

Hoga, Yannick; Karlsson, Martin

Working Paper

Extremal expected shortfall regressions: Inference and an application to health care spending

CINCH series, No. 2026/02

Provided in Cooperation with:

CINCH – Health Economics Research Center, Universität Duisburg-Essen (UDE)

Suggested Citation: Hoga, Yannick; Karlsson, Martin (2025) : Extremal expected shortfall regressions: Inference and an application to health care spending, CINCH series, No. 2026/02, University of Duisburg-Essen, CINCH - Health Economics Research Center, Essen, <https://doi.org/10.17185/dupublico/84966>

This Version is available at:

<https://hdl.handle.net/10419/339383>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

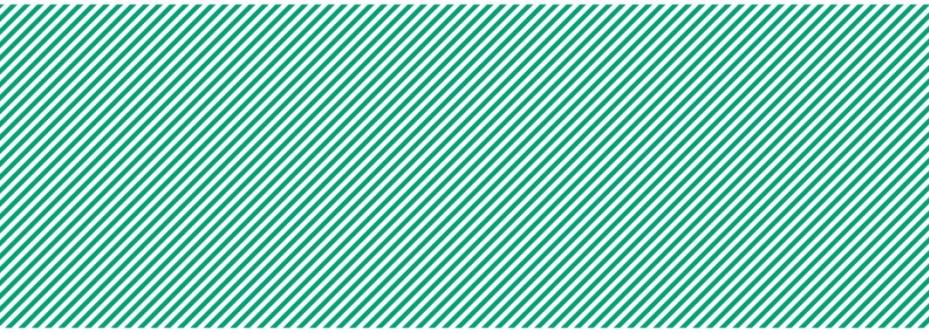
Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



CINCH+

competence in competition + health

series

#2026/02

Yannick Hoga and Martin Karlsson

**Extremal Expected Shortfall Regressions: Inference and
an Application to Health Care Spending**



Imprint

EDITOR-IN-CHIEF

Martin Karlsson, Essen

MANAGING EDITOR

Martin Karlsson, Essen

EDITORIAL BOARD

Boris Augurzky, Essen

Daniel Avdic, Melbourne (AUS)

Jeanette Brosig-Koch, Magdeburg

Stefan Felder, Basel

Annika Herr, Hannover

Nadja Kairies-Schwarz, Düsseldorf

Hendrik Schmitz, Paderborn

Harald Tauchmann, Erlangen-Nürnberg

Jürgen Wasem, Essen

Katharina Blankart, Bern (CHE)

Christoph Kronenberg, Essen

CINCH SERIES

CINCH – Health Economics Research Center
Weststadttürme, Berliner Platz 6-8
45127 Essen

www.cinch.uni-due.de

cinchseries@cinch-essen.de

Phone +49 (0) 201 183 - 3679

Fax +49 (0) 201 183 - 3716

All rights reserved. Essen, Germany, 2025

ISSN 2199-8744 (online)

The working papers published in the Series constitute work in progress circulated to stimulate discussion and critical comments. Views expressed represent exclusively the authors' own opinions and do not necessarily reflect those of the editors.

Yannick Hoga and Martin Karlsson

Extremal Expected Shortfall Regressions: Inference and an Application to Health Care Spending

Yannick Hoga^a and Martin Karlsson^a

Extremal Expected Shortfall Regressions: Inference and an Application to Health Care Spending*

Abstract

This paper proposes feasible inference methods for extremal expected shortfall (ES) regressions. While standard ES regressions consider a fixed probability level, in extremal ES regressions the probability level becomes more extreme as a function of the sample size. We show that in extremal ES regressions the convergence rate of the parameter estimators is no longer root-n (as in standard ES regressions), but depends intricately on the probability level and on the tails of the outcome. For the conditional distribution of the outcome given the covariates, we work under a Pareto-type tail assumption, and we allow for serially dependent observables. We also prove that the “beyond the sample” ES can be estimated consistently by exploiting the Pareto-type tail shape. Monte Carlo simulations show the adequate size of our tests, which are based on self-normalization. Finally, an empirical application to health care spending as a function of temperature deviations demonstrates the practical usefulness of our inference and estimation tools.

Keywords: Expected Shortfall, Extreme Value Theory, Health Insurance, Inference, Self-Normalization

JEL classification: C12 (Hypothesis Testing); C13 (Estimation); C21 (Quantile Regressions); I13 (Health Insurance, Public and Private); Q54 (Climate, Natural Disasters and Their Management, Global Warming)

^a Faculty of Economics and Business Administration, University of Duisburg-Essen, Universitätsstraße 12, D-45117 Essen, Germany, yannick.hoga@vwl.uni-due.de and martin.karlsson@uni-due.de.

*Yannick Hoga gratefully acknowledges support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through grants 460479886, 531866675 and 568876076.

1 Motivation

The expected shortfall (ES) has recently gained prominence in time series econometrics. One reason for this is that ES, as the expected value beyond a high quantile, provides valuable information about tail risk. Moreover, unlike a simple quantile, ES is a coherent (subadditive) risk measure (Artzner et al., 1999). These properties have made ES a widely used risk measure and a cornerstone of financial regulation (Basel Committee on Banking Supervision, 2019), spawning a burgeoning literature on ES modeling for time series; see, for example, Adrian et al. (2019), Dimitriadis and Bayer (2019), Patton et al. (2019), Taylor (2019), Carriero et al. (2022), Delle Monache et al. (2024), and Lopez-Salido and Loria (2024).

Due to its convenient interpretation as a quantile-truncated expectation, the ES gained popularity not only in time series econometrics, but also in microeconometrics. E.g., Wei et al. (2024) and Chen and Yen (2025) analyze the complier treatment effect of a job training program on subsequent earnings, where their emphasis on the ES allows them to specifically focus on the lower tail of the conditional income distribution. As opposed to—inherently local—quantile treatment effects (used, e.g., by Abadie et al., 2002), the ES combines the advantages of being a (truncated) expectation while still retaining its focus on the tails. For further microeconomic applications of ES regressions, we refer to Huang et al. (2025), Yu et al. (2025), and Zhang et al. (2025).

The main contribution of this paper is to extend ES regressions to extreme probability levels τ . In doing so, we draw on extreme value theory (EVT) and—as is customary in that literature—let the probability level converge to unity as the sample size n increases, such that $\tau = \tau_n \rightarrow 1$, as $n \rightarrow \infty$. This setup is more credible in many applications, where the ES is adopted specifically as a *tail* risk measure. For example, in financial risk management, the Basel Committee on Banking Supervision (2019) mandates to report ES estimates at the reasonably extreme 97.5%-level (i.e., $\tau = 0.975$). But our framework may also be more appropriate for macroeconomic applications, where tail risks measured by the

ES have become a key variable in the “at-Risk” literature ([Adrian et al., 2019](#); [Plagborg-Møller et al., 2020](#); [Adams et al., 2021](#); [Delle Monache et al., 2024](#); [Lopez-Salido and Loria, 2024](#)). There, even more moderate probability levels may be seen as reasonably extreme, due to the inherent data scarcity in macroeconomics. Hence, our “vanishing $(1 - \tau)$ ” paradigm may be more appropriate in such cases as well.

In more detail, the main contributions of our paper are threefold. First, we provide tools for feasible inference in extremal (quantile, ES) regressions. Specifically, following [Shao \(2010\)](#) and—in a standard ES regression framework—[Hoga and Schulz \(2025\)](#), we propose to use self-normalization (SN), which sidesteps the need to estimate (or bootstrap) long-run variances in extremal (quantile, ES) regressions. Such long-run variances are often notoriously difficult to estimate in practice ([Kiefer and Vogelsang, 2005](#); [Dimitriadis and Bayer, 2019](#)). This problem is exacerbated in our context because the convergence rate of our extremal ES regression parameter estimator is unknown (instead of the usual \sqrt{n} -rate) and, for valid inference, would have to be estimated. Thanks to SN, this unknown convergence rate drops out of our test statistics, similarly as the long-run variance. In addition to these advantages, SN provides further benefits. First, it is easy to implement, requiring only recursively estimated coefficients as inputs. Second, compared to methods that directly estimate asymptotic variances, SN-based tests typically provide better size control in finite samples ([Shao, 2015](#)).

Our second main aim is to provide a method for consistently estimating the “beyond the sample” ES with $\tau > 1 - 1/n$ (such that the ES corresponds to a mean beyond the largest “within sample” quantile). The idea behind our extremal ES estimator is similar to the [Weissman \(1978\)](#) estimator for large quantiles. We first estimate an “intermediate” ES based on our ES regression. Then, in a second step, we extrapolate from this intermediate ES to the “beyond the sample” ES by exploiting the tail shape implied by our Pareto-type assumption. Monte Carlo simulations show that both our extrapolation-based ES estimates and our inference tools work well in finite samples.

Our third main contribution is to demonstrate the practical relevance of extremal ES regressions in a microeconomic setting with policy significance. We study the relationship between ambient temperature and health care expenditures in Germany, building on prior evidence that expenditures exhibit heavy-tailed distributions (Karlsson et al., 2024). This application complements existing work on health and weather by highlighting how extreme expenditures respond disproportionately to high temperatures and vary across regions. Consistent with theory, we find that spending increases with temperature, with the most extreme expenditures being especially sensitive to high temperatures. By focusing on the tails of the expenditure distribution, our methods reveal patterns that conventional approaches—often based on average effects or binary outcomes—would overlook, underscoring the potential costs of extreme heat under global warming and illustrating the broader utility of extremal ES regressions for applied research.

The only other paper on extremal ES regressions that we are aware of is Huang et al. (2025). They consider extremal non-parametric ES regressions in a potential outcomes framework with a continuous treatment. In particular, they work with independent, identically distributed (i.i.d.) data under an unconfoundedness assumption, which ensures that—given covariates—the treatment is independent of the potential outcomes. In contrast, we allow for serially dependent observables and we do not impose unconfoundedness. We mention that the latter may even be restrictive for i.i.d. data. For instance, in a cross-sectional context with i.i.d. draws from the population, Karlsson et al. (2024) relate the ES of health spending to age—a setting where unconfoundedness clearly does not hold. A further distinguishing feature relative to Huang et al. (2025) is that we do not consider non-parametric regressions, but (linear) parametric ones. Therefore, our framework also incorporates inference, which is important to identify the truly significant regressors for the extreme ES. Finally, from a technical perspective, Huang et al. (2025) do not estimate ES regressions themselves (as we do), but rather work with non-parametric quantile regressions (QRs) from which they infer the ES by exploiting the tail shape.

Since ES regressions like ours are estimated jointly with QRs, our paper relates naturally to the extremal QR literature pioneered by Chernozhukov (2005). Since then, extremal QR was extended in several directions (Chernozhukov and Fernández-Val, 2011; Zhang, 2018, 2021; Li et al., 2022; Daouia et al., 2023; Hou et al., 2024). Among those papers, the one with the most similar focus to ours is that of Zhang (2021). He develops extremal linear QR for time series that are tail dependent, such that heteroskedasticity and autocorrelation (HAC) need to be accounted for in the asymptotic variance of the parameter estimator. Specifically, while for tail independent time series the asymptotic variance of the QR parameter estimator is as in the i.i.d. setting (Chernozhukov, 2005; Chernozhukov and Fernández-Val, 2011), Zhang (2021) shows that for more general tail dependent time series cross-terms appear in the long-run variance. Estimating those long-run variances consistently is a delicate task even in “standard” QR, where we are aware of only three methods (Gregory et al., 2018; Galvao and Yoon, 2024; Hoga and Schulz, 2025), and remains unexplored in extremal QR even by Zhang (2021). Hence, our SN-based methods are the first to allow for valid inference in extremal QR for general (tail dependent) time series, which we view as an important corollary to our main results.

The remainder of the paper proceeds as follows. Section 2 introduces linear ES regressions and the corresponding SN-based inference tools. In Section 3, we propose and investigate (intermediate and extremal) ES estimators. The Monte Carlo simulations are in Section 4, while Section 5 presents the empirical application. The final Section 6 concludes. An Online Appendix contains the proofs of all theoretical results, and additional material for the empirical application.

2 Extremal Expected Shortfall Regressions

In Section 2.1, we first introduce extremal (quantile, ES) regressions. Then, we present our regression estimators in Section 2.2, while Section 2.3 collects our main assumptions.

SN-based inference methods for extremal ES regressions are laid out in Section 2.4.

2.1 The Linear ES Regression Model

Suppose we have a sample $\{(Y_t, \mathbf{X}_t')\}_{t=1, \dots, n}$ of a real-valued dependent variable Y_t and a $(m \times 1)$ -vector of covariates \mathbf{X}_t . In quantile regressions, we want to learn about the conditional quantile of Y_t given the covariates \mathbf{X}_t , i.e., $Q_\tau(Y_t | \mathbf{X}_t) := F_{Y_t|\mathbf{X}_t}^{-1}(\tau)$, where $F_{Y_t|\mathbf{X}_t}(\cdot) = \mathbb{P}\{Y_t \leq \cdot | \mathbf{X}_t\}$. In *joint* (quantile, ES) regressions we are additionally interested in modeling the ES of Y_t conditional on \mathbf{X}_t , i.e., $\text{ES}_\tau(Y_t | \mathbf{X}_t) := \mathbb{E}_t[Y_t | Y_t \geq Q_\tau(Y_t | \mathbf{X}_t)]$, where $\mathbb{E}_t[\cdot] := \mathbb{E}[\cdot | \mathbf{X}_t]$. To do so, we consider the joint linear (quantile, ES) regression

$$Y_t = \mathbf{X}_t' \boldsymbol{\alpha}_{0\tau} + \varepsilon_{t\tau}, \quad Q_\tau(\varepsilon_{t\tau} | \mathbf{X}_t) = 0, \quad (1)$$

$$Y_t = \mathbf{X}_t' \boldsymbol{\beta}_{0\tau} + \xi_{t\tau}, \quad \text{ES}_\tau(\xi_{t\tau} | \mathbf{X}_t) = 0, \quad (2)$$

where $\boldsymbol{\alpha}_{0\tau}$ and $\boldsymbol{\beta}_{0\tau}$ are the true parameters. While *standard* ES regressions consider a fixed $\tau \in (0, 1)$, in our *extremal* ES regressions we let $\tau = \tau_n \in (0, 1)$ depend on the sample size n in a way that $\tau \rightarrow 1$, as $n \rightarrow \infty$ (see Assumption 1 below). The assumption on the errors in (1) implies $Q_\tau(Y_t | \mathbf{X}_t) = \mathbf{X}_t' \boldsymbol{\alpha}_{0\tau}$, such that $\boldsymbol{\alpha}_{0\tau}$ is the coefficient of interest as it measures the impact of \mathbf{X}_t on the τ -th conditional quantile of Y_t . Similarly, the assumption in (2) on the ES regression errors $\xi_{t\tau}$ (viz. $0 = \text{ES}_\tau(\xi_{t\tau} | \mathbf{X}_t) = \mathbb{E}_t[\xi_{t\tau} \geq Q_\tau(\xi_{t\tau} | \mathbf{X}_t)] \stackrel{(2)}{=} \mathbb{E}_t[\xi_{t\tau} | Y_t \geq Q_\tau(Y_t | \mathbf{X}_t)]$) leads to $\text{ES}_\tau(Y_t | \mathbf{X}_t) = \mathbb{E}_t[Y_t | Y_t \geq Q_\tau(Y_t | \mathbf{X}_t)] = \mathbf{X}_t' \boldsymbol{\beta}_{0\tau}$, such that $\boldsymbol{\beta}_{0\tau}$ gauges the influence of the regressors on the conditional ES of Y_t . In this sense, (2) only places a restriction on the relation between Y_t and \mathbf{X}_t *conditional* on a quantile exceedance, $Y_t \geq Q_\tau(Y_t | \mathbf{X}_t)$. By using the subscript t in (1)–(2), we emphasize that $(Y_t, \mathbf{X}_t)'$ can potentially be serially dependent. In cross-sectional contexts, where $\{(Y_t, \mathbf{X}_t')\}_{t=1, \dots, n}$ are i.i.d., the t indicates the individual unit.

The QR in (1) can be viewed as an auxiliary regression, primarily enabling the estimation of the ES regression in (2). It is required because ES itself is not elicitable, meaning that no suitable loss function exists for its direct estimation (Gneiting, 2011; Dimitriadis and Bayer,

2019). However, [Fissler and Ziegel \(2016\)](#) show that the pair (quantile, ES) is elicitable, which allows for joint estimation using their FZ losses, see, e.g., [Dimitriadis and Bayer \(2019\)](#) and [Patton et al. \(2019\)](#). Yet, the FZ losses are generally non-convex ([Fissler, 2017](#), Proposition 4.2.31), making the estimation process computationally challenging. Therefore, in this paper, we take a different tack to estimation, as outlined next.

2.2 Estimating Extremal ES Regression

We adopt the two-step estimator of [Hoga and Schulz \(2025\)](#), which is computationally convenient. Specifically, we estimate the QR parameter vector $\alpha_{0\tau}$ in a first step via

$$\hat{\alpha} = \arg \min_{\alpha \in \mathbb{R}^m} \sum_{t=1}^n \rho(Y_t - \mathbf{X}'_t \alpha),$$

where $\rho(u) := u(\tau - \mathbb{1}_{\{u \leq 0\}})$ is the usual tick loss function. In the second step, we use the following estimator for the ES regression parameters $\beta_{0\tau}$:

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^m} \sum_{t=1}^n \mathbb{1}_{\{Y_t > \mathbf{X}'_t \hat{\alpha}\}} (Y_t - \mathbf{X}'_t \beta)^2. \quad (3)$$

Therefore, $\hat{\beta}$ can be obtained from an OLS regression of Y_t on \mathbf{X}_t for those t , where Y_t exceeds the predicted quantile $\mathbf{X}'_t \hat{\alpha}$. The reasoning behind this estimator is that the ES is a tail mean beyond the quantile and (truncated) mean regressions are estimated by minimizing squared errors. Clearly, the computation of each $\hat{\alpha}$ and $\hat{\beta}$ is based on convex minimization, which contrasts with estimators based on minimizing FZ losses.

As mentioned previously, applications of ES regressions often focus on large probability levels. To reflect this in the theory, we therefore consider our extremal ES regressions with $\tau = \tau_n$ converging to unity, as $n \rightarrow \infty$. In this setting, our goal is to obtain asymptotic distributions that provide more accurate approximations to the actual finite-sample distributions of the ES regression estimators than those obtained in standard ES regressions with fixed τ .

Our self-normalized approach to inference in time series (quantile, ES) regressions requires

parameter estimates that are computed sequentially over the sample, i.e., estimates of $\alpha_{0\tau}$ and $\beta_{0\tau}$ based on the first $\lfloor ns \rfloor$ observations, where $s \in (0, 1]$ and $\lfloor \cdot \rfloor$ denotes the floor function. Therefore, we introduce the estimators

$$\begin{aligned}\widehat{\alpha}(s) &:= \arg \min_{\alpha \in \mathbb{R}^m} \sum_{t=1}^{\lfloor ns \rfloor} \rho(Y_t - \mathbf{X}'_t \alpha), \\ \widehat{\beta}(s) &:= \arg \min_{\beta \in \mathbb{R}^m} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{1}_{\{Y_t > \mathbf{X}'_t \widehat{\alpha}(s)\}} (Y_t - \mathbf{X}'_t \beta)^2.\end{aligned}$$

Obviously, it holds for the full-sample estimate that $\widehat{\alpha}(1) = \widehat{\alpha}$ and $\widehat{\beta}(1) = \widehat{\beta}$. Following [Shao \(2010\)](#), SN-based inference requires *functional* central limit theory for $\widehat{\alpha}(s)$ and $\widehat{\beta}(s)$. To derive this, the next section introduces some regularity conditions.

2.3 Main Assumptions

Define $\|\cdot\|$ to be the Euclidean norm when applied to a vector or a matrix. Furthermore, let $\psi(u) = \tau - \mathbb{1}_{\{u \leq 0\}}$ denote the subgradient of $\rho(u)$, which is equal to the standard gradient of $\rho(u)$ for $u \neq 0$ (i.e., $\psi(u) \stackrel{u \neq 0}{=} \frac{\partial}{\partial u} \rho(u)$). Additionally, set $\psi_*(\varepsilon_{t\tau}, \xi_{t\tau}) = [F(\tau)]^{-1} \mathbb{1}_{\{\varepsilon_{t\tau} > 0\}} \xi_{t\tau}$. Put $F_t(y) := \mathbb{P}\{Y_t \leq y \mid \mathbf{X}_t\}$, $F(y \mid \mathbf{x}) := \mathbb{P}\{Y_t \leq y \mid \mathbf{X}_t = \mathbf{x}\}$ and $F(y) := \mathbb{P}\{Y_t \leq y\}$, where $F(\cdot \mid \mathbf{x})$ and $F(\cdot)$ do not depend on t , because we assume strict stationarity of (Y_t, \mathbf{X}'_t) below. With this, we are ready to state our main assumptions.

Assumption 1 (Probability Level). *The probability level τ satisfies that (i) $\tau \rightarrow 1$ and (ii) $(1 - \tau)n \rightarrow \infty$, as $n \rightarrow \infty$.*

Assumption 2 (Serial Dependence). *The sequence $\{(Y_t, \mathbf{X}'_t)'\}_{t \in \mathbb{N}}$ is strictly stationary and β -mixing with mixing coefficients $\beta(\cdot)$ satisfying*

$$\lim_{n \rightarrow \infty} \frac{n}{r_n} \beta(l_n) = 0$$

for some integer sequences l_n and r_n tending to infinity with $l_n = o(r_n)$ and $r_n = o(k^{1/4})$, as $n \rightarrow \infty$, where the intermediate sequence k is implicitly defined by $\tau = 1 - k/n$.

Assumption 3 (Long-Run Variances). As $n \rightarrow \infty$,

$$\frac{1}{r_n \tau (1 - \tau)} \mathbb{E} \left[\sum_{t=1}^{l_n} \begin{pmatrix} \psi_*(\varepsilon_{t\tau}, \xi_{t\tau}) \mathbf{X}_t \\ \psi(\varepsilon_{t\tau}) \mathbf{X}_t \end{pmatrix} \sum_{t=1}^{l_n} (\psi_*(\varepsilon_{t\tau}, \xi_{t\tau}) \mathbf{X}_t', \psi(\varepsilon_{t\tau}) \mathbf{X}_t') \right] \rightarrow \mathbf{0},$$

$$\frac{1}{r_n \tau (1 - \tau)} \mathbb{E} \left[\sum_{t=1}^{r_n} \begin{pmatrix} \psi_*(\varepsilon_{t\tau}, \xi_{t\tau}) \mathbf{X}_t \\ \psi(\varepsilon_{t\tau}) \mathbf{X}_t \end{pmatrix} \sum_{t=1}^{r_n} (\psi_*(\varepsilon_{t\tau}, \xi_{t\tau}) \mathbf{X}_t', \psi(\varepsilon_{t\tau}) \mathbf{X}_t') \right] \rightarrow \begin{pmatrix} \mathbf{J}_* & \mathbf{C}' \\ \mathbf{C} & \mathbf{J} \end{pmatrix},$$

where the latter limiting matrix is positive definite.

Assumption 4 (Product Moments). The $\mathbb{R}^{m \times m}$ -matrices $\mathbf{D} = \mathbb{E}[\gamma^{-1} h^{-\gamma}(\mathbf{X}_t) \mathbf{X}_t \mathbf{X}_t']$ and $\mathbf{D}_* = \mathbb{E}[\mathbf{X}_t \mathbf{X}_t']$ are positive definite.

Assumption 5 (Support of Covariates). The covariates \mathbf{X}_t have compact support $\mathcal{X} \subset \mathbb{R}^m$.

Assumption 6 (Conditional Distribution of Y_t). (i) $F_t(\cdot)$ is continuous.

(ii) There exists a positive and eventually decreasing function $A(\cdot)$ with $\lim_{n \rightarrow \infty} A(1/[1 - \tau]) = 0$, and a positive continuous function $h(\mathbf{x})$ on \mathcal{X} satisfying

$$0 < \inf_{\mathbf{x} \in \mathcal{X}} h(\mathbf{x}) \leq \sup_{\mathbf{x} \in \mathcal{X}} h(\mathbf{x}) < \infty,$$

such that, as $s \rightarrow \infty$,

$$\sup_{\mathbf{x} \in \mathcal{X}} \left| \frac{1 - F(s | \mathbf{x})}{1 - F(s)} - h(\mathbf{x}) \right| = O\left(A(1/[1 - F(s)])\right).$$

(iii) There exist some $\gamma > 0$, $\xi < 0$ and a function $A_1(\cdot)$ not changing sign eventually, such that $\lim_{n \rightarrow \infty} A_1(1/[1 - \tau]) = 0$, and for any $z > 0$,

$$\lim_{s \rightarrow \infty} \frac{1}{A_1(1/[1 - F(s)])} \left[\frac{1 - F(sz)}{1 - F(s)} - z^{-1/\gamma} \right] = z^{-1/\gamma} \frac{z^{\xi/\gamma} - 1}{\gamma \xi}.$$

Moreover, $\lim_{n \rightarrow \infty} \sqrt{k} A_1(n/k) = 0$ and $\lim_{n \rightarrow \infty} \sqrt{k} A(n/k) = 0$ for the intermediate sequence k defined in Assumption 2.

Assumption 1 implies that $(1 - \tau)$ vanishes asymptotically, but at a slower rate than $1/n$. Therefore, we often speak of working under the “vanishing $(1 - \tau)$ ” paradigm. The

condition on the decay rate of $\beta(\cdot)$ in Assumption 2 is required for our proofs that employ a “big block, small block” approach. Here, r_n indicates the length of the big blocks, and l_n that of the small blocks. Similar conditions can be found in Drees and Rootzén (2010), Hoga (2017, 2018), Davis et al. (2018) and Daouia et al. (2023). Assumption 3, which is similar in spirit to Assumption 3* in Hoga and Schulz (2025), ensures that long-run variances converge appropriately. For i.i.d. data, where the cross-terms vanish, Assumption 3 is almost trivially satisfied. Assumption 4 is similar to the usual full-rank condition on the regressors in standard linear mean regressions. The compact support condition of Assumption 5 is prevalent in the literature on extreme quantile and tail index regressions (Chernozhukov, 2005; Chernozhukov and Fernández-Val, 2011; Wang and Tsai, 2009; Zhang, 2018; Li et al., 2022; Nicolau et al., 2023; Hou et al., 2024). Finally, Assumption 6 imposes Pareto-type tails on the conditional distribution of $Y_t \mid \mathbf{X}_t$, and corresponds to Assumption 1 from Hou et al. (2024) for, in their notation, $N = 1$. Note that Assumption 6 (i) also implies continuity of $F(\cdot)$ by dominated convergence and the law of iterated expectations. In Assumption 6 (iii), γ is called the *extreme value index*, with its inverse $\alpha = 1/\gamma$ being the *tail index*. Larger values of α indicate lighter tails, as moments up to order α exist for $F(\cdot)$. The parameter of second-order regular variation is ξ and it governs the speed of convergence in Assumption 6 (iii) to an exact Pareto tail. Specifically, more negative values of ξ imply a faster convergence, and values closer to 0 a slower convergence. The extreme case $\xi = -\infty$ corresponds to a true Pareto distribution of $F(\cdot)$. Note that Assumptions 5 and 6 together with (2) imply that $\xi_{t\tau}$ possesses finite fourth moments, such that the expectations in Assumption 3 involving $\psi_*(\varepsilon_{t\tau}, \xi_{t\tau})$ exist.

2.4 Inference for Extremal ES Regression

To obtain the weak limit of $\widehat{\beta}(s)$, we utilize a functional version of the well-known Convexity lemma of Knight (1989) that is due to Kato (2009, Theorem 1) and Hoga and Schulz (2025, Lemma 1). As a preliminary step towards that goal, we first derive the functional weak limit

of $\widehat{\boldsymbol{\alpha}}(s)$. To do so, our first main theoretical result, Theorem 1, builds on [Hoga and Schulz \(2025, Lemma 1\)](#). To introduce Theorem 1, we denote by $D_m[a, b]$ the space of \mathbb{R}^m -valued componentwise càdlàg functions on the interval $[a, b]$ ($0 \leq a < b < \infty$), which we endow with the (product) Skorohod topology ([Billingsley, 1999](#)).

Theorem 1. *Under Assumptions 1–6 it holds for any $\epsilon \in (0, 1)$ that, as $n \rightarrow \infty$,*

$$s \frac{\sqrt{n\tau(1-\tau)}}{F^{\leftarrow}(\tau)} [\widehat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau}] \xrightarrow{d} \boldsymbol{\Omega}^{1/2} \mathbf{W}_m(s) \quad \text{in } D_m[\epsilon, 1],$$

where $\mathbf{W}_m(\cdot)$ is a m -variate standard Wiener process, and $\boldsymbol{\Omega} = \mathbf{D}^{-1} \mathbf{J} \mathbf{D}^{-1}$.

Theorem 1 relates as follows to the extant literature on extremal QR. Unlike the results of [Chernozhukov \(2005\)](#), [Chernozhukov and Fernández-Val \(2011\)](#), [Zhang \(2021\)](#), [Daouia et al. \(2023\)](#) and [Hou et al. \(2024\)](#), it provides *functional* central limit theory, which allows for off-the-shelf inference via SN; see Corollary 1 below and the comments surrounding it ([Shao, 2010, 2015](#)). While [Chernozhukov \(2005\)](#) and [Chernozhukov and Fernández-Val \(2011\)](#) impose the strong assumption of tail *independence*, thus assuming the hard-to-estimate cross-terms in \mathbf{J} away, [Zhang \(2021\)](#) deals with tail *dependent* time series and shows that such cross terms arise in this case. However, [Zhang \(2021\)](#) does not provide consistent estimates of the long-run variance matrix. In fact, only recently, [Galvao and Yoon \(2024\)](#) provided HAC-based inference in *non-extremal* QR. Therefore, Theorem 1 is the first result on extremal QR that allows for valid inference for (empirically more realistic) tail dependent time series.

The proof of Theorem 1 reveals that the asymptotic limit of $\widehat{\boldsymbol{\alpha}}(s)$ is (up to pre-multiplication with the matrix \mathbf{D}^{-1}) driven by the asymptotic limit of $[n\tau(1-\tau)]^{-1/2} \sum_{t=1}^{\lfloor ns \rfloor} \psi(\varepsilon_{t\tau}) \mathbf{X}_t$, i.e., the suitably scaled partial sum process of the so-called *generalized errors* $\psi(\varepsilon_{t\tau})$ multiplied by the covariates. The $\psi(\varepsilon_{t\tau})$ are called generalized errors, because—similarly as “usual” regression errors—they have a conditional mean of zero, i.e., $\mathbb{E}_t[\psi(\varepsilon_{t\tau})] = 0$. The counterpart to the QR generalized errors in ES regressions are $\psi_*(\varepsilon_{t\tau}, \xi_{t\tau}) = [F^{\leftarrow}(\tau)]^{-1} \mathbf{1}_{\{\varepsilon_{t\tau} > 0\}} \xi_{t\tau}$, for which also $\mathbb{E}_t[\psi_*(\varepsilon_{t\tau}, \xi_{t\tau})] = 0$.

Theorem 2. Under Assumptions 1–6 and $\gamma \in (0, 1/4)$ it holds for any $\epsilon \in (0, 1)$ that, as $n \rightarrow \infty$,

$$s \frac{\sqrt{n\tau(1-\tau)}}{F^{\leftarrow}(\tau)} [\widehat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}_{0\tau}] \xrightarrow{d} \boldsymbol{\Omega}_*^{1/2} \mathbf{W}_m(s) \quad \text{in } D_m[\epsilon, 1],$$

where $\mathbf{W}_m(\cdot)$ is a m -variate standard Wiener process, and $\boldsymbol{\Omega}_* = \mathbf{D}_*^{-1} \boldsymbol{\Sigma}_* \mathbf{D}_*^{-1}$ with positive definite $\boldsymbol{\Sigma}_* = \mathbf{J}_* + (1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}_*^{-1} \mathbf{C} + [(1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}_*^{-1} \mathbf{C}]' + (1-\gamma)^{-2} \mathbf{D}_* \boldsymbol{\Omega} \mathbf{D}_*$.

The proof of Theorem 2 requires significantly different proof techniques compared with the case of a fixed $\tau \in (0, 1)$ considered, e.g., by [Dimitriadis and Bayer \(2019\)](#), [Barendse \(2023\)](#) and [Hoga and Schulz \(2025\)](#). This is because in our “vanishing $(1-\tau)$ ” setting, standard (F)CLTs, as invoked by the latter authors, no longer apply (see [Hill, 2009](#), Lemma 4.1 and the surrounding discussion). Instead, we employ “big block, small block” arguments, where the small blocks are asymptotically negligible, such that the limit is solely determined by the big blocks. Since these big blocks are separated by small blocks, their dependence vanishes asymptotically, such that they may be regarded as independent over time for the purposes of the proofs.

Compared with Theorem 1, we have to assume $\gamma \in (0, 1/4)$ in Theorem 2. The reason is that the focus in the latter is on ES regressions, that are estimated by minimizing squared error-type losses. In contrast, QRs require weaker moment conditions since they only minimize tick-loss functions. A similar phenomenon appears in extremal QR and extremal expectile regression in [Daouia et al. \(2023, Theorems 2.1 and 3.1\)](#), where expectile regressions require more stringent conditions on γ .

Theorem 2 shows that the estimator $\widehat{\boldsymbol{\beta}}$ may be consistent or inconsistent depending on whether $\sqrt{n\tau(1-\tau)}/F^{\leftarrow}(\tau)$ converges to infinity or zero. It also shows that the asymptotic variance of $\widehat{\boldsymbol{\beta}}$ has the usual sandwich form, with the hard-to-estimate meat matrix $\boldsymbol{\Sigma}_*$ in the middle. By virtue of SN, however, we sidestep the problem of having to estimate $\boldsymbol{\Omega}_*$ and $F^{\leftarrow}(\tau)$ for feasible inference based on Theorem 2. Specifically, applying the continuous mapping theorem (e.g., [Davidson, 1994](#), Theorem 26.13) to Theorem 2 implies the following corollary.

Corollary 1. Let \mathbf{B} be a full rank $(\ell \times m)$ -matrix with $\ell \leq m$, and define $\boldsymbol{\beta}_{0\tau, \mathbf{B}} = \mathbf{B}\boldsymbol{\beta}_{0\tau}$ and $\widehat{\boldsymbol{\beta}}_{\mathbf{B}}(s) = \mathbf{B}\widehat{\boldsymbol{\beta}}(s)$. Under the assumptions of Theorem 2 it holds that, as $n \rightarrow \infty$,

$$n[\widehat{\boldsymbol{\beta}}_{\mathbf{B}}(1) - \boldsymbol{\beta}_{0\tau, \mathbf{B}}]' \mathbf{S}_{n, \widehat{\boldsymbol{\beta}}(\cdot)}^{-1} [\widehat{\boldsymbol{\beta}}_{\mathbf{B}}(1) - \boldsymbol{\beta}_{0\tau, \mathbf{B}}] \xrightarrow{d} \mathbf{W}_m'(1) \mathbf{B}' \mathbf{V}_m^{-1} \mathbf{B} \mathbf{W}_m(1) =: U(\mathbf{B}),$$

where

$$\begin{aligned} \mathbf{S}_{n, \widehat{\boldsymbol{\beta}}(\cdot)} &= \frac{1}{n^2} \sum_{j=\lfloor n\epsilon \rfloor + 1}^n j^2 [\widehat{\boldsymbol{\beta}}_{\mathbf{B}}(j/n) - \widehat{\boldsymbol{\beta}}_{\mathbf{B}}(1)] [\widehat{\boldsymbol{\beta}}_{\mathbf{B}}(j/n) - \widehat{\boldsymbol{\beta}}_{\mathbf{B}}(1)]', \\ \mathbf{V}_m &= \mathbf{B} \int_{\epsilon}^1 [\mathbf{W}_m(s) - s\mathbf{W}_m(1)] [\mathbf{W}_m(s) - s\mathbf{W}_m(1)]' ds \mathbf{B}'. \end{aligned}$$

Proof: Analogous to the proof of Corollary 1 in Hoga and Schulz (2025). \square

The limit $U(\mathbf{B})$ is free of nuisance parameters. Therefore, Corollary 1 allows us to test the null hypothesis $H_0^{\mathcal{E}}: \mathbf{B}\boldsymbol{\beta}_{0\tau} = \boldsymbol{\beta}_{0\tau, \mathbf{B}}^{\circ}$ for some $\boldsymbol{\beta}_{0\tau, \mathbf{B}}^{\circ} \in \mathbb{R}^{\ell}$ via the test statistic

$$\mathcal{E}_n = n[\widehat{\boldsymbol{\beta}}_{\mathbf{B}}(1) - \boldsymbol{\beta}_{0\tau, \mathbf{B}}^{\circ}]' \mathbf{S}_{n, \widehat{\boldsymbol{\beta}}(\cdot)}^{-1} [\widehat{\boldsymbol{\beta}}_{\mathbf{B}}(1) - \boldsymbol{\beta}_{0\tau, \mathbf{B}}^{\circ}].$$

We obtain an asymptotic level- ν test if we reject $H_0^{\mathcal{E}}$ when \mathcal{E}_n exceeds the $(1 - \nu)$ -quantile of $U(\mathbf{B})$. The latter is easy to compute via simulations, because the distribution of $U(\mathbf{B})$ is free of nuisance parameters. Specifically, Table B1 in Appendix B displays some selected $(1 - \nu)$ -quantiles of the limiting distribution $U(\mathbf{B})$ for different m 's. For notational brevity, we suppress the dependence of $U(\mathbf{B})$ on ϵ and m .

Note that our test statistic \mathcal{E}_n does not require plug-in estimates of the asymptotic variance $\boldsymbol{\Omega}_*$ or of the unknown $F^{\leftarrow}(\tau)$ in the convergence rate of Theorem 2. While for classical t -tests these two quantities would have to be estimated, SN serves a double purpose here by eliminating the need to estimate either quantity. This is because the asymptotic variance and the convergence rate cancel each other out in the “numerator” and “denominator” of our test statistic \mathcal{E}_n .

Comparing our results with those of Hoga and Schulz (2025), who consider standard ES regressions with fixed τ , we find the convergence rate and asymptotic variance of the ES

regression parameter estimator to be completely different. E.g., in standard ES regressions we have root- n convergence (Dimitriadis and Bayer, 2019; Barendse, 2023; Hoga and Schulz, 2025). However, in the present setting $\sqrt{n\tau(1-\tau)}/F^{\leftarrow}(\tau)$ -convergence obtains, such that the rate depends on the tails of Y_t through $F^{\leftarrow}(\tau)$. Thus, traditional t -tests valid for fixed τ are invalid in our “vanishing $(1-\tau)$ ” framework. In contrast, our Corollary 1 and Corollary 2 of Hoga and Schulz (2025) show that SN-based inference can proceed in exactly the same manner in both cases, demonstrating the great robustness of SN.

This is similar as for testing hypotheses about the mean. While a classical t -test with standard errors for i.i.d. data would be invalid under serial dependence (the latter requiring HAC corrections), self-normalized inference for the mean á la Shao (2010) works regardless of whether the data is i.i.d. or serially dependent. Our results here show that SN is also capable of providing robustness along a dimension completely different from serial dependence, viz. vanishing $(1-\tau)$.

A result similar to Corollary 1 can also be obtained for $\widehat{\boldsymbol{\alpha}}(\cdot)$. Specifically, the hypothesis $H_0^{\mathcal{Q}}: \mathbf{A}\boldsymbol{\alpha}_{0\tau} = \boldsymbol{\alpha}_{0\tau, \mathbf{A}}^{\circ}$ for some $\boldsymbol{\alpha}_{0\tau, \mathbf{A}}^{\circ} \in \mathbb{R}^{\ell}$ and full rank $\mathbf{A} \in \mathbb{R}^{\ell \times m}$ can be tested via the test statistic

$$\mathcal{Q}_n = n[\widehat{\boldsymbol{\alpha}}_{\mathbf{A}}(1) - \boldsymbol{\alpha}_{0\tau, \mathbf{A}}^{\circ}]' \mathbf{S}_{n, \widehat{\boldsymbol{\alpha}}(\cdot)}^{-1} [\widehat{\boldsymbol{\alpha}}_{\mathbf{A}}(1) - \boldsymbol{\alpha}_{0\tau, \mathbf{A}}^{\circ}],$$

where $\widehat{\boldsymbol{\alpha}}_{\mathbf{A}}(s) := \mathbf{A}\widehat{\boldsymbol{\alpha}}(s)$ and $\mathbf{S}_{n, \widehat{\boldsymbol{\alpha}}(\cdot)}$ is defined in analogy to $\mathbf{S}_{n, \widehat{\boldsymbol{\beta}}(\cdot)}$ from Corollary 1. Similarly as before, \mathcal{Q}_n converges weakly to the asymptotic limit $U(\mathbf{A})$ under the conditions of Theorem 1.

3 ES Estimation

3.1 Intermediate ES Estimation

Once a suitable ES model has been fitted to the data, it may be of interest to estimate the model-implied ES. By (2), an estimator of $\text{ES}_\tau(\mathbf{x}) := \text{ES}_\tau(Y_t \mid \mathbf{X}_t = \mathbf{x})$ is given by

$$\widehat{\text{ES}}_\tau(\mathbf{x}) = \mathbf{x}'\widehat{\boldsymbol{\beta}}. \quad (4)$$

The following theorem shows that this estimator is consistent for the true ES in the sense that $\widehat{\text{ES}}_\tau(\mathbf{x})/\text{ES}_\tau(\mathbf{x}) \xrightarrow{\mathbb{P}} 1$, as $n \rightarrow \infty$.

Theorem 3. *Under the assumptions of Theorem 2 it holds that, as $n \rightarrow \infty$,*

$$\sqrt{n\tau(1-\tau)} \left[\frac{\widehat{\text{ES}}_\tau(\mathbf{x})}{\text{ES}_\tau(\mathbf{x})} - 1 \right] \xrightarrow{d} (1-\gamma)h^{-\gamma}(\mathbf{x})\mathbf{x}'\boldsymbol{\Omega}_*^{1/2}\mathbf{W}_m(1),$$

where $\mathbf{W}_m(\cdot)$ is the Wiener process from Theorem 2.

Theorem 3 covers ES estimation for the intermediate case, where $\tau \rightarrow 1$ but sufficiently slowly to ensure that $n(1-\tau) \rightarrow \infty$. In this sense, the probability level τ still corresponds to a “within sample” region, as $1-\tau$ decays to zero more slowly than $1/n$.

3.2 Extremal ES Estimation

We now consider the case of extreme ES estimation, where we extrapolate outside of the range of the available data. To do so, we consider probability levels $\bar{\tau}$ satisfying that $1-\bar{\tau}$ vanishes more quickly than $1/n$. We specify the decay rate in more detail in Assumption 7 below. Our extrapolation-based ES estimator is based on the fact that under Assumption 6,

$$\text{ES}_{\bar{\tau}}(\mathbf{x}) \approx \left(\frac{1-\bar{\tau}}{1-\tau} \right)^{-\gamma} \text{ES}_\tau(\mathbf{x});$$

see the proof of Theorem 4 for detail. This suggests estimating the “beyond the sample” $\text{ES}_{\bar{\tau}}(\mathbf{x})$ by first estimating the “within sample” $\text{ES}_\tau(\mathbf{x})$ and then using the tail shape γ to

extrapolate to the desired level $\bar{\tau}$. The intermediate $\text{ES}_\tau(\mathbf{x})$ is easily estimable from (4). The other unknown, γ , can be estimated via (e.g.) the Hill (1975) estimator

$$\widehat{\gamma}_H = \widehat{\gamma}_H(k') = \frac{1}{k'} \sum_{i=1}^{k'} \log(Y_{n-i+1:n}/Y_{n-k':n}),$$

where $Y_{1:n} \leq \dots \leq Y_{n:n}$ denote the order statistics of Y_1, \dots, Y_n , and k' is an intermediate sequence of integers satisfying that $k' \rightarrow \infty$ and $k'/n \rightarrow 0$, as $n \rightarrow \infty$. But any other estimator $\widehat{\gamma} = \widehat{\gamma}(k')$ can be used as well. Combining the two components, we propose the extremal ES estimator

$$\widetilde{\text{ES}}_{\bar{\tau}}(\mathbf{x}) = \left(\frac{1 - \bar{\tau}}{1 - \tau} \right)^{-\widehat{\gamma}} \widehat{\text{ES}}_{\tau}(\mathbf{x}). \quad (5)$$

To derive the asymptotic properties of this estimator, we impose the following additional conditions.

Assumption 7. *It holds that $\lim_{n \rightarrow \infty} n(1 - \bar{\tau})/k = 0$ and $\lim_{n \rightarrow \infty} k^{-1/2} \log(n[1 - \bar{\tau}]) = 0$.*

Assumption 8. *There exists an estimator $\widehat{\gamma} = \widehat{\gamma}(k')$, such that $\sqrt{k'}(\widehat{\gamma} - \gamma) \xrightarrow{d} \Gamma$, as $n \rightarrow \infty$, for some limiting distribution Γ and intermediate sequence $k' \sim k$.*

In Assumption 7, the first condition $\lim_{n \rightarrow \infty} n(1 - \bar{\tau})/k = 0$ ensures that $\bar{\tau}$ is more extreme than $\tau = 1 - k/n$. On the other hand, the second condition $\lim_{n \rightarrow \infty} k^{-1/2} \log(n[1 - \bar{\tau}]) = 0$ prevents $1 - \bar{\tau}$ from decaying to zero too fast (i.e., it defines the limit past which extrapolation cannot be carried out). Assumption 8 is a high-level condition. It can be shown to be satisfied by the above Hill (1975) estimator under some mild additional assumptions, such as conditions (C2) and (C3) from Hoga (2017) (where k and X_i in his (C2) and (C3) correspond to k' and Y_t in our notation).

Theorem 4. *Under Assumptions 1–8 and $\gamma \in (0, 1/4)$, it holds that, as $n \rightarrow \infty$,*

$$\frac{\sqrt{n\tau(1 - \tau)}}{\log(k/[n(1 - \bar{\tau})])} \left[\frac{\widetilde{\text{ES}}_{\bar{\tau}}(\mathbf{x})}{\text{ES}_{\tau}(\mathbf{x})} - 1 \right] \xrightarrow{d} \Gamma.$$

The limiting distribution in Theorem 4—unlike that of Theorem 3—is independent of the covariates. This is because the limiting behavior is determined solely by the extreme

value index estimator; see the proof of Theorem 4. Moreover, we see that relative to the intermediate case in Theorem 3, the convergence rate of the extremal ES estimator from Theorem 4 is reduced by the factor $1/\log(k/[n(1-\bar{\tau})]) = o(1)$. In particular, for more extreme extrapolation levels $\bar{\tau}$ (i.e., for $1-\bar{\tau}$ closer to zero), the speed of convergence is further reduced.

Remark 1. It may also be of interest to obtain confidence intervals for the extreme $\text{ES}_{\bar{\tau}}(\mathbf{x})$. This is also possible by using SN, if we strengthen Assumption 8. Specifically, we require subsample estimates $\hat{\gamma}(s) = \hat{\gamma}(s, k')$ based on $Y_1, \dots, Y_{[ns]}$ that satisfy

$$s\sqrt{k'}[\hat{\gamma}(s) - \gamma] \xrightarrow{d} \sigma W(s) \quad \text{in } D[\epsilon, 1],$$

as $n \rightarrow \infty$, for some $\sigma > 0$ and a standard Wiener process $W(\cdot)$. For instance, under the additional above-mentioned conditions (C2)–(C3) in Hoga (2017), this holds true for the Hill (1975) subsample estimator

$$\hat{\gamma}(s, k') = \frac{1}{[k's]} \sum_{i=1}^{[k's]} \log(Y_{[ns]-i+1:[ns]}/Y_{[ns]-[k's]:[ns]})$$

that uses $[k's]$ -many upper order statistics. Arguing similarly as in the proof of Theorem 4, it then follows that, as $n \rightarrow \infty$,

$$s \frac{\sqrt{n\tau(1-\tau)}}{\log(k/[n(1-\bar{\tau})])} \left[\frac{\widetilde{\text{ES}}_{\bar{\tau}}(s, \mathbf{x})}{\text{ES}_{\bar{\tau}}(\mathbf{x})} - 1 \right] \xrightarrow{d} \sigma W(s) \quad \text{in } D[\epsilon, 1]$$

for the extremal ES estimator

$$\widetilde{\text{ES}}_{\bar{\tau}}(s, \mathbf{x}) := \left(\frac{1-\bar{\tau}}{1-\tau} \right)^{-\hat{\gamma}(s)} \widehat{\text{ES}}_{\bar{\tau}}(s, \mathbf{x}) := \left(\frac{1-\bar{\tau}}{1-\tau} \right)^{-\hat{\gamma}(s)} \mathbf{x}' \hat{\boldsymbol{\beta}}(s)$$

that is based on the first $1, \dots, [ns]$ observations. Then, we deduce as in Corollary 1 that

$$\frac{[\widetilde{\text{ES}}_{\bar{\tau}}(\mathbf{x}) - \text{ES}_{\bar{\tau}}(\mathbf{x})]^2}{\frac{1}{n} \sum_{j=[n\epsilon]+1}^n \left(\frac{j}{n}\right)^2 [\widetilde{\text{ES}}_{\bar{\tau}}(j/n, \mathbf{x}) - \widetilde{\text{ES}}_{\bar{\tau}}(1, \mathbf{x})]^2} \xrightarrow{d} \frac{W(1)^2}{\int_{\epsilon}^1 [W(s) - sW(1)]^2} =: U,$$

where the limiting distribution U coincides with that of Corollary 1 for $m = 1$ and $\mathbf{B} = \mathbf{1}$.

This implies the following asymptotic $(1 - \nu)$ -confidence interval for the extremal $\text{ES}_{\bar{\tau}}(\mathbf{x})$:

$$I_{1-\nu} := \left[\widetilde{\text{ES}}_{\bar{\tau}}(\mathbf{x}) \pm \sqrt{\frac{U_{1-\nu}}{n} \sum_{j=\lfloor n\epsilon \rfloor + 1}^n \left(\frac{j}{n}\right)^2 [\widetilde{\text{ES}}_{\bar{\tau}}(j/n, \mathbf{x}) - \widetilde{\text{ES}}_{\bar{\tau}}(1, \mathbf{x})]^2} \right],$$

where $U_{1-\nu}$ denotes the $(1 - \nu)$ -quantile of U ; see Table B1 (row “ $m = 1$ ”) in Appendix B for some selected quantiles of this limiting distribution.

4 Simulations

Here, we present some Monte Carlo simulations. We first describe the data-generating process (DGP) in Section 4.1. For this DGP, we then illustrate the finite-sample properties of our test from Corollary 1 in Section 4.2. Finally, Section 4.3 investigates the accuracy of the intermediate ES estimator $\widehat{\text{ES}}_{\tau}(\mathbf{x})$ and the extremal ES estimator $\widehat{\text{ES}}_{\bar{\tau}}(\mathbf{x})$ from Section 3.

4.1 The Data-Generating Process

We consider a time series regression as a DGP, which exhibits heteroskedasticity and autocorrelation in the errors. The set-up is comparable to that of Dimitriadis and Bayer (2019, Sec. 4.1), except that the regressors and errors are not i.i.d., but follow an AR(1) process here. Specifically, the data $\{(Y_t, \mathbf{X}_t')'\}_{t=1, \dots, n}$ for all simulations is generated as follows.

The stochastic regressor x_t in $\mathbf{X}_t = (1, x_t)'$ follows the transformed autoregression $x_t = \Phi(z_t)$, where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function (c.d.f.) and

$$z_t = \rho_z z_{t-1} + \epsilon_t, \quad |\rho_z| < 1, \quad \epsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1 - \rho_z^2). \quad (6)$$

The variance $1 - \rho_z^2$ of ϵ_t is chosen to imply that $z_t \sim N(0, 1)$ in (6). In particular, $x_t = \Phi(z_t)$ follows a uniform distribution on $[0, 1]$, i.e., $x_t \sim U[0, 1]$. This ensures that our covariates have compact support, as required by Assumption 5.

The dependent variable Y_t is generated from

$$Y_t = \mathbf{X}_t' \boldsymbol{\delta} + (\mathbf{X}_t' \boldsymbol{\eta}) e_t, \quad \boldsymbol{\delta} \in \mathbb{R}^2, \quad \boldsymbol{\eta} \in (0, \infty) \times (0, \infty). \quad (7)$$

To ensure heavy tails of the conditional distribution of $Y_t \mid \mathbf{X}_t$, we draw e_t from a (symmetrized) Burr distribution of type XII. Recall that a Burr(a, κ) distribution of type XII has the c.d.f.

$$G(x) = 1 - (1 + x^a)^{-\kappa}, \quad x > 0, \quad a, \kappa > 0,$$

and quantile function $G^{-1}(u) = [(1-u)^{-1/\kappa} - 1]^{1/a}$. This distribution has extreme value index $\gamma = 1/(a\kappa)$ and second-order parameter $\xi = -a$ in the notation of Assumption 6 (iii) (see [Hua and Joe, 2011](#), Example 2). The symmetrized Burr distribution (i.e., the distribution of $S \cdot X$ for $X \sim \text{Burr}(a, \kappa)$ and Rademacher-distributed S , independent of X) then has quantile function

$$F_{\text{sym-Burr}}^{-1}(u) = \begin{cases} -G^{-1}(1 - 2u), & u < 0.5, \\ G^{-1}(2u - 1), & u \geq 0.5. \end{cases}$$

We now generate the e_t from the transformed autoregression $e_t = F_{\text{sym-Burr}}^{-1}(u_t)$, where $u_t = \Phi(d_t)$ and

$$d_t = \rho d_{t-1} + \nu_t, \quad |\rho| < 1, \quad \nu_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1 - \rho^2). \quad (8)$$

Note that we choose $\text{Var}(\nu_t) = 1 - \rho^2$ in (8) to guarantee that $\text{Var}(d_t) = \frac{\text{Var}(\nu_t)}{1 - \rho^2} = 1$ is independent of ρ . Therefore, $d_t \sim N(0, 1)$ for the Gaussian autoregression in (8) for all $|\rho| < 1$, such that $u_t = \Phi(d_t) \sim U[0, 1]$ and, hence, $e_t \sim F_{\text{sym-Burr}}$ as desired.

Note that the autoregressions in (6) and (8) are geometrically β -mixing by [Liebscher \(2005, Sec. 8\)](#). By independence of d_t and z_t as well as standard mixing equalities, it then also holds that $(Y_t, \mathbf{X}_t)'$ is geometrically β -mixing, such that $\beta(l) \leq Kb^l$ for some $b \in (0, 1)$.

For the time series regression in (7), the true conditional (quantile, ES) are linear functions in \mathbf{X}_t given by

$$Q_\tau(Y_t \mid \mathbf{X}_t) = \mathbf{X}_t' \boldsymbol{\delta} + (\mathbf{X}_t' \boldsymbol{\eta}) Q_\tau(e_t), \quad \text{ES}_\tau(Y_t \mid \mathbf{X}_t) = \mathbf{X}_t' \boldsymbol{\delta} + (\mathbf{X}_t' \boldsymbol{\eta}) \text{ES}_\tau(e_t),$$

where $Q_\tau(e_t) := F_{e_t}^{-1}(\tau)$ denotes the τ -quantile of e_t , and $\text{ES}_\tau(e_t) := \mathbb{E}[e_t \mid e_t \geq Q_\tau(e_t)]$. This implies for the true coefficients $\boldsymbol{\alpha}_{0\tau} = (\alpha_{0\tau,1}, \alpha_{0\tau,2})'$ and $\boldsymbol{\beta}_{0\tau} = (\beta_{0\tau,1}, \beta_{0\tau,2})'$ in (1)–(2) that $\boldsymbol{\alpha}_{0\tau} = \boldsymbol{\delta} + \boldsymbol{\eta} Q_\tau(e_t)$ and $\boldsymbol{\beta}_{0\tau} = \boldsymbol{\delta} + \boldsymbol{\eta} \text{ES}_\tau(e_t)$.

We set $\rho_z = \rho = 0.2$ for the autoregressive coefficients, $\boldsymbol{\delta} = (\delta_1, \delta_2)' = (1, 1)'$, and $\boldsymbol{\eta} = (\eta_1, \eta_2)' = (1, 1)'$. For the parameters of the Burr distribution, we take $(a, \kappa) \in \{(2, 3), (6, 1)\}$. This ensures that $\gamma = 1/(a\kappa) = 1/6$ (such that $\gamma \in (0, 1/4)$ in Assumption 6), yet the second-order parameter $\gamma = -a$ is different across both choices of (a, κ) . The sample sizes we consider are $n \in \{500, 1000, 2000, 4000\}$. Throughout, we use 5000 replications.

4.2 Finite-Sample Properties

We report size and power of our tests based on \mathcal{Q}_n and \mathcal{E}_n from Section 2.4. For both test statistics, we choose $\epsilon = 0.2$ for the trimming parameter. Consistent with the notion that the quantile levels converge to unity for larger samples, we consider $\tau \in \{0.8, 0.85, 0.9\} / \tau \in \{0.85, 0.9, 0.95\} / \tau \in \{0.9, 0.95, 0.975\} / \tau \in \{0.95, 0.975, 0.99\}$ for the sample sizes $n = 500 / n = 1000 / n = 2000 / n = 4000$, respectively. Then, we test the null hypotheses

$$H_0^{\mathcal{Q}}: \boldsymbol{\beta}_{0\tau} = \boldsymbol{\delta} + \boldsymbol{\eta} Q_\tau(e_t) + \Delta \mathbf{1} \quad \text{and} \quad H_0^{\mathcal{E}}: \boldsymbol{\beta}_{0\tau} = \boldsymbol{\delta} + \boldsymbol{\eta} \text{ES}_\tau(e_t) + \Delta \mathbf{1},$$

where $\mathbf{1}$ denotes a vector of ones, and $\Delta \in \mathbb{R}$. We do so using the test statistics \mathcal{Q}_n and \mathcal{E}_n , respectively, with $\mathbf{A} = \mathbf{B} = \mathbf{I}_{2 \times 2}$. To investigate size, we set $\Delta = 0$, and for power we use $\Delta \in \{0.1, 0.2\}$.

Table 1 shows the results. We see that the tests are somewhat liberal under the null ($\Delta = 0$). Keeping n fixed, however, these liberal tendencies are lower for less extreme probability levels. Also, size distortions are lower for ES regressions when compared with the quantile regressions. This is surprising because the former are based on a smaller effective sample size than the latter.

As expected, we find that power increases in $|\Delta|$ (i.e., in the distance from the truth), and it increases in n for *fixed* quantile level τ . Finally, even taking into account the differences

Table 1: Rejection frequencies (in %) for tests based on \mathcal{Q}_n and \mathcal{E}_n

n	(a, κ)	τ	$\Delta = 0$		$\Delta = 0.1$		$\Delta = 0.2$	
			\mathcal{Q}_n	\mathcal{E}_n	\mathcal{Q}_n	\mathcal{E}_n	\mathcal{Q}_n	\mathcal{E}_n
500	(2, 3)	80%	6.86	5.14	43.12	31.70	86.46	69.20
		85%	7.68	6.48	39.52	27.48	81.74	61.12
		90%	8.24	6.02	32.34	22.98	71.26	50.74
	(6, 1)	80%	6.72	5.70	58.16	41.08	94.40	82.08
		85%	7.02	6.80	54.18	34.88	91.96	74.48
		90%	7.98	7.28	45.04	30.08	84.62	61.90
1000	(2, 3)	85%	6.96	6.04	60.48	39.60	95.00	81.52
		90%	6.98	5.32	48.26	31.02	89.80	69.30
		95%	8.26	6.78	32.42	22.30	72.00	48.56
	(6, 1)	85%	6.78	5.66	75.92	50.96	98.80	90.36
		90%	7.24	6.58	65.38	40.64	96.72	79.68
		95%	8.14	7.66	43.66	27.30	84.44	58.66
2000	(2, 3)	90%	6.74	5.48	71.82	45.64	97.96	88.12
		95%	6.54	5.82	50.54	29.68	89.78	66.88
		97.5%	8.86	6.62	32.86	21.34	70.14	44.94
	(6, 1)	90%	7.02	5.94	85.86	57.80	99.52	93.58
		95%	6.64	5.62	63.22	37.14	95.84	75.56
		97.5%	7.58	7.88	40.92	25.34	80.82	52.18
4000	(2, 3)	95%	7.10	5.78	72.00	44.44	97.90	85.56
		97.5%	7.06	6.60	48.28	28.08	88.92	62.58
		99%	9.06	8.16	27.84	18.56	59.48	35.56
	(6, 1)	95%	6.52	5.78	85.52	53.86	99.58	91.70
		97.5%	7.68	6.70	60.54	33.90	94.88	70.72
		99%	8.16	7.70	31.24	21.14	69.56	42.02

NOTE.— The parameters (a, κ) refer to the parameters of the Burr distribution.

in size, power is higher for the quantile regressions (columns \mathcal{Q}_n) than for the ES regressions (columns \mathcal{E}_n), because for the latter the effective sample size is much reduced. The results for size and power are qualitatively unchanged for the two distributions of e_t we consider.

4.3 Intermediate and Extremal ES Estimation

We now compare the performance of our extrapolation-based estimator $\widetilde{\text{ES}}_{\bar{\tau}}(\mathbf{x})$ with the direct regression-based estimate $\widehat{\text{ES}}_{\bar{\tau}}(\mathbf{x})$. We continue to use the DGP from Section 4.1. For the quantile levels, we use $\tau = 0.9$ and $\bar{\tau} = 0.975$ / $\tau = 0.95$ and $\bar{\tau} = 0.99$ / $\tau = 0.975$

Table 2: Bias and RMSE for ES estimators

n	(a, κ)	Bias		RMSE	
		$\widetilde{\text{ES}}_{\bar{\tau}}(\mathbf{x})$	$\widehat{\text{ES}}_{\bar{\tau}}(\mathbf{x})$	$\widetilde{\text{ES}}_{\bar{\tau}}(\mathbf{x})$	$\widehat{\text{ES}}_{\bar{\tau}}(\mathbf{x})$
500	(2, 3)	16.2	-1.5	29.4	25.6
	(6, 1)	10.1	-0.6	22.5	23.9
1000	(2, 3)	13.1	-0.5	31.5	33.3
	(6, 1)	0.8	-1.1	23.3	30.9
2000	(2, 3)	8.3	-0.4	31.4	36.7
	(6, 1)	-3.6	-0.0	26.0	34.1
4000	(2, 3)	3.1	-0.8	34.6	40.6
	(6, 1)	-4.9	-0.2	30.7	38.0

NOTE.— Bias and RMSE for extrapolation-based estimate $\widetilde{\text{ES}}_{\bar{\tau}}(\mathbf{x})$ and direct regression-based estimate $\widehat{\text{ES}}_{\bar{\tau}}(\mathbf{x})$, where $\mathbf{x} = (1, 0.5)'$. All results pre-multiplied with 100 for better readability.

and $\bar{\tau} = 0.995 / \tau = 0.99$ and $\bar{\tau} = 0.9975$ for $n = 500 / n = 1000 / n = 2000 / n = 4000$, respectively. Of course, the τ -levels are only relevant for $\widetilde{\text{ES}}_{\bar{\tau}}(\mathbf{x})$, which extrapolates from the regression-based estimate of $\text{ES}_{\tau}(\mathbf{x})$ to the desired level of $\bar{\tau}$ by exploiting the tail shape, cf. (5). For the estimator of the extreme value index, we use the Hill (1975) estimator $\widehat{\gamma}_H = \widehat{\gamma}_H(k')$ with $k' = \lfloor k \rfloor$. We fix the regressors at $\mathbf{x} = (1, 0.5)'$.

Table 2 displays the results. While the extrapolation-based estimator $\widetilde{\text{ES}}_{\bar{\tau}}(\mathbf{x})$ has a larger bias, the root mean squared error (RMSE) is almost always lower. Therefore, extrapolation invariably incurs a bias, which is, however, overcompensated by a reduction in variance. Of course, the direct estimator $\widehat{\text{ES}}_{\bar{\tau}}(\mathbf{x})$ has an inflated variance because it is computed based on an effective sample size of the order $n(1 - \bar{\tau})$, which is small in all considered cases. Hence, a reduction in variance is key for estimating extremal quantities.

Comparing the results for $(a, \kappa) = (2, 3)$ and $(a, \kappa) = (6, 1)$, we find the following. Recall that in both cases the tails are equally heavy as measured by the extreme value index of $\gamma = 1/6$. However, for $(a, \kappa) = (6, 1)$, the approximation to the Pareto distribution is more accurate as the second order parameter $\xi = -a$ is more negative. This is reflected in the much reduced RMSEs for $\widetilde{\text{ES}}_{\bar{\tau}}(\mathbf{x})$ compared with $\widehat{\text{ES}}_{\bar{\tau}}(\mathbf{x})$ in the case $(a, \kappa) = (6, 1)$ relative to $(a, \kappa) = (2, 3)$. For instance, when $n = 1000$, the RMSEs for both estimators are 31.5 and

33.3 for $(a, \kappa) = (2, 3)$, yet 23.3 and 30.9 for $(a, \kappa) = (6, 1)$, which amounts to a much larger improvement for the extrapolation-based estimator.

Finally, we illustrate the results visually. We do so for $(a, \kappa) = (6, 1)$ by plotting kernel-density estimates (KDEs) of the distribution of our ES estimators. In Figure 1, the results are shown in blue for $\widehat{\text{ES}}_\tau(\mathbf{x})$, in red for the extrapolation-based estimate $\widetilde{\text{ES}}_\tau(\mathbf{x})$, and the dashed red line corresponds to the direct regression-based estimate $\widehat{\text{ES}}_\tau(\mathbf{x})$. We indicate the true values of $\text{ES}_\tau(\mathbf{x})$ and $\text{ES}_{\bar{\tau}}(\mathbf{x})$ by the vertical lines. Figure 1 shows that ES for less extreme quantile levels is easier to estimate, because the KDEs are centered accurately around the true values with little spread. For the more extreme quantile levels, direct regression-based estimates deteriorate by displaying higher variance around the true value than our proposed extrapolation-based estimates.

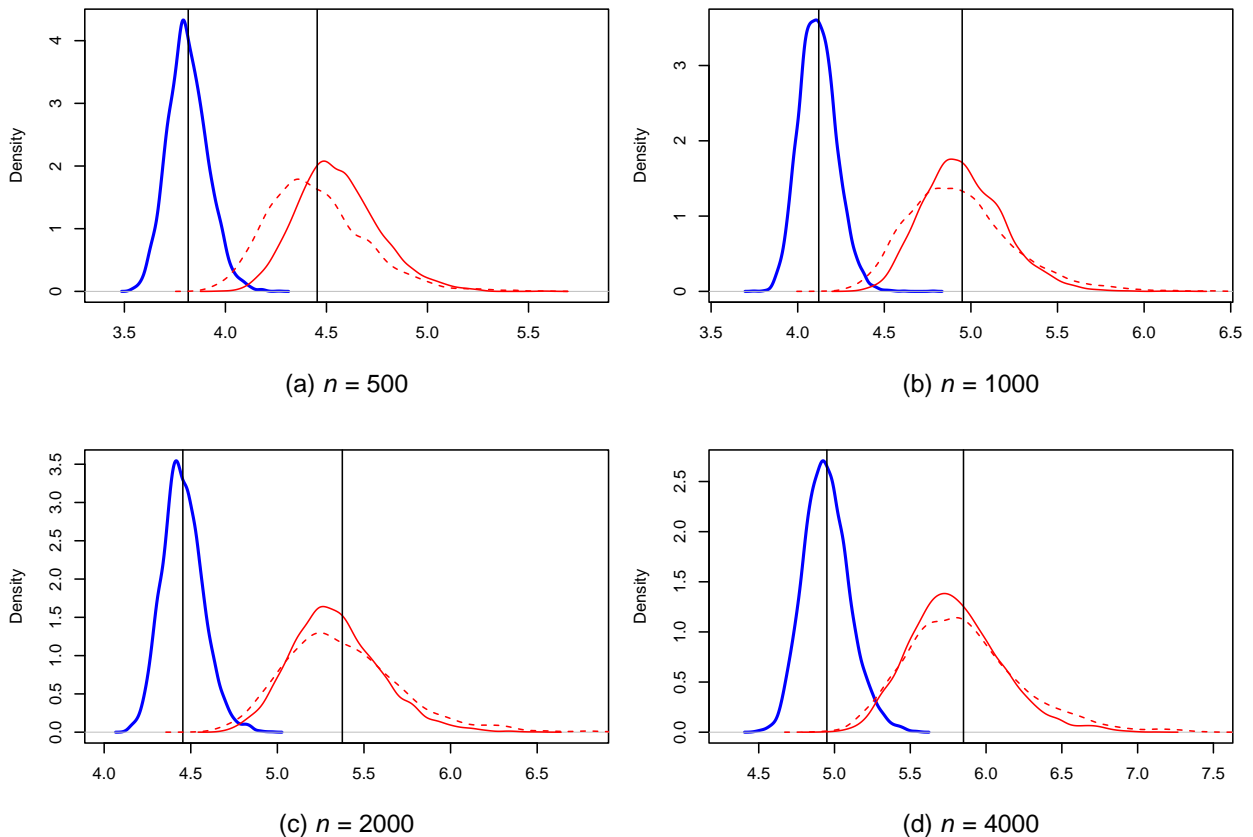


Figure 1: Kernel-density estimates of ES estimators

NOTE.— Kernel-density estimates of $\widehat{\text{ES}}_\tau(\mathbf{x})$ (blue), $\widetilde{\text{ES}}_\tau(\mathbf{x})$ (red) and $\widehat{\text{ES}}_\tau(\mathbf{x})$ (dashed red) for different values of n . True values $\text{ES}_\tau(\mathbf{x})$ and $\text{ES}_{\bar{\tau}}(\mathbf{x})$ indicated by the vertical lines.

5 Empirical Application

To illustrate how the proposed extremal ES methods work in practice, we study the effect of temperature on health care expenditure in Germany. Building on previous evidence that healthcare expenditure exhibits heavy tails (Karlsson et al., 2024), this application sheds light on how the temperature influences health spending. Understanding these tail responses is particularly relevant in the context of climate change, as rising temperatures may exacerbate health risks and associated costs (Deschênes and Greenstone, 2011).

A substantial body of research has documented that ambient temperature significantly affects human health, with consequences for health care utilization. Temperature extremes—both hot and cold—have been linked to elevated mortality and morbidity across settings, reflecting physiological stress on cardiovascular, respiratory, and thermoregulatory systems (e.g., Deschênes and Greenstone, 2011; Barreca et al., 2016). In Germany, comprehensive analyses of hospital admissions and mortality show that extreme heat increases hospitalizations and deaths, with economically meaningful health-related costs on hot days (Karlsson and Ziebarth, 2018). These effects are consistent with physiological mechanisms: high temperatures impair heat dissipation and increase cardiovascular strain, whereas cold stress induces vasoconstriction and respiratory burden; both can precipitate adverse health events.

Other meteorological conditions influence health through related physiological and environmental pathways. Humidity or vapor pressure affects the body’s ability to dissipate heat, as high moisture reduces evaporative cooling and amplifies heat stress, particularly at elevated temperatures (Zhang et al., 2014). Cloud coverage can indirectly affect health by modifying ambient pollution: it reduces solar radiation and alters atmospheric mixing, which may raise near-surface pollutant concentrations (Tai et al., 2010). Short-term exposure to elevated pollution increases hospital admissions and health care utilization (e.g., Chay and Greenstone, 2003; Graff Zivin and Neidell, 2013), suggesting cloud coverage contributes to health risks via pollutant exposure. Including these variables allows us to account for mechanisms that moderate the temperature–expenditure relationship.

5.1 Data

Our analysis combines detailed claims data from a large German private health insurer with daily weather data from the *Deutscher Wetterdienst* (DWD) to study the relationship between temperature and health care expenditure.

Health Care Claims. The claims dataset covers the universe of insurance plans and claims between 2005 and 2011 for policyholders with comprehensive private health insurance (i.e., without supplemental plans). It contains over 2.6 million enrollee-year observations from approximately 620 thousand unique policyholders, with detailed information on premiums, claims, diagnoses, age, gender, and occupational group (Atal et al., 2019, 2025). All monetary values are converted to 2016 Euros.

Sample Selection. We include only policyholders continuously enrolled throughout 2005–2011. Individuals insured under civil servant plans are excluded, as reported expenditures do not reflect full costs. Policyholders with annual deductibles above 500 Euros are also dropped due to incomplete spending data. ZIP codes are used to map individuals to counties (*Kreise*) and federal states (*Bundesländer*). Daily mean expenditure per *Bundesland* is computed by aggregating all claims that start on a given day. This aggregation ensures that the complete costs of an episode are attributed to the correct start date. Because the panel is balanced across years, the daily averages are not affected by compositional changes, apart from ageing.

Health Expenditure. Our dependent variable is the daily average health care expenditure per *Bundesland*. Each claim is attributed to the reported start of treatment date, and daily averages are computed across all policyholders within the *Bundesland*.

Weather Variables. We merge the expenditure data with daily weather observations from the *Deutscher Wetterdienst* (DWD, 2026). The key variable is **maximum daily temperature at 2m height**, averaged across all stations within a *Bundesland* to obtain a

representative measure. In addition, the DWD collects a broad set of routine measurements, including minimum temperature, mean temperature, dew point, relative humidity, cloud coverage, wind speed, wind gusts, and precipitation. We include those variables found to be significant in at least some specifications.

Table 3 provides descriptive statistics for all variables used in the analysis, both for the full dataset and for the subsample of the three largest states, which is the focus of the extreme ES analysis.

Table 3: Descriptive Statistics

	Mean	SD	Min	Max	N
COMPLETE SAMPLE					
Health care expenditure	6.62	6.79	0.00	258.97	40,896
Temperature	13.59	8.56	-14.31	37.67	40,896
Cloud coverage	5.36	2.06	0.00	8.10	40,896
Vapor pressure	9.822	3.964	1.229	22.814	40,896
Time (years)	-0.001	2.020	-3.500	3.497	40,896
THREE LARGEST STATES					
Health care expenditure	7.28	4.90	0.02	66.00	7,668
Temperature	13.81	8.71	-10.27	34.92	7,668
Cloud coverage	5.33	2.01	0.04	8.00	7,668
Vapor pressure	9.706	4.004	1.691	21.230	7,668
Time (years)	-0.001	2.020	-3.500	3.497	7,668

NOTE.— Expenditure is measured in 2016 Euro and represents daily average per *Bundesland*. Time measures the years since 1 July 2008.

5.2 Data Exploration and Specification Checks

Before conducting the extreme ES analysis, we explore the distribution of daily health care expenditure and examine the functional form of its temperature gradient. This exploration serves several purposes. First, we assess whether the distribution of expenditures exhibits heavy-tailed behavior, which motivates the use of tail-extrapolation methods in the ES analysis. Second, we visually inspect the relationship between temperature and spending using binscatter plots, conditioning on key fixed effects. Finally, we estimate a range of OLS specifications to determine which temperature variables (current and lagged) and additional

weather controls are the most important predictors. The results of these steps guide the choice of specification used in the subsequent ES estimation.

We begin with rank-size plots for three large *Bundesländer*, shown in Figure 2, where the logarithm of the rank of each observation is plotted against the logarithm of the corresponding expenditure. These plots provide a simple visual check for heavy-tailed, Pareto-type behavior in the expenditure distribution. Across the three states, the upper tail of the distribution exhibits a roughly linear pattern on the log-log scale, consistent with a Pareto-type tail. This observation supports the use of tail-extrapolation methods in the later ES analysis.

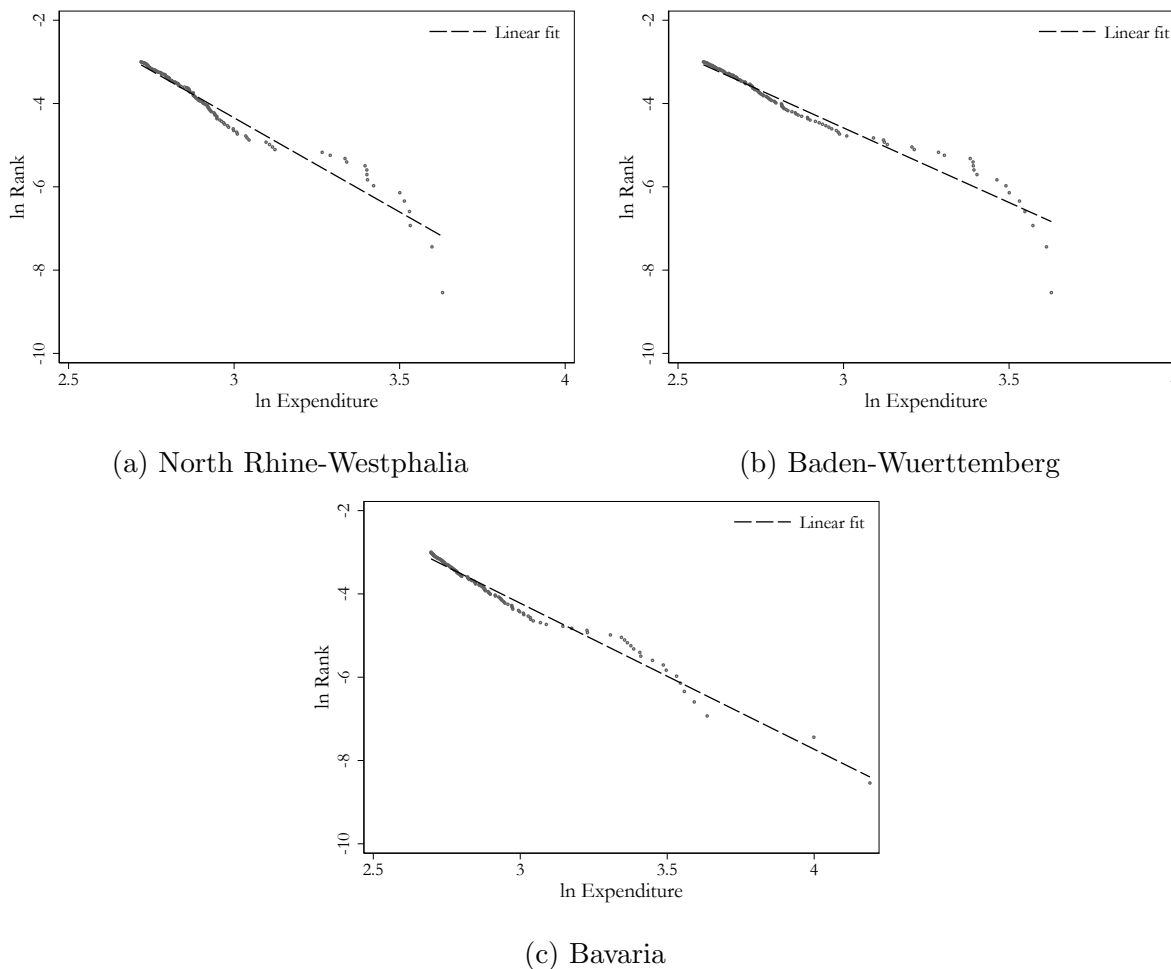


Figure 2: Rank-Size Plots of Natural Logarithms of Rank against Expenditures.

NOTE.— Plots the logarithm of the rank of daily expenditures against the logarithm of total daily expenditures, aggregated at the state level. Expenditures correspond to the total claim amounts attributed to the reported start-of-treatment date.

The materials reported in Appendix C indicate that the expenditure–temperature rela-

relationship is non-linear, roughly quadratic (cf. Figure C1), that current temperature matters more than lagged temperature, and that among additional weather controls, cloud coverage and vapor pressure appear most relevant (cf. Tables C1–C2). Based on these exploratory and specification checks, our ES analyses focus on a model that treats health expenditure as a quadratic function of current temperature, controlling for cloud coverage and vapor pressure, since these variables are significant in some specifications.

5.3 Results

Table 4 presents the results of the extremal ES regressions for the three largest states. We consider the quantile $\tau = 0.9$ — and two specifications: a *short* specification including only the quadratic temperature terms, and a *long* specification additionally controlling for cloud coverage and vapor pressure.

For each *Bundesland* and specification, we report results from the two-step ES estimation procedure. In the first step, we estimate a quantile regression to obtain the conditional quantile. In the second step, we estimate the ES regression parameters via an OLS regression on observations exceeding the predicted quantile. This approach leverages the fact that the ES corresponds to the conditional mean of the tail, while using only convex optimization steps for computational convenience. For comparison, the tables also report full-sample OLS estimates, with columns labeled “OLS”, alongside the QR and ES parameters.

Across states, we find notable heterogeneity in the expenditure–temperature relationship. In the short specification, Bavaria exhibits the strongest impact, while other states display more muted responses. Despite these differences in magnitude, the general shape of the relationship is consistent: in all but one specification across OLS, QR, and ES, the linear term is positive and the quadratic term is negative, supporting the notion of an increasing but concave response to higher temperatures. We also document that the tail parameter γ —the inverse of the tail index α —falls within the theoretically required range 0–0.25 in all cases. This range is particularly relevant because Theorem 2 requires $\gamma \in (0, 1/4)$ for ES

regressions. The pattern of estimates is therefore consistent with the theoretical constraints underlying the method and provides the basis for our extrapolation to the $\bar{\tau} = 0.995$ ES, which is the ultimate quantity of interest in the analysis (Boonen, 2017).

Table 4: ES Results: Impact of Weather Variables – $\tau = 0.9$

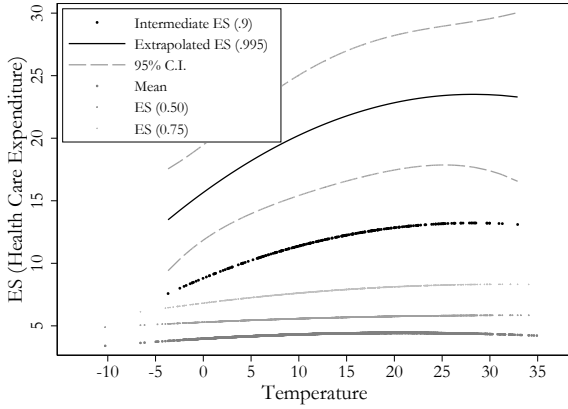
	NRW			BW			BAVARIA		
	OLS	QR	ES	OLS	QR	ES	OLS	QR	ES
SHORT SPECIFICATION									
Temperature	0.0448	0.0532	0.3123**	0.0458	0.0550	0.1809*	0.0752**	0.0528**	0.3501**
Temperature ²	−0.0011	−0.0012	−0.0055	−0.0012	−0.0011	−0.0030	−0.0018*	−0.0017*	−0.0072
n	2,556	2,556	257	2,556	2,556	248	2,556	2,556	258
$\hat{\gamma}$			0.192			0.228			0.202
LONG SPECIFICATION									
Temperature	0.0389	0.0616	0.1914	0.0282	0.0603	0.0580	0.0349	−0.0164	0.1715
Temperature ²	−0.0007	0.0003	−0.0061*	−0.0012	−0.0009	−0.0029	−0.0022**	−0.0007	−0.0056
Cloud coverage	0.0491	0.1630	−0.0984	−0.0035	−0.0020	−0.1590	−0.0378	−0.0146	0.0321
Vapor pressure	0.0047	−0.0735	0.3249	0.0579	−0.0409	0.2645	0.1340*	0.1750	0.5734
n	2,556	2,556	253	2,556	2,556	250	2,556	2,556	251
$\hat{\gamma}$			0.192			0.228			0.202

NOTE.— The table reports estimates from the two-step extremal expected shortfall (ES) regression and, for comparison, ordinary least squares (OLS). For each state (NRW = North Rhine–Westphalia, BW = Baden–Wuerttemberg), the first column shows OLS coefficients, QR” columns show coefficients from a quantile regression at level τ , and ES” columns show coefficients from the ES regression estimated using observations above the predicted quantile. The short specification includes only current temperature (linear and quadratic terms), while the long specification additionally controls for cloud coverage and vapor pressure. ES coefficients capture the marginal impact of a covariate on expected tail expenditure conditional on exceeding the τ -th quantile. Statistical significance is denoted by ***, **, and * for the 1%, 5%, and 10% levels, respectively.

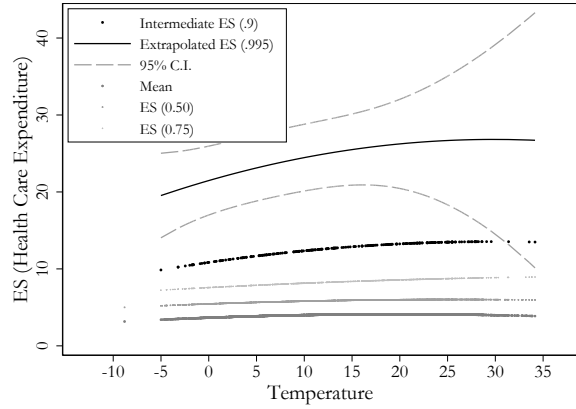
Figure 3 plots the estimated extreme ES for the temperature variable in each *Bundesland*. For every observed value of daily maximum temperature, we first compute the ES at the corresponding within-sample quantile using the fitted ES model (see above). We then extrapolate these values to $\bar{\tau} = 0.995$ using the tail-shape parameter $\hat{\gamma}$ obtained from a Hill-type estimator, as described in Theorem 4.

In each panel, the black scatter markers represent the within-sample ES estimates for $\tau = 0.9$, and the solid black line shows the extrapolated $\bar{\tau} = 0.995$ ES. For comparison, we also plot the conditional mean and the ES curves for $\tau = 0.50$ and $\tau = 0.75$ in lighter shades. These figures provide a visual depiction of tail risk associated with extreme temperature realizations and illustrate how extrapolated ES values increase with temperature.

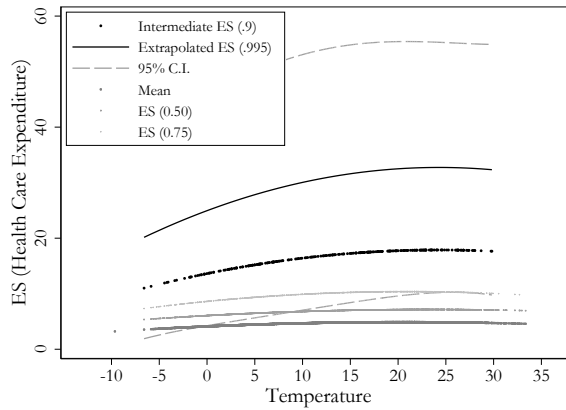
Across states, the figures reveal several consistent patterns. Despite the quadratic speci-



(a) North Rhine-Westphalia



(b) Baden-Wuerttemberg



(c) Bavaria

Figure 3: ES Estimates Health Expenditure – Short Specification, $\tau = 0.90$.

NOTE.— The figure shows the estimated expected shortfall (ES) of daily expenditure by temperature in each state. Circle markers represent the within-sample ES at $\tau = 0.90$. The solid black line shows the extrapolated $\bar{\tau} = 0.995$ ES, based on the Hill-type tail index $\hat{\gamma}$. Temperature is measured in $^{\circ}\text{C}$ as the daily state-level maximum.

fication, the maximum of daily spending is reached near the highest observed temperatures in all states and for all ES levels, indicating that expenditure increases with temperature globally, albeit at a decreasing rate. The relationship is notably steeper for ES estimates at higher τ levels, both within-sample and in the extrapolated $\bar{\tau} = 0.995$ values, highlighting greater sensitivity of extreme expenditures to temperatures. Finally, the confidence intervals indicate substantial heterogeneity in the precision of these estimates across states, with the temperature gradient estimated most precisely in Baden-Wuerttemberg.

Overall, our results support the notion that high temperatures are associated with adverse health outcomes, as reflected in increased health expenditures. Unlike most prior studies

(e.g., [Deschênes and Greenstone, 2011](#)), however, we do not observe a sharp strengthening of the health–temperature gradient at the very highest temperatures. This is consistent with previous evidence suggesting that spikes in poor health outcomes during heat waves are particularly pronounced for mortality, whereas utilization measures tend to be more sensitive to model specification ([Karlsson and Ziebarth, 2018](#)). In this sense, our findings are broadly aligned with the existing literature. At the same time, moving beyond the conditional mean reveals patterns that would otherwise remain hidden: extreme expenditures respond more strongly to high temperatures than the mean or median relationship suggests, highlighting the value of an ES-based analysis for capturing tail risks in health care utilization.

6 Conclusion

This paper contributes to statistical inference in risk measurement by developing novel methods for ES regressions at extreme probability levels. Unlike standard ES regressions that use a fixed probability level τ , we extend the framework to the case where τ approaches 1 as the sample size n increases—a scenario that is often more realistic in practice (e.g., in regulatory contexts such as Basel III or Solvency II). For inference, we follow [Hoga and Schulz \(2025\)](#) by using SN, which—in our current context—has the added benefit of obviating the need to estimate the complex tail-dependent convergence rates of estimators that arise in extreme value settings. Therefore, this paper shows that our SN-based test statistic is “double robust” in the sense that *neither* serial dependence *nor* vanishing $(1 - \tau)$ affect its limiting distribution. The proposed tests demonstrate good finite-sample size in simulations, confirming their practical reliability.

The empirical application illustrates the value of these methods in a microeconomic setting with direct policy relevance. Using detailed German health care claims data, we show that high temperatures disproportionately increase extreme health expenditures, even when average spending responds more moderately. Focusing on the upper tail of the expenditure

distribution, the analysis reveals heterogeneity in temperature sensitivity across regions and uncovers tail risks that mean-based approaches would overlook. These findings highlight how climate-related shocks may translate into rare but severe health care costs, underscoring the importance of tail-focused tools for assessing the economic consequences of extreme weather. Notably, the study also shows that extrapolating ES estimates beyond the observed data—common in risk modeling—can be conducted reliably using the proposed tools.

In summary, our work aims to advance applications of ES regressions by combining extreme value theory, modern inference techniques, and empirical relevance. By doing so, the paper provides a robust, easily implementable, and theoretically sound framework for analyzing extreme risks. It is particularly valuable for regulatory applications, tail-risk modeling, and assessing the impact of rare but severe events.

Disclosure Statement

The authors report there are no competing interests to declare.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- Abadie, A., J. Angrist, and G. Imbens**, “Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings,” *Econometrica*, 2002, *70* (1), 91–117.
- Adams, P. A., T. Adrian, N. Boyarchenko, and D. Giannone**, “Forecasting Macroeconomic Risks,” *International Journal of Forecasting*, 2021, *37* (3), 1173–1191.
- Adrian, T., N. Boyarchenko, and D. Giannone**, “Vulnerable Growth,” *The American Economic Review*, 2019, *109* (4), 1263–1289.
- Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath**, “Coherent Measures of Risk,” *Mathematical Finance*, 1999, *9*, 203–228.
- Atal, J. P., H. Fang, M. Karlsson, and N. R. Ziebarth**, “Exit, Voice, or Loyalty? An Investigation Into Mandated Portability of Front-Loaded Private Health Plans,” *Journal of Risk and Insurance*, 2019, *86* (3), 697–727.
- , —, —, and —, “German Long-Term Health Insurance: Theory Meets Evidence,” *Journal of Political Economy*, 2025, *133* (6), 1840–1885.

- Barendse, S.**, “Efficiently Weighted Estimation of Tail and Interquantile Expectations,” *Preprint*, 2023. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2937665.
- Barreca, A., K. Clay, O. Deschenes, M. Greenstone, and J. S. Shapiro**, “Adapting to Climate Change: The Remarkable Decline in the US Temperature-Mortality Relationship Over the Twentieth Century,” *Journal of Political Economy*, 2016, *124* (1), 105–159.
- Basel Committee on Banking Supervision**, *Minimum Capital Requirements for Market Risk*, Basel: Bank for International Settlements (<https://www.bis.org/bcbs/publ/d457.pdf>), 2019.
- Billingsley, P.**, *Convergence of Probability Measures*, 2nd ed., New York: Wiley, 1999.
- Boonen, T. J.**, “Solvency II Solvency Capital Requirement for Life Insurance Companies Based On Expected Shortfall,” *European Actuarial Journal*, 2017, *7* (2), 405–434.
- Carriero, A., T. E. Clark, and M. Marcellino**, “Nowcasting Tail Risk to Economic Activity at a Weekly Frequency,” *Journal of Applied Econometrics*, 2022, *37* (5), 843–866.
- Chay, K. Y. and M. Greenstone**, “The Impact of Air Pollution on Infant Mortality: Evidence From Geographic Variation in Pollution Shocks Induced by a Recession,” *The Quarterly Journal of Economics*, 2003, *118* (3), 1121–1167.
- Chen, Le-Yu and Yu-Min Yen**, “Estimation of the Local Conditional Tail Average Treatment Effect,” *Journal of Business & Economic Statistics*, 2025, *43* (1), 241–255.
- Chernozhukov, V.**, “Extremal Quantile Regression,” *The Annals of Statistics*, 2005, *33* (2), 806–839.
- and **I. Fernández-Val**, “Inference for Extremal Conditional Quantile Models, With an Application to Market and Birthweight Risks,” *The Review of Economic Studies*, 2011, *78* (2), 559–589.
- Daouia, A., G. Stupfler, and A. Usseglio-Carleve**, “Inference for Extremal Regression With Dependent Heavy-Tailed Data,” *The Annals of Statistics*, 2023, *51* (5), 2040–2066.
- Davidson, J.**, *Stochastic Limit Theory*, Oxford: Oxford University Press, 1994.
- Davis, R. A., H. Drees, J. Segers, and M. Warchoł**, “Inference on the tail process with application to financial time series modeling,” *Journal of Econometrics*, 2018, *205* (2), 508–525.
- Deschênes, O. and M. Greenstone**, “Climate Change, Mortality, and Adaptation: Evidence From Annual Fluctuations in Weather in the US,” *AEJ: Applied Economics*, 2011, *3* (4), 152–185.
- Dimitriadis, T. and S. Bayer**, “A Joint Quantile and Expected Shortfall Regression Framework,” *Electronic Journal of Statistics*, 2019, *13* (1), 1823–1871.
- Drees, H. and H. Rootzén**, “Limit Theorems for Empirical Processes of Cluster Functionals,” *The Annals of Statistics*, 2010, *38* (4), 2145–2186.
- DWD**, “Daily Weather Observations from DWD Weather Stations in Germany,” https://opendata.dwd.de/climate_environment/CDC/ 2026. Data accessed on 16 January 2026. Variables include daily maximum temperature, minimum temperature, mean temperature, dew point, relative humidity, cloud coverage, wind speed, wind gusts, and precipitation.
- Fissler, T.**, “On Higher Order Elicitability and Some Limit Theorems on the Poisson and Wiener Space.” PhD dissertation, University of Bern 2017.
- and **J. F. Ziegel**, “Higher Order Elicitability and Osband’s Principle,” *The Annals of Statistics*, 2016, *44* (4), 1680–1707.

- Galvao, A. F. and J. Yoon**, “HAC Covariance Matrix Estimation in Quantile Regression,” *Journal of the American Statistical Association*, 2024, *119*, 2305–2316.
- Gneiting, T.**, “Making and Evaluating Point Forecasts,” *Journal of the American Statistical Association*, 2011, *106* (494), 746–762.
- Gregory, K. B., S. N. Lahiri, and D. J. Nordman**, “A Smooth Block Bootstrap for Quantile Regression With Time Series,” *The Annals of Statistics*, 2018, *46* (3), 1138–1166.
- Hill, B.**, “A Simple General Approach to Inference About the Tail of a Distribution,” *The Annals of Statistics*, 1975, *3*, 1163–1174.
- Hill, J. B.**, “On Functional Central Limit Theorems for Dependent, Heterogeneous Arrays With Applications to Tail Index and Tail Dependence Estimation,” *Journal of Statistical Planning and Inference*, 2009, *139*, 2091–2110.
- Hoga, Y.**, “Change Point Tests for the Tail Index of β -Mixing Random Variables,” *Econometric Theory*, 2017, *33*, 915–954.
- , “A Structural Break Test for Extremal Dependence in β -Mixing Random Vectors,” *Biometrika*, 2018, *105*, 627–643.
- and **C. Schulz**, “Self-Normalized Inference in (Quantile, Expected Shortfall) Regressions for Time Series,” *Preprint*, 2025, pp. 1–60. <https://arxiv.org/pdf/2502.10065>.
- Hou, Y., X. Leng, L. Peng, and Y. Zhou**, “Panel Quantile Regression for Extreme Risk,” *Journal of Econometrics*, 2024, *240* (1), 105674.
- Hua, L. and H. Joe**, “Second Order Regular Variation and Conditional Tail Expectation of Multiple Risks,” *Insurance: Mathematics and Economics*, 2011, *49*, 537–546.
- Huang, W., S. Li, and L. Peng**, “Estimation and Inference for Extreme Continuous Treatment Effects,” *Journal of Business & Economic Statistics*, 2025, *43*, 822–834.
- Karlsson, M. and N. R. Ziebarth**, “Population Health Effects and Health-Related Costs of Extreme Temperatures: Comprehensive Evidence From Germany,” *Journal of Environmental Economics and Management*, 2018, *91*, 93–117.
- , **Y. Wang, and N. R. Ziebarth**, “Getting the Right Tail Right: Modeling Tails of Health Expenditure Distributions,” *Journal of Health Economics*, 2024, *97*.
- Kato, K.**, “Asymptotics for argmin Processes: Convexity Arguments,” *Journal of Multivariate Analysis*, 2009, *100* (8), 1816–1829.
- Kiefer, N. M. and T. J. Vogelsang**, “A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests,” *Econometric Theory*, 2005, *21* (6), 1130–1164.
- Knight, K.**, “Limit Theory for Autoregressive-Parameter Estimates in an Infinite-Variance Random Walk,” *The Canadian Journal of Statistics*, 1989, *17* (3), 261–278.
- Li, R., C. Leng, and J. You**, “Semiparametric Tail Index Regression,” *Journal of Business & Economic Statistics*, 2022, *40* (1), 82–95.
- Liebscher, E.**, “Towards a Unified Approach for Proving Geometric Ergodicity and Mixing Properties of Nonlinear Autoregressive Processes,” *Journal of Time Series Analysis*, 2005, *26* (5), 669–689.
- Lopez-Salido, D. and F. Loria**, “Inflation at Risk,” *Journal of Monetary Economics*, 2024, *145*. Article 103570.

- Monache, D. Delle, A. De Polis, and I. Petrella**, “Modeling and Forecasting Macroeconomic Downside Risk,” *Journal of Business & Economic Statistics*, 2024, *42* (3), 1010–1025.
- Nicolau, J., P. M. M. Rodrigues, and M. Z. Stoykov**, “Tail Index Estimation in the Presence of Covariates: Stock Returns’ Tail Risk Dynamics,” *Journal of Econometrics*, 2023, *235* (2), 2266–2284.
- Patton, A. J., J. F. Ziegel, and R. Chen**, “Dynamic Semiparametric Models for Expected Shortfall (and Value-at-Risk),” *Journal of Econometrics*, 2019, *211* (2), 388–413.
- Plagborg-Møller, M., L. Reichlin, G. Ricco, and T. Hasenzagl**, “When is Growth at Risk?,” *Brookings Papers on Economic Activity*, 2020, *2020* (1), 167–229.
- Shao, X.**, “A Self-Normalized Approach to Confidence Interval Construction in Time Series,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2010, *72* (3), 343–366.
- , “Self-Normalization for Time Series: A Review of Recent Developments,” *Journal of the American Statistical Association*, 2015, *110*, 1797–1817.
- Tai, A. P. K., L. J. Mickley, and D. J. Jacob**, “Correlations Between Fine Particulate Matter (PM_{2.5}) and Meteorological Variables in the United States: Implications for the Sensitivity of PM_{2.5} to Climate Change,” *Atmospheric Environment*, 2010, *44* (32), 3976–3984.
- Taylor, J. W.**, “Forecasting Value at Risk and Expected Shortfall Using a Semiparametric Approach Based on the Asymmetric Laplace Distribution,” *Journal of Business & Economic Statistics*, 2019, *37*, 121–133.
- Wang, H. and C. L. Tsai**, “Tail Index Regression,” *Journal of the American Statistical Association*, 2009, *104* (487), 1233–1240.
- Wei, B., K. M. Tan, and X. He**, “Estimation of Complier Expected Shortfall Treatment Effects With a Binary Instrumental Variable,” *Journal of Econometrics*, 2024, *238* (2), 105572.
- Weissman, I.**, “Estimation of Parameters and Large Quantiles Based on the k Largest Observations,” *Journal of the American Statistical Association*, 1978, *73*, 812–815.
- Yu, M., Y. Wang, S. Xie, K. M. Tan, and W.-X. Zhou**, “Estimation and Inference for Nonparametric Expected Shortfall Regression Over RKHS,” *Journal of the American Statistical Association*, 2025, *120*, 2314–2325.
- Zhang, K., Y. Li, J. D. Schwartz, and M. S. O’Neill**, “What Weather Variables Are Important in Predicting Heat-Related Mortality? A New Application of Statistical Learning Methods,” *Environmental Research*, 2014, *132*, 350–359.
- Zhang, S., X. He, K. M. Tan, and W.-X. Zhou**, “High-Dimensional Expected Shortfall Regression,” *Journal of the American Statistical Association*, 2025, *120*, 1799–1810.
- Zhang, T.**, “High-Quantile Regression for Tail-Dependent Time Series,” *Biometrika*, 2021, *108* (1), 113–126.
- Zhang, Y.**, “Extremal Quantile Treatment Effects,” *The Annals of Statistics*, 2018, *46* (6B), 3707–3740.
- Zivin, J. Graff and M. Neidell**, “Environment, Health, and Human Capital,” *Journal of Economic Literature*, 2013, *51* (3), 689–730.

Appendix A: Theoretical Proofs and Methodological Details

This appendix contains Sections A.1–A.7. Specifically, in Section A.1 we prove Theorem 1, with supplementary results for that proof being established in Section A.2. Similarly, Section A.3 proves Theorem 2, while additional results required for the proof are relegated to Sections A.4–A.5. Finally, we prove Theorems 3 and 4 in Sections A.6 and A.7, respectively.

Notation

We use the following notational conventions throughout this appendix. The probability space we work on is $(\Omega, \mathcal{F}, \mathbb{P})$. The space $C_m[a, b]$ refers to the set of \mathbb{R}^m -valued functions that are continuous (componentwise) on the interval $[a, b]$, with $0 \leq a < b < \infty$. This space is equipped with the uniform topology. For more details on $C_m[a, b]$ and $D_m[a, b]$ (the latter defined in the main text), we refer the reader to PD86 and Billingsley (1999). For any random variable X , we denote its L_p -norm by $\|X\|_p := \{\mathbb{E}|X|^p\}^{1/p}$. The symbols $o_{\mathbb{P}}(1)$, $O_{\mathbb{P}}(1)$, $o(1)$, and $O(1)$ have their usual meanings, and—unless otherwise specified—are to be interpreted as $n \rightarrow \infty$. Indeed, if not specified otherwise, all convergences are to be understood with respect to $n \rightarrow \infty$. As usual, we take $a_n \asymp b_n$ to mean that $a_n = O(b_n)$ and $b_n = O(a_n)$, as $n \rightarrow \infty$. The identity matrix is denoted by \mathbf{I} , with the dimension clear from context. We denote by $K > 0$ a large positive constant that may change from line to line. We also write $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot \mid \mathbf{X}_t]$ and $\mathbb{P}_t\{\cdot\} = \mathbb{P}\{\cdot \mid \mathbf{X}_t\}$ for short. In the following, all (in)equalities involving conditional expectations only hold almost surely (a.s.), yet for notational economy we do not make this explicit. As is common, $\stackrel{d}{=}$ indicates distributional equality. The undesignated norm $\|\cdot\|$ denotes the Euclidean norm. We exploit without further mention that the Euclidean norm is submultiplicative, i.e., that $\|\mathbf{A}\mathbf{B}\| \leq \|\mathbf{A}\| \cdot \|\mathbf{B}\|$ for conformable matrices \mathbf{A} and \mathbf{B} .

A.1 Proof of Theorem 1

The proof of Theorem 1 (and also that of Theorem 2) crucially relies on the following lemma from Hoga and Schulz (2025), which we repeat here for the reader's convenience. To introduce it, we define $W_n(\mathbf{w}, s, \omega) : \mathbb{R}^m \times [a, b] \times \Omega \rightarrow \mathbb{R}$ ($n \geq 1$) and $W_\infty(\mathbf{w}, s, \omega) : \mathbb{R}^m \times [a, b] \times \Omega \rightarrow \mathbb{R}$ to be random elements. As usual, we suppress their dependence on $\omega \in \Omega$ and abbreviate them as $W_n(\mathbf{w}, s)$ and $W_\infty(\mathbf{w}, s)$, respectively.

Lemma 1 (Hoga and Schulz, 2025, Lemma 1). *Suppose (i) the real-valued random elements $W_n(\mathbf{w}, s)$ ($n \geq 1$) and $W_\infty(\mathbf{w}, s)$ are convex in \mathbf{w} for each s , $W_n(\mathbf{w}, s)$ ($n \geq 1$) is càdlàg in s for each \mathbf{w} and $W_\infty(\mathbf{w}, s)$ is continuous in s for each \mathbf{w} ; (ii) $\mathbf{w}_\infty(s)$ is the unique minimum point of $W_\infty(\cdot, s)$ for each $s \in [a, b]$, and $\mathbf{w}_n(s)$ is the minimum point of $W_n(\cdot, s)$ for each $s \in [a, b]$; (iii) $\mathbf{w}_n(\cdot) \in D_m[a, b]$ ($n \geq 1$) and $\mathbf{w}_\infty(\cdot) \in C_m[a, b]$. If, as $n \rightarrow \infty$,*

$$(W_n(\mathbf{w}_1, \cdot), \dots, W_n(\mathbf{w}_\ell, \cdot))' \xrightarrow{d} (W_\infty(\mathbf{w}_1, \cdot), \dots, W_\infty(\mathbf{w}_\ell, \cdot))' \quad \text{in } D_\ell[a, b] \quad (9)$$

for each $\ell \geq 1$, where $\{\mathbf{w}_1, \mathbf{w}_2, \dots\}$ is a dense subset of \mathbb{R}^m , then, as $n \rightarrow \infty$,

$$\mathbf{w}_n(\cdot) \xrightarrow{d} \mathbf{w}_\infty(\cdot) \quad \text{in } D_m[a, b].$$

Additionally, we require the following two propositions.

Proposition 1. *Under the assumptions of Theorem 1 it holds that, as $n \rightarrow \infty$,*

$$\frac{1}{\sqrt{n\tau(1-\tau)}} \sum_{t=1}^{\lfloor ns \rfloor} \psi(\varepsilon_{t\tau}) \mathbf{X}_t \xrightarrow{d} \mathbf{J}^{1/2} \mathbf{W}(s) \quad \text{in } D_m[0, 1],$$

where $\mathbf{W}(\cdot)$ is a m -variate standard Wiener process.

Proof: See Section A.2. □

Proposition 2. *Under the assumptions of Theorem 1 it holds for every fixed $\mathbf{w} \in \mathbb{R}^m$ that, as $n \rightarrow \infty$,*

$$\sup_{s \in [0, 1]} \left| \frac{1}{F^{\leftarrow}(\tau)} \sum_{t=1}^{\lfloor ns \rfloor} (\varepsilon_{t\tau} - a_n \mathbf{w}' \mathbf{X}_t) (\mathbf{1}_{\{a_n \mathbf{w}' \mathbf{X}_t < \varepsilon_{t\tau} < 0\}} - \mathbf{1}_{\{0 < \varepsilon_{t\tau} < a_n \mathbf{w}' \mathbf{X}_t\}}) - \frac{1}{2} \mathbf{s} \mathbf{w}' \mathbf{D} \mathbf{w} \right| = o_{\mathbb{P}}(1),$$

where $a_n = F^{\leftarrow}(\tau)(n\tau[1-\tau])^{-1/2}$.

Proof: See Section A.2. □

The outline of the proof of Theorem 1 is similar to that of Theorem 1 in Hoga and Schulz (2025). The main difference from our proofs to those of Hoga and Schulz (2025) can be found in the arguments used to verify Propositions 1–2. There, in contrast to the fixed- τ case of Hoga and Schulz (2025), we have to deal with asymptotically degenerate triangular arrays. This necessitates invoking entirely different techniques from “standard” FCLTs for non-degenerate sequences of random variables; see Lemma 4.1 and the surrounding discussion in Hill (2009). More specifically, we employ “big block, small block” arguments to prove Propositions 1–2.

Proof of Theorem 1: The estimator $\widehat{\boldsymbol{\alpha}}(s)$ can equivalently be written as

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}(s) &= \arg \min_{\boldsymbol{\alpha} \in \mathbb{R}^m} \frac{1}{F^{\leftarrow}(\tau)} \sum_{t=1}^{\lfloor ns \rfloor} [\rho(Y_t - \mathbf{X}_t' \boldsymbol{\alpha}) - \rho(\varepsilon_{t\tau})] \\ &= \arg \min_{\boldsymbol{\alpha} \in \mathbb{R}^m} \frac{1}{F^{\leftarrow}(\tau)} \sum_{t=1}^{\lfloor ns \rfloor} [\rho(Y_t - \mathbf{X}_t' \boldsymbol{\alpha}_{0\tau} - \mathbf{X}_t' (\boldsymbol{\alpha} - \boldsymbol{\alpha}_{0\tau})) - \rho(\varepsilon_{t\tau})] \\ &\stackrel{(1)}{=} \arg \min_{\boldsymbol{\alpha} \in \mathbb{R}^m} \frac{1}{F^{\leftarrow}(\tau)} \sum_{t=1}^{\lfloor ns \rfloor} [\rho(\varepsilon_{t\tau} - (\boldsymbol{\alpha} - \boldsymbol{\alpha}_{0\tau})' \mathbf{X}_t) - \rho(\varepsilon_{t\tau})].\end{aligned}$$

Define

$$W_n(\mathbf{w}, s) = \frac{1}{F^{\leftarrow}(\tau)} \sum_{t=1}^{\lfloor ns \rfloor} [\rho(\varepsilon_{t\tau} - a_n \mathbf{w}' \mathbf{X}_t) - \rho(\varepsilon_{t\tau})]. \quad (10)$$

Then, the minimizer $\mathbf{w}_n(s)$ of $W_n(\cdot, s)$ satisfies that

$$\mathbf{w}_n(s) = a_n^{-1} [\widehat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau}],$$

where a_n is defined in Proposition 2.

To derive the weak limit of $\mathbf{w}_n(\cdot)$, we invoke Lemma 1. For this, we first deduce the weak limit of the finite-dimensional distributions of $W_n(\cdot, s)$ to verify (9). Recalling that $\psi(u) = \tau - \mathbf{1}_{\{u \leq 0\}}$ is the subgradient of $\rho(u)$, we have that for $u \neq 0$,

$$\rho(u - v) - \rho(u) = -v\psi(u) + (u - v)[\mathbf{1}_{\{v < u < 0\}} - \mathbf{1}_{\{0 < u < v\}}]. \quad (11)$$

Note that

$$\mathbb{P}\{\exists t \in \mathbb{N}: \varepsilon_{t\tau} = 0\} = \mathbb{P}\left\{\bigcup_{t \in \mathbb{N}} \{\varepsilon_{t\tau} = 0\}\right\} \leq \sum_{t \in \mathbb{N}} \mathbb{P}\{\varepsilon_{t\tau} = 0\} = 0. \quad (12)$$

Here, the final equality follows because

$$\mathbb{P}_t\{\varepsilon_{t\tau} = 0\} = \mathbb{P}_t\{Y_t = \mathbf{X}'_t \boldsymbol{\alpha}_{0\tau}\} = 0$$

by the assumed continuity of $F_t(y) = \mathbb{P}\{Y_t \leq y \mid \mathbf{X}_t\}$.

Due to (12), we may use (10)–(11) to conclude that almost surely

$$\begin{aligned} W_n(\mathbf{w}, s) &= -\mathbf{w}' \frac{1}{\sqrt{n\tau(1-\tau)}} \sum_{t=1}^{\lfloor ns \rfloor} \mathbf{X}_t \psi(\varepsilon_{t\tau}) \\ &\quad + \frac{1}{F^{\leftarrow}(\tau)} \sum_{t=1}^{\lfloor ns \rfloor} (\varepsilon_{t\tau} - a_n \mathbf{w}' \mathbf{X}_t) [\mathbb{1}_{\{a_n \mathbf{w}' \mathbf{X}_t < \varepsilon_{t\tau} < 0\}} - \mathbb{1}_{\{0 < \varepsilon_{t\tau} < a_n \mathbf{w}' \mathbf{X}_t\}}]. \end{aligned}$$

Hence, we obtain from Propositions 1–2 that for each fixed $\mathbf{w} \in \mathbb{R}^m$, as $n \rightarrow \infty$,

$$\begin{aligned} W_n(\mathbf{w}, s) &\xrightarrow{d} -\mathbf{w}' \mathbf{J}^{1/2} \mathbf{W}(s) + \frac{1}{2} \mathbf{w}' s \mathbf{D} \mathbf{w} \\ &= -\mathbf{w}' \mathbf{a}(s) + \frac{1}{2} \mathbf{w}' \mathbf{B}(s) \mathbf{w} \\ &=: W_\infty(\mathbf{w}, s) \quad \text{in } D[\epsilon, 1], \end{aligned} \tag{13}$$

where $\mathbf{a}(s) := \mathbf{J}^{1/2} \mathbf{W}(s)$ and $\mathbf{B}(s) := s \mathbf{D}$. Therefore, a straightforward application of a Cramér–Wold device (e.g., Davidson, 1994, Theorem 29.16) implies

$$(W_n(\mathbf{w}_1, s), \dots, W_n(\mathbf{w}_\ell, s))' \xrightarrow{d} (W_\infty(\mathbf{w}_1, s), \dots, W_\infty(\mathbf{w}_\ell, s))' \quad \text{in } D_\ell[\epsilon, 1]$$

for any $\mathbf{w}_1, \dots, \mathbf{w}_\ell \in \mathbb{R}^m$. Hence, assumption (9) of Lemma 1 is satisfied.

Next, we verify the remaining conditions (i)–(iii) of Lemma 1. The only non-trivial part is to show that $W_\infty(\cdot, s)$ has a unique minimum. A simple exercise in multivariable calculus shows that $W_\infty(\cdot, s)$ in (13) has a minimum for $\mathbf{w}_\infty(s)$ satisfying the first-order condition

$$-\mathbf{a}(s) + \mathbf{B}(s) \mathbf{w}_\infty(s) = \mathbf{0}.$$

For all $s \in [\epsilon, 1]$, $\mathbf{B}(s) = s \mathbf{D}$ is positive definite by Assumption 4, such that $\mathbf{w}_\infty(s)$ is also the unique minimum of $W_\infty(\cdot, s)$ and it satisfies

$$\mathbf{w}_\infty(s) = \mathbf{B}^{-1}(s) \mathbf{a}(s) = \frac{1}{s} \mathbf{D}^{-1} \mathbf{J}^{1/2} \mathbf{W}(s).$$

Therefore, $\mathbf{w}_n(s) \xrightarrow{d} \mathbf{w}_\infty(s)$ in $D_m[\epsilon, 1]$ by Lemma 1, such that the conclusion follows. \square

A.2 Proofs of Propositions 1–2

Proof of Proposition 1: Our proof exploits “big block, small block” arguments. In doing so, we follow, e.g., the proof of Theorem 4 in Hoga (2017). For $s \in [0, 1]$, put $m_n(s) = \lfloor \lfloor ns \rfloor / (r_n + l_n) \rfloor$. Define the big blocks B_j and the small blocks S_j as consecutive blocks of integers of length r_n and l_n , respectively. Hence, $B_1 = \{1, \dots, r_n\}$, $S_1 = \{r_n + 1, \dots, r_n + l_n\}$, $B_2 = \{r_n + l_n + 1, \dots, 2r_n + l_n\}$, $S_2 = \{2r_n + l_n + 1, \dots, 2r_n + 2l_n\}$, etc. To cover the set $\{1, \dots, \lfloor ns \rfloor\}$, we use an equal number of big blocks B_j and small blocks S_j ($j = 1 \dots, m_n(s)$). By a slight abuse of notation, we denote the final block required for full coverage of $\{1, \dots, \lfloor ns \rfloor\}$ by $B_{m_n(s)+1}$. In particular, $|B_{m_n(s)+1}| \leq r_n + l_n - 1$ for the cardinality (since otherwise another big and small block could have been fitted in).

Decompose

$$\frac{1}{\sqrt{n\tau(1-\tau)}} \sum_{t=1}^{\lfloor ns \rfloor} \psi(\varepsilon_{t\tau}) \mathbf{X}_t = \sum_{j=1}^{m_n(s)} \mathbf{Y}_j^B + \sum_{j=1}^{m_n(s)} \mathbf{Y}_j^S + \mathbf{R}_n(s), \quad (14)$$

where

$$\begin{aligned} \mathbf{Y}_j^B &= \sum_{t \in B_j} \frac{1}{\sqrt{n\tau(1-\tau)}} \psi(\varepsilon_{t\tau}) \mathbf{X}_t, \\ \mathbf{Y}_j^S &= \sum_{t \in S_j} \frac{1}{\sqrt{n\tau(1-\tau)}} \psi(\varepsilon_{t\tau}) \mathbf{X}_t, \\ \mathbf{R}_n(s) &= \sum_{t \in B_{m_n(s)+1}} \frac{1}{\sqrt{n\tau(1-\tau)}} \psi(\varepsilon_{t\tau}) \mathbf{X}_t. \end{aligned}$$

Strictly speaking, we should write $\mathbf{Y}_j^B = \mathbf{Y}_{j,n}^B$ (and similarly $\mathbf{Y}_j^S = \mathbf{Y}_{j,n}^S$) to emphasize the fact that the \mathbf{Y}_j^B are in fact triangular arrays, depending on n . However, for notational economy, we suppress the dependence on n .

We now consider each of the three terms on the right-hand side of (14) individually. First, we show that the remainder term $\mathbf{R}_n(s)$ vanishes. Since $|B_{m_n(s)+1}| \leq r_n + l_n - 1$ for the cardinality, we obtain from boundedness of \mathbf{X}_t (see Assumption 5) that a.s.

$$0 \leq \sup_{s \in [0,1]} \|\mathbf{R}_n(s)\| \leq K \frac{r_n + l_n + 1}{\sqrt{n\tau(1-\tau)}} \leq K \frac{r_n}{\sqrt{k}} = o(1),$$

due to $r_n = o(k^{1/4})$ from Assumption 2.

Second, we consider the partial sums $\sum_{j=1}^{m_n(s)} \mathbf{Y}_j^B$ of the big blocks. By the same arguments as in the proof of Theorem 4 in Hoga (2017), the distributional limit of $\sum_{t=1}^{m_n(\cdot)} \mathbf{Y}_j^B$ is the same

as that of $\sum_{j=1}^{m_n(\cdot)} \tilde{\mathbf{Y}}_j^B$, where $\tilde{\mathbf{Y}}_j^B$ is an independent, identically distributed (i.i.d.) sequence with $\tilde{\mathbf{Y}}_1^B \stackrel{d}{=} \mathbf{Y}_1^B$. Now, we apply the functional central limit theorem (FCLT) for martingale difference arrays (MDAs) of [Jacod and Shiryaev \(1987, Theorem 3.33 in Chapter VIII\)](#) to derive the weak limit of $\sum_{j=1}^{m_n(s)} \tilde{\mathbf{Y}}_j^B$. To do so, first note that as a row-wise i.i.d. triangular array, $\{\tilde{\mathbf{Y}}_j^B\}_j$ is in particular a MDA. Next, we check that the conditional variances converge appropriately. Indeed, as $n \rightarrow \infty$,

$$\begin{aligned} \sum_{j=1}^{m_n(s)} \mathbb{E}_{j-1} [\tilde{\mathbf{Y}}_j^B \tilde{\mathbf{Y}}_j^{B'}] &= m_n(s) \mathbb{E} [\tilde{\mathbf{Y}}_1^B \tilde{\mathbf{Y}}_1^{B'}] \\ &= m_n(s) \frac{1}{n\tau(1-\tau)} \mathbb{E} \left[\sum_{t=1}^{r_n} \psi(\varepsilon_{t\tau}) \mathbf{X}_t \sum_{t=1}^{r_n} \psi(\varepsilon_{t\tau}) \mathbf{X}_t' \right] \\ &\longrightarrow s\mathbf{J} \end{aligned}$$

by [Assumption 3](#).

Moreover, from boundedness of \mathbf{X}_t (see [Assumption 5](#)),

$$\|\tilde{\mathbf{Y}}_j^B\| \stackrel{\text{a.s.}}{\leq} K \frac{r_n}{\sqrt{n\tau(1-\tau)}} = o(1).$$

Thus, the conditional Lindeberg condition

$$\sum_{j=1}^{m_n(s)} \mathbb{E}_{j-1} \left[\|\tilde{\mathbf{Y}}_j^B\|^2 \mathbf{1}_{\{\|\tilde{\mathbf{Y}}_j^B\|^2 > \delta^2\}} \right] \xrightarrow{\mathbb{P}} 0$$

is immediate for any $\delta > 0$.

Therefore, [Theorem 3.33 of Jacod and Shiryaev \(1987, Chapter VIII\)](#) implies that, as $n \rightarrow \infty$,

$$\sum_{j=1}^{m_n(s)} \tilde{\mathbf{Y}}_j^B \xrightarrow{d} \mathbf{Y}(s) \quad \text{in } D_m[0, 1],$$

where $\mathbf{Y}(s)$ is a continuous Gaussian martingale with quadratic variation $\langle \mathbf{Y} \rangle_s = s\mathbf{J}$. However, by [Levy's characterization of Wiener processes \(e.g., Jacod and Shiryaev, 1987, Theorem 4.4 in Chapter II\)](#), $\mathbf{Y}(\cdot)$ is a Wiener process with variance-covariance matrix \mathbf{J} . By the above arguments, this then shows that, as $n \rightarrow \infty$,

$$\sum_{j=1}^{m_n(s)} \mathbf{Y}_j^B \xrightarrow{d} \mathbf{J}^{1/2} \mathbf{W}(s) \quad \text{in } D_m[0, 1]$$

for a m -variate standard Wiener process $\mathbf{W}(\cdot)$.

Third, we show negligibility of the small block part $\sum_{j=1}^{m_n(s)} \mathbf{Y}_j^S$ in (14). As for the big blocks, the asymptotic limit of $\sum_{j=1}^{m_n(\cdot)} \mathbf{Y}_j^S$ (if it exists) is the same as that of $\sum_{j=1}^{m_n(\cdot)} \tilde{\mathbf{Y}}_j^S$, where $\{\tilde{\mathbf{Y}}_j^S\}_j$ are i.i.d. with $\tilde{\mathbf{Y}}_1^S \stackrel{d}{=} \mathbf{Y}_1^S$. Now, by Kolmogorov's inequality (e.g., Davidson, 1994, Theorem 15.14) it holds for any generic element \tilde{Y}_j^S of $\tilde{\mathbf{Y}}_j^S$ that

$$\begin{aligned} \mathbb{P}\left\{\sup_{s \in [0,1]} \left| \sum_{j=1}^{m_n(s)} \tilde{Y}_j^S \right| > \varepsilon\right\} &= \mathbb{P}\left\{\max_{1 \leq \ell \leq m_n(1)} \left| \sum_{j=1}^{\ell} \tilde{Y}_j^S \right| > \varepsilon\right\} \\ &\leq \varepsilon^{-2} \sum_{j=1}^{m_n(1)} \text{Var}(\tilde{Y}_j^S) \\ &= \frac{m_n(1)}{\varepsilon^2} \mathbb{E}[\tilde{Y}_j^S \tilde{Y}_j^{S'}] \\ &= o(1) \end{aligned}$$

by Assumption 3. Hence, $\sup_{s \in [0,1]} \left\| \sum_{j=1}^{m_n(s)} \tilde{\mathbf{Y}}_j^S \right\| = o_{\mathbb{P}}(1)$, as $n \rightarrow \infty$, such that (by the above) $\sup_{s \in [0,1]} \left\| \sum_{j=1}^{m_n(s)} \mathbf{Y}_j^S \right\| = o_{\mathbb{P}}(1)$.

Overall, the conclusion follows. \square

The proof of Proposition 2 requires two preliminary lemmas. For the first, recall that $F_t(y) = \mathbb{P}\{Y_t \leq y \mid \mathbf{X}_t\}$.

Lemma 2. *Under the assumptions of Theorem 1 it holds that, as $n \rightarrow \infty$,*

$$\frac{1 - F_t(F_t^{\leftarrow}(\tau) + x)}{1 - \tau} - 1 = -\frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} \cdot \frac{x}{F_t^{\leftarrow}(\tau)} [1 + o(1)] + O\left(A_1(1/[1 - \tau])\right)$$

uniformly in $t \in \mathbb{N}$ and uniformly in x on compact subsets of the form $[a, b]$, where $a < b$ with $a/F_t^{\leftarrow}(\tau) = o(1)$ and $b/F_t^{\leftarrow}(\tau) = o(1)$.

Proof: We follow similar steps as in the proof of Proposition 1 in Hou et al. (2024). Exploiting Assumption 6 (in the second step) and Lemma 2 of Hou et al. (2024) (in the third step), we deduce that

$$\begin{aligned} &\frac{1 - F_t(F_t^{\leftarrow}(\tau) + x)}{1 - \tau} - 1 \\ &= \frac{1 - F_t(F_t^{\leftarrow}(\tau) + x)}{1 - F(F_t^{\leftarrow}(\tau) + x)} \cdot \frac{1 - F(F_t^{\leftarrow}(\tau) + x)}{1 - F(F_t^{\leftarrow}(\tau))} - 1 \end{aligned}$$

$$\begin{aligned}
&= \left[h(\mathbf{X}_t) + O\left(A\left(\frac{1}{1-\tau}\right)\right) \right] \left(\frac{F_t^{\leftarrow}(\tau) + x}{F^{\leftarrow}(\tau)} \right)^{-1/\gamma} \left[1 + O\left(A_1\left(\frac{1}{1-\tau}\right)\right) \right] - 1 \\
&= \left[h(\mathbf{X}_t) + O\left(A\left(\frac{1}{1-\tau}\right)\right) \right] \left[h^\gamma(\mathbf{X}_t) \left\{ 1 + O\left(A\left(\frac{1}{1-\tau}\right)\right) \right\} + \frac{x}{F^{\leftarrow}(\tau)} \right]^{-1/\gamma} \times \\
&\quad \times \left[1 + O\left(A_1\left(\frac{1}{1-\tau}\right)\right) \right] - 1 \quad (15) \\
&= \left[h(\mathbf{X}_t) + O\left(A\left(\frac{1}{1-\tau}\right)\right) \right] \left\{ h^{-1}(\mathbf{X}_t) \left[1 + h^{-\gamma}(\mathbf{X}_t) \frac{x}{F^{\leftarrow}(\tau)} + O\left(A\left(\frac{1}{1-\tau}\right)\right) \right]^{-1/\gamma} \times \right. \\
&\quad \left. \times \left[1 + O\left(A_1\left(\frac{1}{1-\tau}\right)\right) \right] \right\} - 1 \\
&= \left[1 + O\left(A\left(\frac{1}{1-\tau}\right)\right) \right] \left[1 + h^{-\gamma}(\mathbf{X}_t) \frac{x}{F^{\leftarrow}(\tau)} + O\left(A\left(\frac{1}{1-\tau}\right)\right) \right]^{-1/\gamma} \times \\
&\quad \times \left[1 + O\left(A_1\left(\frac{1}{1-\tau}\right)\right) \right] - 1 \\
&= -\frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} \frac{x}{F^{\leftarrow}(\tau)} [1 + o(1)] + O\left(A_1\left(\frac{1}{1-\tau}\right)\right),
\end{aligned}$$

where the final step is due to $x/F^{\leftarrow}(\tau) = o(1)$, a Taylor expansion and Assumption 6. \square

Lemma 3. *Under the assumptions of Theorem 1 it holds for any $r \geq 1$ and $\mathbf{w} \in \mathbb{R}^m$ that, as $n \rightarrow \infty$,*

$$\begin{aligned}
\mathbb{E}_t \left[(\varepsilon_{t\tau} - a_n \mathbf{w}' \mathbf{X}_t)^r \mathbb{1}_{\{a_n \mathbf{w}' \mathbf{X}_t < \varepsilon_{t\tau} < 0\}} \right] &= \mathbb{1}_{\{a_n \mathbf{w}' \mathbf{X}_t < 0\}} \frac{1}{r+1} \frac{1-\tau}{F^{\leftarrow}(\tau)} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} [1 + o(1)] (-a_n \mathbf{w}' \mathbf{X}_t)^{r+1} \\
&\quad + \mathbb{1}_{\{a_n \mathbf{w}' \mathbf{X}_t < 0\}} r(1-\tau) (-a_n \mathbf{w}' \mathbf{X}_t)^r o(k^{-1/2}), \\
\mathbb{E}_t \left[(a_n \mathbf{w}' \mathbf{X}_t - \varepsilon_{t\tau})^r \mathbb{1}_{\{0 < \varepsilon_{t\tau} < a_n \mathbf{w}' \mathbf{X}_t\}} \right] &= \mathbb{1}_{\{0 < a_n \mathbf{w}' \mathbf{X}_t\}} \frac{1}{r+1} \frac{1-\tau}{F^{\leftarrow}(\tau)} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} [1 + o(1)] (a_n \mathbf{w}' \mathbf{X}_t)^{r+1} \\
&\quad + \mathbb{1}_{\{0 < a_n \mathbf{w}' \mathbf{X}_t\}} r(1-\tau) (a_n \mathbf{w}' \mathbf{X}_t)^r o(k^{-1/2})
\end{aligned}$$

uniformly in $t \in \mathbb{N}$.

Proof: Define $F_{\varepsilon_{t\tau} | \mathbf{X}_t}(\cdot)$ to be the cumulative distribution function of $\varepsilon_{t\tau} | \mathbf{X}_t$. In the following, we often exploit that

$$\begin{aligned}
F_{\varepsilon_{t\tau} | \mathbf{X}_t}(x) &= \mathbb{P}\{\varepsilon_{t\tau} \leq x | \mathbf{X}_t\} \\
&= \mathbb{P}\{Y_t \leq \mathbf{X}_t' \boldsymbol{\alpha}_{0\tau} + x | \mathbf{X}_t\} \\
&= F_t(\mathbf{X}_t' \boldsymbol{\alpha}_{0\tau} + x)
\end{aligned}$$

$$= F_t(F_t^{\leftarrow}(\tau) + x). \quad (16)$$

For the second conditional expectation in the lemma, we obtain via Theorem 9.14 in [Davidson \(1994\)](#) that

$$\begin{aligned}
& \mathbb{E}_t \left[(a_n \mathbf{w}' \mathbf{X}_t - \varepsilon_{t\tau})^r \mathbb{1}_{\{0 < \varepsilon_{t\tau} < a_n \mathbf{w}' \mathbf{X}_t\}} \right] \\
&= \mathbb{1}_{\{0 < a_n \mathbf{w}' \mathbf{X}_t\}} \int_0^{(a_n \mathbf{w}' \mathbf{X}_t)^r} \mathbb{P}_t \left\{ (a_n \mathbf{w}' \mathbf{X}_t - \varepsilon_{t\tau})^r \mathbb{1}_{\{0 < \varepsilon_{t\tau} < a_n \mathbf{w}' \mathbf{X}_t\}} > y \right\} dy \\
&= \mathbb{1}_{\{0 < a_n \mathbf{w}' \mathbf{X}_t\}} \int_0^{(a_n \mathbf{w}' \mathbf{X}_t)^r} \mathbb{P}_t \left\{ 0 < \varepsilon_{t\tau} < a_n \mathbf{w}' \mathbf{X}_t - y^{1/r} \right\} dy \\
&= \mathbb{1}_{\{0 < a_n \mathbf{w}' \mathbf{X}_t\}} \int_0^{a_n \mathbf{w}' \mathbf{X}_t} \mathbb{P}_t \{ 0 < \varepsilon_{t\tau} < x \} r (a_n \mathbf{w}' \mathbf{X}_t - x)^{r-1} dx \\
&= r \mathbb{1}_{\{0 < a_n \mathbf{w}' \mathbf{X}_t\}} \int_0^{a_n \mathbf{w}' \mathbf{X}_t} \left[F_t(F_t^{\leftarrow}(\tau) + x) - \tau \right] (a_n \mathbf{w}' \mathbf{X}_t - x)^{r-1} dx \\
&= r \mathbb{1}_{\{0 < a_n \mathbf{w}' \mathbf{X}_t\}} \int_0^{a_n \mathbf{w}' \mathbf{X}_t} (a_n \mathbf{w}' \mathbf{X}_t - x)^{r-1} (\tau - 1) \left[\frac{1 - F_t(F_t^{\leftarrow}(\tau) + x)}{1 - \tau} - 1 \right] dx \\
&= r \mathbb{1}_{\{0 < a_n \mathbf{w}' \mathbf{X}_t\}} \frac{1 - \tau}{F_t^{\leftarrow}(\tau)} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} [1 + o(1)] \int_0^{a_n \mathbf{w}' \mathbf{X}_t} (a_n \mathbf{w}' \mathbf{X}_t - x)^{r-1} x dx \\
&\quad + r(1 - \tau) \mathbb{1}_{\{0 < a_n \mathbf{w}' \mathbf{X}_t\}} \int_0^{a_n \mathbf{w}' \mathbf{X}_t} (a_n \mathbf{w}' \mathbf{X}_t - x)^{r-1} O\left(A_1\left(\frac{1}{1 - \tau}\right)\right) dx \\
&= r \mathbb{1}_{\{0 < a_n \mathbf{w}' \mathbf{X}_t\}} \frac{1 - \tau}{F_t^{\leftarrow}(\tau)} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} [1 + o(1)] \frac{1}{r(r+1)} (a_n \mathbf{w}' \mathbf{X}_t)^{r+1} \\
&\quad + r(1 - \tau) \mathbb{1}_{\{0 < a_n \mathbf{w}' \mathbf{X}_t\}} (a_n \mathbf{w}' \mathbf{X}_t)^r o(k^{-1/2}) \\
&= \mathbb{1}_{\{0 < a_n \mathbf{w}' \mathbf{X}_t\}} \frac{1}{r+1} \frac{1 - \tau}{F_t^{\leftarrow}(\tau)} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} [1 + o(1)] (a_n \mathbf{w}' \mathbf{X}_t)^{r+1} \\
&\quad + \mathbb{1}_{\{0 < a_n \mathbf{w}' \mathbf{X}_t\}} r(1 - \tau) (a_n \mathbf{w}' \mathbf{X}_t)^r o(k^{-1/2}),
\end{aligned}$$

where we used the substitution $x = a_n \mathbf{w}' \mathbf{X}_t - y^{1/r}$ in the third step, and [Lemma 2](#) in the third-to-last step.

The claim for the other conditional expectation follows along similar lines, concluding the proof. \square

Proof of Proposition 2: Define

$$\begin{aligned}
\nu_t(\mathbf{w}) &:= \frac{1}{F_t^{\leftarrow}(\tau)} (\varepsilon_{t\tau} - a_n \mathbf{w}' \mathbf{X}_t) (\mathbb{1}_{\{a_n \mathbf{w}' \mathbf{X}_t < \varepsilon_{t\tau} < 0\}} - \mathbb{1}_{\{0 < \varepsilon_{t\tau} < a_n \mathbf{w}' \mathbf{X}_t\}}), \\
\bar{\nu}_t(\mathbf{w}) &:= \frac{1}{F_t^{\leftarrow}(\tau)} \mathbb{E}_t \left[(\varepsilon_{t\tau} - a_n \mathbf{w}' \mathbf{X}_t) (\mathbb{1}_{\{a_n \mathbf{w}' \mathbf{X}_t < \varepsilon_{t\tau} < 0\}} - \mathbb{1}_{\{0 < \varepsilon_{t\tau} < a_n \mathbf{w}' \mathbf{X}_t\}}) \right],
\end{aligned}$$

$$V_n(\mathbf{w}, s) := \sum_{t=1}^{\lfloor ns \rfloor} \nu_t(\mathbf{w}),$$

$$\bar{V}_n(\mathbf{w}, s) := \sum_{t=1}^{\lfloor ns \rfloor} \bar{\nu}_t(\mathbf{w}),$$

where $a_n = F^{\leftarrow}(\tau)(n\tau[1-\tau])^{-1/2}$. We establish the result by showing that

$$\sup_{s \in [0,1]} \left| \bar{V}_n(\mathbf{w}, s) - \frac{1}{2} s \mathbf{w}' \mathbf{D} \mathbf{w} \right| = o_{\mathbb{P}}(1), \quad (17)$$

$$\sup_{s \in [0,1]} \left| V_n(\mathbf{w}, s) - \bar{V}_n(\mathbf{w}, s) \right| = o_{\mathbb{P}}(1). \quad (18)$$

We first prove (17). To do so, decompose $\bar{V}_n(\mathbf{w}, s)$ via Lemma 3 as follows:

$$\begin{aligned} \bar{V}_n(\mathbf{w}, s) &= \frac{1}{F^{\leftarrow}(\tau)} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{E}_t \left[(\varepsilon_{t\tau} - a_n \mathbf{w}' \mathbf{X}_t) (\mathbb{1}_{\{a_n \mathbf{w}' \mathbf{X}_t < \varepsilon_{t\tau} < 0\}} - \mathbb{1}_{\{0 < \varepsilon_{t\tau} < a_n \mathbf{w}' \mathbf{X}_t\}}) \right] \\ &= \frac{1}{2} \frac{1-\tau}{F^{\leftarrow}(\tau)} \sum_{t=1}^{\lfloor ns \rfloor} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} \cdot \frac{(a_n \mathbf{w}' \mathbf{X}_t)^2}{F^{\leftarrow}(\tau)} [1 + o(1)] + \frac{1-\tau}{F^{\leftarrow}(\tau)} \sum_{t=1}^{\lfloor ns \rfloor} a_n \mathbf{w}' \mathbf{X}_t o(k^{-1/2}) \\ &= \frac{1}{2} \frac{1}{n\tau} \sum_{t=1}^{\lfloor ns \rfloor} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} \mathbf{w}' \mathbf{X}_t \mathbf{X}_t' \mathbf{w} [1 + o(1)] + o(k^{-1/2}) \frac{1}{n} \sum_{t=1}^{\lfloor ns \rfloor} \mathbf{w}' \mathbf{X}_t \\ &= \mathbf{w}' \left(\frac{1}{2} \frac{1}{n} \sum_{t=1}^{\lfloor ns \rfloor} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} \mathbf{X}_t \mathbf{X}_t' \right) \mathbf{w} [1 + o(1)] \frac{1}{\tau} + o(1), \end{aligned}$$

where the penultimate step follows by definition of a_n , and the last step from $|\frac{1}{n} \sum_{t=1}^{\lfloor ns \rfloor} \mathbf{w}' \mathbf{X}_t| \leq K$ for all $s \in [0, 1]$ by boundedness of the \mathbf{X}_t .

Since $h^{-\gamma}(\mathbf{X}_t)$ is a \mathbf{X}_t -measurable random variable, $\{\mathbf{w}' h^{-\gamma}(\mathbf{X}_t) \mathbf{X}_t \mathbf{X}_t' \mathbf{w}\}$ is mixing of the same size as $\{\mathbf{X}_t\}$ (Davidson, 1994, Theorem 14.1). This implies that $\{\mathbf{w}' h^{-\gamma}(\mathbf{X}_t) \mathbf{X}_t \mathbf{X}_t' \mathbf{w} - \mathbf{w}' \mathbb{E}[h^{-\gamma}(\mathbf{X}_t) \mathbf{X}_t \mathbf{X}_t'] \mathbf{w}\}$ is a (zero mean) L_1 -mixingale by arguments in Andrews (1988, Section 3).

The corollary of Hansen (1992, p. 493) and the fact that $1/\tau \rightarrow 1$ therefore give that

$$\sup_{s \in [0,1]} \left| \bar{V}_n(\mathbf{w}, s) - \frac{1}{2} \mathbf{w}' \left\{ \frac{1}{n} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{E}[\gamma^{-1} h^{-\gamma}(\mathbf{X}_t) \mathbf{X}_t \mathbf{X}_t'] \right\} \mathbf{w} \right| \xrightarrow{\mathbb{P}} 0.$$

With this,

$$\sup_{s \in [0,1]} \left| \bar{V}_n(\mathbf{w}, s) - \frac{1}{2} s \mathbf{w}' \mathbf{D} \mathbf{w} \right|$$

$$\begin{aligned}
&\leq \sup_{s \in [0,1]} \left| \bar{V}_n(\mathbf{w}, s) - \frac{1}{2} \mathbf{w}' \left\{ \frac{1}{n} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{E}[\gamma^{-1} h^{-\gamma}(\mathbf{X}_t) \mathbf{X}_t \mathbf{X}_t'] \right\} \mathbf{w} \right| \\
&\quad + \sup_{s \in [0,1]} \left| \frac{1}{2} \mathbf{w}' \left\{ \frac{1}{n} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{E}[\gamma^{-1} h^{-\gamma}(\mathbf{X}_t) \mathbf{X}_t \mathbf{X}_t'] \right\} \mathbf{w} - \frac{1}{2} s \mathbf{w}' \mathbf{D} \mathbf{w} \right| \\
&= o_{\mathbb{P}}(1) + \sup_{s \in [0,1]} \left| \frac{1}{2} \mathbf{w}' \frac{\lfloor ns \rfloor}{n} \left\{ \frac{1}{\lfloor ns \rfloor} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{E}[\gamma^{-1} h^{-\gamma}(\mathbf{X}_t) \mathbf{X}_t \mathbf{X}_t'] \right\} \mathbf{w} - \frac{1}{2} s \mathbf{w}' \mathbf{D} \mathbf{w} \right| \\
&= o_{\mathbb{P}}(1) + o(1),
\end{aligned}$$

where we used Assumption 4 in the last step. This proves (17).

It remains to show (18). To this end, write

$$\begin{aligned}
V_n(\mathbf{w}, s) - \bar{V}_n(\mathbf{w}, s) &= \sum_{t=1}^{\lfloor ns \rfloor} [\nu_t(\mathbf{w}) - \bar{\nu}_t(\mathbf{w})] \\
&= \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{1t}(\mathbf{w}) - \bar{\nu}_{1t}(\mathbf{w})] + \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(\mathbf{w}) - \bar{\nu}_{2t}(\mathbf{w})],
\end{aligned}$$

where

$$\begin{aligned}
\nu_{1t}(\mathbf{w}) &= \frac{1}{F^{\leftarrow}(\tau)} (\varepsilon_{t\tau} - a_n \mathbf{w}' \mathbf{X}_t) \mathbb{1}_{\{a_n \mathbf{w}' \mathbf{X}_t < \varepsilon_{t\tau} < 0\}}, & \bar{\nu}_{1t}(\mathbf{w}) &= \mathbb{E}_t[\nu_{1t}(\mathbf{w})], \\
\nu_{2t}(\mathbf{w}) &= \frac{1}{F^{\leftarrow}(\tau)} (a_n \mathbf{w}' \mathbf{X}_t - \varepsilon_{t\tau}) \mathbb{1}_{\{0 < \varepsilon_{t\tau} < a_n \mathbf{w}' \mathbf{X}_t\}}, & \bar{\nu}_{2t}(\mathbf{w}) &= \mathbb{E}_t[\nu_{2t}(\mathbf{w})].
\end{aligned}$$

Hence, it suffices to prove that $\sup_{s \in [0,1]} |\sum_{t=1}^{\lfloor ns \rfloor} [\nu_{it}(\mathbf{w}) - \bar{\nu}_{it}(\mathbf{w})]| = o_{\mathbb{P}}(1)$ for $i = 1, 2$. We do so only for $i = 2$, as the proof for $i = 1$ is very similar.

Now, fix some $r \geq 1$ and write

$$\begin{aligned}
\mathbb{E}[\nu_{2t}^r(\mathbf{w})] &= [F^{\leftarrow}(\tau)]^{-r} \mathbb{E}[(a_n \mathbf{w}' \mathbf{X}_t - \varepsilon_{t\tau})^r \mathbb{1}_{\{0 < \varepsilon_{t\tau} < a_n \mathbf{w}' \mathbf{X}_t\}}] \\
&= [F^{\leftarrow}(\tau)]^{-r} \mathbb{E}\left[\mathbb{E}_t\{(a_n \mathbf{w}' \mathbf{X}_t - \varepsilon_{t\tau})^r \mathbb{1}_{\{0 < \varepsilon_{t\tau} < a_n \mathbf{w}' \mathbf{X}_t\}}\}\right] \\
&= [F^{\leftarrow}(\tau)]^{-r} \mathbb{E}\left[\mathbb{1}_{\{0 < a_n \mathbf{w}' \mathbf{X}_t\}} \frac{1}{r+1} \frac{1-\tau}{F^{\leftarrow}(\tau)} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} \{1 + o(1)\} (a_n \mathbf{w}' \mathbf{X}_t)^{r+1} \right. \\
&\quad \left. + \mathbb{1}_{\{0 < a_n \mathbf{w}' \mathbf{X}_t\}} r(1-\tau) (a_n \mathbf{w}' \mathbf{X}_t)^r o(k^{-1/2})\right] \\
&\leq K \frac{1-\tau}{[F^{\leftarrow}(\tau)]^{r+1}} a_n^{r+1} + K \frac{1-\tau}{[F^{\leftarrow}(\tau)]^r} a_n^r o(k^{-1/2}) \\
&\leq K \frac{1-\tau}{[n(1-\tau)]^{(r+1)/2}},
\end{aligned}$$

where the third step uses Lemma 3. By similar arguments,

$$\begin{aligned}
\mathbb{E}[\bar{\nu}_{2t}^r(\mathbf{w})] &= [F^{\leftarrow}(\tau)]^{-r} \mathbb{E} \left[\mathbb{E}_t^r \left\{ (a_n \mathbf{w}' \mathbf{X}_t - \varepsilon_{t\tau}) \mathbb{1}_{\{0 < \varepsilon_{t\tau} < a_n \mathbf{w}' \mathbf{X}_t\}} \right\} \right] \\
&= [F^{\leftarrow}(\tau)]^{-r} \mathbb{E} \left[\left\{ \mathbb{1}_{\{0 < a_n \mathbf{w}' \mathbf{X}_t\}} \frac{1 - \tau}{F^{\leftarrow}(\tau)} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} \{1 + o(1)\} \frac{1}{2} (a_n \mathbf{w}' \mathbf{X}_t)^2 \right. \right. \\
&\quad \left. \left. + \mathbb{1}_{\{0 < a_n \mathbf{w}' \mathbf{X}_t\}} r(1 - \tau) (a_n \mathbf{w}' \mathbf{X}_t)^r o(k^{-1/2}) \right\}^r \right] \\
&\leq K [F^{\leftarrow}(\tau)]^{-r} \left[\frac{1 - \tau}{F^{\leftarrow}(\tau)} a_n^2 + \frac{1 - \tau}{k} a_n o(1) \right]^r \\
&\leq K [F^{\leftarrow}(\tau)]^{-r} \left[\frac{F^{\leftarrow}(\tau)}{n} + \frac{F^{\leftarrow}(\tau)}{n} \frac{1}{\sqrt{n(1 - \tau)}} o(1) \right]^r \\
&\leq K n^{-r}.
\end{aligned}$$

Combining the previous two displays gives that

$$\begin{aligned}
\mathbb{E}[|\nu_{2t}(\mathbf{w}) - \bar{\nu}_{2t}(\mathbf{w})|^r] &\leq 2^{r-1} \left\{ \mathbb{E}[\nu_{2t}^r(\mathbf{w})] + \mathbb{E}[\bar{\nu}_{2t}^r(\mathbf{w})] \right\} \\
&\leq K \frac{1 - \tau}{[n(1 - \tau)]^{(r+1)/2}} + K n^{-r} \\
&\leq K \frac{1 - \tau}{[n(1 - \tau)]^{(r+1)/2}}, \tag{19}
\end{aligned}$$

where we used the c_r -inequality (e.g., Davidson, 1994, Theorem 9.28) in the first step, and the fact that

$$n^{-r} = o\left(\frac{1 - \tau}{[n(1 - \tau)]^{(r+1)/2}}\right),$$

as is easy to verify. Now, (19) implies that

$$\|\nu_{2t}(\mathbf{w}) - \bar{\nu}_{2t}(\mathbf{w})\|_r \leq K \frac{(1 - \tau)^{1/r}}{[n(1 - \tau)]^{(1+1/r)/2}} = K \frac{(1 - \tau)^{\frac{1}{2r} - \frac{1}{2}}}{n^{\frac{1}{2} + \frac{1}{2r}}}. \tag{20}$$

With these results at hand, we proceed similarly as in the proof of Proposition 1 and decompose

$$\sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(\mathbf{w}) - \bar{\nu}_{2t}(\mathbf{w})] = \sum_{j=1}^{m_n(s)} \nu_{2j}^B + \sum_{j=1}^{m_n(s)} \nu_{2j}^S + R_n^{\nu^2}(s),$$

where

$$\nu_{2j}^B = \sum_{t \in B_j} [\nu_{2t}(\mathbf{w}) - \bar{\nu}_{2t}(\mathbf{w})],$$

$$\begin{aligned}\nu_{2j}^S &= \sum_{t \in S_j} [\nu_{2t}(\mathbf{w}) - \bar{\nu}_{2t}(\mathbf{w})], \\ R_n^{\nu_2}(s) &= \sum_{t \in B_{m_n(s)+1}} [\nu_{2t}(\mathbf{w}) - \bar{\nu}_{2t}(\mathbf{w})].\end{aligned}$$

Observe that by boundedness of \mathbf{X}_t ,

$$\nu_{2t}(\mathbf{w}) \leq \frac{1}{F^{\leftarrow}(\tau)} a_n \mathbf{w}' \mathbf{X}_t \leq \frac{K}{\sqrt{n(1-\tau)}}.$$

Recalling that $|B_{m_n(s)+1}| \leq r_n + l_n - 1$, it therefore follows that

$$0 \leq \sup_{s \in [0,1]} |R_n^{\nu_2}(s)| \leq K \frac{r_n}{\sqrt{n(1-\tau)}} = K \frac{r_n}{\sqrt{k}} = o(1).$$

By the same arguments as in the proof of Proposition 1, $\sum_{j=1}^{m_n(\cdot)} \nu_{2j}^B$ has the same (functional) asymptotic limit as $\sum_{j=1}^{m_n(\cdot)} \tilde{\nu}_{2j}^B$, where $\{\tilde{\nu}_{2j}^B\}_j$ is i.i.d. with $\tilde{\nu}_{2j}^B \stackrel{d}{=} \nu_{2j}^B$. By Kolmogorov's inequality,

$$\begin{aligned}\mathbb{P}\left\{ \sup_{s \in [0,1]} \left| \sum_{j=1}^{m_n(s)} \tilde{\nu}_{2j}^B \right| > \varepsilon \right\} &= \mathbb{P}\left\{ \max_{1 \leq \ell \leq m_n(1)} \left| \sum_{j=1}^{\ell} \tilde{\nu}_{2j}^B \right| > \varepsilon \right\} \\ &\leq \varepsilon^{-2} \sum_{j=1}^{m_n(1)} \text{Var}(\tilde{\nu}_{2j}^B) \\ &= \varepsilon^{-2} m_n(1) \text{Var}(\tilde{\nu}_{21}^B).\end{aligned}\tag{21}$$

Now, we get from (19) (for $r = 2$) that

$$\begin{aligned}\text{Var}(\tilde{\nu}_{21}^B) &= \text{Var}\left(\sum_{t=1}^{r_n} [\nu_{2t}(\mathbf{w}) - \bar{\nu}_{2t}(\mathbf{w})] \right) \\ &\leq r_n^2 \text{Var}(\nu_{2t}(\mathbf{w}) - \bar{\nu}_{2t}(\mathbf{w})) \\ &\leq K r_n^2 \frac{1-\tau}{[n(1-\tau)]^{3/2}}.\end{aligned}$$

Combining this with (21), we obtain that

$$\mathbb{P}\left\{ \sup_{s \in [0,1]} \left| \sum_{j=1}^{m_n(s)} \tilde{\nu}_{2j}^B \right| > \varepsilon \right\} \leq K \frac{r_n}{r_n} r_n^2 \frac{1-\tau}{[n(1-\tau)]^{3/2}} = K \frac{r_n}{\sqrt{n(1-\tau)}} = K \frac{r_n}{\sqrt{k}} = o(1).$$

Therefore, we get that $\sup_{s \in [0,1]} |\sum_{j=1}^{m_n(s)} \tilde{\nu}_{2j}^B| = o_{\mathbb{P}}(1)$ and, hence, also $\sup_{s \in [0,1]} |\sum_{j=1}^{m_n(s)} \nu_{2j}^B| = o_{\mathbb{P}}(1)$.

By similar arguments, we may deduce that the small blocks vanish as well, i.e.,

$$\sup_{s \in [0,1]} \left| \sum_{j=1}^{m_n(s)} \nu_{2j}^S \right| = o_{\mathbb{P}}(1).$$

Together, we obtain that, as $n \rightarrow \infty$,

$$\sup_{s \in [0,1]} \left| \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(\mathbf{w}) - \bar{\nu}_{2t}(\mathbf{w})] \right| = o_{\mathbb{P}}(1).$$

Similar arguments show that

$$\sup_{s \in [0,1]} \left| \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{1t}(\mathbf{w}) - \bar{\nu}_{1t}(\mathbf{w})] \right| = o_{\mathbb{P}}(1).$$

Combining these two convergences yields the desired result in (18). This finishes the proof. \square

A.3 Proof of Theorem 2

The proof of Theorem 2 requires the four preliminary Propositions 3–6. Propositions 3 and 6 play a role similar to that of Proposition 1 in the proof of Theorem 1. Similarly, Propositions 4 and 5 are the equivalents of Proposition 2 in the proof of Theorem 1.

Proposition 3. *Under the assumptions of Theorem 2 it holds that, as $n \rightarrow \infty$,*

$$\frac{1}{\sqrt{n\tau(1-\tau)}} \sum_{t=1}^{\lfloor ns \rfloor} \psi_*(\varepsilon_{t\tau}, \xi_{t\tau}) \mathbf{X}_t \xrightarrow{d} \mathbf{J}_*^{1/2} \mathbf{W}_*(s) \quad \text{in } D_m[0,1],$$

where $\mathbf{W}_*(\cdot)$ is a m -variate standard Wiener process.

Proof: See Section A.4. \square

Proposition 4. *Under the assumptions of Theorem 2 it holds that, as $n \rightarrow \infty$,*

$$\sup_{s \in [0,1]} \left\| \frac{1}{n(1-\tau)} \sum_{t=1}^{\lfloor ns \rfloor} \mathbf{1}_{\{\varepsilon_{t\tau} > 0\}} \mathbf{X}_t \mathbf{X}_t' - s \mathbf{D}_* \right\| = o_{\mathbb{P}}(1).$$

Proof: See Section A.4. \square

Proposition 5. *Under the assumptions of Theorem 2 it holds that, as $n \rightarrow \infty$,*

$$\sup_{s \in [0,1]} \left\| \frac{1}{n(1-\tau)} \sum_{t=1}^{\lfloor ns \rfloor} [\mathbf{1}_{\{\mathbf{X}_t'(\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau}) < \varepsilon_{t\tau} \leq 0\}} - \mathbf{1}_{\{0 < \varepsilon_{t\tau} \leq \mathbf{X}_t'(\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau})\}}] \mathbf{X}_t \mathbf{X}_t' \right\| = o_{\mathbb{P}}(1).$$

Proof: See Section A.4. □

Proposition 6. *Under the assumptions of Theorem 2 it holds that, as $n \rightarrow \infty$,*

$$\frac{1}{F^{\leftarrow}(\tau)} \frac{1}{\sqrt{n(1-\tau)}} \sum_{t=1}^{\lfloor ns \rfloor} \left[\mathbb{1}_{\{\mathbf{X}'_t(\widehat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau}) < \varepsilon_{t\tau} \leq 0\}} - \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq \mathbf{X}'_t(\widehat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau})\}} \right] \xi_{t\tau} \mathbf{X}_t$$

$$\xrightarrow{d} \frac{1}{1-\gamma} \mathbf{D}_* \boldsymbol{\Omega}^{1/2} \mathbf{W}(s) \quad \text{in } D_m[\varepsilon, 1],$$

where $\mathbf{W}(\cdot)$ is a m -variate standard Wiener process, and $\boldsymbol{\Omega}$ is defined in Theorem 1.

Proof: See Section A.4. □

We are now in a position to prove Theorem 2. The proof of Theorem 2 is similar to that of Theorem 2 in Hoga and Schulz (2025). Once more, it is the proofs of Propositions 3–6 that require significant changes relative to the fixed- τ case considered by Hoga and Schulz (2025). This is because Propositions 3–6 deal with asymptotically degenerate triangular arrays.

Proof of Theorem 2: The first part of the proof is similar to that of Theorem 1. To highlight the similarities, we often overload notation by redefining quantities that already appeared in the proof of Theorem 1, such as $W_n(\mathbf{w}, s)$ and $\mathbf{w}_n(s)$. Note that the estimator $\widehat{\boldsymbol{\beta}}(s)$ can equivalently be written as

$$\begin{aligned} \widehat{\boldsymbol{\beta}}(s) &= \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^m} \frac{1}{[F^{\leftarrow}(\tau)]^2} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{1}_{\{Y_t > \mathbf{X}'_t \widehat{\boldsymbol{\alpha}}(s)\}} \left\{ [Y_t - \mathbf{X}'_t \boldsymbol{\beta}]^2 - \xi_{t\tau}^2 \right\} \\ &= \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^m} \frac{1}{[F^{\leftarrow}(\tau)]^2} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{1}_{\{Y_t > \mathbf{X}'_t \widehat{\boldsymbol{\alpha}}(s)\}} \left\{ [Y_t - \mathbf{X}'_t \boldsymbol{\beta}_{0\tau} - \mathbf{X}'_t (\boldsymbol{\beta} - \boldsymbol{\beta}_{0\tau})]^2 - \xi_{t\tau}^2 \right\} \\ &= \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^m} \frac{1}{[F^{\leftarrow}(\tau)]^2} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{1}_{\{Y_t > \mathbf{X}'_t \widehat{\boldsymbol{\alpha}}(s)\}} \left\{ [\xi_{t\tau} - (\boldsymbol{\beta} - \boldsymbol{\beta}_{0\tau})' \mathbf{X}_t]^2 - \xi_{t\tau}^2 \right\}, \end{aligned}$$

where we used (2) in the final step. Therefore, letting

$$W_n(\mathbf{w}, s) = \frac{1}{[F^{\leftarrow}(\tau)]^2} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{1}_{\{Y_t > \mathbf{X}'_t \widehat{\boldsymbol{\alpha}}(s)\}} \left\{ [\xi_{t\tau} - a_n \mathbf{w}' \mathbf{X}_t]^2 - \xi_{t\tau}^2 \right\},$$

the minimizer $\mathbf{w}_n(s)$ of $W_n(\cdot, s)$ satisfies that

$$\mathbf{w}_n(s) = a_n^{-1} [\widehat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}_{0\tau}].$$

We invoke Lemma 1 to derive the weak limit of $\mathbf{w}_n(\cdot)$. To do so, we again first prove weak convergence of the finite-dimensional distributions of $W_n(\cdot, s)$ to verify (9). To that end, we simplify $W_n(\mathbf{w}, s)$ by exploiting that

$$\begin{aligned}\mathbb{1}_{\{Y_t > \mathbf{X}'_t \hat{\boldsymbol{\alpha}}(s)\}} &= \mathbb{1}_{\{Y_t > \mathbf{X}'_t \boldsymbol{\alpha}_{0\tau} + \mathbf{X}'_t (\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau})\}} - \mathbb{1}_{\{\varepsilon_{t\tau} > 0\}} + \mathbb{1}_{\{\varepsilon_{t\tau} > 0\}} \\ &= \mathbb{1}_{\{\varepsilon_{t\tau} > \mathbf{X}'_t (\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau})\}} - \mathbb{1}_{\{\varepsilon_{t\tau} > 0\}} + \mathbb{1}_{\{\varepsilon_{t\tau} > 0\}} \\ &= \mathbb{1}_{\{\mathbf{X}'_t (\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau}) < \varepsilon_{t\tau} \leq 0\}} - \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq \mathbf{X}'_t (\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau})\}} + \mathbb{1}_{\{\varepsilon_{t\tau} > 0\}}.\end{aligned}$$

Hence,

$$\begin{aligned}W_n(\mathbf{w}, s) &= \frac{1}{[F^{\leftarrow}(\tau)]^2} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{1}_{\{Y_t > \mathbf{X}'_t \hat{\boldsymbol{\alpha}}(s)\}} \left[-2\xi_{t\tau} a_n \mathbf{w}' \mathbf{X}_t + (a_n \mathbf{w}' \mathbf{X}_t)^2 \right] \\ &= \frac{1}{[F^{\leftarrow}(\tau)]^2} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{1}_{\{\varepsilon_{t\tau} > 0\}} \left[-2\xi_{t\tau} a_n \mathbf{w}' \mathbf{X}_t + (a_n \mathbf{w}' \mathbf{X}_t)^2 \right] \\ &\quad + \frac{1}{[F^{\leftarrow}(\tau)]^2} \sum_{t=1}^{\lfloor ns \rfloor} \left[\mathbb{1}_{\{\mathbf{X}'_t (\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau}) < \varepsilon_{t\tau} \leq 0\}} - \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq \mathbf{X}'_t (\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau})\}} \right] \times \\ &\quad \quad \quad \times \left[-2\xi_{t\tau} a_n \mathbf{w}' \mathbf{X}_t + (a_n \mathbf{w}' \mathbf{X}_t)^2 \right] \\ &= -2\mathbf{w}' \frac{1}{\sqrt{n\tau(1-\tau)}} \sum_{t=1}^{\lfloor ns \rfloor} \frac{1}{F^{\leftarrow}(\tau)} \mathbb{1}_{\{\varepsilon_{t\tau} > 0\}} \xi_{t\tau} \mathbf{X}_t + \frac{1}{\tau} \mathbf{w}' \left(\frac{1}{n(1-\tau)} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{1}_{\{\varepsilon_{t\tau} > 0\}} \mathbf{X}_t \mathbf{X}'_t \right) \mathbf{w} \\ &\quad - 2\mathbf{w}' \frac{1}{\sqrt{\tau} F^{\leftarrow}(\tau)} \frac{1}{\sqrt{n(1-\tau)}} \sum_{t=1}^{\lfloor ns \rfloor} \left[\mathbb{1}_{\{\mathbf{X}'_t (\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau}) < \varepsilon_{t\tau} \leq 0\}} - \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq \mathbf{X}'_t (\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau})\}} \right] \xi_{t\tau} \mathbf{X}_t \\ &\quad + \frac{1}{\tau} \mathbf{w}' \left(\frac{1}{n(1-\tau)} \sum_{t=1}^{\lfloor ns \rfloor} \left[\mathbb{1}_{\{\mathbf{X}'_t (\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau}) < \varepsilon_{t\tau} \leq 0\}} - \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq \mathbf{X}'_t (\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau})\}} \right] \mathbf{X}_t \mathbf{X}'_t \right) \mathbf{w}.\end{aligned}$$

We obtain from Propositions 3–6 that, as $n \rightarrow \infty$,

$$\begin{aligned}W_n(\mathbf{w}, s) &\xrightarrow{d} -2\mathbf{w}' \mathbf{J}_*^{1/2} \mathbf{W}_*(s) + s\mathbf{w}' \mathbf{D}_* \mathbf{w} - 2\mathbf{w}' \frac{1}{1-\gamma} \mathbf{D}_* \boldsymbol{\Omega}^{1/2} \mathbf{W}(s) \\ &= -2\mathbf{w}' \mathbf{a}(s) + \mathbf{w}' \mathbf{B}(s) \mathbf{w} \\ &=: W_\infty(\mathbf{w}, s) \quad \text{in } D[\epsilon, 1]\end{aligned}$$

for $\mathbf{a}(s) := \mathbf{J}_*^{1/2} \mathbf{W}_*(s) + \frac{1}{1-\gamma} \mathbf{D}_* \boldsymbol{\Omega}^{1/2} \mathbf{W}(s)$ and $\mathbf{B}(s) := s\mathbf{D}_*$. Similarly as in the proof of Theorem 1, we deduce from this that

$$(W_n(\mathbf{w}_1, s), \dots, W_n(\mathbf{w}_\ell, s))' \xrightarrow{d} (W_\infty(\mathbf{w}_1, s), \dots, W_\infty(\mathbf{w}_\ell, s))' \quad \text{in } D_\ell[\epsilon, 1]$$

for any $\mathbf{w}_1, \dots, \mathbf{w}_\ell \in \mathbb{R}^m$, such that (9) of Lemma 1 is met.

Once more, the only remaining condition of (i)–(iii) in Lemma 1 that is not immediately obvious, is that $W_\infty(\cdot, s)$ has a unique minimum. Because $\mathbf{B}(s)$ is positive definite by Assumption 4 for $s \in [\epsilon, 1]$, it is easy to deduce that $W_\infty(\cdot, s)$ has a unique minimum in $\mathbf{w}_\infty(s)$ satisfying the first-order condition

$$-\mathbf{a}(s) + \mathbf{B}(s)\mathbf{w}_\infty(s) = \mathbf{0},$$

such that

$$\mathbf{w}_\infty(s) = \mathbf{B}^{-1}(s)\mathbf{a}(s) = \frac{1}{s}\mathbf{D}_*^{-1}\left[\mathbf{J}_*^{1/2}\mathbf{W}_*(s) + \frac{1}{1-\gamma}\mathbf{D}_*\boldsymbol{\Omega}^{1/2}\mathbf{W}(s)\right]. \quad (22)$$

Therefore, as $n \rightarrow \infty$,

$$\mathbf{w}_n(s) \xrightarrow{d} \mathbf{w}_\infty(s) \quad \text{in } D_m[\epsilon, 1] \quad (23)$$

by Lemma 1.

In the second step of this proof, it remains to show that $s\mathbf{w}_\infty(s)$ coincides with the limiting distribution given in the statement of the theorem. To that end, we show for

$$\widetilde{\mathbf{W}}(s) := \mathbf{J}_*^{1/2}\mathbf{W}_*(s) + \frac{1}{1-\gamma}\mathbf{D}_*\boldsymbol{\Omega}^{1/2}\mathbf{W}(s)$$

that $\widetilde{\mathbf{W}}(s) \stackrel{d}{=} \boldsymbol{\Sigma}_*^{1/2}\mathbf{W}_m(s)$ for a m -variate standard Wiener process $\mathbf{W}_m(\cdot)$. To do so, it is well-known that it suffices to show for the covariance function that

$$\text{Cov}(\widetilde{\mathbf{W}}(s), \widetilde{\mathbf{W}}(t)) = \text{Cov}(\boldsymbol{\Sigma}_*^{1/2}\mathbf{W}_m(s), \boldsymbol{\Sigma}_*^{1/2}\mathbf{W}_m(t)) = \min\{s, t\}\boldsymbol{\Sigma}_*.$$

Proceeding similarly as in the proofs of Propositions 1 and 3, we obtain using Assumption 3 that

$$\frac{1}{\sqrt{n\tau(1-\tau)}} \sum_{t=1}^{\lfloor ns \rfloor} \begin{pmatrix} \psi_*(\varepsilon_{t\tau}, \xi_{t\tau})\mathbf{X}_t \\ \psi(\varepsilon_{t\tau})\mathbf{X}_t \end{pmatrix} \xrightarrow{d} \overline{\boldsymbol{\Omega}}^{1/2}\overline{\mathbf{W}}(s) \quad \text{in } D_{2m}[0, 1], \quad (24)$$

where $\overline{\mathbf{W}}(\cdot) = (\overline{\mathbf{W}}_1'(\cdot), \overline{\mathbf{W}}_2'(\cdot))'$ is a standard $(2m)$ -dimensional Wiener process, and

$$\overline{\boldsymbol{\Omega}} = \lim_{n \rightarrow \infty} \frac{1}{r_n\tau(1-\tau)} \mathbb{E} \left[\sum_{t=1}^{r_n} \begin{pmatrix} \psi_*(\varepsilon_{t\tau}, \xi_{t\tau})\mathbf{X}_t \\ \psi(\varepsilon_{t\tau})\mathbf{X}_t \end{pmatrix} \sum_{t=1}^{r_n} (\psi_*(\varepsilon_{t\tau}, \xi_{t\tau})\mathbf{X}_t', \psi(\varepsilon_{t\tau})\mathbf{X}_t') \right] = \begin{pmatrix} \mathbf{J}_* & \mathbf{C}' \\ \mathbf{C} & \mathbf{J} \end{pmatrix}$$

is positive definite by Assumption 3. Thus, there exists a Choleski decomposition of the form $\overline{\boldsymbol{\Omega}} = \overline{\boldsymbol{\Omega}}^{1/2}(\overline{\boldsymbol{\Omega}}^{1/2})'$ for a lower triangular $\overline{\boldsymbol{\Omega}}^{1/2}$ (Lütkepohl, 2007, Sec. A.9.3). This lower triangular

matrix is then necessarily of the form

$$\overline{\Omega}^{1/2} = \begin{pmatrix} \mathbf{J}_*^{1/2} & \mathbf{0} \\ \mathbf{C}_{21} & \mathbf{J}_{22} \end{pmatrix},$$

where $\mathbf{J}_*^{1/2}$ is the lower triangular matrix from the Choleski decomposition $\mathbf{J}_* = \mathbf{J}_*^{1/2}(\mathbf{J}_*^{1/2})'$ (which exists by positive definiteness of \mathbf{J}_* in Assumption 3), and the matrices \mathbf{C}_{21} and \mathbf{J}_{22} satisfy that

$$\begin{aligned} \overline{\Omega} &= \overline{\Omega}^{1/2}(\overline{\Omega}^{1/2})' \\ &= \begin{pmatrix} \mathbf{J}_*^{1/2} & \mathbf{0} \\ \mathbf{C}_{21} & \mathbf{J}_{22} \end{pmatrix} \begin{pmatrix} (\mathbf{J}_*^{1/2})' & \mathbf{C}'_{21} \\ \mathbf{0} & \mathbf{J}'_{22} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{J}_* & \mathbf{J}_*^{1/2}\mathbf{C}'_{21} \\ \mathbf{C}_{21}(\mathbf{J}_*^{1/2})' & \mathbf{C}_{21}\mathbf{C}'_{21} + \mathbf{J}_{22}\mathbf{J}'_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{J}_* & \mathbf{C}' \\ \mathbf{C} & \mathbf{J} \end{pmatrix}. \end{aligned} \quad (25)$$

From (24) and the marginal convergences in Propositions 1 and 3, we conclude that

$$\begin{pmatrix} \mathbf{J}_*^{1/2}\mathbf{W}_*(s) \\ \mathbf{J}^{1/2}\mathbf{W}(s) \end{pmatrix} \stackrel{d}{=} \overline{\Omega}^{1/2}\overline{\mathbf{W}}(s) = \begin{pmatrix} \mathbf{J}_*^{1/2} & \mathbf{0} \\ \mathbf{C}_{21} & \mathbf{J}_{22} \end{pmatrix} \begin{pmatrix} \overline{\mathbf{W}}_1(s) \\ \overline{\mathbf{W}}_2(s) \end{pmatrix}.$$

Therefore, also using that $\Omega^{1/2} = (\mathbf{D}^{-1}\mathbf{J}\mathbf{D}^{-1})^{1/2} = \mathbf{D}^{-1}\mathbf{J}^{1/2}$,

$$\begin{aligned} \widetilde{\mathbf{W}}(s) &= \mathbf{J}_*^{1/2}\mathbf{W}_*(s) + (1-\gamma)^{-1}\mathbf{D}_*\Omega^{1/2}\mathbf{J}^{-1/2}\mathbf{J}^{1/2}\mathbf{W}(s) \\ &= \mathbf{J}_*^{1/2}\mathbf{W}_*(s) + (1-\gamma)^{-1}\mathbf{D}_*\mathbf{D}^{-1}\mathbf{J}^{1/2}\mathbf{W}(s) \\ &= \left(\mathbf{I} \mid (1-\gamma)^{-1}\mathbf{D}_*\mathbf{D}^{-1} \right) \begin{pmatrix} \mathbf{J}_*^{1/2}\mathbf{W}_*(s) \\ \mathbf{J}^{1/2}\mathbf{W}(s) \end{pmatrix} \\ &\stackrel{d}{=} \left(\mathbf{I} \mid (1-\gamma)^{-1}\mathbf{D}_*\mathbf{D}^{-1} \right) \begin{pmatrix} \mathbf{J}_*^{1/2} & \mathbf{0} \\ \mathbf{C}_{21} & \mathbf{J}_{22} \end{pmatrix} \begin{pmatrix} \overline{\mathbf{W}}_1(s) \\ \overline{\mathbf{W}}_2(s) \end{pmatrix} \\ &= \left(\mathbf{J}_*^{1/2} + (1-\gamma)^{-1}\mathbf{D}_*\mathbf{D}^{-1}\mathbf{C}_{21} \mid (1-\gamma)^{-1}\mathbf{D}_*\mathbf{D}^{-1}\mathbf{J}_{22} \right) \begin{pmatrix} \overline{\mathbf{W}}_1(s) \\ \overline{\mathbf{W}}_2(s) \end{pmatrix} \\ &= \left(\mathbf{J}_*^{1/2} + (1-\gamma)^{-1}\mathbf{D}_*\mathbf{D}^{-1}\mathbf{C}(\mathbf{J}_*^{-1/2})' \mid (1-\gamma)^{-1}\mathbf{D}_*\mathbf{D}^{-1}\mathbf{J}_{22} \right) \begin{pmatrix} \overline{\mathbf{W}}_1(s) \\ \overline{\mathbf{W}}_2(s) \end{pmatrix}, \end{aligned} \quad (26)$$

where the last step exploits that $\mathbf{C}_{21} = \mathbf{C}(\mathbf{J}_*^{-1/2})'$ from (25). With this and (without loss of generality) $s < t$,

$$\begin{aligned}
\text{Cov}(\widetilde{\mathbf{W}}(s), \widetilde{\mathbf{W}}(t)) &= \text{Cov}(\widetilde{\mathbf{W}}(s), \widetilde{\mathbf{W}}(t) - \widetilde{\mathbf{W}}(s)) + \text{Cov}(\widetilde{\mathbf{W}}(s), \widetilde{\mathbf{W}}(s)) \\
&= \text{Cov}(\widetilde{\mathbf{W}}(s), \widetilde{\mathbf{W}}(s)) \\
&= s \left(\mathbf{J}_*^{1/2} + (1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1} \mathbf{C} (\mathbf{J}_*^{-1/2})' \mid (1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1} \mathbf{J}_{22} \right) \times \\
&\quad \times \begin{pmatrix} (\mathbf{J}_*^{1/2})' + [(1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1} \mathbf{C} (\mathbf{J}_*^{-1/2})']' \\ [(1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1} \mathbf{J}_{22}]' \end{pmatrix} \\
&= s \left\{ \mathbf{J}_* + (1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1} \mathbf{C} + [(1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1} \mathbf{C}]' \right. \\
&\quad \left. + (1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1} \mathbf{C} \mathbf{J}_*^{-1} \mathbf{C}' [(1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1}]' \right. \\
&\quad \left. + (1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1} \mathbf{J}_{22} \mathbf{J}'_{22} [(1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1}]' \right\} \\
&= s \left\{ \mathbf{J}_* + (1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1} \mathbf{C} + [(1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1} \mathbf{C}]' \right. \\
&\quad \left. + (1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1} \mathbf{C} \mathbf{J}_*^{-1} \mathbf{C}' [(1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1}]' \right. \\
&\quad \left. + (1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1} (\mathbf{J} - \mathbf{C} \mathbf{J}_*^{-1} \mathbf{C}') [(1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1}]' \right\} \\
&= s \left\{ \mathbf{J}_* + (1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1} \mathbf{C} + [(1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1} \mathbf{C}]' \right. \\
&\quad \left. + (1-\gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1} \mathbf{J} \mathbf{D}^{-1} (1-\gamma)^{-1} \mathbf{D}_* \right\} \\
&= s \boldsymbol{\Sigma}_*,
\end{aligned}$$

where the second step follows from (26) and the independence of the increments of $(\overline{\mathbf{W}}_1(\cdot), \overline{\mathbf{W}}_2(\cdot))'$, and the fifth step uses that $\mathbf{J}_{22} \mathbf{J}'_{22} = \mathbf{J} - \mathbf{C}_{21} \mathbf{C}'_{21} = \mathbf{J} - \mathbf{C} (\mathbf{J}_*^{-1/2})' \mathbf{J}_*^{-1/2} \mathbf{C}' = \mathbf{J} - \mathbf{C} \mathbf{J}_*^{-1} \mathbf{C}'$ from (25). In conclusion, we obtain that $\text{Cov}(\widetilde{\mathbf{W}}(s), \widetilde{\mathbf{W}}(t)) = \min\{s, t\} \boldsymbol{\Sigma}_*$, which characterizes Wiener processes (e.g., Jacod and Shiryaev, 1987, Theorem 4.4 II), such that $\widetilde{\mathbf{W}}(s) \stackrel{d}{=} \boldsymbol{\Sigma}_*^{1/2} \mathbf{W}_m(s)$ for a m -dimensional standard Wiener process $\mathbf{W}_m(\cdot)$.

Therefore, we conclude from (22) that

$$\mathbf{w}_\infty(s) \stackrel{d}{=} \frac{1}{s} \mathbf{D}_*^{-1} \boldsymbol{\Sigma}_*^{1/2} \mathbf{W}_m(s).$$

From this and (23), the desired convergence of Theorem 2 easily follows.

Finally, from [Lütkepohl \(2007, Rule \(7\) in Sec. A.8.3\)](#) we obtain that

$$\boldsymbol{\Sigma}_* = \left(\mathbf{I} \mid (1 - \gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1} \right) \overline{\boldsymbol{\Omega}} \begin{pmatrix} \mathbf{I} \\ \mathbf{D}^{-1} (1 - \gamma)^{-1} \mathbf{D}_* \end{pmatrix}$$

is positive definite, since $\left(\mathbf{I} \mid (1 - \gamma)^{-1} \mathbf{D}_* \mathbf{D}^{-1} \right)$ obviously has full rank and $\overline{\boldsymbol{\Omega}}$ is positive definite. □

A.4 Proofs of Propositions 3–6

The proof of Proposition 3 requires Lemma 6 as a preliminary result. In turn, the proof of Lemma 6 requires the following two preliminary Lemmas 4–5.

Lemma 4. *Under the assumptions of Theorem 2 it holds for any $r \in (0, 1/\gamma)$ that*

$$\lim_{y \rightarrow \infty} \frac{\mathbb{E}[(Y_t - y)^r \mid Y_t > y, \mathbf{X}_t = \mathbf{x}]}{y^r} = \frac{\Gamma(r + 1)\Gamma(\alpha - r)}{\Gamma(\alpha)},$$

where $\alpha = 1/\gamma$, and $\Gamma(\cdot)$ denotes the Gamma function.

Proof: Define $\overline{F}(\cdot) = 1 - F(\cdot)$ and $\overline{F}(\cdot \mid \mathbf{x}) = 1 - F(\cdot \mid \mathbf{x})$ to be the respective survival functions.

Recall that

$$\frac{\Gamma(r)\Gamma(\alpha - r)}{\Gamma(\alpha)} = \text{B}(r, \alpha - r) = \int_1^\infty (u - 1)^{r-1} u^{-\alpha} du$$

for Euler's beta function $\text{B}(\cdot, \cdot)$. We proceed by combining Assumption 6 with the proof of Proposition 4.5 in [Pan et al. \(2013\)](#). Following said proof, we obtain that, as $y \rightarrow \infty$,

$$\begin{aligned} & \frac{\mathbb{E}[(Y_t - y)^r \mid Y_t > y, \mathbf{X}_t = \mathbf{x}]}{y^r} - \frac{\Gamma(r + 1)\Gamma(\alpha - r)}{\Gamma(\alpha)} \\ &= r \int_1^\infty (u - 1)^{r-1} \frac{\overline{F}(yu \mid \mathbf{x})}{\overline{F}(y \mid \mathbf{x})} du - r \frac{\Gamma(r)\Gamma(\alpha - r)}{\Gamma(\alpha)} \\ &= r \int_1^\infty (u - 1)^{r-1} \left(\frac{\overline{F}(yu \mid \mathbf{x})}{\overline{F}(y \mid \mathbf{x})} - u^{-\alpha} \right) du \\ &= r \int_1^\infty (u - 1)^{r-1} \left(\frac{\overline{F}(yu \mid \mathbf{x})/\overline{F}(yu)}{\overline{F}(y \mid \mathbf{x})/\overline{F}(y)} \frac{\overline{F}(yu)}{\overline{F}(y)} - u^{-\alpha} \right) du \\ &= r \int_1^\infty (u - 1)^{r-1} \left(\frac{h(\mathbf{x}) \left[1 + O(A(1/[1 - F(yu)])) \right]}{h(\mathbf{x}) \left[1 + O(A(1/[1 - F(y)])) \right]} \frac{\overline{F}(yu)}{\overline{F}(y)} - u^{-\alpha} \right) du \\ &= r \int_1^\infty (u - 1)^{r-1} \left(\left[1 + O(A(1/[1 - F(y)])) \right] \frac{\overline{F}(yu)}{\overline{F}(y)} - u^{-\alpha} \right) du \end{aligned}$$

$$\begin{aligned}
&= r \int_1^\infty (u-1)^{r-1} \left(\frac{\bar{F}(yu)}{\bar{F}(y)} - u^{-\alpha} \right) du + o(1)r \int_1^\infty (u-1)^{r-1} \frac{\bar{F}(yu)}{\bar{F}(y)} du \\
&= r \int_1^\infty (u-1)^{r-1} \left(\frac{\bar{F}(yu)}{\bar{F}(y)} - u^{-\alpha} \right) du + o(1)r \int_1^\infty (u-1)^{r-1} \left(\frac{\bar{F}(yu)}{\bar{F}(y)} - u^{-\alpha} \right) du \\
&\quad + o(1)r \int_1^\infty (u-1)^{r-1} u^{-\alpha} du,
\end{aligned}$$

where the fourth step uses Assumption 6 (ii). We now obtain the desired result because

$$\int_1^\infty (u-1)^{r-1} \left(\frac{\bar{F}(yu)}{\bar{F}(y)} - u^{-\alpha} \right) du = o(1)$$

as $y \rightarrow \infty$, which follows from the proof of Proposition 4.5 in Pan et al. (2013). \square

Lemma 5. *Under the assumptions of Theorem 2 it holds that, as $n \rightarrow \infty$,*

$$\sup_{\mathbf{x} \in \mathcal{X}} \left| \frac{\text{ES}_\tau(\mathbf{x})}{\text{ES}_\tau h^\gamma(\mathbf{x})} - 1 \right| = O\left(A(1/[1-\tau])\right),$$

where $\text{ES}_\tau = \frac{1}{1-\tau} \int_\tau^1 F^{\leftarrow}(\vartheta) d\vartheta$ is the ES pertaining to $F(\cdot)$, and $\text{ES}_\tau(\mathbf{x}) = \text{ES}_\tau(Y_t \mid \mathbf{X}_t = \mathbf{x})$ is as defined in Section 3 of the main paper.

Proof: Define $Q_\vartheta(\mathbf{x})$ to be the ϑ -quantile pertaining to $\mathbb{P}\{Y_t \leq \cdot \mid \mathbf{X}_t = \mathbf{x}\}$. Then, it holds that

$$\begin{aligned}
\text{ES}_\tau(\mathbf{x}) &= \frac{1}{1-\tau} \int_\tau^1 Q_\vartheta(\mathbf{x}) d\vartheta \\
&= \frac{1}{1-\tau} \int_\tau^1 F^{\leftarrow}(\vartheta) h^\gamma(\mathbf{x}) \frac{Q_\vartheta(\mathbf{x})}{F^{\leftarrow}(\vartheta) h^\gamma(\mathbf{x})} d\vartheta.
\end{aligned} \tag{27}$$

From the proof of Lemma 2 in Hou et al. (2024) we have for all $\mathbf{x} \in \mathcal{X}$ that

$$\left(1 - \frac{CA(1/[1-\vartheta])}{\underline{g}}\right)^\gamma \leq \frac{Q_\vartheta(\mathbf{x})}{F^{\leftarrow}(\vartheta) h^\gamma(\mathbf{x})} \leq \left(1 + \frac{CA(1/[1-\vartheta])}{\bar{g}}\right)^\gamma \tag{28}$$

for some $\underline{g}, \bar{g}, C > 0$, all independent of \mathbf{x} . By Assumption 6 (ii), $A(\cdot)$ is eventually decreasing, such that for sufficiently large n ,

$$A(1/[1-\tau]) \geq A(1/[1-\vartheta]) \quad \text{for } \vartheta \in [\tau, 1).$$

Hence, from (28),

$$\left(1 - \frac{CA(1/[1-\tau])}{\underline{g}}\right)^\gamma \leq \frac{Q_\vartheta(\mathbf{x})}{F^{\leftarrow}(\vartheta) h^\gamma(\mathbf{x})} \leq \left(1 + \frac{CA(1/[1-\tau])}{\bar{g}}\right)^\gamma.$$

Inserting the right-hand side inequality into (27), we obtain that

$$\begin{aligned} \text{ES}_\tau(\mathbf{x}) &\leq \frac{1}{1-\tau} \int_\tau^1 F^{\leftarrow}(\vartheta) \, \text{d}\vartheta \, h^\gamma(\mathbf{x}) \left(1 + \frac{CA(1/[1-\tau])}{\underline{g}}\right)^\gamma \\ &= \text{ES}_\tau h^\gamma(\mathbf{x}) \left(1 + \frac{CA(1/[1-\tau])}{\underline{g}}\right)^\gamma \end{aligned}$$

and, similarly, using the left-hand side inequality,

$$\text{ES}_\tau(\mathbf{x}) \geq \text{ES}_\tau h^\gamma(\mathbf{x}) \left(1 - \frac{CA(1/[1-\tau])}{\underline{g}}\right)^\gamma.$$

We conclude from these two relations that, uniformly in $\mathbf{x} \in \mathcal{X}$,

$$\frac{\text{ES}_\tau(\mathbf{x})}{\text{ES}_\tau} = h^\gamma(\mathbf{x}) \left[1 + O\left(A(1/[1-\tau])\right)\right],$$

which is just the conclusion. \square

Lemma 6. *Under the assumptions of Theorem 2 it holds for any $r \in (0, 1/\gamma)$ that, as $n \rightarrow \infty$,*

$$\mathbb{E}_t[\mathbf{1}_{\{\varepsilon_{t\tau} > 0\}} |\xi_{t\tau}|^r] \stackrel{a.s.}{\asymp} (1-\tau) [F^{\leftarrow}(\tau)]^r.$$

Proof: For brevity, we set $Q_{t\tau} := Q_\tau(Y_t \mid \mathbf{X}_t) = \mathbf{X}_t' \boldsymbol{\alpha}_{0\tau}$ and $\text{ES}_{t\tau} := \text{ES}_\tau(Y_t \mid \mathbf{X}_t) = \mathbf{X}_t' \boldsymbol{\beta}_{0\tau}$.

Then, it holds that

$$\begin{aligned} \mathbb{E}_t[\mathbf{1}_{\{\varepsilon_{t\tau} > 0\}} |\xi_{t\tau}|^r] &= \mathbb{E}_t[\mathbf{1}_{\{Y_t > \mathbf{X}_t' \boldsymbol{\alpha}_{0\tau}\}} |Y_t - \mathbf{X}_t' \boldsymbol{\beta}_{0\tau}|^r] \\ &= \mathbb{E}_t[\mathbf{1}_{\{Y_t > Q_{t\tau}\}} |Y_t - \text{ES}_{t\tau}|^r] \\ &= \mathbb{E}_t[\mathbf{1}_{\{Q_{t\tau} < Y_t \leq \text{ES}_{t\tau}\}} (\text{ES}_{t\tau} - Y_t)^r] + \mathbb{E}_t[\mathbf{1}_{\{Y_t > \text{ES}_{t\tau}\}} (Y_t - \text{ES}_{t\tau})^r] \\ &=: A_{1t\tau} + B_{1t\tau}. \end{aligned} \tag{29}$$

Before considering $A_{1t\tau}$ and $B_{1t\tau}$, note that from Pan et al. (2013, Theorem 4.2),

$$\lim_{n \rightarrow \infty} \frac{1}{A_1(1/[1 - F(F^{\leftarrow}(\tau))])} \left[\frac{\text{ES}_\tau}{F^{\leftarrow}(\tau)} - \frac{1}{1-\gamma} \right] = \frac{1}{(1/\gamma - 1)(1/\gamma - 1 - \xi)}. \tag{30}$$

By continuity, it holds that

$$F(F^{\leftarrow}(\tau)) = \tau,$$

for sufficiently large n , such that from Assumption 6,

$$A_1\left(1/[1 - F(F^{\leftarrow}(\tau))]\right) = A_1(1/[1 - \tau]) = O(A_1(n/k)) = o(k^{-1/2}).$$

Hence, from (30),

$$\frac{\text{ES}_\tau}{F^{\leftarrow}(\tau)} = \frac{1}{1 - \gamma} + o(k^{-1/2}). \quad (31)$$

Therefore,

$$\begin{aligned} \frac{\text{ES}_{t\tau}}{Q_{t\tau}} &= \frac{\text{ES}_{t\tau} / \text{ES}_\tau}{Q_{t\tau} / F^{\leftarrow}(\tau)} \frac{\text{ES}_\tau}{F^{\leftarrow}(\tau)} \\ &= \frac{h^\gamma(\mathbf{X}_t) \left\{ 1 + O(A(1/[1 - \tau])) \right\}}{h^\gamma(\mathbf{X}_t) \left\{ 1 + O(A(1/[1 - \tau])) \right\}} \frac{\text{ES}_\tau}{F^{\leftarrow}(\tau)} \\ &= \left\{ 1 + O(A(1/[1 - \tau])) \right\} \left[\frac{1}{1 - \gamma} + o(k^{-1/2}) \right] \\ &= \frac{1}{1 - \gamma} + o(k^{-1/2}), \end{aligned} \quad (32)$$

where we used Lemma 5 and Hou et al. (2024, Lemma 2) in the second step, (31) in the third step, and $\sqrt{k}A(1/[1 - \tau]) = \sqrt{k}A(n/k) = o(1)$ from Assumption 6 in the final step. Hence,

$$\begin{aligned} \text{ES}_{t\tau} - Q_{t\tau} &= \frac{\text{ES}_{t\tau} - Q_{t\tau}}{Q_{t\tau}} \frac{Q_{t\tau}}{F^{\leftarrow}(\tau)} F^{\leftarrow}(\tau) \\ &= \left(\frac{\text{ES}_{t\tau}}{Q_{t\tau}} - 1 \right) h^\gamma(\mathbf{X}_t) \left\{ 1 + O(A(1/[1 - \tau])) \right\} F^{\leftarrow}(\tau) \\ &= \left[\frac{\gamma}{1 - \gamma} + o(1) \right] h^\gamma(\mathbf{X}_t) \left\{ 1 + O(A(1/[1 - \tau])) \right\} F^{\leftarrow}(\tau) \\ &= \left[\frac{\gamma}{1 - \gamma} + o(1) \right] h^\gamma(\mathbf{X}_t) F^{\leftarrow}(\tau), \end{aligned} \quad (33)$$

where we used Lemma 2 of Hou et al. (2024) in the second step, and (32) in the third step.

Additionally, we shall exploit

$$\begin{aligned} \text{ES}_{t\tau} &= \left(\frac{\text{ES}_{t\tau}}{Q_{t\tau}} \right) \left(\frac{Q_{t\tau}}{F^{\leftarrow}(\tau)} \right) F^{\leftarrow}(\tau) \\ &= \left[\frac{1}{1 - \gamma} + o(k^{-1/2}) \right] \left(h^\gamma(\mathbf{X}_t) [1 + o(1)] \right) F^{\leftarrow}(\tau) \\ &= \left[\frac{1}{1 - \gamma} + o(k^{-1/2}) \right] h^\gamma(\mathbf{X}_t) F^{\leftarrow}(\tau), \end{aligned} \quad (34)$$

where we again utilized Lemma 2 of Hou et al. (2024) and (32).

Now, for $B_{1t\tau}$ we get with $\alpha = 1/\gamma$ that, as $n \rightarrow \infty$,

$$\begin{aligned}
B_{1t\tau} &= \mathbb{E}_t[\mathbf{1}_{\{Y_t > \text{ES}_{t\tau}\}}(Y_t - \text{ES}_{t\tau})^r] \\
&= \mathbb{P}_t\{Y_t > \text{ES}_{t\tau}\} \mathbb{E}_t[(Y_t - \text{ES}_{t\tau})^r \mid Y_t > \text{ES}_{t\tau}] \\
&= \mathbb{P}_t\{Y_t > \text{ES}_{t\tau}\} \text{ES}_{t\tau}^r \frac{\mathbb{E}_t[(Y_t - \text{ES}_{t\tau})^r \mid Y_t > \text{ES}_{t\tau}]}{\text{ES}_{t\tau}^r} \\
&= \mathbb{P}_t\{Y_t > \text{ES}_{t\tau}\} \text{ES}_{t\tau}^r \frac{\Gamma(r+1)\Gamma(\alpha-r)}{\Gamma(\alpha)} [1 + o(1)]
\end{aligned} \tag{35}$$

by Lemma 4. Moreover,

$$\begin{aligned}
&\mathbb{P}_t\{Y_t > \text{ES}_{t\tau}\} \\
&= \mathbb{P}_t\{Y_t > Q_{t\tau} + (\text{ES}_{t\tau} - Q_{t\tau})\} \\
&= 1 - F_t(F_t^{\leftarrow}(\tau) + (\text{ES}_{t\tau} - Q_{t\tau})) \\
&= (1 - \tau) \left[\frac{1 - F_t(F_t^{\leftarrow}(\tau) + \text{ES}_{t\tau} - Q_{t\tau})}{1 - \tau} - 1 \right] + (1 - \tau) \\
&= (1 - \tau) \left\{ [h(\mathbf{X}_t) + o(1)] \left[h^\gamma(\mathbf{X}_t) + o(1) + \frac{\text{ES}_{t\tau} - Q_{t\tau}}{F_t^{\leftarrow}(\tau)} \right]^{-1/\gamma} [1 + o(1)] - 1 \right\} + (1 - \tau) \\
&= (1 - \tau) [h(\mathbf{X}_t) + o(1)] \left[h^\gamma(\mathbf{X}_t) + o(1) + \left\{ \frac{\gamma}{1 - \gamma} + o(1) \right\} h^\gamma(\mathbf{X}_t) \right]^{-1/\gamma} [1 + o(1)] \\
&\stackrel{\text{a.s.}}{\asymp} (1 - \tau),
\end{aligned}$$

where we used (15) in the fourth step, and (33) in the fifth. Plugging this and (34) into (35), we deduce that

$$B_{1t\tau} \stackrel{\text{a.s.}}{\asymp} (1 - \tau) [F_t^{\leftarrow}(\tau)]^r. \tag{36}$$

For $A_{1t\tau}$, we follow similar steps as in the proof of Lemma 3 to obtain that

$$\begin{aligned}
A_{1t\tau} &= \mathbb{E}_t[\mathbf{1}_{\{Q_{t\tau} < Y_t \leq \text{ES}_{t\tau}\}}(\text{ES}_{t\tau} - Y_t)^r] \\
&= \mathbb{E}_t[\mathbf{1}_{\{0 < \varepsilon_{t\tau} \leq \text{ES}_{t\tau} - Q_{t\tau}\}}(\text{ES}_{t\tau} - Q_{t\tau} - \varepsilon_{t\tau})^r] \\
&= \frac{1}{r+1} \frac{1 - \tau}{F_t^{\leftarrow}(\tau)} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} [1 + o(1)] (\text{ES}_{t\tau} - Q_{t\tau})^{r+1} + r(1 - \tau) (\text{ES}_{t\tau} - Q_{t\tau})^r o(k^{-1/2}).
\end{aligned}$$

Additionally using (33), we get that

$$A_{1t\tau} = \frac{1}{r+1} \frac{1 - \tau}{F_t^{\leftarrow}(\tau)} \left[\frac{\gamma}{1 - \gamma} + o(1) \right]^{r+1} h^{(r+1)\gamma}(\mathbf{X}_t) [F_t^{\leftarrow}(\tau)]^{r+1}$$

$$\begin{aligned}
& + r(1-\tau) \left[\frac{\gamma}{1-\gamma} + o(1) \right]^r h^{r\gamma}(\mathbf{X}_t) [F^{\leftarrow}(\tau)]^r o(k^{-1/2}) \\
& \stackrel{\text{a.s.}}{\asymp} (1-\tau) [F^{\leftarrow}(\tau)]^r.
\end{aligned}$$

Combining this and (36) with (29), we deduce that

$$\mathbb{E}_t [\mathbb{1}_{\{\varepsilon_{t\tau} > 0\}} |\xi_{t\tau}|^r] \stackrel{\text{a.s.}}{\asymp} (1-\tau) [F^{\leftarrow}(\tau)]^r,$$

as desired. \square

Proof of Proposition 3: Slightly overloading notation introduced in the proof of Proposition 1, we decompose

$$\frac{1}{\sqrt{n\tau(1-\tau)}} \sum_{t=1}^{\lfloor ns \rfloor} \psi_*(\varepsilon_{t\tau}, \xi_{t\tau}) \mathbf{X}_t = \sum_{j=1}^{m_n(s)} \mathbf{Y}_j^B + \sum_{j=1}^{m_n(s)} \mathbf{Y}_j^S + \mathbf{R}_n(s), \quad (37)$$

where

$$\begin{aligned}
\mathbf{Y}_j^B &= \sum_{t \in B_j} \frac{1}{\sqrt{n\tau(1-\tau)}} \psi_*(\varepsilon_{t\tau}, \xi_{t\tau}) \mathbf{X}_t, \\
\mathbf{Y}_j^S &= \sum_{t \in S_j} \frac{1}{\sqrt{n\tau(1-\tau)}} \psi_*(\varepsilon_{t\tau}, \xi_{t\tau}) \mathbf{X}_t, \\
\mathbf{R}_n(s) &= \sum_{t \in B_{m_n(s)+1}} \frac{1}{\sqrt{n\tau(1-\tau)}} \psi_*(\varepsilon_{t\tau}, \xi_{t\tau}) \mathbf{X}_t.
\end{aligned}$$

We consider each of the three terms on the right-hand side of (37) individually.

First, we show that $\sup_{s \in [0,1]} \|\mathbf{R}_n(s)\| = o_{\mathbb{P}}(1)$. Since $|B_{m_n(s)+1}| \leq r_n + l_n - 1$ for the cardinality, it suffices to show that, as $n \rightarrow \infty$,

$$\frac{r_n}{\sqrt{n\tau(1-\tau)}} \frac{1}{F^{\leftarrow}(\tau)} \max_{t=1, \dots, n} \|\mathbb{1}_{\{\varepsilon_{t\tau} > 0\}} \xi_{t\tau} \mathbf{X}_t\| = o_{\mathbb{P}}(1).$$

By subadditivity, Markov's inequality and Lemma 6,

$$\begin{aligned}
& \mathbb{P} \left\{ \frac{r_n}{\sqrt{n\tau(1-\tau)}} \frac{1}{F^{\leftarrow}(\tau)} \max_{t=1, \dots, n} \|\mathbb{1}_{\{\varepsilon_{t\tau} > 0\}} \xi_{t\tau} \mathbf{X}_t\| > \varepsilon \right\} \\
& \leq \sum_{t=1}^n \mathbb{P} \left\{ \frac{r_n}{\sqrt{n\tau(1-\tau)}} \frac{1}{F^{\leftarrow}(\tau)} \|\mathbb{1}_{\{\varepsilon_{t\tau} > 0\}} \xi_{t\tau} \mathbf{X}_t\| > \varepsilon \right\} \\
& \leq \sum_{t=1}^n \frac{1}{\varepsilon^4} \mathbb{E} \left[\left\{ \frac{r_n}{\sqrt{n\tau(1-\tau)}} \frac{1}{F^{\leftarrow}(\tau)} \|\mathbb{1}_{\{\varepsilon_{t\tau} > 0\}} \xi_{t\tau} \mathbf{X}_t\| \right\}^4 \right]
\end{aligned}$$

$$\begin{aligned}
&\leq K \frac{r_n^4}{[n\tau(1-\tau)]^2} \frac{1}{[F^{\leftarrow}(\tau)]^4} \sum_{t=1}^n \mathbb{E} \left[\mathbb{E}_t \left\{ \mathbf{1}_{\{\varepsilon_{t\tau} > 0\}} |\xi_{t\tau}|^4 \right\} \|\mathbf{X}_t\|^4 \right] \\
&\leq K \frac{r_n^4}{[n\tau(1-\tau)]^2} \frac{1}{[F^{\leftarrow}(\tau)]^4} nK(1-\tau) [F^{\leftarrow}(\tau)]^4 \\
&\leq K \frac{r_n^4}{n(1-\tau)} \\
&= o(1)
\end{aligned}$$

due to $r_n = o(k^{1/4})$ from Assumption 2.

Second, we consider $\sum_{j=1}^{m_n(s)} \mathbf{Y}_j^B$. Again, the distributional limit of $\sum_{t=1}^{m_n(\cdot)} \mathbf{Y}_j^B$ is the same as that of $\sum_{j=1}^{m_n(\cdot)} \tilde{\mathbf{Y}}_j^B$, where $\{\tilde{\mathbf{Y}}_j^B\}_j$ is an i.i.d. sequence with $\tilde{\mathbf{Y}}_1^B \stackrel{d}{=} \mathbf{Y}_1^B$. Once more, we apply the martingale difference FCLT of [Jacod and Shiryaev \(1987, Theorem 3.33 in Chapter VIII\)](#). The conditional variances converge as follows. By Assumption 3,

$$\begin{aligned}
\sum_{j=1}^{m_n(s)} \mathbb{E}_{j-1} [\tilde{\mathbf{Y}}_j^B \tilde{\mathbf{Y}}_j^{B'}] &= m_n(s) \mathbb{E} [\tilde{\mathbf{Y}}_1^B \tilde{\mathbf{Y}}_1^{B'}] \\
&= m_n(s) \frac{1}{n\tau(1-\tau)} \mathbb{E} \left[\sum_{t=1}^{r_n} \psi_*(\varepsilon_{t\tau}, \xi_{t\tau}) \mathbf{X}_t \sum_{t=1}^{r_n} \psi_*(\varepsilon_{t\tau}, \xi_{t\tau}) \mathbf{X}_t' \right] \longrightarrow s\mathbf{J}_*,
\end{aligned}$$

as $n \rightarrow \infty$. Moreover, the conditional Lindeberg condition

$$\sum_{j=1}^{m_n(s)} \mathbb{E}_{j-1} \left[\|\tilde{\mathbf{Y}}_j^B\|^2 \mathbf{1}_{\{\|\tilde{\mathbf{Y}}_j^B\| > \delta\}} \right] \xrightarrow{\mathbb{P}} 0$$

holds for any $\delta > 0$. To see this, exploit strict stationarity and the Cauchy–Schwarz inequality to obtain that

$$\begin{aligned}
\sum_{j=1}^{m_n(s)} \mathbb{E}_{j-1} \left[\|\tilde{\mathbf{Y}}_j^B\|^2 \mathbf{1}_{\{\|\tilde{\mathbf{Y}}_j^B\| > \delta\}} \right] &= m_n(s) \mathbb{E} \left[\|\tilde{\mathbf{Y}}_1^B\|^2 \mathbf{1}_{\{\|\tilde{\mathbf{Y}}_1^B\| > \delta\}} \right] \\
&\leq m_n(s) \sqrt{\mathbb{E} \left[\|\tilde{\mathbf{Y}}_1^B\|^4 \right] \mathbb{P} \left\{ \|\tilde{\mathbf{Y}}_1^B\| > \delta \right\}}. \tag{38}
\end{aligned}$$

For the expectation in (38), we get using the law of iterated expectations (LIE) and Lemma 6 that

$$\begin{aligned}
\mathbb{E} \left[\|\tilde{\mathbf{Y}}_1^B\|^4 \right] &\leq K \frac{r_n^4}{[n(1-\tau)]^2} \mathbb{E} \left[\|\mathbf{X}_t\|^4 \mathbb{E}_t \left\{ \psi_*^4(\varepsilon_{t\tau}, \xi_{t\tau}) \right\} \right] \\
&\leq K \frac{r_n^4}{[n(1-\tau)]^2} (1-\tau)
\end{aligned}$$

$$= K \frac{r_n^4}{n^2(1-\tau)}. \quad (39)$$

The probability in (38) can be bounded by

$$\begin{aligned} \mathbb{P}\left\{\|\tilde{\mathbf{Y}}_1^B\| > \delta\right\} &\leq \mathbb{P}\left\{\frac{1}{F^{\leftarrow}(\tau)} \frac{1}{\sqrt{n\tau(1-\tau)}} \sum_{t=1}^{r_n} \mathbf{1}_{\{\varepsilon_{t\tau} > 0\}} |\xi_{t\tau}| > \delta\right\} \\ &\leq r_n \mathbb{P}\left\{\mathbf{1}_{\{\varepsilon_{t\tau} > 0\}} |\xi_{t\tau}| > \delta \frac{F^{\leftarrow}(\tau) \sqrt{n(1-\tau)}}{r_n}\right\} \\ &\leq r_n \mathbb{P}\left\{\mathbf{1}_{\{\varepsilon_{t\tau} > 0\}} |\varepsilon_{t\tau} - (\text{ES}_{t\tau} - Q_{t\tau})| > \delta \frac{F^{\leftarrow}(\tau) \sqrt{k}}{r_n}\right\} \\ &\leq r_n \mathbb{P}\left\{\varepsilon_{t\tau} > \frac{\delta F^{\leftarrow}(\tau) \sqrt{k}}{2 r_n}\right\} \\ &\leq r_n \left[1 - F_t\left(F_t^{\leftarrow}(\tau) + \frac{\delta F^{\leftarrow}(\tau) \sqrt{k}}{2 r_n}\right)\right] \\ &= r_n(1-\tau) \frac{1 - F_t\left(F_t^{\leftarrow}(\tau) + \frac{\delta F^{\leftarrow}(\tau) \sqrt{k}}{2 r_n}\right)}{1-\tau} \\ &= r_n(1-\tau) \left[h(\mathbf{X}_t) + o(1)\right] \times \\ &\quad \times \left[h^\gamma(\mathbf{X}_t) \{1 + o(1)\} + \frac{\delta \sqrt{k}}{2 r_n}\right]^{-1/\gamma} [1 + o(1)] \\ &\leq K r_n(1-\tau) \left(\frac{r_n}{\sqrt{k}}\right)^{1/\gamma}, \end{aligned} \quad (40)$$

where we used (1)–(2) in the third step, the fact that $\text{ES}_{t\tau} - Q_{t\tau} \asymp F^{\leftarrow}(\tau) \leq F^{\leftarrow}(\tau) \sqrt{k}/r_n$ from (33) in the fourth step, (16) in the fifth step, and (15) in the seventh step. Plugging (39)–(40) into (38), we get that

$$\begin{aligned} \sum_{j=1}^{m_n(s)} \mathbb{E}_{j-1} \left[\|\tilde{\mathbf{Y}}_j^B\|^2 \mathbf{1}_{\{\|\tilde{\mathbf{Y}}_j^B\| > \delta\}} \right] &\leq K \frac{n}{r_n} \sqrt{\frac{r_n^4}{n^2(1-\tau)} r_n(1-\tau) \left(\frac{r_n}{\sqrt{k}}\right)^{1/\gamma}} \\ &= K \sqrt{\frac{n^2 r_n^4}{r_n^2 n^2} r_n \left(\frac{r_n}{\sqrt{k}}\right)^{1/\gamma}} \\ &= K \sqrt{r_n^3 \left(\frac{r_n}{\sqrt{k}}\right)^{1/\gamma}} \\ &\leq K \sqrt{\frac{r_n^7}{k^2}} \\ &= o(1), \end{aligned}$$

where the fourth step uses that $\gamma \in (0, 1/4)$, and the final one that $r_n = o(k^{1/4})$.

Therefore, Theorem 3.33 of [Jacod and Shiryaev \(1987, Chapter VIII\)](#) implies that, as $n \rightarrow \infty$,

$$\sum_{j=1}^{m_n(s)} \tilde{\mathbf{Y}}_j^B \xrightarrow{d} \mathbf{Y}(s) \quad \text{in } D_m[0, 1],$$

where $\mathbf{Y}(s)$ is a continuous Gaussian martingale with quadratic variation $\langle \mathbf{Y} \rangle_s = s\mathbf{J}_*$. However, by Levy's characterization of Wiener processes (e.g., [Jacod and Shiryaev, 1987, Theorem 4.4](#) in Chapter II), $\mathbf{Y}(\cdot)$ is a Wiener process with covariance matrix \mathbf{J}_* . By the above arguments, this then shows that

$$\sum_{j=1}^{m_n(s)} \mathbf{Y}_j^B \xrightarrow{d} \mathbf{J}_*^{1/2} \mathbf{W}(s) \quad \text{in } D_m[0, 1]$$

for a m -variate standard Wiener process $\mathbf{W}(\cdot)$.

Third, we show negligibility of the small block part $\sum_{j=1}^{m_n(s)} \mathbf{Y}_j^S$. As for the big blocks, the asymptotic limit of $\sum_{j=1}^{m_n(\cdot)} \mathbf{Y}_j^S$ (if it exists) is the same as that of $\sum_{j=1}^{m_n(\cdot)} \tilde{\mathbf{Y}}_j^S$, where $\{\tilde{\mathbf{Y}}_j^S\}_j$ are i.i.d. with $\tilde{\mathbf{Y}}_1^S \stackrel{d}{=} \mathbf{Y}_1^S$. Now, by Kolmogorov's inequality (e.g., [Davidson, 1994, Theorem 15.14](#)) it holds for any generic element \tilde{Y}_j^S of $\tilde{\mathbf{Y}}_j^S$ that

$$\begin{aligned} \mathbb{P} \left\{ \sup_{s \in [0, 1]} \left| \sum_{j=1}^{m_n(s)} \tilde{Y}_j^S \right| > \varepsilon \right\} &= \mathbb{P} \left\{ \max_{1 \leq \ell \leq m_n(1)} \left| \sum_{j=1}^{\ell} \tilde{Y}_j^S \right| > \varepsilon \right\} \\ &\leq \varepsilon^{-2} \sum_{j=1}^{m_n(1)} \text{Var}(\tilde{Y}_j^S) \\ &= o(1) \end{aligned}$$

by Assumption 3. Hence, $\sup_{s \in [0, 1]} \left| \sum_{j=1}^{m_n(s)} \mathbf{Y}_j^S \right| = o_{\mathbb{P}}(1)$, as $n \rightarrow \infty$.

Overall, the conclusion follows. \square

Proof of Proposition 4: Use $\psi(\varepsilon_{t\tau}) = \tau - \mathbf{1}_{\{\varepsilon_{t\tau} \leq 0\}} = \mathbf{1}_{\{\varepsilon_{t\tau} > 0\}} - (1 - \tau)$ to obtain that

$$\begin{aligned} &\sup_{s \in [0, 1]} \left\| \frac{1}{n(1 - \tau)} \sum_{t=1}^{\lfloor ns \rfloor} \mathbf{1}_{\{\varepsilon_{t\tau} > 0\}} \mathbf{X}_t \mathbf{X}_t' - s\mathbf{D}_* \right\| \\ &= \sup_{s \in [0, 1]} \left\| \frac{1}{n(1 - \tau)} \sum_{t=1}^{\lfloor ns \rfloor} \psi(\varepsilon_{t\tau}) \mathbf{X}_t \mathbf{X}_t' + \frac{1}{n} \sum_{t=1}^{\lfloor ns \rfloor} \mathbf{X}_t \mathbf{X}_t' - s\mathbf{D}_* \right\| \\ &\leq \sup_{s \in [0, 1]} \left\| \frac{1}{n(1 - \tau)} \sum_{t=1}^{\lfloor ns \rfloor} \psi(\varepsilon_{t\tau}) \mathbf{X}_t \mathbf{X}_t' \right\| + \sup_{s \in [0, 1]} \left\| \frac{1}{n} \sum_{t=1}^{\lfloor ns \rfloor} \mathbf{X}_t \mathbf{X}_t' - s\mathbf{D}_* \right\|. \end{aligned} \quad (41)$$

We consider the two right-hand side terms separately. First, note that $\{\mathbf{X}_t \mathbf{X}'_t - \mathbb{E}[\mathbf{X}_t \mathbf{X}'_t]\}$ is a mean-zero, α -mixing sequence (since β -mixing implies α -mixing by [Bradley, 2005](#)). Thus, from arguments in [Andrews \(1988, Section 3\)](#), the sequence is an L_1 -mixingale. The corollary of [Hansen \(1992, p. 493\)](#) therefore implies that

$$\sup_{s \in [0,1]} \left\| \frac{1}{n} \sum_{t=1}^{\lfloor ns \rfloor} \mathbf{X}_t \mathbf{X}'_t - s \mathbf{D}_* \right\| = \sup_{s \in [0,1]} \left\| \frac{1}{n} \sum_{t=1}^{\lfloor ns \rfloor} \left\{ \mathbf{X}_t \mathbf{X}'_t - \mathbb{E}[\mathbf{X}_t \mathbf{X}'_t] \right\} \right\| + o(1) = o_{\mathbb{P}}(1), \quad (42)$$

where we also exploited [Assumption 4](#).

Second, to prove

$$\sup_{s \in [0,1]} \left\| \frac{1}{n(1-\tau)} \sum_{t=1}^{\lfloor ns \rfloor} \psi(\varepsilon_{t\tau}) \mathbf{X}_t \mathbf{X}'_t \right\| = o_{\mathbb{P}}(1), \quad (43)$$

we show that an arbitrary element $\frac{1}{n(1-\tau)} \sum_{t=1}^{\lfloor ns \rfloor} \psi(\varepsilon_{t\tau}) X_{it} X_{lt}$ of this random matrix vanishes asymptotically (in probability). To do so, we proceed as in the proof of [Proposition 1](#) and decompose

$$\frac{1}{n(1-\tau)} \sum_{t=1}^{\lfloor ns \rfloor} \psi(\varepsilon_{t\tau}) X_{it} X_{lt} = \sum_{j=1}^{m_n(s)} Y_j^B + \sum_{j=1}^{m_n(s)} Y_j^S + R_n(s),$$

where

$$\begin{aligned} Y_j^B &= \sum_{t \in B_j} \frac{1}{n(1-\tau)} \psi(\varepsilon_{t\tau}) X_{it} X_{lt}, \\ Y_j^S &= \sum_{t \in S_j} \frac{1}{n(1-\tau)} \psi(\varepsilon_{t\tau}) X_{it} X_{lt}, \\ R_n(s) &= \sum_{t \in B_{m_n(s)+1}} \frac{1}{n(1-\tau)} \psi(\varepsilon_{t\tau}) X_{it} X_{lt}. \end{aligned}$$

Boundedness of $\psi(\varepsilon_{t\tau}) X_{it} X_{lt}$ implies that

$$0 \leq \sup_{s \in [0,1]} |R_n(s)| \leq K \frac{r_n}{n(1-\tau)} = o(1).$$

From by now familiar arguments, $\sum_{j=1}^{m_n(\cdot)} Y_j^B$ has the same asymptotic limit (if any) as $\sum_{j=1}^{m_n(\cdot)} \tilde{Y}_j^B$, where \tilde{Y}_j^B are i.i.d. copies of Y_j^B .

We again invoke the FCLT for MDAs of [Jacod and Shiryaev \(1987\)](#). Here, the conditional

variances vanish, since

$$\begin{aligned}
\sum_{j=1}^{m_n(s)} \mathbb{E}_{j-1} [\tilde{Y}_j^B \tilde{Y}_j^B] &= m_n(s) \mathbb{E} [\tilde{Y}_1^B \tilde{Y}_1^B] \\
&= m_n(s) \frac{1}{[n(1-\tau)]^2} \mathbb{E} \left[\left\{ \sum_{t=1}^{r_n} \psi(\varepsilon_{t\tau}) X_{it} X_{lt} \right\} \left\{ \sum_{t=1}^{r_n} \psi(\varepsilon_{t\tau}) X_{it} X_{lt} \right\} \right] \\
&\leq m_n(s) \frac{1}{[n(1-\tau)]^2} r_n^2 \text{Var} (\psi(\varepsilon_{t\tau}) X_{it} X_{lt}) \\
&= m_n(s) \frac{1}{[n(1-\tau)]^2} r_n^2 \mathbb{E} \left[\mathbb{E}_t \{ \psi^2(\varepsilon_{t\tau}) (X_{it} X_{lt})^2 \} \right] \\
&\leq K m_n(s) \frac{1}{[n(1-\tau)]^2} r_n^2 (1-\tau) \\
&= K \frac{r_n}{n(1-\tau)} \\
&= K \frac{r_n}{k} \\
&= o(1).
\end{aligned}$$

Since

$$\|\tilde{Y}_j^B\| \stackrel{\text{a.s.}}{\leq} K \frac{r_n}{n(1-\tau)} = K \frac{r_n}{k} = o(1),$$

the conditional Lindeberg condition is immediate. Therefore, Theorem 3.33 of [Jacod and Shiryaev \(1987, Chapter VIII\)](#) implies that, as $n \rightarrow \infty$,

$$\sum_{j=1}^{m_n(s)} \tilde{Y}_j^B \xrightarrow{d} 0 \quad \text{in } D[0, 1],$$

whence $\sum_{j=1}^{m_n(s)} Y_j^B \xrightarrow{d} 0$ in $D[0, 1]$. This together with the continuous mapping theorem (CMT; e.g., [Davidson, 1994, Theorem 26.13](#)) implies that

$$\sup_{s \in [0, 1]} \left\| \sum_{j=1}^{m_n(s)} Y_j^B \right\| = o_{\mathbb{P}}(1).$$

Similar arguments show that $\sup_{s \in [0, 1]} \left\| \sum_{j=1}^{m_n(s)} Y_j^S \right\| = o_{\mathbb{P}}(1)$, establishing (43).

Combining (42)–(43), the conclusion follows from (41). □

Proof of Proposition 5: We establish the proposition by showing that

$$\sup_{s \in [\epsilon, 1]} \left\| \frac{1}{n(1-\tau)} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{1}_{\{\mathbf{X}'_t(\widehat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau}) < \varepsilon_{t\tau} \leq 0\}} \mathbf{X}_t \mathbf{X}'_t \right\| \xrightarrow{\mathbb{P}} 0$$

and

$$\sup_{s \in [\epsilon, 1]} \left\| \frac{1}{n(1-\tau)} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq \mathbf{X}'_t(\widehat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau})\}} \mathbf{X}_t \mathbf{X}'_t \right\| \xrightarrow{\mathbb{P}} 0.$$

We only prove the latter convergence, as the former follows similarly. Write

$$\begin{aligned} & \mathbb{P} \left\{ \sup_{s \in [\epsilon, 1]} \left\| \frac{1}{n(1-\tau)} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq \mathbf{X}'_t(\widehat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau})\}} \mathbf{X}_t \mathbf{X}'_t \right\| > \varepsilon \right\} \\ &= \mathbb{P} \left\{ \sup_{s \in [\epsilon, 1]} \left\| \frac{1}{n(1-\tau)} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{X}'_t a_n^{-1}(\widehat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau})\}} \mathbf{X}_t \mathbf{X}'_t \right\| > \varepsilon, \right. \\ & \quad \left. \sup_{s \in [\epsilon, 1]} \left\| a_n^{-1}(\widehat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau}) \right\| \leq K \right\} \\ &+ \mathbb{P} \left\{ \sup_{s \in [\epsilon, 1]} \left\| \frac{1}{n(1-\tau)} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{X}'_t a_n^{-1}(\widehat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau})\}} \mathbf{X}_t \mathbf{X}'_t \right\| > \varepsilon, \right. \\ & \quad \left. \sup_{s \in [\epsilon, 1]} \left\| a_n^{-1}(\widehat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau}) \right\| > K \right\} \\ &\leq \mathbb{P} \left\{ \sup_{s \in [\epsilon, 1]} \left\| \frac{1}{n(1-\tau)} \sum_{t=1}^{\lfloor ns \rfloor} \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq K a_n \|\mathbf{X}_t\|\}} \mathbf{X}_t \mathbf{X}'_t \right\| > \varepsilon \right\} + \mathbb{P} \left\{ \sup_{s \in [\epsilon, 1]} \left\| a_n^{-1}(\widehat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau}) \right\| > K \right\} \\ &\leq \mathbb{P} \left\{ \frac{1}{n(1-\tau)} \sum_{t=1}^n \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq K a_n \|\mathbf{X}_t\|\}} \|\mathbf{X}_t\|^2 > \varepsilon \right\} + o(1) \end{aligned} \tag{44}$$

as $n \rightarrow \infty$, followed by $K \rightarrow \infty$, where the final step follows from $\sup_{s \in [\epsilon, 1]} \left\| a_n^{-1}(\widehat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau}) \right\| = O_{\mathbb{P}}(1)$, which—in turn—is a consequence of Theorem 1 and the CMT. In light of this, we only have to show that

$$\frac{1}{n(1-\tau)} \sum_{t=1}^n \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq K a_n \|\mathbf{X}_t\|\}} \|\mathbf{X}_t\|^2 = o_{\mathbb{P}}(1).$$

To do so, we use that

$$\begin{aligned} \mathbb{E}_t \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq K a_n \|\mathbf{X}_t\|\}} \right] &= F_{\varepsilon_t | \mathbf{X}_t}(K a_n \|\mathbf{X}_t\|) - F_{\varepsilon_t | \mathbf{X}_t}(0) \\ &= F_t(F_t^{\leftarrow}(\tau) + K a_n \|\mathbf{X}_t\|) - \tau \\ &= (\tau - 1) \left[\frac{1 - F_t(F_t^{\leftarrow}(\tau) + K a_n \|\mathbf{X}_t\|)}{1 - \tau} - 1 \right] \\ &= \frac{1 - \tau}{F_t^{\leftarrow}(\tau)} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} K a_n \|\mathbf{X}_t\| [1 + o(1)] + (1 - \tau) O\left(A_1(1/[1 - \tau])\right) \end{aligned}$$

$$\begin{aligned}
&\leq K \frac{1-\tau}{F^{\leftarrow}(\tau)} a_n + (1-\tau) o\left(\frac{1}{\sqrt{n(1-\tau)}}\right) \\
&\leq K \sqrt{\frac{1-\tau}{n}},
\end{aligned}$$

where we used (16) in the second step and Lemma 2 in the fourth step. Hence, the LIE and the fact that \mathbf{X}_t is bounded imply

$$\mathbb{E}\left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq K a_n \|\mathbf{X}_t\|\}} \|\mathbf{X}_t\|^2\right] = \mathbb{E}\left[\mathbb{E}_t\left\{\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq K a_n \|\mathbf{X}_t\|\}}\right\} \|\mathbf{X}_t\|^2\right] \leq K \sqrt{\frac{1-\tau}{n}}.$$

Markov's inequality now yields

$$\begin{aligned}
\mathbb{P}\left\{\frac{1}{n(1-\tau)} \sum_{t=1}^n \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq K a_n \|\mathbf{X}_t\|\}} \|\mathbf{X}_t\|^2 > \varepsilon\right\} &\leq \frac{1}{\varepsilon} \frac{1}{n(1-\tau)} \sum_{t=1}^n \mathbb{E}\left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq K a_n \|\mathbf{X}_t\|\}} \|\mathbf{X}_t\|^2\right] \\
&\leq K \frac{1}{n(1-\tau)} \sum_{t=1}^n \sqrt{\frac{1-\tau}{n}} \\
&\leq K \frac{1}{\sqrt{n(1-\tau)}} \\
&= o(1),
\end{aligned}$$

such that the conclusion follows from (44). □

The proof of Proposition 6 requires the preliminary Lemmas 7–8. To introduce these, define

$$\begin{aligned}
\boldsymbol{\nu}_t(\mathbf{v}) &:= \frac{1}{F^{\leftarrow}(\tau)} \frac{1}{\sqrt{n(1-\tau)}} \left[\mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t < \varepsilon_{t\tau} \leq 0\}} - \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} \right] \xi_{t\tau} \mathbf{X}_t, \\
\bar{\boldsymbol{\nu}}_t(\mathbf{v}) &:= \mathbb{E}_t[\boldsymbol{\nu}_t(\mathbf{v})], \\
\mathbf{V}_n(\mathbf{v}, s) &:= \sum_{t=1}^{\lfloor ns \rfloor} \boldsymbol{\nu}_t(\mathbf{v}), \\
\bar{\mathbf{V}}_n(\mathbf{v}, s) &:= \sum_{t=1}^{\lfloor ns \rfloor} \bar{\boldsymbol{\nu}}_t(\mathbf{v}).
\end{aligned}$$

Lemma 7. *Under the assumptions of Theorem 2 it holds for any $K > 0$ that, as $n \rightarrow \infty$,*

$$\sup_{s \in [0,1]} \sup_{\|\mathbf{v}\| \leq K} \left\| \bar{\mathbf{V}}_n(\mathbf{v}, s) - s \frac{1}{1-\gamma} \mathbf{D}_* \mathbf{v} \right\| = o_{\mathbb{P}}(1).$$

Proof: See Section A.5. □

Lemma 8. Under the assumptions of Theorem 2 it holds for any $K > 0$ that, as $n \rightarrow \infty$,

$$\sup_{s \in [0,1]} \sup_{\|\mathbf{v}\| \leq K} \|\mathbf{V}_n(\mathbf{v}, s) - \bar{\mathbf{V}}_n(\mathbf{v}, s)\| = o_{\mathbb{P}}(1).$$

Proof: See Section A.5. □

Proof of Proposition 6: Plugging in $\hat{\mathbf{v}}_n(s) = a_n^{-1}[\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}_{0\tau}]$ for \mathbf{v} in Lemma 7 yields by Theorem 1 that, as $n \rightarrow \infty$,

$$\bar{\mathbf{V}}_n(\hat{\mathbf{v}}_n(s), s) \xrightarrow{d} \frac{1}{1-\gamma} \mathbf{D}_* \boldsymbol{\Omega}^{1/2} \mathbf{W}(s) \quad \text{in } D_m[\epsilon, 1],$$

where $\mathbf{W}(\cdot)$ is a m -variate standard Wiener process. It also holds that

$$\begin{aligned} & \mathbb{P} \left\{ \sup_{s \in [\epsilon, 1]} \|\mathbf{V}_n(\hat{\mathbf{v}}_n(s), s) - \bar{\mathbf{V}}_n(\hat{\mathbf{v}}_n(s), s)\| > \epsilon \right\} \\ &= \mathbb{P} \left\{ \sup_{s \in [\epsilon, 1]} \|\mathbf{V}_n(\hat{\mathbf{v}}_n(s), s) - \bar{\mathbf{V}}_n(\hat{\mathbf{v}}_n(s), s)\| > \epsilon, \sup_{s \in [\epsilon, 1]} \|\hat{\mathbf{v}}_n(s)\| \leq K \right\} \\ & \quad + \mathbb{P} \left\{ \sup_{s \in [\epsilon, 1]} \|\mathbf{V}_n(\hat{\mathbf{v}}_n(s), s) - \bar{\mathbf{V}}_n(\hat{\mathbf{v}}_n(s), s)\| > \epsilon, \sup_{s \in [\epsilon, 1]} \|\hat{\mathbf{v}}_n(s)\| > K \right\} \\ &\leq \mathbb{P} \left\{ \sup_{s \in [\epsilon, 1]} \sup_{\|\mathbf{v}\| \leq K} \|\mathbf{V}_n(\mathbf{v}, s) - \bar{\mathbf{V}}_n(\mathbf{v}, s)\| > \epsilon, \sup_{s \in [\epsilon, 1]} \|\hat{\mathbf{v}}_n(s)\| \leq K \right\} + \mathbb{P} \left\{ \sup_{s \in [\epsilon, 1]} \|\hat{\mathbf{v}}_n(s)\| > K \right\} \\ &\leq \mathbb{P} \left\{ \sup_{s \in [\epsilon, 1]} \sup_{\|\mathbf{v}\| \leq K} \|\mathbf{V}_n(\mathbf{v}, s) - \bar{\mathbf{V}}_n(\mathbf{v}, s)\| > \epsilon \right\} + \mathbb{P} \left\{ \sup_{s \in [\epsilon, 1]} \|\hat{\mathbf{v}}_n(s)\| > K \right\} \\ &= o(1) + o(1) \end{aligned}$$

as $n \rightarrow \infty$, followed by $K \rightarrow \infty$, from Lemma 8 and Theorem 1. Combining the above two displays implies that, as $n \rightarrow \infty$,

$$\begin{aligned} \mathbf{V}_n(\hat{\mathbf{v}}_n(s), s) &= \bar{\mathbf{V}}_n(\hat{\mathbf{v}}_n(s), s) + [\mathbf{V}_n(\hat{\mathbf{v}}_n(s), s) - \bar{\mathbf{V}}_n(\hat{\mathbf{v}}_n(s), s)] \\ &= \bar{\mathbf{V}}_n(\hat{\mathbf{v}}_n(s), s) + o_{\mathbb{P}}(1) \\ &\xrightarrow{d} \frac{1}{1-\gamma} \mathbf{D}_* \boldsymbol{\Omega}^{1/2} \mathbf{W}(s) \quad \text{in } D_m[\epsilon, 1], \end{aligned}$$

which is the claimed result. □

A.5 Proofs of Lemmas 7–8

The proof of Lemma 7 requires the following preliminary result.

Lemma 9. Under the assumptions of Theorem 2 it holds for any $\mathbf{v} \in \mathbb{R}^m$ that, as $n \rightarrow \infty$,

$$\begin{aligned} & \mathbb{E}_t \left[\left(\mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t < \varepsilon_{t\tau} \leq 0\}} - \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} \right) \xi_{t\tau} \right] \\ &= -\frac{1}{2} \frac{F^{\leftarrow}(\tau)}{n\tau} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} (\mathbf{v}' \mathbf{X}_t)^2 [1 + o(1)] + F^{\leftarrow}(\tau) \sqrt{\frac{1-\tau}{n\tau}} \mathbf{v}' \mathbf{X}_t \left[\frac{1}{1-\gamma} + o(1) \right] \\ & \quad + o(1) F^{\leftarrow}(\tau) \left[\frac{\gamma}{1-\gamma} + o(1) \right] h^\gamma(\mathbf{X}_t) \end{aligned}$$

uniformly in $t \in \mathbb{N}$.

Proof: Use (1)–(2) to rewrite the conditional expectation as

$$\begin{aligned} & \mathbb{E}_t \left[\left(\mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t < \varepsilon_{t\tau} \leq 0\}} - \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} \right) \xi_{t\tau} \right] \\ &= \mathbb{E}_t \left[\left(\mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t < \varepsilon_{t\tau} \leq 0\}} - \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} \right) \{ \mathbf{X}'_t (\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau}) + \varepsilon_{t\tau} \} \right] \\ &= \mathbb{E}_t \left[\mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t < \varepsilon_{t\tau} \leq 0\}} \{ \varepsilon_{t\tau} - a_n \mathbf{v}' \mathbf{X}_t + \mathbf{X}'_t (\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau} + a_n \mathbf{v}) \} \right] \\ & \quad - \mathbb{E}_t \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} \{ \varepsilon_{t\tau} - a_n \mathbf{v}' \mathbf{X}_t + \mathbf{X}'_t (\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau} + a_n \mathbf{v}) \} \right] \\ &= \left\{ \mathbb{E}_t \left[\mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t < \varepsilon_{t\tau} \leq 0\}} \{ \varepsilon_{t\tau} - a_n \mathbf{v}' \mathbf{X}_t \} \right] + \mathbb{E}_t \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} \{ a_n \mathbf{v}' \mathbf{X}_t - \varepsilon_{t\tau} \} \right] \right\} \\ & \quad + \mathbb{E}_t \left[\mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t < \varepsilon_{t\tau} \leq 0\}} \right] \mathbf{X}'_t (\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau} + a_n \mathbf{v}) \\ & \quad - \mathbb{E}_t \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} \right] \mathbf{X}'_t (\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau} + a_n \mathbf{v}) \\ &=: A_{2t\tau} + B_{2t\tau} - C_{2t\tau}. \end{aligned} \tag{45}$$

Use Lemma 3 to deduce that

$$\begin{aligned} A_{2t\tau} &= \mathbb{E}_t \left[\mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t < \varepsilon_{t\tau} \leq 0\}} \{ \varepsilon_{t\tau} - a_n \mathbf{v}' \mathbf{X}_t \} \right] + \mathbb{E}_t \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} \{ a_n \mathbf{v}' \mathbf{X}_t - \varepsilon_{t\tau} \} \right] \\ &= \frac{1}{2} \frac{1-\tau}{F^{\leftarrow}(\tau)} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} [1 + o(1)] (a_n \mathbf{v}' \mathbf{X}_t)^2 + (1-\tau) a_n \mathbf{v}' \mathbf{X}_t o(k^{-1/2}) \\ &= \frac{1}{2} \frac{F^{\leftarrow}(\tau)}{n\tau} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} [1 + o(1)] (\mathbf{v}' \mathbf{X}_t)^2 + \frac{F^{\leftarrow}(\tau)}{n} \mathbf{v}' \mathbf{X}_t o(1). \end{aligned} \tag{46}$$

Moreover,

$$\begin{aligned} B_{2t\tau} &= \mathbb{E}_t \left[\mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t < \varepsilon_{t\tau} \leq 0\}} \right] \mathbf{X}'_t (\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau} + a_n \mathbf{v}) \\ &= \mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t < 0\}} \left[F_{\varepsilon_t | \mathbf{X}_t}(0) - F_{\varepsilon_t | \mathbf{X}_t}(a_n \mathbf{v}' \mathbf{X}_t) \right] \mathbf{X}'_t (\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau} + a_n \mathbf{v}) \\ &= -\mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t < 0\}} \left[F_{\varepsilon_t | \mathbf{X}_t}(a_n \mathbf{v}' \mathbf{X}_t) - \tau \right] \mathbf{X}'_t (\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau} + a_n \mathbf{v}) \end{aligned}$$

$$\begin{aligned}
&= -\mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t < 0\}} [F_t(F_t^{\leftarrow}(\tau) + a_n \mathbf{v}' \mathbf{X}_t) - \tau] \mathbf{X}_t'(\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau} + a_n \mathbf{v}) \\
&= -\mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t < 0\}} (\tau - 1) \left[\frac{1 - F_t(F_t^{\leftarrow}(\tau) + a_n \mathbf{v}' \mathbf{X}_t)}{1 - \tau} - 1 \right] \mathbf{X}_t'(\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau} + a_n \mathbf{v}) \\
&= -(1 - \tau) \mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t < 0\}} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} \cdot \frac{a_n \mathbf{v}' \mathbf{X}_t}{F^{\leftarrow}(\tau)} [1 + o(1)] \mathbf{X}_t'(\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau} + a_n \mathbf{v}) \\
&\quad + (1 - \tau) \mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t < 0\}} O\left(A_1(1/[1 - \tau])\right) \mathbf{X}_t'(\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau} + a_n \mathbf{v}), \tag{47}
\end{aligned}$$

where we used (16) in the fourth step, and Lemma 2 in the final step. Similarly,

$$\begin{aligned}
C_{2t\tau} &= \mathbb{E}_t[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}}] \mathbf{X}_t'(\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau} + a_n \mathbf{v}) \\
&= (1 - \tau) \mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t > 0\}} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} \cdot \frac{a_n \mathbf{v}' \mathbf{X}_t}{F^{\leftarrow}(\tau)} [1 + o(1)] \mathbf{X}_t'(\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau} + a_n \mathbf{v}) \\
&\quad + (1 - \tau) \mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t > 0\}} O\left(A_1(1/[1 - \tau])\right) \mathbf{X}_t'(\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau} + a_n \mathbf{v}). \tag{48}
\end{aligned}$$

Combining (47)–(48), we obtain that

$$\begin{aligned}
B_{2t\tau} - C_{2t\tau} &= -\frac{1 - \tau}{F^{\leftarrow}(\tau)} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} a_n \mathbf{v}' \mathbf{X}_t [1 + o(1)] \mathbf{X}_t'(\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau} + a_n \mathbf{v}) \\
&\quad + (1 - \tau) O\left(A_1(1/[1 - \tau])\right) \mathbf{X}_t'(\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau} + a_n \mathbf{v}).
\end{aligned}$$

This expression may be further simplified by recalling from the proof of Lemma 6 that $Q_{t\tau} = Q_\tau(Y_t | \mathbf{X}_t) = \mathbf{X}_t' \boldsymbol{\alpha}_{0\tau}$ and $\text{ES}_{t\tau} = \text{ES}_\tau(Y_t | \mathbf{X}_t) = \mathbf{X}_t' \boldsymbol{\beta}_{0\tau}$, such that (also using (33))

$$\begin{aligned}
\mathbf{X}_t'(\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau}) &= Q_{t\tau} - \text{ES}_{t\tau} \\
&= -F^{\leftarrow}(\tau) \left[\frac{\gamma}{1 - \gamma} + o(1) \right] h^\gamma(\mathbf{X}_t).
\end{aligned}$$

Hence,

$$\begin{aligned}
B_{2t\tau} - C_{2t\tau} &= -\frac{1 - \tau}{F^{\leftarrow}(\tau)} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} (a_n \mathbf{v}' \mathbf{X}_t)^2 [1 + o(1)] \\
&\quad + \frac{1 - \tau}{F^{\leftarrow}(\tau)} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} a_n \mathbf{v}' \mathbf{X}_t [1 + o(1)] F^{\leftarrow}(\tau) \left[\frac{\gamma}{1 - \gamma} + o(1) \right] h^\gamma(\mathbf{X}_t) \\
&\quad - (1 - \tau) O\left(A_1(1/[1 - \tau])\right) F^{\leftarrow}(\tau) \left[\frac{\gamma}{1 - \gamma} + o(1) \right] h^\gamma(\mathbf{X}_t) \\
&\quad + (1 - \tau) O\left(A_1(1/[1 - \tau])\right) a_n \mathbf{v}' \mathbf{X}_t \\
&= -\frac{F^{\leftarrow}(\tau)}{n\tau} \frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma} (\mathbf{v}' \mathbf{X}_t)^2 [1 + o(1)] + F^{\leftarrow}(\tau) \sqrt{\frac{1 - \tau}{n\tau}} \mathbf{v}' \mathbf{X}_t \left[\frac{1}{1 - \gamma} + o(1) \right]
\end{aligned}$$

$$\begin{aligned}
& + o(1)\sqrt{\frac{1-\tau}{n}}F^{\leftarrow}(\tau)\left[\frac{\gamma}{1-\gamma} + o(1)\right]h^\gamma(\mathbf{X}_t) + o(1)\frac{F^{\leftarrow}(\tau)}{n}\mathbf{v}'\mathbf{X}_t \\
& = -\frac{F^{\leftarrow}(\tau)}{n\tau}\frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma}(\mathbf{v}'\mathbf{X}_t)^2[1 + o(1)] + F^{\leftarrow}(\tau)\sqrt{\frac{1-\tau}{n\tau}}\mathbf{v}'\mathbf{X}_t\left[\frac{1}{1-\gamma} + o(1)\right] \\
& \quad + o(1)\sqrt{\frac{1-\tau}{n}}F^{\leftarrow}(\tau)\left[\frac{\gamma}{1-\gamma} + o(1)\right]h^\gamma(\mathbf{X}_t). \tag{49}
\end{aligned}$$

Plugging (46) and (49) into (45) gives that

$$\begin{aligned}
& \mathbb{E}_t\left[\left(\mathbb{1}_{\{a_n\mathbf{v}'\mathbf{X}_t < \varepsilon_{t\tau} \leq 0\}} - \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n\mathbf{v}'\mathbf{X}_t\}}\right)\xi_{t\tau}\right] \\
& = -\frac{1}{2}\frac{F^{\leftarrow}(\tau)}{n\tau}\frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma}(\mathbf{v}'\mathbf{X}_t)^2[1 + o(1)] + \frac{F^{\leftarrow}(\tau)}{n}\mathbf{v}'\mathbf{X}_t o(1) \\
& \quad + F^{\leftarrow}(\tau)\sqrt{\frac{1-\tau}{n\tau}}\mathbf{v}'\mathbf{X}_t\left[\frac{1}{1-\gamma} + o(1)\right] + o(1)F^{\leftarrow}(\tau)\left[\frac{\gamma}{1-\gamma} + o(1)\right]h^\gamma(\mathbf{X}_t),
\end{aligned}$$

from which the conclusion easily follows. \square

Proof of Lemma 7: By the LIE and Lemma 9, we get that

$$\begin{aligned}
\bar{\mathbf{V}}_n(\mathbf{v}, s) & = \frac{1}{F^{\leftarrow}(\tau)}\frac{1}{\sqrt{n(1-\tau)}}\sum_{t=1}^{\lfloor ns \rfloor}\mathbb{E}_t\left[\left(\mathbb{1}_{\{a_n\mathbf{v}'\mathbf{X}_t < \varepsilon_{t\tau} \leq 0\}} - \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n\mathbf{v}'\mathbf{X}_t\}}\right)\xi_{t\tau}\right]\mathbf{X}_t \\
& = -\left(\frac{1}{2}\frac{1}{n\tau}\frac{1}{\sqrt{n(1-\tau)}}\sum_{t=1}^{\lfloor ns \rfloor}\frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma}(\mathbf{v}'\mathbf{X}_t)^2[1 + o(1)]\mathbf{X}_t\right) \\
& \quad + \left(\sqrt{\frac{1-\tau}{n\tau}}\frac{1}{\sqrt{n(1-\tau)}}\sum_{t=1}^{\lfloor ns \rfloor}\left[\frac{1}{1-\gamma} + o(1)\right]\mathbf{X}_t\mathbf{X}_t'\right)\mathbf{v} \\
& \quad + o(1)\frac{1}{n}\sum_{t=1}^{\lfloor ns \rfloor}\left[\frac{\gamma}{1-\gamma} + o(1)\right]h^\gamma(\mathbf{X}_t) \\
& = -\left(\frac{1 + o(1)}{2\tau}\frac{1}{\sqrt{n(1-\tau)}}\frac{1}{n}\sum_{t=1}^{\lfloor ns \rfloor}\frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma}(\mathbf{v}'\mathbf{X}_t)^2\mathbf{X}_t\right) + \frac{1 + o(1)}{1-\gamma}\left(\frac{1}{n}\sum_{t=1}^{\lfloor ns \rfloor}\mathbf{X}_t\mathbf{X}_t'\right)\mathbf{v} \\
& \quad + o(1)\left[\frac{\gamma}{1-\gamma} + o(1)\right]\frac{1}{n}\sum_{t=1}^{\lfloor ns \rfloor}h^\gamma(\mathbf{X}_t) \\
& = o_{\mathbb{P}}(1) - [1 + o(1)]\frac{1}{1-\gamma}\left(\frac{1}{n}\sum_{t=1}^{\lfloor ns \rfloor}\mathbf{X}_t\mathbf{X}_t'\right)\mathbf{v}, \tag{50}
\end{aligned}$$

where we used in the final step that

$$\frac{1}{n}\sum_{t=1}^{\lfloor ns \rfloor}\frac{h^{-\gamma}(\mathbf{X}_t)}{\gamma}(\mathbf{v}'\mathbf{X}_t)^2\mathbf{X}_t = O_{\mathbb{P}}(1) \quad \text{and} \quad \frac{1}{n}\sum_{t=1}^{\lfloor ns \rfloor}h^\gamma(\mathbf{X}_t) = O_{\mathbb{P}}(1)$$

uniformly in $s \in [0, 1]$, due to the boundedness of \mathbf{X}_t , $h^{-\gamma}(\mathbf{X}_t)$ and $h^\gamma(\mathbf{X}_t)$ (from Assumptions 5–6). The conclusion now easily follows from (42) and (50). \square

The proof of Lemma 8 requires the following two Lemmas 10 and 11.

Lemma 10. *Under the assumptions of Theorem 2 it holds for any $\mathbf{v} \in \mathbb{R}^m$ and $r \in (0, 1/\gamma)$ that*

$$\begin{aligned}\mathbb{E}_t \left[\mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t < \varepsilon_{t\tau} \leq 0\}} |\xi_{t\tau}|^r \right] &\leq K \sqrt{\frac{1-\tau}{n}} [F^{\leftarrow}(\tau)]^r, \\ \mathbb{E}_t \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} |\xi_{t\tau}|^r \right] &\leq K \sqrt{\frac{1-\tau}{n}} [F^{\leftarrow}(\tau)]^r.\end{aligned}$$

Proof: We only prove the second claim of the lemma, as the first follows by similar arguments. Recall the definition of $Q_{t\tau}$ and $\text{ES}_{t\tau}$ from the proof of Lemma 6. Then, equating (1) and (2) gives

$$\xi_{t\tau} = \varepsilon_{t\tau} + \mathbf{X}'_t(\boldsymbol{\alpha}_{0\tau} - \boldsymbol{\beta}_{0\tau}) = \varepsilon_{t\tau} - (\text{ES}_{t\tau} - Q_{t\tau}). \quad (51)$$

Therefore,

$$\begin{aligned}\mathbb{E}_t \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} |\xi_{t\tau}|^r \right] &= \mathbb{E}_t \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} |\varepsilon_{t\tau} - (\text{ES}_{t\tau} - Q_{t\tau})|^r \right] \\ &= \mathbb{E}_t \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} \{(\text{ES}_{t\tau} - Q_{t\tau}) - \varepsilon_{t\tau}\}^r \right],\end{aligned}$$

where the last equality holds for sufficiently large n , because $\varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t \leq KF^{\leftarrow}(\tau)/\sqrt{n\tau(1-\tau)}$ (by boundedness of \mathbf{X}_t) grows more slowly than $\text{ES}_{t\tau} - Q_{t\tau} \geq \delta F^{\leftarrow}(\tau)$ for some $\delta > 0$ (by (33) and Assumption 6 (ii)). Hence,

$$\begin{aligned}\mathbb{E}_t \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} |\xi_{t\tau}|^r \right] &\leq (\text{ES}_{t\tau} - Q_{t\tau})^r \mathbb{E}_t \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} \right] \\ &= (\text{ES}_{t\tau} - Q_{t\tau})^r \left[\mathbb{P}_t \{ \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t \} - \mathbb{P}_t \{ \varepsilon_{t\tau} \leq 0 \} \right] \\ &= (\text{ES}_{t\tau} - Q_{t\tau})^r \left[F_t(F_t^{\leftarrow}(\tau) + a_n \mathbf{v}' \mathbf{X}_t) - \tau \right] \\ &= (\text{ES}_{t\tau} - Q_{t\tau})^r (\tau - 1) \left[\frac{1 - F_t(F_t^{\leftarrow}(\tau) + a_n \mathbf{v}' \mathbf{X}_t)}{1 - \tau} - 1 \right] \\ &= (\text{ES}_{t\tau} - Q_{t\tau})^r (1 - \tau) \left[\frac{h^{-\gamma}(\mathbf{X}_t) a_n \mathbf{v}' \mathbf{X}_t}{\gamma F^{\leftarrow}(\tau)} [1 + o(1)] + O\left(A_1(1/[1 - \tau])\right) \right] \\ &\leq K [F^{\leftarrow}(\tau)]^r (1 - \tau) \left[K \frac{1}{\sqrt{n\tau(1-\tau)}} + o(1/\sqrt{n(1-\tau)}) \right] \\ &\leq K \sqrt{\frac{1-\tau}{n}} [F^{\leftarrow}(\tau)]^r,\end{aligned}$$

where we used (16) in the third step, Lemma 2 in the fifth step, and (32) in the penultimate step. \square

Lemma 11. *Under the assumptions of Theorem 2 it holds for any $\mathbf{v} \in \mathbb{R}^m$ that, as $n \rightarrow \infty$,*

$$\sup_{s \in [0,1]} \|\mathbf{V}_n(\mathbf{v}, s) - \bar{\mathbf{V}}_n(\mathbf{v}, s)\| = o_{\mathbb{P}}(1).$$

Proof: Decompose

$$\begin{aligned} \mathbf{V}_n(\mathbf{v}, s) - \bar{\mathbf{V}}_n(\mathbf{v}, s) &= [\mathbf{V}_{1n}(\mathbf{v}, s) - \bar{\mathbf{V}}_{1n}(\mathbf{v}, s)] - [\mathbf{V}_{2n}(\mathbf{v}, s) - \bar{\mathbf{V}}_{2n}(\mathbf{v}, s)] \\ &:= \sum_{t=1}^{\lfloor ns \rfloor} [\boldsymbol{\nu}_{1t}(\mathbf{v}) - \bar{\boldsymbol{\nu}}_{1t}(\mathbf{v})] - \sum_{t=1}^{\lfloor ns \rfloor} [\boldsymbol{\nu}_{2t}(\mathbf{v}) - \bar{\boldsymbol{\nu}}_{2t}(\mathbf{v})], \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\nu}_{1t}(\mathbf{v}) &:= \frac{1}{F^{\leftarrow}(\tau)} \frac{1}{\sqrt{n(1-\tau)}} \mathbb{1}_{\{a_n \mathbf{v}' \mathbf{X}_t < \varepsilon_{t\tau} \leq 0\}} \xi_{t\tau} \mathbf{X}_t, & \bar{\boldsymbol{\nu}}_{1t}(\mathbf{v}) &:= \mathbb{E}_t[\boldsymbol{\nu}_{1t}(\mathbf{v})], \\ \boldsymbol{\nu}_{2t}(\mathbf{v}) &:= \frac{1}{F^{\leftarrow}(\tau)} \frac{1}{\sqrt{n(1-\tau)}} \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} \xi_{t\tau} \mathbf{X}_t, & \bar{\boldsymbol{\nu}}_{2t}(\mathbf{v}) &:= \mathbb{E}_t[\boldsymbol{\nu}_{2t}(\mathbf{v})]. \end{aligned}$$

In view of this decomposition, it suffices to show that for $i = 1, 2$,

$$\sup_{s \in [0,1]} \|\mathbf{V}_{in}(\mathbf{v}, s) - \bar{\mathbf{V}}_{in}(\mathbf{v}, s)\| = o_{\mathbb{P}}(1).$$

We do so only for $i = 2$, as the proof for $i = 1$ is almost identical.

If we denote by X_t a generic component of \mathbf{X}_t and (slightly overloading notation introduced in the proof of Proposition 2) define

$$\nu_{2t}(\mathbf{v}) := \frac{1}{F^{\leftarrow}(\tau)} \frac{1}{\sqrt{n(1-\tau)}} \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} \xi_{t\tau} X_t, \quad \bar{\nu}_{2t}(\mathbf{v}) := \mathbb{E}_t[\nu_{2t}(\mathbf{v})],$$

then we must simply show that

$$\sup_{s \in [0,1]} \left| \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(\mathbf{v}) - \bar{\nu}_{2t}(\mathbf{v})] \right| = o_{\mathbb{P}}(1).$$

For this, we proceed similarly as in the proof of (18). First, we apply the c_r -inequality to deduce that

$$\mathbb{E}[|\nu_{2t}(\mathbf{v}) - \bar{\nu}_{2t}(\mathbf{v})|^r] \leq 2^{r-1} \left\{ \mathbb{E}[|\nu_{2t}(\mathbf{v})|^r] + \mathbb{E}[|\bar{\nu}_{2t}(\mathbf{v})|^r] \right\}. \quad (52)$$

For $\mathbb{E}[|\nu_{2t}(\mathbf{v})|^r]$ we obtain that

$$\begin{aligned}
\mathbb{E}[|\nu_{2t}(\mathbf{v})|^r] &= \frac{1}{[F^{\leftarrow}(\tau)]^r} \frac{1}{[n(1-\tau)]^{r/2}} \mathbb{E}[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} |\xi_{t\tau}|^r |X_t|^r] \\
&= \frac{1}{[F^{\leftarrow}(\tau)]^r} \frac{1}{[n(1-\tau)]^{r/2}} \mathbb{E}[|X_t|^r \mathbb{E}_t\{\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} |\xi_{t\tau}|^r\}] \\
&\leq K \frac{1}{[F^{\leftarrow}(\tau)]^r} \frac{1}{[n(1-\tau)]^{r/2}} \sqrt{\frac{1-\tau}{n}} [F^{\leftarrow}(\tau)]^r \mathbb{E}[|X_t|^r] \\
&\leq K \frac{(1-\tau)^{1/2-r/2}}{n^{1/2+r/2}},
\end{aligned}$$

where the second step follows from the LIE, and the third step from Lemma 10.

For the other expectation on the right-hand side of (52), similar arguments show that

$$\begin{aligned}
\mathbb{E}[|\bar{\nu}_{2t}(\mathbf{v})|^r] &= \frac{1}{[F^{\leftarrow}(\tau)]^r} \frac{1}{[n(1-\tau)]^{r/2}} \mathbb{E}[|\mathbb{E}_t\{\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} \xi_{t\tau} X_t\}|^r] \\
&\leq \frac{1}{[F^{\leftarrow}(\tau)]^r} \frac{1}{[n(1-\tau)]^{r/2}} \mathbb{E}[|X_t|^r |\mathbb{E}_t\{\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} \xi_{t\tau}\}|^r] \\
&\leq K \frac{1}{[F^{\leftarrow}(\tau)]^r} \frac{1}{[n(1-\tau)]^{r/2}} \left(\frac{1-\tau}{n}\right)^{r/2} [F^{\leftarrow}(\tau)]^r \\
&\leq Kn^{-r},
\end{aligned}$$

where we exploited Lemma 10 in the third step. Plugging the previous two results into (52), we deduce that

$$\mathbb{E}[|\nu_{2t}(\mathbf{v}) - \bar{\nu}_{2t}(\mathbf{v})|^r] \leq K \frac{(1-\tau)^{1/2-r/2}}{n^{1/2+r/2}} = K \frac{1-\tau}{[n(1-\tau)]^{(r+1)/2}}, \quad (53)$$

as in (19).

With these results at hand, we once more slightly overload notation introduced in the proof of Proposition 2 and decompose

$$\sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(\mathbf{v}) - \bar{\nu}_{2t}(\mathbf{v})] = \sum_{j=1}^{m_n(s)} \nu_{2j}^B + \sum_{j=1}^{m_n(s)} \nu_{2j}^S + R_n^{\nu_2}(s),$$

where

$$\begin{aligned}
\nu_{2j}^B &= \sum_{t \in B_j} [\nu_{2t}(\mathbf{v}) - \bar{\nu}_{2t}(\mathbf{v})], \\
\nu_{2j}^S &= \sum_{t \in S_j} [\nu_{2t}(\mathbf{v}) - \bar{\nu}_{2t}(\mathbf{v})],
\end{aligned}$$

$$R_n^{\nu_2}(s) = \sum_{t \in B_{m_n(s)+1}} [\nu_{2t}(\mathbf{v}) - \bar{\nu}_{2t}(\mathbf{v})].$$

Observe that by boundedness of \mathbf{X}_t , (33) and (51),

$$\begin{aligned} |\nu_{2t}(\mathbf{v})| &\leq K \frac{1}{F^{\leftarrow}(\tau)} \frac{1}{\sqrt{n(1-\tau)}} \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \mathbf{v}' \mathbf{X}_t\}} |\varepsilon_{t\tau} - (\mathbb{E}S_{t\tau} - Q_{t\tau})| \\ &\leq K \frac{1}{n(1-\tau)} + K \frac{1}{\sqrt{n(1-\tau)}} \\ &\leq K \frac{1}{\sqrt{n(1-\tau)}}. \end{aligned}$$

Recalling that $|B_{m_n(s)+1}| \leq r_n + l_n - 1$, it therefore follows that

$$0 \leq \sup_{s \in [0,1]} |R_n^{\nu_2}(s)| \stackrel{\text{a.s.}}{\leq} K \frac{r_n}{\sqrt{n(1-\tau)}} = K \frac{r_n}{\sqrt{k}} = o(1). \quad (54)$$

By the same arguments as in the proof of Proposition 1, $\sum_{j=1}^{m_n(\cdot)} \nu_{2j}^B$ has the same asymptotic limit as $\sum_{j=1}^{m_n(\cdot)} \tilde{\nu}_{2j}^B$, where $\{\tilde{\nu}_{2j}^B\}_j$ is i.i.d. with $\tilde{\nu}_{2j}^B \stackrel{d}{=} \nu_{2j}^B$. By Kolmogorov's inequality,

$$\begin{aligned} \mathbb{P} \left\{ \sup_{s \in [0,1]} \left| \sum_{j=1}^{m_n(s)} \tilde{\nu}_{2j}^B \right| > \varepsilon \right\} &= \mathbb{P} \left\{ \max_{1 \leq \ell \leq m_n(1)} \left| \sum_{j=1}^{\ell} \tilde{\nu}_{2j}^B \right| > \varepsilon \right\} \\ &\leq \varepsilon^{-2} \sum_{j=1}^{m_n(1)} \text{Var}(\tilde{\nu}_{2j}^B) \\ &= \varepsilon^{-2} m_n(1) \text{Var}(\tilde{\nu}_{21}^B). \end{aligned} \quad (55)$$

Since the same bounds are in force in (19) and (53), we obtain similarly as in the proof of Proposition 1 that $m_n(1) \text{Var}(\tilde{\nu}_{21}^B) = o(1)$, such that, as $n \rightarrow \infty$,

$$\sup_{s \in [0,1]} \left| \sum_{j=1}^{m_n(s)} \nu_{2j}^B \right| = o_{\mathbb{P}}(1).$$

Analogous arguments show that

$$\sup_{s \in [0,1]} \left| \sum_{j=1}^{m_n(s)} \nu_{2j}^S \right| = o_{\mathbb{P}}(1).$$

Combining these two convergences with that in (54) yields the conclusion. \square

Proof of Lemma 8: The proof is similar to that of Lemma 10 in Hoga and Schulz (2025). It is most transparent for scalar-valued $\mathbf{X}_t = X_t$. To reflect this in our notation, we now write

$V_n(v, s) - \bar{V}_n(v, s)$ in place of $\mathbf{V}_n(\mathbf{v}, s) - \bar{\mathbf{V}}_n(\mathbf{v}, s)$ and so on. We omit the extension to vector-valued \mathbf{X}_t , as it is only notationally more complicated. Write

$$\begin{aligned} V_n(v, s) - \bar{V}_n(v, s) &= [V_{1n}(v, s) - \bar{V}_{1n}(v, s)] - [V_{2n}(v, s) - \bar{V}_{2n}(v, s)] \\ &:= \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{1t}(v) - \bar{\nu}_{1t}(v)] - \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(v) - \bar{\nu}_{2t}(v)], \end{aligned}$$

where, similarly as in the proof of Lemma 11,

$$\begin{aligned} \nu_{1t}(v) &:= \frac{1}{F^{\leftarrow}(\tau)} \frac{1}{\sqrt{n(1-\tau)}} \mathbb{1}_{\{a_n v X_t < \varepsilon_{t\tau} \leq 0\}} \xi_{t\tau} X_t, & \bar{\nu}_{1t}(v) &:= \mathbb{E}_t[\nu_{1t}(v)], \\ \nu_{2t}(v) &:= \frac{1}{F^{\leftarrow}(\tau)} \frac{1}{\sqrt{n(1-\tau)}} \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n v X_t\}} \xi_{t\tau} X_t, & \bar{\nu}_{2t}(v) &:= \mathbb{E}_t[\nu_{2t}(v)]. \end{aligned}$$

Therefore, it suffices to show that

$$\sup_{|v| \leq K} \sup_{s \in [0,1]} \left| \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{it}(v) - \bar{\nu}_{it}(v)] \right| = o_{\mathbb{P}}(1)$$

for $i = 1, 2$. Once again, we only do so for $i = 2$, as the proof for $i = 1$ is virtually identical. Fix $\rho > 0$ and assume for ease of presentation that $2K/\rho$ is an integer. Then,

$$\begin{aligned} &\sup_{|v| \leq K} \sup_{s \in [0,1]} \left| \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(v) - \bar{\nu}_{2t}(v)] \right| \\ &\leq \max_{j \in \mathbb{Z}} \sup_{s \in [0,1]} \left| \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(j\rho) - \bar{\nu}_{2t}(j\rho)] \right| \\ &\quad + \sup_{|v_1 - v_2| < \rho} \sup_{s \in [0,1]} \left| \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(v_1) - \bar{\nu}_{2t}(v_1)] - \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(v_2) - \bar{\nu}_{2t}(v_2)] \right|, \end{aligned}$$

where the maximum is taken over the j 's with $j\rho \in [-K, K]$. This implies that

$$\begin{aligned} &\mathbb{P} \left\{ \sup_{|v| \leq K} \sup_{s \in [0,1]} \left| \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(v) - \bar{\nu}_{2t}(v)] \right| > \varepsilon \right\} \\ &\leq \mathbb{P} \left\{ \max_{j \in \mathbb{Z}} \sup_{s \in [0,1]} \left| \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(j\rho) - \bar{\nu}_{2t}(j\rho)] \right| > \frac{\varepsilon}{2} \right\} \\ &\quad + \mathbb{P} \left\{ \sup_{|v_1 - v_2| < \rho} \sup_{s \in [0,1]} \left| \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(v_1) - \bar{\nu}_{2t}(v_1)] - \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(v_2) - \bar{\nu}_{2t}(v_2)] \right| > \frac{\varepsilon}{2} \right\}. \quad (56) \end{aligned}$$

By subadditivity and Lemma 11,

$$\begin{aligned} & \mathbb{P} \left\{ \max_{j \in \mathbb{Z}} \sup_{s \in [0,1]} \left| \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(j\rho) - \bar{\nu}_{2t}(j\rho)] \right| > \frac{\varepsilon}{2} \right\} \\ & \leq \sum_{j: j\rho \in [-K, K]} \mathbb{P} \left\{ \sup_{s \in [0,1]} \left| \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(j\rho) - \bar{\nu}_{2t}(j\rho)] \right| > \frac{\varepsilon}{2} \right\} = o(1). \end{aligned}$$

Therefore, we only have to show that the final right-hand side term in (56) converges to zero.

Note that for $-K \leq v_1, v_2 \leq K$,

$$\begin{aligned} & \sup_{|v_1 - v_2| < \rho} \sup_{s \in [0,1]} \left| \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(v_1) - \bar{\nu}_{2t}(v_1)] - \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(v_2) - \bar{\nu}_{2t}(v_2)] \right| \\ & = \sup_{|v_1 - v_2| < \rho} \sup_{s \in [0,1]} \left| \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(v_1) - \nu_{2t}(v_2)] - \sum_{t=1}^{\lfloor ns \rfloor} [\bar{\nu}_{2t}(v_1) - \bar{\nu}_{2t}(v_2)] \right| \\ & \leq \sup_{|v_1 - v_2| < \rho} \sum_{t=1}^n |\nu_{2t}(v_1) - \nu_{2t}(v_2)| + \sup_{|v_1 - v_2| < \rho} \sum_{t=1}^n |\bar{\nu}_{2t}(v_1) - \bar{\nu}_{2t}(v_2)| \\ & =: A_{3n} + B_{3n}. \end{aligned} \tag{57}$$

For B_{3n} , we obtain the bound

$$B_{3n} \leq \max_{\ell} \frac{1}{F^{\leftarrow}(\tau)} \frac{1}{\sqrt{n(1-\tau)}} \sum_{t=1}^n |X_t| \mathbb{E}_t \left[\left(\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n(\ell+2)\rho X_t\}} - \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \ell \rho X_t\}} \right) |\xi_{t\tau}| \right], \tag{58}$$

where the maximum is taken over the integers ℓ , such that $[\ell\rho, (\ell+2)\rho] \subset [-K, K]$. Now, from arguments in the proof of Lemma 10,

$$\mathbb{E}_t \left[\left(\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n(\ell+2)\rho X_t\}} - \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \ell \rho X_t\}} \right) |\xi_{t\tau}| \right] \leq K \sqrt{\frac{1-\tau}{n}} F^{\leftarrow}(\tau) 2\rho$$

with $K > 0$ independent of ℓ . Plugging this into (58) yields that

$$B_{3n} \leq K\rho \frac{1}{n} \sum_{t=1}^n |X_t| = \rho O_{\mathbb{P}}(1), \tag{59}$$

where $\frac{1}{n} \sum_{t=1}^n |X_t| = O_{\mathbb{P}}(1)$ holds by boundedness of the \mathbf{X}_t . Therefore,

$$\lim_{\rho \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left\{ \sup_{|v_1 - v_2| < \rho} \sum_{t=1}^n |\bar{\nu}_{2t}(v_1) - \bar{\nu}_{2t}(v_2)| > \frac{\varepsilon}{2} \right\} = 0. \tag{60}$$

For A_{3n} , we obtain the bound

$$A_{3n} \leq \max_{\ell} \frac{1}{F^{\leftarrow}(\tau)} \frac{1}{\sqrt{n(1-\tau)}} \sum_{t=1}^n \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n(\ell+2)\rho X_t\}} |\xi_{t\tau} X_t| - \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \ell \rho X_t\}} |\xi_{t\tau} X_t| \right], \tag{61}$$

where the maximum is likewise taken over the integers ℓ , such that $[\ell\rho, (\ell+2)\rho] \subset [-K, K]$. Similar arguments used in the proof of Lemma 11 also give that for each such ℓ ,

$$\begin{aligned} & \frac{1}{F^{\leftarrow}(\tau)} \frac{1}{\sqrt{n(1-\tau)}} \sum_{t=1}^n \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n(\ell+2)\rho X_t\}} |\xi_{t\tau} X_t| - \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \ell \rho X_t\}} |\xi_{t\tau} X_t| \right] \\ & - \left\{ \mathbb{E}_t \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n(\ell+2)\rho X_t\}} |\xi_{t\tau} X_t| \right] - \mathbb{E}_t \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \ell \rho X_t\}} |\xi_{t\tau} X_t| \right] \right\} = o_{\mathbb{P}}(1). \end{aligned} \quad (62)$$

By (58) and (59),

$$\frac{1}{F^{\leftarrow}(\tau)} \frac{1}{\sqrt{n(1-\tau)}} \sum_{t=1}^n \left\{ \mathbb{E}_t \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n(\ell+2)\rho X_t\}} |\xi_{t\tau} X_t| \right] - \mathbb{E}_t \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \ell \rho X_t\}} |\xi_{t\tau} X_t| \right] \right\} = o_{\mathbb{P}}(1),$$

as $n \rightarrow \infty$, followed by $\rho \downarrow 0$. From this and (62),

$$\frac{1}{F^{\leftarrow}(\tau)} \frac{1}{\sqrt{n(1-\tau)}} \sum_{t=1}^n \left[\mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n(\ell+2)\rho X_t\}} |\xi_{t\tau} X_t| - \mathbb{1}_{\{0 < \varepsilon_{t\tau} \leq a_n \ell \rho X_t\}} |\xi_{t\tau} X_t| \right] = o_{\mathbb{P}}(1),$$

as $n \rightarrow \infty$, followed by $\rho \downarrow 0$, such that by (61),

$$\lim_{\rho \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left\{ \sup_{|v_1 - v_2| < \rho} \sum_{t=1}^n |\nu_{2t}(v_1) - \nu_{2t}(v_2)| > \frac{\varepsilon}{2} \right\} = 0.$$

From this, (57) and (60),

$$\sup_{|v_1 - v_2| < \rho} \sup_{s \in [0,1]} \left| \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(v_1) - \bar{\nu}_{2t}(v_1)] - \sum_{t=1}^{\lfloor ns \rfloor} [\nu_{2t}(v_2) - \bar{\nu}_{2t}(v_2)] \right| = o_{\mathbb{P}}(1),$$

as $n \rightarrow \infty$, followed by $\rho \downarrow 0$. Overall, the conclusion follows. \square

A.6 Proof of Theorem 3

Proof of Theorem 3: From (30) and Lemma 5, as $n \rightarrow \infty$,

$$\frac{\text{ES}_{\tau}(\mathbf{x})}{F^{\leftarrow}(\tau)} = \frac{\text{ES}_{\tau}(\mathbf{x})}{\text{ES}_{\tau}} \frac{\text{ES}_{\tau}}{F^{\leftarrow}(\tau)} \rightarrow \frac{1}{1-\gamma} h^{\gamma}(\mathbf{x}). \quad (63)$$

Observing that $\text{ES}_{\tau}(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}_{0\tau}$, we obtain

$$\begin{aligned} \sqrt{n\tau(1-\tau)} \left[\frac{\widehat{\text{ES}}_{\tau}(\mathbf{x})}{\text{ES}_{\tau}(\mathbf{x})} - 1 \right] &= \frac{1}{\text{ES}_{\tau}(\mathbf{x})} \sqrt{n\tau(1-\tau)} [\mathbf{x}'\widehat{\boldsymbol{\beta}} - \mathbf{x}'\boldsymbol{\beta}_{0\tau}] \\ &= \frac{F^{\leftarrow}(\tau)}{\text{ES}_{\tau}(\mathbf{x})} \mathbf{x}' \frac{\sqrt{n\tau(1-\tau)}}{F^{\leftarrow}(\tau)} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{0\tau}) \\ &\xrightarrow{d} (1-\gamma) h^{-\gamma}(\mathbf{x}) \mathbf{x}' \boldsymbol{\Omega}_{*}^{1/2} \mathbf{W}_m(1), \end{aligned}$$

where we have used (4) in the first step, and Theorem 2 and (63) in the final step. \square

A.7 Proof of Theorem 4

Proof of Theorem 4: Define $l_n = (1 - \tau)/(1 - \bar{\tau})$. Recalling that $\tau = 1 - k/n$, it follows from Assumption 7 that $l_n \rightarrow \infty$, as $n \rightarrow \infty$. Decompose

$$\begin{aligned} \frac{\widetilde{\text{ES}}_{\bar{\tau}}(\mathbf{x})}{\text{ES}_{\bar{\tau}}(\mathbf{x})} &= \frac{1}{l_n^\gamma} \frac{\widetilde{\text{ES}}_{\bar{\tau}}(\mathbf{x})}{\widehat{\text{ES}}_{\bar{\tau}}(\mathbf{x})} \cdot \frac{\widehat{\text{ES}}_{\bar{\tau}}(\mathbf{x})}{\text{ES}_{\bar{\tau}}(\mathbf{x})} \cdot l_n^\gamma \frac{\text{ES}_{\bar{\tau}}(\mathbf{x})}{\text{ES}_{\tau}(\mathbf{x})} \\ &= l_n^{\widehat{\gamma} - \gamma} \cdot \frac{\widehat{\text{ES}}_{\bar{\tau}}(\mathbf{x})}{\text{ES}_{\bar{\tau}}(\mathbf{x})} \cdot l_n^\gamma \frac{\text{ES}_{\bar{\tau}}(\mathbf{x})}{\text{ES}_{\tau}(\mathbf{x})} \\ &=: A_{4n} \cdot B_{4n} \cdot C_{4n}. \end{aligned} \tag{64}$$

We consider these three terms separately.

From Theorem 3, as $n \rightarrow \infty$,

$$\sqrt{n\tau(1-\tau)}(B_{1n} - 1) \xrightarrow{d} (1-\gamma)h^{-\gamma}(\mathbf{x})\mathbf{x}'\boldsymbol{\Omega}_*^{1/2}\mathbf{W}_m(1),$$

such that, since $n(1-\tau) = k$,

$$B_{4n} = 1 + O_{\mathbb{P}}(k^{-1/2}). \tag{65}$$

For C_{4n} , we obtain that

$$C_{4n} = l_n^\gamma \frac{F^{\leftarrow}(\tau)}{F^{\leftarrow}(\bar{\tau})} \cdot \frac{\text{ES}_{\tau}/F^{\leftarrow}(\tau)}{\text{ES}_{\bar{\tau}}/F^{\leftarrow}(\bar{\tau})} \cdot \frac{\text{ES}_{\tau}(\mathbf{x})/\text{ES}_{\tau}}{\text{ES}_{\bar{\tau}}(\mathbf{x})/\text{ES}_{\bar{\tau}}}.$$

First note that, as in the proof of Theorem 4.3.8 in de Haan and Ferreira (2006) (which applies thanks to Assumption 6 (iii) and Theorem 2.3.9 in de Haan and Ferreira (2006)),

$$\lim_{n \rightarrow \infty} \frac{l_n^{-\gamma} \frac{F^{\leftarrow}(\bar{\tau})}{F^{\leftarrow}(\tau)} - 1}{A_1(n/k)} = -\frac{1}{\xi}.$$

Therefore,

$$l_n^\gamma \frac{F^{\leftarrow}(\tau)}{F^{\leftarrow}(\bar{\tau})} = 1 + O(A_1(n/k)) = 1 + o(k^{-1/2}). \tag{66}$$

It follows along similar lines as for (31) that $\text{ES}_{\bar{\tau}}/F^{\leftarrow}(\bar{\tau}) = \frac{1}{1-\gamma} + o(k^{-1/2})$. Finally, from Lemma 5

and Assumption 6,

$$\frac{\text{ES}_\tau(\mathbf{x})}{\text{ES}_\tau} = h^\gamma(\mathbf{x})[1 + o(k^{-1/2})] \quad \text{and} \quad \frac{\text{ES}_{\bar{\tau}}(\mathbf{x})}{\text{ES}_{\bar{\tau}}} = h^\gamma(\mathbf{x})[1 + o(k^{-1/2})].$$

This combined with (66) and (31) gives that

$$C_{4n} = 1 + o(k^{-1/2}). \quad (67)$$

For the final term A_{4n} in (64), we use the mean value theorem $f(x) - f(0) = f'(\xi)(x - 0)$ for some mean value ξ between 0 and x . We do so for the function $f(x) = l_n^x / \log(l_n)$ with derivative $f'(x) = l_n^x$ to obtain that

$$\begin{aligned} \frac{\sqrt{k}}{\log(l_n)} (l_n^{\hat{\gamma}} - 1) &= \sqrt{k} [f(\hat{\gamma}) - f(0)] \\ &= \sqrt{k} f'(\vartheta[\hat{\gamma} - \gamma]) (\hat{\gamma} - \gamma) \end{aligned}$$

for some (possibly stochastic) $\vartheta \in (0, 1)$. Assumptions 7–8 imply that $\sqrt{k}(\hat{\gamma} - \gamma) \xrightarrow{d} \Gamma$ and

$$\begin{aligned} f'(\vartheta[\hat{\gamma} - \gamma]) &= l_n^{\vartheta(\hat{\gamma} - \gamma)} \\ &= \exp \{ \vartheta(\hat{\gamma} - \gamma) \log(l_n) \} \\ &= \exp \{ O_{\mathbb{P}}(k^{-1/2}) \log(l_n) \} \\ &= \exp \left\{ O_{\mathbb{P}}(k^{-1/2}) \log(n[1 - \bar{\tau}]/k) \right\} \\ &= 1 + o_{\mathbb{P}}(1). \end{aligned}$$

Hence, we get that, as $n \rightarrow \infty$,

$$\frac{\sqrt{k}}{\log(l_n)} (A_{4n} - 1) = \frac{\sqrt{k}}{\log(l_n)} (l_n^{\hat{\gamma}} - 1) \xrightarrow{d} \Gamma. \quad (68)$$

Putting together (64) with (65), (67) and (68), the conclusion follows. \square

Appendix B: Critical Values

In Table B1, we collect quantiles of the limiting distribution $U(\mathbf{B})$ in Corollary 1, which may serve as critical values for our self-normalized tests. Here, we put $\mathbf{B} = \mathbf{I}_{m \times m}$ and $\epsilon = 0.2$.

Table B1: $(1 - \nu)$ -quantiles of the limiting distribution $U(\mathbf{B})$.

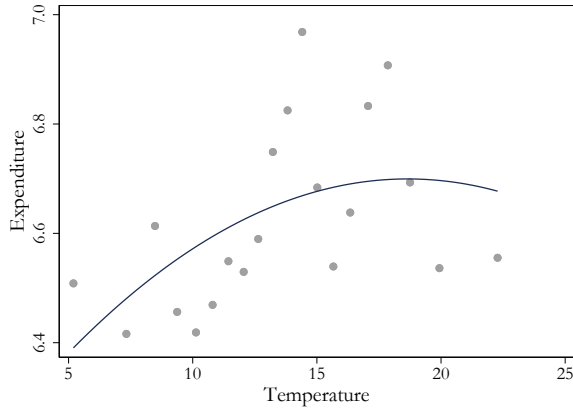
m	$1 - \nu$					
	80%	90%	95%	97.5%	99%	99.5%
1	17.74	34.06	55.64	82.60	126.35	165.75
2	54.99	90.92	133.67	183.41	258.85	323.67
3	109.77	168.36	234.31	308.08	416.17	506.51
4	181.49	265.18	356.61	455.60	597.95	714.09
5	269.72	380.71	498.98	624.16	801.67	943.75
6	374.04	514.05	660.13	813.38	1026.41	1196.59
7	494.01	664.86	840.80	1023.00	1274.00	1471.02
8	629.73	833.11	1039.65	1252.00	1541.31	1768.07
9	780.96	1018.44	1256.99	1499.44	1828.91	2084.83
10	947.30	1220.03	1491.44	1765.59	2134.91	2421.11

NOTE.— Results are for $\mathbf{B} = \mathbf{I}_{m \times m}$ for different values of m and fixed $\epsilon = 0.2$. The critical values have been computed based on 10,000,000 draws from the limiting distribution, where the Brownian motion was approximated on a grid of 100,000 equally spaced points in $[0, 1]$.

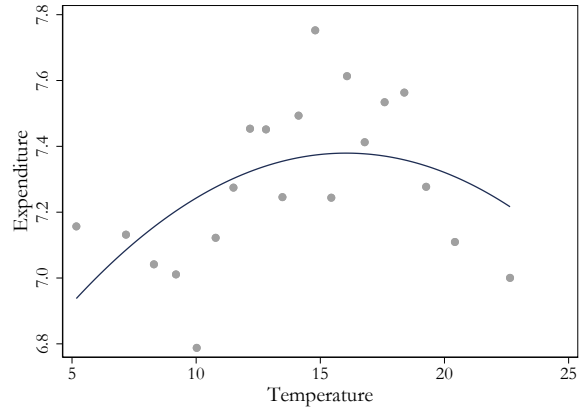
Appendix C: Supplementary Empirical Results

We examine the relationship between health care expenditure and temperature. Figure C1 presents binscatters of daily expenditure against current and lagged temperature. In constructing these plots, we condition on *Bundesland*, a time trend, month, and day-of-week fixed effects, so that the visualized relationship reflects deviations after accounting for these factors. All figures suggest a roughly quadratic relationship between expenditure and temperature, with the non-linearity being particularly pronounced in the three largest *Bundesländer* (Figure C1b). Given the strong serial correlation between current and lagged temperature (correlation coefficient 0.78), it is important to determine which specification of temperature variables is most appropriate, considering that a temperature shock on one day may affect spending on the following day.

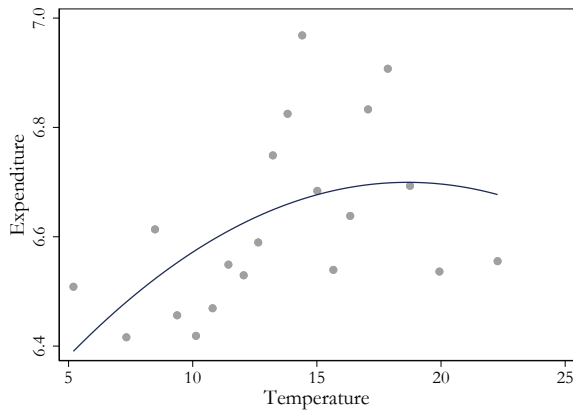
To formally assess the predictive power of temperature variables, we estimate a series of OLS regressions, reported in Tables C1–C2. Specifications include linear and quadratic terms of current and/or lagged temperature, plus additional weather controls (cloud coverage and vapor pressure). All regressions include *Bundesland*, time trends, as well as month and day-of-week fixed effects. F tests are reported at the bottom of each table: the first tests the joint significance of all temperature variables, while the second focuses on lagged temperature terms only. Across specifications, current temperature generally exhibits the strongest explanatory power, as reflected in adjusted R^2 , while



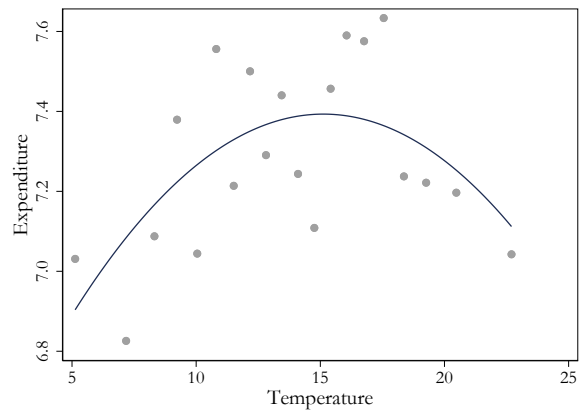
(a) Current: All Federal States



(b) Current: Three Largest



(c) Lagged: All Federal States



(d) Lagged: Three Largest

Figure C1: Expenditure-Temperature Relationship: Binscatter

NOTE.— Temperature is specified as daily max ambient temperature. All binscatters control for *Bundesland*, linear time trend, month, and day-of-week fixed effects.

F tests confirm that lagged temperature contributes little additional predictive power.

Table C1: OLS Results: Impact of Temperature

	(1)	(2)	(3)	(4)	(5)	(6)
Temperature	0.0391*** (0.012)	0.0569*** (0.016)	0.0757*** (0.018)	0.0295 (0.025)		
Temperature ²		-0.0008* (0.000)	-0.0009** (0.000)	0.0005 (0.001)		
Temperature lag			-0.0268** (0.011)	0.0274 (0.024)	-0.0028 (0.010)	0.0419*** (0.015)
Temperature lag ²				-0.0017** (0.001)		-0.0017*** (0.000)
Cloud coverage	0.0571*** (0.021)	0.0446** (0.022)	0.0417* (0.022)	0.0409* (0.022)	0.0092 (0.017)	-0.0007 (0.017)
Vapor pressure	-0.0359 (0.024)	-0.0290 (0.024)	-0.0156 (0.025)	-0.0092 (0.025)	0.0297 (0.020)	0.0350* (0.020)
<i>F</i> test all temp	10.1	6.7	6.3	6.3		
<i>p</i> value all temp	0.001	0.001	0.000	0.000		
<i>F</i> test lag temp			5.5	6.0	0.1	8.1
<i>p</i> value lag temp			0.020	0.002	0.781	0.000
Adj. <i>R</i> ²	0.2697	0.2697	0.2698	0.2699	0.2695	0.2698
<i>N</i>	40,896	40,896	40,896	40,896	40,896	40,896

NOTE.— Each column reports OLS estimates of daily health expenditure on temperature variables for the full sample. Specifications include linear and quadratic terms of current and/or lagged daily temperature, as well as cloud coverage and vapor pressure. All regressions control for Bundesland, year, month, and day-of-week fixed effects. The bottom panel reports *F* tests: the first tests the joint significance of all temperature variables, while the second focuses on lagged temperature terms only.

Table C2: OLS Results: Impact of Temperature – Three Largest

	(1)	(2)	(3)	(4)	(5)	(6)
Temperature	0.0112 (0.017)	0.0373* (0.020)	0.0484** (0.023)	0.0203 (0.033)		
Temperature ²		-0.0014** (0.001)	-0.0014** (0.001)	-0.0006 (0.001)		
Temperature lag			-0.0162 (0.015)	0.0159 (0.032)	-0.0062 (0.013)	0.0324* (0.019)
Temperature lag ²				-0.0010 (0.001)		-0.0015*** (0.001)
Cloud coverage	0.0241 (0.030)	-0.0058 (0.033)	-0.0069 (0.033)	-0.0087 (0.033)	0.0044 (0.023)	-0.0059 (0.023)
Vapor pressure	0.0396 (0.033)	0.0615* (0.035)	0.0715** (0.036)	0.0769** (0.036)	0.0679** (0.027)	0.0771*** (0.027)
<i>F</i> test all temp	0.4	3.0	2.4	2.1		
<i>p</i> value all temp	0.513	0.052	0.068	0.077		
<i>F</i> test lag temp			1.2	1.3	0.2	4.0
<i>p</i> value lag temp			0.272	0.283	0.632	0.019
Adj. <i>R</i> ²	0.5482	0.5484	0.5485	0.5485	0.5482	0.5486
<i>N</i>	7,668	7,668	7,668	7,668	7,668	7,668

NOTE.— Each column reports OLS estimates of daily health expenditure on temperature variables for the three largest Bundesländer. Specifications include linear and quadratic terms of current and/or lagged daily temperature, as well as cloud coverage and vapor pressure. All regressions control for Bundesland, year, month, and day-of-week fixed effects. The bottom panel reports *F* tests: the first tests the joint significance of all temperature variables, while the second focuses on lagged temperature terms only.

Appendix Bibliography

- Andrews, D. W. K.**, “Laws of Large Numbers for Dependent Non-identically Distributed Random Variables,” *Econometric Theory*, 1988, *4* (3), 458–467.
- Billingsley, P.**, *Convergence of Probability Measures*, 2nd ed., New York: Wiley, 1999.
- Bradley, R. C.**, “Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions,” *Probability Surveys*, 2005, *2*, 107–144.
- Davidson, J.**, *Stochastic Limit Theory*, Oxford: Oxford University Press, 1994.
- de Haan, L. and A. Ferreira**, *Extreme Value Theory*, New York: Springer, 2006.
- Hansen, B. E.**, “Convergence to Stochastic Integrals for Dependent Heterogeneous Processes,” *Econometric Theory*, 1992, *8* (4), 489–500.
- Hill, J. B.**, “On Functional Central Limit Theorems for Dependent, Heterogeneous Arrays With Applications to Tail Index and Tail Dependence Estimation,” *Journal of Statistical Planning and Inference*, 2009, *139*, 2091–2110.
- Hoga, Y.**, “Change Point Tests for the Tail Index of β -Mixing Random Variables,” *Econometric Theory*, 2017, *33*, 915–954.
- and **C. Schulz**, “Self-Normalized Inference in (Quantile, Expected Shortfall) Regressions for Time Series,” *Preprint*, 2025, pp. 1–60. <https://arxiv.org/pdf/2502.10065>.
- Hou, Y., X. Leng, L. Peng, and Y. Zhou**, “Panel Quantile Regression for Extreme Risk,” *Journal of Econometrics*, 2024, *240* (1), 105674.
- Jacod, J. and A. N. Shiryaev**, *Limit Theorems for Stochastic Processes*, Berlin: Springer, 1987.
- Lütkepohl, H.**, *New Introduction to Multiple Time Series Analysis*, Berlin: Springer, 2007.
- Pan, X., X. Leng, and T. Hu**, “The Second-Order Version of Karamata’s Theorem With Applications,” *Statistics & Probability Letters*, 2013, *83*, 1397–1403.



CINCH working paper series

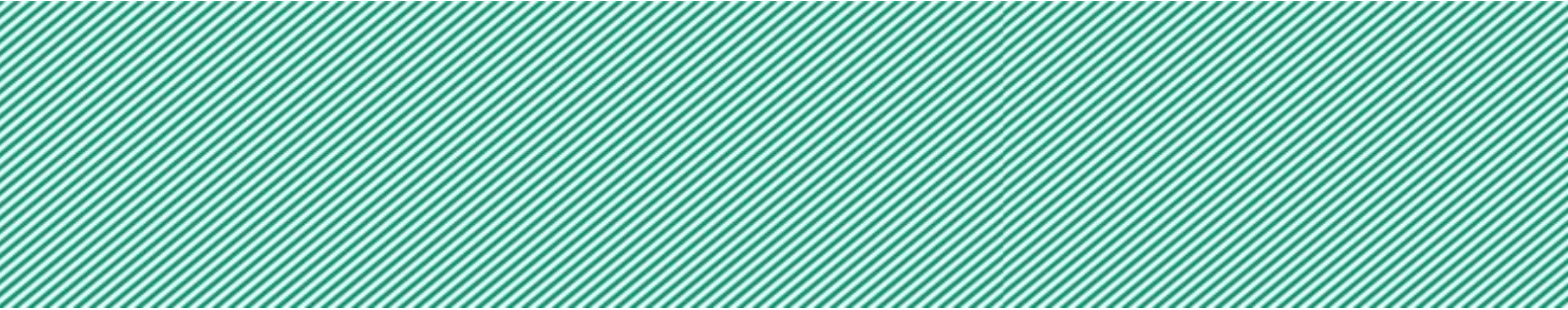
- 1 Halla, Martin and Martina Zweimüller. **Parental Responses to Early Human Capital Shocks:** Evidence from the Chernobyl Accident. CINCH 2014.
- 2 Aparicio, Ainhoa and Libertad González. **Newborn Health and the Business Cycle:** Is it Good to be born in Bad Times? CINCH 2014.
- 3 Robinson, Joshua J. **Sound Body, Sound Mind?:** Asymmetric and Symmetric Fetal Growth Restriction and Human Capital Development. CINCH 2014.
- 4 Bhalotra, Sonia, Martin Karlsson and Therese Nilsson. **Life Expectancy and Mother-Baby Interventions:** Evidence from A Historical Trial. CINCH 2014.
- 5 Goebel, Jan, Christian Krekel, Tim Tiefenbach and Nicolas R. Ziebarth. **Natural Disaster, Environmental Concerns, Well-Being and Policy Action:** The Case of Fukushima. CINCH 2014.
- 6 Avdic, Daniel, **A matter of life and death? Hospital Distance and Quality of Care:** Evidence from Emergency Hospital Closures and Myocardial Infarctions. CINCH 2015.
- 7 Costa-Font, Joan, Martin Karlsson and Henning Øien. **Informal Care and the Great Recession.** CINCH 2015.
- 8 Titus J. Galama and Hans van Kippersluis. **A Theory of Education and Health.** CINCH 2015.
- 9 Dahmann, Sarah. **How Does Education Improve Cognitive Skills?:** Instructional Time versus Timing of Instruction. CINCH 2015.
- 10 Dahmann, Sarah and Silke Anger. **The Impact of Education on Personality:** Evidence from a German High School Reform. CINCH 2015.
- 11 Carbone, Jared C. and Snorre Kverndokk. **Individual Investments in Education and Health.** CINCH 2015.
- 12 Zilic, Ivan. **Effect of forced displacement on health.** CINCH 2015.
- 13 De la Mata, Dolores and Carlos Felipe Gaviria. **Losing Health Insurance When Young:** Impacts on Usage of Medical Services and Health. CINCH 2015.

- 14 Tequame, Miron and Nyasha Tirivayi. **Higher education and fertility: Evidence from a natural experiment in Ethiopia.** CINCH 2015.
- 15 Aoki, Yu and Lualhati Santiago. **Fertility, Health and Education of UK Immigrants: The Role of English Language Skills.** CINCH 2015.
- 16 Rawlings, Samantha B., **Parental education and child health: Evidence from an education reform in China.** CINCH 2015.
- 17 Kamhöfer, Daniel A., Hendrik Schmitz and Matthias Westphal. **Heterogeneity in Marginal Non-monetary Returns to Higher Education.** CINCH 2015.
- 18 Ardila Brenøe, Anne and Ramona Molitor. **Birth Order and Health of Newborns: What Can We Learn from Danish Registry Data?** CINCH 2015.
- 19 Rossi, Pauline. **Strategic Choices in Polygamous Households: Theory and Evidence from Senegal.** CINCH 2016.
- 20 Clarke, Damian and Hanna Mühlrad. **The Impact of Abortion Legalization on Fertility and Maternal Mortality: New Evidence from Mexico.** CINCH 2016.
- 21 Jones, Lauren E. and Nicolas R. Ziebarth. **US Child Safety Seat Laws: Are they Effective, and Who Complies?** CINCH 2016.
- 22 Koppensteiner, Martin Foureaux and Jesse Matheson. **Access to Education and Teenage Pregnancy.** CINCH 2016.
- 23 Hofmann, Sarah M. and Andrea M. Mühlenweg. **Gatekeeping in German Primary Health Care – Impacts on Coordination of Care, Quality Indicators and Ambulatory Costs.** CINCH 2016.
- 24 Sandner, Malte. **Effects of Early Childhood Intervention on Fertility and Maternal Employment: Evidence from a Randomized Controlled Trial.** CINCH 2016.
- 25 Baird, Matthew, Lindsay Daugherty, and Krishna Kumar. **Improving Estimation of Labor Market Disequilibrium through Inclusion of Shortage Indicators.** CINCH 2017.
- 26 Bertoni, Marco, Giorgio Brunello and Gianluca Mazzarella. **Does postponing minimum retirement age improve healthy behaviors before retirement? Evidence from middle-aged Italian workers.** CINCH 2017.
- 27 Berniell, Inés and Jan Bietenbeck. **The Effect of Working Hours on Health.** CINCH 2017.
- 28 Cronin, Christopher, Matthew Forsstrom, and Nicholas Papageorge. **Mental Health, Human Capital and Labor Market Outcomes.** CINCH 2017.

- 29 Kamhöfer, Daniel and Matthias Westphal. **Fertility Effects of College Education:** Evidence from the German Educationl Expansion. CINCH 2017.
- 30 Jones, John Bailey and Yue Li. **The Effects of Collecting Income Taxes on Social Security Benefits.** CINCH 2017.
- 31 Hofmann, Sarah and Andrea Mühlenweg. **Learning Intensity Effects in Students' Mental and Physical Health** – Evidence from a Large Scale Natural Experiment in Germany. CINCH 2017.
- 32 Vollmer, Sebastian and Juditha Wójcik. **The Long-term Consequences of the Global 1918 Influenza Pandemic:** A Systematic Analysis of 117 IPUMS International Census Data Sets. CINCH 2017.
- 33 Thamarapani, Dhanushka, Rockmore, Marc, and Willa Friedman. **The Educational and Fertility Effects of Sibling Deaths.** CINCH 2018.
- 34 Lemmon, Elizabeth. **Utilisation of personal care services in Scotland:** the influence of unpaid carers. CINCH 2018.
- 35 Avdic, Daniel, Büyükdurmus, Tugba, Moscelli, Giuseppe, Pilny, Adam, and Ieva Sriubaite. **Subjective and objective quality reporting and choice of hospital:** Evidence from maternal care services in Germany. CINCH 2018.
- 36 Hentschker, Corinna and Ansgar Wübker. **Quasi-experimental evidence on the effectiveness of heart attack treatment in Germany.** CINCH 2018.
- 37 Pasha, Mochamad, Rockmore, Marc, and Chih Ming Tan. **Early Life Exposure to Above Average Rainfall and Adult Mental Health.** CINCH 2018.
- 38 Elsner, Benjamin and Florian Wozny. **The Human Capital Cost of Radiation:** Long-run Evidence from Exposure outside the Womb. CINCH 2019.
- 39 de la Mata, Dolores and Carlos Felipe Gaviria Garcés. **Exposure to Pollution and Infant Health:** Evidence from Colombia. CINCH 2019
- 40 Besstremyannaya, Galina and Sergei Golovan. **Physicians' altruism in incentives contracts:** Medicare's quality race. CINCH 2019.
- 41 Hayen, Arthur P., Klein, Tobias J., and Martin Salm. **Does the framing of patient cost-sharing incentives matter?** The effects of deductibles vs. no-claim refunds. CINCH 2019.
- 42 Molina, Teresa. **Pollution, Ability, and Gender-Specific Responses to Shocks.** CINCH 2019.

- 43 Fischer, Martin, Gerdtham, Ulf-G, Heckley, Gawain, Karlsson, Martin, Kjellsson, Gustav, and Therese Nilsson. **Education and Health: Long-run Effects of Peers, Tracking and Years.** CINCH 2019.
- 44 Bannenberg, Norman, Førland, Oddvar, Iversen, Tor, Karlsson, Martin, and Henning Øien. **Preventive Home Visits.** CINCH 2019.
- 45 Pestel, Nico and Florian Wozny. **Low Emission Zones for Better Health: Evidence from German Hospitals.** CINCH 2019.
- 46 Bartoli, Paola, Grembi, Veronica, and The Linh Bao Nguyen. **Ethnic Density and Health at Birth.** CINCH 2019.
- 47 Atal, Juan Pablo, Fang, Hanming, Karlsson, Martin, and Nicolas R. Ziebarth. **Long-Term Health Insurance: Theory Meets Evidence.** CINCH 2020.
- 48 Briody, Jonathan. **Parental Unemployment During the Great Recession and Childhood Adiposity.** CINCH 2020.
- 49 Miller, Grant, Paula, Áureo de, and Christine Valente. **Subjective Expectations and Demand for Contraception.** CINCH 2020.
- 50 Kronenberg, Christoph. **New(spaper) Evidence of a Reduction in Suicide Mentions during the 19th-century Gold Rush.** CINCH 2020.
- 51 Avdic, Daniel, Ivets, Maryna, Lagerqvist, Bo, and Ieva Sriubaite. **Providers, Peers and Patients: How do Physicians' Practice Environments Affect Patient Outcomes?** CINCH 2021.
- 52 Alexander Ahammer, Dominik Grübl and Rudolf Winter-Ebmer. **The Health Externalities of Downsizing.** CINCH 2021.
- 53 Emilia Barili, Paola Bertoli, Veronica Grembi, and Veronica Rattini. **COVID Angels Fighting Daily Demons? Mental Health of Healthcare Workers and Religion.** CINCH 2021.
- 54 Signe A. Abrahamsen, Rita Ginja, and Julie Riise. **School Health Programs: Education, Health, and Welfare Dependency of Young Adults.** CINCH 2021.
- 55 Martin Fischer, Martin Karlsson, and Nikolaos Prodromidis. **The Long-Term Effects of Hospital Deliveries.** CINCH 2021.
- 56 Cristina Borra, Libertad González, and David Patiño. **Maternal Age and Infant Health.** CINCH 2021.
- 57 Nicola Barban, Elisabetta De Cao, and Marco Francesconi. **Gene-Environment Effects on Female Fertility.** CINCH 2021.
- 58 Daniel Avdic and Katharina E. Blankart. **A Hard Look at "Soft" Cost-control Measures in Healthcare Organizations: Evidence from Preferred Drug Policies in Germany.** CINCH 2021.

- 59 Sarah Hofmann. **Disease Perception and Preventive Behavior:** The Vaccination Response to Local Measles Outbreaks. CINCH 2022.
- 60 Apostolos Davillas, Victor Hugo de Oliveira, and Andrew M Jones. **A Model of Errors in BMI Based on Self-reported and Measured Anthropometrics with Evidence from Brazilian Data.** CINCH 2022.
- 61 Peter Eibich and Xianhua Zai. **Are the Grandparents Alright?** The Health Consequences of Grandparental Childcare Provision. CINCH 2022.
- 62 Anne Kathrin Heynold, Michael Haylock, and Stefanie Ehmann. **Getting up to Speed:** Informing Prior and Prospective Blood Donors about Supply Uncertainty and Hygiene Measures during the COVID-19 Lockdown. CINCH 2022.
- 63 Elena Ashtari Tafti. **Technology, Skills, and Performance:** The Case of Robots in Surgery. CINCH 2023.
- 64 Kazuya Masuda and Hitoshi Shigeoka. **Education and Later-life Mortality:** Evidence from a School Reform in Japan. CINCH 2023.
- 65 Kai R. Miele. **The Mental Health and Labor Market Effects of Anticipating Job Loss.** CINCH 2024.
- 66 Kaveh Majlesi, Elin Molin, and Paula Roth. **Severe Health Shocks and Financial Well-Being.** CINCH 2024.
- 67 Dora Costa, Lars Olov Bygren, Benedikt Graf, Martin Karlsson, and Joseph Price. **The Economy, the Ghost in Your Gene, and the Escape from Premature Mortality.** CINCH 2025.
- 68 Pablo Muñoz and Cristóbal Otero. **Managers and Public Hospital Performance.** CINCH 2025.
- 69 Christoph Kronenberg, Ulf-Göran Gerdtham, and Martin Karlsson. **Remote Schooling during the Pandemic: A Double Burden for Working Parents in Sweden?.** CINCH 2025.
- 70 Nisha Vernekar and Karan Singhal. **Breaking the Night Barrier: Night-Shift Reforms and Women's Work in India.** CINCH 2025.
- 71 Christoph Kronenberg and Kai R. Miele. **Supply Constraints and Negative Selection: Evidence from Mental Healthcare Markets.** CINCH 2026
- 72 Yannick Hoga and Martin Karlsson. **Extremal Expected Shortfall Regressions: Inference and an Application to Health Care Spending.** CINCH 2026



DuEPublico

Duisburg-Essen Publications online

UNIVERSITÄT
DUISBURG
ESSEN

Offen im Denken

ub | universitäts
bibliothek

Dieser Text wird via DuEPublico, dem Dokumenten- und Publikationsserver der Universität Duisburg-Essen, zur Verfügung gestellt. Die hier veröffentlichte Version der E-Publikation kann von einer eventuell ebenfalls veröffentlichten Verlagsversion abweichen.

DOI: 10.17185/duepublico/84966

URN: urn:nbn:de:hbz:465-20260210-123720-1

Alle Rechte vorbehalten.