



Pekka Malo – Antti Kanto

EVALUATING MULTIVARIATE GARCH MODELS
IN THE NORDIC ELECTRICITY MARKETS

Pekka Malo – Antti Kanto

EVALUATING MULTIVARIATE GARCH MODELS IN THE NORDIC ELECTRICITY MARKETS

Quantitative Methods in Economics and
Management Science

January
2005

HELSINGIN KAUPPAKORKEAKOULU
HELSINKI SCHOOL OF ECONOMICS
WORKING PAPERS
W-382

HELSINGIN KAUPPAKORKEAKOULU
HELSINKI SCHOOL OF ECONOMICS
PL 1210
FIN-00101 HELSINKI
FINLAND

© Pekka Malo, Antti Kanto and
Helsinki School of Economics

ISSN 1235-5674
ISBN 951-791-903-4 (Electronic working paper)

Helsinki School of Economics -
HeSE print 2005

Evaluating Multivariate GARCH models in the Nordic Electricity Markets

*Malo Pekka, *Kanto Antti

January 10, 2005

Abstract

This paper considers a variety of specification tests for multivariate GARCH models that are used in dynamic hedging in the electricity markets. The test statistics include the robust conditional moments tests for sign-size bias along with the recently introduced copula tests for an appropriate dependence structure. We consider this effort worthwhile, since quite often the tests of multivariate GARCH models are easily omitted and the models become selected ad-hoc depending on the results they generate. Hedging performance comparisons, in terms of unconditional and conditional ex-post variance portfolio reduction, are conducted.

1 Introduction

Electricity markets continue to confound international financial economists. With the rapid growth of derivative securities in deregulated electricity markets, the modelling and management of price risk have become important topics for researchers and practitioners. Until early 90's the electricity sector has been a vertically integrated industry, where prices were fixed by regulators as a function of generation and distribution costs. Today there is an increasing number of operating electricity exchanges over the world, where electricity providers compete to sell into the electricity market pools and the distributors purchase electricity from the pool at prices set by the intersection of aggregated demand and supply on an hourly basis.

The deregulation of electricity markets has introduced new elements of uncertainty making understanding and modelling of the electricity price behavior a very challenging task. An important complication that makes the markets difficult to model is the unique nature of electricity as a non-storable commodity [29; 2]. The production and

*Helsinki School of Economics, P.O. Box 1210, FIN-00101, Helsinki, Finland, pekka.malo@hkkk.fi, kanto@hkkk.fi

consumption of electricity have to be continuously balanced so as to prevent the electric power networks from collapsing. Every action taken by a player in the electricity system can affect all other activities on the grid. Failure of a single element can, if not managed properly, cause the subsequent rapid failure of many additional elements, disrupting the entire transmission system. Thus, combined with the characteristics of inelastic supply and demand, the non-storability creates conditions for large price jumps and spikes, that cannot be smoothed out using inventories [11; 28]. Intuitively the jumpy behavior in electricity spot prices is mainly attributed to the fact that a typical regional aggregate supply cost curve for electricity has almost always a kink at certain capacity level and the curve has a steep upward slope beyond that capacity level. A forced outage of a major power plant or a sudden surge in demand will make the regional electricity demand curve to cross the regional supply cost curve at its steep-rise portion thus causing a jump in the price process.

Given this setting the competitive power markets exhibit a level of price volatility unparalleled by traditional commodity markets. Therefore one of the main reasons for electricity generators and producers to trade in futures market is to monitor the volatility of their power portfolios and to minimize the negative impact of adverse fluctuations in electricity markets. However, the fact that the traditional convenience-yield type arbitrage arguments are not applicable in electricity markets, makes pricing of futures contracts and design of hedging strategies particularly challenging. In incomplete markets the starting point for non-arbitrage pricing is that the futures price is the expected value of the future spot price contaminated by some additional noise, and adjusted by a factor, which accounts for the risk preferences of the market [22; 39; 5]. Thus, a speculative component arises in the optimal hedging decision. The traditional assumption that the minimum risk hedge ratio will be the same irrespective of when hedging is undertaken has been often found to be in a sharp contrast with reality [8; 32; 46]. The non-storability of electricity only makes the need to consider time-varying risk-minimizing hedge ratios even more pronounced.

The existence of time-varying risk premia is strongly linked to the hedging strategies. A continuously varying premium impairs the effectiveness of futures market facilities in providing cover from the risk associated with futures trading [50]. For long time many authors [48; 44; 45] have argued that the storability of assets affects the performance of futures markets. Yang and Awokuse [50] have examined risk minimization hedging effectiveness for major storable and non-storable agricultural commodity futures markets. Their findings illustrate the great difference between commodities with different storability characteristics in terms of hedging. They argued that the hedging effectiveness is strong for all storable commodities but weak for all nonstorable commodities under consideration.

Currently there is, however, very little research available on dynamic hedging in the electricity markets. Byström's [9] study is perhaps the only one, where multivariate GARCH models have been applied to analyze short term hedging performance at the Nord Pool. The main thrust of the paper was in comparison of unconditional hedges against conditional hedges estimated using Orthogonal GARCH and Constant Correlation GARCH. His main findings were that hedging appears to provide some benefits, even though no straightforward arbitrage possibilities exist in the electricity

markets. Further he found that the performance of different hedge models depends significantly on whether unconditional or conditional portfolio variances are studied. The simple OLS hedge and the GARCH models managed to reduce both the conditional and unconditional variance, whereas the naive hedge successfully reduced only the unconditional variance. The findings are well in line with the mainstream as the studies of optimal hedging strategies have for long recognized the limits of the conventional hedging, which assumes perfect correlation between the spot and futures price changes [25; 26; 35]. One of the reasons why multivariate GARCH models are increasingly used has been the existence of time-varying risk premia, which effectively drive a wedge between the spot and the futures prices. Thus in a framework of rational agents, the simple hedge strategies can be proved to be highly inefficient if the perfect correlation assumption is removed.

Our study will differ from that of Byström [9] by considering a broader variety of MGARCH models and the primary focus will be more the tests of the different MGARCH model specifications. In particular, in the spirit of Kroner and Ng [31] we aim to consider different sign-size bias and parameter stability tests in order to check the robustness of the selected models. We consider this effort worthwhile, since quite often the tests of multivariate GARCH models are easily omitted and the models become selected ad-hoc depending on the results they generate. Specifically we want to demonstrate that the choice of a multivariate volatility model can lead to substantially different conclusions in estimating the optimal hedge ratio. As argued by Kroner and Ng [31], one should not place too much confidence in only statistically insignificant Ljung-Box statistics when evaluating GARCH models. Since even badly misspecified models can capture the serial correlation in the second moments and give insignificant Ljung-Box test statistics.

Our findings implied that all of the models studied in the paper appeared to be particularly bad at capturing between futures variance and shocks to the spot returns. However, given that the problems were limited to unaccounted asymmetries with the futures variance only, the need to incorporate asymmetric responses to the models is not too acute. Overall, we were surprised by how well all of the models performed as measured by conditional moments tests, even though the models produce quite different estimates for the conditional covariance matrix as indicated by preliminary comparisons. The results from the conditional and unconditional hedging performance tests were well inline with the previous findings of Byström, who has documented that the GARCH hedges outperform the other hedges when evaluated using conditional variances. However, the more interesting findings of the study concerned the selection of the underlying marginal distributions and their dependence structure. Although, we have clearly rejected the assumption of normality of the marginal distributions and found that a Student's t-distribution would be a more appropriate model, the copula statistics performed on the dependence structure provide mixed evidence suggesting that a normal dependence structure would fit better than t-copula. Thus, the problem we have faced in the use of the multivariate GARCH models selected for this study, is that all of them are misspecified in some respect: when using CCORR and diagonal BEKK we have misspecified the marginal distributions by assuming them to be normal, but when using diagonal T-BEKK, T-BEKK or asymmetric T-BEKK we might have

problems with the dependence structure.

The disposition of the paper is following. In the second section we provide a general overview of the most applied multivariate GARCH models that are to be used in the study. The next chapter continues from this by outlining the different test statistics for the model specifications. The fourth chapter presents the data and gives preliminary statistics. The empirical results from the estimation of the models and test results are presented and discussed in the fifth chapter. We conclude this paper in the sixth chapter and propose issues for further research.

2 Multivariate GARCH models in hedging

Based on portfolio theory, hedging with futures contracts can be considered as a portfolio allocation problem in which the investor uses futures as an additional asset to be included into the financial asset space of the economy with the purpose of maximizing his utility function [24; 40]. Assuming that the parameter of risk aversion is large enough and expecting no return on the futures position, we obtain the minimum variance hedge as an optimal hedge ratio for the producer. Therefore, when we will further discuss the minimum-variance ratio we are essentially talking about the risk-minimizing hedge. The optimal hedge ratio (OHR) is defined by

$$OHR_t = \frac{\sigma_{t,sf}}{\sigma_{t,f}^2} \quad (1)$$

where $\sigma_{t,sf}$ denotes the covariance between the spot position and futures position and $\sigma_{t,f}^2$ is the variance of the spot position [32].

Over time various methods have been used to estimate the OHR. One of the earliest techniques has been the OLS based estimation. However, many authors have criticized OLS by drawing attention to the inefficiency of the residuals. Among others Herbst, Kare and Marshall [23] argue that the estimation of the minimum variance hedge ratio suffers from serial correlation problem. One of the techniques to solve for the problem of serial correlation and to account for the important role of cointegration between spot and futures prices has been the use of ECM in the estimation of hedge ratios. In particular Ghosh [19; 18], Lien and Luo [34], and Lien [33] have recognized that ignoring error correction mechanisms yields downward biased estimates for hedge ratios.

In addition to the need of ECM to account for cointegration, Park and Bera [43] have pointed out that the simple regression model is inappropriate to estimate hedge ratios because it ignores the heteroscedasticity often encountered in spot and futures price data. Further Myers and Thompson [41] have argued that the hedge ratio should be modelled through conditional moments that depend on information set available at the time of the hedging decision. In practice, this implies that the hedge ratio should be adjusted continuously based on conditional information. In order to account for the volatility dependencies between spot and futures markets, Autoregressive Conditional Heteroscedasticity (ARCH) and generalized (GARCH) models have been applied [13; 7]. As argued by Sim and Zurbruegg [47], the importance of the time-varying nature of the second moment properties cannot be ignored if a serious attempt is to be made

when examining effective hedge ratios over time. Research that has used cointegration and the GARCH framework includes Koutmos and Tucker [30], Kroner and Sultan [32], Kroner and Ng [31].

In this section we will present the most applied GARCH models in hedging. These include the VECH model of Bollerslev, Engle and Wooldridge [6], the constant correlation model of Bollerslev [8], the factor ARCH model of Engle, Ng, and Rothschild [12], and the BEKK model of Engle and Kroner [14]. One of the latest developments in the era is the dynamic correlation GARCH proposed by Engle and Sheppard [16]. In this study four different GARCH models are described and estimated to calculate the optimal hedge ratios. The models chosen for this purpose include the diagonal BEKK, T-BEKK and Asymmetric T-BEKK models, CCORR and its extension DCC-GARCH. The DCC specification will be used only, if the tests against constant correlation recommend it. Given that the spot and futures prices are cointegrated, the estimation is done using bivariate ECM-GARCH setups.

2.1 Diagonal BEKK and T-BEKK

A general multivariate GARCH model for a k -dimensional process $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{kt})$ is given by

$$\varepsilon_t = \mathbf{z}_t \mathbf{H}_t^{1/2} \quad (2)$$

where \mathbf{z}_t is a k -dimensional iid process with mean zero and identity covariance matrix equal. From these properties of \mathbf{z} and (2) it follows that $\mathbf{E}[\varepsilon \varepsilon' | \Omega_t] = \mathbf{H}_t$. To complete the model, a parametrization for the conditional covariance matrix \mathbf{H}_t needs to be specified.

Bollerslev, Engle, and Wooldridge (1988) suggested a multivariate GARCH model, where matrices \mathbf{A}_1 and \mathbf{B}_1 are diagonal

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}' + \mathbf{A}_1(\varepsilon_{t-1}\varepsilon'_{t-1})\mathbf{A}_1' + \mathbf{B}_1\mathbf{H}_{t-1}\mathbf{B}_1' \quad (3)$$

The number of parameters in the diagonal GARCH(1,1) model equals $3(k(k + 1)/2)$. For the bivariate case, setting all off-diagonal elements to zero, it is seen that 9 parameters remain to be estimated. An additional advantage of the diagonal model is that conditions which ensure that the conditional covariance matrix is positive definite are quite easy to check [1]. The model specification is similar both in diagonal BEKK and diagonal T-BEKK. The only difference being that in T-BEKK the estimation is done assuming a t-distribution instead of normal.

The full BEKK model differs from the diagonal representation in that no diagonality constraints are imposed on the matrices \mathbf{A} and \mathbf{B} . and the number of parameters is $2k^2 + k/2$. The key advantage of this model is that no checks are needed to ensure positive definiteness of the covariance matrix.

2.2 CCORR and DCC-GARCH

Bollerslev [8] put forward an alternative way to simplify GARCH-model, by assuming that the conditional correlations between the elements of ε_t are time-invariant. The model is given by

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t \quad (4)$$

where \mathbf{D}_t is a $(k \times k)$ matrix with the conditional standard deviations on the diagonal, and \mathbf{R} is a $(k \times k)$ matrix containing the correlations. The advantage of this model is that it requires only one matrix inversion for each iteration of the non-linear optimization routine, while others require an inversion for each observation.

Engle and Sheppard [16] proposed a new model that both preserves the ease of estimation of Bollerslev's constant correlation model yet allows for non-constant correlations. This class of MGARCH models differs from the other specifications in that univariate GARCH models are estimated for each asset series, and then, using the standardized residuals resulting from the first step, a time-varying correlation matrix is estimated using a simple specification. The proposed dynamic correlation structure is

$$\mathbf{R}_t = \mathbf{Q}_t^{-1/2} \mathbf{Q}_t \mathbf{Q}_t^{*-1} \quad (5)$$

$$\mathbf{Q}_t = (\mathbf{1} - \sum_{m=1}^M \alpha_m - \sum_{n=1}^N \beta_n) \bar{\mathbf{Q}} + \sum_{m=1}^M \alpha_m (\varepsilon_{t-m} \varepsilon'_{t-m}) + \sum_{n=1}^N \beta_n \mathbf{Q}_{t-n} \quad (6)$$

where $\bar{\mathbf{Q}}$ is the unconditional covariance of the standardized residuals resulting from the first stage of estimation and \mathbf{Q}_t^* is diagonal and each element is equal to the square root of the corresponding element of \mathbf{Q}_t . The coefficients α and β are also obtained from the univariate GARCH models.

2.3 Asymmetric T-BEKK

Detection of asymmetries in the test results has called for the ability of models to capture the asymmetric effects explicitly. One of the basic specifications that are nested within the General Dynamic Covariance Model of Kroner and Ng [31] is the Asymmetric T-BEKK. The model is obtained from the standard BEKK by having an additional quadratic form that is dependent on the outer product of the vector of negative return shocks.

By letting $\eta_{it} = \max[0, -\varepsilon_{it}]$ and $\eta_t = [\eta_{1t}, \dots, \eta_{Nt}]$ the Asymmetric BEKK model is defined as

$$\mathbf{H}_t = \mathbf{C} \mathbf{C}' + \mathbf{A} (\varepsilon_{t-1} \varepsilon'_{t-1}) \mathbf{A}' + \mathbf{B} \mathbf{H}_{t-1} \mathbf{B}' + \mathbf{G} \eta_{t-1} \eta'_{t-1} \mathbf{G}' \quad (7)$$

where the \mathbf{A} , \mathbf{B} , \mathbf{C} , and \mathbf{G} are $(k \times k)$ matrices. Shocks on the downside increase the variance, as well as the covariance through the asymmetric term in \mathbf{H}_t . Similar MGARCH models are estimated by Bekaert and Harvey [3], and Bekaert and Wu [4], among others.

3 Tests of MGARCH specifications

This section is divided into two parts. First we will outline some preliminary test statistics that will help in choosing the underlying distributions. Then, in the second part we will consider the tests available for multivariate models. As argued by Kroner and Ng [31], one should not place too much confidence in only statistically insignificant Ljung-Box statistics when evaluating GARCH models, since even badly misspecified models can capture the serial correlation in the second moments and give insignificant Ljung-Box test statistics. In order to test the models, we will use a set of robust conditional moments tests to detect misspecifications introduced by Wooldridge [49]. In particular they will allow us to study whether there are asymmetric effects that should be taken into account. In addition to the conditional moment tests we will consider the DCC-test proposed by Engle and Sheppard [16], which tests the null of constant correlation against the need for dynamic conditional correlation coefficients when modelling dependencies between the series.

3.1 Preliminary checks

In order to evaluate different underlying distributions we consider goodness of fit tests for the marginal distributions and the copulas. The main focus is on the two most applied distribution models: the normal (gaussian) copula and t-distribution. However, before performing these tests, it must be pointed out that the goodness of fit should not be interpreted so that returns are iid with a given probability density function. Rather, it gives that the tested model provides a reasonably good description of the true return distribution and, therefore, its use is not likely to bias the inference about the shape of the asset return distribution.

In this paper, we will perform the tests against normal distribution, Student's t-distribution and the general error distribution (GED). Given the presence of fat tails, the residuals are standardized by estimating GARCH models using the theoretical distribution under null hypothesis. The algorithm to compute the goodness of fit test is well presented by Harris and Stocker [21].

Having tested the marginal distributions for spot and futures returns, the next step is to evaluate the dependence structures. The most widely used dependence structure is that of multivariate normality or multivariate t-distribution. However, as different parametric copulas lead to models that may have completely different dependence properties, it is important in any empirical application to check whether the chosen parametric copula correctly specifies the dependence structure of the multivariate time series regardless of the marginal distributions of individual assets [10]. Therefore, we will apply the consistent test statistics introduced by Chen, Fan, and Patton [10], which test the null of

$$H_0 : P(g(z_1, \dots, z_d) = 1) = 1$$

where $g(z_1, \dots, z_d)$ is the joint density function of the probability integral transformed random variables z_1, \dots, z_d .

Given that the time series $\{\mathbf{Z}_t\} \equiv (Z_{1t}, \dots, Z_{dt})$ is not observable, since the true distribution functions are unknown, Chen et. al. propose construction of pseudo

observations on \mathbf{Z}_t as follows:

$$\hat{Z}_{1t} = \hat{F}_1(Y_{1t}), \hat{Z}_{jt} = C_{0j}(\hat{F}_j(Y_{jt}); \hat{\alpha} \mid \hat{F}_j(Y_{1t}), \dots, \hat{F}_{j-1}(Y_{j-1,t})),$$

where $j = 2, \dots, d$, $t = 1, \dots, n$, C_{0j} denotes the conditional distribution function of \hat{Z}_{jt} given $(\hat{Z}_{1t}, \dots, \hat{Z}_{j-1,t})$ under H_0 , and $\hat{F}_j(Y)$ is the rescaled empirical distribution function of $F_j(Y)$:

$$\hat{F}_j(Y_j) = \frac{1}{n+1} \sum_{t=1}^n \mathbb{I}(Y_{jt} \leq Y_j)$$

The test is based on

$$\hat{I}_n = \int_0^1 \dots \int_0^1 (\hat{g}(z_1, \dots, z_d) - 1)^2 dz_1 \dots dz_d$$

$$\hat{g}(z_1, \dots, z_d) = \frac{1}{nh^d} \sum_{t=1}^n (\prod_{j=1}^d K_h(z_j, \hat{Z}_{jt}))$$

where $\hat{g}(z_1, \dots, z_d)$ is the kernel estimator of $g(z_1, \dots, z_d)$ constructed from the pseudo observations \hat{Z}_{jt} using boundary kernel K_h defined in Hong and Ling [27]. From these Chen et. al construct their $N(0, 1)$ distributed test statistic as

$$T_{nd} = \frac{nh^{d/2} \hat{I}_n - c_{dn}}{c_d} \quad (8)$$

where

$$c_{dn} = h^{d/2} ((h^{-1} - 2) \int_{-1}^1 k^2(w) dw + 2 \int_0^1 \int_{-1}^x k_x^2(y) dy dz)^d$$

$$\sigma_d^2 = 2\sigma_{-1}^1 (\int_{-1}^1 k(u+v) k(v) dv)^{2d} du^d$$

$$k_x(y) = k(y) / \int_{-1}^x k(u) du.$$

3.2 Robust conditional moments tests

Already Engle and Ng [15] have proposed tests to check whether positive and negative shocks have a different impact on the conditional variance. Then the set of tests has been extended specifically for multivariate GARCH models by Kroner and Ng [31]. Recognizing that a major difference between the models is their asymmetric property a beneficial approach is to partition the $(\varepsilon_{it-1} \varepsilon_{jt-1})$ into first four quadrants corresponding to the following sign combinations of $(\varepsilon_{it-1} \varepsilon_{jt-1})$: $(-, -)$, $(-, +)$, $(+, -)$, and $(+, +)$.

Letting $I(\cdot)$ be an indicator function, we get the following set of misspecification indicators that allow us to test for asymmetry in the shocks

$$\begin{aligned} s_{1t-1} &= I(\varepsilon_{it-1} > 0 \mid \varepsilon_{jt-1} > 0) \\ s_{2t-1} &= I(\varepsilon_{it-1} < 0 \mid \varepsilon_{jt-1} > 0) \\ s_{3t-1} &= I(\varepsilon_{it-1} < 0 \mid \varepsilon_{jt-1} < 0) \\ s_{4t-1} &= I(\varepsilon_{it-1} > 0 \mid \varepsilon_{jt-1} < 0) \end{aligned} \quad (9)$$

In addition to these, we can consider the 'sign indicators' as pointed out by Engle and Ng [15] and the heteroskedastic asymmetry indicators of Kroner and Ng [31].

$$\begin{aligned}
s_{5t-1} &= I(\varepsilon_{it-1} < 0) \\
s_{6t-1} &= I(\varepsilon_{jt-1} < 0) \\
s_{7t-1} &= \varepsilon_{it-1}^2 I(\varepsilon_{it-1} < 0) \\
s_{8t-1} &= \varepsilon_{it-1}^2 I(\varepsilon_{jt-1} < 0) \\
s_{9t-1} &= \varepsilon_{jt-1}^2 I(\varepsilon_{it-1} < 0) \\
s_{10t-1} &= \varepsilon_{jt-1}^2 I(\varepsilon_{jt-1} < 0)
\end{aligned} \tag{10}$$

The indicators 5 and 6 denote the 'sign indicators' of Engle and Ng [15] and the indicators from 7 to 10 test for heteroskedastic asymmetry. The reason to include these indicators is that the effect of the size of a shock on the variances and covariances might also depend on the sign of the shock and possibly the sign of other shocks [31].

When only one time series is being tested, only indicators from 5 to 10 are used. To complete the test, the generalized residuals \hat{z}_t^2 are taken as the dependent variable in the regressions, while the partial derivatives of the conditional variance with respect to the parameters in the original GARCH model along with the misspecification indicators are added as regressors. The generalized residual u_{ijt} is defined to be the (i,j)th element of $\varepsilon_t \varepsilon_t' - H_t$, which are further standardized by h_{ijt} to obtain z_{ijt} . The generalized residuals can be interpreted as the distance between the news impact surface and the points on the scatter plot of $\varepsilon_{it} \varepsilon_{jt}$. If the model is correctly specified, the expected value of the generalized residuals is zero.

The robust test statistic used by Kroner and Ng [31] is constructed as

$$\mathbf{C}_{\text{rcm}} = [(1/T) \sum_{t=1}^T u_{ijt} \lambda_{gt-1}]^2 [(1/T) \sum_{t=1}^T u_{ijt}^2 \lambda_{gt-1}^2]^{-1} \tag{11}$$

where λ_{gt-1} is the residual from a regression of the misspecification indicator s_{gt-1} on the derivatives of h_{ijt} with respect to the parameters of the model. The resulting statistic has an asymptotic $\chi^2(1)$ distribution [49].

3.3 DCC-GARCH test for dynamic correlations

As discussed by Engle and Sheppard [16] testing data for constant correlations has proven to be a difficult problem. In particular, testing for dynamic correlation with data that have time-varying volatilities can result in misleading conclusions and rejection of constant correlation when it is true due to misspecified volatility models.

The test proposed by Engle and Sheppard [16] is the following

$$H_0 : \mathbf{R}_t = \bar{\mathbf{R}}$$

$$H_1 : \text{vech}^u(\mathbf{R}_t) = \text{vech}^u(\bar{\mathbf{R}}) + \beta_1 \text{vech}^u(\mathbf{R}_{t-1}) + \cdots + \beta_p \text{vech}^u(\mathbf{R}_{t-p}).$$

where vech^u is a modified vech which selects only elements above the diagonal. The testing procedure is following. First the univariate GARCH processes are estimated,

and then the residuals are standardized. Once this is done, the correlation of the standardized residuals is estimated, and the vector of univariate standardized residuals is jointly standardized by the symmetric square root decomposition of the $\bar{\mathbf{R}}$. Under the null of constant correlation, these residuals should be iid with a variance-covariance matrix given by \mathbf{I}_k . The artificial regressions will be a regression of the outer products of the residuals on a constant and lagged outer products. The vector autoregression is

$$\mathbf{Y}_t = \alpha + \beta_1 \mathbf{Y}_{t-1} + \cdots + \beta_s \mathbf{Y}_{t-s} + \eta_t \quad (12)$$

where $\mathbf{Y}_t = \text{vech}^{\mathbf{u}}[(\bar{\mathbf{R}}^{-1/2} \mathbf{D}_t^{-1} \varepsilon_t)(\bar{\mathbf{R}}^{-1/2} \mathbf{D}_t^{-1} \varepsilon_t) - \mathbf{I}_k]$ and $\bar{\mathbf{R}}^{-1/2} \mathbf{D}_t^{-1} \varepsilon_t$ is a $k \times 1$ vector of residuals jointly standardized under the null. Under the null the intercept and all of the lag parameters in the model is zero.

4 Data

The data set used in this paper consists of 1699 daily Nord Pool closing prices for spots and futures contracts between Jan-1996 and Nov-2002. The insample period for which the estimation and the specifications tests are done covers the first 1400 observations from Jan-1996 to Aug-2001. In order to achieve some comparability with previous studies, we follow the example of Byström [9] by estimating and performing the tests using only one futures strategy. Since the short-term future contracts are more liquid as well as more correlated with the underlying spot, we will select to use futures with three weeks left to maturity. In order to avoid inclusion of thin market and expiration effects, the futures contract is rolled over one week prior to its expiration.

Table 1 reports some descriptive statistics of the price levels and returns on the spot price and the futures strategy (see fig.1). As reflected by the high Jarque-Bera statistics, the returns series are characterized by strong skewness and kurtosis. Quite expectedly also the null hypothesis of no GARCH effect is readily rejected by the Ljung-Box tests on squared returns. Further, we recognize that the two series are cointegrated, which suggests that an ECM specification is appropriate for modelling. For the Johansen procedure, the trace and the maximum eigenvalue test statistics are considered: the first row gives the statistic for the null hypothesis of no cointegrating vectors $r = 0$ for both tests while the second row shows the values associated with the hypothesis $r \leq 1$ for the trace test and $r = 1$ for the maximum eigenvalue test. The Johansen's tests point out that there is a cointegrating vector at 1% significance level. Visual inspection of the sample autocorrelation functions suggests slight serial correlation (fig. 2). The Ljung-Box statistics support this finding.

5 Results

The estimation is done in two steps. First, we estimate the mean equation to get the residuals, and then we estimate the conditional covariance matrix parameters using maximum likelihood, treating the residuals as observable data. The block diagonality of the information matrix under this setup guarantees that consistency and efficiency

are not lost. This follows the two-step approach of Pagan and Schwert [42], Gallant, Rossi, and Tauchen [17], and Engle and Ng [15].

5.1 Error correction model

Using a sequence of LR-ratio tests for different VAR orders, it appears that a 2-lag ECM specification whitens effectively most of the serial correlation in the returns series. This seemed to be the case even though the second lag autocorrelation is negligible. The remaining 12th order autocorrelation in the futures returns series is generated by the roll over strategy used in construction of the futures price series. As the contract is rolled over every two weeks it is likely to generate high order autocorrelation. Therefore we will consider this as a technical effect rather than something we should hunt down by including higher lag orders. Otherwise the test statistics show no surprises and the null of no GARCH effect can be rejected based on Portmanteau test for squared residuals. The estimated model coefficients and descriptive statistics of the non-standardized residuals are presented in Table 2. The order selection matches with previous studies as Byström [9] used a VAR(2) model for the estimation. The inclusion of an ECM term, however, deviates from the approach chosen by Byström, who excluded the error correction mechanism and rather estimated the system as a plain vector autoregression. However, given the highly significant error correction term, it is well justified to use the ECM model in order to avoid downward bias in hedge ratios.

5.2 Univariate diagnostics and goodness-of-fit

The standard way to model conditional variances is to use GARCH(1,1)-filter with different underlying distributions, which provides an easy way for preliminary evaluation. In Table 4. we have computed standard diagnostics for the univariate GARCH specifications obtained from CCORR. The statistics indicate that the null of no asymmetric effects in the conditional variance is rejected based on the sign and size bias tests proposed by Engle and Ng [15]. But overall the simple GARCH(1,1) specification appears to be quite good, since the rest of the statistics are insignificant and the model parameters are relatively stable. Although, a word of caution is necessary. As often with many tests, the power of the statistics is not very high and rejection of the null hypothesis by one or several of the tests does not give much information concerning which nonlinear GARCH model might be the appropriate alternative. The same problem concerns the direct tests against QGARCH and LSTGARCH.

As a next step, we have tested the goodness-of-fit of different underlying distributions for both marginal distributions and their dependence structure. The results are furnished in Table 5. The panel A provides Kolmogorov-Smirnov tests using t-distribution, GED-distribution and normal distribution to estimate GARCH(1,1) for the spot and futures returns separately. The tests indicate that the null hypothesis of normal distribution is strongly rejected, whereas the null of Student's t-distribution or GED-distribution appears to be a more appropriate choice as a marginal distribution. However, when we move on to test the dependence structure between the two series the findings are somewhat different. As represented in panel B, the copula statistics

proposed by Chen et. al [10] suggest that the normal copula or normal-DCC copula, where the correlation structure is dynamic, would be more appropriate than the Student's t-copula. This finding is quite interesting from the practitioner's point of view, since although we would reject joint normality of returns based on Kolmogorov-Smirnov tests, the bivariate normal copula is still a reasonably good model for the dependence structure. Therefore the choice between using exclusively normal distribution or t-distribution is not clear cut in our case. This is one of the reasons why we have estimated the range of multivariate GARCH models using both normal and Student's t-distribution. However, the finding that a normal copula can be still a viable model for dependence structure even though the marginal distributions would be non-normal, is well documented especially in the case of bivariate models. Recently Malevergne and Sornette [37], among others, have presented evidence that the bivariate normal copula hypothesis cannot be rejected for many pairs of equity returns, but when moving to a larger collections of financial assets Mashal and Zeevi [38] find that a multivariate normal copula is more easily rejected in favor of a multivariate Student's t-copula.

5.3 Comparison of MGARCH models

Having tested the univariate GARCH model, we are ready to compare the range of MGARCH models selected for the study. This section is outlined as follows. First we will shortly present the estimation results of the four models specified in the previous section. Then we will consider the summary statistics of the variance and covariance series obtained from the different models.

In the first MGARCH model the spot and futures returns are modelled within the bivariate conditional correlation (CCORR) framework. The conditional variance equations are specified as GARCH(1,1) for the two series. The estimation results for the model are presented in Table 3. The left hand side of the table presents the parameters for the spot variance and on the right for the futures variance. Judging from the highly significant correlation coefficient the spot and futures returns are highly correlated. The correlation is also quite stable, since the DCC-statistics in panel B do not seem to lend much support for a dynamic correlation coefficient. The residual statistics in the lower part of the table show that the first order GARCH model has been sufficient to remove the serial correlation from the squared residuals. However, the fact that we observe high kurtosis in standardized residuals, suggests that a different distributional hypothesis could prove useful.

Having changed the underlying distribution, we have estimated the diagonal T-BEKK, the full T-BEKK and also Asymmetric T-BEKK to consider the potential asymmetric effects. The estimation results of these models are furnished in Table 6. The descriptive statistics below the table confirm that the assumption of GARCH disturbances removes effectively the autocorrelation in the squared residuals as in the CCORR case. The estimated variance, covariance and hedge ratio series are shown in figures 4 to 6. As a next step we will consider the summary statistics of the variance and covariance series obtained from these models. These summary statistics, including the mean, standard deviation, minimum, and maximum, are reported in Table 7. For further intuition, we have reported the correlations between the estimates in Table 8.

The findings from Tables 7. and 8 are aligned with the argument of Kroner and Ng [31] that the multivariate GARCH models give quite different variance and covariance estimates. For futures variances the CCORR estimates are considerably more volatile than the BEKK estimates, but the situation is reversed, when we consider the covariances; the BEKK models produce a broader range of covariance estimates, as evidenced by the large maximum-minimum range. The effect of the t-distribution is pronounced as higher spot variances and lower futures variances when benchmarked against the models using normal distribution. When considering the estimated hedge ratios from different models, we find that the use of t-distribution appears to lead to smaller hedge ratios and higher standard deviation than the normal distribution. Also the full BEKK models give smaller hedge ratios than the diagonal models. The correlations reported in table 8 evidence the differences between normal and t-distribution quite expectedly, but in addition we find that the covariance estimates from the asymmetric T-BEKK and hence also the hedge ratios are negatively correlated with the rest of the models. This is a somewhat odd finding, as it is difficult to design any good explanation why this should be the case.

The results from robust conditional moments tests as suggested by Kroner and Ng [31] are displayed in Table 9. When taking a look at the broader picture, we could say that from a practitioner's point of view the findings are quite good as only a few misspecifications are detected. Further, in terms of these tests there are no major differences between the various models as judged by the statistics computed for spot variance and covariance series. The spot variance statistics furnished in panel A indicate some heteroskedastic asymmetry problems in CCORR, diagonal BEKK and diagonal T-BEKK as they all appear to suffer from the combination, where a negative futures shock occurs with a large spot shock. The covariance statistics in panel B are even better, as all the other models clear out completely except CCORR, which has heteroskedastic asymmetry problem with negative spot shock combined with large futures innovation. Only futures variance seems to suffer from broader misspecifications. All of the models are misspecified with respect to negative spot returns shocks and combination of positive spot and positive futures returns shocks. CCORR has also the heteroskedastic asymmetry problem as in the case with covariances. Additionally the diagonal models and asymmetric T-BEKK suffer from combination, where positive spot return shock occurs with a negative futures shock. Overall, if we were to rank the models by these test statistics, T-BEKK would be the best candidate for further modelling and CCORR the worst. The greatest problem is that all of the models are bad at capturing the asymmetric relationship between futures and shocks to the spot returns.

5.4 Evaluation of hedging performance

Additionally, following the example of Byström [9] we have compared the hedging performance of the different models using both unconditional and conditional measures. The models selected for the outsample tests are CCORR, Diag-BEKK, Diag T-BEKK, and T-BEKK. Given that the robust conditional moments tests that were performed in the previous section indicated no substantial need for inclusion of asymmetric terms,

we have decided to leave out Asym T-BEKK due to the large computational costs involved in the estimation of the model and the apparent lack of benefits over the traditional models as measured by the conditional moments tests.

In the unconditional method the mean and variance of the returns of the hedged portfolios and the percentage reduction in the variance of the hedged portfolio relative to the unhedged portfolio are calculated as

$$\begin{aligned} r_u &= S_{t+1} - S_t \\ r_h &= r_u - h_t(F_{t+1} - F_t) \\ VR &= 1 - \sigma_{r_h}^2 / \sigma_{r_u}^2 \end{aligned}$$

where S_t , F_t , r_u , r_h , and VR denote the logarithmic spot price, logarithmic futures price, return of unhedged portfolio, return of hedged portfolio, and variance reduction, respectively. The conditional variance method proposed by Byström is based on a similar idea. However, we start by assuming that the true return and variances processes are generated by one of the GARCH models under consideration. This gives us the possibility to compare the relative performance of the data generating GARCH and the other models in minimizing the conditional variance. Hedging the spot portfolio each day and then taking the average of the conditional variance of the spot and the hedge portfolio yields us a measure for each portfolio that should be as small as possible.

The test results are summarized in table 10. Starting with the unconditional variances reported in panel A, we find that all the different hedges, except naive hedge reduce the portfolio variance compared to the spot variance. Somewhat surprisingly, the best hedge in both in-sample and out-of-sample seems to be the ECM-hedge. At this point, it would appear that there is not much point in using the elaborate multivariate models as the simple ECM readily achieves equal or even better results. The findings are quite similar to those made by Byström [9] who studied weekly portfolio returns. However, unlike Byström we do not find the naive hedge better than other hedges - on the contrary it appears to consistently increase the portfolio variances. Further we find a clear difference between the in-sample and out-of-sample performance. The models using Student's t-distribution appear to perform better in the out-of-sample than in-sample.

But once we move to inspect the reduction of the conditional instead of the unconditional variances, the obtained messages are quite different. In panel B, the average conditional variance and the variance reductions (both in-sample and out-of-sample) of the different hedges compared to the open spot position are presented for the four different choices of the underlying covariance matrix. As anticipated, we find that the ECM-hedge and naive hedge both increase the portfolio variance in the in-sample and out-of-sample comparisons as they are not able to account for the time variations. Further, it is a stylized fact that those models that are assumed to generate the underlying covariance matrix perform well. For example, when considering the variance reduction achieved by T-BEKK model and assuming that it also generates the covariance matrix, we are not surprised to find that the model also gives the best results as benchmarked against its competitors. Yet, there are some exceptions. An interesting finding is that the diagonal BEKK appears to perform well throughout the comparisons, whereas CCORR is successful only in case we assume that it is the matrix generating model. Generally also Diag T-BEKK and T-BEKK perform well, yet the

Diag BEKK is even better in the outsample tests. However, none of these models was able to achieve conditional variance reductions greater than 4.97%.

Overall, in the performance of these hedge ratios, the in-sample and out-of-sample comparisons tell a similar story. It is evident that the non-hedged spot position is more volatile than the hedged portfolios. Also use of an appropriate hedging model instead of a naive hedge or no hedge at all is important from variance minimizer's perspective. This finding is well inline with Bystrom's study, who finds that the inclusion of heteroskedasticity and volatility clustering in calculating the hedge ratio clearly contributes towards an optimal hedge when looking at the conditional variance. However, the selection of the model to be used in daily hedging, depends on whether we want to minimize the conditional or unconditional portfolio variance. Based on these hedging performance evaluations only, it is difficult to discriminate between the different MGARCH models. Unconditional tests would recommend the use of either ECM-hedge or T-BEKK based on the out-of-sample tests, whereas the conditional tests suggest that Diagonal BEKK, Diagonal T-BEKK, and T-BEKK are almost equally good candidates for daily hedging. In the light of the robust conditional moments tests and other statistics computed in the previous section, these findings are something that could have well been expected. Given, that all of the models appeared to be relatively well specified in terms of the conditional moments tests, it would have been somewhat surprising to find major differences in the hedging performance tests.

6 Conclusions

The selection of appropriate hedging models has been historically a perennial source of debates. Particularly the electricity markets due to asset non-storability and time-varying risk premia, have struggled with the problem of finding a correctly specified model for dynamic hedging. Thus, in order to get a more thorough insight into the model selection, we have chosen to perform a range of tests to evaluate the most commonly applied MGARCH models - including the robust conditional moments tests along with the recently introduced copula statistics of Chen et al. [10]. The first part of the study is concerned with the goodness-of-fit tests and robust conditional moments tests, where the purpose is to examine whether the models are appropriately specified by testing them against the presence of different kinds of asymmetric effects. The second part of the study focuses more on the hedging performance of the different models based on conditional and unconditional immunized portfolio variances. The purpose is to inspect the sensitivity of hedging results to model selection.

Starting with the goodness-of-fit tests for different underlying distribution candidates and copula statistics, we made a few interesting findings that could guide further studies in the area. These were issues concerned with the choice of underlying marginal distributions and their dependence structure. The problem with the traditional multivariate GARCH models that we have studied so far is that they compel us to use exclusively one distribution to model the whole covariance matrix even though a combinatorial approach could be preferable. In this study we have many times rejected the assumption of normality of the spot and futures series, yet still accepted the use

of normal copula to be a more appropriate choice for the dependence structure than the Student's t-copula. Although, also a t-copula appeared to be adequately good for modelling the dependence structure. Thus, the problem we have faced in the use of the multivariate GARCH models selected for this study, is that all of them are misspecified in some respect: when using CCORR and diagonal BEKK we have misspecified the marginal distributions by assuming them to be normal, but when using diagonal T-BEKK, T-BEKK or asymmetric T-BEKK we have problems with the dependence structure. The model that could prove more effective would be one that combines the best of the both sides: marginal Student's t-distributions with normal copula as the dependence structure. However, having made the decision to focus on the commonly applied MGARCH models, we decided to continue with a collection of models using both gaussian and Student's t-copulas.

The second step was to evaluate the models in the spirit of Kroner and Ng [31] using the robust conditional moments tests for potential asymmetries. In general the findings were good from a practitioner's point of view, since all of the models studied appeared to be relatively well specified. Only futures variance specifications suffered from some asymmetric effects. All of the models appeared to be misspecified with respect to negative spot return shocks and a combination of positive spot and positive futures return shocks. T-BEKK appeared to have the least number of rejections. The diagonal models suffered from the combination where positive spot return shocks occur with negative futures return shocks, and CCORR had heteroskedastic asymmetry problems also with the covariance specification. Overall the tests implied surprisingly weak evidence on unaccounted asymmetries, which suggests that although the models fail to capture some asymmetric responses to shocks the need to incorporate these effects into models is not too acute.

The third step of the study, was to consider the hedging performance differences by comparing the conditional and unconditional immunized portfolio variances. As Byström [9], we find that when comparing the results from the conditional evaluation with the results from the unconditional evaluation, there are both similarities and differences. In both cases, it is quite obvious how hedging in the electricity markets can reduce the variance, but when conditional metrics are used the inclusion of the ARCH effects in calculating the optimal hedge ratio clearly suggests that GARCH models contribute towards the optimal hedge ratios. The evidence was stacked against the naive hedge strategy throughout the comparison and also the simple ECM-hedge and CCORR achieved weaker results than the Diagonal BEKK models and T-BEKK. However, the selection between MGARCH and ECM-hedges depends on whether we want to minimize the conditional or unconditional variance. If conditional variance is considered, the MGARCH approach should be preferred - but if only unconditional variance is considered then the simple ECM-hedge would best out all the rest of the models. Overall, it has shown to be quite difficult to say whether any particular MGARCH model should be preferred. Rather we have been surprised by how little differences the tests have managed to reveal between the models. For a practitioner the message is positive: one just needs to choose the simplest model that is not too badly specified in terms of the conditional moments tests. In this case the choice would be between the Diagonal specifications and the full T-BEKK - each of them yielding

relatively similar hedging benefits.

Yet, for a researcher there is still a lot of work to do in the area. One major issue is the problem of finding an appropriate underlying distribution. The tests performed in this study have readily indicated that neither the gaussian or Student's t-distribution appear to provide full fit. Rather than being a problem of a technical model specification as studied by the robust conditional moments tests, the bigger problem is in the selection of the marginal distributions and their dependence structure. Finally, given that the whole study was flooded with different specifications tests and diagnostics, it became quite clear that a perennial source of new research issues is the design and comparison of new techniques. In particular, the analysis of multivariate non-linear models has been taken up only very recently, and currently there are no generally accepted ideas on how to construct and test such models in the first place. Thus, it seems that much further research is needed on issues as representation, specification, estimation, inference and forecasting for these models.

References

- [1] O. Attanasio. Risk, time varying second moments and market efficiency. *Review of Economic Studies*, 58:479–494, 1991.
- [2] G. Barz and B. Johnson. Modeling the prices of commodities that are costly to store: the case of electricity. Proceedings of the Chicago Risk Management Conference, 5 1998.
- [3] G. Bekaert and C. Harvey. Emerging equity market volatility. *Journal of Financial Economics*, 43:29–77, 1997.
- [4] G. Bekaert and G. Wu. Asymmetric volatility and risk in equity markets. *Review of Financial Studies*, 13:1–42, 2000.
- [5] S. Benninga and C. Oosterhof. Hedging with forwards and puts in complete and incomplete markets. working paper, 2001.
- [6] Engle R.F. Bollerslev, T. and J.M. Wooldridge. A capital asset pricing model with time-varying covariances. *Journal of Econometrics*, 96:116, 1988.
- [7] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.
- [8] T. Bollerslev. Modeling the coherence in short-run nominal exchange rates: A multivariate generalized arch approach. *Review of Economics and Statistics*, 72:498, 1990.
- [9] H. Bystrom. The hedging performance of electricity futures on the nordic power exchange nord pool. *Applied Economics*, 35:1–11, 2003.

- [10] Fan Y. Patton A. Chen, X. Simple tests for models of dependence between multiple financial time series, with applications to u.s. equity returns and exchange rates. working paper, 2004.
- [11] L. Clewlow and C. Strickland. *Energy Derivatives: Pricing and Risk management*. Lacima Publications, 2000.
- [12] Ng V. Engle, R. and M. Rothschild. Asset pricing with a factor-arch structure: empirical estimates for treasury bills. *Journal of Econometrics*, 45:213–237, 1990.
- [13] R. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom. *Econometrica*, 50:987–1007, 1982.
- [14] R. Engle and K. Kroner. Multivariate simultaneous generalized arch. *Econometric Theory*, 11:122–150, 1995.
- [15] R. Engle and V. Ng. Measuring and testing the impact of news on volatility. *Journal of Finance*, 48:1749–1778, 1993.
- [16] R. Engle and K. Sheppard. Theoretical and empirical properties of dynamic conditional correlation multivariate garch. working paper, Department of Economics, UCSD, 2001.
- [17] Rossi P. Tauchen G. Gallant, R. Stock prices and volume. *Review of Financial Studies*, 5:199–242, 1992.
- [18] A. Ghosh. Cointegration and error correction models: Intertemporal causality between index and futures prices. *The Journal of Futures Markets*, 13:193–198, 1993.
- [19] A. Ghosh. Hedging with stock index futures: Estimation and forecasting with error correction model. *The Journal of Futures Markets*, 13:743–752, 1993.
- [20] G. Hagerud. *A new non-linear GARCH model*. PhD thesis, Stockholm School of Economics, 1997.
- [21] J. Harris and H. Stocker. *Handbook of Mathematics and Computational Science*. Springer, New York, 1998.
- [22] J. M. Harrison and S. R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their applications*, 11:215–260, 1981.
- [23] Kare D. Herbst, A. and J. Marshall. A time varying convergence adjusted minimum risk futures hedge ratio. *Advances in Futures and Options Research*, 6:137–155, 1989.
- [24] J. Hirshleifer. *Price theory and applications*. Prentice-Hall, Englewood Cliffs, N.J., 4 edition, 1988.

- [25] P. Holmes. Ex ante hedge ratios and the hedging effectiveness of the ftse-100 stock index futures contract. *Applied Economics Letters*, 2:56–59, 1995.
- [26] P. Holmes. Stock index futures hedging: Hedge ratio estimation, duration effects, expiration effects and hedge ratio stability. *Journal of Business Finance and Accounting*, 23:63–77, 1996.
- [27] Y. Hong and H. Li. Nonparametric specification testing for continuous-time models with applications to spot interest rates. mimeo, Cornell University, 2002.
- [28] R. Huisman and R. Mahieu. Regime jumps in electricity prices. working paper, Erasmus Research Institute of Management (ERIM), Erasmus University Rotterdam, 2001.
- [29] V. Kaminski. The challenge of pricing and risk managing electricity derivatives. *The US Power Market*, pages 149–171, 1997. Risk Publications.
- [30] G. Koutmos and M. Tucker. Temporal relationships and dynamic interactions between spot and futures stock markets. *The Journal of Futures Markets*, 16:55–69, 1996.
- [31] K. Kroner and V. Ng. Modeling asymmetric comovements of asset returns. *Review of Financial Studies*, 11:817–844, 1998.
- [32] K. Kroner and J. Sultan. Time-varying distributions and dynamic hedging with foreign currency futures. *Journal of Financial and Quantitative Analysis*, 28:535–551, 1993.
- [33] D. Lien. The effect of the cointegration relationship on futures hedging: A note. *The Journal of Futures Markets*, 16:773–780, 1996.
- [34] D. Lien and X. Luo. Multi-period hedging in the presence of conditional heteroskedasticity. *The Journal of Futures Markets*, 14:927–955, 1994.
- [35] M. Lindahl. Minimum variance hedge ratios for stock index futures: Duration and expiration effects. *The Journal of Futures Markets*, 12:33–53, 1992.
- [36] S. Lundbergh and T. Terasvirta. Evaluating garch models. *Working papers in Economics and Finance*, 292, 1998b. Stockholm School of Economics.
- [37] Y. Malevergne and D. Sornette. Testing the gaussian copula hypothesis for financial assets dependencies. *Quantitative Finance*, 3:231–250, 2003.
- [38] R. Mashal and A. Zeevi. Beyond correlation: Extreme co-movements between financial assets. manuscript, Columbia University, 2002.
- [39] M. Musiela and M. Rutkowski. *Martingale methods in financial modeling*. Springer-Verlag, 1997.

- [40] R. Myers. Estimating time varying optimal hedge ratios on futures markets. *Journal of Futures Markets*, 11:39–53, 1991.
- [41] R. Myers and S. Thompson. Generalized optimal hedge ratio estimation. *American Journal of Agricultural Economics*, 71:858–868, 1989.
- [42] A. Pagan and G. Schwert. Alternative models for conditional stock volatility. *Journal of Econometrics*, 45:267–290, 1990.
- [43] H. Park and A. Bera. Interest rate volatility, basis, and heteroskedasticity in hedging mortgages. *The American Real Estate and Urban Economics Association*, 15:79–97, 1987.
- [44] A. Peck. Futures markets, supply response and price stability. *Quarterly Journal of Economics*, 90:497–423, 1976.
- [45] W. Purcell and M. Hudson. The economic roles and implications of trade in livestock futures. in *Futures Markets: Regulatory Issues*, A.E. Peck, ed., American Enterprise Institute for Public Policy Research, Washington, 1985.
- [46] E. Rossi and C. Zucca. Hedging interest rate risk with multivariate garch. *Applied Financial Economics*, 6:121–148, 2002.
- [47] A. Sim and R. Zurbruegg. Optimal hedge ratios and alternative hedging strategies in the presence of cointegrated time-varying risks. *The European Journal of Finance*, 7:269–283, 2001.
- [48] J. Skadberg and G. Futrell. An economic appraisal of futures trading in livestock. *Journal of Farm Economics*, 48:1485–1489, 1966.
- [49] J. Wooldridge. A unified approach to robust, regression-based specification tests. *Econometric Theory*, 6:17–43, 1990.
- [50] J. Yang and T. Awokuse. Asset storability and hedging effectiveness in commodity futures markets. working paper, University of Delaware, 2002.

Table 1: Descriptive statistics.

The panel A reports summary statistics and unit root tests of the logarithmic spot and futures prices for the full sample period between Jan-1996 and Nov-2002. ADF and KPSS denote the Augmented Dickey-Fuller test and Kwiatkowski-Phillips-Schmidt-Shin test, respectively. The trend is included in the tests. ADF tests the null hypothesis of a unit root, whereas KPSS tests the null of stationarity. The critical values for ADF are -4.0 and -3.5 at 1% and 5% levels. The corresponding critical values for KPSS are 0.216 and 0.146 for 1% and 5% levels. $Q(k)$ is the Ljung-Box test for autocorrelation. 99% critical value for Jarque-Bera is 9.21. The panel B reports the Johansen cointegration test for spots and futures. For the Johansen's trace test and the eigenvalue test, the first row tests the null of no cointegration and the second row tests the null of one cointegrating vector. The lag orders used in the unit root and cointegration tests were selected using a sequence of Lagrange ratio tests for VARs of different orders. Italic-bold denotes rejection at 1% significance level.

Panel A: Descriptive statistics and unit root tests				
	Spot		Futures	
	Price levels	Returns	Price levels	Returns
Mean	5.973	0.000	4.974	0.000
Standard deviation	0.397	0.082	0.038	0.038
Skewness	-0.159	1.216	0.092	1.154
Excess kurtosis	0.153	37.176	-0.551	11.425
Jarque-Bera Test	8.839	98200.630	23.884	9612.183
$Q(4)$	—	32.456	—	7.094
$Q^2(4)$	—	289.409	—	6.136
$Q(12)$	—	40.328	—	54.663
$Q^2(12)$	—	289.877	—	268.687
ADF	-3.907	-48.734	-1.507	-47.975
KPSS	4.063	0.014	6.869	0.047

Panel B: Cointegration tests		
	Eigenvalue test	Trace test
r=0	113.337	116.621
r=1	3.284	3.284

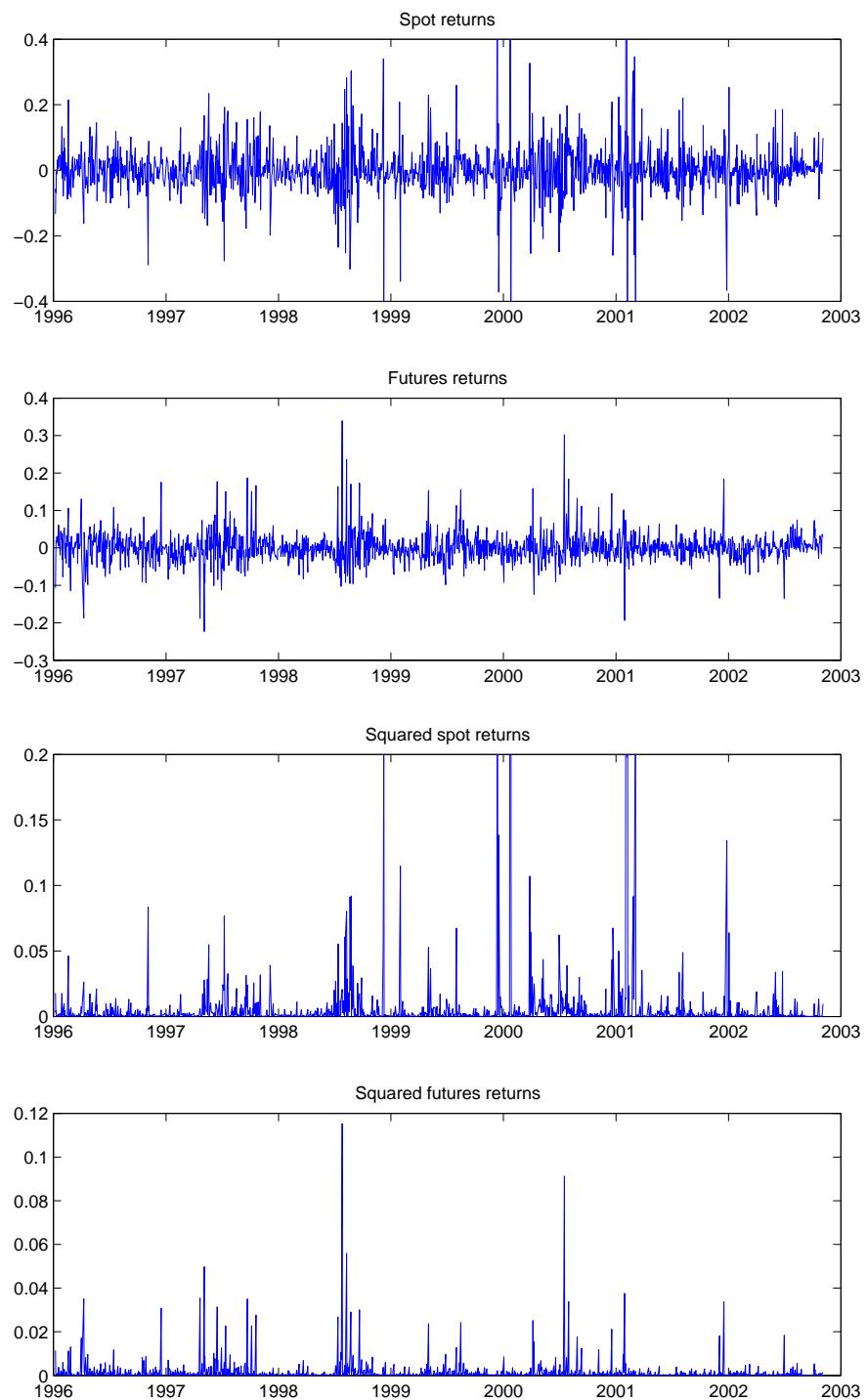


Figure 1: Spot and futures returns

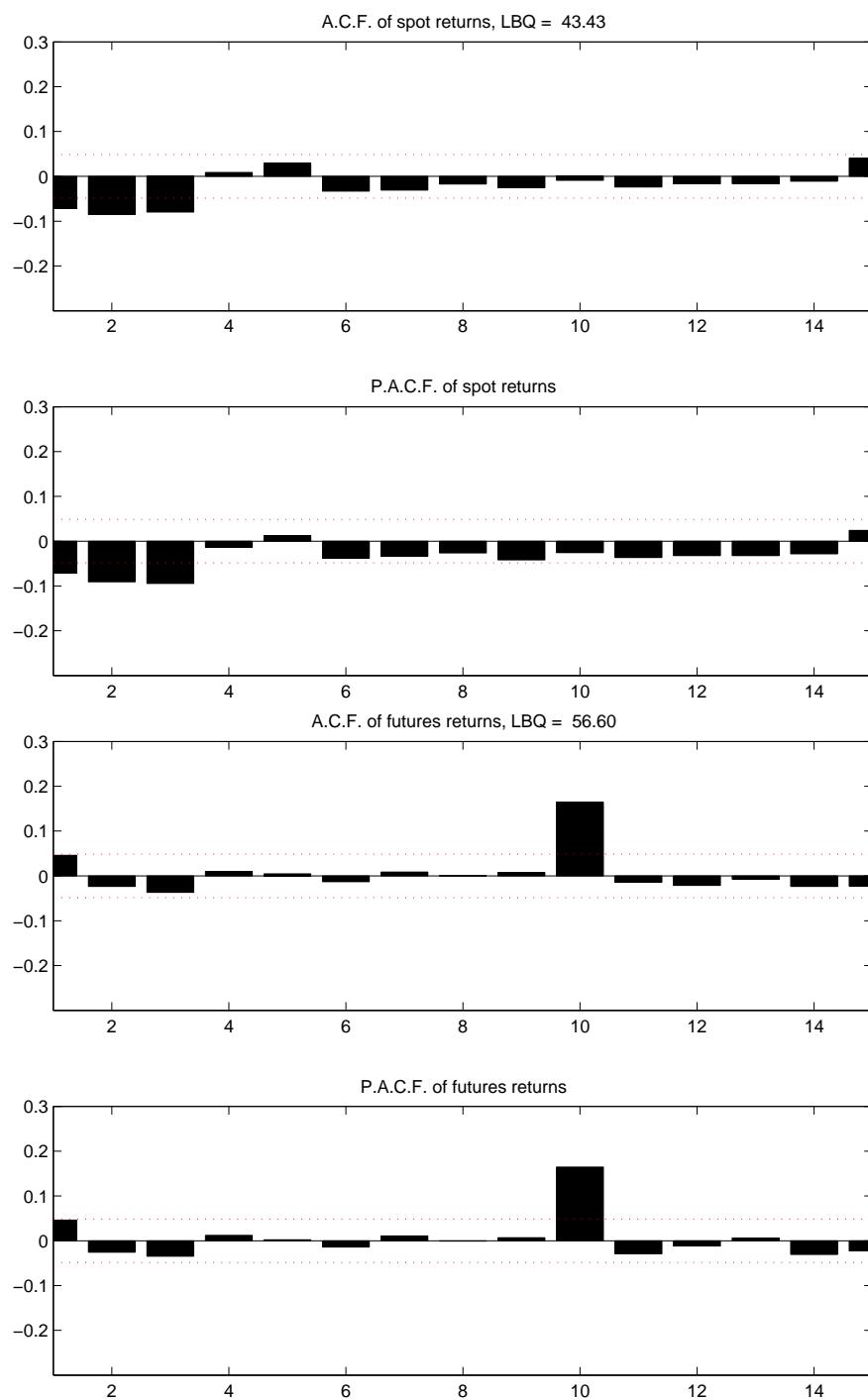


Figure 2: Autocorrelations

Table 2: ECM model estimates

The table reports the estimates of the ECM(2) for the mean in daily spot and futures prices (the first 1400 observations: between Jan-1996 and Aug-2001). The futures strategy is based on buying contracts with three weeks left to maturity and rolling over one week prior to expiration. $Q(k)$ denotes the Ljung-Box test for the first k lags. The model order was selected using a series of Lagrange ratio tests. $Q^2(k)$ is the Ljung-Box test for the squared residuals. 99% critical value for Jarque-Bera normality test is 9.21. Italic-bold denotes rejection at 1% significance level.

Variable	Spot		Futures	
	Estimates	t statist.	Estimates	t statist.
$\Delta \ln S_{t-1}$	-0.052	-1.894	-0.014	-1.064
$\Delta \ln S_{t-2}$	-0.042	-1.613	-0.003	-0.227
$\Delta \ln F_{t-1}$	0.372	6.400	0.029	1.023
$\Delta \ln F_{t-2}$	0.019	0.329	-0.033	-1.147
EC term	-0.021	-9.757	-0.002	-2.196
Constant	0.002	0.900	0.000	0.285
Mean	0.000		0.000	
Stand.dev	0.082		0.040	
Skewness	2.313		1.136	
Excess kurtosis	39.115		10.292	
Jarque-Bera Test	90301.857		6465.328	
$Q(4)$	2.730		4.495	
$Q^2(4)$	121.390		5.730	
$Q(12)$	7.531		57.212	
$Q^2(12)$	121.685		213.697	

Table 3: CCORR and DCC-MGARCH test

The table presents the maximum likelihood parameter estimates for the constant correlation model. The lower part of the table gives standardized residual statistics. Log L denotes the loglikelihood value and Q_k^2 denotes the Ljung-Box test for the remaining serial correlation in the squared residuals. 99% critical value for Ljung-Box is 16.8 and for the Lilliefors normality test 0.0276. The DCC-MGARCH test statistic is χ^2 distributed with nlags+1 degrees of freedom. The pval denotes the probability that the correlation is constant.

Panel A: CCORR				
	Spot		Futures	
	Estimates	t statist.	Estimates	t statist.
ω_{ii}	0.003	15.962	0.000	5096.327
α_{ii}	0.307	0.530	0.100	159.916
β_{ii}	0.335	0.074	0.867	546.339
ρ_{ij}	0.172	54.691	0.172	54.691
$\text{Log}L$	4315.76			
Mean	0.019		-0.010	
Skewness	4.980		0.912	
Excess Kurtosis	79.242		9.300	
Lilliefors Test	0.12		0.09	
$Q(6)$	0.06		4.57	
$Q^2(6)$	0.01		0.16	

Panel B: Test against DCC-MGARCH		
Number of lags	Statistic	P-value
1	0.252	0.882
3	1.087	0.896
6	3.472	0.838
9	16.792	0.079
12	17.352	0.184
15	19.843	0.227

Table 4: Univariate GARCH diagnostics

The table gives diagnostic tests for the univariate GARCH specifications obtained from CCORR. The ARCH and GARCH tests denote the LM-test proposed by Lundbergh and Terasvirta [36]. The sign and size bias tests of standardized residuals are based on Engle and Ng [15]. QGARCH and LSTGARCH denote the direct tests against these alternatives as proposed by Hagerud [20]. The parameters constancy tests denote the Lundbergh and Terasvirta [36] tests against ANST-GARCH.

	Spot		Futures	
	Statistic	P-value	Statistic	P-value
No remaining ARCH	0.022	0.883	1.505	0.220
Higher order GARCH	0.043	0.835	1.347	0.246
GARCH vs. QGARCH	0.221	0.638	2.089	0.148
GARCH vs. LSTGARCH	0.078	0.780	2.147	0.143
Intercept stability	2.292	0.130	1.673	0.196
ARCH parameter stability	0.890	0.346	0.926	0.336
All parameters stability	2.418	0.490	1.955	0.582
Sign Bias	-1.775	0.038	-1.587	0.056
Positive Size Bias	44.434	0.000	31.919	0.000
Negative Size Bias	-2.554	0.005	-13.036	0.000
Sign and Size Bias	960.106	0.000	1018.965	0.000

Table 5: Goodness-of-fit tests for distributions

K-S T-dist, K-S GED-dist and K-S Normal-dist statistics represent the Kolmogorov-Smirnov tests the null that the errors are from the given distribution. The $N(0,1)$ distributed copula statistics are based on the work of Chen et. al. [10].

Panel A: Marginal distribution		
	Spot	Futures
K-S T-dist	0.035 (0.059)	0.030 (0.173)
K-S GED-dist	0.037 (0.047)	0.032 (0.106)
K-S Normal-dist	0.118 (0.000)	0.093 (0.000)

Panel B: Copulas		
	Raw errors	GARCH-filter
Normal copula	0.427 (0.335)	-0.716 (0.237)
Student's t-copula	-1.785 (0.037)	-1.218 (0.112)
Normal-DCC copula	0.601 (0.274)	-0.445 (0.328)

Table 6: BEKK models

Panel A: Parameter estimates of alternative models								
	Diag BEKK	Diag T-BEKK		T-BEKK		Asym	T-BEKK	
ω_{11}	0.050 (0.000)	0.030 (0.005)		0.031 (0.003)		0.029 (0.000)		
ω_{21}	0.003 (0.000)	0.003 (0.000)		0.000 (0.000)		-0.001 (0.000)		
ω_{22}	0.006 (0.000)	0.005 (0.000)		0.006 (0.000)		0.007 (0.000)		
α_{11}	0.524 (0.008)	0.636 (1.970)		0.682 (1.069)		0.640 (0.003)		
α_{12}	- -	- -		- -0.343		-0.599 (0.006)		
α_{21}	- -	- -		- 0.002		-0.021 (0.000)		
α_{22}	0.253 (0.001)	0.188 (0.134)		0.177 (0.059)		-0.141 (0.000)		
β_{11}	0.596 (0.009)	0.717 (0.020)		0.683 (0.009)		0.683 (0.000)		
β_{12}	- -	- -		- 0.095		0.080 (0.001)		
β_{21}	- -	- -		- 0.002		0.012 (0.000)		
β_{22}	0.953 (0.000)	0.973 (0.000)		0.975 (0.000)		0.965 (0.000)		
g_{11}	- -	- -		- -		- 0.286		(0.002)
g_{12}	- -	- -		- -		- -0.286		(0.004)
g_{21}	- -	- -		- -		- -0.008		(0.000)
g_{22}	- -	- -		- -		- 0.170		(0.000)
ν	- -	- 3.176		(98.975)		12.261 (51.115)		3.168 (15.885)
LogL	4305.276		4911.540		4916.935		4924.494	

Panel B: Diagnostics								
	Spot	Futures	Spot	Futures	Spot	Futures	Spot	Futures
$Q(1)$	0.000	0.749	0.015	0.207	0.013	0.151	0.006	0.150
$Q(2)$	0.018	0.827	0.048	1.166	0.042	1.330	0.033	0.667
$Q(5)$	0.051	3.601	0.091	4.644	0.079	5.027	0.070	3.803
$Q(10)$	0.125	51.544	0.193	73.431	0.166	76.937	0.154	66.161
$Q^2(1)$	0.001	0.037	0.001	0.033	0.001	0.031	0.001	0.027
$Q^2(2)$	0.003	0.044	0.002	0.033	0.002	0.031	0.002	0.030
$Q^2(5)$	0.006	0.118	0.005	0.109	0.005	0.106	0.004	0.091
$Q^2(10)$	0.013	2.614	0.010	7.077	0.009	8.061	0.009	5.368
AIC		-8596.551		-9807.079		-9809.871		-9816.987
SIC		-8559.856		-9765.143		-9746.966		-9733.114

AIC = $-2 \log L + 2K$, where K is the number of estimated parameters and $\log L$ is the log-likelihood function.

SIC = $-2 \log L + K \log(T)$, where $T = 1517$ is the sample size and K is the number of estimated parameters and $\log L$ is the log-likelihood function

Table 7: Summary statistics of variances, covariances and hedge ratios

This table gives summary statistics for the variance, covariance and hedge ratio series estimated from the different MGARCH models discussed in the article. All models have been applied to the same dataset between Jan-1996 and Aug-2001. ε_{1t} is the spot return innovation and ε_{2t} is the innovation to the futures return. h_{11t} , h_{22t} , and h_{12t} denote the estimated spot variance, futures variance and their covariance, respectively.

Model	Variable	Mean	SD	Minimum	Maximum
Spot variance					
CCORR	ε_{1t}^2	0.007	0.015	0.004	0.395
Diag BEKK	h_{11t}	0.007	0.014	0.004	0.354
Diag T-BEKK	h_{11t}	0.007	0.023	0.002	0.518
T-BEKK	h_{11t}	0.008	0.025	0.002	0.582
Asym T-BEKK	h_{11t}	0.008	0.025	0.002	0.505
Futures variance					
CCORR	ε_{2t}^2	0.002	0.002	0.001	0.035
Diag BEKK	h_{22t}	0.002	0.001	0.001	0.010
Diag T-BEKK	h_{22t}	0.002	0.001	0.001	0.006
T-BEKK	h_{22t}	0.002	0.001	0.001	0.006
Asym T-BEKK	h_{22t}	0.002	0.001	0.001	0.008
Covariance					
CCORR	ε_{12t}	0.001	0.000	0.000	0.006
Diag BEKK	h_{12t}	0.001	0.001	-0.005	0.010
Diag T-BEKK	h_{12t}	0.001	0.001	-0.005	0.010
T-BEKK	h_{12t}	0.001	0.001	-0.009	0.007
Asym T-BEKK	h_{12t}	0.001	0.001	-0.018	0.010
Hedge ratios					
CCORR	$\varepsilon_{12t}/\varepsilon_{2t}$	0.352	0.165	0.080	3.342
Diag BEKK	h_{12t}/h_{2t}	0.379	0.318	-2.042	3.898
Diag T-BEKK	h_{12t}/h_{2t}	0.315	0.314	-1.901	2.981
T-BEKK	h_{12t}/h_{2t}	0.297	0.349	-2.330	4.267
Asym T-BEKK	h_{12t}/h_{2t}	0.268	0.499	-10.296	3.553

Table 8: Correlations

This table gives correlations between the variance, covariance and hedge ratio series estimated from the different MGARCH models discussed in the article. All models have been applied to the same dataset between Jan-1996 and Aug-2001.

Spot variance		CCORR	Diag BEKK	Diag T-BEKK	T-BEKK	Asym T-BEKK
CCORR	1.000	0.999	0.977	0.983	0.955	
Diag BEKK	0.999	1.000	0.982	0.987	0.961	
Diag T-BEKK	0.977	0.982	1.000	0.995	0.983	
T-BEKK	0.983	0.987	0.995	1.000	0.990	
Asym T-BEKK	0.955	0.961	0.983	0.990	1.000	
Futures variance		CCORR	Diag BEKK	Diag T-BEKK	T-BEKK	Asym T-BEKK
CCORR	1.000	0.586	0.538	0.527	0.527	
Diag BEKK	0.586	1.000	0.968	0.955	0.949	
Diag T-BEKK	0.537	0.968	1.000	0.998	0.978	
T-BEKK	0.527	0.955	0.998	1.000	0.976	
Asym T-BEKK	0.527	0.949	0.978	0.976	1.000	
Covariance		CCORR	Diag BEKK	Diag T-BEKK	T-BEKK	Asym T-BEKK
CCORR	1.000	0.307	0.322	0.218	-0.192	
Diag BEKK	0.307	1.000	0.977	0.803	-0.353	
Diag T-BEKK	0.322	0.977	1.000	0.815	-0.321	
T-BEKK	0.218	0.803	0.815	1.000	-0.604	
Asym T-BEKK	-0.192	-0.353	-0.321	-0.604	1.000	
Hedge ratios		CCORR	Diag BEKK	Diag T-BEKK	T-BEKK	Asym T-BEKK
CCORR	1.000	0.349	0.247	0.470	-0.783	
Diag BEKK	0.349	1.000	0.937	0.847	-0.565	
Diag T-BEKK	0.247	0.937	1.000	0.864	-0.552	
T-BEKK	0.470	0.847	0.864	1.000	-0.780	
Asym T-BEKK	-0.783	-0.565	-0.552	-0.780	1.000	

Table 9: Robust Conditional Moments tests

Panel A: Spot variance					
	CCORR	Diag BEKK	Diag T-BEKK	T-BEKK	Asym T-BEKK
$I(\varepsilon_{1t-1} < 0)$	0.546	0.524	0.794	0.774	0.830
$I(\varepsilon_{2t-1} < 0)$	1.866	1.913	1.517	1.204	1.039
$I(\varepsilon_{1t-1} < 0; \varepsilon_{2t-1} < 0)$	2.110	2.141	1.955	1.119	1.121
$I(\varepsilon_{1t-1} < 0; \varepsilon_{2t-1} > 0)$	0.015	0.024	0.053	0.253	0.238
$I(\varepsilon_{1t-1} > 0; \varepsilon_{2t-1} < 0)$	1.367	1.421	0.920	1.218	0.752
$I(\varepsilon_{1t-1} > 0; \varepsilon_{2t-1} > 0)$	0.818	0.813	0.877	0.931	0.854
$\varepsilon_{1t-1}^2 I(\varepsilon_{1t-1} < 0)$	0.024	0.015	0.732	0.398	0.189
$\varepsilon_{1t-1}^2 I(\varepsilon_{2t-1} < 0)$	3.359	3.005	3.116	1.326	0.558
$\varepsilon_{2t-1}^2 I(\varepsilon_{1t-1} < 0)$	2.082	2.107	0.367	0.076	0.023
$\varepsilon_{2t-1}^2 I(\varepsilon_{2t-1} < 0)$	0.004	0.012	0.132	1.466	0.514
Panel B: Covariance					
	CCORR	Diag BEKK	Diag T-BEKK	T-BEKK	Asym T-BEKK
$I(\varepsilon_{1t-1} < 0)$	0.104	0.771	0.002	1.026	1.081
$I(\varepsilon_{2t-1} < 0)$	0.210	1.132	0.305	1.088	1.219
$I(\varepsilon_{1t-1} < 0; \varepsilon_{2t-1} < 0)$	0.009	2.266	1.102	0.837	1.244
$I(\varepsilon_{1t-1} < 0; \varepsilon_{2t-1} > 0)$	0.168	1.116	0.319	1.060	0.288
$I(\varepsilon_{1t-1} > 0; \varepsilon_{2t-1} < 0)$	0.286	0.784	0.001	1.011	1.281
$I(\varepsilon_{1t-1} > 0; \varepsilon_{2t-1} > 0)$	0.027	0.022	0.000	0.112	0.992
$\varepsilon_{1t-1}^2 I(\varepsilon_{1t-1} < 0)$	0.007	1.534	0.705	0.002	0.865
$\varepsilon_{1t-1}^2 I(\varepsilon_{2t-1} < 0)$	0.272	1.437	0.323	0.095	0.172
$\varepsilon_{2t-1}^2 I(\varepsilon_{1t-1} < 0)$	3.362	0.617	1.504	0.120	0.423
$\varepsilon_{2t-1}^2 I(\varepsilon_{2t-1} < 0)$	0.195	0.826	1.189	0.846	0.402
Panel C: Futures variance					
	CCORR	Diag BEKK	Diag T-BEKK	T-BEKK	Asym T-BEKK
$I(\varepsilon_{1t-1} < 0)$	5.524	5.708	6.492	6.379	6.870
$I(\varepsilon_{2t-1} < 0)$	0.000	0.013	0.010	0.006	0.000
$I(\varepsilon_{1t-1} < 0; \varepsilon_{2t-1} < 0)$	1.058	1.034	1.397	1.293	1.525
$I(\varepsilon_{1t-1} < 0; \varepsilon_{2t-1} > 0)$	1.343	1.503	1.640	2.056	1.541
$I(\varepsilon_{1t-1} > 0; \varepsilon_{2t-1} < 0)$	2.574	3.108	4.007	2.566	5.259
$I(\varepsilon_{1t-1} > 0; \varepsilon_{2t-1} > 0)$	3.746	3.185	2.911	3.735	2.942
$\varepsilon_{1t-1}^2 I(\varepsilon_{1t-1} < 0)$	0.464	0.310	0.367	0.123	0.240
$\varepsilon_{1t-1}^2 I(\varepsilon_{2t-1} < 0)$	1.977	1.627	1.752	1.817	2.233
$\varepsilon_{2t-1}^2 I(\varepsilon_{1t-1} < 0)$	6.382	1.418	1.165	1.573	3.862
$\varepsilon_{2t-1}^2 I(\varepsilon_{2t-1} < 0)$	0.126	0.954	0.168	0.001	0.000

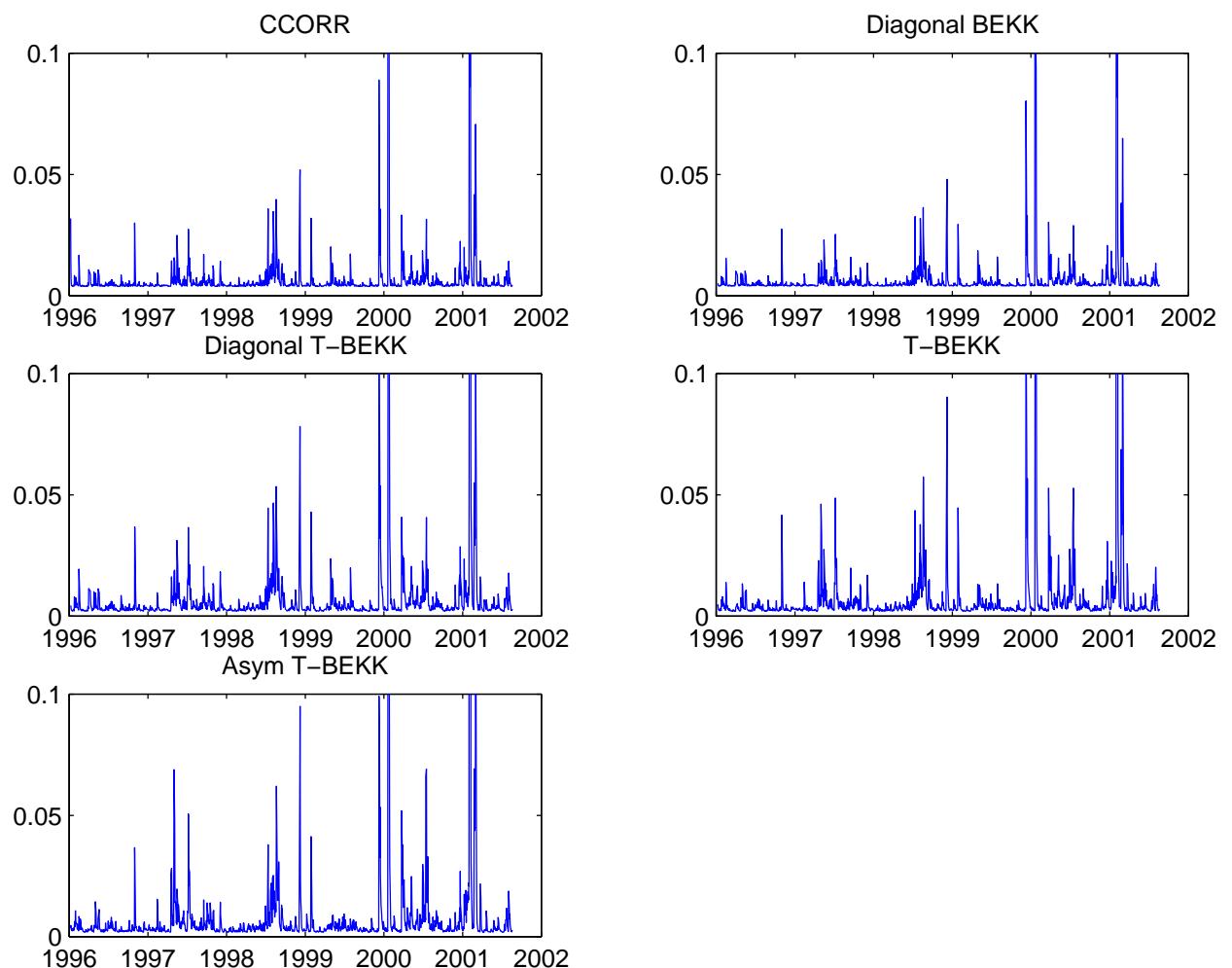


Figure 3: Spot variances

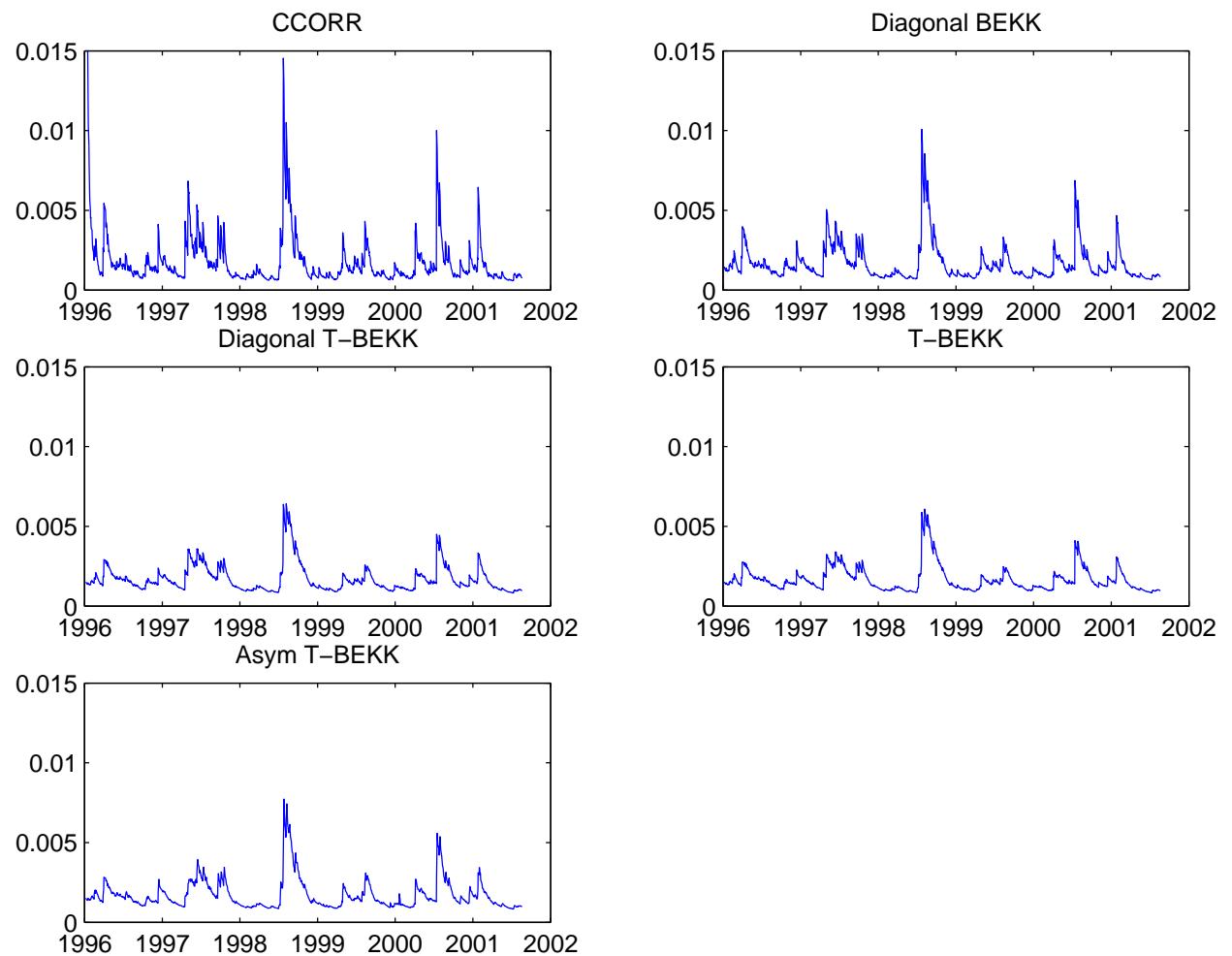


Figure 4: Futures variances

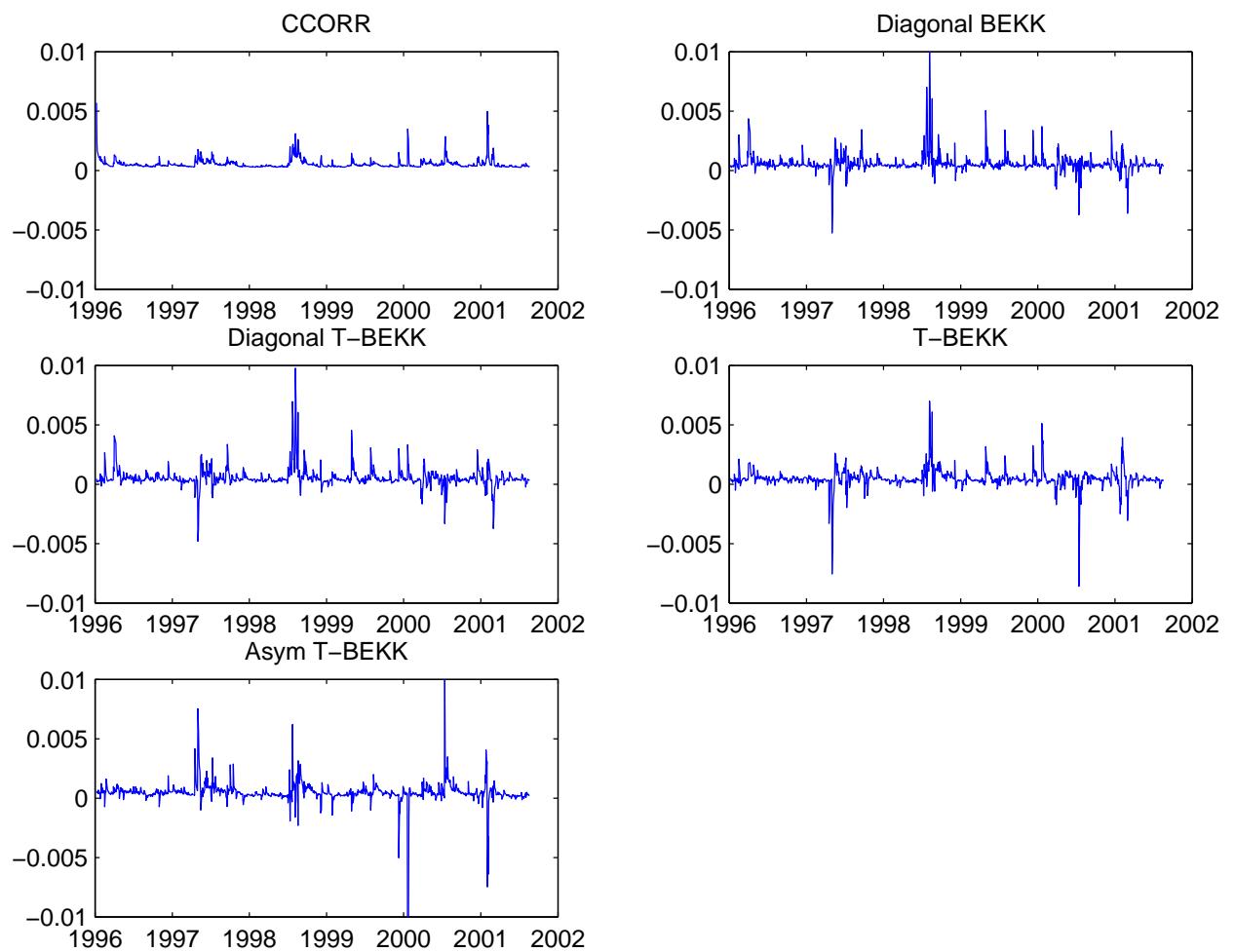


Figure 5: Covariances

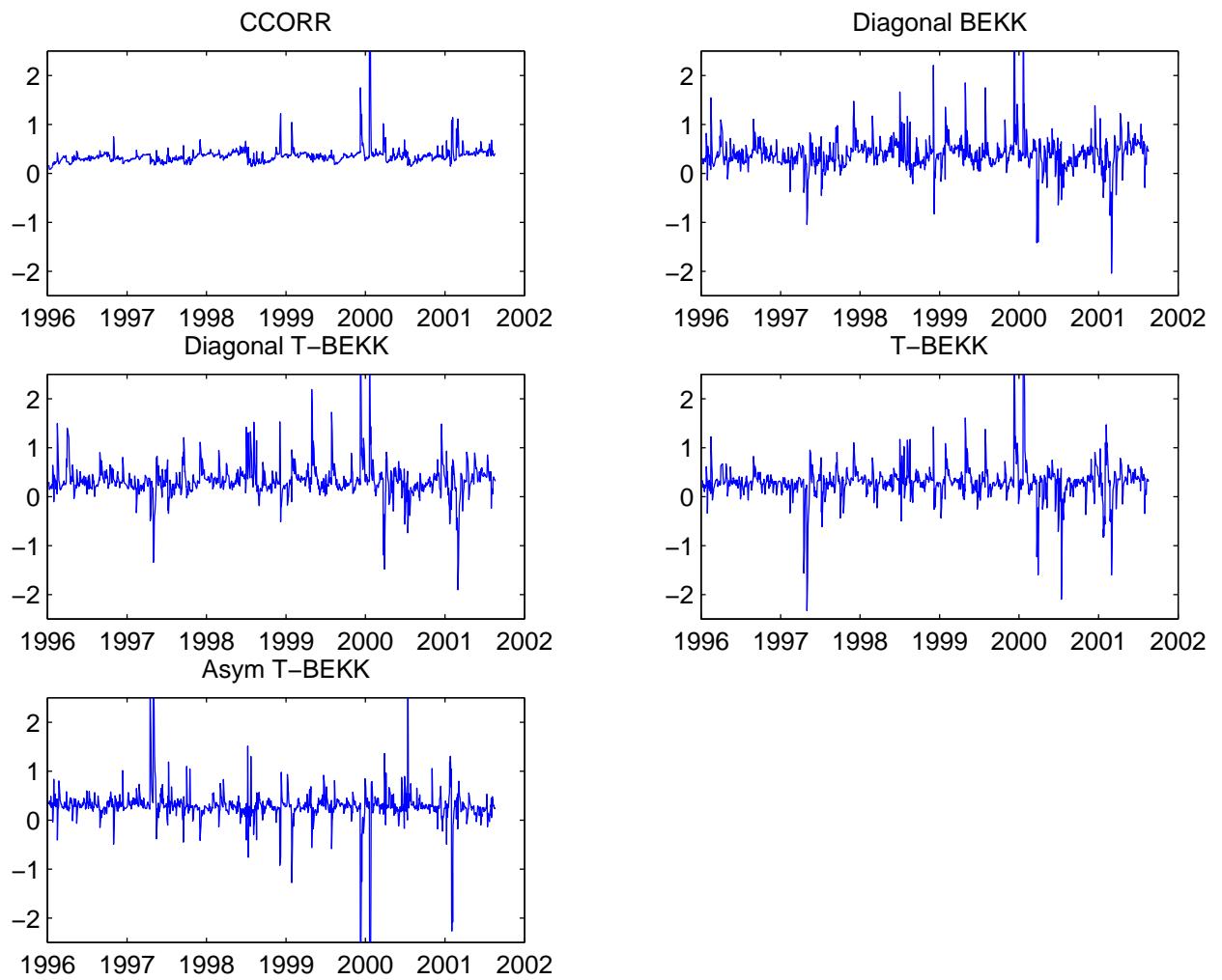


Figure 6: Hedge ratios

Table 10: Portfolio variance tests

Panel A gives the unconditional daily portfolio return variance and percentage variance reduction in the insample and out-of-sample compared to the non-hedged spot position. The insample period covers the first 1400 observations between Jan-1996 and Aug-2001, the outsample period covers the last 299 observations between Aug-2001 and Nov-2002. Panel B gives the daily portfolio return conditional variance (average value over the test period) and percentage reduction of this average conditional variance compared to the non-hedged spot position. The conditional variance tests assume that the true return and variance processes are generated by one of the GARCH models under consideration.

Panel A: Unconditional daily portfolio variances ($\times 100$)						
	Spot	Naive	ECM	CCORR	Diag BEKK	Diag T-BEKK
In-sample						
Variance	0.764	0.800	0.741	0.742	0.755	0.759
% Variance reduction	-	-4.682	3.086	2.944	1.227	0.775
Out-of-sample						
Variance	0.258	0.285	0.251	0.253	0.255	0.252
% Variance reduction	-	-10.802	2.549	1.859	0.996	2.008
Panel B: Conditional daily portfolio variances ($\times 100$)						
	Spot	Naive	ECM	CCORR	Diag BEKK	Diag T-BEKK
In-sample						
Cond.var, CCORR Cov. Matrix	0.688	0.765	0.833	0.668	0.669	0.672
Cond.var, Diag BEKK Cov. Matrix	0.677	0.733	0.801	0.673	0.641	0.643
Cond.var, Diag T-BEKK Cov. Matrix	0.744	0.812	0.877	0.743	0.711	0.710
Cond.var, T-BEKK Cov. Matrix	0.760	0.835	0.894	0.757	0.731	0.733
% Variance reduction, CCORR	-	-11.180	-21.041	2.953	2.870	2.334
% Variance reduction, Diag BEKK	-	-8.319	-18.402	0.547	5.209	4.970
% Variance reduction, Diag T-BEKK	-	-9.163	-17.864	0.222	4.514	4.658
% Variance reduction, T-BEKK	-	-9.926	-17.661	0.281	3.718	3.557
Out-of-sample						
Cond.var, CCORR Cov. Matrix	0.431	0.463	0.505	0.419	0.419	0.421
Cond.var, Diag BEKK Cov. Matrix	0.341	0.373	0.416	0.333	0.326	0.330
Cond.var, Diag T-BEKK Cov. Matrix	0.357	0.395	0.439	0.347	0.342	0.346
Cond.var, T-BEKK Cov. Matrix	0.363	0.414	0.450	0.355	0.350	0.351
% Variance reduction, CCORR	-	-7.405	-17.216	2.798	2.766	2.401
% Variance reduction, Diag BEKK	-	-9.367	-22.086	2.494	4.298	3.297
% Variance reduction, Diag T-BEKK	-	-10.577	-22.982	2.828	4.228	4.119
% Variance reduction, T-BEKK	-	-14.012	-23.854	2.265	3.711	3.328