Economic and financial time series typically exhibit time varying conditional (given the past) standard deviations and correlations. The conditional standard deviation is also called the volatility. Higher volatilities increase the risk of assets, and higher conditional correlations cause an increased risk in portfolios. Therefore, models of time varying volatilities and correlations are essential for risk management. GARCH (Generalized AutoRegressive Conditional Heteroscedastic) processes are dynamic models of conditional standard deviations and correlations. This tutorial begins with univariate GARCH models of conditional variance, including univariate APARCH (Asymmetric Power ARCH) models that feature the leverage effect often seen in asset returns. The leverage effect is the tendency of negative returns to increase the conditional variance more than do positive returns of the same magnitude. Multivariate GARCH models potentially suffer from the curse of dimensionality, because there are \( d(d + 1)/2 \) variances and covariances for a \( d \)-dimensional process, but most multivariate GARCH models reduce the dimensionality in some way. A number of multivariate GARCH models are reviewed: the EWMA model, the diagonal VEC model, the dynamic conditional correlations model, orthogonal GARCH, and the dynamic orthogonal components model. In a case study, the dynamic orthogonal components model was found to provide the best fit.

**INTRODUCTION**

The volatility of a time series \( Y_1, Y_2, \ldots \) at time \( n \) is the conditional standard deviation of \( Y_n \) given \( Y_1, \ldots, Y_{n-1} \). A characteristic feature of economic and financial time series is *volatility clustering* where periods of high and of low volatility occur in the data. Typically, the changes between periods of low, medium, and high volatility do not exhibit any systematic patterns and seem best modeled as occurring randomly.

Volatility clustering is illustrated in Figure 1, which is a time series plot of daily S&P 500 log returns from January 1981 through April 1991. Let \( P_n \) be the price of an asset at time \( n \). The net return on the asset at time \( n \) is the relative change since time \( n - 1 \), specifically \( (P_n - P_{n-1})/P_{n-1} \), and is usually expressed as a percentage. Financial analysts often use the log-return, which is \( \log(P_n/P_{n-1}) \). Using the approximation \( \log(1 + x) \approx x \), it is easy to show that the return and the log-return are close to each other, provided that they are both small, say under 5%, which is typical of daily returns; see Chapter 2 of [1] for more information about returns. In Figure 1, the very high volatility in October 1987 is obvious, especially on and after October 19 (Black Monday) when the return on the S&P 500 index was \(-22.8\%\), and other periods of high or low volatility can also be seen.

Further evidence of volatility clustering can be seen in the sample autocorrelation function (ACF) of the squared returns, which is plotted on the right in Figure 2. The ACF of the returns on the left side of the figure shows little autocorrelation. The squared returns show substantial positive autocorrelation because of volatility clustering, which causes the values of squared returns on days \( t \) and \( t - k \) to be similar, at least for small values of \( k \).

To model volatility clustering, one needs a stochastic process with a nonconstant conditional (given the past) variance. GARCH and stochastic volatility models have this property. GARCH models are easier to implement and are more widely used than stochastic volatility models, so, in this tutorial, we will cover only GARCH models. Stochastic volatility models are used in options pricing; see [2].

We start with univariate GARCH models and then cover the multivariate case. Multivariate GARCH models are of obvious importance to many areas of finance, e.g., managing the...
risk of a portfolio of assets. Even in the case of a single asset, volatility clustering complicates risk management, since the level of risk varies with the conditional standard deviation of the returns on the stock. With a portfolio of stocks, risk management is even more challenging because: (1) the conditional standard deviations of the returns on the stocks in the portfolio will co-evolve; (2) the correlations between the returns will also vary. Multivariate GARCH models accommodate both of these effects.

**UNIVARIATE GARCH MODELS**

To understand GARCH models, it is useful to review ARMA models. Throughout this article, \( \{\epsilon_t\}_{t=-\infty}^{\infty} \) is an independent, identically distributed (iid) sequence with mean zero and finite variance \( \sigma^2_\epsilon \). An ARMA\((p,q)\) model for the process \( Y_t \) is

\[
(Y_t - \mu) = \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \epsilon_t - \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q}
\]

The conditional mean of \( Y_t \), given the past, is \( \mu + \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \). The conditional variance of \( Y_t \) is the constant \( \sigma^2_\epsilon \), which explains why an ARMA process is not suitable for modeling volatility clustering.

In a manner analogous to the way an ARMA process models the conditional mean, a GARCH model expresses the conditional variance at time \( t \) as a linear combination of past values of the conditional variance and of the squared process. The simplest GARCH model is the ARCH(1) process \( a_1, a_2, \ldots \) where

\[
a_t = \sigma_t \epsilon_t,
\]

and

\[
\sigma^2_t = \omega + \alpha_1 a^2_{t-1}.
\]

In this model, a large value of \( |a_{t-1}| \), which is an indication of high volatility of \( a_{t-1} \), increases \( \sigma_t \), the volatility of \( a_t \). For this reason, in an ARCH(1) model volatilities cluster, with higher values of \( \alpha \) causing more volatility clustering. It is required that \( \omega > 0 \) and \( \alpha \geq 0 \) so that the conditional variance stays positive. Also, \( \alpha < 1 \) is required for stationarity, specifically to keep \( \sigma^2_t \) from exploding.

Unfortunately, the ARCH(1) model does not have enough parameters to provide a good fit to many financial time series. A simple solution to this problem is to allow \( \sigma^2_t \) to depend not only on \( a^2_{t-1} \) but also on \( \sigma^2_{t-1} \). This is done with the GARCH(1,1) model where \( a_t = \sigma_t \epsilon_t \) as before, but now

\[
\sigma^2_t = \omega + \alpha_1 a^2_{t-1} + \beta_1 \sigma^2_{t-1}.
\]

The GARCH(1,1) model can be generalized easily; \( \{a_t\}_{t=-\infty}^{\infty} \) is a GARCH\((p,q)\) process if \( a_t = \sigma_t \epsilon_t \), where

\[
\sigma^2_t = \omega + \sum_{i=1}^{p} \alpha_i a^2_{t-i} + \sum_{i=1}^{q} \beta_i \sigma^2_{t-i},
\]

\( \alpha_1, \ldots, \alpha_p \) and \( \beta_1, \ldots, \beta_q \) are all nonnegative, \( \omega > 0 \), and

\[
\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1
\]

for stationarity. Without a constraint, the parameters in the GARCH\((p,q)\) are not uniquely defined because one can multiply each of \( \omega, \alpha_1, \ldots, \alpha_p, \) and \( \beta_1, \ldots, \beta_q \) by some positive constant and divide \( \sigma^2_t \) by that constant without changing the model. To make the parameters uniquely defined, we add the constraint that \( \sigma^2_t \) equals 1.

Let \( \mathcal{F}_n \) be the information set at time \( n \), or, more precisely, the set \( \{\epsilon_n, \epsilon_{n-1}, \ldots\} \) of all information contained in \( \{\epsilon_t : t = -\infty, \ldots, n\} \), or, equivalently, in \( \{a_t : t = -\infty, \ldots, n\} \). Then, for any random variable \( X \), \( E(X|\mathcal{F}_t) \) is the conditional expectation of \( X \) given \( \{\epsilon_t : t = -\infty, \ldots, n-1\} \). It is easy to check that \( E(a_t|\mathcal{F}_{t-1}) = 0 \). More importantly, \( \text{var}(a_t|\mathcal{F}_{t-1}) = \sigma^2_t \), so the conditional variance of \( a_t \) is a linear combination of: (1) past conditional variances; and (2) past values of \( a^2_t \).

Feature (1) keeps the conditional variance smooth while feature (2) allow it to adapt to changes in the observed volatility of the process. Because \( \{\epsilon_t\}_{t=-\infty}^{\infty} \) is stationary, so also is \( \{a_t\}_{t=-\infty}^{\infty} \), and so its unconditional variance is constant, and, in fact, equal to \( \sigma^2_a := \omega/(1 - \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i)) \), as will be shown soon. Here, to simplify notation, we define \( \alpha_i = 0 \) if \( i > p \) and \( \beta_i = 0 \) if \( i > q \). Therefore, a GARCH process has a constant unconditional variance (as it must, since it is a stationary process) but a nonconstant conditional variance—we say that it is homoscedastic but conditionally heteroscedastic. The name “GARCH” is an acronym for “generalized autoregressive conditional heteroscedasticity.” Engle [3] proposed ARCH\((p)\) processes where \( q = 0 \) and the conditional heteroscedasticity is modeled as an autoregression in past values of \( a^2_t \). Bollerslev [4] extended Engle’s idea to GARCH processes, that is, \( q \geq 1 \).

For any \( k \geq 1 \), we have

\[
E(a_{t+k}a_{t-k}) = E\{E(a_{t+1}a_{t-1} | \mathcal{F}_{t-1})\} = E\{a_{t-1}E(a_{t} | \mathcal{F}_{t-1})\} = 0,
\]

Fig. 2. ACF of S&P 500 daily returns and squared returns.
so that \{a_t\}_{t=-\infty} \leq \infty is an uncorrelated process, although it is not an independent sequence since \text{var}(a_t | F_{t-1}) = \sigma_t^2 depends on past values of the process. An uncorrelated process with a constant unconditional mean and variance, such as \{a_t\}_{t=-\infty} \leq \infty is called a weak white noise process.

A GARCH model for \{a_t\}_{t=-\infty} \leq \infty is also an ARMA model for \{a_t^2\}_{t=-\infty} \leq \infty. To see this, let \eta_t = a_t^2 - \sigma_t^2, which is a weak white noise process, and, as before, let \sigma_n^2 = \omega/\{1 - \sum_{i=1}^{\max(p,q)}(\alpha_i + \beta_i)\}. Straightforward algebra shows that

\[ a_t^2 - \sigma_n^2 = \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i)(a_{t-i}^2 - \sigma_{n-i}^2) - \sum_{i=1}^{q} \beta_i \eta_{t-i} + \eta_t, \]

so that \( a_t^2 \) is an ARMA(\max(p,q),q) process with unconditional mean \( \sigma_n^2 \). Therefore, \( \sigma_n^2 \) is the unconditional variance of \( a_t \) as was claimed before.

The ACF of \( \{a_t^2\}_{t=-\infty} \leq \infty \) is readily obtained as the ACF of the corresponding ARMA process. For example, if \( \{a_t\}_{t=-\infty} \leq \infty \) is an ARCH(1) process, then \( \{a_t^2\}_{t=-\infty} \leq \infty \) is an AR(1) process, and therefore the ACF of \( \{a_t^2\}_{t=-\infty} \leq \infty \) is

\[ \rho_{a^2}(1) = \frac{\alpha_1(1 - \alpha_1 \beta_1 - \beta_2^2)}{1 - 2\alpha_1 \beta_1 - \beta_2^2} \]

and

\[ \rho_{a^2}(k) = (\alpha_1 + \beta_1)k^{-1}\rho_{a^2}(1), \quad k \geq 2. \]

These results can be derived from the ACF of the ARMA(1,1) model \( Y_t = \phi Y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t \), which is

\[ \rho_Y(1) = \frac{(1 + \phi \theta)(\phi + \theta)}{1 + \theta^2 + 2\phi \theta} \]

and \( \rho_Y(k) = \phi k^{-1}\rho(1) \), for \( k \geq 2 \), by replacing \( \phi \) by \( \alpha_1 + \beta_1 \) and \( \theta \) by \( -\beta_1 \).

The parameters of a GARCH model can be estimated by maximum likelihood (see [5]), and \( p \) and \( q \) can be selected by minimizing AIC or BIC. AIC and BIC are criteria designed to select models that provide a good fit to the data without an excessive number of parameters; such models are called parsimonious. If \( \theta \) is the vector of all parameters, \( \text{dim}(\theta) \) is the length of \( \theta \), \( \hat{\theta} \) is the MLE, and \( L(\theta) \) is the likelihood, then \( \text{AIC} = -2L(\hat{\theta}) + 2\text{dim}(\theta) \), and \( \text{BIC} = -2L(\hat{\theta}) + \log(n)\text{dim}(\theta) \), where \( n \) is the sample size and \( \log \) is the natural logarithm. Smaller values of AIC or BIC indicate better models. A small value of \(-2L(\hat{\theta})\) indicates a good fit, while \( 2\text{dim}(\theta) \) and \( \log(n)\text{dim}(\theta) \) penalize the number of parameters. Since \( \log(n) > 2 \) in any practical setting, BIC tends to select more parsimonious models than AIC. See Section 5.12 of [1] for more information about AIC and BIC. For the S&P 500 daily returns, all models with \( 1 \leq p \leq 3 \) and \( 1 \leq q \leq 5 \) were compared by AIC and BIC. AIC was minimized by the GARCH(1,4) model, and BIC was minimized by the GARCH(1,2) model. AIC has a tendency to overfit, that is, to select too many parameters, so the GARCH(1,2) model will be used for further analysis.

The standardized residuals are \( a_t/\hat{\sigma}_t \) and estimate the \{\epsilon_t\} process. They can be used for model checking. If the ACF of the squared standardized residuals shows little autocorrelation, then this is a sign of a good fit. For the S&P 500 returns, the ACFs of the standardized residuals and the squared standardized residuals are shown in Figure 3. The squared standardized residuals show little autocorrelation, which is evidence that the GARCH(1,2) model selected by BIC is adequate and the more complex model selected by AIC is not needed.

Many software packages such as SAS and Matlab have functions for fitting GARCH models. The GARCH models in this section, as well as the ARMA/GARCH and APARCH models discussed later, were estimated using the garchFit command in the fgarch package of R [6]. R has the advantages that it is very widely used by statisticians and that it is available free of charge. Examples of R code for fitting GARCH models can be found in [1] and on the web site accompanying that book:

http://people.orie.cornell.edu/~davidr/SDAFE/index.html

![Fig. 3. S&P 500 daily returns. ACFs of the standardized residuals and the squared standardized residuals from a GARCH(1,2) model.](image-url)

**COMBINING GARCH AND ARMA MODELS**

There is some first order autocorrelation in the standardized residuals in Figure 3, and this suggests that the returns are correlated and so are not a GARCH process. This problem is remedied by combining an ARMA model for the conditional mean with a GARCH model for the conditional standard deviation. The combined model, which will be called an ARMA(\( p_A, q_A \))/GARCH(\( p_G, q_G \)) model, is

\[
(\epsilon_t - \mu) = \phi_1(\epsilon_{t-1} - \mu) + \cdots + \phi_{p_A}(\epsilon_{t-p_A} - \mu) + \alpha_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_{q_A} \epsilon_{t-q_A}
\]

where \( \{\epsilon_t\} \) is a GARCH(\( p_G, q_G \)) process.
An ARMA(0,1)/GARCH(1,2) process was fit to the S&P 500 returns and neither the standardized residuals nor squared standardized residuals showed evidence of autocorrelation. The conditional standard deviations estimated by the model are shown in Figure 4 for the second half of 1987. The unconditional standard deviation is shown as a horizontal red line. The conditional standard deviations have a sharp peak on the day after Black Monday. Notice, however, that they had been rising steadily before Black Monday, and on Black Monday the conditional standard deviation was over twice as large as the unconditional standard deviation.

![Graph showing S&P 500 daily returns and estimated conditional standard deviations](image)

**Fig. 4.** S&P 500 daily returns. Estimated conditional standard deviations during the second half of 1987. The horizontal red line shows the unconditional standard deviation. The sharp peak occurs on the day after Black Monday.

**BLACK MONDAY**

How probable was the large negative return on October 19, 1987 when the S&P 500 lost 22.8% of its value? Define a “Black Monday event” as a daily return less than or equal to −22.8%. The probability of a Black Monday event calculated using a Gaussian distribution and the unconditional variance is 4.1 × 10⁻⁹⁸, so we would expect a Black Monday event only once in 9.63 × 10⁹⁴ years! Obviously, something must be wrong with this calculation. In fact, two things are wrong, the Gaussian assumption and the use of the unconditional variance. Black Monday’s return was 21 unconditional standard deviation below 0, but only 9.9 conditional standard deviations below 0. Moreover, daily stock returns have much heavier tails than the Gaussian distribution and have been observed to be approximately t-distributed with between 3 and 8 degrees of freedom (DF). A t-distribution was fit to the standardized residuals and the MLE of DF was 6. The probability that a t₆ random variable is below −9.9 is small, and only one daily return this small can be expected every 1560 years. However, Black Monday is much more unlikely if one ignores the conditional heteroscedasticity and calculates the probability using the t₆-distribution with the unconditional standard deviation—then only one Black Monday event is expected every 21,600 years.

The standardized residuals also show some left skewness, and we should take that into account. If we fit a skew t distribution [7] to the standardized residuals, then the estimated probability of a Black Monday event increases and we can expect one event every 417 years. In summary, if we take into account conditional heteroscedasticity, heavy tails, and left skewness, a Black Monday event changes from being completely implausible to merely a very rare event. We do not expect many Black Monday events, but having one in the 20th century was plausible.

The probabilities reported in this section were calculated using parameters estimated from the entire series, including Black Monday and the period of very high volatility that followed. The results are somewhat different if only the data prior to Black Monday are used. The estimated conditional standard deviation for Black Monday is 37% higher using the entire series compared to using only the data prior to Black Monday. Using only the prior data, the degrees of freedom of the t-distribution fitted to the standardized residuals is 8, rather than 6. A larger degrees of freedom parameter implies lighter tails making extreme returns less likely. One conclusion from these comparisons is that estimation of extreme risks using only data from periods of low to moderate volatility can be misleading. An analyst might be reluctant to use data from 1987 when managing risks, since such data seem too old to be still relevant. However, periods of extreme volatility are rare and perhaps data from that far back are needed to estimate extreme risks.

**APARCH MODELS**

It has often been observed that a large negative return increases volatility more than does a positive return of the same size. This phenomenon is called the leverage effect and makes sense, since we can expect investors to become more nervous after a large negative return than after a positive return of the same magnitude. The GARCH models introduced so far cannot model the leverage effect, because σᵣ is a function only of past values of αᵢ², so information about the sign of αᵢ is lost. To model the leverage effect, the APARCH (asymmetric power ARCH) models replace the square function with a more flexible class of nonnegative functions that includes asymmetric functions. The APARCH(p, q) model for the conditional standard deviation is

$$\sigma_t^\delta = \omega + \sum_{i=1}^{p} \alpha_i (|a_{t-i}| - \gamma_t a_{t-i})^\delta + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^\delta,$$

where δ > 0 and −1 < γ_j < 1, j = 1, . . . , p. The parameters ω, α₁, . . . , αₚ, and β₁, . . . , βₚ satisfy the same constraints as
for a GARCH model. Note that $\delta = 2$ and $\gamma_1 = \cdots = \gamma_p = 0$ give a standard GARCH model. A positive value of $\gamma_j$ indicates a leverage effect, because for a positive $\gamma_j$, as $|x|$ increases, the function $|x| - \gamma x$ increases more quickly when $x$ is negative than when $x$ is positive—this causes the conditional variance to increase more for negative than positive values of $a_{t-i}$.

![Graph](image.png)

**Fig. 5.** The function $|x| - \gamma x$ with $\gamma = 0$ (no leverage) and $\gamma = 0.346$ (MLE for the S&P 500 daily returns).

When an ARMA(0,1)/APARCH(1,2) model was fit to the S&P 500 daily returns, the MLE of $\gamma_1$ was 0.346 with a 95% confidence interval of (0.17, 0.52). Since this interval is well above 0, there is a statistically significant leverage effect. The size of the leverage effect is illustrated in Figure 5 that plots $|x| - \gamma x$ with $\gamma = 0$ and with $\gamma = 0.346$. The MLE of $\delta$ was 1.87 with a 95% confidence interval of (1.3, 2.4), which includes the value $\delta = 2$ of the standard GARCH model. The ARMA(0,1)/APARCH(1,2) had the smallest value of BIC of all models considered, despite having two more parameters ($\gamma_1$ and $\delta$) than the ARMA(0,1)/GARCH(1,2) model and three more parameters than the GARCH(1,2) model.

### Multivariate GARCH Models

As financial asset returns evolve, they tend to move together. Their respective volatilities also tend to move together over time, across both assets and markets. Modeling a time-varying conditional covariance matrix, referred to as the volatility matrix, is important in many financial applications, including asset pricing, portfolio selection, hedging, and risk management. See [8] and [9] for a thorough literature review.

Multivariate volatility models all face major difficulties. First, the curse of dimensionality; there are $d(d + 1)/2$ variances and covariances for a $d$-dimensional process, e.g., 55 for $d = 10$, all of which are latent and may vary with time. Many parameterizations for the evolution of the volatility matrix use such a large number of parameters that estimation becomes infeasible for $d > 10$. In addition to empirical adequacy (i.e., goodness-of-fit of the model to the data), ease and feasibility of estimation are important considerations for any practitioner.

Analogous to positivity constraints in univariate GARCH models, a well-defined multivariate volatility matrix process must be positive-definite at every time point. This property should extend to model-based forecasts as well. From a practical perspective, the inverse of a volatility matrix is often needed for applications. Additionally, a positive conditional variance estimate for a portfolio’s return (i.e., linear combination of asset returns) is essential; fortunately, this is guaranteed by positive definiteness.

### Multivariate Conditional Heteroscedasticity

Figures 6a and 6b are time series plots of daily returns (in percentage) for IBM stock and the Center for Research in Security Prices (CRSP) value-weighted index, including dividends, from January 3, 1989 to December 31, 1998, respectively. Each series clearly exhibits volatility clustering. Let $y_t$ denote the vector time series of these returns.

![Graph](image.png)

**Fig. 6.** Daily returns (in percentage) for (a) IBM stock and (b) the CRSP value-weighted index, including dividends.

Figures 7a and 7b are the sample ACF plots for the IBM stock and CRSP index returns, respectively. These returns will be used as an example for the remainder of the paper. There is some evidence of minor serial correlation at low lags. Cross-dependence between processes is also important in multivariate time series. We first consider the lead-lag linear relationship between pairs of returns. Figure 7c is the
sample cross-correlation function (CCF) between IBM and CRSP. The lag zero estimate for contemporaneous correlation of 0.49 is very significant. There is also some evidence of minor cross-correlation at low lags. A Box-Ljung [10] test can be applied to simultaneously test that the first \( m \) autocorrelations, as well as the lagged cross-correlations in the multivariate case [11], are all zero. The multivariate Box-Ljung test statistic at lag five is \( Q_{d=2}(y_t, m = 5) = 50.12 \), which has a \( p \) value very close to zero. A \( p \) value [1, p. 618] represents the probability of obtaining a test statistic at least as extreme, 50.12 or greater in this case, assuming that the null hypothesis, zero serial correlation for lags \( 1, \ldots, m \) in this case, is true. Here, the \( p \) value very close to zero provides strong evidence to reject the null hypothesis and indicates there is significant serial correlation for the vector process.

![Fig. 7. ACFs for (a) the IBM stock and (b) CRSP index returns; (c) CCF between IBM and CRSP returns.](image)

For simplicity, we use ordinary least squares to fit a vector AR(1) model to remove the minor serial correlation and focus on the conditional variance and covariance. Let \( \hat{e}_t \) denote the estimated residuals from the regression; \( \hat{e}_t \) estimates the innovation process \( e_t \) which is described more fully in the next section. The multivariate Box-Ljung test statistic at lag five is now \( Q_2(\hat{e}_t, 5) = 16.21 \), which has a \( p \) value of 0.704, indicating there is no significant serial correlation in the vector residual process.

Although the innovation process \( e_t \) is serially uncorrelated, Figure 8 shows it is not an independent process. Figures 8a and 8b are sample ACF plots for the squared processes \( e_t^2 \). They both show substantial positive autocorrelation because of the volatility clustering. Figure 8c is the sample CCF for the squared processes; this figure shows there is a dynamic relationship between the squared processes at low lags. Figure 8d is the sample ACF for the product process \( e_{1t} e_{2t} \) and shows that there is also positive autocorrelation in the conditional covariance series. The multivariate volatility models described below attempt to account for these forms of dependence exhibited in the vector time series of innovations.

![Fig. 8. ACFs of squared innovations for (a) IBM and (b) CRSP; (c) CCF between squared innovations; (d) ACF for the product of the innovations.](image)

**Basic Setting**

Let \( y_t = (y_{1t}, \ldots, y_{dt})' \) denote a \( d \)-dimensional vector process and let \( F_t \) denote the information set until time index \( t \), i.e., \( \sigma(y_t, y_{t-1}, \ldots) \). We partition the process as

\[
y_t = \mu_t + e_t,
\]

in which \( \mu_t = \mathbb{E}(y_t | F_{t-1}) \) is the conditional mean and \( e_t \) is the mean zero serially uncorrelated innovation vector. Let \( \Sigma_e = \text{Cov}(e_t) \) be the unconditional covariance matrix of \( e_t \). Let \( \Sigma_t = \text{Cov}(e_t | F_{t-1}) = \text{Cov}(y_t | F_{t-1}) \) denote the conditional covariance matrix. For a stationary process the unconditional mean and covariance matrix are constant, even though the conditional mean and covariance matrix may be non-constant. Multivariate time series modeling is concerned with the time evolutions of \( \mu_t \) and \( \Sigma_t \), the conditional mean and covariance matrix.

In our analysis we assume that \( \mu_t \) follows a vector autoregression \( \mu_t = \phi_0 + \sum_{i=1}^{P} \phi_i y_{t-i} \), where \( P \) is a non-negative integer, and \( \phi_0 \) and the \( \phi_i \) are \( d \times 1 \) and \( d \times d \) coefficient matrices, respectively. We took \( P = 1 \) in the current example. The relationship between the innovation process and the volatility process is defined by

\[
e_t = \Sigma_t^{1/2} \varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} F(0, I_d),
\]

in which \( \Sigma_t^{1/2} \) is the matrix square-root of \( \Sigma_t \) and the standardized iid innovations \( \varepsilon_t \) are from a standardized multivari-
ate distribution \( F \). The models below describe dynamic evolutions for the volatility matrix \( \Sigma_t \). Throughout the rest of this paper, all figures and comparisons are based on in-sample estimates which use the entire sample to calculate. Out-of-sample comparisons are also of interest, but there is no clear metric to measure their adequacy. This is because the conditional distribution is changing over time and there is only one observation available at each time index, making it impossible to directly calculate conditional variances and covariances at each time point.

**EWMA Model**

The simplest matrix generalization of a univariate volatility model is the exponentially weighted moving average (EWMA) model. It is indexed by a single parameter \( \lambda \in (0, 1) \) and is defined by the recursion

\[
\Sigma_t = (1 - \lambda) e_{t-1}' e_{t-1} + \lambda \Sigma_{t-1}
\]

\[
= (1 - \lambda) \sum_{m=1}^{\infty} \lambda^{m-1} e_{t-m}' e_{t-m}.
\]

When the recursion is initialized with a positive-definite (p.d.) matrix the sequence remains p.d. This single parameter model is simple to estimate regardless of the dimension, with large values of \( \lambda \) indicating high persistence in the volatility process. However, the dynamics can be too restrictive in practice, since the component-wise evolutions all have the same discounting factor (or, persistence parameter) \( \lambda \).

Figure 9 shows the in-sample fitted EWMA model for \( \hat{e}_t \) assuming a multivariate standard normal distribution for \( \varepsilon_t \) and using conditional maximum likelihood estimation. The estimated conditional standard deviations are shown in (a) and (d), and the conditional covariances and implied conditional correlations are shown in (b) and (c), respectively. The persistence parameter \( \lambda \) was estimated as 0.985.

**Diagonal VEC Model**

The vectorization model of [12] specifies a vector ARMA type model for a vector process containing all the conditional variances and covariances. The diagonal form of this model (DVEC) simply specifies univariate representations for each component process. Analogous to the GARCH(1,1) model, the (1,1) form of the DVEC model is given by the recursions

\[
\sigma_{ii,t}^2 = \omega_i + \alpha_e e_{i,t-1}' e_{i,t-1} + \beta_i \sigma_{ii,t-1}^2,
\]

\[
\sigma_{ij,t} = \omega_{ij} e_{i,t-1}' e_{j,t-1} + \beta_i e_{i,t-1}^2 + \beta_j e_{j,t-1}^2 + \beta_{ij} \sigma_{ij,t-1},
\]

for \( i \neq j \), in which \( \sigma_{ii,t}^2 = \Sigma_{ii,t} \) and \( \sigma_{ij,t} = \Sigma_{ij,t} = \Sigma_{ji,t} \).

Although univariate models are specified, the model still requires joint estimation of all the parameters. Combining contemporaneous marginal estimates of the conditional variances and covariances does not, in general, assure that the covariance matrix is positive definite (p.d.). Conditions for p.d. estimates from the DVEC model are straight-forward to check, but difficult to incorporate into a constrained estimation problem. Figure 10 shows the fitted DVEC model for \( \hat{e}_t \) using the multivariate standard normal distribution and conditional maximum likelihood estimation. The estimated conditional standard deviations are shown in (a) and (d), and the conditional covariances and implied conditional correlations are shown in (b) and (c), respectively.

A true matrix generalization of the univariate GARCH model is given by [13]; it is commonly refereed to as the BEKK model in reference to the authors. It has a parameterization to assure a positive definite estimated covariance matrix. However, the model parameters are difficult to interpret. Also, the model parameters are difficult to estimate even for very low dimensional processes, so we will not discuss these models further.

**Dynamic Conditional Correlation Models**

Nonlinear combinations of univariate volatility models have been proposed to allow for time-varying correlations, a feature that has been documented in financial applications. Both [14] and [15] generalize the constant correlation model of [16] to allow for such dynamic conditional correlations (DCC).

Analogously to the GARCH(1,1) model, the first order form of the DCC model in [15] may be represented by the
In the form above, the variance components only condition on their own individual lagged returns and not the joint returns. Also, the dynamics for each of the conditional correlations are constrained to have equal persistence parameters, similar to the EWMA model. [However, equal persistence parameters may be more suitable for the scale-free correlation matrix (as in the DCC model) than for the covariance matrix (as in the EWMA model).] An explicit parameterization of the conditional correlation matrix \( R_t \), with flexible dynamics, is just as difficult to estimate in high dimensions as \( \Sigma_t \) itself. Figure 11 shows a fitted DCC model for \( \epsilon_t \) using quasi-maximum likelihood estimation. The estimated conditional standard deviations are shown in (a) and (d), and the conditional covariances and conditional correlations are shown in (b) and (c), respectively.

**Orthogonal GARCH Model**

Several factor and orthogonal models have been proposed to reduce the number of parameters and parameter constraints by imposing a common dynamic structure on the elements of the volatility matrix. The orthogonal GARCH (O-GARCH) model of [17] is among the most popular because of its simplicity. It is assumed that the innovations \( \epsilon_t \) can be decomposed into orthogonal components \( z_t \) via a linear transformation \( U \). This is done in conjunction with principal component

---

**Fig. 10.** A fitted first order DVEC model with \((\omega_1, \alpha_1, \beta_1) = (0.013, 0.023, 0.974), (\omega_2, \alpha_2, \beta_2) = (0.011, 0.046, 0.933), (\omega_{12}, \alpha_{12}, \beta_{12}) = (0.009, 0.033, 0.952)\). The red line in (c) is the sample correlation estimate over the previous six months for comparison.

**Fig. 11.** A fitted first order DCC model with \((\omega_1, \alpha_1, \beta_1) = (0.0741, 0.0552, 0.9233), (\omega_2, \alpha_2, \beta_2) = (0.0206, 0.0834, 0.8823), \) and \( \lambda = 0.9876 \). The red line in (c) is the sample correlation estimate over the previous six months for comparison.
analysis (PCA) as follows. Let $Y$ be the matrix of eigenvectors and $Λ$ the diagonal matrix of the corresponding eigenvalues of $Σ_e$. Then, take $U = Λ^{-1/2} Y'$, and let

$$ z_t = U e_t. $$

The components are constructed such that $Cov\{z_t\} = I_d$. The sample estimate of $Σ_e$ is typically used to estimate $U$.

Next, univariate GARCH(1,1) models are individually fit to each orthogonal component to estimate the conditional covariance $O_t = Cov\{z_t| F_{t-1}\}$. Let

$$ σ^2_{it} = ω_i + α_i z^2_{i,t-1} + β_i σ^2_{it-1}, \quad O_t = \text{diag}\{σ^2_{1t}, \ldots, σ^2_{dt}\}, \quad Σ_t = U^{-1} O_t U^{-1}'. $$

In summary, a linear transformation $U$ is estimated, using PCA, such that the components of $z_t = U e_t$ have zero (unconditional) correlation. It is then also assumed that the conditional correlations of $z_t$ are also zero; however, this is not at all assured to be true. Under this additional stronger assumption, the conditional covariance matrix for $z_t$, $O_t$ is diagonal. For simplicity, univariate models are then fit to model the conditional variance $σ^2_{it}$ for each component of $z_t$.

The main drawback of this model is that the orthogonal components are uncorrelated unconditionally, but they may still be conditionally correlated. The O-GARCH model implicitly assumes the conditional correlations for $z_t$ are zero. Figure 12 shows a fitted O-GARCH model for $e_t$ using PCA followed by univariate conditional maximum likelihood estimation. The estimated conditional standard deviations are shown in (a) and (d), and the conditional covariances and conditional correlations are shown in (b) and (c), respectively. The implied conditional correlations do not appear adequate for this fitted model compared to the sample correlation estimate over the previous six months (used as a proxy for the conditional correlation).

Dynamic Orthogonal Components Models

To properly apply univariate modeling after estimating a linear transformation in the spirit of the O-GARCH model above, the resulting component processes $s_t$ must not only be orthogonal contemporaneously, the conditional correlations must be zero. Additionally, the lagged cross-correlations for the squared components must also be zero. In [18], if the components of a time series $s_t$ satisfy these conditions, then they are called dynamic orthogonal components (DOCs) in volatility.

Let $s_t = (s_{1t}, \ldots, s_{dt})'$ denote a vector time series of DOCs. Without loss of generality, $s_t$ is assumed to be standardized such that $E\{s_{it}\} = 0$ and $Var\{s_{it}\} = 1$ for $i = 1, \ldots, d$. A Ljung-Box statistic, defined below, is used to test for the existence of DOCs in volatility. Including lag zero in the test implies that the conditional covariance between stationary DOCs is zero since the Cauchy-Schwarz inequality gives

$$ |Cov\{s_{it} s_{jt}| s_{i,t-\ell} s_{j,t-\ell}\}| ≤ Var\{s_{it} s_{jt}\} = E\{s^2_{it} s^2_{jt}\}, $$

since $E\{s_{it} s_{jt}\} = E\{s_{i,t-\ell} s_{j,t-\ell}\} = 0$ by the assumption of a DOC model. Let $ρ^2_{i,j}(\ell) = Corr\{s^2_{it}, s^2_{jt}| s_{i,t-\ell}, s_{j,t-\ell}\}$. The joint lag-$m$ null and alternative hypotheses to test for the existence of DOCs in volatility are

$$ H_0: \quad ρ^2_{i,j}(\ell) = 0 \text{ for all } i \neq j, \ell = 0, \ldots, m $$

$$ H_A: \quad ρ^2_{i,j}(\ell) ≠ 0 \text{ for some } i \neq j, \ell = 0, \ldots, m. $$

The corresponding Ljung-Box test statistic is

$$ Q^2_d (m) = n \sum_{i<j} ρ^2_{i,j}(0)^2 + n(n+2) \sum_{k=1}^{m} \sum_{i\neq j} ρ^2_{i,j}(k)^2 / (n-k). $$

Under $H_0$, $Q^2_d (m)$ is asymptotically distributed as a $χ^2$ distribution with $d(d-1)/2 + md(d-1)$ degrees of freedom. The null hypothesis is rejected for a large value of $Q^2_d (m)$. When $H_0$ is rejected, one must seek an alternative modeling procedure.

As expected, the DOCs in volatility hypothesis is rejected for the innovations. The test statistic is $Q^2_2 (e^2, 5) = 356.926$.
with a $p$ value near zero. DOCs in volatility is also rejected for the principal components used in the O-GARCH model, the test statistic is $Q^0_d(z^2, 5) = 135.492$ with a $p$ value near zero. Starting with the uncorrelated principal components $z_t$, [18] propose estimating an orthogonal matrix $W$ such that the components $s_t = Wz_t$ are as close to DOCs in volatility as possible for a particular sample. This is done by minimizing a reweighted version of the Ljung-Box test statistic, defined above, with respect to the separating matrix $W$. The null hypothesis of DOCs in volatility is accepted for the estimated components $s_t$. The test statistic is $Q^0_d(s^2, 5) = 7.845$ with a $p$ value equal to 0.727.

After DOCs are identified, a univariate volatility model is considered for each process $\sigma^2_{it}$. The following model was fit

\[
\begin{align*}
\sigma^2_{it} &= MV_t^{1/2} \epsilon_t, \\
V_t &= \text{diag} \left( \sigma_{1t}^2, \ldots, \sigma_{dt}^2 \right), \quad \epsilon_t \sim t_{n_t}(0, 1) \\
\rho_{it} &= \omega_i + \alpha_i \epsilon_{it-1}^2 + \beta_i \sigma^2_{it-1} \\
\Sigma_i &= MV_tM',
\end{align*}
\]

in which $t_{n_t}(0, 1)$ denotes the standardized Student-$t$ distribution with $n_t$ degrees of freedom. Each $\Sigma_i$ is positive-definite if $\sigma^2_{it} > 0$ for all components. The fundamental motivation is that empirically the dynamics of $\epsilon_t$ can often be well approximated by an invertible linear combination of DOCs $e_t = Ms_t$, in which $M = U^{-1}W$ by definition. In summary, $U$ is estimated by PCA to uncorrelate $e_t$, $W$ is estimated to minimize a reweighted version of $Q^0_d(m)$ defined above (giving more weight to lower lags). $U$ and $W$ are combined to estimate DOCs $s_t$, of which univariate volatility modeling may then be appropriately applied. This approach allows modeling of a $d$-dimensional multivariate volatility process with $d$ univariate volatility models, while greatly reducing both the number of parameters and the computational cost of estimation, and at the same time maintaining adequate empirical performance.

Figure 13 shows a fitted DOCs in volatility GARCH model for $\hat{e}_t$ using generalized decorrelation followed by univariate conditional maximum likelihood estimation. The estimated conditional standard deviations are shown in (a) and (d), and the conditional covariances and implied correlations are shown in (b) and (c), respectively. Unlike the O-GARCH fit, the implied conditional correlations appear adequate compared to the rolling estimator.

**Model Checking**

For a fitted volatility sequence $\hat{\Sigma}_t$, the standardized residuals are defined as

\[
\hat{\epsilon}_t = \hat{\Sigma}_t^{-1/2} \epsilon_t.
\]

To verify the adequacy of a fitted volatility model, lagged cross-correlations of the squared standardized residuals should be zero. The product process $\hat{\epsilon}_{it}\hat{\epsilon}_{jt}$ should also have no serial correlation. Additional diagnostic checks for time series are considered in [11]. Since the standardized residuals are estimated and not observed, all $p$ values given in this section are only approximate.

To check the first condition we can apply a multivariate Box-Ljung test to the squared standardized residuals. For the EWMA model, $Q_2(\hat{\epsilon}^2_t, 5) = 26.40$ with a $p$ value of 0.153, implying no significant serial correlation. For the DVEC model, $Q_2(\hat{\epsilon}^2_t, 5) = 21.14$ with a $p$ value of 0.389, implying no significant serial correlation. For the DCC model, $Q_2(\hat{\epsilon}^2_t, 5) = 10.54$ with a $p$ value 0.957, implying no significant serial correlation. For the O-GARCH model, $Q_2(\hat{\epsilon}^2_t, 5) = 30.77$ with a $p$ value 0.058. In this case, there is some minor evidence of serial correlation. For the DOC in volatility model, $Q_2(\hat{\epsilon}^2_t, 5) = 18.68$ with a $p$ value 0.543, implying no significant serial correlation.

The multivariate Box-Ljung test for the squared standardized residuals is not sensitive to misspecification of the conditional correlation structure. To check this condition, we apply univariate Box-Ljung tests to the product of each pair of standardized residuals. For the EWMA model, $Q(\hat{\epsilon}_{1t}\hat{\epsilon}_{2t}, 5) = 16.45$ with a $p$ value of 0.006. This model has not adequately accounted for the time-varying conditional correlation. For the DVEC model, $Q(\hat{\epsilon}_{1t}\hat{\epsilon}_{2t}, 5) = 7.76$ with a $p$ value of 0.170, implying no significant serial correlation. For the DCC model, $Q(\hat{\epsilon}_{1t}\hat{\epsilon}_{2t}, 5) = 8.37$ with a $p$ value of 0.137, implying

---

**Fig. 13.** A fitted first order DOCs in volatility GARCH model with $(\omega_1, \alpha_1, \beta_1, \nu_1) = (0.0049, 0.0256, 0.9703, 4.3131)$, $(\omega_2, \alpha_2, \beta_2, \nu_2) = (0.0091, 0.0475, 0.9446, 5.0297)$, and $M = ((1.5350, 0.838), (0.0103, 0.7730))$. The red line in (c) is the sample correlation estimate over the previous six months for comparison.
no significant serial correlation. For the O-GARCH model, 
\[ Q(\dot{\epsilon}_{1t}\dot{\epsilon}_{2t}, 5) = 63.09 \] 
with a \( p \) value near zero. This model also fails to account for the observed time-varying conditional correlation. For the DOC in volatility model, 
\[ Q(\dot{\epsilon}_{1t}\dot{\epsilon}_{2t}, 5) = 9.07 \] 
with a \( p \) value of 0.106, implying no significant serial correlation.

**Other Multivariate GARCH Models**

The assumption of a multivariate normal conditional distribution is typically justified by the proof in [19] of the strong consistency of the Gaussian quasi-maximum likelihood for multivariate GARCH models. However, consistent estimates of the first two conditional moments may not be sufficient for some applications and there may be non-negligible bias in small samples. It may be necessary to relax the distributional assumption, at the risk of inconsistent estimates. Use of flexible distributions, able to account for the excess kurtosis or skewness exhibited in the unconditional moments of asset returns, while not significantly increasing the estimation burden, is desirable. However, few multivariate extensions are tractable.

[20] and [21] have proposed copula GARCH models, in which a conditional copula function is specified, extending the conditional dependence beyond conditional correlation. This approach allows for flexible joint distributions in the bivariate case, but its usefulness in higher dimensions has not been studied. In particular, the dynamics of the copula function would also need to be constrained to keep estimation feasible. It is also desirable to incorporate asymmetry into a multivariate model; depending on the application, variances and covariance may react differently to a negative than to a positive innovation. [22] and [5] propose two such multivariate GARCH models which include these so called “leverage effects.”

**FURTHER READING**

Further discussion of univariate GARCH models and more examples can be found in [1]. A more advanced treatment including multivariate GARCH models is available in [23].

**ACKNOWLEDGMENTS**

David Matteson was supported by NSF Grant CMMI-0926814, and David Ruppert was supported by NSF Grant DMS-0805975. The S&P 500 returns are in the SP500 data set and the IBM and CRSP returns are in the CRSPday data set in R’s Ecdat package. The R packages fGarch and mnormt were also used in our analysis.

**AUTHORS**

David S. Matteson (dm484@cornell.edu) received a B.S. in Business from the Carlson School of Management at the University of Minnesota in 2003 with concentrations in Finance, Mathematics, and Statistics, and a Ph.D. in Statistics from the University of Chicago in 2008. He has been a Visiting Assistant Professor in the School of Operations Research and Information Engineering at Cornell University since 2008 and he was a New Researcher Fellow at the Statistical and Applied Mathematical Sciences Institute in Durham, NC from 2009 - 2010. His research interests include time series, dimension reduction, and emergency medical services information systems.

David Ruppert (dr24@cornell.edu) received a B.A. in Mathematics from Cornell University in 1970, an M.A. in Mathematics from the University of Vermont in 1973, and a Ph.D. in Statistics from Michigan State University in 1977. He was Assistant and then Associate Professor in the Department of Statistics at the University of North Carolina at Chapel Hill from 1977 to 1987 and, since 1987, he has been Professor in the School of Operations Research and Information Engineering at Cornell University, where he is now the Andrew Schultz, Jr., Professor of Engineering. He is a Fellow of the American Statistical Association and of the Institute of Mathematical Statistics. He won the Wilcoxon Award in 1986 and is an ISI highly cited researcher. Professor Ruppert is currently Editor of the Electronic Journal of Statistics. He was formerly an Associate Editor for The American Statistician, Journal of the American Statistical Association, Biometrics, and Annals of Statistics and formerly Editor of the IMS’s Lecture Notes–Monographs Series.


**REFERENCES**


