Centre de recherche sur l'emploi et les fluctuations économiques (CREFÉ)
Center for Research on Economic Fluctuations and Employment (CREFE)

Université du Québec à Montréal

Cahier de recherche/Working Paper No. 144

Efficient Estimation of Conditional Asset Pricing Models

Douglas J. Hodgson
UQAM

Keith Vorkink
Brigham Young University

October 2001

Hodgson : Département des sciences économiques, Université du Québec à Montréal, C.P. 8888 Succursale Centre-Ville, Montréal, QC, Canada, H3C 3P8. Tel : 514-987-3000 (ext.4310). Email: hodgson.douglas-james@uqam.ca.

Vorkink: Marriott School of Management, Brigham Young University, Provo, UT, USA, 84602. Tel: 801-378-1765. Email: kpv3@email.byu.edu.

For their helpful comments, we are grateful to the referees, Jeffrey Wooldridge, Bill Brown, Oliver Linton, Rene Garcia, Werner Ploberger, Michael Brandt, and participants in seminars at USC, UCSB, Rice, Maryland, UQAM, BYU, the 2000 winter meetings of the North American Econometric Society, the 2000 EEA summer meetings, and the 2000 Canadian Econometric Study Group. We acknowledge financial support from the Institut de Finance Mathematique de Montréal and the National Science Foundation under CAREER Grant SBR-9701959.

Sommaire:
Nous développons un nouvel estimateur pour les paramètres d'un modèle de GARCH en moyenne (« GARCH-M ») avec plusieurs variables. L'estimateur a l'efficacité semiparamétrique quand les erreurs suivent une loi de probabilité qui est elliptiquement symétrique mais n'aucune autre restriction. Sous les hypothèses de haut niveau, notre estimateur obtient la limite d'efficacité semiparamétrique. L'hypothèse de la symétrie elliptique nous permet d'éviter le problème d'estimer non-paramétriquement une fonction de haut dimension, parce qu'on peut écrire la densité d'un loi elliptique comme un fonction d'une transformation unidimensionnelle de la variable aléatoire multidimensionnelle. Ce cadre est approprié pour analyser des modèles conditionnels des prix des actifs financiers, comme le CAPM conditionnel. Nous appliquons notre méthodologie à l'étude des prix des actions, et nous rendons compte des résultats d'une étude simulation «Monte-Carlo».

Abstract:
A semiparametric efficient estimation procedure is developed for the parameters of multivariate GARCH-in-mean models when the disturbances have a distribution that is assumed to be elliptically symmetric but is otherwise unrestricted. Under high level restrictions, the resulting estimator achieves the asymptotic semiparametric efficiency bound. The elliptical symmetry assumption allows us to avert the curse of dimensionality problem that would otherwise arise in estimating the unknown error distribution. This framework is suitable for the estimation and testing of conditional asset pricing models such as the conditional CAPM, and we apply our estimator in an empirical study of stock prices, with Monte Carlo simulation results also being reported.

Keywords: Capital asset pricing model, elliptical symmetry, semiparametric efficiency, GARCH.

JEL classification: C22

1. INTRODUCTION

Modelling expected returns has permeated much of financial research in the past three decades. The payoffs from a correct relationship between risk and expected return are abundant and include applications to capital budgeting, portfolio performance, event studies and others. The mean-variance model of the risk-return relationship was initially implemented empirically for multivariate data by Bollerslev, Engle, and Wooldridge (1988), who develop a conditional CAPM (C-CAPM) model and associated GARCH-in-mean (GARCH-M) econometric model. A large empirical literature has subsequently developed in this area, generally estimating the models with Gaussian quasi-maximum likelihood estimation (Q-MLE) techniques. Although such techniques typically retain their consistency and asymptotic normality properties in the presence of non-normal data (Bollerslev and Wooldridge (1992)), asymptotic inefficiency and imprecise parameter estimates can occur due to the presence of thick tails in the distributions underlying financial data. We propose a new estimation methodology for the multivariate GARCH-in-mean (GARCH-M) model that is designed to account for excess tail thickness by adopting a flexible distributional assumption of conditional elliptical symmetry. The estimator will achieve the asymptotic semiparametric efficiency bound in the presence of general elliptical symmetry in the data generating process. We apply our estimator to a data set of stock returns and perform asset pricing tests of the conditional capital asset pricing model (C-CAPM).

It has been well documented that stock returns are not independent and identically distributed (*iid)* normal, in particular they tend to exhibit substantial kurtosis and have moments that vary over time (see, for example, Mandelbrot (1963), Fama (1965), Engle (1982), and Bollerslev, Chou, and Kroner (1992)). These phenomena are not unrelated. It is well known that time-varying volatility implies a thick-tailed unconditional distribution. However, as shown in Bollerslev (1987), conditional volatility cannot completely account for the tail behavior of the unconditional distribution in financial returns (see also Diebold (1988), Nelson (1991) and Vorkink (2001)). Accurate description of return distributions should include modelling of both of these properties.

We propose a semiparametric efficient estimator that attempts to improve upon the inefficiencies that may occur in the Gaussian Q-MLE when thick tails are present in the distribution of the standardized innovations to the GARCH-M model. To do so, we assume the distribution of returns to be a member of the class of elliptically symmetric distributions. This class includes those having conditional dependence among higher moments, infinite variance (Cauchy), Student-t and others. For further discussion of elliptical distributions, see Fang, Kotz and Ng (1990), Muirhead (1982), and the Appendix of the present paper. We derive the asymptotic semiparametric efficiency bound for the estimation of our model's parameters in the presence of an unknown elliptically symmetric innovation density, then propose a semiparametric estimator that achieves the bound. This estimator will employ a nonparametric kernel estimator of the unknown innovation density.

This assumption of elliptical symmetry plays an integral role in our estimation methodology and particularly in the estimation of the residual density. We can think of two extreme methods of obtaining an estimate of a leptokurtokic residual density. One would be fitting a fully parametric non-normal distribution to the residuals. Alternatively, the density could be estimated in a fully nonparametric fashion. For example, Drost and Klaassen (1997) propose a semiparametric efficient estimation method for univariate GARCH models that involves nonparametric kernel estimation of the innovation density. However, their method is difficult to extend to a multivariate setting, due to the "curse of dimensionality" problem that the convergence rate of a nonparametric density estimate diminishes rapidly as the dimension of the density increases.

Elliptical symmetry provides a middle ground between a fully specified Q-MLE approach and a fully nonparametric approach. While the density is nonparametrically estimated within the elliptically symmetric class, this restriction allows us to do so without falling prey to the "curse of dimensionality". Specifically, we are able to transform the nonparametric density estimator to one which is always one-dimensional.

This estimator's roots lie in Bickel's (1982) adaptive estimator. Assuming *iid* data, Bickel considered the problem of adaptively estimating mean and covariance parameters in elliptically symmetric location models. He found that under the assumption of elliptical symmetry, the mean could be adaptively estimated and that the covariance parameters could be adaptively estimated up to a scale. Linton (1993) showed that slope parameters can be adaptively estimated in a regression model with ARCH errors when the innovation density is symmetric. In both cases, the innovation density is otherwise unrestricted and is estimated using nonparametric kernel methods. Hodgson, Linton, and Vorkink (henceforth HLV (2001)) have derived adaptive estimators in time series models under the assumption of elliptical symmetry using a nonparametric estimate of the joint innovation density.

HLV (2001) develop an estimator of linear unconditional asset pricing models under elliptical symmetry. Their estimator is fully asymptotically efficient and places no assumptions on the family of return distributions other than that this family is elliptically symmetric. They find that the more general estimator leads to substantially different estimates and conclusions when testing unconditional asset pricing models. However, the treatment of conditional heteroskedasticity is ad hoc which results in potential inefficiencies. The present paper extends this work by parameterizing the conditional heteroskedasticity in the form of a multivariate GARCH-M model with conditionally elliptically symmetric innovation distributions.

Asset pricing theory exists which is consistent with the specification of elliptical symmetry, at least for the case of the one-period unconditional CAPM, although the conditions under which these results would extend to a multiperiod conditional model are not known. Merton (1973) mentions the restrictions that would be required in a multiperiod model to generate one-period ahead mean-variance pricing. It may be possible to show that Merton's (1973) conditions, along with conditional elliptical symmetry, yield such pricing, but we know of no formal results to this effect. In the one-period CAPM, the assumption of normally distributed returns is sufficient for a mean-variance result but not necessary. Chamberlain (1983), Owen and Rabinovitch (1983), and Berk (1997) have obtained mean-variance pricing under the assumption that returns are elliptically symmetric. In fact, Berk (1997) found that elliptical symmetry is the most general distributional assumption that is consistent with mean-variance maximization when consumers are assumed to have concave utility functions. These exact pricing models are more general and consistent with the empirical regularities than their normal distribution counterparts. However, while these theoretical results can be obtained with more general distributional assumptions, estimation of the general model has not been feasible until recently.

Our estimator is specifically designed to be more robust than the Gaussian Q-MLE in the presence of thick tails in the standardized innovations to the GARCH-M model. As shown by Bollerslev and Wooldridge (1992), the Q-MLE has the virtue of being consistent and asymptotically normal for a substantial range of deviations of the innovations from iid normailty. We have not been able to derive similar properties for the semiparametric estimator developed in the present paper, and only know its properties when the assumption of iid elliptical symmetry on the innovations holds. For data where such deviations from this assumption as conditional or unconditional skewness may be present, we can currently only conjecture as to the behaviour of our estimator. Furthermore, empirical and simulation evidence reported below suggests that the efficiency gains of our estimator vis-à-vis the Q-MLE are quite modest for estimation of conditional mean parameters, although the evidence suggests that there may be potential gains in estimating conditional covariance parameters and conditional betas.

The paper is organized as follows. Section 2 introduces the conditional CAPM model that we will be concerned with estimating and testing. In Section 3, we present our multivariate GARCH-M econometric model. Section 4 contains our derivation of the semiparametric efficiency bound for our model and describes a method of feasibly computing an estimator that will achieve the bound. Sections 5 and 6 report empirical and simulation results, respectively, with Section 7 concluding. The Appendix contains a discussion of elliptically symmetric densities and discusses some computational details relating to our estimator.

2. CONDITIONAL RETURN MODELS

It has been shown that the assumption of constant return distributions is not necessary to obtain equilibrium pricing equations. Merton (1973) derived an intertemporal-CAPM which showed how investors would react to changing investment opportunity sets. In an empirical setting Bollerslev, Engle, and Wooldridge (1988) estimated conditional-CAPM covariances assuming that the covariance matrix of returns followed a GARCH-M(1,1) process. They found that, under this model's parameterization, beta and the market price of risk are time-varying. They also show that both returns and volatility are predictable and time-varying. In fact, they are able to predict a larger portion of the variability in returns than the unconditional counterpart (see also Harvey (1991), Buse, Korkie, and Turtle (1994), Braun, Nelson, and Sunier (1995), and Jagannathan and Wang (1996)). We suggest that a natural extension would be to estimate a conditional asset pricing model where the residual distribution is assumed to be thick-tailed relative to the normal distribution and allow some flexibility in the form of the conditional distribution.

We now introduce the conditional-CAPM (C-CAPM) return relationship. Our discussion will closely follow that of Bollerslev, Engle, and Wooldridge (1988), with some variations as suggested by, for example, DeSantis and Gerard (1998). The following equation demonstrates the main relationship of the conditional-CAPM, stating that the excess return on asset i is linear in its covariance with the market portfolio:

$$E_{t-1}\left[R_{i,t}\right] - r_{ft} = \lambda cov_{t-1}\left(R_{m,t}, R_{i,t}\right). \tag{1}$$

We assume that there are n assets in the market, that $R_{i,t}$ is the return on asset i in period t, $R_{m,t}$ is the return on the market portfolio, r_{ft} is the return to a risk-free asset, and that the subscripts on expectations and covariances indicate conditional moments. We note that it would be a straightforward matter to extend the model to allow for multiple factors to influence returns. From (1) we can see that the expected return on the market portfolio is

$$E_{t-1}\left[R_{m,t}\right] - r_{ft} = \lambda var_{t-1}\left(R_{m,t}\right),$$

so that the parameter λ can be interpreted as the market price on risk. Following DeSantis and Gerard (1998), we may treat this parameter either as a constant or as time-varying, in which latter case it can be modelled as being dependent on an $\ell-$vector of state variables $\mathbf{v}_t$, and (1) can be generalized by writing $\lambda = \exp\left(\gamma_0^* + \mathbf{v}_t^T\boldsymbol{\gamma}_1\right)$. In the model with a constant price of risk, we have $\lambda = \exp\left(\gamma_0^*\right)$. We can also write our expected return relationship as

$$E_{t-1}\left[R_{i,t}\right] - r_{ft} = E_{t-1}\left[R_{m,t} - r_{ft}\right]\beta_{i,t},$$

where $\beta_{i,t} = \frac{cov_{t-1}(R_{m,t}, R_{i,t})}{var_{t-1}(R_{m,t})}$ is the conditional "beta" for asset i in period t.

Define the n-vector $\mathbf{r}_t = \mathbf{R}_t - r_{ft}\boldsymbol{\iota}_n$, where $\mathbf{R}_t$ is the vector of returns on the individual assets and $\boldsymbol{\iota}_n$ is an n-vector of ones. Following Bollerslev, Engle, and Wooldridge (1988), let $\boldsymbol{\omega}_{t-1}$ be the n-vector of weights assigned to the assets in computing the "market", so that $R_{m,t} = \boldsymbol{\omega}_{t-1}^T\mathbf{R}_t$. Allowing for a possibly time-varying market price of risk, we may then write our CAPM relationship at time t for our cross-section of assets as

$$E_{t-1}\mathbf{r}_t = \exp\left(\gamma_0^* + \mathbf{v}_t^T\boldsymbol{\gamma}_1\right)\boldsymbol{\Sigma}_t\boldsymbol{\omega}_{t-1}, \tag{2}$$

where $\boldsymbol{\Sigma}_t$ is the covariance matrix of asset returns conditional on information available up to period $t-1$. Note that our vector of conditional betas is given by

$$\boldsymbol{\beta}_t = \frac{\boldsymbol{\Sigma}_t\boldsymbol{\omega}_{t-1}}{\boldsymbol{\omega}_{t-1}^T\boldsymbol{\Sigma}_t\boldsymbol{\omega}_{t-1}}. \tag{3}$$

Estimation of our model will depend on the specification of a model for our conditional covariance matrix.

Testing the C-CAPM typically involves estimating the following model:

$$\mathbf{r}_t = \boldsymbol{\alpha} + \exp\left(\gamma_0^* + \mathbf{v}_t^T\boldsymbol{\gamma}_1\right)\boldsymbol{\Sigma}_t\boldsymbol{\omega}_{t-1} + \mathbf{u}_t \tag{4}$$

where an intercept is included to capture persistent variation in $\mathbf{r}_t$ that is not captured by variation in the market return. One common test of the asset pricing model takes the following form:

$$H_0 : \alpha_i = 0 \quad i = 1, ..., n \tag{5}$$

which implies that no significant excess returns are present in each portfolio's return that cannot be explained by variation in the market portfolio return. This hypothesis can be tested by construction of a standard Wald test

$$J = \tilde{\boldsymbol{\alpha}}'\widehat{\mathbf{Var}}(\tilde{\boldsymbol{\alpha}})\tilde{\boldsymbol{\alpha}}, \tag{6}$$

where $\tilde{\boldsymbol{\alpha}}$ is an estimator and $\widehat{\mathbf{Var}}(\tilde{\boldsymbol{\alpha}})$ estimates its asymptotic covariance matrix. If this statistic deviates significantly from zero, we conclude that the C-CAPM does not fully explain the deviations in returns.

It is also interesting to look at the time series of the implied betas $\beta_{i,t}$ to see if the conditional variance parameterization leads to substantial time variation in the covariation between the asset's return and the market return. For example, Bollerslev, Engle, and Wooldridge (1988) found substantial variation in the implied betas of their estimation of the US stock and bond market. Correct modeling of the time variation in betas and market risk will lead to improved portfolio weights, performance measures, and estimated expected returns.

3. THE ECONOMETRIC MODEL

The regression model we estimate is given in equation (4). In order to arrive at a completely specified econometric model we must specify a model for our conditional covariance matrix $\mathbf{\Sigma}_t$ and our disturbance process $\{\mathbf{u}_t\}$. There is no theory predicting a GARCH model of volatility; however, it is a relatively parsimonious model of time-varying second moments that has been quite successful in capturing the time series behavior of volatility. Our general model of conditional volatility will be the following simplified version of the multivariate GARCH model described in Engle and Kroner (1995):

$$\mathbf{\Sigma}_t = \exp(\kappa)\mathbf{H}_t \tag{7}$$

where

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}^T + \mathbf{A}\mathbf{u}_{t-1}\mathbf{u}_{t-1}^T\mathbf{A} + \mathbf{D}\mathbf{H}_{t-1}\mathbf{D}, \tag{8}$$

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ c_{n1} & \cdots & \cdots & c_{nn} \end{bmatrix}$$

and the matrices $\mathbf{A}$ and $\mathbf{D}$ are diagonal. This model is less general than that developed by Engle and Kroner (1995), in which

$$\mathbf{H}_t = \mathbf{C}^T\mathbf{C} + \sum_{j=1}^{n}\mathbf{A}_j\mathbf{u}_{t-1}\mathbf{u}_{t-1}^T\mathbf{A}_j + \sum_{j=1}^{n}\mathbf{D}_j\mathbf{H}_{t-1}\mathbf{D}_j.$$

We adopt this simplification for computational purposes. Our model still has the generality to allow for systematically time-varying conditional variances and covariances. Other empirical papers such as Bollerslev, Engle, and Wooldridge (1988) and DeSantis and Gerard (1998) employ simplified GARCH-M models. This specification is more general than those of Bollerslev (1987) and Jeantheau (1998) in that it allows for time-varying conditional covariances. We should note that under our assumptions on $\mathbf{A}$ and $\mathbf{D}$, our restriction of the leading term of $\mathbf{C}$ to be unity does not entail any further loss of generality. To see this, note that the conditional variance of the first element of $\mathbf{u}_t$ is

$$var_{t-1}(u_{1t}) = \sigma_{11,t} = \exp(\kappa)\left(1 + a_1^2u_{1,t-1}^2 + d_1^2h_{11,t-1}\right),$$

which is the usual expression for conditional variance in a univariate GARCH model.

To complete our specification of the model, we assume that our regression disturbances $\{\mathbf{u}_t\}$ have the following elliptically symmetric conditional density:

$$p_{t-1}(\mathbf{u}_t) = |\mathbf{\Sigma}_t|^{-\frac{1}{2}} \, \widetilde{g}\left(\mathbf{u}_t^T \mathbf{\Sigma}_t^{-1} \mathbf{u}_t\right). \tag{9}$$

Our objective in this paper is to obtain a semiparametric efficient estimator of the parameters of our model treating the functional form of $\widetilde{g}$ as an unknown infinite dimensional nuisance parameter. The function $\widetilde{g}(\cdot)$ has only a scalar as its argument which plays an important role in the nonparametric estimation of the density. We also define $_{p,t}$ to be the conditional information matrix of $p_{t-1}(\mathbf{u}_t)$; it is proportional to the inverse of $\mathbf{H}_t$ and $\mathbf{\Sigma}_t$. We have $_{p,t} = const \cdot \mathbf{\Sigma}_t^{-1}$, with the constant being greater than or equal to one (it equals one if $p_{t-1}(\mathbf{u}_t)$ is Gaussian). Mitchell (1989) computes the value of the constant for various elliptically symmetric densities.

Note that because we are treating $\widetilde{g}$ as being of unknown functional form, we can also write the density as

$$p_{t-1}(\mathbf{u}_t) = |\mathbf{H}_t|^{-\frac{1}{2}} \, g\left(\mathbf{u}_t^T \mathbf{H}_t^{-1} \mathbf{u}_t\right), \tag{10}$$

where the constant of proportionality relating $\mathbf{H}_t$ and $\mathbf{\Sigma}_t$ has now been absorbed into the function g. This specification follows the example of Linton (1993), who does not consider efficient estimation of κ. Note that $\widetilde{g}(\cdot)$ as defined in (9) is the density function of the *iid* spherically symmetrically distributed random variable $\mathbf{\Sigma}_t^{-1/2}\mathbf{u}_t$ with unit covariance matrix. As defined in (10), $g(\cdot)$ is still the density of an *iid* spherically distributed random variable, but without the restriction of a unit covariance matrix. We shall also not concern ourselves with efficient estimation of κ, and rewrite our regression model as follows:

$$\mathbf{r}_t = \boldsymbol{\alpha} + \exp\left(\gamma_0 + \mathbf{v}_t^T \boldsymbol{\gamma}_1\right) \mathbf{H}_t \boldsymbol{\omega}_{t-1} + \mathbf{u}_t, \tag{1}$$

where $\gamma_0 = \kappa + \gamma_0^*$. We shall not consider semiparameteric efficient estimation of the parameters κ and γ_0^* separately (although in principle it would be possible to do so), but only of their sum γ_0. We justify this parameterization in our case because our parameters of primary interest are the intercept parameter $\boldsymbol{\alpha}$ and the conditional beta vector $\boldsymbol{\beta}_t$. Note that the latter depends only on the parameters of the function $\mathbf{H}_t$ as defined in (7), the reason for this being that

$$\begin{aligned}
\boldsymbol{\beta}_t &= \frac{\mathbf{\Sigma}_t \boldsymbol{\omega}_{t-1}}{\boldsymbol{\omega}_{t-1}^T \mathbf{\Sigma}_t \boldsymbol{\omega}_{t-1}} \\
&= \frac{\exp(\kappa) \mathbf{H}_t \boldsymbol{\omega}_{t-1}}{\boldsymbol{\omega}_{t-1}^T \exp(\kappa) \mathbf{H}_t \boldsymbol{\omega}_{t-1}} \\
&= \frac{\mathbf{H}_t \boldsymbol{\omega}_{t-1}}{\boldsymbol{\omega}_{t-1}^T \mathbf{H}_t \boldsymbol{\omega}_{t-1}}.
\end{aligned}$$

Let $\left(1, \mathbf{c}^T\right)^T = vech(\mathbf{C})$, so that $\mathbf{c}$ is the $\frac{n(n+1)-2}{2}-$vector of unknown elements of $\mathbf{C}$, $\mathbf{a} = diag(\mathbf{A})$, $\mathbf{d} = diag(\mathbf{D})$ and $\boldsymbol{\theta}_2 = \left\{\mathbf{c}^T, \mathbf{a}^T, \mathbf{d}^T\right\}^T$ is the vector of unknown parameters in the conditional covariance function. Note that there are $h_2 = \frac{n(n+5)-2}{2}$ parameters in $\boldsymbol{\theta}_2$. Likewise let the vector of parameters in the conditional mean function be given by

$\boldsymbol{\theta}_1 = \left\{\boldsymbol{\alpha}^T, \boldsymbol{\gamma}^T\right\}^T$, where $\boldsymbol{\gamma} = \left(\gamma_0, \boldsymbol{\gamma}_1^T\right)^T$, so that $\boldsymbol{\theta}_1$ is of dimension $h_1 = n + \ell + 1$. Our $h(= h_1 + h_2)$-dimensional full parameter vector is $\boldsymbol{\theta} \equiv \left(\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T\right)^T$, which has usually been estimated using Q-MLE procedures resulting from a specification of *iid* normality for the normalized disturbance process $\boldsymbol{\varepsilon}_t = \left\{\mathbf{H}_t^{-1/2}\mathbf{u}_t\right\}$. Although few analytical results are available, Bollerslev and Wooldridge (1992) have shown, under high level assumptions, that the Gaussian Q-MLE will be $\sqrt{T}-$consistent and asymptotically normal, even under distributional misspecification. We derive estimators that are asymptotically semiparametrically efficient under our elliptical symmetry assumption (along with high level assumptions similar to those of Bollerslev and Wooldridge (1992)), but without placing additional restrictions on the return distribution. We use a semiparametric Newton-Raphson type estimator following the basic approach of Bickel (1982).

4. EFFICIENT ESTIMATION

Our derivation of a semiparametric efficiency bound for the model described above is given in this section. Following the literature in the area of multivariate GARCH models, we will derive our estimation theory under a set of high-level assumptions. The restrictions which such assumptions imply for the parameters of our model are not known and could presumably be obtained only with great difficulty. This is a endemic problem in multivariate GARCH modelling. Jeantheau (1998) provides a recent example of a consistency result for a multivariate GARCH model which doesn't rely on such high-level assumptions, but at the cost of using a very restrictive parameterization. We assume that our data are stationary and ergodic, that conditional variances are always finite and bounded away from zero, and that the score function has finite variance. Any expectation or derivative taken in the following sections is assumed to exist, and conditions for the consistency and asymptotic normality of the estimators used are assumed to hold. We can apply a result of Brown and Hodgson (2001) to obtain a semiparametric efficiency bound for our model, for which purpose we must make the further assumptions that $g(w)$ is three times differentiable with bounded third derivative, where $w = \varepsilon^T\varepsilon$, that $\ln\left|\mathbf{H}_t(\boldsymbol{\theta})^{-1/2}\right|$ is three times differentiable with respect to $\boldsymbol{\theta}$ with bounded third derivative, and that $\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$ is three times differentiable with respect to $\boldsymbol{\theta}$.

We now turn to the issue of semiparametric efficient estimation of the parameter vector $\boldsymbol{\theta}$. For a fuller discussion of semiparametric efficiency bounds and the related concepts used here, see Newey (1990). We must derive an expression for the *efficient score* for $\boldsymbol{\theta}$, which is the orthogonal complement of the projection of the score for $\boldsymbol{\theta}$ onto the *tangent space*, which is, loosely speaking, the space spanned by all scores for parameterizations $\boldsymbol{\tau}$ of the unknown density $g(\cdot)$ that include the true model of $g(\cdot)$ as special cases. Such a parameterization, which we write as $g\left(\mathbf{u}_t^T(\boldsymbol{\theta})\,\mathbf{H}_t^{-1}(\boldsymbol{\theta})\,\mathbf{u}_t(\boldsymbol{\theta}), \boldsymbol{\tau}\right)$, is known as a *parametric submodel*. For a fuller discussion of semiparametric efficiency bounds and the related concepts used here, see Newey (1990). A semiparametric efficiency bound for our model can be obtained by applying Theorem 1 of Brown and Hodgson (2000), which applies to a class of nonlinear models with elliptical distributions that contains our model. An heuristic derivation of the bound is given below.

The log-likelihood for the aforementioned parametric submodel for a sample of size T, where we follow the usual practice of conditioning on initial conditions whose unconditional density is assumed to have an asymptotically negligible effect on analysis of the likelihood, is

$$\begin{aligned}\ln \mathcal{L}_T(\boldsymbol{\theta},\boldsymbol{\tau}) &= -\frac{1}{2}\sum_{t=1}^{T}\ln|\mathbf{H}_t(\boldsymbol{\theta})| + \sum_{t=1}^{T} g\left(\mathbf{u}_t^T(\boldsymbol{\theta})\mathbf{H}_t^{-1}(\boldsymbol{\theta})\mathbf{u}_t(\boldsymbol{\theta}),\boldsymbol{\tau}\right)\\ &= -\frac{1}{2}\sum_{t=1}^{T}\ln|\mathbf{H}_t(\boldsymbol{\theta})| + \sum_{t=1}^{T} g\left(\boldsymbol{\varepsilon}_t^T(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta}),\boldsymbol{\tau}\right)\\ &= -\frac{1}{2}\sum_{t=1}^{T}\ln|\mathbf{H}_t(\boldsymbol{\theta})| + \sum_{t=1}^{T} g\left(w_t(\boldsymbol{\theta}),\boldsymbol{\tau}\right),\end{aligned}$$

where $w_t(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_t^T(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t(\boldsymbol{\theta})$. The score of the t^{th} observation with respect to the nuisance parameter $\boldsymbol{\tau}$ will be

$$\ell_{t\tau}(\boldsymbol{\theta},\boldsymbol{\tau}) = \frac{g_2(w_t(\boldsymbol{\theta}),\boldsymbol{\tau})}{g(w_t(\boldsymbol{\theta}),\boldsymbol{\tau})},$$

where $g_j(\cdot,\cdot)$ denotes the partial derivative of g with respect to its j^{th} argument, for $j = 1, 2$. Note that because $\{w_t(\boldsymbol{\theta})\}$ is assumed to be an *iid* sequence, so is $\{\ell_{t\tau}(\boldsymbol{\theta},\boldsymbol{\tau})\}$. Similarly, because $w_t(\boldsymbol{\theta})$ is independent of

$$(\mathbf{r}_{t-1},\mathbf{r}_{t-2},...;\mathbf{H}_t(\boldsymbol{\theta}),\mathbf{H}_{t-1}(\boldsymbol{\theta}),...;\mathbf{v}_t,\mathbf{v}_{t-1},...),$$

so is $\ell_{t\tau}(\boldsymbol{\theta},\boldsymbol{\tau})$. Furthermore, we have

$$E\left[\ell_{t\tau}(\boldsymbol{\theta},\boldsymbol{\tau})\right] = E\left[\ell_{t\tau}(\boldsymbol{\theta},\boldsymbol{\tau})\,|\Psi_{t-1}\right] = 0,$$

where we define the $\sigma-$field

$$\Psi_{t-1} = \sigma(\mathbf{r}_{t-1},\mathbf{r}_{t-2},...;\mathbf{H}_t(\boldsymbol{\theta}),\mathbf{H}_{t-1}(\boldsymbol{\theta}),...;\mathbf{v}_t,\mathbf{v}_{t-1},...).$$

The tangent space $\mathcal{T}$ is the infinite-dimensional Hilbert space spanned by all functions having the defining characteristics of $\ell_{t\tau}(\boldsymbol{\theta},\boldsymbol{\tau})$, namely that it is a function only of $\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$ and that it has zero mean:

$$\mathcal{T} = \left\{s\left(\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\right) : E\left[s\left(\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\right)\right] = 0\right\}.$$

The projection of an arbitrary function

$$R(\mathbf{r}_t,\mathbf{r}_{t-1},\mathbf{r}_{t-2},...;\mathbf{H}_t(\boldsymbol{\theta}),\mathbf{H}_{t-1}(\boldsymbol{\theta}),...;\mathbf{v}_t,\mathbf{v}_{t-1},...;\boldsymbol{\varepsilon}_t,\boldsymbol{\varepsilon}_{t-1},...) = R(\mathbf{y}_t)$$

on the tangent space can be shown to be

$$\Pr\left[R(\mathbf{y}_t)\,|\mathcal{T}\right] = E\left[R(\mathbf{y}_t)\,\middle|\,\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\right].$$

In calculating the efficient score for $\boldsymbol{\theta}$, we first consider the score for $\boldsymbol{\theta}$, which for observation t can be written as

$$\ell_{t\theta}(\boldsymbol{\theta},\boldsymbol{\tau}) = -\frac{1}{2}\frac{\partial\ln|\mathbf{H}_t|}{\partial\boldsymbol{\theta}} + 2\frac{\partial\mathbf{u}_t^T}{\partial\boldsymbol{\theta}}\mathbf{H}_t^{-1}\mathbf{u}_t\frac{g_1(\mathbf{w}_t,\boldsymbol{\tau})}{g(\mathbf{w}_t,\boldsymbol{\tau})}$$

$$+\frac{\partial\left(vec\mathbf{H}_t^{-1}\right)^T}{\partial\boldsymbol{\theta}}vec\left(\mathbf{u}_t\mathbf{u}_t^T\right)\frac{g_1\left(\mathbf{w}_t,\boldsymbol{\tau}\right)}{g\left(\mathbf{w}_t,\boldsymbol{\tau}\right)}, \tag{12}$$

where we have suppressed dependence of $\boldsymbol{\Sigma}_t$, w_t, and $\mathbf{u}_t$ on $\boldsymbol{\theta}$ to prevent cluttered notation. In considering the projection of $\ell_{t\theta}\left(\boldsymbol{\theta},\boldsymbol{\tau}\right)$ onto the tangent space, first note that the first two components of $\ell_{t\theta}\left(\boldsymbol{\theta},\boldsymbol{\tau}\right)$ are orthogonal to the nuisance scores $\ell_{t\tau}\left(\boldsymbol{\theta},\boldsymbol{\tau}\right)$ for any parametric submodel and hence are orthogonal to the tangent space. Considering the first component on the RHS of (12), we have

$$E\left[\frac{\partial\ln\left|\mathbf{H}_t\right|}{\partial\boldsymbol{\theta}}\ell_{t\tau}\left(\boldsymbol{\theta},\boldsymbol{\tau}\right)\right]=E\left[\frac{\partial\ln\left|\mathbf{H}_t\right|}{\partial\boldsymbol{\theta}}\right]E\left[\ell_{t\tau}\left(\boldsymbol{\theta},\boldsymbol{\tau}\right)\right]=0,$$

since $E\left[\ell_{t\tau}\left(\boldsymbol{\theta},\boldsymbol{\tau}\right)\right]=0$. Considering now the second component, note that $\frac{\partial\mathbf{u}_t^T}{\partial\boldsymbol{\theta}}$ and $\mathbf{H}_t$ are both measurable with respect to Ψ_{t-1}, yielding

$$E\left[\frac{\partial\mathbf{u}_t^T}{\partial\boldsymbol{\theta}}\mathbf{H}_t^{-1}\mathbf{u}_t\frac{g_1\left(\mathbf{w}_t,\boldsymbol{\tau}\right)}{g\left(\mathbf{w}_t,\boldsymbol{\tau}\right)}\ell_{t\tau}\left(\boldsymbol{\theta},\boldsymbol{\tau}\right)\right]=E\left[\frac{\partial\mathbf{u}_t^T}{\partial\boldsymbol{\theta}}\mathbf{H}_t^{-1/2}\right]E\left[\varepsilon_t\frac{g_1\left(\mathbf{w}_t,\boldsymbol{\tau}\right)}{g\left(\mathbf{w}_t,\boldsymbol{\tau}\right)}\ell_{t\tau}\left(\boldsymbol{\theta},\boldsymbol{\tau}\right)\right]=0,$$

where $E\left[\varepsilon_t\frac{g_1(\mathbf{w}_t,\boldsymbol{\tau})}{g(\mathbf{w}_t,\boldsymbol{\tau})}\ell_{t\tau}\left(\boldsymbol{\theta},\boldsymbol{\tau}\right)\right]=0$ by symmetry. It remains to consider the projection of the third component of the RHS of (12) onto the tangent space, which is given by

$$\begin{aligned}&E\left[\frac{\partial\left(vec\mathbf{H}_t^{-1}\right)^T}{\partial\boldsymbol{\theta}}\left(\mathbf{H}_t^{1/2}\quad\mathbf{H}_t^{1/2}\right)vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T\right)\frac{g'\left(\mathbf{w}_t\right)}{g\left(\mathbf{w}_t\right)}\left|\mathbf{w}_t\right.\right]\\ =\ &E\left[\frac{\partial\left(vec\mathbf{H}_t^{-1}\right)^T}{\partial\boldsymbol{\theta}}\left(\mathbf{H}_t^{1/2}\quad\mathbf{H}_t^{1/2}\right)\right]E\left[vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T\right)\left|\mathbf{w}_t\right.\right]\frac{g'\left(\mathbf{w}_t\right)}{g\left(\mathbf{w}_t\right)},\end{aligned}$$

the equality holding because $\mathbf{H}_t$ is independent of $\boldsymbol{\varepsilon}_t$. Note that $E\left[vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T\right)\left|\mathbf{w}_t\right.\right]\neq E\left[vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T\right)\right]$. Here and in what follows, we drop the nuisance parameter $\boldsymbol{\tau}$ from our notation, since the notion of a parametric submodel has served its purpose and we now concern ourselves with the semiparametric model. The derivative of $g\left(\cdot\right)$ is now denoted by $g'\left(\cdot\right)$.

The projection of the period t score for $\boldsymbol{\theta}$ onto the tangent space is therefore

$$\Pr\left[\ell_{t\theta}\left(\boldsymbol{\theta}\right)\left|\mathcal{T}\right.\right]=E\left[\frac{\partial\left(vec\mathbf{H}_t^{-1}\right)^T}{\partial\boldsymbol{\theta}}\left(\mathbf{H}_t^{1/2}\quad\mathbf{H}_t^{1/2}\right)\right]E\left[vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T\right)\left|\mathbf{w}_t\right.\right]\frac{g'\left(\mathbf{w}_t\right)}{g\left(\mathbf{w}_t\right)}$$

and the period t efficient score for $\boldsymbol{\theta}$ is

$$\begin{aligned}\boldsymbol{\Delta}_{t,T}\left(\boldsymbol{\theta}\right)&=\ell_{t\theta}\left(\boldsymbol{\theta}\right)-\Pr\left[\ell_{t\theta}\left(\boldsymbol{\theta}\right)\left|\mathcal{T}\right.\right]\\&=-\frac{1}{2}\frac{\partial\ln\left|\mathbf{H}_t\right|}{\partial\boldsymbol{\theta}}+\left\{2\frac{\partial\mathbf{u}_t^T}{\partial\boldsymbol{\theta}}\mathbf{H}_t^{-1}\mathbf{u}_t\right.\\&\quad+\frac{\partial\left(vec\mathbf{H}_t^{-1}\right)^T}{\partial\boldsymbol{\theta}}\left(\mathbf{H}_t^{1/2}\quad\mathbf{H}_t^{1/2}\right)vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T\right)\end{aligned} \tag{13}$$

$$-E\left[\frac{\partial\left(vec\mathbf{H}_t^{-1}\right)^T}{\partial\boldsymbol{\theta}}\left(\mathbf{H}_t^{1/2}\quad\mathbf{H}_t^{1/2}\right)\right]E\left[vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T\right)|\mathbf{w}_t\right]\Bigg\}\frac{g'\left(\mathbf{w}_t\right)}{g\left(\mathbf{w}_t\right)}$$

$$=-\frac{1}{2}\frac{\partial\ln\left|\mathbf{H}_t\left(\boldsymbol{\theta}\right)\right|}{\partial\boldsymbol{\theta}}+\boldsymbol{\Gamma}_t\left(\boldsymbol{\theta}\right)\frac{g'\left(\mathbf{w}_t\left(\boldsymbol{\theta}\right)\right)}{g\left(\mathbf{w}_t\left(\boldsymbol{\theta}\right)\right)},$$

where

$$\begin{aligned}\boldsymbol{\Gamma}_t\left(\boldsymbol{\theta}\right) &= 2\frac{\partial\mathbf{u}_t^T\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\mathbf{H}_t^{-1}\left(\boldsymbol{\theta}\right)\mathbf{u}_t\left(\boldsymbol{\theta}\right)\\ &+\frac{\partial\left(vec\mathbf{H}_t^{-1}\left(\boldsymbol{\theta}\right)\right)^T}{\partial\boldsymbol{\theta}}\left(\mathbf{H}_t^{1/2}\left(\boldsymbol{\theta}\right)\quad\mathbf{H}_t^{1/2}\left(\boldsymbol{\theta}\right)\right)vec\left(\boldsymbol{\varepsilon}_t\left(\boldsymbol{\theta}\right)\boldsymbol{\varepsilon}_t^T\left(\boldsymbol{\theta}\right)\right)\\ &-E\left[\frac{\partial\left(vec\mathbf{H}_t^{-1}\left(\boldsymbol{\theta}\right)\right)^T}{\partial\boldsymbol{\theta}}\left(\mathbf{H}_t^{1/2}\left(\boldsymbol{\theta}\right)\quad\mathbf{H}_t^{1/2}\left(\boldsymbol{\theta}\right)\right)\right]E\left[vec\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\right)|\mathbf{w}_t\left(\boldsymbol{\theta}\right)\right].\end{aligned}$$

Our efficient score function for the sample of size T will then be

$$\boldsymbol{\Delta}_T\left(\boldsymbol{\theta}\right)=\sum_{t=1}^{T}\boldsymbol{\Delta}_{t,T}\left(\boldsymbol{\theta}\right), \tag{14}$$

with the semiparametric efficient estimator $\widetilde{\boldsymbol{\theta}}_T^{**}$ being that value $\theta\in\Theta$ that sets the efficient score equal to zero, i.e. such that

$$\boldsymbol{\Delta}_T\left(\widetilde{\boldsymbol{\theta}}_T^{**}\right)=\sum_{t=1}^{T}\boldsymbol{\Delta}_{t,T}\left(\widetilde{\boldsymbol{\theta}}_T^{**}\right)=0.$$

Under the high-level assumptions outlined at the start of this section, the semiparametric efficient estimator will have the following asymptotic distribution:

$$\sqrt{n}\left(\widetilde{\boldsymbol{\theta}}_T^{**}-\boldsymbol{\theta}_0\right)\xrightarrow{d}N\left(\mathbf{0},\mathcal{B}\right), \tag{15}$$

where the semiparametric efficiency bound $\mathcal{B}$ is given by

$$\mathcal{B}=\left\{E\left[\boldsymbol{\Delta}_{t,T}\left(\boldsymbol{\theta}_0\right)\boldsymbol{\Delta}_{t,T}^T\left(\boldsymbol{\theta}_0\right)\right]\right\}^{-1}.$$

Note that under our assumptions, an information matrix equality will hold here, so that

$$E\left[\boldsymbol{\Delta}_{t,T}\left(\boldsymbol{\theta}_0\right)\boldsymbol{\Delta}_{t,T}^T\left(\boldsymbol{\theta}_0\right)\right]=-E\left[\frac{\partial\boldsymbol{\Delta}_{t,T}\left(\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}^T}\right].$$

Note that under any misspecification (such as, for example, the failure of either our *iid* assumption or our elliptical symmetry assumption on the errors) this equality will fail to hold, so the possibility exists of a White (1982)-style specification test, although we do not explore this possibility here.

If we had available a $\sqrt{T}-$consistent preliminary estimator $\widehat{\boldsymbol{\theta}}_T$, the Gaussian Q-MLE for example, and if we furthermore knew the functional form of the density $g\left(\cdot\right)$ and the expectations $E\left[\frac{\partial\left(vec\mathbf{H}_t^{-1}\right)^T}{\partial\boldsymbol{\theta}}\left(\mathbf{H}_t^{1/2} \quad \mathbf{H}_t^{1/2}\right)\right]$ and $E\left[vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T\right)|\mathbf{w}_t\right]$, then we could compute the following one-step iterative estimator, which would be asymptotically equivalent to the semiparametric efficient estimator $\widetilde{\boldsymbol{\theta}}_T^{**}$:

$$\widetilde{\boldsymbol{\theta}}_T^{*} = \widehat{\boldsymbol{\theta}}_T + \left[\sum_{t=1}^{T}\boldsymbol{\Delta}_{t,T}\left(\widehat{\boldsymbol{\theta}}_T\right)\boldsymbol{\Delta}_{t,T}^T\left(\widehat{\boldsymbol{\theta}}_T\right)\right]^{-1}\boldsymbol{\Delta}_T\left(\widehat{\boldsymbol{\theta}}_T\right), \tag{16}$$

with the asymptotic covariance matrix being estimated by

$$\left[T^{-1}\sum_{t=1}^{T}\boldsymbol{\Delta}_{t,T}\left(\widehat{\boldsymbol{\theta}}_T\right)\boldsymbol{\Delta}_{t,T}^T\left(\widehat{\boldsymbol{\theta}}_T\right)\right]^{-1}.$$

As an alternative information estimator in (16) and in the computation of standard errors we could use

$$\left[-T^{-1}\sum_{t=1}^{T}\frac{\partial\boldsymbol{\Delta}_{t,T}\left(\widehat{\boldsymbol{\theta}}_T\right)}{\partial\boldsymbol{\theta}^T}\right]^{-1}.$$

Of course, it will be infeasible to compute $\widetilde{\boldsymbol{\theta}}_T^{*}$ since the aforementioned density and expectation functions are unknown. We must therefore replace these quantities with nonparametric estimates, for which purpose we draw upon existing results of Brown and Hodgson (2001) and HLV (2001). To estimate $E\left[\frac{\partial\left(vec\mathbf{H}_t^{-1}\right)^T}{\partial\boldsymbol{\theta}}\left(\mathbf{H}_t^{1/2} \quad \mathbf{H}_t^{1/2}\right)\right]$, we can use

$$\widehat{E}\left[\frac{\partial\left(vec\mathbf{H}_t^{-1}\right)^T}{\partial\boldsymbol{\theta}}\left(\mathbf{H}_t^{1/2} \quad \mathbf{H}_t^{1/2}\right)\right]$$

$$= T^{-1}\sum_{t=1}^{T}\frac{\partial\left(vec\mathbf{H}_t^{-1}\left(\widehat{\boldsymbol{\theta}}_T\right)\right)^T}{\partial\boldsymbol{\theta}}\left(\mathbf{H}_t^{1/2}\left(\widehat{\boldsymbol{\theta}}_T\right) \quad \mathbf{H}_t^{1/2}\left(\widehat{\boldsymbol{\theta}}_T\right)\right). \tag{17}$$

Note that the derivative $\frac{\partial\left(vec\mathbf{H}_t^{-1}\right)^T}{\partial\boldsymbol{\theta}}$ is difficult to calculate, as are the derivatives $\frac{\partial\mathbf{u}_t^T}{\partial\boldsymbol{\theta}}$ and $\frac{\partial\ln|\mathbf{H}_t|}{\partial\boldsymbol{\theta}}$, which also appear in our expression for the score. These difficulties are discussed in the Appendix. To estimate the conditional expectation function $E\left[vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T\right)|\mathbf{w}_t\right]$, we make use of the fact that, for elliptically symmetric distributions, the random $n-$vector $\boldsymbol{\varepsilon}$ has a distribution (conditional on $w = \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$) that is uniform on the $(n-1)$-dimensional hypersphere with radius $\sqrt{w}$. As Brown and Hodgson (2001) observe, the desired conditional expectation can be estimated to an arbitrarily high degree of precision by

$$\widehat{E}\left[vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T\right)|\mathbf{w}_t\right] = M^{-1}\sum_{i=1}^{M}vec(\boldsymbol{\varepsilon}_i^*\boldsymbol{\varepsilon}_i^{*T}), \tag{18}$$

where $\boldsymbol{\varepsilon}_i^*$, $i = 1, ..., s$ are *iid* draws from the uniform distribution on hypersphere with radius $\sqrt{w_t}$ and M is chosen sufficiently large to achieve the desired degree of precision. The $\boldsymbol{\varepsilon}_i^*$ are easily computed, as pointed out to us by Werner Ploberger. Draw an iid sequence $\{\widetilde{\boldsymbol{\varepsilon}}_i\}_{i=1}^{M}$ from the n-dimensional standard normal, then compute $\boldsymbol{\varepsilon}_i^* = \sqrt{\frac{w_t(\widehat{\boldsymbol{\theta}}_T)}{\widetilde{\boldsymbol{\varepsilon}}_i^T \widetilde{\boldsymbol{\varepsilon}}_i}} \widetilde{\boldsymbol{\varepsilon}}_i$.

We now consider the problem of deriving nonparametric estimates of the functions $g(\cdot)$ and $g'(\cdot)$. We closely follow the discussion in HLV (2001). Using our preliminary estimator $\widehat{\boldsymbol{\theta}}_T$, we compute the standardized residuals $\left\{\boldsymbol{\varepsilon}_t\left(\widehat{\boldsymbol{\theta}}_T\right)\right\}_{t=1}^{T}$ and the sequence of scalars $w_t\left(\widehat{\boldsymbol{\theta}}_T\right) = \boldsymbol{\varepsilon}_t^T\left(\widehat{\boldsymbol{\theta}}_T\right)\boldsymbol{\varepsilon}_t\left(\widehat{\boldsymbol{\theta}}_T\right)$ for every $t = 1, ..., T$. Next, compute the transformation $z_t = \tau(w_t)$, where the transformation $\tau(\cdot)$ belongs to the Box-Cox (1964) family,

$$\tau(w_t) = \frac{w_t^{\zeta} - 1}{\zeta}.$$

We now compute kernel estimates of the density function $\gamma(z)$ of the transformed random variable z, and of its derivative $\gamma'(z)$, and use these estimates to indirectly obtain estimates of the ratio $\frac{g'(w)}{g(w)}$, as described below. Define the kernel $K_{\sigma_T}(\cdot)$, with a bandwidth parameter σ_T, and use the kernel to compute the following estimates:

$$\widehat{\gamma}_t(z) = (T-1)^{-1} \sum_{\substack{s=1 \\ s\neq t}}^{T} K_{\sigma_T}\left(z - z_s\left(\widehat{\boldsymbol{\theta}}_T\right)\right)$$

and

$$\widehat{\gamma}_t'(z) = (T-1)^{-1} \sum_{\substack{s=1 \\ s\neq t}}^{T} K_{\sigma_T}'\left(z - z_s\left(\widehat{\boldsymbol{\theta}}_T\right)\right).$$

We can then use the estimates $\widehat{\gamma}_t(z)$ and $\widehat{\gamma}_t'(z)$ to estimate the ratio $\frac{g'(w)}{g(w)}$ as follows:

$$\frac{\widehat{g}_t'}{\widehat{g}_t}(w_t) = \begin{cases} s(w_t) + \tau'(w_t)\frac{\widehat{\gamma}_t'}{\widehat{\gamma}_t}(z_t) & \text{if trimming conditions hold} \\ 0 & \text{otherwise} \end{cases}, \tag{19}$$

where $s(w) = (1 - n/2)w^{-1} - \frac{J_\tau'}{J_\tau}\{\tau(w)\}\tau'(w)$ and $J_\tau(z) = \left|\frac{\partial\tau^{-1}(z)}{\partial z}\right|$. The trimming conditions referred to in (19) will depend on the kernel employed. For certain kernels, such as the quartic or the logistic, trimming will not be required. In the Appendix, we provide expressions for the trimming conditions in the case where a Gaussian kernel is the one used. Even in this case, very little trimming (i.e. less than one percent of the observations) has been shown, in another context (Hodgson (1998)), to yield semiparametric estimators that work well in Monte Carlo simulations.

Finally, we have our semiparametric estimator for the period t score:

$$\widehat{\boldsymbol{\Delta}}_{t,T}\left(\widehat{\boldsymbol{\theta}}_T\right) = -\frac{1}{2}\frac{\partial \ln\left|\mathbf{H}_t\left(\widehat{\boldsymbol{\theta}}_T\right)\right|}{\partial\boldsymbol{\theta}} + \widehat{\boldsymbol{\Gamma}}_t\left(\widehat{\boldsymbol{\theta}}_T\right)\frac{\widehat{g}_t'}{\widehat{g}_t}\left(w_t\left(\widehat{\boldsymbol{\theta}}_T\right)\right), \tag{20}$$

where the expectation and score estimators are as defined in (17), (18), and (19), and where

$$\widehat{\mathbf{\Gamma}}_t\left(\widehat{\boldsymbol{\theta}}_T\right) = 2\frac{\partial \mathbf{u}_t^T\left(\widehat{\boldsymbol{\theta}}_T\right)}{\partial \boldsymbol{\theta}}\mathbf{H}_t^{-1}\left(\widehat{\boldsymbol{\theta}}_T\right)\mathbf{u}_t\left(\widehat{\boldsymbol{\theta}}_T\right)$$

$$+\frac{\partial\left(vec\mathbf{H}_t^{-1}\left(\widehat{\boldsymbol{\theta}}_T\right)\right)^T}{\partial \boldsymbol{\theta}}\left(\mathbf{H}_t^{1/2}\left(\widehat{\boldsymbol{\theta}}_T\right) \quad \mathbf{H}_t^{1/2}\left(\widehat{\boldsymbol{\theta}}_T\right)\right) vec\left(\boldsymbol{\varepsilon}_t\left(\widehat{\boldsymbol{\theta}}_T\right)\boldsymbol{\varepsilon}_t^T\left(\widehat{\boldsymbol{\theta}}_T\right)\right)$$
$$-\widehat{E}\left[\frac{\partial\left(vec\mathbf{H}_t^{-1}\left(\widehat{\boldsymbol{\theta}}_T\right)\right)^T}{\partial \boldsymbol{\theta}}\left(\mathbf{H}_t^{1/2}\left(\widehat{\boldsymbol{\theta}}_T\right) \quad \mathbf{H}_t^{1/2}\left(\widehat{\boldsymbol{\theta}}_T\right)\right)\right]\widehat{E}\left[vec\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\right)\middle|w_t\left(\widehat{\boldsymbol{\theta}}_T\right)\right].$$

We then have our semiparametric score estimator for the sample of size T :

$$\widehat{\mathbf{\Delta}}_T\left(\widehat{\boldsymbol{\theta}}_T\right) = \sum_{t=1}^{T}\widehat{\mathbf{\Delta}}_{t,T}\left(\widehat{\boldsymbol{\theta}}_T\right).$$

Our last step in deriving a semiparametric efficient estimator of $\boldsymbol{\theta}$ is to come up with a consistent semiparametric estimator of the expected outer product of the score. To this end, note that

$$E\left[\mathbf{\Delta}_{t,T}\left(\boldsymbol{\theta}_0\right)\mathbf{\Delta}_{t,T}\left(\boldsymbol{\theta}_0\right)^T\right]$$
$$= \frac{1}{4}E\left[\frac{\partial \ln\left|\mathbf{H}_t\left(\boldsymbol{\theta}_0\right)\right|}{\partial \boldsymbol{\theta}}\frac{\partial \ln\left|\mathbf{H}_t\left(\boldsymbol{\theta}_0\right)\right|}{\partial \boldsymbol{\theta}^T}\right] + E\left[\mathbf{\Gamma}_t\left(\boldsymbol{\theta}_0\right)\mathbf{\Gamma}_t\left(\boldsymbol{\theta}_0\right)^T\left(\frac{g'}{g}\left(w_t\left(\boldsymbol{\theta}_0\right)\right)\right)^2\right]. \tag{21}$$

To establish (21), note that

$$E\left[\frac{\partial \ln\left|\mathbf{H}_t\left(\boldsymbol{\theta}_0\right)\right|}{\partial \boldsymbol{\theta}}\mathbf{\Gamma}_t\left(\boldsymbol{\theta}_0\right)\frac{g'}{g}\left(w_t\left(\boldsymbol{\theta}_0\right)\right)\right]$$
$$= E\left[\frac{\partial \ln\left|\mathbf{H}_t\left(\boldsymbol{\theta}_0\right)\right|}{\partial \boldsymbol{\theta}}\frac{\partial \mathbf{u}_t^T\left(\boldsymbol{\theta}_0\right)}{\partial \boldsymbol{\theta}}\mathbf{H}_t^{-1/2}\left(\boldsymbol{\theta}_0\right)\right]E\left[\boldsymbol{\varepsilon}_t\left(\boldsymbol{\theta}_0\right)\frac{g'}{g}\left(w_t\left(\boldsymbol{\theta}_0\right)\right)\right]$$
$$+E\left[\frac{\partial\left(vec\mathbf{H}_t^{-1}\left(\boldsymbol{\theta}_0\right)\right)^T}{\partial \boldsymbol{\theta}}\left(\mathbf{H}_t^{1/2}\left(\boldsymbol{\theta}_0\right) \quad \mathbf{H}_t^{1/2}\left(\boldsymbol{\theta}_0\right)\right)\right]$$
$$\cdot E\left[\left(vec\left(\boldsymbol{\varepsilon}_t\left(\boldsymbol{\theta}_0\right)\boldsymbol{\varepsilon}_t^T\left(\boldsymbol{\theta}_0\right)\right) - E\left[vec\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\right)\left|\mathbf{w}_t\left(\boldsymbol{\theta}_0\right)\right.\right]\right)\frac{g'}{g}\left(w_t\left(\boldsymbol{\theta}_0\right)\right)\right].$$

Now, we have

$$E\left[\boldsymbol{\varepsilon}_t\left(\boldsymbol{\theta}_0\right)\frac{g'}{g}\left(w_t\left(\boldsymbol{\theta}_0\right)\right)\right] = E\left[\frac{g'}{g}\left(w_t\left(\boldsymbol{\theta}_0\right)\right)E\left[\boldsymbol{\varepsilon}_t\left(\boldsymbol{\theta}_0\right)\left|w_t\left(\boldsymbol{\theta}_0\right)\right.\right]\right] = 0$$

because $E\left[\boldsymbol{\varepsilon}_t\left(\boldsymbol{\theta}_0\right)\left|w_t\left(\boldsymbol{\theta}_0\right)\right.\right] = 0$. Equation (21) will then follow because

$$E\left[\left(vec\left(\boldsymbol{\varepsilon}_t\left(\boldsymbol{\theta}_0\right)\boldsymbol{\varepsilon}_t^T\left(\boldsymbol{\theta}_0\right)\right) - E\left[vec\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\right)\left|\mathbf{w}_t\left(\boldsymbol{\theta}_0\right)\right.\right]\right)\frac{g'}{g}\left(w_t\left(\boldsymbol{\theta}_0\right)\right)\right]$$

$$= E\left[E\left[\left(vec\left(\boldsymbol{\varepsilon}_t\left(\boldsymbol{\theta}_0\right)\boldsymbol{\varepsilon}_t^T\left(\boldsymbol{\theta}_0\right)\right) - E\left[vec\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\right)|\mathbf{w}_t\left(\boldsymbol{\theta}_0\right)\right]\right)|w_t\left(\boldsymbol{\theta}_0\right)\right]\frac{g'}{g}\left(w_t\left(\boldsymbol{\theta}_0\right)\right)\right]$$

$$= E\left[E\left[\left(vec\left(\boldsymbol{\varepsilon}_t\left(\boldsymbol{\theta}_0\right)\boldsymbol{\varepsilon}_t^T\left(\boldsymbol{\theta}_0\right)|w_t\left(\boldsymbol{\theta}_0\right)\right) - E\left[vec\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\right)|\mathbf{w}_t\left(\boldsymbol{\theta}_0\right)\right]\right)\right]\frac{g'}{g}\left(w_t\left(\boldsymbol{\theta}_0\right)\right)\right] = 0.$$

In place of the unknown expectation

$$E\left[\boldsymbol{\Gamma}_t\left(\boldsymbol{\theta}_0\right)\boldsymbol{\Gamma}_t\left(\boldsymbol{\theta}_0\right)^T\left(\frac{g'}{g}\left(w_t\left(\boldsymbol{\theta}_0\right)\right)\right)^2\right],$$

we can compute the following semiparametric estimator:

$$\widehat{E}\left[\boldsymbol{\Gamma}_t\left(\boldsymbol{\theta}_0\right)\boldsymbol{\Gamma}_t\left(\boldsymbol{\theta}_0\right)^T\left(\frac{g'}{g}\left(w_t\left(\boldsymbol{\theta}_0\right)\right)\right)^2\right] = T^{-1}\sum_{t=1}^{T}\widehat{\boldsymbol{\Gamma}}_t\left(\widehat{\boldsymbol{\theta}}_T\right)\widehat{\boldsymbol{\Gamma}}_t\left(\widehat{\boldsymbol{\theta}}_T\right)^T\left(\frac{\widehat{g}_t'}{\widehat{g}_t}\left(w_t\left(\widehat{\boldsymbol{\theta}}_T\right)\right)\right)^2,$$

where $\frac{\widehat{g}_t'}{\widehat{g}_t}\left(w_t\right)$ is as defined in (19). We then have the resulting information estimator:

$$\widehat{\mathcal{I}}\left(\widehat{\boldsymbol{\theta}}_T\right) = \frac{1}{4}T^{-1}\sum_{t=1}^{T}\frac{\partial\ln\left|\mathbf{H}_t\left(\widehat{\boldsymbol{\theta}}_T\right)\right|}{\partial\boldsymbol{\theta}}\frac{\partial\ln\left|\mathbf{H}_t\left(\widehat{\boldsymbol{\theta}}_T\right)\right|}{\partial\boldsymbol{\theta}^T} + \widehat{E}\left[\boldsymbol{\Gamma}_t\left(\boldsymbol{\theta}_0\right)\boldsymbol{\Gamma}_t\left(\boldsymbol{\theta}_0\right)^T\left(\frac{g'}{g}\left(w_t\left(\boldsymbol{\theta}_0\right)\right)\right)^2\right].$$

Our semiparametric efficient estimator $\widetilde{\boldsymbol{\theta}}_T$ is then computed in the natural manner:

$$\widetilde{\boldsymbol{\theta}}_T = \widehat{\boldsymbol{\theta}}_T - T^{-1}\widehat{\mathcal{I}}\left(\widehat{\boldsymbol{\theta}}_T\right)^{-1}\widehat{\boldsymbol{\Delta}}_T\left(\widehat{\boldsymbol{\theta}}_T\right). \tag{22}$$

The asymptotic covariance matrix of $\widetilde{\boldsymbol{\theta}}_T$ is consistently estimated by $\widehat{\mathcal{I}}^{-1}\left(\widehat{\boldsymbol{\theta}}_T\right)$.

As remarked in the Introduction, we have no analytical results on the asymptotic behavior of $\widetilde{\boldsymbol{\theta}}_T$ when our iid elliptical symmetry assumption fails, an important consideration for data where some form of conditional or unconditional asymmetry may be present. At present, we can only conjecture as to this behavior. Hodgson (2000) analyzes the robustness to misspecification of semiparametric estimators for a much simpler class of model than is considered here, and it is not clear whether analogous arguments can be applied to the GARCH-M model. However, as a crude check on the robustness of our estimator to skewness, we have allowed for skewed innovations in our Monte Carlo experiment reported in Section 6.

5. EMPIRICAL ANALYSIS

Many econometric tests of the CAPM were published shortly after the development of the theory and have consistently found their way into the finance journals ever since (see, for example, Campbell, Lo, and MacKinlay (1997) for a survey). Early empirical work seemed to support the CAPM, see Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973). The primary methodology used in these early works was to perform cross-sectional regressions of mean returns on estimated betas (which were estimated from some preliminary time series regressions) and other putative variables and thus to test the linearity restriction of the theory. The main econometric problem with this approach is the errors in variables problem that arises from the first stage regressions; one approach to this was to group stocks together into portfolios thereby reducing the estimation error. By grouping according to some factor that might also affect returns, like size, one can improve the power of the test. Most modern tests of the CAPM have been based on the multivariate regression model, see for example Gibbons (1982) and Stambaugh (1982).

It has also become apparent that financial asset returns have distributions that are not constant. This has led to the testing of conditional asset pricing models such as the C-CAPM. Bollerslev, Engle, and Wooldridge (1988) and Harvey (1989, 1991) are well known examples where C-CAPM models are estimated and tested, and the possibility of time-varying conditional betas is also investigated by Braun, Nelson, and Sunier (1995). Our approach here is to estimate a C-CAPM using our model discussed previously and test the model in the traditional framework discussed above. The model that we estimate is a simplified version of (11),

$$\mathbf{r}_t = \boldsymbol{\alpha} + \exp(\gamma_0)\,\mathbf{H}_t\boldsymbol{\omega}_{t-1} + \mathbf{u}_t, \tag{23}$$

where we restrict the market price of risk to be constant. This could be easily relaxed by including instruments; however, to keep the analysis simple, we impose this restriction. The hypothesis tests of the C-CAPM we employ are the Wald statistics discussed in Section 2. It is standard in this literature to work with Wald statistics. Linton and Steigerwald (2000) suggest a method of computing nonparametric likelihood ratio statistics when the likelihood is unspecified, but attempts to apply this method in our model and in the unconditional model of HLV (2001) yielded tests with very erratic behavior, so we do not report any LR test results here.

We use our semiparametric procedure to test the C-CAPM on a data set of daily stock returns. Our data set consists of returns generated from the CRSP data set of stock returns and includes daily observations from January 1996 through December 1997, with a sample size of 759. For this time period we construct three portfolios that are generated by sorting firms traded on the NYSE, AMEX, and NASDAQ according to size (market value). On each trading day firms are placed into quartiles according to the NYSE quartile firm size breakpoints. Daily value-weighted returns are then constructed for the firms in each of the quartiles. We construct three portfolios using the quartile returns. The first portfolio consists of a value weighted return of the first two quartiles. We place both of these quartiles into one portfolio primarily due to the small relative market value of these two quartiles. After combining the two quartiles into one portfolio we still find daily relative weights around 1% of the market. The other two portfolios are constructed using the last two quartiles' returns respectively. The returns on these three portfolios and their corresponding market weights are then used to estimate and test the C-CAPM.

Table 1 provides the summary statistics for the annualized portfolio excess returns $\mathbf{r}_t - r_{f,t}$ while Table 2 provides some group statistics on our three portfolio returns as well as residuals from the Q-MLE estimations. Multivariate normality is rejected using either the univariate kurtosis estimates or the Jarque-Bera (1980) tests performed on the individual series reported in Table 1. The multivariate measures of kurtosis also reject normality as seen in Panel A of Table 2. It appears that a substantial portion of the excess kurtosis is generated by time-varying second moments as evidenced by the decline in the test statistic from the unconditional series to the conditionally weighted series. While the GARCH model removes some of the leptokurtosis, the conditional residual distribution still contains sufficient kurtosis to lead to a rejection of normality. Box-Pierce (1970) tests on the squared residuals indicate that autocorrelation may still be present in the second moments. We did not increase the order of our GARCH(1,1) model because of added estimation complexities that would ensue. However, an application of Beran's (1979) test of elliptical symmetry fails to reject the hypothesis that the weighted excess returns and residuals are distributed elliptically symmetric at the 10% level as seen in Panel B of Table 2.

Table 3 reports the results of estimating an unconditional version of (23) using OLS. The OLS estimates are consistent with the empirical literature in that the estimates of β are positive and the estimates of α are close to zero relative to their standard errors. For the Size 1 portfolio the intercept is positive and statistically significant suggesting that excess returns are generated by holding a portfolio of small stocks.

Table 4 reports the results of estimating (23) using Gaussian Q-MLE techniques. We observe some substantial changes in point estimates of the size 1 and size 3 portfolios. The size one intercept increases from 0.108 to 0.163 while the size 3 intercept declines from -.004 to -.166, a substantial increase in absolute value. Estimates of the conditional covariance matrix appear to be consistent with typical results from estimating GARCH models. All portfolios are significantly influenced by both shocks to volatility ($\mathbf{A}$) and memory in volatility ($\mathbf{D}$). We do find that the portfolio of smaller firms are more (less) influenced by shocks (memory) than are the portfolios of larger firms. These are again consistent with stylized facts of these models.

Table 5 reports the results of estimating (23) using our semiparametric estimator with the Bi-Quartic kernel (similar results were obtained with the Gaussian kernel and are not reported to save space). In computing our kernel estimate of the score we make use of Schuster's (1985) correction. See the Appendix for descriptions of the kernels and of Schuster's correction. We use the Box-Cox transformation with $z_t = \tau(w_t) = \frac{w_t^{\zeta}-1}{\zeta}$ with $\zeta = \frac{1}{2n}$ in construction of the semiparametric estimates. This choice of ζ is found by HLV (2001) to yield good results in Monte Carlo experiments in a linear model. We also use separate optimal MISE bandwidth parameters (σ_T) for estimating $\gamma(z)$ and $\gamma'(z)$. In estimating the conditional expectation $\widehat{E}\left[vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T\right)|\mathbf{w}_t\right]$ as given by (18), we set the number of draws M=500. Results were not sensitive to larger choices for this number.

Standard errors tend to fall somewhat when using the semiparametric efficient estimator rather than the Gaussian Q-MLE, and the point estimates of α using the semiparametric estimator tend to be smaller than for their Q-MLE counterparts. The Wald test statistics of the validity of the CAPM, formed from the α estimates, are given in Table 6. For the unconditional CAPM, we find that OLS leads to a marginal rejection of the CAPM at the 5% level. When we look at the tests of the C-CAPM both estimation methods lead to strong rejections of the model with p-values less than .01.

Although the inferences regarding α are quite similar for the two methodologies, we find some potentially interesting differences in the estimated systematic risk as measured by beta ($\boldsymbol{\beta}_t$). These differences are seen in Tables 4 and 5, listing the parameter estimates, or perhaps more easily observed in Figures 1-3. These figures plot the conditional betas for the three portfolios showing the Q-MLE beta as well as the semiparametrically estimated beta using the Bi-Quartic kernel. We observe that the $\beta_{1,t}$ (size 1 portfolio) tends to be higher for the Q-MLE relative to the semiparametric estimates. However, for the other two size portfolios the estimated $\boldsymbol{\beta}_t$ is greater for the semiparametric estimator than for the Q-MLE. We also find that the variability of $\boldsymbol{\beta}_t$ is greater for the Q-MLE than for the semiparametric estimates. This is true for all of the size portfolios but especially for the size 1 and 2 portfolios. For these, the standard deviation of $\boldsymbol{\beta}_t$ is 48% smaller for the semiparametric estimate then for the Q-MLE estimate. On the other hand the standard deviation of $\beta_{3,t}$ is only 2% smaller for the semiparametric estimate.

We also provide graphs of conditional expected returns for the three portfolios. These graphs are found in Figures 4 through 6. We define the conditional expected return to be:

$$E_{t-1}\mathbf{r}_t = \hat{\boldsymbol{\alpha}} + \exp\left(\hat{\gamma}_0\right)\hat{\mathbf{H}}_t\boldsymbol{\omega}_{t-1}.$$

These graphs incorporate both the intercepts and the conditional betas and give a net effect on the parameters of interest for our semiparametric efficient method relative to Q-MLE methods. In general, we find that the semiparametric efficient method leads to estimates of conditional expected return that are greater than Q-MLE methods. These differences are small for the size 1 portfolio but increase as we move to the larger firm portfolios. We observe that the differences in the estimates of the scaled conditional covariance matrix ($\mathbf{H}_t$) tend to dominate differences in the intercept ($\boldsymbol{\alpha}$). In general, the semiparametric method estimates a larger portion of return due to systematic risk and a smaller portion of return coming from unexplained effects relative to Q-MLE.

6. SIMULATIONS

We simulated series of multivariate GARCH(1,1) time series using the following data generating process,

$$\mathbf{r}_t = \boldsymbol{\alpha} + \exp\left(\gamma_0\right)\mathbf{H}_t\boldsymbol{\omega}_{t-1} + \mathbf{u}_t, \tag{24}$$

with

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}^T + \mathbf{A}\mathbf{u}_{t-1}\mathbf{u}_{t-1}^T\mathbf{A}.$$

We set $n = 2$, $T = 759$, and use the following parameterizations:

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1.15 \end{bmatrix}, \mathbf{A} = \begin{bmatrix} .5 & 0 \\ 0 & .25 \end{bmatrix},$$

and $\gamma_0 = -2.75$. This simulation set-up is a simplification of our empirical model, adopted for the purpose of reducing the computer time required to run the simulations. We use the same $\boldsymbol{\omega}_{t-1}$ from our empirical analysis but reduce the dimension to a 2×1 vector by combining the smaller two decile weights into one. We also simulate data using two different $\boldsymbol{\alpha}$ vectors, $\boldsymbol{\alpha} = \{0, 0\}$ under the null and $\boldsymbol{\alpha} = \{-.15, .15\}$ under the alternative.

We add a randomly selected residual ($\mathbf{u}_t$) from some prespecified distribution. We consider a normal, a mixture of normals, a Student-t with 3 degrees of freedom, and Chi-Square with 5 degrees of freedom. The first three distributions are elliptical, while the third is asymmetric and is included as a check on the robustness of our estimator to misspecification. To compute the mixture of normals, we first define the uniform random variable $U \in [0, 1]$.

If $U < (1 - \epsilon)$, then let $\mathbf{u}_t = \sqrt{\kappa_1}\breve{\mathbf{u}}_t$, where $\breve{\mathbf{u}}_t \sim N(0,1)$. Otherwise, we let $\mathbf{u}_t = \sqrt{\kappa_2}\breve{\mathbf{u}}_t$, . The resulting $\mathbf{u}_t$ will follow a mixed normal distribution. We set $\epsilon = .8$, $\kappa_1 = 0.65$ in the simulations, and for all distributions the errors are scaled to have unit variances. We use the same residual in constructing both the alternative and null series. For each simulation we estimate (24) using Q-MLE and the semiparametric efficient estimator. We replicate each simulation 2,000 times for each distribution and report the results of the simulations in Tables 7 - 9.

Table 7 reports bias, standard deviation, and mean squared error (MSE) for the estimators for the four different distributions. For the nonnormal elliptical densities, the semiparametric estimator (SE) yields only slight improvements in estimation of the intercepts, with larger improvements found in the estimation of the conditional variance parameters. This is consistent with our empirical study, where we found that the SE point estimates had greater impact on conditional covariances than on intercepts. Estimation of the risk aversion parameter γ_0 deteriorates when we move from the Q-MLE to the SE, but neither estimator accurately estimates this parameter. We should point out that for the purposes of the present paper, γ_0 is not of substantive interest, as we have focused our attention on testing for zero interecepts and estimating betas, for both of which problems γ_0 can be thought of as a nuisance parameter. Note that in the case of asymmetric errors, the SE provides reasonably good estimates of most of the parameters. This is important because, recalling our earlier comments, we have no theoretical results on the behaviour of the SE under asymmetry, but the simulation results suggest that the SE may be robust to asymmetry. For the case of normal errors we see deterioration in the SE estimation of conditional mean parameters, where MSE are at least twice as large as their MLE counterparts. The SE estimates of conditional variance parameters are much closer to their MLE counterparts.

Table 8 compares the simulation results in estimation of beta. For each simulation, we use the true parameter values and the simulated residuals to construct a time series of 'true' betas $\left(\boldsymbol{\beta}_0^i\right)$, where i indexes the simulation. We compute the average values of $\boldsymbol{\beta}_0^i$ for each portfolio over the time series as well as the standard deviations of $\boldsymbol{\beta}_0^i$, $\left(\sigma_{\boldsymbol{\beta}_{0,j}^i}\right)$ for each simulation. We then define $\beta_{0,j} = \frac{1}{2000}\sum_{i=1}^{2000}\boldsymbol{\beta}_{0,j}^i$, and $\sigma_{\beta_{0,j}} = \frac{1}{2000}\sum_{i=1}^{2000}\sigma_{\boldsymbol{\beta}_{0,j}^i}$ for $j = 1, 2$. We construct this same measure using the parameter and residual estimates from the two estimation methods. The final two measures reported in Table 8 are constructed by looking at absolute and squared differences between the estimated and 'true' conditional betas for each simulation and then averaging over the simulations. One apparent advantage of the semiparametric estimator is that estimated volatilities of the conditional betas are closer to the 'true' beta volatilities than for the Q-MLE estimates. Note that in our empirical application, the SE produces *less* volatile estimated betas than the Q-MLE, while the reverse is true in the simulations. We are not sure why this is the case, althought the simulation set-up is different in a couple of important ways from the empirical model, which may explain the difference in results. The important point to take note of, in our view, is that the SE betas have a volatility that is closer to the true beta volatility than the Q-MLE betas. We also note that the performance of the SE estimated betas for the simulation with normal residuals indicates that there is a loss relative to the MLE estimator. However, the losses for this case are approximately of the same order as the gains in the nonnormal simulations and given the prevalence of nonnormality in the data, the performance of the SE beta estimates are remarkably good.

Table 9 considers the Wald tests of the zero-intercept null hypothesis. We calculate the empirical size and power of the test statistics for the two estimation methods as discussed in Davidson and MacKinnon (1998) using the p-values from each test statistic. The power results are adjusted for any biases in size. The two methods lead to quite similar size and power properties in the asset pricing tests for all of the cases other than normality; the SE being slightly more over-sized but having slightly higher size-corrected power. Both methods lead to reasonably sized tests for the elliptical distributions but over-reject for the asymmetric χ^2_5 distribution. The MLE method has substantially greater power than the SE method for the case of normality.

7. CONCLUSION

We propose a new estimation methodology that captures the nonnormalities of return distributions arising from tail thickness by employing a multivariate GARCH-in-mean model with the flexible distributional assumption of conditional elliptical symmetry. Under high level assumptions, this framework should lead to more efficient estimates than quasi-MLE and should yield more powerful asset pricing tests. We find in empirical and simulation analysis that our estimator does not improve significantly over the Gaussian Q-MLE in the estimation of conditional mean parameters, but that the semiparametric efficient estimates of the conditional betas do improve on the Q-MLE estimates to a degree that may be of potential interest to applied workers.

Further work on the properties of our estimator in the presence of specification failure is suggested. In particular, the work of Harvey and Siddique (1999), among others, suggests that a derivation of the semiparametric efficiency bounds of the GARCH-in-mean model with conditional densities that are not required to be symmetric would be a useful contribution to this research. We have seen in our simulations that the estimator in this paper does not misbehave too badly in the presence of asymmetric errors, but it would be desirable to have an estimator that explicitly accounts for the possibility of asymmetry. Such an estimator would have to take account of the fact that the conditional location is unidentified in the presence of asymmetric errors of unknown distributional form, and would require the use of multivariate generalizations of the approaches taken by Newey and Steigerwald (1997) and Drost and Klaassen (1997), both of which studies analyze univariate GARCH models with possibly asymmetric conditional densities of unknown functional form.

8. APPENDIX

8.1. Elliptical Densities[1]. An n-dimensional random vector u is said to be elliptically distributed about the origin if its density can be written as

$$p(u) = (\det \Sigma)^{-1/2}\, g\left(u'\Sigma^{-1}u\right),$$

where Σ is a positive definite, symmetric matrix that is proportional to the covariance matrix of u (when a finite covariance matrix exists) and is also proportional to the inverse of the

[1]See Fang, Kotz, and Ng (1990) and Owen and Rabinovitch (1983) for further discussion.

information matrix of p. The characteristic function of u is

$$\psi(s) = E\left[\exp\left(is^T u\right)\right] = \phi\left(s^T \Sigma s\right)$$

for some function $\phi(\cdot)$. The standardized n-vector $\varepsilon = \Sigma^{-1/2}u$ is said to be spherically symmetric, with density

$$p(\varepsilon) = g\left(\varepsilon^T \varepsilon\right).$$

Note that the isoprobability contours of the density of the elliptical random variable u will be elliptical in shape, and those of the spherical random variable ε will be spherical (circular in the case of $n = 2$).

Some examples of spherical densities are:

(a) the Gaussian,

$$g(\cdot) = const \cdot \exp\left(-\frac{\varepsilon^T \varepsilon}{2}\right);$$

(b) the Student's t with τ degrees of freedom,

$$g(\cdot) = const \cdot \left(1 + \frac{\varepsilon^T \varepsilon}{\tau}\right)^{-(n+\tau)/2};$$

(c) the Cauchy,

$$g(\cdot) = const \cdot \left(1 + \varepsilon^T \varepsilon\right)^{-(n+1)/2};$$

(d) the logistic,

$$g(\cdot) = const \cdot \exp\left(-\varepsilon^T \varepsilon\right) / \left[1 + \exp\left(-\varepsilon^T \varepsilon\right)\right]^2;$$

(e) and the scale mixed normal,

$$g(\cdot) = const \cdot \int_0^\infty s^{-n/2} \exp\left(-\frac{\varepsilon^T \varepsilon}{2s}\right) dF(s)$$

for some cdf $F(\cdot)$. Note that all the non-Gaussian densities listed here feature thick tails and some of them are popular candidates for modeling tail thickness in empirical work that takes a fully parametric tack.

Elliptical distributions have a few properties that are of interest. First, define the norm $\|\varepsilon\| = \sqrt{\varepsilon^T \varepsilon}$. The random variables $\frac{\varepsilon}{\|\varepsilon\|}$ and $\|\varepsilon\|$ are independent of one another. Furthermore, the random variable $\frac{\varepsilon}{\|\varepsilon\|}$ has a uniform distribution on the $(n-1)$-dimensional unit hypersphere. These two features of elliptical distributions form the basis for Beran's (1979) test for elliptical symmetry, while the latter fact plays a central role in our derivation of the semiparametric efficiency bound for this model, following the results of Brown and Hodgson (2000).

Define the $n^* \times n$ matrix Φ, of rank $n^* \leq n$. Then the n^*-dimensional random variable Φu is elliptically symmetrically distributed with characteristic matrix of $\Phi\Sigma\Phi^T$. Define the partition $u = \left(u_1^T, u_2^T\right)^T$ and partition Σ conformably as

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Then the marginal densities of u_1 and u_2 are of the same form as the joint density of u, with respective characteristic matrices of Σ_{11} and Σ_{22}. The conditional mean can be written as

$$E\left[u_i \left| u_j\right.\right] = \Sigma_{ij}\Sigma_{jj}^{-1}u_j.$$

Furthermore, the density of u_i conditional on u_j will be elliptically symmetric with a characteristic matrix of $\Sigma_{ii} - \Sigma_{ij}\Sigma_{jj}^{-1}\Sigma_{ji}$.

Many of these characteristics of elliptical distributions are well known among economists to apply to the Gaussian density. That they also apply to the more general elliptical family explains why the unconditional CAPM also holds in this case, a point which is discussed in more detail by Owen and Rabinovitch (1983).

8.2. Computation of Derivatives. We remark here on the difficulty of obtaining expressions for the derivatives $\frac{\partial \ln|\mathbf{H}_t|}{\partial \boldsymbol{\theta}}$, $\frac{\partial \mathbf{u}_t^T}{\partial \boldsymbol{\theta}}$ and $\frac{\partial \left(vec\mathbf{H}_t^{-1}\right)^T}{\partial \boldsymbol{\theta}}$. The basic problem is that each of these derivatives involves an infinite recursion, since the expression for $\frac{\partial \mathbf{u}_t^T}{\partial \boldsymbol{\theta}}$, for example, involves $\frac{\partial (\mathbf{H}_t)^T}{\partial \boldsymbol{\theta}}$, which in turn involves $\frac{\partial \mathbf{u}_{t-1}^T}{\partial \boldsymbol{\theta}}$, and so on. Our practical approach is to construct the derivatives by assuming that $\frac{\partial vec(\mathbf{H}_0)}{\partial \boldsymbol{\theta}}$ and $\frac{\partial \mathbf{u}_0}{\partial \boldsymbol{\theta}}$ take on their unconditional values which allow us to obtain $\frac{\partial vec(\mathbf{H}_1)}{\partial \boldsymbol{\theta}}$ and $\frac{\partial \mathbf{u}_1}{\partial \boldsymbol{\theta}}$. Given the derivatives for period one we can construct the same for period two and continue in a likewise manner to construct the derivatives for all T periods. We could also have assumed that $\frac{\partial vec(\mathbf{H}_0)}{\partial \boldsymbol{\theta}}$ and $\frac{\partial \mathbf{u}_0}{\partial \boldsymbol{\theta}}$ are zero, following Drost and Klaassen (1997), but have found in preliminary calculations that the empirical properties of the estimator were quite robust to the assumptions placed on $\frac{\partial vec(\mathbf{H}_0)}{\partial \boldsymbol{\theta}}$ and $\frac{\partial \mathbf{u}_0}{\partial \boldsymbol{\theta}}$. As stated earlier, the asymptotics of the estimator should not depend upon the assumptions of the initial period.

8.3. Kernels and Trimming. The two kernels we consider are the bi-quartic,

$$K_\sigma(u) = \frac{15}{16}\left(1 - \frac{u^2}{\sigma^2}\right)^2 \mathbb{I}\left\{\left|\frac{u}{\sigma}\right| \leq 1\right\},$$

and the Gaussian

$$K_\sigma(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma^2}\right).$$

The bi-quartic is applied without trimming. To establish consistency of the Gaussian kernel estimator, it is sufficient to apply the following trimming conditions, as shown by HLV (2001):

(i) $\widehat{\gamma}_t(z) \geq d_T$,

(ii) $|z| \leq e_T$,

(iii) $|\lambda(z)| \leq b_T$,

(iv) $\left|\rho^{1/2}(z)\widehat{\gamma}_t'(z)\right| \leq c_T\widehat{\gamma}_t(z)$,

where $\rho(z) = w\tau'(w) J_\tau^{-1}(w)$ (recall that $w = \tau^{-1}(z)$), $J_\tau(z) = \left|\frac{\partial \tau^{-1}(z)}{\partial z}\right|$, and $(d/dz)^{-1}\rho^{1/2}(z)$. The constants σ_T, d_T, e_T, b_T, and c_T have the properties that, as $T \to \infty$, we have $\sigma_T \to 0$, $c_T \to \infty$, $e_T \to \infty$, $b_T \to \infty$, $d_T \to 0$, $\sigma_T c_T \to 0$, $e_T\sigma_T^{-3} = o(T)$, and $b_T\sigma_T^{-3} = o(T)$.

8.4. Schuster's Correction. For most standard choices of symmetric kernel, the density estimator $f_T(z)$ typically performs poorly on the right neighborhood of zero. This bias arise because for points x_i in the right neighborhood of 0, the contribution of x_i given by $T^{-1}K_{\sigma_T}(x - x_i)$to $f_T(x)$ extends to points $x \leq 0$ where $f(x) = 0$. A similar bias arise in the multivariate density estimates which impose the elliptical symmetry restriction. This bias creates a volcano like contour in the density estimate. The overflow in weights beyond the lower support of 0 can be corrected by using an estimator which incorporates this additional support constraint information into $f_T(x)$.

Schuster (1985) offers a correction that incorporates this overflow to the region $z < c$, for finite c, back into the region $z \geq c$ by adding a mirror image term $T^{-1}K_{\sigma_T}(z - 2c + z_s)$ to $T^{-1}K_{\sigma_T}(z - z_s)$. The resulting estimator for $z \geq c$ is given by

$$\widehat{\gamma}_t(z) = (T-1)^{-1} \sum_{\substack{s=1 \\ s \neq t}}^{T} \left[K_{\sigma_T}(z - \overline{z}_s) + K_{\sigma_T}(z - 2c + \overline{z}_s) \right].$$

In our case, $c = 0$. Schuster (1985) also proves consistency and asymptotic normality results for this estimator.

References

[1] Beran, R., 1979, Testing for ellipsoidal symmetry of a multivariate density, *Annals of Statistics,* 7:150-162.

[2] Berk, J., 1997, Necessary conditions for the CAPM, *Journal of Economic Theory* 73, 245-257.

[3] Bickel, P. J., 1982, On adaptive estimation, *Annals of Statistics* 10, 647-671.

[4] Black, F., M. Jensen, and M. Scholes, 1972, The capital asset pricing model: some empirical tests, in Jensen, M. (ed), *Studies in the Theory of Capital Markets*, New York, Praeger.

[5] Bollerslev, T., 1987, A conditional heteroskedastic time series model for speculative prices and rates of return, *Review of Economics and Statistics* 69, 542-547.

[6] Bollerslev, T., R. Chou, and K. Kroner, 1992, ARCH modelling in finance: A review of the theory and empirical evidence, *Journal of Econometrics* 52, 5-59.

[7] Bollerslev, T., R. Engle, and J. Wooldridge, 1988, A capital asset pricing model with time varying covariances, *Journal of Political Economy* 96, 116-131.

[8] Bollerslev, T. and J. Wooldridge, 1992, Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, *Econometric Reviews* 11, 143-172.

[9] Box, G. and D. Cox, 1964, An analysis of transformations, *Journal of the Royal Statistical Society, Series B,* 211-264.

[10] Box, G. P., and D. A. Pierce, 1970, Distribution of residual autocorrelations in autoregressive integrated moving average time series models, *Journal of the American Statistical Association*, 65, 1509-1526.

[11] Braun, P.A., Nelson, D.B. and Sunier, A.M. 1995. Good news, bad news, volatility, and betas. *Journal of Finance* 50:1575-1603.

[12] Brown, B.W. and D. Hodgson, 2000, Efficient semiparametric estimation of dynamic nonlinear systems under elliptical symmetry, Unpublished manuscript.

[13] Buse, A., B. Korkie, and H. Turtle, 1994, Tests of conditional asset pricing with time-varying moments and risk prices, *Journal of Financial and Quantitative Analysis*, 29, 15-29.

[14] Campbell, J., W. Lo, and A. C. MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton, Princeton University Press.

[15] Chamberlain, G., 1983, A characterization of the distributions that imply mean-variance utility functions, *Journal of Economic Theory* 29, 185-201.

[16] Davidson, R., and J., MacKinnon, 1998, Graphical methods for investigating the size and power of hypothesis tests, *The Manchester School* 66, 1-26.

[17] De Santis, G., and B. Gerard, 1998, International asset pricing and portfolio diversification with time-varying risk, *Journal of Finance*, forthcoming.

[18] Diebold, F., 1988, *Empirical Modeling of Exchange Rate Dynamics,* New York, Springer-Verlag.

[19] Drost, F. and C. Klaassen, 1997, Efficient estimation in semiparametric GARCH models, *Journal of Econometrics,* 81, 193-221.

[20] Engle, R., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation, *Econometrica,* 50, 987-1008.

[21] Engle, R., and K. Kroner, 1995, Multivariate simultaneous generalized ARCH, *Econometric Theory* 11, 122-150.

[22] Fama, E., 1965, The behavior of stock market prices, *Journal of Business* 38, 34-105.

[23] Fama, E. and J. MacBeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607-636.

[24] Fang, K., S. Kotz, and K. Ng, 1990, *Symmetric Multivariate and Related Distributions,* London, Chapman and Hall.

[25] Gibbons, M., 1982, Multivariate tests of financial models: A new approach. *Journal of Financial Economics* 10, 3-27.

[26] Harvey, C., 1989, Time-varying conditional covariances in tests of asset pricing models, *Journal of Financial Economics* 24, 289-317.

[27] Harvey, C., 1991, The world price of covariance risk, *Journal of Finance* 46, 111-157.

[28] Harvey, C. and A. Siddique, 1999, Autoregressive conditional skewness, *Journal of Financial and Quantitative Analysis*, 34:465-487.

[29] Hodgson, D.J., 1998, Adaptive estimation of cointegrating regressions with ARMA errors, *Journal of Econometrics,* 85:231-267.

[30] Hodgson, D., O. Linton, and K. Vorkink, 2001, Testing the capital asset pricing model efficiently under elliptical symmetry: a semiparametric approach. In press, *Journal of Applied Econometrics*.

[31] Jagannathan, R., and Z. Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3-53.

[32] Jarque, C. and A. Bera, 1980, Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals, *Economics Letters* 6, 255-259.

[33] Jeantheau, T. 1998. Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* 14:70-86.

[34] Linton, O., 1993, Adaptive estimation in ARCH models, *Econometric Theory*, 9, 539-569.

[35] Linton, O.B. and D.G. Steigerwald, 2000, Adaptive testing in ARCH models, *Econometric Reviews,* 19, 145-174.

[36] MacKinlay, A. C., and M. Richardson, 1991, Using generalized method of moments to test mean-variance efficiency, *Journal of Finance* 46, 511-527.

[37] Mandelbrot, B., 1963, The variation of certain speculative prices, *Journal of Business* 36, 394-419.

[38] Mardia, K.V., 1970, Measures of multivariate skewness and kurtosis with applications, *Biometrika* 57, 519-530.

[39] Merton, R., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867-887.

[40] Mitchell, A., 1989, The information matrix, skewness tensor, and α-connections for the general multivariate elliptic distribution, *Annals of the Institute of Mathematical Statistics,* Tokyo, 41, 289-304.

[41] Muirhead, R., 1982, *Aspects of Multivariate Statistical Theory,* New York, Wiley.

[42] Nelson, D., 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59, 347-370.

[43] Newey, W.K., 1990, Semiparametric efficiency bounds, *Journal of Applied Econometrics,* 5:99-135.

[44] Newey, W.K. and D.G. Steigerwald, 1997, Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models, *Econometrica*, 65:587-599.

[45] Owen, J., and R. Rabinovitch, 1983, On the class of elliptical distributions and their applications to the theory of portfolio choice, *Journal of Finance* 38, 745-752.

[46] Schuster, E., 1985, Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics - Theory and Methods* 14, 1123-1136.

[47] Schwert, G., and P. Seguin, 1990, Heteroskedasticity in stock returns, *Journal of Finance* 45, 1129-1155.

[48] Stambaugh, R., 1982, On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis. *Journal of Financial Economics* 10, 237-268.

[49] Vorkink, K., 2001, Return distributions and improved tests of the capital asset pricing model, Working paper, Brigham Young University.

[50] White, H., 1982, Maximum likelihood estimation of misspecified models, *Econometrica* 50:1-25.

[51] Zhou, G., 1993, Asset pricing tests under alternative distributions, *Journal of Finance* 48, 1927-1942.

Table 1. Summary Statistics

Portfolios	Mean	Std. Dev.	min	max	Kurtosis	J-B
Size 1	0.276	1.870	-19.377	8.510	73.313*	11513*
Size 2	0.255	2.231	-21.590	9.548	46.721*	6747*
Size 3	0.332	2.799	-24.301	15.453	45.721*	2890*

Note: This table provides statistical characteristics of the portfolios of excess returns used in our in our empirical analysis. The stock returns are obtained from the CRSP data set. The series are daily returns from Jan. 1995 through Dec. 1997. Estimates of kurtosis have been scaled so that under the assumption of normality the statistics have an asymptotic N(0,1) distribution. J-B refers to the Jarque-Bera (1980) test for normality. *Refers to a rejection of the hypothesis that the given moment is consistent with the Normal distribution at the .01 level.

Table 2. Multivariate Tests of Conditional Normality and Elliptical Symmetry

Panel A: Multivariate Kurtosis Test

	Size Portfolios
Unconditional Returns	17.054*
Unconditional Residuals	17.726*
Conditional Returns	4.395*
Conditional Residuals	6.167*

Panel B: Elliptical Symmetric Test (S_n)

	Size Portfolios
Conditional Returns	1.599[1]
Conditional Residuals	0.426[1]

Note: The test statistics below are Mardia's (1970) multivariate kurtosis measure and Beran's (1979) test for elliptical symmetry. Tests are constructed using the series of portfolio returns ($\mathbf{r}_t$) and residuals (u_t) and where both series are weighted by the matrix $H_t^{1/2}$. The multivariate kurtosis measure has been scaled so that assuming normality the statistic will have an asymptotic N(0,1) distribution. *Indicates a p-value less than .01. [1]Indicates a p-value greater than .1.

Table 3. OLS Estimation of the Unconditional CAPM

	α		β_{im}	
Portfolio	Estimate	Std. Error	Estimate	Std. Error
Size 1	0.108	0.050	0.475	0.018
Size 2	0.032	0.049	0.659	0.018
Size 3	-0.004	0.004	1.031	0.001

Note: Data are from the CRSP data set of stocks listed on the NYSE, AMEX, and NASDAQ. Value-weighted returns are calculated daily from January 1996 through December 1997. Three size portfolios are created according to the previous day's market value of equity. The previous day's NYSE size quartiles are used as the cutoffs for the size portfolios. The CAPM model takes on the following parameterization: $\mathbf{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta} r_{m,t} + \mathbf{u}_t$. Above are results from using OLS estimation methods in estimating the CAPM.

Table 4. Q-ML Estimation of The C-CAPM

Parameter	Size 1	Size 2	Size 3
α	0.163	0.054	-0.166
	(0.028)	(0.023)	(0.043)
γ_0	-2.508		
	(0.173)		
	$\mathbf{C}_1$	$\mathbf{C}_2$	$\mathbf{C}_3$
$\mathbf{C}_1$	1	-	-
	-	-	-
$\mathbf{C}_2$	1.011	0.348	-
	(0.036)	(0.045)	-
$\mathbf{C}_3$	0.807	0.072	-0.001
	(0.057)	(0.091)	(0.050)
$\mathbf{A}$	0.532	0.409	0.263
	(0.033)	(0.028)	(0.021)
$\mathbf{D}$	0.668	0.765	0.920
	(0.022)	(0.023)	(0.007)

Note: Data are from the CRSP data set of stocks listed on the NYSE, AMEX, and NASDAQ. Value-weighted returns are calculated daily from Jan. 1995 through Dec. 1997. Three size portfolios are created according to the previous day's market value of equity. The previous day's NYSE size quartiles are used as the cutoffs for the size portfolios as well as for construction of weights. The first two quartiles are grouped into the first size portfolio with the remaining two quartiles each representing the other two portfolios. The C-CAPM takes on the following parameterization: $\mathbf{r}_t = \boldsymbol{\alpha} + \exp(\gamma_0)\,\mathbf{H}_t\boldsymbol{\omega}_{t-1} + \mathbf{u}_t$, with the following scaled conditional variance parameterization: $\mathbf{H}_t = \mathbf{C}\mathbf{C}^T + \mathbf{A}\mathbf{u}_{t-1}\mathbf{u}_{t-1}^T\mathbf{A} + \mathbf{D}\mathbf{H}_{t-1}\mathbf{D}$. Point estimates of the parameters are reported as well as standard errors which are reported below the point estimates in parantheses.

Table 5. Semiparametric Efficient Estimation of the C-CAPM

Parameter	Size 1	Size 2	Size 3
α	0.145	0.027	-0.269
	(0.022)	(0.018)	(0.036)
γ_0	-2.665		
	(0.133)		
	$\mathbf{C}_1$	$\mathbf{C}_2$	$\mathbf{C}_3$
$\mathbf{C}_1$	1	-	-
	-	-	-
$\mathbf{C}_2$	1.319	0.459	-
	(0.030)	(0.033)	-
$\mathbf{C}_3$	1.127	0.266	0.005
	(0.042)	(0.061)	(0.018)
$\mathbf{A}$	0.507	0.423	0.263
	(0.030)	(0.021)	(0.018)
$\mathbf{D}$	0.720	0.811	0.920
	(0.015)	(0.015)	(0.004)

Note: Data are from the CRSP data set of stocks listed on the NYSE, AMEX, and NASDAQ. Value-weighted returns are calculated daily from Jan. 1995 through Dec. 1997. Three size portfolios are created according to the previous day's market value of equity. The previous day's NYSE size quartiles are used as the cutoffs for the size portfolios as well as for construction of weights. The first two quartiles are grouped into the first size portfolio with the remaining two quartiles each representing the other two portfolios. The C-CAPM takes on the following parameterization: $\mathbf{r}_t = \boldsymbol{\alpha} + \exp(\gamma_0)\,\mathbf{H}_t\boldsymbol{\omega}_{t-1} + \mathbf{u}_t$, with the following scaled conditional variance parameterization: $\mathbf{H}_t = \mathbf{C}\mathbf{C}^T + \mathbf{A}\mathbf{u}_{t-1}\mathbf{u}_{t-1}^T\mathbf{A} + \mathbf{D}\mathbf{H}_{t-1}\mathbf{D}$. Point estimates of the parameters are reported as well as standard errors which are reported below the point estimates in parantheses. Estimates are obtained using the Semiparametric Efficient procedure. We report estimates in this table using the Bi-Quartic Kernel and the transformation function: $\tau(w_t) = (w_t^{\zeta} - 1)/\zeta$.

Table 6. Mean-Variance Efficiency Tests

Unconditional CAPM	J (p-value)
OLS	7.95(0.05)
C-CAPM	
Q-MLE	264.11(0.00)
SE	543.33(0.00)

Note: The test statistics above are constructed using the intercepts from estimating the C-CAPM. Mean-variance efficiency implies that the intercepts are jointly equal to zero: $H_0: \quad \alpha_i = 0 \quad i = 1, \ldots, n.$ The table lists the tests that result from estimating the unconditional CAPM model via OLS, the C-CAPM using Q-MLE techniques as well as efficient procedure using the Bi-Quartic kernel. Under the null that the C-CAPM is the true model, J is distributed asymptotically $\chi^2(3)$. SE corresponds to the semiparametric efficient estimator.

Table 7. Parameter Estimate Results for Simulation Study

Parameter		MN			t_3			χ^2_5			Normal		
		Bias	Std. Dev.	MSE	Bias	Std. Dev.	MSE	Bias	Std. Dev	MSE	Bias	Std. Dev	MSE
α_0	Q-MLE	-0.004	0.053	0.003	-0.011	0.081	0.007	0.133	0.083	0.025	-0.015	0.050	0.003
	SE	0.005	0.052	0.003	-0.020	0.078	0.006	0.160	0.081	0.032	0.025	0.089	0.009
α_1	Q-MLE	-0.007	0.059	0.004	-0.024	0.043	0.002	-.162	0.072	0.031	-0.019	0.079	0.007
	SE	-0.004	0.058	0.003	-0.019	0.044	0.002	-0.160	0.074	0.031	-0.010	0.121	0.015
γ_0	Q-MLE	-0.229	1.006	1.064	-0.431	1.169	1.552	-0.013	1.053	1.109	-0.435	0.906	1.010
	SE	-0.398	1.731	3.154	-0.497	2.404	6.026	-0.196	1.781	3.210	-1.207	2.075	5.762
C_2	Q-MLE	-0.017	0.235	0.056	-0.106	0.286	0.093	-0.030	0.238	0.058	-0.011	0.187	0.035
	SE	0.015	0.223	0.050	-0.097	0.279	0.087	0.006	0.252	0.064	0.068	0.208	0.047
A_1	Q-MLE	-0.170	0.161	0.055	-0.106	0.259	0.078	-0.173	0.173	0.060	-0.186	0.145	0.056
	SE	-0.074	0.157	0.030	-0.076	0.251	0.068	-0.011	0.180	0.033	-0.280	0.162	0.105
A_2	Q-MLE	-0.129	0.119	0.031	0.086	0.216	0.054	-0.123	0.173	0.045	-0.142	0.122	0.035
	SE	-0.111	0.105	0.023	0.151	0.208	0.066	-0.082	0.167	0.034	-0.117	0.164	0.041

Note: This table lists the estimation results from a Monte Carlo study. Four different simulations were performed where the residuals were drawn from three different distributions: mixed normal (MN), t_3, χ^2_5 , and Normal as indicated on the top row. The following parameter values were used in the simulation study: $\alpha_0 = \alpha_1 = 0$, $\gamma_0 = -2.50$, $C_2 = 1.15$, $A_1 = .5$, and $A_2 = .25$. Each series had a length of 759, with 2,000 replications performed for each of the three distributions. Q-MLE and semiparametric efficient (SE) methods are used to estimate the model. The bias (Average estimated value - True value), standard deviation of the parameter estimates, and Mean Squared Error (MSE) are reported.

Table 8. Analysis of β performance from Simulation Study

Distribution	Performance Measure	$\beta_{0,1}$	$\beta_{0,2}$	$\beta_{\text{Q-MLE},1}$	$\beta_{\text{Q-MLE},2}$	$\beta_{\text{SE},1}$	$\beta_{\text{SE},2}$
MN	$\bar{\beta}_t$	1.055	0.929	1.04	0.941	1.066	0.931
	$\sigma_{\bar{\beta}_t}$	0.135	0.166	0.127	0.152	0.132	0.162
	Average $abs\,(\beta_i - \beta_0)$	-	-	0.108	0.132	0.099	0.104
	Average $(\beta_i - \beta_0)^2$	-	-	0.027	0.040	0.025	0.036
t_3	$\bar{\beta}_t$	1.068	0.914	0.993	1.005	1.004	0.983
	$\sigma_{\bar{\beta}_t}$	0.086	0.099	0.060	0.068	0.068	0.079
	Average $abs\,(\beta_i - \beta_0)$	-	-	0.084	0.102	0.078	0.098
	Average $(\beta_i - \beta_0)^2$	-	-	0.014	0.021	0.011	0.018
χ^2_5	$\bar{\beta}_t$	1.056	0.925	1.051	0.935	1.059	0.930
	$\sigma_{\bar{\beta}_t}$	0.123	0.147	0.118	0.133	0.128	0.152
	Average $abs\,(\beta_i - \beta_0)$	-	-	0.108	0.130	0.103	0.125
	Average $(\beta_i - \beta_0)^2$	-		0.028	0.042	0.028	0.039
Normal	$\bar{\beta}_t$	1.066	0.915	1.046	0.940	1.040	0.937
	$\sigma_{\bar{\beta}_t}$	0.004	0.005	0.114	0.139	0.118	0.142
	Average $abs\,(\beta_i - \beta_0)$	-	-	0.092	0.112	0.109	0.136
	Average $(\beta_i - \beta_0)^2$	-	-	0.019	0.028	0.023	0.039

Note: This table lists various measures of βperformance of the Q-MLE and semiparametric efficient estimator. The measure $\bar{\beta}_t$ is constructed by taking the sample average of $\beta_t$for each simulation and then averaging over the simulations. $\sigma_{\bar{\beta}_t}$is constructed by taking the standard deviation of the average $\beta_t s$over the simulations. The notation $\beta_0$corresponds to true beta as opposed to estimated β. The measure Average $abs\,(\beta_i - \beta_0)$ is constructed by taking the average absolute difference between the estimated $\beta_t$and $\beta_0$and then averaging over the simulations. Average $(\beta_i - \beta_0)^2$ reports where squared differences are taken as opposed to absolute differences. SE represents the semiparametric efficient estimator.

Table 9. Conditional Mean-Variance Tests from Simulation Study

	Size		Power (size-corrected)	
Distribution	Q-MLE	SE	Q-MLE	SE
MN	.055	.058	.71	.73
t_3	.054	.057	.66	.66
χ^2_5	.071	.078	.62	.64
Normal	.044	.039	.78	.58

Note: The table above lists results of conditional mean-variance tests as shown in (6) We list both size and power performance at the .05 level. The power reports listed above have been adjusted for any size problems. MN represents the mixed-normal distributional assumption, while SE represents the semiparametric efficient estimator.

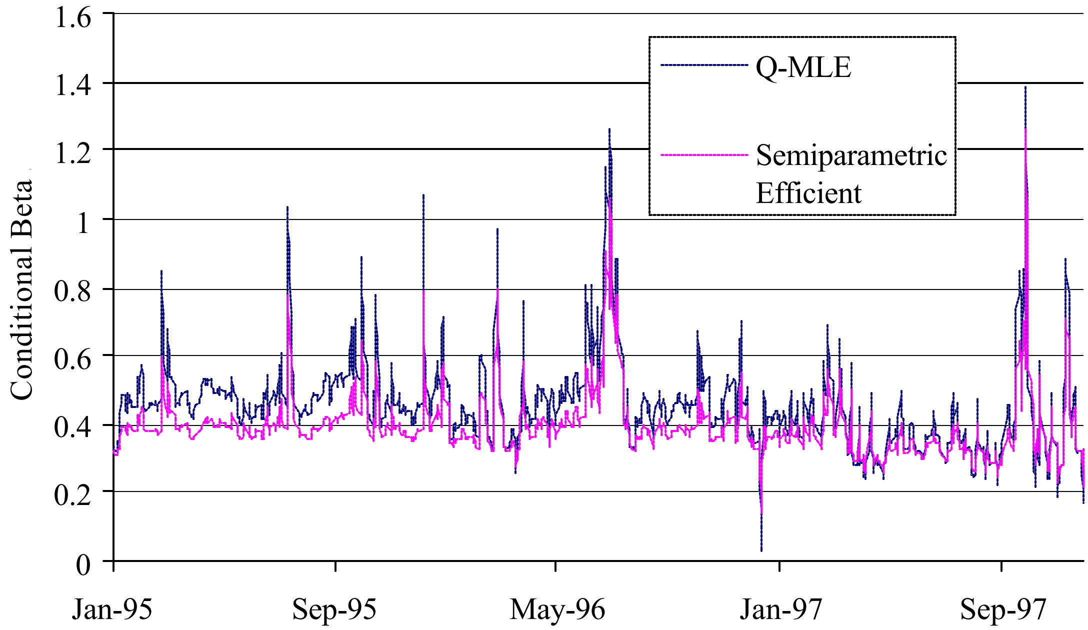

Figure 1: Conditional Beta for Size Portfolio 1

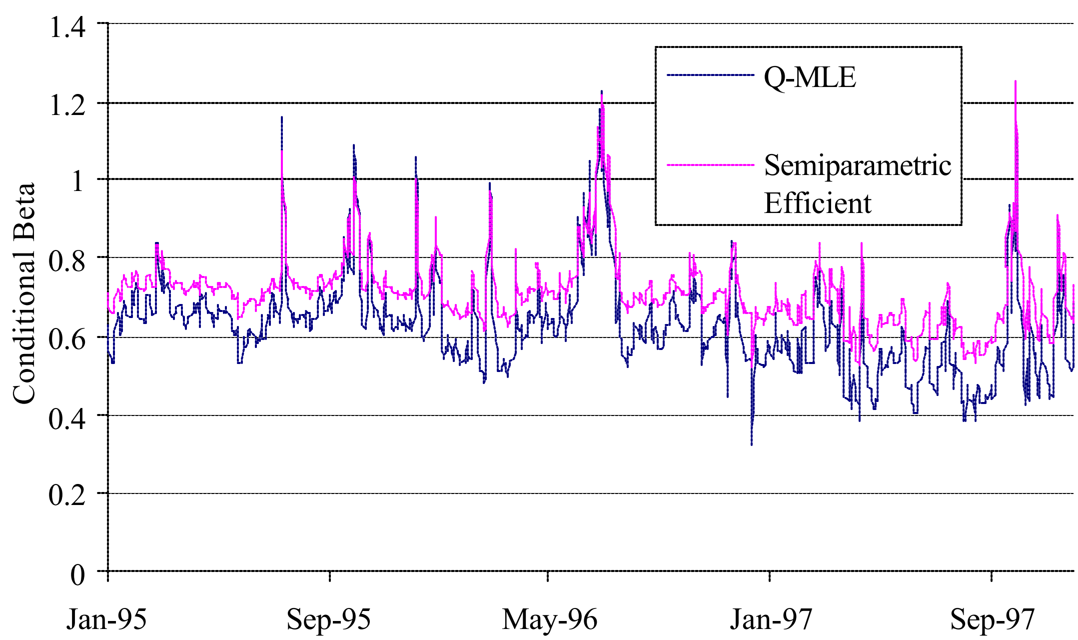

Figure 2: Conditional Beta for Size 2 Portfolio

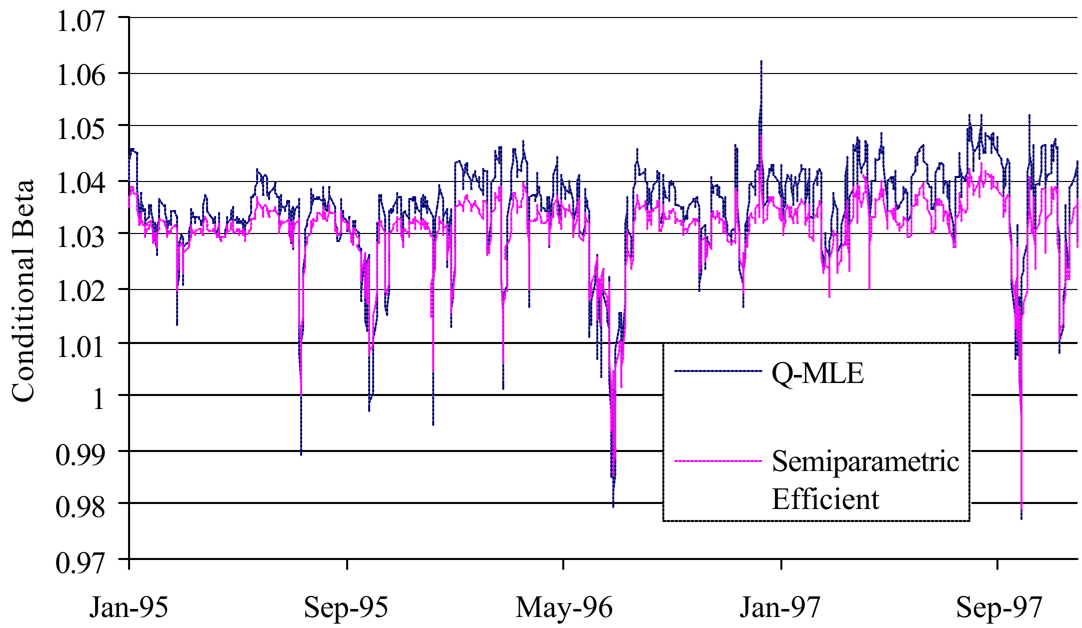

Figure 3: Conditional Beta for Size 3 Portfolio

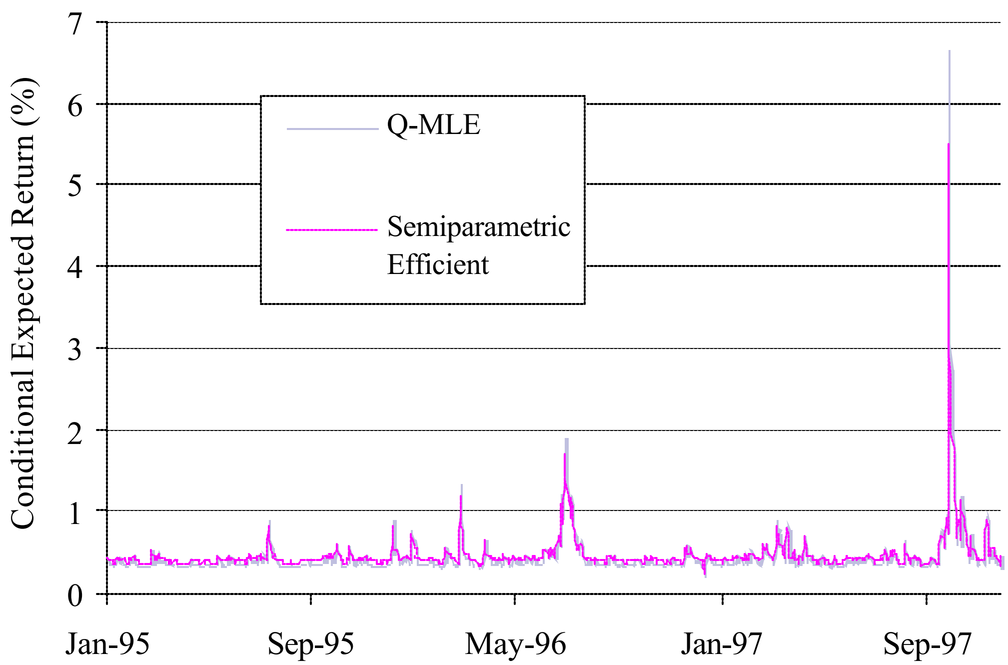

Figure 4: Conditonal Expected Return Size 1 Portfolio

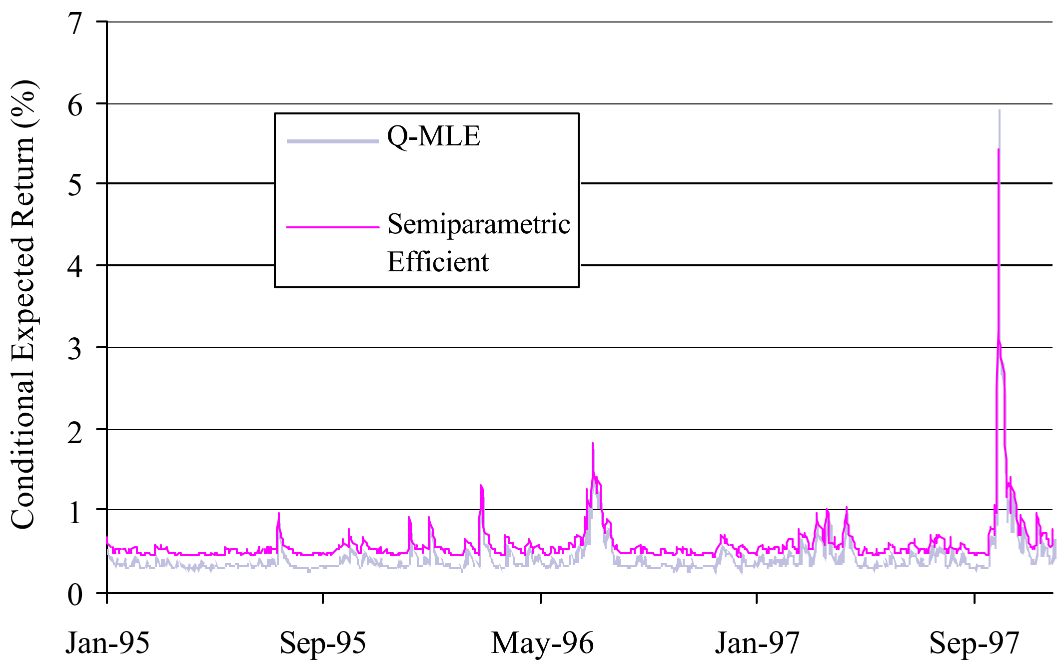

Figure 5: Conditional Expected Return Size 2 Portfolio

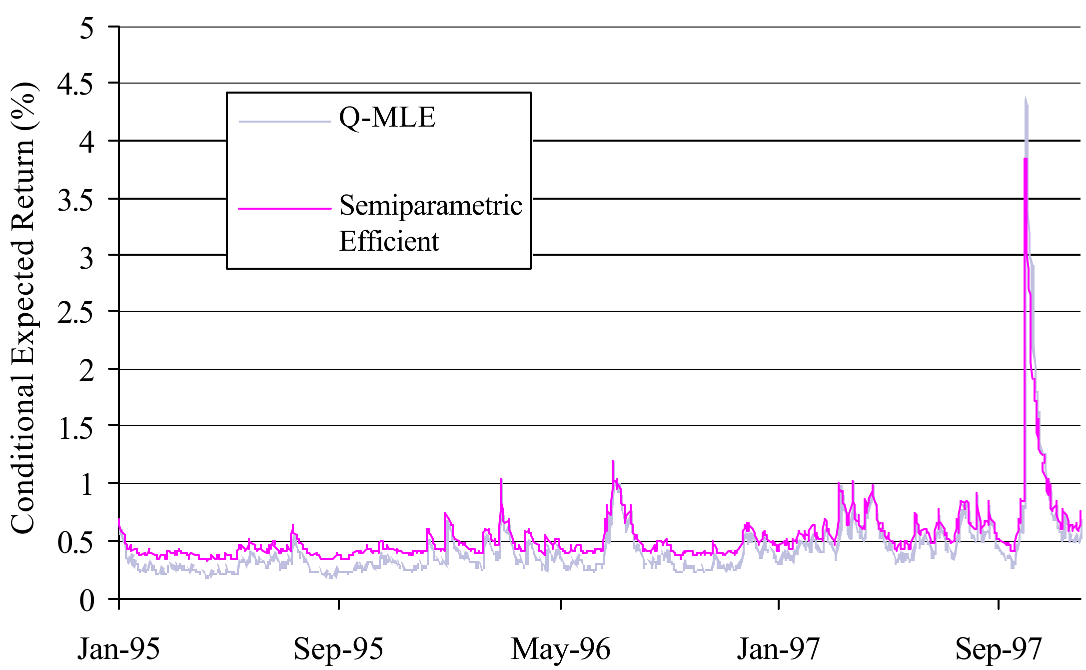

Figure 6: Conditional Expected Return Size 3 Portfolio