ASYMPTOTIC THEORY FOR MODERATE DEVIATIONS FROM THE UNIT BOUNDARY IN QUANTILE AUTOREGRESSIVE TIME SERIES*

Christis Katsouris†

Department of Economics, University of Southampton

March 22, 2022

Abstract

We establish the asymptotic theory in quantile autoregression when the model parameter is specified with respect to moderate deviations from the unit boundary such that $\rho_n = \left(1 + \frac{c}{k_n}\right)$ where $(k_n)_{n \in \mathbb{N}}$ is a nonrandom sequence that diverges at a rate slower than the sample size n. Then, extending the framework proposed by Phillips and Magdalinos (2007), we consider the limit theory for the near-stationary and the near-explosive cases when the model is estimated with a conditional quantile specification function and model parameters are quantile-dependent. Additionally, a Bahadur-type representation and limiting distributions based on the M-estimators of the model parameters are derived. Specifically, we show that the serial correlation coefficient converges in distribution to a ratio of two independent random variables. Monte Carlo simulations illustrate the finite-sample performance of the estimation procedure under investigation.

Keywords: Quantile autoregressive model, moderate deviations, local-to-unity, near-integrated processes, explosive processes, bahadur representation.

1. Introduction

Moderate deviation principles from the unit boundary for quantile regression models are commonly employed when considering the limit distribution of quantile-dependent parameters under regressors nonstationarity. In particular, the development of asymptotic theory for nonstationary quantile time series models has been pioneered by the studies of Koenker and Xiao (2004, 2006) as well as Koenker and Xiao (2002) who investigate estimation and inference aspects for regression quantile models (see, also Hasan and Koenker (1997)). On the contrary, studies for quantile autoregressive time series models that consider moderate deviations within a unified framework allowing to investigate the asymptotic behaviour of estimators with respect to different modes of stability such as stable, unstable and explosive processes has seen less attention in the literature. Therefore, our main objective is to use the moderate deviation principles in order to derive the limiting distribution of the serial correlation coefficient when considering deviations from the unit boundary in the quantile autoregressive model. More precisely, the present paper builds on the framework proposed by Phillips and Magdalinos (2007) (see, also Giraitis and Phillips (2006) and Huang et al. (2014)) that corresponds to the linear autoregressive model under nonstationarity as well as the study of Kong (2015) who develops related limit results for moderate deviations in autoregressive models based on M-estimators.

*Article history: February 2022.

†Ph.D. Candidate, Department of Economics, University of Southampton, Highfield Campus, Southampton, SO17 1BJ, UK. *E-mail Address*: C.Katsouris@soton.ac.uk.

According to Chan et al. (2006): "*The study of the unit root AR(1) model has been actively pursued by statisticians and econometricians alike, and a related question that needs to be addressed is what happens to the limiting distribution of the test statistics when the autoregressive parameter θ_n is close to the unit boundary? Consequently, when the autocorrelation coefficient is expressed with respect to the local unit root specification, what kind of approximation should be used for the distribution of the test statistics?*. To answer this question, Chan and Wei (1987) as well as Phillips (1987a,b) proposed the triangular array framework, so-called nearly nonstationary AR(1) model, and established the limiting distributions of the least squares estimator for θ_n under the assumption that the conditional variance of the model ε_t is finite. Further results of the least squares estimate for the near-integrated AR(1) model are presented by Chan (1988, 1990), Chan and Tran (1989), Knight (1987), Rao (1978), Lai and Wei (1982), Cox and Llatas (1991), Larsson (1995) as well as by Cavaliere (2002).

Additionally, the study of Hui et al. (2022) establishes the asymptotic theory of the ordinary least squares estimator in the explosive first-order Gaussian autoregressive process. Specifically, using a set of deviation inequalities the authors obtain the limit theory of Cramér-type moderate deviations for the explosive and mildly explosive autoregressive processes. Thus, in this paper we focus on establishing the limit theory for moderate deviations from the unit boundary for both the near-stationary and near-explosive cases. Although, White (1958) obtained the limiting distribution of an explosive serial correlation coefficient (see, also Mann and Wald (1943)), limit results for moderate deviations on the explosive side of unity (e.g., mildly explosive case) were only recently established by the studies of Phillips and Magdalinos (2007), Buchmann and Chan (2007) and Aue and Horváth (2007).

Under the Cramér-type moderate deviations framework, there exists positive sequences ν_n and λ_n tending to infinity such that for every $\ell > 0$ as $n \to \infty$ it holds that (e.g., see Hui et al. (2022))

$$\sup_{0 \leq x \leq \ell \lambda_n} \left| \frac{1}{1 - F_n(x)} \mathbb{P}\left(\nu_n \left(\hat{\theta}_n - \theta_n \right) \geq x \right) - 1 \right| \to 0, \tag{1.1}$$

where $F_n(x)$ is the distribution function, satisfying for all $x \in \mathbb{R}$

$$\mathbb{P}\left(\nu_n \left(\hat{\theta}_n - \theta_n \right) \geq x \right) - F_n(x) \overset{p}{\to} 0, \quad \text{as } n \to \infty. \tag{1.2}$$

Therefore, we consider the neighbourhood near the unit boundary, which can be approaching unity from below (near-stationary or near-integrated) or approaching unity from above (near-explosive). Specifically, due to the form of the nonrandom sequence such that, $k_n \equiv n^{\gamma}$, the convergence rate towards to unit boundary is slower than the sample size n. In this case the autocorrelation coefficient approaches 1 at a rate slower than the usual local alternative as n goes to infinity. On the other hand, as $\gamma \to 1$ then $k_n \to n$ which encompasses the conventional local to unit root specification. An alternative form, for the convergence rate include for example the case when $k_n = \sqrt{n}$.

Furthermore to obtain limit results for model parameters based on the M-estimators, we employ the Bahadur representation (see, Bahadur (1966)) that provides a mechanism to facilitate the asymptotic theory analysis since it allows to obtain analytic expressions for quantile-dependent estimators by approximating them with linear forms. Thus, when applied to the quantile autoregressive model it permits to obtain joint convergence results. The limit theory for an autoregressive model which includes a intercept and an autocorrelation coefficient expressed using the local unit root specification is studied by Liu and Liu (2018), however their framework differs from our setting since we consider quantile-dependent parameters. Therefore, this paper builds on and contributes to both the quantile autoregression literature as well as to the literature of moderate deviations from the unit boundary.

1.1. Preliminary Theory

The asymptotic theory for moderate deviations from a unit root in autoregressive models is examined in various studies such as by Fountis and Dickey (1989), Jiang et al. (2015) who focus on the aspect of dependent errors in AR(1) models and Yabe (2012) who obtain limit results for MA(1) time series models[1]. Furthermore, limit results for moderate deviations for M-estimators and quantile processes[2] in autoregressive models are investigated in several studies such as by Jurečková et al. (1988), Knight (1998), Mao and Guo (2019) as well as Kato (2009) where the author develops asymptotics for Lasso quantile-dependent estimators. On the other hand, within the context of quantile autoregressive models related studies which consider moderate deviations from the unit boundary include Lucas (1995), Abadir and Lucas (2000), Ling and McAleer (2004), Chan et al. (2006) and Kong (2015).

Regression asymptotics with roots at or near unity are typically carried out by using autoregressive models with fixed coefficients and then testing for the autoregressive parameter being equal to 1, as pointed out by Dickey and Fuller (1979) (see, also Dickey and Fuller (1981)). More precisely, the idea of developing asymptotics using the near-unit root specification is due to the studies of Cavanagh (1985), Phillips (1987a), Chan and Wei (1987). In particular, Phillips and Magdalinos (2007) showed that $\hat{\theta}_n - \theta_n$ has a $\sqrt{nk_n}$ rate of convergence and a limit normal distribution when $c < 0$,

$$\sqrt{n\kappa_n}\left(\hat{\theta}_n - \theta_n\right) \overset{d}{\to} \mathcal{N}(0, -2c) \tag{1.3}$$

Thus, we are interested to establish a martingale central limit theorem for a normalized version of $\sum_{t=1}^{n} y_{t-1}u_t$ which can give rise to a Gaussian asymptotic distribution for the normalized and centered least squares estimator specifically for the quantile autoregressive model.

A different stream of literature considers a representation of the autoregressive model based on the exponential family with specific canonical parameter. Therefore, by expressing the AR(1) model with respect to the canonical parameters of an exponential family we can obtain insights regarding the asymptotic behaviour of related minimal sufficient statistics. The particular literature is developed under the assumption of a stationary autocorrelation coefficient such that $|\theta| < 1$ for $y_t = \theta y_{t-1} + u_t$. Furthermore, it has been argued that the Efron curvature depends heavily on the AR parameter, especially near the boundary of the parameter space, and increasingly so with increasing sample size[3] (see, Garderen (1999)). In addition, Jansson and Moreira (2006) using the local-unit-root specification of the autocorrelation coefficient, study the properties of predictive regression models under the assumption of persistence regressors using differential geometry and sufficient statistics arguments to establish the asymptotic theory of estimators and test statistics. Lastly, an important aspect worth mentioning is the fact that several studies have demonstrated that the nuisance parameter of persistence cannot be consistently estimated (see, Phillips et al. (2001), Mikusheva (2012) among others). Similarly when considering the quantile autoregressive model, the availability of a consistent estimator for the unknown coefficient of persistence, c, still remains a challenging issue. However, our asymptotic theory analysis which focus on the asymptotic behaviour of the quantile-dependent parameter remains valid despite the absence of such a desirable statistical property.

[1] The MA(1) process is considered to be invertible if and only if the root is away from the unit boundary. Moreover, estimators and test statistics for non-invertible processes behave quite differently from an invertible MA(1) time series model. In particular, the CLT holds for the maximum likelihood estimator for an invertible process while for a non-invertible process the CLT does not hold because the maximum likelihood estimator has a probability mass at unity.

[2] Related theoretical aspects can be found in the book of Csörgő (1983) (see, also Csorgo et al. (1986)).

[3] In other words, when the autocorrelation coefficient is expressed using the moderate deviations form then unit boundary then this can enable the analysis of unit roots and explosive processes, which is necessary to link problems in inference for unit roots to the statistical curvature.

The main research objective of our paper is the asymptotic behaviour of the quantile-dependent estimators. More precisely, we establish the asymptotic theory of estimators in possibly nonstationary quantile autoregressive models for the general near-integrated case assuming that $\theta_n \to 1$ with a rate slower than $1/n$. In particular, Aue and Horváth (2007) develop the limit theory for the serial correlation coefficient in the mildly explosive case with moderate deviations from the unit boundary. Thus, the model parameter satisfies $\theta_n \to 1$ and $n(\theta_n - 1) \to \infty$, as $n \to \infty$, while for large n, $\theta_n > 1$ diverges away from unity but not with the usual convergence rate $\mathcal{O}(1/n)$.

The rest of the paper is organized as follows. Section 2, introduces the framework of moderate deviation principles from the unit boundary in quantile autoregressive processes as well as the main estimation methodology. Section 3, demonstrates the main results of the paper, that is, the limit theory for the near-integrated and near-explosive cases for the quantile autoregression. Section 4 presents a short Monte Carlo simulation study while Section 5 an empirical application. Section 6 concludes.

2. Quantile Autoregression under the Moderate Deviations Framework

Consider the autoregressive AR(1) time series model with an intercept expressed as below

$$y_t = \mu_n + \rho_{n,c} y_{t-1} + \varepsilon_t, \qquad \text{for } \ t = 1, ..., n, \tag{2.1}$$

with

$$\rho_{n,c} = \left(1 + \frac{c}{k_n}\right), \text{ with } \ c \in \mathbb{R} \ \text{ and } \gamma \in (0,1). \tag{2.2}$$

with innovations $(\varepsilon_t)_{t \in \mathbb{N}}$ an *i.i.d* sequence of random variables from a common distribution function F_ε that satisfies regularity conditions for Lipschitz continuity with zero mean and finite variance σ_ε. Moreover, we denote by $\mathcal{F}_t$ the natural filtration $\sigma(\varepsilon_t, \varepsilon_{t-1}, ...)$ and with $\mathbb{E}_{\mathcal{F}_t}[.]$ the corresponding conditional expectation. Furthermore, the convergence rate of the autocorrelation coefficient $\rho_{n,c}$ is such that $k_n := n^\gamma$ where the exponent rate is defined such that $\gamma \in (0,1)$. Specifically, the given convergence rate implies that the $(k_n)_{n \in \mathbb{N}}$ sequence increases to infinity at a slower rate than the sample size such that $k_n = o(n)$ as $n \to \infty$.

Assumption 2.1. *Denote with $y_{0n} = 0$, almost surely, for all $n \in \mathbb{N}$. Then, the innovations $(\varepsilon_t)_{t \in \mathbb{N}}$ forms a sequence of martingale differences such that as $n \to \infty$*

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[\varepsilon_t^2 | \mathcal{F}_{t-1}\right] = 1 + o_p(1), \tag{2.3}$$

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[\varepsilon_t^2 \mathbf{1}\left\{|\varepsilon_t| > n^{1/2} m\right\} | \mathcal{F}_{t-1}\right] = o_p(1), \quad \textit{for all } m > 0, \tag{2.4}$$

where $\mathcal{F}_t := \sigma(\varepsilon_s : 0 \leq s \leq t)$.

Assumption 2.1 provides moment conditions and a uniform integrability condition which induces a restriction on the tail behaviour of the underline distribution of innovations. Furthermore, for the development of the asymptotic theory related conditions are imposed to the noise sequences ε_t such that the corresponding partial sum processes lie in the domain of attraction of functionals of Brownian motions. Then, asymptotic approximations are obtained based on the Ornstein-Uhlenbeck processes.

Assumption 2.2. *The distribution F_ε is in the domain of attraction of a stable law indexed with $\alpha \in (0,2)$, which has a strictly positive density. When $\alpha = 2$ then it corresponds to the case of innovation sequences from a distribution function with finite variance. Furthermore,*

$$\mathbb{E}(\varepsilon_1) = 0, \quad \mathsf{Var}(\varepsilon_1) = \sigma_\varepsilon^2, \quad \textit{and} \quad \mathbb{E}|\varepsilon_1|^{2+m} < \infty, \ \textit{for some } m > 0. \tag{2.5}$$

When $c = 0$ then the AR(1) model is a random walk model with stable innovations. Therefore, under the assumption that $c = 0$, which implies the presence of an integrated process the asymptotic behaviour of the ordinary least squares estimator can be obtained based on Brownian motion functionals. Define with $U_\alpha(r)$ and $V_\alpha(r)$ to be Lévy processes on the space of functions $\mathcal{D}[0,1]$.

Lemma 2.1. *Suppose that ε_t satisfies Assumption 2.2. Then as $n \to \infty$ it holds that*

$$\left(\frac{1}{\nu_n} \sum_{t=1}^{\lfloor nr \rfloor} \varepsilon_t, \frac{1}{\nu_n^2} \sum_{t=1}^{\lfloor nr \rfloor} \varepsilon_t^2 \right) \Rightarrow \Big(U_\alpha(r), V_\alpha(r) \Big) \tag{2.6}$$

where $\big(U_\alpha(r), V_\alpha(r)\big)$ is a Levy process in $\mathcal{D}[0,1]^2$ with index $\alpha \in (0,2)$.

Remark 2.1. *A general class of one-dimensional stochastic processes, are the so-called Lévy processes. Similar to Wiener processes, Lévy processes have right continuous paths with left limits, are initiated from the origin and both have stationary and independent increments (Kyprianou (2014)). Under the i.i.d innovation assumption it can be shown that $V_\alpha(r) = \int_0^r \big(dU_\alpha(s)\big)^2 = [U_\alpha, U_\alpha]_{r\in[0,1]}$ which is the quadratic variation of the Levy process $U_\alpha(r)$. Furthermore, $V_\alpha(r)$ is a stochastic integral. More precisely, when $\alpha = 2$, which corresponds to the finite variance case, it holds that $U_\alpha(r) \equiv W(r)$ for some $0 \leq r \leq 1$, the standard Brownian motion, and $V_\alpha(r) = [W,W]_r = r$ (see, also Cramér (1951)).*

Therefore, any asymptotic results followed by Lemma 2.1 coincide with the standard random walk asymptotics for finite variance models. Specifically, in this paper we focus in the case $\alpha = 2$. Then as $n \to \infty$ the following limit result holds

$$n\left(\hat{\rho}_{n,c} - 1\right) \Rightarrow \frac{\displaystyle\int_0^1 W(r)dW(r)}{\displaystyle\int_0^1 W^2(r)dr}. \tag{2.7}$$

Notice that since $\rho_n = \left(1 + \frac{c}{n^\gamma}\right)$ under the stationary condition which implies that $0 < \rho_n < 1$ then it holds that $-n^\gamma < c < 0$. Therefore, the OLS estimate $\hat{\rho}$ is $n-$consistent, that is, $n\left(\hat{\rho} - \rho\right)$ has a nondegenerate limit distribution depending on c, while $\hat{\mu}$ is $\sqrt{n}-$asymptotically normal. As a result, $\hat{c} = -n(1-\hat{\rho})$ is a natural OLS estimate of the coefficient of persistence c, which is not consistent. Furthermore, Mikusheva (2012) shows that

$$(\hat{c} - c) = \frac{\displaystyle\int_0^1 J_c(r)dW(r)}{\displaystyle\int_0^1 J_c^2(r)dr}. \tag{2.8}$$

The particular asymptotic result demonstrates the well-known conjecture that the OLS estimate of the nuisance parameter of persistence, c, is not consistent. In fact, $\hat{c}$ is asymptotically highly biased to the left, thus the estimated model looks more stationary that it actually is.

In general, the assumption regarding the dependence structure for the disturbance term can affect the limit theory of estimators. In particular, serial correlation in the errors induces an asymptotic bias for $\hat{\rho}_n$ and contributes to the bias of the Gaussian limiting distribution. For the asymptotic theory results of the paper we assume that $\mathcal{D}[0,1]$ is endowed with the Skorokhod topology such that any partial sum processes are measurable for the associated Borel $\sigma-$algebra under the absence of serial correlation. Furthermore, imposing assumptions regarding the properties of the disturbance term ε_t can imply different asymptotic behaviour for estimators given certain modelling conditions. According to Werker and Zhou (2022) the usual procedures established in the literature thus far are based on the assumption of Gaussian innovations and, while their validity has been established under weak assumptions, the asymptotic power of all these procedures cannot go beyond the Gaussian power envelope. Therefore, the asymptotic theory of this paper is established based on optimal test statistics due to the Gaussianity assumption of the innovation processes. Relaxing the Gaussianity assumption requires to apply semiparametric estimation methodologies that is beyond the scope of our study. Thus, the *i.i.d* innovation sequence assumption with finite variance, simplifies the representation of the necessary regularity conditions. Considering the limit theory based on the innovation sequence being a stationary time series under weak dependence, requires further regularity conditions especially when the aim is to derive bounds and corresponding convergence rates based on Berry-Esseen theorems (see, Jirak (2016) and Lahiri and Sun (2009)). Further studies of the proposed econometric environment with the use of moderate deviation principles are presented by Penda et al. (2014) and Proïa (2020).

2.1. Quantile Conditional Estimation Methodology

In this Section we present a more detailed review of the estimation procedure for the quantile autoregressive time series that accommodates the specification of the autocorrelation coefficient with respect to moderate deviations from the unit boundary. We first introduce the quantile estimation method[4] to obtain parameter estimates and then establish the asymptotic theory of this estimator. Denote with $\mu(\tau)$ and $\rho_n(\tau)$ to be the $\tau-$quantile dependent parameters, which are determined based on a conditional quantile specification function

$$\mathsf{Q}_{y_t}\left(\tau|\mathcal{F}_{t-1}\right) := F^{-1}_{y_t|\boldsymbol{x}_{t-1}}(\tau) \equiv \mu(\tau) + \rho_c(\tau) y_{t-1}. \tag{2.9}$$

such that

$$F_{y_t|\boldsymbol{x}_{t-1}}(\tau) := \mathbb{P}\Big(y_t \leq \mathsf{Q}_{y_t}\left(\tau|\mathcal{F}_{t-1}\right) |\mathcal{F}_{t-1}\Big) \equiv \tau. \tag{2.10}$$

for some $\tau \in (0,1)$, where τ denotes the quantile level within a compact set $(0,1)$.

Denote the parameter vector with $\boldsymbol{\vartheta}(\tau) = \big(\mu(\tau), \rho_c(\tau)\big)^\top$ and $\boldsymbol{X}_t = \boldsymbol{D}_n^{-1}(1, y_{t-1})$, where $\boldsymbol{D}_n$ is the normalization matrix which includes the different convergence rates for the model intercept vis-a-vis the slope coefficient. Then, from Koenker and Bassett (1978) and Koenker and Portnoy (1987) the quantile regression estimator is obtained via the following optimization function

$$\widehat{\boldsymbol{\vartheta}}_n(\tau) := \underset{\boldsymbol{\vartheta}(\tau)}{\mathsf{arg\,min}} \sum_{t=1}^{n} \varrho_\tau \Big(y_t - \boldsymbol{\vartheta}(\tau)^\top \boldsymbol{X}_t \Big). \tag{2.11}$$

such that $\widehat{\boldsymbol{\vartheta}}_n(\tau) \equiv \boldsymbol{D}_n\left(\hat{\mu}_n(\tau) - \mu(\tau), \hat{\rho}_{n,c}(\tau) - \rho_c(\tau)\right)$.

[4] A complete treatment of limit results for quantile regressions can be found in the book of Koenker (2005).

Moreover, we denote with $\psi(\mathsf{u})$ to be the left derivative of $\varrho(\mathsf{u})$. In particular, when $\psi(\mathsf{u}) := \mathsf{u}$ is the identity function then the $\widehat{\boldsymbol{\vartheta}}_n$ corresponds to the least squares estimator, while when $\psi(\mathsf{u}) := (1-\tau)$ for $\mathsf{u} \leq 0$ and $\psi(\mathsf{u}) = \tau$ for $\mathsf{u} > 0$, then it corresponds to the quantile regression optimization function and $\widehat{\boldsymbol{\vartheta}}_n$ is the quantile-dependent estimator, the main focus of our paper.

Assumption 2.3. *Suppose that $\mathbb{E}\big[\psi(u_t(\tau))\big] = 0$ and consider the random variable which corresponds to the first derivative around some parameter $\theta \in \mathbb{R}$ such that*

$$\xi := \left| \frac{\partial}{\partial \theta} \mathbb{E}\bigg[\psi\Big(u_1(\tau) - \theta \Big) \bigg] \right|_{\theta = 0}, \quad \text{where } \xi \neq 0. \tag{2.12}$$

where $\mathbb{E}|\psi\left(u_1(\tau)\right)|^{2+m} < \infty$, for some $m > 0$.

Assumption 2.3 ensures that the first derivative is Lipschitz continuous and bounded which corresponds to the first derivative for the expectation of the check function as a random variable evaluated within the neighbourhood of the true parameter vector $\theta = 0$. Furthermore, due to the fact that the quantile autoregressive model we consider in this paper corresponds to a possibly nonstationary time series model, then the asymptotic theory of estimators and corresponding test statistics depends on Brownian motion functionals as introduced with Assumption 2.4 below.

Assumption 2.4. *The following conditions for the innovation sequence hold:*

(i) *The sequence of stationary conditional probability distribution functions denoted with $\{f_{\varepsilon_t(\tau),t-1}(.)\}$ evaluated at zero with a non-degenerate mean function such that $f_{\varepsilon_t(\tau)}(0) := \mathbb{E}\left[f_{\varepsilon_t(\tau),t-1}(0)\right] > 0$ satisfies a Functional Central Limit Theorem (FCLT) expressed as below*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} \Big(f_{\varepsilon_t(\tau),t-1}(0) - \mathbb{E}\left[f_{\varepsilon_t(\tau),t-1}(0) \right] \Big) \Rightarrow B_{f_{\varepsilon_t(\tau)}}(r), \ \text{with } r \in (0,1). \tag{2.13}$$

(ii) *For each t and $\tau \in (0,1)$, $f_{\varepsilon_t(\tau),t-1}(.)$ is uniformly bounded away from zero with a corresponding conditional distribution function $F_t(.)$ which is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$ (see, Neocleous and Portnoy (2008), Goh and Knight (2009) and Lee (2016)).*

Assumption 2.4 gives necessary and sufficient conditions for a functional central limit theorem to hold for the corresponding innovation sequence based on the conditional quantile functional form specification, which is instrumental for deriving the asymptotic behaviour of the quantile-dependent estimators under nonstationarity based on Brownian motion functionals. Therefore, to obtain the model estimates based on the optimization problem (2.11) we apply the Taylor expansion to the check function, such that for a given parameter $\boldsymbol{\delta}(\tau)$ it holds that

$$\varrho_\tau\Big(\varepsilon_t - \boldsymbol{\delta}(\tau)^\top \boldsymbol{X}_t \Big) = \varrho_\tau(\varepsilon_t) - \boldsymbol{\delta}(\tau)^\top \psi(\varepsilon_t) + \varphi_t(\boldsymbol{\delta}(\tau)). \tag{2.14}$$

Remark 2.2. *Notice that for instance the t-ratio for $\rho_{n,c}$ is defined by $\sqrt{\sum_{t=1}^n y_{t-1}^2}\,(\hat{\rho}_{n,c} - \rho_c)$, thus to obtain the limiting distribution of the $t-$test we need to obtain an asymptotic expression for the normalized centered estimator $(\hat{\rho}_{n,c} - \rho_c)$. Furthermore, for sequences such that $\lim_{n\to\infty} n\left(1-\rho_n\right) = 0$, the nearly unstable model behaves asymptotically like the strictly unstable model in which case $\rho = 1$.*

3. Asymptotic Theory

In this section we present the main asymptotic theory results while detailed proofs can be found in the Appendix of the paper. Some important aspects worth emphasizing again is that while the case of near-integrated (NI) processes, such that $c < 0$ and $\gamma = 1$, has been considered before in quantile autoregressive time series (see, Chan et al. (2006)) as well as the case of mildly integrated (MI) such that $c < 0$ and $\gamma \in (0,1)$, the mildly explosive (ME) such that $c > 0$ and $\gamma \in (0,1)$ and the explosive case, such that $c > 0$ and $\gamma = 1$ has not been widely explored before. In particular, all aforementioned cases which correspond to different regions of the parameter space, overcome the singularity problem; the region that the underline stochastic process does not have a solution. More specifically, using the local-to-unity specification in autoregressive processes helps to overcome this gap by considering the limiting distribution for the whole parameter space regardless of the existence of a limit singularity.

Firstly, using the local-to-unity asymptotics our aim is to derive a nuisance-parameter-free limiting distribution which can facilitate statistical inference. In particular, the form of the noncentrality parameter of the χ^2-distribution which is the limiting distribution of test statistics can depend on the initial condition[5] and the form of the autoregressive parameter of the model (see, Theorem 2.2. of Hui et al. (2022)). Secondly, by decomposing the underline stochastic processes into components which include a predictable quadratic variation (see, Magdalinos and Phillips (2009)), allows us to obtain a self-normalized martingale sequence, which is especially useful when deriving the limiting distribution of Wald-type statistics. In other words, Wald statistics constructed with a variance estimator which induced by the predictable quadratic variation ensures that the self-normalization property holds.

Theorem 3.1 (Chan et al. (2006))**.** *Assume that Assumption 2.1-2.4 hold and that the autocorrelation coefficient is expressed as $\rho_{n,c} = \left(1 + \frac{c}{k_n}\right)$. Then, the following limit result hold*

$$\boldsymbol{D}_n \left(\hat{\boldsymbol{\vartheta}}_n(\tau) - \boldsymbol{\vartheta}(\tau) \right) \overset{d}{\to} \frac{1}{f_\varepsilon(F_\varepsilon^{-1}(\varepsilon_t(\tau)))} \Sigma^{-1} \left(W(\tau;1), \int_0^1 J(s) dW(\tau,s) \right)', \tag{3.1}$$

where $\boldsymbol{D}_n = \mathsf{diag}\left(\sqrt{n}, n\right)$ and $\boldsymbol{\vartheta}(\tau) = \left(\mu(\tau), \rho_{n,c}(\tau)\right)$ such that

$$\boldsymbol{\Sigma} := \int_0^1 (1, J_1(s))'(1, J_1(s)) ds \equiv \begin{bmatrix} 1 & \displaystyle\int_0^1 J_1(s) ds \\ \displaystyle\int_0^1 J_1(s)' ds & \displaystyle\int_0^1 J_1(s) J_1(s)' ds \end{bmatrix}. \tag{3.2}$$

Furthermore, it holds that

$$n \Big(\hat{\rho}_{n,c}(\tau) - \rho_{n,c}(\tau) \Big) \overset{d}{\to} \frac{1}{f_\varepsilon(F_\varepsilon^{-1}(\varepsilon_t(\tau)))} \frac{\displaystyle\int_0^1 J_1(s) dW(\tau,s) - W(\tau,1) \int_0^1 J_1(s) ds}{\displaystyle\int_0^1 J_1^2(s) ds - \left(\int_0^1 J_1(s) ds \right)^2} \tag{3.3}$$

and with a suitable normalization it follows that

$$\left\{ \sum_{t=1}^n y_{t-1}^2 - \left(\sum_{t=1}^n y_{t-1} \right)^2 \right\}^{1/2} \Big(\hat{\rho}_{n,c}(\tau) - \rho_{n,c}(\tau) \Big) \overset{d}{\to} \mathcal{N} \left(0, \frac{\tau(1-\tau)}{f_\varepsilon^2(F_\varepsilon^{-1}(\varepsilon_t(\tau)))} \right). \tag{3.4}$$

[5]Notice that since the initial condition of the autoregressive process corresponds to the boundary condition of an ordinary differential equation problem, then the stochastic solution, that is, the limiting distribution under equilibrium conditions will depend on the initial condition (see, Saxena and Alam (1982)).

3.1. Limit theory for near-stationary case

In a similar spirit as in the framework proposed by Phillips and Magdalinos (2007), in order to establish the limit theory of the autocorrelation coefficient within our modelling environment, we consider the asymptotic behaviour of the sample moments that appear in the quantile-dependent estimator separately. However, in contrast to the ordinary least squares estimation, when the model parameters are estimated using the conditional quantile specification function, we employ standard approximation methods (such as the Bahadur representation) from the quantile regression literature to obtain analytical expressions for the quantities of interest. Specifically, in the near-stationary case, which implies that $c < 0$, the limit to the unit boundary is approached from the left of the triangular array. Furthermore, due to the different convergences rate of the model intercept versus the slope parameter we employ the normalization matrices $\boldsymbol{D}_n$ and $\boldsymbol{B}$ as defined below

$$\boldsymbol{D}_n = \begin{pmatrix} \sqrt{n} & 0 \\ 0 & \sqrt{nk_n} \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2/(-2c) \end{pmatrix} \tag{3.5}$$

where $k_n = n^{\gamma}$ and $\gamma \in (0,1)$.

Remark 3.1. *The $n^{-1/2}$ convergence rate corresponds to the model intercept while when $k_n = n^{\gamma}$, then the autocorrelation parameter of the model has a convergence rate of $n^{-\frac{1+\gamma}{2}}$ which is also the rate of convergence that corresponds to a mildly integrated process. Furthermore, in empirical applications in practise we do not know a prior whether the expression $\sqrt{n}\,(\hat{\rho}_n - \rho)$ is positive or negative. However, since we do not partition the parameter space accordingly, the asymptotic theory mainly focus on the near-integrated case and does not cover the mildly explosive or pure explosive since $c < 0$.*

Theorem 3.2. *Under Assumptions 2.2-2.4,*

$$\left(\hat{\mu}_n, \hat{\rho}_{n,c}\right)^{\top} = \left(\mu, \rho_{n,c}\right)^{\top} + \frac{(\boldsymbol{B}\boldsymbol{D}_n)^{-1}}{\xi} \sum_{t=1}^{n} \psi(\varepsilon_t) \boldsymbol{X}_t^{\top} + o_p(1). \tag{3.6}$$

In particular, when $\psi(\mathsf{u}) = \left(\tau - \mathbf{1}\left\{\mathsf{u} \leq 0\right\}\right)$ corresponds to the quantile regression and therefore the above expression reduces to

$$\begin{pmatrix} \hat{\mu}_n(\tau) \\ \hat{\rho}_{c,n}(\tau) \end{pmatrix} = \begin{pmatrix} \mu_n(\tau) \\ \rho_{c,n}(\tau) \end{pmatrix} + \frac{(\boldsymbol{B}\boldsymbol{D}_n)^{-1}}{f_{\varepsilon}\left(F_{\varepsilon}^{-1}(\tau)\right)} \sum_{t=1}^{n} \left(\tau - \mathbf{1}\left\{\varepsilon_t \leq F_{\varepsilon}^{-1}(\tau)\right\}\right) \begin{pmatrix} \dfrac{1}{\sqrt{n}} \\ \dfrac{y_{t-1}}{\sqrt{nk_n}} \end{pmatrix} + o_P(1). \tag{3.7}$$

where $f_{\varepsilon}(x)$ and $F_{\varepsilon}(x)$ denote the probability and cumulative density functions of ε_1, respectively.

Theorem 3.2 provides a Bahadur representation for the parameter vector of the quantile autoregressive time series which includes a model intercept and a slope. In particular for M-regressions a necessary requirement for the functional form is to include a model intercept which can be different than zero. Moreover, the given limit results are employed to derive the asymptotic behaviour of model parameters based on moderate deviations from the unit boundary on the stationary region as summarized by the next theorem. Then, the robust estimation of the sparsity coefficient which depends on the kernel density function can be improve the accuracy of the quantile-dependent model estimates.

Theorem 3.3. *Under Assumptions 2.2-2.4,*

$$\boldsymbol{D}_n\left((\hat{\mu}_n, \hat{\rho}_{n,c}) - (\mu_n, \rho_{n,c})\right) \overset{d}{\to} \mathcal{N}\left(0, \frac{\boldsymbol{B}^{-1} \times \mathbb{E}[\psi^2(\varepsilon_1)]}{\xi}\right). \tag{3.8}$$

In particular, it follows that

(i) If $\psi(\mathsf{u}) = (\tau - \mathbf{1}\{\mathsf{u} \leq 0\})$*, and the pdf* $f(\mathsf{u})$ *of* ε_1 *exists and satisfies* $f_\varepsilon(F_\varepsilon^{-1}(\tau)) > 0$*, then*

$$\frac{\hat{\rho}_{n,c}(\tau) - \rho_c(\tau)}{\sqrt{nk_n}} \overset{d}{\to} \mathcal{N}\left(0, \frac{-2c}{\sigma^2}\frac{\tau(1-\tau)}{f_\varepsilon^2(F_\varepsilon^{-1}(\tau))}\right). \tag{3.9}$$

(ii) If $\psi(\mathsf{u}) = \mathsf{u}$*, then*

$$\frac{\hat{\rho}_{n,c}(\tau) - \rho_c(\tau)}{\sqrt{nk_n}} \overset{d}{\to} \mathcal{N}\left(0, -2c\right). \tag{3.10}$$

Remark 3.2. *The limit results given by Theorem 3.2 summarize the joint asymptotic behaviour of the model intercept and slope from moderate deviations from the unit boundary on the stationary region. In order to prove the above asymptotic results, we employ standard arguments introduced by Pollard (1991) for optimization of convex function relevant to quantile processes. Specifically, by the convexity lemma, if the finite-dimensional distributions of* $\Omega_n(v)$ *converge weakly to those of* $\Omega(v)$*, and* $\Omega(v)$ *has a unique minimum, then the convexity of* $\Omega_n(v)$ *implies that* $\hat{v}$ *converges in distribution to the minimizer of* $\Omega(v)$*. In other words,* $\hat{\rho}_n(\tau)$ *is shown to be weakly consistent, thus to prove that the estimator is asymptotically normally distributed we restrict the spaces* $\mathcal{B}$ *to shrinking neighbourhoods around the true value of the parameter* $\rho(\tau)$ *in order to avoid possible local minima. To do this, we can define the restricted space* $\mathcal{B}_a = \{\rho_n \in \mathcal{B} : \|\beta - \beta(\tau)\| \leq a_n\}$ *where* a_n *some positive sequence.*

3.2. Limit theory for near-explosive case

The near-explosive case corresponds to the nuisance parameters of persistence $c > 0$ and the exponent rate $\gamma \in (0,1)$ or $\gamma = 1$. In particular for the linear autoregressive process $y_t = \theta y_{t-1} + \varepsilon_t$ and an explosive root such that $|\theta| > 1$, a Cauchy limit theory can be derived for the OLS estimator $\hat{\theta}$ as

$$\frac{\theta^n}{\theta^2 - 1}\left(\hat{\theta}_n - \theta\right) \Rightarrow \mathcal{C}, \quad \text{as } n \to \infty. \tag{3.11}$$

More precisely, the seminal study of Anderson (1959) provides examples demonstrating that central limit theory does not apply and the asymptotic distribution of the least squares estimator depends by the distributional assumptions imposed on the innovations which makes inference procedures specifically for purely explosive autoregressions more challenging (see, Magdalinos (2012)). Furthermore, in this direction, Phillips and Magdalinos (2007) consider autoregressive processes under the moderate deviations framework by employing the local-to-unity specification for the autoregression coefficient such that $\theta_n = \left(1 + \frac{c}{n^\gamma}\right), \gamma \in (0,1)$. Therefore, in this case under the assumption of *i.i.d* innovations with finite second moments the following least squares regression theory was proved

$$\frac{1}{2c} n^\gamma \theta_n^n \left(\hat{\theta}_n - \theta\right) \Rightarrow \mathcal{C}, \quad \text{as } n \to \infty. \tag{3.12}$$

On the other hand, for the pure explosive root case such that $|\theta| > 1$ then, the limit distribution of $(\hat{\theta} - \theta)$ is standard Cauchy if it is normalized with $\theta^n/(1-\theta^2)$. However, the limit distribution depends on the distribution of the noise, as was pointed out by Anderson (1959), and hence no central limit theorem applies on the explosive side. Moreover, from empirical data financial applications it can be observed that the parameter θ tends to 1 with increasing sample size. To accommodate this observation, $\theta = \theta_n$ is allowed to depend on n, the number of observations, such that $\theta_n \to 1$ as $n \to \infty$. The process is then referred to as near-integrated. Depending on whether $\theta_n < 1$ or $\theta_n > 1$, it is called near-stationary or mildly explosive. Furthermore, Phillips and Magdalinos (2007) investigated the general parameter case in the near-integrated setting assuming that $\theta_n \to 1$ with a rate slower than $1/n$, the so-called moderate deviations from unity. All aforementioned approaches operate under the assumption of a finite variance along with independent, identically distributed or weakly dependent errors. However, it can be proved that the serial coefficient $\hat{\theta}_n - \theta_n$ has, under a suitable normalization, a limit that consists of a fraction of two independent strictly stable random variables. Therefore, specifically for the quantile autoregressive time series model we consider in our study we employ the following normalization matrices.

$$\boldsymbol{D}_n = \begin{pmatrix} \sqrt{n} & 0 \\ 0 & \rho_{n,c}^n k_n \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 1 & 0 \\ 0 & \mathcal{Z}_3^2/(2c) \end{pmatrix} \tag{3.13}$$

where $\mathcal{Z}_3$ is some normal random variable to be defined below.

Theorem 3.4. *Under the Assumptions 2.1-2.4 it holds that,*

$$(\hat{\mu}, \hat{\rho}_{n,c})^\top = (\mu, \rho_{n,c})^\top + \frac{\left(\sum_{t=1}^n \boldsymbol{X}_t \boldsymbol{X}_t^\top \boldsymbol{D}_n\right)^{-1}}{\xi} \sum_{t=1}^n \psi(\varepsilon_t) \boldsymbol{X}_t^\top + o_p(1). \tag{3.14}$$

Theorem 3.5. *Under Assumptions 2.1-2.4 it holds that,*

$$\boldsymbol{D}_n \big((\hat{\mu}, \hat{\rho}_{n,c}) - (\mu, \rho_{n,c}) \big)^\top \overset{d}{\to} \frac{1}{\xi} \boldsymbol{B}_n^{-1} (\mathcal{Z}_1, \mathcal{Z}_2 \mathcal{Z}_3)^\top. \tag{3.15}$$

In particular, it follows that

(i) If $\psi(\mathsf{u}) = (\tau - \mathbf{1}\{\mathsf{u} \leq 0\})$, and the pdf $f(\mathsf{u})$ of ε_1 exists and satisfies $f_\varepsilon(F_\varepsilon^{-1}(\tau)) > 0$, then

$$\frac{\hat{\rho}_{n,c}(\tau) - \rho_c(\tau)}{\rho_n k_n} \overset{d}{\to} \frac{2c}{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)} \frac{\mathcal{Z}_2}{\mathcal{Z}_3}. \tag{3.16}$$

(ii) If $\psi(\mathsf{u}) = \mathsf{u}$, then

$$\frac{\hat{\rho}_{n,c}(\tau) - \rho_c(\tau)}{2c\rho_n^n k_n} \overset{d}{\to} \frac{\mathcal{Z}_2^*}{\mathcal{Z}_3^*}. \tag{3.17}$$

Therefore, Theorem 3.5 verifies that we indeed obtain the equivalent asymptotic theory results in comparison to the linear autoregressive time series model. Specifically, the autocorrelation coefficient of the nonstationary quantile autoregressive time series model converge into a Cauchy random variate in the case of mildly explosive processes (see, also Aue and Horváth (2007), Phillips and Magdalinos (2007), Magdalinos (2012) and Lee (2018)).

Remark 3.3. *As we see from Theorem 3.5, part 2, the limiting distribution of the normalized and centered estimator is Cauchy, similar to Theorem 4.3 of Phillips and Magdalinos (2007). As a matter of fact when we replace ρ_n by $\rho_{n,c} = \left(1 + \frac{c}{k_n}\right)$ we obtain that $(\rho^2 - 1) = \frac{2c}{k_n}\big[1 + o(1)\big]$. Hence, we see that the normalizations in the Theorem above and the expression derived by White (1958) are asymptotically equivalent as $n \to \infty$. Furthermore, the asymptotic theory for the case of moderate deviations from the unity boundary is not restricted to Gaussian processes. More specifically, the Cauchy limit result applied for $\rho_{n,c} = \left(1 + \frac{c}{k_n}\right)$ and innovations ε_t with finite second moment (e.g., innovations with stable law of attraction). On the other hand, the main difference between the mildly explosive processes given by Theorem 3.5 above and explosive autoregressions with $|\rho| > 1$, occurs due to the different convergence rates of these two cases. In particular, in the case of mildly explosive processes we define the convergence rate such that $k_n = n^{\gamma}$ for some $\gamma \in (0,1)$ while for the case of moderately explosive processes we define with $k_n = n^{\gamma}$ for some $\gamma > 1$.*

3.3. Testing Linear Hypotheses

Consider the autoregressive model

$$y_t = \rho y_{t-1} + \varepsilon_t, \qquad t = 1, ..., n, \tag{3.18}$$

such that $\rho \in [-1, 1]$, is within the stationary region. Then, the usual testing hypothesis of interest is such that, $\mathbb{H}_0 : \rho = \rho_0$. In particular, Dickey and Fuller (1979) showed that the finite sample distribution for ρ in the neighbourhood of unity is very close to the asymptotic unit root case, under the assumption that the error term ϵ_t is normally distributed with finite variance.

Statistical inference for M-estimators of possibly nonstationary time series models (local-to-unit root) is a nonstandard problem due to the presence of nuisance parameters in the limiting distributions of test statistics. However, indeed one of the advantages of M-estimators is that are considered to be robust to outliers since they have a bounded influence function (see, Abadir and Lucas (2000)). Considering now specifically the case of unit root such that $|\rho| = 1$, the asymptotic distribution of the t-statistic denoted by $\mathcal{T}_n(\hat{\rho})$ can be represented by functionals of Wiener processes (see, Dickey and Fuller (1979) and Buchmann and Chan (2007)). Furthermore, the asymptotic distribution of the $t-$statistic based on $M-$estimators, denoted by $\mathcal{T}_{\psi(\tau)}(\hat{\rho})$ depends on the nuisance parameter δ, that is, the correlation between the innovations $\{\epsilon_t\}$ and the pseudo-score function $\psi(\epsilon_t)$ that is employed to define the $M-$estimator. On the other hand, $t-$statistics based on $M-$estimators lead to a reduction in asymptotic MSE relative to LSE for local alternatives to the unit-root null hypothesis[6].

The t-statistic for the null hypothesis $\mathbb{H}_0 : \rho = \rho_0$ is given by

$$t_{\psi} = \frac{\Big(\hat{\rho}_n(\tau) - \rho_n(\tau)\Big)}{\left\{ \left(n^{-1} \sum_{t=1}^{n} \psi_{\tau}(y_t - \hat{\rho}_n(\tau) y_{t-1})^2 \right) \Big/ \left(n^{-1} \sum_{t=1}^{n} \psi'_{\tau}(y_t - \hat{\rho}_n(\tau) y_{t-1}) \right)^2 \right\}^{1/2}}. \tag{3.19}$$

where $\psi'_{\tau}(.)$ denotes the first derivative of the function $\psi_{\tau}(.)$.

[6]Specifically, Magdalinos (2007) consider the approximate Bahadur slopes of test statistics by examining their asymptotic behaviour under the alternative hypothesis.

4. Monte Carlo Simulations

In this section we investigate the finite-sample performance in terms of standard deviations and coverage probabilities of the quantile estimate and the least squares estimate are computed. We consider the autoregressive time series model

$$y_t = \mu + \rho_n y_{t-1} + \varepsilon_t, \quad t = 1, ..., n \tag{4.1}$$

where $\varepsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0,1)$ and μ and ρ the model parameters to be estimated. In particular, since the autocorrelation coefficient is defined as $\rho_{n,c} = \left(1 + \frac{c}{n}\right)$, by definition of the nonstationary autoregressive time series models, we begin by comparing the the performance of the OLS-based estimator and the QR-based estimator for $\tau = 0.5$. The least squares estimate is given by

$$\hat{\rho}^{ols}_{n,c} = \frac{\displaystyle\sum_{t=1}^{n} y_t y_{t-1}}{\displaystyle\sum_{t=1}^{n} y_t^2} \tag{4.2}$$

Therefore, we verify the simulation results presented in the study of Chan et al. (2006) who show that the sample averages of the estimates and the corresponding standard deviations of $\hat{\rho}^{qr}_{n,c}(\tau)$ in comparison to $\hat{\rho}^{ols}_{n,c}$ indicate to perform better in terms of precision and standard errors. In particular, this improvement is shown to be more apparent for finite-samples. To demonstrate this through simulations, we estimate the empirical size of the $t-$test; the size of the test statistic obtained from the simulated finite sample distribution and then compare with the nominal size; the size that would be expected if the finite sample distribution is perfectly approximated by the limiting distribution (significance level $\alpha = 0.05$). The empirical size under the null hypothesis is presented below.

Table goes here.

5. An Empirical Application

We employ the R package developed by Koenker (2012) in order to obtain estimates for the quantile autoregressive model. Furthermore, we compare the time series behaviour of two cryptocurrencies which are considered to be among the 10 largest cryptocurrencies with respect to market capitalization. We focus on the sampling period 01 January 2018 to 20 March 2022. In particular, several papers in the literature employ the framework proposed by Phillips et al. (2015) to detect bubbles in cryptocurrencies which verify that the bubble periods coincide with major events that affected the Bitcoin market. In this paper we focus on comparing the performance of the quantile autoregressive model with respect to the linear autoregressive time series model as well as the testing for predictability using the Student's $t-$statistic applied to the autocorrelation coefficient of the model.

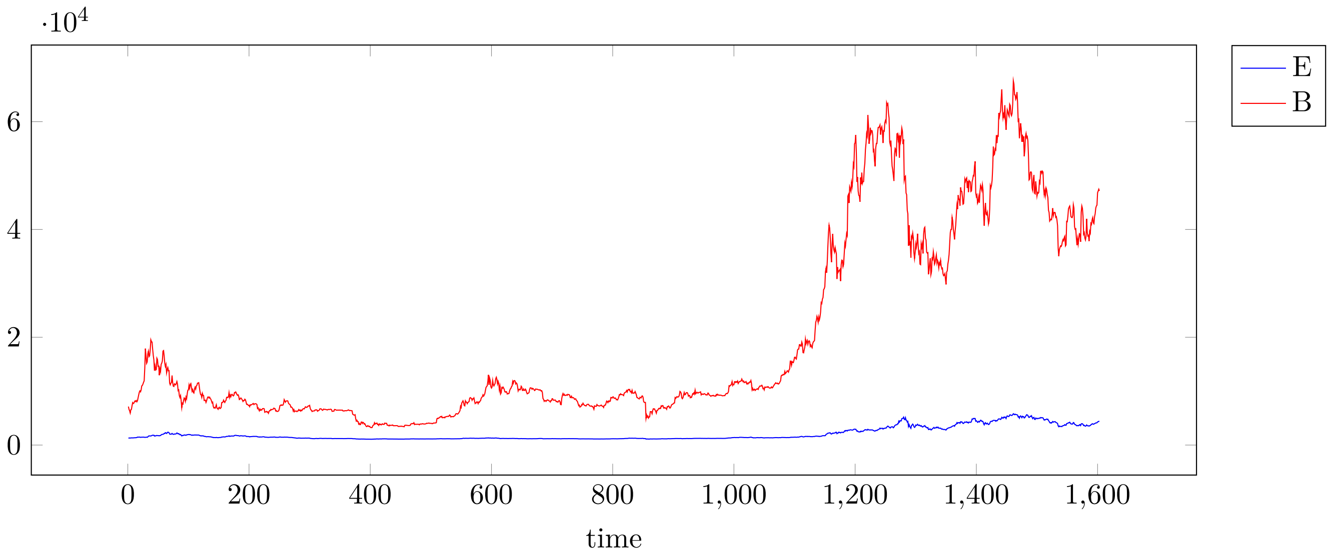

Figure 1 above shows the adjusted closing price (in US dollars) of the cryptocurrencies Bitcoin and Ethereum for 1603 trading days between 01 January 2018 to 20 March 2022.

Table 1 shows the model estimates for the two time series of cryptocurrencies Etherum and Bitcoin that we implement the quantile autoregressive model. To begin with, the linear autoregressive model when with no model intercept has a statistical significant and explosive autocorrelation coefficient while similar results are obtained when we consider the quantile autoregressive model around the median ($\tau = 0.5$), which under the Gaussian innovation assumption should be equivalent to the linear autoregressive model. On the other hand, when considering the quantile autoregressive model at the upper and lower tails of the underline distribution (e.g., $\tau = 0.05$ or $\tau = 0.95$) then moderate deviations from the unit root on the explosive side are more apparent especially when the model is fitted at the $\tau = 0.95$ of the distribution. Therefore, overall there are statistical significant evidence of explosive behaviour in the time series of cryptocurrencies at the tails of the distribution.

Table 1: Estimation Results

		Etherum					
		$\hat{\theta}$	$s.e(\hat{\theta})$	$t-$statistic	$\hat{\theta}$	$s.e(\hat{\theta})$	$t-$statistic
linear	μ				2.9759	2.5803	1.1530
	ρ	1.0002	0.0012	813.6***	0.9990	0.0016	623.3***
QR ($\tau = 0.05$)	μ				0.7857	1.7359	0.4527
	ρ	0.9222	0.0049	188.5***	0.9208	0.0066	140.5***
QR ($\tau = 0.5$)	μ				-0.3981	0.6437	-0.6184
	ρ	1.0023	0.0024	424.5***	1.0029	0.0029	342.5***
QR ($\tau = 0.95$)	μ				0.7496	2.4964	0.3003
	ρ	1.0845	0.0054	200.9***	1.0840	0.0071	153.5***
		Bitcoin					
		$\hat{\theta}$	$s.e(\hat{\theta})$	$t-$statistic	$\hat{\theta}$	$s.e(\hat{\theta})$	$t-$statistic
linear	μ				45.8416	38.1999	1.2000
	ρ	1.0002	0.0010	1021***	0.9989	0.0014	689.2***
QR ($\tau = 0.5$)	μ				29.6022	41.0729	0.7207
	ρ	0.9372	0.0035	270.6***	0.9350	0.0050	186.6***
QR ($\tau = 0.5$)	μ				9.9114	13.6279	0.7273
	ρ	1.0012	0.0014	741.9***	1.0002	0.0020	495.5***
QR ($\tau = 0.5$)	μ				45.4248	66.6845	0.6812
	ρ	1.0707	0.0041	260.1***	1.0679	0.0059	181.4***

Table 1 shows the model estimates for both the linear and quantile autoregressive time series model given by $y_t = \mu + \rho y_{t-1} + \varepsilon_t$, with $t = 1, ..., n$, based on the adjusted closing price (in US dollars) of the cryptocurrencies Bitcoin and Ethereum for 1603 trading days between 01 January 2018 to 20 March 2022.

Remark 5.1. *Notice that for the quantile autoregressive time series the model estimates are obtained using a bootstrap estimation procedure implemented as a build-in option of the R package* `quantreg`*.*

```
tau <- 0.05

### Model 1: Etherum
model.QR_etherum <- rq( e_t ~ e_lag, tau = tau )
model.summary    <- summary( model.QR_etherum, se = "boot", bsmethod= "xy" )

### Model 2: Bitcoin
model.QR_bitcoin <- rq( b_t  ~ b_lag, tau = tau )
model.summary    <- summary( model.QR_bitcoin, se = "boot", bsmethod= "xy" )
```

6. Conclusion

In this paper we consider the asymptotic theory for moderate deviation from the unit boundary in quantile autoregressive time series models. The moderate deviation principles provide a framework for unifying the asymptotic theory by investigating the limiting distribution discontinuities at certain regions of the parameter space. Specifically, in this study we verify the limit results obtained by Phillips and Magdalinos (2007) in the case of the linear autoregressive time series model. In particular, for both the case of near-stationary and near-explosive roots we establish the asymptotic theory of the quantile-dependent estimator which converges into a nuisance-parameter free limiting distribution.

Furthermore, although we do not consider how the presence of serial correlation can affect the limiting distributions under consideration, various studies in the literature consider simple implementations of autocorrelation robust tests such as in Jansson (2004) (see, also Vogelsang (1998)). In particular, a good example is the OLS estimator in a linear regression model with exogenous regressors and an autocorrelated error term. Similar implementations can be considered within the modelling environment of quantile regression models especially of those with possibly nonstationary autoregressive processes with serial correlated innovation terms. In particular, Kiefer et al. (2000), (KVB), demonstrate that the properties of Wald-type statistics can be ameliorated if an inconsistent covariance matrix estimator is used and the critical values are adjusted to accommodate the randomness of the matrix employed in the standardization. Using higher-order asymptotic theory, Jansson (2004) provides an analytical explanation of the encouraging performance of the KVB procedure. An extension of our framework in the regions which unifies all cases such as being in the unstable region with nearly stable or unstable processes such as the explosive and pure explosive processes, is an aspect of ongoing research that the author is actively undertaking. Further research aspects worth mentioning include the investigation of the asymptotic behaviour of quantile autoregressive models when a structural break occurs at an unknown break-point location. A relevant study using moderate deviations principles when testing for structural breaks include the framework proposed by Xu and Pang (2018). Other aspects include comparisons of the performance of $t-$tests with likelihood-based tests as in Nielsen (2001).

Acknowledgements I wish to thank the supervisory committee for helpful discussions and encouragement throughout the PhD programme. All remaining errors are mine.

Funding Financial support from the Vice Chancellor's PhD studentship of the University of Southampton is gratefully acknowledged.

A. Mathematical Appendix

Recall that an $M-$estimator $(\hat{\mu}, \hat{\rho})$ of the parameter vector (μ, ρ) is a minimizer for $\mu \in \mathbb{R}$ and $\rho \in \mathbb{R}$ such that

(A.1)
$$R_n(\mu, \rho) = \frac{1}{n} \sum_{t=1}^{n} \varrho_\tau \left(y_t - \mu - \rho_c x_{t-1} \right).$$

The formulation of the above expression covers the least squares regression when $\rho_\tau(\mathsf{u}) = \mathsf{u}/2$ and the quantile regression such that $\rho_\tau(\mathsf{u}) = \{\tau \mathsf{u} \mathbf{1}(\mathsf{u} > 0) - (1-\tau)\mathsf{u}\mathbf{1}(\mathsf{u} < 0)\}$, for some $0 < \tau < 1$.

A.1. Main Results

A.1.1. Near-Stationary Case $(c < 0)$

Lemma A.1. *Under Assumption 2.2, when $c < 0$ then it holds that*

(A.2)
$$\mathbb{P}\left(\max_{1 \leq t \leq n} y_t^2 \geq \lambda \right) \leq \frac{\mathbb{E}\left[y_n^2 \right]}{\lambda^2}, \ \textit{for some } \lambda > 0.$$

Proof. Notice that Lemma A.1 corresponds to the Kolmogorov, Doob maximal inequality applied to the martingale sequence $(y_n, \mathcal{F}_n)_{n \in \mathbb{N}}$. The proof of Lemma A.1 can be easily obtained by considering the set $\mathcal{S}_s = \{ y_s^2 > \lambda, y_j \leq \lambda, j \leq s \}$ and expanding the expression for the expectation $\mathbb{E}\left[y_n^2 \right]$. The particular result provides a probability bound for the tails of the autoregressive time series which inherits the properties of the stochastic difference equation. □

In addition to Lemma A.1, the following two properties hold:

- If $c < 0$, $\mathbb{E}[y_n^2] = \mathcal{O}(k_n)$.
- $\displaystyle\max_{1 \leq t \leq n} \frac{y_t^2}{n} = o_p(1)$.

Lemma A.2. *Under Assumptions 2.2 - 2.4, it holds that*

(A.3)
$$\sum_{t=1}^{n} \mathbb{E}_{\mathcal{F}_{t-1}} \left[\varphi_{nt}(\boldsymbol{\delta}(\tau)) \right] \overset{p}{\to} \frac{1}{2} \xi \times \boldsymbol{\delta}^\top(\tau) \boldsymbol{B} \boldsymbol{\delta}(\tau).$$

where

(A.4)
$$\xi := \left| \frac{\partial}{\partial \theta} \mathbb{E}\left[\psi\Big(u_1(\tau) - \theta \Big) \right] \right|_{\theta = 0} \quad \textit{and} \quad \boldsymbol{B} = \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2/(-2c) \end{pmatrix}.$$

Proof. By rearranging expression (2.14), taking the conditional expectation and sum over $1 \leq t \leq n$, we obtain the following expression

(A.5)
$$\sum_{t=1}^{n} \mathbb{E}_{\mathcal{F}_{t-1}} \left[\varphi_{nt}(\boldsymbol{\delta}) \right] = \sum_{t=1}^{n} \mathbb{E}_{\mathcal{F}_{t-1}} \left[\varrho_\tau \left(\varepsilon_t - \boldsymbol{\delta}(\tau)^\top \boldsymbol{X}_t \right) - \varrho_\tau(\varepsilon_t) \right].$$

Using Knight (1998)'s identity we have that

$$\varrho_{\tau}(\mathsf{u}_1 - \mathsf{u}_2) - \varrho_{\tau}(\mathsf{u}_1) = \mathsf{u}_2\big(\tau - \mathbf{1}\left\{\mathsf{u}_1 \leq 0\right\}\big) + \mathsf{u}_2 \int_0^1 \Big[\mathbf{1}\{\mathsf{u}_1 \leq \mathsf{u}_2 s\} - \mathbf{1}\{\mathsf{u}_1 \leq 0\}\Big] ds. \tag{A.6}$$

which implies that we can decompose $Z_n(\mathsf{u}, \tau) = Z_n^{(1)}(\mathsf{u}, \tau) + Z_n^{(2)}(\mathsf{u}, \tau)$. Then, it can be proved that

$$\sum_{t=1}^{n} \mathbb{E}_{\mathcal{F}_{t-1}}\Big[\varphi_{nt}(\boldsymbol{\delta})\Big] = \frac{1}{2}\xi \times \boldsymbol{\delta}^{\top} \left(\sum_{t=1}^{n} \boldsymbol{X}_t \boldsymbol{X}_t^{\top} \right) \boldsymbol{\delta} + o_p(1). \tag{A.7}$$

Furthermore, by Theorem 3.2 (a) of Phillips and Magdalinos (2007) it holds that

$$\frac{1}{n k_n} \sum_{t=1}^{n} y_{t-1}^2 \overset{p}{\to} \frac{\sigma^2}{-2c}. \tag{A.8}$$

Therefore, it suffices to show that

$$\frac{1}{n\sqrt{k_n}} \sum_{t=1}^{n} y_{t-1} = o_p(1). \tag{A.9}$$

To see this, we consider the left side of the expression above such that

$$\frac{(1-\rho_n)}{n\sqrt{k_n}(1-\rho_n)} \sum_{t=1}^{n} y_{t-1} = \frac{1}{n\sqrt{k_n}(1-\rho_n)} \sum_{t=1}^{n} (1-\rho_n) y_{t-1} = \frac{1}{n\sqrt{k_n}(1-\rho_n)} \sum_{t=1}^{n} y_{t-1} - \rho_n y_{t-1} \tag{A.10}$$

However, it holds that $y_t = \rho_n y_{t-1} + u_t$, and by rearranging we have that $-\rho_n y_{t-1} = -(y_t - u_t)$. Thus,

$$\begin{aligned} \frac{(1-\rho_n)}{n\sqrt{k_n}(1-\rho_n)} \sum_{t=1}^{n} y_{t-1} &= \frac{1}{n\sqrt{k_n}(1-\rho_n)} \sum_{t=1}^{n} \left[y_{t-1} - (y_t - u_t) \right] \\ &= \frac{1}{n\sqrt{k_n}(c/k_n)} \sum_{t=1}^{n} \left(y_0 - y_n + \sum_{t=1}^{n} u_t \right) = o_p(1). \end{aligned}$$

which shows that $\frac{1}{n\sqrt{k_n}} \sum_{t=1}^{n} y_{t-1} \overset{p}{\to} 0$, converges in probability to zero. ☐

Lemma A.3. *Under Assumptions 2.1-2.4, it holds that*

$$\sum_{t=1}^{n} \varphi_{nt}(\boldsymbol{\delta}(\tau)) \overset{p}{\to} \frac{1}{2}\xi \times \boldsymbol{\delta}(\tau)^{\top} \boldsymbol{B} \boldsymbol{\delta}(\tau). \tag{A.11}$$

Proof. By Lemma A.2, we can show that

$$\sum_{t=1}^{n} \left(\varphi_{nt}(\boldsymbol{\delta}(\tau)) - \mathbb{E}_{t-1}\Big[\varphi_{nt}(\boldsymbol{\delta}(\tau))\Big] \right) = o_p(1). \tag{A.12}$$

Define the set $\mathcal{B}_t(\lambda) := \{\frac{1}{n} y_{t-1}^2 \leq \lambda\}$ for some positive $\lambda \in \mathbb{R}$. Then, (A.12) becomes as below

$$\sum_{t=1}^{n} \left(\varphi_{nt}(\boldsymbol{\delta}) \mathbf{1}\{\mathcal{B}_t(\lambda)\} - \mathbb{E}_{t-1}\Big[\varphi_{nt}(\boldsymbol{\delta}) \mathbf{1}\{\mathcal{B}_t(\lambda)\}\Big] \right) = o_p(1). \tag{A.13}$$

In particular, the expression $\left\{ \varphi_{nt}(\boldsymbol{\delta})\mathbf{1}\{\mathcal{B}_t(\lambda)\} - \mathbb{E}_{t-1}\left[\varphi_{nt}(\boldsymbol{\delta})\mathbf{1}\{\mathcal{B}_t(\lambda)\}\right]\right\}_{t=1}^{n}$ forms a martingale difference sequence, which by the Lenglart's inequality (see, Jacod and Shiryaev (2003)) it follows that

$$\sum_{t=1}^{n}\mathbb{E}_{t-1}\left[\varphi_{nt}^{2}(\boldsymbol{\delta})\mathbf{1}\{\mathcal{B}_t(\lambda)\}\right] = o_p(1). \tag{A.14}$$

Therefore, by definition of φ_{nt} we have that

$$\varphi_{nt}(\boldsymbol{\delta}) = \boldsymbol{\delta}^{\top}\boldsymbol{X}_t\int_0^1\Big(\psi(\tilde{\varepsilon}_t) - \psi\left(\tilde{\varepsilon}_t - \varepsilon_t\boldsymbol{\delta}^{\top}\boldsymbol{X}_t\right)\Big)d\varepsilon. \tag{A.15}$$

Furthermore, by the non-decreasing property of $\psi(x)$ it holds that

$$\sum_{t=1}^{n}\mathbb{E}_{t-1}\left[\varphi_{nt}^{2}(\boldsymbol{\delta})\mathbf{1}\{\mathcal{B}_t(\lambda)\}\right] \leq \max_{1\leq t\leq n}\mathbb{E}_{t-1}\left[\varphi_{nt}^{2}(\boldsymbol{\delta})\mathbf{1}\{\mathcal{B}_t(\lambda)\}\right]\times\left(\sum_{t=1}^{n}\boldsymbol{\delta}^{\top}\boldsymbol{X}_t\boldsymbol{X}_t^{\top}\boldsymbol{\delta}\right) = o_p(1). \tag{A.16}$$

□

Corollary A.1.

$$\sum_{t=1}^{n}\Big(\rho_{\tau}\left(\tilde{\varepsilon}_t - \boldsymbol{\delta}^{\top}\boldsymbol{X}_t\right) - \rho_{\tau}\left(\varepsilon_t\right)\Big) = -\boldsymbol{\delta}^{\top}\sum_{t=1}^{n}\boldsymbol{X}_t\psi\left(\varepsilon_t\right) + \frac{1}{2}\xi\boldsymbol{\delta}^{\top}\boldsymbol{B}\boldsymbol{\delta} + R_n\left(\boldsymbol{\delta}\right) \tag{A.17}$$

with $R_n(\boldsymbol{\delta}) = o_p(1)$ *for a fixed parameter vector* $\boldsymbol{\delta}$ *and* $\max_{\|\boldsymbol{\delta}\|\leq C} R_n(\boldsymbol{\delta}) = o_p(1)$.

Proof. Notice that the function $\varrho(\mathsf{u})$ is convex, therefore we can apply the same argument as that in the proof of Theorem 1 in Pollard (1991) and show that an equivalent solution to the optimization problem is given by the following expression

$$\hat{\gamma} = \sum_{t=1}^{n}\frac{1}{\xi}\boldsymbol{B}^{-1}\psi(\varepsilon_t)\boldsymbol{X}_t^{\top} + o_p(1). \tag{A.18}$$

□

A.1.2. Near-Explosive Case $(c > 0)$

Lemma A.4. *Consider that* $y_1, ..., y_n$ *are random variables generated from the autoregressive process. Then, when* $c > 0$ *it holds that*

$$\mathbb{E}[y_n^2] = o\left(\rho_n^{2n}k_n^2\right) \tag{A.19}$$

In addition to Lemma A.4 the following two results hold

$$\max_{1\leq t\leq n}\left\{\frac{y_t^2}{\rho_{n,c}^{2n}k_n^2}\right\} = o_p(1) \tag{A.20}$$

$$\frac{1}{\sqrt{n}\rho_{n,c}^{n}k_n}\sum_{t=1}^{n}y_{t-1} = o_p(1). \tag{A.21}$$

Lemma A.5. *We consider the following two joint convergence results*

(i). Under Assumptions 2.1-2.4 it holds that

$$\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_{\tau}(\varepsilon_t), \frac{1}{\sqrt{k_n}} \sum_{t=1}^{n} \rho_n^{t-(n-1)} \psi_{\tau}(\varepsilon_t), \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \rho_n^{-t} \varepsilon_t \right)^{\top} \overset{d}{\to} \left(\mathcal{Z}_1, \mathcal{Z}_2, \mathcal{Z}_3 \right)^{\top}. \tag{A.22}$$

where $\left(\mathcal{Z}_1, \mathcal{Z}_2, \mathcal{Z}_3 \right)$ *is a Gaussian random vector with independent components and the finite variance terms given by* $\mathbb{E}\left[\psi_{\tau}^2(\varepsilon_1)\right]$, $\frac{1}{2c}\mathbb{E}\left[\psi_{\tau}^2(\varepsilon_1)\right]$ *and* $\sigma^2/(2c)$, *respectively.*

(ii). Under Assumptions 2.1-2.4 it holds that

$$\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_{\tau}(\varepsilon_t), \frac{1}{\rho_{n,c}^{n} k_n} \sum_{t=1}^{n} y_{t-1} \psi_{\tau}(\varepsilon_t), \frac{1}{\rho_{n,c}^{2n} k_n^2} \sum_{t=1}^{n} y_{t-1}^2 \right)^{\top} \overset{d}{\to} \left(\mathcal{Z}_1, \mathcal{Z}_2 \mathcal{Z}_3, \frac{\mathcal{Z}_3^2}{2c} \right)^{\top}. \tag{A.23}$$

Proof. Recall that the difference equation with no model intercept, that is, $y_t = \rho_{n,c} y_{t-1} + \tilde{\varepsilon}_t$ has a general solution of the form $y_t = \rho_{n,c}^{t} y_0 + \sum_{j=1}^{t} \rho_{n,c}^{t-j} \tilde{\varepsilon}_j$. Similarly, for y_{t-1} by shifting the time index such that $t \mapsto t-1$, then the equivalent general solution is given by the following expression

$$y_{t-1} = \rho_{n,c}^{t-1} y_0 + \sum_{j=1}^{t-1} \rho_{n,c}^{t-1-j} \tilde{\varepsilon}_j \tag{A.24}$$

Thus, by substituting the above expression to the sample moment $\sum_{t=1}^{n} y_{t-1} \psi_{\tau}(\varepsilon_t)$ we obtain that

$$\frac{1}{\rho_{n,c}^{n} k_n} \sum_{t=1}^{n} y_{t-1} \psi_{\tau}(\varepsilon_t) = \frac{y_0}{\rho_{n,c}^{n} k_n} \sum_{t=1}^{n} \rho_{n,c}^{t-1} \psi_{\tau}(\varepsilon_t) + \frac{1}{\rho_{n,c}^{n} k_n} \sum_{t=1}^{n} \left(\sum_{j=1}^{t-1} \rho_{n,c}^{t-j-1} \tilde{\varepsilon}_j \right) \psi_{\tau}(\varepsilon_t) \tag{A.25}$$

Then, since the first term of the above expression is asymptotically negligible by splitting the inner summation of the last term we obtain that

$$\frac{1}{\rho_{n,c}^{n} k_n} \sum_{t=1}^{n} y_{t-1} \psi_{\tau}(\varepsilon_t) = \frac{1}{\rho_{n,c}^{n} k_n} \sum_{t=1}^{n} \left(\sum_{j=1}^{n} \rho_{n,c}^{t-1-j} \tilde{\varepsilon}_j - \sum_{j=t}^{n} \rho_{n,c}^{t-1-j} \tilde{\varepsilon}_j \right) \psi_{\tau}(\varepsilon_t) + o_p(1).$$

Since, $\sum_{j=t}^{n} \rho_{n,c}^{t-1-1} \tilde{\varepsilon}_j \overset{p}{\to} 0$, then it follows that

$$\begin{aligned} \frac{1}{\rho_{n,c}^{n} k_n} \sum_{t=1}^{n} y_{t-1} \psi_{\tau}(\varepsilon_t) &= \frac{1}{\rho_{n,c}^{n} k_n} \sum_{t=1}^{n} \left(\sum_{j=1}^{n} \rho_{n,c}^{t-1-j} \tilde{\varepsilon}_j \right) \psi_{\tau}(\varepsilon_t) + o_p(1) \\ &= \left(\frac{1}{\sqrt{k_n}} \sum_{t=1}^{n} \rho_{n,c}^{t-(n+1)} \psi_{\tau}(\varepsilon_t) \right) \left(\frac{1}{\sqrt{k_n}} \sum_{t=1}^{n} \rho_{n,c}^{-t} \tilde{\varepsilon}_t \right) + o_p(1). \end{aligned}$$

Similarly, it holds that

$$\frac{1}{\rho_{n,c}^{2n} k_n^2} \sum_{t=1}^{n} y_{t-1}^2 = \frac{1}{2c} \left(\frac{1}{\sqrt{k_n}} \sum_{t=1}^{n} \rho_{n,c}^{-t} \tilde{\varepsilon}_t \right) + o_p(1). \tag{A.26}$$

□

Lemma A.6. *The following joint convergence results hold*

(i). Under Assumptions above we have that

$$-\boldsymbol{\vartheta}_n(\tau)^\top \sum_{t=1}^{n} \boldsymbol{X}_t \psi_\tau(\varepsilon_t) + \frac{\xi}{2} \boldsymbol{\vartheta}_n(\tau)^\top \left(\sum_{t=1}^{n} \boldsymbol{X}_t \boldsymbol{X}_t^\top \right) \boldsymbol{\vartheta}_n(\tau) \overset{d}{\to} -\boldsymbol{\vartheta}_n(\tau) \big(\mathcal{Z}_1, \mathcal{Z}_2 \mathcal{Z}_3 \big)^\top + \frac{\xi}{2} \boldsymbol{\vartheta}_n(\tau)^\top \boldsymbol{B} \boldsymbol{\vartheta}_n(\tau)$$

(ii). Under Assumptions above we have that

$$\sum_{t=1}^{n} \left[\varrho_\tau \Big(y_t - \boldsymbol{\vartheta}_n(\tau)^\top \boldsymbol{X}_t \Big) - \varrho_\tau(\varepsilon_t) \right] = -\boldsymbol{\vartheta}_n(\tau)^\top \sum_{t=1}^{n} \boldsymbol{X}_t \psi_\tau(\varepsilon_t) + \frac{\xi}{2} \boldsymbol{\vartheta}_n(\tau)^\top \left(\sum_{t=1}^{n} \boldsymbol{X}_t \boldsymbol{X}_t^\top \right) \boldsymbol{\vartheta}_n(\tau) + R_n \Big(\boldsymbol{\vartheta}_n(\tau) \Big).$$

with $R_n\Big(\boldsymbol{\vartheta}_n(\tau)\Big) = o_p(1)$ *for fixed* $\boldsymbol{\vartheta}_n(\tau)$ *and further*

Proof. In order to prove the uniformity condition we denote with

$$\varphi(\boldsymbol{\vartheta}_n(\tau)) = \frac{1}{2} \xi \times \boldsymbol{\vartheta}_n(\tau)^\top \left(\sum_{t=1}^{n} \boldsymbol{X}_t \boldsymbol{X}_t^\top \right) \boldsymbol{\vartheta}_n(\tau). \tag{A.27}$$

Furthermore, we need to show that

$$\sup_{\|\boldsymbol{\vartheta}_n(\tau)\| \leq C} \left| \sum_{t=1}^{n} \varphi_{nt}(\boldsymbol{\vartheta}_n(\tau)) - \varphi(\boldsymbol{\vartheta}(\tau)) \right| = o_P(1). \tag{A.28}$$

Since $\sum_{t=1}^{n} \boldsymbol{X}_t \boldsymbol{X}_t^\top$ converges in distribution, for any $\lambda > 0$ there exists M large enough, such that,

$$\mathbb{P} \left(\sup_{\|\boldsymbol{\vartheta}_n(\tau)\| \leq C} \left| \sum_{t=1}^{n} \varphi_{nt}(\boldsymbol{\vartheta}_n(\tau)) - \varphi(\boldsymbol{\vartheta}(\tau)) \right| \mathbf{1} \left\{ \left\| \sum_{t=1}^{n} \boldsymbol{X}_t \boldsymbol{X}_t^\top \right\| > M \right\} > \lambda/2 \right) < \lambda^*/2. \tag{A.29}$$

On the other hand, on $\left\{ \left\| \sum_{t=1}^{n} \boldsymbol{X}_t \boldsymbol{X}_t^\top \right\| \right\}$, for any $\lambda > 0$, there exists $\delta > 0$, such that,

$$\sup_{\|\boldsymbol{\vartheta}\| \leq C} \Big| \varphi(\gamma + \vartheta) - \varphi(\gamma) \Big| \leq \lambda. \tag{A.30}$$

Moreover, following the convexity Lemma of Pollard (1991), one can show that

$$\mathbb{P} \left(\sup_{\|\boldsymbol{\vartheta}_n(\tau)\| \leq C} \left| \sum_{t=1}^{n} \varphi_{nt}(\boldsymbol{\vartheta}_n(\tau)) - \varphi(\boldsymbol{\vartheta}(\tau)) \right| \mathbf{1} \left\{ \left\| \sum_{t=1}^{n} \boldsymbol{X}_t \boldsymbol{X}_t^\top \right\| \leq M \right\} > \lambda/2 \right) < \lambda^*/2. \tag{A.31}$$

Therefore, the combination of the above yields the uniformity result of interest. □

A.2. Supplementary Results

Following Koenker (2005), we consider that all parameters share the same monotone behaviour with respect to the quantile level $\tau \in (0,1)$. Then the consistency of quantile dependent parameter can be deduced from the monotonicity of the subgradient as well as a direct consequence of the uniform convergence of the empirical distribution function and the Glivenko-Cantelli Theorem. Therefore, the asymptotic behaviour of $\sqrt{n}\left(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\right)$ follows by considering the following objective function

$$\mathcal{Z}_n(\delta) = \frac{1}{n}\sum_{t=1}^{n} \varrho_\tau\left(\varepsilon_t - \boldsymbol{X}_t^\top \boldsymbol{\delta}/\sqrt{n}\right) - \varrho_\tau(\varepsilon_t) \tag{A.32}$$

where $\varepsilon_t(\tau) = y_t - \boldsymbol{X}_t^\top \boldsymbol{\beta}(\tau)$. The function $\mathcal{Z}_n(\boldsymbol{\delta})$ is convex, and is minimized at $\hat{\boldsymbol{\delta}}_n = \sqrt{n}\left(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\right)$. Therefore, we can show that the limit distribution of $\hat{\boldsymbol{\delta}}_n$ can be determined by the asymptotic distribution of the objective function $\mathcal{Z}_n(\boldsymbol{\delta})$. Furthermore, it follows from the Lindeberg-Feller central limit theorem that $\mathcal{Z}_n^{(1)} \overset{d}{\to} -\boldsymbol{\delta}'\mathcal{W}$, where $\mathcal{W} \overset{d}{\to} \mathcal{N}(0, \tau(1-\tau)\boldsymbol{D}_0)$. Then, following the proof of Theorem 4.1 of Koenker (2005) it holds that (see also derivations in Knight (1998) and Kato (2009))

$$\sum_{t=1}^{n} \mathbb{E}\left[\mathcal{Z}_{nt}^{(2)}(\boldsymbol{\delta})\right] \overset{d}{\to} \frac{1}{2}\boldsymbol{\delta}^\top \boldsymbol{D}_1 \boldsymbol{\delta}. \tag{A.33}$$

Therefore, it can be proved that

$$\mathcal{Z}_n(\boldsymbol{\delta}) \overset{d}{\to} \mathcal{Z}_0(\boldsymbol{\delta}) \equiv -\boldsymbol{\delta}^\top \mathcal{W} + \frac{1}{2}\boldsymbol{\delta}^\top \boldsymbol{D}_1 \boldsymbol{\delta}. \tag{A.34}$$

The convexity of the limiting distribution of $\mathcal{Z}_0(\boldsymbol{\delta})$ ensures that the uniqueness of the minimizer

$$\hat{\delta}_n := \underset{\delta}{\arg\min}\ \mathcal{Z}_n(\boldsymbol{\delta}) \mapsto \hat{\delta}_0 := \underset{\boldsymbol{\delta}}{\arg\min}\ \mathcal{Z}_0(\boldsymbol{\delta}) \tag{A.35}$$

where $\mathcal{Z}_n(\boldsymbol{\delta}) = \mathcal{Z}_n^{(1)}(\boldsymbol{\delta}) + \mathcal{Z}_n^{(2)}(\boldsymbol{\delta})$ and $\mathcal{Z}_0(\boldsymbol{\delta}) = -\boldsymbol{\delta}^\top \mathcal{W} + \frac{1}{2}\boldsymbol{\delta}^\top \boldsymbol{D}_1 \boldsymbol{\delta}$ such that $\hat{\boldsymbol{\delta}}_n = \sqrt{n}\left(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\right)$. A similar approach is followed in the study of Mao and Guo (2019) who prove that for any $\tau \in (0,1)$ the solution of the following expression is obtained by $\frac{\sqrt{n}}{\lambda(n)}\left(\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}(\tau)\right) \in \underset{\mathsf{u}\in\mathbb{R}^p}{\arg\min}\ \mathcal{Z}_n(\mathsf{u}, \tau)$.

Then, it follows that

$$\begin{aligned}\left|\mathcal{Z}_n(\mathsf{u},\tau) - G_n(\mathsf{u},\tau)\right| &= \left|\mathcal{Z}_n^{(2)}(\mathsf{u},\tau) - \frac{\mathsf{u}^\top \boldsymbol{D}\mathsf{u}}{2} f\left(F^{-1}(\tau)\right)\right| \\ &\leq \left|\mathbb{E}\left[\mathcal{Z}_n^{(2)}(\mathsf{u},\tau)\right] - \frac{\mathsf{u}^\top \boldsymbol{D}\mathsf{u}}{2} f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)\right| + \left|\mathcal{Z}_n^{(2)}(\mathsf{u},\tau) - \mathbb{E}\left[\mathcal{Z}_n^{(2)}(\mathsf{u},\tau)\right]\right|\end{aligned}$$

where

$$\mathcal{Z}_n^{(2)}(\mathsf{u},\tau) = \frac{1}{\sqrt{n}\lambda(n)}\sum_{i=1}^{n}\int_0^1 \left(\mathbf{1}\left\{\varepsilon_i \leq F_\varepsilon^{-1}(\tau) + \frac{\lambda(n)\sigma_{in}^{-1} x_{in}^\top \mathsf{u}s}{\sqrt{n}}\right\} - \mathbf{1}\left\{\varepsilon_i \leq F_\varepsilon^{-1}(\tau)\right\}\right) ds. \tag{A.36}$$

Therefore, it can be shown that

$$\sup_{\tau\in[\alpha,1-\alpha]} \left|\mathbb{E}\left[\mathcal{Z}_n^{(2)}(\mathsf{u},\tau)\right] - \frac{\mathsf{u}^\top \boldsymbol{D}\mathsf{u}}{2} f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)\right| \to 0, \quad \text{as } n\to\infty \tag{A.37}$$

Thus, for large n, we further have that

$$\text{(A.38)} \qquad \mathbb{P}\left(\sup_{\tau\in[\alpha,1-\alpha]}\left|\mathcal{Z}_n(\mathsf{u},\tau)-G_n(\mathsf{u},\tau)\right|\geq\lambda\right)\leq\mathbb{P}\left(\sup_{\tau\in[\alpha,1-\alpha]}\left|\mathcal{Z}_n^{(2)}(\mathsf{u},\tau)-\mathbb{E}\left[\mathcal{Z}_n^{(2)}(\mathsf{u},\tau)\right]\right|\geq\frac{\lambda}{2}\right)$$

The following step is to prove that

$$\text{(A.39)} \qquad \limsup_{n\to\infty}\frac{1}{\lambda^2(n)}\log\left\{\mathbb{P}\left(\sup_{\tau\in[\alpha,1-\alpha]}\left|\mathcal{Z}_n^{(2)}(\mathsf{u},\tau)-\mathbb{E}\left[\mathcal{Z}_n^{(2)}(\mathsf{u},\tau)\right]\right|\geq\lambda\right)\right\}=-\infty$$

From the definition of $\mathcal{Z}_n^{(2)}(\mathsf{u},t)$, we obtain that

$$\text{(A.40)} \qquad \mathcal{Z}_n^{(2)}(\mathsf{u},\tau)-\mathbb{E}\left[\mathcal{Z}_n^{(2)}(\mathsf{u},\tau)\right]=\int_0^1\Big(\mathcal{W}_n(\mathsf{u}s,\tau)-\mathcal{W}_n(0,\tau)\Big)ds,$$

where

$$\mathcal{W}_n(r,\tau)=\frac{\sum_{t=1}^n\boldsymbol{X}'_{nt}\mathsf{u}}{\sqrt{n}\lambda(n)}\left[\mathbf{1}\left\{\varepsilon_i\leq F_\varepsilon^{-1}(\tau)+\frac{\lambda(n)\sigma_n^{-1}\boldsymbol{X}'_{nt}r}{\sqrt{n}}\right\}-F_\varepsilon\left(F_\varepsilon^{-1}(\tau)+\frac{\lambda(n)\sigma_n^{-1}\boldsymbol{X}'_{nt}r}{\sqrt{n}}\right)\right]$$

In practise we employ Theorem 1 from Knight (1998). Therefore, we have that

$$\text{(A.41)} \qquad \mathcal{Z}_n^{(1)}(\mathsf{u})=-\frac{1}{\sqrt{n}}\sum_{i=1}^n\boldsymbol{x}_i\mathsf{u}\Big[\mathbf{1}(\epsilon_i>0)-\mathbf{1}(\epsilon_i<0)\Big]$$

and

$$\text{(A.42)} \qquad \mathcal{Z}_n^{(2)}(\mathsf{u})=\frac{2a_n}{\sqrt{n}}\sum_{t=1}^n\int_0^{v_{n_i}}\boldsymbol{X}_t\mathsf{u}\Big[\mathbf{1}(\epsilon_i<s)-\mathbf{1}(\epsilon_i<0)\Big]ds$$

where $v_n=\boldsymbol{X}_t^\top\mathsf{u}/a_n$. Then, by the Lindeberg-Feller central limit theorem, for each u it holds that,

$$\text{(A.43)} \qquad \mathcal{Z}_n^{(1)}(\mathsf{u})\xrightarrow{d}-\mathsf{u}\boldsymbol{W}$$

and the convergence in distribution holds for any finite collection of u's. For $\mathcal{Z}_n^{(2)}(\mathsf{u})$, we have that

$$\text{(A.44)} \qquad \mathcal{Z}_n^{(2)}(\mathsf{u})=\sum_{t=1}^n\mathbb{E}\left[\mathcal{Z}_{nt}^{(2)}(\mathsf{u})\right]+\sum_{t=1}^n\left[\mathcal{Z}_{nt}^{(2)}(\mathsf{u})-\mathbb{E}\left[\mathcal{Z}_{nt}^{(2)}(\mathsf{u})\right]\right].$$

We employ the above orthogonal decomposition when proving the limit results for the estimator of the quantile autoregressive model for moderate deviations from the unit boundary. Notice that the implementation of the corresponding FCLT for the quantile-dependent innovation term is applicable for the *i.i.d* innovation sequence assumption. An extension of the particular results to the case in which innovations have serial correlation via the use of a linear process representation for instance is also possible. Further applications with suitable econometric conditions we could consider within our framework are presented in the study of Doukhan and Louhichi (1999) who consider a different type of weak dependence condition for time series models.

References

Abadir, K. M. and Lucas, A. (2000). Quantiles for t-statistics based on m-estimators of unit roots. *Economics Letters*, 67(2):131–137.

Anderson, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *The Annals of Mathematical Statistics*, pages 676–687.

Aue, A. and Horváth, L. (2007). A limit theorem for mildly explosive autoregression with stable errors. *Econometric Theory*, 23(2):201–220.

Bahadur, R. R. (1966). A note on quantiles in large samples. *The Annals of Mathematical Statistics*, 37(3):577–580.

Buchmann, B. and Chan, N. H. (2007). Asymptotic theory of least squares estimators for nearly unstable processes under strong dependence. *The Annals of statistics*, 35(5):2001–2017.

Cavaliere, G. (2002). Bounded integrated processes and unit root tests. *Statistical Methods and Applications*, 11(1):41–69.

Cavanagh, C. (1985). Roots local to unity. *Manuscript*.

Chan, N. H. (1988). The parameter inference for nearly nonstationary time series. *Journal of the American Statistical Association*, 83(403):857–862.

Chan, N. H. (1990). Inference for near-integrated time series with infinite variance. *Journal of the American Statistical Association*, 85(412):1069–1074.

Chan, N. H., Peng, L., and Qi, Y. (2006). Quantile inference for near-integrated autoregressive time series with infinite variance. *Statistica Sinica*, pages 15–28.

Chan, N. H. and Tran, L. T. (1989). On the first-order autoregressive process with infinite variance. *Econometric Theory*, 5(3):354–362.

Chan, N. H. and Wei, C.-Z. (1987). Asymptotic inference for nearly nonstationary ar (1) processes. *The Annals of Statistics*, pages 1050–1063.

Cox, D. D. and Llatas, I. (1991). Maximum likelihood type estimation for nearly nonstationary autoregressive time series. *The Annals of Statistics*, pages 1109–1128.

Cramér, H. (1951). A contribution to the theory of stochastic processes. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 329–339. University of California Press.

Csörgő, M. (1983). *Quantile processes with statistical applications*. SIAM.

Csorgo, M., Csorgo, S., Horváth, L., and Mason, D. M. (1986). Weighted empirical and quantile processes. *The Annals of Probability*, pages 31–85.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431.

Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: journal of the Econometric Society*, pages 1057–1072.

Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic processes and their applications*, 84(2):313–342.

Fountis, N. G. and Dickey, D. A. (1989). Testing for a unit root nonstationarity in multivariate autoregressive time series. *The Annals of Statistics*, pages 419–428.

Garderen, K. J. v. (1999). Exact geometry of autoregressive models. *Journal of time series analysis*, 20(1):1–21.

Giraitis, L. and Phillips, P. C. (2006). Uniform limit theory for stationary autoregression. *Journal of time series analysis*, 27(1):51–60.

Goh, S. C. and Knight, K. (2009). Nonstandard quantile-regression inference. *Econometric Theory*, 25(5):1415–1432.

Hasan, M. N. and Koenker, R. W. (1997). Robust rank tests of the unit root hypothesis. *Econometrica: Journal of the Econometric Society*, pages 133–161.

Huang, S.-H., Pang, T.-X., and Weng, C. (2014). Limit theory for moderate deviations from a unit root under innovations with a possibly infinite variance. *Methodology and Computing in Applied Probability*, 16(1):187–206.

Hui, J., Yilong, W., and Guangyu, Y. (2022). Deviation inequalities and cramer-type moderate deviations for the explosive autoregressive process. *Bernoulli*, 28(5):1–28.

Jacod, J. and Shiryaev, A. (2003). *Limit theorems for stochastic processes*, volume 288. Berlin: Springer - Verlag.

Jansson, M. (2004). The error in rejection probability of simple autocorrelation robust tests. *Econometrica*, 72(3):937–946.

Jansson, M. and Moreira, M. J. (2006). Optimal inference in regression models with nearly integrated regressors. *Econometrica*, 74(3):681–714.

Jiang, H., Yu, M., and Yang, G. (2015). Moderate deviations for the mildly stationary autoregressive models with dependent errors. *arXiv preprint arXiv:1510.02862*.

Jirak, M. (2016). Berry–esseen theorems under weak dependence. *The Annals of Probability*, 44(3):2024–2063.

Jurečková, J., Kallenberg, W., and Veraverbeke, N. (1988). Moderate and cramér-type large deviation theorems for m-estimators. *Statistics & probability letters*, 6(3):191–199.

Kato, K. (2009). Asymptotics for argmin processes: Convexity arguments. *Journal of Multivariate Analysis*, 100(8):1816–1829.

Kiefer, N. M., Vogelsang, T. J., and Bunzel, H. (2000). Simple robust testing of regression hypotheses. *Econometrica*, 68(3):695–714.

Knight, K. (1987). Rate of convergence of centred estimates of autoregressive parameters for infinite variance autoregressions. *Journal of time series analysis*, 8(1):51–60.

Knight, K. (1998). Limiting distributions for l1 regression estimators under general conditions. *Annals of statistics*, pages 755–770.

Koenker, R. (2005). Quantile regression.

Koenker, R. (2012). quantreg: Quantile regression. r package version 4.98. *Available at cran.r-project.org.*

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

Koenker, R. and Portnoy, S. (1987). L-estimation for linear models. *Journal of the American statistical Association*, 82(399):851–857.

Koenker, R. and Xiao, Z. (2002). Inference on the quantile regression process. *Econometrica*, 70(4):1583–1612.

Koenker, R. and Xiao, Z. (2004). Unit root quantile autoregression inference. *Journal of the American Statistical Association*, 99(467):775–787.

Koenker, R. and Xiao, Z. (2006). Quantile autoregression. *Journal of the American statistical association*, 101(475):980–990.

Kong, X.-B. (2015). M-estimation for moderate deviations from a unit root. *Communications in Statistics-Theory and Methods*, 44(3):476–485.

Kyprianou, A. E. (2014). *Fluctuations of Lévy processes with applications: Introductory Lectures*. Springer Science & Business Media.

Lahiri, S. N. and Sun, S. (2009). A berry–esseen theorem for sample quantiles under weak dependence. *The Annals of Applied Probability*, 19(1):108–126.

Lai, T. L. and Wei, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *The Annals of Statistics*, 10(1):154–166.

Larsson, R. (1995). The asymptotic distributions of some test statistics in near-integrated ar processes. *Econometric Theory*, 11(2):306–330.

Lee, J. H. (2016). Predictive quantile regression with persistent covariates: Ivx-qr approach. *Journal of Econometrics*, 192(1):105–118.

Lee, J. H. (2018). Limit theory for explosive autoregression under conditional heteroskedasticity. *Journal of Statistical Planning and Inference*, 196:30–55.

Ling, S. and McAleer, M. (2004). Regression quantiles for unstable autoregressive models. *Journal of Multivariate Analysis*, 89(2):304–328.

Liu, Q. and Liu, X. (2018). Limit theory for an ar (1) model with intercept and a possible infinite variance. *arXiv preprint arXiv:1802.10299*.

Lucas, A. (1995). Unit root tests based on m estimators. *Econometric Theory*, 11(2):331–346.

Magdalinos, T. (2007). On the inconsistency of the unrestricted estimator of the information matrix near a unit root. *The Econometrics Journal*, 10(2):245–262.

Magdalinos, T. (2012). Mildly explosive autoregression under weak and strong dependence. *Journal of Econometrics*, 169(2):179–187.

Magdalinos, T. and Phillips, P. C. B. (2009). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric Theory*, 25(2):482–526.

Mann, H. B. and Wald, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica, Journal of the Econometric Society*, pages 173–220.

Mao, M. and Guo, W. (2019). Moderate deviations for quantile regression processes. *Communications in Statistics-Theory and Methods*, 48(12):2879–2892.

Mikusheva, A. (2012). One-dimensional inference in autoregressive models with the potential presence of a unit root. *Econometrica*, 80(1):173–212.

Neocleous, T. and Portnoy, S. (2008). On monotonicity of regression quantile functions. *Statistics & probability letters*, 78(10):1226–1229.

Nielsen, B. (2001). The asymptotic distribution of unit root tests of unstable autoregressive processes. *Econometrica*, 69(1):211–219.

Penda, S. V. B., Djellout, H., and Proïa, F. (2014). Moderate deviations for the durbin–watson statistic related to the first-order autoregressive process. *ESAIM: Probability and Statistics*, 18:308–331.

Phillips, P. C., Moon, H. R., and Xiao, Z. (2001). How to estimate autoregressive roots near unity. *Econometric Theory*, 17(1):29–69.

Phillips, P. C., Shi, S., and Yu, J. (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the s&p 500. *International economic review*, 56(4):1043–1078.

Phillips, P. C. B. (1987a). Time series regression with a unit root. *Econometrica: Journal of the Econometric Society*, pages 277–301.

Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika*, 74(3):535–547.

Phillips, P. C. B. and Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, 136(1):115–130.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, 7(2):186–199.

Proïa, F. (2020). Moderate deviations in a class of stable but nearly unstable processes. *Journal of Statistical Planning and Inference*, 208:66–81.

Rao, M. (1978). Asymptotic distribution of an estimator of the boundary parameter of an unstable process. *The Annals of Statistics*, pages 185–190.

Saxena, K. L. and Alam, K. (1982). Estimation of the non-centrality parameter of a chi squared distribution. *The Annals of Statistics*, pages 1012–1016.

Vogelsang, T. J. (1998). Trend function hypothesis testing in the presence of serial correlation. *Econometrica*, pages 123–148.

Werker, B. J. and Zhou, B. (2022). Semiparametric testing with highly persistent predictors. *Journal of Econometrics*, 227(2):347–370.

White, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *The Annals of Mathematical Statistics*, pages 1188–1197.

Xu, C. and Pang, T. (2018). Limit theory for moderate deviations from a unit root with a break in variance. *Communications in Statistics-Theory and Methods*, 47(24):6125–6143.

Yabe, R. (2012). Limiting distribution of the score statistic under moderate deviation from a unit root in ma (1). *Journal of Time Series Analysis*, 33(4):533–541.