Evaluating the Specification of Covariance Models for Large Portfolios

Robert F. Engle[*]
Department of Finance
NYU Stern School of Business

Kevin Sheppard[†]
Department of Economics
and
Oxford-Man Institute of Quantitative Finance
University of Oxford

This version: June 10, 2008

Abstract

Many parameterizations have been introduced to model covariance dynamics. Yet estimating even moderately large models without excessive parameter restrictions, typically through parameter pooling, remains a formidable challenge. Most specifications suffer from a number of issues as the dimension increases, including difficulties ensuring uniformly positive-definite conditional covariances, numerous parameters needing to be simultaneously estimated or many non-linear constraints. This paper examines the classes of models capable of estimating conditional covariances when the cross-section dimension is large, including factor GARCH specifications, restricted vector GARCH processes, dynamic correlation GARCH models and proposes feasible asymmetric extensions for each. Large dimension problems also present difficulties for specification testing. Using up to 50 assets, we evaluate the performance of these specifications out-of-sample employing statistical and economically meaningful criteria against two benchmark models: an exponential smoother and a simple but empirically successful moving average factor model. We find there is value in modeling time-varying covariances, and tightly parameterized models generally perform better when evaluated with economically meaningful criteria.

JEL Classification Codes: C32, C51, C52, C53

Keywords: Covariance Forecasting, Dynamic Correlation, Mean-Variance Portfolios, Multivariate GARCH, Specification Testing, Tracking Portfolios, Volatility

[*]Mailing address: 44 West Fourth St, Ste 9-62, New York, NY 10012 email: rengle@stern.nyu.edu

[†]Corresponding Author, Mailing address: Department of Economics, University of Oxford, Manor Road Building, Manor Road, Oxford, OX1 3UQ, UK, email: kevin.sheppard@economics.ox.ac.uk.

All efforts have been made to ensure the paper is mistake free. Any remaining errors are the sole responsibility of the authors. Software used in this paper is available at `http://www.kevinsheppard.net`.

1 Introduction

While more than two decades have passed since the introduction of the initial multivariate GARCH specification (Kraft & Engle (1982) and Engle, Granger & Kraft (1984)), modeling the conditional covariance of even moderately high dimension systems remains a formidable challenge. Researchers have resigned to estimating many possibly inconsistent low dimension models or adopting implausible and untested parameter restrictions. This paper considers the case where the cross-section dimension of the data is large. Problems of this size can only be attempted using a small sample of the universe of multivariate GARCH parameterizations. These specifications generally make use of a closed-form estimator to reduce the dimensionality of the parameter space.

Initial attempts at modeling conditional covariances focused on providing rich parameterizations without considering scaling. These specifications include the *vec* (Bollerslev, Engle & Wooldridge 1988), BEKK (Engle & Kroner 1995), and latent factor models (Diebold & Nerlove (1989) and Engle, Ng & Rothschild (1990)). These all suffer from the curse of dimensionality; as the cross-section dimension increases, the number of parameters needing to be estimated simultaneously increases *at least* quadratically. The *vec* and the latent factor models are further complicated by the necessity of non-linear parameters constraints.

Recent research has focused on providing parameterizations or estimation strategies which overcome these scaling difficulties. Working in the framework of a scalar BEKK parameterization, Engle & Mezrich (1996) concentrated out the intercept term in the estimation of a scalar BEKK model. They proposed a two-step estimation strategy; the first estimates the long-run covariance of returns and the second, conditioning on this initial estimate and assuming covariance stationarity, fits the parameters governing the covariance dynamics. This reparameterization allows the covariance of very large portfolios to be examined, although the consequences of common parameters across variances and covariance has not been explored. To overcome the limitations of a pooled parameter, the FLEXM specification proposed an alternative estimation strategy for the diagonal *vec* (Ledoit, Santa-Clara & Wolf 2003). This strategy is implemented using only bivariate models, although an auxiliary step is required to ensure conditional covariances are positive definite.

Working in the factor framework, Alexander & Chibumba (1996) suggested using principal components to estimate unobserved heteroskedastic factors. A key assumption, that the factors are conditionally orthogonal, allows for estimation using only univariate volatility model. Engle (2002) extended the constant conditional correlation model (Bollerslev 1990) to allow for both multi-stage estimation and correlation dynamics. This approach combines the empirical success of the univariate GARCH specification for conditional variance models with a variance-targeting-like estimator for conditional correlations. These recent attempts at providing parameterizations suitable for modeling large time-varying covariances have all been found to perform well – at least for some data set. This paper will clarify the relationship between these models and explore their performance using a set of economically relevant criteria with applications in portfolio analysis and risk management.

While these specifications facilitate the estimation of ever larger models, none provide a direct method to capture conditional asymmetries. Kroner & Ng (1998) have introduced a model nesting four common models, the *vec*, BEKK, factor, and constant conditional correlation and allows for theoretically justified and empirically verified asymmetries in conditional covariances. While the asymmetric dynamic covariance model (ADC) specification provides a method to parameterize rich dynamics in variances and/or correlations, estimation of the ADC is nearly impossible with only 3 assets. To bridge the gap between the flexible parameterization of the ADC and models feasible with large portfolios, this paper proposes asymmetric extensions to the core set of models. We examine if allowing for conditional asymmetries, a phenomena widely documented in conditional *volatilities*, is a useful feature for a conditional covariance model.

The difficulties with high dimension conditional covariance models are not limited to estimation. Evaluation is complicated by the large number of covariances when the system dimension is large. Parameter estimation uncertainty, high-dimension parameter spaces and multi-stage estimation makes in-sample testing infeasible. To overcome this difficulty, we use of out-of-sample results to evaluate the specification of these models. We employ economically meaningful tests to address a simple question: Is there value in modeling the conditional covariance of large systems? The performance of alternative specifications is compared to two benchmark models: an EWMA parameterized by RiskMetrics and a sophisticated moving average covariance estimator parameterized in the spirit of Chan, Karceski & Lakonishok (1999) using the Fama-French factors (Fama & French 1993).

The paper is organized as follows. The second section briefly describes the models capable of estimating large time-varying covariance matrices, proposes asymmetric generalizations and explains the two benchmark models. Section three describes the data used in this study and presents some surprising in-sample results. Section four examines the performance of these models and section five concludes and outlines area of future research.

2 Specifications for conditional covariance

This section provides a brief description of the important features of the models examined in this paper. For a more comprehensive overview, including a more detailed discussion of the individual models, see Laurent, Bauwens & Rombouts (2006). Few GARCH specifications are appropriate when the number of assets is more than 5. These can be described as dynamic correlation GARCH models, including the constant conditional correlation model (Bollerslev 1990) and the dynamic conditional correlation model (Engle 2002), orthogonal GARCH (Alexander & Chibumba 1996), vector GARCH with variance targeting (Engle & Mezrich 1996), and a recently introduced estimation strategy for vector GARCH models (Ledoit et al. 2003).

All specification considered in this paper assume that a K by 1 vector *vec* of returns is conditionally normal

$$\mathbf{r}_t|\mathcal{F}_{t-1} \sim N(\mathbf{0}, \mathbf{H}_t) \tag{1}$$

with mean zero and time-varying covariance $\mathbf{H}_t$. The assumption of zero mean is fairly innocuous, since demeaned residuals could be used for estimation of the GARCH instead. A constant conditional mean, typically assumed 0, is widely held as statistically plausible for high-frequency returns. The assumption of conditional normality is not crucial, and when violated these estimators have a standard QML interpretation. The models will differ in how $\mathbf{H}_t$ evolves and the steps necessary to estimate model parameters.

2.1 Orthogonal GARCH

Orthogonal GARCH assumes that returns are generated from a set of m *conditionally* orthogonal factors.

$$\mathbf{r}_t = \mathbf{W}\mathbf{f}_t + \boldsymbol{\epsilon}_t \tag{2}$$

where $\mathbf{W}$ is a $K \times m$ matrix of factor loadings, $\mathbf{f}_t$ is a mean zero vector of factor innovations and $\boldsymbol{\epsilon}_t$ is a vector of residuals. Two assumptions are made to simplify estimation: the factors are conditionally orthogonal ($E[f_{it} f_{jt} | \mathcal{F}_{t-1}] = 0 \; \forall i \neq j$) and the factors are conditionally orthogonal to the errors ($E[f_{it} \epsilon_{jt} | \mathcal{F}_{t-1}] = 0$ $\forall i \in \{1, \ldots, m\} \; \forall j \in \{1, \ldots, K\}$). In addition, we assume the conditional covariance of the residual errors is constant ($E[\epsilon_{it} \epsilon_{jt} | \mathcal{F}_{t-1}] = \omega_{ij}^2 \; \forall i, j \in \{1, \ldots, K\}$). The conditional covariance is given by

$$\mathbf{H}_t = \mathbf{W}\mathbf{F}_t\mathbf{W}' + \boldsymbol{\Omega} \tag{3}$$

where $\mathbf{F}_t$ is the diagonal conditional covariance matrix of the factors and $\boldsymbol{\Omega}$ is the time-invariant matrix of residuals covariances. The conditional variance for the i^{th} factor is parameterized using a standard univariate GARCH specification (Bollerslev 1986)[1]:

$$F_{ii,t} = \omega_i + \alpha_i f_{i,t-1}^2 + \beta_i F_{ii,t-1} \tag{4}$$

One less well understood issue when implementing an OGARCH model is selecting the number of factors. We will address this issue by estimating two OGARCH specifications: one with 3 factors and one with K. The three factor specification is motivated by the empirical asset pricing literature where a limited number of factors, often three, has been found sufficient to explain the cross-sectional variation in asset returns (Fama & French (1993) and Carhart (1997)). We will denote results from the three factor OGARCH model using $\mathbf{O_3}$, and the K factor model using $\mathbf{O_K}$.

2.2 Correlation Multivariate GARCH

The family of correlation multivariate GARCH models all exploit the decomposition of conditional covariances into conditional standard deviations and conditional correlations. Specifically,

$$\mathbf{H}_t \equiv \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t \tag{5}$$

where $\mathbf{D}_t$ is a diagonal matrix of the conditional standard deviations and $\mathbf{R}_t$ is a possibly time-varying correlation matrix. Under certain conditions, this decomposition allows for separate estimation of the volatility and correlation parameters.

The CCC model can be consistently estimated in two steps (Bollerslev 1990). The first step specifies univariate GARCH processes for the conditional variance of each asset series, $r_{i,t}$:

$$h_{i,t} = \omega_i + \alpha_i r_{i,t-1}^2 + \beta h_{i,t-1}. \tag{6}$$

The CCC model assumes that $\mathbf{R}_t = \bar{\mathbf{R}} \; \forall t$. The constant conditional correlation is estimated using the standardized residuals ($\hat{\epsilon}_{i,t} = \frac{r_{i,t}}{\sqrt{\hat{h}_{i,t}}}$) and the usual correlation estimator:

$$\hat{\bar{R}}_{ij} = \frac{\sum_{t=1}^{T} \hat{\epsilon}_{i,t} \hat{\epsilon}_{j,t}}{\sqrt{\sum_{t=1}^{T} \hat{\epsilon}_{i,t}^2 \sum_{t=1}^{T} \hat{\epsilon}_{j,t}^2}}. \tag{7}$$

We will use **CCC** to denote results from the constant conditional correlation model.

The dynamic conditional correlation (DCC) specification extends the CCC to include non-constant correlations. $\mathbf{R}_t$ evolves as a simple ARMA-like process of the standardized residuals. Estimation of $\mathbf{R}_t$ is further

[1]While the specification used could include any number of lags of the squared factor returns, we only consider the simplest specification for each model, corresponding to a single lag of both right-hand-side variables.

simplified using a two-stage process where the intercept, containing all but 2 of the correlation evolution parameters, is concentrated out. The first-stage in estimating the conditional correlations uses equation 7 to estimate the average correlation. The second estimates the scalar parameters which govern the correlation dynamics.

Correlations in a DCC process evolve using two equations:

$$\mathbf{Q}_t = (1-\alpha-\beta)\bar{\mathbf{R}} + a\hat{\boldsymbol{\epsilon}}_{t-1}\hat{\boldsymbol{\epsilon}}'_{t-1} + b\mathbf{Q}_{t-1} \tag{8}$$

$$\mathbf{R}_t = \mathbf{Q}^*_t\mathbf{Q}_t\mathbf{Q}^*_t \tag{9}$$

where $\hat{\boldsymbol{\epsilon}}_t$ is the vector of standardized residuals, a and b are scalars, $\bar{\mathbf{R}} = E[\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}'_t]$ is the unconditional correlation of the standardized residuals,[2] and $\mathbf{Q}^*_t$ is a diagonal matrix composed of the inverse of the square root of the diagonal elements of $\mathbf{Q}_t$ to ensures $\mathbf{R}_t$ is a well defined correlation matrix. We will denote results from the dynamic conditional correlation specification using **DCC**.

2.3 Vector GARCH

The conditional covariance in a diagonal *vec* is given by

$$\mathbf{H}_t = \mathbf{C} + \mathbf{A}\odot\mathbf{r}_{t-1}\mathbf{r}'_{t-1} + \mathbf{B}\odot\mathbf{H}_{t-1} \tag{10}$$

where **A**, **B** and **C** are symmetric $K \times K$ matrices and $\odot$ denotes Hadamard (element-by-element) product. Two modifications of the diagonal *vec* have been proposed to facilitate applications to large portfolios while addressing positive definiteness of conditional covariances: variance targeting and the FLEXM estimation strategy.

2.3.1 Variance targeting scalar *vec*

If all series are restricted to have common scalar innovation and smoothing parameters, conditions for positive definiteness of conditional covariances simplify. The scalar *vec* specification is described by

$$\mathbf{H}_t = \mathbf{C} + \alpha\mathbf{r}_t\mathbf{r}'_t + \beta\mathbf{H}_{t-1} \tag{11}$$

where α and β are scalars. Taking unconditional expectations of 11,

$$\begin{aligned} E[\mathbf{H}_t] &= \mathbf{C} + \alpha\odot E[\mathbf{r}_{t-1}\mathbf{r}'_{t-1}] + \beta\odot E[\mathbf{H}_{t-1}] \Rightarrow \\ \bar{\mathbf{H}} &= \mathbf{C} + \alpha\odot\bar{\mathbf{H}} + \beta\odot\bar{\mathbf{H}} \Rightarrow \\ \mathbf{C} &= (\boldsymbol{\iota}\boldsymbol{\iota}' - \alpha - \beta)\odot\bar{\mathbf{H}} \end{aligned} \tag{12}$$

C can be concentrated out of the evolution of $\mathbf{H}_t$ and replaced with a consistent estimate, $(\boldsymbol{\iota}\boldsymbol{\iota}' - \alpha - \beta)\odot\bar{\mathbf{H}}$ where $\boldsymbol{\iota}$ is a $K \times 1$ vector of ones. Throughout this paper, we will denote this specification as **VT**.

[2]As a parameter of the DGP, $\bar{\mathbf{R}}$ is both the covariance of the standardized residuals as well as the correlation of the standardized residuals. However, in finite samples this will not be true and the covariance of the standardized residuals is used in place of the correlation of the standardized residuals.

2.3.2 FLEXM

Ding & Engle (2001) reparameterized the diagonal vector GARCH process with positive definite parameter matrices to make positive definiteness conditions tractable. After reparameterization, the original diagonal *vec* specification becomes

$$H_t = \tilde{\mathbf{C}}\tilde{\mathbf{C}}' + \tilde{\mathbf{A}}\tilde{\mathbf{A}}' \odot \mathbf{r}_{t-1}\mathbf{r}_{t-1}' + \tilde{\mathbf{B}}\tilde{\mathbf{B}}' \odot \mathbf{H}_{t-1} \tag{13}$$

where $\tilde{\mathbf{C}}$, $\tilde{\mathbf{A}}$, and $\tilde{\mathbf{B}}$ are the Cholesky factors of **C**, **A** and **B** respectively (eq. 10).

Ledoit et al. (2003) have adapted the Ding and Engle representation to allow the estimation of a diagonal *vec* using only univariate and bivariate specifications. Their estimation method for the diagonal *vec* involves first fitting univariate GARCH models to each series' volatility and then estimating bivariate GARCH models for each pair's covariance, conditioning on the initial parameters estimates. In essence, $\frac{K^2+K}{2}$ models are separately estimated. The first stage involves fitting

$$h_{ii,t} = c_{ii} + a_{ii} r_{i,t-1}^2 + b_{ii} h_{i,t-1} \tag{14}$$

for each asset. The second stage then fits the remaining $\frac{K^2-K}{2}$ covariance series using bivariate models conditioning on the two previously fit variance series, $\{\hat{h}_{ii,t}\}$ and $\{\hat{h}_{jj,t}\}$, and the initial parameter estimates using the specification

$$h_{ij,t} = c_{ij} + a_{ij} r_{i,t-1} r_{j,t-1} + b_{ij} h_{ij,t-1} \tag{15}$$

by maximizing the bivariate normal log-likelihood.

Once the $\frac{K^2-K}{2}$ models have been fit and $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$ and $\hat{\mathbf{D}} = \hat{\mathbf{C}} \oslash (\boldsymbol{\iota}\boldsymbol{\iota}' - \hat{\mathbf{B}})$ have been estimated, it is necessary to ensure that each matrix is positive definite. Positive definiteness is ensured by minimizing the Frobenius norm between the estimated parameter matrices and the set of positive definite matrices with the same diagonal elements. For both $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$, positive definiteness is ensured by

$$\begin{aligned} &\min_{\ddot{\mathbf{A}}} \parallel \hat{\mathbf{A}} - \ddot{\mathbf{A}} \parallel \qquad \text{subject to } \hat{a}_{ii} = \ddot{a}_{ii} \text{ and } \ddot{\mathbf{A}} \text{ is positive semi-definite} \\ &\min_{\ddot{\mathbf{B}}} \parallel \hat{\mathbf{B}} - \ddot{\mathbf{B}} \parallel \qquad \text{subject to } \hat{b}_{ii} = \ddot{b}_{ii} \text{ and } \ddot{\mathbf{B}} \text{ is positive semi-definite} \end{aligned} \tag{16}$$

where $\parallel \cdot \parallel$ denotes the Frobenius norm. Once $\ddot{\mathbf{A}}$ and $\ddot{\mathbf{B}}$ have been computed, it is necessary to ensure that $\hat{\mathbf{D}} = \hat{\mathbf{C}} \oslash (\boldsymbol{\iota}\boldsymbol{\iota}' - \ddot{\mathbf{B}})$ is positive semi-definite,

$$\min_{\ddot{\mathbf{D}}} \parallel \hat{\mathbf{D}} - \ddot{\mathbf{D}} \parallel \qquad \text{subject to } \hat{d}_{ii} = \ddot{d}_{ii} \text{ and } \ddot{\mathbf{D}} \text{ is positive semi-definite.} \tag{17}$$

$\ddot{\mathbf{C}}$ is finally computed as $\ddot{\mathbf{D}} \odot (\boldsymbol{\iota}\boldsymbol{\iota}' - \ddot{\mathbf{B}})$ and the fit covariances are generated by

$$\hat{\mathbf{H}}_t = \ddot{\mathbf{C}} + \ddot{\mathbf{A}} \odot \mathbf{r}_{t-1}\mathbf{r}_{t-1}' + \ddot{\mathbf{B}} \odot \hat{\mathbf{H}}_{t-1}. \tag{18}$$

We will use **FLEXM** to denote the multi-step estimation of a diagonal *vec*. While implementation of this estimation strategy is straight forward, the use of the Frobenius is not without caveats. When the estimated parameter matrices are not positive definite, the use of the Frobenius norm makes interpretation of the estimated FLEXM somewhat difficult. Ledoit et al. (2003) point out that non-positive definiteness of the initial estimates $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$ and $\hat{\mathbf{D}}$ is a small sample property and that asymptotically $\ddot{\mathbf{A}} = \hat{\mathbf{A}}$, $\ddot{\mathbf{B}} = \hat{\mathbf{B}}$, and $\ddot{\mathbf{D}} = \hat{\mathbf{D}}$. Unfortunately, for moderately large models, the estimated parameter matrices are rarely positive definite in typical samples sizes. The use of the Frobenius norm further complicates matters as the order of the assets changes the final estimates in small sample using the implementation advocated in Ledoit et al. (2003). Also, some difficulty arises when trying to treat the FLEXM as an M estimator since it does not always possess some standard properties. Specifically, the VT specification, which is nested by the FLEXM, can produce larger log-likelihoods despite having far fewer parameters.

2.4 Feasible Asymmetric Extensions

The models presented thus far provide a broad overview of those capable of estimating conditional covariances of large portfolios. However, none can capture one feature common to many financial time series: conditional asymmetries. Asymmetric increases in equity volatility have been theoretically justified through the leverage effect (Christie 1982) and volatility feedback (Campbell & Hentschel 1992) and have been found in countless empirical studies (see Glosten, Jagannathan & Runkle (1993), Zakoian (1994) and Kroner & Ng (1998), *inter alia.*). All of the models considered in this paper have feasible asymmetric extensions which are only moderately more difficult to implement.

2.4.1 Asymmetric OGARCH

An asymmetric orthogonal GARCH model can be constructed using asymmetric univariate GARCH specifications, such as GJR-GARCH (Glosten et al. 1993), to model the conditional variances of the factors. When using the GJR parameterization for the conditional factor variances, the variance dynamics (eq. 4) become

$$F_{ii,t} = \omega_i + \alpha_i f_{i,t-1}^2 + \gamma_i I_{[f_{i,t-1}<0]} f_{i,t-1}^2 + \beta_i F_{ii,t-1} \tag{19}$$

where $I_{[f_{i,t-1}<0]}$ is a an indicator variable which takes the value of 1 for negative factor returns.[3] Unfortunately, interpretation of asymmetries is difficult since the transformation of the original return to factors obfuscates the nature of the asymmetry. If a factor loading for a particular asset was negative, a negative shock to that asset would actually increase the value of the corresponding factor and could lower volatility. We will again examine the performance of both 3 factor and a K factor model referring to these two specifications as $\mathbf{aO_3}$ and $\mathbf{aO_K}$ respectively.

2.4.2 Asymmetric Constant Correlation GARCH

An asymmetric CCC model can be constructed analogously to the asymmetric OGARCH model by parameterizing the conditional variances using a GJR specification. The CCC variance dynamics (eq. 6) are modified to

$$h_{i,t} = \omega_i + \alpha_i r_{i,t-1}^2 + \gamma_i I_{[r_{i,t-1}<0]} r_{i,t-1}^2 + \beta_i h_{i,t-1} \tag{20}$$

where $I_{[r_{i,t-1}<0]}$ is an indictor for negative returns. We will refer to the asymmetric constant conditional correlation model as **aCCC**.

[3]Normally, in GJR specification $\alpha_i \geq 0$ and $\gamma_i \geq 0$ are assumed. This allows for direct examination of asymmetries consistent with theoretical justifications: that conditional variance subsequent to a negative shock is higher than subsequent to a positive shock of the same magnitude. However, the sign indeterminacy of principal component estimated factors makes this deviation from usual practices necessary.

2.4.3 Asymmetric variance targeting scalar *vec*

The evolution of an asymmetric scalar *vec* model without variance targeting is given by

$$\mathbf{H}_t = \mathbf{C} + \alpha \mathbf{r}_{t-1}\mathbf{r}'_{t-1} + \gamma \boldsymbol{\eta}_{t-1}\boldsymbol{\eta}'_{t-1} + \beta \mathbf{H}_{t-1} \tag{21}$$

where $\boldsymbol{\eta}_t = I_{[\mathbf{r}_t<0]} \odot \mathbf{r}_t$ is a vector and γ is a scalar. $\boldsymbol{\eta}_t\boldsymbol{\eta}'_t$, a quadratic form, will be positive semi-definite and positive definiteness of $\mathbf{H}_t$ can be assured if, in addition to the original parameter restrictions, $\gamma > 0$. Taking unconditional expectations and rearranging, we find

$$\mathbf{C} = (\boldsymbol{\iota}\boldsymbol{\iota}' - \alpha - \beta) \odot E[\mathbf{r}_t\mathbf{r}'_t] - \gamma E[\boldsymbol{\eta}_t\boldsymbol{\eta}'_t] = (\boldsymbol{\iota}\boldsymbol{\iota}' - \alpha - \beta) \odot \bar{\mathbf{H}} - \gamma \bar{\mathbf{N}}. \tag{22}$$

In estimation, $\bar{\mathbf{N}}$ can be replaced by its natural sample analog of $T^{-1}\sum_{t=1}^{T}\hat{\boldsymbol{\eta}}_t\hat{\boldsymbol{\eta}}'_t$. Unlike the symmetric version, conditions for positive definiteness of $\mathbf{H}_t$ are not obvious. Cappiello, Engle & Sheppard (2006) derive a restriction on α, β and γ sufficient to ensure positive definiteness of conditional covariances:

$$\alpha + \beta + \delta\gamma < 1 \tag{23}$$

where δ is the maximum eigenvalue of $\bar{\mathbf{H}}^{-\frac{1}{2}}\bar{\mathbf{N}}\bar{\mathbf{H}}^{-\frac{1}{2}}$. Results from the asymmetric variance targeting model will be denoted by **aVT**.

2.4.4 Asymmetric DCC GARCH

There are two opportunities to introduce asymmetries into a DCC specification. The first, identical to the asymmetries in the aCCC specification, introduces asymmetries by modifying the univariate volatility specifications. We will again specify a GJR-GARCH parameterization for each variance process. A second form of asymmetry can be introduced into the correlation dynamics (Cappiello et al. 2006). We define $\boldsymbol{\varsigma}_{t-1}$ to be the standardized residuals multiplied by a vector of indicators which will take the value of 1 if the corresponding residual was negative. That is $\boldsymbol{\varsigma}_t = I_{[\boldsymbol{\epsilon}_t<0]} \odot \boldsymbol{\epsilon}_t = I_{[\mathbf{r}_t<0]} \odot \frac{\mathbf{r}_t}{\sqrt{\mathbf{h}_t}}$. The dynamics of the correlation in an asymmetric DCC process are described using

$$\mathbf{Q}_t = (1 - \alpha - \beta)\bar{\mathbf{R}} - \gamma\bar{\mathbf{S}} + a\hat{\boldsymbol{\epsilon}}_{t-1}\hat{\boldsymbol{\epsilon}}'_{t-1} + g\hat{\boldsymbol{\varsigma}}_{t-1}\hat{\boldsymbol{\varsigma}}'_{t-1} + b\mathbf{Q}_{t-1} \tag{24}$$

and

$$\mathbf{R}_t = \mathbf{Q}^*_t\mathbf{Q}_t\mathbf{Q}^*_t \tag{25}$$

where $\bar{\mathbf{S}} = E[\boldsymbol{\varsigma}_{t-1}\boldsymbol{\varsigma}'_{t-1}]$ is the unconditional expectation of the outer-product of the standardized residuals when less than zero. Conditions for positive definiteness of the conditional correlations are identical to the asymmetric variance targeting model. Specifically,

$$a + b + \delta g < 1 \tag{26}$$

where δ is the maximum eigenvalue of $\bar{\mathbf{R}}^{-\frac{1}{2}}\bar{\mathbf{S}}\bar{\mathbf{R}}^{-\frac{1}{2}}$ and a, g and b are all greater than or equal to zero. We will denote results from the asymmetric dynamic conditional correlation model using **aDCC**.

2.4.5 Asymmetric FLEXM

An asymmetric FLEXM can be constructed as an analogue to an asymmetric diagonal *vec* using an estimation strategy identical to that of the symmetric FLEXM. The covariance dynamics of an asymmetric diagonal *vec* are given by

$$\mathbf{H}_t = \mathbf{C} + \mathbf{A} \odot \mathbf{r}_{t-1}\mathbf{r}'_{t-1} + \mathbf{G} \odot \boldsymbol{\eta}_{t-1}\boldsymbol{\eta}'_{t-1} + \mathbf{B} \odot \mathbf{H}_{t-1} \tag{27}$$

where the four parameter matrices, **C**, **A**, **G** and **B** are all positive semi-definite. This model will produce positive semi-definite conditional covariances by construction. The asymmetric FLEXM equivalent to this model would specify GJR-GARCH processes for the variance dynamics. Thus,

$$h_{ii,t} = c_{ii} + a_{ii} r^2_{i,t-1} + g_{ii} I_{[r_{i,t-1}<0]} r^2_{i,t-1} + \beta h_{ii,t-1} \tag{28}$$

where $c_{ii} > 0$, $a_{ii} \geq 0$, $g_{ii} \geq 0$ and $b_{ii} \geq 0$. Conditional asymmetries in covariances are parameterized using the equivalent

$$h_{ij,t} = c_{ij} + a_{ij} r_{i,t-1} r_{j,t-1} + g_{ij} I_{[r_{i,t-1}<0, r_{j,t-1}<0]} r_{i,t-1} r_{j,t-1} + \beta h_{ij,t-1} \tag{29}$$

where $I_{[r_{i,t-1}<0, r_{j,t-1}]<0]}$ takes the value of one when both $r_{i,t-1}$ and $r_{j,t-1}$ are less than zero. In each bivariate specification, the original three constraints, along with a new constraint, $0 \leq g_{ij} \leq \sqrt{g_{ii} g_{jj}}$, would be imposed. Again, these are necessary to ensure the four parameter matrices are positive semi-definite and are sufficient to ensure all bivariate covariance matrices are positive semi-definite. Once the $\frac{K^2+K}{2}$ models had been estimated, positive definiteness of the $\hat{\mathbf{A}}$, $\hat{\mathbf{G}}$, $\hat{\mathbf{B}}$ and $\hat{\mathbf{D}} = \hat{\mathbf{C}} \oslash (\boldsymbol{\iota}\boldsymbol{\iota}' - \hat{\mathbf{B}})$ would be ensured using the Frobenius norm procedure. Finally, letting¨denote positive definite versions of the original estimates, the fit conditional covariances are given by

$$\hat{\mathbf{H}}_t = \ddot{\mathbf{C}} + \ddot{\mathbf{A}} \odot \mathbf{r}_{t-1}\mathbf{r}'_{t-1} + \ddot{\mathbf{G}} \odot \boldsymbol{\eta}_{t-1}\boldsymbol{\eta}'_{t-1} + \ddot{\mathbf{B}}\hat{\mathbf{H}}_{t-1}. \tag{30}$$

We will use the notation **aFLEXM** when describing results from this model.

2.5 Model relationships and discussion

While the specifications have been presented independently, there are a number of noteworthy relationships among them. Obvious ones include that the DCC specification nests the CCC specification, and the FLEXM specification nests the VT specification and that all models nest a constant conditional specification. Additionally, all of the proposed asymmetric extensions nest their symmetric counterparts as special cases. There are also some less obvious relationships among the models. The CCC, DCC and FLEXM will produce *identical* variance dynamics, as each specify a standard GARCH process for the assets' conditional variance. Thus, any differences in these models can be reduced to alternative specifications for conditional correlation. When considering the asymmetric versions of these three models, this is no longer the case. In the specification for the conditional variance of asset i in these models,

$$h_{i,t} = c_i + a_i r^2_{i,t-1} + g_i I_{[r_{i,t-1}<0]} r^2_{i,t-1} + \beta h_{i,t-1} \tag{31}$$

there is a subtle but important difference in the necessary restrictions on the parameters. In both the aDCC and the aCCC, the required restrictions are $a_i \geq 0$ and $a_i + g_i \geq 0$. However, in the aFLEXM specification, the second restriction must be replaced by a restriction that $g_i \geq 0$ because the parameter matrix corresponding to the asymmetric shock must be positive definite, ruling out negative values for g_i. In most applications, g_i will be greater than zero since most financial time series exhibit a positive asymmetry with respect to negative shocks and the numerical values of the fit conditional variances will be identical across these three specifications.

Of the six GARCH models discussed, orthogonal GARCH is the simplest to estimate when the number of factors is fixed. Estimating principal components and factor loadings is trivial, requiring only an eigenvalue decomposition of the unconditional covariance matrix and fitting a fixed number of univariate GARCH models. A CCC specification and an orthogonal GARCH with K factors are similarly difficult to estimate. Both require the estimation of K univariate GARCH specifications and employ a computationally simple closed-form estimator for the remaining parameters.

The remaining three models are considerably more difficult to implement.[4] Superficially, it would seem that estimation challenge presented by the FLEXM strategy would be similar to that of the CCC specification as the dimension of the largest estimation problem does not depend on K. However, the number of sub-specification ($\frac{K^2+K}{2}$) needing to be estimated grows rapidly in the FLEXM model. In addition, the minimizing the Frobenius norm can be time consuming when the number of assets is large.

The scalar VT model may also seem superficially easy to estimate. However, two issues complicate the estimation of this model. First, with only two parameters common across all assets, the likelihood is very steep near the optimum. Numerical optimization can require many iterations to reach a stopping condition. Second, estimation requires inverting the $K \times K$ covariance matrix T times for each likelihood evaluation, a costly operation when K is large. Finally, the DCC model combines the optimization effort of the CCC with respect to the univariate specifications with a variance targeting two parameter correlation estimator, similar to the VT model. However, the estimation of the parameters which govern the dynamics in the DCC is generally easier than the estimation of these two parameters in the VT as the likelihood in the DCC is not as steep near the optimum. Of the latter three models discussed, if the order of K is moderate ($K < 25$) the order of difficulty is generally VT, FLEXM, and DCC. However, as K grows ($K > 35$), the order of difficulty in estimation changes to DCC, VT, and finally FLEXM. Finally, the introduction of asymmetries into the conditional covariance specification does not meaningfully increase the estimation difficulty of any of these specifications.

2.6 Benchmark models

In addition to the six GARCH specifications and their asymmetric extensions previously discussed, two benchmark models will be employed to evaluate the performance of these alternative specifications. The first is a widely used exponential smoother parameterized by RiskMetrics. The RiskMetrics smoother is a restricted version of the VT specification where $\alpha + \beta = 1$. The intercept is absent and the covariance dynamics are given by

$$\mathbf{H}_t = (1 - \lambda)\mathbf{r}_{t-1}\mathbf{r}'_{t-1} + \lambda\mathbf{H}_{t-1} \tag{32}$$

where λ is a predetermined parameter not estimated on the data. Using a pre-specified smoothing parameter guarantees that the constructing a RiskMetrics-fit covariance is trivial irrespective of the number of assets. For daily data, RiskMetrics recommends using $\lambda = .94$. While the choice of λ was primarily mo-

[4]While some specifications are *more* difficult to fit, estimation of all 6 models using the full sample (2226 observations) with requires only 15 minutes.

tivated by applications in risk management, the RiskMetrics specification is still a widely used covariance forecast serving multiple purposes. We will denote results from the RiskMetrics specification using **RM**.

The second benchmark model, proposed in Chan et al. (1999), is a sophisticated moving average and has previously demonstrated good performance in forecasting covariances for applications in portfolio optimization. The model is a strict factor model where the factor are chosen from those found to price assets. Chan et. al. argue that the factor structure provides meaningful improvements over unrestricted covariance matrices although they do not consider GARCH models. The choice of a strict factor model ensures that covariance forecasts will be positive definite irrespective of the number of assets included or the length of the sample. The conditional covariance is given by:

$$\mathbf{H}_t = \mathbf{B}_t \mathbf{F}_t \mathbf{B}_t' + \mathbf{\Omega}_t. \quad (33)$$

$\mathbf{F}_t$ is a 252 day moving average covariance of the factors

$$\mathbf{F}_t = \frac{1}{252} \sum_{i=t-252}^{t-1} \mathbf{f}_i \mathbf{f}_i' \quad (34)$$

where $\mathbf{f}_i$ correspond to observable factors. $\mathbf{B}_t$ is the K by number of factors matrix of 252-day rolling window regression coefficients of the returns on the observable factors. Finally, $\mathbf{\Omega}_t$ is a diagonal matrix of residual variances from the rolling 252-day regression of returns on the factors. We will employ a 3 factor version of this estimator using the factors specified in Fama & French (1993).

3 Data and in-sample characteristics

Returns from 50 S&P 500 sector indices will be used to evaluate the performance of these specifications. The choice of assets was motivated by two concerns: finding an economically interesting set of assets and avoiding unnecessary complications which often arise when examining covariances. Cavaglia, Brightman & Aked (2000) show that as financial markets around the globe have become increasingly liberated, the benefits of cross-country diversification have decreased. Further, they argue that the benefits to diversification are higher when considering multiple sectors within one country than through international diversification. Sector indices also avoid synchronization issues which occur in international equity returns.

The S&P 500 sector indices are well diversified value-weighted portfolios for various industry groups. Returns were computed as log-differences using close-to-close prices. Sectors were included based on average market capitalization during the sample period, and sectors which had any unconditional correlation in excess of 80% were excluded.[5] After deletion of sectors which were too correlated, 62 sectors remained. Appendix A contains the list of the 50 indices selected for this paper. Returns were available from January 3, 1995 until October 31, 2003, a total of 2226 observations after removing days the market was closed and were retrieved from DataStream. In addition, the Fama-French factors, VWM, SMB and HML, used in the factor benchmark portfolio were taken from the CRSP database via Kenneth French's website. In fitting the models, the returns were not transformed in any way from their raw values.

[5] The actual procedure for removing series involved iteratively deleting the sector with the maximum correlation until all correlations were under 0.8. Some sectors were extremely correlated ($\rho > .95$) in the original data as the S&P 500 sector industries can overlap. For example, the S&P Leisure Equipment Index and the S&P Leisure Equipment and Products index share common components.

Table 1 contains a brief summary of the data. The returns on these portfolios are typical of equity data although less volatile than individual issues. The annualized mean returns ranged from -6% per year to 24% while annualized standard deviation of returns ranged from a low 17% to a high of 51%. Skewness was generally negative, and kurtosis was uniformly larger than that which would be implied by a normal distribution.

While the paper won't report any model parameters, there were some interesting full sample results. In all cases, the asymmetric models were strictly preferred, based on log-likelihoods, to their symmetric counterparts, although p-value calculation is infeasible with these multi-stage estimators. The DCC innovation parameter was greater than zero, indicating that the data likely exhibited dynamics in correlation and the aDCC model uniformly selected an asymmetry parameter in the correlation evolution equation different from zero. The VT model generally choose a small innovation parameter and high persistence (typical $\alpha = .015$, $\alpha + \beta = .985$) while its asymmetric extension chose an asymmetry parameter different from zero and larger than the symmetric innovation term. The parameters estimated in either variance targeting model differ substantially from those estimated through univariate GARCH specifications where the typical innovation parameter was 0.07 and the typical smoothing parameter was 0.90. We interpret this as evidence that the cross-asset parameter restrictions implicit in the VT models are binding, although it may be a small sample property of VT model parameters.

Full-sample log-likelihoods are presented in table 2. All asymmetric extensions produced higher log-likelihoods. Moreover, if standard likelihood-ratio testing were applicable, the asymmetric parameterizations would all be statistically preferred to their symmetric counterparts. While this is not the case, it does provide limited evidence that asymmetries are a salient feature of conditional covariances and that specifications should attempt to capture them. The DCC models, both symmetric and asymmetric, produced better fits than their constant counterparts while the K-factor orthogonal models were strongly preferred to the 3-factor versions. The FLEXM specifications, while nesting the variance targeting models, produced *lower* full- sample log-likelihoods. This is somewhat troublesome as standard theory indicates the log-likelihood should be larger (with 50 assets, the FLEXM has 2548 more parameters). The likely cause of this was the necessity of the Frobenius norm on the initial estimates.

4 Specification Evaluation

Recent advances have provided a number of methods to evaluate the in-sample fit of covariance specifications. Tests exist to evaluate both absolute performance, such as the robust conditional moment test (Wooldridge (1990) and Wooldridge (1991)), and relative performance, such as the Rivers and Vuong test (Vuong (1989) and Rivers & Vuong (2002)). Unfortunately, these tests require controlling for parameter estimation uncertainty and are generally infeasible in high dimension models where the number of parameters is larger than T.

We make use of out-of-sample performance to evaluate the alternative specifications. Beginning at observation 1000, and continuing until the end of the sample, the models were recursively estimated using an expanding window and one-step ahead forecasts were constructed. The first forecast conditional covariance matrix corresponds to returns at observation 1001 (December 15, 1998) while the last corresponds to returns at observation 2227 (November 1, 2003).

Before specifying a test, it is informative to examine the properties of optimal forecast covariances. Specifically, an optimal forecast should have the property that

$$E[\mathbf{r}_t\mathbf{r}_t'|\mathcal{F}_{t-1}] = \mathbf{H}_{t|t-1} \tag{35}$$

and if $\mathbf{H}_{t|t-1}$ is always invertible,

$$E\left[\mathbf{H}_{t|t-1}^{-1/2}\mathbf{r}_t\mathbf{r}_t'\mathbf{H}_{t|t-1}^{-1/2}|\mathcal{F}_{t-1}\right] = \mathbf{I}_K. \tag{36}$$

Simply put, the outer product of forecast standardized residuals should have mean $\mathbf{I}_K$ and be unpredictable with anything in the time $t-1$ information set, including lagged returns, the forecast itself or any function of these such as commonly used sign indicators. Additionally, optimal properties of forecast residuals extend to portfolio variances using either pre-specified weights or weights estimated from the forecast. Let $\mathbf{w}_t$ be a K by 1 vector of weights composed of a deterministic term plus possibly a $o_p(1)$ term.[6] If a forecast is optimal, it must be the case that

$$E\left[\frac{(\mathbf{w}_t'\mathbf{r_t})^2}{\mathbf{w}_t'\mathbf{H}_{t|t-1}\mathbf{w}_t}|\mathcal{F}_{t-1}\right] = 1. \tag{37}$$

This simple condition for forecast optimality provides an intuitive method to examine alternative specifications. Squared portfolio returns can be standardized by their expected variance and a regression based test can be used to examine whether the conditions of an optimal forecast are true. Define sr_t to be the standardized residual. A hypothesis test, that the model is correctly specified, can be conducted by running the regression

$$sr_t = \alpha + \boldsymbol{\beta}\mathbf{x}_{t-1} + \nu_t \tag{38}$$

where $\mathbf{x}_{t-1}$ is a set of time $t-1$ available instruments. The null hypothesis, $H_0 : \alpha = 1, \boldsymbol{\beta} = \mathbf{0}$ can be tested using standard asymptotic theory, although White (1980) standard errors should be used to control for possible heteroskedasticity in the standardized residuals.

4.1 Static Portfolio Performance

Performance was evaluated by examining the properties of standardized residuals of portfolios formed using three static weight vectors: a single asset portfolio, a two asset long-long portfolio and a two asset long-short portfolio. The single asset portfolios will simply test the ability of the specifications to correctly forecast the conditional variances, ignoring conditional correlations. Both the long-long and the long-short portfolios will extend the single asset results to examine how well the specifications forecast covariance. The long-short is particularly difficult as most fit conditional covariances are positive.

We chose two instruments for the regression analysis: one to detect static misspecification and one to detect dynamic misspecification. The first, a constant, captures failure of the specification to correctly forecast unconditional portfolio variance. The second, a lag of the standardized squared portfolio return, captures neglected dynamics. Beginning with observation 1001, the portfolio variance predicted using the one-step ahead forecast and the square of the realized portfolio return were computed. Forecast standardized residuals were formed as

[6] This assumption will allow the use of either purely predetermined weights, such as an equally weighted portfolio, or weights determined from the forecast covariance matrix, such as minimum variance portfolio weights.

$$sr_t = \frac{(r_t^P)^2}{\mathbf{w}_t' \mathbf{H}_{t|t-1} \mathbf{w}_t} = \frac{(r_t^P)^2}{h_{t|t-1}^P}. \tag{39}$$

A joint test that the forecast is unconditionally correct and not autocorrelated was conducted using a regression. The null hypothesis, $H_0 : \mu = 1, \rho = 0$, can be tested by regressing standardized residuals on a constant and one lag:

$$sr_t = \mu + \rho sr_{t-1} + \nu_t. \tag{40}$$

Because rejection of this null can occur due to either cause, we also examine the two elements of the composite null separately.[7] White (1980) heteroskedasticity robust standard errors were used to control for possible standardized residual heteroskedasticity.

Table 3 contains the results of these three tests for the single asset portfolios. Rather than form a portmanteau statistic, we present the percentage of tests rejecting the null of correct specification at 5%.[8] These results are not promising. The best performance appears to have come from the simple exponential smoother parameterized by RiskMetrics. However, these results are misleading. The RiskMetrics forecast is very volatile due to the large parameter on the innovation term (.06) and the lack of an intercept. This volatility is embedded in the forecast standardized residuals, resulting in low power despite obvious misspecification. Columns 4 and 5 contain the grand average of the standardized residual means and the grand average of the standardized residual autocorrelations. The RM model has larger values in both of these columns than many other specifications and only fails to reject due to model induced variance in the standardized residuals.

Every other specification rejected the joint hypothesis more than a quarter of the time using a 5% test. The symmetric GARCH specification, common to the CCC, DCC and FLEXM, rejected 13 out of 50 series for correct specification and performed slightly better than the GJR-GARCH specification common to the corresponding asymmetric parameterizations. The others, including the VT and orthogonal variants as well as the CKL benchmark model, perform considerably worse. The K factor versions of the orthogonal GARCH perform better than the 3 factor, but their performance, even in capturing the dynamics, was rejected far too often. Interestingly, the CKL model has a close to size rejection rate for the unconditional mean but has obvious deficiencies in modeling the dynamics.

Figure 9 contains a box plot of the standardized residual variance and autocorrelations. The box contains the median and quartiles while the whiskers extend to the minimum and maximum value. The deficiencies of the RM model are more obvious here. The RM model *never* produces a standardized residual series with mean less than 1 and the autocorrelations are generally positive. It is difficult to discern any difference between the symmetric GARCH and the GJR-GARCH process standardized residuals. The average standardized residual variance is slightly better for the GARCH processes while the asymmetric specifications have slightly more symmetry evident in their autocorrelations. The remaining 6 parameterizations perform poorly, particularly in capturing the dynamics, although the CKL benchmark has the best mean coverage coupled with the largest autocorrelations.

Table 4 contains results of the specification tests on the two-asset portfolios. The results for the standardized residuals based on the 1225 long-long portfolios are similar to those of the single asset portfolios.

[7]Specifically, we assume that the portion of the composite *not* being tested is in fact correct. The test for $H_0 : \rho = 0$ is conducted using the regression first subtracting the actual mean. The test for $H_0 : \mu = 0$ is computed using the usual asymptotic t-test for a sample average.

[8]For the specifications where the rejection rate higher than of 5%, portmanteau tests all rejected with p-values less than 0.001.

All models reject correct specification more frequently than size with the exception of the RM specification which again fails to reject due to high variance rather than low mean. Of the models which share common variance forecasts (CCC, DCC, FLEXM and aCCC, aDCC, aFLEXM) the DCC and aDCC models appear to provide better correlation forecasts, although the difference in the rejection rates between the constant and dynamic conditional correlation models is small. The fixed factor orthogonal GARCH specification performance is particularly poor. The CKL benchmark model again provides good unconditional performance while failing to capture covariance dynamics.

The long-short portfolio results, based on 2450 portfolios, are generally worse. The only notable performance improvement was in the FLEXM and aFLEXM specifications. While the joint test rejects far too often for these specifications, they do provide the best description of the dynamics as measured by both the rejection rate on the autocorrelations and the grand mean of the autocorrelations. Figures 2 and 3 contain box plots of the mean and autocorrelations of the standardized residuals for the long-long and long-short portfolios respectively. Apparent in figure 2, all specification except the CKL produce standardized residuals with variance in excess of 1. The RM specification, again despite the failure to reject an average volatility of 1, produces standardized residuals with excess variance in all 1225 portfolios. Figure 3 demonstrates that all specifications have trouble adequately modeling the dynamics of the long-short portfolio return variances. All specifications produced autocorrelations with a lower quartile greater than 0.

Overall, the results of these simple specification tests are disappointing. If a winner had to be declared, the *best* would likely be one of the DCC specifications. These specifications consistently produces the lowest rejection rates across the three tests and three portfolios. However, even flexible parameterizations such as these rejected far too often. The most surprising results come from the highly parameterized FLEXM specifications. The likely cause of the poor performance of these models is the lack of structure in estimating the conditional covariances, necessitating the Frobenius norm minimization procedure to ensure that conditional covariances are positive semi-definite.

4.2 Global Minimum Variance Portfolio Formulation

To examine whether a specification is capable of capturing features which would be useful to investors, we examined the performance of portfolios constructed using the forecast covariance to estimate minimum variance weights. To avoid problems associated with estimating conditional means, we will examine the properties of global minimum variance portfolios constructed under three weighting schemes: unconstrained, a 20% short sales constraint, and a no short-limited exposure portfolio where no more than 8% may be invested in a single asset. In the unconstrained case, the problem can be formulated:

$$\min_{\mathbf{w_t}} \mathbf{w}_t' \mathbf{H}_{t|t-1} \mathbf{w}_t \quad \text{subject to } \mathbf{w_t}' \boldsymbol{\iota} = 1 \tag{41}$$

where $\boldsymbol{\iota}$ is a K by 1 vector of ones. Unconstrained portfolio weights are readily calculated from forecast covariance matrices as

$$\mathbf{w}_t = \frac{\mathbf{H}_{t|t-1}^{-1} \boldsymbol{\iota}}{\boldsymbol{\iota}' \mathbf{H}_{t|t-1}^{-1} \boldsymbol{\iota}}. \tag{42}$$

In the short sales constrained case, an additional constraint, that $|\mathbf{w}_t|' \boldsymbol{\iota} < 1.20$ is imposed, while in the no short-limited exposure case, the additional constraint becomes $0 \leq w_{it} \leq .08$, $i = 1, \ldots, K$. Numerical methods are employed to solve the two constrained problems.

Table 5 presents the results of the portfolios formed using the three weighting methods. For each method, the realized variances, transformed to annualized standard deviation, are presented along with excess variance and a t-test that excess variance is 0. Excess variance was calculated as

$$\frac{1}{T-R}\sum_{t=R+1}^{T}\frac{(\mathbf{w}_t'\mathbf{r}_t)^2}{\mathbf{w}_t\mathbf{H}_t\mathbf{w}_t}-1 \tag{43}$$

where T is the total sample size and R is the last observation used for the initial training period.

Many of the GARCH specifications outperform the tightly parameterized benchmark models. The table contains two noteworthy results. First, the variance targeting models produce the lowest portfolio variance for all specifications despite being among the worst in capturing the features of the conditional covariance in the portfolio standardizes residuals. Second, allowing asymmetries improves the performance of the portfolios.

Unfortunately, despite the positive outcomes evidenced in the table, all specification predict portfolios with much smaller variance than they realize. The closest is off by nearly 50% while most are off by a factor closer to 2. Also evident from the table and consistent with a stylized fact of portfolio allocation, mild constraints improve performance. Even tightly parameterized models, such as the VT specification, benefit (Jagannathan & Ma 2004). Additionally, the excess variances decline as the constraints tighten and the sole case where the null of a correct unconditional forecast cannot be rejected occurs for the asymmetric variance targeting specification in the most constrained portfolio.

Overall, the GMVP results were mixed but encouraging. All of the forecasts, even the highly constrained CKL benchmark model, were overly optimistic in their ability to reduce portfolio variance. It is likely that this is partially due to the unconstrained estimation of the intercept in all of these models. However, additional flexibility in modeling variance dynamics was often helpful. For instance, the DCC performs better than the CCC and the asymmetric parameterizations perform better than their symmetric counterparts in 5 of 6 specifications.

4.3 Tracking Error Minimization

The third metric employed to evaluate the specification of alternative parameterizations is performance in tracking error variance minimization. Simply put, if returns on an asset to be tracked are denoted r_t^0 and a returns on a set of assets used to construct the tracking portfolio are denoted $\mathbf{r}_t$, then the optimal weights are chosen as

$$\min_{\mathbf{w}_t}(r_t^0-\mathbf{w}_t'\mathbf{r_t})^2. \tag{44}$$

Alternatively, these weights correspond to a minimum variance portfolio where the weights are chosen to minimize the variance subject to having a weight of 1 in a single asset. This problem is formulated

$$\min_{\mathbf{w_t}}\mathbf{w_t}'\mathbf{H}_t\mathbf{w_t}\quad\text{subject to } w_{it}=1 \tag{45}$$

and the optimal weights are easily computed from the forecast covariance as

$$\mathbf{w}_t=\frac{\mathbf{H}_{t|t-1}^{-1}\mathbf{e}_i}{\mathbf{e}_i'\mathbf{H}_{t|t-1}^{-1}\mathbf{e}_i}. \tag{46}$$

For each specification, 50 tracking portfolios were formed, one for each asset, using the remaining 49 to track.

Table 6 contains the variance of the tracking errors averaged across all 50 tracking portfolios and transformed to annualized standard deviations. For reference, the variance of the raw returns, transformed into annualized standard deviation, was 30.01%. The table also contains excess variance, defined above, a t-test that the excess variance is 0, and the median rank among the 14 specifications. The rank was computed by sorting the realized tracking error variances for each of the 50 portfolios.

The CCC, DCC, 3-factor orthogonal, and variance targeting parameterizations all performed well. Additionally, asymmetries do not appear to help in forming the tracking error portfolios. The correlation specifications do not reject the null that there is no excess variance in the tracking errors. However, when we consider the ordinal rank of the forecasts, the fixed factor orthogonal model performs the best, producing a median rank of 2. Figure 4 provides a detailed breakdown of the relative ranks of the alternative specifications. Light areas correspond to lower variance while darker areas correspond to higher variance. From the figure, the OGARCH specification with 3 factors has a large number of first or second performances. Unfortunately, it also has a few poor performances with ranks as low as 10. The DCC specification, while not performing at the top as often, never ranks lower than 8^{th} .

4.4 Relative Performance

The specification tests presented thus far have all been absolute. However, the portfolio construction tests can be transformed into relative performance comparisons to examine whether one specification produces significantly smaller portfolio variances. We use the methodology of Diebold & Mariano (1995) to access the relative performance of a subset of the specifications examined. The subset was necessary for two reasons. First, the methodology of Diebold and Mariano is not applicable to strictly nested models. Thus, it is not possible to include both the symmetric and asymmetric versions of the same class. Second, 14 models are simply too many to adequately draw conclusions about relative performance. Based on the previous results, we limited the analysis to the aDCC, aFLEXM, aVT and CKL specifications.

The null hypothesis is that two specifications produce identical portfolio variances. To test this hypothesis, we can examine the properties of the daily squared returns from two competing specifications. If the models produce the same variance, the difference between the squared portfolio returns should be small. Define r_t^{P1} to be the returns from the portfolio weights generated by the first forecast and r_t^{P2} to be the weights generated from the second forecast. We can define the mean difference of the variances to be

$$\hat{\delta} = \frac{\sum_{t=R+1}^{T} (r_t^{P1})^2 - (r_t^{P2})^2}{T - R} \tag{47}$$

and a test statistic, that the mean difference is zero, can be computed from

$$\sqrt{T-R}\frac{\hat{\delta}}{\sqrt{\text{var}(\hat{\delta})}}. \tag{48}$$

Under the null that the average variances are equal, and assuming the in-sample data length grows faster than the out-of-sample data length, Diebold & Mariano (1995) have established this statistic has a standard normal limiting distribution.

The top panel of table 7 contains numerical values from Diebold-Mariano t-tests that the realized portfolio variances were the same. In the unconstrained case, the CKL benchmark model is significantly outperformed by both the aFLEXM and the aVT specifications. However, none of the GARCH specifications is able to out-perform another at conventional levels. In the limited short case, the only interesting result was the aVT's out performance of the CKL benchmark. In the highly constrained case, the aVT's performance was significantly better than the three other specifications. Thus, despite the seemingly differences in the annualized standard deviations of the portfolio returns, the aVT specification produced portfolio variances statistically lower than alternative specifications.

Turning attention to the tracking error problem, the bottom panel of table 7 contains results for the 4 specifications. Subpanel (d) presents results of a Diebold-Mariano test that the variance of the tracking errors for all 50 assets was the same. The aDCC specification performs better than the other three specifications although the difference is insignificant in the case of the aVT specification. The aVT parameterization also significantly outperformed the other two. Subpanel (e) contains the number of tracking portfolios (50 in total) where the null of equal tracking error variance could be rejected using a 5% test. The aDCC outperforms the aFLEXM 49 out of 50 times, the aVT 12 out of 50 times and the CKL 42 out of 50. The aVT outperforms the aFLEXM in 44 out of 50 portfolios, the aDCC specification in 5 out of 50 and the CKL benchmark model in all cases. The CKL and the aFLEXM specifications outperform each other in less then 20% of the cases and never produce tracking error variances significantly lower than the aVT or aDCC specifications.

The results of the relative performance are clear: the asymmetric variance targeting model, despite poor performance in the specification analysis, consistently produced *statistically* superior results. Additionally, the CKL benchmark model was outperformed in all but two of the cases. These results provide strong evidence that there *is* value in modeling conditional covariances for portfolio allocation.[9]

5 Conclusions

Only a small sample of GARCH models are viable for analyzing the covariance of many assets. This paper provides an overview and comparison of the performance of those specifications feasible with 50 or more assets. These specifications all transform a large, difficult estimation problem into one or more simple problems with low dimension parameter spaces using a combination of closed-form estimators, multistage estimation and parameter restrictions. We extend existing specifications to include asymmetries in conditional covariance and compare the performance to two benchmark models using 50 S&P 500 sector indices.

Initial tests, based on the statistical properties of forecast standardized residuals, were disappointing. None of the specifications was able to provide an adequate characterization of the conditional covariance. However, when considered in terms of their ability to capture information useful in portfolio allocation, there were many positive results. A number of the conditional covariance specifications were able to produce superior out-of-sample portfolio allocations relative to a sophisticated benchmark model selected to perform well at this task. Additionally, detailed examination of a select sub-sample of the specifications provided evidence that the difference in portfolio variances was statistically significant.

While these two findings are somewhat contradictory, their is for improvement. Some flaws in the models, specifically the failure of many of the specifications to adequately capture variation in the conditional *variances*, can be easily addressed in the some of the models examined. The correlation models can be trivially enhanced by choosing series specific univariate specifications to better describe the dynamics of the

[9] We have found that obtaining reliable estimates of the parameters in many of these models is difficult. When we did not attempt to robustly estimate the model parameters, the specifications which rely on univariate GARCH (or GJR-GARCH) specifications performed much worse. We have detailed our efforts to ensure that the estimates are located at the global optima in Appendix B.

data. Alternatively, additional lags of innovations or conditional variances may be adequate to improve the properties of the forecast standardized residuals.

While this paper has provided evidence there is value to modeling conditional covariance using large-scale GARCH models, there remain many interesting questions in this area. It would be interesting to examine, in detail, the reason why the variance targeting, and to a lesser extent, the correlation targeting models, performed well in the portfolio problems. The variance targeting specifications are highly constrained and employ a pooled innovation parameter for the volatility series. Is this the root of the strong performance of this class of models? Would extending to common innovation parameter into a DCC specification, such that all of the univariate models are restricted to share a common parameter, improve the performance of the DCC? or the FLEXM? Is there a less *ad hoc* method available than pure parameter pooling to improve asset allocation? For instance, the variance targeting specification and the FLEXM are two ends of a single spectrum. Is there a clever method to choose a point midway between these which uses partial pooling to improve asset allocation performance? How well would these models perform when using lower frequency data (weekly, monthly) and would the performance of highly parameterized specifications deteriorate? We leave these questions for future research.

References

Alexander, C. & Chibumba, A. (1996), Multivariate Orthogonal Factor GARCH. University of Sussex Discussion Papers in Mathematics.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**(3), 307–327.

Bollerslev, T. (1990), 'Modeling the coherence in short run nominal exchange rates: A multivariate generalized ARCH model', *Review of Economics and Statistics* **72**(3), 498–505.

Bollerslev, T., Engle, R. F. & Wooldridge, J. M. (1988), 'A capital asset pricing model with time-varying covariances', *Journal of Political Economy* **96**(1), 116–131.

Campbell, J. Y. & Hentschel, L. (1992), 'No news is good news: An asymetric model of changing volatility in stock returns', *Journal of Financial Economics* **31**(3), 281–318.

Cappiello, L., Engle, R. F. & Sheppard, K. (2006), 'Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns', *Journal of Financial Econometrics* **4**(4), 537–572.

Carhart, M. M. (1997), 'On persistence in mutual find performance', *Journal of Finance* **52**(1), 57–82.

Cavaglia, S., Brightman, C. & Aked, M. (2000), 'On the importance of industry factors: Implications for global portfolio management', *Financial Analysts Journal* **56**(5), 41–54.

Chan, L. K., Karceski, J. & Lakonishok, J. (1999), 'On portfolio optimization: Forecasting covariances and choosing the risk model', *Review of Financial Studies* **12**, 937–974.

Christie, A. (1982), 'The stochastic behavior of common stock variances: Value, leverage and interest rate effects', *Journal of Financial Economics* **10**(4), 407–432.

Diebold, F. X. & Mariano, R. S. (1995), 'Comparing predictive accuracy', *Journal of Business & Economic Statistics* **13**(3), 253–263.

Diebold, F. X. & Nerlove, M. (1989), 'The dynamics of exchange rate volatility: A multivariate latent factor ARCH model', *Journal of Applied Econometrics* **4**(1), 1–21.

Ding, Z. & Engle, R. (2001), 'Large scale conditional matrix modeling, estimation and testing', *Academia Economic Papers* **29**(2), 157–184.

Engle, R. F. (2002), 'Dynamic conditional correlation - a simple class of multivariate GARCH models', *Journal of Business and Economic Statistics* **20**(3), 339–350.

Engle, R. F., Granger, C. W. J. & Kraft, D. (1984), 'Combining competing forecasts of inflation with a bivariate ARCH model', *Journal of Economic Dynamics and Control* **8**(2), 151–165.

Engle, R. F. & Kroner, K. F. (1995), 'Multivariate simultaneous generalized ARCH', *Econometric Theory* **11**(1), 122–150.

Engle, R. F. & Mezrich, J. (1996), 'GARCH for groups', *Risk* **9**(8), 36–40.

Engle, R. F., Ng, V. & Rothschild, M. (1990), 'Asset pricing with a factor ARCH covariance structure: Empirical estimates for treasury bills', *Journal of Econometrics* **45**(2), 213–237.

Fama, E. F. & French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**, 3–56.

Glosten, L., Jagannathan, R. & Runkle, D. (1993), 'On the relationship between the expected value and the volatility of the nominal excess return on stocks', *Journal of Finance* **48**(5), 1779–1801.

Jagannathan, R. & Ma, T. (2004), 'Risk reduction in large portfolio: Why imposing the wrong constraint helps', *Journal of Finance* **58**(4), 1651–1683.

Kraft, D. & Engle, R. F. (1982), ARCH in multiple time series. UCSD Working Paper 82-23.

Kroner, K. E. & Ng, V. K. (1998), 'Modeling asymmetric comovements of asset returns', *Review of Financial Studies* **11**(4), 817–844.

Laurent, S., Bauwens, L. & Rombouts, J. V. K. (2006), 'Multivariate garch models: a survey', *Journal of Applied Econometrics* **21**(1), 79–109.

Ledoit, O., Santa-Clara, P. & Wolf, M. (2003), 'Flexible multivariate GARCH modeling with an application to international stock markets', *Review of Economics and Statistics* **85**(3), 735–747.

Rivers, D. & Vuong, Q. H. (2002), 'Model selection tests for nonlinear dynamic models', *The Econometrics Journal* **5**(1), 1–39.

Vuong, Q. H. (1989), 'Likelihood ratio tests for model selection and non-nested hypotheses', *Econometrica* **57**(2), 307–333.

White, H. (1980), 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *Econometrica* **48**(4), 817–838.

Wooldridge, J. M. (1990), 'A unified approach to robust, regression-based specification tests', *Econometric Theory* **6**(1), 17–43.

Wooldridge, J. M. (1991), 'On the application of robust, regression based diagnostics to models of conditional means and conditional covariances', *Journal of Econometrics* **47**(1), 5–46.

Zakoian, J. M. (1994), 'Threshold heteroskedastic models', *Journal of Economic Dynamics and Control* **18**(5), 931–955.

6 Appendix A: Sector Indices

S&P sector indices used in this study. Names are listed as they appear in DataStream.

- FOOD RETAIL
- PHARM & BIOTECHNOLOGY
- DRUG RETAIL
- FOOD DISTRIBUTORS
- PUBLISHING & PRINTING
- HOME FURNISHINGS
- ELECTRICAL COMP & EQUIP
- BROADCASTING & CABLE TV
- GOLD
- CONTAINERS & PACKAGING
- TOBACCO
- DIVERSIFIED METALS & MINING
- STEEL
- WIRELESS TELECOM SERV
- DEPARTMENT STORES
- CHEMICALS
- FOOD PRODUCTS
- COMMERCIAL PRINTING
- DIVERSIFIED TELECOM SERV
- ELECTRIC UTILITIES
- AGRICULTURAL PRODUCTS
- HOUSEWARES & SPECIALTIES
- CASINOS & GAMING
- COMMERCIAL SERV & SUPP IN
- APPAREL RETAIL
- HEALTH CARE SUPPLIES
- FOREST PRODUCTS
- AEROSPACE & DEFENCE
- ELECTRONIC EQUIP & INSTRUM
- HOMEBUILDING
- BIOTECHNOLOGY
- ENERGY EQUIP & SERV
- FOOTWEAR
- INSURANCE BROKERS
- HEALTH CARE EQUIP
- LEISURE PRODUCTS
- GAS UTILITIES
- CONSUMER DURABLES & APP
- COMMUNI EQUIP
- AUTO COMPONENTS
- AIR FREIGHT & COURIERS
- ALUMINIUM
- BUILDING PRODUCTS
- BEVERAGES IN
- AUTOMOBILE MANUFACTURERS
- BREWERS
- SPECIALTY STORES
- CONSTRUCTION & FARM MACHINE
- APPAREL & ACCESSORIES
- BANKS

7 Appendix B: Recursive Estimation Issues

We have documented that estimation difficulties, primarily in the form of local optima, were commonly found during the estimation of parameters in many of multivariate GARCH specifications. To ensure the results of this paper are not biased due to the use of parameter estimates located at local rather than global optima, we employed three techniques:

- Large grid searches
- Multi-start optimization
- Forward and backward recursive estimation

Large grid searches
For each specification, whether a univariate component of a correlation specification or a complete specification such as a variance targeting model, we evaluated the log-likelihood using a grid of at least 100 points, always including a constant correlation model as a control.
Multi-start optimizations
Using the most likely potential starting values from the grid search, including the constant conditional variance point, the log-likelihood was maximized. These results were saved and any terminal parameters which were sufficiently different from the other estimates were further investigated using a new grid centered around the distinct optima. From these estimates, the time series of likelihood maximizing parameter vectors was constructed.
Forward and Backward recursive estimation
Using the previously found likelihood maximizing parameters, starting at the first pseudo-end-of-sample point t, the model was re-estimated using the likelihood maximizing parameters from $t+1$ and $t-1$. The final set of parameters was chosen to be the maximum of from these two final optimizations and the previous parameter estimates. The forward recursion was repeated for $t+1, t+2, \ldots, T$.

An initial version of the paper did not attempt to robustly estimate the model parameters. Using estimated parameters which were unnecessarily volatile (over time) and not located at global optima resulted in a meaningful change of the final results of this paper. We feel that recursive examination of volatility models, specifically multivariate specifications, must be carefully conducted to avoid arriving at false conclusions about the value of these specifications.

8 Appendix C: Tables

Table 1: Descriptive statistics of the S&P 500 sector data. All statistics were calculated using daily data, although the mean and standard deviation are reported in annualized form. q indicates quantile.

	Mean Return	Std. Dev.	Skewness	Kurtosis
Max.	0.2649	0.5132	0.4626	15.0516
q=.95	0.2079	0.4235	0.3178	11.9808
q=.75	0.1319	0.3324	0.0906	7.9863
Median	0.0706	0.2855	-0.0696	6.8496
q=.25	0.0315	0.2398	-0.2342	6.0719
q=.05	-0.0079	0.1926	-0.7390	4.8717
Min.	-0.0660	0.1741	-0.8373	4.3803

Table 2: Relative log-likelihoods of the 12 specifications using the full sample of 2226 observations. The largest log-likelihood, produced by the asymmetric DCC specification was subtracted from the remaining models. The asymmetric FLEXM and the FLEXM model, which nest the asymmetric variance targeting and the variance targeting models, respectively, produced *lower* log-likelihoods than their nested counterpart. While these order reversals are small-sample effects, they are an important indication of the scaling performance of the FLEXM models. The last two columns report the number of parameters as a function of the number of assets and when $K = 50$.

Model	Relative Log-likelihood	Number of Parameters	Number when $K = 50$
CCC	-476	$\frac{K^2+5K}{2}$	1375
aCCC	-340	$\frac{K^2+7K}{2}$	1425
DCC	-149	$\frac{K^2+5K}{2}+2$	1377
aDCC	-	$\frac{K^2+7K}{2}+3$	1428
FLEXM	-7897	$\frac{3K^2+3K}{2}$	3825
aFLEXM	-5654	$2K^2+2K$	5100
O_3	-7176	$\frac{K^2+7K}{2}+9$	1434
aO_3	-7140	$\frac{K^2+7K}{2}+12$	1437
O_K	-1381	K^2+3K	2650
aO_K	-1275	K^2+4K	2700
VT	-4416	$\frac{K^2+K}{2}+2$	1277
aVT	-4155	$\frac{K^2+K}{2}+3$	1278

Table 3: Specification tests based on single asset portfolios. The first column reports the rejection percentage of a joint hypothesis test that $H_0 : \mu = 1, \rho = 0$ from the regression $sr_t = \mu + \rho sr_{t-1} + \nu_t$ where sr_t denotes the forecast standardized residual. The second column reports the percentage of tests rejecting the null that the model is correctly specified for the *un*conditional mean ($\hat{\mu} = \frac{1}{1227}\sum_{t=1000}^{2226} st_t$). The third column reports the percentage rejecting a null of no autocorrelation ignoring any mean misspecification ($H_0 : \rho = 0$) from the regression $sr_t - \hat{\mu} = \rho(sr_{t-1} - \hat{\mu}) + \nu_t$. All tests were performed using 5% critical values. Finally, the last two columns report the grand mean of the portfolio standardized residuals across all 50 portfolios and the grand mean of the estimated autocorrelations. If the models are well specified, these two columns should have values near 1 and 0 respectively.

Single Asset Portfolios

	$H_0 : \mu = 1$ and $\rho = 0$	$H_0 : \mu = 1$	$H_0 : \rho = 0$	$\bar{\mu}$	$\bar{\rho}$
CCC	0.26	0.28	0.14	1.092	0.011
aCCC	0.28	0.36	0.10	1.096	0.013
DCC	0.26	0.28	0.14	1.092	0.011
aDCC	0.28	0.36	0.10	1.096	0.013
FLEXM	0.26	0.28	0.14	1.092	0.011
aFLEXM	0.28	0.36	0.10	1.096	0.013
O_3	0.82	0.72	0.44	1.307	0.094
aO_3	0.86	0.76	0.48	1.345	0.098
O_K	0.62	0.62	0.24	1.147	0.059
aO_K	0.66	0.62	0.22	1.182	0.061
VT	0.84	0.74	0.60	1.210	0.113
aVT	0.74	0.70	0.52	1.178	0.106
CKL	0.50	0.06	0.64	0.999	0.124
RM	0.02	0.00	0.06	1.095	0.020

Table 4: Specification tests based on 2 asset portfolios. The left 5 columns correspond to the long-long portfolios (1225 in total) while the right 5 columns correspond to the long-short portfolios (2450 in total). All tests were conducted on squared portfolio returns standardized by forecast variances. In each block of 5, the first column reports the rejection percentage of a joint hypothesis test that $H_0 : \mu = 1, \rho = 0$ from the regression $sr_t = \mu + \rho sr_{t-1} + \nu_t$ where sr_t denotes the forecast standardized residual. The second column reports the percentage of tests rejecting the null that the model is correctly specified for the *un*conditional mean ($\hat{\mu} = \frac{1}{1227}\sum_{t=1000}^{2226} st_t$). The third column reports the percentage rejecting a null of no autocorrelation ignoring any mean misspecification ($H_0 : \rho = 0$) from the regression $sr_t - \hat{\mu} = \rho(sr_{t-1} - \hat{\mu}) + \nu_t$. All tests were performed using 5% critical values. Finally, the last two columns report the grand mean of the portfolio standardized residuals across all portfolios and the grand mean of the estimated autocorrelations. If the models are well specified, these two columns should have values of 1 and 0 respectively.

	Long-Long Portfolios					**Long-Short Portfolios**				
	$H_0 : \mu = 1$ and $\rho = 0$	$H_0 : \mu = 1$	$H_0 : \rho = 0$	$\bar{\mu}$	$\bar{\rho}$	$H_0 : \mu = 1$ and $\rho = 0$	$H_0 : \mu = 1$	$H_0 : \rho = 0$	$\bar{\mu}$	$\bar{\rho}$
CCC	0.223	0.230	0.152	1.064	0.003	0.260	0.235	0.155	1.039	0.040
aCCC	0.267	0.269	0.170	1.065	0.001	0.251	0.224	0.150	1.037	0.042
DCC	0.213	0.209	0.156	1.061	0.003	0.233	0.223	0.136	1.041	0.037
aDCC	0.258	0.245	0.174	1.063	0.000	0.232	0.204	0.130	1.038	0.039
FLEXM	0.648	0.727	0.183	1.164	-0.003	0.454	0.444	0.084	0.905	0.024
aFLEXM	0.583	0.653	0.187	1.145	-0.003	0.391	0.367	0.089	0.928	0.030
O_3	0.823	0.793	0.262	1.249	0.060	0.923	0.866	0.742	1.424	0.114
aO_3	0.877	0.850	0.271	1.300	0.066	0.924	0.867	0.743	1.429	0.115
O_K	0.555	0.569	0.120	1.134	0.031	0.640	0.569	0.351	1.159	0.068
aO_K	0.648	0.676	0.111	1.180	0.034	0.657	0.584	0.342	1.170	0.068
VT	0.845	0.806	0.462	1.208	0.105	0.834	0.738	0.694	1.204	0.110
aVT	0.759	0.688	0.411	1.164	0.096	0.820	0.726	0.679	1.196	0.107
CKL	0.477	0.017	0.600	0.995	0.123	0.573	0.045	0.687	0.986	0.108
RM	0.062	0.000	0.099	1.079	0.004	0.031	0.000	0.059	1.097	0.021

Table 5: Results of constructing global minimum variance portfolios using forecast covariance matrices. The three sets of columns correspond performance in one of three portfolios: one formed with unconstrained weights, one formed using a maximum 20% short interest, and one formed imposing a no-short, 8% maximum exposure limit. Each pair of columns contains the realized variance, transformed into annualized standard deviation and the variance of the daily portfolio returns standardized by the forecast variance. Also reported in parentheses are t-stats from a test that the excess variance was zero.

	Unconstrained		**Limited Short**		**No Short**	
	Actual Annualized Std. Dev.	Excess Variance	Actual Annualized Std. Dev.	Excess Variance	Actual Annualized Std. Dev.	Excess Variance
CC	13.93	92	13.52	44	14.33	24
		(9.85)		(6.47)		(4.03)
aCC	13.87	99	13.46	47	14.27	25
		(10.13)		(6.59)		(4.15)
DCC	13.85	88	13.43	42	14.27	22
		(9.58)		(6.24)		(3.79)
aDCC	13.75	93	13.38	45	14.23	24
		(9.86)		(6.48)		(4.03)
FLEXM	13.63	150	13.51	129	14.28	115
		(12.26)		(11.95)		(11.06)
aFLEXM	13.49	124	13.26	101	14.15	90
		(11.87)		(10.47)		(9.61)
O_3	15.67	86	16.07	61	14.74	21
		(8.96)		(6.96)		(3.46)
aO_3	15.53	92	16.07	71	14.72	33
		(9.20)		(7.56)		(4.62)
O_K	19.99	467	14.92	36	14.71	19
		(3.50)		(5.35)		(3.30)
aO_K	37.51	10378	14.95	46	14.65	31
		(1.85)		(6.15)		(4.59)
VT	13.49	47	13.28	28	13.90	19
		(5.86)		(3.94)		(2.75)
aVT	13.44	41	13.12	18	13.81	9
		(5.42)		(2.82)		(1.47)
CKL	14.11	83	13.57	42	14.10	22
		(8.05)		(4.84)		(2.75)
RM	22.04	2314	13.77	197	14.29	98
		(16.27)		(13.58)		(10.29)

Table 6: Average tracking error variances (reported in annualized standard deviations), ratios of realized to predicated variances, and t-tests that the ratio was actually 1. 50 tracking portfolios were constructed from each specification at each point in time, one tracking each asset using the remaining 49. The reported values are grand averages of the 50 tracking error variances, 50 standardized variances, and 50 t-stats. The final column reports the median rank, across the 50 portfolios of the tracking error variances of the tracking error variance among the 14 specification considered.

	Actual Variance	% Excess Variance	T-stat (H_0 : Ratio=1)	Median Rank
CC	0.266	9	1.22	5
aCC	0.267	9	1.24	6
DCC	0.265	10	1.25	4
aDCC	0.266	10	1.28	5
FLEXM	0.288	33	3.73	11
aFLEXM	0.286	29	3.32	10
O_3	0.266	62	5.29	2
aO_3	0.266	62	5.29	2.5
O_K	0.352	88	4.66	12
aO_K	0.674	9976	1.92	14
VT	0.267	44	4.51	6
aVT	0.267	41	4.33	5.5
CKL	0.285	62	6.07	9
RM	0.455	2206	17.8	13

Table 7: Relative performance. The top panel contains test values from Diebold-Mariano tests on the squared portfolio returns from global minimum variance portfolios using three sets of weights: (a) Unconstrained, (b) 20% Short Sales Constraint and (c) No Short, 8% Maximum. A numerical value in a row-column combination indicates that the specification corresponding to the row label out performed the specification which corresponds to the column label. For instance, in subpanel (a), in the row corresponding to the aDCC specification and column corresponding to the CKL benchmark model, the t-test value was 1.173 favoring the aDCC model. The bottom panel contains results of Diebold-Mariano tests on the tracking error portfolio variances. Subpanel (d) contains the value of a t-test that the variances were equal across *all* portfolios. Subpanel (e) contains the number of rejections for the 50 individual tracking portfolios where the variance of the row specification was significantly lower than that of the column specification. For instance, the aDCC model produced statistically smaller variances for 49, 12 and 42 of the tracking portfolios when compared to the aFLEXM, aVT, and CKL specifications respectively. All tests were conducted using 2-sided 95% confidence regions.

Global Minimum Variance Portfolios

Better \ **Worse**	aDCC	aFLEXM	aVT	CKL
aDCC	-	-	-	1.173
aFLEXM	1.350	-	-	2.174
aVT	1.223	0.277	-	4.488
CKL	-	-	-	-

(a): Unconstrained

Better \ **Worse**	aDCC	aFLEXM	aVT	CKL
aDCC	-	-	-	0.676
aFLEXM	0.664	-	-	1.283
aVT	1.168	0.738	-	3.651
CKL	-	-	-	-

(b): Limited Short

Better \ **Worse**	aDCC	aFLEXM	aVT	CKL
aDCC	-	-	-	
aFLEXM	0.784	-	-	-
aVT	2.612	2.303	-	3.460
CKL	0.679	0.258	-	-

(c): No Short, 8% Maximum Weight

Tracking Error Portfolios

Better \ **Worse**	aDCC	aFLEXM	aVT	CKL
aDCC	-	3.672	0.404	3.373
aFLEXM	-	-	-	-
aVT	-	3.093	-	4.709
CKL	-	0.125	-	-

(d): All portfolios

Better \ **Worse**	aDCC	aFLEXM	aVT	CKL
aDCC	-	49	12	42
aFLEXM	0	-	0	6
aVT	5	44	-	50
CKL	0	8	0	-

(d): Individual portfolios

9 Appendix D: Figures

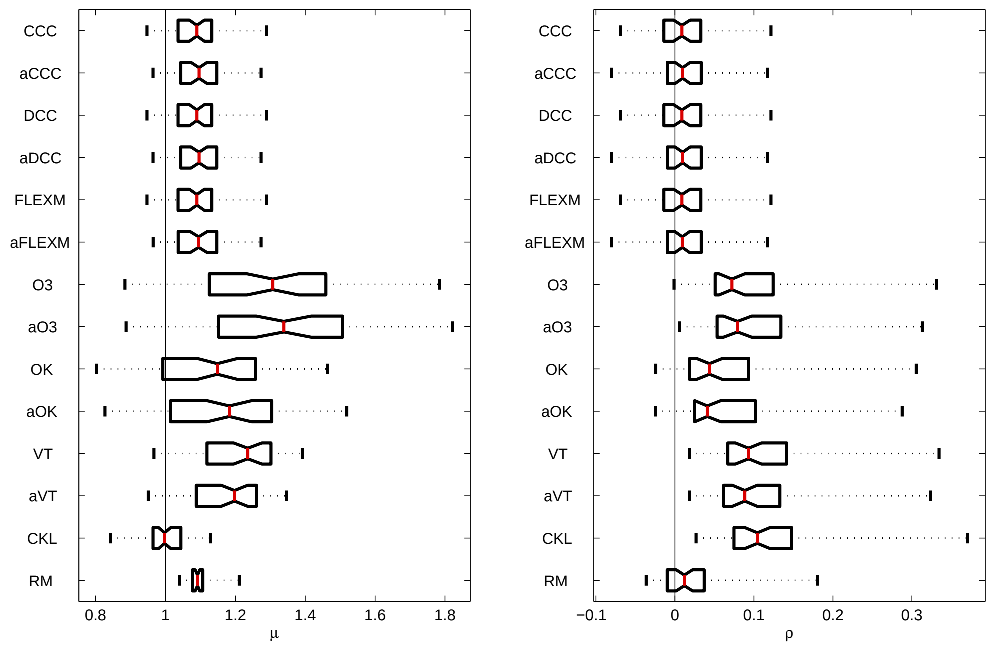

Figure 1: Plot of the average tracking errors and autocorrelations for the standardized residuals based on single asset portfolios. The left plot contains a plot of the min, median, max and quartiles of the 50 standardized residual averages for each forecasting method. These values should be near one if the forecast variance is correct. The right plot contains the min, median, max and quartiles of the first autocorrelation of the standardized residuals for the 50 portfolios. A good forecast should produce autocorrelations near 0.

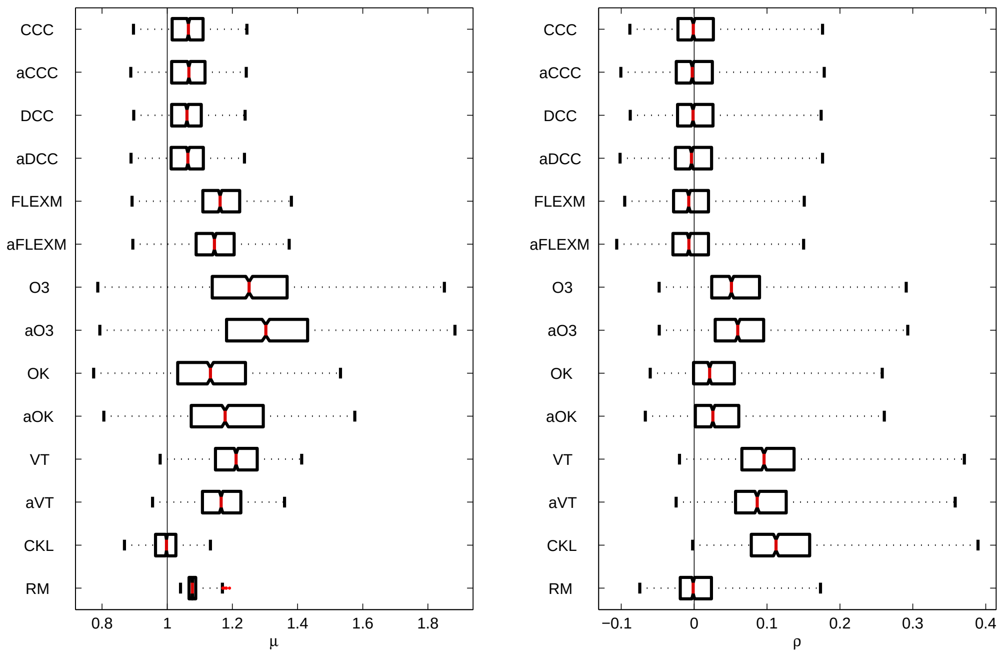

Figure 2: Plot of the average tracking errors and autocorrelations for the standardized residuals based on two asset long-long portfolios. The left plot contains a plot of the min, median, max and quartiles of the standardized residual (1225 unique portfolio) averages for each forecasting method. These values should be near one if the forecast variance is correct. The right plot contains the min, median, max and quartiles of the first autocorrelation of the standardized residuals. A good forecast should produce autocorrelations near 0.

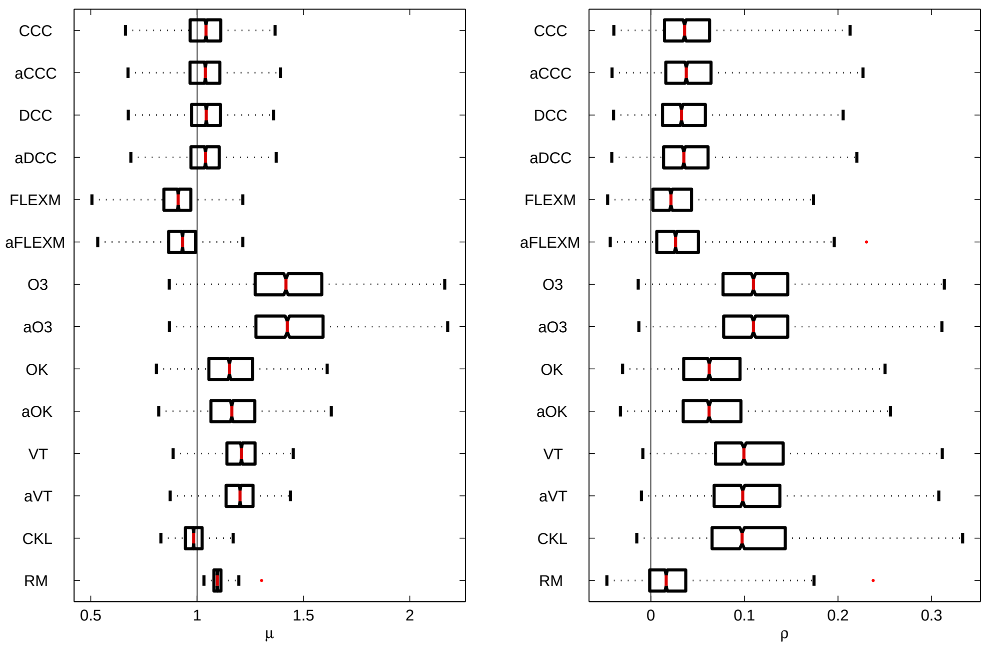

Figure 3: Plot of the average tracking errors and autocorrelations for the standardized residuals based on two asset long-short portfolios. The left plot contains a plot of the min, median, max and quartiles of the standardized residual (2450 unique portfolios) averages for each forecasting method. These values should near be one if the forecast variance is correct. The right plot contains the min, median, max and quartiles of the first autocorrelation of the standardized residuals. A good forecast should produce autocorrelations near 0.

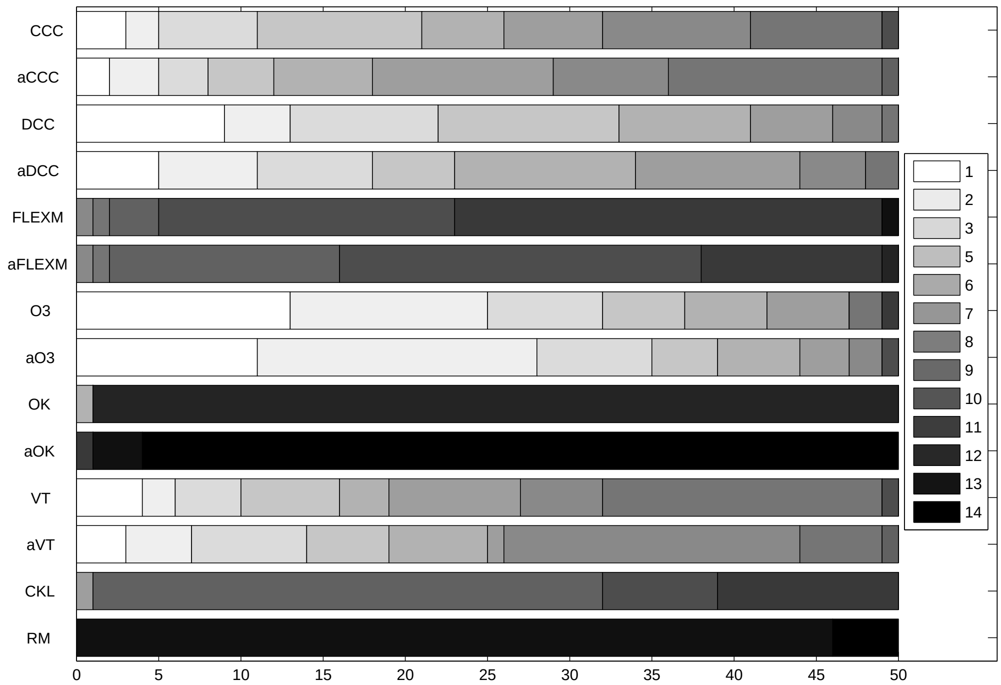

Figure 4: Plot of the *ordinal rank* of the tracking portfolio variances, sorted from lowest to highest variance. For each of the assets (50), tracking portfolios were constructed using one-step ahead forecast covariances from each model. The average tracking error variance for each asset was sorted from one to 14. White areas correspond to the best performance while progressively darker regions correspond to specifications which produced tracking errors with higher variance. For instance, the asymmetric orthogonal GARCH model with K factors produced tracking portfolios with the highest variance in 45 out of 50 opportunities as indicated by the large black bar.