Estimation of a Dynamic Tobit Model with a Unit Root *

Anna Bykhovskaya
Duke University

James A. Duffy
University of Oxford

11th December 2025

Abstract

This paper studies robust estimation in the dynamic Tobit model under local-to-unity (LUR) asymptotics. We show that both Gaussian maximum likelihood (ML) and censored least absolute deviations (CLAD) estimators are consistent, extending results from the stationary case where ordinary least squares (OLS) is inconsistent. The asymptotic distributions of MLE and CLAD are derived; for the short-run parameters they are shown to be Gaussian, yielding standard normal t-statistics. In contrast, although OLS remains consistent under LUR, its t-statistics are not standard normal. These results enable reliable model selection via sequential t-tests based on ML and CLAD, paralleling the linear autoregressive case. Applications to financial and epidemiological time series illustrate their practical relevance.

Keywords: dynamic Tobit, unit root, MLE, CLAD.

*The authors would like to thank Yiting Fu for excellent research assistance.

Contents

1 Introduction

Many macroeconomic and financial time series are highly persistent, often exhibiting the random wandering that is the hallmark of a stochastic trend. While it is relatively straightforward to accommodate these series within the framework of a linear (vector) autoregressive model, even in this setting such persistence creates significant challenges for inference, giving rise to limiting distributions that are non-standard and dependent on nuisance parameters. Accordingly, a substantial literature has developed on inference methods that are broadly robust to temporal dependence, remaining valid for both stationary and stochastically trending data (see e.g. Hansen, 1999; Jansson and Moreira, 2006; Mikusheva, 2007, 2012; Elliott, Müller, and Watson, 2015).

Though this literature has now attained a high degree of maturity, a significant limitation is that it is entirely devoted to linear models. Although nonlinear autoregressive models have been widely studied, this has almost exclusively been in the context of stationary processes (as in e.g. Tong, 1990; Chan, 2009; Terasvirta, Tjøstheim, and Granger, 2010). Indeed, for a long time the only available dynamic models that permitted some conjunction of nonlinearities with stochastic trends were those in which the nonlinearities were confined to the short-run dynamics, i.e. to first differences and equilibrium error components (as in the nonlinear VECM models considered by Balke and Fomby, 1997; Hansen and Seo, 2002; Kristensen and Rahbek, 2010). This excluded, for example, nonlinear autoregressive models with regimes endogenously dependent on the *level* of a stochastically trending series. Only recently has it been shown possible to configure nonlinear (vector) autoregressive models so as to accommodate both nonlinearity in levels and stochastic trends (Cavaliere, 2005; Bykhovskaya and Duffy, 2024; Duffy and Mavroeidis, 2024; Duffy, Mavroeidis, and Wycherley, 2025).

A major impetus for this recent work has come from the desire to adequately model highly persistent series that are subject to occasionally binding constraints, of which the zero lower bound on nominal interest rates is a leading example (Mavroeidis, 2021, Aruoba, Mlikota, Schorfheide, and Villalvazo, 2022). For such series, a plausible descriptive model is the dynamic Tobit, which in a stationary setting has a lengthy pedigree in the empirical analysis of constrained, i.e. censored, time series (see e.g. Demiralp and Jordà, 2002, de Jong and Herrera, 2011, Dong, Schmit, and Kaiser, 2012, Liu, Moon, and Schorfheide, 2019, Brezigar-Masten, Masten, and Volk, 2021, Bykhovskaya, 2023). Bykhovskaya and Duffy (2024) recently determined conditions under which this model may also generate stochastically trending, $I(1)$-like series, and obtained the asymptotic distribution of OLS estimators in a local-to-unity framework. However, while their results are sufficient for the purposes of conducting unit root tests, there is no possibility of extending the validity of OLS beyond their setting, since even its consistency is known to fail for the stationary dynamic Tobit.

The present work is thus motivated by the need to develop methods of estimation and inference, in a *nonlinear* time series setting, that are robust to the degree of persistence of the data generating mechanism. More specifically, we seek methods that enjoy validity for the dynamic Tobit model on a wider domain that allows for both stochastically trending and stationary data, exactly as OLS enjoys in a linear autoregressive model. This leads us to consider two alternative estimators that are known to be consistent and asymptotically normal for the stationary dynamic Tobit: Gaussian maximum likelihood (ML) and Powell's (1984) censored least absolute deviations (CLAD; see de Jong and Herrera, 2011, and Bykhovskaya, 2023).

The principal contribution of this paper is to derive the asymptotic distributions of the ML and CLAD estimators for a dynamic Tobit model, in a local-to-unity framework. When the model is rendered in ADF form, the limiting distribution of the estimator of the sum of the autoregressive coefficients is non-standard, but takes a known form that is suitable for inference. Moreover, estimates of 'short run' coefficients (i.e. those on lagged differences) have an asymptotically Gaussian distribution, so that inferences on these and on certain related functionals, such as short-horizon impulse responses, remain standard irrespective of the degree of persistence of the regressor (exactly as is the case for OLS estimation in a linear autoregressive model). These results facilitate the determination of the model lag order on the basis of sequential t-tests.

At a technical level, our results contribute to the literature on the asymptotics of nonlinear extremum estimators with highly peristent processes (see e.g. Park and Phillips, 2000, 2001; Phillips, Jin, and Hu, 2007; Xiao, 2009; Chan and Wang, 2015). Relative to previous work, the analysis is complicated by two aspects of our problem. Firstly, we consider a nonlinear *autoregressive* model, whereas the literature has been concerned with regression (or discrete choice) models where the regressor x_t is a *linear* unit root process, and the object is to estimate some (possibly) nonlinear transformation of x_t. Secondly, in our analysis of the CLAD estimator, we cannot rely on the convexity of the criterion function, something which greatly facilitates the derivation of the asymptotics of the *uncensored* LAD estimator, as in Pollard (1991), Herce (1996), Li and Li (2009) and Xiao (2009). (By contrast, convexity may be fruitfully exploited in the analysis of the Gaussian MLE.) Nor may approaches employed in the i.i.d. or stationary cases (Powell, 1984; de Jong and Herrera, 2011; Bykhovskaya, 2023) be readily transposed to our setting. Thus deriving the asymptotics of the (centred) CLAD criterion function, in particular, requires some delicate and novel arguments. We expect these arguments will also prove useful for the analysis of other nonlinear extremum estimators, in the presence of highly persistent data.

The remainder of this paper is organised as follows. Section 2 introduces the dynamic Tobit model and our assumptions. Section 3 derives the asymptotic distributions of ML and CLAD estimators. A discussion of the results, and an empirical illustration, is presented in Section 4. Finally, Section 5 concludes. All proofs appear in the appendices.

Notation. All limits are taken as $T \to \infty$ unless otherwise stated. $\overset{p}{\to}$ and $\overset{d}{\to}$ respectively denote convergence in probability and in distribution (weak convergence). We write '$X_T(\lambda) \overset{d}{\to} X(\lambda)$ on $D_{\mathbb{R}^m}[0,1]$' to denote that $\{X_T\}$ converges weakly to X, where these are considered as random elements of $D_{\mathbb{R}^m}[0,1]$, the space of cadlag functions $[0,1] \to \mathbb{R}^m$, equipped with the uniform topology; we denote this as $D[0,1]$ whenever the value of m is clear from the context. $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^m$, and the matrix norm that it induces. For X a random vector and $p \geq 1$, $\|X\|_p := (\mathbb{E}\|X\|^p)^{1/p}$.

2 Model

We suppose $\{y_t\}$ is generated by a dynamic Tobit model of order k, written in augmented Dickey–Fuller (ADF) form as

$$y_t = \left[\alpha_0 + \beta_0 y_{t-1} + \boldsymbol{\phi}_0^{\mathsf{T}} \Delta \boldsymbol{y}_{t-1} + u_t\right]_+, \qquad t = 1, \dots T, \tag{2.1}$$

where $[x]_+ := \max\{x, 0\}$, $\boldsymbol{y}_{t-1} := (y_{t-1}, \dots, y_{t-k+1})^{\mathsf{T}}$, $\boldsymbol{\phi}_0 = (\phi_{1,0}, \dots, \phi_{k-1,0})^{\mathsf{T}}$ and $\Delta y_t := y_t - y_{t-1}$. The model has two attractive features. Firstly, it is Markovian – with the state vector defined by k consecutive lags of y_t – making it well-suited for forecasting. Secondly, the presence of the positive part on the right of the equality enforces a lower bound of zero on y_t, making the model appropriate for series that are constrained to take non-negative values.

We assume that the parameters $\rho_0 := (\alpha_0, \beta_0, \boldsymbol{\phi}_0^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{R}^{k+1}$, innovations $\{u_t\}$ and initial conditions satisfy the following assumptions.

Assumption A1. *$\{y_t\}$ is initialised by (possibly) random variables $\{y_{-k+1}, \dots, y_0\}$. Moreover, $T^{-1/2} y_0 \overset{p}{\to} b_0$ for some $b_0 \geq 0$.*

Assumption A2. *$\{y_t\}$ is generated according to (2.1), where:*

1. *$\{u_t\}_{t \in \mathbb{Z}}$ is independently and identically distributed with $\mathbb{E}u_t = 0$ and $\mathbb{E}u_t^2 = \sigma_0^2$.*

2. *$\alpha_0 = \alpha_{T,0} := T^{-1/2} a_0$ and $\beta_0 = \beta_{T,0} = 1 + T^{-1} c_0$, for some $a_0, c_0 \in \mathbb{R}$.*

Assumption A3. *There exist $\delta_u > 0$ and $C < \infty$ such that:*

1. *$\mathbb{E}|u_t|^{2+\delta_u} < C$.*

2. *$\mathbb{E}|T^{-1/2} y_0|^{2+\delta_u} < C$, and $\mathbb{E}|\Delta y_i|^{2+\delta_u} < C$ for $i \in \{-k+2, \dots, 0\}$.*

Assumption A1 allows the initialisation to be of the same order of magnitude as the process y_t. This is appropriate because the first observation in any sample is unlikely to be the 'true' starting point of the process, but simply the point at which data collection begins. Therefore, it is natural to allow y_0 to be of the same scale as any later y_t. Assumptions A2.1 and A3 are standard conditions on the innovations. Assumption A2.2 ensures that the data is highly persistent, as a consequence of the autoregressive polynomial having a root local to unity.

As shown in Bykhovskaya and Duffy (2024, Appendix C), owing to the nonlinearity of the model additional technical conditions – which go beyond the classical requirements on the stability of the roots of the autoregressive polynomial – are needed to rule out explosive behaviour of the first differences $\{\Delta y_t\}$. Here we ensure this through a constraint on the joint spectral radius of two auxiliary companion-form matrices. For $\delta \in [0, 1]$, define

$$F_\delta := \begin{bmatrix} \phi_{1,0}\delta & \phi_{2,0} & \cdots & \phi_{k-2,0} & \phi_{k-1,0} \\ \delta & 0 & \cdots & 0 & 0 \\ 0 & 1 & & & \\ & & \ddots & & \\ & & & 1 & 0 \end{bmatrix} \tag{2.2}$$

and let $\lambda_{\mathrm{JSR}}(\mathcal{A})$ denote the joint spectral radius of a bounded collection of matrices $\mathcal{A}$, which may be defined as

$$\lambda_{\mathrm{JSR}}(\mathcal{A}) := \limsup_{n\to\infty} \sup_{M\in\mathcal{A}^n} \lambda(M)^{1/n}$$

where $\lambda(M)$ denotes the spectral radius of M, and $\mathcal{A}^n := \{\prod_{i=1}^n A_i \mid A_i \in \mathcal{A}\}$ (cf., Jungers, 2009, Defn. 1.1).

Assumption A4. $\lambda_{\mathrm{JSR}}(\{F_0, F_1\}) < 1$.

Approximate upper bounds for the joint spectral radius (JSR) can be computed numerically to an arbitrarily high degree of accuracy using semidefinite programming (Parrilo and Jadbabaie, 2008), making it possible to verify whether the condition is satisfied for a given set of parameter values (see Duffy, Mavroeidis, and Wycherley, 2023, for a further discussion). Assumption A4 ensures that the process y_t does not exhibit explosive behavior by jointly controlling the dynamics across both the censored ($\delta = 0$) and uncensored ($\delta = 1$) 'regimes'. Without this restriction the interaction between the two regimes could generate a 'bouncing' effect, where hitting the lower bound triggers a strong rebound, potentially leading to exponential growth.

Bykhovskaya and Duffy (2024, Theorem 3.2) show that under Assumptions A1-A4

$$\frac{1}{\sqrt{T}} y_{\lfloor \tau T \rfloor} \xrightarrow{d} \phi(1)^{-1} e^{c_0\tau/\phi(1)} \left(K(\tau) + \sup_{\tau' \le \tau} [-K(\tau')]_+ \right) =: Y(\tau), \tag{2.3}$$

on $D[0,1]$, where

$$K(\tau) \coloneqq \phi(1)b_0 + a_0 \int_0^\tau e^{-c_0 r/\phi(1)}\, \mathrm{d}r + \sigma_0 \int_0^\tau e^{-c_0 r/\phi(1)}\, \mathrm{d}W(r)$$

for $\phi(1) = 1 - \sum_{i=1}^{k-1} \phi_{i,0} > 0$, and $W(\cdot)$ a standard Brownian motion. The limiting distribution of the estimators of α and β, though not of $\boldsymbol{\phi}$, will be shown below to depend on the process $Y(\cdot)$ – a dependence similar to that of the OLS estimators on the limiting Brownian Motion (or Ornstein–Uhlenbeck process) in a linear autoregressive model with a root at (or local to) unity. In fact, the process $K(\cdot)$ – which $Y(\cdot)$ is the constrained counterpart of – corresponds exactly to the limit of a local-to-unity autoregressive process. So surprisingly, despite the Tobit model's added complexity due to censoring, a similar asymptotic structure emerges: the limiting distributions of our estimators will closely resemble those of a linear autoregression, albeit with a nonlinear limiting process $Y(\cdot)$.

3 Estimation and inference

This section introduces two estimation methods – Gaussian maximum likelihood (ML) and censored least absolute deviations (CLAD) – and establishes their asymptotic properties. We show that the estimators of the coefficients $\boldsymbol{\phi}_0$ on the stochastically bounded (though not stationary) lags $\Delta \boldsymbol{y}_{t-1}$ are asymptotically normal under both methods, permitting standard inferences to be drawn. As discussed further in Section 4.2, this contrasts sharply with the behaviour of ordinary least squares estimators, which have some limiting bias (of order $T^{-1/2}$) due to the censoring, even in the presence of a local to unit root.

3.1 Maximum likelihood

To permit estimation by maximum likelihood, we need to make a parametric assumption regarding the distribution of the innovations $\{u_t\}$. A conventional choice – which is particularly attractive for computational reasons, because it yields a loglikelihood with a convex reparametrisation – is for u_t to be Gaussian, as per the following.

Assumption B. *$u_t \sim N[0, \sigma_0^2]$, and $\delta_u > 2$ in A3.*

Note that for $\{u_t\}$ itself, the maintained Gaussianity makes the condition on δ_u redundant; this condition is nonetheless still needed instead to ensure that the initial conditions also have a little more than four finite moments.

Under assumption B, when $\{y_t\}$ is generated according to (2.1), the conditional density of y_t given $(y_{t-1}, \dots, y_{t-k})$, evaluated at the values $y_{t-i} = \mathsf{y}_{t-i}$ for $i \in 0, \dots, k$, is

$$f_{(\alpha,\beta,\boldsymbol{\phi},\sigma)}(\mathsf{y}_t \mid \mathsf{y}_{t-1}, \dots, \mathsf{y}_{t-k}) = \begin{cases} \sigma^{-1}\varphi[\sigma^{-1}(\mathsf{y}_t - \alpha - \beta \mathsf{y}_{t-1} - \boldsymbol{\phi}^\mathsf{T} \Delta \boldsymbol{\mathsf{y}}_{t-1})] & \text{if } \mathsf{y}_t > 0, \\ 1 - \Phi[\sigma^{-1}(\alpha + \beta \mathsf{y}_{t-1} + \boldsymbol{\phi}^\mathsf{T} \Delta \boldsymbol{\mathsf{y}}_{t-1})] & \text{if } \mathsf{y}_t = 0; \end{cases} \tag{3.1}$$

where $\mathsf{y}_{t-1} = (\mathsf{y}_{t-1}, \dots, \mathsf{y}_{t-k+1})^\mathsf{T}$, and $\Phi(x)$ and $\varphi(x)$ denote the standard normal cdf and pdf functions, i.e. $\varphi(x) \coloneqq \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$. We consider a conditional ML estimator that maximises the loglikelihood of the $\{y_t\}_{t=1}^T$ conditional on the initial values $\{y_i\}_{i=-k+1}^0$, i.e. which maximises

$$\mathcal{L}_T(\alpha, \beta, \boldsymbol{\phi}, \sigma) \coloneqq \sum_{t=1}^T \log f_{(\alpha,\beta,\boldsymbol{\phi},\sigma)}(y_t \mid y_{t-1}, \dots, y_{t-k})$$

with respect to $\alpha, \beta, \boldsymbol{\phi}, \sigma$. Define

$$(\hat{\alpha}_T^{\mathsf{M}}, \hat{\beta}_T^{\mathsf{M}}, \hat{\boldsymbol{\phi}}_T^{\mathsf{M}}, \hat{\sigma}_T^{\mathsf{M}}) \in \operatorname*{argmax}_{(\alpha,\beta,\boldsymbol{\phi},\sigma) \in \mathbb{R}^{k+2} \times \mathbb{R}_+} \mathcal{L}_T(\alpha, \beta, \boldsymbol{\phi}, \sigma)$$

to be a sequence of maximisers of $\mathcal{L}_T$. Let $\Omega \coloneqq \operatorname{plim}_{T\to\infty} \frac{1}{T} \sum_{t=1}^T \Delta\mathbf{y}_{t-1} \Delta\mathbf{y}_{t-1}^\mathsf{T}$, which exists as a consequence of Lemma B.4 in Bykhovskaya and Duffy (2024).

Theorem 3.1. *Suppose* A1–A4 *and* B *hold. Then*

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}_T^{\mathsf{M}} - \alpha_0) \\ T(\hat{\beta}_T^{\mathsf{M}} - \beta_0) \end{bmatrix} \xrightarrow[T\to\infty]{d} \sigma_0 \begin{bmatrix} 1 & \int_0^1 Y(\tau)\,\mathrm{d}\tau \\ \int_0^1 Y(\tau)\,\mathrm{d}\tau & \int_0^1 Y^2(\tau)\,\mathrm{d}\tau \end{bmatrix}^{-1} \begin{bmatrix} W(1) \\ \int_0^1 Y(\tau)\,\mathrm{d}W(\tau) \end{bmatrix} \tag{3.2}$$

jointly with

$$T^{1/2} \begin{bmatrix} \hat{\boldsymbol{\phi}}_T^{\mathsf{M}} - \boldsymbol{\phi}_0 \\ (\hat{\sigma}_T^{\mathsf{M}})^2 - \sigma_0^2 \end{bmatrix} \xrightarrow[T\to\infty]{d} \begin{bmatrix} \sigma_0^2 \Omega^{-1} & 0 \\ 0 & (2\sigma_0^2)^{-1} \end{bmatrix}^{1/2} \xi \tag{3.3}$$

where $\xi \sim \mathcal{N}[0, I_k]$ is independent of W (and therefore also Y).

In contrast to the classical autoregressive setting, where the Gaussian MLE is numerically identical to OLS, and therefore has the same asymptotic distribution, Theorem 3.1 and simulations in Section 4.2 show that censoring breaks this equivalence (see Bykhovskaya and Duffy (2024, Theorem 3.4) for the OLS limit theory). Interestingly, though, up to a change of the rescaled limit of y_t, the ML distributions in Theorem 3.1, match their linear analogues, see, e.g., Hamilton (1994, (17.7.25), (17.7.27)).

The asymptotic variance in (3.3) is given by the inverse of the information matrix, i.e., the inverse of the negative Hessian (see Proposition B.1), matching the corresponding result in the stationary case (de Jong and Herrera, 2011, Theorem 3). Unlike in the stationary case, where censoring binds much more frequently ($O(T)$ rather than $O(\sqrt{T})$ times), here we can obtain an explicit form for the Hessian, which is unaffected by censored observations apart from their impact on the asymptotic distribution of y_t in (3.2).

3.2 Censored least absolute deviations

The censored least absolute deviations (CLAD) estimator of (2.1), which we denote as $(\hat{\alpha}_T^{\mathsf{L}}, \hat{\beta}_T^{\mathsf{L}}, \hat{\boldsymbol{\phi}}_T^{\mathsf{L}})$, minimises

$$S_T(\alpha, \beta, \boldsymbol{\phi}) := \sum_{t=1}^{T} \left| y_t - \left[\alpha + \beta y_{t-1} + \boldsymbol{\phi}^{\mathsf{T}} \Delta \boldsymbol{y}_{t-1}\right]_+ \right| \tag{3.4}$$

as per Powell (1984). The presence of the positive part – reflecting Tobit-type censoring as in (2.1) – within the objective function (3.4) renders the criton function non-convex, making the analysis of the CLAD estimator rather more challenging than that of its uncensored counterpart. A further complication arises from the high persistence in $\{y_t\}$, which prevents the application of maximal inequalities or uniform central limit theorems appropriate to stationary processes.

Assumption C.

1. u_t *is continuously distributed, with density* f_u *that is positive at zero, bounded, and continuously differentiable with bounded derivatives,* $\operatorname{med}(u_t) = 0$ *and* $\delta_u > 2$ *in* A3*; and*

2. $(\alpha_0, \beta_0, \boldsymbol{\phi}_0^{\mathsf{T}})^{\mathsf{T}} \in \Pi$, *for some compact* $\Pi \subset \mathbb{R}^{k+2}$*, and* $(0, 1, \boldsymbol{\phi}_0^{\mathsf{T}})^{\mathsf{T}} \in \operatorname{int} \Pi$.

We expect that our requirement that u_t have a little more than four finite moments ($\delta_u > 2$) could be relaxed to two finite moments. Obtaining the limiting distribution in the latter case would necessitate a sharper bound on the order of

$$\sum_{t=1}^{T} \mathbf{1}\{\alpha_0 + \beta_0 y_{t-1} + \boldsymbol{\phi}_0^{\mathsf{T}} \Delta \boldsymbol{y}_{t-1} \leq 0\},$$

which is closely related to the rate at which censored observations occur. The arguments used to prove the following theorem rely on an estimate of $o_p(T)$ for such terms, but this could likely be improved to $O_p(T^{1/2})$.

Theorem 3.2. *Suppose that Assumptions* A1-A4 *and* C *hold. Then*

$$\begin{bmatrix} T^{1/2}(\hat{\alpha}_T^{\mathsf{L}} - \alpha_0) \\ T(\hat{\beta}_T^{\mathsf{L}} - \beta_0) \end{bmatrix} \xrightarrow[T\to\infty]{d} \frac{1}{2 f_u(0)} \begin{bmatrix} 1 & \int_0^1 Y(\tau)\,\mathrm{d}\tau \\ \int_0^1 Y(\tau)\,\mathrm{d}\tau & \int_0^1 Y^2(\tau)\,\mathrm{d}\tau \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{W}(1) \\ \int_0^1 Y(\tau) d\widetilde{W}(\tau) \end{bmatrix} \tag{3.5}$$

jointly with, and independently of

$$\sqrt{T}(\hat{\boldsymbol{\phi}}_T^{\mathsf{L}} - \boldsymbol{\phi}_0) \xrightarrow[T\to\infty]{d} \frac{1}{2 f_u(0)} \mathcal{N}(0, \Omega^{-1}), \tag{3.6}$$

where $\Omega = \mathrm{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \Delta\mathbf{y}_{t-1}\Delta\mathbf{y}_{t-1}^{\mathsf{T}}$ *and* $\widetilde{W}(\cdot)$ *is a* 1*-dimensional Brownian motion, such that the covariance matrix of* $(\sigma_0 W(\cdot), \widetilde{W}(\cdot))^{\mathsf{T}}$ *is*

$$\Sigma = \begin{pmatrix} \sigma_0^2 & \mathbb{E}|u_t| \\ \mathbb{E}|u_t| & 1 \end{pmatrix}.$$

Comparing our asymptotic results with those for (uncensored) LAD in the linear regression setting of Herce (1996, Theorem 1), we see that the structure of our limting distributions closely mirrors its linear counterpart, with the only difference being the limiting processes for $\{y_t\}$ (upon rescaling) that appears in the two distributional limits. This resemblance is particularly surprising given that the observations with $y_t = 0$ cannot be ignored; rather, they play a crucial role in the derivation of Theorem 3.2.

There, the key step involves deriving the first- and second-order terms in the expansion of the CLAD objective function evaluated at an arbitrary parameter vector relative to its value at the true parameters. The first-order term converges to a quadratic form in the parameters, while the second-order term converges to a linear function. This structure enables us to characterize the solution to the limiting optimization problem.

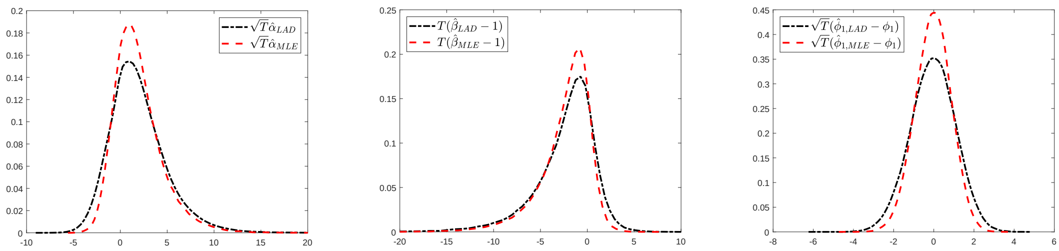
It is also worth comparing the results of Theorem 3.2 with their stationary counterparts de Jong and Herrera (2011, Theorem 5) and Bykhovskaya (2023, Theorem 6). Both share the same factor $\frac{1}{2f_u(0)}$, where the density at zero arises from a Taylor expansion and governs the behavior when y_{t-1} is close to zero. Moreover, the asymptotics for the short-run coefficients $\hat{\boldsymbol{\phi}}^{\mathsf{L}}$ match their stationary analogue, since the number of observations with $y_t = 0$ is small enough that $\frac{1}{T}\sum_t \Delta\mathbf{y}_{t-1}\Delta\mathbf{y}_{t-1}^{\mathsf{T}}$ and $\frac{1}{T}\sum_t \mathbf{1}\{y_t > 0\}\Delta\mathbf{y}_{t-1}\Delta\mathbf{y}_{t-1}^{\mathsf{T}}$ are asymptotically identical.

4 Discussion

4.1 Pros and cons of CLAD and MLE

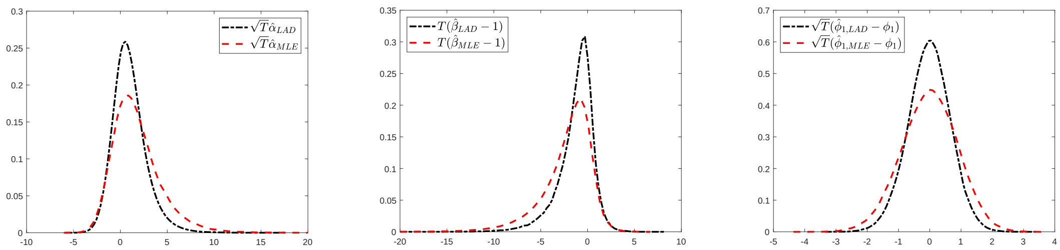
When u_t follows a standard normal distribution, both Theorems 3.1 and 3.2 are applicable, allowing us to compare the variances of the short-run coefficients in MLE and CLAD. In this case, $\frac{1}{2f_u(0)} = \sigma_0\sqrt{\pi/2} \approx 1.25\sigma_0$, implying roughly a 25% increase in the standard deviation of CLAD relative to the (optimal) ML estimator. The increased asymptotic variance of all CLAD coefficients compared to MLE is illustrated in Figure 4.1. We observe that, because the asymptotic distributions of both CLAD and MLE estimators of α and β involve the non-standard limiting process $Y(\cdot)$, they are skewed and centered at positive values for α and negative values for β, with the MLE exhibiting tighter concentration than CLAD. In contrast, the short-run estimators are all centered at zero and approximately normal, though again with CLAD displaying higher variance than MLE.

On the other hand, relative to the (Gaussian) MLE, a major advantage of the CLAD



(a) $\sqrt{T}\hat{\alpha}$. (b) $T(\hat{\beta}-1)$. (c) $\sqrt{T}(\hat{\phi}_1-\phi_1)$.

Figure 4.1: Asymptotic distributions based on MLE and CLAD. Data generating process: $y_t = [y_{t-1} + 0.5\Delta y_{t-1} + u_t]_+$, $y_0 = y_{-1} = 0$, $u_t \sim$ i.i.d. $\mathcal{N}(0,1)$, $T = 1000$, $MC = 100,000$ replications.



(a) $\sqrt{T}\hat{\alpha}$. (b) $T(\hat{\beta}-1)$. (c) $\sqrt{T}(\hat{\phi}_1-\phi_1)$.

Figure 4.2: Asymptotic distributions based on MLE and CLAD when errors follow Laplace distribution. Data generating process: $y_t = [y_{t-1} + 0.5\Delta y_{t-1} + u_t]_+$, $y_0 = y_{-1} = 0$, $u_t \sim$ i.i.d. $Laplace(0, 1/\sqrt{2})$, $T = 1000$, $MC = 100,000$ replications.

estimator is that it is semiparametric, insofar as it does not require a parametric assumption on the distribution of the errors u_t. As illustrated in Section 4.3.2 and previously in de Jong and Herrera (2011); Bykhovskaya (2023), real-world data can deviate substantially from Gaussianity, in which case the CLAD estimator is likely to yield more reliable inferences.

Our simulations suggest that the asymptotic behavior of MLE changes little when the errors are non-Gaussian, so long as they have sufficient moments, consistent with the well-attested good performance of quasi-MLEs. However, depending on the value of $f_u(0)$, CLAD may now be more efficient than the quasi-MLE. For example, if u_t follows a Laplace$(0, 1/\sqrt{2})$ distribution, so that $\mathbb{E}u_t = 0$ and $\mathbb{E}u_t^2 = 1$, we have $f_u(0) = 1/\sqrt{2}$ and $1/2f_u(0) = 1/\sqrt{2} \approx 0.7 < 1$. In this case, the MLE standard deviation for the short-run coefficients is almost 1.5 times larger than that of CLAD. This is illustrated in Figure 4.2, where the behaviour of the MLE estimator is very similar to that observed for Gaussian errors in Figure 4.1, while the CLAD estimator is now much more tightly distributed, reflecting the higher value of the density at zero ($1/\sqrt{2}$ for the Laplace distribution versus $1/\sqrt{2\pi}$ for the standard normal). The same pattern is observed when the errors follow a t-distribution with ν degrees of freedom, where $\nu \in (2, 4.6]$ to ensure that variance is well-defined and $2f_u(0) > 1$.

The main drawbacks of the CLAD estimator come from two aspects. First, the CLAD optimisation problem is non-convex, potentially causing numerical optimisers to converge local rather than global optima, and thus to be sensitive to their initialisation. The Gaussian loglikelihood, by contrast, may be reparameterised so as to render the objective function concave, thereby avoiding this problem.[1] (This may in turn be used to provide reasonable starting values for the optimisation of the CLAD criterion function.) The second drawback stems from the necessity of estimating $f_u(0)$ for the purposes of inference (i.e. to construct confidence intervals or conduct hypothesis tests). We tackle this by means of kernel density estimation, as proposed in Powell (1984, (5.2) and (5.5)), though this has the undesirable consequence of rendering inferences dependent on the choice of a bandwidth parameter.

4.2 Comparison with OLS

It is noteworthy that all three estimators share a common structure in their asymptotic distributions: the inverse of

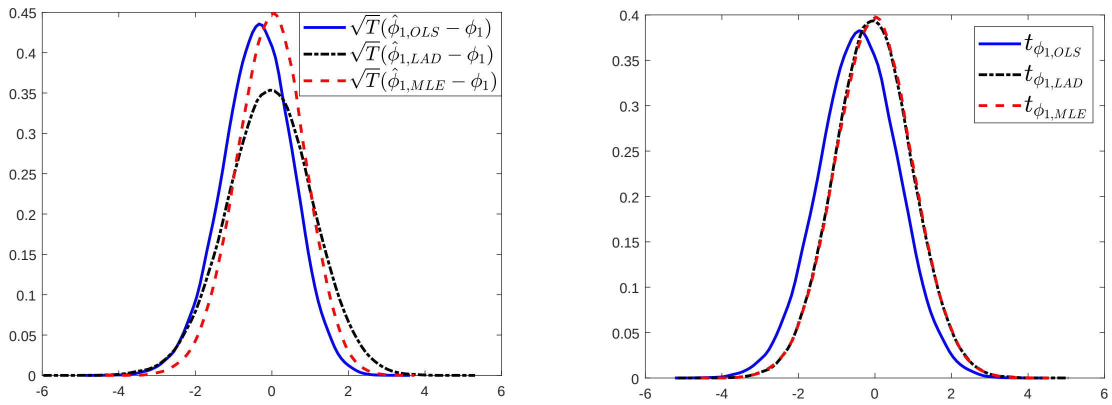
$$\begin{bmatrix} 1 & \int_0^1 Y(\tau)\,\mathrm{d}\tau \\ \int_0^1 Y(\tau)\,\mathrm{d}\tau & \int_0^1 Y^2(\tau)\,\mathrm{d}\tau \end{bmatrix}$$

appears in the asymptotics of the the estimators of (α, β), while the limiting variance of the estimators of $\boldsymbol{\phi}$ depends on the inverse of Ω. This structure mirrors the limit of the properly rescaled OLS regressor signal matrix $X^\intercal X$, for the regressors $(1, y_{t-1}, \Delta\boldsymbol{y}_{t-1}^\intercal)^\intercal$. The similarity arises because, up to proportionality, the first-order Taylor expansions of the estimators coincide; the differences emerge in the behavior of fluctuations, that is, in the second-order terms.

There are two significant advantages to using MLE or CLAD relative to OLS. The first advantage is that the former two are consistent both under stationary and (near) unit root regimes, see e.g., (de Jong and Herrera, 2011; Bykhovskaya, 2023), while OLS is consistent only under the (near) unit root regime (see, e.g., Bykhovskaya (2023, Supplementary material) for inconsistency results and (Bykhovskaya and Duffy, 2024) for consistency results). Since an empirical researcher is typically a priori uncertain about the level of persistence in the data, using MLE or LAD permits valid inferences to be drawn more robustly.

The second advantage is related to the estimation of the 'short-run' parameters ϕ_i, for $i \in \{1, \ldots, k-1\}$, i.e. the coefficients on the lagged differences Δy_{t-i}. As is shown in Theorems 3.1 and 3.2, both the MLE and CLAD estimators of theses are asymptotically normal, leading to standard normal t-statistics. By contrast, as illustrated in Figure 4.3,

[1] As per Olsen (1978), the new parameters are α/σ, β/σ, $\boldsymbol{\phi}/\sigma$, and σ^{-1}.



(a) $\sqrt{T}(\hat{\phi}_1 - \phi_1)$. (b) t-statistic for ϕ_1.

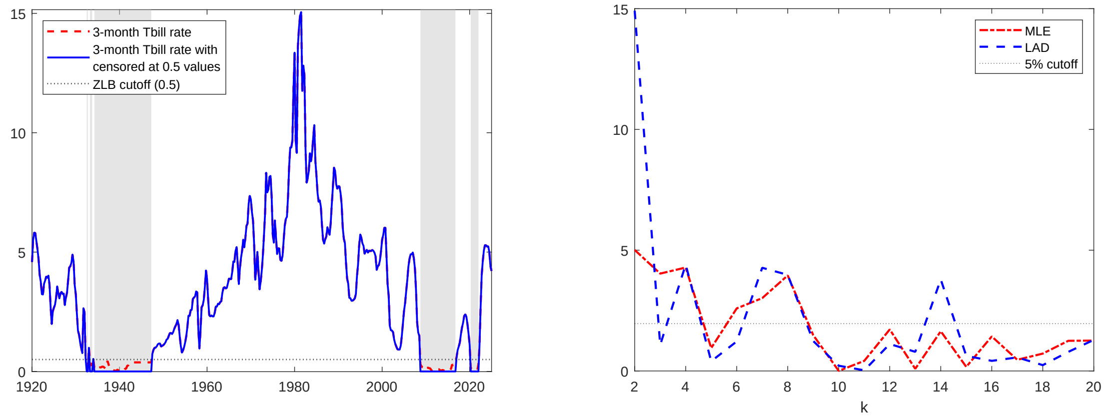
Figure 4.3: Asymptotic distributions of $\sqrt{T}(\hat{\phi}_1 - \phi_1)$ and of t-statistic for ϕ_1 based on three estimation procedures. Data generating process: $y_t = [y_{t-1} + 0.5\Delta y_{t-1} + u_t]_+$, $y_0 = y_{-1} = 0$, $u_t \sim$ i.i.d. $\mathcal{N}(0,1)$, $T = 1000$, $MC = 100,000$ replications.

the OLS asymptotic distribution of $\hat{\phi}_{1,OLS}$ and its corresponding t-statistic (solid blue curves) are not centered at zero, as a consequence of the censoring. In particular, the mean of the OLS t-statistic in Figure 4.3 is approximately -0.4. This negative mean represents the non-degenerate limit of $\frac{1}{\sqrt{T}}\sum_t y_t^- \Delta y_{t-1}$, where $y_t^- = \min\{0, \alpha_0 + \beta_0 y_{t-1} + \boldsymbol{\phi}_0^\mathsf{T} \Delta \boldsymbol{y}_{t-1} + u_t\}$. The intuition is that the positive-part binds $O(\sqrt{T})$ times, so there are $O(\sqrt{T})$ nonzero y_t^-, which, when scaled by $\frac{1}{\sqrt{T}}$, generate a negative bias. In contrast, the simulated distributions of the t-statistics for the ML (red dashed curve) and CLAD (black dash-dotted curve) estimators align closely with the standard normal density.

4.3 Model selection and illustrations

The fact that the t-statistics for ϕ_i are asymptotically standard normal, for both MLE and CLAD, can be used for the purposes of model selection. In particular, one can determine the appropriate lag order k for the model via a sequential test of the null $H_0 : \phi_{k-1} = 0$, starting from some relatively high value $k = k_0$, and then reducing k (by one) so long as the associated t-statistic does not exceed the α-level normal critical value. In contrast, since the OLS t-statistics for the ϕ_i parameters have a non-standard limiting distribution, the naive use of these in such a model selection procedure may lead to an inappropriate choice for k.

This is illustrated below with two data sets. For each we estimate a kth order dynamic Tobit, for various values of k, using MLE and CLAD. For each k, and each estimation method, we compute the corresponding t-statistic for $H_0 : \phi_{k-1} = 0$. The standard errors of the estimators are obtained from Theorems 3.1 and 3.2. The matrix Ω is estimated by its sample analogue, while the density at zero, required for the CLAD estimator, is estimated using a uniform kernel, as recommended by Powell (1984).



(a) 3-month Tbills and ZLB.
Gray area represents ZLB periods.

(b) Absolute values of t-statistics for ϕ_{k-1}.

Figure 4.4: 3-month Tbills and associated Tobit(k) t-statistics.

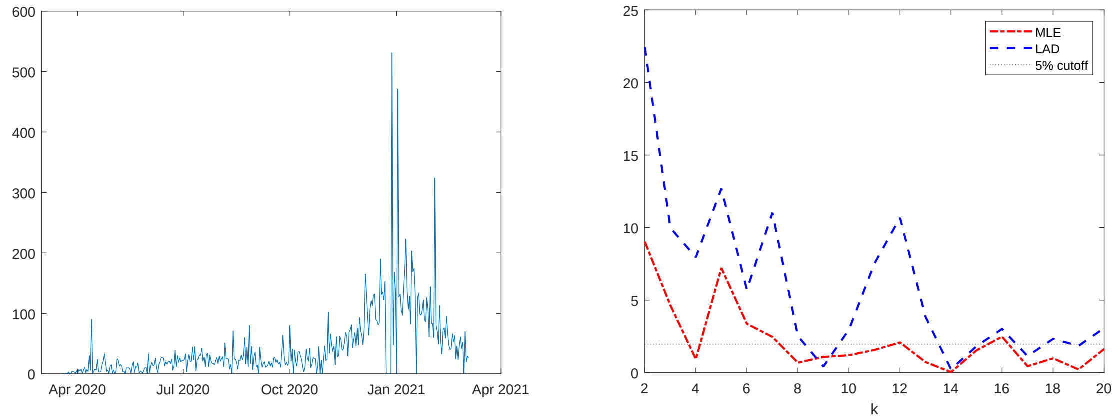
	α	β	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7
MLE	-0.25	1.03	0.39	-0.34	0.32	-0.10	0.13	-0.08	-0.22
			(6.98)	(5.76)	(5.14)	(1.50)	(2.18)	(1.35)	(3.95)
LAD	-0.00	1.00	0.43	-0.09	0.12	0.02	0.04	-0.04	-0.15
			(9.87)	(1.94)	(2.56)	(0.49)	(0.92)	(0.96)	(3.49)

Table 4.1: 3-month Tbills: MLE and LAD estimates with t-statistics for $\hat{\phi}_i$ coefficients.

4.3.1 US Treasury bill rate

Treasury bill rates in the United States have been near the zero lower bound (ZLB) during three historical periods: during the 1930s and 1940s following the Great Depression, during and after the 2007 Global Financial Crisis, and more recently during the COVID-19 pandemic. We use quarterly interest rates on 3-month Treasury bills, with data extending back to the 1920s. The data set is based on Ramey and Zubairy (2018) and extended using FRED data through the first quarter of 2025, so that $T = 421$. We treat values below 0.5 as effectively at the zero lower bound, so that our identified ZLB periods closely align with those in Ramey and Zubairy (2018). Figure 4.4a displays the interest rate data, while Figure 4.4b presents the corresponding t-statistics computed for $k \in \{2, \ldots, 20\}$ (i.e., up to 5 years of quarterly lags).

Comparing the t-statistics in Figure 4.4b with the 5% normal critical value for a two-sided test (1.96), we find that in most cases both estimation methods yield the same results, and point to the model with $k = 8$ as the preferred specification. There are, however, three values of k, 3, 6, and 14, where the methods yield slightly different conclusions, as the CLAD t-statistic falls below 1.96 for $k = 3, 6$ and above it for $k = 14$. We report the estimation results for $k = 8$ in Table 4.1.



(a) Covid-19 cases in Rowan county, NC. (b) Absolute values of t-statistics for ϕ_{k-1}.

Figure 4.5: Daily Covid-19 cases in Rowan county, NC and associated Tobit(k) t-statistics.

4.3.2 Countywide Covid-19 cases

Understanding and predicting the evolution of pandemic case counts is critical for effective public health planning and response. Accurate forecasts can inform timely policy decisions such as implementing social distancing measures, allocating medical resources, or guiding vaccine distribution. This stresses the importance of valid model selection. To illustrate our approach to pandemic model selection, we use Covid-19 case data for Rowan County, NC, obtained from USAFacts (accessed July 15, 2025). Rowan County provides an instructive example due to its moderate population size and relatively localised outbreaks.

We use daily data for 351 consecutive days, stopping before reporting irregularities become prevalent: when missing cases began to be accumulated into subsequent days. The sample begins on March 19, 2020 (the date of the first reported case) and ends on March 4, 2021. Unlike the previous example, we do not apply artificial censoring; the 22 zero observations correspond to days when no positive cases were recorded.

Figure 4.5 summarizes the data and the associated t-statistics. In contrast to the previous example, we now observe substantial disagreement between LAD and MLE, with LAD favoring more complex models. This divergence suggests that the data may not be well approximated by a normal distribution, consistent with findings in the literature showing that Covid-19 case counts deviate significantly from both normality and log-normality (e.g., Gonçalves, Turkman, Geraldes, Marques, and Sousa (2021); Cohen, Davis, and Samorodnitsky (2022)).

5 Conclusion

This paper develops asymptotic theory for maximum likelihood (ML) and censored least absolute deviations (CLAD) estimation in dynamic Tobit models under local unit root (LUR) asymptotics. While earlier work had established consistency (and asymptotic normality) of these estimators in the stationary setting, we show that both ML and CLAD remain consistent in the LUR regime and derive their asymptotic distributions. A key finding is that the short-run parameters are asymptotically normal under both methods, facilitatig standard inference and model selection via sequential t-testing.

These results contrast sharply with the behavior of ordinary least squares (OLS), which, although consistent under LUR, yields t-statistics for the short-run parameters that have an asymptotic bias and non-standard distribution. Consequently, model selection procedures based on OLS can be misleading, especially in settings with censoring and persistent dynamics.

Overall, our results suggest that MLE and LAD offer viable and robust alternatives to OLS in dynamic censored models, with significant advantages for inference and model selection. Future work could extend the analysis to accommodate time-varying volatility, covariates, or alternative forms of censoring commonly encountered in economic and financial data.

A Auxiliary lemmas

Notation. $e_{m,i}$ denotes the ith column of an $m \times m$ identity matrix; when m is clear from the context, we write this simply as e_i. For $S \subset \mathbb{R}^m$, let $L^{\mathrm{ucc}}(S)$ denote the set of functions that are uniformly bounded on compact subsets of S, equipped with the topology of uniform convergence on compacta. (When S is compact, this coincides with the uniform topology.) Let $\mathcal{F}_t := \sigma(\{y_s\}_{s\leq 0}, \{u_s\}_{s\leq t})$ denote the filtration generated by the initial conditions and the innovations $\{u_t\}$.

For ease of reference, we start by presenting some technical results. We will use the following two properties. First, let $y_t^- := [\alpha_{T,0}+\beta_{T,0}y_{t-1}+\boldsymbol{\phi}_0^{\mathsf{T}}\Delta\boldsymbol{y}_{t-1}+u_t]_-$ and $v_t := u_t - y_t^-$, where $[x]_- := \min\{x, 0\}$. Then the evolution of $\{y_t\}$ may also be described by

$$y_t = \alpha_{T,0} + \beta_{T,0}y_{t-1} + \boldsymbol{\phi}_0^{\mathsf{T}}\Delta\boldsymbol{y}_{t-1} + v_t = x_{t-1} + v_t, \tag{A.1}$$

where we have also defined

$$x_{t-1} := \alpha_{T,0} + \beta_{T,0}y_{t-1} + \boldsymbol{\phi}_0^{\mathsf{T}}\Delta\boldsymbol{y}_{t-1}. \tag{A.2}$$

Second, as noted in the proof of Bykhovskaya and Duffy (2024, Lemma B.2), since the polynomial $\phi(z) := 1 - \sum_{i=1}^k \phi_{i,0}z^i$ has all its roots outside the unit circle under A1–A4, it has a well defined inverse $\phi^{-1}(z) \coloneqq \sum_{i=0}^{\infty} \gamma_i z^i$ for all $|z| \leq 1$.

We also define the scaled regressor process:

$$z_{t-1,T} = \begin{bmatrix} T^{-1/2} \\ T^{-1} y_{t-1} \\ T^{-1/2} \Delta \boldsymbol{y}_{t-1}. \end{bmatrix} \tag{A.3}$$

Lemma A.1. *Suppose* A1–A4 *hold. Then*

(i) *there exists a* $C < \infty$ *such that* $\max_{-k+2 \leq t \leq T} \|\Delta y_t\|_{2+\delta_u} < C$;

(ii) $T^{-1} \sum_{t=1}^{T} (\Delta \boldsymbol{y}_{t-1})(\Delta \boldsymbol{y}_{t-1})^{\mathsf{T}} \xrightarrow{p} \Omega$, *where* $\Omega = [\Omega_{ij}]_{i,j=1}^{k-1}$ *is positive definite, with* $\Omega_{ij} = \sigma_0^2 \sum_{n=0}^{\infty} \gamma_n \gamma_{n+|i-j|}$;

(iii) $\sum_{t=1}^{T} y_{t-1} \Delta \boldsymbol{y}_{t-1} = O_p(T)$.

Lemma A.2. *Suppose* A1–A4 *hold,* $M > 0$, *and let* $g(x)$ *be a bounded function such that* $|g(x)| \to 0$ *as* $x \to +\infty$. *Let* $\{\mathcal{Y}_{t,T}\}_{t=1}^{T}$ *be such that* $T^{-1/2} \mathcal{Y}_{\lfloor \tau T \rfloor, T} \xrightarrow{d} Y(\tau)$, *where* $Y(\cdot)$ *is defined in (2.3). Then*

(i) $\int_0^1 \mathbf{1}\{Y(\tau) \leq M\} \, \mathrm{d}\tau \xrightarrow{p} 0$ *as* $M \to 0$;

(ii) $T^{-1} \sum_{t=1}^{T} \mathbf{1}\{\mathcal{Y}_{t-1,T} < M\} \xrightarrow{p} 0$ *as* $T \to \infty$ *and then* $M \to 0$; *and*

(iii) $\mathbb{E}|T^{-1} \sum_{t=1}^{T} g(\mathcal{Y}_{t-1,T})| = o(1)$.

In particular, the preceding holds with $\mathcal{Y}_{t-1,T} = \alpha_{T,0} + \beta_{T,0} y_{t-1} + \boldsymbol{\phi}_0^{\mathsf{T}} \Delta \boldsymbol{y}_{t-1}$.

Lemma A.3. *Suppose* A1–A4 *hold. Then*

(i) $\sum_{t=1}^{T} z_{t-1,T} z_{t-1,T}^{\mathsf{T}} \xrightarrow{d} Q_{ZZ}$ *jointly with* $\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor \tau T \rfloor} u_t \xrightarrow{d} \sigma_0 W(\tau)$, *where*

$$Q_{ZZ} \coloneqq \begin{bmatrix} 1 & \int_0^1 Y(\tau) \, \mathrm{d}\tau & 0 \\ \int_0^1 Y(\tau) \, \mathrm{d}\tau & \int_0^1 Y^2(\tau) \, \mathrm{d}\tau & 0 \\ 0 & 0 & \Omega \end{bmatrix}$$

is a.s. positive definite;

(ii) $\max_{0 \leq t \leq T} \|z_{t,T}\| = O_p(T^{-\delta_u/2(2+\delta_u)}) = o_p(1)$;

(iii) $T^{-1} \sum_{t=1}^{T} \mathbf{1}\{y_t = 0\} = o_p(1)$; *and*

(iv) $\sum_{t=1}^{T} z_{t-1,T} u_t \mathbf{1}\{y_t > 0\} = o_p(T^{1/2})$

Suppose further that A3 *holds with* $\delta_u > 2$. *Then*

(v) $\sum_{t=1}^{T} \|z_{t-1,T}\|^4 = O_p(T^{-1})$;

(vi) $\sum_{t=1}^{T}\mathbf{1}\{y_t=0\}|v_t|^m = o_p(T)$ *for each* $m\in[1,4]$.

Lemma A.4. *Let* $\Phi(x),\varphi(x)$ *denote the standard normal cdf and pdf functions and let* $\lambda(x) \coloneqq \varphi(x)/[1-\Phi(x)]$ *be the inverse Mills ratio. There exists a* $C<\infty$ *such that* $\lambda(x)\leq C(1+[x]_+)$ *and* $|\lambda'(x)|\leq C$ *for all* $x\in\mathbb{R}$.

Proof of Lemma A.1. (i), (ii) and (iii) correspond to Lemma B.2 and parts (iii) and (iv) of Lemma B.4 in Bykhovskaya and Duffy (2024).

Regarding the positive definiteness of Ω asserted in part (i), by defining the stationary autoregressive process

$$w_t = \sum_{i=1}^{k}\phi_i w_{t-i} + u_t$$

and setting $\boldsymbol{w}_t \coloneqq (w_t,\ldots,w_{t-k+1})^\mathsf{T}$, we obtain that $\Omega = \mathbb{E}\boldsymbol{w}_t\boldsymbol{w}_t^\mathsf{T}$ is necessarily positive semi-definite. Further, by the independence of u_t from $\{w_\tau\}_{\tau<t}$, we have for $\boldsymbol{a} = (a_1,\ldots,a_k)^\mathsf{T}$ that $\mathbb{V}(\boldsymbol{a}^\mathsf{T}\boldsymbol{w}_t)\geq\mathbb{V}(a_1u_t) = a_1^2\sigma_0^2$, and so is nonzero so long as $a_1\neq 0$. If $a_1=0$, then we similarly get $\mathbb{V}(\boldsymbol{a}^\mathsf{T}\boldsymbol{w}_t)\geq\mathbb{V}(a_2u_{t-1}) = a_2^2\sigma_0^2$, which is nonzero as long as $a_2\neq 0$. Repeating the same argument, it follows that there does not exist any $\boldsymbol{a}\in\mathbb{R}^k\backslash\{0\}$ such that $\boldsymbol{a}^\mathsf{T}\Omega\boldsymbol{a} = \mathbb{V}(\boldsymbol{a}^\mathsf{T}\boldsymbol{w}_t) = 0$. □

Proof of Lemma A.2. **(i).** Let $c_\phi \coloneqq \phi(1)^{-1}c_0$ and $Y^*(\tau)\coloneqq\phi(1)\mathrm{e}^{-c_\phi\tau}Y(\tau)$, where $\phi(1)>0$ as a consequence of A4, as noted in Bykhovskaya and Duffy (2024, Remark 3.1). Then

$$\begin{aligned}
Y^*(\tau) &= K(\tau) + \sup_{\tau'\leq\tau}[-K(\tau')]_+,\\
K(\tau) &= \phi(1)b_0 + a_0\int_0^\tau e^{-c_\phi r}dr + \sigma_0\int_0^\tau e^{-c_\phi r}dW(r).
\end{aligned}$$

Now consider

$$\begin{aligned}
\int_0^1 \mathbf{1}\{Y(\tau)\leq M\}\,\mathrm{d}\tau &= \int_0^1 \mathbf{1}\{Y^*(\tau)\leq M\phi(1)\mathrm{e}^{-c_\phi\tau}\}\,\mathrm{d}\tau\\
&\leq \int_0^1 \mathbf{1}\{Y^*(\tau)< 2M\phi(1)(1+\mathrm{e}^{-c_\phi})\}\,\mathrm{d}\tau\\
&= \int_0^1 \mathbf{1}\{0\leq Y^*(\tau)< 2M\phi(1)(1+\mathrm{e}^{-c_\phi})\}\,\mathrm{d}\tau
\end{aligned}$$

since $Y^*(\tau)\geq 0$ for all τ. It remains to show that the r.h.s. vanishes as $M\to 0$. To that end, observe that $K(\cdot)$ is a continuous semimartingale, and that the (weakly) increasing process $\sup_{\tau'\leq\tau}[-K(\tau')]_+$ is continuous and has finite variation. Hence $Y^*(\tau)$ is also a continuous semimartingale, with quadratic variation

$$[Y^*(\tau)] = [K(\tau)] = \sigma_0^2\int_0^\tau \mathrm{e}^{-2c_\phi s}\,\mathrm{d}(s) = \frac{\sigma_0^2(1-\mathrm{e}^{-2c_\phi\tau})}{2c_\phi} \coloneqq \mu^*(\tau)$$

and $\mu^*(\tau) = \sigma_0^2\tau$ when $c_0 = 0$. It follows by Corollary VI.1.9 of Revuz and Yor (1999) that, almost surely

$$\int_0^1 \mathbf{1}\{0 \leq Y^*(\tau) < M'\}\,\mathrm{d}\mu^*(\tau) \to 0$$

as $M' \to 0$. Since μ^* and Lebesgue measure on $[0,1]$ are mutually absolutely continuous, it follows that

$$\int_0^1 \mathbf{1}\{0 \leq Y^*(\tau) < M'\}\,\mathrm{d}\tau \to 0$$

as $M' \to 0$, as required.

(ii). Letting $h_M(x)$ be any smooth function such that $\mathbf{1}\{x < M\} \leq h_M(x) \leq \mathbf{1}\{x < 2M\}$, we have by the continuous mapping theorem (CMT) and the result of part (i) that

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}\mathbf{1}\{T^{-1/2}\mathcal{Y}_{t-1,T} < M\} &\leq T^{-1}\sum_{t=1}^{T} h_M(T^{-1/2}\mathcal{Y}_{t-1,T}) \\ &\xrightarrow{d} \int_0^1 h_M[Y(\tau)]\,\mathrm{d}\tau \leq \int_0^1 \mathbf{1}\{Y(\tau) < 2M\}\,\mathrm{d}\tau \xrightarrow{p} 0 \end{aligned}$$

as $T \to \infty$ and then $M \to 0$.

(iii). Let $\epsilon > 0$ and $M > 0$: then for all T sufficiently large, $|g(x)| < \epsilon$ for all $x \geq T^{1/2}M$. Since there exists a $C < \infty$ such that $|g(x)| \leq C$ for all $x \in \mathbb{R}$, it follows that for such T,

$$\left|T^{-1}\sum_{t=1}^{T} g(\mathcal{Y}_{t-1,T})\right| \leq CT^{-1}\sum_{t=1}^{T}\mathbf{1}\{T^{-1/2}\mathcal{Y}_{t-1,T} < M\} + \epsilon. \tag{A.4}$$

Deduce, by the result of part (i), that the r.h.s. of (A.4) is bounded by 2ϵ w.p.a.1 as $T \to \infty$ and then $M \to 0$; since ϵ was arbitrary, it follows that $T^{-1}\sum_{t=1}^{T} g(\mathcal{Y}_{t-1,T}) = o_p(1)$. Since g is bounded, this holds also in L^1.

Regarding the final claim, we need only to note that in this case

$$T^{-1/2}\mathcal{Y}_{\lfloor\tau T\rfloor,T} = T^{-1/2}\alpha_{T,0} + T^{-1/2}\beta_{T,0}y_{\lfloor\tau T\rfloor} + T^{-1/2}\boldsymbol{\phi}_0^{\mathsf{T}}\Delta\boldsymbol{y}_{\lfloor\tau T\rfloor} \xrightarrow{d} Y(\tau)$$

by Lemma A.1(i) and Theorem 3.2 in Bykhovskaya and Duffy (2024). □

Proof of Lemma A.3. **(i).** We have

$$\begin{aligned} \sum_{t=1}^{T} z_{t-1,T}z_{t-1,T}^{\mathsf{T}} &= \begin{bmatrix} 1 & T^{-3/2}\sum_{t=1}^{T} y_{t-1} & T^{-1}\sum_{t=1}^{T}(\Delta\boldsymbol{y}_{t-1})^{\mathsf{T}} \\ T^{-3/2}\sum_{t=1}^{T} y_{t-1} & T^{-2}\sum_{t=1}^{T} y_{t-1}^2 & T^{-3/2}\sum_{t=1}^{T} y_{t-1}(\Delta\boldsymbol{y}_{t-1})^{\mathsf{T}} \\ T^{-1}\sum_{t=1}^{T}\Delta\boldsymbol{y}_{t-1} & T^{-3/2}\sum_{t=1}^{T}(\Delta\boldsymbol{y}_{t-1})y_{t-1} & T^{-1}\sum_{t=1}^{T}(\Delta\boldsymbol{y}_{t-1})(\Delta\boldsymbol{y}_{t-1})^{\mathsf{T}} \end{bmatrix} \\ &\xrightarrow{d} \begin{bmatrix} 1 & \int_0^1 Y(\tau)\,\mathrm{d}\tau & 0 \\ \int_0^1 Y(\tau)\,\mathrm{d}\tau & \int_0^1 Y^2(\tau)\,\mathrm{d}\tau & 0 \\ 0 & 0 & \Omega \end{bmatrix} \end{aligned}$$

by parts (ii) and (iii) of Lemma A.1, Theorem 3.2 in Bykhovskaya and Duffy (2024) and the CMT, noting in particular that $\sum_{t=1}^{T} \Delta \boldsymbol{y}_{t-1} = \boldsymbol{y}_{T-1} - \boldsymbol{y}_0 = O_p(T^{1/2})$. The a.s. positive definiteness of the final matrix follows since Ω is positive definite, by Lemma A.3(ii).

(ii). We have

$$\begin{aligned}\max_{1\le t\le T} \|z_{t,T}\|^2 &\le T^{-1} + T^{-1} \max_{1\le t\le T} (T^{-1/2} y_{t-1})^2 + T^{-1} \max_{1\le t\le T} \|\Delta \boldsymbol{y}_t\|^2 \\ &= O_p(T^{-1}) + O_p(T^{-1+2/(2+\delta_u)}) = O_p(T^{-\delta_u/(2+\delta_u)})\end{aligned}$$

by Theorem 3.2 in Bykhovskaya and Duffy (2024), and Lemma A.1(i).

(iii). Let $\delta > 0$, and $h_\delta(x)$ be any smooth function such that $\mathbf{1}\{0 \le x \le \delta\} \le h_\delta(x) \le \mathbf{1}\{0 \le x \le 2\delta\}$. Then by the CMT and Lemma A.2(i),

$$T^{-1} \sum_{t=1}^{T} \mathbf{1}\{y_t = 0\} \le T^{-1} \sum_{t=1}^{T} h_\delta(T^{-1/2} y_t) \xrightarrow{d} \int_0^1 h_\delta[Y(\tau)] \,\mathrm{d}\tau \le \int_0^1 \mathbf{1}\{0 \le Y(\tau) \le 2\delta\} \,\mathrm{d}\tau \xrightarrow{p} 0$$

as $T \to \infty$ and then $\delta \to 0$.

(iv). Decompose

$$\sum_{t=1}^{T} z_{t-1,T} u_t \mathbf{1}\{y_t > 0\} = \sum_{t=1}^{T} z_{t-1,T} u_t - \sum_{t=1}^{T} z_{t-1,T} u_t \mathbf{1}\{y_t = 0\}.$$

For the second r.h.s. term, two applications of the CS inequality yield

$$\begin{aligned}\sum_{t=1}^{T} \|z_{t-1,T}\| |u_t| \mathbf{1}\{y_t = 0\} &\le \left(\sum_{t=1}^{T} \|z_{t-1,T}\|^2 \right)^{1/2} \left(\sum_{t=1}^{T} |u_t|^4 \right)^{1/4} \left(\sum_{t=1}^{T} \mathbf{1}\{y_t = 0\} \right)^{1/4} \\ &= O_p(1) \cdot O_p(T^{1/4}) \cdot o_p(T^{1/4}) = o_p(T^{1/2})\end{aligned}$$

by the results of parts (i) and (iii). Next, note

$$\sum_{t=1}^{T} z_{t-1,T} u_t = \sum_{t=1}^{T} \begin{bmatrix} T^{-1/2} \\ T^{-1} y_{t-1} \\ T^{-1/2} \Delta \boldsymbol{y}_{t-1} \end{bmatrix} u_t.$$

By Theorem 3.2 in Bykhovskaya and Duffy (2024) and Theorem 2.1 in Liang, Phillips, Wang, and Wang (2016), the second component converges weakly to $\sigma \int_0^1 Y(\tau) \,\mathrm{d}W(\tau)$, and so is $O_p(1)$. The remaining components form a martingale with variance matrix

$$\sigma_0^2 \begin{bmatrix} 1 & 0 \\ 0 & \mathbb{E}(\Delta \boldsymbol{y}_{t-1})(\Delta \boldsymbol{y}_{t-1}^{\mathsf{T}}) \end{bmatrix}$$

which is uniformly bounded by Lemma A.1(i); hence these components are also $O_p(1)$.

(v). Note that there exists a $C < \infty$ (depending only on n) such that for all $x = (x_1, \dots, x_n)^{\mathsf{T}} \in \mathbb{R}^n$, $\|x\|^4 = (\sum_{i=1}^n x_i^2)^2 \le C \sum_{i=1}^n x_i^4$. Therefore, in particular

$$
\begin{aligned}
\sum_{t=1}^{T} \|z_{t-1,T}\|^4 &\le C \sum_{t=1}^{T} \left[T^{-2} + T^{-4}|y_{t-1}|^4 + T^{-2}\|\Delta \boldsymbol{y}_{t-1}\|^4\right] \\
&= CT^{-1}\left[1 + T^{-1}\sum_{t=1}^{T}\left|\frac{y_{t-1}}{T^{1/2}}\right|^4 + T^{-1}\sum_{t=1}^{T}\|\Delta \boldsymbol{y}_{t-1}\|^4\right] \\
&= O_p(T^{-1})
\end{aligned}
$$

since $\mathbb{E}\|\Delta \boldsymbol{y}_{t-1}\|^4$ is uniformly bounded by Lemma A.1(i), and

$$
T^{-1}\sum_{t=1}^{T}\left|\frac{y_{t-1}}{T^{1/2}}\right|^4 \xrightarrow{d} \int_0^1 Y^4(\tau)\,\mathrm{d}\tau
$$

by Theorem 3.2 in Bykhovskaya and Duffy (2024) and the CMT.

(vi). Note that we always have

$$
v_t = u_t - y_t^- \ge u_t
$$

since $-y_t^- \ge 0$, while if $y_t = 0$, then by (A.1)

$$
0 = y_t = \alpha_{0,T} + \beta_{0,T} y_{t-1} + \boldsymbol{\phi}_0^{\mathsf{T}} \Delta \boldsymbol{y}_{t-1} + v_t \ge \alpha_{0,T} + \boldsymbol{\phi}_0^{\mathsf{T}} \Delta \boldsymbol{y}_{t-1} + v_t.
$$

Thus $u_t \le v_t \le -\alpha_{0,T} - \boldsymbol{\phi}_0^{\mathsf{T}} \Delta \boldsymbol{y}_{t-1}$ when $y_t = 0$, whence there exists a $C < \infty$ (depending only on m, a_0 and $\|\boldsymbol{\phi}_0\|$) such that

$$
\begin{aligned}
\mathbb{E}\mathbf{1}\{y_t = 0\}|v_t|^{m+\delta} &\le \mathbb{E}(|u_t| + |\alpha_{0,T}| + \|\boldsymbol{\phi}_0\|\|\Delta \boldsymbol{y}_{t-1}\|)^{m+\delta} \\
&\le C(1 + \mathbb{E}|u_t|^{m+\delta} + \mathbb{E}\|\Delta \boldsymbol{y}_{t-1}\|^{m+\delta}).
\end{aligned}
$$

In view of Lemma A.1(i), we may take $\delta \in (0, \delta_u - 2)$ such that the r.h.s. is uniformly bounded, for all $m \in [1, 4]$. Hence, by Hölder's inequality,

$$
\begin{aligned}
\sum_{t=1}^{T} \mathbf{1}\{y_t = 0\}|v_t|^m &\le \left(\sum_{t=1}^{T} \mathbf{1}\{y_t = 0\}\right)^{\delta/(m+\delta)} \left(\sum_{t=1}^{T} \mathbf{1}\{y_t = 0\}|v_t|^{m+\delta}\right)^{m/(m+\delta)} \\
&= o_p(T^{\delta/(m+\delta)}) \cdot O_p(T^{m/(m+\delta)}) = o_p(T)
\end{aligned}
$$

by the preceding and the result of part (iii). □

Proof of Lemma A.4. Both bounds are an immediate consequence of the bound given in (3) in Sampford (1953). □

B Maximum likelihood

To economise on notation, in this section we shall generally write $\hat{\alpha}_T^{\mathsf{M}}$ as merely $\hat{\alpha}_T$. Define

$$\rho \coloneqq (\alpha, \beta, \boldsymbol{\phi}^{\mathsf{T}})^{\mathsf{T}}, \qquad \vartheta = \sigma^{-1}, \qquad \rho_{T,0} \coloneqq (\alpha_{T,0}, \beta_{T,0}, \boldsymbol{\phi}_0^{\mathsf{T}})^{\mathsf{T}}, \qquad \vartheta_0 = \sigma_0^{-1},$$
$$\pi \coloneqq (\rho^{\mathsf{T}}, \vartheta)^{\mathsf{T}}, \qquad \pi_{T,0} \coloneqq (\rho_{T,0}^{\mathsf{T}}, \vartheta_0)^{\mathsf{T}},$$

where the '0' subscript denotes the true parameter values.

It will be convenient to rewrite the estimation problem in terms of local parameters, defined as:

$$r \coloneqq \begin{bmatrix} a \\ c \\ \boldsymbol{f} \end{bmatrix} = \begin{bmatrix} T^{1/2} & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T^{1/2} \end{bmatrix} \begin{bmatrix} \alpha - \alpha_{T,0} \\ \beta - \beta_{T,0} \\ \boldsymbol{\phi} - \boldsymbol{\phi}_0 \end{bmatrix} \eqqcolon D_{r,T}(\rho - \rho_{T,0}), \qquad h \coloneqq T^{1/2}(\vartheta - \vartheta_0),$$

which we collect together as

$$p \coloneqq \begin{bmatrix} r \\ h \end{bmatrix} = \begin{bmatrix} D_{r,T} & 0 \\ 0 & T^{1/2} \end{bmatrix} \begin{bmatrix} \rho - \rho_{T,0} \\ \vartheta - \vartheta_0 \end{bmatrix} \eqqcolon D_{p,T}(\pi - \pi_{T,0}). \tag{B.1}$$

For each $T \in \mathbb{N}$, there is a bijective mapping between p and π. To reparametrise the model in terms of the former, let

$$\pi_T(p) = \pi_{T,0} + D_{p,T}^{-1} p \qquad \rho_T(r) = \rho_{T,0} + D_{r,T}^{-1} r \qquad \vartheta_T(h) = \vartheta_0 + T^{-1/2} h,$$

denote the model parameters corresponding to given values of $p = (r^{\mathsf{T}}, h)^{\mathsf{T}}$. Since the only constraint on the original parameter space is that $\vartheta > 0$, the parameter space for p is given by $\mathcal{P}_T \coloneqq \mathbb{R}^{k+1} \times \{h \in \mathbb{R} \mid h > -T^{1/2}\vartheta_0\}$; note that for any compact $K \subset \mathbb{R}^{k+2}$, we will have $K \subset \mathcal{P}_T$ for all T sufficiently large.

Recalling (A.1)–(A.3) above, we thus have

$$\begin{aligned} y_t - \alpha - \beta y_{t-1} - \boldsymbol{\phi}^{\mathsf{T}} \Delta \boldsymbol{y}_{t-1} &= v_t + (\alpha_0 - \alpha) + (\beta_0 - \beta) y_{t-1} + (\boldsymbol{\phi}_0 - \boldsymbol{\phi})^{\mathsf{T}} \Delta \boldsymbol{y}_{t-1} \\ &= v_t - (\rho - \rho_0)^{\mathsf{T}} D_{r,T} z_{t-1,T} = v_t - z_{t-1,T}^{\mathsf{T}} r, \end{aligned} \tag{B.2}$$

for $v_t = u_t - y_t^-$, and

$$\begin{aligned} \alpha + \beta y_{t-1} + \boldsymbol{\phi}^{\mathsf{T}} \Delta \boldsymbol{y}_{t-1} &= [\alpha_0 + \beta_0 y_{t-1} + \boldsymbol{\phi}_0^{\mathsf{T}} \Delta \boldsymbol{y}_{t-1}] \\ &\qquad + [(\alpha - \alpha_0) + (\beta - \beta_0) y_{t-1} + (\boldsymbol{\phi} - \boldsymbol{\phi}_0)^{\mathsf{T}} \Delta \boldsymbol{y}_{t-1}] \\ &= x_{t-1} + z_{t-1,T}^{\mathsf{T}} r. \end{aligned} \tag{B.3}$$

B.1 Proof of Theorem 3.1

We may now proceed to analyse the asymptotic behaviour of the loglikelihood. To that end, let

$$g_t(p) := \log f_{\pi_T(p)}(\mathsf{y}_t \mid \mathsf{y}_{t-1}, \ldots, \mathsf{y}_{t-k})$$

denote the conditional density of y_t given $(y_{t-1}, \ldots, y_{t-k})$, when $\pi = \pi_T(p)$. Then in view of (3.1), (B.2) and (B.3), evaluating this at $\mathsf{y}_{t-i} = y_{t-i}$ yields

$$\begin{aligned} g_t(p) &= \mathbf{1}\{y_t > 0\} \log \vartheta_T(h)\varphi[\vartheta_T(h)(y_t - \alpha_T(r) - \beta_T(r)y_{t-1} - \boldsymbol{\phi}_T(r)^{\mathsf{T}}\Delta\boldsymbol{y}_{t-1})] \\ &\qquad + \mathbf{1}\{y_t = 0\} \log\{1 - \Phi[\vartheta_T(h)(\alpha_T(r) + \beta_T(r)y_{t-1} + \boldsymbol{\phi}_T(r)^{\mathsf{T}}\Delta\boldsymbol{y}_{t-1})]\} \\ &= \mathbf{1}\{y_t > 0\} \log \vartheta_T(h)\varphi[\vartheta_T(h)(v_t - z_{t-1,T}^{\mathsf{T}} r)] \\ &\qquad + \mathbf{1}\{y_t = 0\} \log\{1 - \Phi[\vartheta_T(h)(x_{t-1} + z_{t-1,T}^{\mathsf{T}} r)]\} \\ &= [\log \vartheta_T(h) - \tfrac{1}{2}\log 2\pi]\mathbf{1}\{y_t > 0\} - \tfrac{1}{2}\vartheta_T^2(h)(u_t - z_{t-1,T}^{\mathsf{T}} r)^2 \mathbf{1}\{y_t > 0\} \\ &\qquad + \mathbf{1}\{y_t = 0\} \log\{1 - \Phi[\vartheta_T(h)(z_{t-1,T}^{\mathsf{T}} r - v_t)]\}, \end{aligned}$$

since when $y_t > 0$, $y_t^- = 0$, so that $v_t = u_t$, and when $y_t = 0$,

$$0 = y_t = x_{t-1} + v_t$$

so that $v_t = -x_{t-1}$. If $\hat{\pi}$ maximises $\mathcal{L}_T(\pi)$, then $\hat{p} = D_{p,T}(\hat{\pi} - \pi_{T,0})$ maximises

$$\begin{aligned} \ell_T(p) := \mathcal{L}_T[\pi_T(p)] &= \sum_{t=1}^{T} g_t(p) \\ &= N_T[\log \vartheta_T(h) - \tfrac{1}{2}\log 2\pi] - \tfrac{1}{2}\vartheta_T^2(h)\sum_{t=1}^{T}(u_t - z_{t-1,T}^{\mathsf{T}} r)^2 \mathbf{1}\{y_t > 0\} \\ &\qquad + \sum_{t=1}^{T} \mathbf{1}\{y_t = 0\} \log\{1 - \Phi[\vartheta_T(h)(z_{t-1,T}^{\mathsf{T}} r - v_t)]\} \end{aligned}$$

where $N_T := \sum_{t=1}^{T} \mathbf{1}\{y_t > 0\}$.

Let $\mathcal{S}_T(p)$ and $\mathcal{H}_T(p)$ denote the score (gradient) and Hessian of ℓ_T at p. The proof of the following is deferred to Appendix B.2.

Proposition B.1. *Suppose* A1–A4 *and* B *hold. Then jointly with* $\frac{1}{T}\sum_{t=1}^{\lfloor \tau T \rfloor} u_t \xrightarrow{d} \sigma_0 W(\tau)$,

(i) $\mathcal{S}_T(0) \xrightarrow{d} \mathcal{S}$, *and*

(ii) $\mathcal{H}_T(p) \xrightarrow{d} \mathcal{H}$ *on* $L^{\mathrm{ucc}}(\mathbb{R}^{k+2})$

where

$$\mathcal{S} \coloneqq \sigma_0^{-1} \begin{bmatrix} W(1) \\ \int_0^1 Y(\tau)\,\mathrm{d}W(\tau) \\ \Omega^{1/2}\xi_{(1)} \\ 2^{1/2}\sigma_0^2\xi_{(2)} \end{bmatrix} \qquad \mathcal{H} \coloneqq -\sigma_0^{-2} \begin{bmatrix} 1 & \int_0^1 Y(\tau)\,\mathrm{d}\tau & 0 & 0 \\ \int_0^1 Y(\tau)\,\mathrm{d}\tau & \int_0^1 Y^2(\tau)\,\mathrm{d}\tau & 0 & 0 \\ 0 & 0 & \Omega & 0 \\ 0 & 0 & 0 & 2\sigma_0^4 \end{bmatrix}$$

with $\xi_{(1)} \sim N[0, I_k]$, $\xi_{(2)} \sim N[0,1]$ *and* $W(\cdot)$ *being mutually independent.*

Since ℓ_T is twice continuously differentiable on $\mathcal{P}_T$, we have by Taylor's theorem that for every $p \in \mathcal{P}_T$, there exists a $\delta \in [0,1]$ such that

$$\begin{aligned} \ell_T(p) - \ell_T(0) &= \mathcal{S}_T(0)^{\mathsf{T}} p + \tfrac{1}{2} p^{\mathsf{T}} \mathcal{H}_T(\delta p) p \\ &= \mathcal{S}_T(0)^{\mathsf{T}} p + \tfrac{1}{2} p^{\mathsf{T}} \mathcal{H}_T(0) p + p^{\mathsf{T}} R_T(\delta p) p \end{aligned} \tag{B.4}$$

where $R_T(p) \coloneqq \mathcal{H}_T(p) - \mathcal{H}_T(0)$. By Proposition B.1, $R_T(p) \xrightarrow{p} 0$ on $L^{\mathrm{ucc}}(\mathbb{R}^{k+2})$, and hence

$$\ell_T(p) - \ell_T(0) \xrightarrow{d} \mathcal{S}^{\mathsf{T}} p + \tfrac{1}{2} p^{\mathsf{T}} \mathcal{H} p$$

on $L^{\mathrm{ucc}}(\mathbb{R}^{k+2})$. Since $\mathcal{H}$ is a.s. negative definite, the r.h.s. almost surely has an unique maximiser at $p^* \coloneqq -\mathcal{H}^{-1}\mathcal{S}$. With the aid of a convex reparametrisation of ℓ_T, we then have the following, whose proof appears in Appendix B.3

Proposition B.2. *If* $\hat{p}_T = (\hat{r}_T, \hat{h}_T) \in \operatorname{argmax}_{p \in \mathcal{P}_T} \ell_T(p)$ *for all* T, *then* $\hat{p}_T \xrightarrow{d} p^*$.

Since we may evidently take $\hat{p}_T = D_{p,T}(\hat{\pi}_T - \pi_{T,0})$ in the preceding, it follows that

$$D_{p,T}(\hat{\pi}_T - \pi_{T,0}) \xrightarrow{d} -\mathcal{H}^{-1}\mathcal{S}. \qquad \square$$

B.2 Proof of Proposition B.1

Partition the first and second partial derivatives of $g_t(p)$ as

$$\nabla_p g_t(p) = \begin{bmatrix} \nabla_r \\ \nabla_h \end{bmatrix} g_t(p) =: \begin{bmatrix} S_{r,t}(p) \\ S_{h,t}(p) \end{bmatrix} =: S_t(p),$$

$$\nabla_{pp} g_t(p) = \begin{bmatrix} \nabla_{rr} & \nabla_{hr} \\ \nabla_{rh} & \nabla_{hh} \end{bmatrix} g_t(p) = \begin{bmatrix} H_{rr,t}(p) & H_{hr,t}(p) \\ H_{rh,t}(p) & H_{hh,t}(p) \end{bmatrix} =: H_{pp,t}(p).$$

Recalling the definition of the inverse Mills ratio (see the text preceding the statement of Lemma A.4), differentiation and straightforward algebra then yields

$$S_{r,t}(p) = \vartheta_T \left[\vartheta_T (u_t - z_{t-1,T}^{\mathsf{T}} r) \mathbf{1}\{y_t > 0\} - \lambda[\vartheta_T (z_{t-1,T}^{\mathsf{T}} r - v_t)] \mathbf{1}\{y_t = 0\} \right] z_{t-1,T}, \tag{B.5}$$

where we have suppressed the argument of $\vartheta_T(h)$ to reduce notational clutter, and

$$H_{rr,t}(p) = -\vartheta_T^2 \left[\mathbf{1}\{y_t > 0\} + \lambda'[\vartheta_T(z_{t-1,T}^\mathsf{T} r - v_t)]\mathbf{1}\{y_t = 0\}\right] z_{t-1,T} z_{t-1,T}^\mathsf{T}. \tag{B.6}$$

Similarly, differentiating with respect to h yields

$$\begin{aligned} S_{h,t}(p) = T^{-1/2}\{[\vartheta_T^{-1} - \vartheta_T(u_t - z_{t-1,T}^\mathsf{T} r)^2]\mathbf{1}\{y_t > 0\} \\ - (z_{t-1,T}^\mathsf{T} r - v_t)\lambda[\vartheta_T(z_{t-1,T}^\mathsf{T} r - v_t)]\mathbf{1}\{y_t = 0\}\} \end{aligned} \tag{B.7}$$

and

$$\begin{aligned} H_{hh,t}(p) = T^{-1}\{[-\vartheta_T^{-2} - (u_t - z_{t-1,T}^\mathsf{T} r)^2]\mathbf{1}\{y_t > 0\} \\ - (z_{t-1,T}^\mathsf{T} r - v_t)^2\lambda'[\vartheta_T(z_{t-1,T}^\mathsf{T} r - v_t)]\mathbf{1}\{y_t = 0\}\}. \end{aligned} \tag{B.8}$$

Finally,

$$\begin{aligned} H_{rh,t}(p) &= T^{-1/2}\vartheta_T^{-1} S_{r,t}(p) \\ &\quad + T^{-1/2}\vartheta_T z_{t-1,T}[(u_t - z_{t-1,T}^\mathsf{T} r)\mathbf{1}\{y_t > 0\} \\ &\qquad - (z_{t-1,T}^\mathsf{T} r - v_t)\lambda'[\vartheta_T(z_{t-1,T}^\mathsf{T} r - v_t)]\mathbf{1}\{y_t = 0\}]. \end{aligned} \tag{B.9}$$

B.2.1 Proof of part (i)

By construction, the process

$$\mathcal{S}_T(0) := \begin{bmatrix} \mathcal{S}_{r,T}(0) \\ \mathcal{S}_{h,T}(0) \end{bmatrix} := \sum_{t=1}^T \begin{bmatrix} S_{r,t}(0) \\ S_{h,t}(0) \end{bmatrix}$$

is a martingale (see e.g. Hall and Heyde, 1980, Ch. 6). It follows from evaluating (B.5) at $p = 0$ (so that $\vartheta = \vartheta_0$), in particular, that

$$\mathcal{S}_{r,T}(0) = \sum_{t=1}^T z_{t-1,T} m_t,$$

where

$$\begin{aligned} m_t &:= \vartheta_0\left[\vartheta_0 u_t \mathbf{1}\{y_t > 0\} - \lambda[\vartheta_0(-v_t)]\mathbf{1}\{y_t = 0\}\right] \\ &= \vartheta_0\left[\vartheta_0 u_t \mathbf{1}\{y_t > 0\} - \lambda(\vartheta_0 x_{t-1})\mathbf{1}\{y_t = 0\}\right] \end{aligned}$$

is a martingale difference sequence. We first show the following.

Lemma B.3. *Suppose* A1–A4 *and* B *hold. Then*

$$\mathcal{S}_{r,T}(0) = \vartheta_0^2 \sum_{t=1}^{T} z_{t-1,T} u_t + o_p(1) =: \mathcal{M}_{r,T} + o_p(1)$$

Proof. Consider

$$\begin{aligned}\vartheta_0^2 \mathbb{E}_{t-1} u_t^2 \mathbf{1}\{y_t > 0\} &= \vartheta_0^2 \mathbb{E}_{t-1} u_t^2 \mathbf{1}\{x_{t-1} + u_t > 0\} = \mathbb{E}_{t-1}(\vartheta_0 u_t)^2 \mathbf{1}\{\vartheta_0 u_t > -\vartheta_0 x_{t-1}\} \\ &= \int_{-\vartheta_0 x_{t-1}}^{\infty} u^2 \varphi(u)\,\mathrm{d}u = -\vartheta_0 x_{t-1}\varphi(\vartheta_0 x_{t-1}) + \Phi(\vartheta_0 x_{t-1}) \qquad \text{(B.10)}\end{aligned}$$

since $\vartheta_0 u_t \sim N[0,1]$ under B, and because

$$\begin{aligned}\int_{-z}^{\infty} u^2 \varphi(u)\,\mathrm{d}u = -\int_{-z}^{\infty} u\varphi'(u)\,\mathrm{d}u &= -[u\varphi(u)]_{-z}^{\infty} + \int_{-z}^{\infty} \varphi(u)\,\mathrm{d}u \\ &= -z\varphi(-z) + [1 - \Phi(-z)] = -z\varphi(z) + \Phi(z)\end{aligned}$$

via integration by parts. Further, since $y_t = 0$ if and only if $x_{t-1} + u_t \leq 0$,

$$\begin{aligned}\mathbb{E}_{t-1}\lambda^2(\vartheta_0 x_{t-1})\mathbf{1}\{y_t = 0\} &= \lambda^2(\vartheta_0 x_{t-1})\mathbb{E}_{t-1}\mathbf{1}\{\vartheta_0 u_t \leq -\vartheta_0 x_{t-1}\} \\ &= \lambda^2(\vartheta_0 x_{t-1})\Phi(-\vartheta_0 x_{t-1}) = \varphi(\theta_0 x_{t-1})\lambda(\vartheta_0 x_{t-1}). \qquad \text{(B.11)}\end{aligned}$$

Similarly,

$$\begin{aligned}\mathbb{E}_{t-1}\vartheta_0 u_t \lambda(\vartheta_0 x_{t-1})\mathbf{1}\{y_t = 0\} &= \lambda(\vartheta_0 x_{t-1})\mathbb{E}_{t-1}\vartheta_0 u_t \mathbf{1}\{\vartheta_0 u_t \leq -\vartheta_0 x_{t-1}\} \\ &= \lambda(\vartheta_0 x_{t-1}) \int_{-\infty}^{-\vartheta_0 x_{t-1}} u\varphi(u)\,\mathrm{d}u = -\varphi(\vartheta_0 x_{t-1})\lambda(\vartheta_0 x_{t-1}) \qquad \text{(B.12)}\end{aligned}$$

since

$$\int_{-\infty}^{-z} u\varphi(u)\,\mathrm{d}u = -\int_{-\infty}^{-z} \varphi'(u)\,\mathrm{d}u = -\varphi(-z) = -\varphi(z).$$

Since the events $\{y_t > 0\}$ and $\{y_t = 0\}$ are mutually exclusive, it follows from (B.10) and (B.11) that

$$\begin{aligned}\mathbb{E}_{t-1} m_t^2 &= \vartheta_0^2 \left[\vartheta_0^2 \mathbb{E}_{t-1} u_t^2 \mathbf{1}\{y_t > 0\} + \lambda^2(\vartheta_0 x_{t-1})\mathbf{1}\{y_t = 0\}\right] \\ &= \vartheta_0^2 \left[-\vartheta_0 x_{t-1}\varphi(\vartheta_0 x_{t-1}) + \Phi(\vartheta_0 x_{t-1}) + \varphi(\theta_0 x_{t-1})\lambda(\vartheta_0 x_{t-1})\right] =: \vartheta_0^2 \Psi_0(\vartheta_0 x_{t-1})\end{aligned}$$

and from (B.10) and (B.12) that

$$\begin{aligned}\mathbb{E}_{t-1} m_t u_t &= \vartheta_0^2 \mathbb{E}_{t-1} u_t^2 \mathbf{1}\{y_t > 0\} - \vartheta_0 \mathbb{E}_{t-1} u_t \lambda(\vartheta_0 x_{t-1})\mathbf{1}\{y_t = 0\} \\ &= -\vartheta_0 x_{t-1}\varphi(\vartheta_0 x_{t-1}) + \Phi(\vartheta_0 x_{t-1}) + \varphi(\vartheta_0 x_{t-1})\lambda(\vartheta_0 x_{t-1}) = \Psi_0(\vartheta_0 x_{t-1})\end{aligned}$$

where Ψ_0 is bounded, and $\Psi_0(z) \to 1$ as $z \to \infty$. Hence

$$\mathbb{E}_{t-1}(m_t - \vartheta_0^2 u_t)^2 = \mathbb{E}_{t-1}m_t^2 - 2\vartheta_0^2\mathbb{E}_{t-1}m_t u_t + \vartheta_0^4\mathbb{E}_{t-1}u_t^2 = \vartheta_0^2[1 - \Psi_0(\vartheta_0 x_{t-1})]. \quad \text{(B.13)}$$

Finally, we note additionally that for each $\delta > 0$, there exists a $C < \infty$ such that

$$\begin{aligned}\mathbb{E}_{t-1}|m_t|^{2+\delta} &\leq \vartheta_0^{2+\delta} C\left[\vartheta_0^{2+\delta}\mathbb{E}_{t-1}|u_t|^{2+\delta} + \lambda^{2+\delta}(\vartheta_0 x_{t-1})\mathbb{E}_{t-1}\mathbf{1}\{\vartheta_0 u_t \leq -\vartheta_0 x_{t-1}\}\right] \\ &\leq \vartheta_0^{2+\delta} C\left[\vartheta_0^{2+\delta}\mathbb{E}|u_1|^{2+\delta} + \lambda^{2+\delta}(\vartheta_0 x_{t-1})[1 - \Phi(\vartheta_0 x_{t-1})]\right], \qquad \text{(B.14)}\end{aligned}$$

which is bounded by a constant, since in particular

$$\lambda^{2+\delta}(z)[1 - \Phi(z)] = \varphi(z)\lambda^{1+\delta}(z) \to 0$$

as $z \to +\infty$, by Lemma A.4.

Deduce from (B.13) and (B.14) (with $\delta = 2$) that the martingale

$$M_T := \sum_{t=1}^{T} z_{t-1,T}(m_t - \vartheta_0^2 u_t)$$

has conditional variance

$$\begin{aligned}\|\langle M_T\rangle\| &= \left\|\sum_{t=1}^{T}\mathbb{E}_{t-1}(m_t - \vartheta_0^2 u_t)^2 z_{t-1,T}z_{t-1,T}^{\mathsf{T}}\right\| \\ &\leq \vartheta_0^2\sum_{t=1}^{T}[1 - \Psi_0(\vartheta_0 x_{t-1})]\|z_{t-1,T}\|^2 \\ &\leq \vartheta_0^2\left(\sum_{t=1}^{T}[1 - \Psi_0(\vartheta_0 x_{t-1})]^2\right)^{1/2}\left(\sum_{t=1}^{T}\|z_{t-1,T}\|^4\right)^{1/2} = o_p(1)\end{aligned}$$

by Lemmas A.2(iii) and A.3(v). Hence $M_T = o_p(1)$ by Corollary 3.1 in Hall and Heyde (1980). □

Regarding the other component of $\mathcal{S}_T(0)$, we have from evaluating (B.7) at $p = 0$ that

$$\begin{aligned}\mathcal{S}_{h,T}(0) &= T^{-1/2}\sum_{t=1}^{T}[(\vartheta_0^{-1} - \vartheta_0 u_t^2)\mathbf{1}\{y_t > 0\} + v_t\lambda(-\vartheta_0 v_t)\mathbf{1}\{y_t = 0\}] \\ &= T^{-1/2}\sum_{t=1}^{T}[(\vartheta_0^{-1} - \vartheta_0 u_t^2)\mathbf{1}\{y_t > 0\} - x_{t-1}\lambda(\vartheta_0 x_{t-1})\mathbf{1}\{y_t = 0\}].\end{aligned}$$

Lemma B.4. *Suppose* A1–A4 *and* B *hold. Then*

$$\mathcal{S}_{h,T}(0) = T^{-1/2}\sum_{t=1}^{T}(\vartheta_0^{-1} - \vartheta_0 u_t^2) + o_p(1) =: \mathcal{M}_{h,T} + o_p(1)$$

Proof. We have

$$\mathcal{S}_{h,T}(0) = T^{-1/2}\sum_{t=1}^{T}(\vartheta_0^{-1} - \vartheta_0 u_t^2) - T^{-1/2}\sum_{t=1}^{T}\mathbf{1}\{y_t = 0\}[(\vartheta_0^{-1} - \vartheta_0 u_t^2) + x_{t-1}\lambda(\vartheta_0 x_{t-1})]$$
$$=: T^{-1/2}\sum_{t=1}^{T}(\vartheta_0^{-1} - \vartheta_0 u_t^2) + \Delta_T.$$

Since $\mathbb{E}_{t-1}(\vartheta_0^{-1} - \vartheta_0 u_t^2) = 0$, the first r.h.s. term is a martingale, and hence so too is Δ_T. We have the bound

$$\langle \Delta_T \rangle \leq 2T^{-1}\sum_{t=1}^{T}\mathbb{E}_{t-1}\mathbf{1}\{y_t = 0\}[(\vartheta_0^{-1} - \vartheta_0 u_t^2)^2 + x_{t-1}^2\lambda^2(\vartheta_0 x_{t-1})].$$

To show the the r.h.s. is $o_p(1)$ we first note that, similarly to the argument that yielded (B.11) and (B.14) above, for any $\delta \geq 0$,

$$\begin{aligned}\mathbb{E}_{t-1}\mathbf{1}\{y_t = 0\}|x_{t-1}|^{2+\delta}\lambda^{2+\delta}(\vartheta_0 x_{t-1}) &= |x_{t-1}|^{2+\delta}\lambda^{2+\delta}(\vartheta_0 x_{t-1})\Phi(-\vartheta_0 x_{t-1}) \\ &= \vartheta_0^{-(2+\delta)}|\vartheta_0 x_{t-1}|^{2+\delta}\lambda^{1+\delta}(\vartheta_0 x_{t-1})\varphi(\vartheta_0 x_{t-1}) \\ &=: \vartheta_0^{-(2+\delta)}G_\delta(\vartheta_0 x_{t-1}), \end{aligned} \tag{B.15}$$

where G_δ is bounded, and $G_\delta(z) \to 0$ as $z \to +\infty$, by Lemma A.4. Further, by Hölder's inequality and Lemma A.3(iii), for any $\delta > 0$,

$$T^{-1}\sum_{t=1}^{T}\mathbf{1}\{y_t = 0\}(\vartheta_0^{-1} - \vartheta_0 u_t^2)^2$$
$$\leq T^{-1}\left(\sum_{t=1}^{T}\left(\frac{1}{\vartheta_0} - \vartheta_0 u_t^2\right)^{2(1+\delta)}\right)^{1/(1+\delta)}\left(\sum_{t=1}^{T}\mathbf{1}\{y_t = 0\}\right)^{\delta/(1+\delta)} \xrightarrow{p} 0.$$

Since

$$\left\| T^{-1}\sum_{t=1}^{T}\left(\frac{1}{\vartheta_0} - \vartheta_0 u_t^2\right)^2\mathbf{1}\{y_t = 0\}\right\|_{1+\delta} \leq \left\|\left(\frac{1}{\vartheta_0} - \vartheta_0 u_1^2\right)^2\right\|_{1+\delta} < \infty$$

it follows that the preceding convergence also holds in L^1. Hence by the above and Lemma A.2(iii),

$$\langle \Delta_T \rangle \leq o_p(1) + 2\vartheta_0^{-1}T^{-1}\sum_{t=1}^{T}G_0(\vartheta_0 x_{t-1}) = o_p(1).$$

Deduce that $\Delta_T = o_p(1)$ by Corollary 3.1 in Hall and Heyde (1980). □

In view of Lemmas B.3 and B.4, it remains to determine the (joint) limiting distribution of the martingales $\mathcal{M}_{r,T}$ and $\mathcal{M}_{h,T}$, whose definitions appear in the statements of those lemmas. Note

$$\mathcal{M}_{r,T} = \vartheta_0^2 \sum_{t=1}^{T} \begin{bmatrix} T^{-1/2} \\ T^{-1} y_{t-1} \\ T^{-1/2} \Delta \boldsymbol{y}_{t-1} \end{bmatrix} u_t =: \begin{bmatrix} \mathcal{M}_{a,T} \\ \mathcal{M}_{c,T} \\ \mathcal{M}_{\boldsymbol{f},T} \end{bmatrix}.$$

We first consider $\mathcal{M}_{\boldsymbol{f},T}$. By Lemma A.1(ii), this has conditional variance

$$\vartheta_0^4 T^{-1} \sum_{t=1}^{T} \Delta \boldsymbol{y}_{t-1} \Delta \boldsymbol{y}_{t-1}^{\mathsf{T}} \mathbb{E}_{t-1} u_t^2 \xrightarrow{p} \sigma_0^{-2} \Omega,$$

where the convergence holds in probability since the limit is degenerate. Further, by that same result,

$$\sum_{t=1}^{T} \|z_{t-1,T}\|^4 \mathbb{E}_{t-1} |u_t|^4 \leq C \sum_{t=1}^{T} \|z_{t-1,T}\|^4 = O_p(T^{-1})$$

so that $\mathcal{M}_{\boldsymbol{f},T}$ satisfies a conditional Lyapunov condition (and therefore also a conditional Lindeberg condition). Hence by Corollary 3.1 in Hall and Heyde (1980),

$$\mathcal{M}_{\boldsymbol{f},T} \xrightarrow{d} \sigma_0^{-1} \Omega^{1/2} \xi_{(1)}, \tag{B.16}$$

where $\xi_{(1)} \sim N[0, I_k]$. Next, since $\mathcal{M}_{h,T}$ is merely the (scaled) sum of i.i.d. random variables (with finite variances), and

$$\mathbb{E}(\vartheta_0^{-1} - \vartheta_0 u_t^2)^2 = \sigma_0^2 \mathbb{E}(1 - (\sigma_0^{-1} u_t)^2)^2 = \sigma_0^2 (1 - 2 + 3) = 2\sigma_0^2,$$

since $\sigma_0^{-1} u_t \sim N[0, 1]$, it also follows by the central limit theorem that

$$\mathcal{M}_{h,T} \xrightarrow{d} 2^{1/2} \sigma_0 \xi_{(2)} \tag{B.17}$$

for $\xi_{(2)} \sim N[0, 1]$.

Our next step is to establish that the marginal convergences (B.16) and (B.17) hold jointly with the weak convergence of $(\mathcal{M}_{a,T}, \mathcal{M}_{c,T})$, to mutually independent limits. To that end, we first observe that $\mathcal{M}_{a,T} = \vartheta_0^2 U_T(1)$, where $U_T(\tau) := T^{-1} \sum_{t=1}^{\lfloor \tau T \rfloor} u_t$. Now fix $\{\lambda_i\}_{i=1}^m \subset [0, 1]$ with $\lambda_i < \lambda_{i+1}$, and consider the martingale

$$\mathcal{N}_T(\{\lambda_i\}) := T^{-1/2} \sum_{t=1}^{T} \begin{bmatrix} \mathbf{1}\{\lambda_1 < t/T \leq \lambda_2\} u_t \\ \vdots \\ \mathbf{1}\{\lambda_{m-1} < t/T \leq \lambda_m\} u_t \\ \Delta \boldsymbol{y}_{t-1} u_t \\ (\vartheta_0^{-1} - \vartheta_0 u_t^2) \end{bmatrix} = \begin{bmatrix} \mathcal{U}_T(\lambda_1, \lambda_2) \\ \vdots \\ \mathcal{U}_T(\lambda_{m-1}, \lambda_m) \\ \mathcal{M}_{\boldsymbol{f},T} \\ \mathcal{M}_{h,T} \end{bmatrix},$$

where $\mathcal{U}_T(\lambda_1,\lambda_2) = U_T(\lambda_2) - U_T(\lambda_1)$. Since each element of the above satisfies a conditional Lindeberg condition, by Corollary 3.1 in Hall and Heyde (1980), and the Cramér–Wold device, it suffices to show that all their conditional covariances converge in probability to zero. We have

$$\begin{aligned}\langle \mathcal{U}_T(\lambda_{i-1},\lambda_i), \mathcal{M}_{\boldsymbol{f},T}\rangle &= T^{-1}\sum_{t=1}^{T} \mathbf{1}\{\lambda_{i-1} < t/T \le \lambda_i\}\Delta \boldsymbol{y}_{t-1}\mathbb{E}_{t-1}u_t^2 \\ &= \sigma_0^2 T^{-1} \sum_{t=\lfloor \lambda_{i-1}T\rfloor+1}^{\lfloor \lambda_i T\rfloor} \Delta \boldsymbol{y}_{t-1} \\ &= \sigma_0^2 T^{-1}(\boldsymbol{y}_{\lfloor \lambda_i T\rfloor} - \boldsymbol{y}_{\lfloor \lambda_{i-1}T\rfloor}) = O_p(T^{-1/2})\end{aligned}$$

by Theorem 3.2 in Bykhovskaya and Duffy (2024),

$$\langle \mathcal{U}_T(\lambda_{i-1},\lambda_i), \mathcal{M}_{h,T}\rangle = T^{-1}\sum_{t=1}^{T} \mathbf{1}\{\lambda_{i-1} < t/T \le \lambda_i\}\mathbb{E}_{t-1}(\vartheta_0^{-1} - \vartheta_0 u_t^2)u_t = 0,$$

and for $i \neq j$,

$$\langle \mathcal{U}_T(\lambda_{i-1},\lambda_i), \mathcal{U}_T(\lambda_{j-1},\lambda_j)\rangle = 0.$$

It follows that $\mathcal{N}_T(\{\lambda_i\})$ weakly converges to a normal random vector with variance matrix

$$\operatorname{diag}\{\sigma_0^2(\lambda_2-\lambda_1),\ldots,\sigma_0^2(\lambda_m-\lambda_{m-1}),\sigma_0^2\Omega,2\sigma_0^2\}.$$

Deduce that (B.16) and (B.17) hold jointly with $U_T(\cdot) \xrightarrow{d} \sigma_0 W(\cdot)$ (the latter, on $D[0,1]$, by the functional central limit theorem), with $\xi_{(1)}$, $\xi_{(2)}$ and W being mutually independent.

Finally, we note that by Theorem 3.2 in Bykhovskaya and Duffy (2024) and Theorem 2.1 in Liang et al. (2016), the convergences

$$\begin{bmatrix}\mathcal{M}_{a,T}\\ \mathcal{M}_{c,T}\end{bmatrix} = \vartheta_0^2\begin{bmatrix} U_T(1) \\ T^{-1}\sum_{t=1}^{T} y_{t-1}u_t\end{bmatrix} \xrightarrow{d} \sigma_0^{-2}\begin{bmatrix}\sigma_0 W(1) \\ \sigma_0\int_0^1 Y(\tau)\,\mathrm{d}W(\tau)\end{bmatrix} = \sigma_0^{-1}\begin{bmatrix} W(1) \\ \int_0^1 Y(\tau)\,\mathrm{d}W(\tau)\end{bmatrix}$$

hold jointly with $Y_T \xrightarrow{d} Y$ and $U_T \xrightarrow{d} \sigma_0 W$ (on $D[0,1]$). Since Y is a function only of W, it follows that the preceding holds jointly with (B.16) and (B.17), to mutually independent limits. □

B.2.2 Proof of part (ii)

We partition the Hessian as

$$\mathcal{H}_T(p) = \begin{bmatrix}\mathcal{H}_{rr,T}(p) & \mathcal{H}_{hr,T}(p)\\ \mathcal{H}_{rh,T}(p) & \mathcal{H}_{hh,T}(p)\end{bmatrix} := \sum_{t=1}^{T}\begin{bmatrix} H_{rr,t}(p) & H_{hr,t}(p) \\ H_{rh,t}(p) & H_{hh,t}(p)\end{bmatrix},$$

and consider each of these components in turn.

$\boldsymbol{\mathcal{H}_{rr,T}}$. From (B.6), we have

$$\begin{aligned}\mathcal{H}_{rr,T}(p) &= -\sum_{t=1}^{T}\vartheta_T^2(h)\left[\mathbf{1}\{y_t>0\}+\lambda'[\vartheta_T(h)(z_{t-1,T}^{\mathsf{T}}r-v_t)]\mathbf{1}\{y_t=0\}\right]z_{t-1,T}z_{t-1,T}^{\mathsf{T}} \\ &= -\vartheta_T^2(h)\sum_{t=1}^{T}\mathbf{1}\{y_t>0\}z_{t-1,T}z_{t-1,T}^{\mathsf{T}} \\ &\qquad -\vartheta_T^2(h)\sum_{t=1}^{T}\mathbf{1}\{y_t=0\}\lambda'[\vartheta_T(h)(z_{t-1,T}^{\mathsf{T}}r-v_t)]z_{t-1,T}z_{t-1,T}^{\mathsf{T}} \\ &=: \mathcal{H}_{rr,T}^0(p)+\mathcal{R}_{rr,T}(p), \end{aligned} \tag{B.18}$$

where $\mathcal{H}_{rr,T}^0(p)$ depends on p only through h. Further,

$$\sum_{t=1}^{T}\mathbf{1}\{y_t>0\}z_{t-1,T}z_{t-1,T}^{\mathsf{T}}=\sum_{t=1}^{T}z_{t-1,T}z_{t-1,T}^{\mathsf{T}}-\sum_{t=1}^{T}\mathbf{1}\{y_t=0\}z_{t-1,T}z_{t-1,T}^{\mathsf{T}}\xrightarrow{d}Q_{ZZ}$$

by Lemma A.3(i) and the fact that, by the Cauchy–Schwarz (CS) inequality,

$$\left\|\sum_{t=1}^{T}\mathbf{1}\{y_t=0\}z_{t-1,T}z_{t-1,T}^{\mathsf{T}}\right\|^2\le\sum_{t=1}^{T}\mathbf{1}\{y_t=0\}\sum_{t=1}^{T}\|z_{t-1,T}\|^4=o_p(T)\cdot O_p(T^{-1})=o_p(1)$$

by parts (iii) and (v) of Lemma A.3. Since $\vartheta_T^2(h)\xrightarrow{p}\vartheta_0^2$ uniformly on compacta, it therefore follows that $\mathcal{H}_{rr,T}^0(p)\xrightarrow{d}\vartheta_0^2Q_{ZZ}$ on $L^{\mathrm{ucc}}(\mathbb{R}^{k+2})$.

Regarding the second r.h.s. term in (B.18), we note by Lemma A.4 $|\lambda'(x)|$ is bounded, so there exists a $C<\infty$ such that

$$\|\mathcal{R}_{rr,T}(p)\|\le C\vartheta_T^2(h)\sum_{t=1}^{T}\mathbf{1}\{y_t=0\}\|z_{t-1,T}\|^2$$

The sum on the r.h.s. is $o_p(1)$, since parts (iii) and (v) of Lemma A.3, and the CS inequality, yield

$$\left(\sum_{t=1}^{T}\mathbf{1}\{y_t=0\}\|z_{t-1,T}\|^2\right)^2\le\sum_{t=1}^{T}\mathbf{1}\{y_t=0\}\sum_{t=1}^{T}\|z_{t-1,T}\|^4=o_p(1). \tag{B.19}$$

Deduce that $\mathcal{R}_{rr,T}(p)\xrightarrow{p}0$ uniformly on compacta, whence $\mathcal{H}_{rr,T}(p)\xrightarrow{d}\vartheta_0^2Q_{ZZ}$ on $L^{\mathrm{ucc}}(\mathbb{R}^{k+2})$.

$\boldsymbol{\mathcal{H}_{hh,T}}$. From (B.8), we have

$$\begin{aligned}\mathcal{H}_{hh,T}(p) = -T^{-1}\sum_{t=1}^{T}(\vartheta_T^{-2}(h)+u_t^2) &+ T^{-1}\sum_{t=1}^{T}(\vartheta_T^{-2}(h)+u_t^2)\mathbf{1}\{y_t=0\}\\ &+ T^{-1}\sum_{t=1}^{T}[2(z_{t-1,T}^{\mathsf{T}}r)u_t-(z_{t-1,T}^{\mathsf{T}}r)^2]\mathbf{1}\{y_t>0\}\\ &- T^{-1}\sum_{t=1}^{T}(z_{t-1,T}^{\mathsf{T}}r-v_t)^2\lambda'[\vartheta_T(h)(z_{t-1,T}^{\mathsf{T}}r-v_t)]\mathbf{1}\{y_t=0\}\}\\ =: \mathcal{H}_{hh,T}^0(p)+\mathcal{R}_{hh,T}^1(p)&+\mathcal{R}_{hh,T}^2(p)+\mathcal{R}_{hh,T}^3(p),\end{aligned}$$

where, by the law of large numbers (LLN),

$$\mathcal{H}_{hh,T}^0(p) = -\vartheta_T^{-2}(h) - T^{-1}\sum_{t=1}^{T}u_t^2 \xrightarrow{p} -2\sigma_0^2$$

on $L^{\mathrm{ucc}}(\mathbb{R}^{k+2})$. Regarding the $\mathcal{R}_{hh,T}^i$ terms, we first note that by the CS inequality and Lemma A.3(iii)

$$|\mathcal{R}_{hh,T}^1(p)| \leq \vartheta_T^{-2}(h)T^{-1}\sum_{t=1}^{T}\mathbf{1}\{y_t=0\} + T^{-1}\left(\sum_{t=1}^{T}\mathbf{1}\{y_t=0\}\right)^{1/2}\left(\sum_{t=1}^{T}u_t^4\right)^{1/2} \xrightarrow{p} 0$$

uniformly on compacta. Moreover, by parts (i) and (iv) of Lemma A.3,

$$\begin{aligned}|\mathcal{R}_{hh,T}^2(p)| &\leq 2\|r\|\left\|T^{-1}\sum_{t=1}^{T}z_{t-1,T}u_t\mathbf{1}\{y_t>0\}\right\| + \|r\|^2T^{-1}\sum_{t=1}^{T}\|z_{t-1,T}\|^2\\ &= 2\|r\|o_p(T^{-1/2}) + \|r\|^2O_p(T^{-1}) \xrightarrow{p} 0\end{aligned}$$

uniformly on compacta. Finally, by Lemma A.4 there exists a $C<\infty$ such that

$$\begin{aligned}|\mathcal{R}_{hh,T}^3(p)| &\leq CT^{-1}\sum_{t=1}^{T}\mathbf{1}\{y_t=0\}(\|r\|^2\|z_{t-1,T}\|^2+|v_t|^2)\\ &= C\left[\|r\|^2T^{-1}\sum_{t=1}^{T}\mathbf{1}\{y_t=0\}\|z_{t-1,T}\|^2 + T^{-1}\sum_{t=1}^{T}\mathbf{1}\{y_t=0\}|v_t|^2\right].\end{aligned}$$

That both of the sums on the r.h.s. are $o_p(1)$ then follows from (B.19) above, and Lemma A.3(vi). Deduce that $\mathcal{R}_{hh,T}^3(p) \xrightarrow{p} 0$ uniformly on compacta. It follows that $\mathcal{H}_{hh,T}(p) \xrightarrow{p} -2\sigma_0^2$ on $L^{\mathrm{ucc}}(\mathbb{R}^{k+2})$.

$\boldsymbol{\mathcal{H}_{rh,T}}$. We have from (B.9) above that

$$\begin{aligned}\mathcal{H}_{rh,T}(p) &= \vartheta_T^{-1}(h)T^{-1/2}\mathcal{S}_{r,T}(p) \\ &\quad + \vartheta_T(h)T^{-1/2}\sum_{t=1}^{T}\mathbf{1}\{y_t > 0\}z_{t-1,T}(u_t - z_{t-1,T}^{\mathsf{T}}r) \\ &\quad - \vartheta_T(h)T^{-1/2}\sum_{t=1}^{T}\mathbf{1}\{y_t = 0\}z_{t-1,T}(z_{t-1,T}^{\mathsf{T}}r - v_t)\lambda'[\vartheta_T(z_{t-1,T}^{\mathsf{T}}r - v_t)] \\ &=: T^{-1/2}\vartheta_T^{-1}(h)\mathcal{S}_{r,T}(p) + \vartheta_T(h)\mathcal{R}_{r,T}^1(p) + \vartheta_T(h)\mathcal{R}_{r,T}^2(p).\end{aligned}$$

It follows from parts (i) and (iv) of Lemma A.3 that

$$|\mathcal{R}_{r,T}^1(p)| \leq \left\| T^{-1/2}\sum_{t=1}^{T}\mathbf{1}\{y_t > 0\}z_{t-1,T}u_t \right\| + \|r\|^2 T^{-1/2}\sum_{t=1}^{T}\|z_{t-1,T}\|^2 \xrightarrow{p} 0$$

uniformly on compacta. Moreover, by Lemma A.4 there exists a $C < \infty$ such that

$$\begin{aligned}|\mathcal{R}_{r,T}^2(p)| &\leq CT^{-1/2}\sum_{t=1}^{T}\mathbf{1}\{y_t = 0\}\|z_{t-1,T}\|(\|r\|\|z_{t-1,T}\| + |v_t|) \\ &\leq CT^{-1/2}\left[\|r\|\sum_{t=1}^{T}\|z_{t-1,T}^{\mathsf{T}}\|^2 + \sum_{t=1}^{T}\mathbf{1}\{y_t = 0\}\|z_{t-1,T}^{\mathsf{T}}\||v_t|\right]\end{aligned}$$

That each of the sums on the r.h.s. is $o_p(T^{1/2})$ follows from parts (i) and (vi) of Lemma A.3, noting in particular that

$$\sum_{t=1}^{T}\mathbf{1}\{y_t = 0\}\|z_{t-1,T}^{\mathsf{T}}\||v_t| \leq \left(\sum_{t=1}^{T}\|z_{t-1,T}^{\mathsf{T}}\|^2\right)^{1/2}\left(\sum_{t=1}^{T}\mathbf{1}\{y_t = 0\}|v_t|^2\right)^{1/2} = o_p(T^{1/2}) \tag{B.20}$$

by the CS inequality. Deduce that $\mathcal{R}_{r,T}^2(p) \xrightarrow{p} 0$ uniformly on compacta. Finally, we have from (B.5) that

$$\begin{aligned}\mathcal{S}_{r,T}(p) = \vartheta_T^2(h)\sum_{t=1}^{T}z_{t-1,T}(u_t - z_{t-1,T}^{\mathsf{T}}r)\mathbf{1}\{y_t > 0\} \\ - \vartheta_T(h)\sum_{t=1}^{T}z_{t-1,T}\lambda[\vartheta_T(h)(z_{t-1,T}^{\mathsf{T}}r - v_t)]\mathbf{1}\{y_t = 0\}.\end{aligned}$$

It follows from Lemma A.4 that there exists a $C < \infty$ such that

$$\lambda[\vartheta_T(h)(z_{t-1,T}^{\mathsf{T}}r - v_t)] \leq C[1 + \vartheta_T(h)(\|r\|\|z_{t-1,T}\| + |v_t|)].$$

Hence

$$|\mathcal{S}_{r,T}(p)| \leq \vartheta_T^2(h) \left\| \sum_{t=1}^T z_{t-1,T} u_t \mathbf{1}\{y_t > 0\} \right\| + \|r\| \vartheta_T^2(h) \sum_{t=1}^T \|z_{t-1,T}\|^2$$
$$+ C\vartheta_T(h) \Bigg[\sum_{t=1}^T \mathbf{1}\{y_t = 0\} \|z_{t-1,T}\| + \vartheta_T(h) \|r\| \sum_{t=1}^T \|z_{t-1,T}\|^2$$
$$+ \vartheta_T(h) \sum_{t=1}^T \mathbf{1}\{y_t = 0\} \|z_{t-1,T}\| |v_t| \Bigg].$$

That each of the sums on the r.h.s. are $o_p(T^{1/2})$ follows from parts (i) and (vi) of Lemma A.3, (B.20) above, and the fact that

$$\sum_{t=1}^T \mathbf{1}\{y_t = 0\} \|z_{t-1,T}\| \leq \left(\sum_{t=1}^T \mathbf{1}\{y_t = 0\} \right)^{1/2} \left(\sum_{t=1}^T \|z_{t-1,T}\|^2 \right)^{1/2} = o_p(T^{1/2})$$

by the CS inequality. Deduce that $T^{-1/2}\mathcal{S}_{r,T}(p) \xrightarrow{p} 0$ uniformly on compacta. Hence $\mathcal{H}_{rh,T}(p) \xrightarrow{p} 0$ uniformly on compacta.

Joint convergence of $\mathcal{S}_T(0)$ and $\mathcal{H}_T(p)$. It follows from the preceding that $\mathcal{H}_T(p) \xrightarrow{d} \mathcal{H}$ on $L^{\mathrm{ucc}}(\mathbb{R}^{k+2})$. In view of Lemma A.3(i), this convergence holds jointly with $U_T \xrightarrow{d} \sigma_0 W$, with $\mathcal{H}$ being $\sigma(W)$-measurable. We have shown in the proof of part (i) that the convergence $\mathcal{S}_T(0) \xrightarrow{d} \mathcal{S}$ also holds jointly with $U_T \xrightarrow{d} \sigma_0 W$, with each element of $\mathcal{S}$ either being $\sigma(W)$-measurable, or independent of W. Deduce that $\mathcal{S}_T(0) \xrightarrow{d} \mathcal{S}$ jointly with $\mathcal{H}_T(p) \xrightarrow{d} \mathcal{H}$, as claimed. □

B.3 Proof of Proposition B.2

Given $r \in \mathbb{R}^{k+1}$, $h \in \mathbb{R}$ and $T \in \mathbb{N}$, define $\tilde{r} \in \mathbb{R}^{k+1}$ to be such that $r = \vartheta_T^{-1}(h)\tilde{r}$. Similarly, to Olsen (1978), setting $p_T(\tilde{r}, h) := (\vartheta_T^{-1}(h)\tilde{r}^\mathsf{T}, h)^\mathsf{T}$, we have

$$\begin{aligned}
\tilde{\ell}_T(\tilde{r}, h) &:= \ell_T[p_T(\tilde{r}, h)] \\
&= N_T[\log \vartheta_T(h) - \tfrac{1}{2}\log 2\pi] - \tfrac{1}{2}\vartheta_T^2(h) \sum_{t=1}^T (u_t - \vartheta_T^{-1}(h) z_{t-1,T}^\mathsf{T} \tilde{r})^2 \mathbf{1}\{y_t > 0\} \\
&\quad + \sum_{t=1}^T \mathbf{1}\{y_t = 0\} \log\{1 - \Phi[\vartheta_T(h)(\vartheta_T^{-1}(h) z_{t-1,T}^\mathsf{T} \tilde{r} - v_t)]\} \\
&= N_T[\log(\vartheta_0 + T^{-1/2}h) - \tfrac{1}{2}\log 2\pi] - \tfrac{1}{2} \sum_{t=1}^T (\vartheta_0 u_t + hT^{-1/2}u_t - z_{t-1,T}^\mathsf{T} \tilde{r})^2 \mathbf{1}\{y_t > 0\} \\
&\quad + \sum_{t=1}^T \mathbf{1}\{y_t = 0\} \log\{1 - \Phi[(z_{t-1,T}^\mathsf{T} \tilde{r} - hT^{-1/2}v_t - \vartheta_0 v_t)]\}.
\end{aligned}$$

The penultimate line in the preceding display is clearly concave in $(\tilde{r}, h)$, while the concavity of the final line follows from log-concavity of the normal distribution (e.g. Bagnoli and Bergstrom, 2005). Hence $\tilde{\ell}_T$ is concave.

It follows from (B.4) that for every $(\tilde{r}, h) \in \mathbb{R}^{k+2}$ and T sufficiently large, there exists a $\lambda \in [0, 1]$ such that

$$\begin{aligned}\tilde{\ell}_T(\tilde{r}, h) - \tilde{\ell}_T(0, 0) &= \ell_T[p_T(\tilde{r}, h)] - \ell_T(0) \\ &= \mathcal{S}_T(0)^{\mathsf{T}} p_T(\tilde{r}, h) + \tfrac{1}{2} p_T(\tilde{r}, h)^{\mathsf{T}} \mathcal{H}_T(0) p_T(\tilde{r}, h) \\ &\qquad + p_T(\tilde{r}, h)^{\mathsf{T}} \mathcal{R}_T[\lambda p_T(\tilde{r}, h)] p_T(\tilde{r}, h).\end{aligned}$$

Since for each $(\tilde{r}, h) \in \mathbb{R}^{k+2}$,

$$p_T(\tilde{r}, h) = \begin{bmatrix} \vartheta_T^{-1}(h)\tilde{r} \\ h \end{bmatrix} \to \begin{bmatrix} \vartheta_0^{-1}\tilde{r} \\ h \end{bmatrix} = \begin{bmatrix} \vartheta_0^{-1} I_{k+1} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{r} \\ h \end{bmatrix} =: M\tilde{p}$$

and $\mathcal{R}_T(p) \xrightarrow{p} 0$ uniformly on compacta, it follows that

$$\tilde{\ell}_T(\tilde{p}) - \tilde{\ell}_T(0) \xrightarrow{d} \mathcal{S}^{\mathsf{T}}(M\tilde{p}) + \tfrac{1}{2}(M\tilde{p})^{\mathsf{T}} \mathcal{H}(M\tilde{p})$$

in the sense of finite-dimensional convergence. Since both the left and right hand sides of the preceding display are concave, the l.h.s. is maximised at $(\vartheta_T(\hat{h}_T)\hat{r}_T, \hat{h}_T)$ and the r.h.s. almost surely has an unique maximum at $\tilde{p}^* := -M^{-1}\mathcal{H}^{-1}\mathcal{S}$, it follows by Lemma A in Knight (1989) that

$$M^{-1}\hat{p}_T = \begin{bmatrix} \vartheta_T(\hat{h}_T)\hat{r}_T \\ \hat{h}_T \end{bmatrix} + o_p(1) \xrightarrow{d} \tilde{p}^* = -M^{-1}\mathcal{H}^{-1}\mathcal{S}. \qquad \square$$

C Censored least absolute deviations

C.1 Proof of Theorem 3.2

The CLAD estimators $\hat{\rho}_T^{\mathsf{L}} := (\hat{\alpha}_T^{\mathsf{L}}, \hat{\beta}_T^{\mathsf{L}}, \hat{\boldsymbol{\phi}}_T^{\mathsf{L}})$ are minimisers of

$$S_T(\rho) := S_T(\alpha, \beta, \boldsymbol{\phi}) := \sum_{t=1}^{T} |y_t - [\alpha + \beta y_{t-1} + \boldsymbol{\phi}^{\mathsf{T}} \Delta \boldsymbol{y}_{t-1}]_+| \tag{C.1}$$

over Π. To economise on notation, in this section we shall generally write $\hat{\alpha}_T^{\mathsf{L}}$ as merely $\hat{\alpha}_T$, etc. Our first result establishes that $\hat{\beta}_T$ concentrates in a $O_p(T^{-1/2})$ neighbourhood of $\beta_{T,0}$; its proof appears in Appendix C.2 below.

Proposition C.1. *Suppose* A1–A4 *and* C *hold. Then* $T^{1/2}(\hat{\beta}_T - \beta_{T,0}) = O_p(1)$.

In view of the preceding, it will be convenient to work with the following reparametrisation of the model

$$\varpi := \begin{bmatrix} \alpha - \alpha_{T,0} \\ \delta \\ \boldsymbol{\phi} - \boldsymbol{\phi}_0 \end{bmatrix} := \begin{bmatrix} \alpha - \alpha_{T,0} \\ T^{1/2}(\beta - \beta_{T,0}) \\ \boldsymbol{\phi} - \boldsymbol{\phi}_0 \end{bmatrix}.$$

Since Proposition C.1 implies that for any $\epsilon > 0$ we may choose $M < \infty$ such that $\liminf_{T\to\infty} \mathbb{P}\{|\hat{\delta}_T| \leq M\} > 1 - \epsilon/2$, we shall henceforth treat the parameter space $\bar{\Pi}$ for ϖ as though it were compact. (Recall that the parameter spaces for α and $\boldsymbol{\phi}$ are compact under Assumption C.) Similarly to Appendix B, we define the local parameters

$$r := \begin{bmatrix} a \\ c \\ \boldsymbol{f} \end{bmatrix} = \begin{bmatrix} T^{1/2} & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T^{1/2} \end{bmatrix} \begin{bmatrix} \alpha - \alpha_{T,0} \\ \beta - \beta_{T,0} \\ \boldsymbol{\phi} - \boldsymbol{\phi}_0 \end{bmatrix} = T^{1/2} \begin{bmatrix} \alpha - \alpha_{T,0} \\ \delta \\ \boldsymbol{\phi} - \boldsymbol{\phi}_0 \end{bmatrix} = T^{1/2}\varpi,$$

and denote the systematic part of the r.h.s. of the model (2.1), evaluated at the true parameters, as

$$x_{t-1} := \alpha_{T,0} + \beta_{T,0} y_{t-1} + \boldsymbol{\phi}_0^{\mathsf{T}} \Delta \boldsymbol{y}_{t-1}. \tag{C.2}$$

We may then rewrite the part of the LAD criterion function (C.1) that depends on $\rho = (\alpha, \beta, \boldsymbol{\phi}^{\mathsf{T}})^{\mathsf{T}}$ in terms of the local parameters ϖ or r as

$$\alpha + \beta y_{t-1} + \boldsymbol{\phi}^{\mathsf{T}} \Delta \boldsymbol{y}_{t-1} = x_{t-1} + \varpi^{\mathsf{T}} \mathcal{Z}_{t-1,T} = x_{t-1} + r^{\mathsf{T}} z_{t-1,T},$$

where, recalling (A.3) above,

$$z_{t-1,T} = \begin{bmatrix} T^{-1/2} \\ T^{-1} y_{t-1} \\ T^{-1/2} \Delta \boldsymbol{y}_{t-1} \end{bmatrix}, \qquad \mathcal{Z}_{t-1,T} := \begin{bmatrix} 1 \\ T^{-1/2} y_{t-1} \\ \Delta \boldsymbol{y}_{t-1} \end{bmatrix} = T^{1/2} z_{t-1,T}.$$

To establish the rate of convergence and thence the limiting distribution of the CLAD estimator, we will need to consider (appropriately scaled) counterparts of the CLAD criterion in terms of both ϖ and r, defined here (making a slight abuse of notation in reusing S_T) as

$$S_T(\varpi) := \frac{1}{T} \sum_{t=1}^{T} |y_t - [x_{t-1} + \varpi^{\mathsf{T}} \mathcal{Z}_{t-1,T}]_+|, \quad \mathbb{S}_T(r) := \sum_{t=1}^{T} |y_t - [x_{t-1} + r^{\mathsf{T}} z_{t-1,T}]_+|.$$

The first of these will be used to establish the rate of convergence, while the second will help us to derive the limiting distribution. Because $\{y_t\}$ is nonstationary, the appropriate centring for each of these functions is given not by their unconditional expectations, but instead by

$$\bar{S}_T(\varpi) := \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{t-1}|y_t - [x_{t-1} + \varpi^{\mathsf{T}}\mathcal{Z}_{t-1,T}]_+|$$

$$\bar{\mathbb{S}}_T(r) := \sum_{t=1}^{T} \mathbb{E}_{t-1}|y_t - [x_{t-1} + r^{\mathsf{T}} z_{t-1,T}]_+|.$$

Define the associated recentred criterion functions

$$U_T(\varpi) := S_T(\varpi) - \bar{S}_T(\varpi), \qquad \mathbb{U}_T(r) := \mathbb{S}_T(r) - \bar{\mathbb{S}}_T(r),$$

and let $\{e_t\}$ denote the (bounded) martingale difference sequence defined by

$$e_t := \operatorname{sgn}([x_{t-1} + u_t]_+ - [x_{t-1}]_+) - \mathbb{E}_{t-1} \operatorname{sgn}([x_{t-1} + u_t]_+ - [x_{t-1}]_+). \tag{C.3}$$

Our main results on the asymptotics of these functions are the following.

Proposition C.2. *Suppose* A1–A4 *and* C *hold.*

(i) *For every $\epsilon, \delta > 0$, there exists an $\eta > 0$ such that*

$$\liminf_{T\to\infty} \mathbb{P}\left\{\inf_{\|\varpi\|\geq\delta}[\bar{S}_T(\varpi) - \bar{S}_T(0)] > \eta\right\} \geq 1 - \epsilon.$$

(ii) *For every $\epsilon > 0$, there exist $\delta, \eta > 0$ such that*

$$\liminf_{T\to\infty} \mathbb{P}\{\bar{S}_T(\varpi) - \bar{S}_T(0) \geq \eta\|\varpi\|^2,\ \forall\|\varpi\| \leq \delta\} \geq 1 - \epsilon. \tag{C.4}$$

(iii) *Uniformly on compact subsets of $\mathbb{R}^{k+1}$, for Q_{ZZ} defined in Lemma A.3(i),*

$$\bar{\mathbb{S}}_T(r) - \bar{\mathbb{S}}_T(0) \xrightarrow{d} f_u(0) r^{\mathsf{T}} Q_{ZZ} r.$$

Proposition C.3. *Suppose* A1–A4 *and* C *hold. Define $Z_{t,T} := \sum_{s=0}^{t}\|z_{s,T}\|^2$.*

(i) *For every $\epsilon > 0$, there exist $C_\epsilon, K_\epsilon < \infty$ such that $\mathbb{P}\{Z_{T-1,T} > K_\epsilon\} < \epsilon$, and*

$$\mathbb{P}\left\{\sup_{\|\varpi\|\leq\delta} T^{1/2}|U_T(\varpi) - U_T(0)| \geq \kappa,\ Z_{T-1,T} \leq K_\epsilon\right\} < \frac{C_\epsilon \delta}{\kappa} \tag{C.5}$$

for all $\delta, \kappa > 0$, for all T sufficiently large.

(ii) *$\{\mathbb{U}_T(r)\}$ is stochastically equicontinuous on $\mathbb{R}^{k+1}$.*

(iii) *For each* $r = (a, c, \boldsymbol{f}^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{k+1}$,

$$\begin{aligned}\mathbb{U}_T(r) - \mathbb{U}_T(0) &= -r^\mathsf{T} \sum_{t=1}^{T} z_{t-1,T} e_t + o_p(1) \\ &\xrightarrow{d} -r^\mathsf{T}\zeta = -\int_0^1 [a + cY(\tau)] \, \mathrm{d}\widetilde{W}(\tau) - \boldsymbol{f}^\mathsf{T}\xi\end{aligned}$$

where

$$\zeta := \begin{bmatrix}\widetilde{W}(1) & \int_0^1 Y(\tau)\,\mathrm{d}\widetilde{W}(\tau) & \xi^\mathsf{T}\end{bmatrix}^\mathsf{T}$$

for $(\sigma_0 W, \widetilde{W})$ *a bivariate Brownian motion with covariance matrix*

$$\mathbb{V}\begin{bmatrix}\sigma_0 W(1) \\ \widetilde{W}(1)\end{bmatrix} = \begin{bmatrix}\mathbb{E}u_t^2 & \mathbb{E}|u_t| \\ \mathbb{E}|u_t| & 1\end{bmatrix},$$

and $\xi \sim \mathcal{N}[0, \Omega]$ *independent of* $(W, \widetilde{W})$.

We may now finally proceed with the proof.

Proof of Theorem 3.2. We want to derive the limiting distribution of

$$D_{r,T}(\hat{\rho}_T - \rho_{T,0}) = T^{1/2}\hat{\varpi}_T = \hat{r}_T.$$

By definition, $\hat{r}_T$ minimises

$$\mathbb{S}_T(r) - \mathbb{S}_T(0) = [\bar{\mathbb{S}}_T(r) - \bar{\mathbb{S}}_T(0)] + [\mathbb{U}_T(r) - \mathbb{U}_T(0)] \xrightarrow{d} f_u(0) r^\mathsf{T} Q_{ZZ} r - r^\mathsf{T}\zeta \tag{C.6}$$

where the convergence holds on $L^{\mathrm{ucc}}(\mathbb{R}^{k+1})$, by Propositions C.2(iii), C.3(ii) and C.3(iii). The r.h.s. defines a continuous function of r, which since Q_{ZZ} is a.s. positive definite (by Lemma A.3(i)) also has a unique minimum a.s. Therefore, once we have shown below that

$$\hat{r}_T = O_p(1) \tag{C.7}$$

it will follow by the argmax continuous mapping theorem (Theorem 3.2.2 in Van der Vaart and Wellner, 1996) that $\hat{r}_T$ converges in distribution to the minimiser of the r.h.s. of (C.6), i.e.

$$\hat{r}_T \xrightarrow{d} \frac{1}{2 f_u(0)} Q_{ZZ}^{-1}\zeta.$$

It remains to show (C.7): equivalently, that $T^{1/2}\hat{\varpi}_T = O_p(1)$. Although we cannot apply their result directly, the argument used here closely follows the proof of Theorem 3.2.5 in Van der Vaart and Wellner (1996). Let $\epsilon > 0$, and note that, for $Z_{T-1,T}$ as appears in

Proposition C.3(i), we may choose K such that for every $L > 0$,

$$\begin{aligned}\mathbb{P}\{T^{1/2}\|\hat{\varpi}_T\| > L\} &\leq \mathbb{P}\{T^{1/2}\|\hat{\varpi}_T\| > L,\ Z_{T-1,T} \leq K\} + \mathbb{P}\{Z_{T-1,T} > K\} \\ &\leq \mathbb{P}\{T^{1/2}\|\hat{\varpi}_T\| > L,\ Z_{T-1,T} \leq K\} + \epsilon \end{aligned} \tag{C.8}$$

for all T sufficiently large. Further, by defining the 'shells'

$$\bar{\Pi}_{j,T} \coloneqq \{\varpi \in \bar{\Pi} \mid 2^j < T^{1/2}\|\varpi\| \leq 2^{j+1}\}$$

for $j \in \mathbb{N}$, we obtain for any $M \in \mathbb{N}$ and $\gamma > 0$ that

$$\begin{aligned}&\mathbb{P}\{T^{1/2}\|\hat{\varpi}_T\| > 2^M,\ Z_{T-1,T} \leq K\} \\ &\quad \leq \mathbb{P}\left(\bigcup_{\substack{j \geq M \\ 2^j \leq T^{1/2}\gamma}} \{\hat{\varpi}_T \in \bar{\Pi}_{j,T},\ Z_{T-1,T} \leq K\}\right) + \mathbb{P}\{\|\hat{\varpi}_T\| \geq \gamma,\ Z_{T-1,T} \leq K\}. \end{aligned} \tag{C.9}$$

We will now show that the r.h.s. can be made arbitrarily small, for all T sufficiently large, by choice of M.

Regarding the final r.h.s. probability in (C.9), we note that if $\|\hat{\varpi}_T\| \geq \gamma$, then we must have the first inequality in

$$0 \geq \inf_{\|\varpi\| \geq \gamma} [S_T(\varpi) - S_T(0)] \geq \inf_{\|\varpi\| \geq \gamma} [\bar{S}_T(\varpi) - \bar{S}_T(0)] + \inf_{\|\varpi\| \geq \gamma} [U_T(\varpi) - U_T(0)]$$

where the second inequality follows from the decomposition $S_T = \bar{S}_T + U_T$. This further implies the first inequality in

$$\inf_{\|\varpi\| \geq \gamma} [\bar{S}_T(\varpi) - \bar{S}_T(0)] \leq - \inf_{\|\varpi\| \geq \gamma} [U_T(\varpi) - U_T(0)] \leq \sup_{\varpi \in \Pi} |U_T(\varpi) - U_T(0)|$$

whence

$$\begin{aligned}&\mathbb{P}\{\|\hat{\varpi}_T\| \geq \gamma,\ Z_{T-1,T} \leq K\} \\ &\qquad \leq \mathbb{P}\left\{\sup_{\varpi \in \Pi} |U_T(\varpi) - U_T(0)| \geq \inf_{\|\varpi\| \geq \gamma} [\bar{S}_T(\varpi) - \bar{S}_T(0)],\ Z_{T-1,T} \leq K\right\}.\end{aligned}$$

In view of Proposition C.2(i), we may choose $\eta > 0$ such that

$$\mathbb{P}\left\{\inf_{\|\varpi\| \geq \gamma} [\bar{S}_T(\varpi) - \bar{S}_T(0)] > \eta\right\} > 1 - \epsilon/2$$

for all T sufficiently large. Thus

$$\mathbb{P}\{\|\hat{\varpi}_T\| \geq \gamma,\ Z_{T-1,T} \leq K\} \leq \mathbb{P}\left\{\sup_{\varpi\in\Pi}|U_T(\varpi) - U_T(0)| \geq \eta,\ Z_{T-1,T} \leq K\right\} + \mathbb{P}\left\{\inf_{\|\varpi\|\geq\gamma}[\bar{S}_T(\varpi) - \bar{S}_T(0)] \leq \eta\right\}$$
$$\leq \mathbb{P}\left\{\sup_{\varpi\in\Pi}|U_T(\varpi) - U_T(0)| \geq \eta,\ Z_{T-1,T} \leq K\right\} + \epsilon/2$$

for all T sufficiently large. Finally, taking $\kappa = T^{1/2}\eta$ in Proposition C.3(i), and δ sufficiently large that the δ-ball centred at zero contains the whole of $\bar{\Pi}$, we obtain that for some C depending only on ϵ,

$$\mathbb{P}\left\{\sup_{\varpi\in\Pi}|U_T(\varpi) - U_T(0)| \geq \eta,\ Z_{T-1,T} \leq K\right\} < \frac{C\delta}{T^{1/2}\eta} \to 0$$

as $T \to \infty$. Deduce that

$$\limsup_{T\to\infty} \mathbb{P}\{\|\hat{\varpi}_T\| \geq \gamma,\ Z_{T-1,T} \leq K\} < \epsilon. \tag{C.10}$$

We next consider the events in the union on the r.h.s. of (C.9). Observe that the largest value that $\|\varpi\|$ can take if ϖ lies in $\bigcup_{\{j\in\mathbb{N}|2^j\leq T^{1/2}\gamma\}} \bar{\Pi}_{j,T}$ is bounded by 2γ. Without disturbing the conclusion of (C.10), which holds for arbitrary $\gamma > 0$, we may take γ sufficiently small that Proposition C.2(ii) applies for $\delta = 2\gamma$ and the $\epsilon > 0$ given above; i.e. there exists an $\eta > 0$ (in general, different from that given previously) such that for the event

$$A_T := \{\bar{S}_T(\varpi) - \bar{S}_T(0) \geq \eta\|\varpi\|^2,\ \forall\|\varpi\| \leq 2\gamma\} \tag{C.11}$$

we have

$$\liminf_{T\to\infty} \mathbb{P}(A_T) \geq 1 - \epsilon. \tag{C.12}$$

This ensures that $\bar{S}_T(\varpi) - \bar{S}_T(0)$ is minorised, with high probability, by $\eta\|\varpi\|^2$ simultaneously on each of the $\bar{\Pi}_{j,T}$ appearing on the r.h.s. of (C.9). We can accordingly bound

$$\mathbb{P}\left(\bigcup_{\substack{j\geq M\\ 2^j\leq T^{1/2}\gamma}} \{\hat{\varpi}_T \in \bar{\Pi}_{j,T},\ Z_{T-1,T} \leq K\}\right)$$
$$\leq \sum_{\substack{j\geq M\\ 2^j\leq T^{1/2}\gamma}} \mathbb{P}\left(\{\hat{\varpi}_T \in \bar{\Pi}_{j,T},\ Z_{T-1,T} \leq K\} \cap A_T\right) + \mathbb{P}(A_T^c) \tag{C.13}$$

such that the final term is eventually bounded by ϵ, while the minorisation of $\bar{S}_T(\varpi) - \bar{S}_T(0)$ holds for each of the events entering the sum.

Now $\hat{\varpi}_T \in \bar{\Pi}_{j,T}$ implies the first inequality in

$$0 \geq \inf_{\varpi \in \bar{\Pi}_{j,T}} [S_T(\varpi) - S_T(0)] \geq \inf_{\varpi \in \bar{\Pi}_{j,T}} [\bar{S}_T(\varpi) - \bar{S}_T(0)] + \inf_{\varpi \in \bar{\Pi}_{j,T}} [U_T(\varpi) - U_T(0)]$$

whence, in view of $\bar{\Pi}_{j,T} = \{\varpi \in \bar{\Pi} \mid 2^j < T^{1/2}\|\varpi\| \leq 2^{j+1}\}$

$$\inf_{\varpi \in \bar{\Pi}_{j,T}} [\bar{S}_T(\varpi) - \bar{S}_T(0)] \leq \sup_{\|\varpi\| \leq T^{-1/2}2^{j+1}} |U_T(\varpi) - U_T(0)|.$$

Hence, in view of (C.11), and the fact that $\|\varpi\|$ is bounded below by $T^{-1/2}2^j$ for $\varpi \in \bar{\Pi}_{j,T}$,

$$\begin{aligned}
&\mathbb{P}\left(\{\hat{\varpi}_T \in \bar{\Pi}_{j,T},\ Z_{T-1,T} \leq K\} \cap A_T\right) \\
&\quad \leq \mathbb{P}\left(\left\{\sup_{\|\varpi\| \leq T^{-1/2}2^{j+1}} |U_T(\varpi) - U_T(0)| \geq \inf_{\varpi \in \bar{\Pi}_{j,T}} [\bar{S}_T(\varpi) - \bar{S}_T(0)],\ Z_{T-1,T} \leq K\right\} \cap A_T\right) \\
&\quad \leq \mathbb{P}\left\{\sup_{\|\varpi\| \leq T^{-1/2}2^{j+1}} |U_T(\varpi) - U_T(0)| \geq \eta T^{-1}2^{2j},\ Z_{T-1,T} \leq K\right\}
\end{aligned}$$

for all $j \geq M$ such that $2^j \leq T^{1/2}\gamma$, for all T sufficiently large. Finally, we have by Proposition C.3(i) with $\kappa = \eta T^{-1/2}2^{2j}$ and $\delta = T^{-1/2}2^{j+1}$ that for C depending only on ϵ,

$$\mathbb{P}\left\{\sup_{\|\varpi\| \leq T^{-1/2}2^{j+1}} |U_T(\varpi) - U_T(0)| \geq \eta T^{-1}2^{2j},\ Z_{T-1,T} \leq K\right\} \leq \frac{2C}{\eta}2^{-j} \tag{C.14}$$

for all $j \geq M$ such that $2^j \leq T^{1/2}\gamma$, for all T sufficiently large. That is, the preceding bound applies simultaneously to every summand on the r.h.s. of (C.13), for T sufficiently large.

Thus it follows that, for all T sufficiently large,

$$\begin{aligned}
\mathbb{P}\{T^{1/2}\|\hat{\varpi}_T\| > 2^M\} &\leq_{(1)} \mathbb{P}\{T^{1/2}\|\hat{\varpi}_T\| > 2^M,\ Z_{T-1,T} \leq K\} + \epsilon \\
&\leq_{(2)} \sum_{\substack{j \geq M \\ 2^j \leq T^{1/2}\gamma}} \mathbb{P}\left(\{\hat{\varpi}_T \in \bar{\Pi}_{j,T},\ Z_{T-1,T} \leq K\} \cap A_T\right) + \mathbb{P}(A_T^c) \\
&\qquad\qquad + \mathbb{P}\{\|\hat{\varpi}_T\| \geq \gamma,\ Z_{T-1,T} \leq K\} + \epsilon \\
&\leq_{(3)} \sum_{\substack{j \geq M \\ 2^j \leq T^{1/2}\gamma}} \frac{2C}{\eta}2^{-j} + 3\epsilon \leq 2^{-M}\frac{4C}{\eta} + 3\epsilon
\end{aligned}$$

where $\leq_{(1)}$ holds by (C.8) with $L = 2^M$, $\leq_{(2)}$ by (C.9) and (C.13), and $\leq_{(3)}$ by (C.10), (C.12) and (C.14). Since C depends on ϵ but not M, the r.h.s. can be made arbitrarily small by suitable choice of $\epsilon > 0$ and then $M \in \mathbb{N}$. Thus $T^{1/2}\hat{\varpi}_T = O_p(1)$. □

C.2 Proofs of Propositions C.1–C.3

C.2.1 Preliminaries

For future use, we note the following identity (see also Herce, 1996, p. 150):

$$|A - B| - |A| = -\operatorname{sgn}(A)B + \mathbf{1}\{A = 0\}|B| \\ + 2(B - A)[\mathbf{1}\{B > A > 0\} - \mathbf{1}\{B < A < 0\}]. \quad \text{(C.15)}$$

In proving Propositions C.2 and C.3, we will make repeated use of the following, the proof of which appears immediately below.

Lemma C.4. *Suppose* A1–A4 *and* C *hold. Let* w_{t-1} *be* $\mathcal{F}_{t-1}$*-measurable, and define*

$$\ell_{t-1}(w_{t-1}) := [x_{t-1} + w_{t-1}]_+ - [x_{t-1}]_+$$
$$G(b, a) := \int_a^b [F_u(u) - F_u(a)]\,\mathrm{d}u,$$

with the convention that $\int_a^b v(x)\,\mathrm{d}x = -\int_b^a v(x)\,\mathrm{d}x$ *when* $b < a$*. Then*

$$\mathbb{E}_{t-1}\{|y_t - [x_{t-1} + w_{t-1}]_+| - |y_t - [x_{t-1}]_+|\}$$
$$= \int_{-\infty}^{\infty} \{|u - \ell_{t-1}(w_{t-1})| - |u|\} f_u(u - [x_{t-1}]_-)\,\mathrm{d}u \quad \text{(C.16)}$$
$$= 2\ell_{t-1}(w_{t-1})[F_u(-[x_{t-1}]_-) - F_u(0)] + 2G\{\ell_{t-1}(w_{t-1}) - [x_{t-1}]_-, -[x_{t-1}]_-\}. \quad \text{(C.17)}$$

Moreover, there exists a $C < \infty$ *such that*

$$|\ell_{t-1}(w_{t-1})[F_u(-[x_{t-1}]_-) - F_u(0)]| \leq C|w_{t-1}|^2 \mathbf{1}\{x_{t-1} < 0\} \quad \text{(C.18)}$$

and

$$|G\{\ell_{t-1}(w_{t-1}) - [x_{t-1}]_-, -[x_{t-1}]_-\} - \tfrac{1}{2} f_u(-[x_{t-1}]_-)\ell_{t-1}^2(w_{t-1})| \leq C|w_{t-1}|^3. \quad \text{(C.19)}$$

Proof. Letting $(\mathrm{I})_{t-1}$ denote the l.h.s. of (C.16), we have

$$(\mathrm{I})_{t-1} = \int_{-\infty}^{\infty} \{|[x_{t-1} + u]_+ - [x_{t-1} + w_{t-1}]_+| - |[x_{t-1} + u]_+ - [x_{t-1}]_+|\} f_u(u)\,\mathrm{d}u$$
$$= \int_{-x_{t-1}}^{\infty} \{|x_{t-1} + u - [x_{t-1} + w_{t-1}]_+| - |x_{t-1} + u - [x_{t-1}]_+|\} f_u(u)\,\mathrm{d}u$$
$$+ \int_{-\infty}^{-x_{t-1}} \{[x_{t-1} + w_{t-1}]_+ - [x_{t-1}]_+\} f_u(u)\,\mathrm{d}u$$
$$=: (\mathrm{II})_{t-1} + (\mathrm{III})_{t-1}$$

where

$$
\begin{aligned}
(\mathrm{II})_{t-1} = \int_{-\infty}^{\infty} & \{|x_{t-1} + u - [x_{t-1} + w_{t-1}]_+| - |x_{t-1} + u - [x_{t-1}]_+|\} f_u(u)\,\mathrm{d}u \\
& - \int_{-\infty}^{-x_{t-1}} \{|x_{t-1} + u - [x_{t-1} + w_{t-1}]_+| - |x_{t-1} + u - [x_{t-1}]_+|\} f_u(u)\,\mathrm{d}u \\
=: {} & (\mathrm{IV})_{t-1} + (\mathrm{V})_{t-1}.
\end{aligned}
$$

Since $x_{t-1} + u \leq 0$ for all u in the range of integration in $(\mathrm{V})_{t-1}$, we have

$$
\begin{aligned}
(\mathrm{V})_{t-1} &= -\int_{-\infty}^{-x_{t-1}} \{([x_{t-1} + w_{t-1}]_+ - (x_{t-1} + u)) - ([x_{t-1}]_+ - (x_{t-1} + u)\} f_u(u)\,\mathrm{d}u \\
&= -\int_{-\infty}^{-x_{t-1}} \{[x_{t-1} + w_{t-1}]_+ - [x_{t-1}]_+\} f_u(u)\,\mathrm{d}u \\
&= -(\mathrm{III})_{t-1}.
\end{aligned}
$$

Deduce

$$
\begin{aligned}
(\mathrm{I})_{t-1} &= (\mathrm{IV})_{t-1} + (\mathrm{V})_{t-1} + (\mathrm{III})_{t-1} \\
&= \int_{-\infty}^{\infty} \{|x_{t-1} + u - [x_{t-1} + w_{t-1}]_+| - |x_{t-1} + u - [x_{t-1}]_+|\} f_u(u)\,\mathrm{d}u \\
&= \int_{-\infty}^{\infty} \{|u + [x_{t-1}]_- - ([x_{t-1} + w_{t-1}]_+ - [x_{t-1}]_+)| - |u + [x_{t-1}]_-|\} f_u(u)\,\mathrm{d}u \\
&= \int_{-\infty}^{\infty} \{|u - \ell_{t-1}(w_{t-1})| - |u|\} f_u(u - [x_{t-1}]_-)\,\mathrm{d}u \qquad \text{(C.20)}
\end{aligned}
$$

as per (C.16), where the final equality follows by a change of variables.

To prove (C.17), we first apply the identity (C.15) with $A = u$ and $B = \ell_{t-1}(w_{t-1})$ to the term inside the braces in the preceding display. Since $u = 0$ has Lebesgue measure zero, this yields

$$
\begin{aligned}
(\mathrm{I})_{t-1} = {} & -\int_{-\infty}^{\infty} \operatorname{sgn}(u) \ell_{t-1}(w_{t-1}) f_u(u - [x_{t-1}]_-)\,\mathrm{d}u \\
& + 2\int_{\infty}^{-\infty} [\ell_{t-1}(w_{t-1}) - u] \\
& \qquad \cdot [\mathbf{1}\{\ell_{t-1}(w_{t-1}) > u > 0\} - \mathbf{1}\{\ell_{t-1}(w_{t-1}) < u < 0\}] f_u(u - [x_{t-1}]_-)\,\mathrm{d}u \\
=: {} & (\mathrm{VI})_{t-1} + 2(\mathrm{VII})_{t-1}.
\end{aligned}
$$

For $(\mathrm{VI})_{t-1}$, we have

$$
\begin{aligned}
(\mathrm{VI})_{t-1} &= -\ell_{t-1}(w_{t-1}) \left[\int_0^{\infty} f_u(u - [x_{t-1}]_-)\,\mathrm{d}u - \int_{-\infty}^{0} f_u(u - [x_{t-1}]_-)\,\mathrm{d}u \right] \\
&= \ell_{t-1}(w_{t-1}) \{2F_u(-[x_{t-1}]_-) - 1\} = 2\ell_{t-1}(w_{t-1}) \{F_u(-[x_{t-1}]_-) - F_u(0)\}
\end{aligned}
$$

since $F_u(0) = \frac{1}{2}$ by Assumption C; this gives the first r.h.s. term in (C.17). Regarding $(\mathrm{VII})_{t-1}$, following the convention that $\int_a^b v(x)\,\mathrm{d}x = -\int_b^a v(x)\,\mathrm{d}x$ when $b < a$, and then making a change of variables,

$$\begin{aligned}(\mathrm{VII})_{t-1} &= \mathbf{1}\{\ell_{t-1}(w_{t-1}) > 0\} \int_0^{\ell_{t-1}(w_{t-1})} [\ell_{t-1}(w_{t-1}) - u] f_u(u - [x_{t-1}]_-)\,\mathrm{d}u \\ &\qquad - \mathbf{1}\{\ell_{t-1}(w_{t-1}) < 0\} \int_{\ell_{t-1}(w_{t-1})}^0 [\ell_{t-1}(w_{t-1}) - u] f_u(u - [x_{t-1}]_-)\,\mathrm{d}u \\ &= \int_0^{\ell_{t-1}(w_{t-1})} [\ell_{t-1}(w_{t-1}) - u] f_u(u - [x_{t-1}]_-)\,\mathrm{d}u \\ &= \int_{-[x_{t-1}]_-}^{\ell_{t-1}(w_{t-1}) - [x_{t-1}]_-} [\ell_{t-1}(w_{t-1}) - [x_{t-1}]_- - u] f_u(u)\,\mathrm{d}u.\end{aligned}$$

To put this into the required form, note that by integration by parts

$$\int_a^b (b-u) f_u(u)\,\mathrm{d}u = \int_a^b [F_u(u) - F_u(a)]\,\mathrm{d}u = G(b,a).$$

Hence

$$2(\mathrm{VII})_{t-1} = 2G\{\ell_{t-1}(w_{t-1}) - [x_{t-1}]_-, -[x_{t-1}]_-\}$$

corresponds to the second r.h.s. term in (C.17).

It remains to prove (C.18) and (C.19). Regarding (C.18), observe that

$$\ell_{t-1}(w_{t-1})\{F_u(-[x_{t-1}]_-) - F_u(0)\} = \{[x_{t-1} + w_{t-1}]_+ - [x_{t-1}]_+\}\{F_u(-[x_{t-1}]_-) - F_u(0)\}$$

is zero if $x_{t-1} \geq 0$, or if both $x_{t-1} < 0$ and $x_{t-1} + w_t < 0$. Suppose therefore that $x_{t-1} < 0$ and $x_{t-1} + w_t \geq 0$, in which case the preceding is equal to

$$(x_{t-1} + w_{t-1})\{F_u(-x_{t-1}) - F_u(0)\} = (x_{t-1} + w_{t-1}) f_u(-\tilde{x}_{t-1})(-x_{t-1}),$$

where the equality holds by the mean value theorem, for some $\tilde{x}_{t-1} \in [x_{t-1}, 0]$. From $w_t \geq -x_{t-1} > 0$, it follows that $|x_{t-1}| \leq |w_t|$. Since f_u is bounded by Assumption C, it follows that the preceding is bounded by $C|w_{t-1}|^2$, for some $C < \infty$. Hence

$$\begin{aligned}|\ell_{t-1}(w_{t-1})\{F_u(-[x_{t-1}]_-) - F_u(0)\}| &\leq C|w_{t-1}|^2 \mathbf{1}\{x_{t-1} < 0,\ x_{t-1} + w_t \geq 0\} \\ &\leq C|w_{t-1}|^2 \mathbf{1}\{x_{t-1} < 0\}.\end{aligned}$$

Regarding (C.19), observe that a Taylor expansion of $b \mapsto G(b,a)$ around $b = a$ yields

$$G(b,a) = \tfrac{1}{2} f_u(a)(b-a)^2 + \tfrac{1}{3!} f_u'(\tilde{b})(b-a)^3$$

where $\tilde{b}$ lies between a and b; and since f_u' is bounded by Assumption C, there exists a

$C < \infty$ such that

$$|G(b,a) - \tfrac{1}{2}f_u(a)(b-a)^2| \leq C(b-a)^3.$$

Thus (C.19) follows by taking $b = \ell_{t-1}(w_{t-1}) - [x_{t-1}]_-$, $a = -[x_{t-1}]_-$ and noting that $|\ell_{t-1}(w_{t-1})| \leq |w_{t-1}|$. □

C.2.2 Proof of Proposition C.1

Define

$$x_{t-1}(\rho) \coloneqq \alpha + \beta y_{t-1} + \boldsymbol{\phi}^\mathsf{T} \Delta \boldsymbol{y}_{t-1}$$

so that $x_{t-1}(\rho_{T,0}) = x_{t-1}$, where the latter is as in (C.2), and

$$w_t \coloneqq y_t - [x_{t-1}(\rho_{T,0})]_+ = [x_{t-1}(\rho_{T,0}) + u_t]_+ - [x_{t-1}(\rho_{T,0})]_+,$$

so that

$$\begin{aligned} y_t - [x_{t-1}(\rho)]_+ &= (y_t - [x_{t-1}(\rho_{T,0})]_+) - ([x_{t-1}(\rho)]_+ - [x_{t-1}(\rho_{T,0})]_+) \\ &= w_t - ([x_{t-1}(\rho)]_+ - [x_{t-1}(\rho_{T,0})]_+) \\ &\eqqcolon w_t - h_{t-1}(\rho, \rho_{T,0}). \end{aligned}$$

Then we can write the recentred and rescaled LAD criterion function as

$$Q_T(\rho) \coloneqq T^{-1}[S_T(\rho) - S_T(\rho_0)] = T^{-1}\sum_{t=1}^{T}[|w_t - h_{t-1}(\rho, \rho_{T,0})| - |w_t|]. \tag{C.21}$$

Henceforth set $h_{t-1}(\rho) \coloneqq h_{t-1}(\rho, \rho_{T,0})$, for a more compact notation.

Since $x \mapsto [x]_+$ is Lipschitz continuous (with Lipschitz constant equal to unity), $|w_t| \leq |u_t|$. Hence it is always the case that

$$|w_t - h_{t-1}(\rho)| - |w_t| \geq 0 - |w_t| \geq -|u_t|,$$

while if $|h_{t-1}(\rho)| > 3|u_t| \geq 3|w_t|$, then

$$|w_t - h_{t-1}(\rho)| - |w_t| > 2|u_t| - |w_t| \geq |u_t|.$$

Therefore for each summand in (C.21),

$$\begin{aligned} |w_t - h_{t-1}(\rho)| - |w_t| &= (|w_t - h_{t-1}(\rho)| - |w_t|)\mathbf{1}\{|h_{t-1}(\rho)| > 3|u_t|\} \\ &\qquad + (|w_t - h_{t-1}(\rho)| - |w_t|)\mathbf{1}\{|h_{t-1}(\rho)| \leq 3|u_t|\} \\ &\geq |u_t|\mathbf{1}\{|h_{t-1}(\rho)| > 3|u_t|\} - |u_t|\mathbf{1}\{|h_{t-1}(\rho)| \leq 3|u_t|\} \\ &= |u_t| - 2|u_t|\mathbf{1}\{|h_{t-1}(\rho)| \leq 3|u_t|\}. \end{aligned}$$

Hence we have the lower bound

$$
\begin{aligned}
Q_T(\rho) &\geq T^{-1}\sum_{t=1}^{T}|u_t| - 2T^{-1}\sum_{t=1}^{T}|u_t|\mathbf{1}\{|h_{t-1}(\rho)| \leq 3|u_t|\} \\
&\geq T^{-1}\sum_{t=1}^{T}|u_t| - 2\left(T^{-1}\sum_{t=1}^{T}u_t^2\right)^{1/2}\left(1 - T^{-1}\sum_{t=1}^{T}\mathbf{1}\{|h_{t-1}(\rho)| > 3|u_t|\}\right)^{1/2} \\
&=: T^{-1}\sum_{t=1}^{T}|u_t| - 2\left(T^{-1}\sum_{t=1}^{T}u_t^2\right)^{1/2}(1 - K_T(\rho))^{1/2}
\end{aligned}
$$

which depends on ρ only through $K_T(\rho)$.

We shall show below that for *every* non-negative $\kappa_T \to \infty$,

$$
\sup_{\{\rho\in\Pi \,|\, |\beta-\beta_{T,0}|\geq T^{-1/2}\kappa_T\}} |K_T(\rho) - 1| \xrightarrow{p} 0. \tag{C.22}
$$

It then follows by the LLN that

$$
\mathbb{P}\left\{\inf_{\{\rho\in\Pi \,|\, |\beta-\beta_0|\geq T^{-1/2}\kappa_T\}} Q_T(\rho) \geq \tfrac{1}{2}\mathbb{E}|u_1|\right\} \to 1.
$$

Since $Q_T(\rho_{T,0}) = 0$, it follows in particular that

$$
\mathbb{P}\{|\hat{\beta}_T - \beta_{T,0}| \geq \kappa_T T^{-1/2}\} \leq \mathbb{P}\left\{\inf_{\{\rho\in\Pi \,|\, |\beta-\beta_0|\geq T^{-1/2}\kappa_T\}} Q_T(\rho) \leq 0\right\} \to 0.
$$

Deduce that $T^{1/2}(\hat{\beta}_T - \beta_{T,0}) = o_p(\kappa_T)$ for *every* sequence $\kappa_T \to \infty$, whence $T^{1/2}(\hat{\beta}_T - \beta_{T,0}) = O_p(1)$.

It remains to prove (C.22). We first note that since the parameter space Π is compact, there exists a C such that

$$
|(\alpha - \alpha_0) + (\boldsymbol{\phi} - \boldsymbol{\phi}_0)^{\mathsf{T}}\Delta\boldsymbol{y}_{t-1}| \leq C(1 + \|\Delta\boldsymbol{y}_{t-1}\|).
$$

Hence we have the following lower bound,

$$
\begin{aligned}
|h_{t-1}(\rho)| &= |[x_{t-1}(\rho)]_+ - [x_{t-1}(\rho_{T,0})]_+| \\
&\geq |x_{t-1}(\rho) - x_{t-1}(\rho_{T,0})|\mathbf{1}\{x_{t-1}(\rho) > 0\}\mathbf{1}\{x_{t-1}(\rho_{T,0}) > 0\} \\
&= |(\beta - \beta_{T,0})y_{t-1} + (\alpha - \alpha_0) + (\boldsymbol{\phi} - \boldsymbol{\phi}_0)^{\mathsf{T}}\Delta\boldsymbol{y}_{t-1}|\mathbf{1}\{x_{t-1}(\rho) > 0\}\mathbf{1}\{x_{t-1}(\rho_{T,0}) > 0\} \\
&\geq [|\beta - \beta_{T,0}|y_{t-1} - C(1 + \|\Delta\boldsymbol{y}_{t-1}\|)]\mathbf{1}\{x_{t-1}(\rho_{T,0}) > 0\} \\
&\qquad \cdot \mathbf{1}\{\beta y_{t-1} - C(1 + \|\Delta\boldsymbol{y}_{t-1}\|) > 0\},
\end{aligned}
$$

since $y_{t-1} \geq 0$, and so

$$\mathbf{1}\{|h_{t-1}(\rho)| > 3|u_t|\} \geq \mathbf{1}\{|\beta - \beta_{T,0}|y_{t-1} - C(1 + \|\Delta \boldsymbol{y}_{t-1}\|) > 3|u_t|\} \\ \mathbf{1}\{\beta y_{t-1} - C(1 + \|\Delta \boldsymbol{y}_{t-1}\|) > 0\}\mathbf{1}\{x_{t-1}(\rho_0) > 0\}.$$

Observe that the r.h.s. is increasing in $|\beta - \beta_{T,0}|$ and β. Suppose that $\beta \geq \beta_{T,0} + T^{-1/2}\kappa_T$. Then for all such $\rho = (\alpha, \beta, \boldsymbol{\phi})$,

$$\begin{aligned} K_T(\rho) &\geq T^{-1}\sum_{t=1}^{T} \mathbf{1}\{T^{-1/2}y_{t-1} - \kappa_T^{-1}C(1 + \|\Delta \boldsymbol{y}_{t-1}\|) - 3\kappa_T^{-1}|u_t| > 0\} \\ &\qquad \cdot \mathbf{1}\{\beta_{T,0}T^{-1/2}y_{t-1} - T^{-1/2}C(1 + \|\Delta \boldsymbol{y}_{t-1}\|) > 0\}\mathbf{1}\{T^{-1/2}x_{t-1}(\rho_0) > 0\} \\ &\geq T^{-1}\sum_{t=1}^{T} g_M[T^{-1/2}y_{t-1} - \kappa_T^{-1}C(1 + \|\Delta \boldsymbol{y}_{t-1}\|) - 3\kappa_T^{-1}|u_t|] \\ &\qquad \cdot g_M[\beta_{T,0}T^{-1/2}y_{t-1} - T^{-1/2}C(1 + \|\Delta \boldsymbol{y}_{t-1}\|)]g_M[T^{-1/2}x_{t-1}(\rho_0)] \\ &=: L_T^{(1)}(M) \end{aligned} \tag{C.23}$$

where g_M is a Lipschitz function such that $\mathbf{1}\{x > 0\} \geq g_M(x) \geq \mathbf{1}\{x > M\}$. Similarly, if $\beta \in [\frac{1}{2}\beta_{T,0}, \beta_{T,0} - T^{-1/2}\kappa_T]$, we have the lower bound

$$\begin{aligned} K_T(\rho) &\geq T^{-1}\sum_{t=1}^{T} g_M[T^{-1/2}y_{t-1} - \kappa_T^{-1}C(1 + \|\Delta \boldsymbol{y}_{t-1}\|) - 3\kappa_T^{-1}|u_t|] \\ &\qquad \cdot g_M[\tfrac{1}{2}\beta_{T,0}T^{-1/2}y_{t-1} - T^{-1/2}C(1 + \|\Delta \boldsymbol{y}_{t-1}\|)]g_M[T^{-1/2}x_{t-1}(\rho_0)] \\ &=: L_T^{(2)}(M). \end{aligned} \tag{C.24}$$

Finally, suppose $\beta \leq \frac{1}{2}\beta_{T,0}$: we need a slightly different lower bound for $|h_{t-1}(\rho)|$ in this case. We first note that if $a > 0$ and $b < \frac{3}{4}a$, then $|[a]_+ - [b]_+| > \frac{1}{4}a$. Hence

$$\begin{aligned} |h_{t-1}(\rho)| &= |[x_{t-1}(\rho_{T,0})]_+ - [x_{t-1}(\rho)]_+| \\ &\geq \tfrac{1}{4}x_{t-1}(\rho_{T,0})\mathbf{1}\{x_{t-1}(\rho_{T,0}) > 0\}\mathbf{1}\{x_{t-1}(\rho) < \tfrac{3}{4}x_{t-1}(\rho_{T,0})\} \\ &= \tfrac{1}{4}[\alpha_{T,0} + \beta_{T,0}y_{t-1} + \boldsymbol{\phi}_0^\mathsf{T}\Delta \boldsymbol{y}_{t-1}]\mathbf{1}\{x_{t-1}(\rho_{T,0}) > 0\} \\ &\qquad \mathbf{1}\{(\tfrac{3}{4}\beta_{T,0} - \beta)y_{t-1} + (\tfrac{3}{4}\alpha_{T,0} - \alpha) + (\tfrac{3}{4}\boldsymbol{\phi}_0 - \boldsymbol{\phi})^\mathsf{T}\Delta \boldsymbol{y}_{t-1} > 0\} \\ &\geq \tfrac{1}{4}[\beta_{T,0}y_{t-1} + \alpha_{T,0} + \boldsymbol{\phi}_0^\mathsf{T}\Delta \boldsymbol{y}_{t-1}]\mathbf{1}\{x_{t-1}(\rho_{T,0}) > 0\} \\ &\qquad \mathbf{1}\{\tfrac{1}{4}\beta_{T,0}y_{t-1} - C_1(1 + \|\Delta \boldsymbol{y}_{t-1}\|) > 0\}. \end{aligned}$$

where the final inequality holds since $y_{t-1} \geq 0$, and $\frac{3}{4}\beta_{T,0} - \beta \geq \frac{1}{4}\beta_{T,0}$. Therefore,

$$\mathbf{1}\{|h_{t-1}(\rho)| > 3|u_t|\} \geq \mathbf{1}\{\beta_{T,0}T^{-1/2}y_{t-1} + T^{-1/2}\alpha_{T,0} + \boldsymbol{\phi}_0^\mathsf{T}T^{-1/2}\Delta \boldsymbol{y}_{t-1} - 12T^{-1/2}|u_t| > 0\} \\ \cdot \mathbf{1}\{\tfrac{1}{4}\beta_{T,0}T^{-1/2}y_{t-1} - T^{-1/2}C_1(1 + \|\Delta \boldsymbol{y}_{t-1}\|) > 0\}\mathbf{1}\{T^{-1/2}x_{t-1}(\rho_{T,0}) > 0\},$$

whence

$$K_T(\rho) \geq T^{-1} \sum_{t=1}^{T} g_M[\beta_0 T^{-1/2} y_{t-1} + T^{-1/2}\alpha_0 + \boldsymbol{\phi}_0^{\mathsf{T}} T^{-1/2} \Delta \boldsymbol{y}_{t-1} - 12T^{-1/2}|u_t|]$$
$$\cdot g_M[\tfrac{1}{4}\beta_{T,0} T^{-1/2} y_{t-1} - T^{-1/2} C_1(1 + \|\Delta \boldsymbol{y}_{t-1}\|)] g_M[T^{-1/2} x_{t-1}(\rho_{T,0})]$$
$$=: L_T^{(3)}(M). \tag{C.25}$$

It follows from (C.23)–(C.25) that

$$1 \geq K_T(\rho) \geq \min\{L_T^{(i)}(M)\}_{i=1}^{3}$$

for all $\rho \in \Pi$ such that $|\beta - \beta_{T,0}| \geq T^{-1/2}\kappa_T$. Hence it remains to show that $L_T^{(i)}(M) \xrightarrow{p} 1$ for each $i \in \{1, 2, 3\}$. We give the proof only when $i = 1$; the proof in the other cases is analogous. To that end, we note that by the continuous mapping theorem (CMT), Lemma A.1(i), and Theorem 3.2 in Bykhovskaya and Duffy (2024), the finite-dimensional distributions of the process

$$\xi_T(\tau) \coloneqq g_M[T^{-1/2} y_{\lfloor \tau T \rfloor - 1} - \kappa_T^{-1} C(1 + \|\Delta \boldsymbol{y}_{\lfloor \tau T \rfloor - 1}\|) - 3\kappa_T^{-1}|u_{\lfloor \tau T \rfloor}|]$$
$$\cdot g_M[\beta_{T,0} T^{-1/2} y_{\lfloor \tau T \rfloor - 1} - T^{-1/2} C(1 + \|\Delta \boldsymbol{y}_{\lfloor \tau T \rfloor - 1}\|)] g_M[T^{-1/2} x_{\lfloor \tau T \rfloor - 1}(\rho_0)],$$

for $\tau \in [0, 1]$, converge weakly to those of

$$\xi(\tau) \coloneqq g_M^3[Y(\tau)],$$

since $\beta_{T,0} \to 1$. (This cannot be strengthened to weak convergence in $D[0, 1]$, since we may have $\kappa_T \to \infty$ arbitrarily slowly.) Since g_M is bounded by unity, $\sup_{\tau \in [0,1]} \mathbb{E}|\xi_T(\tau)| \leq 1$. Finally, since g_M is Lipschitz and bounded by unity, there exists a constant C_M depending only on M such that

$$\sup_{|\tau' - \tau| \leq \epsilon} \mathbb{E}|\xi_T(\tau') - \xi_T(\tau)|$$
$$\leq C_M \Bigg[\sup_{|\tau' - \tau| \leq \epsilon} \mathbb{E} \min\{|T^{-1/2} y_{\lfloor \tau' T \rfloor - 1} - T^{-1/2} y_{\lfloor \tau T \rfloor - 1}|, 1\}$$
$$+ \max\{\kappa_T^{-1}, T^{-1/2}\} \max_{1 \leq t \leq T} \mathbb{E}\|\Delta \boldsymbol{y}_t\| + \kappa_T^{-1} \max_{1 \leq t \leq T} \mathbb{E}|u_t| \Bigg] \to 0$$

as $T \to \infty$ and then $\epsilon \to 0$, by Lemma A.1(i) and since $T^{-1/2} y_{\lfloor \tau T \rfloor}$ is tight in $D[0, 1]$ by Theorem 3.2 in Bykhovskaya and Duffy (2024). It follows by Theorem IX.7.1 in Gikhman and Skorokhod (1969) that

$$L_T(M) = \int_0^1 \xi_T(\tau)\,\mathrm{d}\tau \xrightarrow{d} \int_0^1 \xi(\tau)\,\mathrm{d}\tau \geq \int_0^1 \mathbf{1}\{Y(\tau) > M\}\,\mathrm{d}\tau \xrightarrow{p} 1$$

as $T \to \infty$ and then $M \to 0$, by Lemma A.2(i). □

C.2.3 Proof of Proposition C.2

(i). Defining

$$\nu(b, a) := \int_{-\infty}^{\infty} \{|u - b| - |u|\} f_u(u - a)\,\mathrm{d}u,$$

we have by Lemma C.4, in particular by (C.16), that

$$\begin{aligned}
\bar{S}_T(\varpi) - \bar{S}_T(0) &= \frac{1}{T}\sum_{t=1}^{T} \nu(\ell_{t-1}(w_{t-1}), [x_{t-1}]_-) \\
&= \frac{1}{T}\sum_{t=1}^{T} \nu(\ell_{t-1}(w_{t-1}), 0) - \frac{1}{T}\sum_{t=1}^{T} \nu(\ell_{t-1}(w_{t-1}), 0)\mathbf{1}\{x_{t-1} < 0\} \\
&\qquad + \frac{1}{T}\sum_{t=1}^{T} \nu(\ell_{t-1}(w_{t-1}), [x_{t-1}]_-)\mathbf{1}\{x_{t-1} < 0\} \\
&=: (\mathrm{I})_T - (\mathrm{II})_T + (\mathrm{III})_T
\end{aligned} \tag{C.26}$$

with $w_{t-1} = \varpi^{\mathsf{T}} \mathcal{Z}_{t-1,T}$. Since $|\nu(b, a)| \leq |b|$ and $|\ell_{t-1}(w_{t-1})| \leq |w_{t-1}|$, it follows that for all $\omega \in \bar{\Pi}$,

$$\begin{aligned}
|(\mathrm{II})_T| &\leq \frac{1}{T}\sum_{t=1}^{T} |\ell_{t-1}(\varpi^{\mathsf{T}}\mathcal{Z}_{t-1,T})|\mathbf{1}\{x_{t-1} < 0\} \leq \frac{1}{T}\sum_{t=1}^{T} |\varpi^{\mathsf{T}}\mathcal{Z}_{t-1,T}|\mathbf{1}\{x_{t-1} < 0\} \\
&\leq_{(1)} C\frac{1}{T}\sum_{t=1}^{T} \|\mathcal{Z}_{t-1,T}\|\mathbf{1}\{x_{t-1} < 0\} \\
&\leq_{(2)} C\left(\frac{1}{T}\sum_{t=1}^{T} \|\mathcal{Z}_{t-1,T}\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\{x_{t-1} < 0\}\right)^{1/2} =_{(3)} o_p(1),
\end{aligned} \tag{C.27}$$

where $\leq_{(1)}$ follows by the compactness of $\bar{\Pi}$, $\leq_{(2)}$ by the Cauchy–Schwarz inequality, and $=_{(3)}$ by Lemmas A.2(ii) and A.3(i). An identical argument shows that $(\mathrm{III})_T = o_p(1)$.

To handle $(\mathrm{I})_T$, we need to consider the map $b \mapsto \nu(b, 0) =: \nu(b)$, which is minimised at $b = \mathrm{med}(u_t) = 0$ by Assumption C. Since $b \mapsto |u - b|$ is differentiable at all $b \in \mathbb{R}\backslash\{u\}$, with bounded derivatives, the function $b \mapsto \nu(b)$ has first derivative

$$\begin{aligned}
\nu'(b) &= -\int_{-\infty}^{\infty} \mathrm{sgn}(u - b) f_u(u)\,\mathrm{d}u = \mathbb{P}\{u_t < b\} - \mathbb{P}\{u_t > b\} \\
&= F_u(b) - [1 - F_u(b)] = 2F_u(b) - 1
\end{aligned}$$

so that $\nu'(b)$ is (weakly) increasing in b, with $\nu'(b) \geq 0$ when $b \geq 0$, and $\nu'(b) \leq 0$ when $b \leq 0$. Since $f_u(0) > 0$ by Assumption C, these inequalities hold strictly for b in a neighbourhood of zero, and so it follows that for any given $\tau > 0$,

$$\nu(b) \geq \gamma_\tau \mathbf{1}\{|b| \geq \tau\}$$

for all $b \in \mathbb{R}$, with $\gamma_\tau \coloneqq \min\{\nu(\tau), \nu(-\tau)\} > 0$. Hence

$$(\mathrm{I})_T = \frac{1}{T}\sum_{t=1}^{T} \nu[\ell_{t-1}(\varpi^\mathsf{T} \mathcal{Z}_{t-1,T})] \geq \gamma_\tau \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\{|\varpi^\mathsf{T} \mathcal{Z}_{t-1,T}| \geq \tau\}. \tag{C.28}$$

It remains to lower bound the sum on the r.h.s.

To that end, we note that by Hölder's inequality,

$$\begin{aligned} &T^{-1}\sum_{t=1}^{T} \mathbf{1}\{|\varpi^\mathsf{T} \mathcal{Z}_{t-1,T}| \geq \tau\} \|\mathcal{Z}_{t-1,T}\|^2 \\ &\qquad \leq \left(T^{-1}\sum_{t=1}^{T} \mathbf{1}\{|\varpi^\mathsf{T} \mathcal{Z}_{t-1,T}| \geq \tau\}\right)^{\delta_u/(2+\delta_u)} \left(T^{-1}\sum_{t=1}^{T} \|\mathcal{Z}_{t-1,T}\|^{2+\delta_u}\right)^{2/(2+\delta_u)} \end{aligned}$$

where

$$\begin{aligned} \xi_T &\coloneqq T^{-1}\sum_{t=1}^{T} \|\mathcal{Z}_{t-1,T}\|^{2+\delta_u} \\ &\qquad \leq C\left(1 + T^{-1}\sum_{t=1}^{T} |T^{-1/2} y_{t-1}|^{2+\delta_u} + T^{-1}\sum_{t=1}^{T} \|\Delta \boldsymbol{y}_{t-1}\|^{2+\delta_u}\right) = O_p(1) \end{aligned}$$

by Lemma A.1(i) and Theorem 3.2 in Bykhovskaya and Duffy (2024). Further, for all $\varpi \in \bar{\Pi}$,

$$\begin{aligned} &T^{-1}\sum_{t=1}^{T} \mathbf{1}\{|\varpi^\mathsf{T} \mathcal{Z}_{t-1,T}| > \tau\} \|\mathcal{Z}_{t-1,T}\|^2 \geq T^{-1}\sum_{t=1}^{T} \mathbf{1}\{|\varpi^\mathsf{T} \mathcal{Z}_{t-1,T}| > \tau\} |\varpi^\mathsf{T} \mathcal{Z}_{t-1,T}|^2 \|\varpi\|^{-2} \\ &= \|\varpi\|^{-2}\left[\varpi^\mathsf{T}\left(T^{-1}\sum_{t=1}^{T} \mathcal{Z}_{t-1,T}\mathcal{Z}_{t-1,T}^\mathsf{T}\right)\varpi - T^{-1}\sum_{t=1}^{T} \mathbf{1}\{|\varpi^\mathsf{T} \mathcal{Z}_{t-1,T}| \leq \tau\} |\varpi^\mathsf{T} \mathcal{Z}_{t-1,T}|^2\right] \\ &\geq \lambda_{\min}(Q_T) - \|\varpi\|^{-2}\tau^2, \end{aligned}$$

where $\lambda_{\min}(A)$ denotes the smallest eigenvalue of the positive semi-definite matrix A, and

$$Q_T \coloneqq T^{-1}\sum_{t=1}^{T} \mathcal{Z}_{t-1,T}\mathcal{Z}_{t-1,T}^\mathsf{T} = \sum_{t=1}^{T} z_{t-1,T} z_{t-1,T}^\mathsf{T} \overset{d}{\to} Q_{ZZ} \tag{C.29}$$

by Lemma A.3(i), where Q_{ZZ} is positive definite. Hence for $\delta > 0$ as in the statement of part (i) the proposition,

$$
\begin{aligned}
\inf_{\|\varpi\|\geq\delta} \frac{1}{T}\sum_{t=1}^{T}&\mathbf{1}\{|\varpi^{\mathsf{T}}\mathcal{Z}_{t-1,T}|\geq\tau\} \\
&\geq \left(\inf_{\|\varpi\|\geq\delta}\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}\{|\varpi^{\mathsf{T}}\mathcal{Z}_{t-1,T}|\geq\tau\}\|\mathcal{Z}_{t-1,T}\|^2\right)^{(2+\delta_u)/\delta_u}\xi_T^{-2/\delta_u} \\
&\geq \left(\lambda_{\min}(Q_T)-\delta^{-2}\tau^2\right)^{(2+\delta_u)/\delta_u}\xi_T^{-2/\delta_u}
\end{aligned}
$$

it follows that for $\epsilon>0$ as in the statement of part (i) of the proposition, we may choose $\kappa,\tau>0$ sufficiently small that

$$
\begin{aligned}
\mathbb{P}\left\{\inf_{\|\varpi\|\geq\delta}\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}\{|\varpi^{\mathsf{T}}\mathcal{Z}_{t-1,T}|\geq\tau\}>\kappa\right\}& \\
\geq \mathbb{P}\left\{\left(\lambda_{\min}(Q_T)-\delta^{-2}\tau^2\right)^{(2+\delta_u)/\delta_u}>\kappa\xi_T^{2/\delta_u}\right\}&>1-\epsilon/2
\end{aligned}
$$

for all T sufficiently large. In view of (C.26), (C.27) and (C.28), it follows that

$$
\liminf_{T\to\infty}\mathbb{P}\left\{\inf_{\|\varpi\|\geq\delta}[\bar{S}_T(\varpi)-\bar{S}_T(0)]>\kappa\gamma_\tau\right\}>1-\epsilon,
$$

whereupon the result follows by taking $\eta=\kappa\gamma_\tau>0$.

(ii). For suitable choices of w_{t-1}, both $\bar{S}_T(\varpi)-\bar{S}_T(0)$ and $\bar{\mathbb{S}}_T(r)-\bar{\mathbb{S}}_T(0)$ can be expressed in terms of

$$
\sum_{t=1}^{T}\mathbb{E}_{t-1}\{|y_t-[x_{t-1}+w_{t-1}]_+|-|y_t-[x_{t-1}]_+| \tag{C.30}
$$

$$
\begin{aligned}
&=2\sum_{t=1}^{T}\ell_{t-1}(w_{t-1})[F_u(-[x_{t-1}]_-)-F_u(0)]+2\sum_{t=1}^{T}G\{\ell_{t-1}(w_{t-1})-[x_{t-1}]_-,-[x_{t-1}]_-\} \\
&=: 2(\mathrm{IV})_T+2(\mathrm{V})_T,
\end{aligned} \tag{C.31}
$$

where the first equality follows by Lemma C.4. That same result also entails that

$$
|(\mathrm{IV})_T|\leq C\sum_{t=1}^{T}|w_{t-1}|^2\mathbf{1}\{x_{t-1}<0\} \tag{C.32}
$$

and

$$
\left|(\mathrm{V})_T-\frac{1}{2}\sum_{t=1}^{T}f_u(-[x_{t-1}]_-)\ell_{t-1}^2(w_{t-1})\right|\leq C\sum_{t=1}^{T}|w_{t-1}|^3. \tag{C.33}
$$

We now use (C.31)–(C.33) to prove parts (ii) and (iii) of the proposition.

Noting that $\bar{S}_T(\varpi) - \bar{S}_T(0)$ is equal to (C.30) times T^{-1}, with $w_{t-1} = \varpi^{\mathsf{T}} \mathcal{Z}_{t-1,T}$, we obtain from (C.31)–(C.33) that

$$\begin{aligned}\bar{S}_T(\varpi) - \bar{S}_T(0) \geq \frac{1}{T}\sum_{t=1}^{T} f_u(-[x_{t-1}]_-)\ell_{t-1}^2(\varpi^{\mathsf{T}} \mathcal{Z}_{t-1,T}) \\ - \|\varpi\|^2 \frac{2C}{T}\sum_{t=1}^{T}\|\mathcal{Z}_{t-1,T}\|^2 \mathbf{1}\{x_{t-1} < 0\} - \|\varpi\|^3 \frac{2C}{T}\sum_{t=1}^{T}\|\mathcal{Z}_{t-1,T}\|^3 \\ =: (\mathrm{VI})_T - (\mathrm{VII})_T - (\mathrm{VIII})_T. \end{aligned} \tag{C.34}$$

To determine the limit of $(\mathrm{VI})_T$, note $\max_{1\leq t\leq T}\|\mathcal{Z}_{t-1,T}\| = O_p(T^{1/(2+\delta_u)})$ by Lemma A.3(ii), and so for $\kappa \coloneqq 1/(2+\delta_u/2) < 1/2$ is bounded by T^{κ}, w.p.a.1. Define

$$\mathbf{1}_{t-1} \coloneqq \mathbf{1}\{x_{t-1} > T^{\kappa},\ \|\mathcal{Z}_{t-1,T}\| \leq T^{\kappa}\}, \qquad \mathbf{1}_{t-1}^c \coloneqq 1 - \mathbf{1}_{t-1}$$

and observe that when $\mathbf{1}_{t-1} = 1$, we must have $x_{t-1} > 0$ and

$$x_{t-1} + \varpi^{\mathsf{T}} \mathcal{Z}_{t-1,T} \geq x_{t-1} - \|\varpi\|\|\mathcal{Z}_{t-1,T}\| > T^{\kappa}(1 - \|\varpi\|),$$

and so the preceding must be non-negative if $\|\varpi\| \leq 1$. Deduce that for all such ϖ,

$$\begin{aligned} f_u(-[x_{t-1}]_-)\ell_{t-1}^2(\varpi^{\mathsf{T}} \mathcal{Z}_{t-1,T})\mathbf{1}_{t-1} &= f_u(-[x_{t-1}]_-)\{[x_{t-1} + \varpi^{\mathsf{T}} \mathcal{Z}_{t-1,T}]_+ - [x_{t-1}]_+\}^2 \mathbf{1}_{t-1} \\ &= f_u(0)\{(x_{t-1} + \varpi^{\mathsf{T}} \mathcal{Z}_{t-1,T}) - x_{t-1}\}^2 \mathbf{1}_{t-1} \\ &= f_u(0)(\varpi^{\mathsf{T}} \mathcal{Z}_{t-1,T})^2 \mathbf{1}_{t-1}. \end{aligned}$$

Hence

$$\begin{aligned}(\mathrm{VI})_T = f_u(0)\frac{1}{T}\sum_{t=1}^{T}(\varpi^{\mathsf{T}} \mathcal{Z}_{t-1,T})^2 + \Bigg[-\frac{1}{T}\sum_{t=1}^{T} f_u(0)(\varpi^{\mathsf{T}} \mathcal{Z}_{t-1,T})^2 \mathbf{1}_{t-1}^c \\ + \frac{1}{T}\sum_{t=1}^{T} f_u(-[x_{t-1}]_-)\ell_{t-1}^2(\varpi^{\mathsf{T}} \mathcal{Z}_{t-1,T})\mathbf{1}_{t-1}^c\Bigg] =: (\mathrm{IX})_T + (\mathrm{X})_T. \end{aligned}$$

Since f_u is bounded, and

$$\ell_{t-1}^2(\varpi^{\mathsf{T}} \mathcal{Z}_{t-1,T}) = |[x_{t-1} + \varpi^{\mathsf{T}} \mathcal{Z}_{t-1,T}]_+ - [x_{t-1}]_+|^2 \leq \|\varpi\|^2\|\mathcal{Z}_{t-1,T}\|^2 = \|\varpi\|^2 T\|z_{t-1,T}\|^2$$

we may bound

$$|(\mathrm{X})_T| \leq C\|\varpi\|^2 \sum_{t=1}^{T}\|z_{t-1,T}\|^2 \mathbf{1}_{t-1}^c$$

$$\leq C\|\varpi\|^2 \left(\sum_{t=1}^{T}\|z_{t-1,T}\|^4\right)^{1/2} \left(\sum_{t=1}^{T}\mathbf{1}_{t-1}^c\right)^{1/2} = \|\varpi\|^2 o_p(1),$$

by the CS inequality, where the final equality holds by Lemma A.3(v), and

$$T^{-1}\sum_{t=1}^{T}\mathbf{1}_{t-1}^c \leq T^{-1}\sum_{t=1}^{T}\mathbf{1}\{T^{-1/2}x_{t-1} \leq T^{\kappa-1/2}\} + T^{-1}\sum_{t=1}^{T}\mathbf{1}\{\|\mathcal{Z}_{t-1,T}\| > T^{\kappa}\}$$
$$\leq_{(1)} T^{-1}\sum_{t=1}^{T}\mathbf{1}\{T^{-1/2}x_{t-1} \leq \gamma\} + \mathbf{1}\left\{\max_{1\leq t\leq T}\|\mathcal{Z}_{t-1,T}\| > T^{\kappa}\right\} \xrightarrow{p}_{(2)} 0 \quad \text{(C.35)}$$

where $\leq_{(1)}$ holds for any fixed $\gamma > 0$, for all T sufficiently large, and $\xrightarrow{p}_{(2)}$ holds as $T \to \infty$ and then $\gamma \to 0$, by Lemma A.2(ii) and the noted stochastic order of $\max_{1\leq t\leq T}\|\mathcal{Z}_{t-1,T}\|$. Recalling the definition of Q_T in (C.29) above, we thus have

$$(\text{VI})_T \geq (\text{IX})_T - \|\varpi\|^2 o_p(1) = f_u(0)\varpi^{\mathsf{T}}[Q_T + o_p(1)]\varpi \quad \text{(C.36)}$$

Returning now to (C.34), we see that by the same argument that was applied to $(\text{X})_T$, with Lemma A.2(ii) playing the role of (C.35), we obtain

$$(\text{VII})_T = C\|\varpi\|^2 o_p(1). \quad \text{(C.37)}$$

Regarding $(\text{VIII})_T$, it follows from Lemma A.1(i) (since $2+\delta_u > 3$ under Assumption C), and Theorem 3.2 in Bykhovskaya and Duffy (2024), that

$$|(\text{VIII})_T| = \|\varpi\|^3 \frac{C}{T}\sum_{t=1}^{T}\|\mathcal{Z}_{t-1,T}\|^3 = \|\varpi\|^3 O_p(1). \quad \text{(C.38)}$$

It follows from (C.34), (C.36), (C.37) and (C.38) that there exist random sequences $\xi_{1,T} = o_p(1)$ and $\xi_{2,T} = O_p(1)$ such that

$$\bar{S}_T(\varpi) - \bar{S}_T(0) \geq f_u(0)\varpi^{\mathsf{T}}[Q_T + \xi_{1,T}]\varpi + \xi_{2,T}\|\varpi\|^3$$
$$\geq f_u(0)[\lambda_{\min}(Q_T) + \xi_{1,T}]\|\varpi\|^2 + \xi_{2,T}\|\varpi\|^3$$

for all $\varpi \in \bar{\Pi}$. Now let $\epsilon > 0$ be as in the statement of part (ii) of the proposition. Since $Q_T \xrightarrow{d} Q_{ZZ}$, which is a.s. positive definite, we may choose $\eta > 0$ sufficiently small that

$$\mathbb{P}\{f_u(0)[\lambda_{\min}(Q_T) + \xi_{1,T}] < 2\eta\} < \epsilon/2$$

for all T sufficiently large. Moreover, we may take $\delta > 0$ such that

$$\mathbb{P}\{\|\varpi\|^3\xi_{2,T} < -\eta\|\varpi\|^2,\ \forall\|\varpi\| \leq \delta\} \leq \mathbb{P}\{|\xi_{2,T}| > \eta\delta^{-1}\} < \epsilon/2,$$

for all T sufficiently large, and therefore

$$\begin{aligned}
&\mathbb{P}\{\bar{S}_T(\varpi) - \bar{S}_T(0) \geq \eta\|\varpi\|^2,\ \forall\|\varpi\| \leq \delta\} \\
&\qquad\qquad \geq \mathbb{P}\{f_u(0)[\lambda_{\min}(Q_T) + \xi_{1,T}]\|\varpi\|^2 \geq 2\eta\|\varpi\|^2,\ \forall\|\varpi\| \leq \delta\} \\
&\qquad\qquad\qquad\qquad - \mathbb{P}\{\|\varpi\|^3\xi_{2,T} < -\eta\|\varpi\|^2,\ \forall\|\varpi\| \leq \delta\} \\
&\qquad\qquad \geq \mathbb{P}\{f_u(0)[\lambda_{\min}(Q_T) + \xi_{1,T}] \geq 2\eta\} - \mathbb{P}\{|\xi_{2,T}| > \eta\delta^{-1}\} > 1 - \epsilon
\end{aligned}$$

for all T sufficiently large.

(iii). Noting in this case that $\bar{\mathbb{S}}_T(r) - \bar{\mathbb{S}}_T(0)$ equals (C.30) with $w_{t-1} = r^\mathsf{T} z_{t-1,T}$, and recalling that

$$|\ell_{t-1}(r^\mathsf{T} z_{t-1,T})| = |[x_{t-1} + r^\mathsf{T} z_{t-1,T}]_+ - [x_{t-1}]_+| \leq |r^\mathsf{T} z_{t-1,T}| \tag{C.39}$$

we obtain from (C.31)–(C.33) that for all $r \in \mathbb{R}^{k+1}$,

$$\begin{aligned}
&\left|[\bar{\mathbb{S}}_T(r) - \bar{\mathbb{S}}_T(0)] - \sum_{t=1}^{T} f_u(-[x_{t-1}]_-)\ell_{t-1}^2(r^\mathsf{T} z_{t-1,T})\right| \qquad\qquad (\text{C.40}) \\
&\leq C\sum_{t=1}^{T}|r^\mathsf{T} z_{t-1,T}|^2\mathbf{1}\{x_{t-1} < 0\} + C\sum_{t=1}^{T}|r^\mathsf{T} z_{t-1,T}|^3 \\
&\leq C\|r\|^2\left(\sum_{t=1}^{T}\|z_{t-1,T}\|^4\right)^{1/2}\left(\sum_{t=1}^{T}\mathbf{1}\{x_{t-1} < 0\}\right)^{1/2} + C\|r\|^3\max_{1\leq s\leq T}\|z_{s-1,T}\|\sum_{t=1}^{T}\|z_{t-1,T}\|^2 \\
&= o_p(1)
\end{aligned}$$

uniformly on compact subsets of $\mathbb{R}^{k+1}$, by the CS inequality, Theorem 3.2 of Bykhovskaya and Duffy (2024) and Lemmas A.2(ii), A.3(i), A.3(ii) and A.3(v).

It remains to consider the l.h.s. of (C.40). Let

$$\mathbf{1}_{t-1}(r) \coloneqq \mathbf{1}\{x_{t-1} \geq 0,\ x_{t-1} + r^\mathsf{T} z_{t-1,T} \geq 0\}, \qquad \mathbf{1}_{t-1}^c(r) \coloneqq 1 - \mathbf{1}_{t-1}(r). \tag{C.41}$$

We note that

$$\begin{aligned}
\mathbf{1}_{t-1}^c(r) &= \mathbf{1}\{x_{t-1} < 0\} + \mathbf{1}\{x_{t-1} \geq 0\}\mathbf{1}\{x_{t-1} + r^\mathsf{T} z_{t-1,T} < 0\} \\
&= \mathbf{1}\{x_{t-1} < 0\} + \mathbf{1}\{r^\mathsf{T} z_{t-1,T} < -x_{t-1} \leq 0\} \\
&\leq \mathbf{1}\{x_{t-1} < 0\} + \mathbf{1}\{|x_{t-1}| \leq \|r^\mathsf{T} z_{t-1,T}\|\} \\
&\leq \mathbf{1}\{x_{t-1} < 0\} + \mathbf{1}\{|x_{t-1}| \leq 1\} + \mathbf{1}\{1 \leq \|r^\mathsf{T} z_{t-1,T}\|\} \\
&\leq 2\cdot\mathbf{1}\{x_{t-1} \leq 1\} + \mathbf{1}\left\{1 \leq \|r\|\sup_{1\leq s\leq T}\|z_{s-1,T}\|\right\} =: 2\cdot\mathbf{1}_{1,t-1}^c + \mathbf{1}_2^c(r),
\end{aligned} \tag{C.42}$$

where

$$\sum_{t=1}^{T} \mathbf{1}_{1,t-1}^{c} = \sum_{t=1}^{T} \mathbf{1}\{x_{t-1} \leq 1\} = o_p(T) \tag{C.43}$$

and, since $\sup_{1\leq s\leq T}\|z_{s-1,T}\| = o_p(1)$, $\mathbf{1}_2^c(r) = o_p(1)$ uniformly on compact subsets of $\mathbb{R}^{k+1}$. Now decompose

$$\begin{aligned}
&\sum_{t=1}^{T} f_u(-[x_{t-1}]_-)\ell_{t-1}^2(r^\mathsf{T} z_{t-1,T}) \\
&= \sum_{t=1}^{T} f_u(0)\ell_{t-1}^2(r^\mathsf{T} z_{t-1,T})\mathbf{1}_{t-1}(r) + \sum_{t=1}^{T} f_u(-[x_{t-1}]_-)\ell_{t-1}^2(r^\mathsf{T} z_{t-1,T})\mathbf{1}_{t-1}^c(r) \\
&= \sum_{t=1}^{T} f_u(0)(r^\mathsf{T} z_{t-1,T})^2 - \sum_{t=1}^{T} f_u(0)(r^\mathsf{T} z_{t-1,T})^2\mathbf{1}_{t-1}^c(r) \\
&\qquad\qquad + \sum_{t=1}^{T} f_u(-[x_{t-1}]_-)\ell_{t-1}^2(r^\mathsf{T} z_{t-1,T})\mathbf{1}_{t-1}^c(r) \\
&=: (\mathrm{XI})_T + (\mathrm{XII})_T + (\mathrm{XIII})_T
\end{aligned} \tag{C.44}$$

where in obtaining the second equality, we have used the fact that

$$\begin{aligned}
\ell_{t-1}^2(r^\mathsf{T} z_{t-1,T})\mathbf{1}_{t-1}(r) &= \{[x_{t-1} + r^\mathsf{T} z_{t-1,T}]_+ - [x_{t-1}]_+\}^2\mathbf{1}\{x_{t-1} \geq 0,\ x_{t-1} + r^\mathsf{T} z_{t-1,T} \geq 0\} \\
&= (r^\mathsf{T} z_{t-1,T})^2\mathbf{1}_{t-1}(r)
\end{aligned}$$

and similarly $f_u(-[x_{t-1}]_-)\mathbf{1}_{t-1}(r) = f_u(0)\mathbf{1}_{t-1}(r)$.

Regarding the final r.h.s. term in (C.44), in view of (C.39) and the boundedness of f_u, there exists a $C < \infty$ such that

$$\begin{aligned}
|(\mathrm{XIII})_T| &\leq C\|r\|^2 \sum_{t=1}^{T} \|z_{t-1,T}\|^2\{2 \cdot \mathbf{1}_{1,t-1}^c + \mathbf{1}_2^c(r)\} \\
&\leq_{(1)} C\|r\|^2 \left[2\left(\sum_{t=1}^{T}\|z_{t-1,T}\|^4\right)^{1/2}\left(\sum_{t=1}^{T}\mathbf{1}_{1,t-1}^c\right)^{1/2} + \mathbf{1}_2^c(r)\sum_{t=1}^{T}\|z_{t-1,T}\|^2\right] \\
&=_{(2)} o_p(1)
\end{aligned} \tag{C.45}$$

uniformly on compact subsets of $\mathbb{R}^{k+1}$, where $\leq_{(1)}$ holds by the CS inequality, and $=_{(2)}$ by Lemmas A.3(i) and A.3(v), (C.43), and the noted asymptotic negligibility (on compacta) of $\mathbf{1}_2^c(r)$. An identical argument yields that $(\mathrm{XII})_T = o_p(1)$, uniformly on compact subsets of $\mathbb{R}^{k+1}$. This leaves the first r.h.s. term in (C.44), for which

$$(\mathrm{XI})_T = f_u(0)r^\mathsf{T}\sum_{t=1}^{T} z_{t-1,T}z_{t-1,T}^\mathsf{T} r \xrightarrow{d} f_u(0)r^\mathsf{T} Q_{ZZ} r$$

uniformly on compact subsets of $\mathbb{R}^{k+1}$, by Lemma A.3(i), as required. □

C.2.4 Proof of Proposition C.3

Define

$$g_t(w_{t-1}) \coloneqq |y_t - [x_{t-1} + w_{t-1}]_+| - |y_t - [x_{t-1}]_+|, \tag{C.46}$$

for which

$$|g_t(w_{t-1}) - g_t(\tilde{w}_{t-1})| \leq |w_{t-1} - \tilde{w}_{t-1}|. \tag{C.47}$$

For suitable choices of w_{t-1}, both $U_T(\varpi) - U_T(0)$ and $\mathbb{U}_T(r) - \mathbb{U}_T(0)$ can be expressed in terms of

$$\sum_{t=1}^{T}[g_t(w_{t-1}) - \mathbb{E}_{t-1}g_t(w_{t-1})].$$

Since $Z_{T-1,T} = \sum_{s=1}^{T}\|z_{s-1,T}\|^2 = O_p(1)$ by Lemma A.3(i), for $\epsilon > 0$ as given in the statement of part (i) of the lemma, we may choose K_ϵ sufficiently large that

$$\limsup_{T\to\infty} \mathbb{P}\{Z_{T-1,T} \geq K_\epsilon\} < \epsilon.$$

For each $T \in \mathbb{N}$, define

$$\varsigma_T \coloneqq \inf\{t \in \mathbb{N} \mid Z_{t,T} > K_\epsilon\}. \tag{C.48}$$

Then for each $t \in \mathbb{N}$,

$$\{\varsigma_T > t\} = \{Z_{t,T} \leq K_\epsilon\} \in \mathcal{F}_t,$$

and so ς_T is a stopping time with respect to $\{\mathcal{F}_t\}$. By construction

$$\sum_{s=1}^{T\wedge\varsigma_T}\|z_{s-1,T}\|^2 = Z_{T\wedge\varsigma_T-1,T} = \mathbf{1}\{\varsigma_T \geq T\}Z_{T-1,T} + \mathbf{1}\{\varsigma_T < T\}Z_{\varsigma_T-1,T} \leq K_\epsilon. \tag{C.49}$$

(i). Define

$$U_{t,T}(\varpi) \coloneqq \frac{1}{T}\sum_{s=1}^{t}\{|y_s - [x_{s-1} + \varpi^\mathsf{T}\mathcal{Z}_{s-1,T}]_+| - \mathbb{E}_{s-1}|y_s - [x_{s-1} + \varpi^\mathsf{T}\mathcal{Z}_{s-1,T}]_+|\} \tag{C.50}$$

so that $U_T(\varpi) = U_{T,T}(\varpi)$. Consider the martingale array $\{M_{t,T}(\varpi), \mathcal{F}_t\}_{t=1}^{T}$ defined by

$$M_{t,T}(\varpi) \coloneqq T^{1/2}[U_{t,T}(\varpi) - U_{t,T}(0)] = \frac{1}{T^{1/2}}\sum_{s=1}^{t}[g_s(\varpi^\mathsf{T}\mathcal{Z}_{s-1,T}) - \mathbb{E}_{s-1}g_s(\varpi^\mathsf{T}\mathcal{Z}_{s-1,T})].$$

Define the stopped process $N_{t,T}(\varpi) := M_{t\wedge\varsigma_T,T}(\varpi)$, and the associated filtration $\mathcal{G}_t := \mathcal{F}_{t\wedge\varsigma_T}$, for $t \in \{1,\ldots,T\}$. Observe that since

$$Z_{T-1,T} \leq K_\epsilon \implies \varsigma_T \geq T \implies N_{T,T}(\varpi) = M_{T,T}(\varpi),\ \forall \varpi \in \bar{\Pi},$$

the event on the l.h.s. of (C.5) may be written as

$$\begin{aligned}\left\{\sup_{\|\varpi\|\leq\delta} |M_{T,T}(\varpi)| \geq \kappa,\ Z_{T-1,T} \leq K_\epsilon\right\} &= \left\{\sup_{\|\varpi\|\leq\delta} |N_{T,T}(\varpi)| \geq \kappa,\ Z_{T-1,T} \leq K_\epsilon\right\} \\ &\subseteq \left\{\sup_{\|\varpi\|\leq\delta} |N_{T,T}(\varpi)| \geq \kappa\right\}.\end{aligned}$$

It remains to control the probability of the r.h.s.

By the stopping time lemma (in particular, the corollary to Theorem 9.3.4 in Chung, 2001), $\{N_{t,T}(\varpi), \mathcal{G}_t\}_{t=1}^T$ is a martingale array for every $\varpi \in \bar{\Pi}$, with increments

$$\begin{aligned}\Delta N_{t,T}(\varpi) &:= N_{t,T}(\varpi) - N_{t-1,T}(\varpi) = M_{t\wedge\varsigma_T,T}(\varpi) - M_{(t-1)\wedge\varsigma_T,T}(\varpi) \\ &= T^{-1/2}[g_t(\varpi^\mathsf{T}\mathcal{Z}_{t-1,T}) - \mathbb{E}_{t-1}g_t(\varpi^\mathsf{T}\mathcal{Z}_{t-1,T})]\mathbf{1}\{\varsigma_T > t-1\}. \qquad \text{(C.51)}\end{aligned}$$

In view of (C.47) and the $\mathcal{F}_{t-1}$-measurability of $\mathcal{Z}_{t-1,T}$, we have that the martingale sum of squares of $N_{t,T}(\varpi) - N_{t,T}(\tilde{\varpi})$, at $t = T$, satisfies

$$\begin{aligned}&\sum_{t=1}^T [\Delta N_{t,T}(\varpi) - \Delta N_{t,T}(\tilde{\varpi})]^2 \\ &\qquad = T^{-1}\sum_{t=1}^T \{[g_t(\varpi^\mathsf{T}\mathcal{Z}_{t-1,T}) - g_t(\tilde{\varpi}^\mathsf{T}\mathcal{Z}_{t-1,T})] \\ &\qquad\qquad - \mathbb{E}_{t-1}[g_t(\varpi^\mathsf{T}\mathcal{Z}_{t-1,T}) - g_t(\tilde{\varpi}^\mathsf{T}\mathcal{Z}_{t-1,T})]\}^2\mathbf{1}\{\varsigma_T > t-1\} \\ &\qquad \leq 4\|\varpi - \tilde{\varpi}\|^2 T^{-1}\sum_{t=1}^T \|\mathcal{Z}_{t-1,T}\|^2\mathbf{1}\{\varsigma_T \geq t\} = 4\|\varpi - \tilde{\varpi}\|^2 Z_{T\wedge\varsigma_T-1,T} \\ &\qquad \leq 4\|\varpi - \tilde{\varpi}\|^2 K_\epsilon, \qquad \text{(C.52)}\end{aligned}$$

where the final equality holds by (C.49).

Therefore by Burkholder's inequality (Theorem 2.10 in Hall and Heyde, 1980), we have for any $\varpi, \tilde{\varpi} \in \bar{\Pi}$ and $j \in \mathbb{N}$ that

$$\begin{aligned}\mathbb{E}|N_{T,T}(\varpi) - N_{T,T}(\tilde{\varpi})|^{2j} &\leq C\mathbb{E}\left|\sum_{t=1}^T [\Delta N_{t,T}(\tilde{\varpi}) - \Delta N_{t,T}(\tilde{\varpi})]^2\right|^j \\ &\leq 4^j C K_\epsilon^j \|\varpi - \tilde{\varpi}\|^{2j}\end{aligned}$$

whence

$$\|N_{T,T}(\varpi) - N_{T,T}(\tilde{\varpi})\|_{2j} \leq (4C^{1/j}K_\epsilon)^{1/2}\|\varpi - \tilde{\varpi}\|.$$

Now fix $j \in \mathbb{N}$ such that $2j > k+1$. Observe that $N_{T,T}(0) = 0$ by construction. By Corollary 2.2.5 of Van der Vaart and Wellner (1996), we therefore have

$$\begin{aligned}\mathbb{E}\sup_{\|\varpi\|\leq\delta}|N_{T,T}(\varpi)| \leq \left\|\sup_{\|\varpi\|\leq\delta}|N_{T,T}(\varpi)|\right\|_{2j} &\leq K'_\epsilon \int_0^{2\delta}(2\delta/u)^{(k+1)/2j}\,\mathrm{d}u \\ &= \delta\cdot 2K'_\epsilon\int_0^1 u^{-(k+1)/2j}\,\mathrm{d}u =: C_\epsilon\delta\end{aligned}$$

where K'_ϵ and C_ϵ depend only on K_ϵ (with k and j fixed), noting that the integral is convergent since $(k+1)/2j < 1$ by choice of j. Hence by Chebyshev's inequality

$$\mathbb{P}\left\{\sup_{\|\varpi\|\leq\delta}|N_{T,T}(\varpi)| \geq \kappa\right\} \leq \frac{\mathbb{E}\sup_{\|\varpi\|\leq\delta}|N_{T,T}(\varpi)|}{\kappa} \leq \frac{C_\epsilon\delta}{\kappa}.$$

(ii). The argument here is similar to that used to prove part (i). Analogously to (C.50), define

$$\mathbb{U}_{t,T}(r) := \sum_{s=1}^{t}\{|y_s - [x_{s-1} + r^\mathsf{T} z_{s-1,T}]_+| - \mathbb{E}_{s-1}|y_s - [x_{s-1} + r^\mathsf{T} z_{s-1,T}]_+|\} \tag{C.53}$$

so that $\mathbb{U}_T(r) = \mathbb{U}_{T,T}(r)$. Defined the stopped process $\mathbb{N}_{t,T}(r) := \mathbb{U}_{t\wedge\varsigma_T,T}(r)$, so that $\{\mathbb{N}_{t,T}(r), \mathcal{G}_t\}_{t=1}^T$ is a martingale array, for each $r \in \mathbb{R}^{k+1}$. Define

$$\mathbb{L}_{t,T}(r) := \mathbb{U}_{t,T}(r) - \mathbb{N}_{t,T}(r) = \mathbb{U}_{t,T}(r) - \mathbb{U}_{t\wedge\varsigma_T,T}(r)$$

so that

$$\begin{aligned}&\mathbb{P}\left\{\sup_{\|r-\tilde{r}\|\leq\delta}|\mathbb{U}_T(r) - \mathbb{U}_T(\tilde{r})| > 2\epsilon\right\} \\ &\quad\leq \mathbb{P}\left\{\sup_{\|r-\tilde{r}\|\leq\delta}|\mathbb{N}_{T,T}(r) - \mathbb{N}_{T,T}(\tilde{r})| > \epsilon\right\} + \mathbb{P}\left\{\sup_{\|r-\tilde{r}\|\leq\delta}|\mathbb{L}_{T,T}(r) - \mathbb{L}_{T,T}(\tilde{r})| > \epsilon\right\}. \end{aligned} \tag{C.54}$$

Thus it suffices to show that, for a suitable choice of $\delta > 0$, both r.h.s. probabilities can eventually be bounded by ϵ. Regarding the second of these, the definition of $\mathbb{L}_{t,T}$ entails that if $\varsigma_T \geq T$, then $\mathbb{L}_{T,T}(r) = 0$ for all $r \in \mathbb{R}^{k+1}$. Hence, recalling the choice of K_ϵ made above,

$$\mathbb{P}\left\{\sup_{\|r-\tilde{r}\|\leq\delta}|\mathbb{L}_{T,T}(r) - \mathbb{L}_{T,T}(\tilde{r})| > \epsilon\right\} \leq 1 - \mathbb{P}\{\varsigma_T \geq T\} = 1 - \mathbb{P}\{Z_{T-1,T} \leq K_\epsilon\} < \epsilon,$$

where the final bound holds for all T sufficiently large.

It remains to bound the first r.h.s. term in (C.54). Similarly to (C.51), we have that the increments of $\mathbb{N}_{t,T}(r)$ are equal to

$$\Delta\mathbb{N}_{t,T}(r) = \mathbb{U}_{t\wedge\varsigma_T,T}(r) - \mathbb{U}_{(t-1)\wedge\varsigma_T,T}(r) = [g_t(r^\mathsf{T} z_{t-1,T}) - \mathbb{E}_{t-1}g_t(r^\mathsf{T} z_{t-1,T})]\mathbf{1}\{\varsigma_T > t-1\}$$

and so, similarly to (C.52),

$$\sum_{t=1}^{T}[\Delta\mathbb{N}_{t,T}(r) - \Delta\mathbb{N}_{t,T}(r)]^2 \leq 4\|r-\tilde{r}\|^2\sum_{t=1}^{T}\|z_{t-1,T}\|^2\mathbf{1}\{\varsigma_T \geq t\} \leq 4K_\epsilon\|r-\tilde{r}\|^2.$$

Hence we have again by Burkholder's inequality that

$$\mathbb{E}|\mathbb{N}_{T,T}(r) - \mathbb{N}_{T,T}(\tilde{r})|^{2j} \leq 4^j C K_\epsilon^j \|r-\tilde{r}\|^{2j}.$$

Taking j such that $2j > k+1$, it follows by Kolmogorov's continuity criterion (Corollary 16.9 in Kallenberg, 2001) that there exists a $\delta > 0$ such that

$$\limsup_{T\to\infty}\mathbb{P}\left\{\sup_{\|r-\tilde{r}\|\leq\delta}|\mathbb{N}_{T,T}(r) - \mathbb{N}_{T,T}(\tilde{r})| > \epsilon\right\} < \epsilon.$$

(iii). Fix an $r \in \mathbb{R}^{k+1}$. Recalling (C.46) above, so that

$$\begin{aligned} g_t(r^\mathsf{T} z_{t-1,T}) &= |y_t - [x_{t-1} + r^\mathsf{T} z_{t-1,T}]_+| - |y_t - [x_{t-1}]_+| \\ &= |(y_t - [x_{t-1}]_+) - ([x_{t-1} + r^\mathsf{T} z_{t-1,T}]_+ - [x_{t-1}]_+)| - |y_t - [x_{t-1}]_+| \end{aligned}$$

and we may write

$$\mathbb{U}_T(r) - \mathbb{U}_T(0) = \sum_{t=1}^{T}[g_t(r^\mathsf{T} z_{t-1,T}) - \mathbb{E}_{t-1}g_t(r^\mathsf{T} z_{t-1,T})]. \tag{C.55}$$

Applying the identity (C.15) to $g_t(r^\mathsf{T} z_{t-1,T})$ with

$$\begin{aligned} A_t &\coloneqq y_t - [x_{t-1}]_+ = [x_{t-1} + u_t]_+ - [x_{t-1}]_+ \\ B_{t-1}(r) &\coloneqq [x_{t-1} + r^\mathsf{T} z_{t-1,T}]_+ - [x_{t-1}]_+, \end{aligned}$$

yields

$$\begin{aligned} g_t(r^\mathsf{T} z_{t-1,T}) &= -\operatorname{sgn}(A_t)B_{t-1}(r) + \mathbf{1}\{A_t = 0\}|B_{t-1}(r)| \\ &\qquad + 2[B_{t-1}(r) - A_t][\mathbf{1}\{B_{t-1}(r) > A_t > 0\} - \mathbf{1}\{B_{t-1}(r) < A_t < 0\}] \\ &=: \upsilon_{1,t} + \upsilon_{2,t} + 2\upsilon_{3,t}. \end{aligned} \tag{C.56}$$

We will show that only v_{1t} contributes non-negligibly to (C.55), as $T \to \infty$. Noting that

$$|B_{t-1}(r)| \leq |r^\mathsf{T} z_{t-1,T}| \tag{C.57}$$

and that $A_t = 0$ with nonzero probability only if $x_{t-1} \leq 0$, we obtain

$$\mathbb{E}_{t-1} v_{2,t}^2 = |B_{t-1}(r)|^2 \mathbb{E}_{t-1} \mathbf{1}\{A_t = 0\} \leq |r^\mathsf{T} z_{t-1,T}|^2 \mathbf{1}\{x_{t-1} \leq 0\}.$$

Further, since $B_{t-1}(r) > A_t > 0$ and $B_{t-1}(r) < A_t < 0$ both imply that

$$|B_{t-1}(r) - A_t| \leq |B_{t-1}(r)| \text{ and } |u_t| \leq |r^\mathsf{T} z_{t-1,T}|,$$

we have

$$\begin{aligned}
\mathbb{E}_{t-1} v_{3,t}^2 &\leq |B_{t-1}(r)|^2 \mathbf{1}\{|u_t| \leq |r^\mathsf{T} z_{t-1,T}|\} \\
&= |B_{t-1}(r)|^2 [F_u(|r^\mathsf{T} z_{t-1,T}|) - F_u(-|r^\mathsf{T} z_{t-1,T}|)] \leq C |r^\mathsf{T} z_{t-1,T}|^3
\end{aligned}$$

for some $C < \infty$ by the mean value theorem, since f_u is bounded. Deduce that

$$\begin{aligned}
&\sum_{t=1}^{T} \mathbb{E}_{t-1}[v_{2,t} + 2v_{3,t} - \mathbb{E}_{t-1}(v_{2,t} + 2v_{3,t})]^2 \leq \sum_{t=1}^{T} \mathbb{E}_{t-1}(v_{2,t}^2 + 4v_{3,t}^2) \\
&\leq \|r\|^2 \sum_{t=1}^{T} \|z_{t-1,T}\|^2 \mathbf{1}\{x_{t-1} \leq 0\} + 8C\|r\|^3 \sum_{t=1}^{T} \|z_{t-1,T}\|^3 \\
&\leq \|r\|^2 \left(\sum_{t=1}^{T} \|z_{t-1,T}\|^4\right)^{1/2} \left(\sum_{t=1}^{T} \mathbf{1}\{x_{t-1} \leq 0\}\right)^{1/2} + 4C\|r\|^3 \max_{1 \leq s \leq T} \|z_{s-1,T}\| \sum_{t=1}^{T} \|z_{t-1,T}\|^2 \\
&= o_p(1)
\end{aligned}$$

by Lemmas A.2(ii), A.3(i), A.3(ii), and A.3(v). It follows by (C.55), (C.56) and Corollary 3.1 of Hall and Heyde (1980) that

$$\begin{aligned}
\mathbb{U}_T(r) - \mathbb{U}_T(0) &= -\sum_{t=1}^{T} B_{t-1}(r)[\operatorname{sgn}(A_t) - \mathbb{E}_{t-1} \operatorname{sgn}(A_t)] + o_p(1) \\
&= -\sum_{t=1}^{T} B_{t-1}(r) e_t + o_p(1)
\end{aligned}$$

for e_t the (bounded) m.d.s. defined in (C.3) above.

It remains to determine the weak limit of the r.h.s. To that end we recall the argument given in the proof of Proposition C.2(iii), that by defining

$$\mathbf{1}_{t-1}(r) \coloneqq \mathbf{1}\{x_{t-1} \geq 0,\ x_{t-1} + r^\mathsf{T} z_{t-1,T} \geq 0\} \qquad \mathbf{1}_{t-1}^c(r) \coloneqq 1 - \mathbf{1}_{t-1}(r)$$

as in (C.41) above, we obtain the bound

$$\mathbf{1}^c_{t-1}(r) \leq 2 \cdot \mathbf{1}\{x_{t-1} \leq 1\} + \mathbf{1}\left\{1 \leq \|r\| \sup_{1 \leq s \leq T} \|z_{s-1,T}\|\right\}$$
$$=: 2 \cdot \mathbf{1}^c_{1,t-1} + \mathbf{1}^c_2(r), \tag{C.58}$$

per (C.42) above. Moreover,

$$B_{t-1}(r)\mathbf{1}_{t-1}(r) = \{[x_{t-1} + r^{\mathsf{T}} z_{t-1,T}]_+ - [x_{t-1}]_+\}\mathbf{1}_{t-1}(r) = (r^{\mathsf{T}} z_{t-1,T})\mathbf{1}_{t-1}(r).$$

Therefore

$$\begin{aligned}\mathbb{U}_T(r) - \mathbb{U}_T(0) &= -\sum_{t=1}^{T} \mathbf{1}_{t-1}(r) B_{t-1}(r) e_t - \sum_{t=1}^{T} \mathbf{1}^c_{t-1}(r) B_{t-1}(r) e_t + o_p(1) \\ &= -\sum_{t=1}^{T} (r^{\mathsf{T}} z_{t-1,T}) e_t + \sum_{t=1}^{T} \mathbf{1}^c_{t-1}(r)[r^{\mathsf{T}} z_{t-1,T} - B_{t-1}(r)] e_t + o_p(1).\end{aligned}$$

To see that the second term is asymptotically negligible, observe that this is a martingale with conditional variance

$$\begin{aligned}\sum_{t=1}^{T} \mathbf{1}^c_{t-1}(r)[r^{\mathsf{T}} z_{t-1,T} - B_{t-1}(r)]^2 \mathbb{E}_{t-1} e_t^2 &\leq_{(1)} C\|r\|^2 \sum_{t=1}^{T} \mathbf{1}^c_{t-1}(r) \|z_{t-1,T}\|^2 \\ &\leq_{(2)} C\|r\|^2 \sum_{t=1}^{T} \|z_{t-1,T}\|^2 \{2 \cdot \mathbf{1}^c_{1,t-1} + \mathbf{1}^c_2(r)\} \\ &=_{(3)} o_p(1)\end{aligned}$$

where $\leq_{(1)}$ follows by (C.57) and the boundedness of e_t, $\leq_{(2)}$ by (C.58) and $=_{(3)}$ by the arguments that yielded (C.45) in the proof of Proposition C.2(iii). Hence by Corollary 3.1 of Hall and Heyde (1980),

$$\mathbb{U}_T(r) - \mathbb{U}_T(0) = -r^{\mathsf{T}} \sum_{t=1}^{T} z_{t-1,T} e_t + o_p(1).$$

The martingale on the r.h.s. of the preceding is

$$\sum_{t=1}^{T} z_{t-1,T} e_t = T^{-1/2} \sum_{t=1}^{T} \begin{bmatrix} 1 \\ T^{-1/2} y_{t-1} \\ \Delta \boldsymbol{y}_{t-1} \end{bmatrix} e_t. \tag{C.59}$$

To determine its weak limit, consider the vector martingale process

$$\mathbb{M}_T(\tau T) := T^{-1/2} \sum_{t=1}^{\lfloor \tau T \rfloor} \begin{bmatrix} u_t \\ e_t \\ \Delta \boldsymbol{y}_{t-1} e_t \end{bmatrix}$$

for $\tau \in [0,1]$. Each of the elements of $\mathbb{M}_T$ satisfy a conditional Lyapunov condition, for every $\tau \in [0,1]$: the first two elements because $\mathbb{E}_{t-1}|u_t|^{2+\delta_u}$ and e_t are respectively bounded, and the final $k-1$ elements because

$$T^{-1-\delta_u/2} \sum_{t=1}^{T} \|\Delta \boldsymbol{y}_{t-1}\|^{2+\delta_u} \mathbb{E}_{t-1}|e_t|^{2+\delta_u} \leq C T^{-\delta_u/2} O_p(1) = o_p(1)$$

by Lemma A.1(i). Therefore, to apply a (functional) CLT to $\mathbb{M}_T$, it suffices to determine the probability limit of the conditional variance process

$$\langle \mathbb{M}_T(\tau T) \rangle = \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \mathbb{E}_{t-1} \begin{bmatrix} u_t^2 & e_t u_t & \Delta \boldsymbol{y}_{t-1}^{\mathsf{T}} e_t u_t \\ e_t u_t & e_t^2 & \Delta \boldsymbol{y}_{t-1}^{\mathsf{T}} e_t^2 \\ \Delta \boldsymbol{y}_{t-1} e_t u_t & \Delta \boldsymbol{y}_{t-1} e_t^2 & \Delta \boldsymbol{y}_{t-1} \Delta \boldsymbol{y}_{t-1}^{\mathsf{T}} e_t^2 \end{bmatrix}, \tag{C.60}$$

for each fixed $\tau \in [0,1]$.

To compute the limit of the r.h.s. of the preceding, note that

$$\mathbf{1}\{x_{t-1} > 0\} \mathbb{E}_{t-1} \operatorname{sgn}([x_{t-1} + u_t]_+ - [x_{t-1}]_+) = \mathbf{1}\{x_{t-1} > 0\} \mathbb{E}_{t-1} \operatorname{sgn}(u_t) = 0$$

and

$$\begin{aligned} \mathbf{1}\{x_{t-1} \leq 0\} \mathbb{E}_{t-1} \operatorname{sgn}([x_{t-1} + u_t]_+ - [x_{t-1}]_+) &= \mathbf{1}\{x_{t-1} \leq 0\} \mathbb{P}_{t-1}\{u_t > -x_{t-1}\} \\ &= \mathbf{1}\{x_{t-1} \leq 0\}[1 - F(-x_{t-1})], \end{aligned}$$

and hence

$$\begin{aligned} \mathbb{E}_{t-1} e_t^2 &= \mathbb{E}_{t-1} \operatorname{sgn}([x_{t-1} + u_t]_+ - [x_{t-1}]_+)^2 - \{\mathbb{E}_{t-1} \operatorname{sgn}([x_{t-1} + u_t]_+ - [x_{t-1}]_+)\}^2 \\ &= 1 - \mathbf{1}\{x_{t-1} \leq 0\}[F(-x_{t-1}) + (1 - F(-x_{t-1}))^2] \\ &=: 1 + H_1(x_{t-1}) \end{aligned} \tag{C.61}$$

where $H_1(x)$ is bounded, and (trivially) $H_1(x) \to 0$ as $x \to +\infty$. Similarly, since

$$\mathbb{E}_{t-1} e_t u_t = \mathbb{E}_{t-1} \operatorname{sgn}([x_{t-1} + u_t]_+ - [x_{t-1}]_+) u_t$$

we obtain

$$\mathbf{1}\{x_{t-1} > 0\}\mathbb{E}_{t-1} e_t u_t = \mathbf{1}\{x_{t-1} > 0\}\mathbb{E}_{t-1} \operatorname{sgn}(u_t) u_t = \mathbf{1}\{x_{t-1} > 0\}\mathbb{E}|u_1|$$

and

$$\begin{aligned}\mathbf{1}\{x_{t-1} \leq 0\}\mathbb{E}_{t-1} e_t u_t &= \mathbf{1}\{x_{t-1} \leq 0\}\mathbb{E}_{t-1} u_t \mathbf{1}\{u_t > -x_{t-1}\} \\ &= \mathbf{1}\{x_{t-1} \leq 0\} \int_{-x_{t-1}}^{\infty} u f_u(u) \, \mathrm{d}u\end{aligned}$$

whence

$$\begin{aligned}\mathbb{E}_{t-1} e_t u_t &= \mathbf{1}\{x_{t-1} > 0\}\mathbb{E}|u_1| + \mathbf{1}\{x_{t-1} \leq 0\} \int_{-x_{t-1}}^{\infty} u f_u(u) \, \mathrm{d}u \\ &= \mathbb{E}|u_1| + \mathbf{1}\{x_{t-1} \leq 0\} \left[\int_{-x_{t-1}}^{\infty} u f_u(u) \, \mathrm{d}u - \mathbb{E}|u_1| \right] \\ &=: \mathbb{E}|u_1| + H_2(x_{t-1}) \qquad \text{(C.62)}\end{aligned}$$

where H_2 is bounded, and (trivially) $H_2(x) \to 0$ as $x \to +\infty$. Deduce that for each $\tau \in [0, 1]$,

$$\frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \mathbb{E}_{t-1} \begin{bmatrix} u_t^2 \\ e_t u_t \\ e_t^2 \end{bmatrix} = \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \begin{bmatrix} \sigma^2 \\ \mathbb{E}|u_1| + H_2(x_{t-1}) \\ 1 + H_1(x_{t-1}) \end{bmatrix} = \tau \begin{bmatrix} \sigma^2 \\ \mathbb{E}|u_1| \\ 1 \end{bmatrix} + o_p(1) \qquad \text{(C.63)}$$

by Lemma A.2(iii). This yields the limit of the upper left 2×2 block of the r.h.s. of (C.60).

Regarding the remaining elements of the r.h.s. of (C.60), we next observe that

$$\begin{aligned}\frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \mathbb{E}_{t-1} \Delta \boldsymbol{y}_{t-1} \begin{bmatrix} e_t u_t \\ e_t^2 \end{bmatrix}^{\mathsf{T}} &= \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \Delta \boldsymbol{y}_{t-1} \mathbb{E}_{t-1} \begin{bmatrix} e_t u_t \\ e_t^2 \end{bmatrix}^{\mathsf{T}} \\ &= \left(\frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \Delta \boldsymbol{y}_{t-1} \right) \begin{bmatrix} \mathbb{E}|u_1| \\ 1 \end{bmatrix}^{\mathsf{T}} + \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \Delta \boldsymbol{y}_{t-1} \begin{bmatrix} H_2(x_{t-1}) \\ H_1(x_{t-1}) \end{bmatrix}^{\mathsf{T}},\end{aligned}$$

by (C.61) and (C.62). Now

$$\frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \Delta \boldsymbol{y}_{t-1} = \frac{1}{T} (\boldsymbol{y}_{\lfloor \tau T \rfloor} - \boldsymbol{y}_0) = O_p(T^{-1/2})$$

by Theorem 3.2 in Bykhovskaya and Duffy (2024), while

$$\left| \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \Delta \boldsymbol{y}_{t-1} \begin{bmatrix} H_2(x_{t-1}) \\ H_1(x_{t-1}) \end{bmatrix}^{\mathsf{T}} \right|$$

$$\leq \left(\frac{1}{T} \sum_{t=1}^{T} \| \Delta \boldsymbol{y}_{t-1} \|^2 \right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^{T} \left\| \begin{bmatrix} H_2(x_{t-1}) \\ H_1(x_{t-1}) \end{bmatrix} \right\|^2 \right)^{1/2} = o_p(1)$$

by Lemmas A.1(ii) and A.2(iii). Hence for each $\tau \in [0, 1]$

$$\frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \mathbb{E}_{t-1} \Delta \boldsymbol{y}_{t-1} \begin{bmatrix} e_t u_t \\ e_t^2 \end{bmatrix}^{\mathsf{T}} = o_p(1). \tag{C.64}$$

Finally, since

$$\frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \mathbb{E}_{t-1} \Delta \boldsymbol{y}_{t-1} \Delta \boldsymbol{y}_{t-1}^{\mathsf{T}} e_t^2 = \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \Delta \boldsymbol{y}_{t-1} \Delta \boldsymbol{y}_{t-1}^{\mathsf{T}} + \frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \Delta \boldsymbol{y}_{t-1} \Delta \boldsymbol{y}_{t-1}^{\mathsf{T}} H_1(x_{t-1})$$

where by Hölder's inequality, the norm of the final term is bounded by

$$\left(\frac{1}{T} \sum_{t=1}^{T} \| \Delta \boldsymbol{y}_{t-1} \|^{2+\delta_u} \right)^{2/(2+\delta_u)} \left(\frac{1}{T} \sum_{t=1}^{T} |H_1(x_{t-1})|^{(2+\delta_u)/\delta_u} \right)^{\delta_u/(2+\delta_u)} = o_p(1)$$

by Lemmas A.1(i) and A.2(iii). Hence, for each $\tau \in [0, 1]$,

$$\frac{1}{T} \sum_{t=1}^{\lfloor \tau T \rfloor} \mathbb{E}_{t-1} \Delta \boldsymbol{y}_{t-1} \Delta \boldsymbol{y}_{t-1}^{\mathsf{T}} e_t^2 \xrightarrow{p} \tau \Omega \tag{C.65}$$

by Lemma A.1(ii).

It follows from (C.63)–(C.65) that for each $\tau \in [0, 1]$,

$$\langle \mathbb{M}_T(\tau T) \rangle \xrightarrow{p} \tau \begin{bmatrix} \sigma^2 & \mathbb{E}|u_1| & 0 \\ \mathbb{E}|u_1| & 1 & 0 \\ 0 & 0 & \Omega \end{bmatrix} =: \tau \Omega_M$$

Therefore Theorem 2.3 in Durrett and Resnick (1978), and the Cramér–Wold device, imply that

$$\mathbb{M}_T(\tau T) = T^{-1/2} \sum_{t=1}^{\lfloor \tau T \rfloor} \begin{bmatrix} u_t \\ e_t \\ \Delta \boldsymbol{y}_{t-1} e_t \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \sigma_0 W(\tau) \\ \widetilde{W}(\tau) \\ \Xi(\tau) \end{bmatrix} =: \mathbb{M}(\tau) \tag{C.66}$$

on $D[0, 1]$, where $\mathbb{M}$ is a $(k + 1)$-dimensional Brownian motion with variance matrix Ω_M: and so, in particular, Ξ is independent of $(W, \widetilde{W})$.

Finally, returning to (C.59), i.e. to

$$\sum_{t=1}^{T} z_{t-1,T} e_t = \sum_{t=1}^{T} \begin{bmatrix} T^{-1/2} e_t \\ T^{-1} y_{t-1} e_t \\ T^{-1/2} \Delta \boldsymbol{y}_{t-1} e_t \end{bmatrix} = \begin{bmatrix} T^{-1/2} \sum_{t=1}^{T} e_t \\ T^{-1} \sum_{t=1}^{T} y_{t-1} e_t \\ \Xi_T(1) \end{bmatrix},$$

we note that by Theorem 3.2 of Bykhovskaya and Duffy (2024) and Theorem 2.1 of Liang et al. (2016),

$$T^{-1} \sum_{t=1}^{T} y_{t-1} e_t \xrightarrow{d} \int_0^1 Y(\tau)\, \mathrm{d}E(\tau)$$

holds jointly with (C.66), where Y is a function only of W. Deduce

$$\sum_{t=1}^{T} z_{t-1,T} e_t \xrightarrow{d} \begin{bmatrix} \widetilde{W}(1) \\ \int_0^1 Y(\tau)\, \mathrm{d}\widetilde{W}(\tau) \\ \Xi(1) \end{bmatrix},$$

where $\Xi(1) \sim \mathcal{N}[0, \Omega]$ independently of $(\widetilde{W}, Y)$. □

References

Aruoba, S. B., M. Mlikota, F. Schorfheide, and S. Villalvazo (2022): "SVARs with occasionally-binding constraints," *Journal of Econometrics*, 231, 477–499.

Bagnoli, M. and T. Bergstrom (2005): "Log-concave probability and its applications," *Economic Theory*, 26, 445.

Balke, N. S. and T. B. Fomby (1997): "Threshold cointegration," *International Economic Review*, 627–645.

Brezigar-Masten, A., I. Masten, and M. Volk (2021): "Modeling credit risk with a Tobit model of days past due," *Journal of Banking & Finance*, 122, 105984.

Bykhovskaya, A. (2023): "Time series approach to the evolution of networks: prediction and estimation," *Journal of Business & Economic Statistics*, 41, 170–183.

Bykhovskaya, A. and J. A. Duffy (2024): "The local to unity dynamic Tobit model," *Journal of Econometrics*, 241, 105764.

Cavaliere, G. (2005): "Limited time series with a unit root," *Econometric Theory*, 21, 907–945.

Chan, K.-S., ed. (2009): *Exploration of a Nonlinear World: an appreciation of Howell Tong's contributions to statistics*, World Scientific.

Chan, N. and Q. Wang (2015): "Nonlinear regressions with nonstationary time series," *Journal of Econometrics*, 185, 182–195.

Chung, K. L. (2001): *A Course in Probability Theory*, Academic Press, 3rd ed.

Cohen, J. E., R. A. Davis, and G. Samorodnitsky (2022): "COVID-19 cases and deaths in the United States follow Taylor's law for heavy-tailed distributions with infinite variance," *Proceedings of the National Academy of Sciences*, 119, e2209234119.

de Jong, R. and A. M. Herrera (2011): "Dynamic censored regression and the Open Market Desk reaction function," *Journal of Business & Economic Statistics*, 29, 228–237.

Demiralp, S. and O. Jordà (2002): "The announcement effect: evidence from Open Market Desk data," *FRBNY Economic Policy Review*, 8, 29–48.

Dong, D., T. M. Schmit, and H. Kaiser (2012): "Modelling household purchasing behaviour to analyse beneficial marketing strategies," *Applied Economics*, 44, 717–725.

Duffy, J. A. and S. Mavroeidis (2024): "Common trends and long-run identification in nonlinear structural VARs," arXiv:2404.05349.

Duffy, J. A., S. Mavroeidis, and S. Wycherley (2023): "Stationarity with Occasionally Binding Constraints," arXiv:2307.06190.

——— (2025): "Cointegration with occasionally binding constraints," *Journal of Econometrics*, in press.

Durrett, R. and S. I. Resnick (1978): "Functional limit theorems for dependent variables," *Annals of Probability*, 829–846.

Elliott, G., U. K. Müller, and M. W. Watson (2015): "Nearly optimal tests when a nuisance parameter is present under the null hypothesis," *Econometrica*, 83, 771–811.

Gikhman, I. I. and A. V. Skorokhod (1969): *Introduction to the Theory of Random Processes*, Philadelphia (USA): W. B. Saunders.

Gonçalves, L., M. A. A. Turkman, C. Geraldes, T. A. Marques, and L. Sousa (2021): "COVID-19: Nothing is Normal in this Pandemic," *Journal of epidemiology and global health*, 11, 146–149.

Hall, P. and C. C. Heyde (1980): *Martingale Limit Theory and Its Application*, New York (USA): Academic Press.

Hamilton, J. D. (1994): *Time Series Analysis*, Princeton, NJ (USA): Princeton University Press.

Hansen, B. E. (1999): "The grid bootstrap and the autoregressive model," *Review of Economics and Statistics*, 81, 594–607.

Hansen, B. E. and B. Seo (2002): "Testing for two-regime threshold cointegration in vector error-correction models," *Journal of Econometrics*, 110, 293–318.

Herce, M. A. (1996): "Asymptotic theory of LAD estimation in a unit root process with finite variance errors," *Econometric Theory*, 12, 129–153.

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

Jungers, R. (2009): *The Joint Spectral Radius: theory and applications*, Springer.

Kallenberg, O. (2001): *Foundations of Modern Probability*, New York (USA): Springer, 2nd ed.

Knight, K. (1989): "Limit theory for autoregressive-parameter estimates in an infinite-variance random walk," *The Canadian Journal of Statistics*, 261–278.

Kristensen, D. and A. Rahbek (2010): "Likelihood-based inference for cointegration with nonlinear error-correction," *Journal of Econometrics*, 158, 78–94.

Li, G. and W. K. Li (2009): "Least absolute deviation estimation for unit root processes with GARCH errors," *Econometric Theory*, 25, 1208–1227.

Liang, H., P. C. B. Phillips, H. Wang, and Q. Wang (2016): "Weak convergence to stochastic integrals for econometric applications," *Econometric Theory*, 32, 1349–1375.

Liu, L., H. R. Moon, and F. Schorfheide (2019): "Forecasting with a panel Tobit model," NBER working paper 28571.

Mavroeidis, S. (2021): "Identification at the zero lower bound," *Econometrica*, 69, 2855–2885.

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

——— (2012): "One-Dimensional Inference in Autoregressive Models With the Potential Presence of a Unit Root," *Econometrica*, 80, 173–212.

Olsen, R. J. (1978): "Note on the uniqueness of the maximum likelihood estimator for the Tobit model," *Econometrica*, 1211–1215.

Park, J. Y. and P. C. B. Phillips (2000): "Nonstationary binary choice," *Econometrica*, 68, 1249–1280.

——— (2001): "Nonlinear regressions with integrated time series," *Econometrica*, 69, 117–161.

Parrilo, P. A. and A. Jadbabaie (2008): "Approximation of the joint spectral radius using sum of squares," *Linear Algebra and its Applications*, 428, 2385–2402.

Phillips, P. C. B., S. Jin, and L. Hu (2007): "Nonstationary discrete choice: A corrigendum and addendum," *Journal of Econometrics*, 141, 1115–1130.

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

Powell, J. L. (1984): "Least absolute deviations estimation for the censored regression model," *Journal of econometrics*, 25, 303–325.

Ramey, V. A. and S. Zubairy (2018): "Government spending multipliers in good times and in bad: evidence from US historical data," *Journal of political economy*, 126, 850–901.

Revuz, D. and M. Yor (1999): *Continuous Martingales and Brownian Motion*, Springer.

Sampford, M. R. (1953): "Some inequalities on Mill's ratio and related functions," *Annals of Mathematical Statistics*, 24, 130–132.

Terasvirta, T., D. Tjøstheim, and C. W. J. Granger (2010): *Modelling Nonlinear Economic Time Series*, O.U.P.

Tong, H. (1990): *Non-linear time series: a dynamical system approach*, O.U.P.

USAFacts (accessed July 15, 2025): "Cases and Deaths," `https://usafacts.org/visualizations/coronavirus-covid-19-spread-map`.

Van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes: with Applications to Statistics*, Springer.

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