Econometric Theory, **22**, 2006, 1052–1090. Printed in the United States of America.
DOI: 10.1017/S0266466606060518

STOCHASTIC UNIT ROOT MODELS

CHRISTIAN GOURIEROUX
University of Toronto
CREST
and
CEPREMAP

CHRISTIAN Y. ROBERT
CNAM
and
CREST

This paper develops a dynamic switching model, with a random walk and a stationary regime, where the time spent in the random walk regime is endogeneously predetermined. More precisely, we assume that the process is recursively defined by $Y_t = \mu + Y_{t-1} + \varepsilon_t$, with stochastic probability $\pi_{rw}(Y_{t-1})$, $Y_t = \mu + \varepsilon_t$, with stochastic probability $1 - \pi_{rw}(Y_{t-1})$, where (ε_t) is a strong white noise and π_{rw} is a nondecreasing function. Then, the dynamics of the process (Y_t), its marginal distribution, and the distribution of the time spent in the unit root regime depend on the pattern of random walk intensity π_{rw} and on the noise distribution F. Moreover, we study the links between the endogeneous switching regime and the degree of persistence of the process (Y_t).

1. INTRODUCTION

The sample autocorrelogram functions (ACF) reported for macroeconomic series as purchasing power parity (PPP) deviations and for financial series concerning either financial returns, exchange rates, interest rates, or volatilities generally feature slow decay rate. This stylized fact gives rise to a large body of applied literature, where the dynamics of the process of interest is described by means of the so-called fractional or long-memory model. The standard form of this model introduced by Granger and Joyeux (1980) is

$$(1 - L)^d Y_t = \mu + \varepsilon_t, \qquad 0 < d \leq 1, \tag{1}$$

where L is the lag operator, d is the differencing order and the errors (ε_t) are independent and identically distributed (i.i.d.) with a nondegenerate distribution. This model provides a large range of (linear) serial dependence. For

Address correspondence to Christian Y. Robert, Laboratoire Finance-Assurance, Timbre J320—Bureau 11.12, 15 Bd Gabriel Péri, 92245 Malakoff Cedex, France; e-mail: chrobert@ensae.fr.

 © 2006 Cambridge University Press 0266-4666/06 $12.00

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

$d \in (0, 0.5)$ the process (Y_t) is both stationary and invertible but is nonstationary if $d > 0.5$. The unit root process is obtained for $d = 1$. When d is positive, the process exhibits long-memory features, because the spectral density at frequency 0 is infinite and the linear autocorrelations of (Y_t) are not absolutely summable.

However, the assumption of i.i.d. error terms is restrictive and has important consequences on the properties of process (Y_t). (i) For instance, the domain of the process is necessarily lower and upper unbounded. This latter property is not compatible with the interpretations or observed evolutions of a large number of macroeconomic or financial variables. Some variables such as risk-free interest rates or volatilities are nonnegative, the exchange rates between European countries entering in the Euro zone can be constrained to stay in a band, and the PPP deviations generally feature mean-reverting behavior. (ii) Moreover, the optimal prediction of Y_t in model (1) involves a large (in fact infinite) number of lagged values of the process, whereas financial practitioners use (nonlinear) prediction formulas depending on a rather small number of lagged values.

Thus, there is a need for models that are compatible with the observed slow decay rate of linear ACF, boundary constraints, and prediction functions with a small number of lagged values. For this purpose, we can introduce nonlinear dynamics that allow for unit root in a given band and ensure nonexplosive behavior by introducing either a reflecting condition or a mean-reverting condition outside the band (see Brooks and Reveiz, 2002; Farnsworth and Bass, 2003). Reflecting Brownian motion is an example of a model including both a unit root characteristic and a boundary effect. However a model based on reflecting Brownian motion has the drawback of assuming a strict barrier. A more flexible specification is the BAND-TAR model introduced by Balke and Fomby (1997) and used by Bec, Ben Salem, and Carrasco (2004) for the analysis of PPP deviations. For an autoregressive dynamic model of order one, the recursive equation is

$$Y_t = \begin{cases} \mu_2 + Y_{t-1} + \varepsilon_t, & \text{if } |Y_{t-1}| \leq \lambda, \\ \mu_1 + \rho Y_{t-1} + \varepsilon_t, & \text{if } |Y_{t-1}| > \lambda, \end{cases} \tag{2}$$

where the errors (ε_t) are i.i.d. such that $\mathbb{E}\varepsilon_t = 0$ and the autoregressive coefficient ρ is smaller than 1 in absolute value. This model has a structural justification in a two-country one-sector model with transportation costs (see Sercu, Uppal, and Van Hulle, 1995): "The costs create a regime of no trade where the PPP relationship does not hold. Outside this area, international goods market arbitrage provides an error correction mechanism which brings national price levels back to equality, apart from a constant term reflecting the cost." The regime dynamics is determined endogeneously according to the lagged value Y_{t-1}.

The autoregressive model (2) with deterministic thresholds can be extended by considering a model with stochastic random walk regime intensity:

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

$$Y_t = \begin{cases} \mu_2 + Y_{t-1} + \varepsilon_t, & \text{with stochastic probability } \pi_{rw}(Y_{t-1}), \\ \mu_1 + \rho Y_{t-1} + \varepsilon_t, & \text{with stochastic probability } 1 - \pi_{rw}(Y_{t-1}), \end{cases} \tag{3}$$

where $\pi_{rw}(Y_{t-1})$ gives the stochastic probability of being in the random walk regime.

(i) Model (2) corresponds to the choice of the indicator function $\pi_{rw}(y) = 1$, if $|y| < \lambda$, $\pi_{rw}(y) = 0$, otherwise.

(ii) Let us now assume for illustration that the regime switching is a consequence of intervention of central banks to change the autoregressive coefficient from 1 (explosive case) to ρ (stationary case). In model (2), deterministic interventions are implicitly assumed whenever $|Y_{t-1}|$ is larger than the given threshold λ. This model can be extended by considering that the central banks intervene when $|Y_{t-1}|$ is too large but not in a systematic way. Their behavior can be represented by a $0 - 1$ variable I_t, where $I_t = 1$ if the intervention occurs and $= 0$ otherwise. A standard specification introduces a stochastic threshold (for a similar idea, see Enders and Granger, 1998) and assumes that $I_t = 1$ if $|Y_{t-1} + \eta_t| > \lambda$ and $= 0$ otherwise, where (η_t) and (ε_t) are independent and (η_t) is a strong white noise with a symmetric continuous distribution. Then, we get model (3) with intensity π_{rw} completely defined by the value of the threshold and by the distribution of η_t:

$$\pi_{rw}(Y_{t-1}) = P(I_t = 0|Y_{t-1}) = P(|Y_{t-1} + \eta_t| \leq \lambda |Y_{t-1}).$$

A special case has been studied recently by Rahbek and Shephard (2002) with $\mu_1 = \mu_2 = 0$,

$$\pi_{rw}(Y_{t-1}) = \frac{1}{1 + \exp(\alpha + \beta Y_{t-1}^2)}, \qquad \beta > 0,$$

and i.i.d. Gaussian errors.[1]

(iii) In fact, the stationarity of PPP deviation can be controlled even if the probability of intervention is not equal to 1 for large PPP deviation, that is, if intensity π_{rw} is such that $\lim_{y\to\infty} \pi_{rw}(y) > 0$. The existence of a small intervention probability for large PPP deviation does not necessarily imply the nonstationarity of the process. However, the distribution of the time spent in the unit root regime, the decay rate of the autocorrelation, and more generally the persistence properties depend on the pattern of intensity π_{rw} and noise distribution F.

All the preceding models assume a symmetric probability of intervention (see Figures 1a corresponding to model (2), 1b corresponding to case (ii), and 1c corresponding to case (iii)). This assumption is approximately satisfied if there is a consensus between the two central banks to avoid PPP deviation, as was the case with European countries before the introduction of the Euro. However, an asymmetric probability of intervention can be more realistic if one

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

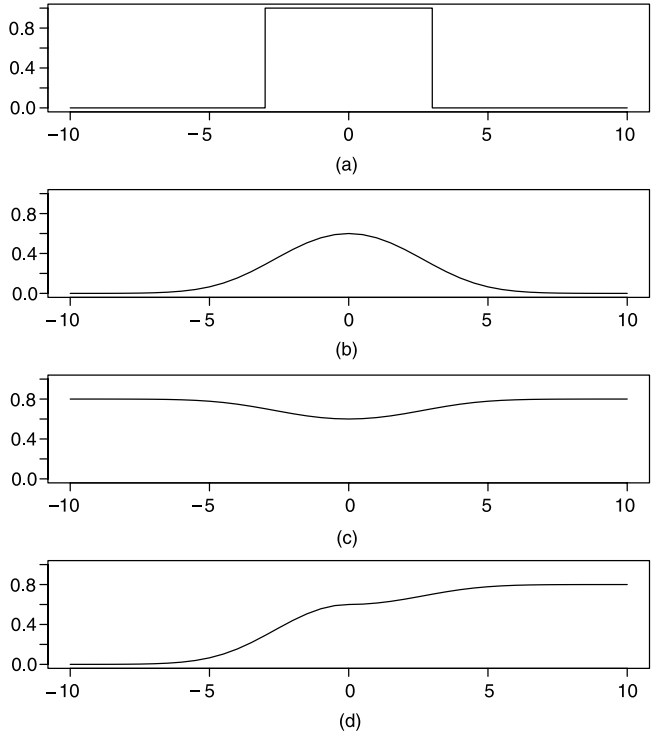


FIGURE 1. Random walk regime intensities.

of the countries is dominant. This is typically the case with the United States when the exchange rate of the Euro versus the dollar increases to facilitate U.S. exports. In this paper, we consider a model that mixes types (ii) and (iii) by considering a probability of intervention that is increasing and such that $\lim_{y\to\infty} \pi_{rw}(y) > 0$ (see Figure 1d).

The model is described in Section 2, with special attention to constant and threshold intensity. In Section 3, we consider an exponential-type distribution for the noise. Indeed, in this framework it is possible to perform analytical computations and derive the stationary distribution of the process (Y_t). Then, we discuss how the tails of this stationary distribution depend on intensity π_{rw}. We provide in Section 4 ergodicity conditions for the process (Y_t). We distinguish the case $\bar{\pi}_{rw} = \lim_{y\to\infty} \pi_{rw}(y) < 1$, where the process is always stationary, and the case $\bar{\pi}_{rw} = 1$, in which joint constraints between π_{rw} and the noise density f have to be introduced. The ergodicity conditions are especially important for statistical inference. They are also important in discussing the asymptotic behavior of the unit root tests or estimators of the fractional order d, usually built under specification (1), when this specification (1) is incorrect and the process

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

satisfies a stochastic unit root model. Section 5 concludes. The proofs are gathered in Appendixes.

2. THE MODEL

2.1. Definition of the Process

The process (Y_t) satisfies the stochastic unit root regime

$$Y_t = \begin{cases} \mu + Y_{t-1} + \varepsilon_t, & \text{with stochastic probability } \pi_{rw}(Y_{t-1}), \\ \mu + \varepsilon_t, & \text{with stochastic probability } 1 - \pi_{rw}(Y_{t-1}), \end{cases} \tag{4}$$

where $t \geq 1$, (ε_t) are i.i.d. zero mean random variables independent of Y_0, and the random walk regime intensity π_{rw} is a càdlàg increasing function with $\lim_{y \to -\infty} \pi_{rw}(y) = \underline{\pi}_{rw}$ and $\lim_{y \to \infty} \pi_{rw}(y) = \bar{\pi}_{rw}$. We assume that ε_t has a density f strictly positive on the whole real line, that the variance of the noise distribution σ^2 is finite, and that $\bar{\pi}_{rw} > 0$ and $\underline{\pi}_{rw} < 1$ to avoid the standard pure white noise or random walk cases. We denote by F the cumulative distribution function (c.d.f.) of ε_t. For expository purposes, we fix $\mu_1 = \mu_2 = \mu$, $\rho = 0$, but the main results are valid for any μ_1, μ_2, and ρ, $|\rho| < 1$ (and will be systematically provided). This process will be called a dynamic switching (D.S.) Markov chain.

At any date t, there are two possible regimes corresponding to either a white noise, when $Y_t = \mu + \varepsilon_t$, or a random walk, when $Y_t = \mu + Y_{t-1} + \varepsilon_t$. This model can be considered as the reduced form of a dynamic model with a latent regime variable. The latent model is

$$Y_t = B_t Y_{t-1} + \mu + \varepsilon_t, \tag{5}$$

where

$$B_t = \begin{cases} 1, & \text{if } Z_t \leq Y_{t-1}, \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

and the variables (Z_t) are i.i.d., independent of the noise (ε_t) with c.d.f. π_{rw}. Thus, the latent variable Z_t can take infinite values with a positive probability. In the example of PPP deviation, the indicator function $B_t = 1 - I_t$, where I_t denotes the indicator of intervention.

We get a linear autoregressive model with stochastic autoregressive coefficient B_t. The regime indicator is endogeneous, because its conditional distribution depends on the lagged value Y_{t-1}. Therefore, the model differs from a model in which (B_t) is an underlying hidden Markov chain independent of the noise (ε_t) as in Hamilton (1989) or Granger and Swanson (1997).

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

Here (Y_t) is a homogeneous Markov process, with a transition density corresponding to a mixture of distributions:

$$l(y_t|y_{t-1}) = \pi_{rw}(y_{t-1})f_\mu(y_t - y_{t-1}) + [1 - \pi_{rw}(y_{t-1})]f_\mu(y_t), \quad \textbf{(7)}$$

where $f_\mu(y) = f(y - \mu)$. The probability density function (p.d.f.) of the stationary distribution is denoted by l if it exists and is unique. The stationary density satisfies the integral equation

$$l(y) = \int [\pi_{rw}(x)f_\mu(y - x) + (1 - \pi_{rw}(x))f_\mu(y)]l(x)\,dx. \quad \textbf{(8)}$$

2.2. The Switching Regime Process

The process $(B_t, t \geq 1)$ provides the sequence of random walk regime indicators. The sequence of durations spent in the two admissible states is denoted by $D_1^0, D_1^1, D_2^0, D_2^1, \ldots$; D_j^0 is the duration of the jth spell in the white noise regime (called regime 0), whereas D_j^1 is the duration of the jth spell in the random walk regime (called regime 1). This sequence may start by either a random walk or a white noise regime according to the initial value Y_0 and the initial drawing B_1 that follows conditional on Y_0, a Bernoulli distribution with parameter $\pi_{rw}(Y_0)$. As an illustration, Figure 2 provides a path of the D.S. Mar-

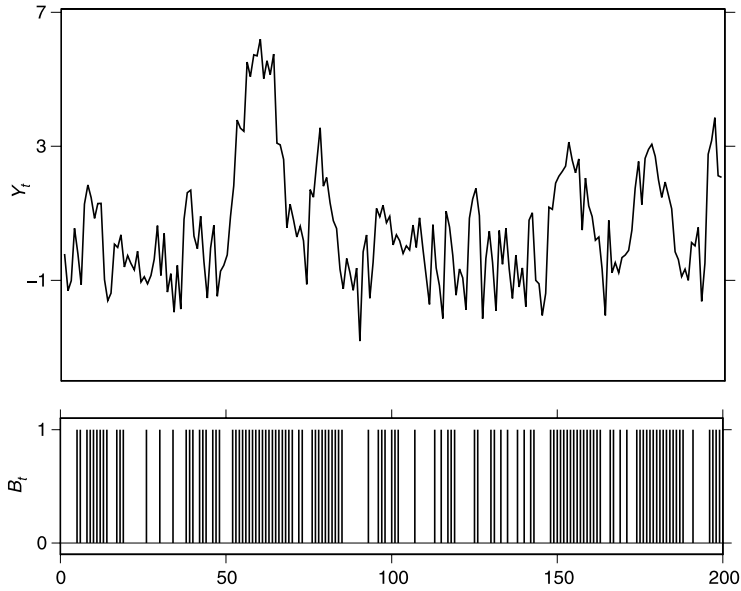


FIGURE 2. Paths of Y and B processes.

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

kov chain jointly with the path of the associated regime indicator process B. The selected stochastic intensity is $\pi_{rw}(y) = \Phi(y)$, where Φ is the standard normal cumulative distribution function, $\mu = -0.2$, whereas the noise is standard normal and the initial value is $Y_0 = 0$. There is no excursion of process (Y_t) in the set of large negative values. Indeed, when the process is negative, there is a high probability of staying in the white noise regime. When process (Y_t) takes large positive values, the path is similar to the path of a random walk with a negative drift. In particular, this path does not come back to zero abruptly because of the high probability of staying in the random walk regime.

Figure 3 provides two paths of the process with different initial values $Y_0 = 12$ (solid line) and $Y_0 = 7$ (dotted line) and the same noise sequences (ε_t) and (Z_t) (see eqn. (6)). Both paths are similar for t larger than the first time when the regime of the process with the largest value changes from the random walk regime to the white noise regime. In the example of Figure 3 where the process with the largest initial value has always been positive before changing of regime, any path with a smaller initial value joins this path at the same date.

Let us now describe some distributional properties of the durations spent in the white noise and random walk regimes, respectively. Let us denote $p_k = \mathbb{E}[\pi_{rw}(S_1)\pi_{rw}(S_2)\dots\pi_{rw}(S_k)]$, $k \geq 1$, where $S_k = \varepsilon_1 + \cdots + \varepsilon_k + k\mu$ is the random walk process with drift. The probabilities p_k, $k \geq 1$, are strictly positive because $\bar{\pi}_{rw} > 0$ and $f > 0$. They define a monotonic nonincreasing sequence.

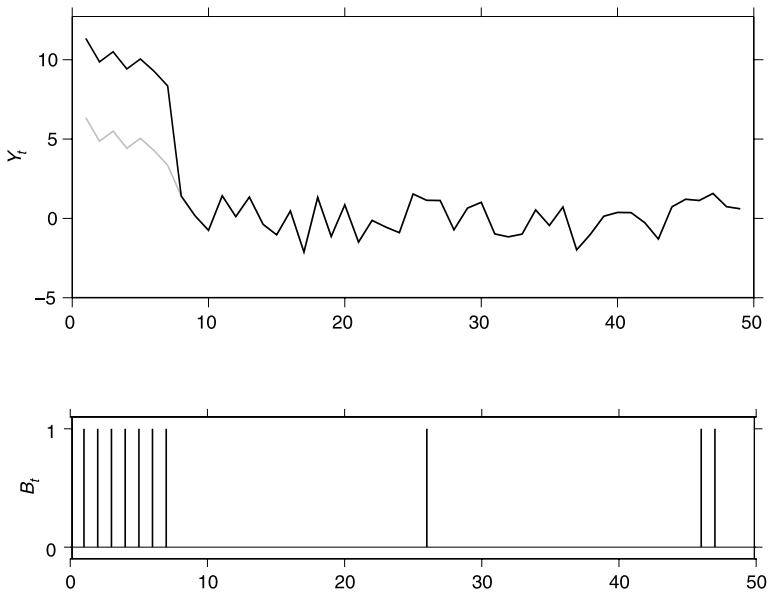


FIGURE 3. Paths of Y for different initial values.

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

PROPOSITION 1.

(i) The durations spent in the random walk regime are a.s. finite if and only if $\lim_{k\to\infty} p_k = 0$.

(ii) If $\lim_{k\to\infty} p_k = 0$, *the durations* $D_1^0, D_1^1, D_2^0, D_2^1, \ldots$ *are independent. Moreover, the durations spent in the white noise regime* D_j^0, $j \geq 2$ *(resp.* D_j^1, $j \geq 2$*) have identical distributions. The distribution is a geometric distribution for* D_2^0*:*

$$P[D_2^0 = d] = p_1(1 - p_1)^{d-1}, \qquad d \geq 1.$$

The distribution of D_2^1 *is given by*

$$P[D_2^1 = d] = \frac{p_d - p_{d+1}}{p_1}, \qquad d \geq 1.$$

Proof. See Appendix A.

The path dependence of the random walk intensity does not imply the serial dependence of durations but only has an effect on their stationary distribution. For instance, if π_{rw} and π_{rw}^* are such that $\pi_{rw}^*(y) \geq \pi_{rw}(y)$, $\forall y$, then the distribution of duration D_2^1 computed with π_{rw}^* stochastically dominates at order one the distribution of D_2^1 computed with π_{rw}, that is, $P^*[D_j^1 \geq d] \geq P[D_j^1 \geq d]$, $\forall d \geq 1$. Note also that the distribution of the duration spent in the white noise regime admits exponential tails, whereas the tail pattern for the duration spent in the random walk regime depends on intensity π_{rw}.

The properties of durations D_j^0 and D_j^1 described in Proposition 1 can be used to simplify the analysis of process (Y_t). More precisely, the discrete time process $\tau(k) = \sum_{j=1}^k (D_j^0 + D_j^1)$ is a renewal process, that is, it admits i.i.d. increments. In this case, the cycles $\{Y_{t+\tau(k-1)}\}_{0\leq t<\tau(k)-\tau(k-1)}$, $k \geq 2$, are also i.i.d. This means that (Y_t) is a regenerative process, for which simplified properties have been given in the literature (Asmussen, 1987). Note also that the set $\mathbb{R} \times \{0\}$ is a proper atom of the chain (Y_t, B_t), that is,

$$P[(Y_t, B_t) \in \cdot | (Y_{t-1}, B_{t-1}) = (y_1, 0)] = P[(Y_t, B_t) \in \cdot | (Y_{t-1}, B_{t-1}) = (y_2, 0)],$$

for all $(y_1,0),(y_2,0) \in \mathbb{R} \times \{0\}$, and the times at which the chain enters the atom are just the regeneration times $\tau(k)$.

In the proposition that follows, we give sufficient conditions for a.s. finite durations spent in the random walk regime.

PROPOSITION 2. *Sufficient conditions for a.s. finite durations spent in the random walk regime are*

(i) $\bar{\pi}_{rw} = \lim_{y\to\infty} \pi_{rw}(y) < 1$;

(ii) $\bar{\pi}_{rw} = 1$ *and* $\mu \leq 0$;

(iii) $\bar{\pi}_{rw} = 1$, $\mu > 0$, *and* $1 - \pi_{rw}(y)$ *is not integrable at* $+\infty$.

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

Proof. Let us note that $A_k = \pi_{rw}(S_1)\pi_{rw}(S_2)\dots\pi_{rw}(S_k)$ is a nonincreasing sequence of nonnegative random variables. Thus, the a.s. convergence of A_k to zero implies the condition $\lim_{k\to\infty} p_k = 0$, by Beppo–Levi's theorem (see Billingsley, 1995, p. 209). Let us now check for the convergence of A_k to zero.

(i) If $\bar{\pi}_{rw} < 1$, we get $\lim_{k\to\infty} A_k \leq \lim_{k\to\infty} (\bar{\pi}_{rw})^k = 0$.

(ii) Let us first assume that $\mu < 0$. The random walk with drift $S_n = \varepsilon_1 + \dots + \varepsilon_n + n\mu$ converges almost surely to $-\infty$, when n increases and the result follows.

Let us now assume that $\mu = 0$. Because $\tau_-(1) = \inf\{n \geq 1 : S_n \leq 0\}$ is a.s. finite (see Asmussen, 1987, Thm. 2.4, p. 169), the descending ladder epochs $\tau_-(n+1) = \inf\{k > \tau_-(n) : S_k \leq S_{\tau_-(n)}\}$ exist and $(\tau_-(n+1) - \tau_-(n))$, $n \geq 1$, are a.s. finite random variables. Then

$$\lim_{k\to\infty} A_k = \lim_{k\to\infty} \prod_{i=1}^{k} \pi_{rw}(S_i) \leq \lim_{n\to\infty} (\pi_{rw}(0))^n = 0, \quad \text{a.s.},$$

because $\pi_{rw}(S_{\tau_-(n)}) \leq \pi_{rw}(0)$.

(iii) Finally, the last sufficient condition results from the law of iterated logarithm, which ensures that for any $\lambda > 1$ (see Billingsley, 1995, p. 155)

$$S_n > n\mu + \lambda(2\sigma^2 n \ln\ln n)^{1/2},$$

a finite number of times a.s. We deduce that a sufficient condition for the a.s. convergence of A_k to zero is

$$\sum_{n\geq 3} \log \pi_{rw}(n\mu + \lambda(2\sigma^2 n \ln\ln n)^{1/2}) = -\infty$$

or equivalently

$$\sum_{n\geq 3} [1 - \pi_{rw}(n\mu + \lambda(2\sigma^2 n \ln\ln n)^{1/2})] = \infty.$$

The result follows. ■

In the proposition that follows, we exhibit lower and upper bounds for probabilities p_n. These bounds can be used to transpose conditions written in terms of p_n into conditions written in terms of π_{rw} and f. When $\underline{\pi}_{rw} > 0$ and $\bar{\pi}_{rw} < 1$, we have $(\underline{\pi}_{rw})^n \leq p_n \leq (\bar{\pi}_{rw})^n$ and a geometric decay for p_n. Some bounds of the probabilities p_n can also be derived in the limiting case $\bar{\pi}_{rw} = 1$.

PROPOSITION 3. *Let us assume* $\bar{\pi}_{rw} = 1$ *and the noise distribution with a Laplace transform existing in a neighborhood of 0.*

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

(i) If $\underline{\pi}_{rw} > 0$*, then, for any* $\lambda > 1$*, there exists a constant* C_λ *such that*

$$p_n > C_\lambda \exp\left(\sum_{i=3}^{n} \log \pi_{rw}(i\mu - \lambda(2\sigma^2 i \log\log i)^{1/2}) \right).$$

(ii) If $\mu > 0$*, then, for any* $\lambda > 1$*, there exists a constant* D_λ *such that*

$$p_n < D_\lambda \exp\left(\sum_{i=1}^{n} \log \pi_{rw}(i\lambda\mu) \right).$$

(iii) If $\mu = 0$*, then, for any* $\lambda > 0$*, there exists a constant* D^*_λ *such that*

$$p_n < D^*_\lambda \exp\left(\sum_{i=1}^{n} \log \pi_{rw}(i\lambda) \right).$$

(iv) If $\mu < 0$*, there exist two scalars* $\varkappa$ *and* $\theta < 1$ *such that*

$$p_n < \varkappa\theta^n.$$

Proof. See Appendix B.

For example, let us assume that $\mu > 0$ and

$$\pi_{rw}(y) = \exp\left(-\frac{c^\tau}{(c+y)^\tau} \right) \mathbb{I}_{[0,\infty[}(y) + \exp(-1)\mathbb{I}_{]-\infty,0[}(y),$$

with $\frac{1}{2} < \tau \leq 1$, $c > 0$. The intensity is such that $\underline{\pi}_{rw} = \exp(-1) > 0$ and $\bar{\pi}_{rw} = 1$. We can apply Proposition 3 (i) and (ii) with $\lambda > 1$.

1. If $\tau = 1$, there exist two constants C_λ and D_λ such that

   $$\frac{C_\lambda}{n^{c/\mu}} < p_n < \frac{D_\lambda}{n^{c/\lambda\mu}}$$

   and (p_n) has a hyperbolic rate of decay. Moreover, the mean of D_2^1,

   $$\mathbb{E}D_2^1 = \sum_{n=1}^{\infty} \frac{n(p_n - p_{n+1})}{p_1} = \sum_{n=1}^{\infty} \frac{p_n}{p_1},$$

   is finite if and only if $c > \mu$.
2. If $\frac{1}{2} < \tau < 1$, there exist two constants C_λ and D_λ such that

   $$C_\lambda \exp\left\{ -\frac{c}{\mu} n^{1-\tau} \right\} < p_n < D_\lambda \exp\left\{ -\frac{c}{\lambda\mu} n^{1-\tau} \right\}.$$

Because the survivor function of the durations spent in the random walk regime is proportional to probability p_n (see Proposition 1), the bounds can be used to analyze the type of the tails of the durations. When $\bar{\pi}_{rw} = 1$ and $\mu > 0$, they

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

are Weibull or Pareto tails in the example (recall that we have thin tails if $\bar{\pi}_{rw} < 1$, or $\bar{\pi}_{rw} = 1$ and $\mu < 0$).

2.3. Special Cases

(i) Constant intensity. When $\pi_{rw}(y) = \pi_c$ is constant, we get exogenous regimes and a special case of hidden Markov process (see MacDonald and Zucchini, 1997). The regime indicators B_t are i.i.d., independent of the noise process. The durations D_j^0, $j \geq 2$, (resp. D_j^1, $j \geq 2$) have both geometric distributions $P(D_2^0 = n) = \pi_c(1 - \pi_c)^{n-1}$, (resp. $P(D_2^1 = n) = (1 - \pi_c)\pi_c^{n-1}$). Therefore, the assumption of exogenous regimes implies thin tails for both duration distributions (if $0 < \pi_c < 1$).

(ii) Threshold intensity. The process reduces to a threshold autoregressive process (see Tong, 1990) when π_{rw} is the indicator function $\pi_{rw}(y) = \mathbb{I}_{[0,+\infty[}(y)$. Indeed, we have $Y_t = \mathbb{I}_{Y_{t-1} \geq 0} Y_{t-1} + \mu + \varepsilon_t$. This process is related to the Lindley process (U_t) defined by

$$U_t = (U_{t-1} + \mu + \varepsilon_t)^+, \qquad t \geq 1, \tag{9}$$

and studied in Feller (1966).

PROPOSITION 4. *When $Y_0 = U_0 = 0$, the process*

$$Y_t = \mathbb{I}_{Y_{t-1} \geq 0} Y_{t-1} + \mu + \varepsilon_t, \qquad t \geq 1,$$

is such that $U_t = Y_t^+$ is a Lindley process.

Proof. The proof is by recursion. The result is obvious for $t = 0$. If it is valid for $t - 1$, we get

$$Y_t = \mathbb{I}_{Y_{t-1} \geq 0} Y_{t-1} + \mu + \varepsilon_t = \mathbb{I}_{U_{t-1} > 0} U_{t-1} + \mu + \varepsilon_t = U_{t-1} + \mu + \varepsilon_t.$$

We deduce $Y_t^+ = (U_{t-1} + \mu + \varepsilon_t)^+ = U_t$. ■

The durations D_j^0, $j \geq 2$, have identical distributions: $P(D_2^0 = n) = (1 - p)p^{n-1}$, where $p = P(S_1 \leq 0)$. The durations D_j^1, $j \geq 2$, are such that $P(D_2^1 = n) = P(\tau = n + 1 | \tau \geq 2)$, where

$$\tau = \inf\{n \geq 1, S_n = \varepsilon_1 + \cdots + \varepsilon_n + n\mu \leq 0\}.$$

Here τ is the time to the first visit of a random walk with drift in the set of negative values. When $\mu < 0$, the first moment of the duration D_2^1 can be derived by applying random walk theory (see, e.g., Asmussen, 1987, p. 176). We have

$$\mathbb{E}\tau = \exp\left\{\sum_{n=1}^{\infty} \frac{1}{n} P(S_n > 0)\right\}.$$

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

Because $P(\tau = 1) = p$, we deduce

$$\mathbb{E}D_2^1 = \frac{\mathbb{E}\tau - 1}{1-p} = \frac{\exp\left\{\sum_{n=1}^{\infty} \frac{1}{n} P(S_n > 0)\right\} - 1}{1-p}.$$

3. EXPONENTIAL-TYPE DISTRIBUTION

Analytical computations can be performed when the noise admits an exponential-type distribution. This is the aim of this section. Let us assume that

$$f_\mu(y) = \begin{cases} \alpha \exp(-\alpha y)/2, & \text{if } y \geq 0, \\ \beta \exp(\beta y)/2, & \text{if } y < 0, \end{cases}$$

where α, β are positive. The mean of this distribution is $\mu = (\alpha^{-1} - \beta^{-1})/2$.

PROPOSITION 5. *If the intensity π_{rw} is differentiable on $\mathbb{R}\backslash\{0\}$, if the stationary distribution of the D.S. Markov chain exists, and if its density l is twice differentiable on $\mathbb{R}\backslash\{0\}$, then it satisfies the second-order linear differential equation*

$$l''(y) - \alpha\beta\mu[2 - \pi_{rw}(y)]l'(y) - \alpha\beta[(1 - \pi_{rw}(y)) - \mu\pi'_{rw}(y)]l(y) = 0,$$

if $y \neq 0$.

Proof. See Appendix C.

Closed form stationary densities can be derived in special cases.
(i) Constant intensity. When $\pi_{rw}(y) = \pi_c$ is constant, we have

$$l''(y) - \alpha\beta\mu(2 - \pi_c)l'(y) - \alpha\beta(1 - \pi_c)l(y) = 0, \quad \text{if } y \neq 0.$$

We get

$$l(y) = A\exp(-\gamma y), \quad \text{if } y > 0, \quad \text{and} \quad l(y) = B\exp(-\delta y), \quad \text{if } y < 0,$$

where

$$\gamma = \frac{-\alpha\beta\mu(2 - \pi_c) + \sqrt{(\alpha\beta\mu(2 - \pi_c))^2 + 4\alpha\beta(1 - \pi_c)}}{2} > 0,$$

$$\delta = \frac{-\alpha\beta\mu(2 - \pi_c) - \sqrt{(\alpha\beta\mu(2 - \pi_c))^2 + 4\alpha\beta(1 - \pi_c)}}{2} < 0.$$

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

The terms γ and δ are the solutions of the equation $\Psi_\varepsilon(u) = (\pi_c)^{-1}$, where

$$\Psi_\varepsilon(u) = \mathbb{E}\exp(-u\varepsilon_1) = \frac{1}{2}\left(\frac{\alpha}{\alpha - u} + \frac{\beta}{\beta + u}\right)$$

is the Laplace transform of the noise distribution.

Thus, for constant intensity $\pi_{rw}(y) = \pi_c$ and exponential-type noise distribution, the stationary distribution of the D.S. Markov chain also has exponential tails. However the left and right tail parameters are related in a complicated way to the tail parameters of the noise α, β and to the random walk regime intensity π_c. For symmetric tails of the noise distribution $\alpha = \beta$, we get symmetric tails for the stationary distribution with $\gamma = -\delta = \alpha\sqrt{1 - \pi_c} < \alpha$. The stochastic random walk regime implies a tail increase when we pass from the noise distribution to the stationary distribution. This result is compatible with the tail increase already observed for conditionally heteroskedastic models (see Engle, 1982) and time deformed processes (see Mandelbrot and Taylor, 1967; Clark, 1973).

(ii) Threshold intensity. When $\pi_{rw}(y) = \mathbb{I}_{[0,+\infty[}(y)$, we also get differential equations with constant coefficients on the positive (resp. negative) real line:

$$l''(y) - \alpha\beta\mu l'(y) = 0, \quad \text{if } y > 0,$$

$$l''(y) - 2\alpha\beta\mu l'(y) - \alpha\beta l(y) = 0, \quad \text{if } y < 0.$$

This differential system admits a density function as solution if and only if $\mu < 0$, which is the stationarity condition of the Lindley process. The solution is given by

$$l(y) = A\exp(-\rho y)\mathbb{I}_{]-\infty,0[}(y) + C\exp(-\gamma y)\mathbb{I}_{[0,\infty[}(y),$$

with

$$\gamma = -\alpha\beta\mu \quad \text{and} \quad \rho = -\alpha\beta\mu - \sqrt{(\alpha\beta\mu)^2 + \alpha\beta}.$$

The tails of the stationary distribution are still of exponential type.

(iii) Model without drift. If $\alpha = \beta = 1$, the drift is zero, and the differential equation becomes

$$l''(y) - (1 - \pi_{rw}(y))l(y) = 0, \quad \text{if } y \neq 0.$$

The right tail of the stationary distribution is necessarily such that l''/l is decreasing, with value in $[0,1]$. As an illustration, we consider subsequently various admissible patterns of the stationary distribution (restricted to the positive real line).

- Exponential tail. If $l''(y) = \gamma^2 A\exp(-\gamma y)[1 + o(1)]$, $0 < \gamma < 1$, for large y, we get $l(y) = A\exp(-\gamma y)[1 + o(1)]$ by using l'Hospital's rule. Then,

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

$l''(y)/l(y) = \gamma^2[1 + o(1)]$ and the intensity π_{rw} tends to $1 - \gamma^2 = \bar{\pi}_{rw} < 1$ at infinity. For instance, an admissible pair is

$$l(y) = b[1 + \exp(-(a + y))]\exp(-\gamma y), \qquad y > 0,$$

and

$$\pi_{rw}(y) = \bar{\pi}_{rw} - \frac{1}{1 + \exp(a + y)}(2\gamma + 1), \qquad y > 0,$$

where $\gamma = \sqrt{1 - \bar{\pi}_{rw}}$, $a > \ln(\gamma(\gamma + 2)/(1 - \gamma^2))$, and $b > 0$.

- Exponential tail with hyperbolic effect. But we can have $l''(y)/l(y) = \gamma^2[1 + o(1)]$ without strict exponential tails. For instance, an admissible pair is

$$l(y) = \frac{b}{(a + y)}\exp(-\gamma y), \qquad y > 0,$$

and

$$\pi_{rw}(y) = \bar{\pi}_{rw} - \frac{2\sqrt{(1 - \bar{\pi}_{rw})}}{(a + y)} - \frac{2}{(a + y)^2}, \qquad y > 0,$$

where $\gamma = \sqrt{1 - \bar{\pi}_{rw}}$, $a > (\sqrt{1 - \bar{\pi}_{rw}} + \sqrt{1 + \bar{\pi}_{rw}})/\bar{\pi}_{rw}$, and $b > 0$.

- Weibull tail. If $l''(y) = Ac^2\tau^2 y^{2(\tau-1)}\exp(-cy^\tau)[1 + o(1)]$, $0 < \tau < 1$, and $c > 0$, for large y, we have $l(y) = A\exp(-cy^\tau)[1 + o(1)]$ and $1 - \pi_{rw}(y) = c^2\tau^2 y^{2(\tau-1)}[1 + o(1)]$. For instance, an admissible pair is

$$l(y) = b\exp[-c(1 + y)^\tau], \qquad y > 0,$$

and

$$\pi_{rw}(y) = 1 - c^2\tau^2(1 + y)^{2(\tau-1)} - (1 - \tau)c\tau(1 + y)^{\tau-2}, \qquad y > 0,$$

with $0 < \tau < 1$, $0 < c < 1$, and $b > 0$.

- Pareto tail. If $l''(y) = Aa(a + 1)y^{-(a+2)}[1 + o(1)]$, $a > 1$, for large y, then $l(y) = Ay^{-a}[1 + o(1)]$ and $1 - \pi_{rw}(y) = a(a + 1)y^{-2}[1 + o(1)]$. For instance, an admissible pair is

$$\pi_{rw}(y) = 1 - \frac{a(a + 1)}{(d + y)^2}, \qquad y > 0, \qquad l(y) = \frac{b}{(d + y)^a}, \qquad y > 0,$$

where $a > 1$, $d^2 > \alpha(\alpha + 1)$, and $b > 0$.

- Gompertz tail. If $l(y) = A\exp(-\exp y)$ $y > 0$, we get $l''(y)/l(y) = \exp[y(\exp y - 1)][1 + o(1)]$, which is not between 0 and 1. Thus, the stationary distribution cannot admit Gompertz tail.

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

The preceding examples show that

1. stationary distributions with exponential tails cannot be obtained if $\bar{\pi}_{rw} = 1$ and $\mu = 0$ but there is a large variety of admissible tails;
2. stationary distributions with very thin tails such as Gompertz tails can never be obtained.

4. ERGODICITY AND PERSISTENCE OF THE PROCESS

In Section 2, we mentioned that the D.S. Markov chain is regenerative, that is, it can be split up into independent cycles, if the time spent in the random walk regime is a.s. finite. Roughly speaking, the process does not explode, and a stationary distribution may exist. In this section, we use Markov chain theory to formalize this discussion, and we study the stochastic properties of the process.

4.1. Ergodicity

General results on Markov processes can be applied to the D.S. Markov chain. Let us denote by ν the Lebesgue measure and by $\mathcal{B}$ the Borel σ-algebra on $\mathbb{R}$. For any measurable set A, we define the time to entry in A:

$$\tau_A = \inf\{t \geq 1 : Y_t \in A\} \tag{10}$$

and the number of visits to A:

$$\eta_A = \sum_{t=1}^{\infty} \mathbb{I}_{\{Y_t \in A\}}. \tag{11}$$

(i) The first property that can be expected from a Markov chain is that any subset of the space can be reached, for any starting value. A Markov chain is ν-irreducible if any set of positive Lebesgue measure can be reached in finite time starting from any initial value. This condition can be written as

$$\nu(A) > 0 \text{ implies } L(y, A) = P[\tau_A < \infty | Y_0 = y] > 0, \quad \forall y \in \mathbb{R}.$$

The D.S. Markov chain is ν-irreducible because the density of ε_1 is strictly positive and

$$L(y, A) \geq P[\tau_A = 1 | Y_0 = y] \geq [1 - \pi(y)]P(\varepsilon_1 + \mu \in A) > 0.$$

(ii) A Markov chain is unstable when it returns only a finite number of times to a given set of "reasonable size." Two definitions of stable chains are commonly used:

(a) the Markov chain (Y_t) is ν-recurrent if it is ν-irreducible and for any set of positive Lebesgue measure the expected number of returns to this set is infinite starting from any initial value. This condition can be written as

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

$$\nu(A) > 0 \text{ implies } U(y, A) = \mathbb{E}[\eta_A | Y_0 = y] = \infty, \quad \forall y \in \mathbb{R}.$$

(b) The Markov chain (Y_t) is Harris-recurrent if it is ν-irreducible and for any set of positive Lebesgue measure the chain returns to the set infinitely often starting from any element in the set. This condition can be written as

$$\nu(A) > 0 \text{ implies } P[\eta_A = \infty | Y_0 = y] = 1, \quad \forall y \in A.$$

A Harris-recurrent Markov chain is ν-recurrent, and, if the chain is ν-recurrent, then it admits a unique invariant measure. Let us now consider the D.S. Markov chain.

LEMMA 1. *If the duration spent in the random walk regime is a.s. finite, the D.S. Markov chain is Harris-recurrent.*

Proof. Indeed, under this condition, the durations spent in the two regimes are a.s. finite, and there is an infinite number of cycles $\{Y_{t+\tau(k-1)}\}_{0 \le t < \tau(k) - \tau(k-1)}$. Then, we have

$$\begin{aligned} &P[\eta_A = \infty | Y_0 = y] \\ &\quad = P\left[\sum_{t=1}^{\infty} \mathbb{I}_{\{Y_t \in A\}} = \infty | Y_0 = y\right] = P\left[\sum_{k=1}^{\infty} \sum_{t=\tau(k-1)+1}^{\tau(k)} \mathbb{I}_{\{Y_t \in A\}} = \infty | Y_0 = y\right] \\ &\quad \ge P\left[\sum_{k=2}^{\infty} \sum_{t=\tau(k-1)+1}^{\tau(k)} \mathbb{I}_{\{Y_t \in A\}} = \infty\right] \ge P\left[\sum_{k=2}^{\infty} \mathbb{I}_{\{(\varepsilon_{\tau(k-1)+1} + \mu) \in A\}} = \infty\right], \end{aligned}$$

where the first inequality uses the independence between cycles corresponding to $k \ge 2$ and the starting value. Because $\varepsilon_{\tau(k-1)+1}$, $k \ge 2$, are i.i.d. and

$$\sum_{k=2}^{\infty} P[(\varepsilon_{\tau(k-1)+1} + \mu) \in A] = \sum_{k=2}^{\infty} P[(\varepsilon_1 + \mu) \in A] = \infty,$$

we conclude by the Borel–Cantelli lemma that

$$P[\eta_A = \infty | Y_0 = y] \ge P\left[\sum_{k=2}^{\infty} \mathbb{I}_{\{(\varepsilon_{\tau(k-1)+1} + \mu) \in A\}} = \infty\right] = 1. \quad \blacksquare$$

Sufficient conditions for Harris-recurrence of the D.S. Markov chain are directly deduced from Proposition 2.

(iii) The Markov chain (Y_t) is positive Harris-recurrent if it is Harris-recurrent and the invariant measure is finite.

Positive Harris-recurrence can be analyzed for the D.S. Markov chain. If the expected duration spent in the random walk regime is finite, the regenerative process (Y_t) admits a limiting probability distribution (see Asmussen, 1987, Thm. 1.2, p. 126) and the Markov process is positive Harris-recurrent. Therefore, suf-

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

ficient conditions for the positive Harris-recurrence of the D.S. Markov chain are directly deduced from Proposition 3.

For instance, we get immediately the corollary that follows.

COROLLARY 1. *Let us assume that* $\bar{\pi}_{rw} = 1$.

(i) The D.S. Markov chain is Harris-recurrent if one of the following conditions holds:
(a) $\mu \leq 0$*;*
(b) $\mu > 0$ *and* $1 - \pi_{rw}(y)$ *is not integrable at* $+\infty$.
(ii) The D.S. Markov chain is positive Harris-recurrent if one of the following conditions holds:
(a) $\mu < 0$*;*
(b) $\mu \geq 0$ *and* $\lim_{y\to\infty}[1 - \pi_{rw}(y)]y = c \in (0;\infty]$ *and* $c > \mu$.

Proof.

(i) It is deduced from Proposition 2.
(ii) If $\mu < 0$, from Proposition 3(iv) we deduce that

$$\mathbb{E}D_2^1 = \sum_{n=1}^{\infty} \frac{p_n}{p_1} < \frac{\varkappa\theta}{(1-\theta)p_1} < \infty.$$

If $\mu > 0$ and $\lim_{y\to\infty}[1 - \pi_{rw}(y)]y = c$ is finite, then, for large n, $p_n < D_\lambda n^{-c/\lambda\mu}$ (see Proposition 3(ii)). Let us choose $1 < \lambda < c\mu^{-1}$. Then, we get $\mathbb{E}D_2^1 < \infty$. A similar discussion can be done if c is not finite or if $\mu = 0$. ■

In the following proposition, we give additional sufficient conditions for positive Harris-recurrence when $\bar{\pi}_{rw} = 1$ and $\mu = 0$ and for null-recurrence[2] when $\bar{\pi}_{rw} = 1$ and $\mu > 0$. The proofs are based on Lyapunov functions (see Meyn and Tweedie, 1993).

PROPOSITION 6.

(i) Let us assume that $\bar{\pi}_{rw} = 1$ *and* $\mu = 0$. *The D.S. Markov chain is positive Harris-recurrent if* $\lim_{y\to\infty}[1 - \pi_{rw}(y)]y^2 = b \in (0;\infty]$ *and* $b > \sigma^2$.
(ii) Let us assume that $\bar{\pi}_{rw} = 1$, $\mu > 0$, *and* $E|\varepsilon_t|^\eta < \infty$ *for some* $\eta > 1$. *The D.S. Markov chain is null-recurrent if* $\lim_{y\to\infty}[1 - \pi_{rw}(y)]y = c > 0$ *and* $\mu > c$.

Proof. See Appendix D.

Condition (i) of Proposition 6 holds for the exponential-type noise distribution without drift ($\alpha = \beta = 1$) discussed in Section 3(iii). It is clearly satisfied

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

if the solution has exponential or Weibull tail. Let us now consider Pareto tail: $l(y) = Ay^{-a}[1 + o(1)]$, where $a > 1$; the intensity π_{rw} satisfies

$$\lim_{y\to\infty}[1 - \pi_{rw}(y)]y^2 = a(a+1) > \sigma^2 = 2.$$

This example shows that the sufficient condition of Proposition 6(i) is rather accurate.

The preceding results can be extended to a situation of stationary autoregressive regime with $0 < \rho < 1$ instead of the white noise regime. In this case, the sufficient conditions for the D.S. Markov chain to be positive Harris-recurrent become[3]

1. $\mu < 0$;
2. $\mu > 0$, $\lim_{y\to\infty}[1 - \pi_{rw}(y)]y = c \in (0;\infty]$, and $c(1-\rho) > \mu$;
3. $\mu = 0$, $\lim_{y\to\infty}[1 - \pi_{rw}(y)]y^2 = b \in (0;\infty]$, and $b(1-\rho^2) > \sigma^2$.

(iv) The Markov chain (Y_t) is geometric ergodic if there exists a scalar $r > 1$ such that

$$\sum_{n=1}^{\infty} r^n \sup_{A\in\mathcal{B}}|P(Y_n \in A|Y_0 = y) - P(Y \in A)| < \infty, \quad \forall y \in \mathbb{R},$$

where Y is a random variable with the stationary distribution.

PROPOSITION 7. *If $\bar{\pi}_{rw} < 1$, the D.S. Markov chain is geometrically ergodic. Moreover, if the distribution of Y_0 is the stationary distribution, the process is β-mixing[4] with mixing coefficients decreasing at a geometric rate.*

Proof. See Appendix D.

4.2. Stationary Distribution

When it exists, the survivor function of the stationary distribution of the D.S. Markov chain can be written as an infinite weighted sum of the survivor functions of S_n, n varying.

PROPOSITION 8. *If the D.S. Markov chain admits a stationary distribution, we have*

$$P(Y > y) = \sum_{n=1}^{\infty} r_n(y)\bar{F}_{\mu}^{*n}(y),$$

where $\bar{F}_{\mu}^{*n}(y) = P(S_n > y)$, $r_n(y) = [1 - \mathbb{E}\pi_{rw}(Y)]\mathbb{E}[\pi_{rw}(S_1)\pi_{rw}(S_2)\dots \pi_{rw}(S_{n-1})|S_n > y]$, if $n > 1$, and $r_1(y) = 1 - \mathbb{E}\pi_{rw}(Y)$, where $\mathbb{E}\pi_{rw}(Y)$ denotes

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

the expectation of $\pi_{rw}(Y)$ with respect to the stationary distribution. Moreover, if the density of the stationary distribution exists, we have[5]

$$l(y) = \sum_{n=1}^{\infty} q_n(y) f_\mu^{*n}(y),$$

where the weights $q_n(y)$ satisfy the recursion

$$q_1(y) f_\mu(y) = [1 - \mathbb{E}\pi_{rw}(Y)] f_\mu(y),$$

$$q_n(y) f_\mu^{*n}(y) = \int \pi_{rw}(x) f_\mu(y - x) q_{n-1}(x) f_\mu^{*n-1}(x)\, dx.$$

Proof. See Appendix E.

The tail behavior of the stationary distribution of the D.S. Markov chain depends not only on the tail behavior of the noise distribution but also on the stochastic intensity through these weights.

For constant regime intensity $\pi_{rw}(y) = \pi_c$, the weights are constant $r_n(y) = q_n(y) = (1 - \pi_c)\pi_c^{n-1}$, and the stationary distribution of the D.S. Markov chain is a compound distribution:

$$P(Y > y) = (1 - \pi_c) \sum_{n=1}^{\infty} \pi_c^{n-1} \bar{F}_\mu^{*n}(y).$$

The following proposition gives a sufficient condition for the existence of moments of the stationary distribution.

PROPOSITION 9. *Assume that $\mu > 0$, $\lim_{y\to\infty}[1 - \pi_{rw}(y)]y = c > \mu$ and that there exists $\alpha > 2$ such that $\mathbb{E}|\varepsilon_1|^\alpha < \infty$. Let $k \leq \alpha$ and $k < c/\mu$. Then $\mathbb{E}|Y|^{k-1} < \infty$.*

Proof. See Appendix E.

The existence of the moments of the stationary distribution depends on the existence of moments of the noise distribution (by mean of α) and of moments of the duration spent in the random walk regime. Indeed, if $\mu > 0$ and $\lim_{y\to\infty}[1 - \pi_{rw}(y)]y = c$, the distribution of the duration spent in the random walk regime D^1 has a Pareto tail with index c/μ (see Proposition 3(iii) and the example following this proposition). In particular, the stationary distribution does not necessarily admit moment at any order even if the noise does.

When the white noise regime is replaced by a stationary autoregressive regime with $\mu_1 = \mu_2 = \mu$ and $0 < \rho < 1$, the conditions of Proposition 9 have to be replaced by the conditions $k \leq \alpha$ and $c(1 - \rho^k) > k\mu$. Therefore, the stationary distribution has fatter tails when ρ increases.

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

4.3. Persistence

By construction, the joint process (Y_t, B_t) is a Markov process, whose history can be summarized by Y_{t-1} only. In this section, we consider the persistence properties of the quantitative and qualitative components (Y_t) and (B_t), respectively. Let us first note that (Y_t) is a Markov process with respect to its own history, whereas the qualitative process alone is not Markovian. Because process (B_t) is a renewal process, its persistence properties are related to the tail behavior of the distribution of the duration spent in the random walk regime. Typically, if this distribution has Pareto tail, that is, if (p_n) features a hyperbolic rate of decay, the autocorrelogramm of (B_t) will also feature asymptotically a hyperbolic rate of decay, which can create persistence effect (see, e.g., Heath, Resnick, and Samorodnitsky, 1998; Granger and Terasvirta, 1999; Gouriéroux and Jasiak, 2001).

The persistence properties of (Y_t) are related to the persistence properties of (B_t). They can be studied by comparing the conditional distribution of Y_t given Y_0 to the marginal one. We first provide the expression of the conditional distribution.

PROPOSITION 10. *The conditional distribution of Y_t given $Y_0 = y_0$ is given by*

$$P(Y_t \in A | Y_0 = y_0)$$
$$= \sum_{n=1}^{t} [1 - \mathbb{E}(\pi_{rw}(Y_{t-n}) | Y_0 = y_0)] \mathbb{E}[\mathbb{I}_{\{S_n \in A\}} \pi_{rw}(S_1) \pi_{rw}(S_2) \dots \pi_{rw}(S_{n-1})]$$
$$+ \mathbb{E}[\mathbb{I}_{\{S_t + y_0 \in A\}} \pi_{rw}(y_0) \pi_{rw}(S_1 + y_0) \dots \pi_{rw}(S_{t-1} + y_0)].$$

Proof. See Appendix E.

When $\bar{\pi}_{rw} < 1$, we have

$$|P(Y_t \in A | Y_0 = y_0) - P(Y \in A)| \leq 2 \bar{\pi}_{rw}^{t-1},$$

and the conditional distribution of Y_t given Y_0 converges to the stationary distribution at a geometric rate. A stronger degree of persistence may be achieved when $\bar{\pi}_{rw} = 1$. As an illustration, Figure 4 displays the paths of both processes when $\pi_{rw}(y) = (1 - y^{-2}) \mathbb{I}_{[1,\infty[}(y)$, whereas the noise distribution is $N(0,1)$ and the initial value is $Y_0 = 0$.

The marginal and cross-autocorrelograms of both processes are provided in Figure 5, where a hyperbolic rate of decay is observed.[6] The sample distribution of durations D_2^1 is displayed in Figure 6, where we also note the slow decrease in the right tail.

The degree of nonlinear persistence of a stationary process may be measured by the way the autocorrelation coefficients of nonlinear functions of the process

$$Corr(g(Y_0); g(Y_t)) \tag{12}$$

approach zero. The proposition that follows provides an asymptotic equivalent of the autocovariance coefficients of $(\pi_{rw}(Y_t))$ when the noise distribution admits

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

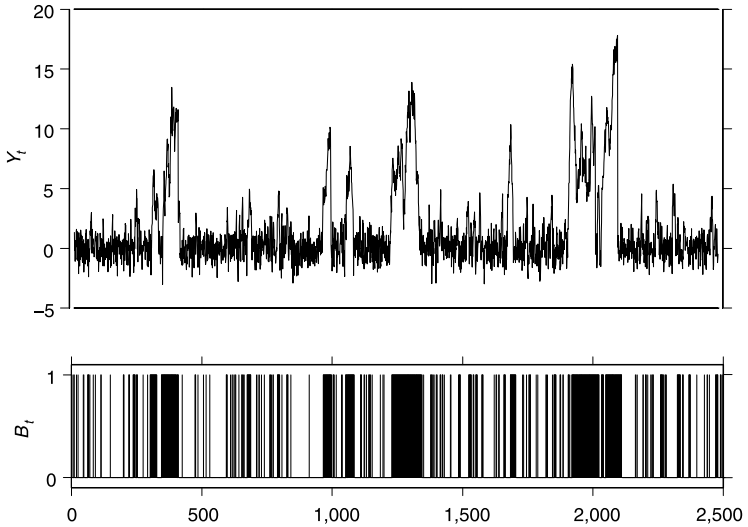


FIGURE 4. Paths of processes (Y_t) and (B_t).

a Laplace transform in a neighborhood of 0 and $\lim_{y\to\infty}[1 - \pi_{rw}(y)]y = c > \mu > 0$.

PROPOSITION 11. *Let us assume that* $\underline{\pi}_{rw} > 0$*, the noise distribution admits a Laplace transform in a neighborhood of* 0*, and* $\lim_{y\to\infty}[1 - \pi_{rw}(y)]y = c$ *is finite such that* $c > \mu > 0$*. If the distribution of* Y_0 *is the stationary distribution, then*

$$c_t = \text{Cov}(\pi_{rw}(Y_0); \pi_{rw}(Y_t)) = [1 - \mathbb{E}\pi_{rw}(Y_0)]^2 \sum_{n=t+1}^{\infty} p_n[1 + o(1)].$$

Proof. See Appendix F.

From Proposition 3, we deduce that we have found a nonlinear function of the D.S. Markov chain with hyperbolic rate of decay of the autocorrelogram (for another example, see Davydov, 1969). Because (p_n/p_1) is the survivor function of D_2^1, the asymptotic equivalent is proportional to $\mathbb{E}[D_2^1 - (t+1)|D_2^1 \geq t+2]$, which is a standard measure of the tail magnitude[7] of D_2^1.

The autocorrelation coefficients (12) may not be precise enough to define appropriately the degree of nonlinear persistence of a stationary process. It appears to be better measured by the ρ-mixing coefficients:

$$\rho_t = \sup\{|Corr(U,V)|; U \in L^2(B_{0,0}), U \in L^2(B_{t,\infty})\},$$

where $B_{i,j} = \sigma\{Y_i, \ldots, Y_j\}$ is the σ-algebra generated by $Y_i, \ldots, Y_j$. The corollary that follows shows that the D.S. Markov chain may have a very high degree of nonlinear dependence.

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

COROLLARY 2. *Let us assume that* $\underline{\pi}_{rw} > 0$*, the noise distribution admits a Laplace transform in a neighborhood of* 0*, and* $\lim_{y\to\infty}[1 - \pi_{rw}(y)]y = c$ *is finite such that* $c > \mu > 0$*. Then* $\rho_t = 1$*, for all* $t \geq 1$.

Proof. The ρ-mixing coefficients of a Markov chain decay to zero at an exponential rate or are identically one (see Bradley, 2005, Thm. 3.3(1)). We have

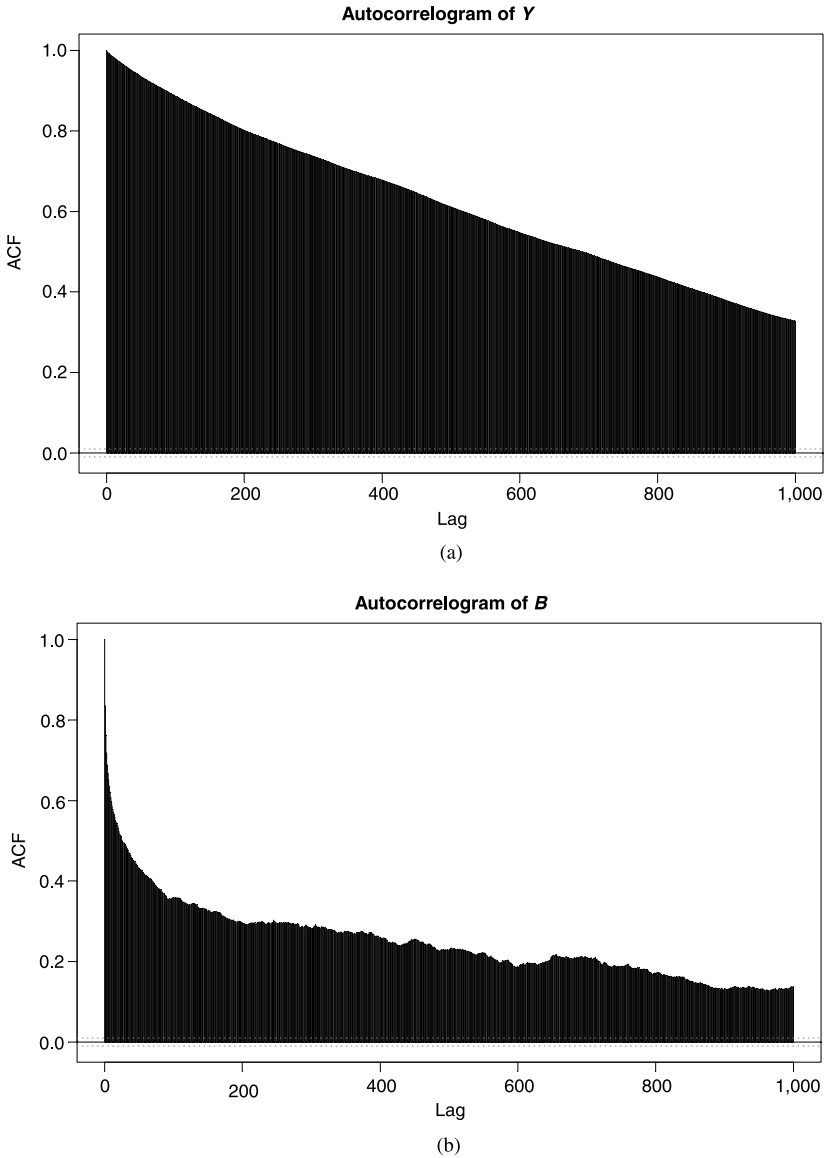
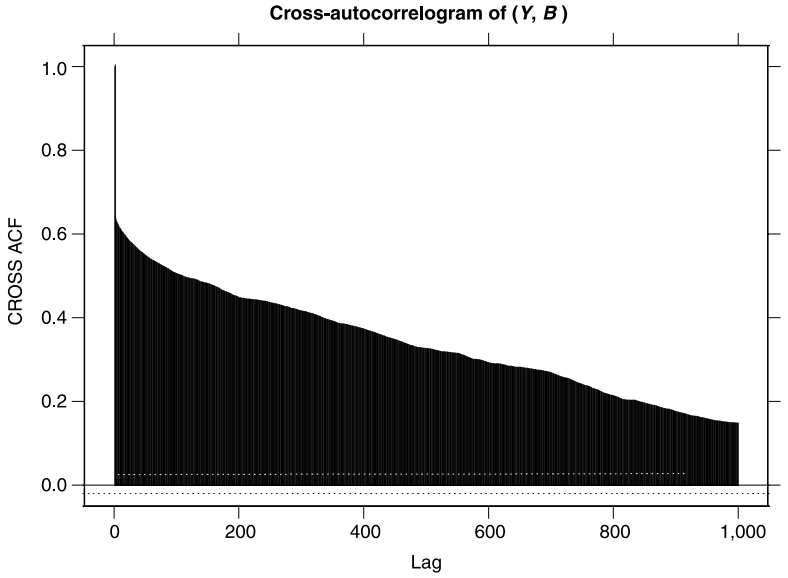


FIGURE 5. Marginal and cross-autocorrelograms. (*Figure continues on next page.*)

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press



(c)

FIGURE 5. *Continued.*

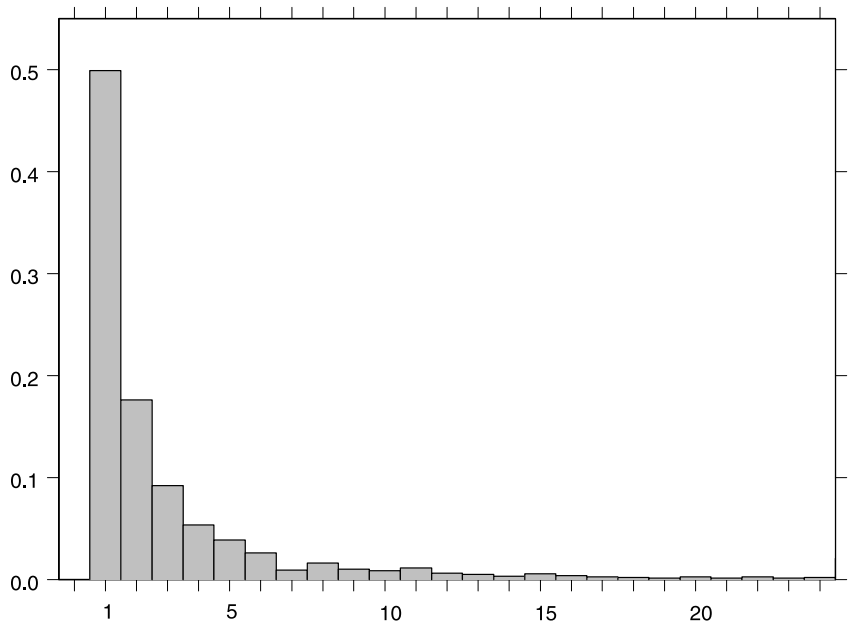


FIGURE 6. Histogram of the duration variable.

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

$\rho_t \geq c_t \geq 0$, and, by Proposition 11, the sequence $(c_t)_{t\geq 0}$ has an hyperbolic rate of decay. Therefore, $\rho_t = 1$ for all $t \geq 0$. ■

5. CONCLUDING REMARKS

Slow decay rates of the linear autocorrelogram are often observed for macroeconomic and financial series. This stylized fact is generally taken into account by considering fractional processes with strong white noise. However, this type of model is not compatible with the domain restrictions on the variables and contradicts the prediction practices. The aim of our paper was to study carefully a class of nonlinear short-memory models able to reconcile stylized facts and practices. The stochastic unit root model is a dynamic model with both a random walk and a stationary regime. Its persistence and fat tail properties depend on the pattern of the random walk regime intensity as a function of the past or equivalently of the tail of the distribution of the duration spent in the random walk regime. The stochastic unit root model is flexible, because it is possible to increase the autoregressive orders in both regimes and to allow unconstrained either the noise distribution or the random walk regime intensity. Moreover, the example of PPP deviation shows that the stochastic unit root model can admit structural interpretations, which is not the case with the standard fractional process. Finally, this model will be very useful for discussing the asymptotic behavior of the unit root tests or estimators of the fractional order d, usually built under specification (1), when this specification (1) is incorrect and the process satisfies a stochastic unit root model (see Shimotsu, 2005).

NOTES

1. It can be checked that this model corresponds to a noise (η_t) with a twice differentiable c.d.f. G satisfying the differential equation $-(G''(y) - G''(0))(G(y) - \frac{1}{2})(\frac{3}{2} - G(y))y/2 = (G'(y) - G'(0))(G(y) - \frac{1}{2}) \times (\frac{3}{2} - G(y)) - 2y(G'(y) - G'(0))^2 G(y)$, for $y > 0$.

2. A Markov chain is null-recurrent if it is Harris-recurrent but not positive Harris-recurrent.

3. The proofs are based on Lyapunov functions and are available on http://www.crest.fr/pageperso/gouriero/gouriero.htm.

4. Let us consider a strictly stationary process $(Y_t, t \geq 0)$. Let $B_{i,j} = \sigma\{Y_i, \ldots, Y_j\}$ be the σ-algebra generated by $Y_i, \ldots, Y_j$. The β-mixing coefficients are defined by

$$\beta(k) = \mathbb{E} \sup_{A \in B_{k,\infty}} |P(A|B_{0,0}) - P(A)|.$$

A process is β-mixing if $\lim_{k\to\infty} \beta(k) = 0$ (see Doukhan, 1994).

5. The nth convolution p.d.f. is defined recursively by

$$f_\mu^{*n}(y) = \int f_\mu(y - x) f_\mu^{*(n-1)}(x)\, dx.$$

6. These theoretical autocorrelograms have been derived by simulations over a large number of replications (10,000).

7. $\mathbb{E}[D_2^1 - (t+1)|D_2^1 \geq t+2] = \sum_{n=t+1}^{\infty} p_n/p_1$ is a standard measure of heaviness of the tail of D_2^1 (see Embrechts, Klüppelberg, and Mikosch, 1997). For instance it corresponds to the tailVaR or expected shortfall proposed to analyze a risk.

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

REFERENCES

Asmussen, S. (1987) *Applied Probability and Queues*. Wiley.

Balke, N.S. & T.B. Fomby (1997) Threshold cointegration. *International Economic Review* 38, 627–646.

Bec, F., M. Ben Salem, & M. Carrasco (2004) Test for unit root versus threshold specification with an application to PPP. *Journal of Business & Economic Statistics* 22, 382–395.

Billingsley, P. (1995) *Probability and Measure*, 3rd ed. Wiley Series in Probability and Mathematical Statistics. Wiley.

Bradley, R.C. (2005) Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* 2, 107–144.

Brooks, C. & A. Reveiz (2002) A model for exchange rates with crawling bands—An application to the Colombian peso. *Journal of Business and Economics* 54, 483–503.

Clark, P.K. (1973) A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 41, 135–156.

Davydov, Y. (1969) On a strong mixing property for Markov chains with a countable number of states. *Soviet Doklady* 187, 825–827.

Doukhan, P. (1994) *Mixing: Properties and Examples.* Lecture Notes in Statistics 85, Springer.

Embrechts, P., C. Klüppelberg, & T. Mikosch (1997) *Modelling Extremal Events*. Applications of Mathematics. Springer.

Enders, W. & C. Granger (1998) Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates. *Journal of Business & Economic Statistics* 16, 304–311.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1008.

Farnsworth, H. & R. Bass (2003) The term structure with credible and semi-credible targeting. *Journal of Finance* 58, 839–866.

Feller, W. (1966) *An Introduction to Probability Theory and Its Applications II*. Wiley.

Gouriéroux, C. & J. Jasiak (2001) Memory and infrequent break. *Economics Letters* 70, 29–41.

Granger, C. & R. Joyeux (1980) An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1, 15–39.

Granger, C. & N. Swanson (1997) An introduction to stochastic unit root processes. *Journal of Econometrics* 80, 35–62.

Granger, C. & J. Terasvirta (1999) Simple nonlinear time series model with misleading linear properties. *Economic Letters* 62, 161–165.

Hamilton, J. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 5, 357–384.

Heath, D., S. Resnick, & G. Samorodnitsky (1998) Heavy tails and long range dependence in on/off processes and associated fluid models. *Mathematics of Operations Research* 23, 145–165.

MacDonald, I.L. & W. Zucchini (1997) *Hidden Markov and Other Models for Discrete-Valued Time Series*. Chapman and Hall.

Mandelbrot, B. & H. Taylor (1967) On the distribution of stock prices differences. *Operational Research* 15, 1057–1062.

Meyn, S.P. & R.L. Tweedie (1993) *Markov Chains and Stochastic Stability*. Springer.

Rahbek, A. & N. Shephard (2002) Inference and Ergodicity in the Autoregressive Conditional Root Model. Discussion paper, University of Oxford.

Sercu, P., R. Uppal, & C. Van Hulle (1995) The exchange rate in the presence of transaction costs: Implications for tests of purchasing power parity. *Journal of Finance* 50, 1309–1319.

Shimotsu, K. (2005) Simple (but effective) Tests of Long Memory versus Structural Breaks. Discussion paper, Queen's University.

Tong, H. (1990) *Non-linear Time Series—A Dynamical System Approach*. Oxford University Press.

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

APPENDIX A: Distribution of the Switching Regime Process (Proof of Proposition 1)

Let us first consider an initial condition Y_0, $B_1 = 0$ (say), which ensures that the first spell corresponds to the white noise regime. The distribution of the $2j$ first durations (conditional on Y_0, $B_1 = 0$) is

$$
\begin{aligned}
&P(D_1^0 = d_1^0, D_1^1 = d_1^1, D_2^0 = d_2^0, D_2^1 = d_2^1, \ldots, D_j^0 = d_j^0, D_j^1 = d_j^1 | Y_0, B_1 = 0) \\
&\quad = P((B_t = 0, 2 \le t \le d_1^0) \cap (B_t = 1, d_1^0 + 1 \le t \le d_1^0 + d_1^1) \ldots \\
&\qquad \cap (B_t = 1, d_1^0 + \cdots + d_j^0 + 1 \le t \le d_1^0 + \cdots + d_j^0 + d_j^1) \\
&\qquad \cap (B_{d_1^0 + d_1^1 + \cdots + d_j^0 + d_j^1 + 1} = 0) | Y_0, B_1 = 0) \\
&\quad = \mathbb{E}[P((B_t = 0, 2 \le t \le d_1^0) \cap (B_t = 1, d_1^0 + 1 \le t \le d_1^0 + d_1^1) \ldots \\
&\qquad \cap (B_t = 1, d_1^0 + \cdots + d_j^0 + 1 \le t \le d_1^0 + \cdots + d_j^0 + d_j^1) \ldots . \\
&\qquad \cap (B_{d_1^0 + d_1^1 + \cdots + d_j^0 + d_j^1 + 1} = 0) | Y_0, B_1 = 0, \varepsilon_t, 1 \le t \le d_1^0 + \cdots + d_j^0 + d_j^1) | Y_0, B_1 = 0].
\end{aligned}
$$

Conditional on the noise (ε_t), the regime indicators are independent with different Bernoulli distributions. We deduce that

$$
\begin{aligned}
&P(D_1^0 = d_1^0, D_1^1 = d_1^1, D_2^0 = d_2^0, D_2^1 = d_2^1, \ldots, D_j^0 = d_j^0, D_j^1 = d_j^1 | Y_0, B_1 = 0) \\
&\quad = \mathbb{E}\left[\prod_{t=2}^{d_1^0} (1 - \pi_{rw}(\varepsilon_{t-1} + \mu)) \prod_{t=d_1^0+1}^{d_1^0+d_1^1} \pi_{rw}(S_{t-1} - S_{d_1^0-1})(1 - \pi_{rw}(S_{d_1^0+d_1^1} - S_{d_1^0-1})) \ldots \right. \\
&\qquad \times \prod_{t=d_1^0+\cdots+d_{j-1}^1+1}^{d_1^0+\cdots+d_j^0} (1 - \pi_{rw}(\varepsilon_{t-1} + \mu)) \prod_{t=d_1^0+\cdots+d_j^0+1}^{d_1^0+\cdots+d_j^0+d_j^1} \pi_{rw}(S_{t-1} - S_{d_1^0+\cdots+d_j^0-1}) \ldots \\
&\qquad \left. \times (1 - \pi_{rw}(S_{d_1^0+\cdots+d_j^0+d_j^1} - S_{d_1^0+\cdots+d_j^0-1})) | Y_0, B_1 = 0 \right];
\end{aligned}
$$

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

(by conditional independence)

$$= \prod_{t=2}^{d_1^0} (1 - \mathbb{E}\pi_{rw}(\varepsilon_{t-1} + \mu))$$

$$\times \left(\mathbb{E} \prod_{t=d_1^0+1}^{d_1^0+d_1^1} \pi_{rw}(S_{t-1} - S_{d_1^0-1}) - \mathbb{E} \prod_{t=d_1^0+1}^{d_1^0+d_1^1+1} \pi_{rw}(S_{t-1} - S_{d_1^0-1}) \right) \cdots$$

$$\times \prod_{t=d_1^0+\cdots+d_{j-1}^0+1}^{d_1^0+\cdots+d_j^0} (1 - \mathbb{E}\pi_{rw}(\varepsilon_{t-1} + \mu))$$

$$\times \left(\mathbb{E} \prod_{t=d_1^0+\cdots+d_j^0+1}^{d_1^0+\cdots+d_j^0+d_j^1} \pi_{rw}(S_{t-1} - S_{d_1^0+\cdots+d_j^0-1}) \right.$$

$$\left. - \mathbb{E} \prod_{t=d_1^0+\cdots+d_j^0+1}^{d_1^0+\cdots+d_j^0+d_j^1+1} \pi_{rw}(S_{t-1} - S_{d_1^0+\cdots+d_j^0-1}) \right);$$

(by the property of the noise)

$$= \prod_{i=1}^{j} (1 - p_1)^{d_i^0 - 1} \prod_{i=1}^{j} (p_{d_i^1} - p_{d_i^1+1}), \quad \text{(by definition of } p_k\text{)}.$$

Therefore, we get

$$P(D_1^0 = d_1^0, D_1^1 = d_1^1, D_2^0 = d_2^0, D_2^1 = d_2^1, \ldots, D_j^0 = d_j^0, D_j^1 = d_j^1 \,|\, Y_0, B_1 = 0)$$

$$= \prod_{i=1}^{j} p_1 (1 - p_1)^{d_i^0 - 1} \prod_{i=1}^{j} \frac{p_{d_i^1} - p_{d_i^1+1}}{p_1}.$$

A similar computation can be performed conditional on Y_0, $B_1 = 1$.

The multiplicative decomposition of the joint duration distribution provides the independence of the durations and their stationary distributions. Finally, the probability that the duration spent in the random walk regime is finite is given by

$$P(D_j^1 < \infty) = \lim_{d_j^1 \to \infty} P(D_j^1 \le d_j^1) = \lim_{d_j^1 \to \infty} \frac{p_1 - p_{d_{j+1}^1}}{p_1}.$$

This probability is equal to one if and only if $\lim_{k\to\infty} p_k = 0$, which is condition (i). ■

APPENDIX B: Bounds for the Probabilities to Stay in the Random Walk Regime (Proof of Proposition 3)

Lower Bound. Let us assume that $\underline{\pi}_{rw} > 0$. The law of iterated logarithm ensures that for any $\lambda > 1$

$$S_n < n\mu - \lambda(2\sigma^2 n \log\log n)^{1/2}, \qquad n \ge 3,$$

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

a finite number of times a.s. Let A be the random set of indices for which the previous inequalities are jointly satisfied and a be the cardinal number of A. Because intensity π_{rw} is a nondecreasing function, we get

$$\prod_{i=1}^{n} \pi_{rw}(S_i) > \underline{\pi}_{rw}^{a+2} \exp\left(\sum_{i \in [[3;n]] \setminus A} \log \pi_{rw}(i\mu - \lambda(2\sigma^2 i \log\log i)^{1/2}) \right)$$

$$\geq \underline{\pi}_{rw}^{a+2} \exp\left(\sum_{i=3}^{n} \log \pi_{rw}(i\mu - \lambda(2\sigma^2 i \log\log i)^{1/2}) \right)$$

and

$$p_n > \mathbb{E}(\underline{\pi}_{rw}^{a+2}) \exp\left(\sum_{i=3}^{n} \log \pi_{rw}(i\mu - \lambda(2\sigma^2 i \log\log i)^{1/2}) \right),$$

where $\mathbb{E}(\underline{\pi}_{rw}^{a+2})$ is strictly positive because a is finite a.s.

Upper Bound. Let us recall some lemmas.

LEMMA A.1 (Feller, 1966, p. 398). *Let $X_1, X_2, \ldots$ be i.i.d. random variables and $\Sigma_n = X_1 + \cdots + X_n$. Let $H_n = \{i : i \leq n, \Sigma_i \geq 0\}$, $C_n = card(H_n)$ be the number of nonnegative terms among $\Sigma_1, \ldots, \Sigma_n$. Then*

$$P(C_n = k) = a_k b_{n-k},$$

where $a_k = P(\Sigma_1 > 0, \ldots, \Sigma_k > 0)$, $b_k = P[\mathrm{Max}(\Sigma_1, \ldots, \Sigma_k) \leq 0]$.

Because C_n increases with n, C_n converges a.s. to a limit C, possibly with infinite value. The term C counts the number of nonnegative terms among $\Sigma_1, \Sigma_2, \ldots$. We have

$$b_{n-k} = P[\mathrm{Max}(\Sigma_1, \ldots, \Sigma_k) \leq 0] = P[X + \mathrm{Max}(0, \Sigma_1, \ldots, \Sigma_{n-k-1}) \leq 0],$$

where X is independent of $X_1, X_2, \ldots$ and has the same distribution as X_n. The random variable $\mathrm{Max}(0, \Sigma_1, \ldots, \Sigma_{n-k-1})$ has the same distribution as U_{n-k-1}, where (U_n) is the Lindley process defined by $U_n = (U_{n-1} + X_n)^+$ (cf. Asmussen, 1987, Prop. 1.1, p. 181). If $\mathbb{E}X_n < 0$, $\mathrm{Max}(0, \Sigma_1, \ldots, \Sigma_{n-k-1})$ weakly converges to a nondegenerate positive distribution as n tends to infinity. Because $P(X_n < 0) > 0$ and $\lim_{n\to\infty} P[\mathrm{Max}(0, \Sigma_1, \ldots, \Sigma_{n-k-1}) = 0] > 0$, we deduce that $\lim_{n\to\infty} b_{n-k} = b > 0$, and the limit C is such that

$$P(C = k) = a_k b.$$

LEMMA A.2 (Chernoff). (See, e.g., Billingsley, 1995, p. 151). *Let $X_1, X_2, \ldots$ be i.i.d. random variables satisfying $\mathbb{E}X_n < 0$ and $P(X_n > 0) > 0$. Let $\Sigma_n = X_1 + \cdots + X_n$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \log P(\Sigma_n \geq 0) = \log \rho,$$

where $\rho = \inf_t \Psi_X(t)$ and $\Psi_X(t)$ is the Laplace transform of X_1.

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

The preceding lemmas are used to derive the upper bound for p_n according to the sign of μ.

(i) Let us first consider the case $\mu > 0$ and denote $X_i = \varepsilon_i - (\lambda - 1)\mu$, where $\lambda > 1$. Note that $\mathbb{E}(X_i) < 0$ and $\Sigma_n = S_n - n\lambda\mu$. We have

$$\begin{aligned}
&\prod_{i=1}^{n} \pi_{rw}(S_i) \\
&\quad = \exp\left(\sum_{i=1}^{n} \log \pi_{rw}(S_i)\right) \\
&\quad \le \exp\left(\sum_{i\in[[1;n]]\setminus H_n} \log \pi_{rw}(i\lambda\mu)\right), \quad \text{where } H_n = \{i : i \le n, S_i \ge i\lambda\mu\}, \\
&\quad \le \exp\left(\sum_{i=C_n+1}^{n} \log \pi_{rw}(i\lambda\mu)\right), \quad \text{because } \pi_{rw} \text{ is nondecreasing}, \\
&\quad \le \exp\left(-\sum_{i=1}^{C} \log \pi_{rw}(i\lambda\mu)\right) \exp\left(\sum_{i=1}^{n} \log \pi_{rw}(i\lambda\mu)\right), \quad \text{because } C_n \le C.
\end{aligned}$$

Then, we get the upper bound:

$$p_n \le \exp\left(\sum_{i=1}^{n} \log \pi_{rw}(i\lambda\mu)\right) \mathbb{E} \exp\left(-\sum_{i=1}^{C} \log \pi_{rw}(i\lambda\mu)\right).$$

It remains to prove that $\mathbb{E}\exp(-\sum_{i=1}^{C} \log \pi_{rw}(i\lambda\mu))$ is finite. For this purpose, let us first consider the Laplace transform of C. From Lemma A.1, we have

$$\mathbb{E}\exp(tC) = b\sum_{k=0}^{\infty} a_k \exp(tk) \le b\sum_{k=0}^{\infty} [P(\Sigma_k \ge 0)\exp(tk)].$$

From Lemma A.2, we get

$$\mathbb{E}\exp(tC) \le b\sum_{k=0}^{\infty} [(\rho'_\lambda)^k \exp(tk)],$$

where $1 > \rho'_\lambda > \rho_\lambda = \inf_t[\Psi_\varepsilon(t)e^{-t(\lambda-1)\mu}]$. We deduce that $\mathbb{E}\exp(tC)$ is finite if $t < -\log \rho'_\lambda$.

Then, we have

$$\begin{aligned}
&\mathbb{E}\exp\left(-\sum_{i=1}^{C} \log \pi_{rw}(i\lambda\mu)\right) \\
&\quad \le \exp\left(-\sum_{i=1}^{N_0} [\log \pi_{rw}(i\lambda\mu) - \log \pi_{rw}(N_0\lambda\mu)]\right) \mathbb{E}\exp[-\log \pi_{rw}(N_0\lambda\mu)C],
\end{aligned}$$

for any integer N_0. The bound is finite if N_0 is sufficiently large, for instance, if $N_0 = \inf\{i : -\log \pi_{rw}(i\lambda\mu) < -\log \rho'_\lambda\}$, which is finite because $\bar{\pi}_{rw} = 1$.

(ii) In the case $\mu = 0$, we select $X_i = \varepsilon_i - \lambda$, where $\lambda > 0$, and conclude as in (i).

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

(iii) In the case $\mu < 0$, let us select $X_i = \varepsilon_i + \kappa\mu$, where $0 < \kappa < 1$. Note that $\mathbb{E}(X_i) < 0$ and $\Sigma_n = S_n - n(1-\kappa)\mu$. Then

$$\begin{aligned}
&\prod_{i=1}^{n} \pi_{rw}(S_i) \\
&\quad = \exp\left(\sum_{i=1}^{n} \log \pi_{rw}(S_i)\right) \\
&\quad \le \exp\left(\sum_{i\in[[1;n]]\setminus H_n} \log \pi_{rw}[i(1-\kappa)\mu]\right), \quad \text{where } H_n = \{i : i \le n, S_i \ge i(1-\kappa)\mu\}, \\
&\quad \le \exp\left(\sum_{i=1}^{n-C_n} \log \pi_{rw}[i(1-\kappa)\mu]\right).
\end{aligned}$$

For any $i \ge 1$ and $t > 0$, $\pi_{rw}[i(1-\kappa)\mu] \le \pi_{rw}(t)$, because π_{rw} is nondecreasing. Therefore:

$$\begin{aligned}
\prod_{i=1}^{n} \pi_{rw}(S_i) &\le \pi_{rw}^n(t)\exp(-[\log \pi_{rw}(t)]C_n) \\
&\le \pi_{rw}^n(t)\exp(-[\log \pi_{rw}(t)]C).
\end{aligned}$$

As in part (i), $\mathbb{E}\exp(-[\log \pi_{rw}(t)]C)$ is finite if $-\log \pi_{rw}(t) < \rho_\kappa = \inf_t[\Psi_\varepsilon(t)e^{t\kappa\mu}]$. Let $N_1 = \inf\{j \ge 1 : -\log \pi_{rw}(j) < \rho_\kappa\}$. We get

$$p_n \le \pi_{rw}^n(N_1)\mathbb{E}\exp[-\log \pi_{rw}(N_1)C].$$

The result follows with $\theta = \pi_{rw}(N_1)$ and $\varkappa = \mathbb{E}\exp[-\log \pi_{rw}(N_1)C]$. ■

APPENDIX C: Stationary Distribution for Exponential Noise (Proof of Proposition 5)

The stationary density l satisfies the integral equation

$$\begin{aligned}
l(y) &= \int_{-\infty}^{\infty} [\pi_{rw}(x)f_\mu(y-x) + (1-\pi_{rw}(x))f_\mu(y)]l(x)\,dx \\
&= \frac{\alpha}{2}\exp(-\alpha y)\left[\mathbb{I}_{[0,\infty[}(y)\left(1 - \int_{-\infty}^{\infty} \pi_{rw}(x)l(x)\,dx\right) + \int_{-\infty}^{y} \pi_{rw}(x)\exp(\alpha x)l(x)\,dx\right] \\
&\quad + \frac{\beta}{2}\exp(\beta y)\left[\mathbb{I}_{]-\infty,0[}(y)\left(1 - \int_{-\infty}^{\infty} \pi_{rw}(x)l(x)\,dx\right) + \int_{y}^{\infty} \pi_{rw}(x)\exp(-\beta x)l(x)\,dx\right].
\end{aligned}$$

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

Let us denote $h(y) = \exp(\alpha y) l(y)$. For positive y, the equation becomes

$$h(y) = \frac{\alpha}{2}\left[\left(1 - \int_{-\infty}^{\infty} \pi_{rw}(x) l(x)\, dx\right) + \int_{-\infty}^{y} \pi_{rw}(x) h(x)\, dx\right]$$

$$+ \frac{\beta}{2} \exp[(\alpha + \beta) y] \int_{y}^{\infty} \pi_{rw}(x) \exp(-(\alpha + \beta) x) h(x)\, dx.$$

By differentiating both sides of the equation, we get

$$h'(y) = \frac{\alpha - \beta}{2} \pi_{rw}(y) h(y) + (\alpha + \beta) \frac{\beta}{2} \exp[(\alpha + \beta) y]$$

$$\times \int_{y}^{\infty} \pi_{rw}(x) \exp[-(\alpha + \beta) x] h(x)\, dx$$

$$= \frac{\alpha - \beta}{2} \pi_{rw}(y) h(y)$$

$$+ (\alpha + \beta)\left[h(y) - \frac{\alpha}{2}\left[\left(1 - \int_{-\infty}^{\infty} \pi_{rw}(x) l(x)\, dx\right) + \int_{-\infty}^{y} \pi_{rw}(x) h(x)\, dx\right]\right]$$

$$= h(y)\left[\frac{(\alpha - \beta)}{2} \pi_{rw}(y) + (\alpha + \beta)\right]$$

$$- \frac{\alpha}{2} (\alpha + \beta)\left[\left(1 - \int_{-\infty}^{\infty} \pi_{rw}(x) l(x)\, dx\right) + \int_{-\infty}^{y} \pi_{rw}(x) h(x)\, dx\right].$$

We deduce

$$h''(y) = h'(y)\left[\frac{(\alpha - \beta)}{2} \pi_{rw}(y) + (\alpha + \beta)\right]$$

$$+ h(y)\left[\frac{(\alpha - \beta)}{2} \pi'_{rw}(y) - \frac{\alpha}{2} (\alpha + \beta) \pi_{rw}(y)\right],$$

which is equivalent to

$$l''(y) - \alpha\beta\mu[2 - \pi_{rw}(y)] l'(y) - \alpha\beta[(1 - \pi_{rw}(y)) - \mu\pi'_{rw}(y)] l(y) = 0.$$

A similar approach can be applied to nonpositive values with $h(y) = \exp(-\beta y) l(y)$. ■

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

APPENDIX D: Ergodicity Property

Proof of Proposition 6. From Theorem 11.0.1 of Meyn and Tweedie (1993), the Markov chain is positive Harris-recurrent if there exist a petite set C, a constant $b < \infty$, and a Lyapunov function $V \geq 0$, finite at some $y \in \mathbb{R}$, such that

$$\Delta V(y) = \mathbb{E}[V(Y_1)|Y_0 = y] - V(y) \leq -1 + d\mathbb{I}_C(y), \quad \forall y \in \mathbb{R}.$$

Recall that a set C is petite if the chain satisfies

$$P(Y_1 \in A|Y_0 = y) \geq \tilde{\nu}(A),$$

for any $y \in C$, any $A \in \mathcal{B}$, where $\tilde{\nu}$ is a nondegenerate measure on $\mathcal{B}$. Because intensity π_{rw} is nondecreasing, for any positive R there exists a constant $\eta(R)$ such that, for any $y \leq R$ and any Borel set A,

$$P(Y_1 \in A|Y_0 = y) \geq \eta(R)P(\varepsilon_1 + \mu \in A).$$

This means that every set of the form $(-\infty; R]$ is petite.

Let us consider the function $V(y) = ay^2$. We have

$$\begin{aligned} a^{-1}\Delta V(y) &= \pi_{rw}(y)(y^2 + \sigma^2) + [1 - \pi_{rw}(y)]\sigma^2 - y^2 \\ &= -[1 - \pi_{rw}(y)]y^2 + \sigma^2. \end{aligned}$$

If b is finite, let us consider $\delta > 0$ such that $b - \sigma^2 - \delta > 0$ and define $a^{-1} = b - \sigma^2 - \delta$. If b is not finite, let us choose $a = 1$. There exists $R > 0$ such that for $y > R$, $\Delta V(y) \leq -1$. Then we have

$$\Delta V(y) \leq -1 + d\mathbb{I}_C(y),$$

where $C = \{y, y \leq R\}$ and $R = \inf\{z : \forall y \geq z, \Delta V(y) \leq -1\}$.

Null-recurrence. From Theorem 11.5.1 of Meyn and Tweedie (1993), the Markov chain is null-recurrent if there exist a petite set C, a Lyapunov function $V \geq 0$ such that

$$\Delta V(y) \geq 0, \quad \forall y \in \mathbb{R}\backslash C,$$

and

$$\sup_{y \in \mathbb{R}} \mathbb{E}[|V(Y_1) - V(Y_0)||Y_0 = y] < \infty.$$

Let us consider the function

$$V(y) = y\mathbb{I}_{\{y>0\}};$$

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

then

$$\Delta V(y) = \pi_{rw}(y)\mathbb{E}[(y+\varepsilon_1+\mu)\mathbb{I}_{\{y+\varepsilon_1+\mu>0\}}] + [1-\pi_{rw}(y)]\mathbb{E}[(\varepsilon_1+\mu)\mathbb{I}_{\{\varepsilon_1+\mu>0\}}] - y\mathbb{I}_{\{y>0\}}.$$

Because

$$\mathbb{E}[(y+\varepsilon_1+\mu)\mathbb{I}_{\{y+\varepsilon_1+\mu>0\}}] = \int_{-\infty}^{+\infty}(y+x)\mathbb{I}_{\{y+x>0\}}f_\mu(x)\,dx = \int_{-y}^{+\infty}(y+x)f_\mu(x)\,dx$$
$$= y+\mu-yP(\varepsilon_1+\mu<-y) - \int_{-\infty}^{-y}xf_\mu(x)\,dx,$$

we have

$$\Delta V(y) = \pi_{rw}(y)\left(y+\mu-yP(\varepsilon_1+\mu<-y) - \int_{-\infty}^{-y}xf_\mu(x)\,dx\right) + [1-\pi_{rw}(y)]\mathbb{E}[(\varepsilon_1+\mu)\mathbb{I}_{\{\varepsilon_1+\mu>0\}}] - y\mathbb{I}_{\{y>0\}}$$
$$= (y-y\mathbb{I}_{\{y>0\}}) + [\pi_{rw}(y)-1]y + \pi_{rw}(y)\mu + [1-\pi_{rw}(y)]\mathbb{E}[(\varepsilon_1+\mu)\mathbb{I}_{\{\varepsilon_1+\mu>0\}}] - \pi_{rw}(y)\left(yP(\varepsilon_1+\mu<-y) + \int_{-\infty}^{-y}xf_\mu(x)\,dx\right).$$

Because $\mathbb{E}|\varepsilon_t|^\eta < \infty$ for some $\eta > 1$, the last terms of the decomposition are asymptotically negligible:

$$\lim_{y\to\infty} yP(\varepsilon_1+\mu<-y) = 0, \qquad \lim_{y\to\infty}\int_{-\infty}^{-y}xf_\mu(x)\,dx = 0.$$

Thus we can focus on the other terms to find the petite set $C = (-\infty;R]$. If $\lim_{y\to\infty}[1-\pi_{rw}(y)]y = c > 0$ and $\mu > c$, then $\lim_{y\to\infty}\Delta V(y) = \mu - c > 0$. There exists $R > 0$ such that for $y > R$, $\Delta V(y) \geq 0$.

Let us now focus on the second condition for a null-recurrent chain. We have

$$\mathbb{E}[|V(Y_1)-V(Y_0)||Y_0=y]$$
$$= \pi_{rw}(y)\mathbb{E}|[(y+\varepsilon_1+\mu)\mathbb{I}_{\{y+\varepsilon_1+\mu>0\}}] - y\mathbb{I}_{\{y>0\}}| + [1-\pi_{rw}(y)]\mathbb{E}|[(\varepsilon_1+\mu)\mathbb{I}_{\{\varepsilon_1+\mu>0\}}] - y\mathbb{I}_{\{y>0\}}|$$
$$\leq \mathbb{E}|\varepsilon_1+\mu| + \mathbb{E}|y(\mathbb{I}_{\{y+\varepsilon_1+\mu>0\}} - \mathbb{I}_{\{y>0\}})| + \mathbb{E}|\varepsilon_1+\mu| + [1-\pi_{rw}(y)]y\mathbb{I}_{\{y>0\}}$$
$$\leq y\mathbb{I}_{\{y>0\}}P(\varepsilon_1+\mu<-y) - y\mathbb{I}_{\{y<0\}}P(\varepsilon_1+\mu>-y) + 2\mathbb{E}|\varepsilon_1+\mu| + [1-\pi_{rw}(y)]y\mathbb{I}_{\{y>0\}}$$

and $\sup_{y\in\mathbb{R}}\mathbb{E}(|V(Y_1)-V(y)||Y_0=y) < \infty$. The result follows. ■

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

Proof of Proposition 7. It is easily seen that the Markov process is aperiodic by using Proposition A1.2 of Tong (1990). Let us now prove that the Markov process is geometrically ergodic. According to Theorem 15.0.1 of Meyn and Tweedie (1993), we have to find a petite set C, scalars $b < \infty$, $\beta > 0$, and a Lyapunov function $V \geq 1$ such that

$$\Delta V(y) = \mathbb{E}[V(Y_1)|Y_0 = y] - V(y) \leq -\beta V(y) + b\mathbb{I}_C(y), \quad \forall y \in \mathbb{R}. \tag{D.1}$$

Let us consider the function

$$V(y) = 1 + |y|;$$

we get

$$\begin{aligned}\mathbb{E}[V(Y_1)|Y_0 = y] &= 1 + \pi_{rw}(y)\mathbb{E}|y + \varepsilon_1 + \mu| + (1 - \pi_{rw}(y))\mathbb{E}|\varepsilon_1 + \mu| \\ &\leq 1 + \pi_{rw}(y)|y| + \mathbb{E}|\varepsilon_1 + \mu| \leq 1 + \bar{\pi}_{rw}|y| + \mathbb{E}|\varepsilon_1 + \mu| \\ &= \bar{\pi}_{rw}V(y) + \mathbb{E}|\varepsilon_1 + \mu| + (1 - \bar{\pi}_{rw}),\end{aligned}$$

and then for $0 < \theta < 1$:

$$\begin{aligned}\Delta V(y) &\leq -(1 - \bar{\pi}_{rw})V(y) + \mathbb{E}|\varepsilon_1 + \mu| + (1 - \bar{\pi}_{rw}) \\ &\leq -(1 - \bar{\pi}_{rw})\theta V(y) + \mathbb{E}|\varepsilon_1 + \mu| + (1 - \bar{\pi}_{rw}) - (1 - \theta)(1 - \bar{\pi}_{rw})V(y) \\ &\leq -\beta V(y) + b\mathbb{I}_C(y),\end{aligned}$$

where $\beta = (1 - \bar{\pi}_{rw})\theta > 0$, $b = \mathbb{E}|\varepsilon_1 + \mu| + (1 - \bar{\pi}_{rw}) > 0$, and C is the set defined by

$$C = \{y; V(y) \leq (\mathbb{E}|\varepsilon_1 + \mu| + (1 - \bar{\pi}_{rw}))/(1 - \theta)(1 - \bar{\pi}_{rw})\}.$$

The set C is not empty, because $[\mathbb{E}|\varepsilon_1 + \mu| + (1 - \bar{\pi}_{rw})]/(1 - \theta)(1 - \bar{\pi}_{rw}) > 1$, and it is easily seen that it is petite. Therefore, there exist scalars $c < \infty$ and $r > 1$ such that

$$\sum_{n=1}^{\infty} r^n \sup_{A \in \mathcal{B}} |P(Y_n \in A|Y_0 = y) - P_Y(A)| < cV(y), \quad \forall y \in \mathbb{R}. \tag{D.2}$$

In particular, because the Lyapunov function V is finite, the condition for geometric ergodicity is satisfied. Moreover, condition (D.2) implies

$$\sup_{A \in \mathcal{B}} |P(Y_n \in A|Y_0 = y) - P(Y_n \in A)| \leq cr^{-n}V(y), \qquad n \geq 1. \tag{D.3}$$

Finally, let us now consider the β-mixing condition. It is well known that a stationary, geometrically ergodic Markov chain is β-mixing with mixing coefficients decaying at a geometric rate (see Doukhan, 1994). ■

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

APPENDIX E: Marginal and Conditional Distributions

Stationary Distribution (Proof of Proposition 8). Let us assume that the distribution of Y_0 is the stationary distribution. Then, for any $t > 0$:

$$
\begin{aligned}
P(Y_t > y) &= \sum_{n=1}^{t} P(Y_t > y, B_t = 1, \ldots, B_{t-n+2} = 1, B_{t-n+1} = 0) \\
&\quad + P(Y_t > y, B_t = 1, \ldots, B_2 = 1, B_1 = 1) \\
&= \sum_{n=1}^{t} P\left(\sum_{k=1}^{n} \varepsilon_{t-k+1} + n\mu > y, B_t = 1, \ldots, B_{t-n+2} = 1, B_{t-n+1} = 0 \right) \\
&\quad + P\left(\sum_{k=1}^{t} \varepsilon_{t-k+1} + t\mu + Y_0 > y, B_t = 1, \ldots, B_2 = 1, B_1 = 1 \right) \\
&= \sum_{n=1}^{t} \mathbb{E}\left(\mathbb{I}_{\left\{ \sum_{k=1}^{n} \varepsilon_{t-k+1} + n\mu > y \right\}} [1 - \pi_{rw}(Y_{t-n})] \pi_{rw}(\varepsilon_{t-n+1} + \mu) \ldots \right. \\
&\qquad \left. \times \pi_{rw}\left(\sum_{k=2}^{n} \varepsilon_{t-k+1} + (n-1)\mu \right) \right) \\
&\quad + \mathbb{E}\left(\mathbb{I}_{\left\{ \sum_{k=1}^{t} \varepsilon_{t-k+1} + t\mu + Y_0 > y \right\}} \pi_{rw}(Y_0) \pi_{rw}(\varepsilon_1 + \mu + Y_0) \ldots \right. \\
&\qquad \left. \times \pi_{rw}\left(\sum_{k=2}^{t} \varepsilon_{t-k+1} + (t-1)\mu + Y_0 \right) \right) \\
&= \sum_{n=1}^{t} [1 - \mathbb{E}\pi_{rw}(Y)] \mathbb{E}[\mathbb{I}_{\{S_n > y\}} \pi_{rw}(S_1) \pi_{rw}(S_2) \ldots \pi_{rw}(S_{n-1})] \\
&\quad + \mathbb{E}[\mathbb{I}_{\{S_t + Y_0 > y\}} \pi_{rw}(Y_0) \pi_{rw}(S_1 + Y_0) \pi_{rw}(S_2 + Y_0) \ldots \pi_{rw}(S_{t-1} + Y_0)].
\end{aligned}
$$

By the dominated convergence theorem, we have

$$
\lim_{t \to \infty} \mathbb{E}[\mathbb{I}_{\{S_t + Y_0 > y\}} \pi_{rw}(Y_0) \pi_{rw}(S_1 + Y_0) \pi_{rw}(S_2 + Y_0) \ldots \pi_{rw}(S_{t-1} + Y_0)] = 0
$$

and

$$
P(Y > y) = \lim_{t \to \infty} P(Y_t > y) = \sum_{n=1}^{\infty} r_n(y) \bar{F}_{\mu}^{*n}(y).
$$

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

The stationary density can also be decomposed as

$$
\begin{aligned}
l(y_t) &= \sum_{n=1}^{\infty} l(y_t, B_t = 1, \ldots, B_{t-n+2} = 1, B_{t-n+1} = 0) \\
&= \sum_{n=1}^{\infty} \int \pi_{rw}(y_{t-1}) f_\mu(y_t - y_{t-1}) \ldots \pi_{rw}(y_{t-n+1}) f_\mu(y_{t-n+2} - y_{t-n+1}) \\
&\quad \times [1 - \pi_{rw}(y_{t-n})] f_\mu(y_{t-n+1}) l(y_{t-n})\, dy_{t-1} \ldots dy_{t-n} \\
&= [1 - \mathbb{E}\pi_{rw}(Y)] \sum_{n=1}^{\infty} \int \pi_{rw}(y_{t-1}) f_\mu(y_t - y_{t-1}) \ldots \pi_{rw}(y_{t-n+1}) f_\mu(y_{t-n+2} - y_{t-n+1}) \\
&\quad \times f_\mu(y_{t-n+1})\, dy_{t-1} \ldots dy_{t-n+1} \\
&=: \sum_{n=1}^{\infty} q_n(y_t) f_\mu^{*n}(y_t), \quad \text{(say)}.
\end{aligned}
$$

The weights q_n can be computed by recursion:

$$
\begin{aligned}
q_1(y) f_\mu(y) &= [1 - \mathbb{E}\pi_{rw}(Y)] f_\mu(y), \\
q_n(y) f_\mu^{*n}(y) &= \int \pi_{rw}(x) f_\mu(y - x) q_{n-1}(x) f_\mu^{*n-1}(x)\, dx,
\end{aligned}
$$

which finishes the proof. ■

Existence of Moments (Proof of Proposition 9). Note that the D.S. Markov chain is ν-irreducible and aperiodic. Let $g \geq 1$ be a function on $\mathbb{R}$. From Theorem 14.0.1 of Meyn and Tweedie (1993), $\int_{-\infty}^{\infty} g(y) P_Y(dy)$ is finite if there exist a petite set C, a constant $b < \infty$, and a nonnegative function V satisfying $V(y_0) < \infty$ for some $y_0 \in \mathbb{R}$, and

$$
\Delta V(y) = \mathbb{E}[V(Y_1) | Y_0 = y] - V(y) \leq -g(y) + b\mathbb{I}_C(y), \qquad y \in \mathbb{R}.
$$

Let us consider the power function $V(y) = |y|^k$, $2 < k \leq \alpha$. Then, we have for $y > 0$

$$
\begin{aligned}
\Delta V(y) &= \pi_{rw}(y) \int_{\mathbb{R}} |x + y|^k f_\mu(x)\, dx + (1 - \pi_{rw}(y)) \mathbb{E}|\varepsilon_1 + \mu|^k - y^k \\
&= \pi_{rw}(y) \int_{-y}^{\infty} (x + y)^k f_\mu(x)\, dx \\
&\quad + \pi_{rw}(y) \int_{-\infty}^{-y} (-(x + y))^k f_\mu(x)\, dx + (1 - \pi_{rw}(y)) \mathbb{E}|\varepsilon_1 + \mu|^k - y^k.
\end{aligned}
$$

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

Let $c > 1$. By derivating $[k]$ times the function

$$f_y(x) = y^k + kxy^{k-1} + \frac{k(k-1)}{2}x^2y^{k-2} + \cdots + \frac{k(k-1)\ldots(k-[k]+1)}{[k]!}x^{[k]}y^{k-[k]} + c|x|^k - (x+y)^k,$$

and by using $f_y^{(j)}(0) = 0$, $j \leq [k]$, it is easily seen that for any $x \geq -y$

$$(x+y)^k \leq y^k + kxy^{k-1} + \frac{k(k-1)}{2}x^2y^{k-2} + \cdots + \frac{k(k-1)\ldots(k-[k]+1)}{[k]!}x^{[k]}y^{k-[k]} + c|x|^k.$$

We deduce that

$$\int_{-y}^{\infty}(x+y)^k f_\mu(x)\,dx \leq y^k P(\varepsilon_1 + \mu > -y) + ky^{k-1}\int_{-y}^{\infty} x f_\mu(x)\,dx + \sum_{j=2}^{[k]} c_j y^{k-j}\mathbb{E}|\varepsilon_1+\mu|^j + c\mathbb{E}|\varepsilon_1+\mu|^k,$$

where $c_j = (\prod_{l=1}^{j-1}(k-l+1))/j!$, and

$$\int_{-y}^{\infty}(x+y)^k f_\mu(x)\,dx \leq y^k + ky^{k-1}\mu - ky^{k-1}\int_{-\infty}^{-y} x f_\mu(x)\,dx + \sum_{j=2}^{[k]} c_j y^{k-j}\mathbb{E}|\varepsilon_1+\mu|^j + c\mathbb{E}|\varepsilon_1+\mu|^k.$$

Because ε_1 admits moments of order $j \leq \alpha$ with $\alpha > 2$, then $\int_{-\infty}^{-y} x f_\mu(x)\,dx = o(y^{-1})$. Now let us remark that

$$\int_{-\infty}^{-y}(-(x+y))^k f_\mu(x)\,dx = \int_0^{\infty} u^k f_\mu(-u-y)\,du = o(1),$$

by the dominated convergence theorem. Then for large y, there exists a constant $d > 0$ such that

$$\Delta V(y) \leq [\pi_{rw}(y) - 1]y^k + k\mu y^{k-1} + dy^{k-2}.$$

Let us recall that $\lim_{y\to\infty}(1 - \pi_{rw}(y))y = c$. Then the previous inequality can be written for large y:

$$\Delta V(y) \leq -(c - k\mu - \eta)y^{k-1} + dy^{k-2},$$

with $c - k\mu > \eta > 0$. We conclude by the same arguments as in the proof of Proposition 6(i).

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

Conditional Distribution (Proof of Proposition 10). The expression of the conditional distribution is deduced from the decomposition that follows:

$$\begin{aligned}P(Y_t \in A|Y_0 = y_0) &= \sum_{n=1}^{t} P(Y_t \in A, B_t = 1, \ldots, B_{t-n+2} = 1, B_{t-n+1} = 0|Y_0 = y_0) \\ &\quad + P(Y_t \in A, B_t = 1, \ldots, B_2 = 1, B_1 = 1|Y_0 = y_0) \\ &= \sum_{n=1}^{t} [1 - \mathbb{E}(\pi_{rw}(Y_{t-n})|Y_0 = y_0)] \\ &\quad \times \mathbb{E}[\mathbb{I}_{\{S_n \in A\}} \pi_{rw}(S_1)\pi_{rw}(S_2)\ldots\pi_{rw}(S_{n-1})] \\ &\quad + \mathbb{E}[\mathbb{I}_{\{S_t + y_0 \in A\}} \pi_{rw}(y_0)\pi_{rw}(S_1 + y_0)\ldots\pi_{rw}(S_{t-1} + y_0)]. \end{aligned}$$

■

APPENDIX F: Proof of Proposition 11

Let $p_0 = 1$. From Proposition 10, we have

$$\begin{aligned}\mathbb{E}(\pi_{rw}(Y_t)|Y_0 = y_0) &= \sum_{n=1}^{t} [1 - \mathbb{E}(\pi_{rw}(Y_{t-n})|Y_0 = y_0)]p_n \\ &\quad + \mathbb{E}[\pi_{rw}(y_0)\pi_{rw}(S_1 + y_0)\ldots\pi_{rw}(S_t + y_0)].\end{aligned}$$

Then, we get

$$\begin{aligned}c_t &= \mathbb{E}[(\pi_{rw}(Y_0) - \mathbb{E}(\pi_{rw}(Y_0)))\mathbb{E}(\pi_{rw}(Y_t)|Y_0)] \\ &= -\sum_{n=1}^{t} c_{t-n}p_n + \operatorname{cov}\left(\pi_{rw}(Y_0); \pi_{rw}(Y_0)\prod_{i=1}^{t}\pi_{rw}(S_i + Y_0)\right)\end{aligned}$$

or equivalently

$$\sum_{n=0}^{t} c_{t-n}p_n = \operatorname{cov}\left(\pi_{rw}(Y_0); \pi_{rw}(Y_0)\prod_{i=1}^{t}\pi_{rw}(S_i + Y_0)\right).$$

Let us introduce the positive increasing function

$$g_t(y_0) = \frac{\pi_{rw}(y_0)\mathbb{E}\left(\prod_{i=1}^{t}\pi_{rw}(S_i + y_0)\right)}{[1 - \mathbb{E}\pi_{rw}(Y_0)]\sum_{n=t+1}^{\infty} p_n}.$$

From Proposition 8, we have

$$\mathbb{E}[\pi_{rw}(Y_0)\pi_{rw}(S_1 + Y_0)\ldots\pi_{rw}(S_t + Y_0)] = [1 - \mathbb{E}\pi_{rw}(Y_0)]\sum_{n=t+1}^{\infty} p_n,$$

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press

and then $\mathbb{E}(g_t(Y_0)) = 1$. Moreover, by using the same arguments as for Proposition 3(i) and (ii), we can show that, for any $y_0 \in \mathbb{R}$,

$$\lim_{t\to\infty} g_t(y_0) = 0.$$

Therefore, g_t can be considered a cumulative weighting function, which weights $+\infty$ for large t. We deduce

$$\lim_{t\to\infty} \mathbb{E}[\pi_{rw}(Y_0)g_t(Y_0)] = \lim_{u\to\infty} \pi_{rw}(u) = 1.$$

Therefore

$$\begin{aligned}
&\operatorname{cov}\left(\pi_{rw}(Y_0); \pi_{rw}(Y_0)\prod_{i=1}^{t}\pi_{rw}(S_i + Y_0)\right)\\
&\quad = \mathbb{E}[\pi_{rw}(Y_0)g_t(Y_0)][1 - \mathbb{E}\pi_{rw}(Y_0)]\sum_{n=t+1}^{\infty} p_n\\
&\qquad - \mathbb{E}\pi_{rw}(Y_0)[1 - \mathbb{E}\pi_{rw}(Y_0)]\sum_{n=t+1}^{\infty} p_n\\
&\quad = [1 - \mathbb{E}\pi_{rw}(Y_0)]^2 \sum_{n=t+1}^{\infty} p_n[1 + o(1)].
\end{aligned}$$

The sequence $(\sum_{n=0}^{t} c_{t-n}p_n)_{t\geq 0}$ is the convolution of the sequences $(p_t)_{t\geq 0}$ and $(c_t)_{t\geq 0}$. From Proposition 3, $(p_t)_{t\geq 0}$ and $(\sum_{n=t}^{\infty} p_n)_{t\geq 0}$ have a hyperbolic rate of decay and $p_t = o(\sum_{n=t}^{\infty} p_n)$. Consequently, the asymptotic behavior of the convolution is given by the asymptotic behavior of the sequence with the heavier "tail." We deduce that

$$c_t = [1 - \mathbb{E}\pi_{rw}(Y_0)]^2 \sum_{n=t+1}^{\infty} p_n[1 + o(1)].$$

■

https://doi.org/10.1017/S0266466606060518 Published online by Cambridge University Press