The limiting distribution of a nonstationary integer valued GARCH(1,1) process

Jon Michel[*]

April 17, 2018

Abstract

We consider the integer valued GARCH(1,1) process of Rydberg and Shephard (1999) defined by the two equation system $Y_n \overset{d}{\sim} \text{Poisson}(\lambda_n)$ and $\lambda_{n+1} = \omega + \alpha Y_n + \beta \lambda_n$. When $\alpha + \beta < 1$ this process has a stationary solution and properties are well understood. In this paper we find the limiting distribution of λ_n and Y_n for the case of $\alpha + \beta = 1$. Using this result, we show some implications for maximum likelihood estimation and nonstationarity testing.

1 Introduction

The integer-valued GARCH (INGARCH) was first introduced by Rydberg and Shephard (1999) as a model of count data to study the number of financial transactions occurring during a small time interval. It was motivated as a discrete approximation to a continuous time model. The model was later generalized by Ferland, Latour, and Oraichi (2006). There stationarity properties of the model were shown. In addition, the name INGARCH was first introduced there and this was motivated by the algebraic similarity between the INGARCH model and the GARCH model of Bollerslev (1986). Properties of this model, and nonlinear variations, were also studied in Fokianos, Rahbek, and Tjøstheim (2009). Stationarity and mixing properties for nonlinear versions were shown in Neumann (2011). Additionally,

---

[*]Department of Economics, Ohio State University, 304 Arps Hall, Columbus, OH 43210, USA, Email `michel.82@osu.edu`.

numerous extensions to this model have been studied, a few examples being Zhu (2011), Wang, Liu, Yao, Davis, and Li (2014), and Fokianos and Tjøstheim (2009). Further references to this model can also be found in the review paper Fokianos (2015).

In this paper we only consider the INGARCH(1,1) case, this model is defined by letting $\lambda_0 = \omega^1$ and for $n \geq 0$

$$Y_n \overset{d}{\sim} \text{Poisson}(\lambda_n) \tag{1}$$

and

$$\lambda_{n+1} = \omega + \alpha Y_n + \beta \lambda_n, \tag{2}$$

for $\omega, \alpha,$ and $\beta > 0$. In all of the previous literature only the case of $\alpha + \beta < 1$ has been considered. In that setting there exists a stationary solution which is β-mixing (see Neumann (2011) for this most general conditions). We will consider the case of $\alpha + \beta = 1$, here we will show that the process is nonstationary and we will show that the process $(Y_{[rN]}/N, \lambda_{[rN]}/N) \Rightarrow (V(r), V(r))$, where $V(r)$ is a stochastic process which satisfies the stochastic differential equation $dV(t) = \omega dt + \alpha \sqrt{V(t)} dW(t)$. To our knowledge, this stochastic process has not appeared elsewhere in the literature.

This result is in sharp contrast with the integrated GARCH (IGARCH) model of Engle and Bollerslev (1986) for which it has been shown (see Nelson (1990)) that for the case of $\alpha + \beta = 1$ the GARCH(1,1) model can have a stationary solution. This shows that while the GARCH(1,1) and INGARCH(1,1) are similarly behaved when $\alpha + \beta < 1$, they have very distinct behavior when $\alpha + \beta = 1$. The behavior is more akin to that of an AR(1) process where the model is stationary when the autoregressive parameter is less than 1 in absolute value, but nonstationary of order $\sqrt{N}$ when the autoregressive parameter is 1 (see, for example, Hamilton (1994) for details).

The techniques used in our proof, to our knowledge, have not appeared in the econometrics literature. We make use of semigroup methods, such as those found in Ethier and Kurtz (2009). Analysis of Markov processes by semigroups was introduced in the seminal papers of Feller (1939), (1951). These methods are mostly used in the analysis of Markov processes coming from stochastic population dynamics, such as problems relating to the

---

[1]Other conventions for λ_1 are sometimes chosen. Often λ_0 is instead selected so that the process (Y_n, λ_n) is stationary. For our results we only require that $\lambda_1 = O_P(1)$, and so we will set it to ω for clarity of exposition.

spread of epidemics or the genetic drift of a population; see Dawson (2017), for example. Other applications of these techniques have been to Markov processes that arise from chemical reactions; as an example, the introductory chapter of Ethier and Kurtz (2009) has an analysis for the Schlögl model of Schlögl (1972). Numerous other applications exist as well, see Ethier and Kurtz (2009) for examples.

Of these various applications, the critical Galton-Watson process, in particular, is most similar to our study of the nonstationary INGARCH(1,1). The critical Galton-Watson process behaves similarly to an INGARCH(1,0), or INARCH(1), model with $\omega = 0$. A Galton-Watson process is defined by the equation

$$N_{n+1} = \sum_{j=1}^{N_n} z_{j,n+1},$$

where $z_{j,n}$ are i.i.d. integer-valued random variables. The process is described as "critical" when $E z_{j,n} = 1$. In the critical case it can be shown (see Dawson (2017) or Ethier and Kurtz (2009)) that if $N_0 = [Nx]$, for some $x > 0$ then $N_{[rN]}/N \Rightarrow N(r)$, where $N(0) = x$ and $N(r)$ satsifies the stochatic differential equation $dN(t) = \sqrt{Var(z_{j,n})}\sqrt{N(t)}dW(t)$.

Additionally, using our weak convergence result we will consider some implications for the maximum likelihood estimation of the nonstationary integer-valued GARCH, nonstationarity testing for count data, and an application to count data coming from measles infections. Somewhat surprisingly, we find that the weak convergence result is not sufficient on its own to characterize the distribution of the maximum likelihood estimator. This is in contrast to other nonstationary settings, such as the unit root model, where the limit distribution is the main tool used when analyzing estimation.

2 Limit Distribution

We let (Y_n, λ_n) satisfy equations (1) and (2) with $\alpha + \beta = 1$. We then have the following results. The proof can be found in the appendix.

Theorem 2.1. *We let $V_N(r) = \lambda_{[rN]}/N$. Then,*

$$V_N(r) \Rightarrow V(r),$$

where $V(0) = 0$ and $V(r)$ satisfies the stochastic differential equation $dV(r) = \omega dt + \alpha \sqrt{V(r)} dW(r)$.

This stochastic process is similar to that of the stochastic differential equation for the Galton-Watson process, except that there is now the additional drift term ωdt.

In addition, since $\lambda_{n+1} = \omega + \alpha Y_n + \beta \lambda_n$ we have that

$$Y_{[rN]} = \alpha^{-1}(\lambda_{[rN]+1} - \beta \lambda_{[rN]} - \omega).$$

Therefore, Theorem 2.1 along with the fact that $\alpha + \beta = 1$ implies the followng result.

Corollary 2.2. *We have that*

$$(Y_{[rN]}/N, \lambda_{[rN]}/N) \Rightarrow (V(r), V(r)),$$

with $V(r)$ is as in Theorem 2.1.

These results are proven rigorously in the appendix; however, we will present the following heuristic for them here. We observe that

$$\lambda_{n+1} = \omega + \alpha Y_n + \beta \lambda_n = \omega + \lambda_n + \alpha(Y_n - \lambda_n)$$

$$= \omega + (\alpha + \beta)\lambda_n + \alpha \sqrt{\lambda_n} \frac{Y_n - \lambda_n}{\sqrt{\lambda_n}} = \omega + \lambda_n + \alpha \sqrt{\lambda_n} \frac{Y_n - \lambda_n}{\sqrt{\lambda_n}}.$$
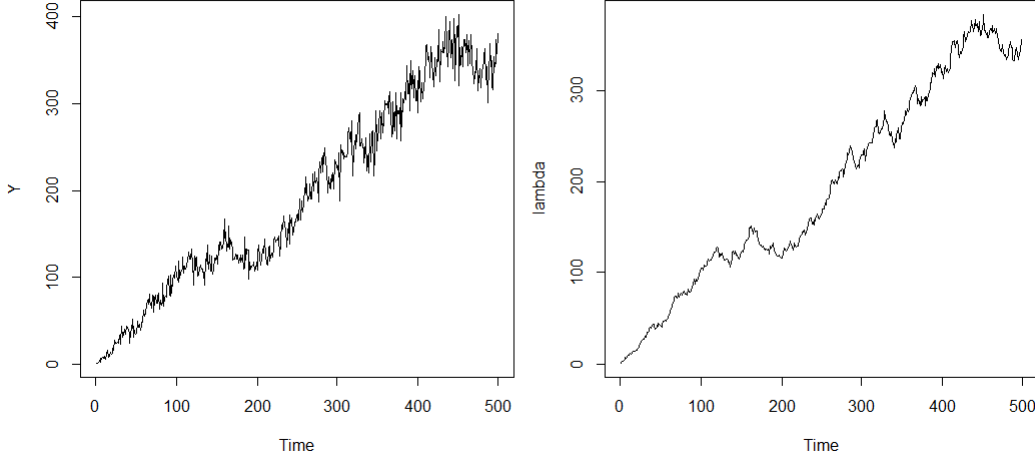
We let $z_n = \frac{Y_n - \lambda_n}{\sqrt{\lambda_n}}$, we observe that this is a martingale difference sequence with conditional variance 1. This implies that $\frac{1}{\sqrt{N}} \sum_{j=1}^{[rN]} z_j \Rightarrow W(r)$, where $W(r)$ is a standard Brownian Motion. Therefore,

$$V_N(\frac{n+1}{N}) - V_N(\frac{n}{N}) = \frac{\omega}{N} + \alpha \sqrt{V_N(\frac{n}{N})} \frac{z_n}{\sqrt{N}}.$$

If we divide by $1/N$ and take limits this suggests that

$$dV(t) \approx \omega \, dt + \alpha \sqrt{V(t)} dW(t).$$

4

A sample path of the process is plotted below using $(\omega, \alpha, \beta) = (1, .3, .7)$ and $N = 500$.

3 Estimation

3.1 Maximum Likelihood Estimation

In this section we discuss implications that Theorem 2.1 has upon maximum likelihood estimation. Throughout we will let $\theta_0 = (\omega_0, \alpha_0, \beta_0)$ denote the true parameters and $\theta = (\omega, \alpha, \beta)$ denote arbitrary parameters. In the stationary setting the likelihood function is given by

$$Q_N(\theta) = \sum_{n=0}^{N} [Y_n \log \lambda_n(\theta) - \lambda_n(\theta) - \log(Y_n!)],$$

where $\lambda_n(\theta)$ is such that $\lambda_0(\theta) = \omega$ and for $n \geq 0$ $\lambda_{n+1}(\theta) = \omega + \alpha Y_n + \beta \lambda_n(\theta)$. In the stationary setting the maximum likelihood estimator can be shown to be asymptotically normal with rate $\sqrt{N}$, see Ferland, Latour, and Oraichi (2006).

In contrast, when $\alpha + \beta = 1$ and the model is non-stationary the maximum likelihood estimation is less straightforward. The main complication is highlighted in the following Lemma.

Lemma 3.1. *For all θ such that θ is in a compact set with $\alpha, \beta \in (0, 1)$ and*

$\omega > 0$ *we have that*

$$\lambda_{[rN]}(\theta)/N \Rightarrow \frac{\alpha}{1 - \beta}V(r),$$

where $V(r)$ is as in Theorem 2.1.

This result is surprising for two reasons. First, the value of ω does not enter into the distribution, Second, for any $\alpha + \beta = 1$ we will have that $\lambda_{[rN]}(\theta)/N \Rightarrow V(r)$, even if $(\alpha, \beta) \neq (\alpha_0, \beta_0)$. This suggests the conjecture that maximum likelihood estimation will be able to correctly estimate $\alpha + \beta$, but not each term individually.

We can use Lemma 3.1 to study the behavior of the likelihood function. We first recall Stirling's approximation that for $k \in \mathbb{N}$ we have

$$\log(k!) = k \log(k) - k + O(\log(k)).$$

Therefore,

$$Q_N(\theta) = \sum_{n=0}^{N}[Y_n \log(\lambda_n(\theta)) - \lambda_n(\theta) - Y_n \log(Y_n) + Y_n + O(\log(Y_n))]$$

$$= \sum_{n=0}^{N}[Y_n \log(\frac{\lambda_n(\theta)}{Y_n}) - \lambda_n(\theta) + Y_n + O(\log(Y_n))].$$

Therefore, Theorem 2.1 and Lemma 3.1 suggest that

$$N^{-2}Q_N(\theta) \Rightarrow [\log(\frac{\alpha}{1 - \beta}) - \frac{\alpha}{1 - \beta} - 1] \int_0^1 V(r)dr. \tag{3}$$

We note that this argument is not fully rigorous as the function $\log(x/y)$ has a pole at $x = 0$ and is undefined at $y = 0$. However a proof that fully addresses this issue is outside the scope of this paper.

Since $V(r) \geq 0$ we can see that the above process will be maximized at $\alpha + \beta = 1$. This implies that the maximum likelihood estimate will be such that $\hat{\alpha} + \hat{\beta} \xrightarrow{P} 1$. This analysis alone does not address the convergence of $\hat{\theta}$ to θ_0 or a possible rate. Motivated by later sections the authors conjecture that $\hat{\alpha} + \hat{\beta}$ will converge to 1 with rate N, however this has not been shown.

We note that the previous Theorem still leaves open the possibility that the maximum likelihood estimate will still be consistent, since it is possible

that the scale factor of N^{-2} hides other higher order behavior of the objective function.

In addition, we note that this type of behavior is not simply an artifact of considering the likelihood function instead of the score function. Define the score function as

$$S_N(\theta) = \sum_{n=0}^{N} (\frac{Y_n}{\lambda_n(\theta)} - 1)\nabla\lambda_n(\theta).$$

Similar to Lemma 3.1 we can show that

$$N^{-1}\nabla\lambda_{[rN]}(\theta) = N^{-1}(\frac{\partial}{\partial\omega}, \frac{\partial}{\partial\alpha}, \frac{\partial}{\partial\beta})\lambda_{[rN]}(\theta) \Rightarrow (0, \frac{1}{1-\beta}V(r), \frac{-\alpha}{(1-\beta)^2}V(r)),$$

and therefore

$$N^{-1}S_N(\theta) \Rightarrow (0, \int_0^1 (\frac{1-\beta}{\alpha}-1)\frac{1}{1-\beta}V(r)dr, \int_0^1 (\frac{1-\beta}{\alpha}-1)\frac{-\alpha}{(1-\beta)^2}V(r)dr).$$

We observe that this process equals 0 for any θ such that $\alpha + \beta = 1$.

Further work is needed to find the statistical properties of the maximum likelihood estimator for the non-stationary integer-valued GARCH process. This is outside of the scope of our paper and so we do not pursue it. Simulation results suggest that the maximum likelihood estimator is still consistent for each parameter, however the rate and limit distributions are unclear.

3.2   Least Squares Estimation

As noted by Ferland, Latour, and Oraichi (2006), Y_n has the ARMA(1,1) representation given by

$$Y_n = \omega_0 + (\alpha_0 + \beta_0)Y_{n-1} + (Y_n - \lambda_n) - \beta_0(Y_{n-1} - \lambda_{n-1})$$

$$= \omega_0 + (\alpha_0 + \beta_0)Y_{n-1} + \sqrt{\lambda_n}z_n - \beta_0\sqrt{\lambda_{n-1}}z_{n-1},$$

where $z_i = \frac{Y_i-\lambda_i}{\sqrt{\lambda_i}}$ is a martingale difference sequence with conditional variance 1. This representation suggests that the least squares regression of Y_{n-1} on Y_n may be fruitful for estimating $\alpha_0 + \beta_0$.

We let $\bar{Y} = N^{-1}\sum_{n=0}^{N} Y_n$ and define

$$\hat{\rho} = \frac{\sum_{n=1}^{N}(Y_n - \bar{Y})(Y_{n-1} - \bar{Y})}{\sum_{n=1}^{N}(Y_{n-1} - \bar{Y})^2}. \tag{4}$$

Then, by Theorem 2.1 we observe that $\hat{\rho} \xrightarrow{P} \frac{\int_0^1 (V(r) - \int_0^1 V(s)ds)^2 dr}{\int_0^1 (V(r) - \int_0^1 V(s)ds)^2 dr} = 1 = \alpha_0 + \beta_0$.

Additionally, we observe that

$$\hat{\rho} - 1 = \frac{\sum_{n=1}^N (Y_n - Y_{n-1})(Y_{n-1} - \bar{Y})}{\sum_{n=1}^N (Y_{n-1} - \bar{Y})^2}$$

$$= \frac{\sum_{n=1}^N (\omega_0 + \sqrt{\lambda_n} z_n - \beta_0 \sqrt{\lambda_{n-1}} z_{n-1})(Y_{n-1} - \bar{Y})}{\sum_{n=1}^N (Y_{n-1} - \bar{Y})^2}.$$

Therefore, using standard techniques analagous to Chan and Wei (1988) one can show that

$$N(\hat{\rho} - 1) \xrightarrow{d} \frac{(1 - \beta_0) \int_0^1 (V(r) - \int_0^1 V(s)ds)\sqrt{V(r)}dW(r)}{\int_0^1 (V(r) - \int_0^1 V(s)ds)^2 dr}$$

$$= \alpha_0 \frac{\int_0^1 (V(r) - \int_0^1 V(s)ds)\sqrt{V(r)}dW(r)}{\int_0^1 (V(r) - \int_0^1 V(s)ds)^2 dr}, \tag{5}$$

where $W(r)$ is the Brownian motion obtained from $N^{-1/2} \sum_{n=0}^{[rN]} z_n \Rightarrow W(r)$. We note that $V(s)$ and $W(s)$ are, in general, not independent. We however do not characterize their exact dependence.

This contrasts with the stationary setting where one can see that $\hat{\rho}$ converges to $\alpha_0 + \beta_0$ with rate $\sqrt{N}$. This is similar to what occurs with the AR(1) process, where the rate is either $\sqrt{N}$ or N depending upon whether or not the process is stationary.

4 A test of non-stationarity with an application

In this section we highlight the use of Theorem 2.1 in determining whether or not an integer-valued GARCH process is stationary. Our main observation is that Theorem 2.1 implies that a KPSS test (see Kwiatkowski et al (1992)) will have power against the alternative that $\alpha_0 + \beta_0 = 1$.

In this section we will consider

$$H_0 : \alpha + \beta < 1$$

8

and

$$H_A : \alpha + \beta = 1.$$

We define $\bar{Y} = N^{-1} \sum_{n=0}^{N} Y_n$, $S_j = \sum_{n=0}^{j} (Y_n - \bar{Y})$, and $\mathcal{K} = \sum_{n=0}^{N} S_n^2$. The KPSS test is based off the dichotomous behavior $\mathcal{K}$ has under H_0 versus H_A.

Under H_0 the behavior of this test is standard since it can be seen that $N^{-1/2} \sum_{n=0}^{[rN]} Y_n \Rightarrow \sigma W(r)$, where $W(r)$ is a standard Brownian motion and σ is the long-run variance of Y_n. This result holds since under H_0 Y_n was shown to be strictly stationary and β-mixing by Neumann (2011). Therefore,

$$N^{-1} S_{[rN]} \Rightarrow \sigma(W(r) - rW(1)),$$

and so

$$\sigma^{-2} N^{-2} \mathcal{K} \xrightarrow{d} \int_0^1 (W(r) - r(W(1))^2 dr.$$

The critical values for this test statistic can be found in Kwiatkowski et al (1992). An HAC estimator (see Newey and West (1987)) for σ^2 will need to be used in practice, since the true σ^2 is unknown.

Under H_A we can use Theorem 2.1 to show that

$$N^{-2} S_{[rN]} \Rightarrow \int_0^r V(s)ds - rV(1).$$

Therefore,

$$N^{-5} \mathcal{K} \xrightarrow{d} \int_0^1 (\int_0^r V(s)ds - rV(1))^2 dr.$$

In addition, Theorem 2.1 also implies that for any HAC estimator of the long-run variance we will have that $\hat{\sigma}^2 = O_P(N^2)$ under standard assumptions regarding the kernel function and bandwidth. Together these properties imply that the asymptotic power of a KPSS test against H_A will go to 1.

4.1 Application

In this section we look at count data coming from the weekly number of reported measles infections in North Rhine-Westphalia from January 2001 to

May 2013 (N=646), this data is found in the R package "tscount", for more details see Liboschik, Fokianos, and Fried (2015). A graph of the data and its autocorrelation function is given below.
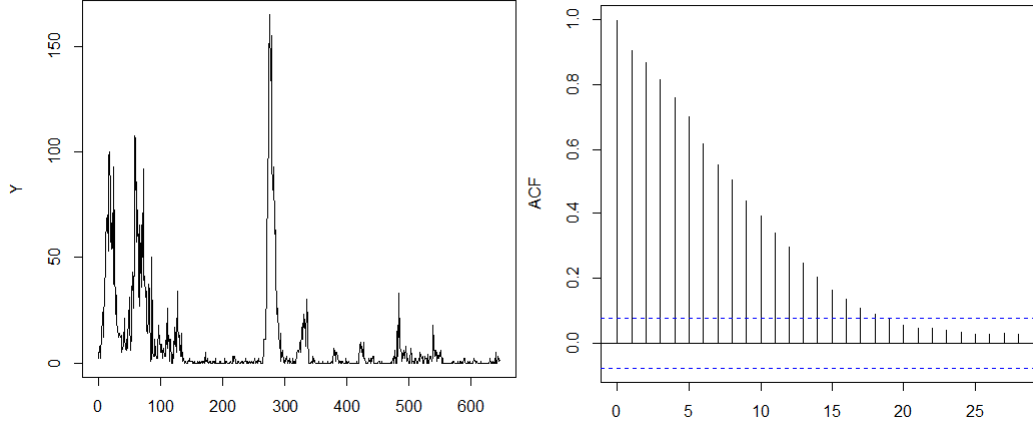
For this data set the maximum likelihood estimation results in

$$(\hat{\omega}, \hat{\alpha}, \hat{\beta}) = (.19, .39, .58),$$

and so $\hat{\alpha} + \hat{\beta} = .97$. As such, it is unclear apriori whether or not this data should be thought of as stationary or nonstationary. Computing the KPSS test gives a test statistic of 1.67. Comparison with the critical values in Kwiatkowski et al (1992) shows that we will reject H_0 even at the 1% level.

We note that this test does not suggest any evidence that the data set behaves like the diffusion process found in Theorem 2.1, but just that it has non-stationarity of some form. In addition, we emphasize that the distribution obtained under H_0 will still hold so long as Y_n is a sequence which satisfies a functional central limit theorem. In this sense the null is only indirectly using the Poisson functional form assumption so as to guarantee a stationary sequence with the necessary properties.

In addition, we note that an alternative test for stationarity based off the null $H_0 : \alpha_0 + \beta_0 = 1$ could be constructed using the least squares estimator $\hat{\rho}$ from Equation (4). A test along these lines would be analagous to a Dickey-Fuller test (see Dickey and Fuller (1979)). We will not fully develop this test here since the limiting distribution depends upon α_0, which we currently can not estimate. We do note, however, for this data set we have $\hat{\rho} = .908$. Given that N is quite large for this data it is likely that this test would reject the null of $\alpha_0 + \beta_0 = 1$ for reasonable values of α_0.

10

In fact, treating $\hat{\alpha}$ as α_0 and conducting simulations for the distribution of $\hat{\rho}$ given by Equation (5) we have that the value $\hat{\rho} = .908$ has a p-value of approximately .014. However, we do note that for a $N = 646$ it appears as though the asymptotic distribution is still quite different from the small sample distribution. The small sample distribution is shown below. This distribution gives $\hat{\rho}$ a p-value of .0612.
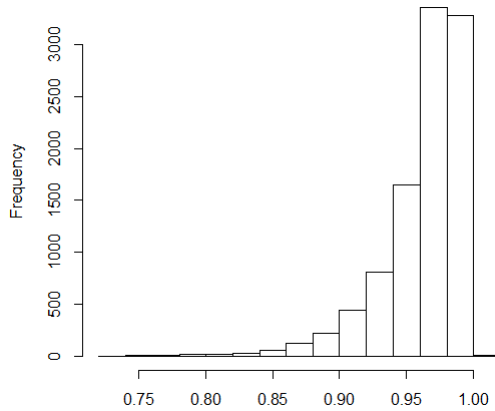
Overall, this analysis suggests that the process is nonstationary, but likely does not follow the distribution of a nonstationary INGARCH.

5 Conclusion

In this paper we find the limit distribution of the integer-valued GARCH model when $\alpha + \beta = 1$. In contrast with the standard GARCH model we find that the process will be non-stationary given these parameters. As such, the distinct between $\alpha + \beta < 1$ and $\alpha + \beta = 1$ should be thought of as being closer to the difference between a stationary autoregressive process and a unit root process. Additional work is still needed to fully understand the asymptotic behavior of the maximum likelihood estimator in this setting.

In addition, it may be of interest to study if the nonstationary INGARCH has implications for more time series. In practice, integer valued time series with large values are often rounded and treated as real valued processes. It is possible that this behavior confuses processes following a nonstationary INGARCH with other models.

11

6 Mathematical appendix

The proof of Theorem 2.1 is based on the semigroup characterization of a Markov process (see, Ethier and Kurtz (2009), for example). Our argument will be similar to the argument used in Theorem 1.3 of Chapter 9 of Ethier and Kurtz (2009) where a weak convergence result for the critical Galton-Watson process is shown.

To begin we first recall some properties of Markov processes. The semigroup characterization of a Markov process, say $X(t)$, is based off the observation that the operators $T(t)$ defined by

$$T(t)f(x) = E[f(X(t))|X(0) = x]$$

generate a one-parameter family of linear operators which forms a semigroup.[2] The fact that this is a semigroup can be seen from the fact that $T(0) = I$ and $T(t + s) = T(s)T(t)$, where the second equation holds by the Markov property.

Given sufficient regularity conditions, we can also define an "infinitesimal generator", or generator, Gf defined by

$$Gf(x) = \lim_{t \to 0} t^{-1}(T(t)f(x) - f(x)).$$

This object can, under certain assumptions, give a complete description of the semigroup $T(t)$, and is often more tractable. If our process $X(t)$ satisfies the stochastic differential equation $dX(t) = b(x)dt + \sigma(x)dW(t)$ then the generator is given by $Gf(x) = b(x)f'(x) + \frac{1}{2}\sigma(x)^2 f''(x)$. This can be seen through a Taylor series expansion.

The benefit of this semigroup approach is that showing the weak convergence $X_n \Rightarrow X$, for X_n a sequence of Markov processes, can often be translated into statements about the semigroups $T(t)$ and $T_n(t)$ or even about the generator A and a "discrete approximation" $A_n = n(T_n(\frac{1}{n}) - I)$. Numerous results of this form can be found in Ethier and Kurtz (2009) and the many references found within.

---

[2]We note that the notation $T(t)f(x)$ is meant to be read as $(T(t)f)(x)$, with the parentheses being implicit. Here $T(t)$ is an operator on the space of functions and $T(t)f$ is therefore a function, which can be evaluated at x.

Proof of Theorem 2.1:

We first note that $V_N(0) = \frac{\lambda_0}{N} = \frac{\omega}{N} \xrightarrow{d} 0$. We let $T(t)$ denote $V(r)$'s associated semigroup and we note that $T(t)$ has the generator $Gf(x) = \omega f'(x) + \frac{\alpha^2}{2}xf''(x)$.

We define $T_N(t)f(x) = E[f(V_N(t))|V_N(0) = x]$ and $G_N f(x) = N(T_N(\frac{1}{N})f(x) - f(x))$. G_N corresponds to the discrete approximation of $T_N(t)$'s generator.

By Corollary 8.9(i) of Chapter 4 of Ethier and Kurtz (2009), showing that $V_N(r) \Rightarrow V(r)$ reduces to showing the following condition:

For every bounded $f : [0, \infty) \to \mathbb{R}$ and $t \in [0, 1]$ we have that

$$\lim_{N \to \infty} \sup_{x \in [0,\infty)} |T_N(t)f(x) - T(t)f(x)| = 0. \tag{6}$$

This is, by Theorem 2.1 of Chapter 8 and Theorem 6.5 of Chapter 1 of Ethier and Kurtz (2009), equivalent to showing that

For every $f \in C_c^\infty[0, \infty)$, the space of infinitely differentiable function on $[0, \infty)$ with compact support, we have the that

$$\lim_{N \to \infty} \sup_{x \in [0,\infty)} |G_N f(x) - Gf(x)| = 0. \tag{7}$$

Alternatively, we show that $\lim_{N \to \infty} \sup_{x \in [0,\infty)} |\varepsilon_N f(x)|$ converges to 0 for each $f \in C_c^\infty[0, \infty)$, where $\varepsilon_N f(x) = G_N f(x) - Gf(x)$.

We first note that for the case of $x = 0$ this is trivial, we therefore consider the case of $x > 0$. We first observe that

$$T_N(\frac{1}{N})f(x) = E[f(\frac{\lambda_1}{N})|\frac{\lambda_0}{N} = x] = E[f(\frac{\omega + \alpha Y_0 + \beta \lambda_0}{N})|\lambda_0 = xN].$$

Motivated by the fact that $E[\frac{\lambda_1}{N}|\lambda_0 = Nx] = \frac{\omega}{N} + x$ we will take the Taylor expansion of $f(\cdot)$ around x. Therefore, $T_N f(x)$ is equal to

$$E[f(x) + f'(x)(\frac{\lambda_1}{N} - x) + \frac{f''(x)}{2}(\frac{\lambda_1}{N} - x)^2 + \frac{f'''(\tilde{x})}{6}(\frac{\lambda_1}{N} - x)^3|\lambda_0 = Nx],$$

for some $\tilde{x}$ between x and λ_1/N.

We observe that

$$\lambda_1 = \omega + \alpha Y_0 + \beta \lambda_0 = \omega + (\alpha + \beta)\lambda_0 + \alpha \sqrt{\lambda_0}\frac{Y_0 - \lambda_0}{\sqrt{\lambda_0}}$$

13

$$= \omega + \lambda_0 + \alpha\sqrt{\lambda_0}z_0,$$

where $z_0 = \frac{Y_0 - \lambda_0}{\sqrt{\lambda_0}}$ is a martingale difference sequence with mean 0, conditional variance 1, $E[z_0^3|\lambda_0 = Nx] = \frac{1}{Nx}$, and $E[z_0^4|\lambda_0 = Nx] = \frac{1+3Nx}{Nx}$.

Therefore, given that $\lambda_0 = Nx$ we have that $\frac{\lambda_1}{N} - x = \frac{\omega}{N} + \alpha\sqrt{\frac{x}{N}}z_0$. Using this information we can see that

$$G_N f(x) = \omega f'(x) + \frac{\alpha^2}{2}xf''(x) + \frac{\omega^2}{N}f''(x) + NE[\frac{f'''(\tilde{x})}{6}(\frac{\omega}{N} + \alpha\sqrt{\frac{x}{N}}z_0)^3|\lambda_0 = Nx].$$

We note that it does not immediately follow that $\sup_{x\in[0,\infty)}|\varepsilon_N(x)| \to 0$ since we must still show that the higher order terms go to 0 uniformly.

To handle this we first note that since $f \in C_c^\infty[0,\infty)$ there exists some C_f such that for all $s \geq C_f$ we have that $f(s) = 0$. In addition, since Y_0 is a positive random variable it follows that $P(\frac{\lambda_1}{N} \geq C|\frac{\lambda_0}{N} = \frac{C}{\beta}) = 1$, for all C and as such $P(f(\frac{\lambda_1}{N}) = 0|\frac{\lambda_0}{N} \geq \frac{C_f}{\beta}) = 1$. Therefore,

$$\sup_{x\in[0,\infty)}|\varepsilon_N f(x)| = \sup_{x\in[0,C_f/\beta]}|\varepsilon_N f(x)|$$

$$= \sup_{x\in[0,C_f/\beta]}|\frac{\omega^2}{N}f''(x) + NE[\frac{f'''(\tilde{x})}{6}(\frac{\omega}{N} + \alpha\sqrt{\frac{x}{N}}z_0)^3|\lambda_0 = Nx]|$$

$$\leq \sup_{x\in[0,C_f/\beta]}(\frac{\omega^2}{N}||f''||_\infty + N\frac{||f'''||_\infty}{6}E[|\frac{\omega}{N} + \alpha\sqrt{\frac{x}{N}}z_0|^3|\lambda_0 = Nx]),$$

where $||f''||_\infty$ and $||f'''||_\infty$ are both finite since $f \in C_c^\infty[0,\infty)$. By Jensen's inequality this is bounded from above by

$$\sup_{x\in[0,C_f/\beta]}(\frac{\omega^2}{N}||f''||_\infty + N\frac{||f'''||_\infty}{6}E[(\frac{\omega}{N} + \alpha\sqrt{\frac{x}{N}}z_0)^4|\lambda_0 = Nx]^{3/4})$$

$$= \sup_{x\in[0,C_f/\beta]}(\frac{\omega^2}{N}||f''||_\infty + N\frac{||f'''||_\infty}{6}(\frac{\omega^4}{N^4} + 6\frac{\omega^2\alpha^2x}{N^3} + 4\frac{\omega\alpha^3x}{N^3} + \frac{\alpha^4x(1+3Nx)}{N^3})^{3/4})$$

Therefore, for all $f \in C_c^\infty[0,\infty)$ we have that

$$\lim_{N\to\infty}\sup_{x\in[0,\infty)}|\varepsilon_N f(x)| = 0.$$

This completes our proof. □

Proof of Lemma 3.1:

We observe that

$$\lambda_{n+1}(\theta) = \omega \sum_{j=0}^{n} \beta^j + \alpha \sum_{j=0}^{n} \beta^j Y_{n-j}.$$

Therefore,

$$\lambda_{[rN]}(\theta)/N = N^{-1}(\omega \sum_{j=0}^{[rN]} \beta^j + \alpha \sum_{j=0}^{[rN]} \beta^j Y_{n-j}).$$

Using Theorem 2.1 we observe that

$$\lambda_{[rN]}(\theta)/N \Rightarrow \frac{\alpha}{1-\beta} V(r).$$

This completes our proof. $\square$

References

[1] Bollerslev, Tim. 1986. *Generalized autoregressive conditional heteroskedasticity.* Journal of econometrics 31, no. 3: 307-327.

[2] Chan, N. H., and C. Z. Wei. 1988. *Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes.* The Annals of Statistics 16, no. 1: 367-401.

[3] Dawson, Donald A. 2017. *Introductory lectures on stochastic population systems.* arXiv preprint arXiv:1705.03781.

[4] Dickey, David A., and Wayne A. Fuller. 1979. *Distribution of the estimators for autoregressive time series with a unit root.* Journal of the American statistical association 74, no. 366a: 427-431.

[5] Engle, Robert F., and Tim Bollerslev. 1986. *Modelling the persistence of conditional variances.* Econometric reviews 5, no. 1: 1-50.

[6] Ethier, Stewart N., and Thomas G. Kurtz. 2009. *Markov processes: characterization and convergence.* Vol. 282. John Wiley & Sons.

[7] Feller, Willy. 1939. *Die grundlagen der volterraschen theorie des kampfes ums dasein in wahrscheinlichkeitstheoretischer behandlung.* Acta Biotheoretica 5, no. 1: 11-40.

[8] Feller, William. 1951. *Diffusion processes in genetics.* In Proceedings of the second Berkeley symposium on mathematical statistics and probability. The Regents of the University of California.

[9] Ferland, Ren, Alain Latour, and Driss Oraichi. 2006. *IntegerValued GARCH Process.* Journal of Time Series Analysis 27, no. 6: 923-942.

[10] Fokianos, Konstantinos 2015. *Statistical analysis of count time series models: A GLM perspective.* Handbook of discrete-valued time series, Handbooks of Modern Statistical Methods: 3-28.

[11] Fokianos, Konstantinos, Anders Rahbek, and Dag Tjøstheim. 2009. *Poisson autoregression.* Journal of the American Statistical Association 104, no. 488: 1430-1439.

[12] Fokianos, Konstantinos, and Dag Tjøstheim. 2011. *Log-linear Poisson autoregression.* Journal of Multivariate Analysis 102, no. 3: 563-578.

[13] Hamilton, James Douglas. 1994. *Time series analysis.* Vol. 2. Princeton: Princeton university press.

[14] Kwiatkowski, Denis, Peter CB Phillips, Peter Schmidt, and Yongcheol Shin. 1992. *Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?.* Journal of econometrics 54, no. 1-3: 159-178.

[15] Liboschik, Tobias, Konstantinos Fokianos, and Roland Fried. 2015 *tscount: An R package for analysis of count time series following generalized linear models.* Universittsbibliothek Dortmund.

[16] Nelson, Daniel B. 1990. *Stationarity and persistence in the GARCH (1, 1) model.* Econometric theory 6, no. 3: 318-334.

[17] Neumann, Michael H. 2011. *Absolute regularity and ergodicity of Poisson count processes.* Bernoulli 17, no. 4 : 1268-1284.

[18] West, Kenneth D., and Whitney K. Newey. 1987 *A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.* Econometrica 55, no. 3: 703-708.

[19] Rydberg, Tina Hviid, and Neil Shephard. 1999. *BIN models for trade-by-trade data. Modelling the number of trades in a fixed interval of time.* Unpublished Paper. Available from the Nuffield College, Oxford Website.

[20] Schlögl, Friedrich. 1972. *Chemical reaction models for non-equilibrium phase transitions.* Zeitschrift fr Physik A Hadrons and Nuclei 253, no. 2: 147-161.

[21] Wang, Chao, Heng Liu, Jian-Feng Yao, Richard A. Davis, and Wai Keung Li. 2014. *Self-excited threshold Poisson autoregression.* Journal of the American Statistical Association 109, no. 506: 777-787.

[22] Zhu, Fukang. 2011. *A negative binomial integervalued GARCH model.* Journal of Time Series Analysis 32, no. 1: 54-67.