# EXISTENCE, UNIQUENESS, STOCHASTIC STABILITY, AND ESTIMATION THEORY OF MULTIVARIATE GARCH MODELS

by

Elisha Rae Hughes

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Mathematics

The University of Utah

August 2011

The University of Utah Graduate School

## STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Elisha Rae Hughes**

has been approved by the following supervisory committee members:

| | | |
|---|---|---|
| **Lajos Horvath** | , Chair | **3/30/2011**<br>Date Approved |
| **Davar Khoshnevisan** | , Member | **4/14/2011**<br>Date Approved |
| **Stewart Ethier** | , Member | **4/27/2011**<br>Date Approved |
| **Nat Smale** | , Member | **4/27/2011**<br>Date Approved |
| **David Kiefer** | , Member | **3/15/2011**<br>Date Approved |

and by **Aaron Bertram** , Chair of

the Department of **Mathematics**

and by Charles A. Wight, Dean of The Graduate School.

# ABSTRACT

Recent economic crises have exposed a critical need for appropriate methods to measure, model, and predict financial volatility. Generalized autoregressive conditional heteroskedastic (GARCH) models have been among the most successful and widely studied tools for this task due to their ability to capture the stylized characteristics of financial data.

Extending the original univariate GARCH processes to the multivariate framework is important because, in many applications, the primary quantity of interest is the interdependence, or covariance, between different univariate processes. Covariances are used for calculations of hedge ratios, betas of CAPM (Capital Asset Pricing Model), portfolio VaR (Value at Risk) estimates, asset weights in portfolios, and to investigate contagion across financial markets.

In Chapter 1 of this dissertation, we briefly review concepts and terminology related to stochastic processes and time series analysis. In Chapter 2, we prove sufficient conditions for existence, uniqueness, and stochastic stability of multivariate GARCH processes. In Chapter 3, we explore the QMLE and VTE methods for estimating multivariate GARCH parameters. We prove sufficient conditions for strong consistency and asymptotic normality of the QMLE and VTE estimators, and we conduct simulation studies to compare the performance of the VTE and QMLE.

For Dino.

# CONTENTS

# CHAPTER 1

# INTRODUCTION AND PRELIMINARIES

## 1.1 Introduction

In this section, we briefly discuss the importance, and background, of multivariate GARCH processes.

### 1.1.1 Motivation

Recent economic crises have exposed a critical need for appropriate methods to measure, model, and predict financial volatility. Generalized autoregressive conditional heteroskedastic (GARCH) models, pioneered by Engle [20] in 1982 and generalized by Bollerslev [9] in 1986, have been among the most successful and widely studied tools for this task due to their ability to capture the stylized characteristics of financial data.

Many of the stylized characteristics of financial data were first put forward in a 1963 paper by Mandelbrot [37], and have subsequently been documented with empirical studies. Some of the most widely observed characteristics include volatility clustering, volatility mean reversion, leptokurtosis (fat tails), and the leverage effect (asymmetry) (see [14], [41]). These stylized characteristics have motivated researchers to abandon the constant variance and normality assumptions imposed by classical econometric models, in favor of the more flexible and general GARCH models.

Extending the original univariate GARCH processes of Engle [20] and Bollerslev [9] to the multivariate framework is important because, in many applications, the primary quantity of interest is the interdependence, or covariance, between different univariate processes. Covariances are used for calculations of hedge ratios, betas of CAPM (Capital Asset Pricing Model), portfolio VaR (Value at Risk) estimates, asset weights in portfolios, and to investigate contagion across financial markets (see [48], [51], [16]).

### 1.1.2 Existence of GARCH

In the univariate case, existence and other statistical and probabilistic properties of GARCH processes are well-established. These results can be found, for instance, in the 2004 paper by Berkes, Horváth, and Kokoszka [4], and in the book by Francq and Zakoian [27].

Although GARCH processes received considerable attention since their introduction by Engle [20] in 1982, the GARCH specification entails a complex probabilistic structure, and existence of univariate GARCH was not established until 1992 when Bougerol and Picard [11] published necessary and sufficient conditions. A brief description of the technique used by Bougerol and Picard [11] is given in Chapter 2 of this dissertation.

Properties of multivariate GARCH processes are only partially known. In the existing literature, sufficient but not necessary conditions for the existence of weakly stationary, strictly stationary, and ergodic solutions have been established in some special cases. For the general BEKK GARCH process, Boussama [12] made waves among GARCH researchers by claiming that he could prove sufficient conditions for existence using Markovian methods combined with recent results in algebraic geometry, primarily those of Mokkadem [39]. He provided only a brief sketch of a proof, and many were skeptical of his claim. In Chapter 2 of this dissertation, we briefly outline the technique suggested by Boussama [12], and we provide a detailed proof that BEKK GARCH processes exist. We use techniques similar to those suggested by Boussama [12], but without the elaborate tools of algebraic geometry.

### 1.1.3 Estimation of GARCH

In the univariate case, many techniques for estimation of GARCH processes have appeared in the literature. Francq and Zakoïan [26] survey the existing univariate GARCH parameter estimation methods and their asymptotic properties. Some of the more popular methods have included least squares estimators, least absolute deviation estimators and $L^p$ estimators. However, estimation by Gaussian quasi-maximum likelihood (QMLE) is perhaps the most popular because it is robust to the distribution of the underlying process, and it is consistent and asymptotically normal without imposing moment conditions on the observed process.

In the multivariate case, parameter estimation research has focused primarily on Gaussian quasi-maximum likelihood estimation (QMLE). Consistency and asymptotic normality of the QMLE were established for models admitting a BEKK representation by Comte and Liebermann [17] under the assumption of independent coordinates for the innovations, and a moment of order eight for the process. Recently, Hafner and Preminger [32] established asymptotic normality of the QMLE under the weaker assumption of a sixth order moment for the observed process.

Despite favorable asymptotic properties, estimation of multivariate GARCH parameters by QMLE is problematic. In practice, QMLE is computationally intense due to the highly nonlinear form of the log-likelihood function, and the large number of parameters which must be estimated in the multivariate framework.

In Chapter 3 of this dissertation, we prove asymptotic normality of the QMLE for the BEKK GARCH representation assuming only a fourth order moment for the process, and we investigate a new variance targeting estimation (VTE) method that reduces the computational intensity of estimation without sacrificing model parameters.

## 1.2 Stochastic Processes

A **stochastic process** is a collection of random variables, $\{X_t : t \in T\}$, defined on some common probability space $(\Omega, \mathcal{F}, P)$, and indexed over some $T \subseteq \mathbb{R}$. Given a stochastic process $\{X_t : t \in T\}$, denote by $\mathcal{T}$ the colection of all vectors $(t_1, ..., t_n)' \in T^n$ such that $t_1 < t_2 < \cdots < t_n$ for $n \in \{1, 2, \ldots\}$. Then the (finite-dimensional) **distribution functions** of $\{X_t : t \in T\}$ are the functions $F_{\mathbf{t}}(\cdot)$, defined for all $\mathbf{t} = (t_1, ..., t_n)' \in \mathcal{T}$, and all $\mathbf{x} = (x_1, ..., x_n)' \in \mathbb{R}^n$, by

$$F_{\mathbf{t}}(\mathbf{x}) := P(X_{t_1} \leq x_1, ..., X_{t_n} \leq x_n).$$

Stochastic processes can be completely described by their distribution functions, but we sometimes limit our characterization to some collection of initial moments; of particular importance are the first and second moments, i.e., the **means** or expected values,

$$\mathbb{E}[X_t] := \mu_t,$$

and the **variances**,

$$\mathbb{V}[X_t] := \mathbb{E}[(X_t - \mu_t)^2] = \mathbb{E}[X_t^2] - \mu_t^2,$$

as well as the **covariances**,

$$\gamma(s,t) := \mathbb{C}\text{ov}(X_s, X_t) = \mathbb{E}[(X_s - \mu_s)(X_t - \mu_t)] = \mathbb{E}(X_s X_t) - \mu_s \mu_t.$$

The covariances of a stochastic process are often called **autocovariances** since they are covariances between random variables of the same stochastic process.

In the special case where all distribution functions of a process are multivariate normal, the process is completely characterized by its first and second moments. For nonnormal processes, the means and autocovariances do not give a complete characterization, but they do give some insight to the temporal dependence structure of the process.

### 1.2.1 Estimation of Autocovariances

The moments of a process are typically estimated from a realization of length $n$, that is to say $X_1, \ldots, X_n$. We estimate the autocovariance function $\gamma(h)$ with the **empirical autocovariance function**, defined for $0 \leq h < n$ by

$$\hat{\gamma}(h) := \frac{1}{n} \sum_{j=1}^{n-h} (X_j - \bar{X})(X_{j+h} - \bar{X}) = \hat{\gamma}(-h),$$

where

$$\bar{X} := \frac{1}{n} \sum_{j=1}^{n} X_j$$

is the empirical mean. Analogously, we define the **empirical autocorrelation function** by

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

for $|h| < n$. The empirical autocovariance estimator is biased, but asymptotically unbiased. It can be made unbiased by replacing $n$ by $n - h$ in the denominator of the estimator, but $\hat{\gamma}(h)$ has the desirable property that the covariance matrix with entry $(i,j) = \hat{\gamma}(i-j)$ is positive semidefinite, where the unbiased counterpart may not be.

## 1.2.2 Stochastic Stability

### 1.2.2.1 Stationarity

A stochastic process $\{X_t : t \in T\}$ is called **mean stationary** if

$$\mathbb{E}(X_t) = \mu_t = \mu$$

is constant and finite for all $t$. It is called **variance stationary** if

$$\mathbb{V}(X_t) = \mathbb{E}[(X_t - \mu_t)^2] = \sigma^2$$

is constant and finite for all $t$, and it is called **covariance stationary** if

$$\mathbb{C}\mathrm{ov}(X_t, X_s) = \gamma(|t - s|)$$

is a function only of the distance between the two random variables. If $\{X_t : t \in T\}$ is both mean stationary, and covariance stationary, then we say that it is **weakly stationary**. In this case we frequently drop the adjective weak, and refer to weak stationarity simply as **stationarity**.

We say that $\{X_t : t = 0, \pm 1, \ldots\}$ is **strictly stationary** if the joint distributions of $(X_{t_1}, \ldots, X_{t_k})'$, and $(X_{t_{1+h}}, \ldots, X_{t_{k+h}})'$ are the same for all integers $t$ and $h$, and all nonnegative integers $k$, i.e., if

$$(X_{t_1}, \ldots, X_{t_k})' \stackrel{d}{=} (X_{t_{1+h}}, \ldots, X_{t_{k+h}})'.$$

Strict stationarity immediately implies that each $X_t$ comes from the same distribution, so if $\mathbb{E}(X_t)$ and $\mathbb{V}(X_t)$ exist, then strict stationarity implies weak stationarity. The converse is not generally true, but if the distributions of a weakly stationary process are multivariate normal, then since the multivariate normal distribution is completely specified by its first and second moments, it is also strictly stationary.

### 1.2.2.2 Ergodicity

Much attention has been devoted to characterizing the dependence between terms of stochastic processes, i.e., the dependence structure of stochastic processes. Do the past states of a process influence its future states? What about the very distant past? The elementary tools of autocovariance and autocorrelation, presented above, are appropriate measures of dependence for many stochastic processes, but when the dependence structure is nonlinear, as it is for GARCH, more sophisticated tools are necessary; to this end, we introduce the concepts of ergodicity and mixing.

The concept of ergodicity is much more general than its limited presentation in this dissertation, and can be extended to nonstationary sequences (see Billingsley [8]).

Many authors (for instance, Wang [49]) define ergodicity only in terms of strictly stationary processes, stating that a strictly stationary processes is **ergodic** if the sample moments calculated from only finitely many indices of a time series converge to the corresponding population moments.

A more precise definition, which can be found in Francq and Zakoïan [27], states that a strictly stationary stochastic process $\{X_t : t = 0, \pm 1, \ldots\}$ is **ergodic** if and only if, for any Borel set $B$ and any integer $k$,

$$\frac{1}{n}\sum_{t=1}^{n} \mathbb{I}_B(X_1, \ldots, X_k) \xrightarrow{a.s} \mathbb{P}\{(X_1, \ldots, X_k) \in B\}. \tag{1.1}$$

The following theorem is a powerful tool for proving results related to ergodicity; it states that measurable transformations of strictly stationary and ergodic processes are again strictly stationary and ergodic.

**Theorem 1** *If* $\{X_t : t = 0, \pm 1, \ldots\}$ *is a strictly stationary and ergodic sequence, and if* $\{Y_t : t = 0, \pm 1, \ldots\}$ *is defined by*

$$Y_t := f(\ldots, X_{t-1}, X_t, X_{t+1}, \ldots),$$

*where* $f$ *is a measurable function from* $\mathbb{R}^\infty$ *into* $\mathbb{R}$*, then* $\{Y_t : t = 0, \pm 1, \ldots\}$ *is also strictly stationary and ergodic.*

**Proof** : See Billingsley [8], Theorem 36.4.

□

#### 1.2.2.3 Mixing

The mixing properties of a stochastic process were introduced by Rosenblatt [44], and are used to characterize different ideas of asymptotic independence between the past and future of a process. We present here two of the most popular mixing coefficients.

The **$\alpha$-mixing coefficient** between two $\sigma$-fields $\mathcal{A}$ and $\mathcal{B}$ is defined by

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

If $\mathbf{Y} = \{Y_t : t = 0, \pm 1, \pm 2, \ldots\}$ is a strictly stationary stochastic process, then for each integer $k$, denote by $\mathcal{F}^k$ the "future" information set $\mathcal{F}^k := \sigma(Y_k, Y_{k+1}, \ldots)$, and denote by $\mathcal{F}_k$ the "past" information set $\mathcal{F}_k := \sigma(Y_k, Y_{k-1}, \ldots,)$. Then, the **$\alpha$-mixing coefficient** of $\mathbf{Y}$ is defined by

$$\alpha_k = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{F}_0, B \in \mathcal{F}^k\}.$$

The process $\mathbf{Y}$ is said to be **$\alpha$-mixing** if $\alpha_k \to 0$ as $k \to \infty$. If $\alpha_k$ tends to zero at an exponential rate, then $\mathbf{Y}$ is said to be **geometrically $\alpha$-mixing**.

The **$\beta$-mixing coefficient** of $\mathbf{Y}$ is defined by

$$\beta_k := \mathbb{E}\left[\sup\{|\mathbb{P}(B|\mathcal{F}_0) - \mathbb{P}(B)| : B \in \mathcal{F}^k\}\right].$$

The process $\mathbf{Y}$ is said to be **$\beta$-mixing** if $\beta_k \to 0$ as $k \to \infty$. If $\beta_k$ tends to zero at an exponential rate, then $\mathbf{Y}$ is said to be **geometrically $\beta$-mixing**. It is easy to see that

$$\alpha_k \leq \beta_k,$$

so that $\beta$-mixing implies $\alpha$-mixing.

#### 1.2.2.4 Martingales

If $\mathbf{Y} = \{Y_t : t = 0, 1, \ldots\}$ is a sequence of real-valued random variables, and $\mathcal{F} = \{\mathcal{F}_t : t = 0, \pm 1, \ldots\}$ is a filtration, then we say that $Y$ is a **martingale** with respect to $\mathcal{F}$ if, for all $t \in \{0, \pm 1, \ldots\}$, we have

(i) $Y_t$ is $\mathcal{F}_t$ measurable,

(ii) $\mathbb{E}|Y_t| < \infty$,

(iii) $\mathbb{E}(Y_{t+1}|\mathcal{F}_t) = Y_t$.

If we merely say that $Y$ is a martingale, then it is implicitly with respect to the filtration $\mathcal{F}_t = \sigma(Y_s : s \leq t)$.

If $N = \{N_t : t = 0, 1, \ldots\}$ is a sequence of real random variables, and $\mathcal{F} = \{\mathcal{F}_t : t = 0, 1, \ldots\}$ is a filtration, then we say that $N$ is a **martingale difference** with respect to $\mathcal{F}$ if, for all $t \in \{0, 1, \ldots\}$, we have

(i) $N_t$ is $\mathcal{F}_t$ measurable,

(ii) $\mathbb{E}|N_t| < \infty$,

(iii) $\mathbb{E}(N_{t+1}|\mathcal{F}_t) = 0$.

Again, if we merely say that $N$ is a martingale difference, then it is implicitly with respect to the filtration $\mathcal{F}_t = \sigma(N_s : s \leq t)$. Note that if $\{Y_t : t \in \mathbb{N}\}$ is a martingale, then setting $N_1 = Y_1$, and $N_t = Y_t - Y_{t-1}$ for $t > 1$ gives a martingale difference, hence the name. Alternatively, if $N = \{N_t : t \in \mathbb{N}\}$ is a martingale difference with respect to $\mathcal{F}$, then setting $Y_t = N_0 + \cdots + N_t$ ensures that $Y = \{Y_t : t \in \mathbb{N}\}$ is a martingale with respect to $\mathcal{F}$.

**Theorem 2** *(The Lindeberg CLT) Suppose that, for each $n \in \{1, 2, \ldots\}$, $\{N_{nk} : k = 1, 2, \ldots\}$ is a square integrable martingale difference with respect to $\{\mathcal{F}_{nk} : k = 1, 2, \ldots\}$. Let $\sigma_{nk}^2 = \mathbb{E}(N_{nk}^2|\mathcal{F}_{n(k-1)})$. Then if*

$$\sum_{k=1}^{n} \sigma_{nk}^2 \xrightarrow{p} \sigma^2$$

*as $n \to \infty$, where $\sigma^2$ is a stricly positive constant, and*

$$\sum_{k=1}^{n} \mathbb{E}(N_{nk}^2 \mathbb{I}_{\{|N_{nk}| \geq \varepsilon\}}) \to 0$$

*as $n \to \infty$ for every $\varepsilon > 0$, then*

$$\sum_{k=1}^{n} N_{nk} \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

*as $n \to \infty$.*

**Proof** : See Billingsley [8], Theorem 35.12.

□

The following corollary applies to GARCH models that can be represented as stationary and ergodic martingale differences.

**Corollary 1** *If $\{N_k : k = 1, 2, \ldots\}$ is a square integrable, stationary and ergodic martingale difference such that $\sigma^2 = \mathbb{V}(N_t) \neq 0$, then*

$$n^{-1/2} \sum_{k=1}^{n} N_k \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

*as $n \to \infty$.*

#### 1.2.2.5 Deterministic and Nondeterministic Processes

Throughout this dissertation, we will see important relationships between stochastic (nondeterministic) processes, and their deterministic (nonrandom) counterparts. A process $\{X_t : t = 0, \pm 1, \ldots\}$ is said to be **deterministic** if, for each $j \in \{1, 2, \ldots\}$, and each $n \in \{0, \pm 1, \ldots\}$, $X_{n+j}$ can be exactly predicted as a function of elements of $\mathcal{M}_n = \overline{\text{span}}\{X_t : -\infty < t \leq n\}$. If $\sigma^2$ is the one-step mean squared error $\sigma^2 = \mathbb{E}|X_{n+1} - P_{\mathcal{M}_n} X_{n+1}|^2$, and $\mathcal{M}_{-\infty}$ is the closed linear subspace $\mathcal{M}_{-\infty} = \bigcap_{n=-\infty}^{\infty} \mathcal{M}_n$, then it follows that the process $\{X_t : t = 0, \pm 1, \ldots\}$ is deterministic if and only if $\sigma^2 = 0$, or equivalently if and only if $X_t \in \mathcal{M}_{-\infty}$ for each $t$.

## 1.3 Time Series

An important class of stochastic processes - random processes that evolve over time - are referred to as time series. In this section, we introduce the predominant process models that are used in time series analysis; in particular, we define GARCH processes, and the closely related ARMA processes. The importance of these types of processes can be seen from a fundamental result that is due to Wold [50], which can be summarized as follows: *any mean zero, weakly stationary and nondeterministic process admits a finite moving average* (MA) *representation.*

It follows that the set of all finite order moving average (MA) processes is dense in the set of nondeterministic and weakly stationary stochastic processes. However, we often require many parameters in the MA model to obtain a good approximation. For this reason, ARMA models were developed, and have been shown to extend the MA models in such a way as to provide good fit with greater parsimony.

### 1.3.1 Univariate Time Series Models

#### 1.3.1.1 The Autoregressive Model (AR)

The autoregressive model of order $p$, denoted AR($p$), is a special type of time series where each observation, $X_t$, can be expressed as a linear function of finitely many past observations plus some random element $\eta_t$. More formally, we say that $\{X_t : t = 0, \pm 1, \ldots\}$ is an AR($p$) process if, for each $t \in \{0, \pm 1, \ldots\}$,

$$X_t = \eta_t + \varphi_1 X_{t-1} + ... + \varphi_p X_{t-p},$$

where $\varphi_1, \varphi_2, ...\varphi_p$ are the parameters of the model and $\{\eta_t : t = 0, \pm 1, \ldots\}$ is a sequence of independent and identically distributed random variables having mean zero and unit variance. The random variable $\eta_t$ is often termed the random shock at time $t$. We say that a random process is strong AR if we require that the random shocks at each point are independent standard normal random variables, and we say that a random process is weak AR we require only that the shocks form a white noise sequence - a sequence of uncorrelated random variables with mean zero and common variance. Note that the autoregressive model is simply a regression of the current value on past values of the series.

#### 1.3.1.2 Moving Average (MA)

The notation MA($q$) refers to the moving average model of order $q$. We say that $\{X_t : t = 0, \pm 1, \ldots\}$ is MA($q$) if, for each $t \in \{0, \pm 1, \ldots\}$,

$$X_t = \eta_t + \theta_1 \eta_{t-1} + ... + \theta_q \eta_{t-q},$$

where $\theta_1, \theta_2, ...\theta_q$ are the parameters of the model and the sequence of random shocks, $\{\eta_t : t = 0, \pm 1, \ldots\}$, consists of independent and identically distributed random variables having mean zero and unit variance. The distinction in the MA model is that these random shocks are propagated to future values of the time series.

#### 1.3.1.3 Autoregressive Moving Average (ARMA)

The notation ARMA($p, q$) refers to the model with $p$ autoregressive terms and $q$ moving average terms. A generalization of the AR($p$) and MA($q$) models, the ARMA model is appropriate when a time series is a function of its own history (the AR part), as well as a series of unobserved shocks (the MA part). We say that $\{X_t\}$ is ARMA$(p, q)$ if, for each $t \in \{0, \pm 1, \ldots\}$,

$$X_t = \eta_t + b_1 X_{t-1} + ... + b_p X_{t-p} + a_1 \eta_{t-1} + ... + a_q \eta_{t-q}.$$

Using the lag operator, $L$, we can write the ARMA$(p, q)$ model more compactly as

$$b(L)X_t = a(L)\eta_t$$

where $b(\cdot)$ and $a(\cdot)$ are polynomials given by

$$b(z) = 1 - b_1 z - \cdots - b_p z^p, \;\; a(z) = 1 + a_1 z + \cdots + a_q z^q.$$

For ARMA models defined as above, most authors assume that the polynomials $b(\cdot)$ and $a(\cdot)$ have no common factors, since otherwise we could define an equivalent process with orders smaller than $(p, q)$ by reducing $b(\cdot)$ and $a(\cdot)$.

The ARMA$(p, q)$ process given by $b(L)X_t = a(L)\eta_t$ is stationary if $b(z) \neq 0$ for all $|z| \leq 1$.

ARMA models are also called Box-Jenkins models due to the iterative, three-stage Box-Jenkins method for finding the best fit model of this form to a given dataset.

#### 1.3.1.4 Autoregressive Conditional Heteroskedasticity (ARCH)

Many time series exhibit changes in variance over time. In particular, stock prices, exchange rates, and other financial phenomena tend to be serially correlated, with periods of volatility appearing in clusters. The ARCH model was developed by Engle [20] to model the variance of forecast errors of heteroskedastic time series - often a sequence of log returns on a stock or asset. We denote observations of these types of time series by $\varepsilon_t$, and we say that $\{\varepsilon_t : t = 0, \pm 1, \ldots\}$ follows an ARCH$(p)$ model if, for each $t \in \{0, \pm 1, \ldots\}$,

(i) $\varepsilon_t = \sigma_t \eta_t$,

(ii) $\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2$,

where $\sigma_t^2$ is the conditional variance of $\varepsilon_t$ given $\mathcal{F}_{t-1}$, $\{\eta_t : t = 0, \pm 1, \ldots\}$ is a sequence of independent, identically distributed random variables with mean zero and unit variance, $\omega > 0$, and $\alpha_i \geq 0$ for $i = 1, .., p$.

The ARCH model is capable of generating sequences with volatility clustering and outliers similar to those observed in financial time series. However, to capture the dynamics of financial time series, ARCH processes of restrictively high orders are often necessary. For greater parsimony, Bollerslev [9] proposed to extend the ARCH model in a manner analogous to that in which the ARMA model extends the AR model; this is the GARCH model described below.

#### 1.3.1.5 Generalized ARCH (GARCH)

The GARCH model is an extension of the ARCH model due to Bollerslev [9]. We say that $\{\varepsilon_t : t = 0, \pm 1, \ldots\}$ follows a GARCH$(p, q)$ model if, for each $t \in \{0, \pm 1, \ldots\}$,

(i) $\varepsilon_t = \sigma_t \eta_t$,

(ii) $\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2$,

where $\omega > 0$, $\alpha_i > 0$ for $i = 1, .., p$ and $\beta_i > 0$ for $i = 1, .., q$.

### 1.3.2 Multivariate Time Series

When interdependence is observed between different univariate time series, it is useful to consider them as components of a vector-valued, multivariate time series. The univariate ARMA model extends naturally to the multivariate VARMA (vector ARMA) model, and the subclass of VAR (vector AR) models have been particularly popular in the econometric literature. This extension has raised new problems and new lines of research including cointegration.

#### 1.3.2.1 Multivariate GARCH

In contrast to ARMA models, the GARCH models do not extend so easily to the multivariate framework. Analogous to the univariate case, we may consider a conditionally heteroskedastic time series $\{\varepsilon_t : t = 0, \pm 1, \ldots\}$ , of dimension $d \times 1$, such that, for each $t \in \{0, \pm 1, \ldots\}$,

$$\varepsilon_t = H_t^{1/2} \eta_t \tag{1.2}$$

where the sequence $\{\eta_t : t = 0, \pm 1, \pm 2, \ldots\}$ consists of independent and identically distributed (i.i.d.) $\mathbb{R}^d$-valued random variables with mean zero and unit covariance.

However, the specification of the multivariate GARCH conditional covariance matrix, $H_t$, is problematic and has been the subject of extensive research. One problem is that the conditional expectation of a vector of dimension $d$ is also a vector of dimension $d$, but the conditional variance is an $d \times d$ matrix, so the number of parameters that must be estimated explodes as the dimension of the process increases. Another difficulty is that any valid model must ensure that the conditional covariance matrix is symmetric and positive definite. Further problems arise if the matrix representation is not unique. Summaries of the existing model specifications, their properties, and limitations can be found in Terasvirta and Silvennoinen [45].

We outline below some of the most popular specifications for the multivariate GARCH conditional covariance matrix.

**1.3.2.1.1 The vech GARCH model.** The vech GARCH model, due to Bollerslev [10], is perhaps the most natural multivariate extension of the univariate GARCH model. This representation makes use of the vech$(\cdot)$ operator, which stacks the columns in the lower triangular part of a square matrix; if $M$ is a square $d \times d$ matrix, then vech$(M)$ is a $d(d+1)/2$ vector, and if $M$ is symmetric, then $M$ can be recovered from vech$(M)$. In the vech representation, the conditional covariance matrix is given, for each $t \in \{0, \pm 1, \ldots\}$, by

$$h_t := \mathcal{C} + \sum_{i=1}^{q} A_i s_{t-i} + \sum_{j=1}^{p} B_j h_{t-j}, \tag{1.3}$$

where $s_t = \text{vech}(\varepsilon_t \varepsilon_t')$, $h_t = \text{vech}(H_t)$, $\mathcal{C}_0 = \text{vech}(C_0)$ for some positive definite $d \times d$ matrix $C_0$, and the coefficients $A_i$ and $B_j$ are positive definite $m \times m$ matrices for $m = d(d+1)/2$. The sequence $\{\eta_t : t = 0, \pm 1, \pm 2, \ldots\}$ consists of independent and identically distributed (i.i.d.) $\mathbb{R}^d$-valued random variables with mean zero and unit covariance.

The vech GARCH model has the advantage of being very flexible and general, but it has the disadvantage that estimation of the parameters is computationally intense. The number of parameters that must be estimated in the model above is $(p+q)[d(d+1)/2]^2 + d(d+1)/2$. A further disadvantage is that this representation does not ensure that $H_t$ will be positive definite for all $t$.

A simplified diagonal vech model, due to Bollerslev [10], assumes that the matrices $A_i$ and $B_j$ are diagonal. In this case, conditions exist for $H_t$ to be positive definite, and the number of parameters that must be estimated is reduced to $(p+q)[d(d+1)/2]$. However, this model is considered too restrictive for most applications since no interaction is allowed between the conditional variances and covariances.

**1.3.2.1.2 The BEKK GARCH model.** The BEKK GARCH model, named after Baba, Engle, Kraft and Kroner in a preliminary version of Engle and Kroner [22], is a restricted version of the vech GARCH model that takes the form

$$H_t := C + \sum_{i=1}^{q} \sum_{k=1}^{\ell_i} \hat{A}_{ik} \varepsilon_{t-i} \varepsilon'_{t-i} \hat{A}'_{ik} + \sum_{j=1}^{p} \sum_{r=1}^{s_j} \hat{B}_{jr} H_{n-j} \hat{B}'_{jr}, \tag{1.4}$$

where the coefficients $A_{ik}$, and $B_{jr}$ are $d \times d$ parameter matrices, and $C$ is a positive definite $d \times d$ matrix.

The BEKK model has the advantage that the conditional covariance matrices, $H_t$, are positive definite by construction. However, problems arise with estimation and identification. Estimation of the BEKK GARCH model is, like the vech GARCH model, computationally intensive due to necessary matrix inversions, and $(p+q)kd^2 + d(d+1)/2$ parameters. For $d > 1$, additional restrictions must be imposed on the coefficient matrices to ensure uniqueness of the parameterization.

A diagonal BEKK model has been proposed but, like the diagonal vech model, it is considered too restrictive for most applications.

**Remark 1** *Engle and Kroner [22] outline sufficient conditions for equivalence of the* vech *and* BEKK GARCH *models, and they note that if $H_t$ admits a* BEKK *representation, then $H_t$ also admits a* vech *representation. Stelzer [46] proves that the converse is not generally true; for $d = 1$ and $d = 2$ the* BEKK *and* vech *representations are equivalent, but for $d \geq 3$ there exist* vech *representations that cannot be written in the* BEKK *form. However, the* vech *representations that cannot be written in the* BEKK *form are necessarily degenerate in the sense that at least one of the parameter matrices maps the half-vectorized positive semidefinite matrices into a strict subset of themselves.*

**1.3.2.1.3 The factor GARCH models.** A number of factor GARCH models have been proposed with motivation in economic theory where returns on assets are assumed to be generated by a number of common components, or factors. The first factor GARCH model was introduced by Engle et al. [21], and can be thought of as an alternative parametrization of the BEKK GARCH model. This model assumes that $H_t$ is generated by $K$ (with $K < d$) not necessarily uncorrelated factors $f_{i,t}$ according to:

$$H_t = \Omega + \sum_{i=1}^{K} \omega_i \omega_i' f_{i,t},$$

where the $\omega_i$, for $i = 1, \dots K$, are linearly independent $d \times 1$ vectors of time invariant factor loadings, or weights, and $\Omega$ is a $d \times d$ positive semidefinite matrix. The factors $f_{i,t}$ are assumed to have a first-order GARCH structure.

**1.3.2.1.4 The CCC GARCH model.** In the CCC GARCH model, of Bollerslev [10], the time-varying conditional covariances are parametrized to be proportional to the product of the corresponding conditional standard deviations. More precisely, this model assumes that the conditional covariance matrix $H_t$ is given by

(i) $H_t = D_t R D_t$,

(ii) $h_t := \omega + \sum_{i=1}^{q} \tilde{A}_i s_{t-i} + \sum_{j=1}^{p} \tilde{B}_j h_{t-j}$,

where $R$ is a correlation matrix, $\omega$ is a $d \times 1$ vector with positive coefficients, $s_t = \mathrm{vech}(\varepsilon_t \varepsilon_t')$, and the coefficients $\tilde{A}_i$ and $\tilde{B}_j$ are $m \times m$ matrices with nonnegative coefficients where $m = d(d+1)/2$.

The conditional covariances are generally nonlinear functions of the components of $s_{t-i}$, and of past values of the components of $H_t$. Thus, the CCC GARCH model is not a restriction of the vech GARCH model, except when $R$ is the identity matrix.

An advantage of the CCC GARCH specification is that positive coefficients for the matrices $\tilde{A}_i$ and $\tilde{B}_j$, and a positive definite choice for $R$, ensure that $H_t$ is positive definite. However, the assumption of constant correlations is arbitrary, and it is not clear whether this assumption is supported by financial data (see Brooks [14]).

# CHAPTER 2

# EXISTENCE, UNIQUENESS, AND STOCHASTIC STABILITY

## 2.1 Introduction

GARCH models have been extremely popular since their introduction by Engle [20] in 1982, and generalization by Bollerslev [9] in 1986, but proving that such processes exist has been a great challenge. Even in the case of univariate GARCH, a proof that such processes exist waited until 1991 when Bougerol and Picard [11] published necessary and sufficient conditions.

The idea behind the 1991 proof of Bougerol and Picard [11] is as follows. GARCH processes may be viewed as special types of Markov processes, i.e., we can group together terms of a GARCH process to create a new process, $\{X_t : t = 0, 1, \ldots\}$, such that the state of the process at time $t$ is conditionally independent of the history of the process before time $s$, given the state of the process at time $s$, for any $s < t$. If the Markov process has a stationary solution, then it can be extended to a two-sided process; this proves existence of the associated GARCH process.

If a Markov process can be written in the form

$$X_{t+1} = F(X_t, \eta_{t+1}),$$

where the sequence $\{\eta_t : t = 0, 1, \ldots\}$ is independent and identically distributed with mean zero and unit variance, and if the Lipschitz property

$$||F(x, \eta_t) - F(y, \eta_t)|| \leq \alpha(\eta_t)||x - y||$$

holds for all possible states $x$ and $y$ of the process, and for some positive function $\alpha$ with $\mathbb{E}(\alpha(\eta_t)^m) < 1$ and $\mathbb{E}(||F(0, \eta_t)||^m) < \infty$ for some real number $m \geq 1$, then the Markov process has a strictly stationary solution.

Bougerol and Picard [11] used this method to show that univariate GARCH processes exist if and only if $\gamma < 0$, where $\gamma$ is the top Lyapunov exponent

$$\gamma = \inf_{0\leq n<\infty} \frac{1}{n+1}\mathbb{E}\log||A_0A_1\cdots A_n||,$$

and $A_t$ is a matrix composed of the coefficients of the process, and the random variable $\eta_t$. However, this method fails in the multivariate case.

Proofs of necessary and sufficient conditions for existence of multivariate GARCH processes do not currently exist in the literature. Engle and Kroner [22] claimed to prove necessary and sufficient conditions for existence of general multivariate GARCH processes, but close inspection reveals that their proof is not correct as it presupposes finite variance for the process. However, some authors (for instance [34], p565) continue to cite their result. For some of the simplest multivariate GARCH specifications, sufficient conditions can be shown using Markovian methods and real analysis. For example, Francq and Zakoïan [26] give a detailed proof of sufficient conditions for strict stationarity of the CCC GARCH model specification due to Bollerslev [10].

Boussama [12] made waves in the econometric community when he published an announcement that, for general multivariate GARCH processes, sufficient conditions for strict stationarity follow from Markovian methods combined with recent results in algebraic geometry, mainly those of Mokkadem [39]. Boussama's [12] article was extremely brief and he provided only a sketch of his proof. Researchers doubted whether Boussama's claim was true; this is evident in the article by Terasvirta and Silvennöinen [45] where they summarize the existing body of knowledge surrounding multivariate GARCH, and they are explicit when mentioning published results that rely on the work of Boussama [12].

Boussama's work can be summarized as follows. A method, similar to that used by Bougerol and Picard [11] in the univariate case, for showing strict stationarity of an irreducible Markov process was developed by Meyn and Tweedie [38], and is based on the Foster-Lyapunov condition:

$$\mathbb{E}[V(X_t)|X_{t-1}=x] \leq \alpha V(x) + b\mathbb{I}_C(x).$$

Here $V \geq 1$ is a Lyapunov function, $0<\alpha<1$, $0<b<\infty$, and $C$ is a so-called small set on which $V$ is bounded.

The Foster-Lyapunov condition requires that the Markov process be irreducible and, in general, this cannot be shown for multivariate GARCH processes. These processes are, however, from Boussama [12], irreducible if we can consider the function $F$ as a composition of a regular map and a diffeomorphism between algebraic varieties, and if we can restrict the process to the Zariski closure of an orbit of the form

$$\bigcup_{k=0}^{\infty}\{F^k(T, u_1, \ldots, u_k) : u_1, \ldots, u_k \in E\},$$

where $T$ is an attracting point for the process, $E$ is the domain of positivity for the density of the random variables $\{\eta_t : t = 0, 1, \ldots\}$, and the function composition $F^k$ is defined by $F^k(x, y_1, \ldots, y_k) := F(F^{k-1}(x, y_1, \ldots, y_{k-1}), y_k)$.

The work presented in this chapter began with a verification, using algebraic geometry, of the claims of Boussama [12]. Consequently, many of the proofs in this chapter are similar to those of Boussama [12]. However, the detailed proofs provided in this chapter eliminate the need for the elaborate machinery of algebraic geometry; we use only probability (especially Markov theory) and basic real analysis.

## 2.2 Preliminaries and Notation

In this chapter, we establish existence, uniqueness, and stability properties of an $\mathbb{R}^d$-valued multivariate GARCH process $\boldsymbol{\varepsilon} := \{\varepsilon_t : t = 0, \pm1, \pm2, \ldots\}$. For each $t$,

$$\varepsilon_t := H_t^{1/2}\eta_t \tag{2.1}$$

where $H_t := \mathbb{E}[\varepsilon_t\varepsilon_t' | \mathcal{F}_{t-1}]$ is the conditional covariance matrix of $\varepsilon_t$ given the sigma algebra $\mathcal{F}_{t-1} := \sigma\{\varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$, and $\{\eta_t : t = 0, \pm1, \pm2, \ldots\}$ is an independent and identically distributed sequence of $\mathbb{R}^d$-valued random variables having mean-zero and unit covariance. We assume that each $H_t$ admits a BEKK representation (see section 1.3 of Chapter 1), given by

$$H_t := C + \sum_{i=1}^{q}\sum_{k=1}^{\ell_i}\hat{A}_{ik}\varepsilon_{t-i}\varepsilon_{t-i}'\hat{A}_{ik}' + \sum_{j=1}^{p}\sum_{r=1}^{s_j}\hat{B}_{jr}H_{n-j}\hat{B}_{jr}', \tag{2.2}$$

and we denote the corresponding vech representation of $H_t$ by

$$h_t := \mathcal{C} + \sum_{i=1}^{q}A_i s_{t-i} + \sum_{j=1}^{p}B_j h_{t-j}, \tag{2.3}$$

where $h_t := \text{vech}(H_t)$, $\mathcal{C} := \text{vech}(C)$, and $s_{t-i} := \text{vech}(\varepsilon_{t-i}\varepsilon_{t-i}')$.

To estabish existence, uniqueness, and stability properties of the process $\boldsymbol{\varepsilon}$, defined by (2.1)-(2.3), we make the following assumptions:

**A1** : The random variables $\eta_t$ admit a density that is nonzero in a neighborhood of the origin.

**A2** : The spectral radius of $\sum_{i=1}^{q} A_i + \sum_{j=1}^{p} B_j$ is less than one.

The aim of this chapter is to prove, in Theorem 7, that under assumptions **A1**-**A2** the process $\boldsymbol{\varepsilon}$ exists and is unique. To obtain results for the process $\boldsymbol{\varepsilon}$ we will analyze a Markov chain $\mathbf{X}$:=$\{X_t : t = 0, 1, 2, \ldots\}$ defined by

$$X_t := (h_t', \ldots, h_{t-p+1}', \varepsilon_t', \ldots, \varepsilon_{t-q+1}')'. \tag{2.4}$$

In section 2.3, we recall some important definitions and results from Markov theory. The results from section 2.3 show that $\boldsymbol{\varepsilon}$ exists if the Markov chain $\mathbf{X}$ defined by (2.4) is aperiodic (Definition 2), $\psi$-irreducible (Definition 1), and satisfies the Foster-Lyapunov drift criteria (Definition 12). Section 2.3.5 establishes some necessary notation and results from linear algebra. In section 2.3.6, we show that the Markov chain $\mathbf{X}$ has a representation of the form $X_t = F(X_{t-1}, \eta_t)$, and we prove some smoothness and invertibility properties of the function $F$. In section 2.3.7 we construct a state space on which $\mathbf{X}$ is $\psi$-irreducible and aperiodic, and we prove that it suffices to consider $\mathbf{X}$ restricted to this state space. Section 2.4 proves that $\mathbf{X}$ satisfies the Foster-Lyapunov drift criteria, and section 2.5 combines results from all previous sections to prove that $\boldsymbol{\varepsilon}$ exists and is unique. Furthermore, we show that $\boldsymbol{\varepsilon}$ is positive Harris recurrent and geometrically ergodic with a strictly stationary solution that is geometrically $\beta$-mixing.

## 2.3 Markov Theory

In this section, we review notation and properties of Markov chains which can be found, for instance, in the book by Meyn and Tweedie [38].

Let $\boldsymbol{\Phi} := \{\Phi_t : t = 0, 1, \ldots\}$ denote a Markov chain taking values in a continuous state space $\mathsf{S}$. Denote the Borel sigma algebra of $\mathsf{S}$ by $\mathcal{B}(\mathsf{S})$, and denote the transition probability kernel (sometimes called a Markov transition function) of $\boldsymbol{\Phi}$ by $\mathcal{P} : \mathsf{S} \times \mathcal{B}(\mathsf{S}) \to \mathbb{R}$. The kernel $\mathcal{P}$ is characterized by the following two properties:

(i) $\mathcal{P}(\cdot, A)$ is measurable and nonnegative for all $A \in \mathcal{B}(\mathsf{S})$,

(ii) $\mathcal{P}(x, \cdot)$ is a probability measure on $(\mathsf{S}, \mathcal{B}(\mathsf{S}))$ for all $x \in \mathsf{S}$.

The $n$-step transition probability kernel, $\mathcal{P}^n$, is defined inductively for nonnegative integers $n$, $x \in \mathsf{S}$, and $A \in \mathcal{B}(\mathsf{S})$ by

$$\mathcal{P}^0(x, A) = \mathbb{I}_x(A).$$

$$\mathcal{P}^n(x, A) = \int_{\mathsf{S}} \mathcal{P}(x, dy)\mathcal{P}^{n-1}(y, A),\ n \geq 1.$$

The kernel $\mathcal{P}^n$ operates on $\sigma$-finite measures $\mu$ on $(\mathsf{S}, \mathcal{B}(\mathsf{S}))$ from the right according to

$$\mu\mathcal{P}^n(A) = \int_{\mathsf{S}} \mu(dx)\mathcal{P}^n(x, A), \qquad A \in \mathcal{B}(\mathsf{S}),$$

and $\mathcal{P}^n$ operates on bounded measurable functions $f$ from the left according to

$$\mathcal{P}^n f(x) = \int_{\mathsf{S}} f(y)\mathcal{P}^n(x, dy), \qquad x \in \mathsf{S}.$$

Note that $\mathcal{P}^n f(x) = \mathbb{E}_x[f(\Phi_n)]$ where $\mathbb{E}_x[\cdot]$ denotes expectation conditional on the event $\{\Phi_0 = x\}$ for $x \in \mathsf{S}$.

### 2.3.1 Irreducibility, Aperiodicity, and Recurrence

The notions of irreducibility, aperiodicity, and recurrence are closely tied to the stability of a Markov chain.

**Definition 1** *A Markov chain $\mathbf{\Phi}$ is called* **$\boldsymbol{\psi}$-irreducible** *if there exists a nontrivial measure $\psi$ on $(\mathsf{S}, \mathcal{B}(\mathsf{S}))$ such that, for all $x \in \mathsf{S}$, and all $A \in \mathcal{B}(\mathsf{S})$, $\psi(A) > 0$ implies that there exists some positive integer $n$, possibly depending on both $A$ and $x$, such that $\mathcal{P}^n(x, A) > 0$.*

**Theorem 3** *Suppose that $\mathbf{\Phi}$ is a $\psi$-irreducible Markov chain on $(\mathsf{S}, \mathcal{B}(\mathsf{S}))$. Then there exists some positive integer $d$ and disjoint sets $D_1, \ldots, D_d \in \mathcal{B}(\mathsf{S})$ (a "d-cycle") such that, for each $i = 0, \ldots, d-1$ (mod $d$), we have*

(i) *for $x \in D_i$, $\mathcal{P}(x, D_{i+1}) = 1$,*

(ii) $\psi\left(\left[\bigcup_{i=1}^{d} D_i\right]^c\right) = 0.$

**Proof** : See Meyn and Tweedie [38], Theorem 5.4.4.

□

**Definition 2** *The largest integer, d, in Theorem 3 is called the* **period** *of* $\mathbf{\Phi}$*. When the period is 1, the chain is said to be* **aperiodic**

**Definition 3** *Suppose* $\mathbf{\Phi} = \{\Phi_t : t = 0, 1, \ldots\}$ *is a Markov chain with state space* $\mathsf{S}$*. Then* $A \in \mathcal{B}(\mathsf{S})$ *is called* **recurrent** *if, for all* $x \in A$*,*

$$\mathbb{E}_x \left[ \sum_{t=1}^{\infty} \mathbb{I}_{\{\Phi_t \in A\}} \right] = \infty.$$

**Definition 4** *If* $\mathbf{\Phi}$ *is a Markov chain with state space* $\mathsf{S}$*, then we say that* $A \in \mathcal{B}(\mathsf{S})$ *is* **Harris recurrent** *if, for every* $x \in A$*,*

$$\mathbb{P} \left[ \sum_{t=1}^{\infty} \mathbb{I}_{\{\Phi_t \in A\}} = \infty | \Phi_0 = x \right] = 1.$$

*The Markov chain* $\mathbf{\Phi}$ *is called Harris recurrent if it is* $\psi$*-irreducible, and if A is Harris recurrent for every* $A \in \mathcal{B}(\mathsf{S})$ *such that* $\psi(A) > 0$*.*

### 2.3.2 Positivity, Ergodicity and Mixing

**Definition 5** *Suppose* $\mathbf{\Phi}$ *is a Markov Chain with transition probability kernel P, and state space* $\mathsf{S}$*. We say that a* $\sigma$*-finite measure* $\pi$ *on* $(\mathsf{S}, \mathcal{B}(\mathsf{S}))$ *is* $\boldsymbol{\mathcal{P}}$ ***− invariant*** *if, for every* $A \in \mathcal{B}(\mathsf{S})$*,*

$$\pi(A) = \int_{\mathsf{S}} \pi(dx)\mathcal{P}(x, A).$$

**Definition 6** *Suppose a Markov chain* $\mathbf{\Phi}$ *is* $\psi$*-irreducible, and admits a* $\mathcal{P}$*-invariant measure* $\pi$*. Then* $\mathbf{\Phi}$ *is called a* **positive** *chain.*

**Remark:** If a Markov chain $\mathbf{\Phi} = \{\Phi_t : t = 0, 1, \ldots\}$ is positive, then $\mathbf{\Phi}$ with initial distribution $\pi$ satisfies

$$(\Phi_{t_1}, \ldots, \Phi_{t_n})' \stackrel{d}{=} (\Phi_{t_1+h}, \ldots, \Phi_{t_n+h})'$$

for all $n \in \{0, 1, \ldots\}$, and all $h, t_1, \ldots, t_n \in \{1, 2, \ldots\}$. In this case, $\mathbf{\Phi}$ has a strictly stationary solution, i.e., the one-tailed process $\mathbf{\Phi}$ can be extended to a strictly stationary, two-tailed process $\tilde{\mathbf{\Phi}} = \{\tilde{\Phi}_t : t = 0, \pm 1, \pm 2, \ldots\}$. Thus, the multivariate GARCH$(p, q)$ process $\boldsymbol{\varepsilon}$, defined by (2.1)-(2.3), exists if the Markov chain $\mathbf{X}$, defined by (2.4), is positive. Furthermore, if $\mathbf{X}$ is positive Harris recurrent (Definition 4) and geometrically ergodic (Definition 7) with a strictly stationary solution that is geometrically $\beta$-mixing (Definition 8), then the same is true of $\boldsymbol{\varepsilon}$.

**Definition 7** *A Markov chain $\mathbf{\Phi}$ with transition probability kernel $\mathcal{P}$ and state space $\mathsf{S}$ is called* **ergodic** *if $\mathbf{\Phi}$ is positive Harris recurrent with invariant probability measure $\pi$, and for all $x \in \mathsf{S}$,*

$$\lim_{n\to\infty} ||\mathcal{P}^n(x, \cdot) - \pi||_{var} = 0$$

*where $||\cdot||_{var}$ denotes the total variation norm, i.e., if $\mu$ is a measure on $(\mathsf{S}, \mathcal{B}(\mathsf{S}))$, then*

$$||\mu||_{var} := \sup_{A\in\mathcal{B}(\mathsf{S})} \mu(A) - \inf_{A\in\mathcal{B}(\mathsf{S})} \mu(A),$$

*or equivalently,*

$$||\mu||_{var} := \sup_{f:|f|\leq 1} |\mu(f)| = \sup_{f:|f|\leq 1} \left| \int_{\mathsf{S}} f(x)\mu(dx) \right|.$$

*If $\mathbf{\Phi}$ is ergodic and for all $x \in \mathsf{S}$ we have*

$$\lim_{n\to\infty} ||\mathcal{P}^n(x, \cdot) - \pi||_{var} = o(r^n)$$

*for some $0 < r < 1$ that is independent of $x$, then we say that $\mathbf{\Phi}$ is* **geometrically ergodic***.*

**Definition 8** *Suppose $\mathbf{Y} = \{Y_t : t = 0, \pm 1, \pm 2, \ldots\}$ is a strictly stationary stochastic process. For each integer $k$, denote by $\mathcal{F}^k$ the "future" information set $\mathcal{F}^k := \sigma(Y_k, Y_{k+1}, \ldots)$, and denote by $\mathcal{F}_k$ the "past" information set $\mathcal{F}_k := \sigma(Y_k, Y_{k-1}, \ldots, )$. Then,*

$$\beta_k := \mathbb{E}\left[\sup\{|P(B|\mathcal{F}_0) - P(B)| : B \in \mathcal{F}^k\}\right]$$

*is called the* **$\boldsymbol{\beta}$-mixing coefficient** *of $\mathbf{Y}$. If*

$$\lim_{k\to\infty} \beta_k = 0,$$

*then* $\mathbf{Y}$ *is called* **$\boldsymbol{\beta}$-mixing**. *If*

$$\lim_{k\to\infty} \beta_k = o(r^n)$$

*for some positive number* $r < 1$, *then we say that* $\mathbf{Y}$ *is* **geometrically $\boldsymbol{\beta}$-mixing**.

### 2.3.3 Small Sets, Petite Sets, and Feller Chains

**Definition 9** *Suppose* $\mathbf{\Phi}$ *is a Markov Chain with state space* $\mathsf{S}$ *and transition probability kernel* $\mathcal{P}$, *and suppose* $K \in \mathcal{B}(\mathsf{S})$. *If there exists a positive integer* $n$ *and a nontrivial measure* $v_n$ *on* $(\mathsf{S}, \mathcal{B}(\mathsf{S}))$ *such that, for all* $A \in \mathcal{B}(\mathsf{S})$ *and all* $x \in K$,

$$\mathcal{P}^n(x, A) \geq v_n(A),$$

*then* $K$ *is said to be a* **small** *set.*

**Definition 10** *Suppose* $\mathbf{\Phi}$ *is a Markov chain with state space* $\mathsf{S}$ *and transition probability kernel* $\mathcal{P}$. *A set* $K \in \mathcal{B}(\mathsf{S})$ *is is called* **petite** *if there exists a sequence* $a = \{a_0, a_1, \ldots\}$ *that sums to 1, and if there exists a nontrivial measure* $v_a$ *on* $\mathcal{B}(\mathsf{S})$ *satisfying*

$$\sum_{n=0}^{\infty} \mathcal{P}^n(x, A) a_n \geq v_a(A),$$

*for all* $x \in K$, *and for all* $A \in \mathcal{B}(\mathsf{S})$.

**Remark:** It is clear from the definitions above that small sets are petite. There are some special cases where small sets and petite sets coincide.

**Theorem 4** *If the Markov chain* $\mathbf{\Phi}$ *is* $\psi$*-irreducible and aperiodic then every petite set is small.*

**Proof** : See Meyn and Tweedie [38] Theorem 5.5.7.

□

**Definition 11** *If* $\mathcal{P}(\cdot, O)$ *is lower semicontinuous for any open set* $O \in \mathcal{B}(\mathsf{S})$, *then the Markov chain* $\mathbf{\Phi}$ *having transition probability kernel* $\mathcal{P}$ *is called a* **Feller** *chain. Equivalently, we say that* $\mathbf{\Phi}$ *has the Feller property.*

**Proposition 1** *Suppose that the Markov chain* $\mathbf{\Phi}$ *is* $\psi$*-irreducible. If* $\mathbf{\Phi}$ *has the Feller property, and the support of* $\psi$ *has nonempty interior, then every compact subset of* $\mathsf{S}$ *is petite.*

**Proof** : See Meyn and Tweedie [38], Proposition 6.2.8 (ii).

□

**Corollary 2** *Suppose that the Markov chain* $\mathbf{\Phi}$ *is* $\psi$*-irreducible and aperiodic. If* $\mathbf{\Phi}$ *has the Feller property, and the support of* $\psi$ *has nonempty interior, then every compact subset of* $\mathsf{S}$ *is small.*

### 2.3.4 The Foster-Lyapunov Drift Criteria

**Definition 12** *Suppose* $\mathbf{\Phi}$ *is a Markov Chain with transition probability kernel* $\mathcal{P}$*, and state space* $\mathsf{S}$*. Then* $\mathbf{\Phi}$ *satifies the* **Foster-Lyapunov drift criteria** *(or drift condition) if, for all* $x \in \mathsf{S}$*,*

$$\mathcal{P}V(x) \leq \alpha V(x) + b\mathbb{I}_K(x), \tag{2.5}$$

*where* $V \geq 1$ *is a so-called Lyapunov function that is finite on* $\mathsf{S}$*,* $\alpha$ *and* $b$ *are real numbers with* $0 < \alpha < 1$*,* $0 < b < \infty$ *, and* $K$ *is a small set in* $\mathsf{S}$ *on which* $V$ *is bounded. Note that any function* $V$ *that satisfies (2.5) is called a Lyapunov function.*

**Theorem 5** *Suppose* $\mathbf{\Phi}$ *is an aperiodic,* $\psi$*-irreducible Markov chain with state space* $\mathsf{S}$ *and transition probability kernel* $\mathcal{P}$*. If the Foster-Lyapunov drift criteria is satisfied, then* $\mathbf{\Phi}$ *is geometrically ergodic, positive Harris recurrent, and the strictly stationary solution* $\{\Phi_t : t = 0, \pm 1, \pm 2, \ldots\}$ *is geometrically* $\beta$*-mixing.*

**Proof** :

(Geometric ergodicity)

Theorem 19.1.3 of Meyn and Tweedie [38] proves that if $\mathbf{\Phi}$ is aperiodic, $\psi$-irreducible, and satisfies the Foster-Lyapunov drift condition, then $\mathbf{\Phi}$ is geometrically ergodic. Moreover, for all $x \in \mathsf{S}$ and for any positive integer $n$,

$$||\mathcal{P}^n(x, \cdot) - \pi||_{var} \leq Rr^n V(x) \tag{2.6}$$

for some constants $0 < R < \infty$, and $0 < r < 1$.

(Harris recurrence)

Theorem 14.2.2 (the Comparison Theorem) of Meyn and Tweedie [38], states that if $W$, $f$, and $s$ are nonnegative functions such that

$$\mathcal{P}W(x) \le W(x) - f(x) + s(x)$$

holds for all $x \in \mathsf{S}$, then for all $x \in \mathsf{S}$, and for any stopping time $\tau$,

$$\mathbb{E}_x\left[\sum_{t=0}^{\tau-1} f(\Phi_t)\right] \le W(x) + \mathbb{E}_x\left[\sum_{t=0}^{\tau-1} s(\Phi_t)\right].$$

We define nonnegative functions $W$ and $s$ on $\mathsf{S}$ by $W(x) := V(x) - 1$, and $s(x) := b\mathbb{I}_K(x)$, and we define $f$ to be the nonnegative constant function $f(x) := 1 - \alpha$. Then, since (2.5) holds, we have for all $x \in \mathsf{S}$

$$\begin{aligned} PW(x) &= PV(x) - 1 \\ &\le \alpha V(x) + b\mathbb{I}_K(x) - 1 \\ &= \alpha V(x) + s(x) - 1 \\ &= \alpha W(x) - f(x) + s(x) \\ &\le W(x) - f(x) + s(x). \end{aligned}$$

Thus, the Comparison Theorem holds for our particular choices of $W$, $f$, and $s$.

Let $\tau_K := \inf\{t \ge 1 : \Phi_t \in K\}$ denote the hitting time of the set $K$ from (2.5). Note that $\tau_K$ is a stopping time, and

$$\mathbb{E}_x[\tau_K] = \mathbb{E}_x\left[\sum_{t=0}^{\tau_K-1} 1\right] = \frac{1}{1-\alpha}\mathbb{E}_x\left[\sum_{t=0}^{\tau_K-1} f(\Phi_t)\right],$$

so the Comparison Theorem implies that, for all $x \in \mathsf{S}$,

$$\mathbb{E}_x[\tau_K] = \frac{1}{1-\alpha}\mathbb{E}_x\left[\sum_{t=0}^{\tau_K-1} f(\Phi_t)\right] \le \frac{1}{1-\alpha}\left(W(x) + \mathbb{E}_x\left[\sum_{t=0}^{\tau_K-1} s(\Phi_t)\right]\right)$$

$$\leq \frac{1}{1-\alpha}\left(V(x) + \underbrace{\mathbb{E}_x\left[\sum_{t=0}^{\tau_K - 1} s(\Phi_t)\right]}_{=b\mathbb{I}_K(x)}\right) < \infty.$$

A nonnegative random variable with finite expectation is finite almost surely, so for all $x \in \mathsf{S}$, $\mathbb{P}(\tau_K < \infty | \Phi_0 = x) = 1$.

Theorem 9.1.7 (ii) of Meyn and Tweedie [38] proves that if for all $x \in \mathsf{S}$, and for some petite set $K$ we have $\mathbb{P}(\tau_K < \infty | \Phi_0 = x) = 1$, then $\mathbf{\Phi}$ is Harris recurrent. From Definitions 9 and 10, small sets are petite, so this proves Harris recurrence of $\mathbf{\Phi}$.

(Positivity)

Harris recurrence implies recurrence and, from Theorem 10.0.1 of Meyn and Tweedie [38], if the chain $\mathbf{\Phi}$ is recurrent, then it admits a unique invariant probability measure $\pi$. Thus $\mathbf{\Phi}$ is positive.

(Geometric $\beta$-mixing)

Theorem 10.0.1 of Meyn and Tweedie [38] also shows that if $\sup_{x \in K} \mathbb{E}_x[\tau_K] < \infty$ holds for some petite set $K$ then $\pi(\mathsf{S}) < \infty$. By assumption, $V$ is bounded on $K$ and so our work above implies

$$\sup_{x \in K} \mathbb{E}_x[\tau_K] \leq \frac{1}{1-\alpha}\left(\sup_{x \in K} V(x) + b\right) < \infty.$$

Thus $\pi(\mathsf{S}) < \infty$.

According to Proposition 1 of Davydov [18], the coefficient of $\beta$-mixing of the strictly stationary process $\{\Phi_t : t = 0, \pm 1, \pm 2, \ldots\}$ is given by

$$\beta_k = \int_{\mathsf{S}} \pi(dx) || \mathcal{P}^k(x, \cdot) - \pi ||_{var}.$$

From (2.6), $\beta_k \leq R r^k \pi(\mathsf{S})$. Since $\pi(\mathsf{S}) < \infty$, $0 < R < \infty$, and $0 < r < 1$, the strictly stationary process is geometrically $\beta$-mixing.

□

### 2.3.5 Linear Algebra

In this section, we establish some necessary notation and results from linear algebra. Unless otherwise noted, matrices and vectors will have real entries. We denote the set of $d \times d$ matrices by $M_d$, the set of symmetric $d \times d$ matrices by $S_d$, the positive semidefinite $d \times d$ matrices by $S_d^+$, and the positive definite $d \times d$ matrices by $S_d^{++}$.

**Lemma 1** *Let $A$, $B$, and $C \in M_d$. Then*

(i) $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$,

(ii) $(A \otimes B)' = A' \otimes B'$,

(iii) *there exist unique $K_d$, $H_d \in M_{d(d+1)/2 \times d^2}$ such that* $\text{vech}(D) = H_d\text{vec}(D)$, $\text{vec}(D) = K_d'\text{vech}(D)$, *and $H_dK_d'$ is the identity matrix in $M_{d(d+1)/2}$ for every $D \in S_d$.*

**Proof** : For a proof of (i), see Lemma 4.3.1 of Johnson and Horn [32]. It is easy to see that (ii) holds by writing out the associated matrices, and (iii) is clear since the vec and vech operators are linear.

□

Next we examine properties of a certain class of linear maps. Let $n$ denote a nonnegative integer, and consider the function $\xi : M_d \to M_d$ defined for all $M \in M_d$ by

$$\xi(M) := \sum_{i=1}^{n} \Upsilon_i M \Upsilon_i'$$

where, for each $i$ in $\{1, \ldots, n\}$, $\Upsilon_i$ is some fixed $d \times d$ matrix.

Using the vec operator, and using Lemma 1, we can consider $\xi$ as a map from $\mathbb{R}^{d^2}$ into $\mathbb{R}^{d^2}$ such that

$$\text{vec}(\xi(M)) = \left( \sum_{i=1}^{n} \Upsilon_i \otimes \Upsilon_i \right) \text{vec}(M).$$

In this context, the map $\xi$ corresponds to left multiplication by the matrix

$$\Upsilon := \sum_{i=1}^{n} \Upsilon_i \otimes \Upsilon_i$$

Note that we have $\xi(S_d) \subseteq S_d$, i.e. the symmetric $d \times d$ matrices are mapped into themselves by $\xi$. We denote by $\tilde{\xi}$ the restriction of $\xi$ to the linear subspace $S_d$. Again using Lemma 1, we have for all $M$ in $S_d$,

$$\mathrm{vech}(\tilde{\xi}(M)) = \mathrm{vech}(\xi(M)) = H_d \mathrm{vec}(\xi(M)) = H_d \Upsilon \mathrm{vec}(M) = H_d \Upsilon K_d' \mathrm{vech}(M).$$

We can identify $S_d$ with $R^{d(d+1)/2}$ using the vech operator, and in this case $\tilde{\xi}$ corresponds to left multiplication by the matrix

$$\tilde{\Upsilon} := H_d \Upsilon K_d'.$$

**Lemma 2** *The follwing statements are equivalent:*

(i) *The spectral radius of $\xi$ is less than one.*

(ii) *The spectral radius of $\tilde{\xi}$ is less than one.*

(iii) *For any $C \in S_d^{++}$, there exists some $H \in S_d^{++}$ such that $H = C + \xi(H)$.*

**Proof** : It is clear that (i) implies (ii) since $\tilde{\xi}$ is a restriction of $\xi$.

To see that (ii) implies (iii), note that if the series $\sum_{n=0}^{\infty} \tilde{\xi}^n$ is convergent with respect to some operator norm, then we can define

$$H := \sum_{n=0}^{\infty} \tilde{\xi}^n(C). \tag{2.7}$$

It is clear that $H$ is symmetric. From the definitions of $\xi$ and $\tilde{\xi}$ we see that for any $M \in S_d^+$, each $\tilde{\xi}^n(M)$ is also an element of $S_d^+$. Thus $H - \tilde{\xi}^0(C) = H - C$ is symmetric and positive semidefinite. This implies that $H$ is positive definite.

From the Definition (2.7) of $H$, and since $\tilde{\xi}$ and $\xi$ coincide on $S_d$, we have

$$H = \sum_{n=0}^{\infty} \xi^n(C) = C + \xi\left(\sum_{n=1}^{\infty} \xi^{n-1}(C)\right) = C + \xi(H).$$

Thus (ii) implies (iii).

Finally, to see that (iii) implies (i), suppose that for any $C \in S_d^{++}$ there exists some $H \in S_d^{++}$ such that $H = C + \xi(H)$. Denote the complex $d \times d$ matrices by $M_d(\mathbb{C})$ and denote the conjugate transpose of a vector $x \in \mathbb{C}^d$ by $\bar{x}$. For every $N \in M_d(\mathbb{C})$ we can define

$$||N||_H := \sup_{x\in\mathbb{C}^d, \bar{x}Hx=1} |\bar{x}Nx|$$

which is a norm on $M_d(\mathbb{C})$ since $H \in S_d^{++}$. Then for all $x \in \mathbb{C}^d$,

$$|\bar{x}Nx| \leq ||N||_H(\bar{x}Hx)$$

Since the unit sphere $\{x \in \mathbb{C}^d : \bar{x}Hx = 1\}$ is compact, there exists for each $N \in M_d(\mathbb{C})$ some vector $x_N \in \mathbb{C}^d$ such that

$$||N||_H = |\bar{x_N}Nx_N|, \text{ and } \bar{x}_NHx_N = 1.$$

Now if $\lambda_\xi$ is an eigenvalue of $\xi$, then there is some nonzero $M \in M_d(\mathbb{C})$ such that

$$\lambda_\xi M = \xi(M) = \sum_{i=1}^n \Upsilon_i M \Upsilon_i'.$$

For every $x \in \mathbb{C}^d$, it follows that

$$\begin{aligned}
|\lambda_\xi||\bar{x}Mx| &= \left|\sum_{i=1}^n \bar{x}\Upsilon_i M \Upsilon_i' x\right| \\
&\leq \sum_{i=1}^n \left|(\overline{\Upsilon_i' x})M(\Upsilon_i' x)\right| \\
&\leq ||M||_H \sum_{i=1}^n \bar{x}\Upsilon_i H \Upsilon_i' x \\
&= ||M||_H \; \bar{x} \underbrace{\left(\sum_{i=1}^n \Upsilon_i H \Upsilon_i'\right)}_{=\xi(H)=H-C} x.
\end{aligned}$$

Now choosing $x_M$ such that $||M||_H = |\bar{x}_M M x_M|$ and $\bar{x}_M H x_M = 1$, we obtain (note $||M||_H \neq 0$),

$$|\lambda_\xi| \leq 1 - \bar{x_M}Cx_M < 1,$$

as desired.

$\square$

In the following, the matrices $\hat{A}_{i,k}$, $\hat{B}_{j,r}$, and $A_i$, $B_j$ are the coefficient matrices defined by (2.1)-(2.3).

**Lemma 3** *The spectral radius of* $\sum_{i=1}^{q} A_i + \sum_{j=1}^{p} B_j$ *is less than one if and only if there exists some positive definite matrix* $H$ *such that*

$$H = C + \sum_{i=1}^{q} \sum_{k=1}^{\ell_i} \hat{A}_{ik} H \hat{A}'_{ik} + \sum_{j=1}^{p} \sum_{r=1}^{s_j} \hat{B}_{jr} H \hat{B}'_{jr}.$$

**Proof** : Consider the map $\xi : M_d \to M_d$ defined for all $M \in M_d$ by

$$\xi(M) = \sum_{i=1}^{q} \sum_{k=1}^{\ell_i} \hat{A}_{ik} M \hat{A}'_{ik} + \sum_{j=1}^{p} \sum_{r=1}^{s_j} \hat{B}_{jr} M \hat{B}'_{jr}.$$

Using the vec operator and Lemma 1, $\xi$ corresponds to multiplication by the matrix

$$\Upsilon = \sum_{i=1}^{q} \sum_{k=1}^{\ell_i} \hat{A}_{ik} \otimes \hat{A}_{ik} + \sum_{j=1}^{p} \sum_{r=1}^{s_j} \hat{B}_{jr} \otimes \hat{B}_{jr}.$$

Letting $\tilde{\xi}$ denote the restriction of $\xi$ to $S_d$ and using the vech operator, it follows from Lemma 1 that $\tilde{\xi}$ corresponds to matrix multiplication by

$$H_d \Upsilon K'_d = \sum_{i=1}^{q} A_i + \sum_{j=1}^{p} B_j.$$

Thus, it follows from Lemma 2 that the spectral radius of $\sum_{i=1}^{q} A_i + \sum_{j=1}^{p} B_j$ is less than one if and only if there exists some $H \in S_d^{++}$ such that

$$H = C + \xi(H) = C + \sum_{i=1}^{q} \sum_{k=1}^{\ell_i} \hat{A}_{ik} H \hat{A}'_{ik} + \sum_{j=1}^{p} \sum_{r=1}^{s_j} \hat{B}_{jr} H \hat{B}'_{jr}.$$

$\square$

**Remark** : With small changes to the proof above we see that, equivalently, the spectral radius of $\sum_{i=1}^{q} A_i + \sum_{j=1}^{p} B_j$ is less than one if and only if there exists some positive definite matrix $H$ such that

$$H = C + \sum_{i=1}^{q} \sum_{k=1}^{\ell_i} \hat{A}'_{ik} H \hat{A}_{ik} + \sum_{j=1}^{p} \sum_{r=1}^{s_j} \hat{B}'_{jr} H \hat{B}_{jr}.$$

**Lemma 4** *If the spectral radius of* $\sum_{i=1}^{q} A_i + \sum_{j=1}^{p} B_j$ *is less than one, then the spectral radius of* $\sum_{j=1}^{p} B_j$ *is less than one, and the spectral radius of* $B$ *is less than* *one, where*

$$B := \begin{pmatrix} B_1 & B_2 & \cdots & B_{p-1} & B_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \ddots & \vdots & 0 \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & I & 0 \end{pmatrix}, \tag{2.8}$$

*and the sub-matrices $I$ are identity matrices with $d(d+1)/2$ rows and columns.*

**Proof** : Suppose the spectral radius of $\sum_{i=1}^{q} A_i + \sum_{j=1}^{p} B_j$ is less than one. Then there exists, due to Lemma 3, some matrix $H \in S_d^{++}$ such that

$$H = C + \sum_{i=1}^{q} \sum_{k=1}^{\ell_i} \hat{A}_{ik} H \hat{A}'_{ik} + \sum_{j=1}^{p} \sum_{r=1}^{s_j} \hat{B}_{jr} H \hat{B}'_{jr}.$$

Define $\tilde{C} := C + \sum_{i=1}^{q} \sum_{k=1}^{\ell_i} \hat{A}_{ik} H \hat{A}'_{ik}$. Then $\tilde{C} \in S_d^{++}$, and

$$H = \tilde{C} + \sum_{j=1}^{p} \sum_{r=1}^{s_j} \hat{B}_{jr} H \hat{B}'_{jr}$$

Using again Lemma 3, it follows that the spectral radius of $\sum_{j=1}^{p} B_j$ is less than one.

Now from Lemma 3, there exists some $\tilde{H} \in S_d^{++}$ such that

$$\tilde{H} = C + \sum_{j=1}^{p} \sum_{r=1}^{s_j} \hat{B}_{jr} \tilde{H} \hat{B}'_{jr}. \tag{2.9}$$

Let $\lambda_B$ denote an eigenvalue of $B$ with eigenvector $h = (h'_1, \ldots, h'_p)'$. Then,

$$\lambda_B h_1 = \sum_{j=1}^{p} B_j h_j, \text{ and } \lambda_B h_j = h_{j-1} \text{ for } 2 \le j \le p.$$

Thus $h_p \neq 0$ (otherwise $h$ would be zero) and

$$\lambda_B^p h_p = \lambda_B(\lambda_B^{p-1} h_p) = \lambda_B h_1 = \sum_{j=1}^{p} B_j h_j = \sum_{j=1}^{p} \lambda_B^{p-j} B_j h_p.$$

Let $M \in S_d$ such that $\mathrm{vech}(M) = h_p$. Then,

$$\lambda_B^p M = \sum_{j=1}^{p} \sum_{r=1}^{s_j} \lambda_B^{p-j} \hat{B}_{jr} M \hat{B}'_{jr}.$$

Define the norm $|| \cdot ||_{\tilde{H}}$, for any $N \in M_d(\mathbb{C})$, as in the proof of Lemma 2 by

$$||N||_{\tilde{H}} := \sup_{x\in\mathbb{C}^d, \bar{x}\tilde{H}x=1} |\bar{x}Nx|.$$

Then, for all $x \in \mathbb{C}^d$,

$$\begin{aligned}
|\lambda_B^p||\bar{x}Mx| &= \left|\sum_{j=1}^{p}\sum_{r=1}^{s_j} \lambda_B^{p-j}\bar{x}\hat{B}_{jr}M\hat{B}'_{jr}x\right| \\
&\leq \sum_{j=1}^{p}\sum_{r=1}^{s_j} |\lambda_B|^{p-j}|\bar{x}\hat{B}_{jr}M\hat{B}'_{jr}x| \\
&\leq ||M||_{\tilde{H}} \sum_{j=1}^{p}\sum_{r=1}^{s_j} |\lambda_B|^{p-j}|\bar{x}\hat{B}_{jr}\tilde{H}\hat{B}'_{jr}x|.
\end{aligned}$$

If we assume (by way of contradiction) that there is an eigenvalue $\lambda_B$ of $B$ with $|\lambda_B| \geq 1$, then choosing the vector $x$ such that $\bar{x}\tilde{H}x = 1$, and $|\bar{x}Mx| = ||M||_{\tilde{H}}$, and using (2.9), we have

$$\begin{aligned}
|\lambda_B|^p &\leq \sum_{j=1}^{p}\sum_{r=1}^{s_j} |\lambda_B|^{p-j}(\bar{x}\hat{B}_{jr}M\hat{B}'_{jr}x) \\
&\leq |\lambda_B|^{p-1}\left[\bar{x}\left(\sum_{j=1}^{p}\sum_{r=1}^{s_j}\hat{B}_{jr}\tilde{H}\hat{B}'_{jr}\right)x\right]. \\
&= |\lambda_B|^{p-1}[\bar{x}(\tilde{H}-C)x] = |\lambda_B|^{p-1}(1-\bar{x}Cx).
\end{aligned}$$

Since $C \in S_d^{++}$, we have $\bar{x}Cx > 0$. Thus $|\lambda_B|^p < |\lambda_B|^{p-1}$, i.e. $|\lambda_B| < 1$ which is a contradiction. Thus the spectral radius of $B$ is less than one.

□

### 2.3.6 Properties of the Markov Chain X

It is clear, from the vech representation (2.3) of a multivariate GARCH$(p, q)$ process, that the Markov chain $\mathbf{X}$ defined by (2.4) will take values in the state space

$$\mathsf{S} := \underbrace{\mathrm{vech}(S_d^{++}) \times \cdots \times \mathrm{vech}(S_d^{++})}_{\text{p factors}} \times \underbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}_{\text{q factors}}. \tag{2.10}$$

However, $\mathbf{X}$ is not irreducible on $\mathsf{S}$. In the next section, we will construct a subspace $\mathsf{A}_+$ of $\mathsf{S}$ on which $\mathbf{X}$ is irreducible, and we will show that it suffices to consider $\mathbf{X}$ restricted to $\mathsf{A}_+$. To this end, we note that

$$X_t = (h'_t, \ldots, h'_{t-p+1}, \varepsilon'_t, \ldots, \varepsilon'_{t-q+1})'$$

can be written as

$$\left(\text{vech}(C) + \sum_{i=1}^{q} A_i \text{vech}(\varepsilon_{t-i}\varepsilon'_{t-i}) + \sum_{j=1}^{p} B_j h_{t-j}, h'_{t-1}, \ldots, h'_{t-p+1}, \varepsilon'_t, \ldots, \varepsilon'_{t-q+1}\right)',$$

and since $\varepsilon_t = H_t^{1/2}\eta_t$, we see that

$$X_t = F(X_{t-1}, \eta_t).$$

More precisely, the function $F$ is defined by

$$F : (X_{t-1}, \eta_t) \mapsto \left(\kappa(X_{t-1}), h'_{t-1}, \ldots, h'_{t-p+1}, g \circ \mathcal{K}(X_{t-1})\eta_t, \varepsilon'_{t-1}, \ldots, \varepsilon'_{t-q+1}\right)', \quad (2.11)$$

where $\kappa$ is the polynomial mapping defined by

$$\kappa : X_{t-1} \mapsto \text{vech}(C) + \textstyle\sum_{i=1}^{q} A_i \text{vech}(\varepsilon_{t-i}\varepsilon'_{t-i}) + \sum_{j=1}^{p} B_j h_{t-j} = h_t,$$

and the function $\mathcal{K}$ is defined by $\text{vech}^{-1}(\kappa)$, i.e., $\mathcal{K}$ is composition of the polynomial $\kappa$ with the map that reverses the vech operator. Note that the function $F$ is polynomial in each coordinate except for the coordinate $p+1$.

#### 2.3.6.1 Properties of the Function F

Here we establish some smoothness and invertibility properties of the function $F$ defined by (2.11). These properties are necessary for proving irreducibility of $\mathbf{X}$, and for showing that $\mathbf{X}$ satisfies the Foster-Lyapunov drift condition.

**Definition 13** *A function $f : U \to \mathbb{R}^n$ (for any positive integer $n$) is called* **smooth** *if $U$ is an open subset of $\mathbb{R}^m$ (for some positive integer $m$), and if $f$ has continuous partial derivatives of all orders (see, for instance, Guillemin and Pollack [29] p.1-3). However, if the domain of $f$ is not open, then partial derivatives may not make sense. For an arbitrary set $X \subseteq \mathbb{R}^m$, we say that $f : X \to \mathbb{R}^n$ is smooth if $f$ may be locally extended to a smooth map on open sets. More specifically, if for each $x \in X$, there exists an open neighborhood $U_x$ of $x$, and a smooth map $\tilde{f} : U_x \to \mathbb{R}^n$ such that $f$ and $\tilde{f}$ agree on $U_x \cap X$.*

**Theorem 6** *The map $g : S_d^{++} \to S_d^{++}$ defined by $g(H) := H^{1/2}$ is a diffeomorphism.*

**Proof** : First note that a positive definite matrix has a unique positive definite square root. This result can be found, for instance, in the book by Horn and Johnson [32], Theorem 7.2.6. Therefore, $g$ is well-defined and bijective.

We can identify the space of symmetric $d \times d$ matrices with real entries, $S_d$, with $\mathbb{R}^m$, where $m := d(d+1)/2$, via the vech operator. If $U$ and $V$ are two simply connected open subsets of $\mathbb{R}^m$, then a differentiable map from $U$ to $V$ is a diffeomorphism if the differential is bijective at each point; we will use this criterion to show that the function $g^{-1}$ is a diffeomorphism.

The differential of $g^{-1}$ is defined, at the point $X$ in $S_d^{++}$, by

$$\begin{aligned} Jg_X^{-1}(H) &= \lim_{t\to 0} \frac{g^{-1}(X+tH) - g^{-1}(X)}{t} \\ &= \lim_{t\to 0} \frac{(X+tH)(X+tH) - XX}{t} \\ &= XH + HX. \end{aligned}$$

It follows from Theorem 1 of Potter [42] that, for any fixed $X$ in $S_d^{++}$,

$$XH + HX = 0$$

implies $H = 0$. Thus, the differential $Jg_X^{-1}$ is injective. As a linear map between vector spaces of the same dimension, it follows that $Jg_X^{-1}$ is bijective for each $X \in S_d^{++}$.

Finally, $S_d^{++}$ is an open and convex (hence simply connected) subset of $\mathbb{R}^m$ from Proposition 2.7 of Arsigny, Fillard, Pennec, and Ayache [2]. This completes the proof.

□

In assumption **A1**, we require that the random variables $\eta_t$ appearing in (2.1) admit a density that is nonzero in a neighborhood of the origin. We will denote the cumulative distribution function of the random variables $\eta_t$ by $\Gamma$, and the corresponding density by $\gamma$. Assumption **A1** implies that $\gamma$ admits a domain of positivity with nonempty interior, and we will denote this domain by $E$, i.e.,

$$E := \{x \in \mathbb{R}^d : \gamma(x) > 0\}. \tag{2.12}$$

Recalling Definition (2.10), where we define the state space $\mathsf{S}$, we may now state the following Corollary to Theorem 6.

**Corollary 3** *$F : \mathsf{S} \times E \to \mathsf{S}$ is smooth.*

**Proof** : First suppose that $\mathsf{S} \times E$, the domain of $F$, is open. In this case, the proof follows from combining the definition of $F$, (2.11), with Theorem 6 above, to see that $F$ is smooth in each coordinate, hence smooth.

More generally, although $\mathsf{S}$ is open as a product of finitely many open sets, we do not require the set $E \subseteq \mathbb{R}^d$ to be open. Thus, the domain $\mathsf{S} \times E$ of $F$ may not be open. In this case, we may consider $\tilde{F}$ defined as $F$ but having domain $\mathsf{S} \times \mathbb{R}^d$. Then $\tilde{F}$ is smooth, and $F$ may be locally extended to $\tilde{F}$.

$\square$

**Corollary 4** *For any fixed $x$ in $\mathsf{S}$, the function $F_x : E \to \mathsf{S}$ defined by $F_x(\cdot) := F(x, \cdot)$ is smooth. Furthermore, the inverse function $F_x^{-1}(\cdot) : F(\mathsf{S}) \to E$ exists and is smooth.*

**Proof** : It is clear that $F_x$ is smooth, since $F$ is smooth in each coordinate.

Regarding the inverse function, since $x$ is fixed, $F_x^{-1}(\cdot)$ simply left multiplies the $p+1$ coordinate of its argument by the matrix inverse of $g \circ \mathcal{K}(x)$. More precisely, if $F_x(\eta) = Y$, then letting $y_{p+1}$ denote the $p+1$ coordinate of $Y$, we have

$$F_x^{-1}(Y) = [g \circ \mathcal{K}(x)]^{-1} y_{p+1}.$$

Matrix inversion is a smooth operation on the space of nonsingular $d \times d$ matrices, so it is in particular smooth on $S_d^{++}$. Thus $F_x^{-1}(\cdot)$ is a composition of smooth maps, hence smooth.

$\square$

**Lemma 5** *For any fixed $x$ in $\mathsf{S}$, $F_x : E \to \mathsf{S}$ is an open map.*

**Proof** : To prove this lemma, we refer to the Invariance of Domain Theorem, due to Brouwer [15]. This theorem states that if $U$ is an open subset of $\mathbb{R}^n$, and if $f : U \to \mathbb{R}^n$ is injective and continuous, then $f$ is an open map.

From Corollary 4, the function $F_x : E \to \mathsf{S}$ is smooth (hence continuous), and invertible (hence injective), for any fixed $x$ in $\mathsf{S}$.

The set $E$ has open subsets since, by assumption **A1**, zero is an interior point of $E$, and for any open subset $U \subseteq E$, the restriction of $F_x$ to $U$ is continuous and injective, so $F_x(U)$ is open in $\mathsf{S}$.

□

**Lemma 6** *For any Borel set $A$ in $\mathsf{S}$, $\mathcal{P}(\cdot, A) = F(\cdot, \Gamma)(A)$*

**Proof** : This lemma follows from a simple change of variables, and the density transformation theorem. See, for instance, Theorem 1.2.1 and Theorem 1.2.2 in Bickel and Doksum [6]. Application of these theorems is justified by Corollary 4. Let $x \in \mathsf{S}$ be arbitrary. Then,

$$\begin{aligned}
\mathcal{P}(x, A) &= \mathbb{E}[\mathbb{I}_A(X_{t+1})|X_t = x] \\
&= \mathbb{E}[\mathbb{I}_A(F(x, \eta_{t+1}))] \\
&= \int_{\mathsf{S}} \mathbb{I}_A(F(x, y))d\Gamma(y) \\
&= \int_{F_x^{-1}(A)} \gamma(y)dy \\
&= \int_A \gamma(F_x^{-1}(u))|J_{F_x^{-1}}(u)|du \\
&= \int_A \gamma'_x(u)du, \text{ where } \gamma'_x \text{ is the density of } F_x(\Gamma) \\
&= F(x, \Gamma)(A).
\end{aligned}$$

□

### 2.3.7 Irreducibility

#### 2.3.7.1 The Set of Attainable States

We define a sequence of maps $\{F^k : k = 1, 2, \ldots\}$ inductively for arbitrary $x$ in $\mathsf{S}$, and arbitrary $e_1, \ldots, e_k$ in $E$, by $F^1(x, e_1) = F(x, e_1)$, and for $k > 1$,

$$F^k(x, e_1, \ldots, e_k) := F(F^{k-1}(x, e_1, \ldots, e_{k-1}), e_k). \tag{2.13}$$

**Definition 14** *For any initial state* $x_0$ *in* $\mathsf{S}$*, we define* $\mathsf{A}_+(x_0)$ *to be the* **set of attainable states** *that can be reached by the Markov process* $\mathbf{X}$ *given that* $X_0 = x_0$*. More formally,*

$$\mathsf{A}_+(x_0) := \bigcup_{k=1}^{\infty} \{F^k(x_1, e_1, \ldots, e_k) : e_1, \ldots, e_k \in E\}. \tag{2.14}$$

Meyn and Tweedie [38] explore, in great detail, the relationship between random processes of the form $X_t = F(X_{t-1}, \eta_t)$, where each $\eta_t$ is a random variable, and the closely related dynamical system that results if each $\eta_t$ is nonrandom. Along these lines, we consider for any fixed $a$ in $E$, and for any initial value $x_0$ in $\mathbb{R}^d$ the nonrandom sequence $\mathbf{X}^a(x_0)$ defined by $X_0^a(x_0) = x_0$, and for $k > 0$,

$$X_k^a(x_0) = F(X_{k-1}^a(x_0), a). \tag{2.15}$$

#### 2.3.7.2 Globally Attracting Points

**Definition 15** *A point* $x^*$ *in* $\mathbb{R}^d$ *is called a* **globally attracting point** *of the Markov chain* $\mathbf{X}$ *if there exists some* $a$ *in* $E$ *such that* $X_k^a(x) \to x^*$ *as* $k \to \infty$ *for every* $x \in \mathsf{S}$*, where* $X_k^a$ *is defined as in* (2.15)*.*

**Lemma 7** *The Markov chain* $\mathbf{X}$ *defined by (2.4) has a globally attracting point.*

**Proof** : Let $x$ in $\mathsf{S}$ be arbitrary, and consider the nonrandom sequence $\mathbf{X}^0(x)$ defined by $X_0^0(x) = x$, and for $k > 0$,

$$X_k^0(x) = F(X_{k-1}^0(x), 0).$$

Denote by $\varepsilon_t^0$, $h_t^0$, and $H_t^0$ the vectors and covariance matrix analogous to $\varepsilon_t$, $h_t$ and $H_t$, respectively, in $\mathbf{X}$. Note that $\varepsilon_t^0 = 0$ for each $t$. Thus,

$$h_t^0 = \mathrm{vech}(C) + \sum_{j=1}^{p} B_j h_{t-j}^0.$$

For all $k > q$, we can write $X_k^0(x)$ as $X_k^0(x) = \mathcal{C} + \mathfrak{B} X_{k-1}^0(x)$, where $\mathcal{C}$ is defined by $\mathcal{C} := (\mathrm{vech}(C)', 0, \ldots, 0)'$ and $\mathfrak{B}$ is a square matrix with $pd(d+1)/2$ rows and columns, given by

$$\mathfrak{B} := \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix},$$

and the matrix $B$ is defined by (2.8).

Since $X_k^0(x) = \mathcal{C} + \mathfrak{B}X_{k-1}^0(x)$, $\lim_{k\to\infty} X_k^0$ exists if and only if the spectral radius of $\mathfrak{B}$ is less than one; the spectral radius of $\mathfrak{B}$ is less than one from **A1** and Lemma 4. Thus $\lim_{k\to\infty} X_k^0 = x^*$ for some $x^*$ in $\mathsf{S}$.

$\square$

A few more things can be said about the attracting point $x^*$. From Lemma 3, there exists some positive definite matrix $H$ such that

$$H = C + \sum_{j=1}^{p} \sum_{r=1}^{s_j} \hat{B}_{jr} H \hat{B}'_{jr}.$$

With $h := \mathrm{vech}(H)$, it follows that

$$h = \mathrm{vech}(C) + \sum_{j=1}^{p} B_j h.$$

Thus, the globally attracting point $x^*$ is given by

$$x^* = (\underbrace{h, \cdots, h}_{\text{p factors}}, \underbrace{0, \cdots, 0}_{\text{qd factors}}) \,. \tag{2.16}$$

To simplify notation in the following, we will denote $\mathsf{A}_+$ by $\mathsf{A}_+(x^*)$. We are now ready to prove that $\mathbf{X}$ is $\psi$-irreducible when restricted to the state space $\mathsf{A}_+$.

**Proposition 2** X *is* $\psi$ *-irreducible and aperiodic when restricted to the state space* $\mathsf{A}_+ := \mathsf{A}_+(x^*)$.

**Proof** : Let $A$ denote an arbitrary Borel subset of $\mathsf{A}_+$. Then, using Lemma 6 and Corollary 4, we have for any $x$ in $\mathsf{A}_+$,

$$\lim_{x\to x^*} \mathcal{P}(x, A) = \lim_{x\to x^*} F(x, \Gamma)(A) = F(x^*, \Gamma)(A) = \mathcal{P}(x^*, A).$$

Define the measure $v$ by $v(A) := \mathcal{P}(x^*, A)$. Then if $v(A) \neq 0$, continuity of the first component of $\mathcal{P}$ ensures that there is some neighborhood $W$ of $x^*$ such that, for all $w$ in $W$,

$$\mathcal{P}(w, A) \geq \frac{v(A)}{2}. \tag{2.17}$$

Choose any subcollection of points $K := \{x_1, \ldots, x_r\}$ in $\mathsf{S}$, and consider the sequences $\mathbf{X}^0(x_1), \ldots, \mathbf{X}^0(x_r)$, defined as in (2.15). Then Lemma 7 implies that $X_k^0(x_i) \to x^*$ as

$k \to \infty$ for each $i$ in $\{1, \ldots, r\}$. Thus, there exists some $\ell > 0$ such that each $X_\ell^0(x_i)$ is in the neighborhood $W$.

Note that $X_\ell^0(x_i) = F^\ell(x_i, 0, \ldots, 0)$, and $F^\ell$ is continuous as a composition of continuous maps, so there exists, for each $i$ in $\{1, \ldots, r\}$, some neighborhood $U_i$ of $(x_i, 0, \ldots, 0)$ such that $F^\ell(U_i) \subseteq W$. Thus, for each $i$ in $\{1, \ldots, r\}$, $U_i$ contains $U_i' \times U_i^0$ where $U_i'$ and $U_i^0$ are neighborhoods of $x_i$ in $W$, and $(0, \ldots, 0)$ in $E^\ell$, respectively. Define

$$U_{(0,\ldots,0)} := \bigcap_{i=1}^{r} U_i^0,$$

which is also a neighborhood of the origin in $E^\ell$ such that $F(x_i, U_{(0,\ldots,0)}) \subseteq W$. Then $U_{(0,\ldots,0)}$ necessarily contains a neighborhood $U_0 \times \cdots \times U_0$, where $U_0$ is some small neighborhood of 0 in $E$, and we have for each $i$ in $\{1, \ldots, r\}$,

$$\mathcal{P}^\ell(x_i, A) \geq \mathbb{P}((\eta_1, \ldots, \eta_\ell) \in U_{(0,\ldots,0)}) = \mathbb{P}(\eta_1 \in U_0)^\ell = \Gamma(U_0)^\ell. \tag{2.18}$$

Using the Chapman-Kolmogorov equations (see Theorem 3.4.2 of Meyn and Tweedie [38]), (2.17), and (2.18), we see that, for each $i$ in $\{1, \ldots, r\}$,

$$\begin{aligned}
\mathcal{P}^{\ell+1}(x_i, A) &= \int_{\mathsf{S}} \mathcal{P}^\ell(x_i, dy)\mathcal{P}(y, A) \\
&\geq \int_W \mathcal{P}^\ell(x_i, dy)\mathcal{P}(y, A) \\
&\geq \frac{v(A)}{2} \int_W \mathcal{P}^\ell(x_i, dy) \\
&= \frac{v(A)}{2} \int_{\mathsf{S}} \mathcal{P}^\ell(x_i, dy)\mathbb{I}_W(y) \\
&= \frac{v(A)}{2} \int_{\mathsf{S}} \mathcal{P}^\ell(x_i, dy)\mathcal{P}^0(y, W) \\
&= \frac{v(A)}{2} \mathcal{P}^\ell(x_i, W) \\
&\geq \frac{v(A)}{2} \Gamma(U_0)^\ell.
\end{aligned}$$

The origin is an interior point of $E$ by assumption, so $U_0$ contains a nonempty open subset of $E$, and thus $\Gamma(U_0) > 0$. This shows that $\mathbf{X}$ is irreducible with respect to the measure $v$, and thus $\mathbf{X}$ is irreducible with respect to some maximal measure $\psi$ such that $v$ is absolutely continuous with respect to $\psi$.

Finally, to see that $\mathbf{X}$ is aperiodic we suppose that $\mathbf{X}$ has period $d$. Then from Theorem 3, there are disjoint Borel sets $D_1, \ldots, D_d$ in $\mathsf{A}_+$ such that $\psi((\bigcup_{i=1}^d D_i)^c) = 0$, and such that for all $i$ in $1, \ldots, d$, and for all $x$ in $D_i$, $\mathcal{P}(x, D_{i(mod(d))+1}) = 1$.

Since $v$ is absolutely continuous with respect to $\psi$, $v((\bigcup_{i=1}^d D_i)^c) = 0$. Thus there is some $D_i$ with positive $v$-measure. Without loss of generality, suppose this is $D_1$. Let $x_1 \in D_1$, and $x_d \in D_d$. Our work above shows that

$$\mathcal{P}^{\ell+1}(x_1, D_1) > 0, \text{ and } \mathcal{P}^{\ell+1}(x_d, D_1) > 0.$$

Thus, the integers $\ell + 1$ and $\ell$ are both divisible by $d$. Therefore $d = 1$.

$\square$

**Lemma 8** *For all positive integers $j$, $F^j(\mathsf{A}_+, E^j) \subseteq \mathsf{A}_+$. Thus $\mathbf{X}$ can be restricted to the set of attainable states $\mathsf{A}_+$.*

**Proof** : Let the positive integer $j$ be arbitrary, and suppose that $y \in F^j(\mathsf{A}_+, E^j)$. Then there exists some $x \in \mathsf{A}_+$ such that $y \in F^j(x, E^j)$.

Since $x \in \mathsf{A}_+$, there is some positive integer $m$ such that $x \in F^m(x^*, E^m)$. Thus $y \in F^j(F^m(x^*, E^m), E^j) = F^{j+m}(x^*, E^{j+m}) \subseteq \mathsf{A}_+$.

$\square$

## 2.4 Checking the Foster-Lyapunov Criteria

In this section, we construct a Lyapunov function $V$ such that $\mathbf{X}$ satisfies the Foster-Lyapunov drift condition. First, we need the following Lemma.

**Lemma 9** *Any compact subset of $\mathsf{A}_+$ is small.*

**Proof** : A Markov chain is said to have the Feller property (Definition 11) if it has a transition probability kernel $\mathcal{P}$ such that $\mathcal{P}(\cdot, O)$ is lower semicontinuous for any open set $O$ in its state space.

From Lemma 6 and Corollary 4, the transition probability kernel $\mathcal{P}$ of $\mathbf{X}$ is continuous as a function of its first component. Thus $\mathcal{P}(\cdot, O)$ is in particular lower semicontinuous for any open set $O \in \mathsf{A}_+$, so $\mathbf{X}$ has the Feller property.

Meyn and Tweedie [38] prove in their Theorem 5.5.7 that if a Markov chain is $\psi$-irreducible and aperiodic on some state space, then every petite set in that state space is small. From Proposition 2, it follows that every petite set in $\mathsf{A}_+$ is small.

Meyn and Tweedie [38] show in Theorem their 6.2.8 (ii) that if a Markov chain has the Feller property, and is $\psi$-irreducible for some measure $\psi$ with nonempty interior, then all compact subsets of the state space are petite.

The measure $v$, defined by $v(A) := \mathcal{P}(x^*, A)$ as in the proof of Proposition 2, is absolutely continuous with respect to the maximal measure $\psi$ such that $\mathbf{X}$ is $\psi$-irreducible. Thus it suffices to show that the support of $v$ has nonempty interior.

For any Borel set $A$ in $\mathsf{A}_+$, $v(A) = \mathcal{P}(x^*, A) = F_{x^*}(\Gamma)(A)$. $\Gamma$ has density $\gamma$, so $v = F_{x^*}(\Gamma)$ has density $\gamma_v$ given by

$$\gamma_v(x) := \gamma(F_{x^*}^{-1}(x))|J_{F_{x^*}^{-1}}(x)|.$$

The domain of positivity of $v$ is

$$\begin{aligned}\{x \in \mathbb{R}^d : \gamma_v(x) > 0\} &= \{x \in \mathbb{R}^d : \gamma(F_{x^*}^{-1}(x)) > 0\} \\ &= \{x \in \mathbb{R}^d : F_{x^*}^{-1}(x) \in E\} \\ &= F_{x^*}(E).\end{aligned}$$

$E$ has nonempty interior by assumption **A1**, and $F_{x^*}$ is an open map from Lemma 5, thus $F_{x^*}(E)$ has nonempty interior and the proof is complete.

□

**Proposition 3** $\mathbf{X}$ *satisfies the Foster-Lyapunov drift condition.*

**Proof** : For ease of exposition, we focus on the case where $p = q = 1$ and $\ell_1 = s_1 = 1$ in the BEKK representation (3.2). We will denote the coefficient matrices $\hat{A}_{11}$ and $\hat{B}_{11}$ of the BEKK representation simply by $\hat{A}$ and $\hat{B}$.

Define matrices $V_1$ and $V_2$ by

$$V_1 := \frac{1}{2}C + \hat{B}H\hat{B}', \qquad V_2 := \frac{1}{2}C + \hat{A}H\hat{A}'.$$

From Lemma 8, $\mathbf{X}$ can be restricted to the state space $\mathsf{A}_+$. Thus, if $X_k$ is an arbitrary term of $\mathbf{X}$, then the realization of $X_k$ is an element of $\mathsf{A}_+$, and we define the Lyapunov function $V$ from $\mathsf{A}_+$ to $[1, \infty)$ by

$$V(X_k) := \mathrm{tr}(V_1 H_k) + \varepsilon_k' V_2 \varepsilon_k + 1.$$

Let $x := (h_{t-1}', \varepsilon_{t-1}')'$ denote an arbitrary point in $\mathsf{A}_+$. Then,

$$\mathbb{E}[V(X_t)|X_{t-1} = x] = \mathbb{E}[\mathrm{tr}(V_1 H_t) + \varepsilon_t' V_2 \varepsilon_t | X_{t-1} = x] + 1.$$

Using the BEKK representation of $H_t$, the right-hand side of the expression above can be written as

$$\mathbb{E}[\varepsilon_t' V_2 \varepsilon_t | X_{t-1} = x] + \mathrm{tr}(V_1 C) + \mathrm{tr}(V_1 \hat{A} \varepsilon_{t-1} \varepsilon_{t-1}' \hat{A}') + \mathrm{tr}(V_1 \hat{B} H_{t-1} \hat{B}') + 1,$$

or equivalently,

$$\mathbb{E}[\varepsilon_t' V_2 \varepsilon_t | X_{t-1} = x] + \mathrm{tr}(V_1 C) + \varepsilon_{t-1}' \hat{A}' V_1 \hat{A} \varepsilon_{t-1} + \mathrm{tr}(\hat{B}' V_1 \hat{B} H_{t-1}) + 1. \tag{2.19}$$

Let us now examine the first term of (2.19). We observe that

$$\begin{aligned}
\mathbb{E}[\varepsilon_t' V_2 \varepsilon_t | X_{t-1} = x] &= \mathbb{E}[\mathrm{tr}(\varepsilon_t (V_2 \varepsilon_t)' | X_{t-1} = x] \\
&= \mathrm{tr}(\mathbb{E}[\varepsilon_t \varepsilon_t' V_2 | X_{t-1} = x] \\
&= \mathrm{tr}(\mathbb{E}[\varepsilon_t \varepsilon_t' | X_{t-1} = x] V_2) \\
&= \mathrm{tr}(H_t V_2) \\
&= \mathrm{tr}(V_2 C) + \mathrm{tr}(V_2 \hat{A} \varepsilon_{t-1} \varepsilon_{t-1}' \hat{A}') + \mathrm{tr}(V_2 \hat{B} H_{t-1} \hat{B}') \\
&= \mathrm{tr}(V_2 C) + \varepsilon_{t-1}' \hat{A}' V_2 \hat{A} \varepsilon_{t-1} + \mathrm{tr}(\hat{B}' V_2 \hat{B} H_{t-1}).
\end{aligned}$$

Combining the expression above with (2.19) gives

$$\begin{aligned}
&\mathbb{E}[V(X_t)|X_{t-1} = x] = \\
&\qquad \mathrm{tr}((V_1 + V_2) C) + \varepsilon_{t-1}' \hat{A}' (V_1 + V_2) \hat{A} \varepsilon_{t-1} + \mathrm{tr}(\hat{B}' (V_1 + V_2) \hat{B} H_{t-1}) + 1.
\end{aligned} \tag{2.20}$$

Note that

$$\hat{B}' (V_1 + V_2) \hat{B} = V_1 - \frac{1}{2} C,$$

and

$$\hat{A}' (V_1 + V_2) \hat{A} = V_2 - \frac{1}{2} C.$$

Thus, $V_k - (1/2)C$ is positive semidefinite for $k = 1, 2$. Define, for $k = 1, 2$,

$$\alpha_k := \max\left\{x'\left(V_k - \frac{1}{2}C\right)x : x \in \mathbb{R}^d, x'V_kx = 1\right\}.$$

Since each $V_k$ is positive definite, it follows that $x'V_ky$ is a well-defined inner product of $x$ with $y$ (for $k = 1, 2$), and thus each $\alpha_k$ is the maximum of a continuous function over a compact set (the unit circle) in $\mathbb{R}^d$. It follows that, (for $k = 1, 2$), there exists some $x_k$ in $\mathbb{R}^d$ such that $x_k'V_kx_k = 1$ and

$$\alpha_k = x_k'\left(V_k - \frac{1}{2}C\right)x_k = 1 - \frac{1}{2}x_k'Cx_k.$$

Let $\alpha_m = \max\{\alpha_1, \alpha_2\}$, and let $V_m$, $x_m$ denote the matrix and vector (respectively) such that

$$\alpha_m = x_m'\left(V_m - \frac{1}{2}C\right)x_m.$$

Then $0 \leq \alpha_m$ since $V_m - (1/2)C$ is positive semidefinite, and $\alpha_m < 1$ since

$$\alpha_m = 1 - \frac{1}{2}(x_k'Cx_k),$$

and $C$ is positive definite. Next we claim, for $k$ in $\{1, 2\}$, that

$$V_k - \frac{1}{2}C \leq \alpha_mV_k. \tag{2.21}$$

To see this, note that $x_m'Cx_m \leq y'Cy$ for all $y$ such that $y'V_ky := ||y||_k^2 = 1$. Thus, for all nonzero $y$ in $\mathbb{R}^d$,

$$\frac{y'Cy}{||y||_k^2} \geq x_m'Cx_m = x_m'Cx_m\frac{y'V_ky}{||y||_k^2},$$

and, for all $y$ in $\mathbb{R}^d$,

$$y'Cy \geq x_m'Cx_my'V_ky.$$

In other words,

$$C \geq x_m'Cx_mV_k.$$

Thus, we have

$$\frac{1}{2}C \geq \frac{1}{2}x_m'Cx_m = V_k - \alpha_mV_k,$$

and therefore (2.21) holds.

If $M$ and $N$ are positive semidefinite matrices such that the product $MN$ is defined, and nonzero, then $\mathrm{tr}(M) \geq 0$, and $\mathrm{tr}(M)\mathrm{tr}(N) \geq \mathrm{tr}(MN) \geq 0$ (see pages 306-307 of Bernstein [5]). Thus,

$$V_k - \frac{1}{2}C \leq \alpha_m V_k$$

implies that, for $k$ in $\{1, 2\}$, and for any positive semidefinite matrix $M$,

$$\mathrm{tr}\left[\left(V_k - \frac{1}{2}C\right) M\right] \leq \alpha_m \mathrm{tr}(V_k M).$$

Applying this to (2.20),

$$\begin{aligned}
\mathbb{E}[V(X_t)|X_{t-1} = x] &= \mathrm{tr}((V_1 + V_2)C) + \varepsilon_{t-1}'\hat{A}'(V_1 + V_2)\hat{A}\varepsilon_{t-1} + \mathrm{tr}(\hat{B}'(V_1 + V_2)\hat{B}) + 1 \\
&= \mathrm{tr}((V_1 - \frac{1}{2}C)H_{t-1}) + \mathrm{tr}[(V_2 - \frac{1}{2}C)\varepsilon_{t-1}\varepsilon_{t-1}')] + \mathrm{tr}(HC) + 1 \\
&\leq \alpha_m \mathrm{tr}(V_1 H_{t-1}) + \alpha_m \mathrm{tr}(V_2 \varepsilon_{t-1}\varepsilon_{t-1}') + \mathrm{tr}(HC) + 1 \\
&= \alpha_m V(x) + \mathrm{tr}(HC) + 1 - \alpha_m.
\end{aligned}$$

Defining $\alpha := (1/2)\alpha_m + (1/2)$ in $[1/2, 1)$, $b := \mathrm{tr}(HC) + 1 - \alpha_m$, we have

$$\mathbb{E}[V(X_t)|X_{t-1} = x] \leq \alpha V(x) + b\mathbb{I}_K(x),$$

where $K$ is the set

$$K := \left\{ x \in \mathsf{A}_+ : 1 \leq V(x) \leq \frac{b}{\alpha - \alpha_m} \right\}.$$

It is clear that $V$ is bounded on $K$, and the proof is finished if we can show that $K$ is a small set. By Lemma 9 it suffices to show that $K$ is compact. It is clear that $K$ is closed in $\mathsf{A}_+$ as the preimage of a closed interval under a continuous map, and we claim that $K$ is bounded hence compact by the Heine-Borel theorem.

To show that $K$ is bounded, we will show that the quantities $||h||_1$, and $||\varepsilon||_2$, are uniformly bounded for all $x := (h', \varepsilon')'$ in $K$ ,where the vector norms $|| \cdot ||_1$ and $|| \cdot ||_2$ are defined by

$$||h||_1 := \sum_{i=1}^{d(d+1)/2} |h_i|,$$

and

$$||\varepsilon||_2 := (\sum_{j=1}^{d} \varepsilon_j^2)^{1/2}.$$

Let $\lambda_{min}(A)$ denote the smallest eigenvalue of a matrix $A$. Proposition 8.4.13 of Bernstein [5] shows that, for any positive semidefinite matrices $A$, and $B$,

$$\lambda_{min}(A)\text{tr}(B) \leq \text{tr}(AB). \tag{2.22}$$

If $x := (h', \varepsilon')' \in K$ then, from the definition of $K$,

$$\text{tr}(HV_1) + \text{tr}(V_2\varepsilon\varepsilon') \leq \frac{b}{\alpha - \alpha_m} + 1,$$

where $H$ is the positive definite matrix such that $\text{vech}(H) = h$. Applying (2.22),

$$\text{tr}(H) + \text{tr}(\varepsilon\varepsilon') \leq \frac{1}{\lambda}\left(\frac{b}{\alpha - \alpha_m} + 1\right) \tag{2.23}$$

where $\lambda := min\{\lambda_{min}(V_1), \lambda_{min}(V_2)\}$. Note that $\lambda > 0$, $\text{tr}(H) > 0$, and $\text{tr}(\varepsilon\varepsilon') \geq 0$ since $V_1$, $V_2$, and $H$ are positive definite, and $\varepsilon\varepsilon'$ is positive semidefinite.

Consider the matrix norms $|| \cdot ||_{m1}$ and $|| \cdot ||_{m2}$ defined, for any matrix $A$, by

$$||A||_{m1} := \sum_{i,j} |A_{i,j}|,$$

and

$$||A||_{m2} := (\sum_{i,j} A_{i,j}^2)^{1/2} = \text{tr}(A'A)^{1/2}.$$

It is clear from the definitions that

$$||\text{vech}(A)||_1 \leq ||A||_{m1}, \text{ and } ||y||_2 \leq ||yy'||_{m2}, \tag{2.24}$$

holds for any matrix $A$, and for any vector $y$. If $A$ is positive semidefinite, then

$$||A||_{m2} \leq (\text{tr}(A')\text{tr}(A))^{1/2} = \text{tr}(A). \tag{2.25}$$

Due to the equivalence of finite-dimensional norms, there exist positive constants $M_1$ and $M_2$ such that, for any matrix $A$,

$$||A||_{m1} \leq M_1||A||_{m2}, \text{ and } ||A||_{m2} \leq M_2||A||_{m1}. \tag{2.26}$$

Thus, applying (respectively) (2.24), (2.26), (2.25), and (2.23), we have

$$||h||_1 \leq ||H||_{m1} \leq M_1||H||_{m2} \leq M_1\mathrm{tr}(H) \leq \frac{M_1}{\lambda}\left(\frac{b}{\alpha-\alpha_m}+1\right),$$

and similarly,

$$||\varepsilon||_1 \leq ||\varepsilon\varepsilon'||_{m2} \leq M_2||\varepsilon\varepsilon'||_{m2} \leq \mathrm{tr}(\varepsilon\varepsilon') \leq \frac{M_2}{\lambda}\left(\frac{b}{\alpha-\alpha_m}+1\right).$$

□

## 2.5 Existence of Multivariate GARCH

**Theorem 7** *Under assumptions* **A1** *and* **A2***, the process* $\boldsymbol{\varepsilon}$ *defined by (2.1) exists and is unique. Furthermore,* $\boldsymbol{\varepsilon}$ *is positive Harris recurrent, geometrically ergodic, and the stricly stationary solution* $\{\varepsilon_t : t = 0, \pm 1, \pm 2, \ldots\}$ *is geometrically* $\beta$*-mixing.*

**Proof** : From the remark following Definition 6, it suffices to show that the Markov chain $\mathbf{X}$ has a strictly stationary solution that is positive Harris recurrent, geometrically ergodic, geometrically $\beta$-mixing, and unique.

From Theorem 5, such a solution exists if $\mathbf{X}$ is $\psi$-irreducible, aperiodic, and satisfies the Foster-Lyapunov condition. From Proposition 2, $\mathbf{X}$ is $\psi$-irreducible and aperiodic when restricted to the state space $\mathsf{A}_+$, and from Lemma 8, it suffices to consider $\mathbf{X}$ restricted to $\mathsf{A}_+$. From Proposition 3, $\mathbf{X}$ (restricted to $\mathsf{A}_+$) satisfies the Foster-Lyapunov drift condition. Thus, the process $\boldsymbol{\varepsilon}$ exists.

The proof is complete upon showing that $\boldsymbol{\varepsilon}$ is unique. To simplify notation, we take $p = q = 1$ in (2.3). Recursive iteration of (2.3) yields

$$\begin{aligned} h_t &= \mathcal{C} + As_{t-1} + Bh_{t-1}, \\ &= \mathcal{C} + As_{t-1} + B\mathcal{C} + BAs_{t-2} + B^2h_{t-2}, \\ &\vdots \\ &= \sum_{i=1}^{\infty} B^{i-1}\left(\mathcal{C} + As_{t-i}\right). \end{aligned}$$

Define, for all integers $t$, and each integer $N > 1$,

$$h_t(N) = \mathcal{C} + \sum_{i=1}^{N} B^{i-1}\left(\mathcal{C} + As_{t-i}\right) + B^{N+1}h_{t-N-1}.$$

Suppose $\tilde{\varepsilon}_t = \Sigma_t^{1/2}\eta_t$ is a strictly stationary solution to (2.1), and let $\sigma_t = \text{vech}(\Sigma_t)$. Then $\sigma_t$ must satisfy the recursion above, i.e., for all $N > 1$,

$$\begin{aligned}\sigma_t &= h_t(N) + B^{N+1}h_{t-N-1} \\ &= \sum_{i=1}^{\infty} B^{i-1}\left(\mathcal{C} + As_{t-i}\right),\end{aligned}$$

and

$$\sigma_t - h_t = \{h_t(N) - h_t\} + B^{N+1}\sigma_{t-N-1}.$$

The term in braces above converges to zero almost surely as $N$ goes to infinity, and since the series $\sum_{i=1}^{\infty} B^{i-1}\left(\mathcal{C} + As_{t-i}\right)$ converges almost surely, $B^{N+1}$ converges to zero almost surely as $N$ goes to infinity.

By stationarity, the distribution of $\sigma_{t-N-1}$ is independent of $N$. Thus, $B^{N+1}\sigma_{t-N-1}$ converges to zero in probability as $N$ goes to infinity, and thus $h_t - \sigma_t$ converges to zero in probability as $N$ goes to infinity. Since the terms are independent of $N$, this implies $\sigma_t = h_t$ almost surely.

$\square$

# CHAPTER 3

# ESTIMATION OF MULTIVARIATE GARCH

## 3.1 Introduction

In the univariate case, numerous techniques have been investigated for parameter estimation of GARCH processes; least squares estimators, least absolute deviation estimators and $L^p$ estimators have appeared in the literature. However, estimation by Gaussian quasi-maximum likelihood (QML) is perhaps the most popular, because it is robust to the distribution of the underlying process, and it is consistent and asymptotically normal without imposing moment conditions on the observed process. Francq and Zakoïan (2009a) survey the existing univariate GARCH parameter estimation methods and their asymptotic properties.

Multivariate GARCH processes are important because, frequently, interdependence is observed between different univariate processes. Covariances are used for calculations of hedge ratios, betas of CAPM (Capital Asset Pricing Model), portfolio VaR (Value at Risk) estimates, and asset weights in portfolios. Additionally, multivariate GARCH models have been used by Carvalho (2007), and Tse and Tsui (2002) to investigate contagion across financial markets.

In the multivariate case, parameter estimation research has focused primarily on Gaussian quasi-maximum likelihood estimation (QMLE). Consistency and asymptotic normality of the QMLE were established for models admitting a BEKK representation by Comte and Liebermann (2003) under the assumption of independent coordinates for the innovations, and a moment of order eight for the process. Recently, Hafner and Preminger (2009a) established asymptotic normality of the QMLE under the weaker assumption of a sixth order moment for the observed process. In this chapter we prove asymptotic normality of QMLE for the BEKK representation assuming only a fourth order moment for the process.

Despite favorable asymptotic properties, estimation of multivariate GARCH parameters by QMLE is problematic. In practice, QMLE is computationally intense due to the highly nonlinear form of the log-likelihood function, and the large number of parameters that must be estimated in the multivariate framework. To reduce the burden of parameter estimation, many authors have proposed restricted versions of the general BEKK and vech models; for instance, various factor and conditional correlation models have been proposed. Silvennoinen and Terasvïrta (2008) survey the most popular of these multivariate GARCH model variants. These restricted models reduce the number of parameters that must be estimated, but parameter reduction inevitably resuts in information loss. We propose a method for reducing the computational intensity of multivariate GARCH models, without reducing the number of model parameters.

Engle and Mezrich [23] proposed a two-step variance targeting estimation (VTE) method to reduce the computational intensity of parameter estimation in the scalar BEKK model of Engle and Kroner [22]. This method is based on a reparametrization of the volatility equation in terms of the long-run variance. A first-step estimate of the long-run variance is computed and, conditioning on this estimate, the remaining parameters are estimated by QML in a second step.

Francq, Horváth, and Zakoïan (2009) established asymptotic properties of the VTE method applied to univariate GARCH models. In this chapter, we establish strong consistency and asymptotic normality for the VTE method applied to multivariate GARCH models. For clarity of exposition, we focus on the GARCH(1,1) case and we include only the chief results in the main sections of this chapter. Detailed proofs are placed in section 3.4.

## 3.2 Notation and Preliminaries

We consider an $\mathbb{R}^d$-valued multivariate GARCH(1,1) process, $\boldsymbol{\varepsilon} = \{\varepsilon_t : t = 0, \pm 1, \pm 2, \ldots\}$, where

$$\varepsilon_t = H_t^{1/2} \eta_t, \tag{3.1}$$

and we assume that each $H_t$ admits a BEKK representation (see section 1.3 of Chapter 1) given by

$$H_t = C + \hat{A}\varepsilon_{t-1}\varepsilon_{t-1}'\hat{A}' + \hat{B}H_{t-1}\hat{B}'. \tag{3.2}$$

The corresponding vech representation is denoted by

$$h_t = \mathcal{C}_0 + A_0 s_{t-1} + B_0 h_{t-1}, \tag{3.3}$$

where $s_t = \text{vech}(\varepsilon_t\varepsilon_t')$, $h_t = \text{vech}(H_t)$, $\mathcal{C}_0 = \text{vech}(C_0) \in \mathbb{R}^m$, $m = d(d+1)/2$, and the sequence $\{\eta_t : t = 0, \pm 1, \pm 2, \ldots\}$ consists of independent and identically distributed (i.i.d.) $\mathbb{R}^d$-valued random variables with mean zero and unit covariance.

The space of $d \times d$ matrices will be denoted by $\mathcal{M}_d$, and the space of positive definite $d \times d$ matrices will be denoted by $S_d^{++}$. We denote that a matrix $X$ is positive semidefinite (or positive definite) by writing $X \geq 0$ (or $X > 0$), and we assume, for any positive definite matrix $X$, that $X^{1/2}$ is the unique positive definite matrix whose square is $X$, i.e. $X^{1/2} > 0$ and $X^{1/2}X^{1/2} = X$. We note that the vech operator is linear and invertible, and we denote its inverse simply by $\text{vech}^{-1}$.

Throughout this chapter, $||\cdot||$ denotes the Euclidean norm for vectors and matrices, i.e.,

$$||X|| := \text{tr}(X'X). \tag{3.4}$$

The spectral radius of any square matrix $X$ is denoted by $\rho(X)$, and the spectral norm for vectors and matrices is denoted by $N(\cdot)$, i.e.,

$$N(X) := \rho(X'X). \tag{3.5}$$

Necessary and sufficient conditions for existence of a unique nonanticipative, weakly stationary, strictly stationary, $\beta$-mixing and ergodic solution to the process $\boldsymbol{\varepsilon}$ described by (3.1)-(3.3) are desirable for estimation theory, but are not currently known.

In the univariate case, Bougerol and Picard [11] extended the results of Nelson [40] to establish necessary and sufficient conditions for strict stationarity and ergodicity in terms of the top Lyapunov exponent of a matrix composed of the innovations and coefficients of the process.

In the multivariate framework, Dennis, Hansen and Rahbek (2002) established sufficient conditions for geometric ergodicity of ARCH($q$) models admitting a BEKK representation. Francq, and Zakoïan (2009) provide a detailed proof of a result due to Ling and McAleer (2003) yielding sufficient conditions for strict stationarity of the CCC-GARCH model of Bollerslev [10]. Hafner and Preminger [31] provide sufficient conditions for strict stationarity and ergodicity of a factor GARCH model. In Chapter 2 of this dissertation, we give a detailed proof, based partly of the work of Boussama [12], that the process $\boldsymbol{\varepsilon}$ described by (3.1)-(3.3) exists and has a unique nonanticipative, weakly stationary, strictly stationary, $\beta$-mixing and ergodic solution under the following assumptions:

**A1** : The matrices $A_0$, $B_0$ and $C_0$ appearing in (3.3) are positive definite.

**A2** : The random variables $\eta_t$ admit a density that is nonzero in a neighborhood of the origin.

**A3** : $\rho(A_0 + B_0)$, the spectral radius of $A_0 + B_0$, is less than one.

**Remark 2** *Condition* **A2** *is a standard condition for proving $\beta$-mixing, but is not generally used for proving stationarity.*

**Remark 3** *In the univariate case, condition* **A3** *is sufficient but not necessary for strict stationarity. (See Berkes, Horváth and Kokoszka (2004).)*

Under assumptions **A1** – **A3**, the long-run (unconditional) variance of $(\varepsilon_t)$ is finite and is given, through recursive iteration of (3.3), by

$$\Gamma_0 := \text{vech}^{-1}(\gamma_0)$$

where

$$\gamma_0 := (I_m - A_0 - B_0)^{-1}\mathcal{C}_0 := K_0^{-1}\mathcal{C}_0.$$

A reparametrization of the volatility equation (3.3) yields

$$h_t = K_0\gamma_0 + A_0 s_{t-1} + (I - K_0 - A_0)h_{t-1}, \tag{3.6}$$

where

$$K_0 + A_0 + B_0 = I_m.$$

This reparametrization shows that the volatility at time $t$, $h_t$, may be interpreted as a weighted average of the long-run variance $\gamma_0$, the square of the last return $s_{t-1} = \mathrm{vech}(\varepsilon_{t-1}\varepsilon'_{t-1})$, and the previous volatility $h_{t-1}$.

### 3.2.1 Gaussian QMLE

In this section, we discuss Gaussian QML estimation for the unknown parameters of a multivariate GARCH process. Estimation by QML can be recommended for multivariate GARCH processes because the estimators are consistent under mild conditions, and we show that they are asymptotically normal under the condition of a fourth order moment; this is an improvement to the sixth order moment condition required in the existing literature. Although QMLE would typically be used to estimate the matrices $A_0$, $B_0$, and $\mathcal{C}_0$ in (3.3), we describe here the (equivalent) QML estimation of the matrices $A_0$, $K_0$, and $\gamma_0$ in the reparametrized model (3.6).

The vector of true, unknown parameters in (3.3) is denoted by

$$\theta_0 := (\gamma_0', \mathrm{vech}(A_0)', \mathrm{vech}(K_0)')' := (\gamma_0', \lambda_0)',$$

and is assumed to exist in some parameter space $\Theta$. Let $(\varepsilon_1, \ldots, \varepsilon_n)$ denote a realization of size $n$ of the unique, nonanticipative and stationary solution to the model (3.1), and denote an arbitrary element of $\Theta$ by

$$\theta := (\gamma', \mathrm{vech}(A)', \mathrm{vech}(K)')' := (\gamma', \lambda)'.$$

The Gaussian quasi-likelihood function is given by

$$\tilde{L}_n(\theta) := \prod_{t=1}^{n} |\tilde{\Sigma}_t|^{-1/2} \exp(-\frac{1}{2}\varepsilon_t'\tilde{\Sigma}_t^{-1}\varepsilon_t),$$

where $\tilde{\Sigma}_t := \tilde{\Sigma}_t(\theta)$ is defined by

$$\mathrm{vech}(\tilde{\Sigma}_t) := \tilde{\sigma}_t^2 := \tilde{\sigma}_t^2(\theta) := K\gamma + As_{t-1} + (I - K - A)\tilde{\sigma}_{t-1}^2, \tag{3.7}$$

given initial values $\varepsilon_0$ and $\tilde{\sigma}_0^2$. The Gaussian QMLE of $\theta_0$ is the location of the maximum of $\tilde{L}_n(\theta)$, i.e.,

$$\hat{\theta}_n^* := \underset{\theta\in\Theta}{\mathrm{argmax}}\tilde{L}_n(\theta) = \underset{\theta\in\Theta}{\mathrm{argmin}}\frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_t(\theta), \tag{3.8}$$

where

$$\tilde{\ell}_t(\theta) := \log(\det \tilde{\Sigma}_t) + \varepsilon_t' \tilde{\Sigma}_t^{-1} \varepsilon_t. \tag{3.9}$$

### 3.2.2 The VTE Method

The VTE method involves (i) reparametrizing the volatility equation (3.3) as (3.6), (ii) estimating $\gamma_0$ by the sample covariance, and then (iii) estimating $\lambda_0 = (\mathrm{vech}(A_0)', \mathrm{vech}(K_0)')'$ by QMLE. Under assumptions **A1** and **A3**, $\lambda_0$ is an element of a parameter space

$$\Lambda \subset \{(\mathrm{vech}(A)', \mathrm{vech}(K)')' : A > 0, K > 0, \rho(I - K) < 1\}. \tag{3.10}$$

We make the additional assumption

> **A4** : The unknown parameter $\lambda_0$ is an interior point of $\Lambda$, and $\Lambda$ is compact.

The sample covariance matrix yields a consistent estimator of $\gamma_0$ via

$$\hat{\gamma}_n := \mathrm{vech}(\hat{\Gamma}_n), \tag{3.11}$$

where

$$\hat{\Gamma}_n := n^{-1} \sum_{t=1}^{n} \varepsilon_t \varepsilon_t'.$$

Since the unknown parameter $\gamma_0$ is replaced by the half-vectorized sample variance $\hat{\gamma}_n$, the variance targeting version of the Gaussian quasi-likelihood function is

$$L_n(\lambda) = \prod_{t=1}^{n} |\tilde{\Sigma}_{t,n}|^{-1/2} \exp(-\frac{1}{2} \varepsilon_t' \tilde{\Sigma}_{t,n}^{-1} \varepsilon_t),$$

where $\tilde{\Sigma}_{t,n} := \tilde{\Sigma}_{t,n}(\lambda)$ is defined by

$$\mathrm{vech}(\tilde{\Sigma}_{t,n}) := \tilde{\sigma}_{t,n} := \tilde{\sigma}_{t,n}(\lambda) := K\hat{\gamma}_n + A s_{t-1} + (I - K - A)\tilde{\sigma}_{t-1,n}^2, \tag{3.12}$$

given initial values $\varepsilon_0$ and $\tilde{\sigma}_{0,n}^2 = \tilde{\sigma}_0^2$. The VTE of $\lambda_0 = (\mathrm{vech}(A_0)', \mathrm{vech}(K_0)')'$ is defined by

$$\hat{\lambda}_n := \underset{\lambda \in \Lambda}{\mathrm{argmin}} \tilde{I}_n(\lambda), \tag{3.13}$$

where

$$\tilde{I}_n(\lambda) := \frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_{t,n}, \tag{3.14}$$

and

$$\tilde{\ell}_{t,n} := \tilde{\ell}_{t,n}(\lambda) := \log(\det \tilde{\Sigma}_{t,n}) + \varepsilon_t' \tilde{\Sigma}_{t,n}^{-1}\varepsilon_t. \tag{3.15}$$

The VTE of $\theta_0 = (\gamma_0', \mathrm{vech}(A_0)', \mathrm{vech}(K_0)')'$ is defined by

$$\hat{\theta}_n = (\hat{\gamma}_n', \hat{\lambda}_n')'. \tag{3.16}$$

Comparing (3.16) above with (3.8), we see that Gaussian QML estimation involves maximizing a function of $[d(d+1)/2]^2 + d(d+1)$ variables. By reparametrizing the volatility equation, and estimating the long-run variance with the sample variance, the VTE method reduces the number of parameters that must be estimated to $[d(d+1)]/2]^2 + d(d+1)/2$. In section 3.3 we compare the estimation times of these two methods.

### 3.2.3 Asymptotic Properties of the VTE

#### 3.2.3.1 Strong Consistency

The results of Comte and Liebermann (2003) imply the strong consistency of the QMLE, $\hat{\theta}_n^*$, but these results do not directly imply consistency of the VTE, $\hat{\theta}_n$, because the VTE is a two-step estimator and cannot be expressed as a function of the QMLE. A detailed proof of the following result can be found in section 3.4.

**Theorem 8** *Under assumptions* **A1 − A4***, the VTE satisfies*

$$\hat{\theta}_n \xrightarrow{a.s} \theta_0$$

*as* $n \to \infty$.

Our proof of strong consistency is analogous to that given by Francq, Horváth, and Zakoïan (2009) in the univariate case, though the multivariate framework introduces additional complexity.

#### 3.2.3.2 Asymptotic Normality

From (3.11) and (3.16), we see that the first component of $\hat{\theta}_n$ is the vech of a sample covariance matrix. Thus, for any positive integer $n$, the VTE, $\hat{\theta}_n$, is an element of the product space of half-vectorized positive definite matrices with the parameter space $\Lambda$ defined in (3.10), and so we may consider the VTE as an element of the parameter space

$$\Theta \subset \mathrm{vech}(\mathcal{M}_d^+) \times \Lambda.$$

In order to establish asymptotic normality of the VTE, we need the additional assumptions:

> **A5** : The unknown parameter $\theta_0$ is an interior point of $\Theta$, and $\Theta$ is compact.
>
> **A6** : $\mathbb{E}||\varepsilon_t||^4 < \infty$.

Assumption **A6** is necessary for asymptotic normality of the sample variance, hence also for the VTE. Assumption **A6** is an improvement to the sixth order moment assumption in the existing literature for asymptotic normality of the Gaussian QMLE, and we prove in Theorem 10 that asymptotic normality of the Gaussian QMLE follows from an argument similar to that which proves asymptotic normality of the VTE. We now state our main result.

**Theorem 9** *Under assumptions* **A1** – **A6**, *the VTE satifies*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, RJR)$$

*as* $n \to \infty$.

The matrix $J$ is given by

$$J := \mathbb{E}\left(\frac{\partial}{\partial \lambda}\tilde{\ell}_t(\theta_0)\frac{\partial}{\partial \lambda'}\tilde{\ell}_t(\theta_0)\right).$$

The matrix $R$ is the left inverse of

$$\tilde{R}_0 := \mathbb{E}\left(\frac{\partial^2}{\partial\theta\partial\lambda}\tilde{\ell}_t(\theta_0)\right).$$

The matrix $\tilde{R}_0$ is the lower right $k \times \ell$ block of

$$\hat{R} := \mathbb{E}\left(\frac{\partial^2}{\partial\theta\partial\theta'}\tilde{\ell}_t(\theta_0)\right). \tag{3.17}$$

**Theorem 10** *Under assumptions* **A1 − A6***, the Gaussian QMLE satifies*

$$\sqrt{n}(\hat{\theta}_n^* - \theta_0) \xrightarrow{d} N(0, \hat{R}^{-1}\hat{J}\hat{R}^{-1})$$

*as* $n \to \infty$*.*

The matrix $\hat{R}$ is defined as in (3.17), and the matrix $\hat{J}$ is given by

$$\hat{J} := \mathbb{E}\left(\frac{\partial}{\partial\theta}\tilde{\ell}_t(\theta_0)\frac{\partial}{\partial\theta'}\tilde{\ell}_t(\theta_0)\right).$$

## 3.3 Simulation Studies

In this section, we compare the performance, and runtime, of the VTE and QMLE estimation methods. We simulated bivariate ARCH(1) processes, with $\rho(A_0)$ taking values 0.3, 0.55, and 0.9, using a method outlined in Francq and Zakoïan (2009). Tables 1-9 detail the sampling distribution of the estimators, based on 100 iterations of ARCH(1) processes of length $n$=500, 5,000, and 10,000, and Table 10 gives an empirical comparison of the runtime of both methods.

Our methods for simulation, estimation by QMLE, and estimation by VTE were implemented in the R statistical environment, version 2.11. These functions are in process for submission to CRAN for public distribution.

Simulation studies have found QMLE estimation to outperform other methods, particularly in the case of normal innovations, see Brooks (2008), Piontek (2004). For this reason, we compare the performance of the VTE against the QMLE for ARCH(1) models with normal innovations. As expected, the VTE shows slightly higher variance, particularly in the terms involving the intercept, since these terms are influenced by error in the sample variance as well as error on the QMLE. However, the overall performance of the VTE is very comparable to that of the QMLE, and the VTE occasionally outperforms the QMLE.

Table 1: Sampling distribution of the QMLE and VTE for ARCH(1) models with n=500, $\eta_t$ standard normal, vech$(A_0)$=(0.3, 0.0, 0.0, 0.3, 0.0, 0.3), $\mathcal{C}_0$=(1.0, 0.0, 1.0).

| true value | estimator | mean | RMSE | min | Q1 | Q2 | Q3 | max | sample variance |
|---|---|---|---|---|---|---|---|---|---|
| 0.3 | QMLE | 0.382 | 0.102 | 0.166 | 0.349 | 0.381 | 0.409 | 0.599 | 0.004 |
| | VTE | 0.365 | 0.103 | 0.131 | 0.320 | 0.364 | 0.424 | 0.561 | 0.006 |
| 0.0 | QMLE | 0.003 | 0.030 | -0.073 | -0.008 | 0.002 | 0.012 | 0.067 | 0.000 |
| | VTE | 0.003 | 0.028 | -0.072 | -0.016 | 0.003 | 0.021 | 0.063 | 0.001 |
| 0.0 | QMLE | 0.001 | 0.004 | 0.000 | 0.000 | 0.000 | 0.001 | 0.012 | 0.000 |
| | VTE | 0.002 | 0.003 | 0.000 | 0.000 | 0.001 | 0.003 | 0.012 | 0.000 |
| 0.3 | QMLE | 0.272 | 0.039 | 0.201 | 0.263 | 0.273 | 0.286 | 0.371 | 0.001 |
| | VTE | 0.266 | 0.054 | 0.161 | 0.236 | 0.267 | 0.297 | 0.359 | 0.002 |
| 0.0 | QMLE | 0.004 | 0.042 | -0.085 | -0.010 | 0.003 | 0.016 | 0.092 | 0.001 |
| | VTE | 0.005 | 0.040 | -0.087 | -0.022 | 0.005 | 0.029 | 0.099 | 0.002 |
| 0.3 | QMLE | 0.196 | 0.110 | 0.102 | 0.180 | 0.194 | 0.211 | 0.324 | 0.001 |
| | VTE | 0.198 | 0.116 | 0.082 | 0.160 | 0.191 | 0.228 | 0.371 | 0.003 |
| 1.0 | QMLE | 0.996 | 0.047 | 0.852 | 0.984 | 1.001 | 1.011 | 1.102 | 0.001 |
| | VTE | 0.982 | 0.091 | 0.722 | 0.916 | 0.994 | 1.048 | 1.206 | 0.008 |
| 0.0 | QMLE | -0.002 | 0.032 | -0.096 | -0.013 | -0.001 | 0.009 | 0.105 | 0.001 |
| | VTE | -0.006 | 0.064 | -0.189 | -0.037 | -0.010 | 0.023 | 0.206 | 0.004 |
| 1.0 | QMLE | 1.009 | 0.043 | 0.908 | 0.998 | 1.010 | 1.025 | 1.102 | 0.001 |
| | VTE | 1.026 | 0.091 | 0.818 | 0.961 | 1.021 | 1.080 | 1.288 | 0.008 |

Table 2: Sampling distribution of the QMLE and VTE for ARCH(1) models with $\eta_t$ standard normal, n=500, vech($A_0$)=(0.55, 0.0, 0.0, 0.55, 0.0, 0.55), $\mathcal{C}_0$=(1.0, 0.0, 1.0).

| true value | estimator | mean | RMSE | min | Q1 | Q2 | Q3 | max | sample variance |
|---|---|---|---|---|---|---|---|---|---|
| 0.55 | QMLE | 0.788 | 0.262 | 0.492 | 0.708 | 0.776 | 0.870 | 1.079 | 0.012 |
| | VTE | 0.640 | 0.118 | 0.427 | 0.596 | 0.632 | 0.681 | 0.879 | 0.006 |
| 0.0 | QMLE | 0.002 | 0.030 | -0.102 | -0.019 | 0.003 | 0.019 | 0.091 | 0.001 |
| | VTE | 0.002 | 0.025 | -0.082 | -0.013 | -0.001 | 0.018 | 0.077 | 0.001 |
| 0.0 | QMLE | 0.001 | 0.002 | 0.000 | 0.000 | 0.000 | 0.001 | 0.012 | 0.000 |
| | VTE | 0.001 | 0.002 | 0.000 | 0.000 | 0.000 | 0.001 | 0.010 | 0.000 |
| 0.55 | QMLE | 0.431 | 0.133 | 0.272 | 0.395 | 0.432 | 0.469 | 0.582 | 0.004 |
| | VTE | 0.390 | 0.168 | 0.267 | 0.356 | 0.389 | 0.413 | 0.544 | 0.003 |
| 0.0 | QMLE | 0.002 | 0.032 | -0.093 | -0.018 | 0.003 | 0.023 | 0.107 | 0.001 |
| | VTE | 0.003 | 0.031 | -0.085 | -0.015 | -0.001 | 0.022 | 0.116 | 0.001 |
| 0.55 | QMLE | 0.239 | 0.316 | 0.107 | 0.204 | 0.233 | 0.273 | 0.356 | 0.003 |
| | VTE | 0.241 | 0.314 | 0.108 | 0.202 | 0.234 | 0.274 | 0.370 | 0.003 |
| 1.0 | QMLE | 0.997 | 0.058 | 0.882 | 0.952 | 0.992 | 1.040 | 1.128 | 0.003 |
| | VTE | 1.016 | 0.112 | 0.798 | 0.943 | 1.003 | 1.099 | 1.284 | 0.012 |
| 0.0 | QMLE | -0.008 | 0.039 | -0.113 | -0.038 | -0.013 | 0.018 | 0.096 | 0.001 |
| | VTE | -0.017 | 0.087 | -0.325 | -0.073 | -0.029 | 0.039 | 0.211 | 0.007 |
| 1.0 | QMLE | 1.041 | 0.057 | 0.940 | 1.017 | 1.042 | 1.067 | 1.164 | 0.002 |
| | VTE | 1.137 | 0.167 | 0.886 | 1.072 | 1.135 | 1.213 | 1.393 | 0.009 |

Table 3: Sampling distribution of the QMLE and VTE for ARCH(1) models with $\eta_t$ standard normal, n=500, vech($A_0$)=(0.9, 0.0, 0.0, 0.9, 0.0, 0.9), $\mathcal{C}_0$=(1.0, 0.0, 1.0).

| true value | estimator | mean | RMSE | min | Q1 | Q2 | Q3 | max | sample variance |
|---|---|---|---|---|---|---|---|---|---|
| 0.9 | QMLE | 1.342 | 0.474 | 0.960 | 1.213 | 1.319 | 1.453 | 1.868 | 0.032 |
| | VTE | 0.857 | 0.077 | 0.713 | 0.817 | 0.856 | 0.900 | 0.993 | 0.004 |
| 0.0 | QMLE | -0.001 | 0.027 | -0.080 | -0.017 | 0.000 | 0.015 | 0.072 | 0.001 |
| | VTE | 0.000 | 0.016 | -0.051 | -0.010 | 0.000 | 0.009 | 0.050 | 0.000 |
| 0.0 | QMLE | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.001 | 0.004 | 0.000 |
| | VTE | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.003 | 0.000 |
| 0.9 | QMLE | 0.550 | 0.356 | 0.338 | 0.508 | 0.559 | 0.583 | 0.678 | 0.004 |
| | VTE | 0.426 | 0.478 | 0.244 | 0.397 | 0.429 | 0.466 | 0.553 | 0.003 |
| 0.0 | QMLE | 0.000 | 0.022 | -0.066 | -0.013 | 0.000 | 0.012 | 0.049 | 0.000 |
| | VTE | 0.000 | 0.016 | -0.050 | -0.009 | 0.000 | 0.010 | 0.051 | 0.000 |
| 0.9 | QMLE | 0.229 | 0.673 | 0.092 | 0.199 | 0.224 | 0.262 | 0.337 | 0.002 |
| | VTE | 0.215 | 0.686 | 0.069 | 0.182 | 0.211 | 0.252 | 0.345 | 0.003 |
| 1.0 | QMLE | 1.005 | 0.064 | 0.760 | 0.965 | 1.000 | 1.051 | 1.176 | 0.004 |
| | VTE | 1.213 | 0.253 | 0.817 | 1.111 | 1.199 | 1.308 | 1.676 | 0.022 |
| 0.0 | QMLE | 0.000 | 0.303 | -0.134 | -0.024 | -0.002 | 0.023 | 0.087 | 0.002 |
| | VTE | 0.004 | 0.312 | -0.376 | -0.056 | 0.000 | 0.071 | 0.289 | 0.010 |
| 1.0 | QMLE | 1.081 | 0.783 | 0.958 | 1.054 | 1.083 | 1.110 | 1.223 | 0.002 |
| | VTE | 1.316 | 1.028 | 0.965 | 1.215 | 1.293 | 1.386 | 2.011 | 0.026 |

Table 4: Sampling distribution of the QMLE and VTE for ARCH(1) models with $\eta_t$ standard normal, n=5,000, vech($A_0$)=(0.3, 0.0, 0.0, 0.3, 0.0, 0.3), $\mathcal{C}_0$=(1.0, 0.0, 1.0).

| true value | estimator | mean | RMSE | min | Q1 | Q2 | Q3 | max | sample variance |
|---|---|---|---|---|---|---|---|---|---|
| 0.3 | QMLE | 0.379 | 0.083 | 0.316 | 0.360 | 0.379 | 0.397 | 0.443 | 0.001 |
| | VTE | 0.363 | 0.068 | 0.300 | 0.344 | 0.363 | 0.379 | 0.423 | 0.001 |
| 0.0 | QMLE | -0.002 | 0.007 | -0.018 | -0.007 | -0.003 | 0.004 | 0.017 | 0.000 |
| | VTE | -0.002 | 0.007 | -0.017 | -0.007 | -0.003 | 0.003 | 0.015 | 0.000 |
| 0.0 | QMLE | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 |
| | VTE | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 |
| 0.3 | QMLE | 0.274 | 0.029 | 0.249 | 0.266 | 0.275 | 0.285 | 0.307 | 0.000 |
| | VTE | 0.271 | 0.032 | 0.245 | 0.262 | 0.272 | 0.281 | 0.302 | 0.000 |
| 0.0 | QMLE | -0.003 | 0.010 | -0.028 | -0.010 | -0.005 | 0.005 | 0.024 | 0.000 |
| | VTE | -0.003 | 0.010 | -0.028 | -0.010 | -0.004 | 0.004 | 0.022 | 0.000 |
| 0.3 | QMLE | 0.199 | 0.102 | 0.163 | 0.189 | 0.198 | 0.211 | 0.234 | 0.000 |
| | VTE | 0.203 | 0.099 | 0.165 | 0.193 | 0.201 | 0.216 | 0.238 | 0.000 |
| 1.0 | QMLE | 0.997 | 0.016 | 0.963 | 0.983 | 0.997 | 1.004 | 1.033 | 0.000 |
| | VTE | 0.987 | 0.033 | 0.920 | 0.961 | 0.987 | 1.005 | 1.059 | 0.001 |
| 0.0 | QMLE | 0.000 | 0.009 | -0.017 | -0.004 | 0.001 | 0.006 | 0.016 | 0.000 |
| | VTE | 0.000 | 0.017 | -0.035 | -0.012 | 0.002 | 0.012 | 0.032 | 0.000 |
| 1.0 | QMLE | 1.011 | 0.017 | 0.977 | 1.004 | 1.012 | 1.020 | 1.035 | 0.000 |
| | VTE | 1.032 | 0.042 | 0.963 | 1.014 | 1.037 | 1.052 | 1.078 | 0.001 |

Table 5: Sampling distribution of the QMLE and VTE for ARCH(1) models with $\eta_t$ standard normal, n=5,000, vech($A_0$)=(0.55, 0.0, 0.0, 0.55, 0.0, 0.55), $\mathcal{C}_0$=(1.0, 0.0, 1.0).

| true value | estimator | mean | RMSE | min | Q1 | Q2 | Q3 | max | sample variance |
|---|---|---|---|---|---|---|---|---|---|
| 0.55 | QMLE | 0.786 | 0.239 | 0.699 | 0.754 | 0.784 | 0.811 | 0.906 | 0.002 |
| | VTE | 0.660 | 0.115 | 0.583 | 0.633 | 0.658 | 0.694 | 0.735 | 0.001 |
| 0.0 | QMLE | 0.003 | 0.011 | -0.014 | -0.006 | 0.004 | 0.011 | 0.025 | 0.000 |
| | VTE | 0.002 | 0.008 | -0.015 | -0.005 | 0.004 | 0.008 | 0.017 | 0.000 |
| 0.0 | QMLE | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 |
| | VTE | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 0.55 | QMLE | 0.431 | 0.133 | 0.272 | 0.395 | 0.432 | 0.469 | 0.582 | 0.004 |
| | VTE | 0.390 | 0.168 | 0.267 | 0.356 | 0.389 | 0.413 | 0.544 | 0.003 |
| 0.0 | QMLE | 0.003 | 0.012 | -0.017 | -0.006 | 0.004 | 0.012 | 0.028 | 0.000 |
| | VTE | 0.003 | 0.010 | -0.018 | -0.006 | 0.005 | 0.010 | 0.023 | 0.000 |
| 0.55 | QMLE | 0.235 | 0.315 | 0.202 | 0.222 | 0.235 | 0.247 | 0.273 | 0.000 |
| | VTE | 0.242 | 0.309 | 0.203 | 0.228 | 0.242 | 0.257 | 0.276 | 0.000 |
| 1.0 | QMLE | 0.994 | 0.019 | 0.956 | 0.980 | 0.994 | 1.007 | 1.058 | 0.000 |
| | VTE | 1.003 | 0.035 | 0.937 | 0.978 | 1.005 | 1.025 | 1.129 | 0.001 |
| 0.0 | QMLE | 0.000 | 0.011 | -0.026 | -0.008 | 0.000 | 0.009 | 0.030 | 0.000 |
| | VTE | -0.001 | 0.025 | -0.057 | -0.017 | 0.001 | 0.016 | 0.058 | 0.001 |
| 1.0 | QMLE | 1.038 | 0.041 | 0.990 | 1.026 | 1.038 | 1.048 | 1.065 | 0.000 |
| | VTE | 1.127 | 0.133 | 1.012 | 1.102 | 1.125 | 1.151 | 1.214 | 0.001 |

Table 6: Sampling distribution of the QMLE and VTE for ARCH(1) models with $\eta_t$ standard normal, n=5,000, vech$(A_0)$=(0.9, 0.0, 0.0, 0.9, 0.0, 0.9), $\mathcal{C}_0$=(1.0, 0.0, 1.0).

| true value | estimator | mean | RMSE | min | Q1 | Q2 | Q3 | max | sample variance |
|---|---|---|---|---|---|---|---|---|---|
| 0.9 | QMLE | 1.337 | 0.440 | 1.248 | 1.300 | 1.331 | 1.378 | 1.486 | 0.003 |
| | VTE | 0.907 | 0.035 | 0.851 | 0.880 | 0.904 | 0.935 | 0.986 | 0.001 |
| 0.0 | QMLE | -0.001 | 0.008 | -0.021 | -0.006 | 0.000 | 0.004 | 0.015 | 0.000 |
| | VTE | 0.000 | 0.016 | -0.051 | -0.010 | 0.000 | 0.009 | 0.050 | 0.000 |
| 0.0 | QMLE | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | VTE | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 0.9 | QMLE | 0.549 | 0.352 | 0.476 | 0.532 | 0.546 | 0.565 | 0.607 | 0.001 |
| | VTE | 0.448 | 0.452 | 0.364 | 0.433 | 0.446 | 0.471 | 0.496 | 0.001 |
| 0.0 | QMLE | -0.001 | 0.007 | -0.018 | -0.005 | 0.000 | 0.004 | 0.012 | 0.000 |
| | VTE | -0.001 | 0.004 | -0.012 | -0.004 | -0.001 | 0.001 | 0.005 | 0.000 |
| 0.9 | QMLE | 0.226 | 0.674 | 0.170 | 0.216 | 0.226 | 0.239 | 0.266 | 0.000 |
| | VTE | 0.222 | 0.678 | 0.152 | 0.209 | 0.220 | 0.244 | 0.256 | 0.001 |
| 1.0 | QMLE | 1.009 | 0.024 | 0.971 | 0.994 | 1.008 | 1.020 | 1.080 | 0.001 |
| | VTE | 1.209 | 0.216 | 1.103 | 1.169 | 1.200 | 1.242 | 1.380 | 0.003 |
| 0.0 | QMLE | 0.001 | 0.012 | -0.032 | -0.008 | -0.001 | 0.011 | 0.026 | 0.000 |
| | VTE | 0.004 | 0.033 | -0.084 | -0.017 | 0.007 | 0.029 | 0.058 | 0.001 |
| 1.0 | QMLE | 1.073 | 0.075 | 1.046 | 1.061 | 1.072 | 1.085 | 1.102 | 0.000 |
| | VTE | 1.287 | 0.291 | 1.191 | 1.261 | 1.287 | 1.318 | 1.385 0.002 | |

Table 7: Sampling distribution of the QMLE and VTE for ARCH(1) models with $\eta_t$ standard normal, n=10,000, vech($A_0$)=(0.3, 0.0, 0.0, 0.3, 0.0, 0.3), $\mathcal{C}_0$=(1.0, 0.0, 1.0).

| true value | estimator | mean | RMSE | min | Q1 | Q2 | Q3 | max | sample variance |
|---|---|---|---|---|---|---|---|---|---|
| 0.3 | QMLE | 0.381 | 0.083 | 0.340 | 0.366 | 0.382 | 0.395 | 0.420 | 0.000 |
| | VTE | 0.363 | 0.065 | 0.326 | 0.348 | 0.364 | 0.375 | 0.396 | 0.000 |
| 0.0 | QMLE | 0.000 | 0.006 | -0.017 | -0.004 | 0.000 | 0.002 | 0.015 | 0.000 |
| | VTE | -0.001 | 0.006 | -0.018 | -0.004 | 0.000 | 0.003 | 0.013 | 0.000 |
| 0.0 | QMLE | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 |
| | VTE | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 |
| 0.3 | QMLE | 0.275 | 0.027 | 0.250 | 0.266 | 0.276 | 0.283 | 0.297 | 0.000 |
| | VTE | 0.271 | 0.031 | 0.247 | 0.264 | 0.273 | 0.279 | 0.292 | 0.000 |
| 0.0 | QMLE | -0.001 | 0.009 | -0.025 | -0.006 | 0.000 | 0.003 | 0.022 | 0.000 |
| | VTE | -0.001 | 0.009 | -0.026 | -0.007 | 0.000 | 0.004 | 0.020 | 0.000 |
| 0.3 | QMLE | 0.200 | 0.101 | 0.169 | 0.192 | 0.201 | 0.206 | 0.224 | 0.000 |
| | VTE | 0.203 | 0.097 | 0.180 | 0.195 | 0.205 | 0.211 | 0.229 | 0.000 |
| 1.0 | QMLE | 0.999 | 0.010 | 0.975 | 0.992 | 0.999 | 1.005 | 1.023 | 0.000 |
| | VTE | 0.991 | 0.022 | 0.941 | 0.978 | 0.992 | 1.004 | 1.023 | 0.000 |
| 0.0 | QMLE | 0.000 | 0.008 | -0.016 | -0.005 | 0.001 | 0.006 | 0.019 | 0.000 |
| | VTE | 0.001 | 0.016 | -0.031 | -0.008 | 0.003 | 0.012 | 0.040 | 0.000 |
| 1.0 | QMLE | 1.011 | 0.014 | 0.983 | 1.005 | 1.010 | 1.017 | 1.030 | 0.000 |
| | VTE | 1.029 | 0.035 | 0.970 | 1.016 | 1.029 | 1.046 | 1.060 | 0.000 |

Table 8: Sampling distribution of the QMLE and VTE for ARCH(1) models with $\eta_t$ standard normal, n=10,000, vech($A_0$)=(0.55, 0.0, 0.0, 0.55, 0.0, 0.55), $\mathcal{C}_0$=(1.0, 0.0, 1.0).

| true value | estimator | mean | RMSE | min | Q1 | Q2 | Q3 | max | sample variance |
|---|---|---|---|---|---|---|---|---|---|
| 0.55 | QMLE | 0.729 | 0.180 | 0.689 | 0.711 | 0.723 | 0.743 | 0.780 | 0.001 |
| | VTE | 0.631 | 0.084 | 0.592 | 0.617 | 0.627 | 0.650 | 0.666 | 0.000 |
| 0.0 | QMLE | 0.000 | 0.007 | -0.009 | -0.005 | 0.000 | 0.005 | 0.020 | 0.000 |
| | VTE | 0.000 | 0.005 | -0.007 | -0.004 | -0.001 | 0.003 | 0.016 | 0.000 |
| 0.0 | QMLE | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 |
| | VTE | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 0.55 | QMLE | 0.480 | 0.071 | 0.462 | 0.470 | 0.482 | 0.487 | 0.495 | 0.000 |
| | VTE | 0.390 | 0.168 | 0.267 | 0.356 | 0.389 | 0.413 | 0.544 | 0.003 |
| 0.0 | QMLE | 0.000 | 0.007 | -0.010 | -0.006 | 0.001 | 0.005 | 0.021 | 0.000 |
| | VTE | 0.000 | 0.007 | -0.008 | -0.005 | -0.001 | 0.004 | 0.019 | 0.000 |
| 0.55 | QMLE | 0.287 | 0.263 | 0.268 | 0.280 | 0.286 | 0.296 | 0.309 | 0.000 |
| | VTE | 0.294 | 0.257 | 0.276 | 0.286 | 0.291 | 0.303 | 0.323 | 0.000 |
| 1.0 | QMLE | 0.996 | 0.009 | 0.973 | 0.991 | 0.996 | 1.002 | 1.009 | 0.000 |
| | VTE | 1.007 | 0.018 | 0.963 | 0.996 | 1.010 | 1.020 | 1.032 | 0.000 |
| 0.0 | QMLE | 0.001 | 0.009 | -0.017 | -0.003 | 0.000 | 0.007 | 0.021 | 0.000 |
| | VTE | 0.002 | 0.019 | -0.038 | -0.008 | 0.000 | 0.013 | 0.044 | 0.000 |
| 1.0 | QMLE | 1.038 | 0.039 | 1.023 | 1.030 | 1.037 | 1.043 | 1.068 | 0.000 |
| | VTE | 1.123 | 0.125 | 1.090 | 1.106 | 1.121 | 1.132 | 1.193 | 0.001 |

Table 9: Sampling distribution of the QMLE and VTE for ARCH(1) models with $\eta_t$ standard normal, n=10,000, vech($A_0$)=(0.9, 0.0, 0.0, 0.9, 0.0, 0.9), $\mathcal{C}_0$=(1.0, 0.0, 1.0).

| true value | estimator | mean | RMSE | min | Q1 | Q2 | Q3 | max | sample variance |
|---|---|---|---|---|---|---|---|---|---|
| 0.9 | QMLE | 1.241 | 0.343 | 1.156 | 1.215 | 1.242 | 1.261 | 1.313 | 0.001 |
| | VTE | 0.903 | 0.034 | 0.847 | 0.879 | 0.896 | 0.921 | 0.991 | 0.001 |
| 0.0 | QMLE | 0.000 | 0.004 | -0.008 | -0.003 | -0.001 | 0.003 | 0.007 | 0.000 |
| | VTE | 0.001 | 0.004 | -0.005 | -0.002 | 0.000 | 0.002 | 0.014 | 0.000 |
| 0.0 | QMLE | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | VTE | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 0.9 | QMLE | 0.653 | 0.248 | 0.628 | 0.642 | 0.652 | 0.658 | 0.692 | 0.000 |
| | VTE | 0.551 | 0.349 | 0.514 | 0.545 | 0.550 | 0.556 | 0.581 | 0.000 |
| 0.0 | QMLE | 0.000 | 0.003 | -0.007 | -0.002 | -0.001 | 0.002 | 0.006 | 0.000 |
| | VTE | 0.001 | 0.004 | -0.005 | -0.002 | 0.000 | 0.002 | 0.013 | 0.000 |
| 0.9 | QMLE | 0.328 | 0.572 | 0.308 | 0.321 | 0.330 | 0.333 | 0.356 | 0.000 |
| | VTE | 0.326 | 0.574 | 0.295 | 0.317 | 0.329 | 0.334 | 0.347 | 0.000 |
| 1.0 | QMLE | 1.009 | 0.024 | 0.971 | 0.994 | 1.008 | 1.020 | 1.080 | 0.001 |
| | VTE | 1.209 | 0.216 | 1.103 | 1.169 | 1.200 | 1.242 | 1.380 | 0.003 |
| 0.0 | QMLE | 0.001 | 0.012 | -0.032 | -0.008 | -0.001 | 0.011 | 0.026 | 0.000 |
| | VTE | 0.004 | 0.033 | -0.084 | -0.017 | 0.007 | 0.029 | 0.058 | 0.001 |
| 1.0 | QMLE | 1.006 | 0.017 | 0.982 | 0.994 | 1.004 | 1.025 | 1.031 | 0.000 |
| | VTE | 1.196 | 0.198 | 1.142 | 1.177 | 1.203 | 1.211 | 1.278 | 0.001 |

In Table 10 we compare the runtime of the VTE and QMLE estimation methods. The times shown represent elapsed time (also known as wall time) as measured in seconds on a 2.66 GHz Intel Core i7 processor. These simulations involved trajectories of length $n$=500, 5,000, and 10,000, and 150 iterations of bivariate ARCH(1) models, with 50 iterations each corresponding to values 0.3, 0.55, and 0.9 for $\rho(A_0)$.

Table 10: Runtime of the QMLE and VTE .

| | VTE | | QMLE | |
|---|---|---|---|---|
| $n$ | mean | median | mean | median |
| 500 | 11.10 | 11.61 | 17.60 | 17.80 |
| 5,000 | 121.11 | 122.51 | 190.53 | 188.90 |
| 10,000 | 249.37 | 267.76 | 365.37 | 373.45 |

## 3.4 Technical Proofs

### 3.4.1 Proof of Strong Consistency in Theorem 8

For any $\lambda \in \Lambda$, one can define the strictly stationary and ergodic process

$$\sigma_t^2 := \sigma_t^2(\lambda) := K\gamma_0 + As_{t-1} + (I - K - A)\sigma_{t-1}^2(\lambda). \tag{3.18}$$

Let $\Sigma_t$ be defined by $\text{vech}(\Sigma_t) := \sigma_t^2$ and, analogous to (3.14) and (3.15), define

$$I_n(\lambda) := \frac{1}{n}\sum_{t=1}^{n} \ell_t, \tag{3.19}$$

where

$$\ell_t := \ell_t(\lambda) := \log(\det \Sigma_t) + \varepsilon_t' \Sigma_t^{-1} \varepsilon_t. \tag{3.20}$$

It follows from the ergodic theorem that the half-vectorized sample variance $\hat{\gamma}_n$ converges almost surely to $\gamma_0$. We will prove that the VTE is strong consistent by establishing the following intermediate results.

(i) $\lim_{n\to\infty} \sup_{\lambda\in\Lambda} |I_n(\lambda) - \tilde{I}_n(\lambda)| = 0$ almost surely,

(ii) if $\sigma_t^2(\lambda) = \sigma_t^2(\lambda_0)$ almost surely, then $\lambda = \lambda_0$,

(iii) if $\lambda \neq \lambda_0$, then $\mathbb{E}\ell_t(\lambda) > \mathbb{E}\ell_t(\lambda_0)$,

(iv) any $\lambda \neq \lambda_0$ has a neighborhood $V(\lambda)$ such that $\liminf_{n\to\infty} \inf_{\lambda^*\in V(\lambda)} \tilde{I}_n(\lambda^*) > \mathbb{E}\ell_1(\lambda_0)$ almost surely.

#### 3.4.1.1 Asymptotic Irrelevance of the Initial Values

Recursive iteration of (3.12) and (3.18) yields, for any positive integer $N$ ,

$$\begin{aligned}
\tilde{\sigma}_{t,n}^2 - \sigma_t^2 &= K(\hat{\gamma}_n - \gamma_0) + B(\tilde{\sigma}_{t-1,n}^2 - \sigma_{t-1}^2) \\
&= K(\hat{\gamma}_n - \gamma_0) + B\left[\sum_{i=1}^{N} B^{i-1}K\hat{\gamma}_n + B^{N+1}\tilde{\sigma}_{t-N-2,n}^2\right] \\
&\quad - B\left[\sum_{i=1}^{N} B^{i-1}K\gamma_0 + B^{N+1}\sigma_{t-N-2}^2\right].
\end{aligned}$$

Taking $N = t - 2$, we have

$$\begin{aligned}
\tilde{\sigma}_{t,n}^2 - \sigma_t^2 &= K(\hat{\gamma}_n - \gamma_0) + B\left[\sum_{i=1}^{t-2} B^{i-1}K\hat{\gamma}_n + B^{t-1}\tilde{\sigma}_0^2\right] - B\left[\sum_{i=1}^{t-2} B^{i-1}K\gamma_0 + B^{t-1}\sigma_0^2\right] \\
&= K(\hat{\gamma}_n - \gamma_0) + \sum_{i=1}^{t-2} B^{i}K\hat{\gamma}_n + B^{t}\tilde{\sigma}_0^2 - \left[\sum_{i=1}^{t-2} B^{i}K\gamma_0 + B^{t}\sigma_0^2\right] \\
&= \sum_{i=0}^{t-2} B^{i}K\hat{\gamma}_n - \sum_{i=0}^{t-2} B^{i}K\gamma_0 + B^{t}\tilde{\sigma}_0^2 - B^{t}\sigma_0^2 \\
&= (I-B)^{-1}(I-B^{t-1})K(\hat{\gamma}_n - \gamma_0) + B^{t}(\tilde{\sigma}_0^2 - \sigma_0^2).
\end{aligned}$$

We proved in Lemma 4 of Chapter 2 that $\rho(I-K) = \rho(A+B) < 1$ implies $\rho(B) < 1$. Using this and compactness of $\Lambda$, there exists some finite positive random variable, say $M_1$, such that

$$||\tilde{\sigma}_{t,n}^2 - \sigma_t^2|| \leq M_1||\hat{\gamma}_n - \gamma_0|| + ||B||^t||\tilde{\sigma}_0^2 - \sigma_0^2||$$

holds uniformly for all $t > 0$.

The initial value $\tilde{\sigma}_0^2$ is a fixed constant, and $\sigma_0^2$ is a measurable function of $\{\varepsilon_u : u \leq 0\}$, so $\sigma_0^2$ depends on neither $n$ nor $t$ and may be considered as a fixed random variable, say $M_2$. Thus,

$$||\tilde{\sigma}_{t,n}^2 - \sigma_t^2|| \leq M_1||\hat{\gamma}_n - \gamma_0|| + M_2||B||^t. \tag{3.21}$$

Furthermore,

$$||\tilde{\sigma}_{t,n}^2 - \sigma_t^2||^2 \leq ||\tilde{\Sigma}_{t,n} - \Sigma_t||^2 \leq 2||\tilde{\sigma}_{t,n}^2 - \sigma_t^2||^2. \tag{3.22}$$

Since $||\hat{\gamma}_n - \gamma_0|| \to 0$ almost surely as $n \to \infty$ and $||B||^t \to 0$ as $t \to \infty$, it follows that,

$$||\tilde{\Sigma}_{t,n} - \Sigma_t|| \xrightarrow{a.s} 0 \tag{3.23}$$

as $n, t \to \infty$. Let $g$ denote the matrix inversion function on $\mathcal{M}_d$, and let $f$ be defined by $f(\cdot) := \log(\det(\cdot))$ on $\mathcal{M}_d$. Then $f$ and $g$ are continuous with respect to the Euclidean norm when restricted to the subspace of positive definite matrices, and

$$\sup_{\lambda\in\Lambda}\left|I_n(\lambda)-\tilde{I}_n(\lambda)\right| \leq \frac{1}{n}\sum_{t=1}^{n}\sup_{\lambda\in\Lambda}\left\{\left|f(\tilde{\Sigma}_{t,n}-f(\Sigma_t)\right|\right\}$$

$$+\frac{1}{n}\sum_{t=1}^{n}||\varepsilon_t||^2\sup_{\lambda\in\Lambda}\left|\left|g(\tilde{\Sigma}_{t,n})-g(\Sigma_t)\right|\right|. \tag{3.24}$$

It follows from (3.23) and continuity of the functions $f$ and $g$ that

$$\left|f(\tilde{\Sigma}_{t,n})-f(\Sigma_t)\right| \xrightarrow{a.s} 0$$

and

$$\left|\left|g(\tilde{\Sigma}_{t,n})-g(\Sigma_t)\right|\right| \xrightarrow{a.s} 0$$

as $n, t \to \infty$. The proof is complete by applying the Cesaro mean lemma to (3.24) if we can show that also

$$||\varepsilon_t||^2\left|\left|g(\tilde{\Sigma}_{t,n})-g(\Sigma_t)\right|\right| \to 0$$

almost surely as $n, t \to \infty$.

The space of real positive definite $d \times d$ matrices, $S_d^{++}$, is an open and convex subset of $\mathcal{M}_d$ according to Proposition 2.7 of Arsigny, Fillard, Pennec, and Ayache (2000). It follows from (3.23) that there exist positive integers $T$ and $N$, and some open neighborhood $V_{T,N}$ such that the closure of $V_{T,N}$ is contained in $S_d^{++}$, and such that $\tilde{\Sigma}_{t,n} \in V_{T,N}$, and $\Sigma_t \in V_{T,N}$ whenever $n > N$ and $t > T$. The function $g$ is smooth on $V_{T,N}$, and so the Jacobian of $g$ is bounded on $V_{T,N}$, i.e., for all $X \in V_{T,N}$ we have

$$||D\tilde{g}(X)|| \leq C_1$$

for some finite constant $C_1$. The mean value inequality yields

$$||g(\tilde{\Sigma}_{t,n})-g(\Sigma_t)|| \leq C_1||\Sigma_{t,n}-\Sigma_t||.$$

Combining the line above with (3.21) and (3.22), we have

$$\lim_{n\to\infty}||\varepsilon_t||^2\left|\left|g(\tilde{\Sigma}_{t,n})-g(\Sigma_t)\right|\right| \leq C_1M_2\,||\varepsilon_t||^2\,||B||^t,$$

which converges to zero almost surely as $t \to \infty$ by the Borel-Cantelli lemma, the Markov inequality, and stationarity of $\boldsymbol{\varepsilon}$ because, for any $\delta > 0$,

$$\sum_{t=1}^{\infty} \mathbb{P}(||\varepsilon_t||^2 ||B||^t > \delta) \leq \sum_{t=1}^{\infty} \frac{\mathbb{E}(||\varepsilon_t||^2)||B||^t}{\delta} < \infty.$$

□

#### 3.4.1.2 Identifiability of the Parameter

The idea for this proof is due to Berkes, Horváth, and Kokoszka (2003). We prove the claim by contradiction. Recursive iteration of (3.18) yields

$$\sigma_t^2(\lambda) = (I - B)^{-1} K \gamma_0 + \sum_{j=0}^{\infty} B^j A s_{t-j-1},$$

thus $\sigma_t^2(\lambda) = \sigma_t^2(\lambda_0)$ implies

$$(I - B)^{-1} K \gamma_0 + \sum_{j=0}^{\infty} B^j A s_{t-j-1} = (I - B_0)^{-1} K_0 \gamma_0 + \sum_{j=0}^{\infty} B_0^j A_0 s_{t-j-1}. \tag{3.25}$$

Let

$$D_0 := (I - B)^{-1} K \gamma_0, \qquad D_0^* := (I - B_0)^{-1} K_0 \gamma_0,$$

and for $i \in \{1, 2, \ldots\}$

$$D_i := B^{i-1} A, \quad D_i^* := B_0^{i-1} A_0.$$

Then we can write (3.25) as

$$D_0 + \sum_{j=1}^{\infty} D_j s_{t-j} = D_0^* + \sum_{j=1}^{\infty} D_j^* s_{t-j}.$$

If $D_i = D_i^*$ for all nonnegative integers $i$, then taking $i = 1$ gives $A = A_0$, and taking $i = 2$ gives $BA = B_0 A$. The matrix $A$ is positive-definite hence invertible, so this implies $B = B_0$ and $\lambda = \lambda_0$.

Suppose (by way of contradiction) that $m > 0$ is the smallest integer such that $D_m \neq D_m^*$ (if $D_j = D_j^*$ for all $j > 0$ then $D_0 = D_0^*$). Then,

$$(D_m - D_m^*) s_{t-m} = (D_0 - D_0^*) + \sum_{j=m+1}^{\infty} (D_j^* - D_j) s_{t-j}. \tag{3.26}$$

The right-hand side of (3.26) is $\mathcal{F}_{t-m-1}$ measurable, so $(D_m - D_m^*) s_{t-m}$ must also be $\mathcal{F}_{t-m-1}$ measurable, but we claim that this leads to a contradiction.

From (3.1), $s_{t-m} = \mathrm{vech}(H_{t-m}^{1/2}\eta_{t-m}\eta_{t-m}'H_{t-m}^{1/2})$. Denote by $h_{i,j}$ the $(i,j)$th entry of $H_{t-m}^{1/2}$, and denote the components of $\eta_{t-m}$ by $\eta_{t-m} = (n_1, \ldots, n_d)'$. The $(i,j)$th entry of $H_{t-m}^{1/2}\eta_{t-m}\eta_{t-m}'H_{t-m}^{1/2}$ is

$$\left(\sum_{\ell=1}^{d} n_\ell h_{i,\ell}\right)\left(\sum_{k=1}^{d} n_k h_{k,j}\right).$$

$H_{t-m}$ is positive definite under assumption **A1**, and we take $H_{t-m}^{1/2}$ as the unique positive definite square root of $H_{t-m}$, so the diagonal elements of $H_{t-m}^{1/2}$ are nonzero. Thus, each coordinate of $s_{t-m}$ is a nontrivial linear combination of the terms $n_\ell n_k$ $(1 \le \ell, k \le d)$. Since $D_m \neq D_m^*$, the vector $(D_m - D_m^*)s_{t-m}$ has at least one nonzero coordinate that is a linear combination of the terms $n_\ell n_k$ $(1 \le \ell, k \le d)$.

The sequence $(\eta_t)$ is i.i.d., and $(D_m - D_m^*)s_{t-m}$ is a measurable and almost surely nonconstant function of $\eta_{t-m}$, so it is not $\mathcal{F}_{t-m-1}$ measurable. This completes the proof of (ii).

$\square$

#### 3.4.1.3 The Limit Criterion is Minimized at the True Value

Using (3.20), using that $\Sigma_t(\lambda_0) = H_t$, and using that $\eta_t$ has unit covariance, we can write

$$\begin{aligned}
\mathbb{E}\ell_t(\lambda_0) &= \mathbb{E}(\log\det\Sigma_t(\lambda_0)) + \mathbb{E}(\varepsilon_t'\Sigma_t^{-1}(\lambda_0)\varepsilon_t) \\
&= \mathbb{E}(\log\det\Sigma_t(\lambda_0)) + \mathbb{E}(\varepsilon_t'H_t^{-1/2}H_t^{-1/2}\varepsilon_t) \\
&= \mathbb{E}(\log\det\Sigma_t(\lambda_0)) + \mathbb{E}(\eta_t'\eta_t) \\
&= \mathbb{E}(\log\det\Sigma_t(\lambda_0)) + d.
\end{aligned}$$

Thus, using that $\eta_t$ is independent of $\mathcal{F}_{t-1}$ while $H_t$ and $\Sigma_t$ are $\mathcal{F}_{t-1}$ measurable,

$$\begin{aligned}
\mathbb{E}\ell_t(\lambda) - \mathbb{E}\ell_t(\lambda_0) &= \mathbb{E}(\log\det\Sigma_t(\lambda) - \log\det\Sigma_t(\lambda_0)) + \mathbb{E}(\varepsilon_t'\Sigma_t^{-1}(\lambda)\varepsilon_t) - d \\
&= \mathbb{E}\log(\det\Sigma_t(\lambda)/\det\Sigma_t(\lambda_0)) + \mathbb{E}[\mathrm{tr}(\varepsilon_t\varepsilon_t'\Sigma_t^{-1}(\lambda))] - d \\
&= \mathbb{E}\log(\det\Sigma_t(\lambda)/\det\Sigma_t(\lambda_0)) + \mathbb{E}[\mathrm{tr}(\eta_t\eta_t'H_t^{1/2}\Sigma_t^{-1}(\lambda)H_t^{1/2})] - d
\end{aligned}$$

$$= \mathbb{E}\log(\det\Sigma_t(\lambda)/\det\Sigma_t(\lambda_0))$$
$$+\mathrm{tr}[\mathbb{E}(\eta_t\eta_t)\mathbb{E}('H_t^{1/2}\Sigma_t^{-1}(\lambda)H_t^{1/2})] - d$$
$$= \mathbb{E}\log(\det\Sigma_t(\lambda)/\det\Sigma_t(\lambda_0))$$
$$+\mathbb{E}[\mathrm{tr}(H_t^{1/2}\Sigma_t^{-1}(\lambda)H_t^{1/2})] - d. \tag{3.27}$$

Abadir and Magnus (2005) prove that if $X$ is a positive definite $d \times d$ matrix, then

$$\log\det X \le \mathrm{tr}(X) - d \tag{3.28}$$

with equality if and only if $X = I_d$. The matrix $H_t^{1/2}\Sigma_t^{-1}(\lambda)H_t^{1/2}$ is positive definite, so it follows from (3.27) and (3.28) that

$$\begin{aligned}\mathbb{E}\ell_t(\lambda) - \mathbb{E}\ell_t(\lambda_0) &\ge \mathbb{E}\log(\det\Sigma_t(\lambda)/\det\Sigma_t(\lambda_0)) + \mathbb{E}\log\det(H_t^{1/2}\Sigma_t^{-1}(\lambda)H_t^{1/2}) \\ &= \mathbb{E}\log(\det\Sigma_t(\lambda)/\det\Sigma_t(\lambda_0)) + \mathbb{E}\log(\det\Sigma_t(\lambda_0)/\det\Sigma_t(\lambda)) \\ &= 0.\end{aligned}$$

Equality holds above if and only if $H_t^{1/2}\Sigma_t^{-1}(\lambda)H_t^{1/2} = I_d$ almost surely, i.e., if and only if $\Sigma_t(\lambda) = \Sigma_t(\lambda_0) = H_t$ almost surely. We proved in part (ii) that $\Sigma_t(\lambda) = \Sigma_t(\lambda_0)$ almost surely implies $\lambda = \lambda_0$. Thus $\lambda \ne \lambda_0$ implies $\mathbb{E}\ell_t(\lambda) - \mathbb{E}\ell_t(\lambda_0) > 0$.

□

#### 3.4.1.4 Compactness of $\Lambda$ and Ergodicity of $\ell_t(\lambda)$

For any $\lambda \in \Lambda$, and any positive integer $k$, let $V_k(\lambda)$ denote intersection of the parameter space $\Lambda$ with the open ball of radius $1/k$ centered at $\lambda$. Then,

$$\begin{aligned}\liminf_{n\to\infty}\inf_{\lambda^*\in V(\lambda)}\tilde{I}_n(\lambda^*) &= \liminf_{n\to\infty}\inf_{\lambda^*\in V(\lambda)}\Big(I_n(\lambda^*) + \tilde{I}_n(\lambda^*) - I_n(\lambda^*)\Big) \\ &= \liminf_{n\to\infty}\left\{\inf_{\lambda^*\in V(\lambda)} I_n(\lambda^*) + \inf_{\lambda^*\in V(\lambda)}\Big(\tilde{I}_n(\lambda^*) - I_n(\lambda^*)\Big)\right\} \\ &= \liminf_{n\to\infty}\left\{\inf_{\lambda^*\in V(\lambda)} I_n(\lambda^*) - \sup_{\lambda^*\in V(\lambda)}\Big(I_n(\lambda^*) - \tilde{I}_n(\lambda^*)\Big)\right\} \\ &\ge \liminf_{n\to\infty}\left\{\inf_{\lambda^*\in V(\lambda)} I_n(\lambda^*)\right\} \\ &\qquad -\limsup_{n\to\infty}\left\{\sup_{\lambda^*\in V(\lambda)}\Big(I_n(\lambda^*) - \tilde{I}_n(\lambda^*)\Big)\right\}\end{aligned}$$

$$\geq \liminf_{n\to\infty} \inf_{\lambda^*\in V(\lambda)} I_n(\lambda^*)$$

$$-\limsup_{n\to\infty} \left\{ \sup_{\lambda^*\in V(\lambda)} \left| I_n(\lambda^*) - \tilde{I}_n(\lambda^*) \right| \right\}.$$

The last term above is zero by part (i). Using (3.19), the ergodic theorem, and part (ii), we have

$$\begin{aligned}
\liminf_{n\to\infty} \inf_{\lambda^*\in V(\lambda)} \tilde{I}_n(\lambda^*) &\geq \liminf_{n\to\infty} \inf_{\lambda^*\in V(\lambda)} I_n(\lambda^*) \\
&= \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \inf_{\lambda^*\in V(\lambda)} \ell_t(\lambda^*) \\
&= \mathbb{E}\left( \inf_{\lambda^*\in V(\lambda)} \ell_1(\lambda^*) \right) \\
&> \mathbb{E}\left(\ell_1(\lambda_0)\right)
\end{aligned}$$

for sufficiently large $k$ whenever $\lambda \neq \lambda_0$. This completes the proof of Theorem 8.

$\square$

### 3.4.2 Proof of Asymptotic Normality in Theorem 9

Under assumptions **A1-A6**, and for any $\theta \in \Theta$, the process defined as in (3.7) by,

$$\mathrm{vech}(\tilde{\Sigma}_t) = \tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\theta) := K\gamma + As_{t-1} + (I - K - A)\tilde{\sigma}_{t-1}^2,$$

is strictly stationary and ergodic. We consider the function $\tilde{\ell}_t$ defined as for the Gaussian QMLE in (3.9) by

$$\tilde{\ell}_t = \tilde{\ell}_t(\theta) = \log(\det \tilde{\Sigma}_t) + \varepsilon_t' \tilde{\Sigma}_t^{-1} \varepsilon_t,$$

and we note that

$$\tilde{\Sigma}_t(\theta_0) = H_t = \mathbb{E}(\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}). \tag{3.29}$$

Write $\lambda = (\lambda_1, \ldots, \lambda_k)'$, $\gamma = (\gamma_1, \ldots, \gamma_m)'$, and $\theta = (\gamma', \lambda)' = (\theta_1, \ldots, \theta_\ell)'$. Then $\ell_{t,n}(\lambda) = \tilde{\ell}_t(\hat{\gamma}_n, \lambda)$ and a Taylor series expansion of the VTE score vector around $\theta_0$ yields

$$(0,\ldots,0)' = n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\lambda}\tilde{\ell}_{t,n}(\hat{\lambda}_n) = n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\lambda}\tilde{\ell}_t(\hat{\theta}_n)$$

$$= n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\lambda}\tilde{\ell}_t(\theta_0) + \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_j\partial\lambda_i}\tilde{\ell}_t(\theta_i^*)\right)\sqrt{n}(\hat{\theta}_n - \theta_0), \tag{3.30}$$

where the $\theta_i^*$ are between $\hat{\theta}_n$ and $\theta_0$. We show in Lemma 10 that

$$n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\lambda}\tilde{\ell}_t(\theta_0) \xrightarrow{d} N(0,J),$$

where

$$J := \mathbb{E}\left(\frac{\partial}{\partial\lambda}\tilde{\ell}_t(\theta_0)\frac{\partial}{\partial\lambda'}\tilde{\ell}_t(\theta_0)\right),$$

and we show in Lemma (12) that

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial^2}{\partial\theta_j\partial\lambda_i}\tilde{\ell}_t(\theta_i^*)\right) \xrightarrow{a.s} \tilde{R}_0,$$

and $\tilde{R}_0$ has left inverse $R$, defined by

$$\tilde{R}_0 := \mathbb{E}\left(\frac{\partial^2}{\partial\theta\partial\lambda}\tilde{\ell}_t(\theta_0)\right).$$

The result follows from solving (3.30) and applying Slutsky's lemma.

□

### 3.4.3 Proof of Asymptotic Normality in Theorem 10

In this section we prove that the Gaussian QMLE,

$$\hat{\theta}_n^* = \underset{\theta\in\Theta}{\operatorname{argmin}}\frac{1}{n}\sum_{t=1}^{n}\tilde{\ell}_t(\theta),$$

is asymptotically normal under assumptions **A1** − **A6**. The proof is analogous to that given in section 3.4.2 for asymptotic normality of the VTE. A Taylor series expansion of the Gaussian QMLE score vector around $\theta_0$ yields

$$(0,\ldots,0)' = n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\tilde{\ell}_t(\hat{\theta}_n^*)$$

$$= n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\tilde{\ell}_t(\theta_0) + \left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial\theta_j\partial\theta_i}\tilde{\ell}_t(\tilde{\theta}_i^*)\right)\sqrt{n}(\hat{\theta}_n^* - \theta_0), \tag{3.31}$$

where the $\tilde{\theta}_i^*$ are between $\hat{\theta}_n^*$ and $\theta_0$. The same argument that proves

$$n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\lambda}\tilde{\ell}_t(\theta_0)\xrightarrow{d}N(0,J)$$

in Lemma 10 also shows that

$$n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\theta}\tilde{\ell}_t(\theta_0)\xrightarrow{d}N(0,\hat{J}),$$

upon replacing each $\lambda$ with $\theta$, so long as $\hat{J}$ is finite. We prove that $\hat{J}$ is finite in Lemma 11. To prove that

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial^2}{\partial\theta_j\partial\theta_i}\tilde{\ell}_t(\tilde{\theta}_i^*)\right)\xrightarrow{a.s}\hat{R}=\mathbb{E}\left(\frac{\partial^2}{\partial\theta\partial\theta}\tilde{\ell}_t(\theta_0)\right),$$

we simply replace each $\lambda$ with $\theta$ in the proof of Lemma 12, and note that $\hat{R}$ is finite by Lemma 13 and invertible by Lemma 14. The result follows from solving (3.31) and applying Slutsky's lemma.

$\square$

### 3.4.4 Technical Lemmas

In this section, we will make repeated use of the following matrix inequalities. For arbitrary matrices $X$ and $Y$,

$$|\mathrm{tr}(XY)|\leq||X||||Y|| \tag{3.32}$$

(see Zhang [52], p.25,213), and the inequality

$$0\leq\mathrm{tr}(XY)\leq\mathrm{tr}(X)\mathrm{tr}(Y) \tag{3.33}$$

holds whenever $X\geq0$ and $Y\geq0$ (Abadir and Magnus (2005), p.329-330).

**Lemma 10** *Under assumptions* **A1 – A6**,

$$n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\lambda}\tilde{\ell}_t(\theta_0)\xrightarrow{d}N(0,J)\ \textit{as}\ n\to\infty.$$

*Proof:* Under **A1 – A2**,

$$\left\{\frac{\partial}{\partial\lambda}\tilde{\ell}_t(\theta_0):t=0,\pm1,\ldots\right\}$$

is a strictly stationary and ergodic sequence. Furthermore, $\tilde{\Sigma}_t$, its inverse, and all of its derivatives, are $\mathcal{F}_{t-1}$ measurable, and $\mathbb{E}(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1})=H_t=\tilde{\Sigma}_t(\theta_0)$. Thus,

$$\mathbb{E}\left(\frac{\partial}{\partial\lambda_i}\tilde{\ell}_t(\theta_0)|\mathcal{F}_{t-1}\right) = \mathbb{E}\left[\frac{\partial}{\partial\lambda_i}\left(\log\det\tilde{\Sigma}_t(\theta_0) + \varepsilon_t'\tilde{\Sigma}_t^{-1}\varepsilon_t\right)|\mathcal{F}_{t-1}\right]$$

$$= \mathbb{E}\left[\mathrm{tr}\left(\tilde{\Sigma}_t^{-1}(\theta_0)\frac{\partial}{\partial\lambda_i}\tilde{\Sigma}_t(\theta_0)\right)\right]$$

$$-\mathbb{E}\left[\mathrm{tr}\left(\varepsilon_t\varepsilon_t'\tilde{\Sigma}_t^{-1}(\theta_0)\frac{\partial}{\partial\lambda_i}\tilde{\Sigma}_t(\theta_0)\tilde{\Sigma}_t^{-1}(\theta_0)\right)|\mathcal{F}_{t-1}\right]$$

$$= \mathrm{tr}\left(\tilde{\Sigma}_t^{-1}(\theta_0)\frac{\partial}{\partial\lambda_i}\tilde{\Sigma}_t(\theta_0)\right)$$

$$-\mathrm{tr}\left[\mathbb{E}\left(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}\right)\tilde{\Sigma}_t^{-1}(\theta_0)\frac{\partial}{\partial\lambda_i}\tilde{\Sigma}_t(\theta_0)\tilde{\Sigma}_t^{-1}(\theta_0)\right]$$

$$= \mathrm{tr}\left(\tilde{\Sigma}_t^{-1}(\theta_0)\frac{\partial}{\partial\lambda_i}\tilde{\Sigma}_t(\theta_0)\right)$$

$$-\mathrm{tr}\left[\tilde{\Sigma}_t(\theta_0)\tilde{\Sigma}_t^{-1}(\theta_0)\frac{\partial}{\partial\lambda_i}\tilde{\Sigma}_t(\theta_0)\tilde{\Sigma}_t^{-1}(\theta_0)\right]$$

$$= 0,$$

and so the score is a martingale difference and we may apply the Martingale CLT (see Billingsley [8], p.788) to obtain

$$n^{-1/2}\sum_{t=1}^{n}\frac{\partial}{\partial\lambda}\tilde{\ell}_t(\theta_0) \xrightarrow{d} N(0,J),$$

where

$$J = \mathbb{E}\left(\frac{\partial}{\partial\lambda}\tilde{\ell}_t(\theta_0)\frac{\partial}{\partial\lambda'}\tilde{\ell}_t(\theta_0)\right)$$

so long as $J$ is finite. The matrix $J$ is a submatrix of

$$\hat{J} = \mathbb{E}\left(\frac{\partial}{\partial\theta}\tilde{\ell}_t(\theta_0)\frac{\partial}{\partial\theta'}\tilde{\ell}_t(\theta_0)\right),$$

so it suffices to show that the matrix $\hat{J}$ is finite, and we prove that $\hat{J}$ is finite in Lemma 11.

□

**Lemma 11** *Under assumptions* **A1 – A6**, *the matrix*

$$\hat{J} = \mathbb{E}\left(\frac{\partial}{\partial\theta}\tilde{\ell}_t(\theta_0)\frac{\partial}{\partial\theta'}\tilde{\ell}_t(\theta_0)\right)$$

*is finite.*

*Proof:* To simplify notation, we denote $\partial/\partial\theta_i(\tilde{\Sigma}_t(\theta_0))$ by $\dot{\Sigma}_{t,i}$, and $\tilde{\Sigma}_t^{-1}(\theta_0)$ by $\tilde{\Sigma}_t^{-1}$. We will use the same notation, in later lemmas, to denote, respectively, $\partial/\partial\theta_i(\tilde{\Sigma}_t(\theta))$ and $\tilde{\Sigma}_t^{-1}(\theta)$. An arbitrary element of $\hat{J}$ has the form

$$\begin{aligned}\frac{\partial}{\partial\theta_i}\tilde{\ell}_t(\theta_0)\frac{\partial}{\partial\theta_j}\tilde{\ell}_t(\theta_0) = {} & \operatorname{tr}\left(\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,i}\right)\operatorname{tr}\left(\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,j}\right) \\ & +\operatorname{tr}\left(\varepsilon_t\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}\right)\operatorname{tr}\left(\varepsilon_t\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,j}\tilde{\Sigma}_t^{-1}\right) \\ & -\operatorname{tr}\left(\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,i}\right)\operatorname{tr}\left(\varepsilon_t\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,j}\tilde{\Sigma}_t^{-1}\right) \\ & -\operatorname{tr}\left(\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,j}\right)\operatorname{tr}\left(\varepsilon_t\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}\right).\end{aligned}$$

Using that $\tilde{\Sigma}_t$, its inverse, and all of its derivatives are $\mathcal{F}_{t-1}$ measurable, and applying (3.29), we see that the expectations of the first and third terms above cancel each other, leaving

$$\begin{aligned}\mathbb{E}\left[\frac{\partial}{\partial\lambda_i}\tilde{\ell}_t(\theta_0)\frac{\partial}{\partial\lambda_j}\tilde{\ell}_t(\theta_0)\right] = {} & \mathbb{E}\left[\operatorname{tr}\left(\varepsilon_t\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}\right)\operatorname{tr}\left(\varepsilon_t\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,j}\tilde{\Sigma}_t^{-1}\right)\right] \\ & -\mathbb{E}\left[\operatorname{tr}\left(\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,j}\right)\operatorname{tr}\left(\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}\right)\right]. \qquad (3.34)\end{aligned}$$

Applying successively (3.32), Lemma 17, and Lemma 16 we see that the second term of (3.34) is bounded. That is,

$$\begin{aligned}\mathbb{E}\left[\operatorname{tr}\left(\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,j}\right)\operatorname{tr}\left(\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}\right)\right] & \leq \mathbb{E}\left(\left|\left|\tilde{\Sigma}_t^{-1}\right|\right|^2\left|\left|\dot{\Sigma}_{t,i}\right|\right|\left|\left|\dot{\Sigma}_{t,j}\right|\right|\right) \\ & \leq M_1^2\mathbb{E}\left(\max_{1\leq i\leq\ell}\left|\left|\dot{\Sigma}_{t,i}\right|\right|^2\right) < \infty. \qquad (3.35)\end{aligned}$$

To prove that the first term of (3.34) is bounded we will use repeatedly the linearity of trace and expectation, and that the trace operator is invariant under cyclic permutations of its argument. Noting that $\tilde{\Sigma}_t$, its inverse, and all of its derivatives are $\mathcal{F}_{t-1}$ measurable we have

$$
\begin{aligned}
&\mathbb{E}\left[\operatorname{tr}\left(\varepsilon_t\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}\right)\operatorname{tr}\left(\varepsilon_t\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,j}\tilde{\Sigma}_t^{-1}\right)\right]\\
&\qquad = \mathbb{E}\left[\operatorname{tr}\left(\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}\varepsilon_t\right)\operatorname{tr}\left(\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,j}\tilde{\Sigma}_t^{-1}\varepsilon_t\right)\right]\\
&\qquad = \mathbb{E}\left[\left(\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}\varepsilon_t\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,j}\tilde{\Sigma}_t^{-1}\varepsilon_t\right)\right]\\
&\qquad = \mathbb{E}\left\{\mathbb{E}\left[\operatorname{tr}\left(\varepsilon_t\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}\varepsilon_t\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,j}\tilde{\Sigma}_t^{-1}\right)|\mathcal{F}_{t-1}\right]\right\}\\
&\qquad = \mathbb{E}\left\{\operatorname{tr}\left(\mathbb{E}\left[\varepsilon_t\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}\right]\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,j}\tilde{\Sigma}_t^{-1}\right)\right\}. \qquad (3.36)
\end{aligned}
$$

Using (3.1), $H_t = \tilde{\Sigma}_t(\theta_0) = \tilde{\Sigma}_t$, and (3.33), we can write (3.36) as

$$
\begin{aligned}
&\mathbb{E}\left\{\operatorname{tr}\left(H_t^{1/2}\mathbb{E}\left[\eta_t\eta_t'H_t^{-1/2}\dot{\Sigma}_{t,i}H_t^{-1/2}\eta_t\eta_t'|\mathcal{F}_{t-1}\right]H_t^{-1/2}\dot{\Sigma}_{t,j}\tilde{\Sigma}_t^{-1}\right)\right\}\\
&\qquad = \mathbb{E}\left\{\operatorname{tr}\left(\mathbb{E}\left[\eta_t\eta_t'H_t^{-1/2}\dot{\Sigma}_{t,i}H_t^{-1/2}\eta_t\eta_t'|\mathcal{F}_{t-1}\right]H_t^{-1/2}\dot{\Sigma}_{t,j}\ H_t^{-1/2}\right)\right\}\\
&\qquad \le \mathbb{E}\left\{\operatorname{tr}\left(\mathbb{E}\left[\eta_t\eta_t'H_t^{-1/2}\dot{\Sigma}_{t,i}H_t^{-1/2}\eta_t\eta_t'|\mathcal{F}_{t-1}\right]\right)\operatorname{tr}\left(H_t^{-1/2}\dot{\Sigma}_{t,j}\ H_t^{-1/2}\right)\right\}\\
&\qquad = \mathbb{E}\left\{\mathbb{E}\left[\operatorname{tr}\left(\eta_t\eta_t'H_t^{-1/2}\dot{\Sigma}_{t,i}H_t^{-1/2}\eta_t\eta_t'\right)|\mathcal{F}_{t-1}\right]\operatorname{tr}\left(H_t^{-1/2}\dot{\Sigma}_{t,j}\ H_t^{-1/2}\right)\right\}\\
&\qquad = \mathbb{E}\left\{\mathbb{E}\left[\operatorname{tr}\left(\eta_t\eta_t'\eta_t\eta_t'H_t^{-1/2}\dot{\Sigma}_{t,i}H_t^{-1/2}\right)|\mathcal{F}_{t-1}\right]\operatorname{tr}\left(H_t^{-1/2}\dot{\Sigma}_{t,j}\ H_t^{-1/2}\right)\right\}\\
&\qquad = \mathbb{E}\left\{\operatorname{tr}\left(\mathbb{E}\left[\eta_t\eta_t'\eta_t\eta_t'|\mathcal{F}_{t-1}\right]H_t^{-1/2}\dot{\Sigma}_{t,i}H_t^{-1/2}\right)\operatorname{tr}\left(H_t^{-1/2}\dot{\Sigma}_{t,j}\ H_t^{-1/2}\right)\right\}\\
&\qquad \le \mathbb{E}\left[\eta_t'\eta_t\eta_t'\eta_t\right]\mathbb{E}\left[\max_{1\le i\le \ell}\left(\operatorname{tr}\left(\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,i}\}\right)^2\right)\right]\\
&\qquad \le \mathbb{E}\left[\eta_t'\eta_t\eta_t'\eta_t\right]\mathbb{E}\left[\max_{1\le i\le \ell}\left(\left|\left|\tilde{\Sigma}_t^{-1}\right|\right|^2\left|\left|\dot{\Sigma}_{t,i}\}\right|\right|^2\right)\right]\\
&\qquad \le M_1^2\mathbb{E}\left[\eta_t'\eta_t\eta_t'\eta_t\right]\mathbb{E}\left[\max_{1\le i\le \ell}\left(\left|\left|\dot{\Sigma}_{t,i}\}\right|\right|^2\right)\right]\\
&\qquad < \infty.
\end{aligned}
$$

The last three lines follows from (3.32), **A6**, Lemma 17 and Lemma 16.

□

**Lemma 12** *Under assumptions* **A1 – A6**

$$\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial^2}{\partial\theta_j\partial\lambda_i}\tilde{\ell}_t(\theta_i^*)\right)\xrightarrow{a.s}\tilde{R}_0=\mathbb{E}\left(\frac{\partial^2}{\partial\theta\partial\lambda}\tilde{\ell}_t(\theta_0)\right)$$

*as $n\to\infty$, and $\tilde{R}_0$ admits a left inverse.*

*Proof:* Following Straumann and Mikosch [47] we use the result of Rao [43], regarding conditions for uniform convergence in the strong law of large numbers. We will apply this result as stated in Theorem 2.7 of Straumann and Mikosch ([47], p. 2456), which we summarize as follows; if $\{v_t : t = 0, \pm1, \ldots\}$ is a stationary and ergodic sequence of random elements with values in the space of continuous functions, equipped with the supremum norm, taking values from a compact set $K \subset \mathbb{R}^m$ into $\mathbb{R}^n$, then uniform convergence in the strong law of large numbers is implied by

$$\mathbb{E}\left(\sup_{s\in K}||v_0(s)||\right).$$

Under assumptions **A1** $-$ **A3**, the sequence

$$\left\{\frac{\partial^2}{\partial\theta\partial\lambda}\tilde{\ell}_t(\theta) : t = 0, \pm1, \ldots\right\} \tag{3.37}$$

is strictly stationary and ergodic for each $\theta \in \Theta$ . Thus, for any compact set $K \subset \Theta$,

$$\lim_{n\to\infty}\sup_{\theta\in K}\left|\left|\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial^2}{\partial\theta\partial\lambda}\tilde{\ell}_t(\theta)\right)-\mathbb{E}\left(\frac{\partial^2}{\partial\theta\partial\lambda}\tilde{\ell}_t(\theta)\right)\right|\right|\xrightarrow{a.s}0 \tag{3.38}$$

is implied by

$$\mathbb{E}\left(\sup_{\theta\in K}\left|\left|\frac{\partial^2}{\partial\theta\partial\lambda}\tilde{\ell}_0(\theta)\right|\right|\right)<\infty. \tag{3.39}$$

The first part of the theorem follows if we can show that (3.39) holds for some compact set $K$ containing $\theta_0$, and by noting that the points $\theta_i^*$ converge almost surely to $\theta_0$ as $n \to \infty$. We show that (3.39) holds by proving the stronger statement

$$\mathbb{E}\left(\sup_{\theta\in V_0}\left|\left|\frac{\partial^2}{\partial\theta\partial\theta}\tilde{\ell}_0(\theta)\right|\right|\right)<\infty, \tag{3.40}$$

where $V_0$ is the neighborhood of $\theta_0$ constructed in Lemma 15. To simplify notation in the following we denote $\partial/\partial\theta_i(\tilde{\Sigma}_0(\theta))$ by $\dot{\Sigma}_i$, $\partial^2/\partial\theta_j\partial\theta_i(\tilde{\Sigma}_0(\theta))$ by $\ddot{\Sigma}_{ij}$, and $\tilde{\Sigma}_0^{-1}(\theta)$ by $\tilde{\Sigma}^{-1}$. An arbitrary element of the matrix appearing in (3.40) takes the form

$$\frac{\partial^2}{\partial\theta_j\partial\theta_i}\tilde{\ell}_0(\theta) = \mathrm{tr}\left(\ddot{\Sigma}_{ij}\tilde{\Sigma}^{-1}\right) + \mathrm{tr}\left(\varepsilon_0\varepsilon_0'\tilde{\Sigma}^{-1}\dot{\Sigma}_j\tilde{\Sigma}^{-1}\dot{\Sigma}_i\tilde{\Sigma}^{-1}\right) + \mathrm{tr}\left(\varepsilon_0\varepsilon_0'\tilde{\Sigma}^{-1}\dot{\Sigma}_i\tilde{\Sigma}^{-1}\dot{\Sigma}_j\tilde{\Sigma}^{-1}\right)$$

$$-\mathrm{tr}\left(\dot{\Sigma}_i\tilde{\Sigma}^{-1}\dot{\Sigma}_j\tilde{\Sigma}^{-1}\right) - \mathrm{tr}\left(\varepsilon_0\varepsilon_0'\tilde{\Sigma}^{-1}\ddot{\Sigma}_{ij}\tilde{\Sigma}^{-1}\right). \qquad (3.41)$$

Regarding the first term of (3.41), it follows from (3.32), Lemma 17 and Lemma 16 that

$$\mathbb{E}\left(\sup_{\theta\in\Theta}\mathrm{tr}\left(\ddot{\Sigma}_{ij}\tilde{\Sigma}^{-1}\right)\right) \le \mathbb{E}\left(\sup_{\theta\in\Theta}\left|\left|\ddot{\Sigma}_{ij}\right|\right|\left|\left|\tilde{\Sigma}^{-1}\right|\right|\right)$$

$$\le M_1\mathbb{E}\left(\sup_{\theta\in\Theta}\left|\left|\ddot{\Sigma}_{ij}\right|\right|\right) < \infty. \qquad (3.42)$$

To prove that the expectation of the second term of (3.41) is bounded uniformly in $\theta$ over $V_0$ we will use repeatedly the linearity of trace and expectation and that the trace operator is invariant under cyclic permutations of its argument. Noting that $H_t$, $\tilde{\Sigma}$, its inverse, and all of its derivatives are $\mathcal{F}_{t-1}$ measurable, and using (3.32), we can write

$$\mathbb{E}\left[\mathrm{tr}\left(\varepsilon_0\varepsilon_0'\tilde{\Sigma}^{-1}\dot{\Sigma}_j\tilde{\Sigma}^{-1}\dot{\Sigma}_i\tilde{\Sigma}^{-1}\right)\right]$$

$$= \mathbb{E}\left[\mathrm{tr}\left(\tilde{\Sigma}^{-1}\varepsilon_0\varepsilon_0'\tilde{\Sigma}^{-1}\dot{\Sigma}_j\tilde{\Sigma}^{-1}\dot{\Sigma}_i\right)\right]$$

$$= \mathbb{E}\left[\mathrm{tr}\left(\tilde{\Sigma}^{-1}\varepsilon_0\varepsilon_0'\tilde{\Sigma}^{-1}\dot{\Sigma}_j\tilde{\Sigma}^{-1}\dot{\Sigma}_i\right)\right]$$

$$= \mathrm{tr}\left[\mathbb{E}\left\{\mathbb{E}\left(\tilde{\Sigma}^{-1}\varepsilon_0\varepsilon_0'\tilde{\Sigma}^{-1}\dot{\Sigma}_j\tilde{\Sigma}^{-1}\dot{\Sigma}_i|\mathcal{F}_{t-1}\right)\right\}\right]$$

$$= \mathrm{tr}\left[\mathbb{E}\left\{\tilde{\Sigma}^{-1}H_0^{1/2}\mathbb{E}\left(\eta_0\eta_0'|\mathcal{F}_{t-1}\right)H_0^{1/2}\tilde{\Sigma}^{-1}\dot{\Sigma}_j\tilde{\Sigma}^{-1}\dot{\Sigma}_i\right\}\right]$$

$$= \mathbb{E}\left[\mathrm{tr}\left(\tilde{\Sigma}^{-1}H_0\tilde{\Sigma}^{-1}\dot{\Sigma}_j\tilde{\Sigma}^{-1}\dot{\Sigma}_i\right)\right]$$

$$= \mathbb{E}\left[\mathrm{tr}\left(\tilde{\Sigma}^{-1/2}\dot{\Sigma}_i\tilde{\Sigma}^{-1}H_0^{1/2}H_0^{1/2}\tilde{\Sigma}^{-1}\dot{\Sigma}_j\tilde{\Sigma}^{-1/2}\right)\right]$$

$$\le \mathbb{E}\left(\left|\left|\tilde{\Sigma}^{-1/2}\dot{\Sigma}_i\tilde{\Sigma}^{-1}H_0^{1/2}\right|\right|\left|\left|H_0^{1/2}\tilde{\Sigma}^{-1}\dot{\Sigma}_j\tilde{\Sigma}^{-1/2}\right|\right|\right), \qquad (3.43)$$

where the last line follows from (3.32).

Using that $||X|| = ||X'||$ holds for arbitrary matrices, and noting that every matrix appearing in (3.43) is symmetric, we can bound (3.43) as

$$
\begin{aligned}
&\mathbb{E}\left(\left|\left|\tilde{\Sigma}^{-1/2}\dot{\Sigma}_i\tilde{\Sigma}^{-1}H_0^{1/2}\right|\right|\left|\left|\tilde{\Sigma}^{-1/2}\dot{\Sigma}_j\tilde{\Sigma}^{-1}H_0^{1/2}\right|\right|\right) \\
&\qquad\qquad \leq \mathbb{E}\left(\max_{1\leq i\leq \ell}\left|\left|\tilde{\Sigma}^{-1/2}\dot{\Sigma}_i\tilde{\Sigma}^{-1}H_0^{1/2}\right|\right|^2\right) \\
&\qquad\qquad = \mathbb{E}\left(\max_{1\leq i\leq \ell}\ \mathrm{tr}\left[\tilde{\Sigma}^{-1/2}\dot{\Sigma}_i\tilde{\Sigma}^{-1}H_0\tilde{\Sigma}^{-1}\dot{\Sigma}_i\tilde{\Sigma}^{-1/2}\right]\right) \\
&\qquad\qquad \leq \mathbb{E}\left(\max_{1\leq i\leq \ell}\ \mathrm{tr}\left[\tilde{\Sigma}^{-1}\right]\mathrm{tr}\left[\dot{\Sigma}_i\tilde{\Sigma}^{-1}H_0\tilde{\Sigma}^{-1}\dot{\Sigma}_i\right]\right) \\
&\qquad\qquad \leq \sqrt{d}M_1\mathbb{E}\left(\max_{1\leq i\leq \ell}\ \mathrm{tr}\left[\tilde{\Sigma}^{-1/2}H_0\tilde{\Sigma}^{-1/2}\tilde{\Sigma}^{-1/2}\dot{\Sigma}_i\dot{\Sigma}_i\tilde{\Sigma}^{-1/2}\right]\right) \\
&\qquad\qquad \leq \sqrt{d}M_1\mathbb{E}\left(\max_{1\leq i\leq \ell}\ \mathrm{tr}\left[\tilde{\Sigma}^{-1}H_0\right]\mathrm{tr}\left[\tilde{\Sigma}^{-1}\dot{\Sigma}_i\dot{\Sigma}_i\right]\right) \\
&\qquad\qquad \leq \sqrt{d}M_1(d+1)\mathbb{E}\left(\max_{1\leq i\leq \ell}\left|\left|\dot{\Sigma}_i\right|\right|^2\left|\left|\tilde{\Sigma}^{-1}\right|\right|\right) \\
&\qquad\qquad \leq \sqrt{d}M_1^2(d+1)\mathbb{E}\left(\max_{1\leq i\leq \ell}\left|\left|\dot{\Sigma}_i\right|\right|^2\right).
\end{aligned}
$$

The last five lines follow, successively, from (3.33), (3.32) and Lemma 17, (3.33), Lemma 15, and Lemma 17. The constants $d$ and $M_1$ do not depend on $\theta$ and it follows from Lemma 15 that

$$
\mathbb{E}\sup_{\theta\in\Theta}\max_{1\leq i\leq \ell}\left|\left|\dot{\Sigma}_i\right|\right|^2<\infty.
$$

Thus the expectation of the second term of (3.41) is bounded uniformly in $\theta$ over $V_0$, and an identical argument bounds the third term.

The fourth term of (3.41) is bounded by (3.32), Lemma 17, and Lemma 16. That is,

$$
\begin{aligned}
\mathbb{E}\left[\sup_{\theta\in\Theta}\mathrm{tr}\left(\dot{\Sigma}_i\tilde{\Sigma}_t^{-1}\dot{\Sigma}_j\tilde{\Sigma}_t^{-1}\right)\right] &\leq \mathbb{E}\left(\sup_{\theta\in\Theta}\left|\left|\dot{\Sigma}_i\right|\right|\left|\left|\dot{\Sigma}_j\right|\right|\left|\left|\tilde{\Sigma}_t^{-1}\right|\right|^2\right) \\
&\leq M_1^2\mathbb{E}\left(\sup_{\theta\in\Theta}\max_{1\leq i\leq \ell}\left|\left|\dot{\Sigma}_i\right|\right|^2\right).
\end{aligned}
$$

To bound the fifth term of (3.41), we use an argument similar to that which bounds the second and third terms.

$$\begin{aligned}
\mathbb{E}\left[\operatorname{tr}\left(\varepsilon_0\varepsilon_0'\tilde{\Sigma}_t^{-1}\ddot{\Sigma}_{ij}\tilde{\Sigma}_t^{-1}\right)\right] &= \mathbb{E}\left\{\mathbb{E}\left[\operatorname{tr}\left(\varepsilon_0\varepsilon_0'\tilde{\Sigma}_t^{-1}\ddot{\Sigma}_{ij}\tilde{\Sigma}_t^{-1}|\mathcal{F}_{\sqcup-\infty}\right)\right]\right\} \\
&= \mathbb{E}\left\{\mathbb{E}\left[\operatorname{tr}\left(\varepsilon_0\varepsilon_0'\tilde{\Sigma}_t^{-1}\ddot{\Sigma}_{ij}\tilde{\Sigma}_t^{-1}|\mathcal{F}_{t-1}\right)\right]\right\} \\
&= \mathbb{E}\left[\operatorname{tr}\left(\tilde{\Sigma}_t^{-1}H_0\tilde{\Sigma}_t^{-1}\ddot{\Sigma}_{ij}\right)\right] \\
&= \mathbb{E}\left[\operatorname{tr}\left(\tilde{\Sigma}_t^{-1/2}H_0\tilde{\Sigma}_t^{-1/2}\tilde{\Sigma}_t^{-1/2}\ddot{\Sigma}_{ij}\tilde{\Sigma}_t^{-1/2}\right)\right] \\
&\leq \mathbb{E}\left[\operatorname{tr}\left(\tilde{\Sigma}_t^{-1/2}H_0\tilde{\Sigma}_t^{-1/2}\right)\operatorname{tr}\left(\tilde{\Sigma}_t^{-1/2}\ddot{\Sigma}_{ij}\tilde{\Sigma}_t^{-1/2}\right)\right] \\
&= \mathbb{E}\left[\operatorname{tr}\left(\tilde{\Sigma}_t^{-1}H_0\right)\operatorname{tr}\left(\tilde{\Sigma}_t^{-1}\ddot{\Sigma}_{ij}\right)\right] \\
&\leq (d+1)\mathbb{E}\left(\left|\left|\tilde{\Sigma}_t^{-1}\right|\right|\left|\left|\ddot{\Sigma}_{ij}\right|\right|\right) \\
&\leq (d+1)M_1\mathbb{E}\left(\sup_{\theta\in\Theta}\left|\left|\ddot{\Sigma}_{ij}\right|\right|\right) \\
&< \infty.
\end{aligned}$$

The last five lines use (3.33), Lemma 15, Lemma 17, and Lemma 16. This completes the proof of (3.40).

The matrix $\tilde{R}_0$ is the lower right $k \times \ell$ block of the matrix $\hat{R}$; thus $\tilde{R}_0$ admits a left inverse if $\hat{R}$ is invertible. Invertibility of $\hat{R}$ follows from Lemma 14 where we prove that $\hat{R}$ is positive definite.

□

**Lemma 13** *Under assumptions* **A1 − A6***, the matrix*

$$\hat{R} = \mathbb{E}\left(\frac{\partial^2}{\partial\theta_j\partial\theta_i}\tilde{\ell}_t(\theta_0)\right)$$

*is finite.*

*Proof:* To simplify notation in the following we denote $\partial/\partial\theta_i(\tilde{\Sigma}_t(\theta_0))$ by $\dot{\Sigma}_{t,i}$, $\tilde{\Sigma}_t^{-1}(\theta_0)$ by $\tilde{\Sigma}_t^{-1}$, and $\partial^2/\partial\theta_j\partial\theta_i(\tilde{\Sigma}_t(\theta_0))$ by $\ddot{\Sigma}_{t,ij}$. The score vector is given by

$$\frac{\partial}{\partial\theta_i}\tilde{\ell}_t(\theta_0) = \operatorname{tr}(\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1} - \varepsilon_t\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}),$$

and the Hessian matrix is given by

$$\frac{\partial^2}{\partial\theta_j\partial\theta_i}\tilde{\ell}_t(\theta_0) = \mathrm{tr}\left(\ddot{\Sigma}_{t,ij}\tilde{\Sigma}_t^{-1} - \dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,j}\tilde{\Sigma}_t^{-1} + \varepsilon_t\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,j}\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}\right)$$

$$-\mathrm{tr}\left(\varepsilon_t\varepsilon_t'\tilde{\Sigma}_t^{-1}\ddot{\Sigma}_{t,ij}\tilde{\Sigma}_t^{-1} + \varepsilon_t\varepsilon_t'\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,j}\tilde{\Sigma}_t^{-1}\right).$$

Examining the five terms on the right-hand side above, we see that since

$$\mathbb{E}(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) = H_t = \tilde{\Sigma}_t(\theta_0),$$

and since $\tilde{\Sigma}_t$, its inverse, and all of its derivatives are $\mathcal{F}_{t-1}$ measurable, the first and fourth terms above cancel each other as do the second and fifth terms. Thus,

$$\mathbb{E}\left(\frac{\partial^2}{\partial\theta_j\partial\theta_i}\tilde{\ell}_t\right) = \mathbb{E}\left[\mathbb{E}\left(\frac{\partial^2}{\partial\theta_j\partial\theta_i}\tilde{\ell}_t|\mathcal{F}_{t-1}\right)\right] = \mathbb{E}\left[\mathrm{tr}\left(\dot{\Sigma}_{t,j}\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}\right)\right].$$

It follows from (3.32), (3.33), Lemma 17, and Lemma 16 that

$$\begin{aligned}
\mathbb{E}\left|\frac{\partial^2}{\partial\theta_j\partial\lambda_i}\tilde{\ell}_t\right| &= \mathbb{E}\left|\mathrm{tr}\left(\dot{\Sigma}_{t,j}\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}\right)\right| \\
&\leq \mathbb{E}\left(||\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}||\,||\dot{\Sigma}_{t,j}\tilde{\Sigma}_t^{-1}||\right) \\
&\leq \mathbb{E}\left(||\dot{\Sigma}_{t,i}||\,||\dot{\Sigma}_{t,j}||\,||\tilde{\Sigma}_t^{-1}||^2\right) \\
&\leq M_1^2\mathbb{E}\left(\sup_{\theta\in\Theta}\max_{1\leq i\leq\ell}||\dot{\Sigma}_{t,i}||^2\right) \\
&< \infty.
\end{aligned}$$

□

**Lemma 14** *Under assumptions* **A1** − **A6**, *the matrix*

$$\hat{R} = \mathbb{E}\left(\frac{\partial^2}{\partial\theta_j\partial\theta_i}\tilde{\ell}_t(\theta_0)\right)$$

*is positive definite.*

*Proof:* We follow the argument of Comte and Lieberman (2003). To simplify notation we denote $\partial/\partial\theta_i(\tilde{\Sigma}_t(\theta_0))$ by $\dot{\Sigma}_{t,i}$, and $\tilde{\Sigma}_t^{-1}(\theta_0)$ by $\tilde{\Sigma}_t^{-1}$. We define $A_i := \tilde{\Sigma}_t^{-1/2}\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1/2}$ for each $1 \leq i \leq \ell$. For arbitrary conformable matrices $X$ and $Y$,

$$\mathrm{tr}(XY) = \mathrm{vec}(X')'\mathrm{vec}(Y).$$

Thus,

$$\hat{R} = \mathbb{E}\left(\frac{\partial^2}{\partial\theta_j\partial\theta_i}\tilde{\ell}_t(\theta_0)\right) = \mathbb{E}\left[\mathrm{tr}\left(\dot{\Sigma}_{t,i}\tilde{\Sigma}_t^{-1}\dot{\Sigma}_{t,j}\tilde{\Sigma}_t^{-1}\right)\right] = \mathbb{E}\left[\mathrm{tr}\left(A_iA_j\right)\right] = \mathrm{vec}(A_i')'\mathrm{vec}(A_j).$$

Since

$$\mathrm{vec}(XYZ) = (Z' \otimes X)\mathrm{vec}(Y),$$

we have

$$\mathrm{vec}(A_i) = \left(\tilde{\Sigma}_t^{-1/2} \otimes \tilde{\Sigma}_t^{-1/2}\right)\mathrm{vec}\left(\dot{\Sigma}_{t,i}\right),$$

and thus

$$\hat{R} = \mathbb{E}\left(P_t'\left(\tilde{\Sigma}_t^{-1} \otimes \tilde{\Sigma}_t^{-1}\right)P_t\right)$$

where the matrix $P_t$ is given by $P_t = (\mathrm{vec}(\dot{\Sigma}_{t,1}), \ldots, \mathrm{vec}(\dot{\Sigma}_{t,\ell}))$. The matrix $(\tilde{\Sigma}_t^{-1} \otimes \tilde{\Sigma}_t^{-1})$ is positive definite as the Kronecker product of positive definite matrices (see Gross (2003), p.355), and thus $\hat{R}$ is positive semidefinite.

Suppose (by way of contradiction) that $\hat{R}$ is singular. Then there is some vector $x$, $x \neq 0$, such that

$$x\mathbb{E}\left(P_t'\left(\tilde{\Sigma}_t^{-1} \otimes \tilde{\Sigma}_t^{-1}\right)P_t\right)x = 0.$$

Thus,

$$\mathbb{E}\left[(P_tx)'\left(\tilde{\Sigma}_t^{-1} \otimes \tilde{\Sigma}_t^{-1}\right)P_tx\right] = 0,$$

which implies

$$P_tx = 0$$

almost surely. From the definition of $P_t$, if $P_tx = 0$ almost surely, then there is some nonzero vector $y$ such that

$$y'\left(\frac{\partial}{\partial\theta}(\tilde{\sigma}_t^2)\right) = 0 \text{ almost surely.}$$

This implies that another representation of $\tilde{\sigma}_t^2$ is possible, contradicting part (ii) of Theorem 8.

□

**Lemma 15** *Under assumptions* **A1-A6** *there exists, for each fixed* $t \in \{0, \pm 1, \pm 2, \ldots\}$, *some neighborhood* $V_t$ *of* $\theta_0$ *in* $\Theta$ *such that*

$$\sup_{\theta \in V_t} \operatorname{tr}\left(\tilde{\Sigma}_t(\theta_0)\tilde{\Sigma}_t^{-1}(\theta)\right) \leq d + 1.$$

*Proof:* Fix any $t \in \{0, \pm 1, \pm 2, \ldots\}$. The mapping $\theta \mapsto \operatorname{tr}(\tilde{\Sigma}_t(\theta_0)\tilde{\Sigma}_t^{-1}(\theta))$ is smooth, and

$$\operatorname{tr}\left(\tilde{\Sigma}_t(\theta_0)\tilde{\Sigma}_t^{-1}(\theta_0)\right) = \operatorname{tr}(I_d) = d,$$

so it is natural to expect that

$$\operatorname{tr}\left(\tilde{\Sigma}_t(\theta_0)\tilde{\Sigma}_t^{-1}(\theta)\right)$$

should take a value close to $d$ whenever $||\theta - \theta_0||$ is small. However, the situation is complicated by the fact that $\tilde{\Sigma}_t$ and $\tilde{\Sigma}_t^{-1}$ are functions not only of $\theta$, but also of the i.i.d. sequence $\{\eta_{t-1}, \eta_{t-2}, \ldots\}$.

For each $k \in \{1, 2, \ldots\}$, we consider $\eta_{t-k}$ as a measurable function from a probability space $(\Omega_k, \mathcal{F}_k, Q_k)$ into $\mathbb{R}^d$. Then, defining $\Omega$ by the cartesian product

$$\Omega := \textstyle\bigotimes_{k=1}^{\infty} \Omega_k,$$

it follows that $\eta := (\eta_{t-1}, \eta_{t-2}, \ldots)$ is a well-defined random variable on the probability space $(\Omega, \mathcal{F}, Q)$ where

$$\mathcal{F} := \textstyle\bigotimes_{k=1}^{\infty} \mathcal{F}_k,$$

and $Q$ is defined by

$$Q(A) = \textstyle\prod_{k=1}^{\infty} Q_k(A_k) \text{ for } A = \bigotimes_{k=1}^{\infty} A_k \in \mathcal{F}$$

(see Fristedt and Gray [27], p.136-140). The function $f : \Theta \times \Omega \to \mathbb{R}$ defined by

$$f(\theta, \omega) = \operatorname{tr}\left(\tilde{\Sigma}_t(\theta_0)\tilde{\Sigma}_t^{-1}(\theta)\right)$$

is smooth in $\theta$, so for each fixed $\omega \in \Omega$ there exists a neighborhood $V_\omega$ of $\theta_0$ such that $f(V_\omega, \omega) \subset (d-1, d+1)$. Define

$$V := \bigcup_{\omega \in \Omega} V_\omega.$$

Then $f(V) \subset (d-1, d+1)$ and, denoting by $V_t$ the projection of $V$ onto $\Theta$, it follows that $V = V_t \times \Omega$ and $V_t$ is an open set containing $\theta_0$ in $\Theta$. Now, for every $\theta \in V_t$ and for every $\omega \in \Omega$ we have $f(\theta, \omega) \in (d-1, d+1)$. Thus

$$\sup_{\theta \in V_t} f(\theta, \omega) \leq d+1$$

holds for all $\omega \in \Omega$, and thus

$$\sup_{\theta \in V_t} \ \mathrm{tr}\left(\tilde{\Sigma}_t(\theta_0)\tilde{\Sigma}_t^{-1}(\theta)\right) \leq d+1$$

holds for every possible realization of the random variable $\tilde{\Sigma}_t(\theta_0)\tilde{\Sigma}_t^{-1}(\theta)$.

$\square$

**Lemma 16** *Under assumptions* **A1** $-$ **A6**, *the derivatives of* $\tilde{\sigma}_t^2$ *satisfy*

(i) $\mathbb{E}\sup_{\theta \in \Theta} \left|\left|\dot{\Sigma}_{t,i}(\theta)\right|\right| < \infty$,

(ii) $\mathbb{E}\sup_{\theta \in \Theta} \left|\left|\dot{\Sigma}_{t,i}(\theta)\right|\right|^2 < \infty$,

(iii) $\mathbb{E}\sup_{\theta \in \Theta} \left|\left|\ddot{\Sigma}_{t,ij}(\theta)\right|\right| < \infty$.

*Proof:* Recall our definition of the spectral norm in (2.5). The following inequalities hold for arbitrary conformable matrices $X$ and $Y$ (see Magnus and Neudecker [35]),

$$||XY|| \leq N(X)||Y||, \quad ||XY|| \leq ||X||N(Y), \quad N(XY) \leq N(X)N(Y). \tag{3.44}$$

Furthermore,

$$N(X \otimes Y)^2 = \rho((X \otimes Y)'(X \otimes Y)) = \rho(X'X)\rho(Y'Y) = N(X)^2 N(Y)^2 \tag{3.45}$$

To simplify notation, we denote $B = I - K - A$, and we define

$$C := I_m \otimes K_m \otimes I_m, \quad \mathfrak{A}_i := (K\gamma + As_{t-i})', \quad a := \mathrm{vech}(A), \quad k := \mathrm{vech}(K),$$

where $K_m$ is the commutation matrix. From our assumptions on $\Theta$, $\rho(B(\theta)) < 1$ for all $\theta \in \Theta$. The eigenvalues of any square matrix are continuous as functions of the matrix coordinates (see Horn and Johnson [32] p.539-540). Thus,

$$\sup_{\theta\in\Theta} \ \rho(B(\theta)) < 1,$$

and

$$r := \sup_{\theta\in\Theta} \ N(B(\theta)) < 1. \tag{3.46}$$

The matrix $C$ is bounded in norm, and we have the following bounds

$$\sup_{\theta\in\Theta} ||K(\theta)|| < \infty, \quad \sup_{\theta\in\Theta} ||\gamma(\theta)|| < \infty, \quad \sup_{\theta\in\Theta} ||A(\theta)|| < \infty, \tag{3.47}$$

because each of the matrices above is the supremum of a continuous function over a compact set. Stationarity of $(\varepsilon_t)$, (3.44) and (3.47) imply

$$\begin{aligned}\mathbb{E}\sup_{\theta\in\Theta}(||\mathfrak{A}_i(\theta)||) &= \mathbb{E}\sup_{\theta\in\Theta}(||\mathfrak{A}_0(\theta)||) \\ &\le \sup_{\theta\in\Theta} ||K(\theta)\gamma(\theta)|| + \sup_{\theta\in\Theta} ||A(\theta)|| \, \mathbb{E} \, ||s_0|| < \infty. \end{aligned}\tag{3.48}$$

The components of $\dot{\Sigma}_{t,i}$ are given by

$$\frac{\partial \tilde{\sigma}_t^2}{\partial \gamma'} = \sum_{i=1}^{\infty} B^{i-1} K,$$

$$\frac{\partial \tilde{\sigma}_t^2}{\partial k'} = \sum_{i=1}^{\infty} \gamma' \otimes B^{i-1} - \sum_{i=1}^{\infty} \sum_{j=1}^{i-1} \mathfrak{A}_i B^{i-j-1} \otimes B^{j-1},$$

and

$$\frac{\partial \tilde{\sigma}_t^2}{\partial a'} = \sum_{i=1}^{\infty} s'_{t-i} \otimes B^{i-1} - \sum_{i=1}^{\infty} \sum_{j=1}^{i-1} \mathfrak{A}_i B^{i-j-1} \otimes B^{j-1}$$

Examining the components of $\dot{\Sigma}_{t,i}$, and applying (3.44)-(3.48) we have

$$\sup_{\theta\in\Theta} \left|\left| \frac{\partial \tilde{\sigma}_t^2}{\partial \gamma'} \right|\right| \le \sum_{i=1}^{\infty} \sup_{\theta\in\Theta} N(B(\theta))^{i-1} \sup_{\theta\in\Theta} ||K(\theta)|| \le \sup_{\theta\in\Theta} ||K(\theta)|| \sum_{i=1}^{\infty} r^{i-1} < \infty,$$

$$\begin{aligned}\mathbb{E}\sup_{\theta\in\Theta} \left|\left| \frac{\partial \tilde{\sigma}_t^2}{\partial k'} \right|\right| &\le \sum_{i=1}^{\infty} \sup_{\theta\in\Theta} ||\gamma(\theta)|| \, N(B)^{i-1} \\ &\quad + \sum_{i=1}^{\infty} \sum_{j=1}^{i-1} \mathbb{E}\sup_{\theta\in\Theta}(||\mathfrak{A}_i(\theta)||) N(B)^{i-j-1} N(B)^{j-1} \\ &\le \ \sup_{\theta\in\Theta} ||\gamma(\theta)|| \sum_{i=1}^{\infty} r^{i-1} + \mathbb{E}\sup_{\theta\in\Theta}(||\mathfrak{A}_0(\theta)||) \sum_{i=1}^{\infty} \sum_{j=1}^{i-1} r^{i-2}\end{aligned}$$

$$\leq \sup_{\theta\in\Theta} ||\gamma(\theta)|| \sum_{i=1}^{\infty} N(B)^{i-1} + \mathbb{E}\sup_{\theta\in\Theta}(||\mathfrak{A}_0(\theta)||) \sum_{i=1}^{\infty}(i-1)r^{i-2}$$

$$< \infty,$$

and

$$\mathbb{E}\sup_{\theta\in\Theta} \left|\left|\frac{\partial\tilde{\sigma}_t^2}{\partial a'}\right|\right| \leq \mathbb{E}\,||s_0|| \sum_{i=1}^{\infty} r^{i-1} + \mathbb{E}\sup_{\theta\in\Theta} ||\mathfrak{A}_0(\theta)|| \sum_{i=1}^{\infty}(i-1)r^{i-2} < \infty.$$

This completes the proof of (i).

To prove (ii), one can bound all products of the terms appearing in the proof of (i). Instead, we prove the claim by writing

$$\frac{\partial}{\partial\theta_i}(\tilde{\sigma}_t^2(\theta)) = \frac{\partial}{\partial\theta_i}((I-B)^{-1})K\gamma + (I-B)^{-1}\frac{\partial}{\partial\theta_i}(K)\gamma$$
$$+ \sum_{j=0}^{\infty} \frac{\partial}{\partial\theta_i}(B^j)As_{t-j-1} + \sum_{j=0}^{\infty} B^j \frac{\partial}{\partial\theta_i}(A)s_{t-j-1}.$$

Using (3.44) and (3.46),

$$\left|\left|\frac{\partial}{\partial\theta_i}(\tilde{\sigma}_t^2(\theta))\right|\right|$$
$$\leq \sup_{\theta\in\Theta}\left\{\left|\left|\frac{\partial}{\partial\theta_i}((I-B(\theta))^{-1}\right|\right| ||K(\theta)\gamma(\theta)|| + ||(I-B(\theta))^{-1}|| \left|\left|\frac{\partial}{\partial\theta_i}K(\theta)\right|\right| ||\gamma(\theta)||\right\}$$
$$+ C_1\sup_{\theta\in\Theta} ||A(\theta)|| \sum_{j=0}^{\infty} jr^{j-1}\, ||s_{t-j-1}|| + \sup_{\theta\in\Theta}\left|\left|\frac{\partial}{\partial\theta_i}(A(\theta))\right|\right| \sum_{j=0}^{\infty} N(B^j)\, ||s_{t-j-1}||\,.$$

Each of the supremums above is bounded as the supremum of a function that is continuous with respect to the Euclidean norm over the compact set $\Theta$. Thus,

$$\left|\left|\frac{\partial}{\partial\theta_i}(\tilde{\sigma}_t^2(\theta))\right|\right| \leq K_1\left(\sum_{j=0}^{\infty} jr^{j-1}\, ||s_{t-j-1}|| + \sum_{j=0}^{\infty} r^j\, ||s_{t-j-1}||\right) \tag{3.49}$$

for some finite constant $K_1$ that depends on neither $\theta$ nor $t$. Stationarity of $(\varepsilon_t)$ implies

$$\mathbb{E}\left|\left|\frac{\partial}{\partial\theta_i}(\tilde{\sigma}_t^2(\theta))\right|\right| \leq K_1\mathbb{E}\,||s_1||\left(\sum_{j=0}^{\infty} jr^{j-1} + \sum_{j=0}^{\infty} r^j\right) := K_2 < \infty. \tag{3.50}$$

By (3.49) we conclude

$$\left|\left|\frac{\partial}{\partial\theta_i}(\tilde{\sigma}_t^2(\theta))\right|\right|^2$$

$$\leq K_1^2\left[\left(\sum_{j=0}^{\infty} jr^{j-1}\left|\left|s_{t-j-1}\right|\right|\right)^2 + 2\left(\sum_{j=0}^{\infty} jr^{j-1}\left|\left|s_{t-j-1}\right|\right|\right)\left(\sum_{j=0}^{\infty} r^j\left|\left|s_{t-j-1}\right|\right|\right)\right]$$

$$+K_1^2\left[\left(\sum_{j=0}^{\infty} r^j\left|\left|s_{t-j-1}\right|\right|\right)^2\right].$$

Thus, using **A6**, we have that

$$\mathbb{E}\left|\left|\frac{\partial}{\partial\theta_i}(\tilde{\sigma}_t^2(\theta))\right|\right|^2$$

is bounded by

$$K_1^2\mathbb{E}\left|\left|s_1\right|\right|^2\left[\left(\sum_{j=0}^{\infty} jr^{j-1}\right) + 2\left(\sum_{j=0}^{\infty} jr^{j-1}\right)\left(\sum_{j=0}^{\infty} r^j\left|\left|s_{t-j-1}\right|\right|\right)^2\left(\sum_{j=0}^{\infty} r^j\right)^2\right] < \infty.$$

This completes the proof of (ii).

The components of $\ddot{\Sigma}_{t,ij}$ are given by

$$\frac{\partial \text{vec}}{\partial a'}\left(\frac{\partial\tilde{\sigma}_t^2}{\partial\gamma'}\right) = -\sum_{i=1}^{\infty}\sum_{j=1}^{i-1} KB^{i-j-1}\otimes B^{j-1},$$

$$\frac{\partial \text{vec}}{\partial k'}\left(\frac{\partial\tilde{\sigma}_t^2}{\partial\gamma'}\right) = \sum_{i=1}^{\infty} I_m\otimes B^{i-1} - \sum_{i=1}^{\infty}\sum_{j=1}^{i-1} KB^{i-j-1}\otimes B^{j-1},$$

$$\frac{\partial \text{vec}}{\partial k'}\left(\frac{\partial\tilde{\sigma}_t^2}{\partial k'}\right) = -\sum_{i=1}^{\infty} C\left[(\gamma\otimes I_{m^2})\left(\sum_{j=1}^{i-1} B^{i-j-1}\otimes B^{j-1}\right)\right]$$

$$-\sum_{i=1}^{\infty}\sum_{j=1}^{i-1} C\left(I_m\otimes \text{vec}B^{j-1}\right)\left(B^{i-j-1}\otimes\gamma' - \sum_{\ell=1}^{i-j-1} B^{i-j-\ell-1}\otimes\mathfrak{A}_i B^{\ell-1}\right)$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{i-1} C\left(\text{vec}\left(\mathfrak{A}_i B^{i-j-1}\right)\otimes I_{m^2}\right)\left(\sum_{\ell=1}^{j-1} B^{j-\ell-1}\otimes B^{\ell-1}\right),$$

$$\frac{\partial \text{vec}}{\partial a'}\left(\frac{\partial\tilde{\sigma}_t^2}{\partial k'}\right) = -\sum_{i=1}^{\infty} C\left[(\gamma\otimes I_{m^2})\left(\sum_{j=1}^{i-1} B^{i-j-1}\otimes B^{j-1}\right)\right]$$

$$-\sum_{i=1}^{\infty}\sum_{j=1}^{i-1} C\left(I_m \otimes \mathrm{vec} B^{j-1}\right)\left(B^{i-j-1}\otimes s'_{t-i} - \sum_{\ell=1}^{i-j-1} B^{i-j-\ell-1}\otimes \mathfrak{A}_i B^{\ell-1}\right)$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{i-1} C\left(\mathrm{vec}\left(\mathfrak{A}_i B^{i-j-1}\right)\otimes I_{m^2}\right)\left(\sum_{\ell=1}^{j-1} B^{j-\ell-1}\otimes B^{\ell-1}\right),$$

$$\frac{\partial \mathrm{vec}}{\partial a'}\left(\frac{\partial \tilde{\sigma}_t^2}{\partial a'}\right) = -\sum_{i=1}^{\infty} C\left[\left(s_{t-i}\otimes I_{m^2}\right)\left(\sum_{j=1}^{i-1} B^{i-j-1}\otimes B^{j-1}\right)\right]$$

$$-\sum_{i=1}^{\infty}\sum_{j=1}^{i-1} C\left(I_m \otimes \mathrm{vec} B^{j-1}\right)\left(B^{i-j-1}\otimes s'_{t-i} - \sum_{\ell=1}^{i-j-1} B^{i-j-\ell-1}\otimes \mathfrak{A}_i B^{\ell-1}\right)$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{i-1} C\left(\mathrm{vec}\left(\mathfrak{A}_i B^{i-j-1}\right)\otimes I_{m^2}\right)\left(\sum_{\ell=1}^{j-1} B^{j-\ell-1}\otimes B^{\ell-1}\right).$$

Applying (3.44)-(3.48) to the components of $\ddot{\Sigma}_{t,ij}$ we have terms identical to those appearing in the proof of (i), plus additional terms of the form

$$\mathbb{E}\sup_{\theta\in\Theta}(||\mathfrak{A}_0(\theta)||)\sum_{i=1}^{\infty}\sum_{j=1}^{i-1}\sum_{\ell=1}^{i-j-1} r^{i-3} = \mathbb{E}\sup_{\theta\in\Theta}(||\mathfrak{A}_0(\theta)||)\sum_{i=1}^{\infty}\sum_{j=1}^{i-1}(i-j-1)r^{i-3}$$

$$\leq \mathbb{E}\sup_{\theta\in\Theta}(||\mathfrak{A}_0(\theta)||)\sum_{i=1}^{\infty} i^2 r^{i-3} < \infty.$$

□

**Lemma 17** *The matrix $\tilde{\Sigma}_t^{-1}$ satisfies $||\tilde{\Sigma}_t^{-1}|| \leq M_1$ for some finite constant $M_1$ that depends on neither $\theta$ nor $t$.*

*Proof:* We follow again the arguments in Comte and Lieberman (2003). Let $X$ and $Y$ denote arbitrary $d \times d$ matrices with $X > 0$, $Y \geq 0$. Then,

$$\begin{aligned}
0 \leq \mathrm{tr}[(X+Y)^{-2}] &= \mathrm{tr}[(X+Y)^{-1}(X+Y)^{-1}] \\
&= ||(X+Y)||^2 \\
&= ||X^{-1/2}(I_d + X^{-1/2}YX^{-1/2})^{-1}X^{-1/2}||^2 \\
&= \mathrm{tr}(X^{-2}(I_d + X^{-1/2}YX^{-1/2})^{-2}) \\
&= ||X^{-2}(I_d + X^{-1/2}YX^{-1/2})^{-2}||^2 \\
&\leq ||X^2||^2||(I_d + X^{-1/2}YX^{-1/2})^{-2}||^2 \\
&= \mathrm{tr}(X^{-4})\mathrm{tr}[(I_d + X^{-1/2}YX^{-1/2})^{-4}]^{1/2}.
\end{aligned}$$

All eigenvalues of $I_d + X^{-1/2}YX^{-1/2}$ are greater than one, so those of its inverse are in (0,1] as are those of any power of the inverse. This implies

$$\mathrm{tr}[(I_d + X^{-1/2}YX^{-1/2})^{-4}] < d,$$

and thus,

$$0 \leq \mathrm{tr}[(X+Y)^{-2}] \leq (\sqrt{d})\mathrm{tr}(X^{-4}). \tag{3.51}$$

From (3.2), each $\tilde{\Sigma}_t$ has a representation of the form

$$\tilde{\Sigma}_t(\theta) = C(\theta) + Y(\theta),$$

where $Y(\theta) \geq 0$, and $C(\theta) > 0$. Using (3.51), and using that $C(\theta)^{-2}$ is continuous with respect to the Euclidean norm over the compact set $\Theta$, we have

$$||\tilde{\Sigma}_t^{-1}(\theta)||^2 \leq (\sqrt{d})||C(\theta)^{-2}|| < M_1^2,$$

for some finite constant $M_1$ that depends on neither $\theta$ nor $t$.

$\square$

## REFERENCES


[1] K. Abadir and J. Magnus, *Matrix Algebra*, Cambridge University Press, Cambridge New York (2005).

[2] V. Arsigny, P. Fillard, X. Pennac, and N. Ayache, *Geometric Means in a Novel Vector Space Structure on Symmetric Positive-Definite Matrices*, SIAM Journal on Matrix Analysis and Applications, **29** (2007), 328-347.

[3] I. Berkes, L. Horváth, and P. Kokoszka, *GARCH Processes: Structure and Estimation*, Bernoulli, **9** (2004), 210-227.

[4] I. Berkes, L. Horváth, and P. Kokoszka, *Probabilistic and Statistical Properties of GARCH Processes*, Fields Institute Communications, American Mathematical Society, (2004), 409-430.

[5] D. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*, Princeton University Press, Princeton, New Jersey (2005).

[6] P. Bickel and K. Docksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, Prentice Hall, Englewood Cliffs, New Jersey (1977).

[7] P. Billingsley, *The Lindeberg-Levy Theorem for Martingales*, Proceedings of the American Mathematical Society, **12** (1961), 788-792.

[8] P. Billingsley, *Probability and Measure*, New York: John Wiley & Sons, Inc. (1995).

[9] T. Bollerslev, *Generalized Autoregressive Conditional Heteroskedasticity*, Journal of Econometrics **31** (1986), 307-327.

[10] T. Bollerslev, *Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model*, Review of Economics and Statistics, **72** (1990), 498-505.

[11] P. Bougerol and N. Picard, *Strict Stationarity of Generalized Auto-Regressive Processes*, Annals of Probability, **20** (1992), 1714-1729.

[12] F. Boussama, *Ergodicite des Chaines de Markov a Valeurs dans une Variete Algebrique: Application aux Modeles GARCH Multivaries*, Comptes Rendus de L'Academie de Sciences Paris, Serie I **343** (2006), 275-278.

[13] P.J. Brockwell and R.A. Davis, *Time Series: Theory and Methods*, New York: Springer, 2nd Edition (1991).

[14] C. Brooks, *Introductory Econometrics for Finance*, Cambridge, 2nd Edition (2008).

[15] L.E.J. Brouwer, *Beweis der Invarianz des n-Dimensionalen Gebiets*, Mathematische Annalen **71** (1912), 305-315.

[16] M. Carvalho, *A Smooth Transition Multivariate GARCH Approach to Contagion*, http://ssrn.com/abstract=1080229, (2007).

[17] F. Comte and O. Lieberman, *Asymptotic Theory for Multivariate GARCH Processes*, Journal of Multivariate Analysis, **84** (2003), 6184.

[18] Y.A. Davydov, *Mixing Conditions for Markov Chains*, Theory of Probability and Applications **18** (1973), 313-328.

[19] J.G. Dennis, E. Hansen, and A. Rahbek, *ARCH Innovations and Their Impact on Cointegration Rank Testing*, Working Paper no. 22, Centre for Analytical Finance, University of Aarhus, (2002).

[20] R.F. Engle, *Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation*, Econometrica, **50** (1982), 987-1007.

[21] R.F. Engle, V.K. Ng, and M. Rothschild, *Asset Pricing with a Factor ARCH Covariance Structure: Empirical Estimates for Treasury Bills.*, Journal of Econometrics **45** (1990), 213-238.

[22] R.F. Engle and K.F. Kroner, *Multivariate Simultaneous Generalized ARCH*, Econometric Theory **11** (1995), 122-150.

[23] R.F. Engle and J. Mezrich, *GARCH for Groups*, Risk **9** (1996), 36-40.

[24] C. Francq, L. Horváth, and J-M. Zakoïan, *Merits and Drawbacks of Variance Targeting in GARCH Models*, Preprint MPRA 15143, University Library of Munich, Germany (2009).

[25] C. Francq and J-M. Zakoïan, *A Tour in the Asymptotic Theory of GARCH Estimation*, Handbook of Financial Time Series, Berlin: Springer (2009).

[26] C. Francq and J-M. Zakoïan, *Modeles GARCH, Structure, Inference Statistique et Applications Financieres*, Economica (2009).

[27] B. Fristedt and L. Gray, *A Modern Approach to Probability Theory*, Birkhäuser Boston (1997).

[28] J. Gross, *Linear Regression*, Springer-Verlag, Berlin (2003).

[29] V. Guillemin and A. Pollack, *Differential Topology*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1974).

[30] C. Hafner and A. Preminger, *Asymptotic Theory for a Factor GARCH Model*, Econometric Theory, **25** (2009a), 336-363 .

[31] C. Hafner and A. Preminger, *On Asymptotic Theory for Multivariate GARCH Models*, Journal of Multivariate Analysis, **100** (2009b), 2044-2054.

[32] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, MA, (1991).

[33] S. Ling and M. McAleer, *Asymptotic Theory for a Vector ARMA-GARCH Model*, Econometric Theory, **19** (2003), 642-674.

[34] H. Lutkepohl, *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin (2005).

[35] J. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons, Inc. (2001).

[36] B. Mandelbrot, *The Variation of Certain Speculative Prices*, Journal of Business, **36** (1963a), 394-419.

[37] B. Mandelbrot, *New Methods in Statistical Economics*, Journal of Political Economy, **61** (1963b), 421-440.

[38] S.P. Meyn and R.L. Tweedie, *Markov Chains and Stochastic Stability*, Springer, London (1993).

[39] A. Mokkadem, *Proprietes de Melange des Processes Autoregressifs Polynomiaux*, Annales de l'Inst. Henri Poincare, Probabilites et Statistiques **50** (1990), 219-260.

[40] D.B. Nelson, *Stationarity and Persistence in the GARCH(1,1) Model*, Econometric Theory **6** (1990), 318-334.

[41] K. Piontek, *Weryfikacja Technik Prognozowania Zmiennouci na Podstawie Szeregow Czasowych*, http://www.kpiontek.ae.wroc.pl/inwest03.pdf (2003).

[42] J. E. Potter, *Matrix Quadratic Solutions*, SIAM Journal on Applied Mathematics **14** (1966), 496501.

[43] R. Rao, *Relations Between Weak and Uniform Convergence of Measures with Applications*, Annals of Mathematical Statistics **33** (1962), 659-680.

[44] M. Rosenblatt, *A Central Limit Theorem and a Strong Mixing Condition*, Proceedings of the National Academy of Science of the USA, **42** (1956), 43-47.

[45] A. Silvennoinen and T. Terasvirta *Multivariate GARCH Models*, Handbook of Financial Time Series, T.G. Anderson et al., Springer, New York (2009), 201-229.

[46] R. Stelzer, *On the Relation Between the Vec and BEKK Multivariate GARCH Models*, Econometric Theory **24** (2008), 1131-1136.

[47] D. Straumann and T. Mikosch, *Quasi-Maximum Likelihood Estimation in Heteroscedastic Time Series: A Stochastic Recurrence Equation Approach*, Annals of Statistics **34** (2006), 2449-2495.

[48] Y. Tse and K. Tsui, *A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model With Time-Varying Correlations*, Journal of Business and Economic Statistics **20** (2002), 351-362.

[49] J. Wang and E. Zivot, *Modeling Financial Time Series with S-PLUS*, Business and Economics (2006).

[50] H. Wold, *A Study in the Analysis of Stationary Time Series*, Uppsala: Almqvist & Wiksell (1938).

[51] A. Worthington and H. Higgs, *Transmission of Equity Returns and Volatility in Asian Developed and Emerging Markets: A Multivariate GARCH Analysis*, International Journal of Finance & Economics, **9** (2004), 71-80.

[52] F. Zhang, *Matrix Theory: Basic Results and Techniques*, Springer-Verlag New York Inc. (1999).