# VOLATILITY MODELS AND THEIR APPLICATION TO OPTIONS PRICING AND RISK MANAGEMENT

by

**A. B. Sharapov**

B.S. Saint-Petersburg State University, 2005

M.S. Saint-Petersburg State University, 2008

M.S. University of Pittsburgh, 2008

M.S. University of Pittsburgh 2009

Submitted to the Graduate Faculty of
the Department of Statistics in partial fulfillment
of the requirements for the degree of

**Master of Science**

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH

STATISTICS DEPARTMENT

This thesis was presented

by

A. B. Sharapov

It was defended on

December 5, 2012

and approved by

David Stoffer, Professor, Department of Statistics

Robert Krafty, Assistant Professor, Department of Statistics

Sungkyu Jung, Assistant Professor,Department of Statistics

Thesis Advisor: David Stoffer, Professor, Department of Statistics

# VOLATILITY MODELS AND THEIR APPLICATION TO OPTIONS PRICING AND RISK MANAGEMENT

A. B. Sharapov, M.S.

University of Pittsburgh, 2012

We look at various volatility models and their applications. Starting from a basic linear GARCH model we proceed to more advanced linear GARCH models involving leverage effects and asymmetry. We also look at some examples of non-linear GARCH models such as TGARCH, smooth transition GARCH and NNGARCH.ML estimation technique is considered. Some applications to options pricing and risk management are presented. Next we turn our attention to discrete and continuous stochastic volatility models. Filtering techniques such as Kalman filter, particle filter are presented and estimation approaches based on filtering as well as efficient method of moments are elaborated on in details. Finally we take a look at the implied volatility surface and some ways of its estimation.

## TABLE OF CONTENTS

## LIST OF FIGURES

# 1.0 INTRODUCTION

For the last $20 - 30$ years the science of financial modeling has drastically developed. These days in order to become a financial modeler it is sometimes required to have a PhD degree from a top university and a highly numerate subject. This is due to a very high complexity of financial models. Different aspects of the behavior of financial returns, volatility or pricing should be taken into account. Nonetheless people are still unable to fully explain some features attributed to very complex derivatives and the ongoing research will continue for years to come.

In this thesis we take a look at one of the most important feature of any derivative security and any financial instrument in general - volatility. Being one of most researched subjects in finance these days it is still not fully understood and new models of volatility appear every year in abundance. To give the reader a sense of breadth of the usage of volatility forecasts let us list a few reasons why finance practitioners make such a big deal out of it.

Firstly, all sorts of derivative securities prices strongly depend on the volatility. Take a simple example of a call option of a stock. The famous Black-Scholes formula says that besides some other factors the price of the option depends on volatility. Being a very simplistic model the BS model is unable to give accurate results for options prices, but nonetheless it gives a great deal of information on what is importance what is not. Being able to predict volatility one can price options more accurately.

Secondly, risk managers look at volatility forecasts on a daily basis. The most popular risk measure VaR depends on volatility, so the ability to quantify risk in directly related to the volatility forecasting.

In this thesis we will look at two types of volatility models: the generalized autoregressive conditional heteroscedasticity models (GARCH) and stochastic volatility models (SV). We well present some theory underlying these models as well as some estimation techniques. We do not claim to cover all aspects of volatility modeling here, since the body of research on this subject is enormous.

We start our presentation with some features of financial returns called stylized facts. For the most part volatility models try to reproduce some of them and the quality of a model sometimes depends of whether a certain fact in explained or not.

## 2.0 STYLIZED FACTS AND PROPERTIES OF FINANCIAL TIME SERIES

The study of statistical properties of financial time series has revealed a wealth of interesting stylized facts [1] which seem to be common to a wide variety of markets, instrument and periods:

1. Absence of autocorrelations: autocorrelations of asset returns are often insignificant, except for very small intraday time scales ($\simeq 20$ minutes) for which microstructure effects come into play.
2. Heavy tails: the (unconditional) distribution of returns seems to display a power-law or Pareto-like tail, with a tail index which is finite, higher than two and less than five for most data sets studied. In particular this excludes stable laws with infinite variance and the normal distribution. However, the precise form of the tails is difficult to determine.
3. Gain/loss asymmetry: one observes a large drawdowns in stock prices and stock index values but not equally large upward movement.
4. Aggregational Gaussianity: as one increases the time scale over which returns are calculated, their distribution looks more and more like a normal distribution. In particular, the shape of the distribution is not the same at different time scales.
5. Intermittency: returns display, at any time scale, a high degree of variability. This is quantified by the presence of irregular bursts in time series of a wide variety of volatility estimators.
6. Volatility clustering: different measures of volatility display a positive autocorrelation over several days, which is quantifies the fact that high-volatility events tend to cluster in time.

7. Conditional heavy tails: even after correcting returns for volatility clustering, the residual time series still exhibit heavy tails. However, the tails are less heavy than in the unconditional distribution of returns.
8. Slow decay of autocorrelation in absolute returns: the autocorrelation function of absolute returns decays slowly as a function of the time lag, roughly as a power law with an exponent $\beta \in [0.2, 0.4]$. This is sometimes interpreted as a sign of long range dependence.
9. Leverage effect: most measures of volatility of an asset are negatively correlated with the returns of the asset.
10. Volume/volatility correlation: trading volume is correlated with all measures of volatility.
11. Asymmetry in time scales: coarse-grained measures of volatility predict fine-scale volatility better than the other round.

In the following chapters we attempt to present various models that have been developed for the last several decades. These models range from simple univariate GARCH(p,q) model to non-linear GARCH model, multivariate GARCH models and finally the Stochastic volatility models. We will also look at some aspect of asset pricing based on those models as well as risk management.

# 3.0 CONDITIONAL HETEROSCEDASTICITY MODELS OF VOLATILITY

## 3.1 LINEAR MODELS FOR CONDITIONAL HETEROSCEDASTICITY.

### 3.1.1 Univariate GARCH.

Autoregressive conditional heteroscedastic models were introduced by Engle (1982) and their GARCH extension is due to Bollerslev. In these models, the key concept is the conditional variance or put in other words, the variance conditional on past information. In the classical GARCH models, the conditional variance is given as a linear function of the squared past values of the series.This particular form is able to capture most of the stylized facts intrinsic to financial time series. At the same time, this model is simple enough to allow for a thorough study of the solutions. In the next section we present the general theory underlying the GARCH models closely following Francq and Zakoian [40, 4].

**3.1.1.1 General theory.** We start with a definition of GARCH processes based on the first two conditional moments.

**Definition.** A process $(\epsilon_t)$ is called a GARCH(p,q) precess if its first two conditional moments exist and satisfy:

- $E(\epsilon_t|\epsilon_u, u < t) = 0$, $t \in \mathbb{Z}$
- There exist constants $\omega, \alpha_i, i = 1, ..., q$ and $\beta_j, j = 1, ..., p$ such that

$$\sigma_t^2 = Var(\epsilon_t|\epsilon_u, u < t) = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \qquad (3.1)$$

Equation (3.1) can be written in a a more compact way as

$$\sigma_t^2 = \omega + \alpha(B)\epsilon_t^2 + \beta(B)\sigma_t^2 \tag{3.2}$$

where $B$ is the standard backshift operator and $\alpha$ and $\beta$ are polynomials of degree $q$ and $p$, respectively:

$$\alpha(B) = \sum_{i=1}^{q} \alpha_i B^i, \quad \beta(B) = \textstyle\sum_{j=1}^{p} \beta_j B^j$$

If $\beta(z) = 0$ we have

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 \tag{3.3}$$

and the precess is called an ARCH(q) precess. By definition, the innovation of the process $\epsilon_t^2$ is the variable $\nu_t = \epsilon_t^2 - \sigma_t^2$. Substituting in (3.1) the variables $\sigma_{t-j}^2$by $\epsilon_{t-j}^2 - \nu_{t-j}$, we get the representation

$$\epsilon_t^2 = \omega + \sum_{i=1}^{r} (\alpha_i + \beta_i)\epsilon_{t-i}^2 + \nu_t - \sum_{j=1}^{p} \beta_j \nu_{t-j} \tag{3.4}$$

where $r = \max(p, q)$, with the convention $\alpha_i = 0, (\beta_j = 0)$ if $i > q, (j > p)$. This equation has the linear structure of an ARMA model, allowing for simple computation of the linear predictions.The ARMA representation will be useful for estimation and identification of GARCH processes.

The above definition does not directly provide a solution process satisfying those conditions. The next definition is more restrictive but allows explicit solutions to be obtained. Let $\eta$ denote a probability distribution with null expectation and unit variance.

**Definition.** Strong GARCH(p,q) process.

Let $(\eta_t)$ be an iid sequence with distribution $\eta$. The process $(\epsilon_t)$ is called a strong GARCH(p,q) (with resprect to the sequence $(\eta_t)$) if

$$\epsilon_t = \sigma_t \eta_t$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{3.5}$$

where the $\alpha_i$ and $\beta_j$ are nonnegative constants and $\omega$ is a strictly positive constant.

Next we turn to stationarity study and identify stationarity conditions without proving them. For detailed proof the reader is advised to refer to the above mentioned reference by Francq and Zakoian. We first consider the GARCH(1,1) model which can be studied more explicitly.

When $p = q = 1$ the model (3.5) has the form

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \end{cases} \tag{3.6}$$

with $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$. Let $a(z) = \alpha z^2 + \beta$.

**Theorem 3.1.1.** *Strict stationarity of the strong GARCH(1,1) process.*

*If*

$$-\infty \leq \gamma := E[\log\{\alpha \eta_t^2 + \beta\}] < 0 \tag{3.7}$$

*then the infinite sum*

$$h_t = \{1 + \sum_{i=1}^{\infty} a(\eta_{t-1})...a(\eta_{t-i})\}\omega \tag{3.8}$$

*converges almost surely and the process* $(\epsilon_t)$ *defined by* $\epsilon_t = \sqrt{h_t}\eta_t$ *is the unique strictly stationary solution of model* (3.6). *This solution is nonanticipative and ergodic. If* $\gamma \geq 0$ *and* $\omega > 0$, *then there exists no strictly stationary solution.*

**Remark.** Condition (3.7) implies $\beta < 1$. Now, if

$$\alpha + \beta < 1$$

then (3.7) is satisfied since by application of Jensen inequality

$$E[\log\{a(\eta_t)\}] \leq \log(E\{a(\eta_t)\}) = \log(\alpha + \beta) < 0$$

**Theorem 3.1.2.** *Second-order stationarity of the GARCH(1,1) process.*

*Let $\omega > 0$. If $\alpha + \beta \geq 1$ , a nonanticipative and second-order stationary solution to the GARCH(1,1) mode does not exist. If $\alpha + \beta < 1$ , the process $\epsilon_t = \sqrt{h_t}\eta_t$ is second-order stationary. More precisely, $\epsilon_t$ is a weak white noise. Moreover, there exists no other second-order stationary and nonanticipative solution.*

*Proof.* If $(\epsilon_t)$ is a GARCH(1,1) process, in the sense of definition 1, which is second-order stationary and nonanticipative, we have

$$E(\epsilon_t^2) = E\{E(\epsilon_t^2|\epsilon_u, u < t)\} = E(\sigma_t^2) = \omega + (\alpha + \beta)E(\epsilon_{t-1}^2)$$

that is,

$$(1 - \alpha - \beta)E(\epsilon_t^2) = \omega$$

Hence, we must have $\alpha + \beta < 1$. In addition , we get $E(\epsilon_t^2) > 0$. Conversely, suppose $\alpha + \beta < 1$. By Remark 1, the strict stationarity condition is satisfied. It is thus sufficient to show the strictly stationary solution defined in $\epsilon_t = \sqrt{h_t}\eta_t$ admits a finite variance. The variable $h_t$ being an increasing limit of positive random variables, the infinite sum and the expectation can be permuted to give

$$E(\epsilon_t^2) = E(h_t) = [1 + \sum_{n=1}^{\infty} E\{a(\eta_{t-1})...a(\eta_{t-n})\}\omega] =$$

$$= [1 + \sum_{n=1}^{\infty}\{Ea(\eta_t)\}^n]\omega = [1 + \sum_{n=1}^{\infty}(\alpha + \beta)^n]\omega = \frac{\omega}{1 - \alpha - \beta}$$

This proves the second-order stationarity of the solution. Moreover, this solution is a white noise because $E(\epsilon_t) = 0$ and for all $h > 0$, $Cov(\epsilon_t, \epsilon_{t-h}) = 0$. One can also prove uniqueness but we refer the reader to the book by Francq and Zakoian. □

Now we turn to the general case of a strong GARCH(p,q) process. We use the following vector representation.

$$z_t = b_t + A_t z_{t-1} \tag{3.9}$$

where

$$b_t = \begin{pmatrix} \omega\eta_t^2 \\ 0 \\ \vdots \\ \omega \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{p+q} \quad z_t = \begin{pmatrix} \epsilon_t^2 \\ \vdots \\ \epsilon_{t-q+1}^2 \\ \sigma_t^2 \\ \vdots \\ \sigma_{t-p+1}^2 \end{pmatrix} \in \mathbb{R}^{p+q}$$

and

$$A_t = \begin{pmatrix} \alpha_1\eta_t^2 & & ... & & \alpha_q\eta_t^2 & \beta_1\eta_t^2 & & ... & & \beta_p\eta_t^2 \\ 1 & 0 & ... & & 0 & 0 & & ... & & 0 \\ 0 & 1 & ... & & 0 & 0 & & ... & & 0 \\ \vdots & \ddots & \ddots & & \vdots & \vdots & \ddots & \ddots & & \vdots \\ 0 & & ... & 1 & 0 & 0 & & ... & 0 & 0 \\ \alpha_1 & & ... & & \alpha_q & \beta_1 & & ... & & \beta_p \\ 0 & & ... & & 0 & 1 & 0 & ... & & 0 \\ 0 & & ... & & 0 & 0 & 1 & ... & & 0 \\ \vdots & \ddots & \ddots & & \vdots & \vdots & \ddots & \ddots & & \vdots \\ 0 & & ... & 0 & 0 & 0 & & ... & 1 & 0 \end{pmatrix}$$

is a $(p+q) \times (p+q)$ matrix.Equation (3.9) defines a first-order vector autoregressive model, with positive and iid matrix coefficients.The distribution of $z_t$ conditional on its infinite past coincides with its distribution of conditional on $z_{t-1}$ only, which means that $(z_t)$ is a Markov process. Model (3.9) is thus called the Markov representation of the GARCH(p,q) model. Iterating (3.9) gives

$$z_t = b_t + \sum_{k=1}^{\infty} A_t A_{t-1} ... A_{t-k+1} b_{t-k} \tag{3.10}$$

provided that the series exists almost surely.

The main tool for studying strict stationarity is the concept of the Lyapunov exponent. Let $A$ be a $(p+q) \times (p+q)$ matrix. The spectral radius of $A$ , denoted by $\rho(A)$, is defined ad the greatest modulus of its eigenvalues. Let $|| \cdot ||$ denote any norm on the space of the $(p+q) \times (p+q)$ matrices.We have the following algebra result:

$$\lim_{t\to\infty} \frac{1}{t} \log ||A^t|| = \log(\rho(A)) \tag{3.11}$$

This property has the following extension to random matrices.

**Theorem 3.1.3.** *Let $\{A_t, t \in \mathbb{Z}\}$ be a strictly stationary end ergodic sequence of random matrices, such that $E(\log^+ ||A_t||)$ is finite. We have*

$$\lim_{t\to\infty} \frac{1}{t} E(\log ||A_t A_{t-1} ... A_1||) = \gamma = \inf_{t\in\mathbb{N}^*} \frac{1}{t} E(\log ||A_t A_{t-1} ... A_1||) \tag{3.12}$$

*$\gamma$ is called the top Lyapunov exponent and $\exp(\gamma)$ is called the spectral radius of the sequence of matrices $\{A_t, t\in \mathbb{Z}\}$. Moreover,*

$$\gamma = \lim_{t\to\infty} a.s \frac{1}{t} \log ||A_t A_{t-1} ... A_1|| \tag{3.13}$$

The next theorem which goes without proof states the necessary conditions for the strict stationarity of GARCH(p,q).

**Theorem 3.1.4.** *Strict stationarity of the GARCH(p,q) model.*

*A necessary and sufficient condition for the existence of a strictly stationary solution to the GARCH(p,q) model is that*

$$\gamma < 0$$

*where $\gamma$ is the top Lyaponov exponent of the sequence $\{A_t, t\in \mathbb{Z}\}$. When the strictly stationary solution exists, it is unique, nonanticipative and ergodic.*

**Theorem 3.1.5.** *Second order stationarity.*

*If there exists a GARCH(p,q) process, in the sense of Definition 1, which is second-order* *stationary and nonanticipative, and if $\omega > 0$, then*

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1 \tag{3.14}$$

*Conversely, if (3.14) holds, the unique strictly stationary solution of model (3.5) is a weak white noise. In addition, there exists no other second-order stationary solution.*

When

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j = 1$$

the model is called an integrated GARCH(p,q) or IGARCH(p,q) model. This name is comes from the unit root in the autoregressive part of representation (3.4) and is introduced by analogy with the integrated ARMA models, ARIMA. However, this analogy can be misleading since there exists no stationary solution of an ARIMA model, whereas in IGARCH model admits a strictly stationary solution under very general conditions.

**Corollary 3.1.6.** *Suppose that the distribution of $\eta_t$ has an unbounded support and has no mass at 0. Then if $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j = 1$, model (3.5) admits a unique strictly stationary solution.*

**3.1.1.2 Identification.** Here we consider the problem of selecting an appropriate GARCH or ARMA-GARCH model for given observations of a centered stationary process. A large part of the finance theory rests on the assumption that prices follow a random walk. The price variation process, $X = (X_t)$, should thus constitute a martingale difference sequence, and should coincide with its innovation process,$\epsilon = (\epsilon_t)$. The first question addressed here will be the test of this property, at least a consequence of it: absence of correlation. The problem is far from trivial because standard tests for non-correlation are actually valid under an independence assumption. Such an assumption is too strong for GARCH processes which are dependent though uncorrelated.

If significant sample autocorrelations are detected in the price variations- in other words, if the random walk assumption cannot be sustained- the practitioner will try to fit an ARMA(P,Q) model to data before using a GARCH(p,q) model for the residuals.

Consider the GARCH(p,q) model

$$\begin{cases} \epsilon_t = \sigma_t \eta \\ \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \end{cases} \tag{3.15}$$

with $\eta_t$ a sequence of iid centered variables with unit variance. We saw that, whatever the orders p and q, the non-anticipative second-order stationary solution of (3.15) is a white noise, that is, a centered process whose theoretical autocorrelation $\rho(h) = 0$ for all $h \neq 0$.

Given observations $\epsilon_1, ..., \epsilon_n$, the theoretical autocorrelations of centered process $(\epsilon_t)$ are generally estimated by the sample autocorrelations (SACRs)

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad \hat{\gamma}(h) = \hat{\gamma}(-h) = n^{-1} \sum_{t=1}^{n-h} \epsilon_t \epsilon_{t+h}$$

for $h = 0, 1, ..., n-1$. If $(\epsilon_t)$ is an iid sequence of centered random variables with finite variance then

$$\sqrt{n}\hat{\rho}(h) \rightarrow N(0,1) \tag{3.16}$$

for all $h \neq 0$. For a strong white noise, the SACRs thus lie between the confidence bounds $\pm 1.96/\sqrt{n}$ with a probability of approximately 95% when $n$ is large. These significance bands are not valid for a weak white noise, in particular, for a GARCH process. Here we show valid asymptotic bands.

Let $\hat{\rho}_m = (\hat{\rho}(1), ..., \hat{\rho}(m))'$ denote the vector of the first $m$ SACRs, based on $n$ observations of the GARCH(p,q) process defined in (3.15). Let $\hat{\gamma}_m = (\hat{\gamma}(1), ..., \hat{\gamma}(m))'$ denote a vector of sample autocovariances (SACVs).

**Theorem 3.1.7.** *Asymptotic distributions of the SACVs and SACRs*

*If $(\epsilon_t)$ is the nonanticipative and stationary solution of the GARCH(p,q) model (3.15) and $E(\epsilon_t^4) < \infty$, then, when $n \rightarrow \infty$,*

$$\sqrt{n}\hat{\gamma}_m \rightarrow N(0, \Sigma_{\hat{\gamma}_m}) \quad \textit{and} \quad \sqrt{n}\hat{\rho}_m \rightarrow N(0, \Sigma_{\hat{\rho}_m} := \{E(\epsilon_t^2)\}^{-2} \Sigma_{\hat{\gamma}_m}),$$

*where*

$$\Sigma_{\hat{\gamma}_m} = \begin{pmatrix} E\epsilon_t^2\epsilon_{t-1}^2 & E\epsilon_t^2\epsilon_{t-1}\epsilon_{t-2} & \cdots & E\epsilon_t^2\epsilon_{t-1}\epsilon_{t-m} \\ E\epsilon_t^2\epsilon_{t-1}\epsilon_{t-2} & E\epsilon_t^2\epsilon_{t-2}^2 & & \vdots \\ \vdots & & \ddots & \\ E\epsilon_t^2\epsilon_{t-1}\epsilon_{t-m} & \cdots & & E\epsilon_t^2\epsilon_{t-m}^2 \end{pmatrix}$$

*is nonsingular. If law of* $(\eta_t)$ *is symmetric then* $\Sigma_{\hat{\gamma}_m}$ *is diagonal.*

A consistent estimator of $\hat{\Sigma}_{\hat{\gamma}_m}$ of $\mathbf{\Sigma_{\hat{\gamma}_m}}$ is obtained by replacing the generic term of $\Sigma_{\hat{\gamma}_m}$ by

$$n^{-1}\sum_{i=1}^{n}\epsilon_t^2\epsilon_{t-i}\epsilon_{t-j}$$

with, by convention, $\epsilon_s = 0$ for $s < 1$. Clearly, $\Sigma_{\hat{\rho}_m} := \hat{\gamma}^{-2}\Sigma_{\hat{\gamma}_m}$ is a consistent estimator of $\Sigma_{\hat{\rho}_m}$ and is almost surely invertible for $n$ large enough. This can be used to construct asymptotic significance bands for the SACRs of a GARCH process.

The standard portmanteau test for checking that the data is a realization of a strong where noise is that of Ljung and Box (1978). It involves computing the statistic

$$Q_m^{LB} := n(n+2)\sum_{n=1}^{m}\hat{\rho}^2(i)/(n-i)$$

and rejecting the strong white noise hypothesis if $Q_m^{LB}$ is greater than the $(1-\alpha)$-quantile of $\chi_m^2$.

Portmanteau tests are constructed for checking noncorrelation, but the asymptotic distribution of the statistics is no longer $\chi_m^2$ when the series departs from the strong white noise assumption. For instance, these tests are not robust to conditional heteroscedasticity. In the GARCH framework, we may wish to simultaneously test the nullity of the first autocorrelations using more robust portmanteau statistics.

**Theorem 3.1.8.** *Corrected portmanteau test in the presence of ARCH*

*Under the assumption of Theorem 5. the portmanteau statistic*

$$Q_m = n\hat{\rho}_m'\hat{\Sigma}_{\hat{\rho}_m}^{-1}\hat{\rho}_m$$

*has an asymptotic* $\chi^2_m$ *distribution.*

Denote by $r_m$ ($\hat{r}_m$) the vector of the $m$ first partial autocorrelations (sample partial autocorrelations (SPACs)) of the process $(\epsilon_t)$. We know that for a weak white noise, the SACRs and SPACs have the same asymptotic distribution. This applies in particular to a GARCH process. Consequently, under the hypothesis of GARCH white noise with a finite fourth-moment, consistent estimators of $\Sigma_{\hat{r}_m}$ are

$$\hat{\Sigma}^{(1)}_{\hat{r}_m} = \hat{\Sigma}_{\hat{\rho}_m} \quad or \quad \hat{\Sigma}^{(2)}_{\hat{r}_m} = \hat{J}_m \hat{\Sigma}_{\hat{\rho}_m} \hat{J}'_m,$$

where $\hat{J}_m$ is the matrix obtained by replacing $\rho_X(1), ..., \rho_X(m)$ by $\hat{\rho}_X(1), ..., \hat{\rho}_X(m)$ in the Jacobian matrix $J_m$ of the mapping $\rho_m \to r_m$, and $\Sigma_{\hat{\rho}_m}$is the consistent estimator of $\hat{\Sigma}_{\hat{\rho}_m}$.

One can test the simultaneous nullity of several theoretical partial autocorrelations using portmanteau tests based on the statistics

$$Q^{r,BP}_m = n\hat{r}'_m \hat{r}_m \quad and \quad Q^r_m = n\hat{r}'_m \left( \hat{\Sigma^{(i)}_{\hat{}\rho_m}} \right)^{-1} \hat{r}_m$$

The statistics $Q^{r,BP}_m, Q^{BP}_m, Q^{LB}_m$ have the same $\chi^2_m$ asymptotic distribution. Under the hypothesis of a pure GARCH process, the statistics $Q^r_m$ and $Q_m$ also have the same $\chi^2_m$ asymptotic distribution.

In case of the weak white noise the standard Barlett formulas are no longer valid. Assuming that the law of $\eta_t$ is symmetric the generalized Barlett formulas are given by

$$\lim_{n\to\infty} nCov\{\hat{\rho}_X(i), \hat{\rho}_X(j)\} = v_{ij} + v^*_{ij}$$

where

$$v_{ij} = \sum_{l=1}^{\infty} \omega_i(l)\omega_j(l), \quad v^*_{ij} = (\kappa_\epsilon - 1) \textstyle\sum_{l=1}^{\infty} \rho_{\epsilon^2}(l)\omega_i(l)\omega_j(l)$$

and

$$\omega_i(l) = \{2\rho_X(i)\rho_X(l) - \rho_X(l+x) - \rho_X(l-i)\}$$

Francq and Zakoian proposed the following algorithm for estimation of generalized Barlett bands.

1. Fit an AR($p_0$) model to the data using an information criterion for the selection of the order $p_0$.
2. Compute the autocorrelations $\rho_1(h)$, $h = 1, 2, ...$, of this AR($p_0$) model.
3. Compute the residuals $e_{p_0+1}, ..., e_n$ of this AR($p_0$)
4. Fit an AR($p_1$) model to the squared residuals $e^2_{p_0+1}, ..., e^2_n$ using an information criterion for $p_1$.
5. Compute the autocorrelations $\rho_2(h)$, $h = 1, 2, ...$, of this AR($p_1$) model.
6. Estimate $\lim_{n\to\infty} nCov\{\hat{\rho}(i), \hat{\rho}(j)\}$ by $v_{ij} + v^*_{ij}$ where

$$\hat{v}_{ij} = \sum_{l=-l_{max}}^{l_{max}} \rho_1(l)[2\rho_1(i)\rho_1(j)\rho_1(l) - 2\rho_1(i)\rho_1(l+j) - 2\rho_1(j)\rho_1(l+i) + \rho_1(l+j-i) + \rho_1(l-j-i)],$$

$$\hat{v}^*_{ij} = \frac{\hat{\gamma}_{\epsilon^2}(0)}{\hat{\gamma}^2_{\epsilon}(0)} \sum_{l=-l_{max}}^{l_{max}} \rho_2(l)[2\rho_1(i)\rho_1(j)\rho_1^2(l) -$$

$$-2\rho_1(j)\rho_1(l)\rho_1(l+i) - 2\rho_1(i)\rho_1(l)\rho_1(l+j) + \rho_1(l+i)\{\rho_1(l+j) + \rho_1(l-j)\}],$$

$$\hat{\gamma}_{\epsilon^2}(0) = \frac{1}{n-p_0} \sum_{t=p_0+1}^{n} e_t^4 - \hat{\gamma}^2_{\epsilon}(0), \quad \hat{\gamma}^2_{\epsilon}(0) = \tfrac{1}{n-p_1} \textstyle\sum_{t=p_0+1}^{n} e_t^2$$

where $l_{max}$ is a truncation parameter, numerically determined so as to have $|\rho_1(l)|$ and $|\rho_2(l)|$ less than a certain tolerance for all $l > l_{max}$.

In cases when distribution of $\eta_t$ is not symmetric, generalized Barlett formulas do not work. The following theorem gives asymptotic results for behavior of SACVs and SACRs for very general linear processes whose innovation is a weak white noise.

**Theorem 3.1.9.** *Let $(X_t)_{t\in\mathbb{Z}}$ be a real stationary process satisfying*

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}, \quad \textstyle\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$$

*where $(\epsilon_t)_{t\in\mathbb{Z}}$is a weak white noise such that $E\epsilon_t^4 < \infty$. Let $\Upsilon_t = X_t(X_t, X_{t+1}, ..., X_{t+m})'$, $\Gamma_\Upsilon(h) = E\Upsilon_t^* \Upsilon_t^{*'}$ and*

$$f_{\Upsilon^*}(\lambda) := \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \Gamma_\Upsilon(h),$$

*the spectral density of the process $\Upsilon^* = (\Upsilon_t^*)$, $\Upsilon_t^* = \Upsilon_t - E\Upsilon_t$. Then we have*

$$\lim_{n\to\infty} nVar\hat{\gamma}_{0:m} := \Sigma_{\hat{\gamma}_{0:m}} = 2\pi f_{\Upsilon^*}(0).$$

Francq and Zakoian propose the following algorithm for its estimation

1. Fit AR(r) model, with $r = 0, 1, ..., R$, to the data $\Upsilon_1 - \bar{\Upsilon}_n, ..., \Upsilon_{n-m} - \bar{\Upsilon}_n$ , where $\bar{\Upsilon}_n = (n-m)^{-1} \sum_{t=1}^{n-m} \Upsilon_t$.
2. Select a value $r_0$by minimizing an information criterion.
3. Take

$$\hat{\sum}_{\hat{\gamma}_{0:m}} = \hat{A}_{r_0}(1)^{-1} \hat{\Sigma}_{r_0} \hat{A}'_{r_0}(1)$$

where for a vector AR(r),

$$A_r(B)Y_t := Y_t - \sum_{i=1}^{r} A_i Y_{t-i} = Z_t$$

and $Z_t$ is white noise with variance $\Sigma_Z$.

Next we consider order determination for ARMA(P,Q). Executing this task by means of SACRs and SPACs is not an easy task. We present here an alternative method called the corner method.

Denote by $D(i,j)$ the $j \times j$ Toepliz matrix

$$D(i,j) = \begin{pmatrix} \rho_X(i) & \rho_X(i-1) & \cdots & \rho_X(i-j+1) \\ \rho_X(i+1) & & & \\ \vdots & & & \\ \rho_X(i+j-1) & \cdots & \rho_X(i+1) & \rho_X(i) \end{pmatrix}$$

and let $\Delta(i,j)$ denote its determinant. Since $\rho_X(h) = \sum_{i=1}^{P} a_i \rho_X(h-i) = 0$, for all $h > Q$, it is clear that $D(i,j)$ is not a full-rank matrix if $i > Q$ and $j > P$. In other words, P and Q are minimal orders if and only if $\Delta(i,j) = 0$, $\forall i > Q$ and $\forall j > P$, $\Delta(i,P) \neq 0$ $\forall i \geq Q$, $\Delta(Q,j) \neq 0$ $\forall j \geq P$.

The minimal orders P and Q can be illustrated by the following table

| $i\backslash j$ | 1 | 2 | . . . | Q | Q+1 | . . . . |
|---|---|---|---|---|---|---|
| 1 | $\rho_1$ | $\rho_2$ | . . . | $\rho_q$ | $\rho_{q+1}$ | |
| $\vdots$ | | | | | | |
| P | | | | × | × | × × × × |
| P+1 | | | | × | 0 | 0 0 0 0 |
| | | | | × | 0 | 0 0 0 0 |

where $\Delta(j,i)$ is at intersection if row $i$ and column $j$,and × denotes a nonzero element.

The entries in this table can be obtained by the following recursive formula

$$\Delta^2(i,j) = \Delta(i+1,j)\Delta(i-1,j) + \Delta(i,j+1)\Delta(i,j-1)$$

and letting $\Delta(i,0) = 1$, $\Delta(i,1) = \rho_X(|i|)$.

Replacing theoretical values by its estimates the orders P and Q are characterized by a corner of small values in the table. However, the notion of 'small' is not precise enough.

It is preferable to consider the studentized statistics defined, for $i = -K, ..., K$ and $j = 0, ..., K - |i| + 1$, by

$$t(i,j) = \sqrt{n}\frac{\hat{\Delta}(i,j)}{\hat{\sigma}_{\hat{\Delta}(i,j)}}, \quad \hat{\sigma}^2_{\hat{\Delta}(i,j)} = \frac{\partial \hat{\Delta}(i,j)}{\partial \rho'_K} \hat{\Sigma}_{\hat{\rho}_K} \frac{\partial \hat{\Delta}(i,j)}{\partial \rho_K}$$

where $\hat{\Sigma}_{\hat{\rho}_K}$ is a consistent estimator of the asymptotic covariance matrix of the first $K$ SACRs, and where

$$\frac{\partial\hat{\Delta}(i,0)}{\partial\rho_X(k)} = 0 \quad i = -K-1,...,K-1 \quad k = 1,...,K$$

$$\frac{\partial\hat{\Delta}(i,1)}{\partial\rho_X(k)} = I_{\{k\}}(|i|) \quad i = -K-1,...,K-1 \quad k = 1,...,K$$

$$\frac{\partial\hat{\Delta}(i,j+1)}{\partial\rho_X(k)} = \frac{2\hat{\Delta}(i,j)\frac{\partial\hat{\Delta}(i,j)}{\partial\rho_X(k)} - \hat{\Delta}(i+1,j)\frac{\partial\hat{\Delta}(i-1,j)}{\partial\rho_X(k)} - \hat{\Delta}(i-1,j)\frac{\partial\hat{\Delta}(i+1,j)}{\partial\rho_X(k)}}{\hat{\Delta}(i,j-1)}$$

$$-\frac{\{\hat{\Delta}(i,j)^2 - \hat{\Delta}(i+1,j)\hat{\Delta}(i-1,j)\}\frac{\partial\hat{\Delta}(i,j-1)}{\partial\rho_X(k)}}{\hat{\Delta}(i,j-1)^2}$$

When $\Delta(i,j) = 0$ the statistic $t(i,j)$ is asymptotically distributed as $N(0,1)$.If, in contrast, $\Delta(i,j) \neq 0$ then $\sqrt{n}|t(i,j)| \to \infty$ a.s. when $n \to \infty$.We can reject the hypothesis of nullity of $\Delta(i,j)$ at level $\alpha$ if $|t(i,j)|$ is beyond the $(1-\alpha/2)$-quantile of a $N(0,1)$.

To identify the orders of a GARCH(p,q) process, one can use the fact that $\epsilon_t^2$ follows an ARMA$(\tilde{P},\tilde{Q})$ with $\tilde{P} = \max(p,q)$, and $\tilde{Q} = p$.

To test linear restrictions on the parameters of a model the most popular tests are the Wald test,the Lagrange multiplier test, and likelihood ration test. Here we present the LM test.

Consider a parametric model, with true parameter value $\theta_0 \in \mathbb{R}^d$, and a null hypothesis

$$H_0 : R\theta_0 = r$$

where $R$ is a given $s \times d$ matrix of full rank $s$, and $r$ is a give $s \times 1$ vector. Under $H_0$ the test statistic is given by

$$LM_n := \frac{1}{n}\frac{\partial}{\partial\theta'}l_n(\hat{\theta}_c)\hat{J}^{-1}\frac{\partial}{\partial\theta}l_n(\hat{\theta}_c)$$

where

$$\hat{\theta} = \arg\sup_{\theta} l_n(\theta) \quad \hat{\theta}^c = \arg\sup\nolimits_{\theta:R\theta=r} l_n(\theta) \quad \hat{J} = -\frac{1}{n}\frac{\partial^2 l_n(\hat{\theta}^c)}{\partial\theta\partial\theta'}$$

asymptotically follows a $\chi_s^2$.

**3.1.1.3 Estimation.** The quasi-likelihood method is particularly relevant for GARCH models because it provides consistent and asymptotically normal estimators for strictly stationary GARCH processes under mild regularity conditions, but with no moment assumptions on the observed process. In this section we study QML method and give explicit formulas for derivatives of likelihood function and the optimization algorithm.

Assume that the observations $\epsilon_1, ..., \epsilon_n$ constitute a realization of a GARCH(p,q) process, more precisely a non-anticipative strictly stationary solution of

$$\begin{cases} \epsilon_t = \sqrt{h_t}\eta_t \\ h_t = \omega_0 + \sum_{i=1}^{q} \alpha_{0i}\epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_{0j}h_{t-j} \end{cases} \tag{3.17}$$

where $(\eta_t)$ is a sequence of iid variables of variance 1, $\omega_0 > 0$, $\alpha_{0i} > 0$, $\beta_{0j} > 0$. The orders $p$ and $q$ are assumed known. The vector of parameters

$$\theta = (\theta_1, ..., \theta_{p+q+1})^{'} := (\omega, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p)^{'} \tag{3.18}$$

belongs to a parameter space of the form

$$\Theta \subset (0, +\infty) \times [0, \infty)^{p+q} \tag{3.19}$$

The true value of the parameter is unknown, and is denoted by

$$\theta_0 = (\omega_0, \alpha_{01}, ..., \alpha_{0q}, \beta_{01}, ..., \beta_{0p})^{'} \tag{3.20}$$

To write the likelihood of the model,a distribution must be specified for the iid variable $\eta_t$. Here we do not make any assumption on the distribution of these variables, but work with a function, called the (Gaussian) quasi-likelihood, which, conditionally in some initial values, coincides with the likelihood when the $\eta_t$ are distributed as standard Gaussian. Later in the discussion we also show how to work with t-distributed $\eta_t$. Given the initial values $\epsilon_0, ..., \epsilon_{1-q}, \tilde{\sigma}_0^2, ..., \tilde{\sigma}_{1-p}^2$ to be specified below, the conditional Gaussian quasi-likelihood is given by

$$L_n(\theta) = L_n(\theta; \epsilon_1, ..., \epsilon_n = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\tilde{\sigma}_t^2}} \exp(-\frac{\epsilon_t^2}{2\tilde{\sigma}_t^2}))$$

where the $\tilde{\sigma}_t^2$ are recursively defined, for $t \geq 1$, by

$$\tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\theta) = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \tilde{\sigma}_{t-j}^2 \tag{3.21}$$

For a given value of $\theta$, under the second-order stationarity assumption, the unconditional variance is a reasonable choice for the unknown initial values:

$$\epsilon_0^2 = ... = \epsilon_{1-q}^2 = \sigma_0^2 = ... = \sigma_{1-p}^2 = \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j} \tag{3.22}$$

Such initial values are, however, not suitable for IGARCH models, in particular, and more generally when the second-order stationarity is not imposed. Indeed, the constant (3.22) would then take negative values for some values of $\theta$. In such a case, suitable initial values are

$$\epsilon_0^2 = ... = \epsilon_{1-q}^2 = \sigma_0^2 = ... = \sigma_{1-p}^2 = \omega \tag{3.23}$$

or

$$\epsilon_0^2 = ... = \epsilon_{1-q}^2 = \sigma_0^2 = ... = \sigma_{1-p}^2 = \epsilon_1^2 \tag{3.24}$$

A QMLE of $\theta$ is defined as any measurable solution $\hat{\theta}_n$ of

$$\hat{\theta}_n = \arg\max(L_n(\theta))$$

Taking the logarithm, it is seen that maximizing the likelihood is equivalent to minimizing, with respect to $\theta$

$$I_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} l_t \tag{3.25}$$

where

$$l_t = l_t(\theta) = \frac{\epsilon_t^2}{\tilde{\sigma}_t^2} + \log(\tilde{\sigma}_t^2)$$

and

$$\tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\theta) = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \tilde{\sigma}_{t-j}^2 \tag{3.26}$$

A QMLE is thus a measurable solution of the equation

$$\hat{\theta}_n = \arg\min(I_n(\theta)) \tag{3.27}$$

**Theorem 3.1.10.** *Strong consistency of the QMLE*

*Let $(\hat{\theta}_n)$ be a sequence of QMLEs satisfying (3.27), with initial conditions (3.23) or (3.24). Under assumptions*

1. *$\theta_0 \in \Theta$ and $\Theta$ is compact*
2. *$\gamma(A_0) < 0$ and for all $\theta \in \Theta$,$\sum_{i=1}^{p} \beta_j < 1$*
3. *$\eta_t^2$ has a nondegenerate distribution and $E(\eta_t^2) = 1$*
4. *If $p > 0$, $A_{\theta_0}(z) = \sum_{i=1}^{q} \alpha_i z^i$ and $B_{\theta_0}(z) = 1 - \sum_{j=1}^{p} \beta_j z^j$ have no common roots, $A_{\theta_0}(1) \neq 1$, and $\alpha_{0q} + \beta_{0p} \neq 0$.*

*almost surely*

$$\hat{\theta}_n \to \theta_0, \quad as \quad n \to \infty$$

**Theorem 3.1.11.** *Asymptotic normality of the QMLE*

*Under assumptions 1-4 and*

1. *$\theta_0 \in \Theta^0$, where $\Theta^0$denotes the interior of $\Theta$.*
2. *$\kappa_\eta = E(\eta_t^4) < \infty$.*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to N(0, (\kappa_\eta - 1)J^{-1})$$

*where*

$$J := E_{\theta_0}\left(\frac{\partial^2 l_t(\theta_0)}{\partial\theta\partial\theta'}\right) = E_{\theta_0}\left(\frac{1}{\sigma_t^4(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial\theta} \frac{\partial \sigma_t^2(\theta_0)}{\partial\theta'}\right) \tag{3.28}$$

*is a positive definite matrix.*

For more details and the proof of these theorems the an interested reader may refer to the book by Francq and Zakoian.

### 3.2 MULTIVARIATE GARCH.

As in the univariate case, we can define multivariate GARCH models by specifying their first two conditional moments. An $\mathbb{R}^{\mathrm{m}}$ -valued GARCH process $\epsilon_t$, with $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{mt})$, must then satisfy, for all $t$,

$$E(\epsilon_t|\epsilon_u, u < t) = 0, Var(\epsilon_t|\epsilon_u, u < t) = H_t$$

The multivariate extension of the notion of the strong GARCH process is based on an equation of the form

$$\epsilon_t = H_t^{1/2}\eta_t \tag{3.29}$$

where $\eta_t$ is a sequence of iid $\mathbb{R}^m$-valued variables with zero mean and identity covariance matrix. The matrix $H_t^{1/2}$can be chosen to be symmetric and positive definite but it can also be chose to be triangular, with positive diagonal elements. The latter choice may be of interest because if, for instance, $H_t^{1/2}$, is chosen to be lower triangular, the first component of $\epsilon_t$ only depends on the first component of $\eta_t$. When $m = 2$, we can thus set

$$\begin{cases} \epsilon_{1t} = h_{11,t}^{1/2}\eta_{1t} \\ \epsilon_{2t} = \frac{h_{12,t}}{h_{11,t}^{1/2}}\eta_{1t} + \left(\frac{h_{11,t}h_{22,t}-h_{12,t}^2}{h_{11,t}}\right)^{1/2}\eta_{2t} \end{cases}$$

Choosing a specification for $H_t$ is obviously more delicate than in the univariate framework because: (i) $H_t$ should be symmetric, and positive definite for all $t$; (ii) the specification should be simple enough, while being of sufficient generality; (iii) the specification should be parsimonious enough to enable feasible estimation.

### 3.2.1 Vector GARCH model.

The vector GARCH model is the most direct generalization of univariate GARCH: every conditional covariance is a function of lagged conditional variances as well as lagged cross-products of all components. Denote by vech($\cdot$) the operator that stacks the columns of the lower triangular part of its argument square matrix. The next definition is a natural extension of the standard GARCH(p,q) specification.

**Definition.** Let $\eta_t$ be a sequence of iid variables with distribution $\eta$. The process $\epsilon_t$ is said to admit a VEC-GARCH(p,q) representation if it satisfies

$$\begin{cases} \epsilon_t = H_t^{1/2}\eta_t \\ vech(H_t) = \omega + \sum_{i=1}^{q} A^{(i)} vech(\epsilon_{t-i}\epsilon_{t-i}^{'}) + \sum_{j=1}^{p} B^{(j)} vech(H_{t-j}) \end{cases}$$

where $\omega$ is a vector of size $m(m+1)/2 \times 1$, and $A^{(i)}$and $B^{(j)}$ are matrices of dimension $m(m+1)/2 \times m(m+1)/2$.

The VEC model potentially has an enormous number of parameters, which can make estimation of the parameters computationally infeasible.

### 3.2.2 Constant conditional correlation models.

Suppose that, for a multivariate GARCH process of the form (3.29), all the past information on $\epsilon_{kt}$, involving all the variables $\epsilon_{l,t-i}$ is summarized in the variable $h_{kk,t}$ with $Eh_{kk,t} = E\epsilon_{kt}^2$. Then, letting $\tilde{\eta}_{kt} = h_{kk,t}^{-1/2}\epsilon_{kt}$, we define for all $k$ a sequence of iid variables with zero mean and unit variance. The variables $\tilde{\eta}_{kt}$are generally correlated, so let $R = Var(\tilde{\eta}_t) = \rho_{kl}$, where $\tilde{\eta}_t = (\tilde{\eta}_{1t}, ..., \tilde{\eta}_{mt})$. The conditional variance of

$$\epsilon_t = diag(h_{11,t}^{1/2}, ..., h_{mm,t}^{1/2})\tilde{\eta}_t$$

is the written as

$$H_t = diag(h_{11,t}^{1/2}, ..., h_{mm,t}^{1/2})R diag(h_{11,t}^{1/2}, ..., h_{mm,t}^{1/2})$$

By construction, the conditional correlation between the components of $\epsilon_t$are time-invariant:

$$\frac{h_{kl,t}}{h_{kk,t}^{1/2}h_{ll,t}^{1/2}} = \rho_{kl}$$

To complete the specification, the dynamics of the conditional variances $h_{kk,t}$, has to be defined. The simplest constant conditional correlations (CCC) model relies on the following univariate GARCH specification:

$$h_{kk,t} = \omega_k + \sum_{i=1}^{q} a_{k,i}\epsilon_{k,t-i}^2 + \sum_{j=1}^{p} b_k h_{kk,t-j} \tag{3.30}$$

In the multivariate framework it seems natural to extend specification (3.30) by allowing $h_{kk,t}$ to depend not only on its own past, but also on the past of all variables $\epsilon_{l,t}$. Set

$$\bar{h}_t = \begin{pmatrix} h_{11,t} \\ \vdots \\ h_{mm,t} \end{pmatrix}, \quad D_t = \begin{pmatrix} \sqrt{h_{11,t}} & 0 & \dots & 0 \\ 0 & \ddots & & \\ \vdots & & \ddots & \\ 0 & \dots & & \sqrt{h_{mm,t}} \end{pmatrix} \quad \bar{\epsilon}_t = \begin{pmatrix} \epsilon_{1t}^2 \\ \vdots \\ \epsilon_{mt}^2 \end{pmatrix}$$

**Definition.** Let $\eta_t$ be a sequence of iid variables with distribution $\eta$. A process $\epsilon_t$ is called CCC-GARCH(p,q) if it satisfies

$$\epsilon_t = H_t^{1/2}\eta_t$$

$$H_t = D_t R D_t$$

$$\bar{h}_t = \bar{\omega} + \sum_{i=1}^{q} A_i \bar{\epsilon}_{t-i} + \sum_{j=1}^{p} B_j \bar{h}_{t-j}$$

where $R$ is a correlation matrix, $\bar{\omega}$ is an $m \times 1$ vector with positive coefficients, and the $A_i$ and $B_j$ are $m \times m$ matrices with nonnegative coefficients.

One advantage of this specification is that a simple condition ensuring the positive definiteness of $H_t$ os obtained though the positive coefficients for the matrices $A_i$ and $B_j$ and the choice of a positive definite matrix $R$.

### 3.2.3 Dynamic conditional correlation models.

Dynamic conditional correlations GARCH (DCC-GARCH) models are an extension of CCC-GARCH, obtained by introducing a dynamic for the conditional correlation. Hence, the constant matrix $R$ is replaced by a matrix $R_t$. Different DCC models are obtained depending on the specification of $R_t$. A simple example is

$$R_t = \theta_1 R + \theta_2 \Psi_{t-1} + \theta_3 R_{t-1}$$

where the $\theta_i$ are positive weights summing to 1, $R$ is a constant correlation matrix, and $\Psi_{t-1}$is empirical correlation matrix of $\epsilon_{t-1}, ..., \epsilon_{t-M}$. The matrix $R_t$is thus a correlation matrix.

Another way of specifying the dynamics of $R_t$ is by setting

$$R_t = diag(Q_t)^{-1/2} Q_t diag(Q_t)^{-1/2}$$

where $diag(Q_t)$ is the diagonal matrix constructed with diagonal elements of $Q_t$, and $Q_t$ is a sequence of covariance matrices. A natural parametrization is

$$Q_t = \theta_1 Q + \theta_2 \epsilon_{t-1} \epsilon_{t-1}^{'} + \theta_3 Q_{t-1}$$

where $Q$ is a covariance matrix.

### 3.2.4 BEKK-GARCH model.

**Definition.** BEKK-GARCH(p,q))

Let $\eta_t$ denote an iid sequence with common distribution $\eta$. The process $\epsilon_t$ is called a strong GARCH(p,q) with respect to the sequence $\eta_t$, if it satisfies

$$\begin{cases} \epsilon_t = H_t^{1/2}\eta_t \\ H_t = \Omega + \sum_{i=1}^{q}\sum_{k=1}^{K} A_{ik}\epsilon_{t-i}\epsilon_{t-i}^{'}A_{ik}^{'} + \sum_{j=1}^{p}\sum_{k=1}^{K} B_{jk}H_{t-j}B_{jk}^{'} \end{cases}$$

where $K$ is an integer, $\Omega$,$A_{ik}$ and $B_{jk}$ are square $m \times m$ matrices, and $\Omega$ is positive definite.

The specification obviously ensures that if the matrices $H_{t-i}$ are almost surely positive definite, then so is $H_t$.

### 3.2.5 Factor GARCH models.

**3.2.5.1 Factor models with idiosyncratic noise.** A very popular model factor model links individual returns $\epsilon_{it}$ to the market return $f_t$ thought a regression model

$$\epsilon_{it} = \beta_i f_t + \eta_{it}$$

The parameter $\beta_i$ can be interpreted as a sensitivity to the factor, and the noise $\eta_{it}$ as a specific risk which is conditionally uncorrelated with $f_t$. It follows that $H_t = \Omega + \lambda_t \boldsymbol{\beta}\boldsymbol{\beta}^{'}$ where $\boldsymbol{\beta}$ is the vector of sensitivities, $\lambda_t$ is the conditional variance of $f_t$ and $\Omega$ is the covariance matrix of the idiosyncratic terms. More generally, assuming the existence of $r$ conditionally uncorrelated factors, we obtain the decomposition

$$H_t = \Omega + \sum_{j=1}^{r} \lambda_{jt}\boldsymbol{\beta}_j\boldsymbol{\beta}_j^{'}$$

It is not restrictive to assume that the factors are linear combinations of the components of $\epsilon_t$. If, in addition, the conditional variances $\lambda_{jt}$ are specified as univariate GARCH, the model remains parsimonious in terms of unknown parameters and the above equation can be reduced to a particular BEKK model.

**3.2.5.2 Principle component GARCH model.** The concept of factor is central to principal component analysis (PCA) and to other methods of exploratory data analysis. PCA relies on decomposing the covariance matrix $V$ of $m$ quantitative variables as $V = P\Lambda P^{'}$, where $\Lambda$ is a diagonal matrix whose elements are the eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m$of $V$, and where $P$ is the orthogonal matrix of the corresponding eigenvectors. The first principal component is the linear combination of the $m$ variables, with weights given by the first column of $P$, which, in some sense, is the factor which best summarizes the set of $m$ variables. There exists $m$ principal components, which are uncorrelated and whose variances $\lambda_1, ..., \lambda_2$ are in decreasing order. It is natural considering this method for extracting the key factors of the volatilities of the $m$ components of $\epsilon_t$.

We obtain a principal component GARCH (PC-GARCH) or orthogonal GARCH (O-GARCH) model by assuming that

$$H_t = P\Lambda P^{'} \tag{3.31}$$

where $P$ is an orthogonal matrix and $\Lambda_t = diag(\lambda_{1t}, ..., \lambda_{mt})$ , where the $\lambda_{it}$ are the volatilities, which can be obtained from univariate GARCH-type models. This is equivalent to assuming

$$\epsilon_t = P\boldsymbol{f}_t$$

where $\boldsymbol{f}_t = P^{'}\epsilon_t$ is the pricipal component vector, whose components are orthogonal factors. If univariate GARCH(1,1) models are used of the factors $f_{it} = \sum_{j=1}^{m} P(j,i)\epsilon_{jt}$ then

$$\lambda_{it} = \omega_i + \alpha_i f_{it-1}^2 + \beta_i \lambda_{it-1}$$

## 3.3 NON-LINEAR MODELS FOR CONDITIONAL HETEROSCEDASTICITY.

In this section, we will review some popular nonlinear GARCH models following Terasvirta. We start off with models which are linear in parameters but can be made nonlinear by assuming a certain unknown quantity in them to be an unknown variable. The most frequently used models of this type are the GJR-GARCH model by Glosten et al. (1993) and the threshold generalized autoregressive conditional heteroscedasticity (TGARCH) model by Rabemananjara and Zakoian (1993) and Zakoian (1994). In applications, the GJR-GARCH model is typically assumed to be a first order GARCH model. It can be generalized to have higher order lags, although is practice, this almost never happens. The model for conditional variance looks like:

$$y_t = \mu_t + \epsilon_t$$

$$\epsilon_t = z_t h_t^{1/2}$$

$$h_t = \alpha_0 + \sum_{j=1}^{q} \left\{ \alpha_j + \kappa_j I(\epsilon_{t-j} < 0) \right\} \epsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{3.32}$$

where $I(A)$ is an indicator function. The idea of this model is to capture the leverage effect present in stock return series. This effect creates asymmetry: a negative shock has a greater impact on the conditional variance than the positive one with the same absolute value.

The GJR-GARCH model can be generalized by extending the asymmetry to the other components of the model. The volatility-switching GARCH or VS-GARCH model by Fornani and Mele (1997) is such an extension. The first order version of this model looks like:

$$h_t = \alpha_0 + \psi_0 sgn(\epsilon_{t-1}) + \left\{ \alpha_1 + \psi_1 sgn(\epsilon_{t-1}) \right\} \epsilon_{t-1}^2 + \left\{ \beta_1 + \psi_2 sgn(\epsilon_{t-1}) \right\} h_{t-1} \tag{3.33}$$

The TGARCH model is similar to (3.33) with one difference: what is being modeled is the conditional standard deviation and not the conditional variance. The model is defined by replacing $h_t$ by its square root and each $\epsilon_{t-j}^2$ by the corresponding absolute value $|\epsilon_{t-j}|$.

### 3.3.1 Nonlinear ARCH and GARCH models.

**3.3.1.1 Engle's nonlinear GARCH model.** The conditional variance in this model has the following form:

$$h_t = \alpha_0 + \alpha_1(\epsilon_{t-1} - \lambda)^2 + \beta_1 h_{t-1}$$

When $\lambda = 0$ , this model collapses into the standard GARCH(1,1) model. These models share the same weak stationarity condition $\alpha_1 + \beta_1 < 1$, and the above equation has $E\epsilon_t^2 = (\alpha_1 + \lambda^2)/(1 - \alpha_1 - \beta_1)$.

**3.3.1.2 Nonlinear ARCH model.** Higgins and Bera (1992) introduced a nonlinear ARCH model (NLARCH) that nests both the standard ARCH model and the logarithmic GARCH model of Pantula (1986) and Geweke (1986). It is an ARCH model with Box-Cox transformed variables:

$$\frac{h_t^\delta}{\delta} = \alpha_0 \frac{\omega^\delta - 1}{\delta} + \alpha_1 \frac{\epsilon_{t-1}^{2\delta} - 1}{\delta} + ... + \alpha_q \frac{\epsilon_{t-q}^{2\delta} - 1}{\delta} \tag{3.34}$$

where $0 \leq \delta \leq 1$, $\omega > 0$, $\alpha_0 > 0$,$\alpha_j \geq 0$ and $\sum_{j=0}^{q} \alpha_j = 1$.

This model has been very rarely used in practice.

**3.3.1.3 Asymmetric power GARCH model.** Ding et al. (1993) introduced the asummetric power GARCH or (APGARCH) model. The first-order APGARCH model has the following form:

$$h_t^\delta = \alpha_0 + \alpha_1(|\epsilon_{t-1}| - \lambda\epsilon_{t-1})^{2\delta} + \beta_1 h_{t-1}^\delta \tag{3.35}$$

where $\alpha_0 > 0$,$\alpha_1 > 0$,$\beta_1 \geq 0$,$\delta > 0$, and $|\lambda| \leq 1$, so it is nonlinear in parameters. Meitz and Saikkonen (2011) considered the special case $\delta = 1$ and called the model the asymmetric GARCH (AGARCH) model. Using the indicator variable, they showed that in this case (3.32) can be rewritten as a GJR-GARCH(1,1) model

$$h_t = \alpha_0 + \alpha_1(1 - \lambda)^2\epsilon_{t-1}^2 + 4\lambda\alpha_1 I(\epsilon_{t-1} < 0)\epsilon_{t-1}^2 + \beta_1 h_{t-1} \tag{3.36}$$

Considering a number of long daily return series,it was found that the autocorrelations $\rho(|\epsilon_t^{2\delta}, |\epsilon_{t-j}|^{2\delta})$ were maximized for $\delta = 1/2$. Fittin the APGARCH model to a long daily S&P 500 return series yielded $\hat{\delta} = 0.72$.

**3.3.1.4 Smooth transition GARCH model.** A generalization can be done to the GJR-GARCH model by replacing the indicator function by a continuous function of its argument and extending the transition to also include the intercept.

$$h_t = \alpha_{10} + \sum_{j=1}^{q} \alpha_{1j}\epsilon_{t-j}^2 + \left(\alpha_{20} + \sum_{j=1}^{q} \alpha_{2j}\epsilon_{t-j}^2\right) G_K(\gamma, c; \epsilon_{t-j}) + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{3.37}$$

where the transition function

$$G_K(\gamma, c; \epsilon_{t-j}) = \left(1 + \exp\left\{-\gamma \prod_{k=1}^{K} (\epsilon_{t-j} - c_k)\right\}\right)^{-1} \tag{3.38}$$

Here $\gamma > 0$ and $c = (c_1, ..., c_K)$.

Smooth transition GARCH models are useful in situations where the assumption of two distinct regimes is too rough an approximation to the asymmetric behavior of conditional variance.

The standard GARCH model has the undesirable property that the estimated model often exaggerates the persistence in volatility. This means that the estimated sum of the $\alpha$and $\beta$ coefficients is close to 1. Overestimated persistence results in poor volatility forecasts in the sense that following a large shock, the forecasts indicated too low a decrease if the conditional variance to more normal levels. In order to find a remedy for this problem, Lanne and Saikkonen (2005) proposed a smooth transition GARCH model, whose first-order version has the form:

$$h_t = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \delta_1 G_1(\theta; h_{t-1}) + \beta_1 h_{t-1} \tag{3.39}$$

In this equation, $G_1(\theta; h_{t-1})$ is a continuous, monotonically increasing bounded function of $h_{t-1}$. Since $h_{t-1} > 0$ almost surely, Lanne and Saikkonen used the cumulative distribution function of the Gamma distribution as the transition function. In empirical examples given in the paper, this parametrization clearly alleviates the problem of exaggerated persistence.

#### 3.3.1.5 Neural network ARCH and GARCH models.

The literature on nonlinear GARCH models also comprises models based on artificial neural networks (ANN) type of specification. The ANN-GARCH model of Donaldson and Kamstra (1997) has the following form:

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} + \sum_{j=1}^{s} \phi_j G(w_{t-j}, \Gamma_j)$$

where the hidden units are defined as follows:

$$G(w_{t-j}, \Gamma_j) = \left(1 + \exp\left\{\gamma_{0j} + \sum_{i=1}^{u} (w_{t-j}^{'} \gamma_{ji})\right\}\right)^{-1}$$

For a user of this model, specification of p,q,s, and u is an important issue, and the authors suggest the use of BIC criterion for this purpose.

A simpler ANN-GARCH model can be obtained by defining the hidden units as in Caulet and Peguin-Feissolle (2000). They give the following ANN-GARCH model:

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} + \sum_{j=1}^{s} \phi_j G(\gamma_{0j} + \epsilon_t^{'} \gamma_j)$$

where

$$G(\gamma_{0j} + \epsilon_t^{'} \gamma_j) = (1 + exp\{\gamma_{0j} + \epsilon_t^{'} \gamma_j\})^{-1}$$

#### 3.3.1.6 Time-varying GARCH.

It has been argued that the assumption of the standard GARCH model having constant parameters may not hold in practice unless the series to be modeled are sufficiently short. On can model such a behavior using the smooth transition GARCH model to fit such a situation. It is done assuming the transition function is a function of time:

$$G_K(\gamma, c; t^*) = \left(1 + \exp\left\{-\gamma \prod_{k=1}^{K} (t^* - c_k)\right\}\right)^{-1}$$

where $t^* = t/T$ is rescaled time and $T$ is the number of observations. The resulting time-varying parameter GARCH or TV-GARCH model has the form:

$$h_t = \alpha_0(t) + \sum_{j=1}^{q} \alpha_j(t)\epsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j(t) h_{t-j}$$

where $\alpha_0(t) = \alpha_{01} + \alpha_{02}G(\gamma, c; t^*)$, $\alpha_j(t) = \alpha_{j1} + \alpha_{j2}G(\gamma, c; t^*)$, and $\beta_j(t) = \beta_{j1} + \beta_{j2}G(\gamma, c; t^*)$.

The TV-GARCH model is non-stationary as the unconditional variance of $\epsilon_t$ varies deterministically over time.

**3.3.1.7 Testing standard GARCH against nonlinear GARCH.** The leading testing principle is the score or Lagrange multiplier principle, because then only the null model has to be estimated. These tests can be carried out in the so-called $TR^2$ form, and under the null hypothesis the test statistic has an asymptotic $\chi^2-$distribution. When the null hypothesis is the standard GARCH model, the test can be carried out in several stages:

1. Estimate the parameters of the GARCH model and compute the residual sum of squares $SSR_0 = \sum_{j=1}^{T}(\epsilon_t^2/\tilde{h_t} - 1)^2$,where $\tilde{h_t}$is the estimated conditional variance at $t$.
2. Regress $\tilde{z_t}^2 = \epsilon_t^2/\tilde{h_t}$ on the gradient of the log-likelihood function and the new variables, and compute the residual sum of squares $SSR_1$ from this auxiliary regression.
3. Form the test statistic

$$T\frac{SSR_0 - SSR_1}{SSR_0} \rightarrow \chi_m^2$$

under the null hypothesis of dimension $m$. When the null model is the standard GARCH, the gradient equals $\tilde{g_t} = \tilde{h_t}^{-1}(\partial h_t/\partial\omega)_0$, where $\omega = (\alpha_0, \alpha_1, ...\alpha_q, \beta_1..., \beta_p)$, and

$$(\partial h_t/\partial\omega)_0 = \tilde{u}_t + \sum_{i=1}^{p} \tilde{\beta}_i(\partial h_{t-i}/\partial\omega)_0$$

with $\tilde{u}_t = (1, \epsilon_{t-1}^2, ..., \epsilon_{t-q}^2, \tilde{h}_{t-1}, ..., \tilde{h}_{t-p})$. The subscript 0 indicates that the partial derivatives are evaluated under $H_0$. The auxiliary regression is thus

$$\tilde{z}_t^2 = a + \tilde{g}_t^{'}\delta_0 + v^{'}\delta_1 + \eta_t$$

## 3.4 REGIME-SWITCHING GARCH.

The idea of the RS approach to modeling asset returns is that the distribution of returns depends on a state of the market. For example, both the level and the time series properties of expected returns and variances may be different in bull and bear markets.

### 3.4.1 The RS-GARCH framework.

Assume that there are $k$ different market regimes and that of the market is in regime $j$ at time $t$, the conditional mean and variance of the return, $r_t$,are given by $\mu_{jt}$ and $\sigma_{jt}^2$, respectively. The RS-GARCH model can the be written in the following form:

$$r_t = \mu_{\Delta_t,t} + \sigma_{\Delta_t,t,}\eta_t$$

where $\Delta_t \in \{1, ..., k\}$ is a variable indicating the market regime at time $t$, and $\eta_t$ is a sequence of i.i.d. random variables with zero mean and unit variance. In many applications, the distribution of $\eta_t$ is taken to be Gaussian, so that the distribution of $r_t$ based on the information that we are in regime $j$ at time $t$, is likewise normal

$$f_{t-1}(r_r|\Delta_t = j) = \phi(r_t; \mu_j, \sigma_j^2) = \frac{1}{\sqrt{2\pi}\sigma_{jt}} \exp\left\{-\frac{(r_t - \mu_{jt})^2}{2\sigma_{jt}^2}\right\} \tag{3.40}$$

where $f_t$ denotes a conditional density based on the return history up to time $t$.

Suppose that the conditional probability for the market being in regime $j$ at time $t$ is $\pi_{jt}$, that is

$$p_{t-1}(\Delta_t = j) = \pi_{jt} \tag{3.41}$$

Then the conditional distribution of $r_t$ is a $k$-component finite normal mixture distribution, with density

$$f_{t-1}(r_r) = \sum_{j=1}^{k} \pi_{jt}\phi(r_t; \mu_{jt}, \sigma_{jt}^2) \tag{3.42}$$

where $\pi_{jt}$ are the mixing weights, and $\phi(r_t; \mu_{jt}, \sigma_{jt}^2)$ are the component densities, with component means $\mu_{jt}$ and component variances $\sigma_{jt}^2$.

The class of finite mixture distributions is known to exhibit considerable flexibility with respect to skewness and excess kurtosis, which are important features of financial return data. Moreover, and in contrast to many other flexible distributions used for that purpose, normal mixtures often provide an economically plausible disaggregation of the stochastic mechanism generating returns, such as the distinction between the bull and bear market dynamics.

### 3.4.2 Modeling the mixing weights.

A particular popular approach to modeling the dynamics of market regimes is the Markov-switching (MS) technique. It formalizes the intuition that market regimes may be persistent; for example, if we are in a bull market currently, then the probability of being in a bull market in the next period will be larger than that if the current regime were a bear market.

It is assumed that the regime process $\{\Delta_t\}$ follows a Markov chain with finite state space $S = \{1, ..., k\}$ and $k \times k$ transition matrix $P$,

$$P = \begin{pmatrix} p_{11} & \cdots & p_{k1} \\ \vdots & \ldots & \vdots \\ p_{1k} & \cdots & p_{kk} \end{pmatrix}$$

where the transition probabilities $p_{ij} = p(\Delta_t = j|\Delta_{t-1} = i)$. Let $\pi_t = (\pi_{1t}, ..., \pi_{kt})$ denote the distribution of the Markov chain at time $t$. It follows from the law of probability that for $j = 1, .., k$

$$\pi_{j,t+1} = p(\Delta_{t+1} = j) = \sum_{i=1}^{k} p(\Delta_t = i)p(\Delta_{t+1} = j|\Delta_t = i) = \sum_{i=1}^{k} \pi_{it}p_{ij}$$

or in matrix form, and then by iteration,

$$\pi_{t+1} = P\pi_t \quad \pi_{t+\tau} = P^{\tau}\pi_t \quad \tau \geq 1,$$

so that the elements of $P^{\tau}$ are the $\tau-$step transition probabilities. Moreover, under general conditions, there exists a stationary and long-run distribution.If regimes are persistent, this will be reflected in rather large diagonal elements of $P$. The degree of persistence can be measured by the magnitude of the second largest eigenvalue of the transition matrix $P$.

A further possibility to model the dynamics of the mixing weights is to make them depend on a set of predetermined variables. For example, in the two-component logistic mixture model, the weight if the first component is determined by

$$\pi_t = \frac{\exp\{\gamma' x_t\}}{1 + \exp(\gamma' x_t)}$$

where $\gamma = (\gamma_0, \gamma_1, ..., \gamma_{p-1})$ is a vector of parameters and $x_t$ is a vector of $p$ predetermined variables.

This can be generalized to more that two components which can lead to

$$\pi_{jt} = \frac{\theta_{jt}}{1 + \sum_{i=1}^{k} \theta_{it}} \quad j = 1, ..., k-1 \quad \pi_{kt} = 1 - \sum_{j=1}^{k-1} \pi_{jt}$$

where

$$\theta_{jt} = \exp\left(\gamma_{0j} + \sum_{i=1}^{u} \gamma_{ij}\epsilon_{t-i} + \sum_{i=1}^{v} \kappa_{ij}\pi_{j,t-i} + \sum_{i=1}^{w} \delta_{ij}|\epsilon_{t-i}|^{d}\right)$$

### 3.4.3 RS-GARCH specification.

There exists different specifications of RS-GARCH models. These have in common that the coefficients of the GARCH equation and thus the conditional variance at time $t$ depend o the current regime $\Delta_t$, and they differ in the way the lagged variance term in the regime-switching GARCH recursions is specified.

In the first version, this term is taken to be the lagged variance conditional on the previous regime, that is, the time series shocks, $\{\epsilon_t\}$, is modeled as

$$\epsilon_t = \sigma_{\Delta_t}\eta_t$$

where the regime-specific conditional variances are

$$\sigma_{jt}^2 = \omega_j + \alpha_j \epsilon_{t-1}^2 + \sigma_{\Delta_{t-1},t-1}^2$$

However, it was observed that maximum likelihood estimation of this specification is not feasible because of the path dependence, and thus MS-ARCH was used rather than MS-GARCH. To see the problem, suppose we want to calculate the likelihood function for the above model. We face the problem that $\Delta_{t-1}$and therefore $\sigma_{\Delta_{t-1},t-1}^2$ is not observable, and so we have to integrate it out. However, $\sigma_{\Delta_{t-1},t-1}^2$ likewise depends on the previous regime, $\Delta_{t-2}$, so that in the end, the conditional variance at time $t$ depends on the entire regime history up to time $t$. Thus the evaluation of the likelihood for a sample of $T$ observations requires integration over all $k^T$ possible regime paths. Recently, it was shown that the MS-GARCH model can be estimated using GMM or MCMC methods.

To circumvent the path dependence, Gray (1996) replaced $\sigma_{\Delta_{t-1},t-1}^2$ with the conditional variance of $\epsilon_{t-1}$, given only the observable information up to time $t-2$. With this information, the conditional distribution of $\epsilon_{t-1}$ is a $k$-component mixture with variance

$$h_{t-1} = \sum_{j=1}^{k} p_{t-2}(\Delta_{t-1} = j)\sigma_{j,t-1}^2$$

where $p_{t-2}(\Delta_{t-1} = j)$ are the conditional regime probabilities implied by the model for the regime process. This quantity $h_{t-1}$is then used instead of $\sigma_{\Delta_{t-1},t-1}^2$ in the regime-specific GARCH equation.

### 3.4.4 Estimation of RS-GARCH.

Since the regimes are not observable, we cannot use the transition probabilities $p_{ij}$ to directly forecast future regimes. However, we can use return history to compute regime inferences once we have estimated the parameters of an MS-GARCH process. These probabilities are also required for the likelihood function. To this end, we define, for each point of time, $t$ , a $k$-dimensional random vector $z_t = (z_{1t}, ..., z_{kt})$ with elements $z_{jt}$ such that

$$z_{jt} = \left\{ \begin{array}{ll} 1 & \Delta_t = j \\ 0 & \Delta_t \neq j \end{array} \right\}$$

Moreover, let $\boldsymbol{\epsilon}_\tau = \{\epsilon_\tau, \epsilon_{\tau-1}, ...\}$ denote the process up to time $\tau$, and let $z_{jt|\tau} = p(z_{jt} = 1|\boldsymbol{\epsilon}_\tau)$, be out probability inference of being in state $j$ and time $t$, based on the process up to time $\tau$, and let $\boldsymbol{z}_{t|\tau} = (z_{1t|\tau}, ..., z_{kt|\tau})'$. Then assuming conditional normality we have

$$\boldsymbol{z}_{t|\tau} = \frac{\boldsymbol{z}_{t|t-1} \odot \boldsymbol{f}_t}{\mathbf{1}_k'(\boldsymbol{z}_{t|t-1} \odot \boldsymbol{f}_t)}$$

$$\boldsymbol{z}_{t+1|t} = \boldsymbol{P}\boldsymbol{z}_{t|t}$$

where $\odot$ denotes element wise multiplication of conformable matrices, and

$$\boldsymbol{f}_t = \begin{pmatrix} \phi(\epsilon_t; \mu_1, \sigma_{1t}^2) \\ \vdots \\ \phi(\epsilon_t; \mu_k, \sigma_{kt}^2) \end{pmatrix} = (2\pi)^{-1/2} \begin{pmatrix} \sigma_{1t}^{-1} \exp\{-(\epsilon_t - \mu_1)^2/(2\sigma_{1t}^2)\} \\ \vdots \\ \sigma_{kt}^{-1} \exp\{-(\epsilon_t - \mu_k)^2/(2\sigma_{kt}^2)\} \end{pmatrix}$$

These equations can be used to calculate regime inferences recursively, and $\tau$-step ahead regime probabilities are obtained as $\boldsymbol{z}_{t+\tau|t} = \boldsymbol{P}^\tau \boldsymbol{z}_{t|t}$

To initialize the the recursion, the stationary distribution of the chain may be used. However, for reasonable long time series, as usually available in financial applications, the choice of the initial distribution will have a negligible impact on actual out-of-sample regime forecasts. The conditional density of $\epsilon_{t+1}$, given the process up to time $t$, is

$$f(\epsilon_{t+1}|\boldsymbol{\epsilon_t}) = \sum_{j=1}^{k} z_{j,t+1|t}\phi(\epsilon_{t+1}; \mu_j, \sigma_{j,t+1}^2) = \mathbf{1}_k'(\boldsymbol{z}_{t+1|t} \odot \boldsymbol{f}_{t+1}) \tag{3.43}$$

and the likelihood function for a sample of size $T$ is

$$\log L = \sum_{t=1}^{T} \log f(\epsilon_t|\boldsymbol{\epsilon}_{t-1}) = \sum_{t=1}^{T} \log[\mathbf{1}_k'(\boldsymbol{z}_{t|t-1} \odot \boldsymbol{f}_t)]$$

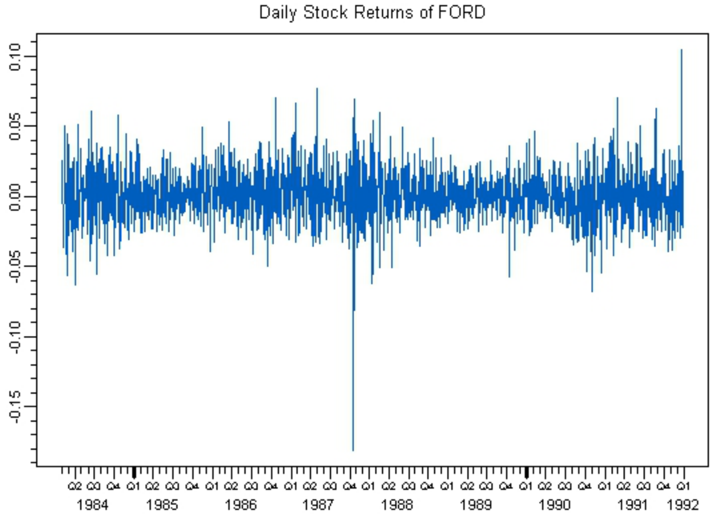

Figure 3.1: Ford stock returns

## 3.5 PRACTICAL ISSUES WITH GARCH.

## 3.6 APPLICATIONS.

### 3.6.1 Analysis of stock data.

In this example, we examine the daily series of Ford stock returns. Although there is little serial correlation in the time series itself, it seems that both large changes and small changes are clustered together, which is typical of many high-frequency macroeconomic and financial time series. To confirm this, we look at the autocorrelation function of Ford returns and its squared returns.

Obviously, there is no autocorrelation in the return series itself, while the squared returns exhibit significant autocorrelation at least up to lag 5. We see that time series of Ford stock returns exhibits time varying conditional heteroscedasticity or volatility clustering.

Testing for ARCH effects using the Langrange Multiplier Test we get p-value which is smaller than the conventional 5% level, so we reject the null hypothesis that there are no

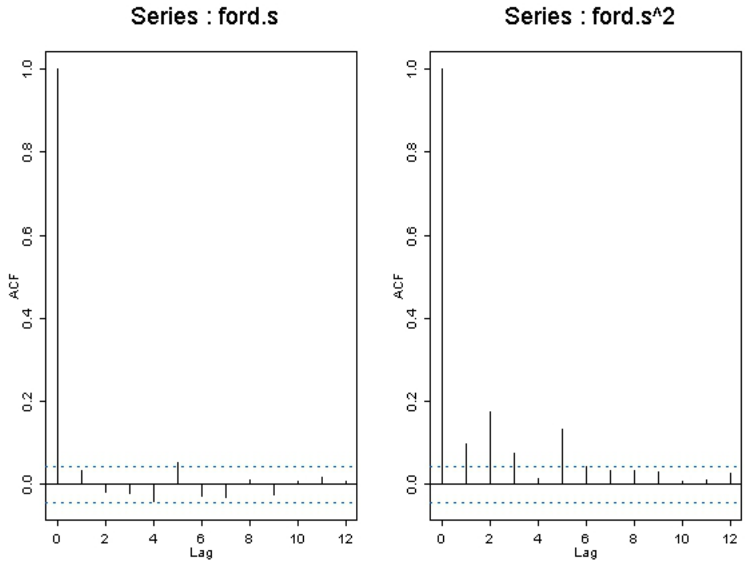

Figure 3.2: Autocorrelation function

ARCH effects.

The model we are going to estimate generally looks like

$$y_y = c + \epsilon_t$$

$$\epsilon_t = z_t \sigma_t$$

$$\sigma_t^2 = a_0 + \sum_{i=1}^{p} a_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2$$

Let us fit the GARCH(1,1) model to the Ford series. We get values for our parameters $c = 7.70e - 04$, $a_0 = 6.534e - 06$, $a_1 = 7.45e - 02$ , and $b_1 = 9.102e - 01$. The sum of $a_1 + b_1 = 0.985$ which indicates a covariance stationary model with a high degree of persistence on the conditional variance. If the model is successful at modeling the serial correlation structure in the conditional mean and conditional variance, then there should be no autocorrelation left in the standardized residuals and squared standardized residuals.

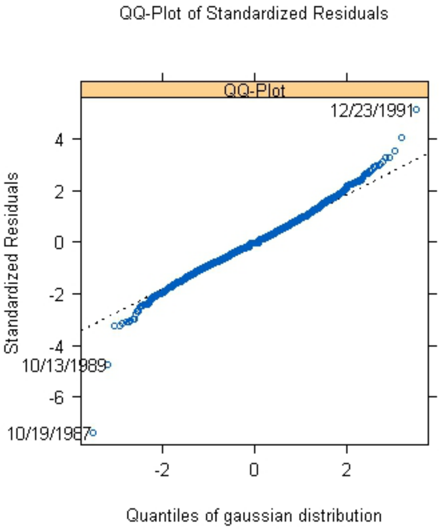

Figure 3.3: QQ plot of residuals when models with the normal distribution

This can be done using Ljung-Box Test. In both cases, the null hypothesis that there are no autocorrelation left cannot be rejected because the p-values in both cases are greater than the conventional 5% level.

The basic GARCH model assumes a normal distribution for the errors $\epsilon_t$. If the model is correctly specified then the estimated standardized residuals $\epsilon_t/\sigma_t$ should behave like a standard normal random variable. We can run the Jarque-Bera or Shapiro-Wilks test for the standardized residuals. However, in this case these tests give opposite conclusions. To get a more decisive conclusion we can use the qq-plot.

In the above example, a normal error distribution has been used. However, given the well known fat tails in financial time series, it may be more desirable to use a distribution which has fatter tails than the normal distribution. We can try to use the Student's t distribution and the Generalized Error Distribution. We estimate the GARCH(1,1) model for the Ford series using the above distributions. The results look much better for Student's Distribution.

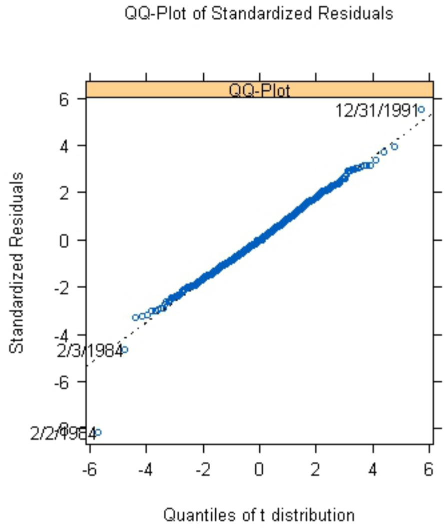

Figure 3.4: QQ plot of residuals when modeled with Student's t distribution

### 3.6.2 Option pricing.

The GARCH process and its variants have gained increasing prominence for modeling financial time series. In this section we discuss the GARCH options pricing approach developed by Duan (1995). The GARCH option pricing model has three distinctive features. First, the GARCH option price is a function of the risk premium embedded in the underlying asset. This contrasts with the standard preference-free option pricing result. Second, the GARCH option pricing model is non-Markovian. In the option pricing literature, the underlying asset value is usually assumed to follow a diffusion process. The standard approach is thus Markovian. Third, the GARCH option pricing model can potentially explain some well-documented systematic biases associated with Black-Scholes model. These biases include underpricing of the OTM options, underpricing of options on low-volatility securities,underpricing of short-maturity options, and the U-shaped implied volatility curve.The GARCH option pricing model also subsumes the Black-Scholes model because the homoscedastic asset return process is a special case of the GARCH model.

Due to complex nature of the GARCH process, a generalized version of risk neutralization, referred to as the local risk-neutral valuation relationship (LRNVR), is called for. The LRNVR stipulates that the one-period ahead conditional variance is invariant with respect to a change to the risk-neutralized pricing measure. This is important because, in the context of the GARCH process, the unconditional variance or any conditional variance beyond one period is not invariant to the change on measures caused by risk neutralization.

#### 3.6.2.1 The GARCH option pricing model.

Consider a discrete-time economy and let $X_t$ be the asset price at time $t$. Its one-period rate of return is assumed to be conditionally log-normally distributed under probability measure $P$. That is,

$$\log(\frac{X_t}{X_{t-1}}) = r + \lambda\sqrt{h_t} - \frac{1}{2}h_t + \epsilon_t$$

where $\epsilon_t$ has mean zero and conditional variance $h_t$ under measure $P$; $r$ is a constant one-period risk-free rate of return (continuously compounded) and $\lambda$the constant unit risk premium. Under conditional log-normality, one plus the conditionally expected return equals $\exp(r + \lambda\sqrt{h_t})$. It thus suggests that $\lambda$can be interpreted as the unit risk premium.

We also assume that $\epsilon_t$ follows a GARCH(p,q) process under measure $P$. Formally

$$\epsilon_t|\phi_{t-1} \sim N(0, h_t)$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}$$

where $\phi_i$ is the information set of all information up to and including time $t$. Using an alternative specification for $h_t$ will not change the basic option pricing results as long as conditional normality remains in place.

In order to develop the GARCH option pricing model, the conventional risk-neutral valuation relationship has to be generalized to accommodate heteroscedasticity of the asset return process. We thus introduce a generalized version of this principle.

**Definition.** A pricing measure $Q$ is said to satisfy the locally risk-neutral valuation relationship (LRNVR) if measure $Q$ is mutually absolutely continuous with respect to measure $P$, $X_t/X_{t-1}|\phi_{t-1}$distributes log-normally under $Q$,

$$E^Q(X_t/X_{t-1}|\phi_{t-1}) = e^r$$

and

$$Var^Q(\log(X_t/X_{t-1})|\phi_{t-1}) = Var^P(\log(X_t/X_{t-1})|\phi_{t-1})$$

almost surely with respect to measure $P$.

In the above definition of the LRNVR, the conditional variances under the two measures are required to be equal. This is desirable because on can observe and hence estimate the conditional variance under $P$.

The implication of LRNVR is presented in the following theorem.

**Theorem 3.6.1.** *The LRNVR implies that, under pricing measure $Q$,*

$$\log(\frac{X_t}{X_{t-1}}) = r - \frac{1}{2}h_t + \xi_t,$$

*where*

$$\xi_t|\phi_{t-1} \sim N(0, h_t)$$

*and*

$$h_t = \alpha_0 + \sum_{i=1}^{q}(\xi_{t-i} - \lambda\sqrt{h_{t-i}})^2 + \sum_{i=1}^{p}\beta_i h_{t-i}$$

*Proof.* Since $X_t/X_{t-1}$ distributes log-normally under measure $Q$, it can be written as

$$\log(\frac{X_t}{X_{t-1}}) = \nu_t + \xi_t$$

where $\nu_t$ is the conditional mean and $\xi_t$ is a $Q$-normal random variable. The conditional mean of $\xi_t$ equals zero and its conditional variance is to be determined. First, we prove that $\nu_t = r - \frac{1}{2}h_t$.

$$E^Q\left(\frac{X_t}{X_{t-1}}|\phi_{t-1}\right) = E^Q(e^{\nu_t+\xi_t}|\phi_{t-1}) = e^{\nu_t+h_t/2}$$

where $h_t = Var^P(\log(X_t/X_{t-1})|\phi_{t-1}) = Var^Q(\log(X_t/X_{t-1})|\phi_{t-1})$ by LRNVR. Since $E^Q(X_t/X_{t-1}|\phi_{t-1}) = e^r$ by LRNVR, it follows that $\nu_t = r - \frac{1}{2}h_t$.It remains to prove that $h_t$can indeed be expressed as stated in the above theorem. By the preceding result, $r + \lambda\sqrt{h_t} - \frac{1}{2}h_t + \epsilon_t = r - \frac{1}{2}h_t + \xi_t$. This implies that $\epsilon_t = \xi_t - \lambda\sqrt{h_t}$. □

This theorem implied the the form of the GARCH(p,q) process remains largely intact with respect to local risk neutralization. The conditional variance process under risk-neutralized pricing measure, in not a GARCH process. The variance innovation is governed by $q$ noncentral chi-square random variables with one degree of freedom, where as the GARCH process under $P$ can be seen as the process governed by $q$ central chi-square innovations. The theorem suggests that the unit risk premium $\lambda$,influences the conditional variance process globally although the risk has been locally neutralized under the pricing measure $Q$. In other words, local risk neutralization is not equivalent to global risk neutralization.

Pricing contingent payoffs requires temporally aggregating one-period asset returns to arrive at a random terminal asset price at some future point in time. The terminal asset price is derived in the following corollary:

**Corollary 3.6.2.**

$$X_T = X_t \exp\left[(T-t)r - \frac{1}{2}\sum_{s=t+1}^{T} h_s + \sum_{s=t+1}^{T} \xi_s\right]$$

**Corollary 3.6.3.** *The discounted asset price process $e^{-rt}X_t$ is a $Q$-martingale.*

**Corollary 3.6.4.** *The option price is given by*

$$C_t^{GH} = e^{-(T-t)r}E^Q[(X_T - K)^+|\phi_t]$$

**Corollary 3.6.5.** *The options delta is given by*

$$\Delta_t^{GH} = e^{-r(T-t)}E^Q\left[\frac{X_T}{X_t}I_{\{X_T \geq K\}}|\phi_t\right]$$

### 3.6.3 Risk management.

VaR has to do with the possible loss of a portfolio in a given time horizon. VaR should be computed using the predictive distribution of future losses, that is, the conditional distribution of the future losses using current information. However, for horizon $h > 1$, this conditional distribution may be hard to obtain.

To be more specific, consider a portfolio whose value at time $t$ is a random variable denoted $V_t$. At horizon $h$, the loss is denoted

$$L_{t,t+h} = -(V_{t+h} - V_t)$$

The distribution of $L_{t,t+h}$ is called the loss distribution. This distribution is used to compute the regulatory capital which allows certain risks to be covered.

**Definition.** The $(1-\alpha)$-quantile of the conditional loss distribution is called the VaR at the level $\alpha$:

$$VaR_{t,h}(\alpha) := inf\{x \in \mathbb{R} | P_t[L_{t,t+h} \leq x] \geq 1-\alpha\},$$

when this quantile is positive. By convention $VaR_{t,h}(\alpha) = 0$ otherwise.

Let introduce the first two moments of $L_{t,t+h}$ conditional on the information available at time $t$:

$$m_{t,t+h} = E(L_{t,t+h}), \quad \sigma^2_{t,t+h} = Var(L_{t,t+h})$$

Suppose that

$$L_{t,t+h} = m_{t,t+h} + \sigma_{t,t+h} L^*_h$$

where $L^*_h$ is a random variable with cumulative distribution function $F_h$. Denote by $F_h^{\leftarrow}$the quantile function of the variable $L^*_h$, defined as the generalized inverse of $F_h$:

$$F_h^{\leftarrow}(\alpha) = inf\{x \in \mathbb{R} | F_h(x) \geq \alpha\}$$

If $F_h$is continuous and strictly increasing, we simply have $F_h^{\leftarrow}(\alpha) = F_h^{-1}(\alpha)$, where $F_h^{-1}$ is the ordinary inverse of $F_h$. In follows that

$$1-\alpha = P_t[VaR_{t,h}(\alpha) \geq m_{t,t+h} + \sigma_{t,t+h}L_h^*] = F_h\left(\frac{VaR_{t,h}(\alpha) - m_{t,t+h}}{\sigma_{t,t+h}}\right)$$

Consequently,

$$VaR_{t,h}(\alpha) = m_{t,t+h} + \sigma_{t,t+h}F_h^{\leftarrow}(1-\alpha).$$

Consider the price of a portfolio, defined as a combination of the prices of $d$ assets, $p_t = a^{'}P_t$ and denote price variation $\Delta P_t = P_t - P_{t-1}$.

$$L_{t,t+h} = -(p_{t+h} - p_t) = -a^{'}\sum_{i=1}^{h}\Delta P_{t+i}$$

**Example.** If the $\Delta P_{t+i}$ are iid $N(m,\Sigma)$ distributed, the law of $L_{t,t+h}$ is $N(-a^{'}mh, a^{'}\Sigma ah)$. It follows then that

$$VaR_{t,h}(\alpha) = -a^{'}mh + \sqrt{a^{'}\Sigma a}\sqrt{h}\Phi^{-1}(1-\alpha)$$

**Example.** Suppose now that

$$\Delta P_t - m = A(\Delta P_{t-1} - m) + U_t \quad U_t \sim N(0,\Sigma)$$

where $A$ is a matrix whose eigenvalues have modulus strictly less than 1. The process $\Delta P_t$ is then stationary with expectation $m$. It can be verified that

$$VaR_{t,h}(\alpha) = a^{'}\mu_{t,h} + \sqrt{a^{'}\Sigma_h a}\Phi^{-1}(1-\alpha)$$

where $A_i = (I - A^i)(I-A)^{-1}$,

$$\mu_{t,h} = -mh - AA_h(\Delta P_t - m) \quad \Sigma_h = \sum_{j=1}^{h} A_{h-j+1}\Sigma A^{'}_{h-j+1}$$

If is often more convenient to work with log-returns $r_t = \Delta\log(p_t)$, assumed to be stationary, than with the price variations. Letting $q_t(h,\alpha)$ be the $\alpha$ -quantile of the conditional distribution of of the future returns $r_{t+1} + \cdots + r_{t+h}$

$$VaR_{t,h}(\alpha) = \{1 - e^{q_t(h,\alpha)}\}p_t$$

Even if VaR is the most widely used risk measure, the choice of an adequate risk measure is an open issue. Var is often criticized for not satisfying, for any distribution of the price variation, the subadditivity property. Subadditivity means that the VaR of two portfolios after they have been merged should be no greater than the sum of their VaRs before they were merged. In other words VaR does not favor diversification.

The simplest estimation method os based on the $K$ last returns at horizon $h$. that is, $r_{t+h-i}(h) = \log(p_{t+h-i}/p_{t-i})$, for $i = h...h+K-1$. These $K$ returns are viewed as scenario for future returns. The nonparametric historical VaR is simply obtained by replacing $q_t(h,\alpha)$ by the empirical $\alpha$-quantile of the last $K$ returns. A parametric version is obtained by fitting a particular distribution to the returns, for example, a Gaussian which amounts to replacing $q_t(h,\alpha)$ by $\hat{\mu} + \hat{\sigma}\Phi^{-1}(\alpha)$. These methods have little theoretical justification.

One can use more sophisticated GARCH-type models.The estimate $VaR_t(1,\alpha)$ it suffices to estimate $q_t(1,\alpha)$ by $\hat{\sigma}_{t+1}\hat{F}^{-1}(\alpha)$, where $\hat{\sigma}^2_t$ is the conditional variance estimated by a GARCH-type model, and $\hat{F}$ is an estimate of the distribution of the normalized residuals. It is important to note that even for a simple Gaussian GARCH(1,1), there is no explicit available formula for computing $q_t(h,\alpha)$ when $h > 1$. In that case one has to use simulations to evaluate the quantile. The follwing procedure may be used:

- Fit a model, for instance GARCH(1,1), on the observed returns $r_t = \epsilon_t$ and deduce the estimate volatility $\hat{\sigma}^2_t$ for $t = 1, ..., n+1$.
- Simulate a large number $N$ of scenarios for $\epsilon_{n+1}, ..., \epsilon_{n+h}$ by iterating independently for $i = 1, ..., N$, the following steps.

1. simulate the values $\eta^{(i)}_{n+1}, ..., \eta^{(i)}_{n+h}$ iid with distribution $\hat{F}$.
2. set $\sigma^{(1)}_{n+1} = \hat{\hat{\sigma}}_{n+1}$ and $\epsilon^{(i)}_{n+1} = \sigma^{(i)}_{n+1}\eta^{(i)}_{n+1}$.
3. for $k = 2, ..., h$, set $\left(\sigma^{(i)}_{n+k}\right)^2 = \hat{\omega} + \hat{\alpha}\left(\epsilon^{(i)}_{n+k-1}\right)^2 + \hat{\beta}\left(\sigma^{(i)}_{n+k-1}\right)^2$ and $\epsilon^{(i)}_{n+k} = \sigma^{(i)}_{n+k}\eta^{(i)}_{n+k}$

- Determine the empirical quantile of the simulations $\epsilon^{(i)}_{n+h}$

# 4.0 DISCRETE STOCHASTIC VOLATILITY MODELS

## 4.1 STATE-SPACE REPRESENTATION. LINEAR, GAUSSIAN MODELS.

### 4.1.1 Filtering.

The linear Gaussian state space model looks like

$$y_t = Z_t \alpha_t + \epsilon_t$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t$$

where $\epsilon_t \sim N(0, H_t)$, $\eta_t \sim N(0, Q_t)$,and $\alpha_1 \sim N(a_1, P_1)$

Let $Y_{t-1}$ denote the set of past observations $y_1, ..., y_{t-1}$. Starting at $t = 1$ and building up the distributions of $\alpha_t$ and $y_t$ recursively, it is easy to show that $p(y_t|\alpha_1, ..., \alpha_t, Y_t) = p(y_t|\alpha_t)$ and $p(\alpha_{t+1}|\alpha_1, ..., \alpha_t, Y_t) = p(\alpha_{t+1}|\alpha_t)$.In this section we derive the Kalman filter for this model for the case where the initial state $\alpha_1$ is $N(a_1, P_1)$ where $a_1$ and $P_1$ are known. Our goal is to calculate the conditional distribution of $\alpha_{t+1}$ given $Y_t$. Since all distributions are normal, conditional distributions of subsets of variables given other subsets of variables are also normal; the required distribution is therefore determined by knowledge of $a_{t+1} = E(\alpha_{t+1}|Y_t)$ and $P_{t+1} = Var(\alpha_{t+1}|Y_t)$. Assume that $\alpha_t$ given $Y_{t-1}$ is $N(a_t, P_t)$. We now show how to calculate $a_{t+1}$ and $P_{t+1}$from $a_t$ and $P_t$ recursively.

Since $\alpha_{t+1} = T_t \alpha_t + R_t \eta_t$, we have

$$a_{t+1} = E(T_t \alpha_t + R_t \eta_t | Y_t) = T_t E(\alpha_t | Y_t) \tag{4.1}$$

$$P_{t+1} = Var(T_t \alpha_t + R_t \eta_t | Y_t) = T_t Var(\alpha_t | Y_t) T_t^{'} + R_t Q_t R_t^{'} \tag{4.2}$$

Let

$$v_t = y_t - E(y_t|Y_{t-1}) = y_t - E(Z_t\alpha_t + \epsilon_t|Y_{t-1}) = y_t - Z_t a_t \tag{4.3}$$

Then $v_t$ is the one-step forecast error of $y_t$ given $Y_{t-1}$.When $Y_{t-1}$ and $v_t$ are fixed then $Y_t$ is fixed and vice versa. Thus $E(\alpha_t|Y_t) = E(\alpha_t|Y_{t-1}, v_t)$. But $E(v_t|Y_{t-1}) = E(y_t - Z_t a_t|Y_{t-1}) = E(Z_t\alpha_t + \epsilon_t - Z_t a_t|Y_{t-1}) = 0$. Consequently, $E(v_t) = 0$ and $Cov(y_j, v_t) = 0$ with $j = 1, ..., t-1$. Using regression we have

$$E(\alpha_t|Y_t) = E(\alpha_t|Y_{t-1}, v_t) = E(\alpha_t|Y_{t-1}) + \frac{Cov(\alpha_t, v_t)}{Var(v_t)} v_t = a_t + M_t F_t^{-1} v_t \tag{4.4}$$

where $M_t = Cov(\alpha_t|v_t)$, $F_t = Var(v_t)$ and $E(\alpha_t|Y_{t-1}) = a_t$ by definition of $a_t$.

Here,

$$M_t = Cov(\alpha_t, v_t) = E(E\{\alpha_t(Z_t\alpha_t + \epsilon_t - Z_t a_t)^{'}|Y_{t-1}\}) = E(E\{\alpha_t(\alpha_t - a_t)^{'} Z_t^{'}|Y_{t-1}\}) = P_t Z_t^{'} \tag{4.5}$$

and

$$F_t = Var(Z_t\alpha_t + \epsilon_t - Z_t a_t) = Z_t P_t Z_t^{'} + H_t \tag{4.6}$$

We assume that $F_t$ is nonsingular; this assumption is normally valid in well-formulated models. Combining the above equations one gets

$$a_{t+1} = T_t a_t + T_t M_t F_t^{-1} v_t = T_t a_t + K_t v_t \tag{4.7}$$

with

$$K_t = T_t M_t F_t^{-1} = T_t P_t Z_t^{'} F_t^{-1} \tag{4.8}$$

We observe that $a_{t+1}$ has been obtained as a linear function of the previous value $a_t$ and $v_t$, the forecast error of $y_t$ given $Y_{t-1}$.

Likewise, using regression approach we can compute the variance. We have

$$Var(\alpha_t|Y_t) = Var(\alpha_t|Y_{t-1}, v_t) = Var(\alpha_t|Y_{t-1}) - Cov(\alpha_t, v_t)Var(v_t)^{-1}Cov(\alpha_t, v_t)^{'} = \quad (4.9)$$

$$= P_t - M_t F_t^{-1} M_t^{'} = P_t - P_t Z_t^{'} F_t^{-1} Z_t P_t \quad (4.10)$$

and

$$P_{t+1} = T_t P_t L_t^{'} + R_t Q_t R_t^{'} \quad (4.11)$$

with

$$L_t = T_t - K_t Z_t \quad (4.12)$$

These recursion formulas constitute the celebrated Kalman filter for out model. They enable us to update out knowledge of the system each time a new observation comes in.

For convenience we collect these filtering equations one more time

$$v_t = y_t - Z_t a_t \quad F_t = Z_t P_t Z_t^{'} + H_t \quad t = 1, ..., n$$

$$K_t = T_t P_t Z_t^{'} F_t^{-1} \quad L_t = T_t - K_t Z_t$$

$$a_{t+1} = T_t a_t + K_t v_t \quad P_{t+1} = T_t P_t L_t^{'} + R_t Q_t R_t^{'}$$

with $a_1$ and $P_1$ as the mean vector and variance matrix of the initial state vector.

### 4.1.2 State smoothing.

We now consider the estimation of $\alpha_t$ given the entire series $y_1, ..., y_n$. Let us denote the stacked vector $(y_1^{'}, ..., y_n^{'})$ by $y$; thus $y$ is $Y_n$ represented as a vector. We shall estimate $\alpha_t$ by its conditional mean $\hat{\alpha}_t = E(\alpha_t|y)$ and we shall also calculate the error variance matrix $V_t = Var(\alpha_t - \hat{\alpha}_t)$. Our approach is to construct recursion for $\hat{\alpha}_t$ and $V_t$ on the assumption that $\alpha_1 \sim N(a_1, P_1)$ where $a_1$ and $P_1$ are known.

The vector $y$ is fixed when $Y_{t-1}$ and $v_t, ..., v_n$are fixed. We therefore have

$$\hat{\alpha}_t = E(\alpha_t|y) = E(\alpha_t|Y_{t-1}, v_t, ..., v_n) = a_t + \sum_{j=t}^{n} Cov(\alpha_t, v_j) F_j^{-1} v_j \tag{4.13}$$

for $t = 1, ..., n$, with $Cov(\alpha_t, v_j) = E(\alpha_t v_j^{'})$. It follows from () that

$$E(\alpha_t v_j^{'}) = E[\alpha_t (Z_j x_j + \epsilon_j)^{'}] = E(\alpha_t x_j^{'}) Z_j^{'} \tag{4.14}$$

Moreover,

$$E(\alpha_t x_t^{'}) = E[E(\alpha_t x_t^{'}|y)] = E[E\{\alpha_t(\alpha_t - a_t)^{'}|y\}] = P_t \tag{4.15}$$

$$E(\alpha_t x_{t+1}^{'}) = E[E\{\alpha_t(L_t x_t + R_t \eta_t - K_t \epsilon_y)^{'}|y\}] = P_t L_t^{'} \tag{4.16}$$

$$E(\alpha_t x_{t+2}^{'}) = P_t L_t^{'} L_{t+1}^{'} \tag{4.17}$$

$$\vdots$$

$$E(\alpha_t x_n^{'}) = P_t L_t^{'} ... L_{n-1}^{'}$$

Substituting it back gives

$$\hat{\alpha}_n = a_n + P_n Z_n^{'} F_n^{-1} v_n \tag{4.18}$$

$$\hat{\alpha}_{n-1} = a_{n-1} + P_{n-1}Z_{n-1}^{'}F_{n-1}^{-1}v_{n-1} + P_{n-1}L_n^{'}Z_n^{'}F_n^{-1}v_n$$

$$\hat{\alpha}_t = a_t + P_tZ_t^{'}F_t^{-1}v_t + P_tL_t^{'}Z_{t+1}^{'}F_{t+1}^{-1}v_{t+1} + ... + P_tL_t^{'}...L_{n-1}^{'}Z_n^{'}F_n^{-1}v_n$$

for $t = n-2, n-3, ..., 1$. We can express the smoothed state vector as

$$\hat{\alpha}_t = a_t + P_tr_{t-1} \tag{4.19}$$

where $r_{t-1} = Z_n^{'}F_n^{-1}v_n$, $r_{n-2} = Z_{n-1}^{'}F_{n-1}^{-1}v_{n-1} + L_{n-1}^{'}Z_n^{'}F_n^{-1}v_n$ and

$$r_{t-1} = Z_t^{'}F_t^{-1}v_t + L_t^{'}Z_{t+1}^{'}F_{t+1}^{-1}v_{t+1} + ... + L_t^{'}L_{t+1}^{'}...L_{n-1}^{'}Z_n^{'}F_n^{-1}v_n \tag{4.20}$$

or

$$r_{t-1} = Z_t^{'}F_t^{-1}v_t + L_t^{'}r_t \tag{4.21}$$

with $r_n = 0$.

Collecting these results gives the recursion for state smoothing,

$$r_{t-1} = Z_t^{'}F_t^{-1}v_t + L_t^{'}r_t \quad \hat{\alpha}_t = a_t + P_tr_{t-1} \quad t = n, ..., 1$$

with $r_n = 0$.

Alternative algorithms for state smoothing have also been proposed. For example, Anderson and Moore (1979) present the so-called classical fixed interval smoother which for our state space model is given by

$$\hat{\alpha}_t = a_{t|t} + P_{t|t}T_t^{'}P_{t+1}^{-1}(\hat{\alpha}_t - a_{t+1}) \qquad t = n, ..., 1$$

where

$$a_{t|t} = a_t + P_tZ_t^{'}F_t^{-1}v_t \qquad P_{t|t} = P_t - P_tZ_t^{'}F_t^{-1}Z_tP_t$$

A recursion formula for calculating $V_t = Var(\alpha_t|y)$ will now be derived. Using regression we get

$$V_t = Var(\alpha_t|Y_{t-1}, v_t, ..., v_n) = P_t - \sum_{j=t}^{n} Cov(\alpha_t, v_j)F_j^{-1}Cov(\alpha_t, v_j)' \qquad (4.22)$$

Using our previous results we obtain

$$V_t = P_t - P_t N_{t-1} P_t$$

where

$$N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' Z_{t+1}' F_{t+1}^{-1} Z_{t+1} L_t + ... + L_t' ... L_{n-1}' Z_n' F_n^{-1} Z_n L_{n-1} ... L_t$$

Using these results we obtain the recursion formula

$$N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t$$

Collecting together all of the previous results we get

$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t \qquad N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t$$

$$\hat{\alpha}_t = a_t + P_t r_{t-1} \qquad V_t = P_t - P_t N_{t-1} P_t$$

### 4.1.3 Forecasting.

Suppose we have a vector of observations $y_1, ..., y_n$ which follow the state space model and we wish to forecast $y_{n+j}$ for $j = 1, ..., J$. For that purpose let us choose the estimate $\bar{y}_{n+j}$which has minimum mean square error matrix given $Y_n$, that is, $\bar{F}_{n+j} = E[(\bar{y}_{n+j} - y_{n+j})(\bar{y}_{n+j} - y_{n+j})'|y]$ is a minimum in the matrix sense for all estimates of $y_{n+j}$.It follows that the minimum mean square error forecast of $y_{n+j}$ given $Y_n$ is the conditional mean $\bar{y}_{n+1} = E(y_{n+j}|y)$

For $j = 1$ the forecast is straightforward. We have $y_{n+1} = Z_{n+1}\alpha_{n+1} + \epsilon_{n+1}$ so

$$\bar{y}_{n+1} = Z_{n+1}E(\alpha_{n+1}|y) = Z_{n+1}a_{n+1}$$

The error variance matrix or mean square error matrix

$$\bar{F}_{n+1} = E[(\bar{y}_{n+1} - y_{n+1})(\bar{y}_{n+1} - y_{n+1})^{'}] = Z_{n+1}P_{n+1}Z_{n+1}^{'} + H_{n+1}$$

is produced by the Kalman filter relation.We now demonstrate that we can generate the forecasts $\bar{y}_{n+j}$ for $j = 2, ..., J$ merely by treating $y_{n+1}, ..., y_{n+J}$ as missing values. Let $\bar{a}_{n+j} = E(\alpha_{n+j}|y)$ and $\bar{P}_{n+j} = E[(\bar{a}_{n+j} - a_{n+j})(\bar{a}_{n+j} - a_{n+j})^{'}|y]$. Since $y_{n+j} = Z_{n+j}\alpha_{n+j} + \epsilon_{n+j}$ we have

$$\bar{y}_{n+j} = Z_{n+j}E(\alpha_{n+j}|y) = Z_{n+j}\bar{a}_{n+j}$$

with mean square error matrix

$$\bar{F}_{n+j} = Z_{n+j}\bar{P}_{n+j}Z_{n+j}^{'} + H_{n+j}$$

We now derive recursions for calculating $\bar{a}_{n+j}$and $\bar{P}_{n+j}$.We have $\alpha_{n+j+1} = T_{n+j}\alpha_{n+j} + R_{n+j}\eta_{n+j}$ so

$$\bar{a}_{n+j+1} = T_{n+j}\bar{a}_{n+j}$$

for $j = 1, ..., J - 1$ and with $\bar{a}_{n+1} = a_{n+1}$.Also

$$\bar{P}_{n+j+1} = T_{n+j}\bar{P}_{n+j}T_{n+j}^{'} + R_{n+j}Q_{n+j}R_{n+j}^{'}$$

## 4.2 STATE SPACE REPRESENTATION. NON-LINEAR AND NON-GAUSSIAN MODELS.

In this section we are closely follow the tutorial on particle filtering by Doucet and Johansen (2008).

### 4.2.1 General setup.

Consider an $\chi$ - valued discrete-time Markov process $\{X_n\}_{n\geq 1}$ such that

$$X_1 \sim \mu(x_1) \tag{4.23}$$

$$X|(X_{n-1} = x_{n-1}) \sim f(x_n|x_{n-1}) \tag{4.24}$$

All the densities are with respect to a dominating measure. We are interested in estimating $\{X_n\}_{n\geq 1}$ but only have access to the $\{Y_n\}_{n\geq 1}$ . We assume that, given $\{X_n\}_{n\geq 1}$ the observations $\{Y_n\}_{n\geq 1}$ are statistically independent and their marginal densities are given by

$$Y_n|(X_n = x_n) \sim g(y_n|x_n) \tag{4.25}$$

Models compatible with the above description are known as hidden Markov models (HMM) or general state-space models. These equations define a Bayesian model in which (4.23),(4.24) define the prior distribution of the process of interest $\{X_n\}_{n\geq 1}$ and (4.26) defines the likelihood function, that is:

$$p(x_{1:n}) = \mu(x_1)\prod_{k=2}^{n} f(x_k|x_{k-1}) \tag{4.26}$$

and

$$p(y_{1:n}|x_{1:n}) = \prod_{k=1}^{n} g(y_k|x_k) \tag{4.27}$$

In such a context, inference about $X_{1:n}$ given a realization of the observations $Y_{1:n} = y_{1:n}$ relies upon the posterior distribution

$$p(x_{1:n}|y_{1:n}) = \frac{p(x_{1:n}, y_{1:n})}{p(y_{1:n})} \tag{4.28}$$

where

$$p(x_{1:n}, y_{1:n}) = p(x_{1:n})p(y_{1:n}|x_{1:n}) \tag{4.29}$$

$$p(y_{1:n}) = \int p(x_{1:n}, y_{1:n})dx_{1:n} \tag{4.30}$$

However, for most non-linear non-Gaussian models, it is not possible to compute these distributions in closed form. Particle methods are a set of flexible and powerful simulation-based algorithms which provide samples approximately distributed according to posterior distributions of the form $p(x_{1:n}|y_{1:n})$ and facilitate the approximate calculation of $p(y_{1:n})$.

The unnormalized posterior distribution $p(x_{1:n}, y_{1:n})$ satisfies

$$p(x_{1:n}, y_{1:n}) = p(x_{1:n-1}, y_{1:n-1})f(x_n|x_{n-1})g(y_n|x_n) \tag{4.31}$$

Consequently, the posterior $p(x_{1:n}, y_{1:n})$ satisfies the following recursion

$$p(x_{1:n}|y_{1:n}) = p(x_{1:n-1}|y_{1:n-1})\frac{f(x_n|x_{n-1})g(y_n|x_n)}{p(y_n|y_{1:n-1})} \tag{4.32}$$

where

$$p(y_n|y_{1:n-1}) = \int p(x_{n-1}|y_{1:n-1})f(x_n|x_{n-1})g(y_n|x_n)dx_{n-1:n} \tag{4.33}$$

It is straightforward to check that we have

$$p(x_n|y_{1:n}) = \frac{g(y_n|x_n)p(x_n|y_{1:n-1})}{p(y_n|y_{1:n-1})} \tag{4.34}$$

where

$$p(x_n|y_{1:n-1}) = \int f(x_n|x_{n-1})p(x_{n-1}|y_{1:n-1})dx_{n-1} \tag{4.35}$$

Equation (4.35) is known as the prediction step and (4.34) is known as the updating step.

### 4.2.2 Sequential Monte Carlo.

SMC methods are a general class of Monte Carlo methods that sample sequentially from a sequence of target probability densities $\{\pi_n(x_{1:n})\}$ of increasing dimensions. Writing

$$\pi_n(x_{1:n}) = \frac{\gamma_n(x_{1:n})}{Z_n} \tag{4.36}$$

we require only that $\gamma_n : \chi^n \to R^+$ is known pointwise; the normalizing constant

$$Z_n = \int \gamma_n(x_{1:n}) dx_{1:n} \tag{4.37}$$

might be unknown. SMC provides an approximation of $\pi_1(x_1)$ and an estimate of $Z_1$ at time 1 then an approximation of $\pi_2(x_{1:2})$ and an estimate of $Z_2$ and so on.

For example, in the context of filtering, we could have $\gamma_n(x_{1:n}) = p(x_{1:n}, y_{1:n})$, $Z_n = p(y_{1:n})$ so $\pi_n(x_{1:n}) = p(x_{1:n}|y_{1:n})$.

#### 4.2.2.1 Basics of Monte Carlo Methods.

Initially, consider approximating a generic pobability density $\pi_n(x_{1:n})$ for some fixed $n$. If we sample $N$ independent random variables, $X^i_{1:n} \sim \pi_n(x_{1:n})$ for $i = 1, ..., N$ , then the Monte Carlo method approximates $\pi_n(x_{1:n})$ by the empirical measure

$$\hat{\pi}_n(x_{1:n}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{X^i_{1:n}}(x_{1:n}) \tag{4.38}$$

Based on this approximation, it is possible to approximate any marginal, say $\pi_n(x_k)$, easily using

$$\hat{\pi}_n(x_k) = \frac{1}{N} \sum_{i=1}^{N} \delta_{X^i_k}(x_k)$$

and the expectation of any test function given by

$$I_n(\phi_n) := \int \phi_n(x_{1:n}) \pi_n(x_{1:n}) dx_{1:n}$$

is estimated by

$$I_n^{MC}(\phi_n) := \frac{1}{N}\sum_{i=1}^{N}\phi_n(X_{1:n}^i)$$

**Problem.**

If $\pi_n(x_{1:n})$ is a complex high-dimensional probability distribution, then we cannot sample from it.

**4.2.2.2 Importance Sampling.** This is a fundamental Monte Carlo method and the basis of all algorithms developed later on. IS relies on the introduction of an importance density $q_n(x_{1:n})$ such that

$$\pi_n(x_{1:n}) > 0 \Rightarrow q_n(x_{1:n}) > 0$$

In this case we have the following IS identities

$$\pi_n(x_{1:n}) = \frac{\omega_n(x_{1:n})q_n(x_{1:n})}{Z_n} \tag{4.39}$$

$$Z_n = \int \omega_n(x_{1:n})q_n(x_{1:n})dx_{1:n} \tag{4.40}$$

where $\omega_n(x_{1:n})$ is the unnormalized weight function

$$\omega_n(x_{1:n}) = \frac{\gamma_n(x_{1:n})}{q_n(x_{1:n})}$$

In particular, we can select an importance density $q_n(x_{1:n})$ from which it is easy to draw samples; e.g. a multivariate Gaussian. Assume we draw $N$ independent samples $X_{1:n}^i \sim q_n(x_{1:n})$ then by inserting the Monte Carlo approximation of $q_n(x_{1:n})$ - that is the empirical measure of the samples $X_{1:n}^i$ - into (4.39) and (4.40) we obtain

$$\hat{\pi}_n(x_{1:n}) = \sum_{i=1}^{N} W_n^i \delta_{X_{1:n}^i}(x_{1:n}) \tag{4.41}$$

$$\hat{Z}_n = \frac{1}{N}\sum_{i=1}^{N}\omega_n(X_{1:n}^i) \tag{4.42}$$

where

$$W_n^i = \frac{\omega_n(X_{1:n}^i)}{\sum_{j=1}^N \omega_n(X_{1:n}^i)} \tag{4.43}$$

If we are interested in computing $I_n(\phi_n)$ , we can also use the estimate

$$I_n^{IS}(\phi_n) = \sum_{i=1}^N W_n^i \phi_n(X_{1:n}^i)$$

Unlike $I_n^{MC}(\phi_n)$, this estimate is biased for finite $N$.

For a given test function, $\phi_n(x_{1:n})$, it is easy to establish the importance distribution minimizing the asymptotic variance of $I_n^{IS}(\phi_n)$. However, such a result is of minimal interest in a filtering context as this distribution depends on $\phi_n(x_{1:n})$ and we are typically interested in the expectations of several test functions. Moreover, even of we were interested in a single test function, say $\phi_n(x_{1:n}) = x_n$, then selecting the optimal importance distribution at time $n$ would have detrimental effects when will try to obtain a sequential version of the algorithms.

A more appropriate approach in this context is to attempt to select the $q_n(x_{1:n})$ which minimizes the variance of the importance weights. Clearly, this variance is minimized for $q_n(x_{1:n}) = \pi_n(x_{1:n})$. We cannot select $q_n(x_{1:n}) = \pi_n(x_{1:n})$ as this is the reason we used US in the first place. However, this simple result indicates that we should aim at selecting an IS distribution which is as close as possible to the target.

**4.2.2.3 Sequential importance Sampling.** We are now going to present an algorithm that admits a fixed computational complexity at each time step in important scenarios. Thus solution involves selecting an importance distribution which has the following structure

$$q_n(x_{1:n}) = q_{n-1}(x_{1:n-1})q_n(x_n|x_{n-1}) = q_1(x_1)\prod_{k=2}^n q_k(x_k|x_{1:k-1}) \tag{4.44}$$

Practically, this means that to obtain particles $X_{1:n}^i \sim q_n(x_{1:n})$ at time $n$, we sample $X_1^i \sim q_1(x_1)$ at time 1 then $X_k^i \sim q_k(x_k|X_{1:k-1}^i)$ at time $k$ for $k = 2, ..., n$. The associated unnormalized weights can be computed recursively using the decomposition

$$\omega_n(x_{1:n}) = \frac{\gamma_n(x_{1:n})}{q_n(x_{1:n})} = \frac{\gamma_{n-1}(x_{n-1})}{q_{n-1}(x_{1:n-1})}\frac{\gamma_n(x_{1:n})}{\gamma_{n-1}(x_{1:n-1})q_n(x_n|x_{1:n-1})} \tag{4.45}$$

which can be written in the form

$$\omega_n(x_{1:n}) = \omega_{n-1}(x_{1:n-1})\alpha_n(x_{1:n}) = \omega_1(x_1)\prod_{k=2}^{n}\alpha_k(x_{1:k})$$

where the incremental importance weight function $\alpha_n(x_{1:n})$ is given by

$$\alpha_n(x_{1:n}) = \frac{\gamma_n(x_{1:n})}{\gamma_{n-1}(x_{1:n-1})q_n(x_n|x_{1:n-1})} \tag{4.46}$$

**4.2.2.4 Resampling.** Resampling is a very intuitive idea which has major practical and theoretical benefits. Consider first an IS approximation $\hat{\pi}_n(x_{1:n})$ of the target distribution $\pi_n(x_{1:n})$. This approximation is based ion weighted samples from $q_n(x_{1:n})$ and does not provide samples approximately distributed according to $\pi_n(x_{1:n})$. To obtain approximate samples from $\pi_n(x_{1:n})$ , we can simply sample from its IS approximation $\hat{\pi}_n(x_{1:n})$ ; that is we select $X^i_{1:n}$ with probability $W^i_n$. This operation is called resampling as it corresponds to sampling from an approximation $\hat{\pi}_n(x_{1:n})$ which was itself obtained by sampling. If we are interested in obtaining $N$ samples from $\hat{\pi}_n(x_{1:n})$ , then we can simply resample $N$ times from $\hat{\pi}_n(x_{1:n})$. This is equivalent to associating a number of offsprings $N^i_n$ with each particle $X^i_{1:n}$ in such a way that $N^{1:N}_n = (N^1_n, ..., N^N_n)$ follow a multinomial distribution with parameter vector $(N, W^{1:N}_n)$ and associating a weight of $1/N$ with each offspring. We approximate $\hat{\pi}_n(x_{1:n})$ by the resampled empirical measure

$$\bar{\pi}_n(x_{1:n}) = \sum_{i=1}^{N}\frac{N^i_n}{N}\delta_{X_{1:n}}(x_{1:n}) \tag{4.47}$$

where $E[N^i_n|W^{1:N}_n] = NW^i_n$. Hence $\bar{\pi}_n(x_{1:n})$ is an unbiased approximation of $\hat{\pi}_n(x_{1:n})$ .

Improved unbiased resampling schemes have been proposed in the literature. These are methods of selecting $N^i_n$ such that the unbiasedness property is preserved, and such that $E[N^i_n|W^{1:N}_n]$ is smaller than that obtained via the multinomial resampling scheme described above. The three most popular algorithms are presented below:

1. Systematic resampling. Sample $U_1 \sim U[0, \frac{1}{N}]$ and define $U_i = U_1 + \frac{i-1}{N}$ for $i = 2, ..., N$, then set $N^i_n = |\left\{U_j : \sum_{k=1}^{i-1} W^k_n \leq U_j \leq \sum_{k=1}^{i} W^k_n\right\}|$ with the convention $\sum_{k=1}^{0} := 0$.

2. Residual resampling. Set $\tilde{N}_n^i = [NW_n^i]$, sample $\bar{N}_n^{1:N}$from a multinomial of parameters $(N, \bar{W}_n^{1:N})$ where $\bar{W}_n^i \propto W_n^i - N^{-1}\tilde{N}_n^i$, then set $N_n^i = \tilde{N}_n^i + \bar{N}_n^i$.
3. Multinomial resampling. Sample $N_n^{1:N}$from a multinomial of parameters $(N, lW_n^{1:N})$.

Resampling allows us to obtain samples distributed approximately according to $\pi_n(x_{1:n})$, but it should be clear that if we are interested in estimating $I_n(\phi_n)$ then we will obtain an estimate with lower variance using $\hat{\pi}_n(x_{1:n})$ that that which we would have obtained by using $\bar{\pi}_n(x_{1:n})$ . By resampling we indeed add some extra noise. However, an important advantage of resampling is that it allows us to remove particles with low weights.

**4.2.2.5 A generic SMC algorithm.** SMC methods are a combination of SIS and resampling. At time 1, we compute the IS approximation $\hat{\pi}_1(x_1)$ of $\pi_1(x_1)$ which is weighted collection of particles $\{W_1^i, X_1^i\}$. Then we use a resampling step to eliminate those particles with low weights and multiply those with high weights. We denote by $\{\frac{1}{N}, \bar{X}_1^i\}$ the collection of equally-weighted resampled particles. Remember that each original particle $X_1^i$ has $N_1^i$ offsprings so there exists $N_1^i$distinct indices $j_1 \neq j_2 \neq ... \neq j_{N_1^i}$ such that $\bar{X}_1^{j_1} = \bar{X}_1^{j_2} = ... = \bar{X}_1^{j_{N_1^i}} = X_1^i$. After the resampling step, we follow the SIS strategy and sample $X_2^i \sim q_2(x_2|\bar{X}_1^i)$.Thus $(\bar{X}_1^i, X_2^i)$ is approximately distributed according to $\pi_1(x_1)q_2(x_2|x_1)$. Hence the corresponding importance weights in this case are simply equal to the incremental weights $\alpha_2(x_{1:2})$. We then resample the particles with respect to the normalized weights and so on.

At any time $n$, this algorithm provides two approximations of $\pi_n(x_{1:n})$. we obtain

$$\hat{\pi}_n(x_{1:n}) = \sum_{i=1}^{N} W_n^i \delta_{X_{1:n}^i}(x_{1:n}) \tag{4.48}$$

after sampling

$$\bar{\pi}_n(x_{1:n}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\bar{X}_{1:n}^i}(x_{1:n}) \tag{4.49}$$

after the resampling step.

As we have already mentioned, resampling has the effect of removing particles with low weights and multiplying particles with high weights. However, this is at the cost of immediately introducing some additional variance. If particles have unnormalized weights with a small variance then the resampling step might be unnecessary. Consequently, in practice, it is more sensible to resample only when the variance of the unnormalized weights is superior to a pre-specified threshold. This is often assessed by looking at the variability of the weights using the so-called Effective Sample Size (ESS) criterion which is given by

$$ESS = \left( \sum_{i=1}^{N} (W_n^i)^2 \right)^{-1} \tag{4.50}$$

### 4.2.3 Particle Filtering.

Remember that is the filtering context, we want to be able to compute a numerical approximation of the distribution $\{p(x_{1:n}|y_{1:n})\}_{n\geq 1}$ sequentially in time. A direct application of the SMC methods described earlier to the sequence of target distributions $\pi_n(x_{1:n}) = p(x_{1:n}|y_{1:n})$ yields a popular class of particle filters. More elaborate sequences of target and proposal distributions yield various more advanced algorithms.

**4.2.3.1 SMC for filtering.** First, consider the simplest case in which the joint density $\gamma_n(x_{1:n}) = p(x_{1:n}, y_{1:n})$ is chosen, yielding $\pi_n(x_{1:n}) = p(x_{1:n}|y_{1:n})$ and $Z_n = p(y_{1:n})$. Practically, it is only necessary to select the importance distribution $q_n(x_n|x_{1:n-1})$. We have seen that in order to minimize the variance of the importance weights at time $n$, we should select $q_n^{opt}(x_n|x_{1:n-1}) = \pi_n(x_n|x_{1:n-1})$ where

$$\pi_n(x_n|x_{1:n-1}) = p(x_n|y_n, x_{n-1}) = \frac{g(y_n|x_n) f(x_n|x_{n-1})}{p(y_n|x_{n-1})} \tag{4.51}$$

and the associated incremental weight is $\alpha_n(x_{1:n}) = p(y_n|x_{n-1})$. In many scenarios, it is not possible to sample from this distribution but we should aim to approximate it. In any case, it shows that we should use an importance distribution of the form

$$q_n(x_n|x_{1:n-1}) = q(x_n|y_n, x_{n-1}) \tag{4.52}$$

and that there is nothing to be gained from building importance distribution depending also upon $(y_{1:n-1}, x_{1:n-2})$ -although, at least in principle, in some settings there may be advantages to using information from subsequent observations if they are available. Incremental weight is given then by

$$\alpha_n(x_{1:n}) = \alpha_n(x_{n-1:n}) = \frac{g(y_n|x_n)f(x_n|x_{n-1})}{q(x_n|y_n, x_{n-1})}$$

We obtain at time $n$

$$\hat{p}(x_{1:n}|y_{1:n}) = \sum_{i=1}^{N} W_n^i \delta_{X_{1:n}^i}(x_{1:n})$$

$$\hat{p}(y_n|y_{1:n-1}) = \sum_{i=1}^{N} W_{n-1}^i \alpha_n(X_{n-1:n}^i)$$

Many techniques have been proposed to design importance distributions $q(x_n|y_n, x_{n-1})$ which approximate $p(x_n|y_n, x_{n-1})$. In particular the use of standard suboptimal filtering techniques such as the Extended Kalman Filter or the Unscented Kalman Filter to obtain importance distributions is very popular in the literature. The use of local optimization techniques to design $q(x_n|y_n, x_{n-1})$ centered around the mode of $p(x_n|y_n, x_{n-1})$ has also been advocated.

### 4.2.4 Auxiliary Particle Filtering.

As was discussed above, the optimal proposal distribution when performing standard particle filtering is $q(x_n|y_n, x_{n-1}) = p(x_n|y_n, x_{n-1})$. Indeed, $\alpha(x_{n-1:n})$ is independent of $x_n$ in this case so it is possible to interchange the order of the sampling and resampling steps. Intuitively, this yields a better approximation of the distribution as it provides a greater number of distinct particles to approximate the target. This is an example of a general principle: resampling, if it is to be applied in a particular iteration, should be performed before, rather than after, any operation that does not influence the importance weights in order to minimize the loss of information.

It is clear that if importance weights are independent of the new state and the proposal distribution corresponds to the marginal distribution of the proposed states then weighting, resampling and then sampling corresponds to a reweighing to correct for the discrepancy between the old and new marginal distribution of the earlier states, resampling to produce an unweighted sample and then generation of the new state from its conditional distribution. This intuition can easily be formalized.

However, in general, the incremental importance weights do depend upon the new states and this straightforward change of order becomes impossible. In a sense, this interchange of sampling and resampling produces an algorithm in which information from the next observation is used to determine which particles should survive resampling at a given time. It is desirable to find methods for making use of this future information in a more general setting, so that we can obtain the same advantage in situations in which it is not possible to make use of the optimal proposal distribution.

The Auxiliary Particle Filter (APF) is an alternative algorithm which does essentially this.It can be shown that the APF can be interpreted as a standard SMC algorithm applied to the following sequence of target distributions

$$\gamma_n(x_{1:n}) = p(x_{1:n}, y_{1:n})\tilde{p}(y_{n+1}|x_n) \tag{4.53}$$

with $\tilde{p}(y_{n+1}|x_n)$ chosen as an approximation of the predictive likelihood $p(y_{n+1}|x_n)$ if it is not known analytically. It follows that $\pi_n(x_{1:n})$ is an approximation of $p(x_{1:n}|y_{1:n+1})$ denoted $\tilde{p}(x_{1:n}|y_{1:n+1})$ given by

$$\pi_n(x_{1:n}) = \tilde{p}(x_{1:n}|y_{1:n+1}) \propto p(x_{1:n}|y_{1:n})\tilde{p}(y_{n+1}|x_n) \tag{4.54}$$

In the APF we also use an importance distribution $q_n(x_n|x_{1:n-1})$ of the form (4.52) which is typically an approximation of (4.51) . Note that (4.51) is different from $\pi_n(x_n|x_{1:n-1})$ in this scenario. Even if we could sample from $\pi_n(x_n|x_{1:n-1})$, one should remember that in this case the object of inference is not $\pi_n(x_{1:n}) = \tilde{p}(x_{1:n}|y_{1:n+1})$ but $p(x_{1:n}|y_{1:n})$. The associated incremental weight is given by

$$\alpha_n(x_{n-1:n}) = \frac{\gamma_n(x_{1:n})}{\gamma_{n-1}(x_{1:n-1})q_n(x_n|x_{1:n-1})} = \frac{g(y_n|x_n)f(x_n|x_{n-1})\tilde{p}(y_{n+1}|x_n)}{\tilde{p}(y_n|x_{n-1})q(x_n|y_n, x_{n-1})} \tag{4.55}$$

Keeping in mind that this algorithm does not approximate the distributions $\{p(x_{1:n}|y_{1:n})\}$ but the distributions $\{\tilde{p}(x_{1:n}|y_{1:n+1})\}$, we use IS to obtain an approximation of $p(x_{1:n}|y_{1:n})$ with

$$\pi_{n-1}(x_{1:n-1})q_n(x_n|x_{1:n-1}) = \tilde{p}(x_{1:n-1}|y_{1:n})q(x_n|y_n, x_{n-1}) \tag{4.56}$$

as the importance distribution. A Monte Carlo approximation of this importance distribution is obtained after the sampling step in the APF and the associated unnormalized weights are given by

$$\tilde{\omega}_n(x_{n-1:n}) = \frac{p(x_{1:n}, y_{1:n})}{\gamma_{n-1}(x_{1:n-1})q_n(x_n|x_{1:n-1})} = \frac{g(y_n|x_n)f(x_n|x_{n-1})}{\tilde{p}(y_n|x_{n-1})q(x_n|y_n, x_{n-1})} \tag{4.57}$$

It follows that we obtain

$$\hat{p}(x_{1:n}|y_{1:n}) = \sum_{i=1}^{N} \tilde{W}_n^i \delta_{X_{1:n}^i}(x_{1:n}) \tag{4.58}$$

$$\hat{p}(y_{1:n}) = \frac{1}{N}\sum_{i=1}^{N} \tilde{\omega}_n(X_{n-1:n}^i) \tag{4.59}$$

where

$$\tilde{W}_n^i \propto \tilde{\omega}_n(X_{n-1:n}^i)$$

or $\tilde{W}_n^i \propto W_{n-1:n}^i \tilde{\omega}_n(X_{n-1:n}^i)$ if resampling was not performed at the end of the previous iteration. Selecting $q_n(x_n|x_{1:n-1}) = p(x_n|y_n, x_{n-1})$ and $\tilde{p}(y_n|x_{n-1}) = p(y_n|n-1)$, when it is possible to do so, leads to so-called "perfect adaption" case. In this case, the APF takes a particularly simple form as $\alpha_n(x_{n-1:n}) = p(y_n|x_{n-1})$ and $\tilde{\omega}_n(x_{n-1:n}) = 1$. This is similar to the algorithm discusses in the previous subsection where the order of the sampling and resampling steps are interchanged.

This simple reinterpretation of the APF shows us several things:

- We should select a distribution $\tilde{p}(x_{1:n}|y_{1:n})$ with thicker tails than $p(x_{1:n}|y_{1:n})$.
- Setting $\tilde{p}(y_n|x_{n-1}) = g(y_n|\mu(x_{n-1}))$ where $\mu$ denotes some point estimate is perhaps unwise as this will not generally satisfy that requirement.

- We should use an approximation of the predictive likelihood which is compatible with the model we are using in the sense that it encodes at least the same degree of uncertainty as the exact model.

### 4.2.5 Limitation of Particle Filters.

The algorithms described earlier suffer from several limitations. It is important to emphasis at this point that, even if the optimal importance distribution $p(x_n|y_n, x_{n-1})$ can be used, this does not guarantee that the SMC algorithms will be efficient. Indeed, if the variance of $p(y_n|x_{n-1})$ is high, then the variance of the resulting approximation will be high. Consequently, it will be necessary to resample very frequently and the particle approximation $\hat{p}(x_{1:n}|y_{1:n})$ of the joint distribution $p(x_{1:n}|y_{1:n})$ will be unreliable. In particular, for $k \ll n$ the marginal distribution $\hat{p}(x_{1:k}|y_{1:n})$ will only by approximated by a few if not a single unique particle because the algorithm will have resampled many times between times $k$ and$n$. One major problem with the approaches discussed above is that only the variables $\{X_n^i\}$ are sampled at time $n$ but the path values $\left\{X_{1:n-1}^i\right\}$remain fixed. An obvious way to improve upon these algorithms would involve not only sampling $\{X_n^i\}$ at time $n$, but also modifying the values of the paths over a fixed lag $\left\{X_{n-L+1:n-1}^i\right\}$ for $L > 1$ in light of the new observation $y_n$; $L$ being fixed or upper bounded to ensure that we have a sequential algorithm. These limitation can be overcome using SMC filtering with MCMC moves or SMC Block Sampling for Filtering. We do not describe these algorithms in the current work.

### 4.2.6 Smoothing.

One problem, which is closely related to filtering, but computationally more challenging for reason which will be apparent later, is known as smoothing. Whereas filtering corresponds estimating the distribution of the current state based upon the observations received up until the current time, smoothing corresponds to estimating the distribution of the state at a particular time given all of the observations up to some later time. The trajectory estimates obtained by such methods, as a result of the additional information available, tend to be smoother than those obtained by filtering. It is intuitive that if estimates of the state at time

$n$ are not required instantly, then better estimation performance is likely to be obtained by taking advantage of a few later observations. Designing efficient sequential algorithms for the solution of this problem is not quite a straightforward as it might seem, but a number of effective strategies have been developed and are described below.

More formally: assume that we have access to the data $y_{1:T}$, and we wish to compute the marginal distribution $\{p(x_n|y_{1:T})\}$where $n = 1, ..., T$ or to sample from $p(x_{1:T}|y_{1:T})$. In principle, the marginals $\{p(x_n|y_{1:T})\}$ could be obtained directly by considering the joint distribution $p(x_{1:T}|y_{1:T})$ and integrating out the variables $(x_{1:n-1}, x_{n+1:T})$ . Extending this reasoning in the context of particle methods, one can simply use the identity $p(x_n|y_{1:T}) = \int p(x_{1:T}|y_{1:T})dx_{1:n-1}dx_{n+1:T}$ and take the same approach which is used in particle filtering: use Monte Carlo algorithms to obtain an approximate characterization of the joint distribution and then use the associated marginal distribution to approximate the distributions of interest. Unfortunately, as is detained below, when $n \ll T$ this strategy id doomed to failure: the marginal distribution $p(x_n|y_{1:n})$ occupies a privileged role within the particle filter framework as it is, in some sense, better characterized than any of the other marginal distributions.

For this reason, it is necessary to develop more sophisticated strategies in order to obtain good smoothing algorithms.There has been much progress in this direction over the past decade. Below, we present two alternative recursions that will prove useful when numerical approximations are required. The key to the success of these recursions is that they rely upon only the marginal filtering distributions $\{p(x_n|y_{1:n})\}$.

**4.2.6.1 Forward-Backward Recursions.** The following decomposition of the joint distribution $p(x_{1:T}|y_{1:T})$

$$p(x_{1:T}|y_{1:T}) = p(x_T|y_{1:T}) \prod_{n=1}^{T-1} p(x_n|x_{n+1}, y_{1:T}) = p(x_T|y_{1:T}) \prod_{n=1}^{T-1} p(x_n|x_{n+1}, y_{1:n}) \qquad (4.60)$$

shows that, conditional on $y_{1:T}$, $\{X_n\}$ is an inhomogeneous Markov process.

Equation (4.60) suggests the following algorithm to sample from $p(x_{1:T}|y_{1:T})$. First compute and store the marginal distributions $\{p(x_n|y_{1:n})\}$ for $n = 1, ..., T$. Then sample $X_T \sim p(x_T|y_{1:n})$ and for $n = T-1, T-2, .., 1$, sample $X_n \sim p(x_n|X_{n+1}, y_{1:n})$ where

$$p(x_n|x_{n+1}, y_{1:n}) = \frac{f(x_{n+1}|x_n)p(x_n|y_{1:n})}{p(x_{n+1}|y_{1:n})} \tag{4.61}$$

It also follows, by integrating out $(x_{1:n-1}, x_{n+1:T})$ in equation (4.60), that

$$p(x_n|y_{1:T}) = p(x_n|y_{1:n}) \int \frac{f(x_{n+1}|x_n)}{p(x_{n+1}|y_{1:n})} p(x_{n+1}|y_{1:T})dx_{n+1} \tag{4.62}$$

So to compute $\{p(x_n|y_{1:T})\}$ ,we simply modify the backward pass and,instead of sampling from $p(x_n|x_{n+1}, y_{1:n})$, we compute $p(x_n|y_{1:T})$ from (4.62).

#### 4.2.6.2 Forward Filtering-Backward Smoothing.

It is possible to obtain an SMC approximation of the forward filering-backward sampling procedure directly by noting that for

$$\hat{p}(x_n|y_{1:n}) = \sum_{i=1}^{N} W_n^i \delta_{X_n^i}(x_n) \tag{4.63}$$

we have

$$\hat{p}(x_n|X_{n+1}, y_{1:n}) = \frac{f(X_{n+1}|x_n)\hat{p}(x_n|y_{1:n})}{\int f(X_{n+1}|x_n)\hat{p}(x_n|y_{1:n})dx_n} = \sum_{i=1}^{N} \frac{W_n^i f(X_{n+1}|X_n^i)\delta_{X_n^i}(x_n)}{\sum_{j=1}^{N} W_n^j f(X_{n+1}|X_n^j)} \tag{4.64}$$

Consequently, the following algorithm generates a sample approximately distributed according to $p(x_{1:T}|y_{1:T})$: first sample $X_T \sim \hat{p}(x_T|y_{1:T})$ and for $n = T-1, T-2, ..., 1$, sample $X_n \sim \hat{p}(x_n|X_{n+1}, y_{1:n})$.

Similarly, we can also provide an SMC approximation of the forward filtering-backward smoothing procedure by direct means. If we denote by

$$\hat{p}(x_n|y_{1:T}) = \sum_{i=1}^{N} W_{n|T}^i \delta_{X_n^i}(x_n) \tag{4.65}$$

the particle approximation of $p(x_n|y_{1:T})$ then, by inserting (4.65) in (4.62), we obtain

$$\hat{p}(x_n|y_{1:T}) = \sum_{i=1}^{N} W_n^i \left[ \sum_{j=1}^{N} W_{n+1|T}^j \frac{f(X_{n+1}^j|X_n^i)}{\sum_{l=1}^{N} W_n^l f(X_{n+1}^j|X_n^l)} \right] \delta_{X_n^i}(x_n) := \sum_{i=1}^{N} W_{n|T}^i \delta_{X_n^i}(x_n) \quad (4.66)$$

## 4.3 APPLICATION TO SV MODELS.

### 4.3.1 Particle filter with SV model.

Recall the state-space model formulation. It consists of two equations: the observation equation and the transition equation which are given by

$$y_n = m_n(x_n, \epsilon_n) \quad (4.67)$$

$$x_n = h_n(x_{n-1}, \eta_n) \quad (4.68)$$

It is assumed that the distributions of the observations and state variables admit density functions with respect to appropriate dominating measures. These densities $p(y_n|x_n;\theta)$ and $p(x_n|x_{n-1};\theta)$ correspond to (4.67) and (4.68) respectively.

Here we call

$$\tilde{\omega}_n = \frac{p(y_n|x_n;\theta)p(x_n|x_{n-1};\theta)}{g_n(x_n|x_{n-1}, y_n;\psi)} \quad (4.69)$$

the incremental weights and $g_n(x_n|x_{n-1}, y_n;\psi)$ is the importance density.

Using the ideas described in the previous paragraphs we can write a particle filter for the SV model

$$y_n = e^{x_n/2}\epsilon_n \quad (4.70)$$

$$x_n = \mu + \phi(x_{n-1} - \mu) + \sigma_\eta \eta_n \quad (4.71)$$

where $\epsilon_n \sim N(0,1)$ and $\eta_n \sim N(0,1)$. We generate an SV model with parameters chosen as $\mu = 0.5$, $\phi = 0.985$, and $\sigma_\eta^2 = 0.04$ and try to use SISR and auxiliary particle filters with this model. To implement the SISR algorithm we select the conditional proposal distribution at each iteration to be the transition density $g_n(x_n|x_{0:n-1}, y_{1:n}; \psi) = p(x_n|x_{n-1}; \theta)$ implied by the dynamics of the model. This means that the incremental weight function is equal to the measurement density $\tilde{\omega}_n = p(y_n|x_n; \theta)$. We use multinomial resampling at each iteration.

We also show the application of the auxiliary particle filter. It is a popular algorithm that is simple to implement and works well in many cases. When proposing new particles at the beginning of each iteration, we would like to use information available in the current observation $y_n$. One calls particles filters that incorporate $y_n$ into their proposal adapted particle filters. In addition, since particles carried over from last period form part of this period's proposal distribution, some of the old particles provide more information about $x_n$ than others.

Pitt and Shepard (1999, 2001) approximate the incremental target distribution in

$$p(x_{0:n}|y_{1:n}) = \frac{p(y_n|x_n;\theta)p(x_n|x_{n-1};\theta)}{p(y_n|y_{1:n-1};\theta)} p(x_{0:n-1}|y_{1:n-1};\theta) \tag{4.72}$$

by

$$p(y_n|x_n;\theta)p(x_n|x_{n-1};\theta) \approx g_{1,n}(y_n|x_n;\psi)g_{2,n}(x_n|x_{n-1};\psi) = \tag{4.73}$$

$$= g_{1,n}(y_n|x_{n-1};\psi)g_{2,n}(x_n|x_{n-1};\psi, y_n)$$

This proposal distribution is decomposed into two parts implying that the sampling of new values $\left\{x_n^{(i)}\right\}_{i=1}^{N}$ from this distribution can be performed in two steps.

Many economic models have a special structure with non-Gaussian measurement densities and linear, Gaussian transition densities. In this case if the measurement density is log-concave, Pitt and Shephard suggest taking $g_{1,n}(y_n|x_{n-1};\theta)$ to be the Taylor series expansion of $\log(p(y_n|x_n;\theta))$ around a point $\mu_n$ and combining it with the transition density $g_{2,n}(x_n|x_{n-1};\theta, y_n) = p(x_n|x_{n-1};\theta, y_n)$.

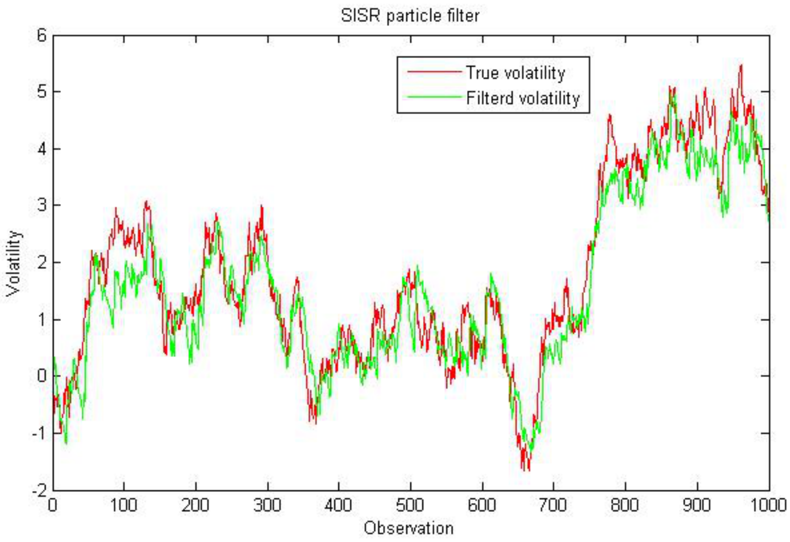

Figure 4.1: SISR filter for SV model

In the settings where one can evaluate $p(y_n|x_{n-1};\theta)$,one can select $g_{1,n}(y_n|x_{n-1};\psi) = p(y_n|x_{n-1};\theta)$ and $g_{2,n}(x_n|x_{n-1};\theta,y_n) = p(x_n|x_{n-1};\theta,y_n)$.

We apply this algorithm to our SV model.

The results are given in the chart below.

**4.3.1.1 Likelihood-based parameter estimation.** The two major issues to consider are computing the maximum likelihood estimator in a computationally efficient way ad its statistical properties once it is computed. Although the particle filter's approximation of the likelihood function at a point $\theta$ is consistent asymptotically in the number of particles, the log-likelihood function is nota continuous function of the parameters. The log-likelihood function is given by

$$\log L(\theta|y_{1:T}) = \log p(y_1,..,y_T;\theta) = \sum_{n=1}^{T} \log p(y_n|y_{1:n-1};\theta) \approx \sum_{n=1}^{T} \log\left[\sum_{i=1}^{N} \omega_{n-1}^{(i)}\tilde{\omega}_n^{(i)}\right] \quad (4.74)$$

This discontinuity is created from the resampling stage within a particle filter and can cause problems for gradient-based optimizers.

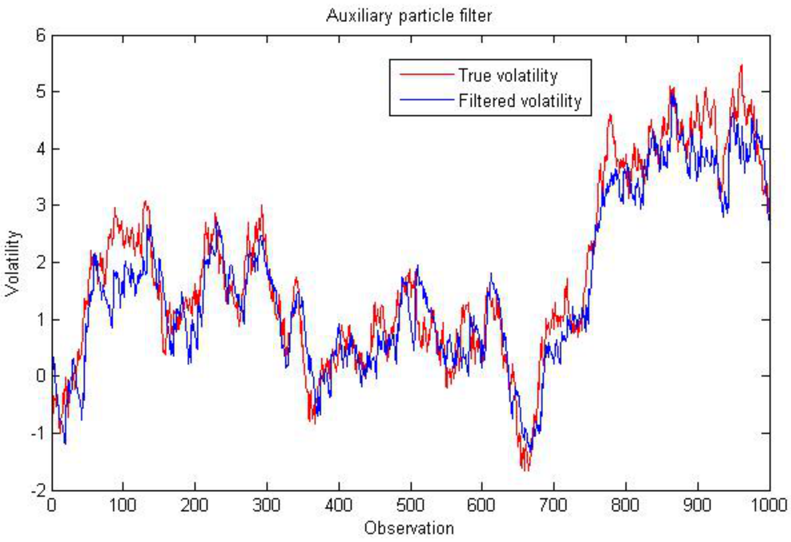

Figure 4.2: Auxiliary filter for SV model

Most of the work on ML estimation using particle filters has focused on using approaches other than gradient-based optimization that avoids the discontinuity problem. These methods include stochastic gradient-based methods, recursive maximum likelihood methods and EM methods.

The main difficulty when using this method is the right choice of the gain sequence. Parameters should also be of the same order of magnitude.

### 4.3.2 SV estimation by the efficient method of moments.

A stochastic volatility model in its basic discrete-time format reads:

$$y_t = \sigma_t \epsilon_t \tag{4.75}$$

$$\ln \sigma_t^2 = \omega + \gamma \ln \sigma_{t-1}^2 + \sigma_\eta \eta_t \tag{4.76}$$

where $\epsilon_t, \eta_t \sim N(0,1)$. This model has served as the benchmark and starting point of the bulk of the econometric literature on stochastic volatility models. This model is in discrete form. Many variations of this model are possible. Here we consider a broad range of models as in van der Sluis (1997), namely:

$$y_t = \sigma_t \epsilon_t \tag{4.77}$$

$$\ln \sigma_t^2 = \omega + \sum_{i=1}^{p} \gamma_i L^i \ln \sigma_t^2 + \sigma_\eta \left( 1 + \sum_{j=1}^{q} \xi_j L^j \right) \eta_t \tag{4.78}$$

$$\begin{bmatrix} \epsilon_t \\ \eta_{t+1} \end{bmatrix} \sim NIID \left( 0, \begin{bmatrix} 1 & \lambda \\ \lambda & 1 \end{bmatrix} \right) \tag{4.79}$$

where $-1 \leq \lambda \leq 1$.

**4.3.2.1 Efficient method of Moments.** Gallant and Tauchen (1996) solve the efficiency problems that moment-based techniques generally have by proposing the efficient method of moments (EMM) technique. The structural model is estimated by using an auxiliary model. The connection between the auxiliary model and the structural model is achieved by means of scores of the auxiliary model, where strict guidelines are given for the choice of the auxiliary model such that maximum likelihood efficiency is attained. In short the EMM methods is as follows. The sequence of densities for the structural model is denoted:

$$\{p_1(x_1|\theta), \{p_t(y_t|x_t, \theta)\}_{t=1}^{\infty}\} \tag{4.80}$$

The sequence of densities for the auxiliary process is denoted as:

$$\{f_1(\omega_1|\beta), \{f_t(y_t|\omega_t, \beta)\}_{t=1}^{\infty}\} \tag{4.81}$$

where $x_t$and $\omega_t$ are observable endogenous variables. In particular, the $x_t$ will be a vector of lagged $y_t$, and the$\omega_t$will also be a vector of lagged $y_t$. We impose Assumptions 1 and 2 in Gallant and Long (1997) on the structural model. These are technical assumptions that imply standard properties of quasi-likelihood estimator and properties of of estimators based on Hermite expansions, which will be explained below. It is important that the structural model is stationary and ergodic. Define

$$m(\theta,\beta)=\int\int\frac{\partial}{\partial\beta}\ln f(y|\omega,\beta)p(y|x,\theta)p(x|\theta)dydx \tag{4.82}$$

which is the expected value score of the auxiliary model under dynamic model. This integral can be approximated by MC techniques:

$$m_N(\theta,\beta)=\frac{1}{N}\sum_{\tau=1}^{N}\ln f(y_\tau(\theta)|\omega_\tau(\theta),\beta) \tag{4.83}$$

where $y_\tau(\theta)$ are drawings from the structural model. Let $n$ denote the sample size. The EMM estimator is defined as:

$$\hat{\theta}_n(I_n):=\arg\min_{\theta\in\Theta} m_N^{'}(\theta,\hat{\beta}_n)\,(I_n)^{-1}\,m_N(\theta,\hat{\beta}_n)$$

where $I_n$ is a weighting matrix and $\hat{\beta}_n$ denotes an estimator for the parameter of the auxiliary model. The optimal weighting matrix here is

$$I_0=\lim_{n\to\infty}V_0\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\frac{\partial}{\partial\beta}\ln f_t(y_t|\omega_t,\beta^*)\right\}\right]$$

where $\beta^*$ is a (pseudo) true value. The small sample pendant is:

$$I_n=V_0\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\frac{\partial}{\partial\beta}\ln f_t(\tilde{y}_t|\tilde{\omega}_t,\hat{\beta}_n)\right\}\right]$$

The auxiliary model is built as follows. The process $y_t(\theta)$ is the process under investigation, $\mu_t(\beta^*)=E_{t-1}[y_t(\theta)]$ is the conditional mean of the auxiliary model, $\sigma_t^2(\beta^*)=Var_{t-1}(y_t(\theta_0)-\mu_t(\beta^*))$ is the conditional variance, and $z_t(\beta^*)=[y_t(\theta_0)-\mu_t(\beta^*)]/\sigma_t(\beta^*)$ is the standardized process. The SNP density now takes the following form:

$$f(y_y;\beta)=\frac{1}{\sqrt{\sigma_t^2}}\frac{[P_K(z_t,x_t)]^2\phi(z_t)}{\int[P_K(u,x_t)]^2\phi(u)du} \tag{4.84}$$

where $\phi$ denoted the standard normal density, $x=(y_{t-1},...,y_{t-L})$ and the polynomials

$$P_K(z,x_t)=\sum_{t=1}^{K_z}\left[\sum_{j=0}^{K_x}a_{ij}x_t^j\right]z^i \tag{4.85}$$

**4.3.2.2 Example of EMM estimation.** We consider the SV model similar to the one considered by Andersen and Sorensen (1997). This model is of the form:

$$y_t = \sigma_t z_t \tag{4.86}$$

$$\ln \sigma_t^2 = \omega + \beta \ln \sigma_t^2 + \sigma_u u_t \tag{4.87}$$

where $(z_t, u_t)$ is iid $N(0, I_2)$. This model was estimated by GMM using twenty four moments. The return series $y_t$ is assumed to be demeaned. For $-1 < \beta < 1$ and $\sigma_u > 0$ the return series, $y_t$, is strictly stationary and ergodic, and conditional moments of any order exist. Let $\omega_t = \ln \sigma_t^2$ so that $\sigma_t = e^{\omega_t/2}$. Then the model can be rewritten as

$$y_t = e^{\omega_t/2} z_t \tag{4.88}$$

$$\omega_t = \omega + \beta \omega_{t-1} + \sigma_u u_t \tag{4.89}$$

To impose stationarity on the log volatility process, the logistic transformation is used to define the autoregressive parameter $\beta$ from the unrestricted parameter $\beta^*$

$$\beta = \frac{e^{\beta^*}}{1 + e^{\beta^*}} \tag{4.90}$$

The logistic transformation restricts $\beta$ to the interval $(0, 1)$. This restriction is reasonable since negative values of $\beta$ are not empirically relevant for asset returns. The unconditional mean of the log volatility process is $\mu = \omega/(1 - \beta)$

We simulate a sample of size $N = 4000$ from the model using parameters $\alpha = -0.147$, $\beta = 0.98$ and $\sigma_u = 0.166$. taken from Andersen, Chung and Sorensen (1999). The parameters are calibrated to match typical daily return data.

Andersen, Chung and Sorensen (1999) study the EMM estimation of this model using an extensive Monte Carlo study. They find that EMM performs substantially better than GMM, and comparably to direct likelihood-based inference procedure. For samples of size 1000 or less they find that a simple Gaussian GARCH(1,1) SNP model is a good choice for a score generator. Only for much larger samples do they find that adding Hermite polynomial terms to the SNP model improves efficiency. Regarding inference, the find that EMM objective function test for overidentifying restriction is remarkably reliable.

The quasi-maximum likelihood estimation of the SNP model utilizes random restarts of the optimizer to avoid getting stuck at a potentia local minimum. EMM converges with high objective function p-value indicating that the data do not reject the single overidentifying restriction implied by the GARCH(1,1) score generator. The estimates are $\omega = -0.2631$, $\sigma_u = 0.1950$ and $\beta = 0.9652$. They are reasonably close to their true values.

GMM and EMM estimates are remarkably similar. However, based on the extensive MC study the above mentioned authors recommend EMM over GMM for the following reasons: (1) EMM estimates are numerically more stable; (2) EMM estimates have smaller root mean square errors; (3) the problems associated with the choice of weighting matrices in the case of GMM are absent in EMM; (4) the EMM test for overidentifying restrictions is more stable; (5) inference regarding the parameters based on EMM t-statistics is more reliable.

Next, we consider EMM estimation of our model for the S&P 500 daily returns using the best fitting score generator. The small p-value $p = 1.916e - 11$ of the final EMM objective value indicates that the SV model is rejected by the S&P 500 returns.

Gallant, Hsieh and Tauchen (1997) consider the general univariate SV model

$$y_t = \mu + \sum_{j=1}^{p} \phi_j y_{t-j} + e^{\omega_t/2} \sigma_z z_t \tag{4.91}$$

$$\omega_t = \sum_{j=1}^{q} \beta_j \omega_{t-j} + \sigma_u u_t \tag{4.92}$$

where $z_t$ and $u_t$ are iid Gaussian random variables with mean zero, unit variance and correlation coefficient $\rho$. The model allows for autoregressive effects in the mean and log volatility. A negative correlation between the innovations to the level and log-volatility allow for the so-called leverage effect.

Fitting this model with leverage to the S&P data produces much better results. In case when $p$=1 and $q = 4$ we get a p-value of almost 0.2. So the model makes sense.

### 4.3.3 Sequential parameter learning.

Assume a Markovian dynamic model for sequentially observed data vector $y_t$, in which the state vector at time $t$ is $x_t$ and the fixed parameter vector is $\theta$. The model is specified at each time $t$ by the observation equation defining the observation density

$$p(y_t|x_t, \theta) \tag{4.93}$$

and the Markovian evolution equation, or state equation, defining the transition density

$$p(x_t|x_{t-1}, \theta) \tag{4.94}$$

Sequential Monte Carlo methods aim to sequentially update Monte Carlo sample approximation to the sequence of posterior distributions $p(x_t, \theta|D_t)$ where $D_t = \{D_{t-1}, y_t\}$ is the information set at time $t$. On observing the new observation $y_{t+1}$ it is desired to produce a sample from the current posterior $p(x_{t+1}, \theta|D_{t+1})$.

We have already considered model where $\theta$was assumed known, so that the focus was entirely on filtering for the state vector. As time evolves to $t+1$ we observe $y_{t+1}$, and want to generate a sample from the posterior $p(x_{t+1}|D_{t+1})$.Theoretically

$$p(x_{t+1}|D_{t+1}) \propto p(y_{t+1}|x_{t+1})p(x_{t+1}|D_t) \tag{4.95}$$

where $p(x_{t+1}|D_t)$ is the prior density of $x_{t+1}$ and $p(y_{t+1}|x_{t+1})$ is the likelihood function.

In the general model, standing at time $t$, we now have a combined sample

$$\left\{x_t^{(j)}, \theta_t^{(j)} : j = 1, ..., N\right\} \tag{4.96}$$

and associated weights

$$\{\omega_t^{(j)} : j = 1, ..., N\} \tag{4.97}$$

representing an importance sample approximation to the time $t$ posterior $p(x_t, \theta|D_t)$ for both parameter and state. Note that the $t$ suffix on the $\theta$ samples here indicate that they are from the time $t$ posterior, not that $\theta$ is time-varying. Time evolves to $t+1$, we observe $y_{t+1}$,and now want to generate a sample from $p(x_{t+1}, \theta|D_{t+1})$. Bayes' theorem gives us

$$p(x_{t+1}, \theta|D_{t+1}) \propto p(y_{t+1}|x_{t+1}, \theta)p(x_{t+1}|\theta, D_t)p(\theta|D_t) \tag{4.98}$$

where $p(\theta|D_t)$ is now an important ingredient.

Consider briefly a model in which $\theta$ is replaced by $\theta_t$at time $t$, and simply include $\theta_t$ in the augmented state vector. Then add an independent, zero-mean normal increment to the parameter at each time. That is,

$$\theta_{t+1} = \theta_t + \xi_{t+1} \tag{4.99}$$

$$\xi_{t+1} \sim N(0, W_{t+1}) \tag{4.100}$$

for some specified variance matrix $W_{t+1}$ and where $\theta_t$ and $\xi_{t+1}$are conditionally independent given $D_t$. With this model, the standard filtering methods for state alone now apply. Among the various issues and drawbacks of this approach, the key on is simply that fixed model parameters are, well, fixed. Pretending that they are in fact time-varying implied an artificial loss of information, resulting in posteriors that are eventually too diffuse.

Understanding the imperative to develop some method of smoothing for approximation of the required density $p(\theta|D_t)$, West (1993) developed kernel smoothing methods that provided the basis for rather effective adaptive importance sampling techniques.

Standing at time $t$, suppose we have current posterior parameter samples $\theta_t^{(j)}$and weights $\omega_t^{(j)}$,providing discrete MC approximation to $p(\theta|D_t)$. Write $\bar{\theta}_t$ and $V_t$ for the MC posterior mean and variance matrix of $p(\theta|D_t)$, computed from the MC sample $\theta_t^{(j)}$with weights $\omega_t^{(j)}$. The smooth kernel density form of West (1993) is given by

$$p(\theta|D_t) \approx \sum_{j=1}^{N} \omega_t^{(j)} N(m_t^{(j)}, h^2 V_t) \tag{4.101}$$

with the following components. First, $N(\cdot|m, S)$ is a multivariate normal density mean $m$ and variance $S$, so that the above density is a mixture of $N(\theta|m_t^{(j)}, h^2 V_t)$ distributions weighted by the sample weights $\omega_t^{(j)}$.Kernel rotation and scaling uses $V_t$, the MC posterior variance, and the overall scale of kernels is a function of the smoothing parameter $h > 0$.

The kernel locations $m_t^{(j)}$are specified using shrinkage rule introduced by West (1993). West introduced the novel idea of shrinkage kernel locations. Take

$$m_t^{(j)} = a\theta_t^{(j)} + (1-a)\bar{\theta}_t \tag{4.102}$$

where $a = \sqrt{1-h^2}$. With these kernel locations, the resulting normal mixture retains the mean $\bar{\theta}_t$and now has the correct variance $V_t$.

The loss of information is explicitly represented by the component of $W_{t+1}$. Now, there is close tie between (159) and (160) and the kernel smoothing approach. To see this, note that the MC approximation to $p(\theta_{t+1}|D_t)$ implied by equation (159) is also a kernel form, namely

$$p(\theta_{t+1}|D_t) \approx \sum_{j=1}^{N} \omega_t^{(j)} N(\theta_{t+1}|\theta_t^{(j)}, W_{t+1}) \tag{4.103}$$

and this is over-dispersed relative to the required variance $V_t$.

There is a way to correct this Liu and West by introducing $A_{t+1} = I - W_{t+1}V_t^{-1}/2$, so that in the case of approximate joint normality of $(\theta_t, \xi_{t+1}|D_t)$, this would imply the conditional normal evolution in which

$$p(\theta_{t+1}|\theta_t) = N(\theta_{t+1}|A_{t+1}\theta_t + (I - A_{t+1})\bar{\theta}_t, (I - A_{t+1}^2)V_t) \tag{4.104}$$

The resulting MC approximation to $p(\theta_{t+1}|D_t)$ is then a generalized kernel form with complicated shrinkage patterns. If one restricts here to the very special case in which the evolution variance matrix is specified using a standard discount factor technique. We can take

$$W_{t+1} = V_t(\frac{1}{\delta} - 1) \tag{4.105}$$

where $\delta$ is a discount factor, typically around $0.95 - 0.99$. In this case, $A_{t+1} = aI$ with

$a = (3\delta - 1)/2\delta$ and the conditional evolution density becomes

$$p(\theta_{t+1}|\theta_t) \sim N(\theta_{t+1}|a\theta_t + (1-a)\theta_t, \bar{h}^2 V_t) \tag{4.106}$$

where $h^2 = 1 - ((3\delta - 1)/2\delta)^2$.

#### 4.3.3.1 Application of Liu and West filter to SV model.

Let $y_t$,for $t = 1, ..., n$ be modeled as

$$y_t|x_t \sim N(0, e^{x_t}) \tag{4.107}$$

$$(x_t|x_{t-1}, \theta) \sim N(\alpha + \beta x_{t-1}, \tau^2) \tag{4.108}$$

where $\theta = (\alpha, \beta, \tau^2)$

We are going to simulate $n = 500$ points, with $\alpha = -0.0031$, $\beta = 0.9951$ and $\tau^2 = 0.0074$ and $x_1 = \alpha/(1-\beta) = -0.632653$.

Prior setup:

$$x_0 \sim N(m_0, C_0)$$

$$\beta \sim N(\beta_0, V_\beta)$$

$$\alpha \sim N(\alpha_0, V_\alpha)$$

$$\tau^2 \sim IG(n_0/2, n_0\tau_0^2/2)$$

where $m_0 = 0.0$, $C_0 = 0.1$, $\alpha_0 = -0.0031$, $V_\alpha = 0.01$, $\beta_0 = 0.9951$, $V_\beta = 0.01$, $n_0 = 3$, and $\tau_0^2 = 0.0074$.

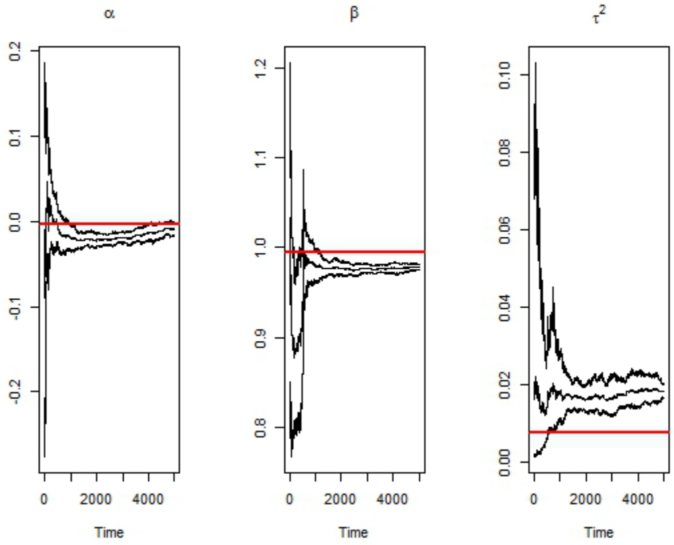

Figure 4.3: Liu and West parameter learning of SV model

## 5.0 CONTINUOUS STOCHASTIC VOLATILITY MODELS.

Stochastic volatility (SV) models are useful because they explain in a self-consistent way why options with different strikes and expirations have different Black-Scholes implied volatilities-that is, the volatility smile. Moreover, unlike alternative models that can fit the smile(such as local volatility model), SV models assume realistic dynamics for the underlying. From hedging perspective, traders who use Black-Scholes model must continuously change the volatility assumption in order to match market prices. Their hedge ratios change accordingly in an uncontrolled way: SV models bring some order into this chaos. Distributions of real returns are highly peaked and fat-tailed relative to the Gaussian distribution. Fat tails and high central peak are characteristics of a mixture of distributions with different variances. That is way variance is modeled as a random variable.

### 5.1 FIRST LOOK AT CONTINUOUS SV MODELS.

Suppose that the stock with the price $S$ and its variance $v = \sigma^2$ are driven by the following stochastic differential equations:

$$dS_t = \mu_t S_t dt + \sqrt{v_t} S_t dW_1 \tag{5.1}$$

$$dv_t = \alpha(S_t, v_t, t)dt + \eta\beta(S_t, v_t, t)\sqrt{v_t} dW_2 \tag{5.2}$$

with

$$< dW_1 dW_2 >= \rho t \tag{5.3}$$

where $\mu_t$ is the deterministic drift, $\eta$ is the volatility of volatility and $\rho$is the correlation coefficient between random stock returns. $dW_1$and $dW_2$are Brownian motions or sometimes called Wiener processes.

In the Black-Scholes case, there is only one source of randomness, the stock price, which can be hedged with a stock. In this case, not only the stock price is random but also its volatility which needs hedging as well. So we set a portfolio $\Pi$ containing the option being priced, whose value is $V(S, v, t)$, a quantity $-\Delta$ of the stock and quantity $-\Delta_1$ of asset whose value $V_1$ also depends on volatility.

We have

$$\Pi = V - \Delta S - \Delta_1 V_1 \tag{5.4}$$

The change in portfolio is given then by

$$d\Pi = \left\{ \frac{\partial V}{\partial t} + \frac{1}{2} v S^2 \frac{\partial^2 V}{\partial S^2} + \rho\eta v \beta S \frac{\partial^2 V}{\partial v \partial S} + \frac{1}{2}\eta^2 v \beta^2 \frac{\partial^2 V}{\partial v^2} \right\} dt$$

$$-\Delta_1 \left\{ \frac{\partial V}{\partial t} + \frac{1}{2} v S^2 \frac{\partial^2 V_1}{\partial S^2} + \rho\eta v \beta S \frac{\partial^2 V_1}{\partial v \partial S} + \frac{1}{2}\eta^2 v \beta^2 \frac{\partial^2 V_1}{\partial v^2} \right\} dt$$

$$+ \left\{ \frac{\partial V}{\partial S} - \Delta_1 \frac{\partial V_1}{\partial S} - \Delta \right\} dS + \left\{ \frac{\partial V}{\partial v} - \Delta_1 \frac{\partial V_1}{\partial v} \right\} dv$$

To make the portfolio instantaneously risk-free, we must choose

$$\frac{\partial V}{\partial S} - \Delta_1 \frac{\partial V_1}{\partial S} - \Delta = 0$$

and

$$\frac{\partial V}{\partial v} - \Delta_1 \frac{\partial V_1}{\partial v} = 0$$

This gives us

$$d\Pi = \left\{ \frac{\partial V}{\partial t} + \frac{1}{2} v S^2 \frac{\partial^2 V}{\partial S^2} + \rho\eta v \beta S \frac{\partial^2 V}{\partial v \partial S} + \frac{1}{2}\eta^2 v \beta^2 \frac{\partial^2 V}{\partial v^2} \right\} dt$$

$$-\Delta_1\left\{\frac{\partial V}{\partial t}+\frac{1}{2}vS^2\frac{\partial^2 V_1}{\partial S^2}+\rho\eta v\beta S\frac{\partial^2 V_1}{\partial v\partial S}+\frac{1}{2}\eta^2 v\beta^2\frac{\partial^2 V_1}{\partial v^2}\right\}dt = r\Pi dt = r(V-\Delta S-\Delta_1 V_1)dt$$

where we explicitly made our portfolio risk by introducing the risk free rate. After some rearrangements one can obtain

$$\frac{\frac{\partial V}{\partial t}+\frac{1}{2}vS^2\frac{\partial^2 V}{\partial S^2}+\rho\eta v\beta S\frac{\partial^2 V}{\partial v\partial S}+\frac{1}{2}\eta^2 v\beta^2\frac{\partial^2 V}{\partial v^2}+rS\frac{\partial V}{\partial S}-rV}{\frac{\partial V}{\partial v}}$$

$$=\frac{\frac{\partial V}{\partial t}+\frac{1}{2}vS^2\frac{\partial^2 V_1}{\partial S^2}+\rho\eta v\beta S\frac{\partial^2 V_1}{\partial v\partial S}+\frac{1}{2}\eta^2 v\beta^2\frac{\partial^2 V_1}{\partial v^2}+rS\frac{\partial V_1}{\partial S}-rV_1}{\frac{\partial V_1}{\partial v}} \tag{5.5}$$

The left-hand side is a function of only $V$ and the right-hand side is the function of only $V_1$. Either side should then be equal to some function of $S$,$v$ and $t$.

$$\frac{\partial V}{\partial t}+\frac{1}{2}vS^2\frac{\partial^2 V}{\partial S^2}+\rho\eta v\beta S\frac{\partial^2 V}{\partial v\partial S}+\frac{1}{2}\eta^2 v\beta^2\frac{\partial^2 V}{\partial v^2}+rS\frac{\partial V}{\partial S}-rV=-(\alpha-\phi\beta\sqrt{v})\frac{\partial V}{\partial v}$$

$\phi(S,v,t)$ is called the market price of volatility risk.

### 5.1.1 Heston model.

The Heston model corresponds to choosing $\alpha(S,v_t,t)=-\lambda(v_t-\bar{v})$ and $\beta(s,v,t)=1$. Then our stochastic processes become

$$dS_t=\mu_t S_t dt+\sqrt{v_t}S_t dW_1 \tag{5.6}$$

$$dv_t=-\lambda(v_t-\bar{v})dt+\eta\sqrt{v_t}dW_2 \tag{5.7}$$

with

$$< dW_1 dW_2 >= \rho t \tag{5.8}$$

The process followed by the instantaneous variance $v_t$ may be recognized as as version of the square root process described by Cox, Ingersoll, and Ross (CIR process).We can now substitute the above values for into the general valuation equation to obtain:

$$\frac{\partial V}{\partial t}+\frac{1}{2}vS^2\frac{\partial^2 V}{\partial S^2}+\rho\eta v\beta S\frac{\partial^2 V}{\partial v\partial S}+\frac{1}{2}\eta^2 v\beta^2\frac{\partial^2 V}{\partial v^2}+rS\frac{\partial V}{\partial S}-rV=\lambda(v_t-\bar{v})\frac{\partial V}{\partial v}$$

We also show the original derivation done by Heston of the solution of the above PDE for the European options.

Before solving this equation with appropriate boundary conditions, we can simplify it by making a series of changes of variables. Let $K$ be the strike price of the option, $T$ time to expiration, $F_{t,T}$ the time $T$ forward price of the stock and $x=\log(F_{t,T}/K)$. Let denote by $C$ the future value to expiration of the option and $\tau=T-t$ is time to expiration. Then the above PDE can be rewritten as

$$-\frac{\partial C}{\partial\tau}+\frac{1}{2}vC_{11}-\frac{1}{2}vC_1+\frac{1}{2}\eta^2 vC_{22}+\rho\eta vC_{12}-\lambda(v-\bar{v})C_2=0$$

where by subscripts 1 and 2 we refer to differentiation with respect to $x$ and $v$ respectively.

According to Duffie, Pan and Singleton (2000), the solution has the form

$$C(x,v,\tau)=K\left\{e^x P_1(x,v,\tau)-P_0(x,v,\tau)\right\} \tag{5.9}$$

Substituting this anzats into the our PDE we get

$$-\frac{\partial P_j}{\partial\tau}+\frac{1}{2}v\frac{\partial^2 P_j}{\partial x^2}-(\frac{1}{2}-j)v\frac{\partial P_j}{\partial x}+\frac{1}{2}\eta^2 v\frac{\partial^2 P_j}{\partial v^2}+\rho\eta v\frac{\partial^2 P_j}{\partial x\partial v}+(a-b_j v)\frac{\partial P_j}{\partial v}=0$$

for $j=0,1$ where

$$a=\lambda\bar{v}\qquad b_j=\lambda-j\rho\eta$$

subject to terminal conditions

$$\lim_{\tau\to 0}P_j(x,v,\tau)=1 \tag{5.10}$$

for $x > 0$ and

$$\lim_{\tau \to 0} P_j(x, v, \tau) = 0 \tag{5.11}$$

for $x \leq 0$.

We solve our equations using Fourier transform technique. Without loading the reader with complex derivations we simply show the final solution.

$$P_j(x, v, \tau) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty du Re \left\{ \frac{\exp\{C_j(u, \tau)\bar{v} + D_j(u, \tau)v + iux\}}{iu} \right\} \tag{5.12}$$

where

$$D(u, \tau) = r_- \frac{1 - e^{-d\tau}}{1 - ge^{-d\tau}}$$

$$C(u, \tau) = \lambda \left\{ r_- \tau - \frac{2}{\eta^2} \log \left( \frac{1 - ge^{-d\tau}}{1 - g} \right) \right\}$$

$$r_\pm = \frac{\beta \pm d}{\eta^2} = \frac{\beta \pm \sqrt{\beta^2 - 4\alpha\gamma}}{2\gamma}$$

$$g = \frac{r_-}{r_+}$$

$$\alpha = -\frac{u^2}{2} - \frac{iu}{2} + iju$$

$$\beta = \lambda - \rho\eta j - \rho\eta iu$$

$$\gamma = \frac{\eta^2}{2}$$

## 5.2 LOCAL VOLATILITY.

Given the computational complexity of stochastic volatility models and the difficulty of fitting parameters of the current prices of vanilla options,practitioners sought a simpler way of pricing exotic options consistently with the volatility skew. The breakthrough came when Dupire (1994) and Dernan and Kani (1994) noted that under the risk-neutrality, there was a unique diffusion process consistent with the distribution of marker prices of options. The correspoding unique state-dependent diffusion coefficient $\sigma_L(S,t)$, consistent with current European options prices, is known as the local volatility function.

### 5.2.1 A review of Dupire's work.

For a given expiration $T$ and current stock price $S_0$,the collection $\{C(S_0,K,T)\}$ of undiscounted option prices of different strikes yields the risk-neutral probability density function $\phi$ of the final spot $S_T$ through the relationship

$$C(S_0,K,T)=\int_K^\infty dS_T\phi(S_T,T;S_0)(S_T-K) \tag{5.13}$$

Differentiating twice with respect to $K$ we get

$$\phi(K,T;S_0)=\frac{\partial^2 C}{\partial K^2}$$

Given the distribution of final spot prices for each time $T$ conditional of some starting spot price $S_0$, Dupire was able to show that there is a unique risk neutral diffusion process which generates these distributions.

Suppose the stock price diffuses with risk-neutral drift $\mu_t=r_t-D_t$ where $r_t$ is the free interest rate and $D_t$is the dividend yield and local volatility $\sigma(S,t)$ according to the equation:

$$\frac{dS}{S}=\mu_t dt+\sigma(S_t,t)dW \tag{5.14}$$

The pseudo-probability density function $\phi(S_T,T'S_0)$ of the final spot at time $T$ evolves according to the Fokker-Planck equation:

$$\frac{1}{2}\frac{\partial^2}{\partial S_T^2}\left(\sigma^2 S_T^2 \phi\right) - S\frac{\partial}{\partial S_T}(\mu S_T \phi) = \frac{\partial \phi}{\partial T} \tag{5.15}$$

Differentiating (5.13) with respect to $T$ gives

$$\frac{\partial C}{\partial T} = \int_K^\infty dS_T \left\{ \frac{\partial^2}{\partial S_T^2}\left(\sigma^2 S_T^2 \phi\right) - \frac{\partial}{\partial S_T}(\mu S_T \phi) \right\} (S_T - K) \tag{5.16}$$

Integrating by parts gives:

$$\frac{\partial C}{\partial T} = \frac{\sigma^2 K^2}{2}\frac{\partial^2 C}{\partial K^2} + \mu(T)\left(-K\frac{\partial C}{\partial K}\right) \tag{5.17}$$

which is the Dupire equation when the underlying stock has risk-neutral drift $\mu$. That is, the forward price of the stock at time $T$ is given by

$$F_T = S_0 \exp\left\{\int_0^T dt \mu_t\right\}$$

Were we to express the option price as a function of the forward $F_T = S_0 \exp\left\{\int_0^T dt\mu(t)\right\}$, we would get the same expression minus the drift term. That is,

$$\frac{\partial C}{\partial T} = \frac{\sigma^2 K^2}{2}\frac{\partial^2 C}{\partial K^2}$$

where $C$ now represents $C(F_T, K, T)$. Inverting this gives

$$\sigma^2(K, T, S_0) = \frac{\frac{\partial C}{\partial T}}{\frac{1}{2}K^2\frac{\partial^2 C}{\partial K^2}} \tag{5.18}$$

We can view this expression as a definition of local volatility function regardless of what kind of process governs the evolution of volatility.

### 5.2.2 Local volatility in terms of implied volatility.

Market prices of options are quoted in terms of Black-Scholes implied volatility $\sigma_{BS}(K, T; S_0)$. In other words, we may write

$$C(S_0, K, T) = C_{BS}(S_0, K, \sigma_{BS}(S_0, K, T), T)$$

It will be more convenient to work in terms of two dimensionless variables:

$$w(S_0, K, T = \sigma_{BS}(K, T; S_0)T$$

and

$$y = \log\left(\frac{K}{F_T}\right)$$

In terms of these variables, the Black-Scholes formula becomes

$$C_{BS}(F_T, y, w) = F_T\left\{N\left(-\frac{y}{\sqrt{w}} + \frac{\sqrt{w}}{2}\right) - e^y N\left(-\frac{y}{\sqrt{w}} - \frac{\sqrt{w}}{2}\right)\right\} \tag{5.19}$$

and the Dupire equation becomes

$$\frac{\partial C}{\partial T} = \frac{v_L}{2}\left\{\frac{\partial^2 C}{\partial y^2} - \frac{\partial C}{\partial y}\right\} + \mu(T)C \tag{5.20}$$

where $v_L = \sigma^2(S_0, K, T)$. After some manipulation with derivatives of the Black-Scholes price and prices of options in terms the implied volatility we arrive at the following result

$$v_L = \frac{\frac{\partial w}{\partial T}}{1 - \frac{y}{w}\frac{\partial w}{\partial y} + \frac{1}{4}\left(-\frac{1}{4} - \frac{1}{w} + \frac{y^2}{w^2}\right)\left(\frac{\partial w}{\partial y}\right)^2 + \frac{1}{2}\frac{\partial^2 w}{\partial y^2}} \tag{5.21}$$

## 5.3 STOCHASTIC VOLATILITY WITH JUMPS.

Assume the stock price follows the SDE

$$dS = \mu S dt + \sigma S dW + (J-1)dq \tag{5.22}$$

where the $dq$ is the Poisson process.

So, once again we, we set up a portfolio $\Pi$ containing the option being priced whose value we denote by $V(S, v, t)$ , a quantity $-\Delta$ of the stock and a quantity $-\Delta_1$ of another asset whose value $V_1$ depends on the jump.

We have

$$\Pi = V - \Delta S - \Delta_1 V_1 \tag{5.23}$$

The change on this portfolio can be found using the Ito's lemma and is given by

$$d\Pi = \left\{ \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right\} dt - \Delta_1 \left\{ \frac{\partial V_1}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V_1}{\partial S^2} \right\} dt + \left\{ \frac{\partial V}{\partial S} - \Delta_1 \frac{\partial V_1}{\partial S} - \Delta \right\} dS^c +$$

$$\{ V(JS,t) - V(S,t) - \Delta_1 (V_1(JS,y) - V_1(S,t)) - \Delta(J-1)S \} \, dq$$

where $S^c(t)$ is the continuous part of $S(t)$.

To make the portfolio risk free, we must choose

$$\frac{\partial V}{\partial S} - \Delta_1 \frac{\partial V_1}{\partial S} - \Delta = 0$$

$$V(JS,t) - V(S,t) - \Delta_1 (V_1(JS,y) - V_1(S,t)) - \Delta(J-1)S = 0$$

This leaves us with

$$d\Pi = \left\{ \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right\} dt - \Delta_1 \left\{ \frac{\partial V_1}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V_1}{\partial S^2} \right\} dt = r\Pi dt = r(V - \Delta S - \Delta_1 V_1) dt \tag{5.24}$$

Collecting all terms with $V$ on one side and terms with $V_1$on the other side we get

$$\frac{\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV}{\delta V - (J-1)S\frac{\partial V}{\partial S}} = \frac{\frac{\partial V_1}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V_1}{\partial S^2} + rS\frac{\partial V_1}{\partial S} - rV_1}{\delta V_1 - (J-1)S\frac{\partial V_1}{\partial S}}$$

where we have defined $\delta V = V(JS,t) - V(S,t)$.

The only way the above expression can be true if each side is equal to a function of $S$ and $t$, which we denote by $-\lambda$.

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV + \lambda \left\{ V(JS,t) - V(S,t) - (J-1)S\frac{\partial V}{\partial S} \right\} = 0$$

### 5.3.1 Risk management.

Volatility is a measure of risk involved in financial and economic decision making and it is a key part of modern financial theory. So far, we have looked at the volatility modeling through time series prism where GARCH and SV models prevail as the cornerstones. Another prospective on volatility modeling can be gained by looking at financial mathematics and in particular derivatives pricing. The celebrated result by Black and Scholes (BS) in 1973 offers a framework for valuation of European style derivatives within a simple set of assumptions. Six parameters enter the pricing formula: the current underlying asset price, the strike price, the expiry date of the option, the riskless interest rate, the dividend yield, and a constant volatility parameter that describes the instantaneous standard deviation of the returns of the log-asset price. The application of the formula, however, faces an obstacle: only its five parameters are known quantities. The last one, the volatility parameter, is unknown.

Going back to our second perspective we can estimate the volatility from option prices. In other words we recover the volatility that the market has priced into a given option. We are interested in what volatility is implied in observed option prices, if the BS model is valid description of market conditions? This reverse perspective is called BS implied volatility. IV exhibits a pronounced curvature across strikes and is also curved across time to maturity but not as much. For a given time to maturity this function has been named smile, and the entire curved surface is called the implied volatility surface (IVS).

IV popularity can be explained for two reasons. One of them is simplicity of BS formula and easiness of communication. Another reason is more fundamental and says that the option implied volatility is a forward looking variable (because option are bets on future development of the underlying asset). IV reflects market expectations on volatility over the remaining life time of the option.

## 5.4 IMPLIED VOLATILITY SURFACE.

Let's recall the form of the BS formula. The price $C(S_t, t)$ of a plain vanilla call is the solution to the PDE with the boundary condition $C(S_T, T) = (S_T - K)^+$. The explicit solution is known as the Black and Scholes formula for calls:

$$C^{BS}(S_t, t, K, T, \sigma, r, \delta) = e^{-\delta\tau} S_t \Phi(d_1) - e^{-r\tau} K \Phi(d_2) \tag{5.25}$$

where

$$d_1 = \frac{\ln(S_t/K) + (r - \delta + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}} \tag{5.26}$$

$$d_2 = d_1 - \sigma\sqrt{\tau} \tag{5.27}$$

and where $\Phi(u) = \int_{-\infty}^{u} \phi(x)dx$ is the cdf of standard normal distribution. $\tau$ is time to maturity.

It is obvious that the BS formula is derived under assumptions that are not likely to be met in reality: frictionless markets, not transaction costs, no price jumps, and constant volatility. Due to the simplicity of the model, any deviation from these assumptions is summarized in one parameter: the IV smile and IVS.

The only unknown parameter in the BS pricing formula is the volatility. Given observed market prices $\tilde{C}_t$, it is therefore natural to define the implied volatility (IV):

$$\tilde{\sigma} : \qquad C^{BS}(S_t, t, K, T, \tilde{\sigma}) - \tilde{C}_t = 0$$

IV is the empirically determined parameter that makes the BS formula fit market prices of options. Since the BS is monotone in $\sigma$, there exists a unique solution $\tilde{\sigma} > 0$. In the derivation of of the BS formula it is assumed that the volatility is constant. IV $\tilde{\sigma}$, however, is a curve across options strikes $K$ and across expiry date $T$. Thus IV is in fact a mapping from time,strike price and expiry days to $\mathbb{R}^+$:

$$\tilde{\sigma} : \qquad (t, K, T) \to \tilde{\sigma}_t(K, T)$$

The mapping is called the implied volatility surface (IVS).

Often it is not convenient to work in absolute variables as expiry dates and strikes. Rather one prefers relative variables, since the analysis becomes independent of expiry effects and movement of the underlying. As a new scale, one typically uses time to maturity $\tau = T - t$ and moneyness. A stock price moneyness can be defined by:

$$\kappa = K/S_t \tag{5.28}$$

We say that an option is at-the-money (ATM) when $\kappa \approx 1$.A call option is called out-of-the-money, OTM, (in-the-money,(ITM)), if $\kappa > 1 (\kappa < 1)$ with the reverse applying to puts. The most fundamental conclusion is that OTM puts and ITM calls are traded at higher prices that the corresponding ATM options. Obviously, the BS model does not properly capture the probability of of large downward movements of the underlying asset price.

### 5.4.1 Static stylized facts.

1. For short time to maturities the smile is very pronounced, while the smile becomes more and more shallow for longer time to maturities.
2. The smile function achieves its minimum in the neighborhood of ATM to near OTM call options.
3. OTM put regions display higher levels of IV than OTM call regions.
4. The volatility of IV is biggest for short maturity options and monotonically declining with time to maturity.
5. Returns of the underlying asset and returns of IV are negatively correlated indicating a leverage effect.
6. IV appears to be mean-reverting.
7. Shocks across the IV are highly correlated.

### 5.4.2 Hedging and risk management.

In the presence of smile, a first obvious challenge is the computation of the relevant hedge ratios. At first glance, an answer may be to insert IV into the BS derivatives in order to compute the hedge ratios for some option positions. This strategy is called an IV compensated BS hedge. However, one should be aware that this strategy can be erroneous, since IV is not necessarily equal to the hedging volatility. Analogously to IV, the hedging volatility, for instance for the delta, is defined by:

$$\tilde{\sigma_h}: \qquad \frac{\partial C^{BS}(S_t, t, K, T, \tilde{\sigma_h})}{\partial S} - \frac{\partial \tilde{C_t}}{\partial S} = 0,$$

which is the volatility that equates the BS delta with the delta of the true model. The hedging volatility is not directly observable.

One can prove that the bias in this approximation is systematic. The bias translates into the following errors in the hedge ratios: for ITM options the use of IV to compute the hedge ratios leads to an underhedge position in the delta, while for OTM options it leads an overhedge position. Only for ATM options this type of a hedge is perfect.

A better strategy due to Lee (2001) includes the stochastic volatility case. Consider

$$\frac{\partial \tilde{C_t}}{\partial K} = \frac{\partial C^{BS}(S_t, t, K, T, \tilde{\sigma})}{\partial K} + \frac{\partial C^{BS}(S_t, t, K, T, \tilde{\sigma})}{\partial \tilde{\sigma}} \frac{\partial \tilde{\sigma}}{\partial K} \tag{5.29}$$

and

$$\frac{\partial \tilde{C_t}}{\partial S} = \frac{\partial C^{BS}(S_t, t, K, T, \tilde{\sigma})}{\partial S} + \frac{\partial C^{BS}(S_t, t, K, T, \tilde{\sigma})}{\partial \tilde{\sigma}} \frac{\partial \tilde{\sigma}}{\partial S} \tag{5.30}$$

Using the second equation we can find that

$$\frac{\partial \tilde{\sigma}}{\partial S} = -\frac{K}{S} \frac{\partial \tilde{\sigma}}{\partial K} \tag{5.31}$$

Thus the corrected hedge ration is:

$$\frac{\partial \tilde{C_t}}{\partial S} = \frac{\partial C^{BS}(S_t, t, K, T, \tilde{\sigma})}{\partial S} - \frac{\partial C^{BS}(S_t, t, K, T, \tilde{\sigma})}{\partial \tilde{\sigma}} \frac{K}{S_t} \frac{\partial \tilde{\sigma}}{\partial K} \tag{5.32}$$

This delta-hedge can be implemented without estimating an underlying stochastic volatility model.

For risk management, other difficulties appear, especially when IV compensated hedge ratios are used. When different BS models apply for different strikes, one may question whether delta and vega risks across different strikes can simply be adds to assess the overall risk in the option book: being a certain amount of dollars delta long in hight strike options, and the same amount delta short in low strike options, need not necessarily imply that the book is delta-neutral.

### 5.4.3 Pricing.

A next challenge is valuing exotic options. The reason is that even the simplest path dependent options, like barrier option, require sophisticated volatility specification. In some cases one knows the explicit formula for the price. However, which IV should we use for pricing. One could use the IV at the strike $K$, the one at the barrier $L$, or some average of both. The problem is the more virulent the more sensitive the exotic option is to volatility.

At this point it becomes clear that, in the presence of the IV, pricing is not sensible without a self-consistent and reliable model. One can use the stochastic volatility models. Another way, which is much closer to the concept of the IVS, is offered by the smile consistent local volatility models. These models rely on a volatility function that is directly backed out of prices of plain vanilla options observed in the market. Thus, the exotic option is priced consistently with the entire IVS. This is a natural approach, especially when the exotic option is to be hedged with plain vanilla options.

### 5.4.4 Predictive capabilities of IV.

In an efficient market, options instantaneously adjust to new information. Thus, IV predictions do not depend on the historical price or volatility series in an adaptive sense. This may be viewed as an advantage of IV type of models. There are two caveats though. First, the test on the forecasting ability of IV is always a joint test of option market efficiency and the option pricing model. Second, given the presence of the smile, one either has to restrict the analysis to ATM options or to find an appropriate weighting scheme of IV across different strikes. The overall consensus of the literature is that IV based predictions do contain a substantial amount of information on future volatility and are better than (only) time series based methods. At the same time, most authors conclude the IV is a biased predictor.

## 5.5 ESTIMATION OF IVS.

Parametric attempts to model the IVS along the strike profile usually employ quadratic specification. However, it seems that these parametric approaches are not capable of capturing the salient features of IVS patterns, and produce biased estimates.

Recently, non- and semi-parametric smoothing techniques for estimating the IVS have been used more and more. The main idea of these methods can be stated as follows: suppose we are given a data set $\{(x_i, y_i)\}_{i=1}^n$. In the context of IVS estimation, this would be some moneyness measures and time to maturity, or either of them, and IV respectively. The goal is to estimate the regression relationship

$$y_i = m(x_i) + \varepsilon_i \tag{5.33}$$

### 5.5.1 Nadaraya-Watson estimator.

For simplicity, consider the univariate model

$$Y = m(X) + \varepsilon \tag{5.34}$$

with unknown regression function $m$. The explanatory variable $X$ and the response variable $Y$ take value in $\mathbb{R}$, have the joint pdf $f(x, y)$ and are independent of $\varepsilon$. The error has the properties $E(\varepsilon|x) = 0$ and $E(\varepsilon^2|x) = \sigma^2(x)$.

Using the definition of conditional expectation we can write

$$m(x) = E(Y|X = x) = \frac{\int y f(x, y) dy}{f_x(x)} \tag{5.35}$$

where $f_x$is the marginal pdf. This form shows that the regression function can be estimated using kernel density estimates of the joint and marginal density.

Suppose we are given the randomly sampled iid data set $\{(x_i, y_i)\}_{i=1}^n$. Then the Nadaraya-Watson estimator is given by:

$$\hat{m}(x) = \frac{n^{-1}\sum_{i=1}^n K_h(x - x_i)y_i}{n^{-1}\sum_{i=1}^n K_h(x - x_i)} \tag{5.36}$$

where $K(u)$ is a kernel function satisfying $\int K(u)du = 1$, $K_h(u) = \frac{1}{h}K(\frac{u}{h})$ and $h$ is called the bandwidth.

One can rewrite the above result as

$$\hat{m}(x) = \frac{1}{n}\sum_{i=1}^n \omega_{i,n}(x)y_i \tag{5.37}$$

where

$$\omega_{i,n}(x) = \frac{K_h(x - x_i)}{n^{-1}\sum_{i=1}^n K_h(x - x_i)} \tag{5.38}$$

Under some regularity conditions, the Nadaraya-Watson estimator is consistent, i.e.

$$\hat{m}(x) \to m(x) \tag{5.39}$$

in probability.

### 5.5.2 Local polynomial smoothing.

Another view on the Nadaraya-Watson estimator can be taken by noting that it can be written as the minimizer of

$$\hat{m}(x) = min\sum_{i=1}^n (y_i - m)^2 K_h(x - x_i) \tag{5.40}$$

Computing the normal equations leads (60) as a solution for $m$. This reveals that Nadaraya-Watson estimator is a special case of fitting a constant in a local neighborhood of

$x$. In local polynomial smoothing this idea is generalized to fitting locally a polynomial of order $p$. This estimator can be formulated in terms of quadratic minimization problem

$$min \sum (y_i - \beta_0 - \beta_1(x - x_i) - ... - \beta_p(x - x_i)^p) K_h(x - x_i) \tag{5.41}$$

The solution to this problem looks like

$$\hat{\beta}(x) = (X^T W X)^{-1} X^T W y \tag{5.42}$$

where

$$X = \begin{pmatrix} 1 & x - x_1 & (x - x_1)^2 & ... & (x - x_1)^p \\ 1 & x - x_2 & (x - x_2)^2 & ... & (x - x_2^p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x - x_n & (x - x_n)^2 & ... & (x - x_n)^p \end{pmatrix} \tag{5.43}$$

and

$$W = \begin{pmatrix} K_h(x - x_1) & 0 & ... & 0 \\ 0 & K_h(x - x_2) & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & K_h(x - x_n) \end{pmatrix} \tag{5.44}$$

An important byproduct of local polynomial estimators is that they provide an easy and efficient way for computing derivatives up to order $(p + 1)$ of the regression function:

$$\hat{m}^{(j)}(x) = j! \hat{\beta}_j(x) \tag{5.45}$$

Another important difficulty with kernel estimator is bandwidth selection. We are not touching this subject here since it is a vast topic. We can only add that it is usually done using cross validation techniques or penalization approaches based on information criteria.

### 5.5.3 Least squares kernel smoothing.

In this section, we discuss a special smoother designed to estimate the IVS. It is a one-step procedure based on a least squares kernel estimator that smoothes IV in the space of option prices. There is no need to invert the BS formula in order to extract IV observations. The LSK estimator is a special case of a general class of estimators, the so-called kernel M-estimators, that has been introduced by Gourieoux.

We first rewrite the BS formula in terms of moneyness metric:

$$C^{BS}(S_t, t, K, T, \sigma, r, \delta) = S_t c^{BS}(\kappa_t, \tau, \sigma, r, \delta) \tag{5.46}$$

where $c^{BS}(\kappa_t, \tau, \sigma, r, \delta) = \Phi(d_1) - \kappa_t e^{-r\tau}\Phi(d_2)$, and $d_1 = \frac{-\ln \kappa_t + (r + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}$, $d_2 = d_1 - \sigma\sqrt{\tau}$. The LSK estimator for the IVS is defined by:

$$\hat{\sigma}(\kappa_t, \tau) = argmin_\sigma \sum_{i=1}^{n} \{\tilde{c}_{t_i} - c^{BS}(\cdot, \sigma)\}^2 \omega(\kappa_{t_i}) K_{(1)}\left(\frac{\kappa_t - \kappa_{t_i}}{h_1}\right) K_{(2)}\left(\frac{\tau - \tau_i}{h_2}\right) \tag{5.47}$$

$K_{(1)}$and $K_{(2)}$ are univariate kernels, and $\omega(\cdot)$ is a weight function, which allows for differential weights of observed option prices. The reason why one incorporated these weights is explained by he fact that ITM options contain a liquidity premium and should be used to a lesser extent.

One can prove using certain assumptions that this IVS estimator is consistent and has asymptotic normality.

# BIBLIOGRAPHY

[1] Empirical properties of asset returns: stylized facts and statistical issues in: Quantitative Finance, Vol 1, No 2, (March 2001) 223-236.

[2] Long range dependence in financial time series, Fractals in Engineering, E Lutton and J Levy Vehel (Eds.), Springer (2005).

[3] Volatility clustering in financial markets, in: A Kirman and G Teyssiere (Eds.): Long memory in economics, Springer (2007), 289-310.

[4] Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes, Christian Francq and Jean-Michel Zakoian.Bernoulli Volume 10, Number 4 (2004), 605-637.

[5] A Tutorial on Particle Filtering and Smoothing:Fifteen years later. Arnaud Doucet and Adam Johansen (2008).

[6] A survey of sequential Monte Carlo methods for economics and finance, Econometric Reviews, (2012), Vol. 31 (3), pp. 245-296

[7] Cappe, O., Godsill, S. J. and Moulines, E. (2007) An overview of e existing methods and recent advances in sequential Monte Carlo, IEEE Proc., 95, 899-924.

[8] Doucet, A., Godsill, S. and Andrieu, C. (2000) On sequential Monte-Carlo sampling methods for Bayesian filtering. Stat. Comput., 10, 197-208.

[9] Doucet, A., De Freitas, N. and Gordon, N. (eds.) (2001) Sequential Monte Carlo Methods in Practice. Springer.

[10] M Briers, A Doucet, and S Maskell. Smoothing algorithms for state-space models. Volume 62, Number 1, 61-89. Annals of the Institute of Statistical Mathematics. December 2010.

[11] Thomas B. Schon, Adrian Wills and Brett Ninness. System Identification of Nonlinear State-Space Models. Automatica, 47(1):39-49, January 2011.

[12] Practical Filtering with Sequential Parameter Learning Polson, N. G., Stroud, J. R. and Muller, P. (2008). Journal of the Royal Statistical Society, Series B, 70, 413-428.

[13] Practical Filtering for Stochastic Volatility Models Stroud, J. R., Polson, N. G. and Muller, P. (2004). State Space and Unobserved Components Models (Harvey et al., eds.), Cambridge University Press, 236-247.

[14] Bayesian Inference for Derivative Prices Polson, N. G. and Stroud, J. R. (2003). Bayesian Statistics 7 (Bernardo et al., eds.), Oxford University Press, 641-650.

[15] Liu and West (2001) Combined parameters and state estimation in simulation-based filtering. In Sequential Monte Carlo Methods in Practice (Eds. A. Doucet, N. de Freitas and N. Gordon). New York: Springer-Verlag, 197-223.

[16] West and Harrison (1997) Bayesian Forecasting and Dynamic Models (2nd edition). New York: Springer-Verlag.

[17] Prado and West (2010) Time Series: Modelling, Computation and Inference. Baton Rouge: Chapman and Hall/CRC.

[18] Petris, Petrone and Campagnoli (2009) Dynamic Linear Models with R. New York: Springer.

[19] Kim, Shephard and Chib (1994) Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. Review of Economic Studies, 65, 361-393.

[20] Johannes and Polson (2009) MCMC methods for Financial Econometrics. In Handbook of Financial Econometrics (Eds Y. Ait-Sahalia and L. Hansen). Oxford: Elsevier, 1-72.

[21] Gamerman and Lopes (2006) MCMC: Stochastic Simulation for Bayesian Inference. Baton Rouge: Chapman and Hall/CRC.

[22] GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study. Journal of Business and Economic Statistics. Torben G. Andersen and Bent E. Sorensen. Vol. 14, No. 3, Jul., 1996, page

[23] Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study. Torben G. Andersen, Hyung-Jin Chung, Bent E. Sorensen. Journal of Econometrics Volume 91, Issue 1, July 1999, Pages 61-87 328 of 328-352.

[24] Estimation of Stochastic Volatility Models with Diagnostics. A.Ronald Gallant, David Hsieh, George Tauchen.Journal of Econometrics Volume 81, Issue 1, November 1997, Pages 159-192.

[25] Gallant, A. R., and Tauchen, G. (1996) ”Which Moments to Match?,” Econometric Theory, 12, 657-681.

[26] Tauchen, G., (1997) ”New Minimum Chi-Square Methods in Empirical Finance,” in Advances in Econometrics, Seventh World Congress, eds. D. Kreps, and K. Wallis, Cambridge UK: Cambridge University Press, 279-317.

[27] Gallant, A. R., and Long, J. R. (1997) ”Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Squared,” Biometrika, 84, 125-141.

[28] Markov chain Monte Carlo methods for stochastic volatility models. Siddhartha Chib,Federico Nardari, Neil Shephard.Journal of Econometrics Volume 108, Issue 2, June 2002, Pages 281-316.

[29] Sequential Monte Carlo Methods for Stochastic Volatility Models with Jumps.

[30] MULTIVARIATE STOCHASTIC VOLATILITY: A REVIEW (2006).Manabu Asai , Michael McAleer , Jun Yu.

[31] Closer Look at the Relation between GARCH and Stochastic Autoregressive Volatility. Jeff Fleming and Chris Kirby.Journal of Financial Econometrics, 2003, vol. 1, issue 3, pages 365-419.

[32] Forecasting Volatility in Financial Markets: A Review. SER-HUANG POON and CLIVE W. J. GRANGER.Journal of Economic Literature Vol. XLI (June 2003) pp. 478-539

[33] Marcucci Juri , ”Forecasting Stock Market Volatility with Regime-Switching GARCH Models”, 2005, Studies in Nonlinear Dynamics and Econometrics, Volume 9, Issue 4, Article 6.

[34] Multivariate GARCH models: a survey. Luc Bauwens, Sebastien Laurent, and Jeroen Rombouts. Journal of Applied Econometrics, 21/1, 79-109.

[35] GO-GARCH: a multivariate generalized orthogonal GARCH model.Roy van der Weide. Journal of Applied Econometrics Special Issue: Modelling and Forecasting Financial Volatility Volume 17, Issue 5, pages 549-564, September/October 2002.

[36] An Artificial Neural Network GARCH Model for International Stock Market Volatility, with R. Glen Donaldson, Journal of Empirical Finance, 4 (1), 17-46, 1997.

[37] An empirical comparison of GARCH option pricing models. Springer journal Review of Derivatives Research. (2005)

[38] Garch option pricing model. Mathematical Finance, Vol.5, No.1 (January 1995).

[39] Semiparametric Modeling of Implied Volatility. Springer (2005).

[40] GARCH Models: Structure, Statistical Inference and Financial Applications. Christian Francq and Jean-Michel Zakoian.Wiley, July 2010, ISBN: 978-0-470-68391-0.