Optimal Portfolios with Parametric Weights*

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

Pedro Santa-Clara
Anderson School
UCLA‡
and NBER

Rossen Valkanov
Anderson School
UCLA§

First Draft: December 2002
This Draft: September 2004

Abstract

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, easily modified and extended, 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. Our approach also suggests a new test of the portfolio choice implications of equilibrium asset pricing models. We illustrate our portfolio optimization approach through an empirical application to all stocks in the CRSP-Compustat universe, exploiting size, value, and momentum effects.

*We thank Michael Brennan, Rob Engle, Ravi Jagannathan, Larry Harris, Francis Longstaff, and seminar participants at CUNY Baruch, NYU, the Chicago Quantitative Alliance, the Los Angeles Quantitative Investment Association, Mellon Capital, Nomura, the Society of Quantitative Analysis, and Wells Capital Management for their comments and suggestions.

†Durham, NC 27708, phone (919)660-1948, e-mail: mbrandt@duke.edu

‡Los Angeles, CA 90095, phone: (310)206-6077, e-mail: pedro.santa-clara@anderson.ucla.edu.

§Los Angeles, CA 90095, phone: (310)825-7246, e-mail: rossen.valkanov@anderson.ucla.edu.

1 Introduction

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 management is currently 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 parameters involved and the need to ensure the positive definiteness of the covariance matrix, but the results of this procedure are also notoriously noisy and unstable (e.g., Michaud, 1989). In reality, the Markowitz approach is therefore implemented along with a number of different fixes, including shrinkage estimation, a factor structure for the covariance matrix, an asset pricing models to relate expected returns in the cross-section, or constraints imposed on the optimized portfolio weights.[2] While these fixes generally improve the properties of the optimized portfolio, they requires substantial resources such as the tools developed by BARRA, Northfield, and others. As a result, formal portfolio optimization based on firm characteristics is seldom implemented by asset managers (with the notable exception of *quants* which are a small part of the profession), even though it has the potential to provide large benefits to the investors.[3]

We propose a simple new approach to equity portfolio optimization based on firm characteristics. We parameterize the weight invested in each stock as a function of the firm's characteristics, with the implicit assumption that these characteristics fully capture all aspects of the joint distribution of returns that are relevant for forming optimal portfolios. We then estimate the coefficients of the portfolio policy by maximizing the utility that would have been obtained by the investor from implementing the policy over the sample period.

Our approach has a number of conceptual advantages. First, we sidestep completely the auxiliary, yet very difficult, step of modeling the joint return distribution and instead focus directly on the object of interest – the optimal portfolio weights. Second, parameterizing the portfolio policy leads to a tremendous reduction in dimensionality. For a problem with N

[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 the variances and covariances of returns out of sample.

[2] See Brandt (2004) for a review of the literature.

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

stocks, the traditional Markowitz approach requires modeling N first and $(N^2+N)/2$ second moments of returns. With preferences other than the simplistic mean-variance kind, the traditional approach involves a practically unmanagable 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. 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 implicity the relation between the characteristics and expected returns, variances, covariances, and even higher order moments of returns, to the extent that these factors all affect the distribution of the optimized portfolio's returns and therefore the investor's 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.

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 (2002) 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 is Brandt and Santa-Clara (2004), who study a similar market-timing problem involving stocks, bonds, and cash by parameterizing the optimal portfolio weights as functions of the predictors. Specifically, they model the weight in each asset class as a separate function of a common set of variables with coefficients that are specific to the asset class. 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 weights 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 stocks, corporate bonds, or currencies.

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 portfolio weight constraints or interactions between the firm characteristics, conditioning the parameters of the portfolio policy on economic predictors, and long-horizon problems. All of these extensions are normative in nature, in that they deal with an investor's partial equilibrium portfolio choice problem. However, our approach also has a positive use for testing the portfolio implications of equilibrium asset pricing models. More concretely, we specify the optimal portfolio weights as the sum of the observed market capitalization weights and optimal deviations from the market cap weights that depend parametrically on the characteristics of the firms. A joint test of whether these optimal deviations from the market cap weights are zero therefore addresses the hypothesis of whether a representative investor with given preferences optimally holds the market portfolio. Failure to reject this hypothesis would identify the investor's preferences as those underlying an equilibrium asset pricing model that is consistent with the joint distribution of returns and firm characteristics (and economic predictors in the case of a time-varying portfolio policy).

To illustrate our approach, we optimize a portfolio of all stocks in the CRSP/Compustat dataset from 1962 through 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 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 large relative 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 10% without short-sale constraints and 3% with short-sale constraints. In the case without short-sale constraints, the benefits are even greater when we take into account interactions between the characteristics, especially interactions between lagged returns and book-to-market ratios, or when we allow for the coefficients of the portfolio policy to depend on the slope of the yield curve (certainty equivalent gains of 18% and 12%, respectively). With short-sale constraints, these two extensions yield relatively little economic benefit relative to the base case. We use an out-of-sample experiment to show that these gains in expected utility are not the result of in-sample overfitting.

The remainder of the paper proceeds as follows. We describe the basic idea and various

extensions of our approach in Section 2. The empirical application is presented in Section 3 and we conclude in Section 4.

2 Methodology

2.1 Basic Idea

Suppose that at each date t there are large number of N_t stocks in the investable universe (our method automatically accommodates the realistic case of a varying number of stocks through time). 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 return $r_{p,t+1}$:

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

We parameterize the optimal portfolio weights as a function of the characteristics:

$$w_{i,t} = \bar{w}_{i,t} + \frac{1}{N_t}\,\theta^\top \hat{x}_{i,t} \tag{2}$$

where $\bar{w}_{i,t}$ is the weight of stock i at date t in a benchmark 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 a zero mean and unit standard deviation across all stocks at date t.

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 $x_{i,t}$ can be non-stationary. 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 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 would result in twice as aggressive allocations, although the investment opportunities are fundamentally unchanged.

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 policy only cares about the characteristics of the stocks, not about the stocks themselves. 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 write the conditional optimization with respect to the portfolio weights w_{it}, equation (1), equivalently as the following unconditional optimization with respect to the coefficients θ:

$$\max_{\theta} \ \mathrm{E}\big[u(r_{p,t+1})\big] = \mathrm{E}\left[u\left(\sum_{i=1}^{N_t}\left(\bar{w}_{i,t} + \frac{1}{N_t}\,\theta^{\top}\hat{x}_{i,t}\right) r_{i,t+1}\right)\right]. \tag{3}$$

We can then estimate the coefficients θ by maximizing the corresponding sample analogue:

$$\begin{aligned} \max_{\theta} \ \frac{1}{T}\sum_{t=0}^{T-1} u(r_{p,t+1}) &= \frac{1}{T}\sum_{t=0}^{T-1} u\left(\sum_{i=1}^{N_t} w_{i,t} r_{i,t+1}\right) \\ &= \frac{1}{T}\sum_{t=0}^{T-1} u\left(\sum_{i=1}^{N_t}\left(\bar{w}_{i,t} + \frac{1}{N_t}\,\theta^{\top}\hat{x}_{i,t}\right) r_{i,t+1}\right), \end{aligned} \tag{4}$$

for some prespecified utility function (e.g., mean-variance, quadratic, or CRRA utility).

Three 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.[4] The computational burden of our approach only

[4]Especially because, for utility functions commonly used and given the linearity of the portfolio policy in the coefficients θ, it is easy to derive analytically the gradient and the hessian of the optimization problem.

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 θ 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 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 these factors all 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 third point, we can approximate the expected utility of the investor with a Taylor series expansion around the riskfree rate r_t^f:

$$\begin{aligned} \mathrm{E}\big[u(r_{p,t+1})\big] \approx\ & u(r_t^f) + u'(r_t^f)\mathrm{E}\big[(r_{p,t+1} - r_t^f)\big] \\ & + u''(r_t^f)\frac{1}{2}\mathrm{E}\big[(r_{p,t+1} - r_t^f)^2\big] + u'''(r_t^f)\frac{1}{6}\mathrm{E}\big[(r_{p,t+1} - r_t^f)^3\big] + ... \end{aligned} \tag{5}$$

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

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

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

To perform a comparable portfolio optimization using the traditional Markowitz approach requires modeling the means, variances, and covariances of all stocks as functions of the 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 parameterize the covariance matrix 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 mean-variance, is practically impossible because it requires modeling not only the conditional skewness and kurtosis of each stock but also the numerous higher-order cross-moments.

2.2 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" estimator $\hat{\theta}$, defined by the optimization problem (4), satisfies the first-order conditions:

$$\frac{1}{T}\sum_{t=0}^{T-1} h(r_{t+1}, x_t; \theta) \equiv \frac{1}{T}\sum_{t=0}^{T-1} u'(r_{p,t+1})(x_t^\top r_{t+1}) = 0 \tag{7}$$

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

$$\Sigma_\theta \equiv \text{AsyVar}[\hat{\theta}] = \frac{1}{T}[G^\top V^{-1} G]^{-1}, \tag{8}$$

where

$$G \equiv \frac{1}{T}\sum_{t=0}^{T-1} \frac{\partial h(r_{t+1}, x_t; \theta)}{\partial \theta} = \frac{1}{T}\sum_{t=0}^{T-1} u''(r_{p,t+1})(x_t^\top r_{t+1})(x_t^\top r_{t+1})^\top \tag{9}$$

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:

$$\frac{1}{T}\sum_{t=0}^{T-1} h(r_{t+1}, x_t; \hat{\theta}) h(r_{t+1}, x_t; \hat{\theta})^\top. \tag{10}$$

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

The resulting estimate of the asymptotic 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 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 be consistent with the relation between the characteristic and conditional return moments. 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 (3) 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 coefficient are constant through time, to condition down the conditional expectation. If either assumption is incorrect, it results in a systematic correlation of the marginal utilities in equation (7) 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 as 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 (7) 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.

2.3 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.

2.3.1 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 accomodate any choice of objective function. The only implicit assumption is that the conditional expected utility maximization problem (1) is 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 ratio, beating or tracking a benchmark, controlling draw-downs, or maintaining a certain value-at-risk (VaR).[5]

In the empirical application we use standard constant relative risk aversion (CRRA) preferences over either wealth or wealth relative to a performance benchmark:

$$u(r_{p,t+1}) = \frac{(1 + r_{p,t+1})^{1-\gamma}}{1-\gamma} \quad \text{or} \quad u(r_{p,t+1}) = \frac{\left(1 + r_t^f + (r_{p,t+1} - r_{b,t+1})\right)^{1-\gamma}}{1-\gamma}, \tag{11}$$

where $r_{b,t+1}$ is the return on an exogenous benchmark portfolio. The advantage of CRRA utility is that it incorporates preferences toward higher-order moments without introducing additional preference parameters. 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.

2.3.2 Portfolio Weight Constraints

By far the most common departure from the basic portfolio choice problem (1) in practice are 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 a no short-sale constraint for long-only equity portfolios. The simplest way to impose this constraint through the portfolio policy is to truncate the portfolio weights in

[5] Benartzi and Thaler (1995), Aït-Sahalia and Brandt (2001), Ang, Bekaert, and Liu (2003), Gomes (2003), among other, 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), Tepla (2001), Basak and Shapiro (2001), and Alexander and Baptista (2002).

equation (2) below at zero. Unfortunately, in doing so the optimal portfolio weights no longer sum to one (setting the negative weights to zero results in an sum of weights greater than one). We therefore need to also renormalize the portfolio weights, resulting in the following specification of the portfolio policy:

$$w_{i,t}^{+} = \frac{\max[0, w_{i,t}]}{\sum_{i=j}^{N_t} \max[0, w_{j,t}]}. \tag{12}$$

One computational problem with this specification of the portfolio policy function is its non-differentiability at $w_{i,t} = 0$. In order to compute the standard errors of the estimated θ from first-order conditions analogous to equation (7), we require first-order derivatives. We overcome this problem in practice by approximating the function $\max[0, y]$ between the two 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. To minimize the effect of this approximation on the estimates, we use a very small $\alpha = 10^{-6}$.

2.3.3 Nonlinearities and Interactions

Although we explicitly specified the portfolio policy 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 $w_{i,t} = \bar{w}_{i,t} + g(y_{i,t})$ for any $g(\cdot)$ that can be spanned by a polynomial expansion in the more basic state variables $y_{i,t}$. Our approach therefore accommodates very general departure of the optimal portfolio weights from the benchmark weights.

Cross-products of the characteristics are a particularly interesting form of non-linearities 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 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 returns. 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.

2.3.4 Time-Varying Coefficients

One of the critical assumptions required for conditioning down the expectation to rewrite the conditional problem (1) as the unconditional problem (3), 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.[6] 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 z_t, we specify the portfolio policy has:

$$w_{i,t} = \bar{w}_{i,t} + \frac{1}{N_t}\theta^\top \left(z_t \otimes x_{i,t}\right) \tag{13}$$

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.

2.3.5 Long-Horizon Problem

The discussion thus far assumed a myopic single-period horizon. However, our approach can in principle be extended to a multiperiod horizon problem as follows. Consider an investor who maximizes the utility of a K-period return:

$$r_{p,t+1\to t+K} = \prod_{k=0}^{K-1} r_{p,t+k+1}, \tag{14}$$

[6]For example, Campbell (1991), Campbell and Shiller (1988), Fama (1990), Fama and French (1988,1989), Hodrick (1992), and Keim and Stambaugh (1986) 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.

where each period's return is given as before by:

$$
\begin{aligned}
r_{p,t+k+1} &= \sum_{i=1}^{N_{t+k}} w_{i,t+k} r_{i,t+k+1} \\
&= \sum_{i=1}^{N_{t+k}} \left(\bar{w}_{i,t+k} + \frac{1}{N_{t+k}} \theta_k^\top x_{i,t+k} \right) r_{i,t+k+1}
\end{aligned}
\tag{15}
$$

with period-specific portfolio policy coefficients θ_k. Following the same steps as in the single-period problem, the investor's conditional expected utility optimization with respect to the portfolio weights $\{\{w_{i,t+k}\}_{i=1}^{N_{t+k}}\}_{k=0}^{K-1}$ can be alternatively expressed as an unconditional expected utility optimization with respect to the portfolio policy coefficients $\{\theta_k\}_{k=0}^{K-1}$. The sample analogue of this unconditional problem is:

$$
\max_{\{\theta\}_{k=0}^{K-1}} \frac{1}{T} \sum_{t=0}^{T-1} u \left(\prod_{k=0}^{K} \left(\sum_{i=1}^{N_{t+k}} \left(\bar{w}_{i,t+k} + \frac{1}{N_{t+k}} \theta_k^\top x_{i,t+k} \right) r_{i,t+k+1} \right) \right). \tag{16}
$$

Expressing the investor's portfolio choice problem in terms of optimizing over portfolio policies, as opposed to over conditional portfolio weights, has the same conceptual benefits in the multiperiod problem as we discussed in the single-period context (e.g., it avoids the return modeling step of the traditional approach). In addition, however, it allows for the universe of investable securities to change over the investor's horizon, meaning that N_{t+k} can be a random variable for $k > 0$. Incorporating this realistic feature of the portfolio problem in the traditional approach is difficult, if not practically impossible, because it requires an explicit model of exchange listings and delistings.

Although theoretically sound, there are at least two practical complications introduced by having multiple periods. First, the number of coefficients increases with the number of periods, raising the risk of in-sample overfitting. This problem can be overcome by imposing a parametric structure on the coefficients of the portfolio policy across periods, such as the restriction that the coefficients are constant or that the coefficients for each period are scaled by some function of the number of periods to the end of the horizon. The second complication is that the average utility in equation (16) is computed with overlapping observations, which can severely distort the statistical inferences based on the iid covariance matrix estimator (10). While it is possible to correct the covariance matrix estimator for the persistence introduces by the overlapping observations, the resulting inferences are still

questionable. The best way to avoid this problem is to use only T/K non-overlapping observations in constructing the sample analogue of the unconditional optimization problem. Doing so unfortunately limits the length of the investment horizon we can realistically consider with our approach given the finite size of the sample.

2.4 Testing Asset Pricing Models

Portfolio choice problems are by nature normative, telling the investor what he or she should do to maximize expected utility. An interesting feature of our approach is that it also has a positive application, helping economists understand better how markets function. Specifically, the cross-sectional relation of expected returns in an equilibrium asset pricing model is typically derived by combining the first-order conditions of the portfolio choice problem of a representative agent with the market clearing condition that this agent has to hold each asset in proportion to its market capitalization weight.

Reversing this argument, given preferences for the representative agent are only consistent with market equilibrium if the agent's implied optimal portfolio weights equal the observed market cap weights. Our approach is ideally suited for testing this implication, since we model explicitly the deviations of the investor's optimal portfolio weights from the observed market cap weights. A rejection of the hypothesis $\theta = 0$ implies that the preferences of the investor do not correspond to those of the representative agent.

To be more specific, any equilibrium asset pricing model (CAPM, equilibrium APT, or CCAPM), requires that the representative investor hold all the assets in the economy in the proportions they exist in the market. Our proposed test makes an assumption about the utility function of the representative investor (e.g., CRRA) and about the functional form of the portfolio policy but does not require that we model the expected returns and risk of stocks (e.g., measuring risk as covariance with the market or covariance with consumption). We just need to find the agent's optimal portfolio and compare it to the market portfolio.[7]

[7] Actually, to the extent that the weight function is misspecified, that makes the test conservative.

3 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 from January 1964 through December 2002. We first describe the data and then present results for the base case, various extensions, and an out-of-sample experiment. Unless otherwise stated, we assume an investor with CRRA preference and a relative risk aversion of five.

3.1 Data

We use monthly firm-level returns from CRSP as well as firm-level characteristics, obtained from the COMPUSTAT/CRSP merged dataset, from January 1964 through December 2002. For each firm in the COMPUSTAT/CRSP dataset, we construct the following variables at the end of each fiscal year: log of 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.[8] We use the standard timing convention of leaving a minimum of six-month lag between the fiscal year-end characteristics and the monthly returns, to insure 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-2 and t-13. 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.

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 3,680, with the fewest firms in Feb 1964 (1,033 firms) and the most firms in Nov 1997 (6,356 firms).

Figure 1 describes the three firm characteristics. The first column plots the cross-sectional means of the (non-standardized) characteristics at each month in our sample. The second column shows the corresponding cross-sectional standard deviations. Recall that the characteristics enter the portfolio policy function in standardized form. The plots in Figure 1

[8] Taking logs makes the cross-section distribution of me and (1+btm) more symmetric and reduces outliers.

can be used to translate given values for the standard 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 riskfree 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 10-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.

3.2 Base Case

Table 1 presents the results for the base case, in which the over- or under-weighting 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, with the policy function in equation (2). The table is divided into four sections describing separately the (i) parameter estimates and standard 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. In addition to the results for the universe of all CRSP stocks, we also present results for a a restricted set of the 500 largest firms. We first focus on the results for all stocks and then point out where the results for the top 500 stocks differ.

The first few rows in Table 1 presents the estimated coefficients of the portfolio policy along with their standard errors. In the case of all stocks, in the second column, the deviations of the optimal weights from the benchmark weights decrease with the firms's market capitalization (size) and increases 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 over-weights small firms, value firms, and past winners and under-weights large firms, growth firms, and past loosers. 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 over-weighting of a stock. All three coefficients are statistically highly significant. We also test whether all three coefficients are jointly equal to zero using a likelihood ratio test, and the p-value of this test is reported in the row labeled "LRT p-value."[9]

[9] When the p-value from the likelihood ratio test is less than 0.001, we report it as 0.000.

The next few rows 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). The average absolute weight of the optimal portfolio is about three times that of the market (0.12% versus 0.04%). Not surprisingly, the active portfolio takes larger positions. However, these positions are not extreme. The average (over time) maximum and minimum weight of the optimal portfolio are 4.36% and -0.32%, respectively, while the corresponding extremes for the market portfolio are 4.63% and 0.00%. The average sum of negative weights in the optimal portfolio is -124%, which implies that the sum of long positions is 224%. Finally, the average fraction of negative weights (shorted stocks) in the optimal portfolio is 0.468. Overall, the optimal portfolio does not reflect unreasonably extreme bets on individual stocks and could well be implemented by a long-short (albeit net long) equity hedge fund.

The following rows characterize the performance of the optimal portfolio relative to the market portfolio. For ease of interpretation, all measures are annualized. The optimal portfolio has a volatility slightly larger than that of the market portfolio, 19% versus 16%, but has an average return of 24.4% as opposed to 12.0%. This translates into a certainty equivalent gain of 10%. Finally, 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 15%, with a low market beta of only 0.440. Dividing the alpha by the residual volatility of 17.7% produces an information ratio of 0.890.

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 market portfolio has a bias towards 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.[10]

[10] 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 an average characteristic of 2 and is short 100% of stocks with average characteristic of 1 has an average characteristic of 3.

Figure 2 plots the time series of the three portfolio characteristics. The characteristics vary over time, but their variability is relatively small and appears stationary. Moreover, the book-to-market characteristic is always larger than the momentum characteristic, which in turn is larger than the size characteristic, which indicates 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. Finally, one might suspect that maintaining such stable portfolio characteristics requires unreasonably large trading activity. Fortunately, this is not the case. The average turnover of the optimized portfolio is only 50% per year, as compared to a average turnover of 12% per year for the market portfolio.

The results are similar for the 500 largest stocks, which are an interesting subset of firms because they are certainly liquid enough to implement an active portfolio strategy in large scale. The most notable difference in the results is that the book-to-market coefficient, while still positive and significant, is about half as large as in the case of all stocks. The optimal allocation is achieved with fewer average number of stocks shorted, 35.8%, and the total short position of the optimal portfolio is only -51.8%. The average return of the optimal portfolio is 20%, or more than 7% larger than the return of the market portfolio. Its standard deviation, at 17.6%, is slightly larger than that of the benchmark. The alpha of the optimal portfolio is 10% with a market beta of 0.704, residual volatility of 13.6% and a information ratio of 0.734. The average characteristics of the optimized portfolio are again tilted toward smaller, value, and winner stocks.

3.3 Extensions

3.3.1 Portfolio Weight Constraints

A large majority of equity portfolio managers faces short-sale constraints. In Table 2, we present the results from estimating the long-only portfolio policy specified in equation (12). As in the unconstrained base 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 equally desirable characteristics for a long-only investor. The coefficients associated with

both of these characteristics are statistically significant, whereas the coefficient associated with the market capitalization of the firm is not. The test for joint significance of all three parameters has a p-value of less than 0.001.

The optimal portfolio still does not involve extreme weights. In fact, the average maximum weight of the optimal portfolio is only 1.5%, which is actually lower than that of the market portfolio. On average, the optimal portfolio invests in only 52% of the stocks. The resulting mean and standard deviation of the portfolio return are 18.6% and 19.5%, respectively, translating into a certainly equivalent gain of 3.1% relative to holding the market portfolio. The α, β, and information ratio of the portfolio are 6.8%, 0.971, and 0.576, respectively. These statistics are quite remarkable, given the long-only constraint. The average characteristics of the optimal portfolio are 0.023, 0.851, and 0.987. The average size of the firms in the optimal portfolio is slightly 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 about one standard deviation above those of the average stock and are also significantly different from those of the market portfolio.

The most interesting comparison is between the long-only portfolio in Table 2 and the unconstrained base case in Table 1. Qualitatively, the results in the two tables are similar, but the long-only constraint results in a lower average portfolio return, slightly higher volatility, and consequently a lower certainty equivalent gain relative to holding the market portfolio. 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 positions in Table 1 is roughly the same as the fraction of stocks not held by the long-portfolio in Table 2. Second, the unconstrained portfolio benefits from using the short positions as leverage to increase the exposure of the long positions.

3.3.2 Interactions

Table 3 includes interactions between the three characteristics in both the constrained and unconstrained portfolio policies. Specifically, we include the three cross-products of a firm's market capitalization, book-to-market ratio, and one-year lagged return, resulting in a total of six characteristics. The interactions do not substantially change the essence of the previous results. Both portfolios are still tilted toward smaller firms, value firms, and past winners.

Examining more closely the estimated coefficients for the interacted characteristics, all three coefficients are statistically significant in the unconstrained case. Relative to the base case, the optimal portfolio policy tilts the weights positively towards large-growth and small-value stocks, large-winner and small-looser stocks, and value-winner and growth-looser stocks. The policy tilts the weights negatively towards the opposite combinations of characteristics. It is important to notice that the θ coefficients can only be interpreted as marginal contributions to the weights, when they are evaluated at the average level of characteristics, much like the coefficients of multivariate regressions cannot be interpreted in isolation. To understand the result of the optimal policy, it is best to examine the resulting average characteristics in the portfolio. In the constrained case, the coefficients on the interaction terms are generally smaller and less significant. The size-momentum interaction does not play a significant role in the constrained portfolio, whereas the value-momentum interaction is comparable to the unconstrained case.

The interactions do not lead to dramatic changes in the distribution of the optimal portfolio weights, but the effect on the distribution of the portfolio return is substantial. Both the unconstrained and constrained portfolios exhibit higher average returns (42.8% and 19.7%), higher certainly equivalents (23.1% and 9.7%), higher alphas (32.4% and 8.1%), and higher information ratios (1.283 and 0.662). Comparing these numbers to the corresponding results without interactions, in tables 1 and 2, the performance improvement is much greater for the unconstrained portfolio. The certainty equivalent increases by 7.8% in the unconstrained case and by only 1.4% in the constrained case.

Besides illustrating the potential for interactions between characteristics to improve the performance of the optimized portfolio, especially in the unconstrained case, this extension also illustrates the ability of our method to accommodate a larger number of characteristics. Given analytical expressions for the gradient and Hessian of the objective function, the estimation with six characteristics takes only marginally longer than in the base case.

3.3.3 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 a separate set of coefficients for months when the yield curve at the beginning of the month is positively sloped (normal) and negatively sloped (inverted). This is especially interesting since inverted yield curves tend to be associated with recessions.

The economic motivation for this particular way of modeling the time-varying coefficients is to let the effect of the characteristics on the joint distribution of returns be different for expansionary and contractionary periods.

The table presents results for unconstrained and constrained portfolios. In both cases, the most dramatic effect of conditioning on the slope of the yield curve is on the role of the firm's size. 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 the coefficient is not statistically different from zero in the unconstrained case). 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 larger in magnitude for the unconstrained case, but they are smaller in magnitude in the constrained case.

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 portfolios have higher average returns, certainly equivalents, alphas, and information ratios than without conditioning. The certainty equivalent improvement is again greater for the unconstrained portfolio, with a gain of 1.8% versus 0.7% for the constrained portfolio.

The average characteristics of the optimal portfolios are the most interesting to analyze. Consider first the unconstrained case. As predicted 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. Turning to the constrained results, notice that the optimal portfolio is not significantly tilted into value and winner firms when the yield curve is inverted, suggesting that during those time the profitability of value and momentum comes primarily from short positions.

3.3.4 Objective Functions

The optimal portfolio policy depends critically on the choice of preferences. The results thus far were obtained assuming CRRA utility with a relative risk aversion of $\gamma = 5$. To get a better sense for the role of this preference assumption, we present in Table 5 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 extreme outcomes. Finally, we also consider a minimum variance objective.

The table presents results for both constrained and unconstrained portfolios, but the discussion is focused on the unconstrained case because two are qualitatively similar. 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 highly significant. As the investor becomes more risk averse and, in the limit, cares only about risk, the characteristics become less important. In fact, for the minimum variance objective, none of the characteristics enter the portfolio policy in a significant way, which means that the investor cannot do better than hold the market portfolio. This suggests that the vast majority of the performance improvement from including the characteristics in the portfolio policy documented above comes from the predictability of mean returns. It is interesting to note that extreme case of $\gamma = 100$ breaks the monotonicity of the results because for such extreme curvature of the utility function, the investor cares only about extreme outcomes.

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 three times larger, which implies that the less risk averse investor takes similar bets but with more leverage. At the other extreme, for a minimum variance objective, the sum of short positions is minimal.

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, volatility, and certainty equivalent are all highest for $\gamma = 1$ and decrease with the investor's level of risk aversion. The certainty equivalent is 36.3% in the log utility case, 15.3% for $\gamma = 5$, and even negative for $\gamma = 100$. For the minimum variance objective, the certainty equivalent is 2.7 percent. The alphas and information ratios follow a similar pattern. The market beta increases with the level of risk aversion because the more risk averse the investor is, the more the optimal portfolio resembles the market portfolio.

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 decreases until, for the minimum variance objective, characteristics of

the optimal portfolio are very similar to those of the market portfolio (see Table 1).

As another example of how the objective function affects the optimal portfolio policy, we present in Table 6 results for a CRRA investor who cares about the return of the portfolio relative to the return on an exogenous benchmark portfolio. The relative performance utility function is given in equation (11). We consider two different benchmark portfolios. The first, in Panel A, is the value-weighted market portfolio and the second, in Panel B, is a constant maturity ten-year Treasury note. In both cases, the portfolio is made up of the 500 largest stocks with or without short-sales. The market portfolio benchmark is a natural choice from the perspective of an index fund. The bond benchmark is motivated by the problem of a pension fund servicing a long-term liability with the return from an equity portfolio.

In Panel A, the estimates of the policy coefficients are quite different from the previous results. Compared to the results in Tables 1 and 2, both the unconstrained and constrained portfolios places substantially larger weights on small firms (in the universe of large firms) and on firms with large positive one-year lagged returns. The role of the book-to-market ratio in the portfolio policy is greatly diminished. The estimated coefficients on size and momentum are large in magnitude and statistically different from zero at conventional levels, whereas the estimated coefficient on book-to-market is neither large nor significant.

To evaluate the performance of the optimized portfolio, we compute the return statistics for the benchmarked return $r_t^f + (r_{p,t+1} - r_{b,t+1})$. The average return is 18.2% in the unconstrained case and 14.5% in the constrained case. The corresponding return volatilities are 13.6% and 8.2%, resulting in certainty equivalents of 13.5% in the constrained case and 12.2% in the unconstrained case. Comparing the average characteristics of the optimized portfolio to those of the value weighted index of the 500 largest firm (in the first column), the unconstrained portfolio is tilted considerably toward small firms as well as toward firms with large positive one-year lagged return. The portfolio only has a slight value bias. With the short-sale constraint imposed, the size tilt is considerably smaller.

The results are yet again very different for the bond benchmark in Panel B. In that case, the estimated coefficients on book-to-market and momentum are statistically significant, while the coefficient on size is not. The benchmarked portfolio returns are quite volatile, due to a relatively small correlation between the stock and bond returns, and the certainty equivalents are, as a result, considerably smaller than in Panel A. The average characteristics confirm that the optimized portfolio is tilted toward value and momentum, with a relatively

little emphasis on small firms.

3.3.5 Long-Horizon Problem

Finally, we present an application of our approach to a long-horizon problem. Table 7 presents the optimal portfolio policy of an agent with a multiperiod objective of one or five years (the one-month results are reproduced for comparison) with monthly rebalancing. The portfolio policies are obtained from problem (16) using non-overlapping returns in the case of the one-year problem but using overlapping annual returns for the five-year problem.

The main results are that the investor still finds it optimal (perhaps more so) to tilt the portfolio toward value and momentum as the horizon increases. This is in contrast to Brennan and Xia (2003), Campbell and Vuolteenaho (2003), and Bansal, Dittmar, and Lundblad (2002), who argue that the value premium in particular can be explained as an ICAPM risk premium. Their argument is that value stocks offer bad returns when investment opportunities are poor. Value stocks would therefore be required to offer a higher expected return to an investor with a long horizon who cares about reinvestment risk.

3.4 Out-of-Sample Performance

While the stellar performance of our approach is unlikely to be due to over-fitting, given that in the base case we optimize a portfolio of 3,680 stocks on average over only three parameters, the most convincing way to establish this fact is through an out-of-sample experiment. Since we lack a sufficiently long truly out-of-sample period, we conduct the following pseudo out-of-sample experiment. We split our sample into two equally long sub-samples from January 1964 through June 1983 and from July 1983 through December 2002, labeled Sample 1 and Sample 2, respectively. For each sub-sample, we estimate the optimal portfolio policy as usual and then use the estimated coefficients to form a portfolio in the sub-sample that was not used in the estimation. More specifically, for each month in Sample 2, we form a portfolio using the policy coefficients estimated from Sample 1 and the firm characteristics observed at the beginning of that month in Sample 2. The returns generated by these portfolios in Sample 2 represent the out-of-sample results for Sample 1. We conduct an analogous out-of-sample experiment, flipping the estimation and evaluation samples, for Sample 2.

Panel A of Table 8 presents the unconstrained results. Before discussing the out-of-sample

performance of our approach, we briefly comment on the sub-sample results. The coefficient estimates are generally consistent across sub-samples. The size coefficient is negative and the book-to-market and momentum coefficients are positive in both sub-samples. The coefficients are individually and jointly significant even with the smaller sample sizes, which implies that our tests are quite powerful. The distributions of the weights and returns of the optimized portfolio are also similar across sub-samples. The certainty equivalent is 15.1% in the first and 17.7% in the second sub-sample. Finally, in both sub-samples the portfolios are tilted substantially toward value and momentum, with momentum playing a larger role in the first sub-sample and book-to-market playing a larger role in the second sub-sample.

Turning to the out-of-sample results, the "Out-of-Sample Fcst." column of Sample 2 summarizes the results from using the coefficients estimated from the first sub-sample to form portfolios in the second sub-sample. The in- and out-of-sample portfolios are remarkably similar in terms of the distributions of the portfolio weights. More importantly, the return statistics are also very similar. For instance, the average return and volatility are 24.5% and 16.3% when using the optimal coefficients (in-sample) versus 26.5% and 22.4% when using the coefficients estimated from Sample 1 (out-of-sample). The resulting certainty equivalents are 17.7% in-sample and 12.1% out-of-sample. It is astonishing that using portfolio policy coefficients estimated as long as 20 years ago results in less than a 6% reduction in certainty equivalent, compared to the (in practice unattainable) in-sample optimum. Moreover, the average characteristics of the in- and out-of-sample portfolios are similar, both with a tilt toward firms that are small, have high book-to-market ratios, and experienced large positive one-year lagged returns. Notice that although the sign of the average size is different for the two portfolios, both are considerably smaller than the average size of the market portfolio, and therefore both portfolios reflect a bias toward small firms

The "Out-of-Sample Fcst." column of Sample 1 summarizes the results from using in the first sub-sample the coefficients estimated from the second sub-sample. The results are qualitatively the same as above. The in- and out-of-sample portfolios are remarkably similar, both in terms of their weights and average characteristics, with comparable performance. The loss in certainty equivalent from not optimizing in-sample is only 3.8% in this case.

We conclude from these results that our approach is likely to perform as well out-of-sample as our in-sample analysis suggests. The same conclusion can be drawn from Panel B of Table 8, which reports the results for the portfolio subject to short-sale constraints. Since the constrained portfolios take less aggressive size, value, and momentum bets, the margin

for error is smaller. As a result, the certainty equivalent loss from not optimizing in-sample is a mere 1.5% for the first sub-sample and 1.1% for the second sub-sample.

4 Conclusion

We proposed a novel approach to optimizing large-scale equity portfolios. Rather than first model the distribution of returns to then solve for the implied optimal portfolio weights, we model directly the optimal portfolio weight in each stock 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. 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 the many features of our approach through an empirical application. Our empirical 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 weights for CRRA utility 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 large positive lagged returns (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 10% without short-sale constraints and 3% with short-sale constraints. Without short-sale constraints, the benefits are even greater when we allow for interactions between characteristics, especially interactions between lagged returns and book-to-market ratios, or when we allow for the coefficients of the portfolio policy to depend on the slope of the yield curve (certainty equivalent gains of 18% and 12%, respectively). With short-sale constraints, these extensions yield relatively little economic benefit compared to the base case. We used an out-of-sample experiment to show that these gains in expected utility are not the result of in-sample overfitting.

Our idea can easily be applied to other asset classes. For instance, we could use a similar approach to form bond portfolios based on the characteristics of the bond (e.g., duration, coupon, rating, leverage) or to form currency portfolios based on the characteristics of the

country pair (e.g., interest rate differential, inflation differential, trade balance).

References

Aït-Sahalia, Yacine, and Michael W. Brandt, 2001, Variable selection for portfolio choice, *Journal of Finance* 56, 1297–1351.

——— , 2002, Portfolio and consumption choice with option-implied state prices, Working Paper, Princeton University.

Alexander, Gordon, and Alexandre 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, 1159–1193.

Ang, Andrew, Geert Bekaert, and Jun Liu, 2003, Why stocks may disappoint, *Journal of Financial Economics* forthcoming.

Bansal, Ravi, Robert Dittmar, and Christian Lundblad, 2002, Consumption, dividends, and the cross-section of equity returns, Working Paper, Duke University.

Basak, Suleyman, and Alexander Shapiro, 2001, Value-at-risk based risk management: optimal policies and asset prices, *Review of Financial Studies* 14, 371–405.

Benartzi, Shlomo, and Richael H. Thaler, 1995, Myopic loss aversion and the equity premium puzzle, *Quarterly Journal of Economics* 110, 73–92.

Brandt, Michael W., 1999, Estimating portfolio and consumption choice: A conditional method of moments approach, *Journal of Finance* 54, 1609–1646.

——— , 2004, Portfolio choice problems, in Yacine Aït-Sahalia, and Lars P. Hansen, ed.: *Handbook of Financial Econometrics.*

——— , and Pedro Santa-Clara, 2004, Dynamic portfolio selection by augmenting the asset space, Working Paper, UCLA.

Brennan, Michael, and Yihong Xia, 2003, Risk and valuation under an intertemporal asset pricing model, *Journal of Business* forthcoming.

Browne, Sid, 1999, Beating a moving target: Optimal portfolio strategies for outperforming a stochastic benchmark, *Finance and Stochastics* 3, 275–294.

Campbell, John, and Tuomo Vuolteenaho, 2003, Bad beta, good beta, Working Paper, Harvard University.

Campbell, John Y., 1991, A variance decomposition for stock returns, *Economic Journal* 101, 157–179.

——— , and Robert J. Shiller, 1988, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661–676.

Chan, K.C., Jason Karceski, and Josef Lakonishok, 1998, The risk and return from factors, *Journal of Financial and Quantitative Analysis* 33, 159–188.

——— , 1999, On portfolio optimization: Forecasting covariances and choosing the risk model, *Review of Financial Studies* 12, 937–974.

Cooper, Michael, Huseyin Gulen, and Maria Vassalou, 2000, Investing in size and book-to-market portfolios using information about the macroeconomy: Some new trading rules, Working Paper, Purdue University.

Daniel, Kent D., and Sheridan Titman, 1999, Market efficiency in an irrational world, *Financial Analyst Journal* 55, 24–50.

Fama, Eugene F., 1990, Stock returns, expected returns, and real activity, *Journal of Finance* 45, 1089–1108.

——— , and Kenneth R. French, 1988, Permanent and temporary components of stock prices, *Journal of Political Economy* 96, 246–273.

——— , 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

——— , 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55–84.

Gomes, Francisco, 2003, Portfolio choice and trading volume with loss-averse investors, *Journal of Business* forthcoming.

Grossman, Sanford J., and Jean-Luc Vila, 1989, Portfolio insurance in complete markets: A note, *Journal of Business* 62, 473–476.

Hansen, Lars Peter, 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1053.

Hodrick, Robert J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial Studies* 5, 257–286.

Jagannathan, Ravi, and Tongshu Ma, 2002, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *Journal of Finance* 58, 1651–1684.

Keim, Donald B., and Robert F. Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* 17, 357–390.

Markowitz, Harry M., 1952, Portfolio selection, *Journal of Finance* 7, 77–91.

Michaud, Robert O., 1989, The markowitz optimization enigma: Is optimized optimal?, *Financial Analyst Journal* 45, 31–42.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Nigmatullin, Eldar A., 2003, Bayesian model averaging for moment conditions models, Working Paper, University of Wisconsin-Madison.

Roy, A.D., 1952, Safety first and the holding of assets, *Econometrica* 20, 431–439.

Tepla, Lucie, 2001, Optimal investment with minimum performance constraints, *Journal of Economic Dynamics and Control* 25, 1629–1645.

Appendix: Data

For each firm in the COMPUSTAT dataset, we construct several variables at the end of fiscal years 1964 to 2002. The first full year of data, 1963, was used to construct lagged values. The exact fiscal year end dates are from CRSP. We merge the COMPUSTAT and CRSP using the COMPUSTAT/CRSP merged database. 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), or (56), or (130), in that order, otherwise zero; balance sheet deferred taxes and investment tax credits (35), otherwise zero; price per share (199), otherwise, taken 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 dataset. Then, we define: book equity (BE) equals total assets minus liabilities plus balance sheet deferred taxes and investment tax credits minus preferred stock value; market equity (ME) equals price per share times shares outstanding; book to market (btm) equals log of one plus BE divided by BE. We omit firms with negative $BEtoME$ ratio. Log market equity (me) is computed as the log of ME. Market leverage (ML) equals liabilities divided by the sum of liabilities minus balance sheet deferred taxes and investment tax credits plus preferred stock value plus me and book leverage (BL) equals liabilities divided by total assets. The monthly firm returns, $r_{i,t}$, are from CRSP. We allow a minimum of six month lag between the fiscal year end of the above variables and the returns to insure that the information form the firms' annual reports is publicly available. From CRSP, we also compute the last twelve-month return (mom), defined as monthly compounded return between months $t-2$ and $t-13$. The firm-level characteristics that we use in the program are me, btm, and mom. After all variables have been created, we omit the smallest 20 percent of firms (i.e., firms in the lowest 20 percentile of me).

Table 1: Base Case

This table shows estimates of the portfolio policy with three characteristics: size (me), book to market ratio (btm), and momentum (mom). In panel "All Stocks," we use all stocks in the merged COMPUSTAT/CRSP database from January 1964 through December 2002. Panel "Top 500 Stocks" includes only the largest 500 companies in each month. The columns labeled "Val.Weighted" and "Opt." display statistics of the market-capitalization weighted portfolio and the optimal portfolio, respectively. The first set of rows shows the estimated coefficients of the portfolio policy with standard errors in parentheses. The p-value of the likelihood ratio test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. The second set of rows shows statistics of the portfolio weights, averaged across time. These statistics include the average absolute portfolio weight ($|w_i|$), the average minimum and maximum portfolio weights (max w_i and min w_i), the average sum of negative weights in the portfolio ($\sum w_i I(w_i < 0)$), and the average fraction of negative weights in the portfolio ($\sum I(w_i \leq 0)/N_t$). The third set of rows displays average portfolio return statistics: average return and standard deviation of returns ($\bar{r}$ and $\sigma(r)$), the certainty equivalent return, 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.

	All Stocks		Top 500 Stocks	
Variable	Val.Weighted	Opt.	Val.Weighted	Opt.
θ_{me}	—	-1.220	—	-1.389
Std.Err.	—	(0.549)	—	(1.174)
θ_{btm}	—	3.466	—	1.557
Std.Err.	—	(0.836)	—	(0.667)
θ_{mom}	—	2.000	—	1.709
Std.Err.	—	(0.590)	—	(0.495)
LRT p-value	—	0.000	—	0.012
$\lvert w_i \rvert \times 100$	0.035	0.121	0.200	0.407
max $w_i \times 100$	4.625	4.363	5.343	4.226
min $w_i \times 100$	0.000	-0.323	0.037	-1.126
$\sum w_i I(w_i < 0)$	0.000	-1.238	0.000	-0.518
$\sum I(w_i \leq 0)/N_t$	0.000	0.468	0.000	0.358
$\bar{r}$	0.120	0.244	0.117	0.200
$\sigma(r)$	0.161	0.190	0.158	0.176
CE(r)	0.052	0.153	0.052	0.115
α	—	0.157	—	0.100
β	—	0.440	—	0.704
$\sigma(\epsilon)$	—	0.177	—	0.136
IR	—	0.890	—	0.734
me	2.095	-0.118	1.527	-0.136
btm	-0.451	3.296	-0.240	1.430
mom	0.012	1.839	-0.026	1.622

Table 2: Short-Sale Constraint

This table shows estimates of the portfolio policy with three characteristics: size (me), book to market ratio (btm), and momentum (mom). The portfolio weights, specified in equation (12), are restricted to be non-negative (long-only positions). In panel "All Stocks," we use all stocks in the merged COMPUSTAT/CRSP database from January 1964 through December 2002. Panel "Top 500 Stocks" includes only the largest 500 companies in each month. The columns labeled "Val.Weighted" and "Opt." display statistics of the market-capitalization weighted portfolio and the optimal portfolio, respectively. The first set of rows shows the estimated coefficients of the portfolio policy with standard errors in parentheses. The p-value of the likelihood ratio test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. The second set of rows shows statistics of the portfolio weights, averaged across time. These statistics include the average absolute portfolio weight ($|w_i|$), the average minimum and maximum portfolio weights (max w_i and min w_i), the average sum of negative weights in the portfolio ($\sum w_i I(w_i < 0)$), and the average fraction of negative weights in the portfolio ($\sum I(w_i \leq 0)/N_t$). The third set of rows displays average portfolio return statistics: average return and standard deviation of returns ($\bar{r}$ and $\sigma(r)$), the certainty equivalent return, 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.

	All Stocks		Top 500 Stocks	
Variable	Val.Weighted	Pos.Opt.	Val.Weighted	Pos.Opt.
θ_{me}	—	-0.634	—	-0.990
Std.Err.	—	(0.492)	—	(0.763)
θ_{btm}	—	4.833	—	1.585
Std.Err.	—	(2.542)	—	(2.846)
θ_{mom}	—	4.554	—	2.188
Std.Err.	—	(2.125)	—	(0.982)
LRT p-value	—	0.000	—	0.000
$\lvert w_i \rvert \times 100$	0.035	0.035	0.200	0.200
max $w_i \times 100$	4.625	1.460	5.343	2.903
min $w_i \times 100$	0.000	0.000	0.037	0.000
$\sum w_i I(w_i < 0)$	0.000	0.000	0.000	0.000
$\sum I(w_i \leq 0)/N_t$	0.000	0.481	0.000	0.373
$\bar{r}$	0.120	0.186	0.117	0.166
$\sigma(r)$	0.161	0.195	0.158	0.167
CE(r)	0.052	0.083	0.052	0.091
α	—	0.068	—	0.052
β	—	0.971	—	0.962
$\sigma(\epsilon)$	—	0.118	—	0.070
IR	—	0.576	—	0.739
me	2.095	0.023	1.527	0.170
btm	-0.451	0.851	-0.240	0.523
mom	0.012	0.987	-0.026	0.918

Table 3: Interactions

This table shows estimates of the portfolio policy with six characteristics: size (me), book to market ratio (btm), momentum (mom), and the cross product of the characteristics me $\times$ btm, me $\times$ mom, and btm $\times$ mom. In the estimation, we use all stocks in the merged COMPUSTAT/CRSP database from January 1964 through December 2002. The column labeled "Val.Weighted" displays the statistics of the market-capitalization weighted portfolio, as a benchmark. The next two columns, "Opt." and "Pos.Opt." characterize the optimal and positive optimal (long-only) portfolios, respectively. The first set of rows shows the estimated coefficients of the portfolio policies (??) and (??) with standard errors in parenthesis. The p-value of the likelihood ratio test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. The second set of rows shows statistics of the portfolio weights, averaged across time. These statistics include the average absolute portfolio weight ($|w_i|$), the average minimum and maximum portfolio weights (max w_i and min w_i), the average sum of negative weights in the portfolio ($\sum w_i I(w_i < 0)$), and the average fraction of negative weights in the portfolio ($\sum I(w_i \leq 0)/N_t$). The third set of rows displays average portfolio return statistics: average return and standard deviation of returns ($\bar{r}$ and $\sigma(r)$), the certainty equivalent return, 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.

	All Stocks		
Variable	Val.Weighted	Opt.	Pos.Opt.
θ_{me}	—	-0.744	-1.848
Std.Err.	—	(0.642)	(0.392)
θ_{btm}	—	4.444	5.957
Std.Err.	—	(0.951)	(0.847)
θ_{mom}	—	4.631	3.800
Std.Err.	—	(1.079)	(1.582)
$\theta_{\text{me}\times\text{btm}}$	—	-1.727	1.180
Std.Err.	—	(0.765)	(0.656)
$\theta_{\text{me}\times\text{mom}}$	—	4.718	0.206
Std.Err.	—	(1.273)	(1.125)
$\theta_{\text{btm}\times\text{mom}}$	—	4.249	4.301
Std.Err.	—	(1.583)	(0.752)
LRT p-value	—	0.000	0.000
$\|w_i\| \times 100$	0.035	0.186	0.035
max $w_i \times 100$	4.625	4.926	1.659
min $w_i \times 100$	0.000	-0.242	0.000
$\sum w_i I(w_i < 0)$	0.000	-1.939	0.000
$\sum I(w_i \leq 0)/N_t$	0.000	0.503	0.567
$\bar{r}$	0.120	0.428	0.197
$\sigma(r)$	0.161	0.278	0.192
CE(r)	0.052	0.231	0.097
α	—	0.324	0.081
β	—	0.742	0.924
$\sigma(\epsilon)$	—	0.252	0.123
IR	—	1.283	0.662
me	2.095	-0.003	-0.137
btm	-0.451	4.714	1.093
mom	0.012	1.446	0.932
me $\times$ btm	-1.142	-5.398	-1.004
me $\times$ mom	0.017	1.864	-0.828
btm $\times$ mom	-0.014	3.281	1.208

Table 4: Conditioning on the Slope of the Yield Curve

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, $I(\text{tsp} \geq 0$ or $I(\text{tsp} < 0$. In the estimation, we use all stocks in the merged COMPUSTAT/CRSP database from January 1964 through December 2002. The column labeled "Val.Weighted" displays the statistics of the market-capitalization weighted portfolio, as a benchmark. The next two columns, "Opt." and "Pos.Opt." characterize the optimal and positive optimal (long-only) portfolios, respectively. The first set of rows shows the estimated coefficients of the portfolio policies with standard errors in parenthesis. The p-value of the likelihood ratio test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. The second set of rows shows statistics of the portfolio weights, averaged across time. These statistics include the average absolute portfolio weight ($|w_i|$), the average minimum and maximum portfolio weights (max w_i and min w_i), the average sum of negative weights in the portfolio ($\sum w_i I(w_i < 0)$), and the average fraction of negative weights in the portfolio ($\sum I(w_i \leq 0)/N_t$). The third set of rows displays average portfolio return statistics: average return and standard deviation of returns ($\bar{r}$ and $\sigma(r)$), the certainty equivalent return, 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.

	All Stocks		
Variable	Val.Weighted	Opt.	Pos.Opt.
$\theta_{\text{me}\times\text{I(tsp}>0)}$	—	-1.918	-4.042
Std.Err.	—	(0.663)	(0.727)
$\theta_{\text{me}\times\text{I(tsp}\leq 0)}$	—	0.952	1.497
Std.Err.	—	(1.060)	(0.609)
$\theta_{\text{btm}\times\text{I(tsp}>0)}$	—	3.270	5.863
Std.Err.	—	(1.029)	(1.353)
$\theta_{\text{btm}\times\text{I(tsp}\leq 0)}$	—	4.965	1.541
Std.Err.	—	(1.823)	(0.704)
$\theta_{\text{mom}\times\text{I(tsp}>0)}$	—	2.128	6.052
Std.Err.	—	(0.713)	(1.249)
$\theta_{\text{mom}\times\text{I(tsp}\leq 0)}$	—	3.499	1.409
Std.Err.	—	(1.334)	(0.513)
LRT p-value	—	0.000	0.000
$\|w_i\| \times 100$	0.035	0.136	0.035
max $w_i \times 100$	4.625	4.370	1.522
min $w_i \times 100$	0.000	-0.389	0.000
$\sum w_i I(w_i < 0)$	0.000	-1.411	0.000
$\sum I(w_i \leq 0)/N_t$	0.000	0.475	0.498
$\bar{r}$	0.120	0.273	0.187
$\sigma(r)$	0.161	0.203	0.194
CE(r)	0.052	0.171	0.090
α	—	0.191	0.072
β	—	0.357	0.932
$\sigma(\epsilon)$	—	0.196	0.124
IR	—	0.977	0.577
me $\times$ I($tsp > 0$)	1.616	-0.556	-0.408
me $\times$ I($tsp \leq 0$)	0.479	0.340	0.329
btm $\times$ I($tsp > 0$)	-0.346	2.565	0.760
btm $\times$ I($tsp \leq 0$)	-0.105	0.909	0.009
mom $\times$ I($tsp > 0$)	0.027	1.534	0.784
mom $\times$ I($tsp \leq 0$)	-0.015	0.693	0.095

Table 5: Risk Aversion

This table shows estimates of the portfolio policy with three characteristics, size (me), book to market ratio (btm), and momentum (mom) using different utility functions. In additions to the $\gamma = 5$ case, reported as a benchmark, we display the results for $\gamma = 1$, $\gamma = 100$, and well as for the case of minimizing the variance. In the estimation, we use all stocks in the merged COMPUSTAT/CRSP database from January 1964 through December 2002. The columns labeled "Opt." and "Pos.Opt." characterize the optimal and positive optimal (long-only) portfolios, respectively. The first set of rows shows the estimated coefficients of the portfolio policies with standard errors in parenthesis. The p-value of the likelihood ratio test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. The second set of rows shows statistics of the portfolio weights, averaged across time. These statistics include the average absolute portfolio weight ($|w_i|$), the average minimum and maximum portfolio weights (max w_i and min w_i), the average sum of negative weights in the portfolio ($\sum w_i I(w_i < 0)$), and the average fraction of negative weights in the portfolio ($\sum I(w_i \leq 0)/N_t$). The third set of rows displays average portfolio return statistics: average return and standard deviation of returns ($\bar{r}$ and $\sigma(r)$), the certainty equivalent return, 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.

Variable	γ=1 Opt.	γ=5 Opt.	γ=100 Opt.	variance Opt.	γ=1 Pos.Opt.	γ=5 Pos.Opt.	γ=100 Pos.Opt.	variance Pos.Opt.
θ_{me}	-6.109	-1.220	0.146	0.104	-2.970	-0.634	-0.033	0.051
Std.Err.	(2.310)	(0.549)	(0.090)	(0.233)	(0.432)	(0.492)	(1.333)	(0.278)
θ_{btm}	6.970	3.466	5.025	0.317	1.699	4.833	0.114	0.135
Std.Err.	(3.748)	(0.836)	(0.244)	(0.387)	(2.145)	(2.542)	(0.418)	(0.611)
θ_{mom}	7.380	2.000	0.483	-0.129	7.353	4.554	-0.047	-0.059
Std.Err.	(1.862)	(0.590)	(0.061)	(0.327)	(1.581)	(2.125)	(0.768)	(0.331)
LRT p-value	0.008	0.000	0.000	0.090	0.000	0.000	0.000	0.000
$\|w_i\|$	0.320	0.121	0.146	0.040	0.035	0.035	0.035	0.035
max w_i	4.433	4.363	4.497	4.629	1.460	1.460	4.575	4.517
min w_i	-0.960	-0.323	-0.350	-0.071	0.000	0.000	0.000	0.000
$\sum w_i I(w_i < 0)$	-4.026	-1.238	-1.594	-0.074	0.000	0.000	0.000	0.000
$\sum I(w_i \leq 0)/N_t$	0.520	0.468	0.477	0.298	0.524	0.481	0.134	0.231
$\bar{r}$	0.546	0.244	0.190	0.118	0.204	0.186	0.121	0.120
$\sigma(r)$	0.598	0.190	0.187	0.156	0.224	0.195	0.159	0.159
CE(r)	0.363	0.153	-0.976	0.027	0.178	0.083	-1.965	0.025
α	0.507	0.157	0.109	0.000	0.081	0.068	0.002	0.001
β	-0.375	0.440	0.345	0.968	1.051	0.971	0.989	0.991
$\sigma(\epsilon)$	0.595	0.177	0.179	0.015	0.148	0.118	0.007	0.005
IR	0.851	0.890	0.609	0.014	0.548	0.576	0.317	0.163
me	-5.932	-0.118	0.744	2.101	-0.197	0.023	2.000	2.031
btm	8.043	3.296	4.508	-0.160	0.232	0.851	-0.333	-0.343
mom	6.958	1.839	0.290	-0.128	1.739	0.987	-0.020	-0.018

Table 6: Benchmarking

This table shows estimates of the portfolio policy with the three characteristics , size (me), book to market ratio (btm), and momentum (mom), estimated using the performance of the portfolio relative to a benchmark portfolio r_{bench}. We consider two benchmark portfolios, the market portfolio and a 10-year bond portfolio, displayed in Panels A and B, respectively. In the estimation, we use the largest 500 companies in each month from January 1964 through December 2002. The column labeled "Val.Weighted" displays the statistics of the market-capitalization weighted portfolio, as a benchmark. The next two columns, "Opt." and "Pos.Opt." characterize the optimal and positive optimal (long-only) portfolios, respectively. The first set of rows shows the estimated coefficients of the portfolio policies with standard errors in parenthesis. The p-value of the likelihood ratio test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. The second set of rows shows statistics of the portfolio weights, averaged across time. These statistics include the average absolute portfolio weight ($|w_i|$), the average minimum and maximum portfolio weights (max w_i and min w_i), the average sum of negative weights in the portfolio ($\sum w_i I(w_i < 0)$), and the average fraction of negative weights in the portfolio ($\sum I(w_i \leq 0)/N_t$). The third set of rows displays average return ($\bar{r}$), standard deviation of returns $\sigma(r)$), and the certainty equivalent return. We also report the alpha, beta, volatility of idiosyncratic shocks, and information ratio of a regression of the portfolio return on the benchmark return. The final set of rows displays the average normalized characteristics of the portfolio relative to the benchmarch.

	Panel A: Equity Market Benchmark			Panel B: Long Bond Benchmark		
	Val.Weighted	Opt.	Pos.Opt.	Val.Weighted	Opt.	Pos.Opt.
θ_{me}	—	-3.717	-2.213	—	-1.252	-0.836
Std.Err.	—	(1.104)	(1.196)	—	(1.238)	(0.658)
θ_{btm}	—	0.314	0.487	—	1.685	2.024
Std.Err.	—	(0.648)	(1.918)	—	(0.680)	(1.097)
θ_{mom}	—	1.924	4.671	—	1.670	2.322
Std.Err.	—	(0.478)	(1.772)	—	(0.513)	(1.182)
LRT p-value	—	0.000	0.000	—	0.016	0.000
$\lvert w_i \rvert \times 100$	0.200	0.557	0.200	0.200	0.413	0.200
max $w_i \times 100$	5.343	3.537	3.163	5.343	4.295	2.759
min $w_i \times 100$	0.037	-1.400	0.000	0.037	-1.146	0.000
$\sum w_i I(w_i < 0)$	0.000	-0.891	0.000	0.000	-0.532	0.000
$\sum I(w_i \leq 0)/N_t$	0.000	0.378	0.451	0.000	0.361	0.390
$\bar{r}$	0.070	0.182	0.145	0.099	0.180	0.146
$\sigma(r)$	0.035	0.136	0.082	0.164	0.173	0.168
CE(r)	0.067	0.135	0.129	0.025	0.092	0.066
α	0.009	0.115	0.076	0.048	0.127	0.094
β	0.986	1.125	1.157	0.409	0.507	0.451
$\sigma(\epsilon)$	0.034	0.135	0.079	0.153	0.166	0.159
IR	0.852	0.852	0.970	0.765	0.765	0.594
me	1.527	-2.273	-0.209	1.527	-0.019	0.185
btm	-0.240	0.535	-0.027	-0.240	1.539	0.632
mom	-0.026	1.959	1.476	-0.026	1.573	0.856

Table 7: Long-Horizon Portfolio Policy

This table shows estimates of a long-horizon portfolio policy with three characteristics, size (me), book to market ratio (btm), and momentum (mom) over horizons of K periods, where K is equal to 1, 12, and 60 months. For each horizon, the estimates are obtained by maximizing average utility over the cumulative, non-overlapping K-period portfolio returns. The results for the "Val.Weighted" and the $K = 1$ case are identical to those presented in the above tables and are shown here for reference. Panels A and B display the results from the unrestricted and positive (long-only) weights. In the estimation, we use all stocks in the merged COMPUSTAT/CRSP database from January 1964 through December 2002. The first set of rows shows the estimated coefficients of the portfolio policy with standard errors in parenthesis. The p-value of the likelihood ratio test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. The second set of rows shows statistics of the portfolio weights, averaged across time. (Cont'd on next page.)

	Panel A: Long and Short Positions			
	Val.Weighted	K=1	K=12	K=60
θ_{me}	—	-1.220	-1.205	-3.504
Std.Err.	—	(0.549)	(0.636)	(1.124)
θ_{btm}	—	3.466	2.623	3.969
Std.Err.	—	(0.836)	(0.837)	(1.928)
θ_{mom}	—	2.000	2.909	3.536
Std.Err.	—	(0.590)	(0.858)	(1.574)
LRT p-value	—	0.000	0.000	0.000
$\lvert w_i \rvert \times 100$	0.035	0.121	0.114	0.144
max $w_i \times 100$	4.625	4.363	4.393	3.712
min $w_i \times 100$	0.000	-0.323	-0.327	-0.415
$\sum w_i I(w_i < 0)$	0.000	-1.238	-1.141	-1.960
$\sum I(w_i \leq 0)/N_t$	0.000	0.468	0.463	0.498
$\bar{r}$	0.120	0.244	0.258	0.321
$\sigma(r)$	0.161	0.190	0.207	0.315
CE(r)	0.052	0.153	0.161	0.365
α	—	0.157	0.167	0.248
β	—	0.440	0.512	0.179
$\sigma(\epsilon)$	—	0.177	0.190	0.314
IR	—	0.890	0.878	0.788
me	2.095	-0.118	0.170	-2.439
btm	-0.451	3.296	2.412	4.423
mom	0.012	1.839	2.783	3.296

Table continued on next page ...

Cont'd from previous page: These statistics include the average absolute portfolio weight ($|w_i|$), the average minimum and maximum portfolio weights (max w_i and min w_i), the average sum of negative weights in the portfolio ($\sum w_i I(w_i < 0)$), and the average fraction of negative weights in the portfolio ($\sum I(w_i \leq 0)/N_t$). The third set of rows displays average portfolio return statistics: average return and standard deviation of returns ($\bar{r}$ and $\sigma(r)$), the certainty equivalent return, 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.

	Panel A: Long-Only Positions			
	Val.Weighted	K=1	K=12	K=60
θ_{me}	—	-0.634	-1.729	-2.169
Std.Err.	—	(0.492)	(0.488)	(1.931)
θ_{btm}	—	4.833	4.306	1.954
Std.Err.	—	(2.542)	(0.696)	(4.765)
θ_{mom}	—	4.554	8.794	9.228
Std.Err.	—	(2.125)	(1.147)	(2.777)
LRT p-value	—	0.000	0.000	0.000
$\lvert w_i \rvert \times 100$	0.035	0.035	0.035	0.029
max $w_i \times 100$	4.625	1.460	1.285	1.254
min $w_i \times 100$	0.000	0.000	0.000	0.000
$\sum w_i I(w_i < 0)$	0.000	0.000	0.000	0.000
$\sum I(w_i \leq 0)/N_t$	0.000	0.481	0.512	0.530
$\bar{r}$	0.120	0.186	0.202	0.190
$\sigma(r)$	0.161	0.195	0.217	0.230
CE(r)	0.052	0.083	0.119	0.175
α	—	0.068	0.080	0.066
β	—	0.971	1.031	1.058
$\sigma(\epsilon)$	—	0.118	0.141	0.148
IR	—	0.576	0.571	0.444
me	2.095	0.023	-0.147	-0.067
btm	-0.451	0.851	0.507	0.153
mom	0.012	0.987	1.609	1.868

Table continued from previous page.

Table 8: Out-of-Sample Performance

This table shows the out-of-sample performance of the portfolio policy with three characteristics, size (me), book to market ratio (btm), and momentum (mom). Panels A and B display the results from the unrestricted and positive (long-only) weights. The original data of all stocks in the merged COMPUSTAT/CRSP database is divided into two equal sub-samples, Sample 1 (January 1964 through June 1983) and Sample 2 (July 1983 through December 2002). In each sub-sample, we display in-sample and out-of-sample statistics. For the in-sample results, the column labeled "Val.Weighted" displays the statistics of the market-capitalization weighted portfolio, whereas the column "Opt." characterizes the optimal portfolio. The out-of-sample forecasts in Sample 1 are obtained using the in-sample estimates of the optimal investment policy function from Sample 2 in forming the optimal portfolio. Similarly, the out-of-sample forecasts in Sample 2 are obtained from the in-sample estimates in Sample 1. The optimal portfolio parameters are not re-estimated in the forecasts. The first set of rows shows the estimated coefficients of the portfolio policies with standard errors in parenthesis. (Cont'd on next page.)

Panel A: Long and Short Positions

	Sample 1			Sample 2		
	In-Sample		Out-of-Sample	In-Sample		Out-of-Sample
	Val.Weighted	Opt.	Fcst.	Val.Weighted	Opt.	Fcst.
θ_{me}	—	-1.707	-0.702	—	-0.702	-1.707
Std.Err.	—	(0.822)	—	—	(0.726)	—
θ_{btm}	—	2.891	4.166	—	4.166	2.891
Std.Err.	—	(1.357)	—	—	(1.084)	—
θ_{mom}	—	3.111	1.346	—	1.346	3.111
Std.Err.	—	(1.056)	—	—	(0.704)	—
LRT p-value	—	0.010	—	—	0.001	—
$\lvert w_i \rvert$	0.049	0.176	0.180	0.021	0.079	0.075
max w_i	6.119	5.661	5.802	3.130	3.019	2.990
min w_i	0.002	-5.158	-4.719	0.000	-1.812	-2.048
$\sum w_i I(w_i < 0)$	0.000	-1.307	-1.359	0.000	-1.363	-1.289
$\sum I(w_i \leq 0)/N_t$	0.000	0.470	0.464	0.000	0.480	0.478
$\bar{r}$	0.103	0.293	0.201	0.136	0.245	0.265
$\sigma(r)$	0.155	0.233	0.194	0.166	0.163	0.224
CE(r)	0.043	0.151	0.113	0.061	0.177	0.121
α	—	0.197	0.110	—	0.177	0.197
β	—	0.795	0.660	—	0.154	0.153
$\sigma(\epsilon)$	—	0.198	0.165	—	0.162	0.223
IR	—	0.994	0.668	—	1.095	0.884
me	2.078	-0.586	0.026	2.111	0.322	-0.228
btm	-0.528	2.732	3.784	-0.375	3.952	2.913
mom	-0.070	2.890	1.044	0.094	1.293	3.034

Table continued on next page ...

Cont'd from previous page: The p-value of the likelihood ratio test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. For the out-of-sample results, we display the coefficients that were used in producing the forecasts. The second set of rows shows statistics of the portfolio weights, averaged across time. These statistics include the average absolute portfolio weight ($|w_i|$), the average minimum and maximum portfolio weights (max w_i and min w_i), the average sum of negative weights in the portfolio ($\sum w_i I(w_i < 0)$), and the average fraction of negative weights in the portfolio ($\sum I(w_i \leq 0)/N_t$). The third set of rows displays average portfolio return statistics: average return and standard deviation of returns ($\bar{r}$ and $\sigma(r)$), the certainty equivalent return, 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.

	Panel B: Long-Only Positions					
	Sample 1			Sample 2		
	In-Sample		Out-of-Sample	In-Sample		Out-of-Sample
	Val.Weighted	Opt.	Fcst.	Val.Weighted	Opt.	Fcst.
θ_{me}	—	-0.821	-0.491	—	-0.351	-0.821
Std.Err.	—	(5.497)	—	—	(1.274)	—
θ_{btm}	—	1.707	1.060	—	0.036	1.707
Std.Err.	—	(8.532)	—	—	(1.929)	—
θ_{mom}	—	3.916	0.789	—	0.520	3.916
Std.Err.	—	(3.498)	—	—	(1.507)	—
LRT p-value	—	0.000	—	—	0.000	—
$\lvert w_i \rvert$	0.049	0.049	0.049	0.021	0.021	0.021
max w_i	6.119	2.749	4.677	3.130	2.925	1.445
min w_i	0.002	0.000	0.000	0.000	0.000	0.000
$\sum w_i I(w_i < 0)$	0.000	0.000	0.000	0.000	0.000	0.000
$\sum I(w_i \leq 0)/N_t$	0.000	0.473	0.356	0.000	0.276	0.472
$\bar{r}$	0.103	0.212	0.160	0.136	0.145	0.164
$\sigma(r)$	0.155	0.208	0.175	0.166	0.163	0.190
CE(r)	0.043	0.098	0.083	0.061	0.072	0.061
α	—	0.103	0.055	—	0.013	0.036
β	—	1.157	1.049	—	0.956	0.903
$\sigma(\epsilon)$	—	0.105	0.063	—	0.035	0.118
IR	—	0.987	0.873	—	0.367	0.307
me	2.078	0.296	0.957	2.111	1.669	0.411
btm	-0.528	0.223	0.279	-0.375	-0.263	0.327
mom	-0.070	1.312	0.371	0.094	0.502	1.405

Table continued from previous page.

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 1964 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 12 month return. The reported means and standard deviations are computed across firms at each point in time.

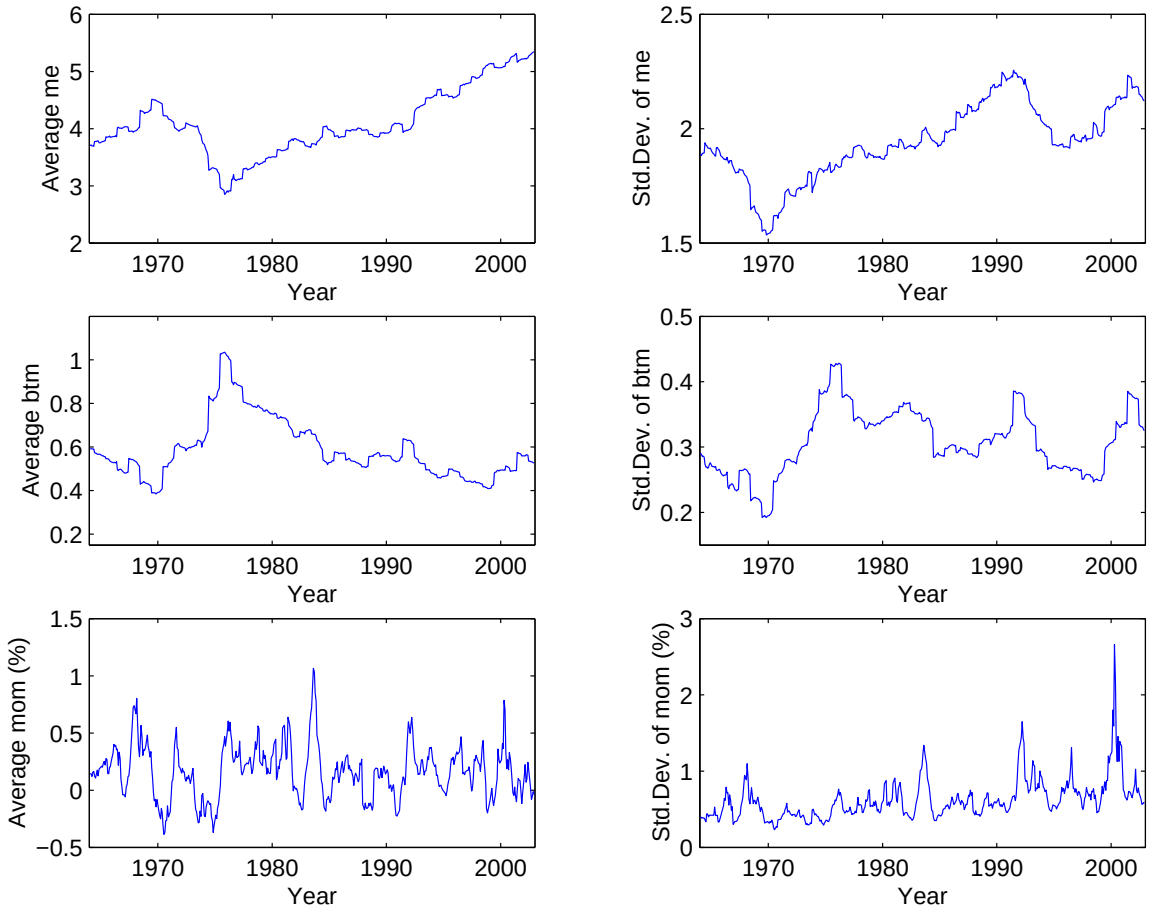
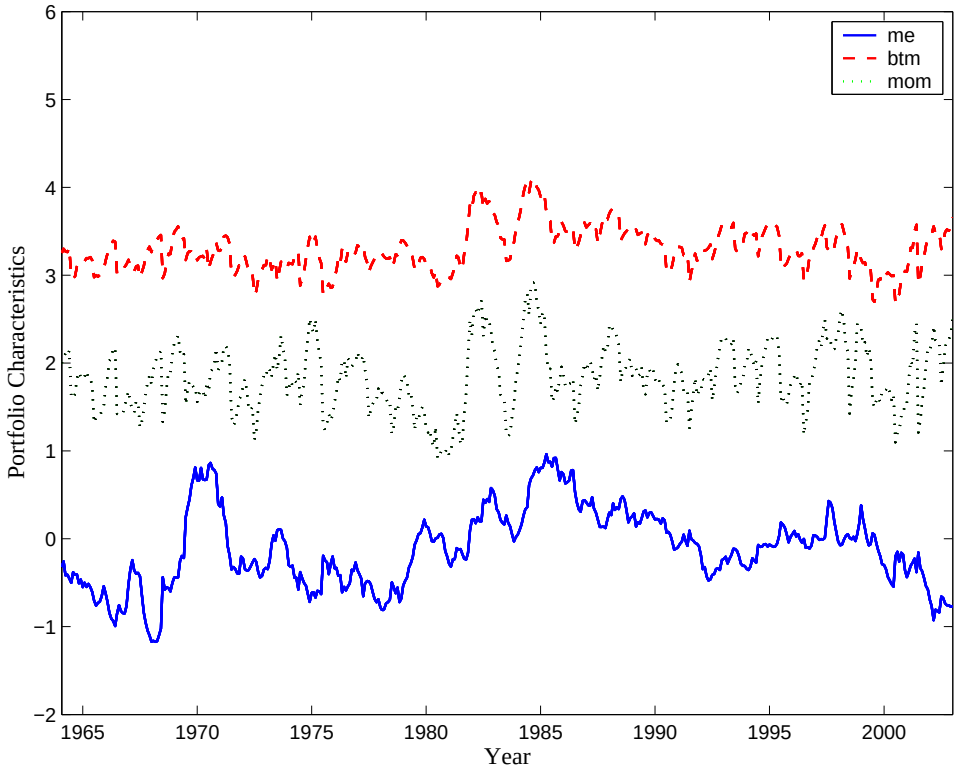


Figure 2: Portfolio Characteristics over Time

The figure displays the portfolio characteristics of policy function (2) 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 1964 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.