

Nonparametric Portfolio Estimation and Asset Allocation

by

Julian James Douglass

B.Sc.E, Queen's University, 1992

M.Sc., The University of British Columbia, 1995

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

The Faculty of Graduate Studies

(Business Administration)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

February, 2009

© Julian James Douglass 2009

Abstract

This thesis comprises two essays that apply nonparametric methods to the estimation of portfolio allocations.

In the first essay, I test the significance to investor welfare of (i) adding additional assets to the portfolio choice set and (ii) conditioning on predictor variables. I estimate unconditional and conditional optimal allocations of a constant relative risk aversion investor by maximizing a nonparametric approximation of the expected utility integral. Investors can improve their expected utility significantly over that of an equities and cash investor by adding portfolios based on the value or momentum premiums into their asset allocation decision. In contrast, neither a size premium portfolio nor a long-term bond portfolio improves expected utility. The significance of predictability is increased by simultaneously conditioning on the two strongest predictors (of eight) studied: the term spread and the gold industry trend.

In the second essay, I formulate a nonparametric estimator that permits combining historical data with a qualitative prior. I investigate the impact of an investor belief, motivated by asset-pricing theory, that optimal allocations are positive. In the estimator construction, I use a Bayesian approach to perturb the probabilities associated with each data point in the empirical distribution to reflect qualitative prior beliefs. In a simulation study and in out-of-sample tests, I find that portfolio estimates conditioned on a belief in the positivity of portfolio weights are significantly more stable than those estimated by an uninformed investor, and that the model performs better in out-of-sample tests than a number of plug-in models. However, the out-of-sample performance lags that of the minimum-variance and $1/N$ policies.

Table of Contents

Abstract	ii
Table of Contents	iii
List of Tables	v
List of Figures	vi
Acknowledgements	viii
Dedication	ix
1 Introduction	1
1.1 Using Inference to Assess the Benefits of Diversification	3
1.2 Prior Disbelief in the Optimality of Short Positions	4
1.3 Nonparametric Portfolio Estimation	5
Bibliography	7
2 Asset Allocation with Value Growth Tilts and Predictability	9
2.1 Introduction	9
2.2 Empirical Framework	16
2.2.1 Investor's Problem	16
2.2.2 Direct Estimator of Conditional Portfolio Policy	17
2.2.3 Measure of Portfolio Performance	19
2.2.4 Data and Utility Assumptions	21
2.3 Unconditional Asset Allocation	23
2.4 Conditional Asset Allocation	28
2.4.1 Predictability of Mean and Variance	28
2.4.2 Predictability in an Asset Allocation Framework	29
2.5 Conditional Asset Allocation with Multiple Assets	32
2.6 Conditioning on Multiple Predictors	37
2.6.1 Direct Estimator for Multiple Predictors	37
2.6.2 Model Selection	43
2.6.3 Certainty Equivalent Returns with Multiple Predictors	44
2.7 Conclusion	44

Bibliography	78
3 Nonparametric Portfolio Estimation with Prior Belief in the Positivity of Portfolio Weights	82
3.1 Introduction	82
3.2 Portfolio Choice Framework	87
3.3 Discrete-Bayesian Estimator	88
3.3.1 Construction of the Discrete-Bayesian Estimator	90
3.3.2 Sampling from the Posterior	93
3.3.3 Mean-Variance Utility	93
3.4 Prior	94
3.5 Relationship to Previous Portfolio Choice Models	97
3.5.1 Plug-in Models	97
3.5.2 Portfolio Resampling	99
3.5.3 Statistical Refinement versus Regularization	102
3.6 Portfolios with Positive Weights Prior	103
3.6.1 Simulation Setup	103
3.6.2 Exploratory Analysis of the Prior Assumptions	104
3.6.3 Monte-Carlo Study	107
3.6.4 Comparison of Discrete-Bayesian Policy with Other Models	109
3.7 Out-of-Sample Study	112
3.8 Conclusion	114
Bibliography	139
4 Conclusion	143
Bibliography	146
Appendices	
A Bandwidth Specification	147
Bibliography	149
B Data Tuning for a Portfolio Choice Problem	150
Bibliography	152
C Sampling from the Posterior	153
Bibliography	156

List of Tables

2.1	Summary Statistics	46
2.2	Correlations of Asset Returns and Predictors	47
2.3	Volatility Correlations	48
2.4	Unconditional Portfolio Weights	49
2.5	Unconditional Certainty Equivalent Returns	50
2.6	Bootstrap Test of Expected Utility Improvement	51
2.7	Abbreviations Used for Conditioning Variables	52
2.8	Returns Regressed on Predictors	53
2.9	Realized Volatility Regressed on Predictors	54
2.10	Portfolio Weights as a Function of Predictor Values	55
2.11	Certainty Equivalent Returns as a Function of Predictor Values	57
2.12	Increase in Certainty Equivalent Returns from Conditioning on a Predictor	59
2.13	Coefficients of Predictive Index	60
2.14	Conditional Certainty Equivalents for Predictive Index	61
3.1	Portfolio Choice Models	116
3.2	Simulation Parameters	117
3.3	Posterior Distribution Given a Simulated Sample	118
3.4	Monte Carlo Statistics for Predictive Distributions	120
3.5	Monte Carlo Averages of Estimated Allocations	121
3.6	Monte-Carlo Average Out-of-Sample Utilities	123
3.7	Monte-Carlo Average Portfolio Allocations for Alternative Port- folio Estimates	124
3.8	Monte-Carlo Average Out-of-Sample Utility for Alternative Port- folio Estimates	125
3.9	Out-of-Sample Utility for Discrete-Bayesian Estimator	126
3.10	Out-of-Sample Utility for Alternative Models	128

List of Figures

2.1	Single Asset Portfolios Conditioned on Dividend Yield and Default Premium	63
2.2	Single Asset Portfolios Conditioned on Term Spread, S&P 500 Trend, and T-bill Rate	64
2.3	Single Asset Portfolios Conditioned on Gold Industry Trend, Realized Volatility, and Inflation	65
2.4	Portfolio Allocations versus Dividend Yield	66
2.5	Portfolio Allocations versus Default Premium	67
2.6	Portfolio Allocations versus Term Spread	68
2.7	Portfolio Allocations versus S&P 500 Trend	69
2.8	Portfolio Allocations versus T-bill Rate	70
2.9	Portfolio Allocations versus Gold Industry Trend	71
2.10	Portfolio Allocations versus Realized Volatility	72
2.11	Portfolio Allocations versus Inflation	73
2.12	CER versus Default Premium and Dividend Yield	74
2.13	CER versus Term Spread and Index Trend	75
2.14	CER versus Tbill Yield and Gold Industry Trend	76
2.15	CER versus versus Inflation and Volatility	77
3.1	Histogram of Posterior Returns for a Simulated Data Set	131
3.2	Histogram of Posterior Weights	132
3.3	Histogram of Posterior Returns with Prior Against Return Bias	133
3.4	Histogram of Posterior Weights with Prior Against Return Bias	134
3.5	Histogram of Posterior Weights with Strong Prior	135
3.6	Histogram of Predictive Return Distribution under Different Priors	136
3.7	Histogram of Asset Weights Across Simulations	137

List of Algorithms

C.1	MCMC Algorithm for Drawing from the Discrete Posterior Distribution	155
-----	---	-----

Acknowledgements

I am greatly indebted to a large number of colleagues, friends, and family who supported me during my years in the Ph.D. program at the Sauder School.

First, I would like to thank the faculty and Ph.D. colleagues of the Finance Division who provided a stimulating environment for research. My committee members, Marcin Kacperczyk, Robert Heinkel, and Shinichi Sakata, provided invaluable input and advice. I would especially like to thank my advisor, Adlai Fisher, for continuous encouragement and support throughout my thesis project. I would also like to thank Murray Carlson for many stimulating conversations, and Tan Wang and Alan Kraus for input during the early stages of this project. I also benefited from guidance from Glen Donaldson, Ron Giammarino, and Maurice Levi.

I also thank my friends and colleagues, particularly, Issouf Soumaré, Lars-Alexander Kuehn, Kyung Shim, Jeff Colpitts, Jason and Jenny Chen, Harjoat Bhamra, and Ali Lazrak.

I owe a great debt to friends and family both in Vancouver and far away. The names are too numerous to list without missing someone. This thesis would not have been possible without their constant support.

Finally, I thank Andrea whose companionship lightened the burden even during some of the most difficult challenges. I look forward to future projects and adventures with her at my side.

To Andrea

Chapter 1

Introduction

The information in historical returns is not sufficient, on its own, to provide useful estimates of optimal portfolio allocations. Previous researchers have documented overwhelming evidence that portfolios estimated by direct substitution of distributions estimated from previous realized returns feature extreme portfolio weights and perform poorly out of sample.¹ Furthermore, accounting for this estimation error in a Bayesian framework that accounts for estimation uncertainty without introducing informative prior beliefs only slightly improves expected portfolio performance.²

Portfolio allocation is a problem that is plagued by estimation error. The objective of portfolio allocation is to select the portfolio that yields the optimal balance between future rewards and risk. The theory of choice under uncertainty yields an elegant mathematical framework for the portfolio problem.³ In turn, allocation models expressed in this framework are tantalizingly straightforward to solve, given the investor's risk aversion and the future distribution of returns.⁴ Yet, in practice, the future distribution of returns is unknown and must be estimated. Asset returns are highly volatile, and return means estimated from time series of past realized returns are very noisy. In addition, optimal allocations are inversely proportional to asset variances for risk averse investors. Thus, optimal portfolios depend on the inverse of the covariance ma-

¹See, for example, Jobson and Korkie (1980), Jobson and Korkie (1981), Best and Grauer (1991), and Chopra and Ziemba (1993).

²Early empirical studies of Bayesian approaches to portfolio choice appear in Bawa et al. (1979) and references therein. See Kan and Zhou (2007) for further discussion and results pertaining to the relative performance of plug-in and diffuse Bayesian portfolio estimators.

³Markowitz (1952) developed the mean-variance formulation of this problem that remains the canonical mathematical formulation of the allocation problem. In this framework, rewards as represented by expected portfolio return is balanced against increased risk from portfolio variance.

⁴A major advantage of the mean-variance model is its analytic tractability. Alternative utility assumptions may lead to less tractable models, but these, too, are readily solved computationally by convex programming algorithms.

trix. In applications, this matrix is often ill conditioned because financial asset returns are highly correlated. As a result, errors in variances and correlations are amplified in the portfolio solution.

This thesis comprises two essays that examine issues in portfolio estimation. The first essay examines the question of asset selection for the tactical allocation problem. Given that portfolio weights are computed from observed data, they are naturally viewed as statistical estimates. Using inference techniques, I test the significance of gains in expected utility from adding assets to a tactical asset allocation portfolio, and of conditioning on predictor variables in the context of tactical asset allocation.

In the second essay, I construct a nonparametric estimator for the portfolio choice problem that incorporates qualitative prior beliefs into the investor's estimation problem. In an empirical exercise, I consider an industry allocation problem. I examine the value of a prior belief that all optimal portfolio weights are positive. The prior is motivated by economic theory that suggests that, given that all industries have positive weight in the market portfolio, that they are unlikely to be priced such that short positions are optimal.

A common theme of both essays is the nonparametric estimation strategy. In the first essay, I evaluate the value of conditioning variables following the nonparametric strategy introduced by Brandt (1999). In the second essay, use Bayesian computational techniques to incorporate qualitative prior beliefs in the investor's problem. As suggested in the pioneering work of Jobson and Korkie (1980), I treat the investor's utility maximization as the statistical objective when evaluating conditioning information. I illustrate the high threshold that must be overcome before we can conclude that conditioning information is helpful to an investor in a statistically meaningful way.

The remainder of the introduction motivates and reviews the empirical contributions, provides background for the estimation strategy, and describes my methodological contribution.

1.1 Using Inference to Assess the Benefits of Diversification

In the first essay, I use an inference approach to examine the benefits of diversification given an optimal portfolio estimated from return data. One of the fundamental corollaries of portfolio theory is that adding additional assets to the set available for diversification improves outcomes.⁵ In the presence of estimation error, there is no guarantee that including the diversifying asset will improve subsequent utility.

For the investor, portfolio estimates are empirical forecasts of their optimal allocations. If an investor can precisely estimate their optimal portfolio then adding an additional asset always improves expected utility⁶ However, adding an asset can also reduce the precision of the optimal portfolio estimator. Consider the very simple example of an investor who is choosing whether to invest solely in the risk free asset or whether to select an optimal portfolio that combines the risky asset and the risk free rate. In the first case the optimal portfolio is trivially and precisely identified since there is only one asset. In the second case, the optimal portfolio depends critically on the investor's estimate of the probability distribution of returns. There is no guarantee that the *estimated* optimal portfolio will have a higher expected utility with respect to the underlying return-generating distribution. I use a hypothesis testing approach to test whether an increase in utility from diversification is large enough to statistically reject the null that the investor would have been better off investing in the original, less diversified portfolio.

Following Brandt (1999), I formulate the portfolio allocation problem itself as a statistical estimation problem. The finite sample estimator of expected utility plays the same role as a statistical loss function⁷, and portfolio weights are parameters to be estimated.

I test the significance of the diversification benefits of various assets for a United States investor. The investor has a choice of assets that includes an

⁵Standard textbook discussions always caution that diversifying beyond one or two dozen equities yields little diversification benefit.

⁶I am assuming the added asset's returns are not perfectly correlated with any portfolio formed from the existing assets.

⁷Actually, statistical optimizations are usually formulated as minimizations so negative utility is the direct analog of a statistical loss function.

equity index, a long-term bond index, and an asset formed by going long assets with high book to market ratio with offsetting short position in assets with low book to market ratios. I find that inclusion of the long-term bond asset does not significantly increase the estimated utility. This is surprising since a bond portfolio is a standard component in practical asset allocation problems. By contrast investment in the value premium significantly improves the expected utility estimate.

Furthermore, I examine the significance of predictability of index returns. I test against the hypothesis that conditioning allocations on predictor variables does not increase expected utility of the investor. I find that this hypothesis cannot be rejected for most of the predictor variables considered. The exceptions are the term spread and the gold industry trend which yield marginally significant improvement in unconditional expected utility.

1.2 Prior Disbelief in the Optimality of Short Positions

The second essay examines whether portfolio estimates are improved when conditioned on additional, economically motivated insights. The idea is to improve the robustness of portfolio estimates by introducing an informative, non-data prior. I accommodate the additional prior by formulating the portfolio estimation problem in a Bayesian setting.

I consider the prior belief that all asset allocations are positive at the optimum. This prior has two motivations. First, the prior is motivated by economic theory. Classic asset pricing models such as the CAPM (Sharpe (1964)) hold that assets are priced such that their market weights reflect the optimal portfolio of an aggregate investor. For assets that have positive market value, the implication is that they will appear with positive weight in the optimal portfolio of individual investors. The second motivation for the no-short-position prior is the importance of short-sales constraints in portfolio allocation applications. While many investment managers impose short-sales constraints due to limitations in their mandates, the presence of such constraints may also reflect the lack of availability of Bayesian technology in standard portfolio selection implementations. As a result, the only convenient method of incorporating a

prior belief is to impose it directly via the use of constraints.

In a simulation study, I examine the impact of a prior disbelief in the optimality of short positions. I consider a five-asset universe and simulate return histories from a generating distribution based on the historical distribution of returns for a five-industry breakdown of the United States equity universe.

I find that the expected out-of-sample performance under the no-short-sales prior improves over the expected out-of-sample performance under a diffuse prior. The model performs similarly to approaches that impose short-sales constraints directly. However, the method does not perform as well as the minimum variance and equal weight (or $1/N$) rules. For the case considered, both of these rules are asymptotically biased. The true optimal minimum variance portfolio includes negative weights while the weights under the $1/N$ rule do not equal the true optimal mean variance weights.

1.3 Nonparametric Portfolio Estimation

I use a nonparametric estimator of portfolio weights. For the inference analysis in the first essay, I employ the formulation of Brandt (1999).⁸ In the second essay, I develop a technique for tuning the data to obtain an empirical distribution of returns that reflects an investor's qualitative (or non-data) priors.⁹ I use Bayesian methods to adapt the probabilities attached to each data point in the return distribution to the non-data information. Because I discretize the set of possible return outcomes to those observed in the historical data, the data informs the setup of the prior, and the resulting estimator is not strictly Bayesian. However, restricting to return outcomes in the investor's data set is an effective means of discretizing a high dimensional space of possible return outcomes to a parsimonious grid. Because the actual data generating points is setting the grid, the included points are naturally concentrated near peaks of the data generating distribution. Effectively, the domain of the posterior is restricted to a set of multinomial distributions. However, the set allows for a

⁸Other applications of Brandt's (1999) method can be found in Aït-Sahalia and Brandt (2001), Paye (2004), and Chapter 2.

⁹Previous applications of data tuning focus on minimally perturbing either the probabilities or the data itself until some constraint on the solution is satisfied. See Braun and Hall (2001) for a discussion of these methods and their application.

broad range of distributional properties and permits effective incorporation of prior information. I integrate the Bayesian posterior distribution of returns by Markov chain Monte Carlo.

In both essays, I combine historical return data with other information. In the first essay, the additional information is the value of a predictor variable while in the second essay the additional information is a qualitative prior. Unsurprisingly, the resulting estimators have a common nonparametric form. In each case, the problem is to choose weights that solve the investor's problem,

$$\max_w E[U(\tilde{r}, w) | R, \mathcal{I}], \quad (1.1)$$

where w is a vector of portfolio weights, \tilde{r} is a vector of unknown future returns, R is a matrix of past asset returns, and \mathcal{I} is the additional information. The resulting nonparametric estimator is given by

$$\max_w \sum_{t=1}^T a_t U(\tilde{r}_t, w), \quad (1.2)$$

where r_t is the t th row of the return matrix R , and the a_t are coefficient weights that depend on the non-return information \mathcal{I} . For the unconditional case with no additional information, the a_t are constant. When conditioning on a predictor variable, the a_t depend on the distance between the historical value of the predictor and the value at time of investment. Finally, in the Bayesian case, the a_t are obtained following integration over the posterior. The incorporation of prior beliefs into the nonparametric portfolio estimator is an important methodological contribution of the second essay.

Bibliography

Aït-Sahalia, Y. and Brandt, M. (2001). Variable selection for portfolio choice. *Journal of Finance*, 56:1297–1351.

Bawa, V. S., Brown, S., and Klein, R. (1979). *Estimation Risk and Optimal Portfolio Choice*. North Holland, Amsterdam.

Best, M. J. and Grauer, R. R. (1991). On the sensitivity of mean-variance efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies*, 4:315–342.

Brandt, M. (1999). Estimating portfolio and consumption choice: A conditional Euler equations approach. *Journal of Finance*, 54:1609–1645.

Braun, W. J. and Hall, P. (2001). Data sharpening for nonparametric inference subject to constraints. *Journal of Computational and Graphical Statistics*, 10(4):786–806.

Chopra, V. K. and Ziemba, W. T. (1993). The effect of errors in means, variances and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19:6–11.

Jobson, J. D. and Korkie, B. M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance*, 36(4):889–908.

Jobson, J. D. and Korkie, R. (1980). Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association*, 75:544–554.

Kan, R. and Zhou, G. (2007). Portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3):621–656.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7:77–91.

Paye, B. (2004). *Essays on Stock Return Predictability and Portfolio Allocation*. PhD thesis, University of California San Diego.

Sharpe, W. F. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance*, 19:425–442.

Chapter 2

Asset Allocation with Value Growth Tilts and Predictability¹⁰

2.1 Introduction

Researchers document evidence of two empirical features of financial asset returns with implications for asset allocation. First, a market index is not sufficient to span systematic risk in the cross-section of equity returns.¹¹ Second, equity and bond index returns can be predicted by variables such as the dividend yield, term spread, and short term interest rate.¹²

An investor is interested in whether conditioning their portfolio policy on either predictability or multi-factor structure improves their expected welfare. More specifically, an investor might ask, “Is the gain in expected utility from conditioning on an empirical feature significant, given the evidence in the data?” A portfolio estimate that conditions on the presence of an additional factor or predictor will always yield a higher expected-utility estimate in-sample, simply from diversification. An important question is whether the utility gain is significant given that the true optimal portfolio is unknown. The investor cannot be certain that the expected utility for an estimated portfolio with a diversifying asset is an improvement over the best estimated portfolio that excludes the additional asset. The actual generating distribution for returns is unknown. In other words, the investor faces a forecasting problem. Adding an asset adds a

¹⁰A version of this chapter will be submitted for publication. Douglass, J., Asset Allocation with Value Growth Tilts and Predictability.

¹¹The presence of multiple factors in equity returns can be observed via principle components analysis (Connor and Korajczyk (1993)).

¹²See Rey (2004) and Ang and Bekaert (2007) for reviews of predictability evidence. Cochrane (1999) reviews investment implications of multifactor pricing and predictability.

parameter to the estimation problem. The extra parameter will permit finding a solution that yields a higher value for the estimation optimand. However, the statistical power is lessened so the out-of-sample performance of the estimator may be lower than under a more parsimonious asset set.

In this chapter, I evaluate the significance of estimated expected utility gains from conditioning on multifactor pricing and predictability. For ease of interpretation, I express expected utility as a certainty equivalent return (CER).¹³ I treat the portfolio optimization problem as a statistical optimization. As such, the expected utility achieved at the optimum is a random variable. I employ bootstrap resampling to estimate the distribution of expected utility for a given policy. I compute p-values for the null hypothesis that investing according to an estimated optimal portfolio that conditions on an empirical feature yields no utility gain over a policy that conditions on asset returns only.¹⁴

My results build on those of Brandt (1999), Aït-Sahalia and Brandt (2001), and Paye (2004). Following Brandt (1999), I estimate allocations in a single step by direct solution of the Euler equation of the investor's problem. The estimation is achieved by replacing the expectation integral with an average over the outcomes for each data point. This nonparametric approach has the advantage of bypassing the need to assume and then estimate a parametric distribution of returns. This eliminates a potential source of misspecification, as well as the problem of explicitly determining which moments of the return distribution are of interest. Direct estimation of portfolio weights leads to a transparent interpretation of portfolio allocations as statistical estimates.

I assess investment significance based on robustness of expected utility gains to estimation error. Treatment of expected utility as a statistical estimate has its roots in the work of Jobson and Korkie (1980). The investor's problem is analogous to a statistical optimization. Consider a comparison with the familiar problem of linear regression. For estimation in an investment framework,

¹³A convention that is standard in the portfolio choice literature. See Kandel and Stambaugh (1996).

¹⁴Goetzmann and Jorion (1993) apply a bootstrap approach to the assessment of gains in R^2 in linear regressions of returns on the dividend yield. Their bootstrap approach differs from mine in that they focus on comparing the predictive regression R^2 with the distribution of regression R^2 s in bootstrap samples constructed to eliminate conditional correlation. I bootstrap from the original data and use the utility difference between the conditional and unconditional strategies to test the null hypothesis.

portfolio weights play the same role as coefficients in a regression, while expected utility is analogous to R^2 . As in regression studies, in which R^2 is rarely accompanied by a measure of precision (Press and Zellner (1978)), estimated expected utility is rarely quoted with a standard error.

In general, expected utility estimates based on finite samples have unknown distribution.¹⁵ There are two common approaches to estimating standard errors under the finite distribution. One can either use a large sample, asymptotic approximation or compute standard errors by a bootstrap procedure. I primarily employ a bootstrap procedure. I obtain each bootstrap sequence by a resampling procedure designed to preserve correlation properties of the data. For each bootstrap sequence, I compute average utility outcomes for optimal policies estimated from the original data. In this manner, I construct an empirical distribution of expected utilities. This becomes an issue for determining standard errors and for hypothesis testing. Wolf (2007), suggests a similar block bootstrap method for testing differences in Sharpe ratios across return series and compares the results with those obtained by asymptotic methods of Jobson and Korkie (1981). Wolf (2007) demonstrates that asymptotic methods are often biased in finite samples and that the bootstrap method performs well in Monte Carlo tests on returns generated by processes with non-IID errors and correlations.¹⁶

In the empirical analysis, I consider the asset allocation problem of a domestic U.S. investor. I consider two empirical questions. First, I examine the diversification benefit from incorporating portfolios that serve as empirical proxies for pricing factors. Multifactor pricing has implications for the role of asset allocation as a mechanism for determining exposure to systematic risk. In a nonparametric factorization of the covariance matrix of equity returns, three or more principle components are required to capture cross sectional variation in returns (Connor and Korajczyk (1993)). The capital asset pricing model suggests the market portfolio as principal factor. However, this portfolio only explains 70-80% of non-diversifiable risk in equity returns. Fama and French

¹⁵Jobson and Korkie (1981), updated by Memmel (2003), present asymptotic results for estimating the distribution of a difference in expected utility under the assumption of mean-variance utility and independent and identically distributed returns.

¹⁶The Jobson and Korkie (1981) test is applied to comparisons of portfolio strategies in out-of-sample tests by DeMiguel et al. (2007) and DeMiguel and Nogales (2007). In contrast, I perform in-sample testing of differences in expected utility for different strategies.

(1996) and Carhart (1997) find a number of hedge portfolios that serve as proxies for priced risk that is not captured by the market portfolio.

I consider three long-minus-short equity portfolios constructed to capture value, size, and momentum premiums, respectively.¹⁷ I find that including a factor portfolio constructed to reflect the return premium between value and growth stocks significantly improves expected utility. This finding corroborates results of Pástor and Stambaugh (2000) and Avramov (2004). They find that an unconstrained investor who is given the opportunity to diversify across value-premium and size-premium factors will take large positions in the value premium. However, they also report large allocations to a size premium portfolio. Furthermore Aït-Sahalia and Brandt (2001) report nontrivial allocations to a long term bond index. I find that neither a factor portfolio based on the size premium nor a long-term bond index provide significant diversification benefit to an equities and cash portfolio. The momentum portfolio provides significant diversification benefit. Finally, I test the null hypothesis that expected utility is not improved by the addition of each factor portfolio in the investor's available asset set. I find that the hypothesis cannot be rejected at the 99% confidence level unless the value-premium or momentum-premium portfolios are included in the investor's asset set.

Second, I study the impact of conditioning on expected utility. Predictability implies state dependence of an investor's optimal portfolio. I estimate conditional and unconditional expected utility from following optimal policies conditioned on a set of eight predictors. I compare the results to conditional and unconditional expected utilities estimated for the unconditionally optimal policy. I find that conditioning on individual predictors leads to a significant improvement in unconditional expected utility. However, utility gains that are conditional on the predictor value are statistically insignificant. My unconditional results are in line with a large literature on the significance of predictability on return regressions.¹⁸ In addition, the unconditional results corroborate results of asymptotic tests by Brandt (1999) for a subset of the predictors studied in this paper.

¹⁷These three portfolios are widely cited factor proxies proposed by Fama and French (1993) (value and size premiums) and Carhart (1997) (momentum premium). As of October 2007, the former had over 2400 citations on Google scholar, while the latter had over 1400.

¹⁸See, for example, Campbell and Yogo (2006).

My results show that optimal portfolios vary substantially with predictor value. However, as noted previously, I find that conditioning on individual predictors does not yield a significant improvement in conditional expected utility. The implication is that conditioning on individual predictors is not economically significant to a myopic investor. This result contrasts with previous literature in which economic significance is linked to conditional variability of portfolio weights. However, following the optimal conditional policy does not increase conditional expected utility significantly over that attained following the optimal unconditional strategy.

Of the eight predictors I consider, seven are taken from the literature.¹⁹The exception is the gold industry trend which measures recent returns on the gold industry. Despite being known for its low correlation with other industries, gold industry returns have not received much attention in previous literature on predictability. The lack of interest might be explained by the lack of significance of this variable in a linear regression versus next period returns. Linear regression is the standard approach to analyzing potential predictors.²⁰ However, gold industry trend proves to be one of the two strongest predictors in a nonparametric estimation of portfolio weights and expected utility. Examination of the portfolio policy as a function of gold industry trend reveals a strong nonlinear relationship that smooths out to a constant under a linear assumption. The significance of gold industry trend is corroborated by findings of Makarov and Papanikolaou (2008). They find evidence that a latent factor that weights heavily on base metal industries helps explain market equity returns.

The result demonstrates that not accounting for potential nonlinearity in the relationship between predictors and expected returns can have important implications for model assessment. This point is also emphasized by Ferson and Siegel (2000) in the context of testing asset pricing models. Rejection of the hypothesis that predictability improves expected utility based on an in-sample study requires a framework in which conditioning information is used with maximum efficiency. The nonparametric model places minimal restrictions on the portfolio policy and is therefore likely to result in a higher expected utility improvement over the unconditional policy when predictability is used

¹⁹See Ang and Bekaert (2007) and references therein.

²⁰Predictors should also have an economic rationale to mitigate data snooping.

in-sample. Thus rejection of the hypothesis that the unconditional policy yields the same expected utility is more challenging, and, therefore, p-values that suggest rejection are more convincing.

My approach to the question of investment significance differs from much of the recent literature, where two approaches are prevalent. First, many studies abstract from estimation uncertainty by calculating optimal allocations from parameters of a model of the expected distribution of returns that are fixed at sample values (i.e., Brennan et al. (1997), Campbell and Viceira (1999), Balduzzi and Lynch (1999), Campbell et al. (2003), and Jurek and Viceira (2005)). These contributions yield insight into potential implications of empirical features for an asset allocator, but do not establish whether the computed impacts are statistically significant.

The second approach seen in the recent literature assumes a Bayesian model of the investor in which estimation uncertainty is incorporated into the predictive distribution of returns. Kandel and Stambaugh (1996) pioneer this approach in a study of the investment impact of predictability evidence.²¹ They conclude that the impact of predictability on investment decisions is significant even though the empirical evidence for predictability is weak. Their conclusions are based on the substantial variability of portfolio weights with predictor value, but they do not provide evidence on the significance of the expected utility gains from using predictor variables.

My results have different implications for the importance of predictor variables when compared to those of Kandel and Stambaugh (1996) and subsequent Bayesian studies.²² The discrepancy can be attributed to differences in the criteria used to assess significance. I treat the portfolio problem as a statistical estimation problem and assess significance in terms of welfare gains using a hypothesis testing approach. I examine whether expected utility gains of adopting optimal strategies are robust to estimation error. In contrast, Kandel and Stambaugh (1996) ask whether investment policies are influenced by predictability but only obtain point values for expected utility gains without addressing the

²¹Kandel and Stambaugh (1996) is a concept study. Barberis (2000) and Wachter and Warusawitharana (2005) apply the ideas to empirical data. Barberis (2000) examine the importance of predictability in the face of estimation risk. Wachter and Warusawitharana (2005) study allocations of an investor with priors biased against predictability.

²²See, for example, Barberis (2000) and Wachter and Warusawitharana (2005).

statistical significance of these gains. However, expected utility estimated in-sample must increase when the investor is given additional information. Thus, whether or not an empiricist uses a Bayesian model of the investor, it is important to address the robustness of estimated expected utility gains before ascribing investment significance.

Finally, I extend the predictability analysis to allow for simultaneous conditioning on multiple predictors. I remain in the nonparametric environment, but follow Aït-Sahalia and Brandt (2001) and reduce the dimensionality of the nonparametric problem by constructing indices from linear combinations of predictors. This semiparametric approach reduces the dimension of the nonparametric estimation problem to one, thereby mitigating problems with the curse of dimensionality that arise in multidimensional nonparametric estimation.

I find that predictability can be significant for an investor with a single-month horizon when multiple predictors are used. I estimate optimal index values for all combinations of four predictors that can be formed from a set of eight predictor variables. This model selection exercise suggests that the best and most robust predictor combination is any combination of predictors that includes the term spread and the gold industry trend. The important predictors differ from those ordinarily employed in studies that focus on the dynamics of investment with time varying investment opportunities. For example, the dividend yield is not a statistically robust component of the optimal predictive index, but is often used in calibrated dynamic models. Of the two variables that play an important role in predicting optimal allocations, the term spread has the longest history in studies of predictability (see Campbell (1987)). Unlike in the single predictor case, the improvement in expected utility for the best multivariate index is marginally significant in some states of the world.

It is important to clarify that the tests in this paper are not primarily aimed at verifying the presence of predictability. The objective of my tests is to evaluate the relative significance of portfolio estimation risk versus the benefits of diversification or conditioning using a metric, expected utility, that reflects investor preferences. In other words, I assume that diversification or conditioning would add value if the true optimal portfolio was known. The testing approach in this paper would also be applicable to tests of predictability. However, the standard errors and p-value statistics would have to be adjusted to address the

data-snooping/joint-testing problem. This analysis is beyond the scope of the this thesis.²³

The estimated expected-utility gains permit weighing the benefit of incorporating factor portfolios into the asset allocation decision versus the benefit of conditioning on predictability. For an equities and cash investor, I find that point estimates of expected utility gains from adding the value premium proxy to the portfolio choice set are effectively equal to the gains from conditioning on the term spread or gold industry trend. The magnitude of the improvement in CER is approximately 50 basis points per month or six percent per year.

2.2 Empirical Framework

I employ an econometric framework based on the portfolio choice problem of a single period investor. The investor's problem is set up as a statistical decision problem with asset allocations as parameters and expected utility as the objective. The allocations are estimated by direct maximization of expected utility. The remainder of this section describes the investment framework and presents the empirical estimator.

2.2.1 Investor's Problem

Consider a single period investor who maximizes the expected value of utility $u(W_{t+1})$ over next period's wealth W_{t+1} . The investor has access to a set of N portfolios for investment. The investor's choice variable is an N -vector of portfolio weights α_t . Expected utility is conditional on a set of predictor variables Z_t . Hence, the investor solves

$$\max_{\alpha_t} E[u(W_{t+1})|Z_t], \quad (2.1)$$

where $u(\cdot)$ is the investor's expected utility over wealth.

The above formulation of the portfolio problem does not explicitly include a budget constraint. Instead of requiring that portfolio weights add to one, the

²³The predictability of stock returns and the precise nature of additional pricing factors remains a topic of current research(Ang and Bekaert (2007)). White (2000) and Dudoit and van der Laan (2007) describe frameworks for testing multiple hypotheses.

asset set is defined such that each asset is a hedging portfolio. Hedging portfolios consist of long positions in some assets and offsetting short positions in other assets such that the net investment is zero. A net long or short position in risky assets is made possible by the inclusion of at least one asset that represents a long position in a portfolio of risky assets and an offsetting short position in the risk-free asset. Given this construction of the asset set, and assuming no other constraints, the vector of portfolio weights α can take on any value in \mathbb{R}^N .

Let random vector R_t be the vector of gross returns on securities from t to $t + 1$. Next period wealth W_{t+1} depends on this period's wealth and portfolio weights along with the intervening period's vector of returns R_{t+1} ,

$$W_{t+1}(W_t, \alpha_t, R_{t+1}) = W_t [R^f + \alpha_t^\top R_{t+1}] \quad (2.2)$$

where R^f is the risk-free rate.

The predictor variables are realizations of a random vector Z_t from an M -dimensional predictor space Λ . The distribution of expected returns is dependent on Z_t . This along with the dependence of utility on returns explains the conditional form of the expectation in the investor's objective function.

The investor's optimal portfolio policy is the solution $\alpha(Z)$ to (2.1). A portfolio policy is a mapping from the predictor space to the space of allowed portfolio weights, $\alpha : \Lambda \rightarrow \mathbb{R}^N$. The first order conditions of the investor's problem are

$$E[u'(W_t(R_f + \alpha_t^\top R_{t+1})) R_{t+1} | Z_t] = 0. \quad (2.3)$$

Hence, the portfolio policy is given by

$$\alpha(Z) = \left\{ \alpha_t : E[u'(W_t(R_f + \alpha_t^\top R_{t+1})) R_{t+1} | Z_t = Z] = 0 \right\}. \quad (2.4)$$

The estimator developed in the following section is based on this equation.

2.2.2 Direct Estimator of Conditional Portfolio Policy

Portfolio policies can be estimated without explicitly modeling return dynamics²⁴. Brandt (1999) adapts the method of moments approach of Hansen and

²⁴The need to model the entire return distribution considerably complicates estimation of optimal portfolio policies by adding an additional set of assumptions beyond those already

Singleton (1982) to this problem. The empirical approach entails replacing the expectation in the first order condition or Euler equation of the investor's problem (2.3) with a consistent estimator. The expectation $E[\cdot|Z_t]$ on the right-hand side of the Euler equation is replaced with a nonparametric estimate $\hat{E}[\cdot|Z_t]$ that converges to the true expectation as $T \rightarrow \infty$. Upon replacing the expectation in (2.3) with its empirical counterpart, an empirical moment condition is obtained and α can be estimated by method of moments.

Operationally, the expectation on the left-hand side of the first-order condition (2.3) is replaced with a historic average

$$\hat{E}[u'(W_s \alpha^\top R_{s+1}) R_{s+1} | Z_t = z] = \frac{1}{T_z} \sum_{\{s: Z_s = Z_t, s \leq t\}} u'(W_s \alpha^\top R_{s+1}) R_{s+1} = 0, \quad (2.5)$$

where T_z is the number of observations at which $Z_t = z$. For Z_t on a continuous domain, the above estimator is infeasible for finite samples. Brandt (1999) shows that a standard nonparametric estimator converges for moment functions that obey reasonable properties. The nonparametric estimator of α in some state z is obtained by weighting each observation according to the similarity of its state with z . Define the weighting function $\omega(\cdot)$ and bandwidth h_T . The nonparametric estimator of the expectation is

$$\hat{E}[u'(W_s \alpha^\top R_{s+1}) R_{s+1} | Z_t = z] = \frac{1}{\tau(h_T, z)} \sum_{s=1}^T \omega\left(\frac{z - z_t}{h_T}\right) u'(W_s \alpha^\top R_{s+1}) R_{s+1}, \quad (2.6)$$

where $\tau(h_T, z)$ is equal to the sum of the weights applied to each observation. Substituting the empirical expectation into the Euler equation yields a set of empirical moment conditions that are satisfied by the optimal portfolio policy.

required to model investor preferences. For most preference models, the precise shape of the portfolio policy is a complicated function of multiple moments of the conditional distribution of returns. This point is discussed extensively in Aït-Sahalia and Brandt (2001). They evaluate the portfolio policy for a variety of utility specifications, including expected utility, ambiguity and loss aversion, and prospect theory preferences. For example, consider the case of an investor with mean-variance utility. In this case, there is an analytic solution for the optimal conditional policy. The optimal policy depends on the ratio of the first two conditional moments of the return distribution. Thus, even in this simple case, in order to study the effects of predictability on portfolio choice by traditional means, one would have to effectively model both the first and second moments as functions of the predictor variables. The problem is exacerbated if allocations depend on higher moments (Harvey et al. (2003), Kacperczyk (2003)) or if allocating across a large number of assets (Brandt et al. (2005)).

The problem of solving the Euler equations can be formulated as a method of moments minimization whereby $\alpha(z)$ is chosen to equate

$$\hat{E}[u'(W_t \alpha_t^\top R_{t+1}) R_{t+1} | Z_t = z]' \hat{E}[u'(W_t \alpha_t^\top R_{t+1}) R_{t+1} | Z_t = z] \quad (2.7)$$

to zero.

2.2.3 Measure of Portfolio Performance

Kandel and Stambaugh (1996) suggest certainty equivalent returns as a useful metric for comparing portfolio performance.²⁵ The certainty equivalent CE is the certain wealth outcome that would be accepted by the investor as an even trade for the risky bet, i.e., $U(CE) = E[u(W_{t+1})]$. When comparing two certainty equivalents, the fractional change is a useful metric. This is defined as $\delta CE = CE_A / CE_B - 1$; CE_A is the certainty equivalent of interest and CE_B is certainty equivalent value of a reference portfolio. Following convention, I refer to δCE as the certainty equivalent return (CER) when the reference portfolio is a 100% investment in the riskless asset.

For unconditional portfolio allocation, the certainty equivalent can be estimated by replacing the expectation with a sample average:

$$\widehat{CE} = U^{-1} \left(\frac{1}{T} \sum_{t=1}^T u(\alpha^\top R_t) \right) \quad (2.8)$$

When using conditioning information, the estimated certainty equivalent is state dependent. Adopting the kernel averaging approach described in the previous section, the conditional certainty equivalent for an investor with unit wealth is

$$\widehat{CE}(z) = U^{-1} \left(\frac{1}{\tau(h_T, z)} \sum_{s=1}^T \omega \left(\frac{z - z_t}{h_T} \right) U(\alpha(Z_t)^\top R_{s+1}) \right). \quad (2.9)$$

I calculate p-values for hypothesis tests and compute standard errors using a bootstrap analysis. I recalculate equivalent statistics for 1000 bootstrap samples. I use the stationary bootstrap procedure of Politis and Romano (1994)

²⁵The use of certainty equivalent to evaluate portfolio performance is a standard approach (Aït-Sahalia and Brandt (2001), Avramov (2004), Brennan et al. (1997)).

to preserve autocorrelation properties of the data in the bootstrap samples.

Many of the hypothesis tests involve comparing expected utility gains achieved by different strategies. I first compute the optimal policies for the two strategies. I then calculate expected utility estimates for each bootstrap draw. If the null hypothesis is that strategy A achieves as high an expected utility as strategy B, then I compute a p-value based on the fraction of draws for which average utility achieved under policy A equals or exceeds that achieved under policy B.

An investor’s ability to exploit the increase in CER from diversification will depend on the amount of data available to the investor and potential CER gain given full knowledge of the generating distribution of returns. To gauge the relative importance of these two factors, I conduct a simulation experiment. I consider the investment benefit of adding an additional risky asset to a two risky asset portfolio. The correlation between the returns of the three assets are based on correlations between the equity index, bond index and HML asset described in section 2.2.4. I vary the Sharpe ratio of the third added asset so that the full-information increase in CER varies between 1 basis point and 100 basis points per period. I simulate 400 return series of length T equal to 120, 240 and 600. For each return series, I estimate an optimal portfolio for the two risky asset and three risky asset cases. I also compute a bootstrap p-value against the hypothesis that the two asset portfolio will achieve a higher CER out of sample.

The results demonstrate that an investor seeking to reject the dominance of the estimated two-asset portfolio would need to be adding an asset with the potential to add greater than 1% return per period given a data set of 120 periods. For a data set of length 600, the null that the estimated portfolio achieves a higher out-of-sample CER can be rejected in the majority of cases as long as the CER is greater than 25 basis points. The latter result is relevant to the unconditional results because the data set used in this study includes over 600 monthly return periods. For results that are conditioned on predictor variables, the effective number of samples is reduced by the kernel averaging method. As a result, larger conditional improvements in CER are required to achieve diversification benefits given the available data for portfolio estimation.²⁶

²⁶Specifically, the Monte-Carlo Average p-values are, given a potential diversification benefit of 1 basis point, 0.30, 0.28, and 0.20 for data series lengths $T = 120, 240$, and 600 respectively. For potential diversification benefit of 25 basis points, the corresponding Monte-Carlo average

In addition to bootstrap errors, I also compute asymptotic standard errors where possible. The asymptotic properties of the nonparametric conditional estimator were given by Brandt (1999). The asymptotic distribution of estimates of portfolio weights are directly analogous to standard convergence results for the method of moments. For conditional and unconditional certainty equivalent returns estimated via (2.8) and (2.9), I compute asymptotic standard errors via the delta method.

2.2.4 Data and Utility Assumptions

I obtain monthly data for asset returns and predictors for the time period January 1952 to December 2006.²⁷ I consider an investor who allocates wealth between an equity index, a bond index, and cash. The investor also has access to hedge portfolios that have been recognized in empirical studies as good proxies for priced risk in the equity market that is not captured by the market portfolio (Fama and French (1996), Carhart (1997)). These consist of a portfolio long high book-to-market (value) stocks and short low book-to-market (growth) stocks, a portfolio long small market capitalization stocks and short large capitalization stocks, and a portfolio positive high momentum stocks and short negative momentum stocks. The market index is the value-weighted index of NYSE, AMEX, and Nasdaq stocks from CRSP. The bond portfolio is an index of long-term government bonds. Returns on this portfolio are obtained from CRSP. Factor portfolios based on value minus growth, small minus big and upward momentum minus downward momentum are the HML (High-book-to-market Minus Low-book-to-market), SMB (Small-stock Minus Big-stock), and UMD (Upward-momentum Minus Downward-momentum) portfolios from Ken French.²⁸ The return of the riskless asset is the yield on treasuries matching

p-values are 0.21, 0.14, and 0.05. For potential diversification benefit of 50 basis points, corresponding Monte-Carlo average p-values are 0.14, 0.06, and 0.009. Finally, for potential diversification benefit of 100 basis points, the corresponding Monte-Carlo average p-values are 0.07, 0.02, and 0.0003.

²⁷This start date is commonly used in calibration studies of predictability and portfolio choice (e.g. Brandt (1999), Campbell and Viceira (1999), Brennan et al. (1997), Ait-Sahalia and Brandt (2001)), and coincides with availability of data for a large set of predictor variables. The time period is subsequent to the 1951 Treasury-Fed accord that allowed independent conduct of monetary policy.

²⁸http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

the investment horizon. For the monthly return horizon, this is the 30-day Treasury-bill yield. Summary statistics for the asset returns are provided in Table 2.1.

Investment opportunities are functions of eight predictor variables that have been cited as able to forecast market returns and volatility. The predictors include the NYSE dividend yield, default premium, term premium, trend in the S&P 500 composite, trend in the 30-day Treasury-bill yield, gold industry returns, value spread, and inflation²⁹. The dividend yield on NYSE stocks is imputed from cum- and ex-dividend returns from CRSP. All interest rate data is obtained from Global Insight with the exception of the T-bill yield. The latter is obtained from CRSP. The default premium is the difference in yield between an index of BAA and AAA rated corporate bonds. The term spread is the difference between the yield on long-term government bonds and three-month Treasury bills. As suggested by Campbell (1991) and Hodrick (1992), I stochastically detrend the 30-day Treasury-bill yield used for prediction. I take the current yield minus a six-month moving average centered six months previously. This eliminates the long-term secular trend in this variable, rising prior to and descending since 1980. The 30-day Treasury-bill yield serves a dual role as both predictor and yield on the risk-free asset for investments with one-month horizon. The trend in the S&P 500 composite is an average over trailing returns of the previous twelve months. Realized volatility is calculated using daily returns over the previous month. Inflation is the monthly change in consumer price index (CPI) for the previous month. Summary statistics for the predictors are provided in Table 2.1. Correlations between returns and lagged predictors are listed in Table 2.2. Correlations between realized volatilities and lagged realized volatilities and predictor variables are listed in Table 2.3.

All the above predictors have received attention in the literature on predictability with the exception of gold industry trend. This variable is constructed from returns on the gold industry portfolio within the 48 industry breakdown of Ken French. This breakdown is based on industry definitions

²⁹Other predictors that have received attention include the ratio of consumption to income (Lettau and Ludvigson (2001)) and measures of market sentiment (Baker and Wurgler (2006)). I restrict attention to predictors derived from market observables (with the exception of inflation). This mitigates possible interpretation issues associated with the timing and availability of non-market based variables.

that begin in June 1964. To extend the series backwards, I construct a gold industry based on membership in the gold industry in June 1964. While there is danger of introducing survivorship bias into the early part of the series, this is overshadowed by the likely loss of information from discarding the returns entirely. For use as predictor, gold industry returns are smoothed by taking a moving average over the trailing twelve months.

For the empirical analysis, I assume an investor with power utility, $U(W) = W^{1-\gamma}/(1-\gamma)$. I use a power utility model because of its appealing properties with respect to the effect of wealth on risk aversion. Power utility engenders decreasing absolute risk aversion. This is more reasonable than the increasing absolute risk aversion of mean-variance utility³⁰. Mean-variance utility, which is often used in practice, can be a good approximation to other utility functions (including power utility). Mean variance is obtained by truncating a power series expansion of the utility at the second power so the quality of the approximation depends on the size of return magnitudes. I consider a monthly investment horizon so some returns in the sample are quite large. In addition, optimal portfolios may be leveraged thereby magnifying portfolio returns.

I vary the risk aversion coefficient γ between one (log utility) and ten. Most of the experiments are performed with risk aversion set to five.

2.3 Unconditional Asset Allocation

I begin the empirical analysis with an examination of the investment gains derived from incorporating portfolios that serve as empirical proxies for pricing factors into the asset allocation portfolio. The section presents an in-sample analysis of the unconditional portfolio choice and certainty equivalent gains for various combinations of assets. To obtain estimates of unconditional policies, I compute solutions to the unconditional version of (2.7).

Table 2.4 lists estimates of unconditional portfolio choices of an investor with power utility. The first row lists portfolio allocations for investors whose investment universe is limited to the equity index. The results are broadly

³⁰Many of the experiments of the next section were repeated with mean-variance utility. Essential features of the results (not reported) are similar to the power utility case. Markowitz (1952) and Arrow (1971) provide theoretical arguments for utility with decreasing absolute risk aversion.

consistent with those of Brandt (1999).³¹ The optimal weights are large. For an investor with risk aversion of five, optimal investment in the index is 58%, rising to 142% at risk aversion of two. These results are consistent with the well-known tendency for estimated allocations to indicate large positions in equities for reasonable levels of risk aversion (Mehra and Prescott (1985)).

The second row of Table 2.4 lists portfolio allocation for an investor who invests in HML only. While this is not a realistic portfolio choice for most investors as it implies short positions in the growth portfolio, the results illustrate the investment appeal of HML's historical return properties. The allocation is almost double that selected by the equities investor. This can be attributed to HML's higher Sharpe ratio (Table 2.1). The mean-variance solution illustrates the source of the difference between the two allocations. For a mean-variance investor, the optimal allocation to a risky asset is proportional to the ratio of mean return to its variance. Using this fact, we expect the ratio of the portfolio weights for the index-only portfolio and the HML portfolio to be $\alpha_{Index}/\alpha_{HML} \approx \mu_{Index}\sigma_{HML}^2/(\sigma_{Index}^2\mu_{HML})$. Based on mean and variances listed in Table 2.1, $\alpha_{Index}/\alpha_{HML} = 0.58$. The ratios between the estimated weights in rows 1 and 2 of Table 2.4 are all within rounding error of 0.56.

The remaining rows of Table 2.4 present allocations across different portfolios. The next two rows show allocation estimates for an equity and bond index portfolio. The allocation to bonds is less than the standard error of the allocation estimate. The boundaries of a 95% confidence interval around the optimal bond allocation are -0.41 and 1.03. In addition, equity allocations in equity and bond portfolios are almost unchanged from their optimum without the bond asset. An equities-only investor achieves little diversification benefit by adding bonds to the asset set. Combining the equity index portfolio with HML leads to higher allocations to the equity index and HML relative optimal holdings in their single asset portfolios. When compared to the allocation for an equities-only investor, the allocation to equities increases by approximately 40% when HML is added to the asset set. These results reflect the large positive premium accorded to value stocks over the fifty-year period encompassed by this data set along with the significant negative correlation between the index and

³¹Brandt (1999) considers a simultaneous portfolio allocation and consumption decision. However, the consumption decision has little effect on portfolio allocation.

HML. Combining the bond asset with the equity index and HML combination has little effect on the portfolio weights of the equity portfolios. In all cases, the allocation to the bond portfolio is less than one standard error from zero and the allocations to the equity portfolios effectively are unchanged.

The large standard errors reported in Table 2.4 reflect considerable uncertainty in return estimates obtained from data. Brandt (1999) suggests using the mean-variance solution for intuition on the source of uncertainty. If we assume known variance then the standard error of the return estimate for the mean-variance solution, and consequently, of the portfolio weights, will be inversely proportional to $(\gamma * \sqrt{T} * \sigma)$, where σ is the standard deviation of returns. The errors calculated using this approximation and sample values in Table 2.1 are of similar size to those given in Table 2.4.

Table 2.5 lists estimated CERs for an investor who follows the optimal unconditional policies for various portfolios. I compute standard errors for CERs using the stationary bootstrap. For each bootstrapped realization of the data, I compute a CER estimate under the assumption that the investor holds the estimated optimal portfolio.³² The certainty equivalent return of an index investor is less than two standard errors above the certainty equivalent return of following a risk-free strategy. A 95% confidence interval around the CER estimate for an equities-only portfolio runs from -5 to positive 37 basis points. The results suggest that an equity index investor may not be able to reject the possibility that the CER of investing in the equity index is no greater than investing solely in the risk-free asset. This level of uncertainty in the CERs might give pause to advocates of straight indexing as an optimal allocation policy.

I also compute the asymptotic standard errors for the CERs reported in Table 2.5. The CER is a function of the realized utilities $u_t = w' * r_t$. The delta method approximation linearizes the variations to approximate the standard error by $(CER'(\bar{u}))^2 var(u)$. I use a Newey-West procedure to account for autocorrelation in the realized utilities when computing the variance. The results are very similar to those obtained by the bootstrap analysis. For example, the standard error of the CER for the equities, bonds, and HML computed by the bootstrap is 0.24 basis points compared per month whereas the asymptotic

³²Alternatively, asymptotic standard errors can be computed by the delta method applied to (2.8). The results are similar and not reported.

standard error given a Newey-West adjustment with 20 lags is 0.25 basis points. Given the similarity of the results obtained by the two methods, I only report bootstrap standard errors in Table 2.5 and subsequent tables.

In contrast to the equity index-only case, the CER of portfolios that include both the equity index and HML portfolios are significantly larger than twice the standard errors (Table 2.5). The portfolio of an HML-only investor has a higher Sharpe ratio in-sample than the equity index, but the CER of an investor who allocates exclusively to this portfolio is less than two standard deviations from the risk-free CER of zero. In contrast, the investor whose portfolio includes equities and HML achieves CER greater than three times the standard error. A 95% confidence interval for CER for the equities and HML investor runs from 0.2 to 1.1 basis points.

The results in Table 2.5 have important implications for the bond asset. Inclusion of the bond portfolio in the asset set has an insignificant impact on CERs of both the equities-only investor and the equities and HML investor. In both cases, the estimated CERs remain unchanged. In addition, there is no shift in the bootstrap confidence intervals. Referring to Table 2.4, the estimated allocations to the bond asset are quite large, but remain insignificant. Consider the case of risk aversion γ equal to five. The estimated bond weight is 30% and 17% in the equities and bonds, and equities, bonds and HML portfolios respectively. The CER results demonstrate that there is no in-sample evidence that an investor achieves any diversification by including the long-term bond asset in their asset allocation portfolio.

Table 2.6 lists bootstrap p-values based on the null hypothesis that adding a particular asset does not improve expected utility. Each p-value is based on 1000 resamples of the joint return series. The left column lists the asset set that is assumed optimal under the null while the middle column lists the asset under consideration for addition to the portfolio. The first and third rows examine the importance of utility gains from investing in the equity index and HML respectively versus the null of holding only the risk-free asset. The p-values are approximately 0.05 in each case, indicating that investing in either portfolio is an improvement over simply holding the risk-free asset. As suspected given the insignificant holdings in the portfolios listed in Table 2.4, the null hypothesis that investing fully in the risk-free asset is preferred to investing in a mix of

bond index and risk-free asset cannot be rejected.

The sixth row lists the p-value of adding the bond index to a portfolio that includes the equity index and HML. The seventh row states the p-value of adding HML to the traditional asset allocation portfolio. The null that the bond index is not additive to an investor's CER cannot be rejected. In contrast, the corresponding null in the HML case is strongly rejected in favor of adding HML to the asset allocation portfolio.

Other portfolios have been suggested as proxies for risk captured by stock returns. Table 2.5 also lists in-sample CERs for an investor who adds two other portfolios to the asset allocation decision. The first is SMB: a portfolio long large stocks and short small stocks. SMB is a member of the oft-cited three factor set of Fama and French (1993). If the investor believes that SMB serves as a proxy for a priced risk factor, then the portfolio choice results (not shown) indicate that the investor will take a marginally significant long position in this asset. However, the increase in CER over that achieved by the optimal equities, bonds, and HML portfolio is insignificant. The p-value for the bootstrap test against the hypothesis that SMB has no diversification benefit is 0.24 (eighth line of Table 2.6), providing further evidence against SMB as a useful diversification factor. The increase in CER is much larger when the momentum factor UMD is added to the asset set. The increase in CER for the equities and HML investor who includes UMD in their asset set is just under twice the standard error. The p-value on the last line of Panel A of Table 2.6 corroborates this observation.

The portfolios that include HML can only be implemented if short-sales are permitted. The short sales occur when short positions on the growth side of the HML portfolio are larger than long positions in the same stocks held in the equity index portfolio. Table 2.4 also list portfolio weights for an investor who cannot hold assets short. Weights in equities and bonds must be greater than zero. To ensure that short positions in the long-short factors are covered, positions in these portfolio are limited to one third of the equity holding.

The asset weights listed in Table 2.4 suggest an increased role for bonds in the equities, bonds, and HML portfolio. With the short sale constraint restricting the investor's ability to leverage the HML portfolio, the optimal bond weight is 42% for an investor with risk aversion of five. This compares to 17% in the unconstrained case. However, the certainty equivalent gain achieved by moving

from the optimal equities and HML portfolio to the optimal portfolio with bonds included is inconsequential. The p-value listed in Table 2.6 for this case (line 5 of Panel B) is close to 0.5. The hypothesis that bonds do not increase CER cannot be rejected even when short sales are not permitted.

2.4 Conditional Asset Allocation

This section examines predictability evidence in an asset allocation framework. The standard approach to predictability evidence is to estimate a statistical model in a linear regression or vector autoregression framework. Returns are regressed on lagged predictor variables. This section considers an alternative approach. I examine the implications of individual predictors on investor welfare. I consider the portfolio problem of a power utility investor whose portfolio choice is between the asset of interest and a risk-free asset.

The first subsection is a brief examination of predictability evidence based on a simple regression analysis. The second subsection looks at predictability evidence through the lens of the optimal asset allocation problem. I consider the allocation problem of an investor with access to a single, risky asset. I evaluate the impact of predictors on portfolio choice by examining the significance of differences between conditional and unconditional allocations estimated by the method of section 2.2.2.

2.4.1 Predictability of Mean and Variance

Panel A of Table 2.8 lists results of univariate least squares regressions of individual asset returns on individual predictor variables, i.e., $r_t = a + b'Z_t$. Because the choice of predictor variables is primarily based on previous literature on market predictability, it is no surprise that levels of significance are reasonably high for regressions involving the equity index return. In each case, a number of predictors yield t-statistics greater than two³³. The least predictable asset is HML; at short horizons, only the Treasury-bill rate yields a t-statistic greater

³³Care must be taken in interpreting t-statistics obtained from regressing a time series on a second, highly persistent time series (Stambaugh (1999)). See Campbell and Yogo (2006) and references therein for a discussion of inference issues associated with regression evidence for stock return predictability.

than one. The value spread does not show evidence of much correlation with future HML returns.

Panel B of Table 2.8 presents results of multivariate regressions. For each asset, next month's return is regressed on the full vector of predictor variables. The results are broadly consistent with those of the previous table, but with dampened significance for individual coefficients due to the simultaneous nature of the regression. For the index portfolio, no one predictor appears particularly significant. By contrast, the regression results indicate term spread and bonds have significant impact for predicting bond returns. Additionally, The Treasury-bill rate stands out as a significant predictor for the HML portfolio.

The R^2 values for the multivariate regressions are highest for the equities and bond indices. Once again, this result is unsurprising given that the predictor variables used are from the long literature on stock market predictability. Thus, while the index and bond portfolios may be more predictable than HML, a data snooping bias in the choice of regressors must also be acknowledged.

Panel A of Table 2.9 displays results of univariate regressions of realized volatility of equity index and HML portfolios respectively on each lagged predictor variable. A number of variables appear to predict realized volatility at the one-month horizon. For example, the t-statistic is greater than two for realized volatility of the equity index regressed on dividend yield, default premium, S&P 500 trend, and value spread. The same holds for HML regressed on dividend yield, S&P 500 trend, and gold industry trend.

Panel B of Table 2.9 shows results of regressing realized volatilities of the equity portfolios on the full set of predictors. Significant regression coefficients are observed for both the equity index and HML. The high R^2 values reflect the high short-term persistence of volatility.

2.4.2 Predictability in an Asset Allocation Framework

In this section, I analyze the impact of predictability on portfolio choice through the lens of the investor's portfolio problem. The analysis provides direct insight into the significance of predictability to portfolio choice. I employ the nonparametric estimator of portfolio policy (Section 2.6) to examine the evidence for predictability of different assets. Consider the portfolio policy of an investor who is able to invest in only one risky asset. The investor's problem is to

choose optimal allocations to the risky asset as a function of predictor value. In this context, the significance of a predictor for portfolio choice is a function of the ratio of the maximum deviation of the conditional allocation from the unconditional allocation and the standard error of the allocation estimates.

I apply this approach to analysis of predictability of the equity index and HML portfolios. Table 2.10 lists conditional allocations for an investor choosing between one risky asset and a risk-free asset. Results in this table can be used to assess the relevance of any one predictor variable for portfolio allocation. Optimal allocations are listed for values at the 20th, 50th and 80th percentiles of the empirical distribution of the predictor variable. Based on differences between allocations at the 20th and 80th percentiles, the greatest in-sample predictability is obtained for the equity index conditioned on the term spread, Treasury-bill rate, inflation rate, and gold industry trend. Interestingly, the gold industry trend does not show up as significant in either univariate or multivariate linear regression results given in Table 2.8. In agreement with regression results, predictability of HML does not appear significant by this portfolio choice metric. The largest difference between allocations at the 80th and 20th percentiles is smaller than those observed for equities.

Comparison of policies at three points of the predictor distribution may miss large variations in the optimal allocations if portfolio policies are nonlinear. Figures 2.1-3 provide a graphical comparison of optimal conditional and unconditional single-asset policies as a function of predictor values. Each figure plots optimal portfolio weight versus standardized predictor value³⁴. The portfolios are estimated for a non-uniform range of percentiles of the empirical distribution. The calculation points are denoted by dots in the figures. Conditional policies are reported for single-asset decisions when the risky asset is the equity index (left panels) or HML (right panels). The equity index policies that condition on the dividend yield, default premium, and term spread are consistent with those documented by Brandt (1999) and, with the T-bill rate added, Paye (2004). The second column presents results for the HML portfolio. The conditional allocation to the equity index as a function of dividend yield (top left panel of Figure 2.1) is broadly consistent with results of Barberis (2000).³⁵

³⁴Predictor values are standardized by subtracting their means and normalizing by their standard deviations.

³⁵See left side of his Figure 5, one period horizon case.

The shaded confidence regions are interpolated between two standard deviation error bars estimated by bootstrap. The vertical bar in the center of each plot depicts bootstrap two standard deviation confidence interval of the unconditional portfolio. Overall, the width of the confidence regions confirms the insignificance of predictability evidence when conditioned on predictor value. The conditional policy rarely moves beyond the limits of the vertical bar on any panel of the three figures. The term spread and gold industry trend are the only two predictors that appear qualitatively significant for equity investment. For each of these predictors, there is a range of predictor values over which the optimal policy is greater than two standard errors from the unconditional policy. Of interest is the strength of predictability of the gold industry trend relative to the regression results presented in Table 2.7. Conditional policies show significant variability even though the coefficient on a linear regression is below standard significance levels. Estimated portfolio policies for gold industry trend vary nonlinearly with predictor value such that a linear fit as is used in the regression would damp out much of the conditional variability.

Evidence for predictability of HML is weaker. This is not unexpected given that the majority of predictors that I consider are sourced from the literature on predictability of the equity index. Gold industry trend, inflation and realized volatility appear marginally significant as predictors for an HML investor. For the other predictors, the conditional variation in the portfolio policies is swamped by the standard errors of the estimates. The data suggest high holdings of HML when the trend in gold industry returns is more than one standard deviation above or below its average trend, and lower holding of HML when trend is near its average. The results for realized volatility suggest decreasing emphasis on value stocks in high volatility scenarios. In contrast, conditional significance of inflation occurs at inflation values close to two standard deviations above its mean. A glance at the histogram for inflation in Figure 2.11 shows low data density for high inflation environments. Hence, the significance in this region is likely to be spurious.

2.5 Conditional Asset Allocation with Multiple Assets

In this section, I study the consequences of predictability for the conventional asset allocation problem – that of choosing between an equity index, long-term bond index and cash. In addition, I examine the importance of predictability when the asset set is expanded to include HML.³⁶ Section 2.3 highlights the significant impact of incorporating the HML portfolio into the asset set used for asset allocation. The results in this section permit a comparison of the utility gain achieved by adding HML as a risk factor proxy to the gain from conditioning on individual predictors.

Figures 2.4-11 depict the portfolio policies of a tactical allocation investor who conditions on a single predictor. There are five panels in each figure laid out in three rows and two columns. Each column depicts results for a different set of assets. First columns show optimal conditional portfolio policies for an investor with access to equity and bond indices. This corresponds to the traditional problem of tactical asset allocation. Second columns show optimal conditional policies for an investor who achieves risk factor diversification by investing in the HML portfolio. All results are obtained assuming risk aversion equal to five. The predictor values are standardized. The shaded regions are two standard deviation error estimates obtained by interpolating between two standard deviation error bars for each point estimate.³⁷ Unconditional portfolio weights are indicated by horizontal lines, with a single cross bar on each plot showing the two standard error range of the unconditional weight.

For the asset allocation problem of allocating across equities, bonds, and cash, conditional variability of portfolio weights is marginally significant for some of the predictors. Based on Figures 2.4-11, the largest deviations of optimal conditional weights from their unconditional level occurs when conditioning on the term spread, S& P 500 trend, T-bill rate, gold industry trend and inflation.

Addition of the HML portfolio to the set of assets included in the portfolio

³⁶I focus on HML rather than UMD. While the UMD portfolio also provides significant utility gains, the practical implementation of UMD strategies requires high asset turnover that leads to high transaction costs.

³⁷Error bars are estimated by stationary bootstrap.

yields two observations. First, as was true for single asset policies for HML (right panels of Figures 2.1-3) the variability in conditional allocations to HML does not appear significant. Second, the presence of the HML asset increases allocations to the equity index. By comparing the top two rows of Figures 2.4 to 2.11, I observe that the effect of adding the HML to the asset set of a index and bond investor is similar to that observed in the unconditional case. The addition of a second, predictable asset increases the total investment in risky assets, but does not influence the conditional variation in asset allocations. Allocations to bonds remain essentially unchanged. The allocations to HML are similar to those obtained by Jurek and Viceira (2005). They calculate a VAR model that includes two of the predictors considered here: the term premium and T-bill rate³⁸. However, they employ a calibration framework and do not evaluate the uncertainty of computed optimal portfolios.

Neither the dividend yield (Figure 2.4) or the default premium (Figure 2.5) yield conditional policies that show significant variation with respect to predictor value. Results in Figure 2.6 suggest underweighting equities when the term spread is low and overweighting bonds when term spread is high. The optimal policy with respect to the S&P 500 trend is to underweight equities and overweight bonds when the trend is greater than one standard deviation below its mean and to underweight bonds when the trend is greater than one standard deviation above its mean.

The downward slope of the relationship between the T-bill rate and optimal allocation to the index (Figure 2.8) shows that post-war U.S. experience validates the common wisdom that equity investment should decrease when short term interest rates rise. The effect of the T-bill rate on HML investment is statistically insignificant, but hints that an investor would have benefited during the in-sample period from holding larger positions in value stocks during periods of very high and very low short term interest rates.

Figures 2.12-15 illustrate the in-sample gain in CERs from conditioning on predictor variables. Results are shown for an investor with risk aversion of five. The solid line is the conditional CER of the conditionally optimal policy as a function of predictor value. The conditional CER is the return equivalent

³⁸Jurek and Viceira (2005) find weak statistical significant evidence of predictability. They employ a parameterized VAR model of returns and an approximate solution to the portfolio choice problem.

of a kernel weighted average over monthly utility outcomes (see equation (9)). Three other curves are plotted for comparison. The dashed curve is the conditional CER of the unconditionally optimal strategy. The dotted line is the unconditional CER that obtains when following the unconditionally optimal strategy. The dash dot line is the unconditional CER from following the optimal conditional strategy. The unconditional CERs are return equivalents of simple time-series averages of realized utilities for each strategy.

By comparing the conditional CER of the conditional and unconditional strategies, we can qualitatively examine the conditional significance of predictability to the asset allocation investor. First, consider results for the bonds and equities portfolio with dividend yield as predictor (left panels of Figure 2.12). The difference between conditional CERs of the conditional strategy (solid line) and unconditional strategy (dashed line) are very small relative to the uncertainty in the conditional CER estimate (shaded region). CER is a direct proxy for expected utility. Thus conditioning on the dividend yield does not improve conditional expected utility when conditioning on the dividend yield.

Similar results hold for all predictors for both the equities and bonds investor (left panels of Figures 2.12-15) and the equities, bonds and HML investor (right panels of Figures 2.12-15). With HML in the asset set, the conditional CER of the unconditional portfolio never leaves the two standard deviation region (right panels of Figures 2.12-15). There are some exceptions near the extremes of the empirical distribution for the case of an equities and bonds investor. At very low term spreads, the conditional CER of the unconditional strategy drops below the two standard error region around the conditional CER of the conditional strategy (top left panel of Figure 2.6). This implies potential benefits from reducing exposures to risky assets when the term spread is small or negative. The CER of the unconditional portfolio also drops below the error region at low values of S&P 500 trend and high values of the gold industry trend and consumer price index. However, these results must be interpreted with caution since the data density at the margins of the empirical distribution is small. This can lead to bias in a bootstrap framework. An empirical distribution is a noisy model of the true distribution in the tail area.

Table 2.11 shows conditional in-sample estimates of CERs attained by following the conditional and unconditional strategies for the portfolios: equities

and bonds; and equities, bonds, and HML. Bootstrap standard errors are given in brackets below each CER. The third line of each entry is a bootstrap estimate of the conditional p-value for the null hypothesis that conditioning does not improve CER. The CER improvement over the unconditional strategy is rarely more than one or two basis points. P-values near 0.5 reflect the lack of significant improvement of conditional expected utility from conditioning on single predictor variables. Gold industry trend and term spread are the only predictors that consistently yield p-values below 0.3.

The difference between significance results for conditional and unconditional hypotheses has consequences for evaluating predictability in an investment context. For an investor with a one-month horizon, the only concern is the significance in the conditional problem. While this paper focuses on the one-month investment horizon, the high persistence of the predictor series means that long horizons are required to ensure a high probability of visiting all predictor states. As a result, even for investment horizons of moderate length, investors will wish to condition on the starting value of the predictor when evaluating predictability evidence.

Table 2.12 lists increases in unconditional CERs from conditioning on individual predictors. The CERs are computed from realized returns from following optimal conditional policies depicted in Figures 2.4 to 2.11. As one might anticipate given the conditional CER results, the greatest CER improvements are attained by conditioning on gold industry trend or term spread. The results can be directly compared to the CERs estimated for the unconditionally optimal policy that are given in Table 2.5. For the conventional asset allocation problem the unconditional CER is 0.19. Only gold and term spread result in unconditional CERs that are more than two standard deviations greater than the unconditional portfolio. For the equities, bonds, and HML case, the unconditional CER estimate is 0.67. In terms of multiples of the standard deviation of the CER estimates, CER gains are less significant than for the case without HML. Only the gold industry trend yields an unconditional certainty equivalent gain of more than two standard deviations.

The results in the first column of Table 2.12 permit comparison of the CER gain from conditioning on individual predictors with that achieved by simply adding HML to the conventional asset allocation. From Table 2.5, the diver-

sification benefit of the HML portfolio is 0.48 basis points per month. CERs achieved by conditioning the conventional asset allocation portfolio on individual conditioning variables are almost all less than the CER estimate for the optimal unconditional portfolio with HML added to the asset set. The lone exception is the CER of the portfolio conditioned on gold industry trend for which the unconditional improvement in CER is equal to the diversification benefit of the HML portfolio.

Beyond indicating whether conditioning improves investment outcomes, the results in Figures 2.12-15 can also be used to determine whether predictors can be used to differentiate between good and poor investment states. Even if conditioning on predictors does not significantly improve investment prospects can we at least identify whether our prospects are good or poor in the current state? Good states feature conditional expected utility significantly greater than the unconditional expected utility of the strategy, and vice versa for bad states.

Using Figures 2.12-15, we can identify potentially good or poor states by identifying predictor values for which the unconditional expected utility of the conditional strategy is outside the error region for the conditional expected utility. For the equities and bonds investor (left panels), some predictors can be used to identify poor investment states. For example, a high up trend in gold industry returns corresponds to low conditional CER relative to unconditional levels. The same is also true for states with negative term spread. Similar regions obtain for a number of the other predictors. When HML is added to the asset set, the predictors lose any ability to differentiate good and poor states, with the notable exception of gold industry trend. Moderately positive upward trend in gold industry return corresponds to poor investment outlook for the asset allocation investor (bottom right panel of Figure 2.14).

The results for the term spread and trend (Figure 2.13), gold industry returns (Figure 2.14), and inflation (Figure 2.15) suggest that conditional policies have significantly higher conditional CERs relative to unconditional policies at the extremes of predictor distributions. For example, the conditional policy indicated for negative term spreads (or inverted yield curve) has significantly higher CER than the corresponding conditional returns of the unconditional policy. However, these results have to be interpreted with caution because the density of the data set decreases towards the extremes of the in-sample range

of the predictor variables.

Whether or not the conditional policy is adopted, the CER in negative term spread (inverted yield curve) environments is significantly less than the unconditional expected utility. This result suggests that an inverted yield curve indicates a state of the economy in which investment outlook is gloomy. However, following a conditional strategy does not mitigate the negative impact of being in this poor state.

2.6 Conditioning on Multiple Predictors

Analysis of the previous section demonstrates that conditioning on individual predictors does not significantly improve investment expectations. In this section, I examine whether this result continues to hold when the investor is able to condition on multiple predictors simultaneously.

2.6.1 Direct Estimator for Multiple Predictors

The nonparametric estimator (2.6) rapidly loses statistical power when applied to a multi-dimensional conditioning vector, i.e. when Z_t is a vector. Given a finite sample, the data density drops exponentially with number of dimensions. As a consequence, there is an exponential drop in rate of convergence with number of dimensions.

Motivated by the implementation of Aït-Sahalia and Brandt (2001), I consider a structural assumption that reduces the dimensionality of the nonparametric estimation. However, in forming the estimator, I make a slightly less restrictive modeling assumption.

Aït-Sahalia and Brandt (2001) reduce the dimensionality of the estimation problem through a partial parameterization of the model of section 2.2. With Z_t representing an M -vector of conditioning variables, the portfolio policy is assumed to depend on a parametric function $\mathcal{Z}(Z_t; \beta)$ of the conditioning variables where β is a vector of parameters. The parametric function projects realizations of Z_t onto a lower dimensional space. Hence, the dimensionality of the nonparametric estimation is reduced at the expense of requiring the estimation of a finite number of unconditional (or global) parameters β . Instead of model-

ing portfolio allocations as nonparametric functions of a vector Z_t defining the state at time t , portfolio allocations are modeled as a nonparametric function of a parametric function $\mathcal{Z}(Z_t; \beta)$ of the predictor variables, where β is a vector of parameters. The function $\mathcal{Z}(\cdot; \beta) : \mathbb{R}^M \rightarrow \mathbb{R}^L$ reduces the dimension of the nonparametric problem (i.e., $L < M$).

The semiparametric approximation mitigates problems with the curse of dimensionality because the nonparametric part of the estimation is with respect to a lower dimension variable, $\mathcal{Z}(Z_t; \beta)$. If we further assume that $\mathcal{Z}(Z_t; \beta) = \beta' Z_t$ where β is an $L \times M$ matrix, then we obtain a model analogous to the semiparametric index model that has received attention in the regression literature.³⁹ In the empirical application to follow, I assume β has one dimension, i.e., $L = 1$.

Under the semiparametric specification, additional predictors are incorporated into the model via the addition of a finite number of parameters per dimension. These parameters add to model complexity, but at a much slower rate than if an additional dimension were added for nonparametric estimation. These parameters, index coefficients β , are global parameters. As a result, they are estimated by optimizing an unconditional objective. Unlike nonparametric portfolio policies, the estimate converges with \sqrt{n} efficiency once an appropriate decision criteria has been specified. For example, in the standard regression framework, index coefficients might be chosen to minimize the squared differences between observed data and the best fit curve for the chosen bandwidth.

I first follow Aït-Sahalia and Brandt (2001) in assuming that the investor's optimal conditional portfolio is independent of information in Z_t that is not captured by $\mathcal{Z}(Z_t, \beta)$. In other words, the hypothesis is that an investor with an information set of infinite size would infer an optimal portfolio that depends only on $\mathcal{Z}(Z_t; \beta)$, i.e.,

$$\arg \max_{\alpha_t} E \left[u(W_t \alpha_t' R_{t+1}) \middle| Z_t \right] \equiv \alpha_t (\beta^\top Z_t), \quad (2.10)$$

where \equiv indicates equality by assumption. Assuming that $\mathcal{Z}(Z_t, \beta) = \beta' Z_t$, the policy of the investor depends only on the subspace Λ of \mathbb{R}^M spanned by the

³⁹In the regression literature, models in which dimensionality of right-hand side vector is reduced through partial parameterization are called semiparametric models. The term single-index model is used when the dimensionality of the parametric problem is reduced to one (see e.g. Hardle et al. (2004)).

column space of β .

Of course, β is unknown and must be estimated. Aït-Sahalia and Brandt (2001) note that a GMM estimator for β can be constructed based on assumption (2.10). The following first order conditions are a direct consequence of (2.10).

$$E \left[u' (W_t \alpha'_t (\beta^\top Z_t) R_{t+1}) R_{t+1} \middle| Z_t \right] = 0. \quad (2.11)$$

Upon multiplying each of the above conditions by a vector of predetermined functions of the conditioning variables $g(Z_t)$ and taking unconditional expectations, we obtain a set of moment conditions. These conditions can be used to estimate β by solving the GMM problem

$$\min_{\beta} E \left[u' (W_t \alpha (\beta' Z_t)' R_{t+1}) R_{t+1} \otimes g(Z_t) \right]^\top W E \left[u' (W_t \alpha (\beta' Z_t)' R_{t+1}) R_{t+1} \otimes g(Z_t) \right] \quad (2.12)$$

$$\beta' \beta = 1 \quad (2.13)$$

where W is a weighting matrix, $g : \mathbb{R}^M \rightarrow \mathbb{R}^{M_g}$ is a predetermined function of the vector of predictor variables. To ensure identifiability, the number of moment conditions must exceed the number of independent index coefficients; i.e., $M_g \times N > L(M - 1)$. The constraint (2.13) arises because the optimal directions are only identified up to scale. As a result, the number of free parameters to be estimated is equal to one less than the number of predictor variables M .⁴⁰

Given assumption (2.10), for known β the investor's optimal portfolio policy is a solution to

$$\alpha_t(\beta^\top Z_t) = \arg \max_{\alpha_t} E \left[u(W_t \alpha'_t R_{t+1}) \middle| \beta' Z_t \right]. \quad (2.14)$$

Thus, with known β , the estimation problem for the portfolio policy is equiva-

⁴⁰I follow Aït-Sahalia and Brandt (2001) in setting the weighting matrix (2.12) proportional to the inverse covariance matrix of the moment conditions $W = \text{Cov}[u'(W_t \alpha (\beta' Z_t)' R_{t+1}) R_{t+1} g(Z_t)]^{-1}$. Aït-Sahalia and Brandt (2001) write out the unconditional utility maximization problem for selecting β and show that the optimal choices of $g(z_t)$ would be the gradient of $\alpha(Z_t' \beta)$ with respect to β . While theoretically optimal, these instruments are not ideal in practice since they require estimates of the derivative of the policy function α with respect to each element of β . Instead, I use instruments that are linear in each predictor variable.

lent to that for a single predictor with $\beta'Z$ taking the role of the single predictor. Consequently, the methods of the previous section can be used for policy estimation.

Aït-Sahalia and Brandt (2001) note that the assumption (2.10) may be sub-optimal. In a second approach, I make a less restrictive assumption. Instead of directly assuming a restricted functional dependence of portfolio policy on predictor variables, I arrive at the same functional form for the portfolio policy by assuming the investor is limited in the number of independent shocks that they have the capacity to condition on. The limitation could arise either due to a bounded information gathering capacity or a bounded ability to process information. In the current context, this assumption limits the investors conditioning information to $\beta^\top Z_t$ where, as before, β projects Z_t onto a lower dimensional subspace of \mathbb{R}^M . Given β , the optimal policy under the current assumption is

$$\alpha(Z_t) \approx \alpha(\beta'Z_t) = \max_{\alpha_t} E \left[u(W_t \alpha_t^\top R_{t+1}) \middle| \beta^\top Z_t \right]. \quad (2.15)$$

This assumption does not limit the information set, just the amount of information that the investor has the ability to use. Hence, the index matrix β is selected by the investor. This is a weaker assumption than that of the previous paragraph. In the previous paragraph, the investor is assumed to condition on all information, but that the resulting portfolio only depends on the projection of predictor shocks onto a lower dimensional subspace. Here, the true optimal policy may depend on the full information subset, but the investor must choose an optimal subspace of the set of possible predictor outcomes upon which to condition their portfolio policy.

The two assumptions differ in their implications for index identification. Assumption (2.15) does not imply that the investor can gain no information from observing shocks along directions perpendicular to those included in β . As a consequence, the conditioning variables are not necessarily instruments of the first-order conditions. Thus, the moment conditions used to construct the GMM objective (2.12) do not necessarily hold. This precludes using the estimator described for the previous case. Instead, the investor chooses directions, the columns of β , that span the subspace of the space of possible predictor vector

outcomes that yields the largest gain in expected utility.

$$\max_{\beta} E \left[\max_{\alpha_t} E [u(W_t \alpha'_t R_{t+1}) | \beta' Z] \right] = \max_{\beta} E \left[u(W_t \alpha(\beta' Z)' R_{t+1}) \right], \quad (2.16)$$

$$\beta' \beta = 1. \quad (2.17)$$

In essence, the investor is choosing the most informative index coefficients given the objective of maximizing utility. In the following I shall refer to (2.15) as the limited-information case, and (2.10) as the restricted case.

To provide further illustration of the differences between the formulations, consider the first-order conditions of the index optimization problem (2.16)

$$E [u'(W_t \alpha(\beta^\top Z_t)^\top R_{t+1}) (\alpha'(\beta^\top Z_t)^\top R_{t+1}) Z_t] = 0. \quad (2.18)$$

Evaluation of the above first-order condition requires not only knowledge of the policy function, but its derivative as well. By contrast the first-order conditions of the restricted case depend only on the policy function itself. Hence, despite the strong assumption upon which it is based, the restricted case may be more efficient in finite samples since the first order condition of its objective does not depend on derivatives of an estimated nonparametric function, making it more amenable to solution by standard optimization techniques. In addition, the restricted approach may be easier to implement in practice given the analytic form of the first-order condition.

Two factors diminish the practical advantage of the restricted case. First, the normalization constraint confines possible index values to a finite domain. As such, grid search is a feasible approach to the index optimization problem. Because grid search can be based on direct evaluation of the objective function, the differentiability of the objective in the restricted case offers less advantage. Second, neither the objective function in (2.16) nor in (2.12) is concave. Because $\alpha(\cdot)$ is a nonparametric function, there is no guarantee that the maximands are concave in β . Consequently, in general some form of grid search is necessary whether or not the optimal portfolio policy is restricted by assumption.

For implementation purposes, the index optimization problems present some solution challenges. For a grid search to be successful, the estimation problem must satisfy two criteria: the space to be searched must be of low dimension

and the curvature of the objective function must be constrained. Data density drops exponentially with the number of grid dimensions. As a result, even on a finite domain, grid search becomes impractical for problems with more than a few dimensions. However, the identification constraint (2.13) reduces the dimensionality of the problem by one. As such, grid search over the objective functions in (2.16) and (2.12) is feasible for problems including up to four predictor variables.

Limitations on the curvature of the objective function with respect to β are required to ensure that a sparse sampling of possible index values will locate the desired extremum. However, reasonable smoothness of these derivatives can be assured by choice of a sufficiently large bandwidth for the portfolio policy. In the applications to follow, this proves to be a non-issue if the bandwidth selection procedure described in Appendix A is employed in determining portfolio policies for given index values. This approach yields objectives that are of sufficient smoothness for estimation on reasonably sparse grids in both the limited information and restricted cases.⁴¹

In addition to dealing with a non-concave objective function, evaluation of the objective functions of the index optimization problems requires nonparametric estimation of portfolio policies for each data point. This renders function evaluation computationally expensive. I deal with this issue by evaluating the $\alpha(\beta'Z)$ at a reasonably small subset of points in the range of $\beta'Z$. Realizations at intermediate points are then obtained by interpolation. Because the smoothness of the policy function is determined by the bandwidth used in the point-by-point optimization, two to four points per bandwidth length along with a polynomial interpolation is sufficient to ensure that numerical errors that are introduced are within statistical uncertainty error.

The global search is implemented over a finite grid. The dimension of the volume to be searched in β space is one less than the dimension of the predictor variable. To keep the global search computationally feasible, I restrict the number of predictors included in the index to four.⁴² This is perhaps less of a limitation than it might appear. Previous research on model selection for return predictability in a linear regression framework has shown that standard

⁴¹For a three dimensional problem, I obtain stable estimated for a grid with $O(10^3)$ points.

⁴²Ait-Sahalia and Brandt (2001) also examine an index of four predictor variables in their empirical application.

model selection procedures rarely select more than four predictors (Bossaerts and Hillion (1999)).

In the empirical analysis that follows, I estimate utilities of the optimal strategies under the assumptions of both the restricted case and the limited information case. In the current application, I do not find a significant difference between the results obtained under the two assumptions. For brevity, I only report results obtained under the limited information assumption.

2.6.2 Model Selection

Turning now to the evaluation of conditional significance of predictability in a multi-predictor setting, I evaluate optimal index values for all seventy combinations of four predictor variables. Index coefficients are estimated by the grid search procedure described in section 6.1.

Table 2.13 displays coefficients estimated for the six combinations of four variables that yield the highest in-sample CER for the equities, bonds, and HML portfolio. The standard errors of the CERs for the equities, bonds, and HML portfolio are between 26 and 29 basis points, and for the case without HML, the standard errors range from 16 to 19 basis points. The difference between the best and sixth best CER is 5 basis points in the no-HML case and 2 basis points for the case with HML.

The term spread and gold industry trend are the dominant predictors. The predictor combinations that produce the top six CERs all include both the term spread and gold industry trend. Of the six other predictors, all but realized equity index volatility appear in at least one of the top six combinations. In all indices, the highest index coefficients are associated with the gold industry trend and term spread predictors. This evidence provides robust support for use of an index consisting of term spread and gold industry trend for estimating allocations. The seventh line of Table 2.13 shows index coefficients estimated for the two-variable index consisting of term spread and gold industry trend. The in-sample CERs achieved using this latter index are only 0.06 and 0.09 less than the maximum CERs achieved with four predictor variables for the no-HML and HML cases respectively. These differences are within one bootstrap standard error and are insignificant.

The last line of Table 2.13 lists index coefficients for an index made up of

the four predictors studied by Aït-Sahalia and Brandt (2001). The term spread dominates this latter index. For the no-HML case, the coefficient on the term spread is four times larger than that on the other three variables. The term spread also boasts the largest coefficient in the case with HML. In agreement with results reported by Aït-Sahalia and Brandt (2001), the term spread and S&P 500 trend variables are the largest index components.

2.6.3 Certainty Equivalent Returns with Multiple Predictors

Table 2.14 lists conditional CERs for strategies conditioned on the particular indices listed in Table 2.13 along with conditional CERs from following the corresponding unconditional policies. Results are calculated for the case with and without HML. For each set of predictors, results are given for index values at the 20th, 50th and 80th percentiles of the index distributions. The first six sets of predictors correspond to the first six predictor combinations listed in Table 2.13. In addition, results are given for the two-predictor index combining term spread and gold industry trend and the four predictor index of Aït-Sahalia and Brandt (2001).

The results can be contrasted with those obtained in the single predictor case. The improvement in conditional CERs from following the conditional policies as opposed to the unconditional policy are significant for the predictor indices that include the term spread and gold industry trend predictors. In contrast, in the single predictor case, the hypotheses that the unconditional policy yielded as high a conditional CER as the conditional policy could not be rejected (see Table 2.10).

2.7 Conclusion

The empirical analysis of this chapter yields three observations: i) the benefits of incorporating a proxy for an additional priced risk factor (the value premium) outweigh the gains from conditioning on individual predictor variables; ii) short horizon investors do not benefit from conditioning on single predictor variables in their portfolio decisions; and iii) short horizon investors may be able

to improve expected outcomes by conditioning portfolio decisions on multiple predictors simultaneously.

In prior studies of predictability, the significance of predictability is based on averaging over long time periods. For predictors with high persistence, the investment significance of predictability is dampened for investors with short horizons. The conditional results presented in this paper help to distinguish predictors that are likely to be of interest to that large contingent of investors that evaluates their performance (or have their performance evaluated) at shorter horizons.

In evaluating the predictability results, the risk of data snooping must be borne in mind. Many of the predictor variables that I examine are extracted from an extensive literature, of which the majority is based on data from the United States covering periods that either include or substantially overlap the sample period considered in this paper. The variables can be legitimately classified as snooped with regard to their ability to predict index returns.

Where my results are additive is in the assessment of investment significance of predictability under the assumption that the predictability is real but that the underlying conditional distribution of returns is unknown. Under this assumption, I demonstrate that simultaneous conditioning on the term spread and gold industry trend increase an investor's expected utility by approximately 50 basis points per month - an amount which is approximately equal to the diversification benefit of incorporating a mechanical portfolio based on the value premium.

A topic for further study is the implications for hedging demands. As horizon increases, the effective sample size shrinks, and statistical power decreases. However, at infinite horizon, a single policy function holds. The sensitivity of this policy to data realization is a promising problem for future exploration using techniques discussed in this paper.

Table 2.1: Summary Statistics

This table shows the following summary statistics for monthly data on asset returns and forecasting variables for the time period January, 1952 to November, 2004: mean, median, standard deviation, skewness, kurtosis, Sharpe ratio, and lag one autocorrelation ρ_1 . The assets are the value weighted CRSP index minus the risk-free interest rate (Index), the HML, SMB and UMD portfolios of Ken French, and returns on an index of long-term bonds minus risk-free rate. The predictor variables are dividend yield (div), default premium (def), term spread (term), S&P 500 trend (trend), the 30-day T-bill rate (tbill), gold industry trend (gold), log of previous month realized volatility (realvol) and inflation (cpi). The mean, median, and standard deviation are annualized.

	mean	median	std	skewness	kurtosis	sharpe	ρ_1
equities	0.071	0.116	0.147	-0.510	5.033	0.136	0.062
SMB	0.022	0.006	0.103	0.577	9.326	0.061	0.063
HML	0.051	0.047	0.094	0.053	5.915	0.152	0.133
RF	0.050	0.048	0.008	1.068	4.643	1.748	0.960
UMD	0.105	0.112	0.129	-0.662	9.199	0.225	-0.034
bonds	0.013	0.000	0.073	0.283	4.604		0.069
div	3.377	3.244	1.101	0.228	2.385		0.986
def	0.934	0.810	0.412	1.463	5.382		0.972
term	1.305	1.210	1.407	-0.126	3.309		0.964
trend	3.617	4.700	8.968	-0.706	3.572		0.923
tbill	0.002	0.012	0.098	-0.492	8.211		0.760
gold	1.129	0.878	2.451	0.695	4.894		0.899
realvol	-2.260	-2.294	0.455	0.349	3.585		0.668
cpi	0.312	0.259	0.231	1.422	4.956		0.992

Table 2.2: Correlations of Asset Returns and Predictors

This table shows the correlation matrix for the predictor variables and subsequent month asset returns. The predictor variables are dividend yield (div), default premium (def), term spread (term), S&P 500 trend (trend), the 30-day T-bill rate (tbill), gold industry trend (gold), log of previous month realized volatility (realvol) and inflation (cpi).

	equities	SMB	HML	UMD	bonds	div	def	term	trend	tbill	gold	realvol	cpi
equities	1.00												
SMB	0.26	1.00											
HML	-0.36	-0.26	1.00										
UMD	-0.06	-0.01	-0.14	1.00									
bonds	0.16	-0.08	0.02	0.04	1.00								
div	0.09	0.05	-0.05	-0.01	0.04	1.00							
def	0.06	0.09	-0.00	-0.04	0.07	0.39	1.00						
term	0.11	0.04	-0.02	0.00	0.11	-0.23	0.10	1.00					
trend	0.01	-0.05	0.03	0.06	-0.12	-0.16	-0.08	0.12	1.00				
tbill	-0.07	-0.07	0.07	-0.06	-0.01	-0.04	-0.25	-0.46	-0.01	1.00			
gold	-0.07	0.07	-0.03	0.01	-0.20	-0.03	0.06	-0.03	0.14	0.07	1.00		
realvol	0.01	0.04	-0.04	-0.03	0.07	-0.12	0.34	0.02	-0.36	-0.04	0.05	1.00	
cpi	-0.07	0.06	-0.01	0.00	-0.03	0.44	0.55	-0.42	-0.27	0.09	0.20	0.26	1.00

Table 2.3: Volatility Correlations

This table shows the correlations of realized volatilities with lagged realized volatilities and predictor variables. The realized volatilities are calculated monthly using within month daily returns. The predictor variables are dividend yield (div), default premium (def), term spread (term), S&P 500 trend (trend), the 30-day T-bill rate (tbill), gold industry trend (gold), log of previous month realized volatility (rvol) and inflation (cpi).

	equities	SMB	HML	equities ₋₁	SMB ₋₁	HML ₋₁
equities	1.00					
SMB	0.81	1.00				
HML	0.69	0.57	1.00			
equities₋₁	0.58	0.43	0.49	1.00		
SMB₋₁	0.50	0.52	0.43	0.81	1.00	
HML₋₁	0.49	0.38	0.74	0.70	0.57	1.00
div	-0.09	-0.15	-0.27	-0.06	-0.13	-0.26
def	0.20	0.11	0.05	0.23	0.16	0.07
term	0.01	0.14	-0.06	0.03	0.15	-0.04
trend	-0.39	-0.13	-0.27	-0.39	-0.15	-0.27
tbill	-0.04	-0.01	-0.05	-0.12	-0.08	-0.10
gold	0.04	0.03	-0.18	0.08	0.06	-0.15
realvol	0.92	0.70	0.65	0.62	0.51	0.50
cpi	0.11	-0.03	-0.04	0.11	-0.03	-0.05

Table 2.4: Unconditional Portfolio Weights

This table shows optimal portfolio allocations as a fraction of wealth. The investor has power utility with risk aversion given by γ . Allocations are listed for an unconstrained investor (Panel A), and for an investor subject to long-only constraints (Panel B). Long-only constraints restrict investment in equities and bonds to be positive, and restrict investment in the HML portfolio to less than 1/3 the allocation to equities in absolute value. Optimal allocations are listed for portfolios made up of the risk-free asset plus, in order from top to bottom, i) the equity index, ii) the HML portfolio, iii) the equity and bond indices, iv) the equity index and the HML portfolio, v) the equity and bond indices plus the HML portfolio. The weight in the risk-free asset is equal to one minus the sum of the allocations to the equity and bond indices. The estimates are obtained by method of moments. Standard errors are calculated based on 1000 stationary bootstrap resamplings of the data set, and are listed in brackets.

	Panel A: Unconstrained				Panel B: No Short Sales			
	$\gamma = 1$	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$	$\gamma = 1$	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$
equities	2.82 [0.83]	1.50 [0.48]	0.61 [0.20]	0.31 [0.10]	1.00 [0.06]	1.00 [0.12]	0.61 [0.19]	0.31 [0.10]
HML	5.15 [1.54]	2.72 [0.87]	1.11 [0.36]	0.56 [0.18]				
equities	2.81 [0.87]	1.47 [0.49]	0.60 [0.20]	0.30 [0.10]	1.00 [0.07]	1.00 [0.13]	0.60 [0.20]	0.30 [0.10]
bonds	1.63 [1.80]	0.78 [0.94]	0.30 [0.38]	0.15 [0.19]	0.00 [0.06]	0.00 [0.11]	0.30 [0.20]	0.15 [0.16]
equities	4.50 [0.81]	2.50 [0.52]	1.03 [0.22]	0.52 [0.11]	1.00 [0.10]	1.00 [0.16]	0.61 [0.21]	0.30 [0.11]
HML	7.52 [1.91]	4.24 [1.08]	1.75 [0.44]	0.88 [0.22]	0.33 [0.18]	0.33 [0.09]	0.20 [0.07]	0.10 [0.04]
equities	4.54 [0.85]	2.49 [0.52]	1.02 [0.22]	0.51 [0.11]	1.00 [0.12]	1.00 [0.18]	0.58 [0.22]	0.29 [0.11]
bonds	1.24 [1.72]	0.49 [0.95]	0.17 [0.39]	0.08 [0.20]	0.00 [0.12]	-0.00 [0.17]	0.42 [0.22]	0.21 [0.18]
HML	7.56 [1.92]	4.23 [1.08]	1.74 [0.44]	0.87 [0.22]	0.33 [0.18]	0.33 [0.09]	0.19 [0.07]	0.10 [0.04]

Table 2.5: Unconditional Certainty Equivalent Returns

Each row of this table lists in-sample estimates of monthly certainty equivalent returns (CERs) attained by allocating to estimated optimal portfolios. Optimal portfolios are estimated for an investor with CRRA utility. Results are computed for four values of risk aversion parameter γ . CERs are given in percent. Allocations are listed for an unconstrained investor (Panel A), and for an investor subject to long-only constraints. Long-only constraints restrict investment in equities and bonds to be positive, and restrict investment in the HML portfolio to less than 1/3 the allocation to equities in absolute value. The left column lists assets in each portfolio. Portfolios include a risk-free asset plus a subset of: an equity index minus risk-free asset (equities), a long-term bond index minus the risk-free asset (bonds), HML, SMB and UMD portfolios provided by Ken French. CERs are estimates given optimal portfolio policies obtained by method of moments. Standard errors are calculated based on 1000 stationary bootstrap resamplings of the data set, and are listed in brackets.

Asset Set	Panel A: Unconstrained				Panel B: No Short Sales			
	$\gamma = 1$	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$	$\gamma = 1$	$\gamma = 2$	$\gamma = 5$	$\gamma = 10$
{equities}	0.86 [0.51]	0.44 [0.27]	0.18 [0.11]	0.09 [0.05]	0.49 [0.17]	0.39 [0.17]	0.18 [0.11]	0.09 [0.05]
{HML}	1.12 [0.63]	0.57 [0.33]	0.23 [0.13]	0.12 [0.07]				
{equities, bonds}	0.93 [0.54]	0.47 [0.28]	0.19 [0.11]	0.10 [0.06]	0.49 [0.17]	0.39 [0.17]	0.19 [0.11]	0.10 [0.06]
{equities, HML}	3.15 [1.04]	1.64 [0.57]	0.66 [0.23]	0.33 [0.12]	0.63 [0.16]	0.55 [0.16]	0.28 [0.10]	0.14 [0.05]
{equities, bonds, HML}	3.19 [1.07]	1.65 [0.58]	0.67 [0.24]	0.33 [0.12]	0.63 [0.16]	0.55 [0.16]	0.28 [0.11]	0.14 [0.05]
{equities, bonds, HML, SMB}	3.49 [1.18]	1.79 [0.62]	0.72 [0.25]	0.36 [0.13]	0.56 [0.17]	0.55 [0.16]	0.29 [0.11]	0.15 [0.05]
{equities, bonds, HML, UMD}	4.95 [1.01]	2.87 [0.62]	1.23 [0.28]	0.63 [0.14]	0.76 [0.17]	0.66 [0.17]	0.35 [0.12]	0.17 [0.06]

Table 2.6: Bootstrap Test of Expected Utility Improvement

This table gives bootstrap p-values for the null hypothesis that adding an asset to an investor's asset set improves expected utility. Starting asset sets are listed in the left column. The second column is the asset being added. Results are computed for a power utility investor with risk aversion of five. The p-values are computed based on expected utility differences computed for 1000 bootstrap samples.

Asset Set	Added Asset	<i>p</i> -value
Panel A: Unconstrained		
{}	{equities}	0.054
{}	{bonds}	0.26
{}	{HML}	0.046
{equities}	{bonds}	0.33
{equities}	{HML}	0.009
{equities, HML}	{bonds}	0.417
{equities, bonds}	{HML}	0.009
{equities, bonds, HML}	{SMB}	0.237
{equities, bonds, HML}	{UMD}	0.001
Panel B: No Short Sales		
{}	{equities}	0.054
{}	{bonds}	0.26
{equities}	{bonds}	0.33
{equities}	{HML}	0
{equities, HML}	{bonds}	0.479
{equities, bonds}	{HML}	0.001
{equities, bonds, HML}	{SMB}	0.171
{equities, bonds, HML}	{UMD}	0.059

Table 2.7: Abbreviations Used for Conditioning Variables

div	dividend yield (log)
def	default premium
term	term spread
trend	S&P 500 trend
tbill	T-bill rate
gold	gold index return
realvol	CRSP index realized volatility (log)
cpi	inflation

Table 2.8: Returns Regressed on Predictors

This table shows coefficient estimates for predictive regressions of realized returns on lagged conditioning variables. Full names of the conditioning variables are given in Table 7. Panel A shows coefficients of linear regressions of next month's returns on a individual predictor variables. Panel B presents coefficients of multivariate regressions of asset returns on the full set of predictor variables. The last column lists R^2 values for the multivariate regressions.

	div	def	term	trend	tbill	gold	realvol	cpi	R^2
	Panel A: univariate regression results								
equities	0.353 [0.180]	0.185 [0.187]	0.376 [0.188]	-0.012 [0.208]	-0.347 [0.164]	-0.333 [0.216]	0.224 [0.175]	-0.229 [0.211]	-
bonds	0.050 [0.102]	0.122 [0.110]	0.217 [0.108]	-0.183 [0.089]	-0.029 [0.096]	-0.348 [0.115]	0.232 [0.076]	-0.054 [0.116]	-
HML	-0.126 [0.131]	0.047 [0.119]	0.009 [0.118]	0.041 [0.115]	0.157 [0.120]	-0.061 [0.145]	-0.047 [0.129]	-0.021 [0.128]	-
	Panel B: multivariate regression results								
equities	0.694 [0.194]	-0.057 [0.268]	0.272 [0.212]	0.167 [0.224]	-0.129 [0.192]	-0.258 [0.211]	0.496 [0.190]	-0.412 [0.271]	0.055
bonds	0.135 [0.089]	0.037 [0.139]	0.291 [0.105]	-0.068 [0.100]	0.156 [0.093]	-0.341 [0.108]	0.240 [0.087]	-0.042 [0.147]	0.060
HML	-0.225 [0.144]	0.256 [0.162]	0.027 [0.151]	-0.025 [0.130]	0.226 [0.134]	-0.083 [0.146]	-0.150 [0.154]	-0.024 [0.178]	0.035

Table 2.9: Realized Volatility Regressed on Predictors

This table shows coefficient estimates for linear regressions of realized monthly volatilities on lagged predictors. Full names of the conditioning variables are given in Table 7. Panel A shows coefficients of linear regressions of realized one-month volatilities of each asset on individual predictor variables. Panel B lists coefficients of multivariate linear regressions of realized volatilities of each asset on the full set of predictor variables. The last column lists R^2 values for the multivariate regressions.

	div	def	term	trend	tbill	gold	realvol	cpi	R^2
	Panel A: univariate regression results								
equities	-0.809 [0.394]	1.137 [0.275]	-0.081 [0.379]	-1.744 [0.405]	-0.458 [0.262]	0.463 [0.452]	3.973 [0.255]	0.745 [0.325]	-
HML	-1.059 [0.267]	0.082 [0.154]	-0.279 [0.181]	-0.722 [0.195]	-0.199 [0.164]	-0.585 [0.243]	1.797 [0.227]	-0.138 [0.166]	-
	Panel B: multivariate regression results								
equities	-1.445 [0.423]	1.519 [0.296]	-0.458 [0.245]	-1.868 [0.367]	-0.528 [0.244]	0.620 [0.311]	2.898 [0.327]	-0.327 [0.371]	0.457
HML	-1.692 [0.302]	0.926 [0.203]	-0.864 [0.188]	-0.734 [0.175]	-0.453 [0.149]	-0.504 [0.159]	1.221 [0.200]	-0.144 [0.220]	0.408

Table 2.10: Portfolio Weights as a Function of Predictor Values

This table shows portfolio allocations conditioned on individual predictor variables. Portfolio weights are estimated by solving a nonparametric approximation to the investor's Euler equation. The optimal strategies are estimated for a power utility investor with risk aversion equal to five and a one-month investment horizon. Portfolio weights are shown for four assets sets: equities only; HML only; equities and bonds; and equities, bonds, and HML. Asymptotic standard errors are given in square brackets.

		{equities}	{bonds}	{equities,bonds}		{equities,bonds,HML}		
div	0.2	0.49	0.35	0.46	0.35	0.90	0.10	1.77
		[0.31]	[0.58]	[0.27]	[0.56]	[0.31]	[0.57]	[0.58]
	0.5	0.44	0.22	0.48	0.06	0.93	-0.14	2.12
		[0.27]	[0.52]	[0.25]	[0.46]	[0.26]	[0.46]	[0.47]
	0.8	0.79	0.10	0.79	-0.22	1.17	-0.18	2.10
		[0.30]	[0.59]	[0.27]	[0.50]	[0.27]	[0.47]	[0.59]
def	0.2	0.53	0.04	0.53	-0.04	0.97	-0.21	1.73
		[0.30]	[0.58]	[0.28]	[0.53]	[0.27]	[0.51]	[0.53]
	0.5	0.46	0.18	0.47	0.13	0.92	-0.02	1.68
		[0.26]	[0.49]	[0.25]	[0.47]	[0.26]	[0.47]	[0.50]
	0.8	0.51	0.89	0.49	0.68	0.86	0.52	1.58
		[0.25]	[0.52]	[0.22]	[0.47]	[0.26]	[0.47]	[0.48]
term	0.2	0.38	0.03	0.46	-0.10	0.86	-0.22	1.39
		[0.32]	[0.59]	[0.30]	[0.53]	[0.30]	[0.51]	[0.47]
	0.5	1.07	0.46	0.98	0.09	1.30	-0.02	1.55
		[0.27]	[0.54]	[0.24]	[0.49]	[0.26]	[0.47]	[0.57]
	0.8	1.10	0.59	1.08	0.50	1.46	0.48	1.82
		[0.39]	[0.58]	[0.32]	[0.58]	[0.34]	[0.55]	[0.63]

		{equities}	{bonds}	{equities,bonds}		{equities,bonds,HML}		
trend	0.2	0.34	0.65	0.34	0.56	0.90	0.29	1.83
		[0.43]	[0.53]	[0.40]	[0.53]	[0.36]	[0.51]	[0.52]
	0.5	0.93	0.41	0.85	0.03	1.23	-0.15	1.84
		[0.24]	[0.56]	[0.23]	[0.51]	[0.25]	[0.49]	[0.60]
	0.8	0.77	-0.09	0.79	-0.16	1.16	-0.27	1.83
		[0.26]	[0.62]	[0.26]	[0.54]	[0.31]	[0.52]	[0.65]
tbill	0.2	0.76	0.84	0.71	0.53	1.03	0.34	1.64
		[0.25]	[0.61]	[0.23]	[0.53]	[0.24]	[0.51]	[0.49]
	0.5	0.60	0.25	0.59	0.21	0.91	0.12	1.62
		[0.22]	[0.48]	[0.22]	[0.46]	[0.24]	[0.47]	[0.51]
	0.8	0.46	0.25	0.46	0.15	0.84	0.07	1.63
		[0.24]	[0.45]	[0.23]	[0.43]	[0.25]	[0.44]	[0.50]
gold	0.2	1.26	0.90	1.15	0.35	1.55	0.22	1.67
		[0.28]	[0.60]	[0.26]	[0.57]	[0.27]	[0.54]	[0.54]
	0.5	1.11	0.20	1.07	-0.22	1.37	-0.24	1.56
		[0.26]	[0.56]	[0.25]	[0.53]	[0.26]	[0.51]	[0.51]
	0.8	0.49	-0.50	0.58	-0.56	0.97	-0.54	1.71
		[0.36]	[0.57]	[0.33]	[0.54]	[0.31]	[0.52]	[0.51]
realvol	0.2	1.06	0.58	0.91	0.09	1.25	-0.25	2.69
		[0.32]	[0.64]	[0.28]	[0.56]	[0.28]	[0.53]	[0.58]
	0.5	0.59	0.22	0.61	-0.07	1.12	-0.28	2.50
		[0.27]	[0.49]	[0.24]	[0.43]	[0.25]	[0.43]	[0.52]
	0.8	0.46	0.00	0.49	-0.01	0.94	-0.05	1.78
		[0.25]	[0.43]	[0.25]	[0.41]	[0.28]	[0.41]	[0.54]
cpi	0.2	0.97	0.62	0.95	0.69	1.33	0.59	1.76
		[0.27]	[0.49]	[0.25]	[0.48]	[0.27]	[0.47]	[0.55]
	0.5	0.76	0.85	0.74	0.70	1.18	0.54	1.74
		[0.26]	[0.45]	[0.25]	[0.43]	[0.27]	[0.44]	[0.51]
	0.8	0.40	0.99	0.37	0.76	0.90	0.46	1.79
		[0.32]	[0.58]	[0.30]	[0.50]	[0.30]	[0.50]	[0.51]

Table 2.11: Certainty Equivalent Returns as a Function of Predictor Values
This table displays certainty equivalent returns (CERs) portfolio strategies conditioned on a single predictor variable. The optimal strategies are estimated for a power utility investor with risk aversion equal to five and a one-month investment horizon. CERs are shown for three possible assets sets: equities only; equities and bonds; and equities, bonds, and HML. The equity and bond assets are hedge funds that are long the appropriate index and short the risk-free asset. The first row shows CER of unconditional strategies. Subsequent rows show conditional CER calculated as kernel averages of utility outcomes at different (standardized) values of predictor variables. Two columns are shown for each set of assets. The first is the conditional CER of an optimal conditional strategy. The second is the conditional CER of the optimal unconditional strategy. Standard errors are calculated by the delta method and given in square brackets.

		{equities,bonds}		{equities,bonds,HML}	
		<i>CER</i>	<i>CER_{unc}</i>	<i>CER</i>	<i>CER_{unc}</i>
{div}	0.2	0.10	0.09	0.60	0.59
		[0.11]	[0.14]	[0.28]	[0.29]
		(0.38)		(0.39)	
	0.5	0.10	0.09	0.66	0.62
		[0.10]	[0.14]	[0.26]	[0.25]
		(0.38)		(0.26)	
	0.8	0.29	0.25	0.81	0.78
		[0.19]	[0.17]	[0.33]	[0.29]
		(0.28)		(0.34)	
{def}	0.2	0.12	0.11	0.59	0.58
		[0.12]	[0.15]	[0.26]	[0.27]
		(0.38)		(0.36)	
	0.5	0.11	0.10	0.56	0.56
		[0.10]	[0.14]	[0.25]	[0.27]
		(0.37)		(0.39)	
	0.8	0.21	0.18	0.58	0.55
		[0.14]	[0.14]	[0.26]	[0.28]
		(0.32)		(0.31)	
{term}	0.2	0.09	0.07	0.42	0.38
		[0.09]	[0.13]	[0.21]	[0.26]
		(0.28)		(0.29)	
	0.5	0.41	0.35	0.73	0.68
		[0.17]	[0.10]	[0.27]	[0.27]
		(0.21)		(0.22)	
	0.8	0.60	0.48	1.06	0.95
		[0.25]	[0.13]	[0.37]	[0.30]
		(0.15)		(0.17)	

		{equities,bonds}		{equities,bonds,HML}	
		<i>CER</i>	<i>CER_{unc}</i>	<i>CER</i>	<i>CER_{unc}</i>
{trend}	0.2	0.11	0.08	0.64	0.62
		[0.12]	[0.18]	[0.27]	[0.27]
		(0.33)		(0.40)	
	0.5	0.29	0.26	0.79	0.77
		[0.15]	[0.11]	[0.29]	[0.27]
		(0.28)		(0.31)	
	0.8	0.28	0.24	0.78	0.75
		[0.15]	[0.12]	[0.33]	[0.31]
		(0.28)		(0.29)	
{tbill}	0.2	0.29	0.27	0.68	0.67
		[0.17]	[0.13]	[0.27]	[0.27]
		(0.38)		(0.43)	
	0.5	0.17	0.17	0.59	0.58
		[0.12]	[0.12]	[0.25]	[0.27]
		(0.46)		(0.40)	
	0.8	0.11	0.10	0.54	0.53
		[0.09]	[0.13]	[0.24]	[0.27]
		(0.35)		(0.34)	
{gold}	0.2	0.56	0.44	0.96	0.84
		[0.23]	[0.12]	[0.31]	[0.27]
		(0.14)		(0.13)	
	0.5	0.42	0.33	0.74	0.67
		[0.18]	[0.11]	[0.26]	[0.25]
		(0.17)		(0.20)	
	0.8	0.15	0.08	0.55	0.49
		[0.13]	[0.13]	[0.25]	[0.26]
		(0.20)		(0.23)	
{realvol}	0.2	0.27	0.24	0.97	0.87
		[0.16]	[0.12]	[0.35]	[0.25]
		(0.29)		(0.19)	
	0.5	0.16	0.14	0.90	0.81
		[0.11]	[0.13]	[0.31]	[0.25]
		(0.34)		(0.17)	
	0.8	0.14	0.12	0.66	0.65
		[0.13]	[0.17]	[0.28]	[0.30]
		(0.32)		(0.35)	
{cpi}	0.2	0.41	0.34	0.87	0.82
		[0.17]	[0.11]	[0.30]	[0.27]
		(0.20)		(0.20)	
	0.5	0.30	0.27	0.78	0.75
		[0.15]	[0.11]	[0.28]	[0.26]
		(0.28)		(0.28)	
	0.8	0.16	0.12	0.65	0.63
		[0.15]	[0.16]	[0.28]	[0.27]
		(0.23)		(0.32)	

Table 2.12: Increase in Certainty Equivalent Returns from Conditioning on a Predictor

This table reports the difference between unconditional CER of optimal policies conditioned on individual predictor variables minus the CER of the optimal unconditional strategy. The standard deviation of CER differences is given in parentheses. P-values reflect the probability of rejecting the null hypothesis that the unconditional certainty equivalent of following the conditional strategy leads to lower certainty equivalent return than that achieved by following the optimal unconditional strategy. Standard deviations and p-values are calculated based on 1000 stationary bootstrap samples.

	{equities,bonds}	{equities,bonds,HML}
div	0.10 [0.05] (0.018)	0.11 [0.06] (0.022)
def	0.06 [0.05] (0.077)	0.11 [0.05] (0.016)
term	0.34 [0.09] (0.000)	0.36 [0.10] (0.000)
trend	0.21 [0.07] (0.000)	0.20 [0.08] (0.000)
tbill	0.09 [0.05] (0.024)	0.14 [0.08] (0.016)
gold	0.49 [0.12] (0.000)	0.55 [0.14] (0.000)
realvol	0.16 [0.06] (0.002)	0.29 [0.10] (0.001)
cpi	0.19 [0.08] (0.004)	0.22 [0.09] (0.002)

Table 2.13: Coefficients of Predictive Index

This table displays estimates of optimal index coefficients for different combinations of predictor variables. Each row is labeled with abbreviations for the predictor variables included in each index. Abbreviations are defined in Table 7. The index coefficients, $\{\beta_i\}$, are estimated for a power utility investor with risk aversion equal to five. Index coefficients are listed for two asset sets: equities and bonds; and equities, bonds and HML. The fifth column for each asset set shows in-sample estimates of CERs achieved by following the optimal portfolio conditioned on the estimated index.

			{equities, bonds}				{equities, bonds, HML}						
			β_1	β_2	β_3	β_4	CE	β_1	β_2	β_3	β_4	CER	
div	term	trend	gold	0.08	0.46	-0.15	-0.31	0.69	0.23	0.38	-0.08	-0.31	1.25
def	term	gold	cpi	0.23	0.38	-0.38	0.00	0.68	0.08	0.38	-0.38	0.15	1.24
term	trend	gold	cpi	0.38	-0.23	-0.38	0.00	0.69	0.38	-0.08	-0.38	0.15	1.24
div	term	tbill	gold	0.08	0.46	-0.00	-0.46	0.69	0.23	0.46	-0.00	-0.31	1.23
div	def	term	gold	0.15	0.15	0.38	-0.31	0.69	0.23	-0.00	0.46	-0.31	1.23
def	term	trend	gold	0.15	0.38	-0.15	-0.31	0.70	0.15	0.38	-0.15	-0.31	1.23
term	trend	tbill	gold	-0.38	0.15	0.00	0.46	0.69	-0.31	0.08	-0.15	0.46	1.21
term	gold			0.54	-0.46			0.68	-0.24	0.76			1.21
div	def	term	trend	0.08	0.08	0.69	-0.15	0.51	-0.08	0.15	0.46	-0.31	0.99

Table 2.14: Conditional Certainty Equivalents for Predictive Index

This table shows certainty equivalent returns (CERs) of portfolio strategies conditioned on predictor indices. Each index is a linear combination of up to four predictor values. The index coefficients are given in Table 12. The optimal strategies are estimated for a power utility investor with risk aversion equal to five and a one-month investment horizon. CERs are shown for two assets sets: equities and bonds; and equities, bonds and HML. The equities and bond assets are hedge funds that are long the appropriate index and short the risk-free asset. Each row shows conditional CER calculated as kernel averages of utility outcomes at different (standardized) values of predictor variables. Standard errors are calculated by the delta method and given in square brackets. Two columns are shown for each set of assets. The first is the conditional CER of an optimal conditional strategy. The second is the conditional CER of the optimal unconditional strategy. Bootstrap p-values are given in parentheses. The p-values are based on the null hypothesis that the unconditional strategy is as efficient in terms of CER as the conditional strategy.

		{equities,bonds}		{equities,bonds,HML}	
		<i>CER</i>	<i>CER_{unc}</i>	<i>CER</i>	<i>CER_{unc}</i>
{div,term,trend,gold}	0.2	0.11	-0.20	0.31	-0.06
		[0.20]	[0.25]	[0.31]	[0.46]
		(0.05)		(0.04)	
	0.5	0.09	0.06	0.30	0.17
		[0.15]	[0.23]	[0.27]	[0.43]
		(0.28)		(0.17)	
	0.8	0.74	0.44	1.51	1.20
		[0.56]	[0.27]	[0.92]	[0.53]
		(0.06)		(0.07)	
{term,gold,vs,cpi}	0.2	0.09	-0.16	1.14	0.92
		[0.18]	[0.25]	[0.66]	[0.47]
		(0.07)		(0.08)	
	0.5	0.08	0.05	0.72	0.68
		[0.14]	[0.23]	[0.46]	[0.43]
		(0.31)		(0.27)	
	0.8	0.72	0.31	0.47	-0.02
		[0.58]	[0.27]	[0.51]	[0.52]
		(0.03)		(0.02)	

Table 2.14: (Continued)

		{equities,bonds}		{equities,bonds,HML}	
		<i>CER</i>	<i>CER_{unc}</i>	<i>CER</i>	<i>CER_{unc}</i>
{def,term,gold,cpi}	0.2	0.22	-0.22	0.32	-0.13
		[0.31]	[0.27]	[0.35]	[0.49]
		(0.02)		(0.03)	
	0.5	0.09	0.06	0.35	0.27
		[0.15]	[0.23]	[0.30]	[0.43]
		(0.26)		(0.23)	
	0.8	0.78	0.48	1.40	1.13
		[0.58]	[0.27]	[0.82]	[0.51]
		(0.07)		(0.08)	
{term,trend,gold,cpi}	0.2	0.10	-0.23	0.31	-0.09
		[0.20]	[0.26]	[0.35]	[0.48]
		(0.06)		(0.03)	
	0.5	0.10	0.08	0.36	0.29
		[0.16]	[0.23]	[0.30]	[0.43]
		(0.31)		(0.24)	
	0.8	0.70	0.40	1.43	1.13
		[0.52]	[0.26]	[0.82]	[0.51]
		(0.07)		(0.07)	
{term,gold,,}	0.2	0.08	-0.02	0.92	0.84
		[0.15]	[0.24]	[0.57]	[0.44]
		(0.15)		(0.19)	
	0.5	0.28	0.24	0.45	0.40
		[0.28]	[0.23]	[0.35]	[0.43]
		(0.30)		(0.29)	
	0.8	1.33	0.64	0.77	-0.24
		[1.02]	[0.36]	[1.13]	[1.03]
		(0.02)		(0.04)	
{div,def,term,trend}	0.2	0.03	-0.16	0.41	0.23
		[0.10]	[0.25]	[0.37]	[0.46]
		(0.13)		(0.11)	
	0.5	0.16	0.14	0.46	0.41
		[0.21]	[0.24]	[0.36]	[0.43]
		(0.33)		(0.28)	
	0.8	0.53	0.41	1.28	1.12
		[0.50]	[0.29]	[0.81]	[0.55]
		(0.20)		(0.17)	

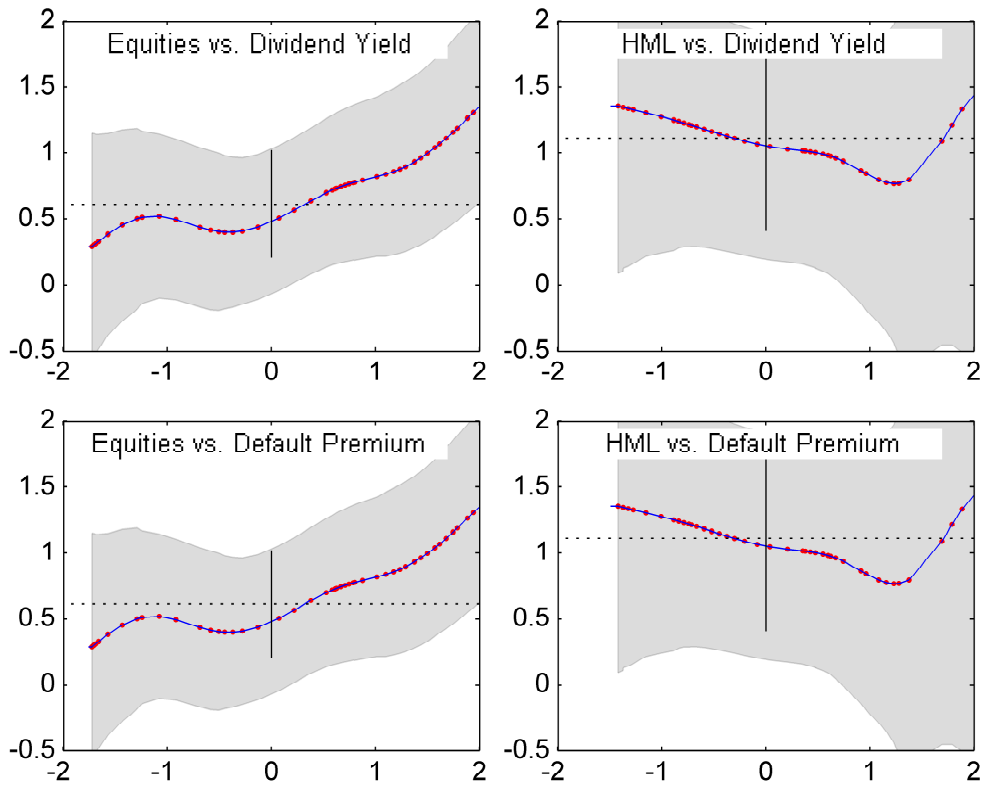


Figure 2.1: Single Asset Portfolios Conditioned on Dividend Yield and Default Premium

Figures 2.1-3 depict conditional, single risky asset allocations as functions of predictor variables. Each row of panels corresponds to a predictor variable. Column one shows index allocation for an equities and cash investor. Column two shows allocations to HML for an investor who otherwise only holds cash. The shaded regions show plus or minus two standard deviation regions. The dotted lines correspond to the unconditional policy. Predictor values are standardized to have mean zero and standard deviation 1.

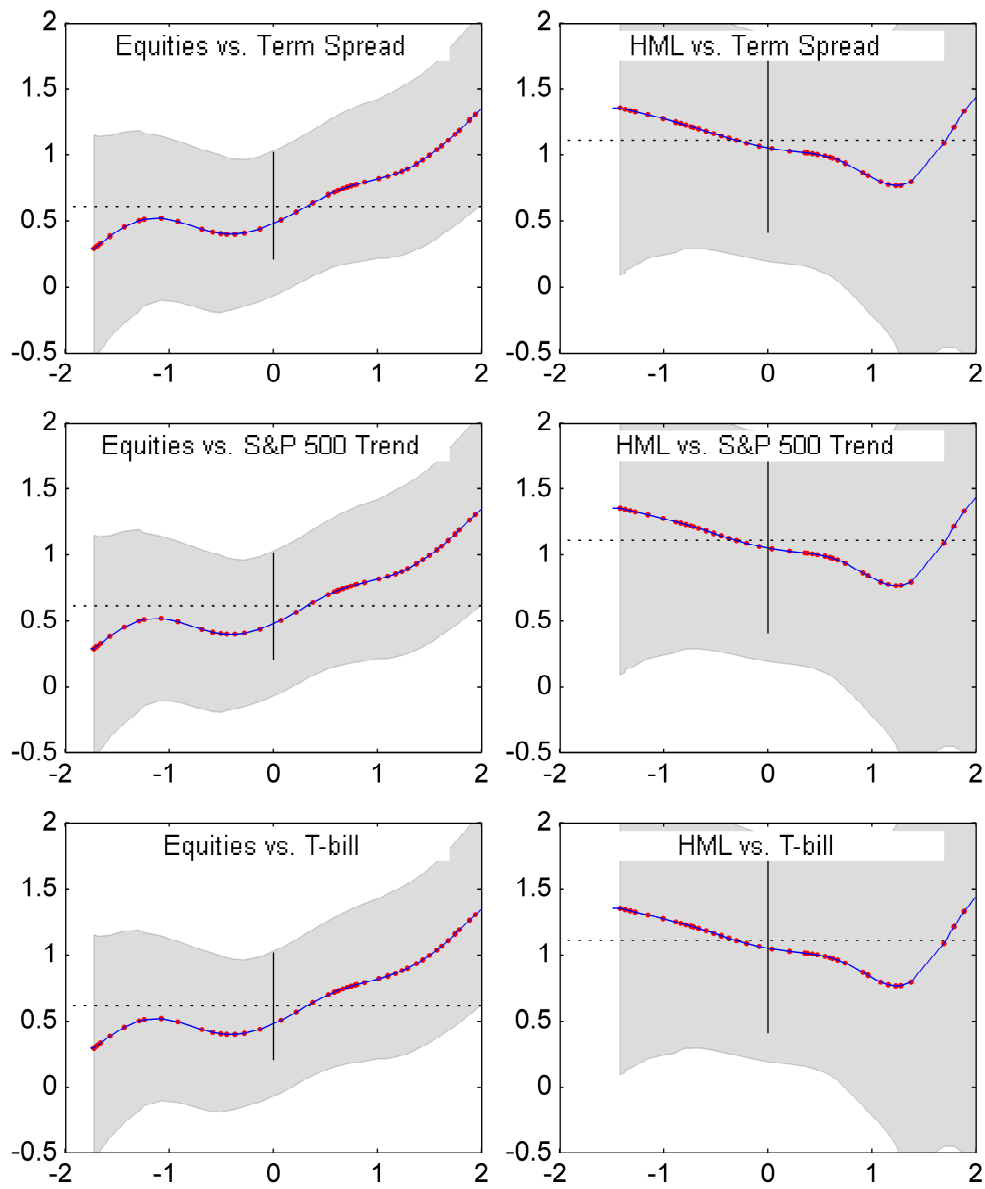


Figure 2.2: Single Asset Portfolios Conditioned on Term Spread, S&P 500 Trend, and T-bill Rate

Conditional, single risky asset allocations as a function of predictor variables. See caption to Figure 2.1 for details.

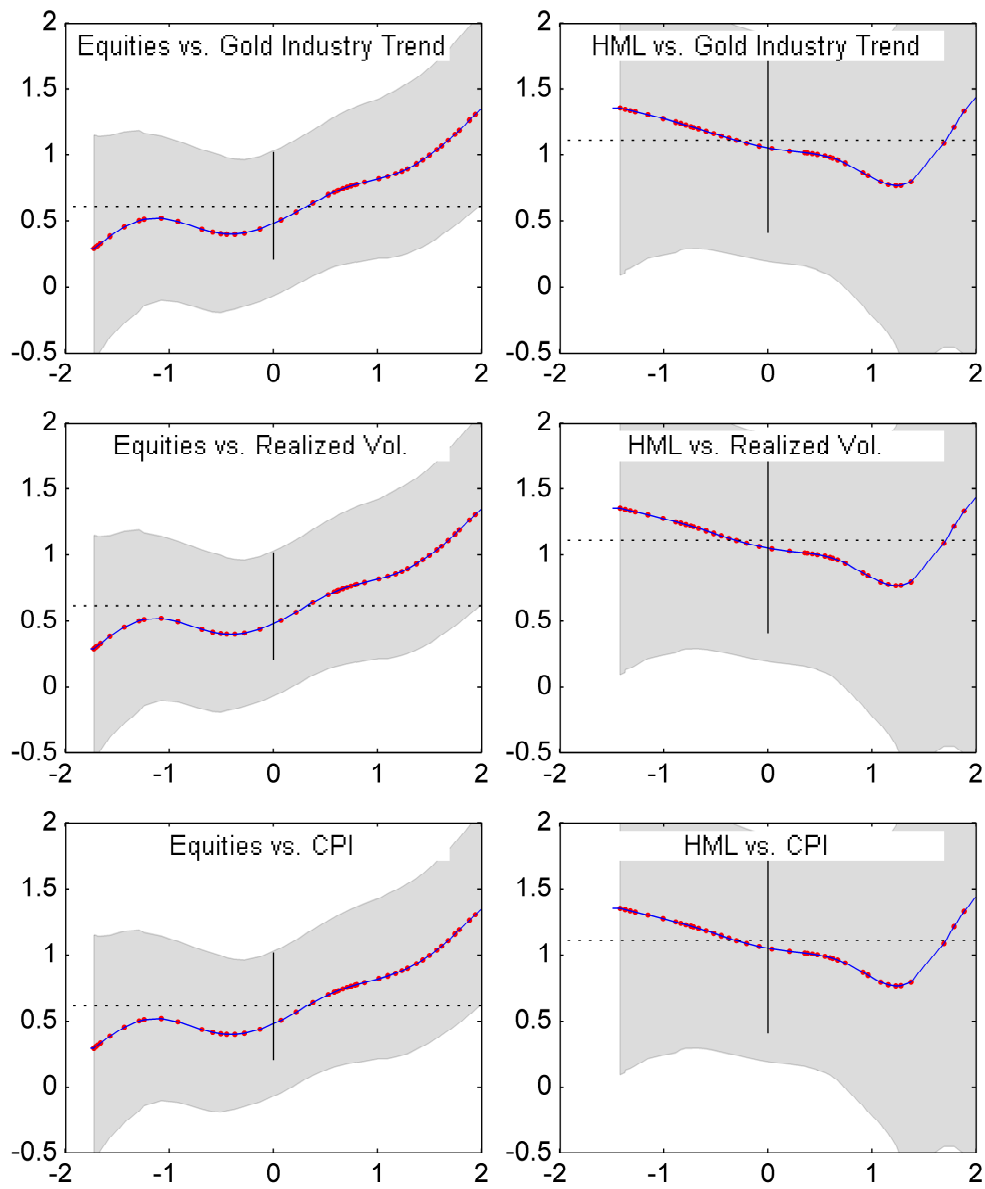


Figure 2.3: Single Asset Portfolios Conditioned on Gold Industry Trend, Realized Volatility, and Inflation

Conditional, single risky asset allocations as functions of predictor variables. See caption to Figure 2.1 for details.

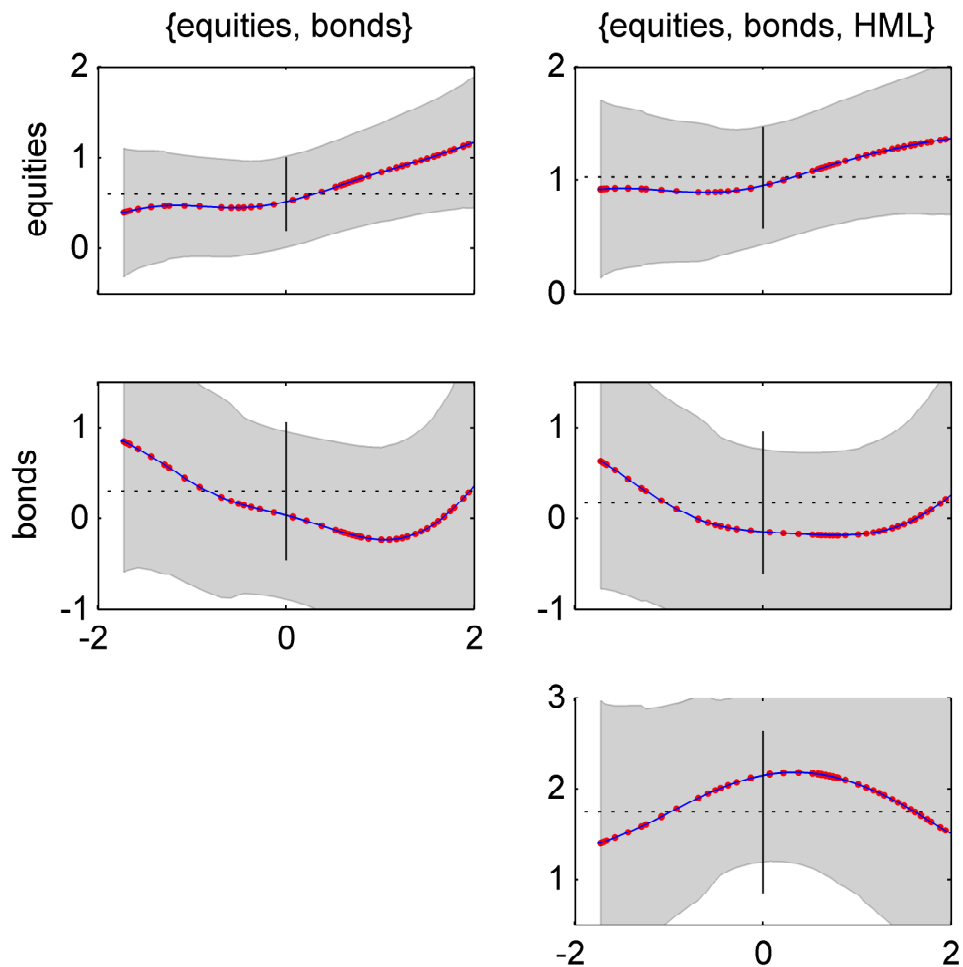


Figure 2.4: Portfolio Allocations versus Dividend Yield

Figures 2.4-11 depict portfolio allocations versus predictor value. Predictor values are standardized to have mean zero and standard deviation one. Each figure shows results for a different predictor. Figure 1 shows results for the dividend yield. Each column shows results for a different set of assets. The first column plots results for an investor who trades in the value weighted CRSP index as well as the CRSP bond index. The second column plots allocations for an investor who also trades in the HML portfolio. The conditional allocations are plotted for risk aversion equal to 5. The horizontal dotted lines in each plot depict unconditional allocations. The solid vertical line segment on each plot marks the two standard error band for the unconditional weights. The lower

left panel of each figure shows a histogram of predictor variable occurrences.

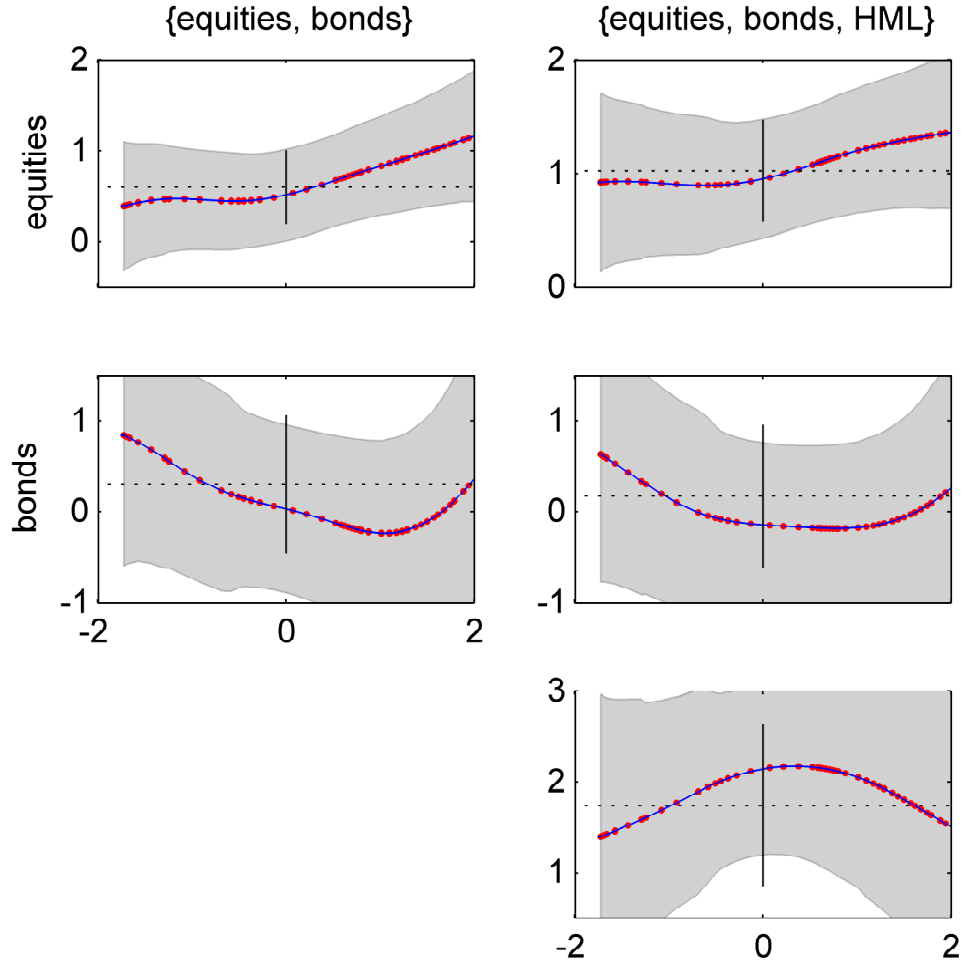


Figure 2.5: Portfolio Allocations versus Default Premium

This figure plots conditional portfolio allocations versus default premium standardized to have mean zero and standard deviation one. See Figure 2.4 for complete description.

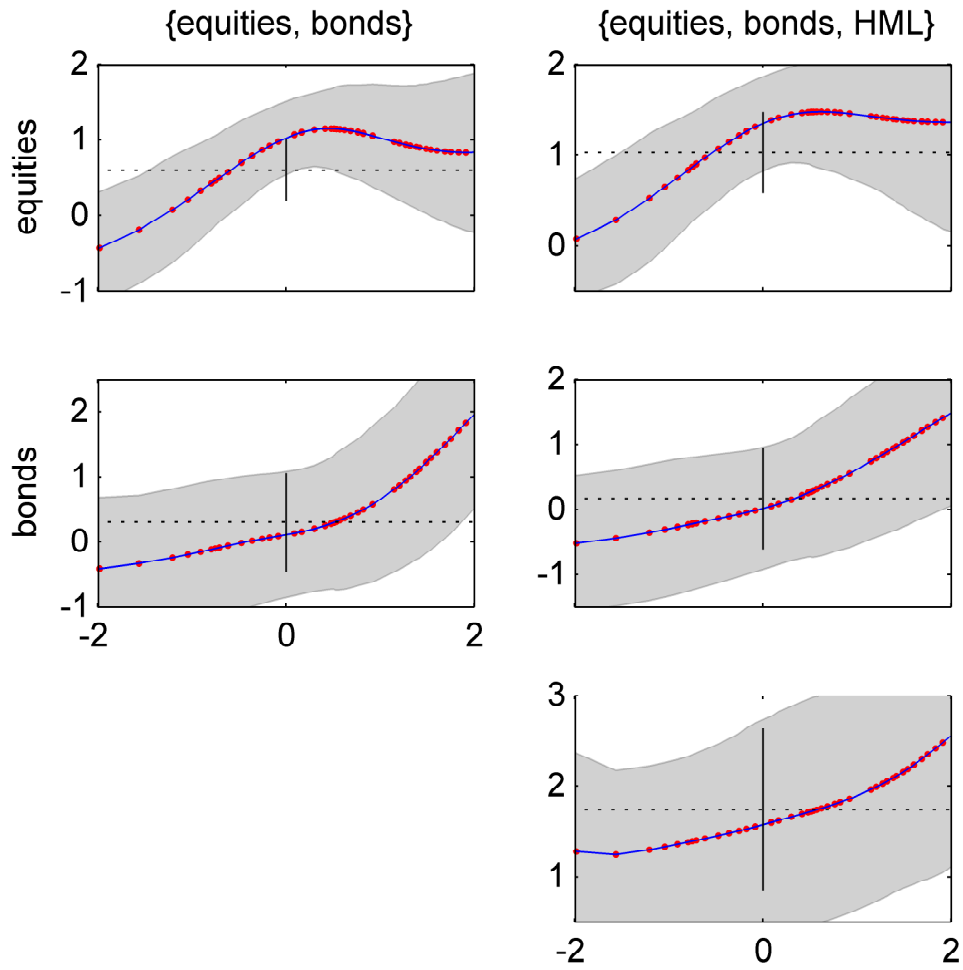


Figure 2.6: Portfolio Allocations versus Term Spread

This figure plots conditional portfolio allocations versus term spread standardized to have mean zero and standard deviation one. See Figure 2.4 for complete description.

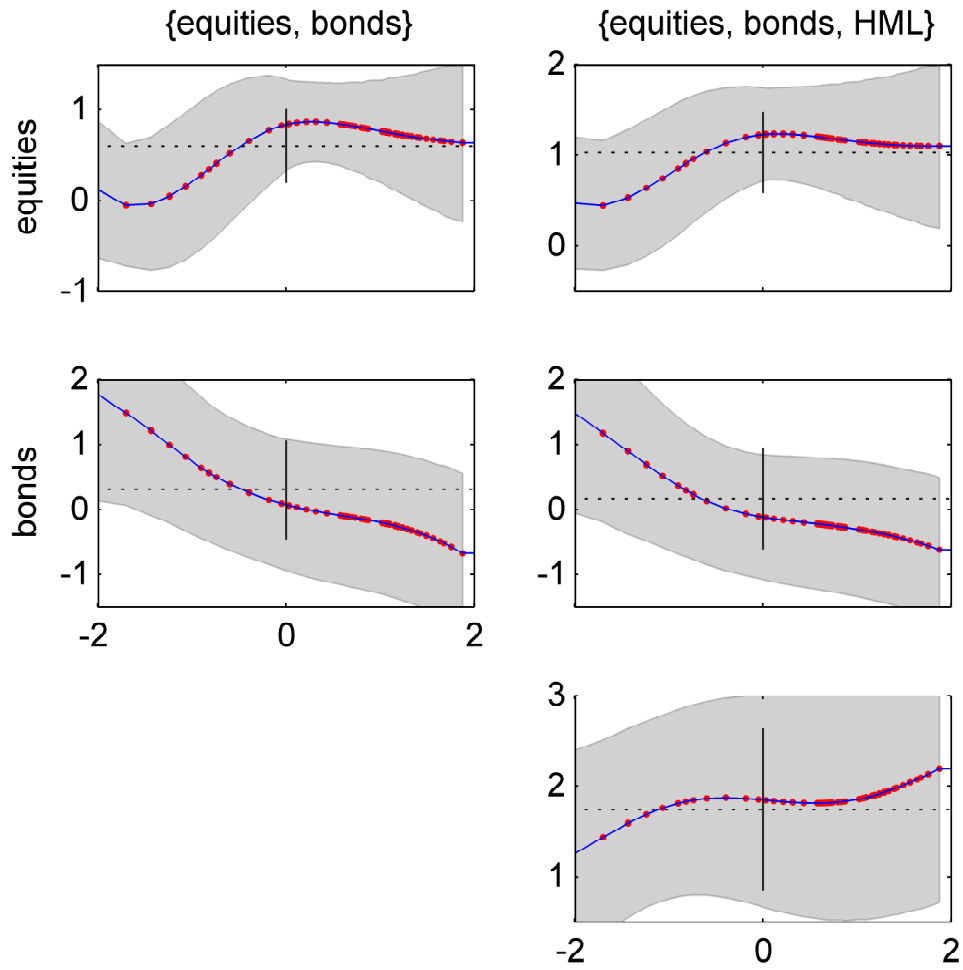


Figure 2.7: Portfolio Allocations versus S&P 500 Trend

This figure plots conditional portfolio allocations versus S&P 500 trend standardized to have mean zero and standard deviation 1. See figure 2.4 for complete description.

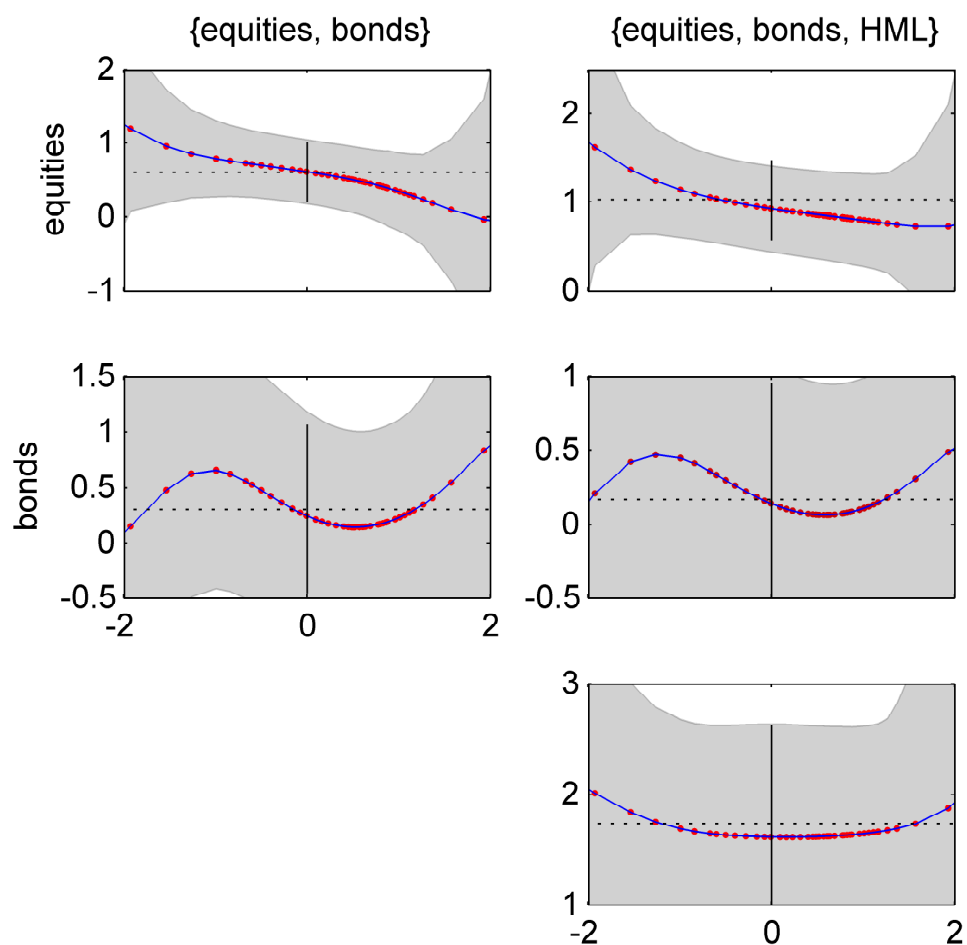


Figure 2.8: Portfolio Allocations versus T-bill Rate

This figure plots conditional portfolio allocations versus 3 month T-bill rate standardized to have mean zero and standard deviation one. See Figure 2.4 for complete description.

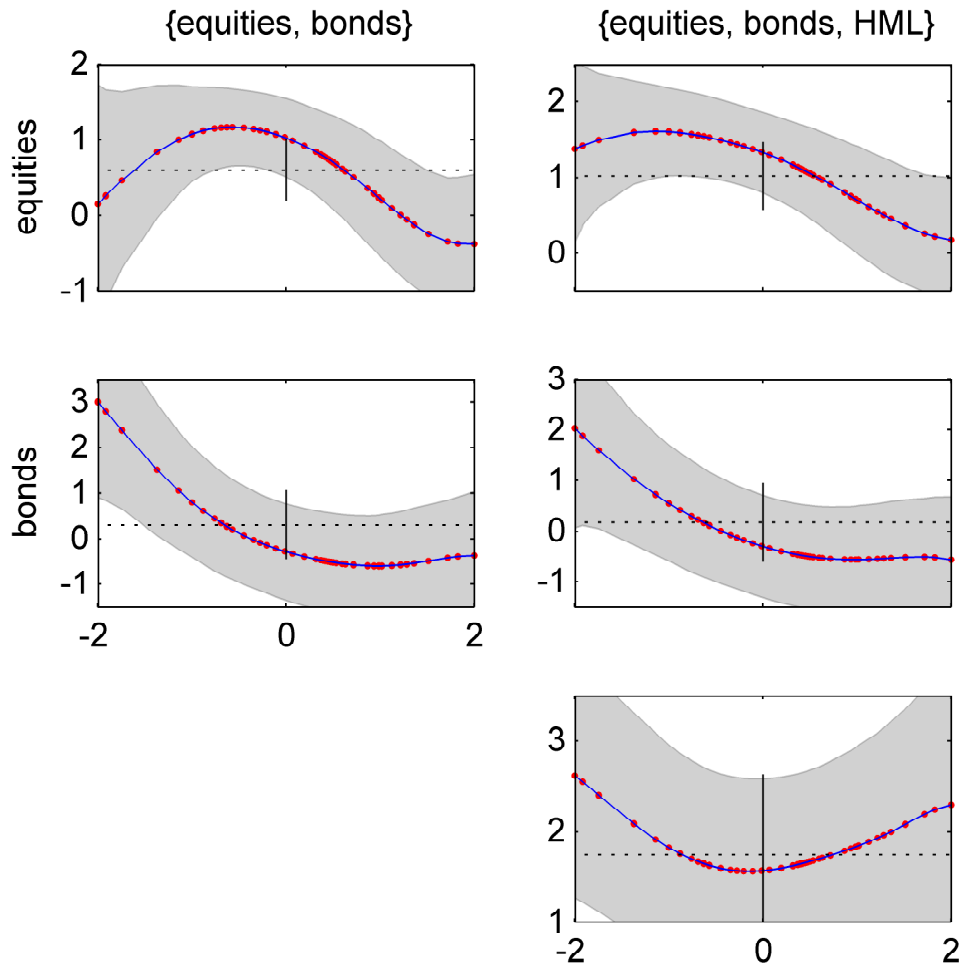


Figure 2.9: Portfolio Allocations versus Gold Industry Trend

This figure plots conditional portfolio allocations versus gold industry trend standardized to have mean zero and standard deviation one. See Figure 2.4 for complete description.

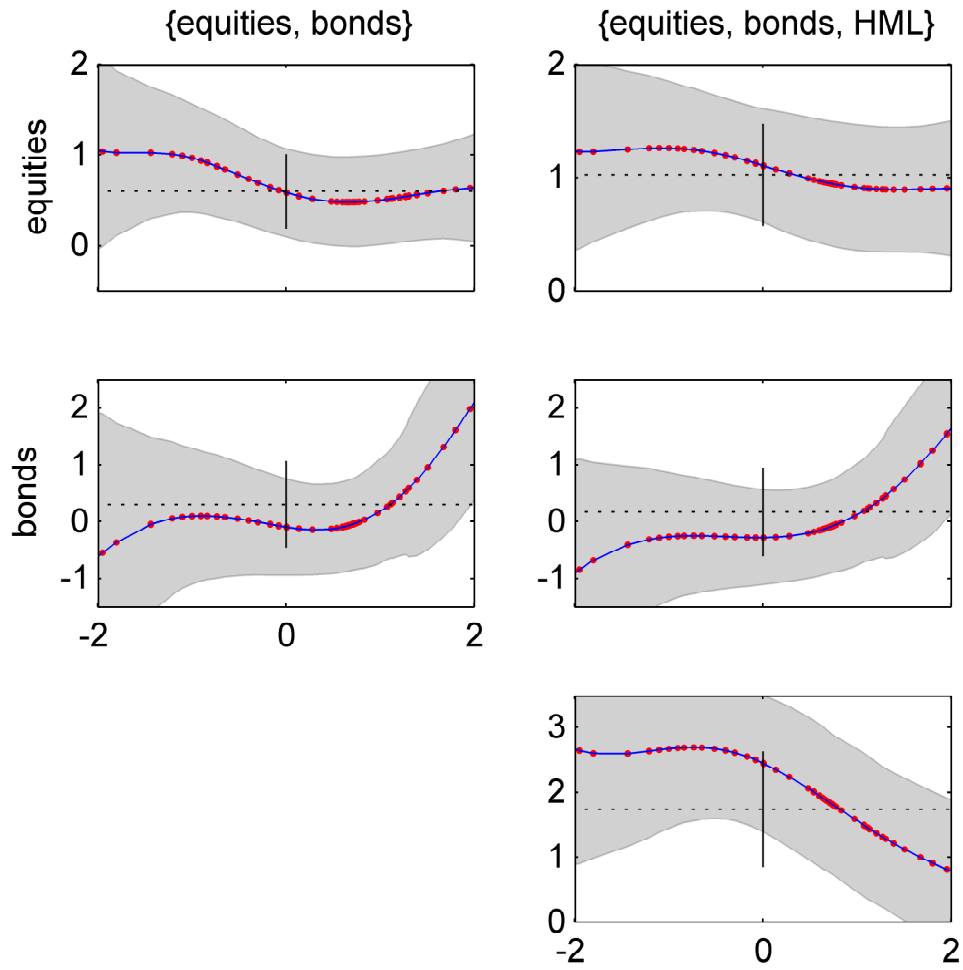


Figure 2.10: Portfolio Allocations versus Realized Volatility

This figure plots conditional portfolio allocations versus realized volatility standardized to have mean zero and standard deviation one. See Figure 2.4 for complete description.

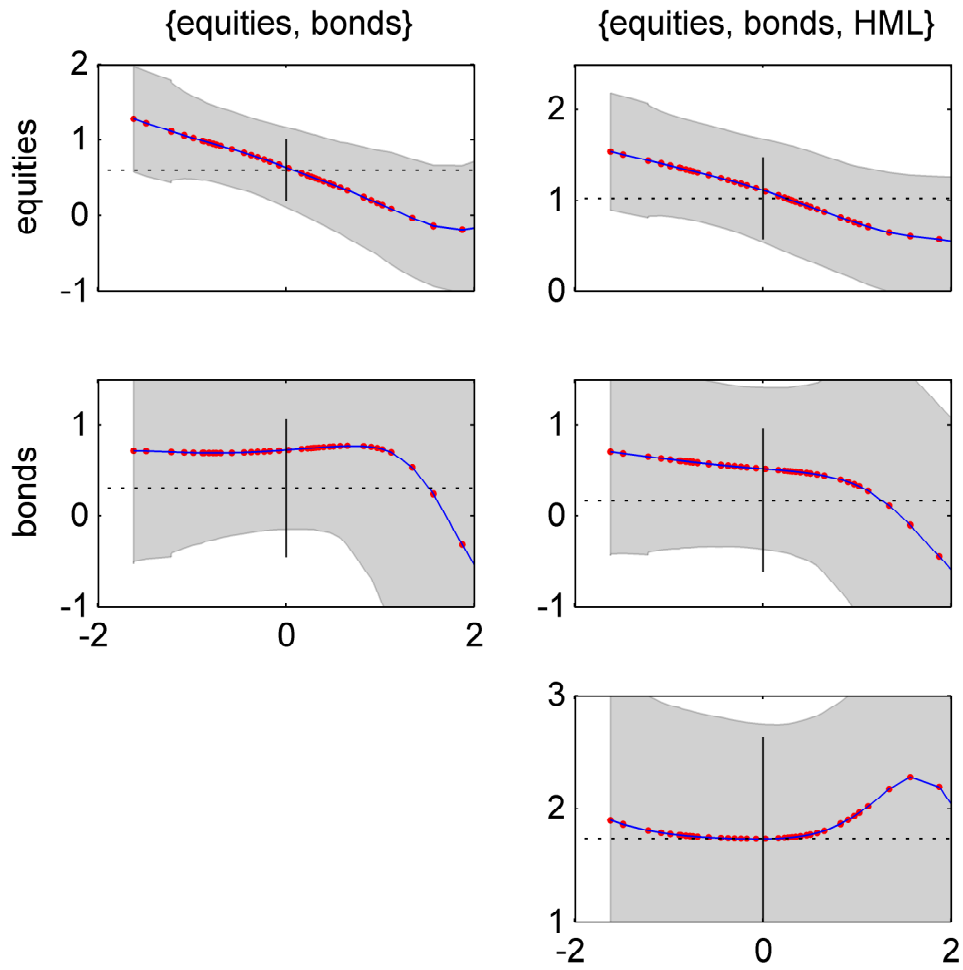


Figure 2.11: Portfolio Allocations versus Inflation

This figure plots conditional portfolio allocations versus inflation standardized to have mean zero and standard deviation one. See Figure 2.4 for complete description.

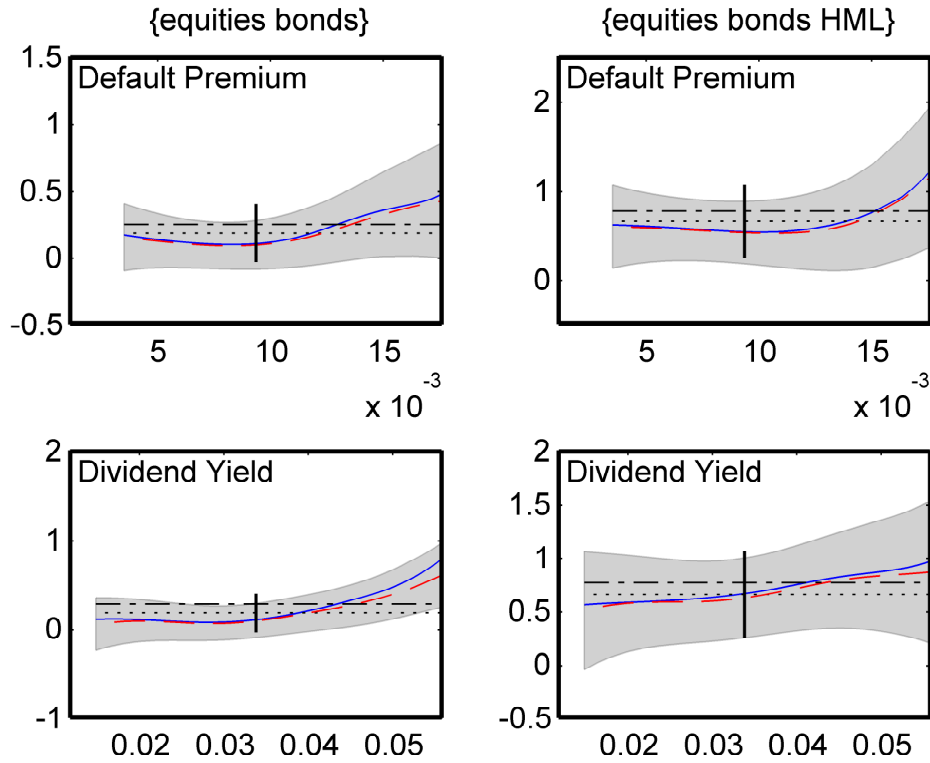


Figure 2.12: CER versus Default Premium and Dividend Yield

Conditional certainty equivalent returns (CERs) as a function of predictor value. The solid curve is the CER for an investor who allocates to the optimal conditional portfolio. The dashed curve is the conditional CER of an investor who allocates to the optimal unconditional portfolio. The dash-dot horizontal line is the unconditional CER attained by following the conditional portfolio. The dotted horizontal line is the unconditional CER of an investor who allocates to the unconditional portfolio. The shaded region is an approximate one standard error region for the conditional CER of the conditional portfolio policy (calculated by the delta method). The vertical segment indicates \pm one standard error above and below the unconditional CER of the unconditional policy.

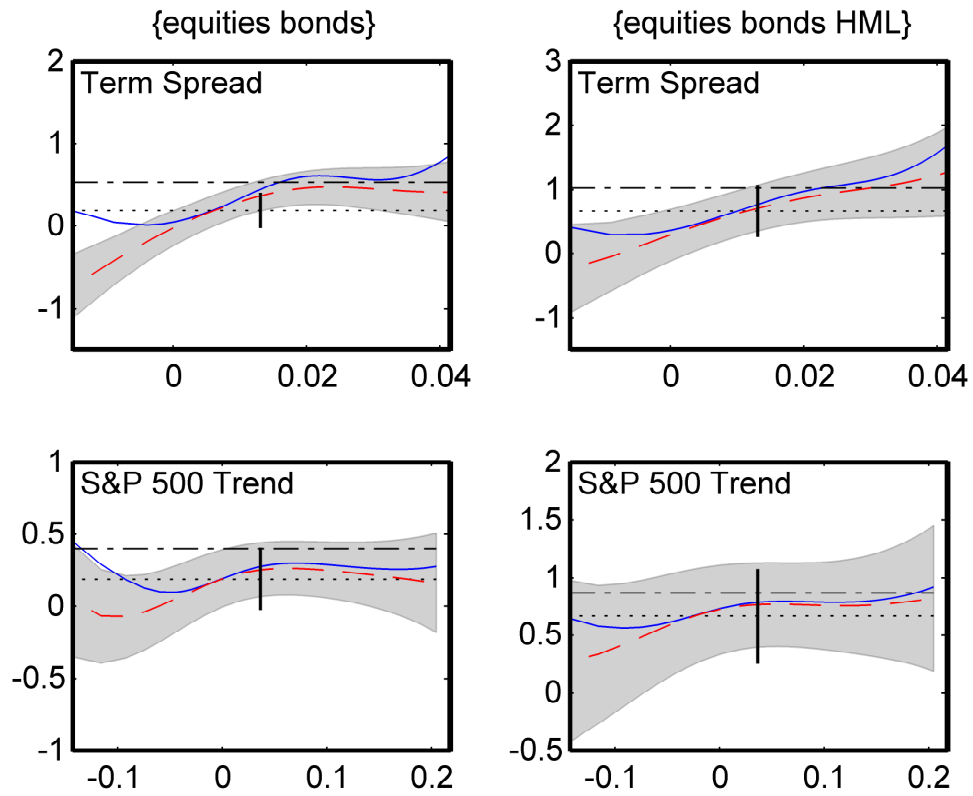


Figure 2.13: CER versus Term Spread and Index Trend

Conditional certainty equivalent returns (CERs) as a function of predictor value. See caption for figure 2.12 for details.

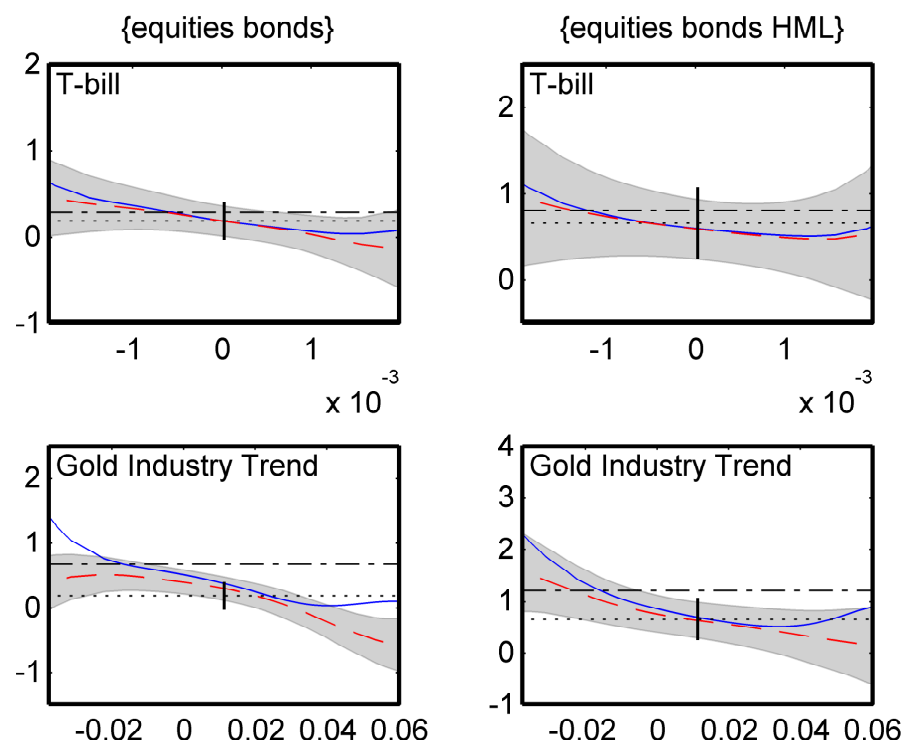


Figure 2.14: CER versus Tbill Yield and Gold Industry Trend

Conditional certainty equivalent returns (CERs) as a function of predictor value. See caption for figure 2.12 for details.

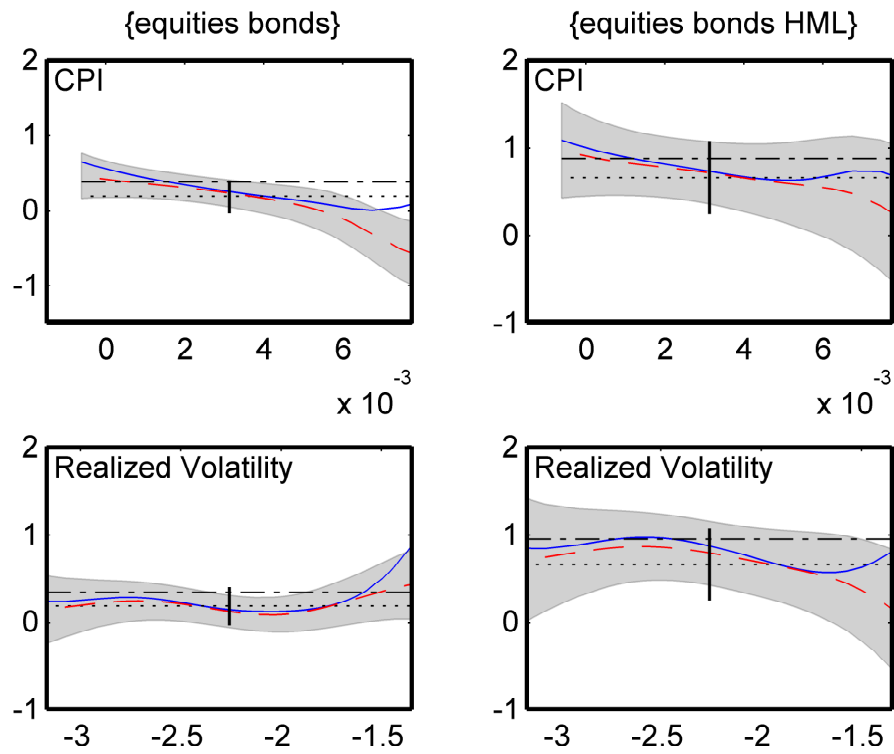


Figure 2.15: CER versus Inflation and Volatility

Conditional certainty equivalent returns (CERs) as a function of predictor value. See caption for figure 2.12 for details.

Bibliography

Aït-Sahalia, Y. and Brandt, M. (2001). Variable selection for portfolio choice. *Journal of Finance*, 56:1297–1351.

Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial Studies*, 20(3):651–707.

Arrow, K. J. (1971). *Essays in the Theory of Risk Bearing*. Markham Publishing, Chicago.

Avramov, D. (2004). Stock return predictability and asset pricing models. *Review of Financial Studies*, 17(3):699–738.

Baker, M. and Wurgler, J. (2006). Investor sentiment and the cross section of expected stock returns. *Journal of Finance*, 61(4):1645–1680.

Balduzzi, P. and Lynch, A. W. (1999). Transaction costs and predictability: Some utility cost calculations. *Journal of Financial Economics*, 52:47–78.

Barberis, N. (2000). Investing for the long run when returns are predictable. *Journal of Finance*, 55:225–264.

Bossaerts, P. and Hillion, P. (1999). Implementing statistical criteria to select return forecasting models. *Review of Financial Studies*, 12(2):405–428.

Brandt, M. (1999). Estimating portfolio and consumption choice: A conditional Euler equations approach. *Journal of Finance*, 54:1609–1645.

Brandt, M., Santa-Clara, P., and Valkanov, R. (2005). Parametric portfolio policies: Exploiting characteristics in the cross section of equity returns. Working Paper.

Brennan, M. J., Schwarz, E. S., and Lagnado, R. (1997). Strategic asset allocation. *Journal of Economic Dynamics and Control*, 21:1377–1403.

- Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics*, 18:373–399.
- Campbell, J. Y. (1991). A variance decomposition for stock returns. *Economic Journal*, 101(405):157–179.
- Campbell, J. Y., Chan, Y. L., and Viceira, L. M. (2003). A multivariate model of strategic asset allocation. *Journal of Financial Economics*, 67:41–80.
- Campbell, J. Y. and Viceira, L. (1999). Consumption and portfolio decisions when expected returns are time varying. *Quarterly Journal of Economics*, 114:433–495.
- Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81:27–60.
- Carhart, M. M. (1997). On the persistence of mutual fund performance. *Journal of Finance*, 52:57–82.
- Cochrane, J. H. (1999). Portfolio advice in a multifactor world. *Economic Perspectives*, 23(3):59–78.
- Connor, G. and Korajczyk, R. A. (1993). A test for the number of factors in an approximate factor model. *Journal of Finance*, 48:1263–1291.
- DeMiguel, V., Garlappi, L., and Uppal, R. (2007). Optimal versus naive diversification: How inefficient is the $1/n$ portfolio strategy? *Review of Financial Studies*, Forthcoming.
- DeMiguel, V. and Nogales, F. (2007). Portfolio selection with robust estimation.
- Dudoit, S. and van der Laan, M. (2007). *Multiple Hypothesis Testing with Applications to Genomics*. Springer.
- Fama, E. and French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33:3–56.
- Fama, E. and French, K. (1996). Multifactor efficiency and multifactor asset pricing. *Journal of Financial and Quantitative Analysis*, 31:441–465.

- Ferson, W. E. and Siegel, A. F. (2000). Testing portfolio efficiency with conditioning information. Working Paper.
- Goetzmann, W. N. and Jorion, P. (1993). Testing the predictive power of dividend yields. *Journal of Finance*, 48(2):663–679.
- Hansen, L. P. and Singleton, K. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica*, 50:1269–1286.
- Hardle, W., Muller, M., Sperlich, S., and Werwatz, A. (2004). *Nonparametric and Semiparametric: Methods: An Introduction*. Springer.
- Harvey, C. R., Liechty, J. C., Liechty, M. W., and Muller, P. (2003). Portfolio selection with higher moments. Working Paper.
- Hodrick, R. J. (1992). Dividend yields and expected stock returns: Alternative procedures for inference. *Review of Financial Studies*, 5(3):357–386.
- Jobson, J. D. and Korkie, B. M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance*, 36(4):889–908.
- Jobson, J. D. and Korkie, R. (1980). Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association*, 75:544–554.
- Jurek, J. and Viceira, L. (2005). Optimal value and growth tilts in long-horizon portfolios. Working Paper.
- Kacperczyk, M. (2003). Asset allocation under distribution uncertainty. Working Paper.
- Kandel, S. and Stambaugh, R. F. (1996). On the predictability of stock returns: An asset allocation perspective. *Journal of Finance*, 51:385–424.
- Lettau, M. and Ludvigson, S. (2001). consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, 56(3):815–849.
- Makarov, I. and Papanikolaou, D. (2008). Sources of systematic risk. Working Paper.

- Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7:77–91.
- Mehra, R. and Prescott, E. C. (1985). The equity premium: a puzzle. *Journal of Monetary Economics*, 15:145–161.
- Memmel, C. (2003). Performance hypothesis testing with the sharpe ratio. *Finance Letters*, 1:21–23.
- Pástor, Ľ. and Stambaugh, R. F. (2000). Comparing asset-pricing models: An investment perspective. *Journal of Financial Economics*, 56:335–381.
- Paye, B. (2004). *Essays on Stock Return Predictability and Portfolio Allocation*. PhD thesis, University of California San Diego.
- Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313.
- Press, S. J. and Zellner, A. (1978). Posterior distribution for the multiple correlation coefficient with fixed regressors. *Journal of Econometrics*, 8:307–321.
- Rey, D. (2004). Stock market predictability: Is it there? A critical review. Working Paper.
- Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, 54:375–421.
- Wachter, J. A. and Warusawitharana, M. (2005). Predictable returns and asset allocation: Should a skeptical investor time the market? Working Paper.
- White, H. (2000). A reality check for data snooping. *Econometrica*, 68:1097–1126.
- Wolf, M. (2007). Robust performance hypothesis testing with the sharpe ratio.

Chapter 3

Nonparametric Portfolio Estimation with Prior Belief in the Positivity of Portfolio Weights⁴³

3.1 Introduction

Short-sales constraints are a common technique for stabilizing portfolio estimates based on historical returns. The imposition of short sales constraints has the effect of penalizing variability in portfolio weights. A disadvantage of this approach is that it results in a biased portfolio estimator. The reward is a more robust estimate that performs better out of sample. However, what if historical returns do not capture all information available to the investor? Any additional information would decrease the variance of portfolio estimates and, hence, decrease the need to bias the portfolio estimator by explicitly constraining the weights vector.

A source of information beyond that provided by historical returns is economic intuition. If investors have beliefs about the properties of portfolio weights, then those beliefs should be incorporated into the investor's information set. One candidate belief can be imputed from the very popularity of short-sales constraint. Given that many assets can be readily shorted either directly or indirectly via exchange traded funds or derivative trades, it seems plausible that this constraint is more reflective of investor bias or beliefs. Furthermore, for many asset allocation problems, the positivity of portfolio allocations is supported by equilibrium asset pricing theory. For example, an investor choosing an allo-

⁴³A version of this chapter will be submitted for publication. Douglass, J., Nonparametric Portfolio Estimation with Prior Belief in the Positivity of Portfolio Weights.

cation across industry portfolios might reflect that all industries have positive market weight and therefore ought to be priced accordingly at equilibrium.

In this chapter, I investigate the impact of conditioning portfolio estimates on a belief in the nonnegativity of portfolio weights. I use Bayesian techniques to construct a nonparametric estimator that can incorporate qualitative investor beliefs. Then, in a simulation study, I examine the properties of estimated weights and evaluate the performance of a portfolio rule based on the conditional estimator.

I construct a conditional nonparametric method-of-moments estimator that incorporates qualitative prior beliefs. The final form of the estimator is analogous to the conditional estimator developed by Brandt (1999).⁴⁴ The difference is that Brandt (1999) conditions portfolio estimates on predictor variables rather than qualitative beliefs. Both estimators approximate the expectation integral in the investor's problem with a probability weighted sum over historical return outcomes. The weights can be viewed as a perturbation of the empirical distribution of returns that leads to the most likely predictive distribution given the conditioning information. The advantages of a nonparametric framework for portfolio choice are discussed by Brandt (1999) and Aït-Sahalia and Brandt (2001). Briefly, the nonparametric framework does not require the assumption of a parametric model of the underlying distribution thereby eliminating a potential source of specification error.

The conditional estimator introduced here is also related to data tuning applications developed by Peter Hall et. al.⁴⁵ Data tuning methods seek to minimally perturb the data set until a nonparametrically estimated quantity satisfies a set of restrictions or constraints. The data set can be perturbed by shifting the data points themselves or by perturbing the relative probability of each data point. These approaches impose prior views on the estimator using constraints. The estimator developed in this paper is akin to data tuning approaches that perturb probabilities. The objective is to obtain a data set that accommodates prior views with a minimal shift of these probabilities. Rather than imposing constraints, I use Bayesian techniques to obtain a discrete predictive distribution that reflects investor priors. The approach allows

⁴⁴Brandt's (1999) approach is applied by Aït-Sahalia and Brandt (2001), Paye (2004) and the author (see Chapter 2).

⁴⁵See, for example, Choi and Hall (1999), Choi et al. (2000) and Braun and Hall (2001).

for a probabilistic specification of prior beliefs. In addition, unlike data tuning algorithms, the resulting discrete-Bayesian estimator is consistent in the sense that the estimator converges in the large-sample limit even if the prior is wrong.

The goal of the estimator construction is to translate qualitative views into a discrete predictive distribution of returns. To accomplish this, I apply Bayes' rule on a discrete subset of the candidate probability distribution functions for returns. I start from a standard formulation of the investor's problem in a Bayesian statistical decision framework (Zellner and Chetty (1965)). In this framework, investor beliefs are modeled as a prior density over a space of candidate probability distributions. The investor combines their prior information with the data likelihood to obtain posterior probabilities for each candidate distribution. The predictive distribution for future returns is then obtained by integrating over the domain of the posterior distribution.

The domain of the posterior is a space of probability distribution functions. In a nonparametric setting, integration over a domain of functions is not computationally tractable without some specification assumptions. The standard approach is to restrict the candidate distributions to a parametric family such as the set of multivariate Gaussians. To retain the nonparametric character of the estimator, I instead restrict the domain of the posterior by discretizing the set of possible return outcomes. I restrict the domain of the posterior to the set of multinomial distributions with nonnegative probabilities on the chosen set.

The final step in the estimator design is the designation of the set of return outcomes. The set must be large enough to allow for a rich set of candidate distributions, yet parsimonious enough to permit sampling. A good candidate is the set of the return realizations in the historical data. This set is a random sample from the return domain that has the useful property of being generated by the data generating process of interest. As a result, this set will likely place a greater number of points where the density is concentrated and the data provides the most information. Furthermore, by discretizing to the return outcomes in the data, the resulting predictive distribution takes the form of a perturbation of the empirical distribution of returns.

One caveat is that the resulting discrete-Bayesian estimator is not strictly Bayesian. By using the data to approximate the posterior domain, I am implicitly restricting the prior domain as well. Despite this, the prior probabilities

attached to each candidate distribution do reflect prior beliefs, and these prior beliefs are effectively incorporated into the posterior.

In a simulation study, I investigate the implications of investor belief in the nonnegativity of portfolio weights on portfolio choice. I model prior views as exponential functions that decay as the conflict between investor views and candidate distributions increases. I examine the implications of the no short views on the predictive distribution of returns and on the portfolio weights.

I consider the example problem of an investor seeking to estimate allocations across a set of five assets that are highly correlated and that have similar distribution characteristics.⁴⁶ As in Kan and Zhou (2007), I use simulations to compare the expected out-of-sample performance of different models. In addition, I conduct out-of-sample experiments on historical data, and compare investment outcomes with a set of portfolio choice models from the literature. I set the parameters of the underlying distribution such that the true optimal allocations are all positive.

I find that, even when blessed with a prior belief in the positivity of portfolio weights that is true for the generating distribution, the expected out-of-sample performance of the estimates struggles to match the performance of the minimum variance and $1/N$ portfolios.

The weak performance of the Bayesian model with informative prior attests to the usefulness of regularization in portfolio forecasting. Regularization is a technique that places smoothness criteria on the vector of estimated weights. The $1/N$ solution is an extreme example of a regularized solution. The $1/N$ policy is optimal if an extremely large penalty on cross-sectional variation in portfolio weights.⁴⁷ For history lengths of up to two-hundred and forty months, the data does not contain enough information to estimate robustly portfolio weights. This is true even after imposing a prior that is both correct and leads to less variation in portfolio weights across assets.

This study adds to several strands of the portfolio choice literature. First, I build on work that has incorporated Bayesian analysis into the portfolio choice problem. The Bayesian framework not only allows for possible prior views, but has the additional benefit of incorporating the impact of distribution uncertainty

⁴⁶The distribution characteristics are based on the historical returns distribution of the five-industry breakdown of the United States equity market provided by Ken French.

⁴⁷Additionally, portfolio weights are assumed to sum to one (DeMiguel et al. (2008)).

on investor decisions. Numerous applications of Bayesian analysis in portfolio choice examine the influence of distribution uncertainty when an investor has uninformative prior beliefs.⁴⁸ Kandel and Stambaugh (1996), and Wachter and Warusawitharana (2005) examine the impact of degrees of belief in return predictability on portfolio decisions.⁴⁹ Pástor (2000) and Pástor and Stambaugh (1999) examine the implications of investor beliefs in an asset pricing model.

This chapter is closely related to that of Chevrier and McCulloch (2008) who also consider the impact of investor views on portfolio composition. This study differs with that of Chevrier and McCulloch (2008) in several significant respects. I employ a modeling approach that is nonparametric with respect to the underlying return distribution. I am able to construct a nonparametric posterior by introducing a simple restriction on the space of possible probability distributions. In contrast, Chevrier and McCulloch (2008) assume the return distribution is normal. While I do assume an investor whose utility depends only on the first two moments of return, I do not restrict the set of underlying distributions to the normal family. As such, my approach could be applied to models with different investor utility such as those considered by Aït-Sahalia and Brandt (2001). In addition, as well as reporting out-of-sample performance tests, I also conduct a simulation study which permits comparing expected out-of-sample performance of the model. The disadvantage of out-of-sample studies is that performance results are based on outcomes for a single sequence of returns. With a simulation experiment, the performance of a portfolio rule can be estimated to arbitrary precision given enough simulated return histories. Finally, I describe the potential for positive weight constraints to introduce upward bias in the predictive return distribution and introduce an adjustment to the prior that mitigates this issue.

This remainder of the chapter is divided into seven sections. In section 3.2, I formulate the portfolio choice problem in a Bayesian decision framework. In section 3.3, I construct a discrete estimator for the investor's optimand. In section 3.4, I quantify the prior beliefs that I incorporate into the investor's problem. Section 3.5 describes the relationship between the portfolio choice model developed in section 3.3 with models from literature. The empirical

⁴⁸See, for example, Brown (1976) and Bawa et al. (1979).

⁴⁹Avramov (2002) and Cremers (2002) use Bayesian methods to help with model selection given a set of predictors.

results are presented and discussed in sections 3.6 and 3.7. Section 3.8 concludes. In Appendix B, I formulate two data tuning approaches to portfolio choice. Appendix C details the Markov chain Monte Carlo (MCMC) algorithm used to resample from the posterior distribution.

3.2 Portfolio Choice Framework

In this section, I formulate the investor's problem as a utility maximization in a Bayesian decision framework. The Bayesian decision framework will be the starting point for development of the resampling estimator developed in section 3.3. The Bayesian formulation of the investor's decision problem has its roots in the work of Zellner and Chetty (1965). Other applications can be found in Brown (1976), Bawa et al. (1979) and Barberis (2000), amongst others.

Consider an investor who must select a portfolio from a universe of N risky assets plus a riskless asset. Define the N vector of risky asset returns r_t^r and the risk-free rate r_t^f from time $t - 1$ to time t . Excess returns on the risky assets are given by $r_t = r_t^r - r_t^f$.

The investor does not know the probability distribution of returns. Instead the investor is Bayesian, and bases portfolio decisions on an information set comprised of historical return data plus prior beliefs. The prior beliefs may be based on economic insights or other non-data knowledge. The prior beliefs and the non-data knowledge are combined to obtain a posterior distribution.

The domain of the posterior distribution is a function space that includes all probability distributions that could possibly describe the data set (prior to introducing any information). This is a very large set. For the portfolio problem, we are concerned with the future distribution of returns. The unrestricted set of potential future distributions includes all N -dimensional functions that integrate to one on \mathbb{R}^N . In applications, one must integrate over this set to evaluate the investor's expectation. This functional integral is an essential feature of Bayesian applications. The evaluation of functional integrals is challenging in applications. Unlike with classical integration of volumes⁵⁰, there is no universal fundamental definition for functional integrals. Thus, in typical applications the set of possible distributions is restricted to a subset that permits evaluation of

⁵⁰Volume is a generalized concept here defined on N dimensions.

a standard integral. I return to this issue and discuss the restriction used in this paper in the next section.

Let $\Omega \subset \mathbb{R}^N$ be the space of possible next period return outcomes $r_{\tau+1}$, and let Δ be the unrestricted space of possible probability distributions of returns. The investor uses his or her available information set to assign posterior probabilities to each possible probability distribution $\delta \in \Delta$. The posterior q_{post} is the product of a likelihood distribution given the available data and a prior distribution q_{prior} that quantifies investor views, i.e.,

$$q_{\text{post}}(\delta|R) \propto l(R|\delta)q_{\text{prior}}(\delta), \quad (3.1)$$

where R is a T by N matrix of historical return data, and the outer integral is a functional integral over the function space Δ .

I assume the investor's utility can be expressed as a von-Neumann-Morgenstern utility function that includes a parameter γ that increases in risk aversion. The investor's problem is to maximize expected utility given the available information.

$$\max_w E[u(W)] = \int_{\Delta} \int_{\Omega} u(r, w) p(r|\delta) dr q_{\text{post}}(\delta|X) d\delta, \quad (3.2)$$

where $p(r|\delta)$ is the probability of return r under probability distribution δ .

Introducing the notation E_{Δ} for the expectation integral over Δ and E_{δ} for the expectation integral over Ω for a given $\delta \in \Delta$, the investor's problem can be written as

$$\max_w E_q[u(r)] = \max_w E_{\Delta} \left[E_{\delta}[u(r)] \right], \quad (3.3)$$

where q is the predictive distribution of returns, i.e., $q(r) = \int_{\Delta} p(r|\delta)q_{\text{post}}(\delta)d\delta$.

3.3 Discrete-Bayesian Estimator

As in many Bayesian applications, solution of the investor's decision problem requires computing an expectation over the space of possible distributions Δ . To render this integration feasible, the domain of possible distributions Δ can be restricted by a model specification. A standard approach is to restrict the set of distributions to a parametric model. For example, the investor might restrict

the space of possible return distributions to the set of multivariate normal distributions. In this case, the posterior is a function of the model parameters: the mean returns and the covariance matrix. In essence, the model specification is a meta-prior that places zero probability on all distributions outside the model space.

In this section, I develop an estimator based on a model specification that does not rely on a parametric specification to achieve a computationally feasible algorithm. In the portfolio choice context, the method extends Brandt’s (1999) nonparametric method of moments to the case with non-data priors. The method also can also be related to a class of “data tuning” (Braun and Hall (2001)) methods developed for conditional smoothing and inference applications in nonparametric statistics.

To provide some intuition for the linkage between the estimator discussed here and previous literature, I anticipate the final form of the estimator constructed below. The approximation to the expectation in the investor’s problem (3.3) reduces to the form,

$$E_q[u(r_{t+1}, w)] \approx \sum_{t=1}^T u(r_t, w) q_t, \quad (3.4)$$

where the q_t , ($t \in \{1 \dots T\}$), are weights that depend on the prior. Like classical nonparametric estimators, the estimator is a sum over historical realizations of the integrand. The estimator departs from the classic estimator in that it incorporates observation weights. These weights represent a discrete approximation to the predictive distribution of returns q that depends on investor priors.

The estimator (3.4) is analogous to the nonparametric estimator of the investor’s problem developed by Brandt (1999). Brandt (1999) also estimates the investor’s expected utility by a weighted sum over historical return outcomes. However, whereas the weights in Brandt (1999) reflect conditioning on return predictors such as the dividend yield, the weights applied here depend on non-data priors. In either case, each of the estimation weights reduces to $1/T$ when no additional information is available.

Incorporation of non-data priors into a nonparametric estimation framework also motivates data tuning methods. Data tuning methods incorporate non-data information by perturbing the empirical distribution of the data subject to a

distance constraint on the perturbation. This is accomplished either by altering the weights applied to each data point in the estimator or by perturbing the data themselves subject to a distance constraint.⁵¹

While the estimator developed in this paper (3.4) has a similar structure to data tuning methods that perturb the probabilities attached to each data point, the methods of arriving at the final estimator differ. The perturbed probabilities used in data tuning are obtained by selecting the highest likelihood distribution for which the optimal decision variables satisfy a constraint.⁵² By contrast, the data weights for the nonparametric Bayesian approach are a reflection of qualitative or non-data prior views, and are obtained by numerical integration over the investor's posterior distribution by Markov chain Monte Carlo. The estimator takes the familiar nonparametric form as a result of the discretization scheme used to model the domain of distributions spanned by the posterior.

3.3.1 Construction of the Discrete-Bayesian Estimator

To construct the discrete-Bayesian estimator (3.4), I start by assuming that we have a finite sample of B draws from the posterior q_{post} . Under this assumption, the finite sample equivalent of the investor's objective in (3.3) is

$$E_{\delta} [u(r_{t+1}, w)] \Big] \approx \frac{1}{B} \sum_{b=1}^B \sum_{t=1}^{T_b} u(r_{bt}, w), \quad (3.5)$$

where $\{\delta_b\}_{b=1}^B$ is the sample of B distributions drawn from the posterior, and $\{r_{bt}\}_{t=1}^{T_b}$ is a draw from δ_b for each b .

To implement the above estimator, we require a feasible means of drawing a set of distributions from the posterior q_{post} . The expression (3.5) is a Monte Carlo approximation to the investor's expectation integral. In theory, the approximation will converge as the number of draws goes to infinity. The challenge

⁵¹Hall and Presnell (1999) introduce a method of perturbing the probability weights applied to each observation. Methods that perturb the data are referred to as data sharpening. (Choi and Hall (1999) and Choi et al. (2000) develop data sharpening methods applicable to nonparametric curve smoothing, density estimation and inference. See Braun and Hall (2001) for further references to related techniques.

⁵²To illustrate the contrast between the Bayesian resampling approach developed below and data tuning methods, I include two data tuning formulations for a sample investment problem with a non data prior. See Appendix B.

is to develop a scheme that is stable and computationally feasible. Given the limited number of historical returns available to the investor, and the high dimensionality of the underlying return distribution, some form of restriction on the space of return distributions is warranted. The usual approach is to restrict the set of possible distributions Δ by introducing a parametric model for the return distribution – often a multivariate normal.⁵³

Here, I depart from the standard approach. I modify the set of candidate distributions Δ by restricting the possible return outcomes to a finite set. This discretizes Δ to a subspace of distributions with nonnegative probabilities at a finite set of points, i.e., a set of multinomial distributions. Unlike parametric models, the approximation places no restriction on the functional form of the distribution. However, the approximation does introduce the problem of selecting a set of return outcomes to form the domain of the candidate distributions. The set must be sufficiently parsimonious for computational feasibility, yet must be dense enough to allow for a rich array of posterior distribution characteristics.

Ideally, the set of possible return outcomes would be chosen to place a greater density of points near peaks of the underlying distribution because the data are more likely to be informative on distribution shape in those regions. But we do not know the distribution in advance. I deal with this issue by setting the discrete set of possible return outcomes equal to the returns realized in the historical data. The underlying data generating process is itself the ideal randomizer for selecting a parsimonious grid of returns with the desired point distribution.

By restricting the set of possible distributions to multinomial distributions with nonnegative probabilities at the realized returns, I obtain an approximation that is analogous to classical nonparametric estimation in which inference is based on the empirical distribution. In the presence of an uninformative prior, the resulting discrete-Bayesian estimator is equivalent to the classical nonparametric estimator, i.e., the probability weights assigned to each observation are equal. One caveat is that the resulting estimator is not strictly Bayesian. The approximation used to allow integration over the posterior domain also estab-

⁵³An exception is Kacperczyk (2003) who studies a Bayesian model in which third and fourth moments of returns are unrestricted.

lishes the domain of the prior. Thus, while the prior likelihoods associated with any distribution are not affected by the use of the return outcomes to set the discretized grid, the prior and the data are no longer independent.

To formalize the construction of the discrete-Bayesian estimator, I define a subset Δ_R of the set of possible return distributions Δ . Each member of Δ_R has finite domain consisting of the T realized returns in the set of realized returns R . Thus the subspace Δ_R is the space of multinomial distributions with nonnegative probabilities p_t at all points in the set of realized returns R and zero probability elsewhere, i.e.,

$$\delta \in \Delta_R \quad \Rightarrow \quad \begin{cases} \delta(r) = \begin{cases} p_t & r \in \{r_1, \dots, r_T\} \\ 0 & \text{otherwise} \end{cases} \\ \sum_{t=1}^T p_t = 1. \end{cases} \quad (3.6)$$

For the purpose of evaluating the investor's problem, I restrict Δ to Δ_R . Each member δ of Δ_R is parameterized by $T - 1$ independent probabilities. The likelihood $l(\delta)$ is the probability of drawing the return sample given a distribution δ . Hence,

$$l(\delta|R) \propto p_1 p_2 \cdots p_{T-1} \left(1 - \sum_{t=1}^T p_t\right), \quad (3.7)$$

where p_t is the probability of draw t under distribution δ . Multiplying by the prior q_{prior} yields the posterior

$$q(\delta) = \left(\prod_{t=1}^{T-1} p_t\right) \left(1 - \sum_{t=1}^{T-1} p_t\right) q_{\text{prior}}(\delta). \quad (3.8)$$

Given a sample of B draws $\delta_b = \{p_{b1}, \dots, p_{bT}\}$ from Δ_R and noting that $p(r_t|\delta_b) = p_{bt}$, the estimator (3.5) reduces to

$$\frac{1}{B} \sum_{b=1}^B \sum_{t=1}^T u(r_t, w) p_{bt}. \quad (3.9)$$

Summing over B simplifies (3.9) to the form given in (3.4). The components q_t of the predictive distribution are given by $q_t = (1/B) \sum_{b=1}^B p_{bt}$. The pre-

dictive distribution reduces to a set of weights applied to each of the sample observations.

3.3.2 Sampling from the Posterior

The posterior distribution (3.8) is a mixture of a multinomial distribution and an as yet unspecified prior. With $T - 1$ parameters, obtaining a draw from the distribution that achieves satisfactory convergence of the investor's expectation is a challenge. One viable approach for such a high dimensional problem is Markov chain Monte Carlo.⁵⁴

To construct the Markov chain, I employ a Metropolis Hastings algorithm. At each step of the chain, I draw a distribution δ from a tractable proposal distribution. The proposal is conditional on the current distribution. This proposal is either accepted or rejected with an acceptance probability that depends on i) the ratio of the value of the posterior distribution at the proposal and the current value, and ii) the ratio of the conditional proposal probabilities. The accept-reject algorithm ensures that as the length of the chain goes to infinity, the number of draws from any region of the posterior's domain is proportional to that region's posterior probability.

I base the calibration of the MCMC algorithm on the convergence with number of draws of the posterior distributions of the mean returns and of the optimal portfolio weights implied by each distribution drawn. I obtain good convergence by combining draws from five independent chains of length 5000. The first half of the draws from each chain are dropped to allow for convergence of the chain to the posterior distribution. Thus, the posterior estimates are based on a total of 12500 draws. I provide full details and an algorithm chart in Appendix C.

3.3.3 Mean-Variance Utility

In the empirical study in section 3.6, I assume quadratic utility over wealth. Under this assumption, the portfolio estimator developed in the previous section simplifies to mean variance form. Working in the mean variance framework allows for comparison of results with those of previous research, as it is the

⁵⁴Gamerman and Lopes (2006) contrast MCMC with more direct simulation methods.

standard model in the portfolio choice literature.⁵⁵ However, application of the portfolio estimator discussed in the previous section is not dependent on this assumption. The fact that the portfolio allocations only depend on the first two moments of the return distribution is purely a function of the utility assumption. The return distribution is modeled without moment assumptions.

Let $\mu(\delta)$ and $\Sigma(\delta)$ be functions that return the mean and covariance matrix respectively of the distribution δ . Under quadratic utility, the expected utility estimator (3.4) is

$$u(r, w) = w' \mu(q) - \frac{\gamma}{2} w' \Sigma(q) w, \quad (3.10)$$

where q is the discrete predictive distribution of returns defined in the previous section.

3.4 Prior

I formulate the investor's beliefs as a quantitative prior. My objective is to examine the consequences of incorporating prior beliefs in the portfolio decision problem. The discrete nonparametric estimator described in section (3.3) allows for specification of arbitrary priors. This is liberating for Bayesian analysis as priors are not limited by analytical requirements for conjugacy with a particular return model. In addition, since the distributions are nonparametric, priors do not have to be formulated in terms of specific distribution parameters. In theory priors should take the form of probability distributions. However the estimator is insensitive to multiplicative constants. Hence, the priors must satisfy two restrictions: they must be positive functions defined on the set of candidate distributions, and multiplication with the data likelihood must result in an integrable posterior distribution.

I model four priors. The first two priors are suggested by Chevrier and McCulloch (2008). The first is a restriction on the global minimum variance portfolio implied by a candidate distribution. The second models the belief that optimal portfolios are positive.

⁵⁵Proposed by Markowitz (1952). Brandt (2005) and DeMiguel et al. (2007) review subsequent developments.

The third prior mitigates a side effect of the second prior. As will be demonstrated in section 3.6.2, the second prior biases posterior mean returns upward. It is implausible that a belief in the positivity of portfolio weights is associated with a belief that historical return means are biased downward. To mitigate this effect, I introduce a prior that models the belief that the mean of the return means or “grand mean” is close to the sample value. Finally, I consider a prior that models the belief that all portfolio weights are smaller than one.

The first prior q_1 is a step function that places zero probability on return distributions that imply a return on the global minimum variance portfolio that is less than the risk-free rate. This is a no-arbitrage condition that states that risky portfolios dominate the risk-free asset in terms of expected return. I quantify the prior by setting

$$q_1(\delta) = \mathbf{1}_{r_{GMV} > r^f} \quad (3.11)$$

where $r_{GMV} = w'_{GMV}\mu(\delta)$ is the return on the global minimum variance portfolio $w_{GMV} = \Sigma^{-1}(\delta)\mathbf{1}$ for distribution δ .

The second prior expresses investor belief that portfolio weights, given the underlying or true distribution of returns, are uniformly positive. This prior is motivated by both asset pricing theory and modeling precedents in practice and research. Asset pricing theory holds that the assets in positive net supply are priced such that the aggregate investor would have positive allocations in an optimal portfolio. If we assume the investor’s preferences are similar to those of the aggregate market, then it follows that the investor would expect optimal portfolio holdings to reflect the holdings of the market.

Short-sales constraints are often employed by investors and practitioners in implementations of portfolio choice models. Mean-variance portfolio weights computed by plug-in methods often feature large negative positions. These positions are attributable to high estimation error and the dominance of one or few eigenvalues in the covariance matrix of returns (Green and Hollifield (1992)). Whether negative weights are justified or not, mean variance methods are often implemented with a constraint against short sales. The constraints might reflect restrictions on investment mandates or trading restrictions. However, the number of assets available for taking short positions is large and continues to grow. In addition, a restrictive investment mandate, while telling us nothing

about the beliefs of an asset manager, indicates bias against short positions on the part of the investor's clients.

Prior skepticism regarding short positions is implemented by the following function:

$$q_2(\delta, c_2) = \exp\left(-c\left(\text{neg}(w_{tan})'w_{tan}\right)^2,\right) \quad (3.12)$$

where c is a decay parameter, w_{tan} is the N vector of portfolio weights in the actual optimal portfolio given distribution δ , and $\text{neg}(\cdot)$ is a vector function that returns a vector of the same length with zeros substituted for positive values and -1 substituted for strictly negative values.⁵⁶ The larger the value of c_2 , the greater is the investor's belief that optimal portfolio weights are greater than zero.

The third prior expresses the belief that the grand mean, or mean of the return means, is unlikely to exceed its sample value. While intuitive, the prior q_2 biases the posterior towards distributions in Δ with expected return vectors that are uniformly higher than the sample mean. The high variance of returns creates regions of the posterior domain characterized by high likelihood and high return means. Distributions with lower return means are discounted by the prior, and these distributions receive lower weight in the posterior. The result is an upward bias in optimal weights that is difficult to justify. The primary motivation for the belief in positive weights is to focus the posterior on candidate distributions with means that are similar to the sample, but whose returns and covariances do not combine to create unrealistic positions. Unrealistic positions are those with high leverage that arise when estimating optimal portfolios for highly correlated assets.⁵⁷

The third prior mitigates this effect. The prior is implemented as

$$q_3(\delta) = \exp\left(-c_3\left(\max(\mu_g(\delta) - \hat{\mu}_g), 0\right)^2\right), \quad (3.13)$$

where $\mu_g(\delta) = (\sum_{i=1}^N \mu_i(\delta) 1)/N$, $\hat{\mu}_g = (\sum_{i=1}^N \hat{\mu}_i)/N$, and c_3 is a parameter.

I also consider a fourth prior that expresses the view that large deviations

⁵⁶ w_{tan} is computed as the solution of $\max_w E_\delta[U(w'r)]$. To eliminate dependence on risk aversion in the prior, tangency weights are scaled to sum to one.

⁵⁷See Green and Hollifield (1992) and Jagannathan and Ma (2003) for studies of gearing in mean variance portfolios.

from the grand mean from the sample value are unlikely whether positive or negative. The prior is expressed as

$$q_3(\delta) = \exp \left(-c_4(\mu_g(\delta) - \hat{\mu}_g)^2 \right). \quad (3.14)$$

Prior q_4 is related to Bayes-Stein estimation. Jorion (1986) shows that a Bayes-Stein estimator that shrinks return means dominates traditional plug-in estimation.

An investor can have one or more of these priors. The investor's overall prior distribution is expressed as $q_{\text{prior}}(\delta; \mathbf{c}) = \prod_{i \in \mathcal{I}} q_i(\delta, c_i)$ where \mathcal{I} is the investor's subset of priors and \mathbf{c} is a vector of the relevant prior coefficients.

3.5 Relationship to Previous Portfolio Choice Models

In this section, I discuss a selection of models for portfolio choice that are related to the estimator employed in the current study. Table 3.1 lists the models considered. Researchers have proposed a vast number of innovative approaches to portfolio choice. The list considered here is by no means comprehensive. See, for example, DeMiguel et al. (2007) and Garlappi et al. (2007) for comparative empirical studies of large sets of models that overlap the set considered here.

The subsections below describe two groups of models. The first group includes plug-in rules. These rules are variants of the classical approach that makes use of sample estimates of return means and variances. The second group consists of portfolio resampling models. I conclude with a brief note on regularization in portfolio choice.

3.5.1 Plug-in Models

Perhaps the most straightforward approach to portfolio estimation is to directly substitute estimated distribution parameters into the investor's problem. For the mean variance model, this amounts to approximating the investors problem

$$\max_w w' \mu - \frac{\gamma}{2} w' \Sigma w \quad (3.15)$$

by

$$\max_w w' \hat{\mu} - \frac{\gamma}{2} w' \hat{\Sigma} w \quad (3.16)$$

where $\hat{\mu} = \frac{1}{T} \sum_{t=1}^T r_t$ and $\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T (r_t - \hat{\mu})(r_t - \hat{\mu})'$ are sample moments. Solutions obtained by the direct plug-in model (3.16) are notoriously unstable and lead to poor out-of-sample performance.⁵⁸

In efforts to achieve more robust performance, researchers have proposed numerous alternative estimators for the distribution moments used in the mean variance model. A large class of these models take the form of plug-in rules that combine the risk-free asset, a sample estimate of the mean variance portfolio, and a sample estimate of the minimum variance portfolio formed without the risk-free asset. Plug-in models assign a portfolio weight vector of the general form

$$w = \frac{\pi_1}{\gamma} \hat{\Sigma}^{-1} \hat{\mu} + \frac{\pi_2}{\gamma} \hat{\Sigma}^{-1} \mathbf{1} \quad (3.17)$$

where π_1 and π_2 are constants. The classic mean variance solution (3.16) is obtained with $\pi_1 = 1$ and $\pi_2 = 0$.⁵⁹

A number of plug-in rules that take the form (3.17) are motivated by statistical considerations. Kan and Zhou (2007) provide a detailed analysis of portfolio models whose solutions take this form. For the most part, these models shrink the estimated portfolio towards the risk-free asset (i.e., $\pi_1 < 1$) or towards the minimum variance portfolio (i.e., $\pi_2 > 0$). I list a subset of these plug-in rules in Table 3.1. The first model is the direct substitution mean variance model. I consider the direct substitution approach because it is a longstanding standard in the literature.

The second plug-in model is obtained following a parametric Bayesian analysis in which the investor assumes normally distributed returns and has a diffuse Jeffrey's prior over the distribution parameters. Kan and Zhou (2007) demonstrate that this model is dominated in terms of out-of-sample performance by some of the plug-in rules that follow. I include this model because it is interesting to compare the parametric Bayesian solution to the discrete, nonparametric

⁵⁸See Best and Grauer (1991).

⁵⁹If the unbiased estimator of Σ is used then $\pi_1 = (T - 1)/T$.

Bayesian approach developed in this paper.

The third through fifth models are developed in Kan and Zhou (2007). Kan and Zhou (2007) evaluate the efficiency of plug-in rules based on their expected out-of-sample utilities under the assumption of normally distributed returns. The third model is the most efficient plug-in rule with a coefficient on the sample portfolio π_1 that is independent of the sample. The fourth and fifth models are plug-in rules that aim to improve efficiency by allowing the coefficients in (3.17) to depend on the sample. The fourth is derived under the restriction $\pi_2 = 0$, and the fifth allows for investment in the sample minimum variance portfolio.

The sixth model is the classical mean variance model with short sales constraints imposed on the investor's problem (3.16). As discussed in the introduction, a common approach is to solve (16) under a constraint against short sales. Solutions obtained with short sale constraints imposed are closely related to those obtained by statistical approaches that shrink portfolios towards the minimum variance portfolio. Because they are straightforward to implement in many optimization packages, models with short sale constraints are very common in investment practice.

The last two models back away entirely from relying on the data to estimate return means. The seventh model is the global minimum variance portfolio. The global minimum variance portfolio has a statistical advantage in small samples in that it is independent of return estimates, and has proven to be a strong performer in many out-of-sample studies, despite being an ad hoc rule. The final entry is the $1/N$ model discussed extensively in Garlappi et al. (2007). This model takes data skepticism to the extreme. The model assumes portfolio weights sum to one and assigns equal weight to each asset.

3.5.2 Portfolio Resampling

Portfolio resampling refers to the recomputation of optimal portfolios for repeated, randomly generated return samples. The return samples can be generated by drawing from a parametric return model estimated from the sample data⁶⁰ or bootstrapping from the empirical distribution of returns. Resampling is a standard technique in statistics for estimating sampling distributions and

⁶⁰See Jorion (1992), Michaud (1998), Markowitz and Usmen (2003) and Harvey et al. (2003).

addressing hypothesis tests (Efron (1979)). Jobson and Korkie (1981) suggest the technique for statistical analysis of portfolio estimates. The technique has been adopted in a number of subsequent studies.⁶¹

Michaud (1998) considers extending the application of portfolio resampling beyond statistical hypothesis testing to decision analysis. He suggests choosing portfolios by averaging across optimal allocations computed for each resampled return draw.⁶² Scherer (2002) and Harvey et al. (2003) point out that the resampling of return distributions is conceptually equivalent to drawing from a Bayesian posterior distribution with an uninformed prior.⁶³ Viewed from a Bayesian perspective, construction of portfolio decisions from averages of resampled portfolios switches the order of the maximization and the integration over the posterior in the investor's portfolio problem (3.2). Using the notation from section 3.3, weights computed by portfolio resampling are an approximation to the integral

$$w = \int_{\Delta} \left[\arg \max_{w_b} \int_{\Omega} u(r, w_{\delta}) p(r|\delta) dr \right] q_{\text{post}}(\delta) d\delta. \quad (3.18)$$

Given B resampled return draws, the finite sample approximation is

$$\hat{w} \approx \frac{1}{B} \sum_{b=1}^B \left[\arg \max_{w_b} \sum_{t=1}^T u(r_{bt}, w_b) p_{bt} \right], \quad (3.19)$$

where δ_b is the empirical distribution associated with the b th resampled return series.

Under mean-variance utility, the portfolio resampling computation (3.18)

⁶¹For example, Jorion (1992) uses resampling to illustrate portfolio sensitivity to estimation risk. Scherer (2002) suggests application of resampling techniques for hypothesis tests of the statistical difference between two portfolios. I apply resampling to test the difference in investment benefits of different asset allocation portfolios in Chapter 2 of this thesis.

⁶²See Scherer (2002) or Meulli (2006) for a complete description of Michaud's algorithm. Herold and Maurer (2006) implements portfolio resampling using the bootstrap.

⁶³For example, as $T \rightarrow \infty$ likelihood properties of distribution moments computed from bootstrap distributions will approach those of moments based on distributions sampled from the posterior discussed in section 3.

simplifies to

$$\frac{1}{B} \sum_{b=1}^B \arg \max_{w_b} \left(\mu(\delta_b)' w_b - \gamma w_b' \Sigma(\delta_b) w_b \right). \quad (3.20)$$

In this case, the solution to the portfolio resampling problem is

$$w_{PS} = \frac{1}{B\gamma} \sum_{t=1}^T \Sigma(\delta_t)^{-1} \mu(\delta_t). \quad (3.21)$$

The portfolio resampling solution is an average of optimal weights implied by a set of distributions drawn from the posterior.

While portfolio resampling procedures are not usually associated with Bayesian analysis, the portfolio resampling heuristic can be readily extended to allow for informed priors. The solution (3.21) can be computed as long as a sampling scheme is available to obtain draws from the resulting posterior distribution. The primary model implemented by Chevrier and McCulloch (2008) is an example of such a scheme.⁶⁴ They analyze portfolio weights obtained as solutions to (3.21) where the set of distributions is drawn from a posterior that incorporates prior beliefs.

In the empirical examples of section 3.6, I compare portfolio allocations estimated via the Bayesian approach to two portfolio resampling models. Both implement the mean variance solution (3.21). The first is a standard portfolio resampling model with uninformed prior and the second incorporates prior beliefs. In both models, I employ the MCMC technique developed in section 3.3 to draw from the posterior. I find that the performance of portfolio resampling with prior is very similar to that of the discrete-Bayesian model.

Finally, I compute results for a heuristic proposed by Michaud (1998) that combines the placement of short sales constraints with portfolio resampling. For this model, each of the resampled portfolio weights are computed with constraints on short sales. For consistency with previous applications, I draw the returns for this last model from a posterior that does not include prior information.

⁶⁴Chevrier and McCulloch (2008) describe the method as the “posterior mean over the implied weights”.

3.5.3 Statistical Refinement versus Regularization

It is useful to distinguish models that are statistical refinements of the classic plug-in approach from those that give up consistency in a statistical sense to regularize (i.e., impose smoothness) on the solution vector. Of the models listed in Table 3.1, the regularizing solutions are those that impose short sale constraints and the $1/N$ rule. The other models, with the exception of the minimum variance portfolio, converge to the true optimal portfolio as the length of the available data set approaches infinity. Statistically motivated models seek to improve allocation estimates by accounting for estimation uncertainty or additional non-data information. Regularization seeks to compensate for data insufficiency by imposing smoothness on the portfolio weights vector.⁶⁵

A major challenge in portfolio estimation is that the amount of data required to achieve a stable estimate is many orders of magnitude larger than is usually available to the investor. As such, even the best statistical estimates of the return distribution yield unstable portfolio estimates. As a result, despite introducing bias, regularized solutions such as those based on constraining short sales (Jagannathan and Ma (2003) and $1/N$ (DeMiguel et al. (2007)) perform well in many applications.

While regularization is effective, ideally an investor would be able to estimate optimal allocations from available information. Previous studies indicate that statistical approaches have a very hard time outperforming the regularized solutions. The simulation analysis of the next section further tests this conclusion by comparing short sales constrained and $1/N$ performance with performance of a Bayesian model that incorporates additional non-data priors in a simulation setup in which the priors are correct.

⁶⁵In general, regularization models result in a Lagrangian with additional terms that penalize variation of the solution vector. For example, a typical regularized Lagrangian for the portfolio problem might have the form

$$E[U(\tilde{r}, w)] - \kappa \|\Phi w\|^p \quad (3.22)$$

where κ is a scaling parameter, p is the degree of the vector norm, and Φ is an $N \times N$ matrix. Jagannathan and Ma (2003) show that the no short sales solution is equivalent to (3.22) with $p = 1$ and Φ equal to a diagonal matrix with the Lagrange multipliers of the short sales constraints on the diagonal. Sufficient conditions for the $1/N$ solution have Φ equal to a demeaning operator, $\Phi = I - \mathbf{1}\mathbf{1}'/N$, $p = 2$ and $\kappa \rightarrow \infty$.

3.6 Portfolios with Positive Weights Prior

In this section, I compare the properties and performance of portfolios estimated using the discrete-Bayesian model for portfolio choice developed in section 3.3 with models from the literature. The empirical analysis is divided into two parts. In the first, I compare the expected out-of-sample performance of the models in a simulation experiment. In the second, I conduct an out-of-sample study using historical returns.

Simulation results are a useful tool for comparing the performance of portfolio models under controlled conditions. Because the underlying return distribution is known to the experimenter, out-of-sample performance for a simulated trial can be computed exactly for any portfolio weight policy.

This study differs from previous simulation studies in that some of the rules I consider assume investors have information beyond that contained in historical data. See Kan and Zhou (2007), Herold and Maurer (2006), and Garlappi et al. (2007), amongst others, for simulation results that examine outcomes for portfolio rules estimated from historical data. This study most directly extends results of Kan and Zhou (2007). I use a simulation procedure that closely parallels their procedure. In addition they provide analytic results for a number of plug-in rules that can be exploited for verifying simulation convergence.

3.6.1 Simulation Setup

In the simulations, I assume a mean-variance investor with relative risk aversion γ equal to five. I generate independent return series from a normal distribution. I consider a five-asset example. I calibrate the covariance matrix to sample values computed for the five-industry data set. In setting the return means, I partially shrink the sample mean vectors towards their grand mean. The degree of shrinkage is high enough to ensure that the optimal portfolio weights of a mean variance investor with full knowledge of the underlying distribution would be strictly positive.⁶⁶ The parameters used in the simulation trial for the five-asset case are listed in Table 3.2.

⁶⁶I set the vector of return means of the data generating process for the simulations to $\mu = \epsilon \hat{\mu} + (1 - \epsilon)(\sum_{i=1}^N \hat{\mu}_i)/N$, with $0 < \epsilon < 1$. I choose a shrinkage coefficient ϵ small enough to ensure that all optimal weights implied by the underlying distribution are greater than zero; i.e., $\hat{\Sigma}^{-1}\mu > 0$.

I employ a simulation strategy similar to that employed by Kan and Zhou (2007). I randomly generate S return series from the underlying distribution. These return series become the “historical” data for S hypothetical investors. Each investor estimates optimal portfolios using the twelve models listed in Table 3.1. I then compute the mean and standard deviation of the out-of-sample utility realizations for each estimated portfolio given their portfolio estimates and the actual distribution parameters.

I calibrate the number of simulations S required to effectively estimate the out-of-sample utility. Given the computational burden of the Markov chain Monte Carlo methods, I wish to minimize S . Kan and Zhou (2007) provide analytic expressions for the out-of-sample utility of the plug-in rules for which π_1 and π_2 are independent of the data. Using these results, I choose an S that is large enough to yield estimates that are precise to within two significant digits when compared with the theoretical result. I find that $S = 250$ is sufficient for this purpose.

3.6.2 Exploratory Analysis of the Prior Assumptions

An important preliminary step in a Bayesian investigation is characterization of the posterior. A prior model designed to quantify an investor belief may have unintended consequences for the posterior. This can result in a posterior model that, while ostensibly capturing the investor belief that is of primary interest, is decidedly unrealistic on another dimension.

Here I contrast the posterior distributions for several different priors. The posteriors are based on a return series of length $T = 120$. The return series is drawn from the five-asset distribution parameterized in Table 3.2. For the analysis, I employ 12500 distributions drawn from the posterior distribution by the MCMC method described in the appendix. I compute summary statistics for each of the 12500 draws. The summary statistics include return means, return standard deviations, and the implied portfolio weights of a mean variance investor with risk aversion coefficient equal to five.

The first panel of Table 3.3 shows return means, return standard deviations, and portfolio weights estimated directly from sample data. The remainder of Table 3.3 reports averages and standard deviations of summary statistics across posterior draws under different prior specifications. The second panel shows the

summary statistics averaged across 12500 distributions drawn without prior. As expected, given that the prior adds no information, the means, returns, and weights are close (within 1%) to values obtained under a diffuse prior assumption on μ and Σ . For a normal model, and given a diffuse prior, the optimal weights are the standard mean variance weights multiplied by $(T - N - 2)/(T + 1)$. In this case, this implies average weights of 0.81, slightly below the value predicted by simulation.

The third panel of Table 3.3 displays results obtained with investor bias against negative portfolio weights incorporated into the prior. For this initial case, the model only incorporates the economically motivated priors described in Section 3.4: the no-arbitrage prior q_1 , and the positivity prior q_2 . I select the parameter c_2 following a qualitative tuning exercise. I consider the posterior distribution obtained when the informed prior is paired with a return sample of length $T = 120$. I find that $c_2 = 20$ yields a set of posterior draws that places less than 5% weight on distributions that imply negative weights in a mean variance optimized portfolio. Hence, this choice of parameter reflects a strong, but not completely dogmatic belief that all portfolio weights are positive.

Under priors q_1 and q_2 , the variability of the means and standard deviations are similar to those observed with no prior. However, comparison of the bottom lines of panels 2 and 3 reveals that implied portfolio weights are far less variable across posterior draws for the informed investor. The prior q_2 informs the implied weights of the mean variance investor without increasing the precision of the individual moment estimates.

Figure 3.1 compares histograms of portfolio weights obtained with no prior with those obtained when priors q_1 and q_2 are implemented. Under the informed prior, the posterior probability that the optimal portfolio weights are less than zero is clearly reduced relative to the no-prior case, as is the variance of the posterior distribution of portfolio weights. The standard deviations of the portfolio weights implied by the posterior distribution are reduced by two to three times upon incorporating a belief in the positivity of portfolio weights.

While the prior q_2 succeeds in incorporating a belief in positive weights, there is a noticeable upward bias in the overall average of the return means. Averaging across the first row of any panel in Table 3.3 yields a cross-asset mean or “grand” mean of the asset returns. The grand mean based on the first

row of panel 3 is 1.15 while that for the second panel is 0.82, a difference of 0.33 percentage points per month. Given a prior that favors positive weights, the posterior places greater weight on distributions with higher return means.

The return bias is illustrated in Figure 3.2. Each panel shows a histogram of the posterior distribution of one asset's return mean (vertical bars). The solid line is a normal distribution fitted to the data. The average posterior return means are shifted to the right relative to their no prior counterparts.

The return bias is a side effect of the positive weights prior. Mean variance analysis tells us that portfolio weights are functions of the mean variance ratio for each asset and the correlation matrix. The positive weights prior informs the relative magnitudes of the mean variance ratios as well as the correlation matrix, but does not directly restrict the absolute magnitudes of either the means or the variances.

The upward shift in return means is not warranted by theory. Skepticism about negative portfolio weights is not usually associated with optimism with respect to return means. Investors are skeptical about negative portfolio weights because of noise in the relative returns of highly correlated assets that create opportunities for long-short trades that are not warranted by the data evidence. To correct the posterior bias of return means, I activate prior q_3 that reflects skepticism that the grand mean of returns is larger than the grand mean of the sample returns. The fourth panel of Table 3.3 lists the posterior characteristics under the revised prior. The averages of the posterior mean returns are shrunk towards the grand mean in the sample, and the standard deviation of these quantities is reduced relative to the posteriors computed without the prior. The primary effect of the shrinkage of the return means is to reduce the overall investment in risky assets. The impact on the posterior distribution of implied portfolio weights is muted.

Figure 3.3 shows the distribution of posterior returns when priors q_1 , q_2 , and q_3 are incorporated into the model. A comparison with Figure 3.2 shows that prior q_3 has a noticeable effect on the return distribution. The information in the prior acts to restrict the posterior distributions of return means. The distributions are shifted towards the grand mean of the sample, and the posterior distribution of the return means has visibly less variance.

Incorporating prior q_3 has an inconsequential effect on the precision of port-

folio weight estimates. Figure 3.4 compares histograms of portfolio weights obtained with no prior with those obtained under prior beliefs that include the no-arbitrage q_1 , positive-weights q_2 , and mean-restriction q_3 priors. The histograms are similar to those shown in Figure 3.2 for the case under priors q_1 and q_2 only. In addition, differences in the averages and standard deviations of the portfolio weights listed in the third and fourth panels of Table 3.3 are inconsequential. Thus, while prior q_3 mitigates return bias introduced by the positive weights prior, it has little impact on the precision of the final allocation estimates.

3.6.3 Monte-Carlo Study

For each of the 250 simulated data sets, I compute the first two moments of the Bayesian predictive distribution of returns under different prior specifications. The top and bottom panels of Table 3.4 list return means and standard deviations, respectively, averaged across simulations. The first and second lines of each panel show results computed for an investor with no prior and with prior that reflects belief in the positivity of portfolio weights, i.e., with prior q_2 activated.⁶⁷ Comparison of the two lines reveals the upward return bias associated with this prior. The mean predictive returns are at least one standard deviation higher than those computed without prior. While not significant in the conventional sense, the upward bias is economically significant. The bias represents an approximate 5% per annum upward shift in expected returns.

The introduction of prior q_3 or q_4 eliminates the upward bias in return means. Results given these priors are listed in lines 3-5 of each panel of the table. Examination of the means of the return standard deviations computed for each simulation (reported in the second panel) reveals the limited impact of priors on second moments of the predictive return distribution. Some reduction in the variance of predictive distributions is expected given that the priors introduce additional information. However, when compared to the no-prior case, the reduction in the standard deviations are small (between 1% and 8%).

Table 3.5 shows simulation averages of portfolio weights estimated under different priors. The first panel shows results estimated using the discrete-

⁶⁷I include the no-arbitrage prior in all cases.

Bayesian model. These results illustrate the additional stability of portfolio weight estimates when priors are incorporated. For example, comparison of lines 1 and 3 of either panel reveals that a model that incorporates priors q_1 , q_2 and q_3 reduces standard deviations of portfolio weights across simulations by a factor of four.

The second panel shows means of allocations for the portfolio resampling approach with prior. For each simulation, I calculate allocation estimates by averaging the optimal weights computed for each distribution drawn from the posterior by MCMC. Comparison of results listed in the first and second panels reveals very little difference in either the average or the variability of portfolio weights across simulations. Thus, despite being an ad hoc decision rule, portfolio weights computed by portfolio resampling are as stable as those obtained by solving the full Bayesian predictive problem when the posterior is based on the same priors.

Table 3.6 shows out-of-sample performance given estimated portfolio weights averaged across simulations. Performance is reported as percent certainty equivalent return per month. As in Table 3.5, I report results for the discrete-Bayesian model and the portfolio resampling approach. The three columns summarize model performance for three sizes of data window: $T \in \{60, 120, 240\}$. The results demonstrate the difficulty of estimating portfolios given just 60 data points. The paucity of information in such a small data set is not sufficient to achieve a risk adjusted portfolio return that exceeds the risk-free rate whether or not the investor has prior information. The certainty equivalent returns in the first column are uniformly negative.

For the longer data histories, the expected risk adjusted return is improved under the positive weight prior. Reading line 2 of Table 3.6, the monthly risk adjusted returns are -0.03 and 0.13 in the $T = 120$ and $T = 240$ cases respectively compared to -0.27 and 0.02 in the no-prior case. However, the upward return bias associated with the positive weights prior is costly. The investor's out-of-sample utility is further improved if the prior also downweights distributions with grand mean greater than that of the sample. For the $T = 120$ and $T = 240$ case, the positive returns are highest when prior q_3 is implemented, as in lines 3 and 4 of either panel of Table 3.6.

3.6.4 Comparison of Discrete-Bayesian Policy with Other Models

Next, I compare the portfolio policies and the expected performance of the discrete-Bayesian model to those of the allocation models listed in Table 3.1. Table 3.7 lists simulation summary statistics while Table 3.8 mirrors Table 3.6 in reporting expected out-of-sample performance of estimated policies.

The first panel of Table 3.7 shows optimal policies of an investor with full knowledge of the first two moments of the underlying distribution. Needless to say, these policies are unachievable in practice. However their properties and performance are a useful reference when evaluating the performance of estimated optimal policies. As in Table 3.5, the average weight across assets is listed in column six. For the full information mean variance portfolio, the average weight assigned across assets is 0.15. This is essentially equal to the simulation mean of the average weights for each of the discrete-Bayesian models examined in Table 3.5 with the exception of the case with positive weights prior instituted without the prior against return bias. In the latter case, the average weight across assets is 5% higher, corresponding to a 25% greater allocation to risky assets.

The second panel of Table 3.7 shows portfolio weights averaged across simulations for eight plug-in rules. The last column show the simulation average of the allocation magnitudes averaged across assets. The difference between mean allocation magnitudes and mean weights reveals the importance of negative weights in the portfolios suggested by the various policies. The first four policies are in decreasing order of the degree to which they account for estimation error by shrinking towards the zero portfolio. As expected, both the mean weight and mean weight magnitude decrease with degree of shrinkage. However, the mean policy weights obtained using the estimated two-fund rule of Kan and Zhou (2007) are substantially lower than those of the full information portfolio.

Kan and Zhou (2007) provide an analytic foundation and simulation results that support the conclusion that their best two-fund and estimated two-fund rules will exhibit better expected out-of-sample performance. However, there is a clear upper bound on their ability to improve performance, because they do not alter the relative magnitudes of the estimated weights. Shrinkage towards the zero portfolio results in smaller negative weights only insofar as the

magnitude of all weights are reduced.

The estimated three-fund rule is an optimal combination of the expected mean variance portfolio, the zero portfolio, and the estimated minimum variance portfolio. This rule allows for shrinkage in the relative magnitudes of portfolio weights by including the estimated minimum variance portfolio as an additional shrinkage target. This effectively allows for some rotation of the covariance matrix and reduction in the relative magnitudes of negative weights. The contrast is best illustrated by comparing the results for the best two-fund and estimated three-fund portfolios (lines 3 and 5 of the second panel of Table 3.7). The average weights for these two portfolios are equal, but the average weight magnitude is 25% lower in the estimated three-fund portfolio. The estimated three-fund policy allocates less to negative positions on average. However, having the minimum variance portfolio as a target portfolio does have disadvantages. The average allocation to the fifth asset is negative and the average allocation to the second asset is well above that of the full information mean variance portfolio. This bias is attributable to the fact that the full information global minimum variance portfolio in this example places extreme weights on these two assets; -0.45 on asset 5 and 0.83 on asset 2, relative to the true mean variance optimal portfolio.

Table 3.8 lists expected out-of-sample performance of the alternative policies for data windows of length 60, 120, and 240. Results for the plug-in rules are listed in the second panel. The expected out-of-sample performance improves with horizon for all policies (except the mechanical $1/N$ rule).

The performance results for the first five mean-variance based rules reflect the conclusions of Kan and Zhou (2007). The two and three-fund policies proposed by Kan and Zhou (2007) show better performance than the classical mean-variance and diffuse Bayesian policies, with the estimated three-fund portfolio showing best performance. The significance of the performance improvement decreases with horizon. None of the policies has expected out-of-sample returns greater than the risk-free rate for the 60 period case. The performance of the estimated three-fund policy is essentially equal to that of the mean variance policy with short sale constraints.

The expected out-of-sample performance of the discrete-Bayesian model with the no bias positive weights prior is higher than that of the estimated

three-fund rule of Kan and Zhou (2007). The difference in performance is obtained by comparing line 3 of Table 3.6 with the results in Table 3.8. For a 120 period horizon, the improvement is seven basis points, dropping to three basis points at the 240 period horizon. The expected performance improvement is small and insignificant relative to the performance standard errors. The increase in expected risk adjusted return is approximately one third of its standard error at both the 60 and 120 period horizons. Based on the improvement in out-of-sample performance alone, the discrete-Bayesian approach is not overwhelmingly superior to the best mean-variance based models of Kan and Zhou (2007).

Stability of estimated weights is also an important consideration. Here, the discrete-Bayesian model is a significant improvement over the estimated two-fund and three-fund rules of Kan and Zhou (2007). The standard deviations of the estimated weights are two to three times greater for the mean variance based policies than they are for the discrete-Bayesian model. Even the mean variance policy with short sales constraints imposed exhibits standard errors in its weight estimates that are approximately one and a half times the standard deviation of the discrete-Bayesian weights.

The last two lines of the second panel of Table 3.8 show expected out-of-sample utility for the minimum variance and $1/N$ policies. Both of these policies exhibit expected out-of-sample performance that is marginally higher than that of the discrete-Bayesian model. The minimum variance and $1/N$ policies mitigate the impact of estimation error by ignoring estimated return means. This approach trades off a major source of estimation error (because return means are difficult to estimate precisely given the high variance of asset returns) for estimator bias. The bias arises because the portfolio weights obtained by minimum variance do not converge to the true optimal weights as sample length approaches infinity. The difference between the realized risk-adjusted return of the full information portfolio without bias and the full information portfolios under the two biased estimates serves as a measure of the potential lost opportunity. For the example problem considered here, an investor gives up twenty-nine or twenty-seven basis points in the full information limit, depending on whether they employ the minimum variance or $1/N$ approaches.

Despite the potential costs, the biased estimators dominate the unbiased

mean variance estimators in the five-asset example. The convergence of the expected out-of-sample utility to the full information value is significantly slower for the unbiased mean variance estimators. In the five-asset example, the expected out-of-sample utility of the minimum variance strategy reaches its limit given a sample length of 120, and is not significantly below the full information limit with a sample length as low as 60. Of course, the expected out-of-sample utility of the $1/N$ strategy is independent of sample size because the data play no role in setting the portfolio allocations. Even for sample sizes as high as $T = 240$, the expected out-of-sample utility of the mean variance estimators included in Table 3.8 does not surpass the values for the minimum variance and $1/N$ portfolios.

The discrete-Bayesian portfolio does better, but only matches the expected out-of-sample performance of the minimum variance portfolio for sample length $T = 240$.

The last panel of Tables 3.7 and 3.8 shows portfolio results and expected out-of-sample utility for the portfolio resampling model with short sales restrictions. This model differs from the portfolio resampling models listed in Table 3.6. In both cases, a final portfolio estimate is obtained by averaging over a larger number of estimated portfolios.

3.7 Out-of-Sample Study

In this section, I examine the out-of-sample performance of the discrete-Bayesian portfolio estimates described in section 3.3. I consider the performance of investments in three asset baskets. The data set for each basket consists of a set of excess returns over the risk-free rate. The first two data sets are monthly returns of five and ten-industry portfolios for the period January 1952 to December 2006. The industry definitions and data are from the web site of Ken French. The third consists of international returns on four country indices from Morgan Stanley Capital International for the period January 1975 to December 2006.

The study design is based on a rolling window framework. The investor is assumed to re-balance their portfolio based on data for the previous T months, where T is either 120 or 240. The rolling window setup is rather arbitrary as

it sets an arbitrary date-based cutoff on the information in historical returns. My main objective with the out-of-sample experiment design is to maintain a connection with the fixed-window simulation analysis, and to ensure that there is no look-ahead bias.

I compute certainty-equivalent returns for the discrete-Bayesian policy, the resampling policy with prior, as well as seven policies of the policies from the literature listed in Table 3.1. Table 3.9 lists the out-of-sample risk-adjusted returns for the two policies based on draws from the discrete-Bayesian posterior, and Table 3.10 lists parallel results for the policies from the literature.

The first and second panels of Tables 3.9 show results for the $T = 120$ and $T = 240$ estimation window respectively. For the $T = 120$ case, the portfolio policy never achieves positive ex-post risk adjusted returns. The five-industry results (column 1) are poor when compared to the results obtained for the simulated case. In the simulation experiment, the expected out-of-sample utility is significantly greater than the realized results over the four time periods listed in Table 3.9. Even with the estimation window expanded to 240 months, the realized risk-adjusted returns are not significantly different from zero (second panel of Table 3.9).

The risk-adjusted returns listed in Table 3.9 can be compared with those obtained for suggested approaches from the previous literature listed in Table 3.10. The discrete-Bayesian policy outperforms the mean variance policy⁶⁸, but does not do as well as the statistically motivated policies of Kan and Zhou (2007). Both the estimated two-fund and the estimated three-fund policies outperform the discrete-Bayesian policy.

Consistent with results in DeMiguel et al. (2007), the best performing policies are the minimum variance and $1/N$ policies. The outperformance of the minimum variance portfolio is an indicator that historical returns data yields exploitable information on return correlations. However, there is a cost to this outperformance as average position magnitudes are approximately twice the size of estimated for the models that constrain short sales (a set that includes the $1/N$ portfolio) as well as for the models listed in Table 3.9 that incorporate priors against short selling.

⁶⁸The poor performance of standard mean variance is well known, and this poor performance extends to the mean variance method with diffuse prior (not reported).

The $1/N$ models and the policies that explicitly forbid short sales (mean variance with no short sales and the direct averaging of resampled weights with no short sales) introduce bias into the portfolio estimation problem. The $1/N$ policy achieves superior risk adjusted returns across asset sets and time periods than the discrete-Bayesian model with informed prior. The short sales constrained policies achieve higher risk adjusted returns for most cases, but the improvement was not statistically significant. These models differ from the discrete-Bayesian policy in that the problems are formulated with exogenous constraints designed to smooth, or ‘regularize’, the variability of the portfolio weights vector. In contrast, the Bayesian approach imposes priors that restrict the space of possible distributions to weights that obey certain properties. In each case, there is a tradeoff of expected bias and variance of the weight estimates. The better performance of the $1/N$ policy relative to the discrete-Bayesian policy demonstrates that costs of estimation error outweigh the smoothing bias for the three asset baskets during the time periods considered in this study.

The out-of-sample tests attest to the usefulness of shrinkage approaches in portfolio forecasting. One could increase the smoothness of the informed Bayesian solution by increasing the confidence in the in prior. However, that approach implicitly conflates the problem of distribution estimation and the problem of selecting the best solution given the available information.

3.8 Conclusion

This paper develops a simple and flexible modeling approach for incorporating general priors into the portfolio estimation problem without requiring the specification of a parametric distribution family. I construct a nonparametric approximation to the predictive distribution of returns by restricting the set of possible return outcomes in candidate distributions under the posterior to returns realized in the sample. This allows me to sample from the posterior by MCMC. Given a resampled set of draws from the posterior, the estimator for the investor’s utility expectation reduces to a simple weighted sum over the sample outcomes.

I apply the discrete-Bayesian estimator to an analysis of a portfolio estima-

tion problem under the prior belief that all portfolio weights are positive. In the simulation analysis, I examine the problem of allocating across five assets with asset correlations greater than 0.5. Such problems are particularly thorny for traditional mean variance analysis as estimation error in the returns can result in solutions that leverage similar assets against each other.

In a simulation analysis, I examine the performance of the discrete-Bayesian model for a five asset problem given an investor with prior belief that positive weights are positive. I find that the discrete-Bayesian policy with informed prior performs on par with models that impose short sale constraints and underperforms the $1/N$ portfolio. Thus, despite introducing bias, smoothing type estimators such as the short-sales constrained and $1/N$ models are as useful to the investor as conditioning on a prior that is true. The benefits of regularizing are further demonstrated in an out-of-sample study on three historical asset sets.

This paper does not extend the analysis of the portfolio allocation problem to the analysis of regularization of estimators in the presence of additional information. The analysis of regularization versus informed Bayesian and statistical estimators is a fertile and open area for future research. The implications of more general forms of regularization is, itself, a new area of research in the portfolio choice field. It will be interesting to address the question of whether portfolio forecasts can be improved by combining portfolio forecasts that incorporate non-data information with regularization of the optimization problem.

Table 3.1: Portfolio Choice Models

This table lists portfolio allocation models that are evaluated in the simulation and out of sample tests.

Model	Description
DISCRETE BAYESIAN MODELS	
Bayesian Full Prior	Portfolio estimated from empirical distribution with probability weights tuned to reflect prior beliefs. Bayesian techniques are used to condition observation weights on priors.
Bayesian No Prior	Portfolio estimated assuming an investor with no prior beliefs.
PLUG-IN MODELS	
Mean variance	Portfolio estimated using sample estimates for mean returns and covariances.
Mean variance with diffuse prior	Mean variance model with estimation uncertainty accounted for via the introduction of diffuse priors for mean and variance.
Best two fund	Portfolio estimated using sample estimates for means and covariance matrix with fixed shrinkage coefficient chosen to optimize expected out of sample performance of the model.
Estimated two fund	Portfolio estimated using sample estimates for means and covariance matrix with shrinkage coefficient estimated from data.
Estimated three fund	Portfolio estimated using sample estimates of means and covariance matrix. Portfolio formed by combining standard mean variance portfolio, minimum variance portfolio and risk-free asset with coefficients estimated from data.
Mean variance - no short	Mean variance portfolio with no-short-sales constraint.
Minimum variance	Global minimum variance portfolio.
1/ N	Portfolio formed by placing equal weights in each asset.
PORTFOLIO RESAMPLING MODELS	
Averaged weights	Portfolio formed by averaging over resampled weights. Each vector of resampled weights is computed by mean variance for using resampled data.
Averaged weights - no short sales	As above but with no-short-sales constraints applied to each weights calculation.

Table 3.2: Simulation Parameters

This table lists parameters of the multivariate normal distribution used to generate return samples for the simulations.

Asset:				
1	2	3	4	5
MEANS				
0.64	0.58	0.66	0.68	0.69
STANDARD DEVIATIONS				
4.44	4.08	5.22	4.99	4.89
CORRELATION MATRIX				
1.00	0.81	0.73	0.70	0.88
0.81	1.00	0.71	0.68	0.87
0.73	0.71	1.00	0.63	0.73
0.70	0.68	0.63	1.00	0.70
0.88	0.87	0.73	0.70	1.00
MEAN VAR. WTS. ($\gamma = 5$)				
0.21	0.25	0.08	0.22	0.01

Table 3.3: Posterior Distribution Given a Simulated Sample: Summary Statistics

This table lists summary statistics for the posterior distribution of asset returns given a simulated data set of length $T = 120$ and prior information. The data set consists of asset returns for five assets and is generated from the multivariate normal distribution described in Table 2. The first panel shows summary statistics of the random return sequence. Subsequent panels show averages of return means and standard deviations as well as implied portfolio weights. The three distribution summary statistics are averaged over 12500 distributions drawn from the posterior. Each panel corresponds to a different prior specification. The second panel is the no prior case. In the next three panels, the prior coefficients $\{c_1, c_2, c_3\}$ reflect investor belief that i) the global minimum variance portfolio has return greater than the risk-free asset, ii) optimal portfolio weights are positive, and iii) the grand mean of returns is not higher than the grand mean of the sample or 5% per annum - whichever is higher.

Table 3.3: (continued)

	Asset:					Avg.
	1	2	3	4	5	
SAMPLE STATISTICS						
<i>mean</i>	0.75	0.77	0.78	0.96	1.07	0.87
<i>std. dev.</i>	4.33	4.24	4.92	4.98	4.64	4.62
<i>weight</i>	-0.96	-0.20	0.08	0.48	1.56	0.19
PRIOR: NONE						
<i>mean</i>	0.71	0.73	0.76	0.94	1.01	0.83
	[0.29]	[0.27]	[0.30]	[0.31]	[0.30]	
<i>std. dev.</i>	4.28	4.17	4.87	4.89	4.60	4.56
	[0.19]	[0.18]	[0.22]	[0.23]	[0.21]	
<i>weight</i>	-0.96	-0.22	0.12	0.53	1.52	0.19
	[0.70]	[0.65]	[0.44]	[0.37]	[0.75]	
PRIOR: $c_1 = \text{YES}$ $c_2 = 20$ $c_3 = 0$						
<i>mean</i>	1.13	1.09	1.11	1.16	1.26	1.15
	[0.24]	[0.23]	[0.24]	[0.26]	[0.28]	
<i>std. dev.</i>	4.24	4.17	4.76	4.85	4.60	4.52
	[0.19]	[0.16]	[0.22]	[0.19]	[0.18]	
<i>weight</i>	0.17	0.20	0.26	0.33	0.52	0.30
	[0.21]	[0.23]	[0.23]	[0.27]	[0.34]	
PRIOR: $c_1 = \text{YES}$ $c_2 = 20$ $c_3 = 100$						
<i>mean</i>	0.80	0.76	0.74	0.84	0.87	0.80
	[0.10]	[0.09]	[0.11]	[0.14]	[0.11]	
<i>std. dev.</i>	4.20	4.11	4.66	4.86	4.50	4.47
	[0.15]	[0.20]	[0.20]	[0.19]	[0.17]	
<i>weight</i>	0.15	0.18	0.15	0.27	0.32	0.21
	[0.18]	[0.19]	[0.17]	[0.22]	[0.25]	
PRIOR: $c_1 = \text{YES}$ $c_2 = 200$ $c_3 = 100$						
<i>mean</i>	0.81	0.78	0.79	0.84	0.89	0.82
	[0.08]	[0.08]	[0.11]	[0.12]	[0.10]	
<i>std. dev.</i>	4.16	4.06	4.66	4.82	4.50	4.44
	[0.15]	[0.19]	[0.19]	[0.17]	[0.17]	
<i>weight</i>	0.16	0.19	0.21	0.24	0.32	0.22
	[0.15]	[0.17]	[0.17]	[0.20]	[0.23]	

Table 3.4: Monte Carlo Statistics for Predictive Distributions

This table lists the average across simulations of summary statistics for the predictive distributions of returns for investors with different prior specifications. The first panel shows predictive return means averaged across simulations. The second panel shows return standard deviations. Simulation standard deviations are listed in square brackets. For each simulation, return series are drawn from the multivariate normal distribution specified in Table 2. The prior coefficients $\{c_1, c_2, c_3\}$ reflect investor beliefs that i) the global minimum variance portfolio has return greater than the risk-free asset, ii) optimal portfolio weights are positive, and iii) the grand mean of returns is not higher than the grand mean of the sample or 5% per annum - whichever is higher.

Prior Model	Asset:					AVG
	1	2	3	4	5	
AVERAGE RETURN MEANS						
<i>No prior</i>	0.66 [0.39]	0.61 [0.36]	0.66 [0.46]	0.70 [0.44]	0.73 [0.43]	0.67
$c_1 = Y \ c_2 = 20$	1.10 [0.28]	1.01 [0.26]	1.18 [0.30]	1.14 [0.31]	1.23 [0.31]	1.13
$c_1 = Y \ c_2 = 20 \ c_3 = 100$	0.67 [0.26]	0.61 [0.24]	0.72 [0.28]	0.70 [0.29]	0.75 [0.29]	0.69
$c_1 = Y \ c_2 = 200 \ c_3 = 100$	0.68 [0.26]	0.62 [0.24]	0.73 [0.29]	0.70 [0.29]	0.76 [0.30]	0.70
$c_1 = Y \ c_2 = 20 \ c_3 = 0 \ c_4 = 100$	0.66 [0.35]	0.61 [0.32]	0.70 [0.39]	0.69 [0.37]	0.74 [0.39]	0.68
AVERAGE RETURN STANDARD DEVIATIONS						
<i>No prior</i>	4.41 [0.28]	4.07 [0.27]	5.18 [0.35]	4.96 [0.31]	4.86 [0.30]	4.70
$c_1 = Y \ c_2 = 20$	4.35 [0.29]	4.02 [0.27]	5.12 [0.36]	4.91 [0.32]	4.80 [0.32]	4.64
$c_1 = Y \ c_2 = 20 \ c_3 = 100$	4.25 [0.27]	3.92 [0.26]	5.01 [0.35]	4.81 [0.31]	4.68 [0.30]	4.53
$c_1 = Y \ c_2 = 200 \ c_3 = 100$	4.25 [0.28]	3.92 [0.26]	5.00 [0.35]	4.80 [0.31]	4.68 [0.31]	4.53
$c_1 = Y \ c_2 = 20 \ c_3 = 0 \ c_4 = 100$	4.14 [0.27]	3.83 [0.26]	4.92 [0.35]	4.71 [0.31]	4.56 [0.29]	4.43

Table 3.5: Monte Carlo Averages of Estimated Allocations

This table lists estimated allocations averaged across return simulations. Results are displayed for different prior specifications. For each simulation, $T = 120$ returns are drawn from the multivariate normal distribution specified in Table 2. A risk aversion coefficient of five is assumed for the portfolio calculations. The first panel shows allocations estimated using the discrete Bayesian model. The portfolio weights are the mean variance optimal weights given the estimated predictive distribution of returns. The second panel shows allocations estimated by averaging portfolio weights implied by a set of distributions sampled from the posterior. Mean variance optimal allocations are computed for each distribution in the investor's posterior draw, and allocations are set to the average of these allocations. Standard deviations across simulations are listed in square brackets. The prior depends on three parameters. c_1 indicates investor belief that the global minimum variance portfolio has higher return than the risk-free rate. c_2 determines the strength of investor belief in the positivity of portfolio weights. c_3 determines the strength of the investor's prior view that the grand mean of the asset returns does not exceed that of the sample.

Table 3.5: (Continued)

Prior Model	Asset					Avg.	Avg.
	1	2	3	4	5	w	$ w $
DISCRETE BAYESIAN							
<i>No prior</i>	0.21 [0.92]	0.33 [1.04]	0.03 [0.64]	0.22 [0.61]	0.06 [1.06]	0.14	0.67
$c_1 = Y \quad c_2 = 20$	0.29 [0.23]	0.34 [0.25]	0.22 [0.21]	0.30 [0.27]	0.23 [0.19]	0.23	0.28
$c_1 = Y \quad c_2 = 20$ $c_3 = 100$	0.18 [0.15]	0.22 [0.16]	0.14 [0.13]	0.20 [0.19]	0.14 [0.13]	0.15	0.18
$c_1 = Y \quad c_2 = 200$ $c_3 = 100$	0.18 [0.13]	0.20 [0.13]	0.15 [0.11]	0.20 [0.17]	0.15 [0.12]	0.15	0.18
$c_1 = Y \quad c_2 = 20$ $c_3 = 0 \quad c_4 = 100$	0.20 [0.20]	0.26 [0.27]	0.13 [0.23]	0.22 [0.27]	0.15 [0.18]	0.16	0.20
PORTFOLIO RESAMPLING							
<i>No prior</i>	0.23 [0.96]	0.34 [1.08]	0.03 [0.66]	0.22 [0.62]	0.05 [1.09]	0.14	0.70
$c_1 = Y \quad c_2 = 20$	0.29 [0.24]	0.34 [0.26]	0.22 [0.21]	0.30 [0.28]	0.23 [0.20]	0.23	0.28
$c_1 = Y \quad c_2 = 20$ $c_3 = 100$	0.19 [0.15]	0.22 [0.16]	0.14 [0.14]	0.20 [0.20]	0.15 [0.13]	0.15	0.18
$c_1 = Y \quad c_2 = 200$ $c_3 = 100$	0.18 [0.13]	0.20 [0.13]	0.15 [0.11]	0.20 [0.17]	0.16 [0.12]	0.15	0.18
$c_1 = Y \quad c_2 = 20$ $c_3 = 0 \quad c_4 = 100$	0.20 [0.20]	0.26 [0.29]	0.13 [0.23]	0.22 [0.32]	0.15 [0.21]	0.16	0.20

Table 3.6: Monte-Carlo Average Out-of-Sample Utilities

This table shows out-of-sample utility averaged across simulations. A risk aversion coefficient of five is assumed for the portfolio calculations. The first panel shows results for allocations estimated using the discrete Bayesian model. The portfolio weights are the mean variance optimal weights given the estimated predictive distribution of returns. The second panel shows results for allocations estimated by portfolio resampling. Mean variance optimal allocations are computed for each distribution in the investor's posterior draw, and allocations are set to the average of these allocations. Standard deviations across simulations are listed in square brackets. The prior depends on three parameters. c_1 indicates investor belief that the global minimum variance portfolio has higher return than the risk-free rate. c_2 determines the strength of investor belief in the positivity of portfolio weights. c_3 determines the strength of the investor's prior view that the grand mean of the asset returns does not exceed that of the sample.

Prior Model	U_{Oos}		
	$T = 60$	$T = 120$	$T = 240$
DISCRETE BAYESIAN			
<i>No prior</i>	-1.01 [1.01]	-0.27 [0.43]	0.02 [0.17]
$c_1 = Y \ c_2 = 20$		-0.03 [0.32]	0.13 [0.15]
$c_1 = Y \ c_2 = 20 \ c_3 = 100$	-0.01 [0.57]	0.16 [0.18]	0.20 [0.09]
$c_1 = Y \ c_2 = 200 \ c_3 = 100$		0.16 [0.18]	0.20 [0.09]
$c_1 = Y \ c_2 = 20 \ c_3 = 0 \ c_4 = 100$	-0.28 [0.95]	0.08 [0.30]	0.16 [0.14]
PORTFOLIO RESAMPLING			
<i>No prior</i>	-1.16 [1.12]	-0.30 [0.45]	0.01 [0.18]
$c_1 = Y \ c_2 = 20$		-0.04 [0.34]	0.13 [0.15]
$c_1 = Y \ c_2 = 20 \ c_3 = 100$	-0.02 [0.60]	0.16 [0.19]	0.20 [0.09]
$c_1 = Y \ c_2 = 200 \ c_3 = 100$		0.16 [0.18]	0.20 [0.09]
$c_1 = Y \ c_2 = 20 \ c_3 = 0 \ c_4 = 100$	-0.31 [1.02]	0.06 [0.37]	0.15 [0.14]

Table 3.7: Monte-Carlo Average Portfolio Allocations for Alternative Portfolio Estimates

This table lists means across simulations of investment weights computed using different allocation models. The data set consists of 120 monthly returns for all five assets in the simulated portfolio. Standard deviations are reported in square brackets.

Model	Asset					Avg. w	Avg. $ w $
	1	2	3	4	5		
FULL INFORMATION							
<i>Mean variance</i>	0.21	0.25	0.08	0.22	0.01	0.15	
<i>Global minimum variance</i>	0.40	0.83	0.05	0.18	-0.45	0.20	
PLUG IN							
<i>Mean variance</i>	0.21 [0.92]	0.32 [1.02]	0.03 [0.63]	0.21 [0.59]	0.07 [1.04]	0.17	0.67
<i>Mean variance: diffuse prior</i>	0.19 [0.86]	0.30 [0.95]	0.03 [0.59]	0.20 [0.56]	0.06 [0.97]	0.16	0.62
<i>Best two fund</i>	0.19 [0.82]	0.29 [0.91]	0.02 [0.57]	0.19 [0.53]	0.06 [0.93]	0.15	0.60
<i>Est. two fund</i>	0.08 [0.44]	0.14 [0.50]	0.01 [0.31]	0.09 [0.29]	0.04 [0.53]	0.07	0.27
<i>Est. three fund</i>	0.25 [0.41]	0.49 [0.53]	0.04 [0.27]	0.15 [0.25]	-0.18 [0.50]	0.15	0.35
<i>Mean variance: no short sales</i>	0.19 [0.33]	0.23 [0.36]	0.14 [0.24]	0.22 [0.32]	0.15 [0.31]	0.19	0.19
<i>Minimum variance</i>	0.41 [0.17]	0.82 [0.17]	0.06 [0.11]	0.18 [0.10]	-0.46 [0.19]	0.20	0.39
$1/N$	0.20 [0.00]	0.20 [0.00]	0.20 [0.00]	0.20 [0.00]	0.20 [0.00]	0.20	0.20
PORTFOLIO RESAMPLING							
<i>Resampling: no short sales</i>	0.20 [0.29]	0.25 [0.31]	0.17 [0.21]	0.24 [0.28]	0.15 [0.26]	0.20	0.20

Table 3.8: Monte-Carlo Average Out-of-Sample Utility for Alternative Portfolio Estimates

This table shows out-of-sample utility averaged across simulations for different portfolio selection models listed in Table 1.

Model	\bar{U}_{OOS}		
	$T = 60$	$T = 120$	$T = 240$
FULL INFORMATION			
<i>Mean variance</i>	0.49	0.49	0.49
<i>Global minimum variance</i>	0.20	0.20	0.20
PLUG IN			
<i>Mean variance</i>	-0.97 [0.98]	-0.25 [0.42]	0.03 [0.17]
<i>Mean variance diffuse prior</i>	-0.66 [0.73]	-0.19 [0.36]	0.04 [0.16]
<i>Best two fund</i>	-0.50 [0.60]	-0.15 [0.33]	0.05 [0.15]
<i>Estimated two fund</i>	-0.04 [0.25]	0.04 [0.17]	0.10 [0.09]
<i>Estimated three fund</i>	-0.05 [0.34]	0.09 [0.19]	0.17 [0.10]
<i>Mean variance - no short sales</i>	-0.13 [0.57]	0.09 [0.21]	0.16 [0.11]
<i>Minimum variance</i>	0.18 [0.06]	0.20 [0.06]	0.20 [0.06]
$1/N$	0.22 [0.06]	0.22 [0.06]	0.22 [0.06]
PORTFOLIO RESAMPLING			
<i>Resampling: no short sales</i>	-0.17 [0.67]	0.09 [0.23]	0.17 [0.11]

Table 3.9: Out-of-Sample Utility for Discrete-Bayesian Estimator

This table shows out-of-sample utilities that would have been realized by an investor using the discrete-Bayesian model with informed prior to set their allocation strategy. The investor has prior belief in the positivity of the portfolio weights. The investor does not believe that the mean of the asset return means is higher than that observed in the return sample. The investor rebalances every three months. The investor's data set at each investment month includes the previous T monthly returns. The three panels correspond to three data sets: the five- and ten-industry portfolios of Fama and French, and a portfolio of four-country indices from Morgan Stanley Capital International.

	T = 120		
	Five	Ten	Four
	Industries	Industries	Countries
JAN1962-JAN2007			
<i>Nonparametric Bayes</i>	-0.47	-0.86	
<i>Resampling with prior</i>	-0.50	-1.06	
JAN1972-JAN2007			
<i>Nonparametric Bayes</i>	-0.26	-0.48	
<i>Resampling with prior</i>	-0.21	-0.74	
JAN1985-DEC2006			
<i>Nonparametric Bayes</i>	-0.22	-0.39	0.03
<i>Resampling with prior</i>	-0.19	-0.49	-0.01
JAN1995-DEC2006			
<i>Nonparametric Bayes</i>	-0.23	-0.60	-0.20
<i>Resampling with prior</i>	-0.30	-0.66	-0.33

Table 3.9: (Continued)

	T = 240		
	Five	Ten	Four
	Industries	Industries	Countries
JAN1972-JAN2007			
<i>Nonparametric Bayes</i>	-0.12	-0.16	
<i>Resampling with prior</i>	-0.07	-0.22	
JAN1985-DEC2006			
<i>Nonparametric Bayes</i>	0.04	-0.00	
<i>Resampling with prior</i>	0.10	-0.08	
JAN1995-DEC2006			
<i>Nonparametric Bayes</i>	0.04	0.08	0.03
<i>Resampling with prior</i>	0.10	0.05	0.07

Table 3.10: Out-of-Sample Utility for Alternative Models

This table shows out-of-sample utilities that would have been realized by an investor using the discrete-Bayesian model with informed prior to set their allocation strategy. The investor has prior belief in the positivity of the portfolio weights. The investor does not believe that the mean of the asset return means is higher than that observed in the return sample. The investor rebalances every three months. The investor's data set at each investment month includes the previous T monthly returns. The three panels correspond to three data sets: the five- and ten-industry portfolios of Fama and French, and a portfolio of four-country indices from Morgan Stanley Capital International.

Table 3.10: (Continued)

	T = 120		
	Five	Ten	Four
	Industries	Industries	Countries
JAN1962-JAN2007			
<i>Mean variance</i>	-0.81	-1.68	
<i>Estimated two fund</i>	-0.20	-0.22	
<i>Estimated three fund</i>	-0.09	-0.06	
<i>Mean variance - no short sales</i>	-0.37	-0.60	
<i>Minimum variance</i>	0.04	0.09	
<i>1/N</i>	0.04	0.09	
<i>Average - no short sales</i>	-0.36	-0.62	
JAN1972-JAN2007			
<i>Mean variance</i>	-0.72	-1.40	
<i>Estimated two fund</i>	-0.10	-0.07	
<i>Estimated three fund</i>	-0.09	-0.03	
<i>Mean variance - no short sales</i>	-0.19	-0.30	
<i>Minimum variance</i>	0.16	0.25	
<i>1/N</i>	0.07	0.13	
<i>Average - no short sales</i>	-0.19	-0.31	
JAN1985-DEC2006			
<i>Mean variance</i>	-0.35	-1.15	-0.19
<i>Estimated two fund</i>	-0.03	-0.04	-0.01
<i>Estimated three fund</i>	-0.00	0.03	0.01
<i>Mean variance - no short sales</i>	-0.07	-0.21	-0.06
<i>Minimum variance</i>	0.35	0.39	0.35
<i>1/N</i>	0.34	0.36	0.29
<i>Average - no short sales</i>	-0.08	-0.24	-0.08
JAN1995-DEC2006			
<i>Mean variance</i>	-0.40	-1.06	-0.30
<i>Estimated two fund</i>	-0.08	-0.08	-0.09
<i>Estimated three fund</i>	-0.07	-0.00	-0.03
<i>Mean variance - no short sales</i>	-0.20	-0.45	-0.19
<i>Minimum variance</i>	0.34	0.27	0.24
<i>1/N</i>	0.40	0.41	0.11
<i>Average - no short sales</i>	-0.22	-0.47	-0.22

Table 3.10: (Continued)

	T = 240		
	Five	Ten	Four
	Industries	Industries	Countries
JAN1972-JAN2007			
<i>Mean variance</i>	-0.30	-0.57	
<i>Estimated two fund</i>	-0.11	-0.10	
<i>Estimated three fund</i>	-0.10	-0.03	
<i>Mean variance - no short sales</i>	-0.10	-0.14	
<i>Minimum variance</i>	0.17	0.27	
<i>1/N</i>	0.07	0.13	
<i>Average - no short sales</i>	-0.07	-0.08	
JAN1985-DEC2006			
<i>Mean variance</i>	0.06	-0.35	
<i>Estimated two fund</i>	0.03	-0.01	
<i>Estimated three fund</i>	0.02	0.05	
<i>Mean variance - no short sales</i>	0.16	0.13	
<i>Minimum variance</i>	0.43	0.44	
<i>1/N</i>	0.34	0.36	
<i>Average - no short sales</i>	0.19	0.20	
JAN1995-DEC2006			
<i>Mean variance</i>	-0.00	-0.48	0.07
<i>Estimated two fund</i>	0.03	-0.05	0.08
<i>Estimated three fund</i>	0.00	0.05	0.06
<i>Mean variance - no short sales</i>	0.20	0.07	0.05
<i>Minimum variance</i>	0.52	0.35	0.16
<i>1/N</i>	0.40	0.41	0.11
<i>Average - no short sales</i>	0.24	0.19	0.05

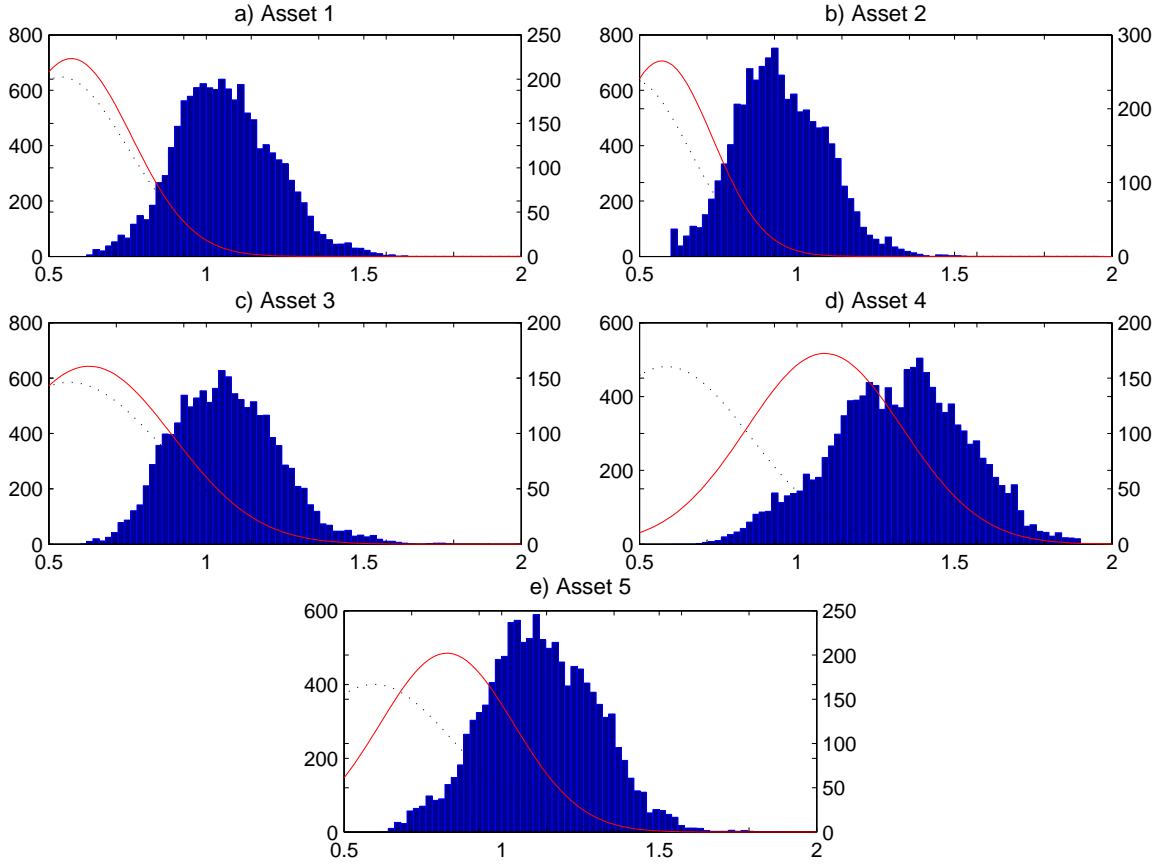


Figure 3.1: Histogram of Posterior Returns for a Simulated Data Set

This figure shows histograms of mean returns across 12500 Markov chain Monte Carlo draws from the posterior distribution of an informed investor. The return data is a simulated series of returns of length $T = 120$ from a five dimensional multivariate normal distribution. The parameters of the underlying distribution are chosen to reflect historical values for the five-industries data set of Fama and French. The investor has a prior that reflects a belief in the positivity of returns. The investors prior coefficients are $c_1 = 20$, $c_2 = 0$, and $c_3 = 0$. The line plots are probability density functions of the returns of the generating distribution (dotted) and the maximum likelihood normal distribution given the sample (solid).

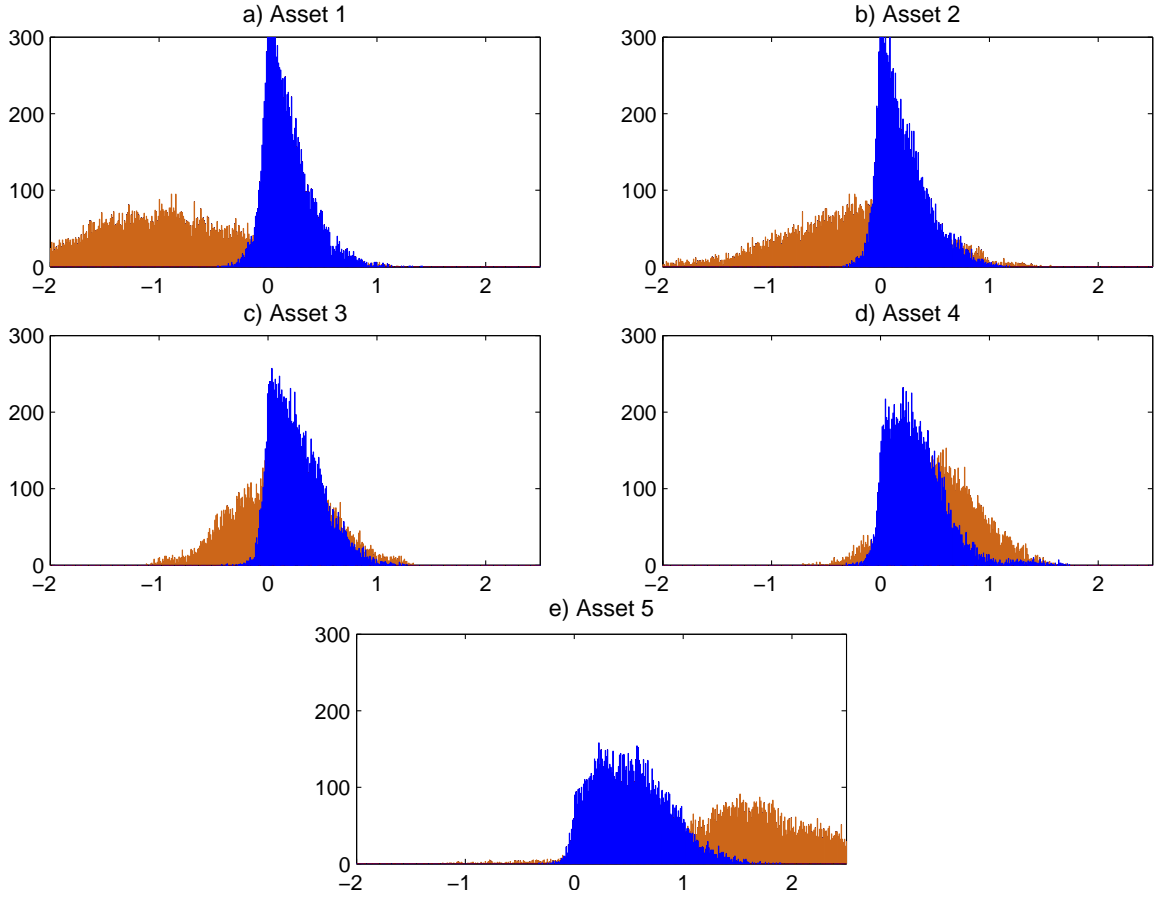


Figure 3.2: Histogram of Posterior Weights

This figure shows histograms of optimal portfolio weights implied by 12500 distributions drawn from the posterior distribution with prior belief in the positivity of portfolio weights. The return data is a simulated series of returns of length $T = 120$ from a five dimensional multivariate normal distribution. The parameters of the underlying distribution are chosen to reflect historical values for the five industries data set of Fama and French. The posterior is sampled via a Markov chain Monte Carlo algorithm. The investor has a prior that reflects a belief in the positivity of returns. The investor's prior coefficients are $c_1 = 1$, $c_2 = 20$, and $c_3 = 0$. The rust colored histograms in the background depict histograms of optimal portfolio weights under the no-arbitrage prior only ($c_1 = 1$, $c_2 = 0$, and $c_3 = 0$).

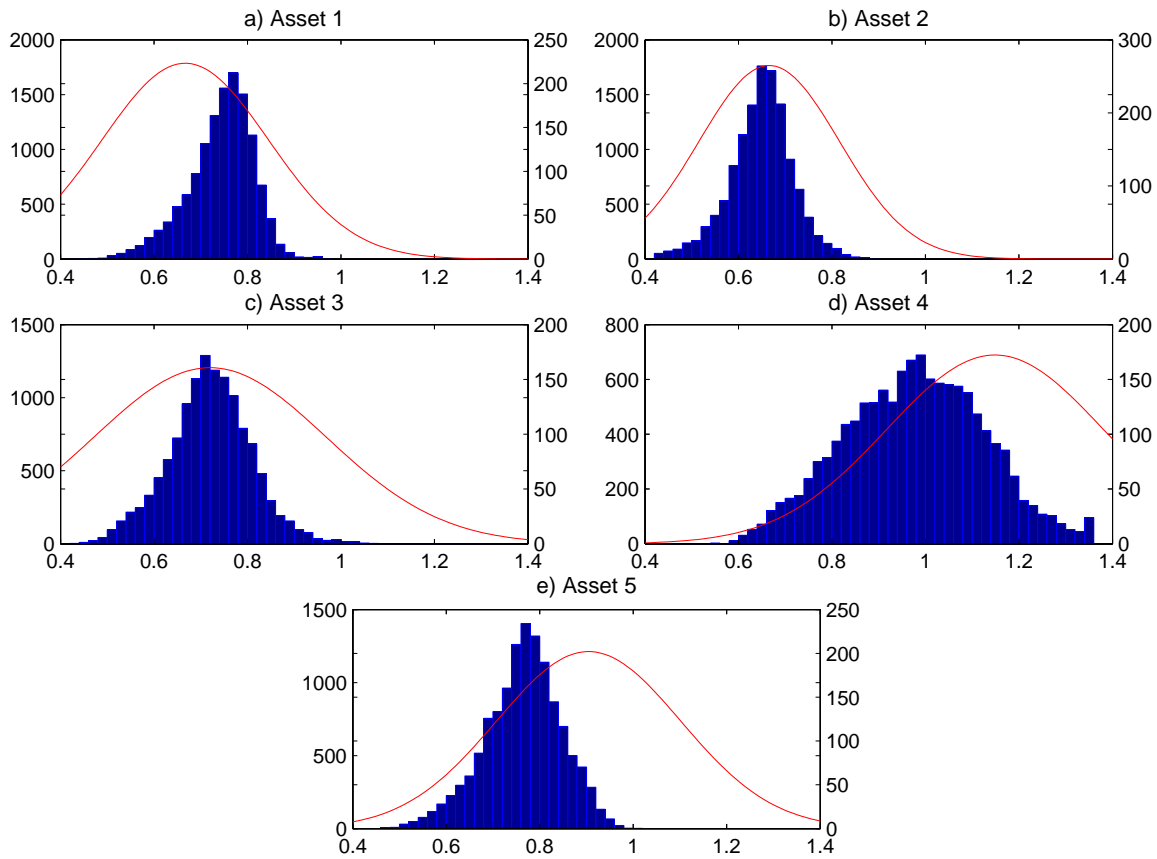


Figure 3.3: Histogram of Posterior Returns with Prior Against Return Bias

This figure shows histograms of mean returns across 12500 Markov chain Monte Carlo draws from the posterior distribution of an informed investor. The return data is a simulated series of returns of length $T = 120$ from a five dimensional multivariate normal distribution. The parameters of the underlying distribution are chosen to reflect historical values for the five industries data set of Fama and French. The investor has a prior that reflects a belief in the positivity of returns. The investors prior coefficients are $c_1 = 1$, $c_2 = 20$, and $c_3 = 100$. The line plots is the maximum-likelihood normal distribution given the sample.

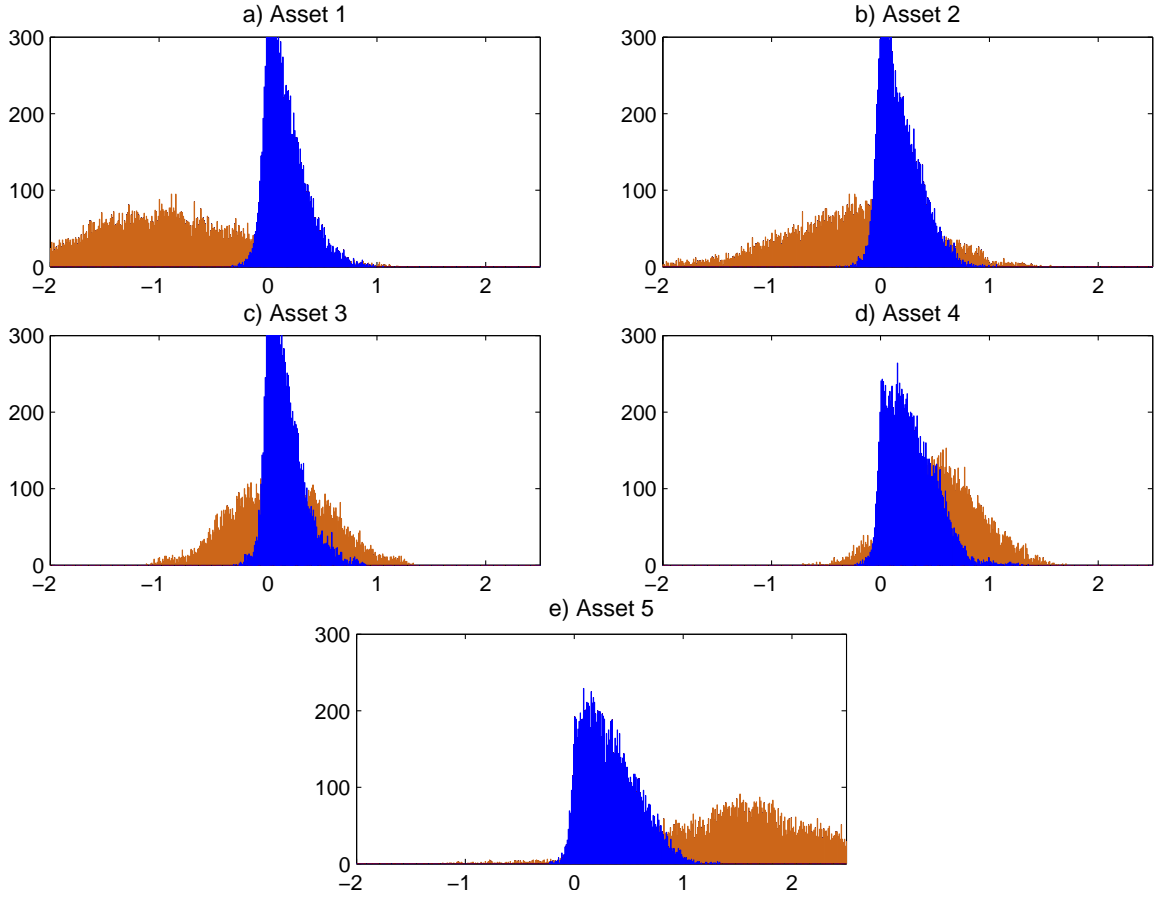


Figure 3.4: Histogram of Posterior Weights with Prior Against Return Bias

This figure shows histograms of weights implied by distributions drawn from the posterior. 12500 Markov chain Monte Carlo draws from the posterior distribution of an informed investor. The return data is a simulated series of returns of length $T = 120$ from a five dimensional multivariate normal distribution. The parameters of the underlying distribution are chosen to reflect historical values for the five-industries data set of Fama and French. The investor has a prior that reflects a belief in the positivity of returns. The investor's prior coefficients are $c_1 = 1$, $c_2 = 20$, and $c_3 = 100$.

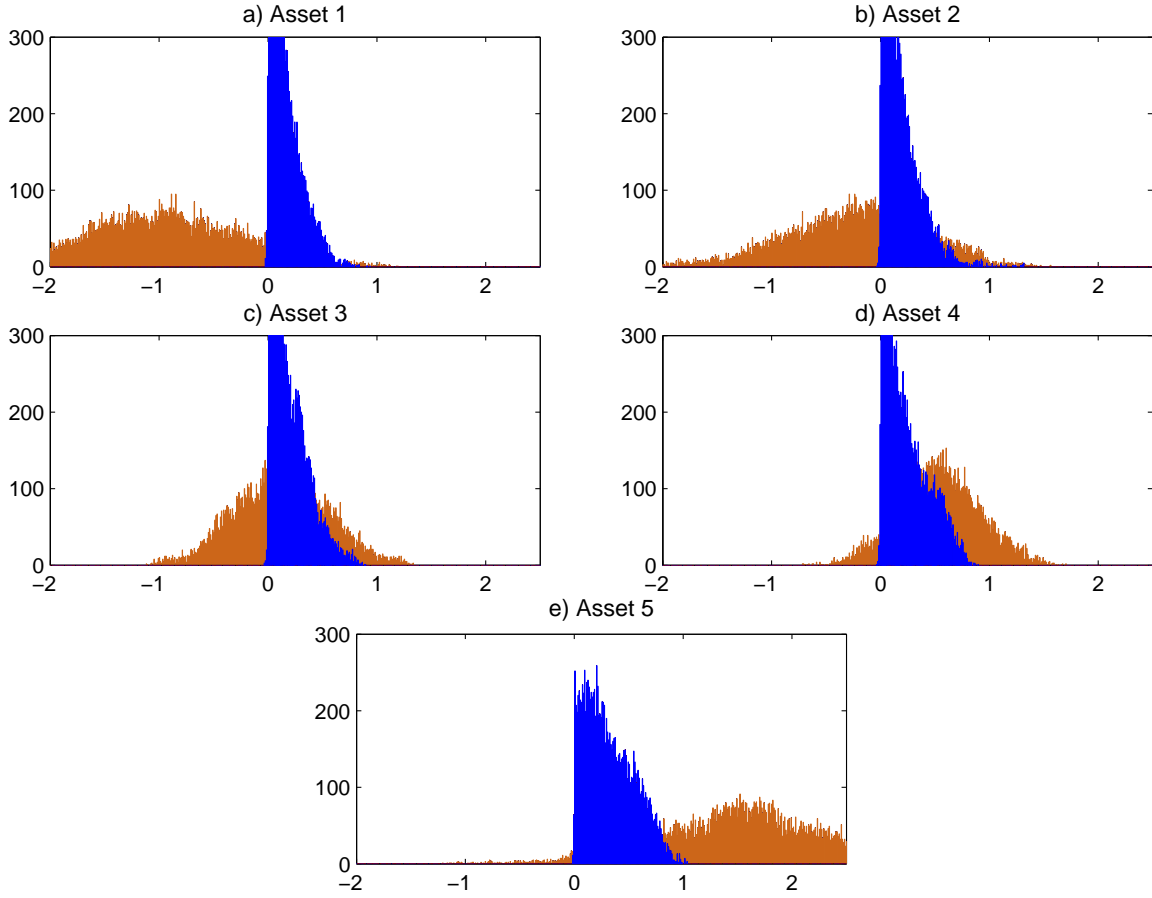


Figure 3.5: Histogram of Posterior Weights with Strong Prior

This figure shows histograms of weights implied by distributions drawn from the posterior. 12500 Markov chain Monte Carlo draws from the posterior distribution of an informed investor. The return data is a simulated series of returns of length $T = 120$ from a five dimensional multivariate normal distribution. The parameters of the underlying distribution are chosen to reflect historical values for the five industries data set of Fama and French. The investor has a prior that reflects a belief in the positivity of returns. The investor's prior coefficients are $c_1 = 200$, $c_2 = 1000$ and $c_3 = 0$.

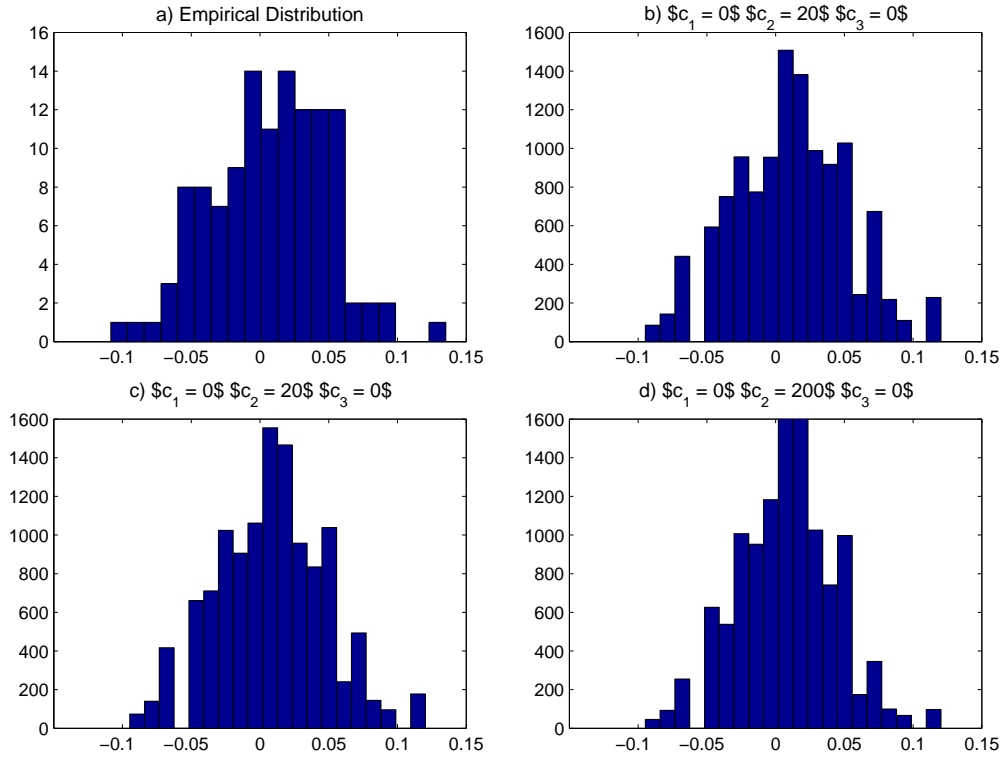


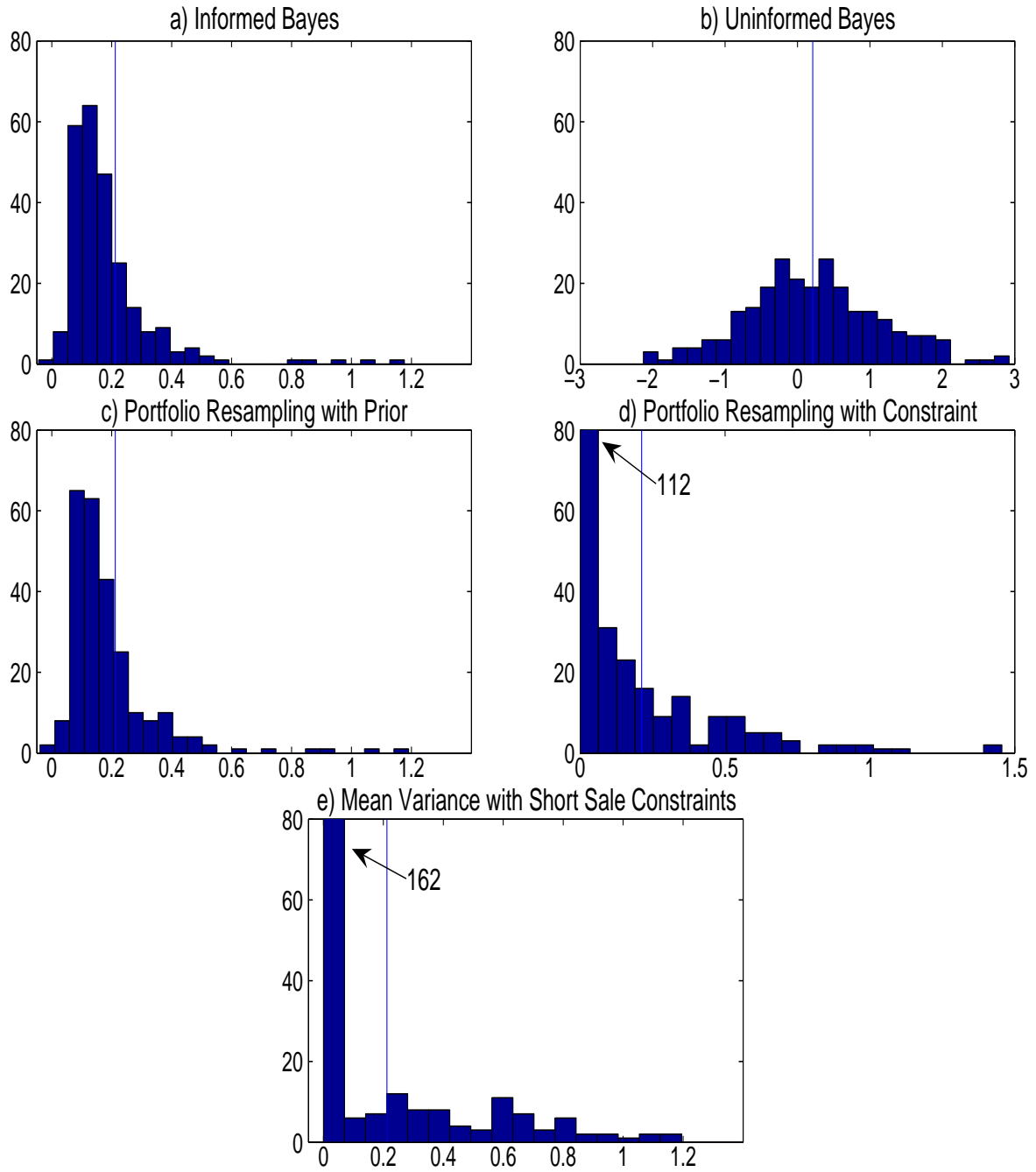
Figure 3.6: Histogram of Predictive Return Distribution under Different Priors

This figure shows histograms of predictive distributions under different prior assumptions for a simulated return series. The predictive distributions are based on observation weights computed by summing over 12500 distributions drawn from the discrete-Bayesian posterior.

Figure 3.7: Histogram of Asset Weights Across Simulations

This figure shows histograms of estimated optimal allocations across a set of 250 simulated historical return series. The figures show an investor's allocation to the first asset in a five asset problem. The returns are simulated from a multivariate normal distribution with mean and covariance parameters that are reflective of those of United States industry portfolios. The investor's risk aversion parameter is 5. For the resampling methods, 12500 distributions are drawn by Markov Chain Monte Carlo from the set of discrete distributions whose domain is the sample returns. Plot (a) shows portfolio weights estimated by an informed Bayesian who has prior views against the presence of negative portfolio weights, and believes the average of the expected returns across assets is close to the average return across assets in the sample. The Bayesian computes expectations by averaging over resampled distributions that take into account their priors. Plot (b) shows portfolio estimated by an uninformed Bayesian. Plots (c) and (d) shows portfolio weights estimated by resampling. For plot(c), the investor uses the same set of distributions as the informed Bayesian, and computes portfolio weights for each draw with short sale constraints imposed, and then takes on average over the 'resampled' weights. For plot(d), the investor uses the same set of distributions as the uninformed Bayesian. Plot (e) shows portfolio weights estimated by an investor who use sample values to proxy for expected returns and means, but imposes short sale constraints on the risky asset weights.

Figure 3.7 (continued)



Bibliography

Aït-Sahalia, Y. and Brandt, M. (2001). Variable selection for portfolio choice. *Journal of Finance*, 56:1297–1351.

Avramov, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics*, 64:423–458.

Barberis, N. (2000). Investing for the long run when returns are predictable. *Journal of Finance*, 55:225–264.

Bawa, V. S., Brown, S., and Klein, R. (1979). *Estimation Risk and Optimal Portfolio Choice*. North Holland, Amsterdam.

Best, M. J. and Grauer, R. R. (1991). On the sensitivity of mean-variance efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies*, 4:315–342.

Brandt, M. (1999). Estimating portfolio and consumption choice: A conditional Euler equations approach. *Journal of Finance*, 54:1609–1645.

Brandt, M. W. (2005). Portfolio choice problems. In Ait-Sahalia, Y. and Hansen, L. P., editors, *Handbook of Financial Econometrics*. Elsevier Science, Amsterdam.

Braun, W. J. and Hall, P. (2001). Data sharpening for nonparametric inference subject to constraints. *Journal of Computational and Graphical Statistics*, 10(4):786–806.

Brown, S. J. (1976). *Optimal Portfolio Choice under Uncertainty: A Bayesian Approach*. PhD thesis, University of Chicago.

Chevrier, T. and McCulloch, R. (2008). Using economic theory to build optimal portfolios. Working Paper.

- Choi, E. and Hall, P. (1999). Data sharpening as a prelude to density estimation. *Biometrika*, 86:941–947.
- Choi, E., Hall, P., and Rousson, V. (2000). Data sharpening methods for bias reduction in nonparametric regression. *The Annals of Statistics*, 28(5):1339–1355.
- Cremers, K. J. M. (2002). Stock return predictability: a Bayesian model selection perspective. *Review of Financial Studies*, 15(4):1223–1249.
- DeMiguel, V., Garlappi, L., Nogales, F., and Uppal, R. (2008). Improving performance by constraining portfolio norms: A generalized approach to portfolio optimization. Working Paper.
- DeMiguel, V., Garlappi, L., and Uppal, R. (2007). Optimal versus naive diversification: How inefficient is the $1/n$ portfolio strategy? *Review of Financial Studies*, Forthcoming.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7(1):1–26.
- Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo*. Chapman and Hall, New York, second edition edition.
- Garlappi, L., Uppal, R., and Wang, T. (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies*, 20(1):41–81.
- Green, R. C. and Hollifield, B. (1992). When will mean-variance efficient portfolios be well diversified? *Journal of Finance*, 47(5):1785–1809.
- Hall, P. and Presnell, B. (1999). Intentionally biased bootstrap methods. *Journal of the Royal Statistical Society Series B*, 61:143–158.
- Harvey, C. R., Liechty, J. C., Liechty, M. W., and Muller, P. (2003). Portfolio selection with higher moments. Working Paper.
- Herold, U. and Maurer, R. (2006). Portfolio choice and estimation risk. a comparison of bayesian and heuristic approaches. *Astin Bulletin*, 36(1):135–160.

- Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 58:1651–1684.
- Jobson, J. D. and Korkie, B. M. (1981). Performance hypothesis testing with the Sharpe and Treynor measures. *Journal of Finance*, 36(4):889–908.
- Jorion, P. (1986). Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21:279–292.
- Jorion, P. (1992). Portfolio optimization in practice. *Financial Analysts Journal*, 48(1):68–74.
- Kacperczyk, M. (2003). Asset allocation under distribution uncertainty. Working Paper.
- Kan, R. and Zhou, G. (2007). Portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3):621–656.
- Kandel, S. and Stambaugh, R. F. (1996). On the predictability of stock returns: An asset allocation perspective. *Journal of Finance*, 51:385–424.
- Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7:77–91.
- Markowitz, H. M. and Usmen, N. (2003). Resampled frontiers versus diffuse bayes: An experiment. *Journal of Investment Management*, 1(4):9–25.
- Meulli, A. (2006). *Risk and Asset Allocation*. Springer, Berlin-Heidelberg-New York.
- Michaud, R. O. (1998). *Efficient Asset Management*. Harvard Business School Press, Boston.
- Pástor, Ľ. (2000). Portfolio selection and asset pricing models. *Journal of Finance*, 55:179–223.
- Pástor, Ľ. and Stambaugh, R. (1999). Costs of equity capital and model mispricing. *Journal of Finance*, 54:67–121.
- Paye, B. (2004). *Essays on Stock Return Predictability and Portfolio Allocation*. PhD thesis, University of California San Diego.

Scherer, B. (2002). Portfolio resampling: Review and critique. *Financial Analysts Journal*, 58:98–109.

Wachter, J. A. and Warusawitharana, M. (2005). Predictable returns and asset allocation: Should a skeptical investor time the market? Working Paper.

Zellner, A. and Chetty, V. K. (1965). Prediction and decision problems in regression models from the bayesian point of view. *Journal of the American Statistical Association*, 60(310):608–616.

Chapter 4

Conclusion

In this thesis, I examine whether supplementing historical returns data yields a significant improvement in risk-adjusted expected return for an investor. In Chapter 2, I find that return variability can severely limit expected gains from diversification and conditioning on predictor variables. In Chapter 3, I demonstrate that, even when granted a prior that is correct, an investor may have trouble attaining risk-adjusted out-of-sample return that matches that of biased estimators such as the $1/N$ model (DeMiguel et al. (2007)). In this chapter, I discuss my results in relation to recent research and describe some promising avenues for future research.

The results of both Chapters 2 and Chapters 3 reflect the difficulty of estimating optimal portfolios from limited information available for return forecasting. Even with correct additional information, it is difficult to outperform biased estimates such as short-sales constrained, minimum-variance and $1/N$ portfolios. However, the results should not discourage an investor from incorporating as much information as they have available into their portfolio choice problem.

It is important to differentiate the underlying motivation and objective of the conditional estimation models of Chapters 2 and 3 and smoothing estimators such as the $1/N$ model. The Bayesian approach aims to obtain the best estimate possible of the underlying distribution of returns by supplementing the information provided by the data.

In contrast, the $1/N$ model and models that impose short sales constraints are examples of smoothing or regularization techniques. Regularization techniques aim to improve the robustness of estimates by smoothing the optimization solution. This smoothing is achieved by adding functions that penalize solution variability to the Lagrangian of the optimization problem.⁶⁹ Statis-

⁶⁹Or, equivalently, placing constraints on the solution (see Jagannathan and Ma (2003)). Hence, short sales constraints are an example of a regularization technique.

ticians use smoothing techniques to deal with estimation problems for which data is limited or highly collinear. Portfolio allocation qualifies as just such a problem.⁷⁰ Asset returns are often highly collinear. For example, DeMiguel et al. (2007) demonstrate that a return history of 250 years would be required to match the $1/N$ strategy for an example problem with 25 assets.

A key distinguishing feature of Bayesian and regularization methods is that Bayesian estimators are consistent whereas regularization may be biased even in large sample. Also, the aim of Bayesian methods is to improve the distribution estimate, whereas regularization aims to improve the robustness of small sample estimates by constraining the solution vector. Recent research has focused on implementations based on one or the other of these approaches. For example, Chevrier and McCulloch (2008), Tu and Zhou (2008), and Chapter 3 of this thesis, consider economically motivated priors in a Bayesian context, whereas DeMiguel et al. (2008) focus on improving estimates by constraining norms of the vector of portfolio weights.

This division between Bayesian portfolio estimates and those based on informed priors is artificial. The objective of the model should be to first obtain the best estimate of the underlying return distribution that is possible given the information at hand, and, second, to then apply an appropriate regularization if the effective sample size given the data and the other information remains too small to obtain consistently robust solutions.

To illustrate, we can add a regularization term to the estimator first discussed in Chapter 1 (see equation 1.2) and applied throughout this thesis.

$$\max_w \sum_{t=1}^T a_t U(\tilde{r}_t, w) - \kappa \|w\|^p, \quad (4.1)$$

where κ is a parameter that determines the strength of the regularization and $p > 0$ determines the norm.⁷¹

It seems likely that the optimal portfolio rule will be one that incorporates all the best available information plus some regularization. This suggests an interesting avenue for future research into both the impact of predictability and

⁷⁰Optimal portfolio weights can be expressed as parameters of a least squares regression (Britten-Jones (1999)).

⁷¹A p-norm regularization is used for illustration. Other regularization functions are possible. See equation 1.2 and surrounding text for definitions of other notation.

the importance of the positive-weights prior. Would out-of-sample performance be improved by an estimator of the form 4.1?

On a related note, the estimation formulation 4.1 suggests an alternative means of assessing the impact of new information on portfolio estimates. The current paradigm for assessing a portfolio rule that incorporates new information is to compare its performance to various ad hoc regularized portfolios such as minimum variance, $1/N$, or short-sales constrained. Instead, one might ask the more subtle question of whether incorporation of the information reduces the degree of regularization required to obtain an admissible solution, i.e., does the new information allow us to obtain equivalent or improved out-of-sample performance as the previous regularized solution but with a reduced weight on the regularization penalty function. Given the inherent instability of the portfolio problem, this softer criterion is likely to be a more suitable basis for inference when analyzing the economic significance of information beyond that provided by historical returns.

Bibliography

Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance*, 54(2):655–671.

Chevrier, T. and McCulloch, R. (2008). Using economic theory to build optimal portfolios. Working Paper.

DeMiguel, V., Garlappi, L., Nogales, F., and Uppal, R. (2008). Improving performance by constraining portfolio norms: A generalized approach to portfolio optimization. Working Paper.

DeMiguel, V., Garlappi, L., and Uppal, R. (2007). Optimal versus naive diversification: How inefficient is the $1/n$ portfolio strategy? *Review of Financial Studies*, Forthcoming.

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 58:1651–1684.

Tu, J. and Zhou, G. (2008). Incorporating economic objectives into bayesian priors: Portfolio choice under parameter uncertainty. Working Paper.

Appendix A

Bandwidth Specification

There is considerable leeway in the specification of the weighting function and bandwidth used in the non-parametric estimator. However, the properties of the solution are much more sensitive to choice of bandwidth than they are to choice of weighting function⁷². I set the weighting function equal to the density function of the normal distribution⁷³.

Bandwidth choice is a classic example of the tradeoff between bias and variance. If the bandwidth is very narrow then few observations receive any weight in the estimator for a given value z of the conditioning variable. This leads to high variance. As the bandwidth is broadened, variance of the estimator drops as more data points receive significant weights, but bias increases as points from states farther away from the z of interest gain more weight in the estimator. A standard approach to balancing this tradeoff is to minimize mean squared error of the estimates. Consistent with Brandt (1999) I consider bandwidths of the form

$$h_T = \lambda \sigma T^{\left(\frac{-1}{K+4}\right)}, \quad (\text{A.1})$$

where σ is the unconditional variance of the state variable, and λ is a complicated function of unobservable functions of the moment conditions and conditioning variable.

In practice, there are a number of methods of choosing λ , and none is ideal. For example Brandt (1999) uses leave-one-out cross validation. For every observation z_t , $\alpha(z_t)$ is estimated both with and without including z_t in the data set. The bandwidth parameter λ is selected that minimizes the sum of the squared differences between the two estimates. My experiments (not reported) indicate that the λ selected by leave-one-out cross validation tends to be overly conservative. Instead I make use of a rule of thumb proposed by Silverman (1986).

⁷²See Hardle (1990) for a summary and discussion of the relative importance of weighting function and bandwidth to asymptotic convergence of nonparametric regression estimators.

⁷³This is a special case of the product of normal densities used by Brandt (1999).

The standard deviation σ in the above formula is replaced by the interquartile range R , and bandwidths are set to $h_T = \lambda R T^{-1/(K+1)}$. I report results for $\lambda = 0.79$ which is consistent with Paye (2004). Recalculation of results of, for example, figures 1-11, for bandwidths within 25% does not significantly change the results.

Bibliography

Brandt, M. (1999). Estimating portfolio and consumption choice: A conditional Euler equations approach. *Journal of Finance*, 54:1609–1645.

Hardle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, New York.

Paye, B. (2004). *Essays on Stock Return Predictability and Portfolio Allocation*. PhD thesis, University of California San Diego.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York.

Appendix B

Data Tuning for a Portfolio Choice Problem

This appendix presents two formulations that make use of data tuning to impose the restriction that all portfolio weights are positive.

Assume an investor seeking to choose allocations to N assets plus a risk free asset. The standard investor's problem is

$$\max_w E[U(\tilde{r}, w)], \quad (\text{B.1})$$

where \tilde{r} is a random return over the investment period and w is a vector of portfolio allocations.

Suppose we wish to find a solution to the above problem under the constraint that all weights are positive. The standard approach is to impose the constraint $w \geq 0$. Data tuning is an alternative approach that seeks to adjust the data set until the solution to the unconstrained problem satisfies the nonnegativity requirement for the portfolio allocations. Two approaches have been proposed.

The first is to shift the probabilities associated with each observation until the most likely multinomial distribution is found that satisfies the constraint. This approach is applied to bootstrap inference by Hall and Presnell (1999). The formulation is given by

$$\max \sum_{t=1}^T \log p_t \quad (\text{B.2})$$

subject to

$$\begin{aligned} w^* &= \arg \max \sum_{t=1}^T p_t U(r_t, w) > 0. \\ \sum_{i=1}^T p_t &= 1 \end{aligned} \tag{B.3}$$

The second approach is to perturb the data vector the minimum distance required to yield a weight vector that satisfies the required constraint. Choi and Hall (1999) and Choi et al. (2000) refer to this latter approach as data sharpening and provide analysis and applications. In this case, the perturbed data set $S = [s_1 : \dots : s_T]$ is found by solving

$$\min_S \mathbf{D}(R - S) \tag{B.4}$$

subject to

$$w^* = \arg \max \sum_{t=1}^T U(s_t, w) > 0,$$

where $\mathbf{D}(\cdot)$ is a distance metric and $R = [r_1 : \dots : r_T]$ is the $N \times T$ matrix of return data.

Bibliography

Choi, E. and Hall, P. (1999). Data sharpening as a prelude to density estimation. *Biometrika*, 86:941–947.

Choi, E., Hall, P., and Rousson, V. (2000). Data sharpening methods for bias reduction in nonparametric regression. *The Annals of Statistics*, 28(5):1339–1355.

Hall, P. and Presnell, B. (1999). Intentionally biased bootstrap methods. *Journal of the Royal Statistical Society Series B*, 61:143–158.

Appendix C

Sampling from the Posterior

I sample from the posterior (3.8) using an implementation of the Metropolis Hastings algorithm.⁷⁴ As described in Gamerman and Lopes (2006), an arbitrary posterior can be sampled by drawing from a Markov chain that has the posterior as a limiting distribution. The Metropolis Hastings is a means of constructing such a chain for arbitrary posteriors. The Markov chain is generated by stepping through the parameter space. At each step a proposal distribution is drawn from a transition kernel $\pi(\xi|\delta)$. The chain steps to the proposed draw with probability given by the acceptance level

$$a(\xi, \delta) = \min \left\{ 1, \frac{q_{\text{post}}(\delta|X)\pi(\xi|\delta)}{q_{\text{post}}(\xi|X)\pi(\delta|\xi)} \right\}. \quad (\text{C.1})$$

If the transition is rejected, the chain remains at the current distribution δ until the next step.

The transition kernel must allow for efficient spanning of a large dimensional parameter space, and be computationally feasible. I construct such a kernel as a mixture of a discrete uniform distribution over the set of observation indices $\{1, \dots, T\}$ and a Beta distribution. The discrete draw $i \sim \text{UNIFORM}(\{1, \dots, T\})$ identifies an individual p_i . The proposal is then formed by drawing a new value of for p_i from a Beta distribution, and adjusting the remaining probabilities in δ such that the summation condition is satisfied.

The transition kernel has some similarity to sequential transition schemes that update individual parameter either one at a time or in blocks. However, while a single p_i the focus of the proposed jump, all the remaining components of the probability vector are also altered to satisfy the summation condition. As such, the kernel describes a standard Metropolis Hastings algorithm.

Let α and β be the parameters of the Beta distribution. The parameters

⁷⁴The high dimensionality of the parameter space renders more traditional direct sampling approaches computationally infeasible.

are set to achieve a distribution whose mode is equal to the current p_i . This is achieved by setting $\alpha = 1 + p_i/v$ and $\beta = 1 - p_i/v + 1/v$. When $\alpha > 0$ and $\beta > 0$, the mode of the β distribution is equal to $(\alpha - 1)/(\alpha + \beta - 2) = p_i$. The variance of the Beta distribution is also a function of its parameters,

$$\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = \frac{\left(1 + \frac{p_i}{v}\right) \left(1 + \frac{p_i}{v} + \frac{1}{v}\right)}{\left(2 + \frac{1}{v}\right) \left(3 + \frac{1}{v}\right)} \approx v \quad (\text{C.2})$$

where the approximate equality holds as long as both v and p_i are small. The latter condition is likely to hold if T is large since, in that case, any one p_i is likely to be small. The choice of v determines the average size of jumps in the parameter space and, by extension, the acceptance rate for proposal distributions. The acceptance rate must be small enough to ensure that the Markov chain does not get stuck at a single point, while at the same time the jumps must be large enough to ensure that the chain visits spans enough of the parameter space to ensure convergence in a reasonable number of iterations. Following recommendations in Gamerman and Lopes (2006), I choose v to obtain an acceptance rate between 0.35 and 0.5.

I use a standard sampling scheme. I generate a small number of Markov chains m . I sample every k step of each chain until I accumulate $2n$ samples from each chain. I drop samples from the first half of each chain to allow convergence of the chain to the limiting distribution. The end result is nm draws that can be used for inference based on the posterior and for computing the predictive distribution. I list pseudocode for the Markov chain Monte Carlo implementation in Algorithm 1.

Using convergence diagnostics described by Gamerman and Lopes [2006, p.196-7], I find that convergence is achieved for the portfolio choice problems considered in Section 3.6 with $m = 5$ and $n = 2500$.

```

for  $c = 1 : \# \text{ of chains}$  do
  Select a random start distribution  $\delta^0 = \{p_1^0, \dots, p_T^0\}$  ;
   $b = 0$ ;
  for  $n = 1 : kb$  do
    Select a random index  $i \in \{1, \dots, T\}$  ;
    Draw  $p_i^* \sim \text{Beta}(\epsilon_i^{n-1}, 1 + \epsilon_i^{n-1}, 1 - \epsilon_i^{n-1} + \epsilon_i^{n-1}/p_i^{n-1})$ 
      where  $\epsilon_i^s = p_i^s/v$  ;
    Compute  $m = \frac{1-p_i^*}{1-p_i^{n-1}}$  ;
    Set  $p_{-i}^* = p_{-i}^{n-1}m$  ;
    Compute posterior at proposal:  $q_{\text{post}}(\delta^*)$  ;
    Compute acceptance probability
      
$$r = \frac{q_{\text{post}}(\delta^*)}{q_{\text{post}}(\delta^{n-1})} = m^{T-1} \frac{p_i^*}{p_i^{n-1}} \frac{q_{\text{post}}(\delta^*)}{q_{\text{post}}(\delta^{n-1})} \frac{\tau(i, n-1, *)}{\tau(i, *, n-1)}$$

      where  $\tau(i, s, l) = \text{Betapdf}(p_i^l, 1 + \epsilon_i^s, 1 - \epsilon_i^s + \epsilon_i^s/p_i^s)$  ;
    Draw  $u \sim U(0, 1)$  ;
    if  $u < r$  then
       $\delta^b = \delta^*$ 
    end
    else
       $\delta^n = \delta^{n-1}$ 
    end
    if  $(n \bmod k) = 0$  then
       $b = b + 1$ ; Store  $\delta^b = \delta^n$ 
    end
  end
end

```

Algorithm 1: MCMC Algorithm for Drawing from the Discrete Posterior Distribution

The above lists the Markov Chain Monte Carlo algorithm for drawing from the posterior distribution of the informed Bayesian investor. The domain of the posterior encompasses discrete distributions that assign positive probabilities to T return outcomes. The notation is as follows. $\text{Beta}(\alpha, \beta)$ is the beta distribution with parameters α and β and $\text{Betapdf}(x, \alpha, \beta)$ is its associated probability distribution function. δ^n is the vector of posterior parameters $\{p_1^n, \dots, p_i^n, \dots, p_T^n\}$ indexed by n . p_{-i}^n is the set of parameters of δ^n with p_i excluded. b is total number of draws to be saved, and k is the frequency that the chain is sampled.

Bibliography

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo*. Chapman and Hall, New York, second edition edition.