Market Proxies, Correlation, and Relative Mean-Variance Efficiency:
Still Living with the Roll Critique[1]

Todd Prono[2]
Federal Reserve Bank of Boston

First Version: April 2006
This Version: March 2009

Abstract

A pricing restriction is developed to test the validity of the CAPM conditional on a prior belief about the correlation between the true market return and the proxy return used in the test. Distinguishing this pricing restriction from competing tests also based upon the relative efficiency of the proxy return is a consideration for the proxy's mismeasurement of the market return. Failure to account for this mismeasurement biases tests of the CAPM towards rejection by overstating the inefficiency of the proxy. A time-varying version of this pricing restriction links mismeasurement of the market return to time-variation in beta.

Keywords: Asset pricing, CAPM, portfolio efficiency, multivariate testing, bootstrap hypothesis testing, triangular systems, endogeneity, identification, GMM, conditional heteroskedasticity, GARCH. JEL codes: C12, C13, C32, G12.

[1] This paper was formerly entitled "The Relative Efficiency of Endogenous Proxies: Still Living with the Roll Critique." Gratitude is owed to Arthur Lewbel, Zhijie Xiao, Christopher F. Baum, Ethan Cohen-Cole, Evan Sekeris, as well as seminar participants at Acadian Asset Management, the Federal Reserve Bank of Boston, the 2007 Summer Meeting of the Econometric Society, and the 2008 Far East and South East Asia Meeting of the Econometric Society for helpful comments and discussions.

Disclaimer: The views expressed in this paper are solely those of the author and do not reflect official positions of the Federal Reserve Bank of Boston or the Federal Reserve System. All errors are my own.

[2] Corresponding Author: Todd Prono, Federal Reserve Bank of Boston, 600 Atlantic Avenue, Boston, MA 02210. (617) 973-3869, todd.prono@bos.frb.org.

1. Introduction

A common feature among many asset pricing models in financial economics is the relation of expected returns on risky securities to the covariance between those securities' returns and an economic aggregate like (the marginal utility of) aggregate wealth or consumption. In empirical work, this economic aggregate (central to the pricing model under consideration) is generally unobservable and so needs to be proxied for. Tests of the given model which, by necessity, are based on the proxy are confronted by a joint hypothesis that complicates the interpretation of a rejection of the model's prediction. In particular, does this rejection signal a violation of the model's result or the poor quality of the proxy chosen to render the model "testable?" In specific regard to the capital asset pricing model (CAPM), the existence of this dual hypothesis led Roll (1977) to conclude that the "theory is not testable unless the exact composition of the true market portfolio is known and used in the tests" (p. 130). This statement led Gibbons, Ross, and Shanken (1989) to develop a statistic for assessing the mean-variance efficiency of a given proxy return and Shanken (1987), based on this result, to propose an indirect method of testing the CAPM conditional on a prior belief about the correlation between the chosen proxy and the market return.[3] A key premise shared by these two works is that the linear relation between security returns and the proxy return can be described by an OLS projection of the former onto the latter. In general, a proxy return mismeasures the market return. This paper explores the effects of this mismeasurement on tests of the degree to which mean-variance efficiency fails (i.e., tests designed to determine just how inefficient a given proxy is), noting that the minimal (or substantial) degree of this failure is interpreted as evidence for (or against) the CAPM.

Suppose that mismeasurement between the proxy and market return is taken to mean that certain components relevant to the market return are excluded from the proxy. Then, the extent to which these components are correlated with the proxy return will determine the extent to which innovations to a linear equation describing security returns conditional on the proxy return will tend to covary with that proxy return, since those innovations will contain the aforementioned omitted components. In other words, mismeasurement renders the proxy return endogenous in a linear equation relating security returns to the proxy. The bottom line is that the resulting structural equation will tend to differ from a projection equation. Ignoring this difference biases tests of the CAPM that are based on the relative efficiency (in a mean-variance sense) of a proxy return towards rejection not because of any failing in the CAPM theory itself but, rather, because the effects of an inefficient proxy are not taken into account.

A test of the CAPM based on the aforementioned structural equation as opposed to an OLS projection requires a consistent estimator for the former in order to render the test empirically feasible. Towards that end, a novel estimator for linear equations with an endogenous regressor proposed by Prono (2008) is utilized. This estimator relies upon the conditional heteroskedasticity of security returns by basing identification on exclusionary restrictions within the function form describing that heteroskedasticity. As such, this estimator is a higher-moment analog to common instrumental variables techniques.

The remainder of this paper is organized as follows. Starting from a rather general pricing model, Section 2 develops a pricing restriction that can be used to test the CAPM conditional on

[3] MacKinlay and Richardson (1991) present a generalization to the Gibbons et al. (1989) statistic, while Kandel and Stambaugh (1987) represent a closely related work to Shanken (1987).

a prior belief about the correlation between the market return and the proxy used in the test. This restriction fully encompasses the effects of potential differences between the market and proxy return. Section 3 reviews conventional tests of relative mean-variance efficiency. Section 4 presents an overview of the econometrics used to identify and estimate the structural equation describing security returns in terms of the proxy return. Section 5 details a method for conducting a test of relative mean-variance efficiency that is based upon the econometric techniques developed in section 4. Section 6 summarizes the results from employing this test, comparing them to the results obtained from the conventional methods reviewed in section 3. Section 7 presents a generalization of the pricing restriction in section 2 that provides a direct link between mismeasurement of the market return and time-variation in beta. Section 8 concludes.

2. Pricing Restriction

Assume there exists an observable risk-free rate. Let r_t be an N-vector of excess security returns, r_{pt} a scalar proxy to the unobservable excess market return r_{mt}, and let the resulting $N+1$ components be linearly independent. Finally, define m_t as a scalar unobservable economic aggregate. Potential examples of m_t include (the marginal utility of) aggregate wealth or consumption. Consider the following pricing model

$$E\left[r_t\right] = Cov\left[m_t,\ r_t\right] \tag{1}$$

that relates expected excess returns to the covariance between excess returns and the economic aggregate. Many pricing models in financial economics can be characterized in terms of (1). For instance, suppose

$$m_t = \left(\frac{E\left[r_{mt}\right]}{\sigma^2\left[r_{mt}\right]}\right) r_{mt}. \tag{2}$$

Then (1) and (2) imply

$$E\left[r_t\right] = \beta E\left[r_{mt}\right]$$

where

$$\beta = \frac{Cov\left[r_{mt},\ r_t\right]}{\sigma^2\left[r_{mt}\right]}$$

which is the familiar CAPM of Sharpe (1964) and Lintner (1965). Alternatively, replacing r_{mt} in (2) with r_{ct}, the excess return on a security (or portfolio of securities) that is perfectly correlated with changes in aggregate consumption, renders (1) and (2) the CCAPM of Breeden (1979). More generally, if m_t can be decomposed into a set of K orthogonal factors where the ith factor f_{it} is weighted by $\frac{E\left[f_{it}\right]}{\sigma^2\left[f_{it}\right]}$, then (1) expresses a multi-beta factor model in the spirit of Ross (1976) and Sharpe (1977). For the purpose of this paper, however, interest is focused on proportionality of the economic aggregate to the market return as given by (2).

The model of (1) and (2) can be expressed as a linear multivariate regression

$$r_t = \alpha + \beta r_{mt} + e_t \tag{3}$$

where $E\left[e_t r_{mt}\right]=0$ and $\alpha=0$. Consider a decomposition of r_{mt} as

$$r_{mt}=\phi_t+\phi_p r_{pt}, \tag{4}$$

which represents a generalization of (17) in Jagannathan and Wang (1996). The variable ϕ_t reflects components to the market return that are excluded from the proxy return. Examples of these components include returns to nontraded assets and/or human capital. Studies by Campbell (1996), Jagannathan and Wang (1996), and Dittmar (2002) note the importance of the return to human capital in pricing expected returns. Substitution of (4) into (3) produces

$$r_t=\gamma+\delta r_{pt}+\widetilde{e}_t \tag{5}$$

where

$$\begin{aligned} \gamma &= \alpha+\beta E\left[\phi_t\right], \quad \delta=\beta\phi_p \\ \widetilde{e}_t &= \beta\widetilde{\phi}_t+e_t \\ \widetilde{\phi}_t &= \phi_t-E\left[\phi_t\right] \end{aligned} \tag{6}$$

(5) expresses a linear multivariate relation between security returns and a proxy return. Suppose $Cov\left[\phi_t,\ r_{pt}\right]\neq 0$; which is to say, for example, that returns to human capital correlate with other observable asset returns. Then $E\left[\widetilde{e}_t r_{pt}\right]\neq 0$, or, equivalently, r_{pt} is an endogenous regressor in (5). As a result, (5) is a structural equation that unlike (3) cannot, necessarily, be treated as a linear projection without loss of generality. The fact that the market return is unobservable and any proxy return, by definition, is incomplete affords this distinction. The effects of this distinction on the efficiency of a proxy return relative to the market return is made explicit in the statement of Proposition 1 below.

According to Cochrane (2001), "*all factor models are derived as specializations of the consumption-based model*" (p. 151). (1) reflects this fact. Empirically-based factor models attempt to tie the discount factor m_t to observable variables. Towards that end, consider a linear projection of m_t onto r_{pt}:

$$m_t=a+br_{pt}+e_{mt} \tag{7}$$

According to Lemma 1 of Shanken (1987), the combination of (5) and (7) implies that

$$Cov\left[\widetilde{e}_t,\ e_{mt}\right]'\Sigma_{\widetilde{e}}^{-1}Cov\left[\widetilde{e}_t,\ e_{mt}\right]\leq\sigma^2\left(m_t\right)\left(1-\rho^2\right) \tag{8}$$

where $\Sigma_{\widetilde{e}}$ is the covariance matrix of $\widetilde{e}_t$, and ρ is the correlation between m_t and r_{pt}. All proofs in this section are given in Appendix A. From Shanken (1987), $Cov\left[\widetilde{e}_t,\ e_{mt}\right]$ "may be interpreted as a vector of deviations from an exact [single] beta expected return relation" (p. 93). (8) places an upper bound on these deviations and is useful in determining a similar bound for deviations from CAPM pricing measured conditional on a proxy return. Proposition 1 formalizes this result in light of the structural equation in (5) and the potential nonzero covariance between $\widetilde{e}_t$ and r_{pt}.

Proposition 1 *Let the pricing model of (1) and (2) hold for all security returns including the proxy return, and consider the structural relationship between security returns and the proxy return* *as given by (5). Define*

$$\theta_p = \frac{E\left[r_{pt}\right]}{\sigma\left[r_{pt}\right]} \tag{9}$$

as the Sharpe performance measure for the proxy return, and

$$\eta = \frac{Cov\left[\widetilde{e}_t,\ r_{pt}\right]}{\sigma^2\left[r_{pt}\right]} \tag{10}$$

as a measure of the degree to which unobservable components to the market return covary with the proxy return. Then,

$$d'\Sigma_{\widetilde{e}}^{-1}d \leq \theta_p^2(\rho^{-2} - 1) \tag{11}$$

where

$$d = E\left[r_t\right] - (\delta + \eta)\, E\left[r_{pt}\right].$$

As stated emphatically by Roll (1977), the pricing restriction of (1) and (2) is not testable because the market return is unobserved. Notice that with the exception of ρ, (11) is constructed in terms of quantities that can be directly estimated from observable data, provided, of course, that (5) can be identified. Proposition 1, therefore, is a testable analog to (1) and (2) conditional on a prior belief about the value of ρ. This proposition provides a means for testing the mean-variance efficiency of a given proxy return relative to the market return.

The proof of Proposition 1 in the appendix demonstrates that

$$\rho = \frac{\theta_p}{\sigma\left[m_t\right]},$$

implying that ρ is strictly positive. Let $\theta_m = \frac{E[r_{mt}]}{\sigma[r_{mt}]}$, the Sharpe performance measure for the market return. Given (2), $\sigma\left[m_t\right] = \theta_m$, and ρ is a ratio of Sharpe performance measures. As a result, ρ is afforded a geometric interpretation in mean-standard deviation space as the ratio of the slope of the security market line passing through the excess proxy return to the slope of the security market line tangent to the minimum variance boundary at the excess market return. This ratio gauges the relative efficiency of the excess proxy return.

Corollary 1 *In (4), suppose that ϕ_t is constant such that $\phi_t = \phi_c$ and ϕ_p is positive. Then*

$$d'\Sigma_{\widetilde{e}}^{-1}d = 0 \tag{12}$$

where $d = \alpha$ from (3) if and only if $\phi_c = 0$.

Under this corollary, the decomposition of the market return in (4) is identical to that given in Jagannathan and Wang (1996). If $\phi_c = 0$, then $\rho = 1$ and (11) holds as an equality to zero. This result provides a basis for tests of mean-variance efficiency like those proposed by Gibbons, Ross, and Shanken (1989) as well as MacKinlay and Richardson (1991) since these tests rely on the assumption that $\phi_t = \phi_c$ so that $Cov\left[\widetilde{e}_t,\ r_{pt}\right] = 0$. For instance, (12) is a statement of the joint null hypothesis

$$H_0 : \alpha = 0; \quad \phi_c = 0, \tag{13}$$

since

$$d = E[r_t] - \delta E[r_{pt}] = \alpha + \beta\phi_c \tag{14}$$

given (5) and (6). Failure to reject this null is a failure to reject direct proportionality between the market and proxy return as well as mean-variance efficiency of the market return. Rejection of this null, on the other hand, is only a rejection of mean-variance efficiency of the proxy return, since either $\phi_c \neq 0$, in which case the proxy return is inefficient because $\rho < 1$, or $\alpha \neq 0$, in which case the proxy and the market return are inefficient, or both.[4] The inability to distinguish between these alternatives illustrates the Roll (1977) critique that the CAPM theory is not directly testable.

If $\phi_t = \phi_c$, then (5) can be treated as a projection equation without loss of generality. In this case, the manner in which Proposition 1 allows for an indirect assessment of the CAPM is parallel to that of Proposition 2 in Shanken (1987). Specifically, if $\alpha = 0$, then $d \neq 0$ because $\phi_c \neq 0$. For a given Sharpe performance measure of the proxy return, the magnitude of this distance d away from zero is bounded from above by the correlation between the market return and the proxy return.[5] Let ρ_o be a prior belief on the true value of ρ. For a given d, $\Sigma_{\widetilde{e}_t}$, and θ_p, let $\overline{\rho}$ be the value of $\rho \in (0,1]$ that, if it exists, satisfies (11). If $\overline{\rho} < \rho_o$, such a result is interpreted as evidence that not only is $\phi_c \neq 0$ but $\alpha \neq 0$ as well. The strength of this evidence increases as $\overline{\rho} - \rho_o$ becomes more negative and is, of course, conditional on the correctness of ρ_o.

Suppose $\phi_t \neq \phi_c$ so that the structural equation in (5) no longer coincides with a projection of r_t onto r_{pt}. The effects of this possibility are explored in the following corollary.

Corollary 2 *Let e_{pt} be the errors from a linear multivariate projection of r_t on r_{pt}, and define Σ_{e_p} as the variance-covariance matrix of these errors. Given (4), $\Sigma_{\widetilde{e}} - \Sigma_{e_p}$ is positive semi-definite.*

According to Corollary 2, $\Sigma_{\widetilde{e}} \geq \Sigma_{e_p}$ and, by extension, $\Sigma_{e_p}^{-1} \geq \Sigma_{\widetilde{e}}^{-1}$. From (52) and (53) in the proof of Corollary 2,

$$d = E[r_t] - (\delta + \eta) E[r_{pt}] = \alpha_p,$$

the vector of constant terms from a linear multivariate projection of r_t on r_{pt}. If (5) is replaced by (51), then the left-hand-side of (11) becomes $\alpha_p' \Sigma_{e_p}^{-1} \alpha_p$. Otherwise, the left-hand-side of (11) is $\alpha_p' \Sigma_{\widetilde{e}}^{-1} \alpha_p$ and

$$\alpha_p' \Sigma_{e_p}^{-1} \alpha_p \geq \alpha_p' \Sigma_{\widetilde{e}}^{-1} \alpha_p.$$

From (55), a case where these two quadratic forms equate is when $\phi_t = \phi_c$. In general, however, the degree to which expected returns deviate from the CAPM prediction measured conditional on a proxy return will tend to be overstated if $\Sigma_{e_p}^{-1}$ is used as a weighting matrix as opposed to $\Sigma_{\widetilde{e}}^{-1}$. As a consequence, $\overline{\rho}$ will tend to be understated. The end result is that treating the relationship between security returns and the proxy return as a projection equation instead of a structural equation will bias test results of the inequality restriction in Proposition 1 towards rejecting the CAPM theory.

Hansen and Jagannathan (1997) criticize model misspecification tests that depend on the variance-covariance matrix of the pricing errors because these tests grant a "reward for sampling error associated with the sampling mean." In reference to the CAPM, this paper argues that higher sampling

[4] If both $\alpha \neq 0$ and $\phi_c \neq 0$, then the market return is not mean-variance efficient. The proxy return is also not mean-variance efficient if β and ϕ_c are both positive or if $\alpha \neq \beta\phi_c$ (see (5) and (6)).

[5] Direct proportionality between the economic aggregate and the market return in (2) leads to $\rho = \frac{Cov[r_{mt}, r_t]}{\sigma[r_{mt}]\sigma[r_t]}$.

error should be accounted for to the extent that it relates to misspecification of the market return. Ignoring this misspecification will bias the test results towards rejecting the theory because of the proxy being used in the test not because of any failing in the theory itself.

If $\phi_t = \phi_c$, then according to (14),

$$\alpha_p = \alpha + \beta\phi_c. \tag{15}$$

In this case, the difference between the alpha proxy and the true alpha is directly proportional to the location-shift in the market return relative to the proxy return. This difference is expected to be positive (negative) if β is positive (negative), since a negative ϕ_c implies that $\rho > 1$ given (50). If $\phi_t \neq \phi_c$, then

$$\alpha_p = \alpha + \beta E\left[\phi_t\right] - \eta E\left[r_{pt}\right] \tag{16}$$

given (53). In this case, the difference between the alpha proxy and the true alpha is ambiguous. Affecting this difference are both the mean of the omitted components as well as the covariance between those components and the proxy return. Provided that the CAPM holds, (15) explains the empirical discovery of "significant" alpha to be a product of the given security's sensitivity to changes in the market return and the souce of inefficiency in the proxy return. (16) adds to this explanation the sensitivity of the proxy return to changes in the unobserved components of the market return.

3. Conventional Tests

Suppose $\phi_t = \phi_c$ in (4), and assume that $e_t \sim N\left(0,\, \Sigma_e\right)$. Let $\widehat{d}$ and $\widehat{\Sigma}_e = \frac{1}{T}\sum\limits_{t=1}^{T} \widehat{e}_t \widehat{e}_t'$ denote estimates of α_p in (15) and Σ_e, respectively, from N separate OLS regressions of r_{it} on r_{pt}, where r_{it} is the ith element of r_t and $t = 1, \ldots, T$. $\widehat{\theta}_p$ is an estimate of the proxy performance measure computed from the sample mean and variance of r_{pt}. Consider the following definitions:

$$Q \equiv \frac{T\widehat{d}'\widehat{\Sigma}_e^{-1}\widehat{d}}{1 + \widehat{\theta}_p^2}; \qquad \lambda \equiv \frac{Td'\Sigma_e^{-1}d}{1 + \widehat{\theta}_p^2}.$$

Gibbons, Ross, and Shanken (1989) show that $\left[N^{-1}\left(T - N - 1\right)/T - 2\right]Q$, conditional on r_{pt}, is distributed as a noncentral F with degrees of freedom N and $T - N - 1$ and non-centrality parameter λ. Multiply both sides of (11) by $T/\left(1 + \widehat{\theta}_p^2\right)$. Then Proposition 1 is equivalent to

$$H_0 : \lambda \leq \frac{T\theta_p^2(\rho^{-2} - 1)}{1 + \widehat{\theta}_p^2}, \tag{17}$$

which establishes an upper-bound on the non-centrality parameter. If $\rho = 1$, then $\lambda = 0$, meaning that under Corollary 1, Q follows a central F distribution. In this case, a test of (13) follows immediately because Q is stated entirely in terms of observable quantities. Suppose, instead, that $\rho < 1$. Then, conditional on a value for θ_p, (17) can be tested for a given value for ρ. Shanken (1987) follows this approach. Alternatively, the value of $\overline{\rho}$ can be found (conditional on a value for θ_p) that satisfies the restriction in (17) at a desired significance level. Following the discussion in section 2, whether ρ_o is greater than (less than) $\overline{\rho}$ then determines whether (17) is rejected (not rejected).

Violations of the normality assumption for e_t are documented in the literature.[6] Numerous studies support Engle's (1982) Autoregressive Conditional Heteroskedasticity (ARCH) and Bollerslev's (1986) Generalized ARCH (GARCH) in security returns.[7] Common specifications of these models assume e_t to be conditionally normal, which (as demonstrated by Milhoj (1985) or Bollerslev (1986)) results in the unconditional distribution of e_t being leptokurtic, although their standardized residuals are still shown to be non-normal. In light of this evidence, the potential for mean-variance efficiency tests like those just described to be sensitive to the normality assumption motivated the search for more robust testing methods. From the results of section 2, it is apparent that normality is not necessary for deriving data-dependent restrictions implied by mean-variance efficiency (or relative efficiency). Rather, such a condition is statistically convenient for determining the distributional properties of the resulting test statistics. With this observation in mind, MacKinlay and Richardson (1991) proposed a GMM-based test that, by construction, is distibution free and able to accommodate general forms of heteroskedasticity. These authors uncovered material differences between their approach and that of Gibbons et al. (1989) at conventional levels of significance.

A unifying restriction of both Gibbons et al. (1989) and MacKinlay and Richardson (1991) is that $\phi_t = \phi_c$. Corollary 2 illustrates how a violation of this assumption could impact a test of either mean-variance efficiency (see Corollary 1) or relative mean-variance efficiency (see Proposition 1). The testing methodology developed in the next section is robust to ϕ_t and is built upon the premise that $\widetilde{e}_t$ follows a GARCH process but one that is not , necessarily, conditionally normal.

4. Econometric Methodology

An empirical investigation into Proposition 1 requires estimation of all quantities, with the exception of ρ, in (11). From the proof of Corollary 2, d is the vector of constant terms from a multivariate projection of r_t onto r_{pt}. As such, the individual elements of d can be estimated following the same approach outlined in section 3. If $\phi_t = \phi_c$, then $\Sigma_{\widetilde{e}} = \Sigma_e$ and can also be estimated in the manner described under section 3. If, on the other hand, $\phi_t \neq \phi_c$, then r_{pt} is an endogenous regressor in (5). Any method for estimating (5) and, hence, $\Sigma_{\widetilde{e}}$ needs to be robust to this endogeneity.

If $Cov\left[\phi_t,\ r_{pt}\right] \neq 0$, then (5) represents a triangular system. In general, such a system is expressed as

$$Y_{1,t} = X_t'\gamma_{1o} + Y_{2,t}\delta_o + \epsilon_{1,t} \tag{18}$$

$$Y_{2,t} = X_t'\gamma_{2o} + \epsilon_{2,t} \tag{19}$$

where $Y_{1,t}$ and $Y_{2,t}$ are observed endogenous variables; X_t is a vector of predetermined variables that can include lags of the endogenous variables, and $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are unobserved errors or shocks. Let $\epsilon_t = \begin{bmatrix}\epsilon_{1,t} & \epsilon_{2,t}\end{bmatrix}'$. The term γ_{1o} refers to the true value of γ_1, with the same interpretation

[6] See Mandelbrot (1963) and Fama (1965) as early examples.

[7] Bodurtha and Mark (1991) find evidence of ARCH in a conditional test of the CAPM.

holding for all other parameter values. In the context of section 2, $Y_{1,t}$ is a given excess security return, $Y_{2,t}$ a proxy to the excess market return, and X_t is inclusive of only a constant term. In general, X_t can also contain forecasting instruments, in which case (18) is analogous to (40) in section 6. In this case, it is assumed that these forecasting instruments apply to both individual security returns and the proxy return.[8] Regardless of the specification for X_t, (18) and (19) are stated such that mean restrictions (e.g., zero restrictions on some of the parameters in γ_{2o}) are not available for identifying the structural form of (18). Under these conditions, the sketch of an identification method for (18) follows together with a proposed estimator. A complete treatment of identification is relegated to Appendix B. Note that (18) makes no explicit use of the error decomposition in (6), meaning that the effect of ϕ_t is considered at the level of $\epsilon_{1,t}$ and does not attempt to isolate or identify properties unique to ϕ_t. The functional form describing the relationship between $\epsilon_{1,t}$ and $\epsilon_{2,t}$ (see (21) below) is sufficiently general to place only minimal constraints on the process governing ϕ_t.

Let Ω_{t-1} be the information available to investors in period $t-1$, and consider $S_{t-1} \subset \Omega_{t-1}$ that is observable to the econometrician. Assume that ϵ_t follows the definition of semi-strong GARCH given in Drost and Nijman (1993) so that

$$E\left[\epsilon_t \mid S_{t-1}\right] = 0, \qquad E\left[\epsilon_t \epsilon_t^{'} \mid S_{t-1}\right] = H_t. \tag{20}$$

Given (20), H_t needs to be positive definite. In order to meet this criterion, parameterize H_t according to the diagonal BEKK specification of Engle and Kroner (1995),

$$\begin{aligned} H_t &= C_o' C_o + \sum_{k=1}^{2} A_{ko}' \epsilon_{t-1} \epsilon_{t-1}' A_{ko} + \sum_{k=1}^{2} B_{ko}' H_{t-1} B_{ko} \\ C_o &= \begin{bmatrix} c_{11o} & 0 \\ c_{21o} & c_{22o} \end{bmatrix} \\ A_{1o} &= \begin{bmatrix} a_{11,1o} & 0 \\ 0 & a_{22,1o} \end{bmatrix}, \qquad A_{2o} = \begin{bmatrix} a_{11,2o} & 0 \\ 0 & 0 \end{bmatrix} \\ B_{1o} &= \begin{bmatrix} b_{11,1o} & 0 \\ 0 & b_{22,1o} \end{bmatrix}, \qquad B_{2o} = \begin{bmatrix} b_{11,2o} & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned} \tag{21}$$

where the parameters c_{11o}, c_{22o}, $a_{22,1o}$, $a_{11,2o}$, $b_{22,1o}$, and $b_{11,2o}$ are strictly positive.[9] Let

$$\begin{aligned} vech\left(H_t\right) &= h_t; \qquad vech\left(\epsilon_t \epsilon_t^{'}\right) = e_t \\ vech\left(C_o' C_o\right) &= \widetilde{C}_o \end{aligned}$$

Throughout this section and Appendix B, $vech\left(\cdot\right)$ denotes the matrix operator that stacks the lower triangle, including the diagonal, of a symmetric matrix into a column vector, and $vec\left(\cdot\right)$ is the

[8] In support of this assumption, it seems difficult to envision an instrument that is strongly related to a given proxy return yet wholly unrelated to an individual security return or porfolio of security returns.

[9] Proposition 2.6 of Engle and Kroner (1995) establishes that the diagonal BEKK model includes all positive definite diagonal GARCH models.

matrix operator that stacks the columns of a matrix into a column vector. Then, given (21),

$$h_t = \widetilde{C}_o + A_o e_{t-1} + B_o h_{t-1}, \tag{22}$$

where $\widetilde{C}_o$ is a 3×1 column vector and A_o and B_o are both 3×3 diagonal matrices. (22) is the bivariate analog to the GARCH(1, 1) model of Bollerslev (1986). Let $i, j = 1, 2$. Following from recursive substitution, (22) restricts $h_{ij,t}$ to be a linear function of $\epsilon_{i,t-\tau}\epsilon_{j,t-\tau}$ for $\tau \geq 1$. If $i \neq j$, then $h_{ij,t}$ does not depend on either $\epsilon^2_{i,t-\tau}$ or $\epsilon^2_{j,t-\tau}$. Alternatively, if $i = j \neq k$ where $k = 1, 2$, then $h_{ij,t}$ does not depend on $\epsilon_{i,t-\tau}\epsilon_{k,t-\tau}$ or $\epsilon^2_{k,t-\tau}$. These zero restrictions on the functional form of $h_{ij,t}$ are what identify (18). An examination of the reduced form to (22) illustrates this result.

Let $R_t = \begin{bmatrix} R_{1,t} & R_{2,t} \end{bmatrix}'$ be the reduced form errors from (18) and (19). Relating these reduced form errors to their structural form counterparts is the equation

$$\epsilon_t = \Delta_o^{-1} R_t, \tag{23}$$

where $\Delta_o = \begin{bmatrix} 1 & \delta_o \\ 0 & 1 \end{bmatrix}$. (23) can be used to solve for the reduced form of H_t in (21). Let $H_{r,t}$ denote this reduced form. In addition

$$vech\left(H_{r,t}\right) = h_{r,t}; \qquad vech\left(R_t R_t'\right) = r_t.$$

Then

$$h_{r,t} = \widetilde{C}_{r,o} + A_{r,o} r_{t-1} + B_{r,o} h_{r,t-1} \tag{24}$$

is the reduced form of (22), where $A_{r,o}$ and $B_{r,o}$ are symmetrically defined. The upper triangle of $A_{r,o}$ ($B_{r,o}$) including the diagonal is composed of nonzero elements. Off diagonal terms in this upper triangle are functions of the structural parameters in A_o (B_o) as well as the parameter δ_o from (18). Let $a_{ij,ro}$ be the element in the ith row and jth column of the matrix $A_{r,o}$, and define $b_{ij,ro}$ in a parallel manner for $B_{r,o}$. Consider the 2×2 partition of A_{ro} beginning with the element $a_{22,ro}$:

$$\begin{bmatrix} a_{22,ro} & a_{23,ro} \\ 0 & a_{33,r}o \end{bmatrix} = \begin{bmatrix} a_{11,1o}a_{22,1o} & \delta_o\left(a^2_{22,1o} - a_{11,1o}a_{22,1o}\right) \\ 0 & a^2_{22,1o} \end{bmatrix}. \tag{25}$$

Then it follows that if $A_{r,o}$ and $B_{r,o}$ are identified, so too is δ_o as

$$\delta_o = \frac{a_{23,ro} + b_{23,ro}}{\left(a_{33,ro} - a_{22,ro}\right) + \left(b_{33,ro} - b_{22,ro}\right)}. \tag{26}$$

The formal identification result is stated as Proposition 2 in Appendix B.

Proposition 2.1 of Iglesias and Phillips (2004) demonstrates that if the structural errors from a triangular system follow a diagonal GARCH process like (21), the reduced form errors, while still GARCH, are no longer diagonal GARCH. (25) illustrates this result. (25) also illustrates that it is precisely this departure from diagonality in the reduced form that identifies (18).

The above identification result is a second-moment analog to exclusion restrictions for γ_{1o}. In (22), A_o and B_o impose zero restrictions on all off-diagonal elements. Suppose instead, that A_o and B_o are fully general and contain no zero elements. Then, the number of reduced form parameters in $A_{r,o}$ ($B_{r,o}$) is less than the corresponding number of structural parameters (i.e., those in A_o (B_o) plus δ_o), and, as a consequence, δ_o is not identified.

The identification result sketched above can be used to define a consistent estimator for (18) and (19). In defining this estimator, two observations are important. First, from (25), identification of δ_o depends on the conditional covariance between $\epsilon_{1,t}$ and $\epsilon_{2,t}$ as well as the conditional variance of $\epsilon_{2,t}$. Given this observation, the proposed estimator is defined in terms of $\overline{e}_t = \begin{bmatrix} \epsilon_{1,t}\epsilon_{2,t} & \epsilon_{2,t}^2 \end{bmatrix}'$. In addition, $\overline{A}_o$ is defined as the 2×2 diagonal matrix formed from the elements a_{22o} and a_{33o} in A_o (see (22)), with $\overline{B}_o$ being similarly defined in terms of the elements of B_o. Finally, let the relationship between $\overline{A}_{r,o}$ and $A_{r,o}$ mirror that between $\overline{A}_o$ and A_o, with the same holding true for $\overline{B}_{r,o}$ and $B_{r,o}$. Second, notice that (17) can be rewritten as

$$\delta_o = \frac{a_{23,ro} + b_{23,ro}}{\left(a_{33,ro} + b_{33,ro}\right) - \left(a_{22,ro} + b_{22,ro}\right)}. \tag{27}$$

The implication of this observation is that identification of δ_o depends on $\overline{A}_{r,o} + \overline{B}_{r,o}$. Separate identification of $\overline{A}_{r,o}$ and $\overline{B}_{r,o}$, while sufficient, is not necessary to identify δ_o.

Following from (20),

$$E\left[X_t \otimes \epsilon_t\right] = 0, \tag{28}$$

where $\otimes$ is the kronecker product. Let $\overline{\Phi}_o = \overline{A}_o + \overline{B}_o$. In addition,

$$E\left[\overline{e}_t\right] = \overline{\sigma}_o \tag{29}$$

where $\overline{\sigma}_o = \begin{bmatrix} \sigma_{12o} & \sigma_{22o} \end{bmatrix}'$, $\sigma_{12o} = \frac{c_{21o}c_{22o}}{1-\phi_{11o}}$ and $\sigma_{22o} = \frac{c_{22o}^2}{1-\phi_{22o}}$. From (60),

$$Cov\left[\overline{e}_t, \overline{e}_{t-\tau}\right] = \overline{\Phi}_o^{\tau-1} Cov\left[\overline{e}_t,\, \overline{e}_{t-1}\right], \qquad \tau \geq 1. \tag{30}$$

Let $\psi = \{\gamma_1,\, \gamma_2,\, \delta,\, c_{21},\, c_{22},\, \phi_{11},\, \phi_{22}\}$, and define Ψ as the set of all possible values for ψ. Given Corollary 3 in Appendix B, the moment conditions in (28)–(30) are uniquely satisfied at $\psi = \psi_o$. Define

$$\epsilon_{1,t} = Y_{1,t} - X_t'\gamma_1 - Y_{2,t}\delta, \quad \epsilon_{2,t} = Y_{1,t} - X_t'\gamma_2,$$

and let

$$g_1 = \widehat{E}\left[X_t \otimes \epsilon_t\right], \qquad g_2 = \widehat{E}\left[\overline{e}_t\right] - \overline{\sigma},$$
$$g_3 = \widehat{Cov}\left(\overline{e}_t, \overline{e}_{t-\tau}\right) - \overline{\Phi}^{\tau-1}\widehat{Cov}\left(\overline{e}_t,\, \overline{e}_{t-1}\right), \quad \tau \geq 2.$$

where $\widehat{E}$ and $\widehat{Cov}$ are the finite sample analogs to expectation and covariance operators, respectively. Let $g = \begin{bmatrix} g_1 & g_2 & vec\left(g_3\right) \end{bmatrix}'$. Then Theorem 2.6 of Newey and McFadden (1994) can be used to establish

$$\widehat{\psi} = \arg\min_{\psi \in \Psi} g'Wg \tag{31}$$

as a consistent estimator, where W is a positive definite weighting matrix. (31) is the estimator for triangular systems given semi-strong GARCH developed in Prono (2008). Monte Carlo studies of

$\widehat{\psi}$ support this result.[10] If $W = I$, then (31) is the single-step GMM estimator–see Hansen (1982).

Let $\zeta^2_{iio} = E\left[\left(\epsilon_{i,t}\epsilon_{2,t} - \sigma_{i2}\right)^2\right]$ for $i = 1, 2$, and define $\widetilde{\zeta}_{ii}$ as a preliminary estimate of ζ_{iio}. Construct $\widetilde{Z} = \begin{bmatrix} \widetilde{\zeta}_{11} & 0 \\ 0 & \widetilde{\zeta}_{22} \end{bmatrix}$. Suppose X_t is a $k \times 1$ vector, and consider the following alternative weighting matrix:

$$W\left(\widetilde{Z}\right) = \begin{bmatrix} I_{2k\times 2k} & \cdots & 0 \\ \vdots & I_{2\times 2} & \vdots \\ 0 & \cdots & \left(\widetilde{Z} \otimes \widetilde{Z}\right)^{-1} \end{bmatrix}. \quad (32)$$

The weights $\left(\widetilde{Z} \otimes \widetilde{Z}\right)^{-1}$ impact the moments that define the autocovariances of $\overline{e}_t$(i.e., g_3), transforming these autocovariances into autocorrelations. Prono (2008) documents improved finite sample properties of (31) if $W = W\left(\widetilde{Z}\right)$ as opposed to $W = I$. For this reason, estimation of (18) and (19) is based on (31) with $W = W\left(\widetilde{Z}\right)$.

Application of (31) to the N separate structural equations implied by (5) produces consistent estimates of the N elements in $\widetilde{e}_t$, which can then be used to estimate $\Sigma_{\widetilde{e}}$. To close this section, $\widehat{\theta}_p = \frac{\widehat{E}[r_{pt}]}{\widehat{\sigma}_{22}}$.

5. Test Methodology

The inequality restriction in (11) implies that

$$\rho \leq \sqrt{\frac{1}{1 + \theta_p^2 d' \Sigma_{\widetilde{e}}^{-1} d}}$$

which identifies an upper bound for ρ since ρ is strictly positive. Define $\xi \equiv \sqrt{\frac{1}{1+\theta_p^2 d' \Sigma_{\widetilde{e}}^{-1} d}}$. Section 4 outlines a methodology for obtaining $\widehat{\xi}$. Following Gibbons, Ross, and Shanken (1989), one method for testing Proposition 1 would be to determine the distribution of ξ and, based on that distribution, determine the value $\overline{\xi}$ such that one is indifferent between rejecting or not rejecting the null hypothesis

$$H_0 : \xi > \overline{\xi} \quad (33)$$

at a significance level of α. If $\overline{\xi} < \rho_o$, this evidence rejects Proposition 1 and, hence, the CAPM. Determining a distribution for ξ, however, would be difficult, owing, in no small part, to the heteroskedastic properties assumed for $\widetilde{e}_t$ (and for which there is substantial empirical evidence) that facilitates its identification. An alternative approach would be to bootstrap a standard error for $\widehat{\xi}$ and use this standard error to determine a value for $\overline{\xi}$ that satisfies (33). This paper adopts the alternative methodology.

Bootstrapping a standard error for $\widehat{\xi}$ requires resampling from the N excess security returns and the excess proxy return used to form the quantities $\widehat{\theta}_p$, $\widehat{d}$, and $\widehat{\Sigma}_{\widetilde{e}}$. Such is a nontrivial exercise

[10]These studies were presented in an earlier version of this paper and are available upon request.

since these returns are not iid and, in fact, their departure from independence (both within and across return series) is a key assumption underlying the estimator that generates $\widehat{\Sigma}_{\widetilde{e}}$. Define

$$\epsilon_t^{(i)} = \begin{bmatrix}\epsilon_{i,t} & \epsilon_{2,t}\end{bmatrix}', \quad i = 1, \ldots, N,$$

where $\epsilon_{i,t} = Y_{i,t} - \gamma_{io} - Y_{2,t}\delta_o$, the errors from the structural equation for the ith security return, and $\epsilon_{2,t}$ is the demeaned proxy return. Suppose that

$$\epsilon_t^{(i)} = \left(H_t^{(i)}\right)^{1/2} V_t^{(i)}, \tag{34}$$

where $H_t^{(i)}$ is the conditional variance-covariance matrix for the ith security return and the proxy return parameterized according to (21), and $V_t^{(i)} = \begin{bmatrix}V_{i,t} & V_{2,t}\end{bmatrix}'$. The vector $V_{i,t}$ is assumed to be iid with mean zero and identity variance-covariance matrix. (34) defines a strong GARCH process, which is implied by (20). Unlike most applications of strong GARCH, however, no particular distribution is assumed for $V_t^{(i)}$. The estimator in (31) supplies $\widehat{\epsilon}_t^{(i)}$. Conditional on this estimate, one can obtain $\widehat{H}_t^{(i)}$. As a result, $\widehat{V}_t^{(i)} = \left(\widehat{H}_t^{(i)}\right)^{-1/2}\widehat{\epsilon}_t^{(i)}$. Bootstrap samples are drawn from $\widehat{V}_t^{(i)}$. Let $\widehat{V}_t^{(i)*}$ be a bootstrap sample. Then $\widehat{\epsilon}_t^{(i)*} = \left(\widehat{H}_t^{(i)*}\right)^{1/2}\widehat{V}_t^{(i)*}$,where $\widehat{H}_t^{(i)*}$ is based upon parameter estimates from the original sample, and

$$\begin{aligned}\widehat{Y}_{2,t}^* &= \widehat{\gamma}_{2o} + \widehat{\epsilon}_{2,t}^*, \\ \widehat{Y}_{i,t}^* &= \widehat{\gamma}_{io} + \widehat{Y}_{2,t}^*\widehat{\delta}_{io} + \widehat{\epsilon}_{i,t}^*, \quad i = 1, \ldots, N,\end{aligned}$$

where $\widehat{\gamma}_{io}$, $\widehat{\gamma}_{2o}$, and $\widehat{\delta}_{io}$ are also obtained from the original sample. The resulting bootstrap series is then used to estimate $\widehat{\xi}^*$ given the estimation method described in section 4.

Define E^* as the expectation operator relative to the bootstrap sample conditional on the original sample, and let

$$g = \frac{1}{T}\sum_{t=1}^{T} g_t.$$

Following Hall and Horowitz (1996), the bootstrap version of the moment conditions in (31) is

$$g_t^* = g_t - E^*\widehat{g}, \tag{35}$$

where $\widehat{g}$ is g evaluated at $\widehat{\psi}$, the parameter estimates from the original data sample. (35) recenters the bootstrap moment conditions such that $E^*g_t^* = 0$. In general, $E^*g_t \neq 0$ when the number of moment conditions exceeds the number of parameters in ψ. If g_t is used as the moment conditions instead of g_t^*, then $\widehat{\psi}^*$ will have different asymptotic properties than $\widehat{\psi}$. In order to avoid this discrepancy,

$$\widehat{\psi}^* = \arg\min_{\psi\in\Psi} g^{*\prime}W^*g^*,$$

where $g^* = \widehat{E}^*g_t^*$ and $W^* = W\left(\widetilde{Z}^*\right)$, the bootstrap analog to (32). The bootstrap standard error of $\widehat{\xi}$ is based on $\widehat{\psi}^*$.

Given a standard error for $\widehat{\xi}$, the asymptotic t statistic

$$\frac{\widehat{\xi}-\overline{\xi}}{\widehat{se}\left(\widehat{\xi}\right)} \tag{36}$$

can be constructed to assess (33) in the manner described. This statistic is asymptotically pivotal with a standard normal asymptotic distribution.[11] According to MacKinnon (2007), bootstrapping (36) will generally lead to an asymptotic refinement. Such a practice is referred to as the double or iterated bootstrap. Implementing the double bootstrap, however, is very computationally expensive. For example, define B_1 as the number of bootstrap iterations used to generate $\widehat{se}\left(\widehat{\xi}\right)$ and B_2 as the number of iterations used to generate the bootstrap distribution of (36). If $B_1 = B_2 = 1000$, then the total number of iterations required for the double bootstrap is approximately 1 million. Given the size of the data samples used to construct $\widehat{\xi}$ (see section 6), the asymptotic distribution of (36) will likely provide a good approximation. As a result, this approximation is used as opposed to the double bootstrap alternative.

6. Test Results

All tests are conducted using size, B/M, and momentum portfolios. The returns are measured weekly (in percentage terms) from 10/6/67 through 9/28/07. Test results consider 20- and 10-year subperiods of this overall date range. The daily 25 size-B/M and 25 size-momentum return files (each 5×5 sorts with breakpoints determined by NYSE quintiles) formed from all securities traded on the NYSE, AMEX, and NASDAQ exchanges are used to construct the weekly return series. Monte Carlo studies of (31) reveal sizable benefits in terms of reduced finite sample bias and increased efficiency from using large sample sizes due to the fact that higher moments are being estimated. In light of this finding, weekly returns are utilized. Further supporting this frequency choice is the fact that weekly returns reduce day-of-the-week and weekend effects as well as the effects of nonsynchronus trading and bid-ask bounce. The size portfolios considered are "Small," "Mid," and "Large." "Small" is the average of the five low-market-cap portfolios, "Mid" the average of the five medium-market cap portfolios, and "Big" the average of the five large-market-cap portfolios. The B/M portfolios considered are "Value," Neutral," and "Growth." Value" is the average of the five high-B/M portfolios, "Neutral" the average of the five middle-B/M portfolios, and "Growth" the average of the five low-B/M portfolios. Finally, the momentum portfolios considered are "Losers," "Draws," and "Winners." "Losers" is the average of the five low-return-sorted portfolios, "Neutral" the average of the five middle-return sorted portfolios, and "Winners" the average of the five high-return-sorted portfolios. The proxy return is the CRSP value-weighted index return formed from all securities traded on the NYSE, AMEX, and NASDAQ exchanges. The risk-free rate is the one-month Treasury bill rate from Ibbotson Associates.

The tests focus on (17) and (33). The former is conducted following the approach developed in Gibbons, Ross, and Shanken (1989), referred to hereafter as GRS, and implemented in Shanken

[11] From MacKinnon (2007), "a test statistic is aymptotically pivotal if its asymptotic distribution does not depend on anything that is unknown" (p.5).

(1987) described in section 3. The latter is conducted following the approach of section 5 under two cases: (1) $\phi_t = \phi_c$, (2) ϕ_t is stochastic. Case 1 will be referred to as Bootstrap Proposition 1 constant (BPC), while case 2 will be referred to as Bootstrap Proposition 1 stochastic (BPS). A comparison of BPC to GRS evidences the effects of conditional heteroskedasticity on a test of mean-variance efficiency. A comparison of BPC to BPS evidences the effects of relating security returns to the proxy return via a projection as opposed to a structural equation. When implementing (1) and (2), the number of lags used in (31) is set to $\tau = 4$. The choice of this lag length is motivated by the frequency of returns as well as the finding in Prono (2008) that higher lag lengths, while successful at reducing the variability of $\widehat{\psi}$ also increases the finite sample bias.

Table 1 (A and B) and 2 (A and B) summarizes results from two 20-year subperiods: (1) 10/6/67 - 9/25/87, (2) 11/6/87 - 9/28/07. Tables 1A and 2A provide summary statistics of the returns used in the tests as well as the alpha proxies (accompanied with heteroskedasticity-corrected standard errors) from individual OLS regressions of those returns on the proxy. Tables 1B and 2B describe the maximum correlation between the proxy return and the market return that still supports the CAPM result according to Proposition 1. Recall that all three tests are based on the inequality restriction in (11). If $\rho < 1$, then a test of this restriction requires a prior belief on the true value of the correlation ρ_o. From Roll (1977), $\rho_o = 0.90$ or above. This value will be used throughout the discussions of the test results.

For the GRS test, if $\rho < 1$, then a test of (17) also requires the true value of θ_p. Possible values for θ_p are taken from Shanken (1987). $\theta_p = 0.52$ is the most likely (or expected) value. $\theta_p = 0.22$ and $\theta_p = 0.86$ are - 1 standard deviations and + 2 standard deviations away from this expected value, respectively. All values of θ_p are annualized for presentation but are expressed in weekly terms when used in the tests. Assuming an annual standard deviation of 20% for the proxy return, $\theta_p = 0.22$ corresponds to a "market" premium of 4.4%, $\theta_p = 0.52$ a "market" premium of 10.4%, and $\theta_p = 0.86$ a "market" premium of 17.2%. This range for θ_p is sufficiently wide to encompass the point estimates for θ_p implied by the different subperiods considered. Finally, $\theta_p = 1.00$ is also reported as a value for the proxy Sharpe ratio that is greater than any conceivable true value.

Since (17) requires θ_p, for comparative purposes θ_p is also treated as known in (33) for the BPC and BPS tests. In addition, however, θ_p is also treated as stochastic, meaning that its value is bootstrapped along with every other random quantity in $\widehat{\xi}$. In the tables, the heading "stochastic" under Panel F: BPC and Panel G: BPS details the results of this more general treatment. Finally, both BPC and BPS are implemented using the test methodology outlined in section 5. The only difference in implementation is that under BPC, $\widehat{\gamma}_{io}$ and $\widehat{\delta}_{io}$ are obtained from OLS regressions while under BPS, these parameter estimates are obtained from (31).

Under Tables 1B and 2B, note that (1) the projection errors appear to be non-normal, characterized by (at times) significant skewness and (often times) excess kurtosis, and (2) there exist apparent differences between the projection and structural errors. These two findings foreshadow differences between the GRS, BPC, and BPS tests. Also under Tables 1B and 2B, a comparison of the implied correlations for constant values of θ_p between the GRS and BPC tests reveals, in general, higher correlations implied by the former. As an example, for the period 10/6/67 - 9/25/87 at $\theta_p = 0.52$, $\overline{\rho} = 0.86$ according to GRS but $\overline{\rho} = 0.74$ according to BPC. For the period 11/6/87 - 9/28/07, the same comparison yields $\overline{\rho} = 1.00$ according to GRS as opposed to $\overline{\rho} = 0.93$ according to BPC. This latter comparison possesses economic significance since the former cannot reject mean-variance efficiency of the proxy (see Corollary 1), while the latter can. These results suggest that GRS tends to under-reject in the presence of conditional heteroskedasticity. MacKinlay and Richardson (1991) document similar empirical findings.

A comparison of the implied correlations for constant values of θ_p between the BPC and BPS tests yields a validation of Corollary 2. Implied correlations are generally higher under the latter, with this result being particularly apparent for the B/M and Momentum portfolios in the more recent period. A similar observation (though, on a more muted scale) can be made when comparing the implied correlations between BPC and BPS given a stochastic θ_p.

The most general test results occur under the stochastic θ_p for BPS. For the time period described by Table 1B, $\widehat{\theta}_p = 0.22$, while for the time period described by Table 2B, $\widehat{\theta}_p \approx 0.52$. Comparing the implied correlations under GRS for each of these two time periods given the corresponding estimated value for θ_p to the correlations under BPS given a stochastic θ_p illustrates the effects of (1) conditional heteroskedasticity, (2) omitted components from the market return, and (3) randomizing θ_p. The combination of these effects leads GRS to over-reject relative to BPS in the sense that GRS implies a lower correlation, at times drastically so (see the corresponding results for the B/M and Momentum portfolios under Table 1B).

The BPS test given a stochastic θ_p cannot reject the null hypothesis that the proxy return is mean-variance efficient for size portfolios in the most-recent 20-year period. Otherwise, the test results do not speak favorably for the CAPM. If $\rho_o = 0.90$, then the result of Proposition 1 is rejected for all remaining time periods and portfolios. The CAPM fares decidedly worse on B/M and momentum portfolios relative to size portfolios and performs the poorest on momentum portfolios. A potential bright-spot emerges when comparing results between the two time periods. Given this comparison, the size of the CAPM errors for all the portfolios considered seems greatly reduced in the most-recent period, since the implied correlations very nearly double. Campbell, Lettau, Malkiel, and Xu (2001) document a significant increase in firm-level volatility relative to market volatility over the period 1962 to 1997. In contrast, this paper documents a significant increase in the ability of the CAPM to price expected returns over a similar period.

As a robustness check, the GRS, BPC, and BPS tests are also applied to three 10-year subperiods: (1) 10/7/77 - 9/25/87, (2) 11/6/87 - 9/26/97, (3) 10/3/97 - 9/28/07.[12] Tables 3 (A and B) through 5 (A and B) summarize the results. These results are largely consistent with those for the two 20-year subperiods discussed above. Namely, for constant values of θ_p, GRS implies higher correlations than BPC; BPC implies lower correlations than BPS, and GRS when evaluated at $\theta_p = \widehat{\theta}_p$ implies lower correlations than BPS evaluated with a stochastic θ_p. Also of note, for the subperiod 11/6/87 - 9/26/87, GRS with $\theta_p = 0.86$ rejects Proposition 1 for B/M portfolios, while BPS with a stochastic θ_p does not. This latter result provides some evidence in favor of the CAPM's ability to price B/M portfolios. In addition, for the subperiod 10/3/97 - 9/28/07, the GRS test with $\theta_p = 0.22$ fails to reject mean-variance efficiency of the proxy return with respect to either the size or B/M portfolios. The BPS test with a stochastic θ_p, on the other hand, rejects mean-variance efficiency of the proxy return with respect to either set of portfolios, and while it does not reject Proposition 1 (and, hence, the CAPM) for the size portfolios since $\overline{\rho} = 0.97$, it does reject Proposition 1 for the B/M portfolios, since $\overline{\rho} = 0.66$.

[12]The period 10/6/67 - 9/30/77 is not considered because the mean of the proxy return is negative.

7. An Extension

This section generalizes Proposition 1 in terms of conditional moment restrictions and, in doing so, links mismeasurement of the market return to time-variation in "beta." In order to develop this generalization, moments for period t conditional on S_{t-1} are labeled with a t subscript as are parameters conditional on S_{t-1}. Consider the following conditional pricing model

$$E_t\left[r_t\right] = Cov_t\left[m_t,\ r_t\right], \tag{37}$$

and assume

$$m_t = \left(\frac{E_t\left[r_{mt}\right]}{\sigma_t^2\left[r_{mt}\right]}\right) r_{mt} \tag{38}$$

so that (37) represents a conditional statement of the CAPM.[13] In addition, assume that

$$\beta = \frac{Cov_t\left[r_{mt},\ r_t\right]}{\sigma_t^2\left[r_{mt}\right]} \tag{39}$$

so that market betas are constant parameters and time variation in expected security returns are driven by changes in the market risk premium. Ferson (1990) asserts that the specification of constant betas "is an important assumption in the context of models with conditional expectations" (p.399).[14]

Consider the following generalization of (5):

$$r_t = \gamma_t + \delta r_{pt} + \widetilde{e}_t \tag{40}$$

where

$$\gamma_t = \alpha + \beta E_t\left[\phi_t\right], \qquad \widetilde{\phi}_t = \phi_t - E_t\left[\phi_t\right].$$

A case for $Cov_t\left[\phi_t,\ r_{pt}\right] \neq 0$ follows the same logic outlined in section 2. (40) affords a general specification for the time-varying mean of security returns.[15] This time variation is linked to time variation in both the expected proxy return and the expected value of the components omitted from that proxy return. In the special case where $\phi_t = \phi_c$, the source of this time variation is limited to the expected proxy return.

Next, consider a linear projection of m_t onto r_{pt} conditional on S_{t-1}:

$$m_t = a + b_t r_{pt} + e_{mt}, \tag{41}$$

[13] Harvey (1989), Bodurtha and Mark (1991), Jagannathan and Wang (1996), Lettau and Ludvigson (2001), Adrian and Franzoni (2004), and Ang and Chen (2007) all consider versions of the CAPM in this form.

[14] In nearly all cases, a conditionally mean-variance efficient portfolio will exist, implying that so too will a single beta model for expected returns. In general, the beta from this model will be time-varying.

[15] There is a consensus in the literature that expected returns are time-varying conditional on a set of forecasting instruments. Potential instruments include (i) lagged values of the proxy return to capture reversion as evidenced in Keim and Stambaugh (1986) and Fama and French (1989) among others, (ii) the term spread as measured by the difference between the 10-year and 3-month yields and advocated by Fama and French (1989), (iii) Moody's BAA - AAA credit spread (see, e.g, Campbell (1996), and (iv) the value spread as measured by the return difference between value and growth stocks (see Campbell and Vuolteenaho (2004)).

where

$$b_t = \frac{Cov_t\left[r_{pt},\ m_t\right]}{\sigma_t^2\left[r_{pt}\right]}.$$

Assume that the correlation between m_t and r_{pt} is constant or, equivalently, that the relative efficiency of the proxy return is constant.[16] Then, a straightforward generalization of Lemma 1 in Appendix A grants that

$$Cov_t\left[\widetilde{e}_t,\ e'_{mt}\right]\Sigma_{\widetilde{e}_t}^{-1}Cov_t\left[\widetilde{e}_t,\ e_{mt}\right] \leq \sigma_t^2\left[m_t\right]\left(1-\rho^2\right) \tag{42}$$

where $\Sigma_{\widetilde{e}_t}^{-1}$ is the variance-covariance matrix of $\widetilde{e}_t$ conditional on S_{t-1}. Given (42), an equally straightforward generalization of Proposition 1 can be stated as

Proposition 3 *Let the pricing model of (37) and (38) hold for all security returns including the proxy return, and consider the structural relationship between security returns and the proxy return as given by (40). Define*

$$\theta_{pt} = \frac{E_t\left[r_{pt}\right]}{\sigma_t\left[r_{pt}\right]}$$

and

$$\eta_t = \frac{Cov_t\left[\widetilde{e}_t,\ r_{pt}\right]}{\sigma_t\left[r_{pt}\right]}.$$

Then,

$$d'\Sigma_{\widetilde{e}_t}^{-1}d \leq \theta_{pt}^2(\rho^{-2}-1) \tag{43}$$

where

$$d = E_t\left[r_t\right] - \left(\delta+\eta_t\right)E_t\left[r_{pt}\right].$$

Proof. See the proof of Proposition 1 in Appendix A, and condition the moments contained therein on S_{t-1}. ■

The deviation vector d from Proposition 2 is an N-vector of constant terms from the following model of r_t:

$$r_t = d + \left(\delta+\eta_t\right)E_t\left[r_{pt}\right] + u_t \tag{44}$$

where $E_t\left[u_t\right] = 0$.[17] From (52) and (53), it follows that the vector of time-varying beta proxies $\beta_{pt} = \delta + \eta_t$. (44) relates time-varying expected security returns to time-varying beta proxies and a time-varying expected proxy return. Proposition 2 establishes an upper bound on deviations from conditional CAPM pricing measured with respect to a proxy return. Given a time path for the Sharpe performance measure of the proxy return, this upper bound is set in terms of the efficiency of that proxy return relative to the market return. Suppose $\phi_t = \phi_c$. Then (40) is a projection of r_t onto $E_t\left[\phi_t\right]$ and r_{pt}. In this case, beta proxies are not time varying since $\eta_t = 0$, and time variation in expected security returns are the result of a time-varying expected proxy return.

[16]Conditioning the right-hand-side of (49) on S_{t-1} and substituting the expression for $\sigma_t\left[m_t\right]$ from (38) produces this latter result.

[17]Noting that $\frac{Cov_t\left[r_t,\ r_{pt}\right]}{\sigma_t^2\left[r_{pt}\right]} = \delta + \eta_t$ given (40), (44) is a vector statement of (4) in Bodurtha and Mark (1991).

Works by Harvey (1989), Bodurtha and Mark (1991), and more recently Adrian and Franzoni (2004) and Ang and Chen (2007) consider time-varying betas for the CAPM. Adrian and Franzoni (2004) and Ang and Chen (2007) stress time-varying betas as meaningful contributors to the improved performance of conditional specifications of the CAPM relative to their unconditional counterparts. By definition, all of these works measure time-variation in betas with respect to a proxy return. In the context of Proposition 3, the finding of Adrian and Franzoni (2004) and Ang and Chen (2007) can be interpreted as supporting evidence of a nonzero covariance between ϕ_t and r_{pt}.

Like its unconditional counterpart, Proposition 3 provides an explanation for the empirical discovery of "significant" alphas that does not invalidate the CAPM theory. In addition, Proposition 3 provides an explanation for the significance of time-varying betas in pricing expected security returns. Like Proposition 1, a principal strength behind Proposition 3 is that with the exception of ρ, all of the quantities in (43) can be directly estimated from observable data. This fact sets up an indirect test of the conditional CAPM in analogous terms to those described in sections 2 and 5. Of course, the set S_{t-1} needs to be specified, as does the relationship of this set to expected proxy returns and the expected value of components omitted from that proxy return. Future research will look to empirically test the conditional pricing restriction of Proposition 3, comparing its performance as measured by $\overline{\rho}$ to (1) the unconditional pricing restriction of Proposition 1 and (2) alternative pricing models like the three-factor model of Fama and French (1992), which can be readily stated in the terms of Proposition 2 in Shanken (1987).

8. Conclusion

This paper develops a new test of the CAPM that accounts for a proxy's mismeasurement of the market return. For a given collection of test assets, conventional tests of the CAPM based on the relative mean-variance efficiency of a given proxy estimate the linear relationship between the returns on those test assets and the return on the proxy by a projection of the former onto the latter. This paper demonstrates that estimating such a projection equation is not without loss of generality. The returns to nontraded assets and the returns to human capital are omitted from common "market"-based proxies. The extent to which these returns correlate with a given proxy will determine the extent to which innovations to the linear equation describing returns to the test assets conditional on the proxy return will tend to covary with the proxy. The resulting structural equation will necessarily differ from the projection equation. A novel estimator is proposed for this structural equation that does not require outside instruments. This estimator is then used to show that the proposed test of relative mean-variance efficiency built upon the aforementioned structural equation differs in economically significant ways from competing tests based upon the projection equation.

Appendix A

Lemma 1 *Consider the structural model in (5) and the linear projection in (7). Then*

$$Cov\left[\widetilde{e}_t,\ e_{mt}\right]' \Sigma_{\widetilde{e}}^{-1} Cov\left[\widetilde{e}_t,\ e_{mt}\right] \leq \sigma^2\left[m_t\right]\left(1-\rho^2\right)$$

where $\Sigma_{\widetilde{e}}$ is the $N \times N$ covariance matrix of $\widetilde{e}_t$, and ρ is the correlation between m_t and r_{pt}.

Proof. Since (7) describes a linear projection of m_t onto r_{pt}, $b = \frac{Cov\left[r_{pt},\ m_t\right]}{\sigma^2\left[r_{pt}\right]}$ and $\sigma^2\left[e_{mt}\right] = \sigma^2\left[m_t\right](1-\rho^2)$. Consider regressing e_{mt} on $\widetilde{e}_t$. The explained variance from that regression is $Cov\left[\widetilde{e}_t,\ e_{mt}\right]' \Sigma_{\widetilde{e}}^{-1} Cov\left[\widetilde{e}_t,\ e_{mt}\right]$, which cannot be greater than $\sigma^2\left[m_t\right](1-\rho^2)$, the total variance of e_{mt}. ■

Proof of Proposition 1 Substitution of (5) into the right-hand-side of (1) produces

$$Cov\left[r_t,\ m_t\right] = \delta Cov\left[r_{pt},\ m_t\right] + Cov\left[\widetilde{e}_t, m_t\right]. \tag{45}$$

Given (7),

$$Cov\left[\widetilde{e}_t,\ m_t\right] = \left(\frac{Cov\left[\widetilde{e}_t,\ r_{pt}\right]}{\sigma^2\left[r_{pt}\right]}\right) Cov\left[r_{pt},\ m_t\right] + Cov\left[\widetilde{e}_t,\ e_{mt}\right]. \tag{46}$$

Combining (45) and (46) produces

$$Cov\left[r_t,\ m_t\right] = (\delta+\eta)\, Cov\left[r_{pt},\ m_t\right] + Cov\left[\widetilde{e}_t,\ e_{mt}\right],$$

where η is defined in (10). Substituting the result into (1) grants the following inequality,

$$d'\Sigma_{\widetilde{e}}^{-1} d \leq \sigma^2\left[m_t\right]\left(1-\rho^2\right) \tag{47}$$

where

$$d = E\left[r_t\right] - (\delta+\eta)\, Cov\left[r_{pt},\ m_t\right]. \tag{48}$$

Next, note that given (7) and (9),

$$\sigma^2\left[m_t\right] = \theta_p^2 + \sigma^2\left[e_{mt}\right].$$

The coefficient of determination from (7) is, therefore, $\frac{\theta_p^2}{\sigma^2\left[m_t\right]}$. Recall that $\rho = \frac{Cov\left[m_t,\ r_{pt}\right]}{\sigma\left[m_t\right]\sigma\left[r_{pt}\right]}$. Given that (1) also holds for the proxy return,

$$\rho = \frac{E\left[r_{pt}\right]}{\sigma\left[m_t\right]\sigma\left[r_{pt}\right]} = \frac{\theta_p}{\sigma\left[m_t\right]}. \tag{49}$$

As a result, the right-hand-side of (47) equals $\theta_p^2(\rho^{-2}-1)$. Finally, (48) can be redefined as $E\left[r_t\right] - (\delta+\eta)\, E\left[r_{pt}\right]$.■

Proof of Corollary 1 Given (2) and (49), $\rho = \frac{\theta_p}{\theta_m}$. Substituting (4) into this result produces

$$\rho = \frac{\phi_p E\left[r_{pt}\right]}{\phi_c + \phi_p E\left[r_{pt}\right]}, \tag{50}$$

from which follows the statement that $\rho = 1$ if and only if $\phi_c = 0$. If $\rho = 1$, then $d'\Sigma_{\widetilde{e}_t}^{-1} d = 0$ in (11). Since $\phi_c = 0$, $\eta = 0$ in (10), and $d = E_t\left[r_t\right] - \delta E_t\left[r_{pt}\right]$. From (5) then follows that $d = \alpha$.■

Proof of Corollary 2 Let

$$r_t = \alpha_p + \beta_p r_{pt} + e_{pt} \tag{51}$$

be a multivariate linear projection of r_t onto r_{pt}, where α_p is an alpha proxy, β_p is a beta proxy, and e_{pt} is a projection error. Then

$$\begin{aligned} \alpha_p &= E\left[r_t\right] - \beta_p E\left[r_{pt}\right] \\ \beta_p &= \frac{Cov\left[r_t,\, r_{pt}\right]}{\sigma^2\left[r_{pt}\right]} \end{aligned} \tag{52}$$

Substitution of (5) into the expression for β_p yields the following relationships between the parameters in (51) and the structural parameters in (5):

$$\begin{aligned} \alpha_p &= \gamma - \eta E\left[r_{pt}\right] \\ \beta_p &= \delta + \eta \end{aligned} \tag{53}$$

where η is defined by (10). Given these relationships,

$$e_{pt} = r_t - \alpha_p - \beta_p r_{pt} = \widetilde{e}_t - \eta \widetilde{r}_{pt}$$

where $\widetilde{r}_{pt} = r_{pt} - E\left[r_{pt}\right]$. It then follows that

$$\Sigma_{e_p} = \Sigma_{\widetilde{e}} - \frac{Cov\left[\widetilde{e}_t,\, \widetilde{r}_{pt}\right] Cov\left[\widetilde{e}_t,\, \widetilde{r}_{pt}\right]'}{\sigma^2\left[\widetilde{r}_{pt}\right]} \tag{54}$$

since given the definition of $\widetilde{r}_{pt}$, $Cov\left[\widetilde{e}_t,\, r_{pt}\right] = Cov\left[\widetilde{e}_t,\, \widetilde{r}_{pt}\right]$ and $\sigma^2\left[r_{pt}\right] = \sigma^2\left[\widetilde{r}_{pt}\right]$. Substitution of the expression for $\widetilde{e}_t$ in (6) into (54) produces

$$\Sigma_{\widetilde{e}} - \Sigma_{e_p} = \left(\frac{Cov\left[\widetilde{\phi}_t,\, \widetilde{r}_{pt}\right]}{\sigma\left[\widetilde{r}_{pt}\right]}\right)^2 \beta\beta'. \tag{55}$$

In general, there exists an x such that $\beta' x = 0$. Let $y = \beta' x$. Then $y'y \geq 0$.■

Appendix B

Identification of the Triangular System

Assumption A1: *$E[X_t X_t']$ and $E[X_t Y_t']$ are finite and identified from the data. $E[X_t X_t']$ is nonsingular.*

Assumption A2: *(i) In (22), the eigenvalues of $A_o + B_o$ are less than one in modulus. (ii) Let $a_{ij,o}$ be the element in the ith row and jth column of the matrix A_o, and similarly define $b_{ij,o}$. $a_{33,o} + b_{33,o} \neq a_{22,o} + b_{22,o}$.*

A1 identifies the reduced form residuals from (18) and (19) as

$$R_{i,t} = Y_{i,t} - X_t' E[X_t X_t']^{-1} E[X_t Y_{i,t}], \quad i = 1, 2.$$

A2(i) defines h_t in (22) (or, equivalently, H_t in (21)) as mean stationary according to Proposition 2.7 of Engle and Kroner (1995). A2(ii) preserves the off-diagonal structure of the reduced form GARCH model necessary for identification as illustrated by (25).

(22) implies that

$$e_t = h_t + \omega_t$$

where $E[\omega_t \mid S_{t-1}] = 0$ and $E[\omega_t \omega_s' \mid S_{t-1}] = 0 \; \forall \; s \neq t$. Let $\overline{e}_t = \begin{bmatrix} \epsilon_{1,t}\epsilon_{2,t} & \epsilon_{2,t}^2 \end{bmatrix}'$, and similarly define $\overline{h}_t$ and $\overline{\omega}_t$ such that

$$\overline{e}_t = \overline{h}_t + \overline{\omega}_t.$$

Assumption A3: *(i) $E[\overline{\omega}_t \overline{\omega}_t'] = \Sigma_{\overline{\omega}} < \infty$. (ii) Define $Cov[\overline{e}_t, \overline{e}_{t-1}] \equiv E\left[(\overline{e}_t - \sigma_e)(\overline{e}_{t-1} - \sigma_e)'\right]$. $Cov[\overline{e}_t, \overline{e}_{t-1}]$ is nonsingular if either $a_{11,1o}$ or $b_{11,1o}$ is nonzero.*

Given A3(i), $\overline{\omega}_t$ is covariance stationary. Lemma 2 demonstrates that A2(i) and A3(i) together determine $\overline{e}_t$ to be covariance stationary. Note that if $a_{11,1o} = b_{11,1o} = 0$, then $Cov(\overline{e}_t, Z_{t-1})$ is singular.

Lemma 2 *Given A2(i) and A3(i), $\overline{e}_t$ is covariance stationary.*

Proof. Given the definitions of $\overline{e}_t$ and $\overline{h}_t$, it follows from (22) that

$$\overline{h}_t = \overline{C}_o + \overline{A}_o \overline{e}_{t-1} + \overline{B}_o \overline{h}_{t-1}. \tag{56}$$

Recall that $\overline{A}_o$ is a 2×2 diagonal matrix formed from the elements $a_{22,o}$ and $a_{33,o}$ in A_o (see (22)) and similarly for $\overline{B}_o$. Recursive substitution into (56) produces

$$\overline{h}_t = \sum_{i=1}^{\infty} \overline{B}_o^{i-1} \left(\overline{C}_o + \overline{A}_o \overline{e}_{t-i}\right). \tag{57}$$

Following the steps outlined in the proof to Proposition 2.7 of Engle and Kroner (1995), (57) can be used to show that

$$E_{t-\tau}\overline{e}_t = \left[I + \left(\overline{A}_o + \overline{B}_o\right) + \cdots + \left(\overline{A}_o + \overline{B}_o\right)^{\tau-2}\right] \overline{C}_o + \left(\overline{A}_o + \overline{B}_o\right)^{\tau-1} \sum_{i=1}^{\infty} \overline{B}_o^{i-1} \left(\overline{C}_o + \overline{A}_o \overline{e}_{t-i-\tau+1}\right)$$

where $E_{t-\tau}$ is the expectations operator conditional on the information set $S_{t-\tau}$. For a square matrix Z, it is well known that $Z^{\tau} \to 0$ as $\tau \to \infty$ if and only if the eigenvalues of Z are less than one in modulus. This same condition grants $(I + Z + \cdots + Z^{\tau-1}) \to (I - Z)^{-1}$ as $\tau \to \infty$ for the appropriately sized identity matrix I. Given A3(i), therefore, $E_{t-\tau}\overline{e}_t \xrightarrow{p} \left[I - \left(\overline{A}_o + \overline{B}_o\right)\right]^{-1} \overline{C}_o$(as $\tau \to \infty$).

Since $\overline{e}_t = \overline{h}_t + \overline{\omega}_t$, where $E\left[\overline{\omega}_t \mid S_{t-1}\right] = 0$, given A4(i),

$$E\left[\overline{e}_t\overline{e}_t^{'}\right] = E\left[\overline{h}_t\overline{h}_t^{'}\right] + \Sigma_{\overline{\omega}}.$$

Let $\overline{\sigma}_o = \left[I - \left(\overline{A}_o + \overline{B}_o\right)\right]^{-1} \overline{C}_o$.

$$\begin{aligned} E\left[\overline{h}_t\overline{h}_t^{'}\right] &= \eta_o + \overline{A}_o E\left[\overline{h}_{t-1}\overline{h}_{t-1}^{'}\right]\overline{A}_o^{'} + \overline{A}_o\Sigma_{\overline{\omega}}\overline{A}_o^{'} + \overline{A}_o E\left[\overline{h}_{t-1}\overline{h}_{t-1}^{'}\right]\overline{B}_o^{'} \\ &\quad + \overline{B}_o E\left[\overline{h}_{t-1}\overline{h}_{t-1}^{'}\right]\overline{A}_o^{'} + \overline{B}_o E\left[\overline{h}_{t-1}\overline{h}_{t-1}^{'}\right]\overline{B}_o^{'} \end{aligned} \tag{58}$$

where $\eta_o = \overline{C}_o\overline{C}_o^{'} + \left(\overline{A}_o + \overline{B}_o\right)\overline{\sigma}_o\overline{C}_o^{'} + \overline{C}_o\overline{\sigma}_o^{'}\left(\overline{A}_o + \overline{B}_o\right)^{'}$. Applying the $vec(\cdot)$ operator to (58) and simplifying yields

$$\begin{aligned} vec\left(E\left[\overline{h}_t\overline{h}_t^{'}\right]\right) &= \eta_o + (D_o)\, vec\left(E\left[\overline{h}_{t-1}\overline{h}_{t-1}^{'}\right]\right) + \left(\overline{A}_o \otimes \overline{A}_o\right) vec\left(\Sigma_{\overline{\omega}}\right) \\ &= [I + D_o]\left(\eta_o + \left(\overline{A}_o \otimes \overline{A}_o\right) vec\left(\Sigma_{\overline{\omega}}\right)\right) + \left(D_o^2\right) vec\left(E\left[\overline{h}_{t-2}\overline{h}_{t-2}^{'}\right]\right) \\ &= \left[I + D_o + D_o^2\right]\left(\eta_o + \left(\overline{A}_o \otimes \overline{A}_o\right) vec\left(\Sigma_{\overline{\omega}}\right)\right) + \left(D_o^3\right) vec\left(E\left[\overline{h}_{t-3}\overline{h}_{t-3}^{'}\right]\right) \\ &= \ldots \\ &= \left[I + D_o + \cdots + D_o^{\tau-1}\right]\left(\eta_o + \left(\overline{A}_o \otimes \overline{A}_o\right) vec\left(\Sigma_{\overline{\omega}}\right)\right) + \left(D_o^{\tau}\right) vec\left(E\left[\overline{h}_{t-\tau}\overline{h}_{t-\tau}^{'}\right]\right) \end{aligned}$$

where $D_o = \left(\overline{A}_o + \overline{B}_o\right) \otimes \left(\overline{A}_o + \overline{B}_o\right)$. Given A3(i), the eigenvalues of D_o are less than one in modulus, granting that $vec\left(E\left[\overline{h}_t\overline{h}_t^{'}\right]\right)$ converges to $[I - D_o]^{-1}\left(\eta_o + \left(\overline{A}_o \otimes \overline{A}_o\right) vec\left(\Sigma_{\overline{\omega}}\right)\right)$ as $\tau \to \infty$.

Note that

$$Cov\left(\overline{e}_t,\, \overline{e}_{t-\tau}\right) = E\left[\overline{e}_t\overline{e}_{t-\tau}^{'}\right] - \overline{\sigma}_o\overline{\sigma}_o^{'}$$

Consider the case where $\tau = 1$.

$$E\left[\overline{e}_t\overline{e}_{t-1}^{'} \mid S_{t-1}\right] = \overline{C}_o\overline{e}_{t-1}^{'} + \overline{A}_o\overline{e}_{t-1}\overline{e}_{t-1}^{'} + \overline{B}_o\overline{h}_{t-1}\overline{e}_{t-1}^{'}.$$

By iterated expectations,

$$E\left[\overline{e}_t\overline{e}_{t-1}^{'}\right] = \overline{C}_o\overline{\sigma}_o^{'} + \left(\overline{A}_o + \overline{B}_o\right)\Sigma_{\overline{h}} + \overline{A}_o\Sigma_{\overline{\omega}}$$

and, as a result,

$$Cov\left(\overline{e}_t,\, \overline{e}_{t-1}\right) = \left(\overline{C}_o - \overline{\sigma}_o\right)\overline{\sigma}_o^{'} + \left(\overline{A}_o + \overline{B}_o\right)\Sigma_{\overline{h}} + \overline{A}_o\Sigma_{\overline{\omega}}$$

where $\Sigma_{\overline{h}} = E\left[\overline{h}_t \overline{h}_t^{'}\right]$. Next, consider the case where $\tau \geq 2$.

$$\begin{aligned}
E\left[\overline{h}_t \mid S_{t-\tau}\right] &= E\left[\overline{C}_o + \overline{A}_o \overline{e}_{t-1} + \overline{B}_o \overline{h}_{t-1} \mid S_{t-\tau}\right] \\
&= \overline{C}_o + \left(\overline{A}_o + \overline{B}_o\right) E\left[\overline{h}_{t-1} \mid S_{t-\tau}\right] \\
&= \left[I + \left(\overline{A}_o + \overline{B}_o\right)\right] \overline{C}_o + \left(\overline{A}_o + \overline{B}_o\right)^2 E\left[\overline{h}_{t-2} \mid S_{t-\tau}\right] \\
&= \ldots \\
&= \left[I + \left(\overline{A}_o + \overline{B}_o\right) + \ldots + \left(\overline{A}_o + \overline{B}_o\right)^{\tau-1}\right] \overline{C}_o + \left(\overline{A}_o + \overline{B}_o\right)^{\tau-1} \left[\overline{A}_o \overline{e}_{t-\tau} + \overline{B}_o \overline{h}_{t-\tau}\right] \\
&= \left[I - \left(\overline{A}_o + \overline{B}_o\right)^{\tau}\right] \overline{\sigma}_o + \left(\overline{A}_o + \overline{B}_o\right)^{\tau-1} \left[\overline{A}_o \overline{e}_{t-\tau} + \overline{B}_o \overline{h}_{t-\tau}\right].
\end{aligned}$$

By iterated expectations,

$$\begin{aligned}
E\left[\overline{e}_t \overline{e}_{t-\tau}^{'}\right] &= E\left[E\left[\overline{e}_t \overline{e}_{t-\tau}^{'} \mid S_{t-\tau}\right]\right] \\
&= E\left[E\left[\overline{h}_t \mid S_{t-\tau}\right] \overline{e}_{t-\tau}^{'}\right] \\
&= \left[I - \left(\overline{A}_o + \overline{B}_o\right)^{\tau}\right] \overline{\sigma}_o \overline{\sigma}_o^{'} + \left(\overline{A}_o + \overline{B}_o\right)^{\tau-1} \left[\left(\overline{A}_o + \overline{B}_o\right) E\left[\overline{h}_{t-\tau} \overline{h}_{t-\tau}^{'}\right] + \overline{A}_o E\left[\overline{\omega}_{t-\tau} \overline{\omega}_{t-\tau}^{'}\right]\right].
\end{aligned}$$

As a result,

$$Cov\left(\overline{e}_t,\ \overline{e}_{t-\tau}\right) = \left(\overline{A}_o + \overline{B}_o\right)^{\tau-1} \left[\left(\overline{A}_o + \overline{B}_o\right)\left(\Sigma_{\overline{h}} - \overline{\sigma}_o \overline{\sigma}_o^{'}\right) + \overline{A}_o \Sigma_{\overline{\omega}}\right] \tag{59}$$

which converges to zero as $\tau \to \infty$, since $\left(\overline{A}_o + \overline{B}_o\right)^{\tau-1} \to 0$ (as $\tau \to \infty$). ■

Proposition 3 *Given A1–A3 for the model of (18) and (19), the structural parameters β_{1o}, β_{2o}, and γ_o are identified.*

Proof. Given A1, $\beta_{2o} = E\left[X_t X_t'\right]^{-1} E\left[X_t Y_{2,t}\right]$. If either $a_{11,1o}$ or $b_{11,1o}$ is nonzero as in A3(ii), then $E\left[\epsilon_{1,t}\epsilon_{2,t} \mid S_{t-1}\right]$ is time-varying. In this case, consider $Cov\left(\overline{e}_t,\ \overline{e}_{t-\tau}\right) = Cov\left(\overline{h}_t,\ \overline{e}_{t-\tau}\right)$. From (59), it follows that

$$Cov\left(\overline{e}_t,\ \overline{e}_{t-\tau}\right) = \left(\overline{A}_o + \overline{B}_o\right)^{\tau-1} Cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right). \tag{60}$$

Given (23), let $\overline{r}_t$ be the reduced form $\overline{e}_t$. Then the reduced form of (60) when $\tau = 2$ is

$$Cov\left(\overline{r}_t,\ \overline{r}_{t-2}\right) = \left(\overline{A}_{ro} + \overline{B}_{ro}\right) Cov\left(\overline{r}_t,\ \overline{r}_{t-1}\right), \tag{61}$$

where the relationship between $\overline{A}_{r0}$ and A_{r0} in (24) is equivalent to the relationship between $\overline{A}_o$ in (56) and A_o in (22). An analogous relationship exists between $\overline{B}_{ro}$ and B_{ro}. Identification of $\overline{A}_{ro} + \overline{B}_{ro}$ follows from the nonsingularity of $Cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right)$. γ_o is then identified as (27).

Next, consider the case where $a_{11,1o} = b_{11,1o} = 0$. Define $Z_{t-1} = \left[\begin{array}{ccc} \epsilon_{2,t-1}^2 & \cdots & \epsilon_{2,t-l}^2 \end{array}\right]'$ for finite $l \geq 1$. Since $E\left[\epsilon_{1,t}\epsilon_{2,t} \mid S_{t-1}\right] = c_{21,0o}c_{22,0o}$, it follows that

$$Cov(\epsilon_{1,t}\epsilon_{2,t},\ Z_{t-1}) = 0. \tag{62}$$

From (23), $\epsilon_{1,t} = R_{1,t} - R_{2,t}\gamma_o$ and $R_{2,t} = \epsilon_{2,t}$. Substitution of these results into (62) produces

$$Cov(R_{1,t}\epsilon_{2,t},\ Z_{t-1}) = Cov(\epsilon_{2,t}^2,\ Z_{t-1})\gamma_o.$$

Let $\Omega = Cov(\epsilon_{2,t}^2,\ Z_{t-1})$, and note that $\Omega \neq 0$ given (21). Then γ_o is identified as $\gamma_o = (\Omega'\Omega)^{-1}\Omega' Cov(R_{1,t}\epsilon_{2,t},\ Z_{t-1})$.

Finally, given identification of γ_o, β_{1o} is identified as $\beta_{1o} = E\left[X_t X_t'\right]^{-1} E\left[X_t\left(Y_{1,t} - Y_{2,t}\gamma_o\right)\right]$. ■

Proposition 3 identifies (18) and (19) given the nuisance parameters in $\overline{C}_o$ and $\overline{\Phi}_o$. A complete treatment of (21) is not necessary to identify the triangular model. Note that the moment conditions in (31) cover both the case where A3(ii) holds as well as the case where $a_{11,1o} = b_{11,1o} = 0$.

Corollary 3 From (29),

$$E\left[\overline{e}_t - \overline{\sigma}_o\right] = 0. \tag{63}$$

From (30),

$$E\left[\left(\overline{e}_t - \overline{\sigma}_o\right)\left(\overline{e}_{t-2} - \overline{\sigma}_o\right)' - \overline{\Phi}_o\left(\overline{e}_t - \overline{\sigma}_o\right)\left(\overline{e}_{t-1} - \overline{\sigma}_o\right)'\right] = 0. \tag{64}$$

Stack the moments of (28), (63), and $vec\left((64)\right)$ into a single vector U. Let

$$\psi = \{\gamma_1,\ \gamma_2,\ \delta,\ c_{21},\ c_{22},\ \phi_{11},\ \phi_{22}\} \in \Psi,$$

and define ψ_o as the true value of ψ. Then $E\left[U\right] = 0$ is uniquely satisfied at $\psi = \psi_o$.

Proof. (28) identifies the reduced form residuals $R_{i,t}$, $i = 1, 2$. Given (23), substituting these residuals into (63) and (64) produces (61). The result that $E\left[U\right] = 0$ is uniquely satisfied at $\psi = \psi_o$ then follows from Proposition 3. ■

References

[1] Adrian, T. and F. Franzoni (2004), "Learning about Beta: Time-Varying Factor Loadings, Expected Returns, and the Conditional CAPM," *Federal Reserve Bank of New York Staff Reports*, no. 193.

[2] Ang, A. and J. Chen (2007), "CAPM over the Long Run: 1926–2001," *Journal of Empirical Finance*, 14, 1-40.

[3] Bodurtha, J.N. and N.C. Mark (1991), "Testing the CAPM with Time-Varying Risks and Returns," *Journal of Finance*, 46, 1485-1505.

[4] Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

[5] Breeden, D.T. (1979) "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 265-296.

[6] Campbell, J.Y. (1996), "Understanding Risk and Return," *Journal of Political Economy*, 104, 298–345.

[7] Campbell, J.Y M. Lettau, B.G. Malkiel and Y. Xu "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk," *Journal of Finance*, 56, 1-43.

[8] Campbell, J.Y. and T. Vuolteenaho (2004), "Bad Beta, Good Beta," *American Economic Review*, 94, 1249-1275.

[9] Cochrane, J. (2001), *Asset Pricing*, Princeton University Press, Princeton, NJ.

[10] Dittmar, R.F. (2002), "Nonlinear Pricing Kernels, Kurtosis Preferences, and Evidence from the Cross Section of Equity Returns," *Journal of Finance*, 57, 369-403.

[11] Drost, F.C. and T.E. Nijman (1993), "Temporal Aggregation of GARCH Processes," *Econometrica*, 61, 909-927.

[12] Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with estimates of the Variance of UK Inflation," *Econometrica*, 50, 987-1008.

[13] Engle, R.F and K.F. Kroner (1995), "Multivariate Simultaneous Generalized GARCH," *Econometric Theory*, 11, 121-150.

[14] Fama, E.F. (1965), "The Behavior of Stock Market Prices," *Journal of Business*, 38, 34-105.

[15] Fama, E.F. and K.R. French (1989), "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics*, 25, 23-49.

[16] Ferson, W.E. (1990), "Are the Latent Variables in Time-Varying Expected Returns Compensation for Consumption Risk?" *Journal of Finance*, 45, 397–429.

[17] Gibbons, M.R., S.A. Ross and J. Shanken (1989), "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57, 1121–1152.

[18] Hall, P. and J.L Horowitz (1996), "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators," *Econometrica*, 64, 891-916.

[19] Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

[20] Hansen, L. and R. Jagannathan (1997), "Assessing Specification Errors in Stochastic Discount Factor Models," *Econometrica*, 55, 587-614.

[21] Harvey, C.R. (1989), "Time-Varying Conditional Covariances in Tests of Asset Pricing Models," *Journal of Financial Economics,* 24, 289-317.

[22] Iglesias, E.M and G.D.A. Phillips (2004), "Simultaneous Equations and Weak Instruments under Conditionally Heteroskedastic Disturbances," unpublished manuscript.

[23] Jagannathan, R. and Z. Wang (1996), "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance*, 51, 3-53.

[24] Kandel, S. and R.F. Stambaugh (1987), "On Correlations and Inferences About Mean-Variance Efficiency," *Journal of Financial Economics*, 18, 61–90.

[25] Keim, D. and R.F. Stambaugh (1986), "Predicting Returns in Stock and Bond Markets," *Journal of Financial Economics*, 17, 357-390.

[26] Lettau, M. and S. Ludvigson (2001), "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying," *Journal of Political Economy*, 109, 1238-1287.

[27] Lintner, J. (1965), "The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, 47, 13–37.

[28] MacKinlay, A.C., and M.P. Richardson (1991), "Using Generalized Method of Moments to Test Mean-Variance Efficiency," *The Journal of Finance*, 46, 511–527.

[29] MacKinnon, J. (2007), "Bootstrap Hypothesis Testing," *Queen's Economics Department Working Paper*, no. 1127.

[30] Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394-419.

[31] Milhoj, A. (1985), "The Moment Structure of ARCH Processes," *Scandinavian Journal of Statistics*, 12, 281-292.

[32] Newey, W.K. and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics, Vol. 4*, ed. by R. Engle and D. McFadden, North Holland: Amsterdam, 2113–2247.

[33] Prono, T. (2008), "GARCH-Based Identification and Estimation of Triangular Systems," *Federal Reserve Bank of Boston QAU Working Paper*, no. 08-04.

[34] Roll, R. (1977), "A Critique of the Asset Pricing Theory's Tests," *Journal of Financial Economics*, 4, 129–176.

[35] Ross, S.A. (1976), "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341-360.

[36] Shanken, J. (1987), "Multivariate Proxies and Asset Pricing Relations: Living with the Roll Critique," *Journal of Financial Economics*, 18, 91–110.

[37] Sharpe, W.F. (1964), "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, 19, 425–442.

[38] Sharpe, W.F. (1977), "The CAPM: A Multi-Beta Interpretation," in Levy H. and M. Sarnat (eds.), *Financial Decision Management Under Uncertainty*, NY, 127-135.

Table 1A

Summary statistics for size, B/M, and momentum portfolios, 10/6/67 - 9/25/87. The portfolio return series are measured weekly (in percentage terms) so that relatively high frequency data is utilized (to estimate higher moments) that reduces day-of-the-week and weekend effects as well as the effects of nonsynchronus trading and bid-ask bounce. The proxy return is the CRSP market-value-weighted index of all securities on the NYSE, AMEX, and NASDAQ exchanges. Security returns are constructed from the 25 size-B/M portfolios and the 25 size-momentum portfolios (each 5×5 sorts with breakpoints determined by NYSE quintiles). "Small" is the average of the five low-market-cap portfolios, "Mid" the average of the five medium-market cap portfolios, and "Big" the average of the five large-market-cap portfolios. "Value" is the average of the five high-B/M portfolios, "Neutral" the average of the five middle-B/M portfolios, and "Growth" the average of the five low-B/M portfolios. Finally, "Losers" is the average of the five low-return-sorted portfolios, "Neutral" the average of the five middle-return sorted portfolios, and "Winners" the average of the five high-return-sorted portfolios.

	Size			B/M			Momentum		
	Small	Mid	Large	Value	Neutral	Growth	Losers	Draws	Growth
Panel A: Excess returns									
mean	0.103	0.120	0.075	0.169	0.108	0.023	-0.034	0.101	0.198
stdev	2.31	2.14	2.03	2.08	1.96	2.64	2.72	1.99	2.44
skew	-0.33	-0.20	0.11	-0.21	-0.18	-0.19	0.32	-0.10	-0.58
kurt	5.12	4.80	4.83	5.18	4.70	4.71	5.84	5.22	4.95
Panel B: Alpha Proxy									
est	0.045	0.060	0.016	0.113	0.053	-0.053	-0.106	0.045	0.132
std error[a]	0.039	0.024	0.012	0.028	0.021	0.025	0.038	0.020	0.031

Notes:
[a]Heteroskedasticity consistent

Table 1B

Test results for size, B/M, and momentum portfolios, 10/6/67 - 9/25/87. Projection errors are the residuals from OLS regressions of security returns on the proxy return. Structural errors are the residuals from linear equations relating security returns to the proxy return, where the residual from each equation and the proxy return is allowed to covary. Gibbons, Ross, and Shanken (1989), or GRS, Bootstrap Proposition 1 constant covariance (BPC), and Bootstrap Proposition 1 stochastic (BPS), are alternative ways of determining the maximum correlation between the CRSP value-weighted proxy return and the market return that supports the CAPM at a 5% significance level. GRS is based on the assumption that the projection errors are normally distributed. BPC is also based on the projection errors, but assumes those errors to follow strong, univariate GARCH(1,1) processes with unknown distributions. BPS is based on the assumption that the structural errors follow strong, univariate GARCH (1,1) processes with unknown distributions.

	Size			B/M			Momentum		
	Small	Mid	Large	Value	Neutral	Growth	Losers	Draws	Winners
Panel C: Projection errors									
skew	0.23	0.22	0.09	0.39	0.48	0.11	0.98	0.39	-0.70
kurt	5.08	5.85	4.15	5.76	8.19	4.45	7.95	5.32	6.83
Panel D: Structural errors									
skew	0.17	0.23	0.00	0.40	0.47	0.10	0.59	0.30	-0.85
kurt	5.67	5.81	5.68	5.99	8.15	4.41	5.97	4.99	6.89
Panel E: GRS[b, c]									
Proxy Sharpe ratio:									
0.22		0.565			0.277			0.236	
0.52		0.855			0.570			0.504	
0.86		0.939			0.753			0.694	
1.00		0.953			0.800			0.746	
Panel F: BPC[b, c]									
Proxy Sharpe ratio:									
0.22		0.410			0.272			0.260	
0.52		0.737			0.551			0.530	
0.86		0.892			0.737			0.715	
1.00		0.924			0.787			0.766	
stochastic		0.581			0.429			0.398	
Panel G: BPS[b, c]									
Proxy Sharpe ratio:									
0.22		0.409			0.271			0.242	
0.52		0.741			0.552			0.511	
0.86		0.896			0.738			0.706	
1.00		0.927			0.787			0.761	
stochastic		0.590			0.440			0.388	

Notes:

[b]Reported correlations do not reject a one-sided test of the true correlation exceeding the reported value at a 5% significance level.

[c]Values for the proxy Sharpe ratio are taken from Shanken (1987). These values are annualized. 0.52 is the expected value; 0.22 and 0.86 are - 1 and + 2 standard deviations away from this expected value, respectively. 1.00 is a value for the proxy Sharpe ratio that is greater than any conceivable true value. Stochastic means that the proxy Sharpe ratio is bootstrapped along with every other estimated quantity in the expression determining an upper bound for the correlation between the proxy and the market return.

Table 2A

Summary statistics for size, B/M, and momentum portfolios, 11/6/87 - 9/28/07. The portfolio return series are measured weekly (in percentage terms) so that relatively high frequency data is utilized (to estimate higher moments) that reduces day-of-the-week and weekend effects as well as the effects of nonsynchronus trading and bid-ask bounce. The proxy return is the CRSP market-value-weighted index of all securities on the NYSE, AMEX, and NASDAQ exchanges. Security returns are constructed from the 25 size-B/M portfolios and the 25 size-momentum portfolios (each 5×5 sorts with breakpoints determined by NYSE quintiles). "Small" is the average of the five low-market-cap portfolios, "Mid" the average of the five medium-market cap portfolios, and "Big" the average of the five large-market-cap portfolios. "Value" is the average of the five high-B/M portfolios, "Neutral" the average of the five middle-B/M portfolios, and "Growth" the average of the five low-B/M portfolios. Finally, "Losers" is the average of the five low-return-sorted portfolios, "Neutral" the average of the five middle-return sorted portfolios, and "Winners" the average of the five high-return-sorted portfolios.

	Size			B/M			Momentum		
	Small	Mid	Large	Value	Neutral	Growth	Losers	Draws	Growth
Panel A: Excess returns									
mean	0.169	0.176	0.157	0.209	0.183	0.103	0.041	0.172	0.315
stdev	2.16	2.09	1.92	1.88	1.83	2.61	2.94	1.74	2.60
skew	-1.04	-0.64	-0.29	-0.91	-0.75	-0.75	0.03	-0.60	-0.89
kurt	11.76	6.64	4.97	9.24	6.66	9.05	7.00	6.72	10.96
Panel B: Alpha Proxy									
est	0.049	0.040	0.027	0.094	0.063	-0.070	-0.129	0.059	0.148
std error[a]	0.044	0.029	0.023	0.032	0.025	0.034	0.056	0.025	0.039

Notes:
[a]Heteroskedasticity consistent

Table 2B

Test results for size, B/M, and momentum portfolios, 11/6/87 - 9/28/07. Projection errors are the residuals from OLS regressions of security returns on the proxy return. Structural errors are the residuals from linear equations relating security returns to the proxy return, where the residual from each equation and the proxy return is allowed to covary. Gibbons, Ross, and Shanken (1989), or GRS, Bootstrap Proposition 1 constant covariance (BPC), and Bootstrap Proposition 1 stochastic (BPS), are alternative ways of determining the maximum correlation between the CRSP value-weighted proxy return and the market return that supports the CAPM at a 5% significance level. GRS is based on the assumption that the projection errors are normally distributed. BPC is also based on the projection errors, but assumes those errors to follow strong, univariate GARCH(1,1) processes with unknown distributions. BPS is based on the assumption that the structural errors follow strong, univariate GARCH (1,1) processes with unknown distributions.

	Size			B/M			Momentum		
	Small	Mid	Large	Value	Neutral	Growth	Losers	Draws	Winners
Panel C: Projection errors									
skew	-0.04	0.14	1.41	-0.08	0.17	-0.07	1.13	0.89	-0.25
kurt	7.56	6.06	19.54	6.68	6.70	7.71	9.22	11.49	6.83
Panel D: Structural errors									
skew	-0.43	-0.33	-0.07	-0.75	-0.63	-0.28	1.07	-0.36	-0.72
kurt	9.72	6.68	5.05	10.20	6.59	8.36	9.24	7.39	10.96
Panel E: GRS[b, c]									
Proxy Sharpe ratio:									
0.22		1.000			0.406			0.258	
0.52		1.000			0.730			0.540	
0.86		1.000			0.870			0.727	
1.00		1.000			0.899			0.776	
Panel F: BPC[b, c]									
Proxy Sharpe ratio:									
0.22		0.639			0.397			0.277	
0.52		0.926			0.698			0.554	
0.86		0.997			0.852			0.745	
1.00		1.000			0.888			0.796	
stochastic		1.000			0.824			0.679	
Panel G: BPS[b, c]									
Proxy Sharpe ratio:									
0.22		0.857			0.442			0.319	
0.52		1.000			0.739			0.622	
0.86		1.000			0.878			0.805	
1.00		1.000			0.908			0.851	
stochastic		1.000			0.841			0.745	

Notes:

[b]Reported correlations do not reject a one-sided test of the true correlation exceeding the reported value at a 5% significance level.

[c]Values for the proxy Sharpe ratio are taken from Shanken (1987). These values are annualized. 0.52 is the expected value; 0.22 and 0.86 are - 1 and + 2 standard deviations away from this expected value, respectively. 1.00 is a value for the proxy Sharpe ratio that is greater than any conceivable true value. Stochastic means that the proxy Sharpe ratio is bootstrapped along with every other estimated quantity in the expression determining an upper bound for the correlation between the proxy and the market return.

Table 3A

Summary statistics for size, B/M, and momentum portfolios, 10/7/77 - 9/25/87. The portfolio return series are measured weekly (in percentage terms) so that relatively high frequency data is utilized (to estimate higher moments) that reduces day-of-the-week and weekend effects as well as the effects of nonsynchronus trading and bid-ask bounce. The proxy return is the CRSP market-value-weighted index of all securities on the NYSE, AMEX, and NASDAQ exchanges. Security returns are constructed from the 25 size-B/M portfolios and the 25 size-momentum portfolios (each 5×5 sorts with breakpoints determined by NYSE quintiles). "Small" is the average of the five low-market-cap portfolios, "Mid" the average of the five medium-market cap portfolios, and "Big" the average of the five large-market-cap portfolios. "Value" is the average of the five high-B/M portfolios, "Neutral" the average of the five middle-B/M portfolios, and "Growth" the average of the five low-B/M portfolios. Finally, "Losers" is the average of the five low-return-sorted portfolios, "Neutral" the average of the five middle-return sorted portfolios, and "Winners" the average of the five high-return-sorted portfolios.

	Size			B/M			Momentum		
	Small	Mid	Large	Value	Neutral	Growth	Losers	Draws	Growth
Panel A: Excess returns									
mean	0.221	0.226	0.151	0.257	0.205	0.140	0.077	0.184	0.316
stdev	1.96	1.98	1.99	1.79	1.83	2.49	2.31	1.82	2.43
skew	-1.08	-0.52	0.04	-0.91	-0.54	-0.30	0.36	-0.39	-0.87
kurt	7.96	5.99	4.51	7.81	5.59	4.91	5.72	5.61	6.59
Panel B: Alpha Proxy									
est	0.091	0.081	-0.002	0.130	0.068	-0.045	-0.081	0.048	0.141
std error[a]	0.048	0.031	0.016	0.035	0.024	0.034	0.048	0.025	0.044

Notes:
[a]Heteroskedasticity consistent

Table 3B

Test results for size, B/M, and momentum portfolios, 10/7/77 - 9/25/87. Projection errors are the residuals from OLS regressions of security returns on the proxy return. Structural errors are the residuals from linear equations relating security returns to the proxy return, where the residual from each equation and the proxy return is allowed to covary. Gibbons, Ross, and Shanken (1989), or GRS, Bootstrap Proposition 1 constant covariance (BPC), and Bootstrap Proposition 1 stochastic (BPS), are alternative ways of determining the maximum correlation between the CRSP value-weighted proxy return and the market return that supports the CAPM at a 5% significance level. GRS is based on the assumption that the projection errors are normally distributed. BPC is also based on the projection errors, but assumes those errors to follow strong, univariate GARCH(1,1) processes with unknown distributions. BPS is based on the assumption that the structural errors follow strong, univariate GARCH (1,1) processes with unknown distributions.

	Size			B/M			Momentum		
	Small	Mid	Large	Value	Neutral	Growth	Losers	Draws	Winners
Panel C: Projection errors									
skew	-0.71	-0.35	0.20	-0.77	-0.63	-0.13	0.95	-0.15	-0.97
kurt	7.00	4.90	3.72	7.64	5.57	3.88	8.35	5.06	7.10
Panel D: Structural errors									
skew	-0.77	-0.29	0.20	-1.06	-0.76	-0.15	1.08	-0.11	-0.55
kurt	7.39	4.58	3.69	9.29	6.23	4.03	9.26	4.91	5.50
Panel E: GRS[b, c]									
Proxy Sharpe ratio:									
0.22		0.437			0.244			0.352	
0.52		0.759			0.518			0.671	
0.86		0.888			0.707			0.831	
1.00		0.913			0.758			0.867	
Panel F: BPC[b, c]									
Proxy Sharpe ratio:									
0.22		0.291			0.220			0.286	
0.52		0.590			0.465			0.577	
0.86		0.785			0.660			0.770	
1.00		0.835			0.719			0.821	
stochastic		0.772			0.614			0.733	
Panel G: BPS[b, c]									
Proxy Sharpe ratio:									
0.22		0.353			0.231			0.316	
0.52		0.645			0.483			0.644	
0.86		0.815			0.672			0.849	
1.00		0.872			0.728			0.899	
stochastic		0.810			0.644			0.795	

Notes:

[b]Reported correlations do not reject a one-sided test of the true correlation exceeding the reported value at a 5% significance level.

[c]Values for the proxy Sharpe ratio are taken from Shanken (1987). These values are annualized. 0.52 is the expected value; 0.22 and 0.86 are - 1 and + 2 standard deviations away from this expected value, respectively. 1.00 is a value for the proxy Sharpe ratio that is greater than any conceivable true value. Stochastic means that the proxy Sharpe ratio is bootstrapped along with every other estimated quantity in the expression determining an upper bound for the correlation between the proxy and the market return.

Table 4A

Summary statistics for size, B/M, and momentum portfolios, 11/6/87 - 9/26/97. The portfolio return series are measured weekly (in percentage terms) so that relatively high frequency data is utilized (to estimate higher moments) that reduces day-of-the-week and weekend effects as well as the effects of nonsynchronus trading and bid-ask bounce. The proxy return is the CRSP market-value-weighted index of all securities on the NYSE, AMEX, and NASDAQ exchanges. Security returns are constructed from the 25 size-B/M portfolios and the 25 size-momentum portfolios (each 5×5 sorts with breakpoints determined by NYSE quintiles). "Small" is the average of the five low-market-cap portfolios, "Mid" the average of the five medium-market cap portfolios, and "Big" the average of the five large-market-cap portfolios. "Value" is the average of the five high-B/M portfolios, "Neutral" the average of the five middle-B/M portfolios, and "Growth" the average of the five low-B/M portfolios. Finally, "Losers" is the average of the five low-return-sorted portfolios, "Neutral" the average of the five middle-return sorted portfolios, and "Winners" the average of the five high-return-sorted portfolios.

	Size			B/M			Momentum		
	Small	Mid	Large	Value	Neutral	Growth	Losers	Draws	Growth
Panel A: Excess returns									
mean	0.172	0.220	0.221	0.242	0.217	0.149	0.075	0.207	0.336
stdev	1.54	1.58	1.71	1.47	1.43	1.95	2.05	1.37	1.95
skew	-0.51	-0.44	-0.10	-0.30	-0.50	-0.35	0.15	-0.48	-0.51
kurt	6.51	5.36	4.14	5.26	5.58	4.91	5.95	5.91	4.39
Panel B: Alpha Proxy									
est	0.029	0.039	0.008	0.079	0.047	-0.074	-0.134	0.044	0.112
std error[a]	0.048	0.032	0.017	0.033	0.023	0.040	0.054	0.023	0.038

Notes:
[a]Heteroskedasticity consistent

Table 4B

Test results for size, B/M, and momentum portfolios, 11/6/87 - 9/26/97. Projection errors are the residuals from OLS regressions of security returns on the proxy return. Structural errors are the residuals from linear equations relating security returns to the proxy return, where the residual from each equation and the proxy return is allowed to covary. Gibbons, Ross, and Shanken (1989), or GRS, Bootstrap Proposition 1 constant covariance (BPC), and Bootstrap Proposition 1 stochastic (BPS), are alternative ways of determining the maximum correlation between the CRSP value-weighted proxy return and the market return that supports the CAPM at a 5% significance level. GRS is based on the assumption that the projection errors are normally distributed. BPC is also based on the projection errors, but assumes those errors to follow strong, univariate GARCH(1,1) processes with unknown distributions. BPS is based on the assumption that the structural errors follow strong, univariate GARCH (1,1) processes with unknown distributions.

	Size			B/M			Momentum		
	Small	Mid	Large	Value	Neutral	Growth	Losers	Draws	Winners
Panel C: Projection errors									
skew	0.24	0.09	0.14	0.88	-0.02	-0.16	0.74	0.29	-0.15
kurt	5.07	3.95	3.17	5.99	5.32	3.83	5.64	5.29	3.81
Panel D: Structural errors									
skew	-0.20	-0.08	0.11	0.86	-0.48	-0.16	0.60	-0.27	-0.18
kurt	6.50	4.74	3.25	6.36	6.13	3.69	6.55	7.02	3.95
Panel E: GRS[b, c]									
Proxy Sharpe ratio:									
0.22		1.000			0.388			0.165	
0.52		1.000			0.712			0.372	
0.86		1.000			0.858			0.553	
1.00		1.000			0.889			0.611	
Panel F: BPC[b, c]									
Proxy Sharpe ratio:									
0.22		0.554			0.310			0.174	
0.52		0.892			0.597			0.385	
0.86		1.000			0.780			0.562	
1.00		1.000			0.828			0.619	
stochastic		1.000			0.941			0.714	
Panel G: BPS[b, c]									
Proxy Sharpe ratio:									
0.22		0.636			0.366			0.172	
0.52		0.971			0.661			0.383	
0.86		1.000			0.826			0.562	
1.00		1.000			0.866			0.619	
stochastic		1.000			0.931			0.690	

Notes:

[b]Reported correlations do not reject a one-sided test of the true correlation exceeding the reported value at a 5% significance level.

[c]Values for the proxy Sharpe ratio are taken from Shanken (1987). These values are annualized. 0.52 is the expected value; 0.22 and 0.86 are - 1 and + 2 standard deviations away from this expected value, respectively. 1.00 is a value for the proxy Sharpe ratio that is greater than any conceivable true value. Stochastic means that the proxy Sharpe ratio is bootstrapped along with every other estimated quantity in the expression determining an upper bound for the correlation between the proxy and the market return.

Table 5A

Summary statistics for size, B/M, and momentum portfolios, 10/3/97 - 9/28/07. The portfolio return series are measured weekly (in percentage terms) so that relatively high frequency data is utilized (to estimate higher moments) that reduces day-of-the-week and weekend effects as well as the effects of nonsynchronus trading and bid-ask bounce. The proxy return is the CRSP market-value-weighted index of all securities on the NYSE, AMEX, and NASDAQ exchanges. Security returns are constructed from the 25 size-B/M portfolios and the 25 size-momentum portfolios (each 5×5 sorts with breakpoints determined by NYSE quintiles). "Small" is the average of the five low-market-cap portfolios, "Mid" the average of the five medium-market cap portfolios, and "Big" the average of the five large-market-cap portfolios. "Value" is the average of the five high-B/M portfolios, "Neutral" the average of the five middle-B/M portfolios, and "Growth" the average of the five low-B/M portfolios. Finally, "Losers" is the average of the five low-return-sorted portfolios, "Neutral" the average of the five middle-return sorted portfolios, and "Winners" the average of the five high-return-sorted portfolios.

	Size			B/M			Momentum		
	Small	Mid	Large	Value	Neutral	Growth	Losers	Draws	Growth
Panel A: Excess returns									
mean	0.166	0.133	0.094	0.177	0.150	0.057	0.007	0.138	0.293
stdev	2.63	2.49	2.11	2.22	2.16	3.12	3.61	2.04	3.12
skew	-1.04	-0.60	-0.36	-1.00	-0.74	-0.76	0.03	-0.57	-0.90
kurt	9.85	5.62	5.00	8.53	5.78	7.96	5.51	5.85	9.91
Panel B: Alpha Proxy									
est	0.087	0.048	0.021	0.107	0.077	-0.052	-0.102	0.070	0.190
std error[a]	0.072	0.048	0.038	0.054	0.043	0.054	0.096	0.043	0.067

Notes:
[a]Heteroskedasticity consistent

Table 5B

Test results for size, B/M, and momentum portfolios, 10/3/97 - 9/28/07. Projection errors are the residuals from OLS regressions of security returns on the proxy return. Structural errors are the residuals from linear equations relating security returns to the proxy return, where the residual from each equation and the proxy return is allowed to covary. Gibbons, Ross, and Shanken (1989), or GRS, Bootstrap Proposition 1 constant covariance (BPC), and Bootstrap Proposition 1 stochastic (BPS), are alternative ways of determining the maximum correlation between the CRSP value-weighted proxy return and the market return that supports the CAPM at a 5% significance level. GRS is based on the assumption that the projection errors are normally distributed. BPC is also based on the projection errors, but assumes those errors to follow strong, univariate GARCH(1,1) processes with unknown distributions. BPS is based on the assumption that the structural errors follow strong, univariate GARCH (1,1) processes with unknown distributions.

	Size			B/M			Momentum		
	Small	Mid	Large	Value	Neutral	Growth	Losers	Draws	Winners
Panel C: Projection errors									
skew	-0.11	0.16	1.15	-0.32	0.15	-0.02	1.09	0.81	-0.29
kurt	6.75	5.51	12.23	5.51	5.16	8.02	7.69	8.72	5.49
Panel D: Structural errors									
skew	-0.82	-0.44	-0.18	-1.01	-0.70	-0.46	0.89	-0.46	-0.84
kurt	9.33	5.62	5.29	9.22	5.59	8.06	7.13	6.11	10.07
Panel E: GRS[b, c]									
Proxy Sharpe ratio:									
0.22		1.000			1.000			0.469	
0.52		1.000			1.000			0.788	
0.86		1.000			1.000			0.904	
1.00		1.000			1.000			0.926	
Panel F: BPC[b, c]									
Proxy Sharpe ratio:									
0.22		0.648			0.427			0.337	
0.52		0.955			0.766			0.645	
0.86		1.000			0.928			0.836	
1.00		1.000			0.960			0.883	
stochastic		0.910			0.714			0.579	
Panel G: BPS[b, c]									
Proxy Sharpe ratio:									
0.22		0.771			0.434			0.391	
0.52		1.000			0.798			0.741	
0.86		1.000			0.975			0.933	
1.00		1.000			1.000			0.974	
stochastic		0.966			0.655			0.597	

Notes:

[b]Reported correlations do not reject a one-sided test of the true correlation exceeding the reported value at a 5% significance level.

[c]Values for the proxy Sharpe ratio are taken from Shanken (1987). These values are annualized. 0.52 is the expected value; 0.22 and 0.86 are - 1 and + 2 standard deviations away from this expected value, respectively. 1.00 is a value for the proxy Sharpe ratio that is greater than any conceivable true value. Stochastic means that the proxy Sharpe ratio is bootstrapped along with every other estimated quantity in the expression determining an upper bound for the correlation between the proxy and the market return.