THE RELATIVE EFFICIENCY OF ENDOGENOUS PROXIES: STILL LIVING WITH THE ROLL CRITIQUE

FEDERAL RESERVE BANK OF BOSTON™

Working Paper No. QAU07-2

Revised: March 2008

Todd Prono
Federal Reserve Bank of Boston

FEDERAL RESERVE BANK OF BOSTON™
QAU
WPS No. QAU07-2
Also listed as FRB Boston Research Department WPS 07-01
REVISED

The Relative Efficiency of Endogenous Proxies:

Still Living with the Roll Critique[1]

Todd Prono[2]
Federal Reserve Bank of Boston

First Version: April 2006
This Version: March 2008

Abstract

An asset pricing restriction that permits inference about the familiar CAPM despite the market return being unobservable is generalized to allow an observable proxy of the market return to be endogenously determined along with the individual asset returns the proxy is supposed to price. Making this allowance reduces the efficiency of the proxy relative to the market return by upwards of 20%. Such a reduction is capable of reversing an inference about the validity of the CAPM theory under the aforementioned pricing restriction. Rendering this pricing restriction feasible empirically is a new method for estimating triangular systems given GARCH errors.

Keywords: CAPM, GARCH, GMM, triangular systems, endogeneity, identification, conditional heteroskedasticity. JEL codes: C13, C32, G12.

[1]Earlier versions of this paper were circulated under the titles "GARCH-Based Identification of Triangular Systems with an Application to the CAPM: Still Living with the Roll Critique," "Endogeneity and the CAPM: Revisiting the Problems with Market Proxies" and "GARCH-Based Identification of Triangular Systems, with an Application to the CAPM." Gratitude is owed to Christopher F. Baum, Arthur Lewbel, and Zhijie Xiao, as well as seminar participants at Acadian Asset Management, the Federal Reserve Bank of Boston, and the 2007 Summer Meetings of the Econometric Society for helpful comments and discussions.

[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

The familiar capital asset pricing model (CAPM) pioneered by Sharpe(1964) and Lintner (1965) proposes a linear relation between expected return and the covariance between an asset's return and the return on the wealth or market portfolio, where this covariance measures the asset's systematic risk. Since the market portfolio includes all stores of value (traded or not), its return is unobservable. Empirical investigations of the CAPM theory, therefore, rely on observable proxies to the market return.[3] Clouding the results of these investigations is the complication that an empirical rejection may signal either a failure of the theory or some misspecification of the proxy. Recognizing this complication 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), since the linear relation between risk and return is equivalent to mean-variance efficiency of the market portfolio. Following Roll's critique, measures of relative efficiency were incorporated into tests of the linear risk return relation to account for the location of a chosen proxy in mean-variance space. Kandel and Stambaugh (1987) and Shanken (1987) developed such tests that rely upon a prior belief about the correlation of the proxy with the market portfolio, where this correlation is the relative efficiency measure. Shanken (1987) describes the intuition behind his approach as follows: "if the statistical evidence of the proxy's inefficiency is sufficiently strong, then the inefficiency of the true market return may indeed be correctly inferred and the CAPM rejected" (p. 92). In other words, if linearity between risk and return is only supported by a low correlation between the proxy and the true market portfolio while the prior for that correlation is high (Roll (1977) argues this correlation to be upwards of 0.90), then this finding is interpreted as evidence against the theory.

This paper further explores the effects of proxy portfolios on a test of the CAPM theory by allowing a given proxy to be endogenously determined along with the asset returns it is supposed to price. Since the tests of both Kandel and Stambaugh (1987) and Shanken (1987) reject the hypothesis that the chosen proxy is the market return given that the CAPM is true, the implication is that the proxy is relatively less efficient and, hence, located somewhere inside the minimum-variance boundary that includes the market return.[4] A result of this implication is that the proxy need not be orthogonal to shocks affecting other asset returns. Unanticipated changes to the returns on either nontraded assets or human capital could produce a nonzero relationship.[5] The question addressed in this paper, therefore, is to what extent does endogeneity impact a proxy's inefficiency, and can this impact be material enough

[3] As early examples, see Black, Jensen, and Scholes (1972), and Fama and MacBeth (1973).

[4] Works by Gibbons, Ross, and Shanken (1989), MacKinlay and Richardson (1991) and Zhou (1991) also reject the efficiency of various proxies.

[5] Regarding the latter, Dittmar (2002) recognizes the importance of including human capital within the market portfolio when testing a pricing kernel consistent with a four-moment CAPM.

to alter the conclusion of a CAPM test based upon the degree of that inefficiency? The answer is that endogeneity can reduce the relative efficiency of a proxy by upwards of 20% and that such a reduction can significantly alter the conclusion of a CAPM test conducted in the spirit of Shanken (1987). Conducting this investigation requires an empirical framework that delivers consistent estimates of both the sensitivities of asset returns to the proxy as well as the resulting pricing errors. The development of this framework contributes to the literature on identifying and estimating triangular systems. An overview is provided below.

Consider the following model:

$$Y_{1,t} = X_t'\beta_1 + Y_{2,t}\gamma + \epsilon_{1,t}, \tag{1}$$

$$Y_{2,t} = X_t'\beta_2 + \epsilon_{2,t}. \tag{2}$$

X_t is a vector of predetermined covariates that can include lags of $Y_t = \begin{bmatrix} Y_{1,t} & Y_{2,t} \end{bmatrix}'$. Let $\epsilon_t = \begin{bmatrix} \epsilon_{1,t} & \epsilon_{2,t} \end{bmatrix}'$ be a vector of innovations, and define S_{t-1} to be the sigma field generated by X_t and its past values, as well as past values of ϵ_t. Assume $E\left[\epsilon_t \mid S_{t-1}\right] = 0$. This assumption identifies equation (2) but, of course, is insufficient for identifying equation (1). Typically, identification of equation (1) follows by imposing equality constraints on coefficients of the mean equations–for example, by setting some of the elements in β_1 to zero or, equivalently, by assuming the availability of instruments. This paper introduces a different approach to identifying equation (1) that restricts the covariances of the errors' second moments.

The reduced form of equation (1) is

$$Y_{1,t} = X_t'\overline{\beta}_1 + R_{1,t},$$

where $\overline{\beta}_1 = \beta_1 + \beta_2\gamma$ and $R_{1,t} = \epsilon_{2,t}\gamma + \epsilon_{1,t}$. If $E\left[\epsilon_t\epsilon_t' \mid S_{t-1}\right] = H_t$, where H_t follows a diagonal bivariate GARCH process, then the reduced-form innovations $R_t = \begin{bmatrix} R_{1,t} & \epsilon_{2,t} \end{bmatrix}'$also follow a bivariate GARCH process, but one where the conditional covariance between $R_{1,t}$ and $\epsilon_{2,t}$ depends not only on lags of $R_{1,t}\epsilon_{2,t}$ but also on lags of $\epsilon_{2,t}^2$.[6] The dynamic covariances of $R_{1,t}\epsilon_{2,t}$ and $\epsilon_{2,t}^2$ identify γ. Existence of these auto- and cross-covariances requires $\epsilon_{2,t}$ to be fourth moment stationary. In order to guarantee that H_t be positive definite, it is parameterized by the BEKK representation of a diagonal bivariate GARCH model as proposed by Engle and Kroner (1995).

The key identifying assumption in this paper is the diagonal bivariate GARCH process for ϵ_t. Apparent in the discussion above, while GARCH passes from the structural innovations to the reduced form, diagonality does not, and it is the deviations from diagonality that produce identification. Bollerslev, Engle, and Wooldridge (1988), in their study of a CAPM

[6]The conditional covariance of the structural innovations ϵ_t only depends on lags of $\epsilon_{1,t}\epsilon_{2,t}$.

with time-varying covariances, find the conditional covariance matrix of asset returns to be strongly autoregressive and well described by a diagonal GARCH model. They further note that any "correctly specified intertemporal asset pricing model ought to take this observed heteroskedastic nature of asset returns into account" (p. 123). This paper does precisely that and goes a step further by using that heteroskedastic nature to identify the loadings on a potentially endogenous market proxy.

The role second-moment restrictions play in identification has a long and established history. Early works by Philip Wright (1928) and Sewall Wright (1921) recognize that increases in the variance reduce the bias inherent in simultaneous equations estimated by OLS. More recent contributions include Klein and Vella (2003), who show that a specific semiparametric functional form of multiplicative heteroskedasticity identifies the triangular system. Works by Sentana and Fiorentini (2001), and Lewbel (2004) discuss identification methods that require a constant conditional covariance between the model's structural errors. This paper's methodology generalizes these works by allowing the covariance to display ARMA-style properties. In addition, identification in this paper does not require a specific distributional assumption for $\epsilon_t \mid S_{t-1}$, as in Sentana and Fiorentini (2001).

The remainder of this paper is organized as follows. Section 2 generalizes the pricing restrictions developed by Shanken (1987) to allow for an endogenous proxy. Section 3 develops the assumptions required for GARCH-based identification of triangular systems. Section 4 discusses estimation of this identification result, and Section 5 conducts a Monte Carlo study of the proposed estimator. Section 6 reviews the methodology used to test the pricing restriction of Section 2. The results from applying this methodology are presented in section 7. Section 8 concludes.

2 Pricing Restriction

For expository convenience, all time subscripts are suppressed. In order to facilitate comparison, variable labels follow Shanken (1987) as closely as possible. Assume there exists an observable risk-free rate. Consider the following model:

$$R = E\left[R \mid S\right] + \Gamma P + e, \quad E\left[R \mid S\right] = BX, \tag{3}$$

$$R_p = E\left[R_p \mid S\right] + P, \quad E\left[R_p \mid S\right] = b_p' X. \tag{4}$$

R is an N-vector of asset returns, and R_p is a scalar proxy to the market return. Assume these $N+1$ returns are nonredundant. X is a K-vector of forecasting instruments for asset returns.[7]

[7]As might be expected, the literature on predicting asset returns is long. Potential candidates for X include the dividend yield as argued by Fama and French (1988) and the term premia in Treasury bill returns investigated by Campbell (1987).

Let $E = \begin{bmatrix} e' & P \end{bmatrix}'$ be a vector of innovations, and define S to be the sigma field generated by X and its past values as well as past values of E. Assume $E[P \mid S] = E[e \mid S] = 0$. Equations (3) and (4) divide all asset returns (including the proxy) into predictable and unpredictable components.[8] The unpredictable component of the proxy is the single factor pricing asset returns. Equation (3) is a vector statement of equation (1). If $E[eP] \neq 0$, then equations (3) and (4) specify a triangular system for asset returns.

Let R_m be the return on some unobservable economic aggregate (the definition of which will be made precise below) with the following specification:

$$R_m = E[R_m \mid S] + m, \quad E[R_m \mid S] = b_m' X, \tag{5}$$

which also allows expected returns to be time-varying. Assume $E[m \mid S] = 0$. Consider the linear regression of m on P:

$$m = bP + e_m. \tag{6}$$

Lemma 1 of Shanken (1987) applies immediately to equations (3) and (6) and is summarized for convenience.

Lemma 1 (Shanken): *Consider equations (3) and (6) and their accompanying assumptions. Then,*

$$cov(e,\ e_m)' \Sigma_e^{-1} cov(e,\ e_m) \leq \sigma^2(m)(1 - \rho^2), \tag{7}$$

where Σ_e is the $N \times N$ covariance matrix of e, $\sigma^2(m)$ the variance of m, and ρ the correlation between m and P. Equation (7) holds as an equality if and only if e_m is an exact linear combination of e.

Proof. See the proof of Lemma 1 in Shanken (1987). The fact that security returns are divided into predictable and unpredictable components bears no effect on the result. ■

Lemma 1 (Shanken) makes no explicit use of R_m as an unobservable economic aggregate. According to Shanken (1987), "the lemma is just the statement that the explained variance in the regression of e_m on e is bounded above by the the total variance of e_m; i.e., the r-squared in this regression is at most one" (p. 93). Lemma 1 (Shanken) together with Lemma 1 in the Appendix is used to support the following pricing restriction, which generalizes the results of Propositions 1 and 2 in Shanken (1987) to (i) allow expected returns to be time-varying and (ii) consider P as an endogenous regressor in equation (3).

Corollary 1: *Assume*

$$E[R \mid S] = r1_N + cov(R,\ m) \tag{8}$$

[8] Ferson (1990) considers a multi-factor version of equations (3) and (4).

and

$$E\left[R_p \mid S\right] = r + cov\left(R_p,\ m\right), \tag{9}$$

where 1_N *is an* N*-vector of ones,* and r *is the observable risk free rate. Let* $\sigma^2(P)$ be the variance of P, and define

$$\theta_p^2 = \frac{\left(E\left[R_p \mid S\right] - r\right)^2}{\sigma^2(P)}. \tag{10}$$

Consider

$$\Delta = cov\left(e,\ P\right)\left(\frac{1}{\sigma^2(P)}\right). \tag{11}$$

Then,

$$d'\Sigma_e^{-1}d \leq \theta_p^2(\rho^{-2} - 1), \tag{12}$$

where

$$d \equiv E\left[R \mid S\right] - r1_N - (\Gamma + \Delta)\left(E\left[R_p \mid S\right] - r\right). \tag{13}$$

The pricing restriction in Corollary 1 equates conditional expected excess returns with the covariance between asset returns and innovations to an unobservable economic aggregate. This equality is assumed for both asset returns and the chosen proxy return. Suppose innovations to the economic aggregate are proportional to m^*, innovations to the true market return R_m^*, with a constant of proportionality equal to $\left(E\left[R_m^* \mid S\right] - r\right)/\sigma^2(m^*)$.[9] Then equation (8) states the CAPM in terms of conditional expected excess returns, and $\sigma(m) = \theta_{m^*}$, where

$$\theta_{m^*} = \frac{E\left[R_m^* \mid S\right] - r}{\sigma(m^*)}.$$

The equality between $\sigma(m)$ and θ_{m^*} is important because it links ρ to the market return. In particular,

$$\rho = \frac{\theta_p}{\sigma(m)} = \frac{\theta_p}{\theta_{m^*}},$$

from the proof to Corollary 1.[10] This link also interprets ρ as a measure of relative efficiency for the proxy in the familiar mean-variance space of portfolio returns, if expected returns are not considered to be time-varying; i.e., the conditioning information in S does not forecast asset returns.[11] Under this case, equation (8) states the CAPM in unconditional terms

[9]Assume $E\left[R_m^* \mid S\right] = \left(b_m^*\right)' X$.

[10]Given this equality, ρ is strictly positive.

[11]If returns are time-invariant, then both conditional and unconditional expected returns are equal.

following the same proportionality argument stated above. Since θ_{m^*} is the Sharpe performance measure for the market return (an efficient portfolio), ρ describes the efficiency of the proxy relative to the market. If $\rho = 1$, then the proxy is the market, and the pricing errors in equation (12) are zero. If $\rho < 1$, then the proxy is located somewhere inside the minimum-variance boundary that includes the market return, and the distance between the proxy and the market bounds the pricing errors from above. The larger are those errors the smaller is ρ, meaning the farther away from the market is the proxy in mean-variance space.

Allowing P to be endogenous in equation (3) decomposes the proxy beta (where beta is afforded its standard definition of $cov\,(R,\ P)\left(\frac{1}{\sigma^2(P)}\right)$) into two components: Γ and Δ. Γ measures the sensitivity of asset returns to movements in the proxy (or innovations to the proxy), while Δ measures the relationship between asset returns and components of the market portfolio that are omitted from the proxy. If the proxy is the market (meaning $\rho = 1$), then $\Delta = 0$ because the market return is efficient; i.e., fully diversified. If the proxy differs from the market (i.e., $\rho < 1$), then Δ could be nonzero. For two proxies $R_{p(1)}$ and $R_{p(2)}$, where $\|\Delta_1\| > \|\Delta_2\|$, relative diversification implies that $\rho_1 < \rho_2$. In this case, $R_{p(1)}$ either omits fewer components than $R_{p(2)}$, or the components omitted from $R_{p(2)}$ bear less of a relationship to the assets being priced.[12]

The errors from equation (13) are similar in form to those derived by Diacogiannis and Feldman (2006), who also modify the theoretical CAPM relation for inefficient portfolios. A key difference (to be fully discussed in Section 6) is that all of the parameters in equation (13) can be estimated directly from observable data.

3 Identification

Consider the triangular system

$$Y_{1,t} = X_t'\beta_{10} + Y_{2,t}\gamma_0 + \epsilon_{1,t}, \tag{14}$$

and

$$Y_{2,t} = X_t'\beta_{20} + \epsilon_{2,t}. \tag{15}$$

Let β_{10} refer to the true value of β_1 and similarly for all other parameters. Assume that the regressors in X_t are ordinary random variables with finite second moments. This section provides identification conditions that require neither the errors $\epsilon_{1,t}$ and $\epsilon_{2,t}$ to be uncorrelated nor equality restrictions to be imposed on β_{10}.

[12] This rationale is not intended to be exhaustive but, rather, consistent with the motivation for why proxies should be considered endogenous.

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: $\epsilon_t = \begin{bmatrix} \epsilon_{1,t} & \epsilon_{2,t} \end{bmatrix}'$, and $S_{t-1} = \{X_t, \ldots, X_{t-p}, \epsilon_{t-1}, \ldots, \epsilon_{t-p}\}$ for some finite p. $E\left[\epsilon_t \mid S_{t-1}\right] = 0$ and $E\left[\epsilon_t \epsilon_t' \mid S_{t-1}\right] = H_t$, where

$$H_t = C_0' C_0 + \sum_{k=1}^{2} A_{k0}' \epsilon_{t-1} \epsilon_{t-1}' A_{k0} + \sum_{k=1}^{2} B_{k0}' H_{t-1} B_{k0}, \tag{16}$$

$C_0 = \begin{bmatrix} c_{10} & 0 \\ c_{20} & c_{30} \end{bmatrix}$, $A_{10} = \begin{bmatrix} a_{11,10} & 0 \\ 0 & a_{22,10} \end{bmatrix}$, $A_{20} = \begin{bmatrix} a_{11,20} & 0 \\ 0 & 0 \end{bmatrix}$, $B_{10} = \begin{bmatrix} b_{11,10} & 0 \\ 0 & b_{22,10} \end{bmatrix}$, and $B_{20} = \begin{bmatrix} b_{11,20} & 0 \\ 0 & 0 \end{bmatrix}$. The parameters c_{10}, c_{20}, $a_{22,10}$, $a_{11,20}$, $b_{22,10}$, and $b_{11,20}$ are strictly positive.

A1 and A2 identify the structural form of equation (15) and the reduced form of equation (14). A2 specifies a bivariate analog to the semi-strong GARCH model defined by Drost and Nijman (1993). Equation (16) defines a first-order diagonal bivariate BEKK model as detailed in Proposition 2.3 of Engle and Kroner (1995). This particular GARCH form is chosen because it establishes H_t as positive definite under very mild conditions. Let $\overline{h}_t = \begin{bmatrix} h_{12,t} & h_{22,t} \end{bmatrix}'$ and $\overline{e}_t = \begin{bmatrix} \epsilon_{1,t}\epsilon_{2,t} & \epsilon_{2,t}^2 \end{bmatrix}'$. Given equation (16),

$$\overline{h}_t = \overline{C}_0 + A_0 \overline{e}_{t-1} + B_0 \overline{h}_{t-1}, \tag{17}$$

where $\overline{C}_0 = \begin{bmatrix} c_{20}c_{30} & c_{30}^2 \end{bmatrix}'$,

$$A_0 = \begin{bmatrix} a_{11,10}a_{22,10} & 0 \\ 0 & a_{22,10}^2 \end{bmatrix}, \quad B_0 = \begin{bmatrix} b_{11,10}b_{22,10} & 0 \\ 0 & b_{22,10}^2 \end{bmatrix}. \tag{18}$$

Equation (17) implies that

$$\overline{e}_t = \overline{h}_t + \overline{w}_t, \tag{19}$$

where $E\left[\overline{w}_t \mid S_{t-1}\right] = 0$ by construction.

Assumption A3: $cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right)$ is nonsingular.

The diagonal GARCH specification of A2 and the nonsingularity condition of A3 are both central to identification. A3 establishes the existence of the auto- and cross-covariances for $R_{1,t}\epsilon_{2,t}$ and $\epsilon_{2,t}^2$ mentioned in the Introduction and is similar in scope to Assumption A3 of

Lewbel (2004). Under A3, the errors $\epsilon_{2,t}$ are fourth moment stationary (see Condition 3 below). Weis (1986), in demonstrating consistency and asymptotic normality of the maximum likelihood estimator for univariate ARCH models, requires this same condition.

A3 is a substantive assumption that holds if and only if the following conditions are satisfied.

Condition C1: The eigenvalues of $A_0 + B_0$ are less less than one in modulus.

Condition C2: $\left|a_{11,10}\right| < 1$ and $\left|b_{11,10}\right| < 1$.

Condition C3: $E\left[\overline{w}_t \overline{w}_t'\right] = \Lambda < \infty$.

Condition C4: $a_{11,10}$ and $b_{11,10}$ are nonzero.

C1–C3 are necessary for $\overline{e}_t$ to be covariance stationary (see the statement and proof of Lemma 4 in the Appendix). C1 is similar to Proposition 2.7 of Engle and Kroner (1995) and determines $\overline{h}_t$ to be mean stationary. C2 is necessary because $h_{11,t}$ is not considered in equation (17). If $\overline{h}_t$ were redefined as $\overline{h}_t = \left[\begin{array}{ccc} h_{11,t} & h_{12,t} & h_{22,t} \end{array}\right]'$ and the definitions of $\overline{e}_t$, $\overline{C}_0$, A_0 and B_0 were adjusted accordingly, then C2 would be unnecessary given C1. C3 is a substantive condition that specifies the existence of a finite fourth moment for $\epsilon_{2,t}$. This condition cannot be made more primitive unless additional structure is provided around the innovations $\overline{w}_t$. For example, suppose ϵ_t follows the strong GARCH model of Drost and Nijman (1993) so that

$$\epsilon_t = H_t^{1/2} V_t, \tag{20}$$

where $V_t \sim iid\, D\left(0,\ I_2\right)$ for some distribution D, and I_2 is the 2×2 identity matrix. Then, C3 holds if and only if $ca_{22,10}^4 + 2a_{22,10}^2 b_{22,10}^2 + b_{22,10}^4 < 1$, where $c = E\left[V_{2,t}^4\right]$. If $V_t \sim N\left(0,\ I_2\right)$, then $c = 3$ and the inequality restriction is provided by Theorem 2 in Bollerslev (1986). Finally, C4 ensures $cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right)$ to be of full rank.

From equation (17),

$$h_{12,t} = \left(c_{20}c_{30}\right) + \left(a_{11,10}a_{22,10}\right)\epsilon_{1,t-1}\epsilon_{2,t-1} + \left(b_{11,10}b_{22,10}\right)h_{12,t-1}. \tag{21}$$

C4 echoes the theme of Sentana and Fiorentini (2001), Rigobon (2002), and Lewbel (2004) in the sense that identification of the triangular system relates to a property of the conditional covariance between the error terms. Unlike these authors, however, who demonstrate this property to be time-invariance, C4 takes a different tack and links identification explicitly to time-variation. The special case of a constant conditional covariance is treated later in this section.

Let $\phi_{120} = a_{11,10}a_{22,10} + b_{11,10}b_{22,10}$ and $\phi_{220} = a_{22,10}^2 + b_{22,10}^2$.

Assumption A4: $\phi_{120} \neq \phi_{220}$.

A4 differentiates the structural parameters governing $h_{12,t}$ from the structural parameters governing $h_{22,t}$. Given A2 and A3, $-1 < \phi_{120} < 1$ and $0 < \phi_{220} < 1$.

Proposition 1. *Let A1–A4 hold for the model of equations (14) and (15). The structural parameters of the triangular system are identified.*

Proof. All proofs, unless otherwise stated, are given in the Appendix. ■

Identification under Proposition 1 is a product of the auto- and cross-covariances of $R_{1,t}\epsilon_{2,t}$ and $\epsilon_{2,t}^2$ dictated by the structural form of H_t in equation (16). Given A1–A2, equation (19) can be solved in reduced form, and given A3, the auto- and cross-covariances from that reduced form exist. Equation (55) specifies the relationship between the first and second order of those auto- and cross-covariances as a function of the parameters from the reduced form of equation (17). Inspection of that function in equation (56) reveals identification of γ_0 given A4. According to Proposition 2.1 of Iglesias and Phillips (2004), if the structural form errors of a triangular system follow a bivariate diagonal GARCH process, the reduced form errrors, while still GARCH, will not (except under severe restrictions) also be diagonal. Equation (56) demonstrates how this departure from diagonality supports identification of the triangular system.

Proposition 1 treats the structural parameters of H_t in equation (16) as nuisance parameters and demonstrates that identification of γ_0 follows from the identification of two composite functions of those parameters, ϕ_{120} and ϕ_{220}. Recall that C4 is a necessary condition for A3, requiring the covariance between $\epsilon_{1,t}$ and $\epsilon_{2,t}$ to be time-varying. Relaxing C4 is the subject of the following proposition.

Proposition 2. *Let A1, A2 where $a_{11,10} = b_{11,10} = 0$, C1 and C3 hold for the model of equations (14) and (15). Define $\phi_{220} = a_{22,10}^2 + b_{22,10}^2$. The structural parameters of the triangular system are identified.*

Identification under Proposition 2 drives from $cov(\epsilon_{1,t}\epsilon_{2,t},\ Z_{t-1}) = 0$, where $Z_{t-1} = \left[\begin{array}{ccc} \epsilon_{2,t-1}^2 & \cdots & \epsilon_{2,t-l}^2 \end{array}\right]'$ for finite $l \geq 1$. The reduced form of this zero covariance restriction produces identification as seen in equation (60). Proposition 1 requires both $\epsilon_{1,t}$ and $\epsilon_{2,t}$ to follow GARCH(1,1) processes. Proposition 2, on the other hand, continues to hold if A2 defines some alternative form of conditional heteroskedasticity for $\epsilon_{1,t}$. The identification result of Proposition 2 depends upon a constant conditional covariance and a GARCH(1,1) process for $\epsilon_{2,t}$. The second-moment dynamics of $\epsilon_{1,t}$ do not play a substantiative role. Theorem 1 of Lewbel (2004) reaches the same conclusion.

Since identification under either Proposition 1 or 2 depends critically on the GARCH structure, rationalizing this structure is an important aspect of the discussion. Prono (2006) explores non-diagonal ARCH models that still support identification of the triangular system. Nelson (1992) examines the ability of misspecified ARCH models to consistently estimate the conditional covariance matrix of certain stochastic processes, finding that for processes well approximated by a diffusion without jumps, the multivariate GARCH(1,1) model provides consistent conditional covariance estimates. Nelson and Foster (1994) build upon this result by formulating the conditions under which the univariate GARCH(1,1) model provides asymptotically optimal conditional variance estimates. In this paper, identification follows from a consistent treatment of the conditional covariance matrix. The consistency results of Nelson (1992) link Proposition 1 to a class of continuous time processes commonly employed in modern finance theory. The asymptotic optimality results of Nelson and Foster (1994) apply directly to Proposition 2.

4 Estimation

Consider estimation of equations (14) and (15) under either Proposition 1 or 2. Let $\epsilon_t = \begin{bmatrix} \epsilon_{1,t} & \epsilon_{2,t} \end{bmatrix}'$ and $\overline{e}_t = \begin{bmatrix} \epsilon_{1,t}\epsilon_{2,t} & \epsilon_{2,t}^2 \end{bmatrix}'$. In addition, $\sigma_{\overline{e}} = \begin{bmatrix} \sigma_{12} & \sigma_{22} \end{bmatrix}'$, where $\sigma_{12} = \frac{c_2 c_3}{1-\phi_{12}}$ and $\sigma_{22} = \frac{c_3^2}{1-\phi_{22}}$. The parameter matrices $\overline{\Phi}$ and Z are defined as $\overline{\Phi} = \begin{bmatrix} \phi_{12} & 0 \\ 0 & \phi_{22} \end{bmatrix}$ and $Z = \begin{bmatrix} \zeta_{120} & 0 \\ 0 & \zeta_{220} \end{bmatrix}$, where

$$\zeta_{i20}^2 = E\left[\left(\epsilon_{i,t}\epsilon_{2,t} - \sigma_{i2}\right)^2\right], \quad i = 1, 2. \tag{22}$$

Finally, define ψ as the set of parameters $\{\beta_1,\ \beta_2,\ \gamma,\ c_2,\ c_3,\ \phi_{12},\ \phi_{22}\}$. Consider the following set of vector functions:

$$U_1\left(\psi,\ Y_t,\ S_{t-1}\right) = X_t \otimes \epsilon_t,$$

$$U_2\left(\psi,\ Y_t,\ S_{t-1}\right) = \overline{e}_t - \sigma_{\overline{e}},$$

$$U_3\left(\psi,\ Y_t,\ S_{t-1}\right) = vec\left[Z^{-1}\left[\left(\overline{e}_t - \sigma_{\overline{e}}\right)\left(\overline{e}_{t-2} - \sigma_{\overline{e}}\right)' - \overline{\Phi}\left(\overline{e}_t - \sigma_{\overline{e}}\right)\left(\overline{e}_{t-1} - \sigma_{\overline{e}}\right)'\right] Z^{-1}\right].$$ [13]

[13] The matrix operator $\otimes$ is the kronecker product. The $vec[\cdot]$ operator stacks the columns of an $(m \times n)$ matrix into an $(mn \times 1)$ vector.

Stack U_1, U_2, and U_3 into a single vector U.

Corollary 2. *Let the assumptions underlying either Proposition 1 or 2 hold for the model of equations (14) and (15). Define ϵ_t, $\overline{e}_t$, $\sigma_{\overline{e}}$, ψ and $U\left(\psi,\ Y_t,\ S_{t-1}\right)$ as above. Denote the set of all possible values that ψ might take on as Ψ, and define ψ_0 to be the true value of ψ. The only value of $\psi \in \Psi$ that satisfies $E\left[U\left(\psi,\ Y_t,\ S_{t-1}\right)\right] = 0$ is $\psi = \psi_0$.*

Corollary 2 nests the results of both Propositions 1 and 2 into a single set of moment conditions. $E\left[U_1\right] = 0$ relates to the conditional means of equations (14) and (15), while $E\left[U_2\right] = 0$ defines the unconditional covariance of $\epsilon_{1,t}$ and $\epsilon_{2,t}$ as well as the unconditional variance of $\epsilon_{2,t}$. Given the parameters ζ_{120} and ζ_{220} defined in equation (22), $E\left[U_3\right] = 0$ describes the auto- and cross-correlations implied by equation (16), nesting a zero auto- and cross-correlation for $\epsilon_{1,t}\epsilon_{2,t}$ as a special case.

From Corollary 2, Hansen's (1982) GMM is a natural choice for estimating ψ. The standard GMM estimator is

$$\widehat{\psi} = \arg\min_{\psi \in \Psi} \left(\frac{1}{T}\sum_{t=1}^{T} U\left(\psi,\ Y_t,\ S_{t-1}\right)\right)' \widehat{W}^{-1} \left(\frac{1}{T}\sum_{t=1}^{T} U\left(\psi,\ Y_t,\ S_{t-1}\right)\right). \tag{23}$$

Consistency results for this estimator given dependent data can be found in Newey and McFadden (1994). Standard regulatory conditions require Ψ to be compact, a fact that needs to be reconciled with A4. Suppose $\left|\phi_{12}\right| \leq \phi_{220}$. Then one such reconciliation might be to define Ψ so that the quotient of the final two elements of every $\psi \in \Psi$ is finite and exclusive of an open neighborhood of one. Asymptotic normality of equation (23) requires $\epsilon_{2,t}$ to be eighth-moment stationary. Alternative estimators with potentially better finite sample properties (for example, Generalized Empirical Likelihood) can be used instead of GMM–see Newey and Smith (2004). If any of the moment conditions in equation (23) are weak, then the alternative distribution theory of Stock and Wright (2000) is potentially applicable. Finally, efficiency gains result if the set of vector functions

$$\begin{aligned} U_{3+p}\left(\psi,\ Y_t,\ S_{t-1}\right) &= vec\left[Z^{-1}\left[\left(\overline{e}_t - \sigma_{\overline{e}}\right)\left(\overline{e}_{t-p} - \sigma_{\overline{e}}\right)' - \overline{\Phi}^{p-1}\left(\overline{e}_t - \sigma_{\overline{e}}\right)\left(\overline{e}_{t-1} - \sigma_{\overline{e}}\right)'\right] Z^{-1}\right], \\ p &= 3, \ldots, Q, \end{aligned}$$

is appended to U. West (2002) discusses this result in the context of AR processes with GARCH errors. These functions involve higher-order autocorrelations from the GARCH process.

Based on the proof to Proposition 1, a less efficient estimator can be constructed using conventional time series software through the following steps.

STEP 1: Estimate equation (15) via maximum likelihood, specifying

$$h_{22,t} = \left(c_3^2\right) + \left(\eta_{22,1}\right)\epsilon_{2,t-1}^2 + \left(\eta_{22,2}\right)h_{22,t-1},$$

as the conditional variance of $\epsilon_{2,t}$, to obtain $\widehat{\epsilon}_{2,t}$ and $\widehat{h}_{22,t}$.

STEP 2: Regress $Y_{1,t}$ on X_t to obtain the reduced-form residuals $\widehat{R}_{1,t}$.

STEP 3: Estimate via maximum likelihood

$$\begin{aligned} R_{1,t}\epsilon_{2,t} &= (c_{12}) + \left(\eta_{12,1}+\eta_{12,4}\right)R_{1,t-1}\epsilon_{2,t-1} + \left(\eta_{12,2}\right)\epsilon_{2,t-1}^2 \\ &\quad + \left(\eta_{12,3}\right)h_{22,t-1} - \left(\eta_{12,4}\right)w_{12,rt-1} + w_{12,rt}. \end{aligned} \tag{24}$$

From the resulting parameter estimates, obtain $\widehat{\gamma}$ as

$$\widehat{\gamma} = \frac{\widehat{\eta}_{12,2}+\widehat{\eta}_{12,3}}{\left(\widehat{\eta}_{22,1}+\widehat{\eta}_{22,2}\right)-\left(\widehat{\eta}_{12,1}+\widehat{\eta}_{12,4}\right)}. \tag{25}$$

STEP 4: Estimate $\widehat{\beta}_1$ as $\widehat{\beta}_1 = \left(\sum_{t=1}^{T} X_t X_t'\right)^{-1} \sum_{t=1}^{T} X_t\left(Y_{1,t}-Y_{2,t}\widehat{\gamma}\right)$. Obtaining estimates of the remaining elements in ψ is straightforward given $\widehat{\epsilon}_{1,t}$ and $\widehat{\epsilon}_{2,t}$.

STEP 1 estimates β_{20} as well as $h_{22,t}$ from equation (17), where $\eta_{22,1} = a_{22,1}^2$, and $\eta_{22,2} = b_{22,1}^2$. Note that the reduced-form expression of $h_{12,t}$ in equation (21) is

$$\begin{aligned} h_{12,rt} &= (c_{120}) + \left(\eta_{12,10}\right)R_{1,t-1}\epsilon_{2,t-1} + \left(\eta_{12,20}\right)\epsilon_{2,t-1}^2 \\ &\quad + \left(\eta_{12,30}\right)h_{22,t-1} + \left(\eta_{12,40}\right)h_{12,rt-1}, \end{aligned} \tag{26}$$

where $c_{120} = (c_{20}c_{30} + \gamma_0 c_{30}^2)$, $\eta_{12,10} = \left(a_{11,10}a_{22,10}\right)$, $\eta_{12,20} = \gamma_0\left(\eta_{22,10}-\eta_{12,10}\right)$, $\eta_{12,30} = \gamma_0\left(\eta_{22,20}-\eta_{12,40}\right)$, and $\eta_{12,40} = \left(b_{11,10}b_{22,10}\right)$. Equation (26) can be rewritten as equation (24) by letting $w_{12,rt} = R_{1,t}\epsilon_{2,t} - h_{12,rt}$. The result is an ARMA$(1,1)$ specification with weakly exogenous covariates $\epsilon_{2,t-1}^2$, and $h_{22,t-1}$. Estimation of this specification is feasible given the products of steps 1 and 2. Equation (25) is the finite sample representation of equation (57). Since equation (56) represents the parameters governing the reduced-forms of $h_{12,t}$ and $h_{22,t}$, equation (25) remains valid under the case of a constant conditional covariance where $\eta_{12,10} = \eta_{12,40} = 0$.

The advantage to steps 1–4 is that they are relatively straightforward to implement. The disadvantage is that convergence can be an issue, since the parameters governing the AR and MA components in equation (24) are likely to be of similar magnitudes. Standard errors are also not available for STEP 3 due to the inclusion of generated regressors from STEP 2. If steps 1 and 2 are estimated simultaneously, then robust standard errors for STEP 3 can be calculated using the theory of two-step estimators–see Newey and McFadden (1994).

In practice, this simple procedure can provide consistent starting values for ψ as well as estimates of ζ^2_{i20} in equation (22) for the GMM estimator.

5 Monte Carlo

The Monte Carlo simulations draw data from the structural model of equations (14) and (15) in the case where $\beta_{10} = \beta_{20} = 0$ and $\gamma_0 = 1$. The vector of errors in equation (20) is parameterized according to equation (16) assuming $V_t \sim N(0,\ I_2)$. Values for the individual elements of A_{k0} and B_{k0} in A2 are selected such that the ARCH terms for $h_{11,t}$, $h_{12,t}$, and $h_{22,t}$ are 0.05, 0.04, and 0.10, while the GARCH terms are 0.90, 0.80, and 0.80.[14] Let $C = vech\,(C_0)$, where C_0 is defined in A2. C is specified for three covariance regimes: low, medium, high. $C_l = (0.22,\ 0.05,\ 0.32)$, $C_m = (0.20,\ 0.10,\ 0.32)$, and $C_h = (0.10,\ 0.20,\ 0.32)$. From equation (22), $\zeta^2_{22} = 2.19$ across all three regimes, while $\zeta^2_{12h} = 1.03$, $\zeta^2_{12m} = 0.91$, $\zeta^2_{12l} = 0.90$.[15] For each of the three regimes, simulations are conducted with $Q = 2,\ 4,\ 8,\ 16$ lags and sample sizes $T = 1260,\ 2520$. These two sample sizes reflect daily returns recorded over 5 and 10 years, respectively. The intent of these simulation exercises is to study the finite sample properties of the single step GMM estimator for varying covariance strengths across different lag lengths and sample sizes.

Table 1 shows results for γ of the Monte Carlo simulations across 5000 trials of $T = 1260,\ 2520$ observations, reporting the following robust measures of central tendency and dispersion: median bias (Med. Bias), median absolute error (MDAE), the difference between the 0.1 and 0.9 quantiles (Decile Range).[16] The standard deviation (SD) of γ is also reported. Although not a robust measure, the standard deviation does give an indication of outliers. In order to avoid initialization effects, all simulations discard the first 200 observations. The parameter values for the simulations satisfy A1–A4. For each simulation trial, the estimates for those values are not restricted to ensure that the same assumptions hold. The starting values for each trial, however, are the true values of the parameters.

At $Q = 2$, estimates of γ are accurate across all levels of covariance between $\epsilon_{1,t}$ and $\epsilon_{2,t}$ (hereafter denoted σ_{12}), with a maximum bias of 0.3%. As Q increases, so too does the bias

[14] These values reflect the high GARCH low ARCH specifications typically implied by asset return data. Furthermore, the specified ARCH and GARCH terms satisfy the inequality restriction of Theorem 2 in Bollerslev (1986) for the existence of a fourth moment for both $\epsilon_{1,t}$ and $\epsilon_{2,t}$.

[15] Values for ζ^2_{12} and ζ^2_{22} are determined through simulation. Series of ϵ_t are constructed from equation (20) with 7560 observations across 10000 trials. For each trial, the variance of $\epsilon_{1,t}\epsilon_{2,t}$ and $\epsilon^2_{2,t}$ is estimated. The values for ζ^2_{12} and ζ^2_{22} are determined as the median of these estimates.

[16] Robust measures are used because of concerns over the existence of moments.

in γ, thus supporting the finding of Newey and Smith (2001) regarding the GMM estimator.[17] The level of this bias tends to be small for low levels of σ_{12}. For example, at $\sigma_{12} = 0.1$, the maximum bias is 7%. The size of this bias, however, increases significantly with σ_{12}. At $\sigma_{12} = 0.2$, the maximum bias is approximately 14%, while at $\sigma_{12} = 0.4$ the maximum bias is 30%. In general, the bias in γ increases with σ_{12} for a given Q. However, for low levels of Q, the difference in bias is less as σ_{12} increases and so too is the level of the bias. Irrespective of Q, the size of the bias decreases with T.

Higher values of Q, while tending to be associated with higher biases in γ, also tend to be associated with fairly substantial declines in the dispersion of γ, evidencing the efficiency gains noted by West (2002). At $Q = 2$, $\sigma_{12} = 0.1$, and $T = 1260$, the decile range for γ is approximately 1.3, while the standard deviation is nearly 1.2. At $Q = 16$, for the same values of σ_{12} and T, the decile range drops to 0.32, while the standard deviation falls to 0.13. This same tendency of reduced dispersion can be seen across the different values of σ_{12} and T as Q increases, though on a more muted scale.

The evidenced bias-variance trade-off regarding Q deserves formal attention. Selection criteria for Q in the spirit of Donald, Imbens, and Newey (2002) would benefit the estimator described in this paper.

6 Methodology

Begin by expressing equations (3) and (4) in terms of excess returns over an observable risk free rate r, letting $Y_{i,t}$ be the ith element of $R - r1_N$ and $Y_{p,t} = R_{p,t} - r$. Individual excess asset returns are then related to excess proxy returns by

$$Y_{i,t} = X_t' \left(\delta_i - \beta_m \gamma_i\right) + Y_{P,t}\gamma_i + \epsilon_{i,t}, \quad i = 1, \ldots, N, \tag{27}$$

where

$$Y_{p,t} = X_t'\beta_p + \epsilon_{p,t}. \tag{28}$$

Consider the special case of X_t only containing a constant term.[18] From equation (3), the vector $\widehat{\Gamma}$ can be obtained by sequentially estimating equations (27) and (28) for each i using

[17]These authors demonstrate that a substantial portion of the bias in the GMM estimator specified with many moment conditions (the magnitude of which is potentially quite large) can be attributed to correlations between U in Corollary 2 and the derivative of the moment function.

[18]Forcasting instruments are not included in X_t for three reasons: (i) such instruments are generally unavailable at the high frequencies best suited for the estimator developed in sections 3 and 4 (see the Monte Carlo results from Section 5), (ii) excluding such instruments expedites the process (yet to be described) of bootstrapping a distribution for $\widehat{d}'\widehat{\Sigma}_e^{-1}\widehat{d}$ in equation (12), (iii) specifying expected excess returns as time-invariant links ρ (see Lemma 1 (Shanken)) to unconditional mean-variance space and focuses attention on the effects of Δ (see Corollary 1) on the proxy's relative position with that space.

the single-step GMM estimator with $W = I$ and the moment conditions specified in Corollary 2. In addition, $Q = 5$.[19] Steps 1–4 in section 4 are used to obtained starting values for the GMM estimator. The vector $\widehat{\Delta}$ from equation (11) can be consistently obtained by sequentially regressing $\epsilon_{i,t}$ on $\epsilon_{p,t}$.[20] A key contribution of Corollary 1 is the decomposition of the proxy beta into components Γ and Δ. The first component represents the traditional interpretation of beta (i.e., the sensitivity of individual asset returns to changes in the market, or proxy, return) and, indeed, Γ would equal the proxy beta if $cov\,(e,\ P) = 0$ (see equation (11)). An empirical question is to what extent can Δ be distinguished as a significant contributor to the decomposition? Recognizing that $\widehat{\Delta}$ is the product of a two-step estimator, standard errors are obtained via the bootstrap. Suppose that each pair of innovations $\epsilon_t = \left[\begin{array}{cc} \epsilon_{i,t} & \epsilon_{p,t} \end{array}\right]'$ for $i = 1, \ldots, N$ can be parameterized according to equation (20), where H_t is specified following A2. Then $\widehat{H}_t$ can be obtained by fitting a bivariate BEKK model to $\widehat{\epsilon}_t$, and the bootstrap can be implemented on the individual elements of $\left(\widehat{H}_t^{-1/2}\right)\widehat{\epsilon}_t$.

Let $\overline{\epsilon}_t = \left[\begin{array}{ccc} \epsilon_{1,t} & \ldots & \epsilon_{N,t} \end{array}\right]'$. From equation (13), $\widehat{d}$ are the pricing errors from a cross-sectional GLS regression of $\left(\frac{1}{T}\sum\limits_{t=1}^{T} R_t - r1_N\right)$ on $\left(\widehat{\Gamma} + \widehat{\Delta}\right)$, using $\widehat{\Sigma}_e = \frac{1}{T}\sum\limits_{t=1}^{T}\widehat{\overline{\epsilon}}_t\widehat{\overline{\epsilon}}_t'$ as the error covariance matrix. In order to test the inequality of equation (12) under the special case that $\Delta = 0$ and $\overline{\epsilon}_t$ are homoskedastic, Shanken (1987) employs a noncentral F distribution. Given that identification of Γ and Δ depends on heteroskedastic errors, such a distribution cannot be applied in testing the pricing restriction of Corollary 1. Instead, the bootstrap will again be employed to determine a distribution for $\widehat{d}'\widehat{\Sigma}_e^{-1}\widehat{d}$. The manner for conducting the bootstrap is an extension of the one described above. In particular, repeat the following steps: (i) bootstrap $\widehat{V}_t$, (ii) reconstruct $\widehat{\epsilon}_t$ using $\widehat{H}_t$, (iii) reconstruct $Y_{i,t}$ and $Y_{p,t}$ using $\widehat{\epsilon}_t$ and the parameter estimates from the single-step GMM estimator.

Thus far, the left-hand side of equation (12) is afforded an empirical treatment. If θ_p^2 is known, an upper bound for ρ that satisfies Corollary 1 at a standard 5% significance level can be determined as that value for which $\theta_P^2(\widehat{\rho}^2 - 1)$ lies above 95% of the bootstrapped

[19]The Monte Carlo study reveals that Q should be kept modest in order to limit the bias in $\widehat{\gamma}_i$. $Q = 5$ seems sensible given this finding as well as the fact that the data is measured at a daily frequency.

[20]Let $\widehat{\eta}_i = \sum\limits_{t=1}^{T}\epsilon_{i,t}\epsilon_{p,t}/\sum\limits_{t=1}^{T}\epsilon_{p,t}^2$, the OLS estimate from a regression of$\epsilon_{i,t}$ on $\epsilon_{p,t}$ without a constant term.

$$p\lim\widehat{\eta}_i = \frac{p\lim\left(\frac{1}{T}\sum\limits_{t=1}^{T}\epsilon_{i,t}\epsilon_{p,t}\right)}{p\lim\left(\frac{1}{T}\sum\limits_{t=1}^{T}\epsilon_{p,t}^2\right)} = \frac{cov\left(\epsilon_{i,t},\ \epsilon_{p,t}\right)}{\sigma^2\left(\epsilon_{p,t}\right)},$$

by the Slutsky and Khinchine's theorem, respectively.

distribution for $\widehat{d}'\widehat{\Sigma}_e^{-1}\widehat{d}$. Values of $\rho > \widehat{\rho}$ do not support the inequality of equation (12), and if the prior for $\rho_0 > \widehat{\rho}$, then the pricing restriction of Corollary 1 is rejected. In practice, of course, θ_p^2 is not known. One can use the point estimate $\widehat{\theta}_p^2 = \frac{1}{T}\sum_{t=1}^{T} Y_{P,t}/\widehat{\sigma}^2\left(Y_{p,t}\right)$–where $\widehat{\sigma}^2\left(Y_{P,t}\right) = \widehat{c}_3^2/\left[1-\left(\widehat{a}_{22,1}^2+\widehat{b}_{22,1}^2\right)\right]$ with $\widehat{c}_3$, $\widehat{a}_{22,1}$, and $\widehat{b}_{22,1}$ defined by A2–as the value for θ_p^2. Since there is likely to be substantial variation in $\widehat{\theta}_p^2$, one can also bootstrap a distribution for $\widehat{\theta}_p^2$ following the steps outlined above and use different values from within that distribution to define a range for θ_p^2. The empirical results (to be discussed) rely on this approach.

In order to assess the impact of non-zero covariances between $\epsilon_{i,t}$ and $\epsilon_{p,t}$ on the relative efficiency of the proxy and determine whether this impact can reverse an inference on the pricing restriction of Corollary 1, the method for bootstrapping $\widehat{d}'\widehat{\Sigma}_e^{-1}\widehat{d}$ and finding the value of ρ that satisfies equation (12) for 95% of the resulting distribution is also conducted assuming $\Delta = 0$. In this case, $\widehat{\Gamma}$ is obtained by sequentially estimating equation (27) by OLS. The other elements in equation (12) are estimated in the same manner described above. Bootstrapping the distribution of $\widehat{d}'\widehat{\Sigma}_e^{-1}\widehat{d}$ also occurs in an analogous fashion, except that $h_{12,t} = 0$ in H_t, meaning that $\widehat{H}_t$ can be constructed from univariate GARCH(1,1) models of $\epsilon_{i,t}$ and $\epsilon_{p,t}$. $Y_{i,t}$ and $Y_{p,t}$ can then be reconstructed using $\widehat{\epsilon}_t$ (which assumes a zero covariance) and the parameter estimates from the OLS estimator.

7. Results

From equations (27) and (28), let $Y_{i,t}$ be the daily excess return on the ith capitalization-based decile of NYSE/AMEX/NASDAQ stocks, and define $Y_{p,t}$ as the daily excess return on the value-weighted index of all NYSE/AMEX/NASDAQ stocks.[21] All stock return data is from CRSP. The risk free rate is the one-month Treasury bill rate from Ibbotson Associates. The time period considered is June 1, 1965 through May 31, 2002. This overall period is subdivided twice: first into 3 subperiods of approximately 12 years each; second into 5 subperiods of approximately 6 years each, with the subperiod covering December 2, 1983 through January 1, 1990 excluded from consideration.[22] The length of the subperiods is chosen to closely correspond with the sample sizes utilized in the Monte Carlo experiment. The first three 6-year subperiods correspond to Subperiods 3–5 in Shanken (1987) in date range, though not in return frequency.[23]

[21]The decile portfolios exclude firms for which a market value cannot be computed at the end of each June. The value-weighted index includes all stocks on the Exchanges on any given day and, therefore, is not a redundant asset.

[22]The exclusion of this subperiod results from the inability to obtain an $\widehat{H}_t$ that conforms to A1–A4, presumably because of the '87 crash.

[23]Shanken considers monthly returns.

Table 2 reports point estimates of the ten elements within $\widehat{\Delta}$ along with their bootstrapped standard errors and 95% confidence intervals for the three 12-year subperiods. In all cases, the bootstrap is implemented for 2000 repetitions. The upper and lower bounds of the 95% confidence intervals are the 2.5 and 97.5 percentiles of the bootstrapped distribution for each $\widehat{\delta}$. Significant $\widehat{\delta}s$ are found in all three subperiods. Across the three subperiods, significant $\widehat{\delta}s$ are found for each of the decile portfolios except one. The highest concentration of significant $\widehat{\delta}s$ is found in the first subperiod, where these $\widehat{\delta}s$ tend to be associated with portfolios of smaller- and medium-sized stocks. These findings show that the proxy beta decomposition of Corollary 1 enjoys empirical support.

Table 3 presents the value of ρ that supports the inequality of equation (12) for $(1-\alpha)\,\%$ of the boostrapped distribution of $\widehat{d}'\widehat{\Sigma}_e^{-1}\widehat{d}$ calculated over 2000 repetitions for four values of θ_p^2 for each of the three 12-year subperiods. Values of the proxy beta (or its two part decomposition) within $\widehat{d}$ are estimated by either OLS or single-step GMM. The four values of θ_p^2 are annualized and reported as Sharpe ratios (i.e., $\theta_p = \left(252 \times \text{daily } \theta_p^2\right)^{1/2}$) to facilitate interpretation. The lowest value, $\theta_p = 0.169$, corresponds to the median of the bootstrapped distribution of $\widehat{\theta}_p^2$ and implies an expected excess proxy return of 3.39% on a standard deviation of 20%.[24] The middle two values, $\theta_p = 0.313$ and $\theta_p = 0.493$, correspond to the point estimate of $\widehat{\theta}_p^2$ and the 95th percentile of the bootstrapped distribution of $\widehat{\theta}_p^2$ and imply expected excess proxy returns of 6.25% and 9.87%, respectively (also on a standard deviation of 20%). The upper value, $\theta_p = 1.00$, implies an expected excess proxy return of 20%. This final value is chosen as an extreme upper bound, since the likelihood of θ_p as high as one seems remote. The time period considered for $\widehat{\theta}_p^2$ is 6/1/65–5/31/02. The bootstrap distribution for $\widehat{\theta}_p^2$ is calculated over 10000 repetitions.

Apparent from Table 3, across the three 12-year subperiods, the pricing restriction of Corollary 1 only holds if the proxy is quite inefficient. This inefficiency is reflected in ρ needing to be very low in order for the inequality of equation (12) to be satisfied. For example, at $\alpha = 0.05$, $\widehat{\rho} = 0.073$ for the smallest value of θ_p in the first subperiod. The highest $\widehat{\rho}$ is recorded in the third subperiod at $\widehat{\rho} = 0.486$ when $\alpha = 0.05$ and $\theta_p = 1$. Across all values of θ_p, $\widehat{\rho}$ increases from the first subperiod to the third. For the GMM estimator, $\widehat{\rho}$ is over 20% higher in the third subperiod than in the first, suggesting a reduction in pricing errors between the subperiods. However, if $\rho_0 \geq 0.90$ as suggested by Roll (1977), then Table 3 provides strong evidence against the CAPM.

Of course, the sound rejection of the CAPM on daily data is not a surprising result. The primary intent of this empirical exercise, however, is not to investigate the level of ρ that satisfies equation (12) but, rather, the relative difference in ρ when Δ is allowed to be nonzero.

[24] A 20% standard deviation is consistent with daily return data.

Table 2 demonstrates many instances where nonzero $\widehat{\delta}s$ are found. The question is, what effect do these nonzero $\widehat{\delta}s$ have on a test of Corollary 1? Table 3 lists the values of ρ obtained by using the OLS and single-step GMM estimators on equation (27). Table 4 summarizes the percentage difference in ρ implied by the GMM estimator relative to OLS. These percentage differences are uniformly negative, supporting the assertion of Section 2 that endogeneity of a proxy should decrease its efficiency relative to the market return. In addition, these differences can imply over a 20% reduction in the relative efficiency of the proxy, where the level of these differences is stable across the different values of θ_p. The subperiod during which these approximate 20% reductions occur also registers the highest concentration of nonzero $\widehat{\delta}s$. To help gauge the impact of a 20% reduction in relative efficiency, suppose that for the same proxy measured at a lower frequency of returns, $\widehat{\rho} = 0.90$ ignoring the effects of Δ. If accounting for Δ reduces $\widehat{\rho}$ to 0.72, then the inference from the test is, indeed, very different. In the latter case, the proxy is found to explain only half $(0.72^2 = 0.52)$ of the variance in the market return if the CAPM is true, whereas in the former case, the proxy is erroneously thought to explain over 80% of the variance.

Tables 5A and 5B report point estimates of the ten elements within $\widehat{\Delta}$ along with their bootstrapped standard errors and 95% confidence intervals for the five 6-year subperiods. Significant $\widehat{\delta}s$ are found in each of the five subperiods except one, while across the five subperiods, significant $\widehat{\delta}s$ are found for each of the decile portfolios. In contrast to the three 12-year subperiods, concentrations of significant $\widehat{\delta}s$ are not limited to the earlier subperiods and when such concentrations do occur, they are not exclusive to small- and medium-sized portfolios.

Table 6 presents the value of ρ that supports the inequality of equation (12) for $(1 - \alpha)\,\%$ of the bootstrapped distribution of $\widehat{d}'\widehat{\Sigma}_e^{-1}\widehat{d}$ calculated over 2000 repetitions for four values of θ_p^2 for each of the five 6-year subperiods. As in Table 3, the level of ρ supporting the pricing restriction in Corollary 1 remains quite low. For $\alpha = 0.05$, the minimum value of $\widehat{\rho}$ found using the single-step GMM estimator is $\widehat{\rho} = 0.05$, while the maximum value is $\widehat{\rho} = 0.391$. There remains an increasing trend in $\widehat{\rho}$ from the first subperiod to the fifth, though the size of this trend is decidedly more modest.

Table 7 presents the percentage difference in ρ implied by the GMM estimator relative to OLS for each of the five subperiods. These percentage differences are negative except in the fourth subperiod, where (though positive) they are close to zero. In the second subperiod, an approximate 20% reduction in the relative efficiency of the proxy is recorded that is relatively stable across all values of θ_p. In the fifth subperiod, a reduction of roughly 10% is recorded that is also stable. A $10\% - 20\%$ reduction in the relative efficiency of a proxy can significantly alter the inference from a test of Corollary 1. Even at the lower end of this range, a $\widehat{\rho} = 0.90$ assuming $\Delta = 0$ drops to $\widehat{\rho} = 0.81$, resulting in a decline in the explained variance of the market return by the proxy from 80% to 66%.

8 Conclusion

Starting from the critique offered by Roll (1977), this paper asks whether an inefficient proxy of the unobservable market return is exogenous to the asset returns it is assigned to price. By generalizing a pricing restriction developed by Shanken (1987), this paper investigates the effect of endogeneity on the relative efficiency of a proxy given that the CAPM is true, in order to determine whether the effect is large enough to reverse an inference about the validity of the CAPM theory. Empirical evidence is provided in support of a decomposition of the familiar proxy beta into two parts: one governing the relationship between asset returns and the proxy, the second governing the relationship between asset returns and components to the market return that are omitted from the proxy. The ability to separately identify and estimate these two parts requires a new estimator reliant upon the the GARCH structure of security returns advocated by Bollerslev, Engle, and Wooldridge (1988) as integral to any correctly specified asset pricing model. Relative to the literature on GARCH-based identification, this estimator distinguishes itself by allowing for time-variation in the conditional covariance. A Monte Carlo study verifies the consistency of this estimator and evidences a bias-variance trade off in the number of lagged instruments used. Controlling for the endogeneity of a proxy is found to reduce the relative efficiency of that proxy by upwards of 20%, suggesting that an endogenous proxy can meaningfully impact a test of the CAPM theory. This result extends beyond the CAPM paradigm to any asset pricing model that specifies the market return as a factor and argues for the use of estimators robust to endogeneity of the proxy in evaluating these models. An interesting extension of this paper would be an investigation into how the three-moment CAPM of Kraus and Litzenberger (1976) or the four-moment CAPM of Dittmar (2002) can be estimated given the inefficiency of any proxy return and what that inefficiency implies.

Appendix

Lemma 1 *Assume*

$$E\left[R \mid S\right] = r1_N + cov\left(R,\ m\right), \tag{29}$$

where 1_N is an N-vector of ones, and r is the observable risk-free rate. Then, there exists a "price of risk" λ that satisfies

$$d'\Sigma_e^{-1}d \leq \sigma^2\left(m\right)\left(1-\rho^2\right), \tag{30}$$

where

$$d \equiv E\left[R \mid S\right] - r1_N - \left(\Gamma + \Delta\right)\lambda, \tag{31}$$

and $\lambda = cov\left(P,\ m\right)$ satisfies equation (30).

A.1. Proof of Lemma 1: From equations (3) and (6),

$$cov\left(R,\ m\right) = \Gamma cov\left(P,\ m\right) + cov\left(e,\ m\right), \tag{32}$$

since $E\left[m \mid S\right] = 0$. From the linear regression in equation (6),

$$cov\left(e,\ m\right) = \left[cov\left(e,\ P\right)\left(\frac{1}{\sigma^2\left(P\right)}\right)\right] cov\left(P,\ m\right) + cov\left(e,\ e_m\right), \tag{33}$$

where $b = \frac{cov(P,\ m)}{\sigma^2(P)}$. Given equation (11), combining (32) and (33) yields

$$cov\left(R,\ m\right) = \left(\Gamma + \Delta\right) cov\left(P,\ m\right) + cov\left(e,\ e_m\right). \tag{34}$$

Substitution of (34) into equation (29) and the result into equation (7) of Lemma 1 (Shanken) produces equation (30), with $\lambda = cov\left(P,\ m\right)$.

A.2. Proof of Corollary 1: From equation (4),

$$cov\left(R_p,\ m\right) = cov\left(P,\ m\right),$$

and, therefore,

$$E\left[R_p \mid S\right] - r = cov\left(P,\ m\right). \tag{35}$$

From equation (6),

$$\sigma^2(m) = b^2\sigma^2(P) + \sigma^2(e_m). \tag{36}$$

Given the definition of b in A.1 and equation (35), (36) simplifies to

$$\sigma^2(m) = \theta_p^2 + \sigma^2(e_m).$$

Hence, the coefficient of determination from equation (6) is $\rho^2 = \frac{\theta_p^2}{\sigma^2(m)}$, and equation (30) in Lemma 1 reduces to (12) and (13).

Lemma 2 *Given C1–C2, the eigenvalues of B_0 in equation (18) are less than one in modulus.*

A.3. Proof of Lemma 2: For the matrix B_0, let $b_{ii,0}$ correspond to the ith element along the principal diagonal, and let λ_j be the jth eigenvalue. Consider ordering the eigenvalues so that

$$\lambda_i = b_{ii,0}, \quad i = 1, 2.$$

Suppose $b_{22,10} \geq 1$. Then $a_{22,10}^2 + b_{22,10}^2 > 1$ since $a_{22,10} > 0$, a contradiction. Therefore, $|\lambda_2| < 1$. Next, note that

$$|\lambda_1| = \left|b_{11,10}b_{22,10}\right| = \left|b_{11,10}\right| b_{22,10} < 1. \tag{37}$$

Lemma 3 *Let I_i be an $i \times i$ identity matrix. Given C1–C2, define*

$$M_0 = \{I_4 - [(A_0 + B_0) \otimes B_0]\}^{-1} (I_2 \otimes B_0 A_0)$$

and

$$N_0 = \{I_4 - [B_0 \otimes (A_0 + B_0)]\}^{-1} (B_0 A_0 \otimes I_2).$$

The matrix $M_0 N_0$ has eigenvalues that are less than one.

A.4. Proof of Lemma 3: C1 and Lemma 2 establish the eigenvalues of $(A_0 + B_0) \otimes B_0$ and $B_0 \otimes (A_0 + B_0)$ to be less than one in modulus. Let $a_{ii,0}$ and $b_{ii,0}$ each correspond to the ith element along the principal diagonal of the matrix A_0 and B_0 in equation (18), respectively. For the matrix $M_0 N_0$, define λ_j as the jth eigenvalue, and consider ordering these eigenvalues so that $\lambda_k$corresponds to the kth element along the principal diagonal of $M_0 N_0$. The matrix $M_0 N_0$ has three distinct eigenvalues: λ_1, $\lambda_2 = \lambda_3$, λ_4.

$$\lambda_1 = \left(\frac{a_{11,0} b_{11,0}}{1 - b_{11,0} \left(a_{11,0} + b_{11,0} \right)} \right)^2 . \tag{38}$$

Suppose $\lambda_1 \geq 1$. Then from equation (38) it follows that

$$\left(1 - b_{11,0}^2\right) \left(1 - \left[2a_{11,0} b_{11,0} + b_{11,0}^2\right]\right) \leq 0. \tag{39}$$

However, the left-hand-side of equation (39) is strictly positive since $\left|b_{11,0}\right| < 1$ (see equation (37)) and

$$2a_{11,0} b_{11,0} + b_{11,0}^2 = \left(a_{11,0} + b_{11,0}\right)^2 - a_{11,0}^2 \leq \left(a_{11,0} + b_{11,0}\right)^2 ,$$

a contradiction. Therefore, $\lambda_1 < 1$. A parallel argument grants that $\lambda_4 < 1$. Simply substitute $a_{22,0}$ for $a_{11,0}$ and $b_{22,0}$ for $b_{11,0}$ and follow the same steps outlined above. Next,

$$\lambda_2 = \left(\frac{a_{11,0} b_{11,0}}{1 - b_{11,0} \left(a_{22,0} + b_{22,0} \right)} \right) \left(\frac{a_{22,0} b_{22,0}}{1 - b_{22,0} \left(a_{11,0} + b_{11,0} \right)} \right) . \tag{40}$$

Suppose $\lambda_2 \geq 1$. From equation (40) it follows that

$$\left(1 - b_{11,0} b_{22,0}\right) \left(1 - \left[a_{22,0} b_{11,0} + b_{22,0} \left(a_{11,0} + b_{11,0}\right)\right]\right) \leq 0. \tag{41}$$

Since

$$\left|b_{11,0} b_{22,0}\right| = \left|b_{11,0}\right| b_{22,0} \leq b_{22,0}$$

and

$$\left|a_{22,0} b_{11,0} + b_{22,0} \left(a_{11,0} + b_{11,0}\right)\right| \leq a_{22,0} \left|b_{11,0}\right| + b_{22,0} \left|a_{11,0} + b_{11,0}\right| \leq a_{22,0} + b_{22,0},$$

however, the left-hand-side of equation (41) is strictly positive, a contradiction. Therefore, $\lambda_2 < 1$.

Lemma 4 $\overline{e}_t$ *is covariance stationary if and only if C1–C3 hold.*

A.5. Proof of Lemma 4: Let I_i be an $i \times i$ identity matrix. From equation (17), $[I_2 - (A_0 + B_0)]\, E[\overline{e}_t] = C_0$. C1 grants that $E[\overline{e}_t] = [I_2 - (A_0 + B_0)]^{-1} C_0$. Given equation (19),

$$cov\left(\overline{e}_t,\ \overline{e}_{t-i}\right) = cov\left(\overline{h}_t,\ \overline{e}_{t-i}\right), \quad i = 1, \ldots, \infty. \tag{42}$$

Recursive substitution into equation (17) reveals

$$\overline{h}_t = \left(I_2 - B_0\right)^{-1} C_0 + \sum_{j=1}^{\infty} B_0^{j-1} A_0 \overline{e}_{t-j}, \tag{43}$$

given Lemma 2. Substitution of equation (43) into (42) finds

$$cov\left(\overline{e}_t,\ \overline{e}_{t-i}\right) = \sum_{j=1}^{\infty} B_0^{j-1} A_0 cov\left(\overline{e}_{t-j},\ \overline{e}_{t-i}\right). \tag{44}$$

For $i = 1, 2$, equation (44) yields

$$cov\left(\overline{e}_t,\ \overline{e}_{t-2}\right) = \left(A_0 + B_0\right) cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right), \tag{45}$$

and, in general,

$$cov\left(\overline{e}_t,\ \overline{e}_{t-i}\right) = \left(A_0 + B_0\right)^{i-1} cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right), \tag{46}$$

given C1. Equation (46) specifies that the higher-order autocovariances of $\overline{e}_t$ exist if and only if the first-order autocovariance of $\overline{e}_t$ exists. Substitution of equation (46) into (44) for $i = 1$ produces

$$cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right) = A_0 var\left(\overline{e}_t\right) + \sum_{j=2}^{\infty} B_0^{j-1} A_0 cov\left(\overline{e}_{t-1},\ \overline{e}_t\right) \left(A_0 + B_0\right)^{j-2'}. \tag{47}$$

Applying the $vec(\cdot)$ operator to both sides of equation (47) yields

$$\begin{aligned} vec\left[cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right)\right] &= \left(I_2 \otimes A_0\right) vec\left[var\left(\overline{e}_t\right)\right] + \\ &\quad \left[\sum_{j=2}^{\infty} \left(A_0 + B_0\right)^{j-2} \otimes B_0^{j-1} A_0\right] vec\left[cov\left(\overline{e}_{t-1},\ \overline{e}_t\right)\right], \end{aligned} \tag{48}$$

while doing the same to the transpose of equation (47) produces

$$\begin{aligned} vec\left[cov\left(\overline{e}_{t-1},\ \overline{e}_t\right)\right] &= (A_0 \otimes I_2)\, vec\left[var\left(\overline{e}_t\right)\right] + \\ &\quad \left[\sum_{j=2}^{\infty} B_0^{j-1} A_0 \otimes (A_0 + B_0)^{j-2}\right] vec\left[cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right)\right]. \end{aligned} \tag{49}$$

In equation (48), $\sum\limits_{j=2}^{\infty} (A_0 + B_0)^{j-2} \otimes B_0^{j-1} A_0 = M_0$ while $\sum\limits_{j=2}^{\infty} B_0^{j-1} A_0 \otimes (A_0 + B_0)^{j-2} = N_0$ in equation (49), where M_0 and N_0 are defined in Lemma 3. Both of these results follow from C1 and Lemma 2. Substitution of equation (49) into equation (48) produces

$$[I_4 - M_0 N_0]\, vec\left[cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right)\right] = P_0 vec\left[var\left(\overline{e}_t\right)\right],$$

where $P_0 = \{I_4 - [(A_0 + B_0) \otimes B_0]\}^{-1} [I_2 - (B_0 \otimes B_0)] (I_2 \otimes A_0)$. Then,

$$vec\left[cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right)\right] = [I_4 - M_0 N_0]^{-1} P_0 vec\left[var\left(\overline{e}_t\right)\right] \tag{50}$$

given Lemma 3. Equation (50) is finite if and only if $vec\left[var\left(\overline{e}_t\right)\right]$ is finite. Since $E\left[\overline{w}_t \mid S_{t-1}\right] = 0$ in equation (19),

$$var\left(\overline{e}_t\right) = var\left(\overline{h}_t\right) + \Lambda.$$

Straight forward (though tedious) matrix algebra reveals that

$$var\left(\overline{h}_t\right) = \{I_4 - [(A_0 + B_0) \otimes (A_0 + B_0)]\}^{-1} (A_0 \otimes A_0)\, \Lambda,$$

given C1. As a result,

$$var\left(\overline{e}_t\right) = Q_0 \Lambda,$$

where $Q_0 = \{I_4 - [(A_0 + B_0) \otimes (A_0 + B_0)]\}^{-1} (A_0 \otimes A_0) + I_4$. Equation (50) then becomes

$$vec\left[cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right)\right] = [I_4 - M_0 N_0]^{-1} P_0 Q_0 vec\left[\Lambda\right], \tag{51}$$

which is finite given C3.

A.6. Proof of Proposition 1: Given A1 and A2,

$$R_{i,t} = Y_{i,t} - X_t' E\left[X_t X_t'\right]^{-1} E\left[X_t Y_{i,t}\right], \quad i = 1, 2, \tag{52}$$

where $R_{2,t} = \epsilon_{2,t}$. Let $R_t = \left[\begin{array}{cc} R_{1,t} & \epsilon_{2,t} \end{array}\right]'$. From equation (52), the relationship between R_t and ϵ_t is

$$R_t = \Gamma_0 \epsilon_t, \tag{53}$$

where $\Gamma_0 = \left[\begin{array}{cc} 1 & \gamma_0 \\ 0 & 1 \end{array}\right]$.Let $\overline{r}_t = \left[\begin{array}{cc} R_{1,t}\epsilon_{2,t} & \epsilon_{2,t}^2 \end{array}\right]'$. Using equation (53), the relationship between $\overline{r}_t$ and $\overline{e}_t$ is

$$\overline{r}_t = \Gamma_0 \overline{e}_t. \tag{54}$$

Substituting equation (54) into (45) from A.5 results in

$$cov\left(\overline{r}_t,\ \overline{r}_{t-2}\right) = \Gamma_0 \left(A_0 + B_0\right) \Gamma_0^{-1} cov\left(\overline{r}_t,\ \overline{r}_{t-1}\right). \tag{55}$$

$\Gamma_0 \left(A_0 + B_0\right) \Gamma_0^{-1}$ is identified by A3. Let $\delta_{ij,r0}$ be the element in the ith row and jth column of

$$\Delta_0 = \Gamma_0 \left(A_0 + B_0\right) \Gamma_0^{-1} = \left[\begin{array}{cc} \phi_{120} & \gamma_0 \left(\phi_{220} - \phi_{120}\right) \\ 0 & \phi_{220} \end{array}\right]. \tag{56}$$

Given A4, γ_0 is identified by

$$\gamma_0 = \frac{\delta_{12,r0}}{\delta_{22,r0} - \delta_{11,r0}}. \tag{57}$$

Given identification of γ_0, β_{10} is identified by $\beta_{10} = E\left[X_t X_t'\right]^{-1} E\left[X_t\left(Y_{1,t} - Y_{2,t}\gamma_0\right)\right]$. β_{20} is identified by $\beta_{20} = E\left[X_t X_t'\right]^{-1} E\left[X_t Y_{2,t}\right]$. The structural innovations $\epsilon_{1,t}$ are identified by $\epsilon_{1,t} = Y_{1,t} - X_t'\beta_{10} - Y_{2,t}\gamma_0$.

A.7. Proof of Proposition 2: Since $a_{11,10} = b_{11,10} = 0$,

$$cov(\epsilon_{1,t}\epsilon_{2,t},\ Z_{t-1}) = 0, \tag{58}$$

where $Z_{t-1} = \left[\begin{array}{ccc} \epsilon_{2,t-1}^2 & \cdots & \epsilon_{2,t-l}^2 \end{array}\right]'$ for finite $l \geq 1$. A1 and A2 identify $R_{1,t}$ and $\epsilon_{2,t}$ as equation (52). From equation (53), write $\epsilon_{1,t}$ as

$$\epsilon_{1,t} = R_{1,t} - \epsilon_{2,t}\gamma_0. \tag{59}$$

Substituting equation (59) into equation (58) yields

$$cov(R_{1,t}\epsilon_{2,t},\ Z_{t-1}) = cov(\epsilon_{2,t}^2,\ Z_{t-1})\gamma_0, \tag{60}$$

where existence of the individual row entries to $cov(\epsilon_{2,t}^2,\ Z_{t-1})$ is established by Lemma 4. Let $\Omega = cov(\epsilon_{2,t}^2,\ Z_{t-1})$. γ_0 is identified as

$$\gamma_0 = (\Omega'\Omega)^{-1}\,\Omega' cov(R_{1,t}\epsilon_{2,t},\ Z_{t-1}).$$

Given identification of γ_0, the identification of β_{10}, β_{20}, and $\epsilon_{1,t}$follows from A.6.

A.8. Proof of Corollary 2: By equations (14)–(16), $U_1 = X_t \otimes \epsilon_t$ and

$$U_3\left(\psi,\ Y_t,\ S_{t-1}\right) = vec\left[Z^{-1}\left[\left(\overline{e}_t - \sigma_{\overline{e}}\right)\left(\overline{e}_{t-2} - \sigma_{\overline{e}}\right)' - \overline{\Phi}\left(\overline{e}_t - \sigma_{\overline{e}}\right)\left(\overline{e}_{t-1} - \sigma_{\overline{e}}\right)'\right] Z^{-1}\right].$$

$E\left[U_2\right] = 0$ means that $E\left[\overline{e}_t\right] = \sigma_{\overline{e}}$, so $E\left[U\right] = 0$ is equivalent to $E\left[X_t \otimes \epsilon_t\right] = 0$ and $cov\left(\overline{e}_t,\ \overline{e}_{t-2}\right) = (A+B)\, cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right)$. Following, then, from either A.6 or A.7, the only $\varphi \in \Psi$ that satisfies $E\left[U\left(\psi, Y_t, S_{t-1}\right)\right] = 0$ is $\varphi = \varphi_0$.

References

[1] Black, F., M.C. Jensen and M.Scholes (1972), "The Capital Asset Pricing Model: Some Empirical Tests," in *Studies in the Theory of Capital Markets*, M.C. Jensen ed., Praeger: New York.

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

[3] Bollerslev, T., R.F. Engle and J.M. Wooldridge (1988), "A Capital Asset Pricing Model with Time-Varying Covariances," *Journal of Political Economy*, 96, 116–131.

[4] Campbell, J.Y. (1987), "Stock Returns and the Term Structure," *Journal of Financial Economics*, 18, 373–399.

[5] Diacogiannis, G. and D. Feldman (2006), "The CAPM Relation for Inefficient Portfolios," unpublished manuscript.

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

[7] Donald, S.G., G. Imbens and W. Newey (2002), "Choosing the Number of Moments in Conditional Moment Restriction Models," unpublished manuscript.

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

[9] Engle, R.F. and K.F. Kroner (1995), "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 122–150.

[10] Fama, E.F. and K.R. French (1988), "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3-25.

[11] Fama, E.F. and J.D. MacBeth (1973), "Risk, Return and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607-636.

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

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

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

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

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

[17] Klein, R. and F. Vella (2003), "Identification and Estimation of the Triangular Simultaneous Equations Model in the Absence of Exclusion Restrictions Through the Presence of Heteroskedasticity," unpublished manuscript.

[18] Kraus, A. and R.H. Litzenberger (1976), "Skewness Preference and the Valuation of Risk Assets," *Journal of Finance*, 4, 1085–1100.

[19] Lewbel, A. (2004), "Identification of Endogenous Heteroskedastic Models," unpublished manuscript.

[20] 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.

[21] 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.

[22] Nelson, D.B. (1992), "Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model," *Journal of Econometrics*, 52, 61–90.

[23] Nelson, D.B. and D.P. Foster (1994), "Asymptotic Filtering Theory for Univariate ARCH Models," *Econometrica*, 62, 1–41.

[24] 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.

[25] Newey, W.K. and R.J. Smith (2004), "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219–255.

[26] Newey, W.K. and R.J. Smith (2001), "Asymptotic Bias and Equivalence of GMM and GEL Estimators," unpublished manuscript.

[27] Prono, T. (2006), "GARCH-based Identification of Simultaneous Systems," unpublished manuscript.

[28] Rigobon, R. (2002), "The Curse of Non-Investment Grade Countries," *Journal of Development Economics*, 69, 423–449.

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

[30] Sentana, E. and G. Fiorentini (2001), "Identification, Estimation, and Testing of Conditionally Heteroskedastic Factor Models," *Journal of Econometrics*, 102, 143–164.

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

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

[33] Stock, J.H. and J.H. Wright (2000), "GMM with Weak Identification," *Econometrica*, 68, 1055–1096.

[34] Weis, A.A. (1986), "Asymptotic Theory for ARCH Models: Estimation and Testing," *Econometric Theory*, 2, 107–131.

[35] West, K.D. (2002), "Efficient GMM Estimation of Weak AR Processes," *Economic Letters*, 75, 415–418.

[36] Wright, P.G. (1928) "The Tariff on Animal and Vegetable Oils," The Institute of Economics, The Macmillan Company, NY.

[37] Wright, S. (1921), "Correlation and Causation," *Journal of Agricultural Research*.

[38] Zhou, G. (1991), "Small Sample Tests of Portfolio Efficiency," *Journal of Financial Economics*, 30, 165–191.

TABLE 1

		T=1260				T=2520			
		Med.		Decile		Med.		Decile	
Q	σ_{12}	Bias	MDAE	Range	SD	Bias	MDAE	Range	SD
	0.1	0.000	0.089	1.275	1.195	0.000	0.071	1.008	0.654
2	0.2	0.000	0.093	1.167	1.223	0.000	0.071	0.922	0.727
	0.4	0.003	0.094	1.038	0.899	0.002	0.082	0.866	0.647
	0.1	0.033	0.120	0.567	0.257	0.028	0.119	0.555	0.245
4	0.2	0.069	0.131	0.567	0.252	0.060	0.124	0.552	0.241
	0.4	0.145	0.171	0.593	0.256	0.129	0.153	0.582	0.252
	0.1	0.057	0.110	0.409	0.166	0.056	0.107	0.387	0.157
8	0.2	0.118	0.136	0.414	0.167	0.111	0.130	0.396	0.160
	0.4	0.252	0.252	0.470	0.185	0.220	0.221	0.460	0.180
	0.1	0.070	0.097	0.319	0.127	0.067	0.091	0.299	0.118
16	0.2	0.143	0.147	0.330	0.130	0.129	0.133	0.314	0.123
	0.4	0.303	0.303	0.381	0.147	0.260	0.260	0.379	0.146

Notes to Table 1: Monte Carlo studies were conducted across 5000 trials for $T = 1260,\ 2520$ observations. Results for the parameter γ are shown. Q denotes the number of lags used in the single step GMM estimator. σ_{12} is the level of covariance between innovations to the given triangular system. The unconditional variance of each innovation is one in all cases. Med. Bias is the median bias of the parameter estimates relative to the true value $\gamma_0 = 1$. MDAE is the median absolute error of the parameter estimates relative to the true value. Decile Range is the difference between the 0.10 and 0.90 quantiles of the parameter estimates, while SD is the standard deviation.

TABLE 2

	6/1/65 - 9/30/77				10/3/77 - 1/31/90				2/1/90 - 5/31/02			
			Conf. Interval				Conf. Interval				Conf. Interval	
	Est.	S.E.	95%		Est.	S.E.	95%		Est.	S.E.	95%	
δ_1	0.436	0.211	0.008	0.716	-0.067	0.393	-0.162	0.181	-0.169	0.079	-0.246	0.041
δ_2	0.371	0.141	0.014	0.517	-0.077	0.192	-0.180	0.299	-0.117	0.091	-0.223	0.135
δ_3	0.292	0.092	0.028	0.369	0.002	0.809	-0.233	0.678	-0.202	0.071	-0.231	0.019
δ_4	0.311	0.095	0.041	0.400	-0.067	0.116	-0.138	0.263	0.216	0.068	-0.010	0.247
δ_5	0.372	0.116	0.027	0.453	-0.030	0.153	-0.128	0.202	-0.114	0.051	-0.166	0.024
δ_6	0.345	0.099	0.046	0.410	-0.025	0.097	-0.092	0.175	-0.082	0.039	-0.122	0.030
δ_7	0.046	0.034	-0.038	0.096	-0.037	0.081	-0.057	0.105	-0.075	0.043	-0.118	0.054
δ_8	0.147	0.044	0.011	0.177	-0.162	0.031	-0.112	0.000	-0.039	0.062	-0.140	0.104
δ_9	0.013	0.013	-0.027	0.026	-0.559	0.110	-0.386	-0.001	0.140	0.053	0.001	0.190
δ_{10}	-0.079	0.029	-0.120	-0.010	0.157	0.044	0.098	0.260	0.119	0.041	0.001	0.142

Notes to Table 2: δ_i for $i = 1, \ldots, 10$ are the individual elements of the vector Δ in equation (11). Est. are the $\widehat{\delta}$s from a regression of $\widehat{\epsilon}_{i,t}$ on $\widehat{\epsilon}_{p,t}$–see equations (27) and (28)–while S.E. are the corresponding bootstrapped standard errors. The 95% confidence intervals are formed from the 2.5 and 97.5 percentiles of the bootstrapped distribution for each $\widehat{\delta}$. In all cases, the bootstrap is implemented for 2000 repetitions.

TABLE 3

		θ_p							
		0.169		0.313		0.493		1.00	
		α		α		α		α	
Period	Estimator	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
6/1/65 – 9/30/77	OLS	0.090	0.082	0.165	0.151	0.256	0.234	0.473	0.439
	GMM	0.073	0.065	0.134	0.119	0.208	0.187	0.396	0.359
10/3/77 - 1/31/90	OLS	0.097	0.088	0.178	0.160	0.275	0.249	0.501	0.461
	GMM	0.091	0.081	0.166	0.149	0.258	0.232	0.475	0.435
2/1/90 - 5/31/02	OLS	0.102	0.091	0.186	0.166	0.286	0.257	0.518	0.474
	GMM	0.094	0.083	0.171	0.153	0.265	0.237	0.486	0.443

TABLE 4

	θ_p							
	0.169		0.313		0.493		1.00	
	α		α		α		α	
Period	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
6/1/65 – 9/30/77	-19.41%	-21.11%	-19.13%	-20.87%	-18.57%	-20.37%	-16.15%	-18.16%
10/3/77 - 1/31/90	-6.67%	-7.12%	-6.53%	-7.00%	-6.26%	-6.77%	-5.16%	-5.77%
2/1/90 - 5/31/02	-8.02%	-8.10%	-7.85%	-7.96%	-7.51%	-7.69%	-6.12%	-6.50%

Notes to Tables 3 and 4: θ_p is an annualized Sharpe ratio (i.e., $\theta_p = \left(252 \times \text{daily } \theta_p^2\right)^{1/2}$). $\theta_p = 0.169$ corresponds to the median of the bootstrapped distribution of $\widehat{\theta}_p^2$ and implies an expected excess proxy return of 3.39%. $\theta_p = 0.313$ corresponds to the point estimate $\widehat{\theta}_p^2$ and implies an excess return of 6.25%. $\theta_p = 0.493$ corresponds to the 95th percentile of the bootstrapped distribution of $\widehat{\theta}_p^2$ and implies an excess return of 9.87%. The time period considered for $\widehat{\theta}_p^2$ is 6/1/65–5/31/02. The bootstrapped distribution of $\widehat{\theta}_p^2$ is calculated over 10000 repetitions. All implied expected excess returns assume an annualized standard deviation of 20%. For a given value of θ_p, Table 3 presents the value of ρ that supports the inequality of equation (12) for $(1-\alpha)\,\%$ of the bootstrapped distribution of $d'\Sigma_e^{-1}d$ calculated over 2000 repetitions. Table 4 presents the $\%$ difference in ρ implied by the GMM estimator relative to OLS.

TABLE 5A

	6/1/65 - 7/30/71				8/2/71 - 9/30/77				10/3/77 - 11/30/83			
			Conf. Interval				Conf. Interval				Conf. Interval	
	Est.	S.E.	95%		Est.	S.E.	95%		Est.	S.E.	95%	
δ_1	-0.040	0.124	-0.306	0.180	0.225	0.197	0.004	0.454	0.023	0.281	-0.164	0.370
δ_2	-0.020	0.049	-0.121	0.070	0.188	0.092	0.018	0.347	-0.024	0.151	-0.168	0.230
δ_3	0.090	0.054	-0.049	0.159	0.181	0.077	0.017	0.306	-0.009	0.142	-0.151	0.217
δ_4	-0.017	0.032	-0.071	0.052	0.172	0.059	0.051	0.276	-0.007	0.106	-0.121	0.159
δ_5	0.049	0.031	-0.021	0.097	0.141	0.045	0.099	0.254	-0.102	0.055	-0.194	0.001
δ_6	0.042	0.015	0.009	0.067	0.176	0.069	0.026	0.291	-0.025	0.047	-0.102	0.076
δ_7	-0.054	0.019	-0.070	0.001	0.046	0.067	-0.106	0.155	-0.066	0.033	-0.117	0.010
δ_8	-0.034	0.025	-0.072	0.027	0.080	0.045	-0.022	0.156	-0.040	0.027	-0.077	0.028
δ_9	-0.049	0.016	-0.064	-0.002	0.007	0.028	-0.056	0.053	-0.029	0.013	-0.051	0.000
δ_{10}	0.072	0.018	0.001	0.063	-0.016	0.039	-0.076	0.034	0.018	0.035	-0.046	0.072

TABLE 5B

	2/1/90 - 3/29/96				4/1/96 - 5/31/02			
			Conf. Interval				Conf. Interval	
	Est.	S.E.	95%		Est.	S.E.	95%	
δ_1	-0.028	0.051	-0.110	0.081	-0.256	0.076	-0.269	-0.001
δ_2	-0.710	0.186	-0.689	-0.014	-0.184	0.085	-0.273	0.024
δ_3	-0.154	0.063	-0.267	-0.012	-0.161	0.075	-0.268	-0.001
δ_4	-0.151	0.051	-0.250	-0.045	-0.189	0.076	-0.275	0.001
δ_5	-0.125	0.044	-0.215	-0.035	-0.129	0.056	-0.195	0.002
δ_6	0.000	0.038	-0.081	0.069	-0.161	0.053	-0.194	-0.001
δ_7	0.001	0.033	-0.057	0.067	-0.138	0.051	-0.186	-0.001
δ_8	0.085	0.024	0.029	0.124	0.136	0.060	-0.050	0.186
δ_9	0.028	0.024	-0.035	0.058	-0.025	0.026	-0.084	0.018
δ_{10}	-0.045	0.010	-0.064	-0.026	0.106	0.042	0.001	0.155

Notes to Table 5A and 5B: δ_i for $i = 1, \ldots, 10$ are the individual elements of the vector Δ in equation (11). Est. are the $\widehat{\delta}$s from a regression of $\widehat{\epsilon}_{i,t}$ on $\widehat{\epsilon}_{p,t}$–see equations (27) and (28)–while S.E. are the corresponding bootstrapped standard errors. The 95% confidence intervals are formed from the 2.5 and 97.5 percentiles of the bootstrapped distribution for each $\widehat{\delta}$. In all cases, the bootstrap is implemented for 2000 repetitions.

TABLE 6

		θ_p							
		0.169		0.313		0.493		1.00	
		α		α		α		α	
Period	Estimator	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
6/1/65 - 7/30/71	OLS	0.062	0.057	0.115	0.105	0.179	0.165	0.346	0.321
	GMM	0.059	0.053	0.109	0.098	0.171	0.153	0.332	0.300
8/2/71 - 9/30/77	OLS	0.071	0.066	0.131	0.120	0.204	0.188	0.389	0.362
	GMM	0.058	0.050	0.106	0.093	0.167	0.145	0.324	0.285
10/3/77 - 11/30/83	OLS	0.054	0.048	0.100	0.089	0.157	0.140	0.306	0.276
	GMM	0.050	0.044	0.093	0.082	0.145	0.129	0.285	0.254
2/1/90 - 3/29/96	OLS	0.072	0.063	0.131	0.116	0.205	0.182	0.390	0.350
	GMM	0.072	0.063	0.132	0.116	0.205	0.182	0.391	0.351
4/1/96 - 5/31/02	OLS	0.069	0.062	0.127	0.113	0.198	0.177	0.379	0.343
	GMM	0.062	0.056	0.113	0.103	0.177	0.161	0.343	0.313

TABLE 7

	θ_p							
	0.169		0.313		0.493		1.00	
	α		α		α		α	
Period	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
6/1/65 - 7/30/71	-4.62%	-7.41%	-4.57%	-7.34%	-4.49%	-7.23%	-4.10%	-6.73%
8/2/71 - 9/30/77	-19.00%	-23.35%	-18.84%	-23.19%	-18.49%	-22.85%	-16.87%	-21.25%
10/3/77 - 11/30/83	-7.43%	-8.23%	-7.39%	-8.19%	-7.30%	-8.11%	-6.83%	-7.69%
2/1/90 - 3/29/96	0.34%	0.22%	0.32%	0.22%	0.31%	0.22%	0.28%	0.20%
4/1/96 - 5/31/02	-10.82%	-9.65%	-10.73%	-9.56%	-10.51%	-9.41%	-9.52%	-8.68%

Notes to Tables 3 and 4: θ_p is an annualized Sharpe ratio (i.e., $\theta_p = \left(252 \times \text{daily } \theta_p^2\right)^{1/2}$). $\theta_p = 0.169$ corresponds to the median of the bootstrapped distribution of $\widehat{\theta}_p^2$ and implies an expected excess proxy return of 3.39%. $\theta_p = 0.313$ corresponds to the point estimate $\widehat{\theta}_p^2$ and implies an excess return of 6.25%. $\theta_p = 0.493$ corresponds to the 95th percentile of the bootstrapped distribution of $\widehat{\theta}_p^2$ and implies an excess return of 9.87%. The time period considered for $\widehat{\theta}_p^2$ is 6/1/65–5/31/02. The bootstrapped distribution of $\widehat{\theta}_p^2$ is calculated over 10000 repetitions. All implied expected excess returns assume an annualized standard deviation of 20%. For a given value of θ_p, Table 3 presents the value of ρ that supports the inequality of equation (12) for $(1-\alpha)\,\%$ of the bootstrapped distribution of $d'\Sigma_e^{-1}d$ calculated over 2000 repetitions. Table 4 presents the $\%$ difference in ρ implied by the GMM estimator relative to OLS.