Surveys in Mathematics and its Applications

ISSN 1842-6298 (electronic), 1843-7265 (print)
Volume 21 (2026), 107 – 156

THE FUNDAMENTALS OF GARCH

Jorge Bustamante and Adrian Pizzinga

Abstract. This paper goes over very first theoretical aspects regarding the GARCH process. Our survey comprises: (i) a literature review that indicates the parts of the founding theory covered by many standard textbooks (and some other parts that are not); and (ii) detailed and classroom level proofs of the existence and uniqueness of a stationary solution for the GARCH defining equations. We aim to contribute to the syllabus of typical graduate level courses on time series analysis, time series econometrics, financial econometrics, statistical methods in finance, risk theory, etc. The target audience includes both students and lecturers from these courses.

1 Introduction

The *autoregressive conditional heteroscedasticity* process (ARCH process, from now on) was introduced in an celebrated article by Robert Engle, as new statistical model for the heteroscedasticity (non-constant variance) of some macroeconomic time series (cf. Engle [18]). Since then, the scientific community has seen an increase in relevance on the subject, with greater emphasis on its applications to financial econometrics. Bollerslev [4] extended Engle's original model (ARCH) to its well-known generalized version (or GARCH), which has proven more adequate for describing volatility of stocks and exchange-rate returns. Other very important articles, on either ARCH or GARCH, are the ones by Weiss [42, 43], Milhøj [31], Engle & Bollerslev [19], Nelson [35, 36] and Bollerslev & Wooldridge [5]. Traditional references on time series analysis with excellent coverage of the GARCH model, along with its notable extensions (EGARCH, ARCH-M, etc.), include Harvey [25], Hamilton [24] and Enders [17]. Important econometrics books are as well worth-citing, such as Campbell et al. [11], Mills [32], Davidson & MacKinnon [15] and Greene [22].

The stochastic difference equations that dictate the GARCH(p, q) dynamics for a discrete time stochastic process $(Y_t)_{t\in\mathbb{Z}}$ read as

2020 Mathematics Subject Classification: 60G10; 60G12; 97K50.
Keywords: GARCH process; stationarity; existence; P-almost surely uniqueness.
This work was supported by Coordenação de Aperfeiçoamento de Pessoal do Nível Superior (CAPES) and Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

$$
\begin{aligned}
Y_t &= \sqrt{h_t}\,\varepsilon_t, \quad \varepsilon_t \sim \mathrm{IID}(0,1),\\
h_t &= \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j},
\end{aligned}
\tag{1.1}
$$

where: (i) p and q define the orders of the process; (ii) $\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$ are parameters generally subjected to the constraints $\alpha_0 > 0$, $\alpha_i \geq 0$ for every $i = 1, \ldots, p$, and $\beta_j \geq 0$ for every $j = 1, \ldots, q$; and (iii) the notation $\mathrm{IID}(0,1)$ means that the random variables of the "input" process $(\varepsilon_t)_{t\in\mathbb{Z}}$ are all independent and identically distributed with zero mean and unity variance. To obtain the original ARCH(p) process from eqs.(1.1), just set $\beta_1 = \ldots = \beta_q = 0$. Standard literature frequently considers more general version of eqs.(1.1), in which $(Y_t)_{t\in\mathbb{Z}}$ follows an autoregressive moving average (ARMA) process with a GARCH error term, together with explanatory variables. Since these extensions are not in this paper's interests, and neither are the theory and methods concerning parameter estimation using time series data, we shall not delve into these topics here.

Now, we begin to analyze the way that most *classroom textbooks* – the ones formerly cited and others that will be discussed later in this paper – cover some very beginning elements of the GARCH theory. The first instance we would like to examine regards the restrictions on the parameters. A closer look at eqs.(1.1) reveals that, in fact, such assumptions are pretty reasonable, since every negative value of h_t must clearly be avoided. But the following two questions we judge to be fair: why exactly these assumptions, and not others possibly a bit less constraining, such as allowing α_0 to take the value 0 (zero), just like the other parameters? Or, perhaps, even more constraining, such as forcing all parameters to be strictly positive? From the practical standpoint, the GARCH process eventually becomes a statistical model to be estimated with actual time series data. Then, our standard statistical theory background reminds us that taking boundary values off the parametric space leads quite generally to convenient asymptotic results for hypothesis testing and other inference purposes – see Rohatgi & Saleh [38], pages 419-421, and Casella & Berger [12], page 516.

Let us discuss another example. It is very frequent in textbooks the derivation of GARCH properties directly from eqs.(1.1), without imposing any further assumption. To illustrate, we recall that, in order to obtain the expected value of the random variable of Y_t in eqs.(1.1) (for any $t \in \mathbb{Z}$), the following is usually done:

$$
\begin{aligned}
\mathrm{E}\,(Y_t) &= \mathrm{E}\left(h_t^{1/2}\varepsilon_t\right) = \mathrm{E}\left(\mathrm{E}\left(h_t^{1/2}\varepsilon_t \mid Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots\right)\right)\\
&= \mathrm{E}\left(h_t^{1/2}\mathrm{E}\,(\varepsilon_t \mid Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots)\right)\\
&= \mathrm{E}\left(h_t^{1/2}\mathrm{E}\,(\varepsilon_t)\right) = \mathrm{E}\left(h_t^{1/2}{\times}0\right) = \mathrm{E}\,(0) = 0.
\end{aligned}
\tag{1.2}
$$

The justifications generally given for eqs.(1.2): the second equality follows from the Law of Iterated Expectations, the third ensues from conditional expectation properties (since h_t depends entirely on $Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots$, it can be regarded as a constant), the fourth comes from independence between ε_t and $Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots$, and the fifth is trivial in view of the assumptions over ε_t in eqs.(1.1).

But, once again, it is worth questioning: even under non-negative parameters, shall the conditional expectation

$$\mathrm{E}\left(Y_t \mid Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots\right) = \mathrm{E}\left(h_t^{1/2}\varepsilon_t \mid Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots\right)$$

always be well-defined? We must notice that, typically, the existence of $\mathrm{E}\left(Y_t\right)$ as a real number, which would guarantee the existence of $\mathrm{E}\left(Y_t \mid Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots\right)$, is not proven (again: *typically*) in time series or econometrics books. As a matter of fact, Shiryaev [39], chapter II, section 7, teaches us that $\mathrm{E}\left(Y_t\right)$ need neither be finite nor even well-defined in order to such a conditional expectation exist. However, to engage in beyond that, let us say through writings like eqs.(1.2), particularly the property used in the third equality, the unconditional expectation (being explicit: unconditional expectation of $Y_t = h_t^{1/2}\varepsilon_t$) has to be more than well-defined – it need be finite[1] (cf. Billingsley [3], Theorem 34.3; or Shiryaev [39], page 216, Property K*; or Chung [14], Theorem 9.1.3). The problem we see is that, frequently and promptly, the GARCH process is assumed to be second-order stationary – and, in the sequel, "facts" like eqs.(1.2) are derived. We nonetheless insist: whether second-order or even strictly, is such process *always* stationary?

Another example: How can one guarantee that ε_t is indeed independent of the random variables $Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots$ in eqs.(1.1)? We once again advocate the question is just pertinent, mainly if we recall that the sufficient condition – specifically the assertion that each one of the random variables $Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots$ depends itself solely on "past terms" of ε_t – is barely proven in the aforementioned references. In other words, the solution of the first difference equation in (1.1) or conditions under which that equation can be solved are not addressed in textbooks[2] (we stress once more: *typically*!).

Actually, this very last sentence leads to a more essential question: why and under what conditions do random variables h_t and, therefore, the whole stochastic processes proposed by both identities in eqs.(1.1) *exist mathematically*? Is the existence of a solution always guaranteed for those difference equations? A fine understanding of this point may well be a good aid to answer our previous questions – besides

[1] As illustrated with the probability exercise solved in section A of the Appendix, even though may $\mathrm{E}\left(\sqrt{h_t}\right)$ be well-defined in $\overline{\mathbb{R}} = [-\infty, +\infty]$ and $\mathrm{E}\left(\varepsilon_t\right)$ be finite, these two conditions, alone, do not imply the existence of $\mathrm{E}\left(Y_t\right) = \mathrm{E}\left(\sqrt{h_t} \times \varepsilon_t\right)$. To prove that the latter holds, more is needed.

[2] One exception is the book Brockwell & Davis [10], chapter 7, where some considerable attention is dedicated to at least the most basic ARCH model of order $p = 1$.

allowing a rigorous understanding of the probably most iconic identity within this whole subject: $h_t = \text{Var}(Y_t \mid Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots)$.

It should be noted, on the other hand, that issues like the above are not really shared by the classic ARCH/GARCH articles. Even though a full review of the latter is not quite the focus of this survey, we still find room to consider at least two remarkable examples:

- Before tackling analytical expressions for ordinary moments – that is, the deepening on theoretical developments such as we have seen in eqs.(1.2) –, Milhøj [31], Theorem 1, establishes the conditions under which the stochastic difference equations, those that define the ARCH process (see eqs.(1.1) again), admit a solution.

- Bollerslev [4], with his GARCH process, also and always stands for the same mathematical diligence. His strategy is initially inspired by the polynomials on the lag operator, well-known from the theory of ARMA models (for the latter, each of the time series books formerly cited in this paper is excellent source). Then, the former approach of Milhøj [31] is duly cited. In fact even more: Bollerslev eventually makes very good use of it.

It is also just fair to mention the literature that has addressed along the decades the questions posed here using pretty general extensions of the GARCH model. Excellent examples are Bougerol & Picard [6], Ling & McAleer [30], Zaffaroni [45], Aue et al. [1] and Doukhan & Neumann [16]. However, the simple fact that each of these papers are indeed very technical may be a deterrent component for some audiences. Especially the first-year graduate students, who in general have not been introduced to the more advanced time series literature.

With all that being said, the main intent of this survey is finally stated: never to replace, but rather to complement usual textbooks used in graduate courses in time series analysis and econometrics, in which concerns the basic elements of GARCH theory. Specifically, we shall not advance to issues or topics other than: (i) proving existence and uniqueness of a GARCH process through direct use of eqs.(1.1); (ii) addressing first two moments (mean and variance); and (iii) discussing, in greater detail, the actual conditions for stationarity – either second-order or strict. In fact, of these three aspects, the first and third (mainly the first) we understand the usual classroom literature lacks on.

Concretely speaking, we dedicate efforts to the following activities:
(a) To prove that a GARCH process – that is, a solution for eqs.(1.1) – indeed exists, provided that the proper restrictions onto the parameters are in effect.
(b) To prove that the same restrictions, besides guaranteeing existence, imply both notions of stationarity. Moreover, under some additional and at times quite natural assumptions, such solution for eqs.(1.1) is in truth unique with probability 1.

(c) To prove the reciprocal: once the existence of a second-order stationary GARCH process is assumed, the parameters must necessarily satisfy the same former restrictions.

Carlos Kubrusly once wrote *"Sketchy proofs seem to have been a perpetual complaint"* (Kubrusly [29], page vii). As we definitely sign for that (and perhaps might you, the reader of this survey), our proofs have been designed to make it to the classroom. For instance, before eventually reaching the full GARCH specification, all tasks (a), (b) and (c) announced above shall be first implemented with the ARCH version of eqs.(1.1). Also with the purpose of simplifying the content, we add footnotes to all phrasings considered more technical. And finally, section B of the Appendix revisits many well-established probability results that help us in proving the main theorems. Such "lemmas" comprise really important material that every probability & statistics graduate student shall eventually be well-versed in.

Even though considered in research articles and some advanced books, many demonstrations included in this survey are frequently absent from standard time series/econometrics classroom literature – and we have tried to address each proof under the most accessible writing we could use. So, from the methodological standpoint, there is not exactly novelty in this manuscript, but rather – and to say again – the intent of reaching either the student or the professor interested in understanding this: the GARCH process is never a heuristic; it is surely a just and solid mathematical model.

One final comment before proceeding. The proofs we are about to present *never require conditional expectations*, such as those quoted in eqs.(1.2). This makes some parts of our derivations, though not exactly faster on one hand, certainly *simpler* from a math standpoint: no conditional expectation theory whatsoever is needed to understand the forthcoming.

Our paper is arranged as follows. In section 2, we expand the bibliographic review, focusing entirely on time series and econometrics books. Section 3 fills with the main theorems – first for ARCH, then for GARCH. Section 4 briefly addresses the *Integrated* GARCH, or IGARCH process, which despite the name *is not* a non-stationary solution for eqs.(1.1)[3].

2 Review of the classroom literature

Since the inception of Engle [18]'s paradigm, the GARCH process (or GARCH *model*, as usually worded in the statistics context) has conquered a spot in virtually every time series analysis or econometrics book. In this section we examine these two major areas of the literature. Both include complete discussions of the intuitions behind the

[3] As a matter of fact, analogously to what we notice for ARMA and ARIMA models, it is not exactly clear how to prove the existence of something such as a "non-stationary GARCH process" on the *whole* discrete time set; that is, for $t \in \mathbb{Z}$.

GARCH analytical expressions, the derivation of first two unconditional moments (including auto-covariance functions), and the (quasi) maximum likelihood estimation theory. And, despite the aforementioned missing on which regards proofs for existence and uniqueness of a stationary solution for eqs.(1.1), textbooks never omit themselves from discussing other aspects that are quite relevant in their own, such as: (a) sufficient conditions for finite third- and fourth-order moments (at times, proofs are given; other times, more technical literature is referred to); (b) the fact that every squared GARCH(p,q) process follows a special kind of ARMA($\max\{p,q\},q$) process; and (c) well-known model extensions, such as the *exponential* GARCH (EGARCH), *threshold* ARCH (TARCH) and ARCH-*in-mean* (ARCH-M) models.

2.1 Time series books

Let us reconsider the three first books cited in section 1 by highlighting some particular details and their didactic contributions to the subject. We begin with Harvey [25], chapter 8, who offers a quite rich coverage of statistical models for handling *"stochastic movements in variance"* – with special attention to ARCH, since the latter and stochastic volatility models *"have proven to be extremely useful in modeling movements in financial time series"*. Prior to addressing in greater detail higher-order ARCH and GARCH equations, Harvey focuses on ARCH(1)'s basic properties: closed expressions for the conditional variance of Y_t given $Y_{t-j}, Y_{t-j-1}, Y_{t-j-2}, \ldots$, for every $j \geq 1$; the unconditional variance of Y_t; the fact that $(Y_t)_{t\in\mathbb{Z}}$ is a *martingale difference* process (that is: $\mathrm{E}\,(Y_t|Y_{t-1}, Y_{t-2}, \ldots) = 0$ for every t); the fact that $\left(Y_t^2\right) \sim \mathrm{AR}(1)$; and that the unconditional kurtosis of Y_t is finite if and only if $\alpha_1^2 < \frac{1}{3}$.

Now, we take a look at the approach of Hamilton [24], chapter 21. Interestingly, Hamilton begins by defining the ARCH(p) process as any white noise $(Y_t)_{t\in\mathbb{Z}}$ whose square follows a stationary AR(p) process. On the course of his theoretical development, it is eventually suggested that eq.(1.1) given in this paper is a better definition, since *"it is often more convenient to use an alternative representation"* for $(Y_t)_{t\in\mathbb{Z}}$. One that *"imposes slightly stronger assumptions about the serial dependence"*. Hamilton also explains that, in order to both possible ARCH definitions work properly – in which concerns $Y_t^2 \sim \mathrm{ARMA}$ –, the fourth unconditional moment of the random variables Y_t must be finite. And then, every probability and statistics aspects, firstly carried with ARCH, is duly extended to the full GARCH specification.

Let us now spend some time with the book by Enders [17], chapter 3, whose starting point is to recognize that any stochastic process with some kind of heteroscedasticity may well follow a defining equation such as

$$Y_t = X_t \varepsilon_t. \tag{2.1}$$

In eq.(2.1), $(\varepsilon_t)_{t\in\mathbb{Z}}$ is the same IID second-order process as we have been assuming, whereas $(X_t)_{t\in\mathbb{Z}}$ has itself a proper serial correlation structure and is such that X_t

and ε_t are independent for any t. The problem of eq.(2.1) resides on its enormous generality. In Enders's words, *"it assumes a specific cause for the changin of variance"*, and in most cases, one *"may not have a firm theoretical reason for selecting"* an adequate $(X_t)_{t\in\mathbb{Z}}$ for some given situation. Therefore, Engle's (1982) rationality is evoked, and mathematical derivations and empirical examples follow similarly to Harvey (1993) and Hamilton (1994).

Three more time series books are worth-discussing. The first is Box et al. [8][4], chapter 10, section 10.2, where the theory and empirical examples ensue much closer to the three books previously reviewed. But two particular and strengthening elements show up: a quite richer literature review and an explicit mention (although still without a proof) that assuming $\sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j < 1$ *"ensures that"* the solution for eq.(1.1) is *"covariance stationary with finite unconditional variance"*.

Brockwell & Davis [10], chapter 7, define the ARCH(p) process much similarly to our eq.(1.1), but with the additional assumption that $(\varepsilon_t) \sim$ IID N$(0,1)$. As already cited in our previous footnote 2, as regards the initial theory, they confine efforts to the basic ARCH(1), using its recursive nature to obtain a proper solution of the defining difference equation. Some fundamental and very interesting properties are therefore explored, such as the fact that ARCH processes, even though white noise, are not IID. Or that the unconditional distribution of Y_t can never be Gaussian, and that the fourth unconditional moment of an ARCH(1) process is finite if and only if $3\alpha_1^2 < 1$. They also cite (yet no proof is given) the following distributional ARCH(1)'s tail property formerly discussed by Engle (1982): for every $\alpha_1 \in (0,1)$, there exists $k \in \mathbb{N}$ such that $\mathrm{E}\left(Y_t^{2k}\right) = \infty$.

Shumway & Stoffer [40], chapter 5, section 5.3, the last time series book to be reviewed in this paper, builds the ARCH model (once again, everything starts with $p = 1$), bearing in mind the concrete and motivating case of a return of some financial asset. In their notation:

$$r_t = \frac{x_t - x_{t-1}}{x_{t-1}},$$

where x_t represents the asset price in time t. They promptly assume that $\alpha_0 > 0$ and $\alpha_1 \geq 0$ in order to derive desirable properties. The reminder of Shumway & Stoffer's presentation goes pretty much the same paths of the former reviewed books, but with greater emphasis on computational aspects. They also briefly discuss the *asymmetric power* ARCH, or APARCH model, which is a quite general specification for dealing with asymmetric effects of positive and negative returns on volatility.

Other time series books that also address the ARCH/GARCH theory and corresponding statistical methods, either with regular whole chapters and sections, or offering at least a brief glimpse (at times a few exercises), are Granger & Newbold

[4]This is actually the most recent edition of the pioneering and influential treatise by George E. P. Box and Gwilym M. Jenkins - cf. Box & Jenkins [7].

[21], Hendry [27], Fuller [20], Chatfield & Xing [13], Tsay [41], Montgomery et al. [34], Neusser [37] and Mills [33].

2.2 Econometrics books

Let us begin with Campbell et al. [11], who develop in their chapter 12 some pretty thorough content on nonlinear models used in financial statistics. Beginning in section 12.1 with model description, examples and statistical tests for nonlinear structures tests, the authors eventually get into time-varying volatility. The first step of their cohesive approach is to establish the specification for a general model and some corresponding properties – the closing one, which is the pretty straightforward (but also pretty realistic in view of stylized facts culled from financial data during decades), is the fact of the unconditional kurtosis of such models are always strictly greater than 3. This all happens in the beginning of section 12.2, whereas the whole subsection 12.2.1 fills with the theory of arguably the two most influential models for conditional variance, namely the duo ARCH/GARCH and the stochastic volatility model. The former clearly receives the higher priority: topics such as stationary (at least a good discussion), extensions to the IGARCH and EGARCH, inclusion of explanatory variables in the variance equation, and parameter estimation (under both Gaussian and non-Gaussian error terms) are all included. Campbell et al. still find the time for GARCH multivariate extensions (cf. subsection 12.2.2) and models that recognize changes on the conditional mean that originate from conditional variance, like the ARCH-M model (cf. 12.2.3).

Now we move to the three books previously mentioned in section 1. Among them, Mills [32] has, as well as Campbell et al. [11], gained the status of a classic financial econometrics treatise. In his chapter 4, Mills discusses important time-varying conditional variance models; the most preeminent one is once again the GARCH model, whose theory derives pretty closely to some of the time series books previously reviewed in subsection 2.1. One interesting distinction though is Mill's effort to bring a stronger financial background; see the connection between the GARCH process and the asset pricing theory that is built in his section 4.4.5.

As regards Davidson & MacKinnon [15], chapter 13, section 13.6, who deliver a quite didactic effort, three aspects are especially distinguishing. At first place, their ARCH's defining equation (13.74) promptly recognizes that all parameters $\alpha_0, \alpha_1, \ldots, \alpha_p$ need to be strictly positive – probably the authors have borne in mind the statistical inference problems that would arise, should the boundary value 0 (zero) be allowed in the parametric space (see again our brief discussion on precisely this point in section 1). Secondly, Davidson & MacKinnon's approach for the estimation of GARCH models is particularly very rich, ranging from fine details of the likelihood function to useful comments about typical numerical drawbacks. The third and last interesting aspect is that the authors offer a guide to the computational simulation of both ARCH and GARCH – an attitude, it is worth to say, they also have when discussing other time series models in their book.

Finally, we review the book by Greene [22], chapter 11, section 8, who makes some interesting comments, such as the resemblance between the ARCH/GARCH conditional variance formula and the MA/ARMA models' defining expressions. One specific contribution by Greene is the space reserved to discuss possible connections between the ARCH-M model and the capital asset pricing model (CAPM) theory (something that Campbell et al. [11] briefly explore as well). As regards the computational aspects about parameter estimation, Greene's excellent coverage pretty much compares to Davidson & MacKinnon [15]'s.

Given that they include a few pages to cover the essentials of ARCH or GARCH, the books by Johnston & DiNardo [28], Hayashi [26], Gujarati et al. [23] and Wooldridge [44] are also worth-mentioning.

3 Existence and uniqueness of GARCH processes

Now, time has come for us to address the main technical points of this paper: to prove that, under suitable and actually quite well-established *sufficient* conditions, the GARCH defining eqs.(1.1) admit a stationary solution; to investigate the additional conditions (if any) that implies uniqueness of such stationary solution; and to deepen on whether the conditions are also *necessary* for any given solution for eq.(1.1) be stationary.

We first tackle the original ARCH process. Promptly after, we move to GARCH. With both, we split our proofs into "if" parts and "only if" parts. The former establish parameters restrictions under which there exists a P-almost surely unique and stationary solution for eq.(1.1), whereas the latter concentrate on showing that such restrictions (at least some of them) necessarily hold once a stationary solution is supposed to exist.

A suggestion (perhaps a reminder) to the reader: section B of the Appendix contains detailed statements for many probability results – such as some famous inequalities and limit theorems – that help us through this whole section 3.

3.1 ARCH

Let us rewrite the defining difference equation for the ARCH(p) model in a more convenient expression:

$$Y_t = \left[\alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2\right]^{1/2} \varepsilon_t = h_t^{1/2} \varepsilon_t. \tag{3.1}$$

In eq.(3.1), each of the coefficients, including α_0, is allowed be any non-negative real number. Having assumed that, we now start our *tour de force*. Let us formally search for a proper solution of eq.(3.1):

$$
\begin{aligned}
h_t &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 h_{t-i} = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 \left[\alpha_0 + \sum_{j=1}^{p} \alpha_j h_{t-i-j} \varepsilon_{t-i-j}^2 \right] \\
&= \alpha_0 + \alpha_0 \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 \sum_{j=1}^{p} \alpha_j h_{t-i-j} \varepsilon_{t-i-j}^2 \\
&= \alpha_0 + \alpha_0 \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 \sum_{j=1}^{p} \alpha_j \varepsilon_{t-i-j}^2 \left[\alpha_0 + \sum_{k=1}^{p} \alpha_k \varepsilon_{t-i-j-k}^2 h_{t-i-j-k} \right] \\
&= \alpha_0 + \alpha_0 \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \alpha_0 \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 \sum_{j=1}^{p} \alpha_j \varepsilon_{t-i-j}^2 \\
&\quad + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 \sum_{j=1}^{p} \alpha_j \varepsilon_{t-i-j}^2 \sum_{k=1}^{p} \alpha_k \varepsilon_{t-i-j-k}^2 \left[\alpha_0 + \sum_{l=1}^{p} \alpha_l \varepsilon_{t-i-j-k-l}^2 h_{t-i-j-k-l} \right] \\
&= \alpha_0 + \alpha_0 \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \alpha_0 \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 \sum_{j=1}^{p} \alpha_j \varepsilon_{t-i-j}^2 \\
&\quad + \alpha_0 \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 \sum_{j=1}^{p} \alpha_j \varepsilon_{t-i-j}^2 \sum_{k=1}^{p} \alpha_k \varepsilon_{t-i-j-k}^2 \\
&\quad + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 \sum_{j=1}^{p} \alpha_j \varepsilon_{t-i-j}^2 \sum_{k=1}^{p} \alpha_k \varepsilon_{t-i-j-k}^2 \sum_{l=1}^{p} \alpha_l \varepsilon_{t-i-j-k-l}^2 h_{t-i-j-k-l} \\
&= \alpha_0 \left[1 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \sum_{j=1}^{p} \alpha_i \alpha_j \varepsilon_{t-i}^2 \varepsilon_{t-i-j}^2 \right] \\
&\quad + \alpha_0 \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} \alpha_i \alpha_j \alpha_k \varepsilon_{t-i}^2 \varepsilon_{t-i-j}^2 \varepsilon_{t-i-j-k}^2 \\
&\quad + \left(\sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} \sum_{l=1}^{p} \alpha_i \alpha_j \alpha_k \alpha_l \varepsilon_{t-i}^2 \varepsilon_{t-i-j}^2 \varepsilon_{t-i-j-k}^2 \varepsilon_{t-i-j-k-l}^2 h_{t-i-j-k-l} \right).
\end{aligned} \tag{3.2}
$$

Changing the indices in eqs.(3.2), we obtain, for each $n \geq 2$,

$$
\begin{aligned}
h_t = \alpha_0 &\left[1 + \sum_{i=1}^{n} \left(\sum_{k_1=1}^{p} \cdots \sum_{k_i=1}^{p} \alpha_{k_1} \ldots \alpha_{k_i} \varepsilon_{t-k_1}^2 \ldots \varepsilon_{t-k_1-\ldots-k_i}^2 \right) \right] \\
&+ \sum_{k_1=1}^{p} \cdots \sum_{k_{n+1}=1}^{p} \alpha_{k_1} \ldots \alpha_{k_{n+1}} \varepsilon_{t-k_1}^2 \ldots \varepsilon_{t-k_1-\ldots-k_{n+1}}^2 h_{t-k_1-\ldots-k_{n+1}}.
\end{aligned} \tag{3.3}
$$

Then, we conjecture that a possible candidate for solving eq.(3.1) might be

$$Y_t^* = \varepsilon_t \sqrt{\alpha_0 \left[1 + \sum_{i=1}^{\infty} \left(\sum_{k_1=1}^{p} \cdots \sum_{k_i=1}^{p} \alpha_{k_1} \ldots \alpha_{k_i} \varepsilon_{t-k_1}^2 \ldots \varepsilon_{t-k_1-\ldots-k_i}^2 \right) \right]}, \; t \in \mathbb{Z}. \tag{3.4}$$

As we see below, eq.(3.4) does quite fine, should one further assumption on the parameters be considered.

Theorem 1. *If $\sum_{j=1}^{p} \alpha_j < 1$, then eq.(3.4) exists as a stochastic process, it solves eq.(3.1), and each of its random variables is P-almost surely finite. Moreover, such process is both second-order and strictly stationary.*

Proof. From now on, assume this additional condition above on the α_i's. Also, fix $t \in \mathbb{Z}$. If we go back to eq.(3.3), it follows that the latter has as a solution

$$h_t^* = \alpha_0 \left[1 + \sum_{i=1}^{\infty} \left(\sum_{k_1=1}^{p} \cdots \sum_{k_i=1}^{p} \alpha_{k_1} \ldots \alpha_{k_i} \varepsilon_{t-k_1}^2 \ldots \varepsilon_{t-k_1-\ldots-k_i}^2 \right) \right], \tag{3.5}$$

and, consequently, $Y_t^* = (h_t^*)^{1/2} \varepsilon_t$ is a solution of eq.(3.1). We leave two detailed but not really difficult derivations to sections C and D of the Appendix.
The fact that $Y_t^* = (h_t^*)^{1/2} \varepsilon_t$ is indeed a random variable follows from noticing that $h_t^* = \lim_{n\to\infty} h_{t,n}^*$, where

$$h_{t,n}^* = \alpha_0 + \alpha_0 \sum_{i=1}^{n} \left(\sum_{k_1=1}^{p} \cdots \sum_{k_i=1}^{p} \alpha_{k_1} \ldots \alpha_{k_i} \varepsilon_{t-k_1}^2 \ldots \varepsilon_{t-k_1-\ldots-k_i}^2 \right), \quad n \in \mathbb{N},$$

is a non-decreasing sequence of non-negative (and certainly well-defined[5]) random variables. So, we just recall that, whenever existing, pointwise limits of sequences of random variables are themselves random variables[6]. This completes the proof of the first two claims of the theorem.
Moving on, we establish finiteness. From the Monotone Convergence Theorem (cf. Shiryaev [39], page 186; or Kubrusly [29], pages 37-38 [7]) and from the hypothesis

[5]Indeed, $h_{t,n}^* = \varphi(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-np})$ for each $n \in \mathbb{N}$, where $\varphi(.)$ is a *Borel measurable function* – for the a precise definition of the latter, see Kubrusly [29], chapter 1. Also, the stochastic process $(\varepsilon_t)_{t\in\mathbb{Z}}$ is always well-defined, according to Shiryaev [39], chapter II, section 9, Theorem 1; or Chung [14], Theorems 3.3.4 and 3.3.6.

[6]Possibly an extended real-valued one – see section B of the Appendix.

[7]It is worth to recall that, from the measure and integration theory standpoint, random variables remain *measurable functions* – and expectations are *integrals*.

that $\varepsilon_t \sim \mathrm{IID}(0,1)$,

$$
\begin{aligned}
0 \leq \mathrm{E}\left(h_t^*\right) &= \mathrm{E}\left(\lim_{n\to\infty} h_{t,n}^*\right) = \lim_{n\to\infty} \mathrm{E}\left(h_{t,n}^*\right) \\
&= \lim_{n\to\infty} \mathrm{E}\left[\alpha_0 + \alpha_0 \sum_{i=1}^{n}\left(\sum_{k_1=1}^{p}\cdots\sum_{k_i=1}^{p} \alpha_{k_1}\ldots\alpha_{k_i}\varepsilon_{t-k_1}^2\ldots\varepsilon_{t-k_1-\ldots-k_i}^2\right)\right] \\
&= \lim_{n\to\infty} \left[\alpha_0 + \alpha_0 \sum_{i=1}^{n}\left(\sum_{k_1=1}^{p}\cdots\sum_{k_i=1}^{p} \alpha_{k_1}\ldots\alpha_{k_i}\mathrm{E}\left(\varepsilon_{t-k_1}^2\right)\ldots\mathrm{E}\left(\varepsilon_{t-k_1-\ldots-k_i}^2\right)\right)\right] \\
&= \lim_{n\to\infty} \alpha_0 \left[1 + \sum_{i=1}^{n}\left(\sum_{k_1=1}^{p}\cdots\sum_{k_i=1}^{p} \alpha_{k_1}\ldots\alpha_{k_i}\right)\right] \\
&= \lim_{n\to\infty} \alpha_0 \left[1 + \sum_{i=1}^{n}\left(\sum_{k_1=1}^{p}\alpha_{k_1}\sum_{k_2=1}^{p}\alpha_{k_2}\cdots\sum_{k_i=1}^{p} \alpha_{k_i}\right)\right] \\
&= \lim_{n\to\infty} \alpha_0 \left[1 + \sum_{i=1}^{n}\underbrace{\left(\alpha_1+\ldots+\alpha_p\right)\left(\alpha_1+\ldots+\alpha_p\right)\ldots\left(\alpha_1+\ldots+\alpha_p\right)}_{i-\text{times}}\right] \\
&= \lim_{n\to\infty} \alpha_0 \left[1 + \sum_{i=1}^{n}\left(\alpha_1+\ldots+\alpha_p\right)^i\right] \\
&= \alpha_0 \left[1 + \sum_{i=1}^{\infty}\left(\alpha_1+\ldots+\alpha_p\right)^i\right] < \infty.
\end{aligned}
$$

The convergence in the last line follows from the hypothesis over the α_i's and, according to Shiryaev [39], page 185, Property J, or Kubrusly [29], page 46, Problem 3.9, implies that h_t^* is finite with probability 1. So, $Y_t^* = \left(h_t^*\right)^{1/2}\varepsilon_t$, besides existing, is finite with probability 1. Furthermore, it is also second-order, since

$$
\mathrm{E}\left[\left(Y_t^*\right)^2\right] = \mathrm{E}\left[h_t^*\varepsilon_t^2\right] = \mathrm{E}\left[h_t^*\right]\mathrm{E}\left[\varepsilon_t^2\right] = \mathrm{E}\left[h_t^*\right] < \infty\ , \tag{3.6}
$$

where the second equality follows from independence between the random variables ε_t and $h_t^* = g\left(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\right)$.
Finally, let us establish that $\left(Y_t^*\right)_{t\in\mathbb{Z}}$ is strictly stationary – which also leads to second-order stationarity, in view of (3.6). Define, for a fixed $n \in \mathbb{N}$, the auxiliary stochastic processes

$$
Y_t^{(n)} = \varepsilon_t \left\{\alpha_0\left[1 + \sum_{i=1}^{n}\left(\sum_{k_1=1}^{p}\cdots\sum_{k_i=1}^{p}\alpha_{k_1}\ldots\alpha_{k_i}\varepsilon_{t-k_1}^2\ldots\varepsilon_{t-k_1-\ldots-k_i}^2\right)\right]\right\}^{1/2}, \qquad t \in \mathbb{Z}\,.
$$

For each $t \in \mathbb{Z}$, $m \in \{0, 1, 2, 3, \ldots\}$ and $h \in \mathbb{Z}$, take $\underline{Y}_t^{(n)} \equiv \left(Y_t^{(n)}, Y_{t-1}^{(n)}, \ldots, Y_{t-m}^{(n)}\right) = g\left(\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-np-m}\right)$ and $\underline{Y}_{t+h}^{(n)} \equiv \left(Y_{t+h}^{(n)}, Y_{t+h-1}^{(n)}, \ldots, Y_{t+h-m}^{(n)}\right)$ $= g\left(\varepsilon_{t+h}, \varepsilon_{t+h-1}, \ldots, \varepsilon_{t+h-np-m}\right)$. Since $\varepsilon_t \sim$ IID, we have $\left(\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-np-m}\right) \stackrel{d}{=} \left(\varepsilon_{t+h}, \varepsilon_{t+h-1}, \ldots, \varepsilon_{t+h-np-m}\right)$, which leads to $\underline{Y}_t^{(n)} \stackrel{d}{=} \underline{Y}_{t+h}^{(n)}$ [8]. But, from what we have formerly seen with regard to proving that Y_t^* is a random variable, it follows that $\underline{Y}_t^{(n)} \stackrel{P-a.s.}{\longrightarrow} \underline{Y}_t^* = \left(Y_t^*, Y_{t-1}^*, \ldots, Y_{t-m}^*\right)$ for each t, which implies that $\underline{Y}_t^{(n)} \stackrel{d}{\longrightarrow} \underline{Y}_t^*$. For these same reasons, it also follows that $\underline{Y}_{t+h}^{(n)} \stackrel{d}{\longrightarrow} \underline{Y}_{t+h}^*$ Combining these facts, we finally obtain $\underline{Y}_t^* \stackrel{d}{=} \underline{Y}_{t+h}^*$. Since t, m and h are arbitrary, stationarity indeed holds under both strict and second-order senses. This concludes the proof. □

It is worth pointing an aspect of the proof above that refers to the final part of our discussion in section 1: generally treated in the GARCH literature under some conditional expectations and related results, eqs.(3.6) have ensued here under quite less technical probabilistic elements.

Is there any other solution for eq.(3.1)? Particularly regarding this question, we shall demonstrate two things. The first is that, under at least one of three relatively general conditions (one of these being a most natural one from a "real-world" standpoint), other existing second-order stationary solutions must all equal eq.(3.4) with probability 1. Secondly (and very interestingly): again with probability 1, and this time under no additional assumption, *any other* available strictly stationary solution and eq.(3.4) also coincide.

Our task needs two auxiliary results. We begin with an inequality on the real line.

Lemma 2. *Fix $j \in \mathbb{N}$. For any non-negative real numbers $x_1, x_2, \ldots, x_n$ ($n \in \mathbb{N}$), $\sqrt[j]{\sum_{i=1}^{n} x_i} \leq \sum_{i=1}^{n} \sqrt[j]{x_i}$.*

Proof. We use math induction in this one. For $n = 1$, we have for every $x_1 \geq 0$

$$\sqrt[j]{\sum_{i=1}^{1} x_i} = \sqrt[j]{x_1} \leq \sqrt[j]{x_1} = \sum_{i=1}^{1} \sqrt[j]{x_i}.$$

Now, take $x_1, x_2 \geq 0$. Notice that

[8] A word on notation: given two random vectors $\underline{X}_1$ and $\underline{X}_2$, we write $\underline{X}_1 \stackrel{d}{=} \underline{X}_2$ to indicate that $\underline{X}_1$ and $\underline{X}_2$ have the *same* probability distribution.

$$\begin{aligned}x_1+x_2 &= x_2+x_1 \le x_2+x_1+\sum_{k=1}^{j-1}\binom{j}{k}\left(x_1^{\frac{1}{j}}\right)^k\left(x_2^{\frac{1}{j}}\right)^{j-k}\\ &= \binom{j}{0}\left(x_1^{\frac{1}{j}}\right)^0\left(x_2^{\frac{1}{j}}\right)^{j-0}+\binom{j}{j}\left(x_1^{\frac{1}{j}}\right)^j\left(x_2^{\frac{1}{j}}\right)^{j-j}\\ &\quad+\sum_{k=1}^{j-1}\binom{j}{k}\left(x_1^{\frac{1}{j}}\right)^k\left(x_2^{\frac{1}{j}}\right)^{j-k}\\ &= \sum_{k=0}^{j}\binom{j}{k}\left(x_1^{\frac{1}{j}}\right)^k\left(x_2^{\frac{1}{j}}\right)^{j-k}=\left(x_1^{\frac{1}{j}}+x_2^{\frac{1}{j}}\right)^j.\end{aligned}$$

Hence, we obtain

$$\sqrt[j]{\sum_{i=1}^{2}x_i}=\sqrt[j]{x_1+x_2}\le\sqrt[j]{\left(x_1^{\frac{1}{j}}+x_2^{\frac{1}{j}}\right)^j}=x_1^{\frac{1}{j}}+x_2^{\frac{1}{j}}=\sum_{i=1}^{2}\sqrt[j]{x_i}. \tag{3.7}$$

Finally, let us assume the inequality in this lemma's statement is true for some specific $n\in\mathbb{N}$. Taking arbitrary $x_1,x_2,\ldots,x_n,x_{n+1}\ge 0$, it follows that

$$\begin{aligned}\sqrt[j]{\sum_{i=1}^{n+1}x_i} &= \sqrt[j]{\left(\sum_{i=1}^{n}x_i\right)+x_{n+1}}\le\left(\sqrt[j]{\sum_{i=1}^{n}x_i}\right)+\sqrt[j]{x_{n+1}}\\ &\le\left(\sum_{i=1}^{n}\sqrt[j]{x_i}\right)+\sqrt[j]{x_{n+1}}=\sum_{i=1}^{n+1}\sqrt[j]{x_i}\,;\end{aligned}$$

the first inequality ensues as an application of (3.7), whereas the second follows from the induction hypothesis. □

The second fact is one of the implications found in the equivalence claimed in Hamilton [24] regarding the inequality (21.1.7) on page 659 of that book. Here, we offer a proof.

Lemma 3. *Let $\alpha_i\ge 0$, $i=1,\ldots,p$, be the parameters of eq.(3.1). If $\sum_{i=1}^{p}\alpha_i<1$, then $\phi(z)=1-\alpha_1 z-\alpha_2 z^2-\ldots-\alpha_p z^p\neq 0$ for any given $z\in\mathbb{C}$ such that $|z|\le 1$ (in other very well-known words: all roots of polynomial $\phi(.)$ lie outside the unit circle).*

Proof. Bearing in mind that $\sum_{i=1}^{p}\alpha_i<1$ is obviously equivalent to $1-\alpha_1-\alpha_2-\ldots-\alpha_p>0$, take $z_0\in\mathbb{C}$ such that $|z_0|\le 1$. This, once again trivially, leads to $-\alpha_i|z_0|^i\ge-\alpha_i$ for $i=1,\ldots,p$. Hence,

$$\begin{aligned}|\phi(z_0)| &= |1-\alpha_1 z_0-\alpha_2 z_0^2-\ldots-\alpha_p z_0^p|\\ &\ge 1-\alpha_1|z_0|-\alpha_2|z_0^2|-\ldots-\alpha_p|z_0^p|\\ &= 1-\alpha_1|z_0|-\alpha_2|z_0|^2-\ldots-\alpha_p|z_0|^p\\ &\ge 1-\alpha_1-\alpha_2-\ldots-\alpha_p>0.\end{aligned}$$

So z_0 is not a root of $\phi(.)$. This completes the proof of the lemma. □

And now, the uniqueness. We begin with the second-order case.

Theorem 4. *The stochastic process $(Y_t^*)_{t\in\mathbb{Z}}$ in (3.4) is $P-$almost surely the unique second-order stationary solution for eq.(3.1), provided that at least one of these three assumptions holds:*

(i) $\sum_{j=1}^{p} \alpha_j < 1$ *and the considered solutions* $\left(\widehat{Y}_t\right)_{t\in\mathbb{Z}}$ *are all causal.*[9]

(ii) *There exists* $\delta \geq 2$ *such that* $\alpha_1^{\frac{1}{\delta}} + \alpha_2^{\frac{1}{\delta}} + \ldots + \alpha_p^{\frac{1}{\delta}} < 1.$[10]

(iii) *There exist* $n_0 \in \mathbb{N}$ *and a strictly positive real number* γ*, such that, for all* $t \in \mathbb{Z}$ *and* $n \geq n_0$*,* $\mathbb{E}\left[\mid Y_{t-j} \mid^4\right] \leq n^{2\gamma}$*,* $j = n+1, \ldots, (n+1)p$*. Also,* $\sqrt{\mathbb{E}\left[\varepsilon_1^4\right]}\,(\alpha_1 + \alpha_2 + \ldots + \alpha_p) < 1.$

Proof. We assume hereafter that at least one of the $\alpha_1, \ldots, \alpha_p$ is strictly positive; otherwise the theorem becomes trivial. Suppose that $\left(\widehat{Y}_t\right)_{t\in\mathbb{Z}}$ is another second-order stationary solution of eq.(3.1). Fixing t and using almost the same writings that precede Theorem 1, we obtain

$$\begin{aligned}\widehat{Y}_t &= \varepsilon_t\sqrt{\alpha_0 + \alpha_1 \widehat{Y}_{t-1}^2 + \ldots + \alpha_p \widehat{Y}_{t-p}^2} \\ &= \varepsilon_t \sqrt{\begin{array}{l}\alpha_0\left[1 + \sum_{i=1}^{n}\left(\sum_{k_1=1}^{p}\cdots\sum_{k_i=1}^{p} \alpha_{k_1}\ldots\alpha_{k_i}\varepsilon_{t-k_1}^2\ldots\varepsilon_{t-k_1\ldots-k_i}^2\right)\right] \\ + \sum_{k_1=1}^{p}\cdots\sum_{k_{n+1}=1}^{p} \alpha_{k_1}\ldots\alpha_{k_{n+1}}\varepsilon_{t-k_1}^2\ldots\varepsilon_{t-k_1-\ldots-k_n}^2 \widehat{Y}_{t-k_1-\ldots-k_{n+1}}^2\end{array}}.\end{aligned} \tag{3.8}$$

Let us examine with greater depth the very last line of eqs.(3.8) above, namely $W_n = \sum_{k_1=1}^{p}\cdots\sum_{k_{n+1}=1}^{p} \alpha_{k_1}\ldots\alpha_{k_{n+1}}\varepsilon_{t-k_1}^2\ldots\varepsilon_{t-k_1-\ldots-k_n}^2 \widehat{Y}_{t-k_1-\ldots-k_{n+1}}^2$. Condition (i), second-order stationarity of $\left(\widehat{Y}_t\right)_{t\in\mathbb{Z}}$ and the assumptions over $(\varepsilon_t)_{t\in\mathbb{Z}}$ imply that

$$\mathrm{E}\left(W_n\right) = (\alpha_1 + \alpha_2 + \ldots + \alpha_p)^{n+1}\,\mathrm{E}\left(\widehat{Y}_1^2\right) \in [0, \infty).$$

[9] This means that, for every $t \in \mathbb{Z}$, $\widehat{Y}_t$ depends only upon ε_j for $j \leq t$.

[10] It is worth-citing that such a restriction is surely stronger than the former given by $\sum_{j=1}^{p} \alpha_j < 1$. Indeed: once imposed, either one implies that $\alpha_i < 1$ for each i.

Therefore, for every $m \in \mathbb{N}$, and from the Markov's inequality (notice: W_n is non-negative),

$$\mathrm{P}\left(|W_n - 0| \geq \frac{1}{m}\right) = \mathrm{P}\left(W_n \geq \frac{1}{m}\right) \leq m\left(\alpha_1 + \ldots + \alpha_p\right)^{n+1} \mathrm{E}\left(\widehat{Y}_1^2\right). \tag{3.9}$$

From (3.9) and the hypothesis over the α_i's,

$$\sum_{n=1}^{\infty} \mathrm{P}\left(|W_n - 0| \geq \frac{1}{m}\right) \leq m\mathrm{E}\left(\widehat{Y}_1^2\right) \sum_{n=1}^{\infty} \left(\alpha_1 + \ldots + \alpha_p\right)^{n+1} < \infty. \tag{3.10}$$

Using (3.10) and the Borel-Cantelli Lemma (cf. Shiryaev [39], page 255; or Rohatgi & Saleh [38], page 281) yields

$$\mathrm{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{l=n}^{\infty} \left(|W_l - 0| \geq \frac{1}{m}\right)\right) = 0,$$

which, in view of De Morgan's law, implies that

$$\begin{aligned}
\mathrm{P}\left(\bigcup_{n=1}^{\infty} \bigcap_{l=n}^{\infty} \left(|W_l - 0| < \frac{1}{m}\right)\right) &= 1 - \mathrm{P}\left(\left(\bigcup_{n=1}^{\infty} \bigcap_{l=n}^{\infty} \left(|W_l - 0| < \frac{1}{m}\right)\right)^{c}\right) \\
&= 1 - \mathrm{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{l=n}^{\infty} \left(|W_l - 0| \geq \frac{1}{m}\right)\right) \\
&= 1 - 0 = 1.
\end{aligned} \tag{3.11}$$

From eqs.(3.11) and from the well-known fact that every countable intersection of events with probability 1 also has probability 1, we finally reach

$$1 = \mathrm{P}\left(\bigcap_{m=1}^{\infty} \left[\bigcup_{n=1}^{\infty} \bigcap_{l=n}^{\infty} |W_l - 0| < \frac{1}{m}\right]\right) = \mathrm{P}\left(\lim_{n\to\infty} W_n = 0\right),$$

that is, $W_n \overset{P-a.s.}{\longrightarrow} 0$ as $n \to \infty$. On the other hand, making $n \to \infty$, the first term at the right-hand side of eq.(3.8), still inside the second squared root, converges pointwise (hence, P-almost surely) to

$$\alpha_0 \left[1 + \sum_{i=1}^{\infty} \left(\sum_{k_1=1}^{p} \cdots \sum_{k_i=1}^{p} \alpha_{k_1} \ldots \alpha_{k_i} \varepsilon_{t-k_1}^2 \ldots \varepsilon_{t-k_1-\ldots-k_i}^2\right)\right].$$

Now, bearing in mind that squared roots preserve P-almost surely convergence, we may conclude that, with probability 1,

$$\begin{aligned} \widehat{Y}_t &= \lim_{n\to\infty} \widehat{Y}_t \\ &= \varepsilon_t \sqrt{\alpha_0 \left[1 + \sum_{i=1}^{\infty} \left(\sum_{k_1=1}^{p} \cdots \sum_{k_i=1}^{p} \alpha_{k_1} \ldots \alpha_{k_i} \varepsilon_{t-k_1}^2 \ldots \varepsilon_{t-k_1-\ldots-k_i}^2\right)\right]}. \end{aligned} \tag{3.12}$$

Combining eqs.(3.12) and eq.(3.4) implies

$$\mathrm{P}\left(\widehat{Y}_t = Y_t^*\right) = 1.$$

Since t is arbitrary, it follows that

$$\mathrm{P}\left((\widehat{Y}_t)_{t\in\mathbb{Z}} = (Y_t^*)_{t\in\mathbb{Z}}\right) = \mathrm{P}\left(\bigcap_{t\in\mathbb{Z}} \left(\widehat{Y}_t = Y_t^*\right)\right) = 1.$$

Now, assume condition (ii) holds and, again, let us consider another second-order stationary solution $\left(\widehat{Y}_t\right)_{t\in\mathbb{Z}}$ (not necessarily causal) for eq.(3.1), as well as eqs.(3.8) and the same definition for W_n. We obtain[11]

$$\begin{aligned} 0 &\le W_n^{\frac{1}{[\delta]}} \\ &= \left(\sum_{k_1=1}^{p} \cdots \sum_{k_{n+1}=1}^{p} \alpha_{k_1} \ldots \alpha_{k_{n+1}} \varepsilon_{t-k_1}^2 \ldots \varepsilon_{t-k_1-\ldots-k_n}^2 \widehat{Y}_{t-k_1-\ldots-k_{n+1}}^2\right)^{\frac{1}{[\delta]}} \\ &\le \sum_{k_1=1}^{p} \cdots \sum_{k_{n+1}=1}^{p} \alpha_{k_1}^{\frac{1}{[\delta]}} \ldots \alpha_{k_{n+1}}^{\frac{1}{[\delta]}} \left|\varepsilon_{t-k_1} \ldots \varepsilon_{t-k_1-\ldots-k_n} \widehat{Y}_{t-k_1-\ldots-k_{n+1}}\right|^{\frac{2}{[\delta]}}, \end{aligned} \tag{3.13}$$

where the second inequality follows from Lemma 2. From (3.13), the hypotheses over $(\varepsilon_t)_{t\in\mathbb{Z}}$ and Jensen's inequality, it follows that

$$\begin{aligned} 0 &\le \mathrm{E}\left(W_n^{\frac{1}{[\delta]}}\right) \\ &\le \sum_{k_1=1}^{p} \cdots \sum_{k_{n+1}=1}^{p} \alpha_{k_1}^{\frac{1}{[\delta]}} \ldots \alpha_{k_{n+1}}^{\frac{1}{[\delta]}} \mathrm{E}\left[\left|\varepsilon_{t-k_1} \ldots \varepsilon_{t-k_1-\ldots-k_n} \widehat{Y}_{t-k_1-\ldots-k_{n+1}}\right|^{\frac{2}{[\delta]}}\right] \\ &\le \sum_{k_1=1}^{p} \cdots \sum_{k_{n+1}=1}^{p} \alpha_{k_1}^{\frac{1}{[\delta]}} \ldots \alpha_{k_{n+1}}^{\frac{1}{[\delta]}} \left[\mathrm{E}\left[\left|\varepsilon_{t-k_1} \ldots \varepsilon_{t-k_1-\ldots-k_n} \widehat{Y}_{t-k_1-\ldots-k_{n+1}}\right|\right]\right]^{\frac{2}{[\delta]}}. \end{aligned} \tag{3.14}$$

[11] Recall that, for any $x \in \mathbb{R}$, $[x] = \max\{z \in \mathbb{Z} : z \le x\}$.

From (3.14) and the Cauchy-Scharwz inequality (and, once again, using the assumptions on $(\varepsilon_t)_{t\in\mathbb{Z}}$),

$$
\begin{aligned}
0 &\leq \mathrm{E}\left(W_n^{\frac{1}{[\delta]}}\right) \\
&\leq \sum_{k_1=1}^{p}\cdots\sum_{k_{n+1}=1}^{p}\alpha_{k_1}^{\frac{1}{[\delta]}}\cdots\alpha_{k_{n+1}}^{\frac{1}{[\delta]}}\left(\sqrt{\mathrm{E}\left(\varepsilon_{t-1}^2\right)\cdots\mathrm{E}\left(\varepsilon_{t-k_1-\ldots-k_n}^2\right)\mathrm{E}\left(\widehat{Y}_{t-k_1-\ldots-k_{n+1}}^2\right)}\right)^{\frac{2}{[\delta]}} \\
&= \sum_{k_1=1}^{p}\cdots\sum_{k_{n+1}=1}^{p}\alpha_{k_1}^{\frac{1}{[\delta]}}\cdots\alpha_{k_{n+1}}^{\frac{1}{[\delta]}}\left(\mathrm{E}\left(\widehat{Y}_{t-k_1-\ldots-k_{n+1}}^2\right)\right)^{\frac{1}{[\delta]}} \\
&= \left(\mathrm{E}\left(\widehat{Y}_1^2\right)\right)^{\frac{1}{[\delta]}}\sum_{k_1=1}^{p}\cdots\sum_{k_{n+1}=1}^{p}\alpha_{k_1}^{\frac{1}{[\delta]}}\cdots\alpha_{k_{n+1}}^{\frac{1}{[\delta]}} \\
&= \left(\mathrm{E}\left(\widehat{Y}_1^2\right)\right)^{\frac{1}{[\delta]}}\left(\alpha_1^{\frac{1}{[\delta]}}+\ldots+\alpha_p^{\frac{1}{[\delta]}}\right)^{n+1} \leq \left(\mathrm{E}\left(\widehat{Y}_1^2\right)\right)^{\frac{1}{[\delta]}}\left(\alpha_1^{\frac{1}{\delta}}+\ldots+\alpha_p^{\frac{1}{\delta}}\right)^{n+1}.
\end{aligned}
$$

The second equality ensues from second-order stationarity, and the last inequality is true in view that $\alpha_i < 1$ for every $i = 1, \ldots, p$ and $[\delta] \leq \delta$, that is, $\frac{1}{\delta} \leq \frac{1}{[\delta]}$, which implies $\alpha_i^{\frac{1}{[\delta]}} \leq \alpha_i^{\frac{1}{\delta}}$ (recall: for every fixed $c \in (0,1)$, $g(x) = c^x$, $x \in \mathbb{R}$, is strictly decreasing). The rest of the proof goes pretty much as we have proceeded with item (i): just replace W_n by $W_n^{\frac{1}{[\delta]}}$ and $(\alpha_1 + \ldots + \alpha_p)^{n+1}$ by $\left(\alpha_1^{\frac{1}{\delta}} + \ldots + \alpha_p^{\frac{1}{\delta}}\right)^{n+1}$, and recall that $W_n^{\frac{1}{[\delta]}} \overset{P-a.s.}{\longrightarrow} 0$ implies $W_n \overset{P-a.s.}{\longrightarrow} 0$.
Finally, assume condition (iii). Using current notation and the same justifications, we obtain, for each $n \geq n_0$,

$$
\begin{aligned}
0 &\leq \mathrm{E}\left(W_n\right) \\
&= \sum_{k_1=1}^{p}\cdots\sum_{k_{n+1}=1}^{p}\alpha_{k_1}\cdots\alpha_{k_{n+1}}\mathrm{E}\left(\left|\varepsilon_{t-k_1}^2\cdots\varepsilon_{t-k_1-\ldots-k_n}^2\right|\left|Y_{t-k_1-\ldots-k_{n+1}}^2\right|\right) \\
&\leq \sum_{k_1=1}^{p}\cdots\sum_{k_{n+1}=1}^{p}\alpha_{k_1}\cdots\alpha_{k_{n+1}}\sqrt{\mathrm{E}\left(\varepsilon_{t-k_1}^4\cdots\varepsilon_{t-k_1-\ldots-k_n}^4\right)\mathrm{E}\left(Y_{t-k_1-\ldots-k_{n+1}}^4\right)} \\
&= \sum_{k_1=1}^{p}\cdots\sum_{k_{n+1}=1}^{p}\alpha_{k_1}\cdots\alpha_{k_{n+1}}\sqrt{\mathrm{E}\left(\varepsilon_{t-k_1}^4\right)\cdots\mathrm{E}\left(\varepsilon_{t-k_1-\ldots-k_n}^4\right)\mathrm{E}\left(Y_{t-k_1-\ldots-k_{n+1}}^4\right)}.
\end{aligned}
\tag{3.15}
$$

Since $\varepsilon_t \sim$ IID, it follow from (3.15) and the first inequality in (iii) that

$$\begin{aligned}
0 \leq \mathrm{E}\left(W_n\right) &\leq \sum_{k_1=1}^{p} \cdots \sum_{k_{n+1}=1}^{p} \alpha_{k_1} \ldots \alpha_{k_{n+1}} \sqrt{\left[\mathrm{E}\left(\varepsilon_1^4\right)\right]^n \mathrm{E}\left(Y_{t-k_1-\ldots-k_{n+1}}^4\right)} \\
&= \left(\sqrt{\mathrm{E}\left(\varepsilon_1^4\right)}\right)^n \sum_{k_1=1}^{p} \cdots \sum_{k_{n+1}=1}^{p} \alpha_{k_1} \ldots \alpha_{k_{n+1}} \sqrt{\mathrm{E}\left(Y_{t-k_1-\ldots-k_{n+1}}^4\right)} \\
&\leq \left(\sqrt{\mathrm{E}\left(\varepsilon_1^4\right)}\right)^n \sum_{k_1=1}^{p} \cdots \sum_{k_{n+1}=1}^{p} \alpha_{k_1} \ldots \alpha_{k_{n+1}} n^\gamma \\
&= n^\gamma \left(\sqrt{\mathrm{E}\left(\varepsilon_1^4\right)}\right)^n \sum_{k_1=1}^{p} \cdots \sum_{k_{n+1}=1}^{p} \alpha_{k_1} \ldots \alpha_{k_{n+1}} \\
&= n^\gamma \left(\sqrt{\mathrm{E}\left(\varepsilon_1^4\right)}\right)^n \left(\alpha_1 + \ldots + \alpha_p\right)^{n+1} \\
&= \frac{n^\gamma}{\sqrt{\mathrm{E}\left(\varepsilon_1^4\right)}} \left(\sqrt{\mathrm{E}\left(\varepsilon_1^4\right)} \left(\alpha_1 + \ldots + \alpha_p\right)\right)^{n+1}.
\end{aligned} \tag{3.16}$$

This last division is certainly a well-defined one, since $\sqrt{\mathrm{E}\left(\varepsilon_1^4\right)}$ is strictly positive (indeed: ε_t is a non-degenerated random variable) and finite (cf. the second inequality in (iii)).
From (3.16) and Markov's inequality, we have, for every $m \in \mathbb{N}$,

$$\mathrm{P}\left(\left|W_n - 0\right| \geq \frac{1}{m}\right) \leq m \mathrm{E}\left(W_n\right) \leq m \frac{n^\gamma}{\sqrt{\mathrm{E}\left(\varepsilon_1^4\right)}} \left(\sqrt{\mathrm{E}\left(\varepsilon_1^4\right)} \left(\alpha_1 + \ldots + \alpha_p\right)\right)^{n+1}.$$

The second inequality in (iii) and the Ratio Test for absolute convergence imply

$$\sum_{n=n_0}^{\infty} n^\gamma \left(\sqrt{\mathrm{E}\left(\varepsilon_1^4\right)} \left(\alpha_1 + \ldots + \alpha_p\right)\right)^{n+1} < \infty.$$

The rest of the proof derives exactly as before. □

To prove uniqueness of eq.(3.4) as a strictly stationary solution of the ARCH equation requires no further assumptions concerning the parameters – even causality is immaterial. Let us now dwell on this. Notice that, if on one hand Lemma 2 has already played its role in Theorem 4, Lemma 3 is about to step in.

Theorem 5. *Assume (only) the restriction $\sum_{j=1}^{p} \alpha_j < 1$. The stochastic process $(Y_t^*)_{t \in \mathbb{Z}}$ in eq.(3.4) is $P-$almost surely the unique strictly stationary solution, either causal or not, for eq.(3.1).*

Proof. The strategy consists on rephrasing eq.(3.1) under the following matrix notation:

$$\begin{bmatrix} \widetilde{h}_t \\ \widetilde{Y}^2_{t-1} \\ \widetilde{Y}^2_{t-2} \\ \vdots \\ \widetilde{Y}^2_{t-(p-1)} \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \widetilde{Y}^2_{t-1} \\ \widetilde{Y}^2_{t-2} \\ \widetilde{Y}^2_{t-3} \\ \vdots \\ \widetilde{Y}^2_{t-(p-1)} \\ \widetilde{Y}^2_{t-p} \end{bmatrix}, \tag{3.17}$$

$$\begin{bmatrix} \widetilde{Y}^2_{t} \\ \widetilde{Y}^2_{t-1} \\ \widetilde{Y}^2_{t-2} \\ \vdots \\ \widetilde{Y}^2_{t-(p-1)} \end{bmatrix} = \begin{bmatrix} \alpha_0\varepsilon_t^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} \alpha_1\varepsilon_t^2 & \alpha_2\varepsilon_t^2 & \cdots & \alpha_{p-1}\varepsilon_t^2 & \alpha_p\varepsilon_t^2 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \widetilde{Y}^2_{t-1} \\ \widetilde{Y}^2_{t-2} \\ \widetilde{Y}^2_{t-3} \\ \vdots \\ \widetilde{Y}^2_{t-(p-1)} \\ \widetilde{Y}^2_{t-p} \end{bmatrix}, \tag{3.18}$$

where $\left(\widetilde{Y}_t\right)_{t\in\mathbb{Z}}$ and $\left(\widetilde{h}_t\right)_{t\in\mathbb{Z}}$ refer to another (possibly existing) strictly stationary solution. Defining

$$\underline{\widetilde{h}}_t = \begin{bmatrix} \widetilde{h}_t \\ \widetilde{Y}^2_{t-1} \\ \widetilde{Y}^2_{t-2} \\ \vdots \\ \widetilde{Y}^2_{t-(p-1)} \end{bmatrix}, \ \underline{\alpha}_0 = \begin{bmatrix} \alpha_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ A = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{p-1} & \alpha_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix},$$

and also[12]

$$\underline{\widetilde{Y}}^2_t = \begin{bmatrix} \widetilde{Y}^2_{t} \\ \widetilde{Y}^2_{t-1} \\ \widetilde{Y}^2_{t-2} \\ \vdots \\ \widetilde{Y}^2_{t-(p-1)} \end{bmatrix}, \ \underline{\alpha}_t = \begin{bmatrix} \alpha_0\varepsilon_t^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ A_t = \begin{bmatrix} \alpha_1\varepsilon_t^2 & \alpha_2\varepsilon_t^2 & \cdots & \alpha_{p-1}\varepsilon_t^2 & \alpha_p\varepsilon_t^2 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix},$$

we can write eq.(3.17) and eq.(3.18) respectively as

$$\underline{\widetilde{h}}_t = \underline{\alpha}_0 + A\underline{\widetilde{Y}}^2_{t-1}, \tag{3.19}$$

[12] Of course, $\underline{Y}^2_t$ is not a "squared random vector". Such a thing would not be even well-defined. Instead, $\underline{Y}^2_t$ is just a notation for a vector that has *squared* random variables along its entries.

$$\underline{\widetilde{Y}}_t^2 = \underline{\alpha}_t + A_t \underline{\widetilde{Y}}_{t-1}^2. \tag{3.20}$$

Using eq.(3.19) and eq.(3.20), we obtain

$$\begin{aligned}
\underline{\widetilde{h}}_t &= \underline{\alpha}_0 + A\underline{\widetilde{Y}}_{t-1}^2 \\
&= \underline{\alpha}_0 + A\left(\underline{\alpha}_{t-1} + A_{t-1}\underline{\widetilde{Y}}_{t-2}^2\right) \\
&= \underline{\alpha}_0 + A\underline{\alpha}_{t-1} + AA_{t-1}\underline{\widetilde{Y}}_{t-2}^2 \\
&= \underline{\alpha}_0 + A\underline{\alpha}_{t-1} + AA_{t-1}\left(\underline{\alpha}_{t-2} + A_{t-2}\underline{\widetilde{Y}}_{t-3}^2\right) \\
&= \underline{\alpha}_0 + A\underline{\alpha}_{t-1} + AA_{t-1}\underline{\alpha}_{t-2} + AA_{t-1}A_{t-2}\underline{\widetilde{Y}}_{t-3}^2 \\
&\vdots \\
&= \underline{\alpha}_0 + A\underline{\alpha}_{t-1} + A\sum_{j=1}^{n-1} A_{t-1}A_{t-2}\dots A_{t-j}\underline{\alpha}_{t-(j+1)} \\
&\quad + AA_{t-1}\dots A_{t-n}\underline{\widetilde{Y}}_{t-(n+1)}^2.
\end{aligned} \tag{3.21}$$

Consider the last parcel of the last right-hand size of (3.21), namely

$$\underline{W}_n = \begin{bmatrix} W_{n,1} \\ W_{n,2} \\ \vdots \\ W_{n,p} \end{bmatrix} = A_{t-1}\dots A_{t-n}\underline{\widetilde{Y}}_{t-(n+1)}^2. \tag{3.22}$$

Since the random matrices $A_{t-1}, A_{t-2}, \dots, A_{t-n}$ are independent and integrable[13] (given that $\varepsilon_t \sim \mathrm{IID}(0,1)$ and the very definition of such matrices), it follows that

$$\mathrm{E}\left(A_{t-1}A_{t-2}\dots A_{t-n}\right) = \mathrm{E}\left(A_{t-1}\right)\mathrm{E}\left(A_{t-2}\right)\dots\mathrm{E}\left(A_{t-n}\right) = AA\dots A = A^n. \tag{3.23}$$

From (3.23) and the hypothesis over $\alpha_1, \alpha_2, \dots, \alpha_p$, it follows from Lemma 3 and the results of Hamilton [24], section 1.2[14], that

$$\mathrm{E}\left(A_{t-1}A_{t-2}\dots A_{t-n}\right) = \left(T\Lambda T^{-1}\right)^n = T\Lambda^n T^{-1}, \tag{3.24}$$

where Λ is a diagonal matrix with the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ of A in the main diagonal (each of them has norm strictly less than 1), and T has the corresponding eigenvectors. We notice that all entries $a_n^{l,k}$ of $A_{t-1}A_{t-2}\dots A_{t-n}$ are non-negative

[13] Should read: with integrable random variables in its entries.

[14] In what follows, we assume the eigenvalues of A are all distinct. The reader possibly interested in the more general case is referred to section E of the Appendix.

random variables and denote by $\tau_{l,k}$ and $\tau^{l,k}$ the entries of T and T^{-1} respectively. Then, for every $m \in \mathbb{N}$ and $l, k = 1, \ldots, p$, Markov's inequality and (3.24) imply that

$$P\left(|a_n^{l,k} - 0| \geq \frac{1}{m}\right) = P\left(a_n^{l,k} \geq \frac{1}{m}\right) \leq m \sum_{i=1}^{p} \lambda_i^n \tau_{li} \tau^{ik}. \tag{3.25}$$

Let us dwell a bit more on the right-hand size of the inequality in (3.25). Using the well-known notations $|\ |$ and $||\ ||$ for the absolute value of real numbers and the norm of complex numbers respectively, it must follows from the triangular inequality that

$$\begin{aligned}\sum_{n=1}^{\infty}\left|\sum_{i=1}^{p}\lambda_i^n\tau_{li}\tau^{ik}\right| &= \sum_{n=1}^{\infty}\left|\left|\sum_{i=1}^{p}\lambda_i^n\tau_{li}\tau^{ik}\right|\right| \ \leq\ \sum_{n=1}^{\infty}\sum_{i=1}^{p} ||\ \lambda_i^n\tau_{li}\tau^{ik}\ || \\ &= \sum_{n=1}^{\infty}\sum_{i=1}^{p} ||\ \lambda_i\ ||^n||\ \tau_{li}\ ||||\ \tau^{ik}\ || \\ &= \sum_{n=1}^{\infty}\left(||\ \lambda_1\ ||^n||\ \tau_{l1}\ ||||\ \tau^{1k}\ || + \ldots + ||\ \lambda_p\ ||^n||\ \tau_{lp}\ ||||\ \tau^{pk}\ ||\right),\end{aligned} \tag{3.26}$$

where the first equality is true since the $\lambda_i^{n+1}\tau_{li}\tau^{ik}$ are real numbers – see (3.23) or (3.24).
Also, for each $i = 1, \ldots, p$,

$$\sum_{n=1}^{\infty} ||\ \lambda_i\ ||^n||\ \tau_{li}\ ||||\ \tau_{ik}\ || \quad = \quad ||\ \tau_{li}\ ||||\ \tau^{ik}\ ||\sum_{n=1}^{\infty}||\ \lambda_i\ ||^n \quad < \quad \infty. \tag{3.27}$$

Combining (3.26) and (3.27):

$$\sum_{n=1}^{\infty}\left|\sum_{i=1}^{p}\lambda_i\tau_{li}\tau^{ik}\right| \quad = \quad \sum_{n=1}^{\infty}\left|\left|\sum_{i=1}^{p}\lambda_i\tau_{li}\tau^{ik}\right|\right| \quad < \quad \infty. \tag{3.28}$$

The same Borel-Cantelli argument used in Theorem 4 also applies here, with (3.25) and (3.28), to yield $a_n^{l,k} \stackrel{P-a.s.}{\longrightarrow} 0$ as $n \to \infty$. And this implies that

$$a_n^{l,k} \stackrel{p}{\longrightarrow} 0. \tag{3.29}$$

On the other hand, since $(\widetilde{Y}_t)_{t\in\mathbb{Z}}$ is strictly stationary, we directly have, for each $t \in \mathbb{Z}$ and $k = 1, \ldots, p$,

$$\widetilde{Y}^2_{t-(n+1)-(k-1)} \stackrel{d}{\longrightarrow} \widetilde{Y}^2_1, \tag{3.30}$$

once again as $n \to \infty$. Slutsky's theorem (cf. Rohatgi & Saleh [38], page 270), used with (3.29) and (3.30), yields $a_n^{l,k}\widetilde{Y}^2_{t-(n+1)-(k-1)} \xrightarrow{d} 0.Y_1^2 = 0$. Therefore, we must necessarily have $a_n^{l,k}\widetilde{Y}^2_{t-(n+1)-(k-1)} \xrightarrow{p} 0$. This, in view of (3.22), implies

$$W_{n,l} = a_n^{l,1}\widetilde{Y}^2_{t-(n+1)} + a_n^{l,2}\widetilde{Y}^2_{t-(n+1)-1} + \ldots + a_n^{l,p}\widetilde{Y}^2_{t-(n+1)-(p-1)} \xrightarrow{p} 0 + 0 + \cdots + 0 = 0. \tag{3.31}$$

Using (3.21) and (3.31), we obtain[15]

$$\underline{\widetilde{h}}_t = \operatorname*{plim}_{n\to\infty} \underline{\widetilde{h}}_t = \underline{\alpha}_0 + A\underline{\alpha}_{t-1} + A\sum_{j=1}^{\infty} A_{t-1}A_{t-2}\ldots A_{t-j}\underline{\alpha}_{t-(j+1)}. \tag{3.32}$$

The fact that $\widetilde{h}_t = \alpha_0\left(1 + \sum_{i=1}^{\infty}\left(\sum_{k_1=1}^{p}\cdots\sum_{k_i=1}^{p}\alpha_{k_1}\ldots\alpha_{k_i}\varepsilon^2_{t-k_1}\ldots\varepsilon^2_{t-k_1-\ldots-k_i}\right)\right)$, which is the first entry of $\underline{\widetilde{h}}_t$ in (3.32), implies

$$\widetilde{Y}_t = \operatorname*{plim}_{n\to\infty} \widetilde{Y}_t = \operatorname*{plim}_{n\to\infty} \varepsilon_t\widetilde{h}_t = \varepsilon_t\sqrt{\alpha_0\left(1 + \sum_{i=1}^{\infty}\left(\sum_{k_1=1}^{p}\cdots\sum_{k_i=1}^{p}\alpha_{k_1}\ldots\alpha_{k_i}\varepsilon^2_{t-k_1}\ldots\varepsilon^2_{t-k_1-\ldots-k_i}\right)\right)}. \tag{3.33}$$

Comparing (3.4) and (3.33), we once again arrive at $\mathrm{P}\left(\widetilde{Y}_t = Y_t^*\right) = 1$ for every $t \in \mathbb{Z}$. Therefore and finally, $\mathrm{P}\left((\widehat{Y}_t)_{t\in\mathbb{Z}} = (Y_t^*)_{t\in\mathbb{Z}}\right) = \mathrm{P}\left(\bigcap_{t\in\mathbb{Z}}\left(\widehat{Y}_t = Y_t^*\right)\right) = 1$. □

What about Theorem 1's reciprocal? Yes, that also holds true. Actually, the textbooks reviewed in section 2 consider the question, and always with enough theoretical care. Thus, the reader might ask: why address, here, this part of the theory? Even though making this survey self-contained is a reason, there are also two more. One is the opportunity to offer another example of a proof, for a ARCH-related result, that is conditional expectation-free. The other is that we shall treat separately two specific cases – the first, which is way more frequent, is addressed right below.

Theorem 6. *Assume eq.(3.1) admits a solution, which is second-order causal stationary. Also, assume that* $\alpha_0 > 0$. *Then,* $\sum_{j=1}^{p}\alpha_j < 1$.

[15] The notation $\operatorname*{plim}_{n\to\infty} \underline{X}_n = \underline{X}$, where the random vectors $\underline{X}_n$ and $\underline{X}$ have of course the same dimension (let us say k), means that $X_{n,i} \xrightarrow{p} X_i$ for each $i = 1, \ldots, k$.

Proof. To ease notation, we simply write $(Y_t)_{t\in\mathbb{Z}}$ for the alleged existing solution of eq.(3.1). Fix $t \in \mathbb{Z}$. Notice that

$$0 \le \mathrm{E}(h_t) = \mathrm{E}(h_t) \times 1 = \mathrm{E}(h_t)\,\mathrm{E}\left(\varepsilon_t^2\right) = \mathrm{E}\left(h_t\varepsilon_t^2\right) = \mathrm{E}\left(Y_t^2\right) < \infty \ ,$$

where the second and third equalities follow from $\varepsilon_t \sim \mathrm{IID}(0,1)$ and Y_t being causal. Hence, Jensen's inequality implies that $0 \le \mathrm{E}\left(h_t^{1/2}\right) \le \sqrt{\mathrm{E}(h_t)} < \infty$.
So, continuing to use both causality and the assumptions on $(\varepsilon_t)_{t\in\mathbb{Z}}$, we obtain

$$\mathrm{E}(Y_t) = \mathrm{E}\left(h_t^{1/2}\varepsilon_t\right) = \mathrm{E}\left(h_t^{1/2}\right)\mathrm{E}(\varepsilon_t) = \mathrm{E}\left(h_t^{1/2}\right) \times 0 = 0,$$

and also

$$\begin{aligned}\mathrm{Var}(Y_1) = \mathrm{Var}(Y_t) = \mathrm{E}\left(Y_t^2\right) &= \mathrm{E}\left(h_t\varepsilon_t^2\right) = \mathrm{E}(h_t)\,\mathrm{E}\left(\varepsilon_t^2\right) = \mathrm{E}(h_t) \\ &= \mathrm{E}\left[\alpha_0 + \sum_{i=1}^{p}\alpha_i Y_{t-i}^2\right] = \alpha_0 + \sum_{i=1}^{p}\alpha_i \mathrm{E}\left(Y_{t-i}^2\right) \\ &= \alpha_0 + \sum_{i=1}^{p}\alpha_i \mathrm{Var}(Y_{t-i}) = \alpha_0 + \sum_{i=1}^{p}\alpha_i \mathrm{Var}(Y_1)\,,\end{aligned}$$

where the first and last equalities follow from second-order stationarity. Therefore,

$$\left(1 - \sum_{i=1}^{p}\alpha_i\right)\mathrm{Var}(Y_1) = \alpha_0\ . \tag{3.34}$$

In view of eq.(3.34), if $\sum_{i=1}^{p}\alpha_i = 1$, then $\alpha_0 = 0$, which contradicts the hypothesis that $\alpha_0 > 0$. So, $\sum_{i=1}^{p}\alpha_i < 1$ or $\sum_{i=1}^{p}\alpha_i > 1$. And we must take the former, since the latter leads to another contradiction, namely $\mathrm{Var}(Y_1) < 0$. □

The following result – which strictly speaking is not exactly a reciprocal of Theorem 1 – regards the second case, namely $\alpha_0 = 0$. Thus, we finally reach an answer to the very first question of this paper, which had to do with this parameter. What is to be learned below is that, before to even consider the aspects regarding asymptotic statistical inference, such restriction *must never be imposed* anytime one is handling the usual: time series data with (at least some) sample variance.

Theorem 7. *If $\alpha_0 = 0$, then the P-almost surely unique second-order stationary causal solution for eq.(3.1) is $Y_t = 0$, $t \in \mathbb{Z}$.*

Proof. First, we notice that $Y_t = 0$, $t \in \mathbb{Z}$, is indeed and trivially a solution for eq.(3.1) in case of $\alpha_0 = 0$. Actually, such solution and $(Y_t^*)_{t\in\mathbb{Z}}$ in eq.(3.4) coincide. Assume there exists another second-order stationary causal solution $\left(Y_t^\dagger\right)_{t\in\mathbb{Z}}$. If $\sum_{i=1}^{p}\alpha_i < 1$, then Theorem 4, item (i), guarantees the uniqueness claimed. Now, assume $\sum_{i=1}^{p} \alpha_i > 1$. Then, eq.(3.8) of Theorem 2's proof, used with $\left(Y_t^{\dagger}\right)_{t\in\mathbb{Z}}$ and $\alpha_0 = 0$, yield

$$\left(Y_t^{\dagger}\right)^2 = \varepsilon_t^2 \left(\sum_{k_1=1}^{p} \dots \sum_{k_n=1}^{p} \alpha_{k_1} \dots \alpha_{k_n} \varepsilon_{t-k_1}^2 \dots \varepsilon_{t-k_1-\dots-k_n}^2 \left(Y_{t-k_1-\dots-k_{n+1}}^{\dagger}\right)^2\right). \tag{3.35}$$

Suppose that, for at least one $t_0 \in \mathbb{Z}$, $\mathrm{P}\left(Y_{t_0}^{\dagger} \neq 0\right) > 0$, or equivalently $\mathrm{P}\left(\left(Y_{t_0}^{\dagger}\right)^2 > 0\right) > 0$. Thus, we have from eq.(3.35):

$$\begin{aligned}
0 &< \mathrm{E}\left[\left(Y_{t_0}^{\dagger}\right)^2\right] \\
&= \mathrm{E}\left[\varepsilon_{t_0}^2 \left(\sum_{k_1=1}^{p} \dots \sum_{k_n=1}^{p} \alpha_{k_1} \dots \alpha_{k_n} \varepsilon_{t_0-k_1}^2 \dots \varepsilon_{t_0-k_1-\dots-k_n}^2 \left(Y_{t_0-k_1-\dots-k_{n+1}}^{\dagger}\right)^2\right)\right] \\
&= \sum_{k_1=1}^{p} \dots \sum_{k_n=1}^{p} \alpha_{k_1} \dots \alpha_{k_n} \mathrm{E}\left[\left(Y_{t_0-k_1-\dots-k_{n+1}}^{\dagger}\right)^2\right] \\
&= \sum_{k_1=1}^{p} \dots \sum_{k_n=1}^{p} \alpha_{k_1} \dots \alpha_{k_n} \mathrm{E}\left[\left(Y_{t_0-k_1-\dots-k_{n+1}}^{\dagger}\right)^2\right] \\
&= \sum_{k_1=1}^{p} \dots \sum_{k_n=1}^{p} \alpha_{k_1} \dots \alpha_{k_n} \mathrm{E}\left[\left(Y_{t_0}^{\dagger}\right)^2\right] \\
&= \mathrm{E}\left[\left(Y_{t_0}^{\dagger}\right)^2\right] \sum_{k_1=1}^{p} \dots \sum_{k_n=1}^{p} \alpha_{k_1} \dots \alpha_{k_n} \\
&= \mathrm{E}\left[\left(Y_{t_0}^{\dagger}\right)^2\right] (\alpha_1 + \dots + \alpha_p)^n ;
\end{aligned} \tag{3.36}$$

the strict inequality follows from Shiryaev [39], page 185, Property H, or Kubrusly [29], page 41, Proposition 37 (indeed: as previously assumed, $\left(Y_{t_0}^{\dagger}\right)^2$ is non-negative and non-degenerated at zero), the second equality comes from the properties of $(\varepsilon_t)_{t\in\mathbb{Z}}$ and from $\left(Y_t^{\dagger}\right)_{t\in\mathbb{Z}}$ being causal, and the fourth one ensues from second-order stationarity. It is clear that (3.36) leads to a contradiction. So, we conclude that $\mathrm{P}\left(Y_t^{\dagger} \neq 0\right) = 0$ for every $t \in \mathbb{Z}$. In other words: P-almost surely, $Y_t^{\dagger} = 0 = Y_t$, for every $t \in \mathbb{Z}$.

Finally, assume that $\sum_{i=1}^{p} \alpha_i = 1$. Let us revisit the existing link between the square of the ARCH process in eq.(3.1) and the usual autoregressive, or AR model, namely

$$\begin{aligned} Y_t^2 = &\quad h_t + Y_t^2 - h_t = \alpha_1 Y_{t-1}^2 + \cdots + \alpha_p Y_{t-p}^2 + (Y_t^2 - h_t) \\ \equiv &\quad \alpha_1 Y_{t-1}^2 + \cdots + \alpha_p Y_{t-p}^2 + \eta_t \ , \end{aligned} \tag{3.37}$$

for each $t \in \mathbb{Z}$. Since $\sum_{i=1}^{p} \alpha_i = 1$ ensures a unit root for the polynomial $\phi(z) = 1 - \alpha_1 z - \alpha_2 z^2 - \ldots - \alpha_p z^p$, $z \in \mathbb{C}$, it follows that, unless $\eta_t = Y_t^2 - h_t = 0$ for each t (which is pretty much equivalent to saying that $Y_t = 0$ for each t), the process described in (3.37) allows no second-order stationary solution whatsoever – see Brockwell & Davis [10], chapters 3 and 4, and Box et al. [8], chapter 4. □

3.2 GARCH

Now we consider the defining difference equation for the GARCH(p, q) model, which for convenience we rewrite again:

$$\begin{aligned} Y_t &= \sqrt{h_t}\varepsilon_t, \\ h_t &= \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}, \end{aligned} \tag{3.38}$$

where $\alpha_0 \geq 0$, $\alpha_i \geq 0$ for $i = 1, \cdots, p$, and $\beta_j \geq 0$ for $j = 1, \cdots, q$. We continue to assume $\varepsilon_t \sim \text{IID}\,(0, 1)$.

Our very first task is to generalize Theorem 1. Notice that, again, Lemma 3 shall be an aid.

Theorem 8. *A sufficient condition for the existence of a both second-order and strictly stationary solution for eqs.* (3.38) *is that* $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$.

Proof. From now on, we assume the restriction above, which, according to Lemma 3, implies that every the root of $1-\beta(z) \equiv 1-\beta_1 z-\beta_2 z^2-\ldots-\beta_q z^q$, $z \in \mathbb{C}$, lies outside the unit circle. Bearing this in mind, we obtain the following equivalences regarding the process $(h_t)_{t\in\mathbb{Z}}$, which use well-known facts about inversion of polynomials on the backward operator B:

$$\begin{aligned} h_t &= \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j} = \alpha_0 + \sum_{i=1}^{p} \alpha_i B^i Y_t^2 + \sum_{j=1}^{q} \beta_j B^j h_t \\ &\equiv \alpha_0 + \alpha(B) Y_t^2 + \beta(B) h_t \\ &\iff [1 - \beta(B)]\, h_t = h_t - \beta(B) h_t = \alpha_0 + \alpha(B) Y_t^2 \\ &\iff h_t = [1 - \beta(B)]^{-1} \alpha_0 + [1 - \beta(B)]^{-1} \alpha(B) Y_t^2 \\ &= \frac{\alpha_0}{1 - \beta(1)} + \delta(B) Y_t^2. \end{aligned} \tag{3.39}$$

In (3.39), we actually have $\delta(z) = \dfrac{\alpha(z)}{1-\beta(z)} = \delta_1 z + \delta_2 z^2 + \delta_3 z^3 + \ldots$ for all z inside the convergence interval of $\dfrac{1}{1-\beta(z)} = 1 + \theta_1 z + \theta_2 z^2 + \ldots$ (see, for instance, Shumway & Stoffer [40], appendix B.2). So, from combining eqs.(3.38) with (3.39), it follows that the GARCH(p,q) process in some kind of ARCH(∞):

$$Y_t = \left(\frac{\alpha_0}{1-\beta(1)} + \sum_{i=1}^{\infty} \delta_i Y_{t-i}^2\right)^{1/2} \varepsilon_t \equiv \tilde{h}_t^{1/2} \varepsilon_t,\ t \in \mathbb{Z}. \tag{3.40}$$

Now, it is a good moment to use every lesson learned from our earlier study on the ARCH process to reach a constructive proof of this theorem.
We begin by preparing a candidate to solve (h_t) in eqs.(3.38) – or, equivalently, to solve $\tilde{h}_t$ in eq.(3.40). As a first step, we define, for each $t \in \mathbb{Z}$,

$$h_t^* = \frac{\alpha_0}{1-\beta(1)} \left\{1 + \sum_{i=1}^{\infty} \left[\sum_{k_1=1}^{\infty} \cdots \sum_{k_i=1}^{\infty} \delta_{k_1} \ldots \delta_{k_i} \varepsilon_{t-k_1}^2 \ldots \varepsilon_{t-k_1-\ldots-k_i}^2\right]\right\}, \tag{3.41}$$

$$Y_t^* = \varepsilon_t \sqrt{\frac{\alpha_0}{1-\beta(1)} \left\{1 + \sum_{i=1}^{\infty} \left[\sum_{k_1=1}^{\infty} \cdots \sum_{k_i=1}^{\infty} \delta_{k_1} \ldots \delta_{k_i} \varepsilon_{t-k_1}^2 \ldots \varepsilon_{t-k_1-\ldots-k_i}^2\right]\right\}}. \tag{3.42}$$

In fact, eq.(3.41) provides a solution for both h_t in eqs.(3.38) and $\tilde{h}_t$ in eq.(3.40), and therefore $Y_t^* = (h_t^*)^{1/2} \varepsilon_t$ proposed in eq.(3.42) solves $(Y_t)_{t\in\mathbb{Z}}$ in both eqs.(3.38) and eq.(3.40). Again, we leave the very lengthy algebra to the Appendix (see section F). More interesting seems to notice the almost perfect match between eq.(3.41) and eq.(3.5); the only (but relevant) difference is that, now in eq.(3.41), $p = \infty$. Hence, even though the proof uses much the same lines as before, some additional and theoretical care in dealing with such "infinite" p is certainly required.
We first recognize that h_t^* is, for each $t \in \mathbb{Z}$, a pointwise double-limit of a non-decreasing and non-negative sequence of random variables – and therefore is itself a random variable. Formally:

$$h_t^* = \lim_{n\to\infty} \lim_{j\to\infty} \left\{\frac{\alpha_0}{1-\beta(1)} \left[1 + \sum_{i=1}^{n} \left(\sum_{k_1=1}^{j} \cdots \sum_{k_i=1}^{j} \delta_{k_1} \ldots \delta_{k_i} \varepsilon_{t-k_1}^2 \ldots \varepsilon_{t-k_1-\ldots-k_i}^2\right)\right]\right\}.$$

Now we handle finiteness:

$$
\begin{aligned}
0 &\leq \mathrm{E}\left(h_t^*\right) \\
&= \mathrm{E}\left(\frac{\alpha_0}{1-\beta(1)}\left\{1+\lim_{n\to\infty}\sum_{i=1}^{n}\left[\lim_{j\to\infty}\sum_{k_1=1}^{j}\dots\sum_{k_i=1}^{j}\delta_{k_1}\dots\delta_{k_i}\varepsilon_{t-k_1}^2\dots\varepsilon_{t-k_1-\ldots-k_i}^2\right]\right\}\right) \\
&= \frac{\alpha_0}{1-\beta(1)}\left\{1+\lim_{n\to\infty}\sum_{i=1}^{n}\left[\lim_{j\to\infty}\sum_{k_1=1}^{j}\dots\sum_{k_i=1}^{j}\delta_{k_1}\dots\delta_{k_i}\mathrm{E}\left(\varepsilon_{t-k_1}^2\right)\dots\mathrm{E}\left(\varepsilon_{t-k_1-\ldots-k_i}^2\right)\right]\right\} \\
&= \frac{\alpha_0}{1-\beta(1)}\left\{1+\lim_{n\to\infty}\sum_{i=1}^{n}\left[\lim_{j\to\infty}\sum_{k_1=1}^{j}\sum_{k_2=2}^{j}\dots\sum_{k_i=1}^{j}\delta_{k_1}\delta_{k_2}\dots\delta_{k_i}\right]\right\} \\
&= \frac{\alpha_0}{1-\beta(1)}\left\{1+\lim_{n\to\infty}\sum_{i=1}^{n}\left[\lim_{j\to\infty}\sum_{k_1=1}^{j}\delta_{k_1}\sum_{k_2=1}^{j}\delta_{k_1}\dots\sum_{k_2=1}^{j}\delta_{k_i}\right]\right\} \\
&= \frac{\alpha_0}{1-\beta(1)}\left\{1+\lim_{n\to\infty}\sum_{i=1}^{n}\left[\sum_{k_1=1}^{\infty}\delta_{k_1}\sum_{k_2=1}^{\infty}\delta_{k_2}\dots\sum_{k_i=1}^{\infty}\delta_{k_i}\right]\right\} \\
&= \frac{\alpha_0}{1-\beta(1)}\left\{1+\sum_{i=1}^{\infty}\left[\underbrace{\delta(1)\delta(1)\dots\delta(1)}_{i\text{ times}}\right]\right\} \\
&= \frac{\alpha_0}{1-\beta(1)}\left\{1+\sum_{i=1}^{\infty}[\delta(1)]^i\right\} < \infty;
\end{aligned}
\tag{3.43}
$$

the second equality comes from a double application of the the Monotone Convergence Theorem and from the hypotheses over $(\varepsilon_t)_{t\in\mathbb{Z}}$, and the very last inequality holds in view that $\delta(1) = \dfrac{\alpha(1)}{1-\beta(1)} < 1$ (since $\alpha(1) < 1-\beta(1)$, or $\alpha(1)+\beta(1) < 1$), which leads therefore to a convergent geometric series. Eqs.(3.43), the fact that h_t^* is non-negative and Jensen's inequality imply that h_t^* is P-almost surely finite, as well as $(h^*)_t^{1/2}$ and Y_t^*. Using all these facts, combined once again with the hypotheses on $(\varepsilon_t)_{t\in\mathbb{Z}}$ (and recall: $Y_t^* = (h_t^*)^{1/2}\,\varepsilon_t,\ t\in\mathbb{Z}$, is causal), we also conclude that Y_t^* is a second order random variable for all $t\in\mathbb{Z}$.
Finally, strictly stationarity (which implies second-order stationarity) ensues very similarly to what we have seen in the proof of Theorem 1. The only one important change is that, for each $n\in\mathbb{N}$, the auxiliary stochastic process this time reads as

$$
Y_t^{(n)} = \left\{\frac{\alpha_0}{1-\beta(1)}\left[1+\sum_{i=1}^{n}\left[\sum_{k_1=1}^{\infty}\dots\sum_{k_i=1}^{\infty}\delta_{k_1}\dots\delta_{k_i}\varepsilon_{t-k_1}^2\dots\varepsilon_{t-k_1-\ldots-k_i}^2\right]\right]\right\}^{1/2}\varepsilon_t\ ,
$$

for all $t \in \mathbb{Z}$. □

Now, we deal with uniqueness – in other words, we are talking the generalization of Theorem 4. For such, it is interesting and useful to gather a closer understanding of the coefficients δ_i that enter eqs.(3.40)-(3.42). The algebra we are looking at occurs as follow[16]. For each z inside the convergence radius of

$$\frac{1}{1-\beta_1 z-\beta_2 z^2-\beta_3 z^3-\ldots-\beta_q z^q}=1+\theta_1 z+\theta_2 z^2+\theta_3 z^3+\ldots,$$

we set

$$\frac{\alpha_1 z+\alpha_2 z^2+\cdots+\alpha_p z^p}{1-\beta_1 z-\beta_2 z^2-\cdots-\beta_q z^q}=\delta_0+\delta_1 z+\delta_2 z^2+\delta_3 z^3+\ldots \tag{3.44}$$

Making $z=0$ on both side of eq.(3.44) implies trivially that $\delta_0=0$. As regards $\delta_1, \delta_2, \delta_3, \ldots$, we begin by noticing that

$$\begin{aligned}
\alpha_1 z+\ldots+\alpha_p z^p &= \left(1-\beta_1 z-\ldots-\beta_q z^q\right)\left(\delta_1 z+\delta_2 z^2+\delta_3 z^3+\ldots\right) \\
&= \delta_1 z+\delta_2 z^2+\delta_3 z^3+\delta_4 z^4+\delta_5 z^5+\ldots \\
&\quad -\beta_1 \delta_1 z^2-\beta_1 \delta_2 z^3-\beta_1 \delta_3 z^4-\beta_1 \delta_4 z^5-\ldots \\
&\quad -\beta_2 \delta_1 z^3-\beta_2 \delta_2 z^4-\beta_2 \delta_3 z^5-\beta_2 \delta_4 z^6+\ldots \\
&\quad \vdots \\
&\quad -\beta_q \delta_1 z^{q+1}-\beta_q \delta_2 z^{q+2}-\beta_q \delta_3 z^{q+3}-\beta_q \delta_4 z^{q+4}-\ldots \\
&= \delta_1 z+\left(\delta_2-\beta_1 \delta_1\right) z^2+\left(\delta_3-\beta_1 \delta_2-\beta_2 \delta_1\right) z^3 \\
&\quad +\left(\delta_4-\beta_1 \delta_3-\beta_2 \delta_2-\beta_3 \delta_1\right) z^4+\ldots+ \\
&\quad \left(\delta_p-\beta_1 \delta_{p-1}-\beta_2 \delta_{p-2}-\ldots-\beta_{p-2} \delta_2-\beta_{p-1} \delta_1\right) z^p \\
&\quad \left(\delta_{p+1}-\beta_1 \delta_p-\beta_2 \delta_{p-1}-\ldots-\beta_p \delta_1\right) z^{p+1}+\ldots
\end{aligned} \tag{3.45}$$

Resorting to the uniqueness of any given power series's coefficients (cf. Bartle & Sherbert [2], section 9.4, Theorem 9.4.13), we obtain from (3.45):

[16] This whole development pretty much resembles the derivation of the coefficients related to both MA(∞) and AR(∞) representations of regular ARMA(p,q) models (cf. Brockwell & Davis [10], section 3.1; and Shumway & Stoffer [40], section 3.1)

$$\begin{aligned}\delta_1 &= \alpha_1,\\ \delta_2 - \beta_1\delta_1 = \alpha_2 &\Rightarrow \delta_2 = \alpha_2 + \beta_1\delta_1,\\ \delta_3 - \beta_1\delta_2 - \beta_2\delta_1 = \alpha_3 &\Rightarrow \delta_3 = \alpha_3 + \beta_1\delta_2 + \beta_2\delta_1,\\ &\vdots\\ \delta_p - \beta_1\delta_{p-1} - \ldots - \beta_{p-1}\delta_1 = \alpha_p &\Rightarrow \delta_p = \alpha_p + \beta_1\delta_{p-1} + \ldots + \beta_{p-1}\delta_1,\\ \delta_{p+1} - \beta_1\delta_p - \ldots - \beta_p\delta_1 = 0 &\Rightarrow \delta_{p+1} = \beta_1\delta_p + \beta_2\delta_{p-1} + \ldots + \beta_p\delta_1,\\ \delta_{p+2} - \beta_1\delta_{p+1} - \ldots - \beta_{p+1}\delta_1 = 0 &\Rightarrow \delta_{p+2} = \beta_1\delta_{p+1} + \ldots + \beta_{p+1}\delta_1,\\ &\vdots\end{aligned} \tag{3.46}$$

and so on. Then, we have proven the following

Lemma 9. (Bollerslev [4], page 310, eq.(5)) *The coefficients of the power series expansion in (3.46), namely $\delta_i, i \in \mathbb{N}$, obeys the recurrence:*

$$\delta_i = \begin{cases} \alpha_i + \sum_{j=1}^{l} \beta_j \delta_{i-j}, & i = 2, 3, \ldots, p, \\ \\ \sum_{j=1}^{l} \beta_j \delta_{i-j}, & i = p+1, p+2, \ldots \end{cases}$$

where $\delta_1 = \alpha_1$ and $l = \min\{q, i-1\}$ for each $i \geq 2$. Also, $\delta_0 = 0$.

Moving on:

Theorem 10. *The stochastic process $(Y_t^*)_{t\in\mathbb{Z}}$ in eq.(3.42) is P-almost surely the unique second stationary solution for eqs.(3.38), provided that at least one of these three conditions holds:*

(i) $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$ *and the considered solutions $\left(\widehat{Y}_t\right)_{t\in\mathbb{Z}}$ are all causal.*

(ii) *There exists $\kappa \geq 2$ such that $\sum_{i=1}^{\infty} \delta_i^{\frac{1}{\kappa}} < 1$, where $\delta_i, i \in \mathbb{N}$, are the coefficients given in Lemma 3.*

(iii) *There exist $n_0 \in \mathbb{N}$ and $\gamma \in \mathbb{R}$, such that $E(\mid Y_{t-j} \mid^4) \leq n^{2\gamma}$, for all $t \in \mathbb{Z}$, $n \geq n_0$ and $j = n+1, \ldots, (n+1)p$. Also, $\sqrt{E(\varepsilon_1^4)}\left(\dfrac{\alpha_1 + \alpha_2 + \ldots + \alpha_p}{1 - \beta_1 - \beta_2 - \ldots - \beta_q}\right) < 1$.*

Proof. Again, at least one of the $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$ is assumed to be strictly positive. Let $(\widehat{Y}_t)_{t\in\mathbb{Z}}$ be another second order stationary solution of eqs.(3.38). Considering eq.(3.40) and proceeding pretty much similarly to what we have done to achieve (3.3) and (3.8), we are led to

$$\widehat{Y}_t = \varepsilon_t \sqrt{\frac{\alpha_0}{1-\beta_1-\ldots-\beta_q}\left[1+\sum_{i=1}^{n}\left(\sum_{k_1=1}^{\infty}\cdots\sum_{k_i=1}^{\infty}\delta_{k_1}\ldots\delta_{k_i}\varepsilon^2_{t-k_1}\ldots\varepsilon^2_{t-k_1-\ldots-k_i}\right)\right] + \sum_{k_1=1}^{\infty}\cdots\sum_{k_{n+1}=1}^{\infty}\delta_{k_1}\ldots\delta_{k_{n+1}}\varepsilon^2_{t-k_1}\ldots\varepsilon^2_{t-k_1-\ldots-k_{n+1}}\widehat{Y}^2_{t-k_1-\ldots-k_{n+1}}}.$$

Just as before, we set

$$W_n = \sum_{k_1=1}^{\infty}\cdots\sum_{k_{n+1}=1}^{\infty}\delta_{k_1}\ldots\delta_{k_{n+1}}\varepsilon^2_{t-k_1}\ldots\varepsilon^2_{t-k_1-\ldots-k_n}\widehat{Y}^2_{t-k_1-\ldots-k_{n+1}}.$$

Condition (i), second order stationary of $\left(\widehat{Y}_t\right)_{t\in\mathbb{Z}}$, the assumptions on $(\varepsilon_t)_{t\in\mathbb{Z}}$ and the Monotone Convergence Theorem imply that

$$\begin{aligned}
\mathrm{E}(W_n) &= \mathrm{E}(\widehat{Y}_1^2)\sum_{k_1=1}^{\infty}\cdots\sum_{k_{n+1}=1}^{\infty}\delta_{k_1}\ldots\delta_{k_{n+1}}\\
&= \mathrm{E}(\widehat{Y}_1^2)\sum_{k_1=1}^{\infty}\delta_{k_1}\ldots\sum_{k_{n+1}=1}^{\infty}\delta_{k_{n+1}}\\
&= \mathrm{E}(\widehat{Y}_1^2)\underbrace{\left(\frac{\alpha_1+\alpha_2+\ldots+\alpha_p}{1-\beta_1-\beta_2-\ldots-\beta_q}\right)\cdots\left(\frac{\alpha_1+\alpha_2+\ldots+\alpha_p}{1-\beta_1-\beta_2-\ldots-\beta_q}\right)}_{n+1 \text{ times}}\\
&= \mathrm{E}(\widehat{Y}_1^2)\left(\frac{\alpha_1+\alpha_2+\ldots+\alpha_p}{1-\beta_1-\beta_2-\ldots-\beta_q}\right)^{n+1}.
\end{aligned}$$

The rest of the proof goes as we have done with Theorem 4.
On the other hand, if condition (ii) holds, we obtain

$$\begin{aligned}
0 &\le \mathrm{E}(W_n^{\frac{1}{[\kappa]}})\\
&= \mathrm{E}\left[\left(\sum_{k_1=1}^{\infty}\cdots\sum_{k_{n+1}=1}^{\infty}\delta_{k_1}\ldots\delta_{k_{n+1}}\varepsilon^2_{t-k_1}\ldots\varepsilon^2_{t-k_1-\ldots-k_n}\widehat{Y}^2_{t-k_1-\ldots-k_{n+1}}\right)^{\frac{1}{[\kappa]}}\right]\\
&\le \mathrm{E}\left[\sum_{k_1=1}^{\infty}\cdots\sum_{k_{n+1}=1}^{\infty}\delta_{k_1}^{\frac{1}{[\kappa]}}\ldots\delta_{k_{n+1}}^{\frac{1}{[\kappa]}}|\varepsilon_{t-k_1}\ldots\varepsilon_{t-k_1-\ldots-k_n}\widehat{Y}_{t-k_1-\ldots-k_{n+1}}|^{\frac{2}{[\kappa]}}\right]\\
&= \sum_{k_1=1}^{\infty}\cdots\sum_{k_{n+1}=1}^{\infty}\delta_{k_1}^{\frac{1}{[\kappa]}}\ldots\delta_{k_{n+1}}^{\frac{1}{[\kappa]}}\mathrm{E}\left(|\,\varepsilon_{t-k_1}\ldots\varepsilon_{t-k_1-\ldots-k_{n+1}}\widehat{Y}_{t-k_1-\ldots-k_{n+1}}\,|^{\frac{2}{[\kappa]}}\right),
\end{aligned} \tag{3.47}$$

where the second inequality follows from Lemma 2, and the second equality is another application of the Monotone Convergence Theorem. From (3.47), Jensen's inequality, Cauchy-Schwarz inequality and second-order stationary of $\left(\widehat{Y}_t\right)_{t\in\mathbb{Z}}$ (also recall that $\varepsilon_t \sim \text{IID}(0,1)$),

$$
\begin{aligned}
0 &\leq \mathrm{E}(W_n^{\frac{1}{[\kappa]}}) \\
&\leq \sum_{k_1=1}^{\infty}\cdots\sum_{k_{n+1}=1}^{\infty}\delta_{k_1}^{\frac{1}{[\kappa]}}\cdots\delta_{k_{n+1}}^{\frac{1}{[\kappa]}}\left(\sqrt{\mathrm{E}(\varepsilon_{t-k_1}^2)\ldots\mathrm{E}(\varepsilon_{t-k_1-\ldots-k_{n+1}}^2)\mathrm{E}(\widehat{Y}_{t-k_1-\ldots-k_{n+1}}^2)}\right)^{\frac{2}{[\kappa]}} \\
&= \left(\mathrm{E}(\widehat{Y}_1^2)\right)^{\frac{1}{[\kappa]}}\left(\delta_1^{\frac{1}{[\kappa]}}+\delta_2^{\frac{1}{[\kappa]}}+\delta_3^{\frac{1}{[\kappa]}}+\ldots\right)^{n+1} \\
&\leq \left(\mathrm{E}(\widehat{Y}_1^2)\right)^{\frac{1}{[\kappa]}}\left(\delta_1^{\frac{1}{\kappa}}+\delta_2^{\frac{1}{\kappa}}+\delta_3^{\frac{1}{\kappa}}+\ldots\right)^{n+1}.
\end{aligned}
$$

One more time, the proof repeats.
As concerns conditions (iii), there is actually no substantial change from the final writings of Theorem 4's proof, except for the need to evoke the Monotone Convergence Theorem in exchanging expected values with infinity sums of non-negative random variables. □

A generalization of Theorem 5 also holds:

Theorem 11. *Assume the restriction $\sum_{i=1}^{p}\alpha_i+\sum_{j=1}^{q}\beta_j<1$. The stochastic process $(Y_t^*)_{t\in\mathbb{Z}}$ in eq.(3.42) is $P-$almost surely the unique strictly stationary solution (either causal or not) for eqs.(3.38).*

Proof. The very same equations in (3.19) and (3.20) work just fine for the GARCH model in eqs.(3.38), but with some changes regarding the matrices definition:

$$
\underline{\widetilde{h}}_t = \begin{bmatrix} \widetilde{h}_t \\ \widetilde{Y}_{t-1}^2 \\ \widetilde{Y}_{t-2}^2 \\ \vdots \\ \widetilde{Y}_{t-(p-1)}^2 \\ \widetilde{h}_{t-1} \\ \widetilde{h}_{t-2} \\ \vdots \\ \widetilde{h}_{t-(q-1)} \\ \widetilde{h}_{t-q} \end{bmatrix}, \qquad \underline{\alpha}_0 = \begin{bmatrix} \alpha_0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix},
$$

$$
A = \begin{bmatrix}
\alpha_1 & \alpha_2 & \cdots & \alpha_{p-1} & \alpha_p & \beta_1 & \beta_2 & \cdots & \beta_{q-1} & \beta_q \\
1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 1
\end{bmatrix},
$$

and

$$
\underline{\widetilde{Y}}_t^2 = \begin{bmatrix}
\widetilde{Y}_t^2 \\ \widetilde{Y}_{t-1}^2 \\ \widetilde{Y}_{t-2}^2 \\ \vdots \\ \widetilde{Y}_{t-(p-1)}^2 \\ \widetilde{h}_t \\ \widetilde{h}_{t-1} \\ \vdots \\ \widetilde{h}_{t-(q-1)}
\end{bmatrix}, \qquad
\underline{\alpha}_t = \begin{bmatrix}
\alpha_0 \varepsilon_t^2 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \tilde{h}_t - \tilde{Y}_{t-p}^2 \\ 0 \\ \vdots \\ 0
\end{bmatrix},
$$

$$
A_t = \begin{bmatrix}
\alpha_1 \varepsilon_t^2 & \alpha_2 \varepsilon_t^2 & \cdots & \alpha_{p-1} \varepsilon_t^2 & \alpha_p \varepsilon_t^2 & \beta_1 \varepsilon_t^2 & \beta_2 \varepsilon_t^2 & \cdots & \beta_{q-1} \varepsilon_t^2 & \beta_q \varepsilon_t^2 \\
1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & 0
\end{bmatrix}.
$$

The remainder matches exactly Theorem 5's proof. □

Two results remain yet to be examined. One is the reciprocal of Theorem 8 for the case of $\alpha_0 > 0$), the other is a generalization of Theorem 7. The proof of the former carries out virtually as much as in Theorem 6 (again: once expendable, conditional expectations never show up), whereas the latter requires us, for the case of $\sum_{i=1}^{p}\alpha_i+\sum_{j=1}^{q}\beta_j>1$, to convince ourselves that (3.36) becomes

$$0< \mathrm{E}\left[\left(Y_{t_0}^{\dagger}\right)^2\right]=\mathrm{E}\left[\left(Y_{t_0}^{\dagger}\right)^2\right]\left(\sum_{i=1}^{p}\alpha_i+\sum_{j=1}^{q}\beta_j\right)^n, \tag{3.48}$$

and, when $\sum_{j=1}^{p}\alpha_j+\sum_{j=1}^{q}\beta_j=1$, to explore the link between GARCH and autoregressive moving average (ARMA) models, which for every $t\in\mathbb{Z}$ reads as

$$Y_t^2=\alpha_0+(\alpha_1+\beta_1)Y_{t-1}^2+\ldots+(\alpha_r+\beta_r)Y_{t-r}^2+\eta_t-\beta_1\eta_{t-1}-\ldots-\beta_q\eta_{t-q}, \tag{3.49}$$

where $r=\max\{p,q\}$, $\alpha_i=0$ for $i>p$, $\beta_j=0$ for $j>q$, and $\eta_t=Y_t^2-h_t$.

We invite you, who have come so far studying this survey, to confirm both (3.48) and (3.49) as well as to detail the proofs of upcoming Theorems 12 and 13. Two hints regarding (3.48):

(i) Work out a solution iterating eqs.(3.38) n times – in the same way we have done with the ARCH process eq.(3.1), in order to derive (3.3). The writings may be quite lengthy but still not complicated or technical.

(ii) After taking expectations on both sides of your solution for eqs.(3.38), recall that

$$\left(\sum_{i=1}^{p}\alpha_i+\sum_{j=1}^{q}\beta_j\right)^{n+1}=\underbrace{\left(\sum_{i=1}^{p}\alpha_i+\sum_{j=1}^{q}\beta_j\right)\left(\sum_{i=1}^{p}\alpha_i+\sum_{j=1}^{q}\beta_j\right)\cdots\left(\sum_{i=1}^{p}\alpha_i+\sum_{j=1}^{q}\beta_j\right)}_{n+1\text{ times}}.$$

Theorem 12. *Assume eqs.(3.38) admit a solution, which is second-order causal stationary. Also, assume that* $\alpha_0>0$. *Then,* $\sum_{i=1}^{p}\alpha_i+\sum_{j=1}^{q}\beta_j<1$.

Theorem 13. *If* $\alpha_0=0$, *then the* P*-almost surely unique second-order stationary causal solution for eqs.(3.38) is* $Y_t=0$, $t\in\mathbb{Z}$.

4 A glimpse at IGARCH

Our survey nears the end with a concise discussion about the Integrated GARCH, or IGARCH process. This model was originally discussed by Engle & Bollerslev [19] and received a very advanced treatment by Daniel B. Nelson, in which concerns theoretical properties of the still nowadays influential IGARCH(1,1) specification (cf. Nelson [35]). The main motivation behind the IGARCH process had been the empirical evidence of a stronger persistence in volatility dynamics of financial time series: parameters estimates of the GARCH model frequently suggested that $\sum_{i=1}^{p}\alpha_i+\sum_{j=1}^{q}\beta_j\approx 1$; for ARCH specifications, $\sum_{i=1}^{p}\alpha_i\approx 1$ with generally appreciable values of p. Actually there was more: keeping in mind the link between

GARCH and ARMA revisited in eqs.(3.37) and (3.49), time series plots and sample auto-correlation functions of squared stock returns used to reveal a persistence substantial to the point of making financial econometricians question whether unit roots might be found.

To fully understand the above preamble, we rephrase eqs.(3.49) after working explicitly the backward operator:

$$\begin{aligned}
h_t &= \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j} = \alpha_0 + \alpha(B) Y_t^2 + \beta(B) h_t \\
&\iff -\alpha(B) Y_t^2 = \alpha_0 - h_t + \beta(B) h_t \\
&\iff Y_t^2 - \alpha(B) Y_t^2 - \beta(B) Y_t^2 = \alpha_0 + Y_t^2 - h_t + \beta(B) h_t - \beta(B) Y_t^2 \\
&\iff (1 - \alpha(B) - \beta(B))\, Y_t^2 = \alpha_0 + \left(Y_t^2 - h_t\right) - \beta(B) \left(Y_t^2 - h_t\right) \\
&\iff (1 - \alpha(B) - \beta(B))\, Y_t^2 = \alpha_0 + \eta_t - \beta(B) \eta_t \\
&\iff (1 - \alpha(B) - \beta(B))\, Y_t^2 = \alpha_0 + (1 - \beta(B))\, \eta_t,
\end{aligned}$$

where $\alpha(z) = \alpha_1 z + \alpha_2 z^2 + \ldots + \alpha_p z^p$ and $\beta(B) = \beta_1 z + \beta_2 z^2 + \ldots + \beta_q z^q$ for any $z \in \mathbb{C}$, and $\eta_t = Y_t^2 - h_t$. Then, eqs.(3.38) define a IGARCH(p, q) process iff the polynomial $1 - \alpha(.) - \beta(.)$ has a root at $z = 1$ of multiplicity one, and the other roots are outside the unit circle. In other words:

$$1 - \alpha(z) - \beta(z) = (1 - z)\phi(z), \quad z \in \mathbb{C},$$

where $\phi(.)$ is a polynomial with all roots outside the unit circle. Or, if we go back to the original writings:
$(Y_t)_{t \in \mathbb{Z}}$ in eqs.(3.38) follows a IGARCH(p, q) process iff $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1$.

Virtually every time series book previously reviewed in section 2 dedicates at least some space to discuss the IGARCH process. See for instance: Harvey [25], page 278; Hamilton [24], page 667; Enders [17], pages 154-155; and Box et al. [8], page 376. However, in which regards the classroom standpoint, we understand the subject still calls for some deeper treatment. The following 3-step road map (essentially the very one taken by us in this manuscript) might well make good theme for a future survey: (i) the study of the conditions that guarantee existence of an IGARCH process; (ii) questions regarding strict stationarity, which is the only possibility left given that, as all four books re-cited in this paragraph affirm (without proving, though), second order moments explode whenever $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1$; and (iii) should it hold, a proof of the P-almost surely uniqueness of IGARCH's difference equations solution.

Acknowledgements: We are indebted to an anonymous Referee, as well as to Cristiano Fernandes, Javier Ribón and Wilson Calmon for constructive remarks and suggestions. Jorge Bustamante is grateful to the financial support from both Coordenação de Aperfeiçoamento de Pessoal do Nível Superior (CAPES) (MSc scholarship code: 001) and Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) (MSc scholarship code: 202.385/2024).

Appendix

The upcoming topics are structured as follows. Section A explores a product of two random variables with well-defined expectation, which itself does not have expectation. Section B briefly reviews some important probabilistic results that are intensively used in section 3. Sections C and D provide two different ways for attaining the general solution of ARCH's difference equation. Section E extends Theorem 5 to scenarios that require a more general matrix diagonalization theory. Finally, the solution of the GARCH process's equations is detailed in section F.

A An interesting counterexample in probability

Consider the following

Claim 14. *Let X and Y two independent random variables. If $E(X)$ is well-defined (but not necessarily finite) and Y is integrable (that is: $E(Y)$ exists and is finite), then $E(XY)$ exists and $E(XY) = E(X)E(Y)$.*

As we shall see, Claim 14 is *false*. Consider two continuous and independent random variables X and Y, whose density functions are

$$f_X(x) = \begin{cases} \frac{1}{x^2}, x \geq 1, \\ 0, \text{otherwise}, \end{cases} \qquad \text{and} \qquad f_Y(y) = \begin{cases} \frac{1}{2}, -1 \leq y \leq 1, \\ 0, \text{otherwise}. \end{cases}$$

(recall: $Y \sim U[-1, 1]$). So, the random vector (X, Y) is continuous with density

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) = \begin{cases} \frac{1}{2x^2}, x \geq 1 \text{ and } -1 \leq y \leq 1, \\ 0, \text{otherwise}. \end{cases}$$

By definition, the expected value of the continuous random variable $Z = XY$ exists in $\overline{R} = [-\infty, +\infty]$ if and only if

$$\int_{-\infty}^{0} z f_Z(z) dz > -\infty \qquad \text{or} \qquad \int_{0}^{\infty} z f_Z(z) dz < \infty.$$

This is equivalent to saying that

$$I_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy I_{\{(r,t) \in \mathbb{R}^2 : rt \geq 0\}}(x, y) f_{X,Y}(x, y) dy dx < \infty \tag{A.1}$$

or

$$I_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyI_{\{(r,t)\in\mathbb{R}^2:rt<0\}}(x,y)f_{X,Y}(x,y)dydx > -\infty. \tag{A.2}$$

Let us deal with the integral in eq.(A.1):

$$\begin{aligned}
I_1 &= \int_{1}^{\infty} \int_{-1}^{1} xyI_{\{(r,t)\in\mathbb{R}^2:rt\geq 0\}}(x,y)\frac{1}{2x^2}dydx \\
&= \int_{1}^{\infty} \int_{-1}^{1} xyI_{\{(r,t)\in\mathbb{R}^2:t\geq 0\}}(x,y)\frac{1}{2x^2}dydx \\
&= \int_{1}^{\infty} \int_{0}^{1} xy\frac{1}{2x^2}dydx \\
&= \frac{1}{2}\int_{1}^{\infty} \frac{1}{x}\left(\int_{0}^{1} ydy\right)dx \\
&= \frac{1}{2}\int_{1}^{\infty} \frac{1}{x}\left.\frac{y^2}{2}\right|_{0}^{1}dx \\
&= \frac{1}{4}\int_{1}^{\infty} \frac{1}{x}dx \\
&= \frac{1}{4}\log x|_{1}^{\infty} \\
&= +\infty.
\end{aligned}$$

Analogously, I_2 in eq.(A.2) diverges to $-\infty$. So, we are lead to conclude: $\mathrm{E}[Z]$ is not well-defined.
On the other hand, both $\mathrm{E}[X]$ and $\mathrm{E}[Y]$ exist in $\overline{\mathbb{R}}$. Indeed, the former is ∞ whereas the latter equals 0.

B Revisiting some classic probability results

In what follows, we consider a general probability space $(\Omega, \mathcal{F}, \mathrm{P})$. The results below split into three major types: (i) *sequence of random variables and convergence results*; (ii) *inequalities involving expectation*; and (iii) *other general results*. We assume that every random variable (properly defined on the aforementioned probability space) is allowed to be *extended* real-valued; that is, it may also yield $+\infty$ or $-\infty$.

B.1 Sequence of random variables and convergence results

Lemma 15. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables that converges pointwise to a function, let us say $X : \Omega \longrightarrow \overline{\mathbb{R}}$. Then X is a itself random variable on the same probability space.*

Lemma 16 (Monotone Convergence Theorem). *Let $(X_n)_{n\in\mathbb{N}}$ be a increasing sequence of non-negative random variables. If $(X_n) \to X$ (either pointwise or P-almost surely), then $E(X_n) \to E(X)$.*

Lemma 17. *If $(X_n)_{n\in\mathbb{N}}$ is a sequence of random variables and X is another random variable, then the following equivalence of events holds:*

$$\left(\lim_{n\to\infty} X_n = X\right) = \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{l=n}^{\infty} \left(\mid X_l - X \mid < \frac{1}{m}\right),$$

where $\left(\lim_{n\to\infty} X_n = X\right) \equiv \{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$.

Lemma 18 (Slutsky's Theorem – multiplication part). *If $X_n \stackrel{d}{\to} X$ and $Y_n \stackrel{p}{\to} c$, then $Y_n X_n \stackrel{d}{\to} cX$.*

Lemma 19 (Continuous Mapping Theorem). *Let $X, X_1, X_2, \ldots$ be m-dimensional random vectors such that $X_n \stackrel{d}{\to} X$, and let $h : \mathbb{R}^m \to \mathbb{R}^k$ be a Borel-measurable function such that its set of discontinuity points D_h satisfies $P(X \in D_h) = 0$. Then, $h(X_n) \stackrel{d}{\to} h(X)$.*

B.2 Inequalities involving expectation

Lemma 20. *Let X and Y two random variables with well-defined expectations. If $X \leq Y$, then $E(X) \leq E(Y)$. In particular: if X is a non-negative random variable, then $E(X) \geq 0$.*

Lemma 21 (Markov's Inequality). *Let X be a non-negative random variable. For all $\lambda > 0$,*

$$P(X \geq \lambda) \leq \frac{E(X)}{\lambda}.$$

Lemma 22 (Jensen's Inequality). *Let $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ be a convex function. If X is an integrable random variable, then $\varphi(E(X))$ is a random variable with well-defined expectation (not necessarily finite) and $\varphi(E(X)) \leq E(\varphi(X))$.*

Lemma 23 (Cauchy-Schwarz Inequality). *Let X and Y be random variable with finite second-order moments. Then,*

$$E(\mid XY \mid) \leq \sqrt{E(X^2)\, E(Y^2)}.$$

B.3 Other general results

Lemma 24. *If X is integrable, then X is finite P-almost surely.*

Lemma 25. *Let X be a non-negative random variable. Then $E(X) = 0$ if only if $X = 0$ P-almost surely.*

Lemma 26 (Borel-Cantelli Lemma – convergence part)**.** *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of events in $\mathcal{F}$. If $\sum_{i=1}^{\infty} \mathrm{P}(A_i) < \infty$, then $\mathrm{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = 0$.*

Lemma 27. *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of events in $\mathcal{F}$ such that $\mathrm{P}(A_n) = 1$ for each $n \in \mathbb{N}$. Then, $\mathrm{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = 1$.*

C Solving eq.(3.1): a first possibility

Let

$$h^*_{t-k_1-\ldots-k_{n+1}} = \alpha_0 + \alpha_0 \sum_{j=1}^{\infty} \left(\sum_{l_1=1}^{p} \cdots \sum_{l_j=1}^{p} \alpha_{l_1} \ldots \alpha_{l_j} \varepsilon^2_{t-k_1-\ldots-k_{n+1}-l_1} \cdots \varepsilon^2_{t-k_1-\ldots-k_{n+1}-l_1-\ldots-l_j} \right). \tag{C.1}$$

Replacing eq.(C.1) in eq.(3.3) (changing i to j, and k to l), we obtain

$$
\begin{aligned}
&\alpha_0\left[1+\sum_{j=1}^{n}\left(\sum_{l_1=1}^{p}\cdots\sum_{l_j=1}^{p}\alpha_{l_1}\ldots\alpha_{l_j}\varepsilon^2_{t-l_1}\ldots\varepsilon^2_{t-l_1-\ldots-l_j}\right)\right]\\
&+\alpha_0\sum_{l_1=1}^{p}\cdots\sum_{l_{n+1}=1}^{p}\alpha_{l_1}\ldots\alpha_{l_{n+1}}\varepsilon^2_{t-l_1}\ldots\varepsilon^2_{t-l_1-\ldots-l_{n+1}}\\
&+\alpha_0\sum_{j=1}^{\infty}\left[\left(\sum_{l_1=1}^{p}\cdots\sum_{l_j=1}^{p}\alpha_{l_1}\ldots\alpha_{l_j}\varepsilon^2_{t-k_1-\ldots-k_{n+1}-l_1}\ldots\varepsilon^2_{t-k_1-\ldots-k_{n+1}-l_1-\ldots-l_j}\right)\right.\\
&\left.\left(\sum_{k_1=1}^{p}\cdots\sum_{k_{n+1}=1}^{p}\alpha_{k_1}\ldots\alpha_{k_{n+1}}\varepsilon^2_{t-k_1}\ldots\varepsilon^2_{t-k_1-\ldots-k_{n+1}}\right)\right]\\
=&\alpha_0\left[1+\sum_{j=1}^{n}\left(\sum_{l_1=1}^{p}\cdots\sum_{l_j=1}^{p}\alpha_{l_1}\ldots\alpha_{l_j}\varepsilon^2_{t-l_1}\ldots\varepsilon^2_{t-l_1-\ldots-l_j}\right)\right]\\
&+\alpha_0\sum_{l_1=1}^{p}\cdots\sum_{l_{n+1}=1}^{p}\alpha_{l_1}\ldots\alpha_{l_{n+1}}\varepsilon^2_{t-l_1}\ldots\alpha^2_{t-l_1-\ldots-l_{n+1}}\\
&+\alpha_0\sum_{j=1}^{\infty}\left[\sum_{k_1=1}^{p}\cdots\sum_{k_{n+1}=1}^{p}\sum_{l_1=1}^{p}\cdots\sum_{l_j=1}^{p}\alpha_{k_1}\ldots\alpha_{k_{n+1}}\alpha_{l_1}\ldots\alpha_{l_j}\varepsilon^2_{t-k_1}\ldots\varepsilon^2_{t-k_1\ldots-k_{n+1}}\right.\\
&\left.\varepsilon^2_{t-k_1\ldots-k_{n+1}-l_1}\ldots\varepsilon^2_{t-k_1\ldots-k_{n+1}-l_1\ldots-l_j}\right]\\
=&\alpha_0+\alpha_0\sum_{j=1}^{\infty}\left(\sum_{l_1=1}^{p}\cdots\sum_{l_j=1}^{p}\alpha_{l_1}\ldots\alpha_{l_j}\varepsilon^2_{t-l_1}\ldots\varepsilon^2_{t-l_1\ldots-l_j}\right)\\
\equiv& h^*_t.
\end{aligned}
$$

D Solving eq.(3.1): another possibility

Let

$$
h^*_{t-i}=\alpha_0+\alpha_0\sum_{j=1}^{\infty}\left(\sum_{k_1=1}^{p}\cdots\sum_{k_j=1}^{p}\alpha_{k_1}\ldots\alpha_{k_j}\varepsilon^2_{t-i-k_1}\ldots\varepsilon^2_{t-i-k_1\ldots-k_j}\right)\;,\quad i=1,\ldots,p. \tag{D.1}
$$

Placing eq.(D.1) in the original expression fot h_t (see eq.(3.1)):

$$
\begin{aligned}
&\alpha_0+\sum_{i=1}^{p}\alpha_i\varepsilon_{t-i}^2h_{t-i}^*\\
&=\alpha_0+\sum_{i=1}^{p}\alpha_i\varepsilon_{t-i}^2\left[\alpha_0+\alpha_0\sum_{j=1}^{\infty}\left(\sum_{k_1=1}^{p}\cdots\sum_{k_j=1}^{p}\alpha_{k_1}\ldots\alpha_{k_j}\varepsilon_{t-i-k_1}^2\ldots\varepsilon_{t-i-k_1-\ldots-k_j}^2\right)\right]\\
&=\alpha_0+\alpha_0\sum_{i=1}^{p}\alpha_i\varepsilon_{t-i}^2+\alpha_0\sum_{j=1}^{\infty}\left[\sum_{i=1}^{p}\sum_{k_1=1}^{p}\cdots\sum_{k_j=1}^{p}\alpha_i\alpha_{k_1}\ldots\alpha_{k_j}\varepsilon_{t-i}^2\varepsilon_{t-i-k_1}^2\ldots\varepsilon_{t-i-k_1-\ldots-k_j}^2\right]\\
&=\alpha_0+\alpha_0\sum_{j=1}^{\infty}\left(\sum_{k_1=1}^{p}\cdots\sum_{k_j=1}^{p}\alpha_{k_1}\ldots\alpha_{k_j}\varepsilon_{t-k_1}^2\ldots\varepsilon_{t-k_1-\ldots-k_j}^2\right)=h_t^*.
\end{aligned}
$$

So, it is straightforward that

$$
\left(\alpha_0+\sum_{i=1}^{p}\alpha_i\left(Y_{t-i}^*\right)^2\right)^{1/2}\varepsilon_t=\left(\alpha_0+\sum_{i=1}^{p}\alpha_ih_t^*\varepsilon_{t-i}^2\right)^{1/2}\varepsilon_t=\left(h_t^*\right)^{1/2}\varepsilon_t=Y_t^*.
$$

E Revisiting Theorem 5's proof

The matrix A used in eq.(3.19) might have some repeated eigenvalues, something that calls for a diagonalization process more general than the one used in eqs.(3.24). We need Jordan blocks for achieving the task. Formally,

$$
A=TJT^{-1}, \tag{E.1}
$$

where

$$
T=\begin{bmatrix}T_{11}&T_{12}&\ldots&T_{1s}\\T_{21}&T_{22}&\ldots&T_{2s}\\\vdots&\vdots&&\vdots\\T_{s1}&T_{s2}&\ldots&T_{ss}\end{bmatrix},\ J=\begin{bmatrix}J_1&0&0&\ldots&0\\0&J_2&0&\ldots&0\\\vdots&\vdots&\vdots&&\vdots\\0&0&0&\ldots&J_R\end{bmatrix},
$$

$$
J_r=\begin{bmatrix}\lambda_r&1&0&\ldots&0&0\\0&\lambda_r&1&\ldots&0&0\\\vdots&\vdots&\vdots&&\vdots&\vdots\\0&0&0&\ldots&\lambda_r&1\\0&0&0&\ldots&0&\lambda_r\end{bmatrix}_{n_r\times n_r},
$$

for each $r=1,\ldots,R$. In the matrices above, λ_r is one of the R *distinct* eigenvalues of A that repeats itself n_r times. Eq.(E.1) leads to

$$A^n = \begin{bmatrix} T_{11} & T_{12} & \dots & T_{1s} \\ T_{21} & T_{22} & \dots & T_{2s} \\ \vdots & \vdots & & \vdots \\ T_{s1} & T_{s2} & \dots & T_{ss} \end{bmatrix} \begin{bmatrix} J_1^n & 0 & 0 & \dots & 0 \\ 0 & J_2^n & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & J_s^n \end{bmatrix} \begin{bmatrix} T^{11} & T^{12} & \dots & T^{1s} \\ T^{21} & T^{22} & \dots & T^{2s} \\ \vdots & \vdots & & \vdots \\ T^{s1} & T^{s2} & \dots & T^{ss} \end{bmatrix}$$

$$= \begin{bmatrix} T_{11}J_1^n & T_{12}J_2^n & T_{13}J_3^n & \dots & T_{1s}TJ_s^n \\ T_{21}J_1^n & T_{22}J_2^n & T_{23}J_3^n & \dots & T_{2s}TJ_s^n \\ \vdots & \vdots & \vdots & & \vdots \\ T_{s1}J_1^n & T_{s2}J_2^n & T_{s3}J_3^n & \dots & T_{ss}TJ_s^n \end{bmatrix} \begin{bmatrix} T^{11} & T^{12} & \dots & T^{1s} \\ T^{21} & T^{22} & \dots & T^{2s} \\ \vdots & \vdots & & \vdots \\ T^{s1} & T^{s2} & \dots & T^{ss} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{r=1}^{s} T_{1r}J_r^nT^{r1} & \sum_{r=1}^{s} T_{1r}J_r^nT^{r2} & \dots & \sum_{r=1}^{s} T_{1r}J_r^nT^{rs} \\ \sum_{r=1}^{s} T_{2r}J_r^nT^{r1} & \sum_{r=1}^{s} T_{2r}J_r^nT^{r2} & \dots & \sum_{r=1}^{s} T_{2r}J_r^nT^{rs} \\ \vdots & \vdots & & \vdots \\ \sum_{r=1}^{s} T_{sr}J_r^nT^{r1} & \sum_{r=1}^{s} T_{sr}J_r^nT^{r2} & \dots & \sum_{r=1}^{s} T_{sr}J_r^nT^{rs} \end{bmatrix}.$$

Now, let $A_{lk}^n = \sum_{r=1}^{s} T_{lr}J_r^nT^{rk}$ and $B_{lk} = T_{lr}J_r^nT^{rk}$, where

$$J_r^n = \begin{bmatrix} \lambda_r^n & \binom{n}{1}\lambda_r^{n-1} & \binom{n}{2}\lambda_r^{n-2} & \dots & \binom{n}{n_r-1}\lambda_r^{n-(n_r-1)} \\ 0 & \lambda_r^n & \binom{n}{1}\lambda_t^{n-1} & \dots & \binom{n}{n_r-2}\lambda_r^{n-(n_r-2)} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \lambda_r^n \end{bmatrix}.$$

Denote by $t_{lr,ij}$ and $t^{rk,ij}$ the elements of T_{lr} and T^{rk}, respectively. It then must follow that

$$B_{lk} = \begin{bmatrix} t_{lr,11} & t_{lr,12} & \dots & t_{lr,1n_r} \\ t_{lr,21} & t_{lr,22} & \dots & t_{lr,2n_r} \\ \vdots & \vdots & & \vdots \\ t_{lr,n_r1} & t_{lr,n_r2} & \dots & t_{lr,n_rn_r} \end{bmatrix} \begin{bmatrix} \lambda_r^n & \binom{n}{1}\lambda_r^{n-1} & \binom{n}{2}\lambda_r^{n-2} & \dots & \binom{n}{n_r-1}\lambda_r^{n-(n_r-1)} \\ 0 & \lambda_r^n & \binom{n}{1}\lambda_t^{n-1} & \dots & \binom{n}{n_r-2}\lambda_r^{n-(n_r-2)} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \lambda_r^n \end{bmatrix}$$

$$\begin{bmatrix} t^{rk,11} & t^{rk,12} & \dots & t^{rk,1n_r} \\ t^{rk,21} & t^{rk,22} & \dots & t^{rk,2n_r} \\ \vdots & \vdots & & \vdots \\ t^{rk,n_r1} & t^{rk,n_r2} & \dots & t^{rk,n_rn_r} \end{bmatrix} \qquad \text{(product of three matrices)}$$

$$= \begin{bmatrix} t_{lr,11}\lambda_r^n & \sum_{j=0}^{1}\binom{n}{j} t_{lr,1(2-j)}\lambda_r^{n-j} & \dots & \sum_{j=0}^{n_r-1}\binom{n}{j} t_{lr,1(n_r-j)}\lambda_r^{n-j} \\ t_{lr,21}\lambda_r^n & \sum_{j=0}^{1}\binom{n}{j} t_{lr,2(2-j)}\lambda_r^{n-j} & \dots & \sum_{j=0}^{n_r-1}\binom{n}{j} t_{lr,2(n_r-j)}\lambda_r^{n-j} \\ \vdots & \vdots & & \vdots \\ t_{lr,n_r1}\lambda_r^n & \sum_{j=0}^{1}\binom{n}{j} t_{lr,n_r(2-j)}\lambda_r^{n-j} & \dots & \sum_{j=0}^{n_r-1}\binom{n}{j} t_{lr,n_r(n_r-j)}\lambda_r^{n-j} \end{bmatrix}$$

$$\begin{bmatrix} t^{rk,11} & t^{rk,12} & \dots & t^{rk,1n_r} \\ t^{rk,21} & t^{rk,22} & \dots & t^{rk,2n_r} \\ \vdots & \vdots & & \vdots \\ t^{rk,n_r1} & t^{rk,n_r2} & \dots & t^{rk,n_rn_r} \end{bmatrix} \qquad \text{(product of two matrices)}$$

$$= \begin{bmatrix} \sum_{q=1}^{n_r}\sum_{j=0}^{q-1}\binom{n}{j} t_{lr,1(q-j)}\lambda_r^{n-j}t^{rk,q1} & \sum_{q=1}^{n_r}\sum_{j=0}^{q-1}\binom{n}{j} t_{lr,1(q-j)}\lambda_r^{n-j}t^{rk,q2} & \dots & \sum_{q=1}^{n_r}\sum_{j=0}^{q-1}\binom{n}{j} t_{lr,1(q-j)}\lambda_r^{n-j}t^{rk,qn_r} \\ \sum_{q=1}^{n_r}\sum_{j=0}^{q-1}\binom{n}{j} t_{lr,2(q-j)}\lambda_r^{n-j}t^{rk,q1} & \sum_{q=1}^{n_r}\sum_{j=0}^{q-1}\binom{n}{j} t_{lr,2(q-j)}\lambda_r^{n-j}t^{rk,q2} & \dots & \sum_{q=1}^{n_r}\sum_{j=0}^{q-1}\binom{n}{j} t_{lr,2(q-j)}\lambda_r^{n-j}t^{rk,qn_r} \\ \vdots & \vdots & & \vdots \\ \sum_{q=1}^{n_r}\sum_{j=0}^{q-1}\binom{n}{j} t_{lr,n_r(q-j)}\lambda_r^{n-j}t^{rk,q1} & \sum_{q=1}^{n_r}\sum_{j=0}^{q-1}\binom{n}{j} t_{lr,n_r(q-j)}\lambda_r^{n-j}t^{rk,q2} & \dots & \sum_{q=1}^{n_r}\sum_{j=0}^{q-1}\binom{n}{j} t_{lr,n_r(q-j)}\lambda_r^{n-j}t^{rk,qn_r} \end{bmatrix}.$$

Now, if $b_{k_1k_2}$ are the elements of B_{lk}, we obtain

$$b_{k_1k_2} = \sum_{q=1}^{n_r}\sum_{j=0}^{q-1}\binom{n}{j} t_{lr,k_1(q-j)}\lambda_r^{n-j}t^{rk,qk_2}.$$

We also define $a_n^{lk} = A_{t-1}A_{t-2}\ldots A_{t-n}$, which, in view of

$$A_{lk}^n = \sum_{r=1}^{s}\sum_{q=1}^{n_r}\sum_{j=0}^{q-1}\binom{n}{j} t_{lr,k_1(q-j)}\lambda_r^{n-j} t^{rk,qk_2},$$

leads to

$$\mathrm{P}\left(\mid a_n^{lk} - 0 \mid \geq \frac{1}{m}\right) = \mathrm{P}\left(a_n^{lk} \geq \frac{1}{m}\right) \leq m\sum_{r=1}^{s}\sum_{q=1}^{n_r}\sum_{j=0}^{q-1}\binom{n}{j} t_{lr,k_1(q-j)}\lambda_r^{n-j} t^{rk,qk_2}.$$

We now prove that $\sum_{n=j}^{\infty} c_n$ converges, where $c_n = \sum_{r=1}^{s}\sum_{q=1}^{n_r}\sum_{j=0}^{q-1}\binom{n}{j} t_{lr,k_1(q-j)}\lambda_r^{n-j} t^{rk,qk_2}$.
We first notice that

$$\mid c_n \mid = \| c_n \| \leq \sum_{r=1}^{s}\sum_{q=1}^{n_r}\sum_{j=0}^{q-1}\binom{n}{j} \| t_{lr,k_1(q-j)} \| \| \lambda_r \|^{n-j} \| t^{rk,qk_2} \|,$$

where $\mid\ \mid$ and $\|\ \|$ are still the norms on $\mathbb{R}$ and $\mathbb{C}$.
If $d_n = \binom{n}{j} \| t_{lr,k_1(q-j)} \| \| \lambda_r \|^{n-j} \| t^{rk,qk_2} \|$, then

$$\begin{aligned}\frac{\mid d_{n+1} \mid}{\mid d_n \mid} &= \frac{\binom{n+1}{j} \| t_{lr,k_1(q-j)} \| \| \lambda_r \|^{n+1-j} \| t^{rk,qk_2} \|}{\binom{n}{j} \| t_{lr,k_1(q-j)} \| \| \lambda_r \|^{n-j} \| t^{rk,qk_2} \|} \\ &= \frac{n+1}{n+1-j} \| \lambda_r \| \longrightarrow \| \lambda_r \| < 1.\end{aligned}$$

Hence, from the Ratio Test,

$$\sum_{n=j}^{\infty}\binom{n}{j} \| t_{lr,k_1(q-j)} \| \| \lambda_r \|^{n-j} \| t^{rk,qk_2} \| < \infty,$$

which finally implies

$$\begin{aligned}\sum_{n=j}^{\infty} \mid c_n \mid = \sum_{n=j}^{\infty} \| c_n \| &\leq \sum_{n=j}^{\infty}\sum_{r=1}^{s}\sum_{q=1}^{n_r}\sum_{j=0}^{q-1}\binom{n}{j} \| t_{lr,k_1(q-j)} \| \| \lambda_r \|^{n-j} \| t^{rk,qk_2} \| \\ &= \sum_{r=1}^{s}\sum_{q=1}^{n_r}\sum_{j=0}^{q-1}\sum_{n=j}^{\infty}\binom{n}{j} \| t_{lr,k_1(q-j)} \| \| \lambda_r \|^{n-j} \| t^{rk,qk_2} \| < \infty.\end{aligned}$$

The rest of the proof remains identical to what comes right after eqs.(3.28).

F Solving eqs.(3.38)

For each z inside the convergence interval of $\delta(.)$, we have:

$$\frac{\alpha(z)}{1-\beta(z)} = \delta(z) \quad \Leftrightarrow \quad \alpha(z) = \delta(z) - \beta(z)\delta(z). \tag{F.1}$$

Let

$$h_t^* = \frac{\alpha_0}{1-\beta(1)} \left\{ 1 + \sum_{i=1}^{\infty} \left[\sum_{k_1=1}^{\infty} \dots \sum_{k_i=1}^{\infty} \delta_{k_1} \dots \delta_{k_i} \varepsilon_{t-k_1}^2 \dots \varepsilon_{t-k_1-\dots-k_i}^2 \right] \right\}, \quad t \in \mathbb{Z}. \tag{F.2}$$

Placing eq.(F.2) in the second equation for original expression for h_t (see eqs.(3.38)):

$$\begin{aligned}
&\alpha_0 + \sum_{l=1}^{p} \alpha_l Y_{t-l}^2 + \sum_{j=1}^{q} \beta_j h_{t-j} = \alpha_0 + \sum_{l=1}^{p} \alpha_l \varepsilon_{t-l}^2 h_{t-l} + \sum_{j=1}^{q} \beta_j h_{t-j} \\
=&\alpha_0 + \sum_{l=1}^{p} \alpha_l \varepsilon_{t-l}^2 \left(\frac{\alpha_0}{1-\beta(1)} \left\{ 1 + \sum_{i=1}^{\infty} \left[\sum_{k_1=1}^{\infty} \dots \sum_{k_i=1}^{\infty} \delta_{k_1} \dots \delta_{k_i} \varepsilon_{t-l-k_1}^2 \dots \varepsilon_{t-l-k_1-\dots-k_i}^2 \right] \right\} \right) \\
&+ \sum_{j=1}^{q} \beta_j \left(\frac{\alpha_0}{1-\beta(1)} \left\{ 1 + \sum_{i=1}^{\infty} \left[\sum_{k_1=1}^{\infty} \dots \sum_{k_i=1}^{\infty} \delta_{k_1} \dots \delta_{k_i} \varepsilon_{t-j-k_1}^2 \dots \varepsilon_{t-j-k_1-\dots-k_i}^2 \right] \right\} \right) \\
=&\alpha_0 + \frac{\alpha_0}{1-\beta(1)} \sum_{l=1}^{p} \alpha_l \varepsilon_{t-l}^2 \\
&+ \frac{\alpha_0}{1-\beta(1)} \sum_{l=1}^{p} \alpha_l \varepsilon_{t-l}^2 \left\{ \sum_{i=1}^{\infty} \left[\sum_{k_1=1}^{\infty} \dots \sum_{k_i=1}^{\infty} \delta_{k_1} \dots \delta_{k_i} \varepsilon_{t-l-k_1}^2 \dots \varepsilon_{t-l-k_1-\dots-k_i}^2 \right] \right\} \\
&+ \frac{\alpha_0}{1-\beta(1)} \sum_{j=1}^{q} \beta_j + \frac{\alpha_0}{1-\beta(1)} \sum_{j=1}^{q} \beta_j \left\{ \sum_{i=1}^{\infty} \left[\sum_{k_1=1}^{\infty} \dots \sum_{k_i=1}^{\infty} \delta_{k_1} \dots \delta_{k_i} \varepsilon_{t-j-k_1}^2 \dots \varepsilon_{t-j-k_1-\dots-k_i}^2 \right] \right\} \\
=&\alpha_0 + \frac{\alpha_0}{1-\beta(1)} \sum_{j=1}^{q} \beta_j + \frac{\alpha_0}{1-\beta(1)} \left\{ \sum_{l=1}^{p} \alpha_l \varepsilon_{t-l}^2 \right. \\
&+ \sum_{l=1}^{p} \alpha_l \varepsilon_{t-l}^2 \left(\sum_{i=1}^{\infty} \left[\sum_{k_1=1}^{\infty} \dots \sum_{k_i=1}^{\infty} \delta_{k_1} \dots \delta_{k_i} \varepsilon_{t-l-k_1}^2 \dots \varepsilon_{t-l-k_1-\dots-k_i}^2 \right] \right) \\
&\left. + \sum_{j=1}^{q} \beta_j \left(\sum_{i=1}^{\infty} \left[\sum_{k_1=1}^{\infty} \dots \sum_{k_i=1}^{\infty} \delta_{k_1} \dots \delta_{k_i} \varepsilon_{t-j-k_1}^2 \dots \varepsilon_{t-j-k_1-\dots-k_i}^2 \right] \right) \right\}
\end{aligned}$$

$$=\frac{\alpha_0}{1-\beta(1)}\left(1-\beta(1)+\sum_{j=1}^{q}\beta_j\right)+\frac{\alpha_0}{1-\beta(1)}\left\{\sum_{l=1}^{\infty}\delta_l\varepsilon_{t-l}^2-\sum_{j=1}^{q}\beta_j\sum_{l=1}^{\infty}\delta_l\varepsilon_{t-l-j}^2\right.$$
$$+\sum_{s=1}^{\infty}\delta_s\varepsilon_{t-s}^2\left(\sum_{i=1}^{\infty}\left[\sum_{k_1=1}^{\infty}\cdots\sum_{k_i=1}^{\infty}\delta_{k_1}\ldots\delta_{k_i}\varepsilon_{t-s-k_1}^2\ldots\varepsilon_{t-s-k_1-\ldots-k_i}^2\right]\right)$$
$$-\sum_{j=1}^{q}\beta_j\sum_{s=1}^{\infty}\delta_s\varepsilon_{t-j-s}^2\left(\sum_{i=1}^{\infty}\left[\sum_{k_1=1}^{\infty}\cdots\sum_{k_i=1}^{\infty}\delta_{k_1}\ldots\delta_{k_i}\varepsilon_{t-j-s-k_1}^2\ldots\varepsilon_{t-j-s-k_1-\ldots-k_i}^2\right]\right)$$
$$\left.+\sum_{j=1}^{q}\beta_j\left(\sum_{i=1}^{\infty}\left[\sum_{k_1=1}^{\infty}\cdots\sum_{k_i=1}^{\infty}\delta_{k_1}\ldots\delta_{k_i}\varepsilon_{t-j-k_1}^2\ldots\varepsilon_{t-j-k_1-\ldots-k_i}^2\right]\right)\right\}$$
$$=\frac{\alpha_0}{1-\beta(1)}$$
$$+\frac{\alpha_0}{1-\beta(1)}\left\{\sum_{l=1}^{\infty}\delta_l\varepsilon_{t-l}^2+\sum_{i=1}^{\infty}\left[\sum_{s=1}^{\infty}\sum_{k_1=1}^{\infty}\cdots\sum_{k_i=1}^{\infty}\delta_s\delta_{k_1}\ldots\delta_{k_i}\varepsilon_{t-s}^2\varepsilon_{t-s-k_1}^2\ldots\varepsilon_{t-s-k_1-\ldots-k_i}^2\right]\right.$$
$$-\sum_{j=1}^{q}\beta_j\sum_{l=1}^{\infty}\delta_l\varepsilon_{t-l-j}^2$$
$$-\sum_{j=1}^{q}\beta_j\left(\sum_{i=1}^{\infty}\left[\sum_{s=1}^{\infty}\sum_{k_1=1}^{\infty}\cdots\sum_{k_i=1}^{\infty}\delta_s\delta_{k_1}\ldots\delta_{k_i}\varepsilon_{t-s-j}^2\varepsilon_{t-s-j-k_1}^2\ldots\varepsilon_{t-s-j-k_1-\ldots-k_i}^2\right]\right)$$
$$\left.+\sum_{j=1}^{q}\beta_j\left(\sum_{i=1}^{\infty}\left[\sum_{k_1=1}^{\infty}\cdots\sum_{k_i=1}^{\infty}\delta_{k_1}\ldots\delta_{k_i}\varepsilon_{t-j-k_1}^2\ldots\varepsilon_{t-j-k_1-\ldots-k_i}^2\right]\right)\right\}=\frac{\alpha_0}{1-\beta(1)}$$
$$-\sum_{j=1}^{q}\beta_j\left(\sum_{i=1}^{\infty}\left[\sum_{k_1=1}^{\infty}\cdots\sum_{k_i=1}^{\infty}\delta_{k_1}\ldots\delta_{k_i}\varepsilon_{t-k_1}^2\ldots\varepsilon_{t-k_1-\ldots-k_i}^2\right]\right)$$
$$\left.+\sum_{j=1}^{q}\beta_j\left(\sum_{i=1}^{\infty}\left[\sum_{k_1=1}^{\infty}\cdots\sum_{k_i=1}^{\infty}\delta_{k_1}\ldots\delta_{k_i}\varepsilon_{t-j-k_1}^2\ldots\varepsilon_{t-j-k_1-\ldots-k_i}^2\right]\right)\right\}$$
$$=\frac{\alpha_0}{1-\beta(1)}+\frac{\alpha_0}{1-\beta(1)}\left\{\sum_{i=1}^{\infty}\left[\sum_{k_1=1}^{\infty}\cdots\sum_{k_i=1}^{\infty}\delta_{k_1}\ldots\delta_{k_i}\varepsilon_{t-k_1}^2\ldots\varepsilon_{t-k_1-\ldots-k_i}^2\right]\right\}$$
$$=\frac{\alpha_0}{1-\beta(1)}\left\{1+\sum_{i=1}^{\infty}\left[\sum_{k_1=1}^{\infty}\cdots\sum_{k_i=1}^{\infty}\delta_{k_1}\ldots\delta_{k_i}\varepsilon_{t-k_1}^2\ldots\varepsilon_{t-k_1-\ldots-k_i}^2\right]\right\}=h_t^*;$$

the fourth equality follows from the equivalence in (F.1).

References

[1] A. Aue, I. Berkes, I., L. Horávth, *Strong Approximation for the Sums of Squares of Augmented GARCH Sequences*, Bernoulli, **12**(4) (2006), 583–608. MR2248229. Zbl 1125.62092. DOI: https://doi.org/10.3150/bj/1155735928.

[2] R. G. Bartle, D. R. Sherbert, *Introduction to real analysis*, 3rd edition, New York, Wiley, 2000. MR1135107. Zbl 0810.26001.

[3] P. Billingsley, *Probability and Measure*, 3rd edition, New York, Wiley, 1995. MR1324786. Zbl 0822.60002.

[4] T. Bollerslev, *Generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics, **31** (1986), 307–327. MR4570274. Zbl 1236.60001. DOI: https://doi.org/10.1016/0304-4076(86)90063-1.

[5] T. Bollerslev, J. Wooldridge. *Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances*, Econometric Reviews, **11**(2) (1992), 143–172. MR1185178. Zbl 0850.62884. DOI: https://doi.org/10.1080/07474939208800229.

[6] P. Bougerol, N. Picard, *Stationarity of GARCH processes and of some nonnegative time series*, Journal of Econometrics, **52** (1992), 115–127. MR1165646. Zbl 0746.62087. DOI: https://doi.org/10.1016/0304-4076(92)90067-2.

[7] G. E. Box, G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Revised edition, Oakland, Holden-Day, 1976. MR436499. Zbl 0363.62069.

[8] G. E. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th edition, Hoboken, Wiley, 2016. Zbl 1317.62001.

[9] P. J. Brockwell, R. A. Davis, *Time Series: Theory and Methods*, 2nd edition, New York, Springer, 1991. MR1093459. Zbl 0709.62080.

[10] P. J. Brockwell, R. A. Davis, *Introduction to Time Series and Forecasting*, 3rd edition, New York, Springer, 2016. MR13526245. Zbl 1355.62001.

[11] J. Y. Campbell, A. W. Lo, A. C. Mackinlay, *The Econometrics of Financial Markets*, Mercer County, Princeton University Press, 1996. Zbl 0927.62113.

[12] G. Casella, R. Berger, *Statistical Inference*, 2nd edition, Duxbury, Duxbury Advanced Series, 2002. MR1051420. Zbl 1554.62002.

[13] C. Chatfield, H. Xing, *The Analysis of Time Series. An introduction with R*, 7th edition, New York, Chapman & Hall, 2019. MR4908498. Zbl 1435.62003.

[14] K. Chung, *A Course in Probability Theory*, 3rd edition, San Diego, Academic Press, 2001. MR1796326. Zbl 0980.60001.

[15] R. Davidson, J. G. MacKinnon, *Econometric Theory and Methods*, New York, Oxford University Press, 2004.

[16] P. Doukhan, M. H. Neumann, *Absolute regularity of semi-contractive GARCH-type processes*, Journal of Applied Probability, **56** (2019), 91–115. MR3981148. Zbl 1418.60029. DOI: https://doi.org/10.1017/jpr.2019.8.

[17] W. Enders, *Applied Econometric Time Series*, 4th edition, Hoboken, John Wiley & Sons, 2014.

[18] R. Engle, *Autoregressive Conditional Heroscedasticity with Estimates of the Variance of United Kingdom Inflation*, Econometrica, **50**(4) (1982), 987–1007. MR666121. Zbl 0491.62099. DOI: https://doi.org/10.2307/1912773.

[19] R. B. Engle, T. Bollerslev, *Modelling the persistence of conditional variances*, Econometric Reviews, **5**(1) (1986), 1–50. MR876792. Zbl 0619.62105. DOI: https://doi.org/10.1080/07474938608800095.

[20] W. A. Fuller, *Introduction to Statistical Time Series*, 2nd edition, Hoboken, Wiley, 1996. MR1365746. Zbl 0851.62057.

[21] C. W. J. Granger, P. Newbold, *Forecasting Economic Time Series*, 2nd edition, London, Academic Press, 1986. Zbl 0642.90001.

[22] W. H. Greene, *Econometric Analysis*, 8th edition, Zamin Pallavaram, Pearson India, 2018.

[23] D. N. Gujarati, D. C. Porter, S. Gunasekar, *Basic Econometrics*, 5th edition, Noida, McGraw-Hill India, 2011.

[24] J. D. Hamilton, *Time Series Analysis*, Princeton, Princeton University Press, 2014. MR1278033. Zbl 0831.62061.

[25] A. C. Harvey,*Time Series Models*. 2nd edition, Hemel Hempstead, Harvester Wheatsheaf, 1993. Zbl 0770.62076.

[26] F. Hayashi, *Econometrics*, Mercer County, Princeton University Press, 2000. MR1881537. Zbl 0994.62107.

[27] D. F. Hendry, *Dynamic Econometrics*, Oxford, Oxford University Press, 1995. MR1351008. Zbl 0897.62127.

[28] J. Johnston, J. E. DiNardo, *Econometric methods*, 4th edition, New York, McGraw-Hill, 1997.

[29] C. S. Kubrusly, *Measure Theory: A First Course*, Boston, Academic Press, 2007. Zbl 1217.28003.

[30] S. Ling, M. McAller, *Stationarity and the existence of moments of a family of GARCH processes*, Journal of Econometrics, **106** (2002), 109–117. MR1875529. Zbl 1043.62073. DOI: https://doi.org/10.1016/S0304-4076(01)00090-2.

[31] A. Milhoj, *The Moment Structure of ARCH Processes*, Scandinavian Journal of Statistics, **12**(4) (1985), 281–292. MR841449. Zbl 0595.62089.

[32] T. C. Mills, *The Econometric Modelling of Financial Time Series*, Cambridge University Press, 1993. Zbl 0861.62078.

[33] T. C. Mills, *Applied Time Series Analysis*, London, Academic Press, 2019.

[34] D. C. Montgomery, C. L. Jennings, M. Kulahci, *Introduction to Time Series Analysis and Forecasting*, 2nd edition, Hoboken, Wiley, 2015. Zbl 1348.91001

[35] D. B. Nelson, *Stationary and persistence in the GARCH(1,1) model*, Econometric Theory, **6** (1990), 318–334.

[36] D. B. Nelson, *Conditional heteroskedasticity in asset returns: a new approach*, Econometrica, **59**(2) (1991), 347–370. MR1097532. Zbl 0722.62069. DOI: https://doi.org/10.2307/2938260.

[37] K. Neusser, *Time Series Econometrics*, New York, Springer, 2016. Zbl 1341.62016.

[38] V. Rohatgi, A. Saleh, *An Introduction to Probability and Statistics*, 2nd edition, New York, John Wiley & Sons, 2001. MR1789794. Zbl 0969.62002.

[39] A. N. Shiryaev, *Probability*, 2nd edition, New York, Springer, 1995. Zbl 0835.60002.

[40] R. H. Shumway, D. S. Stoffer, *Time Series Analysis and Its Applications (With R Examples)*, 4th edition, New York, Springer, 2017. Zbl 1367.62004.

[41] R. S. Tsay, *Analysis of Financial Time Series*, 3rd edition, Hoboken, Wiley, 2010. MR2778591. Zbl 1209.91004.

[42] A. A. Weiss, *ARMA models with ARCH errors*, Journal of Time Series Analysis, **5**(2) (1984), 129–143. Zbl 0549.62079. DOI: https://doi.org/10.1111/j.1467-9892.1984.tb00382.x.

[43] A. A. Weiss, *Asymptotic Theory for ARCH Models: Estimation and Testing*, Econometric Theory, **2**(1) (1986), 107–131.

[44] J. Wooldridge, *Introductory Econometrics: A Modern Approach*, 8th edition, Boston, Cengage Learning, 2025. MR2778591. Zbl 1209.91004.

[45] P. Zaffaroni, *Stationarity and Memory of ARCH(∞) Models*, Econometric Theory, **2**0(1) (2004), 147–160. MR2028355. Zbl 1046.62096.

Jorge Bustamante, ORCID: 0009-0007-7916-9712
Department of Economics
Pontifical Catholic University of Rio de Janeiro, Brazil.
e-mail: jroqueb95@gmail.com

Adrian Pizzinga - Corresponding author, ORCID: 0000-0002-3059-743X
Institute of Mathematics and Statistics
Fluminense Federal University, Brazil.
e-mail: adrianhp@id.uff.br

License

Received: December 27, 2025; Accepted: April 16, 2026;
Published electronically: April 17, 2026.