Dynamic Correlations and Optimal Hedge Ratios

Charles S. Bos[*,a,b] and Phillip Gould[b]

[a]Tinbergen Institute, The Netherlands

[b]Department of Econometrics & O.R.,
VU University Amsterdam, The Netherlands

August 14, 2007

Abstract

The focus of this article is to compare dynamic correlation models for the calculation of minimum variance hedge ratios between pairs of assets. Finding an optimal hedge requires not only knowledge of the variability of both assets, but also of the co-movement between the two assets. For this purpose, use is made of industry standard methods, like the naive hedging or the CAPM approach, more advanced GARCH techniques including estimating BEKK or DCC models and alternatively through the use of unobserved components models. This last set comprises models with stochastically varying variances and/or correlations, and an approximation to these with a single-source-of-error setup.

In order to compare the performance of different models in producing attractive hedging schemes, the reduction in portfolio variance is examined. The data suggests that the most important factor in reducing portfolio variance is the use of a flexible model for time varying volatility, rather than capturing time variation in correlations as closely as possible. Both in simulated and real data series, models incorporating stochastic volatility result in lower risk of the hedged portfolio than GARCH-based variants.

Keywords: Dynamic correlation; multivariate GARCH; stochastic volatility; hedge ratio.

JEL classification: C32, C52, G11

1 Introduction

When combining multiple financial assets into a portfolio, the portfolio return is driven by the dynamics of the underlying asset returns. The asset returns by themselves are volatile, with periods of relatively high and low volatility interchanging. Likewise, when two assets are correlated through time, it can be expected that this correlation is changing over time as well. Taken together, these observations imply that a portfolio of multiple assets which intends to minimise the total risk will have a dynamic ratio of the underlying assets, to account for the relative volatility and the changes in correlations of the assets.

Apart from the construction of optimal portfolios (Markowitz 1952), understanding the dynamics of the volatility and correlation between asset returns is also important for calculations of Value-at-Risk (VaR, Jorion 1997), or for application of the Capital Asset Pricing

[*]Corresponding author: C.S. Bos, Department of Econometrics & O.R., VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Email cbos@feweb.vu.nl, tel/fax: +31 (0)20 598 60 23/+31 (0)20 598 60 20.

Model (CAPM), in which the risk of a stock relative to a market index is assessed (Faff, Hillier, and Hillier 2000).

With the increasing availability of data and computing power, researchers are able to use a range of models and techniques to estimate heteroskedastic time series. The (G)ARCH framework of Engle (1982) and Bollerslev (1986) has been popular in a univariate setting. The stochastic volatility (SV) approach, introduced into the econometric literature by Harvey, Ruiz, and Shephard (1994) and Jacquier, Polson, and Rossi (1994) can also be employed, either using a (quasi) maximum likelihood or a Bayesian approach for the estimation.

In the multivariate GARCH (Bollerslev, Engle, and Wooldridge 1988) and SV (Harvey, Ruiz, and Shephard 1994; Yu and Meyer 2006) literature, correlations are often assumed to be either fixed or follow deterministically from the time variation in variance, in order to simplify the models. Lately, more attention is being paid to freely time varying, dynamic correlations. One difficulty involved in estimating multivariate GARCH models is ensuring positive-definiteness of the covariance matrix. While the BEKK parametrisation of Engle and Kroner (1995) ensures this condition is met, it imposes restrictions on the GARCH structure. The non-linear constraints on the parameter vector also make estimation more difficult. The dynamic conditional correlation (DCC) model of Engle and Sheppard (2001) adapts GARCH models specifically for the estimation of time varying correlations, while also restricting the more general structure. Park and Switzer (1995) estimate models assuming constant correlations and use them to test for time varying hedge ratios in index cash and futures markets. Recently, Pelletier (2006) compares the DCC model to a regime switching dynamic correlation model, with an ARMACH (see Taylor 1986) structure for the variance process.

As in the case of the GARCH model, it is hard to generalise the SV model to allow for time varying correlations between multiple assets; each possible choice for the parametrisation implies a certain restriction in either the space of the possible covariances or correlations. Also, allowing e.g. all correlations to evolve dynamically over time, can lead to a high number of parameters, even for a relatively low number of assets. Therefore, in this article we first limit ourselves to the bivariate case. Yu and Meyer (2006) provide a first attempt at estimating time varying correlations between assets. Our setup, developed independently, corresponds largely in the specification of the model.

This article takes a range of models used in practise and possible extensions which have not yet been applied for estimating the correlation and variance of two assets. With these models, their ability to recover the underlying correlation of the assets is compared, and also the practical effectiveness of the hedging strategies is measured through the standard deviations of hedge portfolio returns. Such a metric is not only relevant for hedge portfolios consisting of cash and futures contracts, but also for constructing and analysing CAPM betas, which are estimated in the same way as time varying hedge ratios.

Section 2 introduces the concepts behind time varying hedge ratios and betas. The different models for time varying correlations are put forward in Section 3. While the dynamic correlation stochastic volatility model appeared once before, in Yu and Meyer (2006), the variant using a single-source-of-error (SSOE, see Ord, Snyder, Koehler, Hyndman, and Leeds 2005) approach is entirely new to the literature. Section 4 starts off describing four data sets in detail, covering data which (i) is simulated, combining GARCH with smoothly varying correlation, (ii) is simulated, specifying an SV process instead of GARCH, (iii) covers the standard CAPM case of daily S&P 500 cash and futures returns and (iv) uses a longer, weekly series of S&P 500 and FTSE 100 returns. The data description is followed by a comparison of the performance of the models in the resulting hedged portfolio standard deviation, in Section 4.2. Section 4.3 discusses the extend to which the different models correctly track the components of the simulated data sets, or provide estimates of the underlying correlation process which seem reasonable in the light of available evidence. Finally, Section 5 gathers the main findings of the comparisons.

2 Time Varying Betas through optimising utility

There is a substantial body of literature dedicated to estimating time varying optimal hedge ratios for simple portfolios consisting of two assets. Brooks, Henry, and Persand (2002) specifically analyse hedge portfolios consisting of positions in cash and futures markets, while Cho and Engle (1999) use similar models to investigate asymmetric time variation in CAPM betas. Though the assets could be a pair of cash and futures, this is not necessary, and a similar hedge could be set up e.g. between the price of kerosene and crude oil futures (Cobbs and Wolf 2004).

The general setup of the problem is as follows. For any two assets at time t, the investor holds β_t units of asset 2 per unit of asset 1, with prices $p_{1\,t}$ and $p_{2\,t}$ respectively. The continuously compounded returns for the two assets are defined as $r_{i\,t} = \log p_{i\,t}/p_{i\,t-1}, i = 1, 2$ with variances $\sigma^2_{i\,t}$ and covariance $\sigma_{12\,t} = \rho_{12\,t}\sigma_{1\,t}\sigma_{2\,t}$.

The rational investor seeks to maximise utility, which is a function of portfolio returns $r_{p\,t} = r_{1\,t} - \beta r_{2\,t}$ and portfolio variance $\sigma^2_{p\,t}$. For portfolio returns, the negative sign before β is used because investors are said to take a short position on e.g. the futures contract and a long position on the underlying stock.

Written in general form the relevant utility function, assuming a standard risk-averse setting with risk-aversion parameter ψ, is:

$$\begin{aligned} U\left[E_{t-1}(r_{p\,t}), \sigma^2_{p\,t}\right] &= E_{t-1}[r_{p\,t}] - \psi\sigma^2_{p\,t} \\ &\approx 0 - \psi(\sigma^2_{1\,t} + \beta^2_{t-1}\sigma^2_{2\,t} - 2\beta_{t-1}\sigma_{12\,t}) \end{aligned}$$

where the last equality is strict when we assume efficient markets and forego discussions on the cost-of-carry and possibly positive expected returns for risk-bearing assets. In this case, a rational agent will seek to maximise this utility function with respect to β_{t-1}, which is equivalent to minimising the variance of portfolio returns. The derived optimal value is

$$\beta^*_{t-1} = \frac{\sigma_{12\,t}}{\sigma^2_{2\,t}} = \rho_t \frac{\sigma_{1\,t}}{\sigma_{2\,t}}. \tag{1}$$

In the time varying hedge ratio literature, the assumption of constant variances is relaxed, which then leads to time variation in the optimal[1] hedge ratio. As discussed in the previous section, different restrictions — such as constant correlations — are sometimes used to facilitate estimation. The next sections describe a range of modelling decisions which imply each their own 'optimal' hedging ratio.

[1]Note how the precise implication of 'optimal' in relation to the hedge ratio depends on the model assumptions.

3 Modelling Correlation

Before a hedged portfolio between two assets can be constructed, it is necessary to model the correlation between the returns. For simplicity, in this article the focus is on a bivariate returns process $r_t = (r_{1\,t}, r_{2\,t})'$, where we assume that the series have been demeaned and any autocorrelation in the series can be neglected. Furthermore, the conditional joint density of the returns is assumed to be normal,

$$r_t | H_t \sim \mathcal{N}(0, H_t), \tag{2}$$

where we condition on the covariance matrix H_t.

When allowing for time variation in this covariance matrix, one can choose to either model the variations and covariation $\sigma^2_{1\,t}, \sigma^2_{2\,t}, \sigma_{12\,t}$ explicitly, specifying

$$H_t = \begin{pmatrix} \sigma^2_{1\,t} & \sigma_{12\,t} \\ \sigma_{12\,t} & \sigma^2_{2\,t} \end{pmatrix}, \tag{3}$$

or alternatively describe the development of the correlation ρ_t over time, using

$$H_t = \begin{pmatrix} \sigma^2_{1\,t} & \rho_t \sigma_{1\,t} \sigma_{2\,t} \\ \rho_t \sigma_{1\,t} \sigma_{2\,t} & \sigma^2_{2\,t} \end{pmatrix}. \tag{4}$$

The difference between these two options basically is whether a change in the volatility of one of the series should also influence the correlation (as in (3)) or not (4).

Subsequent sections provide, in increasing levels of novelty and also of detail, possible model specifications. Section 3.1 provides three baseline models, which look directly at a linear relationship between the returns of the two assets without explicitly modelling a time varying variance process. Section 3.2 moves toward GARCH specifications for the covariance matrix, whereas Section 3.3 introduces novel models with dynamic correlations and stochastic volatility.

3.1 Baseline models with constant variance

For comparison with the models allowing for time varying correlations, three baseline models common to the literature are used. The simplest is the 'naive' model, which hedges the assets fully, fixing $\beta \equiv 1$ throughout the sample.

As two distinct assets are never fully correlated, such a 'naive' hedge is not ideal. The standard approach in the CAPM literature is to look at the time-invariant regression

$$r_{1\,t} = \alpha + \beta r_{2\,t} + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \sigma^2_\epsilon), \tag{5}$$

relating a stock asset with return r_1 to market factor r_2. This gives the second baseline model, indicated as 'CAPM' throughout. As we assumed mean-zero returns, parameter α is fixed at zero.

As a third baseline, some flexibility can be added to the CAPM specification by allowing β to be time varying, as the optimal hedge ratio can also be time varying (see also Equation (1)). Writing the regression in terms of an unobserved random walk β_t gives

$$\begin{aligned} r_{1\,t} &= \beta_t r_{2\,t} + \epsilon_t, & \epsilon_t &\sim \mathcal{N}(0, \sigma^2_\epsilon), \\ \beta_{t+1} &= \beta_t + \eta_t, & \eta_t &\sim \mathcal{N}(0, \sigma^2_\eta). \end{aligned} \tag{6}$$

Such a model in state space (Durbin and Koopman 2001) can be estimated using the Kalman equations, allowing the optimal time varying β_t to be extracted. This model will be referred to below as the time varying regression (TVR) model.

Each of these three models makes its own set of assumptions on the underlying data generating process. Especially, all three baseline models effectively assume constant variances for both assets, with only the TVR model allowing for variation in the relation between the returns. The optimal hedging portfolios that can be constructed on the basis of the outcomes of these models will clearly differ; in Section 4 the results will be compared, to find out which model leads to most optimal hedging performance.

3.2 GARCH specification for volatility

After the inception of the GARCH models (Engle 1982; Bollerslev 1986), multivariate versions quickly appeared. Two distinct specifications are considered here, with the first specifying the development of the covariance directly as in (3), the second directed at modelling correlation (4).

The BEKK model

The BEKK specification for the multivariate GARCH model was introduced in Engle and Kroner (1995, an earlier version of the paper was circulated with authors Baba, Engle, Kraft and Kroner, hence the acronym of the model). The full covariance matrix H_t is related to its past and to past returns through the updating equation

$$H_t = C'C + A'r_{t-1}r'_{t-1}A + B'H_{t-1}B$$

where A and B are symmetric matrices, and C is upper-triangular. By updating H_t as the sum of three positive definite matrices, by construction H_t will stay positive definite itself as well. For stationarity, a non-linear restriction on the parameters in A and B is needed needed to ensure that the eigenvalues of $A + B$ lie within the unit circle.

The DCC model

Where the BEKK model describes a process for the variances and covariances of the returns, the Dynamic Conditional Correlation model (Engle 2002, DCC) explicitly recognises that the interesting dynamics often lie in the realm of the correlations instead of the covariances, and has become a popular multivariate GARCH representation. The elements of the covariance matrix H_t are specified as

$$\sigma^2_{i\,t} = \omega_i + \delta_i \sigma^2_{i\,t-1} + \alpha_i r^2_{i\,t-1}, \tag{7}$$

$$q_{ij\,t} = (1-\lambda)\frac{r_{i\,t-1}}{\sigma_{jt-1}}\frac{r_{i\,t-1}}{\sigma_{jt-1}} + \lambda q_{ij\,t-1}, \tag{8}$$

$$\rho_t = \frac{q_{12\,t}}{\sqrt{q_{11\,t}\; q_{22\,t}}}. \tag{9}$$

For higher-dimensional models, it can be more convenient to express the DCC model directly in matrix format, though for the present bivariate case these univariate expressions clarify how indeed the model separates returns into univariate GARCH (in (7)) and correlation sequences (9). The rescaled correlation coefficient $q_{ij\,t}$ in (8) is a weighted average between past correlation and past standardised returns $r_{i\,t-1}/\sigma_{i\,t-1}, i = 1, 2$. The structure of the equations assures that the covariance matrix stays positive definite. For further details, see also Engle and Sheppard (2001).

3.3 Stochastic specification of the volatility using unobserved components

The multivariate extension of the SV model, for the constant correlation case is already proposed in Harvey, Ruiz, and Shephard (1994). Also a multi-factor SV model is a common extension, as e.g. is done in Chib, Nardari, and Shephard (2006),[2] but to get to a separate time varying correlation factor is more difficult. To our knowledge, the bivariate Dynamic Correlation Stochastic Volatility (DCSV) model used here is only presented before in Yu and Meyer (2006).

Apart from these stochastic volatility models, a quasi maximum likelihood approach using a Single Source of Error Model (SSOE, see Ord, Snyder, Koehler, Hyndman, and Leeds 2005) is presented in a subsequent section.

Dynamic Correlation SV

The starting point for the DCSV model is again the returns equation (2), with the correlation element in the variance matrix H_t of (4) modelled as

$$\rho_t \equiv \frac{\exp(q_t)-1}{\exp(q_t)+1} = 2\frac{1}{\exp(-q_t)+1} - 1, \tag{10}$$

$$q_{t+1} = q_t + \eta_t, \qquad \eta_t \sim \mathcal{N}(0, \sigma^2_\eta). \tag{11}$$

Both log-price returns have SV-type variances, according to

$$\sigma^2_{i\,t} = \exp(h_{i\,t}) \tag{12}$$

$$h_{i\,t+1} = \gamma_i + \phi(h_{i\,t} - \gamma_i) + \xi_{i\,t} \qquad \xi_t = \mathcal{N}\left(0, \Sigma_\xi\right) \tag{13}$$

The two returns are related through the correlation coefficient ρ_t, which varies over time according to a transformed random walk (11). The transformation in (10), which is a rescaled Sigmoid function, assures that the correlation is bounded in (-1, 1).

The correlation ρ_t itself is only observable through the pair of returns at time t itself. Therefore, there is truly little information on ρ_t if not for the time structure imposed on the correlation process. This also implies that it is not useful to specify a more refined process for the correlations, or a different transformation, as the hypothesis of a slightly different specification cannot be tested with acceptable power anyhow. The difference with a strongly simplified model should be detectable, however. As a strong simplification, fixing $\sigma_\eta \equiv 0$ gets us back to the static correlation model with dual stochastic volatility (SCSV). This model is used below for comparison.

The two SV innovations $\xi_{i\,t}$ could display correlation as well, though in this article we limit ourselves to the situation where $\Sigma_\xi \equiv \text{diag}(\sigma^2_{\xi\,1}, \sigma^2_{\xi\,2})$. First of all, the novelty of introducing dynamic correlation in a stochastic volatility model already imposes sufficient technical difficulties. Secondly, it is at present unclear how well identified a correlation coefficient between two innovation processes of unobserved volatility components would be.

[2]With the multi-factor SV setup of Chib, Nardari, and Shephard (2006) it is possible to construct a time varying covariance factor, comparable to the BEKK model, but not an explicitly modelled correlation factor.

Single Source of Error Estimation for SV

The major drawback of the DCSV model of the previous section is that it is fully nonlinear, based on unobserved components. This leads to the need for advanced Bayesian sampling techniques (more on this topic in Section 3.4) to estimate the parameters, variances and correlations.

In the article by Harvey, Ruiz, and Shephard (1994), a quasi maximum likelihood (QML) approach was used to estimate the stochastic volatility model. Their approach entailed linearising the model to a get to a model which can be estimated using standard Kalman filtering techniques (Durbin and Koopman 2001). The linearised version of the univariate SV model for r_t reads

$$\ln r_t^2 = h_t + \nu_t \tag{14}$$
$$h_{t+1} = \gamma + \phi(h_t - \gamma) + \xi_t \tag{13'}$$

where $\nu_t = \ln u_t^2, u_t \sim \mathcal{N}(0, 1)$, has a non-standard density. The density of ν_t is a transformation from the standard normal density of u_t, with mean -1.27 and variance $\pi^2/2$. When the Kalman filter is used in the above setting, effectively a normal approximation to the density of ν_t is used, and hence the resulting estimator is only a quasi-maximum likelihood estimator. On the other hand, the approach has the appeal of being both intuitive and easy to implement. The approach is not efficient, so results will be an approximation of the outcomes of e.g. a Bayesian estimation approach of the exact model.

When correlations ρ_t are introduced as in the DCSV model, a new source of non-linearity appears. A general approach to cope with non-linear state space models is the single source of error (SSOE) framework, presented in detail by Ord, Snyder, Koehler, Hyndman, and Leeds (2005). Where the multiple source of error approach allows separate independent disturbances e_t and ξ_t, the SSOE approach assumes full correlation between the disturbances. Relating the SV disturbance in (13') to the observation disturbance in (14), it takes $\xi_t \equiv \psi\nu_{t-1}$ for some constant parameter ψ. Note how this provides a closer link to the GARCH approach, where also the innovation in the return equation drives deterministically the changes in the variance equation.

To link together a bivariate system of observations and the correlation sequence, the full model is specified as

$$r_{1\,t} = \beta_t r_{2\,t} + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \exp(h_{1\,t})) \tag{6'}$$
$$\ln r_{i\,t}^2 = h_{i\,t} + \nu_{i\,t}, \qquad \nu_{i\,t} \sim \mathcal{N}(-1.27, \frac{1}{2}\pi^2) \tag{14'}$$
$$\beta_t = \rho_t \sqrt{\frac{\exp(h_{1\,t})}{\exp(h_{2\,t})}}, \qquad \rho_t = \frac{\exp(q_t) - 1}{\exp(q_t) + 1}, \tag{1', 10'}$$
$$q_{t+1} = q_t + \psi_0 \epsilon_t, \tag{11'}$$
$$h_{i\,t+1} = \gamma_i + \phi_i(h_{i\,t} - \gamma_i) + \psi_i \nu_{i\,t}. \tag{13''}$$

In this setup, the time varying regression (6) of Section 2 is used to obtain more information on the dynamics of the variance and covariance terms. This equation is combined with two QML observation equations, (14'), while the three state equations describing q_t and $h_{i\,t}$ are linked using smoothing parameters ψ_0, ψ_1, ψ_2 to the disturbances of the observation equations. As mentioned before, the distributional assumption of normality for $\nu_{i\,t}$ is only approximative.

3.4 Estimation of the models

The baseline models of Section 3.1 either do not need estimation (the NAIVE model), can be estimated using OLS (CAPM) or a simple maximum likelihood (ML) setup for a state space model (TVR). Though the likelihood functions of the BEKK and DCC models are more elaborate, these as well allow for a relatively easy ML estimation procedure (see also Engle 2002 for details). The advantage of the SSOE-SV models is that they are able to handle non-linearity easily, as the above set of equations can be solved recursively for a fixed set of parameters. This allows the likelihood function to be calculated in a direct manner, optimising over the parameters using standard methods.

Estimation of the DCSV (and the derived SCSV) model is more demanding, as the likelihood is not available in closed form without integrating out the latent variables $h_{i\,t}, i = 1, 2$ and $\rho_t, t = 1, .., T$, for volatility and correlation. Therefore, estimation is done here using a Bayesian approach with data augmentation, using a Markov chain Monte Carlo method.

The algorithm proceeds, after initialising the parameters $\theta = (\phi_i, \gamma_i, \sigma_{i\,\xi}, \sigma_\eta), i = 1, 2$ and states, by iterating over the following steps:

i Sampling a new vector of ρ, by successively sampling $\rho_t|\rho_{t-1}, \rho_{t+1}, y_t, h_t, \theta$, for $t = 1, .., T$. As this density is not available in closed form, we use a random walk Metropolis-Hastings (MH) step to sample from the posterior density

$$P(q_t|q_{t-1}, q_{t+1}, y_t, h_t, \theta) \propto P(q_t|q_{t-1}, q_{t+1}, \sigma_\eta) \times \mathcal{L}(y_t; h_t, \rho_t(q_t)).$$

Both densities are simple Gaussians, such that sampling a new value of q_t is not difficult. After sampling q_t, it is transformed back to ρ_t;

ii Sampling two new vectors of h jointly, from the density of $h_t|h_{t-1}, h_{t+1}, y_t, \rho_t, \theta$. Again, the density is not easily tractable in closed form, but the posterior is again a combination of the density of $h_t|h_{t-1}, h_{t+1}, \theta$ and the likelihood of the present observation $\mathcal{L}(y_t|h_t, \rho_t, \theta)$.

iii Assuming a prior $\pi(\sigma_\eta) \sim$ IG-1$(\alpha_\eta, \beta_\eta)$, the posterior of σ_η is

$$P(\sigma_\eta|\rho) \sim \text{IG-1}\left(\alpha = \frac{T-1}{2} + \alpha_\eta, \beta = \left(\sum_{t=2}^{T} \frac{(q_t - q_{t-1})^2}{2} + \frac{1}{\beta_\eta}\right)^{-1}\right)$$

iv The parameters $\gamma_i|h_i$ follow a simple normal density, assuming a normal prior with mean μ_γ and variance σ_γ^2;

v The remaining parameters $\phi_i, \sigma_{i\,\xi}$ are sampled per asset i using another MH step, with a random walk normal candidate density.

After performing a sufficient number of iterations this algorithm results in a sample from the posterior density of the parameters θ and states h, ρ. The posterior mode of the parameters θ estimated over the sample is used as input for a particle filter (Pitt and Shephard 1999), to extract filtered estimates of the states $\rho_t, h_t|y_1, .., y_t$, conditioning only on past and present information. This gives a more fair comparison than using the output of the MCMC chain, which describes the distribution of the states $\rho_t, h_t|y_1, .., y_T$ conditional on the full data set.

Table 1: Prior densities and moments

Parameter	Density	Prior parameters		μ	σ_π
σ_η	IG-1	$\alpha_\eta = 1.3$	$\beta_\eta = 100$	0.13	0.13
γ_i	$\mathcal{N}$	$\mu_\gamma = 0$	$\sigma_\gamma = 2$	0	2
ϕ_i	Beta	$\alpha_\phi = 12$	$\beta_\phi = 3$	0.8	0.1
$\sigma_{i\,\xi}$	IG-1	$\alpha_\xi = 1.3$	$\beta_\xi = 300$	0.075	0.074

A Bayesian procedure needs some prior information on the parameters. These priors are fixed using information on the expected size of the parameters based on related research, and taken with sufficient spread to allow the present data to decide on the location of the posterior. The parameters, prior density family, prior parameters and moments of the prior are given in Table 1. For all data series, these same priors are used.

Alternatively, Sandmann and Koopman (1998), with further extensions by Jungbacker and Koopman (2007), provide a classical maximum likelihood approach for models with stochastic variance, applying importance sampling to simulate out the unobserved states. This method could be extended to include the time varying correlation as well.

4 Evaluating Model Performance

In this section the performance of the models is evaluated in four different settings. These settings, or the underlying data, is described in Section 4.1. It is followed by a section comparing the portfolio standard deviation resulting from using the different modelling setups. Section 4.3 continues the comparison by looking at the components for variance and correlation that the best models extract from the data.

4.1 Data

The first two data sets are simulated series, one from a GARCH-type specification and the other from a stochastic volatility data generating process. The parametric settings of the GARCH specification is taken from Engle (2002), where a range of different GARCH models are compared in their ability to extract a time varying correlation structure from bivariate returns. The variances in the covariance matrix H_t of Equation (4) are specified as

$$\sigma^2_{1\,t} = .01 + .05 y^2_{1\,t-1} + .94 \sigma^2_{1\,t-1},$$
$$\sigma^2_{2\,t} = .5 + .2 y^2_{2\,t-1} + .5 \sigma^2_{2\,t-1},$$

implying a GARCH process for both returns. Note how the first GARCH process has high persistence, and the second is of relatively low persistence.

For the time varying correlation, the slowly changing sine specification from the same article is used, with

$$\rho_t = .5 + .4\cos(2\pi t/200). \tag{15}$$

In order to get a more honest comparison, a second data set is generated according to a stochastic volatility setup. The log-volatilities $h_{it} = \log \sigma^2_{i,t}$ follow

$$h_{1\,t} = 1 + .97(h_{1\,t-1} - 1) + \xi_{1\,t}, \qquad \xi_{1,t} = \mathcal{N}(0, 0.059),$$
$$h_{2\,t} = 1 + .7(h_{2\,t-1} - 1) + \xi_{2\,t}, \qquad \xi_{2,t} = \mathcal{N}(0, 0.051).$$

The variances of the volatility disturbances ξ_{it} are chosen such that the long-run variances of the $h_{it}, \sigma^2_{hi} = \sigma^2_{\xi i}/(1-\phi^2_i)$ are 1 and 0.1, respectively. This implies that the first log-volatility is of higher persistence and variance than the second.

Together with specification (15) for the correlation, this provides a second full specification for a possible data generating process.

Data is repeatedly generated from both the GARCH and the SV specification for a period of $N = 1,000$ observations, to obtain 100 data sets of each type. Of the 1,000 observations, 90% are used for estimating the parameters, while 10% are preserved as a hold-out period for assessing out-of-sample performance. Estimation is performed using each of the models; results are disregarded if parameter estimation does not converge. For the Bayesian estimation procedures of the DCSV and SCSV models, a sample of 10,000 parameter vectors is collected after allowing the algorithm to burn in for 1,000 iterations when using simulated data. For the financial data series outlined below (where the estimations are not repeated 100 times), samples 10 times larger are used both for burn-in and for finding the posterior mode.

The latter two data sets compare the S&P 500 index to either its own future, or to the FTSE 100 index. In the first case, daily data over the period January 7, 1998–December 28, 2006 is used, for a total of $N = 2266$ observations. The data are obtained from DiskTrading[3], and concern the daily index or future price at closing time. This data set allows investigating the common situation of an investor trying to hedge his or her risk using a future on the same stock.

The second applied data set compares the S&P 500 as traded in New York to the FTSE 100 index, traded in London. From Yahoo Finance[4], a longer series of prices is available. Of these, the Friday closing price (or last day of the week, in case the market is closed on Friday) is taken to construct a weekly time series over the period January 4, 1985–December 31, 2006, for a total of $N = 1147$ observations. Such a longer time period, with assets trading on two geographically distinct markets, can display the robustness of the models to more variation in correlation.

Table 2 provides the basic statistics on the moments and the correlations of the returns, for one simulation of both DGPs, and for the financial data series. The statistics are split out over the in-sample period consisting of the first 90% of the observations, and the out-of-sample period for the remainder.

The statistics indicate the standard findings of financial return series: A relatively low mean return (which is subtracted before the analysis), some indication of negative skewness, and a kurtosis which tends to be higher than 3. The kurtosis in- and out-of-sample is estimated rather differently for both the simulated SV-with-sine and for the weekly S&P 500 vs. FTSE returns data; this seems to be due to the low number of observations retained for the out-of-sample period.

Note as well how the daily returns on the S&P 500 and its future are highly correlated, whereas the correlation between the weekly S&P 500 and FTSE returns is considerably lower at 0.6–0.7.

4.2 Lowering the portfolio risk

The main idea behind the CAPM is to construct a portfolio which would lower the risk of holding the asset by itself, through hedging. Both in the case of simulated data as in the

[3]See `http://www.disktrading.com`.

[4]See `http://finance.yahoo.com`, symbols `SNP:^GSPC` and `FSI:^FTSE`.

Table 2: Data availability and statistics

	GARCH with sine			
	In-sample		Out-of-sample	
Days	900		100	
	Index	Future	Index	Future
Mean	0.0404	0.0084	-0.0939	-0.0906
St.dev.	0.8560	1.2954	0.7351	0.9787
Skewness	0.0781	-0.0499	0.2007	-0.1695
Kurtosis	3.0826	3.8780	2.6575	2.5969
Avg. corr.	0.4885		0.3359	

	SV with sine			
	In-sample		Out-of-sample	
Days	900		100	
	Index	Future	Index	Future
Mean	0.0373	0.0150	0.0914	-0.0147
St.dev.	2.1481	1.6496	1.2124	1.5353
Skewness	0.4514	-0.0738	0.1191	0.2192
Kurtosis	6.3310	3.0416	3.3796	2.2139
Avg. corr.	0.4309		0.5607	

	S&P 500 versus future			
	In-sample		Out-of-sample	
Days	2039		227	
Period	1998/01/02–2006/02/03		2006/02/06–2006/12/28	
	Index	Future	Index	Future
Mean	0.0130	0.0126	0.0527	0.0550
St.dev.	1.1909	1.2120	0.6250	0.6144
Skewness	0.0031	-0.0609	0.1669	-0.0988
Kurtosis	5.4062	5.7406	4.1332	4.3684
Avg. corr.	0.9690		0.9632	

	S&P 500 versus FTSE 100			
	In-sample		Out-of-sample	
Weeks	1032		115	
Period	1985/01/04–2004/10/15		2004/10/22–2006/12/29	
	S&P	FTSE	S&P	FTSE
Mean	0.1838	0.1287	0.2145	0.2582
St.dev.	2.1870	2.2629	1.3633	1.3987
Skewness	-0.7136	-0.8712	-0.1473	-0.1614
Kurtosis	6.5668	9.9502	2.8372	3.8157
Avg. corr.	0.6111		0.6978	

applications, such a portfolio based on the different models can be constructed to get to the lowest possible portfolio standard deviation. Tables 3–4 display this standard deviation, both in- and out-of-sample, together with the percentage difference with the lowest standard deviation.

For building a low-variance portfolio, it clearly is important to know not only the correlation between the two assets, but also the relative variances. For the simulated data series, the lowest risk is usually obtained using the DCSV model. The SSOESV manages to come close, performing even better on average for the SV-with-sine model over the in-sample period.

Performance of the DCC and BEKK models indicates a portfolio standard deviation of

Table 3: Portfolio standard deviation, simulated data

	$\sigma(r_p)$ GARCH-with-sine in-sample		out-of-sample		non-conv	$\sigma(r_p)$ SV-with-sine in-sample		out-of-sample		non-conv
Naive	1.179	[62.3]	1.176	[57.5]		2.045	[25.1]	1.994	[23.7]	
CAPM	0.815	[12.2]	0.825	[10.5]		1.895	[16.0]	1.819	[12.9]	
TVR	0.788	[8.4]	0.799	[7.0]		1.819	[11.3]	1.754	[8.8]	
BEKK	0.785	[8.0]	0.807	[8.1]	5%	1.776	[8.7]	1.750	[8.6]	17%
DCC	0.796	[9.6]	0.804	[7.7]		1.765	[8.0]	1.714	[6.4]	
DCSV	0.727	[0.0]	0.747	[0.0]		1.670	[2.2]	1.612	[0.0]	
SCSV	0.797	[9.6]	0.815	[9.2]		1.843	[12.8]	1.783	[10.6]	
SSOESV	0.736	[1.3]	0.797	[6.8]		1.634	[0.0]	1.705	[5.8]	5%

Note: The table reports the portfolio standard deviation, both in-sample (first 90% of the data) and out-of-sample, with the percentage difference with the lowest standard deviation reported between square brackets. The remaining columns indicate the percentage non-convergence of the estimation procedure.

around 6-10% worse than for the DCSV model. This holds both for the Sine-with-GARCH data, on the left, as for the Sine-with-SV data. Even though for the DGP with GARCH-type volatility the DCC model is less misspecified than the DCSV model, it does not manage to provide better hedging results. Even the TVR model, which does not explicitly model time variation in variances, manages to come up with an equivalently performing hedging strategy.

Next in order of best-to-worse performing models are the SCSV and CAPM. SCSV improves on CAPM through the time variation in volatility, but both models are severely hampered by not being able to shift the correlation around. Last of the pack is the naive model, fixing the hedge ratio at one. Naturally, this will not be a good strategy in the current settings.

More detailed investigation of the individual simulation results (not reported here) shows that the estimation of both BEKK and SSOESV did not fully converge for each of the sampled data sets. In 17% of the sampled data sets of the Sine-with-SV DGP, no convergence in the BEKK estimation was found, against 5% non-convergence for the SSOESV model. Also, the deterministic adaption rules for variance and correlation of the SSOESV model can result in more extreme estimates for these sequences. As a consequence, the spread of the portfolio standard deviations (not reported here) is considerably higher out-of-sample than that of any of the other approaches.

For a practitioner, the real-world results for the financial data hedging, in Table 4, can be more interesting. For the S&P 500 against the futures data, in the first four columns, the true SV models take the lead, where the generalisation from the SCSV to time varying correlation in the DCSV model does not seem to be as important as the introduction of stochastic volatility itself. The SSOESV model performs worse than even the naive model for this data set. The estimation in this case was found to be troublesome, with a multimodal likelihood function. Reported results are for the case with best likelihood value; in this case however the correlation sequence converged very quickly to a value of 1, even though true correlation between the S&P 500 and its future is slightly lower than that.

The other models result in rather similar portfolio standard deviations, at least for this data set, both in- and out-of-sample. This does indeed imply that also the CAPM approach in this case is performing on par with the BEKK and DCC approaches.

Table 4: Portfolio standard deviation, financial data

	$\sigma(r_p)$ S&P vs Future				$\sigma(r_p)$ S&P vs FTSE			
	in-sample		out-of-sample		in-sample		out-of-sample	
Naive	0.300	[13.5]	0.169	[8.7]	1.963	[20.7]	1.074	[13.6]
CAPM	0.294	[11.4]	0.169	[9.0]	1.731	[6.4]	0.984	[4.1]
TVR	0.295	[11.4]	0.168	[8.7]	1.704	[4.8]	0.984	[4.1]
BEKK	0.295	[11.5]	0.171	[10.5]	1.704	[4.7]	1.030	[9.0]
DCC	0.302	[14.3]	0.170	[9.4]	1.706	[4.9]	0.997	[5.5]
DCSV	0.271	[2.5]	0.159	[2.5]	1.627	[0.0]	0.945	[0.0]
SCSV	0.264	[0.0]	0.155	[0.0]	1.664	[2.3]	0.965	[2.1]
SSOESV	0.315	[19.2]	0.180	[16.3]	1.639	[0.8]	0.962	[1.8]

Note: See Table 3 for an explanation of the entries in the table.

For the S&P 500 vs FTSE data, both DCSV and SSOESV appear to track the volatility and correlation well, resulting in a low hedged portfolio variance, with the static correlation SV model in third place. The multivariate GARCH models, BEKK and DCC, do not deliver a clear advantage over the simpler TVR model, and perform out-of-sample even worse than the simpler CAPM. As the correlation over the longer horizon and for indices on distinct markets is lower than before, the naive approach does not deliver a good hedged portfolio.

To understand why the ordering of model performance comes out this way, the next section considers the estimated components that result from each of the models.

4.3 Component tracking

Though the standard deviation of the resulting hedged portfolio is the most important target of an investor in our setting, one can wonder where these results stem from. Especially it was found that the DCSV model leads to lower portfolio risk in general, with in some cases the SCSV (with its fixed correlation) even winning the contest. Though this means that these stochastic volatility models must be doing a decent job in tracking the correlation and volatilities, does this translate to a closer relation between the estimates and underlying values of the DGP?

For the case of the simulated data sets, a straightforward measure to check is the mean squared error (MSE) of the estimated correlation $\hat{\rho}_t$. Table 5 reports the average MSE over the replications of the simulated data sets. From this table it is clear that the DCC model tracks the sinusoidal form of the correlation best, with the DCSV model following in second place. This implies that it is not only the precision with which the models follow the true ρ sequence which results in lower variability of the hedged portfolio. The manner in which the volatility is filtered is at least as important, also seen from the result in Table 3 where the DCSV model led to improved hedging over the DCC model, even though the DGP was a GARCH-with-sine model fitting in closely with the DCC model.

With the financial data sets, there is no simple measure to compare correlation against. Instead, Figure 1 presents the estimates of underlying volatility, for the case of the S&P 500 index versus the future. The panels show both volatilities σ_1^2 and σ_2^2 (against the left axis), with ρ against the right axis, for the DCSV, BEKK and DCC models.

From this figure it is apparent how DCSV estimates a higher persistence in volatility than each of the other models, with especially DCC allowing volatility to increase more quickly

Table 5: MSE of correlation measure, for simulated data

	MSE($\hat{\rho}$) GARCH-with-sine				MSE($\hat{\rho}$) SV-with-sine			
	in-sample		out-of-sample		in-sample		out-of-sample	
Naive	32.956	[928.1]	33.400	[1125.8]	32.956	[927.6]	33.400	[881.8]
CAPM	8.150	[154.2]	8.127	[198.3]	8.553	[166.7]	8.496	[149.8]
TVR	6.503	[102.9]	5.050	[85.3]	9.208	[187.1]	8.562	[151.7]
BEKK	3.592	[12.1]	3.667	[34.6]	4.164	[29.9]	4.586	[34.8]
DCC	3.206	[0.0]	2.725	[0.0]	3.207	[0.0]	3.402	[0.0]
DCSV	3.412	[6.4]	4.223	[55.0]	3.466	[8.1]	4.282	[25.9]
SCSV	8.199	[155.8]	8.234	[202.2]	8.149	[154.1]	8.176	[140.4]
SSOESV	5.730	[78.7]	5.321	[95.3]	5.825	[81.6]	6.331	[86.1]

Note: The table reports the average mean squared error ($\times 100$) of the estimated correlation, both in-sample (the first 90% of the observations) and out-of-sample (over the last 10% of the simulated data), over 100 iterations of generating data from the DGP and estimating the models. The percentage difference with the lowest MSE is reported between square brackets.

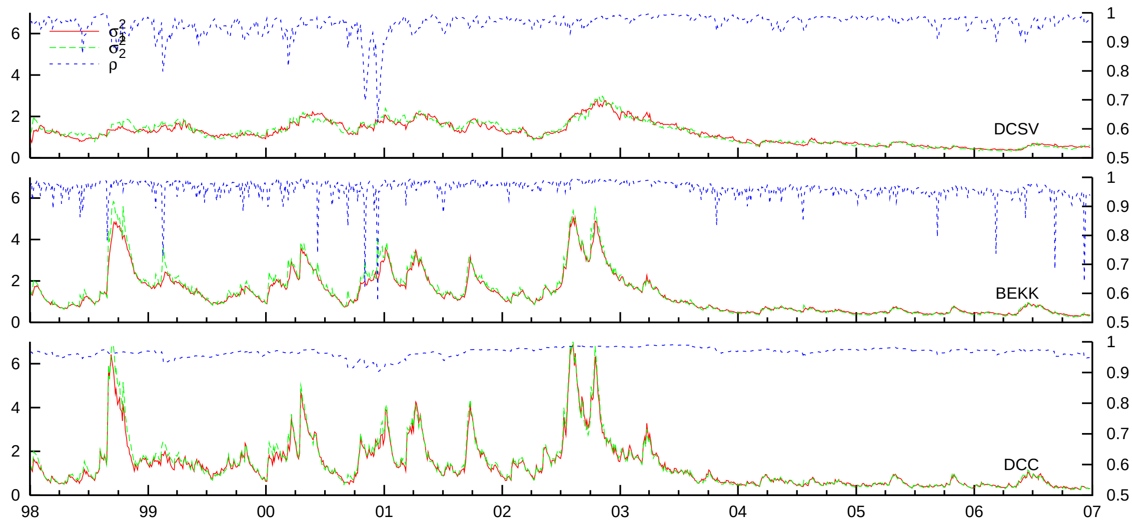

Figure 1: Variances (left hand scale) and correlation (right hand scale) of DCSV, BEKK and DCC models for the S&P 500 vs. futures data

and to higher levels. For the correlation, DCSV allows for some changes in correlation, most pronounced at the end of 2000. The BEKK model estimates more sudden shocks to correlation. The spikes which result seem to be less realistic, as the underlying correlation between the S&P 500 and its future cannot be expected to change so quickly for such short periods. Correlation estimates of the DCC model on the other hand seem very constant, with virtually no change to be found between 1998 and 2007. As correlation is estimated relatively constant by each of these models, simplifying the DCSV to the static correlation SV model did not hamper the hedging at all, as was found before in Table 4.

The estimations using the the S&P 500 versus the FTSE data, over the longer time horizon, resulted in more variability of estimated correlation sequences. Figure 2 shows the underlying variances and correlations for the DCSV, DCC and SSOESV models. Whereas the overall

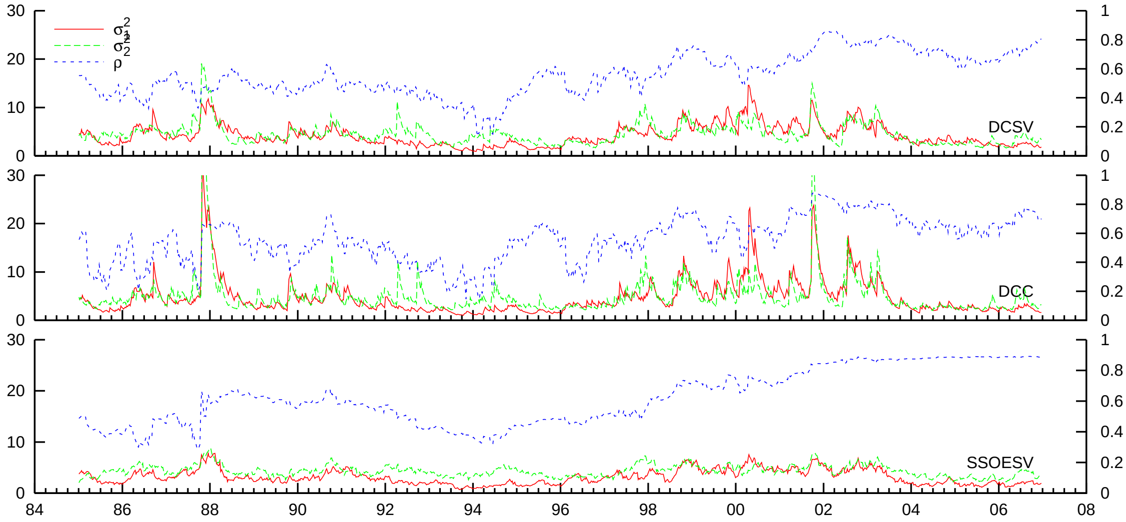

Figure 2: Variances (left hand scale) and correlation (right hand scale) of DCSV, DCC, and SSOESV models for the S&P 500 vs. FTSE data

movement of the correlation is similar between the models, one can see very slow changes in the SSOESV case (with hardly any changes in correlation at all after 2002), and sudden breakdowns in correlations for the DCC results in the center panel. The DCSV takes up an intermediate position, with generally smooth adaptation of correlation to new circumstances, it seems.

The plot of the variances resulting from the DCC model has been clipped at 30; around Black Monday, October 1987, estimated variances jump up towards 68 for this model. Again after September 11, 2001, variances increase strongly for the DCC model. Both other models result in more moderate variance increases in those periods.

5 Conclusions

This article studies the possibility of building a hedge portfolio out of two assets, using industry-standard approaches like the CAPM or with a naive hedge, and compares these to more advanced approaches applying BEKK and DCC GARCH-type models. Furthermore, novel hedging methods are proposed, allowing for time variation in the CAPM using a time varying regression model, or using a latent variable approach with stochastic volatility and stochastic correlation. For comparison, a static correlation SV and an approximate single-source-of-error SV model are introduced. This latter model is entirely novel to the literature.

With each of the models hedging decisions are made for both simulated and actual data series. The relative performance of the models is judged primarily by their ability to minimise the variance of a hedged portfolio. A further look is taken into the estimated variance and correlation components which underlie the hedging decisions, to better understand why certain methods provide better hedging results than others.

If one conclusion is clear from the results, it must be that the DCSV model tends to deliver robust performance throughout the simulated data series and the S&P 500 vs future or FTSE indices. Using stochastic volatility together with dynamic correlation provides a hedging decision which gives either the lowest or second-lowest standard deviation of the portfolio. When the correlation is not moving much over the horizon of the data series, the static correlation SV model performes well. Allowing the correlation to vary over time does not deliver an improved hedging decision.

Among the GARCH-type models, a clear ordering could not be found. Neither the DCC overall outperformed the BEKK nor vice versa, and both trailed the best models in providing a low-variance hedge decision. The DCC model however does provide the best estimates of the correlation sequence, for the simulated data series. For the financial data series, BEKK results in erratic and highly volatile correlation estimates, contrasting the DCC estimates which are far more stable over time. Both DCC and BEKK allow the variance processes to jump up to far higher levels than corresponding SV estimates would imply.

On the S&P 500 vs FTSE data, the SSOESV model performed almost on par with the DCSV model, delivering similar estimates for volatility and correlation as well. This contrasts the finding for the other financial data set, where the SSOESV model was not able to come up with a reasonable estimate of correlation, and performed worse than even the naive hedge. This leads to the conclusion that the SSOESV model is promising, but that care should be taken to ensure the model has indeed converged to a sensible solution. As a first estimate, quick and easy to attain, it however can serve well.

A more definite result of the computation of model-based correlations is the finding that for the S&P 500 vs its future series, the correlation has been high, around 0.97, and nearly constant, throughout the entire sample period 1998–2006. The model-based estimates agree in this respect to a large extent.

For the longer time series concerning the S&P 500 and FTSE indices, the models agree that overall correlation increased, from 0.4-0.6 in 1984 up to 0.8 at the end of 2006. However the path of the correlation is estimated slightly differently depending on the model used.

Only bivariate models have been applied in this article, in order to grasp the importance of modelling correlation correctly in a simple hedge portfolio of two assets. The extension to multiple assets, though interesting in its own right, would obfuscate one of the main findings: For the series considered here, applying a flexible volatility model is at least as important as allowing the correlation to change over time. However, an extension to a comparison of multivariate DCC and DCSV along the lines proposed in Yu and Meyer (2006), effectively following the framework of Engle (2002), could be an interesting possibility.

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988). A capital asset pricing model with time-varying covariances. *The Journal of Political Economy 96*(1), 116–131.

Brooks, C., O. T. Henry, and G. Persand (2002). The effect of asymmetries on optimal hedge ratios. *Journal of Business 75*(2), 333–352.

Chib, S., F. Nardari, and N. Shephard (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics 134*(2), 341–371.

Cho, Y.-H. and R. F. Engle (1999). Time-varying betas and asymmetric effects of news: Empirical analysis of blue chip stocks. Working paper W7330, NBER.

Cobbs, R. and A. Wolf (2004). Jet fuel hedging strategies: Options available for airlines and a survey of industry practices. Unpublished manuscript, `http://www.kellogg.northwestern.edu/research/fimrc/papers/jet_fuel.pdf`.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods.* Oxford: Oxford University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations. *Econometrica 50*, 987–1008.

Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics 20*(3), 339–350.

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

Engle, R. F. and K. Sheppard (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Working paper 8554, NBER.

Faff, R. W., D. Hillier, and J. Hillier (2000). Time varying beta risk: An analysis of alternative modelling techniques. *Journal of Business Finance and Accounting 27*, 523–554.

Harvey, A. C., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies 61*(2), 247–264.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics 12*(4), 371–417.

Jorion, P. (1997). *Value at Risk: The New Benchmark for Controlling Market Risk.* New York: McGraw-Hill.

Jungbacker, B. and S. J. Koopman (2007). Monte Carlo estimation for nonlinear non-Gaussian state space models. *Biometrika*. Forthcoming.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance 7*(1), 77–91.

Ord, J. K., R. D. Snyder, A. B. Koehler, R. J. Hyndman, and M. Leeds (2005). Time series forecasting: The case for the single source of error state space approach. Working Paper 7/05, Monash Econometrics and Business Statistics.

Park, T. and L. Switzer (1995). Time-varying distribution and the optimal hedge ratios for stock index futures. *Applied Financial Economics 5*, 131–137.

Pelletier, D. (2006). Regime switching for dynamic correlations. *Journal of Econometrics 131*(1-2), 445–473.

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association 94*(446), 590–599.

Sandmann, G. and S. J. Koopman (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics 87*, 271–301.

Taylor, S. J. (1986). *Modelling Financial Time Series*. Chichester: Wiley.

Yu, J. and R. Meyer (2006). Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews 25*(2-3), 361–384.