Predictive Quantile Regression with Mixed Roots and Increasing Dimensions: The ALQR Approach

Rui Fan† Ji Hyung Lee‡ Youngki Shin§

Nov 2022

Abstract

In this paper we propose the adaptive lasso for predictive quantile regression (ALQR). Reflecting empirical findings, we allow predictors to have various degrees of persistence and exhibit different signal strengths. The number of predictors is allowed to grow with the sample size. We study regularity conditions under which stationary, local unit root, and cointegrated predictors are present simultaneously. We next show the convergence rates, model selection consistency, and asymptotic distributions of ALQR. We apply the proposed method to the out-of-sample quantile prediction problem of stock returns and find that it outperforms the existing alternatives. We also provide numerical evidence from additional Monte Carlo experiments, supporting the theoretical results.

Keywords: adaptive lasso, cointegration, forecasting, oracle property, quantile regression

JEL classification: C22, C53, C61

1 Introduction

Predictive quantile regression (QR) identifies the impact of predictors on a set of conditional quantiles of a response variable. It provides richer information on the heterogeneous distributional prediction. For example, the conditional quantile prediction of stock returns receives much attention in finance since the tail quantile information has a crucial role in measuring risk. Many

---

*We immensely thank the co-editor, Torben Andersen, the associate editor and three anonymous referees for very constructive comments. We also greatly appreciate helpful comments by Roger Koenker, Xiaofeng Shao, Zhijie Xiao, Viktor Todorov, Kajal Lahiri, the seminar participants at Vanderbilt, Sungkyunkwan, Washington, Syracuse, Northwestern Kellogg, Shanghai University of Finance and Economics, Fudan and Xiamen University. Shin is grateful for financial support by the Social Sciences and Humanities Research Council of Canada (SSHRC-435-2018-0275). This research was enabled in part by support provided by Compute Ontario (https://www.computeontario.ca/) and Compute Canada (www.computecanada.ca).

† Rui Fan: fanr5@rpi.edu. Department of Economics, Rensselaer Polytechnic Institute, Russell Sage Laboratory 4307, 110 8th St., NY 12180, USA

‡ Ji Hyung Lee: jihyung@illinois.edu. Department of Economics, University of Illinois Urbana Champaign, 214 David Kinley Hall, 1407 West Gregory Dr, Urbana, IL 61801, USA

§ Youngki Shin: shiny11@mcmaster.ca. Department of Economics, McMaster University, 1280 Main St. W., Hamilton, ON, Canada, L8S 4L8

1

arXiv:2101.11568v4 [econ.EM]  4 Dec 2022
economic state variables are employed to predict stock returns and the number of candidate predictors is often large. When a large number of predictors are available, researchers encounter the inevitable model selection issue. A good model selection can improve forecasting performance but the opposite can also occur. Considering the importance of model selection in practice, we need constructive guidance for empirical applications.

In this paper we propose the adaptive lasso for predictive quantile regression (ALQR). Although there exists a large volume of literature on predictive mean regression of equity returns (see, e.g. Campbell (1987), Fama and French (1988), Hodrick (1992), Cenesizoglu and Timmermann (2012), Andersen et al. (2020) among others), predictive QR is relatively understudied. Cenesizoglu and Timmermann (2008) is an early paper on predictive QR, and Maynard et al. (2011), Lee (2016), Fan and Lee (2019), Gungor and Luger (2019), and Cai et al. (2022) recently develop inference methods in predictive QR with nonstationarity and heteroskedasticity. The proposed method is different from these approaches. We consider predictive QR with an increasing number of mixed root predictors and address two important problems raised in the stock returns data. First, the prediction power of each predictor can vary over different quantiles. By adapting the lasso, we allow the model selection based on real data not by the researcher’s discretion. Second, the predictors widely used in predicting equity returns are composed of stationary, local unit root, and cointegrated processes. For example, we plot the time-series of two predictors, dividend price ratio (dp) and default yield spread (dfr), in Figure 1. Even a simple eyeballing test easily confirms their different levels of persistence. (We conduct more informative estimation procedures in Section 3). We provide a unified adaptive lasso framework that allows those mixed root predictors. We show that the estimator converges to the true parameter value at different convergence rates and that the faster convergence rates make the adaptive lasso more efficient in the model selection.

Naturally, some technical challenges arise. Since the seminal paper of Tibshirani (1996), the lasso has been intensively studied in various fields of statistical analysis. However, most studies have been focusing on the i.i.d. sample and it has not been a long time since more studies have
been conducted with dependent data (see the references in the related literature section below). Furthermore, we allow nonstationary predictors, which impose an additional difficulty in the formal analysis of the proposed adaptive lasso. We tackle these issues by considering a simple model that only contains unit-root predictors first. Once establishing the desired properties of ALQR in this model, we generalize the model so that it includes all stationary, local unit root, and cointegrated predictors.

The contributions of this paper are two-fold. First, to the best of our knowledge, this is the first paper to study predictive QR with an increasing number of mixed root predictors. We propose the adaptive lasso for predictive quantile regression (ALQR) and derive the convergence rates, model selection consistency and asymptotic distributions of ALQR under some regularity conditions. As a by-product, we also prove that the standard QR estimator is consistent under both mixed roots (including the local unit roots) and the increasing dimension of the predictors, which is new in the literature. Second, we conduct an empirical analysis of the stock returns data and find that ALQR can improve prediction performance over existing alternatives across different quantiles. We apply the ALQR method along with the existing alternatives to the data set. The results confirm that ALQR shows better prediction performance across different quantiles, particularly at higher quantiles. As illustrated in Section 3, ALQR can be readily applicable to other applications.

The rest of the paper is organized as follows. This section finishes with a review of the relevant literature. In section 2, we introduce predictive QR models formally and define the ALQR estimator. In section 3, we investigate the performance of ALQR using the out-of-sample quantile prediction problem of stock returns. We study the theoretical properties of ALQR in sections 4 and 5. In section 7, we conduct some Monte Carlo simulation experiments. Section 7 concludes. All technical proofs are relegated to the appendix.

Related Literature The lasso has been extensively studied for cross-sectional data. Recently, there has been development in lasso procedures with dependent data. Basu and Michailidis (2015) exploit the spectral properties of the stationary time series design and investigate the regularity condition of the lasso that leads to the non-asymptotic bounds and the consistency results. Kock and Callot (2015) investigate the oracle property of the lasso in a stationary vector autoregression model. Adamek et al. (2020) provide an inference procedure based on the debiased/desparsified lasso under the near-epoch dependence assumption. Chernozhukov et al. (2021) propose a penalty selection algorithm with weakly dependent data and the post-selection inference procedure. Wu and Wu (2016) and Wong et al. (2020) analyze the lasso with non-sub-Gaussian processes that allow a heavy-tail distribution. Medeiros and Mendes (2016) show the asymptotic properties of the adaptive lasso for stationary high-dimensional time series models. For the cointegrated models, Kock (2016) shows the oracle property of the adaptive lasso in the autoregression model. Interestingly, he finds that the unit root test can be incorporated into the model selection procedure by adopting the Dickey-Fuller form of autoregression. Liao and Phillips (2015) propose a shrinkage estimator with multiple penalty terms to select the rank of cointegration and estimate the param-
eter simultaneously in a vector error correction model. Liang and Schienle (2019), Zhang et al. (2019), and Onatski and Wang (2018) investigate the same model in the high-dimensional setting.

Koo et al. (2020) recently use lasso to improve the prediction of stock returns. It is shown that lasso significantly reduces forecasting mean squared errors even with a mixture of stationary, unit-root, and cointegrated variables. However, the conventional lasso method may not have model selection consistency and the oracle property as shown by Meinshausen and Bühlmann (2004) and Fan and Li (2001). The adaptive lasso proposed by Zou (2006) improves the performance of the lasso. Instead of imposing the same penalty weight on all candidate parameters, the adaptive lasso penalizes each parameter proportionally to the inverse of its initial estimate. With a proper choice of the tuning parameter $\lambda_n$, the adaptive penalty weights for the irrelevant variables approach infinity, whereas those for the relevant variables converge to constants. Lee et al. (2021) apply the adaptive lasso to a predictive mean regression framework. Similar to Koo et al. (2020), predictors are allowed to have different degrees of persistence and cointegration. Lee et al. (2021) find that the adaptive lasso and a newly proposed twin adaptive lasso outperform the alternative methods in terms of predictor selection consistency and out-of-sample mean squared errors.

Some effort has been also made to investigate model selection and model estimation in QR under the i.i.d. samples. The $\ell_1$-penalized method in the QR framework has been studied for high-dimensional data analysis (see, e.g. Portnoy (1984), Portnoy (1985), Knight and Fu (2000), Koenker (2005), Li and Zhu (2008), Lee et al. (2018) and Belloni and Chernozhukov (2011)). To overcome the problem of inconsistent model selection (Fan et al. (2014); Wang et al. (2012)), recent studies have further considered the adaptive lasso in QR. Wu and Liu (2009) discuss how to conduct model selection for QR models using SCAD and the adaptive lasso method. Zheng et al. (2013) establish the oracle property for an adaptive lasso QR model with heterogeneous error sequences. Zheng et al. (2015) study a globally adaptive lasso method for ultra high-dimensional QR models.

**Notation**

Let $\| \cdot \|$ and $\| \cdot \|_0$ denote $\ell_2$-norm and $\ell_0$-norm, respectively. For a matrix $S$, $\|S\|$ represents the spectral norm. Let $F_a(\cdot)$ and $f_a(\cdot)$ denote a cumulative distribution function (CDF) and a probability density function (pdf) of a generic random variable $a$. Let $S > 0$ denote a generic positive definite matrix $S$. Let $\lambda_{\min}(S)$ and $\lambda_{\max}(S)$ denote the smallest and largest eigenvalues of $S$. We use $O_p(1)$ and $o_p(1)$ when a sequence is bounded in probability and converges to zero in probability, respectively. The $O(1)$ and $o(1)$ denote the non-stochastic counterparts.

## 2 Model and the ALQR Estimator

In this section, we introduce the predictive quantile regression model and the adaptive lasso for quantile regression (ALQR). We will develop the theory for the ALQR in two steps in Section 4. For a better exposition of the theory, we also propose the model and its estimator in two separate cases: (i) unit-root predictors; and (ii) mixed-root predictors.
2.1 QR Model with Unit-Root Predictors

Consider a predictive QR model with unit-root predictors:

$$Q_{y_t} (\tau|F_{t-1}) := \mu_{\tau} + x'_{t-1} \beta_{0\tau}$$
$$x_t := x_{t-1} + v_t,$$

where $Q_{y_t} (\tau|F_{t-1})$ is the conditional $\tau$-quantile of $y_t$, $F_t = \{z_j\}_{j=-\infty}^t$ with $z_j = (y_j, x'_j)$ is a natural filtration, $x_t$ is a $p_n$-dimensional vector of unit-root predictors with a stationary $O_p(1)$ initialization of $v_0 = \sum_{j=0}^{\infty} D_{vj} \epsilon_{t-j}$ following the innovation structure below, and $\beta_{0\tau}$ is the corresponding true parameter vector.

The innovation of unit root predictors, $v_t$, follows a linear process, which is commonly assumed in the predictive regression literature (see Phillips and Lee (2013, 2016), Cai et al. (2022), Lee et al. (2021) for a few recent papers):

$$v_t = \sum_{j=0}^{\infty} D_{vj} \epsilon_{t-j}, \epsilon_t \sim mds(0, \Sigma), \Sigma > 0,$$

$$E||\epsilon_t||^{2+\kappa} < \infty, \kappa > 0,$$

$$D_v0 = I_p, \sum_{j=0}^{\infty} ||D_{vj}|| < \infty, D_v(r) = \sum_{j=0}^{\infty} D_{vjr}, \text{ and } D_v(1) = \sum_{j=0}^{\infty} D_{vj} > 0,$$

$$\Sigma_{vv} = \sum_{h=-\infty}^{\infty} E(v_t v'_{t-h}) = D_v(1) \Sigma D_v(1)' ,$$

where $mds$ denotes martingale difference sequences with respect to the natural filtration.

We allow the dimension of predictors to increase, i.e. $p_n \to \infty$ as $n \to \infty$. To make notation simple, we omit subscript $n$ and use $p$ unless it may cause any confusion. We define $u_{\tau\tau} := y_t - Q_{y_t} (\tau|F_{t-1})$, the deviation of $y_t$ from the conditional $\tau$-quantile. If the CDF of $u_{\tau\tau}$ is continuous, we have

$$F_{u_{\tau\tau}}^{-1} (\tau|F_{t-1}) = F_{u_{\tau\tau}}^{-1} (\tau) = 0, \text{ for all } \tau$$

where $F_{u_{\tau\tau}}^{-1}$ is the inverse CDF of $u_{\tau\tau}$. It holds by construction that $\Pr (u_{\tau\tau} < 0|F_{t-1}) = \tau$, which is equivalent to $F_{u_{\tau\tau}}^{-1} (\tau|F_{t-1}) = 0$. To see the equivalence to the unconditional $\tau$-quantile, we note that

$$\Pr(u_{\tau\tau} < 0) = E[1(u_{\tau\tau} < 0)] = E[E[1(u_{\tau\tau} < 0)|F_{t-1}]] = E[\Pr(u_{\tau\tau} < 0|F_{t-1})] = \tau,$$

where $1(\cdot)$ denotes an indicator function. The equations above imply that both conditional and unconditional $\tau$-quantiles of $u_{\tau\tau}$ are zero for any given $\tau$. However, it does not mean that two distribution functions are the same, i.e. $F_{u_{\tau1\tau}}^{-1} (\tau_2|F_{t-1}) \neq F_{u_{\tau2\tau}}^{-1} (\tau_2)$ for $\tau_1 \neq \tau_2$ in general.
Let $\psi_\tau(u_{t\tau}) := \tau - 1(u_{t\tau} < 0)$. It is easy to find that $\psi_\tau(u_{t\tau})$ is uncorrelated to any predetermined regressor. That is, $\psi_\tau(u_{t\tau})$ is uncorrelated with any past innovations, $v_{t-j}$, for $j \geq 1$:

$$\text{Cov}(\psi_\tau(u_{t\tau}), v_{t-j}) = E[\psi_\tau(u_{t\tau}) \cdot v_{t-j}] - E[\psi_\tau(u_{t\tau})] \cdot E[v_{t-j}] = E[E[\psi_\tau(u_{t\tau}) | F_{t-1}] \cdot v_{t-j}] - E[E[\psi_\tau(u_{t\tau}) | F_{t-1}]] \cdot E[v_{t-j}] = 0,$$

by the law of iterated expectations and the fact that $E(\psi_\tau(u_{t\tau}) | F_{t-1}) = \tau - \text{Pr}(u_{t\tau} < 0 | F_{t-1}) = 0$.

In our settings, however, we allow the QR-induced regression errors $\psi_\tau(u_{t\tau})$ to be contemporaneously correlated with the innovations of unit root sequences. This is commonly assumed in cointegration and predictive regression literature, inducing a potential second order bias arising from the one-sided correlation, see, e.g., Xiao (2009).

We now define the ALQR that minimizes the penalized QR objective function as follows:

$$(\hat{\mu}_{ALQR}^\tau, \hat{\beta}_{ALQR}^\tau) := \arg\min_{\mu \in \mathbb{R}, \beta \in \mathbb{R}^p} n \sum_{t=1}^{n} \rho_\tau(y_t - \mu - x_{t-1}' \beta) + \sum_{j=1}^{p} \lambda_{n,j} |\beta_j|,$$  \hspace{1cm} (2)

where $\rho_\tau(u) := u(\tau - 1(u < 0))$. Following Zou (2006), we define the tuning parameter of the penalty term as

$$\lambda_{n,j} := \frac{\lambda_n}{\omega_j} = \frac{\lambda_n}{|\hat{\beta}_{\tau,j}|}\gamma,$$

where $\lambda_n$ is a sequence converging to infinity, $\omega_j := |\hat{\beta}_{\tau,j}|\gamma$ with a positive constant $\gamma$, and $\hat{\beta}_{\tau,j}$ is a first-step consistent estimator for $\beta_{0\tau,j}$. The penalty $\lambda_{n,j}$, which contrasts with the penalty in the standard lasso, is an adaptive weight and it has an inverse relationship with $\hat{\beta}_{\tau,j}$. The idea of using a consistent first-step estimator as an adaptive weight in $\lambda_{n,j}$ is to avoid over-penalizing important predictors (or under-penalizing irrelevant predictors). In this paper, we use the ordinary QR estimator (Koenker and Bassett (1978)) below as a first-step estimator:

$$(\hat{\mu}_{QR}^\tau, \hat{\beta}_{QR}^\tau) := \arg\min_{\mu \in \mathbb{R}, \beta \in \mathbb{R}^p} n \sum_{t=1}^{n} \rho_\tau(y_t - \mu - x_{t-1}' \beta).$$ \hspace{1cm} (3)

The consistency of the ordinary QR estimator is provided in Section 4.

2.2 QR Model with Mixed Roots

We now introduce a QR model with mixed-root predictors. We assume the predictors of the model have different degrees of persistence: $I(0)$, local unit roots, and cointegration. This model is the most relevant in practice. The model in the previous section can be seen as a special case of it.

Let $z_t$, $x_{c_t}$, and $x_t$ be vectors of stationary,\footnote{In this paper, we use covariance (weak) stationarity unless noted otherwise. The linear process assumption below implies covariance stationarity of $z_t$. In addition, the first element of $z_t$ is 1 so that $\beta_{0\tau,1}$ plays a role of the intercept.} cointegrated, and local unit-root predictors whose
lengths are \( p_z, p_c, \) and \( p_x \) respectively. Let \( p := p_z + p_c + p_x \). Given a sample of \( \{y_t, z_t, x_t'\}_t=1^n \), the QR model with mixed roots is defined as follows:

\[
Q_{y_t} (\tau | F_{t-1}) = z_{t-1}' \beta_{0r}^z + x_{t-1}' \beta_{0r}^x + x_{t-1}' \beta_{0r}^x.
\]

The cointegrated system in \( x_t^c \) has the triangular representation by Phillips (1991): for \( x_t^c \) and \( x_{2t}^c \) whose dimensions are \( p_1 \times 1 \) and \( p_2 \times 1 \), respectively,

\[
Ax_t^c = x_{1t}^c - A_1 x_{2t}^c = v_{1t}^c,
\]

\[
(I_{p_2} - R_2 L)x_{2t}^c = v_{2t}^c,
\]

where \( R_2 = I_{p_2} + c_2/n \) with \( c_2 = \text{diag}(\hat{c}_1, \ldots, \hat{c}_{p_2}) \), \( L \) is the lag operator, and the vector \( v_{1t}^c \) is the vector of \( I(0) \) cointegrating residuals. The cointegrated regressors \( x_t^c \) are decomposed into \( x_{1t}^c \) and \( x_{2t}^c \). The local-to-unity parameter of \( x_{2t}^c \) is assumed to be \( \hat{c}_j \in (-\infty, \infty) \) for \( j = 1, \ldots, p_2 \). Thus, the cointegrated system in \( x_t^c \) includes both stationary and nonstationary local unit root regions. Using (5), we can easily characterize the cointegration relations in \( x_t^c \) by the matrices \( A \equiv (I_{p_1}, -A_1) \) and \( A_1 \).

The near integrated process \( x_t = (x_{1t}, \ldots, x_{p_x}) \) is defined in a similar way:

\[
(I_{p_x} - R_x L)x_t = v_t,
\]

where \( R_x = I_{p_x} + c_x/n \) with \( c_x = \text{diag}(\hat{c}_1, \ldots, \hat{c}_{p_x}) \) and \( \hat{c}_j \in (-\infty, \infty) \). Again, the local-to-unity specification includes the unit root process as a special case when \( \hat{c}_j = 0 \).

Let \( X_t := (z_t^c, x_t^c, x_t') \) and \( \beta^* := (\beta^z, \beta^c, \beta^x)' \). Define a \( p_c \)-dimensional vector \( v_t^c := (v_{1t}^c, v_{2t}^c)' \) and a \( p \)-dimensional vector \( e_t := (z_t^c, v_t^c, v_t)' \). Abusing notation on \( e_t \), we assume that \( e_t \) follows a linear process:

\[
e_t = D_c(L)e_t = \sum_{j=0}^{\infty} D_{ej}e_{t-j},
\]

where \( D_c = I_{p_c} - \sum_{j=0}^{\infty} D_{ej} \) with \( D_{ej} = \text{diag}(\sum_{z} \Sigma_{zz}, \Sigma_{zc} \Sigma_{zv}) > 0, \sum_{z} \Sigma_{zz} \Sigma_{zv} > 0, \Sigma_{zc} \Sigma_{zv} > 0, \Sigma_{zz} > 0, \Sigma_{zc} > 0, \Sigma_{zv} > 0, \Sigma_{vv} > 0, \) and \( E ||e_t||^{2+\kappa} < \infty, \kappa > 0, \)

\[
D_{e0} = I_{p_c} - \sum_{j=0}^{\infty} D_{ej} || < \infty, D_c(z) = \sum_{j=0}^{\infty} D_{ej}z_j, \text{ and } D_c(1) = \sum_{j=0}^{\infty} D_{ej} > 0, \]

\[
\Omega_{ee} = \sum_{h=-\infty}^{\infty} E(e_t e_{t-h}') = D_c(1)\Sigma D_c(1)'.
\]

Note that this assumption implies a covariance stationary \( O_p(1) \) initialization of \( e_0 = (z_0^c, v_0^c, v_0') = \sum_{j=0}^{\infty} D_{ej}e_{-j} \).
Similar to Section 2.1, the QR-induced regression errors, \( \psi_\tau(u_\tau) \), are uncorrelated to any predetermined regressor but allowed to be contemporaneously correlated with the innovations of unit root sequences \( x^2_t \) and \( x_t \), the stationary predictor \( z_t \) as well as the cointegrating residuals \( v^2_{x_\tau} \).

We define the ALQR for Model (4) as follows:

\[
\hat{\beta}_{\tau}^{ALQR} = \arg\min_{\beta^* \in \mathbb{R}^p} \sum_{t=1}^n \rho_\tau(y_t - X'_{t-1}\hat{\beta}^*) + \sum_{j=1}^p \lambda_{n,j}|\hat{\beta}^*_j|, \tag{7}
\]

where \( \lambda_{n,j} = \lambda_n/\omega_j \), \( \omega_j = |\hat{\beta}_{\tau,j}|^\gamma \), \( \gamma \) is a constant, and \( \hat{\beta}_{\tau,j} \) is a first-step consistent estimator. We recommend using the ordinary QR estimator for \( \hat{\beta}_{\tau,j} \). The consistency of QR will be discussed in Section 4.2.

### 3 Quantile Prediction of Stock Returns

In this section, we consider the quantile prediction problem of stock returns and illustrate the usefulness of the proposed ALQR method. We use an updated version of the data set in Welch and Goyal (2008), which ranges from January 1952 to December 2019. It is composed of 816 monthly observations of the US financial and macroeconomic variables. Our goal is to predict the quantiles of the excess stock returns \( (y_t) \) using 12 predictors. The variable names are summarized in Table 1. See also Welch and Goyal (2008) for more details on these variables.

| Notation | Variable Name                  | High Persistence | AR(1) Coefficient |
|----------|--------------------------------|------------------|-------------------|
| \( y \)  | excess stock returns           | No               | 0.102             |
| dp       | dividend price ratio           | Yes              | 0.996             |
| dy       | dividend yield ratio           | Yes              | 0.996             |
| ep       | earnings price ratio           | Yes              | 0.990             |
| bm       | book-to-market ratio           | Yes              | 0.994             |
| dfy      | default yield spread           | Yes              | 0.969             |
| ntis     | net equity expansion           | Yes              | 0.982             |
| lty      | long term yield                | Yes              | 0.995             |
| tbl      | treasury bill rates            | Yes              | 0.991             |
| svar     | stock variance                 | No               | 0.475             |
| dfr      | default return spread          | No               | -0.078            |
| ltr      | long term rate of returns      | No               | 0.040             |
| infl     | inflation                      | No               | 0.545             |

We first check the mixed-root property of the predictors. In Figures 2–3, we plot each predictor using monthly observations. The predictors (dp, dy, ep, bm, dfy, ntis, lty, and tbl) collected in Figure 2 are highly persistent and have quite different patterns from the other four predictors (svar, dfr, ltr and infl) in Figure 3. The first-order autoregression coefficients reported in Table 1 also suggest that predictors have heterogeneous degrees of persistence. We also conduct the Johansen
test for cointegration. The results show that the cointegrating rank is 3 in all of the 804-month rolling windows.\(^2\) Thus, the data fit into the mixed-root model structure discussed in Section 2.2.

We next investigate the performance of ALQR in the quantile prediction problem of stock returns. For comparison, we also apply three alternative estimators in the literature: the QR estimator without selecting predictors (QR), the quantile lasso (LASSO), and the unconditional quantile without using any predictor (QUANT). For the tuning parameter of LASSO and ALQR, we use both the Bayesian information criteria (BIC) and the generalized information criteria (GIC), which are explained in detail in Section 6. We evaluate the performance of quantile prediction using the final prediction error (FPE) and the out-of-sample \(R^2\) following Lu and Su (2015). The FPE measures the out-of-sample quantile prediction errors:

\[
FPE(\tau) = \frac{1}{S} \sum_{s=1}^{S} \rho_{\tau}(y_s - \hat{y}_s),
\]

where \(\hat{y}_s\) is a prediction for \(\tau\)-quantile of \(y_s\) and \(S\) is the number of out-of-sample predictions. Note that FPE(\(\tau\)) averages the quantile loss function. It is different from the standard mean squared prediction error. At each quantile \(\tau\), a smaller FPE implies a better quantile prediction. The second measure of the performance is the out-of-sample \(R^2\):

\[
R^2(\tau) = 1 - \frac{\sum_{s=1}^{S} \rho_{\tau}(y_s - \hat{y}_s)}{\sum_{s=1}^{S} \rho_{\tau}(y_s - \bar{y}_s)},
\]

where \(\bar{y}_s\) is the unconditional \(\tau\)-quantile of \(y\) (QUANT) from the training sample. Note that the out-of-sample \(R^2\) measures the performance of each method relative to QUANT. For example, a positive \(R^2\) indicates that the prediction error is smaller than that of QUANT, and a larger \(R^2\) implies a better prediction. By definition, \(R^2\) of QUANT is always zero.

Tables 2–3 summarize the prediction results with BIC. They are based on the one-step-ahead prediction for the last 12 and 24 periods of the sample, respectively. The results with GIC are similar and left in the appendix. We consider 5 different quantiles, \(\tau = 0.05, 0.1, 0.5, 0.9,\) and 0.95. For each quantile, the best performance results, i.e., the smallest FPE and the highest \(R^2\), are marked in bold. In addition to performance measures, we also report the average number of selected predictors and the corresponding tuning parameter value \(\lambda\) for LASSO and ALQR.

Overall, ALQR shows quite satisfactory results. First, ALQR performs the best in prediction across all quantiles, except for two designs: \(\tau = 0.95\) at the 12-period experiment and \(\tau = 0.05\) at the 24-period experiment. In the former design, LASSO performs slightly better than ALQR but the difference is negligible. In the latter design, QR performs better than the alternatives. Second, the relative performance of ALQR to QUANT improves substantially for higher quantiles. Looking at \(R^2\), we find that ALQR predicts 44% (\(\tau = 0.9\)) and 32% (\(\tau = 0.95\)) better than QUANT at the 12-period design and 28% (\(\tau = 0.9\)) and 31% (\(\tau = 0.95\)) at the 24-period design, respectively.

\(^2\)The Johansen cointegration test results are presented in the appendix, Table D.12.
Figure 2: Time-series Plots of Persistent Predictors

Notes: All plots are based on 816 monthly observations ranging from January 1952 to December 2019. The full predictor names are defined in Table 1.
Figure 3: Time-series Plots of Stationary Predictors

Notes: All plots are based on 816 monthly observations ranging from January 1952 to December 2019. The full predictor names are defined in Table 1.
Table 2: Prediction Results of Stock Returns: 12 one-period-ahead forecasts, BIC

| Quantile (τ) | 0.05 | 0.1  | 0.5  | 0.9  | 0.95 |
|--------------|------|------|------|------|------|
| Final Prediction Error (FPE) |      |      |      |      |      |
| QR           | 0.0047 | 0.0099 | 0.0134 | 0.0078 | 0.0060 |
| LASSO        | **0.0045** | 0.0084 | 0.0123 | 0.0033 | **0.0020** |
| ALQR         | **0.0045** | **0.0083** | **0.0122** | **0.0031** | **0.0020** |
| QUANT        | 0.0046 | **0.0083** | 0.0124 | 0.0055 | 0.0029 |

| Out-of-Sample $R^2$ |      |      |      |      |      |
| QR           | -0.0147 | -0.1930 | -0.0831 | -0.4048 | -1.0666 |
| LASSO        | 0.0219 | -0.0045 | 0.0094 | 0.4115 | **0.3233** |
| ALQR         | **0.0223** | -0.0012 | **0.0167** | **0.4421** | 0.3211 |

| Average # of Selected Predictors |      |      |      |      |      |
| LASSO        | 11.00 | 9.42 | 12.00 | 10.92 | 10.58 |
| ALQR         | 10.00 | 7.83 | 10.00 | 6.08  | 9.25  |

| Tuning Parameter (λ) by BIC |      |      |      |      |      |
| LASSO ($\times10^{-4}$) | 82.07 | 684.27 | 0.04 | 263.25 | 27.66 |
| ALQR ($\times10^{-7}$)  | 5.26  | 89.29  | 1.46 | 283.17 | 0.88  |

Notes: For each quantile, the best performance result is written in bold font.

QR does not show particularly better prediction performance than QUANT, which is in line with the well-known result in the mean stock return prediction literature (see, e.g., Welch and Goyal (2008)). Third, ALQR selects fewer predictors than LASSO on average. This result is consistent with the report in literature that the standard lasso overselects relevant predictors in practice (See, e.g., Wasserman and Roeder (2009)). In Section 4, we prove the oracle properties of ALQR including model selection consistency. In Section 6, we provide further numerical evidence that the number of selected variables by ALQR is closer to the true sparsity via Monte Carlo simulations.

Finally, we make some remarks from the empirical perspective. In Figure 4, we plot the coefficient estimates of ALQR at the 12-period experiment. To make the graph readable, we divide the predictors into two groups (persistent and stationary) and restrict the quantiles into $τ = 0.1, 0.5,$ and 0.9. First, we observe that predictors have heterogeneous effects across quantiles. This provides useful information on predicting the distribution of stock returns. For high stock returns ($τ = 0.9$), default yield spread (dfy), dividend price ratio (dp), dividend yield ratio (dy), stock variance (svar), and inflation (infl) show larger effects. However, long term yield (lty), net equity expansion (ntis), treasury bill rates (tbl), and stock variance (svar) show large coefficients for low stock returns ($τ = 0.1$). Second, the magnitude of the coefficients at the median is relatively smaller than that at both tails. This result coincides with the previous findings in the literature (see, e.g., Welch and Goyal (2008) and Fan and Lee (2019)) that it is more difficult to predict the center part of the stock return distribution. The positive but small $R^2$’s of ALQR at $τ = 0.5$ in Tables 2–3 also reflect
Table 3: Prediction Results of Stock Returns: 24 one-period-ahead forecasts, BIC

| Quantile (τ) | 0.05 | 0.1  | 0.5  | 0.9  | 0.95 |
|--------------|------|------|------|------|------|
| Final Prediction Error (FPE) |      |      |      |      |      |
| QR           | **0.0045** | 0.0102 | 0.0147 | 0.0067 | 0.0050 |
| LASSO        | 0.0058 | 0.0102 | **0.0140** | **0.0045** | **0.0025** |
| ALQR         | 0.0058 | **0.0097** | **0.0140** | **0.0045** | **0.0025** |
| QUANT        | 0.0055 | 0.0098 | 0.0148 | 0.0063 | 0.0036 |

Out-of-Sample $R^2$

|          |      |      |      |      |      |
|----------|------|------|------|------|------|
| QR       | **0.1751** | -0.0388 | 0.0100 | -0.0745 | -0.3938 |
| LASSO    | -0.0590 | -0.0349 | 0.0573 | 0.2767 | 0.2881 |
| ALQR     | -0.0526 | **0.0103** | **0.0577** | **0.2798** | **0.3088** |

Average # of Selected Predictors

|          |      |      |      |      |      |
|----------|------|------|------|------|------|
| LASSO    | 10.88 | 8.96 | 12.00 | 10.63 | 10.71 |
| ALQR     | 9.96  | 8.88 | 10.63 | 10.25 | 9.67  |

Tuning Parameter ($\lambda$) by BIC

|          |      |      |      |      |      |
|----------|------|------|------|------|------|
| LASSO ($\times 10^{-4}$) | 66.33 | 577.58 | 0.30 | **210.13** | 14.76 |
| ALQR ($\times 10^{-7}$)  | 3.74  | 24.70 | 0.65 | 1.55  | 0.59  |

Notes: For each quantile, the best performance result is written in bold font.

such an aspect. Third, the coefficient of stock variance (svar) swings from a large positive value at $\tau = 0.9$ to a large negative number at $\tau = 0.1$. Since the return distribution would spread out when the market is volatile, this result is consistent with the stylized fact in the financial market.

4 Oracle Properties of ALQR Estimators

This section investigates the asymptotic properties of ALQR. We show that the oracle properties (Fan and Li (2001)) of the adaptive lasso remain valid in the QR framework when predictors have both increasing dimensions and various degrees of persistence. This result provides theoretical support for applying the ALQR method to stock returns prediction in Section 3. In Section 4.1, we discuss the case that all predictors are $I(1)$. Section 4.2 then generalizes the results to the case of mixed root predictors.

4.1 Unit-Root Predictors

Consider the QR model in Section 2.1 with unit-root predictors only. The following assumptions provide regularity conditions.

**Assumption 4.1 (Assumption f)** (i) The distribution function of $u_{t\tau}$, $F(\cdot)$, has a continuous density $f(\cdot)$ with $f(a) > 0$ on $\{a : 0 < F(a) < 1\}$. (ii) The derivative of conditional distribution
Figure 4: The ALQR (with BIC) Estimated Coefficients

\( \tau = 0.1 \), nonstationary predictors

\( \tau = 0.1 \), stationary predictors

\( \tau = 0.5 \), nonstationary predictors

\( \tau = 0.5 \), stationary predictors

\( \tau = 0.9 \), nonstationary predictors

\( \tau = 0.9 \), stationary predictors
function $F_{l-1}(a) = \Pr[u_{tr} < a | F_{l-1}]$, which we denote $f_{l-1}(\cdot)$, is continuous and uniformly bounded above by a finite constant $c_f$. (iii) For any sequence $\zeta_n \to F^{-1}(\tau)$, $f_{l-1}(\zeta_n)$ is uniformly integrable, and $E[f_{l-1}^{1+\eta}(F^{-1}(\tau))] < \infty$ for some $\eta > 0$.

For the bound and rate conditions, we define $A_{(t,p)} := E[f_{l-1}(0)x_{t-1}x'_{t-1}]$ and $B_{(t,p)} := E[\psi_r(u_{tr})^2x_{t-1}x'_{t-1}]$.

**Assumption 4.2 (Assumption L1)** For each $t$ and $p$, there exist some constants $c_{A(t,p)}$ and $c_{B(t,p)}$ such that $0 < c_{A(t,p)} \leq \lambda_{\min}(A_{(t,p)})$ and $0 < c_{B(t,p)} \leq \lambda_{\min}(B_{(t,p)})$.

Note that, for each $t$ and $p$, $\lambda_{\max}(A_{(t,p)})/p$ and $\lambda_{\max}(B_{(t,p)})/p$ are bounded above since $|f_{l-1}(\cdot)|$ and $|\psi_r(\cdot)|$ are uniformly bounded and the variance of $x_{t-1}$ is finite by the design of $v_t$. Let $\bar{\tau}_A(t,p) < \infty$ and $\bar{\tau}_B(t,p) < \infty$ be these finite bounds. We next define the following uniform bounds over $t$ for each $n$: $\xi_{A(n,p)} := \min_{1 \leq t \leq n} \xi_{A(t,p)}$, $\xi_{A(n,p)} := \max_{1 \leq t \leq n} \xi_{A(t,p)}$, and $\bar{\tau}_{B(n,p)} := \max_{1 \leq t \leq n} \bar{\tau}_{B(t,p)}$.

**Assumption 4.3 (Assumption U1)** For some $\alpha > 0$, (i) $p = n^\zeta$ with $0 < \alpha \zeta < 1$; (ii) $p^{3/2} \xi_{A(n,p)}^n = o(\xi_{A(n,p)}^n)$; (iii) $\frac{p^{3/2}}{n^{2+\alpha\zeta^2}} = o(1)$.

**Assumption 4.4 (Assumption \lambda 1)** $\lambda_n$ satisfies that (i) $\frac{\lambda_n n^{1/2} \zeta}{n^{1+\alpha\zeta^2}} \xi_{A(n,p)} \to 0$; (ii) $\frac{\lambda_n n^{(1-\alpha)\gamma}}{n^{2+\alpha\zeta^2} \xi_{A(n,p)}} \to \infty$, with $\gamma > 0$.

**Assumption 4.5** Let $q_n = \|A_n\|_0$ be the size of the true active set $A_n$. There exist positive constants $c_q, c_\beta$ and $c_0$ such that (i) $q_n = O(n^{c_\beta})$ with $c_\beta < 1/3$; (ii) $2c_q < c_\beta < 1$ and (iii)

$$n^{(1-c_\beta)/2} \cdot \min_{1 \leq j \leq q_n} |\beta_{0j}| \geq c_0.$$

We make some remarks on these assumptions. Assumption $f$ is a standard restriction in the QR literature on the conditional density of regression residuals. Assumptions L1 and U1 modify the conventional conditions from stationary QR literature to allow for serial dependence in predictors and regression errors. In particular, we modify Assumption A.2 of Lu and Su (2015) to accommodate nonstationary $x_t$. Based on the model settings in Section 2.1, we find that

$$E[x_{t-1}x'_{t-1}] = t \cdot \Sigma_{vv} + o(t),$$

for $t = \lfloor rn \rfloor$ with $r \in (0, 1)$, as $n \to \infty$. Therefore, we impose a similar set of restricted eigenvalue conditions for $E[x_{t-1}x'_{t-1}] / t$. This is a natural extension of the existing conditions in the i.i.d. or stationary QR setup to a nonstationary time series model with an increasing dimension. We allow the upper bounds $\bar{\tau}_{A(t,p)}$, $\bar{\tau}_{B(t,p)}$ and the lower bounds $\xi_{A(t,p)}$, $\xi_{B(t,p)}$ to depend on the dimension of predictors, hence the upper bounds can diverge to infinity and the lower bounds can converge to zero when $p \to \infty$ as $n \to \infty$. However, Assumption U1 (ii) imposes further restrictions on the upper bounds: The rate of convergence in $p$ is slow enough so that $(n/p^\alpha)$-consistency of the
ALQR estimator can be achieved. Specifically, Assumption \( L1 \) prevents high correlation between predictor innovations. Note that \( \lambda_{\min}(E[x_{t-1}x_{t-1}]/t) = 0 \) if there exits perfect multicollinearity between predictor innovations. Also, Assumption \( U1 \) restricts the strength of the contemporaneous correlation in \( v_t \) by limiting nonzero off-diagonal terms in \( \Sigma_{vv} \). In Assumption \( U1 (i) \), the condition \( 0 < \alpha \zeta < 1 \) imposes \( a_0 = \frac{\alpha^2}{n} = \frac{\alpha}{n} \to 0 \), indicating the restriction on the growth rate of the number of parameters. In the appendix, we discuss the technicality of these restrictions on \( p \) and \( \lambda \) in detail. Assumption \( \lambda 1 \) collects a set of rate conditions on \( \lambda_n \) to show the asymptotic properties of the ALQR estimators. Assumption \( \lambda 1 (i) \) is comparable to the condition of \( \lambda_n/\sqrt{n} \to 0 \) in a stationary time series model with fixed dimensions. This condition allows \( (n/p^\alpha) \)-rate under the lasso variable selection. Assumption \( \lambda 1 \) (ii) presents a necessary condition for model selection consistency of ALQR, which depends on the consistency of the initial estimator for \( \beta_{0r} \) (as shown in the appendix, proof of Theorem 4.2). Note that this condition is developed given that the ordinary QR estimator defined in (3) is used as an initial estimator for \( \beta_{0r} \). It can be generalized further since we do not require a certain rate of consistency for the initial estimator. If another consistent estimator is used, condition (ii) may be adjusted to the corresponding rate of consistency. Assumption 4.5 is for the minimum signal strength of the coefficients in the true active set, which we adopted from Sherwood and Wang (2016).

Next we show that the ordinary QR estimator is consistent and can be a qualified initial estimator for ALQR.

**Lemma 4.1 (Consistency of QR Estimator with unit-root predictors)** Under Assumptions 4.1–4.3, the QR estimator defined in (3) satisfies

\[
\| \hat{\beta}_\tau^{QR} - \beta_{0r} \| = O_p \left( \frac{p^\alpha}{n} \right).
\]

It is well-known that the rate of convergence is \( \frac{1}{n^{1/2}} \) with fixed \( p \) and \( \frac{p^{1/2}}{n^{1/2}} \) with increasing \( p \) when predictors are \( I(0) \). This indicates that the information loss (or the degrees of freedom) to estimate the increasing number of parameters is \( p^{1/2} \). It is also known that when the predictors are \( I(1) \), the rate of convergence becomes \( \frac{1}{n} \) with fixed \( p \). The super-consistency is caused by the stronger signal of \( X'X \), where \( X \) is a \( n \times p \) matrix of predictors. To allow for increasing \( p \), Lemma 4.1 extends the existing result for nonstationary QR and finds that the rate becomes \( \frac{p^\alpha}{n} = \frac{p^{1/2}p^{\alpha - 1/2}}{n} \), where the additional rate loss \( p^{\alpha - 1/2} \) comes from the increasing singularity of \( E[x_t x'_t] \) as summarized in Assumptions \( L1 \) and \( U1 \). In contrast, we have \( 0 < [\lambda_{\min}(E[x_t x'_t])] < [\lambda_{\max}(E[x_t x'_t])] < \infty \) for \( I(0) \) predictors with \( p/n \to 0 \).

Taking the QR estimator of Lemma 4.1 as an initial estimator for \( \beta_{0r} \), we show the consistency

---

\(^{3}\)In the i.i.d. mean regression model with highly correlated regressors, it is well known that the lasso performs worse than the ridge estimator. To accommodate the lasso with highly correlated regressors, researchers have also developed variants of the lasso such as the elastic net (Zou and Hastie, 2005) and the group lasso (Yuan and Lin, 2006). As shown in Figure D.1 in the appendix, however, most predictor innovations in our empirical application are not highly correlated to each other and satisfy this assumption. Furthermore, we investigate the performance of the proposed ALQR estimator when predictor innovations are highly correlated in the simulation experiments.
of ALQR and model selection consistency under unit roots.

**Theorem 4.1 (Consistency of ALQR under Unit Roots)** Under Assumptions 4.1–4.4, the ALQR estimator defined in (2) satisfies

$$
\| \hat{\beta}_{ALQR}^{\tau} - \beta_0^{\tau} \| = O_p \left( \frac{p^\alpha}{n} \right).
$$

Theorem 4.1 confirms that both the ALQR estimator and the QR estimator of Lemma 4.1 are \((n/p^\alpha)\)-consistent. With a proper choice of the penalty parameter \(\lambda_n\), the rate of consistency for the ALQR estimate is not affected. As we show in the proof of Theorem 4.1, this estimation rate is possible under Assumption \(\lambda_1\).

**Theorem 4.2 (Model Selection Consistency under Unit Roots)** Let \(\hat{A}_n := \{ j : \hat{\beta}_{ALQR}^{\tau,j} \neq 0 \}\) and \(A_0 := \{ j : \beta_0^{\tau,j} \neq 0 \}\). Under Assumptions 4.1–4.5, it holds that

$$
\Pr(j \in \hat{A}_n) \rightarrow 0, \text{ for } j \notin A_0,
$$

as \(n \rightarrow \infty\).

Theorem 4.2 shows that the ALQR procedure is an oracle procedure. When the predictors with diverging dimensions are all unit root, the adaptive lasso is at least as good as other “oracle” procedures in terms of variable selection. This is a major difference from the standard lasso, which does not enjoy the oracle properties. As in Zou (2006), with a proper choice of \(\lambda_n\), the adaptive lasso procedure can simultaneously estimate relevant coefficients and remove unimportant ones with probability approaching unity. In this study, we further develop the result of Zou (2006) and generalize the adaptive lasso procedure to a nonstationary QR model. Note that the consistency in model selection can be achieved only under Assumption \(\lambda_1\).

### 4.2 Mixed-Root Predictors

In this section, we consider a general QR model discussed in Section 2.2:

$$
Q_{yt}(\tau|\mathcal{F}_{t-1}) = z_{t-1}' \beta_{0z}^{\tau} + x_{1t-1}' \beta_{01}^{c} + x_{2t-1}' \beta_{02}^{c} + x_{t-1}' \beta_{0}^{x} = (z_{t-1}', x_{1t-1}', x_{2t-1}') \beta_{0}^{(0)} + x_{t-1}' \beta_{0}^{(1)}
$$

(10)

where \(X_t = (z_t', x_{1t}', x_{2t}', x_t')\) and \(\beta_{0}^{x} = (\beta_{01z}^{c}, \beta_{011}^{c}, \beta_{022}^{c}, \beta_{02}^{x})'\) are \(p\)-dimensional vectors. This model allows predictors to have heterogeneous degrees of persistence as well as increasing dimensions. The mixed-root predictors in \(X_t\) include stationary \((z_t)\), local unit root \((x_t)\), and cointegrated \((x_{1t}', x_{2t}')\) processes.
To capture the cointegration relation between $x_{1t}^r$ and $x_{2t}^r$, we define a transformation matrix $H$ as follows:

$$H_{(p \times p)} := \begin{pmatrix} I_{p_x} & 0 & 0 \\ 0 & I_{p_1} & 0 \\ 0 & A_1' & I_{p_2} \\ 0 & 0 & 0 \\ I_{p_x} \end{pmatrix}.$$ 

Using $H$, the original model can be rewritten as:

$$Q_\gamma(\tau | F_{t-1}) = X'_{t-1} \beta_{0r} = (X'_{t-1}H^{-1}) H \beta_{0r} \equiv \tilde{x}'_{t-1} \tilde{\beta}_{0r},$$

(11)

where $\tilde{\beta}_{0r} := H \beta_{0r} = (\beta_{0r}', \beta_{0,1r}', (A_1' \beta_{0,1r} + \beta_{0,2r}'), \beta_{0,2r}')'$ and $\tilde{x}'_{t-1} := X'_{t-1}H^{-1} = (z'_t, v'_{c1}, x'_{c1}, x'_1)$.

The following assumptions provide additional regularity conditions for mixed roots. Abusing notation slightly, we will use the same notation for the sequences of bounds. Let $M_n$ be a diagonal matrix of the form

$$M_n := \begin{pmatrix} \sqrt{n}I_r & 0 \\ 0 & I_{p-r} \end{pmatrix}.$$

**Assumption 4.6 (Assumption L2)** There exist $\xi_{A(n,p)}$ and $\xi_{B(n,p)}$ such that:

$$0 < \xi_{A(n,p)} \leq \lambda_{\min}\left(M'_{n} \sum_{t=1}^{n} E \left[ f_{t-1}(0) \tilde{x}_{t-1} \tilde{x}'_{t-1} \right] M_n \right)$$

and

$$0 < \xi_{B(n,p)} \leq \lambda_{\min}\left(M'_{n} \sum_{t=1}^{n} E \left[ \psi_{\tau}(u_{\tau}) \tilde{x}_{t-1} \tilde{x}'_{t-1} \right] M_n \right)$$

as $n \to \infty$.

**Assumption 4.7 (Assumption U2)** There exist $\overline{\xi}_{A(n,p)}$ and $\overline{\xi}_{B(n,p)}$ such that:

$$\lambda_{\max}\left(M'_{n} \sum_{t=1}^{n} E \left[ f_{t-1}(0) \tilde{x}_{t-1} \tilde{x}'_{t-1} \right] M_n \right) \leq \overline{\xi}_{A(n,p)} < \infty$$
and
\[
\lambda_{\text{max}} \left( \frac{M_n' \sum_{t=1}^{n} E \left[ \psi_t (u_{t\tau})^2 \tilde{\epsilon}_{t-1} \tilde{\epsilon}_{t-1}' \right] M_n}{n^2} \right) < 2 \tau_B(n,p) < \infty
\]
as \( n \to \infty \).

**Assumption 4.8 (Assumption λ2)** When \( \lambda_n \to \infty \) as \( n \to \infty \), (i) \( \lambda_n \frac{n^{1/4}}{n^{2^3 \alpha c_{\lambda}(n,p)}} \to 0 \); (ii) 
\[
\lambda_n \frac{n^{1/2 - \alpha c_{\lambda}(n,p)}}{n^{2^3 \alpha c_{\lambda}(n,p)}} \to \infty.
\]

Assumption \( \lambda 2 \) is analogous to Assumption \( \lambda 1 \) in Section 4.1. They are essential regularity conditions used to show ALQR’s consistency in parameter estimation and model selection under mixed roots. Compared with Assumption \( \lambda 1 \), Assumption \( \lambda 2 \) are more restrictive on the choice of \( \lambda_n \), because \( \lambda_n \) of Assumption \( \lambda 2 \) accommodates both stationary and local unit root predictors.

We first show the consistency of QR for the transformed model in (11).

**Lemma 4.2 (Consistency of the transformed QR Estimator under Mixed Roots)** Under Assumptions 4.1, 4.3, 4.6 and 4.7, we have
\[
\| \hat{\beta}^{(0),QR}_c - \tilde{\beta}_0^{(0)} \| = O_p \left( \frac{p^\alpha}{\sqrt{n}} \right), \quad \text{and} \quad \| \hat{\beta}^{(1),QR}_c - \tilde{\beta}_0^{(1)} \| = O_p \left( \frac{p^\alpha}{n} \right).
\]

In a mixed root model, Lemma 4.2 shows that the QR estimators for the \( I(0) \) predictors and \( I(1) \) predictors have different convergence rates. As we discussed in the previous section, the faster convergence rate for the \( I(1) \) predictors comes from the stronger signal of \( E \left[ w_t^{(1)} w_t^{(1)' \tau} \right] \). For the later use of these rates, we define \( a_n^{(0)} := (p^\alpha / \sqrt{n}) \) and \( a_n^{(1)} := (p^\alpha / n) \).

Based on the definition of the transformed model, we know that Lemma 4.2 can further imply consistency of the QR estimator in the original model (4). Since all parameters are identical in model (4) and (11) except \( \tilde{\beta}_2^{c} = A_1' \hat{\beta}_1^{c} + \hat{\beta}_2^{c} \), the consistency of QR estimator in (4) can be verified if we can show that \( \beta_2^{c} \) is an \( a_n^{(0)} \)-consistent estimator for \( \beta_0^{c} \). To see the convergence rate of \( \beta_2^{c} \), we have
\[
\hat{\beta}_2^{c} = \left( \hat{\beta}_2^{c} \right)^{QR} - A_1' \hat{\beta}_1^{c} = \left( \hat{\beta}_2^{c} \right)^{QR} - \beta_0^{c,2} - A_1' \left( \hat{\beta}_1^{c} - \beta_0^{c,1} \right) + \left( \beta_0^{c,2} - A_1' \beta_0^{c,1} \right)
\]
\[
= O_p \left( a_n^{(1)} \right) + O_p \left( a_n^{(0)} \right) + \beta_0^{c,2}
\]
\[
= \beta_0^{c} + O_p \left( a_n^{(0)} \right).
\]

Thus, the \( I(0) \) rate dominates, and \( \| \hat{\beta}_2^{c} - \beta_0^{c,2} \| = O_p \left( a_n^{(0)} \right) \).

These results are summarized in the following corollary.
**Corollary 4.3 (Consistency of the QR Estimator under Mixed Roots)** Under Assumptions 4.1, 4.3, 4.6 and 4.7, we have

\[
\|
\hat{\beta}^{(0),QR^*}_r - \beta^{(0)}_0 \|
= O_p\left(a^{(0)}_n\right), \quad \text{and} \quad
\|
\hat{\beta}^{(1),QR^*}_r - \beta^{(1)}_0 \|
= O_p\left(a^{(1)}_n\right).
\]

Note that \(\beta^{(0)}_0\) is the \((p_z + p_1 + p_2)\)-dimensional QR regression coefficient, and \(\beta^{(1)}_0\) is the \(p_x\)-dimensional QR regression coefficient, as given in Equation (10). We now investigate the oracle properties of ALQR with mixed roots.

**Theorem 4.4 (Consistency of ALQR under Mixed Roots)** Under Assumptions 4.1, 4.3, 4.6-4.8, we have

\[
\|
\hat{\beta}^{(0),ALQR^*}_r - \beta^{(0)}_0 \|
= O_p\left(a^{(0)}_n\right), \quad \text{and} \quad
\|
\hat{\beta}^{(1),ALQR^*}_r - \beta^{(1)}_0 \|
= O_p\left(a^{(1)}_n\right).
\]

**Theorem 4.5 (Model Selection Consistency under Mixed Roots)** Under Assumptions 4.1, 4.3, 4.5-4.8, it holds that

\[
\Pr(j \in \hat{A}_n) \rightarrow 0, \quad \text{for} \ j \notin A_0,
\]

where \(\hat{A}_n = \{j : \hat{\beta}^{ALQR^*}_{r,j} \neq 0\}\) and \(A_0 = \{j : \beta^{(0),r}_{0,j} \neq 0\}\).

It is worth emphasizing that the ALQR procedure does not require practitioners to conduct any pretest to identify different types of predictors. With the proper choice of \(\lambda_n\) along with consistent initial estimates of the parameters, ALQR can simultaneously estimate relevant regression coefficients and provide the consistent variable selection for the mixed root QR model.

## 5 Asymptotic Distributions of ALQR Estimators

In this section, we derive the asymptotic distributions of the proposed ALQR estimators. In QR literature, the Convexity Lemma (Pollard (1991)) typically provides a convenient limit theory, bypassing the stochastic equicontinuity argument, see Section 4 of Koenker (2005). Unfortunately, for the increasing dimension, we cannot use the Convexity Lemma which is defined in \(\mathbb{R}^p\) with a fixed \(p < \infty\). Thus we modify the classical chaining argument of Ruppert and Carroll (1980) to our ALQR framework. Lu and Su (2015) recently modified the proof of Ruppert and Carroll (1980) to show the asymptotic distribution of Jackknife Model Averaging QR with iid regressors of increasing dimensions. We follow a similar proof strategy but substantially refine the results to prove the distributional limit theory with weakly dependent regressors of increasing dimensions, along with the adaptive lasso penalty functions. For the asymptotic distribution with non-cointegrated local unit root predictors, we assume that the number of predictors \((p_x)\) is fixed. Note that \(p = p_z + p_1 + p_2 + p_x\) is still allowed to increase, even with this additional restriction of \(p_x < \infty\).

We now state the additional assumptions we need to prove the asymptotic distributions of the ALQR estimators.
Assumption 5.1 The $p_z \times 1$ vector of stationary predictors $z_t = \sum_{j=0}^{\infty} D_{zj} \epsilon_{t-j}$ satisfies the following condition from Corollary 2 of Withers (1981)

$$\|D_{zj}\| = O(e^{-vj}) \text{ with } v > 0$$

along with the conditions (1), (2), (5) of Withers (1981), provided in the appendix for brevity, see (C.27)-(C.29) in Section C of the appendix. Then $z_t$ is strong mixing with the strong mixing number $\alpha(j) = O(e^{-vj\lambda})$ with another constant $\lambda > 0$.

Assumption 5.2 $\sup_{j \geq 1} E(z_{t-1,j}^8) \leq c_z$ for some $c_z \leq \infty$.

Assumption 5.3 As $n \to \infty$, $\xi_B^{-1/2} \xi_A^{1/2} p^2 n^{-1/2} \to 0$, and $n^{-3} p^{12} \xi_B^{-4} \to 0$.

Assumption 5.4 We require $p_x$ (the number of non-cointegrated local unit root predictors) to be finite: $p_x < \infty$.

Assumption 5.1 is a strong mixing condition with the specific mixing rates to use the Bernstein Inequality of Merlevède et al. (2009). Assumption 5.2 is the moment condition we adopt from Lu and Su (2015). The moment condition is stronger than the typical assumptions when the number of the regressors is fixed. Assumption 5.3 is another set of combined rate conditions required to prove the asymptotic theories of ALQR estimators. Finally, Assumption 5.4 restricts the number of non-cointegrated local unit root predictors, which is also assumed in Koo et al. (2020) who study the mean regressions with increasing dimensions. The distributional QR limit theory with the increasing number of (non-cointegrated) unit-root regressors is an open question, and we leave it for a future research.

We are now ready to provide the asymptotic distributions of ALQR estimators for the coefficients in the true active set, which we denote $\hat{\beta}^{(i),\text{ALQR}}(\tau)(1)$ with $i = 1$ and 2 for $I(0)$ and $I(1)$ predictors, respectively. Let $R$ be a generic constant $l \times p_i$ matrix, where $l$ is fixed and $p_i$ is defined conformably below. Notation $\Rightarrow$ indicates the convergence in distribution.

**Theorem 5.1 (Asymptotic Distributions of ALQR Estimators)** When Assumptions 4.1, 4.3, 4.5-4.8, 5.1-5.4 hold,

(i) For $I(0)$ active predictors (the first $p_z + p_1$ predictors; hence $p_i = p_z + p_1$), the first $q_n$ non-zero elements has the following limit theory:

$$\sqrt{n}R \left[ \hat{\beta}^{(0),\text{ALQR}}(\tau)(1) - \beta^{(0)}_{0\tau}(1) \right] \Rightarrow N \left( 0, RA_{(t,p)}^{-1} B(t,p) A_{(t,p)}^{-1} R' \right)$$

(ii) For the middle $p_2$ predictors (the parameters with respect to the active cointegrated predictors; hence $p_i = p_2$), the first $q_n$ non-zero elements has the following limit theory:

$$\sqrt{n}R \left[ \hat{\beta}^{(1),\text{ALQR}}(\tau)(1) - \beta^{(1)}_{0\tau}(1) \right] \Rightarrow N \left( 0, RA_1' A_{(t,p)}^{-1} B(t,p) A_{(t,p)}^{-1} A_1 R' \right)$$
(iii) For the last \( p_x \) (non-cointegrated) active local unit root predictors, with \( p_x < \infty \), the first \( q_n \) non-zero elements has the following limit theory:

\[
\frac{1}{n} \left[ \hat{\beta}_\tau^{(1),ALQR} - \beta_\tau^{(1)} \right] \overset{\text{d}}{\rightarrow} \left( M_{\beta_\tau}^x \right)^{-1} G_\tau,
\]

where

\[
G_\tau^x = \int J_x(r) dB_{\psi_\tau}, \quad \text{and} \quad M_{\beta_\tau}^x = f_u(0)^{-1} \int J_x(r) J_x(r)' dr
\]

and \( B_{\psi_\tau} := BM(\tau(1 - \tau)) \) is a Brownian motion and \( J_x(r) = \int_0^r e^{(r-s)} dB_x(s) \) is an Ornstein-Uhlenbeck (OU) process.

The result of Theorem 5.1-(i) is in line with Theorem 2 of Medeiros and Mendes (2016) who studied the asymptotic distributions of adaptive lasso estimators in mean regressions. It is also interesting to see that there is an additional nuisance parameter \( A_1 \) (the cointegrating vector) in the asymptotic distribution of ALQR with the cointegrated predictors.

6 Monte Carlo Simulations

In this section, we conduct a set of Monte Carlo simulation studies to evaluate the forecasting performance of the proposed ALQR method. As is motivated by the data patterns of the stock returns application in Section 3, we construct a simulation environment that has 12 predictors. As seen in Table 1, we allow for mix-root predictors and let 8 of them be persistent.

Based on the monthly observations of the stationary predictors (svar, dfr, infl and ltr), we calibrate the structure of \( z_t \) by estimating a VAR(\( p \)) model with a Bayesian information criterion (BIC). An estimated VAR(2) model is obtained as follows:

\[
z_t = \begin{pmatrix}
0.427 & -0.059 & 0 & -0.017 \\
-0.019 & 0.003 & 0.074 \\
0.022 & 0.538 & 0 \\
1 & 0 & 0 & 0
\end{pmatrix} z_{t-1} + \begin{pmatrix}
0.208 & 0 & 0 & 0 \\
0.421 & -0.089 & -0.312 & 0 \\
0.092 & 0.023 & 0.239 & 0 \\
0 & 0 & 1.038 & 0
\end{pmatrix} z_{t-2} + u_{zt},
\]

where \( u_{zt} \sim N(0, 1) \). To identify the cointegrating relations among persistent predictors, we apply the Johansen test. The estimated cointegrating rank is 3. Thus, we generate a set of cointegrating predictors from the following model:

\[
(x_{1t}^c, x_{2t}^c)' = \begin{pmatrix}
0.14 & 0.8495 & 0.0039 & -0.1545 \\
0.19 & 0.8084 & 0.0033 & -0.0507 \\
-1.24 & 1.2438 & 0.9788 & 0.3206 \\
0.02 & -0.0205 & 0.0009 & 0.9835
\end{pmatrix} (x_{1,t-1}^c, x_{2,t-1}^c)' + v_t^c,
\]

where \( v_t^c \sim N(0, 1) \). For unit-root predictors, we use i.i.d. \( N(0, 1) \) innovations with stationary
initializations.

We consider the following scenarios with different numbers of zero coefficients:

(1) 6 non-zero coefficients:
\[
\beta^z = (-0.4, 0, 0, 0.1), \quad (\beta^{c'}, \beta^{x'}) = (0.2, -0.4, 0.1, 0.5, 0, 0, 0, 0);
\]

(2) 8 non-zero coefficients:
\[
\beta^z = (-0.4, 0, 0, 0.1), \quad (\beta^{c'}, \beta^{x'}) = (0.2, -0.4, 0.1, 0.5, 0.2, 0.15, 0, 0);
\]

(3) 12 non-zero coefficients (no zero coefficient):
\[
\beta^z = (-0.4, 0.2, 0.15, 0.1), \quad (\beta^{c'}, \beta^{x'}) = (0.2, -0.4, 0.1, 0.5, 0.2, 0.15, 0.1, -0.05).
\]

The sample size is set to be \( n = 1000 \). The last 12 periods are used for out-of-sample prediction evaluation and the remaining periods are used for in-sample estimation. The quantiles of interest are 0.05, 0.1, 0.5, 0.9, 0.95. The prediction performance is measured by the final prediction error (FPE) and the out-of-sample \( R^2 \) defined in (8) and (9), respectively. The performance measures are computed by averaging over 1000 replications.

We briefly discuss the choice of tuning parameter \( \lambda \). For ALQR, we recommend using the Bayesian information criterion (BIC) proposed by Wang and Leng (2007) or the generalized information criterion (GIC) proposed by Fan and Tang (2013) and Zheng et al. (2015). The objective function for choosing \( \lambda \) is defined as
\[
\log \left( \frac{1}{n} \sum_{t=1}^{n} \rho_\tau(y_t - x_{t-1}'\hat{\beta}_\tau(\lambda)) \right) + \Gamma_n \cdot \hat{S}(\lambda),
\]
where \( \hat{\beta}_\tau(\lambda) \) is ALQR given \( \lambda \), \( \Gamma_n \) is a positive sequence converging to 0, and \( \hat{S}(\lambda) := \|\hat{A}_n\|_0 \) is the number of active predictors selected by \( \hat{\beta}_\tau(\lambda) \). We set \( \Gamma_n = \log(n)/n \) for BIC, and \( \Gamma_n = \log(p) \cdot \log(\log(n))/n \) for GIC and select \( \hat{\lambda} \) that minimizes the information criterion (objective) function. We use the same procedures for LASSO. Different approaches like the k-fold cross-validation and the Akaike information criterion (AIC) are available but it is known that they sometimes fail to effectively identify the true model (see, e.g. Shao (1997); Wang et al. (2007); Zhang et al. (2010)). Zheng et al. (2013) also discussed that the statistical properties of the k-fold CV have not been well understood for high-dimensional regression with heavy-tailed errors, where QR is often applied. On the contrary, Wang et al. (2007) show the model selection consistency of BIC for the fixed dimension and Wang et al. (2009) do for the increasing dimension with \( p < n \). Also, Fan and Tang (2013) and Zheng et al. (2015) show that GIC can identify the underlying true model consistently with probability approaching 1.

Overall, the performance of ALQR is satisfactory and confirms the theory developed in the previous section. Tables 4–6 summarize the simulation results from Scenarios 1–3. The upper and
Table 4: Simulation Results: Scenario 1 (6 non-zero coefficients)

| Quantile ($\tau$) | 0.05 | 0.1  | 0.5  | 0.9  | 0.95 |
|-------------------|------|------|------|------|------|
| Tuning parameter selection: BIC |
| Final Prediction Error (FPE) |
| QR                | 0.1821 | 0.2835 | 0.5807 | 0.2817 | 0.1795 |
| LASSO             | 0.1091 | 0.1813 | 0.4048 | 0.1807 | 0.1084 |
| ALQR              | 0.1068 | 0.1795 | 0.4016 | 0.1793 | 0.1064 |
| QUANT             | 2.7366 | 4.9691 | 12.1063 | 4.8815 | 2.6289 |
| Out-of-Sample $R^2$ |
| QR                | 0.9335 | 0.9430 | 0.9520 | 0.9423 | 0.9317 |
| LASSO             | 0.9601 | 0.9635 | 0.9666 | 0.9630 | 0.9588 |
| ALQR              | 0.9610 | 0.9639 | 0.9668 | 0.9633 | 0.9595 |
| Average # of Selected Predictors |
| LASSO             | 9.80  | 9.39  | 8.99  | 9.43  | 9.74  |
| ALQR              | 6.24  | 6.06  | 5.97  | 6.05  | 6.26  |
| Tuning Parameter ($\lambda$) |
| LASSO ($\times 10^0$) | 8.27  | 15.68 | 36.46 | 15.17 | 8.56  |
| ALQR ($\times 10^{-4}$) | 1.71  | 2.38  | 3.99  | 2.34  | 1.69  |

Tuning parameter selection: GIC

| Final Prediction Error (FPE) |
|-----------------------------|
| QR             | 0.1815 | 0.2834 | 0.5805 | 0.2817 | 0.1793 |
| LASSO          | 0.1084 | 0.1817 | 0.4045 | 0.1805 | 0.1084 |
| ALQR           | 0.1065 | 0.1800 | 0.4017 | 0.1791 | 0.1068 |
| QUANT          | 2.7285 | 4.9162 | 12.0961 | 4.8751 | 2.6290 |
| Out-of-Sample $R^2$ |
| QR             | 0.9335 | 0.9424 | 0.9520 | 0.9422 | 0.9318 |
| LASSO          | 0.9603 | 0.9630 | 0.9666 | 0.9630 | 0.9588 |
| ALQR           | 0.9610 | 0.9634 | 0.9668 | 0.9633 | 0.9594 |
| Average # of Selected Predictors |
| LASSO          | 10.14  | 9.79  | 9.35  | 9.80  | 10.13 |
| ALQR           | 6.80  | 6.36  | 6.11  | 6.35  | 6.76  |
| Tuning Parameter ($\lambda$) |
| LASSO ($\times 10^0$) | 6.21  | 11.66 | 28.39 | 11.54 | 6.32  |
| ALQR ($\times 10^{-4}$) | 1.22  | 1.93  | 3.55  | 1.94  | 1.24  |

24
Table 5: Simulation Results: Scenario 2 (8 non-zero coefficients)

| Quantile (τ) | 0.05 | 0.1 | 0.5 | 0.9 | 0.95 |
|--------------|------|-----|-----|-----|-----|
| Final Prediction Error (FPE) |      |     |     |     |     |
| QR           | 0.1845 | 0.2864 | 0.5882 | 0.2878 | 0.1848 |
| LASSO        | 0.1085 | 0.1818 | 0.4056 | 0.1800 | 0.1078 |
| ALQR         | **0.1074** | **0.1805** | **0.4033** | **0.1794** | **0.1069** |
| QUANT        | 2.6803 | 4.8142 | 12.1974 | 4.8994 | 2.5749 |
| Out-of-Sample $R^2$ |      |     |     |     |     |
| QR           | 0.9312 | 0.9405 | 0.9518 | 0.9413 | 0.9282 |
| LASSO        | 0.9595 | 0.9622 | 0.9667 | 0.9633 | 0.9581 |
| ALQR         | **0.9599** | **0.9625** | **0.9669** | **0.9634** | **0.9585** |
| Average # of Selected Predictors |      |     |     |     |     |
| LASSO        | 10.33 | 10.09 | 9.74 | 10.09 | 10.27 |
| ALQR         | 8.23  | 8.06  | 7.99 | 8.04  | 8.20  |
| Tuning Parameter ($λ$) |      |     |     |     |     |
| LASSO (×10^0) | 6.75 | 12.56 | 29.98 | 12.22 | 7.18 |
| ALQR (×10^{-4}) | 1.17 | 1.60 | 2.54 | 1.61 | 1.18 |

Tuning parameter selection: GIC

| Final Prediction Error (FPE) |      |     |     |     |     |
| QR           | 0.1852 | 0.2876 | 0.5880 | 0.2874 | 0.1848 |
| LASSO        | 0.1090 | 0.1825 | 0.4035 | 0.1799 | 0.1076 |
| ALQR         | **0.1077** | **0.1810** | **0.4021** | **0.1792** | **0.1071** |
| QUANT        | 2.7245 | 4.8459 | 12.2354 | 4.8893 | 2.5796 |
| Out-of-Sample $R^2$ |      |     |     |     |     |
| QR           | 0.9320 | 0.9407 | 0.9519 | 0.9412 | 0.9283 |
| LASSO        | 0.9600 | 0.9623 | 0.9670 | 0.9632 | 0.9583 |
| ALQR         | **0.9605** | **0.9627** | **0.9671** | **0.9633** | **0.9585** |
| Average # of Selected Predictors |      |     |     |     |     |
| LASSO        | 10.58 | 10.37 | 9.98 | 10.34 | 10.56 |
| ALQR         | 8.58  | 8.28  | 8.09 | 8.29  | 8.55  |
| Tuning Parameter ($λ$) |      |     |     |     |     |
| LASSO (×10^0) | 5.18 | 9.33 | 23.82 | 9.46 | 5.27 |
| ALQR (×10^{-4}) | 0.85 | 1.30 | 2.30 | 1.29 | 0.87 |
Table 6: Simulation Results: Scenario 3 (12 non-zero coefficients)

| Quantile (τ) | 0.05 | 0.1  | 0.5  | 0.9  | 0.95 |
|--------------|------|------|------|------|------|

| Final Prediction Error (FPE) |
|-----------------------------|
| QR                  | 0.1927 | 0.2968 | 0.5992 | 0.2964 | 0.1906 |
| LASSO               | 0.1093 | **0.1820** | 0.4046 | **0.1802** | **0.1091** |
| ALQR                | 0.1099 | 0.1825 | **0.4045** | 0.1807 | 0.1092 |
| QUANT               | 2.6958 | 4.9349 | 12.1341 | 4.8949 | 2.6794 |

| Out-of-Sample $R^2$ |
|---------------------|
| QR                  | 0.9285 | 0.9398 | 0.9506 | 0.9395 | 0.9289 |
| LASSO               | **0.9595** | **0.9631** | **0.9667** | **0.9632** | **0.9593** |
| ALQR                | 0.9592 | 0.9630 | **0.9667** | 0.9631 | 0.9592 |

| Average # of Selected Predictors |
|----------------------------------|
| LASSO                            | 11.97 | 11.99 | 12.00 | 11.99 | 11.98 |
| ALQR                             | 11.77 | 11.86 | 11.95 | 11.86 | 11.78 |

| Tuning Parameter (λ) |
|----------------------|
| LASSO (×10⁻¹)        | 2.01  | 1.59  | 0.646 | 1.12  | 1.97  |
| ALQR (×10⁻⁵)         | 1.75  | 1.88  | 1.87  | 1.89  | 1.75  |

Tuning parameter selection: GIC

| Final Prediction Error (FPE) |
|-----------------------------|
| QR                  | 0.1945 | 0.2972 | 0.6002 | 0.2964 | 0.1906 |
| LASSO               | **0.1100** | **0.1818** | 0.4048 | **0.1802** | **0.1091** |
| ALQR                | 0.1105 | 0.1820 | **0.4047** | 0.1803 | 0.1092 |
| QUANT               | 2.6718 | 4.9361 | 12.2134 | 4.8949 | 2.6794 |

| Out-of-Sample $R^2$ |
|---------------------|
| QR                  | 0.9272 | 0.9398 | 0.9509 | 0.9395 | 0.9289 |
| LASSO               | **0.9588** | **0.9632** | **0.9669** | **0.9632** | **0.9593** |
| ALQR                | 0.9586 | 0.9631 | **0.9669** | 0.9632 | 0.9592 |

| Average # of Selected Predictors |
|----------------------------------|
| LASSO                            | 11.98 | 11.99 | 12.00 | 12.00 | 11.98 |
| ALQR                             | 11.86 | 11.92 | 11.98 | 11.93 | 11.85 |

| Tuning Parameter (λ) |
|----------------------|
| LASSO (×10⁻¹)        | 1.42  | 1.24  | 0.64  | 0.90  | 1.45  |
| ALQR (×10⁻⁵)         | 1.34  | 1.39  | 1.27  | 1.36  | 1.40  |
lower panels of each table contain the results of using BIC and GIC, respectively. First, when there are zero coefficients in the model (Scenarios 1 and 2), ALQR performs better than its alternatives in terms of FPE (or $R^2$). As predicted by existing theory in the literature, both LASSO and ALQR show better performance than QR. In the meantime, ALQR performs uniformly, though slightly, better than LASSO. When all predictors have non-zero coefficients (Scenario 3), LASSO performs slightly better than ALQR over all quantiles except $\tau = 0.5$. Second, the simulation results confirm the model selection consistency of ALQR derived in Theorem 4.5. When we take a look at the average number of selected predictors in each table, the results of ALQR are quite close to the number of the true non-zero coefficients (6, 8, and 12 in each scenario). As known in the literature, LASSO mostly overselects them in Scenarios 1–2. Even in Scenario 3 where there are no zero predictors, ALQR selects most of them successfully. These results show the robustness of ALQR in terms of model selection. Finally, both BIC and GIC perform well under different designs and we do not see much difference between these two methods.

To investigate the robustness of the simulation results above, we further conduct simulation experiments with predictors with highly correlated innovations. In section 4, the oracle properties of the ALQR estimator are developed under the assumption that most predictor innovations are not highly correlated to each other. Although our empirical application satisfies this assumption, we provide additional simulation results in Tables D.3-D.11 of the appendix to demonstrate the robustness of ALQR in various settings of correlated predictor innovations. Specifically, we generate the $I(1)$ predictors in Scenario 1-3 with correlated innovations. We consider correlation $\rho = 0.1, 0.5,$ and $0.9$. We find that when the true DGP includes zero coefficients, ALQR is better than other alternative methods in most of the settings. Only when DGP has no zero coefficient ALQR is at risk of underperforming across quantiles, though the performance of the other methods is only slightly better than ALQR. An interesting finding is that ridge regression, commonly recommended when many predictors are highly correlated, has a mixed performance across quantiles. Based on the simulation results, we recommend using ALQR, especially in the case of sparse data. But ridge regression and other methods may be used if sparsity is not found in the data and the quantile of interest is around the center of the distribution.

In sum, the numerical experiments confirm that ALQR can provide satisfactory prediction performance in finite samples. The results are robust over different quantiles and simulation designs. Therefore, we can expect similar results in other quantile prediction applications when the predictors have mixed roots and are composed of $I(0)$, $I(1)$, and cointegrated processes.

### 7 Conclusion

In this paper, we show that the adaptive lasso for quantile regression (ALQR) is attractive in forecasting with stationary and nonstationary predictors as well as cointegrated predictors. The framework is general enough to include mixed roots but ALQR does not require any researchers’ knowledge on the specific structure of each predictor nor the order of integration. In this general
framework, we show that ALQR preserves the oracle properties. These advantages offer substantial convenience and robustness to empirical researchers working with quantile prediction using time series data.

We have focused on the case where the number of covariates, $p$, is allowed to grow as sample size increases, although $p$ is smaller than $n$. This framework justifies a wide range of practical applications in economics, such as the stock return quantile prediction in this paper. It would be an interesting future research to allow $p$ to be even larger than $n$, which has not been studied in a general time series framework with mixed roots.
Appendix

A Proofs for Section 4.1

For simplicity, we remove the intercept terms in Model (1) and define the dequantiled dependent variable (Lee (2016)) as

\[ y_{t\tau} := y_t - \hat{\mu}_{QR}^\tau. \]

Note that, in a mixed-root model, the intercept term is included as one of \( I(0) \) predictors (see Footnote 1), so there is no need to dequantile the dependent variable in Section 4.2.

**Proof of Lemma 4.1:** Let \( a_n = p^\alpha / n \) and \( c \in \mathbb{R}^p \) such that \( ||c|| = C \), where \( C \) is a finite constant. Denote the (unpenalized) quantile objective function as \( Q_n^{QR}(\beta_\tau) \).

To show the result of consistency, it suffices to show that for any \( \epsilon > 0 \), there exists a sufficiently large \( C \) such that

\[
P \left\{ \inf_{||c||=C} Q_n^{QR}(\beta_0\tau + a_n c) > Q_n^{QR}(\beta_0\tau) \right\} \geq 1 - \epsilon. \tag{A.1} \]

This inequality implies that with probability at least \( 1 - \epsilon \), there is a local minimizer \( \tilde{\beta}_\tau \) in the shrinking ball \( \{ \beta_0\tau + a_n c, ||c|| \leq C \} \) such that \( ||\tilde{\beta}_\tau - \beta_0\tau|| = O_p(a_n) \). Thus, the proof is completed if we show that the following term is positive:

\[
Q_n^{QR}(\beta_0\tau + a_n c) - Q_n^{QR}(\beta_0\tau) = \sum_{t=1}^n \rho_\tau(u_{t\tau} - x'_{t-1}a_n c) - \sum_{t=1}^n \rho_\tau(u_{t\tau}) \tag{A.2} \]

By Knight’s Identity,

\[
\sum_{t=1}^n \left[ \rho_\tau(u_{t\tau} - x'_{t-1}a_n c) - \rho_\tau(u_{t\tau}) \right] = -a_n \sum_{t=1}^n x'_{t-1}c \cdot \psi_\tau(u_{t\tau}) + \sum_{t=1}^n \int_0^{x'_{t-1}a_n c} \left( I(u_{t\tau} \leq s) - I(u_{t\tau} \leq 0) \right) ds \]

\[
= -a_n \sum_{t=1}^n x'_{t-1}c \cdot \psi_\tau(u_{t\tau}) + \sum_{t=1}^n \left[ \int_0^{x'_{t-1}a_n c} \left( I(u_{t\tau} \leq s) - I(u_{t\tau} \leq 0) \right) ds + \right. \]

\[
\left. \sum_{t=1}^n \left\{ \int_0^{x'_{t-1}a_n c} \left( I(u_{t\tau} \leq s) - I(u_{t\tau} \leq 0) \right) ds - \right. \]

\[
\left. E \left[ \int_0^{x'_{t-1}a_n c} \left( I(u_{t\tau} \leq s) - I(u_{t\tau} \leq 0) \right) ds \right] \right\} \equiv I_1 + I_2 + I_3.
\]

We will show that \( I_1 \) and \( I_3 \) are dominated by \( I_2 \) and that \( I_2 > 0 \).
First, we derive the upper bound of $I_1$. 

$$E|I_1|^2 = a_n^2 \sum_{t=1}^{n} c^t E \left[ \psi_r(u_{tt})^2 x_{t-1} x'_{t-1} \right] c + 2 \sum_{t=2}^{n} \sum_{k=1}^{t-1} c^t E \left[ \psi_r(u_{tt}) \psi_r(u_{ktt}) x_{t-1} x'_{k-1} \right] c$$

$$= a_n^2 \sum_{t=1}^{n} c^t E \left[ \psi_r(u_{tt})^2 x_{t-1} x'_{t-1} \right] c$$

$$\leq a_n^2 \sum_{t=1}^{n} t \tau_{B(t,p)} C^2$$

$$\leq C^2 a_n^2 \tau_{B(n,p)} \frac{n(n+1)}{2}.$$

The second equality holds since, for $k \leq t - 1$,

$$E \left[ \psi_r(u_{tt})^2 x_{t-1} x'_{t-1} \right] = E \left[ E_{t-1} \left[ \psi_r(u_{tt}) \right] \psi_r(u_{ktt}) x_{t-1} x'_{k-1} \right] = 0.$$

The third inequality holds by the definition of $\tau_{B(t,p)}$. Therefore, the Chebyshev’s inequality implies that

$$I_1 = O_p \left( a_n c_{B(n,p)}^{1/2} n \right) = O_p \left( a_n^2 c_{B(n,p)}^{1/2} n^2 p^{-\alpha} \right).$$

Next, we derive the lower bound of $I_2$. 

$$I_2 = \sum_{t=1}^{n} E \int_{0}^{a_n x'_{t-1} c} (F_{t-1}(s) - F_{t-1}(0)) \, ds$$

$$= \sum_{t=1}^{n} E \int_{0}^{a_n x'_{t-1} c} (f_{t-1}(0) \cdot s) \, ds \{ 1 + o_p(1) \}$$

$$= \frac{1}{2} a_n^2 \sum_{t=1}^{n} c^t E \left[ f_{t-1}(0) x_{t-1} x'_{t-1} \right] c \{ 1 + o(1) \}$$

$$\geq \frac{1}{4} C^2 a_n^2 \tau_{A(n,p)} \frac{n(n+1)}{2}$$

$$= O(a_n^2 n^2 \tau_{A(n,p)}).$$

The first equality holds by the law of iterated expectations and the second does by the Taylor expansion. The inequality holds under Assumption L1.
Finally, we derive the upper bound of $I_3$.

$$Var(I_3) = Var\left(\sum_{t=1}^{n} \int_{0}^{x'_{t-1}a_n c} (1(u_{t\tau} \leq s) - 1(u_{t\tau} \leq 0)) \, ds\right)$$

$$\leq E \left[\left(\sum_{t=1}^{n} \int_{0}^{x'_{t-1}a_n c} (1(u_{t\tau} \leq s) - 1(u_{t\tau} \leq 0)) \, ds\right)^2\right]$$

$$= E \left[\sum_{t=1}^{n} \left(\int_{0}^{x'_{t-1}a_n c} (1(u_{t\tau} \leq s) - 1(u_{t\tau} \leq 0)) \, ds\right)^2 \right]$$

$$+ 2 \sum_{t=2}^{n} \sum_{k=1}^{t-1} \left(\int_{0}^{x'_{t-1}a_n c} (1(u_{t\tau} \leq s) - 1(u_{t\tau} \leq 0)) \, ds\right) \left(\int_{0}^{x'_{k-1}a_n c} (1(u_{k\tau} \leq s) - 1(u_{k\tau} \leq 0)) \, ds\right)$$

$$\leq E \left[\sum_{t=1}^{n} \left(x'_{t-1}a_n c\right)^2 + 2 \sum_{t=2}^{n} \sum_{k=1}^{t-1} \left|x'_{t-1}a_n c\right| \left|x'_{k-1}a_n c\right| \right]$$

$$= a_n^2 \sum_{t=1}^{n} c' E \left[\left|x_{t-1}x'_{t-1}\right|\right] + 2a_n^2 \sum_{t=2}^{n} \sum_{k=1}^{t-1} E \left[\left|x'_{t-1}c\right| \left|x'_{k-1}c\right| \right]$$

$$= V_{3,1} + V_{3,2}$$

Using the similar arguments in $I_1$, we have

$$V_{3,1} \leq a_n^2 C^2 \overline{c}_A(n,p) \frac{n(n+1)}{2} = O(a_n^2 \overline{c}_A(n,p)n^2).$$

By the Cauchy-Schwarz inequality, the definition of $\overline{c}_A(n,p)$, and $t > k$, we have

$$V_{3,2} \leq 2a_n^2 \sum_{t=2}^{n} \sum_{k=1}^{t-1} \sqrt{E \left[\left|x'_{t-1}x_{t-1}\right|\right]} \sqrt{E \left[\left|x'_{t-1}x_{t-1}\right|\right]}$$

$$\leq 2a_n^2 \sum_{t=2}^{n} \sum_{k=1}^{t-1} \sqrt{C^2 \overline{c}_A(t,p)} \frac{c_f}{c_f} \sqrt{C^2 \overline{c}_A(t,p)} \frac{c_f}{c_f}$$

$$\leq 2a_n^2 C^2 \overline{c}_A(n,p) \frac{n}{c_f} \sum_{t=2}^{n} \sum_{k=1}^{t-1} t$$

$$= O(a_n^2 \overline{c}_A(n,p)n^3)$$

Therefore, $Var(I_3) = O(a_n^2 \overline{c}_A(n,p)^3)$, and Chebyshev’s inequality implies that

$$I_3 = O_p(\overline{c}_A^{1/2}(n,p) a_n n^{3/2} p^{-\alpha}) = O_p\left(\overline{c}_A^{1/2}(n,p) a_n^{2} n^{5/2} p^{-\alpha}\right).$$

By Assumptions $U_1$ and $U_2$, we establish the desired result. ■

**Proof of Theorem 4.1**: For simplicity, in this proof, we use $\hat{\beta}_\tau$ to represent the ALQR estimator $\beta_{ALQR}^{\hat{\beta}_\tau}$. Without loss of generality, let the values of $\beta_{0r,1}, \beta_{0r,2}, ..., \beta_{0r,q_n}$ be nonzero and $\beta_{0r,q_n+1}$,

31
$\beta_{0r,qn+2}, \ldots, \beta_{0r,p}$ be zero. Let $E_{t-1}(\cdot) \equiv E(\cdot | \mathcal{F}_{t-1})$.

To show the result of consistency, it suffices to show that for any $\epsilon > 0$, there exists a sufficiently large $C$ such that

$$P \left\{ \inf_{\|c\| = C} Q_n(\beta_{0r} + a_n c) > Q_n(\beta_{0r}) \right\} \geq 1 - \epsilon. \quad (A.3)$$

This inequality implies that with probability at least $1 - \epsilon$, there is a local minimizer $\hat{\beta}_r$ in the shrinking ball $\{\beta_{0r} + a_n c, \|c\| \leq C\}$ such that $\|\hat{\beta}_r - \beta_{0r}\| = O_p(a_n)$.

Since

$$Q_n(\beta_{0r} + a_n c) - Q_n(\beta_{0r}) \quad (A.4)$$

we need to show that $D_1 + D_2$ is positive.

For $D_2$, we know $|\beta_{0r,j} + a_n c_j| - |\beta_{0r,j}| \leq |a_n c_j|$. Given that $\hat{\beta}_r$ is a $(a_n^{-1})$-consistent estimate of $\beta_{0r}$, we have $a_n^{-1}(\hat{\beta}_{r,j} - \beta_{0r,j}) = O_p(1)$, and then $\hat{\beta}_{r,j} = O_p(a_n) + \beta_{0r,j} = o_p(1) + \beta_{0r,j}$. Thus,

$$|D_2| \leq \sum_{j=1}^{qn} \lambda_{n,j} |a_n c_j|$$

$$= \lambda_n a_n \sum_{j=1}^{qn} \left| \frac{1}{|\beta_{r,j}|^\gamma} c_j \right|$$

$$\leq \lambda_n a_n \left( \sum_{j=1}^{qn} \frac{1}{(|\beta_{r,j}|^2)^\gamma} \right)^{1/2} \cdot \|c\|$$
\[
\leq C \lambda_n a_n \left( \sum_{j=1}^{q_n} \frac{1}{(o_p(1) + \beta_{0r,j})^{2\gamma}} \right)^{1/2} \\
= C \lambda_n a_n \left( \frac{q_n}{o_p(1)} \sum_{j=1}^{q_n} 1 \right)^{1/2} \\
= O_p(\lambda_n a_n^{q_n/2}) \\
= O_p(\lambda_n a_n^{p/2}). \tag{A.5}
\]

We next consider \(D_1\). Following the proof of Lemma 4.1, we have that the dominating term of \(D_1\) is \(O_p(a_n^2 n^2 \mathcal{L}(n,p))\) and that it is positive with probability approaching 1. Under Assumption \(\lambda 1\ (i)\), \(D_1\) dominates \(D_2\). We complete the proof of Theorem 4.1. \(\blacksquare\)

**Proof of Theorem 4.2:** From Theorem 4.1 for a sufficiently large constant \(C\), \(\hat{\beta}_\tau^{ALQR}\) is a local minimizer lies in the ball \( \{ \beta_{0r} + a_n c, ||c|| \leq C \} \) with probability approaching 1 and \(a_n = p^{\alpha}/n\). For simplicity, in this proof, we use \(\hat{\beta}_\tau\) to represent the ALQR estimator \(\hat{\beta}_\tau^{ALQR}\).

First, note that the subgradient of the unpenalized objective function, \(s(\beta_\tau) = (s_1(\beta_\tau), ..., s_p(\beta_\tau))'\) is given by (Sherwood and Wang (2016), page 298 and Lemma 1):

\[
s_j(\beta_\tau) = -\sum_{t=1}^{n} x_{t-1,j} \psi_\tau(y_{t\tau} - x_{t-1,\hat{\beta}_\tau}) + \sum_{t=1}^{n} x_{t-1,j} k_t \text{ for } 1 \leq j \leq p,
\]

where \(k_t = 0\) if \(y_{t\tau} - x_{t-1,\beta_\tau} \neq 0\) and \(k_t \in [-\tau, 1 - \tau]\) if \(y_{t\tau} - x_{t-1,\beta_\tau} = 0\).

Let \(\mathcal{D} = \{ t : y_{t\tau} - x_{t-1,\hat{\beta}_\tau} = 0 \}\), then

\[
s_j(\hat{\beta}_\tau) = -\sum_{t=1}^{n} x_{t-1,j} \psi_\tau(y_{t\tau} - x_{t-1,\hat{\beta}_\tau}) + \sum_{t \in \mathcal{D}} x_{t-1,j} (k_t^* + (1 - \tau)),
\]

where \(h_n := \sum_{t \in \mathcal{D}} x_{t-1,j} (k_t^* + (1 - \tau))\) and

\[
k_t^* = \begin{cases} 
0, & \text{if } y_{t\tau} - x_{t-1,\hat{\beta}_\tau} \neq 0 \\
 \in [-\tau, 1 - \tau], & \text{if } y_{t\tau} - x_{t-1,\hat{\beta}_\tau} = 0.
\end{cases}
\]

With probability one (Koenker (2005, Section 2.2)), \(|\mathcal{D}| = p\). Thus \(h_n = O_p(p^{3/2}) = O_p(n^{(3/2)\zeta})\).

Next, define the subgradient of the penalized objective function as \(S_j(\beta_\tau)\):

\[
S_j(\beta_\tau) = -\sum_{t=1}^{n} x_{t-1,j} \psi_\tau(y_{t\tau} - x_{t-1,\beta_\tau}) + h_n + \frac{\lambda_n}{|\hat{\beta}_j_\tau|} sgn(\beta_{j_\tau})
\]

\[
= -\sum_{t=1}^{n} x_{t-1,j} \psi_\tau(x_{t\tau} - x_{t-1,\delta_\tau}) + h_n + \frac{\lambda_n}{|\hat{\beta}_j_\tau|} sgn(\beta_{j_\tau}),
\]

33
where $u_{t\tau} = u_{0t} - F_{u_0}^{-1}(\tau)$ and $\delta_{\tau} = \beta_{\tau} - \beta_{0\tau}$. The subgradient condition requires that at the optimum, $\hat{\beta}_{\tau}$,

$$0 \in S_j(\hat{\beta}_{\tau}).$$

That is,

$$\sum_{t=1}^{n} x_{t-1,j}\psi_{\tau} \left( u_{t\tau} - x_{t-1}'\hat{\delta}_{\tau} \right) - h_n = \frac{\lambda_n}{|\hat{\beta}_{j,\tau}|^\gamma} \text{sgn}(\hat{\beta}_{j,\tau}). \quad (A.6)$$

If $j = 1, \ldots, q_n$, it implies that $|\text{sgn}(\hat{\beta}_{j,\tau})| = 1$. Then we can write the subgradient condition (A.6) as:

$$\left| \sum_{t=1}^{n} x_{t-1,j}\psi_{\tau} \left( u_{t\tau} - \hat{\delta}_{\tau} x_{t-1} \right) - h_n \right| = \frac{\lambda_n}{|\hat{\beta}_{j,\tau}|^\gamma}. \quad (A.7)$$

In the following, we show that this subgradient condition does not hold for $j \notin A_0$, i.e., $j = q_n + 1, \ldots, p$. It suffices to show:

(a) $(n^{\alpha + 2\gamma} c_{A(n,p)}^{-1/2})^{-1} \frac{\lambda_n}{|\hat{\beta}_{j,\tau}|^\gamma} \to \infty$;

(b) $\sum_{t=1}^{n} x_{t-1,j}\psi_{\tau} \left( u_{t\tau} - x_{t-1}'\delta_{\tau} \right) - h_n \leq O_p(n^{\alpha + 2\gamma} c_{A(n,p)}^{-1/2})$.

For (b), we first prove that the first term on the left-hand side of (A.7) is dominated by $O_p(n^{\alpha + 2\gamma} c_{A(n,p)}^{-1/2})$:

$$\sum_{t=1}^{n} x_{t-1,j}\psi_{\tau} \left( u_{t\tau} - \hat{\delta}_{\tau} x_{t-1} \right)$$

$$= \sum_{t=1}^{n} x_{t-1,j} \left[ \psi_{\tau} \left( u_{t\tau} - \delta_{\tau} x_{t-1} \right) - E_{t-1}\psi_{\tau} \left( u_{t\tau} - \delta_{\tau} x_{t-1} \right) - \psi_{\tau} (u_{t\tau}) + E_{t-1}\psi_{\tau} (u_{t\tau}) \right]$$

$$+ \sum_{t=1}^{n} x_{t-1,j} E_{t-1}\psi_{\tau} \left( u_{t\tau} - \delta_{\tau} x_{t-1} \right) + \sum_{t=1}^{n} x_{t-1,j}\psi_{\tau} (u_{t\tau})$$

$$\equiv I_1 + I_2 + I_3. \quad (A.8)$$

For $I_1$, let $I_1 = A + B$, where

$$A \equiv \sum_{t=1}^{n} x_{t-1,j} \left[ \psi_{\tau} \left( u_{t\tau} - \delta_{\tau} x_{t-1} \right) - E_{t-1}\psi_{\tau} \left( u_{t\tau} - \delta_{\tau} x_{t-1} \right) \right],$$

$$B \equiv \sum_{t=1}^{n} x_{t-1,j} \left[ \psi_{\tau} (u_{t\tau}) - E_{t-1}\psi_{\tau} (u_{t\tau}) \right].$$
Note that

\[ E_{t-1}\psi_r(u_{tr} - \delta'_r x_{t-1}) = E_{t-1}\psi_r(u_{tr}) + \left. \frac{\partial E_{t-1}\psi_r(u_{tr} - \delta'_r x_{t-1})}{\partial \delta'_r} \right|_{\delta'_r = 0} \delta'_r + o_p(a_n) \]

\[ = -x'_{t-1}f_{t-1}(0)\delta'_r + o_p(a_n), \quad (A.9) \]

\[ E_{t-1}\psi_r^2(u_{tr} - \delta'_r x_{t-1}) = E_{t-1}\psi_r^2(u_{tr}) + \left. \frac{\partial E_{t-1}\psi_r^2(u_{tr} - \delta'_r x_{t-1})}{\partial \delta'_r} \right|_{\delta'_r = 0} \delta'_r + o_p(a_n) \]

\[ = \tau(1 - \tau) - x'_{t-1}f_{t-1}^2(0)\delta'_r + o_p(a_n). \quad (A.10) \]

Then, we have

\[ E(A) = EE_{t-1}(A) = \left\{ \sum_{t=1}^{n} x_{t-1,j} \left[ E_{t-1}\psi_r(u_{tr} - \delta'_r x_{t-1}) - E_{t-1}\psi_r(u_{tr} - \delta'_r x_{t-1}) \right] \right\} = 0, \]

and

\[ E_{t-1}(A^2) = \sum_{t=1}^{n} x_{t-1,j}^2 E_{t-1}\left[ \psi_r(u_{tr} - \delta'_r x_{t-1}) - E_{t-1}\psi_r(u_{tr} - \delta'_r x_{t-1}) \right]^2 + 0 \]

\[ = \sum_{t=1}^{n} x_{t-1,j}^2 \left[ E_{t-1}\psi_r^2(u_{tr} - \delta'_r x_{t-1}) - (E_{t-1}\psi_r(u_{tr} - \delta'_r x_{t-1}))^2 \right] \]

\[ = \sum_{t=1}^{n} x_{t-1,j}^2 \left( \tau(1 - \tau) - x'_{t-1}f_{t-1}^2(0)\delta'_r + o_p(a_n) \right) \]

\[ - \sum_{t=1}^{n} x_{t-1,j}^2 \left( -x'_{t-1}f_{t-1}(0)\delta'_r + o_p(\gamma_n) \right)^2 \quad \text{(by (A.9) and (A.10))} \]

\[ = O_p(n^2) - \sum_{t=1}^{n} x_{t-1,j}^2 x_{t-1}^2 c_f_{t-1}^2(0)O_p(a_n) + O_p(n^2)O_p(a_n) \]

\[ - \sum_{t=1}^{n} x_{t-1,j}^2 \left( c'x_t x_{t-1} x_{t-1}^2 c_f_{t-1}^2(0)O_p(a_n^2) - 2x'_{t-1}c_{t-1}(0)O_p(a_n)O_p(a_n) + O_p(a_n^2) \right). \]

(since \( \delta'_r = \beta'_r - \beta_{0r} = O_p(a_n) \) and \( x'_{t-1} \delta'_r = x'_{t-1}cO_p(a_n) \))  \quad (A.11) \]

Under Assumption \( f \) and the definition of \( \bar{c}_{A(n,p)} \), we have:

\[ E \left[ \sum_{t=1}^{n} c' f_{t-1}(0)x_{t-1} x'_{t-1} c \right]^{1/2} \leq \sum_{t=1}^{n} E c' f_{t-1}(0)x_{t-1} x'_{t-1} c \leq O(\bar{c}_{A(n,p)}^{1/2}n), \quad \text{(by Jensen’s inequality)} \]
Given the conditions: 0 \leq \alpha < 1, it is easy to verify that the fourth term, \(O(n^{2\alpha \zeta+3} |E_{A(n,p)}^2|)\), dominates the other terms. Thus, \(E(A^2) \leq O(n^{2\alpha \zeta+3} |E_{A(n,p)}^2|)\). The Chebyshev’s inequality implies that

\[ A \leq O_p(n^{\alpha \zeta+(3/2)} |E_{A(n,p)}^2|) \]
For B, it is easy to obtain that

\[ E_{t-1}[\psi_{\tau}(u_{t\tau})] = 0, \quad E_{t-1}[\psi_{\tau}^2(u_{t\tau})] = \tau(1 - \tau). \] (A.14)

Then we can show:

\[
E_{t-1}(B) = \sum_{t=1}^{n} x_{t-1,j} \left[ E_{t-1}\psi_{\tau}(u_{t\tau}) - E_{t-1}\psi_{\tau}(u_{t\tau}) \right] = 0,
\]

\[
E_{t-1}(B^2) = \sum_{t=1}^{n} x_{t-1,j}^2 E_{t-1} \left[ \psi_{\tau}(u_{t\tau}) - E_{t-1}\psi_{\tau}(u_{t\tau}) \right]^2 = \sum_{t=1}^{n} x_{t-1,j}^2 E_{t-1}\psi_{\tau}^2(u_{t\tau}) = O_p(n^2),
\]

which implies that

\[ B \leq O_p(n). \] (A.15)

By the results of (A.13) and (A.15), we have

\[ I_1 \leq O_p(n^{\alpha \varsigma + (3/2)c_3^{1/2}} c_{A(n,p)}^{1/2}). \] (A.16)

For I2, by (A.9),

\[
\sum_{t=1}^{n} x_{t-1,j} E_{t-1}\psi_{\tau}(u_{t\tau} - \delta_{x_{t\tau}} x_{t-1}) = -\sum_{t=1}^{n} x_{t-1,j} x_{t'-1} f_{t-1}(0) \delta_{x_{t\tau}} + o_p(a_n) \sum_{t=1}^{n} x_{t-1,j}.
\]

Thus,

\[
E \left[ \sum_{t=1}^{n} x_{t-1,j} E_{t-1}\psi_{\tau}(u_{t\tau} - \delta_{x_{t\tau}} x_{t-1}) \right] = -E \sum_{t=1}^{n} [x_{t-1,j} x_{t'-1} c f_{t-1}(0) O_p(a_n)] + E[a_p(a_n n^{3/2})]
\]

\[ \leq O(a_n) E \left[ \sum_{t=1}^{n} x_{t-1,j} x_{t'-1} c f_{t-1}(0) \right] + o(a_n n^{3/2})
\]

\[ \leq O(a_n) E \left( \max_{1 \leq t \leq n} |x_{t-1,j} c f_{t-1}(0)| \sum_{t=1}^{n} x_{t-1,j} \right) + o(a_n n^{3/2})
\]

\[ \leq O(a_n) O(n^{3/2}) O(c_{f}^{1/2} c_{A(n,p)}^{1/2} n^{3/2}) + o(a_n n^{3/2})
\]

\[ = O(a_n n^{3/2} c_{A(n,p)}^{1/2}) + o(a_n n^{3/2})
\]

\[ = O(n^{\alpha \varsigma + 2 c_{A(n,p)}^{1/2}}), \] (A.17)

and

\[
E \left[ \left( \sum_{t=1}^{n} x_{t-1,j} E_{t-1}\psi_{\tau}(u_{t\tau} - \delta_{x_{t\tau}} x_{t-1}) \right)^2 \right]
\]
\[
E \left[ -\sum_{t=1}^{n} x_{t-1,j} x'_{t-1} f_{t-1}(0) \delta_{\tau} + o_{p}(a_{n}) \sum_{t=1}^{n} x_{t-1,j} \right]^2 \\
= \ E \left[ \left( \sum_{t=1}^{n} x_{t-1,j} x'_{t-1} f_{t-1}(0) \delta_{\tau} \right)^2 \right] + o_{p}(a_{n}) \sum_{t=1}^{n} x_{t-1,j}^2 \\
- 2 \left( \sum_{t=1}^{n} x_{t-1,j} x'_{t-1} f_{t-1}(0) \delta_{\tau} \right) \cdot \left( o_{p}(a_{n}) \sum_{t=1}^{n} x_{t-1,j} \right) \\
\leq O(a_{n}^2) E \left[ \sum_{t=1}^{n} \left( x_{t-1,j} x'_{t-1} f_{t-1}(0) \right)^2 \right] + 2 \sum_{t=1}^{n-1} \sum_{t=2}^{n} x_{t-1,j} x_{k-1,j} x'_{t-1} f_{t-1}(0) \cdot x'_{k-1} f_{k-1}(0) \\
+ o(a_{n}^2) \left[ O(n^2) + O(n^3) \right] + o(a_{n}^2) E \left[ \left( \max_{1 \leq t \leq n} |x'_{t-1} f_{t-1}(0)| \right) \sum_{t=1}^{n} x_{t-1,j} \cdot O_p(n^{3/2}) \right] \\
\leq O(a_{n}^2) E \left[ \left( \max_{1 \leq t \leq n} |x'_{t-1} f_{t-1}(0)| \right)^2 \sum_{t=1}^{n} x_{t-1,j}^2 \right] + O(a_{n}^2) E \left[ \left( \max_{1 \leq t \leq n} |x'_{t-1} f_{t-1}(0)| \right)^2 \sum_{t=2}^{n} x_{t-1,j} x_{k-1,j} \right] \\
+ o(a_{n}^2 n^3) + o(a_{n}^2) E \left[ \left( \max_{1 \leq t \leq n} |x'_{t-1} f_{t-1}(0)| \right) \sum_{t=1}^{n} x_{t-1,j} \right] \cdot O_p(n^{3/2}) \\
\leq O(a_{n}^2) O(n^2) O(c f \tilde{c}_{A(n,p)} n^3) + O(a_{n}^2) O(n^3) O(c f \tilde{c}_{A(n,p)} n^3) + o(a_{n}^2 n^3) + o(a_{n}^2) O(n^{3/2}) O(n^{3/2}) O(c f \tilde{c}_{A(n,p)} n^{3/2}) \\
= O(a_{n}^2 n^5 \tilde{c}_{A(n,p)}) + O(a_{n}^2 n^6 \tilde{c}_{A(n,p)}) + o(a_{n}^2 n^3) + o(a_{n}^2 n^{9/2} \tilde{c}_{A(n,p)}^{1/2}) \\
= O(a_{n}^2 n^6 \tilde{c}_{A(n,p)}) = O(n^{2\alpha + 4} \tilde{c}_{A(n,p)}), \quad (A.18)
\]
where the second inequality holds using Cauchy-Schwarz inequality and the following result:

\[
\sum_{t=2}^{n} \sum_{k=1}^{t-1} x_{t-1,j} x_{k-1,j} \leq \sum_{t=2}^{n} \left[ x_{t-1,j} \left( \sum_{k=1}^{t-1} x_{k-1,j} \right) \right] \\
\leq \left( \sum_{t=2}^{n} x_{t-1,j}^2 \right)^{1/2} \cdot \left[ \sum_{t=2}^{n} \left( \sum_{k=1}^{t-1} x_{k-1,j} \right)^2 \right]^{1/2} \\
= (O_p(n^2))^{1/2} \cdot \left[ \sum_{t=2}^{n} O_p(t^3) \right]^{1/2} \\
= O_p(n) \cdot O_p(1) \left[ \sum_{t=2}^{n} t^3 \right]^{1/2} \\
= O_p(n) \cdot O_p(1) \left[ O(n^4) \right]^{1/2} \\
= O_p(n^3).
\]

By the results of (A.17) and (A.18), we apply Chebyshev’s inequality and obtain that:

\[
I_2 \leq O_p(n^{\alpha \zeta + 2})^{1/2} A(n,p). \tag{A.19}
\]

For \( I_3 \), it can be shown that

\[
E_{t-1} \sum_{t=1}^{n} x_{t-1,j} \psi_\tau(u_{t\tau}) = \sum_{t=1}^{n} x_{t-1,j} E_{t-1} \psi_\tau(u_{t\tau}) = 0,
\]
\[
E_{t-1} \left[ \sum_{t=1}^{n} x_{t-1,j} \psi_\tau(u_{t\tau}) \right]^2 = \sum_{t=1}^{n} x_{t-1,j}^2 E_{t-1} \psi_\tau^2(u_{t\tau}) + 0 = \tau(1 - \tau) O_p(n^2).
\]

This implies that

\[
I_3 \leq O_p(n). \tag{A.20}
\]

By the results of (A.16), (A.19) and (A.20), we find the upper bound of the first term in (b):

\[
\sum_{t=1}^{n} x_{t-1,j} \psi_\tau(u_{t\tau} - \delta_\tau x_{t-1}) \leq O_p(n^{\alpha \zeta + 2} \tau^{1/2} A(n,p)).
\]

For the second term in (b), under Assumption \( U1 \) (iii), we have

\[
\frac{h_n}{n^{\alpha \zeta + 2} \tau^{1/2} A(n,p)} = \frac{O_p(p^{3/2})}{n^{\alpha \zeta + 2} \tau^{1/2} A(n,p)} = O_p(1) o(1) = o_p(1).
\]
Thus the result in (b) is shown:

\[
\sum_{t=1}^{n} x_{t-1,j} \psi_{\tau} \left( u_{t\tau} - x'_{t-1} \delta_{\tau} \right) - h_n \leq O_p(n^{\alpha \zeta + 2/\gamma} A(n,p)).
\]

(A.21)

For (a), under Assumption \(\lambda 1 \ (ii)\) that \(\frac{\lambda n n^{(1-\alpha) \gamma}}{n^{\alpha \zeta + 2} A(n,p)} \rightarrow \infty\), we obtain:

\[
(n^{\alpha \zeta + 2} A(n,p))^{-1} \frac{\lambda n n^{(1-\alpha) \gamma}}{|\beta_{j,\tau}|^\gamma} = \frac{\lambda n n^{(1-\alpha) \gamma}}{n^{\alpha \zeta + 2} A(n,p) |(n/p)^{\frac{1}{\gamma}} \beta_{j,\tau}|^\gamma} = O_p(1) \rightarrow \infty.
\]

(A.22)

By (A.7), (A.21), and (A.22), we can show that, for any \(j \notin A_0\),

\[
Pr \left( j \in \tilde{A}_n \right) \leq Pr \left( \left| \sum_{t=1}^{n} x_{t-1,j} \psi_{\tau} \left( u_{t\tau} - \tilde{x}'_{t-1} \right) - h_n \right| = \frac{\lambda n}{|\beta_{j,\tau}|^\gamma} \right) = Pr \left( \frac{1}{n^{\alpha \zeta + 2} A(n,p)} \left| \sum_{t=1}^{n} x_{t-1,j} \psi_{\tau} \left( u_{t\tau} - \tilde{x}'_{t-1} \right) - h_n \right| = \frac{\lambda n n^{(1-\alpha) \gamma}}{n^{\alpha \zeta + 2} A(n,p) |(n/p)^{\frac{1}{\gamma}} \beta_{j,\tau}|^\gamma} \right) \rightarrow 0.
\]

B Proofs for Section 4.2

Proof of Lemma 4.2: Let \(Q_n^{QR}(\beta_{\tau})\) be the (unpenalized) QR objective function. To show the consistency of QR estimator, it suffices to show that for any \(\epsilon > 0\), there exists a sufficiently large \(C\) such that

\[
P \left\{ \inf_{||c||=C} Q_n^{QR}(\tilde{\beta}_{\tau} + a_n M_n c) > Q_n^{QR}(\tilde{\beta}_{\tau}) \right\} \geq 1 - \epsilon,
\]

where \(a_n M_n = \left( \begin{array}{cc} a_n^{(0)} I_r & 0 \\ 0 & a_n^{(1)} I_{p-r} \end{array} \right)\) so that \(a_n M_n c = \left( a_n^{(0)} c_1, \ldots, a_n^{(0)} c_r, a_n^{(1)} c_{r+1}, \ldots a_n^{(1)} c_p \right)'\).

As is shown in the proof of Lemma 4.1, this proof can be completed if we show that the following term is positive:

\[
Q_n^{QR}(\tilde{\beta}_{\tau} + a_n M_n c) - Q_n^{QR}(\tilde{\beta}_{\tau}) = \sum_{t=1}^{n} \rho_{\tau}(u_{t\tau} - \tilde{x}'_{t-1} a_n M_n c) - \sum_{t=1}^{n} \rho_{\tau}(u_{t\tau}).
\]
By Knight’s Identity,

\[
\sum_{t=1}^{n} \left[ \rho_{\tau}(u_{t\tau} - \tilde{x}'_{t-1} a_n M_n c) - \rho_{\tau}(u_{t\tau}) \right] = -a_n \sum_{t=1}^{n} \tilde{x}'_{t-1} M_n c \cdot \psi_{\tau}(u_{t\tau}) + \sum_{t=1}^{n} \int_{0}^{\tilde{x}'_{t-1} a_n M_n c} (1(u_{t\tau} \leq s) - 1(u_{t\tau} \leq 0)) \, ds
\]

\[
= -a_n \sum_{t=1}^{n} \tilde{x}'_{t-1} M_n c \cdot \psi_{\tau}(u_{t\tau}) + \sum_{t=1}^{n} E \left[ \int_{0}^{\tilde{x}'_{t-1} a_n M_n c} (1(u_{t\tau} \leq s) - 1(u_{t\tau} \leq 0)) \, ds \right]
\]

\[
+ \sum_{t=1}^{n} \left\{ \int_{0}^{\tilde{x}'_{t-1} a_n M_n c} (1(u_{t\tau} \leq s) - 1(u_{t\tau} \leq 0)) \, ds - E \left[ \int_{0}^{\tilde{x}'_{t-1} a_n M_n c} (1(u_{t\tau} \leq s) - 1(u_{t\tau} \leq 0)) \, ds \right] \right\}
\]

\[
\equiv I_1 + I_2 + I_3.
\]

We will show that \( I_1 \) and \( I_2 \) are dominated by \( I_2 \) and that \( I_2 > 0 \).

First, we derive the upper bound of \( I_1 \). For a large \( n \),

\[
E[I_1] = a_n^2 \sum_{t=1}^{n} c' M_n' E \left[ \psi_{\tau}(u_{t\tau})^2 \tilde{x}_{t-1} \tilde{x}'_{t-1} \right] M_n c + 2a_n^2 \sum_{t=2}^{n} \sum_{k=1}^{t-1} c' M_n' E \left[ \psi_{\tau}(u_{t\tau}) \psi_{\tau}(u_{k\tau}) \tilde{x}_{t-1} \tilde{x}'_{t-1} \right] M_n c
\]

\[
= a_n^2 \sum_{t=1}^{n} c' M_n' E \left[ \psi_{\tau}(u_{t\tau})^2 \tilde{x}_{t-1} \tilde{x}'_{t-1} \right] M_n c
\]

\[
= a_n^2 c' M_n' \sum_{t=1}^{n} E \left[ \psi_{\tau}(u_{t\tau})^2 \tilde{x}_{t-1} \tilde{x}'_{t-1} \right] M_n c
\]

\[
\leq a_n^2 n^2 \varepsilon_{B(n,p)} C^2 \quad \text{, (from Assumption U2)}
\]

so that

\[
I_1 = O \left( a_n n \varepsilon_{B(n,p)}^{1/2} \right).
\]

For \( I_2 \), for a large \( n \),

\[
I_2 = \sum_{t=1}^{n} E \int_{0}^{\tilde{x}'_{t-1} a_n M_n c} (1(u_{t\tau} \leq s) - 1(u_{t\tau} \leq 0)) \, ds
\]

\[
= \sum_{t=1}^{n} E \int_{0}^{\tilde{x}'_{t-1} a_n M_n c} (f_{t-1}(0) \cdot s) \, ds \{1 + o_p(1)\}
\]

\[
= \frac{1}{2} a_n^2 \sum_{t=1}^{n} c' M_n' E \left[ f_{t-1}(0) \tilde{x}_{t-1} \tilde{x}'_{t-1} \right] M_n c \{1 + o_p(1)\}
\]

\[
= \frac{1}{2} a_n^2 c' \left( M_n' \sum_{t=1}^{n} E \left[ f_{t-1}(0) \tilde{x}_{t-1} \tilde{x}'_{t-1} \right] M_n \right) c \{1 + o_p(1)\}
\]

\[
\geq \frac{1}{4} a_n^2 \beta A(n,p) C^2 n^2 \quad \text{by Assumption L2}
\]

\[
= O(a_n^2 n^2 \varepsilon_{A(n,p)}).
\]
Finally, for $I_3$, for a large $n$,

$$Var(I_3) = Var \left( \sum_{t=1}^{n} \int_{0}^{x_{t-1}a_nM_{n,c}} (1(u_{t,r} \leq s) - 1(u_{t,r} \leq 0)) ds \right)$$

$$\leq E \left( \sum_{t=1}^{n} \int_{0}^{x_{t-1}a_nM_{n,c}} (1(u_{t,r} \leq s) - 1(u_{t,r} \leq 0)) ds \right)^2$$

$$\leq a_n^2 \sum_{t=1}^{n} c'M_n'E [x'_{t-1}x_{t-1}'] M_n c + 2a_n^2 \sum_{t=2}^{n} \sum_{k=1}^{t-1} E \left| x'_{t-1}M_n c \right| \left| x'_{k-1}M_n c \right|$$

$$= V_{3,1} + V_{3,2}.$$ 

By the proof of Lemma 4.1, we can easily show that the following bounds hold

$$V_{3,1} \leq O(a_n^2 c A(n,p) n^2), \quad \text{and} \quad V_{3,2} \leq O(a_n^2 c A(n,p) n^3).$$

Based on the results above, we have

$$I_3 = O_p(\bar{c}_A(n,p)a_n n^{3/2}),$$

which establishes the desired result. 

**Remark B.1** Define $H^{-1}M_n = M_n^*$. The normalizing matrix then is

$$\left( \begin{array}{cccc} \sqrt{n}I_{p_2} & 0 & 0 & 0 \\ 0 & \sqrt{n}I_{p_1} & 0 & 0 \\ 0 & 0 & -\sqrt{n}A'_1 & I_{p_2} \\ 0 & 0 & 0 & I_{p_x} \end{array} \right),$$

and

$$a_n^{(1)}M^{*c} = \frac{p^\alpha}{n} \left( \begin{array}{cccc} \sqrt{n}I_{p_2} & 0 & 0 & 0 \\ 0 & \sqrt{n}I_{p_1} & 0 & 0 \\ 0 & 0 & -\sqrt{n}A'_1 & I_{p_2} \\ 0 & 0 & 0 & I_{p_x} \end{array} \right) c = \frac{p^\alpha}{\sqrt{n}} \left( \begin{array}{cccc} 0 & 0 & 0 & 0 \\ 0 & \frac{p^\alpha}{\sqrt{n}I_{p_1}} & 0 & 0 \\ 0 & 0 & -\frac{p^\alpha}{\sqrt{n}A'_1} & \frac{p^\alpha}{n}I_{p_2} \\ 0 & 0 & 0 & \frac{p^\alpha}{n}I_{p_x} \end{array} \right) c$$

where

$$(c^*_1, ..., c^*_p)' = A'_1(c_{p+1}, ..., c_{p+p_1}).$$

Therefore, the reduced rate for the 3rd block component is well accommodated. We define the
convergence rates \( \tilde{a}_{n,j}^* \) by

\[
\tilde{a}_{n,j}^* = \begin{cases} \frac{p_n}{\sqrt{n}} = a_n^{(0)}, & \text{for } j = 1, \ldots, r + p_2, \\ \frac{p_n}{n} = a_n^{(1)}, & \text{for } j = r + p_2 + 1, \ldots, p. \end{cases}
\]

**Proof of Theorem 4.4:** Following the proof of Theorem 4.1, it suffices to show that for any \( \epsilon > 0 \), there exists a sufficiently large \( C \) such that

\[
P\left\{ \inf_{||c|| \leq C} Q_n(\beta_{0\tau} + a_n^{(1)} M_n^* c) > Q_n(\beta_{0\tau}) \right\} \geq 1 - \epsilon,
\]

where \( M_n^* = Q^{-1} M_n \) as in Remark B.1.

This inequality implies that with probability at least \( 1 - \epsilon \), there is a local minimizer \( \hat{\beta}_\tau^* \) in the shrinking ball \( \{\beta_{0\tau} + a_n^{(1)} M_n^* c, ||c|| \leq C\} \) such that \( ||\hat{\beta}_\tau^* - \beta_{0\tau}|| = O_p(\tilde{a}_{n,j}^*) \), where \( \tilde{a}_{n,j}^* \) is the \( j \)-th dominating rates from \( a_n^{(1)} M_n^* c \), i.e.,

\[
\tilde{a}_{n,j}^* = \begin{cases} \frac{p_n}{\sqrt{n}} = a_n^{(0)}, & \text{for } j = 1, \ldots, r + p_2, \\ \frac{p_n}{n} = a_n^{(1)}, & \text{for } j = r + p_2 + 1, \ldots, p. \end{cases}
\]

Then we obtain

\[
Q_n(\beta_{0\tau} + a_n^{(1)} M_n^* c) - Q_n(\beta_{0\tau}) \\
\geq \left( \sum_{t=1}^{n} \rho_\tau(u_{t\tau} - X_{t-1} a_n^{(1)} M_n^* c) - \sum_{t=1}^{n} \rho_\tau(u_{t\tau}) \right) + \frac{q_n}{n} \lambda_{n,j}(||\beta_{0\tau,j} + \tilde{a}_{n,j}^* c_j| - ||\beta_{0\tau,j}||)
\]

\[
= d_1^* + d_2^*.
\]

For \( d_1^* \), similarly to Theorem 4.1, we have

\[
\sum_{t=1}^{n} \rho_\tau(u_{t\tau} - X_{t-1} a_n^{(1)} M_n^* c) - \sum_{t=1}^{n} \rho_\tau(u_{t\tau}) \\
= -a_n^{(1)} \sum_{t=1}^{n} X_{t-1} M_n^* c \cdot \psi_\tau(u_{t\tau}) + \sum_{t=1}^{n} \int_{0}^{X_{t-1} a_n^{(1)} M_n^* c (1(u_{t\tau} \leq s) - 1(u_{t\tau} \leq 0)) ds.
\]

Note that

\[
M_n^r \sum_{t=1}^{n} E \left[ f_{t-1}(0) X_{t-1} X_{t-1} \right] M_n^r = M_n^r (H^{-1})' \sum_{t=1}^{n} E \left[ f_{t-1}(0) X_{t-1} X_{t-1} H^{-1} M_n \right]
\]

\[
= M_n^r \sum_{t=1}^{n} E \left[ f_{t-1}(0) \left( H^{-1} \right)' X_{t-1} X_{t-1} H^{-1} \right] M_n
\]

\[
= M_n^r \sum_{t=1}^{n} E \left[ f_{t-1}(0) \tilde{x}_{t-1} \tilde{x}_{t-1}' \right] M_n,
\]

43
which is controlled by Assumptions $L2$ and $U2$. Thus, using the exactly same proof of Theorem 4.1, the dominating order in $d_1^*$ is

$$
\sum_{t=1}^{n} E \int_{0}^{x_{t-1}^{(a_n(1))} M_n^c} (1(ut \leq s) - 1(u_{t-1} \leq 0)) \, ds \\
= \sum_{t=1}^{n} E \int_{0}^{x_{t-1}^{(a_n(1))} M_n^c} (f_{t-1}(0) \cdot s) \, ds \{1 + o_p(1)\} \\
= \frac{1}{2} \left( a_n^{(1)} \right)^2 \left( \beta_j \sum_{t=1}^{n} E [f_{t-1}(0)X_{t-1}X_{t-1}'] M_n^c \right) c \{1 + o_p(1)\} \\
\geq O_p \left( a_n^{(1)} \right)^2 n^2 e_{A(n,p)} = O_p \left( p^{2\alpha} e_{A(n,p)} \right).
$$

For $d_2^*$, the only differences with the proof of Theorem 4.1 are the rate of divergence in $\lambda_{n,j}$’s for $j = 1, ..., r + p_2$, and for $j = r + p_2 + 1, ..., p$. From Corollary 4.3,

$$
\hat{\beta}_{r,j} = \begin{cases} 
O_p(a_n^{(0)}) + \beta_{0r,j} = o_p(1) + \beta_{0r,j}, & j = 1, ..., r + p_2, \\
O_p(a_n^{(1)}) + \beta_{0r,j} = o_p(1) + \beta_{0r,j}, & j = r + p_2 + 1, ..., p,
\end{cases}
$$

and clearly $a_n^{(0)} > a_n^{(1)}$ for any given $n$. Since $\beta_{0r,j} \neq 0$ for $j = 1, ..., q_n$, we have

$$
\sum_{j=1}^{q_n} \lambda_{n,j}(|\beta_{0r,j} + a_{n,j}^* c_j| - |\beta_{0r,j}|) \leq \sum_{j=1}^{q_n} \lambda_{n,j} a_{n,j}^* |c_j| = \lambda_n \sum_{j=1}^{q_n} a_{n,j}^* |c_j| \frac{1}{|\beta_{r,j}|^{\gamma}} \\
\leq \lambda_n \max_{j} (a_{n,j}^*) \left( \sum_{j=1}^{q_n} \frac{1}{(o_p(1) + \beta_{0r,j})^{2\gamma}} \right)^{1/2} ||c|| \\
= \lambda_n a_n^{(0)} q_n^{1/2} O_p(1) \\
= O_p \left( \lambda_n (p^{\alpha}/\sqrt{n}) p^{1/2} \right) \\
= O_p \left( \lambda_n \frac{n^{3/2 + \alpha}}{n^{1/2}} \right),
$$

so is dominated by $d_1^* = O_p \left( p^{2\alpha} e_{A(n,p)} \right)$ under the condition

$$
\lambda_n \frac{n^{2\alpha}}{n^{1/2} e_{A(n,p)}} = \lambda_n \frac{n^{(3/2 - \alpha)\zeta}}{n^{1/2} e_{A(n,p)}} = \lambda_n \frac{n^{2\alpha}}{n^{1/2 + \alpha \zeta} e_{A(n,p)}} \rightarrow 0.
$$

**Proof of Theorem 4.5:** Among $\hat{\beta}_r^{ALQR^*} = (\hat{\gamma}_{r}, \hat{\gamma}_{1r}, \hat{\gamma}_{2r}, \hat{\gamma}_{3r}, \hat{\gamma}_{4r})'$, we just need to show ALQR sparsity for $\hat{\beta}_{2r}^{*}$, since other parts are shown by Theorem 4.2 and the existing proof for the $I(0)$ case. Among $j = r + 1, ..., r + p_2$, for any $j \notin A_0$, $|\hat{\beta}_{j,r} - \beta_{j,0r}| = |\hat{\beta}_{j,r}^*| = O_p \left( \frac{p^{\alpha}}{\sqrt{n}} \right)$, so
following the proof of Theorem 4.2,

\[
\Pr \left( j \in \hat{\mathcal{A}}_n \right) \\
\leq \Pr \left( \sum_{t=1}^{n} x_{t-1,j} \psi'_\tau \left( u_{t}\tau - \delta'_\tau x_{t-1} \right) - h_n = \frac{\lambda_n}{|\delta'_\tau|^{\gamma}} \right) \\
= \Pr \left( \frac{1}{n^{\alpha\varsigma + 2\frac{1}{2}\gamma} A_{(n,p)}^{-1}} \sum_{t=1}^{n} x_{t-1,j} \psi'_\tau \left( u_{t}\tau - \delta'_\tau x_{t-1} \right) - h_n \right) = \frac{\lambda_n \sqrt{n}}{n^{\alpha\varsigma + 2\frac{1}{2}\gamma} A_{(n,p)}^{-1}} \left[ 1 - \frac{1}{\sqrt{n} \gamma |\delta'_\tau|^{\gamma}} \right] \\
\rightarrow 0,
\]

as long as

\[
\frac{\lambda_n \left( \frac{\sqrt{n}}{p^n} \right)^\gamma}{n^{\alpha\varsigma + 2\frac{1}{2}\gamma} A_{(n,p)}^{-1}} = \frac{\lambda_n n^{(1/2-\alpha\varsigma)\gamma}}{n^{\alpha\varsigma + 2\frac{1}{2}\gamma} A_{(n,p)}^{-1}} \rightarrow \infty.
\]

\[\blacksquare\]

C Proof for Section 5

**Proof of Theorem 5.1:** We first show the limit theory of ALQR estimator with I(0) predictors. For simplicity, we use A and B to represent \( A_{(t,p)} \) and \( B_{(t,p)} \), respectively. For an \( m \times n \) real matrix A, we denote its Frobenius norm as \( ||A||_F := \sqrt{\text{trace}(A^T A)} \). Recall that the I(0) part of our predictors is a linear process of the form \( z_t = \sum_{j=0}^{\infty} D_j \epsilon_{t-j} \) with the conditions given in Section 2.2. In this proof, we use notation \( x_t = \sum_{j=0}^{\infty} D_j \epsilon_{t-j} \) for this I(0) predictors for simplicity.

Moreover, for simple notation, we drop some superscript/subscript and write \( \hat{\delta}^{(0), ALQR*} \) as \( \hat{\delta} \) unless we need some clarity. Also, without loss of generality, we set the first \( q_n \) elements of \( \hat{\delta} \) be non-zero. Let \( \delta = \sqrt{n}(\beta - \beta_0) \) and subscript \( A \) denote elements of the active set. Thus, the oracle (local) estimator is \( \hat{\delta} = (\hat{\delta}_A, 0) \). The oracle property of the ALQR estimator in Section 4 implies that \( \hat{\delta} \) is a minimizer of

\[
\sum_{t=1}^{n} \rho_{\tau} \left( y_t - x'_t \left( n^{-1/2} \delta + \beta_0 \right) \right) + \sum_{j=1}^{p} \lambda_{n,j} \left| n^{-1/2} \delta + \beta_0 \right|
\]

with probability approaching 1. Define

\[
V_j(\delta) = -n^{-1/2} \sum_{t=1}^{n} \psi_{\tau} \left( y_t - x'_t \left( n^{-1/2} \delta + \beta_0 \right) \right) x_{t,j} + n^{-1/2} \lambda_{n,j} \text{sgn} \left( n^{-1/2} \delta_j + \beta_{0,j} \right).
\]

Using vector notation, we have

\[
V(\hat{\delta}) - V(0) = -n^{-1/2} \sum_{t=1}^{n} \left[ \psi_{\tau} \left( y_t - x'_t \left( n^{-1/2} \hat{\delta} + \beta_0 \right) \right) - \psi_{\tau} \left( u_t \right) \right] x_t + n^{-1/2} \omega_n(\hat{\delta}),
\]

where \( \omega_n(\hat{\delta}) = \left( \lambda_{n,1} \left| \hat{\delta}_{\tau,1} \right|^{-\gamma} \text{sgn} \left( n^{-1/2} \delta_1 + \beta_{0,1} \right), \ldots, \lambda_{n,q} \left| \hat{\delta}_{\tau,q} \right|^{-\gamma} \text{sgn} \left( n^{-1/2} \delta_q + \beta_{0,q} \right), 0, \ldots, 0 \right) \).
Note that, except the penalty term $n^{-1/2} \omega_n(\hat{\delta})$,

$$V(\hat{\delta}) - V(0) = \left[ V(\hat{\delta}) - V(0) - E\left[ V(\hat{\delta}) \right] + E\left[ V(0) \right] \right] + \left[ E\left[ V(\hat{\delta}) \right] - E\left[ V(0) \right] + A\hat{\delta} \right] - A\delta$$

and by rearranging

$$A\delta - V(0) = \left[ V(\hat{\delta}) - V(0) - E\left[ V(\hat{\delta}) \right] + E\left[ V(0) \right] \right] + \left[ E\left[ V(\hat{\delta}) \right] - E\left[ V(0) \right] + A\hat{\delta} \right] - V(\hat{\delta}). \quad (C.24)$$

Define the weighted norm $|| \cdot ||_c$ by

$$||A||_c = ||cA||$$

where $c$ is an arbitrary $l \times p$ matrix with $||c|| \leq \sqrt{\frac{1}{2}} B L c$ for a large constant $L_c < \infty$. Since $l$ is fixed, without loss of generality we assume that $l = 1$.

Using this weighted norm $|| \cdot ||_c$, we will show

$$\|A\delta - V(0)\|_c \leq \sup_{||\hat{\delta}|| \leq \sqrt{pL}} \left[ \left\| V(\hat{\delta}) - V(0) - E\left[ V(\hat{\delta}) \right] + E\left[ V(0) \right] \right\|_c + \left\| E\left[ V(\hat{\delta}) \right] - E\left[ V(0) \right] + A\delta \right\|_c + \left\| V(\hat{\delta}) \right\|_c \right] = o_p(1), \quad (C.25)$$

for a large constant $L$.

We denote the terms on the RHS of (C.25) as:

(CA.6) := $\sup_{||\hat{\delta}|| \leq \sqrt{pL}} \left[ \left\| V(\hat{\delta}) - V(0) - E\left[ V(\hat{\delta}) \right] + E\left[ V(0) \right] \right\|_c \right.$

(CA.7) := $\sup_{||\hat{\delta}|| \leq \sqrt{pL}} \left[ \left\| E\left[ V(\hat{\delta}) \right] - E\left[ V(0) \right] + A\delta \right\|_c \right.$

(CA.8) := $\sup_{||\hat{\delta}|| \leq \sqrt{pL}} \left[ \left\| V(\hat{\delta}) \right\|_c \right.$.

For (CA.6), we need to show that for any large constant $L < \infty$,

$$\sup_{||\hat{\delta}|| \leq \sqrt{pL}} \left[ \left\| V(\hat{\delta}) - V(0) - E\left[ V(\hat{\delta}) - V(0) \right] \right\|_c \right] = o_p(1).$$

First, we write $a_t \equiv cx_t = a_t^+ - a_t^-$ where $a_t^+ = max\{a_t, 0\}$ and $a_t^- = max\{-a_t, 0\}$. By Minkowski’s inequality, we obtain

$$\sup_{||\hat{\delta}|| \leq \sqrt{pL}} \left[ \left\| V(\hat{\delta}) - V(0) - E\left[ V(\hat{\delta}) - V(0) \right] \right\|_c \right. \leq \sup_{||\hat{\delta}|| \leq \sqrt{pL}} \left[ \left\| V^+(\hat{\delta}) - V^+(0) - E\left[ V^+(\hat{\delta}) - V^+(0) \right] \right\| + \sup_{||\hat{\delta}|| \leq \sqrt{pL}} \left[ \left\| V^-(\hat{\delta}) - V^-(0) - E\left[ V^-(\hat{\delta}) - V^-(0) \right] \right\|, \quad (C.26)$$
where $V^+(\hat{\delta}) \equiv n^{-1/2} \sum_{t=1}^{n} \psi_t \left( y_t - x_{t-1}' \left( n^{-1/2} \hat{\delta} + \beta_0 \right) \right) a_{t-1}^+$, and $V^-(\hat{\delta})$ is analogously defined. Thus, it suffices to show that each term on the right hand side of (C.26) is $o_p(1)$.

We show the first term of (C.26) is $o_p(1)$. Let $D \equiv \{ \delta \in \mathbb{R}^p : ||\delta|| \leq \sqrt{p}L \}$ for some finite constant $L$. Let $|t|_\infty$ denote the maximum of the absolute values of the coordinates of $t$. We select $N_1 = (2n^2)^p$ grid points, $\delta_1, \ldots, \delta_{N_1}$ and cover $D$ by cubes $D_s = \{ \delta \in \mathbb{R}^p : |\delta - \delta_s|_\infty \leq \delta_{en} \}$ with sides of length $\delta_{en}$ where $\delta_{en} = Lp^{1/2}/n^2$. Since $\psi_t(\cdot)$ is monotone, by the Minkowski’s inequality we obtain that

$$
\sup_{||\delta|| \leq \sqrt{p}L} \left| V^+(\hat{\delta}) - V^+(0) - E \left[ V^+(\hat{\delta}) - V^+(0) \right] \right|
\leq \max_{1 \leq s \leq N_1} \left| V^+(\hat{\delta}) - V^+(0) - E \left[ V^+(\hat{\delta}) - V^+(0) \right] \right|
+ \max_{1 \leq s \leq N_1} \left| \sum_{t=1}^{n} E \left[ \psi_{st}(\delta_{en})a_{t-1}^+ \right] - E \left[ \psi_{st}(\delta_{en})a_{t-1}^+ \right] \right|
+ \max_{1 \leq s \leq N_1} \left| \sum_{t=1}^{n} \left[ \psi_{st}(\delta_{en}) - \psi_{st}(0) \right] a_{t-1}^+ \right| - E \left\{ \left[ \psi_{st}(\delta_{en}) - \psi_{st}(0) \right] a_{t-1}^+ \right\}
\equiv I_1 + I_2 + I_3,
$$

where $\psi_{st}(\delta) = \psi_t \left( u_{t\tau} - n^{-1/2} \hat{\delta} x_{t-1} + n^{-1/2} \hat{\delta} ||x_{t-1}|| \right)$.

For $I_1$, note that

$$
V^+(\hat{\delta}) - V^+(0) - E \left[ V^+(\hat{\delta}) - V^+(0) \right] = n^{-1/2} \sum_{t=1}^{n} \left[ \psi_t \left( y_t - x_{t-1}' \left( n^{-1/2} \hat{\delta} + \beta_0 \right) \right) a_{t-1}^+ - \psi_t \left( y_t - x_{t-1}' \beta_0 \right) a_{t-1}^+ \right]
- n^{-1/2} \sum_{t=1}^{n} E \left[ \psi_t \left( y_t - x_{t-1}' \left( n^{-1/2} \hat{\delta} + \beta_0 \right) \right) a_{t-1}^+ - \psi_t \left( y_t - x_{t-1}' \beta_0 \right) a_{t-1}^+ \right]
= n^{-1} \sum_{t=1}^{n} \eta_{ts}
= n^{-1} \sum_{t=1}^{n} \eta_{ts} 1 \left\{ a_{t-1}^+ \leq e_{1n} \right\} + n^{-1} \sum_{t=1}^{n} \eta_{ts} 1 \left\{ a_{t-1}^+ > e_{1n} \right\}
\equiv D_{1s} + D_{2s},
$$

where $\eta_{ts} = n^{1/2} [\eta_{ts,0} - E(\eta_{ts,0})]$, $\eta_{ts,0} = \left[ \psi_t (u_{t\tau} - n^{-1/2} \hat{\delta} x_{t-1}) - \psi_t (u_{t\tau}) \right] a_{t-1}^+$, and $e_{1n} = (np^{4/2})^{1/8}$. To prove that $I_1 = o_p(1)$, it suffices to show

$$
\max_{1 \leq s \leq N_1} \left| D_{ks} \right| = o_p(1) \quad \text{for } k = 1 \text{ and } 2.
$$

For the I(0) predictors $x_t = \sum_{j=0}^{\infty} D_j e_{t-j}$, we have $D_j = O \left( e^{-v_j} \right)$ with $v > 0$ along with the conditions (1), (2), (5) of Withers (1981): $\{e_j\}_{j=-\infty}^{\infty}$ are independent r.v’s with characteristic
functions \( \{ \phi_j \} \) such that

\[
(1): \quad (2\pi)^{-1} \max_j \int |\phi_j(t)| \, dt < \infty, \quad (C.27)
\]

\[
(2): \quad \max_j E|\epsilon_j|^{\delta} < \infty \text{ for some } \delta > 0, \quad (C.28)
\]

\[
(5): \quad \sup_{m,s,k \geq 1} \sup_{\alpha,\beta,v} \max_t \left| \frac{\partial}{\partial v_t} P(W + v \in \cup_1^s D_j) \right| < \infty, \quad (C.29)
\]

where

\[
D_j = \mathbb{X}_{t=k}^{k+m-1} (\alpha_{jt}, \beta_{jt}), \quad v = (v_k, \ldots, v_{k+m-1}), \quad W = (W_k, \ldots, W_{k+m-1})
\]

with \( W_t = \sum_{j=0}^t D_j \epsilon_{t-j} \) and \( \mathbb{X} \) represents a product space.

As a result, \( x_t \) is strong mixing with \( \alpha(j) = O(e^{-\nu\lambda j}) \) with another constant \( \lambda > 0 \) by Corollary 2 of Withers (1981). By the invariance property of the strong mixing processes, \( \frac{\eta_{ts}}{n^{1/2}} = [\eta_{ts,0} - E(\eta_{ts,0})] \) is also strong mixing with \( \alpha(j) = O(e^{-\nu\lambda j}) \), and

\[
\frac{\eta_{ts}}{n^{1/2}} \mathbb{1} \{ a_{t-1}^+ \leq \varepsilon_1 n \} = [\eta_{ts,0} - E(\eta_{ts,0})] \mathbb{1} \{ a_{t-1}^+ \leq \varepsilon_1 n \} \leq 2\varepsilon_1 n \text{ for all } t \text{ and } s.
\]

From Theorem 2, Equation (2.3) of Merlevède et al. (2009), there is a constant \( C_3 \) depending only on \( \tilde{c} \) such that, for all \( n \geq 2 \),

\[
P \left( \left| \sum_{t=1}^n X_t \right| \geq \eta \right) \leq \exp \left( -\frac{C_3 \eta^2}{v^2 n + M^2 + \eta M (\log n)^2} \right).
\]

with \( v^2 = \sup_{i>0} \left( \text{Var}(X_i) + 2 \sum_{j>i} |\text{Cov}(X_i, X_j)| \right) = \text{Var}(X_0) + 2 \sum_{j=1}^\infty |\text{Cov}(X_j, X_0)| \) (using covariance stationarity).

In our case,

\[
P \left( \left| \sum_{t=1}^n \frac{\eta_{ts}}{n^{1/2}} \mathbb{1} \{ a_{t-1}^+ \leq \varepsilon_1 n \} \right| \geq \varepsilon n^{1/2} \right),
\]

so \( X_t = \frac{\eta_{ts}}{n^{1/2}} \mathbb{1} \{ a_{t-1}^+ \leq \varepsilon_1 n \}, \quad \eta = \varepsilon n^{1/2}, \) and \( M = 2\varepsilon_1 n, \) thereby providing

\[
P \left( \left| \sum_{t=1}^n \frac{\eta_{ts}}{n^{1/2}} \mathbb{1} \{ a_{t-1}^+ \leq \varepsilon_1 n \} \right| \geq \varepsilon n^{1/2} \right) \leq \exp \left( -\frac{C_3 \varepsilon^2 n}{v^2 n + 4\varepsilon_1^2 n + 2\varepsilon_1 \varepsilon n^{1/2} (\log n)^2} \right).
\]

Thus, it suffices to show \( v^2 n \) is slower than \( n \).
Note that

\[ Var(\eta_t 1 \{ a_{t-1}^+ \leq e_{1n} \}) \]
\[ = E[\eta_t^2 1 \{ a_{t-1}^+ \leq e_{1n} \}] - E[\eta_t 1 \{ a_{t-1}^+ \leq e_{1n} \}]^2 \]
\[ \leq E[\eta_t^2 1 \{ a_{t-1}^+ \leq e_{1n} \}] \]
\[ = E[n[\eta_{ts}, 0] - E(\eta_{ts}, 0)]^2 1 \{ a_{t-1}^+ \leq e_{1n} \}] \]
\[ \leq nE[|\eta_{ts}, 0|^2] = nE\left[\left|\left(\psi_\tau(u_{tr} - n^{-1/2}\delta_s x_{t-1}) - \psi_\tau(u_{tr})\right)(a_{t-1}^+)\right|^2\right] \]
\[ = nE\left[\left|\left(F_u\left(n^{-1/2}\delta_s x_{t-1} \mid x_{t-1}\right) - F_u(0 \mid x_{t-1})\right)(a_{t-1}^+)\right|^2\right] \]
\[ = nE\left[\left(n^{-1/2}\delta_s x_{t-1} f_u(0 \mid x_{t-1}) + o_p(n^{-1/2})\right)(a_{t-1}^+)\right]^2 = C n^{1/2} p^{1/2} E[(a_{t-1}^+)^2], \]

where we used the fact \( \|\delta_s x_{t-1}\| \leq O_p(p^{1/2}) \) (see (31) below). Therefore, we have \( Var(X_t) = Var\left(\frac{\eta_{ts}}{n} 1 \{ a_{t-1}^+ \leq e_{1n} \}\right) = \frac{1}{n} Var\left(\eta_{ts} 1 \{ a_{t-1}^+ \leq e_{1n} \}\right) \leq C n^{-3/2} \xi_B p^{3/2}. \)

\[ Var\left(\frac{\eta_{ts}}{n^{1/2}} 1 \{ a_{t-1}^+ \leq e_{1n} \}\right) \leq C n^{-1/2} \xi_B p^{3/2}. \quad (C.30) \]

It remains to show the covariance terms \( \sum_{j=1}^\infty |Cov(X_j, X_0)| \) is same or smaller order than the term in (C.30). By the inequality of the strong mixing processes, see e.g., Corollary 14.3 of Davidson (1994), for \( \frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1, \)

\[ |Cov(X_j, X_0)| \preceq \alpha(j)^{1/r} \|X_j\|_p \|X_0\|_r, \]

where \( \preceq \) is \( \leq \) up to a fixed constant term. Letting \( p = q = 2 + \zeta, \) we have \( r = \frac{2+\zeta}{\zeta}. \) Note that

\[ \|X_j\|_{2+\zeta} = \left(E\left[|X_j|^{2+\zeta}\right]\right)^{\frac{1}{2+\zeta}} \]

and

\[ E\left[|X_j|^{2+\zeta}\right] \]
\[ = E\left[\frac{\eta_{js}}{n^{1/2}} 1 \{ a_{j-1}^+ \leq e_{1n} \}\right]^{2+\zeta} \]
\[ = E\left[\left|\left(\psi_\tau(u_{tr} - n^{-1/2}\delta_s x_{t-1}) - \psi_\tau(u_{tr})\right)(a_{t-1}^+)\right|^2\right] \]
\[ \times 1 \{ a_{j-1}^+ \leq e_{1n} \} \]
\[ \leq E\left[\left|\psi_\tau(u_{tr} - n^{-1/2}\delta_s x_{t-1}) - \psi_\tau(u_{tr})\right|(a_{t-1}^+)\right]^{2+\zeta} \]
\[ = E\left[\left|\psi_\tau(u_{tr} - n^{-1/2}\delta_s x_{t-1}) - \psi_\tau(u_{tr})\right||x_{t-1}\right](a_{t-1}^+)\right]^{2+\zeta} \]

49
From the same argument to derive (C.30), we have
\[
E \left[ \psi_{\tau}(u_{t\tau} - n^{-1/2} \delta'_s x_{t-1}) - \psi_{\tau}(u_{t\tau}) \right] = n^{-1/2} \delta'_s x_{t-1} + o_p \left( n^{-1/2} \right).
\]
Therefore, we get
\[
E \left[ |X_j|^{2+\zeta} \right] \leq n^{-1/2} p^{1/2} E \left[ (a_{t-1}^+) \right]^{2+\zeta}
\]
and
\[
\|X_j\|_{2+\zeta} \leq n^{-\frac{1}{2(2+\zeta)}} p^{\frac{1}{2(2+\zeta)}} \|a_{t-1}^+\|_{2+\zeta}.
\]
The stationarity assumption implies that \(\|X_0\|_{2+\zeta}\) is the same order. Therefore,
\[
|Cov (X_j, X_0)| \leq \alpha(j)^{1/r} n^{-\frac{1}{2(2+\zeta)}} p^{\frac{1}{2(2+\zeta)}} \|a_{t-1}^+\|_{2+\zeta}^2
\]
and
\[
\sum_{j=1}^{\infty} |Cov (X_j, X_0)| \leq C n^{-\frac{1}{2(2+\zeta)}} \xi^{-1} p^{1+\frac{1}{2(2+\zeta)}}.
\]
In other words, the leading term of \(v^2 = \sup_{i>0} \left( Var(X_i) + 2 \sum_{j>i} |Cov (X_i, X_j)| \right)\) is \(Var \left( \frac{n_{ts}}{n^{1/2}} 1 \{ a_{t-1}^+ \leq \varepsilon n \} \right)\), which is \(C n^{-1/2} \xi^{-1} p^{3/2}\). Therefore, we conclude that
\[
P \left( \left| \sum_{t=1}^{n} \frac{n_{ts}}{n^{1/2}} 1 \{ a_{t-1}^+ \leq \varepsilon n \} \right| \geq \varepsilon n^{1/2} \right) \leq \exp \left( -\frac{C_3 \varepsilon^2 n}{C n^{-1/2} \xi^{-1} p^{3/2} + 4 \varepsilon^2 n + 2 \varepsilon n \varepsilon n^{1/2} (\log n)^2} \right) \to 0
\]
using the same argument from Lu and Su (2015, p. 53). Note that, by the weak dependence assumption (asymptotic independence), we achieve the same order of magnitudes (in terms of \(n\)) in the exponent of the Bernstein’s inequality.

The proofs of the other terms \(D_2, I_2\) and \(I_3\) do not rely on the independence assumption, such that the same argument of Lu and Su (2015) can carry over under our additional assumptions. Thus the proofs are omitted.

For (CA.7), we need to show:
\[
\sup_{||\delta|| \leq \sqrt{pL}} ||E[V(\delta)] - E[V(0)] + A\delta||_c = o(1).
\]
By Assumption 4.1 and 4.6,
\[
\sup_{||\delta|| \leq \sqrt{pL}} ||E[V(\delta)] - E[V(0)] + A\delta||_c
\]
\[
= \sup_{||\delta|| \leq \sqrt{pL}} \left| \sum_{t=1}^{n} \frac{n_{ts}}{n^{1/2}} E \left[ \psi_{\tau}(u_t - n^{-1/2} \delta'_s x_{t-1}) \right] x_{t-1} - n^{-1/2} \sum_{t=1}^{n} E \left[ \psi_{\tau}(u_t) x_{t-1} \right] + A\delta \right|_c
\]
\[
= \sup_{||\delta|| \leq \sqrt{pL}} \left| \sum_{t=1}^{n} \left[ \left( \tau - 1 \{ u_t - n^{-1/2} \delta'_s x_{t-1} \leq 0 \} \right) x_{t-1} - \left( \tau - 1 \{ u_t \leq 0 \} \right) x_{t-1} \right] + A\delta \right|_c
\]
\[
= 50
\]
\[
\sup_{\|\delta\| \leq \sqrt{pL}} \|n^{-1/2} \sum_{t=1}^{n} E \left[ \left( -1(u_t - n^{-1/2} \delta' x_{t-1}) + 1(u_t \leq 0) \right) x_{t-1} \right] + A\delta \|_c \\
= \sup_{\|\delta\| \leq \sqrt{pL}} \|n^{-1/2} \sum_{t=1}^{n} E \left[ F(-u_t - n^{-1/2} \delta' x_{t-1} | x_{t-1}) - F(u_t | x_{t-1}) x_{t-1} - F(u_t | x_{t-1}) x_{t-1} \right] - A\delta \|_c \\
= \sup_{\|\delta\| \leq \sqrt{pL}} \|n^{-1/2} \sum_{t=1}^{n} E \left[ n^{-1/2} \delta' x_{t-1} \left( \int_{0}^{1} f(-u_t + s n^{-1/2} \delta' x_{t-1} | x_{t-1}) ds \right) x_{t-1} \right] - E \left[ f(-u_t | x_{t-1}) x_{t-1} x'_{t-1} \right] \delta \|_c \\
= \sup_{\|\delta\| \leq \sqrt{pL}} \|n^{-1} \sum_{t=1}^{n} E \left[ \left( \int_{0}^{1} f(-u_t + s n^{-1/2} \delta' x_{t-1} | x_{t-1}) - f(-u_t | x_{t-1}) \right) ds \right] x_{t-1} x'_{t-1} \delta \|_c \quad \text{(stationary } x_{t-1}) \\
\leq C \sup_{\|\delta\| \leq \sqrt{pL}} n^{-3/2} \sum_{t=1}^{n} E\|\delta' x_{t-1} x_{t-1} x'_{t-1} \delta\|_c \quad \text{(Taylor expansion)} \\
\leq C n^{-3/2} \|c\| \sup_{\|\delta\| \leq \sqrt{pL}} \sum_{t=1}^{n} \left( E\|\delta' x_{t-1}\|_F^2 \right)^{1/2} \left( E\|x_{t-1} x'_{t-1} \delta\|_F^2 \right)^{1/2} \quad \text{(by Cauchy-Schwarz inequality)} \\
\leq C n^{-3/2} \frac{1}{\sqrt{Lc}} \sup_{\|\delta\| \leq \sqrt{pL}} \sum_{t=1}^{n} \left( E\|\delta' x_{t-1}\|_F^2 \right)^{1/2} \left( E\|x_{t-1} x'_{t-1} \delta\|_F^2 \right)^{1/2} \quad \text{(since } \|c\| \leq \frac{1}{\sqrt{Lc}} \text{)} \\
= C n^{-3/2} \frac{1}{\sqrt{Lc}} \sum_{t=1}^{n} O(\tilde{c}_{A}^{1/2} p^{1/2}) O(p^{3/2}) \\
= O(n^{-1/2} \frac{1}{\sqrt{Lc}} \tilde{c}_{A}^{1/2} p^{1/2}) = o_p(1),
\]

where the second last equality uses the facts that

\[
E\|\delta' x_{t-1}\|_F^2 \leq E\|x_{t-1}\|^2 \|\delta\|_F^2 \\
= \lambda_{\max} \left( E[x_{t-1} x'_{t-1}] \right) \|\delta\|_F^2 \\
\leq O(\tilde{c}_{A} p) \quad \text{(Assumption 4.6)} \quad \text{(C.31)}
\]

and

\[
E\|x_{t-1} x'_{t-1} \delta\|_F^2 \leq E\|x_{t-1} x'_{t-1}\|_F^2 \|\delta\|_F^2 \\
= E \left[ \left( \sum_{j=1}^{p} x_{t-1,j}^2 \right)^2 \|\delta\|_F^2 \right] \\
\leq O(p^2) O(p) = O(p^3).
\]

Next, we show (CA.8):

\[
\|V(\hat{\delta})\|_c = o_p(1).
\]
Note that

$$||V(\hat{\delta})||_c = ||n^{-1/2} \sum_{t=1}^{n} \psi_r(u_t - n^{-1/2} \delta' x_{t-1}) x_{t-1}||_c$$

$$= ||n^{-1/2} \sum_{t=1}^{n} \psi_r(y_t - \beta' x_{t-1}) x_{t-1}||_c$$

$$= \left| \sum_{t=1}^{n} \psi_r(y_t - \beta' x_{t-1}) x_{t-1} \right|$$

$$\leq n^{-1/2} \sum_{t=1}^{n} |x_{t-1}| 1(y_t - \beta' x_{t-1} = 0) \text{ (by the proof of Lemma A2 in Ruppert and Carroll (1980))}$$

$$\leq n^{-1/2} \max_{1 \leq t \leq n} |x_{t-1}| p$$

$$= o_p(1),$$

where we use the fact that

$$P\left( \max_{1 \leq t \leq n} |x_{t-1}| \geq n^{1/2} p^{-1} \right) \leq nP\left( |x_{t-1}| \geq n^{1/2} p^{-1} \right) \text{ (by Boole’s inequality)}$$

$$\leq \frac{n E|x_{t-1}|^8}{n^4 p^{-8}} \text{ (by Markov’s inequality)}$$

$$\leq n^{-3} p^8 ||c||^8 E\|x_{t-1}\|^8$$

$$= O(n^{-3} p^8 \xi_B^{-4} p^4) \text{ (since sup}_{j \geq 1} E(x_{t-1,j}^8) \leq c_x \text{ for some } c_x \leq \infty)$$

$$= O(n^{-3} p^{12} \xi_B^{-4})$$

$$= o(1),$$

by the assumption that $n^{-3} p^{12} \xi_B^{-4} \to 0$ as $n \to \infty$.

From (C.25), the limit theory of $A\hat{\delta}$ follows $V(0)$ because the penalty term $n^{-1/2} \omega_n(\delta)$ for the non-zero $q_n$ element is asymptotically negligible under our rate conditions. Note that $V(0) := n^{-1/2} \sum_{t=1}^{n} \psi_r(u_t) x_t$, then under our strong mixing condition of Assumption 5.1, Proposition B.2 of the supplement of Li and Liao (2020) provides:

$$\|V(0) - \tilde{S}_n\|_F = o_p(1),$$

where $\tilde{S}_n$ is a $q_n$-dimensional random vector with distribution $N(0, B)$. Therefore, $\hat{\delta} := \sqrt{n}(\beta - \beta_0)$ converges to a $q_n$-dimensional random vector with distribution $N(0, A^{-1}BA^{-1})$ in $||\cdot||_F$-norm, implying Theorem 5.1-(i), from the relation $\|M\| \leq ||M||_F \leq \sqrt{r} \|M\|$ for a matrix $M$ of rank $r$.

For the proof of Theorem 5.1-(ii), following the argument in Section 4.2, we have

$$R \left[ \beta_r^{(1)QR^*} - \beta_{0r}^{(1)}(1) \right] = R \left[ \beta_r^{(1)QR} - \beta_{0r}^{(1)} \right] - A_1 \left( \tilde{\beta}_r^{(1)QR} - \beta_{0r}^{(1)} \right),$$

52
so that
\[ \sqrt{n}R \left[ \hat{\beta}_{\tau}^{(1),QR*} - \beta_{0\tau}^{(1)} \right] = -\sqrt{n}RA_1' \left( \hat{\beta}_{\tau}^{(1),QR} - \beta_{0\tau}^{(1)} \right) + \frac{1}{\sqrt{n}} nR \left( \hat{\beta}_{\tau}^{(1),QR} - \beta_{0\tau}^{(1)} \right) \]
\[ = -\sqrt{n}RA_1' \left( \hat{\beta}_{\tau}^{(1),QR} - \beta_{0\tau}^{(1)} \right) + O_p \left( \frac{1}{\sqrt{n}} \right) \]
\[ \implies N \left( 0, RA_1'A^{-1}BA^{-1}A_1R' \right). \]

Finally, for the fixed dimension of \( p_x < \infty \), Theorem 5.1-(iii) follows from Lee (2016). Unlike the cointegrated parts of the system, this non-cointegrated local unit root regressors have their own limit theory from the block-wise diagonal structure of \( H \) and \( H^{-1} \) matrix transformations given in Section 4.2.
Table D.1: Prediction Results of Stock Returns: 12 one-period-ahead forecasts, GIC

|          | Quantile (τ) |         |         |         |         |
|----------|--------------|---------|---------|---------|---------|
|          | 0.05 | 0.1   | 0.5   | 0.9   | 0.95   |
| Final Prediction Error (FPE) |         |         |         |         |         |
| QR       | 0.0047 | 0.0099 | 0.0134 | 0.0078 | 0.0060 |
| LASSO    | **0.0045** | **0.0083** | 0.0123 | **0.0032** | **0.0020** |
| ALQR     | **0.0045** | **0.0083** | **0.0122** | **0.0032** | **0.0020** |
| QUANT    | 0.0046 | **0.0083** | 0.0124 | 0.0055 | 0.0029 |

|          |         |         |         |         |         |
| Out-of-Sample $R^2$ |         |         |         |         |         |
| QR       | -0.0147 | -0.1930 | -0.0831 | -0.4048 | -1.0666 |
| LASSO    | 0.0219 | **-0.0009** | 0.0094 | 0.4219 | 0.3233 |
| ALQR     | **0.0223** | -0.0029 | **0.0167** | **0.4288** | 0.3211 |

|          |         |         |         |         |
| Average # of Selected Predictors |         |         |         |         |
| LASSO    | 11.00 | 10.00 | 12.00 | 11.42 | 10.58 |
| ALQR     | 10.00 | 7.92  | 10.00 | 6.33  | 9.25  |

|          |         |         |         |         |
| Tuning Parameter ($\lambda$) by GIC |         |         |         |         |
| LASSO ($\times 10^{-4}$) | 82.07 | 225.20 | 0.04  | 49.17 | 27.66 |
| ALQR ($\times 10^{-7}$)  | 5.26  | 85.12  | 1.46  | 197.67 | 0.88  |
Figure D.1: The Correlation Heat Map of the Persistent Predictors

Note: The correlation heat map is based on 816 monthly observations ranging from January 1952 to December 2019. All correlations are computed using the first difference of predictors. The full predictor names are defined in Table 1.
Table D.2: Prediction Results of Stock Returns: 24 one-period-ahead forecasts, GIC

| Quantile (τ) | 0.05 | 0.1  | 0.5  | 0.9  | 0.95 |
|--------------|------|------|------|------|------|
| Final Prediction Error (FPE) |
| QR           | **0.0045** | 0.0102 | 0.0147 | 0.0067 | 0.0050 |
| LASSO        | 0.0058   | 0.0101 | **0.0140** | **0.0045** | **0.0025** |
| ALQR         | 0.0058   | **0.0097** | **0.0140** | **0.0045** | **0.0025** |
| QUANT        | 0.0055   | 0.0098 | 0.0148 | 0.0063 | 0.0036 |

| Out-of-Sample $R^2$ |
|---------------------|
| QR                  | **0.1751** | -0.0388 | 0.0100 | -0.0745 | -0.3938 |
| LASSO               | -0.0590   | -0.0245 | 0.0573 | 0.2747  | 0.2881  |
| ALQR                | -0.0526   | **0.0103** | **0.0577** | **0.2798** | **0.3088** |

| Average # of Selected Predictors |
|----------------------------------|
| LASSO                            | 10.88   | 9.63  | 12.00 | 10.67  | 10.71  |
| ALQR                             | 9.96    | 8.88  | 10.63 | 10.25  | 9.67   |

| Tuning Parameter (λ) by GIC |
|-----------------------------|
| LASSO ($\times 10^{-4}$)    | 66.33   | 174.85 | 0.30  | 197.25 | 14.76  |
| ALQR ($\times 10^{-7}$)     | 3.74    | 24.70  | 0.65  | 1.55   | 0.59   |
Table D.3: Simulation Results: Scenario 1 with Dependent Predictors

| Quantile ($\tau$) | 0.05 | 0.1  | 0.5  | 0.9  | 0.95 |
|-------------------|------|------|------|------|------|
| Final Prediction Error (FPE) | | | | | |
| QR                | 0.1076 | 0.1809 | 0.4072 | 0.1812 | 0.1081 |
| LASSO             | 0.1073 | 0.1811 | 0.4084 | 0.1809 | 0.1077 |
| ALQR              | **0.1052** | **0.1785** | **0.4048** | **0.1794** | **0.1060** |
| RIDGE             | 0.2288 | 0.3050 | 0.4053 | 0.3031 | 0.2259 |
| QUANT             | 2.2753 | 4.2477 | 11.4435 | 4.3347 | 2.3361 |
| Out-of-Sample $R^2$ | | | | | |
| QR                | 0.9527 | 0.9574 | 0.9644 | 0.9582 | 0.9537 |
| LASSO             | 0.9528 | 0.9574 | 0.9643 | 0.9583 | 0.9539 |
| ALQR              | **0.9537** | **0.9580** | **0.9646** | **0.9586** | **0.9546** |
| RIDGE             | 0.8994 | 0.9282 | 0.9646 | 0.9301 | 0.9033 |
| Average # of Selected Predictors | | | | | |
| LASSO             | 10.42  | 10.15  | 9.67  | 10.14 | 10.47 |
| ALQR              | 6.56   | 6.35   | 6.38  | 6.35  | 6.50  |
| Tuning Parameter ($\lambda$) | | | | | |
| LASSO ($\times 10^0$) | 8.57 | 15.18 | 37.39 | 15.33 | 8.43 |
| ALQR ($\times 10^{-4}$) | 1.68 | 2.39 | 3.81 | 2.43 | 1.73 |
| RIDGE ($\times 10^{-5}$) | 0.8  | 1.7  | 1.6  | 1.7  | 1.0  |

Dependency Rate: $\rho = 0.1$
Table D.4: Simulation Results: Scenario 1 with Dependent Predictors

| Quantile (τ) | 0.05 | 0.1  | 0.5  | 0.9  | 0.95 |
|--------------|------|------|------|------|------|
| Dependency Rate: $\rho = 0.5$ |
| Final Prediction Error (FPE) |
| QR           | 0.1064 | 0.1800 | 0.4074 | 0.1805 | 0.1074 |
| LASSO        | 0.1069 | 0.1808 | 0.4094 | 0.1810 | 0.1078 |
| ALQR         | **0.1053** | **0.1785** | **0.4055** | **0.1790** | **0.1060** |
| RIDGE        | 0.2234 | 0.3046 | 0.4097 | 0.3142 | 0.2296 |
| QUANT        | 2.5896 | 4.6670 | 11.7537 | 4.5306 | 2.4595 |
| Out-of-Sample $R^2$ |
| QR           | 0.9589 | 0.9614 | 0.9653 | 0.9601 | 0.9563 |
| LASSO        | 0.9587 | 0.9613 | 0.9652 | 0.9600 | 0.9562 |
| ALQR         | **0.9593** | **0.9617** | **0.9655** | **0.9605** | **0.9569** |
| RIDGE        | 0.9137 | 0.9347 | **0.9655** | 0.9306 | 0.9067 |
| Average # of Selected Predictors |
| LASSO        | 9.90  | 9.67  | 9.34  | 9.78  | 10.00 |
| ALQR         | 6.46  | 6.32  | 6.33  | 6.29  | 6.46  |
| Tuning Parameter ($\lambda$) |
| LASSO ($\times 10^0$) | 9.82 | 17.20 | 37.28 | 16.12 | 9.37 |
| ALQR ($\times 10^{-4}$) | 1.50 | 2.08 | 3.25 | 2.09 | 1.53 |
| RIDGE ($\times 10^{-5}$) | 0.9  | 1.8  | 2.2  | 1.6  | 0.8  |
Table D.5: Simulation Results: Scenario 1 with Dependent Predictors

| Dependency Rate: $\rho = 0.9$ | Quantile ($\tau$) | 0.05 | 0.1 | 0.5 | 0.9 | 0.95 |
|-------------------------------|-------------------|------|-----|-----|-----|-----|
| Final Prediction Error (FPE)  | QR                | 0.1067 | 0.1804 | 0.4077 | 0.1809 | 0.1083 |
|                               | LASSO             | 0.1060 | 0.1800 | 0.4095 | 0.1810 | 0.1075 |
|                               | ALQR              | 0.1052 | 0.1783 | 0.4055 | 0.1787 | 0.1065 |
|                               | RIDGE             | 0.2157 | 0.2955 | 0.4058 | 0.2990 | 0.2193 |
|                               | QUANT             | 2.2630 | 4.1512 | 11.0832 | 4.3207 | 2.4073 |
| Out-of-Sample $R^2$           | QR                | 0.9528 | 0.9565 | 0.9632 | 0.9581 | 0.9550 |
|                               | LASSO             | 0.9532 | 0.9566 | 0.9630 | 0.9581 | 0.9553 |
|                               | ALQR              | 0.9535 | 0.9570 | 0.9634 | 0.9586 | 0.9558 |
|                               | RIDGE             | 0.9047 | 0.9288 | 0.9634 | 0.9308 | 0.9089 |
| Average # of Selected Predictors | LASSO       | 8.33 | 7.74 | 7.17 | 7.71 | 8.26 |
|                               | ALQR              | 6.17  | 6.01 | 6.07 | 5.97 | 6.13 |
| Tuning Parameter ($\lambda$)  | LASSO ($\times 10^0$) | 13.09 | 23.81 | 52.67 | 23.93 | 13.39 |
|                               | ALQR ($\times 10^{-4}$) | 1.36  | 1.67  | 2.11  | 1.57  | 1.38  |
|                               | RIDGE ($\times 10^{-5}$) | 0.6   | 1.3   | 1.5   | 1.2   | 0.9   |
Table D.6: Simulation Results: Scenario 2 with Dependent Predictors

| Quantile (τ) | 0.05  | 0.1  | 0.5  | 0.9  | 0.95 |
|--------------|-------|------|------|------|------|

Dependency Rate: $\rho = 0.1$

| Final Prediction Error (FPE) | QR      | 0.1075 | 0.1809 | 0.4074 | 0.1811 | 0.1081 |
|------------------------------|---------|--------|--------|--------|--------|--------|
| LASSO                        | 0.1074  | 0.1810 | 0.4087 | 0.1807 | 0.1080 |
| ALQR                         | **0.1067** | **0.1794** | **0.4068** | **0.1799** | **0.1068** |
| RIDGE                        | 0.2363  | 0.3123 | **0.4057** | 0.3123 | 0.2349 |
| QUANT                        | 2.3787  | 4.4253 | 11.8534 | 4.4989 | 2.4148 |

| Out-of-Sample $R^2$          | QR      | 0.9548  | 0.9591 | 0.9656 | 0.9597 | 0.9552 |
|------------------------------|---------|--------|--------|--------|--------|--------|
| LASSO                        | 0.9549  | 0.9591 | 0.9655 | 0.9598 | 0.9553 |
| ALQR                         | **0.9552** | **0.9595** | **0.9657** | **0.9600** | **0.9558** |
| RIDGE                        | 0.9007  | 0.9294 | **0.9658** | 0.9306 | 0.9027 |

| Average # of Selected Predictors | LASSO | 10.82 | 10.67 | 10.35 | 10.66 | 10.86 |
|----------------------------------|-------|------|------|------|------|------|
| ALQR                             | 8.44  | 8.33 | 8.31 | 8.34 | 8.45 |

| Tuning Parameter (λ)             | LASSO ($\times 10^0$) | 6.95 | 12.49 | 30.84 | 12.25 | 6.89 |
|----------------------------------|-----------------------|------|-------|-------|------|------|
| ALQR ($\times 10^{-4}$)         | 1.23                  | 1.61 | 2.52  | 1.62  | 1.18 |
| RIDGE ($\times 10^{-5}$)        | 0.7                   | 1.7  | 1.6   | 1.4   | 0.6  |
Table D.7: Simulation Results: Scenario 2 with Dependent Predictors

| Quantile ($\tau$) | 0.05 | 0.1  | 0.5  | 0.9  | 0.95 |
|-------------------|------|------|------|------|------|

Dependency Rate: $\rho = 0.5$

|                      | Final Prediction Error (FPE) | QR  | LASSO | ALQR | RIDGE | QUANT |
|----------------------|-----------------------------|-----|-------|------|-------|-------|
|                      |                             | 0.1072 | 0.1805 | 0.4081 | 0.1813 | 0.1082 |
|                      |                             | 0.1073 | 0.1811 | 0.4099 | 0.1815 | 0.1085 |
|                      |                             | **0.1068** | **0.1799** | 0.4070 | **0.1805** | **0.1072** |
|                      |                             | 0.2484 | 0.3313 | **0.4063** | 0.3406 | 0.2554 |
|                      |                             | 2.7837 | 5.0151 | 12.9604 | 4.9791 | 2.6430 |

|                      | Out-of-Sample $R^2$ | QR  | LASSO | ALQR | RIDGE | QUANT |
|----------------------|--------------------|-----|-------|------|-------|-------|
|                      |                    | 0.9615 | 0.9640 | 0.9685 | 0.9636 | 0.9591 |
|                      |                    | 0.9614 | 0.9639 | 0.9684 | 0.9635 | 0.9589 |
|                      |                    | **0.9616** | **0.9641** | **0.9686** | **0.9637** | **0.9594** |
|                      |                    | 0.9108 | 0.9339 | **0.9686** | 0.9316 | 0.9034 |

|                      | Average # of Selected Predictors | LASSO | ALQR | QUANT |
|----------------------|----------------------------------|-------|------|-------|
|                      |                                 | 10.58 | 10.45 | 10.27 | 10.54 | 10.66 |
|                      |                                 | 8.33  | 8.32  | 8.32  | 8.27  | 8.39  |

|                      | Tuning Parameter ($\lambda$) | LASSO ($\times 10^0$) | ALQR ($\times 10^{-4}$) | RIDGE ($\times 10^{-5}$) |
|----------------------|-----------------------------|-----------------------|-------------------------|-------------------------|
|                      |                             | 7.23 | 12.75 | 28.74 | 12.08 | 6.94 |
|                      |                             | 1.19 | 1.41  | 2.07  | 1.44  | 1.05 |
|                      |                             | 0.90 | 1.50  | 1.70  | 1.40  | 0.80 |
Table D.8: Simulation Results: Scenario 2 with Dependent Predictors

| Quantile (τ) | 0.05 | 0.1  | 0.5  | 0.9  | 0.95 |
|-------------|------|------|------|------|------|
| Dependency Rate: ρ = 0.9 |
| Final Prediction Error (FPE) |
| QR          | 0.1065 | 0.1803 | 0.4074 | 0.1808 | 0.1082 |
| LASSO       | 0.1065 | 0.1804 | 0.4103 | 0.1817 | 0.1078 |
| ALQR        | 0.1063 | **0.1797** | 0.4065 | **0.1798** | **0.1077** |
| RIDGE       | 0.2571 | 0.3418 | **0.4058** | 0.3469 | 0.2593 |
| QUANT       | 2.7380 | 5.0229 | 13.4626 | 5.2381 | 2.8857 |
| Out-of-Sample $R^2$ |
| QR          | 0.9611 | 0.9641 | 0.9697 | 0.9655 | 0.9625 |
| LASSO       | 0.9611 | 0.9641 | 0.9695 | 0.9653 | **0.9627** |
| ALQR        | **0.9612** | **0.9642** | 0.9698 | **0.9657** | **0.9627** |
| RIDGE       | 0.9061 | 0.9320 | **0.9699** | 0.9338 | 0.9101 |
| Average # of Selected Predictors |
| LASSO       | 9.55  | 9.14  | 8.62  | 9.13  | 9.50  |
| ALQR        | 8.04  | 7.98  | 8.05  | 7.94  | 8.05  |
| Tuning Parameter (λ) |
| LASSO ($\times 10^0$) | 9.47   | 17.32 | 40.78 | 17.24 | 9.63  |
| ALQR ($\times 10^{-4}$) | 1.06   | 1.32  | 1.52  | 1.21  | 0.98  |
| RIDGE ($\times 10^{-5}$) | 0.6    | 1.3   | 1.4   | 1.2   | 0.6   |
Table D.9: Simulation Results: Scenario 3 with Dependent Predictors

| Quantile ($\tau$) | 0.05 | 0.1  | 0.5  | 0.9  | 0.95 |
|-------------------|------|------|------|------|------|
| **Dependency Rate: $\rho = 0.1$** |      |      |      |      |      |
| Final Prediction Error (FPE) |      |      |      |      |      |
| QR                | 0.1076 | 0.1808 | 0.4074 | 0.1811 | 0.1082 |
| LASSO             | 0.1076 | 0.1808 | 0.4074 | 0.1810 | 0.1081 |
| ALQR              | 0.1079 | 0.1812 | 0.4074 | 0.1814 | 0.1086 |
| RIDGE             | 0.2394 | 0.3157 | 0.4074 | 0.1810 | 0.1081 |
| QUANT             | 2.3833 | 4.4624 | 11.9144 | 4.5103 | 2.4318 |
| Out-of-Sample $R^2$ |      |      |      |      |      |
| QR                | 0.9549 | 0.9595 | 0.9658 | 0.9599 | 0.9555 |
| LASSO             | 0.9549 | 0.9595 | 0.9658 | 0.9599 | 0.9555 |
| ALQR              | 0.9547 | 0.9594 | 0.9658 | 0.9598 | 0.9553 |
| RIDGE             | 0.8996 | 0.9292 | 0.9660 | 0.9308 | 0.9035 |
| Average # of Selected Predictors |      |      |      |      |      |
| LASSO             | 12.00 | 12.00 | 12.00 | 12.00 | 12.00 |
| ALQR              | 11.83 | 11.92 | 12.00 | 11.93 | 11.82 |
| Tuning Parameter ($\lambda$) |      |      |      |      |      |
| LASSO (×10$^{-1}$) | 1.67 | 1.82 | 1.92 | 1.80 | 1.73 |
| ALQR (×10$^{-5}$)  | 1.1  | 0.9  | 0.2  | 0.8  | 1.0  |
| RIDGE (×10$^{-5}$) | 0.6  | 1.1  | 1.0  | 1.1  | 0.6  |
Table D.10: Simulation Results: Scenario 3 with Dependent Predictors

| Quantile (τ) | 0.05 | 0.1  | 0.5  | 0.9  | 0.95 |
|-------------|------|------|------|------|------|
| Dependency Rate: ρ = 0.5 |

| Final Prediction Error (FPE) | QR   | LASSO    | ALQR | RIDGE | QUANT |
|-------------------------------|------|----------|------|-------|-------|
| 0.05                          | 0.1065 | 0.1799  | 0.4076 | 0.1804 | 0.1074 |
| 0.1                           | 0.1798 | 0.1805  | 0.4077 | 0.1805 | 0.1073 |
| 0.5                           | 0.4059 | 0.3487  | 0.2597 |       |       |
| 0.9                           | 13.4026 | 5.2607  | 2.8399 |       |       |
| 0.95                          |       |         |       | 2.8938 | 5.2864 |

| Out-of-Sample $R^2$ | QR   | LASSO    | ALQR | RIDGE | QUANT |
|----------------------|------|----------|------|-------|-------|
| 0.05                 | 0.9632 | 0.9660  | 0.9696 | 0.9657 | 0.9622 |
| 0.1                  | 0.9632 | 0.9660  | 0.9696 | 0.9657 | 0.9622 |
| 0.5                  | 0.9630 | 0.9659  | 0.9696 | 0.9656 | 0.9621 |
| 0.9                  | 0.9131 | 0.9360  | 0.9697 | 0.9337 | 0.9086 |
| 0.95                 |       |         |       | 0.9337 | 0.9086 |

| Average # of Selected Predictors | LASSO | ALQR |
|----------------------------------|------|------|
| 0.05                             | 12.00 | 11.78 |
| 0.1                              | 12.00 | 11.90 |
| 0.5                              | 12.00 | 12.00 |
| 0.9                              | 12.00 | 11.91 |
| 0.95                             | 12.00 | 11.76 |

| Tuning Parameter (λ) | LASSO ($×10^{-1}$) | ALQR ($×10^{-5}$) | RIDGE ($×10^{-5}$) |
|----------------------|-------------------|-------------------|-------------------|
| 0.05                 | 1.47              | 1.2               | 0.6               |
| 0.1                  | 1.50              | 1.0               | 1.0               |
| 0.5                  | 1.68              | 0.2               | 1.2               |
| 0.9                  | 1.60              | 0.0               | 1.1               |
| 0.95                 | 1.57              | 1.0               | 0.6               |
Table D.11: Simulation Results: Scenario 3 with Dependent Predictors

| Quantile (τ) | 0.05 | 0.1  | 0.5  | 0.9  | 0.95 |
|-------------|------|------|------|------|------|
| Final Prediction Error (FPE) |
| QR          | 0.1069 | 0.1803 | 0.4079 | 0.1810 | 0.1083 |
| LASSO       | 0.1069 | 0.1803 | 0.4079 | 0.1810 | 0.1083 |
| ALQR        | 0.1075 | 0.1818 | 0.4084 | 0.1823 | 0.1089 |
| RIDGE       | 0.2638 | 0.3498 | 0.4064 | 0.3538 | 0.2660 |
| QUANT       | 2.8993 | 5.1589 | 13.7720 | 5.3428 | 2.9630 |
| Out-of-Sample $R^2$ |
| QR          | 0.9631 | 0.9651 | 0.9704 | 0.9661 | 0.9634 |
| LASSO       | 0.9631 | 0.9651 | 0.9704 | 0.9661 | 0.9634 |
| ALQR        | 0.9629 | 0.9648 | 0.9703 | 0.9659 | 0.9632 |
| RIDGE       | 0.9090 | 0.9322 | 0.9705 | 0.9338 | 0.9102 |
| Average # of Selected Predictors |
| LASSO       | 11.95  | 11.99 | 12.00 | 11.98 | 11.96 |
| ALQR        | 11.01  | 11.28 | 11.74 | 11.31 | 11.07 |
| Tuning Parameter ($\lambda$) |
| LASSO ($\times 10^{-1}$) | 1.51 | 1.25 | 1.02 | 1.20 | 1.49 |
| ALQR ($\times 10^{-5}$) | 2.7  | 3.4  | 2.1  | 3.2  | 2.6  |
| RIDGE ($\times 10^{-5}$) | 0.6  | 1.0  | 1.0  | 1.1  | 0.5  |
### Table D.12: The Johansen Cointegration Test

| Hypothesised number of CE(s) | Trace statistic | 0.05 critical value |
|-----------------------------|-----------------|---------------------|
|                             | (1)  (2)  (3)  (4)  (5)  (6)  (7)  (8)  (9)  (10)  (11)  (12) |
| $r = 0$                     | 2.21  2.40  2.36  2.25  2.26  2.29  2.34  2.42  2.44  2.43  2.47  2.27 | 157.11 |
| $r \leq 1$                  | 2.10  10.66 11.18 10.84 10.52 10.41 10.66 10.91 11.90 10.39 10.04 | 124.25 |
| $r \leq 2$                  | 20.70 20.51 21.52 21.23 21.11 20.64 20.79 21.17 20.75 20.21 19.84 | 90.39 |
| $r \leq 3$                  | 41.47 41.45 41.99 41.25 40.85 40.48 41.01 41.81 41.28 40.87 40.36 | 70.6 |
| $r \leq 4$                  | 83.56 83.44 83.91 83.11 82.61 82.17 82.42 82.91 82.16 81.13 80.63 | 48.28 |
| $r \leq 5$                  | 143.67 142.99 143.71 142.78 142.42 142.09 142.95 142.65 141.95 141.40 141.37 | 31.52 |
| $r \leq 6$                  | 221.96 219.42 219.94 219.65 218.94 218.71 217.84 218.05 216.82 216.47 216.03 | 17.95 |
| $r \leq 7$                  | 355.85 353.34 353.44 352.96 352.51 351.99 352.06 351.59 356.59 356.55 357.69 352.80 | 8.18 |

| Hypothesised number of CE(s) | Maximum eigenvalue statistic | 0.05 critical value |
|------------------------------|------------------------------|---------------------|
|                             | (1)  (2)  (3)  (4)  (5)  (6)  (7)  (8)  (9)  (10)  (11)  (12) |
| $r \leq 0$                  | 2.21  2.40  2.36  2.25  2.26  2.29  2.34  2.42  2.44  2.43  2.47  2.27 | 51.07 |
| $r \leq 1$                  | 8.03  8.26  8.81  8.59  8.26  8.12  8.33  8.49  8.56  7.96  7.57  8.09 | 44.91 |
| $r \leq 2$                  | 10.45 9.84 10.34 10.39 10.59 10.23 10.12 10.26 9.75 9.82 9.80 10.21 | 39.43 |
| $r \leq 3$                  | 20.77 20.94 20.47 20.02 19.74 19.84 20.23 20.23 20.64 20.53 20.65 20.38 | 33.32 |
| $r \leq 4$                  | 42.06 41.99 41.92 41.85 41.76 41.68 41.41 41.10 40.89 40.27 40.27 40.43 | 27.14 |
| $r \leq 5$                  | 60.14 59.55 59.80 59.68 59.80 59.92 59.91 60.04 60.48 60.81 60.76 59.99 | 21.07 |
| $r \leq 6$                  | 78.29 76.43 76.23 76.87 76.78 76.85 76.38 74.89 75.40 74.87 75.07 74.66 | 14.9 |
| $r \leq 7$                  | 133.90 133.92 133.50 133.31 133.31 133.05 133.36 133.75 138.55 139.74 141.23 136.77 | 8.18 |

Notes: 1. $r$ indicates the number of cointegrating vectors. 2. From Column (1) to Column (12), each column states the value of test statistic using data based on the rolling fixed window scheme for 12 one-step-ahead forecasts, respectively.
References

Adamek, R., S. Smeekes, and I. Wilms (2020). Lasso inference for high-dimensional time series. *arXiv preprint arXiv:2007.10952*.

Andersen, T. G., N. Fusari, and V. Todorov (2020). The pricing of tail risk and the equity premium: Evidence from international option markets. *Journal of Business & Economic Statistics 38*(3), 662–678.

Basu, S. and G. Michailidis (2015). Regularized estimation in sparse high-dimensional time series models. *Annals of Statistics 43*(4), 1535–1567.

Belloni, A. and V. Chernozhukov (2011). $\ell_1$-penalized quantile regression in high-dimensional sparse models. *Annals of Statistics 39*(1), 82–130.

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

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics 18*(2), 373–399.

Cenesizoglu, T. and A. Timmermann (2008). Is the distribution of stock returns predictable? *Available at SSRN 1107185*.

Cenesizoglu, T. and A. Timmermann (2012). Do return prediction models add economic value? *Journal of Banking & Finance 36*(11), 2974–2987.

Chernozhukov, V., W. K. Härdle, C. Huang, and W. Wang (2021). Lasso-driven inference in time and space. *Annals of Statistics 49*(3), 1702–1735.

Davidson, J. (1994). *Stochastic limit theory: An introduction for econometricians*. OUP Oxford.

Fama, E. F. and K. R. French (1988). Dividend yields and expected stock returns. *Journal of Financial Economics 22*(1), 3–25.

Fan, J., Y. Fan, and E. Barut (2014). Adaptive robust variable selection. *Annals of Statistics 42*(1), 324.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association 96*(456), 1348–1360.

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

Fan, Y. and C. Y. Tang (2013). Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society: SERIES B: Statistical Methodology*, 531–552.
Gungor, S. and R. Luger (2019). Exact inference in long-horizon predictive quantile regressions with an application to stock returns. *Journal of Financial Econometrics*.

Hodrick, R. J. (1992). Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *The Review of Financial Studies* 5(3), 357–386.

Knight, K. and W. Fu (2000). Asymptotics for lasso-type estimators. *Annals of Statistics*, 1356–1378.

Kock, A. B. (2016). Consistent and conservative model selection with the adaptive lasso in stationary and nonstationary autoregressions. *Econometric Theory* 32(1), 243.

Kock, A. B. and L. Callot (2015). Oracle inequalities for high dimensional vector autoregressions. *Journal of Econometrics* 186(2), 325–344.

Koenker, R. (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica*, 33–50.

Koo, B., H. M. Anderson, M. H. Seo, and W. Yao (2020). High-dimensional predictive regression in the presence of cointegration. *Journal of Econometrics*.

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

Lee, J. H., Z. Shi, and Z. Gao (2021). On lasso for predictive regression. *Journal of Econometrics*.

Lee, S., Y. Liao, M. H. Seo, and Y. Shin (2018). Oracle estimation of a change point in high-dimensional quantile regression. *Journal of the American Statistical Association* 113(523), 1184–1194.

Li, J. and Z. Liao (2020). Uniform nonparametric inference for time series. *Journal of Econometrics* 219(1), 38–51.

Li, Y. and J. Zhu (2008). L1-norm quantile regression. *Journal of Computational and Graphical Statistics* 17(1), 163–185.

Liang, C. and M. Schienle (2019). Determination of vector error correction models in high dimensions. *Journal of Econometrics* 208(2), 418–441.

Liao, Z. and P. C. Phillips (2015). Automated estimation of vector error correction models. *Econometric Theory*, 581–646.

Lu, X. and L. Su (2015). Jackknife model averaging for quantile regressions. *Journal of Econometrics* 188(1), 40–58.
Maynard, A., K. Shimotsu, and Y. Wang (2011). Inference in predictive quantile regressions. *Unpublished manuscript.*

Medeiros, M. C. and E. F. Mendes (2016). l1-regularization of high-dimensional time-series models with non-gaussian and heteroskedastic errors. *Journal of Econometrics 191*(1), 255–271.

Meinshausen, N. and P. Bühlmann (2004). Consistent neighbourhood selection for sparse high-dimensional graphs with the lasso. Seminar für Statistik, Eidgenössische Technische Hochschule (ETH), Zürich.

Merlevède, F., M. Peligrad, E. Rio, et al. (2009). Bernstein inequality and moderate deviations under strong mixing conditions. *High dimensional probability V: the Luminy volume 5*, 273–292.

Onatski, A. and C. Wang (2018). Alternative asymptotics for cointegration tests in large vars. *Econometrica 86*(4), 1465–1478.

Phillips, P. C. (1991). Optimal inference in cointegrated systems. *Econometrica 59*(2), 283–306.

Phillips, P. C. and J. H. Lee (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics 177*(2), 250–264.

Phillips, P. C. and J. H. Lee (2016). Robust econometric inference with mixed integrated and mildly explosive regressors. *Journal of Econometrics 192*(2), 433–450.

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

Portnoy, S. (1984). Asymptotic behavior of m-estimators of p regression parameters when \( p^2/n \) is large. i. consistency. *Annals of Statistics 12*(4), 1298–1309.

Portnoy, S. (1985). Asymptotic behavior of m estimators of p regression parameters when \( p^2/n \) is large; ii. normal approximation. *Annals of Statistics*, 1403–1417.

Ruppert, D. and R. J. Carroll (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association 75*(372), 828–838.

Shao, J. (1997). An asymptotic theory for linear model selection. *Statistica Sinica*, 221–242.

Sherwood, B. and L. Wang (2016). Partially linear additive quantile regression in ultra-high dimension. *Annals of Statistics 44*(1), 288–317.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological) 58*(1), 267–288.

Wang, H. and C. Leng (2007). Unified lasso estimation by least squares approximation. *Journal of the American Statistical Association 102*(479), 1039–1048.
Wang, H., B. Li, and C. Leng (2009). Shrinkage tuning parameter selection with a diverging number of parameters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71(3), 671–683.

Wang, H., G. Li, and G. Jiang (2007). Robust regression shrinkage and consistent variable selection through the lad-lasso. *Journal of Business & Economic Statistics* 25(3), 347–355.

Wang, H., R. Li, and C.-L. Tsai (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika* 94(3), 553–568.

Wang, L., Y. Wu, and R. Li (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association* 107(497), 214–222.

Wasserman, L. and K. Roeder (2009). High dimensional variable selection. *Annals of Statistics* 37(5A), 2178.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies* 21(4), 1455–1508.

Withers, C. (1981). Conditions for linear processes to be strong-mixing. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 57(4), 477–480.

Wong, K. C., Z. Li, and A. Tewari (2020). Lasso guarantees for $\beta$-mixing heavy-tailed time series. *Annals of Statistics* 48(2), 1124–1142.

Wu, W.-B. and Y. N. Wu (2016). Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electronic Journal of Statistics* 10(1), 352–379.

Wu, Y. and Y. Liu (2009). Variable selection in quantile regression. *Statistica Sinica*, 801–817.

Xiao, Z. (2009). Quantile cointegrating regression. *Journal of Econometrics* 150(2), 248–260.

Yuan, M. and Y. Lin (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68(1), 49–67.

Zhang, R., P. Robinson, and Q. Yao (2019). Identifying cointegration by eigenanalysis. *Journal of the American Statistical Association* 114(526), 916–927.

Zhang, Y., R. Li, and C.-L. Tsai (2010). Regularization parameter selections via generalized information criterion. *Journal of the American Statistical Association* 105(489), 312–323.

Zheng, Q., C. Gallagher, and K. Kulasekera (2013). Adaptive penalized quantile regression for high dimensional data. *Journal of Statistical Planning and Inference* 143(6), 1029–1038.

Zheng, Q., L. Peng, and X. He (2015). Globally adaptive quantile regression with ultra-high dimensional data. *Annals of Statistics* 43(5), 2225.
Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* 101(476), 1418–1429.

Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67(2), 301–320.