Chapter 8
Portfolio Performance Assessment: Statistical Issues and Methods for Improvement

Bernell K. Stone

8.1 Introduction: Purposes and Overview

8.1.1 Performance Assessment Problems/Frameworks

Building on Markowitz (1952, 1959) mean–variance portfolio theory and the capital asset pricing model, Treynor (1965), Treynor and Mazuy (1966), Sharpe (1966), and Jensen (1968) set frameworks for portfolio performance assessment. Investment texts now all include chapters summarizing these measures. The crux of these performance assessment frameworks is assessing and explaining the amount that realized return exceeds a fair return for time and risk.

The primary focus of much performance measurement is after-the-fact assessment of how a managed portfolio performed relative to a before-tax fair return for time and risk. A related performance measurement problem is the task of evaluating well methods for active stock selection. The focus here is a full sample backtest of the performance potential of a stock return forecast. The assumed assessment structure is a panel framework for a time series of cross sections rank-ordered into fractile portfolios on the basis of a return forecast.

To assess with high statistical confidence the economic potential of a stock return forecast, the central backtest problem is to ensure that any apparent ability of a return forecast to predict future returns is well isolated from risk, tax, and other nonforecast return variables. The conventional methodology for correcting a cross section of realized returns for variation in risk is a multivariate regression using one

B.K. Stone (✉)
Global Energy Management, 4681 Windsor Drive
e-mail: BKS@BYU.EDU

of the standard return models. Estimating how well-realized returns or realized risky returns are explained by a return forecast and any of the APT or multivariate style models such as the Fama–French three-variable extension of the CAPM is fraught with measurement and specification problems, especially extreme multicollinearity problems.

8.1.2 Purposes

This chapter presents and illustrates the use of an alternative return forecast assessment framework. Rather than estimating a multivariate explanation of how realized returns or realized risky returns depend on a return forecast and other explanatory variables, the proposed alternative suppresses cross-sectional variation in the other explanatory variables by transforming the initial cross section of rank-ordered fractile portfolios into an associated cross section in which every portfolio has the same portfolio-weighted average value of each pertinent explanatory variable.

Response surface/subsurface statistical designs are intended to assess the response (dependent variable) to a treatment. Response surface/subsurface designs are widely used in controlled experiments, e.g., chemical synthesis, petroleum refining, nuclear reaction yield, etc. In fact, response surface/subsurface designs are the preferred statistical design framework for most controlled experiments.[1] Response surface/subsurface designs are also widely used in partially controlled experiments such as the illustrative example in Sect. 8.4.2 of health response to well-controlled variation in drug dosage administered to a patient sample designed to be matched on other sample attributes that could distort apparent response to the drug dosages.

Response surface/subsurface statistical designs and methods are not widely used in economics and finance (or social sciences generally) although response surface/subsurface methods are implicit in the extensive use of matched-sample and partially-matched-sample designs, especially in areas like marketing research and medicine. Regression and related econometric methods are the generally preferred

---

[1]There is a large literature on statistical designs for the empirical estimation of multivariate functional dependencies but primarily focused on controlled or partially controlled studies rather than the observational samples that typically arise in epidemiology, demographics, the social sciences including especially economics and business, medicine, and many physical sciences including astronomy. Treatment response studies were an early statistical design focus and continue to be an ongoing design estimation concern. Because of the early and ongoing concern for treatment response studies, it is common to use the term **response surface methods** to refer to empirical methods for estimating functional dependencies. Most of the statistical literature pertains to controlled or partially controlled experiments in process industries, medicine, and advertising. For background readers are referred to the foundation works by Box (1954) and Box and Draper (1959); to the review article by Myers, Khuri, and Cornell (1989); and to the books on response surface methods: Khuri and Cornell (1996), Box and Draper (1987), and Box, Hunter, and Hunter (2005).

statistical method for assessing return dependencies in asset pricing studies and other return-risk modeling. In addition to goodness of fit measures such as standard forecast error or mean absolute forecast error, the standard procedure for evaluating a return forecast, i.e., for assessing forecast value, is to use regression to evaluate how well the cross-section of realized returns for a pertinent stock sample and time period is explained by the return forecast. As in the case of many return dependency studies, it is standard procedure to rank order the sample stocks on the predicted return or an adjusted return prediction measure such as return in excess of estimated risk. When the realized return response for the rank-ordered sample is estimated via linear regression, the estimated slope coefficient is referred to as the "*information coefficient*." Return forecast value and reliability is indicated by the magnitude and significance of the information coefficient.

The key conceptual insight is to view a stock return forecast as a treatment applied to all the stocks in a sample that is designed to rank order the stocks on the basis of true performance corrected for specified risk, tax, and other return impact variables. Rather than matched controls via sample selection, the control matching for an observation sample of stocks is achieved after the fact by means of a power optimizing mathematical assignment program that transforms the initial cross section of forecast rank-ordered portfolios into an associated rank-ordered cross section matched on key controls.[2]

In addition to presenting the matched control framework, the intent here is a systematic structuring of the major design decisions for a panel study based on a time series of rank-ordered fractile portfolios. Compared to forecast performance assessment using multivariate regression, the matched control framework has substantial efficiency/power benefits. Compared to the multicollinearity distortion in a multivariate regression, control matching on all the pertinent distortion variables means no correlation between the forecast and any of the return control variables. For instance, having the same portfolio average value of a variable such as beta means no cross-sectional variation in beta and therefore no correlation between beta and the return forecast or in fact any other control variable. Control matching ensures that the cross-sectional impact of the forecast is well isolated from any distortion from any of the control variables, because each portfolio in the cross-section has the same security-weighted average value of each impact variable. Since there is no variation in the portfolio-average value of each control variable over the

---

[2]Most of the statistical design literature cited in the previous footnote focuses on controlled and especially partially controlled studies. The ability to adapt response surface methods to observational studies is developed in Stone, Adolphson, and Miller (1993). They extend the use of response surface methods to observational data for those estimation situations in which it is pertinent to group data, e.g., to group observations on individual stocks into portfolios or households into income percentiles. The use of control variables to assess a conditional dependency (response subsurface) is a basic technique in controlled experiments that we adapt in this study to obtain a well-isolated portfolio-level dependency of realized risky return on a return forecast. Fortunately for compatibility with other finance panel studies using rank-ordered grouping, the optimal control matching can be structured as a power/efficiency improvement to the widely used relative rank-ordering used in many return dependency studies.

rank-ordered return cross section, there can be no variation over the cross section in the impact of any matched control to the extent that its average value is accurately measured and is a good summary of the impact of each control on the portfolio return.

Power pertains to correctly assessing sample information. In the context of a full-sample return forecast assessment, there are three pertinent power attributes. The most critical is isolation of the forecast from other return impact variables. The other two pertain to assessing well the magnitude and significance of the apparent forecast response. Control matching ensures complete isolation of the return forecast from the impact of other variables to the extent that control variables are well measured and well summarized by their portfolio average values. Assessing well the magnitude and significance of the forecast response is achieved via the use of a reassignment algorithm that transforms an initial forecast rank ordering into an associated control-matched rank ordering using the reassignment algorithm formulated in Section 8.5. This algorithm optimizes a trade-off between two complementary power instruments – having a wide range of well-ordered forecast values and also having a high level of variable homogeneity within each fractile portfolio. *Variable homogeneity* refers to the extent to which values are close to the portfolio average. Variable homogeneity is measured in this research study by the variance relative to the mean for each fractile portfolio.

8.1.3   Chapter Organization

The rest of this chapter is organized as follows. Section 8.2 distinguishes between standard forecast accuracy measures such as standard forecast error and the ability of a forecast to predict return beyond a fair return for time and risk. Section 8.3 focuses on key statistical designs that determine power and efficiency in estimating performance potential for a well-isolated return forecast. Section 8.4 compares multivariate regression with control matching as alternative ways to assess the performance potential of a stock return forecast. Section 8.5 formulates a mathematical program to transform the starting, presumably correlation-distorted, rank-ordered cross section into an associated cross section of well-ordered, control-matched portfolios with zero correlation with a specified set of control variables. The objective function is to optimize a trade-off between two power measures—cross-sectional range and within-portfolio variable homogeneity on the rank-ordering variable. The decision variable is the amount of stock in each fractile portfolio that is assigned to one or more of the other portfolios. The key constraints are the control matching requirement that each portfolio in the cross section has the same portfolio-weighted average value of the specified control variables.

Sections 8.6, 8.7, and 8.8 provide an illustrative performance potential assessment using an implementation of the eight-variable return forecast model of Bloch, Guerard, Markowitz, Todd, and Xu (1993). Section 8.6 provides an overview of the return forecast model. Section 8.7 defines and discusses a set of firm-specific

control variables. With an emphasis on the elimination of risk and tax effects (both dividend–gain distortion and the tax shield of corporate debt), Sect. 8.8 illustrates the imposition of matched controls for different combinations of control variables. Section 8.9 summarizes conclusions and suggests issues for further research.

8.1.4 Overview of Some Key Results/Conclusions

In addition to optimizing statistical power, the imposition of a combination of risk and tax controls significantly increases statistical efficiency relative to the uncontrolled cross section. For the illustrative forecast model, tax effects are much larger than risk effects as measured by the three Fama–French risk variables: beta, size, and the book-to-market ratio. The three risk variables tend to smooth both the cross section of realized risky returns and especially the cross section of realized standard deviations. Both the return and the realized standard deviation cross sections are nonlinear. The cross section of realize standard deviations is not only nonlinear but highly nonmonotonic. As expected for a forecast designed to identify undervalued stocks (high upside potential with limited downside risk) versus overvalued stocks (limited upside potential with high downside risk), the distribution of realized returns about the average value exhibits significant skewness, negative skewness for low return forecasts, very little for the middle of the distribution, and very large significant skewness for the highest return forecasts.

Both the significant nonmonotonic cross section of realized standard deviations and the significant cross-sectional variation in realized skewness for the illustrative forecast attest to the importance of avoiding restrictive distributional assumptions. Overall, the matched control approach not only ensures a well-isolated return forecast but also provides significant improvements in both statistical efficiency and power compared to estimating a multivariate regression.

8.2 The Problem of Assessing the Performance Potential of a Stock Return Forecast

8.2.1 Forecast Accuracy/Significance Versus Performance Potential

Standard methodology for evaluating a forecast is to see how well actual values correspond to predicted values. For a stock return forecast, this realization versus forecast assessment translates into seeing how well-realized returns correspond to predicted returns. Standard forecast evaluation measures include the standard forecast error and the information coefficient. The standard procedure for computing an information coefficient is to estimate a linear fit of realized returns on predicted returns. The slope measures the information value of the forecast. A slope that

is insignificantly different from zero implies no forecast value. Of course, a significantly negative slope implies negative forecast value. For a significantly positive slope, the larger the estimated slope and the higher the R^2 (the more significant the estimated slope), the better the forecast.

In commenting on a return forecast model of Timmerman (2008a), Brown (2008) asserted that high statistical significance for predicting future returns does not ensure actual ability of a stock return forecast to provide a significant improvement in performance after correcting returns for risk and other systematic return variables. In response, Timmerman (2008b) concurred with Brown's assertion that the appropriate forecast assessment criterion is the ability to create superior portfolio-level performance. However, Timmerman observed that such an assessment was itself problematic. In particular, converting realized return from a forecast into realized return corrected for time and risk means that the assessment is a joint test of the forecast and an assumed fair return model. In questioning the ability to provide a risk correction with a high degree of confidence, Timmerman (2008b) assumes that correcting realized return for risk requires a fair return model with its associated limitations. The issues raised by Timmerman are indicative of a need for a specification-free, distribution-free alternative methodology for obtaining the well-isolated performance potential assessment demanded by Brown (2008).

8.2.2 Key Specification Issue: Eliminating/Controlling for Correlation Distortion

The primary requirement for isolating well true return forecast value from risk, taxes, and other return dependency variables is to ensure that there is no significant distortion from covariation between the return forecast and these other variables. While rank-ordered grouping into fractile portfolios can increase statistical efficiency by mitigating measurement error as discussed further in Sect. 8.3.2, rank-ordered grouping can exacerbate the problem of eliminating correlation distortion. Low sample-level correlation between the return forecast and other return impact variables can be greatly multiplied by the rank-ordered grouping.[3] For a sample of 1000 stocks rank-ordered into deciles, relatively low sample-level correlation coefficients of 0.05 to 0.10 can be magnified to portfolio-level correlation coefficients greater than 0.50 and even as high as 0.80. Getting the measurement error reduction

---

[3]The mathematics for correlation magnification is straightforward. The formula for the correlation coefficient between variables X and Y is covariance(X, Y)/[SD(X)SD(Y)], where SD stands for standard deviation. Ranking on variable X and grouping into fractile portfolios preserve a wide range of values for variable X. However, in each of the portfolios, the individual values of variable Y tend to average out to a value close to the sample average value. Thus, the portfolio-level standard deviation of Y is reduced, and for a very small number of portfolios (e.g., quintiles or deciles), SD(Y) tends to approach zero while the covariance in the numerator declines relatively slowly because of the wide range of portfolio-level values for the ranking variable X.

benefits of rank-ordered grouping clearly requires a statistical design that explicitly deals with correlation distortion.

8.2.3 Eliminating/Controlling for Systematic Tax Effects: Dividends Versus Gains

Tax effects associated with cross-sectional variation in the proportion of return realized as dividends and capital gains[4] are a clear omission from all the standard before-tax APT and multivariate style models. A forecaster seeking to beat any of the before-tax return models can generate a rolling time series of forecasts that exploit the dividend yield tilt, the well-known difference in the before-tax return of dividends and gains reflecting the differential dividend–gain taxation and possible differences in the systematic risk of dividends and gains.

Trying to reflect tax effects by adding a dividend yield term or other dividend–gain mix explanatory term to a before-tax return model can further exacerbate covariation resolution issues. For instance, dividend yield is correlated or partially correlated with beta, size, and especially value variables such as the book–price ratio.

8.3 A Framework for Optimal Statistical Design

8.3.1 Key Design Decisions

Statistical efficiency and power are two complementary dimensions of how well a researcher can extract information from a data sample. Key attributes of information extraction include (1) sample size and how well sample information is used; (2) the usual estimation issues of measurement error, specification error, and omitted variable distortion; and (3) breakdown in the assumptions that underlie estimation–inference tests. Of particular concern in isolating return forecast performance from other return dependency variables is covariation contamination that can significantly distort efforts to isolate the impact of the return forecast from other variable dependencies.

---

[4]Brennen (1970) shows that dividend yield is a significant omitted variable from the CAPM. Rosenberg (1974), Rosenberg and Rudd (1977), Rosenberg and Marathe (1979), Blume (1980), Rosenberg and Rudd (1982) and many subsequent researchers have empirically established the so-called dividend yield tilt. More recent studies include Peterson, Peterson, and Ang (1985), Fama and French (1988) and Pilotte (2003). For an extensive review of both dividend valuation and dividend policy and extensive references in this area, see Lease, Kose, Kalay, Loewenstein, and Sarig (2000).

Given the decisions on the forecast, the pertinent stock sample, and the time frame for the backtest assessment, the major power–efficiency decision is the methodology used to isolate the return forecast from other return impact variables. As already indicated, the conventional methodology is a multivariate regression. The alternative methodology advocated here is the use of matched controls.

8.3.2 The Number of Fractile Portfolios: Measurement Error Versus Power

Two key decisions for rank-ordering into fractile portfolios are the number of fractile portfolios in a cross section and the partitioning rule for deciding on the number of securities in each portfolio. These two decisions are critical for statistical efficiency related to measurement error versus three statistical power attributes: (1) portfolio sample size, (2) cross-sectional forecast range, and (3) within-portfolio variable homogeneity. *Within-portfolio variable homogeneity* refers to how well the portfolio average value represents the collection of stocks within the portfolio. Fewer fractile portfolios and therefore more stocks per portfolio mean that return dependency variables including the rank-ordering return forecast will generally have greater dispersion about the portfolio average value. Within-portfolio variance is a measure of the departure from the mean.[5]

In addition to reducing nonsystematic return variation, the primary reason for grouping stocks into fractile portfolios is to reduce measurement error. Early rank-ordered grouping paradigms established in tests of the capital asset pricing model in Fama and MacBeth (1973) and in Black, Jensen, and Scholes (1972) used rank-ordered grouping into deciles to mitigate beta measurement error. Individual stock betas have relatively large estimation errors and tend to change over time. To the extent that beta measurement errors are independent of each other,[6] estimation measurement error for a portfolio of 50 stocks is reduced to approximately 1/50 of the average stock-level measurement error.

---

[5]Other dispersion measures could be used, e.g., mean absolute deviation or interquartile range. Relative to these measures, within-portfolio variance gives greater weight to extreme departures from the portfolio average.

[6]Since beta measurement errors are known to regress toward the mean, the assumption of uncorrelated measurement errors used in the discussion here is almost certainly too strong for correctly assessing efficiency gains from reducing measurement error by grouping into rank-ordered fractile portfolios. In particular, beta values that are correlated with the forecast are very likely to have systematic variation in beta change values over the cross section. The control matching methodology developed in Sects. 8.4 and 8.5 will mitigate systematic changes in measurement error such as the well-known regression of betas toward the mean. When every portfolio in the cross section has the same beta value, each portfolio will have essentially the same ex post beta change. Hence, control matching provides the benefit of no systematic distortion from beta regression toward the mean.

In designing a study to evaluate forecast value, there is a clear need to trade-off the efficiency gains from measurement error reduction (see for instance Grunfeld and Griliches (1960) and Griliches (1986) for background) versus the information/power loss from reduced sample size and the associated power costs of reduced cross-sectional range (cross-sectional variance) and loss of within-portfolio variable homogeneity.

To illustrate the efficiency/power trade-off, consider partitioning a sample of 1000 stocks into equal-size portfolios of 200, 100, 50, and 20 stocks per portfolio. This partitioning results in 5, 10, 20, and 50 portfolios, respectively. Given that a portfolio of 50 and 20 stocks provides 98 % and 95 % of the measurement error reduction for uncorrelated measurement errors, collapsing the sample size to just ten portfolios (deciles) or even more extremely just five portfolios (quintiles) seems to be a clear case of excessive measurement error reduction relative to lost power from collapse of the sample size.

8.4 Isolation Methodology Alternatives: Multivariate Regression Versus Control Matching

8.4.1 Treatment Response Studies

To motivate intuition, it is useful to think of assessing stock return forecasting performance potential within the broad class of treatment response studies. A stock return forecast is a treatment applied to a sample of stocks to identify misvalued stocks. The assessment problem is to be sure that any apparent ability to rank order a stock sample on the basis of superior return is from the ability to separate true misvaluation from risk, taxes, and possibly other nonforecast return responses.

A *treatment response assessment* is a special case of empirically estimating a functional dependency.[7] The response is assumed to depend on the treatment and other explanatory variables. However, the usual concern in a treatment response assessment is not necessarily estimating the overall response dependency on the

---

[7]There is a large literature on statistical designs for the empirical estimation of multivariate functional dependencies but primarily focused on controlled or partially controlled studies rather than the observational samples that typically arise in epidemiology, demographics, the social sciences including especially economics and business, medicine, and many physical sciences including astronomy. Treatment response studies were an early statistical design focus and continue to be an ongoing design estimation concern. Because of the early and ongoing concern for treatment response studies, it is common to use the term *response surface methods* to refer to empirical methods for estimating functional dependencies. Most of the statistical literature pertains to controlled or partially controlled experiments in process industries, medicine, and advertising. For background readers are referred to the foundation works by Box (1954) and Box and Draper (1959); to the review article by Myers, Khuri, and Cornell (1989) and to the books on response surface methods: Khuri and Cornell (1996), Box and Draper (1987), and Box et al. (2005).

treatment and all other explanatory variables but rather ensuring that the estimated response to the treatment is well isolated from distortion from other explanatory variables. Using regression to estimate a multivariate response dependency is one way to assess treatment response. However, a good statistical design can significantly increase both efficiency and power by creating a series of treatment observations that are matched on the values of one or more of the nontreatment explanatory variables. The matched variables are called *control variables*, often shortened to *controls*. Matched controls eliminate treatment variation from all the controls.

While purely controlled experiments such as petroleum blending–refining and chemical synthesis are a widely studied class of treatment response statistical designs, partially controlled studies such as drug response and product market response to pricing are more pertinent analogs for assessing how realized risky return responds to a stock return forecast.[8]

8.4.2   Intuition Motivation: Isolating Well Treatment Response to Drug Dosage Variation

A test of an anti-inflammatory drug could look at dosage per unit of body weight to assess inflammation reduction. The prototypical treatment response study is conducted on double-blind subsamples, with treatments ranging from no drug (the placebo) and then a range of well-structured dosage increases up to a maximum safe treatment. Each subsample is selected to match each other as much as possible on control variables such as sex, age distribution, obesity distribution, blood pressure, etc. The term "control variable" here refers to attributes of the study population that can impact initial inflammation and inflammation response or otherwise distort assessment of response to the drug dosage. The creation of the subsamples from an initial study population is designed to provide identical values on the key control variables. If all subsamples are matched on all pertinent controls, then inflammation-dosage response is well isolated from variation in the control variables.

Because of initial subsample differences and especially because of withdrawals and disqualification of some of the treatment subjects, the cross section of efficacy assessment subsamples is usually not a perfect match on all controls. Researchers

---

[8]Most of the statistical design literature cited in the previous footnote focuses on controlled and especially partially controlled studies. The ability to adapt response surface methods to observational studies is developed in Stone et al. (1993). They extend the use of response surface methods to observational data for those estimation situations in which it is pertinent to group data, e.g., to group observations on individual stocks into portfolios or households into income percentiles. The use of control variables to assess a conditional dependency (*response subsurface*) is a basic technique in controlled experiments that we adapt in this study to obtain a well-isolated portfolio-level dependency of realized risky return on a return forecast.

assessing dosage response must deal with possible distortion from variation in the unmatched controls.

What is pertinent for intuition about statistical procedure for cross-sectional assessments of portfolio performance potential are alternative ways that researchers can deal with cross-sectional variation in unmatched controls to ensure that drug efficacy assessment is well isolated from distortion in unmatched controls. One alternative is to use a multivariate regression model that attempts to explain observed inflammation changes by a combination of dosage variation and a regression-estimated inflammation response to other variables. Given the very difficult problem of modeling inflammation response to other variables and given the generally small departure from a match on the intended controls, the preferred alternative to a multivariate regression assessment is a transformation of the unmatched cross section back into a new cross section of control-matched subsamples. In drug dosage studies, the sample transformation may be accomplished by holding out some subsample observations. For instance, subsamples with above- and below-average obesity values can remove some above-average and below-average obesity observations from the respective subsamples.

Rather than giving some sample observations a de facto weight of zero, the rematching problem is usually structured as an optimization problem in which observations in each subsample are reweighted. The objective is to find the overall reweighting that minimizes the reduction in sample size (measured as the overall departure of all observation weights from one) while producing a sufficiently near match on each pertinent control variable. The benefit of the control rematching approach is that drug efficacy assessment is reduced to the intended evaluation of a univariate response to the treatment differences. The difficult multivariate regression estimation problem with its associated limitations in functional form modeling has been bypassed. **Rather than spreading the sample observations over the estimation of a multivariate dependency, all of the sample data can be concentrated on the univariate response to the varying dosage treatments**. The statistical properties (explained variation, t-value on treatment, F-stat) of the univariate response almost always dominate the corresponding statistical measures for the multivariate estimation.

The key point of this rather long discussion of a prototypical drug treatment response study is the significant efficiency/power benefits associated with the use of matched controls. The concern in a treatment response assessment is response to the treatment, a conditional univariate dependency. Variation in the controls is a source of distortion in assessing the treatment response. Trying to estimate the control impact involves the unnecessary use of sample data to eliminate distortion from control variable variation. Producing a match on each control avoids this very difficult and generally unnecessary estimation problem and focuses all sample information on the conditional univariate dependency of concern.

In viewing a forecast as a treatment applied to the stock sample in which we want to observe the performance response, the concern is again the estimation of a conditional univariate dependency. The impact of cross-sectional variation in firm-specific values of risk variables, tax effect variables, and possibly other firm impacts

such as differences in growth and profitability can be eliminated from the cross section by transforming the initial rank ordering into an associated cross section that is matched on key return–risk impact variables.

Control matching is especially pertinent to backtest the return response to a forecast using time series of rank-ordered portfolios because grouping stocks into fractile portfolios can magnify sample-level correlation distortion. The magnification is nonlinear so that very low stock-level correlations can be multiplied dramatically when the overall sample is collapsed to a very small number of portfolios. The control-matched cross section has no cross-sectional variation in the control variables and thus zero correlation with the rank-ordering variable.

8.4.3 Transforming a Rank-Ordered Cross Section into a Control-Matched Cross Section

Ranking on forecasted return and grouping into fractile portfolios will produce a set of portfolios ordered on the basis of predicted return. This return cross section will almost certainly have a wide range of forecasted return values. However, each portfolio in the cross section will almost never have the same average values of explanatory variables such as beta or size or the dividend–gain mix or any of the other return impact variables listed in Exhibit 8.1.[9] To the extent values of return impact variables fluctuate randomly about their average value over the cross section, their variation is primarily a source of noise and therefore a source of lost efficiency in assessing the cross-sectional dependency of realized returns on the return forecast score.

Much worse than lost efficiency from random variation in systematic risk, tax, and other return impact variables is the problem of distortion from correlation or partial correlation between these return dependency variables and the return

---

[9]Candidates for control variables are any variable believed to have a significant impact for explaining or predicting the cross-section of realized returns. Classes of return impacts include risk measures, e.g., beta, the book-price ratio, and firm size; tax valuation impacts, e.g., dividend yield, the dividend payout ratio, and possibly the debt tax shield as measured by the percentage of financing that is debt; and attractiveness measures that are indicative of future cashflow generation potential and asset usage efficiency, e.g., growth, return on investment, or sales intensity (sales per dollar of investment). Beta is the standard measure of volatility risk established as a return explanatory variable in formulations of the capital asset pricing model, for instance Sharpe (1964). It is also included in multifactor return modeling, as indicated by the Fama-French series, e.g., Fama and French (1992, 1996 2008a, 2008b). The variables EP and BP are the reciprocals of the price-earnings ratio and the price-book ratio, respectively. Their use as valuation and/or risk variables has been researched extensively beginning with Basu (1977), viewing dependency on the earnings-price ratio primarily as a valuation anomaly but recognizing the possibility that the earnings-price ratio could also be a risk instrument. The tax effect associated with the differential taxation of dividends and capital gains and the debt tax shield are discussed extensively in Sect. 8.7.3.

VARIABLE NAME	SYMBOL	VARIABLE DEFINITION
Beta	β	β = Cov(R$_s$-R$_o$, R$_M$ − R$_o$) / Var(R$_M$ − R$_o$) measured over 3 years of past monthly returns, where R$_o$ is the riskless rate and R$_M$ is return on a market index.
Book-to-Price Ratio (Book-to-Market Ratio)	BP	Ratio of book value to market value. Book value **BP** is the latest (at least two-month back) annual CRSP value for **total common equity.** Market value **MV** is the current market value of common stock (current stock price times number shares outstanding).
Size (Market Value of Common Stock)	S	The market value of common stock at a point in time
Earnings-Price Ratio (Earnings Yield)	EP	The ratio of **Net Income** to **market value**, the reciprocal of the price-earnings ratio
Dividend-Price Ratio (Dividend Yield)	DP	The ratio of **Annual Dividends** to **Market Value**
Financial Structure	FL	The fraction of **Total Investment** provided by debt and preferred stock
Sales Growth	SAG	Five-year average sales growth
Sustainable Growth	SUG	The growth of common equity from retained earnings
Return on Investment	ROI	The ratio of **Operating Income** (before extraordinary income and expenses) to **Total Investment**
Return on Equity	ROE	The ratio of **Net Income** to **Book Value**
Sales Intensity	SI	The ratio of Sales to **Total Investment**

Exhibit 8.1 Summary of control variables

forecast. To the extent that a systematic risk or tax variable is correlated with the return forecast score, the cross-sectional dependence of realized risky return will reflect the well-ordered cross-sectional change in the correlated variable. For instance, if a return forecast were positively correlated with beta, the observed cross section of realized risky returns will include the systematic variation in beta. An apparent increase in realized risky return from the return forecast will also include any return to beta risk bearing. Similarly, if the dividend–gain mix increases systematically with the return forecast, once again, an apparent increase in realized risky return with the return forecast can be a tax tilt in disguise.

The conventional methodology for separating return forecast potential from dependence on other variables is multivariate regression. For instance, if the concern were just to correct for beta risk in the context of the CAPM, a regression that adds the return forecast to the linear market index model as an additional explanatory

variable would resolve the relative importance of beta and the return forecast in explaining the cross section of realized risky returns. For a multivariate risk correction, adding the return forecast to the three-variable Fama–French return model might remove the effect of the three risk variables on the cross section of realized risky returns but clearly incur a multicollinearity problem if the concern is correlation distortion from the Fama–French risk variables. Trying to correct as well for systematic tax effects is even more problematic. For instance, adding a term based on dividend yield to reflect variation in the dividend–gain mix adds another explanatory variable that is correlated with all the Fama–French variables, especially the book–price ratio.[10] Adding more variables to model better the cross section of realized risky returns, for instance, combining the return forecast with all three Fama–French risk variables plus the three tax variables (DP, EP, and FL in Exhibit 8.1), and then additional variables to reflect differences in growth and profitability means using up degrees of freedom while incurring more measurement error and creating an ever worse multicollinearity problem.

As in the previously discussed example of the drug treatment response assessment, isolating well-realized return response to a forecast does not require empirically measuring an overall return dependency but rather ensuring a well-measured return-to-forecast response. In the case of a stock return forecast, the primary isolation requirement is to eliminate distortion from correlated variables in the context of a good assessment design that appropriately trades off efficiency and power.

As in the drug treatment response assessment, the contention here is that the use of matched controls is superior to using a multicollinearity-contaminated multivariate regression when the goal is isolating the impact of a return forecast from other return impact variables. The isolation alternative is to transform the initially rank-ordered cross section into an associated cross section matched on the variables that would be used as explanatory variables in a regression. A control-matched variable has the same impact on each portfolio in the cross section. There is no cross-sectional variation from the matched control and therefore zero correlation distortion.

The drug treatment example was a partially controlled experiment. Each dosage subsamples was selected to match every other treatment subsample on key controls. Rematching to reflect withdrawals and excluded subjects may be accomplished by pruning and more generally by an optimal reweighting of the subjects in each subsample. In the drug dosage study, each subsample represents a well-ordered dosage change. Dosage observations in one subsample would not be mixed with dosage observations from another subsample. In contrast, each fractile portfolio has a distribution of return forecasts summarized by their average. Reassigning stocks is

---

[10]Trying to put before-tax returns on an after-tax basis is fraught with problems. To put the dividend component of return on an after-tax basis requires an estimate of a time-varying marginal tax rate for ordinary income. To put the gain component of return on an after-tax basis requires the determination of the time-varying effective tax rate on capital gains.

an alternative way to obtain a match on key explanatory variables. Thus, rather than either excluding some stocks having extreme values of some control variables[11] or the more general use of portfolio-by portfolio reweighting to produce a transformed cross section matched on key control variables, the approach developed here is a reweighting that allows stocks to be reassigned to adjacent portfolios.

8.5   A Power Optimizing Mathematical Assignment Program

As an example of cross-portfolio reassignment, assume trying to make each portfolio in the cross section have the same average beta value. Cross-portfolio reassignment could move a stock with an above-average beta value into a portfolio whose average beta value is below the population average. At the same time, a stock with a below-average beta value in a below-average beta portfolio could be shifted into an above-average portfolio.

Just to produce a match for each portfolio in the cross section on a single explanatory control variable such as beta clearly is computationally complex for a large stock sample with many fractile portfolios. There is a need for an objective algorithmic procedure to produce the best control-matched transformation. The problem of transforming an initial rank-ordered cross section into the best control-matched cross section can be formulated as a mathematical assignment program. Given an initial rank ordering, the criterion for "best control match" is to optimize three power measures. Covariation distortion is suppressed completely by the control matching while optimizing a trade-off between range and within-portfolio forecast variance.

8.5.1   Overview: Formulating the Mathematical Assignment Program

The assumed input to the control matching algorithm is a rank-ordered grouping into fractile portfolios. In addition to input data, an optimization requires specification of decision variables, an objective function, and constraints.

Given a cross section of fractile portfolios formed by rank ordering on predicted return, the objective of the assignment program is to transform this cross section into an associated control-matched cross section to optimize two complementary attributes of statistical power:

---

[11]In a stock return forecast designed to find misvalued stocks, extreme values of some return variables are very likely the observations of greatest performance potential.

1. Preserving a wide range of well-ordered return forecasts
2. Preserving within-portfolio homogeneity of forecasted return.

The following is a verbal statement of four generic constraints.

1. *Control match restriction*. For each fractile portfolio, make the portfolio average value of each control variable equal to the mean (average) value of that control variable in the sample population.
2. *Preserving initial portfolio size*. In reassigning securities to create the control matching, keep the number of securities in each of the fractile portfolios the same as the number of securities in that portfolio in the initial (starting) rank-ordered cross section.
3. *Full assignment*. Each security must be fully assigned.
4. *No short sales*. There can be no short sales.

The crucial constraints are the control matching restrictions. Preserving initial portfolio size and full use of each security are technical constraints that go with full use of the sample. Prohibiting short sales prevents one return observation from canceling out other return observations. Prohibiting short sales is also consistent with the idea of full use of all sample information in a long-only framework.

8.5.2 *Notation Summary*

The following summarizes notation and defines key variables.

P = number of rank-based portfolios in the cross section
$p = 1$ is the portfolio with the smallest value of the rank-ordering variable
$p = P$ is the portfolio with the largest value of the rank-ordering variable
S = total number of securities being assigned to portfolios
s = security subscript
FS_s = the forecast score for stock s, $s = 1, \ldots, S$
X_{ps} = the fraction of security s assigned to fractile portfolio p, $0 \le X_{ps} \le 1$
F_p = the number of securities in fractile p in the starting rank ordering
C = number of control variables
V_c = control variable c, $c = 1, \ldots, C$
$VTARGET_c$ = the target value for control variable c, $c = 1, \ldots, C$[12]
D_{ps} = a difference measure of the change in rank for stock s when reassigned to portfolio p

---

[12]In this study, the target average value is always the ex ante sample average value.

8.5.3   The Power Optimizing Objective Function

Preserving range and minimizing cross-portfolio mixing are two aspects of statistical power. They are complementary measures in that optimizing one tends to optimize the other. To reflect the relative importance of these two measures, let Φ be a trade-off parameter that defines a relative weighting for range and within-portfolio variance, where $0 < \Phi < 1$. The trade-off between range and within-portfolio variance can be written as

$$\text{OBJECTIVE} = \Phi\,[\text{RANGE MEASURE}] - (1 - \Phi)\,[\text{Within-Portfolio Variances}]$$
(8.1)

For each portfolio in the cross section, the within-portfolio variance is the portfolio-weighted squared deviation of return forecast score from the portfolio mean forecast return score. It is a quadratic function. Thus, minimizing within-portfolio variance, actually minimizing a sum of within-portfolio variances over the cross section, means a quadratic objective function.

In this study in which we assess month-to-month return cross sections in each of the 456 months of 1967–2004, we impose progressively more complete sets of control variables in each month. Obtaining 15 or more control-matched cross sections in 456 months means solving more than 6700 optimization runs. Solving this many quadratic programs would be a computational challenge. However, just as one can approximate well the mean–variance portfolio optimization of Markowitz (1952, 1959) by solving an associated linear programming (LP) approximation to the quadratic program,[13] one can approximate the control matching quadratic optimization by an associated LP objective function.

The LP approximation objective function is

$$\text{Maximize}: \quad \text{LP OBJECTIVE} = \Phi\,[\text{RANGE}] - (1 - \Phi)\,[\text{SHIFTING}] \qquad (8.2)$$

The linear measure SHIFTING is the approximation to variance minimization that we now define.[14] Let D_{ps} be the squared difference in the numerical rank between portfolio p and the natural portfolio rank of security s in the initial rank-order partitioning into fractile portfolios. The set of D_{ps} can be summarized by a symmetric PxS matrix. Squaring the difference means that all values are greater than zero. Squaring the difference also means that large shifts are much worse than small ones. If a stock stays in the initial portfolio, D_{pp} is zero for no shifting. If all or

---

[13]See, for instance, Sharpe (1963, 1967, 1971) and Stone (1973).

[14]It is intuitive that minimizing the amount and distance of cross-portfolio shifting tends to preserve the original within-portfolio forecast distribution including within-portfolio variances. The substance of this approximation is to use portfolio rank-order distance as a substitute for actual return forecast differences. Since we map each return forecast into a near uniform distribution on the (0, 1) interval, we tend to ensure the validity of this approximation.

part of a stock is shifted up or down by one, two, and three portfolios, the respective values of D_{ps} are 1, 4, and 8. Thus, reassignments of two or more portfolios up or down the rank ordering are highly penalized.[15]

If FS_s denotes the value of the forecast score for stock s, then the linear approximation objective function above can be written in terms of assignment variables as

$$\text{Maximize} \qquad \Phi\left[\Sigma_s X_{Ps} FS_s - \Sigma_s X_{1s} FS_s\right] - (1-\Phi)\left[\Sigma_p \Sigma_s X_{ps} D_{ps}\right] \qquad (8.3)$$

The mathematical assignment program can be solved for a range of trade-off values by varying Φ from zero to 1. In the results reported in Sect. 8.8, the value of the trade-off parameter Φ is 0.25. However, experience shows that the solutions are robust to variation in Φ. The reason for the robustness is that these two attributes of statistical power are complementary objectives. Minimizing cross-fractile shifting generally preserves most of the range as well as the distribution of return forecast scores in the starting fractile portfolios.

8.5.4 Control Matching: The Equal Value Constraint for Each Control Variable

Let V_s denote the security s value of a representative control variable. Let VTAR-GET denote the target value of this representative control variable for all P portfolios in the cross section. The representative control constraint can be expressed as

$$\sum_s X_{ps} V_s = \text{VTARGET} \qquad p = 1, \ldots, P \quad \text{and every control variable} \qquad (8.4)$$

8.5.5 Security Usage and Short Sales: Technical Constraints

We impose two generic data usage constraints. The first says that each security must be fully assigned to one or more portfolios, i.e.,

$$\sum_p X_{ps} = 1 \qquad s = 1, \ldots, S \qquad (8.5)$$

---

[15]The changed difference in changed rank is actually a stronger restriction on changing portfolio membership than the quadratic variance change it is approximating. Because the LP shifting measure penalizes very large rank shifts even more than the quadratic, the LP approximation tends to preclude large shifts in rank order even more than the quadratic. However, comparison of the LP and quadratic solutions showed that the LP and quadratic solutions were generally close.

The second security assignment constraint keeps the number of securities in each matched portfolio the same as the number of securities in the corresponding fractile of the starting rank-order partitioning of the distribution of $V1$. Let F_p denote the number of securities in fractile p. Then this restriction is

$$\sum_s X_{ps} = \mathbf{F_p} \qquad p = 1, \ldots, P \tag{8.6}$$

The no short-sale restriction and the natural limitation that no security can be used more than once require

$$0 \leq X_{ps} \leq 1 \qquad s = 1, \ldots, S \quad \text{and} \quad p = 1, \ldots, P \tag{8.7}$$

8.5.6   Synthesis of the Power Optimizing Reassignment Program

Optimization arises in finding the particular reassignment that optimizes a trade-off between preserving the widest possible range of well-ordered portfolio values of forecasted return and also ensuring preservation of within-portfolio homogeneity of forecasted return.

Given the sample of stocks with variable values for each stock in that time period, once we pick a number of portfolios P in the cross section and select a set of control variables, the transformation of the rank-ordered cross section into the control-matched cross section is defined by the optimization program. The mapping from the rank-ordered input cross section into the control-matched output cross section is objective in the sense that the forecaster/researcher exercises no discretion in how stocks are reassigned. The input cross section and the mathematical program determine the output cross section.

The substance of the reassignment process is well understood by knowing input and output. The input is a cross section of fractile portfolios. The rank-ordering variable is the return forecast. The overall output is a cross section of fractile portfolios that are matched on a specified set of controls variables. The mathematical program finds an optimal reassignment of stocks that transforms the input rank-ordered cross section into a new cross section that is matched on the portfolio average values of each control variable.

The input values of the assignment variables are the relative weighting of each stock in each portfolio in the cross section without any controls. The output values of the assignment variables are the relative weighting of each stock in each portfolio in the control-matched cross section.

The relative amount of each stock in each portfolio can be used to compute portfolio average values of pertinent portfolio attributes. Of most concern is the realized risky return for each portfolio. Given a time series of rank-ordered input portfolios and a corresponding time series of control-matched output portfolios, it

is straightforward to compute the average realized risky return for each portfolio before and after controls and then to assess differences associated with the control variables.

8.6 Forecast Model Overview

8.6.1 Selecting an Illustrative Forecast Model

To illustrate well the benefits of using controls to isolate forecast performance from risk, tax, and other nonforecast impacts, a good illustrative forecast model should have a statistically significant dependency of realized risky returns on the return forecast. From the point of view of an illustration, it does not matter if the apparent dependency of realized returns on the stock return forecast is true alpha performance or is from risk, tax distortion, or other nonmodel return impact variables. In fact, when it comes to illustrating forecast isolation methodology, it is actually good if the cross-sectional return dependency is a mixture of effects from the return forecast itself and from systematic risk variables, tax effects, and other nonmodel return performance variables. In effect, to illustrate isolation methodology, it is actually good to have a "dirty return dependency" in the sense that the return dependency includes apparent performance from variables other than the forecast model itself.

The model selected to illustrate the benefits of the control methodology is an eight-variable, fundamental value-focused, rolling horizon return forecast model first published in Bloch, Guerard, Markowitz, Todd, and Xu (1993). We hereafter refer to this return forecast model as the BGMTX return forecast model. In talking about the generic approach of using a regression-estimated weighting of their eight value ratios, we shall refer to the BGMTX forecast approach or BGMTX forecast framework.

In addition to a very rigorous implementation in terms of only using data publicly available well ahead of forming the forecast, BGMTX assessed performance potential by using the model return forecast as the return input for a mean–variance portfolio optimizer. The other inputs to the mean–variance optimization were rolling horizon forecasts of security-level risk parameters.[16] The mean–variance optimizer transformed the rolling horizon return and risk forecasts into a time series of predicted mean–variance efficient portfolios in both Japan (first section, nonfinancial Tokyo Stock Exchange common stocks, January 1975 to December 1990) and the United States (the 1000 largest market-capitalized common stocks, November 1975 to December 1990). BGMTX reports that the mean–variance optimized portfolios

---

[16]For details on the mean–variance optimization used, see Markowitz (1959, 1987).

significantly outperformed benchmark indices even after testing for both survivor and backtest bias.[17]

8.6.2 Overview of the Illustrative Eight-Variable Forecast Model[18]

The BGMTX return forecast model uses a weighted average of eight value ratios[19]:

EP = [earnings per share] / [price per share] = earnings-price ratio
BP = [book value per share] / [price per share] = book-price ratio
CP = [cash flow per share] / [price per share] = cash-price ratio
SP = [net sales per share] / [price per share] = sales-price ratio.
REP = relative earnings-price ratio = EP/ [most recent five-year average value]
RBP = relative earnings-price ratio = BP/ [most recent five-year average value]
RCP = relative earnings-price ratio = CP/ [most recent five-year average value]
RSP = relative earnings-price ratio = SP/ [most recent five-year average value]

The first four ratios are called *current value ratios* in a sense of being the most recently reported values relative to the current price per share. Current value ratios measure value in terms of attractiveness compared to other peer companies. For instance, all other things being equal, a relatively high EP or BP ratio for a stock means that the stock is relatively more value attractive than the peer stocks with lower values for their EP and/or BP ratios.

The last four ratios defined above are *relative value ratios*. The "most recent five-year average value" in the denominator of these four relative value ratios means the five-year average of the ratio in the numerator. The four relative value ratios each

---

[17]Markowitz and Xu (1994) later published the data mining test for backtest bias. Their test allows assessment of the expected difference between the best test model and an average of simulated policies.

[18]BGMTX is a one-step direct forecast of stock returns. The more common return forecast framework is a two-step return forecast in which an analyst predicts both a future value of a variable such as earnings and an associated future value multiple for that variable such as a future price–earnings ratio. These two predictions imply a prediction of future value. Under the assumption that the current price will converge toward this predicted future value, there is an implied prediction of a gain return. Given a prediction of future dividends, there is an implied stock return forecast. For a thorough treatment of the two-step framework and extensive references to the two-step return prediction literature, readers are referred to the CFA study guide by Stowe et al. (2007). Because BGMTX is a direct one-step return prediction following a step-by-step determination of a normalized weighting of current and relative value ratios, it is amenable to a repeatable backtest.

[19]The ratio BP is of course the book-to-market ratio. BP is defined here as the ratio of book value per share to price per share. However, multiplying both numerator and denominator by the number of outstanding shares gives the ratio of book value to market value.

indicates relative attractiveness compared to a company's own past values of the four value ratios in the numerator. Thus, a stock is viewed as attractive not only when it provides a relatively higher earnings' yield than peer companies but also when it provides a high earnings' yield relative to its own past values. If a stock has a high relative EP ratio relative to the stock of peer companies, then that stock has had a greater relative decline in its price–earnings ratio and is thus a relatively "out-of-favor" stock.

These two types of value ratios arise from two complementary ways that fundamental value managers say they use value ratios, namely, (1) attractiveness relative to peer companies and (2) attractiveness relative to a company's own past valuations. In this sense, the relative weighting of these eight value variables can be thought of as a regression-based simulation of the type of fundamental value analysis advocated in works such as Graham and Dodd (1934) and Williams (1938).[20]

8.6.3 Variable Weighting: A Step-By-Step Implementation Summary

Having identified eight ratio variables as potential return predictors, the forecast modeling question is how to use these variables to predict future returns. An obvious way to evaluate relative predictive value is to assess how well they explain recent past returns. BGMTX uses regression to estimate the relative ability of these eight variables to explain past returns. Let R_s denote the return on stock s in a sample of S stocks. A linear regression equation to assess the relative explanatory power of the eight ratio variables is

$$R_s = a_0 + a_1 EP_s + a_2 BP_s + a_3 CP_s + a_4 SP_s + a_5 REP_s$$
$$+ a_6 RBP_s + a_7 RCP_s + a_8 RSP_s + \varepsilon_s \tag{8.8}$$

In the context of a rolling quarterly backtest of the potential benefit of using this type of ratio-based stock return forecast to improve portfolio performance using a mean—variance optimizer, BGMTX creates a time series of rolling one-quarter-ahead return forecasts from the estimated regression coefficients from Eq. (8.8).

---

[20]*Security analysis*, Graham and Dodd (1934), is generally credited with establishing the idea of *value investing*. Graham and Dodd influenced Williams (1938), who made particular reference to their low P/E and net current approaches in *The Theory of Investment Value*. In turn, Williams (1938) influenced Markowitz's thoughts on return and risk as noted in Markowitz (1991). Over the past 25 years, value-focused fundamental analysts and portfolio managers have expanded their value measures from primarily price relative to earnings and price relative to book value to include also price relative to cash flow and even price relative to sales. The choice of the eight fundamental variables in BGMTX reflects this expansion in focus, especially the expansion to include cash and sales ratios.

They use four quarters of past coefficients estimates (four sets of quarterly estimates) as the basis for a relative weighting of the eight value ratios. For each quarter-ahead return forecast, BGMTX develops a relative weighting by first modifying the coefficient estimates as described below to reflect significance and extreme values, then averaging the modified coefficients from the past four quarters, and finally normalizing the averaged coefficient values.

Stone and Guerard (2010) replicate the BGMTX forecast procedure: (1) to test performance after the publication of the model in 1993, (2) to expand the time period and sample size for the model performance potential evaluated, and (3) to resolve questions of whether the apparent performance is at least in part a return for risk or possibly a de facto yield tilt or possibly even from other return impact variables. The question of a de facto risk tilt is especially pertinent because the Fama–French return model includes BP as one of its three risk variables and BP is also one of the BGMTX return forecast variables.

The illustration of the matched control methodology to isolate well forecast performance from risk and other distortions is based on Stone and Guerard (2010). As a post publication test of the original model, the only change that Stone and Guerard (2010) made to the BGMTX return forecast procedure itself is to forecast monthly returns in a rolling month-to-month framework rather than forecasting quarterly returns in a rolling quarter-to-quarter framework.

Detailed below is the step-by-step forecast procedure summary as adapted in Stone and Guerard (2010) for a rolling month-to-month forecast.

1. *Regression coefficient estimation*. With a two-month delay, estimate each month for ten months back the regression coefficients $\{a_0, a_1, \ldots, a_8\}$ of Eq. (8.8) above.
2. *Coefficient modification*. Adjust/modify regression coefficients a_1 to a_8 in each month to reflect significance and/or extreme values in two ways:

   (a) Any coefficient with a *t*-value ≤ 1.96 is set equal to zero.[21]
   (b) Extreme positive values are truncated.

3. *Normalized average*. Average the last ten months adjusted coefficient values and normalize these averages to determine relative weights that sum to one. Let w_i denote the normalized forecast coefficient for the *i*th value variable, $i = 1, \ldots, 8$. The $\{w_i\}$ sum to one.
4. *Update ratio variables*. For each stock in the sample, update the eight value ratios using the current stock price and financial statement variables as reported in Compustat from the "most recent" (at least 2-month back) annual financial statement and current stock prices.
5. *Compute forecasted return*. Use the normalized weights from step 3 and the updated ratios from step 4 to obtain a month-ahead return forecast. If FR$_s$ denotes

---

[21]When regression coefficients with *t*-values ≤ 1.96 are made equal to zero, there are no negative coefficients regardless of significance.

the forecasted return for stock s, then the formula for the forecasted return for stock s is the weighted average of the eight value ratios, i.e.,

$$FR_s = w_1 EP_s + w_2 BP_s + w_3 CP_s + w_4 SP_s + w_5 REP_s + w_6 RBP_s$$
$$+ w_7 RCP_s + w_8 RSP_s \tag{8.9}$$

The forecast formula in Eq. (8.9) is similar to the cross-sectional return regression except that:

1. The regression error term is dropped.
2. There is no intercept coefficient.
3. The regression coefficients in Eq. (8.8) are replaced by the 10-month average of significance-adjusted, outlier-modified, and normalized past coefficient estimates for each variable.

8.7 Control Variables

8.7.1 *Control Constraints*

To assess the performance potential return forecast, it is essential to eliminate any impact from systematic risk, tax effects, or other nonmodel variables such as growth and profitability that could conceivably be the source of apparent performance value. Exhibit 8.1 lists a set of risk, tax, growth, and profitability variables that are candidate control variables.

8.7.2 *Risk Controls: β, BP, and Size*

The first three variables listed in Exhibit 8.1 are the three Fama–French risk variables: beta, book–price, and size. The ex ante beta value used in this study was based on a rolling update using three past years of monthly risky returns (return in excess of the monthly T-bill rate in that month) relative to the risky return on the CRSP index.

The *book–price ratio BP* is the book value per share divided by price per share. The ex ante book–price value is computed using the book value from the most recent financial statement lagged at least two months to allow for the financial statement data to be public information. The price per share is the last closing price in the prior month. Using BP as a risk variable is consistent with the Fama–French risk modeling but conceptually different from the Graham–Dodd use of BP as one of the value ratios that can indicate relative misvaluation of otherwise comparable companies. Given that BP is one of the eight ratio variables in the BGMTX forecast

model, the critical performance question is whether the contribution of BP is a risk effect in disguise or whether it is an indicator of value potential beyond any systematic risk. Rather than the either–or extremes of being either all risk ala Fama–French or all performance value ala fundamental value-focused analysts, the reality is almost certainly a combination of risk and value potential with the critical performance question being the relative amount of risk and value beyond risk at a given point in time. The relative amount of each effect in a cross section of performance-ranked return predictions is almost certain to vary across time. For a researcher trying to assess true value performance potential, resolving these relative contributions is a difficult problem. As we discussed further in Sect. 8.8 illustrating the imposition of risk isolating control variables, it is a very difficult problem to resolve via the conventional multivariate regression assessment but more treatable by the matched control methodology.

The size variable S is simply the market value of outstanding equity, the price per share times the number of outstanding shares. The ex ante value used in this study is based on the price per share at the close of trading in the prior month. While the measurement of size has the least measurement error of the three risk variables, the cross-sectional size distribution is perverse in the sense of having a large number of relatively small cap companies and a small number of very large cap companies. To produce a less extreme size distribution that mitigates the extremely large weight given to the small number of very big companies, Fama–French and other researchers have used *the log of company size* as the size measure in assessing the ability of size to explain the cross section of stock returns. This rather arbitrary nonlinear transformation mitigates but does not cure the heteroscedasticity problem. As in a cross-sectional regression, it matters how size is measured when imposing an equal-size constraint using the power optimizing transformation detailed in Sect. 8.5. In particular, imposing a size control that makes every portfolio in the cross section in a given month have the same average size can mean reassigning a very large company to several portfolios in order to satisfy the equal average size constraint. As in a cross-sectional regression, using the log of size mitigates but does not really cure this size distortion. An alternative used in this study was the creation of a relative size variable. "Relative size" is obtained in a given month by dividing all companies by the size of the largest company. Thus, the relative size variable puts all companies on the interval (0,1). The range in cross-sectional variance is comparable to the range and cross-sectional variance for beta, financial leverage, and growth and clearly less than the range and cross-sectional variance in other control variables such as BP, EP, and DP.

8.7.3 Tax Controls: DP, EP, and FL

The ex ante dividend yield variable DP is an annualized value of the most recent quarterly dividend per share at least 2-months back divided by the share price at the end of the prior month. Because dividends change slowly, there is very

little measurement error in using the ex ante dividend as a predictor of the future dividend. As with BP, most of the uncertainty in DP arises from changes in the price per share. For this reason, the cross-sectional correlation between BP and DP is high.

The primary tax control is DP. With the dividend yield control, every portfolio in the cross section will have the same portfolio average dividend yield. Hence, the DP control means that every portfolio has the same ex ante expectation for ordinary income. With the same dividend yield, the cross section of realized returns becomes a cross section of realized capital gains. Any variation in the dividend–gain mix over the cross section is a capital gain effect. If beta were also controlled, the ex ante CAPM expectation is a flat cross section.

Given a normalized set of weights for the BGMTX return forecast, the higher predicted returns tend to correspond to stocks with higher values of the four current value ratios, BP, EP, CP, and SP. These are all correlated with dividend yield and thus the concern that the apparent return performance may actually be a dividend yield tilt in disguise. Thus, apparent before-tax performance would be significantly reduced or eliminated if returns were put on an after-tax basis.

Adding a control for the earnings yield EP to the DP control tends to improve the ability of the ex ante DP variable to be a good control for the dividend-gain mix. When each portfolio in the cross section has the same average value of both EP and DP, each portfolio has the same ex ante dividend payout ratio, i.e., the same portfolio average value of the dividends–earnings ratio. To the extent that the dividend payout ratio characterizes dividend policy, the combination of the EP and DP controls together means that each portfolio in the cross section has the same portfolio average dividend payout policy.

One further comment on the effect of the DP control alone and especially in combination with the EP and FL controls pertains to the interaction with both size and beta. Stocks having a high dividend yield and high earnings yield tend to be larger companies with lower than average beta values. Hence, imposing the DP control alone and especially the DP and EP controls together tends to move larger and lower beta stocks to lower-ranked portfolios and vice a versa, to move smaller and higher beta stocks to higher-ranked portfolios.

It is common to talk about a value/growth trade-off with the assumption being that high value tends to mean lower growth and vice a versa. **Given validity to the value/growth assumption, value controls like BP, EP, and DP are also de facto growth controls**. Exhibit 8.1 lists two growth control variables: 5-year past sales growth and sustainable growth. When used in addition to the risk and tax controls, it is reasonable to assume that these two controls are simply refining the already established growth control associated with the risk and tax controls. This point is discussed further after illustrating the use of the risk and tax controls.

The financial leverage control FL is the ex ante percentage of nonequity financing. It is measured as one minus the book equity per dollar of total investment, both values being from the most recent annual financial statement at least 2-months

back. Aside from preferred stock, FL measures the percentage of total investment provided by debt financing.

Financial leverage has been included with DP and EP as a tax control. Rather than controlling for the dividend–gain mix, FL is designed to reflect the corporate tax shield associated with debt financing and thus a corporate tax performance impact associated with more return to shareholders and less to the government. Like many controls, FL reflects more than just the tax shield of debt financing. Use of debt involves an increase in both refinancing and interest rate risk. One source of adverse changes in interest rates is a change in inflationary expectations. Hence, the stock-specific FL control has an element of both company and macro risk control in addition to reflecting any valuation effect of the corporate tax shield.

Another potentially important role for the FL control pertains to industry exposure. There is considerable variation across industries in the relative use of debt financing. Thus, controlling for financial leverage tends to be a de facto control on the variation in industry mix across portfolios. A check on industry membership over the uncontrolled cross section of forecast-ranked portfolios compared to cross sections with the FL control imposed indicates a clear but less-than-perfect tendency for the FL control to reduce well concentrations of some industries in subsegments of the uncontrolled cross section.

8.8 Using Control Variables to Isolate Performance Potential

Section 8.8 uses the control variables defined in Sect. 8.7 and the return forecast model summarized in Sect. 8.6 to provide a concrete illustration of the power/efficiency benefits of the matched control methodology for a full sample assessment of the performance potential of a stock return forecast. The purpose here is not to establish value for the BGMTX forecast model per se but rather to use an actual return forecast to illustrate the power/efficiency benefits of the matched control methodology.

This illustration emphasizes the major design decisions that impact statistical power and efficiency.

8.8.1 Alternatives to the Full Sample, Relative Rank-Ordering Framework

From the viewpoint of having a high-power statistical assessment of the performance potential of a stock return forecast, the most important decision is the selection of the assessment framework. One alternative to the full sample relative rank ordering advocated here is the use of the forecast to select a stock portfolio that is then compared to a reference index benchmark. The BGMTX forecast

evaluation was based on the ability of the return forecast along with a risk forecast to generate a frontier of mean–variance optimized portfolios that outperformed a reference index. One problem with this approach is ambiguity with respect to the relative performance contribution of the return forecast versus the risk forecast although good sensitivity analysis can reduce this ambiguity.[22] The significant power disadvantage is limiting the forecast evaluation to the selected stocks, i.e., the stock subsample that has the best expected risk-adjusted return. Not assessing performance potential for the full sample means a loss of information and thus a loss of statistical power.

For the management of active mutual funds and for hedge funds, typical use of return and risk forecasts is not to generate a mean–variance efficient frontier but rather to use a stock return forecast as input to an index–tilt portfolio selection model that seeks to maximize the increase in expected return relative to a benchmark index subject to constraints on tracking error and a maximum tilt away from tracking error-related style variables such as beta, size, value/growth, industry, and country.[23] For organizations operating in an index tilt environment, the standard backtest performance potential assessment is to generate a time series of forecasts over a pertinent past time period and then evaluate the average performance improvement relative to the benchmark or possibly relative to another forecast or even just the past performance of the fund. The comparative assessment of alternative forecast selection approaches is often termed a "performance derby."

While assessing return forecast benefits in the context of the portfolio tilt environment in which the forecast is to be used is clearly an essential step in evaluating the performance potential of a return forecast, such a backtest assessment of constrained portfolio selection is a logical follow-on after first establishing how well the forecast performs in a large sample backtest, at a minimum how well the forecast performs in terms of ability to identify misvalued stocks across at least all the stocks in the benchmark index plus any stocks that are candidates for replacement of benchmark stocks.

There are two problems with skipping a full-sample, relative rank-ordering performance assessment and only assessing benchmark tilt performance. As in the case of evaluating a return forecast via mean–variance portfolio selection, the tilt to a relatively small subset of the stocks in a benchmark index means loss of potential sample information and thus loss of power. The typical benchmark tilt is almost always a small departure from the benchmark, for instance, a 20 % tilt is generally viewed as relatively large with significant tracking risk. Thus, for an S&P 500

---

[22]Michaud (1989, 1998) recognizes that uncertainty about both the return and risk forecasts and other portfolio selection parameters is a source of risk/uncertainty in addition to the inherent uncertainty risk of investment and has formalized a very sophisticated resampling simulation to structure very thorough sensitivity analysis.

[23]See Grinold and Kahn (2000) for a thorough description of the MCI-Barra active tilt frameworks. See Menchero, Korozov, and Shepard (2010) for an updated version of the MCI-Barra equity risk modeling that includes both industry and country factors, a global equity risk factor, and additional style factors for value, size, momentum, etc.

benchmark, the performance of the portfolio is typically more than 80 % benchmark and at most a 20 % tilt.[24] The effective comparison sample is about 100 stocks either predicted to have the best expected return relative to tracking error for overweighted stocks or the worst expected return relative to tracking error for underweighted or excluded stocks. **By focusing on a small subset of the pertinent stocks that is primarily the subset that is predicted to be the extreme best and worst stocks, a benchmark tilt comparison excludes information on the ability to rank order the rest of the pertinent stock universe**. Compared to a full sample relative rank-ordering assessment, just using a benchmark tilt assessment is a low power relatively uninformative performance potential assessment.

The assertion of being uninformative pertains especially to the second problem with using a benchmark tilt assessment to evaluate a stock return forecast, namely, mixing any forecast performance value with the effect of predictions of tracking error and of style alignment/misalignment. Mixing the effect of a return forecast with tracking error predictions and style and industry alignment/misalignment obfuscates information about the forecast itself. While extensive statistical and sensitivity analysis can help separate forecast performance from other factors,[25] the clear best solution to having a high-power assessment of forecast performance potential is to isolate completely return performance from all other return impact variables. The primary function of the matched control embellishment of the relative rank ordering is to ensure a well-isolated return forecast. Use of the power optimizing reassignment programs like that formulated in Sect. 8.5 ensures that power is optimized.

Within the full sample, relative rank-ordering framework, there are alternatives to the power optimizing matched control methodology advocated here. The most common is to use multivariate regression to explain realized returns by a combination of the return forecast and other known return impact variables such as the control candidates developed in Sect. 8.7. Another alternative is to use an endogenous APT to remove all statistically identifiable systematic variation from the return as illustrated in Guerard, Gültekin, and Stone (1997). The merits of matched controls compared to these two alternatives are addressed later in the context of illustrating and evaluating the matched control methodology.

---

[24]The magnitude of a tilt is defined as the absolute value of the difference in the relative weighting of stocks in the benchmark and the tilt portfolio. For instance, if a stock with a weight of 0.5 % is increased to 0.7 %, the tilt change is 0.2 %. If a stock with a relative weight of 0.15 % is excluded completely, the tilt change is 0.15 % to the tilt. The overall tilt percentage is the sum of all the tilt change percentages.

[25]The resampling simulation approach set forth in Michaud (1989, 1998) is again pertinent here.

8.8.2   Stepwise Imposition of Control Constraints: Procedure Overview

In a control-based assessment of how apparent forecast performance value is distorted by interaction with control variables, the starting point is a collection of rank-ordered return forecasts with no controls imposed. Inputting the no-control cross section to the mathematical assignment program for a given set of control variables produces an output cross section in which each portfolio is now matched on the specified set of controls. Comparing before and after cross sections and noting any changes in the cross section enable a forecaster/researcher to assess to what extent apparent performance potential has been distorted by one or more of the control variables. Or, in the case of no significant change, a forecaster/researcher knows that the given set of controls is not distorting apparent performance potential.

Adding controls in a stepwise fashion enables a researcher to explore how the initial rank-ordered cross section changes by systematically removing the effect of a control variable or combination of control variables. This stepwise exploration of how the return dependency changes with changes in combinations of control variables is generally very informative. Because the primary concern here is correcting apparent performance for distortion from risk and tax effects, the stepwise assessment of control impacts focus primarily on cross sections for six sets of controls summarized below:

1. No controls: the initial rank ordering
2. Individual risk controls: beta, book-to-market, and size as individual controls
3. Three risk controls together: beta, book-to-market, and size together
4. Three tax controls together: the earnings–price ratio, the dividend–price ratio, and financial leverage together as a combination control for the dividend–gain mix and other tax effects
5. The combination of risk and tax controls: the three risk and the three tax controls together, six control variables in all.

After in-depth assessment of the effect of risk and tax controls, the impact of growth and profitability controls is assessed. Finally, by removing the effect of the four value ratios BP, EP, CP, and SP, we assess the relative contribution of the four value ratios and the four relative value ratios to forecast performance.

8.8.3   Study Sample and Time Frame

The backtest study period is January 1967 through December 2004. Developing a return forecast for every stock in the backtest sample for January 1967 through December 2004 produces a time series of 456 monthly return forecast cross sections.

The data sample is all nonfinancial common stocks in the intersection of CRSP and Compustat with a book value in excess of $20 million that are included in CRSP for at least three years with monthly return data necessary to compute a 3-year rolling beta and in Compustat for at least five years with all necessary financial statement data. The table below summarizes by year the number of companies in the 1967–2004 backtest study sample.

Year	#Stocks	Year	#Stocks
1967	198	1986	1660
1968	324	1987	1632
1969	422	1988	1580
1970	564	1989	1621
1971	901	1990	1644
1972	966	1991	1671
1973	1058	1992	1742
1974	1108	1993	1845
1975	1037	1994	1921
1976	1329	1995	2003
1977	1495	1996	2057
1978	1651	1997	2193
1979	1701	1998	2238
1980	1703	1999	2331
1981	1757	2000	2284
1982	1734	2001	2256
1983	1698	2002	2305
1984	1714	2003	2318
1985	1676	2004	2238

Because of the sparseness of the Compustat database in the 1964–1966 5-year start-up period required for control variables such as 5-year sales growth, there are only 198 companies in January 1967 and only 324 companies in January 1968. The table shows that the forecast sample size grows rapidly. From 1971 on, there are more than 900 companies in the forecast sample growing to more than 2000 companies by 1995.

The fact that the sample size shows little growth from the 2003 stocks in January 1995 to the 2238 stocks in January 2004 indicates that the large number of new IPOs after the mid-1990s is not producing an increase in the number of sample companies. The fact that our sample does not exhibit the same growth as the cross time increase in publicly listed companies shows that the combination of requiring 5 years of past financial statement data plus the minimum book value restrictions means that we are studying primarily larger more mature companies.

8.8.4   Key Efficiency/Power Design Decision: The Number of Fractile Portfolios

In panel studies using rank-ordered cross sections, many return dependency assessments rank stocks into deciles and in some studies only quintiles. However, in each month of 1967–2004, stocks were ranked into 30 fractile portfolios. As already discussed, having more fractile portfolios pertains to the power/efficiency trade-off. Most of the efficiency benefits of measurement error and omitted variable diversification are accomplished with 20 or fewer stocks in a fractile.

It is pertinent to expand on the greater power benefits of having more fractile portfolios. First, grouping stocks into a fractile portfolio and using the portfolio-weighted average to represent the portfolio value of variables tend to average away information while averaging away measurement error, especially in the tails of the distribution. Second, cross-sectional plots like those in Exhibits 8.2, 8.3, and 8.4 of key performance indicators such as average realized return and realized standard deviation are more useful when there is a high density of data points. Third, when assessing the cross-sectional dependence of realized returns and realized standard deviation cross sections on the return forecast, both efficiency and power are increased from more observations. In particular, regression estimation and related hypothesis testing have much greater statistical power when there are at least 20 observations. Fourth, adjacent portfolios in a control-matched return cross section can be merged together and preserve the control matching without having to resolve the control matching optimization program. For instance, combining adjacent three tuples in the 30-portfolio cross section produces a cross section of matched deciles as done in Sect. 8.8.9. Thus, **it is methodologically better to error on the side of too many portfolios in a cross section rather than too few**.

8.8.5   The Impact of Individual Risk Controls

Exhibits 8.6 shows cross-sectional plots of average realized return and realized standard deviation versus portfolio rank for no controls and for just a beta control for risk. Exhibit 8.3 presents return and risk cross sections for just a size control and just a BP control.

Plot 2.1 is a cross section showing average realized return for the rank-ordered cross section with no controls. If there were no forecast information in the return prediction, the plot would be a random scatter about the overall average return. In contrast to a random scatter, Plot 2.1 shows an overall tendency for realized return to increase with an increase in predicted return.

The rate of increase is clearly nonlinear. For portfolios 1–10, the cross section is noisy and relatively flat. For portfolios 10–20, the realized return increases at a steady rate. For portfolios 20–30, the average realized return tends to increase at an accelerating rate with the largest increases being for portfolios 28, 29, and 30.
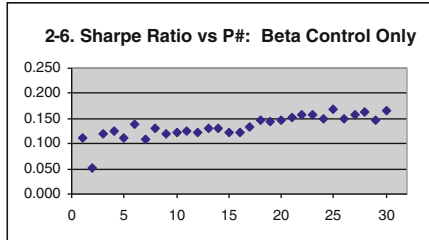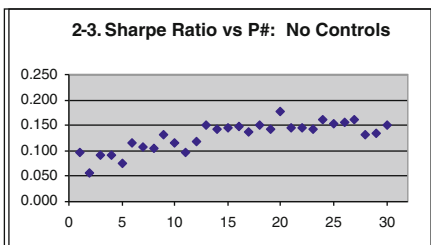
2-1. Risky Returns vs P#: No Controls

2-4. Realized Risky Return: Beta Controls

2-2. Standard Deviation vs P#: No Controls

2-5. Realized SD: Beta Controls Only

2-3. Sharpe Ratio vs P#: No Controls

2-6. Sharpe Ratio vs P#: Beta Control Only
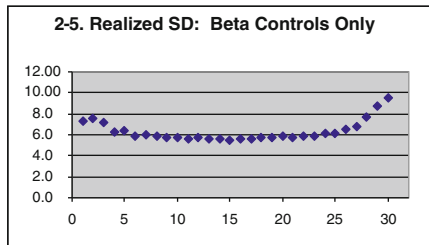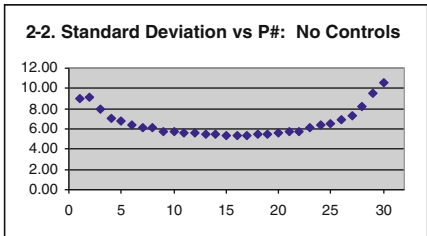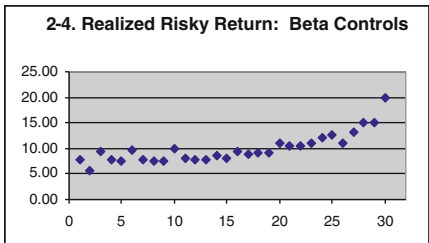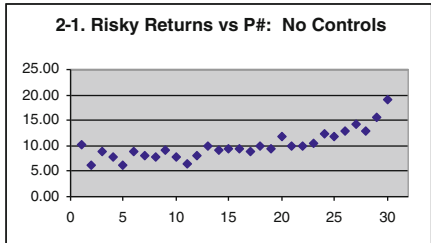
Exhibit 8.2 Risky returns, SD, and Sharpe ratio vs. P# (portfolio #): no controls compared to only a beta control

The realized return range is large and economically significant. The annualized realized return difference between portfolios 30 and 1 is 8.8 %. The difference between the realized return for the upper quintile (average for the top six portfolios) and the realized return on the bottom quintile (average return for the lowest six portfolios) is 6.5 %.

In effect, before imposing controls, Plot 2.1 indicates that the return forecast has limited ability to rank order return performance for the bottom third of the sample other than identifying the bottom third as inferior to the rest of the sample. For portfolios 10–30, the return forecast tends to rank order on average portfolio-level realized risky return. The relative rank-ordering ability is especially good for the top third of the sample, portfolios 21–30. The apparent ability to predict realized risky return improves with portfolio number and seems especially good for the top two portfolios.

The vertical axis in Plot 2.1 is average realized risky return. There is no correction for possible variation in risk. Before imposing controls for risk or

otherwise correcting for any portfolio-to-portfolio variation in risk, basic insight on risk variation is provided by measuring the cross-time variation in realized risk. Plot 2.2 shows the cross-sectional dependence of realized standard deviation on portfolio number for the case of no controls. For each portfolio number, the realized standard deviation is computed in accord with the definition by taking the square root of the mean squared deviation of each annualized monthly return from the long-run average return. The standard deviation cross section in Plot 2.2 is not a random scatter. It is also not the steady increase implied by the assumption that higher predicted returns arises from selecting progressively higher risk stocks. The very steady portfolio-to-portfolio variation pattern is not only nonlinear but also nonmonotonic. If an increase in predicted return were associated with a systematic increase in realized standard deviation risk from either systematic or unsystematic sources, the cross section of realized standard deviations should be increasing with an increase in predicted return. Any interaction between risk and predicted return is more complex than a simple linear association.

The SD cross section in Plot 2.2 is much smoother than the return cross section return in Plot 2.1. Compared to the 8.8 % return range, the SD cross section has a smaller range of just 5.19 %.[26] Most of this range is from the relatively high realized standard deviations for the very low return forecasts and the very high-return forecasts. For the three inner quintiles, portfolio 7 through portfolio 24, all realized standard deviations are within a range of just 0.70 %.

The very smooth, nonmonotonic SD cross section raises questions. One implication is that the very high and the very low return forecasts have greater realized SD risk. One question is whether the greater realized standard deviation risk arises from greater systematic risk or from unsystematic uncertainty or even possibly greater forecast uncertainty for the extreme forecasts. Another related question is whether the higher returns in the upper quintile justify the greater SD. From the viewpoint of both mean–variance efficiency and tracking error control, whether the source is systematic, unsystematic, or greater forecast error is a very important information. Imposing risk control first for beta, size, and BP individually and then in combination can help answer these questions.

8.8.6   CAPM Performance Assessments

Plot 2.4 is a cross section showing average realized return for the rank-ordered cross section with a control for just beta risk but with no other controls imposed. Imposing the same beta control means that every portfolio in the beta-controlled cross section in Plot 2.4 has the same population average value of beta. Since the

---

[26]See Appendix 8.1 for the data to compute the ranges of realized risky return and standard deviations. Appendices 8.2 to 8.6 provide pertinent data for each of the 30 fractile portfolios for the cross sections with other control variables.

rolling three-year risky return beta was calculated relative to the sample average, the ex ante sample average beta value is one in every month.[27]

The textbook *Treynor Performance Measure* is realized risky return divided by the beta of the portfolio. Thus, with all portfolios having a beta of one, **the average realized return cross section in Plot 2.4 is also a Treynor performance cross section**.

Compared to the no-control cross sections, the beta controls tend to smooth somewhat the return cross section but with very little systematic change. In contrast, the beta controls smooth the SD cross section and significantly reduce the range. The overall range is reduced from 5.19 % with no controls to 4.01 %. More significantly, the SD range for the 18 interquartile portfolios is reduced from 0.70 % for no controls to just 0.55 % with beta controls. Given the smoothing and especially the reduced range, the beta control seems to do a good job of correcting portfolios in the three inner quartiles for variation in realized standard deviation risk.

The fact that beta controls reduce but do not eliminate the greater standard deviations for both the very high and very low forecasts is evidence that some of the SD increase for these portfolios is beta related. However, for the highest predicted returns, there seems to be more to realized SD uncertainty than just beta.

The widely used *CAPM alpha* is the realized risky return less beta times the average risky return on the market index portfolio. When beta is one for every portfolio, we have

$$\alpha_p = R_p - \beta_p \text{ (risky index return)} \rightarrow \alpha_p = R_p - \text{(risky index return)} , \ p = 1, \ldots, \ P.$$

When beta is the same for every portfolio in the cross section in Plot 2.4, the CAPM alpha is just an additive constant subtracted from the realized risky returns of each portfolio in the cross section. Thus, **to within an additive constant for the average realized risky return on the market index portfolio, the average realized return cross section in Plot 2.4 with all portfolios having a beta of one is also a CAPM alpha performance cross section**.

The third standard performance measure is the Sharpe ratio. Like the Treynor ratio, the Sharpe ratio assesses risky return relative to the associated risk. Rather than risky return per unit of beta risk, the Sharpe ratio is risky return per unit of standard deviation risk. Plots 2.3 and 2.6 show the cross section of realized Sharpe ratios. As expected given the wide range of realized risky returns and relatively smaller range of realized standard deviations, the overall trend for the Sharpe ratio without any controls and especially with the beta control is a tendency to increase with an increase in predicted return. Given the smoothing from the beta control and especially given the reduced range for realized standard deviations, the beta-controlled Sharpe ratios are smoother and more nearly monotonically increasing than the uncontrolled cross section.

---

[27]Relative to the CRSP index, stocks in the backtest sample had a lower beta, generally about 10 % lower. While the backtest sample excluded generally low beta financial stocks, it also tilted toward larger, more mature companies because of the requirement of inclusion in both Compustat and CRSP for at least five years. This maturity tilt is the reason for the somewhat lower beta values for the backtest sample than for the overall CRSP sample.

Compared to the Treynor and the CAPM alpha, the relatively large realized standard deviations for the highest predicted returns mean that the beta-controlled Sharpe ratio performance assessment is less favorable for the highest predicted returns than either the Treynor or the CAPM alpha.

8.8.7   The Impact of Size and BP Risk Controls

Afterconsidering the CAPM beta-controlled performance, the next logical step is to evaluate the impact of the two remaining Fama–French risk variables—size and BP. Assessing the risk impact of BP is especially pertinent since it is one of the eight predictor variables in the BGMTX return forecast.

Plots 3.1–3.3 in Exhibit 8.3 show the cross sections of realized return, realized standard deviations, and realized Sharpe ratios when a size control is imposed.
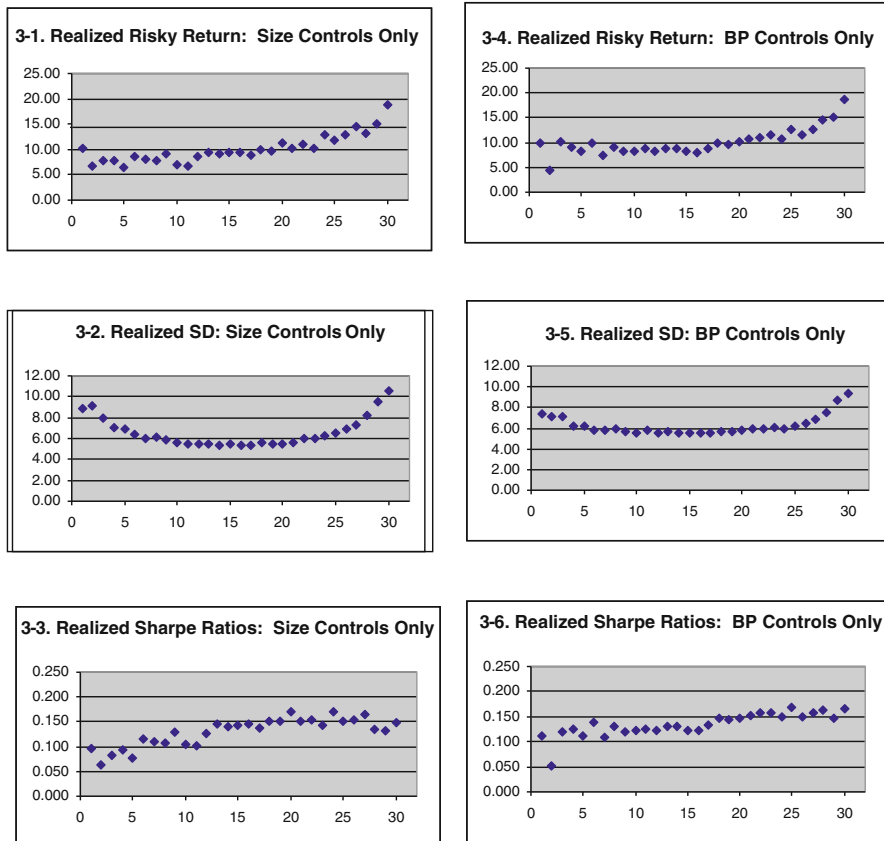
Exhibit 8.3   Risky return, SD, and Sharpe ratio vs. P# (portfolio #) for a size control only and a BP control only

Contrary to what one expects from the tendency of size to be negatively correlated with the four value ratios BP, EP, CP, and SP, the imposition of the size control alone has very little impact relative to the uncontrolled cross sections other than a slight smoothing of each cross section and a modest reduction in the realized return for the highest return forecasts.

Plots 3.4–3.6 in Exhibit 8.3 show the cross sections with just a control for BP. Despite the fact that BP is one of the eight variables in the return forecast model, eliminating any impact of BP on the cross sections has a relatively modest impact on the range of realized returns.

Imposing just the BP control tends to smooth the three cross sections, especially for portfolios ten and higher. Consistent with being a risk variable, imposing just the BP control reduces the range of realized standard deviations.

8.8.8   Imposition of Combinations of Risk and Tax Controls

Exhibits 8.4 and 8.5 repeat plots of the return cross section and SD cross section for no controls and then show the return and SD cross sections, respectively, for three key combinations of controls:

1. The Fama–French risk controls: beta, size, and BP
2. Three tax controls: DP, EP, and FL
3. The combination of the three Fama–French risk controls and the three tax controls
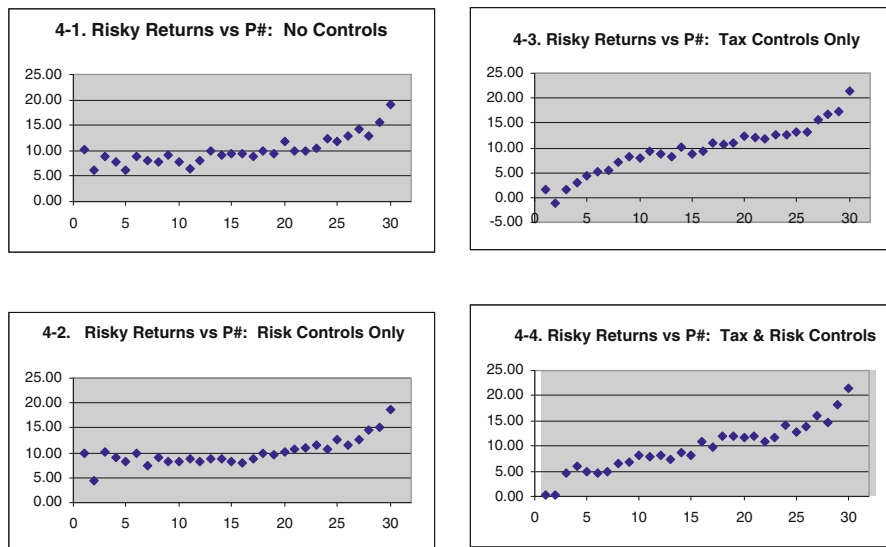
Exhibit 8.4   Risky returns vs. portfolio number: risk only, tax only, and both risk and tax controls together

   Plot 4.2 of Exhibit 8.4 summarizes the realized return cross section when all
three Fama–French risk controls, beta, size, and BP, are imposed together. Recall
that making these three variables into control variables means that the transformed
cross section has shifted securities so that each portfolio in the cross section has the
same portfolio average value of each of these three variables. There is no portfolio-
to-portfolio variation in the value of beta, size, or the book–price ratio. Thus, in
each month, these three variables will have the same contribution to portfolio return.
Realized risky return is now well-isolated from any differential impact from any of
these three risk variables. The portfolio-to-portfolio variation in return must arise
from the forecast variables other than the now-controlled book–price ratio, possibly
taxes, or other return impact variables but not from beta, size, or BP.

   Comparing Plots 4.1 and 4.2 shows a similar range and pattern but with much
less portfolio-to-portfolio variation in average realized return. The net effect of the
three risk controls is to smooth the curve without changing the overall nonlinear
pattern or the range of realized risky returns. Moreover, the smoothing effect makes
the nonlinearity much more pronounced. For the no-control plot and especially risk
control plot, the cross section of average realized returns is flat to slightly declining
for portfolios 1–15. For portfolio 15 on, the cross section has a steady monotonic
increase with the rate of increase being the greatest for the top three portfolios.

   Plot 4.3 of Exhibit 8.4 summarizes the realized return cross section with three tax
controls: EP, DP, and FL. Making each portfolio in the cross section have the same
average value of the dividend price ratio by itself tends to ensure that the percentage
of return realized as dividends is the same as the percentage of return realized as
capital gains. When both EP and DP are the same in every portfolio, this amounts to
each portfolio having the same dividend payout ratio, which is an additional control
on dividend policy. Financial leverage has been included as a tax control to reflect
the tax deductibility of corporate interest payments. The combination of having
the same earnings price ratio and therefore almost the same average earnings for
each of the portfolios plus the same percentage of debt means roughly the same
average percentage of earnings are shielded from taxes. Financial leverage also
tends to reflect both company debt capacity and exposure to interest rate risk and
may reflect some performance and risk beyond the three control variables that we
have characterized as "risk controls."

   Comparing Plot 4.3 with Plots 4.1 and 4.2 indicates significant changes in the
cross section of realized returns compared to no controls and especially compared
to the cross section with all three risk controls together. The largest changes are for
portfolios 1–15. Realized returns are reduced on average and rank-ordered much
better for portfolios 1–15 than for the plot for no controls or also for all three risk
controls together. The overall monotonic increase is now much steadier and more
nearly linear. **Tax effects clearly exhibit very significant systematic variation
over the cross section of forecast rank-ordered portfolios compared to both
the uncontrolled and the risk-controlled cross sections!**

   Plot 4.4 with both risk and tax controls is similar to Plot 4.3. Adding risk controls
to the tax controls does not significantly change the cross-sectional return plot.
Having controls on both DP and β together for the combination of risk and tax
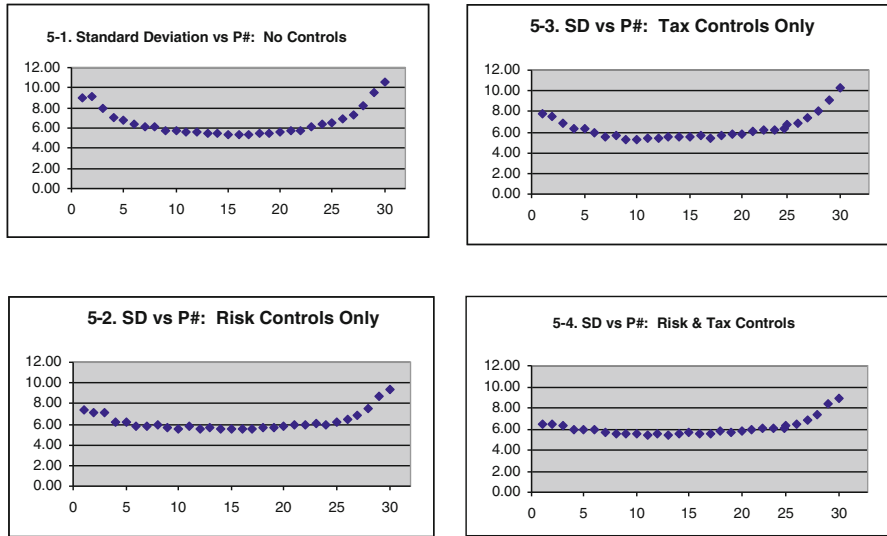
Exhibit 8.5 Standard deviation vs. portfolio number: risk only, tax only, and both risk and tax controls together

controls tends to ensure even more completely that the percentage of returns realized as dividends and as capital gains is the same over the cross section.[28]

Exhibit 8.5 contains plots of realized standard deviation versus portfolio number for no controls, for all three risk controls together, for just the three tax controls together without any risk controls, and then for all three risk and tax controls together. All four plots are nonlinear and nonmonotonic. Plot 5.1 with no controls has a range of realized standard deviation values from a low of about 5.5 % in the

---

[28]It is easy to check how well DP alone and DP and β in combination actually control for cross-sectional variation in the dividend–gain mix. For the forecast rank ordering in this study, in all time periods of 5 years or longer after 1972, the DP control alone does a good job of controlling for cross-sectional variation in the ex post dividend–gain mix. The term "good job" means that the average ex post dividend–gain ratio in each portfolio is very close to the sample average with no systematic variation over the cross section. Controlling for DP and EP together improves the control for variation in the dividend–gain mix by eliminating portfolio-to-portfolio variation and making most portfolios very close to the average. Controlling for DP and β in combination improves the control since high beta tends to be lower dividend payout. Likewise controlling for DP and size in combination improves the control for the dividend–gain mix for similar reasons, namely, the fact that small size tends to be higher beta and often zero or token dividend payout. Thus, the combination of tax controls and risk controls together improves on the tax controls alone in terms of ensuring very little portfolio-to-portfolio variation in the dividend–gain mix, especially for all time periods after 1972. The main caveat is slighter greater variation about the average dividend–gain mix for the three lowest-ranked and the three highest-ranked portfolios. This greater variation about the mean for the low-ranked and high-ranked portfolios is consistent with the much greater realized standard deviation for these portfolios as shown in Exhibit 8.5 as well as the much greater positive skewness for the three highest-ranked portfolios.

middle of the cross section to a high of 11.5 % for portfolio 30, a high-low range of 6 %. The portfolio-to-portfolio changes are remarkably smooth. After portfolio 2, realized standard deviations first smoothly decrease toward the middle of the cross section and then smoothly increase at an accelerating rate with the highest realized standard deviation occurring for portfolio 30.

As expected, adding the three risk controls in Plot 5.2 of Exhibit 8.5 tends to reduce realized risk variation as reflected in the reduced range of realized standard deviations from 6 % to less than 3.5 %. Moreover, realized standard deviation varies by no more than 1 % from portfolio 4 to portfolio 20. Most of the increase in realized standard deviation at the low and high end of the cross section is attributable to skewness, negative skewness for the low end, and significant positive skewness for the high end.

Plot 5.3 for tax controls only is similar to Plot 5.2 for just risk controls except for slightly higher realized standard deviations at the low and high extremes and a slightly greater asymmetry. The fact that both risk and tax controls have similar effects in terms of controlling for realized standard deviation risk is surprising. It suggests risk control impacts from some combination of the dividend–gain mix and possibly financial leverage risk. Plot 5.4 with both risk and tax controls together supports this conjecture of risk control benefit from the three tax controls beyond the risk control provided by beta, book–price, and size. For portfolio 1–21, the cross section of realized standard deviations varies by just a little more than 1 %. For these 21 portfolios, the combination of the conventional risk variables plus the tax controls does an excellent job of controlling for realized risk as measured by realized standard deviation.

As indicated by the skewness data in the Table in Appendix 8.6, the increase in realized standard deviation in Plot 5.4 is well explained by the corresponding increase in significantly positive realized skewness for the highest return portfolios.[29]

Exhibit 8.6 shows cross-sectional plots of the Sharpe ratio for the four control sets. As a synthesis of the respective return and standard deviation plots, they show that the overall increase in returns is outweighed by the very modest increase in realized standard deviations. Hence, the plots in Exhibit 8.6 indicate significant performance potential for the basic BGMTX return forecast framework, with the most pertinent Sharpe ratio cross section being Plot 6.4 since it eliminates distortion from both risk and tax effects.

---

[29]A plot of realized semi-standard deviation for the case of all risk and all tax controls together is flat for the top 25 portfolios in the cross section, strong support for the assertion that the greater standard deviation for the top four portfolios is primarily a positive skewness effect and not downside uncertainty.

The primary purpose of this control matching example is to illustrate the benefits of using matched controls in assessing forecast performance potential rather than to establish value to the eight-variable BGMTX forecast model.[30]

8.8.9   Stepwise Imposition of Risk and Tax Controls: High-Minus-Low Differences

Exhibit 8.7 summarizes high-minus-low returns for major constraint sets. The first column names the constraint set. In addition to the risk and tax controls used in looking at the impact of risk and taxes on the cross-sectional plots, Exhibit 8.6 lists a more detailed stepwise imposition of control matching constraints. In particular, it adds to the risk and tax controls additional controls for growth and profitability.

The next three columns in Exhibit 8.7 give high-minus-low returns. For 30 fractile portfolios in column 2, this high-minus-low value is the long-run average
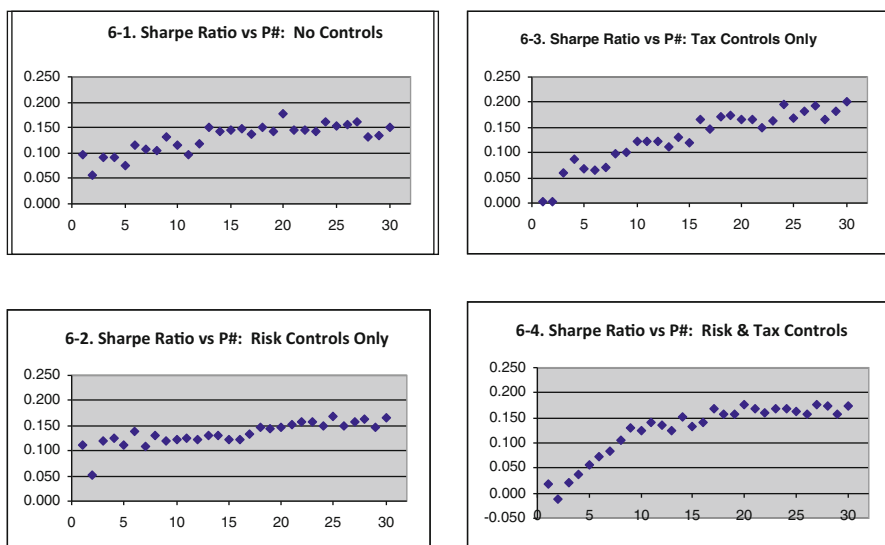
Exhibit 8.6   Sharpe ratio vs. portfolio #: risk only, tax only, and both risk and tax controls together

[30]Forecast value for mean–variance portfolio selection was established in BGMTX for 1978–1990. Guerard, Gultekin, and Stone (1997) added to the evidence of forecast value for the return forecast itself by using an endogenous APT to remove all explainable systematic return. Others have added both growth and momentum to show performance value well after the 1993 publication time.

return on portfolio 30 (the fractile portfolio with the highest forecast score) less the long-run average return on portfolio 1 (the fractile with the lowest forecast score). Column 3 is the average return for the two highest fractiles minus the average return for the two lowest fractiles. Column 4 for deciles[31] is the average of the top three fractiles minus the average of the bottom three fractiles. Since all portfolios in each cross section are matched to the ex ante values of the listed factor controls, **the high-minus-low values are the long-run realized returns on a factor-neutral arbitrage portfolio, i.e., a portfolio that is long in one or more of the top 30 fractile portfolios and short in the corresponding low score portfolios**. It is *factor neutral* in the ex ante values of each of the imposed control variables because each of the portfolios in each cross section has been matched to the sample average value of

CONTROL VARIABLES	30-tiles	Deciles	Quintiles
No Constraints	0.088	0.074	0.065
β(Beta)	0.121	0.091	0.066
S(Size)	0.086	0.075	0.064
BP(Book-to-Market)	0.076	0.068	0.055
β, S, BP	0.087	0.079	0.056
FL, EP, DP	0.196	0.176	0.137
β, S, BP, FL, EP, DP	0.211	0.165	0.128
β, S, DP, FL, Sag5, Sug3	0.239	0.187	0.147
β, S, BP, FL, EP, DP, Sag5, Sug3, ROE	0.267	0.193	0.143
β, S, BP, FL, EP, DP, Sag5, Sug3, ROI	0.250	0.190	0.143
β, S, BP, FL, EP, DP, Sag5, Sug3, SI	0.254	0.194	0.143
β, S, BP, FL, EP, DP, Sag5, Sug3, ROE, SI	0.276	0.198	0.145
β,S,BP,FL,EP,DP,SAG5,SUG3,ROE,SI,CP,SP	0.191	0.142	0.111

Notes:

1. For 30-tiles, the high-minus-low values for each control set are computed by taking the difference between the average realized return for portfolio 30 and for portfolio 1.

2. For deciles, the high-minus-low value is the difference between the average realized returns for the top 3 portfolios minus the average realized return for the bottom 3 portfolios.

3. For quintiles, the high-minus-low value is the difference between the average realized return in the top six portfolios minus the average realized return for the bottom six portfolios.

Exhibit 8.7   High-minus-low values for 1968–2004 average returns: how imposing controls changes the extreme high and low returns

---

[31]With 30 fractile portfolios in the cross sections of conditional return response observations, the difference for the top three returns combined and the bottom three portfolios combined represents a high-minus-low return for the top and bottom deciles of the cross section.

the imposed controls. Therefore, **a long–short combination of any two portfolios has zero ex ante exposure to the imposed controls**.

The high-minus-low values are the annualized average of 458 monthly values. Thus, they indicate the economic significance of the composite value score before any transaction costs for a naive factor-neutral portfolio strategy.

The term "naïve" refers to the fact that these portfolios are formed on the basis of return forecast data alone without using any information about variances, covariances, or higher moments such as skewness. Given that past values of both variance and covariance are fairly good predictions of month-ahead values, use of the value-focused return scores with mean–variance optimization should always produce superior market-neutral hedge portfolios in terms of Sharpe ratios. **The Sharpe ratios reported for these three market-neutral portfolios are lower bounds on the mean–variance optimized factor-neutral portfolios**.

The high-minus-low ranges and associated Sharpe ratios both exhibit a strong dependency on the imposed controls. The return cross section with no controls has a range of 8.8%, the Fama-French three factor control set has a range of 8.7%. In contrast to the is very small change from imposing the three risk controls, imposing the three tax controls results in a high-minus-low range of 19.6%. Imposing the three risk controls and the three tax controls together further increases the 30-fractile high-minus-low range to 21.1%.

Adding growth controls increases the hml to more than 24 %, triples the range for the unranked cross section, and more than doubles the range when the only control variables are the conventional Fama–French three-factor risk instruments. Adding sales intensity and profitability controls further increases the range and improves the Sharpe ratios.

8.8.10 Estimates of the Dependence of the Return and SD Cross Sections on the Return Forecast

Exhibits 8.8, 8.9, and 8.10 summarize regression tests of the ability of the return forecast score to explain the long-run realized cross sections of average returns, Sharpe ratios, standard deviations, and skewness coefficients.

For the stepwise imposition of a series of control constraints, the first table in Exhibit 8.8 summarizes linear regressions of the long-run 456 month average value of realized risky return on return forecast score for a series of progressively more complete sets of control variables.

Control Variables	$R_p = C_0 + C_1(FS_p) + \varepsilon_p$			
	slope	t	R^2	p-value
No Constraints	.079	7.23	.651	<.0001
β(Beta)	.086	7.47	.666	<.0001
S(Size)	.080	7.91	.691	<.0001
BP(Book-to-Market)	.071	7.77	.683	<.0001
β, S, BP	.079	6.44	.597	<.0001
FL, EP, DP	.181	17.81	.919	<.0001
β, S, BP, FL, EP, DP	.192	18.97	.928	<.0001
β, S, DP, FL, Sag5, Sug3	.206	16.4	.906	<.0001
β, S, BP, FL, EP, DP, Sag5, Sug3, ROE	.217	14.57	.884	<.0001
β, S, BP, FL, EP, DP, Sag5, Sug3, ROI	.215	16.44	.906	<.0001
β, S, BP, FL, EP, DP, Sag5, Sug3, SI	.216	15.66	.898	<.0001
β, S, BP, FL, EP, DP, Sag5, Sug3, ROE, SI	.220	14.24	.879	<.0001
β,S,BP,FL,EP,DP,SAG5,SUG3,ROE,SI,CP,SP	.186	16.56	.907	<.0001

Exhibit 8.8 The changing ability of forecast score to explain realized returns and Sharpe ratios

All of the regressions in Exhibit 8.8 have very high R-squared values, large and significant *t*-values, and *p*-values less than 0.0001. Given that all of the regressions have *p*-values less than 0.0001, the change in the *t*-values for the coefficient on return forecast score and the change in *R*-squared are the best indicators of the effect on the cross section of imposing additional control constraints, especially in terms of the extent to which we are obtaining a return forecast dependency that is better isolated from the effect of nonforecast variables. Thus, we focus most of our attention here on the changes in the *t*-values and *R*-squared values as we impose different sets of control constraints.

The cross section of realized returns with no controls and with the three risk controls imposed individually and in combination results in very modest changes in both the estimated slope and the associated *t*-value. For instance, the imposition of the three risk controls together produces a *t*-value on the slope coefficient of 0.597 with an associated *t*-value of 6.44 compared to the no-control case of a slope coefficient of 0.079 and a t-value of 7.23.

The most significant structural feature is the jump in *R*-squared values and *t*-values when we impose the tax controls alone or impose the tax controls along with the three systematic risk factors. Imposing the three tax controls alone produces a t-value of 17.81. The control set with both systematic risk and tax controls has a t-value of 18.97, a clearly significant increase in the slope estimate and its significance.

These results are a surprise! First, the negligible impact of the three Fama–French risk controls on the long-run cross section of realized returns means that apparent return potential is not a systematic risk effect in disguise, at least in terms of the three Fama–French risk variables. The major surprise is the large and very significant tax effect. **Most surprising is the direction of the tax effect on the assessment of performance potential**. In noting the issue of distortion from regularly recurring systematic tax effects, the concern was that the very high correlation of the four current value ratios with dividend yield could mean that apparent forecast potential

could be a dividend yield tilt in disguise rather than finding truly misvalued stocks that would produce superior returns as the market recognized the undervaluation.

Also surprising is the fact that imposing the combination of risk and tax controls means that all portfolios in the cross section have the same portfolio average values of BP and EP, two of the eight variables weighted in the return forecast prediction. In effect, any contribution of these two variables to realized return performance is suppressed for the combination of risk and tax controls. The apparently large and significant performance is from the other six variables. It is pertinent to note that a benefit of the control approach is a straightforward assessment of the relative contribution of one or more of the forecast variables in a multivariate forecasting model. Contrary to much empirical evidence on the value of BP and EP for explaining the cross section of realized returns, it appears that for the cross section based on the BGMTX forecast model, neither BP nor EP are an important part of the very significant performance potential indicated by the plots, the high-minus-low returns, and especially the very significant slope for the linear regression fit. This analysis indicates that most of the forecast potential must arise from the other six variables.

In order to gain more insight on forecast potential, it is useful to add to the risk and tax controls additional controls for growth and profitability and also to assess the impact on the return cross section of suppressing other model variables. The next control sets summarized in Exhibit 8.8 add to the Fama–French risk controls and the set of tax controls two growth controls (5-year sales growth, 3-year sustainable growth) plus a profitability control (either ROE, ROI, or sales intensity SI). In all three cases the estimated slope coefficient increases to more than 0.20, but both the t-value and R^2 decrease slightly. The increased slope coefficient with poorer fit makes sense if one also considers the high-minus-low data in Exhibit 8.7. The effect of adding growth and profitability controls is to increase primarily the very high returns and to decrease the very low returns and to thereby increase the departure from linearity. Hence, there is more range and more performance potential for the very high predicted returns but a departure from linearity and a poorer linear fit.

The final control set adds controls for two more forecast variables: CP and SP. Hence, for this control set there is no cross-sectional impact from any of the four current value ratios: BP, EP, CP, and SP. The large significant slope coefficient is attributable solely to the four relative value ratios. **If we use the estimated slope coefficient as an indicator of overall ability of the return forecast to predict risk-controlled, tax-controlled, growth–profitability controlled realized return predictions, the suppression of CP and SP indicates that the relative value ratios are responsible for about 80 % of the apparent return forecast potential, CP and SP contribute about 20 % of the apparent return forecast potential, and BP and EP seem to contribute very little to the apparent return forecast potential.**

8.8.11 The Cross Sections of Realized Standard Deviations for Different Combinations of Controls

Exhibit 8.9 summarizes regressions of the realized cross-time standard deviations on return forecast score for different combinations of controls.

The coefficient on the linear term C_1 is insignificant until tax controls (FL, EP, and DP) are imposed. The large jump in the t-value with the imposition of tax controls alone or in combination with other variables again indicates that controlling for tax effects is critical to isolate return forecast performance from other distorting return factors. For the cross section of realized standard deviations, it appears systematic tax effects are the most pertinent set of control variables rather than the usual systematic risk variables.

In all of the cross-sectional plots summarizing the dependence of realized standard deviation on portfolio number such as in Exhibit 8.5, the dependence of realized standard deviation on portfolio number is clearly nonlinear and nonmonotonic. In particular, the standard deviations for low portfolio numbers (low forecast scores) and for high portfolio numbers (high forecast scores) were all substantially greater than the standard deviations for the middle of the cross section. Visual inspection of the cross-sectional plots suggests a quadratic dependency. For this reason, the regressions summarized in Exhibit 8.9 designed to assess the cross-sectional dependence of realized standard deviations on return forecast score include a quadratic term as well as a linear term. Because the concern is assessing the impact of below average and above average forecast scores relative to the average, the quadratic dependency is expressed as a squared deviation of return forecast score

Control Variables	$SD_p = C_0 + C_1(FS_p) + C_2(FS_p - mean(FS_p))^2 + \varepsilon_p$					
	C_1	t	C_2	t	Adj R^2	p-value
No Constraints	0.07	2.30	0.02	18.21	0.920	<.0001
β(Beta)	0.11	3.74	0.01	12.45	0.852	<.0001
S(Size)	0.07	2.31	0.02	18.36	0.921	<.0001
BP(Book-to-Market)	0.11	3.89	0.02	22.24	0.946	<.0001
β, S, BP	0.14	5.50	0.02	15.46	0.902	<.0001
FL, EP, DP	0.18	6.46	0.02	17.32	0.928	<.0001
β, S, BP, FL, EP, DP	0.20	9.98	0.01	16.57	0.934	<.0001
β, S, DP, FL, Sag5, Sug3	0.18	7.18	0.01	13.27	0.895	<.0001
β, S, BP, FL, EP, DP, Sag5, Sug3, ROE	0.21	9.24	0.01	14.80	0.918	<.0001
β, S, BP, FL, EP, DP, Sag5, Sug3, ROI	0.20	10.20	0.01	16.20	0.931	<.0001
β, S, BP, FL, EP, DP, Sag5, Sug3, SI	0.21	9.44	0.01	15.04	0.921	<.0001
β, S, BP, FL, EP, DP, Sag5, Sug3, ROE, SI	0.21	9.62	0.01	15.11	0.922	<.0001
β,S,BP,FL,EP,DP,SAG5,SUG3,ROE,SI,CP,SP	0.17	11.06	0.01	17.13	0.936	<.0001

Exhibit 8.9 The ability of forecast score to explain the cross section of standard deviations

from the average forecast score. In the regressions in Exhibit 8.9, the t-values for the quadratic coefficient C_2 are much larger and thus much more significant than the t-values for the linear coefficient C_1. The very high adjusted R-square values for all of the regressions in Exhibit 8.9 strongly indicate that the combination of a linear and quadratic dependency explains most of the cross-sectional variation in the realized standard deviations. The much higher significance for the quadratic term is confirmation of the importance of the nonlinear, nonmonotonic apparently near-quadratic dependency suggested by the cross-sectional plots such as Exhibit 8.5.

8.8.12   The Cross Section of Realized Skewness Coefficients

Exhibit 8.10 is a cross-sectional regression on the long-run realized skewness as measured by the skewness coefficient.

The skewness is significant even with no control constraints. This jump in the *t*-value from imposing the beta control alone suggests that controlling for market movements by means of the beta control increases the isolation of nonsystematic skewness from any market skewness. Interestingly, neither the size control alone nor the book-to-market control alone significantly changes the skewness. However, the three systematic risk controls together increase the increase skewness.

In contrast to the regressions for realized standard deviation, tax controls alone do not seem to help isolate skewness effects. However, the three risk controls plus the three tax controls together do increase the coefficients and associated *t*-values.

Control Variables	$SD_p = C_0 + C_1(FS_p) + C_2(FS_p - mean(FS_p))^2 + \varepsilon_p$							
	C_0	t	C_1	t	C_2	t	Adj R^2	p-value
No Constraints	-1.272	-14.36	0.014	10.7	0.001	13.67	0.911	<.0001
β(Beta)	-1.252	-14.06	0.017	12.13	0.000	8.55	0.883	<.0001
S(Size)	-1.274	-13.42	0.014	9.95	0.001	12.84	0.900	<.0001
BP(Book-to-Market)	-1.375	-15.84	0.016	11.7	0.001	14.68	0.923	<.0001
β, S, BP	-1.340	-21.18	0.018	17.84	0.000	12.14	0.941	<.0001
FL, EP, DP	-1.164	-9.15	0.013	5.92	0.001	9.16	0.822	<.0001
β, S, BP, FL, EP, DP	-1.365	-17.42	0.019	14.07	0.000	7.93	0.907	<.0001
β, S, DP, FL, Sag5, Sug3	-1.347	-21.04	0.018	17.12	0.001	11.15	0.940	<.0001
β, S, BP, FL, EP, DP, Sag5, Sug3, ROE	-1.408	-18.3	0.020	15.07	0.000	8.2	0.916	<.0001
β, S, BP, FL, EP, DP, Sag5, Sug3, ROI	-1.411	-23.11	0.020	19.17	0.000	10	0.945	<.0001
β, S, BP, FL, EP, DP, Sag5, Sug3, SI	-1.427	-19.23	0.020	15.88	0.001	9.24	0.926	<.0001
β, S, BP, FL, EP, DP, Sag5, Sug3, ROE, SI	-1.418	-21.82	0.020	18.03	0.000	10.02	0.940	<.0001
β,S,BP,FL,EP,DP,SAG5,SUG3,ROE,SI,CP,SP	-1.150	-18.14	0.016	14.51	0.000	5.89	0.896	<.0001

Exhibit 8.10   The ability of forecast score to explain the cross section of skewness coefficients

Adding our two growth controls and other company-specific controls does increase significantly the ability of return forecast score to explain the cross-sectional skewness coefficient. We conclude that isolating nonsystematic value-related skewness associated with the illustrative value-focused return forecasting model requires that we isolate these value-related return and risk effects from growth in particular.

In the final set of controls in Exhibit 8.10, we add two more controls for two of the forecast model variables, namely, CP and SP. Given that the control set already contains the BP and EP ratios, the net effect of this final control set is to remove from the cross section the contribution to realized returns of all four current value ratios. Thus, this final set thus measures the skewness response for just the relative value ratios. It is interesting that the *t*-value is the greatest for this set of control constraints. **This high *t*-value for the response of realized skewness to the relative value subset of the eight forecast variables is strong evidence that the significant cross-sectional dependency of skewness on return forecast score is primarily from the relative value ratios rather than the current value ratios**.

Since the relative value ratios measure attractiveness for a company relative to its own past value ratios, this result suggests, or at least is consistent with, company values returning to a moving mean. In other words, this result indicates that when current value ratios are well below their 5-year average value, there is an apparently strong likelihood that the stock price will increase in order to return the company to its recent average valuation. **This result is consistent with the value analyst use of relative value ratios to find stocks with limited downside risk and upside potential**. It also suggest turning point performance, rather than trend (momentum) as the primary source of unpriced, unsystematic skewness.

8.9 Further Research

Having good return forecasts is the primary requirement for successful active portfolio management. Given one or more return forecasting alternatives, the central problem for making decisions about return forecast centers on the ability to conduct high power, high efficiency backtest assessments. The focus here has been to illustrate the use of a mathematical assignment program to optimize the construction of control-matched cross-sections of rank-ordered return forecasts. The central requirement for high-quality return forecast assessments is the ability to isolate the impact of the return forecast from the impact of other return variables. The control matching framework is an alternative to the use of multivariate regression to evaluate return forecasts.

The BGMTX forecast model has been used as an illustrative multivariate return forecast. There are several ways to extend this forecast assessment. A very important question for forecast performance is cross-time consistency. The majority of this chapter has looked at the very long 1967-2004 for time period. Stone and Gerard

(2010) report comparable performance results for this United States sample for four subperiods of 1967–2004.

Additional performance concerns are relative ability to perform in up-markets versus down-markets as well as performance in other markets or market subsectors. Stone and Gerard (2011) use control matching to evaluate performance for Japan including subperiods of net market decline. The Japan assessment shows that the BGMTX forecast model rank orders return performance over extended net market declines in Japan, for instance excellent relative rank ordering for control matched cross-sections when every fractile portfolio except the highest ranked has a negative return.

Tracking error pertains to cross-time consistency. Given a sample of control matched cross-sections for a forecast time period such as the 456 monthly control matched cross-sections used in this illustration of the methodology, a researcher can evaluate tracking error by looking at moving averages, e.g., 12-month, 24-month, and 60-month rolling averages to obtain an indication of one-year, two-year, and five-year performance consistency risk. Observing moving average performance provides insights to not only tracking consistency but also relative performance in net up markets, net down markets, and over market reversals. Relative performance in up, down, and reversal markets provides insight on the extent to which a particular forecast is momentum focused (trend extrapolating) versus turning-point focused. For the BGMTX forecast model used in this illustration of control matching, one-year moving averages indicate that much of the performance value occurred over marker reversals, both transition from a bear to bull market and also transition from a bull market to a bear market. This reversal performance is consistent with the fact that suppressing the effect of the four value ratios indicated that roughly 80% of the high-minus-low performance summarized in Exhibit 8.7 was attributable to the four relative value (return reversal) variables in the BGMTX forecast model.

The key idea implicit in the preceding discussion is many alternative performance assessment insights beyond the statistical tests used in this chapter. Having a time series of control matched forecast cross sections allows a researcher to investigate not only statistical return performance but also cross-sectional risk (tracking error, standard deviations, skewness, etc).

Looking at performance value without noise from other return variables lets a researcher experiment with forecast performance without distortion from other return impact variables. Moreover, looking at performance for alternative control sets provides important information on interaction with other return impact variables. For instance, the fact that the two growth controls had an impact on the high-minus-low performance even after controlling for both risk and tax effects suggests a nonsystematic growth effect and therefore potential value to adding one or more growth variables to the BGMTX forecast model.

8.10   Conclusions

The BGMTX forecast model shows significant return performance potential as established in other studies. Compared to studies that do not use matched controls, the use of risk, tax, growth, and profitability controls provides additional information.

1. Both the return and the realized standard deviations cross sections are nonlinear. The cross section of realized standard deviations is a relatively small range compared to the range in the cross section of realized returns. The cross section of realize standard deviations is not only nonlinear but highly nonmonotonic.
2. The distribution of realized returns about the average value exhibits skewness, negative skewness for low return forecasts, very little for the middle of the distribution, and very large significant skewness for the highest return forecasts.
3. The three risk control variables tend to smooth the cross sections of realized returns; however, risk variables appear not to have a significant effect on the long-run cross section of realized returns. These risk variables are not a source of systematic performance bias.
4. The three risk controls in combination tend to smooth the cross sections of realized standard deviations. More importantly, the three risk controls together reduce the range of realized standard deviations.
5. Tax effects are very significant for the illustrative BGMTX forecast model. Contrary to the hypothesis of apparent return potential being a tax tilt, imposing the three tax controls to eliminate cross-sectional variation in the dividend–gain mix significantly increases the slope and range of the realized return cross section and moderately reduces realized standard deviation.
6. The power optimizing imposition of a combination of risk and tax controls significantly increases statistical efficiency relative to the uncontrolled cross section. Adding growth and profitability controls adds additional value in assessing return potential.
7. Suppressing the four current value ratios in the cross section of realized returns reduces the slope of the realized return cross section only modestly, about 20 %. The relative value ratios are the major source of realized return potential and the significant positive skewness for the high-return forecast part of the cross section.

The main methodology benefit for full sample assessments of performance potential of a return forecast is the control matching framework itself. Using a power optimizing reassignment program like that formulated in Sect. 8.5 makes it possible to eliminate return performance distortion from other return impacting variables without having to make any assumptions about the distribution or form of the functional dependency for any of the control variables. One simply has to assume a possible dependency on the return variables. Imposing the control constraint eliminates any portfolio-to-portfolio impact from the matched controls. The distribution-free, specification-free attribute avoids functional form specification errors and the estimation limitations associated with distributional assumptions.

Compared to using a multivariate regression to estimate jointly the sensitivity of realized returns to the forecast and also to estimate an assumed dependency for all the other possible return impact variables for risk, taxes, growth, profitability, etc., the use of controls to suppress cross-sectional variation in the other variable dependencies means only the need to estimate a univariate dependency on the return forecast under the assumption that any effect of other return impact variables has been suppressed. Concentrating all sample data on a conditional univariate dependency rather than estimating a multivariate dependency means much more efficient, more powerful extraction of sample information relative to making restrictive assumptions to estimate multiple dependencies when the concern is a single well-isolated conditional dependency.

The use of tax controls illustrates the benefit of avoiding functional specification to remove a potential source of realized return distortion. By simply making every portfolio in the cross section have the same dividend yield, dividend payout ratio, and same benefit from the debt tax shield, it is not necessary to estimate time-varying marginal tax rates for dividends and gains. Similarly, using FL as a financial control avoided the need to assess the value of the debt tax shield and the simultaneous need to correct for distortion from other possible but not known valuation effects of financial leverage.

In addition to the benefits of concentrating data on a univariate dependency and avoiding distribution and specification assumptions, the control matching framework completely eliminates bias/distortion from covariability effects between the return forecast and any of the control variables. Complete elimination of covariability contamination is a significant power benefit!

Another efficiency/power design concern is the number of portfolios in the cross section. The efficiency benefit of grouping observations is reduced measurement error and possibly reduced specification and omitted variable error. Grouping observations loses power by reducing the number of sample observations. The point made in this paper is the need to explicitly recognize the relative benefit of reduced measurement error versus loss of power from reduction in the number of sample observations and the associated loss of information from using averages to represent a collection of observations. In the control matching framework, it is possible to consolidate control-matched fractiles and preserve the control matching. This consolidation is illustrated in Sect. 8.8.9 assessing the high-minus-low realized return differences for 30 tiles and the consolidation to 15 tiles and deciles.

Return forecasting is very much an art using knowledge of valuation and statistics. Assessing return forecast performance potential is also an art. The matched control framework aided by an optimal reassignment algorithm is a decision support framework for exploring return performance potential that avoids many limitations of multivariate regression assessments including especially collinearity distortion and the restrictions of distributional and functional form assumptions.

Appendices

Appendix 8.1. Rank-ordered portfolio data: no controls

P#	FS	Rtn%	SD%	Skew	S Ratio
1	2.98	10.29	8.96	0.802	0.096
2	3.90	6.22	9.11	0.279	0.057
3	6.99	8.84	7.96	0.065	0.093
4	11.70	7.72	7.01	−0.231	0.092
5	14.38	6.10	6.81	−0.300	0.075
6	18.02	8.89	6.41	−0.543	0.116
7	21.51	7.94	6.13	−0.509	0.108
8	24.98	7.75	6.09	−0.519	0.106
9	28.05	9.18	5.76	−0.682	0.133
10	31.50	7.85	5.68	−0.593	0.115
11	34.84	6.48	5.54	−0.714	0.097
12	37.98	7.96	5.56	−0.606	0.119
13	41.04	9.92	5.48	−0.591	0.151
14	44.48	9.20	5.43	−0.528	0.141
15	47.98	9.39	5.39	−0.503	0.145
16	51.03	9.42	5.35	−0.384	0.147
17	54.45	8.79	5.35	−0.417	0.137
18	57.89	9.93	5.54	−0.213	0.149
19	60.96	9.48	5.50	−0.209	0.144
20	64.44	11.74	5.56	−0.241	0.176
21	67.87	9.97	5.69	−0.077	0.146
22	70.89	10.04	5.78	0.034	0.145
23	73.96	10.36	6.08	0.462	0.142
24	77.43	12.39	6.35	0.287	0.163
25	80.87	11.95	6.49	0.530	0.154
26	84.41	12.93	6.95	0.421	0.155
27	87.06	14.29	7.35	1.061	0.162
28	91.08	13.03	8.26	0.993	0.131
29	94.29	15.50	9.57	1.271	0.135
30	96.59	19.12	10.62	2.308	0.150

Appendix 8.2. Rank-ordered portfolio data: only a beta control

P#	FS	Rtn%	SD%	Skew	S Ratio
1	4.21	7.77	7.36	−0.248	0.088
2	5.20	5.74	7.59	0.080	0.063
3	8.54	9.37	7.14	−0.447	0.109
4	13.52	7.72	6.22	−0.617	0.103
5	16.19	7.64	6.40	−0.636	0.099
6	19.69	9.61	5.83	−0.565	0.137
7	23.12	7.73	5.98	−0.612	0.108
8	26.27	7.60	5.81	−0.477	0.109
9	29.14	7.56	5.78	−0.507	0.109
10	32.34	9.83	5.68	−0.629	0.144
11	35.49	7.94	5.56	−0.400	0.119
12	38.51	7.73	5.77	−0.541	0.112
13	41.44	7.68	5.61	−0.567	0.114
14	44.72	8.70	5.64	−0.402	0.128
15	48.02	8.08	5.54	−0.484	0.121
16	50.97	9.39	5.65	−0.404	0.139
17	54.24	8.85	5.59	−0.290	0.132
18	57.51	9.04	5.76	−0.034	0.131
19	60.47	9.17	5.75	0.197	0.133
20	63.79	11.05	5.84	−0.012	0.158
21	67.07	10.47	5.78	−0.180	0.151
22	69.96	10.37	5.92	0.149	0.146
23	72.92	10.89	5.85	−0.010	0.155
24	76.15	12.23	6.09	0.042	0.167
25	79.65	12.76	6.07	0.287	0.175
26	83.18	11.10	6.50	0.632	0.142
27	85.85	13.27	6.78	0.247	0.163
28	89.97	15.10	7.65	1.263	0.164
29	93.48	15.11	8.68	1.483	0.145
30	95.94	19.89	9.55	1.585	0.173

Appendix 8.3. Rank-ordered portfolio data: only a size control

P#	FS	Rtn%	SD%	Skew	S Ratio
1	3.28	10.31	8.90	0.914	0.097
2	4.22	6.77	9.08	0.159	0.062
3	7.31	7.73	7.93	0.137	0.081
4	12.00	7.91	7.03	−0.313	0.094
5	14.65	6.35	6.85	−0.200	0.077
6	18.28	8.70	6.33	−0.556	0.115
7	21.86	8.06	6.06	−0.621	0.111
8	25.20	7.79	6.10	−0.538	0.106
9	28.29	9.07	5.84	−0.602	0.130
10	31.67	7.09	5.61	−0.704	0.105
11	34.97	6.80	5.54	−0.640	0.102
12	38.11	8.49	5.54	−0.537	0.128
13	41.16	9.49	5.47	−0.581	0.145
14	44.51	9.05	5.35	−0.604	0.141
15	47.99	9.34	5.44	−0.413	0.143
16	51.09	9.30	5.33	−0.375	0.145
17	54.40	8.84	5.36	−0.478	0.137
18	57.82	10.07	5.58	−0.279	0.150
19	60.90	9.79	5.42	−0.198	0.150
20	64.26	11.35	5.51	−0.196	0.172
21	67.67	10.17	5.62	−0.035	0.151
22	70.68	10.91	5.95	0.266	0.153
23	73.75	10.31	5.96	0.189	0.144
24	77.07	12.86	6.31	0.340	0.170
25	80.63	11.73	6.52	0.322	0.150
26	84.15	12.78	6.96	0.514	0.153
27	86.80	14.43	7.29	1.045	0.165
28	90.80	13.24	8.21	1.017	0.134
29	94.01	15.06	9.58	1.261	0.131
30	96.28	18.87	10.55	2.329	0.149

Appendix 8.4. Rank-ordered portfolio data: only a BP control

P#	FS	Rtn%	SD%	Skew	S Ratio
1	4.03	10.27	8.83	0.805	0.097
2	5.47	7.47	8.89	0.276	0.070
3	9.10	7.66	7.64	−0.064	0.083
4	14.47	7.46	6.83	−0.460	0.091
5	17.26	8.51	6.71	−0.495	0.106
6	21.01	8.60	6.45	−0.292	0.111
7	24.59	9.23	5.94	−0.657	0.129
8	27.75	7.76	5.88	−0.745	0.110
9	30.61	8.45	5.87	−0.772	0.120
10	33.74	8.09	5.67	−0.515	0.119
11	36.75	7.70	5.38	−0.824	0.119
12	39.56	10.29	5.55	−0.547	0.154
13	42.27	9.68	5.51	−0.690	0.146
14	45.27	8.96	5.38	−0.569	0.139
15	48.26	8.93	5.25	−0.594	0.142
16	50.89	9.74	5.35	−0.425	0.152
17	53.84	9.24	5.33	−0.381	0.144
18	56.82	10.67	5.37	−0.310	0.166
19	59.50	10.67	5.51	−0.077	0.161
20	62.58	10.70	5.49	−0.173	0.162
21	65.68	9.64	5.72	−0.105	0.140
22	68.44	10.50	5.83	−0.038	0.150
23	71.36	11.25	6.02	0.177	0.156
24	74.60	10.88	6.30	0.220	0.144
25	78.16	11.75	6.70	0.565	0.146
26	81.86	12.03	6.92	0.373	0.145
27	84.69	13.33	7.15	0.707	0.155
28	89.21	12.84	8.44	1.469	0.127
29	92.95	15.15	9.62	1.342	0.131
30	95.83	17.85	10.65	2.186	0.140

Appendix 8.5. Rank-ordered portfolio data: risk controls only

P#	FS	Rtn%	SD%	Skew	S Ratio
1	4.21	7.77	7.36	−0.248	0.088
2	5.20	5.74	7.59	0.080	0.063
3	8.54	9.37	7.14	−0.447	0.109
4	13.52	7.72	6.22	−0.617	0.103
5	16.19	7.64	6.40	−0.636	0.099
6	19.69	9.61	5.83	−0.565	0.137
7	23.12	7.73	5.98	−0.612	0.108
8	26.27	7.60	5.81	−0.477	0.109
9	29.14	7.56	5.78	−0.507	0.109
10	32.34	9.83	5.68	−0.629	0.144
11	35.49	7.94	5.56	−0.400	0.119
12	38.51	7.73	5.77	−0.541	0.112
13	41.44	7.68	5.61	−0.567	0.114
14	44.72	8.70	5.64	−0.402	0.128
15	48.02	8.08	5.54	−0.484	0.121
16	50.97	9.39	5.65	−0.404	0.139
17	54.24	8.85	5.59	−0.290	0.132
18	57.51	9.04	5.76	−0.034	0.131
19	60.47	9.17	5.75	0.197	0.133
20	63.79	11.05	5.84	−0.012	0.158
21	67.07	10.47	5.78	−0.180	0.151
22	69.96	10.37	5.92	0.149	0.146
23	72.92	10.89	5.85	−0.010	0.155
24	76.15	12.23	6.09	0.042	0.167
25	79.65	12.76	6.07	0.287	0.175
26	83.18	11.10	6.50	0.632	0.142
27	85.85	13.27	6.78	0.247	0.163
28	89.97	15.10	7.65	1.263	0.164
29	93.48	15.11	8.68	1.483	0.145
30	95.94	19.89	9.55	1.585	0.173

Appendix 8.6. Rank-ordered portfolio data: tax controls only

P#	FS	Rtn%	SD%	Skew	S Ratio
1	10.42	0.30	6.50	−0.420	0.004
2	11.42	0.15	6.45	−0.393	0.002
3	15.46	4.50	6.35	−0.600	0.059
4	20.52	6.09	5.89	−0.737	0.086
5	22.84	4.93	5.98	−0.495	0.069
6	25.76	4.61	5.87	−0.728	0.065
7	28.62	4.90	5.67	−0.642	0.072
8	31.02	6.47	5.57	−0.729	0.097
9	33.27	6.75	5.59	−0.490	0.100
10	35.77	8.13	5.60	−0.638	0.121
11	38.14	7.96	5.47	−0.645	0.121
12	40.46	8.12	5.49	−0.587	0.123
13	42.74	7.33	5.47	−0.607	0.112
14	45.30	8.78	5.59	−0.522	0.131
15	47.90	8.26	5.70	−0.377	0.121
16	50.26	10.91	5.51	−0.407	0.165
17	52.94	9.86	5.60	−0.277	0.147
18	55.63	11.84	5.77	−0.077	0.171
19	58.13	12.02	5.72	−0.269	0.175
20	60.96	11.56	5.83	−0.010	0.165
21	63.80	11.84	5.92	0.047	0.167
22	66.32	10.84	6.02	0.019	0.150
23	68.98	11.79	6.07	0.147	0.162
24	71.98	14.23	6.05	0.069	0.196
25	75.37	12.79	6.38	0.144	0.167
26	78.92	13.99	6.39	0.121	0.182
27	81.66	16.00	6.90	0.604	0.193
28	86.26	14.75	7.38	0.515	0.166
29	90.47	18.25	8.38	0.987	0.181
30	93.65	21.40	8.92	1.789	0.200

Appendix 8.7. Rank-ordered portfolio data: risk and tax controls

P#	FS	Rtn%	SD%	Skew	S Ratio
1	10.42	0.30	6.50	−0.420	0.004
2	11.42	0.15	6.45	−0.393	0.002
3	15.46	4.50	6.35	−0.600	0.059
4	20.52	6.09	5.89	−0.737	0.086
5	22.84	4.93	5.98	−0.495	0.069
6	25.76	4.61	5.87	−0.728	0.065
7	28.62	4.90	5.67	−0.642	0.072
8	31.02	6.47	5.57	−0.729	0.097
9	33.27	6.75	5.59	−0.490	0.100
10	35.77	8.13	5.60	−0.638	0.121
11	38.14	7.96	5.47	−0.645	0.121
12	40.46	8.12	5.49	−0.587	0.123
13	42.74	7.33	5.47	−0.607	0.112
14	45.30	8.78	5.59	−0.522	0.131
15	47.90	8.26	5.70	−0.377	0.121
16	50.26	10.91	5.51	−0.407	0.165
17	52.94	9.86	5.60	−0.277	0.147
18	55.63	11.84	5.77	−0.077	0.171
19	58.13	12.02	5.72	−0.269	0.175
20	60.96	11.56	5.83	−0.010	0.165
21	63.80	11.84	5.92	0.047	0.167
22	66.32	10.84	6.02	0.019	0.150
23	68.98	11.79	6.07	0.147	0.162
24	71.98	14.23	6.05	0.069	0.196
25	75.37	12.79	6.38	0.144	0.167
26	78.92	13.99	6.39	0.121	0.182
27	81.66	16.00	6.90	0.604	0.193
28	86.26	14.75	7.38	0.515	0.166
29	90.47	18.25	8.38	0.987	0.181
30	93.65	21.40	8.92	1.789	0.200

References

Basu, S. (1977). Investment performance of common stocks in relations to their price earnings ratios: A test of market efficiency. *Journal of Finance, 32*, 663–682.

Black, F., Jensen, M. C., & Scholes, M. (1972). The capital asset pricing model: Some empirical tests. In M. C. Jensen (Ed.), *Studies in the theory of capital markets* (pp. 79–122). New York: Praeger Publishers.

Blume, M. (1980). Stock return and dividend yield: Some more evidence. *Review of Economics and Statistics, 62*, 567–577.

Box, G. E. P., Hunter, J. S., & Hunter, W. G. (2005). *Statistics for experimenters*. New York: Wiley Interscience.

Box, G. E. P. (1954). The exploration and exploitation of response surfaces. *Biometrics, 10*, 16–60.

Box, G. E. P., & Draper, N. R. (1987). *Empirical model building and response surfaces*. New York: Wiley Interscience.

Box, G. E. P., & Draper, N. R. (1959). A basis for the selection of a response surface design. *Journal of the American Statistical Association, 54*, 622–654.

Brennen, M. (1970). Taxes, market valuation and financial policy. *National Tax Journal, 23*, 417–429.

Brown, S. J. (2008). Elusive return predictability: Discussion. *International Journal of Forecasting, 24*, 19–21.

Conner, G., & Korajczak, R. A. (1988). Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics, 21*, 255–289.

Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics, 22*, 3–25.

Fama, E. F., & French, K. R. (1992). Cross-sectional variation in expected stock returns. *Journal of Finance, 47*, 427–465.

Fama, E. F., & French, K. R. (2008a). Dissecting anomalies. *Journal of Finance, 63*, 1653–1678.

Fama, E. F., & French, K. R. (2008b). Average returns, B/M, and share issues. *Journal of Finance, 63*, 2971–2995.

Fama, E. F., & MacBeth, J. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy, 81*, 607–636.

Graham, B., & Dodd, D. L. (1934). *Security analysis*. New York: McGraw-Hill Book Company.

Griliches, Z. (1986). Economic data issues. In Z. Griliches & M. Intrilagator (Eds.), *Handbook of econometrics*. Amsterdam: North Holland.

Grinold, R. C., & Kahn, R. N. (2000). *Active portfolio management* (2nd ed.). New York: McGraw-Hill.

Grunfeld, Y., & Griliches, Z. (1960). Is aggregation necessarily bad? *Review of Economics and Statistics, 42*, 1–13.

Guerard, J. B., Jr., Gultekin, M. G., & Stone, B. K. (1997). The role of fundamental data and analysts' earnings breadth, forecasts, and revisions in the creation of efficient portfolios. In A. Chen (Ed.), *Research in finance 15*, 69–91.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *Journal of Finance, 23*, 389–416.

Khuri, A. I., & Cornell, J. A. (1996). *Response surfaces: Designs and analysis*. New York: Marcel Dekker, Inc.

Lease, R. C., John, K., Kalay, A., Loewenstein, U., & Sarig, O. H. (2000). *Dividend policy: Its impact on firm value*. Boston: Harvard Business School Press.

Markowitz, H. M. (1987). *Mean-variance analysis in portfolio choice and capital markets*. Oxford: Basil Blackwell.

Markowitz, H. M. (1991). Foundations of portfolio theory. *Journal of Finance, 46*(2), 469–477.

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

Markowitz, H. M. (1959). *Portfolio selection: Efficient diversification of investment*. New York: John Wiley & Sons. Cowles Foundation monograph no. 16.

Markowitz, H. M., & Xu, G. (1994). Data mining corrections. *Journal of Portfolio Management, 21*, 60–69.

Menchero, J., Morozov, A., & Shepard, P. (2010). Global equity risk modeling. In J. B. Guerard Jr. (Ed.), *Chapter 15, Handbook of portfolio construction* (pp. 439–480). New York: Springer.

Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, *45*, #1 (January/February 1989).

Michaud, R. O. (1998). *Efficient asset management*. Boston: Harvard Business School Press.

Myers, R. H., Khuri, A. I., & Cornell, W. H., Jr. (1989). Response surface methodology: 1966–1988. *Technometrics, 31*, 137–157.

Peterson, P., Peterson, D., & Ang, J. (1985). Direct evidence on the marginal rate of taxation on dividend income. *Journal of Financial Economics, 14*, 267–282.

Pilotte, E. A. (2003). Capital gains, dividend yields, and expected inflation. *Journal of Finance, 58*, 447–466.

Rosenberg, B. (1974). Extra-market components of covariance in security returns. *Journal of Financial and Quantitative Analysis, 9*, 263–274.

Rosenberg, B., & Rudd, A. (1982). Factor-related and specific returns of common stocks: Serial correlation and market inefficiency. *Journal of Finance, 37*, 543–554.

Rosenberg, B., & Marathe, V. (1979). Tests of capital asset pricing hypotheses. In H. Levy (Ed.), *Research in finance 1*. Bingley, UK: Emerald Publishing Group Limited.

Rosenberg, B., & Rudd, A. (1977). The yield/beta/residual risk tradeoff. Working paper 66. Research Program in Finance, Institute of Business and Economic Research, University of California, Berkeley.

Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business, 39*, 119–138.

Sharpe, W. F. (1963). A simplified model for portfolio analysis. *Management Science, 9*, 277–293.

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

Sharpe, W. F. (1967). A linear programming algorithm for mutual fund portfolio selection. *Management Science, 13*, 449–510.

Sharpe, W. F. (1971). A linear programming approximation for the general portfolio problem. *Journal of Financial and Quantitative Analysis, 6*, 1263–1275.

Stone, B. K. (1973). A linear programming formulation of the general portfolio selection problem. *Journal of Financial and Quantitative Analysis, 8*, 621–636.

Stone, B. K., Adolphson, D. L., & Miller, T. W. (1993). Optimal data selection and grouping: An extension of traditional response surface methodology to observational studies involving non-controlled empirical data generation. *Advances in Mathematical Programming and Financial Planning, 3*, 39–68.

Stone, B. K., & Guerard, J. B., Jr. (2010). Methodologies for isolating and assessing the portfolio performance potential of stock return forecast models with an illustration. In J. B. Guerard Jr. (Ed.), *Chapter 10, Handbook of portfolio construction* (pp. 259–336). New York: Springer.

Stone, B. K., & Guerard, J. B., Jr. (2011). Methodology for testing stock return forecast models for performance and factor alignment/misalignment: Alpha performance potential for Japan. *Proceedings of the 2011 CRSP Forum*. Chicago: University of Chicago.

Stowe, J. D., Robinson, T. R., Pinto, J. E., & McLeavey, D. W. (2007). *Equity asset valuation*. Hoboken NJ: John Wiley & Sons for CFA Institute and Association for Investment Management and Research.

Timmerman, A. (2008a). Elusive return predictability. *International Journal of Forecasting, 24*, 1–18.

Timmerman, A. (2008b). Reply to discussion of elusive return predictability. *International Journal of Forecasting, 24*, 29–30.

Treynor, J. (1965). How to rate the management of investment funds. *Harvard Business Review, 43*(January-February 1965), 63–75.

Treynor, J., & Mazuy, K. (1966). Can mutual funds outperform the market? *Harvard Business Review, 44*(January-February 1966), 131–136.

Williams, J. B. (1938). *The theory of investment value*. Cambridge: Harvard University Press.