A model for level induced conditional heteroskedasticity

Jon Michel* Robert M. de Jong†

September 29, 2018

Abstract

A class of conditional heteroskedasticity models is introduced and analyzed. This class of models is motivated by the desire to allow the level of a GARCH process to influence the volatility. We show the existence of a unique strictly stationary solution which is β-mixing. The analysis of this model does not rely upon Markov chain methods.

Keywords: Conditional Heteroskedasticity, GARCH, β-mixing, nonlinear time series.

1 Introduction

Since the seminal papers of Engle (1982) and Bollerslev (1986), GARCH models have played a prominent role in the analysis of volatility of financial time series. In the standard GARCH(1,1) model the stock returns or log stock returns ΔY_t are of the form

$$\Delta Y_t = \sigma_t z_t, \tag{1}$$

where z_t is an i.i.d. sequence with mean 0 and variance 1, while σ_t follows the specification

$$\sigma_{t+1}^2 = \omega + \alpha z_t^2 \sigma_t^2 + \beta \sigma_t^2. \tag{2}$$

Numerous extensions of this model have been proposed. Many of these extensions are motivated by a desire for σ_{t+1} to respond asymmetrically to positive and negative ΔY_t, which corresponds to the leverage effect pointed out by Black (1976). See, for example, the EGARCH model of Nelson (1991), the TGARCH model of Zakoïan (1994), and the QGARCH of Sentana (1995).

This paper considers a new model for conditional heteroskedasticity. We propose to let the volatility σ_{t+1}^2 depend upon the "level" $Y_t = \sum_{j=1}^{t} \Delta Y_j$, which in applications often stands

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

†Department of Economics, Ohio State University, 454 Arps Hall, Columbus, OH 43210, USA, Email de-jong.8@osu.edu.

for stock price. Therefore, in such an application, we make volatility depend on whether stock price is relatively high. In the model studied here the volatility will be completely determined by the level and as such our model does not nest the GARCH model. We later discuss a more general model that nests both our model and the GARCH(1,1); however, analysis of this model is quite difficult and outside the scope of this paper.

While, as mentioned, the literature has considered asymmetric responses to large ΔY_t, it is possible that such an analysis confounds the effect of relatively large levels of Y_t with such an asymmetric response. After all, the threat to investors of possibly not being able to meet obligations while the stock price level Y_t is relatively low might increase volatility. Similarly, when stock price nears historic highs, risk aversion and profit taking might affect volatility.

In the option pricing literature, models that account for such an effect have been introduced; see for example the Hobson and Rogers model introduced in Hobson and Rogers (1998). Discussions of this effect can also be found in the popular press. For example, in a July 2017 articled published on Bloomberg.com (https://www.bloomberg.com/view/articles/2017-07-03/what-history-says-about-low-volatility), the author discusses the combination of low volatility and historically high levels in the stock market in the first half of 2017. He notes that historically, the combination of high volatility when stock prices are at high levels is much more prevalent. Additionally, in the next subsection we will provide more empirical motivation through an analysis of recent bitcoin prices.

However, in the GARCH literature, no models appear to exist that allow volatility to depend on level. Given that the dynamics of volatility of stock price levels is given considerable attention in the option pricing literature and the popular press, we feel that the GARCH literature would benefit substantially from the introduction of models that capture this effect.

While this suggestion is intuitively appealing, constructing a model that captures this effect is complicated by the unit root properties of Y_t. We address this issue by introducing a latent bound P_t, which represents a relatively high level for Y_t. We then show that our model can be rewritten in a form that is reminiscent of the Lindley equation. In order to show weak dependence and mixing properties of our model we develop techniques that differ from the now standard techniques of Bougerol and Picard (1992a) and (1992b) or Markov chain methods such as in Meyn and Tweedie (2009). The techniques developed here can, with slight modification, be applied to other stochastic processes where a single value is revisited sufficiently often. An example of this type of model is the nonlinear ARCH model of Saïdi and Zakoïan (2006). We will assume i.i.d. errors throughout, however with slight modifications the techniques used here can be easily extended to weakly dependent errors.

While the analysis of our model requires novel and relatively sophisticated arguments, the idea of allowing volatility to depend on level in a GARCH setting in our opinion deserves to be taken up and developed further, given its obvious empirical relevance. In particular, we feel that a GARCH model that allows volatility to depend on level and nests the GARCH(1,1) would be important. However, given the analytical complexity of such an endeavor, we leave this to future work.

The proofs for all results can be found in our appendix.

1.1 Bitcoin volatility

In this section we consider the volatility of Bitcoin daily prices from 8/20/2017 to 8/20/2018; this data is obtained from the Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org/series/CBBTCUSD). We let $(Y_t)_{t=1}^{T}$ denote the time series of Bitcoin prices. In order to test, in a somewhat ad hoc manner, whether the volatility of Bitcoin prices is influenced by the level of the price we conduct the following analysis.

We let K be a positive integer and for $t \in \{K+1, ..., T\}$ we define $N_t^K = I(Y_t = \max_{j\in\{t-K,...,t\}} Y_j)$. We calculate the ratio

$$\frac{\sum_{t=K+1}^{T-K} N_t^K \ s.d.((Y_{t+1}, ..., Y_{t+K}))^2}{\sum_{t=K+1}^{T-K} N_t^K} \Big/ \frac{1}{T-K} \sum_{t=1}^{T-K} s.d.((Y_{t+1}, ..., Y_{t+K}))^2, \tag{3}$$

where $s.d.(\cdot)$ of a sequence represents the sample standard deviation of the sequence.

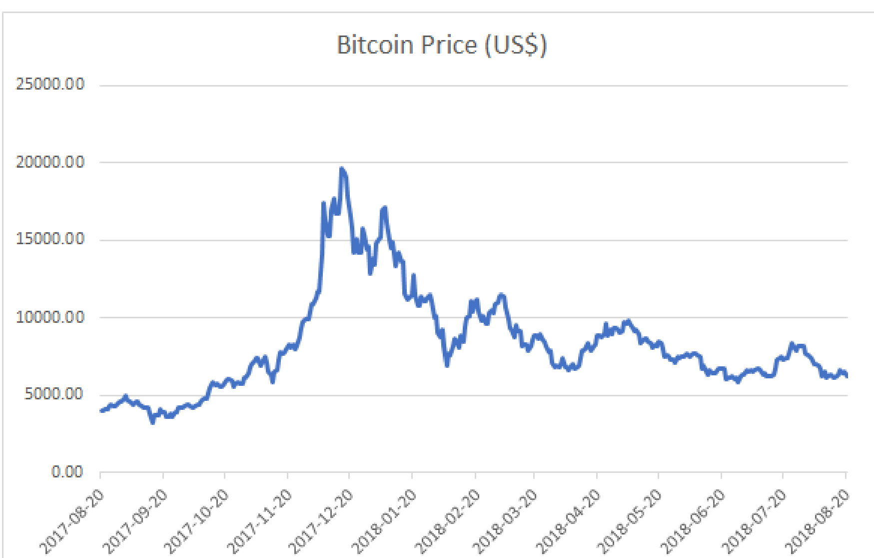
The idea behind this ratio is that it compares the realized volatility in time periods after a "high" level of price with the overall realized volatility in time periods after any level. We note that a value greater than 1 implies higher volatility in the time periods following a "high" level.

Table 1 calculates the statistic of equation (3) for various values of K. The choices for K are representing a time span from one to four weeks. This table suggests that bitcoin prices are more volatile following recent maximums, since the values are all larger than 1.

Table 1: Bitcoin Realized Volatility Post Relative High

K	7	14	21	28
Realized Volatility Ratio	1.38	1.59	1.70	1.81

Additionally, we plot the level of Bitcoin price below. The time series visually appears to have higher volatility near maximums.



2 Model equations

To model a process where volatility is induced by the level we will introduce a latent process P_t, which will represent a threshold for a high versus low level of Y_t. The idea of using a latent process to represent a high or low level has been used elsewhere in the setting of unit root processes and cointegration; see Michel and de Jong (2018a). The conditional heteroskedasticity setting presents different analytical challenges, since exceeding the latent bound now affects volatility, instead of the level.

Let $(z_t)_{t\in\mathbb{Z}}$ be an i.i.d. sequence of random variables with mean 0 and variance 1, and set $Y_0 = 0$ and $P_0 = 0$. Our model equations are

$$Y_t = \sum_{j=1}^{t} \sigma_j z_j \tag{4}$$

where

$$\sigma_{t+1}^2 = \sigma^2 + cI(Y_t \geq P_t) \tag{5}$$

and

$$P_{t+1} = (P_t - a)I(Y_t < P_t) + (Y_{t+1} + \eta)I(Y_t \geq P_t). \tag{6}$$

Here $\eta \geq 0$, $a > 0$, $\sigma^2 > 0$, and $\sigma^2 + c > 0$. We note that $\Delta Y_t = \sigma_t z_t = (\sigma^2 + cI(Y_{t-1} \geq P_{t-1}))z_t$, and so σ_t^2 is the conditional variance of ΔY_t at time t.

Through Equation (5) the level Y_t is allowed to affect the volatility σ_t of the process. A "high" level of Y_t induces a change in volatility in the next time period. The notion of a "high" level is made precise by introducing a latent upper bound P_t. Levels above P_t are considered high; levels below P_t are considered low. In addition, the bound P_t falls by a every time period that Y_t does not cross P_t. This causes our notion of high and low levels to be relative to nearby time periods. We note here that the process σ_t^2 only takes on two distinct values, this is motivated by our desire to study a model where volatility is only induced by high/low levels while remaining tractable.

We note that the bound P_t could instead be placed below Y_t. The analysis of this situation is analogous to the situation considered here.

2.1 Properties of the model

In order to study the above system of equations we first define $D_t = P_t - Y_t$. We observe that Equation (5) can be written in terms of only σ_t and D_t since by Equation (5),

$$\sigma_{t+1}^2 = \sigma^2 + cI(D_t \leq 0), \tag{7}$$

and by Equations (4) and (6), D_t follows the recurrence relation

$$D_{t+1} = (D_t - a - \sigma_{t+1} z_{t+1})I(D_t > 0) + \eta I(D_t \leq 0). \tag{8}$$

Therefore, (σ_{t+1}, D_t) can be studied without reference to Y_t or P_t. In addition, we note that after substitution of Equation (7), Equation (8) is equivalent to

$$D_{t+1} = (D_t - a - \sigma z_{t+1}) I(D_t > 0) + \eta I(D_t \leq 0). \tag{9}$$

This is nearly the Lindley equation $D_{t+1} = \max(0, D_t - a - \sigma z_{t+1})$ from Queueing Theory; see, for example, Baccelli and Brémaud (1994).

To start our analysis, we will first give a stochastic bound for D_t.

Lemma 2.1. *For some $m \in \mathbb{Z}$, and sequence $(\tilde{D}_t)_{t \geq m}$ satisfying the equation $\tilde{D}_{t+1} = (\tilde{D}_t - a - \sigma z_{t+1}) I(\tilde{D}_t > 0) + \eta I(\tilde{D}_t \leq 0)$ and $\tilde{D}_m = 0$, we have that*

$$\tilde{D}_t \leq W_t, \tag{10}$$

where we define $W_t = \max_{k \geq 0}[\eta - ak - \sigma \sum_{j=0}^{k-1} z_{t-j}]$, using the convention that summation over a null set is 0. In addition, W_t is almost surely finite.

W_t is the maximum of a random walk with negative drift. The random walk with drift has been well studied in other contexts; see Janson (1986) and Kiefer and Wolfowitz (1956), for example. Lemma 2.1 allows us to show the following result, which states that for m large enough, σ_t "almost" depends only on $(z_{t-m}, \ldots, z_t)$.

Theorem 2.2. *Let $\mathcal{F}_{t-m}^t$ be the sigma algebra generated by $(z_{t-m}, \ldots, z_t)$. If z_t has full support over $\mathbb{R}$, then for $t \geq 1$ there exists $\tilde{\sigma}_t^m \in \mathcal{F}_{t-m}^t$ such that*

$$\sup_{t \geq 1} P(\sigma_t \neq \tilde{\sigma}_t^m) = \nu(m) \tag{11}$$

and $\nu(m) = O(m^{-r+\varepsilon})$ for any $\varepsilon > 0$ and any $r > 0$ such that $E|\min(0, z_j)|^{r+1} < \infty$.

We note that since z_t was earlier assumed to have a finite variance the above moment condition holds with $r = 1$, however additional moments give a faster rate.

Theorem 2.2 and the fact that σ_t is bounded imply a form of weak dependence known in the econometrics literature as near epoch dependence; see Andrews (1988) and Pötscher and Prucha (1997), for example. This property implies a law of large numbers, and with sufficient rate for $\nu(m)$ implies a central limit theorem. Therefore, σ_t and hence ΔY_t satisfies a central limit theorem under moment conditions on z_t.

We can use Theorem 2.2 to show that there exists a unique strictly stationary (σ_{t+1}^*, D_t^*) when the process is extended to $t \in \mathbb{Z}$.

Theorem 2.3. *If z_t has full support over $\mathbb{R}$ and there exists some $r > 0$ such that $E|\min(0, z_t)|^{r+1} < \infty$, then there exists a unique strictly stationary solution (σ_{t+1}^*, D_t^*) to Equations (7) and (9) for $t \in \mathbb{Z}$.*

It now follows that $\Delta Y_t^* = \sigma_t^* z_t, t \in \mathbb{Z}$, is a strictly stationary sequence. In addition, we can show that ΔY_t^* is β-mixing:

Theorem 2.4. *If z_t has full support over $\mathbb{R}$ and there exists some $r > 0$ such that $E|\min(0, z_t)|^{r+1} < \infty$, then ΔY_t^* is β-mixing with mixing coefficient $\beta_m = O(m^{-r+\varepsilon})$ for any $\varepsilon > 0$.*

This implies that there exists a unique stationary solution for $P_t - Y_t$ and ΔY_t which is β-mixing.

3 Extensions

The model can easily be extended to include multiple lags. The model equations are then

$$Y_t = \sum_{j=1}^{t} \sigma_j z_j,$$

$$\sigma_{t+1}^2 = \sigma^2 + \sum_{k=1}^{K} c_k I(Y_{t+1-k} \geq P_{t+1-k}),$$

and

$$P_{t+1} = (P_t - a)I(Y_t < P_t) + (Y_{t+1} + \eta)I(Y_t \geq P_t).$$

All of the result of the previous section apply to this model with only minor modifications in the proofs. We note that while the σ_t^2 process is modified, the P_t process is unchanged from the model introduced in Section 2. This is done in order to keep the definition of "high" and "low" level unchanged while still allowing the volatility to react differently to the history.

A more general extension would be a model that nests the GARCH(1,1) model. A possible specification for this would be

$$\sigma_{t+1}^2 = \omega + \alpha\sigma_t^2 z_t^2 + \beta\sigma_t^2 + cI(Y_t \geq P_t), \tag{12}$$

with $\omega > 0$, $\alpha, \beta \in (0,1)$ and $\alpha + \beta \leq 1$.

This model cannot be analyzed using the techniques of Theorem 2.2 because σ_t will no longer have a mass point. Therefore, the general analysis of this model is outside the scope of this paper. Further study of this model would be important for assessing the impact of the level on volatility in a GARCH setting.

4 Conclusion

In this paper, we introduced a new model of conditional heteroskedasticity which accounts for the level of process possibly impacting volatility. This model was then shown to have weak dependence properties and a unique stationary solution which is β-mixing.

5 Mathematical Appendix

Proof of Lemma 2.1:

We recall that

$$D_{t+1} = (D_t - a - \sigma z_{t+1})I(D_t > 0) + \eta I(D_t \leq 0)$$

from Equation (8). We now define τ^t to be

$$\tau^t = \max\{j \in \{m, .., t\} : D_j \leq 0\}$$

and we note that τ^t is well-defined since $D_m = 0$. From this definition, we can see that

$$D_{t+1} = \eta - a(t - \tau^t) - \sum_{j=\tau^t+2}^{t+1} \sigma z_j.$$

Here we again use the convention that a sum over an empty index is 0. Since $\tau^t \in \{m, ...t\}$, D_{t+1} is bounded from above by

$$\max_{k\in\{m,...,t\}} [\eta - a(t-k) - \sigma \sum_{j=k+2}^{t+1} z_j] = \max_{k\in\{0,...,t-m\}} [\eta - ak - \sigma \sum_{j=0}^{k-1} z_{t+1-j}].$$

Extending the set from $k \in \{0, ..., t-m\}$ to $\mathbb{N}$ then gives us that $D_{t+1} \leq W_{t+1}$. The fact that W_t is almost surely finite follows from the strong law of large numbers, the i.i.d. property of z_j, $Ez_j = 0$, and $E|z_j| < \infty$. $\square$

Lemma 5.1. *If D_t satisfies Equation (9) and $a + \sigma z_{t-1} \leq 0$, then $D_t = D_{t-1} - a - \sigma z_t$.*

Proof. We will show that $a+\sigma z_{i-1} \leq 0$ implies that $D_{i-1} > 0$ and hence $D_i = D_{i-1} - a - \sigma z_i$ by Equation (9). This will then imply our result. In the case $D_{i-2} > 0$, we will necessarily have $D_{i-1} = D_{i-2} - (a + \sigma z_{i-1}) > 0$ since both $D_{i-2} > 0$ and $a + \sigma z_{i-1} \leq 0$. In the case where $D_{i-2} \leq 0$, we have $D_{i-1} = \eta > 0$. This completes our proof. $\square$

Proof of Theorem 2.2:

We fix a t and $m \in \mathbb{N}^+$. We will construct an approximation $\tilde{I}_t^m$ to $I(D_t \leq 0)$. We observe that when letting $\tilde{\sigma}_t^m = \sqrt{\omega + c\tilde{I}_t^m}$, it then follows that

$$P(\sigma_t \neq \tilde{\sigma}_t^m) = P(I(D_t \leq 0) \neq \tilde{I}_t^m). \tag{13}$$

Therefore, we will focus our attention on approximating $I(D_t \leq 0)$.

We construct $\tilde{I}_t^m$ as follows. Let $\tilde{D}_{t-m-1}^m = 0$ and for $j \in \{t-m, ..., t\}$ let

$$\tilde{D}_j^m = (\tilde{D}_{j-1}^m - a - \sigma z_j) I(\tilde{D}_{j-1}^m > 0) + \eta I(\tilde{D}_{j-1}^m \leq 0). \tag{14}$$

Now define $\tilde{I}_t^m = I(\tilde{D}_t^m \leq 0)$. We observe that if there is some $\tilde{t} \in \{t-m-1, ..., t-1\}$ such that $D_{\tilde{t}} \leq 0$ and $\tilde{D}_{\tilde{t}}^m \leq 0$, then $D_{\tilde{t}+1} = \tilde{D}_{\tilde{t}+1}^m = \eta$, and then for all $j \in \{\tilde{t}+1, ..., t\}$ we have $D_j = \tilde{D}_j^m$, implying that $I(D_t \leq 0) = \tilde{I}_t^m$. Therefore,

$$P(I(D_t \leq 0) \neq \tilde{I}_t^m) \leq P(\nexists \tilde{t} \in \{t-m, ...t-1\} \text{ s.t. } D_{\tilde{t}} \leq 0 \text{ and } \tilde{D}_{\tilde{t}}^m \leq 0) \tag{15}$$

$$= E \prod_{i=t-m}^{t-1} (1 - I(D_i \leq 0) I(\tilde{D}_i^m \leq 0)) = E \prod_{i=t-m}^{t-1} I(D_i > 0 \text{ or } \tilde{D}_i^m > 0). \tag{16}$$

We observe that, by Lemma 5.1, if $a + \sigma z_{i-1} \leq 0$ then $D_i = D_{i-1} - a - \sigma z_{i-1}$ and $\tilde{D}_i = \tilde{D}_{i-1}^m - a - \sigma z_{i-1}$. We therefore bound the expression of Equation (16) by

$$E \prod_{i=t-m}^{t-1} [I(D_{i-1} - a - \sigma z_i > 0 \text{ or } \tilde{D}_{i-1}^m - a - \sigma z_i > 0; a + \sigma z_{i-1} \leq 0) \tag{17}$$

$$+I(a+\sigma z_{i-1}>0)].$$

Since D_j and $\tilde{D}_j^m$ both satisfy the same recurrence relationship of Equation (9) we observe that the proof of Lemma 2.1 also applies to $\tilde{D}_j^m$, and therefore we have that both $D_j \le W_j$ and $\tilde{D}_j^m \le W_j$. An upper bound for the above expression is then

$$E\prod_{i=t-m}^{t-1}[I(W_{i-1}-a-\sigma z_i>0; a+\sigma z_{i-1}\le 0)+I(a+\sigma z_{i-1}>0)]. \tag{18}$$

The terms in the product are not independent of each other, and so we will introduce a separation between them by introducing a sequence b_m such that $b_m \le m$ and $b_m \to \infty$ as $m \to \infty$. Since all product terms in Equation (18) equal 0 or 1, we can bound the expression by

$$E\prod_{j=0}^{[m/b_m]}[I(W_{t-m-1+jb_m}-a-\sigma z_{t-m+jb_m}>0; a+\sigma z_{t-m-1+jb_m}\le 0)+I(a+\sigma z_{t-m-1+jb_m}>0)]. \tag{19}$$

The motivation for this bound is that the product terms will be "nearly" independent for large enough b_m.

For $l \ge 1$ we now define $W_n^l = \max_{k\in\{0,\dots,l\}}[\eta - ak - \sigma\sum_{j=0}^{k-1} z_{n-j}]$. We observe that $W_n^l \in \mathcal{F}_{n-l+1}^n$ and that $W_n^\infty = W_n$ a.s.. The motivation for introducing W_n^l is that we can approximate $W_{t-m-1+jb_m}$ by $W_{t-m-1+jb_m}^{b_m}$ for large enough values of b_m. We therefore bound the previous expression by

$$E\prod_{j=0}^{[m/b_m]}[I(W_{t-m-1+jb_m}^{b_m}-a-\sigma z_{t-m+jb_m}>0; a+\sigma z_{t-m-1+jb_m}\le 0; W_{t-m-1+jb_m}^{b_m}=W_{t-m-1+jb_m}) \tag{20}$$

$$+I(a+\sigma z_{t-m-1+jb_m}>0)$$
$$+I(W_{t-m-1+jb_m}^{b_m}\ne W_{t-m-1+jb_m}; a+\sigma z_{t-m-1+jb_m}\le 0)].$$

We note that at most only one of these indicator functions is ever non-zero, implying that all product terms are 0 or 1. An upper bound for the above term is then

$$E\prod_{j=0}^{[m/b_m]}[I(W_{t-m-1+jb_m}^{b_m}-a-\sigma z_{t-m+jb_m}>0; a+\sigma z_{t-m-1+jb_m}\le 0; W_{t-m-1+jb_m}^{b_m}=W_{t-m-1+jb_m}) \tag{21}$$

$$+I(a+\sigma z_{t-m-1+jb_m}>0)]+E\sum_{j=0}^{[m/b_m]}I(W_{t-m-1+jb_m}^{b_m}\ne W_{t-m-1+jb_m}; a+\sigma z_{t-m-1+jb_m}\le 0).$$

This comes from the fact that $\prod_{i\in S}(I_{1i,}+I_{2i}+I_{3i}) \leq \prod_{i\in S}(I_{1i}+I_{2i}) + \sum_{i\in S} I_{3i}$ for indicator functions I_{ji} satisfying $I_{1i}+I_{2i}+I_{3i} \in \{0,1\}$. We bound the last term by

$$E \prod_{j=0}^{[m/b_m]} [I(W^{b_m}_{t-m-1+jb_m} - a - \sigma z_{t-m+jb_m} > 0; a + \sigma z_{t-m-1+jb_m} \leq 0) + I(a + \sigma z_{t-m-1+jb_m} > 0)]$$

$$+E \sum_{j=0}^{[m/b_m]} I(W^{b_m}_{t-m-1+jb_m} \neq W_{t-m-1+jb_m}).$$

By strict stationarity of the W^l_k terms this equals

$$[E(I(W^{b_m}_{t-m-1} - a - \sigma z_{t-m} > 0; a + \sigma z_{t-m-1+} \leq 0) + I(a + \sigma z_{t-m-1} > 0))]^{[m/b_m]+1}$$

$$+([m/b_m]+1)EI(W^{b_m}_{t-m-1} \neq W_{t-m-1}).$$

By the full support assumption on z_t and the fact that $W^l_n \leq W_n$ for any $l \geq 1$ we have that $E(I(W^{b_m}_{t-m-1} - a - \sigma z_{t-m} > 0; a + \sigma z_{t-m-1} \leq 0) + I(a + \sigma z_{t-m-1} > 0)) \leq E(I(W_{t-m-1} - a - \sigma z_{t-m} > 0; a + \sigma z_{t-m-1} \leq 0) + I(a + \sigma z_{t-m-1} > 0)) = \gamma \in (0,1)$. We note that γ does not depend upon m or t due to the strict stationarity of the terms.

The term $P(W^{b_m}_{t-m-1} \neq W_{t-m-1})$ has been considered before in the study of random walks with negative drift. Theorem 1 of Janson (1986) states that this term is $O(b_m^{-r})$ if $E|\min(0,z_t)|^{r+1} < \infty$ for $r > 0$. Therefore, the above expectation is of order

$$\gamma^{[m/b_m]} + [m/b_m]b_m^{-r}.$$

Letting $b_m = m^{1-\frac{\varepsilon}{1+r}}$ for $\varepsilon > 0$ achieves our desired result.

$\square$

Proof of Theorem 2.3:

We fix $t \in \mathbb{N}$. For each $m \geq 1$ we define the sequence $\tilde{D}^m_j$ for $j \in \{m-1, m, ...\}$ as in the proof of Theorem 2.2. Define the sequence $\tilde{\sigma}^m_j$ for $j \in \{m, m+1, ...\}$ by $\tilde{\sigma}^m_j = \sqrt{\sigma^2 + cI(\tilde{D}^m_{j-1} \leq 0)}$. We will now show that $\lim_{m\to\infty} \tilde{\sigma}^m_t = \tilde{\sigma}_t$ exists and that $\lim_{m\to\infty} \tilde{D}^m_t = \tilde{D}_t$ exists. These will then be strictly stationary because $\tilde{\sigma}_t = f(z_t, z_{t-1}, ...)$ and $\tilde{D}_t = g(z_t, z_{t-1}, ...)$ for some functions f and g.

We will show that the two sequences are Cauchy, i.e. that $\max_{k\geq m} |\tilde{\sigma}^k_t - \tilde{\sigma}^m_t|$ and $\max_{k\geq m} |\tilde{D}^k_t - \tilde{D}^m_t|$ both converge in probability to 0, implying that $\tilde{\sigma}^m_t$ and $\tilde{D}^m_t$ converge a.s. as $m \to \infty$. We will first show this for the sequence $\tilde{\sigma}^m_t$.

We observe that

$$\max_{k\geq m} |\tilde{\sigma}^k_t - \tilde{\sigma}^m_t| = \max_{k\geq m} \sqrt{|c|}\ |I(\tilde{D}^k_t \leq 0) - I(\tilde{D}^m_t \leq 0)|.$$

By the same argument as in Theorem 2.2, if there is some $\tilde{t} \in \{t-m-1, ..., t-1\}$ such that $\tilde{D}^k_{\tilde{t}} \leq 0$ for all $k \geq m$ then we will have that $\tilde{D}^k_t = \tilde{D}^m_t$ for all $k \geq m$, and therefore

$\max_{k\geq m}|I(\tilde{D}_t^k \leq 0) - I(\tilde{D}_t^m \leq 0)| = 0$. We also note that for all k we have $\tilde{D}_j^k \leq W_j$. Therefore, by similar reasoning as in Theorem 2.2,

$$P(\max_{k\geq m}|\tilde{\sigma}_t^k - \tilde{\sigma}_t^m| > 0) \leq P(\max_{k\geq m}\tilde{D}_j^k > 0;\ \text{for all } j \in \{t-m-1, ..., t-1\})$$

$$\leq E\prod_{j=t-m}^{t-1}[I(W_{j-1} - a - \sigma z_j > 0; a + \sigma z_{j-1} \leq 0) + I(a + \sigma z_{j-1} > 0)].$$

As seen in Theorem 2.2, this term is $O(m^{-r+\varepsilon})$. For the sequence $\tilde{D}_t^m$ we observe that the same method applies, since $\max_{k\geq m}|\tilde{D}_t^k - \tilde{D}_t^m|$ equals 0 whenever $\max_{k\geq m}\tilde{D}_j^k < 0$ for some $j \in \{t-m-1, ..., t-1\}$.

The uniqueness now follows from the fact that for any two strictly stationary solutions $(\sigma_{t,1}^*, D_{t,1}^*)$ and $(\sigma_{t,2}^*, D_{t,2}^*)$ we have that for any $m \geq 1$

$$P(\sigma_{t,1}^* \neq \sigma_{t,2}^*) \leq P(\sigma_{t,1}^* \neq \tilde{\sigma}_t^m) + P(\sigma_{t,2}^* \neq \tilde{\sigma}_t^m).$$

Letting $m \to \infty$ shows that the above probability is 0. The same argument holds for $P(D_{t,1}^* \neq D_{t,2}^*)$. □

Proof of Theorem 2.4:

The proof of this result is similar to Theorem 2.5 of Michel and de Jong (2018b), however since we consider i.i.d. errors we can obtain a slightly sharper rate.

By suitably enlarging the probability space we let z_t' be an independent sequence of random variables havng the same distribution of z_t. Define $\hat{z}_t = z_t I(t > 0) + z_t' I(t \leq 0)$; this sequence has the same distribution as z_t. Using the notation of Theorem 2.3 we define

$$\hat{\sigma}_t = f(\hat{z}_t, \hat{z}_{t-1}, ...)$$

and

$$\Delta\hat{Y}_t = \hat{\sigma}_t \hat{z}_t.$$

We observe that $\hat{\sigma}_t$ and $\Delta\hat{Y}_t$ are both strictly stationary solutions.

Let $\mathcal{G}_s^r = \sigma(\Delta Y_s^*, ..., \Delta Y_r^*)$. From the definition of $\Delta\hat{Y}_t$ it is clear that $\Delta\hat{Y}_t$ is independent of $\mathcal{G}_{-\infty}^0$ for all $t \in \mathbb{Z}$.

For $m \geq 1$ we define the event $E_m = \{\Delta Y_j^* = \Delta\hat{Y}_j,\ \text{for all } j \geq m\}$. From the proof of Theorem 2.2 we have that $P(E_m^C) = O(m^{-r+\varepsilon})$. We recall that the β-mixing coefficent is defined as

$$\beta_m = E\sup_{B\in\mathcal{G}_m^\infty}|P(B|\mathcal{G}_{-\infty}^0) - P(B)|.$$

For each $B = \{(\Delta Y_m^*, \Delta Y_{m+1}^*, ...) \in S\} \in \mathcal{G}_m^\infty$ we define $\hat{B} = \{(\Delta\hat{Y}_m, \Delta\hat{Y}_{m+1}, ...) \in S\}$. We then have the following two properties:

1. $B \cap E_m = \hat{B} \cap E_m$. This property follows from the definition of E_m.

2. $P(B) = P(\hat{B})$. This follows from the fact that ΔY_t and $\Delta\hat{Y}_t$ have the same distribution and from the definition of B and $\hat{B}$.

We now observe that

$$|P(B|\mathcal{G}_{-\infty}^0) - P(B)| = |P(B \cap E_m|\mathcal{G}_{-\infty}^0) + P(B \cap E_m^C|\mathcal{G}_{-\infty}^0) - P(B)|.$$

By our first property the above expression equals

$$|P(\hat{B} \cap E_m|\mathcal{G}_{-\infty}^0) + P(B \cap E_m^C|\mathcal{G}_{-\infty}^0) - P(B)|$$

$$= |P(\hat{B}|\mathcal{G}_{-\infty}^0) - P(\hat{B} \cap E_m^C|\mathcal{G}_{-\infty}^0) + P(B \cap E_m^C|\mathcal{G}_{-\infty}^0) - P(B)|$$
$$\leq |P(\hat{B}|\mathcal{G}_{-\infty}^0) - P(B)| + 2P(E_m^C|\mathcal{G}_{-\infty}^0).$$

By our second property the above expression equals

$$|P(\hat{B}|\mathcal{G}_{-\infty}^0) - P(\hat{B})| + 2P(E_m^C|\mathcal{G}_{-\infty}^0)|.$$

Therefore,

$$\beta_m \leq E \sup_{\hat{B} \in \sigma(\Delta\hat{Y}_m, \Delta\hat{Y}_{m+1}, \ldots)} |P(\hat{B}|\mathcal{G}_{-\infty}^0) - P(\hat{B})| + 2EP(E_m^C|\mathcal{G}_{-\infty}^0)$$

$$= E \sup_{\hat{B} \in \sigma(\Delta\hat{Y}_m, \Delta\hat{Y}_{m+1}, \ldots)} |P(\hat{B}|\mathcal{G}_{-\infty}^0) - P(\hat{B})| + 2P(E_m^C).$$

Since $\sigma(\Delta\hat{Y}_m, \Delta\hat{Y}_{m+1}, \ldots)$ is independent of $\mathcal{G}_{-\infty}^0$ and $P(E_m^C) = O(m^{-r+\varepsilon})$, the above expression is $O(m^{-r+\varepsilon})$. This completes our claim. $\square$

References

[1] Andrews, D.W.K. (1988). *Laws of large numbers for dependent non-identically distributed random variables.* Econometric Theory, 4, 458-467.

[2] Baccelli, F., and P. Brémaud. (1994). *Elements of queueing theory: Palm-Martingale calculus and stochastic recurrences.* Berlin: Springer-Verlag.

[3] Black, F (1976). *Studies of Stock Market Volatility Changes.* Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section, 177-181.

[4] Bollerslev, T. (1986). *Generalized autoregressive conditional heteroskedasticity.* Journal of Econometrics, 31, 307327.

[5] Bougerol, P. and N. Picard. (1992a). *Strict stationarity of generalized autoregressive processes.* Annals of Probability, 20, 1714-1730.

[6] Bougerol, P. and N. Picard. (1992b). *Stationarity of GARCH processes and of some nonnegative time series.* Journal of Econometrics, 52, 115-127.

[7] Engle, R.F. (1982). *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation.* Econometrica, 50, 987-1007.

[8] Hobson, D. G., and L.C.G. Rogers. (1998). *Complete models with stochastic volatility.* Mathematical Finance 8(1). 27-48.

[9] Janson, S. (1986). *Moments for first-passage and last-exit times, the minimum, and related quantities for random walks with positive drift.* Advances in Applied Probability. 18 (04), 865-879.

[10] Kiefer, J., and J. Wolfowitz. (1956). *On the Characteristics of the General Queueing Process, with Applications to Random Walk.* The Annals of Mathematical Statistics. 27 (1), 147-161.

[11] Michel, J. and R. M. de Jong. (2018a). *Anxious unit root processes.* Working Paper.

[12] Michel, J. and R.M. de Jong. (2018b). *Mixing properties of the dynamic Tobit model with mixing errors.* Economics Letters 162: 112-115.

[13] Meyn, S. P., and R. L. Tweedie. (2009). *Markov chains and stochastic stability.* Cambridge: Cambridge University Press.

[14] Nelson, D.B. (1991). *Conditional heteroskedasticity in asset returns: A new approach.* Econometrica, 59, 347-370.

[15] Pötscher, B.M. and I.R. Prucha. (1997). Dynamic nonlinear econometric models. Berlin: Springer-Verlag.

[16] Saïdi, Y. and J.M. Zakoïan. (2006). *Stationarity and geometric ergodicity of a class of nonlinear ARCH models.* Annals of Applied Probability, 16(4), 2256-2271.

[17] Sentana, E. (1995). *Quadratic ARCH Models.* The Review of Economic Studies, 62(4), 639-661.

[18] Zakoïan, J.M. (1994). *Threshold heteroskedastic models.* Journal of Economic Dynamics and Control, 18, 931-955.