CONTEMPORANEOUS AGGREGATION OF *GARCH* PROCESSES

by

Paolo Zaffaroni*
Banca d'Italia

Contents:

The Suntory Centre
Suntory and Toyota International Centres for Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel.: 0171-405 7686

Discussion Paper
No. EM/00/378
January 2000

* I thank Karim Abadir, Filippo Altissimo and Domenico Marinucci for insightful discussion and suggestions.

Abstract

We study the impact of large cross-sections of contemporaneous aggregation of GARCH processes and of dynamic GARCH factor models. The results crucially depend on the shape of the cross-sectional distribution of the GARCH coefficients and on the cross-sectional dependence properties of the rescaled innovation. The aggregate maintains the core nonlinearity of a volatility model, uncorrelation in the levels but autocorrelation in the squares, when the rescaled innovation is common across units. The nonlinearity is, however, lost at the aggregate level, when the rescaled innovation is orthogonal across units. This is not a consequence of the usual result of a vanishing importance of purely idiosyncratic risk as, under appropriate conditions, this is simply not fully diversifiable in arbitrary large portfolios. Non-GARCH memory properties arise at the aggregate level. Strict stationarity, ergodicity and finite kurtosis might fail for the aggregate despite the micro GARCH do satisfy these properties. Under no conditions aggregation of GARCH induces long memory conditional heteroskedasticity.

Keywords: Contemporaneous aggregation; GARCH; conditionally heteroskedastic factor models; common and idiosyncratic risk; nonlinearity; memory.

JEL Nos.: C32, C43.

1 Introduction

The ARCH model of Engle (1982) and the GARCH development (Bollerslev 1986) represent the most popular approaches, within the class of nonlinear models, used to describe the conditional heteroskedasticity observed in many financial time series. The chief motivation underlying this success is represented by their excellent widespread performance in fitting the empirical distribution of financial asset returns, synthesized in a number of well known stylized facts; see e.g. Bollerlsev, Chou, and Kroner (1992) and Bollerslev, Engle, and Nelson (1995).

More recent interest in ARCH and GARCH processes has focused on the impact of temporal aggregation (Drost and Nijman 1993) (Drost and Werker 1996) (Corradi 1999), refining the original work of Diebold (1988) and Nelson (1990a). This work was motivated by the need, on one hand, of bridging the discrete time specification of these models with continuous time pricing formulae of modern finance and, on the other, of obtaining coherent volatility forecasts for different time horizons, relevant for practical risk management; see e.g. Christoffersen and Diebold (1997).

Looking at large asset markets naturally prompts the analysis of contemporaneous aggregation, in the sense of cross-sectional (arithmetic) averaging.

A recent strand of empirical research shows that the effect of shocks to the conditional variance of asset returns is very persistent but is eventually absorbed as time passes (see e.g. Gallant, Rossi, and Tauchen (1993)), consistent with the notion of long memory in volatilities (Ding, Granger, and Engle 1993) but not with IGARCH type behaviour. Several volatility models have been proposed to account for this recent stylized fact of asset return dynamics (see e.g. among others Robinson (1991), Baillie, Bollerslev, and Mikkelsen (1996) and Robinson and Zaffaroni (1997)) but its foundation has not been well understood yet.

The impact of contemporaneous aggregation (henceforth aggregation) of linear (ARMA) processes is well known, in particular as a possible source of long memory in observed macroeconomic time series, as noted in Robinson (1978) and further developed in Granger (1980) among others. A systematic analysis is in Lippi and Zaffaroni (1999). The available results valid for ARMA are not readily applicable to a nonlinear framework except when linear models represent a good approximation. Unfortunately this does not apply to GARCH, their main feature being that, unlike linear processes, they exhibit correlated squares and uncorrelated levels. Indeed, in order to employ the linear aggregation results in a nonlinear framework, Ding and Granger (1996) have to consider a suitable parameterization of the aggregate. Under certain parametric distributional assumptions on the GARCH(1,1) coefficients, they shown that long memory might arise for a large number of units. The linear aggregation framework is also used in Andersen and Bollerlsev (1997) in a Stochastic Volatility (SV) setting through the so-called mixture of of distribution hypothesis.

More generally, when the dynamics of single stocks is well described by GARCH behaviour, uncovering the induced statistical properties of a large portfolio is clearly relevant both from a theoretical and practical perspective, e.g. in order to better understand and thus estimate and forecast volatility dynamics of stock return indexes. The GARCH paradigm applies to finite portfolios in the sense that the class of weak GARCH, viz. replacing conditional expectations with linear projections (Drost and Nijman 1993), is closed under aggregation of a small number of units (Nijman and Sentana 1996). Even stronger results apply within a SV framework, as the (finite) portfolio belongs precisely to the same class of the micro SV models (Meddahi and Renault 1998).

Finally, since the development of the CAPM and the APT, common shocks, represented respectively by the 'market' portfolio return or simply by latent common factors, have played a key role in asset pricing theory when facing a large number of assets but there is little doubt that idiosyncratic shocks are an important determinant of assets dynamics. Suitable assumptions, typically expressed by uniform boundedness of the eigenvalues of the idiosyncratic variance-covariance matrix since Chamberlain and Rothschild (1983), however allow to neglect idiosyncratic shocks in large portfolios. It is then crucial to understand the conditions that insure portfolio full diversification of idiosyncratic-driven risk when each asset is modelled by GARCH.

In this paper we study the impact of aggregation of heterogeneous strong GARCH processes when the number of units gets large. Shape and degree of heterogeneity across units' parameters is described by means of a semiparametric specification of the distribution over the GARCH coefficients. Both the case of a homogeneous and heterogeneous, across units, rescaled innovation is considered.

Consider, for sake of exposition, the case of heterogeneous ARCH(1) processes, when both the parameters and the rescaled innovations are potentially varying across units

$$x_{i,t} = z_{i,t}\, \sigma_{i,t}, \tag{1}$$

with

$$\sigma_{i,t}^2 = \omega_i + \alpha_i x_{i,t-1}^2. \tag{2}$$

where we assume that the α_i and the ω_i are i.i.d drawn from a distribution with support $[0, \gamma)$, for a real $0 < \gamma \leq 1$, and $(0, \infty)$ respectively. The $z_{i,t}$ form an i.i.d. $(0, 1)$ sequence which will be, in turn, assumed to be either common ($z_{i,t} = z_t$ say) or mutually orthogonal across units. Given n units, the aggregate process is defined as $1/n \sum_{i=1}^{n} x_{i,t}$. It turns out that the shape near γ of the cross-sectional distribution of the α_i fully characterizes the probabilistic properties of the aggregate as n gets large. A difficulty of the analysis, with respect to the ARMA case, is that care need to be taken due to the GARCH model dichotomy, i.e. the very different statistical properties of the model in the levels and in the squares, respectively. For example, the degree of unbounded kurtosis of the micro processes, dictated by the value of γ plays a crucial role, unlike within the linear framework.

Noteworthy results are that when the distribution of the α_i is sufficiently dense around 1, the aggregate will not vanish even when the $z_{i,t}$ are perfectly orthogonal across units. In fact, as n gets large the aggregate will display the same degree of nonstationarity independently from the nature of the rescaled innovation, i.e. whether common or idiosyncratic across units. On the other hand, the type of cross-sectional dependence of the rescaled innovation has dramatic impact on the degree of nonlinearity of the aggregate. In the purely idiosyncratic case the 'GARCH' nonlinearity is completely washed out whereas is maintained in the common case. Recalling that $\alpha_i < 1/\sqrt{3}$ represents the bounded kurtosis condition for ARCH(1), we show that for a dense enough distribution of the α_i around $1/\sqrt{3}$ the squared aggregate displays unbounded kurtosis even with covariance stationary (hereafter, stationary) squared micro processes. More importantly, allowing $\gamma > 1$, the aggregate might not be strictly stationary and ergodic, unlike the micro GARCH.

Aggregation of heterogeneous GARCH induce memory properties of the aggregate different from the ones of the underlying micro processes. In contrast to a common view, we show that it is not possible to generate long memory conditional heteroskedasticity at the aggregate level. More generally, the aggregate will not belong to the class of weak (and strong) GARCH, in contrast to the small number of units case (Nijman and Sentana 1996). This implies that if GARCH models are fitted to returns on single stocks, a different volatility model is required for a large portfolio, whose statistical properties are here characterized.

The plan of the paper is as follows. In section 2 we focus on the ARCH(1) case. Definitions and assumptions are introduced in section 2.1 and section 2.2 and 2.3 focuses on the case of orthogonal and common rescaled innovation, respectively. Section 3 generalizes the framework discussing GARCH, dynamic GARCH factor models and other extensions. All the results are formally stated in Theorems whose proofs are reported in Appendix A. Section 4 discusses the aggregation implications and possible misspecification from exploiting the ARMA representation of squared GARCH. Section 5 concludes.

2 Aggregation of ARCH(1)

In this section we focus on ARCH(1) micro heterogeneous units as specified in (1) and (2).

2.1 Definitions and assumptions

In order to characterize a framework made by a large number n of units, whose aggregate

$$X_{n,t} = \frac{1}{n}\sum_{i=1}^{n} x_{i,t}, \tag{3}$$

we specify heterogeneity across units by means of a distribution function for the time-invariant ARCH(1) coefficients ω_i, α_i. Time-varying heterogeneity is also allowed, through the rescaled innovation.

Henceforth $\sim$ denotes asymptotic equivalence, i.e. $a(x) \sim b(x)$, as $x \to x_0$, when $a(x)/b(x) \to 1$, and c, C bounded in modulus constants, writing c_θ, C_θ when depending on a parameter θ. We denote by S_δ $(0 < \delta \leq 2)$ a δ-stable random variable with zero location parameter; see e.g. Samorodnitsky and Taqqu (1994)[1].

[1]Using Samorodnitsky and Taqqu (1994) notation, S_δ refers to $S_\delta(\sigma, \beta, 0)$ for real parameters $\sigma \geq 0$ (scale parameter) and $-1 \leq \beta \leq 1$ (skewness parameter). We leave the values for σ, β unspecified although, when $0 < \delta < 1$ we will obtain $\beta = 1$, i.e. a totally skewed to the right δ-stable random variable

Given a real γ with $0 < \gamma \leq 1$, we assume that the ARCH coefficients satisfy the following:

Assumption I (γ)

(i) The ω_j *and the* α_i *are mutually independent for any* i, j.

(ii) The ω_i *are i.i.d. with* $E(\omega_i^2) < \infty$ *and* $0 < c_\omega \leq \omega_i < \infty$ *for any* i *and some constant* c_ω.

(iii) The α_i *have an absolutely continuous distribution in the interval* $[0, \gamma)$, *depending upon the real parameter* b_γ $(b_\gamma > -1)$, *whose density, denoted by* $B(\cdot; b_\gamma)$ *behaves, as* $\alpha_i \to \gamma^-$,

$$B(\alpha_i; b_\gamma) \sim C_{b_\gamma}(\gamma - \alpha_i)^{b_\gamma}, \tag{4}$$

with $C_{b_\gamma} > 0$.

Remark I.1 Assumption $I\,(\gamma)\,(iii)$ represents a mild semiparametric specification of the density function of the α_i, only imposing its behaviour in a neighbourhood of γ. The constraint $b_\gamma > -1$ is the obvious integrability condition. Indeed, it can be alternatively expressed as

$$B(\alpha_i; b_\gamma) = B(\alpha_i)(\gamma - \alpha_i)^{b_\gamma},$$

for any integrable function $B(\cdot)$ defined on $[0, \gamma]$ with $B(\alpha) \sim C$, as $\alpha \to \gamma^-$, with $0 < C < \infty$. A wide variety of parametric specifications $B(\cdot; \theta)$ for $\theta \in \Theta \subset R^p$ is allowed, e.g. as the Beta density.

Remark I.2 A great deal of generality is enhanced replacing (4) with

$$B(\alpha_i; b_\gamma) \sim C_{b_\gamma}(\gamma - \alpha_i)^{b_\gamma} L(1/[\gamma - \alpha_i]), \tag{5}$$

where $L(\cdot)$ denotes a slowly varying function (Zygmund 1977), i.e. for any $x > 0$, $L(tx)/L(t) \to 1$, when $t \to \infty$ (e.g. $L(\cdot) = [log(\cdot)]^\delta$ for any real δ). Focusing on density functions with a pole in γ, the hyperbolic term and the slowly varying function describe all possible behaviours (except for pathological cases). In fact, the integrability constraint rules out the possibility of having a pole at γ with a rate faster than -1. The gap between γ and -1 is naturally filled by slowly varying functions, given that $L(x)x^\delta \to 0$, as $x \to \infty$, for any $L(\cdot)$ and $\delta < 0$ (Yong 1974, Lemma I-11). Qualitatively, considering (5) has no consequence on the results, besides that $L(\cdot)$ will explicitly appear. For simplicity's sake, we will assume throughout the paper that $L(\cdot) = 1$.

Remark I.3 The independence assumption (i) has virtually the same impact of assuming that the ω_i are constant across units. The effect of relaxing such assumption will be discussed in section 3.
Remark I.4 Nelson (1990b) shows that the probabilistic properties of GARCH(1,1) crucially depend on whether (for an arbitrary unit i) ω_i is greater or equal to zero. Assuming a strictly positive c_ω rules out the possibility that, for some i, $\sigma^2_{i,t} \to 0$ almost surely as $t \to \infty$.
Remark I.5 Only covariance stationary ARCH(1) units are considered. It is well known that this implies a considerable restriction on the admissible values for the α_i consistent with strictly stationary ARCH(1) defined by the well-known condition $\alpha_i \, e^{Elog(z^2_{i,t})} < 1$ (Nelson 1990b, Theorem 2) which reduces to $\alpha_i < 2^E$ ($E = 0.577...$ is the Euler constant) for Gaussian $z_{i,t}$. Although our framework can be fairly easily extended to this case, we will only briefly comment in section 3 on the possibility of $\alpha_i \geq 1$ as, in general, this induces explosive behaviours of the aggregate, hiding relevant implications of the aggregation mechanism. The only exception will be made in Theorem 6 were we allow $\gamma > 1$ precisely in order to evaluate the impact of aggregation in terms of the strict stationarity and ergodicity properties.

The impact of aggregation crucially depends on the type of cross-sectional dependence of the rescaled innovation $z_{i,t}$. We will consider two cases, when the $z_{i,t}$ are perfectly orthogonal across units, expressing a source of time-varying heterogeneity, and, in contrast, when they are perfectly correlated across units. For sake of clarity, we will write $z_{i,t} = \epsilon_{i,t}$ and $z_{i,t} = u_t$ respectively in the two cases. The $\epsilon_{i,t}$ and the u_t, called respectively the idiosyncratic and the common shock, satisfy:

Assumption II
(i) The u_t are i.i.d. across t and the $\epsilon_{i,t}$ are i.i.d. across t, i, satisfying $E(u_t) = E(\epsilon_{i,t}) = 0$, $var(u_t) = var(\epsilon_{i,t}) = 1$ and zero fourth-order cumulant.
(ii) The $\{u_t, \, \epsilon_{i,t}\}$ and the $\{\omega_i, \, \alpha_i\}$ mutually independent.
Remark II.1 Both the u_t and the $\epsilon_{i,t}$ behave like standard Gaussian noise up to the fourth moment. The normalizations are made for simplicity's sake but can be easily relaxed as in Nelson (1990b). Setting $\delta_u = E \, |u_t|$ and $\lambda_u = Elog(u_t^2)$, it follows that $\delta_u < (E(u_t^2))^{1/2} = 1$ and $\lambda_u < logE(u_t^2) = 0$.
Remark II.2 Assumption II, in particular the i.i.d. assumption, implies that the micro processes are strong GARCH (Drost and Nijman 1993, Defi-

nition 1).

Hereafter, we will denote the conditional expectation and conditional variance operators, given the GARCH coefficients, by $E_n(\cdot)$ and $\text{var}_n(\cdot)$ respectively. Moreover we will always assume that Assumption II holds without stating this explicitly.

We will denote, for clarity's sake, the aggregate (3) by ${}^E X_{n,t}$ in the idiosyncratic case ($z_{i,t} = \epsilon_{i,t}$) and by ${}^U X_{n,t}$ in the common case ($z_{i,t} = u_t$). Note that no distinction needs to be made between stock and flow variables, unlike the temporal aggregation case (Drost and Nijman 1993).

2.2 Idiosyncratic innovations

In this case ${}^E X_{n,t}$ is given by a sum of purely idiosyncratic components as

$$ {}^E X_{n,t} = \frac{1}{n} \sum_{i=1}^{n} \epsilon_{i,t} \sigma_{i,t}, $$

with $\sigma_{i,t}^2$ given in (2). Simple calculation yields under Assumption II

$$ \text{var}_n({}^E X_{n,t}) = \frac{1}{n^2} \sum_{i=1}^{n} \frac{\omega_i}{1 - \alpha_i}. $$

Theorem 1 *As $n \to \infty$, under Assumption $I(\gamma)$:*
When $\gamma < 1$, uniformly in b_γ,

$$ \text{var}_n({}^E X_{n,t}) \to 0, \quad a.s. $$

When $\gamma = 1$
(i) If $b_1 > 0$, there exists a positive constant C such that a.s.

$$ \text{var}_n({}^E X_{n,t}) \sim C n^{-1}. $$

(ii) If $b_1 = 0$, there exist a positive constant C such that a.s.

$$ \text{var}_n({}^E X_{n,t}) \sim C n^{-1} \log n. $$

(iii) If $b_1 < 0$, setting $\delta = b_1 + 1$, a.s.

$$ \text{var}_n({}^E X_{n,t}) \sim n^{-\frac{2b_1+1}{b_1+1}} S_\delta, $$

with $S_\delta > 0 \, (0 < \delta < 1)$ a.s., where a.s. indicates almost surely.

Proof: See Appendix A.

Let us focus on the case $\gamma = 1$. When $b_1 > -1/2$, Theorem 1 gives the usual result that ${}^{E}X_{n,t}$ converges to zero in mean-square as n tends to infinity. However, for $b_1 < -1/2$ we obtain the rather striking result that the variance of ${}^{E}X_{n,t}$ tends to infinity at rate $n^{-(2b_1+1)/(b_1+1)}$ a.s., violating the usual result on the vanishing importance of idiosyncratic risk at the aggregate level.

When $\alpha_i = \alpha$ for any i, that is when one allows only for time-varying heterogeneity through the $\epsilon_{i,t}$, it easily follows that ${}^{E}X_{n,t} \to 0$ in mean-square for $\gamma < 1$. For $\gamma = 1$, $\mathrm{var}_n({}^{E}X_{n,t})$ is unbounded.

Obviously the ${}^{E}X_{n,t}$ are martingale differences for any n. On the other hand, the 'GARCH' nonlinearity arises once one considers the squared aggregate process ${}^{E}Y_{n,t}$:

$$ {}^{E}Y_{n,t} = \frac{1}{n^2} \sum_{i,j=1}^{n} \epsilon_{i,t}\epsilon_{j,t}\sigma_{i,t}\sigma_{j,t}. $$

From

$$ \mathrm{var}_n({}^{E}Y_{n,t}) = \frac{1}{n^4} \left(\sum_{i=1}^{n} \mathrm{var}_n(\epsilon_{i,t}^2\,\sigma_{i,t}^2) + 2 \sum_{\substack{i,j=1 \\ i \neq j}}^{n} E_n(\epsilon_{i,t}^2\,\sigma_{i,t}^2) E_n(\epsilon_{j,t}^2\,\sigma_{j,t}^2) \right), \quad (6) $$

it follows that the dominating expression of $\mathrm{var}_n({}^{E}Y_{n,t})$ is

$$ E_n(\sigma_{i,t}^4) = \frac{\omega_i^2}{(1 - 3\alpha_i^2)} \frac{(1 + \alpha_i)}{(1 - \alpha_i)}. \quad (7) $$

As originally shown by Engle (1982), a bounded fourth moment requires $\alpha_i^2 < 1/3$ for any $i = 1, \ldots, n$

However, this implies that looking at the correlation structure for ${}^{E}Y_{n,t}$ as $n \to \infty$ becomes irrelevant in light of Theorem 1. When the α_i are bounded below $3^{-1/2}$, the ${}^{E}X_{n,t}$ converge to zero in mean-square for any value of $b_{\sqrt{3}^{-1}}$ and thus the ${}^{E}Y_{n,t}$ converge to zero in probability by Slutsky's theorem.

Theorem 2 *The stationarity condition for ${}^{E}Y_{n,t}$ (cf. Assumption I $(\sqrt{3}^{-1})$) implies that as $n \to \infty$*

$$ {}^{E}Y_{n,t} \to_p 0, $$

uniformly in $b_{\sqrt{3}^{-1}}$.

To allow for micro processes with unbounded kurtosis, we investigate the behaviour of the aggregate of the 'truncated' processes, viz. conditioning on the rescaled innovations defined prior to time zero, $\epsilon_{i,s}$ $(s \leq 0)$,

$$\tilde{x}_{i,t} = \epsilon_{i,t}\tilde{\sigma}_{i,t},$$

with

$$\tilde{\sigma}^2_{i,t} = \omega_i \left(\sum_{k=0}^{t-1} \alpha_i^k \prod_{j=1}^{k} z^2_{i,t-j} \right), \tag{8}$$

recalling that within this section $z_{i,t} = \epsilon_{i,t}$. Let ${}^E\tilde{X}_{n,t} = 1/n \sum_{i=1}^n \tilde{x}_{i,t}$. The conditional ARCH(1) based on (8) is equivalent to the conditional model as defined in Nelson (1990b, eq. (6)) with the initial distribution of $\tilde{\sigma}^2_{i,0}$ being a Dirac mass at zero. Hereafter $\rightarrow_d$ denotes convergence in distribution.

Theorem 3 *As $n \rightarrow \infty$, under Assumption $I(\gamma)$:*
(i) When $\gamma < 1$, uniformly in b_γ, or $\gamma = 1$, with $b_1 > 0$,

$$\sqrt{n}\ {}^E X_{n,t} \rightarrow_d {}^E X_t,$$

where $\{{}^E X_t\}$ is a stationary Gaussian noise N(0, V) with $V = E(\omega/(1-\alpha))$.
(ii) When $\gamma = 1$, with $b_1 < 0$,

$$\sqrt{n}\,{}^E\tilde{X}_{n,t} \rightarrow_d {}^E\tilde{X}_t,$$

where $\{{}^E\tilde{X}_t\}$ is a nonstationary Gaussian noise $N(0, V_t)$ with $V_t \sim c\,t^{-b_1}$ as $t \rightarrow \infty$ for some $0 < c < \infty$.

Proof: See Appendix A.

When the micro GARCH processes are mutually orthogonal the (normalized) aggregate has a normal asymptotic distribution, both in the stationary and in the nonstationary case. This has the strong implication that the 'GARCH' nonlinearity is completely washed out by aggregation. In fact, in both cases the limit process is made by independent (and identically distributed in the stationary case) random variables (r.v.'s). Moreover, in the nonstationary case, weak convergence of the normalized aggregate ${}^E\tilde{X}_{n,t}/\sqrt{var_n({}^E\tilde{X}_{n,t})}$ does not follow. In fact, the distribution of ${}^E\tilde{X}_t - {}^E\tilde{X}_s$ and of ${}^E\tilde{X}_t + {}^E\tilde{X}_s$ coincide for any $t \neq s$, and therefore the erraticness of

${}^{E}\tilde{X}_t - {}^{E}\tilde{X}_s$ cannot controlled for by choosing $|t-s|$ small enough. Formally, a necessary condition for tightness in D, the space of functions defined over [0,1] that are right-continuous and have left-hand limit (Billingsley 1968, pg. 109), does not hold (Billingsley 1968, condition (ii) Theorem 15.2).

Hence weak convergence to some Gaussian probability measure does not follow, although the limit process $\{{}^{E}\tilde{X}_t\}$ is self-similar with index $-b_1$ (Samorodnitsky and Taqqu 1994, Definition 7.1.1). Thus, from Theorem 1 and 3, it follows that the nonlinearity vanishes at the aggregate level, when facing purely idiosyncratic risk, albeit this is not due to the asymptotic degeneracy of aggregate dynamics.

2.3 Common innovations

In this section the aggregate ${}^{U}X_{n,t}$ is not made by a sum of purely idiosyncratic components anymore,

$$ {}^{U}X_{n,t} = \frac{u_t}{n}\sum_{i=1}^{n}\sigma_{i,t}. $$

In fact all the $\sigma_{i,t}$, $i = 1, \ldots, n$ contain the same set of realizations u_s, $s < t$. Indeed, due to the dependence between $\sigma_{i,t}$ and $\sigma_{j,t}$,

$$ \mathrm{var}_n({}^{U}X_{n,t}) = \frac{1}{n^2}\sum_{i,j=1}^{n} E_n(\sigma_{i,t}\sigma_{j,t}). \tag{9} $$

Exploiting the fact that the rescaled innovations are equal across units, a simple way to evaluate the possibility of asymptotic degeneracy of ${}^{U}X_{n,t}$ is based on the absolute first moment:

$$ E_n \mid {}^{U}X_{n,t} \mid = \delta_u \frac{1}{n}\sum_{i=1}^{n} E_n(\sigma_{i,t}). \tag{10} $$

Bounding $E_n(\sigma_{i,t})$ from below and from above as in Nelson (1990b, Theorem 3) yields

$$ \frac{\delta_u}{n}\sum_{i=1}^{n}\frac{\omega_i^{1/2}}{(1-\alpha_i\delta_u^2)^{1/2}} \leq E_n \mid {}^{U}X_{n,t} \mid \leq \frac{\delta_u}{n}\sum_{i=1}^{n}\frac{\omega_i^{1/2}}{(1-\alpha_i^{1/2}\delta_u)}. $$

Under Assumption $I(1)$ $\alpha_i \delta_u^2 < \delta_u^2 < 1$ implying that asymptotically the absolute first moment is bounded and bounded away from zero and thus ${}^U X_{n,t}$ never converges to zero in L_1.

This does not imply that ${}^U X_{n,t}$ is stationary asymptotically but, by Jensen's inequality, that the conditional variance of ${}^U X_{n,t}$ is always bounded away from zero. As in the idiosyncratic case, the stationarity conditions for ${}^U X_{n,t}$ follow from the behaviour of the conditional variance.

Theorem 4 *As $n \to \infty$, under Assumption $I(\gamma)$:*
When $\gamma < 1$, uniformly in b_γ, for some constant C,

$$\mathrm{var}_n({}^U X_{n,t}) \to C, \;\; a.s., \; 0 < \; C \; < \infty.$$

When $\gamma = 1$
(i) If $b_1 > -1/2$,

$$\mathrm{var}_n({}^U X_{n,t}) \sim C.$$

(ii) If $b_1 = -1/2$, there exist positive constants c, C such that a.s.

$$c(log\, n) \leq \mathrm{var}_n({}^U X_{n,t}) \leq C\,(log\, n)^2.$$

(iii) If $b_1 < -1/2$, setting $\delta = -(b_1+1)/b_1$, a.s.

$$\mathrm{var}_n({}^U X_{n,t}) \sim n^{-\frac{2b_1+1}{b_1+1}} S_\delta,$$

with $S_\delta > 0\,(0 < \delta < 1)$ a.s.

Proof: See Appendix A.

Unlike the idiosyncratic case the variance of the ${}^U X_{n,t}$ is always bounded away from zero for any values of b_γ. However, when $b_\gamma < -1/2$ the variance explodes as n becomes large. Moreover, this happens at exactly the same rate as for the variance of the ${}^E X_{n,t}$ so that the conjecture by which common rescaled innovations have stronger impact on aggregate volatility than idiosyncratic ones falls short.

The fact that the asymptotic distribution of the aggregate ${}^U X_{n,t}$ is never degenerate suggests looking at the asymptotic behaviour, in terms of the autocovariance function (ACF), of the square aggregate ${}^U Y_{n,t}$ imposing bounded kurtosis.

Set

$$var_n({}^U Y_{n,t}) = \frac{1}{n^4}\left(3\mathrm{var}_n(\sum_{i=1}^{n} \sigma_{i,t})^2 + 2E_n(\sum_{i=1}^{n} \sigma_{i,t})^4\right). \qquad (11)$$

Theorem 5 *As $n \to \infty$, under Assumption $I(\gamma)$:*
When $\gamma < \sqrt{3}^{-1}$, uniformly in b_γ, for some constant C,

$$\mathrm{var}_n({}^{U}Y_{n,t}) \to C, \ \ a.s., \ 0 < \ C \ < \infty.$$

When $\gamma = \sqrt{3}^{-1}$
(i) If $b_{\sqrt{3}^{-1}} > -3/4$,

$$\mathrm{var}_n({}^{U}Y_{n,t}) \sim C.$$

(ii) If $b_{\sqrt{3}^{-1}} = -3/4$, there exist positive constants c, C such that a.s.

$$c\,(log\, n) \leq \mathrm{var}_n({}^{U}Y_{n,t}) \leq C\,(log\, n)^4.$$

(iii) If $b_{\sqrt{3}^{-1}} < -3/4$, setting $\delta = -(b_{\sqrt{3}^{-1}} + 1)/(3b_{\sqrt{3}^{-1}} + 2)$, a.s.

$$\mathrm{var}_n({}^{U}Y_{n,t}) \sim n^{-\frac{4b_{\sqrt{3}^{-1}}+3}{b_{\sqrt{3}^{-1}}+1}}\, S_\delta,$$

with $S_\delta > 0\,(0 < \delta < 1)$ a.s.

Proof: See Appendix A.

When the α_i are strictly bounded away from $\sqrt{3}^{-1}$ the squared aggregate will always be stationary. However, when $\gamma = \sqrt{3}^{-1}$, the squared aggregate is asymptotically nonstationary for a dense cross-sectional distribution of the α_i around $\sqrt{3}^{-1}$. Hence, the aggregate ${}^{U}X_{n,t}$ might display unbounded kurtosis despite all the micro ARCH(1) have finite kurtosis.

In order to characterize the memory properties of the aggregate without limiting to the bounded kurtosis case, the asymptotic distribution of the aggregate is established.

Hereafter, $\to_1$ denotes convergence in L_1, $\mu_k = E(\alpha_i^k)$ and $E(\omega_i^k) = \rho_k$ for any real k.

Theorem 6 *There exist processes ${}_{(1)}X_{n,t}$, ${}_{(2)}X_{n,t}$ such that*

$$\min[{}_{(1)}X_{n,t}, {}_{(2)}X_{n,t}] \leq {}^{U}X_{n,t} \leq \max[{}_{(1)}X_{n,t}, {}_{(2)}X_{n,t}],\ a.s. \qquad (12)$$

(i) Under Assumption $I\,(\gamma)$, as $n \to \infty$,

$${}_{(1)}X_{n,t} \to_1 {}_{(1)}X_t = u_t\, \rho_{1/2} \left(\sum_{k=0}^{\infty} \mu_{k/2}^2 \prod_{j=1}^{k} u_{t-j}^2 \right)^{1/2}, \qquad (13)$$

$${}_{(2)}X_{n,t} \to_1 {}_{(2)}X_t = u_t\, \rho_{1/2} \left(\sum_{k=0}^{\infty} \mu_{k/2} \prod_{j=1}^{k} |u_{t-j}| \right). \qquad (14)$$

For any real positive γ, ${}_{(1)}X_t$ and ${}_{(2)}X_t$ are bounded (in modulus) a.s., strictly stationary and ergodic if $\gamma e^{\lambda_u} < 1$, uniformly in b_γ. When $\gamma e^{\lambda_u} = 1$, this holds when $b_1 > -1/2$ and $b_1 > 0$ for ${}_{(1)}X_t$, ${}_{(2)}X_t$, respectively.

(ii) The asymptotic stationarity conditions of ${}_{(1)}X_{n,t}$ and ${}_{(2)}X_{n,t}$ and ${}^{U}X_{n,t}$, in the levels and in the squares, coincide (cf. Theorem 4 and 5) and, for $\gamma \leq \sqrt{3}^{-1}$, as $h \to \infty$,

$$cov({}_{(i)}X_t^2, {}_{(i)}X_{t+h}^2) \sim c_i \, \gamma^h \, f_{\gamma, \sqrt{3}^{-1}}(h), \quad i = 1, 2,$$

for constants $0 < c_i < \infty \, (i = 1, 2)$ setting

$$f_{\gamma,\delta}(h) = \begin{cases} h^{-2(b_\gamma+1)}, & \gamma < \delta, \\ h^{-2(b_\delta+1)}(1 + h^{-(2b_\delta+1)}), & \gamma = \delta. \end{cases} \tag{15}$$

Proof: See Appendix A.

The intrinsic nonlinearity of GARCH as well as the cross-sectional dependence of the $\sigma_{i,t}$, do not allow to derive the precise expression for the limit aggregate process. However, for large n the limit of ${}^{U}X_{n,t}$ is enveloped by ${}_{(1)}X_t$ or ${}_{(2)}X_t$, whose statistical properties are more easily obtainable.

In particular both ${}_{(1)}X_t$ and ${}_{(2)}X_t$ display the 'GARCH' nonlinearity. Thus, it does not follow that the nonlinearity characterizing the micro level is washed out by aggregation, in contrast to the idiosyncratic rescaled innovation case. Note that no normalization nor truncation are required, thus deriving the asymptotic distribution of the (plain) aggregate. Remarkably, bounded fourth moment conditions are not imposed, as convergence in L_1 (instead of L_2) is established.

However, the probabilistic properties of the aggregate can be drastically influenced by aggregation. Comparing our result with Nelson (1990b), it follows that when $\gamma e^{\lambda_u} = 1$, the limit aggregate processes, ${}_{(1)}X_t$ and ${}_{(2)}X_t$, might lose the strict stationarity and ergodicity properties characterizing the micro GARCH, for a sufficiently dense distribution of the α_i near $1/e^{\lambda_u} > 1$.

An important implication of Theorem 6 follows with respect to the definitions of strong and weak GARCH. Nijman and Sentana (1996) show that low order weak GARCH are closed under contemporaneous aggregation. The result presented here clearly shows that both strong and weak ARCH are not closed under aggregation for large n. In fact the coefficients μ_k in (13) and (14) cannot be derived by expanding the ratio of polynomials in the lag operator L (cf. (28) in Appendix A for their asymptotic behaviour), e.g. as $(1 + a_1 L + \ldots + a_q L^q)/(1 + b_1 L + \ldots + b_p L^p)$ for given integers $p, q \geq 0$ and constants $a_1, \ldots, a_q, b_1, \ldots, b_p$, as from Drost and Nijman (1993, Definitions 1,2 and 3).

Aggregation induces a change in the memory of the aggregate. Under bounded kurtosis, the squared aggregate displays short memory but of a different type with respect to the memory of micro ARCH. Theorem 6 suggests that a necessary condition for obtaining an hyperbolically decaying ACF, and thus long memory, of the aggregate is $\gamma = 1$. For this purpose, consider the 'truncated' aggregate $\tilde{x}_{i,t} = u_t \tilde{\sigma}_{i,t}$, setting $z_{i,t} = u_t$ in (8). Set ${}^U\tilde{X}_{n,t} = 1/n \sum_{i=1}^n \tilde{x}_{i,t}$ and

$$ {}^U\tilde{y}_{n,t} = \frac{{}^U\tilde{Y}_{n,t}}{\sqrt{\mathrm{var}_n({}^U\tilde{Y}_{n,t})}}, \tag{16} $$

with, as usual, ${}^U\tilde{Y}_{n,t} = {}^U\tilde{X}^2_{n,t}$.

Theorem 7 *There exist processes* $\{{}_{(1)}\tilde{Y}_{n,t}\}$, $\{{}_{(2)}\tilde{Y}_{n,t}\}$ *such that*

$$ {}_{(1)}\tilde{Y}_{n,t} \leq {}^U\tilde{Y}_{n,t} \leq {}_{(2)}\tilde{Y}_{n,t} \ \ a.s. \tag{17} $$

Under Assumption $I(1)$, *uniformly in* b_1, *for* $h = 0, \pm 1, ..,$ *as* $t, n \to \infty$ *(the order is inessential)*

$$ \mathrm{cov}_n({}_{(i)}\tilde{y}_{n,t}, {}_{(i)}\tilde{y}_{n,t+h}) \to 1, \ \ a.s. \ i = 1, 2, $$

setting ${}_{(i)}\tilde{y}_{n,t} = {}_{(i)}\tilde{Y}_{n,t} / \sqrt{\mathrm{var}_n({}_{(i)}\tilde{Y}_{n,t})}$ $(i = 1, 2)$.

Proof: See Appendix A.

Aggregation of strong GARCH does not induce a long memory volatility model for the aggregate[2], in contrast to common wisdom. This is due to

[2]Theorem 7 seems to suggest convergence in distribution of the normalized-truncated squared aggregate to a constant (with respect to time) r.v. This is, however, an artifact, namely the usual result on the misleading information stemming from the moments' convergence, as convergence in distribution to a time-varying random limit holds without need of any normalization (cf. Theorem 6,(i)).

the core nonlinearity of GARCH which, in turn, yields the nonstationary 'innovations' $\prod_{j=1}^{k} u_{t-j}^2$ and $\prod_{j=1}^{k} |u_{t-j}|$ in (13) and (14). On the other hand, under $I(1)$ the coefficients $\mu_{k/2}$ do exhibit the required hyperbolic behaviour, i.e. $\mu_k \sim c\,k^{-(b+1)}$ as $k \to \infty$ for some $0 < c < \infty$ (cf. (28)). Note that renormalizing the u_t such that $E(u_t^4) = 1$ will not change the result as this induces $E(u_t^2) < 1$ (cf. Appendix A for a further discussion).

Hence, the aggregation mechanism cannot invoked to explain the widespread empirical finding of long memory in squared stock return indexes, when maintaining the assumption that single stocks have a GARCH behaviour. This does not mean that aggregation had no effect on the probabilistic properties of the aggregate, as indicated by Theorem 6. For instance GARCH(p, q) are strongly mixing with geometric rate (Davis and Mikosch 1998, Lemma A.2 for ARCH(1)). This result is ultimately based on their Markovian structure which is lost by aggregation.

3 Generalizations

3.1 Aggregation of GARCH

We now discuss extension of the results to the case when the observable micro processes are strong GARCH(1,1)

$$\begin{aligned} x_{i,t} &= z_{i,t}\sigma_{i,t}, \\ \sigma_{i,t}^2 &= \omega_i + \alpha_i x_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2. \end{aligned} \tag{18}$$

We need to introduce the following sequences of r.v.'s

$$\pi_i = (\alpha_i + \beta_i), \tag{19}$$

$$\nu_i = (\alpha_i + \beta_i)^2 + 2\alpha_i^2 = \pi_i^2 + 2\alpha_i^2. \tag{20}$$

The following two conditions replace Assumption $I(\gamma)$.

Assumption III

The ω_i and $\{\alpha_i, \beta_i\}$ satisfy cases (i),(ii) of Assumption $I\,(\gamma)$ and the π_i satisfy $I\,(1)\,(iii)$.

Assumption IV(γ)

The ω_i and $\{\alpha_i, \beta_i\}$ satisfy cases (i),(ii) of Assumption $I\,(\gamma)$ and the ν_i satisfy $I\,(\gamma)\,(iii)$.

Remark IV.1 The cross-sectional distribution of the π_i, the so-called 'persistence' parameter, imparts the statistical properties of the aggregate in the level whereas the cross-sectional distribution of the ν_i imparts the statistical properties of the squared aggregate.
Remark IV.2 When $\beta_i = 0\,(i = 1,..,n)$ yields $\nu_i = 3\alpha_i^2$. In fact, for simplicity's sake, $IV\,(\gamma)$ restricts to the bounded micro kurtosis case and the possibility of $\nu_i \geq 1$ $(\alpha_i > \sqrt{3}^{-1}$ for ARCH(1)) is ruled out.
Remark IV.3 Assumption III and $IV(\gamma)$ can be seen as naturally induced by the distributions of the α_i and β_i, with support $[0, \alpha_{max})$ and$[0, \beta_{max})$, and having probability densities behaving as $C_\alpha(\alpha_{max} - \alpha)^{b_\alpha}$, $\alpha \to 1^-$ and $C_\beta(\beta_{max} - \beta)^{b_\beta}$, $\beta \to 1^-$ for constants $0 < C_\alpha, C_\beta < \infty$ and $b_\alpha, b_\beta > -1$, respectively (cf. Appendix B).

It follows that under Assumption III and $IV\,(\gamma)$ Theorem 1-3, for the case of idiosyncratic innovations, and Theorem 4-7, for the case of common innovations, extend to GARCH(1,1). This contrasts with Ding and Granger (1996) who, considering the case of micro GARCH(1,1), show that only the distribution of the β_i matters and the distribution of the α_i is completely irrelevant. This outcome is due to their particular 'triangular array' structure (cf. their equation (4.16)), which yields a similar yet different definition of aggregation of heterogeneous strong GARCH (cf. (3)). This is confirmed by Theorem 7 which rules out long memory at the aggregate levels, in contrast to Ding and Granger (1996)

However, some care need to be given to the extension of the asymptotic distribution result when the rescaled innovation is common across units (cf. Theorem 6). This is formalized as follows.

Theorem 8 *There exist processes* $\{_{(i)}\tilde{X}_{n,t}\}$, $\{_{(i)}X_{n,t}\}$, $i = 1, 2$ *such that a.s.*

$$\min[_{(1)}\tilde{X}_{n,t},_{(2)}\tilde{X}_{n,t}] \leq \min[_{(1)}X_{n,t},_{(2)}X_{n,t}] \leq {}^{U}X_{n,t} \leq \max[_{(1)}X_{n,t},_{(2)}X_{n,t}] \leq \max[_{(1)}\tilde{X}_{n,t},_{(2)}\tilde{X}_{n,t}].$$

We define $_{(i)}\tilde{X}_t$, $_{(i)}X_t\,(i = 1, 2)$ *in the proof.*

(i) Under Assumption $IV(\gamma)$ *and* $\alpha_{max}^{i/4}\delta_u + \beta_{max}^{i/4} < 1\,(i = 1, 2)$, *as* $n \to \infty$,

$$_{(i)}\tilde{X}_{n,t} \to_1 {}_{(i)}\tilde{X}_t, \;\; i = 1, 2.$$

For any real positive α_{max}, β_{max}, ${}_{(i)}\tilde{X}_t\,(i=1,2)$ *are bounded (in modulus) a.s., strictly stationary and ergodic when* $((\alpha_{max}e^{\lambda_u/2})^{1/i}+\beta_{max}^{1/i})<1\,(i=1,2)$*, uniformly in* b_α, b_β*, respectively. When* $((\alpha_{max}e^{\lambda_u/2})^{1/i}+\beta_{max}^{1/i})=1\ \ (i=1,2)$ *this holds for* $min[b_\alpha,b_\beta]>(i-2)/2\ \ (i=1,2)$.

(ii) The asymptotic stationarity conditions for ${}_{(1)}X_{n,t}$, ${}_{(2)}X_{n,t}$ *and* ${}^{U}X_{n,t}$ *coincide, viz. Assumption* III *with* $b_1>-1/2$ *for the levels and Assumption* $IV(\gamma)$ *with* $\gamma<1$ *or* $\gamma=1$, $b_1>-3/4$ *for the squares. Under the latter conditions, as* $h\to\infty$,

$$cov({}_{(i)}X_t^2,{}_{(i)}X_{t+h}^2)\le c_i\,(\alpha_{max}+\beta_{max})^h\,f_{\gamma,1}(h),\quad i=1,2, \tag{21}$$

for constants $0<c_i<\infty\,(i=1,2)$ *with* $f_{\gamma,1}(h)$ *defined in (15).*

Proof: See Appendix A.

We provide a double envelope for the aggregate. In fact, we exploit the fact that for GARCH(p,q) there exist different nonlinear moving average representations (Zaffaroni 1999), unlike for ARCH(1), and choose the representation suitably depending on whether we need to characterize the asymptotic distribution and the strict stationarity conditions or, more simply, the memory (second-order) properties of the aggregate. Note that in the former case, the ${}_{(i)}\tilde{X}_{n,t}\,(=1,2)$ represent a looser envelope to ${}^{U}X_{n,t}$ as slightly different conditions are required in (i) for ${}_{(1)}\tilde{X}_t$ and ${}_{(2)}\tilde{X}_t$ respectively. This does not apply to the ${}_{(i)}X_{n,t}\,(i=1,2)$ which, as discussed in (ii), exhibit the same second-order asymptotic properties, in turn equal to the ones of ${}^{U}X_{n,t}$, both for the levels and the squares.

When $\beta_{max}=0$, viz. the ARCH(1) case, then assumption $IV(\gamma)\,(0<\gamma\le 1)$ collapses to $I(\gamma)\,(0<\gamma\le\sqrt{3}^{-1}))$. Hence, a comparison with Theorem 6 suggests that the inequality in (21) is in fact sharp.

In contrast to all the other propositions, further conditions had to be imposed beyond $IV(\gamma)$ due to the intrinsic nonlinearity and the greater complexity of higher dimension GARCH models. This holds more generally for aggregation of heterogeneous GARCH(p,q) models, for simplicity's sake reparameterized as GARCH(m,m) for $m=max[p,q]$, given by (18) and

$$\sigma_{i,t}^2=\omega_i+\sum_{j=1}^{m}(\alpha_{i,j}z_{i,t-j}^2+\beta_{i,j})\sigma_{i,t-j}^2.$$

It follows that the cross-sectional distribution of the $\pi_i = \sum_{j=1}^{m}(\alpha_{i,j} + \beta_{i,j})$ defines the probabilistic properties of ${}^{E}X_{n,t}$ and the cross-sectional distribution of the $\nu_i = \sum_{j=1}^{m}(\pi_{i,j}^2 + 2\alpha_{i,j}^2)$ the ones of ${}^{E}Y_{n,t}$.

3.2 Aggregation of conditionally heteroskedastic factor models

This framework can be used to evaluate the impact of aggregation of the components of conditionally heteroskedastic factor models (see e.g. Sentana (1998))

$$x_{i,t} = \beta_{i,1} f_{1,t} + \beta_{i,2} f_{2,t} + .. + \beta_{i,K} f_{K,t} + w_{i,t},\ i = 1, ..., n, \tag{22}$$

where $f_t = (f_{1,t}, ..., f_{K,t})'$ is a vector of K $(n > K)$ unobserved common factors, the $\beta_{i,j}$ $(j = 1, .., K)$ are the associated factor loadings and the $w_{i,t}$ $(i = 1, .., n)$ indicate idiosyncratic r.v.'s, orthogonal to the $f_{j,t}$ $(j = 1, .., K)$.

Setting $w_t = (w_{1,t}, ..., w_{n,t})'$ assume $E_{t-1}(w_t) = 0$, $E_{t-1}(w_t w_t') = \Gamma_t$ and $E_{t-1}(f_t) = 0$, $E_{t-1}(f_t f_t') = \Lambda_t$ with $E_{t-1}(w_t f_t') = 0$, $E_{t-1}(\cdot)$ denoting the expectation operator conditionally on $\{f_s, x_{i,s}, s < t, i = 1, 2, ...\}$. The time variation in Λ_t and Γ_t motivates the denomination of conditionally heteroskedastic factor model. For simplicity's sake we assumed that the factor loadings $\beta_{i,k}$ are time invariant but this could be easily generalized to the case where the factor loadings are time-varying, with $\beta_{i,k,t}$ determined at time $t - 1$.

As showed in Sentana (1998), depending on the specification of the Λ_t and the Γ_t several multivariate volatility models are described by (22), in particular the latent factor model with ARCH factors of Diebold and Nerlove (1989) and the factor GARCH model of Engle (1987). When the $w_{i,t}$ are assumed mutually orthogonal, with $E(\Gamma_t) < \infty$, or mildly correlated across units, (22) is referred to as a conditional exact or approximate K factor structure (Hansen and Richard 1987), generalizing the definition of (unconditional) factor structures (Chamberlain and Rothschild 1983).

Assuming that both the $f_{k,t}$ and the $w_{i,t}$ have time-varying conditional second moment, parameterized as a strong GARCH(p, q), yields for $p = q = 1$

$$vech(\Lambda_t) = \Omega + A_1 vech(f_{t-1} f_{t-1}') + B_1 vech(\Lambda_{t-1}),$$
$$w_{i,t} = z_{i,t}\sigma_{i,t},\ \sigma_{i,t}^2 = \omega_i + \alpha_i w_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2,\ i = 1, .., n,$$

where A_1, B_1 are square matrices and Ω a vector of order $K(K+1)/2$.

Choosing $z_{i,t} = \epsilon_{i,t}$ or alternatively $z_{i,t} = u_t$, satisfying Assumption II, suggests that a conditional exact or approximate K factor structure for the the $x_{i,t}$ could be obtained.

In Chamberlain and Rothschild (1983) the maximum degree of cross-sectional dependence allowed for the $w_{i,t}$ is expressed by boundedness of the maximum eigenvalue of $E(\Gamma_t)$, uniformly in n. This clearly collapses to $lim_{n\to\infty} max_{i=1,..,n} var(w_{i,t}) < \infty$ when Γ_t is diagonal. In our random coefficients framework, when $z_{i,t} = \epsilon_{i,t}$, this requires

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\omega_i}{1-\alpha_i} \to E(\frac{\omega}{1-\alpha}) < \infty \, a.s.$$

as $n \to \infty$. This holds, under Assumption $I(\gamma)$, when either $\gamma < 1$ or $\gamma = 1$ with $b_1 > 0$. Consider the case $z_{i,t} = u_t$ and thus when the $w_{i,t}$ are correlated across units with $E(w_{i,t}w_{j,t}) = E(u_t^2)E(\sigma_{i,t}\sigma_{j,t})$. Under the same above conditions, as $n \to \infty$

$$tr(\Gamma_t) = \sum_{i=1}^{n}\sigma_{i,t}^2 \sim c\, n \quad a.s.,$$

for some $0 < c < \infty$. This is clearly satisfied under the uniformly bounded eigenvalues condition. However, the degree of cross-sectional dependence is too strong as one cannot diversify away the risk induced by the $w_{i,t}$, i.e. $var(1/n\sum_{i=1}^{n} w_{i,t}) \geq c > 0$ for some $0 < c < \infty$, uniformly in b_γ, for any n (cf. Theorem 4). Hence, although setting $z_{i,t} = u_t$ delivers an interesting case of conditionally heteroskedastic factor model with cross-sectional correlated idiosyncratic risk, this rules out the case of (conditional) approximate factor models.

The portfolio, made by $1/n$th of each asset, would then be

$$X_{n,t} = \frac{1}{n}\sum_{i=1}^{n}\beta_i' f_t + \frac{1}{n}\sum_{i=1}^{n} w_{i,t},$$

setting $\beta_i = (\beta_{i,1}, ..., \beta_{i,K})'$. The impact of aggregation on the factors part depends entirely on the cross-sectional properties of the β_i. The case of static factor loadings is straightforward but one could generalize the framework introducing time variation, e.g. if the factor loadings $\beta_{i,t}$ are modelled as

ARMA, then apply the aggregation framework of Lippi and Zaffaroni (1999). On the other hand, using the previous results, we can completely characterize the statistical properties of the idiosyncratic part.

This has sound implications. In particular, when there are no common factors ($K = 0$) an exact zero factor structure would induce a Gaussian model for the portfolio and not a volatility model. Moreover, when the cross-sectional distribution of the α_i is dense around 1 ($b_1 < -1/2$), the average of the idiosyncratic component would not vanish but it rather displays an unbounded variance as n gets large, even in the exact factor structure case. Hence, idiosyncratic risk may not be fully diversifiable even when one can trade a possibly infinite number of assets. This is also relevant when developing statistical inference methods on such nonlinear factor model based on a large cross-sections as the commonly held hypothesis of a vanishing importance of the idiosyncratic part of the portfolio fails.

3.3 Further extensions

(i) When one focuses on strictly stationary micro GARCH, this implies the possibility of inducing explosive behaviours at the aggregate levels when the $\alpha_i \geq 1$. Aggregation of strong IGARCH is an important particular case, where the π_i will have a degenerate distribution at 1.

Although the case $1 < \gamma$ does not appear to be empirically relevant (see e.g. Bollerlsev, Chou, and Kroner (1992)), our framework can easily account for such possibility. Clearly now one needs to evaluate conditional moments not only with respect to the GARCH coefficients but also with respect to past rescaled innovations, viz. consider the truncated aggregates ${}^E\tilde{X}_{n,t}$, ${}^U\tilde{X}_{n,t}$. For example, focusing on the idiosyncratic case, one obtains that as $t, n \to \infty$, under $I(\gamma)$

$$\text{var}_n({}^E\tilde{X}_{n,t}) \sim \gamma^t \left(c\,(log\,log(n\,\mu_{2t}))^{1/2}\,\frac{t^{-b_\gamma/2+1/2}}{n^{3/2}} + c'\frac{t^{-b_\gamma}}{n} \right),$$

for positive constants $0 < c, c' < \infty$. We skip details for simplicity's sake. In contrast to Theorem 1, even allowing $\gamma = 1$ then, irrespective of the value of b_γ, ${}^E\tilde{X}_{n,t}$ converges to zero in mean-square as $t/n \to \infty$ suggesting that the usual result arises when n is large compared with t. A fixed t is an important particular case. When $n \sim ct^{b_\gamma+1}$ for some $c > 0$ the results of Theorem 1 are re-obtained. On the other hand, when $\gamma > 1$, then the rate of divergence is exponential with respect to the time dimension. By the same arguments, the variance of the limit Gaussian process in Theorem 3,(ii) explodes at an exponential rates and it loses the self-similar property. Parallel results can be obtained for the common component case.

(ii) The assumption of independence (cf. remark I.3) between the α_i and the ω_i plays an important role. In fact, the results will be affected by the assumed shape and degree of mutual dependence; e.g. in the extreme case that $\omega_i = \tilde{\omega}_i(1 - \alpha_i^2)$ for some i.i.d. sequence $\{\tilde{\omega}_i\}$ (independent of the α_i), then the usual result arises for any shape of the cross-sectional distribution. More in general, we can assume that the distribution of the ω_i, conditioning on the α_i, behaves as $B(\omega_i \mid \alpha_i) \sim C_\omega (\gamma - \alpha_i)^{b_\omega}$ as $\alpha_i \to \gamma^-$ for some $-1 < b_\omega$, $0 < C_\omega < \infty$. The results would depend on the magnitude of $b_\gamma + b_\omega$.

(iii) We can allow for cross-correlation across units not only through a common rescaled innovation but also by assuming dependence across the α_i. This could be potentially relevant from an economic standpoint. Indeed, the limit laws which this paper is based on and employed in Lemma 1, have been extended to the case of stationary dependent sequences satisfying some form of mixing condition (see references in Samorodnitsky and Taqqu (1994, pg. 575)) and therefore fairly easily adaptable to our framework. There is nonetheless a problem of interpretation in adapting the time series meaning of dependence to a cross-sectional framework, except for the limit case of independence.

4 Exploiting the linear ARMA representation of GARCH

It is well known that any GARCH(p,q) process can be represented as an $ARMA(m, p)$ with $m = max\{p, q\}$ in the squared process (Bollerslev 1986). In our case

$$y_{i,t} = \omega_i + \sum_{j=1}^{m} (\alpha_{i,j} + \beta_{i,j}) y_{i,t-j} - \sum_{h=1}^{q} \beta_{i,h} v_{i,t-h} + v_{i,t}, \tag{23}$$

where $y_{i,t} = x_{i,t}^2$ and the $v_{i,t}$ are a martingale difference sequence given by

$$v_{i,t} = (z_{i,t}^2 - 1)\sigma_{i,t}^2.$$

It follows that, conditioning on the $\alpha_{i,j}, \beta_{i,j}$ and the ω_i, conditional moments of the $y_{i,t}$ can be evaluated straightforwardly, suggesting that the aggregation results for linear ARMA processes can be employed directly. However, the distribution of the shock $v_{i,t}$ is a function of the autoregressive and moving average parameters, unlike from the linear ARMA case. Thus the difference between the idiosyncratic and common components is rather vacuous as, even in the case $z_{i,t} = u_t$, the $v_{i,t}$ are still a function of the index i and thus can no longer be interpreted as common shocks. More importantly, this structure of the non-Gaussian innovations $v_{i,t}$ delivers different expressions for the conditional moments, with respect to the linear ARMA. Therefore the advantage of using the linear representation (23) are clearly nil.

Considering the ARCH(1) case ($p = 0$, $q = 1$ in (23)), setting

$$\dot{Y}_{n,t} = \frac{1}{n^2} \sum_{i=1}^{n} y_{i,t},$$

simple calculations yield

$$E_n(\dot{Y}_{n,t}) = \frac{1}{n^2} \sum_{i=1}^{n} \frac{\omega_i}{1 - \alpha_i},$$

$$\mathrm{var}_n(\dot{Y}_{n,t})$$

$$= \frac{1}{n^4} \sum_{i,j=1}^{n} \frac{\omega_i \omega_j}{(1 - \alpha_i)(1 - \alpha_j)} \frac{E(z_{i,t}^2 - 1)(z_{j,t}^2 - 1)}{[E(z_{i,t} z_{j,t})^2 (1 - \alpha_i \alpha_j) - E(z_{i,t}^2 - 1)(z_{j,t}^2 - 1)]}.$$

Note that the $\dot{Y}_{n,t}$ are not the arithmetic averages of the $y_{i,t}$, as the normalization $1/n^2$ is used. When $z_{i,t} = \epsilon_{i,t}$, $E_n(\dot{Y}_{n,t})$ coincides with $\mathrm{var}_n({}^E X_{n,t})$ whereas $var_n(\dot{Y}_{n,t})$ does not coincide with $\mathrm{var}_n({}^E Y_{n,t})$ (cf. (6)), although the additional terms are of smaller order in n and would not influence the asymptotic results. However, when $z_{i,t} = u_t$ the double product terms, which represent the important difference between the idiosyncratic and common cases, are excluded.

Thus, evaluation of the aggregation mechanism via the ARMA representation, henceforth 'linear aggregation', albeit simpler in terms of calculation of the conditional moments, would yield a misleading inference on the importance of common versus idiosyncratic shocks with respect to the probabilistic properties of the aggregate. Another difficulty of the 'linear aggregation' arises when one needs to evaluate the absolute first moment of the ${}^{U}X_{n,t}$. Comparing (23) with Assumption *III* it follows that the impact of aggregation on GARCH(p, q) is uniquely determined by the autoregressive part of the model, the moving average part having absolutely no influence, in analogy with the case of aggregation of linear ARMA processes; see e.g. Granger (1980).

5 Conclusions

In this paper we establish the impact of contemporaneous aggregation on heterogeneous GARCH processes by means of asymptotic results in the number of micro units, e.g. approximating the dynamics of a large portfolio with GARCH single stocks. The key features of the micro structure are both the shape of the cross sectional distribution of the GARCH coefficients and the degree of cross-sectional dependence of the rescaled innovations. When the micro units are purely idiosyncratic, sufficient conditions for aggregate non-degeneracy do exist, in contrast to the common belief of a vanishing importance of idiosyncratic risk. However, the nonlinearity is lost through aggregation, unlike the case of common rescaled innovations. Even though in the latter case the aggregate is a volatility model, non-GARCH memory properties arise. Long memory conditional heteroskedasticity is ruled out. Unlike the small number of units case, even the class of weak GARCH is not closed under aggregation over a large number of units.

Contemporaneous aggregation, in the asymptotic sense, is an important mechanism and its properties are only partially known even for the most common volatility models. A desirable feature for classes of volatility models would be to obtain long memory versions of the models based on aggregation of short memory ones. This holds for linear processes (Lippi and Zaffaroni 1999) but, as we have seen, fails for GARCH. From an empirical standpoint, both the assumptions and the testable implications, e.g. on the memory of the squared aggregate, can be analyzed using both disaggregate data on single stocks and stock indexes data. These issues are the focus of forthcoming research.

6 Appendix A

We recall that c, C denote arbitrary constants, always bounded but not necessarily the same, the symbol $\sim$ denotes asymptotic equivalence and $P(A)$, 1_A, respectively, the probability and the indicator function of any event A.

We begin with the following lemma which adapts to out framework known results on convergence of normed sums in i.i.d. random variables in the domain of attraction of a possibly non-normal stable distribution.

Lemma 1 *Given a a sequence of i.i.d. positive r.v.'s with probability density $B(\cdot; b)$ defined in the interval $[0, \gamma)$ for real $\gamma > 0$ such that for a real $b\,(-1 < b < \infty)$ and a $C_b > 0$, as $\alpha \to 1^-$,*

$$B(\alpha; b) \sim C_b(\gamma - \alpha)^b\, L(1/(\gamma - \alpha)), \tag{24}$$

where $L(\cdot)$ denotes a slowly varying function. Set $\delta = (b + 1)/k$. Then a.s., as $n \to \infty$,

(1) If $2 < \delta$,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(\gamma - \alpha_i)^k} \sim E(\frac{1}{(\gamma - \alpha)^k}) + n^{-1/2}S_2,$$

(2) If $1 < \delta < 2$,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(\gamma - \alpha_i)^k} \sim E(\frac{1}{(\gamma - \alpha)^k}) + \tilde{L}(n)n^{1/\delta - 1}S_\delta,$$

(3) If $\delta = 1$,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(\gamma - \alpha_i)^k} \sim \hat{L}(n) + \tilde{L}(n)S_1,$$

(4) If $0 < \delta < 1$,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{(\gamma - \alpha_i)^k} \sim \tilde{L}(n)\, n^{1/\delta - 1}S_\delta,$$

where $\tilde{L}(n), \hat{L}(n)$ are slowly varying functions. We recall that $S_\delta\,(0 < \delta \leq 2)$ defines a δ-stable r.v. with zero location parameter (cf. footnote 1), including the case of the normal distribution (S_2). In case (4) S_δ will be a totally skewed to the right δ-stable r.v. with zero location parameter, implying $S_\delta > 0$ a.s.

Proof: Under (24) the $y_i = 1/(\gamma - \alpha_i)^k$ have distribution in the domain of attraction of possibly non-normal stable distribution with index δ. In fact, denoting by $f_y(\cdot)$ the probability density function of the variables y_i,

$$f_y(u) = \frac{1}{k} B(\gamma - u^{-1/k}; b) u^{-1/k-1}, \quad 1/\gamma^k \leq u < \infty,$$

with $f_y(u) \sim u^{-(b+1)/k-1} L(u^{1/k})$ as $u \to \infty$. Therefore, as $n \to \infty$,

$$P(y_i \geq a_n) \sim L(a_n^{1/k})(a_n)^{-(b+1)/k}.$$

For cases (2), (3), (4) the result follows applying any non-Gaussian central limit theorem (CLT) for i.i.d. variates (see e.g. LePage, Woodroofe, and Zinn (1981)) to scaled and normalized partial sums

$$\frac{1}{a_n}(\sum_{i=1}^n y_i - b_i),$$

with the sequence $\{a_n\}$, $\{b_n\}$ defined by

$$nP(y_i > a_n) \to 1, \quad n \to \infty,$$
$$b_n = \int_{1/\gamma^k}^{a_n} x f_y(x) dx.$$

Hence, as $n \to \infty$, one obtains

$$a_n = n^{1/\delta} \, \tilde{L}(n),$$

for a slowly varying function $\tilde{L}(\cdot)$ induced by $L(\cdot)$. Then the result follows noting that for i.i.d. variates convergence in probability implies convergence a.s. with, as $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^n y_i \sim \tilde{L}(n) \, n^{1/\delta - 1} \frac{1}{\tilde{L}(n) \, n^{1/\delta}} \sum_{i=1}^n (y_i - b_i) \; + \; \frac{1}{n} \sum_{i=1}^n b_i, \; a.s.$$

Note that when $\delta < 1$ one can set $b_n = 0$ and when $\delta > 1$ one can use the unconditional expectation, i.e. $b_i = E(y_i) = E(y_1)$.

When $\delta < 1$ we make use of the fact that if $n^{-\beta} Z_n \to_d Z$, as $n \to \infty$, for r.v.'s $\{Z_n\}$, Z and constant $\beta > 0$ with $Z > 0$ $a.s.$, then the Z_n diverge to plus infinity in probability. Under the assumptions made, the limit δ-stable r.v. is totally skewed to the right for $\delta < 1$, and thus its support is included in the half-positive real line (Samorodnitsky and Taqqu 1994, Proposition 1.2.11 and Theorem 1.8.1 (ii)).

When $\delta = 1$, by Yong (1974, Lemma I-11 (1-32')), as $\to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} b_i \sim \hat{L}(n),$$

setting $\hat{L}(n) = L(n^{1/(b+1)}\, \tilde{L}^{1/(b+1)}(n))\, log(n\, \tilde{L}(n))$.

For case (1) a standard CLT for i.i.d. variates (Lindeberg-Lévy) applies.

□

Remark When $L(\cdot) = 1$ then $\tilde{L}(\cdot) = 1$ and $\hat{L}(\cdot) = log(\cdot)$.

Lemma 2 *Under the assumptions of Lemma 1 with $\gamma = 1$, for any integer $p = 1, 2, ..$ and real k, as $n \to \infty$,*

(i)

$$\frac{1}{n^p}\sum_{i_1,..,i_p=1}^{n} \frac{1}{(1-\alpha_{i_1}...\alpha_{i_p})^k} \sim c + C\frac{1}{n}\sum_{i=1}^{n}(1-\alpha_i)^{(p-1)b+(p-1-k)},$$

for constants $0 < c, C < \infty$. The boundedness condition is $pb + (p-k) > 0$.

(ii) When $pb + (p-k) > 0$ for any integer $u > 0$ and $r\,(0 \le r \le p)$ with $s = p - r$, as $n \to \infty$,

$$\frac{1}{n^p}\sum_{i_1,...i_r...,i_p=1}^{n} \frac{\alpha_{i_1}^u..\alpha_{i_r}^u}{(1-\alpha_{i_1}...\alpha_{i_p})^k} \to g_{(r,s)}^{(k)}(u), \quad a.s.$$

where, as $u \to \infty$,

$$g_{(r,s)}^{(k)}(u) \sim c\,(\mu_u)^r(1 + u^{-(sb+s-k)}),$$

for $0 < c < \infty$.

Proof: Case $k < 0$ is trivial. We discuss case $p = 2$ and $k = 1$ as the other cases follow exactly along the same lines.

(i) As $n \to \infty$, for some $0 < c < \infty$,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1-\alpha_i\alpha_j} \to c\int_0^1 (1-t)^b(1-\alpha_j t)^{-1}dt < \infty \;\; a.s.,$$

by $I(1)$. Using Gradshteyn and Ryzhik (1994, # 3.197-3), the integral of the RHS equals

$$B(1,b+1)\,{}_2F_1(1,1;2+b;\alpha_j), \tag{25}$$

where $B(\cdot,\cdot)$ is the Beta function and ${}_2F_1(\cdot,\cdot;\cdot;\cdot)$ denotes the hypergeometric function (Gradshteyn and Ryzhik 1994, section 9.1). Hence, as $\alpha_j \to^- 1$, by Gradshteyn and Ryzhik (1994, # 9.122-1 and # 9.131-1)

$$(25) \sim \left(c\,1_{b>0} + c' log(1-\alpha_j)\,1_{b=0} + c''(1-\alpha_j)^b\,1_{b<0}\right),$$

for $0 < c, c', c'' < \infty$ yielding, as $n \to \infty$,

$$\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{1-\alpha_i\alpha_j} \sim c + C\frac{1}{n}\sum_{i=1}^{n}(1-\alpha_i)^b,$$

for $0 < c, C < \infty$.

(ii) When $r = 2$ ($s = 0$), by Gradshteyn and Ryzhik (1994, # 3.197-3), as $n \to \infty$,

$$\frac{1}{n^2}\sum_{i,j=1}^{n}\frac{\alpha_j^u\alpha_i^u}{1-\alpha_i\alpha_j} \sim \frac{c}{n}\sum_{i=1}^{n}\alpha_i^u B(u+1,b+1){}_2F_1(1,u+1;u+2+b;\alpha_i),$$

for $0 < c < \infty$. Using the aforementioned results of Gradshteyn and Ryzhik (1994), as $\alpha_i \to 1^-$,

$$\begin{aligned}&{}_2F_1(1,u+1;u+2+b;\alpha_i)\\ &\sim 1_{b>0}{}_2F_1(1,u+1;u+2+b;1) + 1_{b<0}(1-\alpha_i)^b{}_2F_1(u+1+b,b+1;u+2+b;1)1_{b<0},\end{aligned}$$

the logarithmic term, arising when $b = 0$, being absorbed by the second one. Simplifying terms yields, as $n \to \infty$,

$$\frac{1}{n^2}\sum_{i,j=1}^{n}\frac{\alpha_j^u\alpha_i^u}{1-\alpha_i\alpha_j} \sim 1_{b>0}\frac{\Gamma(b)\Gamma(u+1)}{\Gamma(u+1+b)}\frac{c}{n}\sum_{i=1}^{n}\alpha_i^u + 1_{b<0}\frac{c'}{n}\sum_{i=1}^{n}\alpha_i^u(1-\alpha_i)^b, \tag{26}$$

for $0 < c, c' < \infty$. By Stirling's formula (Brockwell and Davis 1987, pg. 522)

$$\int_0^\delta u^k(\delta-u)^b \sim C\,\delta^{k+b-1}\,k^{-(b+1)}, \tag{27}$$

as $k \to \infty$, for $b > -1$, $0 < C < \infty$ and $0 < \delta < \infty$. Thus, under $I(\gamma)$ for $0 < c < \infty$, as $k \to \infty$,

$$\mu_k \sim c\,\gamma^k\, k^{-(b+1)}. \tag{28}$$

It follows that the limit (as $n \to \infty$) of both terms on the RHS of (26) are asymptotically equivalent to $u^{-(2b+1)}$, for $u \to \infty$. Recall that to impose stationarity $2b+1 > 0$.

Finally, when $r = 1$ ($s = 1$), as $n \to \infty$,

$$\frac{1}{n^2}\sum_{i,j=1}^{n}\frac{\alpha_i^u}{1-\alpha_i\alpha_j} \sim c\,\frac{1}{n}\sum_{i=1}^{n}\alpha_i^u 1_{b>0} + c'\,\frac{1}{n}\sum_{i=1}^{n}\alpha_i^u(1-\alpha_i)^b 1_{b<0},$$

for $0 < c, c' < \infty$. □

Proof of Theorem 1: Apply Lemma 1 with $k = 1$. □

Proof of Theorem 3: (i) Given the i.i.d.ness of the $x_{i,t}$, the Lindeberg-Lévy CLT applies, as $n \to \infty$. Note that $1/n\sum_{i=1}^{n}\omega_i/(1-\alpha_i)$ converges to $E\left(\omega_i/(1-\alpha_i)\right)$ a.s., bounded when $b_1 > 0$. Easy calculations gives, for any integer $u > 0$ and any n,

$$cov_n\Big(\frac{1}{n^{1/2}}\sum_{i=1}^{n}x_{i,t}, \frac{1}{n^{1/2}}\sum_{i=1}^{n}x_{i,t+u}\Big) = 0.$$

where $cov_n(.,.)$ denotes the covariance operator, conditioning on the ω_i, α_i $(i = 1,..,n)$.

(ii) The convergence in distribution is obtained following (i). From (28), when $b_1 < 0$, as $t \to \infty$,

$$E(\tilde{x}_{i,t}^2) = V_t = E\left(\omega_i\frac{1-\alpha_i^t}{1-\alpha_i}\right) = E(\omega_i)\sum_{k=0}^{t-1}\mu_k \sim c\,t^{-b_1},$$

for some $0 < c < \infty$. □

Proof of Theorem 4: By Schwarz inequality

$$\sigma_{i,t}\sigma_{j,t} \geq \omega_i^{1/2}\omega_j^{1/2}\left(\sum_{a=0}^{\infty}(\alpha_i\alpha_j)^{a/2}\prod_{h=1}^{a}u_{t-h}^2\right),$$

and taking expectations

$$\mathrm{var}_n({}^U X_{n,t}) \geq \frac{c}{n^2}\sum_{i,j=1}^{n}\frac{(\omega_i\omega_j)^{1/2}}{(1-\alpha_i\alpha_j)},$$

and, likewise,

$$\mathrm{var}_n({}^U X_{n,t}) \le \left(\frac{C}{n} \sum_{i=1}^{n} \frac{\omega_i^{1/2}}{(1-\alpha_i)^{1/2}} \right)^2 ,$$

for $0 < c, C < \infty$. Case $\gamma < 1$ easily follows as

$$1/n^2 \sum_{i,j=1}^{n} 1/(1-\alpha_i\alpha_j) = \sum_{k=0}^{\infty} \left(1/n \sum_{i=1}^{n} \alpha_i^k \right)^2 \le 1/(1-\gamma^2) < \infty,$$

and $1/n \sum_{i=1}^{n} \alpha_i^k \to \mu_k$ a.s. uniformly in k. When $\gamma = 1$, apply Lemma 1 with $k = 1$ and $k = 1/2$ respectively. □

Proof of Theorem 5: Bound $E(\sum_{i=1} \sigma_{i,t})^4$, the relevant term in $var_n({}^U Y_{n,t})$, by Schwarz inequality as follows:

$$E(\sum_{i=1}^{n} \sigma_{i,t})^4 \ge \frac{1}{n^4} \sum_{i,j,r,s=1}^{n} \frac{(\omega_i\omega_j\omega_r\omega_s)^{1/2}}{(1-3(\alpha_i\alpha_j\alpha_r\alpha_s)^{1/2})} \frac{(1+\alpha_i\alpha_j\alpha_r\alpha_s)}{(1-\alpha_i\alpha_j\alpha_r\alpha_s)}, \quad (29)$$

$$E(\sum_{i=1}^{n} \sigma_{i,t})^4 \le \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\omega_i^{1/2}(1+\alpha_i)^{1/4}}{(1-\alpha_i)^{1/4}(1-3\alpha_i^2)^{1/4}} \right)^4 . \quad (30)$$

Then apply Lemma 1 with $k = 1/4$ (upper bound) and Lemma 2 with $p = 4$ and $k = 1$ (lower bound). □

Proof of Theorem 6: From two versions of Minkowski's inequality (Hardy, Littlewood, and Polya 1964, generalization of Theorems 24 and 25 to infinite series (Ch.V)), for any sequence $a_{i,j}, i = 1, \dots$ and $j = 1, \dots, n$ one obtains:

$$\left(\sum_{i=0}^{\infty} (\frac{1}{n} \sum_{j=1}^{n} (a_{i,j})^{1/2})^2 \right)^{1/2} \le \frac{1}{n} \sum_{j=1}^{n} \left(\sum_{i=0}^{\infty} a_{i,j} \right)^{1/2} \le \left(\sum_{i=0}^{\infty} \frac{1}{n} \sum_{j=1}^{n} (a_{i,j})^{1/2} \right), \quad (31)$$

yielding (12), where

$$_{(1)}X_{n,t} = u_t \left(\sum_{k=0}^{\infty} (\frac{1}{n} \sum_{i=1}^{n} \omega_i^{1/2} \alpha_i^{k/2})^2 \prod_{j=1}^{k} u_{t-j}^2 \right)^{1/2},$$

$$_{(2)}X_{n,t} = u_t \left(\sum_{k=0}^{\infty} \frac{1}{n} \sum_{i=1}^{n} \omega_i^{1/2} \alpha_i^{k/2} \prod_{j=1}^{k} \mid u_{t-j} \mid \right).$$

(i) Under Assumption $I(\gamma)$, setting $\bar{\chi}_a = (1/n\sum_{i=1}^n \omega_i^{1/2}\alpha_i^a)$ and $\chi_a = E(\bar{\chi}_a)$, using a version of the law of iterated logarithms (Stout 1974, Corollary 5.2.1), i.e. as $n \to \infty$

$$\bar{\chi}_a - \chi_a \sim \frac{2^{1/2}\, var^{1/2}(\omega_i^{1/2}\alpha_i^a)}{n^{1/2}} (\log\log(n\, var(\omega_i^{1/2}\alpha_i^a)))^{1/2}\, a.s., \qquad (32)$$

together with (28) and $var(\omega_i^{1/2}\alpha_i^a) \leq E(\omega_i\alpha_i^{2a})$, yields

$$\begin{aligned} & E_n \mid {}_{(1)}X_{n,t} - {}_{(1)}X_t \mid \\ \leq\; & \delta_u \sum_{k=0}^{\infty} \mid (\bar{\chi}_{\frac{k}{2}})^2 - (\chi_{\frac{k}{2}})^2 \mid^{1/2} \delta_u^k = O_{a.s.}((\frac{\rho_1 log\, log\, n}{n})^{1/4} \sum_{k=0}^{\infty} k^{-(b_\gamma+1)/4}(\gamma^{1/4}\delta_u)^k), \\ & E_n \mid {}_{(2)}X_{n,t} - {}_{(2)}X_t \mid \\ \leq\; & \delta_u \sum_{k=0}^{\infty} \mid \bar{\chi}_{\frac{k}{2}} - \chi_{\frac{k}{2}} \mid \delta_u^k = O_{a.s.}((\frac{\rho_1 log\, log\, n}{n})^{1/2} \sum_{k=0}^{\infty} k^{-(b_\gamma+1)/2}(\gamma^{1/2}\delta_u)^k), \end{aligned}$$

$O_{a.s.}(\cdot)$ denoting an $O(\cdot)$ that holds a.s., where the first inequality is obtained using

$$\mid a^{1/2} - c^{1/2} \mid \leq \mid a - c \mid^{1/2},$$

for any real positive a, c. Recalling that $\chi_k = \rho_{1/2}\mu_k$ by $I(\gamma)$, as $n \to \infty$,

$${}_{(1)}X_{n,t} \to_1 {}_{(1)}X_t, \quad {}_{(2)}X_{n,t} \to_1 {}_{(2)}X_t.$$

Note that this holds for $\gamma^{i/4}\delta_u < 1\,(i = 1, 2)$ or $\gamma^{i/4}\delta_u = 1$ with $b_\gamma > 4/i - 1$ and, thus, for a larger set than the admissible values of γ as stated by Assumption $I(\gamma)$.

Concerning the boundedness, strict stationarity and ergodicity properties, we adapt the proof of Nelson (1990b, Theorem 2). By Dudley (1989, Theorem 8.3.5), with probability one there exists a constant $K < \infty$ such that for all $k > K$

$$\prod_{j=1}^{k} \gamma u_{t-j}^2 = O((\gamma e^{\lambda_u})^{k/2}), \;\; a.s., \qquad (33)$$

using (28).

For ${}_{(2)}X_t$, replacing $\gamma\, u_{t-j}^2$ with $\gamma^{1/2}|u_{t-j}|$ in (33), the same applies noting that $Elog\,|u_t| = \lambda_u/2$. Using (33) in (13) and (14) yields the stated conditions for

$$|\,{}_{(i)}X_t\,| < \infty, \quad a.s. \;\; i = 1, 2.$$

Strict stationarity and ergodicity follows using Stout (1974, Theorem 3.5.8) and Royden (1980, Proposition 5 and Theorem 3), by the same arguments used in Nelson (1990b, pg. 329).

(ii) Consider ${}_{(1)}X_t$. For simplicity's sake, set $\omega_i = 1\,(i = 1,..,n)$ as this is completely innocuous. Then for integer $u > 0$

$$cov({}_{(1)}X_t^2, {}_{(1)}X_{t+u}^2) = \sum_{k,r=0}^{\infty} \mu_{k/2}^2 \mu_{r/2}^2 cov(u_t^2 \prod_{j=1}^{k} u_{t-j}^2, u_{t+u}^2 \prod_{s=1}^{r} u_{t+u-s}^2),$$

and, by means of the cumulants' theorem (Leonov and Shiryaev 1959) one easily obtains

$$cov(u_t^2 \prod_{j=1}^{k} u_{t-j}^2, u_{t+u}^2 \prod_{s=1}^{r} u_{t+u-s}^2) = E(\prod_{s=0}^{u-1} u_{t+u-s}^2) \times \tag{34}$$
$$\left(var(u_t^2) E(\prod_{j=1}^{k} u_{t-j}^2) E(\prod_{s=u+1}^{r} u_{t+u-s}^2) + E(u_t^4) cov(\prod_{j=1}^{k} u_{t-j}^2, \prod_{s=u+1}^{r} u_{t+u-s}^2) \right),$$

taking $r > u - 1$ for otherwise (34) vanishes. Hence using Gradshteyn and Ryzhik (1994, # 3.381-3 and # 8.357) and straightforward evaluation of expectations yields terms such as

$$\sum_{k=0}^{\infty} \mu_{k/2}^2\, \mu_{(k+u)/2}^2 3^k \sim c\, \gamma^u \sum_{k=0}^{\infty} (3\gamma^2)^k k^{-2(b_\gamma+1)}\, (k+u)^{-2(b_\gamma+1)}.$$

for $0 < c < \infty$ as $u \to \infty$. Distinguishing between the two cases $3\gamma^2 = 1$ and $3\gamma^2 < 1$ yields the result, e.g. when $3\gamma^2 = 1$

$$cov({}_{(1)}X_t^2, {}_{(1)}X_{t+u}^2) \sim c\, \mu_{u/2}^2 \left(1 + h^{-(2b_\gamma+1)}\right), \quad u \to \infty,$$

for $0 < c < \infty$.

Using the same arguments, by means of tedious calculations, the same applies to the ACF of ${}_{(2)}X_t^2$, noting that

$${}_{(2)}X_t^2 = {}_{(1)}X_t^2 + u_t^2 \sum_{\substack{k_1 \neq k_2 \\ =0}}^{\infty} \mu_{k_1/2} \mu_{k_2/2} \prod_{j=1}^{k_1} \prod_{s=1}^{k_2} |u_{t-j}| |u_{t+u-s}| \,.$$

□

Proof of Theorem 7: Set $\omega_i = 1\,(i = 1,..,n)$ as this is completely innocuous. Using (31) and squaring yields (17). Recall that

$$
{}_{(1)}\tilde{y}_{n,t} = \frac{1}{\sqrt{\mathrm{var}_n({}_{(1)}\tilde{Y}_{n,t})}} u_t^2 \left(\sum_{k=0}^{t-1} (\frac{1}{n} \sum_{i=1}^{n} \alpha_i^{k/2})^2 \prod_{s=1}^{k} u_{t-s}^2 \right),
$$

$$
{}_{(2)}\tilde{y}_{n,t} = \frac{1}{\sqrt{\mathrm{var}_n({}_{(2)}\tilde{Y}_{n,t})}} u_t^2 \left(\sum_{k=0}^{t-1} (\frac{1}{n} \sum_{i=1}^{n} \alpha_i^{k/2}) \prod_{s=1}^{k} |u_{t-s}| \right)^2.
$$

Let us consider ${}_{(1)}\tilde{y}_{n,t}$. Using the cumulant's theorem yields, for integer $h \geq 0$,

$$
\begin{aligned}
& cov_n({}_{(1)}\tilde{Y}_{n,t}, {}_{(1)}\tilde{Y}_{n,t+h}) \\
&= E(\prod_{j=0}^{h-1} u_{t+h-j}^2) cov_n(u_t^2 \sum_{k=0}^{t-1}(1/n \sum_{i=1}^{n} \alpha_i^{k/2})^2 \prod_{s=1}^{k} u_{t-s}^2, u_t^2 \sum_{r=0}^{t-1}(1/n \sum_{a=1}^{n} \alpha_a^{r/2} \alpha_a^{h/2})^2 \prod_{s=1}^{r} u_{t-s}^2) \\
&= \frac{3}{n^4} \sum_{i,j,a,b=1}^{n} (\alpha_a \alpha_b)^{h/2} \left(\frac{1-(3(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2})^t}{1-3(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2}} - \frac{1-(\alpha_i\alpha_j\alpha_a\alpha_b)^{t/2}}{1-(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2}} \right. \\
&+ \frac{1}{1-(\alpha_i\alpha_j)^{1/2}} \left[\frac{1-(3(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2})^t}{1-3(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2}} - (\alpha_i\alpha_j)^{t/2} \frac{1-(3(\alpha_a\alpha_b)^{1/2})^t}{1-3(\alpha_a\alpha_b)^{1/2}} \right. \\
&- \left. \frac{1-(\alpha_i\alpha_j\alpha_a\alpha_b)^{t/2}}{1-(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2}} + (\alpha_i\alpha_j)^{t/2} \frac{1-(\alpha_a\alpha_b)^{t/2}}{1-(\alpha_a\alpha_b)^{1/2}} \right] \\
&+ \frac{1}{1-(\alpha_a\alpha_b)^{1/2}} \left[\frac{1-(3(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2})^t}{1-3(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2}} - (\alpha_a\alpha_b)^{t/2} \frac{1-(3(\alpha_i\alpha_j)^{1/2})^t}{1-3(\alpha_i\alpha_j)^{1/2}} \right. \\
&- \left. \frac{1-(\alpha_i\alpha_j\alpha_a\alpha_b)^{t/2}}{1-(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2}} + (\alpha_a\alpha_b)^{t/2} \frac{1-(\alpha_i\alpha_j)^{t/2}}{1-(\alpha_i\alpha_j)^{1/2}} \right] \\
&+ \left. \frac{2}{3} \frac{1-(\alpha_a\alpha_b)^{t/2}}{1-(\alpha_a\alpha_b)^{1/2}} \frac{1-(\alpha_i\alpha_j)^{t/2}}{1-(\alpha_i\alpha_j)^{1/2}} \right).
\end{aligned}
$$

Let us consider first the case when t goes to infinity before n. Under $I(1)$, for a sufficiently large n, $3(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2} > 1$ for some set of indexes i, j, a, b. It is absolutely innocuous assuming that this holds for all the summands yielding, as $t \to \infty$,

$$
\begin{aligned}
& cov_n({}_{(1)}\tilde{Y}_{n,t}, {}_{(1)}\tilde{Y}_{n,t+h}) \\
&\sim \frac{3}{n^4} \sum_{i,j,a,b=1}^{n} (\alpha_a\alpha_b)^{h/2} \frac{(3(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2})^t}{3(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2}-1} (1 + \frac{3(\alpha_a\alpha_b)^{1/2}}{3(\alpha_a\alpha_b)^{1/2}-1} + \frac{3(\alpha_i\alpha_j)^{1/2}}{3(\alpha_i\alpha_j)^{1/2}-1}),
\end{aligned}
$$

so that, considering its ratio with $var_n({}_{(1)}\tilde{Y}_{n,t})$, yields

$$\lim_{t\to\infty} cov_n({}_{(1)}\tilde{y}_{n,t}, {}_{(1)}\tilde{y}_{n,t+h}) = \alpha_{(n)}^h, \ \ a.s.,$$

where $\alpha_{(1)} \leq ... \leq \alpha_{(n)}$ indicates the ordered statistics. In fact, $cov_n({}_{(1)}\tilde{y}_{n,t}, {}_{(1)}\tilde{y}_{n,t+h})$ can be expressed as a weighted mean, $1/n^4 \sum_{i,j,a,b=1}^n (\alpha_a\alpha_b)^{h/2} w_{i,j,a,b}(t)$ say, with $w_{(i),(j),(a),(b)}(t) \to 1_{i=j=a=b=n}$ a.s. as $t \to \infty$ when there are no repetitions, i.e. $\alpha_{(i_k-1)} < \alpha_{(i_k)}\, i_k = 1,..,n,\ k = i,j,a,b$. The case of repetitions easily follows. Under $I(1)$, $\alpha_{(n)}$ is a strongly consistent estimator for γ yielding, as $n \to \infty$,

$$\alpha_{(n)}^h \to 1, \ \ a.s.$$

The key factor that drives the result is the particular shape of the nonstationary 'innovations' $\prod_{j=1}^k u_{t-j}^2$. However, if we consider the different normalization such that $E(u_t^4) = 1$, this would imply $E(u_t^2) = \kappa$ (say) with $\kappa < 1$. As a consequence, a factor κ^h appears, due to $E(\prod_{j=0}^{h-1} u_{t+h-j}^2)$, requiring to impose $\gamma = 1/\kappa > 1$ in order to offset the former exponential term. However, when considering the other term of $cov_n({}_{(1)}\tilde{Y}_{n,t}, {}_{(1)}\tilde{Y}_{n,t+h})$ (involving the $\sum_{i,j,a,b=1}^n$), the same arguments just made apply yielding $\alpha_{(n)}^h \to (1/\kappa)^h$ a.s. as $n \to \infty$. Therefore, $\lim_{n\to\infty}\lim_{t\to\infty} cov_n({}_{(1)}\tilde{y}_{n,t}, {}_{(1)}\tilde{y}_{n,t+h}) = 1$ a.s..

Finally, consider the case when n goes to infinity before t. Then

$$\lim_{n\to\infty} cov_n({}_{(1)}\tilde{Y}_{n,t}, {}_{(1)}\tilde{Y}_{n,t+h}) = A(t;h), \ \ a.s.,$$

setting

$$A(t;h) = 3E\Bigg((\alpha_a\alpha_b)^{h/2}\Big(\frac{1-(3(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2})^t}{1-3(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2}} - \frac{1-(\alpha_i\alpha_j\alpha_a\alpha_b)^{t/2}}{1-(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2}}\Big)$$
$$+\frac{1}{1-(\alpha_i\alpha_j)^{1/2}}\left[\frac{1-(3(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2})^t}{1-3(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2}} - (\alpha_i\alpha_j)^{t/2}\frac{1-(3(\alpha_a\alpha_b)^{1/2})^t}{1-3(\alpha_a\alpha_b)^{1/2}}\right.$$
$$\left.-\frac{1-(\alpha_i\alpha_j\alpha_a\alpha_b)^{t/2}}{1-(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2}} + (\alpha_i\alpha_j)^{t/2}\frac{1-(\alpha_a\alpha_b)^{t/2}}{1-(\alpha_a\alpha_b)^{1/2}}\right]$$
$$+\frac{1}{1-(\alpha_a\alpha_b)^{1/2}}\left[\frac{1-(3(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2})^t}{1-3(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2}} - (\alpha_a\alpha_b)^{t/2}\frac{1-(3(\alpha_i\alpha_j)^{1/2})^t}{1-3(\alpha_i\alpha_j)^{1/2}}\right.$$
$$\left.-\frac{1-(\alpha_i\alpha_j\alpha_a\alpha_b)^{t/2}}{1-(\alpha_i\alpha_j\alpha_a\alpha_b)^{1/2}} + (\alpha_a\alpha_b)^{t/2}\frac{1-(\alpha_i\alpha_j)^{t/2}}{1-(\alpha_i\alpha_j)^{1/2}}\right]$$
$$+\frac{2}{3}\frac{1-(\alpha_a\alpha_b)^{t/2}}{1-(\alpha_a\alpha_b)^{1/2}}\frac{1-(\alpha_i\alpha_j)^{t/2}}{1-(\alpha_i\alpha_j)^{1/2}}\Bigg).$$

Using

$$\frac{1-x^s}{1-x} \sim -\,s\,x^s \frac{log(x)}{1-x},\ x>0,\ s\to\infty, \tag{35}$$

by a Taylor expansion in t around 0 and $1-1/x \le log(x) \le x-1,\ \ (x>0)$ yields, as $t\to\infty$,

$$A(t;h) \sim c\,t^2\,3^{t+1}(\mu_{(h+t)/2}\,\mu_{t/2})^2,\ \ h=0,\pm1,...\,a.s.$$

for some constant $0<c<\infty$ (independent of h). Using (28), as $t\to\infty$,

$$A(t;h)/A(t;0) \to 1\ \ a.s.,$$

uniformly in $h=0,\pm1,....$ The same arguments apply to $cov_n({}_{(2)}\tilde{y}_{n,t},{}_{(2)}\tilde{y}_{n,t+h})$.
□

Proof of Theorem 8: Using (31)

$$\min[{}_{(1)}X_{n,t},{}_{(2)}X_{n,t}] \le {}^U X_{n,t} \le \max[{}_{(1)}X_{n,t},{}_{(2)}X_{n,t}],$$

where

$${}_{(1)}X_{n,t} = u_t\left(\sum_{k=0}^{\infty}(\frac{1}{n}\sum_{i=1}^{n}\omega_i^{1/2}\prod_{j=1}^{k}(\beta_i+\alpha_i u_{t-j}^2)^{1/2})^2\right)^{1/2},$$
$${}_{(2)}X_{n,t} = u_t\left(\sum_{k=0}^{\infty}\frac{1}{n}\sum_{i=1}^{n}\omega_i^{1/2}\prod_{j=1}^{k}(\beta_i+\alpha_i u_{t-j}^2)^{1/2}\right).$$

The first and the last inequalities in the first statement of the theorem, involving the ${}_{(i)}\tilde{X}_{n,t}\,(i=1,2)$, follows considering the nonlinear moving average representation of ARCH(∞) of Zaffaroni (1999, section 2) which, for heterogeneous GARCH(1,1), reduces to

$$\sigma_{i,t}^2 = \omega_i \sum_{l=0}^{\infty} N_{i,l}(t), \tag{36}$$

$$N_{i,l} = 1_{l=0} + 1_{l>0}\sum_{k=0}^{l}\alpha_i^k\beta_i^{l-k}\left(\widetilde{\sum}_{(k)}^{(l)} u_{t-j_1}^2 ... u_{t-j_1-...-j_k}^2\right), \tag{37}$$

setting

$$\widetilde{\sum}_{(k)}^{(l)} = 1_{k=0} + 1_{k>0}\sum_{j_1=1}^{l-k+1}\ \sum_{j_2=1}^{l-k+2-j_1} ... \sum_{j_k=1}^{l-j_1-..-j_{k-1}}.$$

For GARCH(1,1), (37) represents an alternative way of writing $\prod_{k=0}^{l}(\beta_i + \alpha_i u_{t-k}^2)$ (cf. Nelson (1990b, eq. (10))). However, (37) is more suitable to our purpose because it neatly separates the contribution of the coefficients and of the rescaled innovations.

Applying Hardy, Littlewood, and Polya (1964, Theorem 24 (2.11.2) and 27 (2.12.2)) yields

$$
\begin{aligned}
{}_{(1)}\tilde{X}_{n,t} &= u_t \left(\sum_{l=0}^{\infty} \sum_{k=0}^{l} \widetilde{\sum}_{(k)}^{(l)} u_{t-j_1}^2 ... u_{t-j_1-...-j_k}^2 \big(\frac{1}{n}\sum_{i=1}^{n} \omega_i^{1/2} \alpha_i^{k/2} \beta_i^{(l-k)/2}\big)^2 \right)^{1/2}, \\
{}_{(2)}\tilde{X}_{n,t} &= u_t \left(\sum_{l=0}^{\infty} \sum_{k=0}^{l} \widetilde{\sum}_{(k)}^{(l)} \mid u_{t-j_1} \mid ... \mid u_{t-j_1-...-j_k} \mid \big(\frac{1}{n}\sum_{i=1}^{n} \omega_i^{1/2} \alpha_i^{k/2} \beta_i^{(l-k)/2}\big) \right),
\end{aligned}
$$

and

$$
\begin{aligned}
{}_{(1)}\tilde{X}_{t} &= u_t \left(\sum_{l=0}^{\infty} \sum_{k=0}^{l} \widetilde{\sum}_{(k)}^{(l)} u_{t-j_1}^2 ... u_{t-j_1-...-j_k}^2 \big(E(\omega_i^{1/2} \alpha_i^{k/2} \beta_i^{(l-k)/2})\big)^2 \right)^{1/2}, \\
{}_{(2)}\tilde{X}_{t} &= u_t \left(\sum_{l=0}^{\infty} \sum_{k=0}^{l} \widetilde{\sum}_{(k)}^{(l)} \mid u_{t-j_1} \mid ... \mid u_{t-j_1-...-j_k} \mid E(\omega_i^{1/2} \alpha_i^{k/2} \beta_i^{(l-k)/2}) \right).
\end{aligned}
$$

(i) We follow the proof of Theorem 6. Applying (32) to the sequence $\{\omega_i^{1/2} \alpha_i^{k/2} \beta_i^{(l-k)/2}\}$ yields, as $n \to \infty$,

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n} \omega_i^{1/2} \alpha_i^{k/2} \beta_i^{(l-k)/2} - E(\omega^{1/2} \alpha^{k/2} \beta^{(l-k)/2})\right| \\
&= O_{a.s}\big(\big(\frac{\rho_1 log\, log\, n}{n}\big)^{1/2} \alpha_{max}^{k/2} \beta_{max}^{(l-k)/2} \, (k+1)^{-(b_\alpha+1)/2} \, (l-k)^{-(b_\beta+1)/2}\big),
\end{aligned}
$$

using (38). Note that $(\alpha_{max} + \beta_{max})^2 + 2\alpha_{max}^2 \leq \gamma$ (cf. (20)).

Considering the case of ${}_{(2)}\tilde{X}_{n,t}$, the result then follows from

$$
\begin{aligned}
& E_n \, |{}_{(2)}\tilde{X}_{n,t} - {}_{(2)}\tilde{X}_t| \\
=\; & O_{a.s.}\left(\big(\frac{log\, log\, n}{n}\big)^{1/2} \sum_{l=0}^{\infty} \Big[\sum_{k=0}^{l} \binom{l}{k} \alpha_{max}^{k/2} \, \delta_u^k \, \beta_{max}^{(k-l)/2} \, k^{-(b_\alpha+1)/2} \, (k-l)^{-(b_\beta+1)/2}\Big] \right) \\
=\; & O_{a.s.}\left(\big(\frac{log\, log\, n}{n}\big)^{1/2} \sum_{l=0}^{\infty} l^{-(min(b_\alpha, b_\beta)+1)/2} (\alpha_{max}^{1/2} \, \delta_u + \beta_{max}^{1/2})^l \right).
\end{aligned}
$$

Imposing IV(γ) is not sufficient per se, unless for small enough γ, as it implies that $0 \leq \beta_i \leq \gamma^{1/2}$ and $0 \leq \alpha_i \leq (-\beta_i + (3\,\gamma - 2\,\beta^2)^{1/2})/3$ without ensuring boundedness of the series. Following the proof of Theorem 6, the result follows for ${}_{(1)}\tilde{X}_{n,t}$ where the stronger condition $\alpha_{max}^{1/4}\delta_u + \beta_{max}^{1/4} < 1$ is needed (recall that, by Assumption III, $0 \leq \alpha_{max} + \beta_{max} \leq 1$). Finally, note that when $\alpha_{max}^{i/4}\delta_u + \beta_{max}^{i/4} = 1\,(i = 1,2)$ then the result follows when $min[b_\alpha, b_\beta] > 4/i - 1\,(i = 1,2)$. These cases do not seem relevant and therefore are not commented in the statement of the theorem.

Boundedness, strict stationarity and ergodicity easily follows by using the same arguments of the proof of Theorem 6, using (33) with $\gamma = 1$ and, as $u \to \infty$,

$$E(\alpha^u) \sim c\,(\alpha_{max})^u\,u^{-(b_\alpha+1)},\;\; E(\beta^u) \sim c'\,(\beta_{max})^u\,u^{-(b_\beta+1)}, \qquad (38)$$

for $0 < c, c' < \infty$, from (28) (cf. remark IV.3 and B.1), with

$$\sum_{k=0}^{l}\binom{l}{k}\alpha_{max}^{k/i}\,e^{\lambda_u k/2i}\,\beta_{max}^{(k-l)/i}k^{-2(b_\alpha+1)/i}\,(k-l)^{-2(b_\beta+1)/i}$$
$$\leq ck^{-2(min(b_\alpha,b_\beta)+1)/i}((\alpha_{max}e^{\lambda_u/2})^{1/i} + \beta_{max}^{1/i})^k\;(i = 1,2).$$

for some $0 < c < \infty$, using $\widetilde{\sum}_{(k)}^{(l)} 1 = \binom{l}{k}$. Note that, due to the greater complexity of GARCH(1,1), the required conditions are much stronger for ${}_{(2)}\tilde{X}_t$ than for ${}_{(1)}\tilde{X}_t$.

(ii) Set $\omega_i = 1\,(i = 1,..,n)$ as this is completely innocuous. Consider ${}_{(1)}X_{n,t}$ and set $\delta_i(t) = (\beta_i + \alpha_i u_t^2)^{1/2}$. Then for any integer $u > 0$

$$cov({}_{(1)}X_{n,t}^2, {}_{(1)}X_{n,t+u}^2) = A + B,$$

with

$$A = A_1 + A_2,$$
$$A_1 = E(u_t^2)\frac{1}{n^4}\sum_{i,ja,b=1}^{n} E(\prod_{s=1}^{u-1}\delta_a(t+u-s)\delta_b(t+u-s))$$
$$\times\sum_{k=u}^{\infty}\left[cov(u_t^2, \delta_a(t)\delta_b(t))E(\prod_{r=1}^{k}\delta_i(t-r)\delta_j(t-r))E(\prod_{h=u+1}^{k}\delta_a(t+u-h)\delta_b(t+u-h))\right],$$
$$A_2 = E(u_t^2)\frac{1}{n^4}\sum_{i,j,a,b=1}^{n} E(\prod_{s=1}^{u-1}\delta_a(t+u-s)\delta_b(t+u-s))$$

$$\times \sum_{k=u}^{\infty} \left[E(u_t^2 \delta_a(t)\delta_b(t)) cov(\prod_{r=1}^{k} \delta_i(t-r)\delta_j(t-r), \prod_{h=u+1}^{k} \delta_a(t+u-h)\delta_b(t+u-h)) \right],$$

and

$$B = B_1 + B_2,$$
$$B_1 = E(u_t^2) \sum_{\substack{k_1 > k_2 \\ k_2 \geq u}}^{\infty} \frac{1}{n^4} \sum_{i,j,a,b=1}^{n} cov(u_t^2 \prod_{r=1}^{k_1} \delta_i(t-r)\delta_j(t-r), \prod_{s=1}^{k_2} \delta_a(t+u-s)\delta_b(t+u-s)),$$
$$B_2 = E(u_t^2) \sum_{\substack{k_2 > k_1 \\ k_2 \geq u}}^{\infty} \frac{1}{n^4} \sum_{i,j,a,b=1}^{n} cov(u_t^2 \prod_{r=1}^{k_1} \delta_i(t-r)\delta_j(t-r), \prod_{s=1}^{k_2} \delta_a(t+u-s)\delta_b(t+u-s)),$$

where for A we have used the cumulants' theorem. Note that the ACF of ${}_{(1)}X_{n,t}^2$ is nonnegative, given that $cov((\beta_i + \alpha_i u_t^2)^{1/2}, (\beta_j + \alpha_j u_t^2)^{1/2}) > 0$ for any i, j, and using

$$cov(\prod_{i=1}^{m} C_i, \prod_{i=1}^{m} D_i) = \sum_{k=1}^{m} \prod_{j=1}^{k-1} E(C_j D_j) cov(C_k, D_k) E(\prod_{j=k+1}^{m} C_j) E(\prod_{j=k+1}^{m} D_j),$$

for sequence of independent r.v.'s $\{C_i, D_i\}$.

By Schwarz inequality for some $0 < c < \infty$

$$A_1 \leq c \left(\frac{1}{n^4} \sum_{i,j,a,b=1}^{n} \frac{(\pi_i \pi_j \pi_a \pi_b)^{u/2}}{1 - (\pi_i \pi_j \pi_a \pi_b)^{1/2}} \right),$$
$$A_2 \leq c \left(\frac{1}{n^4} \sum_{i,j,a,b=1}^{n} \frac{(\pi_i \pi_j \pi_a \pi_b)^{u/2}}{1 - (\nu_i \nu_j \nu_a \nu_b)^{1/4}} \right).$$

For B_1 and B_2, using the cumulants' theorem, for some $0 < c, C < \infty$,

$$B_1 \leq \frac{c}{n^4} \sum_{k_1 > k_2 \geq u}^{\infty} \sum_{a,b,i,j=1}^{n} (\pi_a \pi_b)^{u/2} \Big(cov(u_t^2, \delta_a(t)\delta_b(t)) \, (\pi_i \pi_j)^{k_1/2} (\pi_a \pi_b)^{(k_2 - u)/2}$$
$$+ E(u_t^2 \delta_a(t)\delta_b(t)) \, (\pi_i \pi_j)^{(k_1 - k_2 + u)/2} (\nu_i \nu_j)^{(k_2 - u)/4} (\nu_a \nu_b)^{(k_2 - u)/4} \Big)$$
$$\leq \frac{C}{n^4} \sum_{a,b,i,j=1}^{n} \frac{(\pi_i \pi_j \pi_a \pi_b)^{u/2}}{1 - (\pi_i \pi_j)^{1/2}} \left(\frac{1}{1 - (\pi_a \pi_b \pi_i \pi_j)^{1/2}} + \frac{1}{1 - (\nu_i \nu_j \nu_a \nu_b)^{1/4}} \right).$$

$$B_2 \leq \frac{c}{n^4} \sum_{a,b,i,j=1}^{n} \pi_a^{u/2} \pi_b^{u/2} ($$

$$cov(u_t^2, \delta_a(t)\delta_b(t)) \left[\sum_{k_2 \geq u}^{\infty} \sum_{k_1=0}^{u-1} (\pi_i\pi_j)^{k_1/2} (\pi_a\pi_b)^{(k_2-u)/2} + \sum_{k_2>k_1\geq u}^{\infty} (\pi_i\pi_j)^{k_1/2} (\pi_a\pi_b)^{(k_2-u)/2} \right]$$

$$+E(u_t^2 \delta_a(t)\delta_b(t)) \times$$

$$\left[\sum_{k_1=0}^{u-1} \sum_{k_2=0}^{k_1} (\pi_i\pi_j)^{(k_1-k_2)/2} (\nu_i\nu_j\nu_a\nu_b)^{k_2/4} + \sum_{k_1=0}^{u-1} \sum_{k_2=k_1+1}^{\infty} (\pi_a\pi_b)^{(k_2-k_1)/2} (\nu_i\nu_j\nu_a\nu_b)^{k_1/4} \right.$$

$$\left. + \sum_{k_1=u}^{\infty} \sum_{k_2=k_1+1}^{k_1+u} (\pi_i\pi_j)^{(k_1+u-k_2)/2} (\nu_i\nu_j\nu_a\nu_b)^{(k_2-u)/4} + \sum_{k_1=u}^{\infty} \sum_{k_2=k_1+1+u}^{\infty} (\pi_a\pi_b)^{(k_2-u-k_1)/2} (\nu_i\nu_j\nu_a\nu_b)^{k_1/4} \right] \Bigg)$$

$$\leq \frac{C}{n^4} \sum_{a,b,i,j=1}^{n} (\pi_a\pi_b)^{u/2} \left(\left[\frac{1}{1-(\pi_i\pi_j)^{1/2}} \frac{1}{1-(\pi_a\pi_b)^{1/2}} + \frac{(\pi_i\pi_i)^{u/2}}{1-(\pi_a\pi_b\pi_i\pi_i)^{1/2}} \frac{1}{1-(\pi_a\pi_b)^{1/2}} \right] \right.$$

$$+ \left[\frac{1}{1-(\nu_i\nu_j\nu_a\nu_b)^{1/4}} \left(\frac{1}{1-(\pi_a\pi_b)^{1/2}} + \frac{1}{1-(\pi_i\pi_j)^{1/2}} \right) \right.$$

$$\left. \left. + \frac{(\nu_i\nu_j\nu_a\nu_b)^{u/4}}{1-(\nu_i\nu_j\nu_a\nu_b)^{1/4}} \left(\frac{1}{1-(\pi_a\pi_b)^{1/2}} + \frac{(\pi_a\pi_b)^{1/2}}{(\nu_i\nu_j\nu_a\nu_b)^{1/4}-(\pi_a\pi_b)^{1/2}} \right) \right] \right).$$

Considering the dominating term (in B_2), yields

$$cov({}_{(1)}X_{n,t}^2, {}_{(1)}X_{n,t+u}^2) \leq C \frac{1}{n^4} \sum_{a,b,i,j=1}^{n} \frac{\pi_a^{u/2}\pi_b^{u/2}}{1-(\nu_i\nu_j\nu_a\nu_b)^{1/4}}, \qquad (39)$$

for some $0 < C < \infty$. When $\gamma < 1$ the limit (as $n \to \infty$) of (39) behaves as $(E(\pi_i^{u/2}))^2$, when $u \to \infty$, and thus the result follows applying (28). Recalling that, in the limit, we need to characterize the behaviour near 1 of the ν_i, we use $\pi_i \sim (\alpha_{max} + \beta_{max})\,\nu_i$, as $\nu_i \to 1^-$, when $\gamma = 1$. Lemma 2, with $p = 4$, $r = 2$ concludes.

The same arguments (with tedious calculations) apply to the ACF of ${}_{(2)}X_t^2$, using

$${}_{(2)}X_{n,t}^2 = {}_{(1)}X_{n,t}^2 + u_t^2 \sum_{\substack{k_1>k_2\\=0}}^{\infty} \frac{1}{n^2} \sum_{i,j=1}^{n} \prod_{r=1}^{k_1} \prod_{s=1}^{k_2} \delta_i(t-r)\delta_j(t-s)$$

□

7 Appendix B

In this section we give a set of sufficient conditions on couples of r.v.'s, independent of each other, yielding a convolution with a distribution of the type described in Assumption $I\,(1)$. The following result is certainly known but we could not find a reference.

Lemma 3 *Let α and β be absolutely continuous r.v.'s, independent of each other, defined respectively in the interval $[0,k]$ and $[0,1-k]$ for some constant $0<k<1$. Let us assume that their density functions, labelled $g_\alpha(.)$ and $g_\beta(.)$, satisfy*

$$
\begin{aligned}
&g_\alpha(x) \sim c_\alpha(k-x)^{b_\alpha} \ \ as\ x \to k^-,\\
&g_\beta(y) \sim c_\beta(1-k-y)^{b_\beta} \ \ as\ y \to (1-k)^-,\\
&\ and\ either\\
&\frac{dg_\beta(y)}{dy} \sim c'_\beta(1-k-y)^{b_\beta-1} \ \ as\ y \to (1-k)^-,\\
&\ or\\
&\frac{dg_\alpha(y)}{dy} \sim c'_\alpha(k-y)^{b_\alpha-1} \ \ as\ y \to k^-,\\
&\ with\ bounded\ constants\ c_\alpha>0,\ c_\beta>0,\ |\ c'_\alpha\ |>0,\ |\ c'_\beta\ |>0,\ b_\alpha>-1,\ b_\beta>-1,
\end{aligned}
$$

and the $g_\alpha(\cdot)$, $g_\beta(\cdot)$ and either $dg_\alpha(y)/dy$ or $dg_\beta(y)/dy$ are all bounded everywhere else in their support. Then the r.v. $\pi=\alpha+\beta$, convolution of α and β, is absolutely continuous with probability density $g_\pi(.)$, defined on the support $[0,1]$, satisfying

$$g_\pi(z) \sim c_\pi(1-z)^{b_\pi} \ \ as\ z \to 1^-,$$

with

$$0<c_\pi<\infty \ and\ b_\pi = b_\alpha + b_\beta + 1,$$

and bounded in the interval $[0,1)$.

Proof: The density function of π is given by the following convolution:

$$g_\pi(z) = \int_{-\infty}^{\infty} g_\alpha(z-s)1_{(0\le z-s\le k)}g_\beta(s)1_{(0\le s\le (1-k))}ds.$$

As $z \to 1^-$,

$$g_\pi(z) \sim c \int_{z-k}^{1-k} (k+s-z)^{b_\alpha}\,(1-(k+s))^{b_\beta} ds = c \int_z^1 (u-z)^{b_\alpha}\,(1-u)^{b_\beta} du$$
$$= c\,(1-z)^{b_\alpha} \int_z^1 (1-u)^{b_\beta} du - c b_\alpha \int_z^1 (1-u)^{b_\beta+1} (\tilde{u}-z)^{b_\alpha-1} du, \qquad (40)$$

where the second equality is obtained from the mean value theorem with $\tilde{u} = \theta u + (1-\theta)$ for some $\theta \in (0,1)$. By Schwarz inequality, when $b_\alpha \neq 1/2$, for the second integral in (40), as $z \to 1^-$,

$$\int_z^1 (1-u)^{b_\beta+1} (\tilde{u}-z)^{b_\alpha-1} du \leq \left(\frac{1}{2b_\beta+3} (1-z)^{2b_\beta+3} \right)^{1/2} \left(\frac{1}{\theta} \int_{\theta z+(1-\theta)}^1 (\tilde{u}-z)^{2b_\alpha-2} d\tilde{u} \right)^{1/2}$$
$$\sim c(1-z)^{b_\alpha+b_\beta+1} \left(\frac{1-(1-\theta)^{2b_\alpha-1}}{(2b_\alpha-1)\theta} \right)^{1/2} \sim C(1-z)^{b_\alpha+b_\beta+1},$$

for some $0 < c, C < \infty$, using $\frac{1-(1-\theta)^a}{\theta} \sim a$, as $\theta \to 0^+$, for any real a. The same applies when $b_\beta = 1/2$ using $-\frac{log(1-\theta)}{\theta} \sim 1$, as $\theta \to 0^+$. Thus, the first integral in (40) dominates. When $z < 1$ (strictly) or when $z < min(1-k, k)$ the above expressions are bounded. □

Remark B.1 By (28), as $u \to \infty$,

$$E(\alpha^u) \sim c\, k^u u^{-(b_\alpha+1)}, \;\; E(\beta^u) \sim c'\,(1-k)^u u^{-(b_\beta+1)}, \;\; E(\pi^u) \sim c''\, u^{-(b_\pi+1)},$$

for $0 < c, c', c'' < \infty$.

Remark B.2 Lemma 3 implies the familiar results whereby, when both $0 > b_\alpha > -1/2$ and $0 > b_\beta > -1/2$, the density function $g_\pi(.)$ is continuous even if $g_\alpha(.)$ and $g_\beta(.)$ are not.

Remark B.3 The result can be easily generalized to the case of m r.v.'s, all independent of each other. If each of them has support $[0, c_j]$, $j = 1, \ldots, m$ and b_j denotes the exponent at c_j of the jth r.v., then when all the m r.v.'s satisfy the assumptions of Lemma 3 their convolution will have a density function with support $[0, \sum_{j=1}^m c_j]$ and with exponent $\sum_{j=1}^m b_j + (m-1)$ at $\sum_{j=1}^m c_j$.

References

ANDERSEN, T., AND T. BOLLERLSEV (1997): "Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns," *Journal of Finance*, 52, 975–1005.

BAILLIE, R. T., T. BOLLERSLEV, AND H. O. A. MIKKELSEN (1996): "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 74/1, 3–30.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures.* New York: Wiley.

BOLLERLSEV, T., R. CHOU, AND K. KRONER (1992): "ARCH modeling in finance: a review of the theory and empirical evidence," *Journal of Econometrics*, 52, 5–59.

BOLLERSLEV, T. (1986): "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 302–327.

BOLLERSLEV, T., R. ENGLE, AND D. NELSON (1995): "ARCH models," in *Handbook of Econometrics vol IV*, ed. by R. Engle,and D. McFadden. Amsterdam: North Holland.

BROCKWELL, P., AND R. DAVIS (1987): *Time series: theory and methods.* New York (NY): Springer Verlag.

CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): "Arbitrage, factor structure, and mean-variance analysis on large asset markets," *Econometrica*, 51, 1281–1304.

CHRISTOFFERSEN, P., AND F. DIEBOLD (1997): "How relevant is volatility forecasting for financial risk management," *Preprint* .

CORRADI, V. (1999): "Reconsidering the continuous time limit of the GARCH(1,1) limit," forthcoming *Journal of Econometrics* .

DAVIS, R., AND T. MIKOSCH (1998): "The sample autocorrelations of heavy-tailed processes with applications to ARCH," *Annals of Statistics*, 26, 2049–2080.

Diebold, F. (1988): *Empirical modeling of exchange rates.* New York (NY): Springer Verlag.

Diebold, F., and M. Nerlove (1989): "The dynamics of exchange rate volatility: a multivariate latent factor ARCH model," *Journal of Applied Econometrics*, 4, 1–21.

Ding, Z., and C. Granger (1996): "Modeling volatility persistence of speculative returns: a new approach," *Journal of Econometrics*, 73, 185–215.

Ding, Z., C. W. J. Granger, and R. F. Engle (1993): "A long memory property of stock market and a new model," *Journal of Empirical Finance*, 1, 83–106.

Drost, F., and T. Nijman (1993): "Temporal aggregation of GARCH processes," *Econometrica*, 61/4, 909–927.

Drost, F., and B. Werker (1996): "Closing the GARCH gap: continuous time GARCH modeling," *Journal of Econometrics*, 74, 31–57.

Dudley, R. (1989): *Real analysis and probability.* California: Pacific Grove.

Engle, R. (1987): "Multivariate ARCH with factor structures-cointegration in variance," *Preprint* .

Engle, R. F. (1982): "Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom," *Econometrica*, 50, 987–1007.

Gallant, A., P. Rossi, and G. Tauchen (1993): "Nonlinear dynamic structure," *Econometrica*, 61/4, 871–907.

Gradshteyn, I., and I. Ryzhik (1994): *Table of integrals, series and products.* San Diego: Academic Press, fifth edn.

Granger, C. (1980): "Long memory relationships and the aggregation of dynamic models," *Journal of Econometrics*, 14, 227–238.

HANSEN, L., AND S. RICHARD (1987): "The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models," *Econometrica*, 55, 587–614.

HARDY, G., J. LITTLEWOOD, AND G. POLYA (1964): *Inequalities.* Cambridge: Cambridge University Press.

LEONOV, V., AND A. SHIRYAEV (1959): "On a method of calculation of semi-invariants," *Theory of Probability Applications*, 4, 319–329.

LEPAGE, R., M. WOODROOFE, AND J. ZINN (1981): "Convergence to a stable distribution via order statistics," *The Annals of Probability*, 9/4, 624–632.

LIPPI, M., AND P. ZAFFARONI (1999): "Contemporaneous Aggregation of Linear Dynamic Models in Large Economies," *Preprint* .

MEDDAHI, N., AND E. RENAULT (1998): "Aggregation and marginalization of GARCH and Stochastic Volatility models," *Preprint* .

NELSON, D. (1990a): "ARCH models as diffusion approximation," *Journal of Econometrics*, 45, 7–39.

NELSON, D. (1990b): "Stationarity and persistence in the GARCH(1,1) model," *Econometric Theory*, 6, 318–334.

NIJMAN, T., AND E. SENTANA (1996): "Marginalization and contemporaneous aggregation in multivariate GARCH processes," *Journal of Econometrics*, 71, 71–87.

ROBINSON, P. M. (1978): "Statistical inference for a random coefficient autoregressive model," *Scandinavian Journal of Statistics*, 5, 163–168.

ROBINSON, P. M. (1991): "Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression," *Journal of Econometrics*, 47, 67–84.

ROBINSON, P. M., AND P. ZAFFARONI (1997): "Modelling nonlinearity and long memory in time series," in *Nonlinear Dynamics and Time Series*, ed. by C. Cutler,and D. Kaplan. Providence (NJ): American Mathematical Society.

ROYDEN, H. (1980): *Real analysis.* London: Macmillan.

SAMORODNITSKY, G., AND M. TAQQU (1994): *Stable non-Gaussian Processes: Stochastic Models with Infinite Variance.* New York, London: Chapman and Hall.

SENTANA, E. (1998): "The relation between conditionally heteroskedastic factor models and factor GARCH models," *Econometrics Journal,* 1, 1–9.

STOUT, W. (1974): *Almost sure convergence.* Academic Press.

YONG, C. H. (1974): *Asymptotic Behaviour of Trigonometric Series.* Hong Kong: Chinese University of Hong Kong.

ZAFFARONI, P. (1999): "Strong and weak stationarity of ARCH(∞) models," *Preprint* .

ZYGMUND, A. (1977): *Trigonometric series.* Cambridge: Cambridge University Press.

PAOLO ZAFFARONI
Servizio Studi
Banca d'Italia
Via Nazionale, 91
00184 Roma ITALY
zaffaroni.paolo@insedia.interbusiness.it