THE PREDICTION DISTRIBUTION OF A GARCH(1,1) PROCESS[1,2]

KARIM M. ABADIR, ALESSANDRA LUATI AND PAOLO PARUOLO

DECEMBER 21, 2018

This paper derives the analytic form of the h-step ahead prediction-density of a GARCH(1,1) process under Gaussian innovations, with a possibly asymmetric news impact curve. The analytic form of the density is novel and improves on current methods based on approximations and simulations. The explicit form of the density permits to compute tail probabilities and functionals, such as expected shortfall, that measure risk when the underlying asset return is generated by a GARCH(1,1). The prediction densities are derived for any finite prediction horizon h. For the stationary case, as h increases the prediction density converges to a distribution with Pareto tails which whose form has been already described in the literature. The formulae in the paper characterize the degree of non-gaussianity of the prediction distribution, and the distance between the tails of the finite horizon prediction distribution and the ones of the stationary distribution.

PRELIMINARY: December 21, 2018

KEYWORDS: GARCH(1,1), GJR, prediction density, Pareto tails.

1. INTRODUCTION

Since their introduction in Engle (1982) and Bollerslev (1986), Generalised AutoRegressive Conditional Heteroskedasticity (GARCH) processes have been widely employed in financial econometrics, see e.g. Bollerslev, Russell, and Watson (2010). In their original formulation, the conditional distribution of innovations was typically assumed to be Gaussian.

Empirically, the distribution of stock returns has been studied extensively under the random walk assumption, see e.g. Fama (1965); in this literature, Gaussianity of stock returns has been questioned as too thin-tailed when compared to its empirical counterpart. Gaus-

Karim Abadir. Business School, Imperial College London, UK; e-mail: Email: k.m.abadir@imperial.ac.uk

Alessandra Luati. Department of Statistical Sciences “Paolo Fortunati”, Via Belle Arti 41, University of Bologna, Italy; Email: alessandra.luati@unibo.it

Corresponding author: Paolo Paruolo. European Commission Joint Research Centre, Via E. Fermi 2749, I-3027 Ispra (VA), Italy; Phone: +39 0332785642, Email: paolo.paruolo@ec.europa.eu

[1]JEL classification: C22, C53.

[2]The authors acknowledge useful comments from Enrique Sentana as well as from participants to the Seventh Italian Congress of Econometrics and Empirical Economics 2017, EC^2 2017 in Amsterdam and the NBER-NSF Time Series Conference 2018.

sian GARCH processes can generate uncorrelated, heteroskedastic returns with a stationary distribution with fatter tails than the Gaussian.

GARCH processes can include several lags q of the past squared shocks and several lags p of the past volatility; in practice, however, the GARCH(1,1) model with $p = q = 1$ is often found to offer a good fit for asset returns, and it is usually preferred to GARCH models with more parameters, see Tsay (2010) section 3.5, or Andersen, Bollerslev, Christoffersen, and Diebold (2006), section 3.6. Moreover, many multivariate GARCH models are built on the univariate GARCH(1,1), see e.g. Engle, Ledoit, and Wolf (2017) and references therein. In this sense the GARCH(1,1) is both the prototype and the workhorse of GARCH processes in practice.

GARCH processes map shocks, i.e. news, into the conditional volatility; the function obtained by replacing past conditional volatilities with unconditional ones was called by Engle and Ng (1993) the news-impact-curve (NIC). For GARCH(1,1) processes, this curve yields the same value of volatility for positive and negative shocks, i.e. it is symmetric. Glosten, Jagannanthan, and Runkle (1993) (henceforth GJR) extended the GARCH setup to allow for asymmetric news impact curve responses to negative shocks.

Many measures of risk are functions of the prediction density of asset returns. These measures include the Value at Risk, which is a quantile of the prediction distribution of the asset return, see Jorion (2006), as well as the Expected Shortfall, see Patton, Ziegel, and Chen (2017). The latter is the expected value of the prediction distribution of the asset return in the left tail, between minus infinity and the Value at Risk; this measure has been recently re-emphasised by the Third Basel Accords. Both measures are functionals of the prediction distribution of asset returns, see Arvanitis, Hallam, Post, and Topaloglou (2018).

The prediction distribution of a GARCH(1,1) hence plays an important role for the computation of risk measures in financial applications. This distribution is not known in analytic form beyond the one-step-ahead prediction distribution, which is given by the assumption on innovations used to build the process, see e.g. Andersen, Bollerslev, Christoffersen, and Diebold (2006), page 811.

The unknown analytic form of the prediction density of a GARCH has led econometricians to look for alternative approximate solutions. Alexander, Lazar, and Stanescu (2013) have resorted to approximations based on the first 4 moments of the prediction distribution; see also Baillie and Bollerslev (1992). They use the Cornish-Fisher expansion and the Johnson SU distribution with the same 4 moments.

An alternative to this approach is to simulate from the prediction distribution and to estimate it non-parametrically, e.g. by kernel methods. While consistent, this estimator has the slower rate of convergence typical of kernel density estimators. More recently Delaigle, Meister, and Rombouts (2016) have proposed a non-parametric root-T consistent estimator of the stationary distribution of the (log-)volatility process. Despite a better convergence rate, the non-parametric estimation of the density requires computing time, and does not lead to exact results.

The tail behavior of the stationary distribution of the GARCH(1,1) has been studied extensively, see Mikosch and Starica (2000) and Davis and Mikosch (2009). The tails of the stationary distribution of both the volatility and of the GARCH process x_t are of Pareto type, $\Pr(x_t > u) \approx cu^{-2\kappa}$ say. These properties are based on results for random difference equations and renewal theory obtained in Kesten (1973) and Goldie (1991).

The tail index κ is associated with the number of moments of the stationary distribution, which exist up to order 2κ. Larger values of κ are associated with thinner tails of the stationary distribution; this is interpreted here to mean that the larger the number of moments (i.e. the larger κ) the smaller the distance from the Gaussian distribution, which has an infinite number of moments. κ depends on the coefficient α and β of the GARCH(1,1) process $x_t = \sigma_t\varepsilon_t$, $\sigma_t^2 = \omega + \alpha x_{t-1}^2 + \beta\sigma_{t-1}^2$, as well as on the type of the one-step-ahead distribution. Examples of values of the tail index are given in Davis and Mikosch (2009).

The present paper derives the analytical form of the h-step-ahead prediction density of a GARCH(1,1), allowing for the GJR type with asymmetric NIC. Closed form expressions are given for the prediction density of a GARCH(1,1) process x_t for Gaussian innovations. The results are obtained by marginalizing the joint density of the prediction observations, using integration and special functions, for any prediction horizon $h = 1, 2, \dots$.[1] The formulae are valid for stationary as well as non-stationary GARCH(1,1) processes.

In the case of 2-steps-ahead, the prediction distribution is obtained without imposing constrains on the values of the α and β coefficients. For the h-steps-ahead prediction distribution with $h \geq 3$, a condition on β is required to guarantee integrability of a certain integral; a sufficient condition for this to be satisfied is to have β larger than 0.62, which is a condition often satisfied in practice.

The prediction density is found to be close to a Gaussian density (with appropriate variance) for high values of β/α, and far from it for low values of it. Similarly, large values of β/α are found to be associated to higher values of κ, i.e. smaller distance from the Gaussian distribution for the stationary distribution with Pareto tails $\Pr(x_t > u) \approx cu^{-2\kappa}$.

The rest of the paper is organised as follows. Section 2 describes the general approach for the derivation of the integral. Section 3 states main results, while Section 4 discusses the form of the prediction density when compared with the tails of the stationary distribution. Section 5 concludes. The Appendix contains proofs.

2. THE PREDICTION DENSITY

This section illustrates the construction used to characterise the prediction density as an integral, involving (a product of several copies of) the chosen density of innovations. Consider the asymmetric GARCH(1,1)

$$x_t = \sigma_t\varepsilon_t, \qquad \sigma_t^2 = \omega + \alpha_{t-1}x_{t-1}^2 + \beta\sigma_{t-1}^2, \qquad \alpha_t := \alpha + \lambda 1_{x_t<0} = \alpha + \frac{\lambda}{2}(1 - \varsigma_t) \tag{2.1}$$

[1]The one-step ahead distribution for $h = 1$ is given by construction of the process.

where $\omega, \alpha, \beta > 0$, $\lambda \geq 0$ and $1_{x_t<0} = \frac{1}{2}(1 - \varsigma_t)$ is the indicator function for event $x_t < 0$, and $\varsigma_t := \text{sgn}(\varepsilon_t) = \text{sgn}(x_t)$ is the sign of ε or x_t; these signs are the same because $\sigma_t > 0$. The sequence $\{\varepsilon_t\}$ is assumed to be i.i.d., centered around zero and with Gaussian p.d.f. $f_\varepsilon(\epsilon) := g(\epsilon^2) := (2\pi)^{-\frac{1}{2}} \exp(-\epsilon^2/2)$.

Time $t = 0$ is taken to be the starting time of the forecasts, and it is assumed that one wishes to predict x_h for some $h = 1, 2, 3, \ldots$, conditional on information set at time $t = 0$, taken to consist of observations of x_0 and σ_0. This information set is consistent with observing x_t from minus infinity to time 0 under stationarity. Note also that, because x_0 and σ_0^2 are observed, also σ_1^2 is observed.

Throughout the paper the values taken by the random variables x_t, $z_t := x_t^2$, σ_t^2 are denoted u_t, w_t and s_t^2 respectively, and sometimes the subscript t is omitted if this does not cause ambiguity. The next Lemma reports consequences of the symmetry of the one-step-ahead density g on relevant conditional p.d.f.s. In the Lemma, the following notation is used, $\boldsymbol{z} := (z_1, \ldots, z_{h-1})'$, $\boldsymbol{\varsigma} := (\varsigma_1, \ldots, \varsigma_{h-1})'$; here $\boldsymbol{w} := (w_1, \ldots, w_{h-1})'$, $\boldsymbol{s} := (s_1, \ldots, s_{h-1})'$ denote values of $\boldsymbol{z}$ and $\boldsymbol{\varsigma}$.

LEMMA 2.1 (Densities) *For symmetric $f_\epsilon(e) = g(\epsilon^2)$, $f_{x_t}(\cdot)$ is symmetric, i.e. $f_{x_t}(u) = f_{x_t}(-u)$, $u \in \mathbb{R}$, and it is given by*

$$f_{x_t}(u) = f_{z_t}(u^2)\,|u|\,. \tag{2.2}$$

Moreover, $\Pr(\varsigma_t = \pm 1) = \frac{1}{2}$ *and one has*

$$f_{\boldsymbol{z},z_h|\boldsymbol{\varsigma}}(\boldsymbol{w}, w_h|\boldsymbol{s}) = \prod_{t=1}^{h} \left(w_t\sigma_t^2\right)^{-\frac{1}{2}} g\left(\frac{w_t}{\sigma_t^2}\right) \tag{2.3}$$

where σ_t^2 depends on w_{t-j} (the value of $z_{t-j} = x_{t-j}^2$) and s_{t-j} (the sign of x_{t-j}) for $j = 1, \ldots, t-1$ via (2.1).

Denote the set of all possible $\boldsymbol{\varsigma}$ by $\mathcal{S}$, $\#\mathcal{S} = 2^{h-1}$. Densities are first computed conditionally on $\boldsymbol{\varsigma}$ and later they are marginalized with respect to it. Here, conditioning on $\boldsymbol{\varsigma}$ is relevant only for the GJR case $\lambda \neq 0$.

The basic building block is given by the expression in (2.3). This density can be marginalised with respect to $\boldsymbol{z}$ as follows

$$f_{z_h|\boldsymbol{\varsigma}}(w_h|\boldsymbol{s}) = \int_{\mathbb{R}_+^{h-1}} f_{\boldsymbol{z},z_h|\boldsymbol{\varsigma}}(\boldsymbol{w}, w_h|\boldsymbol{s})\mathrm{d}\boldsymbol{w}. \tag{2.4}$$

Finally, the conditioning with respect to the signs $\boldsymbol{\varsigma}$ is averaged across different configurations, using the mutual independence of the signs ς_{t-j} and the fact that $\Pr(\varsigma_t = \pm 1) = \frac{1}{2}$ for all t, thanks to the symmetry of g. One hence finds

$$f_{z_h}(w_h) = \sum_{\boldsymbol{s}} f_{z_h|\boldsymbol{\varsigma}}(w_h|\boldsymbol{s})\Pr(\boldsymbol{s}) = 2^{-h+1} \sum_{\boldsymbol{s}} f_{z_h|\boldsymbol{\varsigma}}(w_h|\boldsymbol{s}) \tag{2.5}$$

where the sum $\sum_{\boldsymbol{s}}$ is over $s_j \in \{-1,1\}$, for $j = 1,\ldots,h-1$. The prediction density is hence found by combining (2.5), (2.4), (2.3), (2.2).

The next Lemma reports a recursion for the volatility process, that turns out to be useful when solving the integral in (2.4). In the Lemma, for $t = 1,\ldots,h-1$, let $y_t := \alpha_t z_t/(\beta\sigma_t^2) = \alpha_t x_t^2/(\beta\sigma_t^2) = \alpha_t\varepsilon_t^2/\beta$ and $\boldsymbol{y} := (y_1,\ldots,y_{h-1})'$, where $\boldsymbol{v} := (v_1,\ldots,v_{h-1})'$ denotes a value of $\boldsymbol{y}$.

Lemma 2.2 (Volatility and transformations) *The volatility process can also be written*

$$\sigma_t^2 = \omega + \beta(1+y_{t-1})\sigma_{t-1}^2 \qquad y_t := \frac{\alpha_t}{\beta}\varepsilon_t^2.$$

For $h \geq 2$, σ_h^2 has the following recursive expression in terms of y's

$$\begin{aligned}\sigma_h^2 &= \omega + (1+y_{h-1})\,\beta\left\{\omega + (1+y_{h-2})\,\beta\left(\ldots\left(\omega + (1+y_1)\,\beta\sigma_1^2\right)\right)\right\}\\ &= \omega + (1+y_{h-1})\left\{\omega\beta + (1+y_{h-2})\left(\ldots\left(\omega\beta^{h-2} + (1+y_1)\,\beta^{h-1}\sigma_1^2\right)\right)\right\}\end{aligned} \tag{2.6}$$

with $\sigma_1^2 = \omega + \beta\sigma_0^2 + \alpha_0 x_0^2$, which is measurable with respect to the information set at time 0. *Moreover, one has*

$$f_{z_h|\boldsymbol{\varsigma}}(w_h|\boldsymbol{s}) = \left(\frac{\gamma_h}{w_h}\right)^{\frac{1}{2}}\int_{\mathbb{R}_+^{h-1}}\prod_{t=1}^{h-1}\left(v_t^{-\frac{1}{2}}g\left(\frac{\beta}{\alpha_t}v_t\right)\right)\cdot\sigma_h^{-1}g\left(\frac{w_h}{\sigma_h^2}\right)\mathrm{d}\boldsymbol{v}, \tag{2.7}$$

where $\gamma_h := \beta^{h-1}/(\prod_{t=1}^{h-1}\alpha_t)$.

3. Main results

The main results are summarised in Theorem 3.2 below. Before stating the main theorems, an auxiliary assumption is introduced. Define $\theta := \omega/2\sigma_1^2$ with $0 < \theta \leq \frac{1}{2}$ and $\underline{\beta} := \underline{\beta}(\theta) := -\theta + \sqrt{\theta^2 + 2\theta}$.

Assumption 3.1

a. *For $h = 3$, let $\beta \geq \underline{\beta}$;*
b. *For $h > 3$ let $\beta \geq \max(\frac{1}{2}, \underline{\beta})$.*

It can be noted that $\sup_\theta \underline{\beta}(\theta) = \lim_{\theta\to\frac{1}{2}}\underline{\beta}(\theta) = \frac{-1+\sqrt{5}}{2} \approx 0.618\,03$, as $\sigma_1^2 > \omega$. In Figure 1, the area above the curve represents the set $\beta \geq \max(\frac{1}{2}, \underline{\beta})$ for $0 < \theta \leq \frac{1}{2}$.

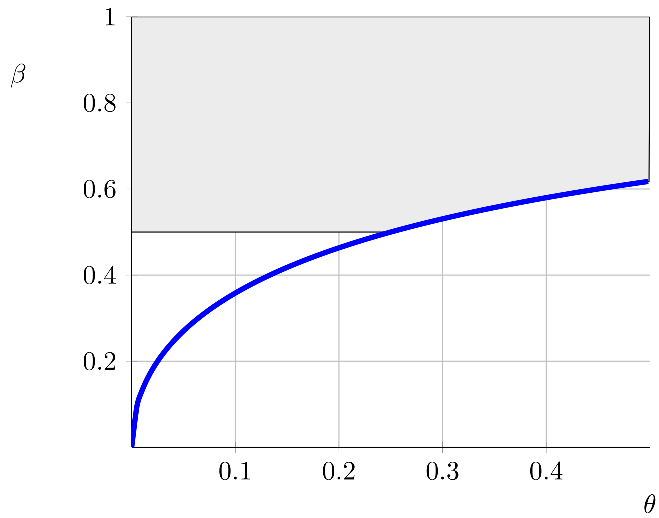
In Theorem 3.2 below, Ψ is the confluent hypergeometric function of the second kind, also known as Tricomi function, see Abadir (1999) and Gradshteyn and Ryzhik (2007), section 9.21, whose integral representation is,

$$\Psi(a,c,z) = \frac{1}{\Gamma(a)}\int_{\mathbb{R}_+}\exp\left(-zt\right)t^{a-1}\left(1+t\right)^{c-a-1}\mathrm{d}t \tag{3.1}$$

with $\mathrm{Re}(a) > 0, \mathrm{Re}(c) > 0$. Moreover, the following notation is used in summations:

$$b_r := \begin{cases} r & r \in \mathbb{N}\\ \infty & r \notin \mathbb{N}\end{cases}$$

FIGURE 1.— $\underline{\beta}$ as a function of θ. Blue line: $\underline{\beta} := \underline{\beta}(\theta) := -\theta + \sqrt{\theta^2 + 2\theta}$. Shaded area: region $\beta \geq \max(\frac{1}{2}, \underline{\beta})$, see Assumption 3.1.



THEOREM 3.2 (GARCH(1,1) prediction density) *Assume that ε_t are i.i.d. $N(0,1)$ and let Assumption 3.1 hold; then one has, for $h \geq 2$,*

$$
\begin{aligned}
f_{z_h}(w_h) &= (2\pi)^{-\frac{h}{2}} \sum_{j=0}^{\infty} \frac{(-2)^{-j}}{j!} w_h^{j-\frac{1}{2}} c_j, \qquad (3.2)\\
f_{x_h}(u_h) &= f_{z_h}(u_h^2)\,|u_h|\,,
\end{aligned}
$$

where $c_j := 2^{-h+1} \sum_{\boldsymbol{s}\in\mathcal{S}} c_{-\frac{1}{2}-j,\boldsymbol{s}}$,

$$
\begin{aligned}
c_{r,\boldsymbol{s}} &:= \sum_{k_1=0}^{b_r} \sum_{k_2=0}^{b_r-k_1} \cdots \sum_{k_{h-2}=0}^{b_r-K_{h-3}} \prod_{t=1}^{h-2} \binom{r-K_{t-1}}{k_t} \omega^{K_{h-2}} \beta^{S_{h-2}} \left(\left(\omega+\beta\sigma_1^2\right)\beta^{h-2}\right)^{r-K_{h-2}+\frac{1}{2}} .\\
&\cdot \left(\sigma_1^2 \alpha_{h-1}\alpha_t\right)^{-\frac{1}{2}} \pi^{\frac{h-1}{2}} p_t\left(r, \frac{\beta}{2\alpha_{h-t}}\right) p_{h-2}\left(r, \frac{\omega+\beta\sigma_1^2}{2\alpha_1\sigma_1^2}\right) \qquad (3.3)
\end{aligned}
$$

where

$$
p_t(r,u) := \Psi\left(\frac{1}{2}, r+\frac{3}{2}-K_t; u\right)
$$

with $\boldsymbol{\varsigma} := (\varsigma_1, \ldots, \varsigma_{h-1})'$, $K_t := \sum_{i=1}^{t} k_i$ *and* $S_t = \sum_{i=1}^{t}(i-1)k_i$, *and empty sums (respectively products) are understood to be equal to* 0 *(respectively equal to* 1*). Recall finally that also* α_t *in* (3.3) *is a function of* $\boldsymbol{s}$.

PROOF: See Appendix. □

Note that in the case when $h = 2$, equation (3.2) holds for any value of β, while for $h = 3$ it holds if and only if $\beta \geq \underline{\beta}$. For $h > 3$, the validity of the (3.2) is guaranteed by the sufficient condition $\beta \geq \max(\frac{1}{2}, \underline{\beta})$, which is, however, not necessary.

The line of proof of Theorem 3.2 is the following: for $h = 2$ the integral is solved by substitution and by using equation (3.1). For $h \geq 3$, subsequent (negative) binomial expansions of expression (2.6) for σ_t^2 are required, whose validity is ensured by the inequality

$$\omega\left(1-\sum_{i=1}^{h-1}\beta^i\right)\le\beta^h\sigma_1^2,$$

which is satisfied under Assumption 3.1, see Lemma 5.1 in the Appendix.

Immediate consequences of Theorem 3.2 are collected in the following corollary.

COROLLARY 3.3 (C.d.f. and moments) *The prediction c.d.f.s of z_h and x_h are given by*

$$F_{z_h}(w_h) = (2\pi)^{-\frac{h}{2}}\sum_{j=0}^{\infty}\frac{(-2)^{-j}}{j!\left(j+\frac{1}{2}\right)}w_h^{j+\frac{1}{2}}c_j,$$

$$F_{x_h}(u_h) = \begin{cases} (2\pi)^{-\frac{h}{2}}\sum_{j=0}^{\infty}\frac{(-2)^{-j}}{j!(2j+1)}u_h^{2j+1}c_j & u_h\ge 0\\ 1-F_{x_h}(-u_h) & u_h<0\end{cases},$$

with moments

$$\mathrm{E}(x_h^{2m})=\mathrm{E}(z_h^m)=2^{m+\frac{3}{2}-\frac{3}{2}h}\pi^{-\frac{h}{2}}\Gamma\left(m+\frac{1}{2}\right)\sum_{\boldsymbol{s}\in\mathcal{S}}c_{m,\boldsymbol{s}},\qquad m=1,2,\ldots$$

where c_j, $c_{m,\boldsymbol{s}}$ are defined in (3.3).

Note that $c_{m,\boldsymbol{s}}$ in the moments calculations are made of finite sums extending to m, involving the Tricomi functions, which do not fall in the logarithmic case as in Theorem 3.2; see Abadir (1999) for the logarithmic case. In fact, $m-k\in\{0,1,\ldots,m\}$ implies that

$$\begin{aligned}\Psi\left(\frac{1}{2};\frac{3}{2}+m-k;\xi\right) &= \frac{\Gamma\left(\frac{1}{2}+m-k\right)}{\sqrt{\pi}}\xi^{-\frac{1}{2}-m+k}\,{}_1F_1\left(-m+k;\frac{1}{2}-m+k;\xi\right)\\ &= \frac{\Gamma\left(m+\frac{1}{2}\right)}{\sqrt{\pi}k!\binom{m-\frac{1}{2}}{k}}\xi^{-\frac{1}{2}-m+k}\sum_{j=0}^{m-k}\frac{\binom{m-k}{j}}{\binom{-\frac{1}{2}+m-k}{j}}\frac{\xi^j}{j!}\end{aligned}$$

is a finite sum.

Some standardised densities of x_h and the corresponding right tails are plotted in Fig. 2 for $h=1,2,3,4$. The curve $h=1$ is the standard Gaussian. Figure 3 shows the predictive densities for $h=2$ and values of β/α that range from to 8.5 ($\alpha=0.1,\beta=0.85$) to 1/8.5 ($\alpha=0.85,\beta=0.1$). Figure 4 shows the tails for asymmetric news impact curves.

One can see that the prediction densities are more similar to a Gaussian when β/α is large.

4. STATIONARY DISTRIBUTION

The limit representation of the random variable x_h in the stationary case can be found in Francq and Zakoian (2010) Theorem 2.1 page 24. The tail behaviour of the limit distribution is reviewed in Mikosch and Starica (2000) and Davis and Mikosch (2009). The tails of the stationary distribution of both the volatility and of the GARCH process x_t are of Pareto type, $\Pr(x_t>u)\approx cu^{-2\kappa}$ say, where $\kappa>0$ is a tail index. These properties are based on results for random difference equations and renewal theory obtained in Kesten (1973) and Goldie (1991).

FIGURE 2.— Prediction densities $f_{x_h}(u_h)$ (left panel) and zoom of the right tails (right panel) for standardised x_h, $h = 1, 2, 3, 4$, $\omega = 0.1, \alpha = 0.1, \beta = 0.7, \sigma_0^2 = 1; x_0^2 = 1, \lambda = 0$. Computations performed in Mathematica.

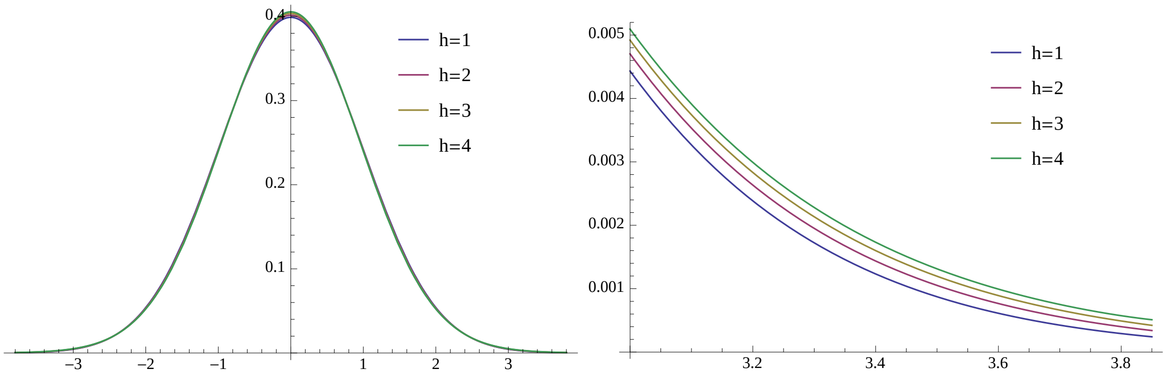
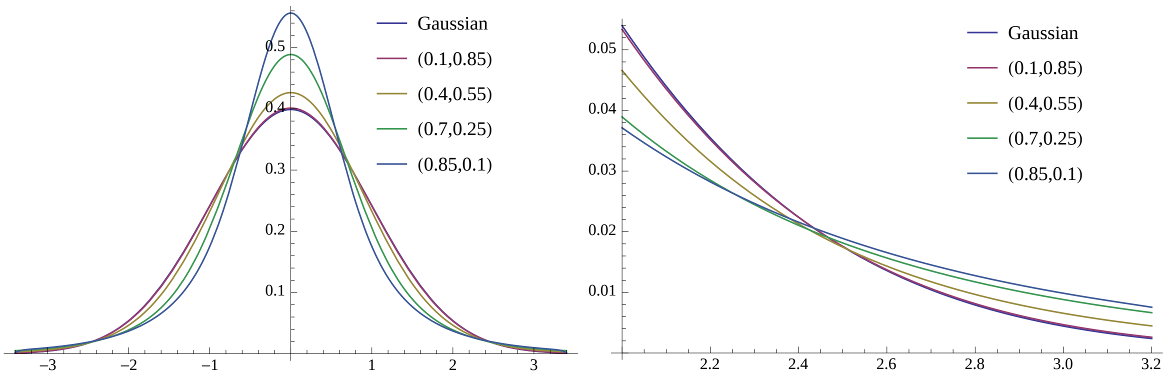


FIGURE 3.— Prediction density $f_{x_2}(u_2)$ for standardised x_2, $\omega = 0.1, \sigma_0^2 = 1; x_0^2 = 1$ varying values of (α, β)



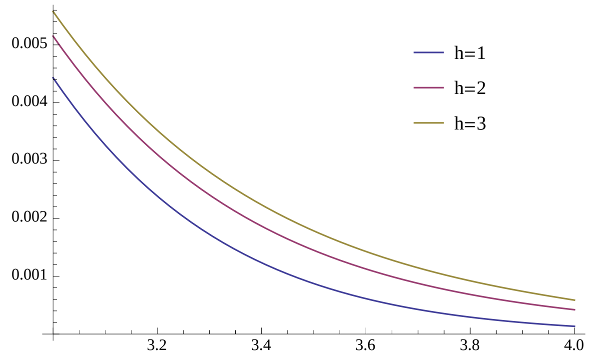
The tail index of the stationary distribution depends on the coefficient α and β of the GARCH(1,1) process x_t as well as on the one-step-ahead distribution. Examples of the tail index are given in Davis and Mikosch (2009); for Gaussian innovations, $\kappa = 14.1$ for $\alpha = \beta = 0.1$, while $\kappa = 1$ for $\alpha = 1 - \beta$.

The index κ is the unique solution of $\mathrm{E}((\alpha\varepsilon_t^2 + \beta)^\kappa) = 1$. When κ is an integer, the expression simplifies to

$$1 = \mathrm{E}((\alpha\varepsilon_t^2 + \beta)^\kappa) = \beta^\kappa \sum_{n=0}^{\kappa} \binom{\kappa}{n} \left(\frac{\beta}{\alpha}\right)^n \mathrm{E}(\varepsilon_t^{2n}), \tag{4.1}$$

see Davis and Mikosch (2009) eq. (10). Substituting the moments $\mathrm{E}(\varepsilon_t^{2n})$ from the χ^2 distribution, and assigning values to β/α over a grid of pre-specified values, one can solve (4.1) for β, and hence for $\alpha = (\beta/\alpha)^{-1}\beta$. This allows to compute (values of) the surface $\kappa(\alpha, \beta)$. Figure 5 reports the level curves of $\kappa(\alpha, \beta)$ as a function of α and β obtained in this way. The figure also reports the lines where β/α is constant. It is seen that, for large values of β/α, κ and β/α increase roughly together. This association is not present for small values of β/α.

FIGURE 4.— Right tail of $f_{x_h}(u_h)$ for standardised x_h, $h = 1, 2, 3$, in blue, red and green respectively, $\omega = 0.25, \alpha = 0.1, \beta = 0.7, \sigma_0^2 = 1; x_0^2 = 1, \lambda = 0.2$ ($h = 1$ is standard Gaussian)



The relation between β/α and fat-tailedness of the prediction density for finite horizon h can be illustrated using the case $h = 2$. From Theorem 3.2,

$$f_{x_2}(u_2) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}_2} \sum_{j=0}^{\infty} \frac{1}{j!} \left(-\frac{1}{2} \frac{u_2^2}{\tilde{\sigma}_2^2} \right)^j \sqrt{z} \Psi \left(\frac{1}{2}, 1 - j; z \right)$$

where $\tilde{\sigma}_2^2 = \omega + \beta\sigma_1^2$ and[2]

$$z := \frac{\omega + \beta\sigma_1^2}{2\alpha\sigma_1^2} = \frac{\tilde{\sigma}_2^2}{2\alpha\sigma_1^2} = \frac{1}{2} \left(\frac{\omega}{\beta\sigma_1^2} + 1 \right) \frac{\beta}{\alpha}.$$

Hence when $\beta/\alpha \to \infty$ one has $z \to \infty$ with $\sqrt{z}\Psi \left(\frac{1}{2}, 1 - j; z\right) = 1 + O(|z|^{-1})$, see Abramowitz and Stegun (1964), eq. 13.1.8, so that all the Tricomi functions Ψ_j, for varying j, tend to one.[3] As a result, when $\beta/\alpha \to \infty$ the prediction distribution converges to a $\mathrm{N}(0, \tilde{\sigma}_2^2)$.

Hence in both the case of the prediction density for $h = 2$ and the stationary distribution, the fat tailedness of the distributions is small for large values of β/α.

5. CONCLUSIONS

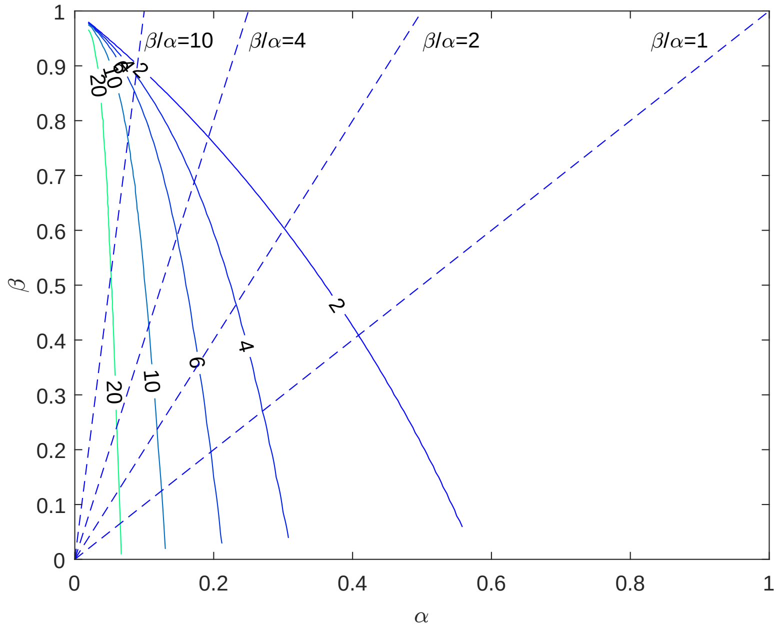
This paper presents the analytical form of the prediction density of a GARCH(1,1) process. This can be used to evaluate the probability of tail events or of quantities that may be of interest for value at risk calculations. This improves on approximation methods based on moments, or on Monte Carlo simulation and estimation.

The techniques in this paper can ge applied also with symmetric innovations density $g(\cdot)$ different from the N(0,1) one. Different densities imply distinct subsequent (negative) binomial expansions of expression (2.6) for σ_t^2, and different auxiliary convergence conditions on the GARCH coefficients, similarly to Assumption 3.1.

[2]The quantity $\tilde{\sigma}_2^2 := \omega + \beta\sigma_1^2$ can be interpreted as the minimum value that $\sigma_2^2 = \omega + (1 + y_1)\beta\sigma_1^2$ can take, in the ideal case when $\alpha = 0$ (thus $y_1 = 0$) and σ_1^2 is given, i.e. $x_2 \sim \mathrm{N}(0, \tilde{\sigma}_2^2)$.

[3]This is unlike in the case for fixed z where the sequence of Ψ_j is decreasing from 1 to 0 for increasing j.

FIGURE 5.— Level curves of κ as a function of α and β in the Gaussian case. Dashed lines represent loci where β/α is constant.



REFERENCES

ABADIR, K. M. (1999) An introduction to hypergeometric functions for economists. *Econometric Reviews*, 18(3), 287–330.

ABRAMOWITZ, M., AND I. STEGUN (1964) *Handbook of mathematical functions*. National Bureau of Standards, Applied Mathematics.

ALEXANDER, C., E. LAZAR, AND S. STANESCU (2013) Forecasting VaR using analytic higher moments for GARCH processes. *International Review of Financial Analysis*, 30, 36 – 45.

ANDERSEN, T., T. BOLLERSLEV, P. F. CHRISTOFFERSEN, AND F. X. DIEBOLD (2006) Volatility and correlation forecasting. in *Handbook of Economic Forecasting, Volume 1*, ed. by G. Elliott, C. W. Granger, and A. Timmermann. Elsevier.

ARVANITIS, S., M. HALLAM, T. POST, AND N. TOPALOGLOU (2018) Stochastic Spanning. *Journal of Business & Economic Statistics*, 0(0), 1–13.

BAILLIE, T. R., AND T. BOLLERSLEV (1992) Prediction in dynamic models with time-dependent conditional variances. *Journal of Econometrics*, 52, 91–113.

BOLLERSLEV, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, 307–327.

BOLLERSLEV, T., J. R. RUSSELL, AND M. W. E. WATSON (2010) *Volatility and time series econometrics: essays in honor of Robert F. Engle*. Oxford University Press.

DAVIS, R., AND T. MIKOSCH (2009) Extreme Value Theory for GARCH Processes. in *Handbook of Financial Time Series*, ed. by T. Andersen, R. Davis, J.-P. Kreiss, and T. Mikosch, pp. 187–200. Springer.

DELAIGLE, A., A. MEISTER, AND J. ROMBOUTS (2016) Root- consistent density estimation in {GARCH} models. *Journal of Econometrics*, 192(1), 55 – 63.

ENGLE, R. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kindom inflation. *Econometrica*, 11, 122–150.

ENGLE, R., AND V. NG (1993) Measuring and Testing the Impact of News on Volatility. *Journal of Finance*, 48, 1749–1778.

Engle, R. F., O. Ledoit, and M. Wolf (2017) Large Dynamic Covariance Matrices. *Journal of Business & Economic Statistics*, 0(0), 1–13.

Fama, E. F. (1965) The Behavior of Stock-Market Prices. *The Journal of Business*, 38(1), 34–105.

Francq, C., and J.-M. Zakoian (2010) *GARCH models*. Wiley.

Glosten, L., R. Jagannanthan, and D. Runkle (1993) On the relation between expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48, 1779–1802.

Goldie, C. M. (1991) Implicit Renewal Theory and Tails of Solutions of Random Equations. *Annals of Applied Probability*, 1, 126–166.

Gradshteyn, I., and I. Ryzhik (2007) *Book of Tables of integrals, series, and products*. 7th ed., Academic Press.

Jorion, P. (2006) *Value at Risk - The New Benchmark for Managing Financial Risk*. McGraw Hill, New York.

Kesten, H. (1973) Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131, 207–248.

Mikosch, T., and C. Starica (2000) Limit Theory for the Sample Autocorrelations and Extremes of a GARCH (1,1) process. *Annals of Statistics*, 28(5), 1427–1451.

Mood, A. M., F. A. Graybill, and D. C. Boes (1974) *Introduction to the Theory of Statistics, 3rd Edition*. Mc Graw-Hill.

Patton, A. J., J. F. Ziegel, and R. Chen (2017) Dynamic Semiparametric Models for Expected Shortfall (and Value-at-Risk). *ArXiv*, https://arxiv.org/abs/1707.05108v1.

Tsay, R. S. (2010) *Analysis of financial time series*. Wiley, 3rd edn.

APPENDIX

Proof of Lemma 2.1: Consider the transformation theorem for $z_h = x_h^2$; from standard results, see e.g. Mood, Graybill, and Boes (1974), page 201, Example 19, one has

$$f_{z_h}(w_h) = \left(\frac{1}{2}\frac{1}{\sqrt{w_h}} f_{x_h}(-\sqrt{w_h}) + \frac{1}{2}\frac{1}{\sqrt{w_h}} f_{x_h}(\sqrt{w_h})\right) 1_{w_h \geq 0}.$$

where 1_A is the indicator function of the event A. Because, by symmetry, one has $f_{x_h}(-\sqrt{w_h}) = f_{x_h}(\sqrt{w_h})$, the expression in the previous display simplifies to $f_{z_h}(w_h) = w_h^{-\frac{1}{2}} f_{x_h}(\sqrt{w_h}) 1_{(w_h \geq 0)}$, or, letting u_h indicate $w_h^{\frac{1}{2}}$, and solving for $f_{x_h}(u_h)$, one finds $f_{x_h}(u_h) = |u_h| \, f_{z_h}(u_h^2)$, which is (2.2). Note that the expression with the absolute value is also valid for $u_h = -\sqrt{w_h}$. This proves (2.2).

Eq. (2.3) follows from definitions. □

Proof of Lemma 2.2: Consider $f_{\boldsymbol{z},z_h|\boldsymbol{\varsigma}}(\boldsymbol{w}, w_h|\boldsymbol{s})$ from (2.3), and consider the transformation of from $\boldsymbol{z}$ to $\boldsymbol{y}$. Observe that the domain of integration remains $\mathbb{R}_+^{h-1}$, that the inverse transformation is $z_t = \beta\sigma_t^2 y_t/\alpha_t$, with Jacobian $\gamma_h \prod_{t=1}^{h-1} \sigma_t^2$, where $\gamma_h := \beta^{h-1}/(\prod_{t=1}^{h-1} \alpha_t)$. Hence one finds

$$f_{\boldsymbol{y},z_h|\boldsymbol{\varsigma}}(\boldsymbol{v}, w_h|\boldsymbol{s}) = 2^{-h}\gamma_h^{\frac{1}{2}} \prod_{t=1}^{h-1}\left(v_t^{-\frac{1}{2}} g\left(\frac{\beta}{\alpha_t} v_t\right)\right) \cdot \left(w_h\sigma_h^2\right)^{-\frac{1}{2}} g\left(\frac{w_h}{\sigma_h^2}\right)$$

from which (2.7) follows, as in (2.4). □

Proof of Theorem 3.2

The proof of Theorem 3.2 is based on the following Lemmas 5.1 and 2.1.

LEMMA 5.1 (Conditions on β) *Let Assumption 3.1.b hold. Then, for any $j \geq 2$*

$$\omega\left(1-\sum_{i=1}^{j-1}\beta^{i}\right)\leq\beta^{j}\sigma_1^2. \tag{5.1}$$

PROOF: of Lemma 5.1. For $j=2$ the inequality (5.1) reads $\beta^2\sigma_1^2+\omega\beta-\omega\geq 0$. Solving the quadratic on the l.h.s. for β one finds two roots, $\beta_1=(-\omega-\sqrt{\omega^2+4\omega\sigma_1^2})/(2\sigma_1^2)<0$ and $\underline{\beta}=(-\omega+\sqrt{\omega^2+4\omega\sigma_1^2})/(2\sigma_1^2)>0$, so that the quadratic is non-negative for $\beta<\beta_1$ or for $\beta>\underline{\beta}$. Because $\beta_1<0$, this holds only when $\beta\geq\underline{\beta}$. This proves that (5.1) is valid for $j=2$ for $\beta\geq\underline{\beta}$ and a fortiori also for $\beta\geq\max\{\frac{1}{2},\underline{\beta}\}$.

An induction approach is used for $j>2$. Assume that (5.1) is valid for some $j=j_0\geq 2$ and $\beta\geq\max\{\frac{1}{2},\underline{\beta}\}$; it can then be shown that (5.1) is valid also replacing j with $j+1$. To see this, take (5.1) for $j=j_0$ and multiply by β. One finds

$$\omega\left(\beta-\sum_{i=1}^{j_0-1}\beta^{i+1}\right)\leq\beta^{j_0+1}\sigma_1^2.$$

Because $\beta\geq\frac{1}{2}$, one has $\omega(1-\beta)\leq\omega\beta$, so that,

$$\omega\left(1-\beta-\sum_{i=1}^{j_0-1}\beta^{i+1}\right)\leq\omega\left(\beta-\sum_{i=1}^{j_0-1}\beta^{i+1}\right)\leq\beta^{j_0+1}\sigma_1^2.$$

Rearranging $1-\beta-\sum_{i=1}^{j_0-1}\beta^{i+1}$ as $1-\sum_{i=1}^{j_0}\beta^{i}$, one finds that (5.1) holds also for $j=j_0+1$. The induction step hence proves that (5.1) holds for any j if $\beta\geq\max\{\frac{1}{2},\underline{\beta}\}$. $\square$

LEMMA 5.2 (Coefficients c_j) *Assume that (5.1) holds for $2\leq j\leq h$; then*

$$\gamma_h^{\frac{1}{2}}\int_{\mathbb{R}_+^{h-1}}\exp\left(-\frac{1}{2}\sum_{t=1}^{h-1}\frac{\beta}{\alpha_t}v_t\right)(\sigma_h^2)^{-\frac{1}{2}-j}\prod_{t=1}^{h-1}\frac{\mathrm{d}v_t}{\sqrt{v_t}} \tag{5.2}$$

equals $c_{-\frac{1}{2}-j,\boldsymbol{s}}$ as defined in (3.3).

PROOF OF LEMMA 5.2: Rewrite (5.2) setting $r=-\frac{1}{2}-j$ as

$$\gamma_h^{\frac{1}{2}}\int_{\mathbb{R}_+^{h-1}}\prod_{t=1}^{h-1}\exp\left(-\frac{1}{2}\frac{\beta}{\alpha_t}v_t\right)(\sigma_h^2)^{r}\frac{\mathrm{d}v_t}{\sqrt{v_t}}. \tag{5.3}$$

Using equation (2.6), for $h = 2$ this expression equals $\gamma_2^{\frac{1}{2}} A_2(r)$ where

$$\begin{aligned} A_h(r) &= \int_{\mathbb{R}_+} \exp\left(-\frac{1}{2}\frac{\beta}{\alpha_1}v_1\right)\left(\omega\beta^{h-2} + (1+v_1)\,\beta^{h-1}\sigma_1^2\right)^r v_1^{-\frac{1}{2}}\mathrm{d}v_1 \\ &= \left(\beta^{h-1}\sigma_1^2\right)^r \int_{\mathbb{R}_+} \exp\left(-\frac{1}{2}\frac{\beta}{\alpha_1}v_1\right)\left(\omega\beta^{-1}\sigma_1^{-2} + 1 + v_1\right)^r v_1^{-\frac{1}{2}}\mathrm{d}v_1 \\ &= \left(\beta^{h-1}\sigma_1^2\right)^r \int_{\mathbb{R}_+} \exp\left(-\frac{1}{2}\frac{\beta}{\alpha_1}bt\right)(b+bt)^r (bt)^{-\frac{1}{2}} b\mathrm{d}t \\ &= \left(\beta^{h-1}\sigma_1^2\right)^r b^{r+\frac{1}{2}} \int_{\mathbb{R}_+} \exp\left(-\frac{1}{2}\frac{\beta}{\alpha_1}bt\right)(1+t)^r t^{-\frac{1}{2}}\mathrm{d}t \end{aligned}$$

where $b := \frac{\omega}{\beta\sigma_1^2} + 1$, $t := \frac{v_1}{b}$, $0 < t < \infty$, $\mathrm{d}v_1 = b\mathrm{d}t$. Hence

$$A_h(r) = \left(\beta^{h-1}\sigma_1^2\right)^r b^{r+\frac{1}{2}}\sqrt{\pi}\,\Psi\left(\frac{1}{2}, r+\frac{3}{2}; \frac{1}{2}\frac{\beta}{\alpha_1}b\right)$$

which follows from equation (3.1). This shows that for $h = 2$, $\gamma_2^{\frac{1}{2}} A_2(-\frac{1}{2} - j) = c_{-\frac{1}{2}-j,s_1}$ for $h = 2$.

Next consider the case $h = 3$, where

$$\sigma_3^2 = \omega + (1+v_2)\left(\omega\beta + (1+v_1)\,\beta^2\sigma_1^2\right),$$

and one wishes to expand $(\sigma_3^2)^r$. Consider the inequality $\omega < \omega\beta + \beta^2\sigma_1^2$, and the associated quadratic equation $\beta^2\sigma_1^2 + \omega\beta - \omega = 0$ in β with solutions $\beta_1 < 0$ and $\underline{\beta}$ as in Assumption 1. One has that for $\beta \geq \underline{\beta}$ one finds $\beta^2\sigma_1^2 + \omega\beta - \omega > 0$, which ensure that $\omega < \omega\beta + \beta^2\sigma_1^2$. Hence for $\beta \geq \underline{\beta}$ one can expand $(\sigma_3^2)^r$ as

$$\left(\sigma_3^2\right)^r = \sum_{j=0}^{\infty}\binom{r}{j}\omega^j (1+v_2)^{r-j}\left(\omega\beta + (1+v_1)\,\beta^2\sigma_1^2\right)^{r-j}.$$

Similarly, for case $h > 3$, one can write (2.6) as

$$\sigma_h^2 = \omega\beta^0 + (1+v_{h-1})\,a_1 \tag{5.4}$$

using the following recursions

$$\begin{aligned} a_{h-2} &:= \omega\beta^{h-2} + (1+v_1)\,\beta^{h-1}\sigma_1^2 \\ a_{h-j} &:= \omega\beta^{h-j} + (1+v_{j-1})\,a_{h-j+1} \qquad j = 3,\ldots,h. \end{aligned} \tag{5.5}$$

In this notation a_{h-2} represents the terms in the inner-most parenthesis in (2.6), a_{h-3} the terms in the second inner-most parentheses in (2.6), etc, up to $\sigma_h^2 = a_0$. It can be shown that condition (5.1) implies that

$$\omega\beta^{h-j} \leq a_{h-j+1} \tag{5.6}$$

in (5.5); in order to prove this, one can start from $j = 2$ and proceed to show that this holds for $j = h$.

Let now $r=-\frac{1}{2}-j$ and apply subsequent binomial expansions to powers of $a_0, a_1, ..., a_{h-3}$ in $(\sigma_h^2)^r$ from (5.4) and (5.5) one finds

$$
\begin{aligned}
(\sigma_h^2)^r &= (\omega\beta^0+(1+v_{h-1})\,a_1)^r=\sum_{k_1=0}^{r}\binom{r}{k_1}\omega^{k_1}\,(1+v_{h-1})^{r-k_1}\,a_1^{r-k_1}\\
&=\sum_{k_1=0}^{r}\sum_{k_2=0}^{r-k_1}\binom{r}{k_1}\binom{r-k_1}{k_2}\omega^{k_1}\,(1+v_{h-1})^{r-k_1}\,\omega^{k_2}\beta^{k_2}\,(1+v_{h-2})^{r-k_1-k_2}\,{a_2}^{r-k_1-k_2}\\
&=\sum_{k_1=0}^{r}\sum_{k_2=0}^{r-k_1}\binom{r}{k_1}\binom{r-k_1}{k_2}\omega^{k_1+k_2}\beta^{k_2}\,(1+v_{h-1})^{r-k_1}\,(1+v_{h-2})^{r-k_1-k_2}\,{a_2}^{r-k_1-k_2}\\
&=\\
&=\sum_{k_1=0}^{r}\sum_{k_2=0}^{r-k_1}\cdots\sum_{k_{h-2}=0}^{r-K_{h-3}}\binom{r}{k_1}\binom{r-k_1}{k_2}\cdots\binom{r-K_{h-3}}{k_{h-2}}\omega^{K_{h-2}}\beta^{S_{h-2}}\cdot\\
&\qquad\cdot\prod_{t=1}^{h-2}(1+v_{h-t})^{r-K_t}\,a_{h-2}^{r-K_{h-2}},
\end{aligned}\tag{5.7}
$$

where the summations are extended to ∞ when r is not an integer; convergence of the series is guaranteed by (5.1). Substituting (5.7) in (5.3) and integrating, one finds

$$
\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\cdots\sum_{k_{h-2}=0}^{\infty}\binom{r}{k_1}\binom{r-k_1}{k_2}\cdots\binom{r-K_{h-3}}{k_{h-2}}\omega^{K_{h-2}}\beta^{S_{h-2}}\gamma_h^{\frac{1}{2}}\prod_{t=1}^{h-2}I_t\cdot A_h(r-K_{h-2})\tag{5.8}
$$

and

$$
I_t:=\int_{\mathbb{R}_+}\exp\left(-\frac{1}{2}\frac{\beta}{\alpha_{h-t}}v_{h-t}\right)(1+v_{h-t})^{r-K_t}\,v_{h-t}^{-\frac{1}{2}}\mathrm{d}v_{h-t}=\sqrt{\pi}\Psi\left(\frac{1}{2},r+\frac{3}{2}-K_t;\frac{1}{2}\frac{\beta}{\alpha_{h-t}}b\right).\tag{5.9}
$$

Finally, replacing $(\gamma_h)^{\frac{1}{2}}=(\beta^{h-1})^{\frac{1}{2}}(\prod_{t=1}^{h-2}\alpha_t\alpha_{h-1})^{-\frac{1}{2}}$ and $b=\frac{\omega+\beta\sigma_1^2}{\beta\sigma_1^2}$ in (5.8) and (5.9) and rearranging, one obtains $c_{r,\boldsymbol{s}}$ as in equation (3.3).

□

Next the proof of Theorem 3.2 is presented.

PROOF OF THEOREM 3.2: The integral to be solved is

$$
f_{z_h|\varsigma_h}(w_h|s_h)=\frac{\sqrt{\gamma_h/w_h}}{(2\pi)^{\frac{h}{2}}}\int_{\mathbb{R}_+^{h-1}}\exp\left(-\frac{1}{2}\left(\sum_{t=1}^{h-1}\frac{\beta}{\alpha_t}v_t+\frac{w_h}{\sigma_h^2}\right)\right)\sigma_h^{-1}\prod_{t=1}^{h-1}\frac{\mathrm{d}v_t}{\sqrt{v_t}}.\tag{5.10}
$$

Expand $\exp(-w_h/(2\sigma_h^2))=\sum_{j=0}^{\infty}\frac{(-w_h/2)^j}{j!}(\sigma_h^2)^{-j}$ and note that

$$
f_{z_h|\varsigma_h}(w_h|s_h)=\frac{\sqrt{w_h}}{(2\pi)^{\frac{h}{2}}}\sum_{j=0}^{\infty}\frac{(-w_h/2)^j}{j!}c_{-\frac{1}{2}-j,\boldsymbol{s}},
$$

where

$$c_{-\frac{1}{2}-j,\boldsymbol{s}} = \gamma_h^{\frac{1}{2}} \int_{\mathbb{R}_+^{h-1}} \exp\left(-\frac{1}{2}\sum_{t=1}^{h-1}\frac{\beta}{\alpha_t}v_t\right)(\sigma_h^2)^{-\frac{1}{2}-j}\prod_{t=1}^{h-1}\frac{\mathrm{d}v_t}{\sqrt{v_t}}$$

which, by Lemma 5.2, also equals the expression (3.3). Marginalizing with respect to $\boldsymbol{\varsigma}$, being all elements in $\mathcal{S}$ equally likely, one finds

$$\begin{aligned} f_{z_h}(w_h) &= 2^{-h+1}\sum_{\boldsymbol{s}\in\mathcal{S}} f_{z_h}(w_h|\boldsymbol{s}) = 2^{-h+1}\sum_{\boldsymbol{s}\in\mathcal{S}} 2^{-h}\frac{\sqrt{w_h}}{(2\pi)^{\frac{h}{2}}}\sum_{j=0}^{\infty}\frac{(-w_h/2)^j}{j!}c_{-\frac{1}{2}-j,\boldsymbol{s}} \\ &= 2^{-h}\frac{\sqrt{w_h}}{(2\pi)^{\frac{h}{2}}}\sum_{j=0}^{\infty}\frac{(-w_h/2)^j}{j!}\left(2^{-h+1}\sum_{\boldsymbol{s}\in\mathcal{S}} c_{-\frac{1}{2}-j,\boldsymbol{\varsigma}_{h-1}}\right) = 2^{-h}\frac{\sqrt{w_h}}{(2\pi)^{\frac{h}{2}}}\sum_{j=0}^{\infty}\frac{(-w_h/2)^j}{j!}c_j \end{aligned}$$

where $c_j := 2^{-h+1}\sum_{\boldsymbol{s}\in\mathcal{S}} c_{-\frac{1}{2}-j,\boldsymbol{s}}$. □

Proof of Corollary 3.3: The c.d.f is found by integrating termwise the p.d.f. The moments are derived as follows. From (5.10) one sees that

$$\mathrm{E}_{z_h|\varsigma_h}(w_h^m|s_h) = \frac{\sqrt{\gamma_h}}{(2\pi)^{\frac{h}{2}}}\int_{\mathbb{R}_+^h} w_h^{m-\frac{1}{2}}\exp\left(-\frac{1}{2}\frac{w_h}{\sigma_h^2}\right)\mathrm{d}w_h\prod_{t=1}^{h-1}\exp\left(-\frac{1}{2}\left(\frac{\beta}{\alpha_t}v_t\right)\right)\sigma_h^{-1}v_t^{-\frac{1}{2}}\mathrm{d}v_t.$$

Recall that

$$\int_{\mathbb{R}_+}\exp\left(-\frac{w}{2\sigma_h^2}\right)w^{m-\frac{1}{2}}\mathrm{d}w = \left(2\sigma_h^2\right)^{m+\frac{1}{2}}\Gamma\left(m+\frac{1}{2}\right)$$

so that

$$\mathrm{E}_{z_h|\varsigma_h}(w_h^m|s_h) = 2^{-\frac{h}{2}+m+\frac{1}{2}}\pi^{-\frac{h}{2}}\gamma_h^{\frac{1}{2}}\Gamma\left(m+\frac{1}{2}\right)\int_{\mathbb{R}_+^{h-1}}\prod_{t=1}^{h-1}\exp\left(-\frac{1}{2}\left(\frac{\beta}{\alpha_t}v_t\right)\right)\left(\sigma_h^2\right)^m v_t^{-\frac{1}{2}}\mathrm{d}v_t.$$

Proceeding as in (5.8) one finds

$$\mathrm{E}_{z_h|\varsigma_h}(w_h^m|s_h) = 2^{-\frac{h}{2}+m+\frac{1}{2}}\pi^{-\frac{h}{2}}\gamma_h^{\frac{1}{2}}\Gamma\left(m+\frac{1}{2}\right)c_{m,\boldsymbol{s}}$$

and hence

$$\mathrm{E}_{z_h}(w_h^m) = 2^{-\frac{3}{2}h+m+\frac{3}{2}}\pi^{-\frac{h}{2}}\gamma_h^{\frac{1}{2}}\Gamma\left(m+\frac{1}{2}\right)\sum_{\boldsymbol{s}\in\mathcal{S}} c_{m,\boldsymbol{s}}.$$

□