

Finding the Stationary Distribution of ARCH Models Through Perfect Simulation

M.S. Carvalho and J.N. Corcoran
University of Colorado *

January 3, 2005

Abstract

We introduce a rapid algorithm for simulating from the stationary distribution of the ARCH(1) model based on many components of existing perfect simulation algorithms. The algorithm itself is based on the Harris coupler of Corcoran and Tweedie which is closely related to the multi-gamma coupler of Murdoch and Green. It exploits a curious monotonicity structure in the ARCH(1) model. Simulation results are provided.

1 Introduction

Autoregressive Conditional Heteroskedasticity (ARCH) models are used to model processes where changes in variance are a function of time. Such techniques have enjoyed wide popularity in financial applications, typically serving as models for asset price volatility. The ARCH econometric model for this, introduced by Engle [6], describes the variance of a time series as another time series. Specifically, it is modeled as an autoregressive (AR) process that is generally taken to be linear.

In this paper, we consider the class of autoregressive processes with ARCH(1) errors given by the stochastic difference equation

$$X_n = \alpha X_{n-1} + \epsilon_n \cdot \sqrt{\beta + \gamma X_{n-1}^2} \quad (1)$$

for some $\alpha \in \mathbb{R}$, $\beta, \gamma > 0$, and $\{\epsilon_n\}$ a sequence of independent and identically distributed (iid) errors.

If the ϵ_n have a density $f_\epsilon(\cdot)$ that is symmetric about zero with support on \mathbb{R} , then it is easy to see that $\{X_n\}$ form a Markov chain with transition density

$$p(x, y) = \frac{1}{\sqrt{\beta + \gamma x^2}} f_\epsilon \left(\frac{y - \alpha x}{\sqrt{\beta + \gamma x^2}} \right). \quad (2)$$

*Postal Address: Department of Applied Mathematics, University of Colorado, Box 526 Boulder CO 80309-0526, USA; email: corcoran@colorado.edu, marcio.carvalho@colorado.edu; phone: 303-492-0685

⁰Keywords: perfect simulation, Harris recurrence ARCH, stochastically dominating processes
AMS Subject classification: 60J10, 65C05, 62F25, 60K30

That is, given the Markov chain is at state x , it will move in one time step into the set $A \subset \mathbb{R}$ with probability $P(x, A)$ where

$$P(x, A) := \int_A p(x, y) dy.$$

For most applications of the ARCH(1) model, it is assumed that the error terms have a normal distribution centered at 0. Throughout this paper, we will assume that $\epsilon_n \stackrel{iid}{\sim} N(0, 1)$ although the techniques described can be extended to more general error distributions.

Our goal is to examine the stationary (invariant) measure of the Markov chain $\{X_n\}$ which we will denote by $\pi(\cdot)$. By definition, $\pi(\cdot)$ satisfies

$$\pi(A) = \int \pi(x) P(x, A) dx, \quad \forall A \in \mathcal{B} \tag{3}$$

where \mathcal{B} is the set of all Borel sets in \mathbb{R} .

For a Markov chain $\{X_n\}$ with the transition density given by (2), it is not possible to explicitly determine parameters for which a stationary distribution exists. Even in the case of normal errors, stationary inducing domains for α , β , and γ are not known except in a few cases. For example, Embrechts et al. [5] have shown existence of a stationary measure for $\alpha = 0$ and $\gamma \in (0, 2e^\Gamma)$ where Γ is Euler's constant. Borkovec and Klüppelberg [2] also give conditions on the existence of $\pi(\cdot)$.

In this paper, we describe a particular perfect simulation algorithm as it relates to this Markov chain to simulate the stationary distribution for parameter values where it is known to exist. This algorithm will also be a useful tool for studying the existence of a stationary distribution for other parameter domains.

2 Perfect Simulation

Perfect simulation, also known as perfect sampling, backward coupling, or coupling-from-the-past, is terminology used to describe a class of Markov chain Monte Carlo (MCMC) algorithms that are free of error caused by convergence issues. They are essentially algorithms that allow the simulator to “see infinity” in a finite amount of time. The essential idea is to find a random epoch $-T$ in the past such that, if we construct sample paths (according to a transition law $P(x, y)$ that is invariant for π) from every point in the state space starting at $-T$, then all paths will have coupled (met) successfully by time zero. The common value of the paths at time zero is a draw from π . Intuitively, it is clear why this result holds with such a random time T . For consider a chain starting at $-\infty$ with the stationary distribution π . At every iteration it maintains the distribution π . But at time $-T$ it must pick *some* value x , and from then on it follows the trajectory from that value. But of course it arrives at the same place at time zero no matter what value x is picked at time $-T$: so the value returned by the algorithm at time zero must itself be a draw from π .

Perfect sampling algorithms can be particularly efficient if the chain is *stochastically monotone* in the sense that paths from lower starting points stay below paths from higher starting

points. In this case, one need only couple sample paths from the “top” and “bottom” of the space, as all other paths will be sandwiched in between. It is possible to generalize one step further to monotone chains on an unbounded state space by considering *stochastically dominating* processes (see Kendall [7]) to bound the journeys of sample paths.

Though the ideas behind perfect simulation existed, (see eg. Asmussen, Glynn, and Thorisson [1]), it was Propp and Wilson [10] who introduced these ideas in a simulational setting. Since this seminal 1996 paper, many variations of the algorithm appeared, addressing details such as coupling on a continuous state space, devising random processes to bound the process of interest on an unbounded state space, and restructuring of the state space in order to realize monotonicity. In this paper, we use a particular algorithm of Corcoran and Tweedie [4] known as the “Harris Coupler” in order to achieve a perfect simulation of the stationary distribution of the ARCH(1) model.

2.1 The Harris Coupler

The Harris Coupler, applicable to *Harris recurrent* chains, was developed by Corcoran and Tweedie [4]. It is a variation of the “multi-gamma coupler”, previously developed by Murdoch and Green [9].

A chain with transition law $P(x, A)$ and stationary measure $\pi(\cdot)$ is called Harris recurrent if, for every A with $\pi(A) > 0$ we have

$$P(X_n \in A \text{ infinitely often}) = 1$$

for all x in the state space of the chain.

A set C is called *small* if the k -step transition law $P^{(k)}(x, A) := P(X_k \in A | X_0 = x)$ is “minorized” for some $0 < \delta < 1$ and for some density ψ in the sense that

$$P^{(k)}(x, A) \geq \delta \Psi(A), \quad x \in C \tag{4}$$

where $\Psi(A) = \int_A \psi(y) dy$. It is known that for aperiodic Harris chains, every set with $\pi(A) > 0$ contains a small set of positive π -measure [8]. We may also consider (4) in terms of densities:

$$p^{(k)}(x, y) \geq \delta \psi(y), \quad x \in C,$$

where $p^{(k)}(x, y)$ is the density corresponding to $P^{(k)}(x, A)$.

Borkovec and Klüppelberg [2] have shown that the Markov chain described by the ARCH(1) model given by (1) is Harris recurrent with $k = 1$ for interval sets of the form $C = [-c, c]$. (Note that when $k = 1$, aperiodicity is ensured and this is sometimes known as the *strongly aperiodic* case [8].)

We may simulate sample paths of a Markov chain satisfying (4) with $k = 1$ as follows. Suppose that the chain is in state x at some time n (ie: $X_n = x$), and we wish to simulate a value y for the chain at time $n + 1$.

- If $x \notin C$, we produce y using the transition law $P(x, dy)$.

- If $x \in C$,
 - with probability δ , we produce y by simulating a value drawn from $\psi(dy)$,
 - with probability $1 - \delta$, we produce y by simulating a value drawn from the “residual” transition law $R(x, dy) := [P(x, dy) - \delta\psi(dy)]/(1 - \delta)$.

It is easy to see that X_{n+1} has the correct marginal distribution with this construction. Note that if $x \in C$, then with probability δ we draw the next value from the distribution with density ψ regardless of the specific value of x . Therefore, in the context of perfect simulation, we can achieve a coupling of all possible sample paths with probability δ in the case that all possible sample paths simultaneously enter C .

3 Perfect Simulation for the ARCH(1) Model

For the purpose of illustration, we consider the ARCH(1) model given by

$$X_n = \epsilon_n \cdot \sqrt{\beta + \gamma X_{n-1}^2} \tag{5}$$

for some $\beta, \gamma > 0$, and $\{\epsilon_n\}$ a sequence of independent and identically distributed normal errors each with mean 0 and variance 1.

For ease of algorithm exposition, we will, in this section, make the (unsatisfactory) assumption that the state space is restricted to an interval $[-K, K]$ where $K > 0$ is presumably large. Specifically, if $X_n > K$, we take $X_n = K$ and if $X_n < -K$, we take $X_n = -K$. In Section 6, we remove this assumption.

3.1 Monotonicity

A Markov chain with transition density $p(x, y)$ and transition law

$$P(x, A) := \int_A p(x, y) dy$$

is said to be *stochastically monotone* if

$$x_1 \leq x_2 \quad \text{implies that} \quad P(x_1, [y, \infty)) \leq P(x_2, [y, \infty)).$$

In terms of simulation, this means that if two sample paths are run from points x_1 and x_2 where $x_1 \leq x_2$, they will maintain their ordering and will not cross. In terms of perfect simulation, where the goal is to follow trajectories of all possible sample paths, stochastic monotonicity allows one to follow the progress of only two bounding sample paths, since all other paths will be “sandwiched” in between.

Suppose that $x_n^{(1)} < x_n^{(2)}$ represent the positions of two sample paths at time n . A natural way to update the sample paths of the ARCH(1) model given by (5) to values at time $n + 1$ would be to generate a value of $\epsilon_{n+1} \sim N(0, 1)$ and to set $x_{n+1}^{(1)} = \epsilon_{n+1} \cdot \sqrt{\beta + \gamma(x_n^{(1)})^2}$

and $x_{n+1}^{(2)} = \epsilon_{n+1} \cdot \sqrt{\beta + \gamma(x_n^{(2)})^2}$. This, however, does not necessarily result in the ordering $x_{n+1}^{(1)} < x_{n+1}^{(2)}$. However, we do, in fact, have a “sandwiching property”. Define, for fixed ϵ_{n+1} ,

$$f_{n+1}(x) = \epsilon_{n+1} \cdot \sqrt{\beta + \gamma x^2},$$

we have that for **any** $x \in [-K, K]$, either

$$0 < f_{n+1}(0) < f_{n+1}(x) < f_{n+1}(K) \quad \text{or} \quad f_{n+1}(K) < f_{n+1}(x) < f_{n+1}(0) < 0.$$

Furthermore, for $0 < x_n^{(1)} < x_n^{(2)} < x_n^{(3)}$ or $x_n^{(1)} < x_n^{(2)} < x_n^{(3)} < 0$, application of f_{n+1} will again result in a value $f_{n+1}(x_n^{(2)})$ that is sandwiched between the values of $f_{n+1}(x_n^{(1)})$ and $f_{n+1}(x_n^{(3)})$, regardless of the ordering of $f_{n+1}(x_n^{(1)})$ and $f_{n+1}(x_n^{(3)})$. So, in order to achieve a coupling of all possible sample paths on $[-K, K]$, we may follow only two paths forward in time. That is, at each time $-n$, $n = 1, 2, \dots$, we will start sample paths in state 0 and in state K until we have reached $n = T$ where T is the first time such that paths started at time $-T$ in states 0 and K end up in the same place at time 0.

3.2 Minorization

We consider a small set of the form $C = [-c, c]$ where $0 < c < K$. The goal is to find a density $\psi(\cdot)$ and a number $0 < \delta < 1$ such that the minorization condition

$$p(x, dy) \geq \delta \cdot \psi(dy)$$

holds uniformly for all $x \in C$.

Recall that $p(x, y)$ is a normal density with mean 0 and variance $\beta + \gamma x^2$. In order to achieve a minorization, we will essentially “push” the “tallest” normal density with mean zero and variance β down, at the origin, to the level at the origin of the “shortest” normal density with mean zero and variance $\beta + \gamma c^2$ for $x \in C$. (See Figure 1.) This “shortest” density is the normal density with mean zero and variance $\beta + \gamma c^2$. That is, we will multiply the $N(0, \beta)$ density by a factor of $\sqrt{2\pi\beta}/\sqrt{2\pi(\beta + \gamma c^2)} = \sqrt{\beta/(\beta + \gamma c^2)}$. It is simple to check, for all $x \in C = [-c, c]$, that

$$p(x, y) \geq \sqrt{\frac{\beta}{\beta + \gamma c^2}} \psi(y)$$

where $\psi(y)$ is the $N(0, \beta)$ density. Hence, we take

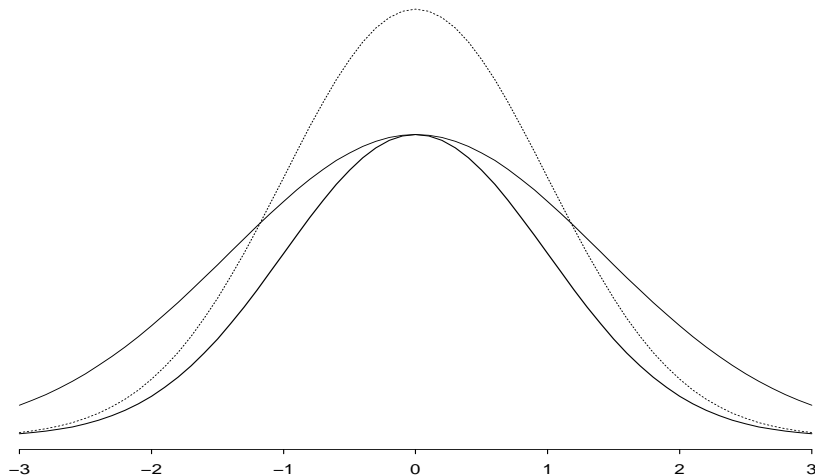
$$\delta = \sqrt{\frac{\beta}{\beta + \gamma c^2}}.$$

3.3 Residual Density

The residual density from which values must be drawn with probability $1 - \delta$ is

$$r(x, y) = \frac{p(x, y) - \delta \cdot \psi(y)}{1 - \delta}$$

Figure 1: Visualization of a Minorization



The tallest curve (dashed line) is the $N(0, \beta)$ density. The middle curve is the $N(0, \beta + \gamma c^2)$ density which is the “shortest” normal density with mean zero and variance $\beta + \gamma x^2$ for $x \in C$. The lowest curve (bold line) is the $N(0, \beta)$ multiplied by $\sqrt{\beta/(\beta + \gamma c^2)}$.

where $p(x, y)$ is the normal density with mean 0 and variance $\beta + \gamma x^2$.

We can easily compute its cumulative distribution function(cdf) with respect to y , obtaining

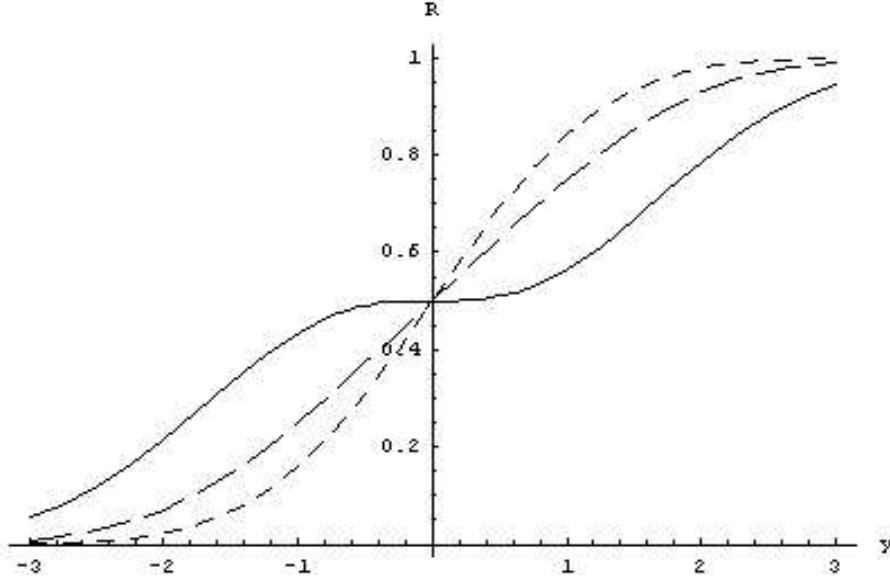
$$R(x, y) = \frac{1}{1 - \delta} \left(\Phi\left(\frac{y}{\sqrt{\beta + \gamma x^2}}\right) - \delta \cdot \Phi\left(\frac{y}{\sqrt{\beta}}\right) \right)$$

where $\Phi(\cdot)$ is the cdf for a standard normal.

To draw a value y from this density for a fixed x , we can use a root finding routine to compute the inverse of this function given some realization v of a uniform random variable V on $(0, 1)$. Thus, $y = R^{-1}(x, v)$. There are better ways to simulate from this density, (for example via an accept/reject algorithm), but the inverse cdf method is the easiest way to preserve the practicable monotonicity of the residual density.

For typical values of $c = \beta = \gamma = 1$, which imply $\delta = \sqrt{2}/2$, we can generate plots of $R(x, y)$ for different values of $x \in [-1, 1]$. Figure 2 shows the curves for $R(0, y)$ (small dash), $R(0.5, y)$ (long dash), $R(1, y)$ (solid line). Given any random number V on $(0, 1)$, we see from the plot that the closer x is to 0, the closer the resulting y is to 0. This suggests that the residual density possesses a sandwich property.

Figure 2: Residual cdf: $R(0, y)$ (small dash), $R(0.5, y)$ (long dash), $R(1, y)$ (solid line)



3.3.1 Residual Monotonicity

In the case where the transition law $P(x, A)$ is stochastically monotone, it is easy to verify that the residual transition law

$$R(x, A) = \int_A r(x, y) dy = \frac{1}{1 - \delta} \int_A [p(x, y) - \delta\varphi(y)] dy$$

is also stochastically monotone.

In the case of the ARCH(1) model, we do not have monotonicity of $P(x, A)$, but we do have the sandwiching property described in Section 3.1. Recall that all possible sample paths are always between the path started from K and the path started from zero. The difference here is that these paths may themselves be crossing. We now investigate formally how this behavior translates to the residual density.

Proposition 3.1 *For values $x \in C$, the residual density $r(x, y)$ possesses the same sandwich property as f_{n+1} .*

Proof:

In order to prove this proposition, we only need to show that $y^{(1)} = R^{-1}(x^{(1)}, v)$ and $y^{(2)} = R^{-1}(x^{(2)}, v)$ imply $y^{(1)} \leq y^{(2)}$ for the case where $1/2 < v < 1$ and $0 < x^{(1)} < x^{(2)} \leq c$. All other cases can be shown analogously. So, given $1/2 < v < 1$ and $0 < x^{(1)} < x^{(2)} \leq c$, the equalities $y^{(1)} = R^{-1}(x^{(1)}, v)$ and $y^{(2)} = R^{-1}(x^{(2)}, v)$ imply that $v = R(x^{(1)}, y^{(1)}) =$

$R(x^{(2)}, y^{(2)})$ and that $y^{(1)}, y^{(2)} > 0$. In generic terms,

$$\frac{1}{2} < v = R(x, y) = \frac{1}{1 - \delta} \left(\Phi\left(\frac{y}{\sqrt{\beta + \gamma x^2}}\right) - \delta \cdot \Phi\left(\frac{y}{\sqrt{\beta}}\right) \right)$$

Suppose that in the above, x increased to x^+ but y decreased to y^- . Then, the value of both Φ integrals above would decrease, but the first one would decrease even more due to two facts: one, its integral's upper limit would have a reduced numerator and increased denominator; two, since the upper limit of the first integral is greater than 0 but less than the upper limit of the second, the value of the Gaussian integrand is greater near the upper limit of the first integral than near the upper limit of the second. Therefore, because of the preceding two facts, the loss of value (area) will be greater in the first Φ integral. So, let the decrease in the first one be denoted by $d^+ > 0$ and the decrease in the second one by $d^- > 0$. Then,

$$R(x^+, y^-) = \frac{1}{1 - \delta} \left(\Phi\left(\frac{y}{\sqrt{\beta + \gamma x^2}}\right) - d^+ + \delta \cdot d^- - \delta \cdot \Phi\left(\frac{y}{\sqrt{\beta}}\right) \right)$$

Since $0 < \delta < 1$ and $d^+ > d^- > 0$, we have that $d^+ > \delta \cdot d^-$, which implies $-d^+ + \delta \cdot d^- < 0$. Therefore, $R(x, y) > R(x^+, y^-)$. So, if the equality $R(x^{(1)}, y^{(1)}) = R(x^{(2)}, y^{(2)})$ must be satisfied for $0 < x^{(1)} < x^{(2)} \leq c$, we cannot have $y^{(2)} < y^{(1)}$. Hence, we must have $y^{(1)} \leq y^{(2)}$, which proves the proposition.

4 Perfect Simulation Algorithm for the ARCH(1) Model

In order to simulate from (5), we need to choose a value $c > 0$ to define the small set used in the minorization. Ideally we choose a value for c that makes the set C large enough so that paths will visit it often, and, at the same time, a value for c that will result in a suitably large value of δ , which is the probability that paths will couple if they are all inside C . In this Section, we restrict the state space to $[-K, K]$ for some given $K > 0$.

Since the state space is uncountable, we will only be monitoring the paths generated by the following two processes: Y_n^+ and Y_n^- , which are, respectively, upper and lower bounding processes to the original X_n . Since all possible values of X_n are in between Y_n^+ and Y_n^- , once these bounding processes couple at some value x_t , we must have that $X_t = x_t$, and from there on, we just need to follow the process X_n from time $n = t$ to time $n = 0$. The value $X_0 = x_0$ will be the resulting draw from the stationary distribution for the restricted ARCH(1) model.

Also, in order to move the processes forward, we will need realizations of the following random variables (with $n = 0, -1, -2, -3, \dots$):

- ϵ_n with a $N(0, 1)$ distribution
- W_n with a $N(0, \beta)$ distribution
- U_n with a uniform distribution on $(0, 1)$

- V_n with a uniform distribution on $(0, 1)$

We generate these realizations as needed, storing them for reuse for forward sweeps from negative time starting points.

We will start our simulation at $n = -2$, by assigning $Y_{-2}^+ = K$ and $Y_{-2}^- = -K$. We then move both processes forward, as described in the following subsections, until $n = 0$. If coupling occurs, (ie: if we at any time draw a value from ψ), the final value $X_0 = x_0$ is the value returned by our algorithm as a draw from the stationary distribution. If coupling did not occur, we start again, but this time at $n = -3$. We once again move the processes forward, and if no coupling occurred by $n = 0$, we keep starting the simulation one step further back, running it forward until $n = 0$. Once coupling occurs, the final value $X_0 = x_0$ is our perfect simulation output. The minimum number of backward time steps to achieve this coupling at or before time zero is our backward coupling time.

4.1 Moving the bounding processes forward before coupling

Before the upper and lower bounding processes have coupled, we can only have one of the following cases:

Case 1: $-K \leq Y_n^- < -c$ and $c < Y_n^+ \leq K$

Case 2: $-K \leq Y_n^- < -c$ and $0 \leq Y_n^+ \leq c$

Case 3: $-c \leq Y_n^- \leq 0$ and $c < Y_n^+ \leq K$

Case 4: $-c \leq Y_n^- \leq 0$ and $0 \leq Y_n^+ \leq c$

(Case 4 is the only case where coupling can occur by moving one step forward)

Moving the processes forward is not as simple as moving the upper and lower processes forward. We must take into account the possible positions of all paths that are sandwiched between the bounding processes. Thus, moving the processes forward involves computing a list of candidate values, and then assigning the largest candidate to Y_{n+1}^+ and the smallest candidate to Y_{n+1}^- . Computing the candidate values depends on which case above we are at step n . A description for each one follows, with y_i representing the i^{th} candidate.

4.1.1 Moving under Case 1

At step n , our chain could be anywhere in between Y_n^- and Y_n^+ , but since not all paths are inside C , coupling cannot occur. Due to monotonicity, any path in between Y_n^- and $-c$ must end up in between $f_{n+1}(Y_n^-)$ and 0. Also, any path in between c and Y_n^+ must end up in between $f_{n+1}(Y_n^+)$ and 0. We must not forget that paths in $[-c, 0]$ will end up either, with probability δ , at a specific location W_{n+1} , or, with probability $1 - \delta$, sandwiched between 0 and the value obtained from the residual distribution applied to $-c$. The same must not be forgotten for paths in $[0, c]$. Finally, we restrict the paths to the interval $[-K, K]$. Taking into account all these details, we can compute candidate values for the bounding processes at step $n + 1$.

We will let $y_1 = f_{n+1}(Y_n^-)$. If $U_{n+1} \leq \delta$, then $y_2 = W_{n+1}$, otherwise, $y_2 = R^{-1}(-c, V_{n+1})$. Also, we will let $y_3 = f_{n+1}(Y_n^+)$. And, if $U_{n+1} \leq \delta$, then $y_4 = W_{n+1}$, otherwise, $y_4 = R^{-1}(c, V_{n+1})$. Now, with our four candidates, we are ready to proceed:

- If all our four candidates are greater than or equal to 0, we let

$$Y_{n+1}^- = 0 \text{ and } Y_{n+1}^+ = \min(K, \max(y_1, y_2, y_3, y_4)).$$

- If all our four candidates are less than or equal to 0, we let

$$Y_{n+1}^- = \max(-K, \min(y_1, y_2, y_3, y_4)) \text{ and } Y_{n+1}^+ = 0.$$

- Otherwise,

$$Y_{n+1}^- = \max(-K, \min(y_1, y_2, y_3, y_4)) \quad \text{and} \quad Y_{n+1}^+ = \min(K, \max(y_1, y_2, y_3, y_4)).$$

4.1.2 Moving under Case 2

At step n , our chain could be anywhere in between Y_n^- and Y_n^+ , but since not all paths are inside C , coupling cannot occur. Due to monotonicity, any path in between Y_n^- and $-c$ must end up in between $f_{n+1}(Y_n^-)$ and 0. We must not forget that paths in $[0, Y_n^+]$ will end up either, with probability δ , at a specific location W_{n+1} , or, with probability $1 - \delta$, sandwiched between 0 and the value obtained from the residual distribution applied to Y_n^+ . The same is true for paths in $[-c, 0]$. As in Case 1, we restrict all paths to the interval $[-K, K]$. Taking into account all these details, we can compute candidate values for the bounding processes at step $n + 1$.

We will let $y_1 = f_{n+1}(Y_n^-)$. If $U_{n+1} \leq \delta$, then $y_2 = W_{n+1}$, otherwise, $y_2 = R^{-1}(-c, V_{n+1})$. Also, if $U_{n+1} \leq \delta$, then $y_4 = W_{n+1}$, otherwise, $y_4 = R^{-1}(Y_n^+, V_{n+1})$. Now, with our three candidates, we are ready to proceed:

- If all our three candidates are greater than or equal to 0, we let

$$Y_{n+1}^- = 0 \text{ and } Y_{n+1}^+ = \min(K, \max(y_1, y_2, y_4)).$$

- If all our three candidates are less than or equal to 0, we must

$$Y_{n+1}^- = \max(-K, \min(y_1, y_2, y_4)) \text{ and } Y_{n+1}^+ = 0.$$

- Otherwise,

$$Y_{n+1}^- = \max(-K, \min(y_1, y_2, y_4)) \text{ and } Y_{n+1}^+ = \min(K, \max(y_1, y_2, y_4)).$$

4.1.3 Moving under Case 3

At step n , our chain could be anywhere in between Y_n^- and Y_n^+ , but since not all paths are inside C , coupling cannot occur. Due to monotonicity, any path in between c and Y_n^+ must end up in between $f_{n+1}(Y_n^+)$ and 0. We must not forget that paths in $[Y_n^-, 0]$ will end up either, with probability δ , at a specific location W_{n+1} , or, with probability $1 - \delta$, sandwiched between 0 and the value obtained from the residual distribution applied to Y_n^- . The same is true for paths in $[0, c]$. Again we restrict paths to the interval $[-K, K]$. Taking into account all these details, we can compute candidate values for the bounding processes at step $n + 1$.

If $U_{n+1} \leq \delta$, then $y_2 = W_{n+1}$, otherwise, $y_2 = R^{-1}(Y_n^-, V_{n+1})$. Also, we will let $y_3 = f_{n+1}(Y_n^+)$. If $U_{n+1} \leq \delta$, then $y_4 = W_{n+1}$, otherwise, $y_4 = R^{-1}(c, V_{n+1})$. Now, with our three candidates, we are ready to proceed:

- If all our three candidates are greater than or equal to 0, we let

$$Y_{n+1}^- = 0 \text{ and } Y_{n+1}^+ = \min(K, \max(y_2, y_3, y_4)).$$

- If all our three candidates are less than or equal to 0, we let

$$Y_{n+1}^- = \max(-K, \min(y_2, y_3, y_4)) \text{ and } Y_{n+1}^+ = 0.$$

- Otherwise,

$$Y_{n+1}^- = \max(-K, \min(y_2, y_3, y_4)) \text{ and } Y_{n+1}^+ = \min(K, \max(y_2, y_3, y_4)).$$

4.1.4 Moving under Case 4 (coupling may occur)

At step n , our chain could be anywhere in between Y_n^- and Y_n^+ , but this time all paths are inside C , thus, coupling can occur and will occur if $U_{n+1} \leq \delta$. Therefore, $X_{n+1} = Y_{n+1}^- = Y_{n+1}^+ = \max(-K, \min(W_{n+1}, K))$. From this point, we need not continue moving the bounding processes forward. We only need to move the original chain from $n + 1$ until 0, according to the rules described in the next subsection.

If coupling did not occur, then we let our candidates be $y_2 = R^{-1}(Y_n^-, V_{n+1})$ and $y_4 = R^{-1}(Y_n^+, V_{n+1})$. Now, with these two candidates, we are ready to proceed:

- If both our candidates are greater than or equal to 0, we let

$$Y_{n+1}^- = 0 \text{ and } Y_{n+1}^+ = \min(K, \max(y_2, y_4)).$$

- If both our candidates are less than or equal to 0, we let

$$Y_{n+1}^- = \max(-K, \min(y_2, y_4)) \text{ and } Y_{n+1}^+ = 0.$$

- Otherwise,

$$Y_{n+1}^- = \max(-K, \min(y_2, y_4)) \text{ and } Y_{n+1}^+ = \min(K, \max(y_2, y_4)).$$

4.2 Moving the chain forward after coupling occurred

Once coupling occurs at step $n = m$, we finish by moving the original chain X_n forward from $n = m$ until $n = 0$. The value obtained for X_0 will be a perfect draw from the stationary distribution of our process of interest. The rules to move the chain forward at this stage are:

- If $X_n \notin C$, then, $X_{n+1} = \max(-K, \min(f_{n+1}(X_n), K))$
- If $X_n \in C$ and $U_{n+1} \leq \delta$, then, $X_{n+1} = \max(-K, \min(W_{n+1}, K))$
- Otherwise, $X_{n+1} = \max(-K, \min(R^{-1}(X_n, V_{n+1}), K))$

5 Sample case for $\beta = 1$ and $\gamma = 1$

Looking back at the model given by (5), we know that solutions exist for $0 \leq \beta$ and $0 \leq \gamma < 3.56$. Due to symmetry, $E[X] = 0$. Trying to compute the second moment, we get that for the second moment to exist, we need $\gamma < 1$, as shown below.

$$E[X^2] = E[\epsilon^2]E[\beta + \gamma X^2] \Rightarrow E[X^2] = \beta + \gamma E[X^2] \Rightarrow E[X^2] = \frac{\beta}{1 - \gamma}$$

The following are trivial cases that can be solved analytically.

- If $\beta > 0$ and $\gamma = 0$, we have that $\pi \sim N(0, \beta)$.
- If $\beta = 0$ and $0 \leq \gamma < 1$, we have that $\pi(x) = \delta_0$ (Dirac's Delta at 0), because both first and second moments are equal to zero.

In order to solve for the non trivial cases, we need to resort to simulation. We can obtain fast and accurate results with the algorithm described in the previous section.

We applied the algorithm described in the previous section to find the stationary distribution of $X_{n+1} = \epsilon_{n+1}\sqrt{1 + X_n^2}$. We simulated 100,000 draws in order to construct a histogram that will have the same shape and form of the density function π of the stationary distribution.

We have chosen (rather arbitrarily) $K = 10$ and $c = 1$. This choice of c results in a value $\delta = \sqrt{2}/2$. As expected, a larger c will result in a smaller δ . That is, if we increase the size of the set in which the paths may couple, we decrease the probability that they will couple.

After 100,000 draws, we were able to construct the histogram in Figure 3, representing the shape of the stationary distribution of the restricted ARCH(1) model.

Figure 4 shows the histogram for the backwards coupling time for the 100,000 represented in Figure 3. The minimum backward coupling time was 116, with a mean was 11.4. This

Figure 3: 100,000 Draws from π for $\alpha = 0$, $\beta = 1$, $\gamma = 1$, $c = 1$, $K = 10$

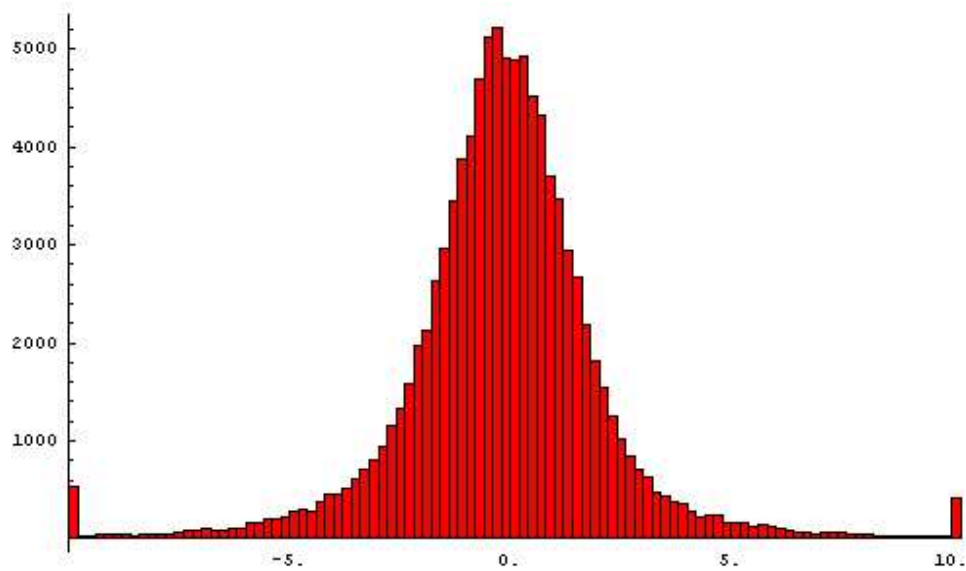
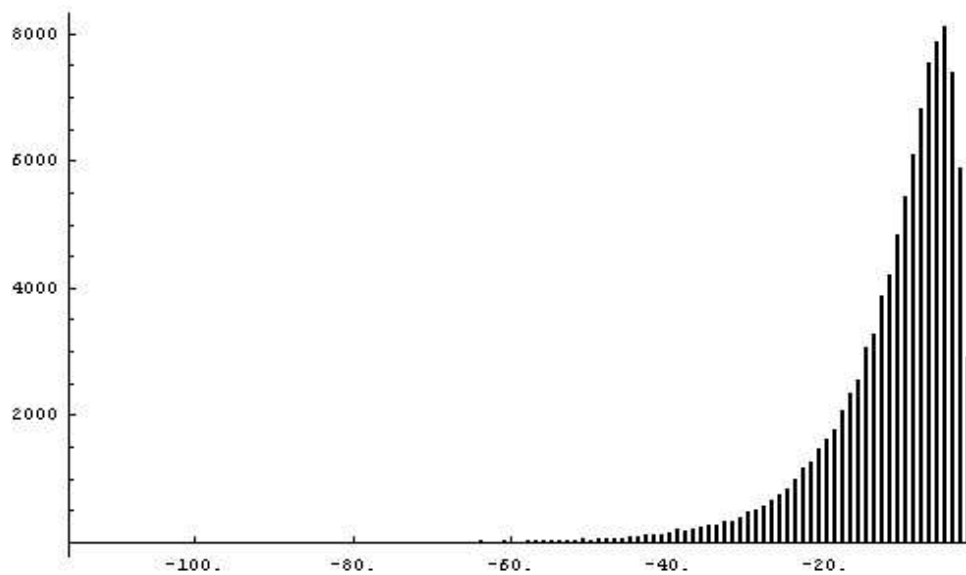


Figure 4: The Distribution of the Backward Coupling Time for the Simulation of π



tells us that on average we need less than 12 backward steps to obtain a perfect draw, with the worst case needing 116 steps for a draw to be obtained.

A simulation was also run for the trivial case $\beta = 1$ and $\gamma = 0$ using 100,000 draws with $c = 1$ and $K = 10$. The empirical cdf obtained was a perfect match with that of a standard normal distribution, as expected.

6 Removing the artificial boundaries

In order to be able to simulate the process with an unbounded state space, we must find some distribution with cdf G such that $\forall x \geq 0, G(x) \leq F(x)$, where F is the cdf of the unknown stationary distribution π .

6.1 A Suggested CDF G

Finding some distribution with cdf G such that for all $x \geq 0, G(x) \leq F(x)$, where F is the cdf of the unknown stationary distribution π is not an easy task.

If $\gamma < 1$, our process $\{X_n\}$ satisfies the conditions of Theorem 8 in Borkovec and Kluppelberg [X], and we have that

$$F(x) \sim 1 - ax^{-k}, \text{ as } x \rightarrow \infty,$$

where

$$a = \frac{1}{2k} \frac{E[|\epsilon\sqrt{\beta + \gamma X^2}|^k - |\epsilon\sqrt{\gamma X^2}|^k]}{E[|\epsilon\sqrt{\gamma}|^k \ln |\epsilon\sqrt{\gamma}|]}$$

and k is the unique positive solution to

$$E[|\epsilon\sqrt{\gamma}|^k] = 1 \Rightarrow \frac{1}{\sqrt{\pi}} (\sqrt{2\gamma})^k \Gamma\left(\frac{k+1}{2}\right) = 1 \quad (6)$$

Equation (6) above can be solved numerically for k . As examples of solutions, $\gamma = 0.5 \Rightarrow k = 4.7303$ and $\gamma = 0.9 \Rightarrow k = 2.3043$.

Once we have k , the problem lies in solving for a . The denominator in the expression for a is easy to handle, but the same cannot be said about the numerator. Since we are looking for some function G that is less than F , it is sufficient come up with an upper bound value b for a and use b to construct G . If we let \tilde{k} be the smallest integer greater than or equal to k , and as long as $\beta \geq \gamma$, we have that

$$\begin{aligned} E[|\epsilon\sqrt{\beta + \gamma X^2}|^k - |\epsilon\sqrt{\gamma X^2}|^k] &\leq E[|\epsilon|^k] \cdot E[|\sqrt{\beta} + \sqrt{\gamma}|X|^k - |\sqrt{\gamma}|X|^k] = \\ &\leq \gamma^{k/2} E[|\epsilon|^k] \cdot E[(\sqrt{\beta/\gamma} + |X|)^k - |X|^k] \end{aligned}$$

Therefore

$$a \leq \frac{1}{2k} \frac{E[|\epsilon|^k] \cdot E[(\sqrt{\beta/\gamma} + |X|)^{\tilde{k}} - |X|^{\tilde{k}}]}{E[|\epsilon|^k \ln |\epsilon\sqrt{\gamma}|]}$$

If we define $G(x) = 1 - bx^{-k}$, we have that $G(x) \leq F(x)$ for large values of x . Since b is any upper bound to a , we can use the above expression to come up with a value for b , as shown in the following example.

Example

Consider the process $X_{n+1} = \epsilon_{n+1} \sqrt{0.9 + 0.9X_n^2}$. Then, $\beta = \gamma = 0.9 \Rightarrow k = 2.30434 \Rightarrow \tilde{k} = 3$. Also, $E[X^2] = 0.9/0.1 = 9$. Thus,

$$\begin{aligned} E[|\epsilon|^k] \cdot E[(\sqrt{\beta/\gamma} + |X|)^{\tilde{k}} - |X|^{\tilde{k}}] &= E[|\epsilon|^k] \cdot E[1 + 3|X| + 3X^2] \leq \\ &\leq E[|\epsilon|^k] \cdot (1 + 3\sqrt{E[X^2]} + 3E[X^2]) \end{aligned}$$

with the last inequality due to Jensen's inequality. Hence, we have

$$a \leq \frac{1}{2k} \frac{E[|\epsilon|^k] \cdot (1 + 9 + 27)}{E[|\epsilon|^k \ln |\epsilon \sqrt{\gamma}|]} \leq 21.1917$$

therefore, by letting $b = 21.2$, we have that $G(x) \leq F(x)$ for large values of x .

Back to the general case, if we are to draw values from G , we will need its inverse, which turns out to be $G^{-1}(x) = (b/(1-x))^{1/k}$. Since we are only drawing from a uniform (0.5,1) to apply the inverse, the function G need only be defined on $[(2b)^{1/k}, \infty)$. Thus, we came up with the following suggested function G to be used in Subsection 6.1:

$$G(x) = 1 - bx^{-k}, \text{ with } x \in [(2b)^{1/k}, \infty), \text{ and } b \text{ some upper bound for } a.$$

We know that the above will have the property $G(x) \leq F(x)$ for "large" values of x , and we make the conjecture that it will hold for all $x \in [(2b)^{1/k}, \infty)$. Empirical studies strongly support this conjecture.

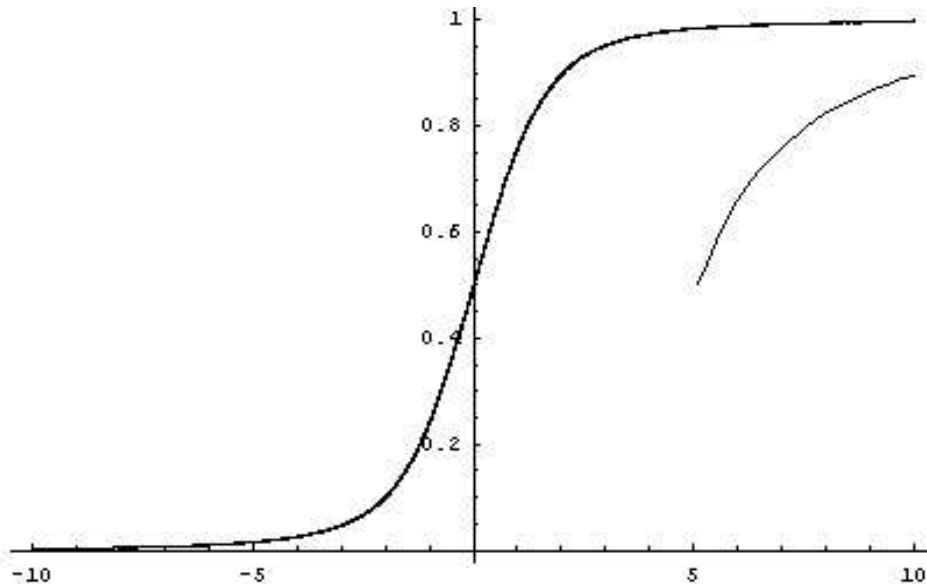
Figure 5 shows the plot (bold) of the cdf obtained by simulating 100,000 draws from the process in the example above ($\beta = \gamma = 0.9$) with $c = 1$ and $K = 10$. It also shows the curve (not bold) for the function $G(x)$ with b as suggested in the example above. We already know that $G(x) \leq F(x)$ for large values of x , and from what we see in Figure 5, further affirms that the relation will also hold for all $x \geq (2b)^{1/k}$.

6.2 An Empirical Unrestricted Perfect Simulation Algorithm

The approach to simulate with an unbounded state space will be an empirical one, as described in this subsection.

We start by simulating the desired process as in Section 4 with the appropriate values for β and γ , making sure they respect the conditions $0 < \gamma < 1$ and $\beta \geq \gamma$. These conditions are necessary for the properties of G , as suggested in the previous subsection, to hold. Reasonable values for c and K must be chosen.

Figure 5: Simulated Curves F (full curve) and G (partial curve)



Once we have our preliminary simulation results, we compute G and compare it to the empirical cdf obtained from the simulation in order to check that G will be below the cdf not only for large x , but also for all other values of x .

We now should look at the histogram, average and largest value of the collection of backwards coupling times from our preliminary simulation. This will tell us how many steps backwards were needed at most in order to achieve coupling and give us an idea of how likely it is for it to take so many more steps back than the largest one observed.

With this information at hand, we will run a second simulation for the same process, this time without the restriction that the paths are always inside $[-K, K]$. The algorithm will be similar to the one described in Section 4, with just a few slight modifications.

We will begin each simulation at some fixed step $t < 0$. We must be confident that coupling will be achieved before we get to step 0. Thus, we will estimate what the value of t should be according to the results obtained for the backward coupling time in the preliminary simulation. We don't want t too large in magnitude, but if it is too close to 0, then we may not achieve coupling in all simulated draws, and the simulation results won't be perfect.

With t chosen, we will draw values for Y_t^- and Y_t^+ from the distribution G .

When moving the bounding processes forward from t until coupling is achieved or step 0 is reached, we no longer need to restrict our bounding paths to $[-K, K]$. The same applies to moving the chain X_n after coupling occurs. Thus, we show below how to move the original chain forward after coupling occurs. By comparing the below with the rules from Subsection 4.2, one can see how to modify the rules described in Subsection 4.1, in order to free the algorithm from the restriction of all paths being inside $[-K, K]$.

- If $X_n \notin C$, then, $X_{n+1} = f_{n+1}(X_n)$
- If $X_n \in C$ and $U_{n+1} \leq \delta$, then, $X_{n+1} = W_{n+1}$
- Otherwise, $X_{n+1} = R^{-1}(X_n, V_{n+1})$

If we are able to successfully draw from each particular simulation (i.e. coupling is achieved in every single simulation), we then have a perfect simulation of the stationary distribution.

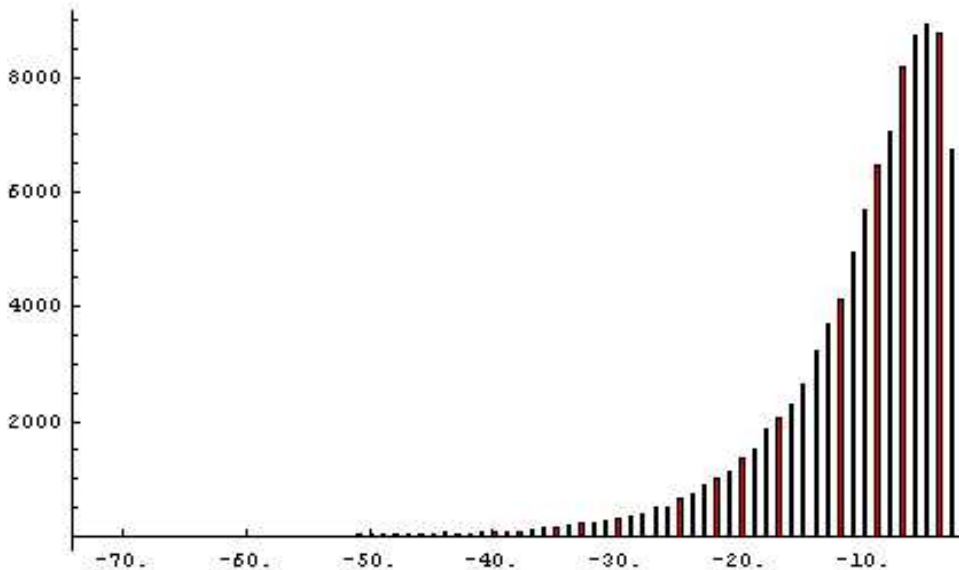
6.3 Sample case for $\beta = 0.9$ and $\gamma = 0.9$

We now use the empirical approach described in the previous subsection to find the stationary distribution of the process considered in the example in Subsection 6.1 ($\beta = 0.9$ and $\gamma = 0.9$).

The preliminary simulation has already been performed (with $c = 1$ and $K = 10$), and in Subsection 6.1 we saw that the suggested $G(x) = 1 - 21.2x^{-2.30434}$ seems to be well below the true cdf for the stationary distribution.

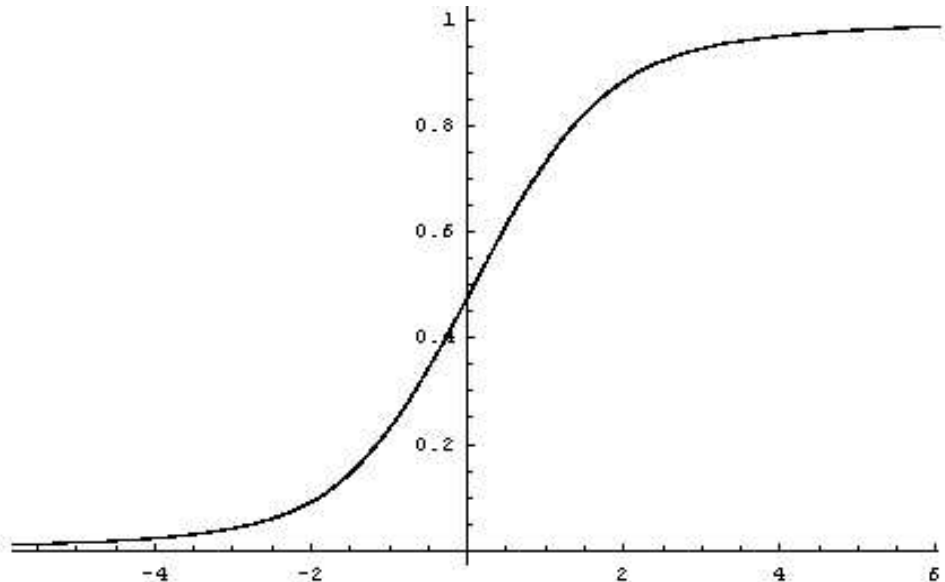
Figure 6 shows the histogram of the backward coupling time for the preliminary simulation. The farther back we had to go in all 100,000 simulations in order to achieve coupling was 74 steps. Based on the shape of the histogram, it looks unlikely that we might need to go back more than 100 steps to achieve coupling. But to be on the safe side, we will run our second simulation (without bounds) starting the upper and lower bounding processes at $t = -200$.

Figure 6: Simulated Distribution of the Backward Coupling Time



Applying the algorithm from the previous section 100,000 times with $t = -200$ gives us the empirical cdf for the stationary distribution shown in Figure 7.

Figure 7: Simulation Output for the CDF of π



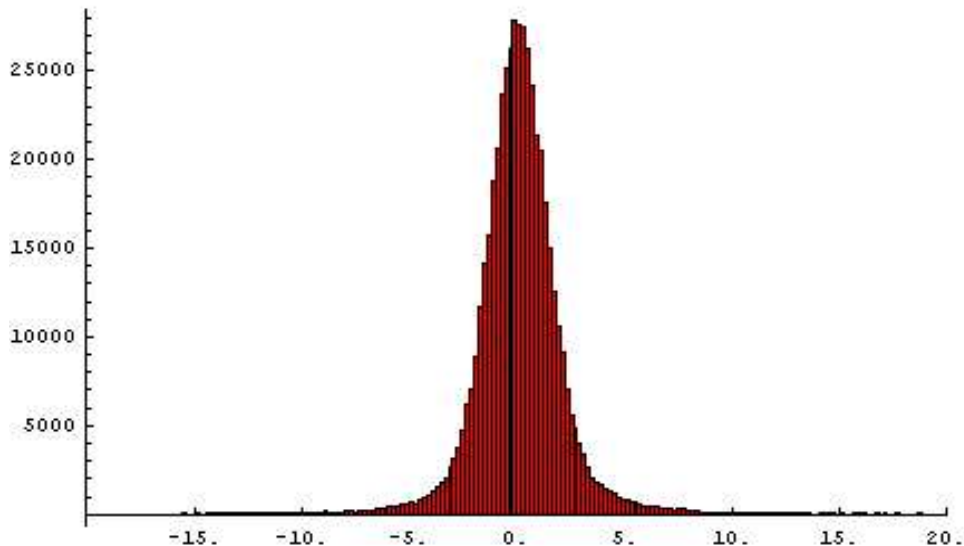
Overlaying the empirical cdf computed with bounds and without bounds does not show much of a difference. But we definitely see better results with the empirical approach. Obviously, with the empirical approach we were able to obtain extreme draws, such as 135, 117, -108, -191. Also, as shown in the example in Subsection 6.1, the second moment of the process is 9. Thus, one would expect a sample standard deviation close to 3. In the preliminary simulation, the sample standard deviation was 2.00 (due to the chain being artificially bounded by $[-K, K]$). For comparison, the sample standard deviation obtained in the second (unbounded) simulation was 2.80.

Figure 8 shows a histogram of the sample obtained with the empirical (and unbounded) approach. The shape of the histogram is that of the density function of the stationary distribution. Although the same shape could have been obtained with the bounded simulation, the empirical approach gives us better results when we are interested in tabulating values for the cdf or when exact draws from the stationary distribution are needed for some experiment.

7 Concluding Remarks

We have introduced a rapid algorithm for simulating from the stationary distribution of the ARCH(1) model based on many components of existing perfect simulation algorithms. Future work includes a simulation exploration of the parameter space in order to identify additional regions where the ARCH(1) stationary distribution exists. Possible extensions of the techniques described in this paper include application to higher dimensional ARCH models (ARCH(p)) models and to generalized ARCH (GARCH) models. Additionally, we

Figure 8: 100,000 Draws from π for $\alpha = 0, \beta = 1, \gamma = 1, c = 1, K = \infty$



intend to obtain a proof of the conjecture that the distribution given by the cdf G in Section 6.1 stochastically dominates that of the target distribution. An obvious question to be posed at this point is that of why we should bother to implement a perfect simulation algorithm for the ARCH model at all given that we are, at this point, forced to use a conjecture and an empirical algorithm. There is mounting empirical evidence that such “psuedo-perfect” backward coupling algorithms outperform (see for example [3]) traditional approaches in terms of both speed and accuracy and are therefore worthwhile endeavors.

References

- [1] S. Asmussen, P. Glynn, and H. Thorisson. Stationary detection in the initial transient problem. *ACM Transactions on Modeling and Computer Simulation*, 2(2):130–157, 1992.
- [2] M. Borkovec and C. Klüppelberg. The tail of the stationary distribution of an autoregressive process with arch(1) errors. *The Annals of Applied Probability*, 00:000–000, 2001.
- [3] J.N. Corcoran and U. Schneider. Pseudo-perfect and adaptive variants of the metropolis-hasting algorithm with an independent candidate density. *Journal of Statistical Computation and Simulation*, to appear, 2004.
- [4] J.N. Corcoran and R.L. Tweedie. Perfect sampling of ergodic harris chains. *Ann. App. Prob.*, 11(2):438–451, 2001.

- [5] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, Heidelberg, 1997.
- [6] R.F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50:987–1008, 1982.
- [7] W.S. Kendall. Perfect simulation for the area-interaction point process. In L. Accardi and C.C. Heyde, editors, *Probability Towards the Year 2000*, pages 218–234. Springer, New York, 1998.
- [8] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, London, 1993.
- [9] D.J. Murdoch and P.J. Green. Exact sampling from a continuous state space. *Scandinavian Journal of Statistics*, 25:483–502, 1998.
- [10] J.G. Propp and D.B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9:223–252, 1996.