this hedge fund is found to be 12.27% (not shown in the table). We can further decompose the hedge fund return as $r_h = q(r_h^+ - r_h^-)$, where r_h^+ is the return on the long part of the hedge fund and r_h^- is the return on the short part, both normalized such that the sum of their weights is one. In this way, q captures the leverage of the long–short portfolio. The average r_h^+ is 20.79% and the average r_h^- is 14.01% so that the return of the hedge fund without leverage, i.e., with one dollar long and one dollar short positions, is 6.78%. These returns compare with the market's return of 11.96% over the same period. We therefore see that the long side of the hedge outperforms the market whereas the short side has roughly the same performance as the market. In fact, the short side could be replaced with a short position in the market portfolio without hurting performance. This is important since it is obviously easier to short the market using futures than it is to hold a short portfolio of stocks. The average return of the entire hedge fund of 12.27% and the returns of the scaled long and short parts imply a leverage q of the long and short positions of the order of 173%.

To describe the composition of the optimized portfolio, we compute for every month the weighted characteristics of the portfolio as $N_t \sum_{i=1}^{N_t} w_{i,t}\hat{x}_{i,t}$. The last three rows of the table compare the average (through time) weighted characteristics of the optimized portfolio to those of the market portfolio. The

[15] The Fama–French three-factor model alpha is 9% and the four-factor alpha, including momentum, is 2.5%. This is consistent with our approach producing a portfolio that loads systematically on the size, value, and momentum factor.

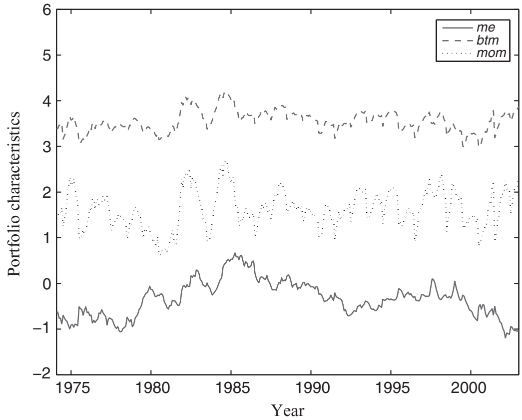

Figure 2
Portfolio characteristics over time
The figure displays the portfolio characteristics of policy function (3) using size, book-to-market, and momentum as firm-specific characteristics. The estimates of θ are obtained using all available CRSP–Compustat stocks from January 1974 to December 2002. The utility function is specified with $\gamma = 5$. The average values of these characteristics are reported in the last three lines of Table 1.

market portfolio has a bias toward very large firms (due to value weighting) and firms with below-average book-to-market ratios (growth), while it is neutral with respect to momentum. In contrast, the optimized portfolio has a slight bias toward small firms and much stronger biases toward high book-to-market ratio (value) firms and past winners. Specifically, the portfolio's book-to-market ratio is more than three standard deviations above the average stock, and the portfolio's momentum is close to two standard deviations above the mean.[16]

Figure 2 plots the time series of the three portfolio characteristics. The characteristics vary over time, but their variability is relatively small and they appear stationary. Moreover, the book-to-market characteristic is always larger than the momentum characteristic, which in turn is larger than the size characteristic, indicating that the optimized portfolio reflects consistent bets through time. While this ordering is also clearly captured in the averages reported in the table, it is comforting to note that the results are systematic and not the product of a few outliers.

While the stellar performance of our approach is unlikely to be due to overfitting since we optimize a portfolio with a large number of stocks over a small number of parameters, the most convincing way to establish its robustness is through an out-of-sample experiment. We use the first ten years of data, from January 1964 until December 1973, to estimate the coefficients of the initial portfolio policy. We then use those parameters to form out-of-sample

[16] In a long–short portfolio, this does not necessarily mean that the typical stock has characteristic values of this order of magnitude. For instance, a portfolio that is long 200% in stocks with a characteristic value of two and is short 100% of stocks with a characteristic value of one has an average characteristic value of three.

monthly portfolios during 1974. At the end of 1974 and of every subsequent year, we reestimate the portfolio policy by enlarging the sample and apply it in every month of the following year. In this way, we estimate the policy with a "telescoping" sample and always apply it out of sample.[17] The standard errors presented are the time-series average of the standard errors from each estimation of the optimal policy.

The out-of-sample results of our parametric portfolio policy are presented in the last column of Table 1. The coefficients on the characteristics are roughly similar to the in-sample estimates, with only an increase of the importance of momentum. All coefficients are still statistically significant, both individually and jointly. The in- and out-of-sample portfolios are also remarkably similar in terms of the distribution of the portfolio weights, consistent with the fact that at least the value and momentum anomalies have been fairly stable throughout our sample period. More importantly, there is not a large deterioration in the return statistics. The certainty equivalent of the portfolio policy is now 11.8%, halfway between those of the market portfolio and of the in-sample policy. The out-of-sample comparison with the equal-weighted portfolio is of particular interest since DeMiguel, Garlappi, and Uppal (2007) have shown that the equal-weighted portfolio generally offers a good compromise between efficiency and robustness out of sample. Our approach substantially improves the efficiency of the portfolio without a significant loss in terms of out-of-sample robustness. We conclude from these results that our approach is likely to perform almost as well out of sample or in real time as our in-sample analysis suggests.

We showed in Equation (10) that the linear portfolio policy is similar to a static choice between long–short portfolios like those constructed by Fama and French (1993) and Carhart (1997). We construct size, book-to-market, and momentum factors based on single sorts of all stocks based on these variables. The factor is then constructed by taking equal-weighted long positions in the stocks belonging to the top 30% and short positions in the stocks in the bottom 30%. The definition of the size, book-to-market, and momentum variables used in the sorts are the same as throughout the rest of the paper. The sample of firms is also the same. Our approach is a little different from the way Fama and French construct these factors, which relies on double sorts on size and book-to-market. We do not follow their approach since that would be equivalent to having interaction terms in the linear policy and would make comparisons more difficult. Then, we simply find the weights on each of the three long–short portfolios that maximize the CRRA utility.[18] Table 2 shows the results. Overall, and as expected, the results are quite similar to the results in Table 1.

[17] The results are not totally out of sample to the extent that the stock characteristics used were known by us to have significant explanatory power for the cross-section of stocks during the entire sample period. There are no simple ways to correct this snooping bias.

[18] Pastor (2000); Pastor and Stambaugh (2000); and Lynch (2001), among others, study the optimal allocation to the Fama–French portfolios.

Table 2
Fama–French portfolios

Variable	VW	In sample FF	Out of sample FF
θ_{me}	–	−0.310	−0.102
Std. err.	–	(0.211)	(0.228)
θ_{btm}	–	0.667	1.190
Std. err.	–	(0.319)	(0.331)
θ_{mom}	–	0.506	0.849
Std. err.	–	(0.186)	(0.194)
LRT p-value	–	0.002	0.006
$\lvert w_i \rvert \times 100$	0.023	0.030	0.049
$\max\ w_i \times 100$	3.678	4.596	6.694
$\min\ w_i \times 100$	0.000	−0.517	−2.167
$\sum w_i I(w_i < 0)$	0.000	−0.146	−0.204
$\sum I(w_i \leq 0)/N_t$	0.000	0.403	0.388
$\sum \lvert w_{i,t} - w_{i,t-1} \rvert$	0.097	0.328	0.484
CE	0.064	0.129	0.095
$\bar{r}$	0.139	0.216	0.240
$\sigma(r)$	0.169	0.178	0.222
SR	0.438	0.847	0.805
α	–	0.104	0.148
β	–	0.627	0.524
$\sigma(\epsilon)$	–	0.143	0.206
IR	–	0.729	0.721
me	2.118	2.311	1.935
btm	−0.418	−0.063	0.286
mom	0.016	0.243	0.398

This table shows results for combinations of the market and three long–short portfolios constructed along the lines of Fama and French, sorted according to size, book-to-market, and momentum, optimized for a power utility function with a relative risk aversion of five. We use data from the merged CRSP–Compustat database from January 1964 to December 2002. In the "out-of-sample" results, we use data until December 1973 to estimate the coefficients of the portfolio policy and then form out-of-sample monthly portfolios using those coefficients in the next year. Every subsequent year, we reestimate the portfolio policy by enlarging the sample. All statistics are reported for the period January 1974 to December 2002. The columns labeled "VW" and "FF" display statistics of the market-capitalization-weighted portfolio and the optimal combination of the market with the long–short portfolios, respectively. The first set of rows shows the estimated coefficients of the portfolio policy with bootstrapped standard errors in parentheses. The bootstrapped p-value of the Wald test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. The "out-of-sample" results display time-series averages of coefficients, standard errors, and p-values. The second set of rows shows statistics of the portfolio weights averaged across time. These statistics include the average absolute portfolio weight, the average minimum and maximum portfolio weights, the average sum of negative weights in the portfolio, the average fraction of negative weights in the portfolio, and the turnover in the portfolio. The third set of rows displays average portfolio return statistics: certainty-equivalent return, average return, standard deviation, and Sharpe ratio of returns; the alpha, beta, and volatility of idiosyncratic shocks of a market model regression; and the information ratio. The final set of rows displays the average normalized characteristics of the portfolio. The average risk-free rate in the sample is 0.061 (annualized).

The differences between the two tables are due to the fact that the Fama–French factors put a weight on each stock proportional to the firm's market capitalization, whereas our linear policy puts a weight that is proportional to the firm's characteristic. Of course, we could easily construct long–short portfolios like those of Fama and French where the weights are proportional to the characteristics. In that case, using our simple linear policy would give exactly the same results as a choice between the factor portfolios.

Notice that the relative differences between our approach and investing in the Fama–French portfolios carry over from the in-sample analysis to the out-of-sample results presented in the last columns of both tables. Weighting stocks by their characteristics, as opposed to equally weighting the top and bottom one-third, improves the in-sample certainty equivalent by 36% and the out-of-sample certainty equivalent by 24%.

2.3 Extensions

2.3.1 Portfolio weight constraints. A large majority of equity portfolio managers face short-sale constraints. In Table 3, we present the results from estimating the long-only portfolio policy specified in Equation (16), again both in and out of sample. As in the unconstrained case, the deviation of the optimal weight from the market portfolio weight decreases with the firm's size, increases with its book-to-market ratio, and increases with its one-year lagged return. Focusing on the portfolio involving the entire universe of stocks, a high book-to-market ratio and large positive one-year lagged return are less desirable characteristics for a long-only investor. The coefficients associated with both of these characteristics are lower in magnitude than in the unrestricted case and are only marginally significant, whereas the coefficient associated with the market capitalization of the firm is not significant. Overall, the significance of the θ coefficients is substantially diminished compared to the unconstrained base case.

The optimal portfolio still does not involve extreme weights. In fact, the average maximum weight of the optimal portfolio is only 1.95%, which is actually lower than that of the market portfolio. On average, the optimal portfolio invests in only 54% of the stocks. The resulting mean and standard deviation of the portfolio return are 19.1% and 18.3%, respectively, translating into a certainly equivalent gain of 3.9% relative to holding the market portfolio. The alpha, beta, and information ratio of the portfolio are 6.2%, 0.86, and 0.56, respectively. These statistics are quite remarkable, given the long-only constraint. Out of sample, the certainty equivalent is 8.1%, showing some small deterioration relative to the in-sample optimum.

The average size of the firms in the optimal portfolio is greater than the size of the average firm but significantly lower than that of the value-weighted market portfolio. The book-to-market ratio and momentum characteristics are less than one standard deviation above those of the average stock and are also significantly different from those of the market portfolio. The results for the optimal long-only portfolio in the universe of the top 500 stocks are qualitatively similar.

The most interesting comparison is between the long-only portfolio in Table 3 and the unconstrained base case in Table 1. The difference in performance is due to two related factors. First, the unconstrained portfolio can exploit both positive and negative forecasts, while the constrained portfolio can only exploit the positive forecasts. Consistent with this argument, the fraction of short

Table 3
Long-only portfolio policy

Variable	VW	In sample PPP	Out of sample PPP
θ_{me}	–	−1.277	0.651
Std. err.	–	(1.217)	(1.510)
θ_{btm}	–	3.215	2.679
Std. err.	–	(1.131)	(1.417)
θ_{mom}	–	1.416	3.780
Std. err.	–	(1.213)	(1.505)
LRT p-value	–	0.045	0.062
$\lvert w_i \rvert \times 100$	0.023	0.023	0.035
$\max\ w_i \times 100$	3.678	1.674	1.952
$\min\ w_i \times 1000$	0.000	0.000	0.000
$\sum w_i I(w_i < 0)$	0.000	0.000	0.000
$\sum I(w_i \leq 0)/N_t$	0.000	0.464	0.464
$\sum \lvert w_{i,t} - w_{i,t-1} \rvert$	0.097	0.241	0.324
CE	0.064	0.103	0.081
$\bar{r}$	0.139	0.191	0.177
$\sigma(r)$	0.169	0.183	0.187
SR	0.438	0.690	0.618
α	–	0.062	0.057
β	–	0.862	0.943
$\sigma(\epsilon)$	–	0.111	0.094
IR	–	0.561	0.601
me	2.118	0.070	0.634
btm	−0.418	0.985	0.345
mom	0.016	0.396	1.106

This table shows estimates of the portfolio policy with long-only weights in Equation (16) with three characteristics: size (me), book-to-market ratio (btm), and momentum (mom), optimized for a power utility function with a relative risk aversion of five. We use data from the merged CRSP–Compustat database from January 1964 to December 2002. In the "out-of-sample" results, we use data until December 1973 to estimate the coefficients of the portfolio policy and then form out-of-sample monthly portfolios using those coefficients in the next year. Every subsequent year, we reestimate the portfolio policy by enlarging the sample. All statistics are reported for the period January 1974 to December 2002. The columns labeled "VW" and "PPP" display statistics of the market-capitalization-weighted portfolio and the optimal parametric portfolio policy, respectively. The first set of rows shows the estimated coefficients of the portfolio policy with bootstrapped standard errors in parentheses. The bootstrapped p-value of the Wald test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. The "out-of-sample" results display time-series averages of coefficients, standard errors, and p-values. The second set of rows shows statistics of the portfolio weights, averaged across time. These statistics include the average absolute portfolio weight, the average minimum and maximum portfolio weights, the average sum of negative weights in the portfolio, the average fraction of negative weights in the portfolio, and the turnover in the portfolio. The third set of rows displays average portfolio return statistics: certainty-equivalent return, average return, standard deviation, and Sharpe ratio of returns; the alpha, beta, and volatility of idiosyncratic shocks of a market model regression; and the information ratio. The final set of rows displays the average normalized characteristics of the portfolio. The average risk-free rate in the sample is 0.061 (annualized).

positions in Table 1 is roughly the same as the fraction of stocks not held by the long portfolio in Table 3. Second, the unconstrained portfolio benefits from using the short positions as leverage to increase the exposure to the long positions.

Interestingly, the tests for joint significance of all three parameters have a p-value around 5%. We therefore cannot reject that the coefficients are jointly zero and that the investor is equally well off holding the market as holding the optimal portfolio. This rejection is consistent with the increase in the standard errors on the coefficients and the smaller gain in certainty equivalent of the restricted optimal portfolio relative to the market. We conclude that short sales constraints have some power in explaining the size, value, and momentum anomalies. An interesting consequence is that market frictions that have constrained investors' ability to short sell stocks (that were more prevalent in the past but that still have an impact) may have limited the arbitraging of the anomalies.

2.3.2 Time-varying coefficients. In Table 4, we allow the coefficients of the portfolio policy to depend on the slope of the yield curve. We estimate different coefficients for months when the yield curve at the beginning of the month is positively sloped (normal) and negatively sloped (inverted). Since inverted yield curves tend to be associated with recessions, letting the portfolio coefficients vary with the yield-curve slope allows the effect of the characteristics on the joint distribution of returns to be different during expansionary and contractionary periods.

We present both in- and out-of-sample results in Table 4. In both cases, the most dramatic effect of conditioning on the slope of the yield curve is on the role of the size of the firm. When the yield curve is upward sloping, the optimal portfolio is tilted toward smaller firms, just as in the base case. When the yield curve is downward sloping, in contrast, the tilt is exactly the opposite, with a positive coefficient (although not statistically different from zero). This is consistent with the common notion that small firms are more affected by economic downturns than larger and more diversified firms. For book-to-market and momentum, the coefficients are generally larger in magnitude when the yield curve slopes down.

Conditioning on the slope of the yield curve does not significantly alter the distribution of the optimal portfolio weights; however, the performance of the portfolio is improved. Both in and out of sample, the portfolios have higher average returns, certainly equivalents, alphas, and information ratios, than without conditioning.

The average characteristics of the optimal portfolios are the most interesting to analyze. Consider the in-sample case. As suggested by the coefficient estimates, the optimal portfolio is tilted toward small stocks when the yield curve is upward sloping. When the yield curve is downward sloping, the portfolio is tilted toward larger stocks and resembles closely the composition of the market portfolio. The average book-to-market and momentum characteristics are both positive and larger when the yield slope is positive. It is interesting to note that although the theta coefficient on book-to-market with an inverted yield curve is very different from the corresponding coefficient in sample, there is no corresponding change in the average characteristic of the portfolio. Intuitively, this arises from the joint distribution of the characteristics conditional on the slope of the yield curve.

Table 4
Conditioning on the slope of the yield curve

Variable	VW	In sample PPP	Out of sample PPP
$\theta_{me\times I(tsp>0)}$	–	−2.168	−1.844
Std. err.	–	(0.706)	(0.745)
$\theta_{me\times I(tsp\leq 0)}$	–	1.684	3.186
Std. err.	–	(1.196)	(1.207)
$\theta_{btm\times I(tsp>0)}$	–	3.197	3.146
Std. err.	–	(1.102)	(1.121)
$\theta_{btm\times I(tsp\leq 0)}$	–	5.830	0.037
Std. err.	–	(2.061)	(0.879)
$\theta_{mom\times I(tsp>0)}$	–	2.023	4.489
Std. err.	–	(0.909)	(1.597)
$\theta_{mom\times I(tsp\leq 0)}$	–	3.705	3.598
Std. err.	–	(1.611)	(1.108)
LRT p-value	–	0.000	0.000
$\|w_i\| \times 100$	0.023	0.091	0.136
max $w_i \times 100$	3.678	3.489	4.392
min $w_i \times 1000$	0.000	−2.619	−0.398
$\sum w_i I(w_i < 0)$	0.000	−1.428	−1.526
$\sum I(w_i \leq 0)/N_t$	0.000	0.476	0.476
$\sum \|w_{i,t} - w_{i,t-1}\|$	0.097	1.295	1.510
CE	0.064	0.194	0.120
$\bar{r}$	0.139	0.293	0.277
$\sigma(r)$	0.169	0.205	0.236
SR	0.438	1.114	0.932
α	–	0.209	0.197
β	–	0.252	0.319
$\sigma(\epsilon)$	–	0.201	0.231
IR	–	1.042	0.851
$me \times \mathrm{I}(tsp > 0)$	1.748	−0.744	−0.430
$me \times \mathrm{I}(tsp \leq 0)$	0.370	0.351	0.129
$btm\ \mathrm{I}(tsp > 0)$	−0.342	2.782	2.544
$btm \times \mathrm{I}(tsp \leq 0)$	−0.076	0.789	0.880
$mom\ \mathrm{I}(tsp > 0)$	0.031	1.583	2.381
$mom \times \mathrm{I}(tsp \leq 0)$	−0.015	0.529	0.740

This table shows estimates of the portfolio policy with the product of three characteristics: size (me), book-to-market ratio (btm), and momentum (mom), and an indicator function of the sign of the slope of the yield curve, optimized for a power utility function with a relative risk aversion of five. We use data from the merged CRSP–Compustat database from January 1964 to December 2002. In the "out-of-sample" results, we use data until December 1973 to estimate the coefficients of the portfolio policy and then form out-of-sample monthly portfolios using those coefficients in the next year. Every subsequent year, we reestimate the portfolio policy by enlarging the sample. All statistics are reported for the period January 1974 to December 2002. The columns labeled "VW" and "PPP" display statistics of the market-capitalization-weighted portfolio and the optimal parametric portfolio policy, respectively. The first set of rows shows the estimated coefficients of the portfolio policy with bootstrapped standard errors in parentheses. The bootstrapped p-value of the Wald test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. The "out-of-sample" results display time-series averages of coefficients, standard errors, and p-values. The second set of rows shows statistics of the portfolio weights averaged across time. These statistics include the average absolute portfolio weight, the average minimum and maximum portfolio weights, the average sum of negative weights in the portfolio, the average fraction of negative weights in the portfolio, and the turnover in the portfolio. The third set of rows displays average portfolio return statistics: certainty-equivalent return, average return, standard deviation, and Sharpe ratio of returns; the alpha, beta, and volatility of idiosyncratic shocks of a market model regression; and the information ratio. The final set of rows displays the average normalized characteristics of the portfolio. The average risk-free rate in the sample is 0.061 (annualized).

Table 5
Varying risk aversion

Variable	VW	In sample PPP $\gamma = 1$	In sample PPP $\gamma = 5$	In sample PPP $\gamma = 100$	Out of sample PPP $\gamma = 1$	Out of sample PPP $\gamma = 5$	Out of sample PPP $\gamma = 100$
θ_{me}	–	−6.242	−1.124	0.026	−7.178	0.651	0.824
Std. err.	–	(2.882)	(0.548)	(0.223)	(4.579)	(0.709)	(0.212)
θ_{btm}	–	7.864	3.611	5.207	8.450	2.679	3.834
Std. err.	–	(3.546)	(0.921)	(0.314)	(3.435)	(1.110)	(0.346)
θ_{mom}	–	6.452	3.057	0.548	11.991	3.780	−0.105
Std. err.	–	(2.915)	(0.743)	(0.483)	(2.971)	(0.914)	(0.205)
LRT p-value	–	0.000	0.000	0.000	0.000	0.000	0.000
$\lvert w_i \rvert \times 100$	0.023	0.217	0.133	0.102	0.422	0.035	0.114
max $w_i \times 100$	3.678	3.623	4.391	3.585	6.106	1.952	4.621
min $w_i \times 100$	0.000	−0.646	−0.386	−0.238	−1.309	0.000	−0.327
$\sum w_i I(w_i < 0)$	0.000	−4.122	−1.447	−1.670	−5.582	0.000	−1.201
$\sum I(w_i \leq 0)/N_t$	0.000	0.522	0.472	0.477	0.529	0.464	0.460
$\sum \lvert w_{i,t} - w_{i,t-1} \rvert$	0.097	3.010	1.341	0.883	4.797	0.324	0.608
CE	0.064	0.360	0.118	−0.989	0.297	0.081	−1.408
$\bar{r}$	0.139	0.534	0.262	0.223	0.673	0.177	0.147
$\sigma(r)$	0.169	0.589	0.223	0.199	0.847	0.187	0.170
SR	0.438	0.796	0.941	0.796	0.726	0.618	0.557
α	–	0.529	0.177	0.141	0.647	0.057	0.058
β	–	−0.817	0.411	0.229	−0.602	0.943	0.465
$\sigma(\epsilon)$	–	0.573	0.214	0.195	0.842	0.094	0.153
IR	–	0.924	0.829	0.723	0.768	0.601	0.380
me	2.118	−6.234	−0.029	0.633	−7.328	0.634	1.763
btm	−0.418	9.060	3.355	4.761	9.633	0.345	3.105
mom	0.016	6.016	2.924	0.391	11.621	1.106	−0.233

This table shows estimates of the linear portfolio policy (3) with three characteristics: size (*me*), book-to-market ratio (*btm*), and momentum (*mom*) optimized for different power utility functions with a relative risk aversion of one, five (as in previous tables), and 100. We use data from the merged CRSP–Compustat database from January 1964 to December 2002. In the “out-of-sample” results, we use data until December 1973 to estimate the coefficients of the portfolio policy and then form out-of-sample monthly portfolios using those coefficients in the next year. Every subsequent year, we reestimate the portfolio policy by enlarging the sample. All statistics are reported for the period January 1974 to December 2002. The columns labeled “VW” and “PPP” display statistics of the market-capitalization-weighted portfolio and the optimal parametric portfolio policy, respectively. The first set of rows shows the estimated coefficients of the portfolio policy with bootstrapped standard errors in parentheses. The bootstrapped p-value of the Wald test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. The “out-of-sample” results display time-series averages of coefficients, standard errors, and p-values. The second set of rows shows statistics of the portfolio weights averaged across time. These statistics include the average absolute portfolio weight, the average minimum and maximum portfolio weights, the average sum of negative weights in the portfolio, the average fraction of negative weights in the portfolio, and the turnover in the portfolio. The third set of rows displays average portfolio return statistics: certainty-equivalent return, average return, standard deviation, and Sharpe ratio of returns; the alpha, beta, and volatility of idiosyncratic shocks of a market model regression; and the information ratio. The final set of rows displays the average normalized characteristics of the portfolio. The average risk-free rate in the sample is 0.061 (annualized). The certainty-equivalent returns of the market portfolio for $\gamma = 1$, $\gamma = 5$, and $\gamma = 100$ are 0.107, 0.052, and −1.976 (annualized), respectively.

2.3.3 Risk aversion. The optimal portfolio policy depends critically on the investor’s preferences. The results thus far were obtained assuming CRRA utility with relative risk aversion $\gamma = 5$. To get a better sense for the role of this utility assumption, we present in Table 5 in- and out-of-sample results for different levels of risk aversion. In addition to $\gamma = 5$, which we report in the table for comparison, we also estimate the optimal portfolio for $\gamma = 1$, corresponding to the popular case of log utility, and $\gamma = 100$, which is extremely high and makes the investor very sensitive to losses.

For small values of γ, the estimates of the coefficients on the firm's size, book-to-market ratio, and one-year lagged return are all large in absolute value and statistically significant. As the investor becomes more risk averse, the coefficients on size and momentum approach zero. This suggests that these characteristics are associated with both mean returns and risk. As risk aversion increases, the investor weighs more the contribution of these characteristics to risk and loads less heavily on them. In contrast, the exposure to book-to-market does not change qualitatively as risk aversion increases. This indicates that this characteristic is more associated with expected return than risk.

The average firm characteristics exhibit the same patterns. For $\gamma = 1$, the portfolio is severely tilted toward firms that are small, value, and winners. As the level of risk aversion increases, the tilting toward small caps and winners decreases. Actually, for $\gamma = 100$, the portfolio holds companies that are 0.6 standard deviations larger than the mean. However, the tilt toward value firms is maintained. Although increasing risk aversion helps in explaining the size and momentum anomalies, it does not explain the value anomaly.

The distribution of the optimal portfolio weights also changes with the level of risk aversion. In particular, an investor with $\gamma = 1$ takes on more and larger negative positions compared to an investor with $\gamma = 5$. The fraction of shorted stocks is only increased by 5%, but the sum of negative weights is almost three times larger, which implies that the less risk-averse investor takes similar bets but with more leverage. Interestingly, the $\gamma = 100$ investor actually uses higher leverage than the investor with $\gamma = 5$. Intuitively, the short positions help by partially hedging the worst performing stocks in the market's lowest return months.

Not surprisingly, the differences in the optimal portfolio weights translate into equally striking differences in the distribution of the optimized portfolio returns. The average return and volatility are highest for $\gamma = 1$ and decrease with the investor's level of risk aversion. For high levels of γ, the curvature of the utility function is such that the average utility across all months is dominated by the utility obtained in the worst month. In this sense, the $\gamma = 100$ preferences correspond closely to a max–min criterion. The portfolio's minimum return (not shown in the table) decreases from -55.60% for $\gamma = 1$, to -19.50% for $\gamma = 5$, and to -13.07% for $\gamma = 100$ (for comparison, the market's minimum return is -21.49%). Table 5 also presents certainty equivalents for the different levels of risk aversion but they cannot be compared with each other. The certainty equivalent for the $\gamma = 100$ case is actually negative, -0.989, but the investor dislikes the market even more, with a certainty equivalent of -1.976.[19]

[19] The certainty equivalents can take values less than -1 because we are taking the monthly certainty equivalents and multiplying them by 12 to express them in annual terms. Of course, the monthly certainty equivalents cannot be less than -1 no matter the degree of risk aversion.

Comparing the in- and out-of-sample results, it appears that our method performs better out of sample for lower levels of risk aversion than for higher levels of risk aversion. This can be seen equivalently from the differences in the policy coefficients, in the performance measures, or in the portfolio characteristics. For example, with $\gamma = 1$, the in-sample average portfolio characteristics are $me = -6.234$, $btm = 9.060$, and $mom = 6.016$; the corresponding out-of-sample average portfolio characteristics are -7.328, 9.633, and 11.621. In contrast, with $\gamma = 100$, the relative differences between the in- and out-of-sample characteristics are much greater: *me* of 0.633 versus 1.763, *btm* of 4.761 versus 3.105, and *mom* of 0.391 versus -0.233. Apparently, risk-minimizing portfolios are less stable than expected return-maximizing portfolios.

2.4 Transaction costs

In this section, we examine the impact of transaction costs on optimal trading policies. As a first approach, we take one-way transaction costs to be constant through time and in the cross-section at 0.5%. However, it is well known that transaction costs vary considerably across stocks, being larger for small caps than for large caps, and have been gradually decreasing over time. This has been noted by Keim and Madhavan (1997); Domowitz, Glen, and Madhavan (2001); and Hasbrouck (2006). To accommodate these empirical facts, we use a second specification for transaction costs, which allows for cross-sectional variation and captures the declining costs over time. The one-way cost of company i at time t is $c_{i,t} = z_{i,t} * T_t$. The variable $z_{i,t} = 0.006 - 0.0025 \times me_{i,t}$ captures the cross-sectional variation in transaction costs with the market cap of the firms. In this specification, $me_{i,t}$ measures the relative size of company i at time t, normalized to be between 0 and 1. In other words, the smallest company has a transaction cost of 0.6%, whereas the largest one has a cost of 0.35%. This is consistent with previous estimates (e.g., Keim and Madhavan 1997 and Hasbrouck 2006). We capture the declining transaction costs over the sample by assuming a trend T_t such that costs in 1974 are four times larger than those in 2002. This again is consistent with the Domowitz, Glen, and Madhavan (2001) and Hasbrouck (2006) papers. The average $c_{i,t}$ at the end of the sample is 0.5%, which is directly comparable to the constant cost case. Figure 3 displays the median, minimum, and maximum trading costs over time.

Table 6 presents the results for the simple linear policy (3) optimized with the two transaction cost scenarios described above. For comparison, we also include the case with no transaction costs from Table 1. We see that transaction costs lead to a slight decrease in the absolute value of the optimal thetas, reflecting the overall higher cost of trading. Note that in the limit with zero thetas, the policy would be equal to the market portfolio and there would be substantially less trading. Indeed, we see that turnover goes down with the increasing levels of trading costs. The effect is not very pronounced because the variables we are using in the portfolio policy are very persistent and induce relatively low levels of turnover compared to the high expected returns they

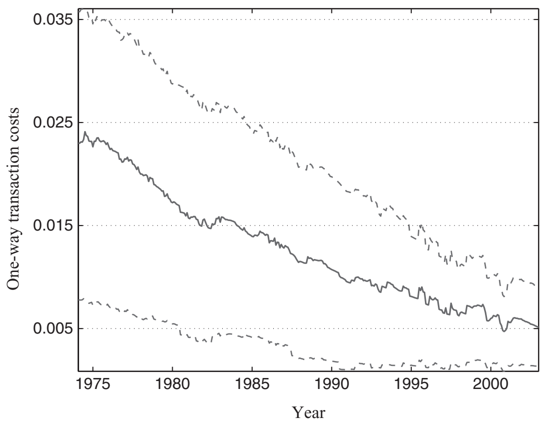

Figure 3
Varying transaction costs
The figure displays the transaction costs described in Section 1.4. The transaction costs are assumed to decline uniformly over time and to decrease with the relative size of the firms. The solid line is the median transaction cost across all stocks over time. The dashed lines are the minimum and maximum transaction costs over time.

generate. With a turnover of 100% per year, average costs of 0.5% generate a trading cost of only 0.5%.

Table 7 shows the results obtained by applying the policy with a no-trade boundary developed in Section 1.4. This policy leads to roughly 30% less turnover than the simple linear policy with a substantial increase in certainty equivalent. This is especially remarkable out of sample, where there is virtually no loss in performance. Intuitively, this is due to the smoothing features of the autoregressive policy. By making weights less volatile through time, this policy essentially becomes more robust out of sample. Note that the weight put on the portfolio from the previous month (α_t) is on average slightly above 50%, reflecting the substantial inertia in the optimal portfolio policy.

Comparing the in-sample results in the first few columns of Tables 6 and 7 to their respective out-of-sample counterparts in the last three columns of each table, we observe that the effect of incorporating transaction costs is roughly equivalent in and out of sample. The underlying reason is that the in- and out-of-sample policies are very similar in the base case. Therefore, the turnover is very similar in the base case, and the effect of incorporating this turnover and the resulting transaction costs are also very similar.

3. Conclusion

We proposed a novel approach to optimizing large-scale equity portfolios. The portfolio weight in each stock is modeled as a function of the firm's characteristics, such as its market capitalization, book-to-market ratio, and lagged return. The coefficients of this function are found by optimizing the investor's average utility of the portfolio's return over a given sample period.

Table 6
Simple portfolio policy with transactions costs

Variable	VW	In-sample PPP $c_{i,t}=$ 0.000	0.005	$f(me_{i,t},t)$	Out-of-sample PPP $c_{i,t}=$ 0.000	0.005	$f(me_{i,t},t)$
θ_{me}	–	−1.451	−1.391	−1.167	−1.124	−1.105	−0.925
Std. err.	–	(0.548)	(0.547)	(0.550)	(0.709)	(0.849)	(0.780)
θ_{btm}	–	3.606	3.557	3.160	3.611	3.618	3.468
Std. err.	–	(0.921)	(0.922)	(0.924)	(1.110)	(1.125)	(1.305)
θ_{mom}	–	1.772	1.651	1.307	3.057	3.028	2.497
Std. err.	–	(0.743)	(0.741)	(0.745)	(0.914)	(0.952)	(0.961)
LRT p-value	–	0.000	0.000	0.000	0.005	0.006	0.008
$\|w_i\| \times 100$	0.023	0.083	0.082	0.074	0.133	0.133	0.084
max $w_i \times 100$	3.678	3.485	3.491	3.508	4.391	4.392	3.535
min $w_i \times 100$	0.000	−0.216	−0.208	−0.183	−0.386	−0.385	−0.233
$\sum w_i I(w_i < 0)$	0.000	−1.279	−1.240	−1.074	−1.447	−1.444	−1.301
$\sum I(w_i \leq 0)/N_t$	0.000	0.472	0.471	0.463	0.472	0.472	0.464
$\sum \|w_i - w_i^h\|$	0.097	0.990	0.942	0.788	1.341	1.333	1.087
CE$\|c_{i,t} = 0.000$	0.064	0.175	0.175	0.173	0.118	0.123	0.122
CE$\|c_{i,t} = 0.005$	–	0.169	0.170	0.168	0.115	0.117	0.118
CE$\|c_{i,t} = f(me_{i,t},t)$	–	0.162	0.162	0.167	0.119	0.120	0.125
$\bar{r}$	0.139	0.262	0.252	0.243	0.262	0.248	0.244
$\sigma(r)$	0.169	0.188	0.183	0.179	0.223	0.220	0.217
SR	0.438	1.048	1.021	0.978	0.941	0.889	0.880
α	–	0.174	0.162	0.151	0.177	0.163	0.152
β	–	0.311	0.328	0.354	0.411	0.416	0.421
$\sigma(\epsilon)$	–	0.181	0.175	0.171	0.214	0.210	0.206
IR	–	0.960	0.929	0.887	0.829	0.774	0.764
me	2.118	−0.337	−0.267	−0.033	−0.029	−0.018	0.252
btm	−0.418	3.553	3.492	3.066	3.355	3.362	3.237
mom	0.016	1.623	1.477	1.279	2.924	2.896	2.369

This table shows estimates of the portfolio policy with three characteristics: size (me), book-to-market ratio (btm), and momentum (mom), as specified in Equation (3) and optimized for a power utility function with a relative risk aversion of five. The utility function is maximized for returns after transaction costs. In the first specification, the proportional transaction costs are 0.5%, constant across stocks and over time. In the second specification, transaction costs vary across stocks and over time as shown in Figure 3. For comparison, we also present results with zero transaction costs. We use data from the merged CRSP–Compustat database from January 1964 to December 2002. In the "out-of-sample" results, we use data until December 1973 to estimate the coefficients of the portfolio policy and then form out-of-sample monthly portfolios using those coefficients in the next year. Every subsequent year, we reestimate the portfolio policy by enlarging the sample. All statistics are reported for the period January 1974 to December 2002. The first set of rows shows the estimated coefficients of the portfolio policy with bootstrapped standard errors in parentheses. The bootstrapped p-value of the Wald test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. The "out-of-sample" results display time-series averages of coefficients, standard errors, and p-values. The second set of rows shows statistics of the portfolio weights, averaged across time. These statistics include the average absolute portfolio weight, the average minimum and maximum portfolio weights, the average sum of negative weights in the portfolio, the average fraction of negative weights in the portfolio, and the turnover in the portfolio. The third set of rows displays average portfolio return statistics: certainty-equivalent returns, average return, standard deviation, and Sharpe ratio of returns; the alpha, beta, and volatility of idiosyncratic shocks of a market model regression; and the information ratio. We compute the certainty-equivalent return for the policy with and without adjustment for transaction costs. The final set of rows displays the average normalized characteristics of the portfolio. The average risk-free rate in the sample is 0.061 (annualized).

We argued that our approach is computationally simple, easily modified and extended, produces sensible portfolio weights, and offers robust performance in and out of sample.

We illustrated many features of our approach through an empirical application to the universe of stocks in the CRSP–Compustat data set. Our empirical

Table 7
Boundary portfolio policy with transactions costs

		In-sample PPP			Out-of-sample PPP		
			$c_{i,t} =$			$c_{i,t} =$	
Variable	VW	0.000	0.005	$f(me_{i,t}, t)$	0.000	0.005	$f(me_{i,t}, t)$
θ_{me}	–	−1.147	−1.133	−0.947	−0.979	−0.946	−0.845
Std. err.	–	(0.561)	(0.537)	(0.523)	(0.577)	(0.571)	(0.568)
θ_{btm}	–	4.432	4.405	4.194	4.264	4.150	4.021
Std. err.	–	(1.137)	(1.124)	(1.077)	(1.153)	(1.232)	(1.247)
θ_{mom}	–	2.366	2.344	2.205	3.587	3.498	3.154
Std. err.	–	(0.964)	(0.846)	(0.824)	(1.007)	(0.943)	(0.948)
$\kappa \times 10^3$	–	0.273	0.282	0.301	0.289	0.294	0.307
Std. err.	–	(0.078)	(0.087)	(0.091)	(0.084)	(0.094)	(0.098)
LRT p-value	–	0.000	0.000	0.000	0.000	0.000	0.000
$\|w_i\| \times 100$	0.023	0.095	0.094	0.089	0.100	0.098	0.097
max $w_i \times 100$	3.678	3.481	3.481	3.495	3.503	3.505	3.504
min $w_i \times 100$	0.000	−0.362	−0.363	−0.356	−0.401	−0.395	−0.398
$\sum w_i I(w_i < 0)$	0.000	−1.516	−1.499	−1.388	−1.636	−1.581	−1.566
$\sum I(w_i \leq 0)/N_t$	0.000	0.478	0.477	0.474	0.476	0.474	0.474
$\sum \|w_i - w_i^h\|$	0.098	0.697	0.676	0.591	0.934	0.888	0.795
CE$\|c_{i,t} = 0.000$	0.064	0.187	0.187	0.187	0.169	0.172	0.176
CE$\|c_{i,t} = 0.005$	–	0.183	0.184	0.183	0.158	0.161	0.159
CE$\|c_{i,t} = f(me_{i,t}, t)$	–	0.177	0.178	0.181	0.160	0.164	0.167
$\bar{r}$	0.139	0.280	0.275	0.266	0.293	0.289	0.284
$\sigma(r)$	0.169	0.195	0.193	0.189	0.223	0.217	0.212
SR	0.438	1.105	1.088	1.065	1.021	1.028	1.032
α	–	0.196	0.190	0.180	0.208	0.202	0.198
β	–	0.266	0.275	0.292	0.272	0.295	0.299
$\sigma(\epsilon)$	–	0.190	0.188	0.183	0.219	0.212	0.206
IR	–	1.028	1.010	0.982	0.950	0.952	0.961
me	2.118	−0.189	−0.164	0.085	0.057	0.123	0.257
btm	−0.418	4.015	3.970	3.670	3.777	3.648	3.469
mom	0.016	1.959	1.927	1.768	3.098	3.002	2.645
avg α	–	0.556	0.572	0.620	0.497	0.514	0.563
min α	–	0.260	0.269	0.300	0.247	0.267	0.294
max α	–	0.828	0.842	0.882	0.740	0.758	0.811

This table shows estimates of the portfolio policy with three characteristics: size (me), book-to-market ratio (btm), and momentum (mom), as specified in Section 1.4 and optimized for a power utility function with a relative risk aversion of five. The utility function is maximized for returns after transaction costs. In the first specification, the proportional transaction costs are 0.5%, constant across stocks and over time. In the second specification, transaction costs vary across stocks and over time as shown in Figure 3. For comparison, we also present results with zero transaction costs. We use data from the merged CRSP–Compustat database from January 1964 to December 2002. In the "out-of-sample" results, we use data until December 1973 to estimate the coefficients of the portfolio policy and then form out-of-sample monthly portfolios using those coefficients in the next year. Every subsequent year, we reestimate the portfolio policy by enlarging the sample. All statistics are reported for the period January 1974 to December 2002. The first set of rows shows the estimated coefficients of the portfolio policy with bootstrapped standard errors in parentheses. The bootstrapped p-value of the Wald test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. The "out-of-sample" results display time-series averages of coefficients, standard errors, and p-values. The second set of rows shows statistics of the portfolio weights averaged across time. These statistics include the average absolute portfolio weight, the average minimum and maximum portfolio weights, the average sum of negative weights in the portfolio, the average fraction of negative weights in the portfolio, and the turnover in the portfolio. The third set of rows displays average portfolio return statistics: certainty-equivalent returns, average return, standard deviation, and Sharpe ratio of returns; the alpha, beta, and volatility of idiosyncratic shocks of a market model regression; and the information ratio. We compute the certainty-equivalent return for the policy with and without adjustment for transaction costs. The final set of rows displays the average normalized characteristics of the portfolio. The average risk-free rate in the sample is 0.061 (annualized).

results document the importance of the firm's market capitalization, book-to-market ratio, and one-year lagged return for explaining deviations of the optimal portfolio for a CRRA investor from the market. Relative to market capitalization weights, the optimal portfolio (with and without short-sale constraints) allocates considerably more wealth to stocks of small firms, firms with high book-to-market ratios (value firms), and firms with large positive lagged returns (past winners). With a relative risk aversion of five, the certainty-equivalent gain from incorporating the firm characteristics, relative to holding the market portfolio, is an annualized 11%. We showed that these results are robust out of sample. Finally, we incorporated transaction costs. We showed that, with a simple policy function that features a no-trade boundary, the portfolio turnover is reduced by up to 50% with only marginal deterioration in performance.

Our idea can easily be applied to other asset classes. We could use a similar approach to form bond portfolios based on bond characteristics (e.g., duration, convexity, coupon rate, credit rating, leverage) or to form currency portfolios based on the characteristics of each country pair (e.g., interest rate and inflation rate differentials, trade balance).

Appendix: Data

For each firm in the CRSP–Compustat data set, we construct several variables at the end of fiscal years 1964 to 2002. The first full year of data, 1963, is used to construct lagged values. The exact fiscal year-end dates are from CRSP. We use the following quantities in the definition of the variables (Compustat data item numbers are in parentheses): total assets (6); liabilities (181); preferred stock value (10, 56, or 130, in that order, otherwise zero); balance sheet deferred taxes and investment tax credits (35, otherwise zero); price per share (from CRSP); and shares outstanding (25, otherwise taken from CRSP). If total assets, liabilities, price, and shares outstanding are missing, the observation is not included in the data set. Then, we define book equity (BE) as equal to total assets minus liabilities plus balance sheet deferred taxes and investment tax credits minus preferred stock value; market equity (ME) as equal to price per share times shares outstanding; book-to-market (*btm*) as equal to the log of one plus book equity divided by market equity. We omit firms with negative book-to-market ratio. Log market equity (*me*) is computed as the log of market equity.

The monthly firm returns are obtained from CRSP. We allow a minimum of a six-month lag between the fiscal year end of the above accounting variables and the returns to ensure that the information from the firms' annual reports would have been publicly available at the time of portfolio formation. From CRSP, we also compute the trailing twelve-month return (*mom*) defined as the monthly compounded return between months $t - 13$ and $t - 2$. After all variables have been created, we eliminate the smallest 20% of firms (i.e., firms in the lowest 20th percentile of *me*).

References

Aït-Sahalia, Y., and M. W. Brandt. 2001. Variable Selection for Portfolio Choice. *Journal of Finance* 56:1297–351.

Akian, M., J. Luis Menaldi, and A. Sulem. 1996. On an Investment-Consumption Model with Transactions Costs. *SIAM Journal of Control and Optimization* 34:329–64.

Alexander, G., and A. Baptista. 2002. Economic Implications of Using a Mean–VaR Model for Portfolio Selection: A Comparison with Mean–Variance Analysis. *Journal of Economic Dynamics and Control* 26:1159–93.

Ang, A., G. Bekaert, and J. Liu. 2005. Why Stocks May Disappoint. *Journal of Financial Economics* 76:471–508.

Asness, C. S., B. Porter, and R. L. Stevens. 2001. Predicting Stock Returns Using Industry-Relative Firm Characteristics. Working Paper, AQR Capital Management.

Basak, S., and A. Shapiro. 2001. Value-at-Risk Based Risk Management: Optimal Policies and Asset Prices. *Review of Financial Studies* 14:371–405.

Benartzi, S., and R. H. Thaler. 1995. Myopic Loss Aversion and the Equity Premium Puzzle. *Quarterly Journal of Economics* 110:73–92.

Black, F., and R. Litterman. 1992. Global Portfolio Optimization. *Financial Analysts Journal* 48:28–43.

Brandt, M. W. 1999. Estimating Portfolio and Consumption Choice: A Conditional Method of Moments Approach. *Journal of Finance* 54:1609–46.

Brandt, M. W. Forthcoming. Portfolio Choice Problems. In Y. Aït-Sahalia and L.P. Hansen (eds.), *Handbook of Financial Econometrics*.

Brandt, M. W., and P. Santa-Clara. 2006. Dynamic Portfolio Selection by Augmenting the Asset Space. *Journal of Finance* 61:2187–217.

Browne, S. 1999. Beating a Moving Target: Optimal Portfolio Strategies for Outperforming a Stochastic Benchmark. *Finance and Stochastics* 3:275–94.

Campbell, J. Y. 1991. A Variance Decomposition for Stock Returns. *Economic Journal* 101:157–79.

Campbell, J. Y., and R. J. Shiller. 1988. Stock Prices, Earnings, and Expected Dividends. *Journal of Finance* 43:661–76.

Carhart, M. M. 1997. On Persistence in Mutual Fund Performance. *Journal of Finance* 52:57–82.

Chan, K. C., J. Karceski, and J. Lakonishok. 1998. The Risk and Return from Factors. *Journal of Financial and Quantitative Analysis* 33:159–88.

Chan, K. C., J. Karceski, and J. Lakonishok. 1999. On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model. *Review of Financial Studies* 12:937–74.

Cooper, M., H. Gulen, and M. Vassalou. 2000. Investing in Size and Book-to-Market Portfolios Using Information About the Macroeconomy: Some New Trading Rules. Working Paper, Purdue University.

Daniel, K. D., and S. Titman. 1999. Market Efficiency in an Irrational World. *Financial Analyst Journal* 55:24–50.

Davis, M. H. A., and A. R. Norman. 1990. Portfolio Selection with Transaction Costs. *Mathematics of Operations Research* 15:676–713.

DeMiguel, V., L. Garlappi, and R. Uppal. Forthcoming. Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *Review of Financial Studies*.

Dixit, A. 1991. A Simplified Treatment of Some Results Concerning Regulated Brownian Motion. *Journal of Economic Dynamics and Control* 15:657–74.

Domowitz, I., J. Glen, and A. Madhavan. 2001. Liquidity, Volatility, and Equity Trading Costs Across Countries and Over Time. *International Finance* 4:221–55.

Dumas, B. 1991. Super Contact and Related Optimality Conditions. *Journal of Economic Dynamics and Control* 15:675–86.

Fama, E. F. 1990. Stock Returns, Expected Returns, and Real Activity. *Journal of Finance* 45:1089–108.

Fama, E. F., and K. R. French. 1988. Permanent and Temporary Components of Stock Prices. *Journal of Political Economy* 96:246–73.

Fama, E. F., and K. R. French. 1989. Business Conditions and Expected Returns on Stocks and Bonds. *Journal of Financial Economics* 25:23–49.

Fama, E. F., and K. R. French. 1993. Common Risk Factors in the Returns of Stocks and Bonds. *Journal of Financial Economics* 33:3–56.

Fama, E. F., and K. R. French. 1996. Multifactor Explanations of Asset Pricing Anomalies. *Journal of Finance* 51:55–84.

Frost, P. A., and J. E. Savarino. 1986. An Empirical Bayes Approach to Efficient Portfolio Selection. *Journal of Financial and Quantitative Analysis* 21:293–305.

Frost, P. A., and J. E. Savarino. 1988. For Better Performance: Constrain Portfolio Weights. *Journal of Portfolio Management* 15:29–34.

Gomes, F. 2005. Portfolio Choice and Trading Volume with Loss-Averse Investors. *Journal of Business* 78:675–706.

Grossman, S. J., and J. Vila. 1989. Portfolio Insurance in Complete Markets: A Note. *Journal of Business* 62:473–76.

Hansen, L. P. 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50:1029–53.

Hasbrouck, J. 2006. Trading Costs and Returns for US Equities: Estimating Effective Costs from Daily Data. Working Paper, NYU.

Hodrick, R. J. 1992. Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement. *Review of Financial Studies* 5:257–86.

Jagannathan, R., and T. Ma. 2002. Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *Journal of Finance* 58:1651–84.

Jobson, J. D., and B. Korkie. 1980. Estimation of Markowitz Efficient Portfolios. *Journal of the American Statistical Association* 75:544–54.

Jobson, J. D., and B. Korkie. 1981. Putting Markowitz Theory to Work. *Journal of Portfolio Management* 7:70–74.

Jorion, P. 1986. Bayes–Stein Estimation for Portfolio Analysis. *Journal of Financial and Quantitative Analysis* 21:279–92.

Kandel, S., and R. F. Stambaugh. 1996. On the Predictability of Stock Returns: An Asset-Allocation Perspective. *Journal of Finance* 51:385–424.

Keim, D. B., and A. Madhavan. 1997. Transactions Costs and Investment Style: An Inter-Exchange Analysis of Institutional Equity Trades. *Journal of Financial Economics* 46:265–92.

Keim, D. B., and R. F. Stambaugh. 1986. Predicting Returns in the Stock and Bond Markets. *Journal of Financial Economics* 17:357–90.

Ledoit, O., and M. Wolf. 2003. Improved Estimation of the Covariance Matrix of Returns with an Application to Portfolio Selection. *Journal of Empirical Finance* 10:603–21.

Ledoit, O., and M. Wolf. 2004. Honey, I Shrunk the Sample Covariance Matrix. *Journal of Portfolio Management* 30:110–19.

Leland, H. E. 2000. Optimal Portfolio Implementation with Transaction Costs and Capital Gains Taxes. Working Paper, University of California–Berkeley.

Lynch, A. W. 2001. Portfolio Choice and Equity Characteristics: Characterizing the Hedging Demands Induced by Return Predictability. *Journal of Financial Economics* 62:67–130.

Magill, M. M., and G. M. Constantinides. 1976. Portfolio Delection with Transaction Costs. *Journal of Economic Theory* 13:245–63.

Markowitz, H. M. 1952. Portfolio Delection. *Journal of Finance* 7:77–91.

Michaud, R. O. 1989. The Markowitz Optimization Enigma: Is Optimized Optimal? *Financial Analyst Journal* 45:31–42.

Newey, W. K., and K. D. West. 1987. A Simple, Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55:703–8.

Nigmatullin, E. A. 2003. Bayesian Model Averaging for Moment Conditions Models. Working Paper, University of Wisconsin–Madison.

Pastor, L. 2000. Portfolio Selection and Asset Pricing Models. *Journal of Finance* 50:179–223.

Pastor, L., and R. F. Stambaugh. 2000. Comparing Asset Pricing Models: An Investment Perspective. *Journal of Financial Economics* 56:335–81.

Pastor, L., and R. F. Stambaugh. 2002. Investing in Equity Mutual Funds. *Journal of Financial Economics* 63:351–80.

Politis, D. N., and J. P. Romano. 1994. The Stationary Bootstrap. *Journal of the American Statistical Association* 89:1303–13.

Roy, A. D. 1952. Safety First and the Holding of Assets. *Econometrica* 20:431–39.

Santa-Clara, P., and A. Saretto. 2006. Option Strategies: Good Deals and Margin Calls. Working Paper, UCLA.

Shreve, S. E., and H. M. Soner. 1994. Optimal Investment and Consumption with Transactions Costs. *Annals of Applied Probability* 4:909–62.

Stambaugh, R. F. 1997. Analyzing Investments Whose Histories Differ in Length. *Journal of Financial Economics* 54:375–421.

Taksar, M., M. J. Klass, and D. Assaf. 1988. A Diffusion Model for Optimal Portfolio Selection in the Presence of Brokerage Fees. *Mathematics of Operations Research* 13:277–94.

Tepla, L. 2001. Optimal Investment with Minimum Performance Constraints. *Journal of Economic Dynamics and Control* 25:1629–45.