Working Papers
Federal Reserve Bank of Boston

No. 07-1

GARCH-Based Identification of Triangular Systems with an Application to the CAPM: Still Living with the Roll Critique

Todd Prono

Abstract:
This paper presents a new method for identifying triangular systems of time-series data. Identification is the product of a bivariate GARCH process. Relative to the literature on GARCH-based identification, this method distinguishes itself both by allowing for a time-varying covariance and by not requiring a complete estimation of the GARCH parameters. Estimation follows OLS and standard univariate GARCH and ARMA techniques, or GMM. A Monte Carlo study of the GMM estimator is provided. The identification method is then applied in testing a conditional version of the CAPM.

Keywords: triangular systems, endogeneity, identification, conditional heteroskedasticity, generalized method of moments, GARCH, GMM, CAPM

JEL Codes: C13, C32, G12

Todd Prono is a Financial Economist at the Federal Reserve Bank of Boston. His email address is todd.prono@bos.frb.org.

This paper, which may be revised, is available on the web site of the Federal Reserve Bank of Boston at http://www.bos.frb.org/economic/wp/index.htm.

Earlier versions of this paper were circulated under the titles "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 for helpful comments and discussions.

This version: **March 2007 (first version: April 2006)**

1 Introduction

This paper presents a new method for identifying the effect of an endogenous regressor in linear models of time-series data. The method is applicable in cases where proper instruments are not available. Identification derives from generalized autoregressive conditional heteroskedasticity (GARCH) in the model's error terms. Relative to the literature basing identification on the GARCH structure, this paper's methodology both generalizes and simplifies the existing work. Regarding the former, the need for a constant covariance between errors is relaxed to allow for time variation. Concerning the latter, identification does not depend on all of the structural parameters from the GARCH model, rather, on only two composite nuisance parameters. Given the reliance on GARCH errors, potential applications for this identification method are found in the asset pricing literature. This paper focuses on testing the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965). An updated version of the test statistic proposed by Shanken (1987) is derived to recognize not only the latent nature of the true market return as argued by Roll (1977), but also the likely endogeneity of any observable proxy.

The proposed identification method is implemented using either OLS and standard univariate GARCH and autoregressive moving average (ARMA) models, or generalized method of moments (GMM). A Monte Carlo study of the GMM estimator is provided. An empirical investigation considers the premise that observable proxies of the true market return are endogenous regressors and provides supporting evidence for this assertion. This paper's methodology is then used to test a conditional version of the CAPM in the spirit of Shanken (1987) and Kandel and Stambaugh (1987). The result is that the CAPM is rejected if the correlation between innovations to an equal-weighted NYSE/AMEX proxy and the true market return exceeds 0.62.[3] The maximum correlation drops to 0.49 if a value-weighted proxy is

[3]The general intuition behind this test is that if the statistical inefficiency of a given proxy to the true market

used instead.

Let $Y_t = \begin{bmatrix} Y_{1,t} & Y_{2,t} \end{bmatrix}'$ be a vector of endogenous variables, $\epsilon_t = \begin{bmatrix} \epsilon_{1,t} & \epsilon_{2,t} \end{bmatrix}'$ a vector of unobservable shocks, and X_t a vector of predetermined covariates that can include lags of the endogenous variables. 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.[4] Consider the following model:

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

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

Equations (1) and (2) define a triangular system. Assume that $E\left[\epsilon_t \mid S_{t-1}\right] = 0$. The moments $E\left[X_t\epsilon_{2,t}\right] = 0$ identify equation (2). The moments $E\left[X_t\epsilon_{1,t}\right] = 0$ alone, however, do not identify equation (1). If, in addition to these moments, $E\left[\epsilon_{1,t}\epsilon_{2,t} \mid S_{t-1}\right] = 0$, then equation (1) is identified and can be estimated by OLS. Alternatively, if some element of the vector β_1 is zero while the corresponding element in β_2 is nonzero, then equation (1) is identified by instrumental variables (IV).

This paper considers identifying equation (1) by restricting the conditional second moments of ϵ_t. Assume $E\left[\epsilon_t\epsilon_t' \mid S_{t-1}\right] = H_t$, where H_t is parameterized according to the diagonal BEKK model of Engle and Kroner (1995).[5] As a vector generalization of the multivariate GARCH model, this parameterization offers a discrete approximation to a diffusion process without jumps, thus linking the identification result in this paper to modern finance theory.[6] Provided additional conditions hold, the zero restrictions imposed by the diagonal

return is gauged to be sufficiently strong, then the inefficiency of the true market return may be correctly inferred and, as a result, the CAPM rejected.

[4]The vector X_t can be assumed to also contain weakly exogenous elements (see Engle, Hendry, and Richard, 1983, for a definition of weak exogeneity). Such an assumption requires S_{t-1} to be redefined so as not to include these elements. Furthermore, any weakly exogenous elements of X_t must be uncorrelated with ϵ_t.

[5]The BEKK model guarantees positive definiteness of H_t under very mild conditions.

[6]See Nelson (1992).

BEKK model grant identification. Examples of these additional conditions include a covariance stationary process for H_t and a finite fourth moment for $\epsilon_{2,t}$.

The problem in identifying equation (1) arises because there are too few moment conditions relative to the number of structural parameters. Assuming $E\left[\epsilon_t\epsilon_t' \mid S_{t-1}\right]$ to be a fully general GARCH process fails to remedy this problem because such an assumption does not deliver additional moment conditions without additional structural parameters.[7] However, restricting some of these structural parameters to be zero via the diagonal model results in additional moment conditions sufficient for identification in an analogous fashion to restricting certain elements of the vector β_1 to be zero in equation (1). Furthermore, the diagonality imposed on H_t, which represents the key identifying assumption of this paper, is testable. When the estimator for the triangular system is GMM, the J statistic provides a joint test of whether ϵ_t is conditionally mean zero and whether H_t is properly specified. The χ^2 difference test of Newey and West (1987) applied to the moments defining H_t tests the fit of H_t alone. A rejection from either test speaks directly against the assumptions identifying equation (1).

This paper is organized into six sections. Section 1.1 shows how the triangular system relates to the CAPM, while section 1.2 provides an overview of the identification method and how this method extends the current literature. Section 2 discusses the necessary conditions for identification and justifies the GARCH structure. Section 3 describes two possible estimation techniques. Section 4 conducts a Monte Carlo study of the proposed GMM estimator. Section 5.1 develops a test statistic for the CAPM that recognizes the possible endogeneity of observable market proxies. Using the identification and estimation techniques developed in sections 2 and 3, section 5.2 provides empirical evidence supporting this endogeneity. Section 5.3 presents the results of the CAPM test, and section 6 concludes.

[7] See Proposition 2.3 of Engle and Kroner (1995) for a description of the most general (where generality is determined by the number of cross-equation restrictions being imposed) BEKK model. For a bivariate model, full generality specifies H_t with 18 structural parameters.

1.1 The CAPM

This section motivates the triangular system in financial economics. The application is a conditional test of the CAPM. Assume there exists an observable risk-free rate. Define $Y_{1,t}$ as the excess return on an arbitrarily chosen security, and let $Y_{2,t}$ be an observable proxy of the true excess market return.[8] Finally, define X_t as a vector of instruments that forecast security returns.[9] Consider the model

$$Y_{1,t} = X_t'\delta + \epsilon_{2,t}\gamma + \epsilon_{1,t}, \tag{3}$$

where $\epsilon_{2,t}$ is the shock to equation (2).[10] Continue to assume that $E\left[\epsilon_t \mid S_{t-1}\right] = 0$.

The model of equation (3) has two salient features. First, it decomposes security returns into predictable and unpredictable components. Second, it prices the first security return relative to a single factor, the unpredictable component of the excess market return proxy. This latter feature relates equation (3) to a conditional version of the CAPM. Substituting equation (2) into (3) and simplifying yields equation (1), with $\beta_1 = \delta - \beta_2\gamma$.

Under the CAPM, the true market return is mean variance efficient (MVE). As first noted by Roll (1977), any observable proxy for the true market return may or may not be MVE. Suppose the given proxy is not MVE.[11] In this case, shocks affecting the proxy could be correlated with shocks affecting other security returns through, for example, unanticipated changes to either international liquidity or investor preferences for nontraded assets.[12] This

[8]Excess returns are measured relative to the risk-free rate.

[9]As might be expected, the literature on predicting asset returns is long. See Ferson, Sarkissian, and Simin (2003) for a review of the proposed instruments and a discussion of the potential shortcomings these instruments face.

[10]Ferson (1990) considers models in the general form of equation (3).

[11]Works by such authors as Roll (1980), Gibbons (1982), Jobson and Korkie (1982), Shanken (1987), Kandel and Stambaugh (1987), Gibbons, Ross, and Shanken (1987), Zhou (1991), and MacKinlay and Richardson (1991) reject the efficiency of various market index proxies.

[12]Miller and Scholes (1972) note the possibility of a nonzero covariance between shocks affecting individual

correlation renders the proxy an endogenous regressor in equation (1) and OLS unsuitable for estimating γ. Furthermore, IV is complicated by the difficulty of finding an instrument correlated with the proxy but not the security.

If $E\left[\epsilon_t\epsilon_t' \mid S_{t-1}\right] = H_t$, where H_t is parameterized by the diagonal BEKK model, then equation (1) is identified. Such a parameterization merely allows the individual elements of the conditional covariance matrix to display ARMA properties. While these properties do not arise from any economic theory, they do offer a "parsimonious approximation to the form of heteroskedasticity typically encountered with economic [and financial] time-series data" (Bollerslev, Engle, and Wooldridge, 1988, p. 119). Bollerslev, Engle, and Wooldridge (1988), in their study of a CAPM with time-varying covariances, find the conditional covariance matrix of asset returns to be strongly autoregressive. 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). Ferson (1985), Ferson, Kandel, and Stambaugh (1987), and Bodurtha and Mark (1991) make similar assertions. This paper extends these authors' argument by demonstrating how the GARCH structure relates to consistent estimation of an individual security's sensitivity to a chosen market proxy.

Of course, devising a means for consistently estimating the relationship between individual security returns and a chosen market proxy does nothing to avoid Roll's (1977) statement that "the [CAPM] theory is not testable unless the exact composition of the true market portfolio is known and used in the tests." Fortunately, as demonstrated by Shanken (1987), estimates of the relationship between individual security and proxy still play a role in testing the CAPM, conditional on a prior belief about the correlation between the true market return and the proxy. The empirical application in this paper is dedicated to performing such a test. Kandel and Stambaugh (1987) discuss a similar testing strategy. Their results serve as an

returns and the market proxy. Fama, Fisher, Jensen, and Roll (1969) find empirical support for a nonzero covariance between the market proxy and shocks to individual returns in the months preceding a split.

additional means of comparison.

1.2 Overview

This paper explores the conditions that identify equation (1) in the case where $\epsilon_t \mid S_{t-1} \sim id\,(0,\ H_t)$.[13] The individual elements of the conditional covariance matrix H_t are parameterized as

$$h_{11,t} = \left(c_1^2 + c_2^2\right) + \left(a_{11,2}^2\right)\epsilon_{1,t-1}^2 + \left(b_{11,2}^2\right)h_{11,t-1}, \tag{4}$$

$$h_{12,t} = \left(c_2 c_3\right) + \left(a_{11,2}a_{22,2}\right)\epsilon_{1,t-1}\epsilon_{2,t-1} + \left(b_{11,2}b_{22,2}\right)h_{12,t-1}, \tag{5}$$

and

$$h_{22,t} = \left(c_3^2\right) + \left(a_{22,1}^2 + a_{22,2}^2\right)\epsilon_{2,t-1}^2 + \left(b_{22,1}^2 + b_{22,2}^2\right)h_{22,t-1}. \tag{6}$$

Equations (4)–(6) illustrate the restrictions a diagonal BEKK model imposes on H_t. Focusing attention on equation (5), the conditional covariance of $\epsilon_{1,t-1}$ and $\epsilon_{2,t-1}$ does not depend on any of the predetermined covariates from either equation (4) or (6). For example, $\epsilon_{1,t-1}^2$, $\epsilon_{2,t-1}^2$, $h_{11,t-1}$, and $h_{22,t-1}$are all excluded as possible explanatory variables. In equation (6), the conditional variance of $\epsilon_{2,t}$ does not depend on any of the predetermined covariates from either equation (4) or (5), since the variables $\epsilon_{1,t-1}^2$, $\epsilon_{1,t-1}\epsilon_{2,t-1}$, $h_{11,t-1}$, and $h_{12,t-1}$ are all omitted. These zero restrictions imposed on the set of possible explanatory variables for $h_{12,t}$ and $h_{22,t}$ are what identify γ in equation (1) of the triangular system. Specifically, these restrictions identify periods of heightened volatility in $\epsilon_{2,t}$ relative to its unconditional mean, and these periods serve as an instrument for $Y_{2,t}$.

Notice the parallel between the identification approach described above and the common method of imposing zero restrictions on individual elements of the vector β_1 in equation (1). The identification approach in this paper transfers the necessary zero restrictions away from

[13] id means identically distributed.

the conditional mean equations (the first moments) and onto the conditional covariance matrix (the second moments). Justifying these restrictions are the empirical findings that support the univariate GARCH(1,1) model among the best predictors of stock market volatility and the link between GARCH and continuous time diffusion processes.[14] In constrast to identification through zero restrictions imposed on β_1, basing identification on the zero restrictions in equations (4)–(6) has the benefit of being testable. Equations (4)–(6) define an entire autocovariance process for $vech\left(\epsilon_t \epsilon_t'\right)$.[15] The fit of that process can be tested at lags $t-l$, where $l>1$.

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 King, Sentana, and Wadhwani (1994), Sentana and Fiorentini (2001), and Lewbel (2004) discuss identification methods that require $h_{12,t}=c$.[16] Given equation (5), this paper's methodology generalizes these works by allowing the conditional covariance to be time-varying. In addition, identification in this paper does not require a specific distributional assumption for $\epsilon_t \mid S_{t-1}$, as in King et al. (1994) and Sentana and Fiorentini (2001).

The current literature that bases identification on the GARCH structure also requires full identification of that structure. King et al. (1994), Engle and Kroner (1995), Sentana and Fiorentini (2001) and Rigobon (2002), all serve as examples. In contrast, the method de-

[14]A diagonal bivariate GARCH model treats each conditional variance and the conditional covariance as a univariate GARCH(1,1) process—see equations (4)–(6).

[15]The $vech\left(\cdot\right)$ operator stacks the lower triangle of an $(n \times n)$ matrix into an $([n\left(n+1\right)/2] \times 1)$ column vector.

[16]King et al. (1994) consider a dynamic version of Ross' (1976) arbitrage pricing theory (APT) as a generalization of the single-factor model discussed in section 1.1.

scribed in this paper requires identification of only two nuisance parameters

$$\phi_{12} = a_{11,2}a_{22,2} + b_{11,2}b_{22,2}$$

and

$$\phi_{22} = a_{22,1}^2 + a_{22,2}^2 + b_{22,1}^2 + b_{22,2}^2,$$

both composite functions of the parameters governing the ARCH and GARCH effects in equations (5) and (6), respectively. This paper's identification result, therefore, economizes on the number of parameters that need to be considered from the GARCH process.

2 Identification

Consider the triangular system

$$Y_{1,t} = X_t'\beta_{10} + Y_{2,t}\gamma_0 + \epsilon_{1,t}, \tag{7}$$

and

$$Y_{2,t} = X_t'\beta_{20} + \epsilon_{2,t}. \tag{8}$$

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.[17]

Equation (8) is identified by OLS. If $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are uncorrelated, then equation (7) is also identified by OLS. If, instead, at least one element of β_{10} is zero while the corresponding element in β_{20} is nonzero, then equation (7) is identified by instrumental variables (IV). This section provides identification conditions that require neither uncorrelated errors nor zero

[17]The inclusion of time trends or deterministic regressors is easily accommodated by replacing the relevant moments with probability limits of sample moments and sample projections.

restrictions on β_{10}. Instead, identification hinges on the parameterization of $E\left[\epsilon_t \epsilon_t' \mid S_{t-1}\right]$. Partial identification of this parameterization is demonstrated.

Assumption 1: $E\left[X_t X_t'\right]$ and $E\left[X_t Y_t'\right]$ are finite and identified from the data. $E\left[X_t X_t'\right]$ is nonsingular.

Assumption 2: $\epsilon_t \mid S_{t-1} \sim id\left(0,\ H_t\right)$, where $\epsilon_t = \begin{bmatrix} \epsilon_{1,t} & \epsilon_{2,t} \end{bmatrix}'$ and

$$H_t = C_0' C_0 + \sum_{k=1}^{2} A_{k0}' \epsilon_{t-1} \epsilon_{t-1}' A_{k0} + \sum_{k=1}^{2} B_{k0}' H_{t-1} B_{k0}. \tag{9}$$

Assumption 3: $C_0 = \begin{bmatrix} c_{10} & 0 \\ c_{20} & c_{30} \end{bmatrix}$. Let $a_{ij,k0}$ and $b_{ij,k0}$ be the element in the ith row and jth column of the matrices A_{k0} and B_{k0}, respectively. $A_{10} = \begin{bmatrix} 0 & 0 \\ 0 & a_{22,10} \end{bmatrix}$, $A_{20} = \begin{bmatrix} a_{11,20} & 0 \\ 0 & a_{22,20} \end{bmatrix}$, $B_{10} = \begin{bmatrix} 0 & 0 \\ 0 & b_{22,10} \end{bmatrix}$, and $B_{20} = \begin{bmatrix} b_{11,20} & 0 \\ 0 & b_{22,20} \end{bmatrix}$. c_1, c_3, $a_{22,1}$, $a_{11,2}$, $b_{22,1}$, and $b_{11,2}$ are strictly positive.

Assumptions 1 and 2 identify the structural innovations to equation (8) and the reduced form innovations to equation (7) as

$$\epsilon_{2,t} = Y_{2,t} - X_t' E\left[X_t X_t'\right]^{-1} E\left[X_t Y_{2,t}\right] \tag{10}$$

and

$$R_{1,t} = Y_{1,t} - X_t' E\left[X_t X_t'\right]^{-1} E\left[X_t Y_{1,t}\right], \tag{11}$$

respectively. Assumptions 2 and 3 describe a fully general diagonal BEKK model with no equivalent representations.[18] By inspection, equation (9) will be positive definite under very mild conditions, thus illustrating the principal advantage of the BEKK model over alternative multivariate GARCH specifications. The fact that all nonzero elements in the matrices A_{k0} and B_{k0} occur along the diagonals evidences why the model of equation (9) and Assumption 3 is termed diagonal. The reduced-form of equation (9) is solved using equations (10) and (11). This reduced-form is shown to identify γ_0.

Assumption 4: $\sum_{k=1}^{2} \left(a_{ii,k0}^2 + b_{ii,k0}^2\right) < 1,\ i = 1, 2.$

Assumption 5: $a_{22,20}$ and $b_{22,20}$ are nonzero.

Assumption 4 defines the model in equation (9) to be covariance stationary and follows from Proposition 2.7 of Engle and Kroner (1995). Assumption 5 ascribes a time-varying covariance to the innovations $\epsilon_{1,t}$ and $\epsilon_{2,t}$ and echoes the theme of King, Sentana, and Wadhwani (1994), Sentana and Fiorentini (2001), Rigobon (2002), and Lewbel (2004) in the sense that identification relates to a property of the conditional covariance. Unlike these authors, however, who demonstrate this property to be time-invariance, Assumption 5 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.

Assumption 6: $cov\left(\bar{e}_t,\ \bar{e}_{t-1}\right)$ is nonsingular, where $\bar{e}_t = \left[\begin{array}{cc} \epsilon_{1,t}\epsilon_{2,t} & \epsilon_{2,t}^2 \end{array}\right]'$.

Assumption 6 imposes higher moment restrictions on the structural innovations to equations (7) and (8) and is similar in scope to Assumption A3 of Lewbel (2004). Principal among

[18] See Proposition 2.3 of Engle and Kroner (1995) for a complete discussion of BEKK model identification and Proposition 2.6 for a specific treatment of the diagonal model.

these restrictions is a finite fourth moment for $\epsilon_{2,t}$, which implies additional constraints on the permissible values of the individual ARCH and GARCH effects defined in Assumption 3. The precise form of these additional constraints depends on the distribution of ϵ_t (see Hamilton (1994) for an example related to the univariate ARCH model with normal errors).

Assumption 7: $\phi_{120} \neq \phi_{220}$, where $\phi_{120} = a_{11,20}a_{22,20} + b_{11,20}b_{22,20}$ and $\phi_{220} = a^2_{22,10} + a^2_{22,20} + b^2_{22,10} + b^2_{22,20}$.

Assumption 7 distinguishes the structural parameters governing time-variation in $h_{12,t}$ and $h_{22,t}$. This restriction, together with Assumption 6, can be likened to the linear independence among time-varying portions of the factor variances in Sentana and Fiorentini (2001). Finally, given Assumptions 3 and 4, $-1 < \phi_{12} < 1$ and $0 < \phi_{22} < 1$.

Proposition 1. *Let Assumptions 1–7 hold for the model of equations (7) and (8). The structural parameters of the triangular system together with c_{20}, c_{30}, ϕ_{120}, and ϕ_{220} are identified.*

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

Proposition 1 treats the structural parameters of H_t in equation (9) and Assumption 3 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}. As a result, identification of γ_0 only depends upon partial identification of H_t.

Lemma 1: *The strict positivity of $a_{11,2}$ and $b_{11,2}$ in Assumption 3 and the nonzero restrictions on* $a_{22,20}$ and $b_{22,20}$ in *Assumption 5 are both necessary conditions for Assumption 6.*

Proposition 1 extends the identification results of King, Sentana, and Wadhwani (1994), Sentana and Fiorentini (2001), and Lewbel (2004) by allowing for a time-varying conditional covariance between $\epsilon_{1,t}$ and $\epsilon_{2,t}$. Lemma 1 highlights this time-varying covariance as a necessary condition for identification. Relaxing this condition is the subject of the following assumptions and proposition.

Assumption 5a: $a_{22,20} = b_{22,20} = 0$.

Assumption 6a: $cov\left(\bar{e}_t,\ \bar{e}_{t-1}\right)$ has rank one.

Given Assumption 5a, the maximum possible rank of $cov\left(\bar{e}_t,\ \bar{e}_{t-1}\right)$ is one. Assumption 6a, therefore, adjusts for the time-invariance of $h_{12,t}$ while maintaining the higher moment restrictions imposed in Assumption 6. Note that Assumption 6a is violated if $\epsilon_{2,t}$ is homoskedastic.

Proposition 2. *Let Assumptions 1–4, 5a, and 6a hold for the model of equations (7) and (8). Define* $\phi_{220} = a^2_{22,10} + a^2_{22,20} + b^2_{22,10} + b^2_{22,20}$. *The structural parameters of the triangular system together with* c_{20}, c_{30}, *and* ϕ_{220} *are identified.*

Given Lemma 1 and equation (9), 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 Assumption 3 defines $a_{11,2}$ and $b_{11,2}$ as nonnegative or, more generally, if some alternative form of conditional heteroskedasticity is assumed for $\epsilon_{1,t}$ besides that given in equation (9). Identification under Proposition 2 depends upon a constant conditional covariance between $\epsilon_{1,t}$ and $\epsilon_{2,t}$ and a GARCH(1,1) process for $\epsilon_{2,t}$. The second-moment dynamics of $\epsilon_{1,t}$ do not play a substantative role. Theorem 1 of Lewbel (2004) reaches this 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. The empirical asset pricing literature supports the univariate GARCH(1,1) model of Bollerslev (1986) as a reasonable parameterization of the conditional heteroskedasticity in stock returns.[19] Equation (9) is a natural extension of this model.

From a more theoretical standpoint, Nelson (1992) examines the ability of misspecified ARCH models to consistently estimate the conditional covariance matrix of certain stochastic processes. He finds 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.

3 Estimation

This section discusses estimation of equations (7) and (8) and partial estimation of equation (9). Two estimation routines are considered. The first is a three-step procedure involving simple OLS regression and standard univariate GARCH and ARMA models. The second is full GMM. The first routine can be used to obtain starting values for the second. For the first routine, define

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

[19]See, for example, Akgiray (1989) and Kim and Kon (1994).

$$\epsilon_{2,t} = Y_{2,t} - X_t'\beta_2,$$

and

$$R_{1,t} = \epsilon_{1,t} + \epsilon_{2,t}\gamma.$$

Consider the following composite parameters from the BEKK model of equation (9) and Assumption 3:

$$\eta_{12,1} = \left(a_{11,2}a_{22,2}\right), \quad \eta_{12,4} = \left(b_{11,2}b_{22,2}\right),$$

$$\eta_{22,1} = \left(a_{22,1}^2 + a_{22,2}^2\right), \quad \eta_{22,2} = \left(b_{22,1}^2 + b_{22,2}^2\right).$$

Let ψ be the set of parameters $\{\beta_1,\ \beta_2,\ \gamma,\ c_2,\ c_3,\ \eta_{12,1},\ \eta_{22,1},\ \eta_{22,2},\ \eta_{12,4}\}$. Finally, define

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

Corollary 1. *Let Assumptions 1–7 or 1–4, 5a, and 6a hold for the model of equations (7) and (8). Define* $\epsilon_{1,t}$, $\epsilon_{2,t}$, $R_{1,t}$, ψ, and $h_{22,t}$ *as above. The following steps consistently estimate the structural parameters of the triangular system together with* c_{20}, c_{30}, ϕ_{120}, *and* ϕ_{220}*: (1) Regress* $Y_{1,t}$ *on* X_t, *(2) Apply maximum likelihood (ML) to equation (8), specifying a univariate GARCH(*$1,1$*) model for* $\epsilon_{2,t}$, *(3) Apply ML to an ARMA(*$1,1$*) model of* $\widehat{R}_{1,t}\widehat{\epsilon}_{2,t}$ *with weakly exogenous covariates* $\widehat{\epsilon}_{2,t-1}^2$ *and* $\widehat{h}_{22,t-1}$.

Corollary 1 nests the results of Propositions 1 and 2. The advantage of Corollary 1 is that it can be implemented with conventional time-series software. The disadvantage is that convergence could be an issue, since the composite parameters governing the ARMA components in step 3 are likely to be of very similar magnitudes.[20] In addition, standard errors

[20] From equation (40) of A.4, $\left(\eta_{12,1} + \eta_{12,4}\right)$ is the AR component, and $\eta_{12,4}$ is the MA component. The parameters $\eta_{12,1}$ and $\eta_{12,4}$ are, respectively, the ARCH and GARCH components of $h_{12,t}$. Typically, the

are 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.[21] Alternatively, bootstrapped standard errors can be estimated for step 3. Separately, if $\epsilon_{2,t}$ is asymmetrically distributed (skewed), a "location parameter" needs to be added to the GARCH specification in step 2, following Newey and Steigerwald (1997).

For the second routine, let $\epsilon_t = \begin{bmatrix} \epsilon_{1,t} & \epsilon_{2,t} \end{bmatrix}'$ and $\overline{e}_t = \begin{bmatrix} \epsilon_{1,t}\epsilon_{2,t} & \epsilon_{2,t}^2 \end{bmatrix}'$. Redefine ψ as the set of parameters $\{\beta_1,\ \beta_2,\ \gamma,\ c_2,\ c_3,\ \phi_{12},\ \phi_{22}\}$. In addition, let $\sigma_{\overline{e}} = \begin{bmatrix} \sigma_{12} & \sigma_{22} \end{bmatrix}'$, where

$$\sigma_{12} = \frac{c_2 c_3}{1-\phi_{12}}$$

and

$$\sigma_{22} = \frac{c_3^2}{1-\phi_{22}}.$$

Finally, let $\overline{\Phi} = \begin{bmatrix} \phi_{12} & 0 \\ 0 & \phi_{22} \end{bmatrix}$. Consider the following set of vector functions:

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

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

$$U_3\left(\psi,\ Y_t,\ S_{t-1}\right) = vec\left[\left(\overline{e}_t - \sigma_{\overline{e}}\right)\left(\overline{e}_{t-2} - \sigma_{\overline{e}}\right)' - \overline{\Phi}\left(\overline{e}_t - \sigma_{\overline{e}}\right)\left(\overline{e}_{t-1} - \sigma_{\overline{e}}\right)'\right].$$ [22]

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

ARCH component tends to be small, rendering $(\eta_{12,1} + \eta_{12,4})$ and $\eta_{12,4}$ comparable. In addition, the GARCH component tends to be large, placing $(\eta_{12,1} + \eta_{12,4})$ and $\eta_{12,4}$ near one.

[21] See Newey and McFadden (1994).

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

Corollary 2. *Let either Assumptions 1–7 or 1–4, 5a, and 6a hold for the model of equations (7) and (8). 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 Propositions 1 and 2 into a single set of moment conditions. $E\left[U_1\right] = 0$ relates to the conditional means of equations (7) and (8). $E\left[U_2\right] = 0$ defines the unconditional covariance of $\epsilon_{1,t}$ and $\epsilon_{2,t}$ as well as the unconditional variance of $\epsilon_{2,t}$. $E\left[U_3\right] = 0$ describes the time-variation of $h_{12,t}$, nesting a constant covariance as a special case, and of the GARCH(1,1) process of $\epsilon_{2,t}$.

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

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

for some positive definite $\widehat{W}$. Let $W_0 = E\left[U\left(\psi,\ Y_t,\ S_{t-1}\right) U\left(\psi,\ Y_t,\ S_{t-1}\right)'\right]$. If the elements of S_{t-1} are stationary and ergodic, and if $\widehat{W} \xrightarrow{p} W_0$, then the resulting GMM estimator is consistent with

$$\sqrt{T}\left(\widehat{\psi} - \psi_0\right) \xrightarrow{d} N\left(0, \left(E\left[\frac{\partial U\left(\psi,\ Y_t,\ S_{t-1}\right)}{\partial \psi}\right]' W_0^{-1} E\left[\frac{\partial U\left(\psi,\ Y_t,\ S_{t-1}\right)}{\partial \psi}\right]\right)^{-1}\right).$$

Efficiency gains result if the set of vector functions

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

is appended to U (West (2002) discusses this result in the context of AR processes with GARCH errors). These functions involve higher-order autocovariances from the GARCH process. Their addition provides two tests for the fit of $h_{12,t}$ and $h_{22,t}$. First, a joint test of whether ϵ_t is mean zero conditional on S_{t-1} and whether $h_{12,t}$ and $h_{22,t}$ are properly specified is afforded by Hansen's J statistic. Second, testing whether $E\left[U_{3+p}\right] = 0$ with a χ^2 difference test judges the fit of $h_{12,t}$ and $h_{22,t}$ alone.

Apparent from the above discussion, there are a large number of potential moment conditions available for estimating equations (7)–(9). While efficiency involves more moments, the use of additional moments can degrade the small-sample properties of the GMM estimator, as shown by Newey and Smith (2001). Donald, Imbens, and Newey (2002) provide mean-squared error (MSE) based criteria for selecting Q. If Corollary 1 is used to obtain $\widehat{h}_{12,rt}$ and $\widehat{h}_{22,t}$, then lags of these estimates together with lags of $\widehat{R}_{1,t}\widehat{\epsilon}_{2,t}$ and $\widehat{\epsilon}^2_{2,t}$ may be used as the instruments in those criteria. Optimal lag selection follows, since the reduced form inherits the lag-order of its structural counterpart. The downside to this procedure is that the effects of generated regressors on the asymptotics of these criteria are unclear. Future research may look to fill this gap.

Moments-based estimators with better small-sample properties (for example, Generalized Empirical Likelihood) can be used instead of GMM to estimate the result of Corollary 1.[23] If the moment conditions are weak, then the alternative distribution theory of Stock and Wright (2000) is applicable. The model and moment selection criteria (MMSC) of Andrews and Lu (1999) can be used to investigate the possibility of structural breaks in the conditional mean equations and the conditional covariance matrix.[24] Finally, the GARCH specification for $\epsilon_{2,t}$ can be evaluated using the tests developed by Lundbergh and Teräsvirta (2002).

[23] See Newey and Smith (2004) as a reference on Generalized Empirical Likelihood (GEL).

[24] See Rigobon (2002) for an example related to sovereign debt.

4 Monte Carlo

The Monte Carlo simulations draw data from the structural model

$$Y_{1,t} = X_{1,t}\left(\delta_1 - \beta_1\gamma\right) + X_{2,t}\left(\delta_2 - \beta_2\gamma\right) + Y_{2,t}\gamma + \epsilon_{1,t}, \tag{12}$$

$$Y_{2,t} = X_{1,t}\beta_1 + X_{2,t}\beta_2 + \epsilon_{2,t}, \tag{13}$$

which corresponds to the single-factor model of section 1.1 with two forecasting instruments. The scalars $X_{1,t}$ and $X_{2,t}$ are scaled standard normal draws with a correlation of 0.28 and first-order autocorrelations of 0.35 and 0.13, respectively. Let $\epsilon_t = \begin{bmatrix}\epsilon_{1,t} & \epsilon_{2,t}\end{bmatrix}'$. The vector ϵ_t is constructed so that $\epsilon_t \mid S_{t-1} \sim N\left(0,\ H_t\right)$, where H_t is parameterized according to equation (9) and Assumption 3. Let

$$\Delta = \begin{bmatrix} \delta_1 & \delta_2 & \beta_1 & \beta_2 & \gamma \end{bmatrix},$$

the vector of parameters from equations (12) and (13), and

$$\Sigma = \begin{bmatrix} c_2 & c_3 & \phi_{12} & \phi_{22} \end{bmatrix},$$

a vector of parameters from Assumption 3. Simulations consider two sets of values for Δ and Σ. The first selects

$$\Delta = \begin{bmatrix} 2.50 & 8.90 & 6.60 & 7.30 & 1.20 \end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix} -0.004 & 0.01 & 0.75 & 0.90 \end{bmatrix},$$

while the second selects

$$\Delta = \left[\begin{array}{ccccc} 5.50 & 6.70 & 6.60 & 7.30 & 0.80 \end{array} \right]$$

and

$$\Sigma = \left[\begin{array}{cccc} 0.006 & 0.01 & 0.75 & 0.90 \end{array} \right].$$

For each of the two sets, simulations are conducted with $Q = 6, 12, 24$, for a total of six Monte Carlo experiments.

The Monte Carlo design attempts to replicate the data environment that is encountered in the empirical exercise of section 5.2. All parameter values and correlations are estimated from the data set described in that section. Values for Δ and Σ are the product of Corollary 1. The first set of values for Δ and Σ consider $Y_{1,t}$ as the return on a portfolio of the smallest capitalization-based decile of NYSE/AMEX stocks and $Y_{2,t}$ as the return on an equal-weighted portfolio of all NYSE/AMEX stocks. The second set considers $Y_{1,t}$ as the return on a portfolio of the largest capitalization-based decile of NYSE/AMEX stocks; $Y_{2,t}$ receives identical treatment as before. Across both sets, $X_{1,t}$ and $X_{2,t}$ are single lagged values of the return spread between two- and one-month Treasury bills as well as the return spread between the lagged two-month and current one-month Treasury bills.[25]

Tables 1 and 2 show results from Monte Carlo simulations of the GMM estimator applied to Corollary 2. In these simulations, the first 200 observations of each series are discarded to avoid initialization effects. Simulations are conducted with $T = 370$ observations across 1000 trials. The number of observations conforms with the sample size used to test the CAPM in section 5. The maximum lag length, Q, used in defining the moment conditions is varied to investigate both the efficiency gains discussed in West (2002) and the size of the bias studied

[25] The scaling of $X_{1,t}$ and $X_{2,t}$ in the simulation exercises adjusts the variances of each to match what is found in the data.

by Newey and Smith (2001). The parameter values for the simulations satisfy Assumptions 3–5, and 7. 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.

In general, the simulation results provide an encouraging assessment of the proposed estimator. The reported biases are of comparable magnitudes to those found by other researchers for standard GMM when applied to data sets of a similar size (see, e.g., Donald, Imbens, and Newey, 2002). The simulation results also support the findings of West (2002). Higher values of Q tend to be associated with lower root mean-squared errors (RMSEs) and tighter interquartile ranges. For all parameter estimates, the mean absolute and median absolute errors tend to decrease with Q, implying that, net of outliers, precision in those estimates depends positively on lag length. In addition, for the parameters in Σ, the biases (both mean and median) diminish as the lag length increases. For the parameters in Δ, on the other hand, longer lag lengths tend to be associated with higher biases, evidencing the results of Newey and Smith (2001). For the estimates of γ, the sign of these biases follows the sign of the unconditional covariance between $\epsilon_{1,t}$ and $\epsilon_{2,t}$. The relative size of these biases, however, tends to be small.

5 Testing the CAPM

This section is divided into three parts. By stating a lemma upon which two corollaries to Propositions 1 and 2 in Shanken (1987) are based, section 5.1 develops a test statistic for the CAPM that recognizes the possible endogeneity of market proxies and relies upon a prior belief about the correlation between innovations to a given proxy and the true market return. Using Corollary 2, section 5.2 estimates the CAPM-style model of section 1.1 with 10 capitalization-based portfolios of stock returns and shows that the resulting market sensitivities can differ significantly from those estimated by OLS. Finally, using the estimates from section 5.2, section 5.3 bootstraps the test statistic developed in section 5.1 to determine what values of ρ, the correlation between innovations to the given proxy and the true market return, reject the CAPM.

5.1 Methodology

For expository convenience, all time subscripts are suppressed. Variable labels follow Shanken (1987), to facilitate comparison. As in section 1.1, assume there exists an observable risk-free rate. Let r_m^* be the true market return, R_p an L-vector of observable proxies to the market return, X a K-vector of predetermined instruments that forecast security returns, R an N-vector of security returns, and $E = \begin{bmatrix} m & P' & e' \end{bmatrix}'$, an $(L+N+1)$-vector of shocks. The sigma field S is defined by X and its past values, as well as past values of R_p, R, and E. Consider the following models for the true market return and its observable proxies:

$$r_m^* = a_{m^*} + b_{m^*} X + m, \tag{14}$$

$$R_p = a_p + B_p X + P. \tag{15}$$

Assume that $E[m \mid S] = E[P \mid S] = 0$. Equations (14) and (15) decompose the true market return and its observable proxies into predictable and unpredictable components. Consider a linear regression of m on P,

$$m = a_m + b_m' P + e_m, \tag{16}$$

and the following model of security returns:

$$R = a + B_R X + \Gamma P + e, \tag{17}$$

where $E[e \mid S] = 0$. Equation (17), expressed in excess returns, is a vector statement of equation (3).

Lemma 2: *Consider equations (14)–(17) and their accompanying assumptions. Then,*

$$cov\,(e,\ e_m)'\,\Sigma_e^{-1} cov\,(e,\ e_m) \leq \sigma^2\,(m)\,(1-\rho^2), \tag{18}$$

where Σ_e is the $N \times N$ unconditional covariance matrix of e, $\sigma^2\,(m)$ the variance of m, and ρ the correlation between m and P. Equation (18) 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 two differences between Lemma 2 and Lemma 1 of Shanken (1987) are that, for the former, security returns are allowed a predictable component, and $cov\,(e,\ P)$ is not assumed to be a zero matrix. Neither of these differences affects the proof. Shanken (1987) interprets $cov\,(e,\ e_m)$ "as a vector of deviations from an exact multibeta expected return relation" (p. 93). Equation (18) sets an upper bound for these deviations.

Corollary 3. *Assume*

$$E\,[R] = r1_N + cov\,(R,\ m)\,, \tag{19}$$

where 1_N is an N-vector of ones. Let r be the observable risk-free rate. Then, there exists an L-vector of "prices of risk" that satisfies

$$d'\Sigma_e^{-1} d \leq \sigma^2\,(m)\,(1-\rho^2), \tag{20}$$

where

$$d \equiv E\,[R] - r1_N - (\Gamma + B_e)\,\delta, \tag{21}$$

B_e is the vector of slope parameters from a multivariate linear regression of e on P. $\delta = cov\,(P,\ m)$ satisfies equation (20).

Corollary 3 generalizes Proposition 1 of Shanken (1987) to recognize the possible endogeneity of market proxies. The risk-return relation of equation (19) holds with a constant of proportionality equal to $\frac{E[r_m^* - r]}{Var[m]}$. In addition, if $\rho = 1$ in equation (20), then $d = 0$, meaning that $E[R]$ is exactly linear in the columns of $(\Gamma + B_e)$.

The next corollary takes the pricing relation of equation (19) and applies it to the vector of market proxies. For use in this corollary, let

$$\theta_p^2 = \left(E\left[R_p\right] - r1_L\right)' \Sigma_p^{-1} \left(E\left[R_p\right] - r1_L\right), \tag{22}$$

where Σ_p is the unconditional covariance matrix of P. Equation (22) describes a Sharpe performance measure in terms of the covariance matrix of proxy innovations P, not proxy returns R_p, as in the usual case. The separating of returns into predictable and unpredictable components establishes this distinction.

Corollary 4. *Assume*

$$E\left[R_p\right] = r1_L + cov\left(R,\ m\right). \tag{23}$$

Then,

$$\rho^2 = \frac{\theta_p^2}{\sigma^2\left(m\right)},$$

and the result of Corollary 3 reduces to

$$d'\Sigma_e^{-1}d \leq \theta_p^2(\rho^{-2} - 1), \tag{24}$$

where

$$d \equiv E\left[R\right] - r1_N - \left(\Gamma + B_e\right)\left(E\left[R_p\right] - r1_L\right). \tag{25}$$

Corollary 4 applies the same modification introduced in Corollary 3 to Proposition 2 of

Shanken (1987). As is the case for Proposition 2, the power of Corollary 4 resides in the fact that, except for ρ, every parameter in equation (24) can be estimated from observable data. Given Corollary 4, a conditional form of the CAPM prices security returns. The presence of X in equation (17) designates expected returns as time-varying. According to Corollary 4, these time-varying expected returns are a linear function of the market risk premium. As a result, Corollary 4 is a stronger statement about security returns than its counterpart, Proposition 2 in Shanken (1987).

5.2 Gamma Estimates

This section considers estimation of Γ in equation (17) using the result of Corollary 2. Let R_p be the return on either the equal-weighted or value-weighted NYSE/AMEX stock index.[26] The security returns are 10 capitalization-based portfolios, with the first portfolio containing the smallest decile of stocks on the NYSE/AMEX, and the last containing the largest. The forecasting instruments are the return spread between two- and one-month Treasury bills as well as the return spread between the lagged two-month and current one-month Treasury bills.[27] All stock return data are from the Center for Research in Security Prices (CRSP). The forecasting instruments are from Fama's six-month bill files. The data set covers the period February 1953 through November 1983, the same period considered by Shanken (1987).

As mentioned in section 5.1, equation (17), expressed in excess returns, is a vector statement of equation (3). Therefore, sequentially estimating the model

$$Y_{i,t} = X_t' \left(\delta_i - \beta_m \gamma_i\right) + Y_{m,t}\gamma_i + \epsilon_{i,t},\ i = 1, \ldots, 10, \tag{26}$$

[26]The ensuing empirical tests consider only the case where $L = 1$ (see the general definition of R_p in section 5.1), so R_p is a scalar random variable.

[27]These instruments, taken from Campbell (1987), are considered because they survive the spurious regression critique of Ferson, Sarkissian, and Simin (2003).

and

$$Y_{m,t} = X_t^{'}\beta_m + \epsilon_{m,t}, \tag{27}$$

where $Y_{i,t}$ is the excess return on the ith capitalization-based portfolio; $Y_{m,t}$ is the excess return on either the equal-weighted or value-weighted NYSE/AMEX stock index, and X_t is a vector of predetermined covariates including a constant, the two aforementioned forecasting instruments, and a dummy variable for the January effect, produces consistent estimates of the market sensitivities (γ_i).

Tables 3 and 4 summarize the results of estimating equations (26) and (27) under Corollary 2, in the case where the market proxy is, respectively, the equal-weighted and value-weighted NYSE/AMEX stock index. Starting values for the GMM estimators were derived using Corollary 1. Estimation sets $Q = 24$. Though the simulation study (section 4) does document some bias in the estimation of γ using higher lag orders, the relative size of that bias is small. In addition, the simulation study also measures some rather sizable efficiency gains across higher lag orders for parameters from the GARCH process. Since the parameters describing the conditional covariance of $\epsilon_{i,t}$ and $\epsilon_{m,t}$ speak directly for or against the consistency of alternative estimators (see the literature review of section 1.2), the higher lag order was chosen.

All gamma estimates are highly significant and of plausible magnitudes. For many of the portfolios, significant values of c_2 evidence a nonzero covariance between $\epsilon_{i,t}$ and $\epsilon_{m,t}$. In general, the sign of that covariance tends to be negative for smaller-capitalization and positive for larger-capitalization stocks. A nontrivial subset of these covariances also appears to be time-varying, as evidenced by significant values for $\widehat{\phi}_{12}$. This time variation seems to be a characteristic of larger-capitalization stocks; however, there is an isolated instance of time variation in the smaller-capitalization issues. The values of $\widehat{\phi}_{22}$ for either proxy reinforce the numerous findings of strong GARCH effects in the innovations to security returns. Finally, none of the time-series models for portfolio returns are rejected according to their J statistics.

Tables 5 and 6 compare the estimates of individual market sensitivities obtained under GMM using the moments from Corollary 2 with those obtained under OLS by simply regressing $Y_{i,t}$ on X_t and $Y_{m,t}$, ignoring any possible endogeneity. The comparison is based on 95-percent confidence intervals constructed from the GMM estimates.[28] As a whole, the results support the endogeneity of the two market proxies. For the equal-weighted proxy, only the OLS estimate for decile portfolio 3 (CAP3) lies inside the associated confidence interval. For the value-weighted proxy, decile portfolios 1, 2, and 4–7 have OLS estimates inside the associated confidence intervals. The result for decile portfolio 4 should be viewed with caution, however, since the hypothesis of a zero covariance is rejected. For all the portfolios for which the GMM estimates differ significantly from their OLS counterparts, the direction of the difference follows the sign of the unconditional covariance between $\epsilon_{i,t}$ and $\epsilon_{m,t}$. Given the properties of the OLS estimator, this result is not surprising.

The implications of the results summarized in Tables 5 and 6 extend beyond tests of the CAPM. For instance, let $Y_{i,t}$ be the return on any portfolio of securities, and interpret γ_i to be the risk-minimizing hedge ratio for the market risk of that portfolio. Traditional practice estimates γ_i from a regression of $Y_{i,t}$ on X_t and $Y_{m,t}$. This section demonstrates that such a practice is likely to result in an inconsistent estimate of the desired hedge ratio. The identification results in this paper, therefore, bear significance on the hedging of portfolio risks.

5.3 Test Results

Let $N = 10$ and $\bar{\epsilon}_t = \begin{bmatrix} \epsilon_{1,t} & \ldots & \epsilon_{N,t} \end{bmatrix}'$, the vector of innovations from equation (26). Equation (25) represents the pricing errors from a cross-sectional GLS regression of

[28] Given Proposition 2, the GMM estimates are consistent if $E\left[\epsilon_{i,t}\epsilon_{m,t} \mid S_{t-1}\right] = 0$.

$(E[R] - r1_N)$ on $(\Gamma + B_e)$, using Σ_e as the error covariance matrix. Consider $\widehat{d}$ to be the estimated errors from such a regression, with $\widehat{\Gamma}$ estimated as in section 5.2, $\widehat{B}_e$ equal to the vector of slope estimates from N separate regressions of $\widehat{\epsilon}_{i,t}$ on $\widehat{\epsilon}_{2,t}$, $i = 1, \ldots, N$, and $\widehat{\Sigma}_e = \frac{1}{T}\sum_{t=1}^{T}\widehat{\epsilon}_t\widehat{\epsilon}_t'$. Likewise, consider $\widehat{\theta}_p$ to be an estimator of the proxy performance measure in equation (22).

The pricing theory of Corollary 4 involves two cases. If $\rho = 1$, meaning the given proxy is the true market portfolio, then $d'\Sigma_e^{-1}d = 0$. If, on the other hand, $0 < \rho < 1$, then $\theta_p^{-2}d'\Sigma_e^{-1}d \leq \rho^{-2} - 1$. Shanken (1987) tests a special case of Corollary 4, by using a noncentral F distribution.[29] In general, such a distribution will not hold for Corollary 4. As a result, consider testing Corollary 4 for $\rho = 1$ by bootstrapping a standard error for $\widehat{d}'\widehat{\Sigma}_e^{-1}\widehat{d}$, using $\widehat{\epsilon}_t$. Table 7 summarizes the results of this test for both the equal-weighted and value-weighted proxies. In both cases, the null of $d'\Sigma_e^{-1}d = 0$ is rejected, meaning that if the pricing theory of Corollary 4 holds, then neither proxy is the true market portfolio.

Next, consider finding the maximum value of $\rho \in (0, 1)$ that supports Corollary 4 by bootstrapping $\widehat{\theta}_p^{-2}\widehat{d}'\widehat{\Sigma}_e^{-1}\widehat{d}$, again using $\widehat{\epsilon}_t$. Table 8 summarizes the results of this test for both the equal-weighted and value-weighted proxies. Under Corollary 4, $\rho^{-2} - 1$ is the upper bound for $\theta_p^{-2}d'\Sigma_e^{-1}d$. Table 8 reports the mass of $\widehat{\theta}_p^{-2}\widehat{d}'\widehat{\Sigma}_e^{-1}\widehat{d}$ lying above $\rho^{-2} - 1$, for different values of ρ. Five percent of the bootstrapped distribution of $\widehat{\theta}_p^{-2}\widehat{d}'\widehat{\Sigma}_e^{-1}\widehat{d}$ lies above $\rho = 0.62$ for the equal-weighted, and $\rho = 0.49$ for the value-weighted market proxy. When interpreted in terms of a standard 5-percent significance level, these results reject the hypothesis that the true correlation between innovations to the equal-weighted proxy and the true market return exceeds 0.62. In terms of the relation between innovations to the value-

[29] Shanken's special case assumes:

1. $b_{m^*} = 0_K$
2. B_p, B_R, and B_e are all zero matrices.

weighted proxy and the true market return, these results reject the correlation's being higher than 0.49. Therefore, if Corollary 4, and, hence, the CAPM, is a valid pricing relation, then innovations to the CRSP equal-weighted index account for only 38 percent $(0.62^2 = 0.38)$ of the variation in the innovations to the true market return. The explained variation of innovations to the true market return drops to 24 percent $(0.49^2 = 0.24)$ if the CRSP value-weighted index is used as the proxy. Given that correlations between market proxies and the true market portfolio are assumed to be quite high, in the range of 0.80 and 0.90, and that forecasting instruments explain a very small percentage of the total variation in security returns, with R^2 values ranging from 0.02 to 0.08, these results do not speak favorably for the CAPM.[30]

The test results presented here are similar to, though uniformly more negative than, what Shanken (1987) and Kandel and Stambaugh (1987) report. Shanken, for instance, finds that the CAPM can be rejected if the "multiple correlation between the true market portfolio and proxy assets exceeds 0.70" (p. 91). Kandel and Stambaugh reach substantially the same conclusion. The significantly lower maximum correlation for the value-weighted proxy reported here is somewhat surprising, despite the fact that a conditional version of the CAPM is being tested. Kandel and Stambaugh do, however, report slightly lower correlations for the value-weighted versus the equal-weighted proxy.

6 Conclusion

This paper presents a new method for identifying triangular systems of time-series data. Identification is the product of a diagonal GARCH process. Relative to the literature on GARCH-based identification, the method discussed in this paper distinguishes itself by both allowing for a time-varying conditional covariance and by not requiring complete estimation

[30] See Roll (1977) as supporting a high correlation between market proxies and the true market return. See Ferson, Sarkissian, and Simin (2003) for a summary of the R^2 values for different forecasting instruments.

of the GARCH parameters. Regarding this latter distinction, only two nuisance parameters from the GARCH model need to be considered. A Monte Carlo study verifies the consistency of the identification method. The method is then applied in testing a conditional version of the CAPM to find that the model of Sharpe (1964) and Lintner (1965) does not seem to adequately describe the cross-sectional variation of expected returns. These results provide a harsher critique of the CAPM than either Shanken (1987) or Kandel and Stambaugh (1987).

Section 5.1 develops a statistic for testing the CAPM that recognizes both the true market return as a latent variable and the potential for observable proxies to be endogenous regressors. Section 5.2 presents empirical evidence supporting the endogeneity of market proxies. As discussed in section 5.3, the test statistic derived in section 5.1 does not deliver favorable news for the CAPM. A key question is whether additional factors can be incorporated into the CAPM framework to avoid having to discard the CAPM theory completely. The intertemporal CAPM of Merton (1973) provides a potential answer. So, too, does the three-moment CAPM of Kraus and Litzenberger (1976), which includes skewness in the market return as a second factor. An interesting investigation would be whether a test statistic in the spirit of section 5.1 can be developed for testing Kraus and Litzenberger's version of the CAPM. The identification method discussed in this paper could be used to estimate such a statistic. The remaining details are left to future research.

Appendix

A.1. Proof of Proposition 1: Given Assumptions 2 and 3, write the equations for $h_{12,t}$ and $h_{22,t}$ in vector-form as

$$\overline{h}_t = C + A\overline{e}_{t-1} + B\overline{h}_{t-1}, \tag{28}$$

where

$$\overline{h}_t = \left[\begin{array}{cc} h_{12,t} & h_{22,t} \end{array} \right]', \quad \overline{e}_t = \left[\begin{array}{cc} \epsilon_{1,t}\epsilon_{2,t} & \epsilon_{2,t}^2 \end{array} \right]',$$

$$C = \left[\begin{array}{cc} c_{20}c_{30} & c_{30}^2 \end{array} \right]',$$

$$A = \left[\begin{array}{cc} a_{11,20}a_{22,20} & 0 \\ 0 & a_{22,10}^2 + a_{22,20}^2 \end{array} \right], \quad B = \left[\begin{array}{cc} b_{11,20}b_{22,20} & 0 \\ 0 & b_{22,10}^2 + b_{22,20}^2 \end{array} \right].$$

From Assumptions 1 and 2, the structural innovations to equation (8) are identified by equation (10), and the reduced form innovations to equation (7) are identified by equation (11). Using equations (10) and (11) to solve the reduced form of (28) yields

$$\overline{h}_{r,t} = C_r + A_r\overline{r}_{t-1} + B_r\overline{h}_{r,t-1}, \tag{29}$$

where

$$\overline{h}_{r,t} = \left[\begin{array}{cc} E\left[R_{1,t}\epsilon_{2,t} \mid S_{t-1}\right] & h_{22,t} \end{array} \right]', \quad \overline{r}_t = \left[\begin{array}{cc} R_{1,t}\epsilon_{2,t} & \epsilon_{2,t}^2 \end{array} \right]',$$

$C_r = \Gamma^{-1}C$, $A_r = \Gamma^{-1}A\Gamma$, $B_r = \Gamma^{-1}B\Gamma$, and $\Gamma = \left[\begin{array}{cc} 1 & -\gamma_0 \\ 0 & 1 \end{array} \right]$. Given Assumption 4, recursive substitution into (29) reveals

$$\overline{h}_{r,t} = (I - B_r)^{-1} C_r + \sum_{j=1}^{\infty} B_r^{j-1} A_r \overline{r}_{t-j}, \tag{30}$$

where I is the (2×2) identity matrix. Equation (29) implies that $\overline{r}_t = \overline{h}_{r,t} + w_{r,t}$, with $E\left[w_{r,t} \mid S_{t-1}\right] = 0$. As a result,

$$cov\left(\overline{r}_t,\ \overline{r}_{t-k}\right) = cov\left(\overline{h}_{r,t},\ \overline{r}_{t-k}\right), \tag{31}$$

for $k = 1, \ldots, \infty$. Substituting (30) into (31) and simplifying for $k = 1, 2$ shows that

$$cov\left(\overline{r}_t,\ \overline{r}_{t-2}\right) = \left(A_r + B_r\right) cov\left(\overline{r}_t,\ \overline{r}_{t-1}\right).$$

Since

$$cov\left(\overline{r}_t,\ \overline{r}_{t-1}\right) = \Gamma^{-1} cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right)\left(\Gamma^{-1}\right)',$$

$\left(A_r + B_r\right)$ is identified by Assumption 6. Let $\delta_{ij,r}$ be the element in the ith row and jth column of

$$\Delta = \left(A_r + B_r\right) = \left[\begin{array}{cc} \phi_{120} & \gamma_0\left(\phi_{220} - \phi_{120}\right) \\ 0 & \phi_{220} \end{array}\right]. \tag{32}$$

Given Assumption 7, γ_0 is identified by $\gamma_0 = \frac{\delta_{12,r}}{\delta_{22,r} - \delta_{11,r}}$. As a result, 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 $\epsilon_{1,t} = Y_{1,t} - X_t'\beta_{10} - Y_{2,t}\gamma_0$. The composite parameters ϕ_{120} and ϕ_{220} are identified from the matrix $(A + B) = cov\left(\overline{e}_t,\ \overline{e}_{t-2}\right) cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right)^{-1}$. The vector of constants C is identified by $C = (I - (A + B)) E\left[\overline{e}_t\right]$. Given Assumption 3, c_{30} is identified. Finally, c_{20} is identified given the identification of c_{30} and the fact that c_{30} is strictly positive.

A.2. Proof of Lemma 1: Suppose $a_{11,20} = b_{11,20} = 0$. Then, $h_{12,t} = c_{20}c_{30}$, and $cov\left(\epsilon_{1,t}^2,\ \epsilon_{i,t-1}\epsilon_{j,t-1}\right) = cov\left(\epsilon_{1,t}\epsilon_{2,t},\ \epsilon_{i,t-1}\epsilon_{j,t-1}\right) = 0\ \forall\, i, j = 1, 2$. As a result, $\left|cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right)\right| = 0$. The same result holds if $a_{22,20} = b_{22,20} = 0$.

A.3. Proof of Proposition 2: Given Assumption 5a, $h_{12,t} = c_{20}c_{30}$. As a result,

$$cov(\epsilon_{1,t}\epsilon_{2,t},\ Z_{t-1}) = 0, \tag{33}$$

where $Z_{t-1} = \begin{bmatrix} \epsilon_{2,t-1}^2 & \cdots & \epsilon_{2,t-l}^2 \end{bmatrix}'$ for some finite $l \geq 1$. Assumptions 1 and 2 identify $\epsilon_{2,t}$ and $R_{1,t}$ as equations (10) and (11), respectively. From equation (11), write $\epsilon_{1,t}$ as

$$\epsilon_{1,t} = R_{1,t} - \epsilon_{2,t}\gamma_0. \tag{34}$$

Substituting equation (34) into equation (33) yields

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

where the individual row entries of $cov(\epsilon_{2,t}^2,\ Z_{t-1})$ are finite and nonzero given Assumptions 2, 3, and 6a. Let $\Lambda = cov(\epsilon_{2,t}^2,\ Z_{t-1})$. γ_0 is identified as

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

β_{10} is identified by $\beta_{10} = E\left[X_t X_t'\right]^{-1} E\left[X_t\left(Y_{1,t} - Y_{2,t}\gamma_0\right)\right]$, and 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 $\epsilon_{1,t} = Y_{1,t} - X_t'\beta_{10} - Y_{2,t}\gamma_0$. From equation (9), recursive substitution into $h_{22,t}$ reveals that

$$h_{22,t} = \frac{c_{30}^2}{1 - \left(b_{22,10}^2 + b_{22,20}^2\right)} + \left(a_{22,10}^2 + a_{22,20}^2\right)\sum_{j=1}^{\infty}\left(b_{22,10}^2 + b_{22,20}^2\right)^{j-1}\epsilon_{2,t-j}^2. \tag{35}$$

Since

$$cov(\epsilon_{2,t}^2,\ \epsilon_{2,t-k}^2) = cov(h_{22,t},\ \epsilon_{2,t-k}^2), \tag{36}$$

for $k = 1, \ldots, \infty$ (see the derivation of equation (31) in A.1.), substituting equation (35) into equation (36) and simplifying for $k = 1, 2$ grants identification of ϕ_{220} as $\phi_{220} = \frac{cov(\epsilon_{2,t}^2,\, \epsilon_{2,t-2}^2)}{cov(\epsilon_{2,t}^2,\, \epsilon_{2,t-1}^2)}$. Given Assumption 3, c_{30} is then identified as $c_{30} = \sqrt{(1-\phi_{220})\, E\left[\epsilon_{2,t}^2\right]}$. Finally, c_{20} is identified by $c_{20} = \frac{E[\epsilon_{1,t}\epsilon_{2,t}]}{c_{30}}$.

A.4. Proof of Corollary 1: Given equation (9) and Assumption 3,

$$h_{12,t} = (c_{20}c_{30}) + (\eta_{12,10})\, \epsilon_{1,t-1}\epsilon_{2,t-1} + (\eta_{12,40})\, h_{12,t-1} \tag{37}$$

and

$$h_{22,t} = (c_{30}^2) + (\eta_{22,10})\, \epsilon_{2,t-1}^2 + (\eta_{22,20})\, h_{22,t-1}. \tag{38}$$

The structural innovations to equation (8) are identified by equation (10), while the reduced-form innovations to equation (7) are identified by equation (11). Using equations (10) and (11) to solve the reduced-form of equation (37) yields

$$h_{12,rt} = (c_{120}) + (\eta_{12,10})\, R_{1,t-1}\epsilon_{2,t-1} + (\eta_{12,20})\, \epsilon_{2,t-1}^2 + (\eta_{12,30})\, h_{22,t-1} + (\eta_{12,40})\, h_{12,rt-1}, \tag{39}$$

where

$$h_{12,rt} = E\left[R_{1,t}\epsilon_{2,t} \mid S_{t-1}\right], \quad c_{120} = \left(c_{20}c_{30} + \gamma_0 c_{30}^2\right),$$

$$\eta_{12,20} = \gamma_0\left(\eta_{22,10} - \eta_{12,10}\right),$$

and

$$\eta_{12,30} = \gamma_0\left(\eta_{22,20} - \eta_{12,40}\right).$$

Equation (39) implies that

$$R_{1,t}\epsilon_{2,t} = (c_{120}) + (\eta_{12,10} + \eta_{12,40})\, R_{1,t-1}\epsilon_{2,t-1} + f\left(S_{t-1},\, \psi_0\right) - (\eta_{12,40})\, w_{12,rt-1} + w_{12,rt}, \tag{40}$$

where

$$f\left(S_{t-1},\ \psi_0\right) = \left(\eta_{12,20}\right)\epsilon_{2,t-1}^2 + \left(\eta_{12,30}\right)h_{22,t-1},$$

and

$$w_{12,rt} = R_{1,t}\epsilon_{2,t} - h_{12,rt}.$$

From equations (38) and (40), γ_0 is identified by

$$\gamma_0 = \frac{\eta_{12,20} + \eta_{12,30}}{\left(\eta_{22,10} + \eta_{22,20}\right) - \left(\eta_{12,10} + \eta_{12,40}\right)}.$$

β_{10} and β_{20} are then identified from A.1. ϕ_{120} and ϕ_{220} are identified as $\phi_{120} = \eta_{12,10} + \eta_{12,40}$ and $\phi_{220} = \eta_{22,10} + \eta_{22,20}$. Identification of C_0 follows from A.1. $R_{1,t}$ in equation (11) are the residuals from an OLS regression of $Y_{1,t}$ on X_t. Equations (8) and (38) are consistently estimated by joint ML. Given $\widehat{R}_{1,t}$, $\widehat{\epsilon}_{2,t}$, and $\widehat{h}_{22,t}$, equation (40) is also consistently estimated by ML.

A.5. Proof of Corollary 2: By equations (7)–(9), $U_1 = X_t \otimes \epsilon_t$ and

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

$E\left[U_2\right] = 0$ 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 $cov\left(\overline{e}_t,\ \overline{e}_{t-2}\right) = (A+B)\,cov\left(\overline{e}_t,\ \overline{e}_{t-1}\right)$, where $(A+B)$ is defined in A.1. Following, then, from either A.1 or A.3, the only $\varphi \in \Psi$ that satisfies $E\left[U\left(\psi, Y_t, S_{t-1}\right)\right] = 0$ is $\varphi = \varphi_0$.

A.6. Proof of Corollary 3: From equations (17),

$$cov\left(R,\ m\right) = \Gamma cov\left(P,\ m\right) + cov\left(e,\ m\right). \tag{41}$$

Given equation (16),

$$cov\left(e,\ m\right)=cov\left(e,\ P\right)\Sigma_{p}^{-1}cov\left(P,\ m\right)+cov\left(e,\ e_{m}\right), \tag{42}$$

where $b_{m}=\Sigma_{p}^{-1}cov\left(P,\ m\right)$, and Σ_{p} is the unconditional covariance matrix of P. Combining (41) and (42) yields

$$cov\left(R,\ m\right)=\left(\Gamma+B_{e}\right)cov\left(P,\ m\right)+cov\left(e,\ e_{m}\right), \tag{43}$$

where $B_{e}=cov\left(e,\ P\right)\Sigma_{p}^{-1}$. Substitution of (43) into equation (19) and the result into equation (18) of Lemma 2 produces equation (20), with $\delta=cov\left(P,\ m\right)$.

A.7. Proof of Corollary 4: From equation (15),

$$cov\left(R_{p},\ m\right)=cov\left(P,\ m\right),$$

and, therefore,

$$E\left[R_{p}\right]-r1_{L}=cov\left(P,\ m\right). \tag{44}$$

From equation (16),

$$\sigma^{2}\left(m\right)=b_{m}^{\prime}\Sigma_{p}b_{m}+\sigma^{2}\left(e_{m}\right). \tag{45}$$

Given the definition of b_{m} in A.6. and equation (44), (45) simplifies to

$$\sigma^{2}\left(m\right)=\theta_{p}^{2}+\sigma^{2}\left(e_{m}\right).$$

Hence, the coefficient of determination from equation (16) is $\rho^{2}=\frac{\theta_{p}^{2}}{\sigma^{2}(m)}$, and equation (20) in Corollary 2 reduces to (24) and (25).

References

[1] Akgiray, V. (1989), "Conditional Heteroskedasticity in Time Series of Stock Returns: Evidence and Forecasts," *Journal of Business*, 62, 55–80.

[2] Andrews, D.W.K. and B. Lu (1999), "Consistent Model and Moment Selection Critieria for GMM Estimation with Application to Dynamic Panel Data," unpublished manuscript.

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

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

[5] 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.

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

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

[8] Engle, R.F., D.F. Hendry and J.F. Richard (1983), "Exogeneity," *Econometrica*, 51, 277–304.

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

[10] Fama, E., L. Fisher, M.C. Jensen and R. Roll (1969), "The Adjustment of Stock Prices to New Information," *International Economic Review*, 10, 1–21.

[11] Ferson, W.E. (1985), "Changes Expected Risk Premiums and Security Risk Measures," unpublished manuscript.

[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] Ferson, W.E., S. Kandel and R.F. Stambaugh (1987), "Tests of Asset Pricing with Time-varying Expected Risk Premiums and Market Betas," *Journal of Finance*, 42, 201–220.

[14] Ferson, W.E., S. Sarkissian and T. Simin (2003), "Spurious Regressions in Financial Economics?" *Journal of Finance*, 58, 1393–1414.

[15] Gibbons, M.R. (1982), "Multivariate Tests of Financial Models: A New Approach," *Journal of Financial Economics*, 10, 3–27.

[16] Gibbons, M.R., S.A. Ross and J. Shanken (1987), "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57, 1121–1152.

[17] Hamilton, J.D. (1994), *Time Series Analysis*, Princeton University Press: Princeton, New Jersey, 659–660.

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

[19] Jobson, J.D. and B. Korkie (1982), "Potential Performance and Tests of Portfolio Efficiency," *Journal of Financial Economics*, 10, 433–466.

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

[21] Kim, D. and S.J. Kon (1994), "Alternative Models for the Conditional Heteroskedasticity of Stock Returns," *Journal of Business*, 67, 563–598.

[22] King, M., E. Sentana and S. Wadhwani (1994), "Volatility and Links Between National Stock Markets," *Econometrica*, 62, 901–933.

[23] 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.

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

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

[26] 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.

[27] Lundbergh, S. and T. Teräsvirta (2002), "Evaluating GARCH models," *Journal of Econometrics*, 110, 417–435.

[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] Merton, R.C. (1973), "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867–887.

[30] 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.

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

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

[34] Newey, W.K. and D.G. Steigerwald (1997), "Asymptotic Bias for Quasi-Maximum-Likelihood Estimators in Conditional Heteroskedastcity Models," *Econometrica*, 65, 587–599.

[35] Newey, W.K. and K.D. West (1987), "Hypothesis Testing with Efficient Method of Moments," *International Economic Review*, 28, 777–787.

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

[37] Prono, T. (2006), "GARCH-based Identification of Endogenous Regressors," unpublished manuscript.

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

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

[40] Roll, R. (1980), "Orthogonal Portfolios," *Journal of Financial and Quantitative Analysis*, 15, 1005–1023.

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

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

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

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

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

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

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

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

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

Table 1
Small-Cap Simulations

	δ_1	$\boldsymbol{\delta}_2$	γ	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	$\mathbf{c}_2$	$\mathbf{c}_3$	$\boldsymbol{\phi}_{12}$	$\boldsymbol{\phi}_{22}$
TRUE	2.50	8.90	1.20	6.60	7.30	−0.004	0.01	0.75	0.90
					Q = 6				
MEAN	2.69	8.67	1.17	6.71	7.17	−0.007	0.014	0.46	0.71
SD	4.90	2.96	0.28	2.55	1.67	0.020	0.007	0.50	0.27
LQ	−0.41	6.74	1.01	4.96	6.10	−0.011	0.009	0.26	0.61
MED	2.64	8.64	1.18	6.74	7.21	−0.003	0.013	0.57	0.79
UQ	5.86	10.74	1.33	8.28	8.19	0.000	0.018	0.76	0.89
RMSE	4.90	2.97	0.28	2.55	1.68	0.020	0.008	0.58	0.33
MAE	3.83	2.32	0.21	2.00	1.30	0.011	0.006	0.39	0.22
MDAE	3.12	1.99	0.16	1.67	1.06	0.005	0.005	0.23	0.11
					Q = 12				
MEAN	2.66	8.64	1.16	6.74	7.15	−0.006	0.011	0.48	0.80
SD	4.90	3.02	0.19	2.55	1.69	0.015	0.005	0.42	0.19
LQ	−0.44	6.58	1.04	4.97	6.00	−0.009	0.008	0.31	0.79
MED	2.57	8.61	1.16	6.80	7.16	−0.003	0.010	0.61	0.86
UQ	5.75	10.64	1.27	8.44	8.27	0.000	0.013	0.75	0.90
RMSE	4.90	3.03	0.19	2.55	1.70	0.015	0.005	0.50	0.22
MAE	3.85	2.37	0.15	2.02	1.33	0.008	0.003	0.33	0.12
MDAE	3.11	1.99	0.12	1.75	1.13	0.004	0.002	0.18	0.05
					Q = 24				
MEAN	2.73	8.62	1.15	6.71	7.13	−0.006	0.009	0.52	0.85
SD	4.79	3.00	0.13	2.47	1.63	0.010	0.004	0.40	0.16
LQ	−0.60	6.78	1.08	5.14	6.15	−0.008	0.007	0.42	0.85
MED	2.67	8.53	1.15	6.80	7.18	−0.003	0.009	0.65	0.90
UQ	5.90	10.50	1.23	8.23	8.17	−0.001	0.010	0.77	0.93
RMSE	4.79	3.01	0.13	2.47	1.63	0.011	0.004	0.46	0.17
MAE	3.81	2.32	0.10	1.89	1.24	0.006	0.003	0.28	0.08
MDAE	3.27	1.96	0.08	1.56	1.02	0.003	0.002	0.14	0.03

Notes: TRUE refers to the true parameter value. Q is the number of lagged terms used to define the moment conditions. MEAN and SD are the mean and standard deviation of the parameter estimates across the simulations. LQ is the lower quartile (bottom 25 percent); MED is the median, and UQ is the upper quartile (the top 25 percent). RMSE, MAE, and MDAE are the root mean-squared error, the mean absolute error, and the median absolute error of the estimates, respectively. Source: Author's calculations.

Table 2

Large-Cap Simulations

	δ_1	$\boldsymbol{\delta}_2$	γ	$\boldsymbol{\beta}_1$	$\boldsymbol{\beta}_2$	$\mathbf{c}_2$	$\mathbf{c}_3$	$\boldsymbol{\phi}_{12}$	$\boldsymbol{\phi}_{22}$
TRUE	5.50	6.70	0.80	6.60	7.30	0.006	0.01	0.75	0.90
					Q = 6				
MEAN	5.69	6.56	0.84	6.73	7.18	0.009	0.014	0.46	0.74
SD	3.52	2.15	0.16	2.65	1.71	0.015	0.007	0.50	0.25
LQ	3.43	5.14	0.76	5.08	6.12	0.001	0.009	0.21	0.68
MED	5.63	6.50	0.83	6.63	7.24	0.006	0.013	0.56	0.82
UQ	7.90	7.96	0.93	8.45	8.27	0.012	0.017	0.77	0.91
RMSE	3.52	2.15	0.16	2.66	1.71	0.016	0.008	0.58	0.29
MAE	2.73	1.67	0.12	2.04	1.31	0.009	0.006	0.39	0.19
MDAE	2.22	1.41	0.08	1.72	1.08	0.005	0.005	0.25	0.09
					Q = 12				
MEAN	5.67	6.48	0.86	6.76	7.11	0.009	0.011	0.49	0.83
SD	3.51	2.19	0.10	2.71	1.74	0.017	0.005	0.44	0.16
LQ	3.32	5.05	0.79	4.93	6.00	0.002	0.008	0.29	0.81
MED	5.63	6.49	0.85	6.72	7.09	0.005	0.010	0.63	0.88
UQ	7.96	7.90	0.91	8.51	8.31	0.010	0.013	0.79	0.92
RMSE	3.51	2.20	0.11	2.71	1.75	0.017	0.005	0.51	0.18
MAE	2.77	1.71	0.09	2.12	1.36	0.007	0.003	0.33	0.10
MDAE	2.30	1.45	0.07	1.77	1.14	0.004	0.002	0.18	0.04
					Q = 24				
MEAN	5.65	6.52	0.87	6.74	7.14	0.008	0.009	0.53	0.87
SD	3.33	2.09	0.08	2.59	1.68	0.010	0.004	0.39	0.14
LQ	3.57	5.27	0.82	5.06	6.12	0.003	0.007	0.39	0.87
MED	5.50	6.58	0.86	6.75	7.13	0.005	0.008	0.65	0.91
UQ	7.81	7.84	0.91	8.37	8.21	0.010	0.010	0.79	0.94
RMSE	3.34	2.10	0.10	2.60	1.68	0.010	0.004	0.45	0.15
MAE	2.57	1.60	0.08	1.99	1.27	0.006	0.003	0.29	0.07
MDAE	2.14	1.30	0.07	1.67	1.03	0.004	0.002	0.15	0.03

Notes: TRUE refers to the true parameter value. Q is the number of lagged terms used to define the moment conditions. MEAN and SD are the mean and standard deviation of the parameter estimates across the simulations. LQ is the lower quartile (bottom 25 percent); MED is the median, and UQ is the upper quartile (the top 25 percent). RMSE, MAE, and MDAE are the root mean-squared error, the mean absolute error, and the median absolute error of the estimates, respectively. Source: Author's calculations.

Table 3
Equal-Weighted Proxy

	γ	c_2	ϕ_{12}	ϕ_{22}	**J stat**	**p-value**
CAP1	1.36	−0.032	0.237	0.972	73.5	0.88
	(0.015)	(0.010)	(0.171)	(0.010)		
CAP2	1.24	−0.019	0.184	0.976	85.4	0.59
	(0.008)	(0.007)	(0.195)	(0.010)		
CAP3	1.10	0.001	0.071	0.950	75.5	0.84
	(0.007)	(0.002)	(0.341)	(0.012)		
CAP4	1.03	0.011	0.079	0.975	103.4	0.14
	(0.004)	(0.004)	(0.175)	(0.009)		
CAP5	0.95	0.024	−0.004	0.956	87.8	0.52
	(0.005)	(0.012)	(0.359)	(0.023)		
CAP6	0.97	0.001	0.945	0.972	84.2	0.63
	(0.005)	(0.000)	(0.012)	(0.008)		
CAP7	0.87	0.0097	0.404	0.978	73.7	0.88
	(0.006)	(0.0047)	(0.233)	(0.008)		
CAP8	0.63	0.118	−0.821	0.966	94.8	0.32
	(0.006)	(0.010)	(0.071)	(0.004)		
CAP9	0.47	0.097	−0.134	0.968	104.1	0.13
	(0.009)	(0.037)	(0.388)	(0.009)		
CAP10	0.41	0.014	0.815	0.979	87.6	0.52
	(0.009)	(0.004)	(0.038)	(0.008)		

Notes: CAPi refers to the value-weighted portfolio of common stocks in the ith size decile. Standard errors of the selected parameter estimates are reported in parentheses. J stat refers to Hansen's (1982) specification test chi-square statistic. The degrees of freedom for all reported J statistics are 89. The p-values for these J statistics are also reported. Source: Author's calculations.

Table 4

Value-Weighted Proxy

	γ	c_2	ϕ_{12}	ϕ_{22}	**J stat**	**p-value**
CAP1	1.27	−0.018	−0.269	0.913	82.2	0.68
	(0.044)	(0.011)	(0.180)	(0.026)		
CAP2	1.26	−0.014	−0.271	0.907	78.0	0.79
	(0.034)	(0.008)	(0.135)	(0.029)		
CAP3	1.33	−0.010	0.022	0.843	91.1	0.42
	(0.032)	(0.004)	(0.109)	(0.040)		
CAP4	1.27	−0.008	−0.028	0.860	95.4	0.30
	(0.028)	(0.004)	(0.129)	(0.040)		
CAP5	1.21	−0.001	−0.069	0.871	84.9	0.60
	(0.026)	(0.003)	(0.098)	(0.034)		
CAP6	1.17	0.000	−0.02	0.899	74.4	0.87
	(0.021)	(0.003)	(0.096)	(0.027)		
CAP7	1.13	0.001	0.33	0.914	82.2	0.68
	(0.018)	(0.002)	(0.101)	(0.024)		
CAP8	1.06	0.004	0.098	0.886	89.5	0.47
	(0.014)	(0.002)	(0.096)	(0.027)		
CAP9	1.11	−0.002	0.447	0.863	96.1	0.29
	(0.010)	(0.001)	(0.097)	(0.026)		
CAP10	0.88	0.005	0.469	0.853	106.7	0.10
	(0.008)	(0.002)	(0.128)	(0.046)		

Notes: CAPi refers to the value-weighted portfolio of common stocks in the ith size decile. Standard errors of the selected parameter estimates are reported in parentheses. J stat refers to Hansen's (1982) specification test chi-square statistic. The degrees of freedom for all reported J statistics are 89. The p-values for these J statistics are also reported. Source: Author's calculations.

Table 5
Equal-Weighted Proxy

	γ		95% C.I.	
Decile	**GMM**	**OLS**	**GMM**	
CAP1	1.36	1.24	1.33	1.39
CAP2	1.24	1.17	1.22	1.25
CAP3	1.10	1.12	1.09	1.11
CAP4	1.03	1.08	1.02	1.04
CAP5	0.95	1.07	0.94	0.95
CAP6	0.97	1.01	0.96	0.98
CAP7	0.87	0.93	0.85	0.88
CAP8	0.63	0.87	0.62	0.64
CAP9	0.47	0.78	0.46	0.49
CAP10	0.41	0.61	0.40	0.43

Table 6
Value-Weighted Proxy

	γ		95% C.I.	
Decile	**GMM**	**OLS**	**GMM**	
CAP1	1.27	1.23	1.19	1.36
CAP2	1.26	1.23	1.20	1.33
CAP3	1.33	1.22	1.26	1.39
CAP4	1.27	1.22	1.21	1.32
CAP5	1.21	1.23	1.16	1.26
CAP6	1.17	1.20	1.13	1.21
CAP7	1.13	1.14	1.09	1.16
CAP8	1.06	1.11	1.03	1.09
CAP9	1.11	1.06	1.09	1.13
CAP10	0.88	0.96	0.87	0.90

Notes: CAPi refers to the value-weighted portfolio of common stocks in the ith size decile. OLS estimates are obtained by regressing $Y_{i,t}$ on X_t and $Y_{m,t}$. GMM estimates use the moments from Corollary 1. Confidence Intervals are based on the GMM estimates. Source: Author's calculations.

Table 7

(A test of whether the equal-weighted or value-weighted stock index is the true market portfolio, given that the CAPM holds.)

	$\widehat{d}'\widehat{\Sigma}_e^{-1}\widehat{d}$	**Std. Error**	**Bias-Corrected 95% C.I.**	
EW	0.034	0.005	0.025	0.042
VW	0.043	0.005	0.034	0.052

Notes: EW is the equal-weighted stock index, VW the value-weighted stock index. The biased-corrected 95% confidence interval is adjusted for the fact that the estimated test statistic does not represent the median of the bootstrapped distribution. Source: Author's calculations.

Table 8

(A test of whether the correlation between innovations to the equal-weighted or value-weighted stock index and the true market portfolio exceeds ρ, given that the CAPM holds.)

		Mass > $\boldsymbol{\rho^{-2}-1}$	
$\boldsymbol{\rho}$	$\boldsymbol{\rho^{-2}-1}$	**EW**	**VW**
0.90	0.23	100.0%	100.0%
0.80	0.56	100.0%	100.0%
0.70	1.04	90.4%	100.0%
0.62	1.44	5.0%	100.0%
0.60	1.52	1.0%	100.0%
0.50	3.00	0.0%	11.7%
0.49	3.16	0.0%	5.0%
0.40	5.25	0.0%	0.0%

Notes: EW is the equal-weighted stock index, VW the value-weighted stock index. The final two columns report the mass of the bootstrapped distribution that lies above $(\rho^{-2}-1)$ for the test statistic estimated with the equal-weighted and value-weighted stock index, respectively. Source: Author's calculations.