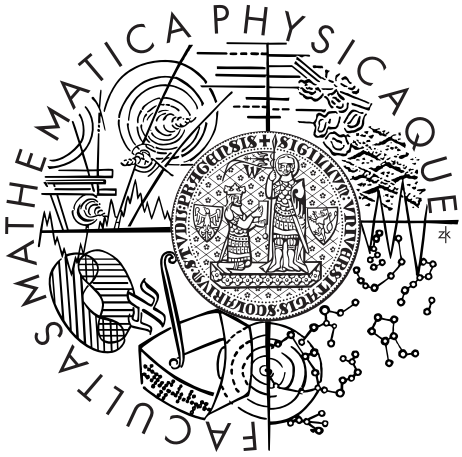
Charles University in Prague

Faculty of Mathematics and Physics

DOCTORAL THESIS



Radek Hendrych

Selected problems of financial time series modelling

Department of Probability and Mathematical Statistics

Supervisor of the doctoral thesis: Prof. RNDr. Tomáš Cipra, DrSc.

Study programme: Mathematics

Specialization: Econometrics and Operational Research

Prague 2015

I would like to express my sincere gratitude to my supervisor Prof. RNDr. Tomáš Cipra, DrSc. for his extraordinarily kind, patient, tireless guidance, and support throughout my doctoral study.

I would like to thank my parents, my girlfriend, and my brother for their unconditional support and trust in me.

Acknowledgements

This work was supported by the grant GA P402/12/G097.
This work was supported by the grant SVV 261315/2011.
This work was supported by the grant SVV 265315/2012.
This work was supported by the grant SVV 267315/2013.
This work was supported by the grant SVV 260105/2014.
This work was supported by the grant SVV 260225/2015.

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

Prague, 21st August 2015 Radek Hendrych

Název práce: Vybrané problémy finančních časových řad

Autor: Radek Hendrych

Katedra: Katedra pravděpodobnosti a matematické statistiky (KPMS)

Vedoucí disertační práce: Prof. RNDr. Tomáš Cipra, DrSc., KPMS

Abstrakt: Předložená disertační práce se věnuje vybraným problémům z oblasti analýzy finančních časových řad. Detailněji se zaměřuje na dva dílčí aspekty modelování podmíněné heteroskedasticity. První část práce prezentuje a diskutuje různé rekurentní odhadové algoritmy navržené pro běžné jednorozměrné modely podmíněné heteroskedasticity, jmenovitě pro procesy typu ARCH, GARCH, RiskMetrics EWMA a GJR-GARCH. Tyto procedury jsou numericky posouzeny prostřednictvím Monte Carlo experimentů a empirických studií vycházejících z reálných dat. Ve druhé části práce je představen inovativní přístup k modelování podmíněných kovariancí (korelací). Odvození navrhované techniky bylo inspirováno ústřední myšlenkou mnohorozměrného ortogonálního GARCH modelu. Principiálně je založeno na vhodném typu lineární dynamické ortogonální transformace, která dále dovoluje aplikovat standardní schéma konstantních podmíněných korelací. Odpovídající model je implementován pomocí nelineární diskrétní stavové reprezentace. Navržený přístup je srovnáván s ostatními, běžně používanými metodami. Ukazuje se jako přinejmenším konkurenceschopný.

Klíčová slova: finanční časové řady, GARCH, podmíněná heteroskedasticita, podmíněné korelace, rekurentní odhady

Title: Selected problems of financial time series modelling

Author: Radek Hendrych

Department: Department of Probability and Mathematical Statistics (DPMS)

Supervisor: Prof. RNDr. Tomáš Cipra, DrSc., DPMS

Abstract: The present dissertation thesis deals with selected problems of financial time series analysis. In particular, it focuses on two fundamental aspects of conditional heteroscedasticity modelling. The first part of the thesis introduces and discusses self-weighted recursive estimation algorithms for several classic univariate conditional heteroscedasticity models, namely for the ARCH, GARCH, RiskMetrics EWMA, and GJR-GARCH processes. Their numerical capabilities are demonstrated by Monte Carlo experiments and real data examples. The second part of the thesis proposes a novel approach to conditional covariance (correlation) modelling. The suggested modelling technique has been inspired by the essential idea of the multivariate orthogonal GARCH method. It is based on a suitable type of linear time-varying orthogonal transformation, which enables to employ the constant conditional correlation scheme. The corresponding model is implemented by using a nonlinear discrete-time state space representation. The proposed approach is compared with other commonly applied models. It demonstrates its competitiveness.

Keywords: conditional correlation, conditional heteroscedasticity, financial time series, GARCH, recursive estimation

Contents

Introduction

Financial time series are an integral part of our everyday lives. Newspapers, television channels, and news websites regularly publish current foreign exchange rates, prices of stocks quoted at various markets, information on returns of mutual funds, current interest and mortgage rates, etc. Financial time series analysis has rightfully attracted attention in recent years. This process has been further intensified by the unforeseen expansion of computer and information technologies. Academics, financial analysts, and other participants in financial markets motivated by various reasons try to understand and describe systematically behaviour of financial series from different perspectives. Refer e.g. to Morgan (1996) and Tsay (2005). Although this highly empirical research field has many specific subdisciplines, one may identify that statistical methods and econometric models characterizing volatility (or the conditional variance) of financial time series are probably the most common. Volatility is, in fact, an elementary indicator of risk, which plays a fundamental role in financial and risk management. Topicality and relevance of this particular issue have motivated this work.

The present thesis focuses on various aspects of univariate and multivariate conditional covariance modelling. It summarizes research results that have been performed during the author's doctoral study. Some of them have been already accepted for publication (Hendrych, 2013, 2014b; Hendrych & Cipra, 2014, 2015); other will be summed up and submitted shortly. See also below. The text of this thesis is organized as follows. After reviewing some basic features of financial time series and fundamental characteristics of volatility in Chapter 1, the work is divided into two separate parts corresponding to two relevant problems of financial time series modelling.

Part I discusses recursive estimation algorithms suitable for selected typical models of conditional heteroscedasticity such as the ARCH or GARCH processes. These models are frequently used to investigate and describe volatility of financial returns systematically. They are routinely estimated by computationally complex estimation methods. The most common is probably the conditional maximum likelihood procedure. However, they are rarely calibrated recursively (i.e. sequentially or on-line). On the other hand, it might be advantageous to adopt numerically effective techniques that could be able to estimate, monitor, and control parameters of the models (or the models themselves) in real time. For instance, one could employ this approach in the case of high-frequency data. Consequently, the main goals of Part I are: (i) to review previously introduced recursive estimation algorithms, (ii) to derive self-weighted one-stage alternatives

applying general recursive identification instruments, and (iii) to examine these methods by means of Monte Carlo simulations and empirical applications.

In particular, Chapter 2 briefly introduces general recursive estimation principles and instruments. Chapter 3 shortly recapitulates the selected classic conditional heteroscedasticity models, namely the ARCH, GARCH, IGARCH, and GJR-GARCH processes. Chapter 4 comments the previously introduced recursive identification schemes suggested for the GARCH processes (Hendrych, 2014b). Chapter 5 concentrates on derivation and justification of various recursive estimators. In detail, Section 5.1 presents two different self-weighted one-stage sequential estimation algorithms for parameters of the GARCH models, which have been considered by Hendrych and Cipra (2015). Section 5.2 discusses various practical aspects of their implementation, for instance, the questions of initialization, factorization, regularization, model identification, or robustification. Section 5.3 reviews theoretical features of the delivered estimation methods. Finally, Section 5.4 introduces several modifications of the considered estimators, which can be used for other conditional heteroscedasticity processes, concretely for the ARCH, RiskMetrics EWMA, and GJR-GARCH models. Chapter 6 investigates behaviour of the suggested recursive methods using Monte Carlo experiments and two empirical applications.

Part II presents a novel approach to conditional covariance modelling (i.e. the subdiscipline of multivariate financial time series analysis), which has been originally proposed by Hendrych (2013) or Hendrych and Cipra (2014). The suggested modelling technique is inspired by the key idea of the multivariate orthogonal GARCH method. This approach models certain linear time-invariant orthogonal combinations of multivariate financial time series components by the constant conditional correlation (CCC) scheme. Here, a suitable linear dynamic orthogonal transformation is recommended instead. It is simply based on the LDL factorization of the conditional covariance matrix. The corresponding model implementation is realized by using nonlinear discrete-time state space models. The introduced procedure has been investigated by extensive Monte Carlo experiments and empirical financial applications. It has been compared with other methods commonly used in this framework. Consequently, the outlined methodology has demonstrated its capabilities and seems to be competitive.

In particular, Chapter 7 briefly recapitulates principles of conditional covariance and correlation modelling. It outlines motivation examples and establishes the modelling framework. Further, it reviews some standard models, which are employed in this particular context. Chapter 8 presents the suggested modelling method and its calibration in more detail. Chapter 9 examines the proposed technique by Monte Carlo experiments and two real data examples.

1. Financial time series modelling

Financial time series modelling is a specific highly empirical subdiscipline of general time series analysis. It is typically concerned with the theory and practice of asset valuation over time. The financial theory (either explicitly, or implicitly) calculates with an element of uncertainty and some special characteristics (see below), which are accordingly accumulated in observed time series. It particularly means that although financial time series analysis is primarily based on the classic concepts described e.g. by Brockwell and Davis (1991), Brockwell and Davis (2002), or Lütkepohl (2005), distinctive theoretical and methodological approaches have been developed in this research field in order to reflect its peculiarities (Franke, Härdle & Hafner, 2011; Tsay, 2005). Standard examples of financial time series are the prices of shares quoted on stock exchanges, various interest rates, or foreign exchange rates amongst different currencies. They may also differ in their frequencies, for instance, one can encounter daily, weekly, monthly, yearly, or high-frequency datasets. The latter one is collected during very short instants, e.g. one second or after each single trade.

Typical financial time series is not (weakly) stationary. The associated autocorrelation coefficient is usually close to one at the first lag. For example, one can hardly consider that the daily foreign exchange rate between the US dollar and Euro currency significantly changes from today to tomorrow. Therefore, most empirical studies treat asset returns instead of asset prices since they operate with more attractive statistical properties. Thus, they are more relevant from the methodological perspective. Additionally, returns of assets introduce a complete and scale-free summary of the investment opportunity from the viewpoint of an average investor. There exist several definitions of asset returns (Tsay, 2005, Chapter 1). However, in practice, one uses mainly:

$$r_t := \log(P_t) - \log(P_{t-1}) \quad \text{or} \quad R_t := \frac{P_t - P_{t-1}}{P_{t-1}}, \tag{1.1}$$

where P_t is the price of an asset at time t. Both introduced returns clearly measure the relative change in price. The *logarithmic return* (*log-return*) r_t and *simple net return* R_t are closely related because:

$$r_t = \log(1 + R_t) \approx R_t. \tag{1.2}$$

Fan and Yao (2005, Section 4.2.7) underlined some important stylized features of (daily) financial (logarithmic) return series, which have been repeatedly identified across all kinds of assets:

Stylized fact 1. (*Leptokurtic distribution*) The financial returns have a leptokurtic distribution, i.e. the distribution with heavier tails than a normal distribution. The empirically estimated kurtosis is mostly greater than three (or zero with respect to the accepted definition of the kurtosis).

Stylized fact 2. (*Volatility clustering*) Volatility (i.e. the conditional standard deviation of asset returns) tends to create clusters. In particular, after a large (small) price change probably occurs a large (small) price change. It means that large volatility changes are likely followed by large volatility changes and that periods of tranquillity alternate with periods of high volatility.

Stylized fact 3. (*Asymmetry*) The distribution of returns is usually slightly negatively skewed. It is commonly explained by asymmetric reactions of traders, who react possibly more strongly to negative than to positive information.

Stylized fact 4. (*Aggregational Gaussianity*) The returns over a long time horizon tend toward a normal distribution according to the central limit law. It can be viewed from the relation for the logarithmic return computed over $k \in \mathbb{N}$ periods, i.e.:

$$\log(P_k) - \log(P_0) = \sum_{t=1}^{k} [\log(P_t) - \log(P_{t-1})] = \sum_{t=1}^{k} r_t. \tag{1.3}$$

Stylized fact 5. (*Long range dependence*) The returns commonly do not exhibit any serial correlation; however, this does not imply that they are independent. On the contrary, both squared and absolute returns often show persistent autocorrelations, which may indicate a possible long-memory dependence in those transformed time series.

Financial time series models usually aim to incorporate at least some of these stylized features to deliver a suitable scheme that respects the empirical findings. See e.g. Chapter 3 of this work, Franke et al. (2011, Chapter 13), or Tsay (2005) for further insights and more relevant references.

Many statistical methods and econometric models of financial time series primarily focus on volatility of an asset return, i.e. on its conditional standard deviation. These models are referred to as *conditional heteroscedasticity models*. Volatility is a major factor with many financial applications in option pricing, risk management, asset allocation under the mean-variance framework (see Example 7.0.1), etc. It can be looked upon as an indicator of risk. Nevertheless,

volatility is not directly observable so that it is hard to represent it systematically. However, it has some conventional characteristics that have been observed across various return series (Tsay, 2005, Section 3.1). Firstly, there usually occur volatility clusters (see Stylized fact 1 above). Secondly, volatility evolves over time almost continuously, i.e. jumps are rare. Thirdly, volatility does not diverge to infinity. Instead, it varies within a fixed range (it is often (weakly) stationary). Fourthly, volatility tends to react differently to a big price increase or decrease (see Stylized fact 3 above). These properties evidently play a substantial role in the volatility modelling framework. See the models discussed in Chapter 3.

From the general perspective, the conditional heteroscedasticity models may be classified into two main groups. The first category tracks the volatility evolution by using an exact function of past information, e.g. the ARCH or GARCH processes. On the contrary, the second category describes volatility by a stochastic equation, i.e. the so-called *stochastic volatility models*. Refer to Tsay (2005, Section 3.2 and others).

Alternatively, one can also emphasize another approach, which originates from different grounds. It accepts the idea that the prices of an underlying asset are governed by an econometric scheme such as the Black-Scholes formula for pricing European options (Tsay, 2005, Chapter 3), which can be further used to deduce the *implied volatility*. However, this approach is possible only under some assumptions, namely that the price follows a geometric Brownian motion, which may be far from reality. Thus, the implied volatility estimates can be imprecise.

The content of the present thesis is closely connected to the already discussed conditional heteroscedasticity models, and thus to volatility analysis. Part I studies recursive estimation techniques for selected classic univariate conditional heteroscedasticity models. They may be applied in many different situations. Part II examines the multivariate extension of the univariate framework and introduces a novel methodology for conditional covariance (correlation) modelling.

Part I

Recursive estimation of selected conditional heteroscedasticity models

2. Recursive estimation methods for financial time series models

The parameters of conditional heteroscedasticity models, the key instruments of systematic financial volatility modelling, are routinely calibrated maximizing associated (conditional) log-likelihood criteria (although other variants exist). See e.g. Fan and Yao (2005, Section 4.2) and the references given therein. On the contrary, they are indeed rarely estimated recursively. Nevertheless, this approach might be advantageous in many (practical) situations, especially in the context of high-frequency data. Particularly, it might be effective to dispose of numerically efficient techniques that could estimate or control models sequentially.

Estimation (identification) methods that comply with this requirement are called *recursive estimation techniques* since the measured input-output data are processed recursively as they become available. Alternatively, one may also use the terms as *on-line*, *real-time*, *adaptive*, or *sequential* (parameter) estimation. To evaluate parameter estimates at a time step, on-line procedures operate only with the current measurement and the parameter estimates delivered in the previous step. It means that an estimate based on data up to time $t-1$ is "slightly" modified to compute the current estimate using the actual observation delivered at time t. It is in sharp contrast to the so-called off-line (batch) methods, in which all recorded data are employed simultaneously to construct the parameter estimates. From this perspective, it might be evident that the recursive estimation algorithms represent a central part of any adaptive system designed primarily for monitoring, controlling, forecasting, or filtering stochastic processes. Here, an action is taken getting the most recent estimates. They are also effective in terms of memory storage and computational complexity since only a modest amount of data is stored, and this amount will not increase with time. These methods can easily handle tracking time-varying parameters since the algorithms might be adapted to respond to data changes and to disregard (gradually) information contained in old data points. Moreover, the recursive estimation techniques can be further applied in fault detection algorithms, which verify whether systems remain stable in some sense (e.g. for detecting structural breaks).

This efficiency can be readily employed in the framework of financial time series analysis. To be more precise, the parameters of selected conditional heteroscedasticity models (see Chapter 3) may be estimated on-line to treat high-frequency data in real time. Otherwise, it is possible to accept such methods to monitor or forecast volatility on-line, to evaluate risk measures (e.g. the Value

at Risk) in real time, to detect eventual faults or structural changes sequentially, to control positivity and to check conditions of stationarity adaptively, etc. The research compiled in Part I of this thesis has been readily motivated by these efforts. Namely, after reviewing selected common conditional heteroscedasticity processes in Chapter 3 and surveying literature in Chapter 4, we shall propose and examine the recursive estimation schemes suitable for these models. They can be undoubtedly applied in many different ways.

Remark 2.0.1 A general estimation method is usually defined as a mapping specified by an (implicitly) given function $\boldsymbol{F}$ from the set of available measurements at time t (denoted as $\boldsymbol{Y}_t$) to a parametric space . In particular, it can be formally written as:

$$\widehat{\boldsymbol{\theta}}_t := \boldsymbol{F}(t, \boldsymbol{Y}_t). \tag{2.1}$$

Such an expression cannot be used in any recursive technique due to its generality since the evaluation of $\boldsymbol{F}$ may involve an unforeseen amount of operations (that may not be terminated at the next sampling moment). Instead of that, a recursive algorithm obviously complies with the following format:

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1} + \alpha_t \boldsymbol{Q}_{\boldsymbol{\theta}}(\boldsymbol{X}_t, \boldsymbol{y}_t), \tag{2.2a}$$

$$\boldsymbol{X}_t = \boldsymbol{X}_{t-1} + \beta_t \boldsymbol{Q}_{\boldsymbol{X}}(\boldsymbol{X}_{t-1}, \boldsymbol{y}_t), \quad t \in \mathbb{N}, \tag{2.2b}$$

where $\boldsymbol{X}_t$ is a vector of fixed dimension that represents some "information state", the functions $\boldsymbol{Q}_{\boldsymbol{\theta}}$ and $\boldsymbol{Q}_{\boldsymbol{X}}$ are explicit expressions that can be evaluated by means of a fixed and a priori known number of operations, and $\{\alpha_t\}_{t\in\mathbb{N}}$ and $\{\beta_t\}_{t\in\mathbb{N}}$ are deterministic sequences of small positive numbers that reflect the relative information value in the latest measurement. Notice that this additive formulation is usually justified by claiming that the information content in the most recent observation $\boldsymbol{y}_t$ is generally small compared to the information already accumulated from the previous measurements. See Ljung (1999, Section 11.1).

Remark 2.0.2 Recursive estimation techniques are usually derived as approximations of off-line methods. Therefore, it may happen that such an approximation reduces accuracy (at least for small finite samples). On the other hand, it is possible to use the on-line algorithms iteratively to remove these inaccuracies and to achieve results that are comparable with those determined by optimizing the corresponding off-line criterion. However, it should be pointed out that a user (commonly) chooses between different off-line or different on-line techniques according to the character of the solved problem.

3. Selected conditional heteroscedasticity models

The conditional heteroscedasticity models provide a systematic framework for volatility modelling. They are capable of respecting (some of) the stylized facts about financial returns and volatility (see Chapter 1). Therefore, they are regarded as fundamental instruments in this field of research. This chapter reviews some of these models jointly with their key features. In particular, the ARCH, GARCH, IGARCH, and GJR-GARCH processes are surveyed in the following sections since they are further studied in Part I of the thesis. Undoubtedly, many other modelling variants might be discussed. An extensive body of academically and practically oriented literature exists in this field of research. For instance, refer to Franke et al. (2011), Fan and Yao (2005), or Tsay (2005) for the references and further insights.

3.1 GARCH model

The pioneering *autoregressive conditional heteroscedasticity (ARCH) model* that conceptually incorporates financial volatility modelling was proposed by Engle (1982). The basic ideas characterizing the ARCH model are that (i) financial returns are serially uncorrelated but dependent, and (ii) this dependence can be described by a simple quadratic function of historical (i.e. lagged) measurements. Although the ARCH model is relatively simple and clearly interpretable (see below), it often requires many parameters to explain adequately the volatility process of asset returns (Tsay, 2005, Section 3.5). Therefore, Bollerslev (1986) suggested a useful extension of this modelling approach known as the *generalized autoregressive conditional heteroscedasticity (GARCH) model*.

The GARCH(p, q) process $\{y_t\}_{t\in\mathbb{Z}}$, $p \in \mathbb{N}$ and $q \in \mathbb{N}_0$, is commonly defined as (Tsay, 2005, Section 3.5):

$$y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{3.1}$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of i. i. d. random variables with zero mean and unit variance (often assumed to be normally distributed), and ω, $\alpha_1, \ldots, \alpha_p$, $\beta_1, \ldots, \beta_q$ are the parameters of the process. The first two conditional moments can be readily calculated as: $\mathbb{E}(y_t|\mathcal{F}_{t-1}) = 0$ and $\mathsf{var}(y_t|\mathcal{F}_{t-1}) = \sigma_t^2$, where $\mathcal{F}_t$ denotes the smallest σ-algebra with respect to which y_s is measurable for all $s \leq t$.

The model specifies a simple parametric function that describes the evolution of volatility. Apparently, sufficient conditions for σ_t^2 being positive are $\omega > 0$, $\alpha_1, \ldots, \alpha_p \geq 0$, $\beta_1, \ldots, \beta_q \geq 0$. If $\beta_1 = \cdots = \beta_q = 0$, the model is reduced to the ARCH(p) case (Tsay, 2005, Section 3.4). Additionally, sufficient conditions for y_t being (weakly) stationary are $\omega > 0$, $\alpha_1, \ldots, \alpha_p \geq 0$, $\beta_1, \ldots, \beta_q \geq 0$, and $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. The stationary GARCH($p, q$) process has the finite variance given by:

$$\mathsf{var}(y_t) = \frac{\omega}{1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j}. \tag{3.2}$$

One can see from (3.1) that large $y_{t-1}^2, \ldots, y_{t-p}^2$ or $\sigma_{t-1}^2, \ldots, \sigma_{t-q}^2$ tend to a large σ_t^2; it generates the well-known behaviour of volatility clustering in financial time series (see Chapter 1). Moreover, if one adopts kurtosis as a measure for heavy tails of distribution, the GARCH model has heavier tails than those of the i. i. d. sequence $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ on which is defined. To this end, let $\kappa_\varepsilon := \mathbb{E}(\varepsilon_t^4)/(\mathbb{E}(\varepsilon_t^2))^2$ denote the kurtosis of the probability distribution of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$. Then, we may compute (assuming that all corresponding moments exist and are finite):

$$\mathbb{E}(y_t^4|\mathcal{F}_{t-1}) = \sigma_t^4 \mathbb{E}(\varepsilon_t^4) = \kappa_\varepsilon [\mathbb{E}(y_t^2|\mathcal{F}_{t-1})]^2. \tag{3.3}$$

Moreover, it follows from the Jensen inequality:

$$\mathbb{E}(y_t^4) = \kappa_\varepsilon \mathbb{E}[\mathbb{E}(y_t^2|\mathcal{F}_{t-1})]^2 \geq \kappa_\varepsilon (\mathbb{E}(y_t^2))^2. \tag{3.4}$$

Consequently, it holds for the kurtosis of $\{y_t\}_{t\in\mathbb{Z}}$ that $\kappa_y := \mathbb{E}(y_t^4)/(\mathbb{E}(y_t^2))^2 \geq \kappa_\varepsilon$, i.e. the tail distribution of the GARCH(p, q) process $\{y_t\}_{t\in\mathbb{Z}}$ is heavier than that of the process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$.

The one-step-ahead forecast of σ_t^2 is calculated as:

$$\widehat{\sigma}_{t+1|t}^2 := \mathbb{E}(\sigma_{t+1}^2|\mathcal{F}_t) = \omega + \sum_{i=1}^{p} \alpha_i y_{t+1-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t+1-j}^2 = \sigma_{t+1}^2. \tag{3.5}$$

The k-step-ahead prediction of σ_t^2 can be evaluated recursively (for $k > 1$). For instance, it holds for the GARCH(1,1) model:

$$\widehat{\sigma}_{t+k|t}^2 := \mathbb{E}(\sigma_{t+k}^2|\mathcal{F}_t) = \omega + (\alpha_1 + \beta_1)\widehat{\sigma}_{t+k-1|t}^2, \tag{3.6}$$

or more precisely:

$$\widehat{\sigma}_{t+k|t}^2 = \frac{\omega[1 - (\alpha_1 + \beta_1)^{k-1}]}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{k-1} \widehat{\sigma}_{t+1|t}^2. \tag{3.7}$$

Therefore, it is clear that $\widehat{\sigma}^2_{t+k|t} \to \omega/(1-\alpha_1-\beta_1)$ as k goes to infinity, provided that $0 \leq \alpha_1 + \beta_1 < 1$. It means that the multi-step-ahead predictions of the GARCH(1,1) process converge to the unconditional variance of $\{y_t\}_{t\in\mathbb{Z}}$ as the forecast horizon increases (supposing that $\mathsf{var}(y_t)$ exists). On the contrary, the ARCH and GARCH models respond equally to positive and negative shocks, which may cause problems for some financial time series (see Chapter 1).

Remark 3.1.1 The ARCH and GARCH models are obviously estimated by maximizing the conditional logarithmic likelihood function (Fan & Yao, 2005, Section 4.2.3). Usually, the normally distributed innovations $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are considered since the quasi-maximum likelihood estimation is still consistent under fairly mild regularity conditions (even if the true distribution is different). However, other calibration alternatives exist. For instance, the Whittle estimator combines the associated ARMA representation (see Section 4.1), its theoretical spectral density, and properties of its periodogram.

3.2 Integrated GARCH model

The *integrated GARCH (IGARCH) model* is an extension of the classic GARCH model. Similarly to ARIMA models, a key feature of this modelling concept is that the impact of past squared shocks $y^2_{t-i} - \sigma^2_{t-i}$ for $i > 0$ on y^2_t is persistent.

The IGARCH(p, q) process $\{y_t\}_{t\in\mathbb{Z}}$, $p \in \mathbb{N}$ and $q \in \mathbb{N}_0$, is commonly defined as (Tsay, 2005, Section 3.6):

$$y_t = \sigma_t \varepsilon_t, \quad \sigma^2_t = \omega + \sum_{i=1}^{p} \alpha_i y^2_{t-i} + \sum_{j=1}^{q} \beta_j \sigma^2_{t-j}, \quad \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1, \tag{3.8}$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of i. i. d. random variables with zero mean and unit variance (often assumed to be normally distributed), and ω, $\alpha_1, \ldots, \alpha_p$, $\beta_1, \ldots, \beta_q$ are the parameters of the process. Apparently, sufficient conditions for σ^2_t being positive are $\omega > 0$, $\alpha_1, \ldots, \alpha_p \geq 0$, $\beta_1, \ldots, \beta_q \geq 0$. The unconditional variance of the process $\{y_t\}_{t\in\mathbb{Z}}$ is not defined under this model. This seems hard to justify for financial return series. From the theoretical point of view, this IGARCH phenomenon may be caused by occasional level shifts in volatility. Under certain conditions, the volatility process is strictly stationary, but not weakly stationary because the first two unconditional moments do not exist. See Tsay (2005, Section 3.6) for further insights and references.

Similarly as in Section 3.1, the k-step-ahead forecast of σ^2_t can be evaluated recursively by using $\widehat{\sigma}^2_{t+1|t} = \sigma^2_{t+1}$ (for $k > 1$). For the most common case, namely

for the IGARCH(1,1) process, one may deduce:

$$\widehat{\sigma}^2_{t+k|t} := \mathbb{E}(\sigma^2_{t+k}|\mathcal{F}_t) = (k-1)\omega + \widehat{\sigma}^2_{t+1|t}. \tag{3.9}$$

The IGARCH(1,1) process with $\omega = 0$, $\alpha_1 = 1 - \lambda$, and $\beta_1 = \lambda$, $\lambda \in (0,1)$ is indeed worth of interest. Particularly, it is defined as:

$$y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = (1-\lambda) y_{t-1}^2 + \lambda \sigma_{t-1}^2, \tag{3.10}$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of i. i. d. random variables with zero mean and unit variance (often assumed to be normally distributed), and $\lambda \in (0,1)$ is the only modelling parameter. In particular, this special IGARCH process is the volatility model used by RiskMetrics (Morgan, 1996), e.g. for calculating the Value at Risk. The model presents, in fact, an exponential smoothing approach to $\{y_t^2\}_{t\in\mathbb{Z}}$ since $\sigma_t^2 = (1-\lambda)\sum_{i=1}^{\infty} \lambda^{i-1} y_{t-i}^2$ by using (3.10) repeatedly. This modelling scheme is usually referred to as the *(RiskMetrics) EWMA model*.

Remark 3.2.1 To calibrate the EWMA model using observations $\{y_1, \ldots, y_T\}$, we usually employ one of the following methods. First, λ is prescribed by experts or users (e.g. the choice 0.94 is obviously recommended for daily data by RiskMetrics). Second, λ is estimated by minimizing the root mean squared error of the forecast inaccuracies $(y_t^2 - \sigma_t^2(\lambda))$ assuming that y_0 and $\sigma_0^2(\lambda)$ are either defined or observed. Third, supposing certain probability distribution of ε_t (the Gaussian innovations are preferred in terms of the consistency of estimates), one may calibrate the parameter λ by maximizing the conditional log-likelihood function (similarly as before, y_0 and $\sigma_0^2(\lambda)$ are supposed to be known).

3.3 GJR-GARCH model

To overcome some weak points of the GARCH and closely related models in handling financial time series, Glosten, Jagannathan and Runkle (1993) proposed another volatility model to reflect leverage effects (see Chapter 1).

Their GJR-GARCH(p,q) process, $p \in \mathbb{N}$ and $q \in \mathbb{N}_0$, is usually formulated as:

$$y_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \omega + \sum_{i=1}^{p} (\alpha_i + \gamma_i I_{t-i}^-) y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{3.11}$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of i. i. d. random variables with zero mean and unit variance (they are symmetrically distributed), and ω, $\alpha_1, \ldots, \alpha_p$, $\gamma_1, \ldots, \gamma_p$, and $\beta_1, \ldots, \beta_q$ are the parameters of the process. Further, I_{t-i}^- denotes an indicator of negative y_{t-i}, i.e. $I_{t-i}^- = 1$ if $y_{t-i} < 0$ and 0 otherwise for $i = 1, \ldots, p$.

Apparently, sufficient conditions for σ_t^2 being positive are $\omega > 0$, $\alpha_i \geq 0$, $\alpha_i + \gamma_i \geq 0$ for all i, and $\beta_j \geq 0$ for all j. Additionally, sufficient conditions for y_t being (weakly) stationary are $\omega > 0$, $\alpha_i \geq 0$, $\alpha_i + \gamma_i \geq 0$ for all i, $\beta_j \geq 0$ for all j, and $\sum_{i=1}^{p}(\alpha_i + \gamma_i/2) + \sum_{j=1}^{q}\beta_j < 1$. The stationary GJR-GARCH(p, q) process has the finite variance given by:

$$\mathsf{var}(y_t) = \frac{\omega}{1 - \sum_{i=1}^{p}(\alpha_i + \gamma_i/2) - \sum_{j=1}^{q}\beta_j}. \tag{3.12}$$

From the outlined representation, it is evident that a positive y_{t-i} contributes $\alpha_i y_{t-i}^2$ to σ_t^2, whereas a negative y_{t-i} has a different impact, namely $(\alpha_i + \gamma_i)y_{t-i}^2$. The zero is used as the threshold to separate the impacts of past observations. One may conclude that this modelling scheme can handle leverage effects. Compare with the list of the stylized features delivered in Chapter 1.

The principle of forecasting remains analogous as before (see Section 3.1). For instance, the k-step-ahead prediction of the GJR-GARCH(1,1) conditional variance σ_t^2 is expressed as (for $k > 1$):

$$\widehat{\sigma}_{t+k|t}^2 := \mathbb{E}(\sigma_{t+k}^2|\mathcal{F}_t) = \omega + (\alpha_1 + \gamma_1/2 + \beta_1)\widehat{\sigma}_{t+k-1|t}^2, \tag{3.13}$$

where $\widehat{\sigma}_{t+1|t}^2$ equals σ_{t+1}^2. Notice that the forecast converges to $\omega/(1 - \alpha_1 - \gamma_1/2 - \beta_1)$ for large k, i.e. to $\mathsf{var}(y_t)$, provided that it exists.

Remark 3.3.1 The GJR-GARCH models are routinely estimated by maximizing the associated conditional log-likelihood criterion assuming suitable distribution of $\{\varepsilon_t\}_{t\in\mathbb{Z}}$. The quasi-maximum likelihood estimator is preferred in practice. See also Remark 3.1.1.

4. Recursive estimation of GARCH models: An overview

This chapter reviews and discusses two closely related sets of previously introduced recursive estimation algorithms developed for the GARCH (or ARCH) modelling parameters. These methods are based on classic recursive estimation (identification) instruments outlined by Ljung (1999), Ljung and Söderström (1983), or Söderström and Stoica (1989). Firstly, Section 4.1 presents the pioneering work by Kierkegaard, Nielsen, Jensen and Madsen (2000), where two complementary on-line estimation procedures have been suggested. Secondly, Section 4.2 addresses attention to a pair of two-stage recursive calibration schemes originally considered by Aknouche and Guerbyenne (2006). The pros and cons of these algorithms are commented. In particular, they have motivated the development of the estimation methods introduced in Chapter 5.

4.1 Methods by Kierkegaard et al.

Kierkegaard et al. (2000) originally suggested two different recursive techniques for estimating the parameters of the GARCH models as eventual alternatives to the common off-line maximum likelihood procedure. To derive these algorithms, the authors employed two associated recursive identification schemes (Ljung, 1999, Chapter 11): (i) the recursive pseudo-linear regression and (ii) the recursive prediction error method.

In particular, the whole computational implementation is based on the following representation of the GARCH(p, q) process $\{y_t\}_{t\in\mathbb{Z}}$:

$$y_t^2 = \sigma_t^2(\boldsymbol{\theta}_0) + v_t, \tag{4.1}$$

where $v_t = \sigma_t^2(\boldsymbol{\theta}_0)(\varepsilon_t^2 - 1)$ and $\boldsymbol{\theta}_0$ denotes the vector of the true values of parameters. Here, $\{v_t\}_{t\in\mathbb{Z}}$ is a sequence of uncorrelated random variables with zero mean. If one additionally assumes that $\mathbb{E}(y_t^4) = const. < \infty$, $\{v_t\}_{t\in\mathbb{Z}}$ is white noise according to Franke et al. (2011, Theorem 13.9).

Equation (4.1) can be rewritten more conveniently as:

$$y_t^2 = \boldsymbol{\varphi}_t^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0 + v_t, \tag{4.2}$$

where

$$\boldsymbol{\theta}_0 = (\omega_0, \alpha_{1,0}, \ldots, \alpha_{p,0}, \beta_{1,0}, \ldots, \beta_{q,0})^\top, \tag{4.3a}$$

$$\boldsymbol{\varphi}_t(\boldsymbol{\theta}_0) = \left(1, y_{t-1}^2, \ldots, y_{t-p}^2, \boldsymbol{\varphi}_{t-1}^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0, \ldots, \boldsymbol{\varphi}_{t-q}^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0\right)^\top, \tag{4.3b}$$

$$\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0 = \omega_0 + \sum_{i=1}^{p} \alpha_{i,0} y_{t-i}^2 + \sum_{j=1}^{q} \beta_{j,0} \boldsymbol{\varphi}_{t-j}^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0. \tag{4.3c}$$

Both recursive pseudo-linear regression and prediction error method can be directly derived from the associated off-line (non-recursive or batch) counterparts (Ljung, 1999, Chapter 11). Assume that observations $\{y_1, \ldots, y_T\}$ are given and that $\{y_{1-p}^2, \ldots, y_0^2\}$ and $\{\boldsymbol{\varphi}_{1-q}^\top(\boldsymbol{\theta})\boldsymbol{\theta}, \ldots, \boldsymbol{\varphi}_0^\top(\boldsymbol{\theta})\boldsymbol{\theta}\}$ are either given or defined, the pseudo-linear regression technique estimates the unknown parameters of (4.2) by solving the following equation for $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^{p+q+1}$ (a parametric space):

$$\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\varphi}_t(\boldsymbol{\theta}) \left[y_t^2 - \boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}\right] = \mathbf{0}. \tag{4.4}$$

It corresponds to the situation, in which the prediction errors $[y_t^2 - \boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}]$ are supposed to be uncorrelated in the given time with the regressors $\boldsymbol{\varphi}_t(\boldsymbol{\theta})$. Similarly, the prediction error method calibrates the model (4.2) by minimizing the sum of the squared prediction errors for $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^{p+q+1}$ (a parametric space). It can be formulated as:

$$\arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{T}\sum_{t=1}^{T} \left(y_t^2 - \boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}\right)^2. \tag{4.5}$$

The recursive pseudo-linear regression algorithm for estimating the parameters of the GARCH(p, q) process is thus defined as (Kierkegaard et al., 2000):

$$\widetilde{\boldsymbol{\theta}}_t = \widetilde{\boldsymbol{\theta}}_{t-1} + \widetilde{\boldsymbol{R}}_t^{-1}\widetilde{\boldsymbol{\varphi}}_t(y_t^2 - \widetilde{\boldsymbol{\varphi}}_t^\top\widetilde{\boldsymbol{\theta}}_{t-1}), \tag{4.6a}$$

$$\widetilde{\boldsymbol{R}}_t = \widetilde{\boldsymbol{R}}_{t-1} + \widetilde{\boldsymbol{\varphi}}_t\widetilde{\boldsymbol{\varphi}}_t^\top,\ t \in \mathbb{N}, \tag{4.6b}$$

where

$$\widetilde{\boldsymbol{\theta}}_t = (\widetilde{\omega}_t, \widetilde{\alpha}_{1,t}, \ldots, \widetilde{\alpha}_{p,t}, \widetilde{\beta}_{1,t}, \ldots, \widetilde{\beta}_{q,t})^\top, \tag{4.7a}$$

$$\widetilde{\boldsymbol{\varphi}}_t = (1, y_{t-1}^2, \ldots, y_{t-p}^2, \widetilde{\boldsymbol{\varphi}}_{t-1}^\top\widetilde{\boldsymbol{\theta}}_{t-2}, \ldots, \widetilde{\boldsymbol{\varphi}}_{t-q}^\top\widetilde{\boldsymbol{\theta}}_{t-q-1})^\top, \tag{4.7b}$$

$$\widetilde{\boldsymbol{\varphi}}_t^\top\widetilde{\boldsymbol{\theta}}_{t-1} = \widetilde{\omega}_{t-1} + \sum_{i=1}^{p} \widetilde{\alpha}_{i,t-1} y_{t-i}^2 + \sum_{j=1}^{q} \widetilde{\beta}_{j,t-1}\widetilde{\boldsymbol{\varphi}}_{t-j}^\top\widetilde{\boldsymbol{\theta}}_{t-j-1}. \tag{4.7c}$$

The recursive prediction error updating scheme for estimating the parameters of the GARCH(p, q) model is given likewise (Kierkegaard et al., 2000):

$$\widetilde{\boldsymbol{\theta}}_t = \widetilde{\boldsymbol{\theta}}_{t-1} + \widetilde{\boldsymbol{R}}_t^{-1}\widetilde{\boldsymbol{\psi}}_t(y_t^2 - \widetilde{\boldsymbol{\varphi}}_t^\top \widetilde{\boldsymbol{\theta}}_{t-1}), \tag{4.8a}$$

$$\widetilde{\boldsymbol{R}}_t = \widetilde{\boldsymbol{R}}_{t-1} + \widetilde{\boldsymbol{\psi}}_t\widetilde{\boldsymbol{\psi}}_t^\top, \tag{4.8b}$$

$$\widetilde{\boldsymbol{\psi}}_{t+1} = \widetilde{\boldsymbol{\varphi}}_{t+1} + \sum_{j=1}^{q} \widetilde{\beta}_{j,t}\widetilde{\boldsymbol{\psi}}_{t+1-j},\ t \in \mathbb{N}, \tag{4.8c}$$

where $\widetilde{\boldsymbol{\theta}}_t$ and $\widetilde{\boldsymbol{\varphi}}_t$ remain as in (4.7). Note that the original algorithm has not included the explicit recurrent formula for computing $\widetilde{\boldsymbol{\psi}}_t$ (Kierkegaard et al., 2000, Equation (25)); therefore, it has been added in accordance with Ljung (1999, Chapter 11). Moreover, it is evident that both described recursive estimation methods are closely related. Namely, they differ only in terms of $\widetilde{\boldsymbol{\varphi}}_t$ and $\widetilde{\boldsymbol{\psi}}_t$. The vector $\widetilde{\boldsymbol{\psi}}_t$ is a recursively evaluated counterpart of $\boldsymbol{\psi}_t(\boldsymbol{\theta})$, where $\boldsymbol{\psi}_t^\top(\boldsymbol{\theta})$ denotes the gradient of the function $\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}$ with respect to $\boldsymbol{\theta}$. See also Section 5.1. This inconspicuous discrepancy is determined by the distinct estimation problems originally inspired by (4.4) and (4.5). The recursive pseudo-linear regression algorithm (4.6) can be also regarded as an approximation of the recursive prediction error method (4.8) by putting $\widetilde{\boldsymbol{\psi}}_t \approx \widetilde{\boldsymbol{\varphi}}_t$. Asymptotically, the recursive estimates converge under the associated assumptions to (local) off-line solutions of the problems (4.4) and (4.5), respectively (Ljung, 1999, Chapter 11).

On the one hand, both on-line estimation techniques are numerically straightforward and easy to implement. On the other hand, they (intuitively) suffer from the loss of efficiency since the original calibration criteria (4.4) and (4.5), from which they are derived, overlook the heteroscedastic structure of the disturbance term v_t in (4.2). In addition, Kierkegaard et al. (2000) omitted to discuss three important aspects of practical implementation. Firstly, it is necessary to adopt a projection algorithm that guarantees the numerical stability of both estimation methods and ensures positivity and stationarity of the GARCH conditional variance. See Sections 4.2 and 5.2. Neglecting this fact, the recursive estimation schemes could theoretically or numerically degenerate (Ljung, 1999, Chapter 11). Secondly, the initial values $\widetilde{\boldsymbol{\theta}}_0$ and $\widetilde{\boldsymbol{R}}_0$ starting both procedures were not discussed. This could be questioned primarily from the practical viewpoint (Hendrych, 2014b). Thirdly, the forgetting factors could be introduced into both procedures to accelerate convergence or to track time-varying parameters.

4.2 Methods by Aknouche and Guerbyenne

Independently, Aknouche and Guerbyenne (2006) proposed a couple of two-stage recursive estimation schemes appropriate for the standard GARCH(p, q) models. The authors extended the ideas introduced by Bose and Mukherjee (2003). However, they focused mainly on the derivation and partly on convergence analysis of the algorithms and not on their numerical evaluation. It can be perceived as a crucial objection (Hendrych, 2014b).

Employing the same representation of the GARCH(p, q) model as in (4.2), the first stages of the recursive estimation methods have been suggested. Similarly as in Section 4.1, the recursive pseudo-linear regression and prediction error estimation schemes have been applied. Nevertheless, both first stages still have ignored the heteroscedasticity of the innovation term v_t from (4.2). See below. Therefore, Aknouche and Guerbyenne (2006) introduced the second stages by using particular weights delivered by the first stages. They should improve the efficiency of the first-stage estimates reflecting the heteroscedastic structure of v_t more precisely. Refer to Section 6.1 for various Monte Carlo experiments.

Lemma 4.2.1 (Matrix inversion lemma) *Let $\boldsymbol{A}$, $\boldsymbol{U}$, $\boldsymbol{C}$, and $\boldsymbol{V}$ be $(n \times n)$, $(n \times k)$, $(k \times k)$, and $(k \times n)$ real matrices, respectively $(k, n \in \mathbb{N})$. It holds:*

$$(\boldsymbol{A}+\boldsymbol{UCV})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}\left(\boldsymbol{C}^{-1}+\boldsymbol{VA}^{-1}\boldsymbol{U}\right)^{-1}\boldsymbol{VA}^{-1}.$$

Proof Here, $(\boldsymbol{A}+\boldsymbol{UCV})\left[\boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}\left(\boldsymbol{C}^{-1}+\boldsymbol{VA}^{-1}\boldsymbol{U}\right)^{-1}\boldsymbol{VA}^{-1}\right] = \boldsymbol{I}$. ∎

To derive the first stage of the recursive pseudo-linear regression estimation algorithm as it was introduced by Aknouche and Guerbyenne (2006), it is sufficient to rearrange (4.6) by applying Lemma 4.2.1, where one puts $\widetilde{\boldsymbol{P}}_t = \widetilde{\boldsymbol{R}}_t^{-1}$, i.e.:

$$\widetilde{\boldsymbol{\theta}}_t = \widetilde{\boldsymbol{\theta}}_{t-1} + \frac{\widetilde{\boldsymbol{P}}_{t-1}\widetilde{\boldsymbol{\varphi}}_t(y_t^2 - \widetilde{\boldsymbol{\varphi}}_t^\top\widetilde{\boldsymbol{\theta}}_{t-1})}{1+\widetilde{\boldsymbol{\varphi}}_t^\top\widetilde{\boldsymbol{P}}_{t-1}\widetilde{\boldsymbol{\varphi}}_t},\ \widetilde{\boldsymbol{\theta}}_0 = \boldsymbol{0}, \tag{4.9a}$$

$$\widetilde{\boldsymbol{P}}_t = \widetilde{\boldsymbol{P}}_{t-1} - \frac{\widetilde{\boldsymbol{P}}_{t-1}\widetilde{\boldsymbol{\varphi}}_t\widetilde{\boldsymbol{\varphi}}_t^\top\widetilde{\boldsymbol{P}}_{t-1}}{1+\widetilde{\boldsymbol{\varphi}}_t^\top\widetilde{\boldsymbol{P}}_{t-1}\widetilde{\boldsymbol{\varphi}}_t},\ \widetilde{\boldsymbol{P}}_0 = \kappa\boldsymbol{I}, \tag{4.9b}$$

where κ is a large positive number, e.g. $\kappa = 10^5$.

The heteroscedasticity of the innovation term v_t defined by (4.2) has not been explicitly taken into account (as in the previous case of (4.6)). For this reason, Aknouche and Guerbyenne (2006) added the second stage by inserting suitable

weights delivered by the first stage as follows:

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1} + \frac{\widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\varphi}}_t(y_t^2 - \widehat{\boldsymbol{\varphi}}_t^\top\widehat{\boldsymbol{\theta}}_{t-1})}{(\widetilde{\boldsymbol{\varphi}}_t^\top\widetilde{\boldsymbol{\theta}}_{t-1})^2 + \widehat{\boldsymbol{\varphi}}_t^\top\widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\varphi}}_t},\ \widehat{\boldsymbol{\theta}}_0 = \mathbf{0}, \tag{4.10a}$$

$$\widehat{\boldsymbol{P}}_t = \widehat{\boldsymbol{P}}_{t-1} - \frac{\widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\varphi}}_t\widehat{\boldsymbol{\varphi}}_t^\top\widehat{\boldsymbol{P}}_{t-1}}{(\widetilde{\boldsymbol{\varphi}}_t^\top\widetilde{\boldsymbol{\theta}}_{t-1})^2 + \widehat{\boldsymbol{\varphi}}_t^\top\widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\varphi}}_t},\ \widehat{\boldsymbol{P}}_0 = \kappa\boldsymbol{I}, \tag{4.10b}$$

where κ is a large positive number, e.g. $\kappa = 10^5$, and where

$$\widehat{\boldsymbol{\theta}}_t = (\widehat{\omega}_t, \widehat{\alpha}_{1,t}, \ldots, \widehat{\alpha}_{p,t}, \widehat{\beta}_{1,t}, \ldots, \widehat{\beta}_{q,t})^\top, \tag{4.11a}$$

$$\widehat{\boldsymbol{\varphi}}_t = (1, y_{t-1}^2, \ldots, y_{t-p}^2, \widehat{\boldsymbol{\varphi}}_{t-1}^\top\widehat{\boldsymbol{\theta}}_{t-2}, \ldots, \widehat{\boldsymbol{\varphi}}_{t-q}^\top\widehat{\boldsymbol{\theta}}_{t-q-1})^\top, \tag{4.11b}$$

$$\widehat{\boldsymbol{\varphi}}_t^\top\widehat{\boldsymbol{\theta}}_{t-1} = \widehat{\omega}_{t-1} + \sum_{i=1}^{p}\widehat{\alpha}_{i,t-1}y_{t-i}^2 + \sum_{j=1}^{q}\widehat{\beta}_{j,t-1}\widehat{\boldsymbol{\varphi}}_{t-j}^\top\widehat{\boldsymbol{\theta}}_{t-j-1}. \tag{4.11c}$$

As was noted in Section 4.1, the recursive pseudo-linear regression can be regarded as an approximation of the more complex recursive prediction error method that reflects the modelling structure more conveniently (Ljung, 1999, Chapter 11). Therefore, one could generalize the preceding two-stage algorithm just using this technique. Similarly as before, the first stage of the recursive prediction error procedure by Aknouche and Guerbyenne (2006) may be directly developed from (4.8) by employing Lemma 4.2.1 using $\widetilde{\boldsymbol{P}}_t = \widetilde{\boldsymbol{R}}_t^{-1}$, i.e.:

$$\widetilde{\boldsymbol{\theta}}_t = \widetilde{\boldsymbol{\theta}}_{t-1} + \frac{\widetilde{\boldsymbol{P}}_{t-1}\widetilde{\boldsymbol{\psi}}_t(y_t^2 - \widetilde{\boldsymbol{\varphi}}_t^\top\widetilde{\boldsymbol{\theta}}_{t-1})}{1 + \widetilde{\boldsymbol{\psi}}_t^\top\widetilde{\boldsymbol{P}}_{t-1}\widetilde{\boldsymbol{\psi}}_t},\ \widetilde{\boldsymbol{\theta}}_0 = \mathbf{0}, \tag{4.12a}$$

$$\widetilde{\boldsymbol{P}}_t = \widetilde{\boldsymbol{P}}_{t-1} - \frac{\widetilde{\boldsymbol{P}}_{t-1}\widetilde{\boldsymbol{\psi}}_t\widetilde{\boldsymbol{\psi}}_t^\top\widetilde{\boldsymbol{P}}_{t-1}}{1 + \widetilde{\boldsymbol{\psi}}_t^\top\widetilde{\boldsymbol{P}}_{t-1}\widetilde{\boldsymbol{\psi}}_t},\ \widetilde{\boldsymbol{P}}_0 = \kappa\boldsymbol{I}, \tag{4.12b}$$

$$\widetilde{\boldsymbol{\psi}}_{t+1} = \widetilde{\boldsymbol{\varphi}}_{t+1} + \sum_{j=1}^{q}\widetilde{\beta}_{j,t}\widetilde{\boldsymbol{\psi}}_{t+1-j}, \tag{4.12c}$$

where the notation remains unchanged.

The second stage can be deduced similarly as before. The term $(\widetilde{\boldsymbol{\varphi}}_t^\top\widetilde{\boldsymbol{\theta}}_{t-1})^2$ computed by using (4.12a) and (4.12b) is included instead of the units in the denominators of (4.12) to reflect the heteroscedasticity of v_t from (4.2) more accurately (Aknouche & Guerbyenne, 2006):

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1} + \frac{\widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\psi}}_t(y_t^2 - \widehat{\boldsymbol{\varphi}}_t^\top\widehat{\boldsymbol{\theta}}_{t-1})}{(\widetilde{\boldsymbol{\varphi}}_t^\top\widetilde{\boldsymbol{\theta}}_{t-1})^2 + \widehat{\boldsymbol{\psi}}_t^\top\widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\psi}}_t},\ \widehat{\boldsymbol{\theta}}_0 = \mathbf{0}, \tag{4.13a}$$

$$\widehat{\boldsymbol{P}}_t = \widehat{\boldsymbol{P}}_{t-1} - \frac{\widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\psi}}_t\widehat{\boldsymbol{\psi}}_t^\top\widehat{\boldsymbol{P}}_{t-1}}{(\widetilde{\boldsymbol{\varphi}}_t^\top\widetilde{\boldsymbol{\theta}}_{t-1})^2 + \widehat{\boldsymbol{\psi}}_t^\top\widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\psi}}_t},\ \widehat{\boldsymbol{P}}_0 = \kappa\boldsymbol{I}, \tag{4.13b}$$

$$\widehat{\boldsymbol{\psi}}_{t+1} = \widehat{\boldsymbol{\varphi}}_{t+1} + \sum_{j=1}^{q} \widehat{\beta}_{j,t} \widehat{\boldsymbol{\psi}}_{t+1-j}, \tag{4.13c}$$

where the notation remains unchanged.

From the practical and theoretical point of view, one should adopt a monitoring procedure that projects the current on-line estimate onto the region of stability to avoid the algorithm degeneration. See Section 5.2.4 for further details. For instance, one can apply the technique originally recommended by Ljung and Söderström (1983, Section 6.6) that completes all presented recursive estimation methods. For example, in the case of (4.9), it may be formulated as:

Step 1. At each time $t \in \mathbb{N}$, set $\mu \in [0, 1)$ and $i = 1$.

Step 2. Check if $\widetilde{\boldsymbol{\theta}}_t$ lies in the defined region of stability denoted as $\boldsymbol{D}_{\mathcal{M}}$.

- If yes, stop.
- Otherwise, go to **Step 3**.

Step 3. Evaluate $\widetilde{\boldsymbol{\theta}}_t = \widetilde{\boldsymbol{\theta}}_{t-1} + \mu^i \frac{\widetilde{\boldsymbol{P}}_{t-1}\widetilde{\boldsymbol{\varphi}}_t(y_t^2 - \widetilde{\boldsymbol{\varphi}}_t^\top \widetilde{\boldsymbol{\theta}}_{t-1})}{1+\widetilde{\boldsymbol{\varphi}}_t^\top \widetilde{\boldsymbol{P}}_{t-1}\widetilde{\boldsymbol{\varphi}}_t}$. Set $i = i + 1$. Skip to **Step 2**.

Evidently, this projection can be simply used with all presented calibration techniques. The term μ^i adequately adjusts the effect of the new observation. However, for $\mu \neq 0$ this approach would decelerate computations. See Section 5.2.4 for the definition of $\boldsymbol{D}_{\mathcal{M}}$. Note that the structure of the stability region was not discussed in detail by Aknouche and Guerbyenne (2006).

Let us stress the fact that the first stages of both two-stages procedures proposed by Aknouche and Guerbyenne (2006) completely coincide with the methods outlined in Section 4.1, which were suggested by Kierkegaard et al. (2000). They have been derived by using the similar model representation and the identical recursive estimation instruments. They differ only in the rearrangement applying Lemma 4.2.1. Introducing the second stages with the particular weights has been motivated by pragmatic econometric approaches (Bose & Mukherjee, 2003). Two-stage estimation methods are relatively common for improving efficiency.

Additionally, one should pay attention to several other points concerning the discussed recursive estimation procedures. Firstly, both the recursive pseudo-linear regression and the recursive prediction error methods in the present form have been originally derived to solve the problems given by (4.4) and (4.5) sequentially. More specifically, they have been primarily formulated for linear ARMA models with homoscedastic disturbances. Although the heteroscedasticity of the term v_t from (4.2) is reflected in the second stages of both algorithms by inserting the specific weights, there still might exist doubts whether this adaptation is sufficient or not. Compare with the general scheme considered in Ljung (1999,

Chapter 11). Secondly, all algorithms are initialized by zero vectors. They apparently lie outside of the region of positivity and stationarity. It means that the positivity and stationarity of the recursively evaluated GARCH conditional variance are not assured in general (according to the accepted projection). It might eventually decelerate or even degenerate the estimation. Refer to Section 6.1 for simulation experiments. Thirdly, the white noise in (4.2) may be distributed non-trivially (even if the original innovations have a common distribution). Fourthly, two-stage procedures are relatively computationally complex in comparison with one-stage ones. Analogous algorithms could be suggested as one-stage by respecting the heteroscedasticity of v_t immediately (see Section 5.1). Nevertheless, the numbers of operations per iteration time corresponding to the one-stage and two-stage procedures are still proportional. Finally, the proposed techniques do not take into account pragmatic approaches to recursive estimation surveyed e.g. by Ljung (1999, Chapter 11). Namely, they usually introduce additional deterministic weights to accelerate the convergence. See Section 5.2.2.

These facts and the results of simulations presented by Hendrych (2014b) have inspired derivation of alternative one-stage self-weighted estimation techniques. These are introduced in Section 5.1. In particular, one has been motivated by the demand for reliable on-line estimation techniques that could be comfortably applied in practice. Moreover, the quality of such sequential estimation procedures should be (asymptotically) comparable with the off-line methods (Ljung & Söderström, 1983, Chapter 4).

Remark 4.2.1 The first (non-weighted) stage of the recursive prediction error estimation algorithm (4.12) may be also viewed from the perspective of the extended Kalman filter (Chui & Chen, 2013, Chapter 8). In particular, assume the following nonlinear state space modelling representation:

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1}, \tag{4.14a}$$

$$y_t^2 = \boldsymbol{\varphi}_t^\top(\boldsymbol{\theta}_t)\boldsymbol{\theta}_t + e_t, \tag{4.14b}$$

where the notation is obvious (see Section 4.1) and $t \in \mathbb{N}$. The error terms $\{e_t\}_{t\in\mathbb{N}}$ are i. i. d. random variables with the normal distribution $N(0, \sigma_e^2)$, $\sigma_e^2 \in (0, \infty)$. The initial state vector $\boldsymbol{\theta}_0$ is supposed to be $N(\widetilde{\boldsymbol{\theta}}_0, \widetilde{\boldsymbol{P}}_0)$ independently of $e_1, e_2, \ldots$, where both quantities $\widetilde{\boldsymbol{\theta}}_0$ and $\widetilde{\boldsymbol{P}}_0$ are assumed to be known. The state equation (4.14a) models the path of estimated parameters. The signal (measurement) equation (4.14b) partly reflects the GARCH(p, q) framework given by (4.2). Here, the simpler homoscedastic normally distributed i. i. d. disturbances are considered. This undoubtedly simplifies the original structure. Remind that the function $\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})$ depends only on $\boldsymbol{\theta}$ and the past measurements. Refer to (4.3).

To provide the associated extended Kalman filtering formulas, the nonlinear term $\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}$ must be linearized using the first-order Taylor expansion around $\widetilde{\boldsymbol{\theta}}$, i.e. $\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta} \approx \boldsymbol{\varphi}_t^\top(\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{\theta}} + \boldsymbol{\psi}_t^\top(\widetilde{\boldsymbol{\theta}})(\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}})$, where $\boldsymbol{\psi}_t^\top(\widetilde{\boldsymbol{\theta}})$ denotes the gradient of the function $\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}$ with respect to $\boldsymbol{\theta}$ evaluated in $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}$.

Define $\widetilde{\boldsymbol{\theta}}_{t|t-1} := \mathbb{E}(\boldsymbol{\theta}_t|\boldsymbol{Y}_{t-1}^2)$ and $\widetilde{\boldsymbol{P}}_{t|t-1} := \mathsf{cov}(\boldsymbol{\theta}_t|\boldsymbol{Y}_{t-1}^2)$, where $\boldsymbol{Y}_t^2$ is the collection of all observations up to and including time t. It is clear that $\widetilde{\boldsymbol{\theta}}_{t|t-1} = \widetilde{\boldsymbol{\theta}}_{t-1|t-1}$ and $\widetilde{\boldsymbol{P}}_{t|t-1} = \widetilde{\boldsymbol{P}}_{t-1|t-1}$ according to (4.14a) using the obvious notation. Applying the already mentioned linearization around $\widetilde{\boldsymbol{\theta}}_{t|t-1}$, one computes:

$$\mathbb{E}(y_t^2|\boldsymbol{Y}_{t-1}^2) \approx \boldsymbol{\varphi}_t^\top(\widetilde{\boldsymbol{\theta}}_{t|t-1})\widetilde{\boldsymbol{\theta}}_{t|t-1}, \tag{4.15a}$$

$$\mathsf{var}(y_t^2|\boldsymbol{Y}_{t-1}^2) \approx \sigma_e^2 + \boldsymbol{\psi}_t^\top(\widetilde{\boldsymbol{\theta}}_{t|t-1})\widetilde{\boldsymbol{P}}_{t|t-1}\boldsymbol{\psi}_t(\widetilde{\boldsymbol{\theta}}_{t|t-1}) =: \widetilde{\boldsymbol{S}}_{t|t-1}, \tag{4.15b}$$

$$\mathsf{cov}(\boldsymbol{\theta}_t, y_t^2|\boldsymbol{Y}_{t-1}^2) \approx \widetilde{\boldsymbol{P}}_{t|t-1}\boldsymbol{\psi}_t(\widetilde{\boldsymbol{\theta}}_{t|t-1}) =: \widetilde{\boldsymbol{M}}_{t|t-1}. \tag{4.15c}$$

The Kalman filter can be derived under the assumption that $\boldsymbol{\theta}_t$ and y_t^2 are jointly conditionally normally distributed given $\boldsymbol{Y}_{t-1}^2$. Namely, suppose that $(\boldsymbol{\theta}_t^\top, y_t^2)^\top|\boldsymbol{Y}_{t-1}^2$ has the normal distribution with this characterization:

$$N\left(\begin{pmatrix} \widetilde{\boldsymbol{\theta}}_{t|t-1} \\ \boldsymbol{\varphi}_t^\top(\widetilde{\boldsymbol{\theta}}_{t|t-1})\widetilde{\boldsymbol{\theta}}_{t|t-1} \end{pmatrix}, \begin{pmatrix} \widetilde{\boldsymbol{P}}_{t|t-1} & \widetilde{\boldsymbol{M}}_{t|t-1} \\ \widetilde{\boldsymbol{M}}_{t|t-1}^\top & \widetilde{\boldsymbol{S}}_{t|t-1} \end{pmatrix}\right). \tag{4.16}$$

Employing the lemma from optimal estimation theory (Durbin & Koopman, 2001, Section 2.13) and setting $\widetilde{\boldsymbol{\theta}}_t := \widetilde{\boldsymbol{\theta}}_{t|t}$ and $\widetilde{\boldsymbol{P}}_t := \widetilde{\boldsymbol{P}}_{t|t}/\sigma_e^2$, one obtains:

$$\widetilde{\boldsymbol{\theta}}_t = \widetilde{\boldsymbol{\theta}}_{t-1} + \frac{\widetilde{\boldsymbol{P}}_{t-1}\boldsymbol{\psi}_t(\widetilde{\boldsymbol{\theta}}_{t-1})(y_t^2 - \boldsymbol{\varphi}_t^\top(\widetilde{\boldsymbol{\theta}}_{t-1})\widetilde{\boldsymbol{\theta}}_{t-1})}{1 + \boldsymbol{\psi}_t^\top(\widetilde{\boldsymbol{\theta}}_{t-1})\widetilde{\boldsymbol{P}}_{t-1}\boldsymbol{\psi}_t(\widetilde{\boldsymbol{\theta}}_{t-1})}, \tag{4.17a}$$

$$\widetilde{\boldsymbol{P}}_t = \widetilde{\boldsymbol{P}}_{t-1} - \frac{\widetilde{\boldsymbol{P}}_{t-1}\boldsymbol{\psi}_t(\widetilde{\boldsymbol{\theta}}_{t-1})\boldsymbol{\psi}_t^\top(\widetilde{\boldsymbol{\theta}}_{t-1})\widetilde{\boldsymbol{P}}_{t-1}}{1 + \boldsymbol{\psi}_t^\top(\widetilde{\boldsymbol{\theta}}_{t-1})\widetilde{\boldsymbol{P}}_{t-1}\boldsymbol{\psi}_t(\widetilde{\boldsymbol{\theta}}_{t-1})}, \ t \in \mathbb{N}. \tag{4.17b}$$

Obviously, the introduced system almost coincides with the algorithm (4.12). However, this particular scheme is not appropriate for recursive estimating since $\boldsymbol{\varphi}_t(\widetilde{\boldsymbol{\theta}}_{t-1})$ and $\boldsymbol{\psi}_t(\widetilde{\boldsymbol{\theta}}_{t-1})$ are not evaluated sequentially. Their calculation at each time indeed requires processing all data up to time t. On the other hand, both compared methods are equivalent accepting several pragmatic approximations for $\boldsymbol{\varphi}_t(\widetilde{\boldsymbol{\theta}}_{t-1})$ and $\boldsymbol{\psi}_t(\widetilde{\boldsymbol{\theta}}_{t-1})$. Naturally, one can consider $\boldsymbol{\varphi}_t(\widetilde{\boldsymbol{\theta}}_{t-1}) \approx \widetilde{\boldsymbol{\varphi}}_t$ and $\boldsymbol{\psi}_t(\widetilde{\boldsymbol{\theta}}_{t-1}) \approx \widetilde{\boldsymbol{\psi}}_t$ similarly as in (4.7).

5. Self-weighted recursive estimation of selected conditional heteroscedasticity models

Chapter 4 has reviewed the distinct variants of recursive procedures available for estimating the parameters of the GARCH(p, q) models. All introduced methods have been based on the particular representation of the GARCH process given by Equation (4.2). They have employed and (partly) modified standard recursive identification schemes originally designed for calibrating linear ARMA processes. Nonetheless, one might apply general calibration instruments to deliver one-stage self-weighted on-line estimation algorithms suitable for various classic conditional heteroscedasticity models. The suggested estimators should naturally respond to the remarks discussed in Section 4.2. They should be applicable in practice.

This chapter is organized as follows. Section 5.1 derives a couple of recursive estimation techniques for calibrating the GARCH parameters. Section 5.2 discusses several practical aspects of implementation (e.g. initialization, regularization, and robustification). Section 5.3 justifies both methods from the theoretical perspective. Finally, Section 5.4 presents various modifications of the adopted estimation procedures, which can calibrate other conditional heteroscedasticity processes, namely the ARCH, RiskMetrics EWMA, and GJR-GARCH models.

5.1 Self-weighted recursive estimation methods for GARCH models: Derivation

The following paragraphs outline an explicit derivation of alternative one-stage self-weighted recursive formulas for estimating the GARCH(p, q) process. Namely, the general recursive identification methods summarized by Ljung (1999), Ljung and Söderström (1983), and Söderström and Stoica (1989) are applied in this context to deliver reliable estimation procedures. See Section 4.2.

Initially, let us consider the loss function that is supposed to be minimized:

$$V_t(\boldsymbol{\theta}) = \gamma_t \sum_{k=1}^{t} \left[\prod_{\ell=k}^{t-1} \lambda_\ell \right] F_k(\boldsymbol{\theta}),\ t \in \mathbb{N},\ \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^{p+q+1}, \tag{5.1}$$

where $\boldsymbol{\Theta}$ is a parametric space, and we accept the convention that $\prod_{\ell=t}^{t-1} \lambda_\ell = 1$ with a sequence of positive real numbers $\{\lambda_\ell\}_{\ell=1}^{\infty}$ such that $\gamma_t = 1/\left[\sum_{k=1}^{t} \prod_{\ell=k}^{t-1} \lambda_\ell\right]$

fulfils $\gamma_t > 0$, $\sum_{t=1}^{\infty} \gamma_t = \infty$, and $\sum_{t=1}^{\infty} \gamma_t^2 < \infty$. The meaning of these deterministic weights is explained in more detail in Section 5.2.2. The remaining function $F_k(\boldsymbol{\theta})$ represents a core criterion associated with the particular estimation problem. To calibrate the standard GARCH(p, q) process $\{y_t\}_{t\in\mathbb{Z}}$ defined by (3.1) with the parameters collected in the vector $\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)^\top$ using a given data sample, one may assume the following criterion function:

$$F_k(\boldsymbol{\theta}) = \frac{y_k^2}{\boldsymbol{\varphi}_k^\top(\boldsymbol{\theta})\boldsymbol{\theta}} + \log\left(\boldsymbol{\varphi}_k^\top(\boldsymbol{\theta})\boldsymbol{\theta}\right), \; k \in \mathbb{N}, \tag{5.2}$$

where

$$\boldsymbol{\varphi}_k(\boldsymbol{\theta}) = \left(1, y_{k-1}^2, \ldots, y_{k-p}^2, \boldsymbol{\varphi}_{k-1}^\top(\boldsymbol{\theta})\boldsymbol{\theta}, \ldots, \boldsymbol{\varphi}_{k-q}^\top(\boldsymbol{\theta})\boldsymbol{\theta}\right)^\top, \tag{5.3a}$$

$$\boldsymbol{\varphi}_k^\top(\boldsymbol{\theta})\boldsymbol{\theta} = \omega + \sum_{i=1}^{p} \alpha_i y_{k-i}^2 + \sum_{j=1}^{q} \beta_j \boldsymbol{\varphi}_{k-j}^\top(\boldsymbol{\theta})\boldsymbol{\theta}. \tag{5.3b}$$

This loss function obviously (almost) coincides with the (negative) conditional log-likelihood function derived assuming the GARCH(p, q) model with normally distributed innovations (Fan & Yao, 2005, Section 4.2). The missing values $\{\boldsymbol{\varphi}_{1-q}^\top(\boldsymbol{\theta})\boldsymbol{\theta}, \ldots, \boldsymbol{\varphi}_0^\top(\boldsymbol{\theta})\boldsymbol{\theta}\}$ starting estimation are usually substituted by appropriate quantities. Refer to Section 5.2.1. Similarly, $\{y_{1-p}^2, \ldots, y_0^2\}$ must be either observed or defined to evaluate initial values of $F_k(\boldsymbol{\theta})$. Nevertheless, the influence of initialization fades away relatively quickly.

Moreover, suppose that $\widehat{\boldsymbol{\theta}}_{t-1}$ minimizes $V_{t-1}(\boldsymbol{\theta})$ and that the minimum point of $V_t(\boldsymbol{\theta})$ is close to $\widehat{\boldsymbol{\theta}}_{t-1}$. Respecting these conditions, it seems reasonable to approximate $V_t(\boldsymbol{\theta})$ by the second-order Taylor expansion around $\widehat{\boldsymbol{\theta}}_{t-1}$ as usual. To be more specific, we can obtain:

$$V_t(\boldsymbol{\theta}) \approx V_t(\widehat{\boldsymbol{\theta}}_{t-1}) + \nabla_{\boldsymbol{\theta}} V_t(\widehat{\boldsymbol{\theta}}_{t-1})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{t-1}) + \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{t-1})^\top \boldsymbol{V}_t''(\widehat{\boldsymbol{\theta}}_{t-1})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_{t-1}), \tag{5.4}$$

where $\nabla_{\boldsymbol{\theta}} V_t(\boldsymbol{\theta})$ and $\boldsymbol{V}_t''(\boldsymbol{\theta})$ stand for the gradient and the Hessian matrix of second-order derivatives of $V_t(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, respectively. The approximation on the right-hand side of (5.4) is a quadratic function of $\boldsymbol{\theta}$, which can be simply minimized with respect to $\boldsymbol{\theta}$. In addition, let the minimum point constitute the new parameter estimate $\widehat{\boldsymbol{\theta}}_t$. Particularly, we obtain:

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1} - \left[\boldsymbol{V}_t''(\widehat{\boldsymbol{\theta}}_{t-1})\right]^{-1} \left[\nabla_{\boldsymbol{\theta}} V_t(\widehat{\boldsymbol{\theta}}_{t-1})\right]^\top. \tag{5.5}$$

Apparently, this formula coincides with one step in the Newton-Raphson algorithm initialized in the point $\widehat{\boldsymbol{\theta}}_{t-1}$.

To evaluate the estimate $\widehat{\boldsymbol{\theta}}_t$ applying (5.5), the gradient and the Hessian matrix of the loss function $V_t(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ must be calculated. The following relations can be simply derived from (5.1):

$$V_t(\boldsymbol{\theta}) = (1-\gamma_t)V_{t-1}(\boldsymbol{\theta}) + \gamma_t F_t(\boldsymbol{\theta}), \tag{5.6a}$$

$$\nabla_{\boldsymbol{\theta}} V_t(\boldsymbol{\theta}) = (1-\gamma_t)\nabla_{\boldsymbol{\theta}} V_{t-1}(\boldsymbol{\theta}) + \gamma_t \nabla_{\boldsymbol{\theta}} F_t(\boldsymbol{\theta}), \tag{5.6b}$$

$$\boldsymbol{V}_t''(\boldsymbol{\theta}) = (1-\gamma_t)\boldsymbol{V}_{t-1}''(\boldsymbol{\theta}) + \gamma_t \boldsymbol{F}_t''(\boldsymbol{\theta}), \tag{5.6c}$$

$$\nabla_{\boldsymbol{\theta}} F_t(\boldsymbol{\theta}) = -\frac{y_t^2 - \boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}}{\left(\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}\right)^2}\boldsymbol{\psi}_t^\top(\boldsymbol{\theta}), \tag{5.6d}$$

$$\boldsymbol{F}_t''(\boldsymbol{\theta}) = \frac{2y_t^2 - \boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}}{\left(\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}\right)^3}\boldsymbol{\psi}_t(\boldsymbol{\theta})\boldsymbol{\psi}_t^\top(\boldsymbol{\theta}) + \frac{\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta} - y_t^2}{\left(\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}\right)^2}\boldsymbol{\Psi}_t(\boldsymbol{\theta}), \tag{5.6e}$$

where $\boldsymbol{\psi}_t^\top(\boldsymbol{\theta})$ and $\boldsymbol{\Psi}_t(\boldsymbol{\theta})$ are the gradient and the Hessian matrix of second-order derivatives of the function $\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}$ with respect to $\boldsymbol{\theta}$, respectively.

In order to transform (5.5) into an effective recursive estimation algorithm, several adjustments must be realized. Firstly, we may put $\nabla_{\boldsymbol{\theta}} V_{t-1}(\widehat{\boldsymbol{\theta}}_{t-1}) = \mathbf{0}$ since $\widehat{\boldsymbol{\theta}}_{t-1}$ is supposed to be the minimum point of $V_{t-1}(\boldsymbol{\theta})$. Secondly, we shall approximate the Hessian matrix $\boldsymbol{V}_t''(\boldsymbol{\theta})$ by an associated matrix denoted as $\boldsymbol{R}_t(\boldsymbol{\theta})$ such that $\mathbb{E}[V_t''(\boldsymbol{\theta}) - \boldsymbol{R}_t(\boldsymbol{\theta})] = \mathbf{0}$ holds for all $t \in \mathbb{N}$ and $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, i.e. for the vector of the true parameters. To be more precise, one can set:

$$\boldsymbol{R}_t(\boldsymbol{\theta}) = (1-\gamma_t)\boldsymbol{R}_{t-1}(\boldsymbol{\theta}) + \gamma_t \widetilde{\boldsymbol{F}}_t''(\boldsymbol{\theta}) \quad \text{with} \quad \widetilde{\boldsymbol{F}}_t''(\boldsymbol{\theta}) = \frac{\boldsymbol{\psi}_t(\boldsymbol{\theta})\boldsymbol{\psi}_t^\top(\boldsymbol{\theta})}{(\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta})^2}. \tag{5.7}$$

Particularly, $\widetilde{\boldsymbol{F}}_t''(\boldsymbol{\theta})$ may replace $\boldsymbol{F}_t''(\boldsymbol{\theta})$ since $\mathbb{E}[\boldsymbol{F}_t''(\boldsymbol{\theta})|\mathcal{F}_{t-1}] = \mathbb{E}[\widetilde{\boldsymbol{F}}_t''(\boldsymbol{\theta})|\mathcal{F}_{t-1}]$ for all $t \in \mathbb{N}$ and $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, i.e. for the vector of the true parameters. Therefore, the evaluation of the second-order derivatives originally included in $\boldsymbol{F}_t''(\boldsymbol{\theta})$ is effectively reduced. It means that the Newton-Raphson algorithm based on the exact Hessian matrix is no longer used. Finally, we shall assume that the supposed approximative matrix $\boldsymbol{R}_t$ varies slowly with $\boldsymbol{\theta}$, i.e. that $\boldsymbol{R}_{t-1}(\widehat{\boldsymbol{\theta}}_{t-1})$ equals $\boldsymbol{R}_{t-1}(\widehat{\boldsymbol{\theta}}_{t-2})$. For further insights, refer to Ljung (1999, Chapter 11) or Söderström and Stoica (1989, Chapter 9).

If we combine the relations in (5.6), (5.7), and the previously suggested modifications together, the consequent recursive formulas can be deduced (using the apparent simplifications in the denotation):

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1} - \gamma_t \boldsymbol{R}_t^{-1}\left[\nabla_{\boldsymbol{\theta}} F_t(\widehat{\boldsymbol{\theta}}_{t-1})\right]^\top, \tag{5.8a}$$

$$\boldsymbol{R}_t = \boldsymbol{R}_{t-1} + \gamma_t\left[\widetilde{\boldsymbol{F}}_t''(\widehat{\boldsymbol{\theta}}_{t-1}) - \boldsymbol{R}_{t-1}\right], \; t \in \mathbb{N}. \tag{5.8b}$$

However, the equations considered in (5.8) are not, in fact, true recursive estimation formulas since the calculation of $F_t(\widehat{\boldsymbol{\theta}}_{t-1})$ and its derivatives for t requires processing all data up to time t. See also Remark 4.2.1. Therefore, we should implement some pragmatic approximations that remove this limitation (Ljung, 1999, Chapter 11). The following relations may be suitable:

$$\boldsymbol{\varphi}_t(\widehat{\boldsymbol{\theta}}_{t-1}) \approx \widehat{\boldsymbol{\varphi}}_t \quad \text{and} \quad \boldsymbol{\psi}_t(\widehat{\boldsymbol{\theta}}_{t-1}) \approx \widehat{\boldsymbol{\psi}}_t, \tag{5.9}$$

where

$$\widehat{\boldsymbol{\varphi}}_t = (1, y_{t-1}^2, \ldots, y_{t-p}^2, \widehat{\boldsymbol{\varphi}}_{t-1}^\top \widehat{\boldsymbol{\theta}}_{t-1}, \ldots, \widehat{\boldsymbol{\varphi}}_{t-q}^\top \widehat{\boldsymbol{\theta}}_{t-q})^\top, \tag{5.10a}$$

$$\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1} = \widehat{\omega}_{t-1} + \sum_{i=1}^{p} \widehat{\alpha}_{i,t-1} y_{t-i}^2 + \sum_{j=1}^{q} \widehat{\beta}_{j,t-1} \widehat{\boldsymbol{\varphi}}_{t-j}^\top \widehat{\boldsymbol{\theta}}_{t-j}, \tag{5.10b}$$

$$\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_t = \widehat{\omega}_t + \sum_{i=1}^{p} \widehat{\alpha}_{i,t} y_{t-i}^2 + \sum_{j=1}^{q} \widehat{\beta}_{j,t} \widehat{\boldsymbol{\varphi}}_{t-j}^\top \widehat{\boldsymbol{\theta}}_{t-j}, \tag{5.10c}$$

and

$$(\widehat{\boldsymbol{\psi}}_t)_k := \begin{cases} 1 + \sum_{j=1}^{q} \widehat{\beta}_{j,t-1} (\widehat{\boldsymbol{\psi}}_{t-j})_k, & k = 1, \\ y_{t-(k-1)}^2 + \sum_{j=1}^{q} \widehat{\beta}_{j,t-1} (\widehat{\boldsymbol{\psi}}_{t-j})_k, & k = 2, \ldots, p+1, \\ (\widehat{\boldsymbol{\varphi}}_t)_k + \sum_{j=1}^{q} \widehat{\beta}_{j,t-1} (\widehat{\boldsymbol{\psi}}_{t-j})_k, & k = p+2, \ldots, p+q+1. \end{cases} \tag{5.11}$$

Equation (5.11) approximating the gradient can be written more concisely as:

$$\widehat{\boldsymbol{\psi}}_t = \widehat{\boldsymbol{\varphi}}_t + \sum_{j=1}^{q} \widehat{\beta}_{j,t-1} \widehat{\boldsymbol{\psi}}_{t-j}. \tag{5.12}$$

Consequently, the preceding discussion results in the recursive algorithm appropriate for estimating the parameters of the standard GARCH(p, q) model. This algorithm has the following form:

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1} + \gamma_t \boldsymbol{R}_t^{-1} \widehat{\boldsymbol{\psi}}_t \frac{y_t^2 - \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1}}{(\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})^2}, \tag{5.13a}$$

$$\boldsymbol{R}_t = \boldsymbol{R}_{t-1} + \gamma_t \left\{ \frac{\widehat{\boldsymbol{\psi}}_t \widehat{\boldsymbol{\psi}}_t^\top}{(\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})^2} - \boldsymbol{R}_{t-1} \right\}, \tag{5.13b}$$

$$\widehat{\boldsymbol{\varphi}}_{t+1} = (1, y_t^2, \ldots, y_{t+1-p}^2, \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_t, \ldots, \widehat{\boldsymbol{\varphi}}_{t+1-q}^\top \widehat{\boldsymbol{\theta}}_{t+1-q})^\top, \tag{5.13c}$$

$$\widehat{\boldsymbol{\psi}}_{t+1} = \widehat{\boldsymbol{\varphi}}_{t+1} + \sum_{j=1}^{q} \widehat{\beta}_{j,t} \widehat{\boldsymbol{\psi}}_{t+1-j}, \; t \in \mathbb{N}. \tag{5.13d}$$

Moreover, if we put $\widehat{\boldsymbol{P}}_t = \gamma_t \boldsymbol{R}_t^{-1}$ (for the definition of λ_t and γ_t, consult (5.1) and Section 5.2.2) and employ Lemma 4.2.1, we shall obtain after rearrangements a computationally more efficient version of the above formulas:

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1} + \frac{\widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\psi}}_t(y_t^2 - \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})}{\lambda_t(\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})^2 + \widehat{\boldsymbol{\psi}}_t^\top \widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\psi}}_t}, \tag{5.14a}$$

$$\widehat{\boldsymbol{P}}_t = \frac{1}{\lambda_t}\left\{\widehat{\boldsymbol{P}}_{t-1} - \frac{\widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\psi}}_t\widehat{\boldsymbol{\psi}}_t^\top \widehat{\boldsymbol{P}}_{t-1}}{\lambda_t(\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})^2 + \widehat{\boldsymbol{\psi}}_t^\top \widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\psi}}_t}\right\}, \tag{5.14b}$$

$$\widehat{\boldsymbol{\varphi}}_{t+1} = (1, y_t^2, \ldots, y_{t+1-p}^2, \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_t, \ldots, \widehat{\boldsymbol{\varphi}}_{t+1-q}^\top \widehat{\boldsymbol{\theta}}_{t+1-q})^\top, \tag{5.14c}$$

$$\widehat{\boldsymbol{\psi}}_{t+1} = \widehat{\boldsymbol{\varphi}}_{t+1} + \sum_{j=1}^{q} \widehat{\beta}_{j,t}\widehat{\boldsymbol{\psi}}_{t+1-j},\ t \in \mathbb{N}. \tag{5.14d}$$

Theoretical analysis of the proposed algorithm can employ the general schemes considered by Ljung and Söderström (1983, Chapter 4); it is rather technical and uses instruments known mainly from the ordinary differential equation theory. Under the corresponding (mostly technical) assumptions, it can be shown that the recursive and the off-line estimates originated from the identical criterion have similar convergence features. Additionally, they are also asymptotically normally distributed (Ljung & Söderström, 1983, Chapter 4). See also Section 5.3.3.

At first sight, the estimation algorithm (5.14) seems to be related to those outlined in Chapter 4. There undoubtedly exist evident connections between these methods. If we simplify the formulas in (5.14) just assuming $\lambda_t \equiv 1$ for all t, we will receive almost similar estimation algorithms. Thus, it may be concluded that the suggested method is, in fact, a *one-stage self-weighted alternative* to the recursive prediction error techniques presented in Chapter 4. Comparing all these recursive schemes mutually, one can summarize that the discussed methods are analogous but operate with the different weights in the denominators of the updating terms and also with the distinct definitions of $\widetilde{\boldsymbol{\varphi}}_t$ and $\widehat{\boldsymbol{\varphi}}_t$ (compare the methods (4.7), (4.11), and (5.10)).

In detail, the proposed self-weighted algorithm (5.14) is connected: (i) to the non-weighted technique (4.8) considered by Kierkegaard et al. (2000), (ii) to both, i.e. the first (non-weighted) and second (weighted), stages of the algorithm (4.12)-(4.13) introduced by Aknouche and Guerbyenne (2006). The modifier "self-weighted" appearing in the name of the suggested method reflects these differences. Note that the nuance between the two definitions of $\widehat{\boldsymbol{\varphi}}_t$ (and $\widetilde{\boldsymbol{\varphi}}_t$) can be omitted since one can easily adapt the algorithms from Chapter 4 to respect (5.10). Additionally, it holds for a large t that $\widetilde{\boldsymbol{\theta}}_{t-1} \approx \widetilde{\boldsymbol{\theta}}_{t-2}$ and $\widehat{\boldsymbol{\theta}}_{t-1} \approx \widehat{\boldsymbol{\theta}}_{t-2}$. It should be also pointed out that the suggested simplified estimation formulas coincide with the stochastic Newton approximation scheme given by Gerencsér,

Orlovits and Torma (2010). Additionally, Gerencsér and Orlovits (2012) have earlier analysed the convergence properties applying stochastic approximations.

From the practical and theoretical viewpoint, it might be interesting to study the self-weighted version of the recursive pseudo-linear regression derived directly from (5.13) by using the obvious approximation $\widehat{\boldsymbol{\psi}}_t \approx \widehat{\boldsymbol{\varphi}}_t$, i.e.:

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1} + \gamma_t \boldsymbol{R}_t^{-1} \widehat{\boldsymbol{\varphi}}_t \frac{y_t^2 - \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1}}{(\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})^2}, \tag{5.15a}$$

$$\boldsymbol{R}_t = \boldsymbol{R}_{t-1} + \gamma_t \left\{ \frac{\widehat{\boldsymbol{\varphi}}_t \widehat{\boldsymbol{\varphi}}_t^\top}{(\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})^2} - \boldsymbol{R}_{t-1} \right\}, \tag{5.15b}$$

$$\widehat{\boldsymbol{\varphi}}_{t+1} = (1, y_t^2, \ldots, y_{t+1-p}^2, \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_t, \ldots, \widehat{\boldsymbol{\varphi}}_{t+1-q}^\top \widehat{\boldsymbol{\theta}}_{t+1-q})^\top, \ t \in \mathbb{N}. \tag{5.15c}$$

Alternatively, by putting $\widehat{\boldsymbol{P}}_t = \gamma_t \boldsymbol{R}_t^{-1}$ and applying Lemma 4.2.1, one obtains:

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1} + \frac{\widehat{\boldsymbol{P}}_{t-1} \widehat{\boldsymbol{\varphi}}_t (y_t^2 - \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})}{\lambda_t (\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})^2 + \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{P}}_{t-1} \widehat{\boldsymbol{\varphi}}_t}, \tag{5.16a}$$

$$\widehat{\boldsymbol{P}}_t = \frac{1}{\lambda_t} \left\{ \widehat{\boldsymbol{P}}_{t-1} - \frac{\widehat{\boldsymbol{P}}_{t-1} \widehat{\boldsymbol{\varphi}}_t \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{P}}_{t-1}}{\lambda_t (\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})^2 + \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{P}}_{t-1} \widehat{\boldsymbol{\varphi}}_t} \right\}, \tag{5.16b}$$

$$\widehat{\boldsymbol{\varphi}}_{t+1} = (1, y_t^2, \ldots, y_{t+1-p}^2, \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_t, \ldots, \widehat{\boldsymbol{\varphi}}_{t+1-q}^\top \widehat{\boldsymbol{\theta}}_{t+1-q})^\top, \ t \in \mathbb{N}. \tag{5.16c}$$

It might be an acceptable numerical alternative to the recursive pseudo-linear regression algorithms introduced in Chapter 4. Similarly as before, all presented recursive pseudo-linear regression estimation procedures are closely related assuming $\lambda_t \equiv 1$ for all t but operate with the different weighting terms and definitions of $\widetilde{\boldsymbol{\varphi}}_t$ and $\widehat{\boldsymbol{\varphi}}_t$. Compare (4.6), (4.9)-(4.10), and simplified (5.16).

Remark 5.1.1 It should be emphasized that the only difference between the recursive pseudo-linear regression method , e.g. (5.16), compared to the recursive prediction error technique, e.g. (5.14), consists in the fact that the approximated gradient $\widehat{\boldsymbol{\psi}}_t$ is substituted by using $\widehat{\boldsymbol{\varphi}}_t$. It is clear that this simplification reduces the algorithm complexity since the gradient $\widehat{\boldsymbol{\psi}}_t$ need not to be evaluated and monitored. However, it necessarily results also in changes of: (i) convergence properties, (ii) asymptotic accuracy, and (iii) transient behaviour.

Firstly, the estimates delivered by the recursive prediction error method converge under more general assumptions and are asymptotically statistically efficient (Ljung & Söderström, 1983, Chapter 4). The key attribute leading to the consistency of the recursive pseudo-linear regression procedure is that a certain filter supporting the mentioned substitution is positive real. This condition is not automatically satisfied. See Ljung and Söderström (1983, Chapter 4) and Section 5.3.4 for further insights.

Secondly, the more complex algorithm, e.g. (5.14), is superior to the simplified one, e.g. (5.16), from the asymptotic accuracy point of view (Ljung & Söderström, 1983, Section 5.9). See experimental results introduced in Section 6.1.

Thirdly, despite the outlined facts, sometimes the pseudo-linear regression technique gives better results for short samples. It might be explained as follows. Processing only a limited number of measurements, the gradient $\widehat{\boldsymbol{\psi}}_t$ is filtered rather poor due to its complex structure (see e.g. the third formula in (5.14)); therefore, the accepted substitution might be more appropriate for short data. Consult Section 5.3.4.

Remark 5.1.2 It is evident that all algorithms considered in Chapter 4, namely (4.6), (4.8), (4.9)-(4.10), and (4.12)-(4.13), can be easily modified to involve the forgetting factor $\{\lambda_t\}_{t\in\mathbb{N}}$ adjusting the convergence speed. Compare with (5.14) or (5.16). See also Section 5.2.2 for further insights.

Remark 5.1.3 The recursive estimation algorithm (5.13) also has an associated off-line counterpart, which can be figured out by applying analogous arguments as before. For simplicity, assume that the sum $\sum_{k=1}^{t} F_k(\boldsymbol{\theta})$ with the previously defined criterion $F_k(\boldsymbol{\theta})$ is supposed to be minimized. It obviously (almost) corresponds to the common negative conditional log-likelihood criterion for the GARCH(p, q) process. The iteration method used in the case of the off-line prediction error estimation is a variant of the gradient-type nonlinear programming scheme. It can be formulated as follows:

$$\widehat{\boldsymbol{\theta}}_{(i+1)} = \widehat{\boldsymbol{\theta}}_{(i)} - \rho_i \boldsymbol{M}_t^{-1}(\widehat{\boldsymbol{\theta}}_{(i)}) \sum_{k=1}^{t} \nabla_{\boldsymbol{\theta}} F_k(\widehat{\boldsymbol{\theta}}_{(i)}),\ i \in \mathbb{N}, \tag{5.17}$$

where $\widehat{\boldsymbol{\theta}}_{(i)}$ is the parameter vector estimate at the ith iteration, $\widehat{\boldsymbol{\theta}}_{(0)}$ is a fixed initial value, ρ_i is a step size parameter accelerating the convergence (initially $\rho_i > 1$, after that $\rho_i \to 1$ as $i \to \infty$), and $\boldsymbol{M}_t(\widehat{\boldsymbol{\theta}}_{(i)})$ is the matrix with the following attractive properties: (i) $\boldsymbol{M}_t(\widehat{\boldsymbol{\theta}}_{(i)})$ is positive definite for all i, (ii) $\mathbb{E}(\boldsymbol{M}_t(\widehat{\boldsymbol{\theta}}_{(i)})) = \mathbb{E}(\sum_{k=1}^{t} \boldsymbol{F}''_k(\widehat{\boldsymbol{\theta}}_{(i)}))$ for all i, (iii) $1/t[\boldsymbol{M}_t(\widehat{\boldsymbol{\theta}}_{(i)}) - \sum_{k=1}^{t} \boldsymbol{F}''_k(\widehat{\boldsymbol{\theta}}_{(i)})] \to \mathbf{0}$ for all i almost surely as $t \to \infty$.

Moore and Weiss (1979) recommended and clarified the choice $\boldsymbol{M}_t(\widehat{\boldsymbol{\theta}}_{(i)}) = \sum_{k=1}^{t} \widetilde{\boldsymbol{F}}''_k(\widehat{\boldsymbol{\theta}}_{(i)})$ (compare with (5.7)). See Moore and Weiss (1979) for more details and the references given therein. Apparently, the iteration scheme (5.17) with this particular matrix $\boldsymbol{M}_t$ is closely related to Equations (5.8).

Remark 5.1.4 The self-weighted sequential estimation algorithm (5.14) can be regarded as a reduced (i.e. approximated) version of the extended Kalman filter. See also Remark 4.2.1 for further insights and the references. In this instance,

a more general nonlinear state space modelling representation must be naturally supposed (Roth & Gustafsson, 2011), i.e.:

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1}, \tag{5.18a}$$

$$y_t^2 = \boldsymbol{\varphi}_t^\top(\boldsymbol{\theta}_t)\boldsymbol{\theta}_t[1 + e_t], \tag{5.18b}$$

where the notation is obvious and $t \in \mathbb{N}$. The disturbance terms $\{e_t\}_{t\in\mathbb{N}}$ are assumed to be i. i. d. random variables with the normal distribution $N(0, \sigma_e^2)$, $\sigma_e^2 \in (0, \infty)$. The initial state vector $\boldsymbol{\theta}_0$ is supposed to be $N(\widehat{\boldsymbol{\theta}}_0, \widehat{\boldsymbol{P}}_0)$ independently of $e_1, e_2, \ldots$, where both hatted quantities are assumed to be known. Similarly as in Remark 4.2.1, the state equation (5.18a) depicts the path of the state vectors, i.e. of the estimated parameters. The signal (measurement) equation (5.18b) represents only a straightforward reorganization of Equation (4.2), where the disturbances $e_t = \varepsilon_t^2 - 1$ are assumed to be normally distributed. Here, the model respects the heteroscedastic structure of the error terms from (4.2). The function $\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})$ readily depends on $\boldsymbol{\theta}$ and the past measurements.

To derive the extended Kalman filtering formulas associated with the state space system (5.18), the nonlinear function $h_t(\boldsymbol{\delta}) = \boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}[1 + e]$ with $\boldsymbol{\delta} = (\boldsymbol{\theta}^\top, e)^\top$ must be linearized applying the first-order Taylor expansion around $\widehat{\boldsymbol{\delta}} = (\widehat{\boldsymbol{\theta}}^\top, \widehat{e})^\top$. Namely, one may use $h_t(\boldsymbol{\delta}) \approx h_t(\widehat{\boldsymbol{\delta}}) + \nabla_{\boldsymbol{\delta}} h_t(\widehat{\boldsymbol{\delta}})(\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}})$, where $\nabla_{\boldsymbol{\delta}} h_t(\widehat{\boldsymbol{\delta}}) = (\nabla_{\boldsymbol{\theta}} h_t(\widehat{\boldsymbol{\delta}}), \nabla_e h_t(\widehat{\boldsymbol{\delta}}))$, $\nabla_{\boldsymbol{\theta}} h_t^\top(\widehat{\boldsymbol{\delta}}) = \boldsymbol{\psi}_t(\widehat{\boldsymbol{\theta}})[1 + \widehat{e}]$, and $\nabla_e h_t^\top(\widehat{\boldsymbol{\delta}}) = \boldsymbol{\varphi}_t^\top(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\theta}}$, respectively. The notation has been explained earlier. Consult also Roth and Gustafsson (2011) for further insights.

Define $\widehat{\boldsymbol{\theta}}_{t|t-1} := \mathbb{E}(\boldsymbol{\theta}_t | \boldsymbol{Y}_{t-1}^2)$ and $\widehat{\boldsymbol{P}}_{t|t-1} := \mathsf{cov}(\boldsymbol{\theta}_t | \boldsymbol{Y}_{t-1}^2)$, where $\boldsymbol{Y}_t^2$ is the collection of all observations up to and including time t. According to (5.18a), it clearly holds that $\widehat{\boldsymbol{\theta}}_{t|t-1} = \widehat{\boldsymbol{\theta}}_{t-1|t-1}$ and $\widehat{\boldsymbol{P}}_{t|t-1} = \widehat{\boldsymbol{P}}_{t-1|t-1}$ using the obvious notation. Employing the linearization of the function $h_t(\boldsymbol{\delta})$ around $\widehat{\boldsymbol{\delta}}_{t|t-1} := \mathbb{E}((\boldsymbol{\theta}_t^\top, e_t)^\top | \boldsymbol{Y}_{t-1}^2) = (\widehat{\boldsymbol{\theta}}_{t|t-1}^\top, 0)^\top$, one obtains:

$$\mathbb{E}(y_t^2 | \boldsymbol{Y}_{t-1}^2) \approx \boldsymbol{\varphi}_t^\top(\widehat{\boldsymbol{\theta}}_{t|t-1})\widehat{\boldsymbol{\theta}}_{t|t-1}, \tag{5.19a}$$

$$\begin{aligned}\mathsf{cov}(y_t^2 | \boldsymbol{Y}_{t-1}^2) \approx{}& \sigma_e^2(\boldsymbol{\varphi}_t^\top(\widehat{\boldsymbol{\theta}}_{t|t-1})\widehat{\boldsymbol{\theta}}_{t|t-1})^2 \\ &+ \boldsymbol{\psi}_t^\top(\widehat{\boldsymbol{\theta}}_{t|t-1})\widehat{\boldsymbol{P}}_{t|t-1}\boldsymbol{\psi}_t(\widehat{\boldsymbol{\theta}}_{t|t-1}) =: \widehat{\boldsymbol{S}}_{t|t-1},\end{aligned} \tag{5.19b}$$

$$\mathsf{cov}(\boldsymbol{\theta}_t, y_t^2 | \boldsymbol{Y}_{t-1}^2) \approx \widehat{\boldsymbol{P}}_{t|t-1}\boldsymbol{\psi}_t(\widehat{\boldsymbol{\theta}}_{t|t-1}) =: \widehat{\boldsymbol{M}}_{t|t-1}. \tag{5.19c}$$

The Kalman filtering scheme can be derived under the condition that $\boldsymbol{\theta}_t$ and y_t^2 are jointly conditionally normally distributed given $\boldsymbol{Y}_{t-1}^2$. Particularly, assume that $(\boldsymbol{\theta}_t^\top, y_t^2)^\top | \boldsymbol{Y}_{t-1}^2$ has the normal distribution with the following specification

respecting the approximations from Equations (5.19):

$$N\left(\begin{pmatrix} \widehat{\boldsymbol{\theta}}_{t|t-1} \\ \boldsymbol{\varphi}_t^\top(\widehat{\boldsymbol{\theta}}_{t|t-1})\widehat{\boldsymbol{\theta}}_{t|t-1} \end{pmatrix}, \begin{pmatrix} \widehat{\boldsymbol{P}}_{t|t-1} & \widehat{\boldsymbol{M}}_{t|t-1} \\ \widehat{\boldsymbol{M}}_{t|t-1}^\top & \widehat{\boldsymbol{S}}_{t|t-1} \end{pmatrix}\right). \tag{5.20}$$

Using the lemma known from optimal estimation theory (Durbin & Koopman, 2001, Section 2.13) and putting $\widehat{\boldsymbol{\theta}}_t := \widehat{\boldsymbol{\theta}}_{t|t}$ and $\widehat{\boldsymbol{P}}_t := \widehat{\boldsymbol{P}}_{t|t}/\sigma_e^2$, $t \in \mathbb{N}$, one obtains:

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1} + \frac{\widehat{\boldsymbol{P}}_{t-1}\boldsymbol{\psi}_t(\widehat{\boldsymbol{\theta}}_{t-1})(y_t^2 - \boldsymbol{\varphi}_t^\top(\widehat{\boldsymbol{\theta}}_{t-1})\widehat{\boldsymbol{\theta}}_{t-1})}{(\boldsymbol{\varphi}_t^\top(\widehat{\boldsymbol{\theta}}_{t-1})\widehat{\boldsymbol{\theta}}_{t-1})^2 + \boldsymbol{\psi}_t^\top(\widehat{\boldsymbol{\theta}}_{t-1})\widehat{\boldsymbol{P}}_{t-1}\boldsymbol{\psi}_t(\widehat{\boldsymbol{\theta}}_{t-1})}, \tag{5.21a}$$

$$\widehat{\boldsymbol{P}}_t = \widehat{\boldsymbol{P}}_{t-1} - \frac{\widehat{\boldsymbol{P}}_{t-1}\boldsymbol{\psi}_t(\widehat{\boldsymbol{\theta}}_{t-1})\boldsymbol{\psi}_t^\top(\widehat{\boldsymbol{\theta}}_{t-1})\widehat{\boldsymbol{P}}_{t-1}}{(\boldsymbol{\varphi}_t^\top(\widehat{\boldsymbol{\theta}}_{t-1})\widehat{\boldsymbol{\theta}}_{t-1})^2 + \boldsymbol{\psi}_t^\top(\widehat{\boldsymbol{\theta}}_{t-1})\widehat{\boldsymbol{P}}_{t-1}\boldsymbol{\psi}_t(\widehat{\boldsymbol{\theta}}_{t-1})}. \tag{5.21b}$$

Apparently, the algorithm (5.21) is linked to the method (5.14). Nevertheless, to provide a fully recursive method, further approximations must be performed since $\boldsymbol{\varphi}_t^\top(\widehat{\boldsymbol{\theta}}_{t-1})$ and $\boldsymbol{\psi}_t^\top(\widehat{\boldsymbol{\theta}}_{t-1})$ are not evaluated recurrently. Their calculation at each time requires processing all data up to time t. Therefore, it seems reasonable to introduce repeatedly the pragmatic simplifications discussed before, i.e. $\boldsymbol{\varphi}_t^\top(\widehat{\boldsymbol{\theta}}_{t-1}) \approx \widehat{\boldsymbol{\varphi}}_t$ and $\boldsymbol{\psi}_t^\top(\widehat{\boldsymbol{\theta}}_{t-1}) \approx \widehat{\boldsymbol{\psi}}_t$ as in (5.9). They transform (5.21) into the estimation procedure that is identical to (5.14) (by putting $\lambda_t \equiv 1$ for all t).

5.2 Self-weighted recursive estimation methods for GARCH models: Practical aspects

Before studying theoretical properties of the estimation algorithms suggested in Section 5.1, we shall clarify several practical aspects necessary for the successful routine implementation of the considered on-line calibration techniques. From the practical point of view, one should primarily address attention to: (i) initialization of both methods, (ii) the accelerating factors γ_t and λ_t controlling convergence speed, (iii) regularization and factorization of the algorithms, and finally (iv) projections ensuring stability (especially, the positivity and stationarity of the conditional variance). Additionally, we shall reflect two other issues that are undoubtedly relevant from this perspective. In particular, the questions of model identification (i.e. determining the model order (p, q)) and algorithm robustification (i.e. reducing the deleterious influence of exceptional measurements) are discussed in greater detail.

5.2.1 Initial conditions

To employ the recursive estimation algorithms (5.14) and (5.16), initial conditions are required for their start-up. The initial value of $\widehat{\boldsymbol{\theta}}_0$ should lie in a stability region. This set can be determined by the corresponding sufficient conditions for the positivity and stationarity of the conditional variance. See its definition in Section 5.2.4. Otherwise, the algorithms could be either decelerated or even degenerated (according to the accepted projection (5.36)). Compare various results of Monte Carlo experiments presented in Section 6.1. To start the proposed recursive estimation formulas (5.14) and (5.16), we recommend the initialization inspired by the `R` package `tseries` (Trapletti & Hornik, 2013). In particular, the first n observations of the GARCH process are used as a burn-in. Namely, one can put $\widehat{\boldsymbol{\theta}}_0 = (\frac{1}{n}\sum_{i=1}^{n} y_i^2[1-(p+q)\varepsilon], \varepsilon, \ldots, \varepsilon)^\top$, where ε is a small positive real number such that $(p+q)\varepsilon < 1$, e.g. 0.05 or 0.1 for $p = q = 1$, and the choice of n should obviously reflect the frequency of observations. The estimation evidently begins after collecting n measurements, i.e. at time $n+1$.

Alternative starting configurations might be considered. They can be delivered using autocorrelation functions, given by experts, etc. Clearly, any accepted choice should respect specifics of on-line estimation. On the other hand, the initialization introduced by Aknouche and Guerbyenne (2006), i.e. $\widetilde{\boldsymbol{\theta}}_0 = \mathbf{0}$ and $\widehat{\boldsymbol{\theta}}_0 = \mathbf{0}$, does not seem suitable in this framework. See Hendrych (2014b) or the results presented in Section 6.1.

$\widehat{\boldsymbol{P}}_0$ is usually set as $\kappa\boldsymbol{I}$ with a positive κ. If κ is small, then the parameter estimates $\widehat{\boldsymbol{\theta}}_t$ will not change too much from $\widehat{\boldsymbol{\theta}}_0$. If it is large enough, e.g. $\kappa = 10^5$, the parameter estimates $\widehat{\boldsymbol{\theta}}_t$ will remote from $\widehat{\boldsymbol{\theta}}_0$ relatively quickly. This particular choice (i.e. $\kappa = 10^5$) associated with no prior information has been also suggested by Aknouche and Guerbyenne (2006). If there exists some prior knowledge about the estimated parameters, the positive elements of diagonal matrix $\widehat{\boldsymbol{P}}_0$ should respect it according to Söderström and Stoica (1989, Chapter 9).

The first q recurrent approximations of the transposed gradient from (5.13d), i.e. $\widehat{\boldsymbol{\psi}}_1, \ldots, \widehat{\boldsymbol{\psi}}_{2-q}$, are chosen as $\widehat{\boldsymbol{\psi}}_1 = \widehat{\boldsymbol{\varphi}}_1$ and $\widehat{\boldsymbol{\psi}}_j = \mathbf{0}$, $j < 1$, since no prior information is available. Finally, the missing elements composing $\widehat{\boldsymbol{\varphi}}_1, \ldots, \widehat{\boldsymbol{\varphi}}_{\max(p,q)}$ must be concretized. Namely, $\{y_{1-p}^2, \ldots, y_0^2\}$ can be either observed or defined, e.g. as a sequence of zeroes. Moreover, the values $\{\widehat{\boldsymbol{\varphi}}_{1-q}^\top \widehat{\boldsymbol{\theta}}_{1-q}, \ldots, \widehat{\boldsymbol{\varphi}}_0^\top \widehat{\boldsymbol{\theta}}_0\}$ are required. These can be quantified in many various ways, e.g. by freezing (i.e. by assuming that all are zeros or small positive numbers), by backcasting as it is usual in some software products (EViews, 2013), or by substituting by suitable sample characteristics. From the sequential estimation viewpoint, it seems the most convenient to use the first option, i.e. the freezing method that puts $\widehat{\boldsymbol{\varphi}}_j^\top \widehat{\boldsymbol{\theta}}_j = c^2$

for all $j = 1 - q, \ldots, 0$ and some small $c \geq 0$. However, it is noteworthy that the choice of these initial conditions is not crucial (Bollerslev, 1986).

5.2.2 Accelerating factors

Furthermore, we should precise the role of the weighting sequences $\{\gamma_t\}_{t\in\mathbb{N}}$ and $\{\lambda_t\}_{t\in\mathbb{N}}$ from the loss function (5.1). The following relations hold:

$$\gamma_1 = 1 \text{ and } \gamma_t = \frac{\gamma_{t-1}}{\gamma_{t-1} + \lambda_t},\ t \geq 2. \tag{5.22}$$

Estimating time-invariant parameters, the *gain sequence* $\{\gamma_t\}_{t\in\mathbb{N}}$ should be selected asymptotically as $1/t$ in regarding to minimal achievable variance of estimates (Ljung & Söderström, 1983, Section 5.6). Nevertheless, in the transient phase (i.e. for small or medium values of t), the gain sequence $\{\gamma_t\}_{t\in\mathbb{N}}$ or equivalently the *forgetting factors* $\{\lambda_t\}_{t\in\mathbb{N}}$ can be chosen in a suitable way to improve significantly the convergence rate. The motivation for applying the gain sequence or forgetting factors is readily apparent from (5.1). Early information could be somewhat misused during the transient phase of estimation and should therefore carry a lower weight in the loss criterion (5.1) compared to later measurements, which are processed in a better way (Ljung, 1999, Chapter 11).

On the one hand, discounting old data requires $\lambda_t < 1$, which obviously corresponds to $\gamma_t > 1/t$. On the other hand, it is preferable to let $\lambda_t \to 1$ as $t \to \infty$, i.e. $t\gamma_t \to 1$ as $t \to \infty$. In order to fulfil these objectives, let $\{\lambda_t\}_{t\in\mathbb{N}}$ grow to 1 exponentially with t as (Ljung & Söderström, 1983, Section 5.6):

$$\lambda_t = \widetilde{\lambda}\lambda_{t-1} + (1 - \widetilde{\lambda}),\ t \in \mathbb{N}, \tag{5.23}$$

where the rate $\widetilde{\lambda}$ and the initial value λ_0 are the tuning variables. In practice, the choice $\widetilde{\lambda} = 0.99$ and $\lambda_0 = 0.95$ has shown to be useful. In the case of a higher order model, it is recommended to let $\{\lambda_t\}_{t\in\mathbb{N}}$ increase more slowly to 1.

Remark 5.2.1 An important aspect of on-line estimation algorithms is their ability to track time-varying parameters. A one way of achieving this is to adapt the forgetting factors or the gain sequence appropriately. It is obviously related to a trade-off between noise sensitivity and tracking ability. Slowly varying parameters can be often followed reasonably well. On the contrary, it is almost impossible to estimate fast changing parameters (Ljung, 1999, Section 11.6).

To reflect time-varying parameters by the recursive method (5.14) and others, the most common choice is to take a constant forgetting factor, i.e.:

$$\lambda_t \equiv \bar{\lambda}\ \in (0, 1),\ t \in \mathbb{N}, \tag{5.24}$$

or, alternatively, the gain sequence in the form (see (5.22)):

$$\gamma_t = \frac{1-\bar{\lambda}}{1-\bar{\lambda}^t},\ t \in \mathbb{N}. \tag{5.25}$$

Thus, the loss function (5.1) is transformed into:

$$V_t(\boldsymbol{\theta}) = \frac{1-\bar{\lambda}}{1-\bar{\lambda}^t} \sum_{k=1}^{t} \bar{\lambda}^{t-k} F_k(\boldsymbol{\theta}),\ t \in \mathbb{N},\ \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^{p+q+1}. \tag{5.26}$$

Old criterion values thus contribute only marginally to the loss function. Concretely, we have:

$$\bar{\lambda}^{t-k} = e^{(t-k)\log(\bar{\lambda})} \approx e^{-(t-k)(1-\bar{\lambda})}. \tag{5.27}$$

It means that the measurements older than $T_0 = 1/(1-\bar{\lambda})$ time units have weights that are less than $e^{-1} \approx 36\%$ of those for the most recent observation. Note that $\gamma_t \to 1/T_0$ as $t \to \infty$. If the system stays approximately constant over T_0 samples, a suitable choice of $\bar{\lambda}$ can be deduced from it. Typically, it is between 0.980 and 0.995 (Ljung, 1999, Section 11.6). See also Section 6.4.2 for results of Monte Carlo experiments.

5.2.3 Regularization and factorization

Although the sequential algorithm (5.14) looks straightforward, some numerical problems may occur during practical computations. However, they can be successfully resolved applying two distinct numerical instruments: (i) regularization and (ii) factorization.

In particular, Equation (5.13b) has introduced the formula for the recursive evaluation of $\boldsymbol{R}_t$. Nevertheless, it may happen that the matrix $\boldsymbol{R}_t$ is (nearly) singular, i.e. $\det(\boldsymbol{R}_t^{-1})$ is either large or goes to infinity as t grows (such matrices are sometimes called *ill-conditioned*). In that case, computing of $\widehat{\boldsymbol{P}}_t = \gamma_t \boldsymbol{R}_t^{-1}$ faces serious numerical difficulties. Consequently, it is necessary to assure that the elements of $\boldsymbol{R}_t^{-1}$ remain bounded, i.e. $\det(\boldsymbol{R}_t^{-1}) < C$ for some $C > 0$. If $\boldsymbol{R}_t \geq \delta \boldsymbol{I}$ for some $\delta > 0$ (i.e. if $\boldsymbol{R}_t - \delta \boldsymbol{I}$ is positive semidefinite), this requirement will be fulfilled. A technique that guarantees the latter condition is called *regularization* (Ljung & Söderström, 1983, Section 6.5).

The matrix $\boldsymbol{R}_t$ presented in (5.13b) is by construction symmetric positive semidefinite. If we add the term $\gamma_t \delta \boldsymbol{I}$ for some $\delta > 0$ to the right-hand side of the equation, i.e.:

$$\boldsymbol{R}_t = \boldsymbol{R}_{t-1} + \gamma_t \left\{ \frac{\widehat{\boldsymbol{\psi}}_t \widehat{\boldsymbol{\psi}}_t^\top}{(\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})^2} + \delta \boldsymbol{I} - \boldsymbol{R}_{t-1} \right\},\ t \in \mathbb{N}, \tag{5.28}$$

$\boldsymbol{R}_t$ will be positive definite, and $\boldsymbol{R}_t \geq \delta \boldsymbol{I}$ for all $t \in \mathbb{N}_0$ when $\boldsymbol{R}_0 \geq \delta \boldsymbol{I}$. This evident analogy of the Levenberg-Marquardt regularization is conceptually simple. The positive δ should be selected small compared to the magnitude of the elements $\widehat{\boldsymbol{\psi}}_t$, say $\delta = 10^{-4}$ (Ljung & Söderström, 1983, Section 6.5). Therefore, the search directions formulated by two distinct definitions of $\boldsymbol{R}_t$ in (5.13b) and (5.28) differ only slightly. Nonetheless, there is a disadvantage of this concept, which is obvious from applying Lemma 4.2.1. The adjusted estimation procedure involves inversion of $(P \times P)$ matrix in each time step, $P = p + q + 1$. This drawback, however, can be partly subdued (Ljung & Söderström, 1983, Section 6.5).

Alternatively, the regularization can be effectively incorporated into *factorization* of $\widehat{\boldsymbol{P}}_t$. Particularly, the evaluation of $\widehat{\boldsymbol{P}}_t$ as in (5.14b) is sensitive to round-off errors, which may gradually accumulate and make $\widehat{\boldsymbol{P}}_t$ indefinite. Therefore, it seems useful to represent $\widehat{\boldsymbol{P}}_t$ in a factorized form to treat with better conditioned matrices (Ljung, 1999, Section 11.7). For instance, we shall assume the Cholesky decomposition of $\widehat{\boldsymbol{P}}_t$, i.e. $\widehat{\boldsymbol{P}}_t = \widehat{\boldsymbol{Q}}_t \widehat{\boldsymbol{Q}}_t^{\top}$ with a lower triangular matrix $\widehat{\boldsymbol{Q}}_t$. The general factorization algorithm (Ljung, 1999, Section 11.7) can be modified for the case of (5.14) as follows:

Step 1. At $t = 0$, initialize the Cholesky decomposition of $\widehat{\boldsymbol{P}}_0$: $\widehat{\boldsymbol{P}}_0 = \widehat{\boldsymbol{Q}}_0 \widehat{\boldsymbol{Q}}_0^{\top}$.

Step 2. At each time $t \in \mathbb{N}$:

a. Form the $(1 + P) \times (1 + P)$ matrix $\boldsymbol{V}_{t-1}$ as:

$$\boldsymbol{V}_{t-1} := \begin{pmatrix} \sqrt{\lambda_t (\widehat{\boldsymbol{\varphi}}_t^{\top} \widehat{\boldsymbol{\theta}}_{t-1})^2} & \boldsymbol{0}^{\top} \\ \widehat{\boldsymbol{Q}}_{t-1}^{\top} \widehat{\boldsymbol{\psi}}_t & \widehat{\boldsymbol{Q}}_{t-1}^{\top} \end{pmatrix}. \tag{5.29}$$

b. Compute the (unique) QR decomposition of $\boldsymbol{V}_{t-1}$: $\boldsymbol{V}_{t-1} = \boldsymbol{S}_{t-1} \boldsymbol{T}_{t-1}$ with an orthogonal matrix $\boldsymbol{S}_{t-1}$ and an upper triangular matrix $\boldsymbol{T}_{t-1}$.

c. Let $\bar{\boldsymbol{\Pi}}_t$, $\bar{\boldsymbol{L}}_t$, and $\bar{\boldsymbol{Q}}_t$ be the (1×1), $(P \times 1)$, and $(P \times P)$ matrices defined by:

$$\boldsymbol{T}_{t-1} := \begin{pmatrix} \bar{\boldsymbol{\Pi}}_t^{\top} & \bar{\boldsymbol{L}}_t^{\top} \\ \boldsymbol{0} & \bar{\boldsymbol{Q}}_t^{\top} \end{pmatrix}. \tag{5.30}$$

Clearly, both $\bar{\boldsymbol{\Pi}}_t$ and $\bar{\boldsymbol{Q}}_t$ are lower triangular from the definition.

d. Finally, it follows:

$$\bar{\boldsymbol{\Pi}}_t \bar{\boldsymbol{\Pi}}_t^{\top} = \lambda_t (\widehat{\boldsymbol{\varphi}}_t^{\top} \widehat{\boldsymbol{\theta}}_{t-1})^2 + \widehat{\boldsymbol{\psi}}_t^{\top} \widehat{\boldsymbol{P}}_{t-1} \widehat{\boldsymbol{\psi}}_t, \tag{5.31a}$$

$$\widehat{\boldsymbol{P}}_t = \widehat{\boldsymbol{Q}}_t \widehat{\boldsymbol{Q}}_t^{\top} = \frac{\bar{\boldsymbol{Q}}_t \bar{\boldsymbol{Q}}_t^{\top}}{\lambda_t}. \tag{5.31b}$$

Verification of this factorization scheme extending (5.14) is simple since:

$$\boldsymbol{T}_{t-1}^{\top}\boldsymbol{T}_{t-1} = \boldsymbol{V}_{t-1}^{\top}\boldsymbol{V}_{t-1}. \tag{5.32}$$

Several numerically effective procedures exist to perform the triangularization step by the QR decomposition (Söderström & Stoica, 1989, Appendix A.3); they can be processed sequentially.

Applying the latter factorization algorithm, the key updating step of the estimation method (5.14) can be reformulated in the associated terms as:

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1} + \bar{\boldsymbol{L}}_t\bar{\boldsymbol{\Pi}}_t^{-1}(y_t^2 - \widehat{\boldsymbol{\varphi}}_t^{\top}\widehat{\boldsymbol{\theta}}_{t-1}),\ t \in \mathbb{N}. \tag{5.33}$$

The factorization algorithm can be further modified similarly as in Ljung and Söderström (1983, Section 6.5) to involve a regularization step. Namely, we may introduce the matrix $\boldsymbol{Q}_t^*$ instead of $\bar{\boldsymbol{Q}}_t$. Both are identical except for the diagonal elements of $\boldsymbol{Q}_t^*$, which are defined by using the diagonal of $\bar{\boldsymbol{Q}}_t$ as:

$$q_{ii,t}^* := \min(D, \bar{q}_{ii,t}), \quad D > 0. \tag{5.34}$$

The positive number D bounds the diagonal elements of $\bar{\boldsymbol{Q}}_t$. It can be chosen properly to satisfy $\det(\boldsymbol{R}_t^{-1}) < C$ for some $C > 0$. Particularly, if we substitute $\bar{\boldsymbol{Q}}_t$ for $\boldsymbol{Q}_t^*$ in the preceding factorization algorithm, we will obtain:

$$\det(\boldsymbol{R}_t^{-1}) = \frac{1}{\gamma_t}\det(\widehat{\boldsymbol{P}}_t) = \frac{1}{\lambda_t\gamma_t}\left(\prod_{i=1}^{P} q_{ii,t}^*\right)^2,\ t \in \mathbb{N}. \tag{5.35}$$

As long as the estimation is well-behaved, the limit in (5.34) is never reached, and $\boldsymbol{R}_t$ remains unchanged. This regularization technique is activated only if it is necessary. This is the main difference from the Levenberg-Marquardt type regularization, where $\boldsymbol{R}_t$ is adjusted at each time step. Note that both discussed procedures of regularization and factorization can be employed also in the context of the recursive pseudo-linear regression algorithm (5.16).

5.2.4 Projection into stability region

As we have mentioned before, the estimated parameters $\widehat{\boldsymbol{\theta}}_t$ should stay inside a compact region of stability denoted as $\boldsymbol{D}_{\mathcal{M}} \subseteq \mathbb{R}^{p+q+1}$ to guarantee non-degenerative estimation (see Section 5.3.2) and to satisfy other modelling constraints. Particularly, one supposes that the GARCH(p, q) parameters should lie in a compact subset of $\{\boldsymbol{\theta} \in \mathbb{R}^{p+q+1} | \omega > 0, \alpha_i \geq 0, \beta_j \geq 0, \sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j < 1\}$, which assures positivity and stationarity of the conditional variance $\sigma_t^2(\boldsymbol{\theta})$. Note

that we have accepted the obvious simplification of the formal notation here; the second and third inequality hold for all i and j. For instance, one may put $\boldsymbol{D}_{\mathcal{M}} := \{\boldsymbol{\theta} \in \mathbb{R}^{p+q+1} | \delta_\omega \leq \omega \leq \Delta_\omega, \alpha_i \geq 0, \beta_j \geq 0, \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j \leq 1 - \delta_{\alpha\beta}\}$ with some suitable $0 < \delta_\omega \leq \Delta_\omega < \infty$ and $0 < \delta_{\alpha\beta} < 1$ given by a user. See also Section 6.1. In fact, it coincides with the off-line estimation case, where the parameters are calibrated by employing corresponding constraints or a penalized objective function. Consult e.g. Trapletti and Hornik (2013). Therefore, one might adopt a projection into $\boldsymbol{D}_{\mathcal{M}}$ to handle this problem (Ljung, 1999, Chapter 11). For instance, the following rule can be implemented:

$$\left[\widehat{\boldsymbol{\theta}}_t\right]_{\boldsymbol{D}_{\mathcal{M}}} = \begin{cases} \widehat{\boldsymbol{\theta}}_t & \text{if } \widehat{\boldsymbol{\theta}}_t \in \boldsymbol{D}_{\mathcal{M}}, \\ \widehat{\boldsymbol{\theta}}_{t-1} & \text{if } \widehat{\boldsymbol{\theta}}_t \notin \boldsymbol{D}_{\mathcal{M}}. \end{cases} \tag{5.36}$$

This simple projection rule is obviously a special case of the one outlined in Section 4.2 (with the particular choice $\mu = 0$), and it can be eventually replaced by any suitable alternative (Ljung, 1999, Chapter 11). Introducing the projection (5.36) involves the extra computational burden, i.e. verifying whether $\widehat{\boldsymbol{\theta}}_t \in \boldsymbol{D}_{\mathcal{M}}$. However, it is necessary for the successful operation of (5.14), (5.16), and others. Additionally, experience has shown that the projection is activated at only a few samples at the beginning of the data sample. The information loss by ignoring certain observations in the sample according to (5.36) is thus moderate.

5.2.5 Model order identification

Before applying any estimation procedure on available data, we are usually supposed to specify a particular modelling framework that is about to be calibrated. In the GARCH case, we should evidently determine the model order (p, q) at first. In many instances, one operates with a priori information about the model specification. In that case, the estimation can start immediately (e.g. applying the GARCH(1,1) model as the benchmark directly). Otherwise, one must employ various instruments to recognize the adequate order (p, q).

Off-line techniques commonly distinguish between two key approaches. The first one uses the ARMA representation of the GARCH process and the standard tools of its identification (i.e. analysis of autocorrelation and partial autocorrelation functions). The second one based on information criteria is preferred by software products since it is routine (the order is selected minimizing an information criterion). For example, the Akaike information criterion is given by $AIC_T(p, q) = -2L_T + 2(p+q+1)$. The Schwarz (Bayesian) information criterion is given by $SIC_T(p, q) = -2L_T + (p + q + 1)\log(T)$. Here, L_T is the value of the maximized conditional log-likelihood function with the $(p+q+1)$ parameters

using T observations. It is clear that both criteria are almost similar. They differ only in the penalty terms.

To determine the GARCH model order (p, q) on-line, we shall implement a non-standard procedure since no general recommendation exists. Nevertheless, one might be inspired by the idea of using information criteria because (i) they are routine and (ii) they can be simply adapted to involve the principles of on-line estimation. Namely, it holds that $L_T(\boldsymbol{\theta}) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T} F_t(\boldsymbol{\theta})$ and thus $L_T(\boldsymbol{\theta}) = L_{T-1}(\boldsymbol{\theta}) - \frac{1}{2}(\log(2\pi) + F_T(\boldsymbol{\theta}))$, where $F_t(\boldsymbol{\theta})$ is defined by (5.2) for some $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^{p+q+1}$. Note that we implicitly suppose that the GARCH process has the normally distributed innovations.

The criterion F_t is evaluated (in its approximative form) recursively during the on-line estimation. Hence, one may compute $\widehat{L}_t$, a recurrent counterpart of L_t, by applying the recursive estimates of F_t and the previous relations as $\widehat{L}_t = \widehat{L}_{t-1} - \frac{1}{2}\left(\log(2\pi) + y_t^2/(\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1}) + \log(\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})\right)$, $t \in \mathbb{N}$ and $\widehat{L}_0 = 0$. Consequently, we recommend the following algorithm to identify the order (p, q) of the GARCH process in real time (sequentially):

Step 1. Select a proper two-dimensional grid of p and q, e.g. $p, q \in \{1, 2\}$.

Step 2. Run parallel estimation using (5.14) or (5.16) for all available grid points (under the same starting conditions). At each time t, compute $\widehat{L}_t^{p,q}$.

Step 3. After T measurements (T is sufficiently large), compare all available $\widehat{AIC}_T(p, q)$ or $\widehat{SIC}_T(p, q)$ evaluated with $\widehat{L}_T^{p,q}$ and the formulas given above. Choose the order (p^*, q^*) with the minimal value of the preferred criterion. Stop the rest estimation processes.

Alternatively, one can implement another version of this algorithm with the function V_T from (5.1) instead of $\sum_{t=1}^{T} F_t$. It reduces the influence of measurements observed during transient phase due to the accelerating factors.

5.2.6 Robustification of estimation algorithms

In practice, it is necessary to be concerned with abnormal observations that may occur in data. They can be caused by many reasons, e.g. by additive innovations, measurement failures, etc. It is evident (e.g. from (5.1)) that abnormal data points (the so-called *outliers*) will influence the model estimation considerably if no specific action is taken. Therefore, if such defects are expected in the data set, one should modify the estimation algorithms to make them more robust. The outliers tend to appear as spikes in the sequence of $\{y_t/\sqrt{\sigma_t^2}\}$, which obviously result in large contributions to the loss function.

There exist various ways how to robustify recursive estimation algorithms. For instance, the criterion function can be modified to become less sensitive to large errors (e.g. by using the Huber functions) as in Ljung and Söderström (1983, Section 5.5). Another way of handling single outliers is based on testing a measurement at each time t. If it is large compared with a given limit, it will be indicated as erroneous and will be substituted by another value. See e.g. Hanzák and Cipra (2011), Ljung and Söderström (1983, Section 5.5), or Romera and Cipra (1995) for more details.

For example, assume the following criterion applicable in the context of the on-line estimation algorithms considered in Sections 4.1, 4.2, 5.1, and others:

$$\left| \frac{y_t}{\sqrt{\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1}}} \right| \leq a_t, \tag{5.37}$$

where $\{a_t\}_{t\in\mathbb{N}}$ is a deterministic sequence of positive real numbers that reflects the character (distribution) of the data set (usually $a_t \equiv a > 0$ for all $t \in \mathbb{N}$). If this condition is satisfied, then the estimation algorithms will remain unchanged. Otherwise, the measurement y_t will be replaced by $\operatorname{sign}(y_t) a_t \sqrt{\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1}}$. Refer to Section 6.4.3, where this concept is investigated by simulations.

5.3 Self-weighted recursive estimation methods for GARCH models: Theoretical properties

In this section, we shall discuss the convergence and asymptotic properties of the recursive estimation algorithms considered in Section 5.1. This analysis may be undoubtedly useful. Theoretical qualities of the proposed estimation techniques (5.14) and (5.16) can be deduced from the asymptotic analysis of the general recursive prediction error estimation algorithm introduced by Ljung and Söderström (1983, Chapter 3 and Section 4.3).

The general recursive prediction error method minimizes this criterion:

$$\bar{\mathbb{E}}\ell(t, \boldsymbol{\theta}, \boldsymbol{\upsilon}_t(\boldsymbol{\theta})) := \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\ell(t, \boldsymbol{\theta}, \boldsymbol{\upsilon}_t(\boldsymbol{\theta})), \tag{5.38}$$

where $\boldsymbol{\upsilon}_t(\boldsymbol{\theta}) = \boldsymbol{y}_t - \widehat{\boldsymbol{y}}_t(\boldsymbol{\theta})$ is the prediction error associated with the current vector measurement $\boldsymbol{y}_t$ and its prediction $\widehat{\boldsymbol{y}}_t(\boldsymbol{\theta})$ based on the parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^d$ (a parametric space), $d \in \mathbb{N}$. Furthermore, $\ell(\cdot,\cdot,\cdot)$ is a real-valued criterion function on $\mathbb{N} \times \mathbb{R}^d \times \mathbb{R}^n$ associated with a concrete estimation task ($d, n \in \mathbb{N}$). This particular problem evidently conforms to the general off-line

prediction error method, which aims at minimizing the following criterion:

$$\frac{1}{T}\sum_{t=1}^{T}\ell(t,\boldsymbol{\theta},\boldsymbol{v}_t(\boldsymbol{\theta})). \tag{5.39}$$

The structure of recursive estimation methods suitable for minimizing (5.38) is outlined in Ljung and Söderström (1983, Section 4.3). In particular, the general recursive estimation algorithm respects the following sequential scheme:

$$\boldsymbol{v}_t = \boldsymbol{y}_t - \widehat{\boldsymbol{y}}_t, \tag{5.40a}$$

$$\boldsymbol{R}_t = \boldsymbol{R}_{t-1} + \gamma_t \boldsymbol{H}(t, \boldsymbol{R}_{t-1}, \widehat{\boldsymbol{\theta}}_{t-1}, \boldsymbol{v}_t, \boldsymbol{\eta}_t), \tag{5.40b}$$

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1} + \gamma_t \boldsymbol{R}_t^{-1} \boldsymbol{h}(t, \widehat{\boldsymbol{\theta}}_{t-1}, \boldsymbol{v}_t, \boldsymbol{\eta}_t), \tag{5.40c}$$

$$\boldsymbol{\xi}_{t+1} = \boldsymbol{A}(\widehat{\boldsymbol{\theta}}_t)\boldsymbol{\xi}_t + \boldsymbol{B}(\widehat{\boldsymbol{\theta}}_t)\boldsymbol{z}_t, \tag{5.40d}$$

$$\begin{pmatrix} \widehat{\boldsymbol{y}}_{t+1} \\ \mathsf{vec}(\boldsymbol{\eta}_{t+1}) \end{pmatrix} = \boldsymbol{C}(\widehat{\boldsymbol{\theta}}_t)\boldsymbol{\xi}_{t+1}, \tag{5.40e}$$

where the notation has been apparently simplified. The $(d \times 1)$ vector $\widehat{\boldsymbol{\theta}}_t$ denotes the sequential estimate of parameters at time t. The functions $\boldsymbol{H}$ and $\boldsymbol{h}$ (in the $(d \times d)$ square matrix and $(d \times 1)$ vector form) are related to the criterion function ℓ (Ljung & Söderström, 1983, Chapter 3). Moreover, $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ denote the matrix functions of $\boldsymbol{\theta}$ (with particular dimensions). The vector $\boldsymbol{z}_t$ contains measurements and/or other exogenous variables. The vector $\boldsymbol{\xi}_t$ and matrix $\boldsymbol{\eta}_t$ jointly form the recurrently evaluated inputs of $\boldsymbol{h}$ and $\boldsymbol{H}$ according to (5.40d) and (5.40e). The matrix operator vec converts a matrix into a vector by stacking all its columns on top of one another. Finally, $\{\gamma_t\}_{t\in\mathbb{N}}$ is the scalar deterministic gain sequence described in Section 5.2.2.

Apparently, the quantities $\boldsymbol{\eta}_t$ and $\boldsymbol{v}_t$ are determined by all previous estimates $\widehat{\boldsymbol{\theta}}_k$, $k = 1, \ldots, t-1$. Thus, the mapping from $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_t\}$ onto $\widehat{\boldsymbol{\theta}}_t$ is fairly complex and difficult to study. See Chapter 2. In the literature, two key approaches are distinguished to treat the asymptotic analysis of (5.40). See Ljung (1999, Chapter 11), Ljung and Söderström (1983, Chapter 4), Moore and Weiss (1979), and others. Firstly, deterministic ordinary differential equations are associated with the scheme (5.40). Their stability properties are examined to investigate theoretical features of (5.40). Secondly, a stochastic Lyapunov function is introduced for the procedure (5.40). After that, martingale theory is applied to survey the convergence qualities of this function (and thus of (5.40)). The first approach is applicable in general; therefore, it will be outlined here. The next section introduces heuristic explanations of the key moments of such analysis (Ljung & Söderström, 1983, Section 4.3).

5.3.1 Heuristic analysis of the general recursive algorithm

The estimation scheme given by (5.40) is a system of recursive stochastic time-varying difference equations. One might heuristically investigate their behaviour when t becomes large enough. Assume $\gamma_t \to 0$ as $t \to \infty$ (see Section 5.2.2), the step size γ_t will be arbitrary small for sufficiently large t. Hence, the estimates $\widehat{\boldsymbol{\theta}}_t$ will vary more and more slowly in accordance to (5.40c).

From Equation (5.40d), we obtain:

$$\boldsymbol{\xi}_t = \sum_{j=0}^{t-1} \left(\prod_{k=j+1}^{t-1} \boldsymbol{A}(\widehat{\boldsymbol{\theta}}_k) \right) \boldsymbol{B}(\widehat{\boldsymbol{\theta}}_j) \boldsymbol{z}_j. \tag{5.41}$$

Suppose that $\widehat{\boldsymbol{\theta}}_k$ belongs to a small neighbourhood of a value $\bar{\boldsymbol{\theta}}$ for $t - K \leq k \leq t-1$, $K \in \mathbb{N}$, such that $\bar{\boldsymbol{\theta}} \in \boldsymbol{D}_{\mathcal{S}}$. Let $\boldsymbol{D}_{\mathcal{S}}$ be a set of all $\boldsymbol{\theta} \in \mathbb{R}^d$ such that all eigenvalues of $\boldsymbol{A}(\boldsymbol{\theta})$ are strictly inside the unit circle. If the neighbourhood is small enough, we can simplify:

$$\prod_{k=t-K}^{t-1} \boldsymbol{A}(\widehat{\boldsymbol{\theta}}_k) \approx \boldsymbol{A}(\bar{\boldsymbol{\theta}})^K, \tag{5.42}$$

which has apparently a norm smaller than $C\lambda^K$ for some $\lambda \in (0,1)$ and $C > 0$. For K large enough, we may approximate (5.41) as:

$$\boldsymbol{\xi}_t \approx \sum_{j=t-K}^{t-1} \boldsymbol{A}(\bar{\boldsymbol{\theta}})^{t-j-1} \boldsymbol{B}(\bar{\boldsymbol{\theta}}) \boldsymbol{z}_j. \tag{5.43}$$

We can add terms $\boldsymbol{A}(\bar{\boldsymbol{\theta}})^{t-j}\boldsymbol{B}(\bar{\boldsymbol{\theta}})\boldsymbol{z}_j$ for $j < t-K$ to this sum since $\boldsymbol{A}(\bar{\boldsymbol{\theta}})$ is stable (as discussed above); they represent only a small change. Hence, we have:

$$\boldsymbol{\xi}_t \approx \boldsymbol{\xi}_t(\bar{\boldsymbol{\theta}}) := \sum_{j=0}^{t-1} \boldsymbol{A}(\bar{\boldsymbol{\theta}})^{t-j-1} \boldsymbol{B}(\bar{\boldsymbol{\theta}}) \boldsymbol{z}_j, \tag{5.44}$$

which can be expressed in the recursive manner as:

$$\boldsymbol{\xi}_{t+1}(\bar{\boldsymbol{\theta}}) = \boldsymbol{A}(\bar{\boldsymbol{\theta}})\boldsymbol{\xi}_t(\bar{\boldsymbol{\theta}}) + \boldsymbol{B}(\bar{\boldsymbol{\theta}})\boldsymbol{z}_t,\ \boldsymbol{\xi}_0(\bar{\boldsymbol{\theta}}) = \boldsymbol{0}. \tag{5.45}$$

As a consequence, we accept the following approximations:

$$\widehat{\boldsymbol{y}}_t \approx \widehat{\boldsymbol{y}}_t(\bar{\boldsymbol{\theta}}),\ \boldsymbol{\eta}_t \approx \boldsymbol{\eta}_t(\bar{\boldsymbol{\theta}}),\ \boldsymbol{\upsilon}_t \approx \boldsymbol{\upsilon}_t(\bar{\boldsymbol{\theta}}),\ \begin{pmatrix} \widehat{\boldsymbol{y}}_t(\bar{\boldsymbol{\theta}}) \\ \mathsf{vec}(\boldsymbol{\eta}_t(\bar{\boldsymbol{\theta}})) \end{pmatrix} = \boldsymbol{C}(\bar{\boldsymbol{\theta}})\boldsymbol{\xi}_t(\bar{\boldsymbol{\theta}}). \tag{5.46}$$

If $\widehat{\boldsymbol{\theta}}_t$ is close to $\bar{\boldsymbol{\theta}}$, $\boldsymbol{R}_t$ is close to $\bar{\boldsymbol{R}}$, and t is large enough, the previous approximations indicate that Equations (5.40b)-(5.40c) approximately behave like:

$$\boldsymbol{R}_t \approx \boldsymbol{R}_{t-1} + \gamma_t \boldsymbol{H}(t, \bar{\boldsymbol{R}}, \bar{\boldsymbol{\theta}}, \boldsymbol{\upsilon}_t(\bar{\boldsymbol{\theta}}), \boldsymbol{\eta}_t(\bar{\boldsymbol{\theta}})), \tag{5.47a}$$

$$\widehat{\boldsymbol{\theta}}_t \approx \widehat{\boldsymbol{\theta}}_{t-1} + \gamma_t \bar{\boldsymbol{R}}^{-1} \boldsymbol{h}(t, \bar{\boldsymbol{\theta}}, \boldsymbol{\upsilon}_t(\bar{\boldsymbol{\theta}}), \boldsymbol{\eta}_t(\bar{\boldsymbol{\theta}})). \tag{5.47b}$$

Introduce the following expected values:

$$\boldsymbol{f}(\bar{\boldsymbol{\theta}}) := \mathbb{E}\left[\boldsymbol{h}(t, \bar{\boldsymbol{\theta}}, \boldsymbol{\upsilon}_t(\bar{\boldsymbol{\theta}}), \boldsymbol{\eta}_t(\bar{\boldsymbol{\theta}}))\right], \tag{5.48a}$$

$$\boldsymbol{F}(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{R}}) := \mathbb{E}\left[\boldsymbol{H}(t, \bar{\boldsymbol{R}}, \bar{\boldsymbol{\theta}}, \boldsymbol{\upsilon}_t(\bar{\boldsymbol{\theta}}), \boldsymbol{\eta}_t(\bar{\boldsymbol{\theta}}))\right], \tag{5.48b}$$

where the expected values are taken over $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_t\}$. Since t is large enough, one has neglected the transients in (5.45) and assumed the right-hand sides of Equations (5.48) to be time-invariant. Thus, we can get:

$$\boldsymbol{R}_t \approx \boldsymbol{R}_{t-1} + \gamma_t \boldsymbol{F}(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{R}}) + \gamma_t \boldsymbol{\varepsilon}_t^{\boldsymbol{R}}, \tag{5.49a}$$

$$\widehat{\boldsymbol{\theta}}_t \approx \widehat{\boldsymbol{\theta}}_{t-1} + \gamma_t \bar{\boldsymbol{R}}^{-1} \boldsymbol{f}(\bar{\boldsymbol{\theta}}) + \gamma_t \boldsymbol{\varepsilon}_t^{\boldsymbol{\theta}}, \tag{5.49b}$$

where $\{\boldsymbol{\varepsilon}_t^{\boldsymbol{R}}\}$ and $\{\boldsymbol{\varepsilon}_t^{\boldsymbol{\theta}}\}$ are zero-mean random variables. Next, let $\Delta\tau$ be a small number and let t', t be defined as $\sum_{k=t}^{t'} \gamma_k = \Delta\tau$. If $\widehat{\boldsymbol{\theta}}_t = \bar{\boldsymbol{\theta}}$ and $\boldsymbol{R}_t = \bar{\boldsymbol{R}}$, we shall obtain from the previous formulas:

$$\boldsymbol{R}_{t'} \approx \bar{\boldsymbol{R}} + \Delta\tau \boldsymbol{F}(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{R}}) + \sum_{k=t}^{t'} \gamma_k \boldsymbol{\varepsilon}_k^{\boldsymbol{R}}, \tag{5.50a}$$

$$\widehat{\boldsymbol{\theta}}_{t'} \approx \bar{\boldsymbol{\theta}} + \Delta\tau \bar{\boldsymbol{R}}^{-1} \boldsymbol{f}(\bar{\boldsymbol{\theta}}) + \sum_{k=t}^{t'} \gamma_k \boldsymbol{\varepsilon}_k^{\boldsymbol{\theta}}, \tag{5.50b}$$

where the last terms in both equalities are sums of random variables with zero mean; therefore, their contribution will be small (compared to the second terms). Hence, we can accept this simplification and adapt these formulas as follows:

$$\boldsymbol{R}_{t'} \approx \bar{\boldsymbol{R}} + \Delta\tau \boldsymbol{F}(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{R}}), \tag{5.51a}$$

$$\widehat{\boldsymbol{\theta}}_{t'} \approx \bar{\boldsymbol{\theta}} + \Delta\tau \bar{\boldsymbol{R}}^{-1} \boldsymbol{f}(\bar{\boldsymbol{\theta}}). \tag{5.51b}$$

Changing time scale according to the definition of $\Delta\tau$ (i.e. $t \to \tau$, $t' \to \tau + \Delta\tau$ for some τ), Equations (5.51a) and (5.51b) can be regarded as a scheme to solve the following system of differential equations (for small $\Delta\tau$):

$$\frac{d}{d\tau} \boldsymbol{\theta}_D(\tau) = \boldsymbol{R}_D^{-1}(\tau) \boldsymbol{f}(\boldsymbol{\theta}_D(\tau)), \tag{5.52a}$$

$$\frac{d}{d\tau}\boldsymbol{R}_D(\tau) = \boldsymbol{F}(\boldsymbol{\theta}_D(\tau), \boldsymbol{R}_D(\tau)), \tag{5.52b}$$

where the subscript D distinguishes the solution of these equations from the variables in the general recursive prediction error method algorithm (5.40). The discussed arguments advise that if it holds for some large t_0:

$$\widehat{\boldsymbol{\theta}}_{t_0} = \boldsymbol{\theta}_D(\tau_0),\ \boldsymbol{R}_{t_0} = \boldsymbol{R}_D(\tau_0),\ \text{and}\ \sum_{k=1}^{t_0}\gamma_k = \tau_0, \tag{5.53}$$

then it is satisfied for $t > t_0$:

$$\widehat{\boldsymbol{\theta}}_t \approx \boldsymbol{\theta}_D(\tau),\ \boldsymbol{R}_t \approx \boldsymbol{R}_D(\tau),\ \text{and}\ \sum_{k=1}^{t}\gamma_k = \tau. \tag{5.54}$$

The heuristic analysis has shown that the general recursive algorithm outlined by (5.40) can be asymptotically linked to the system of the differential equations (5.52) and that the estimates $\widehat{\boldsymbol{\theta}}_t$ should follow the trajectories of these equations (Ljung & Söderström, 1983, Section 4.3). It means that one might employ the theory of ordinary differential equations as a background of convergence analysis of recursive estimation methods given by (5.40).

5.3.2 Formal analysis of the general recursive algorithm

Before formulating a convergence theorem for the recursive estimation algorithm (5.40), we should introduce some assumptions. See Ljung and Söderström (1983, Section 4.3) or Moore and Weiss (1979) for further insights.

(M1) $\boldsymbol{D}_{\mathcal{M}} \subseteq \mathbb{R}^d$ is a compact set such that: $\boldsymbol{\theta} \in \boldsymbol{D}_{\mathcal{M}} \subseteq \mathbb{R}^d \Rightarrow \boldsymbol{A}(\boldsymbol{\theta})$ has all eigenvalues strictly inside the unit circle.

(M2) The matrices $\boldsymbol{A}(\boldsymbol{\theta})$, $\boldsymbol{B}(\boldsymbol{\theta})$, and $\boldsymbol{C}(\boldsymbol{\theta})$ are continuously differentiable with respect to $\boldsymbol{\theta}$, for all $\boldsymbol{\theta} \in \boldsymbol{D}_{\mathcal{M}}$.

(Cr1) The function $\boldsymbol{h}(t, \boldsymbol{\theta}, \boldsymbol{\upsilon}, \boldsymbol{\eta})$ is differentiable with respect to $\boldsymbol{\theta}$, $\boldsymbol{\upsilon}$, and $\boldsymbol{\eta}$. For some $C \in (0, \infty)$ and all $\boldsymbol{\theta} \in \boldsymbol{D}_{\mathcal{M}}$, the following conditions hold:

$$\|\boldsymbol{h}(t, \boldsymbol{\theta}, \boldsymbol{\upsilon}, \boldsymbol{\eta})\| + \|\boldsymbol{h}'_{\boldsymbol{\theta}}(t, \boldsymbol{\theta}, \boldsymbol{\upsilon}, \boldsymbol{\eta})\| \leq C(1 + \|\boldsymbol{\upsilon}\|^2 + \|\boldsymbol{\eta}\|^2),$$
$$\|\boldsymbol{h}'_{\boldsymbol{\upsilon}}(t, \boldsymbol{\theta}, \boldsymbol{\upsilon}, \boldsymbol{\eta})\| + \|\boldsymbol{h}'_{\boldsymbol{\eta}}(t, \boldsymbol{\theta}, \boldsymbol{\upsilon}, \boldsymbol{\eta})\| \leq C(1 + \|\boldsymbol{\upsilon}\| + \|\boldsymbol{\eta}\|),$$

where $\|\cdot\|$ denotes vector and matrix norms. Furthermore, $\boldsymbol{h}'_{\boldsymbol{\theta}}$, $\boldsymbol{h}'_{\boldsymbol{\upsilon}}$, and $\boldsymbol{h}'_{\boldsymbol{\eta}}$ are the first derivatives of $\boldsymbol{h}$ with respect to $\boldsymbol{\theta}$, $\boldsymbol{\upsilon}$, and $\boldsymbol{\eta}$, respectively.

(Cr2) The function $\boldsymbol{H}(t, \boldsymbol{R}, \boldsymbol{\theta}, \boldsymbol{\upsilon}, \boldsymbol{\eta})$ is differentiable with respect to $\boldsymbol{R}$, $\boldsymbol{\theta}$, $\boldsymbol{\upsilon}$, and $\boldsymbol{\eta}$. For some $C \in (0, \infty)$ and all $\boldsymbol{\theta} \in \boldsymbol{D}_{\mathcal{M}}$, these inequalities hold:

$$\begin{aligned}
\|\boldsymbol{H}(t, \boldsymbol{R}, \boldsymbol{\theta}, \boldsymbol{\upsilon}, \boldsymbol{\eta})\| &\leq C(1 + \|\boldsymbol{\upsilon}\|^2 + \|\boldsymbol{\eta}\|^2 + \|\boldsymbol{R}\|), \\
\|\boldsymbol{H}'_{\boldsymbol{R}}(t, \boldsymbol{R}, \boldsymbol{\theta}, \boldsymbol{\upsilon}, \boldsymbol{\eta})\| + \|\boldsymbol{H}'_{\boldsymbol{\theta}}(t, \boldsymbol{R}, \boldsymbol{\theta}, \boldsymbol{\upsilon}, \boldsymbol{\eta})\| &\leq C(1 + \|\boldsymbol{\upsilon}\|^2 + \|\boldsymbol{\eta}\|^2), \\
\|\boldsymbol{H}'_{\boldsymbol{\upsilon}}(t, \boldsymbol{R}, \boldsymbol{\theta}, \boldsymbol{\upsilon}, \boldsymbol{\eta})\| + \|\boldsymbol{H}'_{\boldsymbol{\eta}}(t, \boldsymbol{R}, \boldsymbol{\theta}, \boldsymbol{\upsilon}, \boldsymbol{\eta})\| &\leq C(1 + \|\boldsymbol{\upsilon}\| + \|\boldsymbol{\eta}\|),
\end{aligned}$$

where $\boldsymbol{H}'_{\boldsymbol{R}}$, $\boldsymbol{H}'_{\boldsymbol{\theta}}$, $\boldsymbol{H}'_{\boldsymbol{\upsilon}}$, and $\boldsymbol{H}'_{\boldsymbol{\eta}}$ are the first derivatives of $\boldsymbol{H}$ with respect to $\boldsymbol{R}$, $\boldsymbol{\theta}$, $\boldsymbol{\upsilon}$, and $\boldsymbol{\eta}$, respectively.

(R1) The matrix $\boldsymbol{R}_t$ is symmetric and $\boldsymbol{R}_t \geq \delta \boldsymbol{I}$ for all t and some $\delta > 0$.

(G1) It holds $\lim_{t \to \infty} t\gamma_t = 1$.

(A1) The data sequence $\{\boldsymbol{z}_t\}_{t \in \mathbb{N}}$ is such that the following limits exist (almost surely) for all $\boldsymbol{\theta} \in \boldsymbol{D}_{\mathcal{M}}$:

(A1a) $\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{h}(t, \boldsymbol{\theta}, \boldsymbol{\upsilon}_t(\boldsymbol{\theta}), \boldsymbol{\eta}_t(\boldsymbol{\theta})) =: \boldsymbol{f}(\boldsymbol{\theta})$,

(A1b) $\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{H}(t, \boldsymbol{R}, \boldsymbol{\theta}, \boldsymbol{\upsilon}_t(\boldsymbol{\theta}), \boldsymbol{\eta}_t(\boldsymbol{\theta})) =: \boldsymbol{F}(\boldsymbol{\theta}, \boldsymbol{R})$,

(A1c) $\lim_{T \to \infty} \sup \frac{1}{T} \sum_{t=1}^{T} [1 + \|\boldsymbol{z}_t\|]^3 < \infty$.

The variables $\boldsymbol{\upsilon}_t(\boldsymbol{\theta})$, $\widehat{\boldsymbol{y}}_t(\boldsymbol{\theta})$, and $\boldsymbol{\eta}_t(\boldsymbol{\theta})$ are obtained from $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_t\}$ by the following linear filter:

$$\begin{aligned}
\boldsymbol{\xi}_{t+1}(\boldsymbol{\theta}) &= \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{\xi}_t(\boldsymbol{\theta}) + \boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{z}_t, \ \boldsymbol{\xi}_0(\boldsymbol{\theta}) = \boldsymbol{0}, \\
\begin{pmatrix} \widehat{\boldsymbol{y}}_t(\boldsymbol{\theta}) \\ \mathsf{vec}(\boldsymbol{\eta}_t(\boldsymbol{\theta})) \end{pmatrix} &= \boldsymbol{C}(\boldsymbol{\theta})\boldsymbol{\xi}_t(\boldsymbol{\theta}), \ \boldsymbol{\upsilon}_t(\boldsymbol{\theta}) = \boldsymbol{y}_t - \widehat{\boldsymbol{y}}_t(\boldsymbol{\theta}).
\end{aligned}$$

Remark 5.3.1 The condition (M1) defines $\boldsymbol{D}_{\mathcal{M}}$ as a compact subset of the region $\boldsymbol{D}_{\mathcal{S}}$ defined in Section 5.3.1. It should be pointed out that (M1) assures only that Equation (5.40d) is stable (in an evident sense). It does not impose any serious restriction on the model. Particularly, it does not consider the structure of the parametric space $\boldsymbol{\Theta}$. Nevertheless, this does not result in any limitation since the additive constraints may be of course incorporated into $\boldsymbol{D}_{\mathcal{M}}$.

The assumptions (Cr1) and (Cr2) are smoothness conditions. The assumptions (R1) and (G1) are not restrictive since (R1) can be guaranteed by using any regularization technique (see Section 5.2.3), and (G1) depends on the user's choice only (see Section 5.2.2).

The conditions in (A1) are certainly implicit (Ljung & Söderström, 1983, Section 4.3). The type of the intended convergence depends on the character of the sequence $\{\boldsymbol{z}_t\}_{t\in\mathbb{N}}$; one can distinguish between deterministic or stochastic variants. Moreover, the limits in (A1a) and (A1b) represent "average" adjustments that would be made in the algorithm if $\boldsymbol{\theta}$ was fixed as a constant. The heuristic analysis has applied exactly this concept (see Section 5.3.1).

Note that the condition (A1a) will be satisfied almost surely as t goes to infinity, for instance, if the following assumptions hold (let $\boldsymbol{h}_t := \boldsymbol{h}(t, \boldsymbol{\theta}, \boldsymbol{\upsilon}_t(\boldsymbol{\theta}), \boldsymbol{\eta}_t(\boldsymbol{\theta}))$): (i) $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}[\boldsymbol{h}_t - \mathbb{E}(\boldsymbol{h}_t)] = \boldsymbol{0}$ almost surely, (ii) $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(\boldsymbol{h}_t) = \boldsymbol{f}(\boldsymbol{\theta})$. The first assumption will be fulfilled, for example, if $\boldsymbol{h}_t$ and $\boldsymbol{h}_s$ are asymptotically independent for a fixed $\boldsymbol{\theta} \in \boldsymbol{D}_{\mathcal{M}}$ when $|t-s|$ is large (Ljung & Söderström, 1983, Section 4.3). The condition (A1b) can be handled similarly.

Theorem 5.3.1 *Consider the estimation algorithm* (5.40). *Assume that it includes a projection to keep* $\widehat{\boldsymbol{\theta}}_t$ *inside* $\boldsymbol{D}_{\mathcal{M}}$ *and to assure a bounded subsequence of* $\{\boldsymbol{\xi}_t\}_{t\in\mathbb{N}}$ *(that may depend on the realization of* $\{\boldsymbol{z}_t\}_{t\in\mathbb{N}}$*). Assume that (M1), (M2), (Cr1), (Cr2), (R1), and (G1) hold and that (A1) holds (almost surely). Suppose that there exists a positive function* $V(\boldsymbol{\theta}, \boldsymbol{R})$ *such that:*

$$\frac{d}{d\tau}V(\boldsymbol{\theta}_D(\tau), \boldsymbol{R}_D(\tau)) \leq 0 \quad \textit{for all } \boldsymbol{\theta}_D \in \boldsymbol{D}_{\mathcal{M}} \tag{5.58}$$

when it is evaluated along solutions of these ordinary differential equations:

$$\frac{d}{d\tau}\boldsymbol{\theta}_D(\tau) = \boldsymbol{R}_D^{-1}(\tau)\boldsymbol{f}(\boldsymbol{\theta}_D(\tau)), \tag{5.59a}$$

$$\frac{d}{d\tau}\boldsymbol{R}_D(\tau) = \boldsymbol{F}(\boldsymbol{\theta}_D(\tau), \boldsymbol{R}_D(\tau)), \tag{5.59b}$$

where $\boldsymbol{f}$ *and* $\boldsymbol{F}$ *are defined by (A1). Let* $\boldsymbol{D}_{\mathcal{C}} = \{(\boldsymbol{\theta}, \boldsymbol{R}) | \frac{d}{d\tau}V(\boldsymbol{\theta}_D(\tau), \boldsymbol{R}_D(\tau)) = 0\}$.

Then as $t \to \infty$ *either* $\{(\widehat{\boldsymbol{\theta}}_t, \boldsymbol{R}_t)\}$ *tend to* $\boldsymbol{D}_{\mathcal{C}}$ *or* $\{\widehat{\boldsymbol{\theta}}_t\}$ *tends to the boundary of* $\boldsymbol{D}_{\mathcal{M}}$ *(almost surely).*

Proof See Ljung and Söderström (1983, Section 4.3) for the references. ∎

5.3.3 Remarks on convergence analysis: The self-weighted recursive prediction error method algorithm

The self-weighted recursive estimation algorithm (5.13) designed for estimating the GARCH models can be clearly incorporated into the general recursive calibration scheme (5.40). Particularly, one identifies that: (i) $\ell(t, \boldsymbol{\theta}, \boldsymbol{\upsilon}_t(\boldsymbol{\theta}))$ equals $F_t(\boldsymbol{\theta})$ given by (5.2), (ii) $\boldsymbol{h}(t, \boldsymbol{\theta}, \boldsymbol{\upsilon}_t(\boldsymbol{\theta}), \boldsymbol{\eta}_t(\boldsymbol{\theta}))$ corresponds to $-[\nabla_{\boldsymbol{\theta}}\ell(t, \boldsymbol{\theta}, \boldsymbol{\upsilon}_t(\boldsymbol{\theta}))]^\top$, and (iii) $\boldsymbol{H}(t, \boldsymbol{R}, \boldsymbol{\theta}, \boldsymbol{\upsilon}_t(\boldsymbol{\theta}), \boldsymbol{\eta}_t(\boldsymbol{\theta}))$ is clearly defined by (5.13b). See also Section 5.2.

Therefore, Theorem 5.3.1 may be employed to verify the convergence features of the recursive estimates delivered by (5.13). For instance, in the case of the GARCH$(1,1)$ process $\{y_t\}_{t\in\mathbb{Z}}$, we can write in (5.40):

$$\begin{aligned}
&\boldsymbol{\upsilon}_t = y_t^2 - \widehat{y}_t^2, \quad \widehat{y}_t^2 = \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1} = \left(\boldsymbol{C}(\widehat{\boldsymbol{\theta}}_{t-1})\boldsymbol{\xi}_t\right)_1, \\
&\boldsymbol{\xi}_t = \left(\widehat{\boldsymbol{\varphi}}_t^\top, \widehat{\boldsymbol{\psi}}_t^\top\right)^\top, \\
&\boldsymbol{z}_t = (1, y_t^2)^\top, \\
&\boldsymbol{\eta}_t = \left(\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1}, \widehat{\boldsymbol{\psi}}_t^\top\right)^\top,
\end{aligned}$$

and thus:

$$\begin{aligned}
\boldsymbol{h}(t, \widehat{\boldsymbol{\theta}}_{t-1}, \boldsymbol{\upsilon}_t, \boldsymbol{\eta}_t) &= \frac{\boldsymbol{\upsilon}_t}{[(1,0,0,0)\boldsymbol{\eta}_t]^2}\boldsymbol{T}\boldsymbol{\eta}_t, \\
\boldsymbol{H}(t, \boldsymbol{R}_{t-1}, \widehat{\boldsymbol{\theta}}_{t-1}, \boldsymbol{\upsilon}_t, \boldsymbol{\eta}_t) &= \frac{(\boldsymbol{T}\boldsymbol{\eta}_t)(\boldsymbol{T}\boldsymbol{\eta}_t)^\top}{[(1,0,0,0)\boldsymbol{\eta}_t]^2} - \boldsymbol{R}_{t-1},
\end{aligned}$$

where

$$\boldsymbol{A}(\widehat{\boldsymbol{\theta}}_t) = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
\widehat{\omega}_t & \widehat{\alpha}_{1,t} & \widehat{\beta}_{1,t} & 0 & 0 & 0 \\
0 & 0 & 0 & \widehat{\beta}_{1,t} & 0 & 0 \\
0 & 0 & 0 & 0 & \widehat{\beta}_{1,t} & 0 \\
\widehat{\omega}_t & \widehat{\alpha}_{1,t} & \widehat{\beta}_{1,t} & 0 & 0 & \widehat{\beta}_{1,t}
\end{pmatrix}, \; \boldsymbol{B}(\widehat{\boldsymbol{\theta}}_t) = \begin{pmatrix}
1 & 0 \\
0 & 1 \\
0 & 0 \\
1 & 0 \\
0 & 1 \\
0 & 0
\end{pmatrix},$$

and

$$\boldsymbol{C}(\widehat{\boldsymbol{\theta}}_t) = \begin{pmatrix}
\widehat{\omega}_t & \widehat{\alpha}_{1,t} & \widehat{\beta}_{1,t} & 0 & 0 & 0 \\
\widehat{\omega}_t & \widehat{\alpha}_{1,t} & \widehat{\beta}_{1,t} & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}, \; \boldsymbol{T} = \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}.$$

The previous expressions can be easily generalized into the GARCH(p,q) form.

If the conditions of Theorem 5.3.1 are satisfied for these particular settings and the following limits:

$$\widetilde{V}(\boldsymbol{\theta}) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\ell(t, \boldsymbol{\theta}, \boldsymbol{\upsilon}_t(\boldsymbol{\theta})), \tag{5.62a}$$

$$-\boldsymbol{f}(\boldsymbol{\theta}) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\ell(t, \boldsymbol{\theta}, \boldsymbol{\upsilon}_t(\boldsymbol{\theta}))\right]^\top, \tag{5.62b}$$

$$\boldsymbol{F}(\boldsymbol{\theta}, \boldsymbol{R}) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\boldsymbol{H}(t, \boldsymbol{R}, \boldsymbol{\theta}, \boldsymbol{\upsilon}_t(\boldsymbol{\theta}), \boldsymbol{\eta}_t(\boldsymbol{\theta})) \tag{5.62c}$$

exist for all $\boldsymbol{\theta} \in \boldsymbol{D}_{\mathcal{M}}$, one can consider $\widetilde{V}(\boldsymbol{\theta})$ as a candidate for the function V appearing in Theorem 5.3.1. See Section 5.2.2 for the definition of the gain sequence, Section 5.2.3 for the regularization techniques, and Section 5.2.4 for the adopted projection and the structure of $\boldsymbol{D}_{\mathcal{M}}$. With appropriate assumptions (i.e. that $[\nabla_{\boldsymbol{\theta}} \ell(t, \boldsymbol{\theta}, \boldsymbol{\upsilon}_t(\boldsymbol{\theta}))]^\top$ is dominated by a $\boldsymbol{\theta}$-independent integrable function and that the second derivative of $\mathbb{E}\ell(t, \boldsymbol{\theta}, \boldsymbol{\upsilon}_t(\boldsymbol{\theta}))$ is bounded), we can establish:

$$\boldsymbol{f}^\top(\boldsymbol{\theta}) = -\nabla_{\boldsymbol{\theta}} \widetilde{V}(\boldsymbol{\theta}). \tag{5.63}$$

To verify that $\widetilde{V}(\boldsymbol{\theta})$ is the function satisfying the assumptions of Theorem 5.3.1, we should study the associated ordinary differential equations:

$$\begin{aligned}
\frac{d}{d\tau}\boldsymbol{\theta}_D(\tau) &= \boldsymbol{R}_D^{-1}(\tau)\boldsymbol{f}(\boldsymbol{\theta}_D(\tau)) = -\boldsymbol{R}_D^{-1}(\tau)\left[\nabla_{\boldsymbol{\theta}} \widetilde{V}(\boldsymbol{\theta}_D(\tau))\right]^\top, \\
\frac{d}{d\tau}\boldsymbol{R}_D(\tau) &= \boldsymbol{F}(\boldsymbol{\theta}_D(\tau), \boldsymbol{R}_D(\tau)).
\end{aligned}$$

Thus, along trajectories of these equations, we shall obtain:

$$\begin{aligned}
\frac{d}{d\tau}\widetilde{V}(\boldsymbol{\theta}_D(\tau)) &= \nabla_{\boldsymbol{\theta}} \widetilde{V}(\boldsymbol{\theta}_D(\tau)) \frac{d}{d\tau}\boldsymbol{\theta}_D(\tau) \\
&= -\nabla_{\boldsymbol{\theta}} \widetilde{V}(\boldsymbol{\theta}_D(\tau)) \boldsymbol{R}_D^{-1}(\tau)\left[\nabla_{\boldsymbol{\theta}} \widetilde{V}(\boldsymbol{\theta}_D(\tau))\right]^\top \leq 0,
\end{aligned}$$

which follows from the condition (R1). The equality is fulfilled for $\boldsymbol{\theta}_D(\tau) \in \boldsymbol{D}_{\mathcal{C}}$. In that case, we may write $\boldsymbol{D}_{\mathcal{C}}$ in this form:

$$\boldsymbol{D}_{\mathcal{C}} = \left\{ \boldsymbol{\theta} \,\middle|\, \left[\nabla_{\boldsymbol{\theta}} \widetilde{V}(\boldsymbol{\theta})\right]^\top = \boldsymbol{0} \right\}, \tag{5.66}$$

or in more detail as:

$$\boldsymbol{D}_{\mathcal{C}} = \left\{ \boldsymbol{\theta} \,\middle|\, \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[\frac{y_t^2 - \boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}}{\left(\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}\right)^2} \boldsymbol{\psi}_t(\boldsymbol{\theta}) \right] = \boldsymbol{0} \right\}. \tag{5.67}$$

In respect to Theorem 5.3.1, the estimates $\widehat{\boldsymbol{\theta}}_t$ converge either to a stationary point of the criterion function $\widetilde{V}$ or to the boundary of $\boldsymbol{D}_{\mathcal{M}}$ almost surely as t goes to infinity. This result can be strengthened to convergence to a local minimum of the function $\widetilde{V}(\boldsymbol{\theta})$, see e.g. Moore and Weiss (1979) or Remark 5.3.2.

Recall that $\widetilde{V}(\boldsymbol{\theta})$ is also the limit of the off-line criterion (5.39) under weak conditions according to Ljung and Söderström (1983, Section 4.4.2) when the

limits in (5.62) exist. Compare also with Equation (5.1) applying the assumption (G1). The off-line estimates are usually found by local numerical minimization of the criterion (5.1) (or (5.39)). See Remark 5.1.3. Consequently, the convergence properties of the recursive prediction error method estimates and their off-line counterparts will coincide. Refer to Ljung and Söderström (1983, Section 4.4) or Moore and Weiss (1979). It should be pointed out that the off-line criterion (5.1) (or (5.39)) is readily associated with the negative conditional logarithmic likelihood function for estimating the GARCH(p, q) processes.

The drawback of this result is, however, that it cannot guarantee convergence to the global minimum of the loss function (similarly as the off-line estimators). Namely, if other local minima exist, the estimate may converge to one of them.

Remark 5.3.2 Moreover, we can examine the stationary points of the associated ordinary differential equations (under the given assumptions). In particular, these equations are given in the following form (see above):

$$\frac{d}{d\tau}\boldsymbol{\theta}_D(\tau) = -\boldsymbol{R}_D^{-1}(\tau)\left[\nabla_{\boldsymbol{\theta}}\widetilde{V}(\boldsymbol{\theta}_D(\tau))\right]^{\top}, \tag{5.68a}$$

$$\frac{d}{d\tau}\boldsymbol{R}_D(\tau) = \boldsymbol{G}(\boldsymbol{\theta}_D(\tau)) - \boldsymbol{R}_D(\tau), \tag{5.68b}$$

where one clearly puts $\boldsymbol{G}(\boldsymbol{\theta}) := \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\boldsymbol{\psi}_t(\boldsymbol{\theta})\boldsymbol{\psi}_t^{\top}(\boldsymbol{\theta})/(\boldsymbol{\varphi}_t^{\top}(\boldsymbol{\theta})\boldsymbol{\theta})^2\right]$ and $\boldsymbol{F}(\boldsymbol{\theta}, \boldsymbol{R}) := \boldsymbol{G}(\boldsymbol{\theta}) - \boldsymbol{R}$. The stationary points of these equations are defined as:

$$\left\{(\boldsymbol{\theta}, \boldsymbol{R}) \,\middle|\, \boldsymbol{f}(\boldsymbol{\theta}) = -\left[\nabla_{\boldsymbol{\theta}}\widetilde{V}(\boldsymbol{\theta})\right]^{\top} = \boldsymbol{0},\ \boldsymbol{R} = \boldsymbol{G}(\boldsymbol{\theta})\right\}. \tag{5.69}$$

After linearizing (5.68) around a stationary point $(\boldsymbol{\theta}^*, \boldsymbol{R}^*)$, we shall get:

$$\frac{d}{d\tau}\left(\boldsymbol{\theta}(\tau) - \boldsymbol{\theta}^*\right) = \left[\boldsymbol{G}(\boldsymbol{\theta}^*)\right]^{-1}\widetilde{\boldsymbol{H}}(\boldsymbol{\theta}^*)\left(\boldsymbol{\theta}(\tau) - \boldsymbol{\theta}^*\right),$$
$$\frac{d}{d\tau}\left(\boldsymbol{R}(\tau) - \boldsymbol{R}^*\right) = -\left(\boldsymbol{R}(\tau) - \boldsymbol{R}^*\right) + \boldsymbol{G}'(\boldsymbol{\theta}^*)\left(\boldsymbol{\theta}(\tau) - \boldsymbol{\theta}^*\right),$$

where $\widetilde{\boldsymbol{H}}(\boldsymbol{\theta}^*) = \frac{d}{d\boldsymbol{\theta}}\boldsymbol{f}(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$ and the term $\boldsymbol{G}'(\boldsymbol{\theta}^*)\left(\boldsymbol{\theta}(\tau) - \boldsymbol{\theta}^*\right)$ stands for the second term in the Taylor expansion of $\boldsymbol{G}(\boldsymbol{\theta}^* + (\boldsymbol{\theta}(\tau) - \boldsymbol{\theta}^*))$.

The stationary point $(\boldsymbol{\theta}^*, \boldsymbol{R}^*)$ is stable if and only if the matrix $\boldsymbol{L}(\boldsymbol{\theta}^*) := \left[\boldsymbol{G}(\boldsymbol{\theta}^*)\right]^{-1}\widetilde{\boldsymbol{H}}(\boldsymbol{\theta}^*)$ has all eigenvalues in the left half-plane (Söderström & Stoica, 1989, Section 9.6). According to Ljung and Söderström (1983, Results 4.1 and 4.3), it might be concluded that $\widehat{\boldsymbol{\theta}}_t$ can only converge to a value $\boldsymbol{\theta}^*$ such that $\left[\nabla_{\boldsymbol{\theta}}\widetilde{V}(\boldsymbol{\theta}^*)\right]^{\top} = \boldsymbol{0}$ and such that the matrix $\boldsymbol{L}(\boldsymbol{\theta}^*)$ has all eigenvalues in the left half-plane. This result evidently provides a tool for investigating the necessary conditions for (local) convergence. A desired convergence point of the studied esti-

mation algorithm must be a stable stationary point of the corresponding ordinary differential equations. This can be used to improve the convergence analysis by excluding unstable stationary points from $\boldsymbol{D}_{\mathcal{C}}$. In other words, the points that are not local minima of $\widetilde{V}$ can be excluded from $\boldsymbol{D}_{\mathcal{C}}$.

Remark 5.3.3 The convergence properties of the investigated recursive estimates have been established; therefore, the next question is to describe the asymptotic behaviour of the discussed estimation scheme. Assume that the vector of the true parameters $\boldsymbol{\theta}_0 \in \boldsymbol{D}_{\mathcal{M}}$ satisfies that $y_t^2/(\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0) - 1 = e_t^0$ is (strict) white noise (compare with (4.2)). Let $\widetilde{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_0$ and $\bar{\boldsymbol{R}}_t = \frac{1}{\gamma_t}\boldsymbol{R}_t$. After some computation using all formulas in (5.13), we shall obtain:

$$\bar{\boldsymbol{R}}_t\widetilde{\boldsymbol{\theta}}_t = \lambda_t\bar{\boldsymbol{R}}_{t-1}\widetilde{\boldsymbol{\theta}}_{t-1} + \frac{y_t^2 - \widehat{\boldsymbol{\varphi}}_t^\top\widehat{\boldsymbol{\theta}}_{t-1}}{(\widehat{\boldsymbol{\varphi}}_t^\top\widehat{\boldsymbol{\theta}}_{t-1})^2}\widehat{\boldsymbol{\psi}}_t + \frac{\widehat{\boldsymbol{\psi}}_t\widehat{\boldsymbol{\psi}}_t^\top}{(\widehat{\boldsymbol{\varphi}}_t^\top\widehat{\boldsymbol{\theta}}_{t-1})^2}\widetilde{\boldsymbol{\theta}}_{t-1}.$$

This expression can be summed from time $k = 0$ to t with $\beta(t,k) = \prod_{j=k+1}^{t}\lambda_j$ and the convention $\beta(t,t) = 1$, i.e.:

$$\begin{aligned}\bar{\boldsymbol{R}}_t\widetilde{\boldsymbol{\theta}}_t &= \beta(t,0)\bar{\boldsymbol{R}}_0\widetilde{\boldsymbol{\theta}}_0 + \sum_{k=1}^{t}\beta(t,k)\left[\frac{y_k^2 - \widehat{\boldsymbol{\varphi}}_k^\top\widehat{\boldsymbol{\theta}}_{k-1}}{(\widehat{\boldsymbol{\varphi}}_k^\top\widehat{\boldsymbol{\theta}}_{k-1})^2}\widehat{\boldsymbol{\psi}}_k + \frac{\widehat{\boldsymbol{\psi}}_k\widehat{\boldsymbol{\psi}}_k^\top}{(\widehat{\boldsymbol{\varphi}}_k^\top\widehat{\boldsymbol{\theta}}_{k-1})^2}\widetilde{\boldsymbol{\theta}}_{k-1}\right] \\ &= \beta(t,0)\bar{\boldsymbol{R}}_0\widetilde{\boldsymbol{\theta}}_0 + \sum_{k=1}^{t}\beta(t,k)\frac{\boldsymbol{\psi}_k(\boldsymbol{\theta}_0)}{\boldsymbol{\varphi}_k^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0}\left(\frac{y_k^2}{\boldsymbol{\varphi}_k^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0} - 1\right) + \boldsymbol{S}_t,\end{aligned}$$

where

$$\boldsymbol{S}_t = \sum_{k=1}^{t}\beta(t,k)\left[\frac{y_k^2 - \widehat{\boldsymbol{\varphi}}_k^\top\widehat{\boldsymbol{\theta}}_{k-1}}{(\widehat{\boldsymbol{\varphi}}_k^\top\widehat{\boldsymbol{\theta}}_{k-1})^2}\widehat{\boldsymbol{\psi}}_k - \frac{y_k^2 - \boldsymbol{\varphi}_k^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0}{(\boldsymbol{\varphi}_k^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0)^2}\boldsymbol{\psi}_k(\boldsymbol{\theta}_0) + \frac{\widehat{\boldsymbol{\psi}}_k\widehat{\boldsymbol{\psi}}_k^\top}{(\widehat{\boldsymbol{\varphi}}_k^\top\widehat{\boldsymbol{\theta}}_{k-1})^2}\widetilde{\boldsymbol{\theta}}_{k-1}\right].$$

Employing the initial suppositions, one can simplify:

$$\begin{aligned}\frac{y_k^2 - \widehat{\boldsymbol{\varphi}}_k^\top\widehat{\boldsymbol{\theta}}_{k-1}}{(\widehat{\boldsymbol{\varphi}}_k^\top\widehat{\boldsymbol{\theta}}_{k-1})^2}\widehat{\boldsymbol{\psi}}_k - \frac{y_k^2 - \boldsymbol{\varphi}_k^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0}{(\boldsymbol{\varphi}_k^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0)^2}\boldsymbol{\psi}_k(\boldsymbol{\theta}_0) &\approx [\nabla_{\boldsymbol{\theta}}F_k(\boldsymbol{\theta}_0)]^\top - \left[\nabla_{\boldsymbol{\theta}}F_k(\widehat{\boldsymbol{\theta}}_{k-1})\right]^\top \\ &\approx -\frac{\widehat{\boldsymbol{\psi}}_k\widehat{\boldsymbol{\psi}}_k^\top}{(\widehat{\boldsymbol{\varphi}}_k^\top\widehat{\boldsymbol{\theta}}_{k-1})^2}\widetilde{\boldsymbol{\theta}}_{k-1}.\end{aligned}$$

The first approximation follows the fact that $\widehat{\boldsymbol{\varphi}}_k$ and $\widehat{\boldsymbol{\psi}}_k$ estimate $\boldsymbol{\varphi}_k(\widehat{\boldsymbol{\theta}}_{k-1})$ and $\boldsymbol{\psi}_k(\widehat{\boldsymbol{\theta}}_{k-1})$ according to the adopted estimation framework. The second one is the consequence of the Taylor expansion using the approximation $\boldsymbol{F}_k''(\boldsymbol{\theta}) \approx \widetilde{\boldsymbol{F}}_k''(\boldsymbol{\theta})$, which has been accepted for $\boldsymbol{\theta}$ that are close to $\boldsymbol{\theta}_0$. See Section 5.1. Apparently, the sum $\boldsymbol{S}_t$ is likely negligible compared to other terms. Similarly, $\beta(t,0)\bar{\boldsymbol{R}}_0\widetilde{\boldsymbol{\theta}}_0$

should also be negligible. Therefore, we may write:

$$\bar{\boldsymbol{R}}_t \widetilde{\boldsymbol{\theta}}_t \approx \sum_{k=1}^{t} \beta(t,k) \frac{\boldsymbol{\psi}_k(\boldsymbol{\theta}_0)}{\boldsymbol{\varphi}_k^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0} e_k^0.$$

Assuming that $\widehat{\boldsymbol{\theta}}_{k-1}$ is close to $\boldsymbol{\theta}_0$ "asymptotically most of time", we can replace $\widehat{\boldsymbol{\psi}}_k$ by $\boldsymbol{\psi}_k(\boldsymbol{\theta}_0)$ and $\widehat{\boldsymbol{\varphi}}_k^\top \widehat{\boldsymbol{\theta}}_{k-1}$ by $\boldsymbol{\varphi}_k^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0$ without too much error and constitute $\bar{\boldsymbol{R}}_k^0$ by these quantities as in (5.13) (Ljung, 1999, Chapter 11). Consequently, one instantly obtains:

$$\left(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_0\right) \approx \left[\bar{\boldsymbol{R}}_t^0\right]^{-1} \sum_{k=1}^{t} \beta(t,k) \frac{\boldsymbol{\psi}_k(\boldsymbol{\theta}_0)}{\boldsymbol{\varphi}_k^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0} e_k^0. \tag{5.73}$$

This approximation is exactly the same asymptotic expression as for the off-line estimation case as was found by Ljung (1999, Chapter 11) or Ljung and Söderström (1983, Section 4.4). Refer also to Remark 5.1.3. Thus, the asymptotic properties of the on-line and off-line techniques will coincide. This discussion has been undoubtedly heuristic. On the contrary, it has presented the key steps of the asymptotic analysis of the suggested recursive estimator. The formal justification of the outlined approximations has been formed in the cited literature.

Recall that the off-line estimate minimizing (5.1) has the following asymptotic distribution (Ljung & Söderström, 1983, Section 4.4 and Theorem 4.5):

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_{[T]} - \boldsymbol{\theta}_0) \sim AsN(\boldsymbol{0}, \Gamma \boldsymbol{P}), \tag{5.74}$$

where $\widehat{\boldsymbol{\theta}}_{[T]}$ denotes the off-line prediction error method estimate evaluated using T corresponding observations (see e.g. Remark 5.1.3), $\widehat{\boldsymbol{\theta}}_{[T]} \to \boldsymbol{\theta}_0$ (the true value) almost surely as $T \to \infty$, Γ is defined as $\Gamma := \lim_{T\to\infty} T \sum_{t=1}^{T} \beta^2(T,t)$, and $\boldsymbol{P} := [\boldsymbol{G}(\boldsymbol{\theta}_0)]^{-1} \boldsymbol{Q}(\boldsymbol{\theta}_0) [\boldsymbol{G}(\boldsymbol{\theta}_0)]^{-1}$. Here, one (repeatedly) defines:

$$\boldsymbol{G}(\boldsymbol{\theta}_0) := \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[\boldsymbol{F}_t''(\boldsymbol{\theta}_0)\right] = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[\frac{\boldsymbol{\psi}_t(\boldsymbol{\theta}_0)\boldsymbol{\psi}_t^\top(\boldsymbol{\theta}_0)}{(\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0)^2}\right], \tag{5.75a}$$

$$\begin{aligned}\boldsymbol{Q}(\boldsymbol{\theta}_0) &:= \lim_{T\to\infty} \frac{1}{T} \sum_{s,t=1}^{T} \mathbb{E}\left[\nabla_{\boldsymbol{\theta}} F_t(\boldsymbol{\theta}_0)\right]^\top \left[\nabla_{\boldsymbol{\theta}} F_s(\boldsymbol{\theta}_0)\right] \\ &= \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[\frac{(\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0 - y_t^2)^2}{(\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0)^4} \boldsymbol{\psi}_t(\boldsymbol{\theta}_0)\boldsymbol{\psi}_t^\top(\boldsymbol{\theta}_0)\right],\end{aligned} \tag{5.75b}$$

which respect the discussion in Section 5.1 and Ljung and Söderström (1983, Section 4.4). One should clearly assume that $\boldsymbol{G}(\boldsymbol{\theta}_0)$ is invertible and that the limits defining $\boldsymbol{G}(\boldsymbol{\theta}_0)$ and $\boldsymbol{Q}(\boldsymbol{\theta}_0)$ exist.

5.3.4 Remarks on convergence analysis: The self-weighted recursive pseudo-linear regression algorithm

The self-weighted recursive pseudo-linear regression method (5.15) apparently coincides with the general recursive calibration scheme (5.40) in the analogous manner as before. However, the convergence analysis of this estimation procedure partially differs from the previous one presented in Section 5.3.3 due to the adopted approximation of the gradient $\widehat{\boldsymbol{\psi}}_t$. Compare Equations (5.13) and (5.15).

Assume that the conditions given by Theorem 5.3.1 are satisfied and postulate a certain error model (Ljung, 1999, Chapter 11):

$$\frac{y_t^2}{\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}} - 1 = H_0(B)\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})(\boldsymbol{\theta}_0 - \boldsymbol{\theta})/[\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}] + e_t \tag{5.76}$$

for some casual and strictly stable filter $H_0(B)$, where B denotes the lag operator, and some value $\boldsymbol{\theta}_0$, where $\{e_t\}_{t\in\mathbb{Z}}$ is a sequence of zero-mean random variables such that e_t is independent of $\boldsymbol{\varphi}_s(\boldsymbol{\theta})/(\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta})$ for all $s \leq t$ and all $\boldsymbol{\theta} \in \boldsymbol{D}_{\mathcal{M}}$. Furthermore, suppose that $\mathrm{Re}[H_0(e^{i\omega})] > \frac{1}{2}$ for all $\omega \in (-\pi, \pi]$ and that $\bar{\mathbb{E}}[\boldsymbol{\varphi}_t(\boldsymbol{\theta})\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})/(\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta})^2] =: \boldsymbol{G}(\boldsymbol{\theta})$ exists for all $\boldsymbol{\theta} \in \boldsymbol{D}_{\mathcal{M}}$. The operator $\bar{\mathbb{E}}(\boldsymbol{x})$ is defined as $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(\boldsymbol{x})$ for some random variable $\boldsymbol{x}$. Thus, the functions appearing in the ordinary differential equations associated with the algorithm become by using (5.15) and (5.76):

$$\begin{aligned}
\boldsymbol{f}(\boldsymbol{\theta}) &= \bar{\mathbb{E}}\left[\left(\frac{y_t^2}{\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}} - 1\right)\frac{\boldsymbol{\varphi}_t(\boldsymbol{\theta})}{\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}}\right] \\
&= \bar{\mathbb{E}}\left[\frac{\boldsymbol{\varphi}_t(\boldsymbol{\theta})\widetilde{\boldsymbol{\varphi}}_t^\top(\boldsymbol{\theta})}{(\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta})^2}(\boldsymbol{\theta}_0 - \boldsymbol{\theta})\right] = \widetilde{\boldsymbol{G}}(\boldsymbol{\theta})(\boldsymbol{\theta}_0 - \boldsymbol{\theta}), \qquad (5.77a) \\
\boldsymbol{F}(\boldsymbol{\theta}, \boldsymbol{R}) &= \boldsymbol{G}(\boldsymbol{\theta}) - \boldsymbol{R}, \qquad (5.77b)
\end{aligned}$$

where $\widetilde{\boldsymbol{\varphi}}_t(\boldsymbol{\theta}) = H_0(B)\boldsymbol{\varphi}_t(\boldsymbol{\theta})$ and $\widetilde{\boldsymbol{G}}(\boldsymbol{\theta}) = \bar{\mathbb{E}}[\boldsymbol{\varphi}_t(\boldsymbol{\theta})\widetilde{\boldsymbol{\varphi}}_t^\top(\boldsymbol{\theta})/(\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta})^2]$ (provided that all limits exist for all $\boldsymbol{\theta} \in \boldsymbol{D}_{\mathcal{M}}$). The corresponding ordinary differential equations are thus formulated as follows:

$$\begin{aligned}
\frac{d}{d\tau}\boldsymbol{\theta}_D(\tau) &= \boldsymbol{R}_D^{-1}(\tau)\widetilde{\boldsymbol{G}}(\boldsymbol{\theta}_D(\tau))\left[\boldsymbol{\theta}_0 - \boldsymbol{\theta}_D(\tau)\right], \qquad (5.78a) \\
\frac{d}{d\tau}\boldsymbol{R}_D(\tau) &= \boldsymbol{G}(\boldsymbol{\theta}_D(\tau)) - \boldsymbol{R}_D(\tau). \qquad (5.78b)
\end{aligned}$$

Assume $\widetilde{V}(\boldsymbol{\theta}, \boldsymbol{R}) = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \boldsymbol{R}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ as a candidate for the function V from Theorem 5.3.1. Clearly, $\frac{d}{d\tau}\widetilde{V}(\boldsymbol{\theta}_D(\tau), \boldsymbol{R}_D(\tau))$ is equal to:

$$-(\boldsymbol{\theta}_D(\tau) - \boldsymbol{\theta}_0)^\top \left[\widetilde{\boldsymbol{G}}(\boldsymbol{\theta}_D(\tau)) + \widetilde{\boldsymbol{G}}^\top(\boldsymbol{\theta}_D(\tau)) - \boldsymbol{G}(\boldsymbol{\theta}_D(\tau)) + \boldsymbol{R}_D(\tau)\right](\boldsymbol{\theta}_D(\tau) - \boldsymbol{\theta}_0).$$

It means that $\widetilde{V}$ will be a suitable function fulfilling Theorem 5.3.1 if the matrix $\boldsymbol{Z} = \boldsymbol{Z}(\boldsymbol{\theta}, \boldsymbol{R}) := \widetilde{\boldsymbol{G}}(\boldsymbol{\theta}) + \widetilde{\boldsymbol{G}}^{\top}(\boldsymbol{\theta}) - \boldsymbol{G}(\boldsymbol{\theta}) + \boldsymbol{R}$ is positive semidefinite. Remind that $\boldsymbol{R}$ fulfils the condition (R1). Moreover, let $\boldsymbol{x}$ be an arbitrary real vector of the corresponding dimension. Put $\boldsymbol{x}(\boldsymbol{\theta}) := \boldsymbol{x}^{\top}\boldsymbol{\varphi}_t(\boldsymbol{\theta})/(\boldsymbol{\varphi}_t^{\top}(\boldsymbol{\theta})\boldsymbol{\theta})$ and analogously $\widetilde{\boldsymbol{x}}(\boldsymbol{\theta}) := \boldsymbol{x}^{\top}\widetilde{\boldsymbol{\varphi}}_t(\boldsymbol{\theta})/(\boldsymbol{\varphi}_t^{\top}(\boldsymbol{\theta})\boldsymbol{\theta})$. Hence, we may compute:

$$\begin{aligned}\boldsymbol{x}^{\top}\boldsymbol{Z}\boldsymbol{x} &= 2\bar{\mathbb{E}}\left[\boldsymbol{x}(\boldsymbol{\theta})\left(\widetilde{\boldsymbol{x}}^{\top}(\boldsymbol{\theta}) - 0.5\boldsymbol{x}^{\top}(\boldsymbol{\theta})\right)\right] + \boldsymbol{x}^{\top}\boldsymbol{R}\boldsymbol{x} \\ &= 2\int_{-\pi}^{\pi}\boldsymbol{\Phi}_{\boldsymbol{x}}(\omega)\mathrm{Re}\left[H_0(e^{i\omega}) - \frac{1}{2}\right]d\omega + \boldsymbol{x}^{\top}\boldsymbol{R}\boldsymbol{x} \geq 0,\end{aligned}$$

where $\boldsymbol{\Phi}_{\boldsymbol{x}}(\omega)$ denotes the spectral density of $\boldsymbol{x}(\boldsymbol{\theta})$. See also Ljung and Söderström (1983, Section 4.5.2). Equality holds only for $\boldsymbol{\Phi}_{\boldsymbol{x}}(\omega) \equiv 0$, i.e. for $\bar{\mathbb{E}}[\boldsymbol{x}(\boldsymbol{\theta})]^2 = 0$. Following Theorem 5.3.1, we can set:

$$\boldsymbol{D}_{\mathcal{C}} = \left\{\boldsymbol{\theta}\,\middle|\,\bar{\mathbb{E}}\left[(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^{\top}\boldsymbol{\varphi}_t(\boldsymbol{\theta})/(\boldsymbol{\varphi}_t^{\top}(\boldsymbol{\theta})\boldsymbol{\theta})\right]^2 = 0\right\}, \tag{5.80}$$

i.e. $\boldsymbol{x}(\boldsymbol{\theta})$ is given by the particular choice of $\boldsymbol{x}$, namely $\boldsymbol{x} := \boldsymbol{\theta} - \boldsymbol{\theta}_0$. Employing (5.76), $\boldsymbol{D}_{\mathcal{C}}$ can be reformulated as $\{\boldsymbol{\theta}|\bar{\mathbb{E}}[y_t^2/(\boldsymbol{\varphi}_t^{\top}(\boldsymbol{\theta})\boldsymbol{\theta}) - 1 - e_t]^2 = 0\}$. According to Theorem 5.3.1, one may conclude that the estimates $\widehat{\boldsymbol{\theta}}_t$ converge either to $\boldsymbol{D}_{\mathcal{C}}$ or to the boundary of $\boldsymbol{D}_{\mathcal{M}}$ (almost surely as t grows to infinity).

Remark 5.3.4 From (5.13), it holds that $\widehat{\boldsymbol{\psi}}_{t+1} = \widehat{\boldsymbol{\varphi}}_{t+1} + \sum_{j=1}^{q}\widehat{\beta}_{j,t}\widehat{\boldsymbol{\psi}}_{t+1-j}$, i.e. $(B^0 - \sum_{j=1}^{q}\widehat{\beta}_{j,t}B^j)\widehat{\boldsymbol{\psi}}_{t+1} = \widehat{\boldsymbol{\varphi}}_{t+1}$. Thus, one can set $H_0(z) = \frac{1}{C_0(z)}$ in (5.76), where $C_0(z) = 1 - \sum_{j=1}^{q}\beta_{j,0}z^j$ using the true parameters (Ljung, 1999, Chapter 11).

Remark 5.3.5 The ordinary differential equations (5.78a) and (5.78b) associated with the estimation algorithm (5.15) may be linearized around a stationary point $(\boldsymbol{\theta}^*, \boldsymbol{R}^*)$ as in Remark 5.3.2. It has been found that the stability properties of the linearized equations are determined by the matrix $-[\boldsymbol{G}(\boldsymbol{\theta}^*)]^{-1}\widetilde{\boldsymbol{G}}(\boldsymbol{\theta}^*)$. If this matrix has eigenvalues in the right half-plane, then $\widehat{\boldsymbol{\theta}}_t$ cannot converge to a stationary point $\boldsymbol{\theta}^*$ (Söderström & Stoica, 1989, Section 9.6). In some special cases, the eigenvalues of the matrix can be expressed explicitly; therefore, one can establish conditions when the estimate $\widehat{\boldsymbol{\theta}}_t$ cannot converge to its true value.

Remark 5.3.6 Analysis of asymptotic properties of the recursive algorithm (5.15) is not as straightforward as for the recursive scheme (5.13). From Remark 5.3.3, it has been obvious the crucial property that $-\widehat{\boldsymbol{\psi}}_t\widehat{\boldsymbol{\psi}}_t^{\top}/(\widehat{\boldsymbol{\varphi}}_t^{\top}\widehat{\boldsymbol{\theta}}_{t-1})^2$ is approximately the Jacobian matrix of $[(y_t^2 - \widehat{\boldsymbol{\varphi}}_t^{\top}\widehat{\boldsymbol{\theta}}_{t-1})\widehat{\boldsymbol{\psi}}_t]^{\top}/(\widehat{\boldsymbol{\varphi}}_t^{\top}\widehat{\boldsymbol{\theta}}_{t-1})^2$. When $\widehat{\boldsymbol{\psi}}_t$ is replaced by $\widehat{\boldsymbol{\varphi}}_t$, the heuristic analysis loses the key argument. The asymptotic features of the recursive pseudo-linear regression algorithms have not been established in general (Ljung & Söderström, 1983, Section 4.5.5).

5.4 Modifications of self-weighted recursive estimation algorithms

In Section 5.1, two distinct one-stage self-weighted recursive estimation methods have been suggested for estimating the parameters of the GARCH models. The algorithms have been derived by using the general identification instruments and the loss function introduced by (5.1). Section 5.2 has discussed several practical aspects of the implementation of such techniques (e.g. the initialization in Section 5.2.1 or the robustification in Section 5.2.6). Section 5.3 has remarked on theoretical features of the presented procedures, namely on the convergence and asymptotic properties.

With regards to the generality of the core criterion $F_t(\boldsymbol{\theta})$ defined by Equation (5.2), one can simply replicate all foregoing steps to derive and justify the following one-stage self-weighted recursive estimation procedures for the selected classic (linear) conditional heteroscedasticity models. Here, the linearity means that $\sigma_t^2(\boldsymbol{\theta})$ is defined as a linear function of previous observations and conditional variances. See Chapter 3 for further details. Particularly, the (negative) conditional log-likelihood criteria for those models (assuming the normally distributed disturbances) almost coincide since one may express $\sigma_t^2(\boldsymbol{\theta})$ as $\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}$. Only the components of the vectors $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}_t(\boldsymbol{\theta})$ differ accordingly. See below. Therefore, the on-line estimation techniques originally proposed for the GARCH parameters can be updated for calibrating the selected conditional heteroscedasticity models. The following sections introduce such modifications for: (i) the ARCH models, (ii) the RiskMetrics EWMA models, and finally (iii) the GJR-GARCH models.

5.4.1 Recursive estimation of ARCH models

Updating the recursive schemes (5.14) and (5.16) for the case of the ARCH(p) process $\{y_t\}_{t\in\mathbb{Z}}$ given by (3.1) is relatively straightforward since one clearly puts $\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \alpha_p)^\top$ and $\boldsymbol{\varphi}_t(\boldsymbol{\theta}) = (1, y_{t-1}^2, \ldots, y_{t-p}^2)^\top$. Here, it is evident that $\boldsymbol{\varphi}_t(\boldsymbol{\theta})$ depends only on the previous measurements (and not on the value of $\boldsymbol{\theta}$). See Section 3.1 for more details. Additionally, it is possible to verify that:

$$\boldsymbol{\psi}_t(\boldsymbol{\theta}) = \left[\nabla_{\boldsymbol{\theta}}\boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}\right]^\top = \boldsymbol{\varphi}_t(\boldsymbol{\theta}). \tag{5.81}$$

Therefore, the recursive prediction error method and the recursive pseudo-linear regression algorithm coincide for the ARCH(p) model. Consequently, the

ARCH parameters may be recurrently estimated using the following procedure:

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1} + \frac{\widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\varphi}}_t(y_t^2 - \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})}{\lambda_t(\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})^2 + \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\varphi}}_t}, \tag{5.82a}$$

$$\widehat{\boldsymbol{P}}_t = \frac{1}{\lambda_t}\left\{\widehat{\boldsymbol{P}}_{t-1} - \frac{\widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\varphi}}_t\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{P}}_{t-1}}{\lambda_t(\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})^2 + \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\varphi}}_t}\right\}, \tag{5.82b}$$

$$\widehat{\boldsymbol{\varphi}}_{t+1} = (1, y_t^2, \ldots, y_{t+1-p}^2)^\top,\ t \in \mathbb{N}. \tag{5.82c}$$

We recommend initializing this algorithm with: (i) $\widehat{\boldsymbol{P}}_0 = \kappa \boldsymbol{I}$ for a large positive κ (e.g. $\kappa = 10^5$), (ii) $\widehat{\boldsymbol{\theta}}_0$ is given by a burn-in, e.g. $\widehat{\boldsymbol{\theta}}_0 = (\frac{1}{n}\sum_{i=1}^n y_i^2(1 - p\varepsilon), \varepsilon, \ldots, \varepsilon)^\top$ for some $0 < \varepsilon < 1/p$ and an adequate $n \in \mathbb{N}$, finally (iii) the values $\{y_{1-p}^2, \ldots, y_0^2\}$ are either defined or observed. The set defining the projection (5.36) is a compact subset of $\{\boldsymbol{\theta} \in \mathbb{R}^{p+1} | \omega > 0, \alpha_i \geq 0, \sum_{i=1}^p \alpha_i < 1\}$, which ensures stationarity and positivity of $\sigma_t^2(\boldsymbol{\theta})$. For instance, one may put $\boldsymbol{D}_{\mathcal{M}} := \{\boldsymbol{\theta} \in \mathbb{R}^{p+1} | \delta_\omega \leq \omega \leq \Delta_\omega, \alpha_i \geq 0, \sum_{i=1}^p \alpha_i \leq 1 - \delta_\alpha\}$ with some given $0 < \delta_\omega \leq \Delta_\omega < \infty$ and $0 < \delta_\alpha < 1$. Other practical aspects and theoretical properties can be simply deduced from Sections 5.2 and 5.3, respectively. Section 6.2 demonstrates capabilities of this method by means of numerical simulations.

5.4.2 Recursive estimation of EWMA models

To adjust the recursive calibration algorithms (5.14) and (5.16) to follow the RiskMetrics EWMA process $\{y_t\}_{t\in\mathbb{Z}}$ defined by (3.10), one can recognize that formally $\boldsymbol{\theta} = (1 - \lambda, \lambda)^\top$ and $\boldsymbol{\varphi}_t(\boldsymbol{\theta}) = (y_{t-1}^2, \boldsymbol{\varphi}_{t-1}^\top(\boldsymbol{\theta})\boldsymbol{\theta})^\top$. See Section 3.2.

According to the structure of $\boldsymbol{\theta}$, we may formulate the simplified version of the self-weighted recursive prediction error method algorithm as follows:

$$\widehat{\lambda}_t = \widehat{\lambda}_{t-1} + \frac{\widehat{p}_{t-1}(y_t^2 - \widehat{\sigma}_t^2)\widehat{\sigma}_t^{2'}}{\lambda_t(\widehat{\sigma}_t^2)^2 + (\widehat{\sigma}_t^{2'})^2\widehat{p}_{t-1}}, \tag{5.83a}$$

$$\widehat{p}_t = \frac{1}{\lambda_t}\left\{\widehat{p}_{t-1} - \frac{\widehat{p}_{t-1}^2(\widehat{\sigma}_t^{2'})^2}{\lambda_t(\widehat{\sigma}_t^2)^2 + (\widehat{\sigma}_t^{2'})^2\widehat{p}_{t-1}}\right\}, \tag{5.83b}$$

$$\widehat{\sigma}_{t+1}^2 = (1 - \widehat{\lambda}_t)y_t^2 + \widehat{\lambda}_t\widehat{\sigma}_t^2, \tag{5.83c}$$

$$\widehat{\sigma}_{t+1}^{2'} = -y_t^2 + \widehat{\sigma}_t^2 + \widehat{\lambda}_t\widehat{\sigma}_t^{2'},\ t \in \mathbb{N}, \tag{5.83d}$$

where for the sake of clarity $\widehat{\lambda}_t$ denotes the recursive estimate of the EWMA modelling parameter λ, and λ_t stands for the forgetting factor as usual. Similarly as in (5.16), we could consider the recursive pseudo-linear regression form of the previous algorithm. However, it seems as a needless simplification in the case of this elementary recursive scheme.

We recommend starting the algorithm (5.83) with the following settings: (i) $\widehat{p}_0$ is a large positive number, e.g. $\widehat{p}_0 = 10^5$, (ii) $\widehat{\lambda}_0$ can be taken from (0,1), e.g. as 0.94 as is usually preferred by RiskMetrics (for daily data), finally (iii) $\widehat{\sigma}_1^2$ is a positive number, and $\widehat{\sigma}_1^{2'}$ is equal to zero. The set defining the necessary projection (5.36) is a compact subset of $\{\lambda \in \mathbb{R} | \lambda \in (0,1)\}$, which guarantees positivity of $\sigma_t^2(\lambda)$. For instance, one may put $\boldsymbol{D}_{\mathcal{M}} := \{\lambda \in \mathbb{R} | \lambda \in [\delta_\lambda, 1-\delta_\lambda]\}$ with some prescribed $\delta_\lambda \in (0, 1/2]$. Other practical aspects and theoretical properties can be simply deduced from Sections 5.2 and 5.3, respectively. Section 6.4.1 numerically investigates this estimation technique.

5.4.3 Recursive estimation of GJR-GARCH models

Section 3.3 has introduced the GJR-GARCH process $\{y_t\}_{t\in\mathbb{Z}}$ as a GARCH variant that includes leverage terms for modelling asymmetric volatility clustering. In the GJR-GARCH representation (3.11), large negative changes are more likely to be clustered than positive ones. See also Chapter 1. Similarly as before, it is possible to adjust the recursive formulas (5.14) and (5.16) to estimate the parameters of the GJR-GARCH(p,q) model. Particularly, we put:

$$\boldsymbol{\theta} = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \gamma_1, \ldots, \gamma_p)^\top, \tag{5.84a}$$

$$\boldsymbol{\varphi}_t(\boldsymbol{\theta}) = \big(1, y_{t-1}^2, \ldots, y_{t-p}^2, \boldsymbol{\varphi}_{t-1}^\top(\boldsymbol{\theta})\boldsymbol{\theta}, \ldots, \boldsymbol{\varphi}_{t-q}^\top(\boldsymbol{\theta})\boldsymbol{\theta}, \tag{5.84b}$$

$$y_{t-1}^2 I_{t-1}^-, \ldots, y_{t-p}^2 I_{t-p}^-\big)^\top, \tag{5.84c}$$

where $I_{t-i}^- = 1$ if $y_{t-i} < 0$ and 0 otherwise for $i = 1, \ldots, p$. See Section 3.3 for further insights. The transposed gradient $\boldsymbol{\psi}_t(\boldsymbol{\theta})$ can be readily evaluated as:

$$\boldsymbol{\psi}_t(\boldsymbol{\theta}) = \big[\nabla_{\boldsymbol{\theta}} \boldsymbol{\varphi}_t^\top(\boldsymbol{\theta})\boldsymbol{\theta}\big]^\top = \boldsymbol{\varphi}_t(\boldsymbol{\theta}) + \sum_{j=1}^{q} \beta_j \boldsymbol{\psi}_{t-j}(\boldsymbol{\theta}). \tag{5.85}$$

Therefore, the self-weighted recursive prediction error method adequate for the GJR-GARCH(p,q) model is given in the following form:

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1} + \frac{\widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\psi}}_t(y_t^2 - \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})}{\lambda_t(\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})^2 + \widehat{\boldsymbol{\psi}}_t^\top \widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\psi}}_t}, \tag{5.86a}$$

$$\widehat{\boldsymbol{P}}_t = \frac{1}{\lambda_t}\left\{\widehat{\boldsymbol{P}}_{t-1} - \frac{\widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\psi}}_t\widehat{\boldsymbol{\psi}}_t^\top\widehat{\boldsymbol{P}}_{t-1}}{\lambda_t(\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})^2 + \widehat{\boldsymbol{\psi}}_t^\top \widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\psi}}_t}\right\}, \tag{5.86b}$$

$$\widehat{\boldsymbol{\varphi}}_{t+1} = (1, y_t^2, \ldots, y_{t+1-p}^2, \widehat{\boldsymbol{\varphi}}_t^\top\widehat{\boldsymbol{\theta}}_t, \ldots, \widehat{\boldsymbol{\varphi}}_{t+1-q}^\top\widehat{\boldsymbol{\theta}}_{t+1-q},$$
$$y_t^2 I_t^-, \ldots, y_{t+1-p}^2 I_{t+1-p}^-)^\top, \tag{5.86c}$$

$$\widehat{\boldsymbol{\psi}}_{t+1} = \widehat{\boldsymbol{\varphi}}_{t+1} + \sum_{j=1}^{q} \widehat{\beta}_{j,t} \widehat{\boldsymbol{\psi}}_{t+1-j},\ t \in \mathbb{N}. \tag{5.86d}$$

The numerically more effective alternative, i.e. the recursive pseudo-linear regression, is thus formulated as follows by accepting $\widehat{\boldsymbol{\psi}}_t \approx \widehat{\boldsymbol{\varphi}}_t$:

$$\widehat{\boldsymbol{\theta}}_t = \widehat{\boldsymbol{\theta}}_{t-1} + \frac{\widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\varphi}}_t(y_t^2 - \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})}{\lambda_t(\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})^2 + \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\varphi}}_t}, \tag{5.87a}$$

$$\widehat{\boldsymbol{P}}_t = \frac{1}{\lambda_t}\left\{\widehat{\boldsymbol{P}}_{t-1} - \frac{\widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\varphi}}_t\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{P}}_{t-1}}{\lambda_t(\widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_{t-1})^2 + \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{P}}_{t-1}\widehat{\boldsymbol{\varphi}}_t}\right\}, \tag{5.87b}$$

$$\begin{aligned}\widehat{\boldsymbol{\varphi}}_{t+1} = (1, y_t^2, \ldots, y_{t+1-p}^2, \widehat{\boldsymbol{\varphi}}_t^\top \widehat{\boldsymbol{\theta}}_t, \ldots, \widehat{\boldsymbol{\varphi}}_{t+1-q}^\top \widehat{\boldsymbol{\theta}}_{t+1-q},\\ y_t^2 I_t^-, \ldots, y_{t+1-p}^2 I_{t+1-p}^-)^\top,\ t \in \mathbb{N}.\end{aligned} \tag{5.87c}$$

Initial conditions for both algorithms may be: (i) $\widehat{\boldsymbol{P}}_0$ equals $\kappa \boldsymbol{I}$ for a large positive κ (e.g. $\kappa = 10^5$) and (ii) $\widehat{\boldsymbol{\theta}}_0$ is given by a burn-in, e.g. $\widehat{\boldsymbol{\theta}}_0 = (\frac{1}{n}\sum_{i=1}^n y_i^2[1-(p+q)\varepsilon], \varepsilon, \ldots, \varepsilon, 0, \ldots, 0)^\top$ for some $0 < \varepsilon < 1/(p+q)$, an adequate $n \in \mathbb{N}$, and $\widehat{\gamma}_{i,0} = 0$ for all $i = 1, \ldots, p$. The remaining initial quantities are selected similarly as in Section 5.2.1. The set defining the projection (5.36) is a compact subset of $\{\boldsymbol{\theta} \in \mathbb{R}^{2p+q+1} | \omega > 0, \alpha_i \geq 0, \beta_j \geq 0, \alpha_i + \gamma_i \geq 0, \sum_{i=1}^p (\alpha_i + \gamma_i/2) + \sum_{j=1}^q \beta_j < 1\}$, which ensures stationarity and positivity of $\sigma_t^2(\boldsymbol{\theta})$. For instance, one may use $\boldsymbol{D}_{\mathcal{M}} := \{\boldsymbol{\theta} \in \mathbb{R}^{2p+q+1} | \delta_\omega \leq \omega \leq \Delta_\omega, \alpha_i \geq 0, \beta_j \geq 0, \alpha_i + \gamma_i \geq 0, \sum_{i=1}^p (\alpha_i + \gamma_i/2) + \sum_{j=1}^q \beta_j \leq 1 - \delta_{\alpha\beta\gamma}\}$ with some suitable $0 < \delta_\omega \leq \Delta_\omega < \infty$ and $0 < \delta_{\alpha\beta\gamma} < 1$. Other practical aspects and theoretical properties can be simply transferred from Sections 5.2 and 5.3, respectively. Section 6.3 demonstrates numerical capabilities of these estimation methods.

Remark 5.4.1 The derivation scheme presented in Section 5.1 may be generalized in order to deliver reliable recursive estimation techniques also for other (either linear or nonlinear) conditional heteroscedasticity models (i.e. $\sigma_t^2(\boldsymbol{\theta})$ is either or not a linear function of previous measurements and conditional variances), e.g. the asymmetric power ARCH (APARCH), the exponential GARCH (EGARCH), and other analogous processes (Tsay, 2005, Chapter 3).

6. Numerical studies

Chapters 4 and 5 have surveyed the different recursive methods designed for estimating parameters of the selected classic conditional heteroscedasticity models. Especially, we have focused on the GARCH model and its variants. Although many theoretical and practical aspects concerning these approaches have been discussed thoroughly (see Sections 4.1, 4.2, 5.1, 5.2, and 5.3), they usually do not provide any detail how large t has to be for ensuring the results to be (reasonably) applicable in practice. To get some insight into the convergence rate, the transient behaviour, and the finite-sample properties, the analysis should be complemented by numerical studies (Ljung & Söderström, 1983, Section 4.1).

Therefore, this section verifies the introduced recurrent estimation formulas by means of various numerical experiments. Sections 6.1, 6.2, 6.3, and 6.4.1 examine the convergence performance of the algorithms developed for calibrating the GARCH, ARCH, GJR-GARCH, and RiskMetrics EWMA processes, respectively, using Monte Carlo simulations. Section 6.4.2 investigates the ability of the estimation technique considered for the EWMA model to track parameters that vary over time. Moreover, Section 6.4.3 demonstrates the adequacy of the robustification suggested in Section 5.2.6 in the context of the EWMA modelling framework. Finally, Sections 6.5 and 6.6 employ some of the introduced techniques to analyse two different empirical datasets in detail. Firstly, the PX index (daily) returns are studied by applying the instruments discussed in Section 6.5. The outputs show that the outlined concepts can be helpful also in the context of low-frequency data. Secondly, volatility of high-frequency data (tick EUR/USD exchange rates) is monitored and predicted in real time. This example illustrates one of the possible practical applications of the recursive estimation methodology.

6.1 On-line estimation of GARCH models: Monte Carlo study

This section examines various recursive estimation algorithms developed for calibrating the parameters of the GARCH processes using Monte Carlo experiments. See Section 3.1. Namely, we shall compare: (i) both two-stage on-line calibration methods suggested by Aknouche and Guerbyenne (2006) summarized by Equations (4.9)-(4.10) and (4.12)-(4.13), (ii) both proposed one-stage self-weighted recursive variants (5.14) and (5.16) suggested in Section 5.1. Note that the recurrent procedures developed by Kierkegaard et al. (2000) delivered in Section 4.1

are not involved since they do not sufficiently reflect the heteroscedastic structure of the innovation term in (4.2) as we remarked before.

All estimation procedures have been compared by performing various simulation experiments in the framework of the most common GARCH(1,1) process with normally distributed innovations. The calibration schemes introduced in Section 4.2 were adjusted to produce fair comparison by adding the identical forgetting factor λ_t, using the same initial conditions, applying the similar regularization and factorization techniques, and employing the coincident projection. See below. As we indicated before, the considered recursive estimation methods have demonstrated different convergence behaviour. This conclusion should be taken into consideration in empirical applications.

In the presented Monte Carlo study, various simulations were realized. In detail, time series with various configurations of the GARCH(1,1) parameters $\boldsymbol{\theta} = (\omega, \alpha_1, \beta_1)^\top$ were repeatedly generated to test the described algorithms. Different values of κ appearing in $\widetilde{\boldsymbol{P}}_0$ or $\widehat{\boldsymbol{P}}_0$ and various combinations of $\widetilde{\lambda}$ and λ_0 generating the deterministic forgetting factor sequence (that grows to one as t goes to infinity) were examined. However, the obtained conclusions remained almost identical. Therefore, only three representative settings are reviewed below.

Additionally, we have inspected the numerical behaviour of the methods using two different groups of the initial conditions for $\widetilde{\boldsymbol{\theta}}_0$ and $\widehat{\boldsymbol{\theta}}_0$. Firstly, the starting values recommended in Section 4.2 were slightly modified since the initialization vectors $(0, 0, 0)^\top$ do not lie in the region of positivity and stationarity. See the discussion in Section 4.2. Accordingly, the original settings could influence the quality of convergence or its speed since the accepted projection algorithm does not ensure shifting the current estimate into the region of stability $\boldsymbol{D}_{\mathcal{M}}$. See below. Therefore, all estimation algorithms were initialized by $(10^{-3}, 10^{-3}, 10^{-3})^\top$, i.e. $\widetilde{\boldsymbol{\theta}}_0 = \widehat{\boldsymbol{\theta}}_0 \approx \mathbf{0}$. Secondly, the initialization proposed in Section 5.2.1 was employed in the described way with the burn-in $n = 60$ and $\varepsilon = 0.1$. Namely, the first $n = 60$ observations produced $\widetilde{\boldsymbol{\theta}}_0$ or $\widehat{\boldsymbol{\theta}}_0$, and after that the proper estimation procedures were started. For example, $n = 60$ corresponds to a one-minute dataset supposing one-second observations.

In particular, three GARCH(1,1) processes (3.1) with Gaussian innovations were replicated 1000 times with the following specification of the true parameters:

(i) $(\omega, \alpha_1, \beta_1) = (0.30, 0.05, 0.80)$,

(ii) $(\omega, \alpha_1, \beta_1) = (0.50, 0.20, 0.50)$,

(iii) $(\omega, \alpha_1, \beta_1) = (0.05, 0.05, 0.94)$.

Obviously, these parameters determine the GARCH(1,1) models with different levels of persistence and unconditional variances. We investigated the simulated

time series of the length 10000, i.e. approximately three-hours datasets working with one-second data. All computations were conducted in the statistical software `R` by introducing original algorithms. The estimation procedures defined by (4.9)-(4.10), (4.12)-(4.13), (5.14), and finally (5.16) were implemented using these conditions: (i) $\widetilde{\boldsymbol{P}}_0 = \widehat{\boldsymbol{P}}_0 = 10^5\boldsymbol{I}$, (ii) the two variants of $\widetilde{\boldsymbol{\theta}}_0$ and $\widehat{\boldsymbol{\theta}}_0$ discussed in the previous paragraph, (iii) $\widetilde{\boldsymbol{\varphi}}_1 = \widehat{\boldsymbol{\varphi}}_1 = (1,0,0)^\top$, $\widetilde{\boldsymbol{\psi}}_1 = \widetilde{\boldsymbol{\varphi}}_1$, and $\widehat{\boldsymbol{\psi}}_1 = \widehat{\boldsymbol{\varphi}}_1$. The examined algorithms were adjusted to adopt the similar principles of practical implementation (see above and Section 5.2). Except for the already mentioned initial values, all tested estimation schemes included the forgetting factor sequence given as $\lambda_t = 0.99\lambda_{t-1} + (1 - 0.99)$, $\lambda_0 = 0.95$ and $t \in \mathbb{N}$. They were adapted to respect the factorization and regularization described in Section 5.2.3. The projection necessary for controlling each estimate was defined by (5.36) with $\boldsymbol{D}_{\mathcal{M}} = \{\boldsymbol{\theta} \in \mathbb{R}^3 | 10^{-9} \leq \omega \leq 10^3, \alpha_1 \geq 0, \beta_1 \geq 0, \alpha_1 + \beta_1 \leq 1 - 10^{-9}\}$, which ensures positivity and stationarity of the conditional variance. Correspondingly, we should produce fair comparisons since all estimation techniques satisfied the equal conditions.

Consequently, we have compared the following four methods under the conditions described in the foregoing paragraphs: (i) the two-stage recursive pseudo-linear regression method (2S-RPLR) given by (4.9) and (4.10), (ii) the two-stage recursive prediction error method (2S-RPEM) delivered by (4.12) and (4.13), (iii) the self-weighted recursive pseudo-linear regression algorithm (SW-RPLR) as in (5.16), and finally (iv) the self-weighted recursive prediction error method algorithm (SW-RPEM) introduced by (5.14).

Tables 6.1, 6.2, and 6.3 summarize the results of the described Monte Carlo experiments. In particular, the medians of absolute errors of the final estimates are presented here (the different initializations are clearly distinguished). Note that this specific measure of performance is used because there were observed a few (degenerated) estimates that achieved the boundary of $\boldsymbol{D}_{\mathcal{M}}$. They could negatively influence any performance characteristic based on the arithmetic mean. Firstly, it is clear that the self-weighted recursive algorithms suggested in Section 5.1 are truly competitive. Secondly, comparing all delivered results, one may conclude that the initialization originally recommended for the methods discussed in Section 4.2 should be reconsidered (Hendrych, 2014b). On the other hand, the initialization introduced in Section 5.2.1 has improved the convergence behaviour. Consequently, the self-weighted recursive prediction error method algorithm initialized by the burn-in could be regarded as the best technique in the tested Monte Carlo framework.

Figures 6.1, 6.2, and 6.3 illustrate the convergence of the self-weighted recursive prediction error estimates derived in Section 5.1 to the true values jointly with

their decreasing variances. Namely, we simulated 1000 time series of the length 20000 and stopped the recursive estimation process at the times 1000, 5000, 10000, and 20000 (the other experimental conditions remained as before). The results displayed by the associated boxplots respect the remarks in Section 5.3. On the contrary, it is apparent that the convergence speed differs amongst these cases (Ljung, 1999).

Method	$\tilde{\boldsymbol{\theta}}_0 = \hat{\boldsymbol{\theta}}_0 \approx \mathbf{0}$			$\tilde{\boldsymbol{\theta}}_0 = \hat{\boldsymbol{\theta}}_0$ given by burn-in		
	ω	α_1	β_1	ω	α_1	β_1
2S-RPLR	0.61798	0.04999	0.79999	0.27927	0.00928	0.14722
2S-RPEM	0.61798	0.04999	0.79999	0.10057	0.00828	0.05575
SW-RPLR	0.61798	0.01616	0.31128	0.21856	0.00800	0.11236
SW-RPEM	0.61798	0.02468	0.32527	**0.06902**	**0.00660**	**0.03871**

Table 6.1: Simulation results for the GARCH(1,1) process of the length 10000 with $(\omega, \alpha_1, \beta_1) = (0.30, 0.05, 0.80)$. Medians of absolute errors of the final estimates.

Method	$\tilde{\boldsymbol{\theta}}_0 = \hat{\boldsymbol{\theta}}_0 \approx \mathbf{0}$			$\tilde{\boldsymbol{\theta}}_0 = \hat{\boldsymbol{\theta}}_0$ given by burn-in		
	ω	α_1	β_1	ω	α_1	β_1
2S-RPLR	0.48686	0.19999	0.49999	0.05050	0.01126	0.03641
2S-RPEM	0.48686	0.19999	0.49999	0.03982	0.01148	0.02953
SW-RPLR	0.48686	0.03102	0.34632	0.04620	**0.01034**	0.03500
SW-RPEM	0.48685	0.08615	0.41641	**0.03649**	0.01062	**0.02812**

Table 6.2: Simulation results for the GARCH(1,1) process of the length 10000 with $(\omega, \alpha_1, \beta_1) = (0.50, 0.20, 0.50)$. Medians of absolute errors of the final estimates.

Method	$\tilde{\boldsymbol{\theta}}_0 = \hat{\boldsymbol{\theta}}_0 \approx \mathbf{0}$			$\tilde{\boldsymbol{\theta}}_0 = \hat{\boldsymbol{\theta}}_0$ given by burn-in		
	ω	α_1	β_1	ω	α_1	β_1
2S-RPLR	1.99601	0.04999	0.93999	0.14409	0.02083	0.05376
2S-RPEM	1.99601	0.04999	0.93999	0.01813	0.00601	0.01027
SW-RPLR	1.99601	0.04280	0.45648	0.07882	0.01252	0.02897
SW-RPEM	1.99601	0.04995	0.49192	**0.00955**	**0.00397**	**0.00529**

Table 6.3: Simulation results for the GARCH(1,1) process of the length 10000 with $(\omega, \alpha_1, \beta_1) = (0.05, 0.05, 0.94)$. Medians of absolute errors of the final estimates.

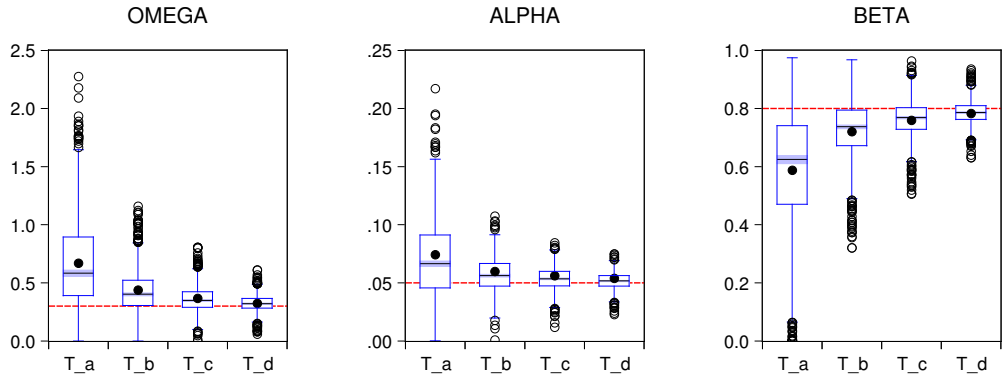


Figure 6.1: Boxplots of the SW-RPEM estimates of the GARCH(1,1) process with $(\omega, \alpha_1, \beta_1) = (0.30, 0.05, 0.80)$. The recursive estimation process was stopped at the times $T_a = 1000$, $T_b = 5000$, $T_c = 10000$, and $T_d = 20000$.

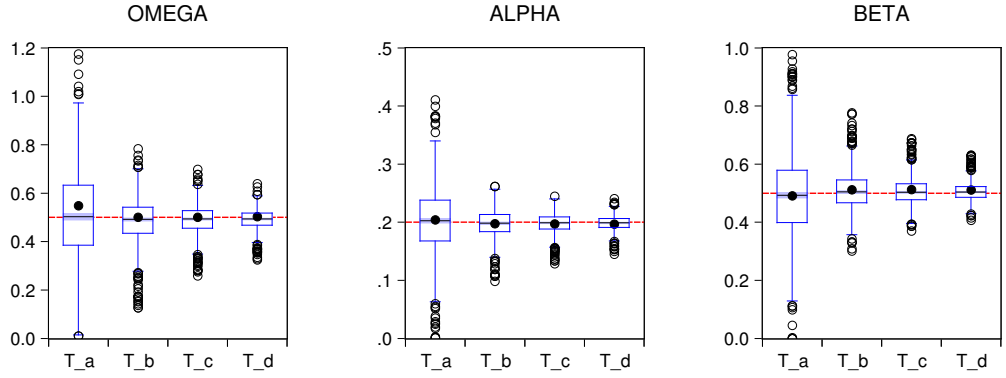


Figure 6.2: Boxplots of the SW-RPEM estimates of the GARCH(1,1) process with $(\omega, \alpha_1, \beta_1) = (0.50, 0.20, 0.50)$. The recursive estimation process was stopped at the times $T_a = 1000$, $T_b = 5000$, $T_c = 10000$, and $T_d = 20000$.

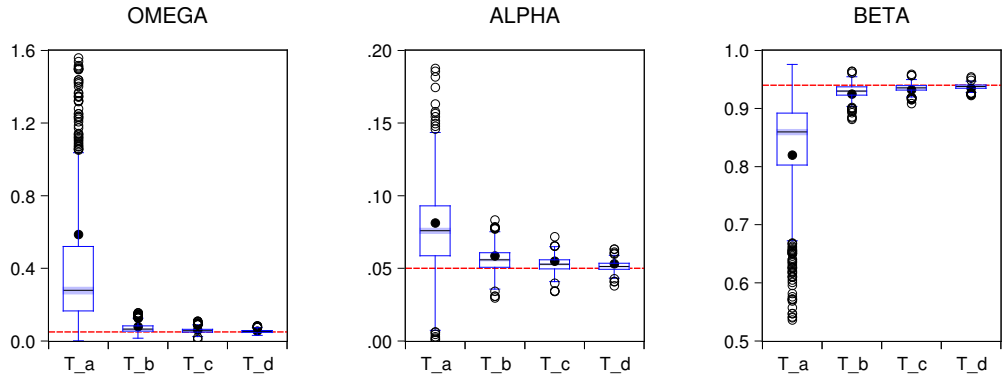


Figure 6.3: Boxplots of the SW-RPEM estimates of the GARCH(1,1) process with $(\omega, \alpha_1, \beta_1) = (0.05, 0.05, 0.94)$. The recursive estimation process was stopped at the times $T_a = 1000$, $T_b = 5000$, $T_c = 10000$, and $T_d = 20000$.

6.2 On-line estimation of ARCH models: Monte Carlo study

This section deals with analogous Monte Carlo experiments as Section 6.1. We shall demonstrate that the suggested recursive algorithm (5.82) is capable of estimating the ARCH parameters suitably. See also Section 3.1. Only the self-weighted recursive prediction error method (SW-RPEM) is investigated since it is completely identical to the recursive pseudo-linear regression algorithm due to the structure of the data generating process. Various simulations were performed using many different configurations of the true parameters. Results remained almost similar. Therefore, only three representative cases are presented below.

In particular, three ARCH(1) processes (3.1) with Gaussian disturbances of the length 20000 were replicated 1000 times with the following true parameters: (i) $(\omega, \alpha_1) = (0.10, 0.05)$, (ii) $(\omega, \alpha_1) = (0.10, 0.50)$, and (iii) $(\omega, \alpha_1) = (0.10, 0.95)$. These sets of parameters determine the ARCH(1) processes with various properties, e.g. with different unconditional variances. All computations were conducted in the statistical software `R` by introducing original computational procedures. The on-line estimation algorithm (5.82) was employed under these conditions: (i) $\widehat{\boldsymbol{P}}_0 = 10^5 \boldsymbol{I}$, (ii) $\widehat{\boldsymbol{\theta}}_0$ was given by the burn-in with $n = 60$ and $\varepsilon = 0.1$ (see Section 5.4.1), (iii) $\widehat{\boldsymbol{\varphi}}_1 = (1, 0)^\top$, (iv) the method was factorized and regularized as in Section 5.2.3, (v) the forgetting factors were defined as $\lambda_t = 0.99\lambda_{t-1} + 0.01$, $\lambda_0 = 0.95$ and $t \in \mathbb{N}$, and finally (vi) the projection was defined by (5.36) with $\boldsymbol{D}_{\mathcal{M}} = \{\boldsymbol{\theta} \in \mathbb{R}^2 | 10^{-9} \leq \omega \leq 10^3, 0 \leq \alpha_1 \leq 1 - 10^{-9}\}$, which assures positivity and stationarity of the ARCH conditional variance.

Figures 6.4, 6.5, and 6.6 summarize the results of the Monte Carlo analysis. The estimation was stopped at times $T_a = 1000$, $T_b = 5000$, $T_c = 10000$, and $T_d = 20000$. The corresponding estimates have been studied by means of the standard boxplots. It is evident that they converge to their true counterparts with decreasing variances as T grows. Apparently, the speed of convergence differs amongst these cases. However, this is in accordance with theoretical findings (Ljung, 1999, Chapter 11). From the displayed figures, one may conclude that the proposed recursive estimator (5.82) can calibrate the ARCH processes adaptively.

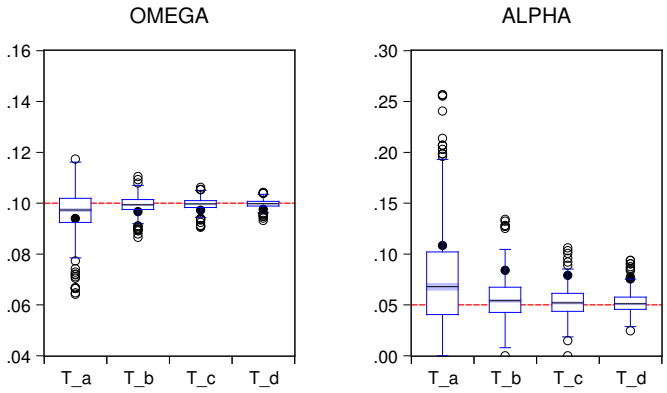


Figure 6.4: Boxplots of the SW-RPEM estimates of the ARCH(1) process with $(\omega, \alpha_1) = (0.10, 0.05)$. The recursive estimation process was stopped at the times $T_a = 1000$, $T_b = 5000$, $T_c = 10000$, and $T_d = 20000$.

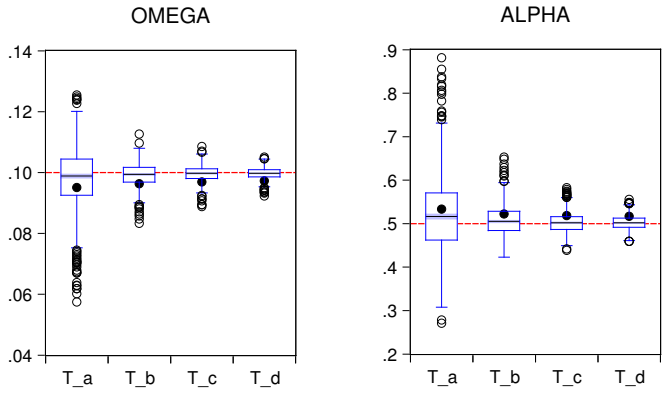


Figure 6.5: Boxplots of the SW-RPEM estimates of the ARCH(1) process with $(\omega, \alpha_1) = (0.10, 0.50)$. The recursive estimation process was stopped at the times $T_a = 1000$, $T_b = 5000$, $T_c = 10000$, and $T_d = 20000$.

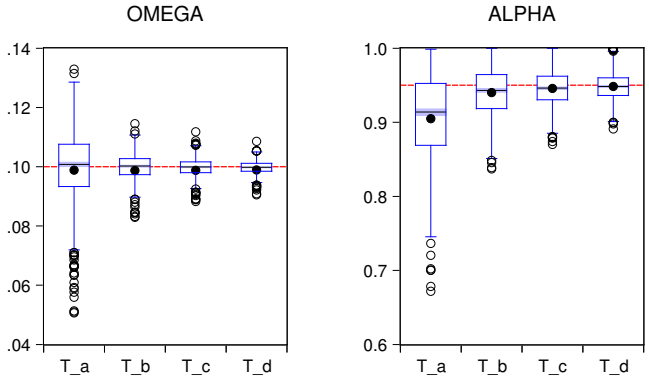


Figure 6.6: Boxplots of the SW-RPEM estimates of the ARCH(1) process with $(\omega, \alpha_1) = (0.10, 0.95)$. The recursive estimation process was stopped at the times $T_a = 1000$, $T_b = 5000$, $T_c = 10000$, and $T_d = 20000$.

6.3 On-line estimation of GJR-GARCH models: Monte Carlo study

Similarly as in Sections 6.1 and 6.2, we shall examine the numerical behaviour of the on-line estimates designed for the GJR-GARCH parameters. See also Section 3.3. Namely, we shall compare two versions of the recurrent estimators, i.e. the self-weighted recursive prediction error method (SW-RPEM) and the self-weighted recursive pseudo-linear regression algorithm (SW-RPLR) given by (5.86) and (5.87), using simulations. Monte Carlo experiments were realized and verified under various experimental settings. The suggested techniques have behaved in line with the expectations. Therefore, it may be interesting to study differences between the results introduced in Section 6.1 and those delivered by applying the estimation procedures for the GJR-GARCH models under the completely same arrangements.

Particularly, three GJR-GARCH(1,1) processes (3.11) with Gaussian innovations of the length 10000 were replicated 1000 times with the following true parameters:

(i) $(\omega, \alpha_1, \beta_1, \gamma_1) = (0.30, 0.05, 0.80, 0)$,

(ii) $(\omega, \alpha_1, \beta_1, \gamma_1) = (0.50, 0.20, 0.50, 0)$,

(iii) $(\omega, \alpha_1, \beta_1, \gamma_1) = (0.05, 0.05, 0.94, 0)$.

Obviously, these parameters undoubtedly generate the GARCH(1,1) processes, which have been studied in Section 6.1. Thus, we shall investigate whether the on-line methods available for estimating the GJR-GARCH models deliver results comparable with those introduced in Section 6.1 (i.e. how the additional parameter does influence their numerical performance). The estimation schemes (5.86) and (5.87) were implemented in the statistical software `R` under these conditions: (i) $\widehat{\boldsymbol{P}}_0 = 10^5\boldsymbol{I}$, (ii) $\widehat{\boldsymbol{\theta}}_0$ was given by the burn-in as $\big(\frac{1}{60}\sum_{i=1}^{60} y_i^2(1 - 2 \cdot 0.1, 0.1, 0.1, 0)\big)^\top$, (iii) $\widehat{\boldsymbol{\varphi}}_1 = \widehat{\boldsymbol{\psi}}_1 = (1, 0, 0, 0)^\top$, (iv) the methods were factorized and regularized as in Section 5.2.3, (v) the forgetting factors were defined as $\lambda_t = 0.99\lambda_{t-1} + 0.01$, $\lambda_0 = 0.95$ and $t \in \mathbb{N}$, and finally (vi) the projection was given by (5.36) with $\boldsymbol{D}_{\mathcal{M}} = \{\boldsymbol{\theta} \in \mathbb{R}^4 | 10^{-9} \leq \omega \leq 10^3, \alpha_1 \geq 0, \beta_1 \geq 0, \alpha_1 + \gamma_1 \geq 0, \alpha_1 + \beta_1 + 0.5\gamma_1 \leq 1 - 10^{-9}\}$, which guarantees positivity and stationarity of the GJR-GARCH conditional variance.

Table 6.4 displays the results of the simulation study. Similarly as in Section 6.1, the medians of absolute errors of the final estimates are presented therein. One can see that the numbers in this table and Tables 6.1, 6.2, and

6.3 are comparable. Therefore, it may be summed up that the algorithms estimating the GJR-GARCH models are reliable (from this viewpoint). According to expectations, the procedure (5.86) outperforms the technique given by (5.87).

Additionally, Figure 6.7, 6.8, and 6.9 illustrate the convergence of the self-weighted recursive prediction error method algorithm. In particular, one thousand GJR-GARCH(1,1) processes of the length 20000 were simulated, and the estimation was stopped at the times $T_a = 1000$, $T_b = 5000$, $T_c = 10000$, and $T_d = 20000$. Other experimental conditions remained unchanged. The algorithm has repeatedly demonstrated desirable properties.

True values	*Method*	ω	α_1	β_1	γ_1
(i)	SW-RPLR	0.36647	0.01079	0.19137	0.01361
	SW-RPEM	**0.10403**	**0.00887**	**0.05886**	**0.00849**
(ii)	SW-RPLR	0.04898	0.01453	0.03609	0.01814
	SW-RPEM	**0.03775**	**0.01323**	**0.02896**	**0.01481**
(iii)	SW-RPLR	0.10636	0.01950	0.04262	0.01410
	SW-RPEM	**0.01033**	**0.00511**	**0.00622**	**0.00500**

Table 6.4: Simulation results for the GJR-GARCH(1,1) process of the length 10000 with the various true parameters. Medians of absolute errors of the final estimates.

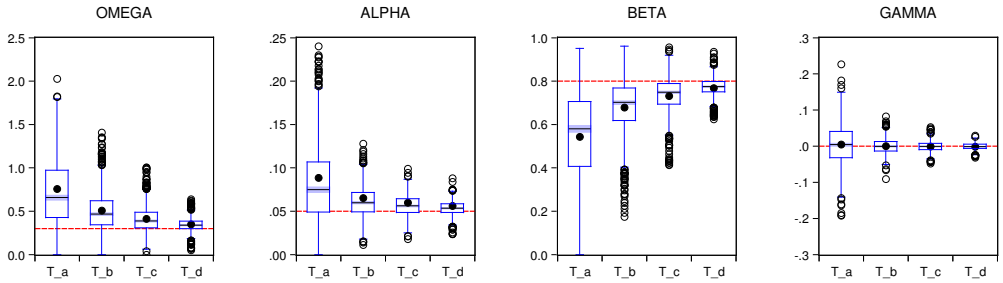


Figure 6.7: Boxplots of the SW-RPEM estimates of the GJR-GARCH(1,1) process with $(\omega, \alpha_1, \beta_1, \gamma_1) = (0.30, 0.05, 0.80, 0)$. The recursive estimation process was stopped at $T_a = 1000$, $T_b = 5000$, $T_c = 10000$, and $T_d = 20000$.

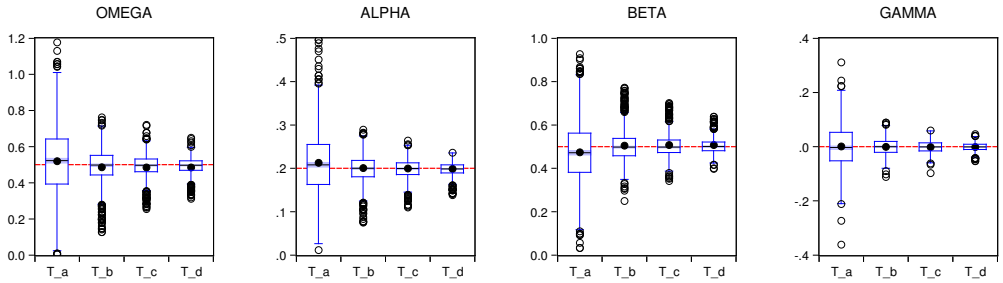


Figure 6.8: Boxplots of the SW-RPEM estimates of the GJR-GARCH(1,1) process with $(\omega, \alpha_1, \beta_1, \gamma_1) = (0.50, 0.20, 0.50, 0)$. The recursive estimation process was stopped at $T_a = 1000$, $T_b = 5000$, $T_c = 10000$, and $T_d = 20000$.

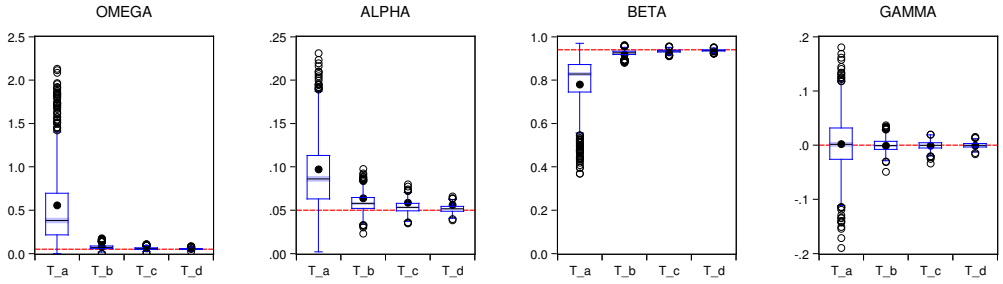


Figure 6.9: Boxplots of the SW-RPEM estimates of the GJR-GARCH(1,1) process with $(\omega, \alpha_1, \beta_1, \gamma_1) = (0.05, 0.05, 0.94, 0)$. The recursive estimation process was stopped at $T_a = 1000$, $T_b = 5000$, $T_c = 10000$, and $T_d = 20000$.

6.4 On-line estimation of EWMA models: Monte Carlo study

This section examines the numerical effectiveness of the proposed recursive estimation method (5.83) designed for calibrating the RiskMetrics EWMA model from Section 3.2. Firstly, simulations study the convergence properties of the suggested technique in the same manner as in Sections 6.1, 6.2, and 6.3. Secondly, we shall test the adaptability of the algorithm to a time-varying EWMA modelling parameter. Thirdly, we shall analyse the performance of the robustification proposed in Section 5.2.6.

6.4.1 Convergence behaviour

The following paragraphs summarize investigating of the suggested estimation procedure given by (5.83) by means of Monte Carlo experiments. Various numerical experiments were analysed with almost analogous results. Therefore, three representative cases are reviewed here. Particularly, we replicated three EWMA processes (3.10) with Gaussian disturbances of the length 20000 with three distinct parameters λ (namely, 0.90, 0.94, and 0.99) in order to study the convergence properties of the tested technique. One thousand repetitions were generated. The chosen length corresponds to approximately six-hour dataset working with one-second data. All computations were conducted in the statistical software `R` by introducing original methods.

Figure 6.10 exemplifies the effectiveness of the one-stage self-weighted recursive prediction error procedure (SW-RPEM) defined by (5.83) specified by the consequent recommendations. Namely, we put: (i) $\widehat{p}_0 = 10^5$, (ii) $\widehat{\lambda}_0 = 0.94$, (iii) $\widehat{\sigma}_1^2$ was the sample variance of several first observations and $\widehat{\sigma}_1^{2\prime}$ equalled zero, (iv) the forgetting factors were given as $\lambda_t = 0.99\lambda_{t-1} + 0.01$, $\lambda_0 = 0.95$ and $t \in \mathbb{N}$ (i.e. it gradually grows to 1). The projection (5.36) used $\boldsymbol{D}_{\mathcal{M}} = \{\lambda \in \mathbb{R} | \lambda \in [10^{-9}, 1 - 10^{-9}]\}$. The estimation process was stopped at the times $T_a = 1000$, $T_b = 3000$, $T_c = 5000$, and $T_d = 10000$; the current estimates were always stored. Figure 6.10 plots sample characteristics of these estimates using the standard boxplots for each stopping time. It is apparent that the estimates converge to the true values jointly with decreasing variances. Thus, one might conclude that the tested self-weighted recursive method (5.83) is capable of estimating the EWMA parameter in accordance with Ljung (1999, Chapter 11).

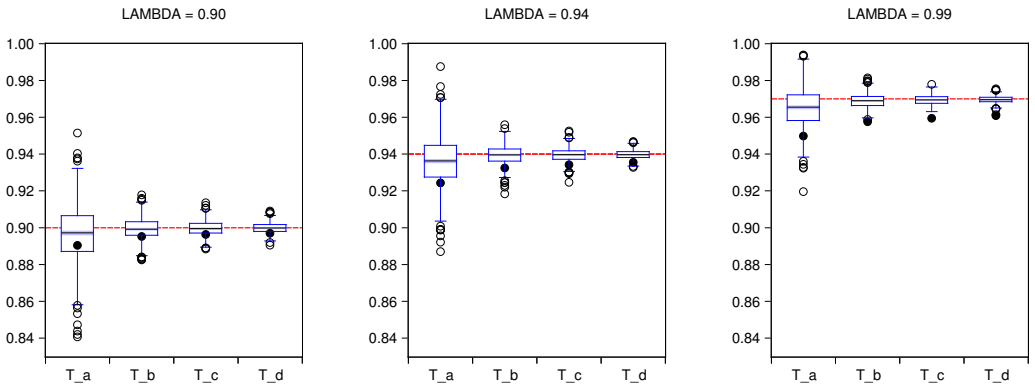


Figure 6.10: Boxplots of the SW-RPEM estimates of the EWMA parameter ($\lambda = 0.90$, 0.94, and 0.99, respectively). The recursive estimation process was stopped at $T_a = 1000$, $T_b = 5000$, $T_c = 10000$, and $T_d = 20000$.

6.4.2 Tracking time-varying parameters

Furthermore, one may examine how the different procedures can handle estimating the RiskMetrics EWMA model with the parameter λ that varies over time. The performance of the on-line and off-line methods may be compared. Namely, we replicated two EWMA processes (3.10) with Gaussian innovations of the length 10000 with the time-varying parameters. One thousand repetitions were generated. More specifically, we tested these two parameter settings: (i) $\lambda = 0.94$ for $t = 1, \ldots, 5000$ and $\lambda = 0.99$ otherwise, (ii) $\lambda = 0.94$ for $t = 1, \ldots, 5000$, $\lambda = 0.99$ for $t = 5001, \ldots, 7500$, and $\lambda = 0.97$ otherwise. The following estimation procedures were conducted: (i) the recursive estimation scheme (5.83) with the forgetting factor growing to 1 (the same as above), (ii) the recursive estimation scheme (5.83) with the constant forgetting factor $\lambda_t \equiv 0.995$ for all t, (iii) the recursive estimation scheme (5.83) with the constant forgetting factor $\lambda_t \equiv 0.997$ for all t, (iv) the recursive estimation scheme (5.83) with the constant forgetting factor $\lambda_t \equiv 0.999$ for all t, and finally (v) the off-line conditional maximum likelihood method suitable for the normally distributed EWMA model. Other experimental conditions remained as before.

Figures 6.11 and 6.12 survey these experiments and the corresponding estimates in detail. To be more precise, medians of available estimates were computed at each time t, and they are presented in both figures. Clearly, the off-line conditional maximum likelihood estimates demonstrate the least accuracy from this point of view. The self-weighted recursive algorithm (5.83) with the forgetting factor growing to one slowly accepts the underlying change in the parameter. Finally, the on-line calibration procedures with the constant forgetting factors are able to track the change more conveniently. The closer the forgetting factor is to

one, the more conservative estimates are, i.e. less volatile but also less precise. See Remark 5.2.1. To conclude, the recursive estimation scheme (5.83) with the particular forgetting factors has outperformed the off-line estimator since this is not adapted to reflect any time change in the RiskMetrics EWMA parameter. In this analysis, the on-line algorithms have always behaved more favourably.

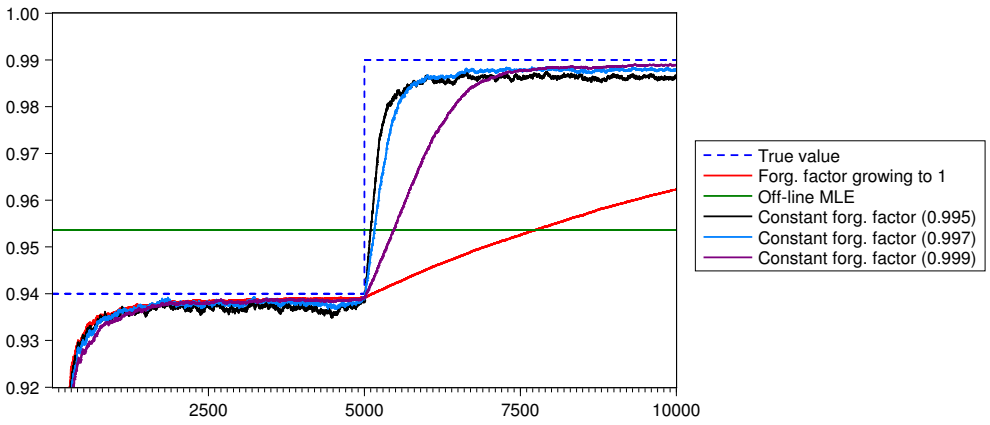


Figure 6.11: Medians of various estimates of the time-varying EWMA parameter.

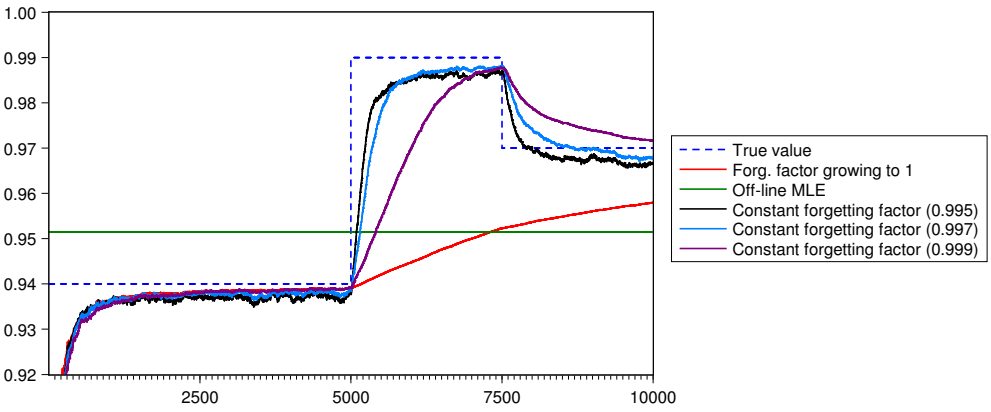


Figure 6.12: Medians of various estimates of the time-varying EWMA parameter.

6.4.3 Robustification

Section 5.2.6 has suggested a clear truncation concept, which should solve eventual problems with additive outliers in data. This section studies the approach using two demonstrative numerical examples, which can justify its adequacy.

Before analysing the results, we shall introduce the experimental framework. We replicated one thousand EWMA processes (3.10) of the length 10000 with normally distributed innovations and $\lambda = 0.94$. After each simulation, we produced

two different time series incorporating additive outliers; we inserted $y_{2500} = 10$ in the first case, and $y_{7500} = 10$ in the second instance. Other experimental conditions remained as before. In both instances, we have studied two versions of the estimation scheme (5.83), i.e. the first one with the forgetting factor increasing to one (defined as before) and the second one with the constant forgetting factor 0.997. Additionally, we have distinguished between three estimation variants of both approaches: without any robustification and with the robustification as was outlined in Section 5.2.6 using $a_t \equiv u_{0.99}$ and $a_t \equiv u_{0.9999}$. Recall that u_α denotes the corresponding α-quantile of the standard normal distribution.

Figures 6.13 and 6.14 display medians of the distinct on-line estimates (calculated at each time step). Apparently, the additive outliers have had disturbingly destructive influence on the estimation from this viewpoint when no robustification has been applied. By employing the recommended robustification, the results look more favourably. However, one should use a less conservative a_t (e.g. as $u_{0.9999}$) in (5.37) not to restrict on-line estimation excessively, especially during the transient phase.

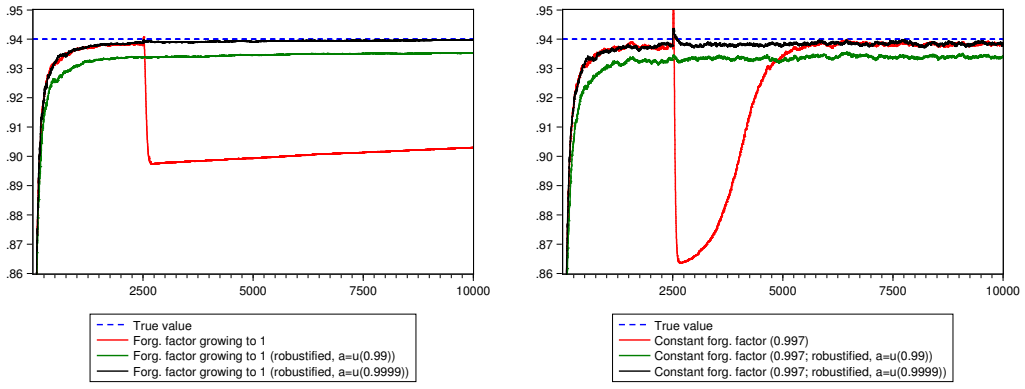


Figure 6.13: Medians of the EWMA parameter estimates (the outlier at $t = 2500$).

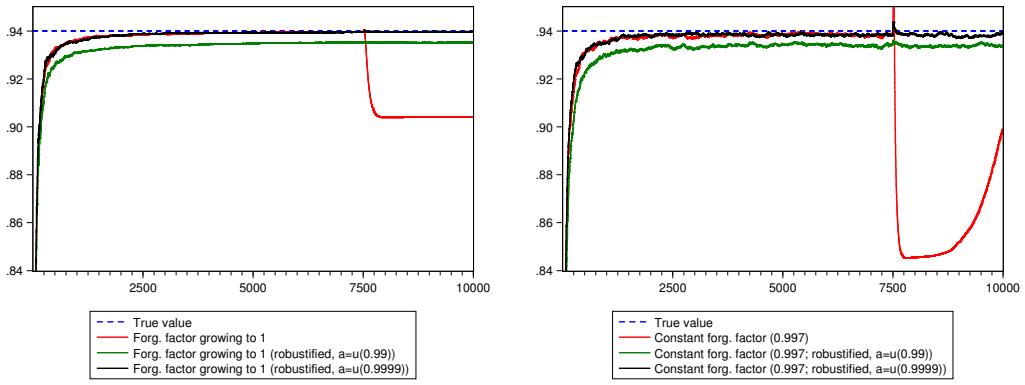


Figure 6.14: Medians of the EWMA parameter estimates (the outlier at $t = 7500$).

6.5 Empirical analysis of the PX index

In this section, the central stock index of Prague Stock Exchange (the PX index) is studied by the RiskMetrics EWMA modelling framework using all instruments investigated in Section 6.4. In particular, we consider the pragmatic argument that it may be advantageous to analyse daily financial returns by conditional heteroscedasticity models with parameters varying over time (Trešl, 2011). This modelling framework is usually implemented ad hoc. For instance, one may utilize the rolling window estimates, where usual (off-line) calibration methods are applied within a data window of some (limited) width; this window is moved systematically over time, and thus generates time-varying paths of estimates. However, tracking time-changing parameters can be methodically provided by recursive estimation techniques by employing constant forgetting factors as was discussed earlier (see Remark 5.2.1). Therefore, we have decided to examine the suggested estimation scheme (5.83) under distinct conditions in this context.

The PX index (ISIN XC0009698371) is an official market-cap weighted stock index composed of the most liquid shares traded on the Prague Stock Exchange. In particular, it is a price index of blue chips issues, which is calculated in real-time and weighted by market capitalization. Dividends are not considered. A new value of the PX index is delivered by a particular formula; it reflects each single price change of index constituents. The maximum weight for a share issue is 20% on a decisive day. A portfolio of core issues is variable, and it can be restructured quarterly (*Wiener Borse*, 2015).

The PX index was launched on 5th April 1994 (originally known as PX-50). Its base was composed of the fifty most significant share issues operating on the Prague Stock Exchange. The opening base value was fixed on 1000. The number of basic issues has been variable since December 2001. In March 2006, the PX index was officially introduced. It took over the whole history of the replaced index PX-50 continuing in its development. In March 2015, the PX base contained fourteen issues. The top five stocks had approximately 85% share of market capitalization in the portfolio. The majority of capitalization was allocated in banking, energy, and insurance sectors. Further details (including historical data) can be found on the website of *Prague Stock Exchange* (2015).

Figure 6.15 graphs all historical daily closing quotes and associated logarithmic returns of the PX index until 31st March 2015 (i.e. 5248 observations). The minimal value 316 occurred on 8th October 1998 after the Russian financial crisis. The maximal observation 1936 was achieved on 29th October 2007. It is visible that the crisis year 2008 was truly exceptional (Hendrych, 2014a).

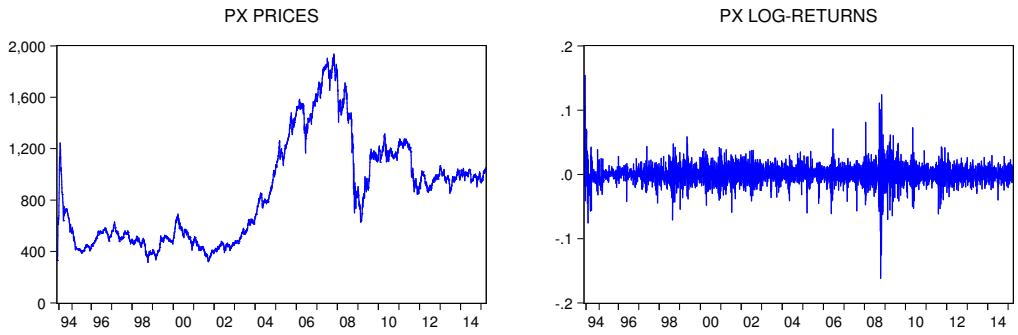


Figure 6.15: The PX index historical closing quotes and associated log-returns.

Figure 6.16 shows different off-line and on-line estimates of the EWMA modelling parameter for the PX index log-returns. Namely, the given financial returns are studied by means of the following methods: (i) the recursive algorithm (5.83) with the forgetting factor growing to one (defined as in Section 6.4) with the robustification (5.37) using $a_t \equiv u_{0.9999}$ and without it, (ii) the recursive algorithm (5.83) with the constant forgetting factor 0.997 with the robustification (5.37) using $a_t \equiv u_{0.9999}$ and without it, (iii) the off-line conditional maximum likelihood method. The on-line estimation was initialized as in Section 6.4.

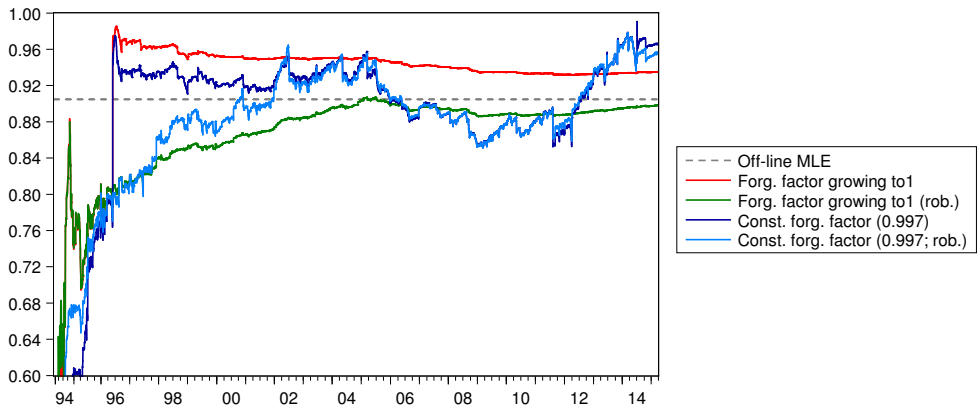


Figure 6.16: Different estimates of the EWMA parameter (the PX log-returns).

At first glance, the accepted robustification eliminates the influence of additive outliers (especially the one occurring in the first part of the dataset). Therefore, the robustified methods are more reliable from this perspective. Moreover, the recursive estimates calculated by using the constant forgetting factor apparently fluctuate around the off-line one. One can discover several trends, which obviously correspond to the overall development of the PX index closing quotes (see Figure 6.15). The on-line estimates evaluated by the growing forgetting factors are more rigid than the ones with the constant forgetting factors; compare with

Section 6.4.2. Note that all mentioned recursive estimates are less reliable at the beginning of the observed time series since these methods are initialized therein.

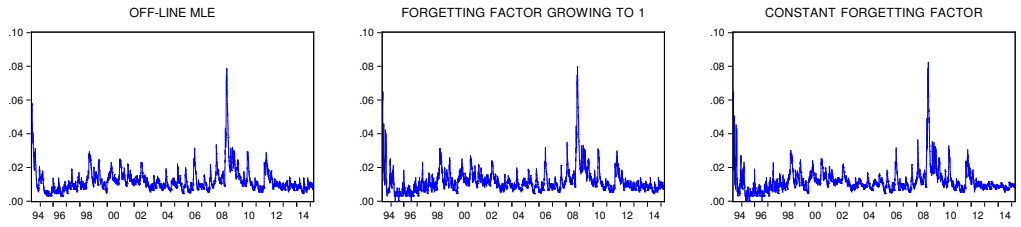


Figure 6.17: Different estimates of volatility (the PX log-returns).

Figure 6.17 presents the estimated volatilities of the logarithmic returns of the PX index. According to the preceding discussion, we have primarily examined the off-line method and the two robustified recursive procedures for calibrating the RiskMetrics EWMA parameter. At first sight, one could conclude that all considered outputs follow analogous trends. The estimates based on the off-line conditional likelihood procedure and the suggested recursive algorithm seem to be closely related. To distinguish which of the studied calibration techniques do offer better predictions of volatility of financial returns, it would be necessary to employ additional criteria that study this issue from the financial management point of view (Patton, 2011). Alternatively, one may compare the achieved values of the (conditional) log-likelihood function associated with this particular estimation problem (or equivalently contrast some information criteria). The calculated log-likelihoods are recapitulated in Table 6.5. The complete and truncated samples are considered computing the introduced numbers. Namely, first 10%, 20%, 30%, 40%, and 50% observations were cut off in order to verify the adequacy since the recursive methods are less stable during the initial phase of estimation (see above). Consequently, the calibration algorithm (5.83) introduced in Section 5.4.2 is visibly competitive. It can be employed in the PX index daily data context.

Sample/Method	Off-line MLE	On-line ($\lambda_t \nearrow 1$)	On-line ($\lambda_t \equiv 0.997$)
Complete	**15819.37**	15751.06	15755.63
Truncated (10%)	14191.75	14178.27	**14191.80**
Truncated (20%)	12470.94	12460.19	**12473.38**
Truncated (30%)	11001.93	10994.66	**11007.17**
Truncated (40%)	9525.31	9521.00	**9530.27**
Truncated (50%)	7858.61	7856.11	**7862.97**

Table 6.5: Values of log-likelihood functions corresponding to estimates in Figure 6.17.

6.6 On-line volatility monitoring

Monte Carlo experiments reviewed in the preceding sections have demonstrated that the suggested one-stage self-weighted recursive prediction error methods (in the severe proposed variants introduced in Chapter 5) are generally capable of estimating the particular modelling parameters recursively. This approach has been empirically examined in the context of the PX index (daily) data, where it has successfully tracked the time-varying EWMA modelling parameter. However, the primary goal of the proposed recursive estimation algorithms is to treat high-frequency data that cannot be managed in real time by mutual off-line methods just due to their frequencies. Therefore, we shall demonstrate the possible utility of this methodological instrument examining this context.

In practice, it might be desirable to monitor or forecast volatility of financial returns in real time since volatility represents a crucial measure of risk with a broad portfolio of empirical applications. For instance, it can determine a solution of asset allocation or hedging problems; it can help to construct an optimal portfolio or to evaluate risk measures (e.g. the Value at Risk).

For illustrative purposes, let us assume the high-frequency tick data concerning the exchange rate EUR/USD on 17th July 2014 (i.e. 27896 observations). Refer also to Figure 6.18. Such (historical) datasets are usually provided by standard trading software. In particular, we have analysed the log-returns in the form $y_t = 100 \log(P_t/P_{t-1})$, where P_t denotes the exchange rate at time t, by applying the GJR-GARCH(1,1) with the normally distributed innovations. This simplification is strongly motivated by practice since this model may be regarded as the benchmark (similarly to the GARCH(1,1) model, which does not reflect an eventual leverage effect). Moreover, there exist pieces of evidence that the given dataset should be modelled by using the conditional heteroscedasticity framework (e.g. the high kurtosis or the results of the Ljung-Box tests indicating autocorrelated squared financial returns). Particularly, we applied the recursive scheme (5.86) on the data (in accordance with the accepted modelling scheme). We used the robustification recommended in Section 5.2.6, namely defined by the truncation (5.37) with $a_t \equiv u_{0.9999}$. The method was initialized as in Section 6.3. One may contrast two distinct versions, i.e. the first one with the constant forgetting factor ($\lambda_t \equiv 0.997$ for all t) and the second one with the forgetting factor growing to 1 ($\lambda_t = 0.99\lambda_{t-1} + 0.01$, $\lambda_0 = 0.95$ and $t \in \mathbb{N}$). See Remark 5.2.1 for explaining the motivation for doing this.

Finally, Figure 6.19 compares on-line estimated volatilities computed recursively. As before, one might conclude that there exist severe analogous trends in both cases. Nonetheless, the volatility estimated by the constant forgetting

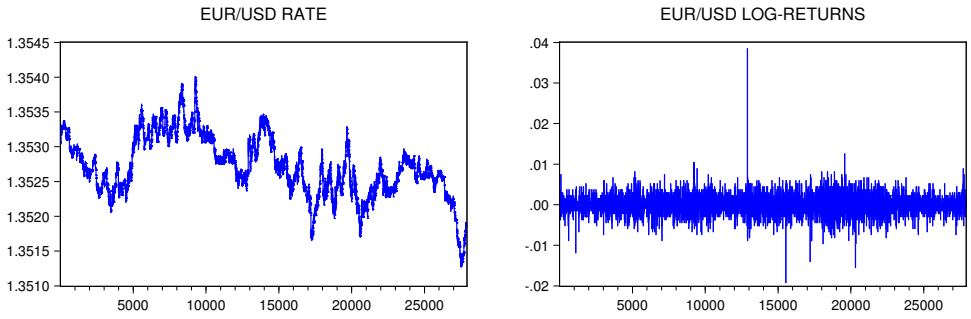


Figure 6.18: EUR/USD exchange rates and associated log-returns.

factor seems to change more progressively. Since the proposed method is computationally efficient in terms of its complexity and required memory, it is possible to track both variants (with the distinct forgetting factors) simultaneously and compare their performance on-line by applying some suitable criteria (e.g. the corresponding analogies of information criteria introduced in Section 5.2.5).

The calibrated volatilities may be employed in specific financial problems, which go beyond the scope of this thesis. The primary goal of this subsection has been to demonstrate that the proposed recursive estimation procedures can monitor time-varying volatility, to check its positivity and stationarity sequentially (by using a simple projection as in (5.36)). It is in sharp contrast to the off-line estimation methods.

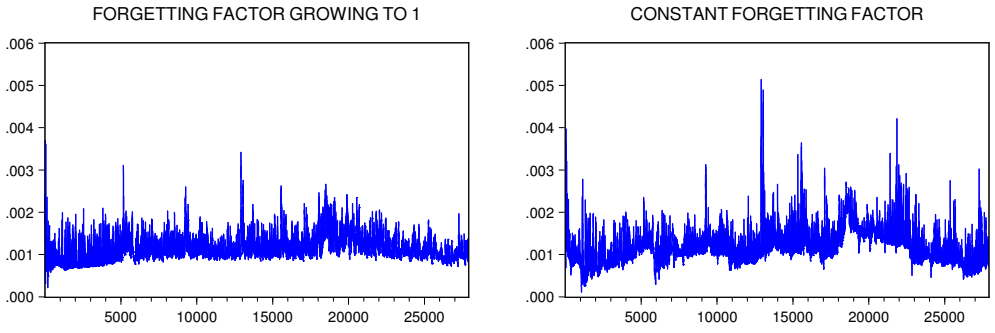


Figure 6.19: Different estimates of volatility (the EUR/USD log-returns).

Part II

Conditional covariance and correlation modelling

7. Conditional covariance modelling: An overview

The conditional heteroscedasticity models considered in Chapter 3 (and many others cited and elaborated in the literature) calibrate volatility of financial returns independently of other financial time series. However, this univariate approach may be inefficient in many situations of common econometric and financial practice. To be more precise, it cannot reflect the spillover of volatility amongst various financial assets, and it cannot describe their mutual correlations.

Conditional correlations are crucial inputs for many problems of financial and risk management. Hedges rely on estimates of volatilities and conditional correlations between the returns of the (risky) assets in the hedging portfolio. If they are supposed to vary over time, the hedge ratio should be adjusted to incorporate the most recent information. Similarly, structured financial products based on a portfolio of underlying assets are usually priced with respect to conditional correlations amongst the underlying returns. Moreover, asset allocation or risk assessment also require knowledge of such (inter)connections. For instance, construction of an optimal portfolio with a set of constraints requests the conditional covariance matrix of all assets involved in this portfolio. The correlation structure of assets is a fundamental feature of this investment decision since it is instrumental in determining the risk (Engle, 2009, Chapter 1). One might conclude that investigating conditional covariances and correlations is indeed worth of interest from the perspective of banks, financial analysts, investors, and other participants in financial markets.

Analysis of these significant quantities is undoubtedly an important part of multivariate financial time series modelling also from the theoretical point of view. One should primarily highlight at least two reasons amongst others. Firstly, financial time series (univariate or multivariate) require applying specific model concepts that respect their characteristic features. Consult the stylized facts about financial returns and volatility discussed in Chapter 1. Secondly, several constraints are naturally linked to estimation of the considered quantities of interest. Namely, each conditional covariance matrix must be symmetric and positive (semi)definite. Additionally, each conditional correlation matrix must have unit diagonal elements, and other components must lie within the interval $[-1, 1]$. These conditions might indisputably incorporate tough limitations into calibration, especially in the case of higher dimensions. Therefore, these restrictions must be seriously taken into account by the modelling framework. Consequently,

various model representations have been implemented to simplify, transform, or (completely/partly) eliminate these constraints. See Section 7.2.

Remind that conditional correlation modelling is inherently related to conditional covariance modelling. One can essentially distinguish between the *explicit* and *implicit* modelling schemes. The first category involves representations that explicitly evaluate conditional correlations. The second one estimates conditional correlations implicitly as normalizations of conditional covariances (see below).

The relevance of the discussed issue is evident from the substantial body of literature in this research field. Many academically or practically oriented publications treat this topic from different viewpoints. See the works by Aielli (2013), Alp and Demetrescu (2010), Bauwens, Laurent and Rombouts (2006), Bollerslev (1990), Engle (2002), Engle (2009), Engle and Colacito (2006), Hafner and Reznikova (2012), Rossi and Spazzini (2010), and the references given therein.

To illustrate an essential importance of (conditional) covariances and correlations in common financial practice, we shall introduce demonstrative examples.

Example 7.0.1 Let $\boldsymbol{r}_t = (r_{1,t}, \ldots, r_{N,t})^\top$ be the vector of N risky financial returns, r_f be the risk-free rate, and $\boldsymbol{w} = (w_1, \ldots, w_N)^\top$ be the portfolio weights. The portfolio return at time t is defined by these terms as follows:

$$r_{port,t} := \sum_{i=1}^{N} w_i r_{i,t} + \left(1 - \sum_{i=1}^{N} w_i\right) r_f = \boldsymbol{w}^\top (\boldsymbol{r}_t - r_f \mathbf{1}) + r_f, \tag{7.1}$$

where $\mathbf{1}$ denotes the vector of units of the corresponding length.

Assuming $\mathbb{E}(\boldsymbol{r}_t - r_f\mathbf{1}) = \boldsymbol{\mu}$ and $\mathsf{cov}(\boldsymbol{r}_t) = \boldsymbol{\Sigma} = (\sigma_{ij})_{i,j=1}^N$ ($\boldsymbol{\Sigma}$ is regular), the portfolio mean and variance are concisely expressed as:

$$\mathbb{E}(r_{port,t}) = \boldsymbol{w}^\top \boldsymbol{\mu} + r_f \quad \text{and} \quad \mathsf{var}(r_{port,t}) = \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}. \tag{7.2}$$

In the classic Markovitz portfolio theory, the risk of a portfolio is given as the expected variance of the portfolio over a given time period. This is compared with the expected return on the portfolio over the same period (Engle, 2009, Section 2.4). Supposing that the portfolio can be either long or short in all assets, the optimal portfolio problem can be solved evaluating:

$$\min_{\boldsymbol{w} \in \mathbb{R}^N} \mathsf{var}(r_{port,t}), \quad \text{subject to} \quad \mathbb{E}(r_{port,t}) \geq \mu_0, \tag{7.3}$$

where a priori known $\mu_0 \in \mathbb{R}$ denotes the required return of the portfolio.

This minimization task has the analytical solution (provided that it exists):

$$\boldsymbol{w}_{opt} = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}(\mu_0 - r_f). \tag{7.4}$$

It is apparent that the individual variances and mutual correlations between the risky assets influence the structure of the optimal portfolio. If all risky assets have the identical expected return (i.e. $\boldsymbol{\mu} = \tilde{\mu}\mathbf{1}$ for some $\tilde{\mu} \in \mathbb{R}$) and there is no riskless rate (i.e. $\boldsymbol{w}^\top \mathbf{1} = 1$, $r_f \equiv 0$, and $\tilde{\mu} = \mu_0$), then the portfolio weights will be chosen to minimize variance. Namely, the weights of this *minimum-variance portfolio* are determined by:

$$\boldsymbol{w}_{opt}^{MV} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}}. \tag{7.5}$$

Engle and Colacito (2006) employed the conditional moments instead of their unconditional counterparts in this particular framework (conditioning by the information accumulated up to and including time $t-1$ denoted as $\mathcal{F}_{t-1}$, see Section 7.1). Accepting this concept, one may simply deduce the minimum-variance bivariate portfolio of two risky assets with an equal expected return (if there is no risk-free rate):

$$r_{port,t} = w_t r_{1,t} + (1 - w_t) r_{2,t}, \tag{7.6a}$$

$$w_t = \frac{h_{22,t} - h_{12,t}}{h_{11,t} + h_{22,t} - 2h_{12,t}}, \tag{7.6b}$$

$$\boldsymbol{H}_t = (h_{ij,t})_{i,j=1}^2 = \mathsf{cov}(\boldsymbol{r}_t | \mathcal{F}_{t-1}). \tag{7.6c}$$

It means that the optimal composition of the portfolio is varying over time based on the elements of the conditional covariance matrix $\boldsymbol{H}_t$. The advantage of this approach is that it uses the time-varying portfolio weights. It might lead to portfolios with smaller variances compared with the portfolio driven by constant weights (Engle, 2009, Chapter 9).

Example 7.0.2 The problem of hedging is closely related to the context outlined in Example 7.0.1. Namely, it can be incorporated into this framework. Consider that the first risky asset in the portfolio (without loss of generality) must be held and remaining risky assets are held primarily to reduce the risk. In particular, it means that an investor wants to create the portfolio with the minimal risk (which is measured by the portfolio variance). It leads to the following optimization problem (Engle, 2009, Section 2.4):

$$\min_{\boldsymbol{w} \in \mathbb{R}^N} \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}, \quad \text{subject to} \quad w_1 = 1,\ \mathbf{1}^\top \boldsymbol{w} = 1. \tag{7.7}$$

For instance, the optimal hedged bivariate portfolio is given as:

$$r_{port,t} = r_{1,t} - w_2 r_{2,t}, \quad w_2 = \frac{\sigma_{12}}{\sigma_{22}} = \rho_{12}\frac{\sqrt{\sigma_{11}}}{\sqrt{\sigma_{22}}}, \tag{7.8}$$

where ρ_{12} denotes the bivariate correlation between $r_{1,t}$ and $r_{2,t}$.

Similarly as in Example 7.0.1, this approach can be generalized assuming the conditional mean and covariances. Accordingly, one can reformulate the previous bivariate portfolio described by (7.8) as:

$$r_{port,t} = r_{1,t} - w_t r_{2,t}, \quad w_t = \frac{h_{12,t}}{h_{22,t}} = \rho_{12,t}\frac{\sqrt{h_{11,t}}}{\sqrt{h_{22,t}}}, \tag{7.9}$$

where the notation remains as in Example 7.0.1 and $\rho_{12,t}$ is the bivariate conditional correlation between the returns $r_{1,t}$ and $r_{2,t}$.

Both examples discussed above should have demonstrated that the (either unconditional or conditional) covariance and correlation matrices are indeed significant inputs for essential financial tasks. This fact has motivated the research presented in Part II of the thesis. Notice that more sophisticated variants of the problems solved above might be evidently considered.

Remark 7.0.1 Engle (2009, Section 1.4) briefly introduced an interesting economic model of correlations with the aim to show how correlations in financial returns are interrelated with correlations in the news taken as any information that changes forecasts of future payments.

Let P_t be the price per share, D_t be the dividend per share. Continuously compounded returns are commonly defined as:

$$r_{t+1} := \log(P_{t+1} + D_{t+1}) - \log(P_t). \tag{7.10}$$

Employing the log-linearization used by Engle (2009, Section 1.4), they can be approximated as:

$$r_{t+1} \approx k + \rho p_{t+1} + (1-\rho)d_{t+1} - p_t, \tag{7.11}$$

where $p_t = \log(P_t)$ and $d_t = \log(D_t)$, k is a constant of the linearization, and $\rho \in (0,1)$ is the discount rate. Note that this approximation behaves conveniently if the ratio of the two components is small and (relatively) constant.

The unexpected return at time t defined by $r_t - \mathbb{E}(r_t|\mathcal{F}_{t-1})$, where $\mathcal{F}_{t-1}$ denotes the information set available at time $t-1$, can be rewritten using (7.11) as:

$$r_t - \mathbb{E}(r_t|\mathcal{F}_{t-1}) \approx \rho[p_t - \mathbb{E}(p_t|\mathcal{F}_{t-1})] + (1-\rho)[d_t - \mathbb{E}(d_t|\mathcal{F}_{t-1})]. \tag{7.12}$$

It means that the unexpected return consists of two components: the unexpected movement of price and surprise in the dividend. Moreover, Equation (7.11) can be solved for p_t (under the corresponding conditions), i.e. p_t may be written as a function of future returns and dividends; this solution can be substituted into (7.12). The unexpected return can be thus schematically summarized by the following relation:

$$r_t - \mathbb{E}(r_t|\mathcal{F}_{t-1}) = \eta_t^d - \eta_t^r, \tag{7.13}$$

where the innovations η_t^d and η_t^r comprise the news about the (future) dividends and returns. Therefore, the conditional variance of an asset return r_t is simply driven by (Engle, 2009, Section 1.4):

$$\mathsf{var}(r_t|\mathcal{F}_{t-1}) = \mathsf{var}(\eta_t^d|\mathcal{F}_{t-1}) + \mathsf{var}(\eta_t^r|\mathcal{F}_{t-1}) - 2\mathsf{cov}(\eta_t^d, \eta_t^r|\mathcal{F}_{t-1}). \tag{7.14}$$

In the case of multivariate returns, the situation is mostly analogous. Thus, the correlations in returns result either from correlations amongst dividend news or correlations amongst returns or mutual correlations between both of them.

7.1 Model framework

This section introduces a general multivariate model framework, which is commonly employed in the described context. It can be simply regarded as a straightforward extension of the univariate case referred in Chapter 3. Particularly, consider a multivariate stochastic vector process $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ of dimension $(n \times 1)$. Denote $\mathcal{F}_t$ the σ-algebra generated by observed time series up to and including time t, i.e. $\mathcal{F}_t = \sigma(\boldsymbol{X}_s, s \leq t)$ is the smallest σ-algebra with respect to which $\boldsymbol{X}_s$ is measurable for all $s \leq t$, $s, t \in \mathbb{Z}$.

In this framework, assume the following model:

$$\boldsymbol{X}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{Z}_t, \tag{7.15}$$

where $\boldsymbol{H}_t = (h_{ij,t})_{i,j=1}^n$ is the $(n \times n)$ symmetric positive definite conditional covariance matrix of $\boldsymbol{X}_t$ given $\mathcal{F}_{t-1}$. Furthermore, one supposes that $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ is an $(n\times 1)$ i. i. d. stochastic vector process with the following moments: $\mathbb{E}(\boldsymbol{Z}_t) = \boldsymbol{0}$ and $\mathsf{cov}(\boldsymbol{Z}_t) = \boldsymbol{I}$, where $\boldsymbol{I}$ denotes the corresponding identity matrix.

In the model (7.15), the moments of $\boldsymbol{X}_t$ can be readily evaluated as:

$$\mathbb{E}(\boldsymbol{X}_t|\mathcal{F}_{t-1}) = \boldsymbol{0}, \quad \mathsf{cov}(\boldsymbol{X}_t|\mathcal{F}_{t-1}) = \boldsymbol{H}_t^{1/2}(\boldsymbol{H}_t^{1/2})^\top = \boldsymbol{H}_t, \tag{7.16a}$$

$$\mathbb{E}(\boldsymbol{X}_t) = \boldsymbol{0}, \quad \mathsf{cov}(\boldsymbol{X}_t) = \mathbb{E}(\boldsymbol{H}_t), \quad \mathsf{cov}(\boldsymbol{X}_t, \boldsymbol{X}_{t+h}) = \boldsymbol{0},\ h \in \mathbb{Z}\backslash\{0\}. \tag{7.16b}$$

Here, $\boldsymbol{H}_t^{1/2}$ is any $(n \times n)$ positive definite matrix such that the last equality in (7.16a) holds. The considered decomposition of $\boldsymbol{H}_t$ may be delivered, for instance, by the Cholesky factorization as it is common (Bauwens et al., 2006).

Moreover, the conditional correlation matrix of $\boldsymbol{X}_t$ given $\mathcal{F}_{t-1}$ denoted as $\boldsymbol{R}_t = (\rho_{ij,t})_{i,j=1}^n$ can be evaluated by computing the normalization of the conditional covariance matrix $\boldsymbol{H}_t$, i.e. $\boldsymbol{R}_t = \mathsf{diag}(\boldsymbol{H}_t)^{-1/2}\boldsymbol{H}_t\mathsf{diag}(\boldsymbol{H}_t)^{-1/2}$. Note that the matrix operator diag creates a diagonal matrix extracting associated elements of an original matrix (and all non-diagonal components are substituted by zeroes). All correlations defined in this way must clearly lie within the interval $[-1, 1]$, and they are based on information accumulated up to the previous period.

7.2 Common models of conditional covariances

The demand for reliable estimates of conditional covariances and correlations amongst financial assets has motivated and determined development of countless different modelling approaches. This section briefly recapitulates several frequent techniques that are commonly applied in the discussed context. The literature has been recently surveyed by Bauwens et al. (2006) or Engle (2009, Chapter 3). Particularly, Engle (2009, Section 3.7) concisely introduced many other representations that have been proposed and employed in this framework. They include semi-parametric, factor, stochastic volatility methods, etc.

7.2.1 Basic approaches

The most widely used covariance (correlation) matrix estimators are the simplest ones. They apply elementary modelling schemes.

One should start with the *multivariate moving average* covariance matrix (often called the *historical* covariance matrix) given by:

$$\boldsymbol{H}_t^{MA} = \frac{1}{M}\sum_{s=1}^{M}\boldsymbol{X}_{t-s}\boldsymbol{X}_{t-s}^{\top}, \quad M \geq 2. \tag{7.17}$$

This particular modelling scheme prescribes the equal weight $1/M$ to all observations less than M periods in the past and the zero weight to older measurements. Apparently, the conditional covariances may be expressed as:

$$h_{ij,t}^{MA} = \frac{1}{M}\sum_{s=1}^{M}X_{i,t-s}X_{j,t-s}, \quad M \geq 2,\ i,j = 1,\ldots,n, \tag{7.18}$$

hence, the conditional correlations are described as:

$$\rho_{ij,t}^{MA} = \frac{\sum_{s=1}^{M} X_{i,t-s} X_{j,t-s}}{\sqrt{\sum_{s=1}^{M} X_{i,t-s}^2} \sqrt{\sum_{s=1}^{M} X_{j,t-s}^2}}, \quad M \geq 2,\ i,j = 1,\ldots,n. \tag{7.19}$$

The *multivariate exponentially weighted moving average smoother* designed by RiskMetrics is defined as (Morgan, 1996):

$$\boldsymbol{H}_t^{EWMA} = (1-\lambda)\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}^{\top} + \lambda \boldsymbol{H}_{t-1}^{EWMA}, \quad \lambda \in (0,1). \tag{7.20}$$

This scheme uses the geometrically decreasing weights based on the parameter λ since the following relations hold (for all $i,j = 1,\ldots,n$):

$$h_{ij,t}^{EWMA} = (1-\lambda)X_{i,t-1}X_{j,t-1} + \lambda h_{ij,t-1}^{EWMA} = (1-\lambda)\sum_{s=1}^{\infty} \lambda^{s-1} X_{i,t-s}X_{j,t-s}. \tag{7.21}$$

Both formulas (7.17) and (7.20) ensure the symmetric positive semidefinite covariance matrices (that are positive definite under additional, relatively weak conditions). Each representation contains only the single (tuning) parameter. However, these unknown parameters are often not estimated but are simply prescribed by experts or users. For instance, one usually puts the length of the moving average as $M = 100$ and the smoothing constant as $\lambda = 0.94$ for daily data (Engle, 2009, Section 3.1). Alternatively, common statistical inference procedures may be applied in this context. Both methods must be properly initialized at first. For example, the moving average estimator can be started after collecting at least M measurements. The exponential smoother can be initiated using the sample covariance matrix of an initial data segment. Both models are apparently very simple; therefore, they are frequently applied in practice.

7.2.2 Multivariate GARCH models

Multivariate GARCH models are natural generalizations that have been suggested and used in the considered context. Many different specifications have been proposed according to the key principles of this framework (Engle, 2009, Section 3.2). Therefore, it suffices to outline how these models are typically formulated. Their key features are analogous to the univariate case introduced and discussed in Chapter 3. See Bauwens et al. (2006) or Engle (2009, Section 3.3) for further insights and more references.

The most general expression of this modelling class is the *VEC*(P,Q) *representation* of $\boldsymbol{H}_t$ given by:

$$\mathsf{vec}(\boldsymbol{H}_t) = \mathsf{vec}(\boldsymbol{\Omega}) + \sum_{p=1}^{P} \boldsymbol{A}_p \mathsf{vec}(\boldsymbol{X}_{t-p}\boldsymbol{X}_{t-p}^{\top}) + \sum_{q=1}^{Q} \boldsymbol{B}_q \mathsf{vec}(\boldsymbol{H}_{t-q}), \tag{7.22}$$

where $\boldsymbol{\Omega}$ is the $(n \times n)$ matrix of parameters, $\boldsymbol{A}_p$ and $\boldsymbol{B}_q$ are the $(n^2 \times n^2)$ matrices of parameters (all of them must respect the symmetry of $\boldsymbol{H}_t$), $P \in \mathbb{N}$ and $Q \in \mathbb{N}_0$. The matrix operator vec converts a matrix to a vector by stacking all its columns on top of one another. If the matrix is symmetric, then there will be many duplications. For this reason, the VEC scheme is sometimes replaced by the corresponding *VECH*(P,Q) *model*, where the operator vec is substituted by its inherent counterpart vech. The matrix operator vech reorganizes a symmetric matrix into one vector by stacking columns of the lower triangular part of this matrix on top of one another. In the context of Equation (7.22), it particularly means that the $(n \times n)$ matrix $\boldsymbol{H}_t$ is modified to the $(n(n+1)/2 \times 1)$ vector $\mathsf{vech}(\boldsymbol{H}_t)$ instead of the $(n^2 \times 1)$ vector $\mathsf{vec}(\boldsymbol{H}_t)$, etc. Nevertheless, both considered representations have identical features.

The VEC model describes the dependence between each element of $\boldsymbol{H}_t$ and squares and cross products of past returns and lagged covariances. However, the conditional covariance matrix $\boldsymbol{H}_t$ is not positive definite without imposing further restrictions. Additionally, this model potentially includes an enormous number of free parameters. Therefore, it is convenient to tighten the structure of (7.22) to reduce the number of parameters and to guarantee the positive definite $\boldsymbol{H}_t$. Notice that the sufficient conditions for (weak) stationarity may be deduced from the matrices $\boldsymbol{A}_p$ and $\boldsymbol{B}_q$ (Engle, 2009, Theorem 3.1).

The extremely modest parameterizations of the VEC model (7.22) are the moving average given by (7.17) and the exponential smoother delivered by (7.20). It is evident by using: $P = M$, $Q = 0$, $\boldsymbol{\Omega} \equiv \boldsymbol{0}$, $\boldsymbol{A}_p = (1/M)\boldsymbol{I}$ for all p (in the first case); $P = Q = 1$, $\boldsymbol{\Omega} \equiv \boldsymbol{0}$, $\boldsymbol{A}_1 = (1-\lambda)\boldsymbol{I}$, $\boldsymbol{B}_1 = \lambda \boldsymbol{I}$ (in the second case). A less restrictive modelling subclass involves the *diagonal* VEC(P,Q) representation (i.e. the matrices $\boldsymbol{A}_p$ and $\boldsymbol{B}_q$ are restricted to be diagonal). In this model, Equation (7.22) can be concisely reformulated as:

$$\boldsymbol{H}_t = \boldsymbol{\Omega} + \sum_{p=1}^{P} \bar{\boldsymbol{A}}_p \odot (\boldsymbol{X}_{t-p}\boldsymbol{X}_{t-p}^{\top}) + \sum_{q=1}^{Q} \bar{\boldsymbol{B}}_q \odot \boldsymbol{H}_{t-q}, \tag{7.23}$$

where $\bar{\boldsymbol{A}}_p$ and $\bar{\boldsymbol{B}}_q$ are the $(n \times n)$ matrices of parameters that correspond to the diagonals of $\boldsymbol{A}_p$ and $\boldsymbol{B}_q$ from (7.22), respectively. Namely, these matrices are implied by the relations $\boldsymbol{A}_p = \mathsf{diagv}[\mathsf{vec}(\bar{\boldsymbol{A}}_p)]$ and $\boldsymbol{B}_q = \mathsf{diagv}[\mathsf{vec}(\bar{\boldsymbol{B}}_q)]$, where

the operator diagv transforms a vector into a diagonal matrix (i.e. the vector creates the diagonal of this matrix). The Hadamard product $\odot$ of two general $(m \times n)$ matrices is defined as the $(m \times n)$ matrix of element-by-element products. Notice that if all matrices $\bar{\boldsymbol{A}}_p$ and $\bar{\boldsymbol{B}}_q$ are positive semidefinite and if $\boldsymbol{\Omega}$ is positive definite, then $\boldsymbol{H}_t$ will be positive definite for all t (Engle, 2009, Lemma 3.4). The diagonal VEC model is usually applied under the following (more restrictive) parameterizations: (i) the *scalar-diagonal* model ($\bar{\boldsymbol{A}}_p = \bar{\alpha}_p \mathbf{1}\mathbf{1}^\top$, $\bar{\alpha}_p \geq 0$, and $\bar{\boldsymbol{B}}_q = \bar{\beta}_q \mathbf{1}\mathbf{1}^\top$, $\bar{\beta}_q \geq 0$), (ii) the *vector-diagonal* model ($\bar{\boldsymbol{A}}_p = \bar{\boldsymbol{a}}_p \bar{\boldsymbol{a}}_p^\top$ and $\bar{\boldsymbol{B}}_q = \bar{\boldsymbol{b}}_q \bar{\boldsymbol{b}}_q^\top$, where $\bar{\boldsymbol{a}}_p$ and $\bar{\boldsymbol{b}}_q$ are the $(n \times 1)$ vectors), (iii) the *matrix-diagonal* ($\bar{\boldsymbol{A}}_p = \bar{\boldsymbol{\Upsilon}}_p \bar{\boldsymbol{\Upsilon}}_p^\top$ and $\bar{\boldsymbol{B}}_q = \bar{\boldsymbol{\Psi}}_q \bar{\boldsymbol{\Psi}}_q^\top$, where $\bar{\boldsymbol{\Upsilon}}_p$ and $\bar{\boldsymbol{\Psi}}_q$ are the $(n \times n)$ matrices). In these three cases, the positive semidefiniteness of $\boldsymbol{H}_t$ is guaranteed (if $\boldsymbol{\Omega}$ is positive definite).

Another general multivariate GARCH representation that can ensure the positive definiteness may be established (Engle, 2009, Section 3.3). Namely, the *BEKK*(P, Q, R) *representation* of $\boldsymbol{H}_t$ is formulated as:

$$\boldsymbol{H}_t = \boldsymbol{\Omega} + \sum_{p=1}^{P} \sum_{r=1}^{R} \widetilde{\boldsymbol{A}}_{p,r}^\top \boldsymbol{X}_{t-p} \boldsymbol{X}_{t-p}^\top \widetilde{\boldsymbol{A}}_{p,r} + \sum_{q=1}^{Q} \sum_{r=1}^{R} \widetilde{\boldsymbol{B}}_{q,r}^\top \boldsymbol{H}_{t-q} \widetilde{\boldsymbol{B}}_{q,r}, \tag{7.24}$$

where $\boldsymbol{\Omega}$, $\widetilde{\boldsymbol{A}}_{p,r}$, and $\widetilde{\boldsymbol{B}}_{q,r}$ are the $(n \times n)$ matrices of parameters, $P \in \mathbb{N}$, $Q \in \mathbb{N}_0$, and $R \in \mathbb{N}$. The matrix $\boldsymbol{\Omega}$ must be symmetric. The conditional covariance matrix is clearly positive definite if $\boldsymbol{\Omega}$ is positive definite (since the remaining terms are positive semidefinite). The summation limit R determines the generality of conditional covariances.

Various special versions of this scheme have been discussed in the literature. The matrices of coefficients can be full rank, symmetric, reduced rank, diagonal, etc. None of these eventualities affects the positive definiteness and symmetry. For instance, one may assume the *scalar* BEKK(1,1,1) model with $\widetilde{\boldsymbol{A}}_{1,1} = \widetilde{\alpha}_{1,1} \boldsymbol{I}$, $\widetilde{\alpha}_{1,1} \geq 0$, and $\widetilde{\boldsymbol{B}}_{1,1} = \widetilde{\beta}_{1,1} \boldsymbol{I}$, $\widetilde{\beta}_{1,1} \geq 0$, or the *diagonal* BEKK(1,1,1), where $\widetilde{\boldsymbol{A}}_{1,1}$ and $\widetilde{\boldsymbol{B}}_{1,1}$ are diagonal with non-negative elements (to avoid observationally equivalent structures).

Note also that the BEKK model is a special VEC model in accordance with Engle (2009, Lemma 3.5). Remind that estimation of all these models is usually provided by the conditional maximum likelihood procedure assuming a certain distribution of $\boldsymbol{Z}_t$ (the multivariate standard normal distribution is commonly supposed). See Bauwens et al. (2006) for further insights.

7.2.3 Constant conditional correlations

Another alternative to conventional multivariate GARCH models was proposed by Bollerslev (1990). The *constant conditional correlation (CCC) model* decom-

poses the matrix $\boldsymbol{H}_t$ as follows:

$$\boldsymbol{H}_t = \mathsf{diag}(\boldsymbol{H}_t)^{1/2}\, \boldsymbol{R}\, \mathsf{diag}(\boldsymbol{H}_t)^{1/2}, \tag{7.25}$$

where $\boldsymbol{R} = (\rho_{ij})_{i,j=1}^{n}$ is the $(n \times n)$ conditional correlation matrix, i.e. a symmetric positive definite matrix with $\rho_{ii} = 1$ for all i and $\rho_{ij} \in [-1, 1]$ for all $i \neq j$.

In this model, the conditional correlations between each pair of assets are restricted to be time invariant. The conditional variances of the individual components of $\boldsymbol{X}_t$ described by $\mathsf{diag}(\boldsymbol{H}_t)$ can follow any process, but the conditional covariances must keep the conditional correlations constant. In particular, $h_{ii,t}$ usually follows a univariate GARCH process (for all i).

Estimation is commonly provided by the conditional maximum likelihood procedure (assuming a certain distribution function of $\boldsymbol{Z}_t$). It is possible to accept the two-stage approach. The first stage calibrates univariate models of the conditional variances. The second stage simply evaluates the sample correlation matrix of the standardized residuals, which are given as $\mathsf{diag}(\boldsymbol{H}_t)^{-1/2}\boldsymbol{X}_t$.

7.2.4 Dynamic conditional correlations

The assumption of constant conditional correlations adopted in the previous section may seem too restrictive. Engle and Sheppard (2001) and Engle (2002) extended the CCC modelling framework into a more general scheme. The *dynamic conditional correlation (DCC) model* defines $\boldsymbol{H}_t$ as follows:

$$\boldsymbol{H}_t = \mathsf{diag}(\boldsymbol{H}_t)^{1/2}\, \boldsymbol{R}_t\, \mathsf{diag}(\boldsymbol{H}_t)^{1/2}, \tag{7.26a}$$

$$\boldsymbol{R}_t = \mathsf{diag}(\boldsymbol{Q}_t)^{-1/2}\, \boldsymbol{Q}_t\, \mathsf{diag}(\boldsymbol{Q}_t)^{-1/2}, \tag{7.26b}$$

$$\boldsymbol{Q}_t = \boldsymbol{\Omega} + \alpha \left[\mathsf{diag}(\boldsymbol{H}_t)^{-1/2}\boldsymbol{X}_t\right] \left[\mathsf{diag}(\boldsymbol{H}_t)^{-1/2}\boldsymbol{X}_t\right]^{\top} + \beta \boldsymbol{Q}_{t-1}, \tag{7.26c}$$

where α and β are the scalar parameters, $\boldsymbol{\Omega}$ is the $(n \times n)$ matrix of parameters, and $\boldsymbol{Q}_t$ is the $(n \times n)$ matrix defined using the standardized residuals as above. Similarly as before, the individual conditional variances accumulated in $\mathsf{diag}(\boldsymbol{H}_t)$ can follow any suitable process; usually, the GARCH(p, q) specification is supposed. If $\boldsymbol{Q}_t$ is positive definite, then so is $\boldsymbol{R}_t$, and consequently also $\boldsymbol{H}_t$. To ensure the positive definite $\boldsymbol{Q}_t$, it is sufficient to assume that $\alpha \geq 0$, $\beta \geq 0$, $\alpha + \beta < 1$ and that $\boldsymbol{\Omega}$ is positive definite (Engle, 2002).

The DCC decomposition delivered by (7.26) and specified by the latter conditions is called *mean-reverting*. The structure of $\boldsymbol{Q}_t$ is an evident analogy of the scalar-diagonal multivariate GARCH model; compare with Equation (7.23). Undoubtedly, there naturally exist various alternatives to this particular modelling representation. Refer to Section 7.2.2. For instance, Engle (2002) also suggested

the *integrated* DCC model as follows ($\lambda \in (0,1)$):

$$\boldsymbol{H}_t = \mathsf{diag}(\boldsymbol{H}_t)^{1/2}\, \boldsymbol{R}_t\, \mathsf{diag}(\boldsymbol{H}_t)^{1/2}, \tag{7.27a}$$

$$\boldsymbol{R}_t = \mathsf{diag}(\boldsymbol{Q}_t)^{-1/2}\, \boldsymbol{Q}_t\, \mathsf{diag}(\boldsymbol{Q}_t)^{-1/2}, \tag{7.27b}$$

$$\boldsymbol{Q}_t = (1-\lambda)\left[\mathsf{diag}(\boldsymbol{H}_t)^{-1/2}\boldsymbol{X}_t\right]\left[\mathsf{diag}(\boldsymbol{H}_t)^{-1/2}\boldsymbol{X}_t\right]^\top + \lambda \boldsymbol{Q}_{t-1}. \tag{7.27c}$$

This particular relation for $\boldsymbol{Q}_t$ is a direct analogy of the exponential smoother discussed in Section 7.2.1. The process describing $\boldsymbol{Q}_t$ has the unit root. Thus, the covariances have no tendency to revert to a constant level. It contrasts with the mean-reverting scheme (7.26).

These models are commonly estimated by the conditional maximum likelihood technique (the multivariate Gaussian innovations $\boldsymbol{Z}_t$ are commonly supposed). As can be seen (Engle, 2002), the two-stage method can be adopted (by decomposing the conditional log-likelihood criterion based on the Gaussian innovations). The first stage calibrates the individual conditional variance terms. The second one investigates the behaviour of the standardized residuals. See Engle (2002) or Engle (2009, Section 4.4) for more details.

Note that there exists a modification of the DCC model that evaluates $\boldsymbol{R}_t$ directly without the normalization of $\boldsymbol{Q}_t$ (Tse & Tsui, 2002). In that case, it is necessary to ensure that $\boldsymbol{R}_t$ has the properties of correlation matrices by different instruments. On the other hand, both modelling concepts are almost similar. See Bauwens et al. (2006) or Engle (2009, Section 4.3).

7.2.5 Orthogonal GARCH models

Another possible generalization of the CCC model supposes that a non-singular linear combination of components of $\boldsymbol{X}_t$ respects a CCC structure. Particularly, assume that there exists an $(n \times n)$ constant regular matrix $\boldsymbol{P}$ such that:

$$\mathsf{cov}(\boldsymbol{P}\boldsymbol{X}_t|\mathcal{F}_{t-1}) = \boldsymbol{D}_t^{1/2}\, \boldsymbol{R}\, \boldsymbol{D}_t^{1/2}, \tag{7.28}$$

where $\boldsymbol{D}_t$ is the $(n\times n)$ diagonal matrix involving individual conditional variances of the components of $\boldsymbol{P}\boldsymbol{X}_t$ on its diagonal, i.e. $\boldsymbol{D}_t = \mathsf{diag}[\mathsf{cov}(\boldsymbol{P}\boldsymbol{X}_t|\mathcal{F}_{t-1})]$, and $\boldsymbol{R}$ is the $(n\times n)$ constant correlation matrix associated with the transformation $\boldsymbol{P}\boldsymbol{X}_t$. See also Section 7.2.3. Moreover, one may express the conditional covariance matrix of $\boldsymbol{X}_t$ as:

$$\boldsymbol{H}_t = \boldsymbol{P}^{-1}\, \boldsymbol{D}_t^{1/2}\, \boldsymbol{R}\, \boldsymbol{D}_t^{1/2}\, (\boldsymbol{P}^\top)^{-1}. \tag{7.29}$$

It is evident that the covariances of $\boldsymbol{X}_t$ are not driven by the CCC model in this modelling framework, even though $\boldsymbol{P}\boldsymbol{X}_t$ are (unless $\boldsymbol{P}$ is diagonal). The trans-

formation matrix $\boldsymbol{P}$ can readily involve unknown parameters to be estimated. Most often, the assumption that $\boldsymbol{R} = \boldsymbol{I}$ is established in this context. Remark that the distinct choices of $\boldsymbol{P}$ apparently lead to different conditional covariance (correlation) models. Refer to Bauwens et al. (2006) for more examples.

The most widely used version of this approach advocated by Alexander (2002) is called the *orthogonal GARCH (OGARCH)* or *principal component GARCH model*. This technique creates unconditionally uncorrelated linear combinations of the elements of $\boldsymbol{X}_t$. The univariate GARCH models are calibrated for all of these transformations. Consequently, the full conditional covariance matrix is reconstructed by assuming that all conditional correlations are zero.

To be more precise, let $\mathsf{cov}(\boldsymbol{X}_t) = \boldsymbol{\Sigma}$ be a symmetric positive definite matrix with the associated spectral decomposition $\boldsymbol{\Sigma} = \boldsymbol{P}^{-1}\boldsymbol{\Lambda}(\boldsymbol{P}^{\top})^{-1}$, where $\boldsymbol{P}^{-1}$ is the matrix of eigenvectors of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ is the diagonal matrix with the corresponding eigenvalues. Assume that the conditional covariance matrix of the transformation $\boldsymbol{P}\boldsymbol{X}_t$ satisfies the following relation:

$$\mathsf{cov}(\boldsymbol{P}\boldsymbol{X}_t|\mathcal{F}_{t-1}) = \boldsymbol{D}_t, \quad \text{i.e.} \quad \boldsymbol{H}_t = \boldsymbol{P}^{-1}\,\boldsymbol{D}_t\,(\boldsymbol{P}^{\top})^{-1}, \tag{7.30}$$

where each component of the diagonal matrix $\boldsymbol{D}_t$ follows a univariate GARCH process. Moreover, it holds $\mathbb{E}(\boldsymbol{D}_t) = \boldsymbol{\Lambda}$.

A closely related alternative considers that $\boldsymbol{X}_t$ is transformed to have unit unconditional variance at first so that the eigenvectors and eigenvalues are computed from the unconditional correlation matrix of $\boldsymbol{X}_t$. This particularly means that $\mathsf{diag}(\boldsymbol{\Sigma})^{-1/2}\boldsymbol{\Sigma}\mathsf{diag}(\boldsymbol{\Sigma})^{-1/2}$ is spectrally decomposed as $\widetilde{\boldsymbol{P}}^{-1}\widetilde{\boldsymbol{\Lambda}}(\widetilde{\boldsymbol{P}}^{\top})^{-1}$, where the notation remains as before (except for the tildes). The OGARCH model supposes that the conditional covariance matrix of the transformation $\widetilde{\boldsymbol{P}}\mathsf{diag}(\boldsymbol{\Sigma})^{-1/2}\boldsymbol{X}_t$ is given by:

$$\mathsf{cov}(\widetilde{\boldsymbol{P}}\mathsf{diag}(\boldsymbol{\Sigma})^{-1/2}\boldsymbol{X}_t|\mathcal{F}_{t-1}) = \boldsymbol{D}_t, \tag{7.31}$$

where each component of the diagonal matrix $\boldsymbol{D}_t$ follows a univariate GARCH model. The conditional covariance matrix of $\boldsymbol{X}_t$ can be calculated as:

$$\boldsymbol{H}_t = \mathsf{diag}(\boldsymbol{\Sigma})^{1/2}\,\widetilde{\boldsymbol{P}}^{-1}\,\boldsymbol{D}_t\,(\widetilde{\boldsymbol{P}}^{\top})^{-1}\,\mathsf{diag}(\boldsymbol{\Sigma})^{1/2}. \tag{7.32}$$

Both previous OGARCH representations are usually calibrated in two steps. The first one computes the sample covariance (correlation) matrix and its spectral decomposition from the given data sample. The second step estimates univariate GARCH models of the delivered linear combinations, then constructs the conditional covariance matrix.

8. Conditional covariance modelling: An approach based on state space representations

The modelling techniques introduced in Section 7.2.5 have motivated development of another methodological approach to conditional covariances and correlations. To be more precise, the essential idea that a non-singular linear combination of elements of $\boldsymbol{X}_t$ follows the CCC representation has been replicated and further extended in a specific way. In particular, the recommended orthogonal transformation is non-constant. It varies over time. Thus, it should reflect the structure of multivariate time series more accurately. Compare with Equation (7.28).

This chapter proposes a novel approach to conditional covariance modelling. In particular, a class of multivariate generalized autoregressive conditional heteroscedasticity models is suggested. It is based on a dynamic orthogonal transformation of $\boldsymbol{X}_t$, which is derived by the LDL factorization of the conditional covariance matrix. As was indicated before, one supposes that the transformed time series respects the constant conditional correlation (CCC) scheme, where all associated conditional correlations are assumed to be zero. Refer to Equation (7.29) and the remarks below it. Moreover, the time-varying transformation can be specified by nonlinear discrete-time state space representations under corresponding assumptions (amongst other alternatives). The three following sections outline the suggested modelling framework in greater detail.

Section 8.1 recapitulates properties of the LDL matrix decomposition and derives the transformation scheme by employing it. Section 8.2 introduces the concrete model implementation. Section 8.3 discusses different calibration methods for this particular modelling technique.

8.1 LDL factorization

Following the algebraic theory, each real symmetric positive definite matrix has a unique LDL decomposition (Harville, 1997, Chapter 14). Let the conditional covariance matrix $\boldsymbol{H}_t$ defined by (7.15) have the LDL reparameterization in the standard form, i.e. given by:

$$\boldsymbol{H}_t = \boldsymbol{L}_t \boldsymbol{D}_t \boldsymbol{L}_t^\top, \tag{8.1}$$

where $\boldsymbol{L}_t = (\ell_{ij,t})_{i,j=1}^n$ is the $(n \times n)$ lower triangular matrix with the unit diagonal and $\boldsymbol{D}_t = (d_{ij,t})_{i,j=1}^n$ is the $(n \times n)$ diagonal matrix with positive elements $d_{ii,t}$ on its diagonal. Particularly, $\det(\boldsymbol{L}_t) = 1$, $\boldsymbol{L}_t$ is invertible, and the inverted matrix $\boldsymbol{L}_t^{-1} = (\ell_t^{ij})_{i,j=1}^n$ is also the $(n \times n)$ lower triangular matrix with the unit diagonal elements. It is noteworthy that the decomposition (8.1) evidently requires no additional (parameter) constraints for $\boldsymbol{H}_t$ being symmetric and positive definite since this is guaranteed by the LDL structure. Furthermore, one can readily see that the LDL and Cholesky factorization of $\boldsymbol{H}_t$ are closely related:

$$\boldsymbol{H}_t = \boldsymbol{L}_t \boldsymbol{D}_t \boldsymbol{L}_t^\top = (\boldsymbol{L}_t \boldsymbol{D}_t^{1/2})(\boldsymbol{L}_t \boldsymbol{D}_t^{1/2})^\top = \boldsymbol{H}_t^{1/2} (\boldsymbol{H}_t^{1/2})^\top, \tag{8.2}$$

where $\boldsymbol{H}_t^{1/2}$ is the $(n \times n)$ lower triangular matrix with positive diagonal terms.

The LDL factorization (8.1) delivers uniquely determined recurrent relations for the elements of $\boldsymbol{H}_t$ evaluated by the components of $\boldsymbol{L}_t$ and $\boldsymbol{D}_t$ (Harville, 1997, Chapter 14). More precisely, one can easily derive the following formulas for conditional covariances and correlations:

$$h_{ii,t} = \mathsf{var}(X_{i,t}|\mathcal{F}_{t-1}) = \sum_{v=1}^{i} \ell_{iv,t}^2 d_{vv,t}, \quad i = 1, \ldots, n, \tag{8.3a}$$

$$h_{ij,t} = \mathsf{cov}(X_{i,t}, X_{j,t}|\mathcal{F}_{t-1}) = \sum_{v=1}^{j} \ell_{iv,t} \ell_{jv,t} d_{vv,t}, \quad j < i,\ i = 2, \ldots, n, \tag{8.3b}$$

$$\rho_{ij,t} = \mathsf{corr}(X_{i,t}, X_{j,t}|\mathcal{F}_{t-1}) = \frac{h_{ij,t}}{\sqrt{h_{ii,t}}\sqrt{h_{jj,t}}}, \quad i, j = 1, \ldots, n, \tag{8.3c}$$

where one puts $\ell_{vv,t} = 1$, $v = 1, \ldots, n$.

The matrix $\boldsymbol{L}_t$ naturally provides the required orthogonal transformation of $\boldsymbol{X}_t$. It may be specified as (according to Equations (7.15) and (8.2)):

$$\boldsymbol{L}_t^{-1} \boldsymbol{X}_t = \boldsymbol{D}_t^{1/2} \boldsymbol{Z}_t. \tag{8.4}$$

The conditional and unconditional moments of the transformed stochastic process $\boldsymbol{L}_t^{-1} \boldsymbol{X}_t$ can be calculated under the given assumptions as follows:

$$\mathbb{E}(\boldsymbol{L}_t^{-1} \boldsymbol{X}_t|\mathcal{F}_{t-1}) = \boldsymbol{0}, \quad \mathsf{cov}(\boldsymbol{L}_t^{-1} \boldsymbol{X}_t|\mathcal{F}_{t-1}) = \boldsymbol{D}_t, \tag{8.5a}$$

$$\mathbb{E}(\boldsymbol{L}_t^{-1} \boldsymbol{X}_t) = \boldsymbol{0}, \quad \mathsf{cov}(\boldsymbol{L}_t^{-1} \boldsymbol{X}_t) = \mathbb{E}(\boldsymbol{D}_t), \tag{8.5b}$$

$$\mathsf{cov}(\boldsymbol{L}_t^{-1} \boldsymbol{X}_t, \boldsymbol{L}_{t+h}^{-1} \boldsymbol{X}_{t+h}) = \boldsymbol{0}, \quad h \in \mathbb{Z} \backslash \{0\}. \tag{8.5c}$$

Note that the relations given by Equation (8.5a) are in accordance with the motivation outlined in the introduction of this chapter. Namely, the suggested dynamic transformation of $\boldsymbol{X}_t$ follows the framework considered in Section 7.2.5; it cor-

responds to the interpretation of Equation (7.28) assuming $\boldsymbol{R} = \boldsymbol{I}$. Moreover, Equation (8.4) can be concisely rewritten as:

$$\boldsymbol{X}_t = -\left(\boldsymbol{L}_t^{-1} - \boldsymbol{I}\right)\boldsymbol{X}_t + \boldsymbol{D}_t^{1/2}\boldsymbol{Z}_t, \tag{8.6}$$

or more explicitly as:

$$\begin{aligned} X_{1,t} &= \sqrt{d_{11,t}}Z_{1,t}, \\ X_{2,t} &= -\ell_t^{21}X_{1,t} + \sqrt{d_{22,t}}Z_{2,t}, \\ &\vdots \\ X_{n,t} &= -\sum_{i=1}^{n-1}\ell_t^{ni}X_{i,t} + \sqrt{d_{nn,t}}Z_{n,t}. \end{aligned} \tag{8.7}$$

All terms containing $X_{j,t}$, $j = 1, \ldots, i-1$, on the right hand side of the ith equation in (8.7) are uncorrelated with the error term $\sqrt{d_{ii,t}}Z_{i,t}$, $i = 2, \ldots, n$, and the components of $\boldsymbol{D}_t^{1/2}\boldsymbol{Z}_t$ are simultaneously uncorrelated. Therefore, the system (8.7) is a direct analogy of the fully recursive simultaneous-equations model with no exogenous variables. It is consistently and asymptotically efficiently estimated by the equation-by-equation maximum likelihood procedures (supposing the standard normal distribution). See e.g. Greene (2003, Section 15.6).

Note that the proposed transformation is apparently ordering dependent, i.e. $X_{1,t}$ influences $X_{2,t}, \ldots, X_{n,t}$, $X_{2,t}$ influences $X_{3,t}, \ldots, X_{n,t}$, etc. It follows from the character of the considered decomposition, which is commonly employed in statistics. Remind that the analogous question also occurs in other modelling methods when the usually accepted Cholesky factorization (and implicitly also the LDL decomposition) is adopted to produce $\boldsymbol{H}_t^{1/2}$ in the model (7.15). See e.g. Engle (2002, 2009) or Bauwens et al. (2006). Despite this fact, these methods and modelling schemes have been successfully applied in practice.

8.2 Model implementation

The LDL factorization of the conditional covariance matrix $\boldsymbol{H}_t$ translates the original model (7.15) into the system (8.7) with the properties studied in Section 8.1. Namely, the ith equation of this system ($i \geq 2$) can be simply viewed as the linear regression of X_i on $X_1, \ldots, X_{i-1}$ with the particular coefficients and conditional heteroscedasticity variance. Moreover, the structure of (8.7) is analogous to the fully recursive systems of linear regression equations with uncorrelated errors (Greene, 2003, Section 15.6).

This arrangement evokes miscellaneous methodological instruments, which might be applicable in this specific context. For instance, one can implement a nonlinear discrete-time state space modelling representation, where the time-varying transformation coefficients follow an unobservable process and Equations (8.7) determine the signal (measurement) equations of this state space scheme. This particular approach has been successfully employed in many different situations. See Durbin and Koopman (2001) or Brockwell and Davis (2002).

According to (8.7) and the previous discussion, the particular model implementation may be formulated as follows:

$$X_{1,t} = \sqrt{d_{11,t}} Z_{1,t}, \tag{8.8a}$$
$$d_{11,t} = c_1 + a_1 X_{1,t-1}^2 + b_1 d_{11,t-1}, \tag{8.8b}$$

$$\boldsymbol{\beta}_{t+1}^i = \boldsymbol{\mu}_i + \boldsymbol{\Phi}_i \boldsymbol{\beta}_t^i + \boldsymbol{\varepsilon}_{t+1}^i, \tag{8.8c}$$
$$X_{i,t} = -(X_{1,t}, \ldots, X_{i-1,t}) \boldsymbol{\beta}_t^i + \sqrt{d_{ii,t}} Z_{i,t}, \tag{8.8d}$$
$$d_{ii,t} = c_i + a_i \left(X_{i,t-1} + (X_{1,t-1}, \ldots, X_{i-1,t-1}) \boldsymbol{\beta}_{t-1}^i \right)^2 + b_i d_{ii,t-1}, \tag{8.8e}$$

where $\boldsymbol{\beta}_t^i = (\ell_t^{i1}, \ldots, \ell_t^{i,i-1})^\top$, $i = 2, \ldots, n$. Further, a_i, b_i, and c_i are scalar parameters, $i = 1, \ldots, n$. In (8.8c), $\boldsymbol{\mu}_i$ is the vector of $(i-1)$ parameters and $\boldsymbol{\Phi}_i$ denotes the $((i-1) \times (i-1))$ matrix of parameters, $i = 2, \ldots, n$. A sequence of i. i. d. random vectors $\{\boldsymbol{\varepsilon}_t^i\}_{t \in \mathbb{Z}}$ with finite second moments fulfils $\mathbb{E}(\boldsymbol{\varepsilon}_t^i) = \mathbf{0}$ and $\mathsf{cov}(\boldsymbol{\varepsilon}_t^i) = \boldsymbol{M}_i$, where $\boldsymbol{M}_i$ is the $((i-1) \times (i-1))$ matrix of parameters, $i = 2, \ldots, n$. Moreover, one supposes that $\{\boldsymbol{\varepsilon}_t^i\}_{t \in \mathbb{Z}}$ and $\{\boldsymbol{\varepsilon}_t^j\}_{t \in \mathbb{Z}}$ are mutually and serially independent at all times for $i, j = 2, \ldots, n$, $i \neq j$, and that the same holds for $\{Z_{i,t}\}_{t \in \mathbb{Z}}$ and $\{\boldsymbol{\varepsilon}_{t+1}^j\}_{t \in \mathbb{Z}}$, $i = 1, \ldots, n$, $j = 2 \ldots, n$. Finally, $\{Z_{i,t}\}_{t \in \mathbb{Z}}$ and $\{\boldsymbol{\varepsilon}_{t+1}^j\}_{t \in \mathbb{Z}}$ are assumed to be uncorrelated with $\boldsymbol{\beta}_t^k, \boldsymbol{\beta}_{t-1}^k, \ldots$, $i = 1, \ldots, n$, $j, k = 2 \ldots, n$.

The structure of the introduced model and the role of the included parameters should be further clarified. The formulas (8.8a) and (8.8d) entirely respect the transformation equations introduced in (8.7). The terms $d_{ii,t}$ in (8.8b) or (8.8e), i.e. the diagonal components of $\boldsymbol{D}_t$, are modelled by the GARCH(1,1) process with the associated parameters a_i, b_i, and c_i. It is motivated by the joint modelling practice. Implicitly, this brings several restrictions on the parameters c_i, a_i, and b_i, e.g. $c_i > 0$, $a_i \geq 0$, $b_i \geq 0$, and $a_i + b_i < 1$ are sufficient conditions for $d_{ii,t}$ being positive and (weakly) stationary (Tsay, 2005, Section 3.5). The vector $\boldsymbol{\beta}_t^i$ appearing in (8.8d) includes the unknown elements of the ith row of the matrix $\boldsymbol{L}_t^{-1}$. As was previously remarked, Equation (8.8d) can be interpreted as the regression problem with the randomly varying coefficients. The states $\boldsymbol{\beta}_t^i$ are

represented by (8.8c) as unobservable vector autoregressive processes driven by the parameters in $\boldsymbol{\mu}_i$, $\boldsymbol{\Phi}_i$, and $\boldsymbol{M}_i$. Recall that the condition $\det(\boldsymbol{I} - z\boldsymbol{\Phi}_i) \neq 0$, $|z| \leq 1$, ensures the stationary states (Lütkepohl, 2005, Chapter 2).

The model clearly includes a relatively high number of parameters in its general form. On the other hand, it is possible to adopt various reasonable restrictions that effectively reduce the number of parameters as in the case of the VEC model, the BEKK model, etc. Refer also to Sections 7.2.2 and 9.2.

The motivation for formulating this model has been purely pragmatic, e.g. the off-diagonal elements of $\boldsymbol{\Phi}_i$ have no straightforward interpretation. The suggested modelling framework undoubtedly reduces the general principles involved in the initially considered model described in Section 7.1. It introduces the particular modelling regime similarly to other common representations; see Section 7.2. Note that there are several resemblances to the stochastic volatility approach (Tsay, 2005, Sections 3.12 and 12.7).

Furthermore, there surely exists a variety of possible model generalizations: (i) $d_{ii,t}$ can be modelled by another (more complex) GARCH type process, (ii) the model formula for $\boldsymbol{\beta}_t^i$ can be extended, e.g. by assuming a higher lag, by taking a (stochastically) time-varying counterpart of the matrix $\boldsymbol{M}_i$, or (iii) $\boldsymbol{\beta}_t^i$ may be expressed by an entirely different state process, etc. On the other hand, the suggested model in the proposed form has empirically demonstrated its capabilities and competitiveness. See Chapter 9 for numerical demonstrations.

8.3 Model calibration

To construct an estimator of $\boldsymbol{H}_t$ by applying the transformation determined by the suggested model (8.8), the state space representation must be calibrated under some distributional assumptions using the given sample $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T\}$. Suppose that the processes $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ and $\{\boldsymbol{\varepsilon}_t^i\}_{t\in\mathbb{Z}}$ are Gaussian (for all i). Each of the n modelling subsystems is calibrated separately according to the accepted covariance structure and given assumptions.

The conditional maximum likelihood procedure available for the standard GARCH(1,1) model estimates the parameters involved in Equations (8.8a) and (8.8b), respectively. See e.g. Tsay (2005, Section 3.5).

Calibration of the remaining $(n-1)$ state space models described by Equations (8.8c)-(8.8e) is less straightforward due to the presence of the stochastically time-varying parameter $d_{ii,t}$, which is not linear in the observations and states. Additionally, the signal equation (8.8d) cannot be simply linearized as, e.g., in the case of the stochastic volatility model (Durbin & Koopman, 2001, Section 10.6).

One could consider calibration procedures suitable for nonlinear state space methods. A relatively extensive body of literature concerns the analysis of nonlinear state space models involving contributions on the extended Kalman filter, the unscented filter, the particle filter, the importance sampling, or Markov Chain Monte Carlo techniques. Consult Anderson and Moore (1979), Durbin and Koopman (2001), Chui and Chen (2013), and the references given therein. However, many calibration procedures related to nonlinear state space modelling techniques are either based on computationally intensive simulation methods (e.g. the Monte Carlo integration) or model approximations. The latter approach seems to be favourable in our context; the dimension of $\boldsymbol{X}_t$ is usually relatively high. Therefore, numerically more effective methodological instruments are preferred accepting the obvious trade-off.

If the proposed GARCH(1,1) updating scheme of the stochastically time-varying variance $d_{ii,t}$ is supposed, the linear Gaussian state space methodology may be implemented into the calibration process as recommended in the literature (Creal, Koopman & Lucas, 2008; Harvey, Ruiz & Santana, 1992). To be more specific, if $d_{ii,t}$ is replaced by an appropriately defined recursive approximation $\widetilde{d}_{ii,t}$ driven only by past observations, the Kalman recursive formulas for predicting, filtering, and smoothing associated with the linear Gaussian state space model can be used to calibrate the proposed model. Refer to Durbin and Koopman (2001, Chapter 4) for the Kalman recursions. In particular, the approximative term $\widetilde{d}_{ii,t}$ has the identical structure to $d_{ii,t}$, but it depends on $\widetilde{\boldsymbol{\beta}}^i_{t-1|t-2}$ and $\widetilde{d}_{ii,t-1}$ instead of $\boldsymbol{\beta}^i_{t-1}$ and $d_{ii,t-1}$, respectively, where $\widetilde{\boldsymbol{\beta}}^i_{t-1|t-2}$ denotes the corresponding predicted state calculated recursively by the classic Kalman recursions. See below. It means that $\widetilde{d}_{ii,t}$ can be effectively calculated jointly with the other updating formulas with respect to the outlined approximative feature. In order to calibrate the ith state space submodel, the sequentially updated predicted states $\widetilde{\boldsymbol{\beta}}^i_{t|t-1}$ and filtered states $\widetilde{\boldsymbol{\beta}}^i_{t|t}$ are obtained by the following recursive relations (Durbin & Koopman, 2001, Section 4.2):

$$\widetilde{v}^i_t = X_{i,t} + (X_{1,t}, \ldots, X_{i-1,t})\, \widetilde{\boldsymbol{\beta}}^i_{t|t-1}, \tag{8.9a}$$

$$\widetilde{\boldsymbol{\beta}}^i_{t|t} = \widetilde{\boldsymbol{\beta}}^i_{t|t-1} + \boldsymbol{S}^i_t \widetilde{v}^i_t / T^i_t, \tag{8.9b}$$

$$\widetilde{\boldsymbol{\beta}}^i_{t+1|t} = \boldsymbol{\mu}_i + \boldsymbol{\Phi}_i \widetilde{\boldsymbol{\beta}}^i_{t|t}, \tag{8.9c}$$

$$\widetilde{d}_{ii,t+1} = c_i + a_i \left[X_{i,t} + (X_{1,t}, \ldots, X_{i-1,t})\, \widetilde{\boldsymbol{\beta}}^i_{t|t-1} \right]^2 + b_i \widetilde{d}_{ii,t}, \tag{8.9d}$$

$$T^i_t = (X_{1,t}, \ldots, X_{i-1,t})\, \boldsymbol{P}^i_{t|t-1}\, (X_{1,t}, \ldots, X_{i-1,t})^\top + \widetilde{d}_{ii,t}, \tag{8.9e}$$

$$\boldsymbol{S}^i_t = \boldsymbol{P}^i_{t|t-1}\, (X_{1,t}, \ldots, X_{i-1,t})^\top, \tag{8.9f}$$

$$\boldsymbol{P}^i_{t|t} = \boldsymbol{P}^i_{t|t-1} - \boldsymbol{S}^i_t (\boldsymbol{S}^i_t)^\top / T^i_t, \tag{8.9g}$$

$$\boldsymbol{P}^i_{t+1|t} = \boldsymbol{\Phi}_i \boldsymbol{P}^i_{t|t} \boldsymbol{\Phi}_i^\top + \boldsymbol{M}_i, \quad t = 1, \ldots, T, \tag{8.9h}$$

$$\widetilde{\boldsymbol{\beta}}^i_{1|0} = \boldsymbol{0}, \quad \boldsymbol{P}^i_{1|0} = \kappa \boldsymbol{I},\ \kappa > 0, \quad \widetilde{d}_{ii,1} = C \geq 0. \tag{8.9i}$$

The fixed-interval smoothed states $\widetilde{\boldsymbol{\beta}}^i_{t|T}$ that complete the previous formulas are computed as follows (Durbin & Koopman, 2001, Section 4.3):

$$\widetilde{\boldsymbol{\beta}}^i_{t|T} = \widetilde{\boldsymbol{\beta}}^i_{t|t} + \boldsymbol{P}^i_{t|t} \boldsymbol{\Phi}_i^\top (\boldsymbol{P}^i_{t+1|t})^{-1} \left[\widetilde{\boldsymbol{\beta}}^i_{t+1|T} - \widetilde{\boldsymbol{\beta}}^i_{t+1|t} \right], \quad t = T, \ldots, 1. \tag{8.10}$$

Notice that the fixed-interval smoother requires inverting the matrix $\boldsymbol{P}^i_{t+1|t}$. This numerical drawback can be eliminated by adjusting the algorithm as in Durbin and Koopman (2001, Section 4.3.1). Consequently, the estimator of the conditional covariance matrix $\boldsymbol{H}_t$ is constructed using $\widetilde{d}_{ii,t}$ and the corresponding (filtered or smoothed) estimated states with respect to the considered LDL decomposition. Compare with Equation (7.29).

Accepting pragmatic reasons as many other authors, the normality has been assumed in the model. Consequently, the unknown parameters of the ith modelling formula ($i \geq 2$) can be estimated applying the corresponding maximum likelihood method associated with the linear Gaussian state space model using the previously described approximation of $d_{ii,t}$ and the diffuse prior settings, i.e. $\boldsymbol{\beta}^i_1 \sim N(\boldsymbol{0}, \kappa \boldsymbol{I})$, where $\kappa \to \infty$. See Durbin and Koopman (2001, Chapter 5). In practice, the κ is an arbitrarily chosen large positive number, e.g. $\kappa = 10^5$. Alternatively, these computational algorithms may be initialized as in Koopman (1997). The resulting method consists in maximizing the following likelihood criterion to estimate the unknown parameters of the ith state space submodel:

$$-\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T} \log|T^i_t| - \frac{1}{2}\sum_{t=1}^{T} \left(\widetilde{v}^i_t\right)^2 / T^i_t. \tag{8.11}$$

For further insights, see Durbin and Koopman (2001, Chapter 7).

In the more general case of non-Gaussian distributions, the process of estimation is no longer comfortable. It can be partly viewed as a limitation of the suggested methodological approach. On the other hand, the normality is commonly assumed in the most state space methods.

9. Numerical studies

This chapter numerically examines the suggested modelling approach introduced in Chapter 8 and compares it with other methods, which are commonly applied in the considered context. Firstly, the behaviour and accuracy of the different modelling schemes are investigated using extensive Monte Carlo experiments. Secondly, two empirical examples are studied in more detail, namely (i) bivariate correlations between stock and bond indices and (ii) conditional correlation links amongst the portfolio of six selected non-Euro EU27 currencies.

9.1 Monte Carlo experiments

Models of conditional covariances may be investigated primarily from the conditional correlation modelling perspective. As was mentioned in Chapter 7, the conditional correlations are important measures in financial practice with many successful empirical applications. See Examples 7.0.1 and 7.0.2. Therefore, one might investigate the performance of various modelling techniques by focusing primarily on the correlations as crucial financial inputs. This approach was advocated by Engle (2002).

In the following Monte Carlo experiments, true correlation structures are supposed to be known. Thus, they can be compared with their estimators by several criteria, namely in terms of simple goodness-of-fit statistics, different diagnostic tests, and tests on portfolio returns. All presented computations have been conducted in EViews 8.0 by implementing original computational procedures for the method proposed in Chapter 8.

9.1.1 Simulation framework

Following Engle (2002), the simple bivariate version of the model (7.15) has been considered for the purposes of this simulation study. In particular, the entire analysis has been reduced to the situation of the (2×2) matrix $\boldsymbol{H}_t$ using the notation introduced in Sections 7.1 and 8.1. The i. i. d. innovations $\boldsymbol{Z}_t$ were sampled from the multivariate standard normal distribution $N(\boldsymbol{0}, \boldsymbol{I})$ generating $T = 1000$ and $T = 5000$ observations. One thousand repetitions were produced in both cases. The chosen lengths correspond approximately to four and twenty years of daily data, respectively.

The following schemes for the correlation function $\rho_t := h_{21,t}/\sqrt{h_{11,t}h_{22,t}}$, $t = 1, \ldots, T$, are considered:

(i) $\rho_t^1 = 0.5$;

(ii) $\rho_t^2 = 0.25 + 0.7\cos\left(\frac{\pi t}{125}\right)$;

(iii) $\rho_t^3 = 0.25 + 0.7\cos\left(\frac{\pi t}{20}\right)$;

(iv) $\rho_t^4 = 0.5 + 0.4 \cdot 1_{[t>T/2]}$, $1_{[t>T/2]}$ equals 1 for $t > T/2$ and 0 otherwise;

(v) $\rho_t^5 = \mathrm{mod}(t, 200)/200$, $\mathrm{mod}(\cdot, 200)$ denotes the modulo operator (namely, the remainder after division of $\cdot$ by 200);

(vi) ρ_t^6 is a scale-adjusted realization of an ARMA(1,1) process with $N(0,1)$ innovations;

(vii) ρ_t^7 is a scale-adjusted realization of an ARIMA(1,1,1) process with $N(0,1)$ innovations.

Such correlations were chosen because they exhibit jump changes, gradual changes, periodical behaviour without changes, and random evolution. The conditional volatilities have been evaluated according to the following scheme:

$$h_{11,t} = 0.01 + 0.05X_{1,t-1}^2 + 0.94h_{11,t-1}, \tag{9.1a}$$

$$h_{22,t} = 0.50 + 0.20X_{2,t-1}^2 + 0.50h_{22,t-1}, \tag{9.1b}$$

$$X_{1,t} = \sqrt{h_{11,t}}Z_{1,t}, \tag{9.1c}$$

$$X_{2,t} = \frac{h_{21,t}}{\sqrt{h_{11,t}}}Z_{1,t} + \sqrt{h_{22,t} - \left(\frac{h_{21,t}}{\sqrt{h_{11,t}}}\right)^2}Z_{2,t}, \quad t = 1, \ldots, T, \tag{9.1d}$$

where the last two equations originate from (7.15) and the Cholesky factorization of the conditional covariance matrix $\boldsymbol{H}_t$. The first conditional variance series is highly persistent while the second one not. Remark that other experiments were also conducted. For instance, various error distributions or data-generating parameters were tested. Nonetheless, the results have remained mostly similar.

The following modelling techniques have been compared: (i) the multivariate moving average (7.17) with $M = 100$ (*MA100*); (ii) the multivariate exponentially weighted moving average (7.20) with $\lambda = 0.94$ (*EWMA0.94*); (iii) the scalar BEKK(1,1,1) model given by (7.24) and the discussion below (*sBEKK*); (iv) the diagonal BEKK(1,1,1) model given by (7.24) and the discussion below (*dBEKK*); (v) the constant conditional correlations (7.25) with the univariate GARCH(1,1) processes (*CCC*); (vi) the mean-reverting dynamic conditional correlations (7.26) with the univariate GARCH(1,1) processes (*DCCmr*); (vii) the integrated dynamic conditional correlations (7.27) with the univariate GARCH(1,1) processes (*DCCint*); (viii) the model (8.8) with $\mu_2 = 0$ and $\Phi_2 = 1$, i.e. β_t^2 has followed the random walk, using filtered states (*SSpaceRWf*); (ix) the

model (8.8) with $\mu_2 = 0$ and $\Phi_2 = 1$, i.e. β_t^2 has followed the random walk, using smoothed states (*SSpaceRWs*); (x) the fully parameterized model (8.8) using filtered states (*SSpaceFf*); (xi) the fully parameterized model (8.8) using smoothed states (*SSpaceFs*). It should be also pointed out that the ordering problem discussed thoroughly in Section 8.1 has been tested in various experiments. In fact, the results have not been substantially affected by these changes.

The alternative correlation estimators have been examined by means of several performance measures. Firstly, the estimated correlations $\widehat{\rho}_t$ have been compared with their true counterparts ρ_t using the mean absolute error, which is simply defined as:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|\rho_t - \widehat{\rho}_t|. \tag{9.2}$$

More specifically, one can contrast (i) the total sums of all obtained mean absolute errors and (ii) the means of all obtained mean absolute errors.

Secondly, the multivariate Ljung-Box test has been considered as another performance measure. It tests the standardized residual series defined by the estimate of $\boldsymbol{H}_t$ as $\widehat{\boldsymbol{Z}}_t = \widehat{\boldsymbol{H}}_t^{-1/2}\boldsymbol{X}_t$ for the presence of serial correlations (Tsay, 2005, Section 8.1.4). Seven lags are used. Thus, the number of rejections (using the corresponding 5% critical value) may be the performance measure of the distinct estimators. The more rejections mean the more evidence that the calculated standardized residuals involve remaining time-varying structures.

Thirdly, the test for autocorrelations of the squared standardized residuals (a variant of the ARCH LM test) has been delivered (Engle, 1982). The test has been computed as the Lagrange multiplier (LM) test from the regression of $\widehat{Z}_{1,t}^2$ and $\widehat{Z}_{2,t}^2$ on five lags of the squared and cross products of the standardized residuals and an intercept. In particular, the number of rejections (using the corresponding 5% critical value) can be the performance measure. The more rejections mean the more evidence that the squared standardized residuals contain remaining time-varying structures.

Finally, another performance measure has been supposed. It has been based on testing on portfolio returns, $\boldsymbol{w}_t^\top\boldsymbol{X}_t$, where $\boldsymbol{w}_t$ is an $\mathcal{F}_{t-1}$-measurable $(n \times 1)$ vector of portfolio weights. Two distinct types of weights have been considered: the *equally weighted portfolio* (EWP), i.e. $\widehat{\boldsymbol{w}}_t = \mathbf{1}/n$, $\mathbf{1}$ is the $(n \times 1)$ vector of units, and the *minimum variance portfolio* (MVP), i.e. $\widehat{\boldsymbol{w}}_t = (\widehat{\boldsymbol{H}}_t^{-1}\mathbf{1})/(\mathbf{1}^\top\widehat{\boldsymbol{H}}_t^{-1}\mathbf{1})$. Compare with Example 7.0.1. Particularly, one can employ the LM test of the ARCH effects in this context. This test is based on the property that the series $\{(\boldsymbol{w}_t^\top\boldsymbol{X}_t)^2/(\boldsymbol{w}_t^\top\boldsymbol{H}_t\boldsymbol{w}_t)\}$ should not exhibit serial correlations (Engle, 1982). Recall that the conditional variance of $\boldsymbol{w}_t^\top\boldsymbol{X}_t$ given $\mathcal{F}_{t-1}$ is evidently $\boldsymbol{w}_t^\top\boldsymbol{H}_t\boldsymbol{w}_t$. The null hypothesis that $\{(\widehat{\boldsymbol{w}}_t^\top\boldsymbol{X}_t)^2/(\widehat{\boldsymbol{w}}_t^\top\widehat{\boldsymbol{H}}_t\widehat{\boldsymbol{w}}_t)\}$ is serially uncorrelated is tested.

Five lags are used here. Similarly as before, the number of rejections (using the corresponding 5% critical value) may serve as the performance measure evaluating the accuracy of the estimators.

Apparently, the mentioned performance measures based on the mean absolute errors have investigated the estimators of conditional correlations. The remaining suggested criteria have revised the adequacy of the estimated conditional covariance structures. The entire Monte Carlo study was repeated five times with various sets of random numbers; one additional experiment was conducted with innovations following the multivariate t-distribution with four degrees of freedom. The results were mostly similar. Therefore, one can accept the drawn conclusions of this study with relative confidence. See Section 9.1.2.

9.1.2 Simulation results

The following paragraphs review the results of the described Monte Carlo experiments. The performance measures introduced in the previous section have been applied. Namely, Figures 9.1 and 9.2 present the total sums of all mean absolute errors for both assumed sample lengths $T = 1000$ and $T = 5000$. In both cases, the estimators based on the proposed state space representation (employing smoothed states) outperform the others. The estimators based on filtered states are beaten only by the mean-reverting dynamic conditional correlation model.

Moreover, Tables 9.1 and 9.2 present the results for the means of all mean absolute errors for both supposed lengths $T = 1000$ and $T = 5000$ according to the different data-generating correlation functions. If the means of all errors are compared across all cases, the previous remarks on the proposed state space modelling technique will remain valid without any substantial changes. In four instances, ρ_t^2, ρ_t^3, ρ_t^5, and ρ_t^7, it delivers the best fit. In the case of ρ_t^4 and ρ_t^6, it is overcome by the other models but the differences are relatively small. In contrast, the suggested technique is not entirely adequate for the constant correlation ρ_t^1.

Tables 9.3 and 9.4 show the results of the Ljung-Box test measure for both sample lengths. Thus, the proposed modelling class based on the state space representation leads to reasonable results overall. It seems to be competitive.

In Tables 9.5 and 9.6, the first component squared standardized residuals have been tested for remaining autocorrelations till five lags for both lengths $T = 1000$ and $T = 5000$. See Section 9.1.1. Overall, the suggested models (with the different calibration techniques) offer comparable results to the others. Similarly, in Tables 9.7 and 9.8, the second component squared standardized residuals have been tested for remaining autocorrelations for both lengths $T = 1000$ and $T = 5000$. See Section 9.1.1. One can reproduce the similar conclusions as before.

Finally, Tables 9.9, 9.10, 9.11, and 9.12 compare the results of the ARCH LM tests on defined portfolio returns for both considered portfolio weights and both sample lengths. See Section 9.1.1. The state space estimators are comparable with the others in all cases not delivering significantly different results.

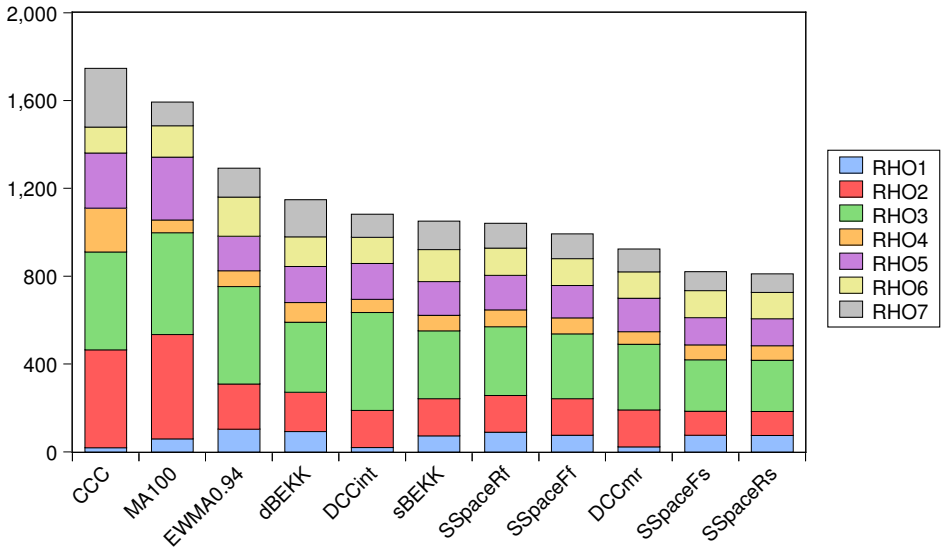


Figure 9.1: Total sums of all mean absolute errors ($T = 1000$).

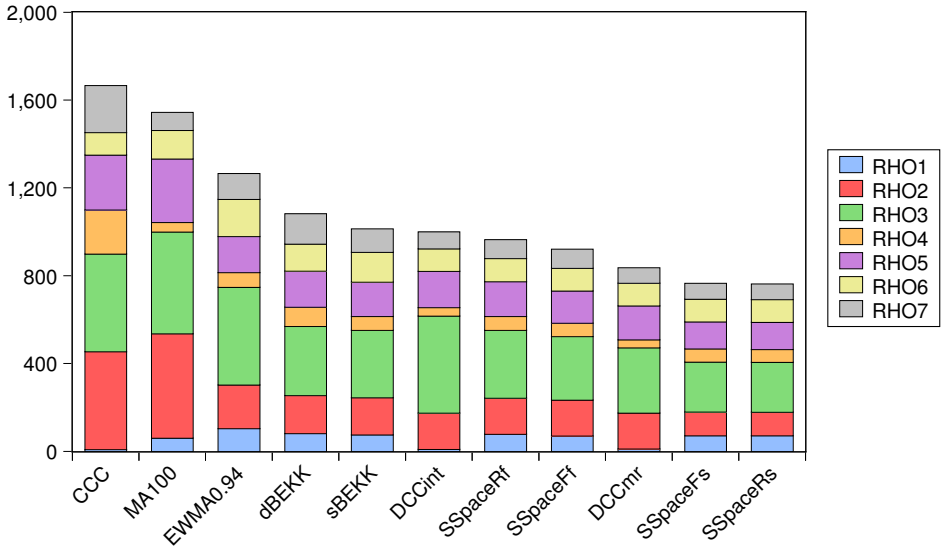


Figure 9.2: Total sums of all mean absolute errors ($T = 5000$).

The results achieved by the suggested modelling technique can be concisely summarized as follows. On the one hand, the state space model with β_t^2 represented by the random walk is indeed straightforward, and it is favourable due to

a lower number of parameters. On the other hand, one might expect the mean-reverting property in the given multivariate financial time series framework. In the case of the filtering estimation, the state space model with the fully parameterized states β_t^2 has demonstrated better results than the one with the random walk states. The differences between state space models using the smoothing estimation are indeed small. Thus, one considers applying both of them.

To conclude this section, the outlined Monte Carlo study has shown that the proposed conditional covariance modelling class (see Chapter 8) is at least comparable and competitive with the other commonly used techniques (according to the given performance measures, which were also used by Engle (2002)). Hence, one can state that the suggested methodology can be accepted in the supposed modelling framework.

Model	ρ_t^1	ρ_t^2	ρ_t^3	ρ_t^4	ρ_t^5	ρ_t^6	ρ_t^7	*Mean*
MA100	0.060	0.475	0.463	**0.059**	0.286	0.143	0.107	0.228
EWMA0.94	0.104	0.206	0.443	0.072	0.158	0.178	0.131	0.185
sBEKK	0.073	0.169	0.309	0.071	0.153	0.146	0.129	0.150
dBEKK	0.093	0.179	0.318	0.091	0.164	0.135	0.169	0.164
CCC	**0.019**	0.446	0.446	0.200	0.251	**0.118**	0.266	0.250
DCCint	0.020	0.170	0.446	0.060	0.163	0.119	0.105	0.155
DCCmr	0.024	0.168	0.299	0.060	0.153	0.120	0.104	0.133
SSpaceRf	0.090	0.167	0.313	0.077	0.158	0.124	0.112	0.149
SSpaceRs	0.075	**0.109**	**0.233**	0.067	**0.123**	0.120	**0.084**	**0.116**
SSpaceFf	0.076	0.167	0.295	0.073	0.148	0.122	0.112	0.142
SSpaceFs	0.077	0.110	0.234	0.068	0.124	0.123	0.086	0.117

Table 9.1: Means of the mean absolute errors ($T = 1000$).

Model	ρ_t^1	ρ_t^2	ρ_t^3	ρ_t^4	ρ_t^5	ρ_t^6	ρ_t^7	*Mean*
MA100	0.060	0.476	0.463	0.043	0.290	0.130	0.082	0.221
EWMA0.94	0.104	0.199	0.444	0.067	0.164	0.170	0.117	0.181
sBEKK	0.075	0.169	0.308	0.063	0.156	0.136	0.107	0.145
dBEKK	0.081	0.174	0.315	0.087	0.164	0.124	0.138	0.155
CCC	**0.009**	0.446	0.445	0.200	0.250	**0.103**	0.214	0.238
DCCint	0.009	0.166	0.445	**0.038**	0.165	0.103	0.077	0.143
DCCmr	0.011	0.164	0.297	0.038	0.155	0.103	0.077	0.122
SSpaceRf	0.078	0.165	0.309	0.063	0.158	0.105	0.086	0.138
SSpaceRs	0.071	**0.107**	0.228	0.059	0.124	0.103	**0.071**	**0.109**
SSpaceFf	0.070	0.163	0.289	0.061	0.146	0.103	0.086	0.131
SSpaceFs	0.072	0.108	**0.227**	0.059	**0.124**	0.103	0.071	0.109

Table 9.2: Means of the mean absolute errors ($T = 5000$).

Model	ρ_t^1	ρ_t^2	ρ_t^3	ρ_t^4	ρ_t^5	ρ_t^6	ρ_t^7	*Mean*
MA100	0.096	0.123	0.116	0.118	0.107	0.092	0.091	0.106
EWMA0.94	0.055	0.061	0.059	0.068	0.082	0.055	0.048	0.061
sBEKK	0.059	0.041	**0.041**	0.061	0.059	0.054	0.043	0.051
dBEKK	0.055	0.039	0.043	0.067	0.056	0.052	0.046	0.051
CCC	0.047	0.058	0.072	0.101	0.068	0.045	0.047	0.063
DCCint	0.047	0.056	0.072	0.059	0.075	0.045	0.046	0.057
DCCmr	0.048	0.056	0.060	0.055	0.060	0.044	**0.042**	0.052
SSpaceRf	0.046	0.041	0.057	0.056	0.053	0.046	0.046	0.049
SSpaceRs	0.048	0.047	0.059	**0.054**	**0.050**	**0.044**	0.046	0.050
SSpaceFf	**0.045**	**0.038**	0.059	0.059	0.052	0.045	0.045	**0.049**
SSpaceFs	0.048	0.047	0.058	0.055	0.057	0.045	0.046	0.051

Table 9.3: Mean fractions of 5% rejections of the Ljung-Box tests ($T = 1000$).

Model	ρ_t^1	ρ_t^2	ρ_t^3	ρ_t^4	ρ_t^5	ρ_t^6	ρ_t^7	*Mean*
MA100	0.105	0.135	0.109	0.112	0.114	0.090	0.095	0.109
EWMA0.94	0.066	0.078	0.061	0.067	0.089	0.073	0.064	0.071
sBEKK	0.070	0.058	0.041	0.068	0.068	0.074	0.065	0.063
dBEKK	0.065	0.055	**0.037**	0.074	**0.065**	0.068	0.053	0.060
CCC	**0.061**	0.065	0.063	0.095	0.091	0.059	0.065	0.071
DCCint	0.062	0.072	0.064	0.055	0.095	0.059	**0.052**	0.066
DCCmr	0.062	0.062	0.051	**0.055**	0.077	0.061	0.053	0.060
SSpaceRf	0.066	**0.049**	0.049	0.059	0.068	0.060	0.057	**0.058**
SSpaceRs	0.065	0.054	0.057	0.058	0.070	0.061	0.057	0.060
SSpaceFf	0.065	0.050	0.050	0.060	0.071	**0.059**	0.058	0.059
SSpaceFs	0.065	0.056	0.057	0.057	0.068	0.060	0.055	0.060

Table 9.4: Mean fractions of 5% rejections of the Ljung-Box tests ($T = 5000$).

Model	ρ_t^1	ρ_t^2	ρ_t^3	ρ_t^4	ρ_t^5	ρ_t^6	ρ_t^7	*Mean*
MA100	0.689	0.988	0.884	0.904	0.849	0.599	0.577	0.784
EWMA0.94	0.191	0.314	0.315	0.292	0.510	0.065	0.066	0.250
sBEKK	0.199	0.077	0.122	0.256	0.112	0.084	0.060	0.130
dBEKK	0.239	0.080	0.130	0.244	0.107	0.108	0.063	0.139
CCC	0.027	**0.023**	0.033	0.027	0.030	0.028	0.032	0.028
DCCint	**0.027**	0.034	0.033	**0.019**	0.039	**0.028**	0.031	0.030
DCCmr	0.027	0.027	0.033	0.020	0.031	0.030	0.032	0.029
SSpaceRf	0.031	0.032	0.031	0.023	**0.027**	0.028	**0.030**	0.029
SSpaceRs	0.032	0.041	0.042	0.022	0.035	0.030	0.035	0.034
SSpaceFf	0.030	0.031	**0.027**	0.020	0.030	0.029	0.032	**0.028**
SSpaceFs	0.032	0.042	0.035	0.025	0.039	0.031	0.034	0.034

Table 9.5: Mean fractions of 5% rejections of the ARCH tests on $\widehat{Z}_{1,t}^2$ ($T = 1000$).

Model	ρ_t^1	ρ_t^2	ρ_t^3	ρ_t^4	ρ_t^5	ρ_t^6	ρ_t^7	*Mean*
MA100	1.000	1.000	1.000	1.000	1.000	0.997	0.998	0.999
EWMA0.94	0.634	0.937	0.998	0.944	0.942	0.095	0.327	0.697
sBEKK	0.694	0.235	0.859	0.905	0.240	0.118	0.287	0.477
dBEKK	0.947	0.134	0.884	0.883	0.227	0.273	0.260	0.515
CCC	0.026	**0.023**	0.022	0.026	0.031	0.019	0.028	0.025
DCCint	0.026	0.023	0.022	0.029	0.042	0.019	0.026	0.027
DCCmr	0.024	0.024	0.023	**0.026**	0.031	0.018	**0.026**	**0.024**
SSpaceRf	0.022	0.025	0.025	0.030	0.029	0.018	0.028	0.025
SSpaceRs	**0.020**	0.049	0.042	0.042	0.060	0.018	0.032	0.038
SSpaceFf	0.022	0.027	**0.021**	0.030	**0.027**	**0.017**	0.030	0.025
SSpaceFs	0.022	0.052	0.038	0.051	0.073	**0.017**	0.032	0.041

Table 9.6: Mean fractions of 5% rejections of the ARCH tests on $\widehat{Z}_{1,t}^2$ ($T = 5000$).

Model	ρ_t^1	ρ_t^2	ρ_t^3	ρ_t^4	ρ_t^5	ρ_t^6	ρ_t^7	*Mean*
MA100	0.972	0.970	0.964	0.977	0.978	0.975	0.974	0.973
EWMA0.94	0.792	0.759	0.773	0.779	0.780	0.787	0.785	0.779
sBEKK	0.869	0.403	**0.078**	0.774	0.434	0.853	0.869	0.612
dBEKK	**0.065**	0.172	0.084	0.458	0.132	**0.045**	0.246	0.172
CCC	0.074	0.718	0.583	0.999	0.860	0.065	0.070	0.481
DCCint	0.074	0.344	0.575	0.168	0.617	0.063	0.067	0.273
DCCmr	0.067	0.281	0.313	0.153	0.300	0.062	0.069	0.178
SSpaceRf	0.130	0.042	0.647	**0.118**	0.149	0.047	0.052	0.169
SSpaceRs	0.135	0.048	0.506	0.143	0.177	0.051	0.056	0.159
SSpaceFf	0.148	**0.036**	0.411	0.126	**0.102**	0.049	**0.050**	**0.132**
SSpaceFs	0.138	0.043	0.341	0.155	0.163	0.049	0.054	0.135

Table 9.7: Mean fractions of 5% rejections of the ARCH tests on $\widehat{Z}_{2,t}^2$ ($T = 1000$).

Model	ρ_t^1	ρ_t^2	ρ_t^3	ρ_t^4	ρ_t^5	ρ_t^6	ρ_t^7	*Mean*
MA100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EWMA0.94	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
sBEKK	1.000	0.994	0.771	1.000	0.998	1.000	1.000	0.966
dBEKK	0.082	0.946	0.890	0.915	0.788	0.069	0.815	0.644
CCC	0.043	1.000	1.000	1.000	1.000	0.048	0.580	0.667
DCCint	0.043	0.914	1.000	0.112	0.998	0.048	0.087	0.457
DCCmr	**0.038**	0.842	0.952	**0.097**	0.785	0.048	**0.084**	0.412
SSpaceRf	0.580	0.086	0.999	0.708	0.558	0.042	0.187	0.452
SSpaceRs	0.572	**0.054**	0.964	0.783	0.599	**0.041**	0.193	0.458
SSpaceFf	0.612	0.076	0.971	0.655	**0.306**	0.050	0.172	**0.406**
SSpaceFs	0.532	0.066	**0.784**	0.774	0.600	0.050	0.187	0.428

Table 9.8: Mean fractions of 5% rejections of the ARCH tests on $\widehat{Z}_{2,t}^2$ ($T = 5000$).

Model	ρ_t^1	ρ_t^2	ρ_t^3	ρ_t^4	ρ_t^5	ρ_t^6	ρ_t^7	*Mean*
MA100	0.514	0.641	0.635	0.519	0.570	0.471	0.463	0.545
EWMA0.94	0.156	0.180	0.238	0.175	0.182	0.159	0.155	0.178
sBEKK	0.194	0.068	**0.025**	0.165	0.066	0.181	0.217	0.131
dBEKK	0.059	0.054	0.047	0.104	0.050	**0.060**	0.102	0.068
CCC	0.067	0.262	0.216	0.076	0.118	0.077	0.216	0.147
DCCint	0.066	0.051	0.211	0.064	0.055	0.077	0.072	0.085
DCCmr	0.063	0.042	0.064	0.066	0.050	0.065	0.073	0.060
SSpaceRf	0.069	0.050	0.099	0.051	0.054	0.074	0.076	0.068
SSpaceRs	0.054	0.041	0.068	0.049	0.045	0.074	**0.059**	**0.056**
SSpaceFf	0.059	0.062	0.095	**0.049**	0.059	0.068	0.076	0.067
SSpaceFs	**0.052**	**0.041**	0.079	**0.049**	**0.045**	0.067	0.060	0.056

Table 9.9: Mean fractions of 5% rejections of the ARCH tests (EWP, $T = 1000$).

Model	ρ_t^1	ρ_t^2	ρ_t^3	ρ_t^4	ρ_t^5	ρ_t^6	ρ_t^7	*Mean*
MA100	0.954	0.991	0.992	0.976	0.986	0.936	0.954	0.970
EWMA0.94	0.366	0.363	0.512	0.392	0.391	0.254	0.336	0.373
sBEKK	0.369	0.066	**0.020**	0.474	0.091	0.302	0.477	0.257
dBEKK	0.071	0.050	0.024	0.221	0.054	**0.046**	0.154	0.089
CCC	0.078	0.543	0.446	0.116	0.203	0.080	0.279	0.249
DCCint	0.078	0.050	0.448	0.062	0.055	0.080	0.078	0.121
DCCmr	0.073	**0.032**	0.051	0.060	**0.037**	0.061	0.077	**0.056**
SSpaceRf	0.066	0.056	0.154	0.053	0.051	0.075	0.081	0.077
SSpaceRs	0.059	0.048	0.074	**0.037**	0.046	0.081	0.061	0.058
SSpaceFf	0.064	0.064	0.126	0.057	0.058	0.068	0.073	0.073
SSpaceFs	**0.050**	0.051	0.084	0.041	0.050	0.065	**0.054**	0.056

Table 9.10: Mean fractions of 5% rejections of the ARCH tests (EWP, $T = 5000$).

Model	ρ_t^1	ρ_t^2	ρ_t^3	ρ_t^4	ρ_t^5	ρ_t^6	ρ_t^7	*Mean*
MA100	0.343	0.500	0.523	0.391	0.421	0.292	0.286	0.394
EWMA0.94	**0.063**	0.092	0.112	0.068	0.069	**0.057**	**0.063**	0.075
sBEKK	0.089	0.033	**0.027**	0.059	**0.028**	0.077	0.075	**0.055**
dBEKK	0.092	0.042	0.028	0.059	0.032	0.071	0.069	0.056
CCC	0.085	0.228	0.188	0.090	0.092	0.080	0.255	0.145
DCCint	0.085	0.109	0.186	0.107	0.104	0.081	0.081	0.107
DCCmr	0.080	0.101	0.099	0.104	0.089	0.065	0.082	0.089
SSpaceRf	0.096	0.040	0.101	0.062	0.036	0.080	0.072	0.070
SSpaceRs	0.080	**0.030**	0.062	**0.048**	0.081	0.081	0.068	0.064
SSpaceFf	0.101	0.039	0.061	0.065	0.032	0.080	0.075	0.065
SSpaceFs	0.088	0.030	0.055	0.050	0.075	0.076	0.065	0.063

Table 9.11: Mean fractions of 5% rejections of the ARCH tests (MVP, $T = 1000$).

Model	ρ_t^1	ρ_t^2	ρ_t^3	ρ_t^4	ρ_t^5	ρ_t^6	ρ_t^7	*Mean*
MA100	0.721	0.961	0.967	0.815	0.880	0.562	0.651	0.794
EWMA0.94	0.075	0.087	0.168	0.095	0.073	**0.052**	**0.050**	0.086
sBEKK	0.098	**0.021**	0.023	0.106	**0.022**	0.074	0.097	0.063
dBEKK	0.100	0.032	**0.020**	**0.078**	0.025	0.060	0.056	**0.053**
CCC	0.082	0.474	0.384	0.087	0.136	0.096	0.321	0.226
DCCint	0.081	0.120	0.384	0.111	0.117	0.097	0.095	0.143
DCCmr	0.082	0.126	0.106	0.109	0.100	0.077	0.091	0.099
SSpaceRf	0.123	0.030	0.111	0.143	0.032	0.095	0.077	0.087
SSpaceRs	0.089	0.036	0.055	0.086	0.182	0.095	0.061	0.086
SSpaceFf	0.110	0.033	0.060	0.134	0.037	0.077	0.078	0.076
SSpaceFs	**0.065**	0.039	0.048	0.084	0.150	0.071	0.061	0.074

Table 9.12: Mean fractions of 5% rejections of the ARCH tests (MVP, $T = 5000$).

9.2 Empirical analyses

To examine the empirical performance of the suggested state space modelling approach to conditional covariances and correlations, two empirical applications are considered. They compare the different modelling techniques through real data examples. Firstly, the daily correlations between logarithmic returns of the stock and bond indices are investigated. Secondly, the daily correlations amongst logarithmic returns in the portfolio of six non-Euro EU27 currencies are evaluated and interpreted.

9.2.1 Stocks and bonds

In the first example, all techniques mentioned in Section 9.1 are compared by studying time-varying correlation links between logarithmic returns of the particular stock and bond indices. From the general perspective, there is no consensus about how stocks and long-term bonds are related. Short-run correlations are regularly affected, e.g. by new announcements. Long-run correlations between these two types of assets should be state dependent, e.g. driven by macroeconomic factors. The way how the correlation links respond to such factors may vary over time (Engle, 2002). Therefore, we shall analyse the daily logarithmic returns of the S&P 500 index and 30-year bond futures. The corresponding data sample from 3rd January 1990 to 30th September 2013 was collected from Yahoo! Finance (2013) as *GSPC* and *TYX* quotations. Table 9.13 delivers some basic sample characteristics.

The dataset has been examined under the identical conditions as in Section 9.1. The estimated conditional correlations are displayed in Figure 9.3. It indicates numerous similarities amongst various correlation models. For instance, one can see the analogies between the state space modelling class and

Statistics	*S&P 500*	*30Y Bonds*
Mean	0.00026	-0.00013
Median	0.00055	0.00000
Maximum	0.10957	0.07512
Minimum	-0.09470	-0.08582
Std. dev.	0.01164	0.01154
Skewness	-0.23378	-0.01909
Kurtosis	11.51722	7.92449

Table 9.13: Sample characteristics of the stock and bond log-returns.

the dynamic conditional correlations. The EWMA and especially the BEKK estimators seem to be more volatile. On the contrary, the main trends in these correlations are comparable with the other estimators. It should be highlighted that the constant conditional correlation is not truly competitive because it remains constant, namely $\widehat{\rho}_t = 0.03944$. The time-varying correlations are mostly negative during the 90's, rather positive after the year 2000, and positive at the end of the observed period with the decreasing tendency in the last analysed year.

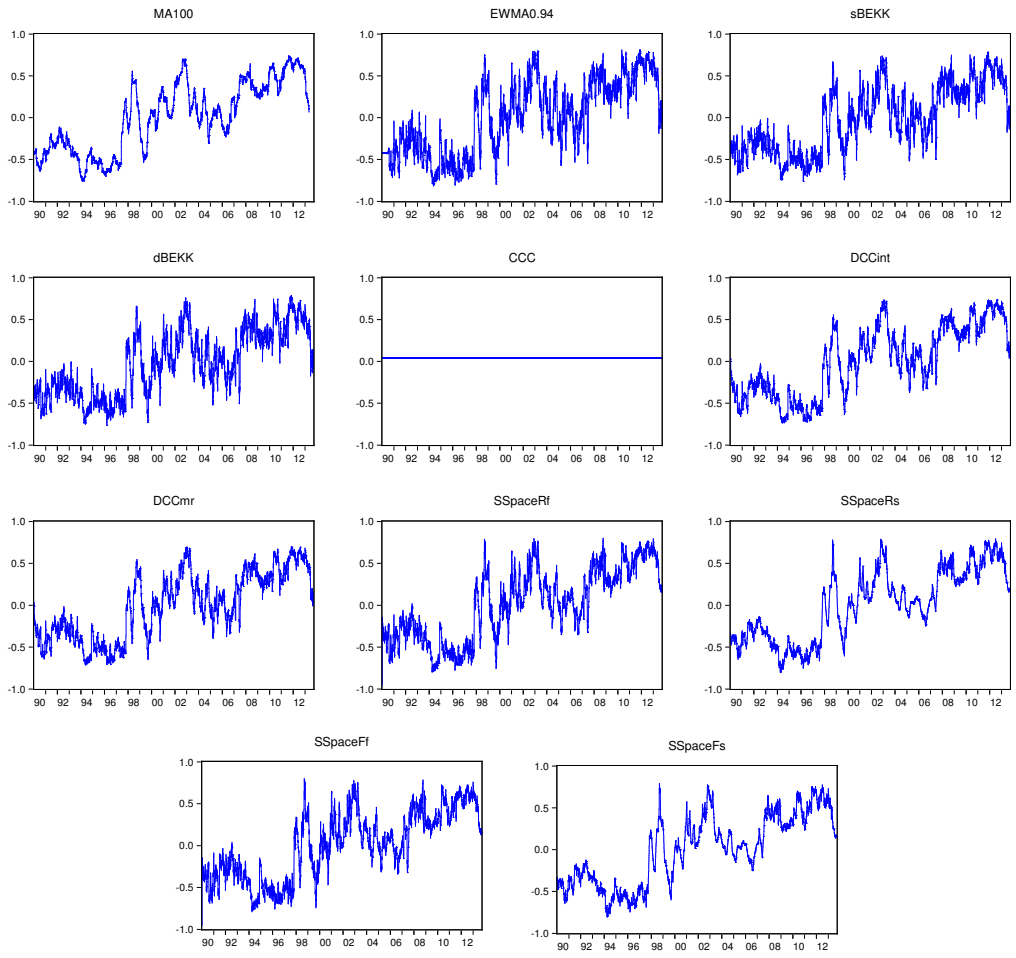


Figure 9.3: Estimated correlations between the S&P 500 index and 30Y Bonds.

Model	MA100	EWMA0.94	sBEKK	dBEKK	DCCint	DCCmr	SSpaceRf	SSpaceRs	SSpaceFf	SSpaceFs
MA100	1.000	0.829	0.831	0.829	0.846	0.843	0.848	0.917	0.843	0.914
EWMA0.94	0.829	1.000	0.997	0.996	0.980	0.986	0.985	0.955	0.985	0.955
sBEKK	0.831	0.997	1.000	1.000	0.983	0.988	0.986	0.958	0.985	0.957
dBEKK	0.829	0.996	1.000	1.000	0.982	0.987	0.984	0.956	0.984	0.956
DCCint	0.846	0.980	0.983	0.982	1.000	0.999	0.988	0.964	0.986	0.963
DCCmr	0.843	0.986	0.988	0.987	0.999	1.000	0.990	0.964	0.988	0.963
SSpaceRf	0.848	0.985	0.986	0.984	0.988	0.990	1.000	0.971	0.999	0.971
SSpaceRs	0.917	0.955	0.958	0.956	0.964	0.964	0.971	1.000	0.970	1.000
SSpaceFf	0.843	0.985	0.985	0.984	0.986	0.988	0.999	0.970	1.000	0.970
SSpaceFs	0.914	0.955	0.957	0.956	0.963	0.963	0.971	1.000	0.970	1.000

Table 9.14: Sample correlations of the estimated time-varying correlations between the S&P 500 index and 30Y bond futures.

Table 9.14 contains the sample correlation matrix of the estimated time-varying correlations. One may identify that the correlations based on the state space modelling approach (8.8) are strongly positively correlated with the other modelling schemes. This is also obvious from Figure 9.3. The smoothed estimators are less variable than the filtered ones according to the character of estimation. Furthermore, Table 9.15 delivers some of the performance measures from Section 9.1.1 for all considered models. One can view that the proposed methods based on the state space representation are comparable with the other commonly applied procedures. Note that the first component squared standardized residuals $\widehat{Z}_{1,t}^2$ are not sufficiently explained by any of the considered models; see the achieved p-values of the ARCH LM test in Table 9.15. On the one hand, it could be resolved by taking a more sophisticated version of the underlying GARCH model into account (with respect to the fact that $\widehat{Z}_{1,t} = X_{1,t}/\sqrt{\widehat{h}_{11,t}}$ by following the Cholesky factorization of $\widehat{\boldsymbol{H}}_t$). On the other hand, for the purposes of this comparison, it suffices to demonstrate that the proposed state space models have analogous empirical features.

Model	$Q(7)$	*ARCH 1*	*ARCH 2*	*ARCH EWP*	*ARCH MVP*
MA100	0.004	0.000	0.000	0.000	0.120
EWMA0.94	0.054	0.001	0.244	0.471	1.000
sBEKK	0.210	0.000	0.239	0.114	0.982
dBEKK	0.230	0.000	0.214	0.168	0.993
CCC	0.034	0.000	0.342	0.000	0.604
DCCint	0.270	0.001	0.365	0.050	0.909
DCCmr	0.261	0.000	0.397	0.062	0.950
SSpaceRf	0.260	0.000	0.314	0.005	0.997
SSpaceRs	0.177	0.005	0.208	0.445	0.999
SSpaceFf	0.224	0.000	0.430	0.007	0.995
SSpaceFs	0.157	0.004	0.304	0.446	0.999

Table 9.15: Comparison of the alternative conditional correlation models: the achieved p-values of the Ljung-Box statistics Q with 7 lags and the ARCH tests on $\widehat{Z}_{1,t}^2$ (ARCH 1), on $\widehat{Z}_{2,t}^2$ (ARCH 2), on EWP returns (ARCH EWP), and on MVP returns (ARCH MVP), all with 5 lags.

Finally, an out-of-sample experiment has been conducted to verify the stability of the delivered estimators. Two samples are assumed: (i) the full data set and (ii) the reduced data sample (last 250 observations, i.e. approximately one year, are omitted). Then, the alternative conditional correlation estimators could be compared contrasting their behaviour in these samples. The moving averages and the exponentially weighted moving averages do not change due to their definition and calibration. See Section 7.2. All remaining estimators (except for constant conditional correlations) show their capabilities by demonstrating stable paths

of time-varying correlation when one compares the estimates based on both supposed samples. Figure 9.4 graphs an example of the estimator obtained by the proposed fully parameterized state space model with smoothed states. The estimators seem to be stable, i.e. the calibrated correlations (almost) coincide. Sample correlations and calculated statistics for the reduced data set remain nearly the same as in the case of the full data sample.

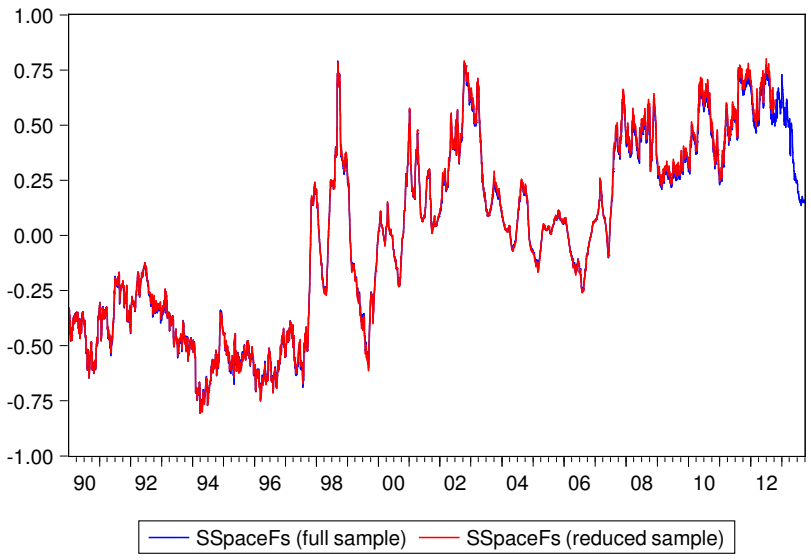


Figure 9.4: The full and reduced data set correlation estimators.

9.2.2 Non-Euro EU27 currencies

In the second example, the exchange rates of selected non-Euro EU27 currencies are examined. In the EU27, 17 member countries used the Euro currency. Other 3 states (Denmark, Latvia, and Lithuania) were members of the ERM II regime (the European Exchange Rate Mechanism II, where the national currencies are allowed to fluctuate around their assigned value with respect to limiting bounds). The Bulgarian Lev was pegged with the Euro. For these reasons, the portfolio of the six remaining non-Euro EU27 currencies has been taken into account, i.e. the Czech crown (CZK), the British pound sterling (GBP), the Hungarian forint (HUF), the Polish zloty (PLN), the Romanian leu (RON), and the Swedish krona (SEK). In this example, the logarithmic returns of the bilateral exchange rates from 3rd January 2000 to 30th September 2013 with the Euro as the denominator were collected (European Central Bank, 2013). Table 9.16 delivers some sample characteristics. The conditional correlation links can be investigated by the techniques assumed and tested in Section 9.1.

Statistics	*PLN*	*HUF*	*RON*	*GBP*	*SEK*	*CZK*
Mean	0.00000	0.00004	0.00025	0.00009	0.00001	-0.00010
Median	-0.00026	-0.00004	0.00000	0.00000	0.00000	-0.00012
Maximum	0.04164	0.05069	0.05001	0.03461	0.02784	0.03165
Minimum	-0.03680	-0.03389	-0.03210	-0.02657	-0.02260	-0.03274
Std. dev.	0.00652	0.00599	0.00537	0.00507	0.00424	0.00394
Skewness	0.45502	0.73601	0.85576	0.33814	0.23366	0.08058
Kurtosis	7.64032	11.40755	10.86120	6.85313	6.41883	9.02455

Table 9.16: Sample characteristics of the exchange rate log-returns.

Only selected graphical outputs of the conditional correlation estimators between the non-Euro EU27 currencies are interpreted here due to the limited space. For instance, three Visegrad countries may be analysed in greater detail, i.e. the Czech crown (CZK), the Hungarian forint (HUF), and the Polish zloty (PLN). Indeed, these currencies have shown interesting correlation links. Figure 9.5 presents the estimated time-varying correlations between the non-Euro Visegrad countries. Particularly, only the mean-reverting dynamic conditional correlation model and the proposed state space model (8.8) using states following the random walk (i.e. $\boldsymbol{\mu}_i = \mathbf{0}$ and $\boldsymbol{\Phi}_i = \boldsymbol{I}$ for all i) with the diagonal matrices $\boldsymbol{M}_i$ for all i (the calibration is provided by the smoothed states) are displayed according to the results obtained in Sections 9.1.2 and 9.2.1. The remaining methods were also tested; they have shown analogous features as before. One should comment the relatively high positive correlations between the Hungarian forint and the Polish zloty, and rather positive correlations between the remaining currencies. Notice that the differences between the two different estimators are clearly caused by the distinct modelling structures, which describe the behaviour of the conditional covariances in the different manners. However, the paths of estimated time-varying correlations indeed respect the analogous trends; consult Figure 9.5. It is also interesting that there are visible peaks in 2007 in the correlations between CZK and PLN, CZK and HUF, respectively. For a short time segment, the logarithmic returns of these currencies were negatively correlated. It might be likely justified by the mortgage and consequent financial crisis.

Finally, Figure 9.6 draws all mutual correlations between the six non-Euro EU27 currencies (as before, only the two models are considered). Apparently, one can identify similar trends, i.e. both methods have delivered analogous estimates (roughly speaking). It is curious that there are peaks (or some changes) in (after) 2007 only in some graphs. Some correlations seem to be undisturbed by the crisis. This might be further studied from the macroeconomic viewpoint.

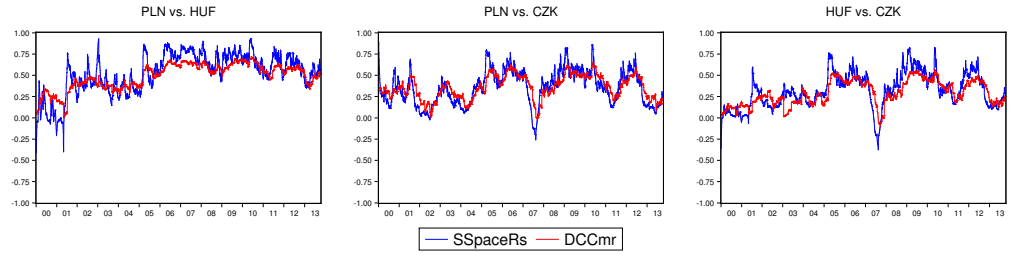


Figure 9.5: Estimated conditional correlations (the non-Euro Visegrad currencies).

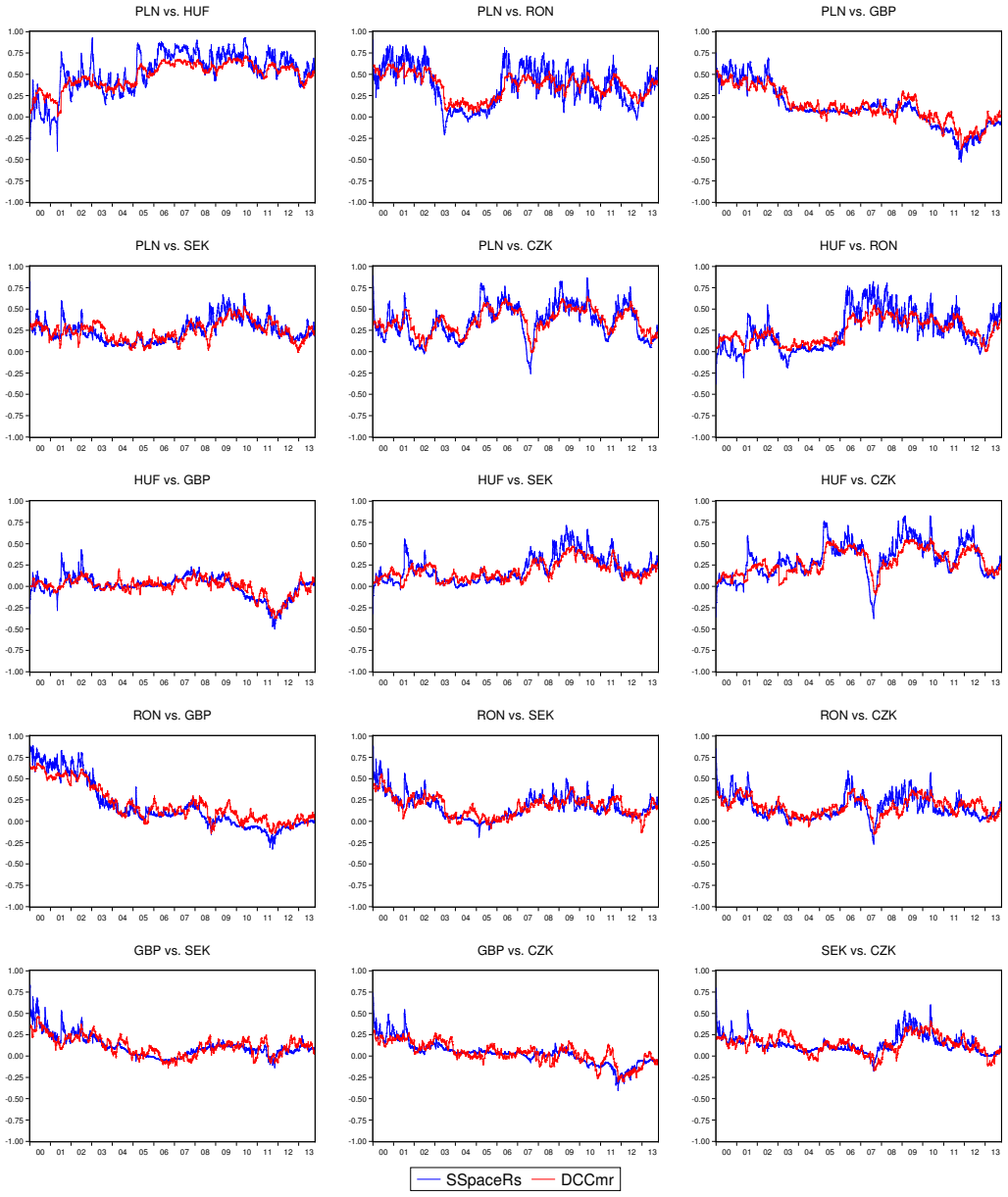


Figure 9.6: Estimated conditional correlations (the non-Euro EU27 currencies).

Conclusion

The present dissertation thesis elaborated on two different issues from the sphere of conditional covariance modelling, which is a substantial subdiscipline of financial time series analysis. Both the univariate and multivariate modelling frameworks have been studied from the different perspectives. This work, in fact, summarized and recapitulated the previously presented pieces of research, whose outputs have been accepted for publication (Hendrych, 2013, 2014b; Hendrych & Cipra, 2014, 2015). Moreover, several other (unpublished) results have been introduced; namely, see Sections 5.2.5, 5.2.6, 5.4.1, 5.4.2, 5.4.3, and the associated parts of Chapter 6. These particular findings will be submitted shortly.

The text was divided into two separate parts corresponding to two significant problems of financial time series modelling. Part I discussed recursive estimation algorithms for selected classic conditional heteroscedasticity models, concretely for the ARCH, GARCH, RiskMetrics EWMA, and GJR-GARCH processes. In greater detail, the self-weighted one-stage sequential estimation schemes have been suggested and theoretically justified; we have addressed special attention to the practical aspects of implementation. This effort was primarily motivated by the two following facts. Firstly, the character of the previously considered on-line calibration techniques recapitulated in Chapter 4 has naturally encouraged their revising and extending. Secondly, there exists a lingering demand for reliable sequential estimators of various financial time series models since these are processed effectively with low memory usage. The suggested self-weighted one-stage recursive estimation procedures have been examined by Monte Carlo experiments and two real data examples. The proposed concepts have shown their numerical capabilities. They have also demonstrated that they might be applied in various empirical contexts. Further research may be focused on employing these methods in the detection of structural (model) changes or breaks; simultaneously, other financial datasets may be analysed.

Part II introduced a novel approach to conditional covariance and correlation modelling. It has been inspired by the key principle of the multivariate orthogonal GARCH method, i.e. by the idea of a linear time-invariant orthogonal transformation of observed multivariate time series. The proposed linear time-varying orthogonal transformation has been based on the LDL decomposition of the conditional covariance matrix. Subsequently, it has enabled to apply the CCC structure on the transformed series. This particular approach has been implemented and calibrated by using the nonlinear discrete-time state space modelling framework. It has demonstrated its capabilities compared with other common

methods utilized in the considered context. Namely, extensive Monte Carlo experiments and two real data examples have shown that the suggested modelling scheme is competitive amongst others. Further research may be focused on interconnecting all concepts discussed in this thesis. In particular, sequentially estimated multivariate conditional covariance (correlation) models might be truly useful. Recursive algorithms are usually relatively simple to operate; they are numerically effective. Therefore, such methods might be favourable in analysing and systematic processing substantially large portfolios of risky assets, where one must primarily prefer computationally (highly) efficient approaches.

Bibliography

Aielli, G. P. (2013). Dynamic conditional correlation: On properties and estimation. *Journal of Business & Economic Statistics*, *31*(3), 282–299. doi: 10.1080/07350015.2013.771027

Aknouche, A. & Guerbyenne, H. (2006). Recursive estimation of GARCH models. *Communications in Statistics - Simulation and Computation*, *35*(4), 925–938. doi: 10.1080/03610910600880328

Alexander, C. (2002). Principal component models for generating large GARCH covariance matrices. *Economic Notes*, *31*(2), 337–359. doi: 10.1111/1468-0300.00089

Alp, T. & Demetrescu, M. (2010). Joint forecasts of Dow Jones stocks under general multivariate loss function. *Computational Statistics & Data Analysis*, *54*(11), 2360–2371. doi: 10.1016/j.csda.2009.09.027

Anderson, B. & Moore, J. (1979). *Optimal filtering*. Englewood Cliffs, NJ: Prentice Hall.

Bauwens, L., Laurent, S. & Rombouts, J. (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, *21*(1), 79–109. doi: 10.1002/jae.842

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. doi: 10.1016/0304-4076(86)90063-1

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics*, *72*(3), 498–505. doi: 10.2307/2109358

Bose, A. & Mukherjee, K. (2003). Estimating the ARCH parameters by solving linear equations. *Journal of Time Series Analysis*, *24*(2), 127–136. doi: 10.1111/1467-9892.00296

Brockwell, P. J. & Davis, R. A. (1991). *Time series: Theory and methods*. New York, NY: Springer.

Brockwell, P. J. & Davis, R. A. (2002). *Introduction to time series and forecasting*. New York, NY: Springer.

Chui, C. & Chen, G. (2013). *Kalman filtering with real-time applications*. Berlin, Germany: Springer.

Creal, D., Koopman, S. & Lucas, A. (2008). *A general framework for observation driven time-varying parameter models* (Discussion Paper No. 08-108/4). Tinbergen Institute. Retrieved February 3, 2014, from `http://ssrn.com/abstract=1297183`

Durbin, J. & Koopman, S. J. (2001). *Time series analysis by state space methods*. New York, NY: Oxford University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987–1007. doi: 10.2307/1912773

Engle, R. F. (2002). Dynamic conditional correlation. *Journal of Business & Economic Statistics*, *20*(3), 339–350. doi: 10.1198/073500102288618487

Engle, R. F. (2009). *Anticipating correlations: A new paradigm for risk management*. Princeton, NJ: Princeton University Press.

Engle, R. F. & Colacito, R. (2006). Testing and valuing dynamic correlations for asset allocation. *Journal of Business & Economic Statistics*, *24*(2), 238–253. doi: 10.1198/073500106000000017

Engle, R. F. & Sheppard, K. (2001). *Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH* (Working Paper No. 8554). National Bureau of Economic Research. Retrieved May 17, 2013, from `http://www.nber.org/papers/w8554`

European Central Bank. (2013). *Exchange rates*. Retrieved October 5, 2013, from `http://sdw.ecb.europa.eu/browse.do?node=2018779`

EViews. (2013). Eviews 8 user's guide [Computer software manual]. Irvine, CA.

Fan, J. & Yao, Q. (2005). *Nonlinear time series: Nonparametric and parametric methods*. New York, NY: Springer.

Franke, J., Härdle, W. K. & Hafner, C. M. (2011). *Statistics of financial markets: An introduction*. Berlin, Germany: Springer.

Gerencsér, L. & Orlovits, Z. (2012). Real time estimation of stochastic volatility process. *Annals of Operations Research*, *200*(November 2012), 223–246. doi: 10.1007/s10479-011-0976-2

Gerencsér, L., Orlovits, Z. & Torma, B. (2010). Recursive estimation of GARCH processes. In A. Edelmayer (Ed.), *Proceedings of the 19th International Symposium on Mathematical Theory and Systems - MTNS* (pp. 2415–2422). Budapest, Hungary.

Glosten, L. R., Jagannathan, R. & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, *48*(5), 1779–1801. doi: 10.1111/j.1540-6261.1993.tb05128.x

Greene, W. H. (2003). *Econometric analysis*. New York, NY: Prentice Hall.

Hafner, C. M. & Reznikova, O. (2012). On the estimation of dynamic conditional correlation models. *Computational Statistics & Data Analysis*, *56*(11), 3533–3545. doi: 10.1016/j.csda.2010.09.022

Hanzák, T. & Cipra, T. (2011). Exponential smoothing for time series with outliers. *Kybernetika*, *47*(2), 165–178.

Harvey, A., Ruiz, E. & Santana, E. (1992). Unobserved component time series models with ARCH disturbances. *Journal of Econometrics*, *52*(1-2), 129–157. doi: 10.1016/0304-4076(92)90068-3

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*. Secaucus, NJ: Springer.

Hendrych, R. (2013). Another view on time-varying correlations: The case of stocks and bonds. In H. Vojáčková (Ed.), *Proceedings of the 31st International Conference Mathematical Methods in Economics* (pp. 249–254). Jihlava, Czech Republic: College of Polytechnics.

Hendrych, R. (2014a). On comparing various modelling schemes: The case of the Prague Stock Exchange index. In T. Löster & T. Pavelka (Eds.), *The 8th international days of statistics and economics: Conference proceedings* (pp. 456–465). Slaný, Czech Republic: Melandrium.

Hendrych, R. (2014b). Recursive estimators of GARCH models: Selected problems. In J. Talašová, J. Stoklasa & T. Talášek (Eds.), *Proceedings of the 32nd International Conference Mathematical Methods in Economics* (pp. 281–286). Olomouc, Czech Republic: Palacký University.

Hendrych, R. & Cipra, T. (2014). On conditional covariance modelling: An approach using state space models. *Computational Statistics & Data Analysis*. doi: 10.1016/j.csda.2014.09.019

Hendrych, R. & Cipra, T. (2015). Self-weighted recursive estimation of GARCH models. *Communications in Statistics - Simulation and Computation*. doi: 10.1080/03610918.2015.1053924

Kierkegaard, J., Nielsen, J., Jensen, L. & Madsen, H. (2000). *Estimating GARCH models using recursive methods.* Retrieved August 7, 2014, from `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.703`

Koopman, S. (1997). Exact initial Kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association*, *92*(440), 1630–1638. doi: 10.2307/2965434

Ljung, L. (1999). *System identification: Theory for the user*. Upper Saddle River, NJ: Prentice Hall.

Ljung, L. & Söderström, T. (1983). *Theory and practice of recursive identification*. Cambridge, MA: MIT Press.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. New York, NY: Springer.

Moore, J. & Weiss, H. (1979). Recursive prediction error methods for adaptive estimation. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(4), 197–205. doi: 10.1109/TSMC.1979.4310182

Morgan, J. P. (1996). *Riskmetrics - technical document* (4th ed.). New York, NY: Morgan Guaranty Trust Company.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1), 246–256. doi: 10.1016/j.jeconom .2010.03.034

Prague stock exchange. (2015). Retrieved May 3, 2015, from `http://pse.cz`

Romera, R. & Cipra, T. (1995). On practical implementation of robust Kalman filtering. *Communications in Statistics - Simulation and Computation*, *24*(2), 461–488. doi: 10.1080/03610919508813252

Rossi, E. & Spazzini, F. (2010). Model and distribution uncertainty in multivariate GARCH estimation: A Monte Carlo analysis. *Computational Statistics & Data Analysis*, *54*(11), 2786–2800. doi: 10.1016/j.csda.2009.06.004

Roth, M. & Gustafsson, F. (2011). An efficient implementation of the second order extended Kalman filter. In *Proceedings of the 14th International Conference on Information Fusion (FUSION)* (pp. 1913–1918). Chicago, IL: IEEE.

Söderström, T. & Stoica, P. (1989). *System identification.* New York, NY: Prentice Hall.

Trapletti, A. & Hornik, K. (2013). tseries: Time series analysis and computational finance [Computer software manual]. Retrieved September 7, 2014, from `http://CRAN.R-project.org/package=tseries`

Trešl, J. (2011). Srovnání vybraných metod predikce změn trendu indexu PX. *Politická ekonomie*, *2011*(2), 184–204.

Tsay, R. S. (2005). *Analysis of financial time series.* Hoboken, NJ: Wiley-Interscience.

Tse, Y. K. & Tsui, A. K. C. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics*, *20*(3), 351–362. doi: 10.1198/073500102288618496

Wiener borse. (2015). Retrieved May 3, 2015, from `http://wienerborse.at`

Yahoo! Finance. (2013). *Quotes for ^GSPC and ^TYX.* Retrieved October 5, 2013, from `http://finance.yahoo.com`

List of Abbreviations

AIC	Akaike information criterion
APARCH	Asymmetric power ARCH
ARCH	Autoregressive conditional heteroscedasticity
ARIMA	Autoregressive integrated moving average
ARMA	Autoregressive moving average
BEKK	Baba, Engle, Kraft, and Kroner
CCC	Constant conditional correlation
DCC	Dynamic conditional correlation
EGARCH	Exponential GARCH
EWMA	Exponentially weighted moving average
EWP	Equally weighted portfolio
GARCH	Generalized autoregressive conditional heteroscedasticity
GJR-GARCH	Glosten, Jagannathan, and Runkle - GARCH
IGARCH	Integrated GARCH
LM	Lagrange multiplier
MA	Moving average
MAE	Mean absolute error
MVP	Minimum variance portfolio
OGARCH	Multivariate orthogonal GARCH
SIC	Schwarz information criterion
SW-RPEM	Self-weighted (one-stage) recursive prediction error method
SW-RPLR	Self-weighted (one-stage) recursive pseudo-linear regression
2S-RPEM	Two-stage recursive prediction error method
2S-RPLR	Two-stage recursive pseudo-linear regression

List of Notations

$\mathbb{N}$, $\mathbb{N}_0$	Positive integer numbers, non-negative integer numbers
$\mathbb{Z}$	Integer numbers
$\mathbb{R}$	Real numbers
$\boldsymbol{A}$, $\boldsymbol{a}$	Matrix, column vector
$\boldsymbol{A}^\top$, $\boldsymbol{a}^\top$	Transpose of matrix, transpose of column vector
$\|\|\boldsymbol{A}\|\|$, $\|\|\boldsymbol{a}\|\|$	Matrix norm, vector norm
$\boldsymbol{A}^{-1}$	Inversion of matrix
$\det(\boldsymbol{A})$	Determinant of matrix
$\boldsymbol{A} > \boldsymbol{B}$	$\boldsymbol{A} - \boldsymbol{B}$ is positive definite matrix
$\boldsymbol{A} \geq \boldsymbol{B}$	$\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite matrix
$\boldsymbol{I}$, $\boldsymbol{1}$	Identity matrix, vector of units
vec	Vectorization operator (for general matrices)
vech	Vectorization operator (for symmetric matrices)
diag	Matrix operator extracting diagonal matrix
diagv	Vector operator creating diagonal matrix
$\boldsymbol{f}'$, ∇f	First-order derivative of $\boldsymbol{f}$, gradient of f
$\boldsymbol{f}''$	Hessian matrix of second-order derivatives of f
$N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$AsN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Asymptotic $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
$X \sim \mathcal{L}$	Random variable X has distribution $\mathcal{L}$
u_α	α-quantile of $N(0, 1)$
$x \approx y$	x approximately equals y
$x \equiv y$	x equals y
$x := y$	x is defined as y
$x \to y$	x tends to y
$\mathrm{mod}(x, y)$	Remainder after division of x by y
$\mathrm{Re}(x)$	Real part of complex number x
$1_{[A]}$	Indicator of event A
$\mathbb{E}(\cdot)$, $\mathbb{E}(\cdot\|\cdot)$	Expected value, conditional expected value
$\mathsf{var}(\cdot)$, $\mathsf{var}(\cdot\|\cdot)$	Variance, conditional variance
$\mathsf{cov}(\cdot)$, $\mathsf{cov}(\cdot\|\cdot)$	Covariance, conditional covariance
$\mathsf{corr}(\cdot)$, $\mathsf{corr}(\cdot\|\cdot)$	Correlation, conditional correlation
I_t^-	Indicator of negativity of time series y_t at time t
B	Backshift time series operator ($B^i y_t = y_{t-i}$, $i \in \mathbb{N}_0$)
$\mathcal{F}_t$	σ-algebra (information set) available at time t
t	Time variable (integer value)

List of Figures

List of Tables