Anders Rahbek and Rasmus Søndergaard Pedersen
University of Copenhagen October 2020

Part VI
Stochastic Volatility: An Introduction

VI.1 Introduction

The models we have discussed so far have all been defined by specifying explicitly the conditional distribution of the present observation (log-return), x_t, given past observations $x_{t-1}, \ldots, x_1$ as in the rich class of GARCH models, where

$$x_t = \sigma_t z_t,$$

with z_t *i.i.d.* $(0,1)$ and σ_t^2 a function of the past observations $x_{t-1}, \ldots, x_1$.

Now we focus on stochastic volatility (SV) models where σ_t (or σ_t^2) is an unobservable, or so-called latent, exogenous process. Typically we will assume that (σ_t) satisfies

$$(\sigma_t | \sigma_{t-1}, \sigma_{t-2}, ...) \overset{d}{=} (\sigma_t | \sigma_{t-1}), \tag{VI.1}$$

that is, it is a Markov chain. We will assume further that the volatility process (σ_t) is independent of the z_t's in the specification of x_t. Note that this excludes modeling the leverage effect as in e.g. the GJR-ARCH model, where negative returns are allowed to have a different effect on σ_t than the positive.

Example VI.1 *Given an* $i.i.d.\,(0,1)$ *sequence* (z_t)*, we say that* (x_t) *is a stochastic volatility (SV) model if*

$$x_t = \sigma_t z_t,$$

where (σ_t) *is a stochastic process that is independent of* (z_t) *and* $Var(\sigma_t | x_{t-1}, ...) > 0$*. The volatility* σ_t *satisfies (VI.1) and* σ_t *is positive and real-valued* ($\sigma_t \in \mathbb{R}_+$)*, or* σ_t *takes values in some finite and discrete set, say* $A = \{h_1, h_2\}$*. Note that ARCH processes do not fit into this definition as* σ_t *depends on* x_{t-1} *and therefore on* z_{t-1}. □

In the following sections we go through some issues related to the use of SV models using that SV models are examples of nonlinear time series driven by exogenous Markov chains as described above. The positive message is that dependence (weak mixing, geometric ergodicity) properties of the observable series x_t may usually be addressed only by studying the latent chain (σ_t). This makes it relatively easy to derive asymptotic properties of estimators based on method of moments.

However, as we shall see, the likelihood function is not directly computable and therefore maximum likelihood analysis for these kind of models is not straightforward. We discuss some general filtering techniques that allow us to evaluate the log-likelihood function (approximately).

VI.2 Stochastic volatility models

We list below some consequences of the definition of the SV-model given in Example VI.1. It is assumed that the innovations (z_t) follow a distribution with a strictly positive density, f, on $\mathbb{R}$ as in the case of Gaussian or Student's t-distributed z_t.

Property 1: If (σ_t) is (strictly) stationary so is (x_t) as by definition $x_t = \sigma_t z_t$ and (σ_t) and (z_t) are independent.

Property 2: If (σ_t) is stationary then x_t follows a mixture distribution of the density f with respect to the invariant distribution of (σ_t), cf. Example VI.2.2.

Property 3: The observations (x_t) are uncorrelated – provided of course that correlations are well-defined, or equivalently, second order moments exist.

Property 4: Given, or conditional on, $(\sigma_T, \ldots, \sigma_1)$, then x_t and x_{t+h} are independent. The conditional distribution of x_t given $(\sigma_T, \ldots, \sigma_1)$, denoted $x_t|\sigma_T, \ldots, \sigma_1$, is equal to the distribution of $x_t|\sigma_t$.

These properties characterize the (*hidden Markov) process* (x_t, σ_t).

Our main focus is on two classical examples of a stochastic volatility models, introduced by Taylor and discussed extensively in Taylor (1986, 2005). Much like the development of ARCH models, the purpose was to construct a time series model with a changing volatility that was able to match marginal distributions and correlation structures found in typical financial time series. Taking as offset the multiplicative model of the form

$$x_t = \sigma_t z_t,$$

modelling (σ_t) as an autoregressive model turned out to be a suitable compromise between simplicity and the property of replicating empirical findings of real data. In addition, the case where σ_t can take only two (or more) different values corresponding to states, is also of interest, both from a theoretical and an empirical point of view.

VI.2.1 The log-normal SV model

The first example of a stochastic volatility model is the log-normal SV model as introduced in Taylor (1986). Here the observations are given by

$$x_t = \sigma_t z_t = \exp(\log(\sigma_t)) z_t, \tag{VI.2}$$

where (z_t) is $i.i.d.\mathrm{N}(0,1)$ and independent of the volatility process (σ_t), which is unobservable. It is further assumed that $h_t = \log(\sigma_t)$ follows a first-order autoregressive (AR(1)) process,

$$h_t = (1-\phi)\alpha + \phi h_{t-1} + \eta_t,$$

with (η_t) $i.i.d.\mathrm{N}\big(0, \sigma_\eta^2\big)$, $\sigma_\eta^2 = \beta^2(1-\phi^2)$, and $\alpha, \beta \in \mathbb{R}$, $\phi \in (-1,1)$. Since the distribution of h_t is Gaussian (when h_1 is Gaussian) then the volatility $\sigma_t = \exp(h_t)$ marginally follows a log-normal distribution leading to the terminology *log-normal SV* model, or simply *log-SV*, for the model defined by (VI.2). We refer to Part V for additional details about and properties of the log-normal distribution. □

VI.2.2 A two-state SV model

Consider again the process in Example VI.1 given by

$$x_t = \sigma_t z_t, \quad t = 1, 2, \ldots$$

where (z_t) is $i.i.d.\mathrm{N}(0,1)$ and independent of the volatility process (σ_t). In the log-normal SV model, the unobserved volatility σ_t is continuous taking positive real values, $\sigma_t \in \mathbb{R}_+$, whereas in the two-state SV model it can take only two positve values, say h_1 and h_2, corresponding to *states* 1 and 2 respectively. In general, SV models can be formulated for any finite number of states, but to keep things simple (and because the two-state case is widely applied) we focus on the two-state case here.

Let s_t be a process that takes values in $\{1,2\}$. A classic *mixture Gaussian model* is obtained by letting

$$\sigma_t = h_{s_t}$$

and letting s_t be *i.i.d.* with $P(s_t = 1) = 1 - P(s_t = 2) = p$, that is

$$(s_t | s_{t-1}, ..., s_1) \stackrel{d}{=} s_t.$$

However, this does not replicate the realizations of real-world log-returns x_t very well, as the model essentially rules out volatility clustering due to the memoryless state variable (and hence volatility). Instead we consider the popular Markov switching SV model where s_t is a Markov chain (see below), in which

$$(s_t | s_{t-1}, ..., s_1) \stackrel{d}{=} (s_t | s_{t-1}).$$

A different class of models such as the threshold or the so-called ACR (autoregressive conditional root) volatility models, is given by letting s_t depend on past returns, that is for example,

$$(s_t | s_{t-1}, ..., s_1, x_{t-1}, ...x_1) \stackrel{d}{=} (s_t | x_{t-1}).$$

However, such observation switching models violates our initial set-up where σ_t is independent of z_t and we discuss this elsewhere.

VI.2.3 A note on Markov chains

Let s_t be a random variable that takes values in $\{1, 2\}$. The process $(s_t)_{t=1,2,...}$ is a Markov chain,

$$(s_t | s_{t-1}, ..., s_1) \stackrel{d}{=} (s_t | s_{t-1}), \tag{VI.3}$$

and we specify the switches between states 1 and 2 by the parameters $(p_{ij})_{i,j=1,2}$

$$p_{ij} := P(s_t = j | s_{t-1} = i) = \text{"}P_{\text{begin, end state}}\text{"}.$$

By construction, it follows that

$$p_{11} + p_{12} = P(s_t = 1 | s_{t-1} = 1) + P(s_t = 2 | s_{t-1} = 1) = 1,$$

and likewise $p_{22} + p_{12} = 1$. When discussing the properties of s_t it is convenient to collect the transition probabilities in the so-called transition matrix P,

$$P = \begin{pmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{pmatrix} = \begin{pmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{pmatrix}.$$

VI.2.3.1 Some simple considerations and calculations

A way to interpret this, which may also be useful for simulations of s_t, given a known p_{ij}, is to let S_t (as opposed to s_t) be a bivariate variable with two values corresponding to s_t :

$$S_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ if } s_t = 1, \quad \text{and} \quad S_t = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ if } s_t = 2. \tag{VI.4}$$

Then, with say $s_t = 1$, one finds by definition,

$$\begin{aligned} P\left(s_{t+1} = 1 | s_t = 1\right) &= p_{11} \\ P\left(s_{t+1} = 2 | s_t = 1\right) &= p_{12} = 1 - p_{11} \end{aligned}$$

or simply,

$$\begin{pmatrix} P\left(s_{t+1} = 1 | s_t = 1\right) \\ P\left(s_{t+1} = 2 | s_t = 1\right) \end{pmatrix} = PS_t = \begin{pmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} p_{11} \\ 1 - p_{11} \end{pmatrix}$$

Thus if $s_t = 1$, then $s_{t+1} = 1$ with probability p_{11} and $s_{t+1} = 2$ with probability $1 - p_{11} = p_{12}$.

Next, using repeatedly that

$$P\left(X \in A | Y \in B\right) = \frac{P\left(X \in A, Y \in B\right)}{P\left(Y \in B\right)}$$

together with (VI.3), we can find the probability that s_t when initiated in $s_t = 1$ ends in state 1 after two steps,

$$\begin{aligned} &P\left(s_{t+2} = 1 | s_t = 1\right) \\ &= P\left(s_{t+2} = 1, s_t = 1\right) / P\left(s_t = 1\right) \\ &= \left(P\left(s_{t+2} = 1, s_{t+1} = 1, s_t = 1\right) + P\left(s_{t+2} = 1, s_{t+1} = 2, s_t = 1\right)\right) / P\left(s_t = 1\right) \\ &= P\left(s_{t+2} = 1 | s_{t+1} = 1\right) P\left(s_{t+1} = 1 | s_t = 1\right) + P\left(s_{t+2} = 1 | s_{t+1} = 2\right) P\left(s_{t+1} = 2 | s_t = 1\right) \\ &= p_{11} p_{11} + p_{21} p_{12}. \end{aligned}$$

Similarly,

$$P\left(s_{t+2} = 2 | s_t = 1\right) = p_{12} p_{11} + p_{22} p_{12}.$$

We note that the same could have been obtained by simply using the transition matrix P as follows,

$$\begin{pmatrix} P\left(s_{t+2} = 1 | s_t = 1\right) \\ P\left(s_{t+2} = 2 | s_t = 1\right) \end{pmatrix} = P\left(PS_t\right) = P^2 S_t.$$

Likewise, for any k number of steps,

$$\begin{pmatrix} P\left(s_{t+k}=1|s_t\right) \\ P\left(s_{t+k}=2|s_t\right) \end{pmatrix} = P^k S_t \tag{VI.5}$$

with S_t defined in (VI.4). That is, we can compute the probability for s_{t+k} taking values 1 or 2 given any starting value for s_t by simple *multiplication* of the transition matrix P.

VI.2.3.2 Weak mixing and Geometric Ergodicity

The regularity conditions we used for establishing that the ARCH processes, x_t say, were weakly mixing, were that the transition density, $f\left(x_t|x_{t-1}\right)$ was sufficiently well-behaved (i.e. continuous or in the case of threshold models, positive and bounded on intervals). Analogously when studying the properties of the finite state Markov chain s_t we need the transition matrix P to have certain properties. These properties, as implied by the transition matrix P for the finite state Markov chain, namely *irreducibility* and *aperiodicity* (both explained below), are exactly the ones implied by the conditions on the transition density $f\left(\cdot|\cdot\right)$ for the x_t Markov chain with a 'continuum of states'.

First of all, we need that s_t when entering state 1 (or 2) is not staying there. This would be implied by for example $p_{11}=1$ such that $p_{12}=0$, in which case the chain is said to have an absorbing state 1 (as it would never enter state 2 again) and s_t is a *reducible* Markov chain. We need the chain to be *irreducible,* that is given any starting state $s_t=i$, then for some k,

$$P\left(s_{t+k}=j|s_t=i\right)>0,$$

for any $j=1,2$. This is implied by the condition that,

$$p_{11},\ p_{22}<1. \tag{VI.6}$$

Secondly, we need that the irreducible Markov chain is not *periodic* as would be the case if $p_{21}=p_{12}=1$, or

$$P=\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

In this case with $s_t=1$, then $s_{t+2}=s_{t+4}=...=1$ while $s_{t+1}=s_{t+3}=...=2$. That is, s_t with a 'period of 2' returns to 1 (and to 2 with the same period). For the irreducible chain to also be *aperiodic,* we need in addition to (VI.6) that,

$$p_{11}+p_{22}>0.$$

Next, application of the drift-criterion to the irreducible and aperiodic chain is vacuous (as it would be also for the case where s_t has any finite number of states) and as a result the two conditions imply that s_t is weakly mixing (all moments finite), or geometrically ergodic.

Alternatively, a simple way to state the conditions is that, given that the irreducibility condition (VI.6) holds, the eigenvalues of P satisfy $\lambda_1 = 1$ and $|\lambda_2| < 1$. To see this, observe that the characteristic equation of P is

$$\det(\lambda I_2 - P) = (\lambda - 1)(\lambda + 1 - (p_{11} + p_{22})) = 0.$$

Note that the fact that each column of P sums to one, implies that one eigenvalue equals unity.

VI.2.3.3 Stationary distribution

Recall that weakly mixing implies that there exists a stationary distribution for the process considered. As in the AR(1) case, where this can be found directly, this can also be found directly for the 2-state Markov chain. The stationary distribution of s_t is characterized by the unconditional probabilities,

$$v = \begin{pmatrix} P(s_t = 1) \\ P(s_t = 2) \end{pmatrix} = \begin{pmatrix} P(s_t = 1) \\ 1 - P(s_t = 1) \end{pmatrix} = \begin{pmatrix} P(s_{t+k} = 1) \\ P(s_{t+k} = 2) \end{pmatrix},$$

for all k. In particular, we have by definition (recall also the way we compute moments for the AR(1) process when we impose stationarity),

$$\begin{aligned} \begin{pmatrix} P(s_{t+1} = 1) \\ P(s_{t+2} = 2) \end{pmatrix} &= \begin{pmatrix} P(s_{t+1} = 1 | s_t = 1) P(s_t = 1) + P(s_{t+1} = 1 | s_t = 2) P(s_t = 2) \\ P(s_{t+1} = 2 | s_t = 1) P(s_t = 1) + P(s_{t+1} = 2 | s_t = 2) P(s_t = 2) \end{pmatrix} \\ &= \begin{pmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{pmatrix} \begin{pmatrix} P(s_t = 1) \\ P(s_t = 2) \end{pmatrix} \\ &= P \begin{pmatrix} P(s_t = 1) \\ P(s_t = 2) \end{pmatrix}, \end{aligned}$$

or with $v = (P(s_t = 1), P(s_t = 2))'$,

$$v = Pv.$$

That is v, is the eigenvector of P corresponding to the eigenvalue $\lambda_1 = 1$ of P. Tedious calculations give that

$$v = \tfrac{1}{2 - p_{11} - p_{22}} \begin{pmatrix} 1 - p_{22} \\ 1 - p_{11} \end{pmatrix}. \tag{VI.7}$$

This can be useful for defining starting values in an algorithm for obtaining the MLE of p_{ij}.

Note that it can be shown that also (the Perron–Frobenius theorem),

$$\lim_{k\to\infty} P^k = v\,(1,1)\,,$$

which is a (2×2) matrix of rank one, with each column containing the stationary invariant distribution.

VI.2.3.4 Markov chain with N states

We end by briefly stating a summary for the case where $s_t \in \{1,2,...,N\}$ for some finite N. For this general case we define irreducibility as follows: For any $i,j=1,..,N$ there exists an integer m_{ij} (potentially depending on i and j) such that

$$P\left(s_{t+m_{ij}} = i|s_t = j\right) > 0.$$

Again defining the transition matrix, P by (observe the transpose),

$$P' = (p_{ij})_{i,j=1,2,...,N}\,,$$

then s_t is weakly mixing provided that it is irreducible and that the eigenvalues of P satisfy, $\lambda_1 = 1$, and $|\lambda_i| < 1$ for $i = 2,...,N$. The stationary distribution is similarly given by the eigenvector v corresponding to $\lambda_1 = 1$, or equivalently,

$$\lim_{k\to\infty} P^k = v\,(1,1,...,1)\,,$$

where $v = (P(s_t = 1),...P(s_t = N))'$. Specifically, with $0_{N\times 1} = (0,\ldots,0)'$ and $\iota = (1,\ldots,1)'$ respectively N-dimensional column vectors of zeros and ones, and noting that $\iota' v = 1$, it holds that

$$v = (A'A)^{-1}A'\begin{pmatrix} 0_{N\times 1} \\ 1 \end{pmatrix}, \tag{VI.8}$$

where $A = (I_N - P', \iota)'$.

VI.3 Weakly mixing and the drift criterion for SV models

In order to discuss the properties of the return process $x_t = \sigma_t z_t$, we first observe that x_t itself is not a Markov chain while for the joint evolution of the return process x_t and the volatility σ_t, we have by assumption,

$$((x_t,\sigma_t)\,|\,(x_{t-1},\sigma_{t-1})\,,...,(x_1,\sigma_1)) \overset{d}{=} ((x_t,\sigma_t)\,|\,(x_{t-1},\sigma_{t-1}))\,.$$

And moreover, the joint distribution of (x_t, σ_t) given (x_{t-1}, σ_{t-1}) is characterized by the distributions of

$$(x_t|\sigma_t) \quad \text{and} \quad (\sigma_t|\sigma_{t-1}).$$

In the case where both have densities, as in the log-normal SV model, this can be written in terms of densities as,

$$f((x_t, \sigma_t) \,|\, (x_{t-1}, \sigma_{t-1})) = f(x_t|\sigma_t) \, f(\sigma_t|\sigma_{t-1}).$$

Thus if we establish that σ_t is weakly mixing, then x_t should 'inherit the properties'. More precisely, the joint process (x_t, σ_t) should be weakly mixing, provided that σ_t is.

VI.3.1 Two-state SV model

Turn first the case where σ_t is switching between two states as given by the s_t process. We saw that verification of weak mixing for the two-state – and indeed N-state – case of σ_t was straightforward reflecting that σ_t could only take a finite number of values.

It then follows by Chen and Carrasco (2002, Proposition 4) that the joint Markov process, (x_t, σ_t) is weakly mixing if P satisfies the regularity conditions just discussed, see also Genon-Catalot et al. (2002) and Meitz and Saikkonen (2008).

VI.3.2 Log-SV models

The alternative log-normal SV model is an example of a SV model where σ_t is real and positive, $\sigma_t \in \mathbb{R}_+ = (0, \infty)$, as opposed to the finite state case. We could proceed as for the two-state SV model and quote results in the just mentioned reference Carrasco and Chen (2002) and also Meitz and Saikkonen (2008) where it is shown that under general conditions that if σ_t satisfies some drift criterion, so does (x_t, σ_t).

However, we can establish this ourselves directly using our developed theory. That is, we can apply the drift criterion in Part I, Assumptions I.3.1-I.3.2 and Theorem I.3.2. In order to allow for some flexibility in the choice of models for σ_t we shall formulate the result such that the log-normal case is a specific example. Hence we restate the SV model as:

Assumption VI.1 *Consider the SV model as given by*

$$x_t = \sigma_t z_t, \quad t = 1, 2, ...T.$$

Assume (i) that z_t is i.i.d.$(0,1)$ with some continuous density f_z. Assume furthermore, (ii) that the unobservable σ_t is real and positive, $\sigma_t \in \mathbb{R}_+$, with

$$(\sigma_t | \sigma_{t-1}, \sigma_{t-2}, ...) \overset{d}{=} (\sigma_t | \sigma_{t-1}),$$

and such that the transition density $f_\sigma(\sigma_t|\sigma_{t-1})$ is continuous. Finally, assume (iii) that σ_t and $(z_1, \ldots, z_t)$ are independent.

Now we shall assume that we have used a drift criterion with a drift function δ to establish that σ_t is weakly mixing - this is often straightforward as we shall also see below for the log-normal case.

Theorem VI.1 *Let $x_t = \sigma_t z_t$ be a SV model satisfying Assumption VI.1, with σ_t a Markov chain on $\mathbb{R}_+ = (0, \infty)$ which satisfies a drift criterion with drift function δ. The drift function is assumed to be bounded on closed intervals of the form $[b, B]$ in $\mathbb{R}_+$ and such that*

$$\lim_{\sigma \to \infty} \delta(\sigma) = \infty \qquad \text{and} \qquad \lim_{\sigma \to 0} \delta(\sigma) = \infty.$$

Then (x_t, σ_t) satisfies the drift criterion with drift function,

$$d(x, \sigma) = \delta(\sigma) + x^2/\sigma^2. \tag{VI.9}$$

Hence (x_t, σ_t) is weakly mixing and geometric ergodic, provided Assumption I.3.1 in Part I[1] holds for the joint process (x_t, σ_t).

Proof:
First of all, by Assumption VI.1, (x_t, σ_t) satisfies Assumption I.3.1 since

$$((x_t, \sigma_t) \,|\, (x_{t-1}, \sigma_{t-1}), ..., (x_1, \sigma_1)) \overset{d}{=} ((x_t, \sigma_t) \,|\, (x_{t-1}, \sigma_{t-1})),$$

and the joint distribution of (x_t, σ_t) given (x_{t-1}, σ_{t-1}) is characterized by

$$f((x_t, \sigma_t) \,|\, (x_{t-1}, \sigma_{t-1})) = f(x_t|\sigma_t) f_\sigma(\sigma_t|\sigma_{t-1}) = \frac{1}{|\sigma_t|} f_z\left(\frac{x_t}{\sigma_t}\right) f_\sigma(\sigma_t|\sigma_{t-1}).$$

Next, with $\delta : \mathbb{R} \to [1, \infty]$ the drift function for σ_t, we note that for there are $0 < m < M < \infty$, $\in \mathbb{R}_+$, such that $E[\delta(\sigma_t)|\sigma_{t-1} = \sigma] \leq C$ for $\sigma \in [m, M]$, while

$$E[\delta(\sigma_t)|\sigma_{t-1} = \sigma] \leq \phi\delta(\sigma),$$

[1]or a similar condition as Assumption II.1.1

for $\sigma \notin [m, M]$ and $\phi < 1$, see Assumption I.3.2.

Applying the proposed drift function $d(\cdot)$ for (x_t, σ_t) we get,

$$\begin{aligned} & E[d(x_t, \sigma_t)|(x_{t-1}, \sigma_{t-1}) = (x, \sigma)] \\ & = E[\delta(\sigma_t) + x_t^2/\sigma_t^2|(x_{t-1}, \sigma_{t-1})] \\ & = E[\delta(\sigma_t)|\sigma_{t-1} = \sigma] + E\left(z_t^2\right) \\ & \leq \phi\delta(\sigma) + C + 1 \\ & = \left[\frac{\phi\delta(\sigma) + C + 1}{\delta(\sigma) + x^2/\sigma^2}\right] d(x, \sigma). \end{aligned} \tag{VI.10}$$

We need to establish that the term

$$\left[\frac{\phi\delta(\sigma) + C + 1}{\delta(\sigma) + x^2/\sigma^2}\right] \leq \rho$$

for some $\rho < 1$, and with $\sigma \notin [r, R]$ and $x \notin [-X, X]$, where $r, R, X > 0$.

For $\phi < \rho < 1$ we can choose $r, R > 0$ such that for $\sigma \notin [r, R]$ it holds that

$$\frac{\phi\delta(\sigma) + C + 1}{\delta(\sigma) + x^2/\sigma^2} \leq \frac{\phi\delta(\sigma) + C + 1}{\delta(\sigma)} = \phi + \frac{C + 1}{\delta(\sigma)} \leq \rho.$$

Similarly, since δ is bounded on compact subsets of $\mathbb{R}_+$ we may choose $X > 0$ such that for $\sigma \in [r, R]$ and $x_t^2 > X^2$ we have,

$$\frac{\phi\delta(\sigma) + C + 1}{\delta(\sigma) + x^2/\sigma^2} \leq \rho.$$

Finally, for $(\sigma, x) \in [r, R] \times [-X, X]$,

$$E[d(x_t, \sigma_t)|(\sigma_{t-1}, y_{t-1})] \leq K.$$

□

Examples of drift functions δ that satisfy the assumptions of the theorem are for example,

$$\delta(\sigma) = 1 + \sigma^2 + \sigma^{-2} \quad \text{and} \quad \delta(\sigma) = 1 + \left(\log \sigma^2\right)^2. \tag{VI.11}$$

VI.3.2.1 The log-normal SV

For the log-normal case, the volatility (σ_t) is a Markov chain with transition density which is continuous and hence bounded on closed intervals in $\mathbb{R}_+$ since $\sigma_t|\sigma_{t-1}$ is log-normal distributed.

With drift function $\delta(\sigma) = 1 + (\log \sigma_t)^2 = 1 + h_t^2$, where $h_t = \log(\sigma_t)$, follows a first order autoregressive AR(1)) process, $h_t = \gamma + \phi h_{t-1} + \eta_t$ with $\gamma := (1-\phi)\alpha$, $\sigma_\eta^2 := \beta^2(1-\phi^2)$, and $\phi \in (-1,1)$, we get directly that

$$\begin{aligned} E(\delta(\sigma_t)|\sigma_{t-1} = \sigma) &= 1 + \gamma^2 + \phi^2 h^2 + \sigma_\eta^2 + 2\gamma\phi h \\ &= \left(\frac{1 + \gamma^2 + \phi h^2 + \sigma_\eta^2 + 2\gamma\phi h}{1+h^2}\right)\delta(\sigma) \end{aligned}$$

with $h = \log(\sigma^2)$. Hence, since $|\phi| < 1$, then for $\sigma \notin [r, R]$, $r, R > 0$ and hence $h \notin [\log r^2, \log R^2]$,

$$\left(\frac{1 + \gamma^2 + \phi^2 h^2 + \sigma_\eta^2 + 2\gamma\phi h}{1+h^2}\right) \leq \rho,$$

with $|\phi| < \rho < 1$. From Theorem VI.1 we deduce that (x_t, σ_t) is geometrically ergodic in this case and hence satisfies the LLN and CLT for weakly mixing processes.

As a consequence the marginal process, (x_t), admits a stationary distribution. To find the invariant distribution remember that (x_t) is stationary if (σ_t), or equivalently (h_t), is stationary. From well known facts about the AR(1) process we know that in the Gaussian case, the invariant distribution for h_t is a Gaussian distribution with

$$E(h_t) = \alpha, \quad V(h_t) = \beta^2. \tag{VI.12}$$

Moreover, (h_t) is a Gaussian process with covariances given by

$$Cov(h_t, h_{t+k}) = \phi^k \beta, \quad k \geq 0.$$

Thus the stationary or invariant distribution of (x_t, σ_t) has density (see the discussion on the log-normal distribution),

$$\begin{aligned} f(\sigma, y) &= f(y|\sigma) f(\sigma) \\ &= (\frac{1}{\sigma} f_z(y/\sigma)) \left(\frac{1}{\sigma}\frac{1}{\sqrt{2\pi\beta}} \exp\left(-\frac{1}{2\beta}(\log(\sigma) - \alpha)^2\right)\right). \end{aligned}$$

and the stationary version of (x_t) satisfies the LLN.

In particular, the marginal process (x_t) is stationary with an invariant distribution given by the density

$$\phi(x) = \int_{\mathbb{R}} f(\sigma, x) d\sigma.$$

For future reference we note that the $k-$th order moment of x_t exists if $E|z_t|^k < \infty$ in which case it holds that (using that σ_t is log-normal)

$$Ex_t^k = \exp\left(\beta\frac{k^2}{2} + k\alpha\right) Ez_t^k.$$

References

Carrasco, M. and X. Chen, 2002, "β-mixing and Moment properties of Various GARCH, Stochastic Volatility and ACD Models", *Econometric Theory*,

Taylor, S.J., 2005, *Asset Price Dynamics, Volatility and Prediction*, Princeton University Press

Taylor, S.J., 1986, *Modelling Financial Time Series*, John Wiley, Chichester.

Genon-Catalot, V., T. Jeantheau, and C. Laredo, 2000, "Stochastic Volatility Models as Hidden Markov Models and Statistical Applications", *Bernoulli* 6, 1051-1079.

Meitz, M. and P. Saikkonen, 2008, "Ergodicity, Mixing and Existence of Moments for a class of Markov Models with Applications to GARCH and ACD Models", *Econometric Theory.*