

MULTIVARIATE VOLATILITY

Zhejiang University

March 2008

Luc Bauwens

Université catholique de Louvain

Center for Operations Research and Econometrics

Financial returns

- Let p_{it} denote the price of asset i at date t .
- The (one-period, i.e. from $t - 1$ to t) return is defined by

$$y_{it} = \log p_{it} - \log p_{i,t-1} = \Delta \log p_{it} \approx \frac{p_{it} - p_{i,t-1}}{p_{i,t-1}}$$

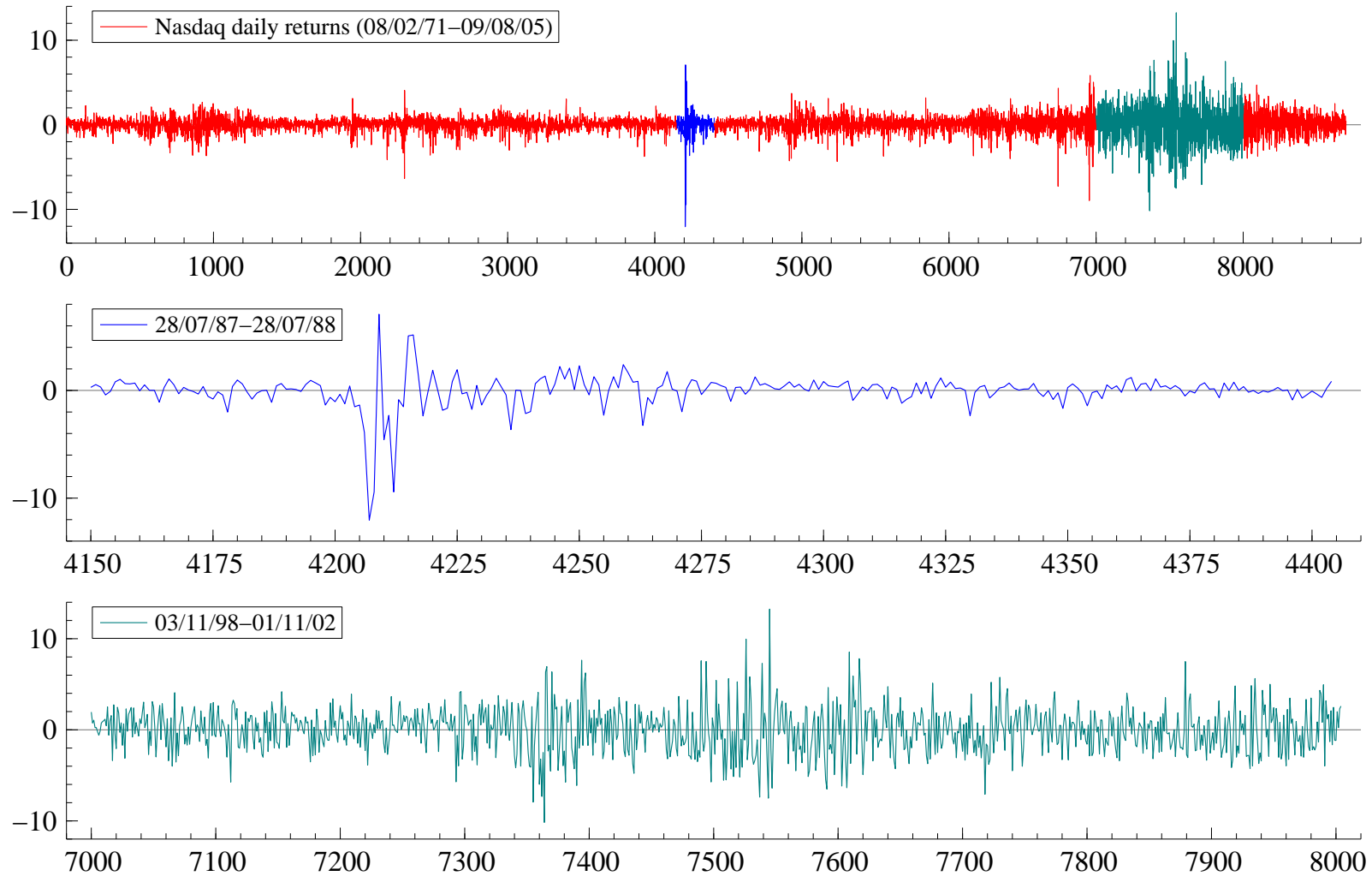
This may be multiplied by 100 to express the return in percentage.

- The frequency of a return series refers to the length of the time period between t and $t - 1$.
Low frequency (LF): less than weekly, e.g. monthly.
High frequency (HF): weekly to daily.
Ultra high frequency (UHF): intra-daily (less than daily).

Notations

$$p_t = \begin{pmatrix} p_{1t} \\ p_{2t} \\ \vdots \\ p_{Nt} \end{pmatrix} \quad y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Nt} \end{pmatrix} = \Delta \log p_t$$

Plots of Nasdaq daily returns



Volatility

- The volatility of an asset alludes to the fact that the asset price and therefore its associated return are random. Randomness implies we can never be sure of the future value of the asset (or the return).
- Technically, the volatility of an asset is the standard deviation (or the variance) of the return considered as a random variable.
- Likewise, the co-volatility between two assets is the covariance between the two corresponding returns.
- Returns and (co-)volatilities are defined relatively to a time period of some length (day, week, year, ...).

Volatility at different frequencies

- $p_t, p_{t+1}, \dots, p_{t+5}$: price vectors at the end of 6 days.
NB: $p_t = (p'_{1t} \ p_{2t} \ \dots \ p_{Nt})'$, and $y_t = (y_{1t} \ y_{2t} \ \dots \ y_{Nt})$;
- $y_t, y_{t+1}, \dots, y_{t+5}$: 5 corresponding return vectors.
- The 5-day return vector, between days $t + 1$ and $t + 5$, is
 $y_{t+5}^w = y_{t+1} + y_{t+2} + y_{t+3} + y_{t+4} + y_{t+5} = \log p_{t+5} - \log p_t$.
- If the daily returns are not autocorrelated, have the same mean vector μ and the same variance matrix Σ :
$$E(y_{t+5}^w) = 5\mu \text{ and } \text{Var}(y_{t+5}^w) = 5\Sigma.$$
- The volatility at a given frequency is (approximately) equal to k times the volatility at the k times higher frequency.

What is Σ ?

- $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \dots & \sigma_{N1} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \dots & \sigma_{N2} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \dots & \sigma_{N3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \sigma_{N3} & \dots & \sigma_N^2 \end{pmatrix} = \mathbf{E}[(y - \mu)(y - \mu)'].$

Σ must be a positive-definite matrix.

- Remember that

$$\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$$

where ρ_{ij} , the correlation coefficient between y_i and y_j , is comprised between -1 and +1. This is a necessary, but not a sufficient condition for Σ to be positive-definite.

Univariate example: Nasdaq returns

	daily	5-days	25 days
Nb. Obs.	8 700	1 740	348
Mean	0.035	0.177	0.88
Variance	1.44	8.58	51.6

$$0.177/0.035 \simeq 0.88/0.177 \simeq 5$$

$$8.58/1.44 \simeq 51.6/8.58 \simeq 2.45 \simeq 6 > 5$$

Volatility is latent

- Volatility is inherently not observable.
- Like returns, it evolves randomly through time. This does not imply that it is not predictable.
- There is a double uncertainty in financial markets: about returns, and about the volatility of returns.
- Volatility can be indirectly measured, since returns are observed. A measure can be "model-free" or "model-based".
- How to infer about volatility from returns has been and will be the main research theme of financial econometrics for long.

Historical measures

Rolling windows empirical variance matrix based on r most recent returns:

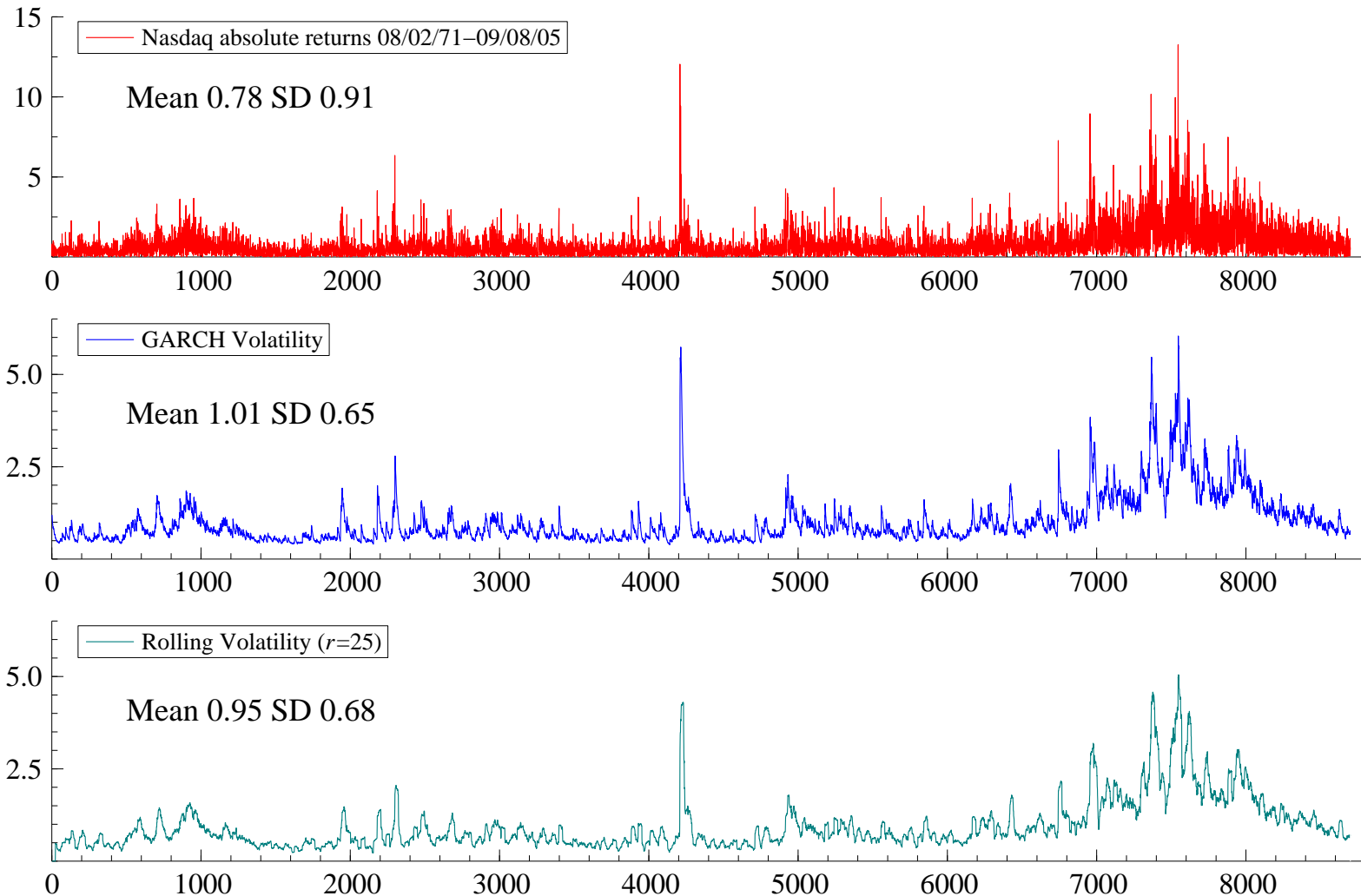
$$\hat{\sigma}_{ijt}(r) = \frac{1}{r} \sum_{l=0}^{r-1} (y_{i,t-l} - \bar{y}_{it})(y_{j,t-l} - \bar{y}_{jt}) \quad \text{where} \quad \bar{y}_{it} = \frac{1}{r} \sum_{l=0}^{r-1} y_{i,t-l}.$$

In matrix notation,

$$\hat{\Sigma}_t(r) = \frac{1}{r} \sum_{l=0}^{r-1} (y_{t-l} - \bar{y}_t)(y_{t-l} - \bar{y}_t)'$$

The choice of r is arbitrary, but $r > N$ is required for $\hat{\Sigma}_t(r)$ to be positive-definite (PD).

Nasdaq daily volatilities: GARCH and rolling



Correlations: GARCH and rolling: 0.96
Absolute returns and any other: 0.56

Realized volatility

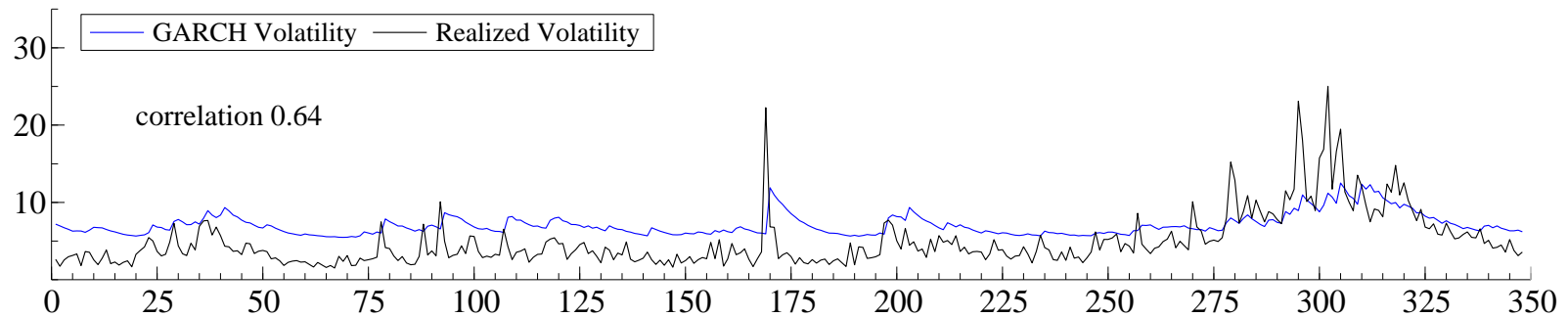
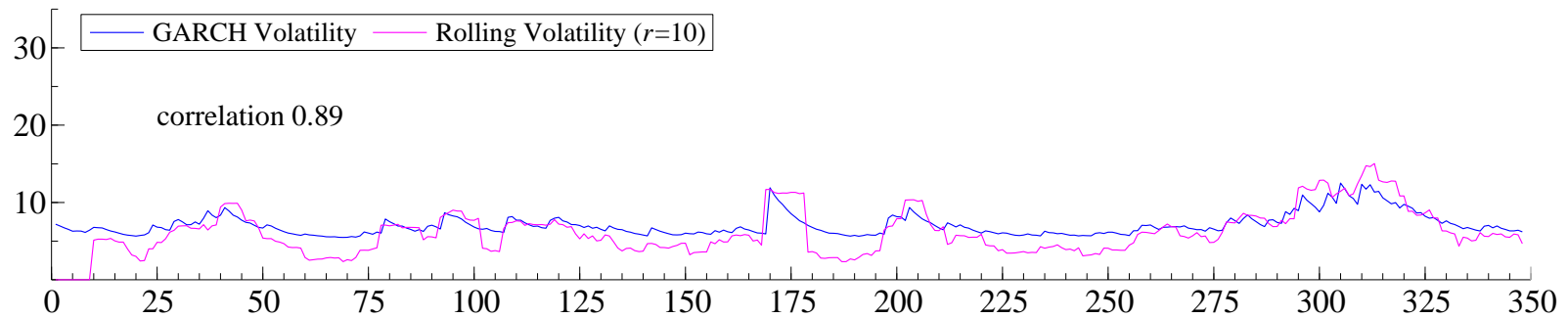
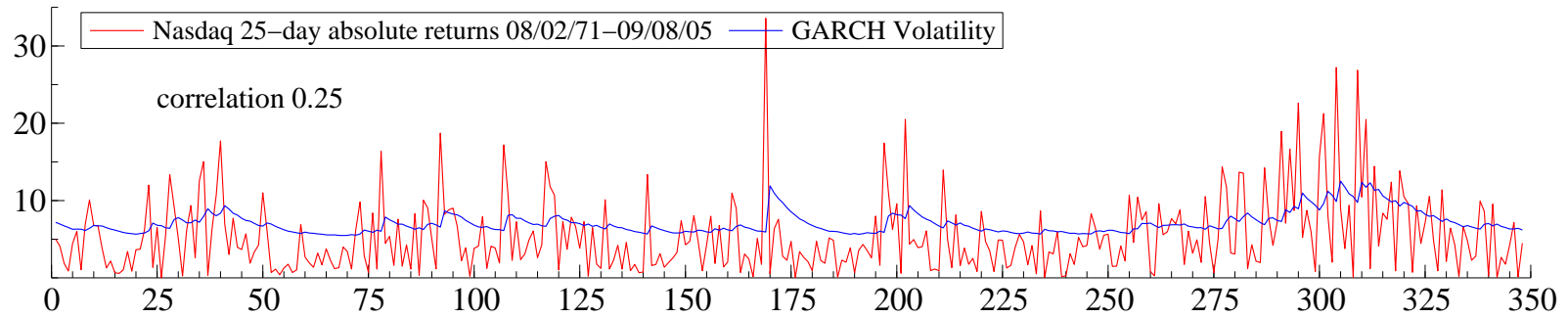
- The realized variance matrix is an estimator of the volatility matrix Σ_t at a given frequency (e.g. daily), based on higher frequency returns (e.g. every 15 minutes within the day).
- $y_{t1}, y_{t2}, \dots, y_{tk}$: k intra-daily return vectors for N assets in day t . The daily return vector is $y_t = \sum_{h=1}^k y_{th}$.
- If the intra-daily return vectors are IID,

$\Sigma_t = \text{Var}(y_t) = k \text{Var}(y_{th})$, and since

$$\frac{1}{k} \sum_{h=1}^k (y_{th} - \bar{y}_t)(y_{th} - \bar{y}_t)' \xrightarrow{p} \text{Var}(y_{th}), \text{ with } \bar{y}_t = \frac{1}{k} \sum_{h=1}^k y_{th}$$

$$\Rightarrow \sum_{h=1}^k (y_{th} - \bar{y}_t)(y_{th} - \bar{y}_t)' \approx \underbrace{\sum_{h=1}^k y_{th} y_{th}'}_{RVOL_t} \xrightarrow{p} \Sigma_t.$$

Nasdaq 25-day volatilities



Prediction using model-free measures

- These measures can be used for prediction.
- Either use date t measure as a prediction of date $t + 1$ volatility.
- Or use a model for predicting these measures, from time series of their past values. Such a model should ensure that the predicted variance matrices are PD.
- A VAR model for the elements of $\hat{\Sigma}_t(r)$ or $RVOL_t$ is not satisfactory in this respect, but a VAR model for their Cholesky factorization is fine.
- The WAR (Wishart autoregressive) process of Gouriéroux et al. ensures PD matrices.

Model-based measures of volatility

- RiskMetrics.
- Multivariate GARCH models.
- Multivariate stochastic volatility models.
- Each type of model specifies Σ_t as a parametric function of past returns and maybe other information. Given the relevant information and an estimate of the parameters, prediction is feasible.
- Again, the challenge of ensuring a PD variance matrix arises.

MOTIVATIONS FOR MULTIVARIATE VOLATILITY PREDICTION

PART 1

- Portfolio allocation
- Value-at-Risk in a multivariate context

Portfolios

- Consider an investment, worth W euros, in a portfolio made up of N securities.
- Let w_i denote the share of W invested in asset i .
Let $w' = (w_1 \ w_2 \ \dots \ w_N)$ be the vector of shares, so that $\sum_{i=1}^N w_i = 1$.
- Let $y = (y_1 \ y_2 \ \dots \ y_N)'$ be the vector of returns, $\mu = (\mu_1 \ \mu_2 \ \dots \ \mu_N)$ the vector of expected returns, and Σ the variance(-covariance) matrix of y . Then,

$$y_w = w'y \quad (\text{portfolio return})$$

$$E(y_w) := \mu_w = w'\mu \quad (\text{portfolio expected return})$$

$$\text{Var}(y_w) := \sigma_w^2 = w'\Sigma w \quad (\text{variance of portfolio return}).$$

Benefit of diversification

- It is folk wisdom that "one should not put one's eggs in the same basket". Wealth should be diversified among assets to reduce the risk..
- For example, with two securities held in proportions $w_1 = 1 - w_2$ and w_2 ,

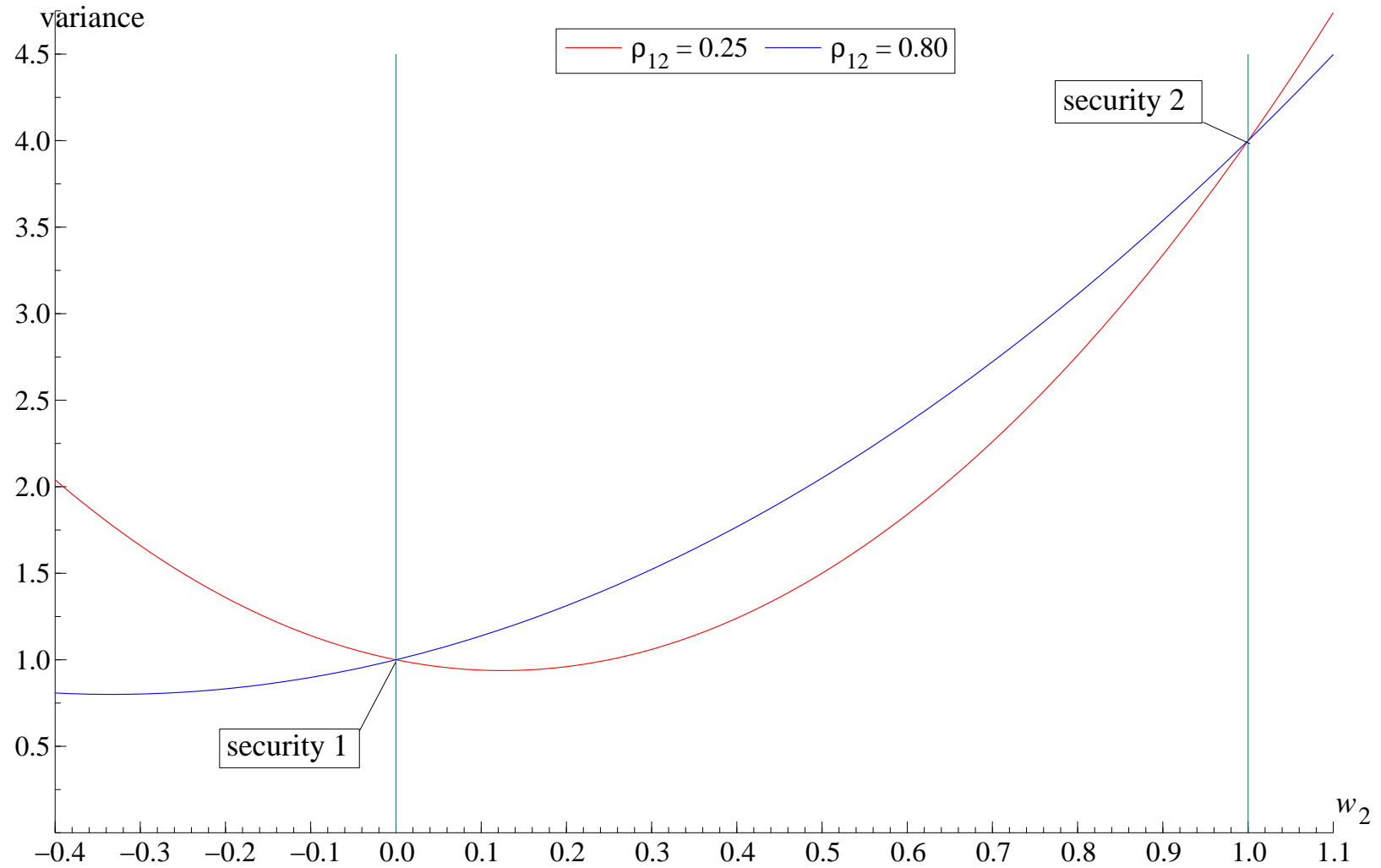
$$\sigma_w^2 = (1 - w_2)^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2(1 - w_2)w_2 \rho_{12} \sigma_1 \sigma_2$$

is smallest at

$$\bar{w}_2 = \frac{\sigma_1^2 - \rho_{12} \sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho_{12} \sigma_1 \sigma_2}.$$

which can be positive or negative. A negative value corresponds to short-selling.

Graphical illustration



Optimal portfolio choice

- An investor may choose his allocation w so as to maximize the expected return ($w'\mu$) of the portfolio under the constraints that i) the variance ($w'\Sigma w$) does not exceed some risk threshold $\bar{\sigma}^2$ and ii) $\sum w_i = 1$.
- Equivalently, the investor may choose w so as to maximize the mean-variance utility function $w'\mu - \frac{1}{2}\theta w'\Sigma w$ subject to $\sum w_i = 1$, where $\theta > 0$ is a parameter that represents the relative importance which the investor attaches to the risk relative to the expected return.
- The two problems have the same solution. To any possible value of the risk-aversion parameter θ corresponds a value of $\bar{\sigma}^2$: increasing the risk-aversion corresponds to decreasing the maximum risk threshold.

The optimal solution

- The optimal weight vector is a function of θ , μ and Σ :

$$w_* = \frac{1}{\theta} \Sigma^{-1} \left[\mu - \frac{\iota' \Sigma^{-1} \mu}{\iota' \Sigma^{-1} \iota} \iota \right] - \frac{\Sigma^{-1} \iota}{\iota' \Sigma^{-1} \iota}$$

where $\iota' = (1 \ 1 \ \dots \ 1)$ (a vector of dimension N).

- To implement this solution, the investor must know μ and Σ . They must be estimated and predicted.

Implementation

- Even if μ and Σ were constant, with N large, a lot of information is needed. For $N = 100$, we need 100 expected returns, and Σ has 5050 elements!
- We know that μ and Σ are changing through time. Multivariate, dynamic, approaches are needed.
- L. Bauwens, E. Rengifo and W. Ben Omrane (2006), "Intra-daily FX optimal portfolio allocation", CORE DP 2006/10.

PART 1

- Portfolio allocation
- Value-at-Risk in a multivariate context

Value-at-Risk (VaR) of a portfolio

- The VaR of a portfolio of financial assets is a quantitative measure of risk attached to the value of the portfolio.
- A bank holding a portfolio incurs the risk of a loss of value over a certain time horizon.
- The Basel banking regulations require banks to set aside a certain amount of capital as a reserve against such risks. Banks must nowadays evaluate and predict the VaR of their financial assets.

Value-at-Risk

- A bank might evaluate the ten day VaR of its portfolio at 20 millions euros at the 95% statistical confidence level (or, equivalently, at the 5% risk level). In other words, there is 5-in-100 chances that a loss greater than 20 millions will occur after 10 days, under normal market conditions.
- This amount is based on a statistical risk level (like 5%), a time horizon (like ten days), and a model for evaluating the risk.
- The VaR is derived from a quantile of the (conditional) distribution of the portfolio return.

Definition of VaR

- The VaR at level α of a portfolio worth one unit of investment is the absolute value of the α -% left quantile of the return distribution:

$$VaR_w(\alpha) = -q_w(\alpha) = -\inf_x \{x : \Pr(y_w \leq x) \geq \alpha\},$$

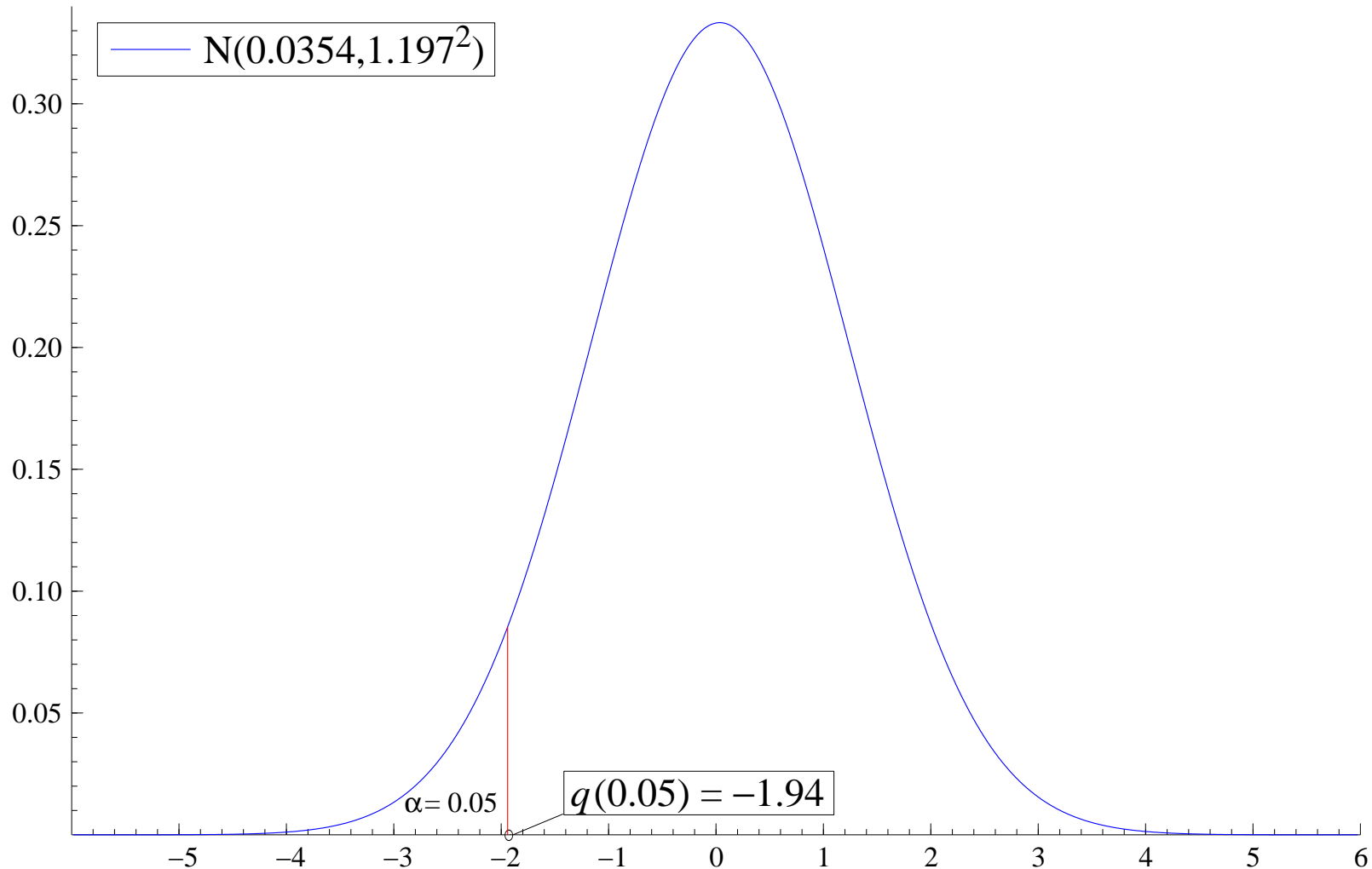
- For example, if $y_w \sim N(\mu_w, \sigma_w^2)$,

$$VaR_w(\alpha) = -(\mu_w + z_\alpha \sigma_w) = -\left(w' \mu + z_\alpha \sqrt{w' \Sigma w}\right),$$

with z_α the α % left-quantile of the $N(0,1)$ distribution.

- Covariances appear implicitly or explicitly in the VaR of a diversified portfolio.

VaR example (normal distribution)



$$VaR(0.05) = -(\underbrace{0.0354}_{\mu} - \underbrace{1.645}_{z_{0.05}} \times \underbrace{1.197}_{\sigma}) = 1.94,$$

Diversification

- It is beneficial to account for the covariances (or correlations) in computing the VaR. Indeed, VaR_w is smaller than the sum of the individual VaR measures, also called the undiversified VaR.
- For example, with two assets invested in positive proportions w_1 and w_2 of one unit, $VaR_i = -w_i(\mu_i + z_\alpha\sigma_i)$,
 $VaR_1 + VaR_2 = -[w_1\mu_1 + w_2\mu_2 + z_\alpha(w_1\sigma_1 + w_2\sigma_2)]$

$$VaR_w =$$

$$- \left[w_1\mu_1 + w_2\mu_2 + z_\alpha \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2\rho_{12}} \right]$$

Hence $VaR_1 + VaR_2 \geq VaR_w$ since $z_\alpha < 0$ and

$$\sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2\rho_{12}} \leq w_1\sigma_1 + w_2\sigma_2 \text{ when } \rho_{12} \leq 1.$$

Example

For $\sigma_1 = \sigma_2$, and $w_1 = w_2 = 0.5$,

$$\frac{VaR_w}{VaR_1 + VaR_2} = \sqrt{\frac{1 + \rho_{12}}{2}}$$

ρ_{12}	$\frac{VaR_w}{VaR_1 + VaR_2}$
1	1
0.5	0.87
0	0.71
-0.5	0.50
-1	0

Univariate modeling

- A univariate volatility model for $y_{w,t}$ can be estimated for a given weight vector, from which $\mu_{w,t}$ and $\sigma_{w,t}$ can be used in the VaR formula: $VaR_{w,t}(\alpha) = -(\mu_{w,t} + z_{\alpha}\sigma_{w,t})$.
- This univariate approach is directly dependent on the portfolio allocation (w), and it requires us to redo the volatility modeling every time the weight vector is changed if we want to study the impact of changing w on $VaR_{w,t}$.
- This approach is not convenient for risk management, i.e. to perform a sensitivity analysis and assess the impact of diversification.

Multivariate modeling

- If a multivariate model is estimated for y_t , with conditional means μ_t and conditional variance matrix Σ_t , the multivariate distribution of the returns can be directly used to compute the VaR of any portfolio, e.g.
$$VaR_{w,t}(\alpha) = - (w' \mu_t + z_\alpha \sqrt{w' \Sigma_t w}).$$
- Thus, there is no need to re-estimate the model for a different weight vector.
- The multivariate approach is limited in practice by the fact that multivariate models are difficult to estimate for a large number of assets. However, progress is under way.
- L. Bauwens and S. Laurent (2005), "A new class of multivariate skew densities, with application to GARCH models", JBES 23/3, 346-354.

MULTIVARIATE VOLATILITY: MODELS and PREDICTION

PART 2

- RiskMetrics
- Multivariate GARCH Models

Change of notation: H_t for Σ_t ; h_{iit} for σ_{it}^2 ; h_{ijt} for σ_{ijt} .

RiskMetrics (exponential smoothing)

- JP Morgan's RiskMetrics system for market risk management uses a simple model of volatility:

$$\begin{aligned} H_t &= (1 - \gamma) \sum_{l=1}^{\infty} \gamma^{l-1} y_{t-l} y'_{t-l} \\ &= \gamma H_{t-1} + (1 - \gamma) y_{t-1} y'_{t-1}, \end{aligned}$$

where $\gamma = 0.94$ (for daily returns) is proposed as a rule of thumb value to avoid estimation.

- For prediction h steps ahead, H_t is the natural predictor:

$$\text{Var}(y_{t+h} | I_t) = H_t,$$

so that there is no reversion to a mean level as h increases.

RiskMetrics: extensions

- A different persistence parameter γ for each element of H_t could be considered.

Let $\Gamma = (\gamma_{ij})$, with $0 < \gamma_{ij} < 1$ for all i and j , and U_N be a matrix of ones ($N \times N$):

$$H_t = \Gamma \odot H_{t-1} + (U_N - \Gamma) \odot y_{t-1}y'_{t-1},$$

Both Γ and $U_N - \Gamma$ must be PD to ensure that H_t is PD.

- Correlations are more persistent than variances if e.g. $\gamma_{ij} = \gamma_c$ for all i and j and $\gamma_{ii} = \gamma_v$ for all i , with $\gamma_c > \gamma_v$.
- Assuming that $y_t | y_{t-1}, y_{t-2}, \dots \sim N(0, H_t)$ and the sequence is conditionally independent, γ (or Γ) can be estimated by ML.

PART 2

- RiskMetrics
- Multivariate GARCH Models

Reference: survey by Bauwens, Laurent and Rombouts
(*Journal of Applied Econometrics* 21/1, 2006).

MGARCH models: definition

A dynamic model with time-varying means, variances and covariances for the N components of $y_t = (y_{1t}, \dots, y_{Nt})'$:

$$y_t = \mu_t + \epsilon_t$$

$$\epsilon_t = H_t^{1/2} z_t, \quad H_t^{1/2} \text{ a } N \times N \text{ matrix}$$

$$z_t \text{ IID, } \mathbf{E}(z_t) = 0, \quad \text{Var}(z_t) = I_N$$

$$\mu_t = \mathbf{E}(y_t | I_{t-1}) = \mathbf{E}_{t-1}(y_t)$$

$$H_t = H_t^{1/2} (H_t^{1/2})' = \text{Var}(y_t | I_{t-1}) = \text{Var}_{t-1}(y_t)$$

where I_{t-1} is the information available at time $t - 1$, at least containing $\{y_{t-1}, y_{t-2} \dots\}$.

An assumption about the joint distribution of z_t completes the definition.

Remarks

- $H_t^{1/2}$ is any $N \times N$ matrix such that H_t is the conditional variance matrix of y_t (e.g. $H_t^{1/2}$ may be obtained by a Cholesky factorization of H_t).
- μ_t and H_t depend on unknown parameters θ but are otherwise known (parametric model), hence sometimes we write explicitly $\mu_t(\theta)$ and $H_t(\theta)$. For example, μ_t may be a VARMA model.
- The " z_t is IID" assumption may be relaxed to " z_t is a martingale difference sequence (MDS) with respect to I_{t-1} ", e.g. to show that $\text{Var}_{t-1}(y_t) = H_t$. However, for ML estimation, the IID assumption is relevant.

Prediction of/ or using volatility

- It follows directly from the model definition that H_{t+1} is the best, in the mean-square error sense, one-step ahead variance matrix predictor, since $H_{t+1} = \text{Var}(y_{t+1}|I_t)$.
- The h -step ahead predictor can also be obtained by computing $\text{Var}(y_{t+h}|I_t)$ as a function of the information set and the model parameters.
- Based on a specific assumption for the distribution of z_t , which implies a distribution of $y_{t+1}|I_t$, point, interval and density forecasts of y_{t+1} can also be deduced from the model. This can be extended to h -step ahead forecasts. Evidently, such forecasts will usually depend on the volatility matrix. Their computation may require numerical simulations.

Three challenges

- State conditions on the parameters θ such that that $H_t > 0 \forall t$, given that $H_0 > 0$.
- Avoid too many parameters (to keep estimation feasible), but maintain enough flexibility in the dynamics of H_t .
- Find $\Sigma = \text{Var}(y_t) = \text{E}(H_t)$ and the conditions for the existence of Σ and for weak stationarity of y_t .

For ease of exposition, we make $H_t = (h_{ijt})$ a function of one lag of ϵ_t and one lag of itself, i.e. we review MGARCH(1,1) models.

Typology of MGARCH models

- Models that specify conditional variances and covariances as linear functions of past: multivariate generalizations of univariate GARCH(1,1) model
 $h_t = \omega + \alpha \epsilon_{t-1}^2 + \gamma h_{t-1}$: VEC, BEKK, factor-GARCH models.
- Models that specify conditional variances and correlations: CCC, DCC.

VEC(1,1) (Bollerslev, Engle, and Wooldridge, 1988)

In the bivariate case,

$$h_{11t} = c_1 + a_{11}\epsilon_{1,t-1}^2 + a_{12}\epsilon_{1,t-1}\epsilon_{2,t-1} + a_{13}\epsilon_{2,t-1}^2 \\ + g_{11}h_{11,t-1} + g_{12}h_{21,t-1} + g_{13}h_{22,t-1}$$

$$h_{21t} = c_2 + a_{21}\epsilon_{1,t-1}^2 + a_{22}\epsilon_{1,t-1}\epsilon_{2,t-1} + a_{23}\epsilon_{2,t-1}^2 \\ + g_{21}h_{11,t-1} + g_{22}h_{21,t-1} + g_{23}h_{22,t-1}$$

$$h_{22t} = c_3 + a_{31}\epsilon_{1,t-1}^2 + a_{32}\epsilon_{1,t-1}\epsilon_{2,t-1} + a_{33}\epsilon_{2,t-1}^2 \\ + g_{31}h_{11,t-1} + g_{32}h_{21,t-1} + g_{33}h_{22,t-1}.$$

Matrix notation for bivariate VEC(1,1)

$$\begin{bmatrix} h_{11t} \\ h_{21t} \\ h_{22t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{21,t-1} \\ h_{22,t-1} \end{bmatrix} .$$

\Rightarrow the numbers of parameters is of order N^4
(for $N = 2, 3, 4$ it is equal to 21, 78, 210 respectively).

General VEC(1,1)

In this model, h_{ijt} is a linear function of the lagged squared errors, cross products of errors, and lagged values of all the elements of H_t . The $VEC(1,1)$ is defined as:

$$\text{vech } H_t = c + A \text{vech } (\epsilon_{t-1} \epsilon'_{t-1}) + G \text{vech } H_{t-1},$$

where c is a $N^* \times 1$ vector of parameters [with $N^* = N(N + 1)/2$],

and A and G are $N^* \times N^*$ matrices of parameters.

Vech and vec operators

- **vech** is the operator that stacks the lower triangle of a $N \times N$ matrix as an $N(N + 1)/2 \times 1$ vector:

$$\text{vech } H_t = (h_{11t}, h_{21t}, h_{22t}, h_{31t}, \dots, h_{NNt})'.$$

- **vec** is the operator that stacks a matrix as a column vector:

$$\text{vec } H_t = (h_{11t}, h_{21t}, \dots, h_{N1t}, h_{12t}, h_{22t}, \dots, h_{NNt})'.$$

Stationarity conditions: VEC

- The VEC(1,1) model

$$\text{vech } H_t = c + A \text{vech } (\epsilon_{t-1} \epsilon'_{t-1}) + G \text{vech } H_{t-1}$$

can be written as a VARMA(1,1) model for $\text{vech } (\epsilon_t \epsilon'_t)$:

$$\text{vech } (\epsilon_t \epsilon'_t) = c + (A + G) \text{vech } (\epsilon_{t-1} \epsilon'_{t-1}) + w_t - G w_{t-1},$$

where $w_t = \text{vech } (\epsilon_{t-1} \epsilon'_{t-1} - H_t)$ is a martingale difference.

- \Rightarrow The VEC(1,1) model is weakly stationary if the eigenvalues of $A + G$ are less than 1 in modulus. Then,

$$\text{vech } \Sigma = E(\text{vech } H_t) = (I_{N^*} - A - G)^{-1} c,$$

and the h -step ahead predictor of the volatility matrix tends to the mean volatility level Σ as h increases.

Volatility prediction with VEC

- Using I_t as information set, the 1-step ahead predictor is

$$\text{vech } H_{t+1|t} = c + A \text{vech } (\epsilon_t \epsilon_t') + G \text{vech } H_{t|t-1}$$

- The 2-step ahead predictor is

$$\begin{aligned} \text{vech } H_{t+2|t} &= c + A \underbrace{\text{E}_t[\text{vech } (\epsilon_{t+1} \epsilon_{t+1}')]_{\text{vech } H_{t+1|t}}}_{\text{vech } H_{t+1|t}} + G \text{vech } H_{t+1|t} \\ &= c + (A + G) \text{vech } H_{t+1|t} \end{aligned}$$

Volatility prediction with VEC

- The 3-step ahead predictor is

$$\begin{aligned}\text{vech } H_{t+3|t} &= c + A \underbrace{\mathbf{E}_t[\text{vech } (\epsilon_{t+2}\epsilon'_{t+2})]}_{c+(A+G)\text{vech } H_{t+1|t}} + G \underbrace{\mathbf{E}_t[\text{vech } H_{t+2|t}]}_{c+(A+G)\text{vech } H_{t+1|t}} \\ &= c + (A + G)c + (A + G)^2 \text{vech } H_{t+1|t}\end{aligned}$$

- The h -step ahead predictor

$$\begin{aligned}\text{vech } H_{t+h|t} &= \sum_{j=0}^{h-2} (A + G)^j c + (A + G)^{h-1} \text{vech } H_{t+1|t} \\ &\rightarrow \sum_{j=0}^{\infty} (A + G)^j c = (I_{N^*} - A - G)^{-1} c = \text{vech } \Sigma\end{aligned}$$

Positivity conditions for VEC (Gouriéroux, 1997)

One can also write

$$h_{11t} = c_1 + a_{11}\epsilon_{1,t-1}^2 + a_{12}\epsilon_{1,t-1}\epsilon_{2,t-1} + a_{13}\epsilon_{2,t-1}^2 \\ + g_{11}h_{11,t-1} + g_{12}h_{21,t-1} + g_{13}h_{22,t-1}$$

like

$$h_{11t} = c_1 + \begin{pmatrix} \epsilon_{1,t-1} & \epsilon_{2,t-1} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12}/2 \\ a_{12}/2 & a_{13} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix} \\ + \mathbf{E}_{t-2} \left[\begin{pmatrix} \epsilon_{1,t-1} & \epsilon_{2,t-1} \end{pmatrix} \begin{pmatrix} g_{11} & g_{12}/2 \\ g_{12}/2 & g_{13} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix} \right]$$

as $h_{ii,t-1} = \mathbf{E}_{t-2}[\epsilon_{i,t-1}^2]$ ($i = 1, 2$), and $h_{21,t-1} = \mathbf{E}_{t-2}[\epsilon_{1,t-1}\epsilon_{2,t-1}]$.

Positivity conditions for VEC

$$h_{12t} = c_2 + \begin{pmatrix} \epsilon_{1,t-1} & \epsilon_{2,t-1} \end{pmatrix} \begin{pmatrix} a_{21} & a_{22}/2 \\ a_{22}/2 & a_{23} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix} \\ + \mathbf{E}_{t-2} \left[\begin{pmatrix} \epsilon_{1,t-1} & \epsilon_{2,t-1} \end{pmatrix} \begin{pmatrix} g_{21} & g_{22}/2 \\ g_{22}/2 & g_{23} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix} \right]$$

$$h_{22t} = c_3 + \begin{pmatrix} \epsilon_{1,t-1} & \epsilon_{2,t-1} \end{pmatrix} \begin{pmatrix} a_{31} & a_{32}/2 \\ a_{32}/2 & a_{33} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix} \\ + \mathbf{E}_{t-2} \left[\begin{pmatrix} \epsilon_{1,t-1} & \epsilon_{2,t-1} \end{pmatrix} \begin{pmatrix} g_{31} & g_{32}/2 \\ g_{32}/2 & g_{33} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{pmatrix} \right]$$

Positivity conditions for VEC

Putting the different parts together:

$$H_t = \begin{pmatrix} c_1 & c_2 \\ c_2 & c_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t-1} & \epsilon_{2,t-1} & 0 & 0 \\ 0 & 0 & \epsilon_{1,t-1} & \epsilon_{2,t-1} \end{pmatrix} \\ \times \begin{pmatrix} a_{11} & a_{12}/2 & a_{21} & a_{22}/2 \\ a_{12}/2 & a_{13} & a_{22}/2 & a_{23} \\ a_{21} & a_{22}/2 & a_{31} & a_{32}/2 \\ a_{22}/2 & a_{23} & a_{32}/2 & a_{33} \end{pmatrix} \begin{pmatrix} \epsilon_{1,t-1} & 0 \\ \epsilon_{2,t-1} & 0 \\ 0 & \epsilon_{1,t-1} \\ 0 & \epsilon_{2,t-1} \end{pmatrix} \\ + \mathbf{E}_{t-2} \left[(I_2 \otimes \epsilon'_{t-1}) \tilde{G} (I_2 \otimes \epsilon_{t-1}) \right]$$

We denote by \tilde{A} the above 4×4 matrix, and by \tilde{G} the matrix built in the same way from G .

Positivity conditions for VEC

- A general matrix (rather than vech) expression of H_t in the VEC(1,1) case is:

$$H_t = C + (I_N \otimes \epsilon'_{t-1}) \tilde{A} (I_N \otimes \epsilon_{t-1}) \\ + \mathbf{E}_{t-2} \left[(I_N \otimes \epsilon'_{t-1}) \tilde{G} (I_N \otimes \epsilon_{t-1}) \right]$$

- Hence, sufficient conditions for positivity of H_t are that $C \geq 0$, $\tilde{A} \geq 0$, $\tilde{G} \geq 0$, with at least one strict inequality.
- These restrictions may be imposed in estimation by using Cholesky factorizations. Usually they are not imposed, but they can be checked after unrestricted estimation.

Diagonal and Scalar VEC

- To reduce the number of parameters, BEW (1988) suggest the diagonal VEC (DVEC) model in which the A and G matrices are diagonal.
- Each variance h_{iit} depends only on its own past squared error $\epsilon_{i,t-1}^2$ and its own lag $h_{ii,t-1}$.
- Each covariance h_{ijt} depends only on its own past cross-products of errors $\epsilon_{i,t-1}\epsilon_{j,t-1}$, and its own lag.
- Quite restrictive: no "spillover effect".
- Big reduction: 9 parameters instead of 21 when $N=2$; 18 instead of 78 when $N=3$...
- Scalar VEC: $A = aU$ and $G = gU$ where a and g are scalars and U is a matrix of ones.

Diagonal bivariate VEC(1,1)

$$\begin{bmatrix} h_{11t} \\ h_{21t} \\ h_{22t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} + \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{2,t-1}^2 \end{bmatrix} \\ + \begin{bmatrix} g_{11} & 0 & 0 \\ 0 & g_{22} & 0 \\ 0 & 0 & g_{33} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{21,t-1} \\ h_{22,t-1} \end{bmatrix}$$

or

$$\begin{aligned} h_{11t} &= c_1 + a_{11}\epsilon_{1,t-1}^2 + g_{11}h_{11,t-1} \\ h_{21t} &= c_2 + a_{22}\epsilon_{1,t-1}\epsilon_{2,t-1} + g_{22}h_{21,t-1} \\ h_{22t} &= c_3 + a_{33}\epsilon_{2,t-1}^2 + g_{33}h_{22,t-1}. \end{aligned}$$

Diagonal VEC(1,1)

Equivalently,

$$\begin{pmatrix} h_{11t} & h_{21t} \\ h_{21t} & h_{22t} \end{pmatrix} = \begin{pmatrix} c_1 & c_2 \\ c_2 & c_3 \end{pmatrix} + \begin{pmatrix} a_{11} & a_{22} \\ a_{22} & a_{33} \end{pmatrix} \odot \begin{pmatrix} \epsilon_{1,t-1}^2 & \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{2,t-1}\epsilon_{1,t-1} & \epsilon_{2,t-1}^2 \end{pmatrix} \\ + \begin{pmatrix} g_{11} & g_{22} \\ g_{22} & g_{33} \end{pmatrix} \odot \begin{pmatrix} h_{11,t-1} & h_{21,t-1} \\ h_{21,t-1} & h_{22,t-1} \end{pmatrix}$$

In general,

$$H_t = C + A_d \odot (\epsilon_{t-1}\epsilon'_{t-1}) + G_d \odot H_{t-1}$$

If C , A_d , and G_d are PD, H_t is PD.

NB: The simple RiskMetrics formula imposes

$$C = 0, A_d = (1 - \gamma)U_N \text{ and } G_d = \gamma U_N.$$

Bivariate BEKK(1,1,1) (Engle and Kroner, 1995)

$$\begin{bmatrix} h_{11t} & h_{21t} \\ h_{21t} & h_{22t} \end{bmatrix} = \begin{bmatrix} c_{11}^* & 0 \\ c_{21}^* & c_{22}^* \end{bmatrix} \begin{bmatrix} c_{11}^* & c_{21}^* \\ 0 & c_{22}^* \end{bmatrix} \\ + \begin{bmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \end{bmatrix}' \begin{bmatrix} \epsilon_{1,t-1}^2 & \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{2,t-1}\epsilon_{1,t-1} & \epsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} a_{11}^* & a_{12}^* \\ a_{21}^* & a_{22}^* \end{bmatrix} \\ + \begin{bmatrix} g_{11}^* & g_{12}^* \\ g_{21}^* & g_{22}^* \end{bmatrix}' \begin{bmatrix} h_{11,t-1} & h_{21,t-1} \\ h_{21,t-1} & h_{22,t-1} \end{bmatrix} \begin{bmatrix} g_{11}^* & g_{12}^* \\ g_{21}^* & g_{22}^* \end{bmatrix}.$$

→ 11 parameters, against 21 in the VEC model.

Bivariate BEKK(1,1,1)

Same linear structure as in VEC model...

$$\begin{aligned}h_{11t} &= \omega_{11} \\ &+ a_{11}^{*2} \epsilon_{1,t-1}^2 + 2a_{11}^* a_{21}^* \epsilon_{1,t-1} \epsilon_{2,t-1} + a_{21}^{*2} \epsilon_{2,t-1}^2 \\ &+ g_{11}^{*2} h_{11,t-1} + 2g_{11}^* g_{21}^* h_{21,t-1} + g_{21}^{*2} h_{22,t-1}\end{aligned}$$

$$\begin{aligned}h_{21t} &= \omega_{21} \\ &+ a_{11}^* a_{12}^* \epsilon_{1,t-1}^2 + (a_{11}^* a_{22}^* + a_{12}^* a_{21}^*) \epsilon_{1,t-1} \epsilon_{2,t-1} + a_{22}^* a_{21}^* \epsilon_{2,t-1}^2 \\ &+ g_{11}^* g_{12}^* h_{11,t-1} + (g_{11}^* g_{22}^* + g_{21}^* g_{21}^*) h_{21,t-1} + g_{22}^* g_{21}^* h_{22,t-1}\end{aligned}$$

... but constraints on parameters (compare with bivariate VEC).

General BEKK(1,1,K)

- The BEKK(1, 1, K) model is defined as:

$$H_t = C^{*'} C^* + \sum_{k=1}^K A_k^{*'} \epsilon_{t-1} \epsilon_{t-1}' A_k^* + \sum_{k=1}^K G_k^{*'} H_{t-1} G_k^*.$$

where C^* , A_k^* and G_k^* are $N \times N$ matrices of parameters but C^* is upper triangular. One can write as well

$$C^{*'} C^* = \Omega > 0.$$

- Positivity of H_t is automatically guaranteed (if $H_0 \geq 0$).

Remarks

- Interpretation of the basic parameters not obvious, see previous equations.
- By increasing K , one makes the specification more flexible (e.g. for $N = K = 2$, there are 19 parameters, against 21 in the bivariate VEC).
- Diagonal BEKK model: take A_k^* and G_k^* as diagonal matrices. It is a restricted DVEC model (check the covariance equation to see the restrictions).
- One can define a scalar BEKK model: $A_k^* = a_k U$,
 $G_k^* = g_k U$.

Stationarity conditions: BEKK

- The BEKK(1,1,1) model

$H_t = \Omega + A^* \epsilon_{t-1} \epsilon'_{t-1} A^* + G^* H_{t-1} G^*$ can be written as a VEC model (subject to restrictions) using the formula $\text{vec}(ABC) = (C' \otimes A) \text{vec} B$, so that

$$\text{vec} H_t = \text{vec} \Omega + (A^* \otimes A^*)' \text{vec} (\epsilon_{t-1} \epsilon'_{t-1}) + (G^* \otimes G^*)' \text{vec} H_{t-1}.$$

- Hence, the BEKK model is weakly stationary if the eigenvalues of $(A^* \otimes A^*) + (G^* \otimes G^*)$ are smaller than 1 in modulus, and then

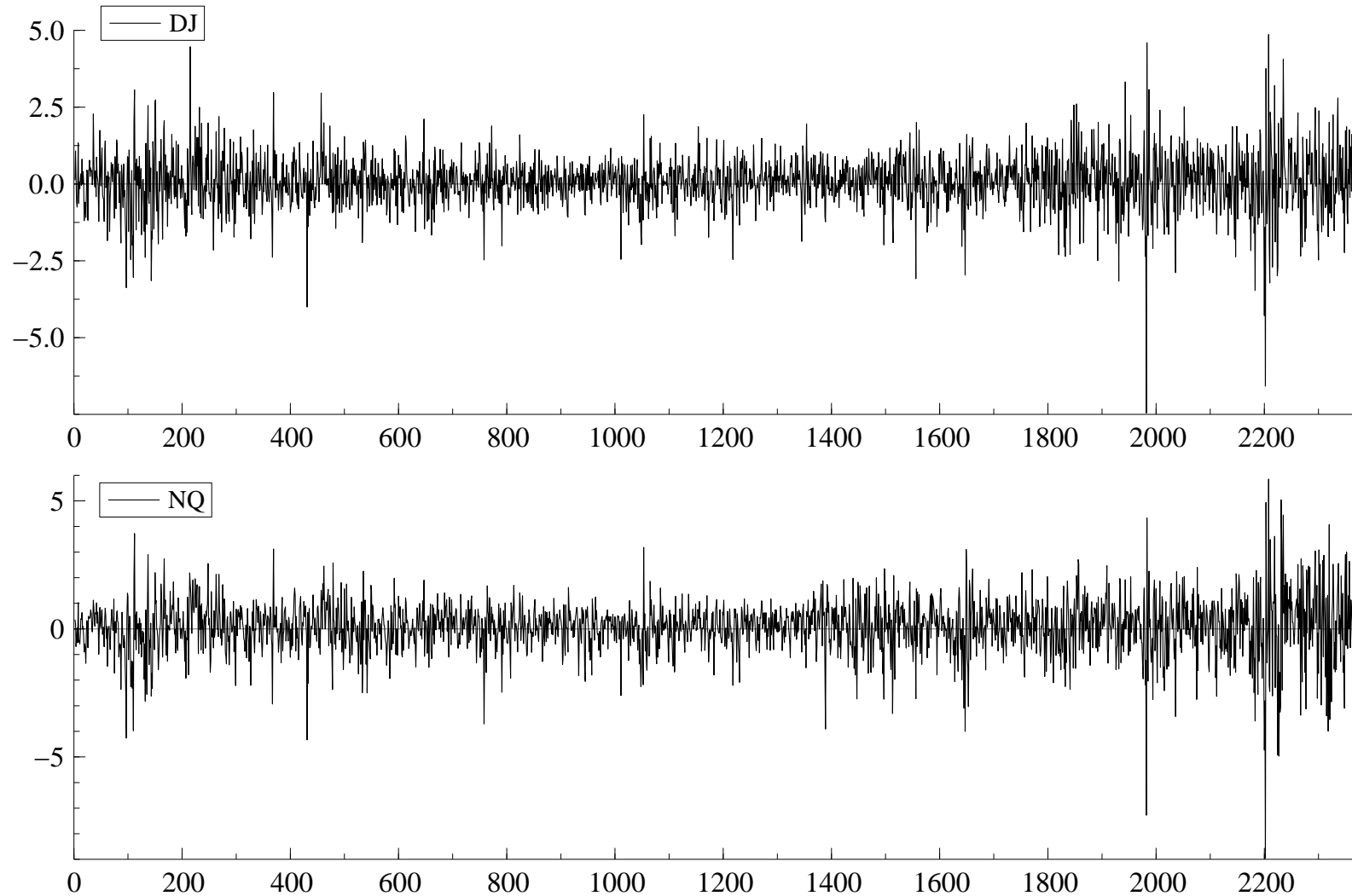
$$\text{vec} \Sigma = E(\text{vec} H_t) = \left(I_{N^2} - (A^* \otimes A^*)' - (G^* \otimes G^*)' \right)^{-1} \text{vec} \Omega.$$

Stylized property

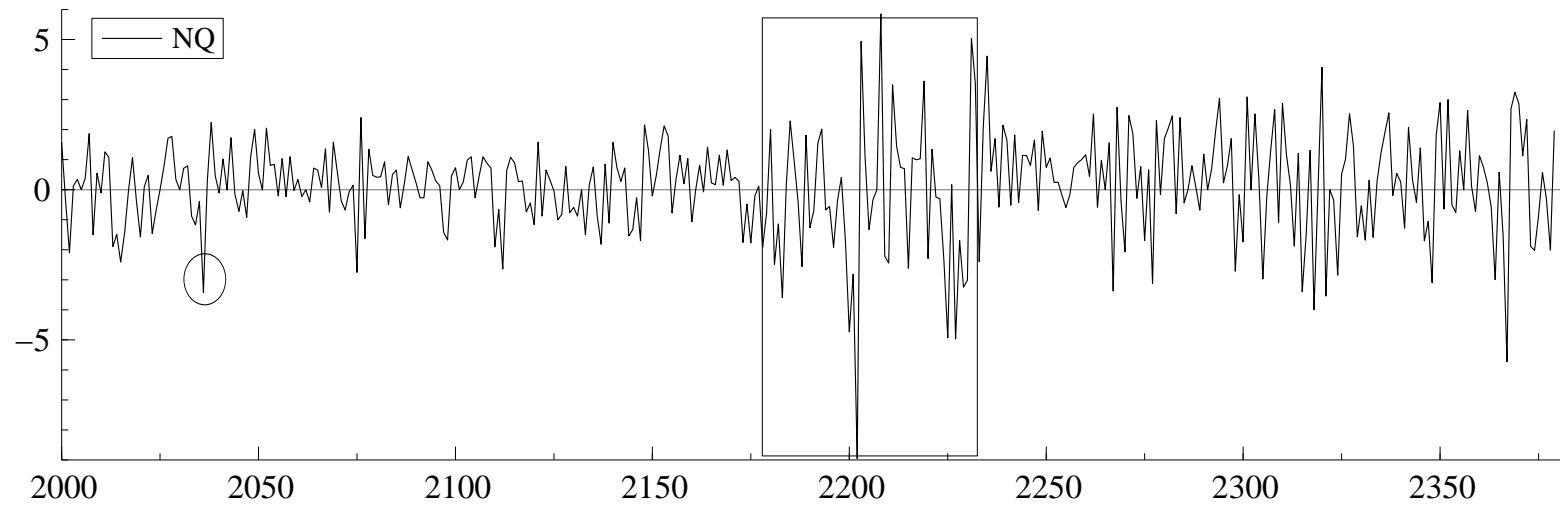
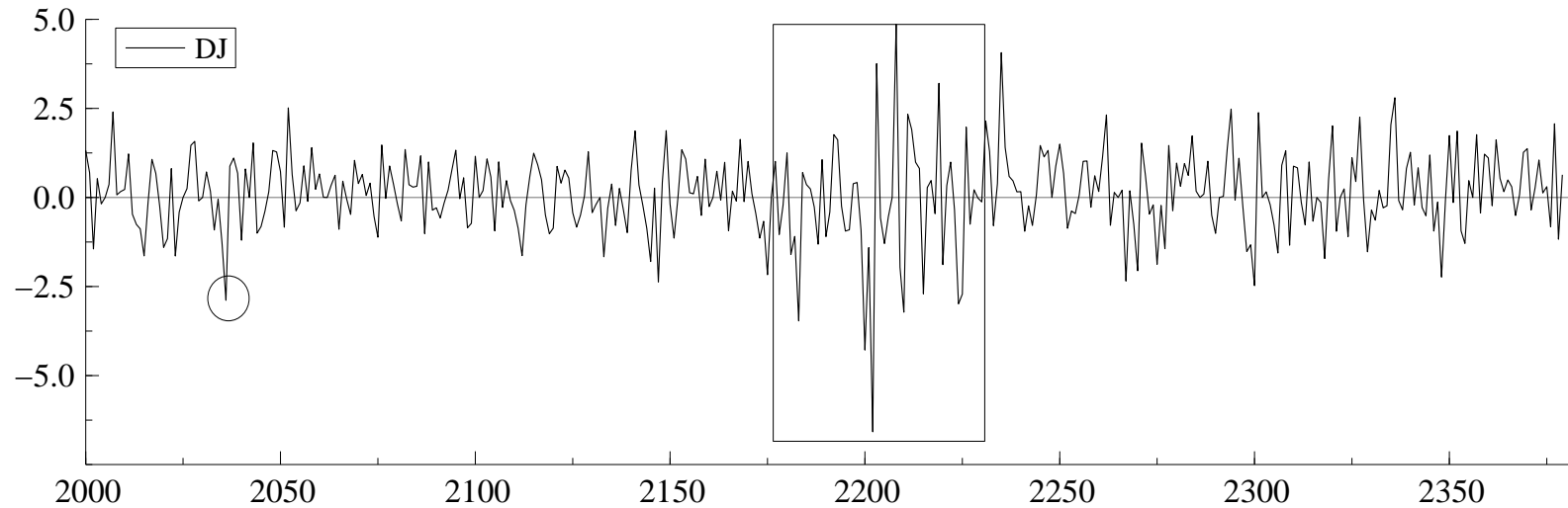
- Black (1976) observed that
“..there is a lot of commonality in volatility changes across stocks: a 1% market volatility change typically implies a 1 % volatility change for each stock. Well, perhaps the high volatility stocks are somewhat more sensitive to market volatility changes than the low volatility stocks. In general it seems fair to say that when stock volatilities change, they all tend to change in the same direction.”
- This type of commonality has also been observed for exchange rate volatility and for the volatility of bonds of different maturities.

Dow Jones and Nasdaq daily returns

Daily returns, DOW Jones (DJ) and NASDAQ (NQ) indices, 03/26/1990 - 03/23/2000



A sub-period



One-factor MGARCH model

- Let

$$y_t - \mu_t = \epsilon_t = \lambda f_t + e_t,$$

and assume that

i) f_t , the common shock (a scalar random variable) and e_t , the idiosyncratic shocks (a $N \times 1$ vector), are uncorrelated,

ii) $e_t \sim IID(0, \Omega^*)$,

iii) $E_{t-1}(f_t) = 0$, and $\text{Var}_{t-1}(f_t) = h_t = \omega + \alpha f_{t-1}^2 + \beta h_{t-1}$,
we get

$$H_t = \Omega^* + \lambda \lambda' h_t.$$

- The elements of H_t follow the same dynamics, determined by h_t . How to make this more flexible?
- Weak stationarity requires that $\alpha + \beta < 1 \forall k$.

One-factor MGARCH model

- The factor can be observable, such as $f_t = w' \epsilon_t$ where w is a vector of parameters, or a market index return.
- The factor may be considered as unobservable, which makes the model a stochastic volatility model (rendering estimation much more complex).
- The conditional correlations implied by this model

$$\rho_{ijt} = \frac{\lambda_i \lambda_j h_t + \omega_{ij}^*}{(\lambda_i^2 h_t + \omega_{ii}^*)^{1/2} (\lambda_j^2 h_t + \omega_{jj}^*)^{1/2}}$$

tend to 1 as h_t increases, if $\lambda_i \lambda_j > 0$. This corresponds to an empirical regularity.

Bivariate one-factor-GARCH(1,1)

$$h_{11t} = \omega_{11}^* + \lambda_1^2 h_t$$

$$h_{21t} = \omega_{21}^* + \lambda_1 \lambda_2 h_t$$

$$h_{22t} = \omega_{22}^* + \lambda_2^2 h_t$$

where

$$\lambda_2 = (1 - w_1 \lambda_1) / (1 - w_1)$$

$$h_t = \omega + \alpha^2 f_{t-1}^2 + \beta^2 h_{t-1}$$

$$f_t = w_1 \epsilon_{1t} + (1 - w_1) \epsilon_{2t}.$$

K-factor-GARCH(1,1) (Engle, Ng, and Rotschild, 1990b)

The *K*-Factor-GARCH(1,1) model can be viewed as a particular BEKK(1,1,*K*) model:

$$H_t = \Omega + \sum_{k=1}^K \alpha_k \lambda_k w_k' \epsilon_{t-1} \epsilon_{t-1}' w_k \lambda_k' + \sum_{k=1}^K \beta_k \lambda_k w_k' H_{t-1} w_k \lambda_k'$$

i.e. A_k^* and G_k^* are replaced by rank one matrices that are proportional to one another. The $N \times 1$ vectors λ_k and w_k are subject to the identifying restrictions:

$$w_k' \lambda_i = \begin{cases} 0 & \text{for } k \neq i \\ 1 & \text{for } k = i, \end{cases} \quad \sum_{n=1}^N w_{kn} = 1.$$

Other Factor-GARCH models

- The orthogonal GARCH model (Kariya, 1988, Alexander and Chibumba, 1997) and the generalized orthogonal GARCH models of van der Weide (2002) and Vrontos et al. (2003) are also Factor-GARCH models.
- Lanne and Saikkonen (2005) propose "A Multivariate Generalized Orthogonal Factor GARCH Model", as an interesting alternative to the previous models.
- See also Doz and Renault (2004), "Conditionally heteroskedastic factor models: identification and instrumental variables estimation".

Number of parameters

	H_t	# parameters for $N = 2, 3, 4, 5$
<i>VEC(1,1)</i>	$h_t = c + A\eta_{t-1} + Gh_{t-1}$ $h_t = \text{vech}(H_t) \quad \eta_t = \text{vech}(\epsilon_t \epsilon_t')$	$\frac{N(N+1)(N(N+1)+1)}{2}$ 21, 78, 210, 465
<i>BEKK(1,1,1)</i>	$H_t = C^*{}' C^* + A^*{}' \epsilon_{t-1} \epsilon_{t-1}' A^* + G^*{}' H_{t-1} G^*$	$\frac{N(5N+1)}{2}$ 11, 24, 42, 65
<i>F-GARCH(1,1,1)</i>	$H_t = \Omega + \lambda \lambda' (\alpha^2 w' \epsilon_{t-1} \epsilon_{t-1}' w + \beta^2 w' H_{t-1} w)$ $w' \lambda = 1, \quad \sum_{n=1}^N w_n = 1$	$\frac{N(N+5)}{2}$ 7, 12, 18, 25

Conditional correlations

- H_t can always be written as

$$H_t = D_t R_t D_t,$$

$$D_t = \text{diag} (h_{11t}^{1/2} \dots h_{NNt}^{1/2}),$$

$$R_t = (\rho_{ijt}), \quad \text{with } \rho_{iit} = 1.$$

R_t is the $N \times N$ matrix of conditional correlations, and h_{iit} is a conditional variance. Hence,

$$h_{ijt} = \rho_{ijt} \sqrt{h_{iit} h_{jjt}} \quad \forall i \neq j.$$

- Each h_{iit} can be defined individually as a univariate GARCH model.
- Positivity of H_t follows from positivity of R_t and of all h_{iit} .

Bivariate case

$$\underbrace{\begin{pmatrix} h_{11t} & h_{21t} \\ h_{21t} & h_{22t} \end{pmatrix}}_{H_t} = \underbrace{\begin{pmatrix} \sqrt{h_{11t}} & 0 \\ 0 & \sqrt{h_{22t}} \end{pmatrix}}_{D_t} \underbrace{\begin{pmatrix} 1 & \rho_{21t} \\ \rho_{21t} & 1 \end{pmatrix}}_{R_t} \underbrace{\begin{pmatrix} \sqrt{h_{11t}} & 0 \\ 0 & \sqrt{h_{22t}} \end{pmatrix}}_{D_t}$$

since $h_{21t} = \rho_{21t} \sqrt{h_{11t}} \sqrt{h_{22t}}$.

CCC (Bollerslev, 1990)

In this case,

$$R_t = R = (\rho_{ij}), \quad \rho_{ii} = 1,$$

i.e. "constant conditional correlations" (CCC). Hence,

$$h_{ijt} = \rho_{ij} \sqrt{h_{iit} h_{jjt}} \quad \forall i \neq j,$$

and thus the dynamics of the covariance is determined only by the dynamics of the two conditional variances.

NB: there are $N(N - 1)/2$ parameters in R .

DCC of Tse & Tsui (2002)

DCC for "dynamic conditional correlations".

$DCC_T(M)$:

$$\begin{aligned} R_t &= (1 - \theta_1 - \theta_2)R + \theta_1 \Psi_{t-1} + \theta_2 R_{t-1} & (1) \\ \psi_{ij,t-1} &= \frac{\sum_{m=1}^M u_{i,t-m} u_{j,t-m}}{\sqrt{(\sum_{m=1}^M u_{i,t-m}^2)(\sum_{h=1}^M u_{j,t-m}^2)}} \\ u_{it} &= \epsilon_{it} / \sqrt{h_{iit}} \end{aligned}$$

with $\theta_1, \theta_2 > 0$ and $\theta_1 + \theta_2 < 1$. R is like in CCC.

Notice that $\psi_{ii,t-1} = 1 \forall i$ by construction.

Remarks

- Ψ_{t-1} is the sample correlation matrix of ϵ_τ for $\tau = t - M, t - M + 1, \dots, t - 1$. A necessary condition to ensure positivity of Ψ_{t-1} is that $M \geq N$.
- R_t is a weighted average of correlation matrices (R, Ψ_{t-1}, R_{t-1}). Hence, $R_t > 0$ if any of the three components is > 0 .
- If $\theta_1 = \theta_2 = 0$, the CCC model is obtained. Hence one can test for CCC against $DCC_T(M)$.

DCC of Engle (2002)

$DCC_E(1, 1)$:

$$R_t = (\text{diag } Q_t)^{-1/2} Q_t (\text{diag } Q_t)^{-1/2}. \quad (2)$$

Q_t is a $N \times N$ matrix, symmetric and > 0 , given by

$$Q_t = (1 - \theta_1 - \theta_2) \bar{Q} + \theta_1 u_{t-1} u'_{t-1} + \theta_2 Q_{t-1}, \quad (3)$$

where $u_t = (u_{1t} \dots u_{Nt})'$, $u_{it} = \epsilon_{it} / \sqrt{h_{iit}}$,

\bar{Q} is a $N \times N$ matrix, symmetric and > 0 , of parameters, and θ_1 and θ_2 positive parameters satisfying $\theta_1 + \theta_2 < 1$,
 $\Rightarrow Q_t > 0$ and $R_t > 0$.

Remarks

- Q_t is the covariance matrix of u_t , since q_{iit} is not equal to 1 by construction. Then it is transformed into a correlation matrix by (2).
- If $\theta_1 = \theta_2 = 0$, and $\bar{q}_{ii} = 1$, the CCC model is obtained. Hence one can test for CCC against $DCC_E(M)$.
- In both DCC models, all the correlations obey the same dynamics. This saves a lot of parameters, compared to VEC and BEKK models, but is quite restrictive (especially when N is large).

Comparison

The correlation coefficient in the bivariate case:
for the $DCC_T(M)$,

$$\rho_{12t} = (1 - \theta_1 - \theta_2)\rho_{12} + \theta_2\rho_{12,t-1} + \theta_1 \frac{\sum_{m=1}^M u_{1,t-m}u_{2,t-m}}{\sqrt{(\sum_{m=1}^M u_{1,t-m}^2)(\sum_{h=1}^M u_{2,t-m}^2)}}$$

and for the $DCC_E(1, 1)$, $\rho_{12t} =$

$$\frac{(1 - \theta_1 - \theta_2)\bar{q}_{12} + \theta_1 u_{1,t-1}u_{2,t-1} + \theta_2 q_{12,t-1}}{\sqrt{\left((1 - \theta_1 - \theta_2)\bar{q}_{11} + \theta_1 u_{1,t-1}^2 + \theta_2 q_{11,t-1}\right) \left((1 - \theta_1 - \theta_2)\bar{q}_{22} + \theta_1 u_{2,t-1}^2 + \theta_2 q_{22,t-1}\right)}}$$

Number of parameters

	$H_t = D_t R_t D_t$ $D_t = \text{diag} (h_{11t}^{1/2} \dots h_{NNt}^{1/2})$	# parameters for $N = 2, 3, 4, 5$
<i>CCC</i>	$R_t = R$	$\frac{N(N+5)}{2}$ 7, 12, 18, 25
<i>DCC_T(M)</i>	$R_t = (1 - \theta_1 - \theta_2)R + \theta_1 \Psi_{t-1} + \theta_2 R_{t-1}$ $\psi_{ij,t-1} = \frac{\sum_{m=1}^M u_{i,t-m} u_{j,t-m}}{\sqrt{(\sum_{m=1}^M u_{i,t-m}^2)(\sum_{m=1}^M u_{j,t-m}^2)}}$	$\frac{(N+1)(N+4)}{2}$ 9, 14, 20, 27
<i>DCC_E(1, 1)</i>	$R_t = (\text{diag } Q_t)^{-1/2} Q_t (\text{diag } Q_t)^{-1/2}$ $Q_t = (1 - \theta_1 - \theta_2) \bar{Q} + \theta_1 u_{t-1} u'_{t-1} + \theta_2 Q_{t-1}$	$\frac{(N+1)(N+4)}{2}$ 9, 14, 20, 27

Extensions of DCC

- Recent and ongoing research aims at specifying more flexible dynamic correlations, avoiding the common dynamics restriction of all correlations.
- References:
 - Billio, Caporin, and Gobbo (2003): a block-structure of DCC.
 - Hafner and Franses (2003), "A Generalized Dynamic Conditional Correlation Model for Many Asset Returns".
 - Palandri (2005), "Sequential Conditional Correlations: Inference and Evaluation".
- Copula-MGARCH models combine GARCH for variances and copula for conditional dependence.
 - Patton (2000), Jondeau and Rockinger (2001).

ML Estimation

- ML requires an assumption about the density of z_t , denoted $g(z_t; \eta)$, where η is an additional parameter vector.
- \Rightarrow Maximize with respect to (θ, η) the function

$$L_T(\theta, \eta) = \sum_{t=1}^T \log f(y_t | \theta, \eta, I_{t-1}),$$

with

$$f(y_t | \theta, \eta, I_{t-1}) = |H_t|^{-1/2} g \left(H_t^{-1/2} (y_t - \mu_t); \eta \right),$$

where the dependence with respect to θ occurs through μ_t and H_t .

Gaussian likelihood

- In many cases, $z_t \sim N(0, I_N)$ is assumed (hence η is empty). Then, neglecting a constant,

$$L_T(\theta) = -\frac{1}{2} \sum_{t=1}^T \left(\log |H_t| + (y_t - \mu_t)' H_t^{-1} (y_t - \mu_t) \right). \quad (4)$$

- This Gaussian log-lik. provides the QML estimator that is consistent for θ even if the true density is not $N(0, I_N)$ (if μ_t and H_t are correctly specified).
- However, this QML estimator is less efficient than the ML estimator that would be obtained using the log-lik. based on the true density.

Remarks

- For financial returns, normality is not realistic, like for univariate GARCH models.
- For financial applications (such as computing the VaR), it is important to use the most correct assumption about the density. Hence, normality is not useful in some applications...
- Alternative distributions: multivariate Student (to account for excess conditional kurtosis), multivariate skewed-Student or mixture of two multivariate Gaussian densities (for conditional skewness), generalized hyperbolic distribution, ...
- Danger of this approach: if the assumption is not correct, inconsistency of the estimator results. To what extent?

Student density

- The multivariate Student density, denoted $ST(0, I_N, \nu)$ (ν corresponds to η), is

$$g(z_t|\nu) = \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) [\pi(\nu-2)]^{\frac{N}{2}}} \left[1 + \frac{z_t' z_t}{\nu-2}\right]^{-\frac{N+\nu}{2}}, \quad (5)$$

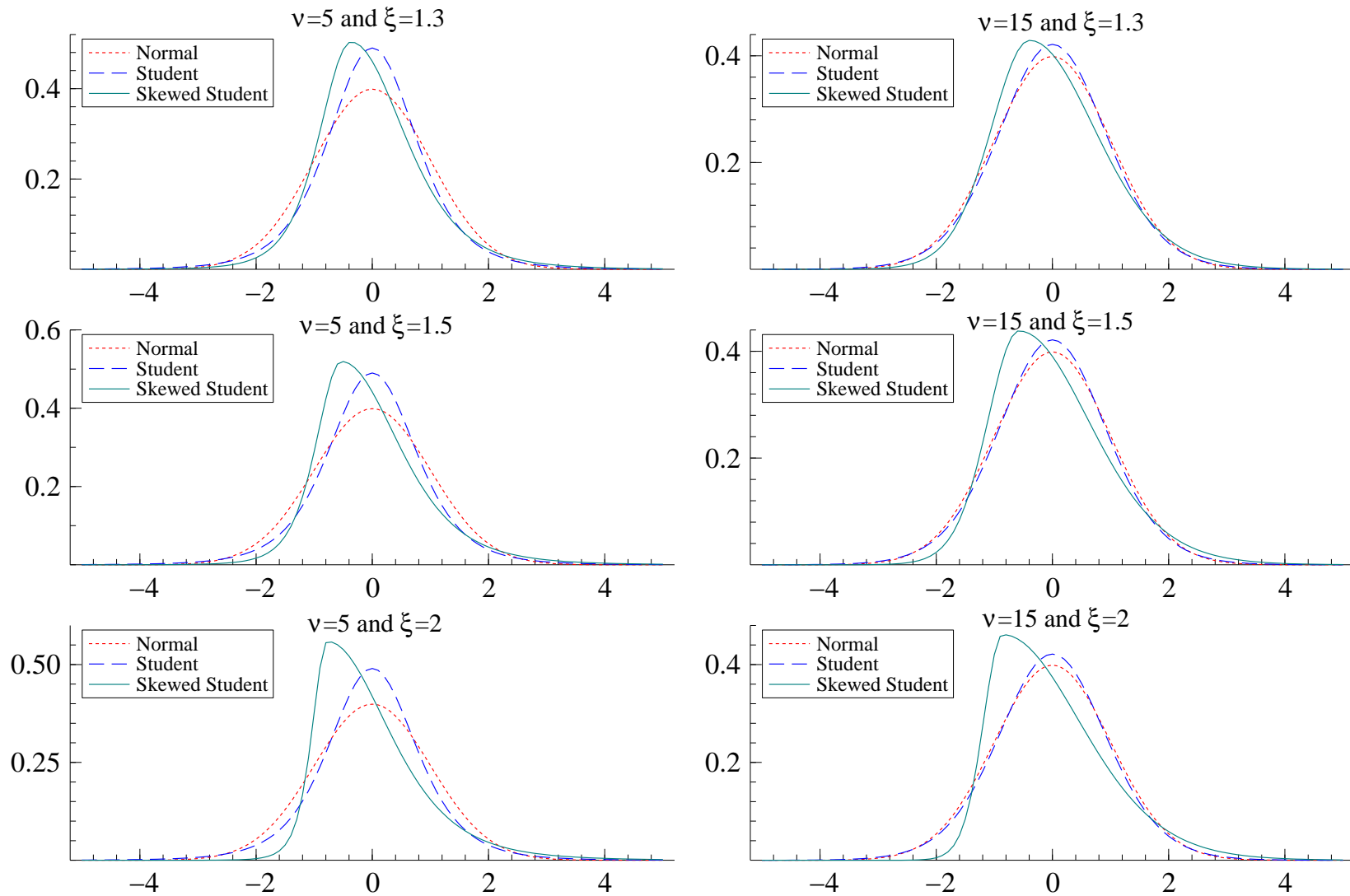
where $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$ is the Gamma function.

- Here we impose $\nu > 2$, and $\text{Var}(z_t) = I_N$ while $E(z_t) = 0$. Although uncorrelated, the elements of z_t are not independent.
- When $\nu \rightarrow \infty$, $ST(0, I_N, \nu) \rightarrow N(0, I_N)$.
- When $\nu \rightarrow 2$, the tails of the density become thicker and thicker.

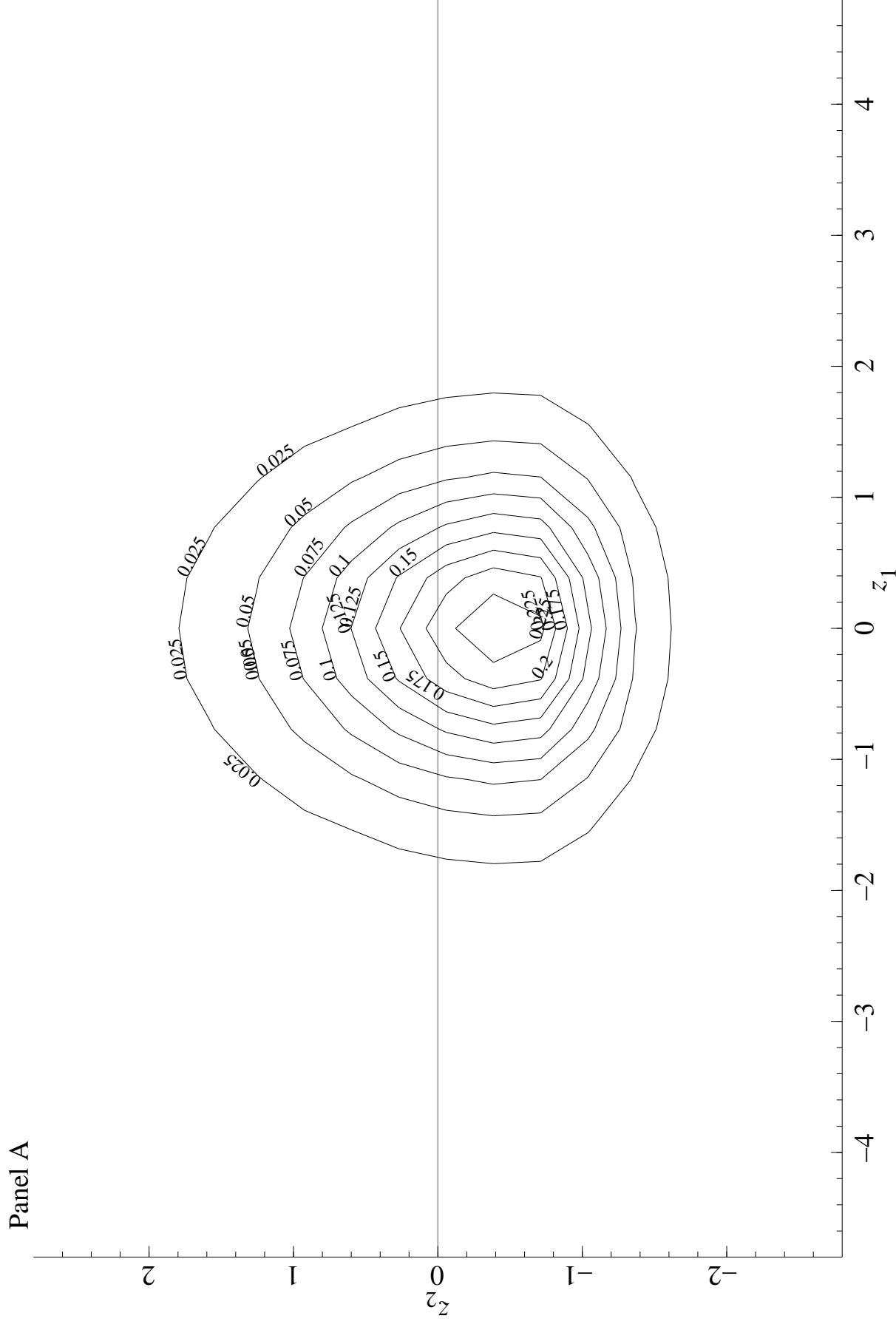
Skewed-Student density (Bauwens and Laurent, 2005)

- The multivariate skewed-Student density, denoted $SKST(0, I_N, \xi, \nu)$ (ξ and ν correspond to η), is a skewed version of the $ST(0, I_N, \nu)$.
- $\xi = (\xi_1 \dots \xi_N)$ is a vector of skewness parameters, with $\xi_i > 0$ for all i .
- ξ_i governs the skewness of z_{it} since $\xi_i^2 = \frac{\text{Prob}(z_{it} > \text{mode})}{\text{Prob}(z_{it} < \text{mode})}$.
If $0 < \xi_i < 1$, the marginal of z_{it} is left-skewed, while if $\xi_i > 1$, it is right-skewed.
- If $\xi_i = 1 \forall i$, the skewed-Student reduces to the Student density. Hence, one can test the null hypothesis of symmetry.

Univariate skewed-Student densities



Contours of SKST $\xi_1 = 1, \xi_2 = 1.3, \nu = 6$



Another way to get multivariate distributions

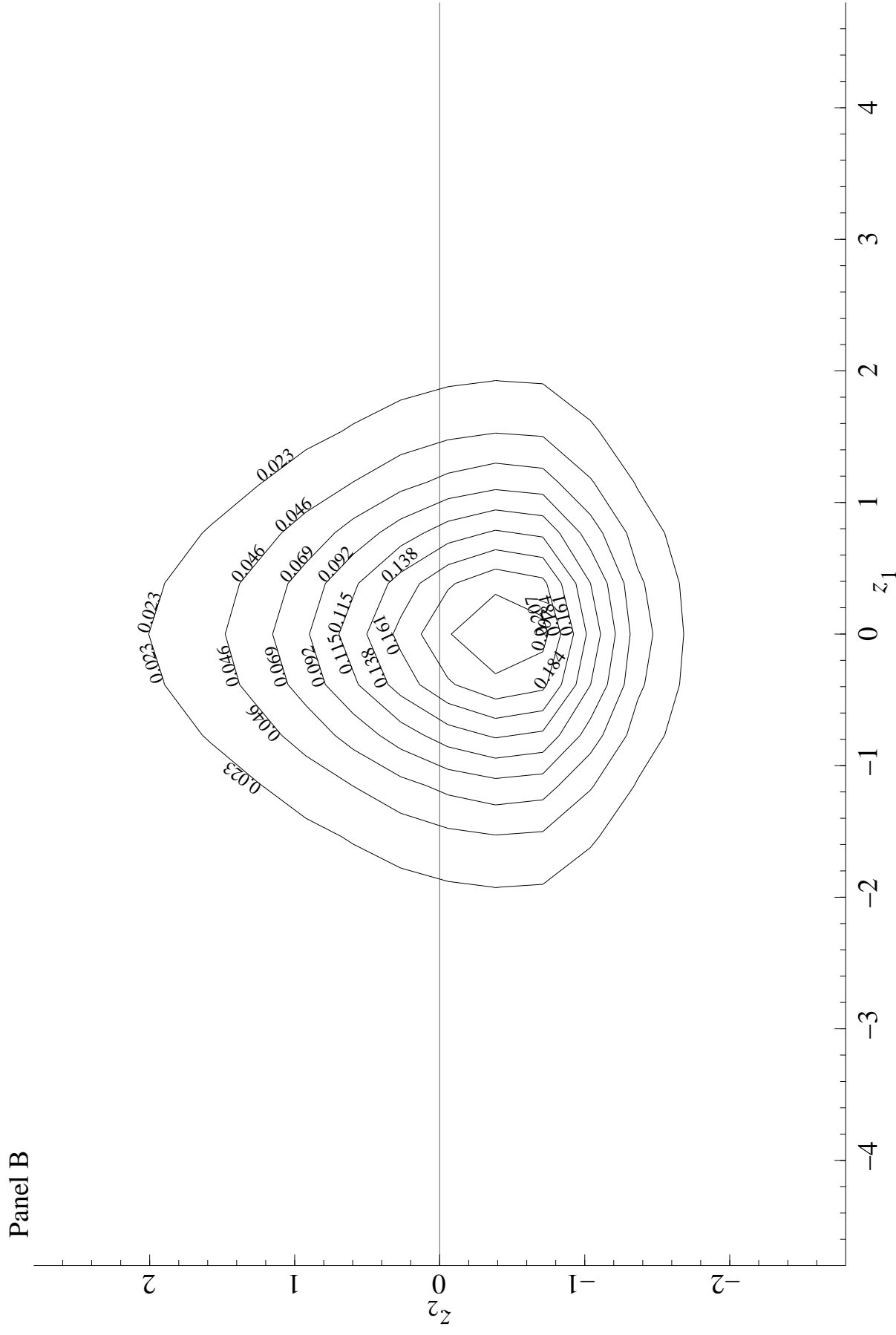
One can also define the density of z_t as the product of independent univariate densities for each element of z_t :

- univariate Student $ST(0, 1, \nu_i)$ (with their own degrees of freedom, not the same for each marginal);
- univariate skewed-Student $SKST(0, 1, \xi_i, \nu_i)$ (Bauwens and Laurent, 2005);
- GED(ν_i).

Not much implemented up to now...

This allows more flexibility, but may render estimation more difficult since there is more parameters in η .

SKST-IC $\xi_1 = 1, \xi_2 = 1.3, \nu_1 = \nu_2 = 6$



Asymptotic properties of ML & QML

- Consistency of QMLE is shown (Bollerslev and Wooldridge, 1992; Jeantheau, 1998).
- Asymptotic normality "assumed" in practice (or shown using high level assumptions).
- Hence, in practice one does inference as usual (asymptotic χ^2 Wald and likelihood ratio tests).
- Recent work on these issues in univariate GARCH models has shown that usual asymptotics does not necessarily hold if z_t does not have moments of low order (4 at least). See Hall and Yao (2003).

Two step estimation of DCC models

- This approach uses the Gaussian likelihood (ML under the normality assumption, or QML otherwise).
- Substituting $H_t = D_t R_t D_t$ in the Gaussian log-likelihood function (4) gives:

$$L_T(\theta) = -\frac{1}{2} \sum_{t=1}^T (\log | D_t R_t D_t | + u_t' R_t^{-1} u_t) \quad (6)$$

where $u_t = D_t^{-1}(y_t - \mu_t)$, so that

$$u_t' R_t^{-1} u_t = (y_t - \mu_t)' D_t^{-1} R_t^{-1} D_t^{-1} (y_t - \mu_t).$$

Two step estimation of DCC models

Hence, we can write:

$$\begin{aligned}L_T(\theta) &= -\frac{1}{2} \sum_{t=1}^T (\log | D_t R_t D_t | + u_t' R_t^{-1} u_t) \\ &= -\frac{1}{2} \sum_{t=1}^T (2 \log | D_t | + u_t' u_t) \leftarrow \mathbf{Q1L_T}(\theta_1^*) \\ &\quad -\frac{1}{2} \sum_{t=1}^T (\log | R_t | + u_t' R_t^{-1} u_t - u_t' u_t) \leftarrow \mathbf{Q2L_T}(\theta_1^*, \theta_2^*)\end{aligned}$$

θ_1^* : parameters of the conditional variances (D_t),

θ_2^* : parameters of the conditional correlations (R_t).

Two step estimation of DCC models

- $L_T(\theta) = L_T(\theta_1^*, \theta_2^*) = Q_1 L_T(\theta_1^*) + Q_2 L_T(\theta_1^*, \theta_2^*)$.
- First step: estimate θ_1^* by

$$\hat{\theta}_1^* = \operatorname{argmax} Q_1 L_T(\theta_1^*)$$

Easy: estimate separate univariate GARCH models if there is no spillover effects in conditional variances.

- Second step: estimate θ_2^* by

$$\hat{\theta}_2^* = \operatorname{argmax} Q_2 L_T(\hat{\theta}_1^*, \theta_2^*).$$

Easy, since many parameters are fixed in this step.

Remarks

- These estimators are consistent but not efficient asymptotically, since some information is sacrificed (about θ_1^* in the first step).
- The variance matrix of the estimator $\hat{\theta}_2^*$ has to be adjusted to take account of the first step (see Engle, 2002 and Newey and McFadden, 1994) but this is not important for VaR forecasts.

Example of Bauwens and Laurent (2005)

- 2 datasets of daily returns:
 - 3 stocks: Alcoa (AA), Caterpillar Inc. (CAT), and Walt Disney Company (DIS), from January 1990 to May 2002.
 - 3 exchange rates with respect to US dollar: euro (DM before euro period), yen, and British pound, from January 1989 to February 2001.
- Conditional means: AR(0) or AR(1) with constant.
- Conditional variances: GARCH(1,1) for exchange rates, and GJR(1,1) for stocks:
$$h_t = \omega + \alpha \epsilon_{t-1}^2 + \gamma \epsilon_{t-1}^2 I_{\{\epsilon_{t-1} < 0\}} + \beta h_{t-1}.$$
- DCC model of Engle (2002) for conditional correlation matrix, with skewed-Student distribution.

Partial estimation results

	AA-CAT-DIS		EUR-YEN-GBP	
	1 step	2 steps	1 step	2 steps
θ_1	0.0088 (0.0021)	0.0095	0.0303 (0.0033)	0.0294
θ_2	0.9846 (0.0047)	0.9837	0.9684 (0.0037)	0.9689
$\log \xi_1$	0.1050 (0.0257)	0.0977	-0.0875 (0.0242)	-0.0724
$\log \xi_2$	0.0786 (0.0263)	0.0698	0.0987 (0.0253)	0.0983
$\log \xi_3$	0.0667 (0.0276)	0.0591	-0.0677 (0.0238)	-0.0353
ν	7.2858 (0.5335)	7.4020	6.1928 (0.3960)	6.4896
Sample size	3113		3066	
$LRT(CCC)$	58.41		895.62	
$LRT(SY)$	34.58		33.45	

Note: For each parameter, the table reports the one step ML estimate and its standard error (in parantheses). The estimate of the two-step approach is also reported.

$LRT(CCC)$ and $LRT(SY)$ are respectively $\chi^2(2)$ and $\chi^2(3)$ likelihood ratio statistics for the hypotheses of constant correlations and symmetry with respect to the Student density.

Variance and correlation targeting

- The constant part of H_t , if unrestricted, contains $N(N + 1)/2$ parameters, a number that increases fastly with N .
- This constant part, or a function of it, can sometimes be estimated consistently without doing ML or QML. Then this consistent estimate can be substituted for the corresponding parameter matrix in the (quasi-)log-likelihood function, rendering maximization easier by the reduction in the number of parameters.
- These estimators are consistent but not efficient asymptotically, since some information is sacrificed.
- Correlation targeting: a similar argument can be applied to estimate R in (1) and \bar{Q} in (3).

Example of variance targeting

- In the VEC model, we know that $\text{vech } \Sigma = E(\text{vech } H_t) = (I_{N^*} - A - G)^{-1}c$. Hence we can write the VEC model as $h_t = (I_{N^*} - A - G)(\text{vech } \Sigma) + A\eta_{t-1} + Gh_{t-1}$.
- A consistent estimator of Σ is

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t \hat{\epsilon}_t'$$

where $\hat{\epsilon}_t = y_t - \hat{\mu}_t$, with $\hat{\mu}_t$ a consistent estimator of μ_t (usually easily available, e.g. by OLS or QML).

- Hence we estimate A and G from $h_t = (I_{N^*} - A - G)(\text{vech } \hat{\Sigma}) + A\eta_{t-1} + Gh_{t-1}$.

Limits of MGARCH models

- Curse of dimensionality, but Engle and Sheppard estimate a DCC model for 100 series, and Palandri (2005) estimates a more flexible model (SCC) for 69 series.
- Estimation software not yet enough developed. MG@RCH under development by S. Laurent and others (similar to G@RCH, belongs to OxMetrics). Otherwise, RATS, FinMetrics in S+, Fanpac in GAUSS include some models.

New developments in MGARCH models

- More flexible DCC and factor models.
- Finite mixtures of GARCH models.
- Semi-parametric estimation.