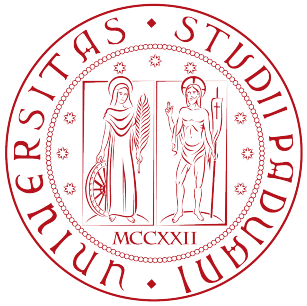


Università degli Studi di Padova

Department of Information Engineering

Master's Degree
in Automation Engineering

Structure and Identification of GARCH models

Student:
Matteo Terzi

Supervisor:
Em. Prof. Giorgio
Picci

Academic Year 2014/2015

ABSTRACT

In this thesis we shall study in some depth the structure and identification of various GARCH models. These models were introduced to describe the volatile behavior of empirical time series and to try to give an explanation of the phenomenon. In order to improve the predictions of volatility, various univariate and multivariate extensions of the basic Garch model have been introduced. In particular, our concern is focused on the state space framework, which overcomes the difficulties we may encounter studying Multivariate GARCH models.

We shall study the structure and the statistical properties of Garch models such as stochastic stability, predictive capacity and the identification procedures. A set of simulations are performed in order to verify if the empirical findings is consistent with the theory exposed. We shall see that state space models combined with a Garch framework outperform the standard GARCH models.

ACKNOWLEDGEMENTS

First and foremost I offer my gratitude to my supervisor, Em.Prof. Giorgio Picci, who followed me in the writing of this thesis. I consider it an honor to have worked with him, even if I am the only responsible for any error contained in this thesis.

In my daily work I have been supported by my girlfriend Giorgia, all my friends and my sisters, Eleonora and Melissa. I am especially thankful to Umberto and Mattia, called Carle by the friends, for having been close during this intense phase.

Finally, I am thankful to my parents who have given me the opportunity of an education from the best institutions and support throughout my life.

Dedicated to Giorgia, my companion in life.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 WHY FORECAST VOLATILITY?

In finance, volatility is a measure of how much the values of a temporal series may vary over time. It may for example be is a measure of the standard deviation of asset returns. Understanding the volatility behavior and, in particular, forecast it on the future, it has been intensely studied in mathematical finance, both empirically and theoretically. Modeling volatility is important for investment, risk management, trading, and for academic researcher to understand market dynamics. Risk management, almost entirely, consists on measuring the potential losses of a portfolio, and estimating these losses requires an estimate of future volatility. Indeed, the estimate of future volatility is important to derive option prices and is important to Value-at-Risk model in order to produce a risk measure. Another field of application, perhaps the most challenging, is trading. Option traders, evaluate their own strategies by forecasting the amount of volatility of the price underlying an option. Actually, in modern markets, it is also possible to trade volatility directly, through the use of derivative securities such as options and variance swaps and thus give a prediction on volatility becomes very important to traders to make profits. Volatility is also important in the bank industry. In fact, interest rates are volatile. High interest rate volatility not only can compromise the ability of the central bank to conduct monetary policy, but can also affect the ability of financial market to discern the monetary policy stance.

Until 1980's, the financial literature assumes a constant one-period forecast variance. Empirical observations show some peculiar characteristics of volatility:

- clustering and persistence phenomena for which there exist high volatility periods followed by low volatility periods;
- mean reverting (see for example Hillebrand);
- asymmetry, that is, negative return generates higher volatility than positive return.

Given these characteristics, the assumption of a constant one-period forecast is implausible. In particular, correlation of square returns on historical series causes clustering and persistence. Figure Figure 1 shows an example of this phenomenon. In or-

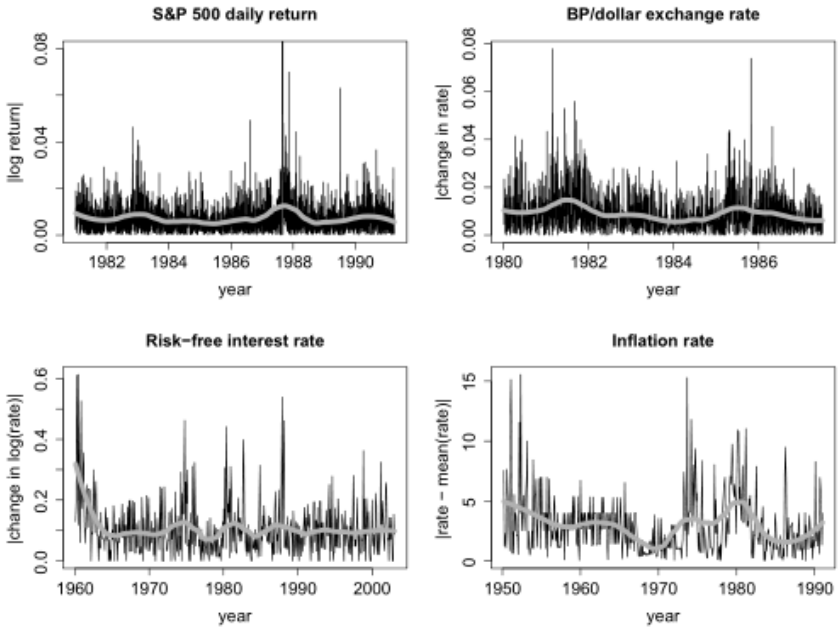


Figure 1: Examples of financial markets and economic data with time-varying volatility: (a) absolute values of S&P 500 log returns; (b) absolute values of changes in the BP/dollar exchange rate; (c) absolute values of changes in the log of the risk-free interest rate; (d) absolute deviations of the inflation rate from its mean

der to explain this evidence Engle [11] introduces a new class of stochastic processes called *AutoRegressive Conditional Heteroscedastic* (ARCH). These processes are zero mean, serially uncorrelated processes with constant uncondition variance and noncostant

variance conditional on the past. Successively, Bollerslev et al. [1] extended ARCH to *Generalized AutoRegressive Conditional Heteroscedastic* (GARCH) to permit the volatility to also depend on volatility past values. Both these papers show the consistency and normality of maximum or quasi-maximum likelihood estimators and, moreover provide a test to verify the presence of GARCH components on time series. A drawback of these models was their univariate structure. Indeed, in real scenarios, different economic variables are correlated with each other and an univariate modeling can be inappropriate. A multivariate framework is introduced by Engle and Kroner [12] to take in account possible variable correlations.

In parallel, other branches of mathematical finance, considered state space models as, for example Harvey [15]. State space models were used since 1960's (see [17]) in control engineering because of the powerful estimation algorithm called *Kalman filter*. Within the state space approach forecasting algorithm can be based on Kalman filtering.

Harvey, Ruiz, and Sentana [14] incorporate ARCH and GARCH disturbances in time series models with unobserved components, that is, in a state space approach, and then analyze the implications this has for estimation. Most state space models are however based on intuitive statistical grounds without a strong finance theory support. For example the Nelson-Siegel model proposed by Nelson and Siegel [20]. Its popularity is due to parsimony, ease of estimation and to the fact that there is some underlying economic interpretation in the three factors it is based on, which represent level, slope and curvature of the yield curve. An issue of this model is that it does not ensure, theoretically, absence of arbitrage. As an extension of Nielson-Siegel model, [8] introduce the Dynamic Nelson-Siegel (DNS) model by estimating the classical one with time-varying factors and model them using (V)AR specifications. In addition Diebold, Rudebusch, and Aruoba [9] rewrite the DNS model in state space form and shows the forecasting performances are better than those of standard time series models. However, Christensen, Diebold, and Rudebusch [6] derive the Nelson-Siegel model under absence of the riskless arbitrage assumption and introduce the Arbitrage Free Nelson-Siegel (AFNS) model, thereby reducing,

at least partially, the leak of absence of theoretical ground. Koopman, Mallee, and van der Wel [18] introduces the DNS-GARCH models, assuming time-varying volatility. In particular they assume the errors follow a GARCH dynamic. In such a way, it allows the model to capture latent exogenous shocks that affect the entire yield curve and are not captured by the three factor structure of the level, slope and curvature factors. This expansion increases the flexibility of the term structure model and enables it to better fit more complex shapes of the yield curve, as Koopman, Mallee, and van der Wel [18] show by plotting some fitted curves. They find that allowing for time-varying volatility significantly increases the likelihood value relative to the traditional DNS model. In this thesis we will extend DNS-GARCH models permitting to errors to follow also asymmetric GARCH. Indeed, as we already said, volatility reacts differently to positive shocks than negative. Thus, we aspect that allowing for asymmetric response of the variance of the common component to shocks turns out to increase in-sample fit of the time-varying volatility. We will consider the GJR-GARCH and E-GARCH for asymmetry. Moreover, it is reasonable to think that macroeconomic and financial variables influences volatility .

In this thesis we will analyze the structure and identification of various GARCH models. In particular we will point our attention to state space GARCH models. It is important to understand how good are the predictions and fitness with respect to other time series models, already known. Random walk forecasts turn out to be difficult to beat in the short term, as also noted by Duffee [10]. For the medium and long term the DNS models with time-varying volatility components seem to be able to significantly outperform the naive forecasting method at the short end of the yield curve. However, in the long end of the curve the random walk forecasts are relatively accurate and stay very hard to beat. As well known, parsimony is it very important in identification. It turn out that the DNS model with a common shock component in the factors, which has the smallest number of parameters among the time-varying volatility models, performs best when forecasting is concerned. Moreover, the smaller number of parameters the smaller the variance of estimate.

The thesis is organized as follows. Chapter 2 introduces univariate and multivariate GARCH models, state space models with an introduction of Gaussian state space model and Kalman filter. Chapter 3 studies stochastic stability of all the GARCH models presented. Chapter 4 discusses identification algorithms showing how a state space approach is powerful instrument for estimation. Finally, Chapter 5 compares the performance, in term of goodness of fit and forecasting, of the presented models.

2

STYLIZED FACTS AND MODELS

FINANCIAL TIME SERIES

In this thesis we shall consider the mathematical description of financial time series using stochastic models. In contrast with engineering, modeling financial time series presents same difficulties which make their analysis more complex. Although there are some statistical regularities (called *stylized facts*) which are common to a large number of financial series quite independently on their nature, this complexity is due to the variety of instruments, such as stocks, interest rates etc.

Let p_t denote the price of an asset at time t and let $y_t = \log(p_t/p_{t-1})$ the log return. The reason why log returns are often used is that they are independent of monetary units, which facilitates comparisons between assets. The are some empirical properties to be noticed of financial series that explains the theory used is this thesis.

A. **Nonstationarity of price series.** Sample paths of prices are close to a random walk. Moreover, they are compatible with second-order stationarity but not strict stationary assumption. Figure 2 shows log returns of CAC index.

B. **Absence of autocorrelation for the price variations**. The series of price variations generally display small autocorrelation, that is, they are close to white noise.

C. **Autocorrelations of squared price returns**. In general squared return y_t^2 are strongly correlated. This property implies that the white noise is not an i.i.d. process for example be a martingale difference. Figure 3 shows this property.

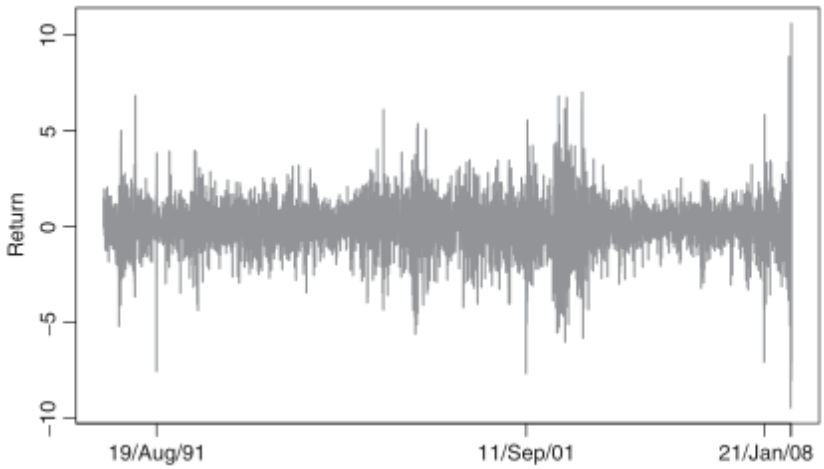


Figure 2: CAC 40 log-returns.

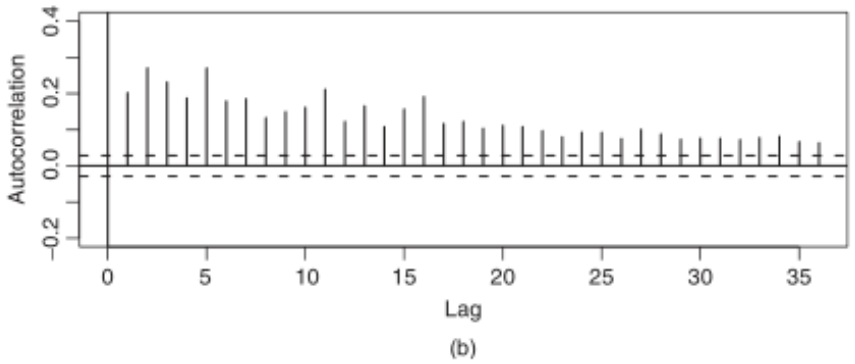


Figure 3: CAC 40: sample squared returns (January 2, 2008 to October 15, 2008).

D. **Volatility clustering**. In general high-volatility subperiods are followed by low-volatility periods. This property is recurrent but not periodic. In particular, it is not compatible with homoscedastic marginal distribution for the returns.

E. **Fat-tailed distributions**. It can be shown that empirical distribution does not resemble a Gaussian distribution. In particular, the densities have fat tails and are sharply picked at zero (*leptokurtic*). To measure this fact, there exists the kurtosis coefficient defined as the ratio of the sample forth-order moment to the squared sample variance.

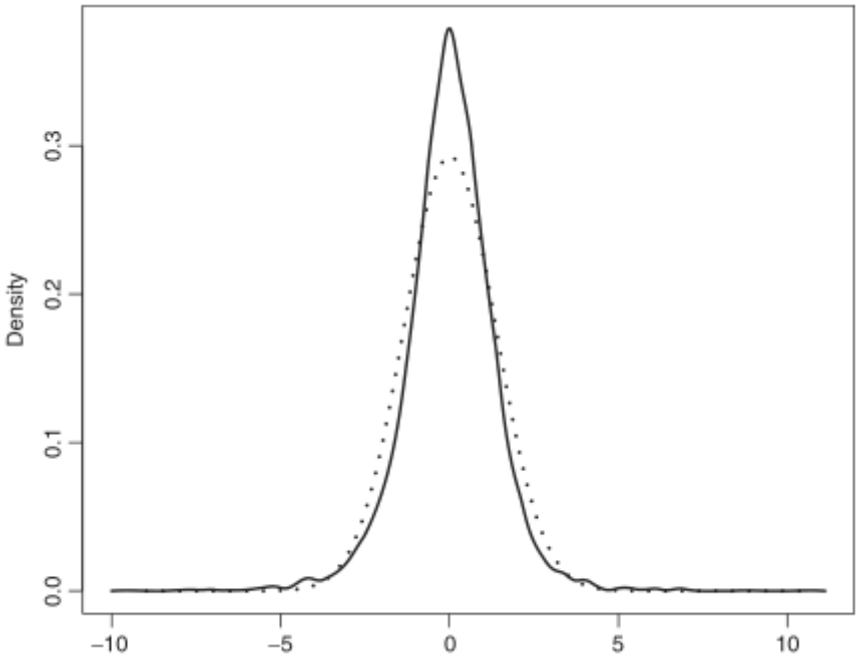


Figure 4: Kernel estimation of the CAC 40 return density versus Gaussian density (dotted line).

F. **Leverage effects**. This fact was noticed by Black, and involves an asymmetry of the impact of past positive and negative values on the current volatility. Negative returns (corresponding to price decreases) tend to increase volatility by a larger amount than positive returns (price increases) of the same magnitude.

G. **Seasonality**.

As already said, this empirical evidence is incompatible with standard time series, for example ARMA models. The fact that large absolute returns tend to be followed by large absolute returns is not compatible with costant conditional variance assumption. This is called *conditional heteroscedasticity*:

$$Var(y_t \mid \mathbf{y}^t) \neq \text{ const.}$$

where $\mathbf{y}^t$ is the strict past of $\{y_t\}$. Suppose y_t are interest rates or log-returns. In order to account for the peculiarities of financial series let consider the model

$$y_t = \sigma_t \varepsilon_t \tag{1}$$

where

1. σ_t is measurable with respect to a σ-field $\mathcal{F}_{t-1}$;
2. $\{\varepsilon_t\}$ is a iid zero mean, unit variance process, independent of $\mathcal{F}_{t-1}$ and $\{y_s,\ s < t\}$;
3. $\sigma > 0$.

All this implies that $\{y(t)\}$ is a d-martingale, which in particular implies that

$$E(y_t \mid \mathcal{F}_{t-1}) = 0, \qquad E(y_t^2 | \mathcal{F}_{t-1}) = \sigma_t^2$$

where the random variable σ_t^2 (or σ_t) is called *volatility* of y_t. Moreover we have

$$E(y_t) = E(\sigma_t)E(\varepsilon_t) = 0$$

for all t and

$$\begin{aligned} Cov(y_t, y_{t-s}) &= Ey_t y_{t-s} = E(E(y_t y_{t-s}) \mid \mathcal{F}_{t-1}) \\ &= E(\varepsilon_t)E(\sigma_t y_{t-s}) = 0, \quad \forall s > 0, \end{aligned}$$

that is, $\{y_t\}$ is second-order white noise process. Finally, the kurtosis coefficient of y_t is related to that of ε_t, denoted κ_ε and it is given by

$$\frac{Ey_t^4}{(Ey_t^2)^2} = \kappa_\varepsilon \left(1 + \frac{Var(\sigma_t^2)}{\left(E\sigma_t^2\right)^2}\right).$$

.

UNIVARIATE GARCH

An important class of time series that verifies the properties above are the so-called GARCH models.

An early definition, Engle [11] is the ARCH model, while the GARCH model is introduced by [1] as a generalized ARCH. For the sake of clarity, we will give only the definition of a general GARCH(p,q).

Definition 1 (GARCH(p,q) process)**.** A process y_t is called a GARCH(p,q) process if it satisfies:

A. $E(y_t|y_s, s < t) = 0, \quad \forall s > 0,$

B. There exist constants ω, α_i, $i = 1, \dots, q$ and β_j, $j = 1, \dots, p$ such that

$$\sigma_t^2 = Var(y_t \mid y_s, s < t) = \omega + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{2}$$

The last equation is rewritten as

$$\sigma_t^2 = \omega + \mathcal{A}(z^{-1})y_t^2 + \mathcal{B}(z^{-1}\sigma_t^2), \quad t \in \mathbb{Z}$$

where

$$\mathcal{A}(z^{-1}) = \sum_{i=1}^{q} \alpha_i z^{-i}, \qquad \mathcal{B}(z^{-1}) = \sum_{i=1}^{p} \beta_i z^{-i}$$

The definition in Definition 1 does not give guarantees that GARCH(p,q) processes are well-defined. On needs to show that the formula defines a process of finite variance of all t. This will be addressed later in Theorem 1.

Let $\{\varepsilon_t\}$ be an iid sequence with zero mean and unit variance.

Theorem 1. The process $\{y_t\}$ GARCH(p,q) if

$$\begin{cases} y_t = \sigma_t \varepsilon_t \\ \sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \end{cases} \tag{3}$$

where $\alpha_i \geqslant 0$, $\beta_j \geqslant 0$ and $\omega > 0$.

Proof. The Definition 1 implies that process $\{y_t\}$ is a d-martingale with a variance following the law

$$\sigma_t^2 = \omega + \mathcal{A}(z^{-1})y_t^2 + \mathcal{B}(z^{-1}\sigma_t^2), \quad t \in \mathbb{Z}$$

The process $\sigma_t\varepsilon_t$ is a d-martingale. In fact, σ_t process depends on the past and ε_t is i.i.d. and thus it can be easily shown that a the product with a i.i.d is a d-martingale itself satisfying the conditions:

1. $\mathcal{F}_t \subset \mathcal{F}_{t+1}$;
2. y_t is $\mathcal{F}_t$-measurable;
3. $\mathrm{E}|y_t| = \mathrm{E}\sigma_t|\varepsilon_t| = \mathrm{E}\sigma_t\left\{\mathrm{E}|\varepsilon_t|\middle|\mathcal{F}_{t-1}\right\} = 0 < \infty$;
4. $\mathrm{E}\left(y_{t+1}|\mathcal{F}_t\right) = \sigma_{t+1}\mathrm{E}\left(\varepsilon_{t+1}|\mathcal{F}_t\right) = 0$.

And these conditions are compatible with Definition 1. Finally, the existence conditions are related to the positivity constraint on σ^2. In order to guarantee σ_t^2 to be positive, all coefficients $\omega, \alpha_i, \beta_j$ must be positive. In particular, ω is strictly positive in order to avoid the eventuality $\sigma_t^2 = 0$. □

Figure 5 shows a simulation of a GARCH(1,1). It can be proved that the forth-order moment doesn't exist. This is reflected by the presence of large absolute values. In addition, when $\beta \to 1$, a shock on volatility has a persistent effect. Here we briefly give a condition for the existence of $2m$-th moment of a GARCH(1,1)

Theorem 2 (Order existence)**.** The $2m$-th order exists if and only if

$$\mu(\alpha, \beta, m) = \sum_{j=0}^{m} \binom{m}{j} a_j \alpha^j \beta^{m-j} < 1$$

where

$$\omega = 1, \quad a_j = \prod_{i=1}^{j}(2j-1), \quad j = 1...m$$

Then the second-order moment exists if and only if $\beta + \alpha < 1$ and it results $\mathbb{E}y_t^2 = \dfrac{\omega}{1-\alpha-\beta} > \omega$. In Section 2.0.1, we will see that this is also the condition for second-order stationary.

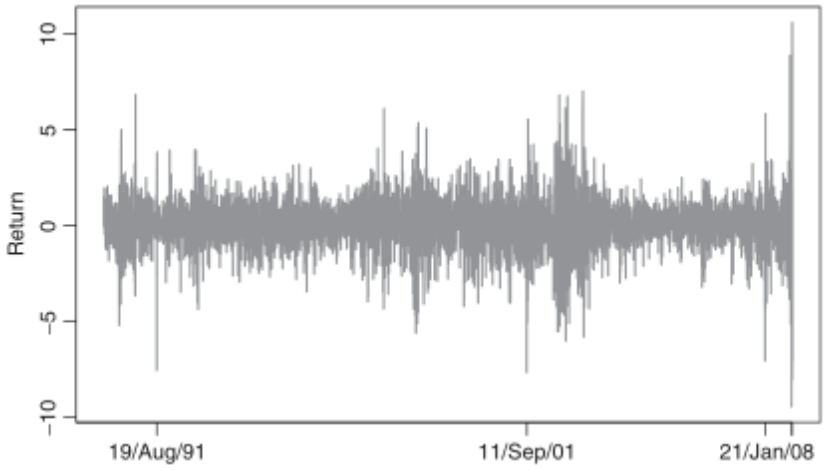


Figure 5: Simulation of 500 observation of GARCH(1,1) process with $\omega = 1$, $\alpha 00.7$, $\beta = 0.2$ and $\varepsilon_r \sim N(0,1)$.

A GARCH(1,1) process satisfies the squared correlation properties of financial series. Indeed,

$$E y_t^2 y_{t-s}^2 = C(\alpha + \beta)^s$$

where C is a constant.

2.0.1 *Stationarity and Ergodicity*

In this thesis, we will focus on GARCH(1,1) case since the model used on the simulations considers only a GARCH(1,1) model error

$$\begin{cases} y_t = \sigma_t^{\frac{1}{2}} \varepsilon_t \qquad \varepsilon_t \quad iid(0,1) \\ h_t = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 \end{cases} \tag{4}$$

with $\omega > 0$, $\alpha \geqslant 0$ and $\beta \geqslant 0$. The following theorem gives a necessary and sufficient condition for strict stationarity.

Theorem 3 (Strict stationarity)**.** If

$$\gamma := E \log\{\alpha \varepsilon_t^2 + \beta\} < 0,$$

then

$$h_t = \Big\{1 + \sum_{t=1}^{\infty} \Big(\alpha\varepsilon_{t-1}^2 + \beta\Big) \dots \Big(\alpha\varepsilon_{t-i}^2 + \beta\Big) \Big\}\omega$$

converges a.s. and the process $\{y_t\}$ is the unique strictly stationary solution of model (4) and it also ergodic. If $\gamma \geqslant 0$ and $\omega > 0$, there exists no strictly stationary solution.

The proof of stationarity shows that the sum

$$h_t = \Big\{ \sum_{t=0}^{\infty} \Big(\alpha\varepsilon_{t-1}^2 + \beta\Big) \dots \Big(\alpha\varepsilon_{t-i}^2 + \beta\Big) \Big\}\omega$$

converges a.s.. Being ε_t is i.i.d. and thus ergodic, the ergodicity of y_t follows by the relation

$$y_t = \sqrt{h_t}\varepsilon_t = \left\{ \sum_{t=0}^{\infty} \Big(\alpha\varepsilon_{t-1}^2 + \beta\Big) \dots \Big(\alpha\varepsilon_{t-i}^2 + \beta\Big) \omega \right\} \varepsilon_t$$

and Theorem A.2.

Proof. First we note that $\mathrm{E}\log^+(\alpha\varepsilon_t^2 + \beta) \leqslant \mathrm{E}(\alpha\varepsilon_t^2 + \beta) = \alpha + \beta$ and so γ is always defined. For $n \geqslant 1$ it can easily found

$$\begin{aligned} \sigma_t^2 = \omega\Big\{1 - \sum_{i=1}^{n} \Big(\alpha\varepsilon_{t-1}^2 + \beta\Big) \dots \Big(\alpha\varepsilon_{t-i}^2 + \beta\Big) \Big\} \\ + \Big(\alpha\varepsilon_{t-1}^2 + \beta\Big) \dots \Big(\alpha\varepsilon_{t-n-1}^2 + \beta\Big) \sigma_{t-n-1}^2 \\ := h_t(n) + \Big(\alpha\varepsilon_{t-1}^2 + \beta\Big) \dots \Big(\alpha\varepsilon_{t-n-1}^2 + \beta\Big) \sigma_{t-n-1}^2 \end{aligned} \tag{5}$$

Since the elements of the series are non-negative the limit process $h_t = \lim_{n\to\infty} h_t(n)$ exists and it assumes value in $[0, +\infty]$. We have to show it is finite. Letting n go to $+\infty$ in

$$h_t(n) = \omega + \Big(\alpha\varepsilon_{t-1}^2 + \beta\Big) h_{t-1}(n)$$

we obtain

$$h_t = \omega + \Big(\alpha\varepsilon_{t-1}^2 + \beta\Big) h_{t-1}$$

Now suppose that $\gamma < 0$. We have

$$\left\{ \left(\alpha\varepsilon_{t-1}^2 + \beta\right) \dots \left(\alpha\varepsilon_{t-n}^2 + \beta\right) \right\}^{1/n} = \exp\left\{ \frac{1}{n} \sum_{i=1}^{n} \log\left(\alpha\varepsilon_{t-i}^2 + \beta\right) \right\} \to e^{\gamma} \quad \text{a.s.}$$

as $n \to \infty$ by the application of the strong law of large numbers. Applying the Cauchy rule, the series

$$h_t = \left\{ \sum_{t=0}^{\infty} \left(\alpha\varepsilon_{t-1}^2 + \beta\right) \dots \left(\alpha\varepsilon_{t-i}^2 + \beta\right) \right\} \omega$$

converges almost surely in $\mathbb{R}$. Then, the process

$$y_t = \sqrt{h_t}\varepsilon_t = \left\{ \sum_{t=0}^{\infty} \left(\alpha\varepsilon_{t-1}^2 + \beta\right) \dots \left(\alpha\varepsilon_{t-i}^2 + \beta\right) \right\} \omega \qquad (6)$$

is strictly stationary and ergodic (see Section A.1). If $\gamma \geqslant 0$ the sum

$$\sum_{i=1}^{n} \left(\alpha\varepsilon_{t-1}^2 + \beta\right) \dots \left(\alpha\varepsilon_{t-i}^2 + \beta\right) \Big\} \to +\infty \qquad \text{as } n \to \infty$$

and there exists no finite solution of Equation 4. □

In real scenario it is difficult to verify strict stationarity on generated data. Often it is easier to verify the second order condition. Thus, now we give a sufficient and necessary condition for Second-Order stationarity.

Theorem 4. Let $\omega > 0$. If $\alpha + \beta \geqslant 1$ there no exists no second-order stationary solution to the GARCH(1,1) model. If $\alpha + \beta < 1$, the process y_t defined by (4) is second-order stationary. In addiction, y_t is a second-order white noise process. The solution is unique.

We will show only the existence of the stationary solution. For a detailed proof see [13].

Proof. If we have a second-order solution of (4) then

$$E(y_t^2) = E\{E(y_t^2|y_s, s < t)\} = E(\sigma_t^2) = \omega + (\alpha + \beta)Ey_{t-1}^2$$

that is,

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

Hence, we must have $\alpha+\beta < 1$ since $\omega > 0$. Conversely, suppose $\alpha+\beta < 1$. Then there exists the limit sum h_t and we have

$$\begin{aligned} Ey_t^2 = Eh_t &= E\left\{1+\sum_{i=1}^{\infty}\left(\alpha\varepsilon_{t-1}^2+\beta\right)\ldots\left(\alpha\varepsilon_{t-i}^2+\beta\right)\right\}\omega \\ &= E\left\{1+\sum_{i=1}^{\infty}\left(E\left(\alpha\varepsilon_{t-1}^2+\beta\right)\right)^i\right\}\omega \\ &= \left\{1+\sum_{i=1}^{\infty}(\alpha+\beta)^n\right\}\omega \\ &= \frac{\omega}{1-(\alpha+\beta)} \end{aligned}$$

The condition $\alpha+\beta<1$ it is easy to verify practically. □

Analogously it can be shown the condition for stationary for GARCH(p,q). The proof of second-order stationarity follows hand in hand the proof of Theorem 4. Define the GARCH(p,q) models as

$$\begin{cases} y_t = \sigma_t\varepsilon_t \\ \sigma_t^2 = \omega+\sum_{i=1}^{q}\alpha_i y_{t-i}^2+\sum_{j=1}^{p}\beta_j\sigma_{t-j}^2 \end{cases} \tag{7}$$

where $\alpha_i \geqslant 0, \beta_j \geqslant 0$ and $\omega > 0$. The following theorem gives a necessary and sufficient condition for second order stationarity of GARCH(p,q) process.

Theorem 5. If there exists a GARCH(p,q) process, which is second-stationary, and if $\omega > 0$, then

$$\sum_{i=1}^{q}\alpha_i+\sum_{i=j}^{p}\beta_j < 1. \tag{8}$$

Conversely, if holds (8), the unique stricly stationary solution of model (7) is a weak white noise (and thus is second-order stationary).

Proof. Subsequent substitutions yields

$$\begin{aligned} h_t &= \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 h_{t-i} + \sum_{i=1}^{p} \beta_i h_{t-i} \\ &= \omega + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 \left(\omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 h_{t-i-j} + \sum_{i=1}^{p} \beta_i h_{t-i-j} \right) \\ &\quad + \sum_{j=1}^{p} \beta_j \left(\omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 h_{t-i-j} + \sum_{i=1}^{p} \beta_i h_{t-i-j} \right) \\ &\vdots \\ &= \omega \sum_{k=0}^{\infty} M(t,k) \end{aligned} \tag{9}$$

where

$$\begin{aligned} M(t,0) &= 1, \\ M(t,1) &= \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \end{aligned}$$

and in general

$$M(t,k+1) = \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 M(t-i,k) + \sum_{i=1}^{p} \beta_i M(t-i,k). \tag{10}$$

Since ε_t^2 is i.i.d., we have

$$E\left(M(t,k)\right) = E\left(M(s,k)\right) \qquad \text{for all} \quad k,t,s. \tag{11}$$

From Equation 10 and Equation 11 it follows

$$\begin{aligned}
\mathrm{E}\left(M(t,k+1)\right) &= \left(\sum_{i=1}^{q}\alpha_i + \sum_{i=1}^{p}\beta_i\right)\mathrm{E}\left(M(t,k)\right) \\
&\vdots \\
&= \left(\sum_{i=1}^{q}\alpha_i + \sum_{i=1}^{p}\beta_i\right)^{k+1}\mathrm{E}\left(M(t,0)\right) \\
&= \left(\sum_{i=1}^{q}\alpha_i + \sum_{i=1}^{p}\beta_i\right)^{k+1}.
\end{aligned}$$

Finally,

$$\begin{aligned}
\mathrm{E}\left(y_t^2\right) &= \omega\mathrm{E}\left(\sum_{k=0}^{\infty}M(t,k)\right) \\
&= \omega\sum_{k=0}^{\infty}\mathrm{E}\left(M(t,k)\right) \\
&= \frac{\omega}{1-\sum_{i=1}^{q}\alpha_i - \sum_{i=1}^{p}\beta_i}
\end{aligned}$$

if and only if

$$\sum_{i=1}^{q}\alpha_i - \sum_{i=1}^{p}\beta_i < 1,$$

and y_t^2 converges almost surely. □

2.1 MULTIVARIATE GARCH

In the real world scenario, financial variables are not independent but are correlated each other. This implies the use of multivariate processes. Now we define a multivariate GARCH process.

Definition 2. A N-dimensional multivariate GARCH process $\boldsymbol{\epsilon}_t|\mathcal{F}_{t-1}$[1] is given by

$$\boldsymbol{\epsilon}_t = \mathbf{z}_t \mathbf{H}_t^{\frac{1}{2}} \tag{12}$$

where $\mathbf{z}_t$ is a zero mean N-dimensional i.i.d. process with covariance matrix $\mathbf{I}_N$.

One easily obtains $E(\boldsymbol{\epsilon}_t|\Omega_{t-1}) = 0$ e $E(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'|\Omega_{t-1}) = \mathbf{H}_t$. The general extension of a GARCH(p,q) is the *vech model* proposed by Bollerslev et al. [2]. It uses vech(·) operator, which stacks all the non-redundant element of a matrix $N \times N$ in a vector $N(N+1)/2 \times 1$. Then the matrix $\mathbf{H}_t$ is transformed in

$$vech(\mathbf{H}_t) = \boldsymbol{\omega}^* + \sum_{i=1}^{q} \mathbf{A}_i^* vech(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}') + \sum_{j=1}^{p} \mathbf{B}_j vech(\mathbf{H}_{t-j}) \tag{13}$$

$$vech(\mathbf{H}_t) = \boldsymbol{\omega}^* + \mathbf{A}(z^{-1}) vech(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}') + \mathbf{B}(z^{-1}) vech(\mathbf{H}_{t-1}) \tag{14}$$

where $\boldsymbol{\omega}^* = vech(\boldsymbol{\Omega})$ is a vector of $N(N+1)/2 \times 1$ parameters while $\mathbf{A}_i^*$ e $\mathbf{B}_j^*$ are $N(N+1)/2 \times N(N+1)/2$ matrices. This model has two drawbacks: in order to guarantee $\mathbf{H}_t > 0$ is it necessary impose constraints on $\mathbf{A}$ and $\mathbf{B}$, and, more important, there are $\frac{N(N+1)}{2}\left[1 + (p+q)\frac{N(N+1)}{2}\right]$ parameters to be estimated. For instance, with $N = 5$ and $p = q = 1$ there are 465 parameters to be estimated. This fact makes his use inconvenient.

A more general model is the *BEEK model* proposed by Engle and Kroner [12] and Kroner. It intrinsically imposes $\mathbf{H}_t$. Let us consider the following model

$$\mathbf{H}_t = \mathbf{C}\mathbf{C}' + \sum_{k=1}^{K}\sum_{i=1}^{q} \mathbf{A}_{ik}\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-i}'\mathbf{A}_{ik}' + \sum_{k=1}^{K}\sum_{i=1}^{p} \mathbf{B}_{ik}\mathbf{H}_{t-i}\mathbf{B}_{ik}' \tag{15}$$

where $\mathbf{C}$,$\mathbf{A}_{ik}$ e $\mathbf{B}_{ik}$ are $N \times N$ matrices. The following proposition gives a condition for the positivity of $\mathbf{H}_t$ in a BEKK-GARCH(p,q).

1 $\mathcal{F}_{t-1}$ represents all the past information

Theorem 6. If $\mathbf{H}_0, \mathbf{H}_1, ..., \mathbf{H}_{p+1}$ are all definite positive, then the BEKK with $K = 1$ imposes $\mathbf{H}_t > 0$ for all ϵ_t values if $\mathbf{C}$ is full rank or if each $\mathbf{B}_i$ is full rank.

For stationarity conditionso for *vech model* and *BEKK model* see Engle and Kroner [12]. Unfortunately, also BEKK-GARCH has a number of parameters of order N^2 causing computation problems. Although, there are approaches of simplification for estimation, it is preferable use state space approaches.

LINEAL GAUSSIAN STATE SPACE MODELS AND THE KALMAN FILTER

Before introducing state space GARCH models we present LGSSM and Kalman filter, since they will be used in this thesis. The estimation procedure of this models is reported in Chapter 4. A state space models consists of a *state equation* and *observation equation*. While the state equation formulates the dynamics of the state variables, the observation equation relates the observed variables to the unobserved state vector. The state vector can contain trend, seasonal, cycle and regression components plus an error term. With these models can be used a very power instruments both for estimation and prediction of latent variable, named Kalman filter. Named after [17] the Kalman filter is a recursive algorithm that computes estimates for the unobserved components at time t, based on the available information at the same date. Suppose to consider the state space model

$$\begin{cases} \mathbf{x}(t+1) = A_t\mathbf{x}(t) + B_t\mathbf{v}(t) \\ \mathbf{y}(t) = C_t\mathbf{x}(t) + \mathbf{w}(t) \end{cases} \tag{16}$$

$$E(\mathbf{v}(t)\mathbf{v}(\tau)') = \begin{cases} Q & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases}$$

$$E(\mathbf{w}(t)\mathbf{w}(\tau)') = \begin{cases} R & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases}$$

with the notation:

- $\mathbf{y}(t)$ denotes an $p \times 1$ vector of observable variables.

- $\mathbf{x}(t)$ denotes a possibly unobserved $n \times 1$ state vector.
- $\mathbf{v}(t)$ and $\mathbf{w}(t)$ are white noise i.i.d with zero mean.

In addiction we will make the assumptions:

- The disturbances $\mathbf{v}(t)$ and $\mathbf{w}(t)$ are assumed to be uncorrelated at all leads and lags

$$E(\mathbf{v}(t)\mathbf{w}(\tau)') = 0 \qquad \text{for all } t \text{ and } \tau$$

- The initial state and the disturbances are uncorrelated, i.e.

$$E\{\mathbf{x}_0 \left[\mathbf{v}(t)', \mathbf{w}(t)'\right]\} = 0, \qquad \forall t \geqslant t_0$$

The matrices A_t, B_t, C_t, Q_t and H_t are in general time variant. For our application they are constant. Usually, at least some of the elements of the system matrices A_t and B_t in the state equation and C_t , Q_t and R_t in the observation equation depend on a vector θ unknown parameters.

The Kalman filter and smoother

Once a model is written into state space form, the Kalman filter can be employed to compute optimal forecasts of the mean and covariance matrix of the normally distributed state vector $\mathbf{x}_{t+1}$, based on the available information through time t. More precisely it produces estimate minimizing the mean squared error. The Kalman filter can be used for estimation by *filtering* and *smoothing*: filtering uses only the information up to time t while filtering uses the entire set of information in the sample.

Filtering

Now let assume A, B, C, R and Q are constants. Let us denote

$$\mathcal{F}_t \equiv (\mathbf{y}_t, \mathbf{y}_{t-1}, ..., \mathbf{y}_1, \mathbf{x}_t, \mathbf{x}_{t-1}, ..., \mathbf{x}_1)$$

and denote

$$\hat{\mathbf{x}}_{t+1|t} \equiv \widehat{E}(\mathbf{x}_{t+1}|\mathcal{F}_t)$$
$$\hat{\mathbf{x}}_{t|t} \equiv \widehat{E}(\mathbf{x}_t|\mathcal{F}_t)$$

$$\mathbf{P}_{t+1|t} \equiv E\left(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t}\right)\left(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t}\right)'$$

It can be demonstrated that the equation describing the Kalman filter are

$$\hat{\mathbf{x}}_{t+1|t} = A\hat{\mathbf{x}}_{t|t-1} + K(t)\left(\mathbf{y}(t) - C\hat{\mathbf{x}}_{t|t-1}\right)$$

$$K(t) = AP_{t|t-1}C'(CP_{t|t-1}C' + R)^{-1}$$

$$P_{t+1|t} = A\left(P_{t|t-1} - P_{t|t-1}C'(CP_{t|t-1}C' + R)^{-1}CP_{t|t-1}\right) + Q$$

The $N \times 1$ vector

$$\boldsymbol{\varepsilon}_t = \mathbf{y}_t - E(\mathbf{y}_t|\mathcal{F}_{t-1}) = \mathbf{y}_t - C\hat{\mathbf{x}}_{t|t-1}$$

is the one-step ahead prediction error of $\mathbf{y}_t$ given $\mathcal{F}_{t-1}$ and it is called *innovation* since

$$E(\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}) = E(\boldsymbol{\varepsilon}_t), \qquad Cov(\mathbf{y}_s, \boldsymbol{\varepsilon}_t) = 0$$

for $s = 1, \dots, t-1$. In general the prediction error $\boldsymbol{\varepsilon}_t$ is a d-martingale.

It is also important to understand in which conditions the Kalman filter admits a steady-state representations. It is immediate to see that the process is stationary in $t \geqslant 0$ if only if A is "stable". In this case P_0 is the solution of Lyapunov equation

$$P_0 = AP_0A' + Q$$

In these hypothesis we have $\Sigma(t) = E\mathbf{x}(t)\mathbf{x}(t)' = P_0, \forall t \geqslant 0$. Then we immediately have

Theorem 7. Let us denote $\hat{\mathbf{x}}(t) = \hat{\mathbf{x}}(t|t-1)$. If the matrix A in Equation 16 is asymptotically stable, then the estimate

$$\hat{\mathbf{x}}(t) = E(\mathbf{x}(t)|\mathbf{y}_0, \dots, \mathbf{y}(t-1))$$

converge, as $t_0 \to -\infty$, in quadratic mean, towards the limit $\hat{\mathbf{x}}_\infty(t)$, that is the Wiener-Kolmogorov predictor of $\mathbf{x}(t)$ conditional to the infinite past of the process $\mathbf{y}$. Them then there exists the limit

$$\lim_{t_0 \to -\infty} P(t) = P_\infty$$

Finally, P_∞ is the solution of A.R.E.

$$P = A\left[P - PC'(CPC' + R)^{-1}CP\right]A' + Q$$

It must be noticed that it is true only if $\boldsymbol{v}(t)$ and $\boldsymbol{w}(t)$ are uncorrelated. Otherwise we have to substitute A with $F = A - SR^{-1}C$ and Q with $\tilde{Q} = Q - SR^{-1}S'$ where $S = E\boldsymbol{v}(t)\boldsymbol{w}(t)'$.
The estimate $\hat{\boldsymbol{x}}_\infty(t)$ satisfies the equation (steady-state)

$$\hat{\boldsymbol{x}}_\infty(t+1) = A - K_\infty C\hat{\boldsymbol{x}}_\infty(t) + K_\infty \boldsymbol{y}_t \tag{17}$$

If Equation 17 represents a completely reachable dynamic system and the spectrum $S_{\boldsymbol{y}}(e^{j\omega}) > 0$, then all the eigenvalues are inside the unit circle and the steady-state predictor is asymptotically stable.

In general we have the following theorem (if $S = 0$).

Theorem 8 (Fundamental theorem of Kalman filter)**.** Necessary and sufficient condition for

1. there exists a unique solution $\bar{P} = \bar{P}'$ of A.R.E.,
2. $\bar{P}$ is stabilizing,
3. $\lim_{t\to\infty} P(t) = \bar{P}$ for every $P_0 = P_0' \geqslant 0$, is that (A,C) è detectable and (A,Q) is stabilizable.

If we assume stationarity, the initial conditions is set as

$$E\boldsymbol{x}(0|-1) = E(\boldsymbol{x}) = 0$$

$$P(0|-1) = P_0$$

where P_0 is the solution of

$$P_0 = AP_0A' + Q$$

In general, for non-stationary process is a good procedure to choice $P_0 = kI$ with k very large. In the steady-state, the Kalman predictor is

$$\hat{\boldsymbol{x}}(t+1) = A\hat{\boldsymbol{x}}(t) + K_\infty \boldsymbol{e}(t) \tag{18}$$

$$\boldsymbol{y}(t) = C\hat{\boldsymbol{x}}(t) + \boldsymbol{e}(t) \tag{19}$$

where $\boldsymbol{e}(t) = \boldsymbol{y}(t) - C\boldsymbol{x}(t)$ define the steady-state innovation.

In case of non-Gaussian disturbances, the Kalman filter is no longer guaranteed to yield the conditional mean of the state vector. However, it nevertheless represents an optimal estimator in the sense that no other linear estimator has a smaller MSE. For a detailed discussion see Picci [21].

2.2 STATE SPACE GARCH

In parallel to multivariate GARCH, state space framework is considered. Ruiz e Sentana (1992) Harvey et al. [14] introduces a state space model with ARCH/GARCH errors. Given the observation vector $N \times 1$ y_t we consider the model without regression factors

$$\boldsymbol{x}_t = A\boldsymbol{\alpha}_{t-1} + B\eta_t + \boldsymbol{\nu}_t \tag{20}$$

$$\boldsymbol{y}_t = C\boldsymbol{\alpha}_t + D\boldsymbol{\epsilon}_t + \boldsymbol{w}_t \tag{21}$$

where $\boldsymbol{x}_t$ is $m \times 1$ state vector. The matrices C $(N \times m)$ e A $(m \times m)$. The disturbances $\boldsymbol{\nu}_t$ $(N \times 1)$ and $\boldsymbol{\omega}_t$ $(N \times 1)$ are $\boldsymbol{\nu}_t \sim \text{NID}(0, Q)$ e $\boldsymbol{w}_t \sim \text{NID}(0, R)$. Moreover ϵ_t, $\boldsymbol{\nu}_t$ and $\boldsymbol{w}_t$ are mutually independent. The GARCH effect is introduced through the scalar disturbances ϵ_t e η_t:

$$\epsilon_t = h_t^{1/2}\eta_{1t} \qquad \text{e} \qquad \eta_t = q_t^{1/2}\eta_{2t}$$

where $\eta_{1t} \sim \text{NID}(0, 1)$ and $\eta_{2t} \sim \text{NID}(0, 1)$ are uncorrelated. In addition we have

$$h_t = \omega_1 + \alpha_1\epsilon_{t-1}^2 + \beta_1 h_{t-1} \tag{22}$$

$$q_t = \omega_2 + \alpha_2\eta_{t-1}^2 + \beta_2 q_{t-1} \tag{23}$$

It is assumed $\alpha_1 + \alpha_2 < 1$ e $\beta_1 + \beta_2 < 1$. The estimation procedure is in Chapter 4. However, see Harvey, Ruiz, and Sentana [14] for a complete discussion.

2.3 DNS-GARCH MODELS

In this section we will introduce DNS state space model. These model are particular linear Gaussian models introduced to explain evidences on the behavior of the yields. We chose to use this model rather than general Gaussian state space model because they are, how we will see, suitable for explaining empirical facts.

2.3.1 *Dynamic Nelson-Siegel Model*

Diebold and Li in [8] introduce the DNS model to model the yield curve,period-by-period, as a three-dimensional parameter evolving dynamically. The N yields $y_t(\tau_i)$ for $i = 1, \dots, N$ at time $t = 1, \dots, T$, where τ_i is the maturity time is modeled by

$$y_t(\tau_i) = \beta_{1t} + \beta_{2t}\left(\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i}\right) + \beta_{3t}\left(\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i}\right) + \varepsilon_{i,t} \tag{24}$$

$$\varepsilon_{i,t} \sim N(0, \sigma^2 I_N),$$

where the coefficients β_{it} represent the factors level, slope and curvature, respectively. The parameter λ_t affects the exponential decay rate; small values of λ_t produce slow decay and imply a better fit of the curve at long maturities, while large values of λ_t produce fast decay and imply a better fit the curve at short maturities. Furthermore, it also governs where the loading on β_{3t} achieves its maximum.

The challenge of yield curve modelling is to identify a framework under which the modelled yield curves will provide a sufficiently accurate fit to endless permutations of possible yield curve shapes and structures without over complicating the solution in the presence of assumed additive observation error. The Nelson Siegel three factor model makes a valuable contribution to this pursuit through the use of an elegant and easily interpreted method. shapes and structures without over complicating the solution in the presence of assumed additive observation error.

To give an interpretations of factors β_{it} we consider the behavior over the maturity time τ. Actually, for high values of τ we have

$$y_t(\tau_i) \to \beta_{1t}$$

and thus, β_{1t} may be viewed as the long-term factor, while the loading factor

$$\left(\frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i}\right) \to 0$$

meaning that β_{2t} can be viewed as the short-term factor. At last, the factor loading of β_{3t}

$$\left(\frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i}\right) \to 0$$

as maturity is high or low, and this means β_{3t} affects the mid-term of yield curve. To estimate this model, [9] introduce a unified state-space modeling approach that permits to simultaneously fit the yield curve at each point in time and estimate the underlying dynamics of the factors. To specify the autocorrelations of the three coefficients they introduce the state equation of dimension 3

$$\boldsymbol{x}_{t+1} = (I-\Phi)\boldsymbol{\mu} + \Phi\boldsymbol{x}_t + \boldsymbol{v}_t \qquad \boldsymbol{v}_t \sim \text{NID}(0, Q), \tag{25}$$

where $\mathbf{x}_{t+1} = (\beta_{1t}, \beta_{2t}, \beta_{3t})$, $\boldsymbol{\mu}$ is a 3×1 costant vector, Φ (3×3) is the coefficient matrix, $\boldsymbol{v}_t$ is the disturbance error with variance Q e initial condition $\mathbf{x}_0 \sim N(\boldsymbol{\mu}, \Sigma_x)$ with variance Σ_x that solves $X - \Phi X \Phi' = Q$. The measurement equation is given by

$$\boldsymbol{y}_t = C(\lambda)\boldsymbol{x}_t + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim \text{NID}(0, R), \qquad t = 1, \ldots, T, \tag{26}$$

with the observations vector $\boldsymbol{y}_t = \left[y_t(\tau_1), ..., y_t(\tau_N)\right]'$, the disturbances vector $\boldsymbol{w}_t = \left[w_{1t}, ..., w_{Nt}\right]'$, and the matrices $\mathbf{C}(\lambda)$ is defined by

$$C_{ij}(\lambda) = \begin{cases} 1, & j = 1 \\ (1-e^{-\lambda\cdot\tau_i})/\lambda\cdot\tau_i, & j = 2 \\ (1-e^{-\lambda\cdot\tau_i} - \lambda\cdot\tau_i e^{-\lambda\cdot\tau_i}/\lambda\cdot\tau_i), & j = 3 \end{cases}$$

The parameter μ carries information about the mean of process $\{x(t)\}$. In general, without loss of generality, we can suppose the $\{y(t)\}$ and therefore $\{x(t)\}$ is a zero mean process. This assumption allows us to estimate one parameter less. The model to be estimated become

$$\boldsymbol{x}_{t+1} = A\boldsymbol{x}_t + \boldsymbol{\nu}_t \qquad \boldsymbol{\nu}_t \sim NID(0, Q), \tag{27}$$

$$\boldsymbol{y}_t = C(\lambda)\boldsymbol{x}_t + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim NID(0, R), \qquad t = 1, \ldots, T, \tag{28}$$

2.3.2 *DNS-GARCH*

Koopman in [18] introduces the GARCH component as

$$\begin{aligned} \boldsymbol{y}_t &= C(\lambda)\boldsymbol{x}_t + \boldsymbol{\varepsilon}_t \\ \boldsymbol{\varepsilon}_t &= D\varepsilon_{1t} + \boldsymbol{w}_t \end{aligned} \tag{29}$$

where D is $N \times 1$ vector, ε_t^* is a scalar disturbance and $\boldsymbol{\omega}_t$ is a $N \times 1$ disturbances vector. For identification issues we can impose $DD' = 1$. The disturbances are mutually independent and their distributions are given by

$$\varepsilon_{1t} \sim NID(0, h_t), \qquad \boldsymbol{w}_t \sim NID(0, R), \qquad t = 1, ..., T,$$

where R typically diagonal while h_t is given by

$$h_{t+1} = \omega_1 + \alpha_1 \varepsilon_{1t}^2 + \beta_1 h_t \qquad t = 1, ..., T, \tag{30}$$

with ω, α, β unknown e such that to satisfy the stationarity. The matrix variance of $\boldsymbol{y}_t$ is $C(\lambda)\Sigma_x C(\lambda)' + \Sigma_\varepsilon(h_t)$ where Σ_x satisfies $A\Sigma_x A' - \Sigma_x = Q$.

As an alternative, the volatility component can be incorporated in the state equation. In this case, the volatility component indirectly influences the yields through the state as

$$\begin{aligned} \boldsymbol{x}_{t+1} &= A\boldsymbol{x}_t + \boldsymbol{\eta}_t \\ \boldsymbol{\eta}_t &= B\varepsilon_{2t} + \boldsymbol{\nu}_t \\ \boldsymbol{\nu}_t \sim N(0, Q), &\qquad \varepsilon_{2t}|\mathcal{F}_{t-1} \sim N(0, q_t) \end{aligned} \tag{31}$$

where B and $\boldsymbol{\nu}_t$ are 3×1 vectors of loadings and noise terms, and ε_{2t} carries the garch component and q_t has the formulation

$$q_{t+1} = \omega_2 + \alpha_2 \varepsilon_{2t}^2 + \beta_2 q_t \qquad t = 1, ..., T.$$

The matrices B and D take into account the sensitivity of the different yields in respect with the GARCH disturbance. It has been empirically found that short maturity yields are more sensitive to common shocks than long maturity yields.

In the following part of this section we shall give the conditions for stationarity of DNS-XGARCH model defined by

$$\begin{cases} \boldsymbol{x}_{t+1} = A\boldsymbol{x}_t + B\varepsilon_{2,t} + \boldsymbol{\nu}_t \\ h_t = \omega + \alpha\varepsilon_{2,t-1}^2 + \beta h_{t-1} \qquad t = 1, ..., T, \\ \boldsymbol{y}_t = C(\lambda)\boldsymbol{x}_t + \boldsymbol{w}_t, \end{cases} \tag{32}$$

$$\boldsymbol{\nu}_t \sim N(0, Q), \quad \varepsilon_{2,t} \mid \mathcal{F}_{t-1} \sim N(0, h_t), \quad \boldsymbol{w}_t \sim NID(0, R)$$

Theorem 9. Let $\omega > 0$, $\alpha \geqslant 0$, $\beta \geqslant 0$ and suppose the initial time to be $t_0 = 0$. The the processes $\{\boldsymbol{x}(t)\}$ and $\{\boldsymbol{y}(t)\}$ are jointly asymptotically second-order stationary as $t \to \infty$ if and only if are satisfied

1. all the eigenvalues of A is strictly less than 1,
2. $\alpha + \beta < 1$, that is, the process $\varepsilon_{2,t}$ is second-order stationary.

The variance matrix $\bar{\Sigma}_x := \lim_{t\to\infty} \Sigma_x(t)$ is the asymptotic state variance matrix.

Proof. The proof consists on showing that the matrix $\Sigma_x(t)$ converges at $\bar{\Sigma}_x$.

The variance matrix of the state process at $t + 1$ is

$$\Sigma_x(t+1) = A\Sigma_x(t)A' + B\Sigma_\varepsilon(t)B' + Q$$

where

$$\begin{aligned} \Sigma_\varepsilon(t) = E\varepsilon_t^{*2} = Eh_t^2 &= \omega + \alpha E\varepsilon_{2,t-1}^2 + \beta Eh_{t-1} \\ &= \omega + \alpha\Sigma_\eta(t-1) + \beta\Sigma_\varepsilon(t-1) \\ &= \omega + (\alpha+\beta)\Sigma_\varepsilon(t-1). \end{aligned}$$

Thus, we have

$$\Sigma_\varepsilon(t) = (\alpha+\beta)^t\Sigma_\varepsilon(0) + \omega\sum_{i=0}^{t-1}(\alpha+\beta)^i \tag{33}$$

and we obtain

$$\begin{aligned}\Sigma_x(t) &= A^t\Sigma_x(0)(A')^t \\ &\quad + \sum_{k=0}^{t-1} A^{t-1-k}B\Sigma_\varepsilon(k)B'(A')^{t-1-k} + \sum_{k=0}^{t-1} A^{t-1-k}Q(A')^{t-1-k}.\end{aligned}$$

Using (33) it follows

$$\begin{aligned}\Sigma_x(t) &= A^t\Sigma_x(0)(A')^t \\ &\quad + \sum_{k=0}^{t-1} A^{t-1-k}B\left\{(\alpha+\beta)^k\Sigma_\eta(0) + \omega\sum_{i=0}^{k-1}(\alpha+\beta)^i\right\}B'(A')^{t-1-k} \\ &\quad + \sum_{k=0}^{t-1} A^{t-1-k}\Sigma_\nu(A')^{t-1-k}\end{aligned}$$

$$\begin{aligned}\Sigma_x(t) = A^t\Sigma_x(0)(A')^t &+ \sum_{k=0}^{t-1} A^{t-1-k}B(\alpha+\beta)^k\Sigma_\eta(0)B'(A')^{t-1-k} \\ &+ \omega\sum_{k=0}^{t-1}\sum_{i=0}^{k-1}(\alpha+\beta)^iA^{t-1-k}BB'(A')^{t-1-k} \\ &+ \sum_{k=0}^{t-1} A^{t-1-k}\Sigma_\nu(A')^{t-1-k}\end{aligned}$$

Now suppose that $\alpha+\beta < 1$. This implies

$$\begin{aligned}(\alpha+\beta)^k &< 1 \qquad \forall k \geqslant 0 \\ \sum_{i=0}^{k-1}(\alpha+\beta)^i &< \sum_{i=0}^{\infty}(\alpha+\beta)^i = \frac{1}{1-(\alpha+\beta)}\end{aligned}$$

and hence it results

$$\begin{aligned}\Sigma_x(t) < A^t\Sigma_x(0)(A')^t &\\ + \sum_{k=0}^{t-1} A^{t-1-k}B\Sigma_\eta(0)B'(A')^{t-1-k} &\\ + \frac{\omega}{1-(\alpha+\beta)}\sum_{k=0}^{t-1} A^{t-1-k}BB'(A')^{t-1-k} &\\ + \sum_{k=0}^{t-1} A^{t-1-k}\Sigma_\nu(A')^{t-1-k} &\end{aligned}$$

Now suppose A is stable. We have, as $t \to \infty$,

$$\begin{aligned}\lim_{t\to\infty}\Sigma_x(t) \leqslant \sum_{k=0}^{\infty} A^{t-1-k}B\Sigma_\eta(0)B'(A')^{t-1-k} &\\ + \frac{\omega}{1-(\alpha+\beta)}\sum_{k=0}^{\infty} A^{t-1-k}BB'(A')^{t-1-k} &\\ + \sum_{k=0}^{\infty} A^{t-1-k}\Sigma_\nu(A')^{t-1-k}. &\end{aligned}$$

Since the matrix A is stable all the series converge and this proves that

$$0 < \lim_{t\to\infty}\Sigma_x(t) < \infty$$

□

The stationarity conditions of DNS-YGARCH model are the same of those of DNS-XGARCH. In addiction, if the process $\{x(t)\}$ is gaussian, the DNS-YGARCH model is also ergodic.

2.3.3 *SSGARCH*

Here we will introduce two GARCH models in state space framework.

The first model is given by

$$\begin{cases} x(t+1) = ax(t) + \nu(t) & \{\nu(t)\} \quad \text{i.i.d.} \quad Var(\nu) = \sigma_\nu^2 \\ y(t) = \sigma(t)x(t) \\ \sigma(t)^2 = \omega + \sum_{i=1}^{q} \alpha_i y(t-1)^2 + \sum_{j=1}^{p} \beta_j \sigma(t-1)^2 \end{cases} \tag{34}$$

First we are interested to see if this model is effectively a GARCH model, that is, consistent with Definition 1. The following theorem shows that the model Equation 34 is partially a GARCH model.

Theorem 10. Suppose the process $\{\nu(t)\}$ to be i.i.d. with zero mean and variance σ_ν^2. Then, the model defined by Equation 34 is a GARCH process according to the Definition 1 if we substitute the condition $E(y_t \mid \mathcal{F}_{t-1}) = 0$ with

$$E(y_t \mid \mathcal{F}_{t-1}) = k(\mathcal{F}_{t-1}) \qquad \exists k$$

where k depends on past observations. By this substitution it follows that the process $\{y_t\}$ it is not a *d-martingale* but zero mean white noise process.

Proof. Let define $\mathcal{F}_{t-1}$ as the space of past observations of process $\{y_t\}$. The following property are satisfied:

1. $\mathcal{F}_t \subset \mathcal{F}_{t+1}$;
2. $\{y_t,\ t \in \mathbb{Z}\}$ is $\mathcal{F}_t$-measurable;
3. $E|y(t)| < \infty$.

The third condition follows by

$$E|y(t)| = E|\sigma(t)| \cdot |x(t)| < KE|x(t)| < \infty \qquad \exists\, K.$$

provided that $\{x(t)\}$ is Gaussian.
Finally, we have

$$E\big[y(t) \mid \mathcal{F}_{t-1}\big] = \sigma(t)E[x(t) \mid \mathcal{F}_{t-1}]$$

where $E[x(t) \mid \mathcal{F}_{t-1}]$ is provided by the Kalman filter and it is different from zero, in general.
The conditional variance is

$$E[y_t^2 \mid |\mathcal{F}_{t-1}] = E[\sigma(t)^2 x(t)^2 \mid \mathcal{F}_{t-1}] = \sigma(t)^2 E[x(t)^2 \mid \mathcal{F}_{t-1}]$$

where $E[x(t)^2 \mid \mathcal{F}_{t-1}]$ is provided by the Kalman filter. This result is compatible with time-varying conditional variance hypothesis. The difference with the variance defined in Definition 1 is the term $E[x(t)^2 \big| \mathcal{F}_{t-1}]$ which can be interpreted as a time-varying coefficient before the heteroscedastic variance.
Moreover in Theorem 11 we proof the unconditional variance of process $\{y(t)\}$ is constant under the stationarity assumption. □

The following theorem gives a necessary and sufficient condition for the second-order stationarity for the GARCH(1,1) model. It can merely generalize with some more calculus to GARCH(p,q) process.

Theorem 11. Suppose $\{v(t)\}$ to be i.i.d.. If are satisfied the conditions

$$\alpha\sigma_x^2 + \beta < 1, \qquad |a| < 1$$

where σ_x is the stationary state variance of process $\{x(t)\}$, then then process $y(t)$ is second-order stationary.

Proof. For the stationarity of $\{y(t)\}$ the process $\{x(t)\}$ must be stationary, that is, the required condition is

$$|a| < 1.$$

Under this condition we have

$$\sigma_x^2 = \frac{\sigma_v^2}{1 - a^2} > \sigma_v^2 \qquad \text{as } t \to \infty$$

Then, asymptotically, the unconditional variance of process $y(t)$ is

$$Ey(t)^2 = \sigma_x^2 E\sigma(t)^2$$

Thus we need to evaluate $E\sigma(t)^2$. It follows that

$$
\begin{aligned}
E\sigma(t+1)^2 &= \omega + E\sigma(t)^2 E\left(\alpha x(t)^2 + \beta\right) \\
E\sigma(t+1)^2 &= \omega + E\sigma(t)^2\left(\alpha\sigma_x^2 + \beta\right) \\
E\sigma(t)^2 &= \omega \sum_{k=0}^{\infty}\left(\alpha\sigma_x^2 + \beta\right)^k \qquad \text{as} \quad t \to \infty
\end{aligned}
$$

and if and only if $\left(\alpha\sigma_x^2 + \beta\right) < 1$ the stationary variance is

$$E\sigma(t)^2 = \frac{\omega}{1 - (\alpha\sigma_x^2 + \beta)}$$

Finally we have

$$Ey(t)^2 = \frac{\sigma_x^2\omega}{1 - (\alpha\sigma_x^2 + \beta)}.$$

□

Let suppose $y(0) = 0$. The process $y(t)$ can be rewritten avoiding the state space form as

$$y(t) = \sigma(t)\sum_{k=0}^{t-1} a^k \nu(t-1-k) \qquad \text{as } t \to \infty$$

In particular, we have

$$
\begin{aligned}
y(t) &= \sigma(t)\sum_{k=0}^{t-1} a^k \nu(t-1-k) \\
y(t+1) &= \sigma(t+1)\sum_{k=0}^{t} a^k \nu(t-k) \\
&= \sigma(t+1)\{\sum_{k=0}^{t-1} a^k \nu(t-k) + \nu(t)\}
\end{aligned}
$$

and thus it can be easily found the auto-regressive form

$$y(t+1) = a\frac{\sigma(t+1)}{\sigma(t)}y(t) + \sigma(t+1)\nu(t)$$

Defining $\gamma(t+1) = a\frac{\sigma(t+1)}{\sigma(t)}$ we have

$$y(t+1) = \gamma(t+1)y(t) + \sigma(t+1)\nu(t) \tag{35}$$

where $\gamma(t+1)$, $\sigma(t+1)$ are known at instant t and $y(t)$ and $\nu(t)$ are uncorrelated. Thus, the one-step prediction is given by

$$y(t+1|t) = \gamma(t+1)y(t)$$

The conditional one-step variance prediction is

$$\begin{aligned} E[y(t+1)^2|t] &= \gamma(t+1)^2 E[y(t)^2|t] + \sigma(t+1)^2 E[\nu(t)^2|t] \\ &= \gamma(t+1)^2 E[y(t)^2|t] + \sigma_\nu^2 \sigma(t+1)^2 \\ &= \gamma(t+1)^2 y(t)^2 + \sigma_\nu^2 \sigma(t+1)^2. \end{aligned} \tag{36}$$

The second model we propose is given by

$$\begin{cases} x(t+1) = ax(t) + \nu(t) & \{\nu(t)\} \quad \text{i.i.d.}(0, \sigma_\nu^2) \\ y(t) = cx(t) + \sigma(t)\varepsilon(t) & \{\varepsilon(t)\} \quad \text{i.i.d.} \quad (0, \sigma_\varepsilon^2) \\ \sigma(t)^2 = \omega + \alpha y(t-1)^2 + \beta\sigma(t-1)^2 \end{cases} \tag{37}$$

The following theorem gives a necessary and sufficient condition for the stationarity.

Theorem 12. The model Equation 37 is second-order stationary if and only if

$$\alpha\sigma_\varepsilon^2 + \beta < 1$$

and the stationary variance is

$$Ey(t)^2 = c^2\sigma_x^2 + \sigma_\varepsilon^2 \frac{\omega + \alpha c^2 \sigma_x^2}{1 - (\alpha\sigma_\varepsilon^2 + \beta)}$$

Proof. If $|a| < 1$ the process $\{x(t)\}$ is asymptotically stationary and then the stationary variance is

$$\begin{aligned} Ey(t)^2 &= c^2\sigma_x^2 + E\sigma(t)^2\sigma_\varepsilon^2 \\ E\sigma(t)^2 &= \omega + \alpha Ey(t-1)^2 + \beta E\sigma(t-1)^2 \end{aligned}$$

which yields

$$\begin{aligned} E\sigma(t)^2 &= \omega + \alpha\left\{c^2\sigma_x^2 + E\sigma(t-1)^2\sigma_\varepsilon^2\right\} + \beta E\sigma(t-1)^2 \\ &= \omega + \alpha c^2\sigma_x^2 + (\alpha\sigma_\varepsilon^2 + \beta)E\sigma(t-1)^2 \end{aligned}$$

Letting $t \to \infty$ we obtain

$$\begin{aligned} E\sigma(t)^2 &= \left(\omega + \alpha c^2\sigma_x^2\right)\sum_{i=0}^{\infty}(\alpha\sigma_\varepsilon^2 + \beta)^i \\ &= \frac{\omega + \alpha c^2\sigma_x^2}{1 - (\alpha\sigma_\varepsilon^2 + \beta)} \end{aligned}$$

if and only if $\alpha\sigma_\varepsilon^2 + \beta < 1$. It follows that

$$Ey(t)^2 = c^2\sigma_x^2 + \sigma_\varepsilon^2\frac{\omega + \alpha c^2\sigma_x^2}{1 - (\alpha\sigma_\varepsilon^2 + \beta)}$$

□

Concerning all the state space models presented in this section a remark have to be done: all these models are conditionally Gaussian on the past and obviously the term the term that the Kalman filter minimizes is

$$E\left[(x - \hat{x})(x - \hat{x})'\middle|\mathcal{F}_{t-1}\right]$$

with the same formulas for prediction and updating. See Chen and Kumar [5] for details.

3

STOCHASTIC STABILITY

In this section we will explore the stochastic stability of GARCH models. In particular we will study the geometrical ergodicity and the β-mixing properties. This last properties require the an heavy use of Markov Chian theory. We will refer to Meyn and Tweedie [19].

3.1 GEOMETRIC ERGODICITY AND MIXING

In this section we shall study the Geometric Ergodicity and Mixing properties of GARCH models. We will need elements of Markov chain theory, well explained in [19].

Let be X_t a Markovian process evolving on state space E with a σ-field $\mathcal{F}$. Our objective is study the converge of the probability $P_\mu(X)(X_t \in \cdot)$ to a probability $\pi(\cdot)$ independent on the initial probability μ, as $t \to \infty$. We need μ to be invariant, i.e stationary, and this is satisfied if there exists a probability measure π such that, $\forall \pi$,

$$\forall B \in \mathcal{F}, \quad P_\mu(X_t \to B) \to \pi(B), \quad \text{ast} \to +\infty.$$

A chain $\{X_t\}$ is said to be *ergodic* if

$$\|P^t(x, \cdot) - \pi\| \to 0, \qquad \text{ast} \to +\infty \tag{38}$$

, where $\|\cdot\|$ is the total variation norm. This condition is satisfied if the chain is irreducible, aperiodic and admits a invariant probability π, for almost all $x \in E$. The mixing property is stronger than ergodicity. In addiction, condition (38) does not give any information on convergence rate. We will define a stronger notion of ergodicity that is *geometric ergodicity*, that gives a geometrical

rate convergence. The chain $\{X_t\}$ is called *geometrically ergodic* if there exists $\rho \in (0,1)$ such that

$$\forall x \in E, \quad \frac{1}{\rho^t}\|P^t(x,\cdot) - \pi\| \to 0, \qquad \text{ast} \to +\infty. \tag{39}$$

Geometric ergodicity imply α and β-*mixing* (see Section A.3). Davydov in [7] showed that for an ergodic Markov chain $\{X_t\}$, of invariant probability measure π,

$$\beta_x(k) = \int \|P^k(x,\cdot) - \pi\|\pi(dx).$$

Then, if Equation 39 holds, that is $\{X_t\}$ is geoemtrically stationary,

$$\beta_x(k) \to 0 \qquad \text{as } t \to \infty$$

which is the condition for β-mixing. In general, it is not easy to verify the properties of recurrence, existence of an invariant probability measure and geometric ergodicity. The follow theorem provides a way to prove geometric ergodicity for several Markov processes.

Theorem 13 (Feign and Tweedie)**.** Assuming that:

1. $\{X_t\}$ is a Feller chain;
2. $\{X_t\}$ is ϕ-irreducible;
3. there exists a compact set $A \subset E$ such that $\phi(A) > 0$ and a continuous function $V : E \to \mathbb{R}^+$ satisfying

$$V(x) \geqslant 1, \forall x \in A, \tag{40}$$

and for $\delta > 0$,

$$E\{V(X_t)|X_{t-1} = x\} \leqslant (1-\delta)V(x), \quad \forall x \notin A. \tag{41}$$

Then $\{X_t\}$ is geometrically ergodic.

The condition (3.) is named *drift criterion*. Indeed, V can be seen as a energy function and when the chain is outside the center A of the state space, the energy is dissipated on average. When the chain lies inside A, the energy function is bounded

since A is compact and V is continue. For some models, the drift criterion is too restrictive because is based on in only one step transitions. Meyn and Tweedie [19] adapt the criteria to transition in n steps.

Theorem 14 (Geometric ergodicity criterion). Assuming that:

1. $\{X_t\}$ is a Feller chain;
2. $\{X_t\}$ is ϕ-irreducible;
3. there exists a compact set $A \subset E$ such that $\phi(A) > 0$, an integer $n \geqslant 1$ and a continuous function $V : E \to \mathbb{R}^+$ satisfying

$$V(x) \geqslant 1, \forall x \in A, \tag{42}$$

and for $\delta > 0$ and $b > 0$,

$$E\{V(X_{t+n})|X_{t-1} = x\} \leqslant (1-\delta)V(x), \quad \forall x \notin A \tag{43}$$
$$E\{V(X_{t+n})|X_{t-1} = x\} \leqslant b \quad \forall x \in A \tag{44}$$

Then $\{X_t\}$ is geometrically ergodic.

The compactness condition of A can be replaced by a small set if the function V is bounded on A.

3.1.1 *Geometric ergodicity of GARCH*

We will focus on GARCH(1,1) since only ths is used on state space framework. Consider the GARCH(1,1) model

$$\begin{cases} y_t = \sigma_t^{\frac{1}{2}} \varepsilon_t \\ h_t = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 \end{cases} \tag{45}$$

where $\omega > 0$, $\alpha \geqslant 0$, $\beta \geqslant 0$ and $\{\varepsilon_t\}$ as usual. Let us make an assumption

Assumption 1. The law P_ε of process $\{\varepsilon_t\}$ is absolutely continuous, of density f with respect to the Lebesgue measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We assume that

$$\varepsilon^0 := \inf\{\varepsilon | \varepsilon > 0, f(\varepsilon) > 0\} = \inf\{-\varepsilon | \varepsilon < 0, f(\varepsilon) > 0\},$$

and that there exists $\tau > 0$ such that

$$\left(-\varepsilon^0 - \tau, -\varepsilon^0\right) \cup \left(\varepsilon^0, \varepsilon^0 + \tau\right) \subset \{f > 0\}.$$

This imply $\varepsilon^0 = 0$.

Now we are ready to present the following theorem.

Theorem 15. Under Assumption 1 and if

$$E \log(\alpha \varepsilon_t^2 + \beta) < 0,$$

the strictly stationary solution of the GARCH(1,1) model (45) is such that the Markov chain $\{\sigma_t\}$ is geometrically ergodic and the process $\{y_t\}$ is geometrically β-mixing.

4

IDENTIFICATION

A relevant part of finance theory still assumes that the prices follow a random walk process. If significant sample autocorrelations are detected in the price variations it means the random walk hypothesis drops. As we saw in previous chapter the second-order stationary solution of a GARCH process is a white noise, that is, with theoretical autocorrelation such that $\rho(h) = Ey_t y_{t+h}/Wy_t^2 = 0 \quad \forall h \neq 0$. Thus, the first step is to verify the absence of correlation. The second step consists on checking for correlation in the squared returns. Given observations $y_1, ..., y_N$ the theoretical autocorrelations of a centered process $\{y_t\}$ are generally estimated by the sample autocorrelations (SACRs)

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

with

$$\hat{\gamma}(h) = \hat{\gamma}(-h) = \frac{1}{N}\sum_{i=1}^{N-h} y_t y_{t+h} \qquad h = 0, 1, \ldots, N-1.$$

while for sample autocorrelations of squared return we have

$$\hat{\gamma}(h) = \hat{\gamma}(-h) = \frac{1}{N}\sum_{i=1}^{N-h} y_t^2 y_{t+h}^2 \qquad h = 0, 1, \ldots, N-1.$$

If we find that the variance process exhibits some correlation, a GARCH framework should be considered. In order to quantify the correlation we use formal hypothesis tests, such as the Ljung-Box-Pierce Q-test and Engle's ARCH test.
Under the null hypothesis that a time series is a random sequence of Gaussian disturbances (i.e., no ARCH effects exist), this test statistic is also asymptotically Chi-Square distributed. Third step consists on estimating the model parameters.

In order to provide estimates of the parameters we use the concept of Maximum Likelihood.
For pure GARCH models we will use the Quasi Maximum Likelihood Estimator (QMLE): term "Quasi" refers to the non-Gaussianity of noise $\varepsilon(t)$. Indeed, in general is a i.i.d process.
The estimation of GARCH models will be discussed in Section 4.2. In Section 4.3 we will discuss the estimation problem of DNS-GARCH models.

For state space models with GARCH noise and SSGARCH discussed in Section 2.2 we must use a slightly different framework. As it is, we cope with the state equation representing the unobserved latent process. Being the state value unknown we can not directly employ the Maximum Likelihood Maximization. A well known technique, dealing with this problem, is Expectation-Maximization (EM) algorithm. Basically, given an initial guess of vector parameters θ, it is an iterative procedure that follows the steps:

- perform the E-step, that is, apply the optimal linear prediction by the Kalman filter;
- perform the M-step, that is, find θ such that it maximizes the log-likelihood function $L(\cdot, \theta)$
- repeat steps 1 and 2 with the new current estimation of θ until the algorithm converges.

Finally, the last step (see Section 4.5) concerns the validation of the fitted model and in more detail we will compare the residuals, conditional standard deviations, and returns.

First of all, we consider the property of white noise of the general stationary GARCH(p,q) process.

4.1 VERIFICATION OF WHITE NOISE PROPERTY

Consider the usual GARCH(p,q) model

$$\begin{cases} y_t = \sqrt{h_t}\varepsilon_t \\ h_t = \omega_0 + \sum_{i=1}^{q} \alpha_{0i} y_{t-i}^2 + \sum_{j=1}^{p} \beta_{0j} h_{t-j}, \qquad \forall t \in \mathbb{Z}, \end{cases} \tag{46}$$

with $\{\varepsilon_t\}$ a i.i.d process and satisfying the second-order assumption.

Consider $\hat{\rho}_m = (\hat{\rho}(1), \ldots, \hat{\rho}(m))'$ and $\hat{\gamma}_m = (\hat{\gamma}(1), \ldots, \hat{\gamma}(m))'$ with obvious meaning of notation given the N observations of process y_t. Then, it follows

Theorem 16. If $\{y_t\}$ is the stationary sulution of the GARCH(p,q) model and if Ey_t^2 is finite, then, when $N \to \infty$,

$$\sqrt{N}\hat{\gamma}_m \to N\left(0, \Sigma_{\hat{\gamma}_m}\right) \quad \text{and} \quad \sqrt{N}\hat{\rho}_m \to N\left(0, \Sigma_{\hat{\rho}_m}\right)$$

where

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

is nonsingular and

$$\Sigma_{\hat{\rho}_m} = \frac{1}{\left(Ey_t^2\right)^2}\Sigma_{\hat{\gamma}_m}.$$

If the law of $\{\varepsilon_t\}$ is symmetric then $\Sigma_{\hat{\gamma}_m}$ is diagonal.

The sample autocorrelation function (ACF) and partial autocorrelation function (PACF) are useful qualitative tools to assess the presence of autocorrelation at individual lags. The Ljung-Box Q-test is a more quantitative way to test for autocorrelation at multiple lags jointly. The choice of m affects test performance. If N is the length of the observed time series, choosing $m = \ln(N)$ is recommended. The Ljung-Box test statistic is given by

$$Q(m) := N(N+2)\sum_{i=h}^{m} \frac{\hat{\rho}^2(h)}{N-h}$$

and it rejects the strong white noise hypothesis if Q is greater than the $(1-\alpha)$-quantile of $\chi(m)^2$ since, under the null hypothesis, $Q(m)$ follows a $\chi(m)^2$ distribution. Unfortunately this test is not robust to condition heteroscedasticity since it assume the

process is strong white noise. It can be shown that under assumption of Theorem 16 the statistic

$$Q(m) = N\hat{\rho}'_m \Sigma^{-1}_{\hat{\rho}_m} \hat{\rho}_m$$

asymptotically follows a $\chi(m)^2$ distribution.

4.2 QMLE FOR PURE GARCH AND ARMA-GARCH

QMLE is the most used method to estimate GARCH because it provides good asymptotic behaviors as consistency and normality under the assumption of stationarity and some other mild assumptions.

In Section 4.2.1 will be presented this method for pure GARCH Although the variable ϵ_t is i.i.d. and not in general gaussian, it will be demonstrated that the gaussian log-likelihood provides consist estimates and asymptotic normality. The terms "Quasi" refer to the assumption that ϵ_t is gaussian.

4.2.1 *QLME for GARCH models*

Suppose the observations $y_1, .., y_N$ follow a stationary GARCH(p,q) process defined as

$$\begin{cases} y(t) = \sqrt{h(t)}\varepsilon(t) \\ h(t) = \omega_0 + \sum_{i=1}^{q} \alpha_{0i} y(t-i)^2 + \sum_{j=1}^{p} \beta_{0j} h(t-j), \qquad \forall t \in \mathbb{Z}, \end{cases} \tag{47}$$

where $\epsilon(t)$ is i.i.d with variance 1, $\omega_0 > 0$, $\alpha_{0i} \geqslant 0$ and $\beta_{0j} \geqslant 0$. Assume the orders p and q are known. The vector of parameters to be estimated is

$$\theta = (\omega, \alpha_1, ..., \alpha_q, \beta_1, .., \beta_p)'$$

and it belongs to the parameter space

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

The notation θ_0 refers to the true parameters, that is

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

For the sake of clarity, given a time-varying variable $f(t)$ let use the notation $f(t) = f_t$. Given the initial conditions $y_0, \dots, y_{1-q}$ and $\tilde{h}_0, \dots, \tilde{h}_{1-p}$, the conditional Gaussian likelihood is given by

$$L_N(\theta) = \prod_{t=1}^{N} \frac{1}{\sqrt{2\pi\tilde{h}_t}} \exp\left(-\frac{y_t^2}{2\tilde{h}_t}\right)$$

where, for $t \geqslant 1$, $\tilde{h}_t$ are defined as in Equation 47. If process is wide stationary, the unknown initial values can be chosen as the unconditional variance and so

$$y_0^2 = \cdots = y_{1-q}^2 = \tilde{h}_0 = \cdots = \tilde{h}_{1-p} = \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j} \tag{48}$$

Defining $\mathbf{x} = (y_1, \dots, y_N)$, the estimated parameter $\hat{\theta}(\mathbf{x}) \in \Theta$ maximizes $L(\theta)$, that is

$$\hat{\theta}_N = \arg\max_{\theta \in \Theta} L_N(\theta). \tag{49}$$

Under gaussian assumption and taking the log-likelihood, it results that the function to be maximized is given by

$$-\frac{1}{2} \sum_{i=1}^{N} \left(\log \tilde{h}_t + \frac{y_t^2}{\tilde{h}_t}\right)$$

and it is equivalent to minimizing, with respect to θ,

$$\tilde{l}(\mathbf{x}, \theta) = \frac{1}{N} \sum_{i=1}^{N} \left(\log \tilde{h}_t + \frac{y_t^2}{\tilde{h}_t}\right) \tag{50}$$

In order to find $\hat{\theta}$, it is necessary to cancel the derivative of $\tilde{l}(\mathbf{x}, \theta)$, that is,

$$\frac{1}{N}\sum_{t=1}^{N}\left(y_t^2 - \tilde{h}_t\right)\frac{1}{\tilde{h}_t^2}\frac{\partial \tilde{h}_t}{\partial \theta} = 0. \tag{51}$$

Similarly, define

$$l(\mathbf{x}, \theta) = \frac{1}{N}\sum_{i=1}^{N}\left(\log h_t + \frac{y_t^2}{h_t}\right)$$

.

To show the strong consistency, we make the following assumption.

A1: $\theta_0 \in \Theta$ and Θ is compact.

A2: $\sum_{j=1}^{p}\beta_j < 1 \quad \forall\theta \in \Theta$.

A3: ε_t^2 has nondegenerate distribution and $E\varepsilon_t^2 = 1$.

A4: $\sum_{i=1}^{q}\alpha_i z^i$ and $\sum_{i=1}^{q}\beta_i z^i$ are coprime, $\sum_{i=1}^{q}\alpha_i \neq 0$, and $\alpha_{0q} + \beta_{0p} \neq 0$.

Under these assumptions, it follows

Theorem 17 (Consistency of QMLE)**.**] Let $\hat{\theta}_N$ be the estimator satisfying Equation 49 with the initial conditions Equation 48, then

$$\hat{\theta}_N \to \theta_0, \quad \text{as } n \to \infty \quad \text{a.s.}$$

Proof. Let rewrite the GARCH component of model Equation 47 in the auto regressive form

$$\boldsymbol{h_t} = \boldsymbol{c_t} + B\boldsymbol{h_{t-1}} \tag{52}$$

where

$$\boldsymbol{h_t} = \begin{bmatrix} h_t \\ h_{t-1} \\ \vdots \\ h_{t-p+1} \end{bmatrix}, \qquad \boldsymbol{c_t} = \begin{bmatrix} \omega + \sum_{i=1}^{q}\alpha_i y_{t-i}^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$B = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_p \\ 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 0 & \cdots & 1 & 0 \end{bmatrix}$$

We will establish the following intermediate results.

1. $\lim_{N\to\infty} \sup_{\theta\in\Theta} |l_N(\theta) - \tilde{l}_N(\theta)| = 0, \quad$ a.s.
2. $\exists t \in \mathbb{Z}$ such that $h_t(\theta) = h_t(\theta_0)$ a.s. $\Rightarrow \theta = \theta_0$
3. $E_{\theta_0}|l_t(\theta_0)| < \infty$, and if $\theta \neq \theta_0$, $E_{\theta_0} l_t(\theta) > E_{\theta_0} l_t(\theta_0)$.
4. For any $\theta \neq \theta_0$, there exists a neighborhood $V(\theta)$ such that

$$\lim_{N\to\infty} \inf_{\theta^*\in V(\theta)} \tilde{l}_N(\theta^*) > E_{\theta_0} l_1(\theta_0) \quad \text{a.s.}$$

1 .**Asymptotic irrelevance of the initial values.** Iterating (52) we obtain

$$\boldsymbol{h_t} = \sum_{k=0}^{\infty} B^k \boldsymbol{c}_{t-k}.$$

Let $\boldsymbol{h_t}$ be the vector obtained by replacing $\tilde{h_{t-i}}$ by h_{t-i}, and let $\boldsymbol{c_t}$ be the vector obtained by replacing $y_0^2, \ldots, y_{t-q}^2$ by by the initial values (48).
Then, it is easy to see that almost surely

$$\sup_{\theta\in\Theta} \|\boldsymbol{h}_t - \tilde{\boldsymbol{h}}_t\| = \sup_{\theta\in\Theta} \left\| \sum_{i=1}^{q} B^{t-k}(\boldsymbol{c}_t - \tilde{\boldsymbol{c}}_t) + B^t(\boldsymbol{h}_0 - \tilde{\boldsymbol{h}}_0) \right\| \leqslant K\rho^t \quad \forall t. \tag{53}$$

where K is a opportune constant and $\rho < 1$. The last inequality follows by the structure of B and assumption A2. For $x > 0$ we

have $\log x \leqslant x - 1$. It follows that, for $x, y > 0$, $\left|\log \frac{x}{y}\right| \leqslant \frac{|x-y|}{\min(x,y)}$. We thus have almost surely, using (53),

$$\begin{aligned}\sup_{\theta\in\Theta} \left|l_N(\theta) - \tilde{l}_N(\theta)\right| &\leqslant \frac{1}{N}\sum_{i=1}^{N} \sup_{\theta_0\in\Theta} \left\{ \left|\frac{\tilde{h}_t - h_t}{\tilde{h}_t h_t}\right| y_t^2 + \left|\log\left(\frac{h_t^2}{\tilde{h}_t^2}\right)\right| \right\} \\ &\leqslant \frac{K}{N}\left\{\sup_{\theta\in\Theta}\frac{1}{\omega^2}\right\}\sum_{t=1}^{N}\rho^t y_t^2 + \frac{K}{N}\left\{\sup_{\theta\in\Theta}\frac{1}{\omega}\right\}\sum_{t=1}^{N}\rho^t \end{aligned} \tag{54}$$

Given that $\sum_{t=1}^{\infty}\rho^t = \frac{1}{1-\rho}$ we have $\frac{1}{N}\sum_{t=1}^{N}\rho^t \to 0$ as $N \to \infty$. Then, from assumption A1, using the Borel-Cantelli's lemma and the Markov's Inequality we are allowed to show $\rho^t y_t^2 \to 0$ a.s. Finally, applying the first Cesaro's theorem we have

$$\frac{1}{N}\sum_{t=1}^{N}\rho^t y_t^2 \to 0$$

and point 1 follows.

2. **Identifiability of the parameter.** Assume that $h_t(\theta) = h_t(\theta_0)$ a.s.. Let define

$$\mathcal{A}_\theta(z) = \sum_{i=1}^{q}\alpha_i z^i, \qquad \mathcal{B}_\theta(z) = \sum_{i=1}^{q}\beta_i z^i. \tag{55}$$

Then we obtain

$$\left\{\frac{\mathcal{A}_\theta(z)}{\mathcal{B}_{\theta_0}(z)} - \frac{\mathcal{A}_{\theta_0}(z)}{\mathcal{B}_\theta(z)}\right\} y_t^2 = \frac{\omega_0}{\mathcal{B}_{\theta_0}(1)} - \frac{\omega_0}{\mathcal{B}_\theta(1)} \quad \text{a.s.} \quad \forall t.$$

If the operator between braces were not null, then there would exist a constant linear combination of the y_{t-i}^2, $i \geqslant 0$. Thus the linear innovation of the process (y_t^2) would be equal to zero. Then, under assumption A3,

$$y_t^2 - E_{\theta_0}(y_t^2 | y_{t-1}^2, \dots) = h_t(\theta_0)(\varepsilon_t^2 - 1) \neq 0, \qquad \text{with not null probability.}$$

We thus have

$$\frac{\mathcal{A}_\theta(z)}{\mathcal{B}_{\theta_0}(z)} = \frac{\mathcal{A}_{\theta_0}(z)}{\mathcal{B}_{\theta_0}(z)}, \qquad \forall |z| \leqslant 1 \quad \text{and} \quad \frac{\omega}{\mathcal{B}_\theta(1)} = \frac{\omega}{\mathcal{B}_{\theta_0}(1)}. \tag{56}$$

Under assumption A4 it follows that $\mathcal{A}_\theta(z) = \mathcal{A}_{\theta_0}(z)$, $\mathcal{B}_\theta(z) = \mathcal{B}_{\theta_0}(z)$ and $\omega = \omega_0$.

3. It can be shown that $\mathrm{E}_{\theta_0} l_t(\theta_0)$ is well defined in $\mathbb{R}$. Since for all $x > 0$, $\log x \leqslant x - 1$ with equality iff $x = 1$, we obtain

$$\mathrm{E}_{\theta_0} l_t(\theta) - \mathrm{E}_{\theta_0} l_t(\theta_0) = \mathrm{E}_{\theta_0} \log \frac{h_t(\theta)}{h_t(\theta_0)} + \mathrm{E}_{\theta_0} \frac{h_t(\theta_0)}{h_t(\theta)} \varepsilon_t^2 - \mathrm{E}_{\theta_0} \varepsilon_t^2 \tag{57}$$

$$= \mathrm{E}_{\theta_0} \log \frac{h_t(\theta)}{h_t(\theta_0)} + \mathrm{E}_{\theta_0} \frac{h_t(\theta_0)}{h_t(\theta)} - 1 \tag{58}$$

$$\geqslant \mathrm{E}_{\theta_0} \left\{ \log \frac{h_t(\theta)}{h_t(\theta_0)} + \log \frac{h_t(\theta_0)}{h_t(\theta)} \right\} = 0 \tag{59}$$

with equality if and only if $\frac{h_t(\theta_0)}{h_t(\theta)} = 1$ a.s., that is, for point 3 if and only if $\theta = \theta_0$.

4. **Compactness of Θ and ergodicity of** $l_t(\theta)$. See Francq and Zakoïan [13]. □

It can be shown the Asymptotic Normality of the estimator. Let consider the following assumptions

A5: θ belongs to the interior of Θ.

A6: $\nu = \mathrm{E}\varepsilon_t^4 < \infty$.

Then, it follows the theorem

Theorem 18 (Asymptotic Normality of QMLE)**.** Under assumption A1-A6,

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, (\nu - 1) J^{-1}\right)$$

where

$$J := \mathrm{E}_{\theta_0} \left(\frac{\partial l_t(\theta_0)}{\partial \theta \partial \theta'} \right) = \mathrm{E}_{\theta_0} \left(\frac{1}{h_t^2(\theta_0)} \frac{\partial h_t(\theta_0)}{\partial \theta} \frac{\partial h_t(\theta_0)}{\partial \theta'} \right)$$

is a positive definite matrix.

Proof. The proof is split in several parts. It shows the integrability of the derivatives of the criterion at θ_0, the invertibility of J, uniform integrability of the third-order derivatives of the criterion, asymptotic decrease of the effect of the initial values , the CLT for martingale increments and finally the ergodic theorem. □

4.3 PARAMETER ESTIMATION: A KALMAN FILTER APPROACH

So far, we have presented the QMLE estimation for time series. Now we will extend this results to state space models introducing the EM algorithm. Here we will not specify the distribution of noises, but we consider it as i.i.d. white noise process. In can be shown (for example see Wu [24]) that under general condition the EM algorithm converges providing consistent estimates of parameters. In our case, we note that the state-space models we will employ are in general non-Gaussian but they are conditionally Gaussian to the strict past.

4.3.1 *EM algorithm with Kalman filter*

Let us now, a brief explanation of EM algorithms, in the general linear state space model. Suppose to consider the state space model

$$\begin{cases} \mathbf{x}(t+1) = A\mathbf{x}(t) + \mathbf{v}(t) \\ \mathbf{y}(t) = C\mathbf{x}(t) + \mathbf{w}(t) \end{cases}$$

$$E(\mathbf{v}(t)\mathbf{v}(\tau)') = \begin{cases} Q & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases}$$

$$E(\mathbf{w}(t)\mathbf{w}(\tau)') = \begin{cases} R & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases}$$

with the notation:

- $\mathbf{y}(t)$ denotes an $p \times 1$ vector of observable variables.

- $\mathbf{x}(t)$ denotes a possibly unobserved $n \times 1$ state vector.
- $\mathbf{v}(t)$ and $\mathbf{w}(t)$ are white noise i.i.d with zero mean.

We will make the assumptions:

- The disturbances $\mathbf{v}(t)$ and $\mathbf{w}(t)$ are assumed to be uncorrelated at all leads and lags

$$E(\mathbf{v}(t)\mathbf{w}(\tau)') = 0 \qquad \text{for all } t \text{ and } \tau$$

- The initial state and the disturbances are uncorrelated, i.e.

$$E\{\mathbf{x}_0 \left[\mathbf{v}(t)', \mathbf{w}(t)'\right]\} = 0, \qquad \forall t \geqslant t_0$$

- A has all stable eigenvalues, i.e., the process is stationary.

Let denote

$$\mathcal{F}_t \equiv (\mathbf{y}_t, \mathbf{y}_{t-1}, ..., \mathbf{y}_1, \mathbf{x}_t, \mathbf{x}_{t-1}, ..., \mathbf{x}_1)$$

and denote

$$\hat{\mathbf{x}}_{t+1|t} \equiv \widehat{E}(\mathbf{x}_{t+1}|\mathcal{F}_t)$$

$$\hat{\mathbf{x}}_{t|t} \equiv \widehat{E}(\mathbf{x}_t|\mathcal{F}_t)$$

$$\mathbf{P}_{t+1|t} \equiv E\left(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t}\right)\left(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t}\right)'$$

It can be demonstrated that the equation describing the Kalman filter are

$$\hat{\mathbf{x}}_{t+1|t} = A\hat{\mathbf{x}}_{t|t-1} + K(t)\left(\mathbf{y}(t) - C\hat{\mathbf{x}}_{t|t-1}\right)$$

$$K(t) = AP_{t|t-1}C'(CP_{t|t-1}C' + R)^{-1}$$

$$P_{t+1|t} = A\left(P_{t|t-1} - P_{t|t-1}C'(CP_{t|t-1}C' + R)^{-1}CP_{t|t-1}\right) + Q$$

Given the stationarity the initial conditions is set as

$$E\mathbf{x}(0|-1) = E(\mathbf{x}) = 0$$

$$P(0|-1) = \Sigma$$

where Σ is the solution of

$$\Sigma = A\Sigma A' + Q$$

Besides the predictions, the Kalman filter can be used to identify the unknown parameters A, C, D, Q, R. This approach is based on the iterative maximization of an opportune likelihood function. In order to be clear, let suppose T observations $(y_1, \dots, y_T)$ are available. Once obtained the Kalman predictions we have to find the parameters maximizing the log-likelihood given all the information, that is, the likelihood function

$$L(\mathbf{y}_1, \dots, \mathbf{y}_T; \theta) = \prod_{t=1}^{T} f(\mathbf{y}_t | \mathcal{F}_{t-1}; \theta)$$

If the disturbances $\{\mathbf{v}, \mathbf{w}\}_{t=1}^{T}$ are Gaussian then the conditional distribution of $\mathbf{y}(t)$ is Gaussian:

$$\mathbf{y}(t)|\mathcal{F}_{t-1} \sim N\Big((C\hat{\mathbf{x}}_{t|t-1}), CP_{t|t-1}C' + R\Big)$$

and the likelihood function is

$$f_{\mathbf{y}_t|\mathcal{F}_{t-1}}(\mathbf{y}(t)|\mathbf{x}(t), \mathcal{F}) = (2\pi)^{-n/2}|CP_{t|t-1}C' + R|^{1/2}$$
$$\times \exp\left\{ -\frac{1}{2} \underbrace{\big(\mathbf{y}(t) - C\hat{x}_{t|t-1}\big)'}_{\mathbf{z}(t)'} \underbrace{(CP_{t|t-1}C' + R)^{-1}}_{F(t)^{-1}} \underbrace{\big(\mathbf{y}(t) - C\hat{x}_{t|t-1}\big)}_{\mathbf{z}(t)} \right\}$$

And the sample log-likelihood function is

$$\begin{aligned} \log\big(f_{\mathbf{Y}_t|\mathbf{X}_t,\mathcal{F}}(\mathbf{y}_t|\mathcal{F}_{t-1})\big) &= \sum_{t=1}^{T} \log f_{\mathbf{y}_t|\mathcal{F}_{t-1}}\big(\mathbf{y}_t|\mathcal{F}\_t - 1 \texttt{bigr}\big) \\ &= -\frac{nT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log|F_t| - \frac{1}{2}\sum_{t=1}^{T}\mathbf{z}_t' F_t^{-1}\mathbf{z}_t. \end{aligned} \tag{60}$$

The identification algorithm consists of the follow step:

1. Construct a guess of initial numerical values for state space parameters A,C,D,Q,R.

2. Iterate on the Kalman filter

$$\hat{\mathbf{x}}(t+1|t) = A\hat{\mathbf{x}}(t|t-1) + K(t)\Big(\mathbf{y}_t - C\hat{\mathbf{x}}(t|t-1)\Big)$$

$$K(t) = AP(t|t-1)C'(CP(t|t-1)C' + R)^{-1}$$

$$\Lambda(t) = C'(CP(t|t-1)C' + R)$$

$$P(t+1|t) = A\Big(P(t|t-1) - P(t|t-1)C'\Lambda(t)^{-1}CP(t|t-1)\Big) + Q$$

to obtain estimates of sequences $\left\{\hat{\mathbf{x}}(t+1|t)\right\}_{t=1}^{T}$ and $\left\{P(t+1|t)\right\}_{t=1}^{T}$.

3. Use the sequences $\left\{\hat{\mathbf{x}}_{t+1|t}\right\}_{t=1}^{T}$ and $\left\{P_{t+1|t}\right\}_{t=1}^{T}$ to calculate the value for the log likelihood function.

4. Use numerical optimization methods to make better guesses as to the state space parameters until the log-likelihood function is maximized.

5. Go to step 2.

It is important to note that there must be constraints on A, C, D, Q, and R due to identification problems. In this thesis we use the MATLAB Optimization package to maximize the log-likelihood function. It implies Nelder-Mead algorithm. Moreover, each on step 2 we can calculate the information matrix $N \times N$ ($NumerOfParameters = N$) with elements (i, j)

$$I_{ij}(\theta) = \frac{1}{2}\sum_t \left\{ \mathrm{tr}\left[F_t^{-1}\frac{\partial F_t}{\partial \theta_i}F_t^{-1}\frac{\partial F_t}{\partial \theta_j}\right]\right\} + E\sum_t \left(\frac{\partial z_t}{\partial \theta_i}\right)' F_t^{-1}\frac{\partial z_t}{\partial \theta_i}, \quad i,j \in \{1,\dots,N\} \tag{61}$$

and by dropping the expectation operator from this expression we obtain a result that is asymptotically equivalent and in our case easier to evaluate. If we remember the Cramer-Rao inequality

$$Var(\hat{\theta}) \geqslant I(\theta)^{-1}$$

and that the maximum likelihood asymptotically satisfies the equality we obtain, as $T \to \infty$

$$Var(\hat{\theta}) = I(\theta)^{-1}.$$

In this way we found an estimate of estimator variance. This gives us an insight about the estimator significance. Unfortunately, we need to evaluate the derivatives of F_t and z_t. They result

$$\frac{\partial z_t}{\partial \theta_i} = -C\frac{\partial x_{t|t-1}}{\partial \theta_i} - \frac{\partial C}{\theta_i}\mathbf{x}_{t|t-1}$$
$$\frac{\partial F_t}{\partial \theta_j} = \frac{\partial C}{\partial \theta_j}P_{t|t-1}C' + C\frac{\partial P_{t|t-1}}{\partial \theta_i}C' + CP_{t|t-1}\frac{\partial C'}{\theta_i}$$
$$\frac{\partial \mathbf{x}_{t|t-1}}{\partial \theta_i} = \frac{\partial A}{\partial \theta_i}\mathbf{x}(t-1|t-1) + A\frac{\partial \mathbf{x}_{t-1|t-1}}{\partial \theta_i}\left(+\frac{\mu}{\partial \theta_i}\right)$$
$$\frac{\partial P_{t|t-1}}{\partial \theta_i} = \frac{\partial A}{\theta_i}P_{t-1|t-1}A' + A\frac{\partial \mathbf{x}_{t-1|t-1}}{\theta_i}A' + AP_{t-1|t-1}\frac{\partial A}{\theta_i} + \frac{\partial Q}{\theta_i}$$

$$\frac{\partial \mathbf{x}_{t|t}}{\theta_i} = \frac{\partial \mathbf{x}_{t|t-1}}{\partial \theta_i} + \frac{\partial \mathbf{P}_{t|t-1}}{\partial \theta_i}C'F_t^{-1}z_t + P_{t|t-1}\frac{\partial C}{\partial \theta_i}F_t^{-1}z_t -$$
$$P_{t|t-1}C'F_t^{-1}\frac{\partial F_t}{\partial \theta_i}F_t^{-1} + P_{t|t-1}C'F_t^{-1}\frac{\partial z_t}{\partial \theta_i}$$

4.4 DIEBOLD-NIELSON-SIEGEL MODELS

In this section we will consider the identification problem of state space models with GARCH(1,1) errors. We will concentrate on DNS-GARCH models but all the discussion below can be banally generalized on more general system state space system.

Now let consider the DNS model Equation 27- Equation 28 (with the factor λ to be constant over time) rewritten with control theory parameters notation. Since we are interested to estimate the volatility, let insert $\varepsilon_{1,t}$ in the state vector forming the

augmented state $z(t) = (x(t)', \varepsilon_1(t))'$. The augmented model is given by

$$z(t+1) = \begin{bmatrix} A & 0_3 \\ 0_3' & 0 \end{bmatrix} z(t) + \begin{bmatrix} v(t) \\ \varepsilon_1(t+1) \end{bmatrix} \qquad t = 0, \dots, T-1 \tag{62}$$

that rewritten is

$$z(t+1) = \bar{A} z(t) + \begin{bmatrix} v(t) \\ \varepsilon_1(t+1) \end{bmatrix} \tag{63}$$

$$\begin{bmatrix} v(t) \\ \varepsilon_1(t+1) \mid y^t \end{bmatrix} \sim N\left(0, \begin{bmatrix} Q & 0 \\ 0 & h(t+1) \end{bmatrix}\right)$$

where all parameters are those presented above. As we can see there is no input but a mean coefficient is present. Thus the state equation we are considering is The measurement equation is given by

$$\begin{aligned} y(t) &= C(\lambda)x(t) + D\varepsilon_1(t) \\ &= \begin{bmatrix} C(\lambda) & D \end{bmatrix} z(t) \\ &= \bar{C} z(t) + w(t) \qquad w \sim N(0, R) \qquad t = 0, ..., T-1 \end{aligned} \tag{64}$$

If the volatility component appears in the state equation the augmented model becomes

$$\begin{cases} z(t+1) = \begin{bmatrix} A & 0_3 \\ 0_3' & 0 \end{bmatrix} z(t) + \begin{bmatrix} I & B \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v(t) \\ \varepsilon_2(t+1) \end{bmatrix} \\ y(t) = \begin{bmatrix} C(\lambda) & 0 \end{bmatrix} z(t) + \varepsilon(t) \end{cases} \tag{65}$$

$$w(t) \sim N(0, R), \qquad \begin{bmatrix} v(t+1) \\ \varepsilon_2(t+1) \mid y^t \end{bmatrix} \sim N\left(0, \begin{bmatrix} Q & 0 \\ 0 & q(t+1) \end{bmatrix}\right)$$

where $z(t) = (x'(t), \varepsilon_2(t))'$.

Let rewrite the Equation 62-Equation 64 and Equation 65 with the usual notation

$$\begin{cases} \boldsymbol{z}(t+1) = \bar{A}\boldsymbol{z}(t) + G\boldsymbol{n}(t+1) & \boldsymbol{n}(t+1)|\boldsymbol{y}^t \sim N(0, \bar{Q}(t+1)) \\ \boldsymbol{y}(t) = \bar{C}z(t) + \boldsymbol{w}_t & \boldsymbol{w}(t) \sim N(0, R) \end{cases}$$

When the measurement equation has the GARCH component we have

$$G = I_4, \qquad \bar{C} = \begin{bmatrix} C(\lambda) & D \end{bmatrix}$$

$$\bar{Q} = \begin{bmatrix} Q & 0 \\ 0 & h(t+1) \end{bmatrix}, \qquad \boldsymbol{n}(t+1) = \begin{bmatrix} \boldsymbol{v}(t) \\ \varepsilon_1(t+1) \end{bmatrix}$$

while when the GARCH component appears in the state equation we have

$$G = \begin{bmatrix} I & B \\ 0 & 1 \end{bmatrix}, \qquad \bar{C} = \begin{bmatrix} C(\lambda) & D \end{bmatrix},$$

$$Q = \begin{bmatrix} Q & 0 \\ 0 & q(t+1) \end{bmatrix}, \qquad \boldsymbol{n}(t+1) = \begin{bmatrix} \boldsymbol{v}(t) \\ \varepsilon_2(t+1) \end{bmatrix}$$

The prediction step is given by

$$\hat{\boldsymbol{z}}(t+1 \mid t) = \bar{A}\hat{\boldsymbol{z}}(t \mid t)$$

$$\hat{\boldsymbol{z}}(t+1 \mid t) = \bar{A}\hat{\boldsymbol{z}}(t \mid t-1) + K(t)\left(\boldsymbol{y}_t - \bar{C}\hat{\boldsymbol{z}}(t \mid t-1)\right)$$

$$K(t) = \bar{A}P(t \mid t-1)\bar{C}'(\bar{C}P(t \mid t-1)\bar{C}' + R)^{-1}$$

$$\begin{aligned} P_{t+1|t} &= \bar{A}P(t \mid t)\bar{A}' + G\bar{Q}G' \\ &= A\left\{P(t \mid t-1)\left[I - \bar{C}'\Lambda^{-1}(t)\bar{C}P(t \mid t-1)\right]\right\}\bar{A}' + G\bar{Q}G' \end{aligned}$$

and the update step is given by

$$\hat{\boldsymbol{z}}(t+1 \mid t+1) = \bar{A}\hat{\boldsymbol{z}}(t \mid t-1) + L(t+1)\left[\boldsymbol{y}(t+1) - \bar{C}A\hat{\boldsymbol{z}}(t \mid t)\right]$$

$$P(t+1 \mid t+1) = \left[I - L(t+1)\bar{C}\right]P(t+1 \mid t)$$

where $L(t+1)$ is called *filter gain* and it is

$$L(t+1) = P(t+1 \mid t)\bar{C}'\Lambda(t+1)^{-1}$$

Since $h(t+1)$ is a function of the unobserved value $\varepsilon_1(t)$ and its past values, we will not be able to compute the necessary value of $h(t+1)$ at time t. If past values of the disturbances $\varepsilon_1(t)$ were directly observable, the model would be a conditionally Gaussian one and so the Kalman filter it will be quasi-optimal. Since the Kalman filter produces the minimum mean square linear estimator (conditional on the past), a natural solution would be to take the expectations of latent variables in (30)

$$h(t) = \omega + \alpha\,\mathbb{E}\left[\varepsilon^{*2}(t) \mid \mathbf{y}^t\right] + \beta h(t-1) \qquad t = 1, ..., T-1$$

and $\mathbb{E}\left[\varepsilon_1^2(t) \mid \mathbf{y}^t\right]$ is precisely the estimate provided by the Kalman filter. Indeed,

$$\varepsilon_1(t-1) = \hat{\varepsilon}_1(t-1 \mid t-1) + \tilde{\varepsilon}_1(t-1)$$

is the decomposition in estimate and estimation error; it can easily demonstrated that these two terms are uncorrelated and so we obtain

$$\begin{aligned}
&\mathbb{E}\left[\varepsilon_1^2(t-1) \mid \mathbf{y}^{t-1}\right] \\
&= \mathbb{E}\left[\varepsilon_1(t-1) \mid \mathbf{y}^{t-1}\right]^2 + \mathbb{E}\left(\varepsilon_1(t-1) - \mathbb{E}\left[\varepsilon_1(t-1) \mid \mathbf{y}^{t-1}\right]\right)^2 \\
&= \mathbb{E}\,\hat{\varepsilon}_1(t-1 \mid t-1)^2 + \mathbb{E}\,\tilde{\varepsilon}_1(t-1)^2
\end{aligned}$$

where $\mathbb{E}\,\hat{\varepsilon}_1(t-1 \mid t-1)^2$ is the third element of $\hat{z}(t-1 \mid t-1)$ and $\mathbb{E}\,\tilde{\varepsilon}_1(t-1)^2$ is the element in the position $(3,3)$ of $P(t-1 \mid t-1)$. The initial conditions of the Kalman filter are given by

$$z(0 \mid -1) = \mathbb{E}\,z(t) = 0$$

and

$$P(0 \mid -1) = \begin{bmatrix} \Sigma_x & 0 \\ 0 & h_0 \end{bmatrix}$$

where $h_0 = \frac{\omega}{1-\alpha-\beta}$ is the unconditional variance, the covariance matrix of ε_t is given by

$$\Sigma_\varepsilon(h_t) = Eh_t DD' + R. \tag{66}$$

and the matrix variance Σ_x is the solution of

$$A\Sigma_x A' - \Sigma_x = Q.$$

It is important to notice that the covariance matrix of $\varepsilon_1(t)$ it is not $h_t DD' + R$ how Koopman claims. Actually, this last one term represents the covariance matrix based on past information, that is

$$Var(\varepsilon(t) \mid \varepsilon_1(t)) = h_t DD' + R$$

with h_t deterministic. Once produced the estimates of $z(t)$ by the Kalman filter at time $t = 0, \ldots, T-1$, we will use these ones in the likelihood function in order to estimate the parameters. Thus, the likelihood function to be maximized is

$$l_T(\theta) = -\frac{nT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log|F(t \mid t-1)| \tag{67}$$

$$-\frac{1}{2}\sum_{t=1}^{T}\mathbf{r}(t \mid t-1)'F(t \mid t-1)^{-1}\mathbf{r}(t \mid t-1). \tag{68}$$

where

$$\mathbf{r}(t \mid t-1) = \mathbf{y}(t) - C\mathbf{x}(t \mid t-1), \qquad F(t \mid t-1) = CP(t \mid t-1)C' + R$$

and the vector of parameters are $\theta = (A, \lambda, Q, R, D, \omega, \alpha, \beta)$ or $\theta = (A, \lambda, Q, R, B, \omega, \alpha,\ beta)$ respectively in dependence of the location of GARCH component. The last model, named as DNS-XYGARCH, is given by

$$\mathbf{x}(t+1) = A\mathbf{x}(t) + B\varepsilon_2(t) + \boldsymbol{\nu}(t) \tag{69}$$

$$y(t) = C(\lambda)\mathbf{x}(t) + D\varepsilon_1(t) + \boldsymbol{w}(t) \tag{70}$$

Defining

$$\eta(t) = B\varepsilon_2(t) + \boldsymbol{v}(t)$$
$$\varepsilon(t) = D\varepsilon_1(t) + \boldsymbol{w}(t)$$

we have

$$\eta(t) \mid \boldsymbol{x}^t, \boldsymbol{y}^t \sim N(0, Q + q_t BB')$$
$$\varepsilon(t) \mid \boldsymbol{x}^t, \boldsymbol{y}^t \sim N(0, R + h_t DD')$$

Then we apply the Kalman filter to the model

$$\boldsymbol{z}(t+1) = \bar{A}\boldsymbol{z}(t) + G\boldsymbol{n}(t+1) \tag{71}$$
$$\boldsymbol{y}(t) = \bar{C}\boldsymbol{z}(t) + \boldsymbol{w}(t) \tag{72}$$

where

$$\boldsymbol{z}(t) = \begin{bmatrix} \boldsymbol{x}(t) \\ \varepsilon_1(t) \\ \eta_2(t) \end{bmatrix}, \bar{A} = \begin{bmatrix} A & 0_3 & 0_3 \\ 0_3' & 0 & 0 \\ 0_3' & 0 & 0 \end{bmatrix},$$

$$\boldsymbol{n}(t+1) = \begin{bmatrix} \boldsymbol{v}(t) \\ \varepsilon_1(t) \\ \varepsilon_2(t) \end{bmatrix}, \qquad G = \begin{bmatrix} I_3 & 0_3 & B \\ 0_3' & 1 & 0 \\ 0_3' & 0 & 1 \end{bmatrix}, \qquad \bar{C} = \begin{bmatrix} C & D & 0 \end{bmatrix}$$

In order to avoid identification issues a restriction is required. Koopman in [18] suggests the normalization $DD' = 1$ but we choose to fix ω at a very small value close to zero. The actual choice for the constraint to avoid identification problems is irrelevant to the results of the analysis since all methods are equivalent to a scaling factor.
Finally, we would aspect the parameters to be estimated consistently. Indeed, the DNS-GARCH is estimated with the EM algorithm, that, under stationary and other mild conditions, provides consistent estimates.

We remember that the estimation of state space models presented in Section 2.2 is performed by the EM algorithm and that the E-step evaluates the optimal estimate minimizing the *conditional* MSE.

4.5 VALIDATION

Once obtained the estimated models we want to search for the most suitable model. In order to do this we can simply compare different models. But the question is: What to compare? Normally there exists a number of ways to evaluate a model that, in general, depends on the application. We are interested to goodness of forecasting, i.e, how well the model is capable of reproducing and predicting the data (volatility). Thus we will work with *k-step ahead model predictions* $\hat{y}_k(t|m)$ where m is the model used. We will use a reasonable way to compare models, that is, we will compare the cost function

$$J_k(m) = \frac{1}{T}\sum_{t=1}^{T} |y_\tau(t) - \hat{y}_{\tau,k}(t|m)|^2 \tag{73}$$

where τ is the maturity time considered and $y(t)$ are the yields. It represents the mean square prediction k-ahead error. To give a measure of this fit we define

$$R^2 = 1 - \frac{J_k(m)}{\frac{1}{T}\sum_{t=1}^{T} |y(t)|^2} \tag{74}$$

where $y(t)$ has been detrended to zero mean. Given that the quantity $J_k(m)$ will depend on the actual data for which the comparison is made, it is natural to consider the expectation of the model fit with respect to θ_T

$$\bar{J}_k(\mathcal{M}) = E\bar{J}_k(\mathcal{M}(\hat{\theta}_T))$$

A common error is to use the *identification data* to assess the quality measure for the model, both in model comparison and in validation of single model. This error is also made by [18]. While the validation question will be treated in a while, the use of identification data comports that observation J_k is not an unbiased estimate of $\bar{J}_k$. On other hand, using *validation data* yields an correct estimate of $\bar{J}_k$. Thus we will compare the quantity

$$J_k(m) = \frac{1}{T}\sum_{t=1}^{T} |y_\tau(t) - \hat{y}_{\tau,k}(t|\hat{\theta_T})|^2 \tag{75}$$

also known as Root Mean Squared Error (RMSE). There exist vary model structure selection criteria, such as, AIC, BIC, and MDL. We will consider AIC and BIC.

The AIC criterion is defined by the formula

$$2p - 2\ln(L)$$

where p is the number of parameters and L is the value of likelihood. Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. Hence AIC not only rewards goodness of fit, but also includes a penalty that is an increasing function of the number of estimated parameters. The penalty discourages overfitting (increasing the number of parameters in the model almost always improves the goodness of the fit). In practice we will use the corrected AIC criteria, i.e.,

$$AIC_c = AIC + \frac{2k(k+1)}{T-k-1}$$

where T denote the sample size. It is strongly recommended to use AIC_c. For example see [3]. The BIC criteria formula is given by

$$BIC = -2\ln L + k\ln(T)$$

with the usual notations. The BIC criterion is similar to AIC, but it tends to penalize free parameters more strongly than does AIC.

Other comparison criteria we will use is LR-statistic (*Likelihood Ratio*). We note that this statistic is reasonable only if we compare two models $\mathcal{M}_1$ and $\mathcal{M}_2$ such that $\mathcal{M}_1 \subset \mathcal{M}_2$. This is why the likelihood ratio test compares specifications of nested models by assessing the significance of restrictions to an extended model with unrestricted parameters. In our application, we will use this test to verify GARCH components on noise errors. The test uses the following algorithm:

- Maximize the log-likelihood function $l(\theta)$ under the restricted and unrestricted model assumptions. Denote the MLEs for the restricted and unrestricted models $\hat{\theta}_{CON}$ and $\hat{theta}$,respectively. In our case $\hat{\theta}_{CON}$ is the parameter evaluated under the null hypothesis that GARCH component is assent, i.e, where the model has Gaussian noise.

- Evaluate the loglikelihood objective function at the restricted and unrestricted MLEs, i.e., $\hat{l}_{CON} = l(\hat{\theta}_{CON})$ and $\hat{l} = l(\hat{\theta})$
- Compute the likelihood ratio test statistic, $LR = 2\left(\hat{l} - \hat{l}_{CON}\right)$.
- If LR exceeds a critical value (C_α) relative to its asymptotic distribution, then reject the null, restricted model in favor of the alternative, unrestricted model. Under the null hypothesis, LR is χ^2_d distributed with d degrees of freedom. The degrees of freedom for the test (d) is the number of restricted parameters and the significance level of the test α determines the critical value (C_α).

Normally, once the model is fitted, correlation of the standardized innovations are quantified and compared using Q-test and ARCH test. For ARMA-GARCH and state-space models with GARCH innovation, for example DNS-GARCH, the validation procedures is slightly different. The main difference consists on the role of innovations, meaning that, in the DNS-GARCH case, the GARCH component represents the model innovation driving the returns and unobserved processes. Thus, we cannot directly use tests such as Q-test or Engle's test to identify heteroscedasticity on volatility. In order to validate the GARCH innovation hypothesis, we will first fit the homoscedasticity model to the reduced form and then, we test the heteroscedasticity with the tests proposed above.

Although in time series models, such as pure-GARCH or ARMA-GARCH, the GARCH component can be identified analyzing the residual with the data to be fitted, in state space models the innovations can be either on state equation or measurement equation. In other hand, we have available only the process $\{y_t\}$ while the state process is unobserved and we cannot calculate the residuals. In order to test the GARCH component on innovation we propose this steps:

1. fit the homoscedasticity model to the reduced form, that is, without regarding of an eventual GARCH components;
2. if an autocorrelations on squared sresiduals is identified, proceed to fit the model with the GARCH component only on state equation.

3. if an autocorrelations on squared residuals is identified again and tests validate the heteroscedasticity hypothesis, it means the GARCH component can also be in measurement equation.

4. fit the model with GARCH components in both state and measurement equations.

As we know the parameter estimation procedure selects the best model within the chosen model structure and the constraints. The problem of validation can be summarized on the following questions:

- Does the model agree sufficiently well with the observed data?
- Is the model good enough for our purpose?
- Does the model describe the "true system"?

We shall focus on question 1. To do this will use the *residual analysis*.

The residual is the part of the data that the model could not reproduce, i.e.

$$\varepsilon(t) = y(t) - \hat{y}(t|\hat{\theta}_T) \tag{76}$$

where is the 1-ahead step prediction. Let us considered the quantity

$$R_\varepsilon^T(\tau) = \frac{1}{T}\sum_{t=\tau}^{T} \varepsilon(t)\varepsilon(t-\tau) \tag{77}$$

where $0 \leqslant \tau \leqslant \tau_{MAX}$ and τ_{MAX} is chosen opportunely small. If we find this quantity to be not small for $h \neq 0$, i.e., there is a correlation among the residuals themselves. This means that $y(t)$ could have been better predicted. If a model well describe to true model we will aspect the residual error to be white. Here we present briefly the *Whiteness Test*.

We can now understand that the number $R_\varepsilon^T(t)$ carries information about whether the residuals can be regarded as white.

Indeed, suppose $\{\varepsilon(t)\}$ is a white noise process with zero mean and variance λ. Then it can be shown that

$$\frac{1}{\sqrt{T}} \sum_{t=m}^{T} \varepsilon(t) \begin{bmatrix} \varepsilon(t-1) \\ \vdots \\ \varepsilon(t-m) \end{bmatrix} \to N(0, \lambda^2 I) \tag{78}$$

and the $k-$th row of this vector is $\sqrt{N} R_\varepsilon^T(k)$. Therefore, if ε is white we can show that

$$\frac{N}{R_\varepsilon^T(0)} \sum_{\tau=1}^{m} R_\varepsilon^T(\tau)^2 \to \chi^2(m) \quad \text{in law.}$$

This can be regarded as a test statistic. We refuse the null hypothesis, i.e. hypothesis of whiteness with significance α if statistic values are bigger than k_α where $P_{\chi^2(m)} = \alpha$.

We shall also discuss the *goodness of fit*. Naturally, this is strictly correlated with the residuals, in particular with the quantity Equation 75. We define the fitness value as

$$\text{fit} = 100 \left(1 - \frac{\|y - \hat{y}_k(t|\theta^T)\|}{\|y - \text{mean}(y)\|} \right) \tag{79}$$

where the quantity $\|y - \hat{y}_k(t|\theta^T)\|$ is the RMSE. Finally, we will analyze also the filtered errors, that is, the difference between the observed yield and the filtered estimate obtained from the Kalman filter. We newly remark that all this work must be done on *validation data* and not as does Koopman et al. [18].

5 SIMULATIONS

This chapter is concerned about the simulations of the considered models. Before the estimation process, in Section 5.1 we shall analyze the data used to perform the simulations in order to infer some styled facts. In particular, our interest in focused on to verify heteroscedasticity. In section Section 5.2 we shall present and discuss the estimation results of various models. In section Section 5.3 we shall discuss the predictive capacity and the validation question.

5.1 DATA STATISTICS

In this thesis we will use monthly constant maturity yields of US government zero-coupon bonds obtain from the Federal Bank dataset. This dataset consists on monthly yields[1] for the the period from January 1994 until December 2013. We consider the maturities of 3, 6, 12, 24, 60, 120, 240 months. We will not consider 1-month maturity since they are only available from 2011 onwards. Identification data starts from January 1994 until December 2003 while validation data start from January 2004 until December 2013. In some figure will consider only the period 2007-2013 as period 2004-2007 has no significance for our purposes, that is, we are interested in volatile periods. Figure 6 consists on the so called cross section of the yields over the sample period. As we can see interest rates has heavily varied over time. The long term trend is downwards and short term interest rates approach zero. It is interesting to see two sudden variation, namely in mid-end 2011 and 2008: they corresponds to twin towers' attack and financial crisis, respectively.

At this point a spontaneous question arises: Can we already say anything about the heroscedasticity in the data? As we discussed in Section 4.5 we can verify the presence of GARCH ef-

1 The monthly yields are simply the mean of daily yields.

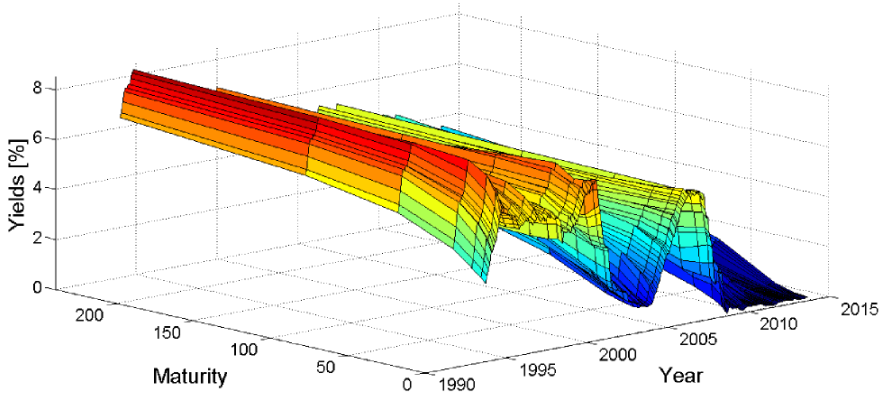


Figure 6: Cross section of yields from 1994 until December, 2013. The most significant variations are in correspondence of Twin Towers attack 2001 and mortgage crisis 2008, respectively.

fect by the analysis of returns and squared returns. First, let us convert the yields in log returns. The 3, 12, 60, 120-months log returns are presented in Figure 7 and their ACF is presented in Figure 8. The autocorrelation slowly decreases and this indicates that we are nearby the non-stationarity region. This will be confirmed later. Without regarding the squared return analysis, we can see evident clustering phenomenon on returns. This tell us that high probably there exists an ARCH behavior, at least. Thus, let us analyze the ACF of the squared returns to verify this evidence. In Figure 9 we present the ACF of squared returns. They strongly confirm our hypothesis, that is, there is a GARCH component on time series. In order to also give a formal answer, in Table 1 we report the Q-test results on squared returns for the respective maturities with a complete rejection of homoscedasticity hypothesis.

As we saw in Section 2.3.1, since the first component of $C(\lambda)$ is the only one that equals one as $\tau \to \infty$, its corresponding β_{1t} coefficient is linked with the long-term interest rate. Let define

$$M_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau - e^{-\lambda\tau}}\right)$$

Then, by defining the slope of the yield curve as $M_t(\infty) - M_t(0)$, it is easy to verify that the slope converges to $-\beta_{2t}$ for a given t. Finally, the shape of the yield can be defined by $[M_t(\tau_1) - M_t(0)] -$

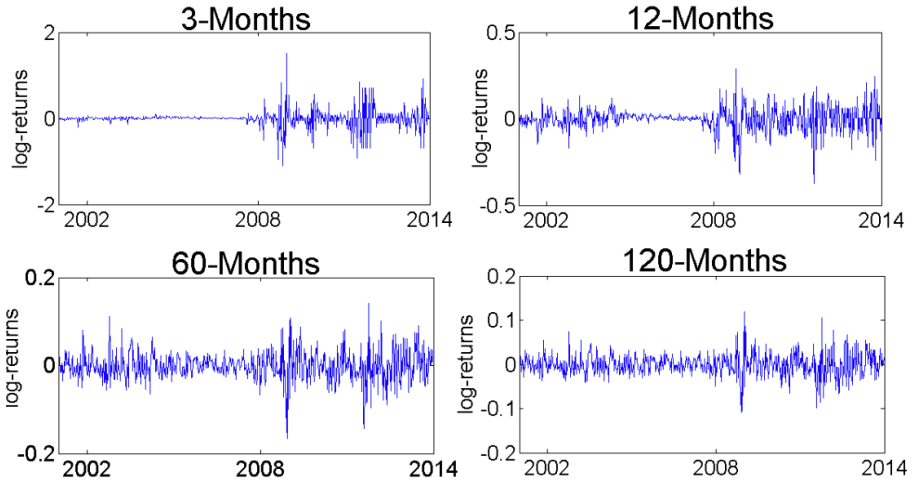
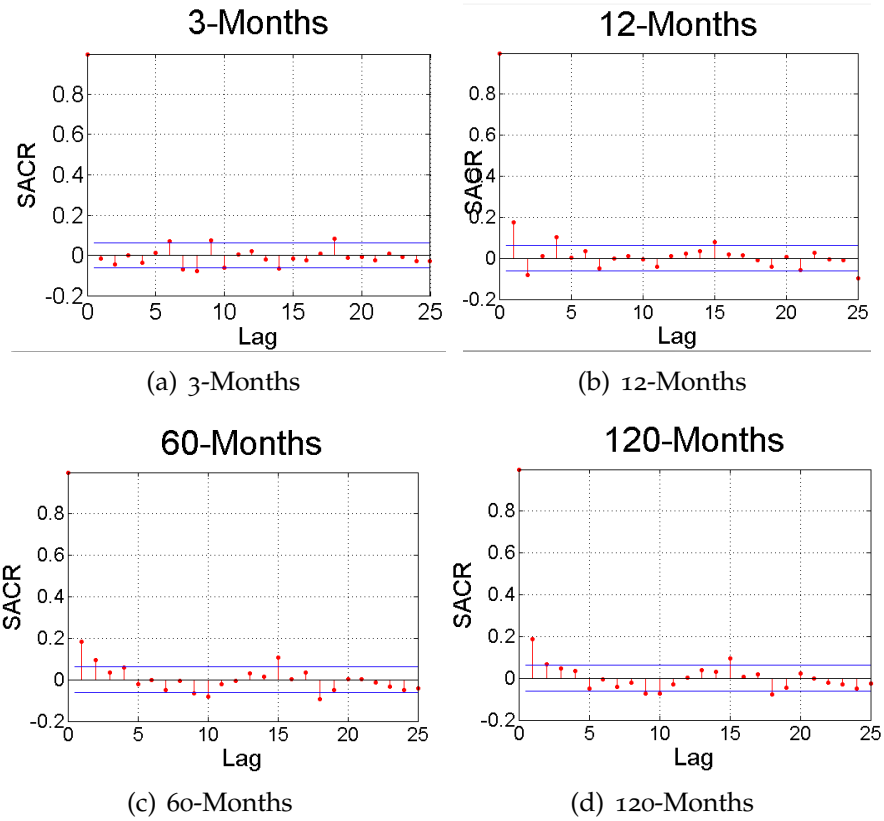


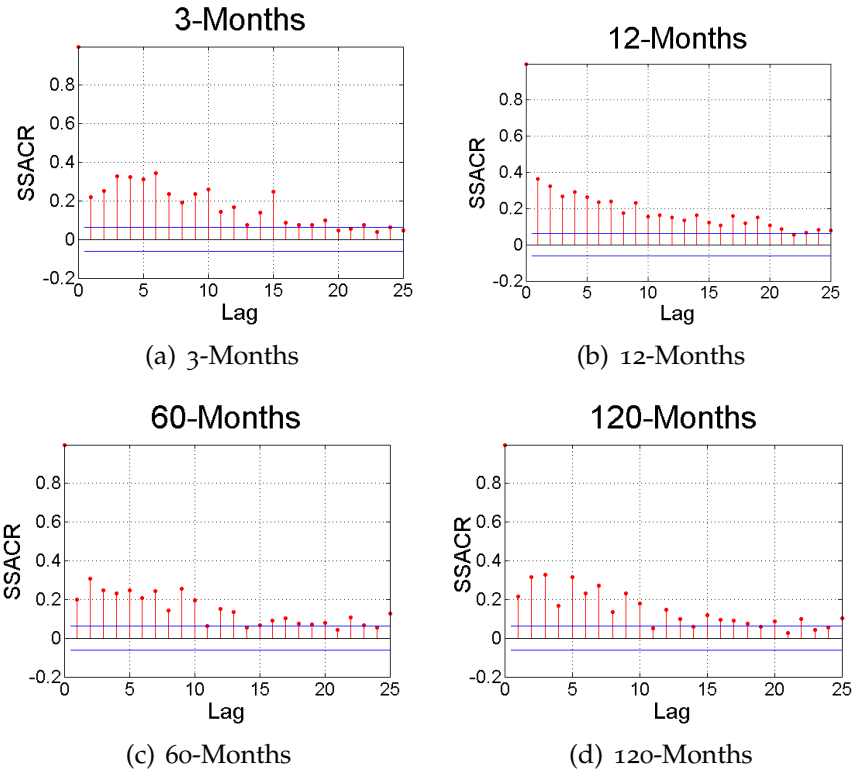
Figure 7: Log return of interest yields for 3, 12, 60, 120-months maturities from 2001 until December 2013.



(a) 3-Months (b) 12-Months

(c) 60-Months (d) 120-Months

Figure 8: Sample autocorrelation of *returns* for 6, 12, 60, 120-months maturities.



(a) 3-Months (b) 12-Months

(c) 60-Months (d) 120-Months

Figure 9: Sample autocorrelation of *squared returns* for 6, 12, 60, 120-months maturities.

Maturity	**Statistic**	
(Months)	Q-test	k_α
3	967	31.4
6	1382	31.4
12	928	31.4
24	460	31.4
60	732	31.4
120	656	31.4
240	731	31.4

Table 1: Test statistic of squared residuals, k_α is the critical value. The significance is $\alpha = 0.5$ and the number of lags is $m = 20$.

$[M_t(\infty) - M_t(\tau_1)]$ for a medium maturation τ_1 , say, two years, and for a given t. It can be shown that this shape approximately approaches β_{3t}. Thus, the longest maturity approximates the level of yields curve. Based on this heuristic we define the slope factor as $y_t(240) - y_t(3)$, where $y_t(\tau)$ is the yield under maturity τ (in months). The curvature, proxies the mid-term behavior and so we can define it as $[y_t(24) - y_t(3)] - [y_t(240) - y_t(24)]$. Shorter maturity yields are more volatile than longer maturities. This can be seen by the standard deviation decreasing. The only exception is for 6-month yields, as also remarked by Koopman et al. [18]. In addiction, the high correlations for all maturities at different horizons indicates persistence of the yields dynamic. This persistence tends to be more remarkable with long-term maturity. Also Slope and Curvature present high persistence.

5.2 MODEL FITTING ON IDENTIFICATION AND VALIDATION DATA

In Section 5.2.1 we shall discuss the estimation of the standard DNS model, that is, without any GARCH component. In Section 5.2.2 we shall argue the introduction of GARCH shocks on noise errors.

5.2.1 *Pure DNS model*

The first model we analyze is the AR(1) DNS model, that is,

$$\boldsymbol{x}_{t+1} = \begin{pmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{pmatrix} \boldsymbol{x}_t + \boldsymbol{v}_t \qquad \boldsymbol{v}_t \sim \mathrm{NID}(0, Q)$$

$$\boldsymbol{y}_t = C(\lambda)\boldsymbol{x}_t + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim \mathrm{NID}(0, \sigma_r^2 I_7), \qquad t = 1, \ldots, T,$$

The covariance matrix Q is restricted to be diagonal. We analyze 7 maturities and so the matrix C is 7×3 dimensional. The estimation results are presented in Table 2. The diagonal values are very close to one, reflecting the presence of high persistence in the yields and empirical factor proxies. The parameter λ is estimated as 0.0436 with an standard error of .00007: this means

that the estimate has high significance. The fact that the eigenvalue are so close to one, advise us to use a less parsimonious model accounting for the possibility of eventual correlation between the factors. The measurement errors covariance matrix $R = \sigma_r^2 I = 0.01I$ with a standard error 0.007. In Table 3 we

(a) Estimates of A and μ. The standard errors are shown in subscript.

A	$\beta_{1,t-1}$	$\beta_{2,t-1}$	$\beta_{3,t-1}$
$\beta_{1,t}$	$0.99_{0.03}$	0	0
$\beta_{2,t}$	0	$0.99_{0.02}$	0
$\beta_{3,t}$	0	0	$0.985_{0.05}$

(b) Estimate of diagonal matrix Q.

Q	$\beta_{1,t-1}$	$\beta_{2,t-1}$	$\beta_{3,t-1}$
$\beta_{1,t}$	$0.22_{0.009}$	0	0
$\beta_{2,t}$	0	$0.09_{0.001}$	0
$\beta_{3,t}$	0	0	$0.27_{0.05}$

Table 2: Estimation of the AR(1) DNS model. Panel 5.1(a) reports the coefficient estimates of A. Panel 5.1(b) reports the estimate of covariance matrix Q.

present the estimation results of the standard DNS model, that is, with A unconstrained, Q symmetric ed semidefinite positive and R diagonal definite positive. In order to ensure positivity and symmetry on Q and R we use the Cholesky factorization $Q = qq'$ and $R = r^2$ with r diagonal of dimension 7×7. As we can see the diagonal values of A are less close to one than AR(1)-DNS model. In addiction, the log-likelihood function (see Table 5) is higher than that of AR(1) and the lagged value of the third factor, which proxies for the curvature, has a significant influence on the slope factor. In Figure 5.10(a) and Figure 5.10(b), the empirical level, slope, curvature and the filtered latent factors of DNS model are respectively reported. In order to show the that the β_i represent the three main characteristics of the term structure of interest rates, in Figure 5.10(c) we present the filtered latent factors of DNS model obtained by the Kalman filter) compared with the empirical factors. In particular we plot

$-\beta_{2t}$ and a scaled β_{3t}. In Figure 13 we present the filtered esti-

(a) Estimates of A and . The standard errors are shown in subscript.

A	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$
$\beta_{1.t-1}$	$0.94_{0.03}$	$0.002_{0.01}$	$0.011_{0.01}$
$\beta_{2,t-1}$	$-0.03_{0.02}$	$0.92_{0.02}$	$0.07_{0.02}$
$\beta_{3,t-1}$	$0.12_{0.07}$	$0.06_{0.03}$	$0.83_{0}.04$

(b) Estimate of Q. Since Q is forced to be symmetric we report only the upper triangular matrix (6 parameters to be estimated).

Q	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$
$\beta_{1.t}$	$0.07_{0.008}$	$-0.06_{0.009}$	$0.018_{0.01}$
$\beta_{2,t}$		$0.11_{0.01}$	$-0.03_{0.02}$
$\beta_{3,t}$			$0.54_{0.05}$

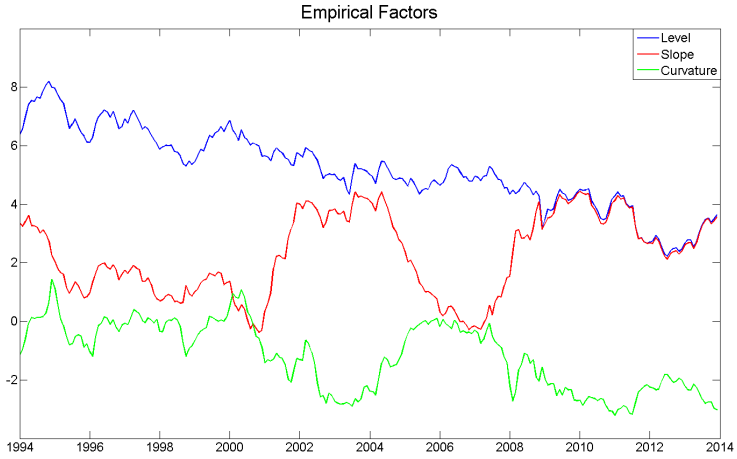
Table 3: Estimation of the standard DNS model. Panel 5.2(a) reports the estimate of A. Panel 5.2(b) reports the estimate of covariance matrix Q.

mates of the yields compared with the reference yields (that is the identification data). In particular, in Figure 12 we plot the filtered estimates of the yields for 4 maturities. In order to verify the fitting property of validation data, in Figure 14 we plot the filtered yields on the validation data (from January 2004 until 2014).

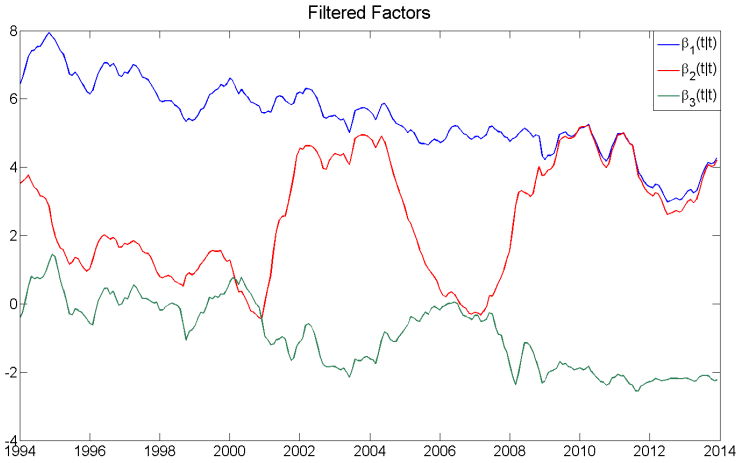
As we can see, the goodness of fit is less than that obtained using identification data. This behavior is reasonable and give us the a better description on real fitting capacity.

5.2.2 *DNS-GARCH models*

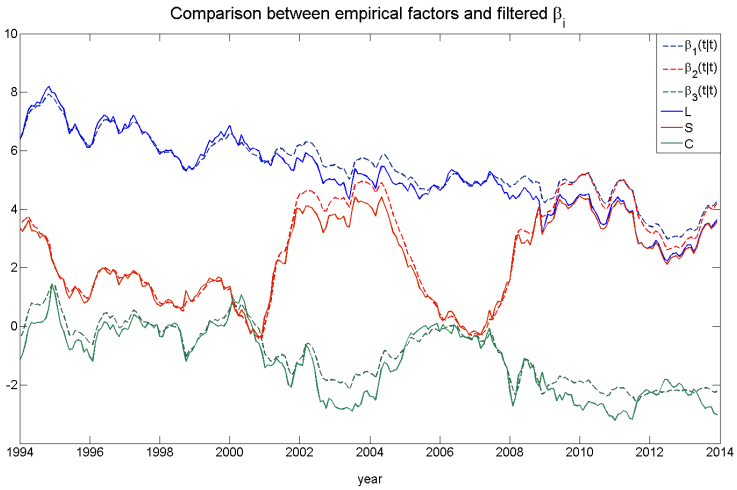
In this subsection, we will estimate the DNS model with GARCH errors. The likelihood maximization algorithm used to estimate the DNS model it is not suitable for DNS-GARCH estimation, since we need to constraints the stationarity condition on GARCH process, that is, $\alpha + \beta < 1$. For identification purposes we fix $\omega = 0.0001$. Adding a GARCH component on errors, should



(a) Empirical Factors (L, S, C).



(b) Filtered latent factors.



(c) Filtered latent factors.

Figure 10: Figure 5.10(a): empirical Level, Slope, Curvature. $L = y(3M)$, $S = y(120M) - y(3M)$, $C = 2y(24M) - y(120M) - y(3M)$. Figure 5.10(c): filtered β_i.

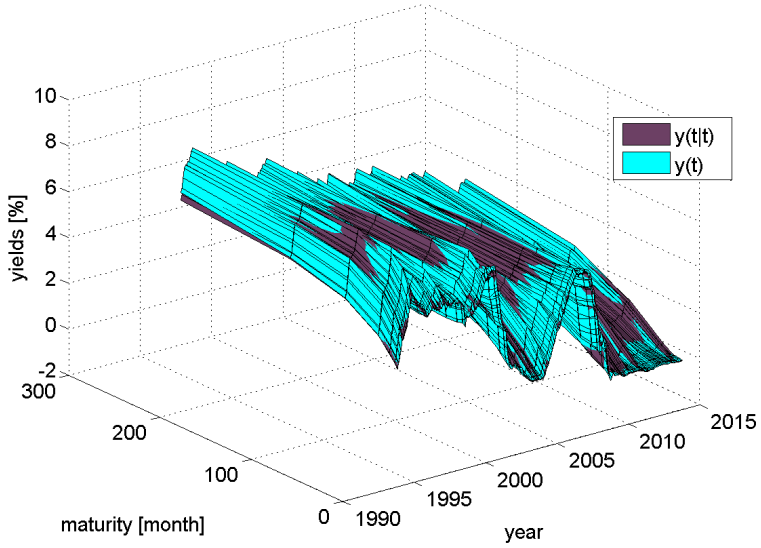
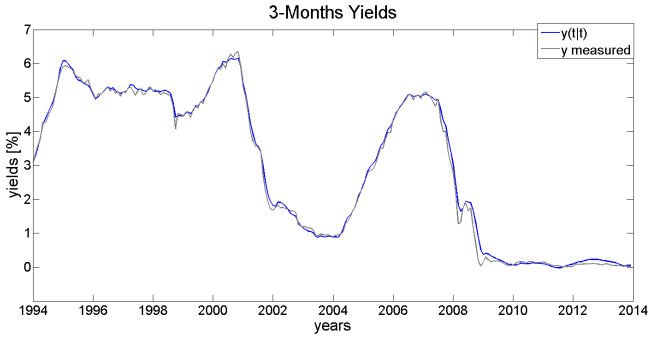


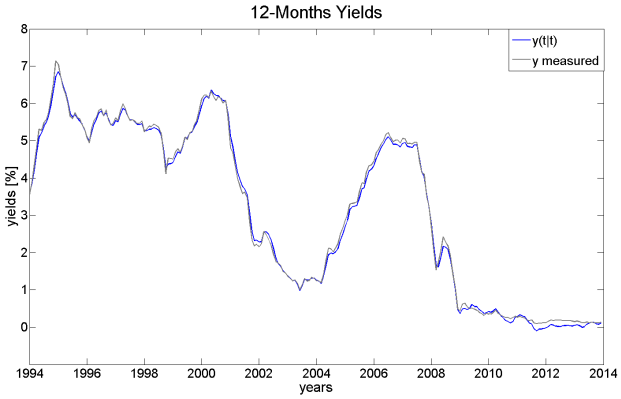
Figure 11: Filtered Yields for all the maturities.

permit us to capture the clustering phenomena. In particular, when the GARCH volatility of the common shock component is low and relatively constant, the measurement error is close to a white noise process. In other hand, when volatility increases, latent shocks affect the yield curve and the measurement error cannot be characterized as such. This specification could indicate that the three factor structure does not suffice in fitting the shape of the yield curve during these periods in time. Table 4 shows the estimated parameter of DNS-GARCH. The estimated parameter λ is 0.0596. How we can see some estimate has high standard errors meaning that the estimate is not significant. In Figure 15 and Figure 16 we report the estimate of matrices D and the volatility, respectively.

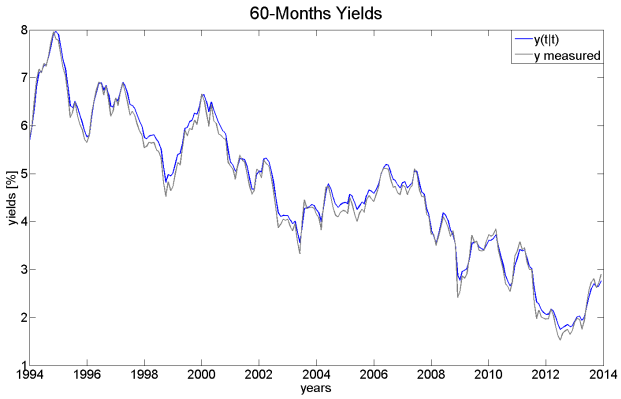
We also implemented DNS with GARCH in the state equation (DNS-XGARCH) and different GARCH specification as T-GARCH and E-GARCH. Here we will not report the results since GARCH on state equation doesn't considerately improve the standard DNS performances. Finally, in Table 5 we report the likelihood function of each model and other statistics such as, LR statistic, BIC and AIC. The LR test rejects the null hypothesis, that is, the omoscedasticity, with a statistic equal to 610 and a critical value equal to 5.99.



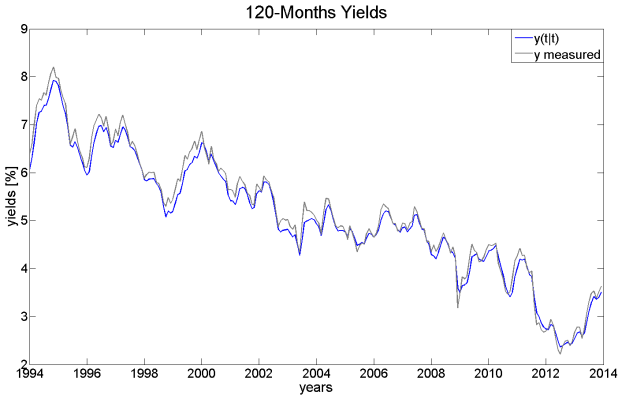
(a) 3-Months yields estimation



(b) 12-Months yields estimation



(c) 60-Months yields estimation



(d) 120-Months yields estimation

Figure 12: Filtered Yields for vary maturities.

(a) Estimates of A and μ. The standard errors are shown in subscript.

A	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$
$\beta_{1.t-1}$	$0.99_{0.04}$	$-0.004_{0.006}$	$0.0017_{0.01}$
$\beta_{2,t-1}$	$-0.002_{0.02}$	$0.98_{0.02}$	$0.017_{0.01}$
$\beta_{3,t-1}$	$0.088_{0.07}$	$0.005_{0.03}$	$0.98_{0.04}$

(b) Estimate of Q. Since Q is forced to be symmetric we report only the upper triangular matrix (6 parameters to be estimated).

Q	$\beta_{1,t}$	$\beta_{2,t}$	$\beta_{3,t}$
$\beta_{1.t}$	$0.05_{0.008}$	$-0.06_{0.009}$	$0.045_{0.01}$
$\beta_{2,t}$		$0.116_{0.01}$	$-0.08_{0.02}$
$\beta_{3,t}$			$0.59_{0.06}$

(c) α and β estimates.

α	β
$0.12_{0.14}$	$0.82_{0.06}$

Table 4: Estimation of the DNS-GARCH model. Panel 5.3(a) reports the estimates of A coefficients. Panel 5.3(b) reports the estimate of covariance matrix Q. Panel 5.3(c) reports the estimated garch components α and β.

	$\mathbf{l}(\theta)$	$\lvert\theta\rvert$	AIC	BIC	LR
AR(1)-DNS	11	1150	-2278	-2271	
DNS	1472	26	-2918	-2721	
DNS-GARCH	1777	35	-3484	3456	610

Table 5: Statistic of estimated models. LR statistic is evaluate between the unconstrained model (DNS-GARCH) and the constrained model DNS. The symbol $|\cdot|$ indicates the cardinality, that is, the number of parameter to be estimated.

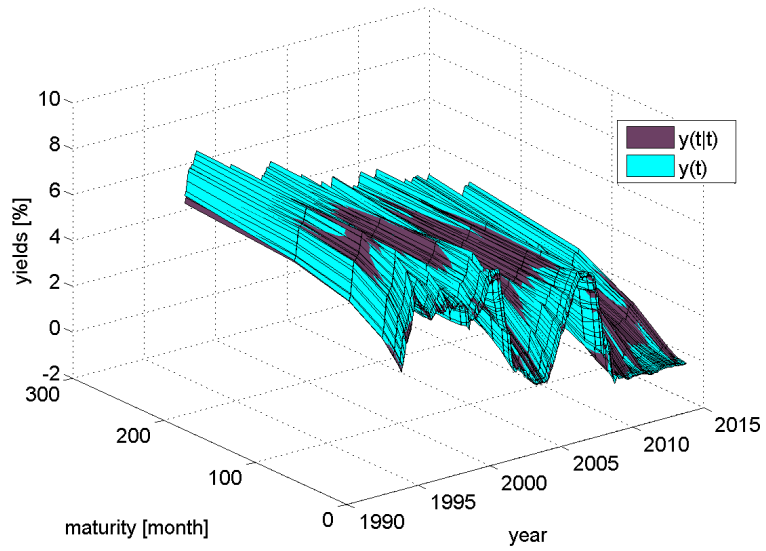


Figure 13: Filtered Yields for all the maturities on validation data.

5.3 MODEL FORECASTING ON VALIDATION DATA

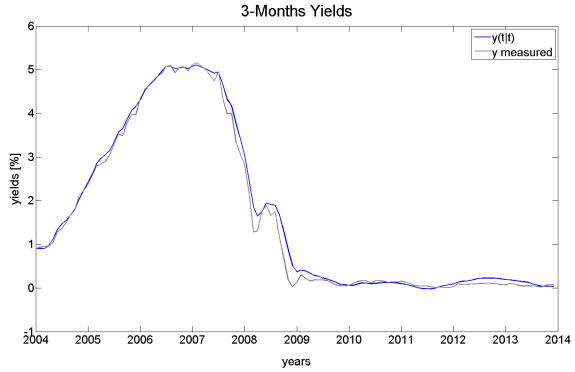
In this section we shall study the forecasting capacity of DNS and DNS-GARCH models. In particular, we will compare their performance in order to establish whether introducing GARCH errors increases the quality of predictions.

The prediction of k-step ahead (in our case one step is equal to one months) can be obtained by the iteration of Kalman equations presented in Section 4.3. The k-months ahead prediction is given by

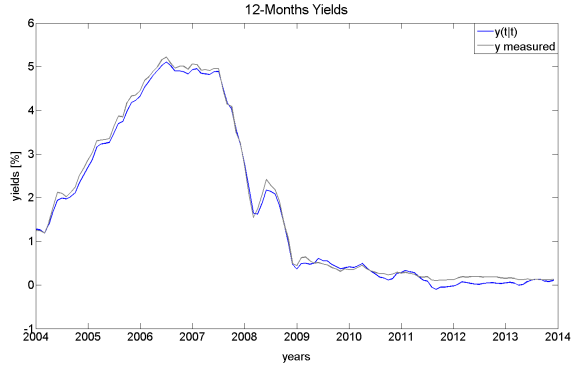
$$x(t+k|t) = A^k x(t|t) \tag{80}$$

$$y(t+k|t) = Cx(t+k|t) \tag{81}$$

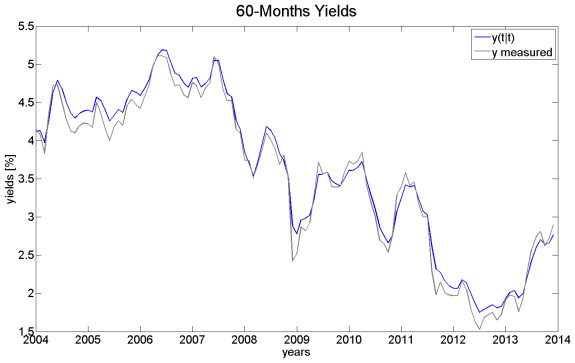
These equation are the same for both the DNS and DNS-GARCH models. This is due to the fact that the expectation of the garch component in the measurement equation is equal to zero. Thus, it doesn't directly influence the prediction steps. Actually, the time-varying volatility is accounted for in the Kalman filter steps and therefore it affects the estimates. Hence, the common shock has an indirect influence on the predictions. For the sake of clarity, in this thesis, we consider only 1, 3, 6, 12 months ahead



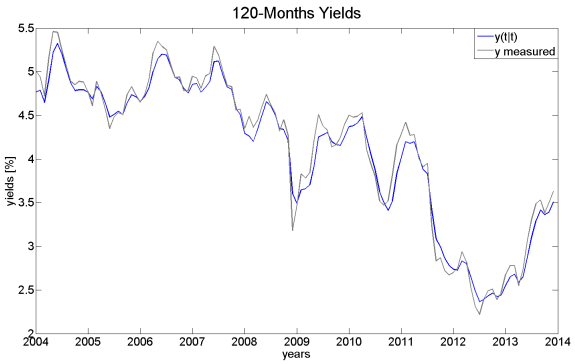
(a) 3-Months yields estimation



(b) 12-Months yields estimation



(c) 60-Months yields estimation



(d) 120-Months yields estimation

Figure 14: Filtered Yields for vary maturities on validation data.

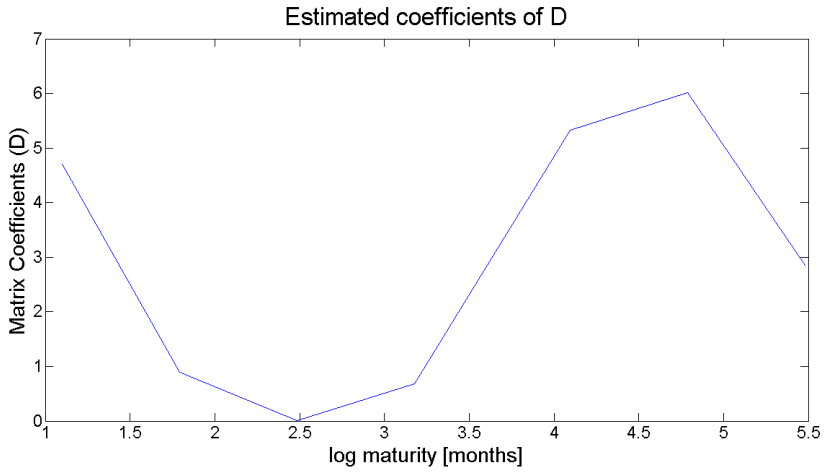


Figure 15: Estimate of D coefficients. The x-axis represent the maturities in log scale.

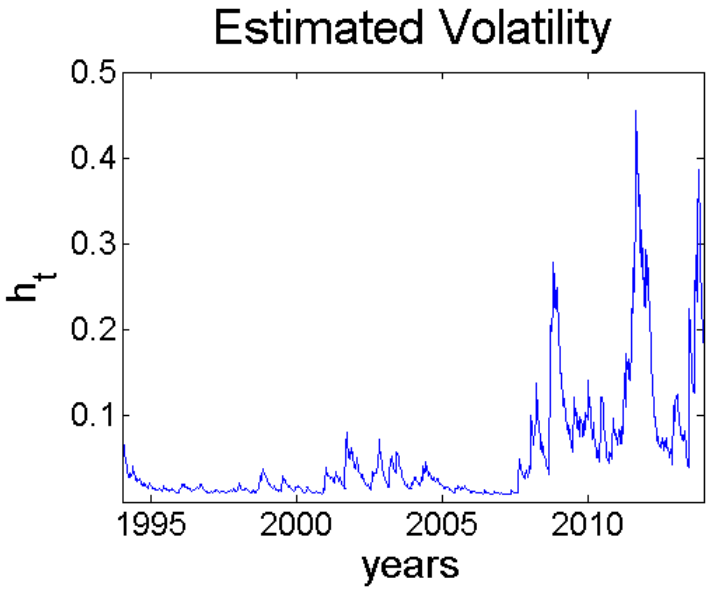


Figure 16: Estimated volatility. The peak in years 2008-2009 is due to the subprime mortgages crisis.

predictions for the maturities. In Table 6 we report the goodness of fit of the predictions and RMSE for all the maturities. For instance, in Figure 17 we show the comparison between DNS and DNS-GARCH performance for 1-month ahead prediction. The fitting statistics confirm us that when prediction horizon

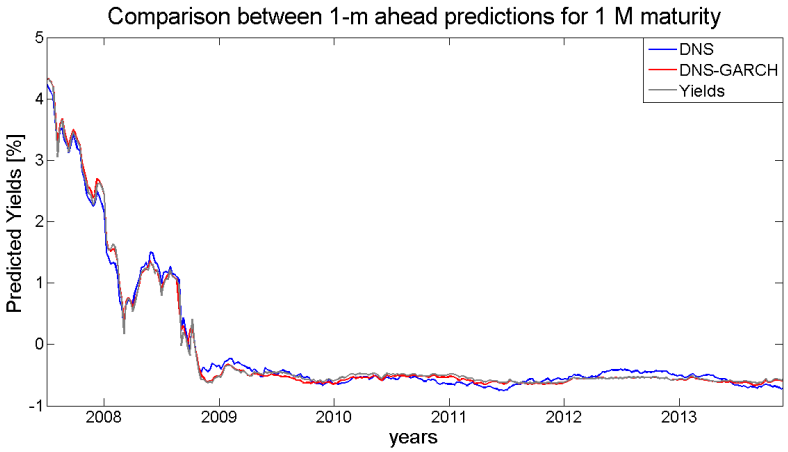


Figure 17: Comparison of 1-month ahead predictions between DNS and DNS-GARCH models. The gray line indicates the true yields at the prediction instant.

increases the goodness of fit decreases. In particular, the worst case is verified when the maturity is equal to 120 months. The best fit is obtained on 1 years-ahead prediction in both the models. The DNS-GARCH model outperforms the standard DNS model: it better predicts yields on high volatility time periods. In low volatility time periods the DNS-GARCH and DNS performances are very similar. This is what we expected because when the volatility is low the GARCH component on error in irrelevant in respect with the white noise.

The presence of GARCH components on error is confirmed by Table 7, in which the ARCH-tests of DNS model squared residuals are reported. For all the maturities, the omoscedasticity hypothesis is rejected.

Finally, we fit a simple model AR(1)-GARCH(1,1)

$$\begin{cases} y_t = a y_{t-1} + \varepsilon_t, & |a| < 1, \varepsilon_t \text{i.i.d} \\ \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 & \alpha + \beta < 1, \omega > 0 \end{cases} \tag{82}$$

FIT [%]

	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7
DNS-GARCH							
1-M	88.74	93.13	93.23	89.51	83.36	84.68	77.30
3-M	79.51	86.80	86.80	81.71	75.38	72.14	65.11
6-M	71.00	78.19	77.76	72.07	65.90	58.62	51.29
12-M	50.58	58.69	57.49	51.05	48.25	40.64	35.50
DNS							
1-M	86.64	88.1	92.1	89.42	82.78	84.08	77.01
3-M	77.1	84.77	84.12	81.70	74.75	71.76	64.98
6-M	69.17	77.10	71.22	65.39	58.55	51.44	51.44
12-M	49.1	58.25	57.79	51.43	48.18	41.14	36.18

RMSE

	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7
DNS-GARCH							
1-M	1.6	1.58	1.51	2.18	3.21	2.49	3.40
3-M	3.39	2.92	2.85	3.65	4.62	4.45	5.17
6-M	5.82	4.55	4.52	5.25	6.13	6.43	7.10
12-M	8.64	7.55	7.61	8.16	8.64	8.83	9.14
DNS							
1-M	2.53	2.22	2.01	1.81	3.32	2.20	2.97
3-M	4.79	3.78	3.33	3.51	5.06	4.38	4.51
6-M	6.84	5.60	4.94	5.12	6.69	6.39	6.41
12-M	9.36	8.10	7.56	7.77	9.09	8.83	8.67

Table 6: Fitting and RMSE on validation data for DNS and DNS-GARCH models. We consider the $1, 3, 6, 12$-months ahead predictions. The goodness of fit is measure as $100\left(1-\sqrt{\frac{\sum_{t=1}^{T-k}(y(t+k)-\hat{y}(t+k|t,\theta))^2}{\sum_{t=1}^{T-k}y(t+k)^2}}\right)$ where k is the horizon in months and $y=(y(1),\ldots,y(T))$ is detrended.

Maturity	**Statistic**	
(Months)	ARCH-test	k_α
3	250	3.8
6	293	3.8
12	269	3.8
24	311	3.8
60	326	3.8
120	330	3.8
240	303	3.8

Table 7: Test statistic of DNS model squared residuals, k_α is the critical value. The significance is $\alpha = 0.5$ and the number of lags is $m = 20$.

ω	α	β
0.000035	0.33	0.65
0.00001	0.22	0.77
0.000006	0.12	0.87
0.000005	0.079	0.91
0.000005	0.057	0.93
0.000004	0.043	0.94
0.000003	0.026	0.96

Table 8: Estimated parameters of AR(1)-GARCH(1,1) model.

for each maturity in order to understand if the ARCH(1) component α and the GARCH(1) β components are similar to those provided by DNS-GARCH estimation. The estimated parameters for the AR-GARCH model are reported in Table 8, which confirm the similarity with the DNS parameter estimations. The parameters in Table 8 are similar to those evaluated for DNS-GARCH model. It is, interesting to note the β factor increases, that is, the *persistence* increases as the maturity increases. In Figure 18 we finally present the volatility (that is the conditional standard deviations) and standardized residuals for vary maturities. As we expect, the volatility on the yields decreases as the maturity time increases.

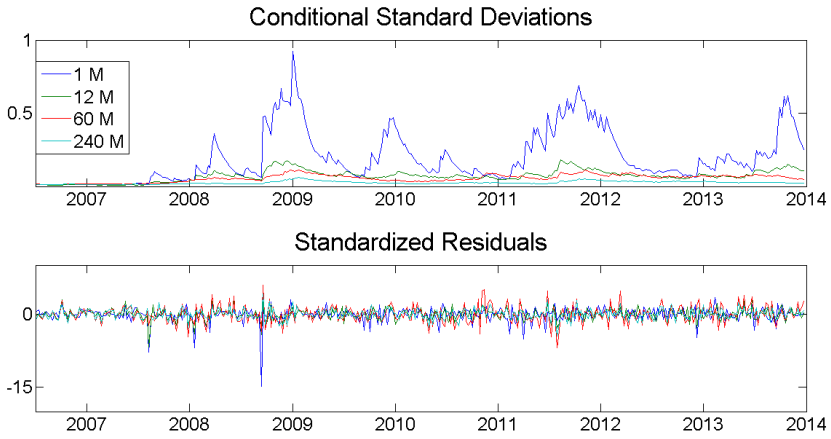


Figure 18: Condition Standard deviations and standardized residual of AR(1)-GARCH(1,1) model for 1, 12, 60, 240 maturities.

The prediction of yields provided by AR(1)-GARCH(1,1) model is (see Equation 93)

$$y_{t+k|t} = a^{k+1} y_{t-1} \tag{83}$$

$$\tag{84}$$

while the variance prediction is

$$Var(y_{t+k}|y_t) = \frac{\omega(1-a^{2(k+1)})}{\{1-(\alpha+\beta)(1-a^2)\}} + \left\{\sigma_t^2 - \frac{\omega}{1-(\alpha+\beta)}\right\} \frac{a^{2(k+1)} - (\alpha+\beta)^{k+1}}{a^2-(\alpha+\beta)} \tag{85}$$

with $a^2 \neq \alpha+\beta$. In Table 9 we present the goodness of fit of AR(1)-GARCH(1,1) model 1-month ahead prediction for 1 month maturity. As we can see, DNS-GARCH model better predicts the dynamic of yields.

FIT [%]

	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7
AR(1)-GARCH(1,1)							
1-M	26.63	35.6	32.7	30.39	27.19	25.4	22.62

RMSE

	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7
AR(1)-GARCH(1,1)							
1-M	16.27	14.78	15.04	14.45	14.03	12.1	11.6

Table 9: Fitting and RMSE on validation data for AR(1)-GARCH(1,1) model for 1-month ahead prediction. The goodness of fit is measure as $100\left(1-\sqrt{\frac{\sum_{t=1}^{T-k}(y(t+k)-\hat{y}(t+k|t,\theta))^2}{\sum_{t=1}^{T-k}y(t+k)^2}}\right)$ where k is the horizon in months and $y = (y(1),\ldots,y(T))$ is detrended.

6

CONCLUSIONS

The main purpose of this thesis was to introduce models suitable to predict the volatility of returns. These models have to explain some recurrent phenomena, called *stylized facts*, exhibited by most of the time series (asset returns, bond returns, etc). The class of models which explains these facts, and discussed in this thesis, are the GARCH models. The importance of this models lays on the prediction of its variance which represents the *volatility*. Firstly, in Chapter 2 we presented the univariate GARCH(p,q) models and we studied its properties such as existence of moments and stationarity. Unfortunately, an univariate framework is not sufficient to deal with the real financial scenario. Engle and Kroner [12] introduced Multivariate models in order to take in account of this drawback. Besides the difficulties in producing theoretical studies, these models suffer a heavy drawback, namely, an enormous number of parameters have to be estimated even with low-order multivariate GARCH models. These difficulties have led us to use state space model framework, yet intensely employed in engineering and physics fields. These models are particularly suitable for filtering and optimal predictions provided by the Kalman filter, minimizing the MSE.

In Section 2.2 we introduced the following state space models: DNS-GARCH models, that is, a linear Gaussian state space model with noise process following a GARCH(1,1) process, and SSGARCH, namely, GARCH models in space space formulation.

Then we provided the stationarity conditions. Concerning the DNS-GARCH model we have shown the stationary condition are the stability of transition matrix and the stationarity of GARCH process describing the error dynamic. In other words, the stationarity is established by the Gaussian state space model and the the GARCH(1,1) model independently. Similarly, SSGARCH models require the transition coefficients to be stable (that is, less than 1 in the univariate case) but unlike the DNS-model the

second condition relative the GARCH dynamic depens on the unconditional asymptotic state variance σ_x^2. The larger σ_x^2 the smaller must be the coefficients α and β.

In Chapter 3 we analyzed stochastic stability of univariate GARCH models in terms of mixing and geometric ergodicity. Mixing regards the description of dependency between past and future, while ergodicity is an instrument useful to show consistency of identification algorithm. As we have seen, even for simple univariate GARCH model mixing property is difficult to study as it requires a strong Markov Chain theory.

In Chapter 4 we studied the identification problem.
More specifically, we presented the QMLE for univariate GARCH models and the EM algorithm for SSGARCH models. There are several alternatives to QML method to estimate GARCH and ARMA-GARCH models. We studied QML method because it is the most commonly used estimation method. Its most attractive characteristic is that the asymptotic properties (consistency and normality) are valid under mild conditions. On the other hand it presents some drawbacks, the most important of which is that the estimator is not explicit but it requires a numerical maximization. Another drawback is the inefficiency.
We identified state space models, such as DNS and DNS-GARCH models with the EM algorithm. These method provides consistent estimators under very mild conditions, but a numeric likelihood maximization is needed. In addiction, we employed constrained minimization in order to impose stationarity on DNS-GARCH model. We used the *interior point* algorithm. Its approach to constrained minimization is to solve a sequence of approximate minimization problems. If the original problem is

$$\min_x f(x),\ h(x) = 0 \text{ and } g(x) \leqslant 0 \tag{86}$$

then for each $\mu > 0$, the approximate problem is given by

$$\begin{aligned} \min_{x,s} f_\mu(x,s) = \min_{x,s} f(x) - \mu \sum_i \ln(s_i), \\ h(x) = 0 \text{ and } g(x) + s = 0. \end{aligned} \tag{87}$$

As μ decreases to zero, the minimum of f_μ should approach the minimum of f. The added logarithmic term is called a *bar-*

rier function. A complete description of this method is given by Byrd et al. [4] and Waltz et al. [23]. To solve the approximate problem the algorithm uses one of two main types of step at each iteration:

- A *direct* step in (x, s). This step attempts to solve the KKT equations

$$\nabla_x L(x, \lambda) = 0, \qquad \lambda_{g_i} g_i(x) = 0 \quad \forall i, \tag{88}$$

  with the constraints

$$\begin{cases} g(x) \leqslant 0, \\ h(x) = 0, \\ \lambda_{g,i} \geqslant 0. \end{cases} \tag{89}$$

  where L is the Lagrangian and $\lambda = \begin{bmatrix} \lambda_g & \lambda_h \end{bmatrix}$ is the vector of multipliers such that

$$L(x, \lambda) = f(x) + \sum \lambda_{g,i} g_i(x) + \sum \lambda_{h,i} h_i(x). \tag{90}$$

- A conjugate gradient step, using trust region.

At each iteration the algorithm decreases a merit function, such as

$$f_\mu(x, s) + \nu \|(h(x), g(x) + s)\|.$$

The parameter ν may increase with iteration number in order to force the solution towards feasibility. If an attempted step does not decrease the merit function, the algorithm rejects the attempted step, and attempts a new step. The main drawback of this algorithm is the expensive computational cost.

As regards the simulations, in Chapter 5 we have shown the data analyzed, that is, the US treasury bonds, present clustering volatility phenomena according to GARCH models. The identification step corroborates these findings, proving that introducing a GARCH component on standard DNS models increases the likelihood function and improves the goodness of fit. Moreover,

using DNS-GARCH models we are able to both predict yields and volatility yields (σ_t^2) by consecutive iterations of

$$\sigma_{t+k|t}^2 = \omega + \alpha\varepsilon_{t+k-1|t}^2 + \beta\sigma_{t+k-1|t}^2.$$

that yields

$$\sigma_{t+k|t-1}^2 = \omega\left(1 + \cdots + \beta^{k-1}\right) + \alpha\sum_{i=0}^{k-1}\beta^{k-i-1}\varepsilon_{t+1|t-1} + \beta^k\sigma_t^2$$

where σ_t^2 is deterministic and $\varepsilon_{t+i|t-1}^2$ for $i = 0, \ldots, k$ are evaluated by the Kalman filter equations. The main difference with the volatility predicted in the other SSGARCH models is the ARCH variable. In DNS-GARCH model the ARCH variable is given by ε_t^2, i.e., the noise component while in SSGARCH it is the yield, i.e., y_t.
Moreover we found the DNS-GARCH proxies very well empirical evidences (Level, Slope and Factor) and can better explain hig volatility time periods than the DNS model.
We also compared DNS-GARCH model with the AR(1)-GARCH models estimated by the QML method and we have shown the DNS-GARCH model outperforms the AR(1)-GARCH(1,1) models for short maturities. In the empirical simulations we did not find particular problems. The main problem we could encounter was the parameter initialization. For DNS-GARCH models, we performed a multistart procedure in order to avoid to be attracted by a local minim. A disadvantage we encountered is the computational time, especially in the DNS-GARCH model estimation.

6.0.1 *Further research*

Finally let us represent the GARCH(1,1) model in the state space framework. It is given by

$$\begin{aligned} y_t &= \sigma_t\varepsilon_t \\ \sigma_t^2 &= \omega + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2 \end{aligned}$$

with ε_t i.i.d..

Let define $\sigma_t^2 = x_1(t)$ and $\varepsilon_t = x_2(t)$. Then the GARCH(1,1) can be rewritten as

$$\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \end{bmatrix} = \begin{bmatrix} \omega + \alpha x_1(t) x_2(t)^2 + \beta x_1(t) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \varepsilon(t+1)$$
$$y(t) = \sqrt{x_1(t)} x_2(t)$$

This model is non-linear and it needs of the Extended Kalman filter (EKF) framework. The estimation of the model is provided by the EM algorithm with the only difference that the model used by the EM algorithm is the linearization of the non-linear system. Using a Kalman filter framework we are able to estimate efficiently the GARCH(1,1) model.
It is interesting for further works, studying multivariate GARCH written in state space model framework. Although these models would be non-linear they would allow to avoid the complications encountered with Multivariate GARCH models. Successively, an important step would be studying stochastic stability of these models.

A

SOME DEFINITIONS AND THEOREMS

A.1 ERGODICITY

Definition A.1 (Stationarity). A process $\{y_t\}$ is said to be strictly *stationary* if all its finite order distributions are invariant to temporal shifting, that is, for each n,

$$F_n(x_1, \ldots, x_n, t_1 + \Delta, \ldots, t_n + \Delta) = F_n(x_1, \ldots, t_n)$$

identically in $x_1, \ldots, x_n, t_1, \ldots, t_n$, for all $\Delta \in \mathbb{Z}$.
In particular, it is said to be second-order stationary if

$$F_2(x_1, x_2, t_1 + \Delta, t_2 + \Delta) = F_2(x_1, x_2, t_1, t_2).$$

This means that the second-order mutual distribution $F_2(x_1, x_2, t_1, t_2)$ of variables $y(t_1), y(t_2)$, depends only on temporal interval $\tau = t_1 - t_2$. Therefore, the process mean $Ey(t)$ is is constant over time and the covariance matrix

$$\Sigma(t_1, t_2) = E\left[(\mathbf{y}(t_1) - E\mathbf{y}(t_1))\right]\left[(\mathbf{y}(t_2) - E\mathbf{y}(t_2))\right]'$$

depends only on temporal distance τ.

A stationary process is said to be *ergodic* if it satisfies the strong law of large numbers.

Definition A.2 (Ergodicity). A strictly stationary process $\{y_t\}$, $t \in \mathbb{Z}$, is said to be ergodic if and only if, for any Borel set B and any integer k,

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{1}_B(y_t, y_{t+1}, \ldots, y_{t+k}) \to \mathbb{P}\{(y_t, y_{t+1}, \ldots, y_{t+k}) \in B\}$$

with probability 1.
In particular it is said to be second-order ergodic if and only if the sample mean

$$\bar{y}_T = \frac{1}{T}\sum_{t=1}^{T} y(t) \to Ey(t)$$

and the sample variance

$$S_T(\tau) = \frac{1}{T}\sum_{t=1}^{T} [y(t+\tau) - \bar{y}_T][y(t) - \bar{y}_T] \to \Sigma(\tau)$$

with probability 1.

Theorem A.1. If $\{y_t\}$, $t \in \mathbb{Z}$, is an ergodic process and if $\{y_t\}$ is defined by

$$y_t = f(\dots, y_{t-1}, y_t, y_t, y_{t+1}, \dots)$$

where f is a measurable function from $\mathbb{R}^\infty \to R$, then y_t is also ergodic.

In particular, if $\{y_t\}$ is governed by the law

$$y_t = a y_{t-1} + e_t, \qquad |a| < 1, \qquad e_t \quad iid(0, \sigma^2), \tag{91}$$

then the theorem shows that $\{y_t\}$, $\{(y_{t-1}e_t)\}$ and y_{t-1}^2 are also ergodic.

Theorem A.2 (Ergodic theorem). If $\{y_t\}$ is ergodic, if f is measurable and if $f(y) \in L^1(y)$ then

$$\frac{1}{T}\sum_{t=1}^{T} f(\dots, y_{t-1}, y_t, y_t, y_{t+1}, \dots) \to Ef(\dots, y_{t-1}, y_t, y_t, y_{t+1}, \dots)$$

almost surely.

A.2 MARTINGALE DIFFERENCE

Let $\{\mathcal{F}_t; t \in \mathbb{Z}\}$be asuccession of σ-fields such that $\mathcal{F}_{t-1} \subset \mathcal{F}_t$. The stochastic process $\{y(t), t \in \mathbb{Z}\}$ is a *martingale difference* (d-martingale) in respect with $\{\mathcal{F}_t\}$ if and only if

1. y_t is $\mathcal{F}_t$-measurable;
2. $E|y_t| < \infty$;
3. $E(y_{t+1}|\mathcal{F}_t) = 0$.

The first condition and second conditions yield

$$E\{y_t|\mathcal{F}_s\} = 0 \qquad \forall\, s < t.$$

Obviously a d-martingale is always a zero mean process.
The following theorem applies to GARCH processes, which are ergodic martingale differences.

Theorem A.3 (Billingsley, 1961). If $(y_t, \mathcal{F}_t)$ is an ergodic sequence of squared integrable martingale differences such that $\sigma_y^2 = Var(y_t) \neq 0$, then

$$\frac{1}{T}\sum_{t=1}^{T} y_t \to \mathcal{N}\left(0, \sigma_y^2\right) \qquad \text{in law.}$$

A.3 MIXING

Mixing assumptions, introduced by Rosenblatt [22], are used to convey different ideas of asymptotic independence between remote past and remote future of a process. We present here α-mixing and β-mixing coefficients.

A.3.1 *α-Mixing and β-Mixing*

The strong mixing (α-mixing) coefficient between two σ-fields $\mathcal{A}$ and $\mathcal{B}$ is defined by

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A\in\mathcal{A}, B\in\mathcal{B}} |\mathcal{P}(A \cap B) - \mathcal{P}(A)\mathcal{P}(B)|.$$

It is clear that:

- if $\mathcal{A}$ are independent then $\alpha(\mathcal{A}, \mathcal{B}) = 0$;
- $0 \leqslant \alpha(\mathcal{A}, \mathcal{B}) \leqslant \frac{1}{4}$;

- $\alpha(\mathcal{A}, \mathcal{A}) > 0$ provided that $\mathcal{A}$ is nontrivial;
- $\alpha(\mathcal{A}', \mathcal{B}') \geqslant \alpha(\mathcal{A}, \mathcal{B})$ provided that $\mathcal{A}' \subset \mathcal{A}$ and $\mathcal{B}' \subset \mathcal{B}$.

For a process $\{y_t\}$ we have

$$\alpha_y(k) = \sup_t \alpha\{\sigma(y_s, s \leqslant t), \sigma(y_s, s \geqslant t+k)\}.$$

If $\{y_t\}$ is stationary, we can omit the term $\sup_t$ and thus we have

$$\begin{aligned}\alpha_y(k) &= \sup_{A\in\mathcal{A}, B\in\mathcal{B}} |\mathcal{P}(A \cap B) - \mathcal{P}(A)\mathcal{P}(B)| \\ &= \sup_{f,g} \left|\mathrm{Cov}(f(\ldots, y_{-1}, y_0), g(y_k, y_{k+1}, \ldots))\right|\end{aligned}$$

where the supremum is taken on set of measurable functions f and g such that $|f| < 1$, $|g| < 1$. The process $\{y_t\}$ is said to be *α-mixing*, if $\alpha_y(k) \to 0$ as $k \to \infty$.s If $\alpha_y(k)$ decreases to zero at an exponential rate, then $\{y_t\}$ is said to be *geometrically α-mixing*.

The β-coefficients of a stationary process $\{y_t\}$ are defined by

$$\begin{aligned}\beta_t(k) &= \mathrm{E} \sup_{B\in\sigma(y_s, s\geqslant k)} \left|\mathcal{P}(B|\sigma(y_s, s \leqslant 0)) - \mathcal{P}(B)\right| \\ &= \frac{1}{2} \sup \sum_{i=1}^{P} \sum_{j=1}^{Q} \left|\mathcal{P}(A_i \cap B_j) - \mathcal{P}(A_i)\mathcal{P}(B_j)\right|\end{aligned}$$

where the sup is taken among all the pairs of partitions $\{A_1, \ldots, A_P\}$ and $\{B_1, \ldots, B_Q\}$ such that $A_i \in \sigma(y_s, s \leqslant 0)\}$ for all i and $B_j \in \sigma(X_s, s \geqslant k)$ for all j. The process $\{y_t\}$ is said to be β-mixing if $\beta_y(k) \to 0$ as $k \to \infty$. We have

$$\alpha_y(k) \leqslant \beta_y(k),$$

meaning that β-mixing implies α-mixing.

A.4 THEORETICAL PREDICTION OF GARCH PROCESSES

Let consider the GARCH(p,q) stationary process defined in Equation 7. We have already shown that the optimal prediction in L^2

sense of y_t given its past is 0. The GARCH processes are interesting because we can predict the squared process. Defined the past of process as $\mathcal{F}_{t-1}$ and given the prediction horizon k we obtain recursively

$$\begin{aligned} E(y_{t+k}^2|\mathcal{F}_{t-1}) &= E(\sigma_{t+k}^2|\mathcal{F}_{t-1}) \\ &= \omega + \sum_{i=1}^{q} \alpha_i E(y_{t+k-i}^2|\mathcal{F}_{t-1}) + \sum_{j=1}^{p} \beta_j E(\sigma_{t+k-i}^2|\mathcal{F}_{t-1}), \end{aligned} \tag{92}$$

with

$$\begin{cases} E(y_{t+k-i}^2|\mathcal{F}_{t-1}) = E(\sigma_{t+k-i}^2|\mathcal{F}_{t-1}), & i \leqslant h, \\ E(y_{t+k-i}^2|\mathcal{F}_{t-1}) = y_{t+k-i}^2, & i > h, \\ E(\sigma_{t+k-i}^2|\mathcal{F}_{t-1}) = \sigma_{t+k-i}^2, & i \geqslant h. \end{cases}$$

These predictions are also optimal linear predictions of $\{y_t^2\}$.

For $AR(1) - GARCH(1,1)$ process defined by

$$\begin{cases} y_t = a y_{t-1} + \varepsilon_t \\ \varepsilon_t = \sigma_t \eta_t \\ \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \end{cases} \tag{93}$$

where $\omega > 0$, $\alpha \geqslant 0$, $\beta \geqslant 0$, $\alpha + \beta < 1$, $|a| < 1$. For $k \geqslant 0$, we have

$$y_{t+k} = \varepsilon_{t+k} + a\varepsilon_{t+k-1} + \cdots + a^k \varepsilon_t + a^{k+1} y_{t-1}$$

and hence

$$E(y_{t+k}|y_s,\ \mathcal{F}_{t-1}) = a^{k+1} y_{t-1}.$$

The conditional variance results,

$$\begin{aligned} Var(y_{t+h}|\ \mathcal{F}_{t-1}) &= Var\left(\sum_{i=0}^{k} a^{k-i} \varepsilon_{t+i}|\ \mathcal{F}_{t-1}\right) \\ &= \sum_{i=0}^{k} a^{2(k-i)} Var(y_{t+i}|\ \mathcal{F}_{t-1}). \end{aligned}$$

Given

$$\begin{aligned}\mathrm{Var}(y_{t+i}|\,\mathcal{F}_{t-1}) &= \mathrm{E}(\sigma^2_{t+i}|\,\mathcal{F}_{t-1})\\ &= \omega + (\alpha+\beta)\mathrm{E}(\sigma^2_{t+i-1}|\,\mathcal{F}_{t-1})\\ &= \omega[1+\cdots(\alpha+\beta)^{i-1}] + (\alpha+\beta)^i\sigma^2_t,\end{aligned}$$

we have

$$\mathrm{Var}(y_{t+i}|\,\mathcal{F}_{t-1}) = \omega\frac{1-(\alpha+\beta)^i}{1-(\alpha+\beta)} + (\alpha+\beta)^i\sigma^2_t, \qquad \text{for all } i \geqslant 0.$$

Finally,

$$\begin{aligned}&\mathrm{Var}(y_{t+h}|\,\mathcal{F}_{t-1})\\ &= \sum_{i=0}^{k} a^{2(k-i)}\frac{\omega}{1-(\alpha+\beta)}\\ &\qquad + \left\{\sigma^2_t - \frac{\omega}{1-(\alpha+\beta)}\right\}\sum_{i=0}^{k}(\alpha+\beta)^i a^{2(h-i)}\\ &= \frac{\omega(1-a^{2(h+1)})}{[1-(\alpha+\beta)](1-a^2)}\\ &\qquad + \left\{\sigma^2_t - \frac{\omega}{1-(\alpha+\beta)}\right\}\frac{a^{2(h+1)}-(\alpha+\beta)^{h+1})}{a^2-(\alpha+\beta)}\end{aligned}$$

with $a^2 \neq \alpha+\beta$.

B

MATLAB CODE

In the following we report some code functions used in this thesis.
The following file is `DNS.m`. It uses the function `kalmanRecursionDNS` in order to provide an estimate on DNS model.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%MAIN%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clear all
close all
load data/OriginalData.mat
maturity = [3 6 12 24 60 120 240]';
y=yieldsTOT(1:600,:); %240x7q
%% DNS-AR-Qdiag-Req
% NoP = 3 + 3 + 3 + 1 + 1 = 11
A_0 = [0.9 0 0;0 0.9 0;0 0 0.9];
mu = [3 -3 -3]';
Q = [0.5 0 0 ;0 0.5 0;0 0 0.5];
r_0 = 0.5;
R = diag(repmat(r_0,7,1));
lambda_0= 0.05;
vp0 = [reshape(diag(A_0),[3,1]); mu; reshape(diag(Q),[3,1]); r_0; lambda_0];

%% optimization
options=optimset('Display','iter','TolFun',10^(-25),'TolX',10^-...
8,'MaxFunEvals',10000,'MaxIter',32);
% Z = zeros(11,11);
% Z(1,1) = .99;
% Z(2,2) = .99;
% Z(3,3) = .99;
% b = [1 1 1 inf inf inf inf inf inf inf inf inf]';
lambda_0 = 0.05;
vp0 = [reshape(diag(A_0),[3,1]); mu; reshape(diag(Q),[3,1]); r_0; lambda_0];
ub = [.99 .99 .99 inf inf inf inf inf inf inf inf inf]';
lb = [-inf -inf -inf -inf -inf -inf -inf -inf -inf -inf 0.01];
```

```
[vpEstAR1,fvalAR1] =
fmincon(@(vp) ...
-kalmanRecursionDNS(vp,y,maturity,'DNS-AR-Qdiag-Req'),...
 vp0,[],[],[],[],lb,ub,[],options);

A = diag(vpEstAR1(1:3));
mu = vpEstAR1(4:6);
Q = diag(vpEstAR1(7:9))^2;
% lambda = vpEstAR1(11);
R  = diag(repmat(vpEstAR1(10),length(maturity),1))^2;
% % h = hessian(@(vp) kalmanRecursionDNS(vp,y,maturity,...
'DNS-AR-Qdiag-Req'), vpEst);
% % I = -h; %fisher
I = fisherMatrix(y,maturity,'DNS-AR-Qdiag-Req',vpEstAR1);
vc = inv(I);
stderr = sqrt(diag(vc));
%% DNS-AR-Qdiag-Rdiag
% NoP = 9 + 3 + 6 + 7 + 1 = 26
A_0 = [0.9 0 0.001;-0.01 0.92 0.05;0.1 0.05 0.85];
mu_0 = [4 -4 -4]';
Q_0 = [0.05 0 0 ;0 0.124 0;0 0 0.5];
R_0 = 0.1*eye(7);
lambda_0= 0.05;
vpQ0 = [0.05 0.05 0.02 0.1 -0.05 0.59]';
vp0 = [reshape(A_0,[9,1]); mu_0; vpQ0;
 repmat(0.1,length(maturity),1); lambda_0];

%% optimization
options=optimset('Display','iter','TolFun',10^(-10),...
'TolX',10^-6,'MaxFunEvals',10000,'MaxIter',...
150,'Algorithm','interior-point' );
ub = inf*ones(26,1);
ub(1) = 0.99;
ub(5) = 0.99;
ub(9) = 0.99;
lb = -20*ones(26,1);
lb(26) = 0.01;
[vpEst,fval] = fmincon(@(vp) -...
kalmanRecursionDNS(vp,y,maturity,' '),...
         vp0,[],[],[],[],lb,ub,[],options);
```

```
% [vpEst,fval] = fmincon(@(vp) -
kalmanRecursionDNS(vp,y,maturity,''),vp0,options);

A = reshape(vpEst(1:9),[3,3]);
mu = vpEst(10:12);
a = vpEst(13:18);
q = triu(ones(3),0);
q(q==1) = a;
q=q'; %lower triang
Q=q*q'; %3x3 - 6 params
% Q = [vpEst(13:15)'; 0 vpEst(16:17)'; 0 0 vpEst(18)];
R = diag(vpEst(19:25).^2);
lambda = vpEst(26);
% C=[ones(NoM,1),((1-exp(-lambda*maturity))./
(lambda*maturity)),((1-exp(-...
% lambda*maturity))./(lambda*maturity))-...
%  exp(-lambda*maturity)];
% % h = HessMp(@(vp) kalmanRecursionDNS(vp,y,maturity,''),
vpEst);
% % h = hessian(@(vp) kalmanRecursionDNS(vp,y,maturity,''),
 vpEst);
% % I = fisherMatrix(y,maturity,' ',vpEst); %fisher
% % vc = inv(I);
% % stderr = sqrt(diag(vc));
```

The following file is `DNS-GARCH.m`. It uses `kalmanRecursionDNSGARCH` function in order ro provide an estimate of DNS-GARCH model.

```
clear all
close all
load data/OriginalDataW.mat
maturity = [3 6 12 24 60 120 240]';
y=yieldsTOT(1:700,:); %240x7
% y=yieldsTOT;
%% DNS-GARCH
% NoP = 9 + 3 + 6 + 7 + 1 + 7 + 2 = 35
A_0 = [0.95 0.001 0.003;-0.05 0.92 0.05;0.1 0.05 0.85];
mu_0 = [4.5 -4 -4 0]';
omega = 0.0001;
alpha0= 0.7;
beta0 = 0.2;
Q_0 = [0.055 -0.059 0.051 0;0 0.124 0 0;0 0 0.5 0;...
0 0 0 omega/(1-(alpha0 + beta0))];
R_0 = 0.1*eye(7);
lambda_0= 0.05;
% C=[ones(NoM,1),((1-exp(-lambda*maturity))./...
(lambda*maturity)),((1-exp(-...
% lambda*maturity))./(lambda*maturity))-...
%  exp(-lambda*maturity)];
vpQ0 = [0.05 -0.05 0.05 0.1 -0.05 0.6]';
Lambda_0 = [1 0.5 0.8 1.2 1.1 1.1 0.05]';
vp0 = [reshape(A_0,[9,1]); mu_0(1:3); vpQ0;...
repmat(0.1,length(maturity),1);...
            lambda_0; Lambda_0; [alpha0 beta0]'];

%% optimization
options=optimset('Display','iter','TolFun',10^(-8),...
'TolX',10^-...
6,'MaxFunEvals',10000,'MaxIter',150,'Algorithm'...
,'interior-point','AlwaysHonorConstraints','bounds');
ub = inf*ones(35,1);
ub(1) = 0.99;
ub(5) = 0.99;
ub(9) = 0.99;
ub(34) = 0.9;
% ub(27:33) = 10;
```

```
lb = -inf*ones(35,1);
lb(26) = 0.03;
lb(34) = 0.5;
lb(35) = 0.05;
lb(27:33) = 0.001;
Z = zeros(35,35);
Z(1,34:35) = 1;
b = ones(35,1);
% b(1) = 1;
[vpEst,fval] = fmincon(@(vp) -...
kalmanRecursionDNSGARCH(vp,y,maturity),...
                vp0,Z,b,[],[],lb,ub,[],options);
NoM = 7;
A = reshape(vpEst(1:9),[3,3]);
mu = vpEst(10:12);
a = vpEst(13:18);
q = triu(ones(3),0);
q(q==1) = a;
q=q'; %lower triang
Q=q*q'; %3x3 - 6 params
% Q = [vpEst(13:15)'; 0 vpEst(16:17)'; 0 0 vpEst(18)];
R = diag(vpEst(19:25).^2);
lambda = vpEst(26);
Lambda = vpEst(27:33);
alpha = vpEst(34);
beta = vpEst(35);
C=[ones(NoM,1),((1-exp(-lambda*maturity))./...
(lambda*maturity)),((1-exp(-...
lambda*maturity))./(lambda*maturity))-...
 exp(-lambda*maturity) Lambda];
% h = HessMp(@(vp) kalmanRecursionDNS(vp,y,maturity,''), vpEst);
% h = hessian(@(vp) kalmanRecursionDNS(vp,y,maturity,''), vpEst);
% I = fisherMatrix(y,maturity,' ',vpEst); %fisher
% vc = inv(I);
% stderr = sqrt(diag(vc));
```

The following code is the function `kalmanRecursionDNS`

```
function L = kalmanRecursionDNS(vp,y,maturity,type)
NoM = 7;
if strcmp(type,'DNS-AR-Qdiag-Req')
    A = vec2mat(diag(vp(1:3)),3);
    mu = vp(4:6);
    q = diag(vp(7:9));
    Q = q*q';
    R = eye(NoM)*vp(10)^2;
    lambda = vp(11);
    NoP = 11;

else
    A = reshape(vp(1:9),[3,3]);
    mu = vp(10:12);
    a = vp(13:18);
    q = triu(ones(3),0);
    q(q==1) = a;
    q=q'; %lower triang
    Q=q*q'; %3x3 - 6 params
    R = diag(vp(19:25))*diag(vp(19:25))';
    lambda = vp(26);
    NoP = 26;
end
C=[ones(NoM,1),((1-exp(-lambda*maturity))./...
(lambda*maturity)),((1-exp(-...
lambda*maturity))./(lambda*maturity))-...
 exp(-lambda*maturity)];
%kalmanRec
ss = 3; %scalar = 3
[T ~] = size(y); T=240;
y=y';
x = zeros(ss, T); %3x240 matrix of L,S,C
V = zeros(ss, ss, T); %3x3x240 matrix of V(Y);
loglik = 0;
L = 0;

%derivate matrici di stato

for t=1:T
```

```
    init_x=mu;
%     init_V=inv(eye(ss)-A*A')*Q;
%     init_V=A*V(:,:,t)*A'+Q;
        init_V =  dlyap(A,Q);
%init_V=eye(length(Q));
    if t==1
        prevx = init_x;
        prevV = init_V;
    else
        prevx = x(:,t-1); %E[ X | y_t-1 ] prior mean
        prevV = V(:,:,t-1);%Cov[ X | y_t-1] prior covariance
%         dxprev = dx(:,t);
%         dVprev = dV(:,:,t);
    end
    %prediction step
    xpred(:,t) = A*prevx + (eye(ss)-A)*mu; %(I-A)u+AX - E[X_t | X_t-1]
    Vpred(:,:,t) = A*prevV*A' + Q; %A*V*A' + Q - E[V_t | V_t-1]
    %update step
    % errors from measurement eq. (innovation)
    e(:,t) = y(:,t) - C*xpred(:,t);
    n=7; %scalar = 7
    %7x7 B*SIGMA*B' or V(e|x_t...) on measurement eq
    S = C*Vpred(:,:,t)*C' + R+0.001*eye(7);
    Sinv= inv(S); %7x7
    K = Vpred(:,:,t)*C'*Sinv; % Kalman gain matrix
    xnew(:,t) = xpred(:,t) + K*e(:,t); %E[ X | y_t ] updated mean
    %Cov[ X | y_t] updated covariance
    Vnew(:,:,t) = (eye(ss) - K*C)*Vpred(:,:,t);

    x(:,t)=xnew(:,t); %update timestep
    V(:,:,t)=Vnew(:,:,t); %update timestep
    loglik(:,t)=((-n/2)*log(pi)-0.5*log(det(S))-0.5*(e(:,t)'*Sinv*e(:,t)));
    L= L + loglik(:,t);

end
end
```

The following code is the function kalmanRecursionDNS

```
function L = kalmanRecursionDNSGARCH(vp,y,maturity)
NoM = 7;
    A1 = reshape(vp(1:9),[3,3]);
    A = [A1 zeros(3,1);zeros(1,3) 0];
    mu = vp(10:12);
    a = vp(13:18);
    q = triu(ones(3),0);
    q(q==1) = a;
    q=q'; %lower triang
    Q1=q*q'; %3x3 - 6 params
    omega  =0.0001;
    alpha = vp(34);
    beta = vp(35);
   %Q = [Q1 zeros(3,1);zeros(1,3) omega/(1 - (alpha + beta))];
R = diag(vp(19:25))*diag(vp(19:25))';
    lambda = vp(26);
    Lambda = vp(27:33);
C=[ones(NoM,1),((1-exp(-lambda*maturity))./...
(lambda*maturity)),((1-exp(-...
lambda*maturity))./(lambda*maturity))-...
 exp(-lambda*maturity) Lambda];
%kalmanRec
ss = 4; %scalar = 4
[T os] = size(y); %OS=7, T=240;
y=y';
x = zeros(ss, T); %4x240 matrix of L,S,C
V = zeros(ss, ss, T); %4x4x240 matrix of V(Y);
loglik = 0;
L = 0;
h1 = omega/(1-alpha - beta);
for t=1:T
    init_x=[mu;0];
    init_V1 = dlyap(A1,Q1);
    init_Q = [Q1 zeros(3,1);zeros(1,3) h1];
    init_V =  [init_V1 zeros(3,1);
                zeros(1,3) h1];
    if t==1
        prevx = init_x;
        prevV = init_V;
```

```
        Q(:,:,t) = init_Q;
        h(t) = h1;
    else
        prevx = x(:,t-1); %E[ X | y_t-1 ] prior mean
        prevV = V(:,:,t-1);%Cov[ X | y_t-1] prior covariance
%         dxprev = dx(:,t);
%         dVprev = dV(:,:,t);
    end
    %prediction step
     if t > 1
        eps = prevV;
        h(t) = omega + alpha*(eps(4,4)+prevx(4)^2) + beta*h(t-1);
%   h(t) = omega + alpha*(eps(4,4) + prev) + beta*h(t-1);
     end
    Q(:,:,t) = [Q1 zeros(3,1);
                zeros(1,3) h(t)];
                %(I-A)u+AX - E[X_t | X_t-1]
    xpred(:,t) = A*prevx + (eye(ss)-A)*[mu;0];
    %A*V*A' + Q - E[V_t | V_t-1]
    Vpred(:,:,t) = A*prevV*A' + Q(:,:,t);
    %update step
    % errors from measurement eq. (innovation)
    e(:,t) = y(:,t) - C*xpred(:,t);
    n=7; %scalar = 7
    %7x7 B*SIGMA*B' or V(e|x_t...) on measurement eq
    S = C*Vpred(:,:,t)*C' + R + 0.0001*eye(7);
    Sinv = inv(S); %7x7
    K = Vpred(:,:,t)*C'*Sinv; % Kalman gain matrix
    %E[ X | y_t ] updated mean
    xnew(:,t) = xpred(:,t) + K*e(:,t);
    %Cov[ X | y_t] updated covariance
    Vnew(:,:,t) = (eye(ss) - K*C)*Vpred(:,:,t);
    x(:,t)=xnew(:,t); %update timestep
    V(:,:,t)=Vnew(:,:,t); %update timestep
    loglik(:,t)=((-n/2)*log(pi)-0.5*log(det(S))-0.5*(e(:,t)'*Sinv*e(:,t)));
    L= L + loglik(:,t);
end
save('volatility','h');
end
```

BIBLIOGRAPHY

[1] T. Bollerslev, R.F. Y. Chou, and K. Kroner. Arch modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics*, pages 5–59, 1986.

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

[3] K.P. Burnham and D.R. Anderson. *Model Selection and Inference: A Practical Information-theoretic Approach*. Intelligence, SS. of Lncs; 1501. Springer, 1998. ISBN 9780387985046. URL http://books.google.it/books?id=0HHwAAAAMAAJ.

[4] Richard H. Byrd, Mary E. Hribar, and Jorge Nocedal. An interior point algorithm for large-scale nonlinear programming. *SIAM J. on Optimization*, 9(4):877–900, April 1999. ISSN 1052-6234. doi: 10.1137/S1052623497325107. URL http://dx.doi.org/10.1137/S1052623497325107.

[5] Han-fu Chen and P. R. Kumar. On kalman filtering for conditionally gaussian systems with random matrices, 1989.

[6] J. Christensen, F. Diebold, and G. Rudebusch. The affine arbitrage-free class of nelson-siegel term structure models. *Journal of Econometrics*, pages 4–20, 2011.

[7] J.A. Davydov. Convergence of distributions generated by stationary stochastic processes. *Theory Probab. Appl.*, 13: 691–696, 1968. ISSN 0040-585X; 1095-7219/e. doi: 10.1137/1113086.

[8] F. X. Diebold and C. Li. Forecasting the term structure of government bond yields. *Journal of Econometrics*, pages 337–364, 2006.

[9] F. X. Diebold, S. Rudebusch, and S. Aruoba. The macroeconomy and the yield curve. *Journal of Econometrics*, pages 309–338, 2006.

[10] Gregory R. Duffee. Term premia and interest rate forecasts in affine models. *JOURNAL OF FINANCE*, 57(1):405–443, 2002.

[11] R. F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, pages 987–1007, 1982.

[12] R. F. Engle and K. F. Kroner. Multivariate simultaneous generalized arch. *Econometric Theory*, pages 122–150, 1995.

[13] Christian Francq and Jean-Michel Zakoïan. Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli*, 10(4):605–637, 08 2004. doi: 10.3150/bj/1093265632. URL http://dx.doi.org/10.3150/bj/1093265632.

[14] A Harvey, E Ruiz, and E Sentana. Unobserved component time series models with arch disturbances. *Journal of Econometrics*, 1992.

[15] A. C. Harvey. *Time Series Models*. Philip Allan and Humanities Press, 1981.

[16] Eric Hillebrand. A mean-reversion theory of stock-market crashes., 2003.

[17] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[18] S. Koopman, M. Mallee, and M. van der Wel. Analyzing the term structure of interest rates using dynamic nelson-siegel model with time-varying parameters. *Journal of Business and Economic Statistics*, pages 329–343, 2010.

[19] Sean Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, New York, NY, USA, 2nd edition, 2009. ISBN 0521731828, 9780521731829.

[20] C. Nelson and A. Siegel. Parsimonious modelling of yield curves. *Journal of Business*, pages 473–489, 1987.

[21] G. Picci. *Filtraggio statistico (Wiener, Levinson, Kalman) e applicazioni*. Progetto Libreria, 2007. ISBN 9788887331882. URL http://books.google.it/books?id=Ivb8GAAACAAJ.

[22] Murray Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, 09 1956. doi: 10.1214/aoms/1177728190. URL http://dx.doi.org/10.1214/aoms/1177728190.

[23] R. A. Waltz, J. L. Morales, J. Nocedal, and D. Orban. An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Math. Program.*, 107(3):391–408, July 2006. ISSN 0025-5610. doi: 10.1007/s10107-004-0560-5. URL http://dx.doi.org/10.1007/s10107-004-0560-5.

[24] C. F. Jeff Wu. On the convergence properties of the em algorithm. *The Annals of Statistics*, 11(1):95–103, 03 1983. doi: 10.1214/aos/1176346060. URL http://dx.doi.org/10.1214/aos/1176346060.