Asymptotic Theory for Unit Root Moderate Deviations in Quantile Autoregressions and Predictive Regressions∗

Christis Katsouris

University of Southampton and University of Exeter

Working Paper

August 22, 2023

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 functional form and model parameters are quantile-dependent. A Bahadur-type representation and limiting distributions based on the M-estimators of the model parameters are derived. 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.

JEL classification: C22

∗Article history: Previous draft titled: "Asymptotic Theory for Moderate Deviations from the Unit Boundary in Quantile Autoregressive Time Series", 05 April 2022. This draft: August 2023. I am grateful to Xiaohui Luì, Tassos Magdalinos, Jean-Yves Pitarakis and Jose Olmo for helpful discussions.

Lecturer in Economics, University of Exeter Business School, Exeter EX4 4PU, United Kingdom. Email: christiskatsouris@gmail.com.
1 Introduction

Moderate deviation principles from the unit boundary for quantile autoregressions 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)). Specifically, studies for quantile autoregressive regressions that consider moderate deviations within a unified framework allowing to investigate the asymptotic behaviour of estimators with respect to different regimes 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 autoregressive coefficient when considering deviations from the unit boundary under a conditional quantile functional form. 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) and Wang et al. (2022) who develop limit theory for moderate deviations in autoregressive models based on M-estimators.

In this line of literature, the two relevant research questions are summarized by 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 $\theta_n$ is close to the unit boundary? Consequently, when the autoregression coefficient is expressed with respect to the local-to-unity parametization, what kind of approximation should be used for the distribution of the test statistics?"

The seminal studies of Chan and Wei (1987) and Phillips (1987a,b) tackle exactly these nonstandard statistical problems via their triangular array framework. The framework of the nearly nonstationary AR(1) model allows to establish the limiting distributions of the least squares estimator for $\theta_n$ under the assumption that the conditional variance of the model is finite. Moreover, the properties of the least squares estimator when the true stochastic process is nearly integrated are investigated by Chan (1988, 1990), Chan and Tran (1989), Knight (1987), Rao (1978), Lai and Wei (1982), Cox and Llatas (1991), Larsson (1995), Cavaliere (2002), Buchmann and Chan (2007), Phillips and Magdalinos (2007) and Duffy and Kasparis (2021) among others. Furthermore, the asymptotic theory for moderate deviations from a unit root in autoregressive models is presented 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.
On the other hand, the properties of nonstationary autoregressive models for the case of an explosive autoregressive coefficient is also of interest. In particular, White (1958) obtained the limit distribution of an explosive serial correlation coefficient (see, also Mann and Wald (1943)). Limit theory for moderate deviations on the explosive side of unity (e.g., mildly explosive case) were developed by Buchmann and Chan (2007), Aue and Horváth (2007), Magdalinos (2012), Arvanitis and Magdalinos (2018), Oh et al. (2018), Lee (2018), Próia (2020), Yu and Kejriwal (2021), Hirukawa and Lee (2021), Liu et al. (2022). Recently, Hui et al. (2022) derived the limit theory for the ordinary least squares estimator in the explosive first-order Gaussian autoregressive process using a set of deviation inequalities\(^1\). 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 \( \mathcal{O}(1/n) \).

Under the Cramér-type moderate deviations framework, there exists positive sequences \( \nu_n \) and \( \lambda_n \) 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 \( x \in \mathbb{R} \)

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

We focus on 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). Due to the form of the nonrandom sequence such that, \( k_n \equiv n^\gamma \), the convergence rate towards unity is slower than the sample size \( n \). In this case the autoregression coefficient approaches 1 at a rate slower than the usual local alternative as \( n \) goes to infinity. When, as \( \gamma \to 1 \) then \( k_n \to n \) which encompasses the conventional local-to-unity parametrization. An alternative form for the convergence rate include the case when \( k_n = \sqrt{n} \). To obtain limit results for model parameters based on M-estimation, we employ the Bahadur representation (see, Bahadur (1966)) that provides a mechanism to facilitate the asymptotic theory, which allows to obtain analytic expressions for quantile-dependent estimators by approximating them with linear forms. Although we do not consider how the presence of serial correlation can affect the limiting distributions, various studies in the literature consider simple implementations of autoregression robust tests as in Jansson (2004) (see, also Vogelsang (1998)). 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. 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.

---

\(^1\)The authors obtain the limit theory of Cramér-type moderate deviations for the explosive and mildly explosive autoregressive processes.
1.1 Literature Review

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 one, as pointed out by Dickey and Fuller (1979) (see, also Dickey and Fuller (1981)). The idea of developing asymptotics using the local-to-unity parametrization is due to the studies of Cavanagh (1985), Phillips (1987a), Chan and Wei (1987). 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 $c < 0$ for $\theta_n = \left(1 + \frac{c}{n}\right)$,

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

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. In that case, by expressing the AR(1) model with respect to the canonical parameters of an exponential family one can establish the asymptotic behaviour of related minimal sufficient statistics\(^2\) (see, Jansson and Moreira (2006)). The particular literature is developed under the assumption of a stationary autoregression 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)). Specifically, this implies that when the true parameter value is explosive rather than being stationary, then the asymptotic theory of estimators and test statistics based on OLS estimation are driven by the distribution of the innovations $u_t$. On the other hand, comparing the cases of Gaussian innovations against non-Gaussian innovations (e.g., heavy tailed errors) and an explosive autoregressive parameter then the asymptotic theory of the OLS estimator in these two cases will not necessarily be identical.

Various studies in the literature consider limit results for moderate deviations for M-estimators and quantile processes\(^4\) in autoregressive models include among others Jurečková et al. (1988), Knight (1998), Mao and Guo (2019) as well as Kato (2009) who develops asymptotics for Lasso quantile-dependent estimators. Limit theory for moderate deviations from the unit root in the context of quantile autoregressive models are established in the studies of Lucas (1995), Abadir and Lucas (2000), Ling and McAleer (2004), Koenker and Xiao (2004), Chan et al. (2006), Kong (2015), Zhou and Lin (2015), Wang et al. (2022) and Fu et al. (2022).

\(^2\)In particular, Jansson and Moreira (2006) using the local-to-unity parametrization of the autoregression coefficient study the properties of predictive regressions under the assumption of persistence regressors using differential geometry and sufficient statistics arguments to establish the asymptotic theory of estimators and test statistics.

\(^3\)Using the local-to-unity parametrization allows the analysis of unit roots and explosive processes, which is necessary to link problems in inference for unit roots to the statistical curvature.

\(^4\)Related theoretical aspects can be found in the book of Csörgő (1983) (see, also Csorgo et al. (1986)).
Another important aspect worth mentioning, is the fact that several studies 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. The limit theory for an autoregressive model which includes a intercept and an autoregression coefficient expressed using the local-to-unity parametrization is studied by Liu and Liu (2018) and Hwang and Pang (2009), however their framework differs from our setting since we consider quantile-dependent parameters. This article, builds on and contributes to both the quantile autoregression literature as well as to the literature of moderate deviations from the unit boundary. We focus on establishing the limit theory for moderate deviations from the unit boundary for both the near-stationary and near-explosive cases in quantile autoregression and quantile predictive regressions. Our main objective is to develop a unified framework for the asymptotic behaviour of the quantile-dependent estimators across the whole spectrum of nonstationarity regimes, including the general near-integrated case assuming that \( \theta_n \to 1 \) with a rate slower than \( 1/n \).

Throughout the paper, we assume that all random elements are defined within a probability space denoted with the triple \( (\Omega, \mathcal{F}, P) \). All limits are taken as \( n \to \infty \), where \( n \) is the sample size. Denote with \( D([0,1]) \) to be the set of functions on \([0,1]\) that are right continuous and have left limits, equipped with the Skorokhod metric. Then, the symbol ” \( \Rightarrow \) ” is used to denote the weak convergence of the associated probability measures as \( n \to \infty \). The symbol \( D \to \) and \( P \to \) are employed to denote convergence in distribution and convergence in probability respectively. Moreover, we denote with \( O_P(.) \) and \( o_P(.) \) the stochastic order of convergence in probability (Billingsley (1968)).

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. Section 2.3, 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 ?? presents a short Monte Carlo simulation study while Section ?? an empirical application. Section 4 concludes.

## 2 Moderate Deviations from the Unit Boundary Framework

### 2.1 Main Terminology, Definitions and Assumptions

Consider the first order autoregressive process with expressed as below

\[
x_t^* = \theta_n x_{t-1}^* + u_t, \quad \text{for } t = 1, ..., n,
\]

with a possible sample-size dependent autoregressive root \( \theta_n \) and \( (u_t)_{t \in \mathbb{N}} \) is the innovation sequence\(^5\).

\(^5\)The innovations \( (u_t)_{t \in \mathbb{N}} \) is an \( i.i.d \) sequence of random variables from a common distribution function \( F_u \) that satisfies regularity conditions for Lipschitz continuity with zero mean and finite variance \( \sigma_u \).
Moreover, we allow for the presence of a non-zero intercept by considering the process

\begin{equation}
    x_t = \mu + x_t^*, \quad \text{with} \quad x_{0t} = \sum_{j=1}^{n} \theta_n^{t-j} u_j. 
\end{equation}

where \( x_{0t} \) represents the solution of the autoregressive process with \( \mu = 0 \) and \( X_0 = 0 \), the initial condition of the recursion. Furthermore, since we are interested about the asymptotic stability of the autoregressive process near the unit boundary we allow a local-to-unity parametrization as

\begin{equation}
    \theta_n = \left( 1 + \frac{c}{k_n} \right), \quad \text{with} \quad c \in \mathbb{R}, \quad k_n = n^\gamma \quad \text{and} \quad \gamma = 0, \ \gamma \in (0, 1) \text{ or } \gamma = 1.
\end{equation}

The convergence rate of the autoregression coefficient \( \theta_n \) 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 (Persistence).** Consider the autoregressive process \( X_t = \theta_n X_{t-1} + \varepsilon_t \) with \( (\varepsilon_t)_{t\in\mathbb{N}} \). Then, consider the following limit to determine the degree of persistence for the processes

\begin{equation}
    \zeta_n := \lim_{n \to +\infty} n (\theta_n - 1) \rightarrow \zeta
\end{equation}

- **P.1 nearly stable processes:** if \( (\theta_n)_{n\in\mathbb{N}} \) is such that \( \zeta = -\infty \) and it holds that \( \theta_n \to |\theta| < 1 \).
- **P.2 nearly unstable processes:** if \( (\theta_n)_{n\in\mathbb{N}} \) is such that \( \zeta \equiv c \in \mathbb{R} \) and it holds that \( \theta_n \to \theta = 1 \).
- **P.3 nearly explosive processes:** if \( (\theta_n)_{n\in\mathbb{N}} \) is such that \( \zeta = +\infty \) and it holds that \( \theta_n \to |\theta| > 1 \).

**Assumption 2.2.** Denote with \( x_{0n} = 0 \), 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 \( n \to \infty \)

\begin{equation}
    \frac{1}{n} \sum_{t=1}^{n} \mathbb{E} \left[ \varepsilon_t^2 \mid \mathcal{F}_{t-1} \right] = 1 + o_p(1),
\end{equation}

\begin{equation}
    \frac{1}{n} \sum_{t=1}^{n} \mathbb{E} \left[ \varepsilon_t^2 1 \{ |\varepsilon_t | > n^{1/2} m \} \mid \mathcal{F}_{t-1} \right] = o_p(1), \quad \text{for all } m > 0,
\end{equation}

where \( \mathcal{F}_t := \sigma(\varepsilon_s : 0 \leq s \leq t) \).
Remark 2.1. Assumption 2.2 provides moment conditions and a uniform integrability condition which induces a restriction on the tail behaviour of the underline distribution of innovations. For asymptotic theory analysis purposes related conditions are imposed to the noise sequences $\varepsilon_t$ such that the corresponding partial sum processes lie in the domain of attraction of functionals of Brownian motions. Thus asymptotic approximations are obtained based on the Ornstein-Uhlenbeck process.

Extending and verifying the existence of Cauchy limiting distribution theory in mildly explosive autoregression under various conditions on the innovation structure (stationary conditional heteroscedastic errors, anti-persistence errors - Lui et al. (2021), non-Gaussian errors, serially correlated errors - Lui et al. (2018), errors with possibly infinite variance - Liu et al. (2021) and Wang et al. (2022), mixing innovations - Oh et al. (2018) and Liu et al. (2022)) doesn’t necessarily imply the presence of a unified theory across the spectrum of nonstationarity as per Assumption 1. In other words, the fact that there is a limiting distribution discontinuity across these different regions of the parameter space of $\theta_n$, despite that the separate Gaussian and Cauchy limit results are found to be robust against different properties in the innovation structure, still requires considering whether a unified distribution-free inference technique can be applied regardless of: (i) the presence of a model intercept, (ii) the innovation sequence properties and (iii) the region of the parameter space.

Furthermore, based on extensive empirical applications using the dataset of Härdle et al. (2016), where cross-sectional specific autoregressive and predictive regressions were fitted on stock returns with predictors the macroeconomic variable of the dataset, we have observed that in cases where the estimated autoregressive coefficient is above the unit boundary (i.e., on the mildly explosive side), the estimation method for the model coefficient of the predictive regression model remains unchanged if one uses the existing IVX instrumentation approach proposed by Magdalinos and Phillips (2009) (see, also Kostakis et al. (2015)). This observation motivated us to consider an alternative instrumentation procedure when the estimated autoregressive coefficient is within that region is explosive.
More precisely, the novel instrumentation procedure proposed by Magdalinos and Petrova (2022) does exactly that, while theoretical and empirical illustrations are given for the linear autoregressive and predictive regression models. Specifically, the asymptotic theory of estimators, test statistics and corresponding confidence intervals can be constructed using a common distribution without imposing additional assumptions. These properties are found to hold specifically for autoregressive and predictive regression models based on a conditional mean functional form while the purpose of this paper is to extend those to the case of a conditional quantile functional form.

The class $P.1$ implies that $\theta_n$ is near to the unit boundary under there is no local-to-unity parametrization, thus the limit tends to $-\infty$. For the class $P.2$ we have that $\theta_n \to 1$ when specified as $\theta_n = \left(1 + \frac{c}{k_n}\right)$, where $k_n = n^\gamma$ with $\gamma = 0$, $\gamma = 1$ or $\gamma \in (0, 1)$. Thus, when $\theta_n = \left(1 + \frac{c}{k_n}\right)$,

$$
\zeta_n := \lim_{n \to +\infty} n (\theta_n - 1) = \lim_{n \to +\infty} n \left(1 + \frac{c}{k_n}\right) - 1 = \lim_{n \to +\infty} \left\{\frac{n}{k_n}\right\} c
$$

Assume that $c < 0$ then it holds that

A. If $k_n = n^\gamma$ with $\gamma = 1$ ($k_n = n$), then $\zeta \equiv c \in \mathbb{R}$, which falls in the class of nearly-unstable processes (near-nonstationary).

B. If $k_n = n^\gamma$ with $\gamma \in (0, 1)$, then $\zeta = -\infty$. In this case, although we are in the region of the 'black circle' within the 'blue zone' then we have a mildly integrated process which falls in the class of nearly-stable processes (near-stationary).

C. If $k_n = n^\gamma$ with $\gamma = 0$ ($k_n = 1$), then $\zeta = -\infty$ which falls in the class of nearly-stable processes (near-stationary).

Assume that $c > 0$ then it holds that

A. If $k_n = n^\gamma$ with $\gamma = 1$ ($k_n = n$), then $\zeta \equiv c \in \mathbb{R}$, which falls in the class of nearly-unstable processes (near-nonstationary).

B. If $k_n = n^\gamma$ with $\gamma \in (0, 1)$, then $\zeta = +\infty$. In this case, although we are in the region of the 'black circle' within the 'red zone' then we have a mildly explosive process which falls in the class of nearly-explosive processes.

C. If $k_n = n^\gamma$ with $\gamma = 0$ ($k_n = 1$), then $\zeta = +\infty$ which falls in the class of nearly-explosive processes.

Assume that $c = 0$ then we have a pure unit root process.
Persistence Classes:

I. Pure Stationary Processes: $\rho_n \to \rho \in (-1, 1)$.

II. Near Unit Root Processes: $\rho_n \to \rho = 1$ where $\rho_n$ is specified as $\rho_n = \left(1 + \frac{c}{k_n}\right)$, $k_n = n^\gamma$.

For all the case below we can employ the original IVX instrument.

II.a. Near-stationary processes: $c < 0, \gamma = 1$ (near-integrated)

\[
\lim_{n \to +\infty} \rho_n := \lim_{n \to +\infty} \left(1 + \frac{c}{n}\right) = 1 + \lim_{n \to +\infty} \left\{\frac{c}{n}\right\} = -\infty, \text{ when } c < 0. 
\]

II.b. Mildly-integrated processes: $c < 0, \gamma \in (0, 1)$

\[
\lim_{n \to +\infty} \rho_n := \lim_{n \to +\infty} \left(1 + \frac{c}{n}\right) = 1 + \lim_{n \to +\infty} \left\{\frac{c}{n}\right\} = -\infty, \text{ when } c < 0. 
\]

II.c. Stationary processes: $c < 0, \gamma = 0$

\[
\lim_{n \to +\infty} \rho_n := \lim_{n \to +\infty} (1 + c) = (1 + c) \in \mathbb{R} 
\]

II.a'. Near-explosive processes: $c > 0, \gamma = 1$

\[
\lim_{n \to +\infty} \rho_n := \lim_{n \to +\infty} \left(1 + \frac{c}{n}\right) = 1 + \lim_{n \to +\infty} \left\{\frac{c}{n}\right\} = +\infty, \text{ when } c > 0. 
\]

II.b'. Mildly-explosive process: $c > 0, \gamma \in (0, 1)$

\[
\lim_{n \to +\infty} \rho_n := \lim_{n \to +\infty} \left(1 + \frac{c}{n}\right) = 1 + \lim_{n \to +\infty} \left\{\frac{c}{n}\right\} = +\infty, \text{ when } c > 0. 
\]

II.c'. Stationary process: $c > 0, \gamma = 0$

\[
\lim_{n \to +\infty} \rho_n := \lim_{n \to +\infty} (1 + c) = (1 + c) \in \mathbb{R} 
\]

II.d. Pure Unit Root processes: $c = 0$ regardless if $\gamma = 1$ or $\gamma \in (0, 1)$

\[
\lim_{n \to +\infty} \rho_n := \lim_{n \to +\infty} \left(1 + \frac{0}{n}\right) = 1 \in \mathbb{R}. 
\]

III. Pure Explosive Processes: $\rho_n \to \rho > 1$.

Thus, in this case to construct the IVX instrument we need to convert the process to be mildly explosive, i.e., so that is within the red region.
Example 1. Consider the classification of nonstationarity given by Benke and Pap (2021).

\[ \begin{align*}
X_k &= \beta X_{k-1} + \varepsilon_k, \quad X_0 = 0. 
\end{align*} \] (2.15)

- In the case that \(|\beta| < 1\), the sequence \(\{\hat{\beta}_n\}_{n \in \mathbb{N}}\) is asymptotically normal such that

\[ \sqrt{n} \left( \hat{\beta}_n - \beta \right) \xrightarrow{d} N \left( 0, 1 - \beta^2 \right), \quad \text{as} \quad n \to \infty. \] (2.16)

- In the case that \(|\beta| = 1\), then the sequence \(\{\hat{\beta}_n\}_{n \in \mathbb{N}}\) has a limit distribution that depends on functionals of Brownian motion such that

\[ n \left( \hat{\beta}_n - \beta \right) \xrightarrow{d} \int_0^1 W(r) dW(r) \left/ \int_0^1 W(r)^2 dr \right., \quad \text{as} \quad n \to \infty. \] (2.17)

- A sequence \(\left( X_{\beta^{(n)}} \right)_{k \in \mathbb{N}}\) for \(n \in \mathbb{N}\) of an AR(1) process is called nearly unstable if \(\beta^{(n)} \to \beta\) as \(n \to \infty\), where \(|\beta| = 1\). Moreover, in the case that \(\beta = 1\) and \(\beta^{(n)} = \left( 1 + \frac{c}{n} \right)\) with some \(c \in \mathbb{R}\),

\[ n \left( \hat{\beta}^{(n)}_n - \beta^{(n)} \right) \xrightarrow{d} \int_0^1 J(r) dW(r) \left/ \int_0^1 J(r)^2 dr \right., \quad \text{as} \quad n \to \infty. \] (2.18)

such that \(\{J_c(r)\}_{r \in [0,1]}\) is a continuous AR(1) process, that is, an OU process defined as a unique strong solution of the SDE such that

\[ \begin{align*}
\frac{dJ_c(r)}{r} &= c J_c(r) dr + dW(r), \quad r \in [0,1] \\
Y_c(r) &= 0 
\end{align*} \] (2.19)

Furthermore, if we consider \(c\) to be a parameter instead of \(\beta^{(n)}\), then the LSE of \(c = n \left( \beta^{(n)} - 1 \right)\) is given by the following expression:

\[ \hat{c}_n = n \left( \hat{\beta}^{(n)}_n - 1 \right) = n \left( \hat{\beta}^{(n)}_n - \beta^{(n)} \right) + c, \] (2.20)

\[ \hat{c}_n \xrightarrow{d} \int_0^1 J_c(r) dW(r) \left/ \int_0^1 J_c(r)^2 dr \right. + c = \int_0^1 J_c(r) dJ_c(r) \left/ \int_0^1 J_c(r)^2 dr \right., \] (2.21)

where the limit distribution above turns out to be the maximum likelihood estimator (MLE) of the parameter \(c\) for the expression that gives the stochastic differential equation based on a sample \(\{J_c(r)\}_{r \in [0,1]}\).
Example 2. Consider the following processes:

\[(2.22) \quad x_t = \mu + X_t\]
\[(2.23) \quad X_t = \rho X_{t-1} + \varepsilon_t\]

which implies that

\[(2.24) \quad x_t = \mu (1 - \rho_n) + \rho_n x_{t-1} + \varepsilon_t\]

Moreover, denote with

\[x^\mu_t = (x_t - \bar{x}_t), \quad \bar{x}_t = \frac{1}{n} \sum_{t=1}^{n} x_t, \quad x^\mu_{t-1} = (x_{t-1} - \bar{x}_{t-1}), \quad \bar{x}_{t-1} = \frac{1}{n} \sum_{t=1}^{n} x_{t-1}.\]

Equivalently, for \(\mu \neq 0\), we consider the autoregressive model \(x^\mu_t = \rho_n x^\mu_{t-1} + \varepsilon^\mu_t\), which implies

\[(2.25) \quad \hat{\rho}_n = \left(\frac{n \sum_{t=1}^{n} x^\mu_{t-1}}{n \sum_{t=1}^{n} x^\mu_t x^\mu_{t-1}}\right)^{-1}, \quad \varepsilon^\mu_t = x^\mu_t - \hat{\rho}_n x^\mu_{t-1}\]

Therefore in this article we focus on the quantile autoregressive model with a model intercept the particular modeling framework motivate us to investigate the interplay between near-zero intercept\(^6\) and stochastic persistence by deriving the asymptotic distribution.

Assumption 2.3. The distribution \(F_{\varepsilon}\) 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,

\[(2.26) \quad \mathbb{E}(\varepsilon_1) = 0, \quad \text{Var}(\varepsilon_1) = \sigma^2_{\varepsilon}, \quad \text{and} \quad \mathbb{E}|\varepsilon_1|^{2+m} < \infty, \text{ for some } m > 0.\]

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 \(D[0, 1]\).

Lemma 2.1. Suppose that \(\varepsilon_t\) satisfies Assumption 2.3. Then as \(n \to \infty\) it holds that

\[(2.27) \quad \left(\sum_{i=1}^{[nr]} \frac{\varepsilon_i}{a_n}, \sum_{i=1}^{[nr]} \frac{\varepsilon_i^2}{a_n^2}\right) \Rightarrow \left(U_\alpha(r), V_\alpha(r)\right)\]

where \(\left(U_\alpha(r), V_\alpha(r)\right)\) is a Levy process in \(D[0, 1]^2\) with index \(\alpha \in (0, 2)\).

\(^6\)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)).
Remark 2.2. 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 $V_{\alpha}(r) = \int_0^r (dU_{\alpha}(s))^2 = [U_{\alpha}, U_{\alpha}]_{r \in [0,1]}$ which is the quadratic variation of the Levy process $U_{\alpha}(r)$. Furthermore, $V_{\alpha}(r)$ is a stochastic integral. 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)).

Any asymptotic results followed by Lemma 2.1 coincide with the standard random walk asymptotics for finite variance models (we focus in the case $\alpha = 2$). Then as $n \to \infty$, it holds that

\begin{equation}
(2.28) \quad n(\hat{\rho}_{n,c} - 1) \Rightarrow \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W^2(r)dr}.
\end{equation}

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 $-n^{\gamma} < c < 0$. The OLS estimate $\hat{\rho}$ is $n-$consistent, that is, $n(\hat{\rho} - \rho)$ 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 the OLS estimate of the coefficient of persistence $c$, which is not consistent. Mikusheva (2012) shows

\begin{equation}
(2.29) \quad (\hat{c} - c) = \frac{\int_0^1 J_c(r)dW(r)}{\int_0^1 J_c^2(r)dr}.
\end{equation}

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. Roughly speaking, 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 we assume that $D[0,1]$ 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. Imposing assumptions regarding the properties of the disturbance term $\varepsilon_t$ 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. Relaxing the Gaussianity assumption requires to apply semiparametric estimation methodologies 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.
**2.2 Quantile Conditional Estimation**

In this Section we discuss in more details the estimation procedure for the quantile autoregressive time series that accommodates the parametrization of the autoregression coefficient with respect to moderate deviations from the unit boundary. We first introduce the quantile estimation method\(^7\) to obtain parameter estimates and then establish the asymptotic theory of this estimator for the autoregressive model denoted with \(y_t = \mu + \rho y_{t-1} + u_t\) (with a slight change of notation).

Denote with \(\mu(\tau)\) and \(\rho_n(\tau)\) to be the \(\tau\)-quantile dependent parameters, which are determined based on a conditional quantile functional form as below:

\[
Q_{y_t}(\tau | F_{t-1}) := F_{y_t|x_{t-1}}^{-1}(\tau) \equiv \mu(\tau) + \rho_c(\tau)y_{t-1}.
\]

\[
F_{y_t|x_{t-1}}(\tau) := \mathbb{P}(y_t \leq Q_{y_t}(\tau | F_{t-1}) | F_{t-1}) \equiv \tau.
\]

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

Denote the parameter vector with \(\theta(\tau) = (\mu(\tau), \rho(\tau))^\top\) and \(X_t = D_n^{-1}(1, y_{t-1})\), where \(D_n\) 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

\[
(2.32) \quad \hat{\theta}_n(\tau) := \arg \min_{\theta(\tau)} \sum_{t=1}^{n} \varphi_{\tau}(y_t - \theta(\tau)^\top X_t).
\]

such that \(\hat{\theta}_n(\tau) \equiv D_n(\bar{\mu}_n(\tau) - \mu(\tau), \bar{\rho}_{n,c}(\tau) - \rho_c(\tau))\). Moreover, we denote with \(\psi(u)\) to be the left derivative of \(\varphi(u)\). In particular, when \(\psi(u) := u\) is the identity function then the \(\hat{\theta}_n\) corresponds to the least squares estimator, while when \(\psi(u) := (1 - \tau)\) for \(u \leq 0\) and \(\psi(u) = \tau\) for \(u > 0\), then it corresponds to the optimization function and \(\hat{\theta}_n\) is the quantile-dependent estimator.

**Assumption 2.4.** Suppose that \(\mathbb{E}[\psi(u_1(\tau))] = 0\) and consider the random variable which corresponds to the first derivative around some parameter \(\theta \in \mathbb{R}\) such that

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

where \(\mathbb{E}[\psi(u_1(\tau))]^{2+m} < \infty\), for some \(m > 0\).

Assumption 2.4 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 2.5 below.

\(^7\)A complete treatment of limit results for quantile regressions can be found in the book of Koenker (2005).
Assumption 2.5. The following conditions for the innovation sequence hold:

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

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

(ii) For each \( t \) and \( \tau \in (0,1) \), \( f_{\varepsilon_t(t),t-1(.)} \) 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)).

Remark 2.3. Assumption 2.5 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, 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 (2.32) we apply the Taylor expansion to the check function, such that for a given parameter \( \delta(\tau) \) it holds that

\[
\varrho_t \left( \varepsilon_t - \delta(\tau)^\top X_t \right) = \varrho_t(\varepsilon_t) - \delta(\tau)^\top \psi(\varepsilon_t) + \varphi_t(\delta(\tau)).
\]  

Remark 2.4. 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 \( \lim_{n \to \infty} n (1 - \rho_n) = 0 \), the nearly unstable model behaves asymptotically like the strictly unstable model in which case \( \rho = 1 \).

2.3 Large-Sample 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. Using the local-to-unity parametrization in autoregressive processes overcomes this gap by considering the limiting distribution\(^8\) for the whole parameter space regardless of the existence of a limit singularity.

\(^8\)The form of the noncentrality parameter of the \( \chi^2 \)-distribution depends on the initial condition and the form of the autoregressive parameter of the model (see, Theorem 2.2. of Hui et al. (2022)).
Using the local-to-unity asymptotics our aim is to derive a nuisance-parameter-free limiting distribution which can facilitate statistical inference. Therefore, by decomposing the underline stochastic processes into components which include a predictable quadratic variation, 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 is induced by the predictable quadratic variation ensures that the self-normalization property holds. When asymptotic theory of the autoregressive parameter that corresponds to a particular region of the parameter space is nonstandard, then this implies that the limit distribution of the MLE of the model can be represented as the MLE of a parameter of a process satisfying the stochastic differential equation for the OU process (see, Benke and Pap (2021)). Thus, the limiting distributions of the model estimator and the corresponding test statistic are functions of the local-to-unity parameter which is not consistently estimable since \((\hat{c} - c) = O_p(1)\).

Although in this article we only consider point inference procedures, in the case of interval inference, that is, when concerned with the construction of confidence interval for unknown model parameters then several studies discuss the possibility of obtaining uniform inference procedures when the nonstationary autoregressive model is expressed using the local-to-unity parametrization (see, Mikusheva (2007), Phillips (2014) and Magdalinos and Petrova (2022)). In this article we focus on bridging the gap in the robust and uniform inference literature on quantile autoregressions and quantile predictive regression models by considering that the instrumentation procedure depends on the region of the parameter space. In other words, although the original IVX method proposed by Phillips and Magdalinos (2009) and Kostakis et al. (2015) it is found to be robust in LUR moderate deviations from the unit boundary, it is clear than in explosive and mildly explosive regimes the same IVX instrumentation might not work so well.

**Theorem 2.1** (Chan et al. (2006)). Assume that Assumption 2.2-2.5 hold and that the autoregression coefficient is expressed as \(\rho_{n,c} = (1 + \frac{c}{k_n})\). Then, the following limit result hold

\[
D_n \left( \hat{\vartheta}_n(\tau) - \vartheta(\tau) \right) \xrightarrow{d} \frac{1}{f_{\varepsilon}(F_{-1}(\varepsilon_t(\tau)))} \Sigma^{-1} \left( W(\tau; 1), \int_0^1 J(s)dW(\tau, s) \right),
\]

\[
\frac{1}{f_{\varepsilon}(F_{-1}(\varepsilon_t(\tau)))} \int_0^1 J_1(s)dW(\tau, s) - W(\tau, 1) \int_0^1 J_1(s)ds \int_0^1 J_2^2(s)ds - \left( \int_0^1 J_1(s)ds \right)^2
\]

where \(D_n = \text{diag} (\sqrt{n}, n)\) and \(\vartheta(\tau) = (\mu(\tau), \rho_{n,c}(\tau))\) such that

\[
\Sigma := \int_0^1 \begin{pmatrix} 1, J_1(s) \end{pmatrix}' \begin{pmatrix} 1, J_1(s) \end{pmatrix} ds \equiv \begin{bmatrix} 1 & \int_0^1 J_1(s)ds \\ \int_0^1 J_1(s)ds & \int_0^1 J_1(s)J_1(s)ds \end{bmatrix}.
\]
(2.39) \[ \left\{ \sum_{t=1}^{n} y_{t-1}^2 - \left( \sum_{t=1}^{n} y_t \right)^2 \right\}^{1/2} \left( \hat{\rho}_{n,c}(\tau) - \rho_{n,c}(\tau) \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{\tau(1-\tau)}{f_\varepsilon^2 \left( F_{\varepsilon}^{-1}(\varepsilon_t(\tau)) \right)} \right). \]

2.3.1 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 autoregression 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 functional form, 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 \( D_n \) and \( B \) as defined below

(2.40) \[ D_n = \begin{pmatrix} \sqrt{n} & 0 \\ 0 & \sqrt{n k_n} \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2 / (-2c) \end{pmatrix} \]

where \( k_n = n^\gamma \) and \( \gamma \in (0, 1) \). The \( n^{-1/2} \) convergence rate corresponds to the model intercept while when \( k_n = n^\gamma \), then the autoregression parameter of the model has a convergence rate of \( n^{-1+2\gamma} \) 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 2.2.** Under Assumptions 2.3-2.5,

(2.41) \[ \left( \hat{\mu}_n, \hat{\rho}_{n,c} \right)^\top = \left( \mu, \rho_{n,c} \right)^\top + \frac{(BD_n)^{-1}}{\xi} \sum_{t=1}^{n} \psi(\varepsilon_t) X_t^\top + o_p(1). \]

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

(2.42) \[ \begin{pmatrix} \hat{\mu}_n(\tau) \\ \hat{\rho}_{c,n}(\tau) \end{pmatrix} = \begin{pmatrix} \mu_n(\tau) \\ \rho_{c,n}(\tau) \end{pmatrix} + \frac{(BD_n)^{-1}}{f_\varepsilon \left( F_{\varepsilon}^{-1}(\tau) \right)} \sum_{t=1}^{n} \left( \tau - 1 \{ \varepsilon_t \leq F_{\varepsilon}^{-1}(\varepsilon_t) \} \right) \frac{1}{\sqrt{n}} \frac{y_{t-1}}{\sqrt{n k_n}} + o_p(1). \]

where \( f_\varepsilon(x) \) and \( F_\varepsilon(x) \) denote the probability and cumulative density functions of \( \varepsilon_1 \), respectively.
Remark 2.5. Theorem 2.2 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 2.3. Under Assumptions 2.3-2.5,

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

In particular, it follows that

(i) If \( \psi(u) = (\tau - 1 \{u \leq 0\}) \), and the pdf \( f(u) \) of \( \varepsilon_1 \) exists and satisfies \( f_{\varepsilon}(F_{\varepsilon}^{-1}(\tau)) > 0 \), then

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

(ii) If \( \psi(u) = u \), then

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

Remark 2.6. The limit results given by Theorem 2.2 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 \( B \) 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 \( B_a = \{ \rho_n \in B \| \beta - \beta(\tau) \| \leq a_n \} \) where \( \{a_n\} \) is some positive sequence.
2.3.2 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

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

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 parametrization 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

\[
\frac{1}{2c} n^\gamma \theta^n \left( \hat{\theta}_n - \theta \right) \Rightarrow C, \quad \text{as } n \to \infty.
\]

On the other hand, for the pure explosive root case such that \( |\theta| > 1 \) then, the limit distribution of \( \left( \hat{\theta} - \theta \right) \) 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 \( \theta \) 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.

\[
D_n = \begin{pmatrix} \sqrt{n} & 0 \\ 0 & \rho_{n,c}^{\alpha} \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 0 & \mathcal{Z}_n^2 / (2c) \end{pmatrix}
\]
where $Z_3$ is some normal random variable to be defined below.

**Theorem 2.4.** Under the Assumptions 2.2-2.5 it holds that,

$$
(\hat{\mu}, \hat{\rho}_{n,c})^\top = (\mu, \rho_{n,c})^\top + \frac{\left(\sum_{t=1}^n X_t X_t^\top D_n\right)^{-1}}{\xi} \sum_{t=1}^n \psi(\varepsilon_t) X_t^\top + o_p(1).
$$

**Theorem 2.5.** Under Assumptions 2.2-2.5 it holds that,

$$
D_n \left( (\hat{\mu}, \hat{\rho}_{n,c}) - (\mu, \rho_{n,c}) \right)^\top \xrightarrow{d} \frac{1}{\xi} B_n^{-1} \left( Z_1, Z_2, Z_3 \right)^\top.
$$

In particular, it follows that

1. If $\psi(u) = \left( \tau - 1 \{ u \leq 0 \} \right)$, and the pdf $f(u)$ of $\varepsilon_1$ exists and satisfies $f_\varepsilon(F_\varepsilon^{-1}(\tau)) > 0$, then

$$
\frac{\hat{\rho}_{n,c}(\tau) - \rho_c(\tau)}{\rho_n k_n} \xrightarrow{d} \frac{2c}{f_\varepsilon(F_\varepsilon^{-1}(\tau))} Z_2.
$$

2. If $\psi(u) = u$, then

$$
\frac{\hat{\rho}_{n,c}(\tau) - \rho_c(\tau)}{2c \rho_n^2 k_n} \xrightarrow{d} \frac{Z_2^*}{Z_3^*}.
$$

Therefore, Theorem 2.5 verifies that we indeed obtain the equivalent asymptotic theory results in comparison to the linear autoregressive time series model. Specifically, the autoregression 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 2.7.** As we see from Theorem 2.5, 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 $\rho_n$ by $\rho_{n,c} = \left( 1 + \frac{c}{k_n} \right)$ we obtain that $(\rho^2 - 1) = \frac{2c}{k_n} \left[ 1 + o(1) \right]$. 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 $\varepsilon_t$ 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 2.5 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 $\gamma' > 1$. 

18
2.4 Testing Linear Hypotheses

Consider the autoregressive model

\begin{equation}
    y_t = \rho y_{t-1} + \varepsilon_t, \quad t = 1, \ldots, n,
\end{equation}

such that \( \rho \in [-1, 1] \), is within the stationary region. Then, the usual testing hypothesis of interest is such that, \( H_0 : \rho = \rho_0 \). In particular, Dickey and Fuller (1979) showed that the finite sample distribution for \( \rho \) in the neighbourhood of unity is very close to the asymptotic unit root case, under the assumption that the error term \( \varepsilon_t \) 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 \( T_n(\hat{\rho}) \) can be represented by functionals of Wiener processes (see, Dickey and Fuller (1979) and Buchmann and Chan (2007)). Thus, the asymptotic distribution of the t-statistic based on M-estimators, denoted by \( T_{\psi(\tau)}(\hat{\rho}) \) depends on the nuisance parameter \( \delta \), that is, the correlation between the innovations \( \{\varepsilon_t\} \) and the pseudo-score function \( \psi(\varepsilon_t) \) 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.

The t-statistic for the null hypothesis \( H_0 : \rho = \rho_0 \) is given by

\begin{equation}
    t_{\psi} = \frac{\left( \hat{\rho}_n(\tau) - \rho_n(\tau) \right)}{\left\{ \left( n^{-1} \sum_{t=1}^{n} \psi_{\tau}(y_t - \hat{\rho}_n(\tau)y_{t-1})^2 \right) / \left( n^{-1} \sum_{t=1}^{n} \psi_{\tau}'(y_t - \hat{\rho}_n(\tau)y_{t-1})^2 \right) \right\}^{1/2}},
\end{equation}

where \( \psi_{\tau}'(.) \) denotes the first derivative of the function \( \psi_{\tau}(.) \).
3 Asymptotic theory for Quantile Predictive Regression

In this section we unify the theory for the quantile predictive regression model while presenting the corresponding asymptotic theory for the quantile autoregressive process we introduced in the previous section with our novel pruned-based endogenous instrumentation approach.

3.1 Model specification and assumptions

Consider the following predictive regression model

\begin{align}
\ y_t &= \alpha + \beta_n x_{t-1} + \varepsilon_t \\
\ x_t &= \mu + \phi_n x_{t-1} + u_t \\
\end{align}

Various studies in the literature consider inference methodologies for nearly-integrated processes. Specifically for the conditional quantile functional form Lee (2016) propose a framework for inference in quantile predictive regression models using the IVX instrumentation of Phillips and Magdalinos (2009). However, robust inference in the near-explosive region for both the quantile autoregression and the predictive regression has not been examined previously in the literature.

In particular, OLS-based inference on $\beta_n$ for nearly-explosive processes and specifically purely-explosive suffers from the same problem as OLS-based inference on $\rho_n$, with standard inference applying only under i.i.d Gaussian innovations $\varepsilon_t$. Therefore, the inference procedure proposed in this paper for the quantile-dependent coefficient of $\beta$, in the quantile predictive regression model can accommodate regressions with time series properties along the entire spectrum of autoregressive processes. Although our setting considers a model with only one regressor, therefore further theory is needed in the case of multiple regressors that correspond to either the same or different persistence class. Specifically, one can establish its asymptotic validity uniformly over the autoregressive regime and regardless of the distributional assumptions of the innovations $\varepsilon_t$ and $u_t$. Robust and uniform inference in quantile predictive regression models has been studied by Maynard et al. (2023), Lee (2016), Fan and Lee (2019), Cai et al. (2022) and more recently Liu et al. (2023).

The main idea with the proposed endogenously generated instrumentation is to decompose the the two mutually disjoint parameter space regions and then develop an asymptotic theory for both these regions. In other words, the main intuition behind this is that although the dependence structure of innovation sequences is not affected by considering two different regions of the parameter space, it does however have an effect on the asymptotic behaviour of the corresponding estimators which bridge the gap in both the explosive as well as the near-nonstationary cases. The proposed instrumentation method is based on the framework of Magdalinos and Petrova (2022) and combines the nearly stable with the nearly explosive processes (near stationary/near explosive), thereby implying an adaptive and uniform inference technique for quantile autoregressions and quantile predictive regression models regardless of the peristence properties of regressors or whether the autoregressive and predictive equations include a model intercept.
3.2 Pruned-based endogenous instrumentation approach

In this paper we propose a novel pruned-based endogenous instrumentation approach based on the original instrumentation methodology proposed in the paper of Phillips and Magdalinos (2009) (see, also Kostakis et al. (2015)). The procedure we propose is similar to the IV instrument proposed in the recent paper of Magdalinos and Petrova (2022), although our motivation is to unify inference in autoregressions when considering a conditional quantile functional form, that is, is applied specifically to quantile autoregressive processes. We call our novel estimator IVX-P which can be employed in either linear or quantile conditional functional forms with univariate or multivariate regressors. In the literature such estimators mainly were concerned with Stein-type estimators, however our study is the first to consider an estimator which is obtained using properties of the underline stochastic processes both within the admissible parameter space as well as outside the usual parameter space.

Thus, the proposed pruned-based instrumental estimator implies a data driven instrument selection such that $B_n = 1 \{ n \left( \hat{\vartheta}^{ols} - 1 \right) \leq 0 \}$. More specifically, the chosen $\rho_{nz}$ is such that

\begin{equation}
\vartheta_{nz} = \varphi_{1n} 1 \{ B_n \} + \varphi_{2n} 1 \{ B_n^c \}
\end{equation}

\begin{equation}
\varphi_{1n} = \left( 1 - \frac{1}{\kappa_n} \right), \quad \varphi_{2n} = \left( 1 + \frac{1}{\kappa_n} \right)
\end{equation}

where $\kappa_{1n}, \kappa_{2n} \to \infty$ with $\kappa_{1n}/n \to 0$ and $\kappa_{2n}/n \to 0$. Then, the combined instrument process is

\begin{equation}
z_t = \vartheta_{nz} z_{t-1} + \tilde{u}_t, \quad \text{where} \quad \tilde{u}_t = \Delta x_t 1 \{ B_n \} + \hat{u}_t 1 \{ B_n^c \}
\end{equation}

In other words, we have an orthogonal decomposition such that $z_t = z_{1t} 1 \{ B_n \} + z_{2t} 1 \{ B_n^c \}$ where

\begin{equation}
z_{1t} = \varphi_{1n} z_{1t-1} + \Delta x_t
\end{equation}

\begin{equation}
z_{2t} = \varphi_{2n} z_{2t-1} + \hat{u}_t
\end{equation}

Therefore, for all cases it holds that

\begin{equation}
n \left( \hat{\vartheta}^{ols} - 1 \right) = c n \frac{1}{\kappa_n} \left( 1 + \varepsilon_n \right), \quad \varepsilon_n \overset{p}{\to} 0
\end{equation}

\begin{equation}
\overset{p}{\to} \text{sign}(c) \times \infty
\end{equation}

where $z_t^{IVX-P}$ represents the IVX pruned-based estimator. A key result that we are aiming to illustrate is the Asymptotic Mixed Gaussianity (AMG) property of the IVX-P estimator.
3.2.1 Instrument Construction

Successful instrumentation based on a combined near-stationary/near-explosive process requires statistical information separating the near-stationary autoregressive class from the near-explosive class asymptotically. In other words, the advantage of the inference procedure proposed in this paper over existing procedures is that it is valid for any \( \rho_n \rightarrow \rho \in (0, +\infty) \), which includes all three parameter regions of interest of empirical interest. Practical implementation of our instrumentation procedure requires a choice for \( \phi \) parameter regions of interest. Specifically, choosing \((\phi_{1n})_{n \in \mathbb{N}}\) and \((\phi_{2n})_{n \in \mathbb{N}}\) with \( n (\phi_{1n} - 1) \rightarrow \infty \) as \( n \rightarrow \infty \) and \( \phi_{2n} \rightarrow 1 \) with \( n (\phi_{1n} - 1) \rightarrow +\infty \).

\[
(3.9) \quad \phi_{1n} = \left(1 - \frac{1}{n^{\gamma_1}}\right) \quad \text{and} \quad \phi_{2n} = \left(1 + \frac{1}{n^{\gamma_2}}\right)
\]

reduces to the problem of selecting the values for \( \gamma_1 \) and \( \gamma_2 \).

We construct the instrumentation procedure based on a min-max optimality guarantee. In other words, we have that \( \tilde{z}_{1t} \) can be asymptotically approximated by a near-stationary process such that

\[
(3.10) \quad \tilde{z}_{1t} = \phi_{1n} z_{1t-1} + u_t = \sum_{j=1}^{t} \phi_{1n}^{t-j} u_j.
\]

In particular, when \( \rho_n \) is closer to 1 that \( \phi_{1n} \) then \( \tilde{z}_{1t} \) reduces asymptotically to the original process \( x_t \). Furthermore, the instrument \( \tilde{z}_{2t} \) is always approximated by a mildly explosive process such that

\[
(3.11) \quad z_{2t} = \phi_{2n} z_{2t-1} + u_t = \sum_{j=1}^{t} \phi_{2n}^{t-j} u_j.
\]

Therefore, from the above decomposition we see that sample moments involving the near-stationary instrument \( \tilde{z}_{1t} \) will contribute asymptotically when the original process \( x_t \) belongs to the persistence classes \( P.1-P.2 \) (i.e., near stationary and near-nonstationary), whereas sample moments involving the mildly-explosive instrument \( \tilde{z}_{2t} \) will make an asymptotic contribution for autoregressions that belong to persistence classes \( P.2-P.3 \).

Moreover, under Assumption 4 we denote the autocovariance function and long-run variance of \( (u_t) \) by \( \gamma_u(\cdot) \) and \( \omega^2 = \sum_{k=-\infty}^{+\infty} \gamma_u(k) = C(1)^2 \sigma^2 \) respectively and let

\[
(3.12) \quad \gamma_n = \sum_{k=1}^{+\infty} \rho_n^{k-1} \gamma_u(k) \quad \text{and} \quad \Gamma = \sum_{k=1}^{+\infty} \rho^{k-1} \gamma_u(k).
\]

Thus, we have that \( \rho_N \rightarrow \rho \) and \( \Gamma = \lim_{n \rightarrow \infty} \Gamma_n \) exists by the dominated convergence theorem since \( \sum_{k=1}^{+\infty} |\gamma_u(k)| < +\infty \). Notice that when \( \rho = 1, \Gamma = \sum_{k=1}^{+\infty} \gamma_u(k) \) is the one-sided long-run covariance of \( (u_t)_{t \in \mathbb{N}} \). Denote with \( W(t) \) to be the standard Brownian motion on \([0, 1]\) and \( B(t) = \omega W(t) \) and define the Ornstein-Uhlenbeck processes below

\[
(3.13) \quad W_c(t) = \int_{0}^{t} e^{c(t-s)} dW(s) \quad \text{and} \quad J_c(t) = \int_{0}^{t} e^{c(t-s)} dB(s)
\]
4 Conclusion

In this paper we consider the asymptotic theory for moderate deviation from the unit boundary in quantile autoregressive and quantile predictive regression models. Using the moderate deviation principles we unify the asymptotic theory with a modified endogenous instrumentation procedure, without inducing 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.

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) as is presented by Katsouris (2023).
### A  Moderate Deviations in Mildly Integrated and Explosive Cases

Generally, it is found that quantile estimates are more robust than the ordinary least squares estimate when the underlying series is heavy-tailed. Strictly speaking, the moderate deviations from unity process is a triangular array such that \( \{y_{nt} : 1 \leq t \leq n\} \). Notice that we consider the quantile estimation in the present paper and provide the asymptotic distribution for the quantile estimate. Following the framework presented by Wang et al. (2022). In particular, consider the first-order autoregressive model as below

\[
x_t = \rho_n x_{t-1} + u_t, \quad t = 1, ..., n, \quad \rho_n = \left(1 + \frac{c}{k_n}\right),
\]

We consider the truncated second moment \( \ell(x) = E[u_1^2 \mathbb{1}_{\{|u_1| \leq x\}}] \) is a slowly varying function of \( x \) at \( \infty \), which implies that \( \lim_{x \to \infty} \ell(tx)/\ell(x) = 1 \) for all \( t > 0 \).

**Mildly Integrated Case:**

The following vector weak convergence holds:

\[
\left(1 \sqrt{\frac{1}{n}} \sum_{t=1}^{n} \psi_{\tau}(u_{t\tau}), \frac{1}{\sqrt{n k_n \ell(\eta_n)}} \sum_{t=1}^{n} x_{t-1} \psi_{\tau}(u_{t\tau}) \right) \Rightarrow (S(\tau), T(\tau)),
\]

where

\[
S(\tau) \sim \mathcal{N}(0, \tau(1 - \tau)), \quad \text{and} \quad T(t) \sim \mathcal{N}\left(0, -\frac{\tau(1 - \tau)}{2c}\right)
\]

are independent random variables.

**Proof.** Therefore, it suffices to show that for any \( a, b \in \mathbb{R} \),

\[
\frac{a}{\sqrt{n}} \sum_{t=1}^{n} \psi_{\tau}(u_{t\tau}) + \frac{b}{\sqrt{n k_n \ell(\eta_n)}} \sum_{t=1}^{n} x_{t-1} \psi_{\tau}(u_{t\tau}) \Rightarrow aS(\tau) + bT(\tau)
\]

\[
= \mathcal{N}\left(0, a^2\tau(1 - \tau) - \frac{b^2}{2c}\tau(1 - \tau)\right).
\]

Therefore, we denote with

\[
\sum_{t=1}^{n} \varepsilon_{nt} := \frac{a}{\sqrt{n}} \sum_{t=1}^{n} \psi_{\tau}(u_{t\tau}) + \frac{b}{\sqrt{n k_n \ell(\eta_n)}} \sum_{t=1}^{n} x_{t-1} \psi_{\tau}(u_{t\tau})
\]

where

\[
\varepsilon_{nt} := \frac{a}{\sqrt{n}} \psi_{\tau}(u_{t\tau}) + \frac{b}{\sqrt{n k_n \ell(\eta_n)}} x_{t-1} \psi_{\tau}(u_{t\tau}), \quad 1 \leq t \leq n.
\]
Then $\varepsilon_{nt}$ is a martingale difference with respect to the filtration $\mathcal{F}_{nt} = \sigma(y_0, u_1, ..., u_t)$ and therefore the conditional variance of the martingale array $\sum_{t=1}^{n} \varepsilon_{nt}$ is given by

$$
\sum_{t=1}^{n} \mathbb{E} \left[ \varepsilon_{nt}^2 | \mathcal{F}_{nt-1} \right] = \sum_{t=1}^{n} \left( \frac{a}{\sqrt{n}} \tau(1 - \tau) + \frac{b}{\sqrt{nk_n \ell(\eta_n)}} x_{t-1}^2 \tau(1 - \tau) + \frac{2ab}{n \sqrt{k_n \ell(\eta_n)}} x_{t-1} \tau(1 - \tau) \right)
$$

$$
\xrightarrow{p} a^2 \tau(1 - \tau) - \frac{b^2}{2c} \tau(1 - \tau),
$$

which holds since $\mathbb{E} \left[ \psi^2(\tau u_{\tau}) \right] = \tau(1 - \tau)$ and $\frac{1}{nk_n \ell(\eta_n)} \sum_{t=1}^{n} x_{t-1}^2 \xrightarrow{p} \frac{1}{2c}$ and due to the fact that

(A.6) \[ \frac{1}{n \sqrt{k_n \ell(\eta_n)}} \sum_{t=1}^{n} x_{t-1} = o_p(1). \]

\[ \square \]

**Proof.**

$$
\frac{1}{n \sqrt{k_n \ell(\eta_n)}} \sum_{t=1}^{n} x_{t-1} = o_p(1).
$$

By expanding the expression we obtain

$$
\frac{1}{n \sqrt{k_n \ell(\eta_n)}} \sum_{t=1}^{n} x_{t-1} = \frac{x_0}{n \sqrt{k_n \ell(\eta_n)}} \sum_{t=1}^{n} \rho_{t-1}^{1} + \frac{1}{n \sqrt{k_n \ell(\eta_n)}} \sum_{t=1}^{n} \sum_{j=1}^{t-1} \rho_{t-1-j}^{1} u_{j} = K_{n1} + K_{n2}.
$$

It is easy to show that $K_{n1} = o_p(1)$, therefore we need to show that $K_{n2} = o_p(1)$. Thus, by adopting the truncation approach, we rewrite the expression as below

(A.7) \[ K_{n2} = \frac{1}{n \sqrt{k_n \ell(\eta_n)}} \sum_{t=1}^{n} \sum_{j=1}^{t-1} \rho_{t-1-j}^{1} u_{j}^{(1)} + \frac{1}{n \sqrt{k_n \ell(\eta_n)}} \sum_{t=1}^{n} \sum_{j=1}^{t-1} \rho_{t-1-j}^{1} u_{j}^{(1)} u_{j}^{(2)} := I + II. \]

We have that

$$
\mathbb{E} \left[ I^2 \right] = \frac{1}{n^2 k_n \ell(\eta_n)} \mathbb{E} \left[ \sum_{j=1}^{n-1} \sum_{t=j+1}^{n} \rho_{t-1-j}^{1} u_{j}^{(1)} \right]^2
$$

$$
= \frac{1}{n^2 k_n \ell(\eta_n)} \sum_{j=1}^{n-1} \left( \sum_{t=j+1}^{n} \rho_{t-1-j}^{1} \right)^2 \ell(\eta_n) \left( 1 + o(1) \right)
$$

$$
\leq \frac{A}{n^2 k_n} \sum_{j=1}^{n} \left( 1 - \rho_{n-j}^{1} \right) \leq \frac{A}{n^2 k_n (1 - \rho_n)^2} = O \left( \frac{k_n}{n} \right) = o(1).
$$

25
and

\[
E [|II|] \leq \frac{2}{n\sqrt{k_n \ell(\eta_n)}} o\left(\frac{\ell(\eta_n)}{\eta_n}\right) \sum_{t=1}^{n-1} \sum_{j=1}^{n-1} \rho_n^{t-j} \leq \frac{2}{n\sqrt{k_n \ell(\eta_n)}} o\left(\frac{\ell(\eta_n)}{\eta_n}\right) \frac{n}{1 - \rho_n} = o\left(\sqrt{\frac{k_n}{n}}\right) = o(1),
\]

By using the martingale limit theorem, the proof is completed if the Linderberg condition

\[
\sum_{t=1}^{n} E \left[\varepsilon_{nt}^2 1\{|\varepsilon_{nt}| > \eta\} \mid \mathcal{F}_{nt-1}\right] = o_p(1), \quad \text{for any } \eta > 0.
\]  
(A.8)

We have that, for all \(\eta > 0\)

\[
\sum_{t=1}^{n} E \left[\varepsilon_{nt}^2 1\{|\varepsilon_{nt}| > \eta\} \mid \mathcal{F}_{nt-1}\right] = \sum_{t=1}^{n} E \left[\left(\frac{a}{\sqrt{n}} \psi_T(\mu_T) + \frac{b}{\sqrt{n k_n \ell(\eta_n)}} x_{t-1} \psi_T(\mu_T)\right)^2 \{\varepsilon_{nt} > \eta\} \mid \mathcal{F}_{nt-1}\right] \leq 2 \sum_{t=1}^{n} E \left[\left(\frac{a^2}{n} \psi_T^2(\mu_T) + \frac{b}{n k_n \ell(\eta_n)} x_{t-1}^2 \psi_T^2(\mu_T)\right) \{\varepsilon_{nt} > \eta\} \mid \mathcal{F}_{nt-1}\right] \leq \left(2a^2 \tau(1 - \tau) + 2b^2 \tau(1 - \tau) \cdot \frac{1}{n k_n \ell(\eta_n)} \sum_{t=1}^{n} \sum_{t=1}^{n} y_{t-1}^2\right) \cdot \max_{1 \leq t \leq n} E \left[1\{|\varepsilon_{nt}| > \eta\} \mid \mathcal{F}_{nt-1}\right].
\]

Since it holds that \(\frac{1}{n k_n \ell(\eta_n)} \sum_{t=1}^{n} x_{t-1}^2 = O_p(1)\), then it suffices to show that

\[
\max_{1 \leq t \leq n} E \left[1\{|\varepsilon_{nt}| > \eta\} \mid \mathcal{F}_{nt-1}\right] = o_p(1).
\]  
(A.9)

Next, by applying the Chebyshev inequality we obtain that

\[
E \left[1\{|\varepsilon_{nt}| > \eta\} \mid \mathcal{F}_{nt-1}\right] \leq \frac{E \left[\varepsilon_{nt}^2 |\mathcal{F}_{nt-1}\right]}{\eta^2} \leq \frac{2a^2 \tau(1 - \tau)}{n \eta^2} + \frac{2b^2 \tau(1 - \tau)}{n \eta^2} \cdot \frac{x_{t-1}^2}{n k_n \ell(\eta_n)}.
\]  
(A.10)

Then the previous expression holds since \(\max_{1 \leq t \leq n} \frac{x_{t-1}}{\sqrt{n k_n \ell(\eta_n)}} = o_p(1)\).
Proof of Theorem 3.1:

Using the Taylor’s expansion we have that:

\[
\sum_{t=1}^{n} \mathbb{E} \left[ \xi_t(\mathbf{v}) \right] = \sum_{t=1}^{n} \int_{0}^{\mathbf{v}^\top D_n^{-1} x_t} \mathbb{E} \left[ 1 \{ u_t \leq \beta(\tau) + s \} - 1 \{ u_t \leq \beta(\tau) \} \right] ds \\
= \sum_{t=1}^{n} \int_{0}^{\mathbf{v}^\top D_n^{-1} x_t} \{ F(\beta(\tau) + s) - F(\beta(\tau)) \} ds \\
= \sum_{t=1}^{n} \int_{0}^{\mathbf{v}^\top D_n^{-1} x_t} \left\{ s \cdot f(\beta(\tau)) + \frac{1}{2} s^2 \cdot f'(s^*) \right\} ds \\
= \frac{f(\beta(\tau))}{2} \cdot \mathbf{v}^\top \left( \sum_{t=1}^{n} D_n^{-1} x_t x_t^\top D_n^{-1} \right) \mathbf{v} + \frac{1}{2} \sum_{t=1}^{n} \int_{0}^{\mathbf{v}^\top D_n^{-1} x_t} s^2 \cdot f'(s^*) ds,
\]

where \( s^* \in (\beta(\tau), \beta(\tau) + s) \). We have that

\[
(A.11) \quad \sum_{t=1}^{n} D_n^{-1} x_t x_t^\top D_n^{-1} = \begin{pmatrix} 1 & \sum_{t=1}^{n} x_{t-1} x_t \\ \sum_{t=1}^{n} x_{t-1} x_t & \frac{1}{nk_n(\eta)} \sum_{t=1}^{n} x_{t-1}^2 \end{pmatrix} \Rightarrow \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & -\frac{1}{2c} \end{pmatrix}.
\]

Hence, we can prove that

\[
(A.12) \quad \sum_{t=1}^{n} \mathbb{E} \left[ \xi_t(\mathbf{v}) | \mathcal{F}_{nt-1} \right] = \frac{f(\beta(\tau))}{2} \cdot \mathbf{v}^\top \Sigma + o_p(1).
\]

Notice that \( \sum_{t=1}^{n} \left( \xi_t(\mathbf{v}) - \mathbb{E} \left[ \xi_t(\mathbf{v}) | \mathcal{F}_{nt-1} \right] \right) \) is a martingale difference sequence and since it also holds that \( \sum_{t=1}^{n} \mathbb{E} \left[ \xi_t^2(\mathbf{v}) | \mathcal{F}_{nt-1} \right] \) then it follows that \( \sum_{t=1}^{n} \left( \xi_t(\mathbf{v}) - \mathbb{E} \left[ \xi_t(\mathbf{v}) | \mathcal{F}_{nt-1} \right] \right) = o_p(1). \)

Therefore, it holds that

\[
(A.13) \quad Z_n(\mathbf{v}) := -\mathbf{v}^\top R_n(\tau) + \frac{f(\beta(\tau))}{2} \cdot \mathbf{v}^\top \Sigma + o_p(1).
\]

Moreover, the following joint weakly convergence results for the two functionals:

\[
(A.14) \quad R_n(\tau) := \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_\tau(u_{tr}), \frac{1}{\sqrt{nk_n(\eta)}} \sum_{t=1}^{n} y_{t-1} \psi_\tau(u_{tr}) \right)^\top \Rightarrow (S(\tau), J(\tau))^\top.
\]

Thus, it holds that

\[
(A.15) \quad Z_n(\mathbf{v}) := -\mathbf{v}^\top \cdot (S(\tau), J(\tau))^\top + \frac{f(\beta(\tau))}{2} \cdot \mathbf{v}^\top \Sigma.
\]

Since \( Z_n(\mathbf{v}) \) has convex sample paths, it implies uniform convergence on compact sets.
Proof of Theorem 4.1:

By the Crámer-Wold principle, it suffices to show for any \( a, b, c \in \mathbb{R} \),

\[
\alpha J_n(\tau) + \beta K_n(\tau) + \gamma L_n \Rightarrow \alpha J(\tau) + \beta K(\tau) + \gamma L
\]

In particular, we rewrite

\[
\alpha J_n(\tau) + \beta K_n(\tau) + \gamma L_n \\
= \sum_{t=1}^{n} \left[ \left( \frac{\alpha}{\sqrt{n}} + \frac{\beta}{\sqrt{k_n}} \rho_n^{-(n-t)-1} \right) \psi_\tau(u_{t\tau}) + \frac{\gamma}{\sqrt{k_n \ell(\eta_n)}} \sum_{t=1}^{n} \rho_n^{-t} u_t^{(2)} \right] =: R_{n1} + R_{n2}.
\]

Thus, it holds that

\[
\mathbb{E}[R_{n2}] \leq \frac{A}{\sqrt{k_n \ell(\eta_n)}} o \left( \frac{\ell(\eta_n)}{\eta_n} \right) 1 - \rho_n^{-n} \leq Ac \sqrt{k_n} o \left( \frac{\sqrt{\ell(\eta_n)}}{\eta_n} \right) \leq Ac \sqrt{k_n} o \left( \frac{1}{\sqrt{n}} \right) = o(1),
\]

therefore, \( R_{n2} = o_p(1) \). Furthermore, to show that

\[
R_{n1} \overset{d}{\rightarrow} \mathcal{N} \left( 0, \alpha^2 \tau(1 - \tau) + \frac{\beta^2}{2c} \tau(1 - \tau) + \frac{\gamma^2}{2c} \right)
\]

In particular, denote with \( R_{n1} =: \sum_{t=1}^{n} \delta_{nt} \), where

\[
\delta_{nt} = \left\{ \frac{\alpha}{\sqrt{n}} \psi_\tau(u_{t\tau}) + \frac{\beta}{\sqrt{k_n}} \rho_n^{-(n-t)-1} \psi_\tau(u_{t\tau}) + \frac{\gamma}{\sqrt{k_n \ell(\eta_n)}} \sum_{t=1}^{n} \rho_n^{-t} u_t^{(1)} \right\}
\]

Then \( \{\delta_{nt}\} \) are independent random variables with \( \mathbb{E}(\delta_{nt}) = 0 \). Then, the variance of the \( m.d.s \) is

\[
\mathbb{E} \left( \sum_{t=1}^{n} \delta_{nt} \right)^2 = \mathbb{E} \left[ \sum_{t=1}^{n} \left( \frac{\alpha}{\sqrt{n}} \psi_\tau(u_{t\tau}) \right)^2 + \sum_{t=1}^{n} \left( \frac{\beta}{\sqrt{k_n}} \rho_n^{-(n-t)-1} \psi_\tau(u_{t\tau}) \right)^2 + \frac{\gamma^2}{k_n \ell(\eta_n)} \sum_{t=1}^{n} \rho_n^{-2t} u_t^{(1)^2} \right]
\]

\[
+ 2 \mathbb{E} \left[ \sum_{t=1}^{n} \left( \frac{\alpha}{\sqrt{n}} + \frac{\beta}{\sqrt{k_n}} \rho_n^{-(n-t)-1} \right) \psi_\tau(u_{t\tau}) \frac{\gamma}{\sqrt{k_n \ell(\eta_n)}} \sum_{t=1}^{n} \rho_n^{-2t} u_t^{(1)} \right]
\]

\[
= \alpha^2 \tau(1 - \tau) + \beta^2 \tau(1 - \tau) \frac{1}{k_n} \sum_{t=1}^{n} \rho_n^{-2(n-t)-1} + \frac{\gamma^2}{k_n} \sum_{t=1}^{n} \rho_n^{-2t} (1 + o_p(1))
\]

\[
+ 2\alpha \beta \tau(1 - \tau) \cdot \frac{1}{\sqrt{k_n}} \sum_{t=1}^{n} \rho_n^{-(n-t)-1} + \mathbb{E} \left[ \frac{2\alpha \gamma}{n k_n \ell(\eta_n)} \sum_{t=1}^{n} \rho_n^{-t} \psi_\tau(u_{t\tau}) u_t^{(1)} \right]
\]

\[
+ \mathbb{E} \left[ \frac{2\beta \gamma}{k_n \sqrt{n \ell(\eta_n)}} \sum_{t=1}^{n} \rho_n^{-(n+1)} \psi_\tau(u_{t\tau}) u_t^{(1)} \right]
\]

\[
\text{IV}
\]

\[
\text{III}
\]

\[
\text{II}
\]

\[
\text{I}
\]
Notice that we have that

\[
\frac{1}{\sqrt{n k_n^2}} \sum_{t=1}^{n} \rho_n^{-(n-t)-1} = \frac{1}{\sqrt{n k_n^2}} \frac{1 - \rho_n^{-n}}{\rho_n - 1} \to 0. \tag{A.20}
\]

Moreover, by the triangular inequality it holds that

\[
E \left| \psi_T(u_{\tau}) \frac{u_t^{(1)}}{\sqrt{\ell(\eta_n)}} \right| \leq \sqrt{E \left[ \psi_T^2(u_{\tau}) \right]} \sqrt{E \left[ \left( \frac{u_t^{(1)}}{\sqrt{\ell(\eta_n)}} \right)^2 \right]} = O(1), \tag{A.21}
\]

Therefore, it holds that

\[
|III| \leq E \left| \frac{2 \alpha \gamma}{\sqrt{n k_n}} \sum_{t=1}^{n} \rho_n^{-t} \psi_T(u_{\tau}) \frac{u_t^{(1)}}{\sqrt{\ell(\eta_n)}} \right| \leq O(1), \frac{1}{\sqrt{n k_n}} \sum_{t=1}^{n} \rho_n^{-t} = o(1), \tag{A.22}
\]

\[
|IV| \leq E \left| \frac{2 \beta \gamma}{k_n} \sum_{t=1}^{n} \rho_n^{-(n+1)} \psi_T(u_{\tau}) \frac{u_t^{(1)}}{\sqrt{\ell(\eta_n)}} \right| \leq O(1), \frac{n}{k_n} \rho_n^{-(n+1)} = o(1). \tag{A.23}
\]

Hence, it holds that

\[
E \left( \sum_{t=1}^{n} \delta_{nt} \right)^2 = \alpha^2 \tau(1 - \tau) + \frac{\beta^2}{2c} \tau(1 - \tau) + \frac{\gamma^2}{2c} + o_p(1)
\]

In other words, since we have shown that the conditional Lindeberg condition for the martingale difference array \( \sum_{t=1}^{n} \delta_{nt} \) holds and we determine its conditional variance, then the proof of the theorem is completed.

**Proof of Lemma 4.2:**

\[
\rho_n^{-2n+1} \frac{k_n^{-n+1}}{\ell(\eta_n)} \sum_{t=1}^{n} x_{t-1} u_t = \frac{\rho_n^{-n+1} x_0}{\sqrt{k_n \ell(\eta_n)}} \frac{1}{\sqrt{k_n \ell(\eta_n)}} \sum_{t=1}^{n} \rho_n^{-(n-t)-1} u_t + \frac{\rho_n^{-2n+1} x_0}{k_n \ell(\eta_n)} \sum_{t=1}^{n} \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) u_t
\]

\[
= \rho_n^{-2n+1} x_0 \sum_{t=1}^{n} \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) u_t + o_p(1),
\]

since \( x_0 = o_p(\sqrt{k_n}) \) and \( \rho_n^n = o(k_n/n) \) and \( \frac{1}{\sqrt{k_n \ell(\eta_n)}} \sum_{t=1}^{n} \rho_n^{-(n-t)-1} u_t \overset{d}{=} L_n \overset{d}{=} L \). Therefore, it suffices to show that

\[
\frac{\rho_n^{-2n+1}}{k_n \ell(\eta_n)} \sum_{t=1}^{n} \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) u_t = o_p(1). \tag{A.24}
\]
Moreover, we rewrite the following expression

\[ \sum_{j=1}^{n} \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) u_t = \frac{\rho_n^{2n+1}}{k_n \ell(\eta_n)} \sum_{t=1}^{n} \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) u_t(1) + \frac{\rho_n^{2n+1}}{k_n \ell(\eta_n)} \sum_{t=1}^{n} \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) u_t(2) \]

\[ + \frac{\rho_n^{2n+1}}{k_n \ell(\eta_n)} \sum_{t=1}^{n} \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) u_t(1) + \frac{\rho_n^{2n+1}}{k_n \ell(\eta_n)} \sum_{t=1}^{n} \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) u_t(2) \]

\[ =: S_{n1} + S_{n2} + S_{n3} + S_{n4}. \]

Therefore by Lemma A.1 we can obtain the following result

\[ \mathbb{E} [S_{n1}]^2 = \mathbb{E} \left[ \frac{\rho_n^{2n+1}}{k_n \ell(\eta_n)} \sum_{t=1}^{n} \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) u_t(1) \right]^2 \]

\[ \leq \frac{A \rho_n^{4n+2}}{k_n^2} \sum_{t=1}^{n} \sum_{j=1}^{t-1} \rho_n^{2(t-1-j)} \]

\[ = \frac{A \rho_n^{4n+2}}{k_n^2 (\rho_n^2 - 1)} \left[ \sum_{j=1}^{n} \rho_n^{2(t-1)} - n \right] = O \left( \rho_n^{-2n} \right) = o(1), \]

which shows that \( S_{n1} = o_p(1) \). Moreover, it holds that

\[ S_{n2} = \frac{\rho_n^{2n+1}}{k_n \ell(\eta_n)} \sum_{t=1}^{n} \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) u_t 1\{ |u_t| > \eta_n \} + o \left( \ell(\eta_n) \right) \frac{\rho_n^{2n+1}}{k_n \ell(\eta_n)} \sum_{t=1}^{n} \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j(1) =: V + VI. \]

Therefore, using the Cauchy-Schwarz inequality, we obtain the following result

\[ |V| \leq \frac{\rho_n^{2n+1}}{k_n \ell(\eta_n)} \left[ \sum_{t=1}^{n} u_t^2 \right] \left[ \sum_{t=1}^{n} \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) \right]^2 1\{ |u_t| > \eta_t \} \]

\[ = \sqrt{n} \frac{\rho_n^{2n+1}}{k_n \ell(\eta_n)} \sqrt{n} \frac{\rho_n^{2n+1}}{k_n \ell(\eta_n)} \sum_{t=1}^{n} \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) \right] \leq \frac{\sqrt{n} \rho_n^{2n+1}}{k_n \ell(\eta_n)} \left[ \sum_{t=1}^{n} \sum_{j=1}^{t-1} \rho_n^{2(t-1-j)} \ell(\eta_n) \left[ 1 + o(1) \right] \right] \mathbb{P}( |u_t| > \eta_t ) \]

\[ \leq \frac{\sqrt{n} \rho_n^{2n+1}}{k_n \ell(\eta_n)} \sqrt{k_n^2 \ell(\eta_n) o \left( \frac{1}{n} \right) = o(1),} \]

which together with \( \sum_{t=1}^{n} u_t^2 / n \ell(\eta_n) \overset{p}{\to} 1 \) implies that \( V = o_p(1) \).
Therefore, Lemma A.1 yields the following result

$$E \mid T_{n1} \mid \leq \left| \frac{\rho_n^{-n-1}}{k_n} \sum_{t=1}^{n} \frac{\sqrt{k_n}}{\ell(\eta_n)} \left| \frac{u_{t}^{(1)}}{\psi_{\tau}(u_{t\tau})} \right| \right| = O \left( \frac{\rho_n^{-n-1} n}{k_n} \right) = o(1),$$

$$E \mid T_{n2} \mid \leq \left| \frac{\rho_n^{-n-1}}{k_n \sqrt{\ell(\eta_n)}} \sum_{t=1}^{n} \frac{u_{t}^{(2)}}{\psi_{\tau}(u_{t\tau})} \right| \leq \frac{\rho_n^{-n-1}}{k_n \sqrt{\ell(\eta_n)}} n A. o \left( \frac{\sqrt{\ell(\eta_n)}}{\eta_n} \right) = o(1).$$

Furthermore, we have that

$$E \left( T_{n3}^2 \right) = \frac{\rho_n^{-2n}}{k_n \ell(\eta_n)} \tau(1 - \tau) \sum_{t=1}^{n} \sum_{j=t+1}^{n} \rho_n^{2(t-j-1)} \ell(\eta_n) \left[ 1 + o(1) \right]$$

$$\leq A. \frac{\tau(1 - \tau) \rho_n^{-2n-2}}{k_n^2 \rho_n^2 - 1} \frac{1}{n} = O \left( \frac{\rho_n^{-2n} n}{k_n} \right) = o(1).$$

Therefore, the quantities $T_{n1}, T_{n2}, T_{n3}$ are all $o_p(1)$. Next, we also show that $T_{n4} = o_p(1)$. In particular, it holds that

$$T_{n4} = \frac{\rho_n^{-n}}{k_n \sqrt{\ell(\eta_n)}} \sum_{t=1}^{n} \left( \sum_{j=t+1}^{n} \rho_n^{t-j} u_j 1 \{ |u_j| > \eta_n \} \right) \psi_{\tau}(u_{t\tau}) + o \left( \frac{\sqrt{\ell(\eta_n)}}{\eta_n} \right) \frac{\rho_n^{-n}}{k_n \sqrt{\ell(\eta_n)}} \sum_{t=1}^{n} \sum_{j=t+1}^{n} \rho_n^{t-j} \psi_{\tau}(u_{t\tau}).$$

$$= \frac{\rho_n^{-2n}}{k_n^2 \ell(\eta_n)} \sum_{t=1}^{n} x_{t-1}^2 = \frac{1}{k_n^2 \ell(\eta_n)} \left( \frac{\rho_n^{-2n}}{\ell(\eta_n)} \right) \frac{1}{k_n (\rho_n^2 - 1)} \left[ \rho_n^{-2n} (x_n^2 - x_0^2) - 2 \rho_n^{-2n+1} \sum_{t=1}^{n} x_{t-1} u_t - \rho_n^{-2n} \sum_{t=1}^{n} u_t^2 \right]$$

$$= \frac{1}{k_n (\rho_n^2 - 1)} \left[ \rho_n^{-2n} \sum_{t=1}^{n} x_{t-1}^2 - 2 \rho_n^{-2n+1} \sum_{t=1}^{n} x_{t-1} u_t - \rho_n^{-2n} \sum_{t=1}^{n} u_t^2 \right] + o_p(\rho_n^{-2n}).$$

Since $k_n (\rho_n^2 - 1) \to 2c$, $\rho_n^{-n} = o \left( \frac{k_n}{n} \right)$,

(A.25)

$$\frac{\rho_n^{-2n} \sum_{t=1}^{n} u_t^2}{k_n \ell(\eta_n)} = O_p \left( \frac{n}{k_n} \rho_n^{-2n} \right) = o_p(1).$$

Note that we also have that $\rho_n^{-2n+1} \sum_{t=1}^{n} x_{t-1} u_t = o_p(1)$, then we obtain that

$$\frac{\rho_n^{-2n}}{k_n^2 \ell(\eta_n)} \sum_{t=1}^{n} x_{t-1}^2 = \frac{1}{k_n (\rho_n^2 - 1)} \left( \frac{\rho_n^{-n}}{\sqrt{k_n \ell(\eta_n)}} \right)^2 + o_p(1)$$

$$= \frac{1}{k_n (\rho_n^2 - 1)} \left\{ \frac{x_0}{\sqrt{k_n \ell(\eta_n)}} + \frac{1}{\sqrt{k_n \ell(\eta_n)}} \sum_{t=1}^{n} \rho_n^{-t} u_t \right\}^2 + o_p(1) = \frac{1}{2c} L_n^2 + o_p(1).$$

31
Moreover, by Lemma we obtain that
\[
\frac{\rho_n^{-n}}{k_n \sqrt{\ell (\eta_n)}} \sum_{t=1}^{n} x_{t-1} \psi_\tau (u_{tr}) = \frac{x_0}{\sqrt{k_n \ell (\eta_n)}} \cdot \frac{1}{\sqrt{k_n \ell (\eta_n)}} \sum_{t=1}^{n} \rho_n^{-(n-t-1)} \psi_\tau (u_{tr}) + \frac{\rho_n^{-n}}{k_n \sqrt{\ell (\eta_n)}} \sum_{t=1}^{n} \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) \psi_\tau (u_{tr})
\]
\[
= \frac{x_0}{\sqrt{k_n \ell (\eta_n)}} K_n (\tau) + \frac{\rho_n^{-n}}{k_n \sqrt{\ell (\eta_n)}} \sum_{t=1}^{n} \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) \psi_\tau (u_{tr}) + o_p (1)
\]
\[
= \frac{\rho_n^{-n}}{k_n \sqrt{\ell (\eta_n)}} \sum_{t=1}^{n} \left( \sum_{j=1}^{t} \rho_n^{t-1-j} u_j \right) \psi_\tau (u_{tr}) - \frac{\rho_n^{-n}}{k_n \sqrt{\ell (\eta_n)}} \sum_{t=1}^{n} \left( \sum_{j=t}^{n} \rho_n^{t-1-j} u_j \right) \psi_\tau (u_{tr}) + o_p (1)
\]
\[
= \left( \frac{1}{\sqrt{k_n \ell (\eta_n)}} \sum_{t=1}^{n} \rho_n^{-(n-t+1)} \psi_\tau (u_{tr}) \right) \cdot \left( \frac{1}{\sqrt{k_n \ell (\eta_n)}} \sum_{j=1}^{n} \rho_n^{-j} u_j \right) + o_p (1),
\]
\[
= K_n (\tau) L_n + o_p (1).
\]

Furthermore, for \( \forall \ x > 0 \), we have that
\[
\mathbb{P} \left( \max_{1 \leq t \leq n} \left| \frac{\rho_n^{-n}}{k_n \sqrt{\ell (\eta_n)}} x_{t-1} \right| > x \right)
\]
\[
= \mathbb{P} \left( \max_{1 \leq t \leq n} \left| \frac{\rho_n^{-n}}{\sqrt{\ell (\eta_n)}} \left( \rho_n^{t-1} x_0 + \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) \right| > x k_n \right)
\]
\[
\leq \mathbb{P} \left( \max_{1 \leq t \leq n} \rho_n^{t-n-1} \left| \frac{x_0}{\sqrt{\ell (\eta_n)}} > \frac{k_n x}{3} \right\} + \mathbb{P} \left( \max_{1 \leq t \leq n-1} \left| \sum_{j=1}^{t-1} \rho_n^{-j} \frac{u_j^{(1)}}{\sqrt{\ell (\eta_n)}} \right| > \frac{k_n x}{3} \right)
\]
\[
+ \mathbb{P} \left( \max_{1 \leq t \leq n} \left| \sum_{j=1}^{t-1} \rho_n^{-j} \frac{u_j^{(2)}}{\sqrt{\ell (\eta_n)}} \right| > \frac{k_n x}{3} \right)
\]
\[
\leq \mathbb{P} \left( \max_{1 \leq t \leq n} \left| \frac{x_0}{\sqrt{\ell (\eta_n)}} > \frac{k_n x}{3} \right\} + \mathbb{P} \left( \max_{1 \leq t \leq n} \left| \sum_{j=1}^{t-1} \rho_n^{-j} \frac{u_j^{(1)}}{\sqrt{\ell (\eta_n)}} \right| > \frac{k_n x}{3} \right) + \mathbb{P} \left( \max_{1 \leq t \leq n} \left| \sum_{j=1}^{t-1} \rho_n^{-j} \frac{u_j^{(2)}}{\sqrt{\ell (\eta_n)}} \right| > \frac{k_n x}{3} \right).
\]

Notice that we have that
\[
\mathbb{P} \left( \max_{1 \leq t \leq n} \left| \sum_{j=1}^{t-1} \rho_n^{-j} u_j^{(2)} \right| > \frac{k_n \sqrt{\ell (\eta_n)} x}{3} \right)
\]
\[
\leq \mathbb{E} \left( \sum_{j=1}^{t-1} \rho_n^{-j} u_j^{(2)} \right) \leq o \left( \frac{\ell (\eta_n)}{\eta_n} \right) \cdot \frac{A}{(\rho_n - 1) k_n \sqrt{\ell (\eta_n)}} = o \left( \frac{\sqrt{\ell (\eta_n)}}{\eta_n} \right) \leq o \left( \frac{1}{\sqrt{n}} \right) = o (1).
\]
Therefore, we have that

\[
P \left( \max_{1 \leq t \leq n} \left| \frac{\rho_n^{-n} x_{t-1}}{k_n \ell(\eta_n)} \right| > x \right) \leq \text{all the above} \rightarrow 0.
\]

(A.26)

**Proof.** To prove Lemma 1, using the Cramer-Wold device, it suffices to show that

\[
aX_n + bY_n \overset{d}{\to} \mathcal{N} \left( 0, \frac{(a^2 + b^2)\sigma^2}{2c} \right), \text{ for any } a, b \in \mathbb{R}.
\]

(A.27)

We rewrite \( aX_n + bY_n = \sum_{i=1}^{n} \zeta_{ni} \), where

\[
\zeta_{ni} = \frac{1}{\sqrt{n}} \left[ a \rho_n^{-i} + b \rho_n^{-(n-i)-1} \right] u_i, \quad 1 \leq i \leq n,
\]

(A.28)

Denote with \( r_n, q_n, q_n \) be sequences of positive integers such that

\[
r_n (p_n + q_n) \leq n < (r_n + 1)(p_n + q_n)
\]

and \( r_n \sim n^{1-\nu/2}, p_n \sim n^{\nu/2} - n^{\nu/4} \) and \( q_n \sim n^{\nu/4} \). Notice that here notation \( a_n \sim b_n \) means \( \frac{a_n}{b_n} \to 1 \) as \( n \to \infty \). Therefore, we write

\[
\sum_{t=1}^{n} \zeta_{ni} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} V_j + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} W_j + \frac{1}{\sqrt{n}} R_n,
\]

(A.30)

Since \( \{V_1^*, ..., V_{r_n}^*\} \) are independent, we have that as \( n \to \infty \),

\[
\mathbb{E} \left( \sum_{j=1}^{r_n} \frac{V_j^*}{\sqrt{n}} \right)^2 \to \frac{(a^2 + b^2)\sigma^2}{2c}.
\]

(A.31)

Next, we have as \( n \to \infty \)

\[
\frac{p_n}{n^\gamma} \sum_{j=1}^{(p_n-1)} \sum_{\ell=-\ell-1}^{\ell} \sum_{k=1}^{\ell} \left( I_{n,j,\ell,k} + I_{n,\ell} \right) \to \frac{a^2 + b^2}{2c},
\]

(A.32)

which implies that the LHS is uniformly bounded by a positive real number \( K_1 \).

\[\square\]

**Remark A.1.** Notice that when the model parameter is such that \( |\rho_n| < 1 \), then the influence of the initial value of \( \epsilon_0 \) vanishes as \( t \) grows. Therefore, it is convenient to take \( \epsilon_0 = 0 \). However, the error committed by imposing \( \epsilon_0 \) becomes serious when \( |\rho_n| > 1 \).
Proof of Lemma 4.3:

The following joint weakly functional convergence result holds:

\[
\left( J_n(\tau), \frac{\rho_n y_t}{k_n^2 \ell(\eta_n)} \sum_{i=1}^n y_{t-i} \right), \frac{\rho_n}{\sqrt{k_n^2 \ell(\eta_n)}} \sum_{i=1}^n y_{t-i} \psi_{\tau}(u_{t\tau}) \right) \Rightarrow \left( J(\tau), \frac{1}{2c} L^2, K(\tau)L \right).
\]

and it also holds that

\[
\max_{1 \leq t \leq n} \left| \frac{\rho_n}{\sqrt{k_n^2 \ell(\eta_n)}} y_{t-1} \right| \overset{p}{\rightarrow}.
\]

Notice that the above vector-valued functional implies that the second coordinate is a functional of the population model parameter that corresponds to the data generating process being estimated without the presence of quantile-dependent parameters.

Proof.

\[
\frac{\rho_n^{-2n+1}}{k_n \ell(\eta_n)} \sum_{i=1}^n y_{t-1} u_t = \frac{\rho_n^{-n+1} y_0}{\sqrt{k_n \ell(\eta_n)}} \frac{1}{\sqrt{k_n^2 \ell(\eta_n)}} \sum_{i=1}^n \rho_n^{-(n-t)-1} u_t + \frac{\rho_n^{-2n+1}}{k_n \ell(\eta_n)} \sum_{i=1}^n \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) u_t
\]

\[
= \frac{\rho_n^{-2n+1}}{k_n \ell(\eta_n)} \sum_{i=1}^n \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) u_t + o_p(1).
\]

Therefore, it suffices to show that the following probability bound result holds:

\[
\frac{\rho_n y_t}{k_n \ell(\eta_n)} \sum_{i=1}^n \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) u_t = o_p(1).
\]

Here we have that

\[
\frac{\rho_n^{-n}}{k_n \sqrt{\ell(\eta_n)}} \sum_{i=1}^n y_{t-1} \psi_{\tau}(u_t) = \frac{y_0}{k_n \sqrt{\ell(\eta_n)}} \cdot \frac{1}{\sqrt{k_n}} \sum_{i=1}^n \rho_n^{-(n-t+1)} \psi_{\tau}(u_t) + \frac{\rho_n^{-n}}{k_n \sqrt{\ell(\eta_n)}} \sum_{i=1}^n \left( \sum_{j=1}^{t-1} \rho_n^{t-1-j} u_j \right) \psi_{\tau}(u_t)
\]

\[
= \left( \frac{1}{\sqrt{k_n}} \sum_{i=1}^n \rho_n^{-(n-t+1)} \psi_{\tau}(u_t) \right) \cdot \left( \frac{1}{\sqrt{k_n \ell(\eta_n)}} \sum_{i=1}^n \rho_n^{j} u_j \right) + o_p(1)
\]

\[
K_n(\tau)L_n + o_p(1).
\]

Therefore, it holds that

\[
\mathbb{P} \left( \max_{1 \leq t \leq n} \left| \frac{\rho_n}{k_n \sqrt{\ell(\eta_n)}} y_{t-1} \right| > x \right) \rightarrow 0.
\]
B Main Results

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

$$R_n(\mu, \rho) = \frac{1}{n} \sum_{t=1}^{n} \varphi_t(y_t - \mu - \rho c x_{t-1}).$$

(B.1)

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

B.1 Near-Stationary Case ($c < 0$)

Lemma B.1. Under Assumption 2.3, when $c < 0$ then it holds that

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

(B.2)

Proof. Notice that Lemma B.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 B.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 $\mathbb{E}[y_n^2]$. The particular result provides a probability bound for the tails of the autoregressive time series which inherits the properties of the stochastic difference equation. $\square$

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

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

Lemma B.2. Under Assumptions 2.3 - 2.5, it holds that

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

(B.3)

where

$$\xi := \left| \frac{\partial}{\partial \theta} \mathbb{E}[\psi(u_1(\tau) - \theta)] \right|_{\theta = 0} \text{ and } B = \begin{pmatrix} 1 & 0 \\ 0 & \sigma^2 / (-2c) \end{pmatrix}.$$
Proof. By rearranging expression (2.35), taking the conditional expectation and sum over \(1 \leq t \leq n\), we obtain the following expression

\[
\sum_{t=1}^{n} \mathbb{E}_{\mathcal{F}_{t-1}} \left[ \varphi_{nt}(\delta) \right] = \sum_{t=1}^{n} \mathbb{E}_{\mathcal{F}_{t-1}} \left[ \varphi_{t} (\varepsilon_{t} - \delta(\tau)^{\top} X_{t}) - \varphi_{t} (\varepsilon_{t}) \right].
\]

Using Knight (1998)'s identity we have that

\[
\varphi_{t}(u_{1} - u_{2}) - \varphi_{t}(u_{1}) = u_{2} \left( \tau - \mathbb{1}\{u_{1} \leq 0\} \right) + u_{2} \int_{0}^{1} \left[ \mathbb{1}\{u_{1} \leq u_{2}s\} - \mathbb{1}\{u_{1} \leq 0\} \right] ds.
\]

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

\[
\sum_{t=1}^{n} \mathbb{E}_{\mathcal{F}_{t-1}} \left[ \varphi_{nt}(\delta) \right] = \frac{1}{2} \xi \times \delta^{\top} \left( n \sum_{t=1}^{n} X_{t} X_{t}^{\top} \right) \delta + o_{p}(1).
\]

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

\[
\frac{1}{n k_{n}} \sum_{t=1}^{n} y_{t-1}^{2} \overset{p}{\rightarrow} \frac{\sigma^{2}}{-2c}.
\]

Therefore, it suffices to show that

\[
\frac{1}{n \sqrt{k_{n}}} \sum_{t=1}^{n} y_{t-1} = o_{p}(1).
\]

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

\[
\frac{1}{n \sqrt{k_{n}}} \sum_{t=1}^{n} y_{t-1} = \frac{1}{n \sqrt{k_{n}} (1 - \rho_{n})} \sum_{t=1}^{n} (1 - \rho_{n}) y_{t-1} = \frac{1}{n \sqrt{k_{n}} (1 - \rho_{n})} \sum_{t=1}^{n} y_{t-1} - \rho_{n} y_{t-1}
\]

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,

\[
\frac{1}{n \sqrt{k_{n}} (1 - \rho_{n})} \sum_{t=1}^{n} y_{t-1} = \frac{1}{n \sqrt{k_{n}} (1 - \rho_{n})} \sum_{t=1}^{n} \left[ y_{t-1} - (y_{t} - u_{t}) \right]
\]

\[
= \frac{1}{n \sqrt{k_{n}} (c/k_{n})} \sum_{t=1}^{n} \left( y_{0} - y_{n} + \sum_{t=1}^{n} u_{t} \right) = o_{p}(1).
\]

which shows that \(\frac{1}{n \sqrt{k_{n}}} \sum_{t=1}^{n} y_{t-1} \overset{p}{\rightarrow} 0\), converges in probability to zero.

Lemma B.3. Under Assumptions 2.2-2.5, it holds that

\[
\sum_{t=1}^{n} \varphi_{nt} \left( \delta(\tau) \right) \overset{p}{\rightarrow} \frac{1}{2} \xi \times \delta(\tau)^{\top} B \delta(\tau).
\]
Proof. By Lemma B.2, we can show that

\begin{equation}
\sum_{t=1}^{n} \left( \varphi_{nt}(\delta(\tau)) - \mathbb{E}_{t-1} \left[ \varphi_{nt}(\delta(\tau)) \right] \right) = o_p(1). \tag{B.12}
\end{equation}

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

\begin{equation}
\sum_{t=1}^{n} \left( \varphi_{nt}(\delta) \mathbb{1}\{\mathcal{B}_t(\lambda)\} - \mathbb{E}_{t-1} \left[ \varphi_{nt}(\delta) \mathbb{1}\{\mathcal{B}_t(\lambda)\} \right] \right) = o_p(1). \tag{B.13}
\end{equation}

In particular, the expression

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

Therefore, by definition of \( \varphi_{nt} \) we have that

\begin{equation}
\varphi_{nt}(\delta) = \delta^\top X_t \int_0^1 \left( \psi(\tilde{\varepsilon}_t) - \psi(\varepsilon_t - \varepsilon_t^\top \delta^\top X_t) \right) d\varepsilon. \tag{B.15}
\end{equation}

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

\begin{equation}
\sum_{t=1}^{n} \mathbb{E}_{t-1} \left[ \varphi_{nt}^2(\delta) \mathbb{1}\{\mathcal{B}_t(\lambda)\} \right] \leq \max_{1 \leq t \leq n} \mathbb{E}_{t-1} \left[ \varphi_{nt}^2(\delta) \mathbb{1}\{\mathcal{B}_t(\lambda)\} \right] \times \left( \sum_{t=1}^{n} \delta^\top X_t X_t^\top \delta \right) = o_p(1). \tag{B.16}
\end{equation}

\[ \square \]

Corollary B.1.

\begin{equation}
\sum_{t=1}^{n} \left( \rho_{\tau} \left( \varepsilon_t - X_t^\top \delta \right) - \rho_{\tau}(\varepsilon_t) \right) = -\delta^\top \sum_{t=1}^{n} X_t \psi(\varepsilon_t) + \frac{1}{2} \xi \delta^\top B \delta + R_n(\delta) \tag{B.17}
\end{equation}

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

Proof. Notice that the function \( \varrho(u) \) 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

\begin{equation}
\hat{\gamma} = \sum_{t=1}^{n} \frac{1}{\xi} \left( \int \psi(\varepsilon_t) X_t^\top + o_p(1). \tag{B.18}
\end{equation}

\[ \square \]
B.2 Near-Explosive Case \((c > 0)\)

Lemma B.4. Consider that \(y_1, \ldots, y_n\) are random variables generated from the autoregressive process. Then, when \(c > 0\) it holds that

\[
E\left[y_n^2\right] = o\left(p_n^2 k_n^2\right) \quad (B.19)
\]

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

\[
\max_{1 \leq t \leq n} \left\{ \frac{y_t^2}{\rho_n^2 k_n^2} \right\} = o_p(1) \quad (B.20)
\]

\[
\frac{1}{\sqrt{n\rho_n^2 k_n^2}} \sum_{t=1}^{n} y_{t-1} = o_p(1). \quad (B.21)
\]

Lemma B.5. We consider the following two joint convergence results

(i). Under Assumptions 2.2-2.5 it holds that

\[
\left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_t(\varepsilon_t), \frac{1}{\rho_n^2 k_n^2} \sum_{t=1}^{n} \rho_n^{t-1}(n-1) \psi_t(\varepsilon_t), \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \rho_n^{-t} \psi_t(\varepsilon_t) \right) \xrightarrow{d} \left( Z_1, Z_2, Z_3 \right). \quad (B.22)
\]

where \(Z_1, Z_2, Z_3\) is a Gaussian random vector with independent components and the finite variance terms given by \(E[\psi_t^2(\varepsilon_1)], \frac{1}{2c} E[\psi_t^2(\varepsilon_1)]\) and \(\sigma^2/(2c)\), respectively.

(ii). Under Assumptions 2.2-2.5 it holds that

\[
\left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_t(\varepsilon_t), \frac{1}{\rho_n^2 k_n^2} \sum_{t=1}^{n} y_{t-1} \psi_t(\varepsilon_t), \frac{1}{\rho_n^2 k_n^2} \sum_{t=1}^{n} y_{t-1}^2 \psi_t(\varepsilon_t) \right) \xrightarrow{d} \left( Z_1, \sum_{t=1}^{n} y_{t-1} \psi_t(\varepsilon_t) \right)\]

\[
\left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_t(\varepsilon_t), \frac{1}{\rho_n^2 k_n^2} \sum_{t=1}^{n} \rho_n^{t-1} \sum_{j=1}^{t-1} \rho_n^{t-j-1} \varepsilon_j \right) \xrightarrow{d} \left( Z_1, \sum_{t=1}^{n} y_{t-1} \psi_t(\varepsilon_t) \right). \quad (B.23)
\]

Proof. Recall that the difference equation with no model intercept, that is, \(y_t = \rho_n y_{t-1} + \varepsilon_t\) has a general solution of the form \(y_t = \rho_n^{t-1} y_0 + \sum_{j=1}^{t-1} \rho_n^{t-1-j} \varepsilon_j\). 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

\[
y_{t-1} = \rho_n^{t-1} y_0 + \sum_{j=1}^{t-1} \rho_n^{t-1-j} \varepsilon_j \quad (B.24)
\]

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

\[
\frac{1}{\rho_n^2 k_n^2} \sum_{t=1}^{n} y_{t-1} \psi_t(\varepsilon_t) = \frac{y_0}{\rho_n^2 k_n^2} \sum_{t=1}^{n} \rho_n^{t-1} \psi_t(\varepsilon_t) + \frac{1}{\rho_n^2 k_n^2} \sum_{t=1}^{n} \left( \sum_{j=1}^{t-1} \rho_n^{t-j-1} \varepsilon_j \right) \psi_t(\varepsilon_t) \quad (B.25)
\]
Then, since the first term of the above expression is asymptotically negligible by splitting the inner summation of the last term we obtain that

\[
\frac{1}{\rho_{n,c}^n k_n} \sum_{t=1}^{n} y_{t-1} \psi_t(\varepsilon_t) = \frac{1}{\rho_{n,c}^n k_n} \sum_{t=1}^{n} \left( \sum_{j=1}^{n} \rho_{n,c}^{-1-j} \xi_j - \sum_{j=t}^{n} \rho_{n,c}^{-1-j} \xi_j \right) \psi_t(\varepsilon_t) + o_p(1).
\]

Since, \( \sum_{j=t}^{n} \rho_{n,c}^{-1-j} \xi_j \overset{p}{\to} 0 \), then it follows that

\[
\frac{1}{\rho_{n,c}^n k_n} \sum_{t=1}^{n} y_{t-1} \psi_t(\varepsilon_t) = \frac{1}{\rho_{n,c}^n k_n} \sum_{t=1}^{n} \left( \sum_{j=1}^{n} \rho_{n,c}^{-1-j} \xi_j \right) \psi_t(\varepsilon_t) + o_p(1)
\]

\[
= \left( \frac{1}{\sqrt{k_n}} \sum_{t=1}^{n} \rho_{n,c}^{-t(n+1)} \psi_t(\varepsilon_t) \right) \left( \frac{1}{\sqrt{k_n}} \sum_{t=1}^{n} \rho_{n,c}^{-t} \xi_t \right) + o_p(1).
\]

Similarly, it holds that

\[
(B.26) \quad \frac{1}{\rho_{n,c}^n k_n^2} \sum_{t=1}^{n} y_{t-1}^2 = \frac{1}{2c} \left( \frac{1}{\sqrt{k_n}} \sum_{t=1}^{n} \rho_{n,c}^{-t} \xi_t \right) + o_p(1).
\]

Lemma B.6. The following joint convergence results hold

(i). Under Assumptions above we have that

\[
-\vartheta_n(\tau) \top \sum_{t=1}^{n} X_t \psi_t(\varepsilon_t) + \frac{\xi}{2} \vartheta_n(\tau) \top \left( \sum_{t=1}^{n} X_t X_t \top \right) \vartheta_n(\tau) \overset{d}{\to} -\vartheta_n(\tau) \left( Z_1, Z_2, Z_3 \right) \top + \frac{\xi}{2} \vartheta_n(\tau) \top B \vartheta_n(\tau)
\]

(ii). Under Assumptions above we have that

\[
\sum_{t=1}^{n} \left[ \vartheta_t \left( y_t - \vartheta_n(\tau) \top X_t \right) - \vartheta_t(\varepsilon_t) \right] = -\vartheta_n(\tau) \top \sum_{t=1}^{n} X_t \psi_t(\varepsilon_t) + \frac{\xi}{2} \vartheta_n(\tau) \top \left( \sum_{t=1}^{n} X_t X_t \top \right) \vartheta_n(\tau) + R_n \left( \vartheta_n(\tau) \right)
\]

with \( R_n \left( \vartheta_n(\tau) \right) = o_p(1) \) for fixed \( \vartheta_n(\tau) \) and further

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

\[
(B.27) \quad \phi \left( \vartheta_n(\tau) \right) = \frac{1}{2} \xi \times \vartheta_n(\tau) \top \left( \sum_{t=1}^{n} X_t X_t \top \right) \vartheta_n(\tau).
\]

Furthermore, we need to show that

\[
(B.28) \quad \sup_{\|\vartheta_n(\tau)\| \leq C} \left| \sum_{t=1}^{n} \phi \left( \vartheta_n(\tau) \right) - \phi \left( \vartheta(\tau) \right) \right| = o_P(1).
\]
Since \( \sum_{t=1}^{n} X_t X_t^\top \) converges in distribution, for any \( \lambda > 0 \) there exists \( M \) large enough, such that,

\[
P \left( \sup_{\|\vartheta_n(\tau)\| \leq C} \left| \varphi_{nt}(\vartheta_n(\tau)) - \varphi(\vartheta(\tau)) \right| I \left\{ \left\| \sum_{t=1}^{n} X_t X_t^\top \right\| > M \right\} > \lambda / 2 \right) < \lambda^*/2.
\]

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

\[
\sup_{\|\vartheta\| \leq C} \left| \varphi(\gamma + \vartheta) - \varphi(\gamma) \right| \leq \lambda.
\]

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

\[
P \left( \sup_{\|\vartheta_n(\tau)\| \leq C} \left| \varphi_{nt}(\vartheta_n(\tau)) - \varphi(\vartheta(\tau)) \right| I \left\{ \left\| \sum_{t=1}^{n} X_t X_t^\top \right\| \leq M \right\} > \lambda / 2 \right) < \lambda^*/2.
\]

Therefore, the combination of the above yields the uniformity result of interest. \( \Box \)

### B.3 Weak Convergence of Functionals 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 \( \sqrt{n} \left( \hat{\beta}(\tau) - \beta(\tau) \right) \) follows by considering the following objective function

\[
Z_n(\delta) = \frac{1}{n} \sum_{t=1}^{n} \varrho_{\tau} \left( \varepsilon_t - X_t^\top \delta / \sqrt{n} \right) - \varrho_{\tau} \left( \varepsilon_t \right)
\]

where \( \varepsilon_t(\tau) = y_t - X_t^\top \beta(\tau) \). The function \( Z_n(\delta) \) is convex, and is minimized at \( \hat{\delta}_n = \sqrt{n} \left( \hat{\beta}(\tau) - \beta(\tau) \right) \).

Therefore, we can show that the limit distribution of \( \hat{\delta}_n \) can be determined by the asymptotic distribution of the objective function \( Z_n(\delta) \). Furthermore, it follows from the Lindeberg-Feller central limit theorem that \( Z_n^{(1)} \xrightarrow{d} -\delta^\top W \), where \( W \xrightarrow{d} \mathcal{N}(0, \tau(1-\tau) D_0) \). Then, following the proof of Theorem 4.1 of Koenker (2005) it holds that (see also derivations in Knight (1998) and Kato (2009))

\[
\sum_{t=1}^{n} \mathbb{E} \left[ Z_{nt}^{(2)}(\delta) \right] \xrightarrow{d} \frac{1}{2} \delta^\top D_1 \delta.
\]

Therefore, it can be proved that

\[
Z_n(\delta) \xrightarrow{d} Z_0(\delta) \equiv -\delta^\top W + \frac{1}{2} \delta^\top D_1 \delta.
\]
The convexity of the limiting distribution of $Z_0(\delta)$ ensures that the uniqueness of the minimizer

\[(B.35)\quad \hat{\delta}_n := \arg \min_\delta Z_n(\delta) \mapsto \hat{\delta}_0 := \arg \min_\delta Z_0(\delta)\]

where $Z_n(\delta) = Z_n^{(1)}(\delta) + Z_n^{(2)}(\delta)$ and $Z_0(\delta) = -\delta^T \mathcal{W} + \frac{1}{2} \delta^T D_1 \delta$ such that $\hat{\delta}_n = \sqrt{n} \left( \hat{\beta}(\tau) - \beta(\tau) \right)$. 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 $\sqrt{n} \left( \hat{\beta}(\tau) - \beta(\tau) \right) \in \arg \min_{u \in \mathbb{R}^p} Z_n(u, \tau)$. Then, it follows that

\[
\left| Z_n(u, \tau) - G_n(u, \tau) \right| = \left| Z_n^{(2)}(u, \tau) - \frac{u^T D u}{2} f \left( F^{-1}(\tau) \right) \right|
\]

\[
\leq \left| E \left[ Z_n^{(2)}(u, \tau) - \frac{u^T D u}{2} f \left( F^{-1}(\tau) \right) \right] \right| + \left| Z_n^{(2)}(u, \tau) - E \left[ Z_n^{(2)}(u, \tau) \right] \right|
\]

where

\[(B.36)\quad Z_n^{(2)}(u, \tau) = \frac{1}{\sqrt{n} \lambda(n)} \sum_{i=1}^n \int_0^1 \left( \mathbf{1} \left\{ \varepsilon_i \leq F^{-1}_\varepsilon(\tau) + \frac{\lambda(n) \sigma_n^{-1} x_n^T u s}{\sqrt{n}} \right\} - \mathbf{1} \left\{ \varepsilon_i \leq F^{-1}_\varepsilon(\tau) \right\} \right) ds.
\]

Therefore, it can be shown that

\[(B.37)\quad \sup_{\tau \in [\alpha, 1-\alpha]} \left| E \left[ Z_n^{(2)}(u, \tau) - \frac{u^T D u}{2} f \left( F^{-1}(\tau) \right) \right] \right| \to 0, \quad \text{as } n \to \infty
\]

Thus, for large $n$, we further have that

\[(B.38)\quad \mathbb{P} \left( \sup_{\tau \in [\alpha, 1-\alpha]} \left| Z_n(u, \tau) - G_n(u, \tau) \right| \geq \lambda \right) \leq \mathbb{P} \left( \sup_{\tau \in [\alpha, 1-\alpha]} \left| Z_n^{(2)}(u, \tau) - E \left[ Z_n^{(2)}(u, \tau) \right] \right| \geq \frac{\lambda}{2} \right)
\]

The following step is to prove that

\[(B.39)\quad \limsup_{n \to \infty} \frac{1}{\lambda^2(n)} \log \left\{ \mathbb{P} \left( \sup_{\tau \in [\alpha, 1-\alpha]} \left| Z_n^{(2)}(u, \tau) - E \left[ Z_n^{(2)}(u, \tau) \right] \right| \geq \lambda \right) \right\} = -\infty
\]

From the definition of $Z_n^{(2)}(u, t)$, we obtain that

\[(B.40)\quad Z_n^{(2)}(u, \tau) - E \left[ Z_n^{(2)}(u, \tau) \right] = \int_0^1 \left( \mathcal{W}_n(us, \tau) - \mathcal{W}_n(0, \tau) \right) ds,
\]

where

\[
\mathcal{W}_n(r, \tau) = \frac{\sum_{t=1}^n X_n't u}{\sqrt{n} \lambda(n)} \left[ \mathbf{1} \left\{ \varepsilon_i \leq F^{-1}_\varepsilon(\tau) + \frac{\lambda(n) \sigma_n^{-1} X_n't r}{\sqrt{n}} \right\} - F^{-1}_\varepsilon(\tau) + \frac{\lambda(n) \sigma_n^{-1} X_n't r}{\sqrt{n}} \right] - F^{-1}_\varepsilon(\tau) \]
In practise we employ Theorem 1 from Knight (1998). Therefore, we have that

\begin{equation}
Z_n^{(1)}(u) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i u \left[ 1(\epsilon_i > 0) - 1(\epsilon_i < 0) \right]
\end{equation}

and

\begin{equation}
Z_n^{(2)}(u) = \frac{2a_n}{\sqrt{n}} \sum_{t=1}^{n} \int_{0}^{v_n} X_t u \left[ 1(\epsilon_i < s) - 1(\epsilon_i < 0) \right] ds
\end{equation}

where \( v_n = X_t^\top u/a_n \).

Then, by the Lindeberg-Feller central limit theorem, for each \( u \) it holds that,

\begin{equation}
Z_n^{(1)}(u) \xrightarrow{d} -uW
\end{equation}

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

\begin{equation}
Z_n^{(2)}(u) = \sum_{t=1}^{n} \mathbb{E} \left[ Z_{nt}^{(2)}(u) \right] + \sum_{t=1}^{n} \left[ Z_{nt}^{(2)}(u) - \mathbb{E} \left[ Z_{nt}^{(2)}(u) \right] \right].
\end{equation}

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.

Arvanitis, S. and Magdalinos, T. (2018). Mildly explosive autoregression under stationary conditional heteroskedasticity. *Journal of Time Series Analysis*, 39(6):892–908.

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.

Benke, J. M. and Pap, G. (2021). Nearly unstable family of stochastic processes given by stochastic differential equations with time delay. *Journal of Statistical Planning and Inference*, 211:1–11.

Billingsley, P. (1968). *Convergence of probability measures*. John Wiley & Sons.

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.

Cai, Z., Chen, H., and Liao, X. (2022). A new robust inference for predictive quantile regression. *Journal of Econometrics*.

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.

Duffy, J. A. and Kasparis, I. (2021). Estimation and inference in the presence of fractional d= 1/2 and weakly nonstationary processes. *The Annals of Statistics*, 49(2):1195–1217.

Fan, R. and Lee, J. H. (2019). Predictive quantile regressions under persistence and conditional heteroskedasticity. *Journal of Econometrics*, 213(1):261–280.

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.

Fu, K.-A., Ni, J., and Dong, Y. (2022). Cqr-based inference for the infinite-variance nearly nonstationary autoregressive models. *Lithuanian Mathematical Journal*, pages 1–9.

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.

Härdle, W. K., Wang, W., and Yu, L. (2016). Tenet: Tail-event driven network risk. *Journal of Econometrics*, 192(2):499–513.

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.

Hirukawa, J. and Lee, S. (2021). Asymptotic properties of mildly explosive processes with locally stationary disturbance. *Metrika*, 84(4):511–534.

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 cramér-type moderate deviations for the explosive autoregressive process. *Bernoulli*, 28(5):1–28.

Hwang, K.-S. and Pang, T.-X. (2009). Asymptotic inference for nearly nonstationary ar (1) processes with possibly infinite variance. *Statistics & probability letters*, 79(22):2374–2379.

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*.

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.

Katsouris, C. (2023). Structural break detection in quantile predictive regression models with persistent covariates. *arXiv preprint arXiv:2302.05193*.

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. 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.

Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *The Review of Financial Studies*, 28(5):1506–1553.

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

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
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.*

Liu, Q.-m., Liao, G.-l., and Zhang, R.-m. (2021). Quantile inference for nonstationary processes with infinite variance innovations. *Applied Mathematics-A Journal of Chinese Universities, 36*(3):443–461.

Liu, X., Li, X., Gao, M., and Yang, W. (2022). Mildly explosive autoregression with strong mixing errors. *Entropy, 24*(12):1730.

Liu, X., Long, W., Peng, L., and Yang, B. (2023). A unified inference for predictive quantile regression. *Journal of the American Statistical Association, pages* 1–15.

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

Lui, Y. L., Xiao, W., and Yu, J. (2018). Mild-explosive and local-to-mild-explosive autoregressions with serially correlated errors.

Lui, Y. L., Xiao, W., and Yu, J. (2021). Mildly explosive autoregression with anti-persistent errors. *Oxford Bulletin of Economics and Statistics, 83*(2):518–539.

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

Magdalinos, T. and Petrova, K. (2022). Uniform and distribution-free inference with general autoregressive processes.

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.

Maynard, A., Shimotsu, K., and Kuriyama, N. (2023). Inference in predictive quantile regressions. *arXiv preprint arXiv:2306.00296.*

Mikusheva, A. (2007). Uniform inference in autoregressive models. *Econometrica, 75*(5):1411–1452.

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.

Oh, H., Lee, S., and Chan, N. H. (2018). Mildly explosive autoregression with mixing innovations. *Journal of the Korean Statistical Society, 47*(1):41–53.

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. 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. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica*, 82(3):1177–1195.

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

Phillips, P. C. B. and Magdalinos, T. (2009). Econometric inference in the vicinity of unity. *Singapore Management University, CoFie Working Paper*, 7.

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.

Wang, X., Tang, X., and Song, Y. (2022). Asymptotics of m-estimators for moderate deviations from a unit root model with possibly infinite variance. *Communications in Statistics-Theory and Methods*, pages 1–18.

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.

Yu, X. and Kejriwal, M. (2021). Inference in mildly explosive autoregressions under unconditional heteroskedasticity. *Available at SSRN 4081663*.

Zhou, Z. and Lin, Z. (2015). Quantile inference for moderate deviations from a unit root model with infinite variance. *Journal of the Korean Statistical Society*, 44(2):280–294.