For the **Handbook of Financial Econometrics and Statistics** (2012), *C.F. Lee, Editor*

Optimal Orthogonal Portfolios with Conditioning Information

First draft August 10, 2007, this revision: August 28, 2011

ABSTRACT

Optimal orthogonal portfolios are a central feature of tests of asset pricing models and are important in active portfolio management problems. The portfolios combine with a benchmark portfolio to form ex ante mean variance efficient portfolios. This paper derives and characterizes optimal orthogonal portfolios in the presence of conditioning information in the form of a set of lagged instruments. In this setting, studied by Hansen and Richard (1987), the conditioning information is used to optimize with respect to the unconditional moments. We present an empirical illustration of the properties of the optimal orthogonal portfolios. From an asset pricing perspective, a standard stock market index is far from efficient when portfolios trade based on lagged interest rates and dividend yields. From an active portfolio management perspective, the example shows that a strong tilt toward bonds improves the efficiency of equity portfolios.

Our analytical results provide economic interpretation for test statistics like the Wald test or multivariate *F* test used in asset pricing research. The empirical applications in this paper make use of regression and maximum likelihood parameter estimation, as well as method of moments estimation. We form maximum likelihood estimates of nonlinear functions as the functions evaluated at the maximum likelihood parameter estimates.

KEY WORDS: Asset Pricing Tests, Optimal Portfolios, Portfolio Management, Method of Moments, Maximum Likelihood, Parametric Bootstrap.

1. Introduction

The optimal orthogonal portfolio, also known as the most mispriced portfolio or the active portfolio, is a central concept in asset pricing tests and in modern portfolio management. In asset pricing problems, it represents the difference between the performance of a benchmark portfolio and the maximum potential performance in a sample of assets (Jobson and Korkie, 1982). In modern portfolio management it shows how to actively tilt away from a given benchmark portfolio to achieve portfolio efficiency (Gibbons, Ross and Shanken (1989) or Grinold and Kan (1992).

Optimal orthogonal portfolios are studied by Roll (1980), MacKinlay (1995), Campbell, Lo and MacKinlay (1997) and others. However, these studies restrict the analysis to a setting where the portfolio weights are fixed over time. In contrast, studies in asset pricing use predetermined variables to model conditional expected returns, correlations and volatility. Portfolio weights may be functions of the predetermined variables, and they will generally vary over time. Quantitative portfolio managers routinely use conditioning information in optimized portfolio strategies. Therefore, it is important to understand optimal orthogonal portfolios in a conditional setting.

This paper derives, characterizes and illustrates optimal orthogonal portfolios in a conditional setting. The setting is one where the conditional means and variances of returns are time-varying, and optimal time-varying portfolio weights achieve unconditional mean-variance efficiency with respect to the information, as described by Hansen and Richard (1987) and

Ferson and Siegel (2001).[1] Ferson and Siegel (2001) argue that this setting is interesting from the perspective of active portfolio management, where the client can't observe the information that a portfolio manager may have. Ferson and Siegel (2003, 2009) show that this setting is also interesting from the perspective of testing asset pricing models, but they do not develop the optimal orthogonal portfolio.

We show that the optimal orthogonal portfolio has time-varying weights and we derive the weights in closed form. The portfolio weights for unconditionally efficient portfolios in the presence of conditioning information are derived by Ferson and Siegel (2001). They consider the case with no risk-free asset and the case with a fixed risk-free asset whose return is constant over time. We generalize these solutions to the case with a "conditionally" risk-free asset whose return is known at the beginning of the period, and is thus included in the lagged conditioning information, and may vary over time. We derive solutions for the optimal orthogonal portfolios with conditioning information, including cases where there is no risk-free asset, a constant risk-free rate, or a time-varying conditionally risk-free rate. We show that a "law of conservation of squared Sharpe ratios" holds, implying that the optimal orthogonal portfolio's squared unconditional Sharpe ratio is the difference between that of the benchmark portfolio and the maximum unconditional squared Sharpe ratio that is possible using the assets and conditioning information. Empirical examples illustrate the performance of the optimal orthogonal portfolios with conditioning information and the behavior of the portfolio weights.

---

[1] An alternative is to study conditional efficiency, where the weights minimize the conditional variance. This may be handled by simply reinterpreting the classical analysis.

Section 2 briefly reviews optimal orthogonal portfolios in the classical case with no conditioning information. Section 3 describes the setting for our analysis with conditioning information. Section 4 presents the main theoretical results and Section 5 presents our empirical examples. Section 6 concludes. The Appendix includes the proofs of the main results and a Methodology Appendix describes our empirical examples in detail, including a general description of the parametric bootstrap.

2. Optimal Orthogonal Portfolios: The Classical Case

In the classical case portfolio weights are fixed constants over time and there is no conditioning information. Optimal orthogonal portfolios are tied to mean variance efficiency. Mean variance efficient portfolios maximize the expected return, given the variance of the return. Since Markowitz (1952) and Sharpe (1964), such portfolios have been at the core of financial economics.

The mean variance efficiency of a given portfolio can be described using a system of time-series regressions. If $r_t = R_t - \gamma_0$ is the vector of N *excess* returns at time t, measured in excess of a given risk-free or zero-beta return, γ_0, and $r_{pt} = R_{pt} - \gamma_0$ is the excess return on a benchmark portfolio, the regression system is

$$r_t = \alpha + \beta r_{pt} + u_t; \quad t = 1,...,T , \tag{1}$$

where T is the number of time-series observations, β is the N-vector of regression slopes or betas and α is the N-vector of intercepts or alphas. The tested portfolio r_{pt} is represented among the returns in r_t, so the covariance matrix of the residuals in (1) is singular (r_{pt} might be included explicitly, or might be a fixed-weight portfolio of the assets in r_t). The portfolio r_p is minimum-variance efficient and has the given zero-beta return only if $\alpha = 0$.[2] The mean variance efficiency of a given portfolio is of normative investment interest, as an efficient portfolio maximizes a concave utility function defined solely over the mean and variance of the portfolio return, as would follow from normally distributed returns in a single-period model, for example. Equation (1) may be interpreted as referring to multiple factor portfolios, where r_p is a *K*-vector and β is an $N \times K$ matrix. Then, the benchmark portfolio is a linear combination of the *K* returns in the vector r_p (e.g., Shanken (1987), Gibbons, Ross and Shanken, 1989).

Definition: The *most mispriced (or optimal orthogonal) portfolio with respect to* r_p*, when there is no conditioning information, has excess return* $r_c = x_c' r$*, where the weights* x_c *satisfy*:

$$x_c = \underset{x}{\text{Arg Max}} \frac{(x'\alpha)^2}{Var(x'r)} \tag{2}$$

It is clear from the definition in (2) why the portfolio is referred to as the most mispriced. The vector α captures the "mispricing" of the tested asset returns in (1) when evaluated using the benchmark r_p, and the portfolio x_c has the largest squared alpha relative to its variance. This interpretation also reveals why the portfolio is of central interest in active portfolio management. Given a benchmark portfolio r_p, an active portfolio manager places bets by deviating from the

[2] Note the distinction between *minimum variance efficient* portfolios, which minimize the variance for the given mean return, and *mean variance efficient*, which maximize the mean return given its variance. The latter set of portfolios is a subset of the former, typically depicted as the positively-sloped portion of the minimum variance efficient boundary when graphed with mean return on the *y*-axis and standard deviation or variance of return on the *x*-axis. The portfolio r_p is mean variance efficient when $\alpha = 0$ and $E(r_p)$ exceeds the expected excess return of the global minimum-variance portfolio.

portfolio weights that define the benchmark. The manager is rewarded for bets that deliver higher returns and penalized for increasing the volatility. The portfolio in (2) describes the active bets that achieve the largest amount of extra return for the variance. Thus, the solution is also referred to as the *active* portfolio by Gibbons, Ross and Shanken (1989). (See Grinold and Kan (1992) for an in-depth treatment of modern portfolio management.)

In the classical case of fixed portfolio weights the solution to (2) is given by $x_c = \left(\underline{1}'V^{-1}\alpha\right)^{-1} V^{-1}\alpha$, where $\underline{1}$ is an *N*-vector of ones and $V = Cov(r)$, the covariance matrix of the returns. Using this solution, several well-known properties of the optimal orthogonal portfolio follow.[3] For example, a combination of the portfolio r_c and the benchmark portfolio r_p is *optimal*; that is, minimum variance efficient (Jobson and Korkie, 1982). The portfolio is *orthogonal* to the benchmark portfolio in the sense that $Cov\left(x_c'r, r_p\right) = 0$.

The optimal orthogonal portfolio is central for the interpretation of tests of portfolio efficiency. Classical test statistics for the hypothesis that $\alpha = 0$ in Equation (1) can be written in terms of squared Sharpe ratios (e.g., Jobson and Korkie, 1982). Consider the Wald Statistic:

$$W = T\hat{\alpha}'[Cov(\hat{\alpha})]^{-1}\hat{\alpha} = T\left(\frac{\hat{S}^2(R) - \hat{S}^2(R_p)}{1+\hat{S}^2(R_p)}\right) \sim \chi^2(N) \tag{3}$$

where $\hat{\alpha}$ is the OLS or maximum likelihood (ML) estimator of α (after removing r_p or another asset from the vector *r* to avoid singularity of the covariance matrix) and $Cov(\hat{\alpha})$ is its

---

[3] See Roll (1980), Gibbons, Ross and Shanken (1989), MacKinlay (1995) and Lo and MacKinlay (1997) for analyses of optimal orthogonal portfolios in the classical case with no conditioning information.

asymptotic covariance matrix. Upper case *R*'s refer to gross returns and lower case *r*'s refer to returns in excess of the zero beta rate. The term $\hat{S}^2\left(R_p\right)$ is the sample value of the squared Sharpe ratio of R_p when the zero beta rate is γ_0 so that $S^2\left(R_p\right) = \left[E\left(r_p\right)/\sigma\left(r_p\right)\right]^2$. The term $\hat{S}^2\left(R\right)$ is the sample value of the maximum squared Sharpe ratio that can be obtained by portfolios of the assets in *R* (including R_p):

$$S^2(R) = \max_x \left\{ \frac{[E(x'r)]^2}{Var(x'r)} \right\}. \tag{4}$$

The Wald statistic has an asymptotic chi-squared distribution with *N* degrees of freedom, under the null hypothesis that R_p is efficient with the given zero-beta return. Scaled with a degrees-of-freedom adjustment, the statistic has an *F* distribution under normally distributed returns (Gibbons, Ross and Shanken, 1989).

It can be shown that the squared Sharpe ratios can be decomposed using the optimal orthogonal portfolio as $S^2\left(R\right) = S^2\left(R_p\right) + S^2\left(R_c\right)$. A similar decomposition holds at the sample values. This decomposition, a "law of conservation of squared Sharpe ratios" is used by Jobson and Korkie (1982) to derive the second equality in (3). Since the Sharpe ratio is the slope of a line in the mean-standard deviation space, Equation (3) suggests a graphical representation for the statistic in the sample mean standard deviation space. It measures the distance between the sample mean-standard deviation frontier and the location of the tested portfolio, inside the frontier. This distance is proportional to the squared Sharpe ratio of the optimal orthogonal portfolio. Kandel (1984), Roll (1985), Gibbons, Ross and Shanken (1989) and Kandel and Stambaugh (1987, 1989) further develop this interpretation.

3. The Conditional Setting

We use conditioning information in a setting similar to that of Hansen and Richard (1987) and Ferson and Siegel (2001), where minimum variance efficiency is defined in terms of the unconditional means and variances of the portfolios that result from the use of conditioning information. Ferson and Siegel (2009) refer to this as efficiency *with respect to the information*, *Z*. This setting has proven useful in asset pricing tests (Ferson and Siegel 2003, 2009, Bekaert and Liu, 2004), in forming hedging portfolios (Ferson, Siegel and Xu, 2006) and in portfolio management problems (Ahkbar, Devraj and Stremme 2007, Chiang, 2009). We study the optimal orthogonal portfolio in this setting. The distinction between mean variance efficiency and minimum variance efficiency, as in the classical setting, applies in this setting as well.

Consider a portfolio of N assets with gross returns, R_{t+1}, where the weights that determine the portfolio at time t are functions of the information, Z_t. The gross return on such a portfolio with weight $x(Z_t)$, is $x'(Z_t)R_{t+1}$. The restrictions on the portfolio weight function are that the weights must sum to 1 (almost surely in Z_t), and that the unconditional expected value and second moments of the portfolio return are well defined. Consider now all portfolio returns that may be formed, for a given set of asset returns R_{t+1} and given conditioning information, Z_t, with well-defined first and second moments. This set determines a mean-standard deviation frontier, as shown by Hansen and Richard (1987). This frontier depicts the *unconditional* means versus the *unconditional* standard deviations of the portfolio returns. A portfolio is defined to be efficient with respect to the information Z_t if and only if it is on this mean standard deviation frontier.

Ferson and Siegel (2001) derive solutions for efficient-with-respect to *Z* portfolios in closed form. They consider the case with no risk-free asset and the case with a fixed risk-free asset whose return is constant over time. In Theorem 1 of the Appendix, we derive the solution for the case with a risk-free asset whose return is known at the beginning of the period, and is thus included in the information *Z*, and may vary over time. In this case the variation in the risk-free rate over time affects the unconditional variance of the portfolio return.

Ferson and Siegel (2001) argue that efficiency with respect to the information is especially relevant in a portfolio management context. It is reasonable to assume that the portfolio manager has more information about asset returns than the client. Assume that the client desires an unconditionally mean-variance efficient portfolio. The manager observes conditioning information that is relevant about future returns, and by conditioning on this information he can expand the investor's opportunity set. The manager maximizes the investor's mean variance opportunity set by using his information to maximize the unconditional mean for a given unconditional variance. The efficient-with-respect to *Z* strategy is therefore the strategy that the investor would wish the portfolio manager to use.

Ferson and Siegel (2009) show how asset pricing theories make statements about portfolios that are efficient with respect to information *Z*, and develop tests of the hypothesis that a portfolio is efficient with respect to *Z*. The optimal orthogonal portfolio with respect to information *Z* is a useful concept in these portfolio efficiency tests. We begin the analysis with a result from Hansen and Richard (1987).

<u>Proposition 1</u>: [Hansen and Richard, 1987, Corollary 3.1.] Given *N* asset gross returns, R_{t+1}, a given portfolio with gross return $R_{p,t+1}$ is *minimum-variance efficient with respect*

to the information Z_t if and only if Equation (5) is satisfied (equivalently, there exists constants γ_0 and γ_1 such that Equation (6) is satisfied) for all $x(Z_t)$ such that $x'(Z_t)\underline{1}=1$ almost surely, where the relevant unconditional expectations exist and are finite:

$$Var\left(R_{p,t+1}\right) \le Var\left[x'(Z_t)R_{t+1}\right] \quad \text{if} \quad E\left(R_{p,t+1}\right) = E\left[x'(Z_t)R_{t+1}\right] \tag{5}$$

$$E\left[x'(Z_t)R_{t+1}\right] = \gamma_0 + \gamma_1 Cov\left[x'(Z_t)R_{t+1}, R_{p,t+1}\right]. \tag{6}$$

Equation (5) is the *definition* of efficiency with respect to *Z*. It states that $R_{p,t+1}$ is on the minimum variance boundary formed by all possible portfolios that use the assets in *R* and the conditioning information. Equation (6) states that the familiar expected return - covariance relation from Fama (1973) and Roll (1977) must hold with respect to the efficient portfolio. In Equation (6), the coefficients γ_0 and γ_1 are fixed scalars that do not depend on the functions $x(\cdot)$ or the realizations of Z_t .

4. The Main Results

The optimal orthogonal portfolio with conditioning information plays roles analogous to the classical setting with no conditioning information. Thus, for example, restricting the maximization in Equation (4) to *fixed-weight* portfolios where *x* is a constant vector, we obtain efficiency in the classical case. In contrast, an efficient portfolio with respect to the information *Z* maximizes the squared Sharpe ratio over *all portfolio weight functions*, $x(Z)$. Maximizing over a larger set of weights expands the investment opportunity set and produces a larger maximum Sharpe ratio.

With conditioning information the optimal orthogonal portfolio's weight function is time-varying and we derive this portfolio weight for three cases. First, with no risk-free asset in which case a fixed unconditional "zero-beta" rate γ_0 is arbitrarily chosen. By varying the zero-beta rate the solutions can describe any point on the efficient-with-respect-to-*Z* frontier. Second, we consider a conditionally time-varying risk-free asset whose return $R_f = R_f(Z)$ is measureable and thus known as part of the information set *Z* so that $Var\left[R_f(Z) | Z\right] = 0$, but which is unconditionally risky in the sense that $Var\left[R_f(Z)\right] > 0$. Here we again choose an arbitrary zero-beta rate γ_0 to describe the frontier. In the third case, there exists an unconditional risk-free asset with fixed return $R_f = \gamma_0$. In this case, the efficient-with-respect-to-*Z* frontier becomes a line passing through γ_0 (at risk zero) and the point representing the mean and standard deviation of a particular portfolio strategy's return.

We consider portfolios formed from the risky assets using weights $x_q = x_q(Z)$ where the weights must sum to 1 (for all *Z*) when there is no risk-free asset. This constraint is relaxed when there is a conditional or unconditional risk-free asset (where the implicit weight in the risk-free asset is then set at 1 minus the sum of the weights in the risky assets). When there is no risk-free asset the portfolio return is $R_{q,t+1} = x_q'(Z_t) R_{t+1}$, and when there is a risk-free asset the portfolio return is $R_{q,t+1} = R_f(Z_t) + x_q'(Z_t)\left[R_{t+1} - R_f(Z_t)\underline{1}\right]$ whether or not $R_f(Z)$ is constant. In either case we denote the (unconditional) portfolio mean $\mu_q = E(R_q)$ and variance $\sigma_q^2 = Var(R_q)$. When there exists a conditional time-varying risk-free asset, the portfolio takes advantage of the ability to adapt both the percentage invested in risky assets and their portfolio weights in response to the information *Z*. This may be interpreted as "market timing" and security selection, respectively. We define the optimal orthogonal portfolio with respect to a given benchmark portfolio *P* formed from the risky assets using (possibly) time-varying weights $x_p = x_p(Z)$.

Definition: The *most mispriced (or optimal orthogonal) portfolio,* R_c*, with respect to the benchmark portfolio* R_p *and conditioning information Z*, with portfolio weight denoted $x_c(Z)$, uses the conditioning information to maximize α_c^2/σ_c^2 where the unconditional variance of R_c is σ_c^2, the unconditional mean is $\mu_c = E(R_c)$, the unconditional alpha of R_c with respect to R_p is $\alpha_c = \mu_c - \left[\gamma_0 + (\mu_p - \gamma_0)\sigma_{cp}/\sigma_p^2\right]$, the zero-beta rate is γ_0, and the unconditional covariance is $\sigma_{cp} = Cov(R_c, R_p)$.

Proposition 2: The unique most mispriced (or optimal orthogonal) portfolio R_c with respect to a given benchmark portfolio R_p (with weights x_p, and expected return μ_p), conditioning information *Z* and given zero-beta rate, γ_0, has the following portfolio weight in each of the following cases. If there is no risk-free rate, then the weights conditionally sum to 1 as defined by

$$x_c(Z) = A\left\{\frac{\Lambda \underline{1}}{\underline{1}'\Lambda\underline{1}} + \left[(c+1)\mu_s + b\right]\left(\Lambda - \frac{\Lambda\underline{1}\underline{1}'\Lambda}{\underline{1}'\Lambda\underline{1}}\right)\mu(Z)\right\} + Bx_p\,, \tag{7}$$

while if there is a risk-free rate $R_f(Z)$ (which may be either constant or time varying, but if constant[4] then we must choose $\gamma_0 = R_f$ along with any $\mu_s \neq R_f$) and the solution is:

$$x_c(Z) = A\left[(c+1)\mu_s + b - R_f\right]Q\left[\mu(Z) - R_f \underline{1}\right] + Bx_p, \quad (8)$$

where

$$A = \frac{(\mu_s - \gamma_0)/\sigma_s^2}{(\mu_s - \gamma_0)/\sigma_s^2 - (\mu_p - \gamma_0)/\sigma_p^2}, \quad (9)$$

$$B = -\frac{(\mu_p - \gamma_0)/\sigma_p^2}{(\mu_s - \gamma_0)/\sigma_s^2 - (\mu_p - \gamma_0)/\sigma_p^2}, \quad (10)$$

$$\mu_s = -(a + b\gamma_0)/(b + c\gamma_0), \quad (11)$$

$$\sigma_s^2 = a + 2b\mu_s + c\mu_s^2, \quad (12)$$

$$\begin{aligned} Q = Q(Z) &\equiv \left\{E\left[(R - R_f\underline{1})(R - R_f\underline{1})' \middle| Z\right]\right\}^{-1} \\ &= \left\{\left[\mu(Z) - R_f\underline{1}\right]\left[\mu(Z) - R_f\underline{1}\right]' + \Sigma_\varepsilon(Z)\right\}^{-1}, \end{aligned} \quad (13)$$

$$\Lambda = \Lambda(Z) \equiv \left\{E\left[RR' \middle| Z\right]\right\}^{-1} = \left[\mu(Z)\mu'(Z) + \Sigma_\varepsilon(Z)\right]^{-1}. \quad (14)$$

The constants *a*, *b*, and *c* are defined in the Appendix in Theorem 1 (when there exists a risk-free rate) and in Theorem 2 (when there is no risk-free rate).

---

[4] Equation (11) cannot be used to determine μ_s when $R_f(Z)$ is almost surely constant due to division by zero and, in this case, every choice $\mu_s \neq R_f$ uses the same (rescaled) portfolio of risky assets $Q\left[\mu(Z) - R_f\underline{1}\right]$ in the formation of an efficient portfolio $x_s(Z)$.

Proof: See the Appendix.

The term μ_s represents the unconditional expected return of the efficient-with-respect to Z portfolio, R_s, that maximizes the squared unconditional Sharpe ratio in (4) over all portfolio weight functions with respect to the given value of γ_0. When there exists a risk-free rate $R_f(Z)$ (that may be time-varying because its value is included in the information set Z at the beginning of the period) the conditional mean variance boundary, the tangency intercept $R_f(Z)$, and thus the location of the conditionally mean variance efficient portfolio may vary over time as $R_f(Z)$ varies. When there is no risk-free rate, or when the risk-free rate is time varying, we use the parameter γ_0 to determine a fixed location on the (curved) unconditionally efficient-with-respect to Z boundary; however, when there is a fixed risk-free rate this boundary is a degenerate hyperbola and every portfolio on the upper line is efficient. We next show that the optimal orthogonal portfolio R_c can be formed by combining the benchmark portfolio R_p with the efficient-with-respect to Z portfolio R_s, and from this result it then follows that R_p and R_c can be combined to produce the efficient-with-respect to Z portfolio R_s.

Proposition 3: The most mispriced or optimal orthogonal portfolio R_c may be found as a fixed linear combination of the benchmark portfolio R_p and the efficient-with-respect to Z portfolio, R_s (that maximizes the squared Sharpe ratio for the given zero beta rate, γ_0) as follows:

$$R_c = \frac{\left[\left(\mu_s - \gamma_0\right)/\sigma_s^2\right] R_s - \left[\left(\mu_p - \gamma_0\right)/\sigma_p^2\right] R_p}{\left[\left(\mu_s - \gamma_0\right)/\sigma_s^2\right] - \left[\left(\mu_p - \gamma_0\right)/\sigma_p^2\right]} \quad (15)$$

or

$$R_c = \frac{\sigma_p^2 R_s - \sigma_{ps} R_p}{\sigma_p^2 - \sigma_{ps}} = AR_s + BR_p \quad \text{with} \quad A + B = 1, \quad (16)$$

where we assume that R_s and R_p are not perfectly correlated and that $\sigma_p^2 \neq \sigma_{sp}$

Proof: See the Appendix.

Propositions 2 and 3 extend the concept of the active or optimal orthogonal portfolio to the setting of efficiency with respect to given conditioning information. Given a benchmark portfolio R_p, its optimal orthogonal portfolio with respect to *Z* shows how to tilt away from the benchmark weights to obtain efficiency with respect to *Z*. The portfolio R_c has weights that depend on *Z*. Thus, the optimal tilt away from a benchmark uses the manager's information Z in a dynamic way.

Equation (16) shows how the optimal orthogonal portfolio R_c can be formed by combining an efficient-with-respect to *Z* portfolio R_s with R_p. The portfolio R_c is the regression error of R_s projected on R_p, normalized so that the weights sum to 1, as can be seen by solving Equation (16) for R_s. Thus, the portfolio R_c is uncorrelated with R_p.

We defined the most mispriced portfolio with conditioning information as maximizing the squared alpha relative to the unconditional variance of R_c. Since R_c is orthogonal to R_p, its residual variance in regression (1) is the same as its total variance. Thus, we can think of the optimal orthogonal portfolio as maximizing alpha given its residual variance among all orthogonal portfolios.

Given a benchmark portfolio with return R_p, the optimal orthogonal portfolio with conditioning information is useful for active portfolio management. It might seem natural for a manager with information *Z* to simply reinterpret the classical analysis, where all the moments are the conditional moments given *Z*. This, in fact, is the interpretation that much of the literature on active portfolio management has used (e.g., Jorion (2003), Roll, 1982). This approach produces *conditionally* mean variance efficient portfolios given *Z*. However, as shown by Dybvig and Ross (1985), a conditionally efficient portfolio is likely to be seen as inefficient from the (unconditional) perspective of a client without access to the information *Z*. The optimal orthogonal portfolio describes the active portfolio bets that deliver optimal performance from the client's perspective.

Let $S^2(R)$ be the maximum squared Sharpe ratio obtained by the efficient with respect to *Z* portfolio, and let R_c be the optimal orthogonal portfolio with respect to R_p and information *Z*.

Proposition 4: "*Law of Conservation of Squared Sharpe Ratios*." For a given zero beta or risk-free rate, if $S_s^2 = S^2(R)$ is the maximum squared Sharpe ratio obtained by portfolios $x(Z)$ and R_c is the optimal orthogonal portfolio with respect to R_p and information *Z* at that zero beta rate, then $S_s^2 = S_p^2 + S_c^2$, where S_i^2 denotes the squared Sharpe ratio of portfolio *i*.

Proof: (See the Appendix.)

Proposition 4 shows that if we test the hypothesis that a portfolio is efficient with respect to *Z* using versions of the test statistic in (3), as developed in Ferson and Siegel (2009), then the role of the optimal orthogonal portfolio with information *Z* is analogous to the role of the optimal orthogonal portfolio in the case of the classical test statistics. The Sharpe ratio of the optimal orthogonal portfolio with conditioning information indicates how far the tested benchmark portfolio is from the efficient-with-respect to *Z* frontier. The squared Sharpe ratio of the optimal orthogonal portfolio is the numerator of the test statistic. This numerator and thus the test statistic is zero only if the tested portfolio is efficient in the sample, and it grows larger as the tested portfolio is further from efficiency.

5. Empirical Examples

We present empirical examples using data on portfolios of common stocks, where the firms are grouped according to conventional criteria. We use the returns of common stocks sorted according to market capitalization and book-to-market ratios, focusing on value-weighted decile portfolios of small-capitalization stocks, value stocks (high book/market) and growth stocks (low book/market), as provided on Ken French's web site. We also include a long-term Government bond return, splicing the Ibbotson Associates 20 year US Government bond return series for 1931-1971, with the CRSP greater than 120 month US Government bond return after 1971. The market portfolio, measured as the CRSP value-weighted stock return index, is the benchmark or tested portfolio, R_p. The risk-free return is the return from rolling over one-month Treasury bills from CRSP. We use its sample average, 3.8% per year, as the fixed zero-beta rate in all of the examples. As conditioning information in *Z*, its return is lagged by one year. All of the returns are discretely-compounded annual returns, and the sample period is 1931-2007.

The lagged instruments, *Z*, are the lagged Treasury return and the log of the market price/dividend ratio at the end of the previous year. In calculating the price/dividend ratio, the stock price is the real price of the S&P500 Index, and the dividend is the real dividends accruing to the Index over the past year. These data are from Robert Shiller's website.

We treat the risk-free asset in three distinctly different ways to highlight the three different versions of our solutions for the optimal orthogonal portfolios. In the first case the risk-free rate is assumed to be a fixed constant. Here we do not include the Treasury return as a lagged instrument and we set γ_0 = 3.8% to be the average Treasury return during the sample. The target mean μ_s of the efficient-with-respect to *Z* portfolio is set equal to the sample mean return of the market index in this case, which determines the amount of leverage the portfolio uses at the fixed risk-free rate.

In the second case no risk-free asset exists. Here we again set γ_0 = 3.8% to pick a point on the mean variance boundary, and we do not allow the portfolio to take a position in a risk-free asset. We do allow the lagged Treasury return as conditioning information in *Z*, which highlights the information in lagged Treasury returns about the future risky asset returns. In the third case there is a conditionally risk-free return that is contained in *Z*. Here we use the lagged Treasury return in the conditioning information, and we allow the portfolio to trade the subsequent Treasury return in addition to the other risky assets. The subsequent Treasury return is not really known *ex ante* as the formula assumes, but its correlation with the lagged return is 0.92 during our sample, and this allows us to implement the time-varying risk-free rate example in a somewhat realistic way. (In reality, there is no *ex ante* risk-free asset given the importance of inflation risk.) In practical terms, this example highlights the effects of "market timing," or varying the amount of "cash" in the portfolio, in addition to varying the allocation among the risky assets.

Table 1 summarizes the results. The rows show results for the benchmark index (Market), three equity portfolios and the government bond return. The CAPM α refers to the intercept in the regression (1), of the portfolio returns in excess of the Treasury bill returns, on the excess return of the market index. The small stock portfolio has the largest alpha, at 3.95% per year, while the growth stock portfolio has a negative alpha. The symbol σ_u refers to the standard deviation of the regression residuals. The small stock portfolio has the largest σ_u or nonmarket risk, at more than 25% per year.

The bottom four rows of Panel A summarize the optimal orthogonal portfolios when the market index is the benchmark. The fixed-weight portfolio R_c uses no conditioning information. Its alpha is larger than any of the separate assets, at 4.66% per year, and its residual standard deviation is also relatively large, at 16.5% per year. Since the portfolio is orthogonal to the market index, its residual standard deviation is the same as its total standard deviation, or volatility of return. The ratio of the alpha to the residual volatility is known as the Appraisal ratio (Treynor and Black, 1973) or the Information ratio (Grinold and Kan, 1992). Optimal orthogonal portfolios try to maximize the square of this ratio. The fixed-weight portfolio R_c delivers an information ratio of 0.282, substantially larger than those of the small stock or value portfolios, at 0.155 and 0.177 respectively, and also larger than the bond portfolio, which has an information ratio of 0.183 by virtue of its relatively small volatility.

Table 1 summarizes performance statistics for the optimal orthogonal portfolios with conditioning information. There are three versions with: (1) a fixed risk-free rate, (2) no risk-free rate and (3), a time-varying conditional risk-free rate. The information ratios in all three cases are larger than those of the individual assets or the fixed-weight active portfolio, which illustrates the potential value of using conditioning information explicitly in a portfolio management context (see also, Chiang, 2009). The improvement over fixed weights is modest for the case with no risk-free asset. However, the information ratio is about three times as large, at 0.955, in the example with a time-varying risk-free rate. This illustrates the potential usefulness of interest rate information and the ability to hold "cash" as a function of market conditions, in a portfolio management context.

The averages over time of the optimal orthogonal portfolios' weights on the risky assets are shown in the right hand columns of Panel A. The weights are normalized to sum to 1.0. The weights x_c of the optimal orthogonal portfolios combine with the benchmark (whose weights, x_p, are 100% in the market index) to determine an efficient portfolio. The overall efficient portfolio weights x_s therefore vary over time and depend on how x_c and x_p are combined, which is determined by the coefficients A and B in equations (9) and (10). The estimated weights are as follows. In the fixed risk-free rate case, $x_s = 0.60\, x_c + 0.40\, x_p$. In the no risk-free rate case, $x_s = 0.23\, x_c + 0.77\, x_p$. In the time-varying risk-free rate case, $x_s = 0.44\, x_c + 0.56\, x_p$. Thus, the efficient portfolio is formed as a convex combination of the benchmark and the active portfolio, with reasonable weights in each case.

The four right hand columns of Panel A show that two of the active portfolios take short positions in the market benchmark, indicating an optimal tilt away from the market index. The fixed-weight portfolio R_c takes an extreme short position of -126%, while on average the portfolio using the conditioning information but no risk-free asset takes a position of -21% in the market index. These short positions finance large long positions in the US government bond, and also long positions (in most cases) in small stocks, value stocks and growth stocks. It is interesting that all the portfolios tilt positively, although by small amounts, into growth stocks even though growth stocks have negative CAPM alphas. This occurs because of the correlations among the asset classes.

All the versions of the optimal orthogonal portfolio suggest strong tilts into government bonds. The bond tilt is the most extreme for the fixed-weight solution, at 141%, and is relatively modest, at 43.2%, for the portfolio assuming a time-varying conditional risk-free rate. This makes sense, as that portfolio can hold short-term Treasuries in addition to government bonds. The large weights in bonds reflect various features of the data and the value of the zero beta rate. With larger values for the zero beta rate, provided that the rate remains below the mean of the global minimum variance portfolio of risky assets, the optimal orthogonal portfolios become more aggressive as the target expected return of the efficient-with-respect-to-*Z* portfolio increases.

The squared Sharpe ratios in Panel B of Table 1 indicate how far the stock market index is from efficiency. The squared Sharpe ratio for the market is 0.189, measured relative to the fixed zero beta rate of 3.8%. The market portfolio's squared Sharpe ratio is slightly smaller, at 0.182, for returns measured in excess of a time-varying risk-free rate. This is because the negative covariance between the risk-free rate and stock returns increases the variance of the excess return. For the fixed-weight orthogonal portfolio R_c the squared Sharpe ratio is 0.079 and for the portfolios using conditioning information, it varies between 0.088 and 0.909.

According to the *Law of Conservation of Squared Sharpe ratios* in Proposition 4, the sum of the index and optimal orthogonal portfolios' squared Sharpe ratios is the squared slope of the tangency from the given zero beta rate to the relevant mean variance frontier. The sum is 0.267

when no conditioning information is used, and is 0.304-1.091 when the information is used. With a fixed risk-free rate we condition only on the lagged dividend yield, which is a relatively weak predictor for stock returns (see Ferson, Sarkissian and Simin, 2003). However, in the case with no risk-free rate the portfolio is not allowed to hold short-term Treasuries, which substantially weakens the performance. In the time-varying risk-free rate case the portfolio strategy is allowed to "market time" by holding short-term Treasuries. In the other two cases we use the information in the lagged Treasury rate, which is a relatively strong predictor, and the efficient-with-respect to *Z* boundary is far above the mean variance boundary that ignores the conditioning information. (Ferson and Siegel, 2009, present an analysis of the statistical significance of differences like these.)

Table 1 suggests that the portfolio weights of the fixed-weight R_c portfolio are extreme and would not be realistic in practical portfolio management settings. This reflects well-known issues with mean classical variance optimal solutions in practice (e.g. Michaud, 1989, and Siegel and Woodgate, 2007). To obtain practical portfolio weights in the classical mean variance problem it is generally necessary to constrain the weights (e.g. Frost and Savarino, 1988) or shrink them towards a benchmark (e.g. Jorion (2003) or Jagannathan and Ma, 2003). In this context note that the optimal orthogonal portfolios using *Z* take less extreme positions on average than the fixed weight solution, yet still are able to generate larger information ratios.

Figures 1 and 2 present time-series plots of the weights for the optimal orthogonal portfolios using *Z*. Like the fixed-weight R_c case, the weights in the case with a fixed risk-free rate assumption appear too volatile and noisy to be of practical interest, likely reflecting the poor predictive ability based solely on the dividend yield. We do not plot them here to save space. The weights in the other two examples are shown in Figures 1 and 2. They generally vary relatively smoothly over time, suggesting that they would not involve prohibitive trading costs in practice.

Figure 1 depicts the weights for the optimal orthogonal portfolio in the case with no risk-free asset. The market index weights are negative through much of the sample and near −25%, indicating that the overall efficient-with-respect to *Z* portfolio keeps about 71% of its money in the market index for much of the sample, computed as (0.23)(−0.25) + 0.77 using the estimated weights for the no risk-free rate case, $x_s = 0.23\, x_c + 0.77\, x_p$, as reported earlier. Starting in the mid-1990s the market weights decrease to around −50% for the optimal orthogonal portfolio weight and 66% for the overall efficient-with-respect to *Z* portfolio. The weight of this optimal orthogonal portfolio in small stocks is positive for much of the sample, but turns slightly negative in the early 1970s, then positive again in the early 1990s. The strategy shorts value stocks in the 1930s and 1940s, but holds positive positions through most of the rest of the sample. The government bond gets the largest weight, starting near 80% and growing sharply in the 1990s to near 100% in the latter parts of the sample.

Figure 2 depicts the weights for the optimal orthogonal portfolio in the case with a time-varying, conditional risk-free rate. This strategy keeps positive weights in the market index until 1999, then it holds small short positions for most of the rest of the sample. The weight in small stocks is positive for much of the sample, turning slightly negative in the early 1970s and then positive again in the early 1990s. The strategy holds value stocks long until the early 2000's and shorts growth stocks during much of the 1980s. The government bond again gets the largest weight, starting at 32% and growing to almost 70% at the end of the sample.

6. Conclusions

This paper derives, characterizes and illustrates optimal orthogonal portfolios in the presence of conditioning information in the form of a set of lagged instruments. Optimal orthogonal portfolios combine with a benchmark portfolio to form mean variance efficient portfolios. We generalize previously published solutions for optimal orthogonal portfolios with information to include the case of a time-varying, but conditionally known, risk-free asset return.

In a portfolio management context, it is reasonable to assume that the portfolio manager has more information about asset returns than the client. The manager observes information about future returns, and by conditioning on this information can expand the investor's opportunity set. Starting from a given benchmark portfolio the optimal orthogonal portfolio with conditioning information describes the active portfolio bets that, when combined with the benchmark, deliver mean variance optimal performance from the uninformed client's perspective.

We present empirical examples using a broad stock market index as the benchmark and portfolios featuring small capitalization, value and growth stocks and a long-term US government bond return. We examine three versions of the optimal orthogonal portfolio with (1) a fixed risk-free rate, (2) no risk-free rate and (3), a time-varying conditional risk-free rate. The optimal orthogonal portfolios with conditioning information have larger information ratios than the orthogonal portfolio that does not use the conditioning information. At the same time, they takes less extreme positions than the fixed weight solution.

From an asset pricing perspective, a standard stock market index is far from efficient when portfolios trade based on lagged interest rates and dividend yields. From an active portfolio management perspective, the example shows that a strong tilt toward bonds improves the efficiency of equity portfolios. Our results should be useful in future asset pricing and portfolio management applications.

Appendix

Efficient Portfolio Solutions

Portfolio weights for efficient portfolios in the presence of conditioning information are derived by Ferson and Siegel (2001). They consider the case with no risk-free asset and the case with a fixed risk-free asset whose return is constant over time. In Theorem 1 we generalize to consider the case with a risk-free asset whose return is known at the beginning of the period, and thus is included in the information *Z*, and may vary over time. We then, in Theorem 2, reproduce the case with no risk-free asset from Ferson and Siegel (2001) for future reference.

Consider *N* risky assets with returns *R*. In $N\times 1$ column-vector notation, we have

$$R = \mu(Z) + \varepsilon . \tag{17}$$

The noise term ε is assumed to have conditional mean zero given *Z* and nonsingular conditional covariance matrix $\Sigma_\varepsilon(Z)$. The conditional expected return vector is $\mu(Z) = E(R\,|\,Z)$. Let the $1\times N$ row vector $x'(Z) = (x_1(Z),\ldots,x_N(Z))$ denote the portfolio share invested in each of the *N* risky assets, investing (or borrowing) at the risk-free rate the amount $1 - x'(Z)\underline{1}$, where $\underline{1} \equiv (1,\ldots,1)'$ denotes the column vector of ones. We allow for a conditional risk-free asset returning $R_f = R_f(Z)$. The return on the portfolio is $R_s = R_f + x'(Z)\left(R - R_f\underline{1}\right)$, with unconditional expectation and variance as follows:

$$\mu_s = E\left(R_f\right) + E\left\{x'(Z)\left[\mu(Z) - R_f\underline{1}\right]\right\}, \tag{18}$$

$$\sigma_s^2 = E\left(R_s^2\right) - \mu_s^2 = E\left[E\left(R_s^2\,|\,Z\right)\right] - \mu_s^2 , \tag{19}$$

$$\sigma_s^2 = E\left(R_f^2\right) + E\left[x'(Z)Q^{-1}x(Z)\right] + 2E\left\{R_f x'(Z)\left[\mu(Z) - R_f\underline{1}\right]\right\} - \mu_s^2 , \tag{20}$$

where we have defined the $N\times N$ matrix

$$Q = Q(Z) \equiv \left\{ E\left[\left(R - R_f \underline{1}\right)\left(R - R_f \underline{1}\right)' \middle| Z \right] \right\}^{-1} = \left\{ \left[\mu(Z) - R_f \underline{1}\right]\left[\mu(Z) - R_f \underline{1}\right]' + \Sigma_\varepsilon(Z) \right\}^{-1}. \quad (21)$$

Also, define the constants:

$$\zeta \equiv E\left\{ \left[\mu(Z) - R_f \underline{1}\right]' Q \left[\mu(Z) - R_f \underline{1}\right] \right\} \quad (22)$$

$$\varphi \equiv E\left\{ R_f \left[\mu(Z) - R_f \underline{1}\right]' Q \left[\mu(Z) - R_f \underline{1}\right] \right\}, \quad \text{and} \quad (23)$$

$$\psi \equiv E\left\{ R_f^2 \left[\mu(Z) - R_f \underline{1}\right]' Q \left[\mu(Z) - R_f \underline{1}\right] \right\}. \quad (24)$$

<u>Theorem 1:</u> Given a target unconditional expected return μ_s, N risky assets, instruments Z, and a conditional risk-free asset with rate $R_f = R_f(Z)$ that may vary over time, the unique portfolio having minimum unconditional variance is determined by the weights:

$$\begin{aligned} x_s(Z) &= \left(\frac{\mu_s - E(R_f) + \varphi}{\zeta} - R_f \right) Q\left[\mu(Z) - R_f \underline{1}\right] \\ &= \left[(c+1)\mu_s + b - R_f \right] Q\left[\mu(Z) - R_f \underline{1}\right] \end{aligned}, \quad (25)$$

and the optimal portfolio variance is:

$$\sigma_s^2 = a + 2b\mu_s + c\mu_s^2 \quad (26)$$

where $a = E\left(R_f^2\right) + \frac{\left[E(R_f) - \varphi\right]^2}{\zeta} - \psi$, $b = \frac{\varphi - E(R_f)}{\zeta}$, and $c = \frac{1}{\zeta} - 1$. When the risk-free asset return is constant, then these formulas simplify to Theorem 2 of Ferson and Siegel (2001) with

$$x_s(Z) = \frac{\mu_s - R_f}{\zeta} Q\left[\mu(Z) - R_f \underline{1}\right], \tag{27}$$

and with optimal portfolio variance $\sigma_s^2 = \frac{1-\zeta}{\zeta}\left(\mu_s - R_f\right)^2$.

<u>Proof:</u> Our objective is to minimize, over the choice of $x_s(Z)$, the portfolio variance $Var(R_s)$ subject to $E(R_s) = \mu_s$, where $R_s = R_f + x'(Z)\left(R - R_f \underline{1}\right)$ and the variance is given by Equation (20). We form the Lagrangian:

$$\begin{aligned} L\left[x(Z)\right] = E\left[x'(Z)Q^{-1}x(Z)\right] + 2E\left\{R_f x'(Z)\left[\mu(Z) - R_f \underline{1}\right]\right\} \\ +2\lambda E\left\{\mu_s - R_f - x'(Z)\left[\mu(Z) - R_f \underline{1}\right]\right\} \end{aligned} \tag{28}$$

and proceed using a perturbation argument. Let $q(Z) = x(Z) + dy(Z)$, where $x(Z)$ is the conjectured optimal solution, $y(Z)$ is any regular function of Z and d is a scalar. Optimality of $x(Z)$ follows when the partial derivative of $L\left[q(Z)\right]$ with respect to d is identically zero when evaluated at $d = 0$. Thus,

$$0 = E\left(y'(Z)\left\{Q^{-1}x(Z) + \left(R_f - \lambda\right)\left[\mu(Z) - R_f \underline{1}\right]\right\}\right) \tag{29}$$

for all functions $y(Z)$, which implies that $Q^{-1}x(Z) + \left(R_f - \lambda\right)\left[\mu(Z) - R_f \underline{1}\right] = 0$ almost surely in Z. Solve this expression for $x(Z)$ to obtain Equation (25), where the Lagrange multiplier λ is evaluated by solving for the target mean, μ_s. The expression for the optimal portfolio variance follows by substituting the optimal weight function into Equation (20). Formulas for fixed R_f then follow directly. QED.

When the risk-free asset's return is time varying and contained in the information set Z at the beginning of the portfolio formation period, the *conditional* mean variance efficient boundary varies over time with the value of $R_f(Z)$ along with the conditional asset means and covariances. In this case a zero-beta parameter, γ_0, may be chosen to fix a point on the *unconditionally* efficient-with-respect to *Z* boundary. The choice of the zero beta parameter corresponds to the choice of a target unconditional expected return μ_s. For a given value of γ_0, the target mean maximizes the squared Sharpe ratio $(\mu_s - \gamma_0)^2/\sigma_s^2$ along the mean variance boundary described by Equation (26), which implies $\mu_s = -(a + b\gamma_0)/(b + c\gamma_0)$.

When there is a risk-free asset that is constant over time, the unconditionally efficient-with-respect to *Z* boundary is linear (a degenerate hyperbola) and reaches the risk-free asset at zero risk. In this case we use $\gamma_0 = R_f$ and can obtain any μ_s larger or smaller than R_f, levering the efficient portfolio up or down with positions in the risk-free asset.

When there is no risk-free asset, we define portfolio *s* by letting $x' = x'(Z) = [x_1(Z),...,x_N(Z)]$ denote the shares invested in each of the *N* risky assets, with the constraint that the weights sum to 1.0 almost surely in Z. The return on this portfolio, $R_s = x'(Z)R$, has expectation and variance as follows:

$$\mu_s = E\left[x'(Z)\mu(Z)\right], \tag{30}$$

$$\sigma_s^2 = E\left\{x'(Z)\Lambda^{-1}x(Z)\right\} - \mu_s^2, \tag{31}$$

where we have defined the $N\times N$ matrix

$$\Lambda = \Lambda(Z) \equiv \left\{E\left[RR'|Z\right]\right\}^{-1} = \left[\mu(Z)\mu'(Z) + \Sigma_\varepsilon(Z)\right]^{-1}. \tag{32}$$

Also, define the constants:

$$\delta_1 = E\left(\frac{1}{\underset{-}{1}'\Lambda\underset{-}{1}}\right), \tag{33}$$

$$\delta_2 = E\left(\frac{\underset{-}{1}'\Lambda\mu(Z)}{\underset{-}{1}'\Lambda\underset{-}{1}}\right), \quad \text{and} \tag{34}$$

$$\delta_3 = E\left[\mu'(Z)\left(\Lambda - \frac{\Lambda\underset{-}{1}\underset{-}{1}'\Lambda}{\underset{-}{1}'\Lambda\underset{-}{1}}\right)\mu(Z)\right]. \tag{35}$$

<u>Theorem 2:</u> (Ferson and Siegel, 2001, Theorem 3) Given *N* risky assets and no risk-free asset, the unique portfolio having minimum unconditional variance and unconditional expected return μ_s, is determined by the weights:

$$\begin{aligned} x_s'(Z) &= \frac{\underset{-}{1}'\Lambda}{\underset{-}{1}'\Lambda\underset{-}{1}} + \frac{\mu_s - \delta_2}{\delta_3}\mu'(Z)\left(\Lambda - \frac{\Lambda\underset{-}{1}\underset{-}{1}'\Lambda}{\underset{-}{1}'\Lambda\underset{-}{1}}\right) \\ &= \frac{\underset{-}{1}'\Lambda}{\underset{-}{1}'\Lambda\underset{-}{1}} + \left[(c+1)\mu_s + b\right]\mu'(Z)\left(\Lambda - \frac{\Lambda\underset{-}{1}\underset{-}{1}'\Lambda}{\underset{-}{1}'\Lambda\underset{-}{1}}\right) \end{aligned}, \tag{36}$$

and the optimal portfolio variance is:

$$\sigma_s^2 = a + 2b\mu_s + c\mu_s^2, \tag{37}$$

where $a = \delta_1 + \delta_2^2/\delta_3$, $b = -\delta_2/\delta_3$, and $c = (1-\delta_3)/\delta_3$.

The efficient-with-respect to *Z* boundary is formed by varying the value of the target mean return μ_s in (36). Note that the second term on the right-hand side of Equation (36) is proportional to the vector of weights of an excess return, or zero net investment portfolio (post multiplying that term by a vector of ones implies that the weights sum to zero). The first term in (36) is the weight of the global minimum conditional second moment portfolio. Thus, Equation

(36) illustrates two-fund separation: Any efficient-with-respect to *Z* portfolio can be found as a combination of the global minimum conditional second moment portfolio and some weight on the unconditionally efficient excess return described by the second term.

Proofs of Propositions:

Proof of Propositions 2 and 3:

We begin with Proposition 3. To see that (15) and (16) are equivalent, use the fact that efficiency of R_s implies that $\mu_p = \gamma_0 + \frac{\sigma_{ps}}{\sigma_s^2}(\mu_s - \gamma_0)$ and thus $\frac{\mu_s - \gamma_0}{\sigma_s^2} = \frac{\mu_p - \gamma_0}{\sigma_{ps}}$. Next, given any portfolio $R_q \neq R_c$, we will show that $\alpha_q^2 / \sigma_q^2 < \alpha_c^2 / \sigma_c^2$. Beginning with Equation (16), we compute:

$$\sigma_{cs} = \frac{\sigma_p^2 \sigma_s^2 - \sigma_{ps}^2}{\sigma_p^2 - \sigma_{ps}}, \text{ and} \tag{38}$$

$$\sigma_c^2 = \frac{\sigma_p^2 \left(\sigma_p^2 \sigma_s^2 - \sigma_{ps}^2\right)}{\left(\sigma_p^2 - \sigma_{ps}\right)^2}, \tag{39}$$

The efficiency of R_s implies that $\mu_c = \gamma_0 + \frac{\sigma_{cs}}{\sigma_s^2}(\mu_s - \gamma_0)$, $\mu_p = \gamma_0 + \frac{\sigma_{ps}}{\sigma_s^2}(\mu_s - \gamma_0)$ and $\mu_q = \gamma_0 + \frac{\sigma_{qs}}{\sigma_s^2}(\mu_s - \gamma_0)$. Substituting these expressions and using the fact that $\sigma_{cp} = 0$, which follows from (16), we compute:

$$\frac{\alpha_c^2}{\sigma_c^2}-\frac{\alpha_q^2}{\sigma_q^2}=\frac{\left(\mu_c-[\gamma_o+(\mu_p-\gamma_o)\sigma_{cp}/\sigma_p{}^2]\right)^2}{\sigma_c^2}-\frac{\left(\mu_q-[\gamma_o+(\mu_p-\gamma_o)\sigma_{qp}/\sigma_p{}^2]\right)^2}{\sigma_q^2}$$
$$=\left(\mu_s-\gamma_0\right)^2\left(\frac{\sigma_{cs}^2}{\sigma_s^4\sigma_c^2}-\frac{\left(\sigma_{qs}\sigma_p^2-\sigma_{ps}\sigma_{qp}\right)^2}{\sigma_s^4\sigma_p^4\sigma_q^2}\right)$$
$$=\left(\mu_s-\gamma_0\right)^2\left(\frac{\sigma_p^2\sigma_s^2-\sigma_{ps}^2}{\sigma_s^4\sigma_p^2}-\frac{\left(\sigma_{qs}^2\sigma_p^4+\sigma_{ps}^2\sigma_{qp}^2-2\sigma_{qs}\sigma_p^2\sigma_{ps}\sigma_{qp}\right)}{\sigma_s^4\sigma_p^4\sigma_q^2}\right) \tag{40}$$
$$=\frac{\left(\mu_s-\gamma_0\right)^2}{\sigma_s^4\sigma_p^4\sigma_q^2}\left(\sigma_p^4\sigma_s^2\sigma_q^2-\sigma_p^2\sigma_q^2\sigma_{ps}^2-\sigma_{qs}^2\sigma_p^4-\sigma_{ps}^2\sigma_{qp}^2+2\sigma_p^2\sigma_{qs}\sigma_{ps}\sigma_{qp}\right)$$

We now use the fact that $\sigma_{sp}^2<\sigma_s^2\sigma_p^2$ (because, by assumption, R_p and R_s are not perfectly correlated) to see that:

$$\frac{\alpha_c^2}{\sigma_c^2}-\frac{\alpha_q^2}{\sigma_q^2}\geq\frac{\left(\mu_s-\gamma_0\right)^2}{\sigma_p^4\sigma_q^2\sigma_s^4}\left[\sigma_p^4\sigma_q^2\sigma_s^2-\sigma_p^2\sigma_q^2\sigma_{sp}^2-\sigma_p^4\sigma_{qs}^2-\sigma_{qp}^2\left(\sigma_s^2\sigma_p^2\right)+2\sigma_p^2\sigma_{qs}\sigma_{qp}\sigma_{ps}\right]$$
$$=\frac{\left(\mu_s-\gamma_0\right)^2}{\sigma_p^2\sigma_q^2\sigma_s^4}\left(\sigma_p^2\sigma_q^2\sigma_s^2-\sigma_q^2\sigma_{sp}^2-\sigma_p^2\sigma_{qs}^2-\sigma_{qp}^2\sigma_s^2+2\sigma_{qs}\sigma_{qp}\sigma_{ps}\right)\geq 0 \tag{41}$$

where the final inequality follows from recognizing that the variance-covariance terms in parentheses are equal to the determinant of the (necessarily non-negative definite) covariance matrix of $\left(R_p,R_q,R_s\right)$. This establishes the maximal property of R_c.

To show uniqueness, note further that the inequality will be strict (and we will have $\alpha_c^2/\sigma_c^2-\alpha_q^2/\sigma_q^2>0$) unless we have both of the following conditions corresponding to the two inequalities in the final calculation: (1) $\sigma_{qp}=0$ so that R_q and R_p are orthogonal and (2) the covariance matrix of $\left(R_p,R_q,R_s\right)$ is singular so that R_q is a linear combination of R_p and R_s. However, there is only one portfolio orthogonal to R_p that can be formed as a linear combination

$\lambda R_p + (1-\lambda) R_s$, and this solution is R_c. This establishes Proposition 3, which holds in the case of both Theorem 1 and Theorem 2; that is, whether or not there is a conditionally risk-free asset.

The expressions in Proposition 2 for the optimal weights, $x_c(Z)$, follow from substituting portfolio weights from Theorem 1 (if there exists a conditional risk-free asset that may be time varying) or Theorem 2 (otherwise) into Equation (15) and noting that the constants A and B represent the combining portfolio weights implied by (15) as $R_c = AR_s + BR_p$. Substituting, we see that Equation (15) implies that the portfolio weight function $Ax_s(Z) + Bx_p(Z)$ generates returns R_c, completing the proof of Proposition 2. QED.

Proof of Proposition 4:

Let R_s denote the efficient-with-respect to *Z* portfolio corresponding to zero-beta rate γ_0. The efficiency of R_s implies that we may substitute $\mu_p - \gamma_0 = (\mu_s - \gamma_0)\frac{\sigma_{ps}}{\sigma_s^2}$ and $\mu_c - \gamma_0 = (\mu_s - \gamma_0)\frac{\sigma_{cs}}{\sigma_s^2}$ to find

$$S_p^2 + S_c^2 = \frac{(\mu_p - \gamma_0)^2}{\sigma_p^2} + \frac{(\mu_c - \gamma_0)^2}{\sigma_c^2} = \frac{(\mu_s - \gamma_0)^2}{\sigma_s^4}\left(\frac{\sigma_{ps}^2}{\sigma_p^2} + \frac{\sigma_{cs}^2}{\sigma_c^2}\right). \tag{42}$$

Next, substituting Equations (38) and (39) for σ_{cs} and σ_c^2 we verify that this expression reduces to S_s^2. QED.

Methodology Appendix

We estimate the conditional mean functions, $\mu(Z)$, by ordinary least squares regressions of the returns on the lagged values of the conditioning variables. On the assumption that the conditional mean returns are linear functions of *Z*, these are the optimal generalized method of moments (GMM, see Hansen 1982) estimators. The covariance matrix of the residuals is used as the estimate of $\Sigma_\varepsilon(Z)$, which is assumed to be constant. These are the maximum likelihood estimates (MLE) under joint normality of (R,Z). In general the conditional covariance matrix of the returns given *Z* will be time-varying as a function of *Z*, as in conditional heteroskedasticity. Ferson and Siegel (2003) model conditional heteroskedasticity in alternative ways and find using parametric bootstrap simulations that this increases the tendency of the efficient-with-respect to *Z* portfolio weights to behave conservatively in the face of extreme realizations of *Z*.

The optimal orthogonal portfolio weights in Table 1 and Figure 1 are estimated from Equations (7) and (8) in the text where, in the time-varying risk-free rate case, the Treasury bill return is assumed to be conditionally risk-free in (8). The benchmark portfolio x_p is a vector with a 1.0 in the place of the Market index and zeros elsewhere. The matrix *Q* is estimated by using the MLE estimates of $\mu(Z)$ and $\Sigma_\varepsilon(Z)$ in the function given by Equation (13). The parameter μ_p is estimated as the sample excess return on the Market index, and γ_0 is the sample mean of the Treasury return, 3.8%.

The Parametric Bootstrap

The parametric bootstrap is a special case of the simple, or nonparametric bootstrap, itself an example of a resampling scheme. Introduced by Efron (1979), the bootstrap is useful when we wish to conduct statistical inferences but when we either don't have an analytical formula for the sampling variation of a statistic, don't wish to assume normality or some other convenient distribution that allows for an analytical formula or have a sample too small to trust asymptotic distribution theory. The basic idea is to build a sampling distribution by resampling from the data at hand. In the simplest example we have some statistic that we have estimated from a sample and we want to know its sampling distribution. We resample from the original data, randomly with replacement, to generate an artificial sample of the same size, and we compute the statistic on the artificial sample. Repeating this many times, the histogram of the statistics computed on the artificial samples is an estimate of the sampling distribution for the original statistic. This distribution can be used to estimate standard errors, confidence intervals, etc. We can think of the bootstrap samples as being related to the original sample as the original sample is to the population. There are many variations on the bootstrap, and a good overview is provided by Efron and Tibshirani (1993).

In the simple, or nonparametric bootstrap no assumptions are made about the form of the distribution. It is assumed, however, that the sample accurately reflects the underlying population distribution, and this is critical for reliable inferences. For example, suppose that the true distribution was a uniform on $[0, M]$. In a sample drawn from this distribution, the maximum value is likely to be smaller than M, so that the bootstrap will likely understate the true variability of the data. This problem is obviously worse if the original sample has fewer observations. If the data are contaminated with measurement errors, in contrast, the extent of the true variability can be overstated. Even with large sample sizes the bootstrap can be unreliable. For example, if the true distribution has infinite variance, the bootstrap distribution for the sample mean is inconsistent (Athreya, 1987).

With a parametric bootstrap, we can sometimes do better than with a nonparametric bootstrap, where "do better" means for example, obtain more accurate confidence intervals (e.g. Andrews, 2001). The idea of the parametric bootstrap is to use some of the parametric structure of the data. This might be as simple as assuming the form of the probability distribution. For example, assuming that the data are independent and normally distributed, we can generate artificial samples from a normal distribution using the sample mean and variance as the parameters. This is not exactly the right thing to do, because we should be sampling from a population with the true parameter values, not their estimated values. But, if the estimates of the mean and variance are good enough, we should be able to obtain reliable inferences.

To illustrate and further suggest the flexibility of the parametric bootstrap, consider an example, similar to the setting in our paper, where we have a regression of stock returns on a vector of lagged instruments, Z, which are highly persistent over time. Obviously, sampling from the Z's randomly with replacement would destroy their strong time-series dependence. Time-series dependence can be accommodated in a nonparametric way by using a block bootstrap. Here, we sample randomly a block of consecutive observations, where the block length is set to capture the extent of memory in the data.

In order to capture the time series dependence of the lagged Z in a parametric bootstrap, we can model the lagged instruments as vector AR(1), for example, retaining the estimator of the

AR(1) coefficient and the model residual, which we call the shocks, U_z. Regressing the future stock returns on the lagged instruments, we retain the regression coefficient and the residuals, which we call the shocks, U_r. We generate a sample of artificial data, with the same length as the original sample, as follows. We concatenate the shocks as $v = (U_z, U_r)$. Resampling rows from *v*, randomly with replacement, retains the covariances between the stock return shocks and the instrument Z's shocks. This can be important for capturing features like the lagged stochastic regressor bias described by Stambaugh (1999). Drawing an initial row from *v*, we take the U_z shock and add it to the initial value of the Z (perhaps, drawn from the unconditional sample distribution) to produce the first lagged instrument vector, Z_{t-1}. We draw another row from *v* and construct the first observation of the remaining data as follows. The stocks' returns are formed by adding the U_r shock to βz_{t-1}, where β is the "true" regression coefficient that defines the conditional expected return. The contemporaneous values of the Zs are formed by multiplying the VAR coefficient by Z_{t-1} and adding the shock, U_z. The next observation is generated by taking the previous contemporaneous value of the Z as z_{t-1} and repeating the process. In this way, the Z values are built up recursively, which captures their strong serial correlation.

*Wayne E. Ferson is the Ivadelle and Theodore Johnson Chair in Banking and Finance at the Marshall School, University of Southern California, 3670 Trousdale Parkway, Suite 308, Los Angeles, CA 90089-0804, [phone (213) 740-5615, email ferson@marshall.usc.edu], and a Research Associate, National Bureau of Economic Research. Andrew F. Siegel is the Grant I. Butterbaugh Professor of Finance and Business Economics, also of Information Systems and Operations Management, and is Adjunct Professor of Statistics, Foster School of Business, University of Washington, Box 353226, Seattle, WA. 98195-3226 [phone: (206)543-4476, fax: 685-9392, e-mail asiegel@uw.edu].

References

Abhyankar, Abhay, Devraj Basu and Alexander Streme, 2007, Portfolio efficiency and discount factor bounds with conditioning information: An empirical study, working paper, Warwick Business School.

Andrews, D.W.K, 2001, Higher order improvements of the parametric bootstrap for markov processes, working paper, Yale University.

Athreya, K.B., 1987, Bootstrap of the mean in the infinite variance case, *Annals of Statistics* 15, 724-731.

Bekaert, Geert, and Jun Liu, 2004, Conditioning information and variance bounds on pricing kernels. *Review of Financial Studies* 17, 339-378.

Campbell, John, Lo, Andrew and AC MacKinlay, 1997, The econometrics of financial markets, Princeton University Press, Princeton NJ.

Chiang, Ethan, 2009, Modern portfolio management with conditioning information, working paper, UNC Charlotte.

Dybvig, Philip and Stephen Ross, 1985, Differential information and performance measurement using a security market line, *Journal of Finance* 40, 383-399.

Efron, Bradley, 1979, Bootstrap methods: Another looks at the jackknife, *Annals of Statistics* 7, 1-26.

Efron, Bradley and R.J. Tibshirani, 1993, *Introduction to the bootstrap*, New York, Chapman and Hall.

Fama, Eugene F., 1973, A note on the market model and the two-parameter model, *Journal of Finance* 28, 1181-1185.

Ferson, Wayne E., Sergei Sarkissian and Timothy Simin, 2003, Spurious regressions in financial economics? *Journal of Finance* 58, 1393-1414.

Ferson, Wayne E. and Andrew F. Siegel, 2001, The efficient use of conditioning information in portfolios. *Journal of Finance* 56, 967-982.

Ferson, Wayne E. and Andrew F. Siegel, 2003, Stochastic discount factor bounds with conditioning information, *Review of Financial Studies* 16, 567-595.

Ferson, Wayne E. and Andrew F. Siegel, 2009, Testing portfolio efficiency with conditioning information, *Review of Financial Studies* 22, 2735-2758*.*

Ferson, Wayne E., Andrew F. Siegel and Pisun Xu, 2006, Mimicking portfolios with conditioning information, *Journal of Financial and Quantitative Analysis* 41, 607-635.

Frost, Peter A. and James E. Savarino, 1988, For better performance: Constrain portfolio weights. *Journal of Portfolio Management* 15, 29-34.

Gibbons, Michael R., Stephen A. Ross and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121-1152.

Grinold, Richard C. and Ronald N. Kahn, 1992, Active portfolio management, 2nd Edition, McGraw Hill, NY, NY.

Hansen, Lars P., and Scott F. Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica* 55, No. 3, 587-613.

Jagannathan, Ravi and Tongshu Ma, 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *Journal of Finance* 58, 1651-1684.

Jobson, J. D., and Bob Korkie, 1982, Potential performance and tests of portfolio efficiency, *Journal of Financial Economics* 10, 433-466.

Jorion, Philippe, 1992, Portfolio optimization in practice, *Financial Analysts Journal* 48, 68-74.

Jorion, Philippe, 2003, Portfolio optimization with tracking error constraints, *Financial Analysts Journal* 59, 70-82.

Kandel, Shmuel A., 1984, The likelihood ratio test statistic of mean variance efficiency without a riskless asset, *Journal of Financial Economics* 13, 575-592.

Kandel, Shmuel A. and Robert F. Stambaugh, 1987, On correlations and inferences about mean-variance efficiency, *Journal of Financial Economics* 18, 61-90.

Kandel, Shmuel A. and Robert F. Stambaugh, 1989, A mean-variance framework for tests of asset pricing models, *Review of Financial Studies* 2, 125-156.

MacKinlay, A. Craig, 1987, On multivariate tests of the CAPM. *Journal of Financial Economics* 18, 341-371.

MacKinlay, A. Craig, 1995, Multifactor models do not explain deviations from the CAPM, *Journal of Financial Economics* 38, 3-28.

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

Michaud, Richard O., 1989, The Markowitz optimization enigma: Is ‘optimized’ optimal? *Financial Analysts Journal* 45, 31-42.

Roll, Richard R., 1985, A note on the geometry of Shanken's CSR T^2 test for mean/variance efficiency, *Journal of Financial Economics* 14, 349-357.

Roll, Richard R., 1980, Orthogonal portfolios, *Journal of Financial and Quantitative Analysis* 15, 1005-1023.

Roll, Richard R, 1982, A mean variance analysis of tracking error, *Journal of Portfolio Management* 18, 13-22.

Roll, Richard R., 1977, A critique of the asset pricing theory's tests - part 1: On past and potential testability of the theory, *Journal of Financial Economics* 4, 129-176.

Shanken, Jay, 1987, Multivariate proxies and asset pricing relations: Living with the Roll critique, *Journal of Financial Economics* 18, 91-110.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425-442.

Siegel, Andrew F., and Artemiza Woodgate, 2007, Performance of portfolios optimized with estimation error, *Management Science* **53**, 1005-1015.

Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375-421.

Treynor, J.L. and Fisher Black, 1973, How to use security analysis to improve portfolio selection, *Journal of Business* **46**, 66-86.

Table 1.

Annual returns on portfolios of common stocks and long term US government bonds cover the 1931-2007 period. The CAPM α refers to the intercept in the regression Equation (1), of the portfolio return in excess of a zero beta parameter, on that of the broad value-weighted stock market index (Market Index). The symbol σ_u refers to the standard deviation of the regression residuals. The average active portfolio weights are the weights of the most mispriced or optimal orthogonal portfolios, shown under various assumptions about the risk-free rate, R_f. The fixed portfolio, R_c, ignores the conditioning information. The alphas and residual standard deviations are annual percentage units. The information ratio is the ratio α / σ_u.

Panel A: Optimal Orthogonal Portfolios

Asset	CAPM α	σ_u	Information Ratio	fixed R_c	Average Active Portfolio weights Avg $\left[X_c(Z)\right]$ Fixed R_f	No R_f	Varying R_f
Market Index	0	0	0.000	-1.260	0.179	-0.210	0.308
Small Stocks	3.95	25.5	0.155	0.353	-0.262	0.133	0.097
Value Stocks	3.14	17.8	0.177	0.345	0.137	0.079	0.142
Growth Stocks	-1.00	8.3	-0.121	0.155	0.028	0.098	0.020
Bonds	1.66	9.1	0.183	1.410	0.918	0.900	0.432
Fixed R_c	4.66	16.5	0.282	1.000			
$X_c(Z)$ Portfolios:							
Fixed R_f	8.66	25.5	0.340				
No R_f	3.12	10.5	0.297				
Varying R_f	8.77	9.2	0.955				

Panel B: Squared Sharpe Ratios

Case	$S^2(R_p)$	$S^2(R_c)$	Sum
Fixed R_c	0.189	0.079	0.267
$X_c(Z)$ Portfolios:			
Fixed R_f	0.189	0.115	0.304
No R_f	0.189	0.088	0.277
Varying R_f	0.182	0.909	1.191

Figure 1: Optimal Orthogonal Portfolio Weights: The No Risk-free Rate Case

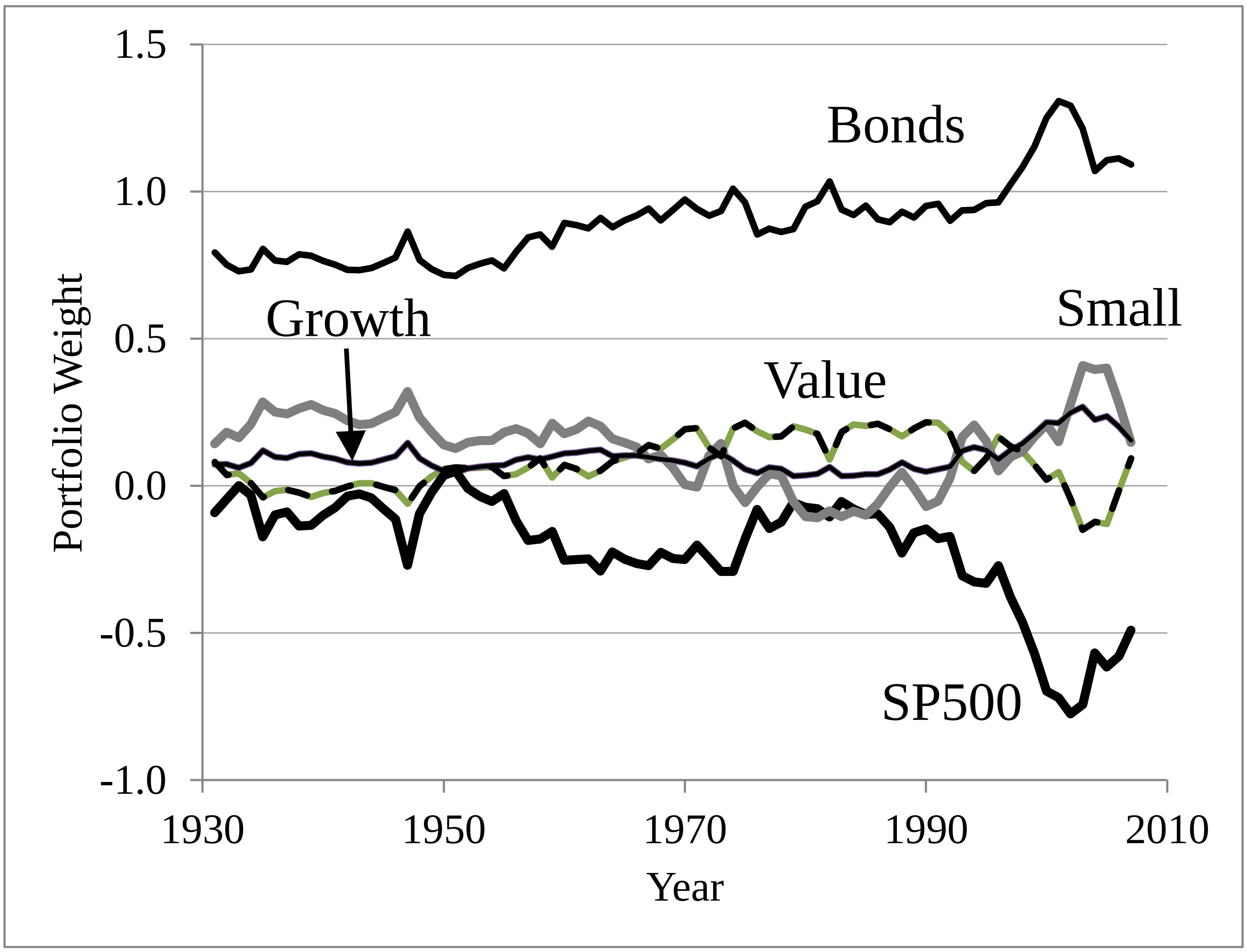

Figure 2: Optimal Orthogonal Portfolio Weights: The Time-Varying Risk-Free Rate Case