Econometric Theory, **24**, 2008, 1137–1148. Printed in the United States of America.
doi:10.1017/S0266466608080572

REGIME-SWITCHING AUTOREGRESSIVE COEFFICIENTS AND THE ASYMPTOTICS FOR UNIT ROOT TESTS

GIUSEPPE CAVALIERE
Università di Bologna

ILIYAN GEORGIEV
Universidade Nova de Lisboa

Most of the asymptotic results for Markov regime-switching models with possible unit roots are based on specifications implying that the number of regime switches grows to infinity as the sample size increases. Conversely, in this note we derive some new asymptotic results for the case of Markov regime switches that are infrequent in the sense that their number is bounded in probability, even asymptotically. This is achieved by (inversely) relating the probability of regime switching to the sample size. The proposed asymptotic theory is applied to a well-known stochastic unit root model, where the dynamics of the observed variable switches between a unit root regime and a stationary regime.

1. PROBLEM AND MOTIVATION

Since the seminal work of Perron (1989), the analysis of the finite-sample and asymptotic properties of unit root (UR) tests under regime switches has attracted considerable interest among econometricians. Although most of the authors initially focused on cases where the switches are fixed and nonrandom, the literature has been extended to cover also regime switches occurring according to a random process. A leading example is given by the so-called Markov regime switches introduced by Hamilton (1989); i.e., models where some of the parameters switch over time according to a (generally) latent Markov chain; see Nelson, Piger, and Zivot (2001), Rahbek and Shephard (2002), Gourieroux and Roberts (2006), and references therein.

Most of the asymptotic theories for Markov regime-switching models with possible unit roots are based on specifications implying that the number of regime switches grows to infinity as the sample size increases (see, e.g., Nelson et al., 2001; Psaradakis, 2001; Cavaliere, 2003). The results of these theories, how-

We are grateful to a referee for his comments on an earlier version of this paper. Address correspondence to Iliyan Georgiev, Faculdade de Economia, Universidade Nova de Lisboa, Campus de Campolide, PT 1099-032, Lisboa, Portugal; e-mail: igeorgiev@fe.unl.pt.

© 2008 Cambridge University Press 0266-4666/08 $15.00

https://doi.org/10.1017/S0266466608080572 Published online by Cambridge University Press

ever, may provide poor approximations in finite samples, given the empirical evidence that only a small number of regime switches occurs in economic time series: cf. Balke and Fomby (1991a, 1991b, 1994) and Diebold and Inoue (2001).

The aim of this note is to derive some new asymptotic results for UR tests in the presence of Markov regime switches that are "infrequent" in the sense that their number is bounded in probability as the sample size diverges. This is achieved by (inversely) relating the probability of regime switching to the sample size, as advocated by, among others, Diebold and Inoue (2001). By doing so, we propose an asymptotic theory that seems to be useful in approximating the empirically relevant cases.

To illustrate the proposed asymptotic theory, we provide an application to a well-known stochastic UR model from the class discussed by Gourieroux and Roberts (2006). In this model, the dynamics of the observed variable switches between a "unit root" regime and a "stationary" regime. The model has been found of interest in applications; see, among others, Ang and Bekaert (2002). We will show that, if the regime switches are infrequent, then the so-called coefficient Dickey–Fuller (DF) test of the UR null hypothesis is not consistent against the alternative of Markov switching, in contrast with the theoretical results reported, e.g., by Ang and Bekaert (2002) for the case of frequent switches. Not surprisingly, the test will reject with increasing limiting probability as the autoregressive (AR) root in the stationary regime gets further away from unity.

2. ASYMPTOTICS FOR MARKOV CHAINS WITH INFREQUENT SWITCHES

We start by considering a sequence (indexed by T) of discrete-time, two-state Markov chains $\{S_{T,t}\}_{t=0}^{T}$ where, for each t, $S_{T,t} \in \{0,1\}$, with initial state $S_{T,0} = 0$. The two-state assumption is made for expositional purposes; the generalization to an arbitrary finite state space is straightforward. We denote the one-step transition probability matrix by $Q_T := \{p_{T,ij}\}$, $i,j \in \{0,1\}$, where, instead of taking Q_T constant, we specify it as follows:

$$p_{T,00} := P\{S_{T,t+1} = 0 \,|\, S_{T,t} = 0\} = 1 - \alpha T^{-1}, \tag{1}$$

$$p_{T,11} := P\{S_{T,t+1} = 1 \,|\, S_{T,t} = 1\} = 1 - \beta T^{-1} \tag{2}$$

with $t = 0, \ldots, T-1$ and $T > \alpha, \beta > 0$. This specification is a special case of Assumption (b) in Diebold and Inoue (2001), where the T^{-1} term is replaced with the more general $T^{-\delta_i}$, $\delta_i > 0$. The logic behind (1) and (2) is that, by properly letting the probabilities of a regime switch (i.e., $1 - p_{T,ii}$) decline as T increases, we are able to bound (in probability) the number of switches.

The next lemma, which will be used in the subsequent analysis, states that the Markov chain $\{S_{T,t}\}$, after rescaling the time axis, converges weakly to a continuous-time Markov process in the càdlàg space $D[0,1]$ endowed with the Skorohod topology.[1] We denote by $\lfloor \cdot \rfloor$ the integer-part function.

https://doi.org/10.1017/S0266466608080572 Published online by Cambridge University Press

LEMMA 1. *Let* $\{S(u): u \in [0,1]\}$ *be the continuous-time Markov process with state space* $\{0,1\}$*, initial state* $S(0) = 0$ *a.s., and distribution determined by the equations*

$$P(S(u+h) = 1 | S(u) = 0) = \frac{\alpha}{\alpha+\beta}(1 - e^{-(\alpha+\beta)h})$$
$$P(S(u+h) = 1 | S(u) = 1) = \frac{1}{\alpha+\beta}(\alpha + \beta e^{-(\alpha+\beta)h}) \qquad u, u+h \in [0,1], \quad h > 0.$$

Then, as $T \to \infty$*, it holds that* $S_{T,\lfloor T\cdot \rfloor} \xrightarrow{w} S(\cdot)$ *in* $D[0,1]$.

Remarks.

(a) The process $S(\cdot)$ has a finite expected number of switches (see, e.g., Råde, 1976), and hence its number of switches is almost surely (a.s.) finite. As a consequence, the smallest sojourn time of $S(\cdot)$ is a.s. positive. Using this fact and the expressions for the transition probabilities from Lemma 1, it is readily seen that $S(\cdot)$ is a.s. continuous at one. Further, the enumerated properties of $S(\cdot)$, Lemma 1, and the continuous mapping theorem (CMT) imply that the number of switches and the smallest sojourn time of $S_{T,\lfloor T\cdot \rfloor}$ (as a process on $[0,1]$) converge weakly to those of $S(\cdot)$ as $T \to \infty$. Thus, $S_{T,t}$ (as a process on $0, \ldots, T$) has a stochastically bounded number of switches, and its smallest sojourn time diverges in probability to ∞ at the rate of T.

(b) To emphasize the difference between the case considered here and the well-known case of constant transition probabilities (where Q_T is fixed and independent of T), recall that in the latter case, under rather general conditions[2] objects such as $T^{-1}\sum_{t=1}^{T} S_{T,t}$ converge in probability to a constant as T diverges. Conversely, in the case considered here, it follows from the CMT that $T^{-1}\sum_{t=1}^{T} S_{T,t}$ converges weakly to the random variable $\int_0^1 S(u)\,du$, the total time that $S(\cdot)$ spends in state one.

3. APPLICATION TO A STOCHASTIC UR PROCESS

Following Ang and Bekaert (2002) and Nelson et al. (2001), we consider the two-regime AR process:

$$Y_t = \begin{cases} \mu_0 + Y_{t-1} + \varepsilon_t & \text{if } S_{T,t} = 0 \\ \mu_1 + \theta Y_{t-1} + \varepsilon_t & \text{if } S_{T,t} = 1 \end{cases} \qquad t = 1, \ldots, T, \tag{3}$$

where $|\theta| < 1$, $Y_0 = O_p(1)$, and $\{\varepsilon_t\}$ is $\mathrm{IID}(0, \sigma^2)$ and is independent of $\{S_{T,t}\}$.[3] Equation (3) defines a "stochastic unit root" process in the sense that the AR root switches between one (in the UR regime) and a stable value (in the "stationary" regime); cf. Rahbek and Shephard (2002) and Gourieroux and Roberts (2006).

https://doi.org/10.1017/S0266466608080572 Published online by Cambridge University Press

If the latent Markov chain $S_{T,t}$ had constant transition probabilities, a sufficient condition for (asymptotic) covariance stationarity of Y_t would be that the ergodic probability associated with the stationary regime is nonzero; see Ang and Bekaert (2002). It is not difficult to prove that in this case UR tests such as the well-known DF tests would reject the UR null hypothesis with probability approaching one as T diverges.

However, in the case where the regime switches are infrequent, this result no longer holds. To see this fact, consider the simple (constant-corrected) DF coefficient test, i.e., a test that rejects for small values of the statistic

$$Z_\theta := T \frac{\sum_{t=1}^{T} (Y_{t-1} - \bar{Y}_{-1}) \Delta Y_t}{\sum_{t=1}^{T} (Y_{t-1} - \bar{Y}_{-1})^2}, \tag{4}$$

where $\bar{Y}_{-1} := T^{-1} \sum_{t=1}^{T} Y_{t-1}$. The asymptotic distribution of Z_θ under the data generating process (3) is obtained in Theorem 1. Notice that we restrict μ_0 to be equal to zero to rule out linear time trends in the UR regime. Also, we write stochastic integrals $\int_0^1 X(u)\, dY(u)$ and pathwise integrals $\int_0^1 X(u)\, du$ as $\int X dY$ and $\int X$, respectively.

THEOREM 1. *For the limiting process S from Lemma 1, let* $\tau(u)$, $u \in [0,1]$, *denote the time of the last regime switch up to time u, or* 0 *if no such switches have occurred; i.e.,* $\tau(0) := 0$ *and*

$$\tau(u) := \sup\{s : s \in (0,u],\, S(s) - S(s-) \neq 0\}, \qquad u \in (0,1],$$

with $\sup \varnothing := 0$. *Let also R be the process defined by*

$$R(u) := (1 - S(u))(B(u) - B(\tau(u))), \qquad u \in [0,1], \tag{5}$$

where B is a standard Brownian motion on $[0,1]$, *independent of S. If the data* $\{Y_t\}_{t=1}^{T}$ *are drawn from (3) with* $\mu_0 = 0$, *the following convergence holds as* $T \to \infty$:

$$Z_\theta \xrightarrow{w} \frac{\int R dB - R(1)\bar{R}}{\int (R - \bar{R})^2} - \frac{1}{1+\theta} \frac{\sum_{u \in (0,1]} \{\Delta R(u)\}^2 + \bar{S}}{\int (R - \bar{R})^2}, \tag{6}$$

where $\bar{R} := \int R$, $\bar{S} := \int S$, *and* $\Delta R(u) := R(u) - R(u-)$, $u \in (0,1]$.

Remarks.

(a) A key implication of the theorem is that infrequent regime switches invalidate the consistency of the DF test against alternatives such as the

https://doi.org/10.1017/S0266466608080572 Published online by Cambridge University Press

stochastic UR model (3), in contrast to the case of frequent regime switches (whose number diverges to infinity in probability at the rate of T). In particular, the limiting rejection probability, instead of being equal to one, depends on the model parameters; see the discussion that follows.

(b) Differently from the usual UR case, the asymptotic distribution of the DF statistic does not depend on a standard Brownian motion only. Instead, it depends on the càdlàg process R (see eqn. (5)), which switches between independent Brownian motions starting at zero (in regimes with $S = 0$) and the zero process (in regimes with $S = 1$).[4] Hence, the trajectories of R are discontinuous at switches from $S = 0$ to $S = 1$ but not at switches from $S = 1$ to $S = 0$. When no regime switches occur, the DF asymptotic distribution obtains; specifically, conditionally on no switches, we have $R = B$ (by the independence of B and S) and $\bar{S} = 0$, which implies that the limiting expression in (6) is the well-known DF functional $\int(B - \bar{B})\, dB / \int(B - \bar{B})^2$.

(c) The second term on the right-hand side of (6) is a contribution from the stationary regime. It is seen that, as θ gets further away from unity, this term shifts the asymptotic distribution of Z_θ to the left, hence increasing (other things being equal) the limiting rejection probability of the UR test. This follows from the nonnegativity of both $\sum_{u\in(0,1]}\{\Delta R(u)\}^2$ (the sum of squared discontinuities of R) and $\bar{S}$ (the time spent in the stationary regime).

(d) Because the magnitude order of Y_T depends only on the last realized regime, demeaning has an effect on the numerator of (6) only through that regime. To see why, note that the numerator of the DF statistic may be expressed, after normalizing by T, as $T^{-1}\sum_{t=1}^{T} Y_{t-1}\Delta Y_t - (T^{-1/2}\overline{Y}_{-1})(T^{-1/2}Y_T) + o_p(1)$, where $T^{-1/2}\overline{Y}_{-1} \xrightarrow{w} \sigma\bar{R}$ (see the proof of Theorem 1 in the Appendix), and the $o_p(1)$ term contains a contribution of Y_0. At the end of a nonstationary regime the value of Y is nonnegligible relative to $T^{1/2}$, whereas at the end of the following stationary regime that value becomes negligible. Thus, if the last regime is stationary, $T^{-1/2}Y_T$ is negligible and demeaning has a negligible effect on the numerator, whereas in the case of a nonstationary last regime, $T^{-1/2}Y_T$ is essentially determined by the cumulated innovations of that regime.

(e) Alternative representations of the limiting distribution in (6) can be obtained using the identities $\int RdB = \int RdR + \sum_{u\in(0,1]}\{\Delta R(u)\}^2$ and $\int RdR = \frac{1}{2}\{R^2(1) - \sum_{u\in(0,1]}\{\Delta R(u)\}^2 - (1 - \bar{S})\}$. The stochastic integral $\int RdR$ is well defined, for R is a semimartingale with finite-variation component $M(\cdot) := \sum_{u\in(0,\cdot]}\Delta R(u)$ and martingale component $R - M$.

(f) The denominator in (6) is a.s. nonzero by the assumption that the initial regime is nonstationary and because S has a.s. positive sojourn time in that regime. Without the assumption that S is initialized at zero, the theorem would hold conditionally on at least one regime switch.

https://doi.org/10.1017/S0266466608080572 Published online by Cambridge University Press

NOTES

1. All proofs are collected in the Appendix.
2. See, e.g., Lin (1997).
3. Although Y_t is an array of the form $Y_{T,t}$, $t = 1, \ldots, T$, we drop the T subscript.
4. Notice that, although the limiting distribution is expressed in terms of R, the process $T^{-1/2}Y_{\lfloor T\cdot \rfloor}$ need not converge weakly to σR.

REFERENCES

Ang, A. & G. Bekaert (2002) Regime switches in interest rates. *Journal of Business & Economics Statistics* 20, 163–182.

Balke, N.S. & T.B. Fomby (1991a) Shifting trends, segmented trends, and infrequent permanent shocks. *Journal of Monetary Economics* 28, 61–85.

Balke, N.S. & T.B. Fomby (1991b) Infrequent permanent shocks and the finite-sample performance of unit root tests. *Economics Letters* 36, 269–273.

Balke, N.S. & T.B. Fomby (1994) Large shocks, small shocks and economic fluctuations: Outliers in macroeconomic time series. *Journal of Applied Econometrics* 9, 181–200.

Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley.

Cavaliere, G. (2003) Asymptotics for unit root tests under Markov regime-switching. *Econometrics Journal* 6, 193–216.

Diebold, F.X. & A. Inoue (2001) Long memory and regime switching. *Journal of Econometrics* 105, 131–159.

Gourieroux, C. & C.Y. Roberts (2006) Stochastic unit root models. *Econometric Theory* 22, 1052–1090.

Hamilton, J.D. (1989) A new approach to the economic analysis of non-stationary time series and the business cycle. *Econometrica* 57, 357–384.

Kurtz, T. & P. Protter (1991) Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability* 19, 1035–1070.

Lin, Z. (1997) The invariance principle for some class of Markov chains. *Theory of Probability and Its Applications* 44, 136–140.

Nelson, C.R., J. Piger, & E. Zivot (2001) Markov regime-switching and unit root tests. *Journal of Business & Economic Statistics* 19, 404–415.

Perron, P. (1989) The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* 57, 1361–1401.

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.

Psaradakis, Z. (2001) Markov level shifts and the unit root hypothesis. *Econometrics Journal* 4, 226–242.

Råde, L. (1976) The two-state Markov process and additional events. *American Mathematical Monthly* 83, 354–356.

Rahbek, A. & N. Shephard (2002) Inference and Ergodicity in the Autoregressive Conditional Root Model. Economics Working paper 2002-W7, Nuffield College.

APPENDIX

Proof of Lemma 1. As is standard (see Billingsley, 1968), we prove weak convergence in $D[0,1]$ by establishing finite-dimensional convergence and tightness.

From expressions (1) and (2) for the transition probabilities, it follows that

$$(\alpha+\beta)Q_T^{\lfloor T(u+h)\rfloor-\lfloor Tu\rfloor} = \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + E_T(u,h)\begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix},$$

https://doi.org/10.1017/S0266466608080572 Published online by Cambridge University Press

where $E_T(u,h) := (1-(\alpha+\beta)/T)^{\lfloor T(u+h)\rfloor-\lfloor Tu\rfloor} \to e^{-(\alpha+\beta)h}$ as $T\to\infty$. The matrix convergence can be written component-wise as

$$P(S_{T,\lfloor T(u+h)\rfloor}=j\,|\,S_{T,\lfloor Tu\rfloor}=i)\to P(S(u+h)=j\,|\,S(u)=i),\qquad i,j\in\{0,1\},\tag{A.1}$$

where the probabilities on the right-hand side are as given in Lemma 1. Furthermore, in view of the Markov property, for any $0\le u_1<\cdots<u_k\le 1$ and any $b=(b_1,\dots,b_k)\in\{0,1\}^k$ it holds that

$$P((S_{T,\lfloor Tu_1\rfloor},\dots,S_{T,\lfloor Tu_k\rfloor})=b)=\prod_{i=1}^{k}P(S_{T,\lfloor Tu_i\rfloor}=b_i\,|\,S_{T,\lfloor Tu_{i-1}\rfloor}=b_{i-1}),$$

with $u_0:=0=:b_0$, and a similar equality holds for the limiting process, which in conjunction with (A.1) implies finite-dimensional convergence.

Tightness follows from Theorem 15.6 of Billingsley (1968), as the process $S(\cdot)$ is a.s. continuous at one. Indeed, let $K:=\alpha\beta/(\alpha+\beta)^2$, and for $u_1<u<u_2$, let $\mathcal{E}_T(u_1,u,u_2):=\mathrm{E}\{|S_{T,\lfloor Tu\rfloor}-S_{T,\lfloor Tu_1\rfloor}|\,|S_{T,\lfloor Tu_2\rfloor}-S_{T,\lfloor Tu\rfloor}|\}$. Then

$$\begin{aligned}\mathcal{E}_T(u_1,u,u_2)&=P((S_{T,\lfloor Tu_1\rfloor},S_{T,\lfloor Tu\rfloor},S_{T,\lfloor Tu_2\rfloor})\in\{(0,1,0),(1,0,1)\})\\&=K\{1-E_T(u_1,u-u_1)\}\{1-E_T(u,u_2-u)\}\\&<K\{1-e^{-2(\alpha+\beta)(u-u_1)}\}\{1-e^{-2(\alpha+\beta)(u_2-u)}\}\\&\le K\{1-e^{-(\alpha+\beta)(u_2-u_1)}\}^2=Ke^{-2(\alpha+\beta)u_2}\{e^{(\alpha+\beta)u_2}-e^{(\alpha+\beta)u_1}\}^2\\&<\{F(u_2)-F(u_1)\}^2,\end{aligned}$$

the first inequality holding for large enough T. Here $F(\cdot):=K^{1/2}e^{(\alpha+\beta)(\cdot)}$, which satisfies the requirements of Theorem 15.6 in Billingsley (1968). This completes the proof. ■

For reference in the proof of Theorem 1, we formulate a corollary of Lemma 1. Let $D^*[0,1]$ be the subspace of $D[0,1]$ consisting of càdlàg functions with codomain $\{0,1\}$, with finitely many discontinuities, and that are continuous at one. For $s\in D^*[0,1]$, let $\tau^s(0):=0$ and

$$\tau^s(u):=\sup\{v:v\in(0,u],\,s(v)-s(v-)\neq 0\},\qquad u\in(0,1],\tag{A.2}$$

with $\sup\varnothing:=0$. Finally, let τ_T^S be the process associated pathwise to $S_{T,\lfloor T\cdot\rfloor}$ via (A.2).

COROLLARY 1. *For the process R_T defined by*

$$R_T(u):=(1-S_{T,\lfloor Tu\rfloor})\left(T^{-1/2}\sum_{t=1}^{\lfloor Tu\rfloor}\varepsilon_t-T^{-1/2}\sum_{t=1}^{\lfloor T\tau_T^S(u)\rfloor}\varepsilon_t\right),\qquad u\in[0,1],$$

and the process R defined in Theorem 1, it holds that $(S_{T,\lfloor T\cdot\rfloor},T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\varepsilon_t,R_T)\xrightarrow{w}(S,\sigma B,\sigma R)$ in $D_3[0,1]$, the space of càdlàg functions with values in $\mathbb{R}^3$, endowed with the Skorohod topology.

Proof. On the one hand, Donsker's theorem, Lemma 1, and the independence of $S_{T,\lfloor T\cdot\rfloor}$ and $T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\varepsilon_t$ imply that $(S_{T,\lfloor T\cdot\rfloor},T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor}\varepsilon_t)\xrightarrow{w}(S,\sigma B)$ in $D^{\times 2}[0,1]$ ($D^{\times m}[0,1]$

https://doi.org/10.1017/S0266466608080572 Published online by Cambridge University Press

denotes the m-fold product space of $D[0,1]$). On the other hand, the following map from $D^{\times 2}[0,1]$ to $D^{\times 3}[0,1]$ is continuous on $D^*[0,1] \times C[0,1]$:

$$(s,w) \mapsto (s,w,r), \quad \text{with } r(u) = (1 - s(u))(w(u) - w(\tau^s(u))), \qquad u \in [0,1].$$

This is not hard to check given the special form of $s \in D^*[0,1]$ and the uniform continuity of $w \in C[0,1]$. As $D^*[0,1] \times C[0,1]$ supports $(S, \sigma B)$, it follows from the CMT that $(S_{T,\lfloor T\cdot \rfloor}, T^{-1/2} \sum_{t=1}^{\lfloor T\cdot \rfloor} \varepsilon_t, R_T) \xrightarrow{w} (S, \sigma B, \sigma R)$ in $D^{\times 3}[0,1]$. The convergence extends to $D_3[0,1]$ because, first, B is continuous, and second, the jump dates of R are a subset of those of S. ■

Finally, before we present a proof of Theorem 1, we justify two magnitude orders. Let N_T denote the total number of regime switches of $\{S_{T,t}\}$. Let also τ_i, $i = 1, \ldots, N_T$, denote the times of the regime switches in increasing order, with $\tau_0 := 0$, $\tau_{N_T+1} := T + 1$, so that over the subsamples $\{\tau_i, \ldots, \tau_{i+1} - 1\}$, $i = 0, \ldots, N_T$, the regime is constant. Then

$$\max_{t \le T} |Y_t| = O_P(T^{1/2}) \quad \text{and} \quad \max_{0 \le i \le N_T} |(1 - S_{\tau_i}) Y_{\tau_i}| = o_P(T^{1/2}). \tag{A.3}$$

To see this, equation (3) with $\mu_0 = 0$ can be written as

$$Y_t = \nu_t + \theta_t Y_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, T,$$

where $\theta_t := \theta S_{T,t} + 1 - S_{T,t}$, $\nu_t := \mu(1-\theta) S_{T,t}$ and $\mu := \mu_1/(1-\theta)$. Thus, for $\tau_i \le t < \tau_{i+1}$, $i = 0, \ldots, N_T$, it holds that

$$Y_t = \theta_t^{t-\tau_i} Y_{\tau_i} + (1 - \theta_t^{t-\tau_i})\mu + \sum_{j=0}^{t-\tau_i - 1} \theta_t^j \varepsilon_{t-j}.$$

From here it can be concluded by induction that, as Y_0 is assumed $O_P(1)$, $\max_{\tau_i \le t < \tau_{i+1}} |Y_t| = O_P(T^{1/2})$ for $i = 0, \ldots, N_T$. As $N_T = O_P(1)$, also $\max_{t \le T} |Y_t| = O_P(T^{1/2})$.

To justify the second relation in (A.3), for $\tau_i \le t < \tau_{i+1}$ we use the evaluation

$$\max_{t \le T} \left| \sum_{j=0}^{t-\tau_i - 1} \theta_t^j \varepsilon_{t-j} - (1 - S_t) \sum_{j=0}^{t-\tau_i - 1} \varepsilon_{t-j} \right| = \max_{t \le T} \left| S_t \sum_{j=0}^{t-\tau_i - 1} \theta^i \varepsilon_{t-j} \right|$$

$$\le (1 - |\theta|)^{-1} \max_{t \le T} |\varepsilon_t| = o_P(T^{1/2}), \tag{A.4}$$

with the last equality following from ε_t being $\mathrm{IID}(0, \sigma^2)$ with $\sigma^2 < \infty$; see, e.g., Phillips and Solo (1992, p. 978). Thus, at the end of every stationary regime (i.e., for $i > 0$ such that $S_{\tau_i - 1} = 1$),

$$|Y_{\tau_i - 1}| = \left| \theta^{\tau_i - \tau_{i-1} - 1} (Y_{\tau_{i-1}} - \mu) + \mu + \sum_{j=0}^{\tau_i - \tau_{i-1} - 2} \theta^j \varepsilon_{\tau_i - 1 - j} \right|$$

$$\le |\theta|^{\delta_T - 1} \left(\max_{t \le T} |Y_t| + |\mu| \right) + |\mu| + (1 - |\theta|)^{-1} \max_{t \le T} |\varepsilon_t|, \tag{A.5}$$

https://doi.org/10.1017/S0266466608080572 Published online by Cambridge University Press

where $\delta_T := \min_{0\le i\le N_T}(\tau_{i+1} - \tau_i)$ is the minimal sojourn time of $\{S_{T,t}\}$. Because $|\theta|^{\delta_T} \xrightarrow{P} 0$ exponentially fast ($\delta_T \xrightarrow{P} \infty$ at the rate of T), it follows from (A.4) and (A.5) that $\max_{1\le i\le N_T}|(1 - S_{\tau_i})Y_{\tau_i-1}| = o_P(T^{1/2})$. Finally, because $(1 - S_{\tau_i})Y_{\tau_i} = (1 - S_{\tau_i})(Y_{\tau_i-1} + \varepsilon_{\tau_i})$, $\max_{t\le T}|\varepsilon_t| = o_P(T^{1/2})$, and $Y_0 = O_P(1)$, the second relation in (A.3) obtains.

Proof of Theorem 1. We find separately the weak limits of $T^{-2}\sum_{t=1}^T(Y_{t-1} - \overline{Y}_{-1})^2$ and $T^{-1}\sum_{t=1}^T(Y_{t-1} - \overline{Y}_{-1})\Delta Y_t$. The following evaluation helps in the derivation of both limits. Omitting for brevity the T subscript, let $\{a_t\}$ be any random sequence or array. Introduce $\upsilon(t) := \max_{0\le i\le N_T}\{\tau_i : \tau_i \le t\}$, $1 \le t \le T$, so that the process τ_T^S associated to $S_{\lfloor T\cdot\rfloor}$ via (A.2) satisfies $\tau_T^S(u) = \upsilon(\lfloor Tu\rfloor)/T$. Then, because $N_T = O_P(1)$, we have

$$\left|\sum_{t=1}^T S_t\theta^{t-\upsilon(t)}a_t\right| \le \max_{t\le T}|S_t a_t|\sum_{t=1}^T S_t|\theta|^{t-\upsilon(t)} < \left(\max_{t\le T}|S_t a_t|\right)N_T/(1 - |\theta|)$$

$$= O_P\left(\max_{t\le T}|S_t a_t|\right). \tag{A.6}$$

We decompose next Y_t into $Y_t = \langle Y_t\rangle + \lfloor Y_t\rceil$, where

$$\langle Y_t\rangle := S_t\mu + \sum_{j=0}^{t-\upsilon(t)-1}\theta_t^j\varepsilon_{t-j} \quad\text{and}\quad \lfloor Y_t\rceil := \theta_t^{t-\upsilon(t)}(Y_{\upsilon(t)} - S_t\mu), \tag{A.7}$$

so that $Y_t = \langle Y_t\rangle + \lfloor Y_t\rceil$. Note that

$$\sum_{t=1}^T|\lfloor Y_t\rceil| \le \left(\max_{t\le T}|Y_t| + |\mu|\right)\sum_{t=1}^T S_t|\theta|^{t-\upsilon(t)} + \sum_{t=1}^T(1 - S_t)|Y_{\upsilon(t)}| = o_P(T^{3/2}) \tag{A.8}$$

by (A.3) and (A.6). Thus, $T^{-2}\sum_{t=1}^T Y_{t-1}^2 = T^{-2}\sum_{t=1}^T\langle Y_{t-1}\rangle^2 + o_P(1)$, because

$$\left|\sum_{t=1}^T Y_{t-1}^2 - \sum_{t=1}^T\langle Y_{t-1}\rangle^2\right| = \left|\sum_{t=0}^{T-1}(2Y_t - \lfloor Y_t\rceil)\lfloor Y_t\rceil\right|$$

$$\le \left(3\max_{t\le T}|Y_t| + |\mu|\right)\sum_{t=1}^T|\lfloor Y_t\rceil| = o_P(T^2).$$

As $\tau_T^S(u) = \upsilon(\lfloor Tu\rfloor)/T$, from (A.4) and (A.7) it follows that

$$\max_{t\le T}|\langle Y_t\rangle - R_T(t/T)| = \max_{t\le T}\left|\langle Y_t\rangle - (1 - S_t)\sum_{j=0}^{t-\upsilon(t)-1}\varepsilon_{t-j}\right|$$

$$\le |\mu| + o_P(T^{1/2}) = o_P(T^{1/2}). \tag{A.9}$$

Together with Corollary 1, this yields $T^{-1/2}\langle Y_{T,\lfloor T\cdot\rfloor}\rangle \xrightarrow{w} \sigma R$, and by the CMT, $T^{-2}\sum_{t=1}^T\langle Y_{t-1}\rangle^2 \xrightarrow{w} \sigma^2\int R^2$, which is also the weak limit of $T^{-2}\sum_{t=1}^T Y_{t-1}^2$ in view of the argument in the preceding paragraph. Similarly, from (A.8) and the CMT it follows that $T^{-1/2}\overline{Y}_{-1} = T^{-3/2}\sum_{t=1}^T\langle Y_{t-1}\rangle + o_P(1) \xrightarrow{w} \sigma\bar{R}$. Finally, from the last two conclusions, the convergence $T^{-2}\sum_{t=1}^T(Y_{t-1} - \overline{Y}_{-1})^2 \xrightarrow{w} \sigma^2\int(R - \bar{R})^2$ obtains.

https://doi.org/10.1017/S0266466608080572 Published online by Cambridge University Press

Consider now $T^{-1}\sum_{t=1}^{T} Y_{t-1}\Delta Y_t$. As $\Delta Y_t = \varepsilon_t + (\theta-1)S_t(Y_{t-1}-\mu)$, it holds that

$$\sum_{t=1}^{T} Y_{t-1}\Delta Y_t = \sum_{t=1}^{T} Y_{t-1}\varepsilon_t + (\theta-1)\sum_{t=1}^{T} S_t(Y_{t-1}-\mu)^2 + (\theta-1)\mu\sum_{t=1}^{T} S_t(Y_{t-1}-\mu).$$

Here $S_t(Y_{t-1}-\mu) = S_t\{\theta^{t-v(t)}(Y_{v(t)-1}-\mu) + \sum_{j=0}^{t-1-v(t)}\theta^j\varepsilon_{t-1-j}\}$, so that $\sum_{t=1}^{T} S_t(Y_{t-1}-\mu) = b_1 + b_2 = o_P(T)$, with (a) $b_1 := \sum_{t=1}^{T} S_t\theta^{t-v(t)}(Y_{v(t)-1}-\mu) = O_P(T^{1/2})$ by (A.3) and (A.6), and (b) $b_2 := \sum_{t=1}^{T} S_t\sum_{j=0}^{t-1-v(t)}\theta^j\varepsilon_{t-1-j} = o_P(T)$ because $\mathrm{E}(b_2)=0$ and $\mathrm{E}(b_2^2) = O(T)$. Therefore,

$$T^{-1}\sum_{t=1}^{T} Y_{t-1}\Delta Y_t = T^{-1}\sum_{t=1}^{T} Y_{t-1}\varepsilon_t + (\theta-1)T^{-1}\sum_{t=1}^{T} S_t(Y_{t-1}-\mu)^2 + o_P(1). \tag{A.10}$$

On the one hand, $\sum_{t=1}^{T}\langle Y_{t-1}\rangle\varepsilon_t = \sum_{t=1}^{T} R_T((t-1)/T)\varepsilon_t + b_3$, where $b_3 = o_P(T)$ similarly to b_2. On the other hand, $\sum_{t=1}^{T}\lfloor Y_{t-1}\rceil\varepsilon_t = b_4 + b_5$, where $b_4 := \sum_{t=1}^{T} S_{t-1}\lfloor Y_{t-1}\rceil\varepsilon_t$ satisfies (see (A.7))

$$|b_4| = \left|\sum_{t=0}^{T-1} S_t\theta^{t-v(t)}(Y_{v(t)}-\mu)\varepsilon_{t+1}\right| \le \left(\max_{t\le T}|Y_t| + |\mu|\right)\left(\max_{t\le T}|\varepsilon_t|\right)O_P(1) = o_P(T)$$

by (A.3) and (A.6), and $b_5 := \sum_{t=1}^{T}(1-S_{t-1})\lfloor Y_{t-1}\rceil\varepsilon_t$ satisfies

$$|b_5| = \left|\sum_{i=1}^{N_T}(1-S_{\tau_i})Y_{\tau_i}\sum_{t=\tau_i+1}^{\tau_{i+1}}\varepsilon_t\right| \le 2\max_{s\le T}\left|\sum_{t=1}^{s}\varepsilon_t\right|\sum_{i=1}^{N_T}|(1-S_{\tau_i})Y_{\tau_i}| = o_P(T)$$

by (A.3), and because $N_T = O_P(1)$ and $\max_{s\le T}|T^{-1/2}\sum_{t=1}^{s}\varepsilon_t|$ converges weakly. Thus,

$$T^{-1}\sum_{t=1}^{T} Y_{t-1}\varepsilon_t = T^{-1}\sum_{t=1}^{T} R_T\left(\frac{t-1}{T}\right)\varepsilon_t + o_P(1) \xrightarrow{w} \sigma^2\int RdB \tag{A.11}$$

by Theorem 2.2 of Kurtz and Protter (1991) and Corollary 1.

The last summation in (A.10) can be written as

$$\sum_{t=1}^{T} S_t(Y_{t-1}-\mu)^2 = \sum_{t=1}^{T} S_t\theta^{2(t-v(t))}(Y_{v(t)-1}-\mu)^2 + \sum_{t=1}^{T} S_t\left(\sum_{j=0}^{t-1-v(t)}\theta^j\varepsilon_{t-1-j}\right)^2 + 2c_1, \tag{A.12}$$

where $c_1 := \sum_{t=1}^{T}(S_t\theta^{t-v(t)}(Y_{v(t)-1}-\mu)\sum_{j=0}^{t-1-v(t)}\theta^j\varepsilon_{t-1-j}) = o_P(T)$ by (A.6) and the relation

$$\max_{t\le T}\left|(Y_{v(t)-1}-\mu)\sum_{j=0}^{t-1-v(t)}\theta^j\varepsilon_{t-1-j}\right| \le \left(\max_{t\le T}|Y_t| + |\mu|\right)\left(\max_{t\le T}|\varepsilon_t|\right)(1-|\theta|)^{-1}$$

$$= o_P(T),$$

https://doi.org/10.1017/S0266466608080572 Published online by Cambridge University Press

which follows from (A.3) and the relation $\max_{t\le T}|\varepsilon_t| = o_P(T^{1/2})$. Further,

$$\sum_{t=1}^{T} S_t\theta^{2(t-v(t))}(Y_{v(t)-1}-\mu)^2 = (1-\theta^2)^{-1}\sum_{i=1}^{N_T} S_{\tau_i}Y_{\tau_i-1}^2 - (1-\theta^2)^{-1}c_2 + \mu c_3, \qquad \textbf{(A.13)}$$

where $c_2 := \sum_{i=1}^{N_T} S_{\tau_i}\theta^{2(\tau_{i+1}-\tau_i)}Y_{\tau_i-1}^2 \le \theta^{2\delta_T}N_T(\max_{t\le T}|Y_t|)^2 = o_P(1)$ because $\theta^{2\delta_T}\xrightarrow{P}0$ exponentially (recall that δ_T is the smallest sojourn time of $\{S_{T,t}\}$) and $c_3 := \sum_{t=1}^{T} S_t\theta^{2(t-v(t))}(\mu - 2Y_{v(t)-1}) = O_P(T^{1/2})$ by (A.3) and (A.6). Next,

$$\sum_{i=1}^{N_T} S_{\tau_i}|Y_{\tau_i-1}^2 - \langle Y_{\tau_i-1}\rangle^2| \le \left(\max_{t\le T}|Y_t| + \max_{t\le T}|\langle Y_t\rangle|\right)\sum_{i=1}^{N_T} S_{\tau_i}|\lfloor Y_{\tau_i-1}\rfloor| = o_P(T) \qquad \textbf{(A.14)}$$

by (A.3), by the convergence $\max_{t\le T}|T^{-1/2}\langle Y_t\rangle| \xrightarrow{w} \max_{u\in[0,1]}|R(u)|$, and because, using (A.7),

$$\sum_{i=1}^{N_T} S_{\tau_i}|\lfloor Y_{\tau_i-1}\rfloor| = \sum_{i=1}^{N_T}(1-S_{\tau_i-1})|Y_{v(\tau_i-1)}| \le N_T \max_{1\le i\le N_T}\{(1-S_{\tau_i-1})|Y_{\tau_i-1}|\} = o_P(T^{1/2}),$$

again by (A.3). Jointly (A.13) and (A.14) give

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} S_t\theta^{2(t-v(t))}(Y_{v(t)-1}-\mu)^2 &= T^{-1}(1-\theta^2)^{-1}\sum_{i=1}^{N_T}(1-S_{\tau_i-1})\langle Y_{\tau_i-1}\rangle^2 + o_P(1) \\ &\xrightarrow{w} \sigma^2(1-\theta^2)^{-1}\int (1-S)R^2\,dS \\ &= \sigma^2(1-\theta^2)^{-1}\sum_{u\in(0,1]}\{\Delta R(u)\}^2, \end{aligned} \qquad \textbf{(A.15)}$$

with the convergence following from Theorem 2.7 of Kurtz and Protter (1991), Corollary 1, and (A.9). Finally,

$$\sum_{t=1}^{T} S_t\left(\sum_{j=0}^{t-1-v(t)}\theta^j\varepsilon_{t-1-j}\right)^2 = \sum_{t=1}^{T} S_t\left(\sum_{j=0}^{t-1}\theta^j\varepsilon_{t-1-j}\right)^2 - c_4,$$

where $c_4 := \sum_{t=1}^{T} S_t\sum_{j=t-v(t)}^{t-1}\theta^j\varepsilon_{t-1-j}(\sum_{j=0}^{t-1-v(t)}\theta^j\varepsilon_{t-1-j} + \sum_{j=0}^{t-1}\theta^j\varepsilon_{t-1-j})$ has absolute value bounded by $2(\max_{t\le T}|\varepsilon_t|)^2(\sum_{j=0}^{\infty}|\theta|^j)^2\sum_{t=1}^{T} S_t|\theta|^{t-v(t)} = o_P(T)$ in view of (A.6). Let $\Omega_{\lfloor Tu\rfloor} := T^{-1}\sum_{t=1}^{\lfloor Tu\rfloor}(\sum_{j=0}^{t-1}\theta^j\varepsilon_{t-1-j})^2$. As $\sup_{u\in[0,1]}|\Omega_{\lfloor Tu\rfloor} - (1-\theta^2)^{-1}\sigma^2u| \xrightarrow{P} 0$ (the law of large numbers convergence is uniform by monotonicity of all the involved functions and by continuity of the limiting function), it follows that

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} S_t\left(\sum_{j=0}^{t-1}\theta^j\varepsilon_{t-1-j}\right)^2 &= \sum_{t=1}^{T} S_t\,\Delta\Omega_t = \sigma^2(1-\theta^2)^{-1}\left(T^{-1}\sum_{t=1}^{T} S_t\right) + o_P(1) \\ &\xrightarrow{w} \sigma^2(1-\theta^2)^{-1}\bar{S}, \end{aligned} \qquad \textbf{(A.16)}$$

with the convergence following from the CMT.

https://doi.org/10.1017/S0266466608080572 Published online by Cambridge University Press

Now from (A.10)–(A.12), using (A.15) and (A.16), it obtains that

$$T^{-1}\sum_{t=1}^{T} Y_{t-1}\,\Delta Y_t \xrightarrow{w} \sigma^2 \int RdB - \sigma^2(1+\theta)^{-1}\left[\sum_{u\in(0,1]}\{\Delta R(u)\}^2 + \bar{S}\right].$$

In view of a standard joint convergence argument, subtracting $(T^{-1/2}Y_T)(T^{-1/2}\bar{Y}_{-1}) \xrightarrow{w} \sigma^2 R(1)\bar{R}$ and dividing by $T^{-2}\sum_{t=1}^{T}(Y_{t-1} - \bar{Y}_{-1})^2 \xrightarrow{w} \sigma^2 \int (R - \bar{R})^2$ completes the proof. ■

https://doi.org/10.1017/S0266466608080572 Published online by Cambridge University Press