Università Commerciale
Luigi Bocconi

The Instability of Correlations: Measurement and the Implications for Market Risk

Prof. Massimo Guidolin

20541 – Advanced Quantitative Methods for Asset Pricing and Structuring

Winter/Spring 2019

Plan of the Lecture

$$\text{Covariance matrix} = \underbrace{\begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_n \end{bmatrix}}_{\substack{\text{Diagonal matrix with} \\ \text{standard deviations in} \\ \text{the diagonal (and} \\ \text{zeros in the other} \\ \text{cells)}}} \text{x} \underbrace{\begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \cdots & \cdots & 1 \end{bmatrix}}_{\text{Correlation matrix}} \text{x} \underbrace{\begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_n \end{bmatrix}}_{\substack{\text{Diagonal matrix with} \\ \text{standard deviations in} \\ \text{the diagonal (and zeros} \\ \text{in the other cells) –} \\ \text{same as the one before}}}$$

- Overview and Introduction: why Econometrics?
- Merton's credit risk model and the use of correlations as inputs
- From multivariate ARCH models to DCC models
- Markov switching models, alone or in ARCH combo
- Stochastic Volatility Models

Overview and Introduction

- Two key methods to forecast credit correlations: using empirical joint default rates and/or rating migrations; structural credit risk modelling based on underlying correlations

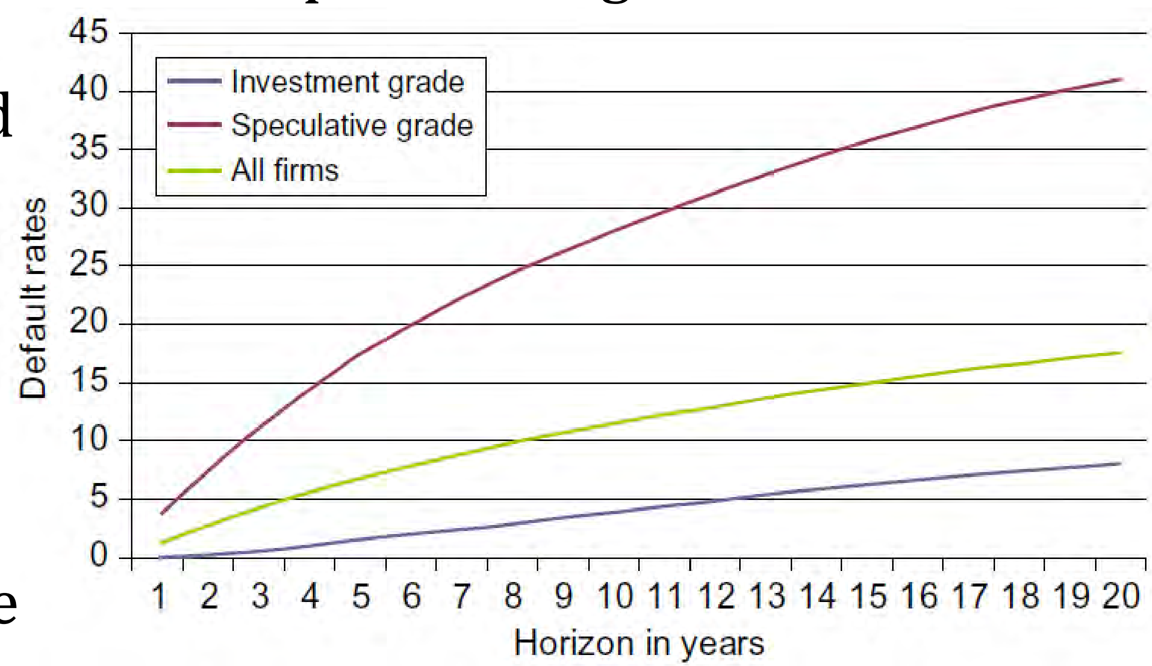
- The interaction—often captured by correlations, i.e., statistical measures of linear association—of different credit risk positions and instruments is at heart of all model-based internal rating systems
- Correspondingly, the trading of OTC derivatives that allow market participants to directly trade credit risk correlations is growing
  - E.g., collateralized debt obligations (CDOs)
- While the prices of such instruments should reflect the assessment of credit risk correlations by traders, the pricing of these derivatives ought to depend on forecasts of correlations, creating a logical loop
- There are two major approaches to estimating and forecasting credit correlations, both subject to problems and limitations

① Estimation from empirical default rates or rating migrations

- Approach difficult because of scarcity of joint default or migrations

Overview and Introduction

- Credit rating agencies such as Moody's and Standard & Poor's maintain databases of corporate defaults through time
  - In Moody's definition corporate default is triggered by one of three events: (1) a missed or delayed interest or principal payment, (2) a bankruptcy filing, or (3) a distressed exchange where old debt is exchanged for new debt that represents a smaller obligation
- The average corporate default rate for speculative grade US firms was 2.78% per year during the entire 1920–2014 period



- For investment grade firms the average was just 0.15% per year
- These very cumulative rates show that in some rating clusters, the defaults are rare
  - E.g., over a 20-year horizon there is an 8.4% probability of an investment grade firm to default

Overview and Introduction

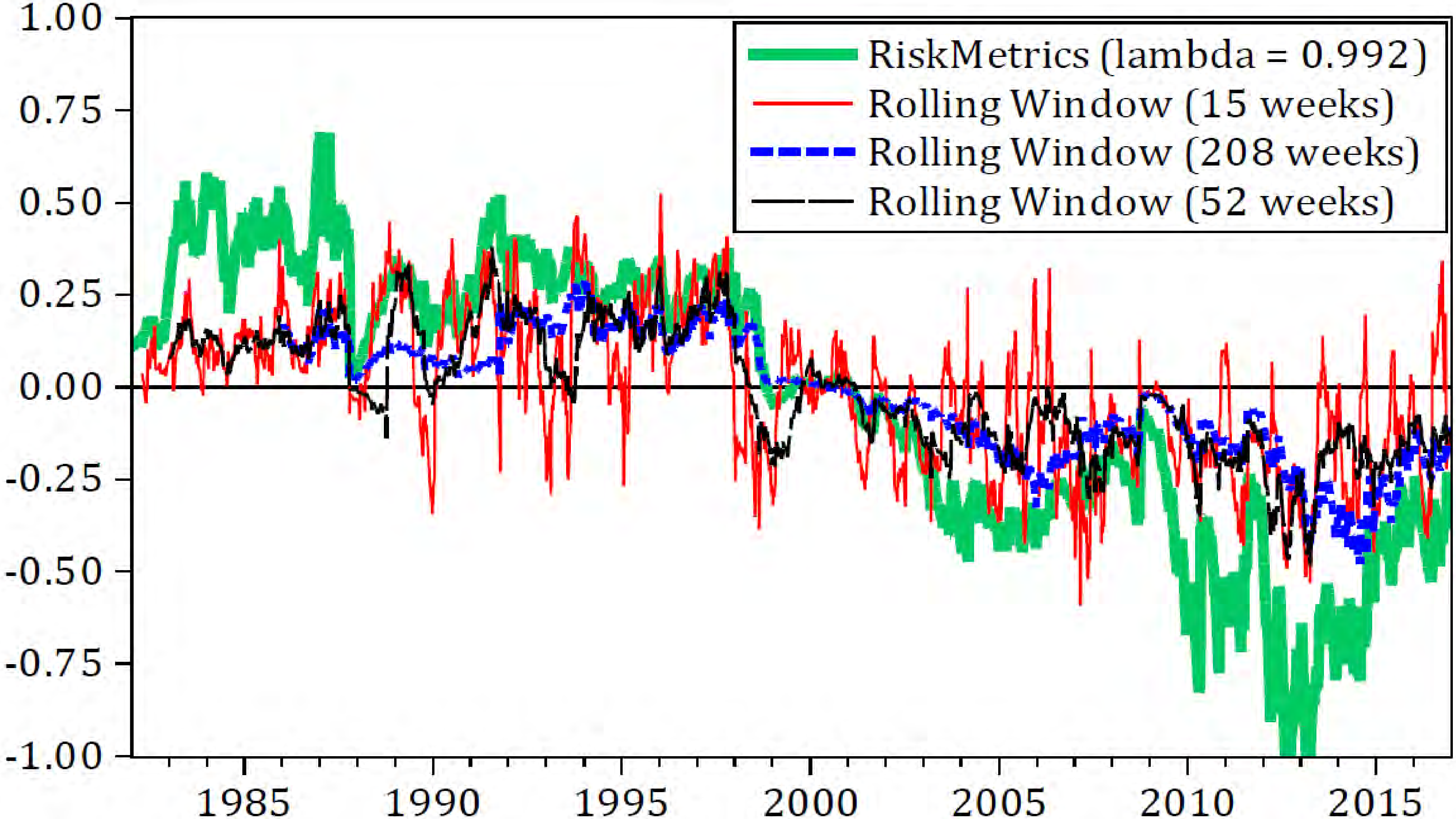
- Underlying equity correlations are strongly time varying and they can be easily predicted using time series methods

② Extensions of structural credit risk models à la Merton (1974) from a univariate to a multivariate framework

- Practitioners frequently use equity correlations as proxies for asset correlations, with corrections to reflect the fact that stock returns may be affected by factors unrelated to credit risk
- This gives them one key advantage: the methods are flexible enough to capture the fact that correlations are strongly time-varying
- This is where econometrics comes in
  - Through simple and yet flexible ways: factor models in which regression betas capture commonalities that drive credit risk correlations
  - Or through latent factors/effects
    - _ Multivariate GARCH models
    - _ Dynamic Conditional Correlation Models
    - _ Markov and Regime Switching Models
    - _ Dynamic Copula Models

Overview and Introduction

- Underlying equity correlations are strongly time varying and they can be easily predicted using time series methods



·Rolling Window and RiskMetrics Stock-Bond Correlations

Where Do Correlations Enter? Merton's Model

- In Merton's model, equity is equivalent to a long call option on the assets of the firm

- Consider the situation where we are exposed to the risk that a particular firm defaults
    - This risk could arise from the fact that we own stock in the firm, or it could be that we have lent the firm cash
    - Or because the firm is a counterparty in a derivative transaction with us
- Assume that the firm is financed with debt and equity and all the debt expires at time t+T
- The face value of the debt is D and it is fixed; the future asset value of the firm, A_{t+T}, is uncertain
- By first principles, at t+T when the company's debt is due the firm will continue to operate if $A_{t+T} \geq$ D but the firm's debt holders will declare the firm bankrupt if $A_{t+T} <$ D and the firm will go into default
- Because the shareholders are residual claimants, to stock holders the firm is worth $E_{t+T} = \max\{A_{t+T} - D, 0\}$ the payoff of a call with strike D

Where Do Correlations Enter? Merton's Model

- In Merton's model, debt is equivalent to a short put option on the assets of the firm

- Note that the asset value of the firm is the risky variable
- Assuming that asset volatility and the risk-free rate are constant, and assuming that the log asset value is normally distributed we get that by Black-Scholes:

$$E_t = A_t \Phi(d) - D\exp(-r_f T)\Phi\left(d - \sigma_A\sqrt{T}\right) \quad d = \frac{\ln(A_t/D) + (r_f + \sigma_A^2/2)T}{\sigma_A\sqrt{T}}$$

- Because $A_{t+T} = D_{t+T} + \max\{A_{t+T} - D, 0\}$ where D_{t+T} is the mkt. value of debt, the debt holders look as if they sold a put option with strike equal to the face value of the debt and go long in the risk-free asset:

$$\begin{aligned} D_{t+T} &= A_{t+T} - \max\{A_{t+T} - D, 0\} \\ &= D - \max\{D - A_{t+T}, 0\} = e^{-r_f T} D\Phi\left(d - \sigma_A\sqrt{T}\right) - A_t\Phi(-d) \end{aligned}$$

from Black-Scholes' standard formula

- For the model to be implemented we need asset volatility σ_A and the mkt. value of the asset A_t to be solved from one identity, $E_t = S_t N_t$ (share price x number of shares of stock)

Where Do Correlations Enter? Merton's Model

- When Merton's model is used to price corporate debt and assess default risk, this occurs "off stock returns properties"

and 2 equations: $\boxed{E_t = A_t \Phi(d) - D\exp(-r_f T)\,\Phi\left(d - \sigma_A \sqrt{T}\right)}$ $\boxed{\sigma E_t = \Phi(d)\,\sigma_A A_t}$

- One of the key inputs is then stock returns volatility, σ
- A powerful feature of Merton's model is that we can use it to price corporate debt on firms even without observing the asset value as long as time series of stock returns are available
- The risk-neutral probability of default is the prob. that the put option is exercised: $\Pr(A_{t+T} < D) = 1 - \Phi\left(d - \sigma_A \sqrt{T}\right) = \Phi\left(\sigma_A \sqrt{T} - d\right)$
  - This probability of default is constructed from the risk-neutral distribution (where assets grow at the risk-free rate) of asset values and so it may well be different from the actual physical probability
- Yet the physical default probability can be easily derived by modifying the drift of asset returns, for given volatility
- Default risk is also measured in terms of distance to default: $dd = d - \sigma_A \sqrt{T} = \dfrac{\ln(A_t/D) + \left(r_f - \sigma_A^2/2\right)T}{\sigma_A \sqrt{T}}$

Where Do Correlations Enter? Merton's Model

- Estimating and forecasting correlations for different firms/ stocks is a key input in Merton's model ptf applications

- In practice, we need to manage the credit risk of portfolios of debt
- Default is a highly nonlinear event and is correlated across firms and so credit risk is likely to impose limits on diversification benefits
- Certain credit derivatives, such as CDOs, depend on the correlation of defaults that we need to model
- Consider a multivariate version of Merton's model in which the asset value of each firm *i* is log normally distributed:

$$\ln\left(A_{i,t+T}\right) = \ln A_{i,t} + r_f T - \frac{1}{2}\sigma^2_{A,i} T + \sigma_{A,i}\sqrt{T} z_{i,t+T}$$

where $z_{i,t+T}$ is a standard normal variable

- Using same steps as before, the probability of default for the ptf is.

$$\Pr\left(A_{t+T} < D\right) = 1 - \Phi\left(d - \sigma_A\sqrt{T}\right) = \Phi\left(\sigma_A\sqrt{T} - d\right)$$

$$dd = d - \sigma_A\sqrt{T} = \frac{\ln\left(A_t/D\right) + \left(r_f - \sigma^2_A/2\right)T}{\sigma_A\sqrt{T}}$$

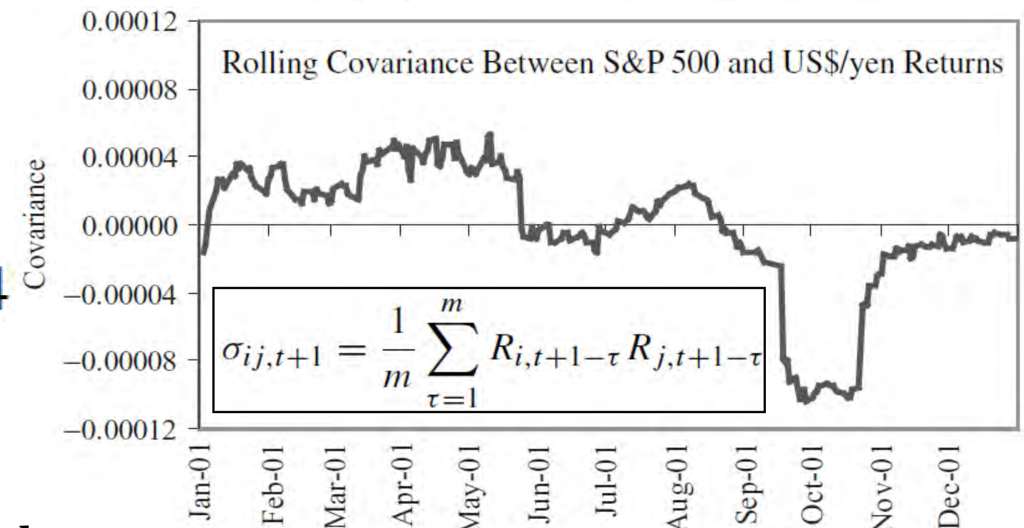
Depend on correlation matrix of equity returns of different firms

Modeling GARCH Conditional Covariances

- The simplest idea is to build time-varying estimates of covariances using rolling (moving) averages,

$$\sigma_{ij,t+1} = \frac{1}{m}\sum_{\tau=1}^{m} R_{i,t+1-\tau} R_{j,t+1-\tau}$$

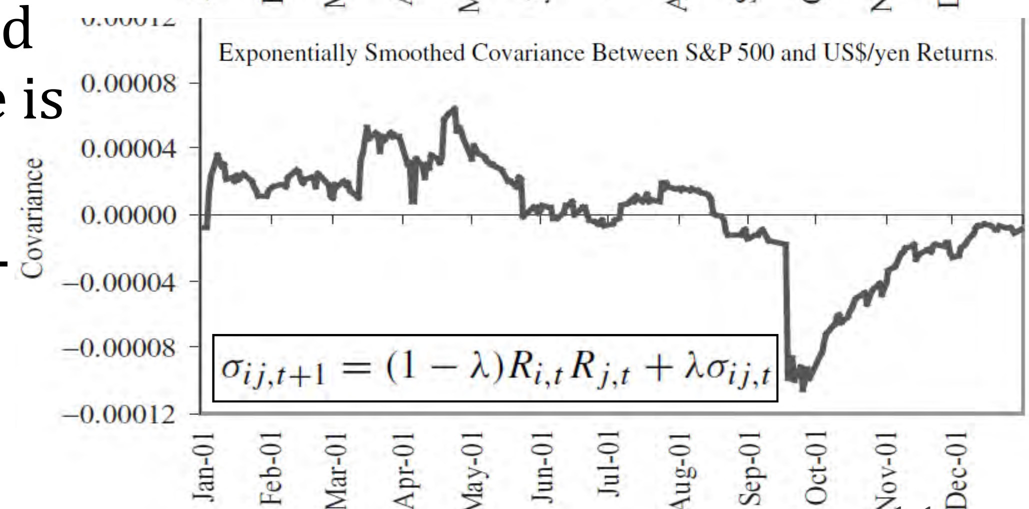
  - Not really satisfactory because the choice of *m* is problematic



- We can use simple exponential smoother on covariances $\lambda = .94$

$$\sigma_{ij,t+1} = \boxed{(1-\lambda)} R_{i,t} R_{j,t} + \boxed{\lambda} \sigma_{ij,t}$$

  - The restriction that coefficient $(1 - \lambda)$ on the cross products and λ on past covariance sum to one is not necessarily desirable
  - It implies that there is no mean-reversion in covariance
  - A high covariance will remain high forever!



Modeling GARCH Conditional Covariances

- GARCH models may be fruitfully extended to modelling covariances, even though restrictions are needed to keep the covariance matrix (semi) positive definite

- This follows from the fact that the model can be written in GARCH(1,1) form as $\sigma_{ij,t+1} = \omega_{ij} + \alpha R_{1,t}R_{2,t} + \beta\sigma_{ij,t}$ which is to be compared to:

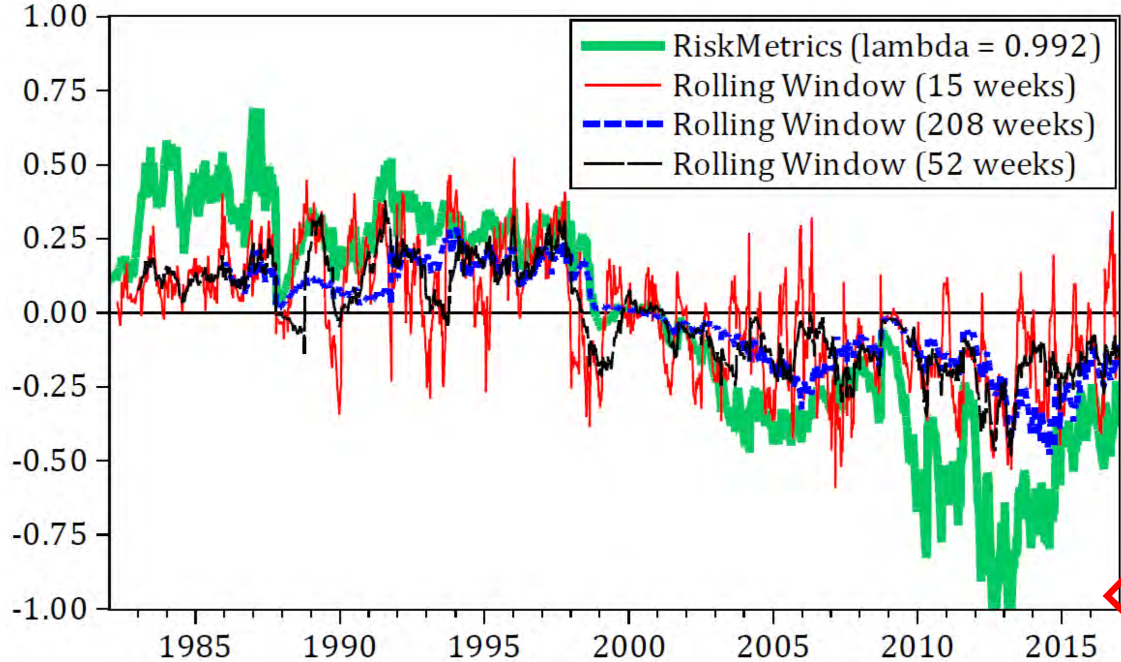
$$\sigma_{ij,t+1} = (1-\lambda)R_{1,t}R_{2,t} + \lambda\sigma_{ij,t} \quad \Rightarrow \omega_{ij} = 0, \alpha = 1-\lambda, \lambda = \beta, \alpha + \beta = 1$$

  so that, as a result, $E[\sigma_{ij,t+1}] = \omega_{ij} /(1- \alpha - \beta)$ fails to be defined
- The next step is then rather obvious: let's not restrict $\alpha = (1 - \beta)$ in the GARCH(1,1) type model for conditional covariance:

$$\sigma_{ij,t+1} = \omega_{ij} + \alpha R_{i,t} R_{j,t} + \beta\sigma_{ij,t}$$

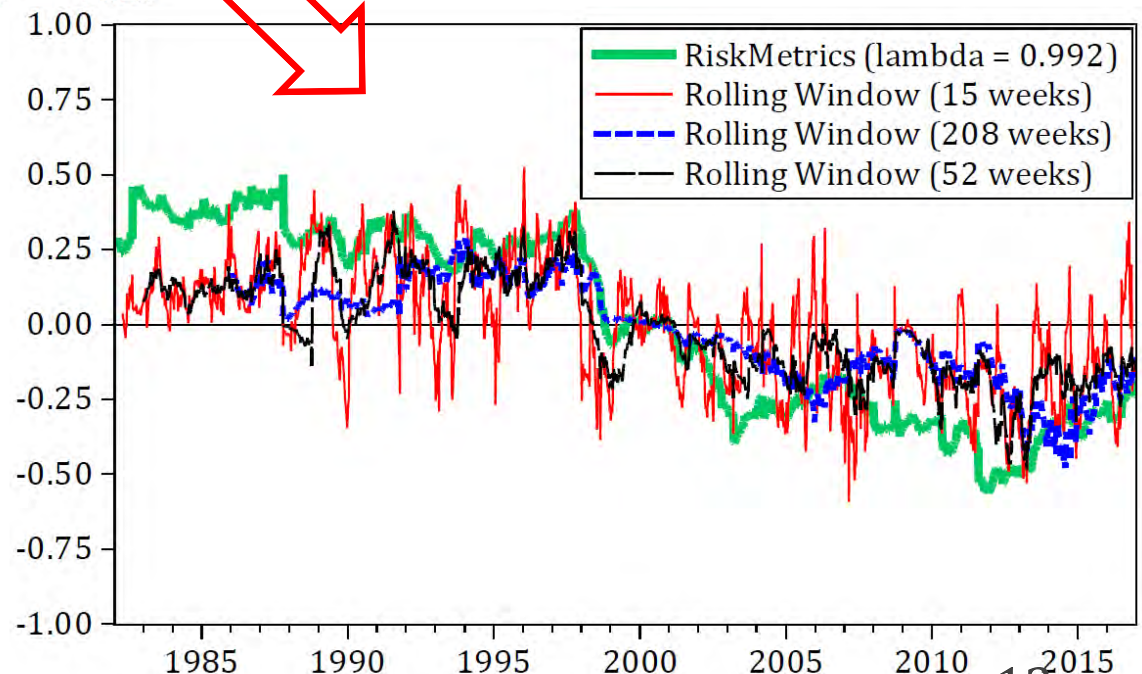
- When $|\alpha + \beta| < 1$, the process is stationary and the unconditional covariance will equal $\omega_{ij} /(1- \alpha - \beta)$
  - Why are we restricting α and β to NOT depend on the specific pair of securities/assets examined?
  - Setting α and β (and λ) not to depend on *i* and *j* yields good outcomes

Modeling GARCH Conditional Covariances



Rolling Window and RiskMetrics Stock-Bond Correlations

Imposing SPD restrictions…



Rolling Window and RiskMetrics Stock-Bond Correlations

Modeling GARCH Conditional Covariances

- Restricting parameters to not depend on *i* and *j* guarantees that the resulting covariance matrix that collects GARCH(1,1) variances and covariance is (semi) positive definite, i.e., that for all possible vectors *w*, $$w'\Sigma_{t+1}w \geq 0$$
- Why is that relevant? Well, just recall that $\sigma^2_{PF,t+1} = w'\Sigma_{t+1}w$
  - This SPD condition is ensured by estimating volatilities and covariances in an "internally consistent fashion"
  - Sufficient condition for internal consistency is the use of the same λ for every volatility and covariance in exponential smoothing
  - Similarly, using a GARCH(1,1) model with α and β identical across variances and covariances is sufficient
  - Not clear that the persistence parameters λ, α, and β should be the same for all variances and covariances: need to develop better models
- Idea: if we could jointly model variances and covariances, then we could impose restrictions that are less heroic, at the cost of higher mathematical complexity

Multivariate GARCH Models: VECH(1,1)

- A Vech GARC(1,1) model is based on the idea of modeling covariance matrices as column-stacked vectors

- Multivariate GARCH models are in spirit similar to their univariate counterparts, except that they also specify conditional covariance functions, i.e., covariances directly move over time
- Several different multivariate GARCH formulations, e.g., VECH, the diagonal VECH and the BEKK models
  - Below it is assumed for simplicity that there are n = 2 assets
- A VECH(1,1) model is specified as:

$$VECH(H_t) = C + AVECH(\Xi_{t-1}\Xi'_{t-1}) + BVECH(H_{t-1}) \quad \Xi_t|\psi_{t-1} \sim N(0, H_t)$$

  - H_t is a conditional covariance matrix, Ξ_{t-1} is an innovation (disturbance) vector, ψ_{t-1} represents the information set at time t – 1
  - C is a 3 × 1 parameter vector, A and B are 3 × 3 parameter matrices and VECH(·) denotes the column-stacking operator applied to the upper portion of the symmetric matrix H_t
  - The model requires the estimation of 21 parameters, a lot!

Multivariate GARCH Models : VECH(1,1)

- How the VECH operator works is shown below:

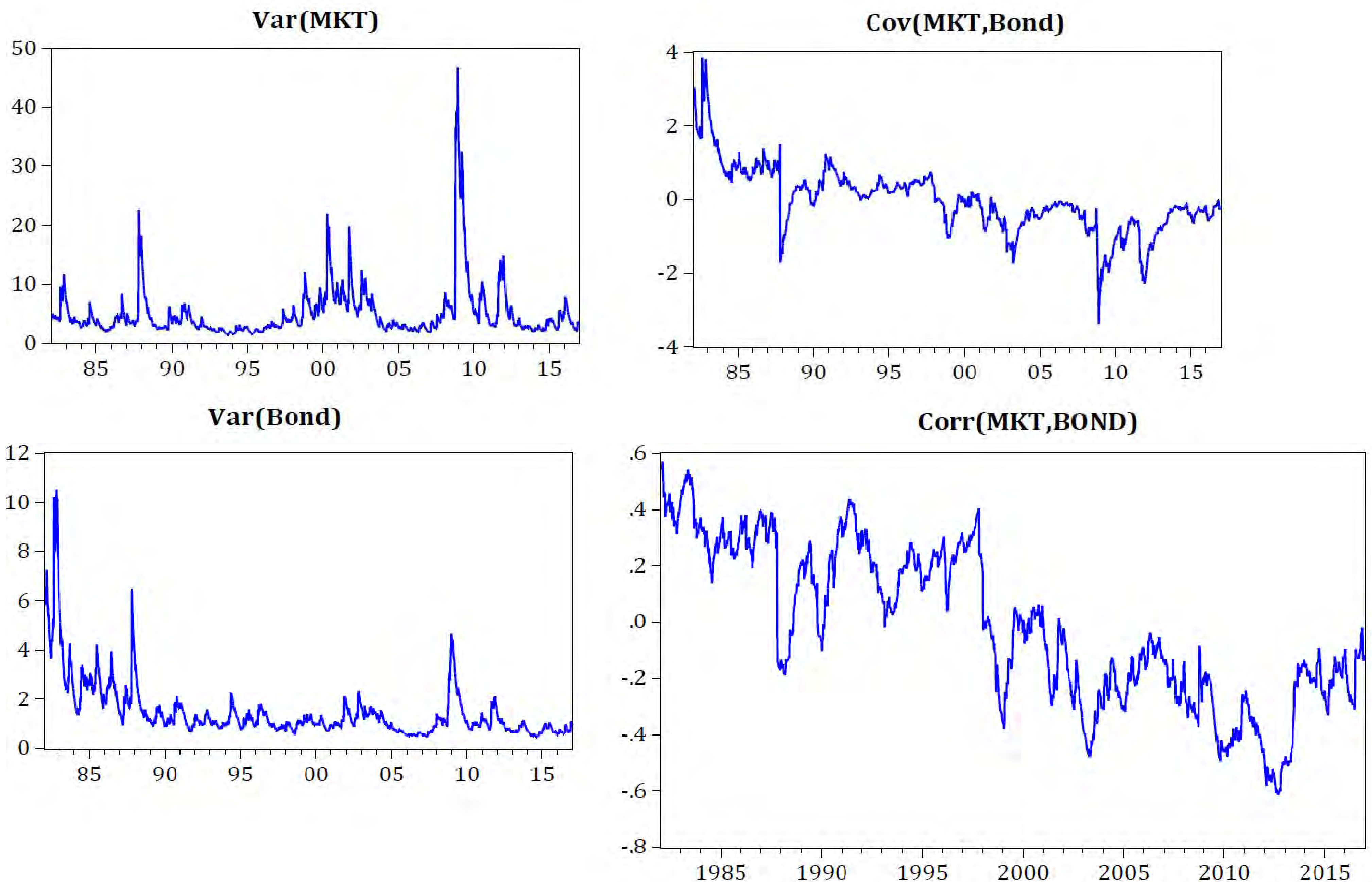
$$VECH(\Xi_t \Xi_t') = VECH\left(\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}\begin{bmatrix} u_{1t} & u_{2t} \end{bmatrix}\right) = VECH\begin{pmatrix} u_{1t}^2 & u_{1t}u_{2t} \\ u_{1t}u_{2t} & u_{2t}^2 \end{pmatrix} = \begin{bmatrix} u_{1t}^2 \\ u_{2t}^2 \\ u_{1t}u_{2t} \end{bmatrix}$$

- The elements for the case *n* = 2 are written out below:

$$\begin{aligned} h_{11t} &= c_{11} + a_{11}u_{1t-1}^2 + a_{12}u_{2t-1}^2 + a_{13}u_{1t-1}u_{2t-1} + b_{11}h_{11t-1} \\ &\quad + b_{12}h_{22t-1} + b_{13}h_{12t-1} \\ h_{22t} &= c_{21} + a_{21}u_{1t-1}^2 + a_{22}u_{2t-1}^2 + a_{23}u_{1t-1}u_{2t-1} + b_{21}h_{11t-1} \\ &\quad + b_{22}h_{22t-1} + b_{23}h_{12t-1} \\ h_{12t} &= c_{31} + a_{31}u_{1t-1}^2 + a_{32}u_{2t-1}^2 + a_{33}u_{1t-1}u_{2t-1} + b_{31}h_{11t-1} \\ &\quad + b_{32}h_{22t-1} + b_{33}h_{12t-1} \end{aligned}$$

- Conditional variances and conditional covariances depend on the lagged values of all of the conditional variances of, and conditional covariances between, all of the asset returns in the series, as well as the lagged squared errors and the error cross-products
- As *n* increases, the estimation of the VECH model quickly becomes infeasible

Example of Predicted Stock-Bond Correlations from VECH



Multivariate GARCH: Diagonal VECH and BEKK

- A Diagonal Vech GARC(1,1) model is a VECH(1,1) in which the matrices of parameters are restricted to be diagonal

- VECH conditional covariance matrix may be restricted so that A and B are assumed to be diagonal
  - This reduces the number of parameters to be estimated to 9 (A and B each have 3 elements) and the model, known as a diagonal VECH, is:

$$h_{ij,t} = \omega_{ij} + \alpha_{ij} u_{i,t-1} u_{j,t-1} + \beta_{ij} h_{ij,t-1} \quad \text{for } i, j = 1, 2$$

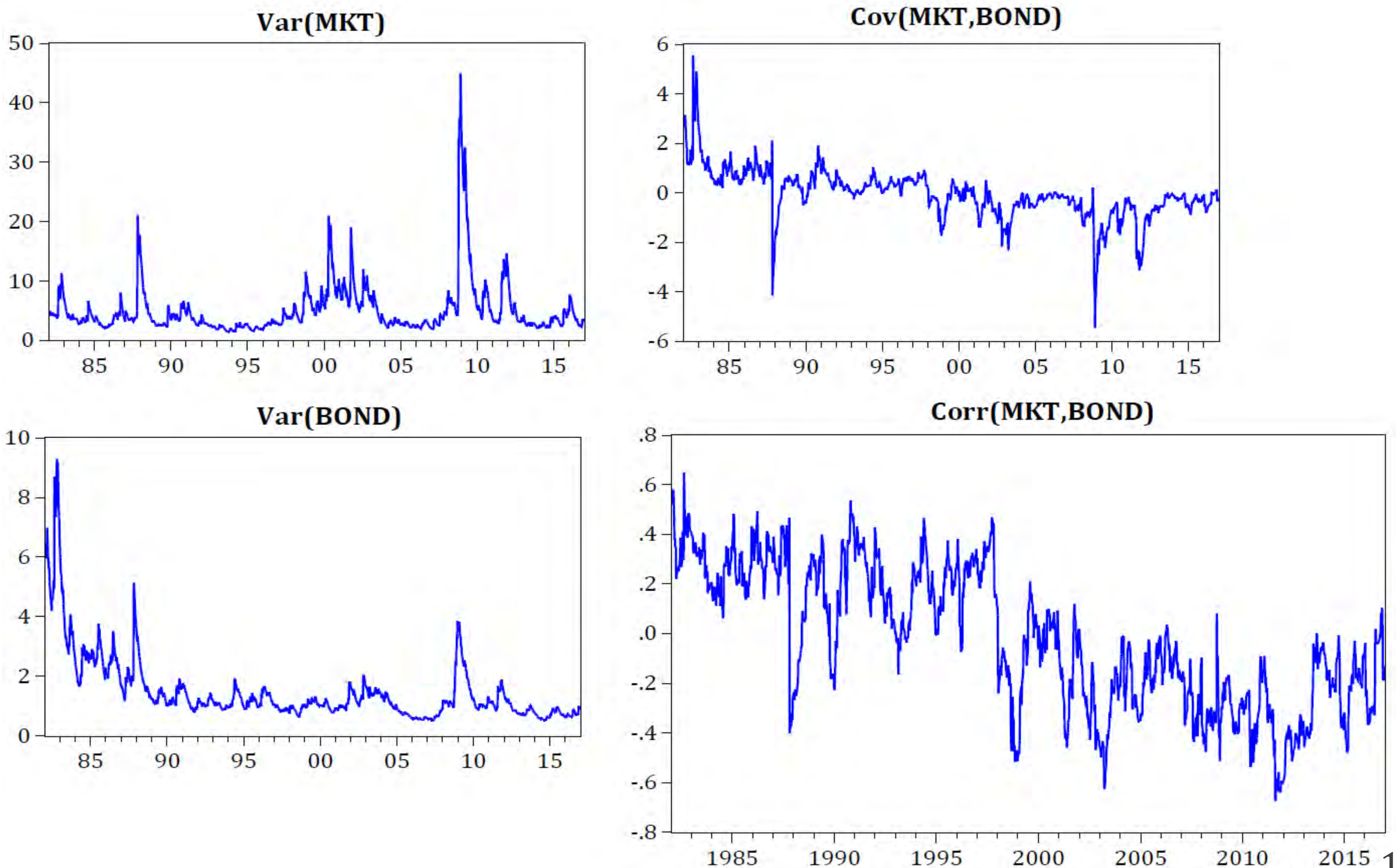
- A disadvantage of the VECH model is that there is no guarantee of a positive semi-definite covariance matrix
  - It is this property which ensures that, whatever the weight of each asset in the portfolio, an estimated value-at-risk is always positive
- The BEKK model addresses the difficulty with VECH of ensuring that the *H* matrix is always positive definite:

$$H_t = W'W + A'H_{t-1}A + B'\Xi_{t-1}\Xi'_{t-1}B$$

  - A and B are 2 × 2 matrices of parameters and W is upper triangular

Multivariate GARCH: Diagonal VECH and BEKK

- The positive definiteness of the covariance matrix is ensured owing to the quadratic nature of the terms on the equation's RHS



Multivariate GARCH: Pros and Cons

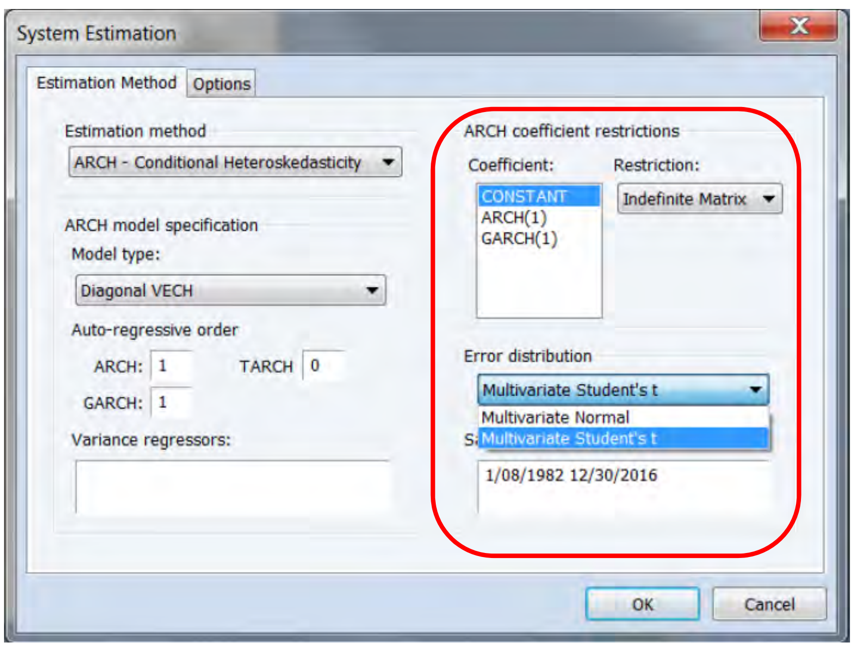
- Multivariate GARCH models present pros and cons:

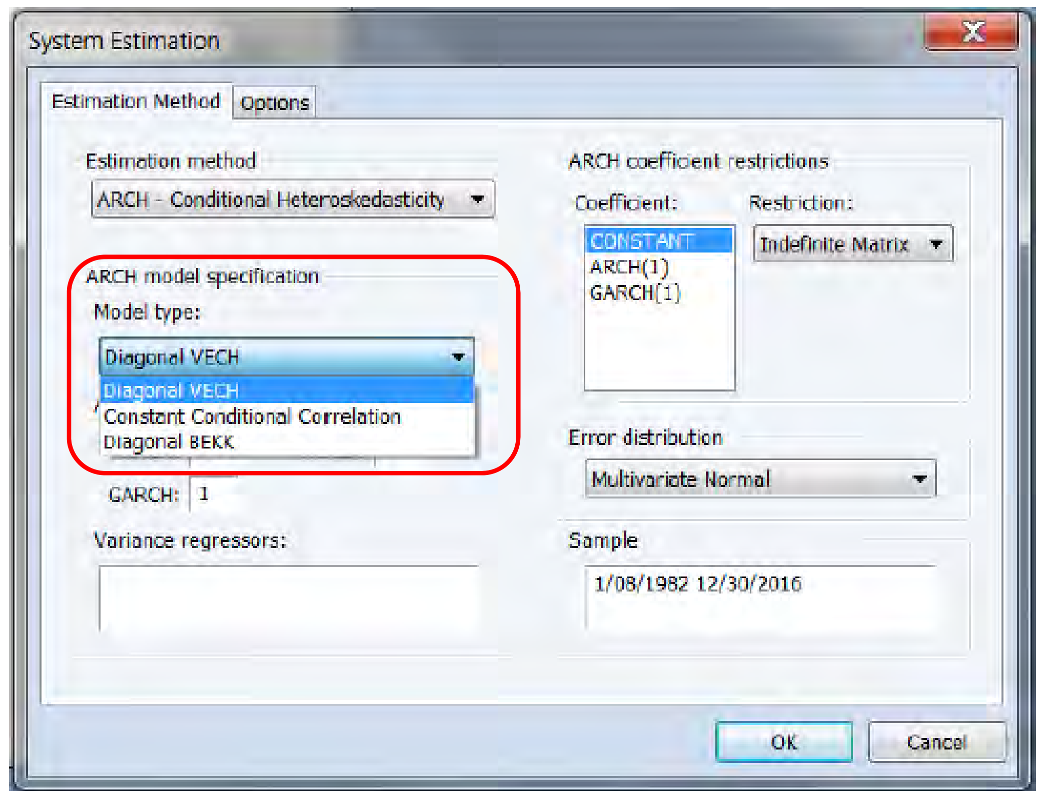
⊕ Transparent restrictions to impose positivity of predicted variances, predicted correlations in [-1, +1], and statistical properties

  - The key one is stationarity of the resulting process

⊕ Apart from restrictions, the full form of the VEC, Full-Rank Multi, and BEKK models allow all second moment to depend on shocks to all assets/industries/series

⊕ Increasingly easy to apply, in spite of the issues listed below





Multivariate GARCH: Pros and Cons

- Multivariate GARCH models present pros and cons:

⊖ They tend to suffer from a curse of parameter dimensionality, unless tight restrictions are imposed, fully general models easily carry hundreds of parameters to estimate!

  - For instance, BEKK is characterized by $0.5N(N+1)(1+p+q)=O(N^2)$ parameters to estimate
  - With N = 100 in a simple BEKK(1,1), these are 15,150 parameters
  - Even if you have 500 obs. per series, (500 x 100)/15150 = 3.33, less than 4 observations per parameter!
  - Bad news: BEKK is known to be “relatively” parsimonious

⊖ Because of the many parameters, numerical convergence within ML estimation algorithms represents an issue

⊖ When the model is as rich as possible, most of the (thousands of) estimated parameters fail to have any interpretation

⊖ Subject to some inconsistencies – only BEKK guarantees that a ptf. of BEKK assets follows a BEKK process

Dynamic Conditional Correlation Models

- The dynamic conditional correlation approach is based on the eigenvalue-eigenvector decomposition $\Sigma_{t+1} = D_{t+1}\Gamma_{t+1}D_{t+1}$
  - For instance, one interesting (worrisome) phenomenon is that all correlations tend to "skyrocket" during market crisis (bear)
  - Skyrocket is a way to speak: you do recall that a correlation,
  $$\rho_{ij,t+1} = \sigma_{ij,t+1}/(\sigma_{i,t+1}\sigma_{j,t+1})$$
  belongs to [-1, 1]
  - A first, intuitive but mechanical approach consists in applying GARCH models to both variances and covariances in the definition of conditional correlation, e.g.: $\sigma_{ij,t+1} = (1-\lambda)R_{i,t}R_{j,t} + \lambda\sigma_{ij,t}, \quad \text{for all } i, j$
  $$\Rightarrow \quad \rho_{ij,t+1} = \frac{(1-\lambda)R_{i,t}R_{j,t} + \lambda\sigma_{ij,t}}{\sqrt{((1-\lambda)R_{i,t}^2 + \lambda\sigma_{i,t}^2)((1-\lambda)R_{j,t}^2 + \lambda\sigma_{j,t}^2)}}$$
- A more fruitful approach still start from the decomposition $\sigma_{ij,t+1} = \sigma_{i,t+1}\sigma_{j,t+1}\rho_{ij,t+1}$ but it generalizes it to matrix form:
$$\boxed{\Sigma_{t+1} = D_{t+1}\Gamma_{t+1}D_{t+1}}$$

Dynamic Conditional Correlation Models

- The DCC approach is based on two steps: modelling the volatility of each individual asset; modelling the covariances of standardized residuals from the first step
  - Here D_{t+1} is a matrix of standard deviations, $\sigma_{i,t+1}$, on the ith diagonal and zero everywhere else
  - Γ_{t+1} is a matrix of correlations, $\rho_{ij,t+1}$ with ones on the diagonal
  - E.g., for n = 2:

$$\Sigma_{t+1} = \begin{bmatrix} \sigma^2_{1,t+1} & \sigma_{12,t+1} \\ \sigma_{12,t+1} & \sigma^2_{2,t+1} \end{bmatrix}$$
$$= \begin{bmatrix} \sigma_{1,t+1} & 0 \\ 0 & \sigma_{2,t+1} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12,t+1} \\ \rho_{12,t+1} & 1 \end{bmatrix} \begin{bmatrix} \sigma_{1,t+1} & 0 \\ 0 & \sigma_{2,t+1} \end{bmatrix}$$

- At this point we proceed in two steps:

① Volatilities of each asset are estimated through GARCH or one of the other methods considered in first part of the course

② Model conditional covariances of standardized returns derived from the first step $z_{i,t+1} = R_{i,t+1}/\sigma_{i,t+1} \quad \text{for all } i$

Dynamic Conditional Correlation Models

- The DCC approach is based on applying GARCH/RiskMetrics-type models to auxiliary variables that ensure $\rho_{ij,t+1} \in [-1,1]$
  - Luckily, the conditional covariance of the $z_{i,t+1}$ variables equals the conditional correlation of the raw returns:

$$E_t(z_{i,t+1}z_{j,t+1}) = E_t((R_{i,t+1}/\sigma_{i,t+1})(R_{j,t+1}/\sigma_{j,t+1}))$$
$$= E_t(R_{i,t+1}R_{j,t+1})/(\sigma_{i,t+1}\sigma_{j,t+1} = \sigma_{ij,t+1}/(\sigma_{i,t+1}\sigma_{j,t+1})$$
$$= \rho_{ij,t+1}, \text{ for all } i,\ j$$

  - You need to use an auxiliary variable $q_{ij,t+1}$ to be updated to be able to compute conditional correlations:

$$\rho_{ij,t+1} = \frac{q_{ij,t+1}}{\sqrt{q_{ii,t+1}q_{jj,t+1}}}$$

  - Why a need for the $q_{ij,t+1}$ auxiliary variable? Because being able to use the ratio above ensures $\rho_{ij,t+1}$ falls in the interval [-1,1]
  - At this point write a dynamic model for the conditional value for $q_{ij,t+1}$, like: $$q_{ij,t+1} = (1-\lambda)(z_{i,t}z_{j,t}) + \lambda q_{ij,t}, \quad \text{for all } i,\ j$$

    in this case of exponential smoothing with parameter λ

Dynamic Conditional Correlation Models

- An obvious alternative is a GARCH-type dynamic model:
$$q_{ij,t+1} = \bar{\rho}_{ij} + \alpha(z_{i,t}z_{j,t} - \bar{\rho}_{ij}) + \beta(q_{ij,t} - \bar{\rho}_{ij}) \text{ for all } i,\ j$$
- Notice that the correlation persistence parameters α and β are common across *i* and *j*: the persistence of the correlation between any two assets in the portfolio is the same.
- It does not, however, imply that the level of the correlations at any time is the same across pairs of assets
- Why the restriction? Usual reason: to guarantee $\rho_{ij,t+1} \in [-1,1]$

▪ DCC models are enjoying a massive popularity because they are easy to implement in 3 steps:
  - First, all the individual variances are estimated one by one
  - Second, the returns are standardized and the unconditional correlation matrix is estimated
  - Third, the correlation persistence parameters α and β (or λ) are estimated

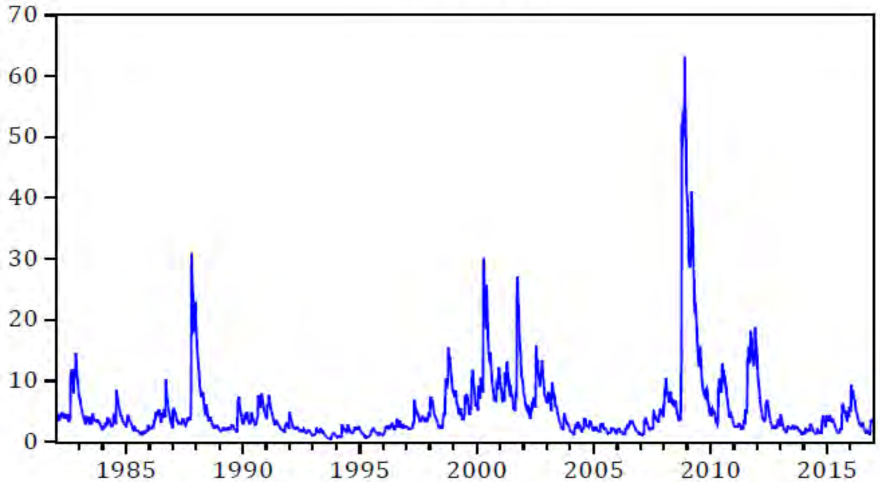
Predicted Stock-Bond Correlations from DCC RiskMetrics

$$MKT_{t+1} = \underset{(0.000)}{0.582} + \underset{(0.008)}{0.094}\, Bond_t + \varepsilon_{MKT,t}$$

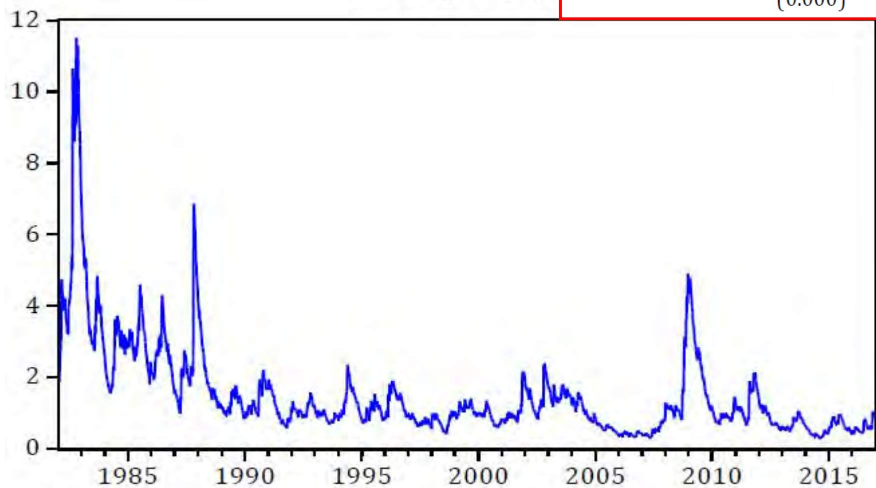
$$Bond_{t+1} = \underset{(0.169)}{0.032} + \underset{(0.000)}{0.230}\, Bond_t + \varepsilon_{Bond,t+1} \qquad [\varepsilon_{MKT,t+1} \;\; \varepsilon_{Bond,t+1}]' \text{ IID } N(\mathbf{0}, \mathbf{I}_2)$$

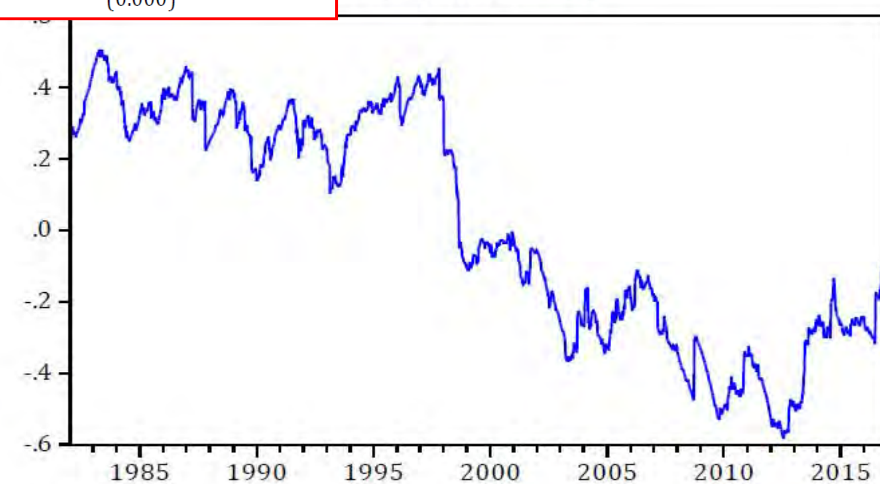
$$\sigma^2_{MKT,t+1|t} = \underset{(0.000)}{0.112}\, \varepsilon^2_{MKT,t} + \underset{(0.000)}{0.888}\, \sigma^2_{MKT,t|t-1} \cdot$$

$$\sigma^2_{Bond,t+1|t} = \underset{(0.000)}{0.066}\, \varepsilon^2_{Bond,t} + \underset{(0.000)}{0.934}\, \sigma^2_{Bond,t|t-1},$$



$$q_{MKT-Bond,t+1} = \underset{(0.000)}{0.014}\, \hat{z}^{RiskM}_{MKT,t}\, \hat{z}^{RiskM}_{Bond,t} + \underset{(0.000)}{0.986}\, q_{MKT-Bond,t}$$

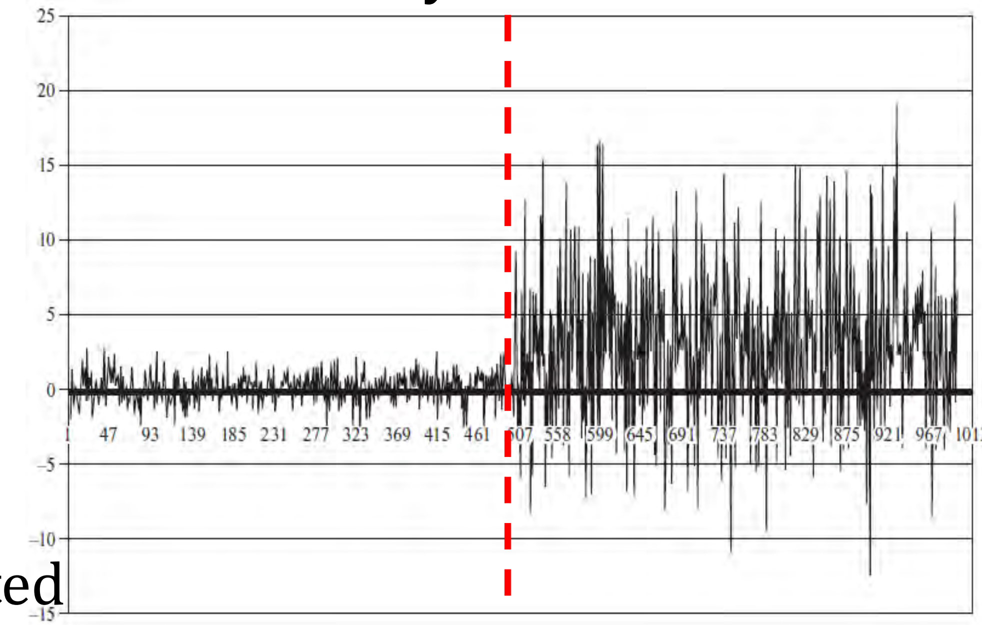




Regime Switching Models

- Financial time series are typically subject to structural instability, in the form of either breaks or regimes

- Many financial and economic time series undergo episodes in which the behavior of the series changes dramatically
  - The behavior of a series could change over time in terms of its mean value, its volatility, or its persistence
- The behavior may change once and for all, usually known as a structural break
- Or it may change for a period of time before reverting back to its original behavior or switching to yet another style of behavior; this is a regime shift or regime switch
  - Substantial changes in the properties of a series are attributed to large-scale events, such as wars, financial panics, changes in government policy (e.g., the introduction of an inflation target), etc.

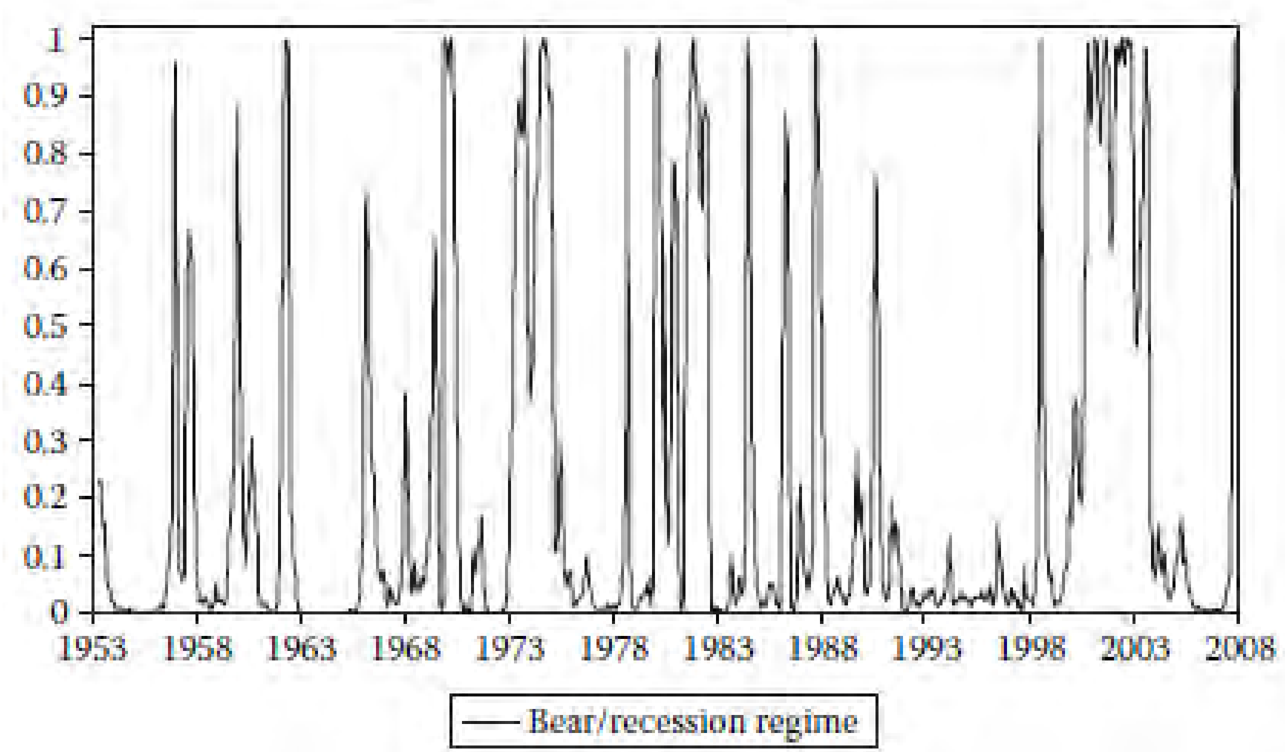


Threshold, Markov, and Dummy Switching Models

- Three classes of models
  1. Deterministic dummy multiple regression/VARMA models
  2. Threshold VAR models
  3. Markov switching VAR models
- In the first case, switches are deterministic and pre-determined and therefore not useful to forecast
- In the other two cases, regime switches are stochastic and endogenously determined from the data
- Markov switching models (MSMs) are the most popular class of non-linear models that can be found in finance
- Under a MSM there are k regimes: y_t switches regime according to some (possibly unobserved) variable, S_t, that takes integer values
  - If $S_t = 1$, the process is in regime 1 at time t, and if $S_t = 2$, in regime 2
  - In credit risk management applications, y_t is likely to be a vector of asset returns

Markov Switching Models

- In a MS model, the process followed by y_t switches over time according to one of k values taken by a discrete variable S_t
- Model also appealing to capture time-varying covariances and correlations, either in combo with GARCH or alone
- Typical output consists of time-varying parameters plus a classification, either in real time (filtered) or full sample (smoothed) of the states the different observations came from



Markov Switching Models

- In MSMs, the state variable follows a *q*th order Markov process and is often assumed to be unobservable

- Movements of state btw. regimes are governed by a Markov process such that: $P[a < y_t \leq b \mid y_1, y_2, \ldots, y_{t-1}] = P[a < y_t \leq b \mid y_{t-1}]$
  - The probability distribution of the state at t depends only on the state at t −1 and not on the states that were passed through at t −2, t −3, . . .
  - Markov processes are not path-dependent
  - The model's strength lies in its flexibility, being capable of capturing changes in the variance btw. states, as well as changes in the mean
- In the most typical implementation, the unobserved state variable, z_t, follows a first-order Markov process with transition probs.:
- p_{ij} = probability of being in regime j, give that the system was in regime i during the previous period

$$\mathrm{prob}[z_t = 1 | z_{t-1} = 1] = p_{11}$$
$$\mathrm{prob}[z_t = 2 | z_{t-1} = 1] = 1 - p_{11}$$
$$\mathrm{prob}[z_t = 2 | z_{t-1} = 2] = p_{22}$$
$$\mathrm{prob}[z_t = 1 | z_{t-1} = 2] = 1 - p_{22}$$

Multivariate Markov Switching Models

- The methods described above can be easily extended when $\mathbf{R}_t$ is a vector that collects returns on N assets/firms/sub-portfolios
- The general MSVAR(m, p) model becomes:

$$\mathbf{R}_{t+1} = \boldsymbol{\mu}_{St+1} + \boldsymbol{\Phi}_{1,St+1}\mathbf{R}_t + \boldsymbol{\Phi}_{2,St+1}\mathbf{R}_{t-1} + \ldots + \boldsymbol{\Phi}_{p,St+1}\mathbf{R}_{t-p} + \boldsymbol{\epsilon}_{t+1},$$

$$\boldsymbol{\epsilon}_{t+1} \sim N(0, \boldsymbol{\Omega}_{St+1}), \quad S_{t+1} = 1, 2, \ldots, ,$$

- In principle, the vector of intercepts $\boldsymbol{\mu}$, the p VAR matrices $\mathbf{A}_1$, $\mathbf{A}_2$, …, $\mathbf{A}_p$ and the covariance matrix $\boldsymbol{\Omega}$ all become regime-dependent
- The portion $\mathbf{R}_{t+1} = \boldsymbol{\mu}_{St+1} + \ldots$ "re-bases" the process when regime shifts occur; the portion $\mathbf{R}_{t+1} = \ldots + \boldsymbol{\Phi}_{1,St+1}\mathbf{R}_t + \boldsymbol{\Phi}_{2,St+1}\mathbf{R}_{t-1} + \ldots + \boldsymbol{\Phi}_{p,St+1}\mathbf{R}_{t-p}$ capture shifts in cross- and own-serial correlations of returns
- Of course, in our case we care chiefly but not only for modelling and predicting regime shifts in the covariance matrix, $\boldsymbol{\Omega}$
  - These models are particularly useful when extended to capture the dynamics not of $\mathbf{R}_t$ only, but of some variable $\mathbf{y}_{t+1} \equiv [\mathbf{R}_{t+1}\ \mathbf{z}_{t+1}]'$ where $\mathbf{z}_{t+1}$ is a vector that collects predictors of subsequent returns

MSVAR Models and Contagion Dynamics

- Guidolin and Ono (2006) model US stock and bond returns as predicted by past inflation, interest rates, industrial production, etc.
- This is a natural conduit to perform stress tests reflecting aggregate macroeconomic conditions

- MSVAR models are particularly suitable to model and study contagion dynamics
  - It answers the question of whether it is possible to use performance in any market/ asset to forecast what will happen in other markets
  - E.g., do performance today in the European financial sector drive the performance of industrial firm the week after?
- In MSVARH(m, p) models, 3 types of contagion effects:

1 Simultaneous, through the off-diagonal elements of $\mathbf{\Omega}_{t+1}$ that capture the dynamics across regimes of correlations

2 Dynamic and linear, through the VAR components

3 Dynamic and nonlinear, through the fact that the regime variable that drives the process of all variables in $\mathbf{y}_{t+1}$ is common to all variables

MSVAR Models and Contagion Dynamics

- Let's examine this issue with a N = 2 example in which $\mathbf{R}_t \equiv [R_t^{US}\ R_t^{Italy}]'$: how can Italian traders be affected by US ones?

$$\begin{bmatrix} R_{t+1}^{US} \\ R_{t+1}^{Italy} \end{bmatrix} = \begin{bmatrix} \mu_{S_{t+1}}^{US} \\ \mu_{S_{t+1}}^{Italy} \end{bmatrix} + \begin{bmatrix} a_{S_{t+1}}^{US,US} & a_{S_{t+1}}^{US,IT} \\ a_{S_{t+1}}^{IT,US} & a_{S_{t+1}}^{IT,IT} \end{bmatrix} \begin{bmatrix} R_t^{US} \\ R_t^{Italy} \end{bmatrix} +$$

$$+ \begin{bmatrix} \sigma_{S_{t+1}}^{US,US} & 0 \\ \rho_{S_{t+1}}^{US,IT} \sigma_{S_{t+1}}^{IT,IT} & \sqrt{1-(\rho_{S_{t+1}}^{US,IT})^2}\sigma_{S_{t+1}}^{IT,IT} \end{bmatrix} \begin{bmatrix} \epsilon_{t+1}^{US} \\ \epsilon_{t+1}^{Italy} \end{bmatrix}$$

$$= \begin{bmatrix} \mu_{S_{t+1}}^{US} + a_{S_{t+1}}^{US,US} R_t^{US} + a_{S_{t+1}}^{US,IT} R_t^{Italy} + \sigma_{S_{t+1}}^{US,US} \epsilon_{t+1}^{US} \\ \mu_{S_{t+1}}^{Italy} + a_{S_{t+1}}^{IT,US} R_t^{US} + a_{S_{t+1}}^{IT,IT} R_t^{Italy} + \rho_{S_{t+1}}^{US,IT} \sigma_{S_{t+1}}^{IT,IT} \epsilon_{t+1}^{US} + \sqrt{1-(\rho_{S_{t+1}}^{US,Italy})^2}\sigma_{S_{t+1}}^{IT,IT} \epsilon_{t+1}^{IT} \end{bmatrix}$$

- Italian returns are affected by contemporaneous US shocks with coefficient $\rho^{US,IT}_{St+1}$ (this is a regime switching correlation)
- Italian returns are linearly affected by past US returns with coefficient $a^{IT,SU}_{St+1}$
- Italian returns switch (also μ^{IT}_{St+1}, $a^{IT,IT}_{St+1}$, and $\sqrt{1-(\rho_{S_{t+1}}^{US,Italy})^2}\sigma_{S_{t+1}}^{IT,IT}$) according to a state variable S_t which is the same that drives switches in US returns, which may cause nonlinear contagion

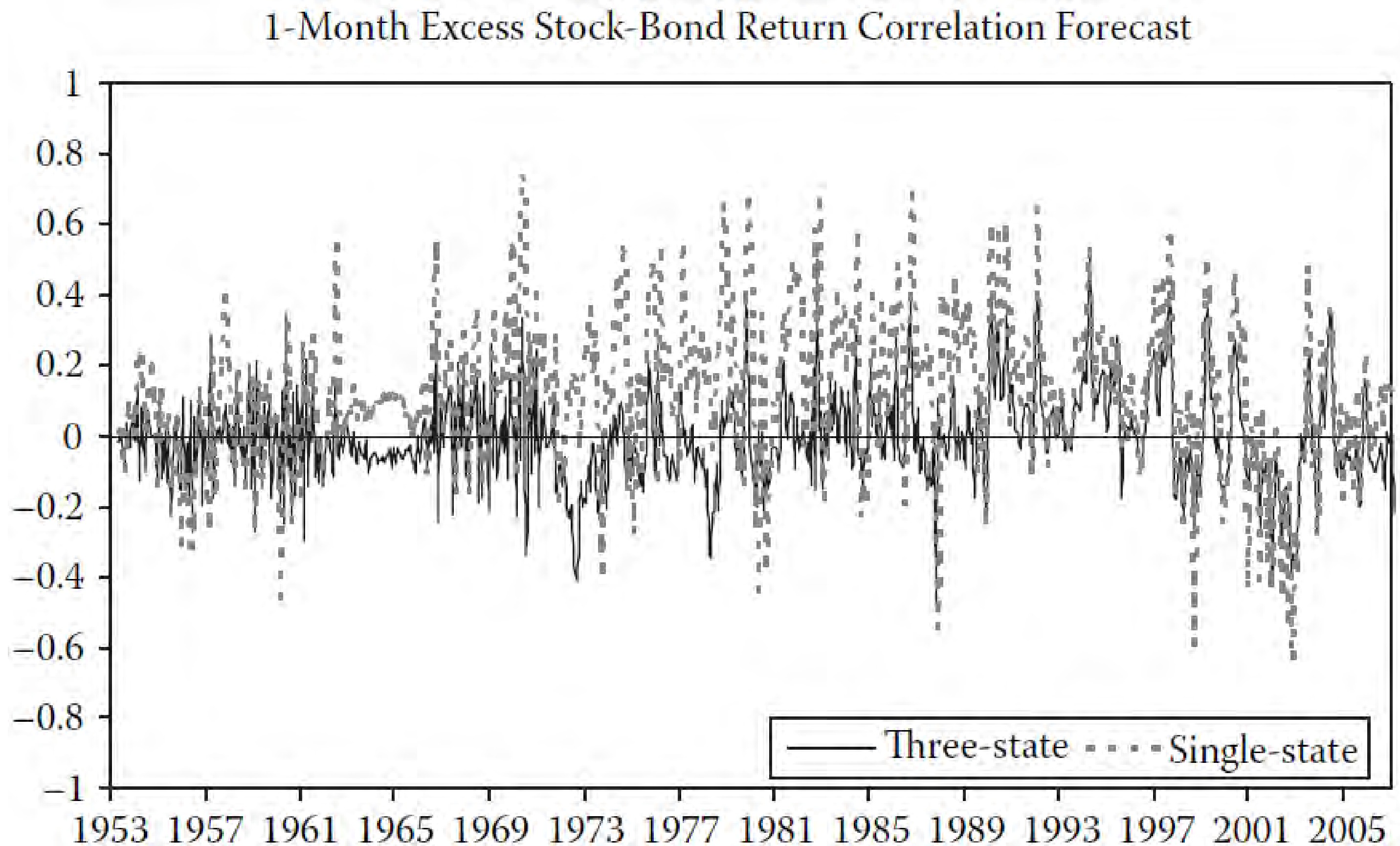
Markov Switching ARCH

- One last form of contagion has been explored: suppose that in our example R^{US}_t is driven by the Markov state S_t^{US} and R_t^{Italy} by S_t^{Italy}
- A form of interesting and testable contagion is when $S_t^{Italy} = S_{t-1}^{US}$ i.e., the market state in Italy today is driven by the market state in the US as of last period
- Previous examples stress an implied capability of MS models: to capture and forecast time-varying variances and correlations, similarly to ARCH and DCC models
- Although at some (relatively low) frequencies, MS directly competes with GARCH, at high (daily, weekly) frequencies MS, ARCH, DCC, and t-student variants are compatible
- For instance, a MS VECH GARCH model is:

$$\mathbf{r}_t = \boldsymbol{\mu}_{S_t} + \sum_{j=1}^{p} \mathbf{B}_{j,S_t}\mathbf{r}_{t-j} + \mathbf{u}_t \quad \mathbf{u}_t \sim F(\mathbf{0}, \mathbf{H}_{t,S_t}; \mathbf{v}_{S_t}),\ E[\mathbf{u}_t] = \mathbf{0},\ Var[\mathbf{u}_t] = \mathbf{H}_{t,S_t}$$

$$\mathbf{H}_{t,S_t} = \mathbf{A}_{0,S_t}\mathbf{A}'_{0,S_t} + \sum_{j=1}^{q_1}\left(\mathbf{A}_{j,S_t}\mathbf{A}'_{j,S_t}\right)\mathbf{u}_{t-j}\mathbf{u}'_{t-j} + \sum_{j=1}^{q_2}\mathfrak{I}_{t-j}\odot(\boldsymbol{\Upsilon}_{j,S_t}\boldsymbol{\Upsilon}'_{j,S_t})\mathbf{u}_{t-j}\mathbf{u}'_{t-j},$$

Markov Switching ARCH



	Three-State t-ARCH(3)	Single-State t-ARCH(3)	IID Homeskedastic
	(Pseudo Out-of-) Sample 1983:12 – 2007:11 (Recursive Estimates Only)		
Root mean squared forecast error	**10.922**	11.292	11.204
Bias	0.699	**−0.133**	−1.171
Standard deviation of forecast errors	**10.899**	11.291	11.143
Mean absolute error	5.385	5.522	**5.310**

Markov Switching ARCH Models: Pros and Cons

⊕ The full flexibility of non-linear models, degenerating in semiparametric fitting power when $m \to \infty$

⊕ Strong economic intuition in terms of business cycles, equity and bond market "phases" (bull and bear), and stress scenarios

- A stress scenario may be simply represented as an abrupt transition to a "bad regime"
- One can "load up" on the scenario by simply making it very persistent

⊕ Possible to connect regimes and their duration to exogenous economic factors

⊕ Nests as special cases standard VAR (linear) and ARCH models

⊖ They imply a large number of parameters: even in the parsimonious DCC GARCH case, these will be at least $m(N+N^2p+3N+3)$

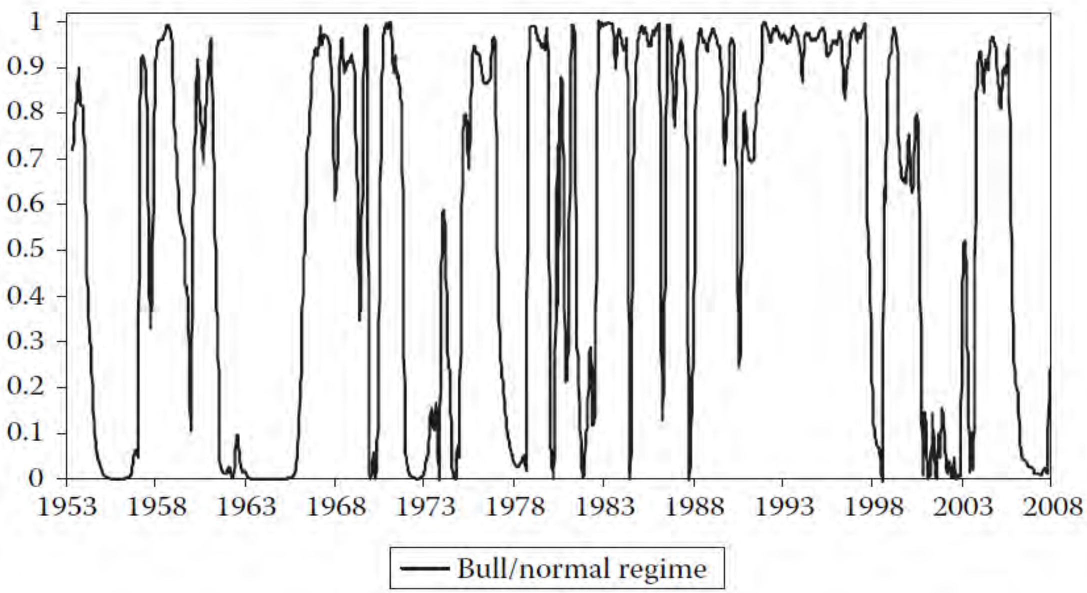
- E.g., with 100 assets, 3 regimes, p = 0, a GARCH(1,1), these are 1,209!

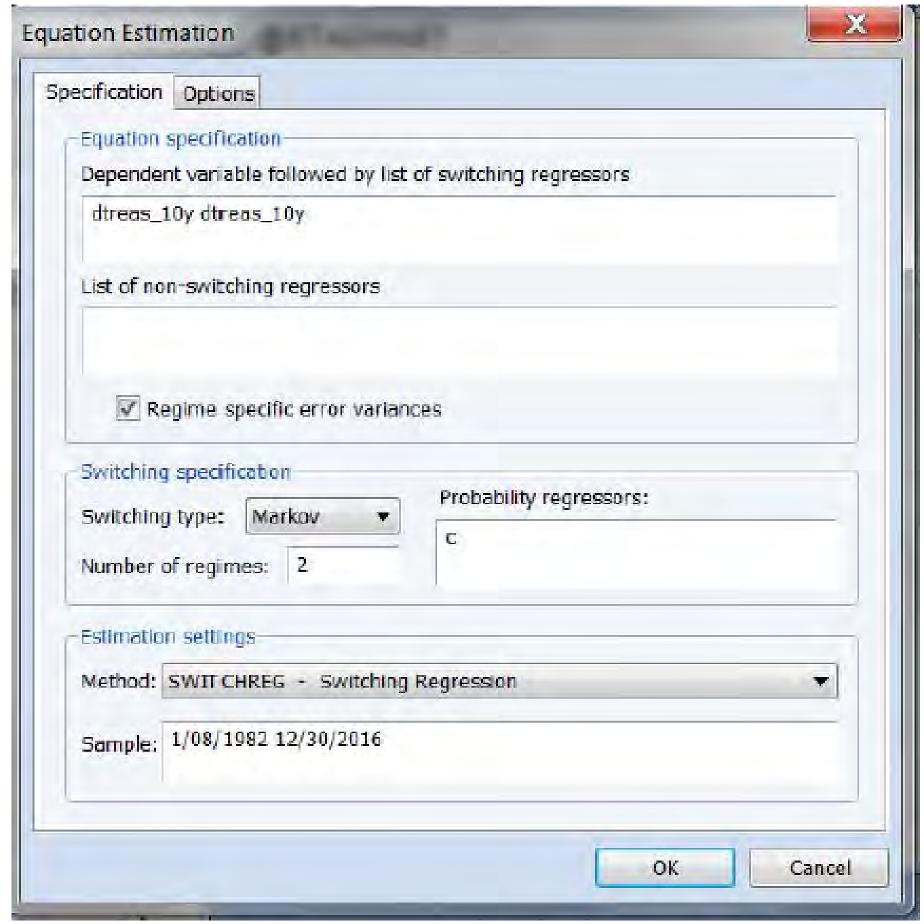
⊖ Even in the age of EM-MLE and fast computing, estimation may be slow and initial conditions will matter

⊖ Actual forecasting performance remain under scrutiny

Markov Switching ARCH Models: Pros and Cons

⊖ The regime remains latent and as such at best one can infer (filter or smooth) probabilities concerning its nature, with all the uncertainty of the case





⊖ Their availability in "software packaged form" is not uniform and especially the multivariate case will need further development

Multi-Factor Heteroskedasticity and Stochastic Volatility

Under stochastic volatility (SV) volatility is random with its own shocks, not necessarily related to functions of past asset returns

- The key feature of (G)ARCH is that it explicitly models the conditional variance only as a function of past returns
  - The one-step-ahead prediction approach to volatility modeling is convenient as it immediately delivers the likelihood function as the product of one-step-ahead predictive densities
- However, in many applications the model is not flexible enough
- Moreover, GARCH does not (always) have a straigthforward continuous time limit, which is often required in derivative pricing
- Under SV approach, volatility (and covariance) is a random variable characterized by shocks that do not simply reflect past returns
- Therefore the model is a multi- (at least, two-) factor one
  - The predictive distribution of returns is specified indirectly, via the structure of the model, rather than directly
  - Often, the predictive distribution can only be evaluated numerically

Multi-Factor Heteroskedasticity and Stochastic Volatility

- The distinguishing feature of SV specifications is that volatility, being inherently unobservable and subject to its own random shocks, is not measurable with respect to observable information
- An estimate of current volatility state must be filtered from a noisy environment
- SV models provide the basis for realistic, state-of the-art modeling of option process
  - Hull and White (1987) assumed that volatility risk was unrewarded and showed that SV models could produce smiles and skews in option prices, which are frequently observed in market data

Brownian motion

$$d\sigma^2(t) = v(\sigma^2(t)) \cdot dt + \kappa(\sigma^2(t)) \cdot dW_{2,t}$$

  - Of course people who have studied derivative pricing before 1987 may not know that!
  - See http://janroman.dhis.org/finance/Statistics/Stochastic%20Implied%20Trees.pdf, it's a GS research note and from 1997
- How does one take care of the unobservable nature of volatility in SV models? Using the Kalman filter or its extensions

The Kalman Filter

Kalman filter is a recursively updated algorithm to filter, forecast and estimate the process of latent variables in dynamic models

- Consider the simple univariate model:

$$R_t = \gamma_1 + \gamma_2 h_t + \varepsilon_t \quad (\star)$$

$$h_t = \gamma_3 + \gamma_4 h_{t-1} + \eta_t \quad (\star\star)$$

$$\varepsilon_t \text{ IID } N(0, v_1) \text{ and } \eta_t \text{ IID } N(0, v_2)$$

$$\mathrm{Cov}(v_1, v_2) = 0$$

- h_t is unobservable and we can at simulate it from its stochastic process
- ($\star$) is the measurement or observation equation whereas ($\star\star$) is the state or transition equation
- ($\star$)+ ($\star\star$) = state-space representation of the model
  - Assume that the initial values of h_0 and w_0 are given and fixed
- The Kalman filter exploits normality of the shocks and is organized around 3 steps: filtering, forecasting, updating, and estimation
- In the forecasting step, we predict $h_{t|t-1}$ and $w_{t|t-1}$; under normality of the shocks these are optimal forecasts in a MSE sense:

$$h_{t|t-1} = \gamma_3 + \gamma_4 h_{t-1}$$

$$\underbrace{w_{t|t-1}}_{\text{Variance of } h_t} = \gamma_4^2 w_{t-1} + v_2$$

- The second step is updating: at time t, we have a new observation on the variable, R_t; we can compute the prediction error u_t

The Kalman Filter

$$u_t = R_t - \gamma_1 - \gamma_2 h_{t|t-1} = (\gamma_1 + \gamma_2 h_t + \varepsilon_t) - \gamma_1 - \gamma_2 h_{t|t-1} = \gamma_2(h_t - h_{t|t-1}) + \varepsilon_t$$

- The variance of u_t (prediction error var), ψ_t, is: $\psi_t = \gamma_2^2 w_{t|t-1} + v_1$
- The update of h_t and its variance w_t as follows:

$$h_t = h_{t|t-1} + \frac{\gamma_2 w_{t|t-1}}{\psi_t} u_t \qquad w_t = w_{t|t-1} + \frac{(\gamma_2 w_{t|t-1})^2}{\psi_t}$$

  - These equations are conditionally unbiased and efficient estimators of the latent variables and can be used for subsequent forecasting step
- To estimate the parameters γ_1, γ_2, γ_3, γ_4, v_1, and v_2, we use the maximum likelihood method: the (partial, conditioning on given initial values h_0 and w_0) log-likelihood function is:

$$\ell(\gamma_1, \gamma_2, \gamma_3, \gamma_4, v_1, v_2) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T} \ln \psi_t - \frac{1}{2}\sum_{t=1}^{T} \frac{u_t^2}{\psi_t}$$

$$\propto -\frac{1}{2}\sum_{t=1}^{T} \ln(\gamma_2^2\gamma_4^2 w_{t-1} + \gamma_2^2 v_2 + v_1) - \frac{1}{2}\sum_{t=1}^{T} \frac{(R_t - \gamma_1 - \gamma_2\gamma_3 - \gamma_2\gamma_4 h_{t-1})^2}{\gamma_2^2\gamma_4^2 w_{t-1} + \gamma_2^2 v_2 + v_1}$$

- The log-likelihood can be recursively evaluated by iterating the forecasting and updating steps between t = 1 and T, also yielding (filtered) time series of latent variables, chiefly $\{h_t\}$

The Kalman Filter

- These methods can be extended to linear state-space representations of the dynamics of the Nx1 vector $\mathbf{y}_t$:

$$\boldsymbol{y}_t = \boldsymbol{\gamma}_t + \boldsymbol{\Gamma}_t \boldsymbol{h}_t + \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{h}_t = \boldsymbol{\delta}_t + \boldsymbol{\Delta}_t \boldsymbol{h}_{t-1} + \boldsymbol{\eta}_t$$

$$\boldsymbol{\Omega}_t = Var\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_t & \boldsymbol{G}_t \\ \boldsymbol{G'}_t & \boldsymbol{Q}_t \end{bmatrix}$$

  - It is often assumed that such a VAR is first-order, stationary, and ergodic by imposing standard conditions
  - However, the Kalman filter is general enough to deal with the case of a nonstationary VAR model for the hidden state, for instance, when $\boldsymbol{h}_t = \boldsymbol{\delta}_t + \boldsymbol{h}_{t-1} + \boldsymbol{\eta}_t$
  - The disturbances are serially independent, with contemporaneous covariance structure $\boldsymbol{\Omega}_t$
- Asset pricing theory contends that asset prices reflect the discounted value of future expected cash flows $\Rightarrow$ all news relevant for either discount rates or cash flows should shift prices
- Since news appear almost continuously, if the process is stationary sense, it will yield a relationship between news arrivals, market activity, and return volatility that in limit, from CLT, becomes:

$$R_t | S_t \sim N(\mu S_t, \sigma^2 S_t)$$

From the Normal Mixture Model to SV

The continuous time limit of a normal mixture model in which the positive intensity follows a subordinated process is a SV model

- S_t is a positive intensity process reflecting the rate of news arrivals

- This is a normal mixture model, where the S_t process governs or "mixes" the scale of the distribution across periods
  - Clark (1973) used trading volume as a proxy for the intensity variable, motivated by high correlation between return volatility and volume
  - If S_t is constant, the mixture model degenerates to a simple Gaussian IID process for returns which is at odds with the empirical evidence
- Therefore, S_t is typically assumed to follow a separate (subordinated) stochastic process with random innovations
  - In each period returns are subject to two separate shocks, namely the idiosyncratic error associated with the (normal) return distribution, and also a shock to the variance process, S_t
- This endows the return process with genuine SV, reflecting the random intensity of news arrivals

From the Normal Mixture Model to SV

- Typically assumed that only returns, transactions, and quotes are observable, but not the actual value of S_t itself $\Rightarrow$ the variance/ diffusion process σ^2 in $R_t|S_t \sim N(\mu S_t, \sigma^2 S_t)$ cannot be identified
- Because it cannot be identified, it is typical to set $\sigma^2 = 1$

- The structural randomness and unobserved nature of the news arrival process makes the true mean and variance series latent
- The variation in the information flow induces a fat-tailed unconditional distribution for returns: days with many news display more rapid fluctuations and trading than days with a low news count
- If the S_t process is positively serially correlated, then shocks to the conditional mean and variance for returns will be persistent
- For concreteness, we follow Harvey and Shephard (1996) and consider a simple SDE for the log of the price (P(t)) of an asset:

Standard normal

$$d \ln P(t) = \frac{dP(t)}{P(t)} = \mu \cdot dt + \sigma(t) dz_{1,t}$$

Cont. Compounded return

- A simple discretization delivers: $R_t \equiv \Delta \ln P(t) = \mu + \sigma_t dz_{1,t}$

A Constant Drift AR(1) Stochastic Volatility Model

A log-normal, two-factor SVM implies zero skewness, arbitrarily large excess kurtosis, and non-Gaussian (mixture) asset returns

- The process followed by the log-conditional variance is AR(1):

$$h_t \equiv \ln \sigma_t^2 = \gamma_1 + \gamma_2 h_{t-1} + dz_{2,t}, \quad dz_{2,t} \text{ IID } N(0, \sigma_{z_2}^2)$$

  for (0 <) $|\gamma_2| < 1 \Rightarrow \sigma_t = \exp(0.5h_t)$, long-run log-variance $\gamma_1/(1 - \gamma_2)$
  - Andersen et al. (2003) have shown that empirically the log-variance process can be well approximated by a normal distribution
- Clear why we may refer to this SVM as a two-factor model: there are two shocks, two random drivers of asset returns:
  1. $dz_{1,t}$ which enters directly in the model, in a conditionally linear fashion
  2. $dz_{2,t}$ which enters instead through the (transformed, nonlinear) process for the multiplicative factor σ_t
- Taylor (2008) proves that if $|\gamma_2| < 1$, the log-run, unconditional variance of continuously compounded returns is given by:

$$\bar{\sigma}^2 = \exp\left[\frac{1}{2}\frac{\sigma_{z_2}^2}{1-\gamma_2^2}\right]$$

A Constant Drift AR(1) Stochastic Volatility Model

- While the model is symmetric and returns display zero skewness, the unconditional kurtosis is: $Kurt(R_t) = Kurt(dz_{1t})\exp\left[\frac{\sigma_{z_2}^2}{1-\gamma_2^2}\right]$
  - E.g., in the normal case, $Kurt(R_t) = 3\exp\left[\sigma_{z_2}^2/(1-\gamma_2^2)\right]$
- Although returns are uncorrelated, they are far from independent because the dynamics of the series appear in the squared returns:

$$Corr(R_t^2, R_{t-\tau}^2) = \frac{\exp[(\sigma_{z_2}^2/(1-\gamma_2^2))\gamma_2^\tau] - 1}{Kurt(dz_{1t})\exp[(\sigma_{z_2}^2/1-\gamma_2^2)] - 1} \quad \tau = 1, 2, \ldots$$

  - Even though the expression shows that this is not correct, γ_2 is considered the driver of persistence of the autocorrelations of squares
- Differently from GARCH, even when the shocks are assumed to be Gaussian, the distribution of time *t* returns conditional on past observations up to time *t*+1 is not normal
- The SVM for log-variance provides a mixture of normal distributions, i.e., a weighted sum of normal densities, with random weights
- Linearizing the first SDE around μ = 0 (an empirically plausible assumption for high-frequency data), we have $R_t = \exp(0.5h_t)dz_{1,t}$

A Constant Drift AR(1) Stochastic Volatility Model

- This implies: $\ln R_t^2 = h_t + \ln(dz_{1,t})^2$
- The distribution of $(\ln dz_{1,t})^2$ is therefore a logarithmic χ^2 with 1 degree of freedom and therefore with an expectation of -1.27 and

$$\begin{aligned}\ln R_t^2 &= \gamma_0 + E[\ln(dz_{1,t})^2] + h_t + \underbrace{(\ln(dz_{1,t})^2 - E[\ln(dz_{1,t})^2])}_{\upsilon_t} \\ &= (\gamma_0 - 1.27) + h_t + \upsilon_t,\end{aligned}$$

  - γ_0 is a small sample correction that can be dropped for large estimation samples, and υ_t is a zero mean χ^2 shock with 1 degree of freedom such that $Var[\upsilon_t] = 0.5\pi^2 \simeq 4.93$
- The equations that we want to estimate is then the following:

$$\begin{aligned}\ln R_t^2 &= (\gamma_0 - 1.27) + h_t + \upsilon_t \quad \upsilon_t \text{ IID } \chi_1^2 && \boxed{\text{(Measurement equation)}} \\ h_t &= \gamma_1 + \gamma_2 h_{t-1} + dz_{2,t} \quad dz_{2,t} \text{ IID } N(0, \sigma_{z_2}^2) && \boxed{\text{(State equation)}}\end{aligned}$$

- Assume that υ_t and $dz_{2,t}$ are uncorrelated, that is, shocks to the variance of returns carry no information on shocks to the level
- The two equations are now in appropriate form for the Kalman filter to be applied, subject to an approximation

A Constant Drift AR(1) Stochastic Volatility Model

A log-normal, two-factor SV model can be estimated by approximate quasi-ML using the Kalman filter

- Comparison btw. the SVM and a standard GARCH(1,1) based on US value-weighted CRSP stock returns over a 1963-2016 daily sample
  - The initial state, here represented by log-variance, is solved to be -1.92, simply exploiting the AR(1) nature of the state equation when υ_t IID χ^2 is replaced by a υ_t IID N(0, 4.935) approximation
  - QML estimation in EViews yields (p-values are computed using the estimated Hessian matrix and are in parenthesis):

$$\ln R_t^2 = (\underset{(0.998)}{-0.005} - 1.27) + h_t + \upsilon_t \quad \upsilon_t \text{ IID } N(0, 4.935)$$

$$h_t = \underset{(0.716)}{-0.004} + \underset{(0.000)}{0.993} h_{t-1} + dz_{2,t} \quad dz_{2,t} \text{ IID } N(0, \exp(\underset{(0.000)}{-4.629}))$$

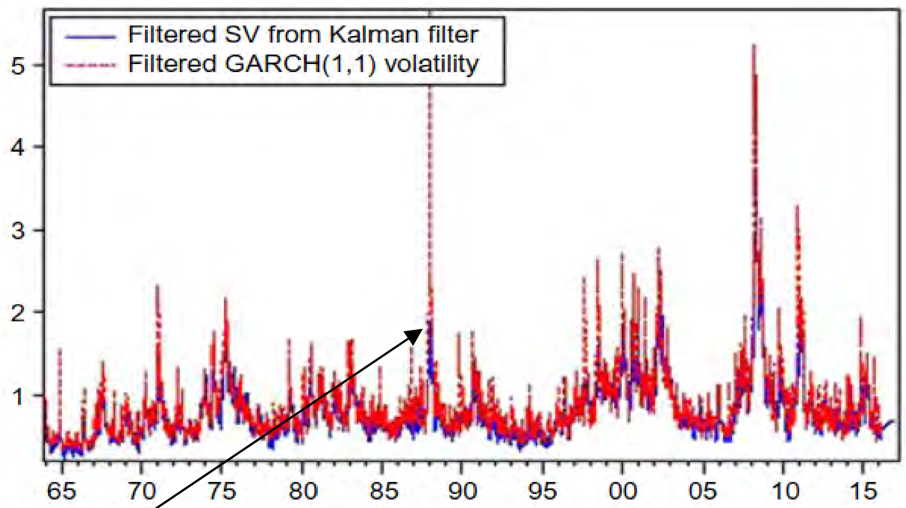
  - The model is covariance stationary: a Wald test of the null hypothesis $\gamma_2 = 1$ is rejected with a p-value of 0.001
  - On the same data, MLE of a Gaussian GARCH(1,1) model gives:

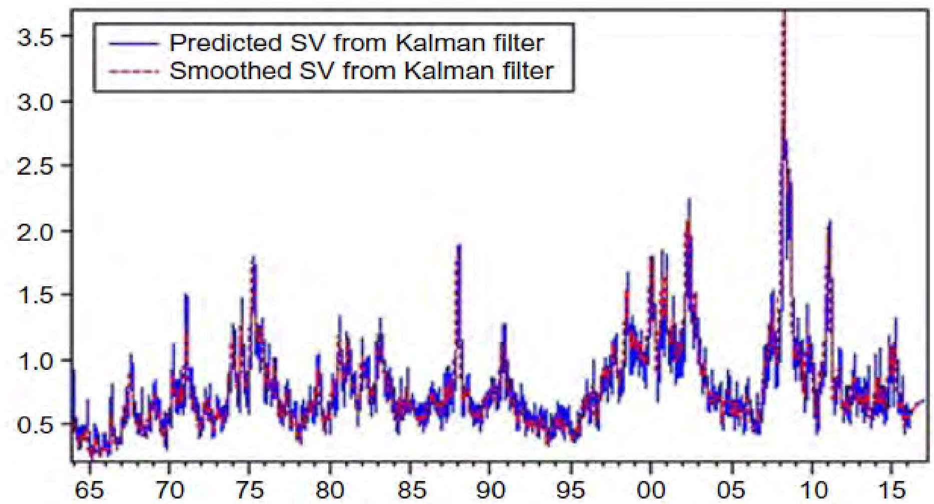
$$R_{t+1} = \underset{(0.000)}{0.065} + \sigma_{t+1|t} dz_{1,t+1} \quad dz_{1,t+1} \text{ IID } N(0,1)$$

$$\sigma^2_{t+1|t} = \underset{(0.000)}{0.008} + \underset{(0.000)}{0.090} \sigma^2_{t|t-1} (dz_{1,t})^2 + \underset{(0.000)}{0.904} \sigma^2_{t|t-1}.$$

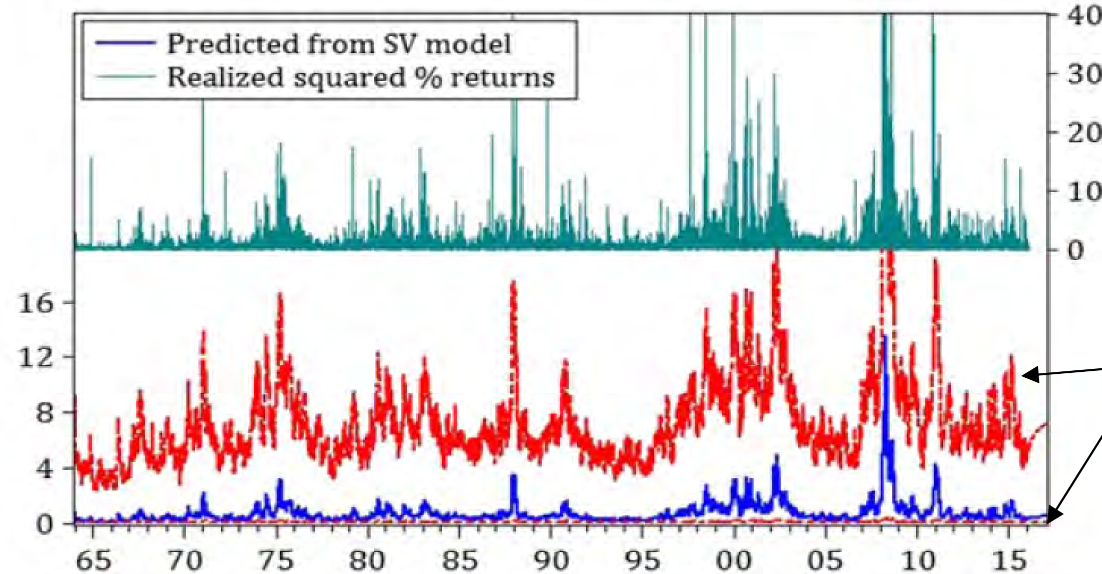
  - The implied GARCH persistence is almost identical, 0.994

A Constant Drift AR(1) Stochastic Volatility Model





SVM does not pick up Oct 1987 as a volatility spike



95% confidence interval

- Ruiz (1994) shows that the QMLE is consistent and asymptotically normal, but also considerably inefficient

AR(1) Stochastic Volatility Model with Leverage

SV models may be extended to incorporate leverage effects

- The reason is that the approximation may occasionally be poor because, when returns are very close to 0, the log-squared transformation yields large negative numbers

- Possible to modify the SVM to account for the plausible fact that when a large shock to $dz_{1,t}$ generates large positive or negative returns (hence a large squared shock υ_t), it becomes more likely that also shocks driving the time-varying variance of $dz_{2,t}$ are large
- We add a new parameter ρ to capture just such leverage effect:

$$\ln R_t^2 = (\gamma_0 - 1.27) + h_t + \upsilon_t, \quad \upsilon_t \text{ IID } \chi_1^2$$
$$h_t = \gamma_1 + \gamma_2 h_{t-1} + dz_{2,t}, \quad dz_{2,t} \text{ IID } N(0, \sigma_{z_2}^2), \quad \rho = Cov[\upsilon_t, dz_{2,t}].$$

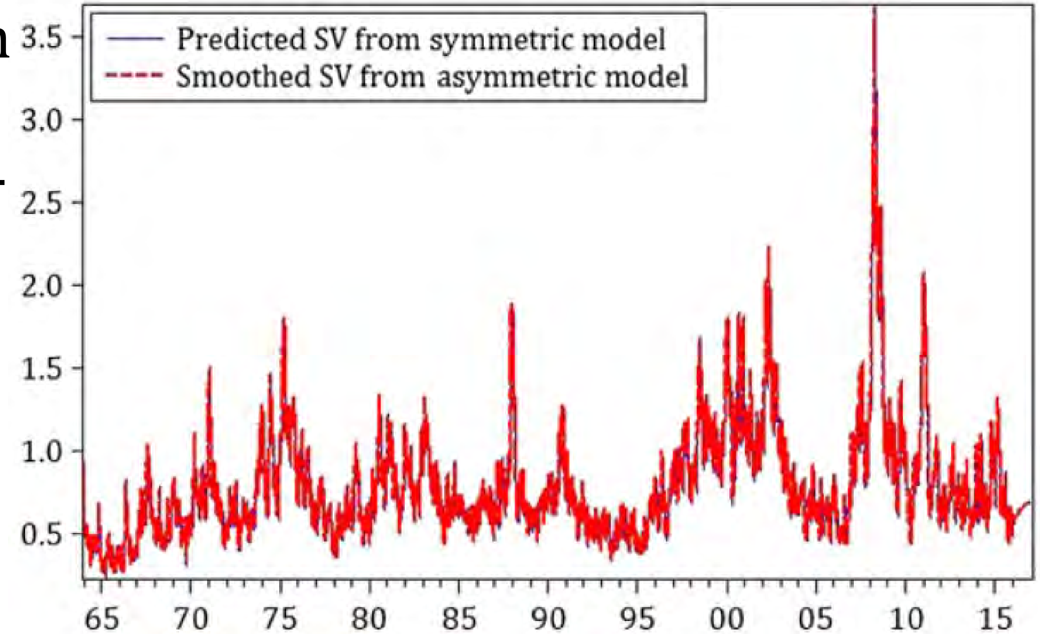
- Estimation on daily S&P 500 returns yields:

$$\ln R_t^2 = (\underset{(0.998)}{-0.022} - 1.27) + h_t + \upsilon_t, \quad \upsilon_t \text{ IID } N(0, 4.935)$$
$$h_t = \underset{(0.944)}{-0.004} + \underset{(0.000)}{0.994}\, h_{t-1} + dz_{2,t}, \quad dz_{2,t} \text{ IID } N(0, \exp(\underset{(0.000)}{-4.562})), \quad \widehat{\rho} = \underset{(0.025)}{0.143}$$

  - This model yields a maximized log-lik of -30,535.9, which exceeds by 2.6 the maximized log-lik of -30,538.5 when $\rho = 0$

AR(1) Stochastic Volatility Model with Leverage

- The Hannan-Quinn information criterion (modestly) values the additional parameter represen-ted by ρ, because it is 4.50117 in the case of the asymmetric, 5-parameter SVM and 4.50122 in the case of the symmetric 4-parameter SVM



- A comparison btw the 1-day-predicted latent volatility from the SVMs reveals modest differences

▪ A SV model with $\rho \neq 0$ allows for non-zero skewness in returns

▪ One drawback of the Kalman filter approach is that the finite sample properties can be poor because the error terms are highly non-Gaussian, as shown by Andersen et al. (1999)
  - Lo (1988) warns that the common approach of estimating parameters of an Ito process by applying ML to a discretization of the SDE yields inconsistent estimators
  - He characterizes the likelihood function as a solution to a PDE

Extending Stochastic Volatility Models

Empirical fit of SV may be enriched with jumps and mean-reversion

- Empirical research in the early 2000s has shown that more complicated volatility dynamics compared to the standard SVM may be required, such as:

$$dP(t) = \mu P(t)\, dt + \sigma(t) P(t)\, dW_{1,t}$$
$$d \ln \sigma^2(t) = \beta(v - \ln \sigma^2(t))\, dt + \kappa dW_{2,t}$$

- Or else

$$dP(t) = (\mu - \lambda\bar{\xi}) P(t)\, dt + \sigma P(t)\, dW_{1,t} + \xi(t)\, dq_t$$
$$\Pr(dq_t = 1) = \lambda \cdot dt, \qquad \xi(t) \sim N(\bar{\xi}, \sigma_\xi^2),$$

- The Brownian motions $dW_{1,t}$ and $dW_{2,t}$ are potentially correlated
- q follows a Poisson jump process uncorrelated with the Brownian motion $dW_{1,t}$ and is parameterized by a constant jump intensity λ
  - λξ compensates for the price of jump risk
  - The scaling factor ξ(t) denotes the normally distributed magnitude of the jump in the return process when a jump occurs at time t, so that jumps can be both positive and negative
- GARCH and SV models are not orthogonal: a few specific types of SVMs may be discretized to become ARCH and the continuous-time limit of a few specific GARCH models is a precise type of SVM!

From GARCH to Stochastic Volatility Models and Back

EGARCH(1,1) models are the continuous limit of log-SV models

- Nelson (1990) shows that starting from discrete GARCH(1,1) process

$$\ln P_{t+\Delta t} = \ln P_t + (\mu - 0.5\sigma_t^2)\Delta t + \sqrt{\Delta t}\sigma_t dz_{1,t+\Delta t}$$
$$\ln \sigma_{t+\Delta t}^2 = \ln \sigma_t^2 + \beta(v - \ln \sigma_t^2)\Delta t + \sqrt{\Delta t} g(dz_{1,t+\Delta t})$$

  if one picks

$$g\left(dz_{1,t+\Delta t}\right) = \rho\kappa dz_{1,t+\Delta t} + \kappa\sqrt{\frac{1-\rho^2}{1-(2/\pi)}}\left(|dz_{1,t+\Delta t}| - \sqrt{2/\pi}\right)$$

  is the limit as $\Delta t \rightarrow 0$ of the log-SV model and it is a EGARCH(1,1)
- This is why EGARCH models are so meaningful
- Alternatively,

$$\ln P_{t+\Delta t} = \mu + \ln P_t + \sqrt{\Delta t}\sigma_t dz_{1,t+\Delta t} = \mu + \ln P_t + \varepsilon_{t+\Delta t}$$
$$\sigma_{t+\Delta t}^2 = \omega\Delta t + (1 - \theta\Delta t - \alpha\sqrt{\Delta t})\sigma_t^2 + \alpha\Delta t\sigma_t^2(dz_{1,t+\Delta t})^2$$
$$= \omega\Delta t + (1 - \theta\Delta t - \alpha\sqrt{\Delta t})\sigma_t^2 + \alpha\sqrt{\Delta t}\varepsilon_{t+\Delta t}^2,$$

  a special GARCH(1,1) converges as $\Delta t \rightarrow 0$ ($dW_{1,t} \perp dW_{2,t}$) to

$$dP(t) = \mu P(t)\cdot dt + \sigma(t)P(t)\cdot dW_{1,t}$$
$$d\sigma^2(t) = (\omega - \theta\sigma^2(t))\cdot dt + \sqrt{2}\alpha\sigma^2(t)\cdot dW_{2,t}.$$

  a mean reverting SV model in levels, called Hull and White's model

From GARCH to Stochastic Volatility Models and Back

GBM and hence BS obtain from a homoskedastic GARCH(0,0) model

- Note that $\omega = \theta = \alpha = 0$ so that $\ln P_{t+\Delta t} = \mu + \ln P_t + \varepsilon_{t+\Delta t}$ with $\sigma^2_{t+\Delta t} = \sigma^2_t = \sigma^2$ has a continuous-time limit represented by the classical, Black-Scholes Geometric Brownian motion-with drift model:

$$dP(t) = \mu P(t) \cdot dt + \sigma P(t) \cdot dW_{1,t} \qquad d\sigma^2(t) = 0.$$

- This implies that when in this framework the data want to tell us that options ought be priced by Black-Scholes, they will be able to tell us, no need for us to impose it!
- Exercises that use SV models with jumps to jointly fit the time series of the asset returns and the cross-section of option prices has established that SVJ dramatically improves the fit vs. Black-Scholes)
  - SV alone has a first-order effect and low-frequency, rare jumps further enhance performance by generating fatter tails in the return distribution and reducing the pricing error for short-dated options
  - If volatility follows a pure diffusion, the implied continuous sample path may be incapable of generating a sufficiently distribution over short horizons to justify the observed prices of derivatives