Alma Mater Studiorum - University of Bologna

DOTTORATO DI RICERCA IN
Scienze Statistiche

XXXI Ciclo

Settore Concorsuale di afferenza: 13/D1
Settore Scientifico disciplinare: SECS-S/01

Threshold Autoregressive Moving-Average models: probabilistic structure, statistical aspects and applications

Presentata da
Greta Goracci

Supervisore
Prof. Simone Giannerini

Co-supervisori
Prof. Kung-Sik Chan
Prof. Howell Tong

Coordinatore Dottorato
Prof. Alessandra Luati

Contents

List of Figures

List of Tables

Preface

The thesis analyses threshold autoregressive moving-average models (TARMA). They are an extension of the well known threshold autoregressive models (TAR) as to allow serially dependent noise. In linear time series analysis, the moving-average extension of the autoregressive model yields a parsimonious model for linear time series analysis. Similarly, the TARMA model may provide a parsimonious model for non-linear time series analysis. The systematic study of TARMA models presents several challenges. In particular, the main issue concerns their probabilistic structure, namely, establishing the parametric conditions for ergodicity, due to the long-standing open problem of proving irreducibility. As the conditions for the stationarity of TARMA models were unknown, it has not been possible to develop the inferential aspects and use them in real applications. The thesis solves the probabilistic problems for the first order TARMA model, and, therefore, makes a first step to enable the practical application of TARMA models. The results allow us to develop a powerful unit root test for both linear and non-linear processes. One of the most serious drawbacks that affect unit root tests is the size distortion in presence of dependent errors, especially of the moving-average kind. To the best of our knowledge, all the proposals that address the problem of the size distortion due to MA processes do not consider non-linear alternatives. On the other hand, tests that have a non-linear specification in the alternative hypothesis do not deal with such issue and, as we will show, their size is severely biased. We use TARMA models to develop a novel unit root test based upon Lagrange multipliers that does not suffer from size distortions and, at the same time, allows for a wide and flexible non-linear alternative. We prove that our supLM test is consistent, it is similar (w.r.t to the MA parameter) and it is nuisance parameters free. Moreover, in addition to the asymptotic version of the test we propose a wild bootstrap version with very good properties in terms of size and power. The final part of the thesis is devoted to a preliminary empirical investigation regarding the parsimony of TARMA models.
The results are incorporated into the following articles:

> Chan and Goracci [2017]: *K.-S. Chan, G. Goracci (2018) "Necessary and sufficient conditions for the ergodicity of first-order threshold autoregressive moving-average processes", submitted.*
>
> Chan et al. [2018]: *K.-S. Chan, S. Giannerini, G. Goracci, H. Tong (2018) "Unit-root testing for linear and non-linear alternatives: a TARMA based approach", Technical Report.*
>
> Goracci [2018] *G. Goracci (2018) "On the parsimony of TARMA models for non-linear time series", Working paper.*

The thesis is organized as follows: in Chapter 1, we present a review on threshold models in time series analysis. In Chapter 2, we provide an essential theoretical background that contains the main mathematical tools used throughout the thesis. In Chapter 3 we present the derivation of the probabilistic structure of first order TARMA models, while, in Chapter 4, we present a novel unit root test for both linear and non-linear alternatives based on TARMA models. Finally, in Chapter 5, we perform an empirical investigation on the parsimony of TARMA models and use them to analyse the Canadian lynx time series.

Chapter 1

Threshold models

Threshold autoregressive models are a class of non-linear time series models that can describe many dynamical phenomena. They are based on the threshold principle: when the phenomenon crosses a certain threshold then it changes qualitatively. For instance, in Figures 1.1 and 1.2 there are two thresholds that identify three regimes. The green lower regime, the red middle regime and the yellow upper regime. They were introduced by Tong [1978]. Due to his contributions, in 2007, he has been awarded the Guy medal in Silver of the Royal Statistical Society. In a seminal work, Tong and Lim [1980] presented the threshold autoregressive (TAR) model, defined as follows:

l-TAR(p):

$$X_t = \begin{cases} \phi_{1,0} + \sum_{i=1}^{p} \phi_{1,i} X_{t-i} + \varepsilon_t & \text{if } X_{t-d} \leq r_1 \\ \phi_{2,0} + \sum_{i=1}^{p} \phi_{2,i} X_{t-i} + \varepsilon_t & \text{if } r_1 \leq X_{t-d} \leq r_2 \\ \vdots & \vdots \\ \phi_{l,0} + \sum_{i=1}^{p} \phi_{l,i} X_{t-i} + \varepsilon_t & \text{if } X_{t-d} > r_{l-1} \end{cases} \tag{1.1}$$

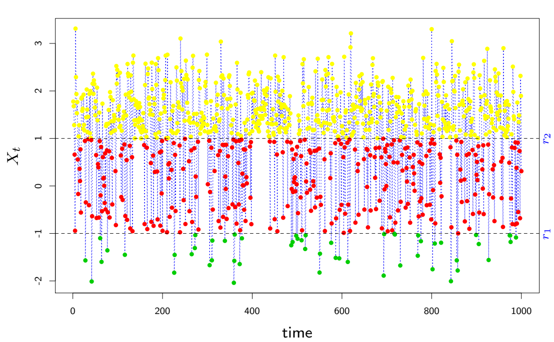

Figure 1.1: TARMA with 3 regimes: time plot

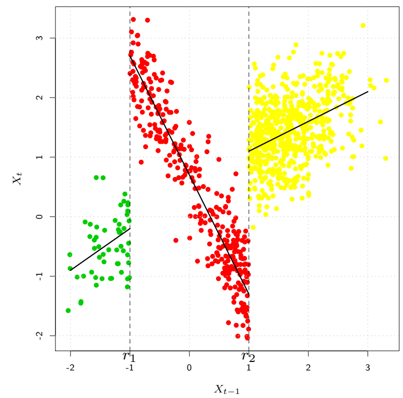

Figure 1.2: TARMA with 3 regimes: lag plot

Since the introduction of the TAR model, many useful variations have been developed and found diverse applications in non-linear time series analysis [Tong, 1990, 2007, Cryer and Chan, 2008, Chan, 2009, Tong, 2011, Chan et al., 2017], [see also Hansen, 2011, and references therein for a review]. Among them, the threshold moving-average model (TMA) and the threshold autoregressive moving-average model (TARMA) play a relevant role.

l-TMA(q):

$$X_t = \begin{cases} \varepsilon_t + \sum_{j=1}^{q} \theta_{1,j}\varepsilon_{t-j} & \text{if } X_{t-d} \le r_1 \\ \varepsilon_t + \sum_{j=1}^{q} \theta_{2,j}\varepsilon_{t-j} & \text{if } r_1 \le X_{t-d} \le r_2 \\ \vdots & \vdots \\ \varepsilon_t + \sum_{j=1}^{q} \theta_{l,j}\varepsilon_{t-j} & \text{if } X_{t-d} > r_{l-1} \end{cases} \tag{1.2}$$

l-TARMA(p,q):

$$X_t = \begin{cases} \phi_{1,0} + \sum_{i=1}^{p} \phi_{1,i}X_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_{1,j}\varepsilon_{t-j} & \text{if } X_{t-d} \le r_1 \\ \phi_{2,0} + \sum_{i=1}^{p} \phi_{2,i}X_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_{2,j}\varepsilon_{t-j} & \text{if } r_1 \le X_{t-d} \le r_2 \\ \vdots & \vdots \\ \phi_{l,0} + \sum_{i=1}^{p} \phi_{l,i}X_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_{l,j}\varepsilon_{t-j} & \text{if } X_{t-d} > r_{l-1}, \end{cases} \tag{1.3}$$

where (i) l is the number of regimes; (ii) p and q are the autoregressive and moving-average orders, respectively; (iii) $\phi_i, i = 1, \ldots, p$ and $\theta_j, j = 1, \ldots, q$ are the autoregressive and moving-average parameters, respectively; (iv) d the delay parameter; (v) $r_1 < r_2 < \ldots < r_{l-1}$ are the threshold parameters. The process $\{\varepsilon_t\}$ is the error term (innovation) and it is assumed to be a sequence of independent and identically distributed (iid) random variables. The orders p and q can be regime specific; for simplicity they are generally assumed identical across them.

Models (1.1), (1.2) and (1.3) are piecewise-linear, since the sub-models are linear in each regime. They are an extension of the corresponding well known AR, MA, ARMA models, allowing the parameters to change across regimes. It is well known that, in the linear case, ARMA models may provide more parsimonious models with respect to AR or MA models. Analogously, TARMA models may describe a wide range of dynamical phenomena with few parameters. Therefore, they possess a great descriptive power but the theory behind them is rather complex and the research has been stuck for 25 years due to unsolved probabilistic and statistical challenges. In particular, the conditions for ergodicity are necessary to develop the theoretical aspect related to parameter estimation and hypothesis testing. In the following, we start with a review of the relevant literature.

1.1 Probabilistic properties and parameter estimation of threshold models in time series

TAR models were introduced in Tong and Lim [1980] where the authors derived some sufficient conditions for their ergodicity. Moreover, they proposed several applications in order to show their descriptive power. Petruccelli and Woolford [1984] analyzed the simplest TAR model with one threshold equal to zero, $d = 1$, $p = 1$ and without intercepts, i.e.:

$$X_t = \begin{cases} \phi_{1,1} X_{t-1} + \varepsilon_t & \text{if } X_{t-1} \leq 0 \\ \phi_{2,1} X_{t-1} + \varepsilon_t & \text{if } X_{t-1} > 0. \end{cases} \tag{1.4}$$

They assumed $\{\varepsilon_t\}$ to be a sequence of iid random variables each having a strictly positive density on $\mathbb{R}$ and zero mean. Moreover, for each t, ε_t is independent from $X_0, X_1, \ldots, X_t$. They showed that the ergodic region is unbounded. In fact, they proved that the process defined in (1.4) is ergodic if and only if

$$\phi_{1,1} < 1, \quad \phi_{2,1} < 1 \quad \text{and } \phi_{1,1}\phi_{2,1} < 1. \tag{1.5}$$

Therefore, non-linearity allows us more freedom in the choice of parameters with respect to the linear AR(1) case where the ergodic region is bounded: the magnitude of each autoregressive parameter must be less then one. Moreover, under the assumption that $E\left[|\varepsilon_t|^{2+\zeta}\right] < \infty$ for some $\zeta > 0$, they proved the consistency and the asymptotic normality of the least square estimators for $\phi_{1,1}$ and $\phi_{2,1}$. Finally, they proposed a test for the null hypothesis $\phi_{1,1} = \phi_{2,1}$ (AR versus TAR). Chan et al. [1985] focused on (1.1) with $d = 1$, $p = 1$, i.e.:

$$X_t = \begin{cases} \phi_{1,0} + \phi_{1,1}X_{t-1} + \sigma_1\varepsilon_t & \text{if } X_{t-1} \leq r_1 \\ \phi_{2,0} + \phi_{2,1}X_{t-1} + \sigma_2\varepsilon_t & \text{if } r_1 < X_{t-1} \leq r_2 \\ \vdots & \vdots \\ \phi_{l,0} + \phi_{l,1}X_{t-1} + \sigma_l\varepsilon_t & \text{if } X_{t-1} > r_{l-1}, \end{cases} \tag{1.6}$$

where $\{\varepsilon_t\}$ is a zero mean iid process with strictly positive density over the real line and σ_i, $1 \leq i \leq l$, are positive parameters. Note that the error variance can change across the regimes. They obtained that the process defined in (1.6) is ergodic if and only if one of the following conditions holds:

$$\begin{aligned} \phi_{1,1} < 1, \quad & \phi_{l,1} < 1 \quad \text{and} \quad \phi_{1,1}\phi_{l,1} < 1, \\ \phi_{1,1} = 1, \quad & \phi_{l,1} < 1 \quad \text{and} \quad \phi_{1,0} > 0, \\ \phi_{1,1} < 1, \quad & \phi_{l,1} = 1 \quad \text{and} \quad \phi_{l,0} < 0, \\ \phi_{1,1} = 1, \quad & \phi_{l,1} = 1 \quad \text{and} \quad \phi_{l,0} < 0 < \phi_{1,0}, \\ \phi_{1,1}\phi_{l,1} = 1, \quad & \phi_{1,1} < 0 \quad \text{and} \quad \phi_{l,0} + \phi_{l,1}\phi_{1,0} > 0. \end{aligned}$$

Hence, the ergodicity of the process depends only on the two extreme regimes. Moreover, the conditions are less restrictive than those found in Petruccelli and Woolford [1984]. This is due to the presence of the intercept parameters that allow to include the boundaries of the previous region. Assuming that the innovations have finite absolute moment of order K, for some K, they proved that if

$$\phi_{1,1} < 1, \quad \phi_{l,1} < 1 \quad \text{and} \quad \phi_{1,1}\phi_{l,1} < 1$$

then the invariant probability distribution of the chain has finite K-th moment and it is geometrically ergodic. Furthermore, under the assumptions that (i) $\{X_t\}$ has a stationary distribution with finite second moment; (ii) $\sigma_i^2 = E[\varepsilon_{i,t}^2] < \infty$, for $i = 1, \ldots, l$ and (iii) r_i, $1 \leq i \leq l$ are known, they established strong consistency of the estimators of $\phi_{i,0}$ and $\phi_{i,1}$ and σ_i^2, for $i = 1, \ldots, l$ and the asymptotic normality for the estimator of $\phi_{i,0}$ and $\phi_{i,1}$. Guo and Petruccelli [1991] derived the complete classification of the parameter space of (1.6) into parametric regions over which such model is either transient or recurrent, and the recurrence region is further subdivided into regions of null recurrence or positive recurrence, or even geometric recurrence. Hence Model 1.1 has been well studied when $p = 1$ and $d = 1$. The analysis is less complete if $p \neq 1$ and $d \neq 1$, even if some results are available. Chen and Tsay [1991] extended the result of Petruccelli and Woolford [1984] to the model with a general positive delay parameter, i.e.:

$$X_t = \begin{cases} \phi_{1,1}X_{t-1} + \varepsilon_t & \text{if } X_{t-d} \leq 0 \\ \phi_{2,1}X_{t-1} + \varepsilon_t & \text{if } X_{t-d} > 0, \end{cases} \tag{1.7}$$

where $\{\varepsilon_t\}$ are iid random variables with absolutely continuous marginal distribution and positive probability density function over the real line and $E[\varepsilon_t] < \infty$, for each t. They proved that the process defined in (1.7) is ergodic if and only if

$$\phi_{1,1} < 1, \quad \phi_{2,1} < 1, \quad \phi_{1,1}^{s(d)} \phi_{2,1}^{t(d)} < 1 \quad \text{and} \quad \phi_{1,1}^{t(d)} \phi_{2,1}^{s(d)} < 1,$$

where $s(d)$ and $t(d)$ are nonnegative integers depending on d and $s(d)$ and $t(d)$ are odd and even numbers, respectively. They also provided the necessary and sufficient conditions for geometric ergodicity. Ling et al. [2007] investigated the first-order threshold moving-average model. They obtained a sufficient condition for a unique strictly stationary and ergodic solution of the model without the need to check for irreducibility. They also established the necessary and sufficient conditions for the invertibility of first-order TMA models. Furthermore, they discussed the extension of their results to the first-order multiple threshold moving-average models and higher-order threshold moving-average models. Li and Ling [2012] focused on the estimation of TMA. They showed that the estimator of the threshold is consistent and its limiting distribution is related to a two-sided compound Poisson process, whereas the estimators of other coefficients are strongly consistent and asymptotically normal.

The techniques developed for the TMA models paved the way for a systematic study of the much more challenging TARMA models. Liu and Susko [1992] focused on the TARMA(1,q):

$$X_t = \begin{cases} \phi_{1,0} + \phi_{1,1} X_{t-1} + \varepsilon_t + \sum_{j=1}^{q} \theta_{1,j} \varepsilon_{t-j} & \text{if } X_{t-d} \le r_1 \\ \phi_{2,0} + \sum_{i=1}^{p} \phi_{2,1} X_{t-1} + \varepsilon_t + \sum_{j=1}^{q} \theta_{2,j} \varepsilon_{t-j} & \text{if } r_1 \le X_{t-d} \le r_2 \\ \vdots & \vdots \\ \phi_{l,0} + \sum_{i=1}^{p} \phi_{l,1} X_{t-1} + \varepsilon_t + \sum_{j=1}^{q} \theta_{l,j} \varepsilon_{t-j} & \text{if } X_{t-d} > r_{l-1}, \end{cases}$$

where $\{\varepsilon_t\}$ is a iid process of zero mean random variables having finite absolute moment of order K, for some K. They were not able to prove the irreducibility of the process, but proved that it is ergodic if

$$\phi_{1,1} < 1, \quad \phi_{l,1} < 1 \quad \text{and } \phi_{1,1}\phi_{l,1} < 1. \tag{1.8}$$

Moreover, they showed that $\phi_{1,1} \le 1$ and $\phi_{l,1} \le 1$ are necessary conditions for ergodicity. They partially answered to the conjecture that the stationarity of the TARMA is based just on their autoregressive part as is the case for ARMA models. Ling [1999] derived the sufficient conditions for the stationarity and finiteness of the moments of TARMA models but they were not able to prove their irreducibility. Therefore, they hypothesized the TARMA to be irreducible and derived some conditions for ergodicity.

To the best of our knowledge, there are no other theoretical contributions on the probabilistic properties of TARMA models and the above contributions are the starting point for solving the long standing riddle of the regions of ergodicity. This will be presented in Chapter 3.

1.2 Unit-root tests for non-linear alternatives

Testing for the presence of a unit root in time series has important practical implications and is witnessed by the vast amount of literature devoted to the problem. Indeed, the fact that there are entire books dedicated to the exercise [Patterson, 2010, 2011, 2012, Choi, 2015] gives the idea of the many facets and of the non-trivial theoretical and practical issues related to it [see also Haldrup and Jansson, 2006, and references therein]. A general advice that might be drawn from the large amount of investigations is that unit root tests should never be applied without prior knowledge upon the process that might have generated the data. In fact, the theoretical framework and the distribution of the associated test statistics change according to the alternative hypothesis. In any case, as also pointed out in [Choi, 2015, Sec. 3.6], it remains unclear whether, besides ascertaining the presence of a unit root, these tests can be used to decide in favour of a particular model specification.

Recently, attention has been given to unit root tests against non-linear stationary alternatives. The simulation studies of Balke and Fomby [1997], Pippenger and Goering [1993] and Taylor [2001] showed that the Dickey-Fuller test [DF hereafter Dickey and Fuller, 1979], loses power against a non-linear alternative. Therefore, several different approaches have been proposed to derive more powerful tests in this respect. One possibility is to use a non-linear model, such as the TAR model, as the alternative hypothesis. The TAR model is particularly appealing in this context as it can incorporate one or more regimes with a unit root and still be globally stationary. It can provide a key interpretation in terms of a stationary non-linear process of many series that were deemed as non-stationary.

One of the main theoretical problems arising from testing for a unit root against a TAR model is that the threshold parameters are absent under the null hypothesis so that it is hard to derive the null asymptotic distribution of the statistics. Usually, this issue is addressed in two different steps: i) the null distribution of the statistic is derived assuming the threshold fixed and ii) it is considered either the supremum or the average or the exponential average of the statistic built on a set of possible values for the thresholds. Therefore, the choice of such set plays a crucial role. Enders and Granger [1998] suggested to use a F-type statistic to test a random walk against a two-regime SETAR model with fixed threshold, but the simulation studies showed that their test does not improve with respect to the DF test. Bec et al. [2004] focused on the three-regime SETAR with thresholds symmetrically chosen. They proposed a supWald test, computed on a set of possible values for the thresholds, derived its asymptotic null distribution and proved that it is nuisance parameter free. The model tested is a stationary three-regime SETAR model of the kind

$$\Delta X_t = \begin{cases} \alpha_{11}\Delta X_{t-1} + \cdots + \alpha_{1p}\Delta X_{t-p+1} + \mu_1 + \rho_1 X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} \leq -r \\ \alpha_{21}\Delta X_{t-1} + \cdots + \alpha_{2p}\Delta X_{t-p+1} + \mu_2 + \rho_2 X_{t-1} + \varepsilon_t, & \text{if } |X_{t-1}| < r \\ \alpha_{31}\Delta X_{t-1} + \cdots + \alpha_{3p}\Delta X_{t-p+1} + \mu_3 + \rho_3 X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} \geq -r \end{cases} \tag{1.9}$$

where $\varepsilon_t \sim$ i.i.d.$N(0,1)$ and the null hypothesis results:

$$H_0 : \rho_1 = \rho_2 = \rho_3 = 0.$$

They proved that a three-regime SETAR with unit root in the middle regime is stationary and mixing under mild conditions. Moreover, they derived the asymptotic distribution of the statistics under H_0'

$$H_0' : \rho_1 = \rho_2 = \rho_3 = 0; \mu_1 = \mu_2 = \mu_3 = 0; \alpha_{ij} = 0 \text{ for } i = 1,2,3 \; j = 1, \dots, p.$$

They proposed three different test statistics but they suggested to use the following:

$$\mathbf{BBC} = \sup_{r \in [\underline{r}, \overline{r}]^\intercal} LR_n(r), \quad \text{where} \quad LR_n(r) = n \log \frac{\tilde{\sigma}^2}{\hat{\sigma}^2} \tag{1.10}$$

and $\tilde{\sigma}^2$ and $\hat{\sigma}^2$ are the residual variances from the OLS estimated model (1.9) with and without the restriction $\rho_1 = \rho_2 = \rho_3 = 0$, respectively. Also, Kapetanios and Shin [2006] tested against a three-regime SETAR model with the middle regime (usually named corridor regime) constrained to be a random walk. They proposed a Wald statistic and proved that it is not dependent on the threshold parameters. The thresholds grid is selected such that the corridor regime has a finite width both under the null and the alternative hypothesis. The model tested is a stationary three-regime SETAR model of the kind

$$\Delta X_t = \begin{cases} \beta_1 X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} \le r_1 \\ \varepsilon_t, & \text{if } r_1 < X_{t-1} \le r_2 \\ \beta_2 X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} > r_2 \end{cases} \tag{1.11}$$

where $\varepsilon_t \sim$ iid$N(0,1)$ and the system of hypothesis results:

$$\begin{cases} H_0 : & \beta_1 = \beta_2 = 0 \\ H_1 : & \beta_1 < 0; \beta_2 < 0. \end{cases}$$

The three test statistics are based upon the Wald statistic:

$$\mathcal{W}_{(r_1,r_2)} = \hat{\boldsymbol{\beta}}^\intercal V(\hat{\boldsymbol{\beta}})^{-1} \hat{\boldsymbol{\beta}}$$

where $\hat{\boldsymbol{\beta}}$ is the OLS estimator of $\boldsymbol{\beta} = (\beta_1, \beta_2)^\intercal$. In particular:

$$\mathbf{KSs} = \sup_{i \in \Gamma} \mathcal{W}^{(i)}_{(r_1,r_2)}; \quad \mathbf{KSa} = \frac{1}{\#\Gamma} \sum_{i \in \Gamma} \mathcal{W}^{(i)}_{(r_1,r_2)}; \quad \mathbf{KSe} = \frac{1}{\#\Gamma} \sum_{i \in \Gamma} \exp\left(\frac{\mathcal{W}^{(i)}_{(r_1,r_2)}}{2} \right);$$

here $\mathcal{W}^{(i)}_{(r_1,r_2)}$ is the Wald statistic obtained from the i-th point of the threshold parameters grid set Γ. Bec et al. [2008a] used the same model considered in Bec et al. [2004], but revisited the choice of the set for the thresholds. They proposed an adaptive procedure such that the set is bounded under the null hypothesis and unbounded under the alternative. Other contributions include Seo [2008], Park and Shintani [2016], de Jong et al. [2007].

The aforementioned authors used models where the threshold variable coincides with the lagged dependent variable. Among the works based on TAR models with possibly exogenous threshold variables we mention Caner and Hansen [2001] that examined a two-regime TAR(p) with an autoregressive unit root. They treated a statistical test to analyse simultaneously both non-linearity and non-stationarity. In particular, they studied Wald tests for a threshold effect (for non-linearity) and Wald and t tests for unit roots. Their statistic has an asymptotic null distribution with two components: one that reflects the unit root and deterministic trends and it is free of nuisance parameters; and one that is identical to the empirical process found in the stationary case, and is nuisance-parameter dependent. Moreover, they proposed bootstrap procedures to approximate the sampling distribution. Also, Enders and Granger [1998] used an auxiliary model where the threshold variable is taken to be the first differences of the dependent variable. Moreover, Giordano et al. [2017] presented a Wald test based on a double threshold process where the innovation process has a threshold structure.

One of the most serious drawbacks that affect unit root tests is the size distortion in presence of dependent errors, especially of the moving-average kind [for a detailed account see Ch. 6 and 9 in Patterson, 2011]. Such distortion is particularly evident when the root of the moving-average polynomial of the first differenced series is large and positive[1]. Hosseinkouchack and Hassler [2016] proposed a variance ratio-type unit root test and compared it with several recent tests, some of which are nuisance parameter free. The results show clearly that the size distortion affects practically all the tests and persists for a sample size of 1000. The same behaviour is observed in Cook [2010] where the practical usefulness of the so called robust range unit root tests is questioned.

A first way to cope with the issue of the size distortion is to augment the DF regression with a number of lagged predictors (say p) as to model the moving-average component. The problem with this approach is that it is bound to an arbitrary lag selection step that can lead either to a severe power loss or to no size correction at all [see e.g. Agiakloglou and Newbold, 1999]. Ng and Perron [2001] proposed unit root tests and a modified information criterion (MIC) for lag selection in the ADF regression that reduces considerably the size distortion while maintaining good power properties. A slight modification of the aforementioned tests that improves over non local alternatives was proposed in Perron and Qu [2007]. Contrary to the results reported in Ng and Perron [2001], Hosseinkouchack and Hassler [2016] found no size improvement when they apply the modified information criterion to the GLS-DF test proposed in Elliott et al. [1996]. The modified test statistics proposed in Ng and Perron [2001] are bound to the estimation of the (autoregressive) spectral density at frequency zero. An heteroskedasticity-robust version of the MIC criterion and a wild bootstrap

[1] The parametrization we use for an ARMA(p, q) process is the following: $(1 - B\phi_1 - \cdots - B^p\phi_p)X_t = (1 - B\theta_1 - \cdots - B^p\theta_q)\varepsilon_t$, where $B^pX_t = X_{t-p}$.

ADF test were proposed in Cavaliere et al. [2015], where the authors showed that the MAIC criterion overestimates the lag order in a number of situations and advocated the use of their modified criterion both under homoskedasticity and heteroskedasticity. Paparoditis and Politis [2018] found a theoretical justification for the apparent contradictory behaviour of the ADF test in the presence of correlated data. They proved that the asymptotic distribution of the ADF statistic under the null hypothesis is valid under very general assumptions regarding the innovation process, namely, it is valid with the innovations following a zero mean, second order stationary (linear or non-linear) process having a continuous and strictly positive spectral density. Despite this flexibility, they found that the finite sample distribution of the statistic might differ consistently from the asymptotic counterpart in virtue of the low rate of convergence of the estimator for the autoregressive parameter which is of the order of $O(\sqrt{n/p})$. This is markedly different from the expected order $O(n)$ under the null, and from the order $O(\sqrt{n})$ of other test statistics under the alternative hypothesis. Moreover, as p diverges, the slow rate of convergence under the alternative is due to the lagged predictors of the ADF regression becoming asymptotically collinear. A second approach to deal with ARMA-type errors in unit root tests is to incorporate them directly in the parametric specification of the model, [for an account see also Ch. 7 of Patterson, 2011]. This idea was developed in Said and Dickey [1985] where the limiting distributions of non-linear least squares regression estimators of the parameters of an ARIMA$(p, 1, q)$ model were derived. In Galbraith and Zinde-Walsh [1999], the authors derived the asymptotic distribution of the ADF statistics in MA processes. More recently, Davidson [2010] analyzed the problem from the point of view of bootstrap testing and proposed a resampling procedure to address the issue.

As mentioned in the preface, those tests that address the problem of the size distortion do not consider non-linear alternatives. On the other hand, tests that have a non-linear specification in the alternative hypothesis do not deal with such issue. This motivated us to study a unit root test having good properties in terms of size and power against non-linear alternatives. As we will show in Chapter 4, TARMA models are the key to addressing such challenge.

Chapter 2

Mathematical tools

In this chapter, we present some mathematical tools used to develop the main results of the thesis. In Section 2.1 we introduce the essential concepts of Markov Chain theory. We focus on the notion of irreducibility and the classification of the long-run probabilistic behaviour of an irreducible Markov chain. In Section 2.2 we summarize the theory concerning the convergence of probability measures in metric spaces dwelling on the convergence of stochastic elements in the functional space $\mathcal{D}$ and the convergence of stochastic integrals. In Section 2.3, we introduce the definition of contiguity and the concept of change of measure. For a detailed treatment of these topics see Meyn and Tweedie [2012] for the Markov Chain theory; Billingsley [1968] for the convergence of probability measures in metric spaces and Øksendal [2003] for stochastic differential equations.

2.1 Classification of a φ-irreducible Markov chain

Let $\mathcal{X}$ and $\mathcal{B}(\mathcal{X})$ be a general set and a countably generated σ-field on it, respectively. We assume that $\mathcal{X}$ is a topological space and $\mathcal{B}(\mathcal{X})$ the Borel σ-field. Moreover, consider $\mathbf{X} = \{X_t\}_{t\in\mathbb{Z}^+}$ to be a stochastic process evolving in $\mathcal{X}$ and $P^n(x, A)$ the probability that the chain reaches A from x in n steps, with $x \in \mathcal{X}$ and $A \subset \mathcal{X}$. In the following, we use P_x (or E_x) to indicate that the probability (expectation) is computed with the starting point equal to x, i.e. $X_0 = x$.

Definition 1. The family of probabilities $P = \{P(x, A), x \in \mathcal{X}, A \subset \mathcal{X}\}$ is a *transition probability kernel* or *Markov transition functional* if:

(i) for each $A \subset \mathcal{X}$, $P(\cdot, A)$ is a non negative measurable function on $\mathcal{X}$;

(ii) for each $x \in \mathcal{X}$, $P(x, \cdot)$ is a probability measure on $\mathcal{B}(\mathcal{X})$.

Definition 2. The process $\mathbf{X} = \{X_t\}_{t\in\mathbb{Z}^+}$ is a *time-homogeneous Markov chain* if

$$\text{Prob}(X_{t+n} \in A \mid X_s, s < t, X_t = x) = P^n(x, A).$$

The independence of P^n on the values X_s, $s < t$, given known value of X_t, is named the *Markov property*, while the independence between P^n and t *time-homogeneity property*.

Now, we introduce the key concept of irreducibility that refers to the possibility for the chain to reach every part of $\mathcal{X}$ with non zero measure, from every starting point.

Definition 3. The process $\mathbf{X} = \{X_t\}_{t\in\mathbb{Z}^+}$ is *φ-irreducible* if there exists a measure φ on $\mathcal{B}(\mathcal{X})$ such that, for every $x \in \mathcal{X}$ and $A \subset \mathcal{X}$ with $\varphi(A) > 0$, the probability that the chain reaches A starting from x is positive. The measure φ is called *irreducibility measure.*

Definition 4. Let ψ be a measure on $\mathcal{B}(\mathcal{X})$. ψ is the *maximal irreducibility measure* for the chain $\mathbf{X}$ if

(i) $\mathbf{X}$ is ψ-irreducible;

(ii) if $\mathbf{X}$ is φ-irreducible for some φ, then φ is absolutely continuous w.r.t. ψ ($\varphi \ll \psi$), i.e.:

$$\psi(A) = 0 \text{ implies } \varphi(A) = 0 \text{ for each } A \in \mathcal{B}(\mathcal{X}).$$

We set

$$\mathcal{B}^+(\mathcal{X}) = \{A \in \mathcal{B}(\mathcal{X}) : \psi(A) > 0\}.$$

Irreducibility gives an important idea on the behaviour of the chain. In fact, if the chain is φ-irreducible, it seems that small change in the initial position do not change substantially the behaviour of the chain, allowing to reach the same set. The definition guarantees that non negligible sets A, namely such that $\varphi(A) \neq 0$, are always reached by the chain. In the analysis of the long-run probabilistic behaviour of a Markov chain, establishing the φ-irreducibility of the chain is the primary step. In fact, the problem of classifying the long-run probabilistic behaviour of a Markov chain can be posed only if it is φ-irreducible. We recall that, in the context of chains evolving in a countable space, an irreducible Markov chain is characterized by the solidarity property in that all its states have the same classification. Hence, it is meaningful to classify the whole chain. When the state-space is a general set $\mathcal{X}$, there is a similar solidarity property: if the chain is φ-irreducible then there exists a partition of $\mathcal{X}$ whose elements have the same classification. These sets are called *small sets* (or *status sets* or *petite sets*). Therefore, if the chain is φ-irreducible, it is sufficient to analyze only the behaviour of such sets in order to derive the probabilistic structure of the whole chain. In addition, several powerful theorems to detect the probabilistic structure of a chain require it to be φ-irreducible (e.g. the Tweedie's criteria, see below).

Irreducibility can be studied via the concept of reachability in Control Theory. In particular, it is needed to (i) analyze the deterministic control model associated to $\{X_t\}$, where $\{\varepsilon_t\}$ is replaced by a deterministic real-valued sequence $\{u_t\}$; (ii) derive the *reachable set* which consists of all states y such that for each state x, there exists a positive integer m and a sequence, $u_t, \ldots, u_{t+m-1}$, such that $X_t = x$ and $X_{t+m} = y$; (*iii*) prove that for each $k \in \mathbb{N}$ the k-step transition probability density of $\{X_t\}$ is positive over the set of states that can be reached by the associated deterministic chain in k steps. This ensures that if the reachable set Ω has positive Lebesgue measure, then $\{X_t\}$ is irreducible w.r.t. the Lebesgue measure restricted on Ω.

An important aspect in the analysis of a Markov chain is to study its stability. One can refer to the stability in different ways: (*i*) the property of returning at the starting point (weak stability, irreducibility), (*ii*) the property of visiting the non null-measure sets with probability one or in a finite mean time (mild stability, recurrence), (*iii*) the convergence of the n-step transition probability to a invariant measure (strong stability, ergodicity).

Definition 5. A ψ-irreducible chain is *recurrent* if for each $A \in \mathcal{B}^+(\mathcal{X})$, it holds that:

$$\sum_n P^n(x, A) = \infty.$$

Otherwise, it is *transient*.

The notion of recurrence is based on the expected value of the random variable η_A, which counts the number of visits to a set A, i.e.:

$$\eta_A = \sum_{n=1}^{\infty} I\left(X_n \in A\right).$$

Hence, a ψ-irreducible chain is recurrent if $E_x\left(\eta_A\right) = \infty$, for each $A \in \mathcal{B}^+(\mathcal{X})$. Now, we introduce a stronger concept of recurrence, named Harris recurrence, that considers the event that the chain enters in A infinitely often or, equivalently, $\eta_A = \infty$.

Definition 6. A ψ-irreducible chain is *Harris recurrent* if for each $A \in \mathcal{B}^+(\mathcal{X})$, it holds that:

$$P_x\left(\eta_A = \infty\right) = 1, \quad \text{for each } x \in A.$$

We focus on recurrent chains and divide them into positive recurrent and null recurrent. The strongest possible form of stability is that the distribution of X_n does not change as n assumes different values. If this is the case, then the Markov property implies that the finite dimensional distributions of $\mathbf{X}$ are invariant under translation in time. In this respect, we introduce the concept of invariant measure.

Definition 7. A σ-finite measure π on $\mathcal{B}^+(\mathcal{X})$ is an *invariant measure* if it satisfies the following balance equation:

$$\pi(A) = \int_{\mathcal{X}} \pi(dx) P(x, A), \quad \text{for each } A \in \mathcal{B}(\mathcal{X}).$$

A key characterization of recurrent chains is that they admit a unique (up to a multiplication constant) invariant measure π. If the invariant measure is finite, then it may be normalized as to obtain a stationary probability measure, otherwise it is not possible to build such stationary probability measure. This leads us to the following classification of recurrent chains.

Definition 8. A ψ-irreducible recurrent chain is *positive recurrent* if it admits an invariant probability measure π. Otherwise, it is *null recurrent*.

Definition 9. A ψ-irreducible chain is *ergodic* if it is positive recurrent and aperiodic, i.e.

$$\gcd\{n | P^n(x, x) > 0\} = 1, \quad \text{for each } x \in \mathcal{X}.$$

Essentially, ergodicity means that the sequence of transition probability kernels $P^n(x, \cdot)$ converges to the invariant measure π, as n increases, i.e.:

$$\lim_{n \to \infty} \|P^n(x, \cdot) - \pi\| = 0,$$

where $\| \cdot \|$ is the *total variation norm*, defined as:

$$\|\mu\| = \sup_{A \in \mathcal{B}(\mathcal{X})} \mu(A) - \inf_{A \in \mathcal{B}(\mathcal{X})} \mu(A), \text{ where } \mu \text{ is a measure on } \mathcal{B}(\mathcal{X}).$$

Obviously, such convergence of the transition probability kernels to the invariant measures can occur with different rates. We examine when the convergence takes place at uniform geometric rate.

Definition 10. A ψ-irreducible ergodic chain is *geometrically ergodic* if there exists $r > 1$ such that:

$$\sum_{n=1}^{\infty} r^n \|P^n(x, \cdot) - \pi\| < \infty.$$

As stated before, the long-run probabilistic behaviour of a φ-irreducible chain can be derived studying its small sets. In this respect, the most powerful tools are the so called Tweedie's criteria that allow to classify a φ-irreducible chain. Essentially they are based on two steps: (i) proving that any compact set (e.g. the set of the form $[-M, M]$) is a small set and (ii) finding the energy functions that satisfy such criteria. Below, we report the definition of a small set, a useful criterion to prove if a set is a small set and some Tweedie's criteria.

Definition 11. A set $C \in \mathcal{B}(\mathcal{X})$ is called ν_m*-small set* if there exists a $m > 0$ and a non-trivial measure ν_m on $\mathcal{B}(\mathcal{X})$ such that

$$P^m(x, B) \geq \nu_m(B), \text{ for all } x \in C \text{ and } B \in \mathcal{B}(\mathcal{X}).$$

Proposition 12. *(from [Nummelin, 2004, Proposition 2.11]) A set C is a small set if there exists a set D, with $\mu(D) > 0$ and a non-negative integer $L < \infty$ such that:*

$$\inf_{x \in [-M, M]} \sum_{n=0}^{L} P^n(x, C) > 0, \qquad \textit{for each } C \subseteq D, \textit{ with } \mu(C) > 0.$$

Proposition 13. *A ψ-irreducible chain is recurrent if there exist a small set A and a function $g(\cdot) : \mathbb{R} \to [0, +\infty]$ such that*

(i) $E[g(V_{t+1})|V_t = x] \leq g(x)$, *for all* $x \in A^c$;
(ii) $g(x) > \sup_{y\in[-M,M]} g(y)$, *for all* $x \in A^c$;
(iii) $B_n = \{y \in \mathbb{R} : g(y) \leq n\}$ *is a small set for all sufficient large* n.

Proposition 14. *A ψ-irreducible is null if there exist $\delta > 0$, a non-negative function $g(\cdot)$ and a set A, with $\mu(A) > 0$ and $\mu(A^c) > 0$, such that:*

(i) $E[g(V_{t+1})|V_t = x] \geq g(x)$, *for each* $x \in A^c$;
(ii) $E[|g(V_{t+1}) - g(x)| \quad | \quad V_t = x] \leq \delta$, *for each* $x \in \mathbb{R}$;
(iii) $g(x) > \sup_{y\in A} g(y)$, *for each* $x \in A^c$.

Proposition 15. *A ψ-irreducible chain is ergodic if there exist a small set A and a function $g(\cdot) : \mathbb{R} \to [0, +\infty]$ and some constants $K, \gamma > 0$, such that*

(i) $E[g(V_{t+1})|V_t = x] \leq g(x) - \gamma$, *for all* $x \in A^c$;
(ii) $E[g(V_{t+1})|V_t = x] \leq K$, *for all* $x \in A$.

Proposition 16. *A ψ-irreducible chain is geometrically ergodic if there exist a function $g(\cdot) : \mathbb{R} \to [1, +\infty)$ and some constants $M, K, \gamma > 0$, such that $[-M, M]$ is a small set and*

(i) $E[g(V_{t+1})|V_t = x] \leq (1 - \gamma)g(x)$, *for all* $x \in [-M, M]^c$;
(ii) $E[g(V_{t+1})|V_t = x] \leq K$, *for all* $x \in [-M, M]$.

2.2 Convergence of probability measures in a metric spaces

Definition 17. Given a set $\mathcal{X}$ and a function $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, the pair $(\mathcal{X}, d)$ is a *metric space* if d satisfies the following conditions:

$d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$, for each $x, y \in \mathcal{X}$;

$d(x, y) = d(y, x)$, for each $x, y \in \mathcal{X}$;

$d(x, y) \leq d(x, z) + d(z, y)$.

Below, let $(\mathcal{X}, d)$ be a metric space and $\mathcal{B}(\mathcal{X})$ the σ-field generated by the open sets of $\mathcal{X}$. For notational convenience, we will write $\mathcal{B}$ instead of $\mathcal{B}(\mathcal{X})$.

Definition 18. A function $P : \mathcal{B} \to [0, 1]$ is a *probability measure* if

(i) $P(\emptyset) = 0$;

(ii) $P(\mathcal{X}) = 1$;

(iii) given $\{A_n\}_{n\in\mathbb{N}}$ such that $A_i \cap A_j = \emptyset$, for each $i \neq j$ then

$$P\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} P(A_n).$$

The triplet $(\mathcal{X}, \mathcal{B}, P)$ is called *probability space.*

Definition 19. Let P_n and P be probability measures on $(\mathcal{X}, \mathcal{B})$. We say that P_n *converges weakly* to P (and we write $P_n \rightsquigarrow P$) if

$$\int_{\mathcal{X}} f dP_n \to \int_{\mathcal{X}} f dP \text{ for every bounded continuous real function } f \text{ on } \mathcal{X}.$$

We are interested in studying the weak convergence in the space $\mathcal{D}$.

2.2.1 Weak convergence in $\mathcal{D}$

The space $\mathcal{D}$ is the space of functions $x = x(t)$ that are continuous or present a discontinuity of the first kind (i.e. $x(t-)$ and $x(t+)$ exist but differ and $x(t)$ lies between them). This functional space is suitable to describe processes that contain jumps, as the Wiener process. A measurable map X from a probability space to a metric space is called *random element.* Note that if the codomain of X is $\mathbb{R}$ then it is a random variable. Now, we introduce the random elements of the space $\mathcal{D}$, that are called *random functions.*
Let $(\mathcal{X}, \mathcal{B}, P)$ be a probability space and consider the map

$$\begin{aligned} X : (\mathcal{X}, \mathcal{B}, P) &\to \mathcal{D} \\ \omega &\mapsto X_t(\omega), \text{ with } X_t(\omega) : \mathbf{T} \to \mathbb{R}. \end{aligned}$$

In many applications, the index set $\mathbf{T}$ is of the form $[0, T]$. Usually, $X_t(\omega)$ is named *sample path* and, for notational convenience, we set $X_t = X_t(\omega)$. Now, consider a sequence of random functions $\{X^n\}_{n\in\mathbb{N}}$. We indicate the corresponding sequence of trajectories $\{X_t^n\}$, $t \in \mathbf{T}$ and $n \in \mathbb{N}$. Note that X_t is a random process and so $\{X_t^n\}$ is a sequence of stochastic processes.

Definition 20. A sequence $\{X^n\}$, $n \in \mathbb{N}$, of random elements *converges in distribution* to the random element X $\left(\text{we write } X^n \xrightarrow[n\to\infty]{\text{D}} X\right)$ if $P_n \rightsquigarrow P$, where P_n and P are the distributions of X^n and X, respectively.

We report below a useful criterion to prove the convergence of a sequence of random functions. To this aim, we introduce the key concept of tightness.

Definition 21. A probability measure P on $(\mathcal{X}, \mathcal{B})$ is *tight* if, for each positive ε, there exists a compact set K such that $P(K) > 1 - \varepsilon$.

Another useful definition of tightness is the following: a sequence is tight if each subsequence admits a converging sub-subsequence.

Proposition 22. *(Kunita [1986]) A sequence of stochastic processes* $\{X_t^n\}$, $t \in [0,T]$, $n \in \mathbb{N}$, *right continuous with the left hand limits is tight if there are positive constants* K, γ *and* α *not depending on* n *such that*

$$E[|X_t^n - X_{t_1}^n|^\gamma |X_{t_2}^n - X_t^n|^\gamma] \le K|t_2 - t_1|^{1+\alpha}, \qquad 0 \le t_1 < t < t_2 \le T, \quad (2.1)$$

$$E[|X_t^n|^\gamma] \le K, \qquad 0 \le t \le T. \quad (2.2)$$

The convergence in distribution of a sequence of random elements, lying in a metric space, follows by proving the convergence of the finite-dimensional distributions and the asymptotic tightness of the sequence. It is required that for each $\epsilon > 0$ the set $\mathbf{T}$ can be partitioned into finitely many sets $\mathbf{T}_1, \ldots, \mathbf{T}_k$ such that (asymptotically) the variation of the trajectories X_t^n is smaller than ϵ on every set $\mathbf{T}_i$, with large probability. Then, the behaviour of the process can be described by the behaviour of the marginal vectors $\left(X_{t_1}^n, \ldots, X_{t_k}^n\right)$ for arbitrary fixed points $t_i \in \mathbf{T}_i$. If these margins converge, then the processes converge.

Proposition 23. *(van der Vaart [1998], Theorem 18.14, p. 261) A sequence of random functions* $\{X^n\}$, $n \in \mathbb{N}$ *converges in distribution to a tight random element* X *if and only if the following conditions hold.*

(i) The sequence $(X_{t_1}^n, \ldots, X_{t_k}^n)$ *converges in distribution in* $\mathbb{R}^k$ *for every finite set of points* $t_1, \ldots, t_k$ *in* $\mathbf{T}$.

(ii) The sequence of stochastic processes $\{X_t^n\}$, $t \in \mathbf{T}$, *is asymptotically tight or, equivalently, for every* $\epsilon, \eta > 0$ *there exists a partition of* $\mathbf{T}$ *in finitely many sets* $\mathbf{T}_1, \ldots, \mathbf{T}_k$ *such that:*

$$\limsup_{n \to \infty} P\left(\sup_{i \in \{1,\ldots,k\}} \sup_{t_1, t_2 \in \mathbf{T}_i} \left| X_{t_1}^n - X_{t_2}^n \right| \ge \epsilon \right) \le \eta.$$

2.2.2 Convergence of stochastic integrals

Definition 24. A process H_t is *simple predictable* if

$$H_t = H_0 I_{\{0\}}(t) + \sum_{i=1}^n H_i I_{(T_i, T_{i+1}]}(t),$$

where $0 = T_1 \le T_2 \le \ldots, \le T_n \le T_{n+1} < \infty$ is a sequence of stopping times, $H_i \in \mathfrak{F}_{T_i}$, $|H_i| < \infty$ a.s. for each $1 \le i \le n$.

Let $\mathfrak{P}$ be the collection of simple predictable process and we topologize it by the uniform convergence. Moreover, set $\mathcal{L}^0$ to be the set of a.s. finite random variables.

Definition 25. A process X_t is a *semimartingale* if the map $\mathcal{I}_X : \mathfrak{P} \to \mathcal{L}^0$ is continuous on compact time sets, where

$$\mathcal{I}_X = H_0 X_0 + \sum_{i=1}^n H_i \left(X_{\min\{t, T_{i+1}\}} - X_{\min\{t, T_i\}} \right), \quad H \in \mathfrak{P},$$

Set $\mathfrak{D}$ to be the space of adapted processes whose paths are right continuous with left limits.

Definition 26. Consider the processes $X_t \in \mathfrak{D}$ and $H_t \in \mathfrak{P}$. The *stochastic integral* of H_t w.r.t X_t is defined as the following process:

$$\int H dX = H_0 X_0 + \sum_{i=1}^{n} H_i \left(X_{\min\{t,T_{i+1}\}} - X_{\min\{t,T_i\}} \right).$$

Remember that $\mathcal{D}$ is the space of functions $x = x(t)$ right continuous with left limits. Therefore, if a process X_t belongs to $\mathfrak{D}$ then its paths x live in $\mathcal{D}$.

Definition 27. A sequence of functions $\{x_n(t)\}_{n\in\mathbb{N}}$, $x_n(t) \in \mathcal{D}$, *converges in the Skorohod topology* to $x(t) \in \mathcal{D}$ if there exists a sequence of functions $\lambda_n : \mathbb{R}_+ \to \mathbb{R}_+$ such that:

λ_n is an increasing, bijective function, for each n;

$\lambda_n(t)$ converges to t uniformly as n increases;

$x_n(\lambda_n(t))$ converges to $x(t)$ uniformly on compacts as n increases.

If $\{x_n\}$ converges to x uniformly on compacts then, for large enough n, the jumps of x_n must occur at the same times as those of x and the jump sizes of x_n converge to the corresponding sizes of x. If $\{x_n\}$ converges to x in the Skorohod topology then the jump sizes of x_n converge to the corresponding sizes of x, but they do not need to occur at the same time. It suffices that the times of occurrence of the jumps converge.

Definition 28. A sequence of semimartingales $\{X_t^n\}$, $n \in \mathbb{N}$ and $t \in \mathbf{T}$, is said to be *uniformly tight* if for each $u > 0$, the set

$$\left\{ \int_0^u H_{s^-}^n dX_s^n,\ H_t^n \in \mathfrak{P}^n, |H_t^n| \leq 1,\ n \in \mathbb{N}, t \in \mathbf{T} \right\}$$

is stochastically bounded, uniformly in n.

Proposition 29. *If (H^n, X^n) converges in distribution in the Skorohod topology to (H, X) and if $\{X^n\}$, $n \in \mathbb{N}$, is a sequence of semimartingales uniformly tight then there exists a filtration $\mathbb{H}$ such that X is an $\mathbb{H}$ semimartingale and $(H^n, X^n, \int H_-^n dX^n)$ converges in distribution to $(H, X, \int H_- dX)$.*

2.3 Contiguity and change of measure

2.3.1 Likelihood ratio

Let P and Q be two measures on a measurable space $(\mathcal{X}, \mathcal{B})$ whose densities are p and q, respectively. Their supports are indicated as

$$\mathcal{X}_P = \{p > 0\} \quad \text{and} \quad \mathcal{X}_Q = \{q > 0\}.$$

Definition 30. We say that:

Q is *absolutely continuous* w.r.t. P (and we write $Q \ll P$) if

$$P(A) = 0 \text{ implies } Q(A) = 0, \text{ for each } A \in \mathcal{B}.$$

P and Q are *orthogonal* (and we write $Q \perp P$) if $\mathcal{X}$ can be partitioned as

$$\mathcal{X} = \mathcal{X}_P \cup \mathcal{X}_Q, \text{ with } \mathcal{X}_P \cap \mathcal{X}_Q = \emptyset \text{ and } P(\mathcal{X}_Q) = Q(\mathcal{X}_P) = 0.$$

Now, given $A \in \mathcal{B}$, define the following two measures:

Absolutely continuous part of Q w.r.t. P:

$$Q^a(A) = Q(A \cap \{p > 0\}).$$

Orthogonal part of Q w.r.t. P:

$$Q^\perp(A) = Q(A \cap \{p = 0\}).$$

It holds that the measure Q can be written as the sum $Q = Q^a + Q^\perp$, which is called the *Lebesgue decomposition* of Q w.r.t. P. It is easy to prove that $Q^a \ll P$ and $Q^\perp \perp P$.
Moreover, we have that:

$$Q^a(A) = \int_A \frac{q}{p} dP.$$

Therefore, the density of Q^a w.r.t. P is $\frac{q}{p} := \frac{dQ}{dP}$, that is named the Radon-Nikodym density or *Likelihood ratio*. The Likelihood ratio can be seen as a random variable $\frac{dQ}{dP} : \mathcal{X} \to [0, \infty)$ and it is of interest to study its law under P.

2.3.2 Contiguity

If a probability measure Q is absolutely continuous w.r.t. P, then the Q-law of a random vector $X : \mathcal{X} \to \mathbb{R}^k$, can be derived from the knowledge of the P-law and the Likelihood ratio, by applying the following formula:

$$E_Q f(X) = E_P f(X) \frac{dQ}{dP}.$$

Now, consider the asymptotic version of such problem. Given two sequences of measures $\{Q_n\}$ and $\{P_n\}$, $n \in \mathbb{N}$, under which conditions can a Q_n-limit law be obtained from a P_n-limit law? The answer is based of the concept of contiguity between measures.

Definition 31. Given two sequences of probability measures $\{P_n\}$ and $\{Q_n\}$ on measurable spaces $(\mathcal{X}_n, \mathcal{B}_n)$, $n \in \mathbb{N}$, we say that

$\{Q_n\}$ is *contiguous* w.r.t. $\{P_n\}$ (and we write $Q_n \triangleleft P_n$) if for each sequence of measurable sets $\{A_n\}$:

$$\lim_{n\to\infty} P_n(A_n) = 0 \text{ implies } \lim_{n\to\infty} Q_n(A_n) = 0.$$

$\{Q_n\}$ and $\{P_n\}$ are *mutually contiguous* (and we write $Q_n \triangleleft \triangleright P_n$) if $P_n \triangleleft Q_n$ and $Q_n \triangleleft P_n$.

The following Lemma gives a powerful tool to address the problem of proving the contiguity without using directly the definition.

Proposition 32. ***(Le Cam's first Lemma)*** *Let $\{P_n\}$ and $\{Q_n\}$, $n \in \mathbb{N}$, be two sequences of probability measures on measurable spaces $(\mathcal{X}_n, \mathcal{B}_n)$. Then the following statements are equivalent:*

1. $Q_n \triangleleft P_n$.
2. *If* $\frac{dP_n}{dQ_n} \xrightarrow[n\to\infty]{Q_n} U$ *along a subsequence, then* $P(U > 0) = 1$.
3. *If* $\frac{dQ_n}{dP_n} \xrightarrow[n\to\infty]{P_n} V$ *along a subsequence, then* $E(V) = 1$.
4. *For any statistics* $T_n : \mathcal{X}_n \to \mathbb{R}^k$*, if* $T_n \xrightarrow[n\to\infty]{P_n} 0$ *then* $T_n \xrightarrow[n\to\infty]{Q_n} 0$.

The following theorem answers the question above.

Proposition 33. *Let $\{P_n\}$ and $\{Q_n\}$, $n \in \mathbb{N}$, be two sequences of probability measures on measurable spaces $(\mathcal{X}_n, \mathcal{B}_n)$ and let $X_n : \mathcal{X}_n \to \mathbb{R}^k$ be a sequence of random vectors. Suppose that $Q_n \triangleleft P_n$ and*

$$\left(X_n, \frac{dQ_n}{dP_n} \right) \xrightarrow[n\to\infty]{P_n} (X, V).$$

Then $L(B) = EI_B(X)V$ defines a probability measure and $X_n \xrightarrow[n\to\infty]{Q_n} L$.

2.3.3 Change of probability measure

Definition 34. A stochastic process W_t is called *Wiener process* if it satisfies the following properties:

(i) $W_0 = 0$ a.s..

(ii) For every $0 < t_1 < t_2 \leq t_3 < t_4$, the increments $(W_{t_4} - W_{t_3})$ and $(W_{t_2} - W_{t_1})$ are independent random variables.

(iii) For every $0 < t_1 < t_2$, $(W_{t_2} - W_{t_1}) \sim N(0, t_2 - t_1)$.

(iv) W_t is continuous in t with probability 1.

Definition 35. We define X_t to be an *Ito's process* if it satisfies the following differential equation:

$$dX_t = \mu_t dt + \gamma_t dW_t,$$

where:

(i) μ_t and γ_t, $t \in [0,T]$, are processes adapted to the filtration $\mathfrak{I}_t$.

(ii) $P\left[\int_0^T |\mu_s|\, ds < \infty\right] = 1$.

(iii) $P\left[\int_0^T |\gamma_s^2|\, ds < \infty\right] = 1$.

Therefore:

$$X_t = \int_0^t \mu_s ds + \int_0^t \gamma_s dW_s.$$

Proposition 36. *Let P be a probability measure on $(\mathcal{X}, \mathcal{B})$ and X_t, $t \in [0,T]$, an Ito's process defined as follows:*

$$X_t = \int_0^s \gamma_s dW_s, \qquad \text{where } W_s \text{ is the } P\text{-Wiener process.}$$

If the Novikov's condition is fulfilled, i.e.:

$$E_P\left[e^{\frac{1}{2}\int_0^T \gamma_s^2 ds}\right] < \infty,$$

then there exists a probability measure Q on $(\mathcal{X}, \mathcal{B})$ such that:

1. $\frac{dQ}{dP} = e^{X_t - \frac{1}{2}[X]_t}$, *where $[X]_t$ is the quadratic variation of the process X_t, i.e.*

$$[X]_t = \int_0^T \gamma_s^2 ds.$$

2. *The process $\tilde{W}_t$ is the Q-Wiener process, where:*

$$\tilde{W}_t = W_t - \int_0^t \gamma_s ds.$$

Chapter 3

On the probabilistic structure of TARMA processes

This chapter analyses the long-run probabilistic behaviour of first-order TARMA models. This includes deriving the necessary and sufficient conditions for the models to be ergodic and studying their recurrence/transience properties. This problem has attracted much interest in the literature [Brockwell et al., 1992, Liu and Susko, 1992, Ling, 1999, Ling et al., 2007]. The recurrence/transience properties of a TAR model have been well-studied. In particular, results from [Petruccelli and Woolford, 1984, Chan et al., 1985, Guo and Petruccelli, 1991] together provide a complete classification of the parameter space of a first-order TAR model into parametric regions over which the TAR model is either transient or recurrent, and the recurrence region is further subdivided into regions of null recurrence or positive recurrence, or even geometric recurrence. The classification is less complete in the higher-order case, but see Chan and Tong [1985], Tong [1990], Tjøstheim [1990], Cline [2009].

Since the TARMA process admits a Markovian representation, Markov chain techniques (for instance, various drift criteria, for classifying continuous-state-space Markov chains Tweedie [1975, 1976], Chan and Tong [1985], Tjøstheim [1990], Tong [1990], An and Huang [1996], An and Chen [1997], Nummelin [2004], Cline [2009], Meyn and Tweedie [2012]) provide a natural approach for studying the long-run probabilistic behaviour of the TARMA model. However, such an approach requires the Markov chain to be irreducible. Except under some special cases (e.g. identical moving-average noise process across regimes studied by Brockwell et al. [1992]), the problem of characterizing when a TARMA process admits an irreducibile Markovian representation turns out to be very hard, even under the assumption that the probability density function (pdf) of the independent and identically distributed (iid) innovations driving the TARMA process is positive everywhere. Consequently, other approaches have been employed to study the stationarity and ergodicity of the TARMA model. Liu and Susko [1992] derived sufficient conditions for a TARMA model to be ergodic, under the assumption of irreducibility which, however, was not shown to hold. They underlined that the ergodicity depends only on the autoregressive parts in the two outermost regimes. Ling [1999] found some sufficient condition for the ergodicity of the TARMA model. However, this condition is much stronger than the necessary and sufficient condition for the ergodicity of a TAR(1) model. He concluded by remarking that *"It might be possible to find some weaker conditions, but it seems not to be an easy task"*. In more recent work, Ling et al. [2007] obtained sufficient conditions for stationarity and ergodicity of the TMA model without proving its irreducibility. Also, they emphasized the difficulty of establishing irreducibility for a TARMA model.

Here, our contribution is twofold. First, in Section 3.1 we introduce a novel Markovian representation of the TARMA process. For the first-order TARMA model, denoted as TARMA(1,1) model to be defined in Section 3.1, the 2-dimensional Markovian representation so constructed contains an embedded 1-dimensional Markov chain with which we are able to prove in Section 3.2 that all invertible (see below for definition) TARMA(1,1) models are irreducible under a very general distributional assumption on the iid innovations, specifically (C2) below. Moreover, the Markov chain can be shown in Section 3.3 to be aperiodic. A key observation is that the embedded 1-dimensional Markov chain and the 2-dimensional Markov chain share identical transience/recurrence properties. Leveraging on the 1-dimensional embedded Markov chain, we then derive a complete classification of the parameter space of a TARMA(1,1) model into parametric regions over which the model is either transient or recurrent, and the recurrence region is further subdivided into regions of null recurrence or positive recurrence, or even geometric recurrence. By combining the results on aperiodicity and positive recurrence, we derive a set of necessary and sufficient conditions for ergodicity, and deduce conditions for geometric ergodicity. For clarity of presentation, all results and proofs are first presented for a two-regime TARMA(1,1) model, with extension to multiple regimes briefly outlined in Section 3.5, showing that the irreducibility, transience/recurrence of a multiple-regime TARMA(1,1) model generally depend only on the autoregressive parameters of the two outermost regimes.

3.1 A Markovian representation

The p-th order two-regime threshold autoregressive moving-average TARMA(p, p) process $\{X_t\}$ satisfies the following difference equation:

$$X_t = \begin{cases} \phi_{1,0} + \sum_{j=1}^{p} \phi_{1,j} X_{t-j} + \varepsilon_t + \sum_{j=1}^{p} \theta_{1,j}\varepsilon_{t-j}, & \text{if } X_{t-d} \le r \\ \phi_{2,0} + \sum_{j=1}^{p} \phi_{2,j} X_{t-j} + \varepsilon_t + \sum_{j=1}^{p} \theta_{2,j}\varepsilon_{t-j}, & \text{otherwise,} \end{cases} \tag{3.1}$$

where the ϕ_i and θ_i with $i = 1, \ldots, p$ are the autoregressive and moving-average coefficients, respectively, r is the real-valued threshold parameter and the positive integer d is the delay parameter. The dynamics of a TARMA model then switches between two autoregressive moving-average (ARMA) sub-models depending on whether or not $X_{t-d} \le r$. The process is said to fall into the first regime whenever $X_{t-d} \le r$, otherwise it is in the second regime. With no loss of generality, it is assumed that $d \le p$. The innovation term $\{\varepsilon_t\}$ is a sequence of iid random variables, satisfying the following conditions:

(C1) The innovations admit finite k-th absolute moment, where $k \ge 1$ is a constant, and are of zero mean.

(C2) The innovations have an absolutely continuous probability distribution whose probability density function (pdf), denoted as $\rho(\cdot)$, is positive and continuous on $\mathbb{R}$.

(C2) can be relaxed as to require the common probability distribution of the innovations having an absolutely continuous component with a positive, continuous pdf. It can be readily checked that all the results derived below continue to hold with (C2) relaxed. The preceding formulation of the TARMA model admits several generalizations: (i) the model can be made conditionally heteroscedastic by changing the coefficient of ε_t in (3.1) from unity to some regime-specific coefficient, (ii) the AR and MA orders need not to be identical and may be regime specific, (iii) there may be more than two regimes, and (iv) the regimes may be delineated by a threshold variable more complex than X_{t-d}, for instance, $X_{t-d} - X_{t-d-1}$.

To avoid certain degeneracy to be explained below, we shall assume that, for $i = 1, 2$, the AR characteristic polynomial $1 - \sum_{j=1}^{p} \phi_{i,j} x^j$ and the MA characteristic polynomial $1 + \sum_{j=1}^{p} \theta_{i,j} x^j$, of the ith regime, have no common roots, for $i = 1, 2$. In particular, for the case of $p = 1$, the assumption of the AR and MA polynomials of each regime sharing no common roots is equivalent to the condition that $\phi_{i,1} + \theta_{i,1} \neq 0, i = 1, 2$. For linear ARMA models, if the AR polynomial and the MA polynomial share some common roots, then their common parts can be canceled out, thereby reducing the AR and MA orders by the number of common roots. For instance in the case of an ARMA(1,1) model, the AR and MA polynomials sharing a common root implies that the ARMA(1,1) model is actually a white-noise model; hence the 1-step (as well as all higher step) ahead predictor (conditional mean) is simply the stationary mean, i.e., the 1-step ahead predictor remains the same even after observing a new data case. For TARMA models, the AR and MA polynomials of a certain regime having a common root in some regime per se does not result in a lower-order ARMA sub-model in that regime, but some degeneracy would occur. For instance, Eqn (3.3) demonstrates that for the TARMA(1,1) model, $\phi_{i1} + \theta_{i,1} = 0$ implies that upon observing a new data case, the next 1-step ahead predictor is a deterministic function of its current counterpart, with non-zero probability. To sum up, the assumption of no common roots shared by the AR and MA polynomials in each regime of a TARMA model will be maintained throughout the paper. It is an interesting future research problem concerning the relaxation of the no-common-root assumption.

It is advantageous to study the probabilistic properties of a TARMA process via some Markovian representation. Specifically, we shall express X_t as some functional of a Markov chain. One such representation is motivated by decomposing the time-series observation as the "one-step-ahead" predictor plus innovation (see Akaike [1974] for a related predictor-based Markovian representation for an ARMA process), i.e., $X_t = V_t + \varepsilon_t$. In particular, the V's are driven by the following difference equation:

$$V_t = \begin{cases} \phi_{1,0} + \sum_{j=1}^{p} \phi_{1,j} V_{t-j} + \sum_{j=1}^{p} (\phi_{1,j} + \theta_{1,j}) \varepsilon_{t-j}, & \text{if } V_{t-d} + \varepsilon_{t-d} \le r \\ \phi_{2,0} + \sum_{j=1}^{p} \phi_{2,j} V_{t-j} + \sum_{j=1}^{p} (\phi_{2,j} + \theta_{2,j}) \varepsilon_{t-j}, & \text{otherwise.} \end{cases} \tag{3.2}$$

Let

$$D_t = (\phi_0(t), \underbrace{0, \ldots, 0}_{p-1}, \varepsilon_t, \underbrace{0, \ldots, 0}_{p-1})^{\intercal}.$$

Moreover, for any positive integers ℓ and m, let $\mathcal{I}_{\ell \times \ell}$ and $\mathbf{0}_{\ell \times m}$ be the $\ell \times \ell$ identity matrix and the $\ell \times m$ null matrix, respectively. Define A_t, B_t, C to be $p \times p$ matrices, which for $p = 1$, $A_t = \phi_i(t)$, $B_t = \phi_i(t) + \theta_i(t)$ and $C = 0$ and otherwise:

$$A_t = \left(\begin{array}{c|c} \multicolumn{2}{c}{\phi_1(t) \quad \phi_2(t) \quad \cdots \quad \phi_p(t)} \\ \hline \mathcal{I}_{(p-1)\times(p-1)} & \mathbf{0}_{(p-1)\times 1} \end{array} \right),$$

$$B_t = \left(\begin{array}{c} \phi_1(t) + \theta_1(t) \quad \phi_2(t) + \theta_2(t) \quad \cdots \quad \phi_p(t) + \theta_p(t) \\ \hline \mathbf{0}_{(p-1)\times p} \end{array} \right),$$

$$C = \left(\begin{array}{c|c} \multicolumn{2}{c}{\mathbf{0}_{1\times p}} \\ \hline \mathcal{I}_{(p-1)\times(p-1)} & \mathbf{0}_{(p-1)\times 1} \end{array} \right).$$

Hence, the $\{Y_t\}$ process satisfies the following stochastic difference equation:

$$Y_t = \begin{pmatrix} A_t & B_t \\ \mathbf{0}_{p\times p} & C \end{pmatrix} Y_{t-1} + D_t.$$

It is a Markov chain with the $2p$-th dimensional Euclidean space as the state space, and X_t is a linear function of Y_t:

$$X_t = V_t + \varepsilon_t = (1, \underbrace{0, \ldots, 0}_{p-1}, 1, \underbrace{0, \ldots, 0}_{p-1}) Y_t.$$

Set $u^{\intercal} = (1, \underbrace{0, \ldots, 0}_{p-1}, 1, \underbrace{0, \ldots, 0}_{p-1})$, since $X_t = Y_{1,t} + Y_{p+1,t} = u^{\intercal} Y_t$, it follows that

$$Y_t = \begin{cases} A_{10} + A_{11} Y_{t-1} + \varepsilon_t A_{12}, & \text{if } u^{\intercal} Y_{t-1} \le r \\ A_{20} + A_{21} Y_{t-1} + \varepsilon_t A_{22}, & \text{otherwise,} \end{cases}$$

where A_{i0} and A_{i2} are the $2p$-dimensional vectors with all components equal to zero except their first and the $p+1$-th components: $A_{i0}[1] = \phi_{i,0}$, while $A_{i2}[p+1] = 1$, where $A_{i0}[1]$ denotes the first component of A_{i0}, etc. Moreover, for any positive integers ℓ and m, let $\mathcal{I}_{\ell\times\ell}$ and $\mathbf{0}_{\ell\times m}$ be the $\ell\times\ell$ identity matrix and the $\ell\times m$ null matrix, respectively. We define the $2p\times 2p$ matrix A_{i1} as follows:

$$A_{i1} = \left(\begin{array}{cccc|cccc}
\phi_{i,1} & \phi_{i,2} & \cdots & \phi_{i,p} & (\phi_{i,1}+\theta_{i,1}) & (\phi_{i,2}+\theta_{i,2}) & \cdots & (\phi_{i,p}+\theta_{i,p}) \\
\hline
\multicolumn{1}{c|}{\mathcal{I}_{(p-1)\times(p-1)}} & \multicolumn{7}{c}{\mathbf{0}_{(p-1)\times(p+1)}} \\
\hline
\multicolumn{8}{c}{\mathbf{0}_{1\times 2p}} \\
\hline
\multicolumn{5}{c|}{\mathbf{0}_{(p-1)\times p}} & \multicolumn{2}{c|}{\mathcal{I}_{(p-1)\times(p-1)}} & \mathbf{0}_{(p-1)\times 1}
\end{array}\right).$$

For instance, if $p=4$, these matrices become

$$A_{i0} = \begin{pmatrix} \phi_{i,0} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \qquad A_{i2} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$A_{i1} = \begin{pmatrix}
\phi_{i,1} & \phi_{i,2} & \phi_{i,3} & \phi_{i,4} & (\phi_{i,1}+\theta_{i,1}) & (\phi_{i,2}+\theta_{i,2}) & (\phi_{i,3}+\theta_{i,3}) & (\phi_{i,4}+\theta_{i,4}) \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}.$$

The probabilistic analysis of the TARMA(1,1) model (with $d=1$) can be much simplified by noting that (i) $\{V_t\}$ is itself a Markov chain:

$$V_t = \begin{cases} \phi_{1,0} + \phi_{1,1}V_{t-1} + (\phi_{1,1}+\theta_{1,1})\varepsilon_{t-1}, & \text{if } V_{t-1} \le r - \varepsilon_{t-1} \\ \phi_{2,0} + \phi_{2,1}V_{t-1} + (\phi_{2,1}+\theta_{2,1})\varepsilon_{t-1}, & \text{otherwise.} \end{cases} \tag{3.3}$$

and (ii) if $\{V_t\}$ is irreducible (aperiodic, transient, null or positive recurrent, (geometrically) ergodic), so is $\{Y_t\}$, which are proved at the end of this section. Thus, for studying the probabilistic properties of the TARMA(1,1) model, it suffices to study the corresponding properties of the 1-dimensional Markov chain $\{V_t\}$. The V process differs from the well-studied TAR(1) model in that its threshold is random, being $r - \varepsilon_{t-1}$ instead of the fixed threshold r in the TAR(1) model, and non-linearly associated with the driving noise term, which underlines the difficulty in checking the irreducibility, even for the TARMA(1,1) model. In the next section, we leverage on the unit dimensionality of V_t to provide a systematic analysis of its irreducibility under (C2) and the following condition:

(C3) If $(\phi_{1,1} + \theta_{1,1})(\phi_{2,1} + \theta_{2,1}) < 0$, neither of the following three parametric conditions on the ϕ's and θ's hold: (i) $\theta_{i,1} < -1$, $i = 1, 2$; (ii) for some $1 \leq i \neq j \leq 2$, $\theta_{i,1} = -1$, $\theta_{j,1} < -1$ and $\text{sign}(\phi_{1,1} + \theta_{1,1}) \times h_i(0) \leq 0$; (iii) $\theta_{i,1} = -1$ and $\text{sign}(\phi_{1,1} + \theta_{1,1}) \times h_i(0) \leq 0, i = 1, 2$, where $h_i(0) = (\phi_{i,1} + \theta_{i,1})r + \phi_{i,0}$.

We note that condition (C3) is a mild condition, which is a necessary condition in statistical inference with TARMA(1,1) model, for the following reason. For model diagnostics or forecasting with a TARMA(1,1) model, it is necessary to (approximately) reconstruct $\varepsilon_t, t \leq n$, from the observed data $X_1, \ldots, X_n$, given the model parameter such that ε_t differs from its reconstructed version by an error of $o_p(1)$ as t and $n \to \infty$. A TARMA(1,1) model is said to be invertible if the preceding reconstruction condition holds. It can be checked that under any of the conditions (i) – (iii) in (C3), the TARMA(1,1) model is non-invertible, see [Chan and Tong, 2010]. Hence (C3) holds for any invertible TARMA(1,1) model, even though some non-invertible TARMA(1,1) models satisfy (C3), e.g., $(\phi_{1,1} + \theta_{1,1})(\phi_{2,1} + \theta_{2,1}) > 0$ and $\theta_{i,1} \geq 1, i = 1, 2$. Due to the random threshold, the proof techniques used in establishing the recurrence properties of the V process is somewhat different from those for the TAR(1) model with a deterministic threshold.

In the rest of this section, we state several results indicating that the irreducibility and recurrence/transience properties of $\{Y_t\}$ can be inferred from those of the embedded Markov chain $\{V_t\}$. Denote by F_ε the common probability measure of the iid $\{\varepsilon_t\}$. Note that assumptions (C1)–(C3) are not required for the validity of the following results.

Proposition 37. *If* $\{V_t\}$ *is irreducible, then so is* $\{Y_t\}$.

Proof. Suppose $\{V_t\}$ is μ-irreducible. Since $\{\varepsilon_t\}$ is an iid process, it is readily seen that $\{Y_t\}$ is irreducible w.r.t. $\mu \times F_\varepsilon$. □

Henceforth in this section, assume that $\{V_t\}$ is μ-irreducible and hence $\{Y_t\}$ is $\mu \times F_\varepsilon$-irreducible.

Proof. The result follows immediately from the independence of the two components of the process, V_t and ε_t, and from the assumption on the error term which has positive density on the whole real line. □

Proposition 38. *$\{V_t\}$ is recurrent (positive recurrent, null recurrent, transient) if and only if so is $\{Y_t\}$.*

Proof. We first prove the necessity part. Let $y_0 = (v_0, e_0)^\intercal$ be any initial vector. For any Borel set of the form $A = A_1 \times A_2$ with $\mu(A_1) > 0$ and $F_\varepsilon(A_2) > 0$ and any integer $t > 0$,

$$P(Y_t \in A|Y_0 = y_0) = F_\varepsilon(A_2)P(V_t \in A_1|V_0 = v_0). \tag{3.4}$$

Suppose $\{V_t\}$ is recurrent, hence it admits a non-trivial (σ-finite) invariant measure, say π. Consequently, $\pi \times F_\varepsilon$ is a non-trivial invariant measure for $\{Y_t\}$. Since $\mathbb{R} = \cup_{n\in\mathbb{Z}}[n, n+1)$, there exists an integer n such that $[n, n+1)$ has positive μ-measure. But then $\pi([n, n+1)) > 0$, which follows from (i) the formula that for any Borel set B, invariance of π implies that $(1-r)^{-1}r\pi(B) = \int G_r(x, B)\pi(dx)$, where $0 \le r < 1$, $G_r(x, B) = \sum_{t=1}^{\infty} r^t P^t(x, B)$ and $P^t(x, B)$ is the t-th step transition probability kernel of $\{V_t\}$, and (ii) irreducibility implies that for any $0 < r < 1$ and B with $\mu(B) > 0$, $G_r(x, B) > 0$, for all $x \in \mathbb{R}$. We claim that there exists an interval $[a, b) \subset [n, n+1)$ such that $0 < \pi([a, b)) < \infty$. If $\pi([n, n+1)) < \infty$, then we can take $[a, b) = [n, n+1)$. Otherwise, decompose $[n, n+1)$ into $[n, n+0.5) \cup [n+0.5, n+1)$ and if one of the two sub-intervals has finite, positive π-measure, then take it as $[a, b)$. Otherwise, one of the two sub-intervals has infinite π-measure, with which we can repeatedly carry out the splitting procedure to find a sub-interval with positive, finite π-measure. The process must end in finding such a sub-interval in a finite number of steps, otherwise we construct a decreasing sequence of nesting intervals converging to an atom of infinite π-measure, which is impossible because π is a σ-finite measure. With no loss of generality, $0 < \pi([n, n+1)) < \infty$. Then $(\pi \times F_\varepsilon)([n, n+1) \times R) = \pi([n, n+1))$ is finite and positive, entailing that $C = [n, n+1) \times R$ is a status set for $\{Y_t\}$, by [Tweedie, 1976, Proposition 5.1]. Lemma 5.1 of [Tweedie, 1976] implies that the recurrence (positive recurrence) of the chain $\{Y_t\}$ can be deduced from that of any status set, and, in particular, that of C.

Below, write $G(\cdot, \cdot)$ for $G_1(\cdot, \cdot)$. The $G(\cdot, \cdot)$ for $\{V_t\}$ will be denoted as $G_V(\cdot, \cdot)$; similarly $G_Y(\cdot, \cdot)$ for $\{Y_t\}$. Analogously defined are $P_V(\cdot, \cdot)$ and $P_Y(\cdot, \cdot)$. If $\{V_t\}$ is recurrent, then $[n, n+1)$ is a recurrent set for $\{V_t\}$, hence

$$G_V(x, [n, n+1)) \equiv \infty \text{ for all } x,$$

and so

$$G_Y((x, \varepsilon)^\intercal, [n, n+1) \times \mathbb{R}) = G_V(x, [n, n+1)) \equiv \infty, \text{ for all } (x, \varepsilon)^\intercal.$$

Therefore, C is recurrent, and hence $\{Y_t\}$ is recurrent. If $\{V_t\}$ is weakly positive, then $[n, n+1)$ is weakly positive, i.e.,

$$\limsup_{t\to\infty} P_V^t(x, [n, n+1)) > 0 \text{ for all } x,$$

but then

$$\limsup_{t\to\infty} P_Y^t((x, \varepsilon)^\intercal, [n, n+1) \times \mathbb{R}) = \limsup_{t\to\infty} P_V^t(x, [n, n+1)) > 0,$$

for all $(x, \varepsilon)^\intercal$. Therefore, C is weakly positive, hence $\{Y_t\}$ is positive recurrent.

It follows from Tweedie [1976] (Propositions 3.5 and 4.2) that if $\{V_t\}$ is null recurrent (transient), then there is a sequence $B(j) \uparrow \mathbb{R}$, each of the $B(j)'s$ is strongly null (transient), i.e., for all x, $P_V^t(x, B(j)) \to 0$, as $t \to \infty$ $(G_V(x, B(j)) < \infty, \forall x)$. But then $B(j) \times \mathbb{R} \uparrow \mathbb{R}^2$, with the $B(j) \times \mathbb{R}$ clearly strongly null (transient) for $\{Y_t\}$. Hence $\{Y_t\}$ is null recurrent (transient), [Tweedie, 1976, Lemma 5.1].

The sufficiency part is quite clear. For instance, if $\{Y_t\}$ is recurrent, $\{V_t\}$ must be recurrent, otherwise it is transient so that $\{Y_t\}$ is transient, leading to a contradiction. This completes the proof. □

Proposition 39. *$\{V_t\}$ is periodic of period $d \geq 1$ if and only if so is $\{Y_t\}$.*

Proof. Suppose $\{V_t\}$ is periodic of period d so that there exists a periodic partition, i.e., Borel sets $D_i, i = 1, \ldots, d$ such that (i) $\cup_{i=1}^d D_d$ is a partition of $\mathbb{R}$ up to a μ-null set, (ii) each D has positive μ-measure and (iii)

$$P\left(V_{t+1} \in D_{i+1 \pmod d} | V_t \in D_i\right) = 1,$$

for all $i = 1, \ldots, d$. But then

$$P\left(Y_{t+1} \in D_{i+1 \pmod d} \times \mathbb{R} | Y_t \in D_i \times \mathbb{R}\right) = 1,$$

for all $i = 1, \ldots, d$, indicating that $\{Y_t\}$ is periodic of period d. The converse is clear since $\{\varepsilon_t\}$ is iid, so that the sets of any periodic partition of the state space for $\{Y_t\}$ must comprise sets of the form $D_i \times \mathbb{R}$. □

As a corollary, $\{Y_t\}$ is aperiodic if $\{V_t\}$ is aperiodic.

Proposition 40. *$\{V_t\}$ is (geometrically) ergodic if and only if $\{Y_t\}$ is (geometrically) ergodic.*

Proof. Ergodicity of $\{V_t\}$ means that it is positive recurrent and aperiodic, so $\{Y_t\}$ is positive recurrent and aperiodic and hence ergodic, by the earlier two propositions. Geometric ergodicity quantifies that the marginal distribution of the Markov chain converges to the invariant probability measure at a geometric rate in total variation norm. Specifically, geometric ergodicity of $\{V_t\}$ means that there exist a constant $0 \leq r < 1$ and a non-negative π_Y–integrable function, say $h(\cdot)$, such that for any initial value v_0,

$$\sup_B |P(V_t \in B | V_0 = v_0) - \pi_V(B)| \leq r^t h(v_0)$$

where π_V is the invariant probability measure of $\{V_t\}$ and the supremum is taken over all Borel sets B. To prove that $\{Y_t\}$ is geometrically ergodic, it suffices to derive an analogous inequality for its invariant probability measure, say, π_Y, which we do as follows. Let $A \subseteq \mathbb{R}^2$ be any Borel set. For any $e \in \mathbb{R}$, let $A_e = \{x \in \mathbb{R} : (x, e)^\intercal \in A\}$. Then for any initial vector $y_0 = (v_0, e_0)^\intercal$,

$$\begin{aligned} & \sup_A |P(Y_t \in A | Y_0 = y_0) - \pi_Y(A)| \\ \leq & \sup_A |E\{P(V_t \in A_{\varepsilon_t} | V_0 = v_0)\} - E\{\pi_V(A_{\varepsilon_t})\}| \\ \leq & \sup_A E\{|P(V_t \in A_{\varepsilon_t} | V_0 = v_0) - \pi_V(A_{\varepsilon_t})|\} \\ \leq & \sup_B |P(V_t \in B | V_0 = v_0) - \pi_V(B)| \leq r^t h(v_0). \end{aligned}$$

As $h(\cdot)$ is clearly π_Y–integrable, we conclude that Y is geometrically ergodic. The converse of the result is trivial. □

3.2 Irreducibility of TARMA(1,1) processes

In this section, we study the irreducibility of the Markov chain $\{V_t\}$ defined by (3.3), via the concept of reachability in Control Theory [Meyn and Tweedie, 2012, Meyn, 1989, Meyn and Caines, 1989]. We first analyze the deterministic control model associated to $\{V_t\}$, where $\{\varepsilon_t\}$ is replaced by a deterministic real-valued sequence $\{u_t\}$. A key concept is the reachable set which is the collection of all states that can be reached by the deterministic system from any state via a finite sequence of inputs, i.e. the reachable set, denoted by Ω, consists of all states y such that for all state x, there exist a positive integer m and a sequence, $u_t, \ldots, u_{t+m-1}$, such that $V_t = x$ and $V_{t+m} = y$. (Here, we use the general index t to stand for a non-negative integer, even though, without loss of generality, we can set $t = 0$.) We then build a link between the irreducibility of the Markov chain $\{V_t\}$ and the reachable set of its associated deterministic chain by proving that under Assumption (C2), for each $k \in \mathbb{N}$ the k-step transition probability density of $\{V_t\}$ is positive over the set of states that can be reached by the associated deterministic chain in k steps. This ensures that if the reachable set Ω has positive Lebesgue measure, then $\{V_t\}$ is irreducible w.r.t. the Lebesgue measure restricted on Ω.

We begin with studying the set of states that are reachable in one step. Given an initial value x, the real line can be decomposed as follows:

$$\mathbb{R} = G_x \cup M_x,$$

where G_x includes the points that can be reached by the V process in one step from x and $M_x = G_x^c$, i.e.

$$G_x = \{y \in \mathbb{R} : V_t = x \quad \text{and} \quad \exists \quad \varepsilon_t \quad \text{such that} \quad V_{t+1} = y\}.$$

Now, we derive the set G_x. In order for $V_t = x$ and $V_{t+1} = y$, we exhibit a realization of the variable ε_t, i.e., u, such that *either*

(i)
$$\begin{cases} \phi_{1,0} + \phi_{1,1}x + (\phi_{1,1} + \theta_{1,1})u = y \\ x \leq r - u \end{cases}$$

	$\phi_{1,1}+\theta_{1,1}>0$	$\phi_{1,1}+\theta_{1,1}<0$
$\phi_{2,1}+\theta_{2,1}>0$	$G_x=(-\infty,h_1(x)]\cup(h_2(x),+\infty)$	$G_x=[h_1(x),+\infty)\cup(h_2(x),+\infty)$
$\phi_{2,1}+\theta_{2,1}<0$	$G_x=(-\infty,h_1(x)]\cup(-\infty,h_2(x))$	$G_x=(-\infty,h_2(x))\cup[h_1(x),+\infty)$

Table 3.1: Definition of the set G_x with respect to the signs of $\phi_{1,1}+\theta_{1,1}>0$ and $\phi_{2,1}+\theta_{2,1}>0$.

which is equivalent to

$$\begin{cases} u=\frac{y-\phi_{1,0}-\phi_{1,1}x}{\phi_{1,1}+\theta_{1,1}} \\ u\le r-x, \end{cases}$$

or

(ii)

$$\begin{cases} \phi_{2,0}+\phi_{2,1}x+(\phi_{2,1}+\theta_{2,1})u=y \\ x>r-u \end{cases}$$

which is equivalent to

$$\begin{cases} u=\frac{y-\phi_{2,0}-\phi_{2,1}x}{\phi_{2,1}+\theta_{2,1}} \\ u>r-x. \end{cases}$$

Therefore $G_x\neq\emptyset$ if and only if at least one of the following two conditions holds:

$$\frac{y-\phi_{1,0}-\phi_{1,1}x}{\phi_{1,1}+\theta_{1,1}}\le r-x;$$
$$\frac{y-\phi_{2,0}-\phi_{2,1}x}{\phi_{2,1}+\theta_{2,1}}>r-x.$$

Consequently:

$$G_x=\left\{y\in\mathbb{R}:\frac{y-\phi_{1,0}-\phi_{1,1}x}{\phi_{1,1}+\theta_{1,1}}\le r-x\quad\text{or}\quad\frac{y-\phi_{2,0}-\phi_{2,1}x}{\phi_{2,1}+\theta_{2,1}}>r-x\right\}.$$

Obviously, the set G_x assumes different forms according to the sign of $(\phi_{1,1}+\theta_{1,1})$ and $(\phi_{2,1}+\theta_{2,1})$, as shown in Table 1, where:

$$h_1(x)=(\phi_{1,1}+\theta_{1,1})r+\phi_{1,0}-\theta_{1,1}x\quad\text{and}\quad h_2(x)=(\phi_{2,1}+\theta_{2,1})r+\phi_{2,0}-\theta_{2,1}x.$$

The two cases in the (off) diagonal of the table are analogous. In fact when $\phi_{1,1}+\theta_{1,1}$ and $\phi_{2,1}+\theta_{2,1}$ have the same sign G_x is the union of two non-nesting half lines, while, in the other case, G_x is a half line. We now state a main result on the irreducibility of the two-regime TARMA(1,1) model.

Theorem 41. *Let $\{V_t\}$ be defined in (3.3). Table 3.2 delineates the reachable set of the associated deterministic system assuming that $\{\varepsilon_t\}$ were deterministic inputs, according to the parametric conditions. The reachable set is either an infinite interval or an empty set. In particular, Ω is empty if and only if $(\phi_{1,1}+$*

Conditions on parameters	Reachable set
$(\phi_{1,1}+\theta_{1,1})(\phi_{2,1}+\theta_{2,1})>0$	$\mathbb{R}$
$(\phi_{1,1}+\theta_{1,1})(\phi_{2,1}+\theta_{2,1})<0$ $\exists\, i$ such that $\theta_{i,1}>0$	$\mathbb{R}$
$\phi_{1,1}+\theta_{1,1}>0, \phi_{2,1}+\theta_{2,1}<0$ $\exists\, i$ s.t. $\theta_{i,1}=-1$ and $h_i(0)>0$	$\mathbb{R}$
$\phi_{1,1}+\theta_{1,1}<0, \phi_{2,1}+\theta_{2,1}>0$ $\exists\, i$ s.t. $\theta_{i,1}=-1$ and $h_i(0)<0$	$\mathbb{R}$
$\phi_{1,1}+\theta_{1,1}>0, \phi_{2,1}+\theta_{2,1}<0$ $\exists\, i$ s.t. $-1<\theta_{i,1}\le 0$	$(-\infty,c)$ with $c\ge \frac{h_i(0)}{1+\theta_{i,1}}$
$\phi_{1,1}+\theta_{1,1}<0, \phi_{2,1}+\theta_{2,1}>0$ $\exists\, i\ -1<\theta_{i,1}\le 0$	$(c,+\infty)$ with $c\le \frac{h_i(0)}{1+\theta_{i,1}}$
$\phi_{1,1}+\theta_{1,1}>0, \phi_{2,1}+\theta_{2,1}<0$ $\exists\, i\neq j$ s.t. $\theta_{i,1}<-1$ and $-1<\theta_{j,1}\le 0$	$\mathbb{R}$ if $\frac{h_i(0)}{1+\theta_{i,1}}<\frac{h_j(0)}{1+\theta_{j,1}}$ $\left(-\infty,\frac{h_j(0)}{1+\theta_{j,1}}\right)$ otherwise
$\phi_{1,1}+\theta_{1,1}<0, \phi_{2,1}+\theta_{2,1}>0$ $\exists\, i\neq j$ s.t. $\theta_{i,1}<-1$ and $-1<\theta_{j,1}\le 0$	$\mathbb{R}$ if $\frac{h_j(0)}{1+\theta_{j,1}}<\frac{h_i(0)}{1+\theta_{i,1}}$ $\left(\frac{h_j(0)}{1+\theta_{j,1}},+\infty\right)$ otherwise
Otherwise	$\emptyset$

Table 3.2: Reachable set according to the conditions on the parameters.

$\theta_{1,1})(\phi_{2,1}+\theta_{2,1})<0$ and it holds that either (i) $\theta_{i,1}<-1$, $i=1,2$ or (ii) for some $1\le i\neq j\le 2$, $\theta_{i,1}=-1$, $\theta_{j,1}<-1$ and sign$(\phi_{1,1}+\theta_{1,1})\times h_i(0)\le 0$ or (iii) $\theta_{i,1}=-1$ and sign$(\phi_{1,1}+\theta_{1,1})\times h_i(0)\le 0, i=1,2$. If the reachable set is non-empty, i.e., (C3) holds, and $\{\varepsilon_t\}$ are iid random variables satisfying condition (C2), then $\{V_t\}$ is irreducible w.r.t. μ, the Lebesgue measure restricted to the reachable set.

We prove the preceding theorem in the remainder of this section, as follows. Theorems 42, 43 and 44 below derive the reachable set under different parametric conditions. Propositions 46 and 45 show that the Markov chain $\{V_t\}$ is μ-irreducible where μ is the Lebesgue measure restricted to the reachable set. In the following theorems we derive the conditions on the parameters such that $\{V_t\}$ is controllable. Moreover, we compute a bound of the minimum number of steps needed to reach any target point y.

Theorem 42. *Let $\{V_t\}$ be as defined in (3.3). Suppose there exists $i\in\{1,2\}$, such that $\theta_{i,1}=0$. Then, the reachable set Ω is:*

$$\begin{array}{ll} (-\infty,h_1(0)] & \text{if } \phi_{1,1}+\theta_{1,1}>0 \text{ and } \theta_{1,1}=0;\\ [h_1(0),+\infty) & \text{if } \phi_{1,1}+\theta_{1,1}<0 \text{ and } \theta_{1,1}=0;\\ (h_2(0),+\infty) & \text{if } \phi_{2,1}+\theta_{2,1}>0 \text{ and } \theta_{2,1}=0;\\ (-\infty,h_2(0)) & \text{if } \phi_{2,1}+\theta_{2,1}<0 \text{ and } \theta_{2,1}=0. \end{array}$$

Proof. The result readily follows upon noting that $\Omega \subseteq G_x$ and it does not depend on x. ☐

Theorem 43. *Let $\{V_t\}$ be as defined in (3.3). Assume $\theta_{i,1} \neq 0$, $i \in \{1,2\}$. Then, the chain can reach any state in $\mathbb{R}$ from any state in at most two steps if either of the following two conditions holds:*

(43.1) $(\phi_{1,1} + \theta_{1,1})(\phi_{2,1} + \theta_{2,1}) > 0$;

(43.2) $(\phi_{1,1} + \theta_{1,1})(\phi_{2,1} + \theta_{2,1}) < 0$, *and there exists* $i \in \{1,2\}$ *s.t.* $\theta_{i,1} > 0$.

Proof. It suffices to prove that, for any $x \in \mathbb{R}$, every z in M_x can be reached by the chain from x in one step from a point of G_x, i.e., there exist w and y satisfying at least one of the following two systems of equations.

$$\begin{cases} \phi_{1,0} + \phi_{1,1}y + (\phi_{1,1} + \theta_{1,1})w = z \\ w \leq r - y \\ y \in G_x; \end{cases} \tag{3.5}$$

$$\begin{cases} \phi_{2,0} + \phi_{2,1}y + (\phi_{2,1} + \theta_{2,1})w = z \\ w > r - y \\ y \in G_x. \end{cases} \tag{3.6}$$

Suppose (43.1) holds and assume $\phi_{1,1} + \theta_{1,1} > 0$ and $\phi_{2,1} + \theta_{2,1} > 0$. From Table 1, if $h_2(x) \leq h_1(x)$ then the set M_x is empty hence the whole space can be covered in one step, otherwise

$$M_x = \{z \in \mathbb{R} : h_1(x) < z \leq h_2(x)\}.$$

Since $\theta_{1,1} \neq 0$, we claim that for every z in M_x, there exist w and y satisfying (3.5), which can be seen as follows. Let then $z \in M_x$ be fixed. Indeed, for fixed x and $z \in M_x$, if (3.5) admits a solution pair (y, w), routine algebra yields

$$z \leq \phi_{1,0} + (\phi_{1,1} + \theta_{1,1})r - \theta_{1,1}y. \tag{3.7}$$

On the other hand, if the preceding equation admits a solution $y \in G_x$, then the positivity of $\phi_{1,1} + \theta_{1,1}$ implies the existence of $w \in \mathbb{R}$ such that (y, w) is a solution of (3.5), on noting that the right side of (3.7) can be written as $\phi_{1,0} + \phi_{1,1}y + (\phi_{1,1} + \theta_{1,1})(r - y)$. Because $\theta_{1,1} \neq 0$ and G_x contains all states of sufficiently large magnitude, (3.7) always admits a solution $y \in G_x$, hence the claim is established. On the other hand, if $\theta_{2,1} \neq 0$, it can similarly shown that (3.6) admits a solution (y, w) iff the following equation has a solution $y \in G_x$:

$$z > \phi_{2,0} + (\phi_{2,1} + \theta_{2,1})r - \theta_{2,1}y. \tag{3.8}$$

Clearly, (3.8) always admits a solution $y \in G_x$. The proof for the case that $\phi_{1,1} + \theta_{1,1} < 0$ and $\phi_{2,1} + \theta_{2,1} < 0$ is similar and hence omitted. Next suppose condition (43.2) holds. We outline the proof under the condition that $\phi_{1,1} + \theta_{1,1} > 0$, $\phi_{2,1} + \theta_{2,1} < 0$ and there exists $i \in \{1,2\}$ such that $\theta_{i,1} > 0$, as the proof for the other case is similar. Given x, without loss of generality we can suppose $h_1(x) > h_2(x)$, hence

$$M_x = \{z \in \mathbb{R} : z > h_1(x)\}.$$

If $\theta_{1,1} > 0$ then (3.5) always admits a solution because (3.7) admits a solution $y \in G_x$. If $\theta_{2,1} > 0$ then (3.6) always admits a solution because the inequality

$$z < \phi_{2,0} + (\phi_{2,1} + \theta_{2,1})r - \theta_{2,1}y \tag{3.9}$$

always admits a solution $y \in G_x$. The proof for the case that $\phi_{1,1} + \theta_{1,1} < 0$, $\phi_{2,1} + \theta_{2,1} > 0$ and there exists $i \in \{1, 2\}$ such that $\theta_{i,1} > 0$ is similar and omitted. □

Under the hypothesis of the previous theorem, we have proved that the chain can reach from anywhere every point in the real line in at most two steps, so the reachable set is the entire real line. This implies that, under additional conditions, $\{V_t\}$ is irreducible w.r.t the Lebesgue measure on $\mathbb{R}$. Now, we derive the reachable set under the condition that $(\phi_{1,1} + \theta_{1,1})(\phi_{2,1} + \theta_{2,1}) < 0$ and $\theta_{i,1} < 0$, for each $i \in \{1, 2\}$. Without loss of generality, assume $\phi_{1,1} + \theta_{1,1} > 0$ and $\phi_{2,1} + \theta_{2,1} < 0$, otherwise we will consider $\{-V_t\}$. In this case, we show, below, that either the chain can reach the whole real line but in a greater and variable number of steps or the reachable set is restricted to the half line

$$(-\infty, c), \tag{3.10}$$

where $c \in \mathbb{R}$. Given two points x and z, let $m = m(x, z)$ be the number of steps needed for the chain to go from x to z. In the following theorem, we derive the conditions on the parameters such that $\{V_t\}$ can reach any point in $(-\infty, c)$ or $(c, +\infty)$ from any initial value. Moreover, we compute the value of c and an upper bound of m.

Theorem 44. *Let $\{V_t\}$ be defined by (3.3) with $(\phi_{1,1}+\theta_{1,1}) > 0$, $(\phi_{2,1}+\theta_{2,1}) < 0$ and $\theta_{i,1} < 0$, for each $i \in \{1, 2\}$.*

(44.1) Suppose there exists $i \in \{1, 2\}$ such that $\theta_{i,1} = -1$ and $h_i(0) > 0$. Then the reachable set is $(-\infty, +\infty)$.

(44.2) Suppose $\theta_{i,1} \in (-1, 0)$, for some $i = 1, 2$. Then the reachable set is $(-\infty, c)$ where

$$c \geq \frac{h_i(0)}{1 + \theta_{i,1}}.$$

(44.3) Suppose there exist $1 \leq i \neq j \leq 2$ such that $\theta_{i,1} \in (-\infty, -1)$ and $\theta_{j,1} \in (-1, 0)$. Then the reachable set is $(-\infty, c)$ where

$$c = \begin{cases} +\infty, & \text{if } \frac{h_i(0)}{1+\theta_{i,1}} < \frac{h_j(0)}{1+\theta_{j,1}} \\ \frac{h_j(0)}{1+\theta_{j,1}} & \end{cases}$$

(44.4) Otherwise, the reachable set is an empty set.

Proof. Recall G_x denotes the states which the chain can reach in 1 step from x. Under the conditions that $(\phi_{1,1}+\theta_{1,1})>0$, $(\phi_{2,1}+\theta_{2,1})<0$ and $\theta_{i,1}<0$, for each $i\in\{1,2\}$, Table 3.1 indicates that G_x is an one-sided infinite interval with the finite right end-point equal to $c(x)=\max\{h_1(x),h_2(x)\}$. Clearly, for each x, $c(x)\in G_x$ if and only if $h_1(x)\geq h_2(x)$. If $c(x)>x$, the chain may potentially reach any state from any initial state. Thus, it is useful to define the following functions:

$$d_i(x)=h_i(x)-x=(\phi_{i,1}+\theta_{i,1})r+\phi_{i,0}-(1+\theta_{i,1})x, \qquad i\text{=1,2}.$$

If $d_i(x)>0$, it represents the maximum distance the chain can move in one step to its right hand side via regime i. If $d_i(x)<0$, for all i, then the chain can not move to the right side of its initial value. Now we divide the proof in different cases, (44.1)-(44.3), as stated in the theorem. In each of these we derive the reachable set to be of the form $(-\infty,c)$. Moreover, we prove that for each x and for each $b<c$ there exist an integer $m=m(x,b)$ and a sequence of values $u_t,u_{t+1},\ldots,u_{t+m-1}$ such that $V_{t+m}=b$, given $V_t=x$.

Case (44.1). In this case we have that $d_i(x)=d_i(0)>0$, for all x. It is clear that, starting from any initial state, $\{V_t\}$ can reach, in one step, any state on its left side and reach any state on its right side with distance not more than $d_i(0)$. Hence, it can reach any state b on its right side in m steps where

$$m=\left\lfloor\frac{b-x}{d_i(0)}\right\rfloor+1.$$

Case (44.2). In this case $d_i(x)>0$ if and only if $x<\frac{h_i(0)}{1+\theta_{i,1}}$. Indeed, starting from $x<\frac{h_i(0)}{1+\theta_{i,1}}$, the chain can reach any point $b<\frac{h_i(0)}{1+\theta_{i,1}}$ in m steps, where

$$m\leq\left\lfloor\frac{b-x}{d_i(b)}\right\rfloor+1.$$

On the other hand if $x\geq\frac{h_i(0)}{1+\theta_{i,1}}$, then there exists a point $y\in G_x\cap(-\infty,b)$. Hence $\{V_t\}$ can reach any point $b<\frac{h_i(0)}{1+\theta_{i,1}}$ as before. The nature of the reachable set is fully determined by the value $\theta_{j,1}, j\neq i$. For $\theta_{j,1}\in(0,-1)$, the preceding reasoning entails that $c=\max_{i=1,2}\frac{h_i(0)}{1+\theta_{i,1}}$. If $\theta_{j,1}=-1$ and $h_j(0)>0$ then $c=\infty$, from Case (44.1). But if $h_j(0)\leq 0$, then $c=\frac{h_i(0)}{1+\theta_{i,1}}$. The case of $\theta_{j,1}<-1$ is covered below.

Case (44.3). In this case, the functions $d_i(x)$ and $d_j(x)$ have the opposite slope: $d_i(x)$ is an increasing function, while $d_j(x)$ is a decreasing function. If $\frac{(\phi_{i,1}+\theta_{i,1})r+\phi_{i,0}}{1+\theta_{i,1}}<\frac{(\phi_{j,1}+\theta_{j,1})r+\phi_{j,0}}{1+\theta_{j,1}}$ then there exists $\eta>0$, such that $\max\{d_1(x),d_2(x)\}>\eta$. Hence, the result follows from using a similar argument for proving Case (44.1). Otherwise, from any starting point, $\{V_t\}$ can reach any point $b<\frac{h_j(0)}{1+\theta_{j,1}}$, by using a similar argument for Case (44.2).

Case (44.4) follows readily from arguments laid out in the previous three cases. □

Proposition 45. *Let $\{V_t\}$ be defined by (3.3). Suppose condition (C2) holds. Then, for each $x \in \mathbb{R}$ and $y \in G_x$, the 1-step transition probability density of $\{V_t\}$ is given by*

$$
\begin{aligned}
p(x,y) &= \frac{\rho\left(\frac{y-\phi_{1,1}x-\phi_{1,0}}{\phi_{1,1}+\theta_{1,1}}\right)}{|\phi_{1,1}+\theta_{1,1}|} I_{G_x^1}(y) + \left[\frac{\rho\left(\frac{y-\phi_{1,1}x-\phi_{1,0}}{\phi_{1,1}+\theta_{1,1}}\right)}{|\phi_{1,1}+\theta_{1,1}|} + \frac{\rho\left(\frac{y-\phi_{2,1}x-\phi_{2,0}}{\phi_{2,1}+\theta_{2,1}}\right)}{|\phi_{2,1}+\theta_{2,1}|}\right] I_{G_x^2}(y) \\
&\quad + \frac{\rho\left(\frac{y-\phi_{2,1}x-\phi_{2,0}}{\phi_{2,1}+\theta_{2,1}}\right)}{|\phi_{2,1}+\theta_{2,1}|} I_{G_x^3}(y) \\
&= \frac{\rho\left(\frac{y-\phi_{1,1}x-\phi_{1,0}}{\phi_{1,1}+\theta_{1,1}}\right)}{|\phi_{1,1}+\theta_{1,1}|} I_{G_x^1 \cup G_x^2}(y) + \frac{\rho\left(\frac{y-\phi_{2,1}x-\phi_{2,0}}{\phi_{2,1}+\theta_{2,1}}\right)}{|\phi_{2,1}+\theta_{2,1}|} I_{G_x^1 \cup G_x^3}(y),
\end{aligned}
$$

where I is the indicator function and G_x^i with $i = 1,2,3$, defined below, is a partition of G_x:

$$
\begin{aligned}
G_x^1 &= \left\{ y \in \mathbb{R} : \frac{y-\phi_{1,1}x-\phi_{1,0}}{\phi_{1,1}+\theta_{1,1}} \le r-x, \frac{y-\phi_{2,1}x-\phi_{2,0}}{\phi_{2,1}+\theta_{2,1}} \le r-x \right\}; \\
G_x^2 &= \left\{ y \in \mathbb{R} : \frac{y-\phi_{1,1}x-\phi_{1,0}}{\phi_{1,1}+\theta_{1,1}} \le r-x, \frac{y-\phi_{2,1}x-\phi_{2,0}}{\phi_{2,1}+\theta_{2,1}} > r-x \right\}; \qquad (3.11) \\
G_x^3 &= \left\{ y \in \mathbb{R} : \frac{y-\phi_{1,1}x-\phi_{1,0}}{\phi_{1,1}+\theta_{1,1}} > r-x, \frac{y-\phi_{2,1}x-\phi_{2,0}}{\phi_{2,1}+\theta_{2,1}} > r-x \right\}.
\end{aligned}
$$

Proof. The theorem is proved only when $\phi_{1,1}+\theta_{1,1} > 0$ and $\phi_{2,1}+\theta_{2,1} > 0$, since in the other cases the argument is similar. It is clear that $p(x,y) = 0$, for each $y \in G_x^c$. Therefore, given $y \in G_x$, we compute $P(V_{t+1} \le y \mid V_t = x)$ and, hence, derive the first step transition pdf $p(x,y)$.

$$
\begin{aligned}
P(V_{t+1} \le y \mid V_t = x) &= P\left(\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon_t \le y, \varepsilon_t \le r-x\right) \\
&\quad + P\left(\phi_{2,0}+\phi_{2,1}x+(\phi_{2,1}+\theta_{2,1})\varepsilon_t \le y, \varepsilon_t > r-x\right) \\
&= P\left(\varepsilon_t \le \frac{y-\phi_{1,1}x-\phi_{1,0}}{\phi_{1,1}+\theta_{1,1}}, \varepsilon_t \le r-x\right) \\
&\quad + P\left(\varepsilon_t \le \frac{y-\phi_{2,1}x-\phi_{2,0}}{\phi_{2,1}+\theta_{2,1}}, \varepsilon_t > r-x\right).
\end{aligned}
$$

According to the position of $\frac{y-\phi_{1,1}x-\phi_{1,0}}{\phi_{1,1}+\theta_{1,1}}$ and $\frac{y-\phi_{2,1}x-\phi_{2,0}}{\phi_{2,1}+\theta_{2,1}}$ w.r.t. $r-x$, the set G_x is decomposed into three disjoint sets:

$$
G_x = G_x^1 \cup G_x^2 \cup G_x^3,
$$

where $G_x^i, i = 1,2,3$ are defined in (3.11).

For $y \in G_x^1$, we have

$$\begin{aligned} P(V_{t+1} \le y \mid V_t = x) &= P\left(\varepsilon_t \le \frac{y - \phi_{1,1}x - \phi_{1,0}}{\phi_{1,1} + \theta_{1,1}}\right) \\ &= \int_{-\infty}^{\frac{y-\phi_{1,1}x-\phi_{1,0}}{\phi_{1,1}+\theta_{1,1}}} \rho(u)du. \end{aligned}$$

Consequently,

$$\begin{aligned} p(x,y) &= \rho\left(\frac{y - \phi_{1,1}x - \phi_{1,0}}{\phi_{1,1} + \theta_{1,1}}\right) \cdot \frac{\partial\left(\frac{y-\phi_{1,1}x-\phi_{1,0}}{\phi_{1,1}+\theta_{1,1}}\right)}{\partial y} \\ &= \frac{\rho\left(\frac{y-\phi_{1,1}x-\phi_{1,0}}{\phi_{1,1}+\theta_{1,1}}\right)}{\phi_{1,1} + \theta_{1,1}}. \end{aligned}$$

Similarly,

$$p(x,y) = \begin{cases} \frac{\rho\left(\frac{y-\phi_{1,1}x-\phi_{1,0}}{\phi_{1,1}+\theta_{1,1}}\right)}{\phi_{1,1}+\theta_{1,1}} + \frac{\rho\left(\frac{y-\phi_{2,1}x-\phi_{2,0}}{\phi_{2,1}+\theta_{2,1}}\right)}{\phi_{2,1}+\theta_{2,1}}, & y \in G_x^2 \\ \frac{\rho\left(\frac{y-\phi_{2,1}x-\phi_{2,0}}{\phi_{2,1}+\theta_{2,1}}\right)}{\phi_{2,1}+\theta_{2,1}}, & y \in G_x^3. \end{cases}$$

Therefore, the proposition follows from summing the 1-step probability densities over all these cases. □

For each integer $k \ge 2$ the k-step probability density of V_t is

$$p^k(x,z) = \int_{G_{x,k-1}} p^{k-1}(x,y)p(y,z)dy,$$

for $z \in G_{x,k}$ and zero otherwise, where $G_{x,k}$ is the set reachable from x in k steps and $G_{x,1} = G_x$.

Corollary 46. *Assume (C2) holds. For each integer $k \ge 1$ and each $x \in \mathbb{R}$, the k-step probability density of $\{V_t\}$ is positive on the set of states reachable in k steps from each $x \in \mathbb{R}$.*

Proof. The claim follows immediately by mathematical induction and since the density $\rho(\cdot)$ is positive over $\mathbb{R}$. □

3.3 Aperiodicity of TARMA(1,1) processes

Proposition 47. *Let $\{V_t\}$ be defined by (3.3). Suppose (C2) holds and $\{V_t\}$ is μ-irreducible, where μ is the Lebesgue measure restricted to the reachable set. Then $\{V_t\}$ is aperiodic.*

Proof. We prove by contradiction. Suppose the chain is periodic, with period $d \ge 2$, then there exists a d-cycle of μ-positive sets $D_1, \ldots, D_d$ such that:

(i) $D_i \cap D_j = \emptyset$, for each $i \neq j$;

(ii) $N = \left[\bigcup_{i=1}^{d} D_i\right]^c$ has null μ-measure;

(iii) $P(x, D_{i+1}) = 1$, for each $x \in D_i$, with $i = 1, \dots, d \mod(d)$.

Suppose $x \in D_1$. From (iii), $\mu\left(G_x \cap D_2\right) > 0$ and $\mu\left(G_x \cap D_i\right) = 0$, for each $i \neq 2$. Now consider $y \in D_2$; as before we have $\mu\left(G_y \cap D_3\right) > 0$ and $\mu\left(G_y \cap D_i\right) = 0$, for each $i \neq 3 \mod(d)$. From Table 1, $G_x \cap G_y$ is either a half line or the union of two half lines, which has non-empty intersection with the reachable set. Therefore, $\mu\left(G_x \cap G_y\right) > 0$. Furthermore,

$$\mu\left(G_x \cap G_y\right) = \mu\left(\left(G_x \cap G_y\right) \cap \left[\bigcup_{i=1}^{d} D_i \cup N\right]\right) = \sum_{i=1}^{d} \mu\left(G_x \cap G_y \cap D_i\right) > 0.$$

Hence, there exists an integer $i \in \{1, \dots, d\}$, such that $\mu\left(G_x \cap G_y \cap D_i\right) > 0$. This is a contradiction since $\mu\left(G_x \cap D_i\right) = 0$, for each $i \neq 2$ and $\mu\left(G_y \cap D_i\right) = 0$, for each $i \neq 3 \mod(d)$, therefore the result is proved. □

3.4 Recurrence properties of the TARMA(1,1) processes

In this section we provide a complete parametric classification of the long-run probabilistic behaviour of $\{Y_t\}$ for the 2-regime TARMA(1,1) model with unit delay, i.e., $d = 1$. Specifically, assuming (C1)–(C3), we provide the parametric condition for recurrence versus transience, and finer sub-division of recurrence into null recurrence versus positive recurrence and even geometric ergodicity. Such a complete classification is enabled by the fact that the univariate time series $\{V_t\}$ is a one-dimensional Markov chain and that it resembles the standard TAR(1) model except for having a random threshold $r - \varepsilon_{t-1}$ instead of a fixed threshold r. In particular, under conditions (C1)–(C3), we show that the long-run probabilistic behaviour (transience versus recurrence etc.) of a TARMA(1,1) process driven by (3.3) is same as that of the associated TAR(1) process $\{\tilde{V}_t\}$ satisfying the following equation:

$$\tilde{V}_t = \begin{cases} \phi_{1,0} + \phi_{1,1}\tilde{V}_{t-1} + (\phi_{1,1} + \theta_{1,1})\varepsilon_{t-1}, & \text{if } \tilde{V}_{t-1} \leq r \\ \phi_{2,0} + \phi_{2,1}\tilde{V}_{t-1} + (\phi_{2,1} + \theta_{2,1})\varepsilon_{t-1}, & \text{otherwise.} \end{cases} \tag{3.12}$$

The classification is proved mainly via the general drift criteria for the recurrence classification of a continuous-state-space Markov chain [Tweedie, 1976]. Our general approach is to verify the drift criteria for $\{V_t\}$ with a test (generalized energy) function that works for $\{\tilde{V}_t\}$, and demonstrate that the random threshold has a negligible impact on the drift criteria. In several cases noted below, we develop new test functions that makes it easier to verify the relevant drift criteria.

Theorem 48. *Let* $\{V_t\}$ *defined by (3.3). Suppose (C1), with* $k \geq 1$*, (C2) and (C3) hold. Then the following characterization holds.*
Part 1. $\{V_t\}$ *is geometrically ergodic if it satisfies (48.1):*

(48.1) $\phi_{1,1}\phi_{2,1} < 1,\ \phi_{1,1} < 1,\ \phi_{2,1} < 1.$

Part 2. $\{V_t\}$ *is ergodic if and only if it satisfies one of (48.2)-(48.6):*

(48.2) $\phi_{1,1}\phi_{2,1} < 1,\ \phi_{1,1} < 1,\ \phi_{2,1} < 1;$

(48.3) $\phi_{1,1} = 1,\ \phi_{2,1} < 1,\ \phi_{1,0} > 0;$

(48.4) $\phi_{1,1} < 1,\ \phi_{2,1} = 1,\ \phi_{2,0} < 0;$

(48.5) $\phi_{1,1} = 1,\ \phi_{2,1} = 1,\ \phi_{2,0} < 0 < \phi_{1,0};$

(48.6) $\phi_{1,1} < 0,\ \phi_{1,1}\phi_{2,1} = 1,\ \phi_{2,0} + \phi_{2,1}\phi_{1,0} > 0.$

Part 3. *If furthermore there exists* $k > 2$ *such that (C1) holds then* $\{V_t\}$ *is null recurrent if and only if it satisfies one of (48.7)-(48.11):*

(48.7) $\phi_{1,1} = 1,\ \phi_{2,1} = 1,\ \phi_{2,0} = 0,\ \phi_{1,0} \geq 0;$

(48.8) $\phi_{1,1} = 1,\ \phi_{2,1} = 1,\ \phi_{2,0} < 0,\ \phi_{1,0} = 0;$

(48.9) $\phi_{1,1} < 1,\ \phi_{2,1} = 1,\ \phi_{2,0} = 0;$

(48.10) $\phi_{1,1} = 1,\ \phi_{2,1} < 1,\ \phi_{1,0} = 0;$

(48.11) $\phi_{1,1} < 0,\ \phi_{1,1}\phi_{2,1} = 1,\ \phi_{2,0} + \phi_{2,1}\phi_{1,0} = 0.$

Part 4. *If furthermore there exists* $k \geq 2$ *such that (C1) holds then* $\{V_t\}$ *is transient if and only if it satisfies one of (48.12)-(48.17):*

(48.12) $\phi_{1,1} > 1;$

(48.13) $\phi_{2,1} > 1;$

(48.14) $\phi_{1,1} < 0,\ \phi_{1,1}\phi_{2,1} > 1;$

(48.15) $\phi_{1,1} < 0,\ \phi_{1,1}\phi_{2,1} = 1,\ \phi_{2,0} + \phi_{2,1}\phi_{1,0} < 0;$

(48.16) $\phi_{1,1} \leq 1,\ \phi_{2,1} = 1,\ \phi_{2,0} > 0;$

(48.17) $\phi_{1,1} = 1,\ \phi_{2,1} \leq 1,\ \phi_{1,0} < 0.$

We remark that ergodicity of $\{V_t\}$ implies that of $\{Y_t\}$ and hence they admit a unique stationary distribution. In particular, under (C1)–(C3), (48.2)–(48.6) are the necessary and sufficient conditions for the stationarity of the TARMA(1,1) model. Moreover, under the conditions imposed in Part 1 of Theorem 48, the stationary distribution of $\{V_t\}$ has finite k-th absolute moments and so has X_t [see Tweedie, 1983a]. The Central Limit Theorem is applicable under some further conditions, see Chan [1993] for details. Before providing the proof of the preceding classification results, we establish some preliminary lemmas. Note that, with no loss of generality, the threshold can and will be assumed to be zero. Otherwise, we can consider $\{X_t - r\}$ which is a TARMA(1,1) process with zero threshold, and all the parametric conditions stated in Theorem 48 are shift invariant. We first prove that any compact set $[-M, M]$ of sufficiently large M is a small set, which is required in verifying the aforementioned drift criteria.

Lemma 49. *Suppose the conditions stated in Theorem 48 hold. Then, every set of the form $[-M, M]$ is a small set, where $M > 0$ is arbitrary if the support of the irreducibility measure μ is the real line, and otherwise greater than $|c|$ where c is listed in Table 3.2.*

Proof. The conditions on M implies that $[-M, M]$ has positive μ-measure. Its smallness then follows from [Nummelin, 2004, Proposition 2.11], if we find a set D, with $\mu(D) > 0$ and a non-negative integer $L < +\infty$ such that:

$$\inf_{x\in[-M,M]} \sum_{n=0}^{L} P^n(x, C) > 0, \qquad \text{for each } C \subseteq D, \text{ with } \mu(C) > 0.$$

We consider different cases according to the hypotheses on the parameters. In each case, we show that there exists a set D such that $P^1(x, C) \geq \eta > 0$, for each $C \subseteq D$ with $\mu(C) > 0$. We analyze only the case in which the irreducible measure is the Lebesgue measure on $\mathbb{R}$. Otherwise, we consider the intersection between D and the reachable set defined in Table 2. The proof is divided into two cases according to the sign of $\phi_{1,1} + \theta_{1,1}$.

Case 1. Assume $\phi_{1,1} + \theta_{1,1} > 0$ and define the set $D = (-\infty, d)$, with

$$d = \begin{cases} h_1(+M), & \text{if } \theta_{1,1} > 0 \\ h_1(-M), & \text{if } \theta_{1,1} < 0 \\ h_1(0), & \text{if } \theta_{1,1} = 0. \end{cases}$$

It is clear that $D \subset (-\infty, h_1(x)] \subset G_x$, for each $x \in [-M, M]$. Now, let $x \in [-M, M]$ and $C \subseteq D$ with $\mu(C) > 0$. Since C has a positive Lebesgue measure then there exists an interval $[a, b]$, such that $\mu([a, b] \cap C) > 0$. We claim that there exists $\delta > 0$ such that $p(x, y) > \delta$, for each $y \in [a, b]$. Assuming the validity of the claim, we have:

$$\begin{aligned} P^1(x, C) &= \int_{G_x \cap C} p(x, y) dy = \int_C p(x, y) dy \\ &> \int_{[a,b]\cap C} \delta dy = \delta\mu([a, b] \cap C) = \eta > 0. \end{aligned}$$

Now, we verify the claim. for each $y \in (-\infty, h_1(x)]$, it holds that

$$p(x,y) = \begin{cases} \frac{\rho\left(\frac{y-\phi_{1,1}x-\phi_{1,0}}{\phi_{1,1}+\theta_{1,1}}\right)}{\phi_{1,1}+\theta_{1,1}}, & \text{if } h_1(x) < h_2(x) \\ \frac{\rho\left(\frac{y-\phi_{1,1}x-\phi_{1,0}}{\phi_{1,1}+\theta_{1,1}}\right)}{\phi_{1,1}+\theta_{1,1}} + \frac{\rho\left(\frac{y-\phi_{2,1}x-\phi_{2,0}}{\phi_{2,1}+\theta_{2,1}}\right)}{\phi_{2,1}+\theta_{2,1}} I_{(h_2(x),h_1(x)]}(y), & \text{if } h_1(x) \geq h_2(x) \end{cases}$$

$$\geq p^\star(x,y) = \frac{\rho\left(\frac{y-\phi_{1,1}x-\phi_{1,0}}{\phi_{1,1}+\theta_{1,1}}\right)}{\phi_{1,1}+\theta_{1,1}},$$

hence the validity of the claim since $p^\star(x,y)$, as a function of y, is continuous and positive over G_x, therefore attaining a positive minimum over $[a,b]$.

Case 2. Assume $\phi_{1,1} + \theta_{1,1} < 0$. The argument is the same of Case 1, where $D = (d, +\infty)$, with

$$d = \begin{cases} h_1(-M), & \text{if } \theta_{1,1} > 0 \\ h_1(+M), & \text{if } \theta_{1,1} < 0 \\ h_1(0), & \text{if } \theta_{1,1} = 0. \end{cases}$$

□

Next, we present three lemmas that furnish technical tools useful for verifying the drift criteria, with the first two essentially being Lemmas 1 and 2 in Guo and Petruccelli [1991], and hence their proofs are omitted.

Lemma 50. *Let η be a random variable, s any positive number and t any real number. If furthermore η has finite second moment, then for any events $A \subseteq \{s + t\eta > 0\}$ and $B \subseteq \{-s + t\eta > 0\}$,*

$$\begin{aligned} E[\ln(s+t\eta)I_A] \leq \quad & P(A)\ln(s) + (t/s)E[\eta I_A] \\ & -\{t^2/(2s^2)\}E[\eta^2 I_{\{A\cap\{t\eta<0\}\}}]. \end{aligned} \tag{3.13}$$

If η admits finite first moment,

$$E[\ln(-s+t\eta)I_B] \leq P(B)\{\ln(s) - 2\} + (t/s)E[\eta I_B]. \tag{3.14}$$

Lemma 51. *Let η be a random variable with distribution function G, $E(\eta) = 0$ and $E(\eta^2) < +\infty$. Moreover, let t, c, u_2 and v_2 be positive numbers, and let $s_1 \geq s_2$ and u_1, v_1, s be real numbers. Then the following hold*

$$\lim_{x\to-\infty} xE[\eta I_{\{\eta<s+tx\}}] = \lim_{x\to+\infty} xE[\eta I_{\{\eta>s+tx]\}}] = 0;$$

$$\lim_{x\to-\infty} xE[\eta I_{\{\eta>s+tx\}}] = \lim_{x\to+\infty} xE[\eta I_{\{\eta<s+tx\}}] = 0;$$

$$\limsup_{x\to-\infty} x^2[-G(s_1+tx)\ln(u_1-u_2x) + G(s_2+tx)\{\ln(v_1-v_2x)-c)\}] \leq 0;$$

$$\limsup_{x\to+\infty} x^2[-\{1-G(s_2+tx)\}\ln(v_1+v_2x) + \{1-G(s_1+tx)\}\{\ln(u_1+u_2x)-c\}] \leq 0.$$

We also need the following technical result to prove that the random threshold has a negligible impact on the drift criteria.

Lemma 52. *Let η be a random variable with $E(\eta) = 0$ and $E(|\eta|^k) < \infty$ for some $k > 0$. Let t be any positive number and let s_1 and s_2 be any real numbers. Then it holds that*

$$\lim_{x\to+\infty} x^k P(s_1 + tx + s_2\eta < 0) = 0; \tag{3.15}$$

$$\lim_{x\to+\infty} x^k P(s_1 - tx + s_2\eta > 0) = 0. \tag{3.16}$$

Proof. The results are trivial if $s_2 = 0$. So, without loss of generality, assume $s_2 \neq 0$. Equation (3.15) holds because

$$\begin{aligned}
0 &\le \liminf_{x\to+\infty} x^k P(s_1 + tx + s_2\eta < 0) \\
&\le \limsup_{x\to+\infty} x^k P(s_1 + tx + s_2\eta < 0) \\
&\le \lim_{x\to+\infty} E\left[\{-s_1/t - (s_2/t)\eta\}^k I_{\{s_1+tx+s_2\eta<0\}}\right] \\
&= \begin{array}{ll} \lim_{x\to+\infty} E\left[\{-s_1/t - (s_2/t)\eta\}^k I_{\{\eta<-s_1/s_2-(t/s_2)x\}}\right], & \text{if } s_2 > 0 \\ \lim_{x\to+\infty} E\left[\{-s_1/t - (s_2/t)\eta\}^k I_{\{\eta>-s_1/s_2-(t/s_2)x\}}\right], & \text{if } s_2 < 0. \end{array}
\end{aligned}$$

These latter limits exist and they are equal to zero, thanks to the Lebesgue dominated convergence theorem. Eqn (3.16) can be similarly proved. □

Proof of Part 1 of Theorem 48.

From Lemma 49, every set of the form $[-M, M]$ for all sufficiently large M is a small set. Hence, the result follows by verifying the drift criterion in Tweedie [1983b], Theorem 3, i.e., there exist a function $g(\cdot) : \mathbb{R} \to [1, +\infty)$ and some constants $M, K, \gamma > 0$, such that $[-M, M]$ is a small set and

$$\begin{array}{lll} (i) & E[g(V_{t+1})|V_t = x] \le (1-\gamma)g(x), & \text{for all } x \in [-M, M]^c; \\ (ii) & E[g(V_{t+1})|V_t = x] \le K, & \text{for all } x \in [-M, M]. \end{array}$$

It can be seen in the proof that M can be chosen to be arbitrarily large, hence $[-M, M]$ is small. The test function $g(\cdot)$ is of the following form

$$g(x) = \begin{cases} b^k|x|^k + 1, & x < 0 \\ a^k x^k + 1, & x \ge 0 \end{cases}$$

with a and b positive constants listed in Table 3.3. Conditions (i) and (ii) are satisfied with V_t there replaced by $\tilde{V}_t$, [Chan et al., 1985, Theorem 2.3]. We claim that

$$E[g(V_{t+1})|V_t = x] \le E[g(\tilde{V}_{t+1})|\tilde{V}_t = x] + \tau(x) \tag{3.17}$$

where $\tau(x)$ is a bounded function and $\tau(x) = o(1)$ as $|x| \to \infty$, which implies that conditions (i) and (ii) hold with a possibly smaller $\gamma > 0$ and a possibly larger $K > 0$. Thus it remains to verify (3.17). For simplicity, we do so for the case that $0 \leq \phi_{1,1} < 1$, $\phi_{2,1} \leq -1$, as the proof is similar for other cases. Consider

$$\begin{aligned}
E[g(V_{t+1})|V_t = x] = & \int_{-\infty}^{-x} \left[a^k \{\phi_{1,0} + \phi_{1,1}x + (\phi_{1,1} + \theta_{1,1})\varepsilon\}^k + 1\right] \\
& \times I_{\{\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon \geq 0\}} \rho(\varepsilon) d\varepsilon \\
& + \int_{-x}^{+\infty} \left[a^k \{\phi_{2,0} + \phi_{2,1}x + (\phi_{2,1} + \theta_{2,1})\varepsilon\}^k + 1\right] \\
& \times I_{\{\phi_{2,0}+\phi_{2,1}x+(\phi_{2,1}+\theta_{2,1})\varepsilon \geq 0\}} \rho(\varepsilon) d\varepsilon \\
& + \int_{-\infty}^{-x} \left[b^k |\phi_{1,0} + \phi_{1,1}x + (\phi_{1,1} + \theta_{1,1})\varepsilon|^k + 1\right] \\
& \times I_{\{\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon < 0\}} \rho(\varepsilon) d\varepsilon \\
& + \int_{-x}^{+\infty} \left[b^k |\phi_{2,0} + \phi_{2,1}x + (\phi_{2,1} + \theta_{2,1})\varepsilon|^k + 1\right] \\
& \times I_{\{\phi_{2,0}+\phi_{2,1}x+(\phi_{2,1}+\theta_{2,1})\varepsilon < 0\}} \rho(\varepsilon) d\varepsilon
\end{aligned} \tag{3.18}$$

It holds that all the integrals in Eqn (3.18), except the last one, tends to zero as $x \to \infty$. Consequently, it is readily seen that they are bounded over $x > 0$. We only verify that the first integral on the right hand side of Eqn (3.18) tends to zero as $x \to \infty$, as the zero limits of the second and third integrals can be similarly shown. The limit of the first integral can be expressed as

$$\begin{aligned}
& \lim_{x \to +\infty} E\left[\left[a^k \{\phi_{1,0} + \phi_{1,1}x + (\phi_{1,1} + \theta_{1,1})\varepsilon_t\}^k + 1\right] I_{\{\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon_t \geq 0,\, \varepsilon_t \leq -x\}}\right] \\
\leq & \lim_{x \to +\infty} E\left[a^k |\{\phi_{1,0} + \phi_{1,1}x + (\phi_{1,1} + \theta_{1,1})\varepsilon_t\}^k + 1| I_{\{\varepsilon_t \leq -x\}}\right] \\
\leq & \lim_{x \to +\infty} E\left[[a^k 2^k \{|\phi_{1,1}|^k x^k + |\phi_{1,0} + (\phi_{1,1} + \theta_{1,1})\varepsilon_t|^k\} + 1] I_{\{\varepsilon_t \leq -x\}}\right] \\
= & \lim_{x \to +\infty} 2^k a^k |\phi_{1,1}|^k x^k P(\varepsilon_t \leq -x) \\
+ & \lim_{x \to +\infty} E\left[\{2^k a^k |\phi_{1,0} + (\phi_{1,1} + \theta_{1,1})\varepsilon_t|^k + 1\} I_{\{\varepsilon_t \leq -x\}}\right].
\end{aligned}$$

Because $E[|\varepsilon_t|^k] < \infty$, it follows from Lemma 52 and the Lebesgue dominated convergence theorem that the last two limits are zero, hence the claim holds. It can be similarly proved that all the integrals in Eqn (3.18), except the third

Conditions on parameters	Constants a and b in the test function
$0 \leq \phi_{1,1} < 1$ $\phi_{2,1} \leq -1$	$a = 1$ $b < \|\phi_{2,1}\|^{-1}$
$\phi_{1,1} \leq -1$ $0 \leq \phi_{2,1} < 1$	$a < \|\phi_{1,1}\|^{-1}$ $b = 1$
$-1 \leq \phi_{1,1} < 0$ $\phi_{2,1} \leq -1$ $\phi_{1,1}\phi_{2,1} < 1$	$a = 1$ $\|\phi_{1,1}\| < b < \|\phi_{2,1}\|^{-1}$
$\phi_{1,1} \leq -1$ $-1 \leq \phi_{2,1} < 0$ $\phi_{1,1}\phi_{2,1} < 1$	$a = 1$ $\|\phi_{2,1}\| < b < \|\phi_{1,1}\|^{-1}$
$\|\phi_{1,1}\| < 1$ $\|\phi_{2,1}\| < 1$	$a = 1$ $b = 1$

Table 3.3: Test functions in the drift criterion for the geometric ergodicity.

one, tends to zero as $x \to -\infty$. Thus (3.17) follows by noting that

$$\begin{aligned}
E[g(\tilde{V}_{t+1})|\tilde{V}_t = x] = I_{\{x\leq 0\}} \int_{-\infty}^{\infty} &\left[a^k\{\phi_{1,0} + \phi_{1,1}x + (\phi_{1,1} + \theta_{1,1})\varepsilon\}^k + 1\right] \\
&\times I_{\{\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon\geq 0\}}\rho(\varepsilon)d\varepsilon \\
&+ I_{\{x>0\}} \int_{-\infty}^{+\infty} \left[a^k\{\phi_{2,0} + \phi_{2,1}x + (\phi_{2,1} + \theta_{2,1})\varepsilon\}^k + 1\right] \\
&\times I_{\{\phi_{2,0}+\phi_{2,1}x+(\phi_{2,1}+\theta_{2,1})\varepsilon\geq 0\}}\rho(\varepsilon)d\varepsilon \\
&+ I_{\{x\leq 0\}} \int_{-\infty}^{\infty} \left[b^k|\phi_{1,0} + \phi_{1,1}x + (\phi_{1,1} + \theta_{1,1})\varepsilon|^k + 1\right] \\
&\times I_{\{\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon< 0\}}\rho(\varepsilon)d\varepsilon \\
&+ I_{\{x>0\}} \int_{-\infty}^{+\infty} \left[b^k|\phi_{2,0} + \phi_{2,1}x + (\phi_{2,1} + \theta_{2,1})\varepsilon|^k + 1\right] \\
&\times I_{\{\phi_{2,0}+\phi_{2,1}x+(\phi_{2,1}+\theta_{2,1})\varepsilon< 0\}}\rho(\varepsilon)d\varepsilon,
\end{aligned}$$

which completes the proof.

Proof of Part 2 of Theorem 48.

In view of the proof of the preceding part and since geometric ergodicity implies ergodicity, we need only verify the ergodicity of $\{V_t\}$ under any one of (48.3)-(48.6). The results follows from [Tweedie, 1976, Theorem 9.1(i)] if there exist a small set of the form $[-M, M]$ (c.f. the remark on M in the proof of Part 1 of Theorem 48) and a function $g(\cdot) : \mathbb{R} \to [0, +\infty]$ and some constants $K, \gamma > 0$, such that

$$\begin{array}{lll} \text{(i)} & E[g(V_{t+1})|V_t = x] \leq g(x) - \gamma, & \text{for all } x \in [-M, M]^c; \\ \text{(ii)} & E[g(V_{t+1})|V_t = x] \leq K, & \text{for all } x \in [-M, M]. \end{array}$$

For (48.3)-(48.5), the function $g(x)$ takes the form

$$g(x) = \begin{cases} -bx, & x < 0 \\ ax, & x \geq 0 \end{cases}$$

with a and b being positive constants as determined in [Chan et al., 1985, Lemma 2.2]. For (48.6), $g(x)$ takes the form

$$g(x) = \begin{cases} -x + d, & x < 0 \\ -\phi_{2,1}x + c, & x \geq 0 \end{cases}$$

where c, d are two positive constants chosen such that $d \geq |\phi_{2,0}|$ and

$$-\phi_{2,0} + d < c < \phi_{2,1}\phi_{1,0} + d,$$

which is feasible because $\phi_{2,0} + \phi_{2,1}\phi_{1,0} > 0$. Conditions (i) and (ii) hold with g so defined for $\{\tilde{V}_t\}$, c.f. Lemma 2.2 in Chan et al. [1985] although we have constructed a new g function for (48.6) while that used in Chan et al. [1985] is based on $\{\tilde{V}_{2t}\}$. Moreover, (3.17) can be readily seen to hold so the drift criterion using the same g function is satisfied for $\{V_t\}$ as well.

Proof of Part 3 of Theorem 48.

The proof is divided into two steps. First, we show that the chain is recurrent, and then we prove that it is not ergodic under these conditions and so $\{V_t\}$ must be null recurrent. We start by proving the recurrence of the chain, using the drift criterion in [Tweedie, 1976, Theorem 10.2]. The result follows if there exist a small set of the form $[-M, M]$ (c.f. the remark on M in the proof of Part 1 of Theorem 48) and a function $g(\cdot) : \mathbb{R} \to [0, +\infty]$ such that

$$\begin{aligned} &(i) E[g(V_{t+1})|V_t = x] \leq g(x), \text{ for all } x \in [-M, M]^c; \\ &(ii) g(x) > \sup_{y \in [-M,M]} g(y), \text{ for all } x \in [-M, M]^c; \\ &(iii) B_n = \{y \in \mathbb{R} : g(y) \leq n\} \text{ is a small set for all sufficient large } n. \end{aligned}$$

Following [Guo and Petruccelli, 1991], the test function $g(\cdot)$ takes the following form:

$$g(x) = \begin{cases} \ln(-bx + \beta), & x < -M, \\ 0 & -M \leq x < M, \\ \ln(ax + \alpha), & x \geq M \end{cases}$$

with a and b positive constants and α and β determined in [Guo and Petruccelli, 1991, lemma 3]. The test function so defined satisfies conditions (i)-(iii) with

V_t there replaced by $\tilde{V}_t$ [Guo and Petruccelli, 1991, lemma 3]; specifically they showed that

$$E[g(\tilde{V}_t)|\tilde{V}_t = x] \leq g(x) - \kappa x^{-2} + o(x^{-2}),$$

where κ is a positive constant and $o(x^{-2})$ is an expression that multiplied by x^2 tends to zero as $x \to +\infty$ or $x \to -\infty$. Hence, for the drift criterion to hold for $\{V_t\}$, it suffices to show that

$$E[g(V_t)|V_t = x] \leq E[g(\tilde{V}_t)|\tilde{V}_t = x] + \tau(x), \tag{3.19}$$

with $\tau(x) = o(x^{-2})$.

We prove (3.19) only in the case that $\phi_{1,1} = 1$, $\phi_{2,1} = 1$, $\phi_{2,0} = 0$, $\phi_{1,0} \geq 0$, since the proof is similar for other cases. Set $a = b = \alpha = \beta = 1$ and consider

$$\begin{aligned} E[g(V_{t+1})|V_t = x] \quad &= \textstyle\int_{-\infty}^{-x} \ln\{1 + \phi_{1,0} + \phi_{1,1}x + (\phi_{1,1} + \theta_{1,1})\varepsilon\} \\ &\quad \times I_{\{\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon>0\}}\rho(\varepsilon)d\varepsilon \\ &\quad + \textstyle\int_{-x}^{+\infty} \ln\{1 + \phi_{2,0} + \phi_{2,1}x + (\phi_{2,1} + \theta_{2,1})\varepsilon\} \\ &\quad \times I_{\{\phi_{2,0}+\phi_{2,1}x+(\phi_{2,1}+\theta_{2,1})\varepsilon>0\}}\rho(\varepsilon)d\varepsilon \\ &\quad + \textstyle\int_{-\infty}^{-x} \ln[1 - \{\phi_{1,0} + \phi_{1,1}x + (\phi_{1,1} + \theta_{1,1})\varepsilon\}] \\ &\quad \times I_{\{\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon<0\}}\rho(\varepsilon)d\varepsilon \\ &\quad + \textstyle\int_{-x}^{+\infty} \ln[1 - \{\phi_{2,0} + \phi_{2,1}x + (\phi_{2,1} + \theta_{2,1})\varepsilon\}] \\ &\quad \times I_{\{\phi_{2,0}+\phi_{2,1}x+(\phi_{2,1}+\theta_{2,1})\varepsilon<0\}}\rho(\varepsilon)d\varepsilon. \end{aligned} \tag{3.20}$$

For positive x, the validity of (3.19) follows from the following claim and after some routine algebra:

$$\begin{aligned} &\lim_{x\to+\infty} x^2 \int_{-\infty}^{-x} \ln\{1 + \phi_{1,0} + \phi_{1,1}x + (\phi_{1,1} + \theta_{1,1})\varepsilon\} \\ &\qquad \times I_{\{\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon>0\}}\rho(\varepsilon)d\varepsilon = 0, \\ &\lim_{x\to+\infty} x^2 \int_{-\infty}^{-x} \ln[1 - \{\phi_{1,0} + \phi_{1,1}x + (\phi_{1,1} + \theta_{1,1})\varepsilon\}] \\ &\qquad \times I_{\{\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon<0\}}\rho(\varepsilon)d\varepsilon = 0. \end{aligned}$$

We only verify the first limit as the other one can be similarly shown. Applying (3.13) to the expression in the first limit, we have that

$$\begin{aligned} &x^2 E\left[\ln\{1 + \phi_{1,0} + \phi_{1,1}x + (\phi_{1,1} + \theta_{1,1})\varepsilon_t\} I_{\{\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon_t>0,\, \varepsilon_t\leq -x\}}\right] \\ \leq &x^2 P\left(\phi_{1,0} + \phi_{1,1}x + (\phi_{1,1} + \theta_{1,1})\varepsilon_t > 0,\, \varepsilon_t \leq -x\right) \ln(1 + \phi_{1,0} + \phi_{1,1}x) \\ &+ x^2(\phi_{1,1} + \theta_{1,1})(1 + \phi_{1,0} + \phi_{1,1}x)^{-1} E\left[\varepsilon_t I_{\{\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon_t>0,\, \varepsilon_t\leq -x\}}\right] \\ &- x^2(\phi_{1,1} + \theta_{1,1})^2(1 + \phi_{1,0} + \phi_{1,1}x)^{-2} 2^{-1} \\ &\quad \times E\left[\varepsilon_t^2 I_{\{\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon_t>0,\, \varepsilon_t\leq -x,\, (\phi_{1,1}+\theta_{1,1})\varepsilon_t<0\}}\right]. \end{aligned}$$

We prove separately that the three addends tend to zero. Consider the first addend. Recall $E[|\varepsilon_t|^k] < \infty$ for some $k > 2$. Let $\delta > 0$ be such that $2 + \delta < k$.

$$
\begin{aligned}
0 &\le \liminf_{x\to+\infty} x^2 \ln(1+\phi_{1,0}+\phi_{1,1}x) P\left(\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon_t > 0,\, \varepsilon_t \le -x\right)\\
&\le \limsup_{x\to+\infty} x^2 \ln(1+\phi_{1,0}+\phi_{1,1}x) P\left(\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon_t > 0,\, \varepsilon_t \le -x\right)\\
&\le \lim_{x\to+\infty} x^2 \ln(1+\phi_{1,0}+\phi_{1,1}x) P\left(\varepsilon_t \le -x\right)\\
&= \lim_{x\to+\infty} x^{2+\delta} \ln(1+\phi_{1,0}+\phi_{1,1}x) x^{-\delta} P\left(\varepsilon_t \le -x\right) = 0,
\end{aligned}
$$

from Lemma 52. Hence the first addend tends to 0 as $x \to +\infty$. It is clear that the second addend is non-negative for all sufficiently large x. In order to prove that it tends to 0 as $x \to +\infty$, it suffices to show that

$$
\begin{aligned}
&\limsup_{x\to+\infty} x^2(\phi_{1,1}+\theta_{1,1})(1+\phi_{1,0}+\phi_{1,1}x)^{-1} E\left[\varepsilon_t I_{\{\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon_t>0,\,\varepsilon_t\le -x\}}\right]\\
&\le \lim_{x\to+\infty} (\phi_{1,1}+\theta_{1,1})[(1+\phi_{1,0})/x+\phi_{1,1}]^{-1} x E\left[\varepsilon_t I_{\{\varepsilon_t\le -x\}}\right] = 0,
\end{aligned}
$$

from Lemma 51. Since the variance of ε_t is finite, the Lebesgue dominated convergence theorem implies that the limit of the third addend is zero. Since $x \to +\infty$, we can consider x to be positive and therefore it follows that

$$
\begin{aligned}
E[g(\tilde{V}_{t+1})|\tilde{V}_t = x] = &\int_{-\infty}^{+\infty} \ln\{1+\phi_{2,0}+\phi_{2,1}x+(\phi_{2,1}+\theta_{2,1})\varepsilon\}\\
&\times I_{\{\phi_{2,0}+\phi_{2,1}x+(\phi_{2,1}+\theta_{2,1})\varepsilon>0\}}\rho(\varepsilon)d\varepsilon\\
&+\int_{-\infty}^{+\infty} \ln[1-\{\phi_{2,0}+\phi_{2,1}x+(\phi_{2,1}+\theta_{2,1})\varepsilon\}]\\
&\times I_{\{\phi_{2,0}+\phi_{2,1}x+(\phi_{2,1}+\theta_{2,1})\varepsilon<0\}}\rho(\varepsilon)d\varepsilon.
\end{aligned}
$$

That (3.19) holds for non-positive x can be similarly proved. The same argument holds also for x tending to $-\infty$ and so the proof is completed. Now, we prove the nullity of the chain. First we prove it when (48.11) holds. We use the drift criterion in Tweedie [1976], Theorem 9.1(ii). The chain is null if there exist $\delta > 0$, a non-negative function $g(\cdot)$ and a set A, with $\mu(A) > 0$ and $\mu(A^c) > 0$, such that:

$$
\begin{array}{ll}
\text{(i) } E[g(V_{t+1})|V_t = x] \ge g(x), & \text{for each } x \in A^c;\\
\text{(ii) } E[|g(V_{t+1}) - g(x)| \quad | \quad V_t = x] \le \delta, & \text{for each } x \in \mathbb{R};\\
\text{(iii) } g(x) > \sup_{y\in A} g(y), & \text{for each } x \in A^c.
\end{array}
$$

Let the test function be

$$
g(x) = \begin{cases} -bx+\beta, & x<0\\ ax+\alpha, & x\ge 0. \end{cases}
$$

Set a, b, α and β as in Chan et al. [1985],Lemma 2.3, i.e., satisfying the following conditions:

1. $\phi_{1,1} = -ba^{-1}$ and $\phi_{2,1} = -ab^{-1}$;

2. $a\phi_{1,0} \geq \beta - \alpha \geq b\phi_{2,0}$.

The test function so defined satisfies Conditions (i)-(iii) with V_t replaced by $\tilde{V}_t$ Chan et al. [1985], Lemma 2.3. It is not hard to show that (3.17) holds. In fact, consider

$$\begin{aligned} E[g(V_{t+1})|V_t = x] = &\int_{-\infty}^{-x} [a\{\phi_{1,0} + \phi_{1,1}x + (\phi_{1,1} + \theta_{1,1})\varepsilon\} + \alpha] \\ &\times I_{\{\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon \geq 0\}}\rho(\varepsilon)d\varepsilon \\ &+ \int_{-x}^{+\infty} [a\{\phi_{2,0} + \phi_{2,1}x + (\phi_{2,1} + \theta_{2,1})\varepsilon\} + \alpha] \\ &\times I_{\{\phi_{2,0}+\phi_{2,1}x+(\phi_{2,1}+\theta_{2,1})\varepsilon \geq 0\}}\rho(\varepsilon)d\varepsilon \\ &+ \int_{-\infty}^{-x} [-b\{\phi_{1,0} + \phi_{1,1}x + (\phi_{1,1} + \theta_{1,1})\varepsilon\} + \beta] \\ &\times I_{\{\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon < 0\}}\rho(\varepsilon)d\varepsilon \\ &+ \int_{-x}^{+\infty} [-b\{\phi_{2,0} + \phi_{2,1}x + (\phi_{2,1} + \theta_{2,1})\varepsilon\} + \beta] \\ &\times I_{\{\phi_{2,0}+\phi_{2,1}x+(\phi_{2,1}+\theta_{2,1})\varepsilon < 0\}}\rho(\varepsilon)d\varepsilon. \end{aligned} \tag{3.21}$$

Applying Lemma 52, with $k = 1$, and the Lebesgue dominated convergence theorem, it is readily seen that all the integrals in Eqn (3.21), except the last (first) one, tends to zero as $x \to \infty$ $(x \to -\infty)$.
Finally, we prove the nullity of the chain when:

$$(\phi_{1,1} = 1 \quad \text{and} \quad \phi_{1,0} = 0) \qquad \text{or} \qquad (\phi_{2,1} = 1 \quad \text{and} \quad \phi_{2,0} = 0),$$

with an approach different from the random-walk argument in Chan et al. [1985] that is invalidated by the random threshold. It is easy to prove that the Conditions (i)-(iii) hold by choosing $g(\cdot)$ as follows. If $\phi_{1,1} = 1$ and $\phi_{1,0} = 0$, take $A = (0, +\infty)$ and

$$g(x) = \begin{cases} -x, & x < 0 \\ 0, & x \geq 0 \end{cases}$$

if $\phi_{2,1} = 1$ and $\phi_{2,0} = 0$, take $A = (-\infty, 0)$ and

$$g(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0. \end{cases}$$

Proof of Part 4 of Theorem 48.

The result follows from Tweedie [1976], Theorem 11.3(ii), if there exist a bounded non-negative function $g(\cdot)$, a set A, with $\mu(A) > 0$ and $\mu(A^c) > 0$, such that

$$\begin{aligned} (i) &\quad E[g(V_{t+1})|V_t = x] \geq g(x), \qquad \text{for all } x \in A^c; \\ (ii) &\quad g(x) > \sup_{y \in A} g(y), \qquad \text{for all } x \in A^c. \end{aligned}$$

We provide some details for the case (48.15). Since $\phi_{2,1} < 0$ and $\phi_{1,1}\phi_{2,1} = 1$, there exist two positive constants, a and b such that

$$\phi_{1,1} = -ba^{-1} \quad \text{and} \quad \phi_{2,1} = -ab^{-1}.$$

Since $\phi_{2,0} + \phi_{2,1}\phi_{1,0} < 0$, there exist two real constants, α and β, such that

$$-a\phi_{1,0} < a\alpha + b\beta < -b\phi_{2,0}.$$

Moreover, let c be a positive constant such that

$$ca^{-1} - \alpha > 0 \quad \text{and} \quad -cb^{-1} - \beta < 0.$$

Consider the test function $g(\cdot)$ defined in Guo and Petruccelli [1991], lemma 4, that is,

$$g(x) = \begin{cases} 1 + 1/\{b(x+\beta)\}, & x < -cb^{-1} - \beta \\ 1 - 1/c, & -cb^{-1} - \beta < x < ca^{-1} - \alpha \\ 1 - 1/\{a(x+\alpha)\}, & x > ca^{-1} - \alpha. \end{cases}$$

As (3.19) can be readily shown by applying Lemma 52, with $k = 2$, the proof follows from that of Guo and Petruccelli [1991], lemma 4. For other cases, the proof is similar, so we only spell out the form of the test function and the corresponding set A. The test functions constructed for the cases (48.12–13, 48.16–17) are new.

Case (48.12). Let c be a positive constant such that $c - 1 > 0$. Then, take $A = [-c-1, +\infty)$ and define

$$g(x) = \begin{cases} 1 + 1/(x+1), & x < -c-1 \\ 1 - 1/c, & x \geq -c-1. \end{cases}$$

Case (48.13). Let c be a positive constant such that $c - 1 > 0$. Then, take $A = (-\infty, c-1]$ and define

$$g(x) = \begin{cases} 1 - 1/c, & x \leq c-1 \\ 1 - 1/(x+1), & x > c-1. \end{cases}$$

Case (48.14). Let a be a positive constant such that $-\phi_{1,1}^{-1} < a < -\phi_{2,1}$ (note that this is attainable since $\phi_{1,1} < 0$ and $\phi_{1,1}\phi_{2,1} > 1$) and let c be a positive constant such that $c/a - 1 > 0$. Then, take $A = [-c-1, c/a-1]$ and define

$$g(x) = \begin{cases} 1 + 1/(x+1), & x \leq -c-1 \\ 1 - 1/c, & -c-1 \leq x \leq c/a - 1 \\ 1 - 1/\{a(x+1)\}, & x > c/a - 1. \end{cases}$$

Case (48.16). Let c be a positive constant such that $c - 1 > 0$. Then, take $A = (-\infty, c-1]$ and define

$$g(x) = \begin{cases} 1 - 1/c, & x \leq c-1 \\ 1 - 1/(x+1), & x > c-1. \end{cases}$$

Case (48.17). Let c be a positive constant such that $c - 1 > 0$. Then, take $A = [-c-1, +\infty)$ and define

$$g(x) = \begin{cases} 1 + 1/(x+1), & x < -c-1 \\ 1 - 1/c, & x \geq -c-1. \end{cases}$$

3.5 Extension to multiple-regime TARMA(1,1) processes

We now extend the results from the 2-regime TARMA(1,1) model to the multiple-regime TARMA(1,1) model. More generally, the m-regime TARMA(p,p) process satisfies the following difference equation:

$$X_t = \sum_{i=1}^{m} \left\{ \phi_{i,0} + \sum_{j=1}^{p} \phi_{i,j} X_{t-j} + \varepsilon_t + \sum_{j=1}^{p} \theta_{i,j} \varepsilon_{t-j} \right\} \times I(r_{i-1} < X_{t-d} \leq r_i)$$

where $-\infty = r_0 < r_1 < \cdots < r_{m-1} < r_m = \infty$. The Markovian representation detailed in Section 3.1 can be readily lifted to the multiple-regime TARMA process. Moreover, Propositions 37–40 continue to hold for multiple-regime TARMA(1,1) processes so that we can study the long-run behaviour of $\{X_t\}$ by studying those of $\{V_t\}$. We now state two main results.

Theorem 53. *Let $\{X_t\}$ be a multiple-regime TARMA(1,1) process. Suppose (C2) holds. Then, $\{V_t\}$ is irreducible and aperiodic if (i) (C3) holds with the coefficients of the second regime there replaced by those of the last regime, or (ii) if $\theta_{i,1} = 0$ for some i.*

Theorem 54. *Let $\{X_t\}$ be a multiple-regime TARMA(1,1) process. Suppose (C1) and (C2) hold, and either (C3) holds for the outermost two regimes of $\{X_t\}$ or one of its regimes is defined by an AR(1) model, i.e., $\theta_{i,1} = 0$ for some i. Then the classification results stated in Theorem 48 hold for $\{V_t\}$ with the parametric conditions pertaining to the outermost two regimes.*

We outline the proofs in the case with three regimes:

$$V_t = \begin{cases} \phi_{1,0} + \phi_{1,1}V_t + (\phi_{1,1} + \theta_{1,1})\varepsilon_t, & \text{if } V_t \le r_1 - \varepsilon_t \\ \phi_{2,0} + \phi_{2,1}V_t + (\phi_{2,1} + \theta_{2,1})\varepsilon_t, & \text{if } r_1 - \varepsilon_t < V_{t-1} \le r_2 - \varepsilon_t \\ \phi_{3,0} + \phi_{3,1}V_t + (\phi_{3,1} + \theta_{3,1})\varepsilon_t, & \text{if } V_t > r_2 - \varepsilon_t. \end{cases} \quad (3.22)$$

We derive the set G_x, the reachable set in 1 step from x, as follows. In order for $V_t = x$ and $V_{t+1} = y$, we exhibit a realization of the variable ε_t, i.e., u, such that at least one of the following systems admits a solution.

(i)

$$\begin{cases} \phi_{1,0} + \phi_{1,1}x + (\phi_{1,1} + \theta_{1,1})u = y \\ x \le r_1 - u; \end{cases}$$

(ii)

$$\begin{cases} \phi_{2,0} + \phi_{2,1}x + (\phi_{2,1} + \theta_{2,1})u = y \\ r_1 - u < x \le r_2 - u; \end{cases} .$$

(iii)

$$\begin{cases} \phi_{3,0} + \phi_{3,1}x + (\phi_{3,1} + \theta_{3,1})u = y \\ x > r_2 - u. \end{cases} .$$

Define the following functions:

$$\begin{aligned} h_1(x) &= (\phi_{1,1} + \theta_{1,1})r_1 + \phi_{1,0} - \theta_{1,1}x; \\ h_2(x) &= (\phi_{2,1} + \theta_{2,1})r_1 + \phi_{2,0} - \theta_{2,1}x; \\ h_3(x) &= (\phi_{2,1} + \theta_{2,1})r_2 + \phi_{2,0} - \theta_{2,1}x; \\ h_4(x) &= (\phi_{3,1} + \theta_{3,1})r_2 + \phi_{3,0} - \theta_{3,1}x. \end{aligned}$$

The set G_x has a different form according to the signs of $\phi_{i,1}+\theta_{i,1}$, $i = 1, 2, 3$, as summarized in the Table 3.4. Obviously, there are some cases in which $G_x = \mathbb{R}$. For instance, if $\phi_{i,1}+\theta_{i,1} > 0$, $i = 1, 2, 3$, then $G_x = \mathbb{R}$ if one of the two following conditions holds:

(1) $h_1(x) > h_4(x)$;

(2) $h_1(x) > h_2(x)$ and $h_3(x) > h_4(x)$.

In the other cases, note that

$$\begin{aligned} G_x &\subset (-\infty, h_1(x)] \cup (h_4(x), +\infty), \text{if } \phi_{1,1} + \theta_{1,1} > 0 \text{ and } \phi_{3,1} + \theta_{3,1} > 0; \\ G_x &\subset (-\infty, h_1(x)] \cup (-\infty, h_4(x)), \text{if } \phi_{1,1} + \theta_{1,1} > 0 \text{ and } \phi_{3,1} + \theta_{3,1} < 0; \\ G_x &\subset [h_1(x), +\infty) \cup (h_4(x), +\infty), \text{if } \phi_{1,1} + \theta_{1,1} < 0 \text{ and } \phi_{3,1} + \theta_{3,1} > 0; \\ G_x &\subset [h_1(x), +\infty) \cup (-\infty, h_4(x)), \text{if } \phi_{1,1} + \theta_{1,1} < 0 \text{ and } \phi_{3,1} + \theta_{3,1} < 0. \end{aligned}$$

$\phi_{1,1}+\theta_{1,1}$	$\phi_{2,1}+\theta_{2,1}$	$\phi_{3,1}+\theta_{3,1}$	1-step Reachable set
+	+	+	$(-\infty,h_1(x)]\cup(h_2(x),h_3(x)]\cup(h_4(x),+\infty)$
+	+	-	$(-\infty,h_1(x)]\cup(h_2(x),h_3(x)]\cup(-\infty,h_4(x))$
+	-	+	$(-\infty,h_1(x)]\cup[h_3(x),h_2(x))\cup(h_4(x),+\infty)$
-	+	+	$[h_1(x),+\infty)\cup(h_2(x),h_3(x)]\cup(h_4(x),+\infty)$
+	-	-	$(-\infty,h_1(x)]\cup[h_3(x),h_2(x))\cup(-\infty,h_4(x))$
-	+	-	$[h_1(x),+\infty)\cup(h_2(x),h_3(x)]\cup(-\infty,h_4(x))$
-	-	+	$[h_1(x),+\infty)\cup[h_3(x),h_2(x))\cup(h_4(x),+\infty)$
-	-	-	$[h_1(x),+\infty)\cup[h_3(x),h_2(x))\cup(-\infty,h_4(x))$

Table 3.4: G_x according to the parametric conditions.

Therefore if either $\theta_{1,1}\neq 0$ or $\theta_{2,1}\neq 0$ then we can apply the same arguments used in the case of two regimes to study the irreducibility. To sum up, the three-regime TARMA(1,1) model is irreducible if $\phi_{1,0}$, $\phi_{3,0}$, $\phi_{1,1}$, $\phi_{3,1}$, $\theta_{1,1}$ and $\theta_{3,1}$ satisfy the modified condition (C3) (c.f. Theorem 41). On the other hand, irreducibility holds automatically if any $\theta_{i,1}=0$, because of the existence of an infinite interval that is a subset of the reachable set $\Omega\subseteq G_x$ that does not depend on x. For instance, if $\theta_{2,1}=0$, it follows that:

$$\Omega=(h_2(0),h_3(0)],\qquad \text{if } \phi_{2,1}>0;$$
$$\Omega=[h_3(0),h_2(0)),\qquad \text{if } \phi_{2,1}<0.$$

It is then readily checked that $\{V_t\}$ is aperiodic as Proposition 47 holds regardless of the number of regimes. The classification of long-run probabilistic behaviour depends only on the parameters in the extreme regimes, as is also the case for the TAR(1) model [Chan et al., 1985]. For example, when we apply the drift criterion for verifying the geometric ergodicity, the expectation $E[g(V_{t+1})|V_t=x]$ can be decomposed into a sum of three integrals:

$$\begin{aligned}E[g(V_{t+1})|V_t=x] &= \int_{-\infty}^{r_1-x} g(\phi_{1,0}+\phi_{1,1}x+(\phi_{1,1}+\theta_{1,1})\varepsilon)\rho(\varepsilon)d\varepsilon\\ &+\int_{r_1-x}^{r_2-x} g(\phi_{2,0}+\phi_{2,1}x+(\phi_{2,1}+\theta_{2,1})\varepsilon)\rho(\varepsilon)d\varepsilon\\ &+\int_{r_2-x}^{+\infty} g(\phi_{3,0}+\phi_{3,1}x+(\phi_{3,1}+\theta_{3,1})\varepsilon)\rho(\varepsilon)d\varepsilon.\end{aligned}$$

By using the Lebesgue dominated convergence theorem, it is readily checked that the intermediate integral tend to zero as $x\to+\infty$ (or $x\to-\infty$), resulting in the equation being akin to the case of two regimes so that we can lift the proof techniques used in the two-regime case to multiple regimes. Obviously, the same argument holds if there are more than three regimes and can be modified appropriately for the other drift criteria. This completes the outline of the two proofs.

Chapter 4

Unit-root testing for linear and non-linear alternatives: a TARMA based approach

4.1 Introduction

In this Chapter we propose a novel unit root test that does not suffer from size distortions and, at the same time, allows for a wide and flexible non-linear alternative and is robust against heteroskedasticity. This is made possible by the recent results on the probabilistic structure of the TARMA model derived in Chan and Goracci [2017]. We specify an IMA$(1, 1)$ model as the null hypothesis and a TARMA$(1, 1)$ with a unit root regime as the alternative. As we will show, both the IMA(1,1) and the TARMA(1,1) are able to encompass a wide range of stationary and non-stationary linear and non-linear models. In particular, the IMA(1,1) formalizes the exponential smoothing approach, a popular general-purpose forecasting tool (see Gardner [1985], Holt [2004], Hyndman et al. [2008], Chatfield [2000]) that predates the Box-Jenkins approach of forecasting via the ARIMA models. It is based on the simple idea that past data contain relevant information for predicting future values, with the relevance declining with time at some geometric rate quantified by the MA coefficient.

We propose a supremum Lagrange Multiplier test statistic (supLM) and derive its asymptotic distribution both under the null hypothesis and local alternatives. We prove that the test is consistent and free of nuisance parameters. Moreover, we also prove that it is similar in that its distribution does not depend on the value of the MA parameter. The derivation of the asymptotic theory is highly non-standard and the approach is completely new and not based on existing results e.g. those of Park and Phillips [2001]. We also introduce a wild bootstrap version of the supLM statistic. We perform a simulation study where we compare our proposals with existing tests where the alternative hypothesis is that of a threshold model. In general, the size of such tests is severely biased in a number of cases so that their use in practical applications remains questionable if no additional information on the data generating process is available. Also, the comparison includes the best performing unit root tests to date, where the alternative hypothesis does not specify explicitly a non-linear process. The striking evidence from the simulations confirms that rejection of the null hypothesis does not necessarily imply a non-linear specification so that, as also suggested in Choi [2015], unit root tests should not be applied without previous knowledge on the series. Finally, we apply our tests to the monthly real exchange rates from September 1973 to December 1998 for a set of European countries. We are able to reject with confidence the null hypothesis for Germany and find a plausible TARMA fit that might help shedding some light on the PPP puzzle.

The chapter is structured as follows. In Section 4.2 we present the parametrization of TARMA models that will be used in the theoretical derivation of our test. In particular, we describe the TARMA(1,1) parametrization under the alternative hypothesis that reduces to the IMA(1,1) process under the null. In Section 4.3 we present the supLM test statistic. We derive its asymptotic distribution both under the null hypothesis and local alternatives. Moreover, we present the wild bootstrap version of our test statistic. In Section 4.4 we perform a large scale simulation study to show the performance of the tests in finite samples, in terms of size and power, and compare them with existing proposals. Section 4.5 contains the empirical illustration where we apply the tests to the pre-Euro monthly real exchange rates of a set countries. Finally, in Section 4.6 we outline in details all the proofs.

4.2 TARMA models in the unit-root setting

The two-regime TARMA(1,1) model specifies that the time series $\{X_t, t = 0, 1, \cdots\}$ satisfies the following equation:

$$X_t = \begin{cases} \phi_{1,0} + \phi_{1,1}X_{t-1} + \theta_{1,0}\varepsilon_t - \theta_{1,1}\varepsilon_{t-1}, & \text{if } X_{t-d} \leq r \\ \phi_{2,0} + \phi_{2,1}X_{t-1} + \theta_{2,0}\varepsilon_t - \theta_{2,1}\varepsilon_{t-1} & \text{otherwise,} \end{cases} \tag{4.1}$$

where the innovations $\{\varepsilon_t\}$ are iid random variables of zero mean and variance σ^2, ε_t is independent of past X's, i.e., $X_{t-j}, j = 1, \cdots, t$; the delay d is a positive integer which, for simplicity, is taken to be 1 henceforth; r is the real-valued threshold parameter; the ϕ's and θ's are unknown coefficients. The preceding TARMA model is not identifiable without further parametric constraints. One way to ensure model identifiability is to set $\sigma^2 = 1$. The TARMA process is said to fall into the first (second) regime if $X_{t-1} \leq r$ ($X_{t-1} > r$). The TARMA model specifies that the data-generating mechanism switches between two ARMA(1,1) sub-models depending on whether the threshold variable X_{t-1} exceeds the threshold r. Clearly, the TARMA(1,1) model subsumes the IMA(1,1) model. In order to develop our test, we introduce the following constrained TARMA(1,1) model that states a general hypothesis including both the

IMA(1,1) model and against which a certain competitive direction of non-linear departure:

$$X_t = \begin{cases} \tilde{\phi}_{1,0} + \tilde{\phi}_{1,1} X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, & \text{if } X_{t-d} \le r \\ \tilde{\phi}_{2,0} + X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} & \text{otherwise,} \end{cases} \tag{4.2}$$

where the innovations are re-parameterized to make the innovation variance a positive parameter. The preceding (constrained) TARMA(1,1) model assumes that the sub-model in the second (upper) regime is an IMA(1,1) model while the sub-model in the lower regime is a general ARMA(1,1) model.

Statistical inference with a TARMA model hinges on whether the model is invertible in the sense that the innovations can be asymptotically recovered were the model parameters known [Chan and Tong, 2010]. Note that model (4.2) is invertible if $|\theta| < 1$ [Chan and Tong, 2010]. Henceforth, θ is assumed to lie inside $(-1, 1)$. Assuming the innovations admit a positive, continuous probability density function (pdf) of finite second moment, Chan and Goracci [2017] showed that model (4.2) is an ergodic Markov chain if and only if $\tilde{\phi}_{2,0} < 0$ and either (i) $\tilde{\phi}_{1,1} < 1$, or (ii) $\tilde{\phi}_{1,1} = 1$, $\tilde{\phi}_{1,0} > 0$; ergodicity then implies that the TARMA(1,1) model admits a unique stationary distribution. Furthermore, Chan and Goracci [2017] provides a complete classification of the parametric regions of the TARMA(1,1) model into sub-regions of ergodicity, null recurrence and transience: the (constrained) TARMA(1,1) defined by (4.2) is null-recurrent if (iii) $\tilde{\phi}_{2,0} \ge 0$ or (iv) $\tilde{\phi}_{1,1} = 1, \tilde{\phi}_{1,0} < 0$; if none of (i)–(iv) hold, then the model is transient. Thus model (4.2) is a rich model that encompasses both linear or non-linear processes spanning a wide spectrum of long-run behaviours including ergodicity, null recurrence and transience.
Model (4.2) can be re-parametrized as follows:

$$X_t = \phi_0 + X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} + \{\phi_{1,0} + \phi_{1,1} X_{t-1}\} \times I(X_{t-1} \le r),$$

where $\phi_0 = \tilde{\phi}_{2,0}$, $\phi_{1,0} = \tilde{\phi}_{1,0} - \tilde{\phi}_{2,0}$ and $\phi_{1,1} = \tilde{\phi}_{1,1} - 1$.

4.3 Lagrange multiplier test

Let $\{X_t, t = 1, \ldots, n\}$, be a time series and assume that, for each t, X_t satisfies the difference equation:

$$H: \qquad X_t = \phi_0 + X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} + \{\phi_{1,0} + \phi_{1,1} X_{t-1}\} \times I(X_{t-1} \le r), \tag{4.3}$$

where the parameters and the innovations are those defined above. Our interest is in testing whether $\phi_{1,0} = \phi_{1,1} = 0$, i.e., the data are generated by the IMA(1,1) model with an intercept term:

$$H_0: \qquad X_t = \phi_0 + X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}. \tag{4.4}$$

If the intercept $\phi_0 \neq 0$, then the IMA(1,1) process superimposes on a linear trend. If no such linear trend is apparent in the data, it is reasonable to omit the intercept from the IMA(1,1) model. However, the intercept terms on the two regimes of any competing stationary TARMA(1,1) model will be required to model the mean of the data. Indeed, even for mean-deleted data, the intercept terms of the TARMA(1,1) model are not necessarily zero. Hence, the intercept terms are retained in the constrained TARMA model.

Under the null hypothesis, the threshold parameter is absent thereby complicating the test. Our approach is to develop a Lagrange multiplier test statistic for H_0 with the threshold parameter fixed at some r. Denote the test statistic as $T_{n,r}$. Since r is unknown and indeed undefined under H_0, we shall compute $T_{n,r}$ for all r over some data-driven interval, say, $[r_L, r_U]$ with the end points being some percentiles of the observed data. For instance, r_L is the 20 percentile and r_U the 80 percentile. Then the overall test statistic $T_n = \sup_{r\in[r_L,r_U]} T_{n,r}$. Besides taking supremum, other approaches including integration can be employed to derive an overall test statistic.

For fixed r, the Lagrange multiplier test is developed based on the Gaussian likelihood conditional on X_0:

$$\ell = -\log(\sigma^2 2\pi) \times n/2 - \sum_{t=1}^{n} \varepsilon_t^2/(2\sigma^2), \tag{4.5}$$

where

$$\varepsilon_t = X_t - [\phi_0 + X_{t-1} + \{\phi_{1,0} + \phi_{1,1}X_{t-1}\} \times I(X_{t-1} \le r)] + \theta\varepsilon_{t-1}. \tag{4.6}$$

Let $\boldsymbol{\psi} = (\phi_0, \theta, \sigma^2, \phi_{1,0}, \phi_{1,1})^\intercal$, with its components denoted by $\psi_j, j = 1, 2, \ldots, 5$. Denote by $\boldsymbol{\psi}_1$ the sub-vector comprising the first three components of $\boldsymbol{\psi}$, and $\boldsymbol{\psi}_2$ the sub-vector consisting of the remaining two components, i.e.:

$$\begin{aligned}\boldsymbol{\psi}_1 &= \left(\phi_0, \theta, \sigma^2\right)^\intercal \\ \boldsymbol{\psi}_2 &= (\phi_{1,0}, \phi_{1,1})^\intercal .\end{aligned}$$

Consider the score and the Fisher information matrix partitioned according to $\boldsymbol{\psi}_i, i = 1, 2$ into

$$\begin{aligned}\frac{\partial\ell}{\partial\boldsymbol{\psi}} &= \begin{pmatrix} \frac{\partial\ell}{\partial\boldsymbol{\psi}_1} \\ \frac{\partial\ell}{\partial\boldsymbol{\psi}_2} \end{pmatrix} \\ I_n(\tau) &= \begin{pmatrix} I_{1,1,n}(\tau) & I_{1,2,n}(\tau) \\ I_{2,1,n}(\tau) & I_{2,2,n}(\tau) \end{pmatrix}.\end{aligned}$$

The IMA(1,1) model under the null hypothesis can be estimated by solving the score equation $\frac{\partial\ell}{\partial\boldsymbol{\psi}_1} = 0$, yielding $\hat{\boldsymbol{\psi}}_1 = \hat{\boldsymbol{\psi}}_{1,n} = (\hat{\phi}_{0,n}, \hat{\theta}_n, \hat{\sigma}_n^2)^\intercal$. Let $\frac{\partial\hat{\ell}}{\partial\boldsymbol{\psi}_2}$ be equal to $\frac{\partial\ell}{\partial\boldsymbol{\psi}_2}$ evaluated at $\boldsymbol{\psi}_1 = \hat{\boldsymbol{\psi}}_1$ and $\boldsymbol{\psi}_2 = 0$. Similarly, $\hat{I}_{i,j,n}(\tau)$ is obtained by evaluating $I_{i,j,n}(\tau)$, $i, j = 1, 2$ at $\boldsymbol{\psi}_1 = \hat{\boldsymbol{\psi}}_1$ and $\boldsymbol{\psi}_2 = 0$.

Our test statistic is

$$T_n = \sup_{r\in[r_L,r_U]} T_{n,r}; \tag{4.7}$$

$$T_{n,r} = \left(\frac{\partial\hat{\ell}}{\partial\boldsymbol{\psi}_2}\right)^{\intercal}\left(\hat{I}_{2,2,n}(\tau) - \hat{I}_{2,1,n}(\tau)\hat{I}^{-1}_{1,1,n}(\tau)\hat{I}_{1,2,n}(\tau)\right)^{-1}\frac{\partial\hat{\ell}}{\partial\boldsymbol{\psi}_2}. \tag{4.8}$$

Some notation and conventions, to be adopted throughout, follow. Unless stated otherwise, all the expectations are taken under the true probability distribution for which H_0 holds. Let τ be such that $r = \sqrt{n}\sigma(1-\theta)\tau$ and $\{W_t, t = 1, \dots, n\}$ be the standard Wiener process. Set

$$I(\tau) = \begin{pmatrix} I_{1,1}(\tau) & I^{\intercal}_{2,1}(\tau) \\ I_{2,1}(\tau) & I_{2,2}(\tau) \end{pmatrix},$$

where

$$I_{1,1}(\tau) = \begin{pmatrix} \sigma^2(1-\theta)^2 & 0 & 0 \\ 0 & \cdot & 0 \\ 0 & 0 & \cdot \end{pmatrix};$$

$$I_{2,1}(\tau) = \begin{pmatrix} \frac{1}{(1-\theta)^2\sigma^2}\int_0^1 I(W_s \le \tau)ds & 0 & 0 \\ \frac{1}{(1-\theta)\sigma}\int_0^1 W_s I(W_s \le \tau)ds & 0 & 0 \end{pmatrix};$$

$$I_{2,2}(\tau) = \begin{pmatrix} \frac{1}{(1-\theta)^2\sigma^2}\int_0^1 I(W_s \le \tau)ds & \frac{1}{(1-\theta)\sigma}\int_0^1 W_s I(W_s \le \tau)ds \\ \frac{1}{(1-\theta)\sigma}\int_0^1 W_s I(W_s \le \tau)ds & \int_0^1 W_s^2 I(W_s \le \tau)ds \end{pmatrix}.$$

Moreover, define the following two matrices:

$$Q_n = \begin{pmatrix} n^{-1/2} & 0 \\ 0 & n^{-1} \end{pmatrix} \text{ and } P_n = \begin{pmatrix} n^{-1/2} & 0 & 0 \\ 0 & n^{-1/2} & 0 \\ 0 & 0 & \cdot \end{pmatrix}.$$

4.3.1 The asymptotic null distribution

We now derive the asymptotic distribution of both $T_{n,r}$ and T_n under the null hypothesis of an IMA(1,1) model with zero intercept:

$$X_t = X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}. \tag{4.9}$$

For simplicity, we assume Gaussian innovations. From (4.5), it follows that the components of the score vector are

$$\frac{\partial\ell}{\partial\boldsymbol{\psi}_1} = \left(\sum_{t=1}^{n}\frac{\varepsilon_t}{\sigma^2}\sum_{j=0}^{t-1}\theta^j, \sum_{t=1}^{n}\frac{\varepsilon_t}{\sigma^2}\sum_{j=0}^{t-1}\theta^j\varepsilon_{t-1-j}, \sum_{t=1}^{n}\frac{\varepsilon_t^2-\sigma^2}{\sigma^4}\right)^{\intercal},$$

$$\frac{\partial\ell}{\partial\boldsymbol{\psi}_2} = \left(\sum_{t=1}^{n}\frac{\varepsilon_t}{\sigma^2}\sum_{j=0}^{t-1}\theta^j I\left(\frac{X_{t-1-j}}{\sqrt{n}\sigma(1-\theta)} \le \tau\right), \sum_{t=1}^{n}\frac{\varepsilon_t}{\sigma^2}\sum_{j=0}^{t-1}\theta^j X_{t-1-j} I\left(\frac{X_{t-1-j}}{\sqrt{n}\sigma(1-\theta)} \le \tau\right)\right)^{\intercal},$$

while

$$\operatorname{diag}(I_{1,1,n}(\tau)) = \left(\sum_{t=1}^{n} \frac{1}{\sigma^2} \left(\sum_{j=0}^{t-1} \theta^j \right)^2, \sum_{t=1}^{n} \frac{1}{\sigma^2} \left(\sum_{j=0}^{t-1} \theta^j \varepsilon_{t-1-j} \right)^2, \sum_{t=1}^{n} \frac{(\varepsilon_t^2 - \sigma^2)^2}{4\sigma^8} \right)^{\intercal},$$

the first column of $I_{2,1,n}(\tau)$ is

$$\begin{pmatrix} \sum_{t=1}^{n} \frac{1}{\sigma^2} \left(\sum_{j=0}^{t-1} \theta^j \right) \left(\sum_{j=0}^{t-1} \theta^j I \left(\frac{X_{t-1-j}}{\sqrt{n}\sigma(1-\theta)} \leq \tau \right) \right) \\ \sum_{t=1}^{n} \frac{1}{\sigma^2} \left(\sum_{j=0}^{t-1} \theta^j \right) \left(\sum_{j=0}^{t-1} \theta^j X_{t-1-j} I \left(\frac{X_{t-1-j}}{\sqrt{n}\sigma(1-\theta)} \leq \tau \right) \right) \end{pmatrix};$$

$I_{2,2,n}(\tau)$ has the diagonal equal to

$$\begin{pmatrix} \sum_{t=1}^{n} \frac{1}{\sigma^2} \left(\sum_{j=0}^{t-1} \theta^j I \left(\frac{X_{t-1-j}}{\sqrt{n}\sigma(1-\theta)} \leq \tau \right) \right)^2 \\ \sum_{t=1}^{n} \frac{1}{\sigma^2} \left(\sum_{j=0}^{t-1} \theta^j X_{t-1-j} I \left(\frac{X_{t-1-j}}{\sqrt{n}\sigma(1-\theta)} \leq \tau \right) \right)^2 \end{pmatrix}$$

and the off-diagonal elements coincide and are equal to

$$\sum_{t=1}^{n} \frac{1}{\sigma^2} \left(\sum_{j=0}^{t-1} \theta^j I \left(\frac{X_{t-1-j}}{\sqrt{n}\sigma(1-\theta)} \leq \tau \right) \right) \left(\sum_{j=0}^{t-1} \theta^j X_{t-1-j} I \left(\frac{X_{t-1-j}}{\sqrt{n}\sigma(1-\theta)} \leq r \right) \right).$$

Note that, in the above derivation, we exploited the fact that, asymptotically, ε_t^2/σ^2 can be replaced by 1. The next proposition shows that, as n increases, the distribution of $\frac{\partial \hat{\ell}}{\partial \boldsymbol{\psi}_2}$ can be derived from those of $\frac{\partial \ell}{\partial \boldsymbol{\psi}_1}$ and $\frac{\partial \ell}{\partial \boldsymbol{\psi}_2}$. This is of fundamental importance in the development of the asymptotic theory since it allows us to replace $\hat{\boldsymbol{\psi}}$ with $\boldsymbol{\psi}$.

Proposition 55. *It holds that*

$$\begin{aligned} \frac{\partial \hat{\ell}}{\partial \boldsymbol{\psi}_2} &\approx (I_{2,2,n}^{-1}(\tau))^{-1} \left(I_{2,1,n}^{-1}(\tau) \frac{\partial \ell}{\partial \boldsymbol{\psi}_1} + I_{2,2,n}^{-1}(\tau) \frac{\partial \ell}{\partial \boldsymbol{\psi}_2} \right) \\ &= \frac{\partial \ell}{\partial \boldsymbol{\psi}_2} - I_{2,1,n}(\tau)(I_{1,1,n})^{-1}(\tau) \frac{\partial \ell}{\partial \boldsymbol{\psi}_1}, \end{aligned}$$

where $I_{2,2,n}^{-1}(\tau)$ denotes the $(2,2)$ block of the inverse of $I_n(\tau)$ that is partitioned according to $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ etc., and the RHS of the preceding equation is evaluated at the true value under the null hypothesis with $\phi_0 = 0$.

We now focus on the asymptotic distribution of $Q_n \frac{\partial \hat{\ell}}{\partial \boldsymbol{\psi}_2}$. Since

$$I_{2,1}(\tau) I_{1,1}^{-1}(\tau) = \begin{pmatrix} \frac{1}{(1-\theta)^2\sigma^2} \int_0^1 I(W_s \leq \tau) ds & 0 & 0 \\ \frac{1}{(1-\theta)\sigma} \int_0^1 W_s I(W_s \leq \tau) ds & 0 & 0 \end{pmatrix} \begin{pmatrix} \sigma^2(1-\theta)^2 & 0 & 0 \\ 0 & \cdot & 0 \\ 0 & 0 & \cdot \end{pmatrix}$$

$$= \begin{pmatrix} \int_0^1 I(W_s \leq \tau) ds & 0 & 0 \\ \sigma(1-\theta) \int_0^1 W_s I(W_s \leq \tau) ds & 0 & 0 \end{pmatrix},$$

the only elements that play a role in the limiting distribution are

$$\nabla_n^1(\tau) = \frac{1}{\sqrt{n}}\frac{\partial \ell}{\partial \phi_0} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\varepsilon_t}{\sigma}\sum_{j=0}^{t-1}\theta^j; \tag{4.10}$$

$$\nabla_n^2(\tau) = \frac{1}{\sqrt{n}}\frac{\partial \ell}{\partial \phi_{1,0}} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\varepsilon_t}{\sigma}\sum_{j=0}^{t-1}\theta^j I\left(\frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma}\leq\tau\right); \tag{4.11}$$

$$\nabla_n^3(\tau) = \frac{1}{n}\frac{\partial \ell}{\partial \phi_{1,1}} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\varepsilon_t}{\sigma}\sum_{j=0}^{t-1}\theta^j \frac{X_{t-1-j}}{\sqrt{n}\sigma} I\left(\frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma}\leq\tau\right). \tag{4.12}$$

Let $\mathcal{D}_{\mathbb{R}}(-b,b)$, $b>0$ be the space of functions from $(-b,b)$ to $\mathbb{R}$ that are right continuous with left-hand limits. $\mathcal{D}_{\mathbb{R}}(-b,b)$ is equipped with the topology of uniform convergence on compact sets. $\nabla_n^i(\tau)$, $i=1,2,3$ are random elements living in $\mathcal{D}_{\mathbb{R}}(-b,b)$. In order to establish their weak convergence, we need to prove the following propositions.

Proposition 56. *Under the null hypothesis, for every fixed τ, it holds that*

$$\nabla_n^1(\tau) \xrightarrow[n\to\infty]{d} \frac{1}{(1-\theta)\sigma}\int_0^1 dW_s, \tag{4.13}$$

$$\nabla_n^2(\tau) \xrightarrow[n\to\infty]{d} \frac{1}{(1-\theta)\sigma}\int_0^1 I(W_s\leq\tau)dW_s, \tag{4.14}$$

$$\nabla_n^3(\tau) \xrightarrow[n\to\infty]{d} \int_0^1 W_s I(W_s\leq\tau)dW_s. \tag{4.15}$$

Proposition 57. *Under the null hypothesis $\{\nabla_n^i(\tau), -b\leq\tau\leq b\}$, $i=1,2,3$, $b>0$ are tight.*

Now we are able to prove the following

Theorem 58. *Let $\nabla_n(\tau) = \left(\nabla_n^1(\tau),\nabla_n^2(\tau),\nabla_n^3(\tau)\right)^\intercal$ and*

$$\Xi(\tau) = \left(\frac{1}{(1-\theta)\sigma}\int_0^1 dW_s, \frac{1}{(1-\theta)\sigma}\int_0^1 I(W_s\leq\tau)dW_s, \int_0^1 W_s I(W_s\leq\tau)dW_s\right)^\intercal.$$

Under the null hypothesis, $\{\nabla_n(\tau)\}$ converges weakly to $\Xi(\tau)$ in the space $\mathcal{D}_{\mathbb{R}}(-b,b)$.

Proof. Propositions (56) and (57) allow to apply Theorem 18.14, p. 261 of van der Vaart [1998] so that the result follows. □

The following theorem allows to derive the weak limit of functions of I_n.

Theorem 59. *Let K_n be the 5 by 5 diagonal matrix where $diag(K_n) = (\sqrt{n},\sqrt{n},\sqrt{n},\sqrt{n},n)$. Under the null hypothesis, it holds that*

$$\sup_{\tau\in[a,b]}\|K_n^{-1}I_n(\tau)K_n^{-1} - I(\tau)\| \xrightarrow[n\to\infty]{p} 0,$$

where $-\infty < a < b < \infty$, *and* $\|\cdot\|$ *is the* L^2 *matrix norm (the Frobenius' norm, i.e.* $\|A\| = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m}|a_{ij}|^2}$, *where* A *is a* $n \times m$ *matrix.)*

The following theorem is the main result of the section. Theorems 58 and 59 are used to derive the asymptotic distribution of the supLM statistic under the null hypothesis.

Theorem 60. *Let* $H(\tau) = (\int_0^1 dW_s, \int_0^1 I(W_s \leq \tau)dW_s, \int_0^1 W_s I(W_s \leq \tau)dW_s)^\intercal$ *and*

$$\Lambda(\tau) = \begin{pmatrix} 1 & \int_0^1 I(W_s \leq \tau)ds & \int_0^1 W_s I(W_s \leq \tau)ds \\ \int_0^1 I(W_s \leq \tau)ds & \int_0^1 I(W_s \leq \tau)ds & \int_0^1 W_s I(W_s \leq \tau)ds \\ \int_0^1 W_s I(W_s \leq \tau)ds & \int_0^1 W_s I(W_s \leq \tau)ds & \int_0^1 W_s^2 I(W_s \leq \tau)ds \end{pmatrix}$$

Let Λ_τ *be partitioned into a* 2×2 *block matrix with the* $(2,2)$*-th block being* 2×2. *Similarly partitioned is* $H_\tau = (H_1(\tau), H_2(\tau))^\intercal$. *Then, the asymptotic null distribution of* $T_{n,r}$ *is the same as that of* $\left\|((\Lambda_\tau^{-1})_{2,2})^{1/2}\left(H_2(\tau) - \Lambda_{2,1}(\tau)H_1(\tau)\right)\right\|^2$ *where* $\|\cdot\|$ *is the Euclidean vector norm. Hence, the asymptotic null distribution of* $T_n = \sup\{T_{n,r}, r \in [\sqrt{n}(1+\theta)r_L, \sqrt{n}(1+\theta)r_U]\}$ *converges in distribution to*

$$\sup_{\tau\in[r_L,r_U]}\left\|((\Lambda_\tau^{-1})_{2,2})^{1/2}\left(H_2(\tau) - \Lambda_{2,1}(\tau)H_1(\tau)\right)\right\|^2$$

which is parameter-free under the null distribution with $\phi_0 = 0$.

Proof. The result readily follows by routine algebra. □

4.3.2 The distribution under local alternatives

In this section we derive the asymptotic distribution of the supLM statistic under a sequence of local alternatives and prove the consistency of associated test. For each n we have the null hypothesis
$H_{0,n}$: $(X_0, X_1, \ldots, X_n)$ follows the model

$$X_t = X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}.$$

versus the alternative hypothesis
$H_{1,n}$: $(X_0, X_1, \ldots, X_n)$ follows the model

$$X_t = \begin{cases} \frac{h_{1,0}}{\sqrt{n}} + \left(1 + \frac{h_{1,1}}{n}\right)X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} & \text{if } \frac{X_{t-1}}{\sigma\sqrt{n}(1-\theta)} \leq \tau_0 \\ \frac{h_{2,0}}{\sqrt{n}} + \left(1 + \frac{h_{2,1}}{n}\right)X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1} & \text{if } \frac{X_{t-1}}{\sigma\sqrt{n}(1-\theta)} > \tau_0. \end{cases}$$

where $\mathbf{h} = (h_{1,0}, h_{2,0}, h_{1,1}, h_{2,1})^\intercal$ is a fixed vector and τ_0 is a fixed scalar. In the following we develop the theory for $h_{2,0} = h_{2,1} = 0$ and, for notational convenience, we set $h_{1,0} = h_0$ and $h_{1,1} = h_1$. We constrain h_1 to be negative so $\{X_t\}$ is a stationary process in the lower regime. Moreover, we assume h_0 to be negative. Let $P_{0,n}$ and $P_{1,n}$ be the probability measures of $(X_0, X_1, \ldots, X_n)$ under H_0 and H_1, respectively.

Proposition 61. *It holds that:*

1. *If* $h_1 \in \left(-\frac{\pi}{2(1-\theta)}, 0\right)$ *then* $\{P_{1,n}\}$ *is contiguous to* $\{P_{0,n}\}$.

2. *Under the null hypothesis, as* n *increases, the log likelihood ratio* $\log \frac{dP_{1,n}}{dP_{0,n}}$ *converges to* $X_t - 1/2[X_t]$, *where* $[X_t]$ *is the quadratic variation of* X_t *and*

$$X_t = \int_0^t \left\{ \frac{h_0}{\sigma} I\left(W_s \le \tau_0\right) + (1-\theta) h_1 W_s I\left(W_s \le \tau_0\right) \right\} dW_s.$$

Let $\{W_t\}$ and $\{\tilde{W}_t\}$ be the $P_{0,n}$-Wiener process and $P_{1,n}$-Wiener process, respectively. In the following proposition we derive the distribution of $\{W_t\}$ under $P_{1,n}$.

Proposition 62. *Assume that* H_1 *holds. Then,* W_t *satisfies the following threshold stochastic differential equation on* $[0,1]$*:*

$$\begin{aligned} dW_t &= d\tilde{W}_t + \left\{ \frac{h_0}{\sigma} I\left(W_t \le \tau_0\right) + (1-\theta) h_1 W_t I\left(W_t \le \tau_0\right) \right\} dt \\ &= \begin{cases} d\tilde{W}_t + \left[\frac{h_0}{\sigma} + (1-\theta) h_1 W_t\right] dt & \text{if } W_t \le \tau_0 \\ d\tilde{W}_t & \text{if } W_t > \tau_0 \end{cases} \end{aligned}$$

Let $\tilde{\theta} = -(1-\theta)h_1$ and $\tilde{\mu} = -\frac{h_0}{\sigma(1-\theta)h_1}$. In the following Corollary, we prove that as $\tilde{\mu} \to -\infty$, $\{W_t\}$ behaves as a stochastic differential equation without threshold.

Corollary 63. *Suppose* $\tilde{\mu} \to -\infty$ *and* $\tau_0 > 0$. *Then, except for an asymptotically negligible event,* W_t *is the following Ornstein-Uhlenbeck process over* $[0,1]$*:*

$$dW_t = \tilde{\theta}\left(\tilde{\mu} - W_t\right) dt + d\tilde{W}_t. \tag{4.16}$$

Note that, for each t,

$$W_t = \tilde{\mu}(1 - e^{-\tilde{\theta}t}) + \int_0^t e^{-\tilde{\theta}(t-s)} d\tilde{W}_t \sim N\left(\tilde{\mu}(1 - e^{-\tilde{\theta}t}), \frac{1 - e^{-2\tilde{\theta}t}}{2\tilde{\theta}\tilde{\mu}^2}\right)$$

. Now, consider the SDE 4.16. The normalized process $\left\{\frac{W_t}{|\tilde{\mu}|}\right\}$ satisfies the following equation:

$$d\frac{W_t}{|\tilde{\mu}|} = -\tilde{\theta}\left(1 + \frac{W_t}{|\tilde{\mu}|}\right) dt + d\frac{\tilde{W}_t}{|\tilde{\mu}|}. \tag{4.17}$$

Proposition 64. *Define the deterministic process* $\{\mathcal{G}_t\}$, *such that*

$$\mathcal{G}_t = e^{-\tilde{\theta}t} - 1.$$

It holds that

$$\left\{\frac{W_t}{|\tilde{\mu}|}, 0 \le t \le 1\right\} \xrightarrow[|\tilde{\mu}|\to\infty]{p} \{\mathcal{G}_t, 0 \le t \le 1\} \tag{4.18}$$

i.e.

$$\sup_{0\le t\le 1}\left|\frac{W_t}{|\tilde{\mu}|} - \mathcal{G}_t\right| \xrightarrow[|\tilde{\mu}|\to\infty]{p} 0.$$

Now, consider our test statistic

$$T = \sup_{\tau\in[r_L, r_U]} L(\tau)^\intercal \Delta(\tau)^{-1} L(\tau),$$

where

$$\begin{aligned}
L(\tau) &= (H_2(\tau) - \Lambda_{2,1}(\tau)H_1(\tau))^\intercal\,;\\
\Delta(\tau) &= \left(\Lambda_{22}(\tau) - \Lambda_{21}(\tau)\Lambda_{11}^{-1}(\tau)\Lambda_{12}(\tau)\right);\\
H(\tau) &= \left(\int_0^1 dW_s, \int_0^1 I(W_s \le \tau)dW_s, \int_0^1 W_s I(W_s \le \tau)dW_s\right)^\intercal;\\
\Lambda_{22}(\tau) &= \begin{pmatrix} \int_0^1 I(W_s \le \tau)ds & \int_0^1 W_s I(W_s \le \tau)ds \\ \int_0^1 W_s I(W_s \le \tau)ds & \int_0^1 W_s^2 I(W_s \le \tau)ds \end{pmatrix};\\
\Lambda_{21}(\tau) &= \Lambda_{12}(\tau)^\intercal = \left(\int_0^1 I(W_s \le \tau)ds, \int_0^1 W_s I(W_s \le \tau)ds\right)^\intercal;\\
\Lambda_{11}(\tau) &= (1).
\end{aligned}$$

Proposition 65. *Let* $q_i\left(\frac{W_s}{|\tilde{\mu}|}, 0 \le s \le 1\right)$ *and* $q_i\left(\mathcal{G}_s, 0 \le s \le 1\right)$, $i = 1, 2$, *be two quantiles of the processes* $\left\{\frac{W_s}{|\tilde{\mu}|}\right\}$ *and* $\{\mathcal{G}_s\}$, *respectively. It holds that*

$$q_i\left(\frac{W_s}{|\tilde{\mu}|}, 0 \le s \le 1\right) \xrightarrow[|\tilde{\mu}|\to\infty]{p} q_i\left(\mathcal{G}_s, 0 \le s \le 1\right).$$

Moreover,

$$\begin{aligned}
\int_0^1 I(W_s \le \tau)ds &\xrightarrow[|\tilde{\mu}|\to\infty]{p} \int_0^1 I(\mathcal{G}_s \le \tilde{\tau})ds;\\
\int_0^1 \frac{W_s}{|\tilde{\mu}|} I(W_s \le \tau)ds &\xrightarrow[|\tilde{\mu}|\to\infty]{p} \int_0^1 \mathcal{G}_s I(\mathcal{G}_s \le \tilde{\tau})ds;\\
\int_0^1 \frac{W_s^2}{|\tilde{\mu}|^2} I(W_s \le \tau)ds &\xrightarrow[|\tilde{\mu}|\to\infty]{p} \int_0^1 \mathcal{G}_s^2 I(\mathcal{G}_s \le \tilde{\tau}),
\end{aligned}$$

where, given a constant $\tilde{s}$*:*

$$\begin{aligned}
\tau &= \tilde{s}q_1\left(\frac{W_s}{|\tilde{\mu}|}, 0 \le s \le 1\right) + (1-\tilde{s})q_2\left(\frac{W_s}{|\tilde{\mu}|}, 0 \le s \le 1\right)\\
\tilde{\tau} &= \tilde{s}q_1\left(\mathcal{G}_s, 0 \le s \le 1\right) + (1-\tilde{s})q_2\left(\mathcal{G}_s, 0 \le s \le 1\right).
\end{aligned}$$

Let

$$a = \int_0^1 I\left(\mathcal{G}_s \leq \tilde{\tau}\right) ds; \quad b = \mathcal{G}_s I\left(\mathcal{G}_s \leq \tilde{\tau}\right) ds; \quad c = \mathcal{G}_s^2 I\left(\mathcal{G}_s \leq \tilde{\tau}\right) ds.$$

In the following proposition we establish the orders of the various quantities defining the test statistic 4.7.

Proposition 66. *It holds that:*

$$\int_0^1 I(W_s \leq \tau) ds = O(1); \quad \int_0^1 W_s I(W_s \leq \tau) ds = O(|\tilde{\mu}|); \quad \int_0^1 W_s^2 I(W_s \leq \tau) ds = O(|\tilde{\mu}|^2);$$

$$\int_0^1 I\left(W_s \leq \tau\right) dW_s - \left[\int_0^1 I\left(W_s \leq \tau\right) ds\right] \int_0^1 dW_s = O_p\left(|\tilde{\mu}|^{-1}\right);$$

$$\int_0^1 W_s I\left(W_s \leq \tau\right) dW_s - \left[\int_0^1 W_s I\left(W_s \leq \tau\right) ds\right] \int_0^1 dW_s = O_p\left(|\tilde{\mu}|^{-2}\right).$$

Moreover, let

$$R_{\tilde{\mu}} = \begin{pmatrix} 1 & 0 \\ 0 & \tilde{\mu} \end{pmatrix},$$

then:

$$\lim_{\tilde{\mu} \to -\infty} R_{\tilde{\mu}} \left(\Lambda_{22}(\tau) - \Lambda_{21}(\tau)\Lambda_{11}^{-1}(\tau)\Lambda_{12}(\tau)\right)^{-1} R_{\tilde{\mu}} = \tilde{\Delta}(\tau) = \begin{pmatrix} \frac{c-b^2}{(1-a)(ac-b^2)} & -\frac{b}{ac-b^2} \\ -\frac{b}{ac-b^2} & \frac{a}{ac-b^2} \end{pmatrix},$$

and $\tilde{\Delta}(\tau)$ is invertible.

Based on the above results we are able to assess the asymptotic behaviour of the supLM statistic and prove the consistency of the associated test.

Theorem 67. *It holds that, under the local alternatives, as $\tilde{\mu} \to -\infty$, the statistic T_n diverges.*

4.3.3 A wild bootstrap approach

In this section we introduce a wild bootstrap version of our supLM statistic. This bootstrap scheme has proved to deliver valid inference under heteroskedastic disturbances Liu [1988], Mammen [1993], Davidson and Flachaire [2008]. As also shown in Cavaliere and Taylor [2008] in the context of unit root testing, the wild bootstrap is able to correctly reproduce the first-order limiting null distribution of the statistics in case of non-stationary volatility. The algorithm has the following structure:

1. Compute $\tilde{X}_t = X_t - \hat{\boldsymbol{\beta}}^\intercal \mathbf{d_t}$, where $\mathbf{d_t}$ is a vector of deterministic components and $\hat{\boldsymbol{\beta}}$ is obtained through either OLS or GLS detrending;

2. Obtain $\hat{\theta}$, the maximum likelihood estimate for θ and the residuals $\hat{e}_t$ from the following IMA(1,1) model: $\tilde{X}_t = \tilde{X}_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}$;

3. Compute wild bootstrap errors: $\hat{e}_t^* = \hat{e}_t \eta_t$, where η_t is a random variable such that $E(\eta_t) = 0$ and $E(\eta_t^2) = 1$. In this paper we use the Rademacher variable $\eta_t = (1, -1)$ with probability $(1/2, 1/2)$. We also experimented with η_t being standard Gaussian leading to no significant differences.

4. Obtain the bootstrap sample

$$\hat{X}_t^* = \sum_{j=i}^{t} \left(\hat{e}_j^* - \hat{\theta} \hat{e}_{j-1}^* \right),$$

and compute the supLM bootstrap statistic T_n^* upon it.

5. Repeat steps 3–4, B times as to obtain replications of the bootstrap test statistic, T_n^{*b}, $b = 1, \ldots B$ and derive the bootstrap p-value as

$$B^{-1} \sum_{b=1}^{B} I(T_n^{*b} \geq T_n),$$

where $I(\cdot)$ is the indicator function and T_n is the value of the supLM statistic computed on the original sample.

4.4 Finite sample performance

In this section we present a Monte Carlo simulation to investigate the finite sample performance of our supLM tests and compare them with existing unit-root tests. We have simulated from a plethora of data generating processes (hereafter DGPs), both linear and non-linear. In Table 4.1 we present the set of integrated DGPs used to assess the empirical size of the tests. Models 1–7 are linear integrated processes. In particular, models 2,4,6,7 are linear processes close to non-invertibility and model 6 is an ARIMA(1,1,1) process where there is near-cancellation of the MA and AR polynomials. Models 8–12 are non-linear integrated processes. Table 4.2 reports the stationary DGPs used to compute the empirical power. Models 15–20 are linear and stationary; in particular, models 19–20 are the stationary counterparts of models 6–7. Models 21–24 are two-regime TAR models whereas models 25–26 follow a three-regime TAR. Models 27 and 28 are the heteroskedastic counterparts of models 22 and 23, respectively. Also, models 29–32 are the stationary versions of models 9–12. The upper percentage points of the asymptotic null distribution of the test statistic T_n can be obtained by simulation. Since we have proved the asymptotic similarity of the test, we derived the quantiles of the distribution of T_n from 20000 independent replications of an IMA(1,1) model of sample size 5000, with $\theta = \phi_0 = 0$ and iid standard Gaussian innovations, with the threshold searched between the 15th and 85th percentiles. In practice, since the finite sample distribution of T_n might depend upon θ and differ from the asymptotic one we have simulated the null distributions for the sample sizes in use. Moreover, since we have found that the finite sample distribution of T_n

linear	
01. RW	$X_t = X_{t-1} + \varepsilon_t$
02. IMA11.1	$\Delta X_t = \varepsilon_t - 0.9\varepsilon_{t-1}$
03. IMA11.2	$\Delta X_t = \varepsilon_t + 0.5\varepsilon_{t-1}$
04. IMA11.3	$\Delta X_t = \varepsilon_t + 0.9\varepsilon_{t-1}$
05. ARI5.1	$\Delta X_t = -0.6\Delta X_{t-1} - 0.4\Delta X_{t-2} - 0.3\Delta X_{t-3} - 0.4\Delta X_{t-4} - 0.5\Delta X_{t-5} + \varepsilon_t$
06. ARIMA111.1	$\Delta X_t = 0.7\Delta X_{t-1} - 0.9\varepsilon_{t-1} + \varepsilon_t$
07. ARIMA111.2	$\Delta X_t = 0.7\Delta X_{t-1} + 0.9\varepsilon_{t-1} + \varepsilon_t$
non-linear	
08. TARI1.1	$\Delta X_t = \begin{cases} -2X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} \leq r \\ 0.3X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} > r \end{cases}$
09. ARI-GARCH.1	$X_t = X_{t-1} + \varepsilon_t \sigma_t$, where $\sigma_t^2 = 0.05 + 0.90\varepsilon_{t-1}^2 + 0.05\sigma_{t-1}^2$
10. ARI-GARCH.2	$X_t = X_{t-1} + \varepsilon_t \sigma_t$, where $\sigma_t^2 = 0.05 + 0.30\varepsilon_{t-1}^2 + 0.65\sigma_{t-1}^2$
11. NLIMA.1	$\Delta X_t = \varepsilon_t - 0.8\varepsilon_{t-1}^2$
12. NLIMA.2	$\Delta X_t = \varepsilon_t + 0.8\varepsilon_{t-1}^2$

Table 4.1: Integrated data generating processes used to investigate the size of the tests. Unless otherwise stated $r = 0$ and $\{\varepsilon_t\}$ follows a standard Gaussian white noise.

changes appreciably only when $|\theta|$ is close to one we have adopted the following, conservative, approach: if $|\hat{\theta}| > 0.3$ we use the quantiles of the simulated null with $\theta = \text{sign}(\hat{\theta}) \cdot 0.9$. We denote our asymptotic test with sLM. The wild bootstrap version of the test is denoted with sLM.b. The set of competing tests can be divided in those whose alternative is a threshold autoregressive model and those that do not specify explicitly a non-linear alternative. As for the former set, we have implemented the tests by *i*) Kapetanios and Shin [2006] (KS): both the asymptotic and the bootstrap version. Of the three statistics proposed we report in the paper the one suggested by the authors, i.e. the average of the exponential of the Wald statistic over the threshold range. The results for remaining statistics are reported in the supplementary material; *ii*) Enders and Granger [1998] (EG): we have implemented the two statistics for the estimated constant and report the results for panel C, whereas the results for the statistic relative to panel D are relegated to the supplementary material; *iii*) Bec et al. [2004] (BBC): we have implemented the suggested asymptotic supLR statistic and a bootstrap version of it along the scheme detailed in Kapetanios and Shin [2006]. Note that the bootstrap test is a novel implementation not present in literature. We have decided not to include the tests by Caner and Hansen [2001] since the threshold variable there is taken as the first difference of X_t. Preliminary investigations showed that their performance is similar to the tests by Bec et al. [2004]. The second set of implemented unit-root tests includes the tests by: *i*) Dickey and Fuller [1979]: the ADF test; *ii*) Ng and Perron [2001]: the class of M tests with GLS detrending and MIC criterion for selecting the lags of the ADF regression. In particular, we implemented all the tests proposed and reviewed in the aforementioned article and report the $\bar{M}Z_\alpha^{\text{GLS}}$ which is essentially the M test proposed in Perron and Ng [1996] where the data have been GLS detrended and the MAIC criterion is used (see also Eq. (3) in Ng and Perron [2001]). We denote this test as $\bar{M}^{\text{GLS}}$. Also, we report the results for the $\bar{M}P_t^{\text{GLS}}$ the modified feasible point optimal tests (see also Eq. (9) in Ng and Perron [2001]) and the GLS detrended version of the ADF test (denoted with ADF^{GLS}); *iii*) The test MZ_α^{GLS} as proposed by Perron and Qu [2007], which is essentially the same as $\bar{M}Z_\alpha^{\text{GLS}}$ but the lag of the ADF regression is selected upon OLS detrended data. We denote it with M^{GLS}. We selected the above tests by virtue of their good performance, the results for the remaining tests can be found in the supplementary material.

linear	
13. AR1.1	$X_t = -0.9X_{t-1} + \varepsilon_t$
14. AR1.2	$X_t = \quad 0.9X_{t-1} + \varepsilon_t$
15. AR1.3	$X_t = -0.6X_{t-1} + \varepsilon_t$
16. AR1.4	$X_t = \quad 0.6X_{t-1} + \varepsilon_t$
17. ARMA11.1	$X_t = 0.7X_{t-1} - 0.9\varepsilon_{t-1} + \varepsilon_t$
18. ARMA11.2	$X_t = 0.7X_{t-1} + 0.9\varepsilon_{t-1} + \varepsilon_t$
non-linear	
19. TAR1.1	$X_t = \begin{cases} 0.6X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} \leq r \\ 0.35X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} > r \end{cases}$
20. TAR1.2	$X_t = \begin{cases} 0.6X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} \leq r \\ -0.35X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} > r \end{cases}$
21. TAR1.3	$X_t = \begin{cases} -2X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} \leq r \\ 0.3X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} > r \end{cases}$
22. TAR3.1	$X_t = \begin{cases} 0.3X_{t-1} - 0.7X_{t-2} + 0.6X_{t-3} + \varepsilon_t, & \text{if } X_{t-1} \leq r \\ -0.3X_{t-1} + 0.7X_{t-2} - 0.6X_{t-3} + \varepsilon_t, & \text{if } X_{t-1} > r \end{cases}$
23. 3TAR1.1	$X_t = \begin{cases} 0.3 + 0.5X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} \leq -1 \\ 0.3 + X_{t-1} + \varepsilon_t, & \text{if } -1 > X_{t-1} \leq 1 \\ 0.3 + 0.5X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} > 1 \end{cases}$
24. 3TAR2.1	$X_t = \begin{cases} -3.9 + X_{t-1} - 0.3X_{t-2} + \varepsilon_t, & \text{if } X_{t-1} \leq -10 \\ 1.3X_{t-1} - 0.3X_{t-2} + \varepsilon_t, & \text{if } -10 > X_{t-1} \leq 10 \\ 3.9 + X_{t-1} - 0.3X_{t-2} + \varepsilon_t, & \text{if } X_{t-1} > 10 \end{cases}$
25. TAR1h.1	$X_t = \begin{cases} 0.6X_{t-1} + 1 \cdot \varepsilon_t, & \text{if } X_{t-1} \leq r \\ -0.35X_{t-1} + 1.5 \cdot \varepsilon_t, & \text{if } X_{t-1} > r \end{cases}$
26. TAR1h.2	$X_t = \begin{cases} -2X_{t-1} + 1 \cdot \varepsilon_t, & \text{if } X_{t-1} \leq r \\ 0.3X_{t-1} + 1.5 \cdot \varepsilon_t, & \text{if } X_{t-1} > r \end{cases}$
27. NLMA.1	$X_t = \varepsilon_t - 0.8\varepsilon_{t-1}^2$
28. NLMA.2	$X_t = \varepsilon_t + 0.8\varepsilon_{t-1}^2$
29. AR-GARCH.1	$X_t = 0.9\,X_{t-1} + \varepsilon_t\,\sigma_t, \quad \text{where } \sigma_t^2 = 0.05 + 0.90\varepsilon_{t-1}^2 + 0.05\sigma_{t-1}^2$
30. AR-GARCH.2	$X_t = 0.9\,X_{t-1} + \varepsilon_t\,\sigma_t, \quad \text{where } \sigma_t^2 = 0.05 + 0.30\varepsilon_{t-1}^2 + 0.65\sigma_{t-1}^2$

Table 4.2: Stationary data generating processes used to investigate the power of the tests. Unless otherwise stated $r = 0$ and $\{\varepsilon_t\}$ follows a standard Gaussian white noise.

The sample sizes considered are 100, 300 and 500. The rejection percentages are derived with a nominal size $\alpha = 0.05$ and based upon 1000 replications and $B = 500$ bootstrap resamples for the bootstrap tests. Preliminary experiments proved that these are sufficient to ensure Monte Carlo stability in our setting. In Table 4.3 we show the empirical size of the tests, derived from the rejection percentages computed upon the integrated processes of Table 4.1. The table for $n = 500$ does not differ significantly from that with $n = 300$ and is reported in the appendix. Notably, all the tests that have a TAR model as the alternative hypothesis (last five columns of the table) break down completely in several cases, especially for models 2, 5, 6, 7, not only when the MA parameter is close to -1. Moreover, note that the bias gets worse for increasing sample size. These results suggest that such tests cannot be used without previous knowledge or a preliminary investigation upon the data at hand. Our supLM tests are generally well behaved in terms of size. In particular, the asymptotic test sLM is oversized only for the conditional heteroskedastic models 08-09. The wild bootstrap version sLM.b is never oversized. The three M tests are generally comparable in size, except for model 02 (IMA(1,1) model with $\theta = -0.9$) for which the M^{GLS} test of Perron and Qu [2007] results oversized for $n = 100$. Also, the ADF test is severely oversized for models 02 and 05 and the GLS version partly overcomes the problem, at least for model 05. As mentioned earlier, model 06 is an ARIMA(1,1,1) model with near cancellation of the AR and MA polynomials. This is the instance where all the tests fail to achieve the correct size.

The empirical power of the test is presented in Table 4.4 that reports the rejection percentages computed on the stationary processes listed in Table 4.2 for $n = 300$. The cases $n = 100, 500$ are reported in the appendix. Notably, the supLM test is uniformly more powerful than the M tests $\bar{M}^{\mathrm{GLS}}$, MP_T and the $A\bar{D}F^{\mathrm{GLS}}$ test, except for the ARGARCH model (models 29,30). Our test is almost always more powerful than the M^{GLS} of Perron and Qu [2007], both for linear and for threshold processes, except for the ARMA(1,1) model with near cancellation and for model 20. Our test is less powerful than the M^{GLS} for the non-linear MA (models 25 and 26), a borderline case where the representation in terms of a TARMA process might not provide a good approximation. In general, our tests seems to lose power in presence of heteroskedasticity (models 27–30), even though this depends on the values of the parameters. In fact, for models 28 and 30 the power is superior or comparable to that of the M tests. An important evidence emerging from the results is that our supLM tests do have power against all the alternatives and in every setting. This is not the case for the M tests; for instance, for model 1, the tests $\bar{M}^{\mathrm{GLS}}$, MP_T and $A\bar{D}F^{\mathrm{GLS}}$ have practically zero power even for $n = 300$. This is even more evident when the DGP is a TARMA model, as shown in the next section.

$n = 100$	sLM	sLM.bI	sLM.bT	$\bar{M}^{\text{GLS}}$	M^{GLS}	MP_T	ADF	ADF$^{\text{GLS}}$	KSe	KSe.b	BBC	BBC.b	EGc
01. RW	5.1	5.5	5.2	4.3	4.4	4.0	4.4	4.6	7.5	5.0	3.2	6.0	6.2
02. IMA11.1	6.3	5.4	7.8	7.2	18.5	7.2	74.5	17.6	100.0	99.7	91.2	94.0	100.0
03. IMA11.2	4.4	4.7	3.7	5.5	5.3	4.8	3.1	4.7	5.7	4.4	7.3	11.4	4.5
04. IMA11.3	5.3	6.0	7.2	8.0	6.7	6.9	2.7	3.4	7.1	5.0	11.5	15.2	7.7
05. ARI5.1	1.1	5.3	3.6	0.6	1.6	0.6	49.0	5.4	93.8	89.5	42.6	52.2	94.2
06. ARIMA111.1	18.4	22.4	21.3	27.4	33.2	25.6	21.3	29.3	59.2	48.9	13.8	21.1	61.2
07. ARIMA111.2	7.4	6.3	6.0	11.3	9.9	10.1	2.5	3.1	37.3	32.9	24.1	29.8	29.8
08. ARIGARCH.1	18.2	4.8	9.7	5.1	5.1	4.2	5.1	4.7	14.5	10.7	26.1	27.2	11.2
09. ARIGARCH.2	11.6	5.2	7.7	5.7	5.8	4.9	7.4	5.6	13.6	10.0	11.8	14.4	11.0
10. TARI1.1	1.3	1.3	1.6	1.2	1.1	1.2	3.9	0.0	99.8	99.3	0.3	0.6	99.7
11. NLIMA.1	8.0	4.8	6.9	0.4	0.1	0.1	5.1	0.0	100.0	99.8	5.0	6.7	97.7
12. NLIMA.2	6.9	3.0	5.3	0.1	0.0	0.1	4.7	0.0	99.9	99.6	4.4	5.3	97.8
$n = 300$	sLM	sLM.bI	sLM.bT	$\bar{M}^{\text{GLS}}$	M^{GLS}	MP_T	ADF	ADF$^{\text{GLS}}$	KSe	KSe.b	BBC	BBC.b	EGc
01. RW	4.2	4.2	4.1	5.3	5.3	5.1	4.2	5.6	5.5	3.8	3.8	5.8	4.3
02. IMA11.1	4.0	3.6	2.0	2.0	2.6	2.0	87.3	15.8	100.0	100.0	99.4	99.6	100.0
03. IMA11.2	2.7	4.5	4.5	4.2	4.2	3.8	4.3	3.8	4.9	3.0	6.7	10.2	4.9
04. IMA11.3	6.8	7.3	8.9	8.1	7.6	7.4	3.2	6.1	6.0	4.1	15.9	19.7	6.3
05. ARI5.1	2.5	5.6	5.7	0.6	0.6	0.9	4.1	4.5	98.0	94.9	72.2	77.5	96.1
06. ARIMA111.1	50.1	52.1	51.6	18.2	18.6	17.1	43.8	19.2	82.4	75.6	44.6	54.8	81.7
07. ARIMA111.2	7.7	7.2	6.8	7.1	6.4	6.3	3.5	3.7	32.6	27.7	29.0	34.6	28.4
08. ARIGARCH.1	23.0	5.9	8.8	5.1	5.1	4.6	4.9	4.8	16.9	12.9	31.5	29.1	10.8
09. ARIGARCH.2	14.1	5.9	6.5	6.4	6.3	6.2	6.8	5.8	12.3	8.4	18.5	20.3	9.3
10. TARI1.1	1.8	1.9	2.6	0.0	0.0	0.0	4.7	0.0	100.0	100.0	0.1	0.2	100.0
11. NLIMA.1	4.7	2.6	3.3	0.0	0.0	0.0	5.2	0.0	100.0	100.0	2.2	2.1	100.0
12. NLIMA.2	4.0	2.1	2.7	0.0	0.0	0.0	4.6	0.0	100.0	100.0	1.8	1.7	100.0

Table 4.3: Empirical size at $\alpha = 0.05$. Rejection percentages from integrated DGPs.

$n = 300$	sLM	sLM.bI	sLM.bT	$\bar{M}^{\text{GLS}}$	M^{GLS}	MP_T	ADF	ADF$^{\text{GLS}}$	KSe	KSe.b	BBC	BBC.b	EGc
13. AR1.1	99.5	100.0	100.0	0.8	28.2	0.8	100.0	10.5	100.0	100.0	100.0	100.0	100.0
14. AR1.2	82.9	83.6	83.2	78.1	82.1	77.0	73.4	78.1	99.9	99.4	36.1	44.5	99.4
15. AR1.3	86.1	84.7	97.8	5.9	40.4	5.7	100.0	22.4	100.0	100.0	100.0	100.0	100.0
16. AR1.4	100.0	100.0	100.0	54.3	80.4	53.9	100.0	68.3	100.0	100.0	100.0	100.0	100.0
17. ARMA11.1	96.5	80.1	85.7	14.8	100.0	14.8	100.0	30.0	100.0	100.0	100.0	100.0	100.0
18. ARMA11.2	100.0	100.0	100.0	63.6	74.4	63.2	98.8	70.1	100.0	100.0	100.0	100.0	100.0
19. TAR1.1	97.4	98.1	98.6	46.0	77.9	45.6	100.0	61.0	100.0	100.0	100.0	100.0	100.0
20. TAR1.2	55.3	50.6	68.8	32.1	71.0	31.2	100.0	53.3	100.0	100.0	100.0	100.0	100.0
21. TAR1.3	85.2	82.6	97.7	11.7	76.5	11.8	100.0	36.8	100.0	100.0	100.0	100.0	100.0
22. TAR3.1	65.7	49.4	41.1	10.5	29.5	10.4	98.6	39.8	100.0	100.0	100.0	100.0	100.0
23. 3TAR1.1	99.6	99.7	99.7	47.7	79.6	46.8	100.0	62.8	100.0	100.0	100.0	100.0	100.0
24. 3TAR2.1	84.6	83.7	83.2	19.6	21.8	19.2	24.3	19.9	65.8	58.5	95.1	96.2	39.1
25. NLMA.1	43.5	30.7	59.4	22.3	66.5	22.3	100.0	48.3	100.0	100.0	100.0	100.0	100.0
26. NLMA.2	46.8	30.9	58.2	20.9	67.2	20.7	100.0	48.5	100.0	100.0	100.0	100.0	100.0
27. TAR1h.1	52.6	42.6	66.7	27.5	71.5	27.8	100.0	49.3	100.0	100.0	100.0	100.0	100.0
28. TAR1h.2	86.9	85.8	97.4	9.7	70.8	9.9	100.0	32.0	100.0	100.0	100.0	100.0	100.0
29. ARGARCH.1	64.1	28.2	28.8	81.7	83.9	79.9	70.9	82.0	97.3	95.3	66.0	64.3	94.3
30. ARGARCH.2	73.2	46.9	46.8	76.3	79.2	74.8	70.6	76.5	98.6	97.3	59.7	63.3	96.4

Table 4.4: Empirical power at $\alpha = 0.05$. Rejection percentages from stationary DGPs with $n = 300$.

4.4.1 TARMA models

We simulated data from the following TARMA(1,1) model:

$$X_t = \begin{cases} \phi_{1,0} + \phi_{1,1}X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, & \text{if } X_{t-d} \leq r \\ \phi_{2,0} + \phi_{2,1}X_{t-1} + \varepsilon_t - \theta\varepsilon_{t-1}, & \text{otherwise,} \end{cases} \tag{4.19}$$

where $(\phi_{1,0}, \phi_{1,1}, \phi_{2,0}, \phi_{2,1}) = t \times (0, 0.7, -0.02, 0.99) + (1 - t) \times (0, 1, 0, 1)$ with t increasing from 0 to 1.5 with increments 0.5. When $t = 0$, the model is an IMA(1,1) model with zero intercept, while the model becomes a stationary TARMA(1,1) model with $t > 0$, of increasing disparity from the IMA(1,1) model with increasing t. The empirical size of the tests is displayed in Table 4.5. Note that in this instance we have partitioned the set of tests according to their nature: either asymptotic or bootstrap. Moreover, we have increased the number of the Monte Carlo replications to 10000 for the asymptotic tests. As also shown in Section 4.4, the ADF, the KS, the BBC and the EG tests are severely oversized also for $n = 500$. Clearly, the sLM.b test is the only test that shows a correct size in all the settings, whereas the both sLM and the M class of tests show some bias, albeit small.

The size-corrected power and of the asymptotic tests and the empirical power of the bootstrap tests is presented in Table 4.6. The results for $n = 500$ are similar to the those for $n = 300$ and are reported in the appendix. Note that, for the sake of readability, we have kept the rows corresponding to $t = 0$, that contain the empirical size for the bootstrap tests and the corrected 5% size otherwise. The case $n = 300$ (lower panel) shows that our supLM tests are always more powerful than the other tests, especially as t increases. For instance, when $t = 1.5$ the sLM test has almost double the power of M tests. The same trend is observed for $n = 100$ when t is large but the power is comparable to that of the M tests for smaller values of t. The power (be it size-corrected or bootstrap) of the ADF, KS, BBC, EG tests reflects the severe size bias and should not be considered.

4.5 A real application: testing the PPP hypothesis

In this section we apply our supLM tests to the post-Bretton Woods and pre-euro real exchange rates of a panel of European countries. The idea is to contribute to the widely debated issue of the power of purchase parity (PPP) and show that the TARMA model can be a useful tool to this aim. As mentioned in the introduction, based on macroeconomic theory, there is some consensus on the fact that price gaps (measured in a common currency) for the same goods in different countries should rapidly disappear. However, the empirical evidence points to a strong persistence and unit root tests generally fail to reject the null hypothesis of a random walk. As also pointed out in Taylor [2001] this can be ascribed to two factors. First, the way economic data are produced or aggregated may lead a severe bias in the inference based upon them. This is

	asymptotic									bootstrap			
θ	sLM	$\bar{M}^{\mathrm{GLS}}$	M^{GLS}	MP_T	ADF	$\mathrm{ADF}^{\mathrm{GLS}}$	KSe	BBC	EGc	sLM.bI	sLM.bT	KSe.b	BBC.b
$n = 100$													
-0.9	2.2	7.7	7.0	7.1	2.5	3.8	8.1	11.2	7.1	5.1	7.3	4.9	16.0
0.5	1.7	5.6	5.9	5.1	6.7	7.4	64.5	10.2	57.5	5.2	4.9	58.6	16.4
0.9	11.3	6.5	17.7	6.4	77.9	17.8	100.0	92.4	100.0	5.7	8.5	99.8	95.5
$n = 300$													
-0.9	5.5	6.7	6.3	6.1	3.3	4.2	6.3	14.0	6.5	5.3	7.9	3.8	20.6
0.5	2.3	5.5	5.4	5.1	5.4	5.8	74.5	19.0	61.1	4.9	4.7	67.7	23.8
0.9	4.9	1.9	2.4	1.9	86.0	15.8	100.0	99.7	100.0	4.9	3.3	100.0	99.8
$n = 500$													
-0.9	8.1	6.4	6.1	6.0	7.4	4.7	5.7	16.0	6.1	5.5	6.4	4.0	18.4
0.5	2.5	5.2	5.1	4.8	5.1	5.3	78.4	23.7	62.3	4.5	4.2	71.7	27.1
0.9	3.3	1.3	1.4	1.4	83.2	14.5	100.0	99.9	100.0	5.4	4.5	100.0	100.0

Table 4.5: Empirical size at $\alpha = 0.05$. Rejection percentages from the TARMA$(1,1)$ model of Eq.4.19.

$n = 100$	asymptotic									bootstrap			
t ; θ	sLM	$\bar{M}^{\text{GLS}}$	M^{GLS}	MP_T	ADF	ADF$^{\text{GLS}}$	KSe	BBC	EGc	sLM.bI	sLM.bT	KSe.b	BBC.b
0.0;-0.9	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.1	7.3	4.9	16.0
0.5;-0.9	9.5	8.7	9.1	9.1	6.0	9.9	2.2	7.1	1.8	9.4	11.7	4.0	22.6
1.0;-0.9	16.0	9.6	10.4	10.3	7.1	12.1	2.8	9.0	1.5	17.3	21.7	3.6	26.6
1.5;-0.9	28.1	10.7	12.1	11.6	8.3	15.0	5.8	12.7	2.1	29.6	34.8	7.5	33.8
0.0;0.5	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.2	4.9	58.6	16.4
0.5;0.5	7.5	11.2	11.8	11.3	6.9	11.5	18.8	8.8	14.3	7.7	6.8	76.5	22.7
1.0;0.5	12.4	15.7	17.2	15.9	9.1	16.7	29.6	14.1	24.4	10.6	10.1	84.2	32.6
1.5;0.5	20.4	18.1	21.0	18.7	12.3	19.8	42.3	21.0	36.0	20.6	21.5	88.6	42.9
0.0;0.9	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.7	8.5	99.8	95.5
0.5;0.9	7.4	8.1	9.0	8.0	5.8	8.1	2.8	10.0	8.6	5.4	11.9	99.9	99.5
1.0;0.9	12.7	9.9	12.7	9.8	7.0	9.6	3.4	13.6	12.0	11.8	20.4	100.0	100.0
1.5;0.9	22.4	12.2	15.2	12.2	8.0	12.1	4.6	15.7	13.9	23.5	29.9	100.0	100.0

$n = 300$	asymptotic									bootstrap			
t ; θ	sLM	$\bar{M}^{\text{GLS}}$	M^{GLS}	MP_T	ADF	ADF$^{\text{GLS}}$	KSe	BBC	EGc	sLM.bI	sLM.bT	KSe.b	BBC.b
0.0;-0.9	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.3	7.9	3.8	20.6
0.5;-0.9	25.7	17.6	17.8	18.2	10.2	19.4	5.1	16.5	1.6	24.1	28.1	4.6	38.5
1.0;-0.9	52.5	26.7	26.9	27.7	15.8	30.4	17.4	31.9	3.8	54.7	59.3	19.7	62.1
1.5;-0.9	77.1	33.5	34.0	35.1	22.1	38.2	36.9	50.1	8.2	76.2	79.9	37.4	76.6
0.0;0.5	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	4.9	4.7	67.7	23.8
0.5;0.5	20.6	25.0	25.1	24.5	12.9	25.6	42.9	18.8	35.3	21.2	21.2	92.5	53.1
1.0;0.5	50.1	39.5	40.1	39.2	26.2	40.9	70.9	45.1	65.8	48.2	48.0	96.9	80.6
1.5;0.5	76.9	49.8	51.9	49.9	43.1	53.1	88.9	72.5	88.0	77.0	77.7	99.2	94.9
0.0;0.9	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	4.9	3.3	100.0	99.8
0.5;0.9	24.8	16.2	19.9	16.0	14.9	18.2	6.4	34.6	29.6	14.3	15.8	100.0	100.0
1.0;0.9	62.8	22.9	34.5	22.5	32.8	26.5	12.4	63.6	52.4	36.1	43.0	100.0	100.0
1.5;0.9	86.3	25.3	44.9	25.0	47.8	29.5	23.5	77.2	65.8	61.7	72.8	100.0	100.0

Table 4.6: Size corrected power of the asymptotic tests and empirical power of the bootstrap tests. Simulation from the TARMA(1,1) model of Eq.4.19.

	PT	DE	FR	BE	AT	GB	NL	IT
sLM	0.167	0.002	0.126	0.900	0.329	0.318	0.900	0.874
sLM.bI	0.384	0.009	0.292	0.833	0.417	0.259	0.802	0.836
sLM.bT	0.379	0.013	0.283	0.837	0.440	0.243	0.817	0.880
$\bar{M}^{\text{GLS}}$	.	.	.	.	.	.	✓	.
M^{GLS}	.	.	.	.	.	.	✓	.
MP_T	.	.	.	.	.	.	✓	.
ADF	.	.	.	.	.	.	.	.
ADF$^{\text{GLS}}$	.	.	.	.	.	.	.	.
KSe	.	.	.	.	.	.	.	.
KSe.b	.	.	.	.	.	.	.	.
BBC	.	✓	.	.	.	.	.	✓
BBC.b	.	✓	.	.	.	.	.	✓
EGc	.	.	.	.	.	.	.	.

Table 4.7: Results of the application of the set of unit root tests on the 8 monthly series of real exchange rates. The first three rows report the p-value for the supLM tests whereas the remaining rows show the checkmark ✓if the test results significant at 5%.

also noted in Pelagatti and Colombo [2015] where the authors show that real exchange rates based on the consumer price index do not preserve the possible stationarity properties of the ratios. A second factor is represented by the incorrect linear specification for the price dynamics. Indeed, the presence of trading costs implies that the mechanisms governing price adjustments are non-linear and threshold autoregressive models provide a solution to the problem by allowing a "band of inaction" random walk regime where arbitrage does not occur, and other regimes where the mean reversion takes place so that the model is globally stationary [see Bec et al., 2004, and references therein for further discussion]. For a review on how TAR models are used to analyse the exchange rates dynamics see also Hansen [2011]. Among other approaches, Bec et al. [2008b], Gourieroux and Robert [2006] introduce switching models to incorporate the possibility that the threshold that defines the regimes where arbitrage takes place is a random variable.

We consider the monthly log real exchange rates for the following countries: Portugal (PT), Germany (DE), France (FR), Belgium (BE), Austria (AT), Great Britain (GB), Netherland (NL), Italy (IT). The series range from 1973:09 to 1998:12 ($n = 304$) and are produced by the Bank of International Settlements (BIS) by taking the geometric weighted average of a basket of bilateral exchange rates (27 economies), adjusted with the corresponding relative consumer prices. Such weights are constructed from manufacturing trade flows as to encompass both third-market competition and direct bilateral trade through a double-weighting scheme. See Klau and Fung [2006] and `https://www.bis.org/` for more details on the construction of the indexes.

Table 4.7 reports the results of the application of the battery of unit root

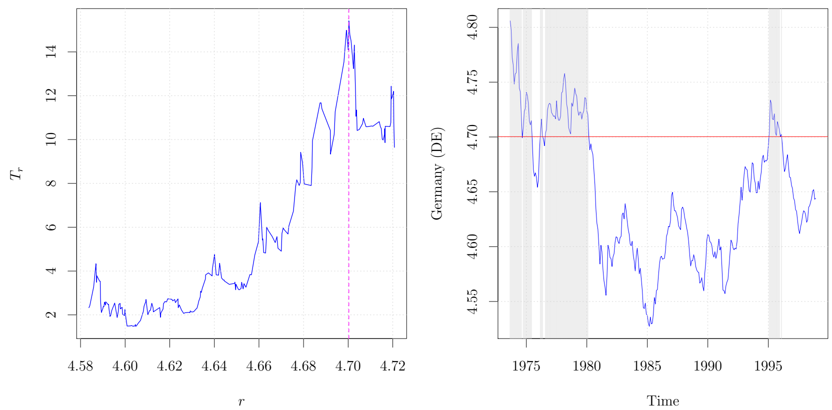

Figure 4.1: Left: values of the LM statistic T_r over the threshold grid. The value that maximizes T_r, $\hat{r} = 4.700$, is taken as the estimated threshold and is indicated with a dashed line. Right: time series of real monthly exchange rates for Germany. The estimated threshold is indicated with a horizontal red line. The gray shaded area indicates the periods associated to the upper regime.

tests described in the previous section on the 8 monthly series of real exchange rates. The first three rows show the p-values from our supLM tests. To enhance readability, the remaining rows show a checkmark ✓ if the corresponding test rejects the null hypothesis at level 5%. Based upon our tests, we can reject the null hypothesis with some confidence for Germany (DE). Interestingly, with the exception of the BBC tests, all the other tests fail to reject and the finding is somehow consistent with that of Bec et al. [2004] where the authors rejected for the pairwise real exchange rates of Germany versus France, Italy, Belgium, Netherland and Portugal. Indeed, the BBC tests reject also for Italy but our tests do not and this might be due to the oversize of the BBC tests. Moreover, as shown in Section 4.4.1, the M tests may have very little power against some TARMA alternatives and this explains their failure to reject the null hypothesis. This result raises the question whether a TARMA model is plausible for the series for Germany. Hence, we fit the following TARMA(1,1) model

$$X_t = \begin{cases} \phi_{1,0} + \phi_{1,1} X_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}, & \text{if } X_{t-1} > r \\ \phi_{2,0} + \phi_{2,1} X_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}, & \text{if } X_{t-1} \leq r \end{cases} \tag{4.20}$$

In Figure 4.1(left) we plot the values of the LM statistic T_r computed over a threshold grid that ranges from the 15th to the 85th percentiles of the data. The estimated threshold $\hat{r} = 4.700$, that maximizes T_r, is also the value that minimizes the AIC criterion over the same grid. In the right panel of the figure we plot the time series of the monthly real exchange rates for Germany (DE) where we have indicated the selected threshold with a red line. The gray shaded area indicates the months associated to the upper regime. The parameter estimates are presented in Table 4.8 and point to a lower regime with a

	θ	$\phi_{1,0}$	$\phi_{1,1}$	$\phi_{2,0}$	$\phi_{2,1}$
estimate	0.305	-1.250	0.735	-0.149	0.968
s.e.	(0.055)	(0.281)	(0.059)	(0.088)	(0.019)

Table 4.8: Parameter estimates from the TARMA(1,1) fit of Eq. 4.20 on the monthly real exchange rates for Germany (DE) with $\hat{r} = 4.700$.

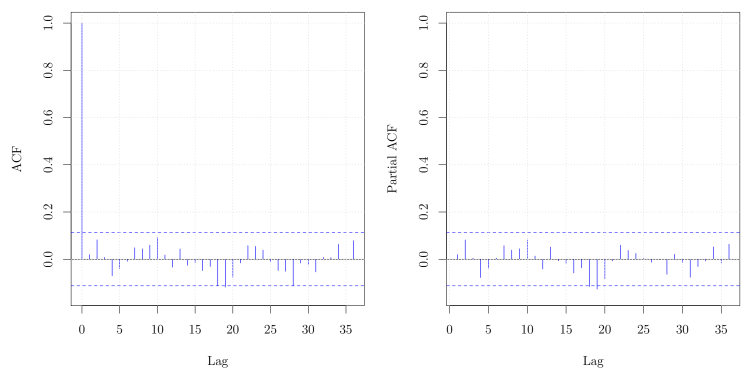

Figure 4.2: Global and partial correlograms of the residuals from the TARMA fit for the time series of monthly real exchange rates for Germany.

possible unit root and an upper regime where the slope is strictly smaller than 1. This is consistent with the idea of a non-linear adjustment mechanism that activates when the rate crosses the threshold. Figure 4.1(right) shows that the intervention regime is visited mostly before 1980 and after 1995 and this is in general agreement with the results of Bec et al. [2004] and Bec et al. [2008b] obtained on the real exchange rate series of French Franc against Deutsche Mark. The MA parameter θ greatly enhances the fitting ability of the model while retaining parsimony. This is witnessed by the diagnostics computed on the residuals. Figure 4.2 shows the global and partial sample autocorrelations up to 36 months, computed on the residuals of the TARMA model. Clearly, there is no sign of residual correlation. In order to rule out the possibility of non-linear serial dependence in the residuals we have computed the test based upon the entropy measure S_ρ described in Giannerini et al. [2015]. The results are shown in Figure 4.3(left) where the rejection bands correspond to the null hypothesis of serial independence at level 95% (green dashed line) and 99% (blue dashed line) up to 24 months. The results confirm that the residuals do not present any kind of dependence. We complete the diagnostic analysis by looking at the quantile-quantile plot of the residuals and by computing the Shapiro-Wilk's normality test. This is shown in Figure 4.3(right). The results do not point to important deviations from normality and confirm the goodness of the proposed fit.

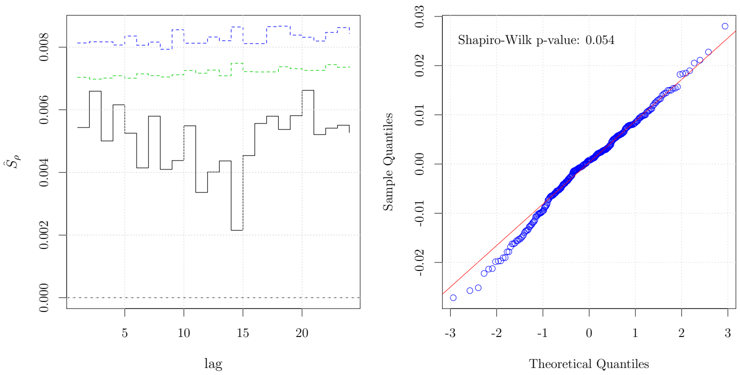

Figure 4.3: Entropy measure S_ρ computed on the residuals from the TARMA fit for the time series of monthly real exchange rates for Germany. The confidence bands at 95% (green) and 99%(blue) correspond to the null hypothesis of serial independence (left) and qqplot of the residuals with the p-value of the Shapiro-Wilk's normality test (right).

4.6 Proofs

Proof of Proposition 56

We give the proof of (4.15), via the following uniform approximation argument [Pollard, 2012, Example 11, p.70]. Let $G, G_1, G_2, \ldots$ be a sequence of random elements in a metric space $(\mathcal{X}, d)$, with the support of G being a separable set of completely regular elements. Suppose for each $\epsilon > 0, \delta > 0$, there exist approximating random elements $AG, AG_1, AG_2, \ldots$ such that

(*i*) $P^*\{d(G, AG) > \epsilon\} < \delta$;

(*ii*) $\limsup P^*\{d(G_n, AG_n\} > \epsilon\} < \delta$;

(*iii*) $AG_n \rightsquigarrow AG$,

where $P^*(\cdot)$ denotes the outer probability measure of the enclosed expression and $\rightsquigarrow$ the weak convergence. Then $G_n \rightsquigarrow G$, as $n \to \infty$. The complete regularity condition holds in our current setting, first for the case of the sample space being some Euclidean space and also for the case of some product space of $\mathcal{D}[0,1]$ with the limiting distribution concentrated on the corresponding $\mathcal{C}[0,1]$ product sub-space; in both cases, the sample space is equipped with the metric induced by the supremum norm. Hence, we focus on verifying conditions (i)–(iii) below. Recall that

$$\frac{1}{n}\frac{\partial \ell}{\partial \phi_{1,1}} = \sum_{t=1}^{n} \frac{1}{\sqrt{n}} \frac{\varepsilon_t}{\sigma} \frac{1}{1-\theta B} \left\{ \frac{X_{t-1}}{\sqrt{n}\sigma} I\left(\frac{X_{t-1}}{\sqrt{n}(1-\theta)\sigma} \leq \tau \right) \right\}.$$

Let $A_{n,k} = \sum_{t=k+1}^{n} \frac{1}{\sqrt{n}} \frac{\varepsilon_t}{\sigma} \sum_{j=0}^{k} \theta^j \frac{X_{t-1-j}}{\sqrt{n}\sigma} I\left(\frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma} \le \tau\right)$. We claim that for any fixed positive integer k and as $n \to \infty$,

$$A_{n,k} \rightsquigarrow (1-\theta^{k+1}) \times \int_0^1 W_s I(W_s \le \tau) dW_s, \tag{4.21}$$

which will be verified later. We shall check conditions (i)–(iii) with $G_n = \frac{1}{n}\frac{\partial \ell}{\partial \phi_{1,1}}$, $AG_n = A_{n,k}$, $G = \int_0^1 W_s I(W_s \le \tau) dW_s$ and $AG = (1-\theta^{k+1})G$; indeed condition (iii) obtains due to (4.21). Clearly, (i) holds by Slutsky's theorem. It remains to show (ii), which can be done by first bounding the difference

$$\begin{aligned} D_{n,k} &= \frac{1}{n}\frac{\partial \ell}{\partial \phi_{1,1}} - A_{n,k} \\ &= \sum_{t=1}^{k} \frac{1}{\sqrt{n}} \frac{\varepsilon_t}{\sigma} \sum_{j=0}^{t-1} \theta^j \left\{ \frac{X_{t-1-j}}{\sqrt{n}\sigma} I\left(\frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma} \le \tau\right)\right\} + \\ &\quad \sum_{t=k+1}^{n} \frac{1}{\sqrt{n}} \frac{\varepsilon_t}{\sigma} \sum_{j=k+1}^{t-1} \theta^j \left\{ \frac{X_{t-1-j}}{\sqrt{n}\sigma} I\left(\frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma} \le \tau\right)\right\}. \end{aligned}$$

The summands of $D_{n,k}$ form a martingale difference sequence with respect to the σ-algebra $\mathcal{F}_t$ generated by the innovations $\varepsilon_{t-j}, j \ge 0$, hence $D_{n,k}$ is of zero mean and Jensen's inequality implies that its variance is bounded by

$$\begin{aligned} \mathrm{var}(D_{n,k}) &= \sum_{t=1}^{k} \frac{1}{n} E\left| \sum_{j=0}^{t-1} \theta^j \left\{ \frac{X_{t-1-j}}{\sqrt{n}\sigma} I\left(\frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma} \le \tau\right)\right\}\right|^2 + \\ &\quad \sum_{k+1}^{n} \frac{1}{n} E\left| \sum_{j=k+1}^{t-1} \theta^j \left\{ \frac{X_{t-1-j}}{\sqrt{n}\sigma} I\left(\frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma} \le \tau\right)\right\}\right|^2 \\ &\le \sum_{t=1}^{k} \frac{1}{n} E\left(\sum_{j=0}^{t-1} |\theta|^j \frac{|X_{t-1-j}|}{\sqrt{n}\sigma}\right)^2 + \sum_{k+1}^{n} \frac{1}{n} E\left(\sum_{j=k+1}^{t-1} |\theta|^j \frac{|X_{t-1-j}|}{\sqrt{n}\sigma}\right)^2 \\ &\le \sum_{t=1}^{k} \frac{1}{n} E\{\sum_{j=0}^{t-1} |\theta|^j X_{t-1-j}^2/(n\sigma^2)\}/(1-|\theta|) + \\ &\quad \sum_{k+1}^{n} \frac{1}{n} E\{ \sum_{j=k+1}^{t-1} |\theta|^j X_{t-1-j}^2)/(n\sigma^2)\}/(1-|\theta|) \\ &\le \frac{K}{(1-|\theta|)^2} \left\{ \sum_{t=1}^{k} \frac{t-1}{n^2} + |\theta|^{k+1} \sum_{j=k+1}^{n} \frac{t-1}{n^2} \right\} \\ &\le \frac{K}{(1-|\theta|)^2} \left\{ \frac{k(k-1)}{2n^2} + |\theta|^{k+1}\frac{1}{2} \right\} \end{aligned} \tag{4.22}$$

where $K > 0$ is a constant such that $E(X_t^2) = \sigma^2\{1+\theta^2+(t-1)(1-\theta)^2\} \le tK\sigma^2$. Since the true θ is less than 1 in magnitude, (4.22) indicates that for any positive ϵ, by choosing k sufficiently large and then letting $n \to \infty$, $P(|D_{n,k}| \le \epsilon) \to 1$. Thus, (ii) holds by Markov's inequality.

It remains to verify (4.21). The claim (4.21) would follow readily from Theorem 7.3 of Kurtz and Protter [1996] were the step function $I(x \le \tau)$ a continuous function. Unfortunately, this is not the case but it is discontinuous only at τ. The idea of proof is to approximate the step function by a net of smooth functions, say $G_\delta(x)$, such that $|G_\delta(x) - I(x \le \tau)| \le H_\delta(x)$ with the bound $H_\delta(x)$ being uniformly bounded, continuous functions and with support inside $[\tau-\delta, \tau+\delta]$. Define

$$A_{n,k,\delta} = \sum_{t=k+1}^{n} \frac{1}{\sqrt{n}} \frac{\varepsilon_t}{\sigma} \sum_{j=0}^{k} \theta^j \frac{X_{t-1-j}}{\sqrt{n}\sigma} G_\delta\left(\frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma}\right).$$

Then, for fixed k and δ, $A_{n,k,\delta} \rightsquigarrow (1-\theta^{k+1}) \int_0^1 W_s G_\delta(W_s) dW_s$, as $n \to \infty$. Consider

$$A_{n,k} - A_{n,k,\delta} = \sum_{t=k+1}^{n} \frac{1}{\sqrt{n}} \frac{\varepsilon_t}{\sigma} \sum_{j=0}^{k} \theta^j \frac{X_{t-1-j}}{\sqrt{n}\sigma} \left\{ I\left(\frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma} \le r\right) - G_\delta\left(\frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma}\right)\right\},$$

whose summands form a martingale difference sequence, so it is of zero mean and its variance can be bounded as follows:

$$E(A_{n,k} - A_{n,k,\delta})^2 \le \frac{1}{1-|\theta|} \sum_{t=k+1}^{n} \frac{1}{n} E\left\{\sum_{j=0}^{k} |\theta|^j \left\{\frac{X_{t-1-j}}{\sqrt{n}\sigma}\right\}^2 H_\delta^2\left(\frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma}\right)\right\}$$

$$\le \frac{(1-\theta)^2 \max(|r-\delta|^2, |r+\delta|^2)}{1-|\theta|} E\left\{\sum_{t=k+1}^{n} \frac{1}{n} \sum_{j=0}^{k} |\theta|^j H_\delta^2\left(\frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma}\right)\right\}.$$

On the other hand, it follows from Theorem 7.3 of Kurtz and Protter [1996] that

$$\sum_{t=k+1}^{n} \frac{1}{n} \sum_{j=0}^{k} |\theta|^j H_\delta^2\left(\frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma}\right) \rightsquigarrow \frac{1-|\theta|^{k+1}}{1-|\theta|} \int_0^1 H_\delta^2(W_s) ds. \tag{4.23}$$

Since $H_\delta^2(\cdot)$ is continuous, uniformly bounded, say, by $K > 0$, and its support lies inside $[r-\delta, r+\delta]$, the expectation of the LHS of (4.23) converges to

$$\begin{aligned} E\int_0^1 H_\delta^2(W_s) ds &\le K \int_0^1 P(W_s \in [\tau-\delta, \tau+\delta]) ds \\ &\le K\left\{\nu + \int_\nu^1 P(W_s \in [\tau-\delta, \tau+\delta]) ds\right\} \\ &\le K\left\{\nu + \int_\nu^1 \int_{\tau-\delta}^{\tau+\delta} \frac{1}{\sqrt{2\pi s}} \exp(-y^2/(2s)) dy ds\right\}, \end{aligned}$$

where $0 < \tau < 1$ can be chosen to be an arbitrary small, fixed number, and then the double integral having a bounded integrand can be made arbitrarily small by rendering $\delta > 0$ small. Hence, for any fixed $\epsilon > 0$ and $\gamma > 0$ it holds that for all sufficiently small δ, $P(|A_{n,k} - A_{n,k,\delta}| > \epsilon) < \gamma$ for all sufficiently large n. Also, the difference $\int_0^1 W_s G_\delta(W_s) dW_s - \int_0^1 W_s I(W_s \le r) dW_s = \int_0^1 W_s \{G_\delta(W_s) - I(W_s \le r)\} dW_s$ is of zero mean and variance equal to

$$\int_0^1 E[W_s^2 \{G_\delta(W_s) - I(W_s \le \tau)\}^2] ds \le \int_0^1 E[W_s^2 H_\delta^2(W_s)] ds$$
$$\le K \max(|\tau - \delta|^2, |\tau + \delta|^2) \int_0^1 P(W_s \in [\tau - \delta, \tau + \delta]) ds$$

which can be similarly shown to be made arbitrarily small for all sufficiently small δ. Thus, the claim (4.21) can be verified, by routine arguments [see e.g. Pollard, 2012, Example 11, p.70].

Proof of Proposition 57

Since, $\tau_1 \le \frac{X_{t-1-j}}{\sigma\sqrt{n}(1-\theta)} \le \tau_2$, then the process defined by (4.12) differs from that defined in (4.11) just for a multiplicative constant. Therefore we just need to prove the tightness of (4.11). First, we prove that the tightness of (4.12) is implied by the tightness of $\{\mathcal{T}(\tau), a \le \tau \le b\}$, where

$$\mathcal{T}(\tau) = \sum_{t=2}^{n} \frac{1}{\sqrt{n}} \frac{\varepsilon_t}{\sigma} \sum_{j=0}^{t-2} \theta^j I\left(\frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma} \le \tau\right). \tag{4.24}$$

In fact, noting that $I\left(\frac{X_0}{\sqrt{n}(1-\theta)\sigma} \le \tau\right)$ is no longer a random variable, such implication immediately follows since the variable

$$\nabla_2(\tau) - \mathcal{T}(\tau) = \frac{1}{\sqrt{n}} \frac{\varepsilon_1}{\sigma} I\left(\frac{X_0}{\sqrt{n}(1-\theta)\sigma} \le \tau\right)$$
$$+ \sum_{t=2}^{n} \frac{1}{\sqrt{n}} \frac{\varepsilon_t}{\sigma} \theta^{t-1} I\left(\frac{X_0}{\sqrt{n}(1-\theta)\sigma} \le \tau\right)$$

is of zero mean and its variance tends to zero as n increases. Hence, we focus on verify the tightness of $\{\mathcal{T}(\tau), a \le \tau \le b\}$ below. Following the argument of Theorem 2.2 in Chan [1990] and theorem 22.1 in Billingsley [1968] (see also pp. 94 of Billingsley [1968]), we shall show that there exists a constant $C > 0$ such that, for any fixed $a \le \tau_1 < \tau_2 \le b$,

$$E\left[\left|\sum_{t=2}^{n} \frac{1}{\sqrt{n}} \frac{\varepsilon_t}{\sigma} \sum_{j=0}^{t-2} \theta^j I\left(\tau_1 \le \frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma} \le \tau_2\right)\right|^4\right] \le C(\tau_2 - \tau_1)^2$$

Essentially, we need to compute the following expectation:

$$\frac{1}{\sigma^4 n^2}\sum_{t_1,t_2,t_3,t_4} E\left[\varepsilon_{t_1}\varepsilon_{t_2}\varepsilon_{t_3}\varepsilon_{t_4}\times I(t_1)I(t_2)I(t_3)I(t_4)\right]$$

where

$$I(s)=\sum_{j=0}^{s-2}\theta^j I\left(\tau_1\le \frac{X_{s-1-j}}{\sqrt{n}(1-\theta)\sigma}\le\tau_2\right).$$

Below, K denote a general constant that may depend on θ and vary from occurrence to occurrence. Assume that $t=\max_{i\in\{1,2,3,4\}} t_i$. Since $E[\varepsilon_t]=E[\varepsilon_t^3]=0$, the law of iterated expectations implies that

$$\begin{aligned}&\frac{1}{\sigma^4 n^2}\sum_{t_1,t_2,t_3,t_4} E\left[\varepsilon_{t_1}\varepsilon_{t_2}\varepsilon_{t_3}\varepsilon_{t_4}\times I(t_1)I(t_2)I(t_3)I(t_4)\right]\\ =&\frac{1}{\sigma^4 n^2}\sum_{t=2}^{n-1}E\left[\varepsilon_t^4 I^4(t)\right]+\frac{K}{\sigma^4 n^2}\sum_{t=2}^{n-1}\sum_{u,v<t}E\left[\varepsilon_t^2\varepsilon_u\varepsilon_v I^2(t)I(u)I(v)\right]\end{aligned}$$

We verify separately that there exists two constants $C_1, C_2>0$ such that

$$\frac{1}{\sigma^4 n^2}\sum_{t=2}^{n-1}E\left[\varepsilon_t^4 I^4(t)\right]\le\frac{C_1}{n}(\tau_2-\tau_1) \tag{4.25}$$

$$\frac{K}{\sigma^4 n^2}\sum_{t=3}^{n-1}\sum_{u,v<t}E\left[\varepsilon_t^2\varepsilon_u\varepsilon_v I^2(t)I(u)I(v)\right]\le C_2(\tau_2-\tau_1)^2 \tag{4.26}$$

By using the low of iterated expectations, the Jensen's inequality and the fact the $E\left[\varepsilon_t^4\right]$ and $\sum_{j=0}^{\infty}|\theta|^j$ can be absorbed into K, we have:

$$\begin{aligned}&\frac{1}{\sigma^4 n^2}\sum_{t=2}^{n-1}E\left[\varepsilon_t^4 I^4(t)\right]\\ =&\frac{K}{n^2}\sum_{t=2}^{n-1}E\left[\varepsilon_t^4\left\{\sum_{j=0}^{t-2}\theta^j I\left(\tau_1\le\frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma}\le\tau_2\right)\right\}^4\right]\\ \le&\frac{K}{n^2}\sum_{t=2}^{n-1}E\left[\sum_{j=0}^{t-2}|\theta|^j I\left(\tau_1\le\frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma}\le\tau_2\right)\right]\\ =&\frac{K}{n^2}\sum_{t=2}^{n-1}\sum_{j=0}^{t-2}|\theta|^j P\left(\tau_1\le\frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma}\le\tau_2\right)\\ \le&\frac{K}{n}\sum_{j=0}^{n-3}|\theta|^j\frac{1}{n}\sum_{s=0}^{n-1}P\left(\tau_1\le\frac{X_{s+1}}{\sqrt{n}(1-\theta)\sigma}\le\tau_2\right).\end{aligned}$$

The last inequality holds since we have fixed j and done the change of variables $s+1 = t-1-j \to s = t-2-j$. Since $\frac{X_s}{\sqrt{n}\sigma(1-\theta)} \sim N\left(0, \frac{(1-\theta)^2(s-1)+1+\theta^2}{n(1-\theta)^2}\right)$ and $e^{-\frac{y^2 n(1-\theta)^2}{2\{(1-\theta)^2 s+1+\theta^2\}}} < 1$, the previous expression is bounded by

$$\frac{K}{n}\sum_{j=0}^{n-3}|\theta|^j \frac{1}{n}\sum_{s=0}^{n-1}\left[\int_{\tau_1}^{\tau_2}\left\{\frac{\sqrt{n(1-\theta)^2}}{\sqrt{2\pi\left\{(1-\theta)^2 s+1+\theta^2\right\}}}\right\}dy\right]$$
$$=\frac{K(\tau_2-\tau_1)}{n}\sum_{j=0}^{n-3}|\theta|^j \frac{1}{n}\sum_{s=0}^{n-1}\sqrt{\frac{1}{\frac{s}{n}+\frac{1+\theta^2}{n(1-\theta)^2}}} \leq \frac{K(\tau_2-\tau_1)}{n}\sum_{j=0}^{n-3}|\theta|^j \frac{1}{n}\sum_{s=0}^{n-1}\sqrt{\frac{1}{\frac{s+1/2}{n}}}.$$

Let $\zeta(t) = t^{-1/2}$, with $t \in [0,1]$. Since $\frac{\partial^2\zeta}{\partial t^2} > 0$, for each $t \in [0,1]$, the middle point rule implies that

$$\frac{K(\tau_2-\tau_1)}{n}\sum_{j=0}^{n-3}|\theta|^j \frac{1}{n}\sum_{s=0}^{n-1}\sqrt{\frac{1}{\frac{s+1/2}{n}}}$$
$$\leq \frac{K(\tau_2-\tau_1)}{n}\sum_{j=0}^{n-3}|\theta|^j \int_0^1 t^{-1/2}dt \leq \frac{K(\tau_2-\tau_1)}{n}.$$

Hence, Condition 4.25 holds. Below, we verify Condition 4.26.

$$\frac{K}{\sigma^4 n^2}\sum_{t=3}^{n-1}\sum_{u,v<t} E\left[\varepsilon_t^2\varepsilon_u\varepsilon_v I^2(t)I(u)I(v)\right] = \frac{K}{n^2}\sum_{t=3}^{n-1} E\left[I^2(t)\left\{\sum_{u<t}\varepsilon_u I(u)\right\}^2\right]$$
$$=\frac{K}{n^2}\sum_{t=3}^{n-1} E\left[\left\{\sum_{j=0}^{t-2}\theta^j I\left(\tau_1 \leq \frac{X_{t-1-j}}{\sqrt{n}(1-\theta)\sigma} \leq \tau_2\right)\right\}^2\left\{\sum_{u=2}^{t-1}\varepsilon_u\sum_{j=0}^{u-2}\theta^j I\left(\tau_1 \leq \frac{X_{u-1-j}}{\sqrt{n}(1-\theta)\sigma} \leq \tau_2\right)\right\}^2\right]$$
$$=\frac{K}{n^2}\sum_{t=3}^{n-1} E\left[\left\{\sum_{u=2}^{t-1}\sum_{j_1=0}^{t-1-u}\sum_{j_2=0}^{u-2}\theta^{j_1+j_2} I\left(\tau_1 \leq \frac{X_{t-1-j_1}}{\sqrt{n}(1-\theta)\sigma} \leq \tau_2\right)\varepsilon_u I\left(\tau_1 \leq \frac{X_{u-1-j_2}}{\sqrt{n}(1-\theta)\sigma} \leq \tau_2\right)\right\}^2\right]$$
$$+\frac{K}{n^2}\sum_{t=3}^{n-1} E\left[\left\{\sum_{u=2}^{t-1}\sum_{j_1=t-u}^{t-2}\sum_{j_2=0}^{u-2}\theta^{j_1+j_2} I\left(\tau_1 \leq \frac{X_{t-1-j_1}}{\sqrt{n}(1-\theta)\sigma} \leq \tau_2\right)\varepsilon_u I\left(\tau_1 \leq \frac{X_{u-1-j_2}}{\sqrt{n}(1-\theta)\sigma} \leq \tau_2\right)\right\}^2\right]$$

Now, we consider the two addends separately. Since $j_1 > t-u \to t-1-j_1 < u$, the latter one is a sequence of martingale difference and applying a similar argument of that previously used we obtain:

$$\frac{K}{n^2}\sum_{t=3}^{n-1} E\left[\left\{\sum_{u=2}^{t-1}\sum_{j_1=t-u}^{t-2}\sum_{j_2=0}^{u-2}\theta^{j_1+j_2} I\left(\tau_1 \le \frac{X_{t-1-j_1}}{\sqrt{n}(1-\theta)\sigma} \le \tau_2\right)\varepsilon_u I\left(\tau_1 \le \frac{X_{u-1-j_2}}{\sqrt{n}(1-\theta)\sigma} \le \tau_2\right)\right\}^2\right]$$

$$\le \frac{K}{n^2}\sum_{t=3}^{n-1}\sum_{u=2}^{t-1} E\left[\varepsilon_u^2\left\{\sum_{j_1=t-u}^{t-2}|\theta|^{j_1}\sum_{j_2=0}^{u-2}|\theta|^{j_2} I\left(\tau_1 \le \frac{X_{t-1-j_1}}{\sqrt{n}(1-\theta)\sigma} \le \tau_2\right) I\left(\tau_1 \le \frac{X_{u-1-j_2}}{\sqrt{n}(1-\theta)\sigma} \le \tau_2\right)\right\}^2\right]$$

$$\le \frac{K}{n^2}\sum_{t=3}^{n-1}\sum_{u=2}^{t-1}\sum_{j_1=t-u}^{t-2}|\theta|^{j_1}\sum_{j_2=0}^{u-2}|\theta|^{j_2} P\left(\tau_1 \le \frac{X_{t-1-j_1}}{\sqrt{n}(1-\theta)\sigma} \le \tau_2, \tau_1 \le \frac{X_{u-1-j_2}}{\sqrt{n}(1-\theta)\sigma} \le \tau_2\right)$$

$$= \frac{K}{n^2}\sum_{j_2=0}^{n-1}|\theta|^{j_2}\sum_{j_1=1}^{n-1}|\theta|^{j_1}\sum_{s_2=0}^{n-1}\sum_{s_1=0}^{n-1} P\left(\tau_1 \le \frac{X_{s_1+1}}{\sqrt{n}(1-\theta)\sigma} \le \tau_2, \tau_1 \le \frac{X_{s_2+1}}{\sqrt{n}(1-\theta)\sigma} \le \tau_2\right)$$

$$\le \frac{K}{n^2}\sum_{j_2=0}^{n-1}|\theta|^{j_2}\sum_{j_1=1}^{n-1}|\theta|^{j_1}\sum_{s_2=0}^{n-1}\sum_{s_1=0}^{n-1}$$

$$\int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2}\frac{n(1-\theta)^2}{\sqrt{\{(1-\theta)^2 s_1 + 1 + \theta^2\}\{(1-\theta)^2 s_2 + 1 + \theta^2\} - \{(1-\theta)^2\min(s_1,s_2)+1\}^2}}dy_1 dy_2$$

$$= \frac{K(\tau_2-\tau_1)^2}{n^2}\sum_{j_2=0}^{n-1}|\theta|^{j_2}\sum_{j_1=1}^{n-1}|\theta|^{j_1}\sum_{s_2=0}^{n-1}\sum_{s_1=0}^{n-1}$$

$$\times\frac{n(1-\theta)^2}{\sqrt{\{(1-\theta)^2 s_1 + 1 + \theta^2\}\{(1-\theta)^2 s_2 + 1 + \theta^2\} - \{(1-\theta)^2\min(s_1,s_2)+1\}^2}}$$

$$\le K(\tau_2-\tau_1)^2\sum_{j_2=0}^{n-1}|\theta|^{j_2}\sum_{j_1=1}^{n-1}|\theta|^{j_1}\frac{1}{n^2}\sum_{s_2=0}^{n-1}\sum_{s_1=0}^{n-1}\frac{1}{\sqrt{\frac{s_1+1/2}{n}\frac{s_2+1/2}{n} - \left(\frac{\min(s_1,s_2)+1/2}{n}\right)^2}}.$$

Since $\int_0^1\int_0^1 \frac{1}{\sqrt{xy-\min^2(x,y)}}dxdy = 2\pi$, the summation

$$\frac{1}{n^2}\sum_{s_2=0}^{n-1}\sum_{s_1=0}^{n-1}\frac{1}{\sqrt{\frac{s_1+1/2}{n}\frac{s_2+1/2}{n} - \left(\frac{\min(s_1,s_2)+1/2}{n}\right)^2}}$$

is bounded and this implies:

$$K(\tau_2-\tau_1)^2\sum_{j_2=0}^{n-1}|\theta|^{2j_2}\sum_{j_1=1}^{n-1}|\theta|^{2j_1}\frac{1}{n^2}\sum_{s_2=0}^{n-1}\sum_{s_1=0}^{n-1}\frac{1}{\sqrt{\frac{s_1+1/2}{n}\frac{s_2+1/2}{n} - \left(\frac{s_1+1/2}{n}\right)^2}}$$

$$\le K(\tau_2-\tau_1)^2.$$

Therefore, it remains to verify that:

$$\frac{K}{n^2}\sum_{t=3}^{n-1}E\left[\left\{\sum_{u=2}^{t-1}\sum_{j_1=0}^{t-1-u}\sum_{j_2=0}^{u-2}\theta^{j_1+j_2}I\left(\tau_1\leq\frac{X_{t-1-j_1}}{\sqrt{n}(1-\theta)\sigma}\leq\tau_2\right)\varepsilon_u I\left(\tau_1\leq\frac{X_{u-1-j_2}}{\sqrt{n}(1-\theta)\sigma}\leq\tau_2\right)\right\}^2\right]$$
$$\leq K(\tau_2-\tau_1)^2.$$

The entity is composed by the square terms and the double products. Firstly, we consider the sum of square terms and apply the Jensen's inequality:

$$\frac{K}{n^2}\sum_{t=3}^{n-1}E\left[\sum_{u=2}^{t-1}\varepsilon_u^2\left\{\sum_{j_1=0}^{t-1-u}\theta^{j_1}I\left(\tau_1\leq\frac{X_{t-1-j_1}}{\sqrt{n}(1-\theta)\sigma}\leq\tau_2\right)\right\}^2\right.$$
$$\left.\times\left\{\sum_{j_2=0}^{u-2}\theta^{j_2}I\left(\tau_1\leq\frac{X_{u-1-j_2}}{\sqrt{n}(1-\theta)\sigma}\leq\tau_2\right)\right\}^2\right]$$
$$=\frac{K}{n^2}\sum_{j_1=0}^{\cdot}|\theta|^{j_1}\sum_{j_2=0}^{\cdot}|\theta|^{j_2}\sum_{t=j_1+3}^{n-1}\sum_{u=j_2+2}^{n-2}E\left[\varepsilon_u^2 I\left(\tau_1\leq\frac{X_{t-1-j_1}}{\sqrt{n}(1-\theta)\sigma}\leq\tau_2\right)I\left(\tau_1\leq\frac{X_{u-1-j_2}}{\sqrt{n}(1-\theta)\sigma}\leq\tau_2\right)\right]$$
$$=\frac{K}{n^2}\sum_{j_1=0}^{\cdot}|\theta|^{j_1}\sum_{j_2=0}^{\cdot}|\theta|^{j_2}\sum_{t=j_1+3}^{n-1}\sum_{u=j_2+2}^{n-2}\int_{-\infty}^{+\infty}\int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2}\varepsilon_u^2 f(y_1,y_2,y_3)dy_1dy_2dy_3,$$

where $f(y_1,y_2,y_3)$ is the probability density function of the random vector

$$\left(\frac{X_{t-1-j_1}}{\sigma\sqrt{n}(1-\theta)},\varepsilon_u,\frac{X_{u-1-j_2}}{\sigma\sqrt{n}(1-\theta)}\right)^{\intercal}.$$

It holds that

$$E\left[\varepsilon_u^2\middle|\frac{X_{s_1}}{(1-\theta)^2\sigma\sqrt{n}},\frac{X_{s_3}}{(1-\theta)^2\sigma\sqrt{n}}\right]$$
$$=\sigma^2-\sigma^2\frac{(1-\theta)^4 s_1+2(1-\theta)^2\theta}{s_1(s_3-s_1)(1-\theta)^4+2\theta s_3(1-\theta)^2+3\theta^2}$$
$$+\left(\frac{\sqrt{n}\sigma(1-\theta)^2}{s_1(s_3-s_1)(1-\theta)^4+2\theta s_3(1-\theta)^2+3\theta^2}\right.$$
$$\left.\times\left[\frac{X_{s_3}}{\sigma\sqrt{n}(1-\theta)}\left\{s_1(1-\theta)^2+2\theta\right\}-\frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)}\left\{s_1(1-\theta)^2+\theta\right\}\right]\right)^2.$$

Hence,

$$
\begin{aligned}
&\int_{-\infty}^{+\infty}\int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2}\varepsilon_u^2 f(y_1,y_2,y_3)dy_1dy_2dy_3\\
=&\int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2}\sigma^2 f(y_1,y_3)dy_1dy_3\\
-&\sigma^2\frac{(1-\theta)^4s_1+2(1-\theta)^2\theta}{s_1(s_3-s_1)(1-\theta)^4+2\theta s_3(1-\theta)^2+3\theta^2}\int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2} f(y_1,y_3)dy_1dy_3\\
+&\left(\frac{\sqrt{n}\sigma(1-\theta)^2}{s_1(s_3-s_1)(1-\theta)^4+2\theta s_3(1-\theta)^2+3\theta^2}\right)^2\\
\times&\int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2}\left(\left[y_3\left\{s_1(1-\theta)^2+2\theta\right\}-y_1\left\{s_1(1-\theta)^2+\theta\right\}\right]\right)^2 f(y_1,y_3)dy_1dy_3,\\
\leq&\int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2}\sigma^2 f(y_1,y_3)dy_1dy_3\\
+&\sigma^2\frac{|(1-\theta)^4s_1+2(1-\theta)^2\theta|}{|s_1(s_3-s_1)(1-\theta)^4+2\theta s_3(1-\theta)^2+3\theta^2|}\int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2} f(y_1,y_3)dy_1dy_3\\
+&\left(\frac{\sqrt{n}\sigma(1-\theta)^2}{s_1(s_3-s_1)(1-\theta)^4+2\theta s_3(1-\theta)^2+3\theta^2}\right)^2\\
\times&\int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2}\left(\left[y_3\left\{s_1(1-\theta)^2+2\theta\right\}-y_1\left\{s_1(1-\theta)^2+\theta\right\}\right]\right)^2 f(y_1,y_3)dy_1dy_3\\
\leq&\int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2}\sigma^2 f(y_1,y_3)dy_1dy_3\\
+&\left(\frac{\sqrt{n}\sigma(1-\theta)^2}{s_1(s_3-s_1)(1-\theta)^4+2\theta s_3(1-\theta)^2+3\theta^2}\right)^2\\
\times&\int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2}\left(\left[y_3\left\{s_1(1-\theta)^2+2\theta\right\}-y_1\left\{s_1(1-\theta)^2+\theta\right\}\right]\right)^2 f(y_1,y_3)dy_1dy_3,
\end{aligned}
$$

where $f(y_1,y_3)$ is the probability density function of the normal distributed random vector $\left[\frac{X_{s_1}}{\sqrt{n}\sigma(1-\theta)^2},\frac{X_{s_3}}{\sqrt{n}\sigma(1-\theta)^2}\right]$. The last inequality holds because $s_1(s_3-s_1)(1-\theta)^4+2\theta+s_3(1-\theta)^2+3\theta^2>0$. We prove that each of the previous two integrals can be reduced to an integrable function in the variable $x=\frac{s_1}{n}$ and $y=\frac{s_3}{n}$. We have already shown that this holds for the $\int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2}\sigma^2 f(y_1,y_3)dy_1dy_3$ hence consider the second integral. For notation convenience, set

$$
\begin{aligned}
D_1 &= s_1(s_3-s_1)(1-\theta)^4+2\theta s_3(1-\theta)^2+3\theta^2\\
D_2 &= \left[y_3^2\{s_1(1-\theta)^2+2\theta\}+y_1^2\{s_3(1-\theta)^2+\theta\}-2y_1y_3\{s_1(1-\theta)^2\theta\}\right].
\end{aligned}
$$

Note that D_1/n^2 is the determinat of the variance - covariance matrix of the normal distributed random vector $\left[\frac{X_{s_1}}{\sqrt{n}\sigma(1-\theta)^2}, \frac{X_{s_3}}{\sqrt{n}\sigma(1-\theta)^2}\right]$. Since

$$
\begin{aligned}
& y_3^2\{s_1(1-\theta)^2+2\theta\}^2 + y_1^2\{s_1(1-\theta)^2+\theta\}^2 - 2y_1y_3\{s_1(1-\theta)^2+2\theta\}\{s_1(1-\theta)^2+\theta\} \\
=& \{s_1(1-\theta)^2+2\theta\}\left[y_3^2\{s_1(1-\theta)^2+2\theta\} + y_1^2\frac{\{s_1(1-\theta)^2+\theta\}^2}{s_1(1-\theta)^2+2\theta} - 2y_1y_3\{s_1(1-\theta)^2\theta\}\right] \\
\leq& \{s_1(1-\theta)^2+2\theta\}\left[y_3^2\{s_1(1-\theta)^2+2\theta\} + y_1^2\{s_3(1-\theta)^2+\theta\} - 2y_1y_3\{s_1(1-\theta)^2\theta\}\right] \\
=& \{s_1(1-\theta)^2+2\theta\}D_2,
\end{aligned}
$$

we have that

$$
\begin{aligned}
& \left(\frac{\sqrt{n}\sigma(1-\theta)^2}{s_1(s_3-s_1)(1-\theta)^4+2\theta s_3(1-\theta)^2+3\theta^2}\right)^2 \\
& \times \int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2}\left(\left[y_3\left\{s_1(1-\theta)^2+2\theta\right\} - y_1\left\{s_1(1-\theta)^2+\theta\right\}\right]\right)^2 f(y_1,y_3)dy_1dy_3 \\
\leq& \frac{n\sigma^2(1-\theta)^4}{D_1^2}\int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2}\{s_1(1-\theta)^2+2\theta\}D_2\frac{n(1-\theta)^2}{\sqrt{D_1}}\exp\left\{-\frac{n(1-\theta)^2}{D_1}D_2\right\}dy_1dy_3 \\
=& \frac{n\sigma^2(1-\theta)^4}{\sqrt{D_1}}\frac{\{s_1(1-\theta)^2+2\theta\}}{D_1}\int_{\tau_1}^{\tau_2}\int_{\tau_1}^{\tau_2}D_2\frac{n(1-\theta)^2}{D_1}\exp\left\{-D_2\frac{n(1-\theta)^2}{D_1}\right\}dy_1dy_3 \\
\leq& K(\tau_2-\tau_1)^2\frac{n}{\sqrt{D_1}}\frac{\{s_1(1-\theta)^2+2\theta\}}{D_1} \\
\leq& K(\tau_2-\tau_1)^2\frac{n}{\sqrt{D_1}}.
\end{aligned}
$$

The last inequality comes because

$$
\begin{aligned}
& \frac{\{s_1(1-\theta)^2+2\theta\}}{D_1} \\
\leq& \frac{(1-\theta)^2+2\theta}{(s_3-s_1)(1-\theta)^4+2\theta\frac{s_3}{s_1(1-\theta)^2}+3\frac{\theta^2}{s_1}}
\end{aligned}
$$

that is bounded by a constant since $s_3 > s_1$. Hence, the results immediately follows since $\int_0^1\int_0^y \frac{x}{y\sqrt{x(y-x)}}dxdy = \frac{\pi}{2}$ (see Appendix B). Now, we focus on the double products. The inequality $2|\varepsilon_u\varepsilon_v| \leq \varepsilon_u^2 + \varepsilon_v^2$ implies that:

$$\frac{K}{n^2}\sum_{t=3}^{n-1} E\left[\sum_{u=2}^{t-1}\sum_{v<u}\varepsilon_u\varepsilon_v\left\{\sum_{j_1=0}^{t-1-u}\theta^{j_1} I\left(\tau_1 \leq \frac{X_{t-1-j_1}}{\sqrt{n}(1-\theta)\sigma}\leq\tau_2\right)\sum_{j_2=0}^{u-2}\theta^{j_2} I\left(\tau_1 \leq \frac{X_{u-1-j_2}}{\sqrt{n}(1-\theta)\sigma}\leq\tau_2\right)\right\}\right.$$
$$\left.\times\left\{\sum_{j_3=0}^{t-1-v}\theta^{j_3} I\left(\tau_1 \leq \frac{X_{t-1-j_3}}{\sqrt{n}(1-\theta)\sigma}\leq\tau_2\right)\sum_{j_4=0}^{v-2}\theta^{j_4} I\left(\tau_1 \leq \frac{X_{v-1-j_4}}{\sqrt{n}(1-\theta)\sigma}\leq\tau_2\right)\right\}\right]$$
$$\leq\frac{K}{n^2}\sum_{t=3}^{n-1} E\left[\sum_{u=2}^{t-1}\sum_{v<u}|\varepsilon_u\varepsilon_v|\right.$$
$$\left.\times\left\{\sum_{j_1=0}^{t-1-u}|\theta|^{j_1} I\left(\tau_1 \leq \frac{X_{t-1-j_1}}{\sqrt{n}(1-\theta)\sigma}\leq\tau_2\right)\sum_{j_4=0}^{v-2}|\theta|^{j_4} I\left(\tau_1 \leq \frac{X_{v-1-j_4}}{\sqrt{n}(1-\theta)\sigma}\leq\tau_2\right)\right\}\right]$$
$$\leq\frac{K}{n^3}n\sum_{t=3}^{n-1} E\left[\sum_{u=2}^{t-1}\sum_{v<u}(\varepsilon_u^2+\varepsilon_v^2)\right.$$
$$\left.\times\left\{\sum_{j_1=0}^{t-1-u}|\theta|^{j_1} I\left(\tau_1 \leq \frac{X_{t-1-j_1}}{\sqrt{n}(1-\theta)\sigma}\leq\tau_2\right)\sum_{j_4=0}^{v-2}|\theta|^{j_4} I\left(\tau_1 \leq \frac{X_{v-1-j_4}}{\sqrt{n}(1-\theta)\sigma}\leq\tau_2\right)\right\}\right].$$

It is easy to see that the same argument used in the previous case holds and so Condition 4.26 is completely verified. The tightness is proved if we show that there exists a constant $u > 0$ such that

$$E\left[\left|\frac{1}{\sqrt{n}}\sum_{t=2}^{n-1}\varepsilon_t\sum_{j=0}^{t-2}\theta^j I\left(r_i \leq \frac{X_{t-1-j}}{\sqrt{n}\sigma(1-\theta)}\leq r\right)\right|\right] \leq u\sqrt{n}, \text{ for } r_1 \leq r \leq r_i + u.$$

The condition is satisfied since

$$E\left[\left|\frac{1}{\sqrt{n}}\sum_{t=2}^{n-1}\varepsilon_t\sum_{j=0}^{t-2}\theta^j I\left(r_i \leq \frac{X_{t-1-j}}{\sqrt{n}\sigma(1-\theta)}\leq r\right)\right|\right]$$
$$\leq\frac{K}{\sqrt{n}}\sum_{t=2}^{n-1}E\left|\varepsilon_t\right| \leq u\sqrt{n}.$$

Now, consider the partition $[\tau_1 = r_0, r_1, r_2, \ldots, r_{L-1}, r_L = \tau_2]$ such that

$$r_i = r_{i-1} + u, \text{ with } 0 \leq i \leq L-1 \text{ and } r_L - r_{L-1} \leq u.$$

The proof is completed using the same argument of which used in Theorem 2.2 of Chan [1990] and Theorem 22.1 of Billingsley [1968].

Proof of Theorem 59

Let $F_n(\tau) = K_n^{-1}I_nK_n^{-1} - I$. It suffices to demonstrate the desiderated convergence entrywise. By using Theorem 7.3 in Kurtz and Protter [1996] and

Example 11 in Pollard [2012] (the argument is the same of that in proof of Proposition 56), it holds that $F_n(\tau) \to 0$ in probability, for any fixed τ. In fact, when the limiting variable is a constant-valued random variable (i.e. a constant) the weak convergence is equivalent of the convergence in probability. Fix $a < b$ and consider a grid $a = \tau_0 < \tau_1 < \ldots < \tau_m = b$ with equal mesh size, i.e. $\tau_i - \tau_{i-1} \equiv c$, for some $c > 0$. It holds that

$$\sup_{\tau \in [\tau_{i-1}, \tau_i]} \|F_n(\tau) - F_n(\tau_{i-1})\| \leq C_n \quad \text{for all } i.$$

Moreover, $E(C_n) \to 0$ as $c \to 0$. Because for any $\tau \in [a, b]$, there exist an i such that $\tau_{i-1} \leq \tau \leq \tau_i$ and hence

$$F_n(\tau) = F_n(\tau) - F_n(\tau_{i-1}) + F_n(\tau_{i-1})$$
$$\sup_{\tau \in [a,b]} \|F_n(\tau)\| \leq \max_{i=0,\ldots,m} F_n(\tau_i) + C_n.$$

The results follows since for fixed m, $\max_{i=0,\ldots,m} F_n(\tau_i) \to 0$ in probability and $E(C_n) \to 0$ as $c \to 0$ in probability.

Proof of Proposition 55

Since of the MLE admits the asymptotic representation $P_n^{-1}\left(\hat{\boldsymbol{\psi}}_{1,n} - \boldsymbol{\psi}\right) = (I_{1,1}(\tau))^{-1} P_n \frac{\partial \ell}{\partial \boldsymbol{\psi}_1} + o_P(1)$, the essence in deriving the limiting distribution of the proposed test is to demonstrate that

$$Q_n \frac{\partial \hat{\ell}}{\partial \boldsymbol{\psi}_2} = Q_n \frac{\partial \ell}{\partial \boldsymbol{\psi}_2} - I_{2,1}(\tau) P_n^{-1}\left(\hat{\boldsymbol{\psi}}_{1,n} - \boldsymbol{\psi}\right) + o_p(1), \tag{4.27}$$

For notation simplicity, we will omit n in the subindex of $\hat{\boldsymbol{\psi}}_{1,n}$. We prove (4.27) componentwise, i.e. we show that:

$$\frac{1}{\sqrt{n}} \frac{\partial \hat{\ell}}{\partial \phi_{1,0}} = \frac{1}{\sqrt{n}} \frac{\partial \ell}{\partial \phi_{1,0}} - \left\{ \frac{1}{(1-\theta)^2 \sigma^2} \int_0^1 I(W_s \leq \tau) ds \right\} \sqrt{n}\left(\hat{\phi}_0 - \phi_0\right) + o_p(1)$$
$$\frac{1}{n} \frac{\partial \hat{\ell}}{\partial \phi_{1,1}} = \frac{1}{n} \frac{\partial \ell}{\partial \phi_{1,1}} - \left\{ \frac{1}{(1-\theta)\sigma} \int_0^1 W_s I(W_s \leq \tau) ds \right\} \sqrt{n}\left(\hat{\phi}_0 - \phi_0\right) + o_p(1).$$

We prove only the first equality, since the proof of the second one uses the same argument. To this aim, remember that

$$\frac{\partial \hat{\ell}}{\partial \phi_{1,0}} = -\sum_{t=1}^{n} \frac{\hat{\varepsilon}_t}{\hat{\sigma}^2} \frac{\partial \hat{\varepsilon}_t}{\partial \phi_{1,0}} \quad \text{and} \quad \frac{\partial \ell}{\partial \phi_{1,0}} = -\sum_{t=1}^{n} \frac{\varepsilon_t}{\sigma^2} \frac{\partial \varepsilon_t}{\partial \phi_{1,0}}.$$

By routine algebra it is readily checked that

$$\hat{\varepsilon}_t - \varepsilon_t = (\phi_0 - \hat{\phi}_0) \sum_{j=0}^{t-1} \hat{\theta}^j + (\hat{\theta} - \theta) \sum_{j=0}^{t-1} \theta^j \varepsilon_{t-1-j} + \hat{\theta}^t \varepsilon_0;$$
$$\frac{\partial \hat{\varepsilon}_t}{\partial \phi_{1,0}} - \frac{\partial \varepsilon_t}{\partial \phi_{1,0}} = (\hat{\theta} - \theta) \sum_{j=0}^{t-1} \theta^j \frac{\partial \varepsilon_{t-1-j}}{\partial \phi_{1,0}} - \hat{\theta}^t \frac{\partial \varepsilon_0}{\partial \phi_{1,0}}.$$

In fact:

$$\begin{aligned}\hat{\varepsilon}_t - \varepsilon_t &= \left(\phi_0 - \hat{\phi}_0\right) + \hat{\theta}\hat{\varepsilon}_{t-1} - \theta\varepsilon_{t-1} \\ &= \hat{\theta}\left(\hat{\varepsilon}_{t-1} - \varepsilon_{t-1}\right) + (\hat{\theta} - \theta)\varepsilon_{t-1} + \left(\phi_0 - \hat{\phi}_0\right) \\ &\vdots \\ &= (\hat{\theta} - \theta)(\varepsilon_{t-1} + \hat{\theta}\varepsilon_{t-2} + \hat{\theta}^2\varepsilon_{t-3} + \cdots + \hat{\theta}^{t-1}\varepsilon_0) \\ &+ (\phi_0 - \hat{\phi}_0)(1 + \hat{\theta} + \hat{\theta}^2 + \cdots + \hat{\theta}^{t-1}) + \hat{\theta}^t\varepsilon_0 \\ &= (\phi_0 - \hat{\phi}_0)\sum_{j=0}^{t-1}\hat{\theta}^j + (\hat{\theta} - \theta)\sum_{j=0}^{t-1}\theta^j\varepsilon_{t-1-j} - \hat{\theta}^t\varepsilon_0.\end{aligned}$$

$$\begin{aligned}\frac{\partial\hat{\varepsilon}_t}{\partial\phi_{1,0}} - \frac{\partial\varepsilon_t}{\partial\phi_{1,0}} &= \hat{\theta}\frac{\partial\hat{\varepsilon}_{t-1}}{\partial\phi_{1,0}} - \theta\frac{\partial\varepsilon_{t-1}}{\partial\phi_{1,0}} \\ &= \hat{\theta}\left(\frac{\partial\hat{\varepsilon}_{t-1}}{\partial\phi_{1,0}} - \frac{\partial\varepsilon_{t-1}}{\partial\phi_{1,0}}\right) + (\hat{\theta} - \theta)\frac{\partial\varepsilon_{t-1}}{\partial\phi_{1,0}} \\ &\vdots \\ &= (\hat{\theta} - \theta)\left(\frac{\partial\varepsilon_{t-1}}{\partial\phi_{1,0}} + \hat{\theta}\frac{\partial\varepsilon_{t-2}}{\partial\phi_{1,0}} + \cdots + \hat{\theta}^{t-1}\frac{\partial\varepsilon_0}{\partial\phi_{1,0}}\right) - \hat{\theta}^t\frac{\partial\varepsilon_0}{\partial\phi_{1,0}} \\ &= (\hat{\theta} - \theta)\sum_{j=0}^{t-1}\theta^j\frac{\partial\varepsilon_{t-1-j}}{\partial\phi_{1,0}} - \hat{\theta}^t\frac{\partial\varepsilon_0}{\partial\phi_{1,0}}.\end{aligned}$$

Since $\frac{1}{\hat{\sigma}^2} - \frac{1}{\sigma^2} = \frac{\sigma^2 - \hat{\sigma}^2}{\sigma^2\hat{\sigma}^2} = O_p(n^{-1/2})$, we have that:

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial\hat{\ell}}{\partial\phi_{1,0}} &= -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\hat{\varepsilon}_t}{\sigma^2}\frac{\partial\hat{\varepsilon}_t}{\partial\phi_{1,0}} + O(n^{-1/2}) \\ &= \frac{1}{\sqrt{n}}\frac{\partial\ell}{\partial\phi_{1,0}} \\ &+ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\varepsilon_t}{\sigma^2}\frac{\partial\varepsilon_t}{\partial\phi_{1,0}} - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\hat{\varepsilon}_t}{\sigma^2}\frac{\partial\hat{\varepsilon}_t}{\partial\phi_{1,0}} + O(n^{-1/2}).\end{aligned}$$

We focus on the term which is not negligible and therefore the result is proved if we show that:

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\varepsilon_t}{\sigma^2}\frac{\partial\varepsilon_t}{\partial\phi_{1,0}} - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\hat{\varepsilon}_t}{\sigma^2}\frac{\partial\hat{\varepsilon}_t}{\partial\phi_{1,0}} = -\sqrt{n}\left(\hat{\phi}_0 - \phi_0\right)\frac{1}{(1-\theta)^2\sigma^2}\int_0^1 I\left(W_s \leq \tau\right)ds + o_p(1).$$

It holds that

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\varepsilon_t}{\sigma^2}\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}-\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\hat{\varepsilon}_t}{\sigma^2}\frac{\partial\hat{\varepsilon}_t}{\partial\phi_{1,0}}\\
=&\frac{\sqrt{n}}{n\sigma^2}\sum_{t=1}^{n}\left\{\varepsilon_t\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}-\hat{\varepsilon}_t\frac{\partial\hat{\varepsilon}_t}{\partial\phi_{1,0}}\right\}\\
=&\frac{\sqrt{n}}{n\sigma^2}\sum_{t=1}^{n}\left\{\varepsilon_t\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}-\hat{\varepsilon}_t\frac{\partial\hat{\varepsilon}_t}{\partial\phi_{1,0}}+\left(\hat{\phi}_0-\phi_0\right)\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\frac{\partial\varepsilon_t}{\partial\phi_0}\right\}\\
-&\sqrt{n}\left(\hat{\phi}_0-\phi_0\right)\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\sigma^2}\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\frac{\partial\varepsilon_t}{\partial\phi_0}.
\end{aligned}
$$

Since

$$
\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\sigma^2}\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\frac{\partial\varepsilon_t}{\partial\phi_0}\rightsquigarrow\frac{1}{(1-\theta)^2\sigma^2}\int_0^1 I\left(W_s\leq\tau\right)ds,
$$

it remains to verify that

$$
\frac{\sqrt{n}}{n\sigma^2}\sum_{t=1}^{n}\left\{\varepsilon_t\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}-\hat{\varepsilon}_t\frac{\partial\hat{\varepsilon}_t}{\partial\phi_{1,0}}+\left(\hat{\phi}_0-\phi_0\right)\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\frac{\partial\varepsilon_t}{\partial\phi_0}\right\}=o_p(1).
$$

It holds that

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\varepsilon_t}{\sigma^2}\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}-\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\hat{\varepsilon}_t}{\sigma^2}\frac{\partial\hat{\varepsilon}_t}{\partial\phi_{1,0}}\\
=&\frac{\sqrt{n}}{n\sigma^2}\sum_{t=1}^{n}\left\{\varepsilon_t\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}-\hat{\varepsilon}_t\frac{\partial\hat{\varepsilon}_t}{\partial\phi_{1,0}}+\left(\hat{\phi}_0-\phi_0\right)\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\frac{\partial\varepsilon_t}{\partial\phi_0}\right\}\\
-&\sqrt{n}\left(\hat{\phi}_0-\phi_0\right)\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\sigma^2}\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\frac{\partial\varepsilon_t}{\partial\phi_0}.
\end{aligned}
$$

Since

$$
\frac{1}{n}\sum_{t=1}^{n}\frac{1}{\sigma^2}\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\frac{\partial\varepsilon_t}{\partial\phi_0}\rightsquigarrow\frac{1}{(1-\theta)^2\sigma^2}\int_0^1 I\left(W_s\leq\tau\right)ds,
$$

it remains to verify that

$$
\frac{\sqrt{n}}{n\sigma^2}\sum_{t=1}^{n}\left\{\varepsilon_t\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}-\hat{\varepsilon}_t\frac{\partial\hat{\varepsilon}_t}{\partial\phi_{1,0}}+\left(\hat{\phi}_0-\phi_0\right)\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\frac{\partial\varepsilon_t}{\partial\phi_0}\right\}=o_p(1).
$$

It holds that

$$\begin{aligned}
&\frac{\sqrt{n}}{n\sigma^2}\sum_{t=1}^{n}\left\{\varepsilon_t\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}-\hat{\varepsilon}_t\frac{\partial\hat{\varepsilon}_t}{\partial\phi_{1,0}}+\left(\hat{\phi}_0-\phi_0\right)\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\frac{\partial\varepsilon_t}{\partial\phi_0}\right\}\\
=&\frac{1}{\sqrt{n}\sigma^2}\sum_{t=1}^{n}\left\{\varepsilon_t\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}-\hat{\varepsilon}_t\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}+\hat{\varepsilon}_t\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}-\hat{\varepsilon}_t\frac{\partial\hat{\varepsilon}_t}{\partial\phi_{1,0}}+\left(\hat{\phi}_0-\phi_0\right)\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\frac{\partial\varepsilon_t}{\partial\phi_0}\right\}\\
=&\frac{1}{\sqrt{n}\sigma^2}\sum_{t=1}^{n}\left\{\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\left(\hat{\phi}_0-\phi_0\right)\sum_{j=0}^{t-1}\hat{\theta}^j-\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\left(\hat{\theta}-\theta\right)\sum_{j=0}^{t-1}\theta^j\varepsilon_{t-1-j}-\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\hat{\theta}^t\varepsilon_0\right.\\
&\left.-\hat{\varepsilon}_t\left[(\hat{\theta}-\theta)\sum_{j=0}^{t-1}\theta^j\frac{\partial\varepsilon_{t-1-j}}{\partial\phi_{1,0}}-\hat{\theta}^t\frac{\partial\varepsilon_0}{\partial\phi_{1,0}}\right]+\left(\hat{\phi}_0-\phi_0\right)\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\frac{\partial\varepsilon_t}{\partial\phi_0}\right\}
\end{aligned}$$

Since $\frac{1}{\sqrt{n}}\sum_{j=0}^{t-1}\hat{\theta}^j=\frac{1}{\sqrt{n}}\sum_{j=0}^{t-1}\theta^j+O(n^{-1/2})$, then

$$\frac{1}{\sqrt{n}\sigma^2}\left\{\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\left(\hat{\phi}_0-\phi_0\right)\sum_{j=0}^{t-1}\hat{\theta}^j+\left(\hat{\phi}_0-\phi_0\right)\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\frac{\partial\varepsilon_t}{\partial\phi_0}\right\}=o_p(1).$$

We claim that:

$$\begin{aligned}
\frac{1}{\sqrt{n}\sigma^2}\sum_{t=1}^{n}&\left\{-\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\left(\hat{\theta}-\theta\right)\sum_{j=0}^{t-1}\theta^j\varepsilon_{t-1-j}-\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\hat{\theta}^t\varepsilon_0\right.\\
&\left.-\hat{\varepsilon}_t\left[(\hat{\theta}-\theta)\sum_{j=0}^{t-1}\theta^j\frac{\partial\varepsilon_{t-1-j}}{\partial\phi_{1,0}}-\hat{\theta}^t\frac{\partial\varepsilon_0}{\partial\phi_{1,0}}\right]\right\}=o_p(1).
\end{aligned}$$

As each it is possible to show that each of the terms is negligible by using the same argument, we below prove

$$\frac{1}{\sqrt{n}\sigma^2}\sum_{t=1}^{n}-\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\left(\hat{\theta}-\theta\right)\sum_{j=0}^{t-1}\theta^j\varepsilon_{t-1-j}=o_p(1).$$

Since $(\hat{\theta}-\theta)=O_p(n^{-1/2})$, it holds that

$$\begin{aligned}
&\frac{1}{\sqrt{n}\sigma^2}\sum_{t=1}^{n}-\frac{\partial\varepsilon_t}{\partial\phi_{1,0}}\left(\hat{\theta}-\theta\right)\sum_{j=0}^{t-1}\theta^j\varepsilon_{t-1-j}\\
=&O_p(1)\times\frac{1}{n\sigma^2}\sum_{t=1}^{n}\left\{\sum_{j=0}^{t-1}\theta^jI\left(X_{t-1-j}\le r\right)\right\}\left\{\sum_{j=0}^{t-1}\theta^j\varepsilon_{t-1-j}\right\}.
\end{aligned}$$

The result follows if:

$$N(\tau)=\frac{1}{n\sigma^2}\sum_{t=1}^{n}\left\{\sum_{j=0}^{t-1}\theta^jI\left(\frac{X_{t-1-j}}{\sqrt{n}\sigma(1-\theta)}\le\tau\right)\right\}\left\{\sum_{j=0}^{t-1}\theta^j\varepsilon_{t-1-j}\right\}=o_p(1).$$

It suffices to verify that

(i) $N(\tau)$ is asymptotically tight;

(ii) $E\left[\left(\frac{1}{n\sigma^2}\sum_{t=1}^{n}\left\{\sum_{j=0}^{t-1}\theta^j I\left(\frac{X_{t-1-j}}{\sqrt{n}\sigma(1-\theta)}\leq\tau\right)\right\}\left\{\sum_{j=0}^{t-1}\theta^j\varepsilon_{t-1-j}\right\}\right)^2\right]\xrightarrow[n\to\infty]{}0.$

To prove (ii), note that:

$$
\begin{aligned}
&E\left[\left(\frac{1}{n\sigma^2}\sum_{t=1}^{n}\left\{\sum_{j=0}^{t-1}\theta^j I\left(\frac{X_{t-1-j}}{\sqrt{n}\sigma(1-\theta)}\leq\tau\right)\right\}\left\{\sum_{j=0}^{t-1}\theta^j\varepsilon_{t-1-j}\right\}\right)^2\right]\\
=&\frac{1}{n^2\sigma^4}E\left[\sum_{t=1}^{n}\left\{\sum_{j_1=0}^{t-1}\theta^{j_1}I\left(\frac{X_{t-1-j_1}}{\sqrt{n}\sigma(1-\theta)}\leq\tau\right)\right\}^2\left\{\sum_{j_2=0}^{t-1}\theta^{j_2}\varepsilon_{t-1-j_2}\right\}^2\right]\\
+&\frac{1}{n^2\sigma^4}E\left[\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\left\{\sum_{j_1=0}^{t_1-1}\theta^{j_1}I\left(\frac{X_{t_1-1-j_1}}{\sqrt{n}\sigma(1-\theta)}\leq\tau\right)\right\}\left\{\sum_{j_2=0}^{t_1-1}\theta^{j_2}\varepsilon_{t_1-1-j_2}\right\}\right.\\
\times&\left.\left\{\sum_{j_3=0}^{t_2-1}\theta^{j_3}I\left(\frac{X_{t_2-1-j_3}}{\sqrt{n}\sigma(1-\theta)}\leq\tau\right)\right\}\left\{\sum_{j_4=0}^{t_2-1}\theta^{j_4}\varepsilon_{t_2-1-j_4}\right\}\right].
\end{aligned}
$$

We prove separately that the two expectations are $o_p(1)$.

$$
\begin{aligned}
&\frac{1}{n^2\sigma^4}E\left[\sum_{t=1}^{n}\left\{\sum_{j_1=0}^{t-1}\theta^{j_1}I\left(\frac{X_{t-1-j_1}}{\sqrt{n}\sigma(1-\theta)}\leq\tau\right)\right\}^2\left\{\sum_{j_2=0}^{t-1}\theta^{j_2}\varepsilon_{t-1-j_2}\right\}^2\right]\\
\leq&\frac{1}{n^2\sigma^4}E\left[\sum_{t=1}^{n}\left\{\sum_{j_1=0}^{t-1}|\theta|^{j_1}I\left(\frac{X_{t-1-j_1}}{\sqrt{n}\sigma(1-\theta)}\leq\tau\right)\right\}^2\left\{\sum_{j_2=0}^{t-1}|\theta|^{j_2}|\varepsilon_{t-1-j_2}|\right\}^2\right]\\
\leq&\frac{K}{n^2\sigma^4}E\left[\sum_{t=1}^{n}\left\{\sum_{j_1=0}^{t-1}|\theta|^{j_1}\right\}\left\{\sum_{j_2=0}^{t-1}|\theta|^{j_2}\varepsilon^2_{t-1-j_2}\right\}\right]\leq\frac{K}{n^2\sigma^4}E\left[\sum_{t=1}^{n}\left\{\sum_{j_2=0}^{t-1}|\theta|^{j_2}\varepsilon^2_{t-1-j_2}\right\}\right]\\
=&\frac{K}{n^2\sigma^4}\sum_{t=1}^{n}\sum_{j_2=0}^{t-1}|\theta|^{j_2}E\left[\varepsilon^2_{t-1-j_2}\right]=\frac{K}{n^2\sigma^4}\sum_{t=1}^{n}\sum_{j_2=0}^{t-1}|\theta|^{j_2}=o_p(1).
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
&\frac{1}{n^2\sigma^4}E\left[\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\left\{\sum_{j_1=0}^{t_1-1}\theta^{j_1}I\left(\frac{X_{t_1-1-j_1}}{\sqrt{n}\sigma(1-\theta)}\leq\tau\right)\right\}\left\{\sum_{j_2=0}^{t_1-1}\theta^{j_2}\varepsilon_{t_1-1-j_2}\right\}\right.\\
&\left.\times\left\{\sum_{j_3=0}^{t_2-1}\theta_3^{j}I\left(\frac{X_{t_2-1-j_3}}{\sqrt{n}\sigma(1-\theta)}\leq\tau\right)\right\}\left\{\sum_{j_4=0}^{t_2-1}\theta_4^{j}\varepsilon_{t_2-1-j_4}\right\}\right]\\
\leq&\frac{1}{n^2\sigma^4}E\left[\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\left\{\sum_{j_1=0}^{t_1-1}|\theta|^{j_1}I\left(\frac{X_{t_1-1-j_1}}{\sqrt{n}\sigma(1-\theta)}\leq\tau\right)\right\}\left\{\sum_{j_2=0}^{t_1-1}|\theta|^{j_2}|\varepsilon_{t_1-1-j_2}|\right\}\right.\\
&\left.\times\left\{\sum_{j_3=0}^{t_2-1}|\theta|^{j_3}I\left(\frac{X_{t_2-1-j_3}}{\sqrt{n}\sigma(1-\theta)}\leq\tau\right)\right\}\left\{\sum_{j_4=0}^{t_2-1}|\theta|^{j_4}|\varepsilon_{t_2-1-j_4}|\right\}\right]\\
\leq&\frac{K}{n^2\sigma^4}E\left[\sum_{j_1=0}^{n-1}|\theta|^{j_1}\sum_{j_2=0}^{n-1}|\theta|^{j_2}\sum_{j_3=0}^{n-1}|\theta|^{j_3}\sum_{j_4=0}^{n-1}|\theta|^{j_4}\right.\\
&\left.\times\sum_{s_1=1}^{n-1}\sum_{s_2=1}^{n-1}|\varepsilon_{s_1}\varepsilon_{s_2}|I\left(\frac{X_{s_1}}{\sqrt{n}\sigma(1-\theta)}\leq\tau\right)I\left(\frac{X_{s_2}}{\sqrt{n}\sigma(1-\theta)}\leq\tau\right)\right]\\
\leq&\frac{K}{n^2\sigma^4}E\left[\sum_{j_1=0}^{n-1}|\theta|^{j_1}\sum_{j_2=0}^{n-1}|\theta|^{j_2}\sum_{j_3=0}^{n-1}|\theta|^{j_3}\sum_{j_4=0}^{n-1}|\theta|^{j_4}\right.\\
&\left.\times\sum_{s_1=1}^{n-1}\sum_{s_2=1}^{n-1}\varepsilon_{s_1}^2I\left(\frac{X_{s_1}}{\sqrt{n}\sigma(1-\theta)}\leq\tau\right)I\left(\frac{X_{s_2}}{\sqrt{n}\sigma(1-\theta)}\leq\tau\right)\right]\\
+&\frac{K}{n^2\sigma^4}E\left[\sum_{j_1=0}^{n-1}|\theta|^{j_1}\sum_{j_2=0}^{n-1}|\theta|^{j_2}\sum_{j_3=0}^{n-1}|\theta|^{j_3}\sum_{j_4=0}^{n-1}|\theta|^{j_4}\right.\\
&\left.\times\sum_{s_1=1}^{n-1}\sum_{s_2=1}^{n-1}\varepsilon_{s_2}^2I\left(\frac{X_{s_1}}{\sqrt{n}\sigma(1-\theta)}\leq\tau\right)I\left(\frac{X_{s_2}}{\sqrt{n}\sigma(1-\theta)}\leq\tau\right)\right].
\end{aligned}
$$

The result follows by using the same argument used in the proof Proposition 57 upon noting that

$$\lim_{n\to\infty}\int_{-\infty}^{\tau}\int_{-\infty}^{\tau}\frac{1}{n^\alpha}dy_1dy_2=0,\qquad\text{for each }\alpha>2.$$

Finally the proof of the tightness is equal to which derived in Proposition 57 and hence the result is completely verified.

Proof of Proposition 61

Let Λ_n be the logarithm of the likelihood ratio $\log \frac{dP_{1,n}}{dP_{0,n}}$. Under the null hypothesis, it holds that

$$\begin{aligned}
\Lambda_n =& -\frac{1}{2\sigma^2}\sum_{t=1}^{n}\left[X_t - X_{t-1} - \left(\frac{h_0}{\sqrt{n}} + \frac{h_1}{n}X_{t-1}\right) I\left(\frac{X_{t-1}}{\sigma\sqrt{n}(1-\theta)} \le \tau_0\right) + \theta\varepsilon_{t-1}\right]^2 \\
&+ \frac{1}{2\sigma^2}\sum_{t=1}^{n}[X_t - X_{t-1} + \theta\varepsilon_{t-1}]^2 \\
&= -\frac{1}{2\sigma^2}\sum_{t=1}^{n}[X_t - X_{t-1} + \theta\varepsilon_{t-1}]^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{n}\left[\left\{\frac{h_0}{\sqrt{n}} + \frac{h_1}{n}X_{t-1}\right\}^2 I\left(\frac{X_{t-1}}{\sigma\sqrt{n}(1-\theta)} \le \tau_0\right)\right] \\
&+ \frac{1}{\sigma^2}\sum_{t=1}^{n}\left[\{X_t - X_{t-1} + \theta\varepsilon_{t-1}\}\left\{\frac{h_0}{\sqrt{n}} + \frac{h_1}{n}X_{t-1}\right\} I\left(\frac{X_{t-1}}{\sigma\sqrt{n}(1-\theta)} \le \tau_0\right)\right] \\
&+ \frac{1}{2\sigma^2}\sum_{t=1}^{n}[X_t - X_{t-1} + \theta\varepsilon_{t-1}]^2 \\
=& -\frac{1}{2\sigma^2}\sum_{t=1}^{n}\left[\left\{\frac{h_0^2}{n} + \frac{h_1^2}{n}\frac{X_{t-1}^2}{n} + \frac{2h_0h_1}{n}\frac{X_{t-1}}{\sqrt{n}}\right\} I\left(\frac{X_{t-1}}{\sigma\sqrt{n}(1-\theta)} \le \tau_0\right)\right] \\
&+ \frac{1}{\sigma^2}\sum_{t=1}^{n}\left[\left\{h_0\frac{\varepsilon_t}{\sqrt{n}} + h_1\frac{\varepsilon_t}{\sqrt{n}}\frac{X_{t-1}}{\sqrt{n}}\right\} I\left(\frac{X_{t-1}}{\sigma\sqrt{n}(1-\theta)} \le \tau_0\right)\right].
\end{aligned}$$

The last inequality holds because, under the null hypothesis, $X_t - X_{t-1} + \theta\varepsilon_{t-1} = \varepsilon_t$. As n tends to infinity Λ_n converges weakly to

$$\begin{aligned}
\bar{\Lambda} =& -\frac{h_0^2}{2\sigma^2}\int_0^1 I(W_t \le \tau_0)\,dt - \frac{(1-\theta)^2 h_1^2}{2}\int_0^1 W_t^2 I(W_t \le \tau_0)\,dt \\
&- \frac{(1-\theta)h_1h_0}{\sigma}\int_0^1 W_t I(W_t \le \tau_0)\,dt \\
&+ \frac{h_0}{\sigma}\int_0^1 I(W_t \le \tau_0)\,dW_t + (1-\theta)h_1\int_0^1 W_t I(W_t \le \tau_0)\,dW_t.
\end{aligned}$$

It is easy to see that $\bar{\Lambda}$ can be written in terms of the Ito's process $\{X_t\}$ as follows:

$$\begin{aligned}
\bar{\Lambda} &= X_t - \frac{1}{2}[X_t], \\
X_t &= \int_0^t \delta_s dW_s
\end{aligned}$$

with $\delta_s = \frac{h_0}{\sigma}I(W_s \le \tau_0) + (1-\theta)h_1 W_s I(W_s \le \tau_0)$.
Set $\Upsilon(t) = \exp\left(X_t - \frac{1}{2}[X]_t\right) = \lim_{n\to\infty}\frac{dP_{1,n}}{dP_{0,n}}$. Note that $\Upsilon(t)$ is the stochastic exponential of the process $\{X_t\}$ and we claim that it is a martingale. Since

$\Upsilon(0) = 1$, $E[\Upsilon(t)] = 1$, for each t and therefore Le Cam's first lemma [van der Vaart, 1998, Lemma 6, pag 88] is fulfilled. Hence the proof is completed if we verify the martingale claim. From Kazamaki [1977], it suffices to verify that

$$E\left[e^{\frac{1}{2}\int_0^1 \delta_s^2 ds}\right] < \infty.$$

Let $C_1 = \frac{(1-\theta)^2 h_1^2}{2}$ and $C_2 = \frac{2(1-\theta)|h_0 h_1|}{\sigma}$ and consider

$$\begin{aligned}
E\left[e^{\frac{1}{2}\int_0^1 \delta_s^2 ds}\right] =& E\left[\exp\left\{\frac{1}{2}\int_0^1 \left[\frac{h_0^2}{\sigma^2} I\left(W_s \le \tau_0\right) + (1-\theta)^2 h_1^2 W_s^2 I\left(W_s \le \tau_0\right)\right.\right.\right. \\
&\left.\left.\left. +2(1-\theta)\frac{h_0 h_1}{\sigma} W_s I\left(W_s \le \tau_0\right)\right] ds\right\}\right] \\
\le& KE\left[\exp\left\{C_1 \int_0^1 W_s^2 ds + C_2 \sqrt{\int_0^1 W_s^2 ds}\right\}\right] \\
=& K\int_0^\infty \exp\left\{C_1 y + C_2\sqrt{y}\right\} dF(y),
\end{aligned}$$

where $K = e^{h_0^2/\sigma^2}$ and $F(y)$ is cdf of the random variable $Y = \int_0^1 W_s^2 ds$. The last integral is finite if and only if $\int_1^\infty \exp\left\{C_1 y + C_2\sqrt{y}\right\} dF(y)$ is finite. Since $\sqrt{y} = o(y)$, it suffices to prove that $\int_1^\infty \exp\left\{C_1 y\right\} dF(y) < \infty$. It holds that:

$$\begin{aligned}
&\int_1^\infty e^{C_1 y} dF(y) \le \int_1^\infty e^{C_1 y} dF(y) \\
=& \int_1^\infty \left[e^{C_1 y} - 1\right] dF(y) + \int_1^\infty dF(y) \\
=& \int_0^\infty \int_0^y \frac{1}{C_1} e^{C_1 x} dx dF(y) + 1 = \frac{1}{C_1}\int_1^\infty \int_x^\infty dF(y) e^{C_1 x} dx + 1.
\end{aligned}$$

Therefore, it suffices to verify that

$$\int_0^\infty \bar{F}(x) e^{C_1 x} dx < \infty, \quad \text{where } \bar{F}(x) = 1 - F(x).$$

From Li [1992, lemma 2], it follows that

$$\bar{F}(x) \sim K x^{-1/2} e^{-\frac{x}{2\lambda}}, \quad \text{with } \lambda = \frac{4}{\pi^2}.$$

Since $h_1 \in \left(-\frac{\pi}{2(1-\theta)}, 0\right)$, the integral $\int_0^\infty x^{-1/2} e^{\left(-\frac{\pi^2}{8} + C_1\right)x} dx$ is finite and this completes the proof.

Proof of Proposition 62

The result readily follows by applying Girsanov's formula [Girsanov, 1960].

Proof of Corollary 63

Since $W_0 = 0$ and $\tau_0 > 0$, the process starts in the lower regime, i.e.:

$$dW_0 = \tilde{\theta}\left(\tilde{\mu} - W_0\right)dt + d\tilde{W}_0 = \tilde{\theta}\tilde{\mu}dt + d\tilde{W}_0.$$

Let $\bar{t}$ be the stopping time that indicates the first time at which the process jumps in the upper regime:

$$\bar{t} = \inf\left\{t \in [0,1] | W_t > \tau_0\right\}.$$

We claim that $P(\bar{t} < 1)$ tends to zero as $\tilde{\mu}$ tends to $-\infty$. In fact, since $\tilde{\mu}$ tends to $-\infty$, it is possible to choose an arbitrary fixed $\lambda > 0$ and a sufficient small $\tilde{\mu}$ such that $\tilde{\mu} + \lambda < \tau_0$. Hence, by using the Chebyshev's inequality, it follows that:

$$\begin{aligned} P(W_t > \tau_0) &< P(W_t > \tilde{\mu} + \lambda) \\ &< P(W_t > \tilde{\mu} + \lambda \quad \cup \quad W_t < \tilde{\mu} - \lambda) \\ &= P(|W_t - \tilde{\mu}| > \lambda) < \frac{1}{\lambda^2}. \end{aligned}$$

Since λ is arbitrary, the statement is verified.

Proof of Proposition 64

We prove that the stochastic differential equation (4.17) converges in probability to the deterministic ordinary differential equation (ODE) of which the function $\mathcal{G}_s$, $0 \le s \le 1$, is the solution. Consider the solution of (4.16)

$$\begin{aligned} W_t &= \tilde{\mu}\left(1 - e^{-\tilde{\theta}t}\right) + \int_0^t e^{-\tilde{\theta}(t-s)}d\tilde{W}_s \\ \Rightarrow \frac{W_t}{|\tilde{\mu}|} &= \mathcal{G}_s + \int_0^t e^{-\tilde{\theta}(t-s)}d\frac{\tilde{W}_s}{|\tilde{\mu}|}. \end{aligned}$$

The result then follows since

$$\int_0^t e^{-\tilde{\theta}(t-s)}d\frac{\tilde{W}_s}{|\tilde{\mu}|} \xrightarrow[|\tilde{\mu}|\to\infty]{p} 0$$

because

$$\int_0^t e^{-\tilde{\theta}(t-s)}d\frac{\tilde{W}_s}{|\tilde{\mu}|} = e^{-\tilde{\theta}(t-s)} \times \frac{1}{|\tilde{\mu}|}\int_0^t d\tilde{W},$$

with mean

$$E\left[\frac{1}{|\tilde{\mu}|}\int_0^t d\tilde{W}\right] = 0$$

and variance

$$\text{Var}\left[\frac{1}{|\tilde{\mu}|}\int_0^t d\tilde{W}\right] = \frac{t}{|\tilde{\mu}|} \xrightarrow[|\tilde{\mu}|\to\infty]{} 0.$$

Proof of Proposition 65

We start proving that for $i = 1, 2$

$$q_i\left(\frac{W_s}{|\tilde{\mu}|}, 0 \le s \le 1\right) \xrightarrow[|\tilde{\mu}|\to\infty]{p} q_i\left(\mathcal{G}_s, 0 \le s \le 1\right).$$

Given $\alpha \in [0,1]$, and let $q_\alpha\left(\mathcal{G}_s, 0 \le s \le 1\right)$ be the upper $\alpha \times 100$ quantile. It holds that

$$\begin{aligned}&\ell\left(\{s \in [0,1] | \mathcal{G}_s > q_\alpha\left(\mathcal{G}_s, 0 \le s \le 1\right)\}\right) = \alpha\\ &\ell\left(\{s \in [0,1] | \mathcal{G}_s \le q_\alpha\left(\mathcal{G}_s, 0 \le s \le 1\right)\}\right) = 1 - \alpha.\end{aligned}$$

Since $\left(\frac{W_s}{|\tilde{\mu}|}, 0 \le s \le 1\right) \xrightarrow[|\tilde{\mu}|\to\infty]{p} \left(\mathcal{G}_s, 0 \le s \le 1\right)$ uniformly in probability , it holds in probability that for all sufficiently large $|\tilde{\mu}|$ such that for each $\xi > 0$

$$\begin{aligned}&\ell\left(\left\{s \in [0,1] | \frac{W_s}{|\tilde{\mu}|} < q_{\alpha-\xi}\left(\mathcal{G}_s, 0 \le s \le 1\right)\right\}\right) > 1 - \alpha;\\ &\ell\left(\left\{s \in [0,1] | \frac{W_s}{|\tilde{\mu}|} > q_{\alpha+\xi}\left(\mathcal{G}_s, 0 \le s \le 1\right)\right\}\right) > \alpha.\end{aligned}$$

Therefore,

$$q_{\alpha+\xi}\left(\mathcal{G}_s, 0 \le s \le 1\right) \le q_\alpha\left(\frac{W_s}{|\tilde{\mu}|}, 0 \le s \le 1\right) \le q_{\alpha-\xi}\left(\mathcal{G}_s, 0 \le s \le 1\right).$$

Since $q_\alpha\left(\mathcal{G}_s, 0 \le s \le 1\right)$ is continuous in α and ξ can be chosen arbitrarily small the proof is completed. Now, we prove that

$$\int_0^1 I(W_s \le \tau) ds \xrightarrow[|\tilde{\mu}|\to\infty]{p} \int_0^1 I(\mathcal{G}_s \le \tilde{\tau}) ds.$$

It suffices to show that, given $\tilde{\tau}$, it follows that

$$\int_0^1 I\left(\frac{W_s}{|\tilde{\mu}|} \le \tilde{\tau}\right) ds \xrightarrow[|\tilde{\mu}|\to\infty]{p} \int_0^1 I(\mathcal{G}_s \le \tilde{\tau}) ds. \tag{4.28}$$

In fact, let $q_1\left(W_s, 0 \le s \le 1\right)$ and $q_2\left(W_s, 0 \le s \le 1\right)$ be two fixed quantiles for $\{W_s, 0 \le s \le 1\}$ and choose $\tau \in [q_1, q_2]$. Therefore, there exists $\tilde{s}$ such that

$\tau = \tilde{s}q_1\left(W_s, 0 \le s \le 1\right) + (1-\tilde{s})q_2\left(W_s, 0 \le s \le 1\right)$. We have that

$$\begin{aligned}
&\int_0^1 I\left(W_s \le \tau\right) ds \\
=& \int_0^1 I\left(W_s \le \tilde{s}q_1\left(W_s, 0 \le s \le 1\right) + (1-\tilde{s})q_2\left(W_s, 0 \le s \le 1\right)\right) ds \\
=& \int_0^1 I\left(|\tilde{\mu}|\frac{W_s}{|\tilde{\mu}|} \le \tilde{s}q_1\left(|\tilde{\mu}|\frac{W_s}{|\tilde{\mu}|}, 0 \le s \le 1\right) + (1-\tilde{s})q_2\left(|\tilde{\mu}|\frac{W_s}{|\tilde{\mu}|}, 0 \le s \le 1\right)\right) ds \\
=& \int_0^1 I\left(|\tilde{\mu}|\frac{W_s}{|\tilde{\mu}|} \le \tilde{s}|\tilde{\mu}|q_1\left(\frac{W_s}{|\tilde{\mu}|}, 0 \le s \le 1\right) + (1-\tilde{s})|\tilde{\mu}|q_2\left(\frac{W_s}{|\tilde{\mu}|}, 0 \le s \le 1\right)\right) ds \\
=& \int_0^1 I\left(\frac{W_s}{|\tilde{\mu}|} \le \tilde{s}q_1\left(\frac{W_s}{|\tilde{\mu}|}, 0 \le s \le 1\right) + (1-\tilde{s})q_2\left(\frac{W_s}{|\tilde{\mu}|}, 0 \le s \le 1\right)\right) ds.
\end{aligned}$$

On the other hand, (4.28) would immediately follows if the step function $I(x \le \tau)$ was a continuous function. Unfortunately, this is not the case but it is discontinuous only at $\tilde{\tau}$. We use the following uniform approximation argument [Pollard, 2012, Example 11, p.70]: Let $G, G_1, G_2, \ldots$ be a sequence of random elements in a metric space $(\mathcal{X}, d)$, with the support of G being a separable set of completely regular elements. Suppose for each $\epsilon > 0, \delta > 0$, there exist approximating random elements $AG, AG_1, AG_2, \ldots$ such that

(i) $P^*\{d(G, AG) > \epsilon\} < \delta$;

(ii) $\limsup_{n\to\infty} P^*\{d(G_n, AG_n\} > \epsilon\} < \delta$;

(iii) $AG_n \rightsquigarrow AG$,

where the notation $\rightsquigarrow$ denotes weak convergence, $P^*(\cdot)$ denotes the outer probability measure of the enclosed expression. Then $G_n \rightsquigarrow G$, as $n \to \infty$. Given a fixed and positive number $0 < \delta < 1$, define the continuous function $G_\delta(x)$ such that $|G_\delta(x) - I(x \le \tau)| \le H_\delta(x)$ with the bound $H_\delta(x)$ being uniformly bounded, continuous functions and with support inside $[\tilde{\tau} - \delta, \tilde{\tau} + \delta]$. We have that

$$\begin{aligned}
&\left|\int_0^1 I\left(\frac{W_s}{|\tilde{\mu}|} \le \tilde{\tau}\right) ds - \int_0^1 G_\delta\left(\frac{W_s}{|\tilde{\mu}|}\right) ds\right| \\
\le& \int_0^1 \left|I\left(\frac{W_s}{|\tilde{\mu}|} \le \tilde{\tau}\right) - G_\delta\left(\frac{W_s}{|\tilde{\mu}|}\right)\right| ds \le \int_0^1 H_\delta\left(\frac{W_s}{|\tilde{\mu}|}\right) ds.
\end{aligned}$$

For each fixed positive number δ,

$$P\left(\frac{W_s}{|\tilde{\mu}|} \in (\tilde{\tau} - \delta, \tilde{\tau} + \delta)\right) \xrightarrow[|\tilde{\mu}|\to\infty]{} 0$$

uniformly for each $0 \le t \le 1$ such that $e^{-\tilde{\theta}t} - 1 \notin (\tilde{\tau} - 2\delta, \tilde{\tau} + 2\delta)$. Note that the quantity $\tilde{\tau} + 1$ is positive. we prove this claim by contradiction. Since

$\tilde{\tau}$ is a quantile of the distribution implies by $\mathcal{G}_t, 0 \leq t \leq 1$, it holds that $P\left(e^{-\tilde{\theta}} - 1 \leq \tilde{\tau}\right) > 0$ and therefore $P\left(e^{-\tilde{\theta}} \leq \tilde{\tau} + 1\right) > 0$ that is impossible if $\tilde{\tau} + 1 < 0$.
Now, we compute the expectation of the previous integral

$$E\left[\int_0^1 H_\delta\left(\frac{W_s}{|\tilde{\mu}|}\right) ds\right] = \int_0^1 E\left[H_\delta\left(\frac{W_s}{|\tilde{\mu}|}\right)\right] ds$$

$$\leq K \int_0^1 P\left(\frac{W_s}{|\tilde{\mu}|} \in (\tilde{\tau} - \delta, \tilde{\tau} + \delta)\right) ds$$

$$= K \int_{a_\delta}^{b_\delta} P\left(\frac{W_s}{|\tilde{\mu}|} \in (\tilde{\tau} - \delta, \tilde{\tau} + \delta)\right) ds + K \int_{[0,1]\cap[a_\delta, b_\delta]^c} P\left(\frac{W_s}{|\tilde{\mu}|} \in (\tilde{\tau} - \delta, \tilde{\tau} + \delta)\right) ds$$

with $0 < K < \infty$ being the upper bound of all H_δ, $0 < \delta < 1$ and

$$a_\delta = \min\left(\max\left(0, -\frac{1}{\tilde{\theta}} \log(1 + \tilde{\tau} + 2\delta)\right), 1\right),$$

$$b_\delta = \min\left(\max\left(0, -\frac{1}{\tilde{\theta}} \log(1 + \tilde{\tau} - 2\delta)\right), 1\right).$$

Therefore for each fixed $\delta > 0$, $P\left(\frac{W_s}{|\tilde{\mu}|} \in (\tilde{\tau} - \delta, \tilde{\tau} + \delta)\right) \xrightarrow[|\tilde{\mu}|\to\infty]{} 0$ uniformly on $s \in [0,1] \cap [a_\delta, b_\delta]^c$ since $e^{-\tilde{\theta}s} - 1 \in (\tilde{\tau} - 2\delta, \tilde{\tau} + 2\delta)$ if and only if $s \in [a_\delta, b_\delta]$.
Hence

$$K \int_{[0,1]\cap[a_\delta, b_\delta]^c} P\left(\frac{W_s}{|\tilde{\mu}|} \in (\tilde{\tau} - \delta, \tilde{\tau} + \delta)\right) ds \xrightarrow[|\tilde{\mu}|\to\infty]{} 0$$

So that, for each $\xi > 0$, it is possible to chose $|\tilde{\mu}|$ sufficiently large such that $K \int_{[0,1]\cap[a_\delta, b_\delta]^c} P\left(\frac{W_s}{|\tilde{\mu}|} \in (\tilde{\tau} - \delta, \tilde{\tau} + \delta)\right) ds \leq \xi/2$. Moreover, we can take δ sufficiently small such that $1 + \tilde{\tau} - 2\delta > 0$ and $K(b_\delta - a_\delta) \leq \frac{\xi}{2}$. Hence $E\left[\int_0^1 H_\delta\left(\frac{W_s}{|\tilde{\mu}|}\right) ds\right] < \xi$. Since $\xi > 0$ is arbitrary, so the Markov's inequality implies that

$$\left\{\int_0^1 I\left(\frac{W_s}{|\tilde{\mu}|} \leq \tilde{\tau}\right) ds - \int_0^1 G_\delta\left(\frac{W_s}{|\tilde{\mu}|}\right) ds\right\} = o_p(1).$$

Now, we show that, for each $\eta > 0$

$$\left|\int_0^1 I\left(\mathcal{G}_s \leq \tilde{\tau}\right) ds - \int_0^1 G_\delta\left(\mathcal{G}_s\right) ds\right| < \eta. \tag{4.29}$$

It holds that

$$\left|\int_0^1 \left\{I\left(\mathcal{G}_s \leq \tilde{\tau}\right) - G_\delta\left(\mathcal{G}_s\right)\right\} ds\right|$$

$$\leq \int_0^1 \left|I\left(\mathcal{G}_s \leq \tilde{\tau}\right) - G_\delta\left(\mathcal{G}_s\right)\right| ds \leq \int_0^1 H_\delta\left(\mathcal{G}_s\right) ds.$$

Obviously, if $\mathcal{G}_s \in [\tilde{\tau} - 2\delta, \tilde{\tau} + 2\delta] \Rightarrow \mathcal{G}_s \in [\tilde{\tau} - \delta, \tilde{\tau} + \delta] \Rightarrow H_\delta(\mathcal{G}_s) \neq 0$. Set $A_\delta = \left[-\frac{1}{\tilde{\theta}} \log(\tilde{\tau} + 1 + 2\delta), -\frac{1}{\tilde{\theta}} \log(\tilde{\tau} + 1 - 2\delta)\right]$. Note that $\mathcal{G}_s \in [\tilde{\tau} - 2\delta, \tilde{\tau} + 2\delta]$ if and only if $s \in A_\delta$. Choose δ sufficient small such that $\tilde{\tau} + 1 - 2\delta > 0$ Therefore

$$
\begin{aligned}
&\int_0^1 H_\delta(\mathcal{G}_s)\, ds \leq \int_{[0,1]\cap A_\delta} H_\delta(\mathcal{G}_s) ds \leq \int_{A_\delta} H_\delta(\mathcal{G}_s) ds \\
&\leq K \int_{A_\delta} ds = K\left(-\frac{1}{\tilde{\theta}} \log(\tilde{\tau} + 1 - 2\delta) + \frac{1}{\tilde{\theta}} \log(\tilde{\tau} + 1 + 2\delta)\right).
\end{aligned}
$$

So, since we can choose $\delta > 0$ sufficiently small to make

$$
\left(-\frac{1}{\tilde{\theta}} \log(\tilde{\tau} + 1 - 2\delta) + \frac{1}{\tilde{\theta}} \log(\tilde{\tau} + 1 + 2\delta)\right)
$$

arbitrary small, (4.29) is proved. The proof of (ii) is completed upon applying the continuous mapping theorem that implies

$$
\int_0^1 G_\delta\left(\frac{W_s}{|\tilde{\mu}|}\right) ds \xrightarrow[|\tilde{\mu}|\to\infty]{p} \int_0^1 G_\delta(\mathcal{G}_s)\, ds.
$$

By using the same argument if follows that

$$
\begin{aligned}
\int_0^1 \frac{W_s}{|\tilde{\mu}|} I(W_s \leq \tau) ds &\xrightarrow[|\tilde{\mu}|\to\infty]{p} \int_0^1 \mathcal{G}_s I(\mathcal{G}_s \leq \tilde{\tau}) ds; \\
\int_0^1 \frac{W_s^2}{|\tilde{\mu}|^2} I(W_s \leq \tau) ds &\xrightarrow[|\tilde{\mu}|\to\infty]{p} \int_0^1 \mathcal{G}_s^2 I(\mathcal{G}_s \leq \tilde{\tau}).
\end{aligned}
$$

In fact,

$$
\begin{aligned}
\int_0^1 E\left[\frac{W_s}{|\tilde{\mu}|} H_\delta\left(\frac{W_s}{|\tilde{\mu}|}\right) ds\right] &\leq K \max\{|\tilde{\tau} - \delta|, |\tilde{\tau} + \delta|\} \int_0^1 P\left(\frac{W_s}{|\tilde{\mu}|} \in (\tilde{\tau} - \delta, \tilde{\tau} + \delta)\right) ds \\
\int_0^1 E\left[\left(\frac{W_s}{|\tilde{\mu}|}\right)^2 H_\delta\left(\frac{W_s}{|\tilde{\mu}|}\right) ds\right] &\leq K \max\left\{|\tilde{\tau} - \delta|^2, |\tilde{\tau} + \delta|^2\right\} \int_0^1 P\left(\frac{W_s}{|\tilde{\mu}|} \in (\tilde{\tau} - \delta, \tilde{\tau} + \delta)\right) ds.
\end{aligned}
$$

Proof of Proposition 66

The proof of the equalities in the first line readily follows upon noting that the indicators functions can not be always equal to zero since $\tilde{\tau}$ is selected in a interval based on the quantiles of the data. Now, we prove the equality in the second line. We have that

$$
\begin{aligned}
&\frac{1}{|\tilde{\mu}|}\left\{\int_0^1 I\left(W_s \le \tau\right) dW_s - \left[\int_0^1 I\left(W_s \le \tau\right) ds\right]\int_0^1 dW_s\right\}\\
&-\left[\int_0^1 I\left(W_s \le \tau\right) ds\right]\int_0^1 \left\{\tilde{\theta}(\tilde{\mu}-W_s)ds + d\tilde{W}_s\right\}\Bigg\}\\
&=\int_0^1 I\left(\frac{W_s}{|\tilde{\mu}|}\le\tilde{\tau}\right)\left\{-\tilde{\theta}\left(1+\frac{W_s}{|\tilde{\mu}|}\right)ds + d\frac{\tilde{W}_s}{|\tilde{\mu}|}\right\}\\
&-\left[\int_0^1 I\left(\frac{W_s}{|\tilde{\mu}|}\le\tilde{\tau}\right)ds\right]\int_0^1\left\{-\tilde{\theta}\left(1+\frac{W_s}{|\tilde{\mu}|}\right)ds + d\frac{\tilde{W}_s}{|\tilde{\mu}|}\right\}\\
&=-\,\tilde{\theta}\int_0^1\left(1+\frac{W_s}{|\tilde{\mu}|}\right)I\left(\frac{W_s}{|\tilde{\mu}|}\le\tilde{\tau}\right)ds + \int_0^1 I\left(\frac{W_s}{|\tilde{\mu}|}\le\tilde{\tau}\right)d\frac{\tilde{W}_s}{|\tilde{\mu}|}\\
&+\tilde{\theta}\left[\int_0^1 I\left(\frac{W_s}{|\tilde{\mu}|}\le\tilde{\tau}\right)ds\right]\int_0^1\left(1+\frac{W_s}{|\tilde{\mu}|}\right)ds - \left[I\left(\frac{W_s}{|\tilde{\mu}|}\le\tilde{\tau}\right)ds\right]\int_0^1 d\frac{\tilde{W}_s}{|\tilde{\mu}|}\\
&=\tilde{\theta}\left[\int_0^1 I\left(\frac{W_s}{|\tilde{\mu}|}\le\tilde{\tau}\right)ds\right]\int_0^1\left(1+\frac{W_s}{|\tilde{\mu}|}\right)ds - \tilde{\theta}\int_0^1\left(1+\frac{W_s}{|\tilde{\mu}|}\right)I\left(\frac{W_s}{|\tilde{\mu}|}\le\tilde{\tau}\right)ds + o_p(1)
\end{aligned}
$$

The last equality holds since

$$
\begin{aligned}
E\left[\int_0^1 I\left(\frac{W_s}{|\tilde{\mu}|}\le\tilde{\tau}\right)d\frac{\tilde{W}_s}{|\tilde{\mu}|}\right] &= E\left[\int_0^1 d\frac{\tilde{W}_s}{|\tilde{\mu}|}\right]=0;\\
\operatorname{Var}\left[\int_0^1 I\left(\frac{W_s}{|\tilde{\mu}|}\le\tilde{\tau}\right)d\frac{\tilde{W}_s}{|\tilde{\mu}|}\right] &= \frac{1}{\tilde{\mu}^2}E\left[\int_0^1 I\left(\frac{W_s}{|\tilde{\mu}|}\le\tilde{\tau}\right)ds\right]\xrightarrow[|\tilde{\mu}|\to\infty]{p} 0
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
&\tilde{\theta}\left[\int_0^1 I\left(\frac{W_s}{|\tilde{\mu}|}\le\tilde{\tau}\right)ds\right]\int_0^1\left(1+\frac{W_s}{|\tilde{\mu}|}\right)ds - \tilde{\theta}\int_0^1\left(1+\frac{W_s}{|\tilde{\mu}|}\right)I\left(\frac{W_s}{|\tilde{\mu}|}\le\tilde{\tau}\right)ds\\
\xrightarrow[|\tilde{\mu}|\to\infty]{p}&\tilde{\theta}\left\{\int_0^1 e^{-\tilde{\theta}s}ds\int_0^1 I\left(\mathcal{G}_s\le\tilde{\tau}\right)ds - \int_0^1 e^{-\tilde{\theta}s}I\left(\mathcal{G}_s\le\tilde{\tau}\right)ds\right\}
\end{aligned}
$$

Now we show that

$$\int_0^1 e^{-\tilde{\theta}s}ds\int_0^1 I\left(\mathcal{G}_s\le\tilde{\tau}\right)ds - \int_0^1 e^{-\tilde{\theta}s}I\left(\mathcal{G}_s\le\tilde{\tau}\right)ds > 0.$$

In fact, let $A=\{s\in[0,1]|\mathcal{G}_s\le\tilde{\tau}\}$ and $\ell(\cdot)$ the length of the enclosed interval. Note that, since $\mathcal{G}_s$ is a decreasing negative function over $[0,1]$, $A=[\bar{s},1]$ where $\bar{s}=\min_{s\in[0,1]}\{s|\mathcal{G}_s\le\tilde{\tau}\}$.

It holds that

$$
\begin{aligned}
&\tilde{\theta}\left\{\int_0^1 e^{-\tilde{\theta}s}ds\int_0^1 I\left(\mathcal{G}_s\leq\tilde{\tau}\right)ds-\int_0^1 e^{-\tilde{\theta}s}I\left(\mathcal{G}_s\leq\tilde{\tau}\right)ds\right\}\\
=&\left(\int_A e^{-\tilde{\theta}s}ds+\int_{A^c}e^{-\tilde{\theta}s}ds\right)\ell(A)-\int_A e^{-\tilde{\theta}s}ds\\
=&\ell(A)\int_{A^c}e^{-\tilde{\theta}s}ds-\{1-\ell(A)\}\int_A e^{-\tilde{\theta}s}ds.
\end{aligned}
$$

Since $e^{-\tilde{\theta}s}$ is a decreasing positive function on [0,1], there exists a positive real value γ such that

$$
\begin{aligned}
&\ell(A)\int_{A^c}e^{-\tilde{\theta}s}ds-\{1-\ell(A)\}\int_A e^{-\tilde{\theta}s}ds\\
>&\gamma\ell(A)\ell(A^c)-\{1-\ell(A)\}\gamma\ell(A)=0.
\end{aligned}
$$

Therefore the first claim is proved. In the same manner, it is possible to verify the last equality. We have:

$$
\begin{aligned}
&\frac{1}{\tilde{\mu}^2}\left\{\int_0^1 W_sI\left(W_s\leq\tau\right)dW_s-\left[\int_0^1 W_sI\left(W_s\leq\tau\right)ds\right]\int_0^1 dW_s\right\}\\
\xrightarrow[|\tilde{\mu}|\to\infty]{p}&\tilde{\theta}\left\{\int_0^1\mathcal{G}_sI\left(\mathcal{G}_s\leq\tilde{\tau}\right)ds\int_0^1\left(1+\mathcal{G}_s\right)ds-\int_0^1\mathcal{G}_s\left(1+\mathcal{G}_s\right)I\left(\mathcal{G}_s\leq\tilde{\tau}\right)ds\right\}
\end{aligned}
$$

Below, we show that

$$
\int_0^1\mathcal{G}_sI\left(\mathcal{G}_s\leq\tilde{\tau}\right)ds\int_0^1\left(1+\mathcal{G}_s\right)ds-\int_0^1\mathcal{G}_s\left(1+\mathcal{G}_s\right)I\left(\mathcal{G}_s\leq\tilde{\tau}\right)ds\neq 0.
$$

Let $A=\{s\in[0,1]|\mathcal{G}_s\leq\tilde{\tau}\}=[a,1]$, where $0<a<1$. Therefore we have to prove that

$$
\int_A(e^{-\theta s}-1)ds\int_0^1 e^{-\theta s}ds-\int_A e^{-\theta s}(e^{-\theta s}-1)ds\neq 0
$$

that is equivalent to show that

$$
\int_A(1-e^{-\theta s})ds\int_0^1 e^{-\theta s}ds-\int_A e^{-\theta s}(1-e^{-\theta s})ds\neq 0 \tag{4.30}
$$

The mean value theorem for definite integrals implies that there exists $b\in A$ such that

$$
\int_A e^{-\theta s}(1-e^{-\theta s})ds=e^{-\theta b}\int_A(1-e^{-\theta s})ds.
$$

Therefore

$$\begin{aligned}
&\int_A (1-e^{-\theta s})ds \int_0^1 e^{-\theta s}ds - \int_A e^{-\theta s}(1-e^{-\theta s})ds \\
=&\int_A (1-e^{-\theta s})ds \int_0^1 e^{-\theta s}ds - e^{-\theta b}\int_A (1-e^{-\theta s})ds \\
=&\int_A (1-e^{-\theta s})ds \left\{\int_0^1 e^{-\theta s}ds - e^{-\theta b}\right\}
\end{aligned}$$

Hence, (4.30) is verified if we show that

$$\int_0^1 e^{-\theta s}ds - e^{-\theta b} \neq 0.$$

We prove that

$$\int_0^1 e^{-\theta s}ds - e^{-\theta b} > 0. \tag{4.31}$$

In fact, it holds that

$$\int_0^1 e^{-\theta s}ds - e^{-\theta b} = \frac{1-e^{-\theta}}{\theta} - e^{-b\theta} = \frac{1-e^{-\theta}-\theta e^{-\theta b}}{\theta},$$

that is positive if and only if

$$\begin{aligned}
&1-e^{-\theta}-\theta e^{-\theta b} > 0 \Leftrightarrow -\theta e^{-\theta b} > e^{-\theta}-1 \Leftrightarrow \theta e^{-\theta b} < 1-e^{-\theta} \\
\Leftrightarrow& e^{-\theta b} < \frac{1-e^{-\theta}}{\theta} \Leftrightarrow -\theta b < \ln\left(\frac{1-e^{-\theta}}{\theta}\right) \Leftrightarrow b > \frac{1}{\theta}\ln\left(\frac{1-e^{-\theta}}{\theta}\right).
\end{aligned}$$

Hence, Condition 4.31 is verified we show that the function $z(\theta) = \frac{1}{\theta}\ln\left(\frac{1-e^{-\theta}}{\theta}\right)$ is negative for each $\theta > 0$. But this is readily checked because

$$\begin{aligned}
&\lim_{\theta\to 0^+} z(\theta) = 0, \\
&\frac{\partial z(\theta)}{\partial\theta} = e^{-\theta} - 1 < 0, \quad \text{for each } \theta > 0.
\end{aligned}$$

Finally, we prove the invertibility of $\tilde{\Delta}(\tau)$:

$$\begin{pmatrix} \frac{c-b^2}{(1-a)(ac-b^2)} & -\frac{b}{ac-b^2} \\ -\frac{b}{ac-b^2} & \frac{a}{ac-b^2} \end{pmatrix}$$

Its determinant results

$$\begin{aligned}
&\frac{c-b^2}{(1-a)(ac-b^2)} \times \frac{a}{ac-b^2} - \frac{b^2}{(ac-b^2)^2} \\
=&\frac{a(c-b^2)-b^2(1-a)}{(1-a)(ac-b^2)^2} = \frac{1}{(1-a)(ac-b^2)}.
\end{aligned}$$

Note that $1-a$ is positive since $\tilde{\tau}$ is selected in a interval based on the quantiles of the data. It suffices to show that $c-b^2>0$ and $ac-b^2>0$. But this readily follows upon using

$$c-b^2>0 \text{ and } ac-b^2>0.$$

the Cauchy-Schwarz inequality, i.e., for each $f,g\in\mathcal{L}^2$

$$\left|\int_0^1 f(s)g(s)ds\right|^2 \le \int_0^1 |f(s)|^2\,ds \times \int_0^1 |g(s)|^2\,ds$$

For the fist case take $f(s)=\mathcal{G}_s I\left(\mathcal{G}_s\le\tilde{\tau}\right)$ and $g(s)=I\left(\mathcal{G}_s\le\tilde{\tau}\right)$, while for the second case $f(s)=\mathcal{G}_s I\left(\mathcal{G}_s\le\tilde{\tau}\right)$ and $g(s)=1$.

Appendix

Appendix A

In this appendix we derive the distribution, used in several proofs, of:

$$\frac{X_s}{\sqrt{n}\sigma(1-\theta)};\ \left(\frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)},\varepsilon_{s_2},\frac{X_{s_3}}{\sigma\sqrt{n}(1-\theta)}\right)^{\intercal};\ \varepsilon_{s_2}\left|\frac{X_{s_1}}{\sqrt{n}\sigma(1-\theta)},\frac{X_{s_3}}{\sqrt{n}\sigma(1-\theta)}\right.,$$

where $s,s_1,s_2,s_3=1,\ldots,n$ and $s_1<s_2<s_3$.

Proposition 68. *Let $X_s=X_{s-1}+\varepsilon_s-\theta\varepsilon_{s-1}$, with $\varepsilon_s\sim N(0,\sigma^2)$ and assume $X_1=0$. Then it holds that*

$$\frac{X_s}{\sqrt{n}\sigma(1-\theta)}\sim N\left(0,\frac{(1-\theta)^2(s-1)+1+\theta^2}{n(1-\theta)^2}\right). \tag{4.32}$$

Proof. We have that:

$$\begin{aligned}
X_1 &= \varepsilon_1-\theta\varepsilon_0;\\
X_s &= \varepsilon_s+(1-\theta)\sum_{j=1}^{s-1}\varepsilon_j-\theta\varepsilon_0, \qquad \text{for each } s\ge 2;\\
X_{s_2} &= X_{s_2-1}+\varepsilon_{s_2}-\theta\varepsilon_{s_2-1};\\
X_{s_2} &= X_{s_1}+(1-\theta)\sum_{j=s_1+1}^{s_2-1}\varepsilon_j+\varepsilon_{s_2}-\theta\varepsilon_{s_1} \qquad \text{for each } s_1<s_2-1.
\end{aligned}$$

$$\begin{aligned}
V\left(X_s\right) &= V\left(\varepsilon_s+(1-\theta)\sum_{j=1}^{s-1}\varepsilon_j-\theta\varepsilon_0\right)\\
&= V(\varepsilon_s)+(1-\theta)^2\sum_{j=1}^{s-1}V(\varepsilon_j)+\theta^2V(\varepsilon_0)\\
&= \left\{(1-\theta)^2(s-1)+1+\theta^2\right\}\sigma^2
\end{aligned}$$

and this completes the proof. □

Proposition 69. *Let* $X_s = X_{s-1} + \varepsilon_s - \theta\varepsilon_{s-1}$*, with* $\varepsilon_s \sim N(0, \sigma^2)$ *and assume* $X_1 = 0$*. Given* $s_1 < s_2 \leq s_3$*, consider the three-dimension random vector*

$$\left(\frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)}, \varepsilon_{s_2}, \frac{X_{s_3}}{\sigma\sqrt{n}(1-\theta)} \right)^{\intercal}.$$

It has a three-dimensional normal distribution with zero mean vector and variance-covariance matrix:

$$\begin{pmatrix} \frac{s_1(1-\theta)^2+2\theta}{n(1-\theta)^2} & 0 & \frac{s_1(1-\theta)^2+\theta}{n(1-\theta)^2} \\ 0 & \sigma^2 & \frac{\sigma}{\sqrt{n}} \\ \frac{s_1(1-\theta)^2+\theta}{n(1-\theta)^2} & \frac{\sigma}{\sqrt{n}} & \frac{s_3(1-\theta)^2+2\theta}{n(1-\theta)^2} \end{pmatrix}$$

Proof. We start computing the mean:

$$E\left[\begin{pmatrix} \frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)} \\ \varepsilon_{s_2} \\ \frac{X_{s_3}}{\sigma\sqrt{n}(1-\theta)} \end{pmatrix} \right] = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The variance-covariance matrix is

$$\begin{pmatrix} V\left(\frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)}\right) & \text{Cov}\left(\frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)}, \varepsilon_{s_2}\right) & \text{Cov}\left(\frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)}, \frac{X_{s_3}}{\sigma\sqrt{n}(1-\theta)}\right) \\ \text{Cov}\left(\frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)}, \varepsilon_{s_2}\right) & V\left(\varepsilon_{s_2}\right) & \text{Cov}\left(\frac{X_{s_3}}{\sigma\sqrt{n}(1-\theta)}, \varepsilon_{s_2}\right) \\ \text{Cov}\left(\frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)}, \frac{X_{s_3}}{\sigma\sqrt{n}(1-\theta)}\right) & \text{Cov}\left(\frac{X_{s_3}}{\sigma\sqrt{n}(1-\theta)}, \varepsilon_{s_2}\right) & V\left(\frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)}\right) \end{pmatrix}$$

Here we compute the $\text{Cov}\left(\frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)}, \frac{X_{s_3}}{\sigma\sqrt{n}(1-\theta)}\right)$, where $s_2 < s_3$

$$\begin{aligned} \text{Cov}\left(\frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)}, \frac{X_{s_3}}{\sigma\sqrt{n}(1-\theta)}\right) &= \frac{1}{\sigma^2(1-\theta)^2 n}\text{Cov}\left(X_{s_1}, X_{s_1} + (1-\theta)\sum_{j=s_1+1}^{s_3-1}\varepsilon_j + \varepsilon_{s_3} - \theta\varepsilon_{s_1}\right) \\ &= \frac{1}{\sigma^2(1-\theta)^2 n}\left\{\text{Cov}\left(X_{s_1}, X_{s_1}\right) - \theta\text{Cov}\left(X_{s_1}, \varepsilon_{s_1}\right)\right\} \\ &= \frac{1}{\sigma^2(1-\theta)^2 n}\left[\sigma^2\left\{(1-\theta)^2(s_1-1) + 1 + \theta^2\right\} - \theta\sigma^2\right] \\ &= \frac{(1-\theta)^2(s_1-1) + 1 + \theta^2 - \theta}{(1-\theta)^2 n} \\ &= \frac{(1-\theta)^2(s_1-1) + 1 + \theta^2 - 2\theta + \theta}{(1-\theta)^2 n} \\ &= \frac{(1-\theta)^2(s_1-1) + (1-\theta)^2 + \theta}{(1-\theta)^2 n} \\ &= \frac{(1-\theta)^2(s_1-1+1)}{(1-\theta)^2 n} + \frac{\theta}{(1-\theta)^2 n} \\ &= \frac{s_1}{n} + \frac{\theta}{(1-\theta)^2 n}. \end{aligned}$$

Therefore the variance-covariance matrix results

$$\begin{pmatrix} \frac{s_1}{n}+\frac{2\theta}{n(1-\theta)^2} & 0 & \frac{s_1}{n}+\frac{\theta}{n(1-\theta)^2} \\ 0 & \sigma^2 & \frac{\sigma}{\sqrt{n}} \\ \frac{s_1}{n}+\frac{\theta}{n(1-\theta)^2} & \frac{\sigma}{\sqrt{n}} & \frac{s_3}{n}+\frac{2\theta}{n(1-\theta)^2} \end{pmatrix} = \begin{pmatrix} \frac{s_1(1-\theta)^2+2\theta}{n(1-\theta)^2} & 0 & \frac{s_1(1-\theta)^2+\theta}{n(1-\theta)^2} \\ 0 & \sigma^2 & \frac{\sigma}{\sqrt{n}} \\ \frac{s_1(1-\theta)^2+\theta}{n(1-\theta)^2} & \frac{\sigma}{\sqrt{n}} & \frac{s_3(1-\theta)^2+2\theta}{n(1-\theta)^2} \end{pmatrix}$$

Note that the case with $s_2 = s_3$ can be done in the identic manner because the variance and covariance matrix changes just for a constant. In fact, when $s_2 = s_3$, we have that

$$\begin{pmatrix} \frac{s_1}{n}+\frac{2\theta}{n(1-\theta)^2} & 0 & \frac{s_1}{n}+\frac{\theta}{n(1-\theta)^2} \\ 0 & \sigma^2 & \frac{\sigma}{(1-\theta)\sqrt{n}} \\ \frac{s_1}{n}+\frac{\theta}{n(1-\theta)^2} & \frac{\sigma}{(1-\theta)\sqrt{n}} & \frac{s_3}{n}+\frac{2\theta}{n(1-\theta)^2} \end{pmatrix}.$$

Hence, the statement is completely verified. □

Proposition 70. *Let $X_s = X_{s-1} + \varepsilon_s - \theta\varepsilon_{s-1}$, with $\varepsilon_s \sim N(0, \sigma^2)$ and assume $X_1 = 0$. Then it holds that*

$$\varepsilon_{s_2} \left| \frac{X_{s_3}}{\sigma\sqrt{n}(1-\theta)}, \frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)} \right.$$

is a univariate normal distributed random variable with mean equal

$$\frac{\sqrt{n}\sigma(1-\theta)^2}{s_1(s_3-s_1)(1-\theta)^4+2\theta s_3(1-\theta)^2+3\theta^2}$$
$$\times \left[\frac{X_{s_3}}{\sigma\sqrt{n}(1-\theta)}\left\{s_1(1-\theta)^2+2\theta\right\} - \frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)}\left\{s_1(1-\theta)^2+\theta\right\}\right]$$

and variance equal

$$\sigma^2 - \sigma^2\frac{(1-\theta)^4 s_1 + 2(1-\theta)^2\theta}{s_1(s_3-s_1)(1-\theta)^4+2\theta s_3(1-\theta)^2+3\theta^2}.$$

Proof. We will use the following result concerning the k-dimensional normal random vector: Let $X \in \mathbb{R}^k$ has a multivariate normal distribution, whose mean is μ_X and the variance-covariance matrix Σ. Consider

$$X = (X_1, X_2)^\intercal,\ X_1 \in \mathbb{R}^{k_1},\ X_2 \in \mathbb{R}^{k_2} \text{ and } k_1 + k_2 = k.$$
$$\mu_X = \begin{pmatrix} \mu_{X_1} \\ \mu_{X_2} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{pmatrix}.$$

Then $X_2|X_1$ is distributed as a k_2-dimensional normal random vector with

$$E[X_2|X_1] = \Sigma_{2,1}\Sigma_{1,1}^{-1}(x_1 - \mu_{X_1}) + \mu_{X_2}$$
$$\Sigma_{X_2|X_1} = \Sigma_{2,2} - \Sigma_{2,1}\Sigma_{1,1}^{-1}\Sigma_{1,2}.$$

This result implies that $\varepsilon_{s_2}\left|\frac{X_{s_3}}{\sigma\sqrt{n}(1-\theta)},\frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)}\right.$ is a univariate normal distributed random variable with

$$\begin{aligned}
& E\left[\varepsilon_{s_2}\left|\frac{X_{s_3}}{\sigma\sqrt{n}(1-\theta)},\frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)}\right.\right] \\
=& \frac{n(1-\theta)^2}{s_1(s_3-s_1)(1-\theta)^4+2\theta s_3(1-\theta)^2+3\theta^2} \\
\times& \left[\frac{X_{s_3}}{\sigma\sqrt{n}(1-\theta)}\frac{\left\{s_1(1-\theta)^2+2\theta\right\}\sigma}{\sqrt{n}}-\frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)}\frac{\left\{s_1(1-\theta)^2+\theta\right\}\sigma}{\sqrt{n}}\right] \\
=& \frac{\sqrt{n}\sigma(1-\theta)^2}{s_1(s_3-s_1)(1-\theta)^4+2\theta s_3(1-\theta)^2+3\theta^2} \\
\times& \left[\frac{X_{s_3}}{\sigma\sqrt{n}(1-\theta)}\left\{s_1(1-\theta)^2+2\theta\right\}-\frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)}\left\{s_1(1-\theta)^2+\theta\right\}\right]
\end{aligned}$$

$$\begin{aligned}
& V\left[\varepsilon_{s_2}\left|\frac{X_{s_3}}{\sigma\sqrt{n}(1-\theta)},\frac{X_{s_1}}{\sigma\sqrt{n}(1-\theta)}\right.\right] \\
=& \sigma^2-\sigma^2\frac{(1-\theta)^4 s_1+2(1-\theta)^2\theta}{s_1(s_3-s_1)(1-\theta)^4+2\theta s_3(1-\theta)^2+3\theta^2}
\end{aligned}$$

and, therefore, the proposition is proved. □

Appendix B

Here we report some integrals used in the proofs.

1. $\int_0^1\int_0^1\frac{1}{\sqrt{xy-\min^2(x,y)}}dxdy=2\pi$. In fact:

$$\begin{aligned}
& \int_0^1\int_0^1\frac{1}{\sqrt{xy-\min^2(x,y)}}dxdy \\
=& 2\int_0^1\int_x^1\frac{1}{\sqrt{xy-x^2}}dydx \\
=& 2\int_0^1\frac{1}{\sqrt{x}}\int_x^1\frac{1}{\sqrt{y-x}}dydx \\
=& 4\int_0^1\frac{\sqrt{1-x}}{\sqrt{x}}dx \\
=& 8\int_0^1\sqrt{1-t^2}dt \\
=& 8\left[\frac{1}{2}\arcsin t+t\sqrt{1-t^2}\right]_{t=0}^{t=1} \\
=& 2\pi.
\end{aligned}$$

2. $\int_0^1 \int_0^y \frac{x}{y\sqrt{x(y-x)}} dxdy = \frac{\pi}{2}$. In fact:

$$\begin{aligned}
&\text{we do the change of variables } t = \sqrt{y-x} \text{ that implies}\\
&t^2 = y - x \Rightarrow x = y - t^2\\
\Rightarrow & dx = -2tdt \Rightarrow dt = \frac{-dx}{2t} = \frac{-dx}{2\sqrt{y-x}}\\
&\text{when } x = y \text{ we have } t = 0; \text{ when } x = 0 \text{ we have } t = \sqrt{y}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\int_0^1 \int_0^y \frac{x}{y\sqrt{x(y-x)}} dxdy &= \int_0^1 \frac{1}{y} \int_0^y (-2) \frac{\sqrt{x}}{(-2)\sqrt{y-x}} dxdy\\
&= \int_0^1 \frac{-2}{y} \int_{\sqrt{y}}^0 \sqrt{y-t^2} dtdy\\
&= \int_0^1 \frac{2}{y} \int_0^{\sqrt{y}} \sqrt{y-t^2} dtdy\\
&= \int_0^1 \frac{2}{\sqrt{y}} \int_0^{\sqrt{y}} \sqrt{\frac{y-t^2}{y}} dtdy\\
&= \int_0^1 \frac{2}{\sqrt{y}} \int_0^{\sqrt{y}} \sqrt{1 - \frac{t^2}{y}} dtdy.
\end{aligned}$$

Now, we do another change of variables:

$w = \frac{t}{\sqrt{y}} \Rightarrow dw = \frac{1}{\sqrt{y}} dt$ and $w^2 = \frac{t^2}{y}$

when $t = 0$, we have $w = 0$ and when $t = \sqrt{y}$ we have $t = 1$. Therefore:

$$\begin{aligned}
&\int_0^1 \frac{2}{\sqrt{y}} \int_0^{\sqrt{y}} \sqrt{1 - \frac{t^2}{y}} dtdy\\
=& \int_0^1 2 \int_0^1 \sqrt{1-w^2} dwdt\\
=& 2\int_0^1 dy \int_0^1 \sqrt{1-w^2} dwdt\\
=& 2\int_0^1 \sqrt{1-w^2} dwdt.
\end{aligned}$$

From the previous computation we know that $\int_0^1 \sqrt{1-w^2} dw = \frac{\pi}{4}$, therefore

$$\int_0^1 \int_0^y \frac{x}{y\sqrt{x(y-x)}} = \frac{\pi}{2}.$$

Chapter 5

On the parsimony of TARMA models

The well known Wold decomposition states that any stationary process admits a MA(∞) representation (for more details see Brockwell and Davis [2001]). This implies the duality between MA and AR processes so that, for instance, an MA(1) admits a AR(∞) representation and viceversa. On the operational strand, this leads to the fact that ARMA(1,1) processes can describe and fit well data coming from AR(p) processes with p large. Interestingly, Bickel and Bühlmann [1997] proved that the class of linear processes, that admit an AR(∞) representation can approximate virtually everything provided the autoregressive order is sufficiently large, and this is at the base of the so called sieve bootstrap. The ARMA(1,1) process is notable since it retains the approximating capability of AR(∞) process with just two parameters. In the same spirit, we investigate whether a similar representation holds in non-linear case. In particular, we focus on the possible duality between the TARMA(1,1) model and TAR(p) models with large p. The aim of this chapter is twofold. First, we show that if we generate data from several TARMA(1,1) processes and fit TAR(p) models with p chosen using information criteria then, in many cases, the selected order can be very large. This indicates that TARMA(1,1) models may exhibit complex high-dimensional dynamic behaviour with parsimony. Second, we assess the descriptive power of TARMA(1,1) models by generating data from a plethora of data generating processes and fitting both TARMA(1,1) and AR(p) models with p possibly large. The results show that, in presence of MA components, the TARMA model outperforms AR models even for linear processes. This is especially true for integrated processes and for nonlinear processes. In Section 5.3 we analyse the Canadian lynx time series by using TARMA models. Even if the preliminary results point at TAR models as having the best fit, the TARMA(1,1) model allows to shed additional light on the lynx data. Further investigations are definitely required.

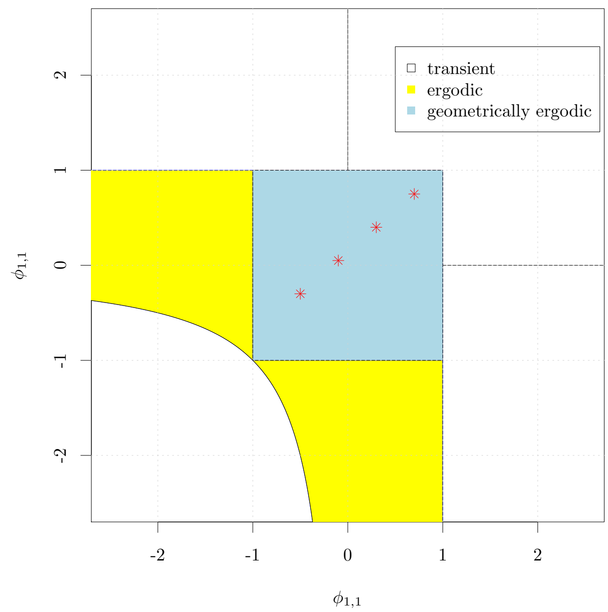

Figure 5.1: Ergodicity region

5.1 Order selection

We have generated $B = 10000$ time series of length $n = 100, 200$ from several TARMA models with different parameters. The TARMA(1,1,2) we consider is

$$X_t = \begin{cases} \phi_{1,0} + \phi_{1,1}X_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}, & \text{if } X_{t-d} \leq r \\ \phi_{2,0} + \phi_{2,1}X_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}, & \text{otherwise,} \end{cases} \tag{5.1}$$

with the parameters chosen as follows:

$$(\phi_{1,0}, \phi_{1,1}, \phi_{2,0}, \phi_{2,1}) = (0, 0.3, 0, 0.4) \times t + (0, -0.5, 0, -0.3) \times (1 - t),$$
$$\text{with } t = 0, 0.5, 1, 1.5 \text{ and } \theta = -0.9, -0.65, -0.5, 0.5, 0.65, 0.9.$$

The parameters belong to the region where the process is geometrically ergodic. They are indicated with red stars in Figure 5.1. We fitted a two-regime TAR over each time series by selecting the order that minimizes the AIC. Let p_1 and p_2 be the orders selected in the first and second regime, respectively. The maximum value they can assume is 15 and we set $p = \max\{p_1, p_2\}$. Table 5.1 and 5.2 show the frequency distribution (percentage) of p for each choice of the parameters for $n = 100$ and $n = 200$, respectively. Analogously, we report in Table 5.3 and Table 5.4 (Table 5.5 and Table 5.6) the same distributions for the selected order in the first (second) regime. Each row corresponds to a distribution for a set of parameters. In agreement with the linear case, when the processes are close to the region of non-invertibility (i.e. $|\theta|$ close to 1) then the selected order tends to be very high. In Table 5.7 we aggregate the distributions of Tables 5.1-5.2 according to the selected order $p \leq 5$ or $p > 5$. Clearly, even when $\theta = 0$, i.e. the DGP is a TAR(1,1), the selected order is greater than 5 most of the times. This raises the question of the suitability of the AIC for model selection in the non-linear setting.

$n = 100$	p														
t ; θ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.0;-0.9	0	5	6	9	7	8	6	7	5	6	5	6	7	9	13
0.5;-0.9	0	1	2	7	6	8	7	7	6	7	6	8	8	11	15
1.0;-0.9	0	0	1	5	5	9	7	8	6	7	6	8	9	12	17
1.5;-0.9	0	0	1	4	5	9	7	9	6	7	7	8	9	12	17
0.0;-0.65	8	16	9	8	5	5	4	4	4	4	4	5	5	8	12
0.5;-0.65	1	12	11	11	7	6	5	5	4	4	4	5	5	7	12
1.0;-0.65	0	5	10	14	8	7	6	5	5	4	5	5	6	8	13
1.5;-0.65	0	3	10	14	8	8	5	6	5	4	5	5	6	8	12
0.0;-0.5	15	12	7	6	5	5	4	4	4	4	4	5	5	8	12
0.5;-0.5	6	19	9	8	5	4	4	4	4	4	4	4	6	7	12
1.0;-0.5	1	16	12	10	6	5	5	4	4	4	5	4	5	7	12
1.5;-0.5	0	12	13	12	6	6	4	4	4	4	4	5	6	7	12
0.0;0.0	21	10	8	6	5	4	4	4	3	4	3	5	5	7	12
0.5;0.0	16	10	8	6	5	4	4	4	4	4	4	5	5	7	13
1.0;0.0	18	10	7	6	5	5	4	4	3	3	4	4	6	8	12
1.5;0.0	25	11	7	5	4	4	4	3	3	3	4	4	5	6	11
0.0;0.5	1	13	13	10	7	5	4	5	4	4	4	5	6	7	13
0.5;0.5	6	15	12	8	6	5	4	4	4	4	3	4	5	7	12
1.0;0.5	13	11	8	7	5	5	4	4	4	4	4	5	6	8	13
1.5;0.5	13	14	8	7	5	5	4	4	3	4	4	4	6	7	12
0.0;0.65	0	4	11	13	8	7	6	5	4	5	5	5	6	8	13
0.5;0.65	1	8	12	11	8	6	5	5	4	4	4	5	5	8	12
1.0;0.65	7	12	11	8	7	6	4	4	4	4	4	5	5	8	12
1.5;0.65	16	10	7	6	5	4	4	4	4	4	4	5	6	8	13
0.0;0.9	0	0	2	6	8	9	7	8	6	7	6	6	8	10	16
0.5;0.9	0	0	3	6	7	8	7	7	7	6	6	7	8	10	16
1.0;0.9	0	3	6	8	7	8	6	6	6	6	6	6	7	10	16
1.5;0.9	8	8	8	7	5	6	5	5	4	5	5	5	6	9	14

Table 5.1: Distribution of $p = \max_{i\in\{1,2\}} p_i$ when $n = 100$

$n = 200$	p														
t ; θ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.0;-0.9	0	0	2	5	6	9	8	9	7	9	7	8	7	9	13
0.5;-0.9	0	0	0	1	3	7	6	10	8	10	8	10	9	12	16
1.0;-0.9	0	0	0	0	1	4	6	9	8	11	9	10	10	14	17
1.5;-0.9	0	0	0	0	1	4	5	9	8	11	9	11	10	14	19
0.0;-0.65	3	15	12	11	7	6	5	4	4	4	4	4	4	6	11
0.5;-0.65	0	4	10	15	9	9	6	6	4	5	4	5	5	6	11
1.0;-0.65	0	1	6	14	11	10	7	6	5	5	5	5	6	7	12
1.5;-0.65	0	0	3	13	11	11	7	7	5	6	5	6	6	7	11
0.0;-0.5	14	13	8	7	6	5	4	4	4	4	4	5	5	6	11
0.5;-0.5	1	17	13	10	7	5	4	5	4	4	4	4	5	6	10
1.0;-0.5	0	8	14	13	8	7	5	5	4	4	4	4	5	7	10
1.5;-0.5	0	4	14	16	8	8	5	5	5	4	4	4	6	6	11
0.0;0.0	24	11	8	6	5	5	4	4	4	3	4	4	5	6	10
0.5;0.0	16	10	8	7	5	5	4	4	4	4	4	4	5	7	12
1.0;0.0	21	11	8	6	5	5	4	4	3	4	4	5	4	6	10
1.5;0.0	22	11	7	6	6	4	4	4	4	4	4	4	5	6	10
0.0;0.5	0	6	16	13	9	7	5	5	4	4	4	5	5	7	10
0.5;0.5	1	13	15	11	7	6	5	4	4	4	4	4	5	6	11
1.0;0.5	10	13	10	8	6	5	4	4	4	4	4	4	5	7	12
1.5;0.5	7	18	12	8	6	5	4	4	4	4	4	4	5	6	11
0.0;0.65	0	1	6	14	12	10	8	6	5	5	4	5	6	7	11
0.5;0.65	0	2	11	14	10	9	6	6	5	5	4	5	5	7	10
1.0;0.65	2	11	12	11	8	6	5	5	4	4	4	4	5	7	10
1.5;0.65	14	12	9	6	6	4	4	4	4	4	4	5	5	7	11
0.0;0.9	0	0	0	1	4	8	10	11	9	9	8	8	9	10	14
0.5;0.9	0	0	0	1	4	6	8	9	9	9	8	9	9	11	17
1.0;0.9	0	0	1	3	5	7	8	9	8	8	8	8	9	10	15
1.5;0.9	3	6	8	9	7	8	7	6	5	6	5	5	6	8	12

Table 5.2: Distribution of $p = \max_{i \in \{1,2\}} p_i$ when $n = 200$

$n = 100$	p_1														
t ; θ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.0;-0.9	7	12	10	10	7	8	5	5	4	4	3	4	4	6	9
0.5;-0.9	3	10	10	12	8	8	6	6	4	4	4	5	5	6	8
1.0;-0.9	1	8	9	12	9	10	6	6	4	5	4	5	5	7	9
1.5;-0.9	0	5	8	13	9	10	7	7	5	5	4	5	5	7	9
0.0;-0.65	15	18	9	8	5	4	4	4	4	3	3	4	4	6	10
0.5;-0.65	10	22	13	11	6	5	4	3	3	3	3	3	3	5	8
1.0;-0.65	5	23	15	13	7	6	4	3	3	3	3	3	3	5	7
1.5;-0.65	2	20	17	15	7	6	4	3	3	3	3	3	3	4	7
0.0;-0.5	20	14	8	6	5	4	4	4	3	3	4	4	4	6	10
0.5;-0.5	18	23	9	7	5	4	3	3	3	2	3	3	4	5	8
1.0;-0.5	14	28	12	8	5	4	3	2	2	3	3	3	3	4	6
1.5;-0.5	6	31	14	10	5	4	3	3	2	2	2	3	3	4	7
0.0;0.0	41	10	7	5	4	3	3	3	2	2	2	3	3	5	7
0.5;0.0	22	12	8	6	5	4	3	4	3	4	3	4	5	6	11
1.0;0.0	31	11	7	5	5	4	3	3	3	3	3	3	4	6	9
1.5;0.0	48	11	6	4	3	3	2	2	2	2	2	2	3	4	6
0.0;0.5	16	23	14	9	5	4	3	3	2	2	2	3	3	4	7
0.5;0.5	23	20	11	7	5	4	3	3	3	3	2	3	3	4	7
1.0;0.5	23	13	8	6	5	4	3	3	3	3	3	4	5	6	10
1.5;0.5	24	15	8	6	4	4	4	3	3	3	3	3	5	5	8
0.0;0.65	7	17	17	12	7	6	4	3	3	3	3	3	4	4	7
0.5;0.65	12	18	15	10	6	5	4	4	3	3	3	3	3	5	7
1.0;0.65	20	15	11	8	6	4	3	3	3	3	3	3	4	5	9
1.5;0.65	22	12	8	6	5	4	3	4	3	3	4	3	5	7	11
0.0;0.9	2	8	11	13	10	9	6	6	4	5	4	4	4	6	9
0.5;0.9	3	9	10	11	9	8	6	5	5	5	4	5	5	6	9
1.0;0.9	8	12	11	10	7	6	5	5	4	4	4	4	4	6	9
1.5;0.9	15	12	9	8	5	5	4	4	3	4	4	4	4	7	11

Table 5.3: Distribution of p_1 when $n = 100$

$n = 200$	p_1														
t ; θ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.0;-0.9	3	6	8	10	9	10	7	8	5	6	4	5	4	6	8
0.5;-0.9	1	4	5	8	8	11	8	9	7	7	5	6	5	7	9
1.0;-0.9	1	3	3	7	7	10	9	10	8	8	6	7	6	7	9
1.5;-0.9	0	2	3	6	6	10	9	10	8	8	6	7	6	8	10
0.0;-0.65	14	20	13	10	6	5	4	3	3	3	2	3	3	4	7
0.5;-0.65	5	16	16	16	8	6	4	4	3	3	3	3	3	3	6
1.0;-0.65	2	11	16	18	10	8	5	4	3	3	3	3	3	4	6
1.5;-0.65	1	8	14	19	12	10	5	5	3	3	3	3	3	4	6
0.0;-0.5	21	15	8	7	5	5	3	3	4	3	3	3	4	5	10
0.5;-0.5	13	26	14	10	5	4	3	3	2	2	2	3	3	4	6
1.0;-0.5	5	24	18	12	6	5	3	3	3	2	2	2	3	4	5
1.5;-0.5	2	21	22	15	6	5	4	3	3	2	2	2	3	3	6
0.0;0.0	45	11	6	5	4	3	3	2	2	2	2	2	3	3	6
0.5;0.0	25	12	8	7	5	4	4	4	3	3	3	3	4	5	10
1.0;0.0	40	11	7	5	4	4	3	2	2	3	3	3	3	4	6
1.5;0.0	47	10	6	4	4	3	3	2	2	2	2	2	2	4	6
0.0;0.5	7	20	21	12	7	5	3	3	2	3	2	3	3	4	5
0.5;0.5	14	23	16	9	5	4	3	3	2	3	2	3	3	4	6
1.0;0.5	25	17	10	7	5	4	3	3	3	3	3	3	3	5	8
1.5;0.5	20	23	12	7	5	4	3	3	2	3	2	2	3	4	7
0.0;0.65	3	9	17	18	12	8	5	4	3	3	3	3	3	4	6
0.5;0.65	5	13	18	16	10	7	4	4	3	3	2	3	3	4	6
1.0;0.65	14	19	14	11	7	5	4	3	3	3	3	3	3	4	6
1.5;0.65	21	15	9	7	5	4	4	4	3	3	3	3	4	6	9
0.0;0.9	1	2	4	8	10	11	10	10	7	7	5	5	5	6	8
0.5;0.9	2	3	5	8	9	9	9	9	8	7	5	6	6	7	9
1.0;0.9	3	6	6	9	9	10	9	8	6	6	5	5	6	6	8
1.5;0.9	11	12	10	9	8	7	5	4	4	4	4	4	4	6	8

Table 5.4: Distribution of p_1 when $n = 200$

$n = 100$	p_2														
t ; θ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.0;-0.9	6	17	10	12	7	7	5	6	4	4	3	4	4	6	7
0.5;-0.9	2	11	9	13	8	9	6	6	4	5	4	5	4	6	8
1.0;-0.9	1	7	9	13	8	10	6	7	5	5	4	5	5	6	8
1.5;-0.9	0	5	8	13	9	11	7	8	5	5	4	5	5	7	9
0.0;-0.65	18	21	10	8	5	4	4	3	3	3	3	3	3	5	7
0.5;-0.65	10	26	13	11	5	5	3	3	3	3	3	3	3	4	6
1.0;-0.65	4	22	16	14	7	6	4	3	3	2	2	3	3	4	6
1.5;-0.65	1	20	18	16	7	6	4	4	3	3	3	3	3	4	6
0.0;-0.5	24	14	9	7	5	4	3	3	3	3	3	4	4	6	9
0.5;-0.5	19	26	10	7	5	3	3	2	3	3	2	3	4	4	6
1.0;-0.5	11	31	13	9	4	4	3	3	2	3	3	3	3	3	6
1.5;-0.5	6	32	15	10	5	4	3	3	3	2	2	3	3	4	6
0.0;0.0	35	11	8	5	5	3	3	3	3	3	2	4	3	5	7
0.5;0.0	23	12	9	6	5	4	3	4	3	3	3	4	4	6	9
1.0;0.0	37	11	7	5	4	4	3	3	2	2	3	3	4	5	8
1.5;0.0	50	11	6	4	3	3	2	2	2	2	2	2	2	3	6
0.0;0.5	11	29	14	9	5	4	3	3	2	2	2	3	3	4	6
0.5;0.5	16	24	13	8	5	4	3	3	3	3	2	3	3	5	7
1.0;0.5	20	14	9	7	5	5	4	4	3	3	3	3	5	6	9
1.5;0.5	26	16	9	6	5	4	3	3	2	2	3	3	4	5	8
0.0;0.65	4	22	17	13	7	5	4	3	3	3	3	3	3	4	6
0.5;0.65	8	22	16	11	7	5	4	3	3	2	3	3	3	4	7
1.0;0.65	15	18	12	9	6	5	3	4	3	3	3	3	4	5	8
1.5;0.65	24	13	8	6	5	4	4	3	3	3	3	4	4	6	9
0.0;0.9	1	9	12	14	10	9	6	6	4	5	4	4	5	5	8
0.5;0.9	2	11	11	12	9	8	6	6	5	4	4	5	5	5	8
1.0;0.9	5	12	11	11	8	8	5	5	4	4	4	4	5	5	9
1.5;0.9	15	13	10	8	5	6	4	4	4	4	4	4	5	6	10

Table 5.5: Distribution of p_2 when $n = 100$

$n = 200$	p_2														
t ; θ	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.0;-0.9	2	8	7	11	9	10	8	8	5	6	4	5	4	5	7
0.5;-0.9	1	4	5	7	8	10	8	10	7	7	6	6	5	7	8
1.0;-0.9	0	3	3	7	7	10	8	10	8	8	6	7	6	8	9
1.5;-0.9	0	2	3	6	6	10	9	11	8	9	6	7	6	7	9
0.0;-0.65	15	25	13	10	6	4	3	3	3	3	3	2	3	3	6
0.5;-0.65	3	18	17	17	8	7	4	4	3	3	2	3	3	3	6
1.0;-0.65	1	11	15	19	12	8	5	4	3	3	2	3	3	3	6
1.5;-0.65	1	8	15	21	12	10	5	4	4	3	3	3	3	4	5
0.0;-0.5	27	16	9	6	5	4	3	3	3	2	3	3	4	4	7
0.5;-0.5	10	30	15	9	5	4	3	3	3	2	2	2	3	3	5
1.0;-0.5	4	25	20	13	7	5	4	3	2	2	2	2	3	4	5
1.5;-0.5	2	21	22	15	7	6	3	3	3	3	2	2	3	3	5
0.0;0.0	45	11	6	5	4	3	3	2	2	2	2	2	3	3	5
0.5;0.0	24	13	9	6	5	4	4	3	3	3	3	4	4	6	8
1.0;0.0	44	11	7	5	4	3	3	2	2	2	2	3	2	3	6
1.5;0.0	49	11	6	5	4	3	3	2	2	2	2	2	3	3	5
0.0;0.5	4	25	20	12	7	5	3	3	2	2	2	2	3	3	5
0.5;0.5	9	27	17	10	6	4	3	3	3	2	2	2	3	3	6
1.0;0.5	20	17	11	7	6	4	4	3	3	3	3	3	4	5	8
1.5;0.5	23	23	12	7	5	4	3	3	2	2	2	3	3	4	6
0.0;0.65	1	12	17	18	11	8	6	4	3	3	2	2	3	4	5
0.5;0.65	2	15	19	16	9	7	5	4	3	3	3	3	3	4	5
1.0;0.65	11	20	16	11	7	5	4	3	3	3	2	3	3	4	6
1.5;0.65	24	15	10	6	5	4	3	3	3	3	3	4	4	5	7
0.0;0.9	0	4	5	9	12	12	10	9	7	6	5	5	5	5	7
0.5;0.9	1	4	6	7	9	11	9	9	7	7	6	6	5	6	8
1.0;0.9	2	6	7	9	9	9	8	8	7	6	5	5	5	6	8
1.5;0.9	10	13	11	10	8	7	5	5	4	4	4	4	4	5	7

Table 5.6: Distribution of p_2 when $n = 200$

	$n = 100$		$n = 200$	
t ; θ	$p \leq 5$	$p > 5$	$p \leq 5$	$p > 5$
0.0;-0.9	27	72	13	72
0.5;-0.9	16	83	4	83
1.0;-0.9	11	89	1	89
1.5;-0.9	10	91	1	91
0.0;-0.65	46	55	48	55
0.5;-0.65	42	57	38	57
1.0;-0.65	37	64	32	64
1.5;-0.65	35	64	27	64
0.0;-0.5	45	55	48	55
0.5;-0.5	47	53	48	53
1.0;-0.5	45	55	43	55
1.5;-0.5	43	56	42	56
0.0;0.0	50	51	54	51
0.5;0.0	45	54	46	54
1.0;0.0	46	53	51	53
1.5;0.0	52	47	52	47
0.0;0.5	44	57	44	57
0.5;0.5	47	52	47	52
1.0;0.5	44	57	47	57
1.5;0.5	47	53	51	53
0.0;0.65	36	64	33	64
0.5;0.65	40	58	37	58
1.0;0.65	45	56	44	56
1.5;0.65	44	56	47	56
0.0;0.9	16	83	5	83
0.5;0.9	16	82	5	82
1.0;0.9	24	77	9	77
1.5;0.9	36	64	33	64

Table 5.7: Aggregated distribution of p aggregated according to $p \leq 5$ and $p > 5$, when $n = 100, 200$

5.2 The descriptive power of TARMA models

As mentioned, Wold decomposition implies that, under mild assumptions, a stationary process admits an AR representation. Hence, an AR(p) model with large p should fit well time series coming from several different DGPs. In this section we compare the descriptive power of TARMA models with respect to AR models. We simulate from 13 DGPs from those described in Table 4.1. Some of these are either linear or non-linear, stationary or non-stationary. Then, we fit the best AR(p) with p ranging from 1 to 10, selected through the AIC and compare it with the two-regime TARMA(1,1). The results are reported in Table 5.8 that shows the percentages of model selection between the best AR(p) and the TARMA model. In general, when the DGP is not stationary (integrated) then the TARMA(1,1) model provides a better fit. Clearly, the MA component plays a role in determining the fitting inability of AR models. The ARI5.1 model is fit well by the AR(6) model as shown in Table 5.9. The TARMA model is also preferred when the DGP is non-linear with the exception of the TAR3.1 model where, again, the AR(5) or AR(6) models seem to provide a better fit. Note that a two-regime TAR(3) has 6 parameters to be estimated. Interestingly, even when the DGP is linear, a TARMA model provides a better fit nearly half of the times. As for non linear moving-average models (NLMA), the TARMA model is always preferred, whereas, as expected, for the AR-GARCH model an AR of low order outperforms the TARMA. The percentage distribution of the selected AR order p is reported in Table 5.9. Note that the modal selected p results 10 both for the IMA(1,1) and the NLMA models.

	$n = 100$		$n = 300$		$n = 500$	
	AR	TARMA	AR	TARMA	AR	TARMA
IMA11.1	18	82	6	94	2	98
IMA11.3	18	82	9	91	7	93
ARI5.1	73	27	75	25	72	28
ARIMA111.1	37	63	37	63	57	43
AR1.1	54	46	57	43	58	42
AR1.2	53	47	57	43	58	42
ARMA11.1	46	54	21	79	11	89
TAR1.1	51	49	24	76	10	90
TAR3.1	95	5	100	0	100	0
3TAR1.1	9	91	0	100	0	100
NLMA.1	0	100	0	100	0	100
NLMA.2	0	100	0	100	0	100
ARGARCH.1	60	40	66	34	68	32

Table 5.8: Percentages of model selection between AR and TARMA for different DGPs and $n = 100, 300, 500$.

	p									
$n = 100$	1	2	3	4	5	6	7	8	9	10
IMA11.1	5	5	9	14	18	14	13	9	8	5
IMA11.3	0	5	1	10	11	19	13	17	10	14
ARI5.1	1	7	11	3	6	51	9	5	4	3
ARIMA111.1	52	17	12	7	5	3	2	1	1	1
AR1.1	72	12	5	4	2	2	1	1	1	1
AR1.2	72	11	6	4	2	2	1	1	1	1
ARMA11.1	22	21	14	13	8	7	4	4	3	3
TAR1.1	71	11	6	4	2	2	1	1	1	1
TAR3.1	0	0	4	12	48	19	6	5	3	2
3TAR1.1	54	14	10	6	5	3	3	2	2	2
NLMA.1	8	10	15	13	13	11	10	8	7	6
NLMA.2	7	10	15	13	14	10	10	7	7	6
ARGARCH.1	38	23	15	8	6	3	3	1	1	1
$n = 300$	1	2	3	4	5	6	7	8	9	10
IMA11.1	0	0	1	2	4	8	14	19	25	28
IMA11.3	0	4	0	2	0	5	7	18	22	41
ARI5.1	1	3	8	3	13	52	9	5	3	3
ARIMA111.1	9	10	14	13	14	11	10	7	6	5
AR1.1	72	11	6	3	2	2	1	1	1	1
AR1.2	72	11	6	4	2	2	1	1	1	1
ARMA11.1	3	7	11	14	14	14	10	10	9	9
TAR1.1	70	12	6	4	2	2	1	1	1	1
TAR3.1	0	0	0	0	34	35	6	12	6	6
3TAR1.1	27	13	11	9	8	7	7	6	6	7
NLMA.1	0	0	1	2	5	9	14	17	23	30
NLMA.2	0	0	1	2	5	9	14	17	23	29
ARGARCH.1	24	18	16	13	9	6	5	4	3	2
$n = 500$	1	2	3	4	5	6	7	8	9	10
IMA11.1	0	0	0	0	1	2	6	13	26	51
IMA11.3	0	5	0	2	0	2	2	10	21	58
ARI5.1	1	3	5	3	18	51	9	5	3	2
ARIMA111.1	1	3	6	10	13	14	14	13	13	13
AR1.1	72	12	6	4	2	2	1	1	1	1
AR1.2	72	11	6	3	2	2	1	1	1	1
ARMA11.1	0	1	4	8	12	15	15	15	14	16
TAR1.1	68	12	7	4	3	2	2	1	1	1
TAR3.1	0	0	0	0	17	41	6	17	8	10
3TAR1.1	12	9	10	8	9	9	9	9	11	13
NLMA.1	0	0	0	0	1	2	6	13	26	51
NLMA.2	0	0	0	0	1	2	7	13	26	51
ARGARCH.1	18	15	16	13	10	8	6	5	4	4

Table 5.9: Percentages of selected AR order p for different DGPs and $n = 100, 300, 500$.

5.3 The Canadian lynx time series

In this section we use TARMA models to fit the *Canadian lynx* data. This data set, as well as the ongoing Wolf's sunspot numbers, has attracted great attention among non-linear time series analysts. The Canadian lynx data set is the annual record of the number of the Canadian lynx trapped in the Mackenzie River district of the North-West Canada for the period 1821-1934 inclusively. The population cycle of these animals has received wide attention especially from biologists due to the regularity in the hunted quantities by the Hudson's Bay Company that has been using them for a long period to produce furs. The data set and further useful materials for the analysis are reported in Elton and Nicholson [1942]. These data are the total fur return, or total sales, from the London archives of the aforementioned company and is a proxy of the dynamics of the population size. There is a time lag between the year in which a lynx was trapped and the year in which its fur was sold and this complicates the analysis. Since TAR models may be seen as the discrete time version of continuous time prey-predator models, [Tong, 1990, Chapter 7, Section 7.2] suggested to fit a TAR model and showed that it was the most appropriate among several alternative models proposed in literature. Here, we revisit the problem by using TARMA models.

In Figure 5.2 we show both the raw (left) and the log10 transformed series (right). Figure 5.3 shows the 2-dimensional lag plots of y_t versus y_{t-1} (left) and y_{t-2} (right), whereas in Figure 5.4 we present the 3d lag plot of $(y_{t-1}, y_{t-1}, y_{t-2})$. These plots highlight the non-linear oscillatory nature of the lynx time series.

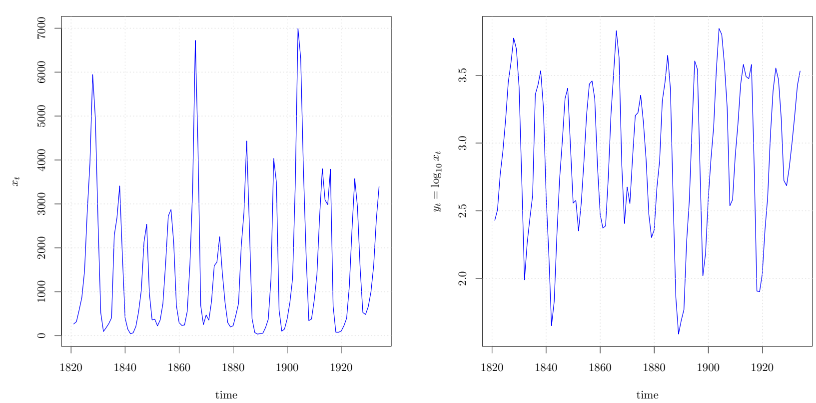

Figure 5.2: Time series of the Canadian lynx time series, from 1821 to 1934. (Left) raw time series. (Right) log-transformed series.

We have fitted several two-regime TAR and TARMA models, including those proposed in literature. Some of them include a single, common intercept rather than one intercept per-regime. Note that in TARMA models the MA parameter θ is also common. A summary of the models, that includes information criteria is presented in Table 5.10. The threshold that minimises the AIC criterion results $r = 3.310$ whereas the value 3.1163, which was used in Tong and Lim [1980] has been used for comparison purposes. This is also reported in Figure 5.5 that shows the values of the AIC versus the threshold r. The minimizer of the AIC is indicated with a blue dashed line whereas the value 3.1163 is indicated in green. Note that, as reported in Tong [1990], 3.1163 corresponds to the antimode but the minimizer of the AIC $r = 3.310$ is closer to the dominant mode of the distribution. This is shown in the right panel of Figure 5.5.

The presence of a single intercept term in indicated with an asterisk in Table 5.10, and the lags column indicates the subset of lags included in the two-regime fit (0 stands for intercept). The best model in terms of AIC is the TAR(9)*, whereas, according to the BIC, the TAR(2)* results the best fit. Note that, both of them have a common intercept and this is consistent with the underlying biology. Moreover, the TAR(9)* may reflect the 9-year cycle of the lynx population and the fit resulted superior with respect to those with lag 7. The TARMA(2,1)* with common intercept outperforms the original TAR(2) (also proposed in [Tong, 1990, Section 7.2]) but results slightly inferior to the TAR models with common intercept.

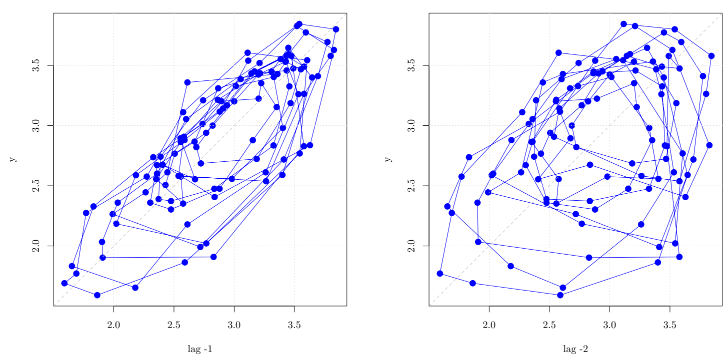

Figure 5.3: Lag plot of the time series of the Canadian lynx. (Left) Lag 1. (Right) Lag 2.

The parameter estimates are presented in Table 5.11 where the set is partitioned in common parameters and parameters belonging to the lower or upper regime. One appreciable difference between the fits with $r = 3.1163$ and r chosen as the minimizer of the AIC is that in the latter case the two intercepts are not so different (see models TAR(2) and TARMA(2,1)). This suggests to refit the models with a single, common intercept and this improves the fit considerably (see models TAR(2)* and TARMA(2,1)*). The inclusion of lag 9 improves further the fit. Note that when lag 9 was added to the TARMA model, then the parameter θ becomes negligible.

The diagnostics for the TAR(2)* fit are presented in Figure 5.6 and Figure 5.7. In particular, Figure 5.6 shows the global and partial correlograms for

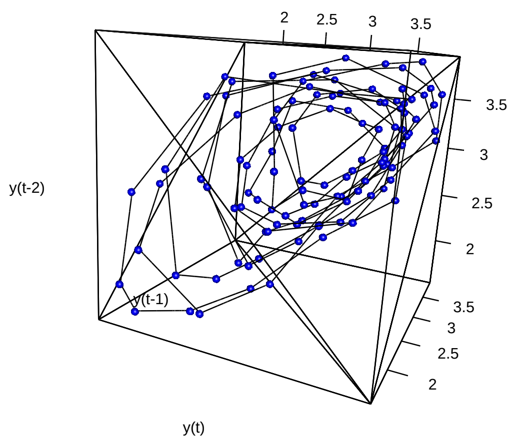

Figure 5.4: Three-dimensional lag plot of the time series of the Canadian lynx.

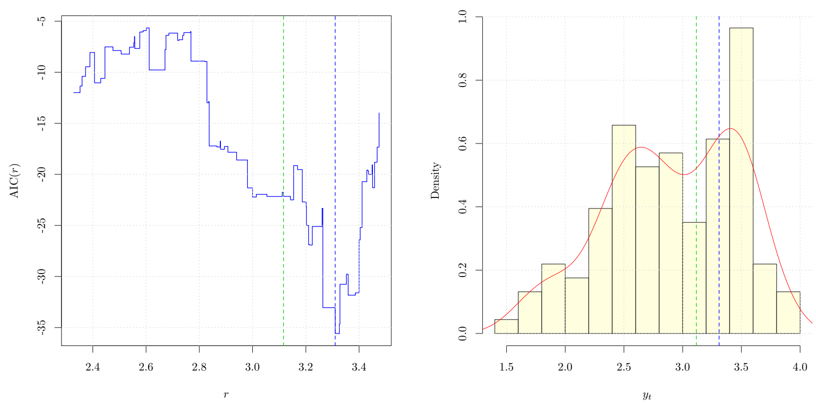

Figure 5.5: (Left) values of the AIC versus the threshold values r. (Right) histogram of y_t with a kernel density estimate superimposed in red. In both panels, the value that minimizes the criterion is indicated with a blue dashed line, while the value used in Tong and Lim [1980] is reported as a green line.

MODEL	common intercept	lags	AIC	BIC	threshold
TARMA(2,1)		0,1,2	-33.459	-9.711	3.310
TARMA(2,1)		0,1,2	-28.522	-4.774	3.116
TARMA(2,1)*	*	1,2	-35.459	-14.429	3.310
TARMA(9,1)*	*	1,2,9	-37.495	-11.609	3.310
TAR(2)		0,1,2	-34.017	-12.987	3.310
TAR(2)		0,1,2	-29.279	-8.249	3.116
TAR(2)*	*	1,2	-35.579	-17.268	3.310
TAR(9)*	*	1,2,9	-39.469	-16.238	3.310

Table 5.10: TAR vs TARMA: lags, AIC, BIC and threshold values. The lags indicate the subset of lagged variables for which two regimes are estimated. A zero in the lag indicates a model with two intercepts, otherwise only one common intercept is estimated.

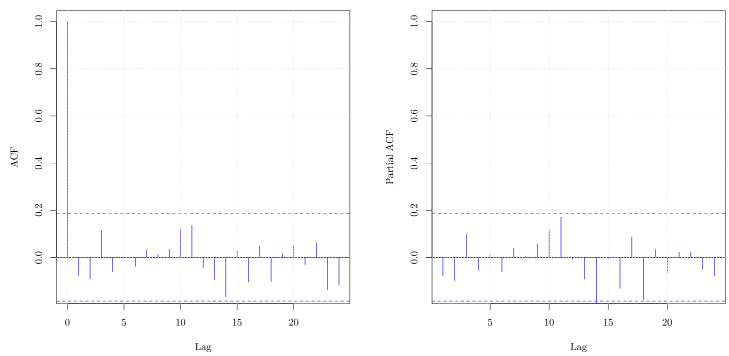

Figure 5.6: Correlograms of the residuals from the TAR(2)* fit for the time series of Canadian lynx.

the residuals and these do not show any structure. The result is reinforced by the entropy measure S_ρ [Giannerini et al., 2015] computed up to lag 12 Figure 5.7(left), and by the qqplot, shown in Figure 5.7(right), together with the Shapiro-Wilk test for normality. We have performed the same analysis for the residuals of the TARMA(2,1)* fit. The results are presented in Figure 5.8 and Figure 5.9. Not that, in this case, the entropy measure, points at a significant dependence at lag 9 in the residuals Figure 5.9(right). If we analyse the residuals of the TARMA(9,1)*, then such dependence is not present anymore, see Figure 5.11(right).

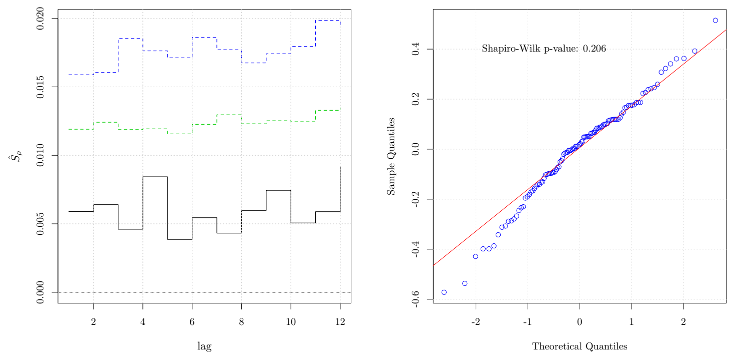

Figure 5.7: Entropy measure S_ρ computed on the residuals from the TAR(2)* fit for the time series of the Canadian lynx. The confidence bands at 95% (green) and 99%(blue) correspond to the null hypothesis of serial independence (left) and qqplot of the residuals with the p-value of the Shapiro-Wilk's normality test. (right).

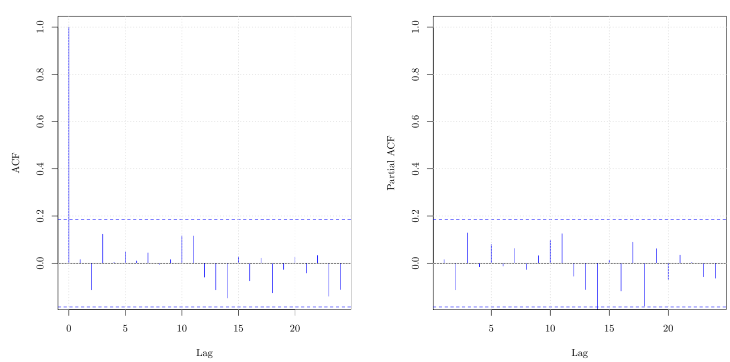

Figure 5.8: Correlograms of the residuals from the TARMA(2,1)* fit for the time series of Canadian lynx.

	common		lower				upper			
MODEL	θ	ϕ_0	$\phi_{1,0}$	$\phi_{1,1}$	$\phi_{1,2}$	$\phi_{1,9}$	$\phi_{2,0}$	$\phi_{2,1}$	$\phi_{2,2}$	$\phi_{2,9}$
TARMA(2,1)	-0.199		0.586	1.341	-0.506		0.608	1.665	-0.916	
s.e.	(0.152)		(0.122)	(0.079)	(0.089)		(0.913)	(0.115)	(0.257)	
TARMA(2,1)	-0.166		0.499	1.303	-0.428		1.964	1.593	-1.225	
s.e.	(0.139)		(0.152)	(0.085)	(0.102)		(0.611)	(0.100)	(0.167)	
TARMA(2,1)*	-0.201	0.586		1.341	-0.507			1.665	-0.911	
s.e.	(0.133)	(0.122)		(0.075)	(0.084)			(0.112)	(0.113)	
TARMA(9,1)*	0.028	0.262		1.117	-0.307	0.135		1.356	-0.680	0.200
s.e.	(0.178)	(0.203)		(0.137)	(0.128)	(0.067)		(0.142)	(0.130)	(0.077)
TAR(2)			0.588	1.264	-0.428		1.166	1.599	-1.012	
s.e.			(0.141)	(0.064)	(0.076)		(0.861)	(0.107)	(0.260)	
TAR(2)			0.503	1.238	-0.363		2.314	1.528	-1.263	
s.e.			(0.169)	(0.072)	(0.093)		(0.563)	(0.090)	(0.173)	
TAR(2)*		0.603		1.263	-0.433			1.602	-0.855	
s.e.		(0.139)		(0.064)	(0.076)			(0.107)	(0.110)	
TAR(9)*		0.275		1.134	-0.322	0.129		1.367	-0.690	0.194
s.e.		(0.182)		(0.081)	(0.084)	(0.052)		(0.123)	(0.113)	(0.067)

Table 5.11: TAR vs TARMA: estimated parameters divided in common parameters, upper regime and lower regime. Models with an asterisk contain a common intercept term.

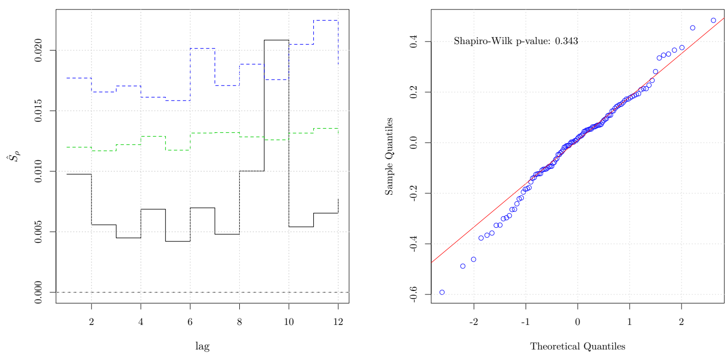

Figure 5.9: Entropy measure S_ρ computed on the residuals from the TARMA(2,1)* fit for the time series of the Canadian lynx. The confidence bands at 95% (green) and 99%(blue) correspond to the null hypothesis of serial independence (left) and qqplot of the residuals with the p-value of the Shapiro-Wilk's normality test. (right).

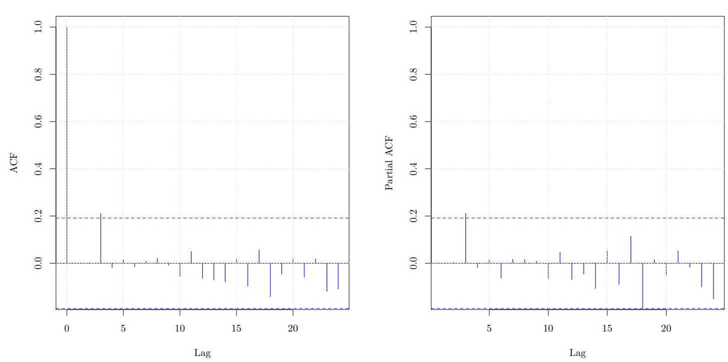

Figure 5.10: Correlograms of the residuals from the TARMA(9,1)* fit for the time series of Canadian lynx.

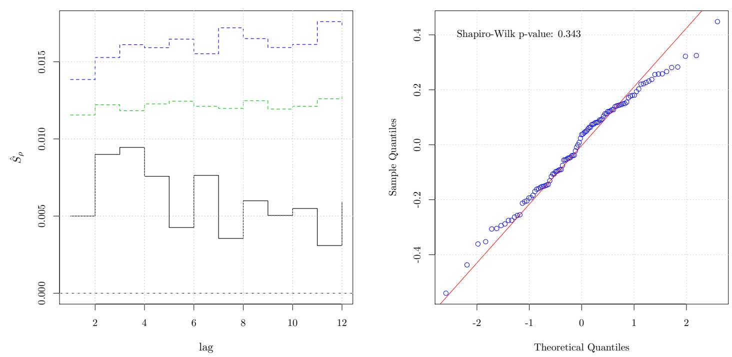

Figure 5.11: Entropy measure S_ρ computed on the residuals from the TARMA(9,1)* fit for the time series of the Canadian lynx. The confidence bands at 95% (green) and 99%(blue) correspond to the null hypothesis of serial independence (left) and qqplot of the residuals with the p-value of the Shapiro-Wilk's normality test. (right).

Bibliography

C. Agiakloglou and P. Newbold. Empirical evidence on Dickey-Fuller-type tests. *Journal of Time Series Analysis*, 13(6):471–483, 1999. doi: 10.1111/j.1467-9892.1992.tb00121.x.

H. Akaike. Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Annals of the Institute of Statistical Mathematics*, 26(1):363–387, Dec 1974. doi: 10.1007/BF02479833. URL https://doi.org/10.1007/BF02479833.

H.Z. An and S.G. Chen. A note on the ergodicity of non-linear autoregressive model. *Statistics & Probability Letters*, 34(4):365–372, 1997.

H.Z. An and F.C. Huang. The geometrical ergodicity of nonlinear autoregressive models. *Statistica Sinica*, 6(4):943–956, 1996.

N.S. Balke and T.B. Fomby. Threshold cointegration. *International economic review*, pages 627–645, 1997.

F. Bec, M. Ben Salem, and M. Carrasco. Tests for unit-root versus threshold specification with an application to the purchasing power parity relationship. *Journal of Business & Economic Statistics*, 22(4):382–395, 2004.

F. Bec, A. Guay, and E. Guerre. Adaptive consistent unit-root tests based on autoregressive threshold model. *Journal of Econometrics*, 142(1):94–133, 2008a.

F. Bec, A. Rahbek, and N. Shephard. The ACR Model: a multivariate dynamic mixture autoregression. *Oxford Bulletin of Economics and Statistics*, 70(5):583–618, 2008b.

P.J. Bickel and P. Bühlmann. Closure of linear processes. *Journal of Theoretical Probability*, 10(2):445–479, Apr 1997. ISSN 1572-9230. doi: 10.1023/A:1022616601841. URL https://doi.org/10.1023/A:1022616601841.

P. Billingsley. *Convergence of probability measure*. Wiley, New York, 1968.

P.J. Brockwell and R.A. Davis. *Time series: theory and methods*. Springer, 2001.

P.J. Brockwell, J. Liu, and R.L. Tweedie. On the existence of stationary threshold autoregressive moving-average processes. *Journal of Time Series Analysis*, 13(2):95–107, 1992.

M. Caner and B.E. Hansen. Threshold autoregression with a unit root. *Econometrica*, 69(6):1555–1596, 2001.

G. Cavaliere and A.M.R. Taylor. Bootstrap unit root tests for time series with nonstationary volatility. *Econometric Theory*, 24(1):43–71, 2008.

G. Cavaliere, P.C.B. Phillips, S. Smeekes, and A.M.R. Taylor. Lag length selection for unit root tests in the presence of nonstationary volatility. *Econometric Reviews*, 34(4):512–536, 2015. doi: 10.1080/07474938.2013.808065.

K.-S. Chan. Testing for threshold autoregression. *Ann. Statist.*, 18(4):1886–1894, 12 1990. doi: 10.1214/aos/1176347886. URL https://doi.org/10.1214/aos/1176347886.

K.-S. Chan. On the central limit theorem for an ergodic markov chain. *Stochastic Processes and their Applications*, 47(1):113–117, 1993.

K.-S. Chan. *Exploration of a Nonlinear World: An Appreciation of Howell Tong's Contributions to Statistics*. World Scientific, 2009.

K.-S. Chan and G. Goracci. Recurrence properties of the first-order Threshold Autoregressive Moving-average model. Technical report, University of Iowa, 2017.

K.-S. Chan and H. Tong. On the use of the deterministic lyapunov function for the ergodicity of stochastic difference equations. *Advances in applied probability*, 17(3):666–678, 1985.

K.-S. Chan and H. Tong. A note on the invertibility of nonlinear ARMA models. *Journal of Statistical Planning and Inference*, 140(12):3709–3714, 2010.

K.-S. Chan, J.D. Petruccelli, H. Tong, and S.W. Woolford. A multiple-threshold AR (1) model. *Journal of Applied Probability*, 22(02):267–279, 1985.

K.-S. Chan, B.E. Hansen, and A. Timmermann. Guest editors' introduction: Regime switching and threshold models. *Journal of Business & Economic Statistics*, 35(2):159–161, 2017. doi: 10.1080/07350015.2017.1236521. URL http://dx.doi.org/10.1080/07350015.2017.1236521.

K.-S. Chan, S. Giannerini, G. Goracci, and H. Tong. Unit-root tests for linear and non-linear alternatives: a tarma based approach. Technical report, University of Iowa, 2018.

C. Chatfield. *Time-series forecasting*. CRC Press, 2000.

R. Chen and R.S. Tsay. On the ergodicity of TAR(1) processes. *Ann. Appl. Probab.*, 1(4):613–634, 11 1991.

I. Choi. *Almost all about unit roots: foundations, developments, and applications*. Themes in Modern Econometrics. Cambridge University Press, 2015. ISBN 9781316300589.

D.B.H. Cline. Thoughts on the connection between threshold time series models and dynamical systems. In K-S. Chan, editor, *Exploration of a nonlinear world. An Appreciation of Howell Tong's Contributions to Statistics*, pages 165–181. World Scientific, Singapore, 2009.

S. Cook. Serial correlation, drift and range unit root testing. *Applied Economics Letters*, 17(10):939–944, 2010. doi: 10.1080/13504850802660367.

J.D. Cryer and K.-S. Chan. *Time Series Analysis: With Applications in R (2nd edition)*. Springer, New York, 2008.

R. Davidson. Size distortion of bootstrap tests: an example from unit root testing. *Review of Economic Analysis*, 2(2):169–193, June 2010.

R. Davidson and E. Flachaire. The wild bootstrap, tamed at last. *Journal of Econometrics*, 146(1):162–169, 2008. doi: https://doi.org/10.1016/j.jeconom.2008.08.003.

R.M. de Jong, C-H. Wang, and Y. Bae. Correlation robust threshold unit root tests. Mimeo, Ohio State University, Michigan, 2007.

D.A. Dickey and W.A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431, 1979.

G. Elliott, T.J. Rothenberg, and J.H. Stock. Efficient tests for an autoregressive unit root. *Econometrica*, 64(4):813–836, 1996. URL http://www.jstor.org/stable/2171846.

C. Elton and M. Nicholson. The ten-year cycle in numbers of the lynx in canada. *Journal of Animal Ecology*, 11(2):215–244, 1942. URL http://www.jstor.org/stable/1358.

W. Enders and C.W.J. Granger. Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates. *Journal of Business & Economic Statistics*, 16(3):304–311, 1998.

J.W. Galbraith and V. Zinde-Walsh. On the distributions of Augmented Dickey—Fuller statistics in processes with moving average components. *Journal of Econometrics*, 93(1):25–47, 1999. doi: https://doi.org/10.1016/S0304-4076(98)00097-9.

E.S. Gardner. Exponential smoothing: the state of the art. *Journal of forecasting*, 4(1):1–28, 1985.

S. Giannerini, E. Maasoumi, and E. Bee Dagum. Entropy testing for nonlinear serial dependence in time series. *Biometrika*, 102:661–675, 2015. URL http://biomet.oxfordjournals.org/content/102/3/661.abstract.

F. Giordano, M. Niglio, and C.D. Vitale. Unit root testing in presence of a double threshold process. *Methodology and Computing in Applied Probability*, 19(2):539–556, 2017.

I. Girsanov. On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory of Probability & Its Applications*, 5(3):285–301, 1960. doi: 10.1137/1105027.

G. Goracci. On the parsimony of TARMA models for non linear time series. Technical report, University of Iowa, 2018.

C. Gourieroux and C.Y. Robert. Stochastic unit root models. *Econometric Theory*, 22(6):1052–1090, 2006. URL http://www.jstor.org/stable/4093213.

M. Guo and J.D. Petruccelli. On the null recurrence and transience of a first-order setar model. *Journal of Applied Probability*, 28(3):584–592, 1991.

N. Haldrup and M. Jansson. Improving size and power in unit root testing. In K. Patterson and T.C. Mills, editors, *Palgrave Handbook of Econometrics: Volume 1: Econometric Theory*, pages 252–277. Palgrave Macmillan, 2006.

B.E. Hansen. Threshold autoregression in economics. *Statistics and its Interface*, 4(2):123–127, 2011.

C.C. Holt. Forecasting seasonals and trends by exponentially weighted moving averages. *International journal of forecasting*, 20(1):5–10, 2004.

M. Hosseinkouchack and U. Hassler. Powerful unit root tests free of nuisance parameters. *Journal of Time Series Analysis*, 37(4):533–554, 2016. doi: 10.1111/jtsa.12172.

R. Hyndman, A.B. Koehler, J.K. Ord, and R.D. Snyder. *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media, 2008.

G. Kapetanios and Y. Shin. Unit root tests in three-regime SETAR models. *The Econometrics Journal*, 9(2):252–278, 2006.

N. Kazamaki. On a problem of girsanov. *Tohoku Math. J. (2)*, 29(4):597–600, 1977. doi: 10.2748/tmj/1178240496. URL https://doi.org/10.2748/tmj/1178240496.

M. Klau and S.S. Fung. The new bis effective exchange rate indices. Bis quarterly review, Bank of International Settlements, March 2006.

H. Kunita. Tightness of probability measures in $d([0,t];c)$ and $d([0,t];d)$. *J. Math. Soc. Japan*, 38(2):309–334, 04 1986. doi: 10.2969/jmsj/03820309. URL https://doi.org/10.2969/jmsj/03820309.

T.G. Kurtz and P.E. Protter. Weak convergence of stochastic integrals and differential equations. *Lecture notes in mathematics-Springer Verlag*, pages 1–41, 1996.

D. Li and S. Ling. On the least squares estimation of multiple-regime threshold autoregressive models. 167(1):240–253, 2012.

Wenbo V. Li. Limit theorems for the square integral of brownian motion and its increments. *Stochastic Processes and their Applications*, 41(2):223–239, 1992. ISSN 0304-4149.

S. Ling. On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. *J. Appl. Probab.*, 36(3):688–705, 09 1999.

S. Ling, H. Tong, and D. Li. Ergodicity and invertibility of threshold moving-average models. *Bernoulli*, 13(1):161–168, 2007.

J. Liu and E. Susko. On strict stationarity and ergodicity of a non-linear ARMA model. *Journal of Applied Probability*, 29(2):363–373, 1992.

R.Y. Liu. Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.*, 16 (4):1696–1708, 12 1988. doi: 10.1214/aos/1176351062. URL https://doi.org/10.1214/aos/1176351062.

E. Mammen. Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, 21(1):255–285, 1993. ISSN 00905364. URL http://www.jstor.org/stable/3035590.

S.P. Meyn. Ergodic theorems for discrete time stochastic systems using a stochastic lyapunov function. *SIAM Journal on Control and Optimization*, 27(6):1409–1439, 1989.

S.P. Meyn and P.E. Caines. *Stochastic controllability and stochastic Lyapunov functions with applications to adaptive and nonlinear systems*, pages 235–257. Springer Berlin Heidelberg, Berlin, Heidelberg, 1989.

S.P. Meyn and R.L. Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.

S. Ng and P. Perron. Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69(6):1519–1554, 2001. doi: 10.1111/1468-0262.00256.

E. Nummelin. *General irreducible Markov chains and non-negative operators*, volume 83. Cambridge University Press, 2004.

B. Øksendal. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003. doi: 10.1007/978-3-642-14394-6_5. URL https://doi.org/10.1007/978-3-642-14394-6_5.

E. Paparoditis and D.N. Politis. The asymptotic size and power of the augmented dickey–fuller test for a unit root. *Econometric Reviews*, 37(9):955–973, 2018. doi: 10.1080/00927872.2016.1178887.

J.Y. Park and P.C.B. Phillips. Nonlinear regressions with integrated time series. *Econometrica*, 69(1):117–161, 2001.

J.Y. Park and M. Shintani. Testing for a unit root against transitional autoregressive models. *International Economic Review*, 57(2):635–664, 2016.

K. Patterson. *A Primer for Unit Root Testing*. Palgrave Texts in Econometrics. Palgrave Macmillan UK, 2010. ISBN 9780230248458.

K. Patterson. *Unit Root Tests in Time Series Volume 1: Key Concepts and Problems*. Palgrave Texts in Econometrics. Palgrave Macmillan UK, 2011. ISBN 9780230250246.

K. Patterson. *Unit Root Tests in Time Series Volume 2: Extensions and Developments*. Palgrave Texts in Econometrics. Palgrave Macmillan, 2012. ISBN 9780230250260.

M. Pelagatti and E. Colombo. On the empirical failure of purchasing power parity tests. *Journal of Applied Econometrics*, 30(6):904–923, 2015. doi: 10.1002/jae.2418.

P. Perron and S. Ng. Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *The Review of Economic Studies*, 63(3):435–463, 1996.

P. Perron and Z. Qu. A simple modification to improve the finite sample properties of Ng and Perron's unit root tests. *Economics Letters*, 94(1):12–19, 2007. ISSN 0165-1765.

J.D. Petruccelli and S.W. Woolford. A threshold AR (1) model. *Journal of Applied Probability*, 21(02):270–286, 1984.

M.K. Pippenger and G.E. Goering. Practitioners corner: a note on the empirical power of unit root tests under threshold processes. *Oxford Bulletin of Economics and Statistics*, 55(4):473–481, 1993.

D. Pollard. *Convergence of stochastic processes*. Springer Science & Business Media, 2012.

S.E. Said and D.A. Dickey. Hypothesis testing in ARIMA(p,1,q) models. *Journal of the American Statistical Association*, 80(390):369–374, 1985.

M.H. Seo. Unit root test in a threshold autoregression: asymptotic theory and residual-based block bootstrap. *Econometric Theory*, 24(6):1699–1716, 2008.

A.M. Taylor. Potential pitfalls for the purchasing-power-parity puzzle? sampling and specification biases in mean-reversion tests of the law of one price. *Econometrica*, 69(2):473–498, 2001.

D. Tjøstheim. Non-linear time series and markov chains. *Advances in Applied Probability*, 22(3):587–611, 1990.

H. Tong. On a threshold model. In C.H. Chen, editor, *Pattern recognition and signal processing*, NATO ASI Series E: Applied Sc.(29), pages 575–586. Sijthoff & Noordhoff, Amsterdam, 1978.

H. Tong. *Non-linear Time Series: A Dynamical System Approach*. Clarendon Press, 1990.

H. Tong. Birth of the threshold time series model. *Statistica Sinica*, 17(1):8–14, 2007.

H. Tong. Threshold models in time series analysis–30 years on. *Statistics and its Interface*, 4(2):107–118, 2011.

H. Tong and K.S. Lim. Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 245–292, 1980.

R.L. Tweedie. Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stochastic Processes and their Applications*, 3(4):385–403, 1975.

R.L. Tweedie. Criteria for classifying general Markov chains. *Advances in Applied Probability*, 8(4):737–771, 1976.

R.L. Tweedie. The existence of moments for stationary markov chains. *Journal of Applied Probability*, 20(1):191–196, 1983a.

R.L. Tweedie. The existence of moments for stationary Markov chains. *Journal of Applied Probability*, 20(1):191–196, 1983b.

A.W. van der Vaart. *Asymptotic statistics.* Cambridge series in statistical and probabilistic Mathematics, Cambridge University Press, 1998.