Appendix A. Supplementary appendix

In this appendix, we prove the main propositions and theorems in the main text. The derivations will rely on technical lemmas which are proved in the technical lemma appendix.

Appendix A.1. Low-dimensional

Let $\gamma_t^\top = (\alpha_{0,t}^\top, h\alpha_{1,t}^\top)$, and define:

$$\Theta_{t,T}(\gamma_t) = T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} \; \rho \left(y_{s+1} - \alpha_{0,t}^\top X_s - \alpha_{1,t}^\top \left(\frac{s-t}{T} \right) X_s \right)$$

and:

$$\Psi_{t,T}(\gamma_t) = T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} \; \psi \left(y_{s+1} - \alpha_{0,t}^\top X_s - \alpha_{1,t}^\top \left(\frac{s-t}{T} \right) X_s \right) \begin{bmatrix} -X_s \\ -\left(\frac{s-t}{Th} \right) X_s \end{bmatrix}.$$

$\Theta_{t,T}(\gamma_t^0)$ and $\Psi_{t,T}(\gamma_t^0)$ are defined analogously. In the following theoretical statements and their respective derivations, the negative sign in $\Psi_{t,T}$ does not affect the key results as we are often either considering its absolute value or studying its asymptotic behavior. Hence, we will often exclude it for notational convenience.

We begin by stating an approximate expansion result for the objective function Θ, before providing a proof for Proposition 1

Lemma A.1. *Under the assumptions of A.1-5, we have*

$$\sup_{\|\gamma_t - \gamma_t^0\| \le c} \left| \Theta_{t,T}(\gamma_t) - \Theta_{t,T}(\gamma_t^0) - (\gamma_t - \gamma_t^0)\Psi_{t,T}(\gamma_t^0) - \frac{1}{2}(\gamma_t - \gamma_t^0)^\top H_t(\gamma_t - \gamma_t^0) \right| = o_p(1), \tag{A.1}$$

where

$$H_t = \begin{bmatrix} M(t/T) & \mathbf{0} \\ \mathbf{0} & \mu_2 M(t/T) \end{bmatrix}.$$

Proof. See Appendix B. □

Proof of Proposition 1

Let $B(\gamma_t^0, r)$ be an open ball around γ_t^0 with radius $r > 0$. We want to show that

$$\lim_{T\to\infty} P(\inf_{\gamma_t \in B(\gamma_t^0, r)} \Theta_{t,T}(\gamma_t) \ge \Theta_{t,T}(\gamma_t^0)) = 1, \tag{A.2}$$

for arbitrarily positive r. By Lemma $A.1$, we have

$$\Theta_{t,T}(\gamma_t) = \Theta_{t,T}(\gamma_t^0) + (\gamma_t - \gamma_t^0)\Psi_{t,T}(\gamma_t^0) + \frac{1}{2}(\gamma_t - \gamma_t^0)^\top H_t(\gamma_t - \gamma_t^0) + o_p(1).$$

To obtain (A.2), we need to verify:

$$\Psi_{t,T}(\gamma_t^0) = o_p(1), \tag{A.3}$$

and

$$(\gamma_t - \gamma_t^0)^\top H_t(\gamma_t - \gamma_t^0) > 0 \tag{A.4}$$

with high probability. For (A.3), we first define

$$R_{s,t}(X_s) \equiv \alpha_{0,s}^{0\top} X_s - \alpha_{0,t}^{0\top} X_s - \alpha_{1,t}^{0\top} X_s \left(\frac{s-t}{T} \right),$$

then we can rewrite

$$\begin{aligned}
\Psi_{t,T}(\gamma_t^0) &= T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} \psi\left(y_{s+1} - \alpha_{0,t}^{0\top} X_s - \alpha_{1,t}^{0\top} X_s \left(\frac{s-t}{T} \right) \right) \begin{bmatrix} X_s \\ \left(\frac{s-t}{Th}\right) X_s \end{bmatrix} \\
&= T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} \psi\left(y_{s+1} - \alpha_{0,s}^{0\top} X_s + \underbrace{\alpha_{0,s}^{0\top} X_s - \alpha_{0,t}^{0\top} X_s - \alpha_{1,t}^{0\top} X_s \left(\frac{s-t}{T} \right)}_{=R_{s,t}(X_s)} \right) \begin{bmatrix} X_s \\ \left(\frac{s-t}{Th}\right) X_s \end{bmatrix} \\
&= T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} \Big[\psi(\varepsilon_{s+1} + R_{s,t}(X_s)) - \psi(\varepsilon_{s+1}) \Big] \begin{bmatrix} X_s \\ \left(\frac{s-t}{Th}\right) X_s \end{bmatrix} \\
&+ T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} \psi(\varepsilon_{s+1}) \begin{bmatrix} X_s \\ \left(\frac{s-t}{Th}\right) X_s \end{bmatrix} \\
&\equiv g_{t,T} + \tilde{g}_{t,T}.
\end{aligned}$$

We will study $g_{t,T}$ variable-by-variable. Define, for $l = 0, 1$ and $i = 1, ..., d$,

$$g_{t,T,l}(X_{si}) = T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} \Big[\psi(\varepsilon_{s+1} + R_{s,t}(X_s)) - \psi(\varepsilon_{s+1}) \Big] \left(\frac{s-t}{Th} \right)^l X_{si}$$

and analogously for $\tilde{g}_{t,T,l}(X_{si})$. We consider for $l = 0$:

$$\begin{aligned}
E|g_{t,T,0}(X_{si})| &= T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} E \left| \Big[\psi(\varepsilon_{s+1} + R_{s,t}(X_s)) - \psi(\varepsilon_{s+1}) \Big] X_{si} \right| \\
&\leq T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} E \left[\Big\{ \psi(\varepsilon_{s+1} + R_{s,t}(X_s)) - \psi(\varepsilon_{s+1}) \Big\}^2 X_{si}^2 \right]^{1/2}
\end{aligned}$$

$$= T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} E\left[E\left[\left\{ \psi(\varepsilon_{s+1} + R_{s,t}(X_s)) - \psi(\varepsilon_{s+1}) \right\}^2 \middle| X_s \right] X_{si}^2 \right]^{1/2}$$

$$\leq T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} E\left[M_2(|R_{s,t}(X_s)|) X_{si}^2 \right]^{1/2} \leq T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} E\left[c|R_{s,t}(X_s)| X_{si}^2 \right]^{1/2}$$

where the 2 last inequalities follow from A.2(iii): $M_2(|R_{s,t}(X_s)|) = O(|R_{s,t}(X_s)|)$. Next, notice that $\alpha_0(s/T) = \alpha_0(t/T) + \alpha_1(t/T)[(s-t)/T] + \alpha_2(t/T)[(s-t)/T]^2 + o(h^2)$, and therefore we conclude that $E|g_{t,T,0}(X_{si})| = o(h)$. When $l = 1$, we can follow the same procedure to arrive at a smaller order, so we get $E|g_{t,T}| = o(h)$. Then, by Markov's inequality, we get $|g_{t,T}| = o_p(1)$.

For $\tilde{g}_{t,T,0}(X_{si}) = T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} \psi(\varepsilon_{s+1}) X_{si}$:

$$Var(\tilde{g}_{t,T,0}(X_{si})) = (Th)^{-2} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} Var(X_{si}\psi(\varepsilon_{s+1})) k^2\left(\frac{s-t}{Th}\right)$$

$$+ 2(Th)^{-2} \sum_{t-\lfloor Th \rfloor \leq s < j \leq t+\lfloor Th \rfloor} Cov(X_{si}\psi(\varepsilon_{s+1}), X_{ji}\psi(\varepsilon_{j+1})) k\left(\frac{s-t}{Th}\right) k\left(\frac{j-t}{Th}\right).$$

For the second term, note that when $s = t - \kappa$ and $l = t + \kappa$ for $\kappa = 1, .., \lfloor Th \rfloor$, the covariance is the variance due to data reflection. Hence, for these terms, the argument below can be applied. For the other true covariances, we can apply Lemma $C.3$ to show that their sum is o(1). Now, We focus on the variances in the first term. Note that $Var(X_{si}\psi(\varepsilon_{s+1})) = E(X_{si}^2 \psi(\varepsilon_{s+1})^2) + o(1)$, where $E(\psi(\varepsilon_{s+1})X_{si}) = o(1)$ follows from A.2(ii), which also implies that $E(\tilde{g}_{t,T,0}(X_{si})) = o(1)$. We have

$$(Th)^{-2} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} Var(X_{si}\psi(\varepsilon_{s+1})) k^2\left(\frac{s-t}{Th}\right) = \frac{1}{Th} Var(X_{ti}\psi(\varepsilon_{t+1})) \int k^2(u) du + o(1) = o(1).$$

Therefore by Chebyshev's inequality, we get that $|\tilde{g}_{t,T}| = o_p(1)$, and we conclude that $\Psi_{t,T}(\gamma_t^0) = o_p(1)$.

Next, we verify (A.4). By A4(i), the matrix H_t is positive definite, then the minimum eigenvalue of H_t, $\lambda_{min}(H_t)$, is strictly positive. So with high probability,

$$0 < \inf_{\gamma_t \in B(\gamma_t^0, r)} \|\gamma_t - \gamma_t^0\| \lambda_{min}(H_t) \leq \inf_{\gamma_t \in B(\gamma_t^0, r)} (\gamma_t - \gamma_t^0)^\top H_t (\gamma_t - \gamma_t^0).$$

Together with (A.3) and (A.4), we can establish (A.2). This implies that a local minimizer of $\Theta_{t,T}(\gamma_t)$ exists in $B(\gamma_t^0, r)$. By the convexity of the objective function, the local minimizer is thus the global minimizer.

□

In order to prove Proposition 2, we will rely on the following lemmas. Under the conditions of Proposition 2, the following hold:

Lemma A.2. *We have*

$$\Psi_{t,T}(\hat{\gamma}_t) - \Psi_{t,T}(\gamma_t^0) - H_t(\hat{\gamma}_t - \gamma_t^0) = o_p(1/\sqrt{Th}).$$

Proof. Proof omitted as it is similar to Lemma $A.1$.

Lemma A.3. *The long-run variance, as defined in Proposition 2, exists:* $\Omega(\tau) < \infty$ *for all* $\tau \in [0, 1]$

Proof. See Appendix B. □

Lemma A.4. *Define for* $l = 0, 1$,

$$\tilde{g}_{t,T,l} = \frac{1}{Th} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k\left(\frac{s-t}{Th}\right) \psi(\varepsilon_{s+1}) X_s \left(\frac{s-t}{Th}\right)^l.$$

Then $\tilde{g}_{t,T} = [\tilde{g}_{t,T,0}^\top, \tilde{g}_{t,T,1}^\top]^\top$, *and*

$$ThVar(\tilde{g}_{t,T}) = \begin{bmatrix} 2\Omega(t/T)\nu_0 & 0 \\ 0 & 2\Omega(t/T)\nu_2 \end{bmatrix} + o(1) \equiv V_\gamma(t/T) + o(1),$$

where $\nu_j = \int u^j k^2(u) du$.

Proof. The derivation follows Lemma 2 of Chen and Maung (2023), but with ε_{t+1} replaced with $\psi(\varepsilon_{t+1})$ in all the corresponding moments, and the covariance inequality in Lemma $C.3$ for τ-mixing processes is used instead of β-mixing. □

Lemma A.5. *We have*

$$\sqrt{Th}\tilde{g}_{t,T} = \frac{1}{\sqrt{Th}} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} \psi(\varepsilon_{s+1}) k\left(\frac{s-t}{Th}\right) \begin{bmatrix} X_s \\ \left(\frac{s-t}{Th}\right) X_s \end{bmatrix} \to^d N(0, V_\gamma(t/T)).$$

Proof. See Appendix B. □

Proof of Proposition 2

We start from Lemma $A.2$. We know that $\Psi_{t,T}(\hat{\gamma}_t) = 0$ by construction, and $\Psi_{t,T}(\gamma_t^0) = g_{t,T} + \tilde{g}_{t,T}$, whose definitions are from the proof of Proposition 1. To find the bias:

$$E[g_{t,T}] = E\left(T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t}\{\psi(\varepsilon_{s+1} + R_{s,t}(X_s)) - \psi(\varepsilon_{s+1})\} \begin{bmatrix} X_s \\ \left(\frac{s-t}{Th}\right) X_s \end{bmatrix}\right)$$

$$= E\Bigg(T^{-1}\sum_{\substack{s=t-\lfloor Th\rfloor\\ s\neq t}}^{t+\lfloor Th\rfloor} k_{s,t}\{\underbrace{E[\psi(\varepsilon_{s+1}+R_{s,t}(X_s))|X_s]}_{=M_1(X_s)R_{s,t}(X_s)+o(R_{s,t}(X_s))} - \underbrace{E[\psi(\varepsilon_{s+1})|X_s]}_{=o(1)}\}\begin{bmatrix} X_s \\ \left(\frac{s-t}{Th}\right)X_s \end{bmatrix}\Bigg)$$

$$= E\Bigg(T^{-1}\sum_{\substack{s=t-\lfloor Th\rfloor\\ s\neq t}}^{t+\lfloor Th\rfloor} k_{s,t}M_1(X_s)R_{s,t}(X_s)\begin{bmatrix} X_s \\ \left(\frac{s-t}{Th}\right)X_s \end{bmatrix}\Bigg) + o(1)$$

$$= \begin{bmatrix} M(t/T)\frac{\beta''(t/T)}{2}h^2\int u^2k(u)du \\ 0 \end{bmatrix} + o(1),$$

where we have used $R_{s,t}(X_s) = \frac{1}{2}\beta''(t/T)^\top X_s[(s-t)/T]^2 + o(h^2)$, and $\beta''(t/T)$ is the second derivative of $\beta(t/T)$. The second half of the vector is 0 because $\int u^3k(u)du = 0$ by the symmetric kernel assumption. Along with the variance results in the proof of Proposition 1,

$$g_{t,T} \to^p \begin{bmatrix} M(t/T)\frac{\beta''(t/T)}{2}h^2\int u^2k(u)du \\ 0 \end{bmatrix}.$$

So from Lemma $A.2$, we get

$$0 - H_t^{-1}g_{t,T} - (\hat{\gamma}_t - \gamma_t^0) + o_p(1/\sqrt{Th}) = H_t^{-1}\tilde{g}_{t,T}.$$

By multiplying throughout by $\sqrt{Th}$ and applying Lemma $A.5$ for $\tau \in [0,1]$, we get:

$$\sqrt{Th}\Bigg\{(\hat{\gamma}(\tau) - \gamma(\tau)^0) - \begin{bmatrix} \frac{\beta''(\tau)}{2}h^2\mu_2 \\ 0 \end{bmatrix} + o_p(1/\sqrt{Th})\Bigg\} \to^d N(0, H(\tau)^{-1}V_\gamma(\tau)H(\tau)^{-1}).$$

□

Appendix A.2. High-dimensional

Appendix A.2.1. RSC Condition for asymmetric squared loss

The proof of Proposition 3 follows a similar approach to that of Wang and He (2024) for the lin-lin loss function but is overall more intricate for at least three reasons. First, our time-varying estimation relies on the local-linear kernel approach, which necessitates more meticulous and involved calculations. Additionally, the asymmetric squared loss function introduces a complicated non-Lipschitz continuous subgradient, making the analysis significantly more challenging. Finally, our investigation considers weakly dependent or τ-mixing data, as opposed to independent random variables, which complicates the empirical process theory required in the proofs.

Before we discuss the proof strategy, we first consider a specific subgradient of the asymmetric squared loss function. In particular, when $\rho_q(u) = u^2|q - 1_{u<0}|$, we have:

$$\Psi_{t,T}(\gamma_t^0) = -T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s\neq t}}^{t+\lfloor Th \rfloor} k_{s,t}\ \psi_q\left(y_{s+1} - \alpha_{0,t}^{0\top} X_s - \alpha_{1,t}^{0\top}\left(\frac{s-t}{T}\right) X_s\right) \begin{bmatrix} X_s \\ \left(\frac{s-t}{Th}\right) X_s \end{bmatrix}$$

$$= -T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s\neq t}}^{t+\lfloor Th \rfloor} k_{s,t}\ \psi_q\Big(\underbrace{y_{s+1} - \alpha_{0,s}^{0\top} X_s}_{=\varepsilon_{s+1}} + \underbrace{(\alpha_{0,s}^0 - \alpha_{0,t}^0)^\top X_s - \alpha_{1,t}^{0\top}\left(\frac{s-t}{T}\right) X_s}_{=R_{s,t}(X_s)}\Big) Z_{s,t}$$

$$= -2T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s\neq t}}^{t+\lfloor Th \rfloor} k_{s,t}[\varepsilon_{s+1} + R_{s,t}(X_s)](q1_{\{\varepsilon_{s+1}+R_{s,t}(X_s)>0\}} + (1-q)1_{\{\varepsilon_{s+1}+R_{s,t}(X_s)<0\}})Z_{s,t}$$

where $Z_{s,t} = [X_s^\top, \{(s-t)/Th\}X_s^\top]^\top$. Note that $R_{s,t}(X_s) = O_p(s_T h^2)$.

Next, define $\mathcal{E}_+ \equiv 1_{\{\varepsilon_{s+1}+R_{s,t}(X_s)>0\}} - 1_{\{\varepsilon_{s+1}>0\}}$ and $\mathcal{E}_- \equiv 1_{\{\varepsilon_{s+1}+R_{s,t}(X_s)<0\}} - 1_{\{\varepsilon_{s+1}<0\}}$. Then, we have

$$\Psi_{t,T}(\gamma_t^0) = -2T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s\neq t}}^{t+\lfloor Th \rfloor} k_{s,t}\varepsilon_{s+1}(q1_{\{\varepsilon_{s+1}>0\}} + (1-q)1_{\{\varepsilon_{s+1}<0\}})Z_{s,t}$$

$$+ [-2T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s\neq t}}^{t+\lfloor Th \rfloor} k_{s,t}\varepsilon_{s+1} q\mathcal{E}_+ Z_{s,t}] + [-2T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s\neq t}}^{t+\lfloor Th \rfloor} k_{s,t}\varepsilon_{s+1}(1-q)\mathcal{E}_- Z_{s,t}]$$

$$+ [-2T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s\neq t}}^{t+\lfloor Th \rfloor} k_{s,t}R_{s,t}(X_s)(q1_{\{\varepsilon_{s+1}+R_{s,t}(X_s)>0\}} + (1-q)1_{\{\varepsilon_{s+1}+R_{s,t}(X_s)<0\}})Z_{s,t}].$$

Note that $\mathcal{E}_+$ and $\mathcal{E}_-$ are non-zero only when $|\varepsilon_{s+1}| \leq |R_{s,t}(X_s)|$, so $P(\mathcal{E} \neq 0) = P(|\varepsilon_{s+1}| \leq |R_{s,t}(X_s)|)$. Since $|R_{s,t}(X_s)| = O_p(s_T h^2)$, there exists a constant $C > 0$ such that $P(|\varepsilon_{s+1}| \leq |R_{s,t}(X_s)|) \leq P(|\varepsilon_{s+1}| \leq C s_T h^2) = \int_{-Ch^2}^{Ch^2} f_{(s+1)/T}(u)du \leq 2Ch^2 M = O(s_T h^2)$ where M is an upper bound of the density which is finite as given in assumption A.2. Hence for arbitrary $c > 0$, $P(|\mathcal{E}| > c) \leq P(\mathcal{E} \neq 0) = O(s_T h^2) = o(1)$ by condition A.8(i), and thus $\mathcal{E} = o_p(1)$. We conclude similarly for $\mathcal{E}_-$. Hence, the second and third terms are smaller order relative to the first term due to the product with an additional vanishing term. The final term is even smaller by a factor of $s_T h^2$. Note that a similar approximation follows for $\Psi_{t,T}(\gamma_t^0 + \zeta)$, however, because $\zeta^\top Z_{s,t}$ appears in the indicator function, we can use the boundedness of the density and also the sub-Gaussian assumption of the $Z_{s,t}$ in assumption A.6 to bind $P(|\varepsilon_{s+1} - \zeta^\top Z_{s,t}| \leq Ch^2)$.

Following the expression in Proposition 3, on the set $\{\zeta = (\zeta_1^\top, \zeta_2^\top)^\top : \|\zeta\| \leq 1\}$ we arrive at:

$$\Psi_{t,T}(\gamma_t^0 + \zeta) - \Psi_{t,T}(\gamma_t^0)$$

$$= 2T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} Z_{s,t} \Big\{ \varepsilon_{s+1} (q 1_{\{\varepsilon_{s+1} > 0\}} + (1-q) 1_{\{\varepsilon_{s+1} < 0\}})$$

$$- [\varepsilon_{s+1} - \zeta^\top Z_{s,t}](q 1_{\{\varepsilon_{s+1} - \zeta^\top Z_{s,t} > 0\}} + (1-q) 1_{\{\varepsilon_{s+1} - \zeta^\top Z_{s,t} < 0\}}) \Big\} + o_p(1).$$

This object (and the inner product with ζ) can be summarized via the following four cases.

Case 1: $\varepsilon_{s+1} > \zeta^\top Z_{s,t}$ and $\varepsilon_{s+1} > 0$.

$$[\Psi_{t,T}(\gamma_t^0 + \zeta) - \Psi_{t,T}(\gamma_t^0)]^\top \zeta = 2qT^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} (\zeta^\top Z_{s,t})^2;$$

Case 2: $\varepsilon_{s+1} > \zeta^\top Z_{s,t}$ and $\varepsilon_{s+1} < 0$.

$$[\Psi_{t,T}(\gamma_t^0 + \zeta) - \Psi_{t,T}(\gamma_t^0)]^\top \zeta = 2T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} (\zeta^\top Z_{s,t}) \{ -(\varepsilon_{s+1} - \zeta^\top Z_{s,t}) q + \varepsilon_{s+1}(1-q) \};$$

Case 3: $\varepsilon_{s+1} < \zeta^\top Z_{s,t}$ and $\varepsilon_{s+1} > 0$.

$$[\Psi_{t,T}(\gamma_t^0 + \zeta) - \Psi_{t,T}(\gamma_t^0)]^\top \zeta = 2T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} (\zeta^\top Z_{s,t}) \{ -(\varepsilon_{s+1} - \zeta^\top Z_{s,t})(1-q) + \varepsilon_{s+1} q \};$$

Case 4: $\varepsilon_{s+1} < \zeta^\top Z_{s,t}$ and $\varepsilon_{s+1} < 0$.

$$[\Psi_{t,T}(\gamma_t^0 + \zeta) - \Psi_{t,T}(\gamma_t^0)]^\top \zeta = 2(1-q)T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} (\zeta^\top Z_{s,t})^2.$$

Hence,

$$[\Psi_{t,T}(\gamma_t^0 + \zeta) - \Psi_{t,T}(\gamma_t^0)]^\top \zeta = 2T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} \Big\{ q(\zeta^\top Z_{s,t})^2 1_{\{\text{Case 1}\}} + (1-q)(\zeta^\top Z_{s,t})^2 1_{\{\text{Case 4}\}}$$

$$+ (\zeta^\top Z_{s,t}) \{ -(\varepsilon_{s+1} - \zeta^\top Z_{s,t}) q + \varepsilon_{s+1}(1-q) \} 1_{\{\text{Case 2}\}} + (\zeta^\top Z_{s,t}) \{ -(\varepsilon_{s+1} - \zeta^\top Z_{s,t})(1-q) + \varepsilon_{s+1} q \} 1_{\{\text{Case 3}\}} \Big\}$$

$$\geq 2T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} \Big\{ 0 + 0 + \varepsilon_{s+1}^2 (1-q) 1_{\{\text{Case 2}\}} + \varepsilon_{s+1}^2 q 1_{\{\text{Case 3}\}} \Big\}$$

$$\geq 2T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} \varepsilon_{s+1}^2 \min\{q, 1-q\} \geq \frac{1}{3} m_q T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} \varepsilon_{s+1}^2,$$

where $m_q \equiv \min\{q, 1-q\}$. The second inequality requires more explanation. For case 2, we have noted that

$$
\begin{aligned}
(\zeta^\top Z_{s,t})\{-(\varepsilon_{s+1} - \zeta^\top Z_{s,t})q + \varepsilon_{s+1}(1-q)\} &= q(\zeta^\top Z_{s,t})^2 - q(\zeta^\top Z_{s,t})\varepsilon_{s+1} + (1-q)(\zeta^\top Z_{s,t})\varepsilon_{s+1} \\
&\geq q(\zeta^\top Z_{s,t})^2 - q(\zeta^\top Z_{s,t})^2 + (1-q)\varepsilon_{s+1}^2 = (1-q)\varepsilon_{s+1}^2,
\end{aligned}
$$

since for $a, b < 0$ and $a > b$, then $-ab > -b^2$ and $ab > a^2$. The situation is very similar for case 3,

$$
\begin{aligned}
(\zeta^\top Z_{s,t})\{-(\varepsilon_{s+1} - \zeta^\top Z_{s,t})(1-q) + \varepsilon_{s+1}q\} &= -(1-q)(\zeta^\top Z_{s,t})\varepsilon_{s+1} + (1-q)(\zeta^\top Z_{s,t})^2 + q(\zeta^\top Z_{s,t})\varepsilon_{s+1} \\
&\geq -(1-q)(\zeta^\top Z_{s,t})^2 + (1-q)(\zeta^\top Z_{s,t})^2 + q\varepsilon_{s+1}^2 = q\varepsilon_{s+1}^2,
\end{aligned}
$$

and noting that for $a, b > 0$ and $a < b$, we have $ab < b^2$ and $-ab > -b^2$.

Subsequently, we consider the following two functions, modified from Wang and He (2024), that are bounded in $[0, 1]$:

$$
\varphi_{e^2}(u) = \begin{cases} 1, & \text{if } u > 2e^2, \\ -1 + \frac{u}{e^2}, & \text{if } e^2 \leq u \leq 2e^2, \\ 0, & \text{otherwise,} \end{cases}
$$

and

$$
\nu_a(u) = \begin{cases} 1, & \text{if } |u| < \frac{a}{2}, \\ 2 - \frac{2|u|}{a}, & \text{if } 0 < \frac{a}{2} \leq |u| \leq a, \\ 0, & \text{otherwise.} \end{cases}
$$

Finally, we have that

$$
\begin{aligned}
[\Psi_{t,T}(\gamma_t^0 + \zeta) - \Psi_{t,T}(\gamma_t^0)]^\top \zeta &\geq \frac{1}{3} m_q T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} \varepsilon_{s+1}^2 \\
&\geq \frac{1}{3} m_q T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t} \varepsilon_{s+1}^2 \varphi_{\varepsilon_{s+1}^2}(\zeta^\top Z_{s,t}) \nu_{b\|\zeta\|}(\zeta^\top Z_{s,t}) \\
&\equiv P_{t,T,b}(\zeta), \qquad \text{(A.5)}
\end{aligned}
$$

where b is a positive constant. Our proof of Proposition 3 involves studying this lower bound. We will achieve this in the following three lemmas. The first lemma solves the problem caused by the lack of Lipschizt continuity experienced by the subgradient of the asymmetric squared loss function.

Lemma A.6. *Let* $g(\varepsilon_{s+1}^2, Z_{s,t}) \equiv \frac{1}{3} m_q \varepsilon_{s+1}^2 \varphi_{\varepsilon_{s+1}^2}(\zeta^\top Z_{s,t}) \nu_{b\|\zeta\|}(\zeta^\top Z_{s,t})$ *for a positive constant* b *and* $\|\zeta\| \leq 1$*. Then* $g(\cdot)$ *is a 1-Lipschitz map in* $(\varepsilon_{s+1}^2, Z_{s,t})$ *with respect to the* ℓ_1 *norm, and hence a* τ*-mixing process under the conditions of Proposition* 3*.*

Proof. See Appendix B. □

Next, we describe an auxiliary lemma required for deriving a concentration inequality for $P_{t,T,b}(\zeta)$.

Lemma A.7. *Under the conditions of Proposition* 3*, and consider a constant* $b_0 > 0$ *such that for all* $b \geq b_0$*,*

$$E[(\zeta^\top Z_{s,t})^2 1_{\{|\zeta^\top Z_{s,t}| > b\|\zeta\|/2\}}] \leq 0.5 E[(\zeta^\top Z_{s,t})^2] \tag{A.6}$$

holds. Then there exists a positive constant u^* *such that uniformly on* $\{\|\zeta\| \leq 1\}$ *and for all* $b \geq b_0$*, we have*

$$E[P_{t,T,b}(\zeta)] \geq u^* \|\zeta\|^2 + o(1),$$

where $P_{t,T,b}(\zeta)$ *is the lower bound defined in* (A.5)*.*

Proof. See Appendix B. □

Finally, we have,

Lemma A.8. *Define* $S_\theta = \{\zeta : \|\zeta\| = \theta\}$ *where* $0 < \theta \leq 1$*, and* $\Gamma_J = \{\zeta : \zeta \in S_\theta, \|\zeta\|_1 \leq J\|\zeta\|_2, \forall J > 0\}$*, and*

$$Z_{t,T,J} = \sup_{\zeta \in \Gamma_J} |P_{t,T,b}(\zeta) - E[P_{t,T,b}(\zeta)]|$$

for any $b \geq b_0$*. For notational convenience, let the events in Assumption A.6 be written as* $B_{X,t,T} \equiv \{\max_{1 \leq j \leq m_T} \hat{\sigma}_{X_j,t,T}^2 \leq c_X\}$*. Then, there exists a positive constants* C^*, C_Z, m^*, b *and* $1 > \alpha > 1/r_\tau$ *such that on* $B_{X,t,T}$ *and assuming the conditions of Proposition* 3*, we have*

$$P\left(Z_{t,T,J} \geq C^* J\theta \sqrt{\frac{log(m_T)}{(Th)^{1-\alpha}}}\right) \leq \exp\left(-\frac{C^{*2} J^2}{144 m^{*2} b^2} \log(m_T)\right) + \frac{C_Z}{J} \frac{(Th)^{1/2 - \alpha(r_\tau + 1/2)}}{\sqrt{\log(m_T)}}.$$

Proof. See Appendix B. □

Proof of Proposition 3

The proof follows very similarly to that of Theorem 1 in Wang and He (2024) but we use Lemma $A.6$ to Lemma $A.8$ above instead and we exclude a full derivation. However, we note that $[\Psi_{t,T}(\gamma_t^0 + \zeta) - \Psi_{t,T}(\gamma_t^0)]^\top \zeta \geq a_1 \|\zeta\|^2 - a_2 \sqrt{\frac{\log m_T}{(Th)^{1-\alpha}}} \|\zeta\|_1$ holds uniformly over $\{\zeta : \|\zeta\| \leq 1\}$ with probability at least $1 - \Delta_T - k_1 \exp(-k_2 \log m_T) - 2C_Z \frac{(Th)^{1/2 - \alpha(r_\tau + 1/2)}}{\sqrt{\log(m_T)}}$, where $k_1 = 1 + \exp(-k_2 \log 2)$ and $k_2 = \frac{C^{*2}}{72 m^{*2} b^2}$. A difference with Wang and He (2024) is the additional final term in the probability which is a result of the mixing approximation. It is nonetheless $o(1)$ as $T \to \infty$. □

Appendix A.2.2. Consistency proof

Lemma A.9. *Define*

$$\Psi_{t,T,j}(\gamma_t^0) = T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t}\ \psi\left(y_{s+1} - \alpha_{0,t}^{0\top} X_s - \alpha_{1,t}^{0\top}\left(\frac{s-t}{T}\right) X_s\right) Z_{s,t,j},$$

where $Z_{s,t,j}$ is the j-th element of the vector $[-X_s^\top, -(\frac{s-t}{Th})X_s^\top]$. Under the conditions of Proposition 3*, for every $\delta \in (0,1)$, we have*

$$P\left(\max_{1\leq j\leq 2m_T} |\Psi_{t,T,j}(\gamma_t^0)| \leq C\left(\left(\frac{m_T}{\delta (Th)^{\kappa-1}}\right)^{1/\kappa} \vee \sqrt{\frac{\log(24 m_T/\delta)}{(Th)}} \vee \left(\frac{m_T}{\delta (Th)^{3/2}}\right)^{1/2}\right)\right) \geq 1-\delta.$$

Proof. See Appendix B. □

The stated lemma can be seen as the "deviation bound" as in Wong et al. (2020) and will be required to derive the estimation error bounds in Theorem 1.

Before we begin our proofs for the main results, we first provide a characterization of the local solutions to the SCAD problems similar to that of Wang et al. (2012); Wang and He (2024). Our work here extends the result from the specific case of the quantile loss function to that of general loss functions. Since we have a potentially non-smooth loss function along with the non-convex SCAD penalties, the usual approach of relying on the KKT conditions to show that the resulting estimate is a minimizer might not be applicable. Instead, we rely on the following result from Tao and An (1997).

Let $f(\gamma)$ and $g(\gamma)$ be convex functions with subdifferentials $\partial f(\gamma)$ and $\partial g(\gamma)$. Let γ^* be a point with neighborhood Γ. If $\partial f(\gamma) \cap \partial g(\gamma^*) \neq \varnothing$ for all $\gamma \in \Gamma \cap \text{dom } g$ where dom $g = \{\gamma : g(\gamma) < \infty\}$, then γ^* is a local minimizer of $f(\gamma) - g(\gamma)$.

This result is relevant because we can indeed represent the penalized objective function $\mathcal{L}(\alpha_{0,t}^\top, h\alpha_{1,t}^\top) \equiv \mathcal{L}(\gamma_t)$ as the difference of two convex functions. In particular, $\mathcal{L}(\gamma_t) = f(\gamma_t) - g(\gamma_t)$ where

$$f(\gamma_t) = T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t}\ \rho\left(y_{s+1} - \alpha_{0,t}^\top X_s - h\alpha_{1,t}^\top\left(\frac{s-t}{Th}\right) X_s\right) + \lambda_0 \sum_{j=1}^{m_T} |\alpha_{0,t,j}| + \lambda_1 \sum_{j=1}^{m_T} |h\alpha_{1,t,j}|, \quad \text{(A.7)}$$

$$g(\gamma_t) = \sum_{j=1}^{m_T} (G_{\lambda_0}(\alpha_{0,t,j}) + G_{\lambda_1}(h\alpha_{1,t,j})),$$

$$G_{\lambda_i}(h^i \alpha_{i,t,j}) = \left(\frac{(h^i\alpha_{i,t,j})^2 - 2\lambda_i |h^i \alpha_{i,t,j}| + \lambda_i^2}{2(a-1)}\right) 1_{\{\lambda_i \leq |h^i\alpha_{i,t,j}| \leq a\lambda_i\}} + \left(\lambda_i |h^i\alpha_{i,t,j}| - (a+1)\lambda_i^2/2\right) 1_{\{|h^i\alpha_{i,t,j}| > a\lambda_i\}},$$

$(i = 0, 1)$

and a is set at 3.7. Furthermore, write the derivative as $G'_{\lambda_i}(h^i\alpha_{i,t,j})$ and if we stack it up, we have $G'(\gamma_t) = (G'_{\lambda_0}(\alpha_{0,t,1}), ..., G'_{\lambda_0}(\alpha_{0,t,m_T}), G'_{\lambda_1}(h\alpha_{1,t,1}), ..., G'_{\lambda_1}(h\alpha_{1,t,m_T}))^\top$.

In lieu of the result from Tao and An (1997), we consider stationary points $\hat{\gamma}_t$ that satisfy

$$\nabla f(\hat{\gamma}_t) - G'(\hat{\gamma}_t) = 0 \tag{A.8}$$

where $\nabla f(\hat{\gamma}_t)$ refers to a subgradient in $\partial f(\gamma_t)$ evaluated at $\hat{\gamma}_t$. By Lemma 1 of Wang and He (2024), the solutions also satisfy

$$\Psi_{t,T}(\hat{\gamma}_t) + s_\lambda(\hat{\gamma}_t) - G'(\hat{\gamma}_t) = 0, \tag{A.9}$$

where $s_\lambda(\gamma_t) = (\lambda_0 sign(\alpha_{0,t,1}), ..., \lambda_0 sign(\alpha_{0,t,m_T}), \lambda_1 sign(h\alpha_{1,t,1}), ..., \lambda_1 sign(h\alpha_{1,t,m_T}))^\top$, and $\Psi_{t,T}(\gamma_t)$ is defined at the beginning of Appendix A.1 (i.e. it is a subgradient of the unpenalized loss function).

We are now ready to provide a proof of Theorem 1 in a similar fashion to Wang and He (2024).

Proof of Theorem 1

In order to utilize the RSC condition in assumption A.9, we first need to establish that the discrepancy between the estimator and the true value $\hat{\gamma}_t - \gamma_t^0 \equiv \hat{\zeta}_t$ is ≤ 1 with high probability. We will conduct our analysis assuming that the following events are satisfied: (i) $\Lambda_T \equiv \{\lambda \geq \max_j |\Psi_{t,T,j}(\gamma_t^0)|\}$, and (ii) $B_T^{(1)} \equiv \{[\Psi_{t,T}(\hat{\gamma}_t) - \Psi_{t,T}(\gamma_t^0)]^\top \hat{\zeta}_t \geq a_1\|\hat{\zeta}_t\|_2 - a_2\sqrt{\frac{\log m_T}{(Th)^{1-\alpha}}}\|\hat{\zeta}_t\|_1\}$ [13]. For (i), by the rate assumption on λ in A.8 and by Lemma $A.9$, we have $P(\Lambda_T) \geq 1 - \delta$. For (ii), following the argument of lemma $C.4$ in Wang and He (2024), and assumption A.9, we have $P(B_T^{(1)}) \geq 1 - Q_T$. Hence $P(\Lambda_T \cap B_T^{(1)}) \geq 1 - \delta - Q_T$. We want to show that $P(\|\hat{\zeta}_t\|_2 \leq 1) \geq 1 - \delta - Q_T$ and we do so by contradiction.

First suppose $\|\hat{\zeta}_t\|_2 > 1$. Then since $\hat{\gamma}_t$ satisfies (A.9), we have

$$[s(\hat{\gamma}_t) - G'(\hat{\gamma}_t)]^\top (\gamma_t^0 - \hat{\gamma}_t) = -\Psi_{t,T}(\hat{\gamma}_t)^\top (\gamma_t^0 - \hat{\gamma}_t),$$

and we get

$$[G'(\hat{\gamma}_t) - s(\hat{\gamma}_t)]^\top \hat{\zeta}_t = \Psi_{t,T}(\hat{\gamma}_t)^\top \hat{\zeta}_t \tag{A.10}$$

$$[G'(\hat{\gamma}_t) - s(\hat{\gamma}_t) - \Psi_{t,T}(\gamma_t^0)]^\top \hat{\zeta}_t = (\Psi_{t,T}(\hat{\gamma}_t) - \Psi_{t,T}(\gamma_t^0))^\top \hat{\zeta}_t \geq a_1\|\hat{\zeta}_t\|_2 - a_2\sqrt{\frac{\log m_T}{(Th)^{1-\alpha}}}\|\hat{\zeta}_t\|_1,$$

where the inequality is a result of conditioning on $B_T^{(1)}$. By Hölder's inequality, the LHS of the equation above is bounded by

$$[G'(\hat{\gamma}_t) - s(\hat{\gamma}_t) - \Psi_{t,T}(\gamma_t^0)]^\top \hat{\zeta}_t \leq (\|G'(\hat{\gamma}_t) - s(\hat{\gamma}_t)\|_\infty + \|\Psi_{t,T}(\gamma_t^0)\|_\infty)\|\hat{\zeta}_t\|_1.$$

[13]Note that the event in (ii) is not the same as the RSC condition. The subtle difference is due to the lack of the square on $\|\hat{\zeta}_t\|_2$.

Note that by construction, $p_\lambda(\gamma_t) \equiv \sum_{j=1}^{m_T} p_{\lambda_0}(|\alpha_{0,t,j}|) + \sum_{j=1}^{m_T} p_{\lambda_1}(h|\alpha_{1,t,j}|)$ satisfies $p_\lambda(\gamma_t) = (\lambda_0 \sum_{j=1}^{m_T} |\alpha_{0,t,j}| + \lambda_1 \sum_{j=1}^{m_T} |h\alpha_{1,t,j}|) - g(\gamma_t)$ (see definition of $f(\gamma_t)$ and $g(\gamma_t)$ in the discussion before the proof). Hence, $\|G'(\hat{\gamma}_t) - s(\hat{\gamma}_t)\|_\infty = \|\partial p_\lambda(\hat{\gamma}_t)\|\infty \leq \lambda$ where the bound is due to Lemma 4(a) of Loh and Wainwright (2015) and the requirement that λ_0 and λ_1 are proportional to λ. Furthermore, on the event Λ_T we have $\|\Psi_{t,T}(\gamma_t^0)\|_\infty \leq \lambda$. Hence,

$$a_1\|\hat{\zeta}_t\|_2 - a_2 \underbrace{\sqrt{\frac{\log m_T}{(Th)^{1-\alpha}}}}_{\leq \lambda} \|\hat{\zeta}_t\|_1 \leq [G'(\hat{\gamma}_t) - s(\hat{\gamma}_t)) - \Psi_{t,T}(\gamma_t^0)]^\top \hat{\zeta}_t \leq 2\lambda \|\hat{\zeta}_t\|_1,$$

and

$$a_1\|\hat{\zeta}_t\|_2 \leq (2 + a_2)\lambda\|\hat{\zeta}_t\|_1 < 2\kappa(2 + a_2)\lambda,$$

since both $\|\hat{\gamma}_t\|_1$ and $\|\gamma_t^0\|_1$ are $< \kappa$. Then, $2\kappa(2 + a_2)\lambda = 2\kappa(2 + a_2)\tilde{c}H_T(\delta)$ and since $H_T(\delta) < \frac{a_1}{2\kappa\tilde{c}(2+a_2)}$, we get $a_1\|\hat{\zeta}_t\|_2 < a_1$, which is a contradiction since we have assumed that $\|\hat{\zeta}_t\|_2 > 1$. Therefore, label the event $V_T = \{\|\hat{\zeta}_t\|_2 \leq 1\}$, and we know that $P(V_T) \geq 1 - \delta - Q_T$.

Next, label the event implied in assumption A.9 as

$$B_T^{(2)} \equiv \left\{ (\Psi_{t,T}(\hat{\gamma}_t) - \Psi_{t,T}(\gamma_t^0))^\top \hat{\zeta}_t \geq a_1\|\hat{\zeta}_t\|_2^2 - a_2\sqrt{\frac{\log m_T}{(Th)^{1-\alpha}}}\|\hat{\zeta}_t\|_1 \right\}.$$

We have by assumption A.9, on V_T, $P(B_T^{(2)}) \geq 1 - Q_T$. We will now condition our analysis on the event $\Lambda_T \cap V_T \cap B_T^{(2)}$ and we know $P(\Lambda_T \cap V_T \cap B_T^{(2)}) \geq 1 - 2\delta - 2Q_T$.

We can now use the RSC condition. First, recall the definition of $p_{\lambda,\mu}(\gamma_t)$ given in (6). Then, by convexity,

$$\begin{aligned}
&p_{\lambda,\mu}(\gamma_t^0) - p_{\lambda,\mu}(\hat{\gamma}_t) \\
&= \underbrace{\lambda\|\gamma_t^0\|_1 - g(\gamma_t^0)}_{=p_\lambda(\gamma_t^0)} + \frac{\mu}{2}\|\gamma_t^0\|_2^2 - \{\underbrace{\lambda\|\hat{\gamma}_t\|_1 - g(\hat{\gamma}_t)}_{=p_\lambda(\hat{\gamma}_t)} + \frac{\mu}{2}\|\hat{\gamma}_t\|_2^2\} \\
&\geq [\mu\hat{\gamma}_t - G'(\hat{\gamma}_t) + s(\hat{\gamma}_t)]^\top (\gamma_t^0 - \hat{\gamma}_t) \\
&= [-G'(\hat{\gamma}_t) + s(\hat{\gamma}_t)]^\top (\gamma_t^0 - \hat{\gamma}_t) + \mu\hat{\gamma}_t^\top (\gamma_t^0 - \hat{\gamma}_t).
\end{aligned}$$

Since $\|\gamma_t^0\|_2^2 - \|\hat{\gamma}_t\|_2^2 = \|\gamma_t^0 - \hat{\gamma}_t\|_2^2 + 2\hat{\gamma}_t^\top(\gamma_t^0 - \hat{\gamma}_t)$, we have

$$p_\lambda(\gamma_t^0) - p_\lambda(\hat{\gamma}_t) + \frac{\mu}{2}\|\gamma_t^0 - \hat{\gamma}_t\|_2^2 \geq [-G'(\hat{\gamma}_t) + s(\hat{\gamma}_t)]^\top (\gamma_t^0 - \hat{\gamma}_t) = [G'(\hat{\gamma}_t) - s(\hat{\gamma}_t)]^\top \hat{\zeta}_t,$$

or by multiplying -1,

$$-[G'(\hat{\gamma}_t) - s(\hat{\gamma}_t)]^\top \hat{\zeta}_t = -\Psi_{t,T}(\hat{\gamma}_t)^\top \hat{\zeta}_t \geq p_\lambda(\hat{\gamma}_t) - p_\lambda(\gamma_t^0) - \frac{\mu}{2}\|\gamma_t^0 - \hat{\gamma}_t\|_2^2, \tag{A.11}$$

where the equality is due to (A.10). Next, from the RSC condition:

$$-\Psi_{t,T}(\gamma_t^0)^\top\hat{\zeta}_t \geq -\Psi_{t,T}(\hat{\gamma}_t)^\top\hat{\zeta}_t + a_1\|\hat{\zeta}_t\|_2^2 - a_2\sqrt{\frac{\log m_T}{(Th)^{1-\alpha}}}\|\hat{\zeta}_t\|_1$$
$$\geq p_\lambda(\hat{\gamma}_t) - p_\lambda(\gamma_t^0) - \frac{\mu}{2}\|\hat{\zeta}_t\|_2^2 + a_1\|\hat{\zeta}_t\|_2^2 - a_2\sqrt{\frac{\log m_T}{(Th)^{1-\alpha}}}\|\hat{\zeta}_t\|_1,$$

where we have used (A.11) in the last line. Again by Hölder's inequality and Lemma $A.9$:

$$\lambda\|\hat{\zeta}_t\|_1 \geq \|\Psi_{t,T}(\gamma_t^0)\|_\infty\|\hat{\zeta}_t\|_1 \geq p_\lambda(\hat{\gamma}_t) - p_\lambda(\gamma_t^0) + (a_1 - \frac{\mu}{2})\|\hat{\zeta}_t\|_2^2 - a_2\sqrt{\frac{\log m_T}{(Th)^{1-\alpha}}}\|\hat{\zeta}_t\|_1.$$

Next, by the Lipschitz-continuity of p_λ (see lemma 4 of Loh and Wainwright, 2015), we get $|p_\lambda(\hat{\gamma}_t) - p_\lambda(\gamma_t^0)| \leq \lambda\|\hat{\zeta}_t\|_1$ or $-|p_\lambda(\hat{\gamma}_t) - p_\lambda(\gamma_t^0)| \geq -\lambda\|\hat{\zeta}_t\|_1$, and since for all $x \in \mathbb{R}$, $x \geq -|x|$, we have

$$3\lambda\|\hat{\zeta}_t\|_1 \geq 2\lambda\|\hat{\zeta}_t\|_1 + a_2\sqrt{\frac{\log m_T}{(Th)^{1-\alpha}}}\|\hat{\zeta}_t\|_1 \geq (a_1 - \frac{\mu}{2})\|\hat{\zeta}_t\|_2^2,$$

given the conditions on λ. Then, by a standard calculation, we can show that $\|\hat{\zeta}_t\|_1 \leq 4\sqrt{s_T}\|\hat{\zeta}_t\|_2$ (for e.g. by using corollary 1 of Wang and He, 2024), and therefore we conclude that

$$\frac{12}{(a_1 - \frac{\mu}{2})}\lambda\sqrt{s_T} \geq \|\hat{\zeta}_t\|_2,$$

is valid conditional on the event $\Lambda_T \cap V_T \cap B_T^{(2)}$. □

Appendix A.2.3. Selection proof

We first introduce a restricted problem termed as the *biased oracle problem*. This is similar to the original penalized problem in (5) but with the identity of 'relevant' forecasts (and the gradient terms) known prior to solving the problem.

Recall that the first $s_{0,T}$ forecasts are deemed 'relevant' along with the first $s_{1,T}$ gradient terms from the local linear expansion. Define $\alpha_{0,t}^{\mathcal{S}}$ and $\alpha_{1,t}^{\mathcal{S}}$ to be the sub-vectors that represent the first $s_{0,T}$ and $s_{1,T}$ elements of $\alpha_{0,t}$ and $\alpha_{1,t}$ respectively. Furthermore, for notational convenience, we will establish the index set of relevant variables to be $\mathcal{S} \equiv \{1, ..., s_{0,T}, m_T + 1, ..., m_T + s_{1,T}\}$. Variables that have zero coefficients (i.e. not relevant) therefore belong to the index set $\mathcal{S}^c$. Note that both sets are actually indexed by T but we avoid annotating as such to save on notation. Solutions to the biased oracle problem are given by

$$\widetilde{\gamma}_t^{\mathcal{S}} = (\widetilde{\alpha}_{0,t}^{\mathcal{S}}, h\widetilde{\alpha}_{1,t}^{\mathcal{S}})$$
$$= \underset{(\alpha_{0,t}^{\mathcal{S}}, h\alpha_{1,t}^{\mathcal{S}})\in\mathbb{R}^{\mathcal{S}}}{\arg\min}\; T^{-1}\sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} k_{s,t}\; \rho\left(y_{s+1} - \alpha_{0,t}^{\top\mathcal{S}}\{X_s\}_{\mathcal{S}} - h\alpha_{1,t}^{\top\mathcal{S}}\left\{\left(\frac{s-t}{Th}\right)X_s\right\}_{\mathcal{S}}\right)$$

$$+\sum_{j=1}^{s_{0,T}} p_{\lambda_0}(|\alpha_{0,t,j}|) + \sum_{j=1}^{s_{1,T}} p_{\lambda_1}(|h\alpha_{1,t,j}|), \tag{A.12}$$

where $\{X_s\}_{\mathcal{S}}$ and $\left\{\left(\frac{s-t}{Th}\right) X_s\right\}_{\mathcal{S}}$ correspond to the sub-vectors with non-zero coefficients. Furthermore, the solutions satisfy:

$$\Psi^{\mathcal{S}}_{t,T}(\widetilde{\gamma}^{\mathcal{S}}_t) + s^{\mathcal{S}}_{\lambda}(\widetilde{\gamma}^{\mathcal{S}}_t) - G'^{\mathcal{S}}(\widetilde{\gamma}^{\mathcal{S}}_t) = 0, \tag{A.13}$$

where

$$\Psi^{\mathcal{S}}_{t,T}(\gamma^{\mathcal{S}}_t) = T^{-1} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} k_{s,t}\ \psi\left(y_{s+1} - \alpha^{\mathcal{S}\top}_{0,t}\{X_s\}_{\mathcal{S}} - \alpha^{\mathcal{S}\top}_{1,t}\left\{\left(\frac{s-t}{Th}\right) X_s\right\}_{\mathcal{S}}\right) \begin{bmatrix} -\{X_s\}_{\mathcal{S}} \\ -\left\{\left(\frac{s-t}{Th}\right) X_s\right\}_{\mathcal{S}} \end{bmatrix},$$

and both $s^{\mathcal{S}}_{\lambda}(\gamma^{\mathcal{S}}_t)$ and $G'^{\mathcal{S}}(\widetilde{\gamma}^{\mathcal{S}}_t)$ are analogous to $s_{\lambda}(\gamma_t)$ and $G'(\gamma_t)$ from (A.9) but containing only the relevant variables.

In order to discuss the properties of $\widetilde{\gamma}^{\mathcal{S}}_t$ in relation to the non-oracle penalized problem, we define the following vector:

$$\widehat{\gamma}^{\mathcal{S}}_t = (\underbrace{\widetilde{\alpha}^{\mathcal{S}\top}_{0,t}, \mathbf{0}^{\top}_{(m_T - s_{0,T})\times 1}}_{\equiv \widehat{\alpha}^{\mathcal{S}\top}_{0,t}}, \underbrace{h\widetilde{\alpha}^{\mathcal{S}\top}_{1,t}, \mathbf{0}^{\top}_{(m_T - s_{1,T})\times 1}}_{\equiv h\widehat{\alpha}^{\mathcal{S}\top}_{1,t}})^{\top}, \tag{A.14}$$

where $\mathbf{0}_{(m_T - s_{0,T})\times 1}$ and $\mathbf{0}_{(m_T - s_{1,T})\times 1}$ are zero vectors of corresponding lengths.

We now detail the proof strategy of Theorem 2 as follows:

1. Determine that $\widehat{\gamma}^{\mathcal{S}}_t$ is consistent for γ^0_t (Lemma A.10);
2. Using the results from Tao and An (1997), show that $\widehat{\gamma}^{\mathcal{S}}_t$ is also a local minimizer of the original penalized problem in (5) (Lemma A.12);
3. Show that any stationary point, $\hat{\gamma}_t$, to the original penalized problem is supported on $\mathcal{S}$. Furthermore, we establish that $\widehat{\gamma}^{\mathcal{S}}_t$ is a unique stationary point in the biased oracle problem in (A.12) and by extension, $\hat{\gamma}_t$ is also a unique stationary point to the program in (5) (Lemma A.13);
4. Finally, we argue that $\hat{\gamma}_t$ is equivalent to the solution to the *unbiased oracle problem* which is (A.12) without any penalty terms.

These steps are detailed in the following list of lemmas:

Lemma A.10. *Under the conditions of and definitions given in Theorem* 1*, we have with high probability,*

$$\|\widehat{\gamma}^{\mathcal{S}}_t - \gamma^0_t\|_2 \leq c^*\lambda\sqrt{s_T}.$$

.

Proof. This proof follows almost exactly the derivation of Theorem 1 although we use (A.13) instead of (A.9) as our starting point. The RSC condition as stated in assumption A.9 is still applicable because

$$(\Psi_{t,T}(\widehat{\gamma}_t^{\mathcal{S}}) - \Psi_{t,T}(\gamma_t^0))^\top[\widehat{\gamma}_t^{\mathcal{S}} - \gamma_t^0] = (\Psi_{t,T}^{\mathcal{S}}(\widetilde{\gamma}_t^{\mathcal{S}}) - \Psi_{t,T}^{\mathcal{S}}(\gamma_t^{0,\mathcal{S}}))^\top[\widetilde{\gamma}_t^{\mathcal{S}} - \gamma_t^{0,\mathcal{S}}],$$

where $\gamma_t^{0,\mathcal{S}}$ refers to the true non-zero parameters because $\widehat{\gamma}_{t,j}^{\mathcal{S}} - \gamma_{t,j}^0 = 0$ for all $j \in \mathcal{S}^c$. □

Next,

Lemma A.11. *Under the conditions of Theorem 1 and assumption A.10, we have with high probability*

(i) $|\widehat{\gamma}_{t,j}^{\mathcal{S}}| \geq (a + a^*)\sqrt{s_T}\lambda$ *for all* $j \in \mathcal{S}$ *where* a *is the constant in the SCAD penalty, and* $a^* > 0$ *is an arbitrary constant;*

(ii) $\Psi_{t,T,j}(\widehat{\gamma}_t^{\mathcal{S}}) = 0$ *for all* $j \in \mathcal{S}$*;*

(iii) $|\Psi_{t,T,j}(\widehat{\gamma}_t^{\mathcal{S}})| \leq C\lambda$ *for all* $j \in \mathcal{S}^c$ *and* $C > 0$*.*

These results are required to show the following:

Lemma A.12. *With high probability,* $\widehat{\gamma}_t^{\mathcal{S}}$ *is a local minimizer of* (5)*.*

The proofs of both lemmas are relegated to Appendix B.

Lemma A.13. *Let* $\hat{\gamma}_t$ *be any stationary point of the original penalized loss function* (5)*. Then with high probability,* $\hat{\gamma}_t$ *is supported on* $\mathcal{S}$*. Furthermore,* $\hat{\gamma}_t$ *is unique.*

Proof. See Appendix B. □

Proof of Theorem 2 From the results of Lemma $A.13$, $\hat{\gamma}_t = (\widetilde{\gamma}_t^{\mathcal{S}\top}, 0_{\mathcal{S}^c}^\top)^\top$ where $\widetilde{\gamma}_t^{\mathcal{S}\top}$ is the unique solution to (A.12) (see proof of Lemma $A.13$ for details). In the proof of Lemma $A.11$, we showed that $s_\lambda^{\mathcal{S}}(\widetilde{\gamma}_t^{\mathcal{S}}) = G'^{\mathcal{S}}(\widetilde{\gamma}_t^{\mathcal{S}})$ and thus $\Psi_{t,T}^{\mathcal{S}}(\widetilde{\gamma}_t^{\mathcal{S}}) = 0$. By Lemma 1 of Loh and Wainwright (2017), the *unbiased* oracle program in (10) is strictly convex on $\mathbb{R}^{\mathcal{S}}$, and hence $\Psi_{t,T}^{\mathcal{S}}(\widetilde{\gamma}_t^{\mathcal{S}}) = 0$ implies that $\widetilde{\gamma}_t^{\mathcal{S}}$ is the unique global minimizer of the unbiased oracle problem. Hence, $\widetilde{\gamma}_t^{\mathcal{S}} = \hat{\gamma}_t^{\mathcal{O},\mathcal{S}}$ where recall that $\hat{\gamma}_t^{\mathcal{O},\mathcal{S}}$ is the unique stationary point of (10). Therefore $\hat{\gamma}_t = (\widetilde{\gamma}_t^{\mathcal{S}\top}, 0_{\mathcal{S}^c}^\top)^\top = (\hat{\gamma}_t^{\mathcal{O},\mathcal{S}}, 0_{\mathcal{S}^c}^\top)^\top = \hat{\gamma}_t^{\mathcal{O}}$. □

Appendix B. Proofs of Lemmas

Proof of Lemma $A.1$

Write

$$\Theta_{t,T}(\gamma_t) - \Theta_{t,T}(\gamma_t^0) - (\gamma_t - \gamma_t^0)\Psi_{t,T}(\gamma_t^0) - E\left[\Theta_{t,T}(\gamma_t) - \Theta_{t,T}(\gamma_t^0) - (\gamma_t - \gamma_t^0)\Psi_{t,T}(\gamma_t^0)\right]$$
$$\equiv T^{-1}\sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} U_{s,t} - E(U_{s,t}) \qquad \text{(B.1)}$$

where

$$U_{s,t} = k_{s,t}\Bigg[\rho\bigg(y_{s+1} - \alpha_{0,t}^\top X_s - \alpha_{1,t}^\top\bigg(\frac{s-t}{T}\bigg)X_s\bigg) - \rho\bigg(y_{s+1} - \alpha_{0,t}^{0\top} X_s - \alpha_{1,t}^{0\top}\bigg(\frac{s-t}{T}\bigg)X_s\bigg)$$
$$- (\gamma_t - \gamma_t^0)^\top \begin{bmatrix} X_s \\ \left(\frac{s-t}{Th}\right)X_s \end{bmatrix} \psi\bigg(y_{s+1} - \alpha_{0,t}^{0\top} X_s - \alpha_{1,t}^{0\top}\bigg(\frac{s-t}{T}\bigg)X_s\bigg)\Bigg].$$

We want to show that (B.1) is $o_p(1)$. To do so, consider

$$Var\Bigg(T^{-1}\sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} U_{s,t}\Bigg) = T^{-2}\sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} Var(U_{s,t}) + 2T^{-2}\sum_{\substack{t-\lfloor Th\rfloor \le s<l\le t+\lfloor Th\rfloor \\ s,l\neq t}} Cov(U_{s,t}, U_{l,t})$$
$$= V_{1,s,t} + V_{2,s,t} \qquad \text{(B.2)}$$

where $V_{i,s,t}(i = 1, 2, 3)$ refers to the corresponding summands above. We start with $V_{1,s,t}$. First note that

$$|U_{s,t}| \le \Bigg|k_{s,t}\Bigg[\rho\bigg(y_{s+1} - \alpha_{0,t}^\top X_s - \alpha_{1,t}^\top\bigg(\frac{s-t}{T}\bigg)X_s\bigg) - \rho\bigg(y_{s+1} - \alpha_{0,t}^{0\top} X_s - \alpha_{1,t}^{0\top}\bigg(\frac{s-t}{T}\bigg)X_s\bigg)$$
$$- (\gamma_t - \gamma_t^0)^\top \begin{bmatrix} X_s \\ \left(\frac{s-t}{Th}\right)X_s \end{bmatrix} \psi\bigg(y_{s+1} - \alpha_{0,t}^{0\top} X_s - \alpha_{1,t}^{0\top}\bigg(\frac{s-t}{T}\bigg)X_s\bigg)\Bigg]\Bigg|$$
$$\le \Bigg|k_{s,t}(\gamma_t - \gamma_t^0)^\top \begin{bmatrix} X_s \\ \left(\frac{s-t}{Th}\right)X_s \end{bmatrix} \Bigg[\psi\bigg(y_{s+1} - \alpha_{0,t}^\top X_s - \alpha_{1,t}^\top\bigg(\frac{s-t}{T}\bigg)X_s\bigg) - \psi\bigg(y_{s+1} - \alpha_{0,t}^{0\top} X_s$$
$$- \alpha_{1,t}^{0\top}\bigg(\frac{s-t}{T}\bigg)X_s\bigg)\Bigg]\Bigg| \qquad \text{(B.3)}$$

where the last inequality follows from the convexity of ρ. Since,

$$T^{-2}\sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} Var(U_{s,t}) \le T^{-2}\sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} E(U_{s,t}^2)$$

$$\leq T^{-2}\sum_{\substack{s=t-\lfloor Th\rfloor\\ s\neq t}}^{t+\lfloor Th\rfloor} E\Bigg(k_{s,t}^2\Bigg\{(\gamma_t-\gamma_t^0)^\top\begin{bmatrix}X_s\\ \left(\frac{s-t}{Th}\right)X_s\end{bmatrix}\Bigg\}^2\Bigg[\psi\Bigg(y_{s+1}-\alpha_{0,t}^\top X_s-\alpha_{1,t}^\top\left(\frac{s-t}{T}\right)X_s\Bigg)-\psi\Bigg(y_{s+1}-\alpha_{0,t}^{0\top}X_s$$

$$-\alpha_{1,t}^{0\top}\left(\frac{s-t}{T}\right)X_s\Bigg)\Bigg]^2\Bigg)$$

$$=T^{-2}\sum_{\substack{s=t-\lfloor Th\rfloor\\ s\neq t}}^{t+\lfloor Th\rfloor} E\Bigg(k_{s,t}^2\Bigg\{(\gamma_t-\gamma_t^0)^\top\begin{bmatrix}X_s\\ \left(\frac{s-t}{Th}\right)X_s\end{bmatrix}\Bigg\}^2 E\Bigg[\Bigg[\psi\Bigg(y_{s+1}-\alpha_{0,t}^\top X_s-\alpha_{1,t}^\top\left(\frac{s-t}{T}\right)X_s\Bigg)-\psi\Bigg(y_{s+1}-\alpha_{0,t}^{0\top}X_s$$

$$-\alpha_{1,t}^{0\top}\left(\frac{s-t}{T}\right)X_s\Bigg)\Bigg]^2\Bigg|X_s\Bigg]\Bigg). \tag{B.4}$$

Let us add and subtract $\psi(\varepsilon_{s+1})$ into the conditional expectation:

$$E\Bigg[\Bigg[\psi\Bigg(y_{s+1}-\alpha_{0,t}^\top X_s-\alpha_{1,t}^\top\left(\frac{s-t}{T}\right)X_s\Bigg)-\psi(\varepsilon_{s+1})-\psi\Bigg(y_{s+1}-\alpha_{0,t}^{0\top}X_s-\alpha_{1,t}^{0\top}\left(\frac{s-t}{T}\right)X_s\Bigg)+\psi(\varepsilon_{s+1})\Bigg]^2\Bigg|X_s\Bigg]. \tag{B.5}$$

Next we study the 2nd half of the conditional expectation with the help of Assumption A.2(iii):

$$\begin{aligned}&E\Bigg[\Bigg(\psi\Bigg(y_{s+1}-\alpha_{0,t}^{0\top}X_s-\alpha_{1,t}^{0\top}\left(\frac{s-t}{T}\right)X_s\Bigg)-\psi(\varepsilon_{s+1})\Bigg)^2\Bigg|X_s\Bigg]\\ &\leq M_2\Bigg(|(\alpha_{0s}^0-\alpha_{0t}^0)^\top X_s-\alpha_{1t}^{0\top}(s-t)/TX_s|\Bigg).\end{aligned} \tag{B.6}$$

For the first part of the conditional expectation, we have

$$\begin{aligned}&E\Bigg[\Bigg(\psi\Bigg(y_{s+1}-\alpha_{0,t}^{\top}X_s-\alpha_{1,t}^{\top}\left(\frac{s-t}{T}\right)X_s\Bigg)-\psi(\varepsilon_{s+1})\Bigg)^2\Bigg|X_s\Bigg]\\ &\leq M_2\Bigg(|(\alpha_{0t}^0-\alpha_{0t})^\top X_s+(\alpha_{0s}^0-\alpha_{0t}^0)^\top X_s-\alpha_{1t}^{\top}(s-t)/TX_s|\Bigg).\end{aligned} \tag{B.7}$$

Terms involving $(\alpha_{0s}^0-\alpha_{0t}^0)$ are smaller order, so using the definition of $M_2(|\epsilon|)=O(|\epsilon|)$ in A.2(iii), and for some constant $C^*>0$:

$$\text{(B.4)}\leq C^*T^{-2}\sum_{\substack{s=t-\lfloor Th\rfloor\\ s\neq t}}^{t+\lfloor Th\rfloor} k_{s,t}^2 E\Bigg(\Bigg\{(\gamma_t-\gamma_t^0)^\top\begin{bmatrix}X_s\\ (\frac{s-t}{Th})X_s\end{bmatrix}\Bigg\}^3\Bigg)+o(1). \tag{B.8}$$

Given our moment assumptions and noting that $\|\gamma_t-\gamma_t^0\|\leq c$, we can see that the expectation in the expression is $O(1)$.

Finally, using the Riemann sum approximation $\frac{1}{(Th)^2}\sum_{\substack{s=t-\lfloor Th\rfloor\\ s\neq t}}^{t+\lfloor Th\rfloor} K^2\left(\frac{s-t}{Th}\right)=\frac{1}{Th}\int_{-1}^1 K^2(z)dz+O(\frac{1}{Th})$ we conclude that

$$T^{-2}\sum_{\substack{s=t-\lfloor Th\rfloor\\ s\neq t}}^{t+\lfloor Th\rfloor} Var(U_{s,t})=O((Th)^{-1}).$$

Next for $V_{2,s,t}$ defined in (B.2), note that because of the data reflection, when $s = t - \kappa$ and $l = t + \kappa$ for $\kappa = 1, ..., \lfloor Th \rfloor$, we have that $Cov(U_{s,t}, U_{l,t}) = Var(U_{s,t})$. Hence, define the following set of indices $\Theta_{\{s\sim l\}} = \{s, l : t - \lfloor Th \rfloor \leq s < l \leq t + \lfloor Th \rfloor, s = t - \kappa, l = t + \kappa, \kappa = 1, ..., \lfloor Th \rfloor\}$ and, we can further split this term:

$$V_{2,s,t} = 2T^{-2} \sum_{(s,l)\in\Theta_{\{s\sim l\}}} \underbrace{Cov(U_{s,t}, U_{l,t})}_{=Var(U_{s,t})} + 2T^{-2} \sum_{(s,l)\notin\Theta_{\{s\sim l\}}} Cov(U_{s,t}, U_{l,t}) = V_{2,1,s,t} + V_{2,2,s,t}.$$

The argument for the 1st term is exactly the same as the above, which implies that $V_{2,1,s,t} = O((Th)^{-1})$. For $V_{2,2,s,t}$, we can show with mixing arguments (see Lemma $C.3$) that it is of the same order.

Therefore, by Chebyshev's inequality, we get that

$$\begin{aligned} &\Theta_{t,T}(\gamma_t) - \Theta_{t,T}(\gamma_t^0) - (\gamma_t - \gamma_t^0)\Psi_{t,T}(\gamma_t^0) - E\left[\Theta_{t,T}(\gamma_t) - \Theta_{t,T}(\gamma_t^0) - (\gamma_t - \gamma_t^0)\Psi_{t,T}(\gamma_t^0)\right] \\ &= O_p((Th)^{-1/2}) = o_p(1). \end{aligned} \tag{B.9}$$

Next, we apply lemma 1 of Bai et al. (1992) to obtain:

$$E(\Theta_{t,T}(\gamma_t) - \Theta_{t,T}(\gamma_t^0)) = \frac{1}{2}(\gamma_t - \gamma_t^0)^\top H_t(\gamma_t - \gamma_t^0) + o(1). \tag{B.10}$$

Similar to (B.7), we can conclude that $E(\Psi_{t,T}(\gamma_t^0)) = o(1)$. Piecing this together with (B.10):

$$E(\Theta_{t,T}(\gamma_t) - \Theta_{t,T}(\gamma_t^0) - (\gamma_t - \gamma_t^0)\Psi_{t,T}(\gamma_t^0)) = \frac{1}{2}(\gamma_t - \gamma_t^0)^\top H_t(\gamma_t - \gamma_t^0) + o(1).$$

Then together with (B.9) and convexity of (B.1), we get the desired result. □

Proof of Lemma $A.3$

Recall that $\Omega(t/T) = \sum_{j=\infty}^{\infty} \Gamma_j(t/T)$ where $\Gamma_j(t/T) = Cov(X_t\psi(\varepsilon_{t+1}), X_{t+j}\psi(\varepsilon_{t+1+j}))$. By Lemma $C.3$,

$$\Gamma_j(t/T) \leq C\tau_j^{*\frac{R-2}{R-1}} < \infty$$

for some constant $C < \infty$ and $R > 2$. Here, we follow the argument in Lemma $C.3$ and let $R = pq/(p+q) > 2$ where p, and q are defined in assumption A.3. Furthermore, $\tau_j^{*\frac{(R-2)}{(R-1)}} \leq C \cdot j^{-\varphi^* \frac{R-2}{R-1}}$.

Returning to the long-run variance, because the sum is symmetric we focus on the positive side:

$$\sum_{j=1}^{\infty} \Gamma_j(t/T) \leq C \sum_{j=1}^{\infty} j^{-\varphi^* \frac{R-2}{R-1}}.$$

To show convergence, we can use a comparison test with respect to a p-harmonic series. The result follows by noting that $\varphi^* > \frac{R-1}{R-2}$ as given by A.3(ii) so $\varphi^* \frac{R-2}{R-1} > 1$. □

Proof of Lemma $A.5$

To prove the normality result, we use the triangular array CLT of Neumann (2013). This CLT is also used in Babii et al. (2024), and is attractive because of its minimal dependence requirements. The application involves verifying conditions (2.1) to (2.4) in Theorem 2.1 of Neumann (2013) which we replicate here.

Theorem 2.1 of Neumann (2013) Suppose $(X_{T,t})_{t=1,...,T}$, $T \in \mathbb{N}$ is a triangular scheme of random variables with $E[X_{T,t}] = 0$ and $\sum_{t=1}^{T} E[X_{T,t}^2] \leq \nu_0$ for all T, t and some $\nu_0 < \infty$. Assume that as $T \to \infty$, we have

$$\sigma_T^2 \equiv Var(X_{T,1} + ... + X_{T,T}) \to \sigma^2 \in [0, \infty) \tag{B.11}$$

and

$$\sum_{t=1}^{T} E[X_{T,t}^2 1_{\{|X_{T,t}|>e\}}] \to 0 \tag{B.12}$$

holds for all $e > 0$. Furthermore, assume that there exists a summable sequence $(\theta_k)_{k\in\mathbb{N}}$ such that for all $u \in \mathbb{N}$ and all indices $1 \leq s_1 \leq s_2 < ... < s_u < s_u + k = j_1 \leq j_2 \leq T$, the following upper bounds for covariances hold true: for all measurable functions $g : \mathbb{R}^u \to \mathbb{R}$ with $\|g\|_\infty = sup_{x\in\mathbb{R}^u} |g(x)| \leq 1$,

$$|Cov(g(X_{T,s_1}, ..., X_{T,s_u})X_{T,s_u}, X_{T,j_1})| \leq (E[X_{T,s_u}^2] + E[X_{T,j_1}^2] + T^{-1})\theta_k \tag{B.13}$$

and

$$|Cov(g(X_{T,s_1}, ..., X_{T,s_u}), X_{T,j_1}X_{T,j_2})| \leq (E[X_{T,j_1}^2] + E[X_{T,j_2}^2] + T^{-1})\theta_k. \tag{B.14}$$

Then we have,

$$X_{T,1} + ... + X_{T,T} \to^d N(0, \sigma^2).$$

$\square$.

Define:

$$v^\top \frac{1}{\sqrt{Th}} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s\neq t}}^{t+\lfloor Th \rfloor} \psi(\varepsilon_{s+1}) K\left(\frac{s-t}{Th}\right) \begin{bmatrix} X_s \\ \left(\frac{s-t}{Th}\right) X_s \end{bmatrix} \equiv \frac{1}{\sqrt{Th}} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s\neq t}}^{t+\lfloor Th \rfloor} \psi(\varepsilon_{s+1}) K\left(\frac{s-t}{Th}\right) v^\top Z_{s,t}$$

where v is a unit vector of conformable dimension and $Z_{s,t} = (X_s^\top, (s - t/Th)X_s^\top)^\top$. The multiplication with the unit vector is to facilitate the application of Crámer-Wold.

Verification of (B.11). Condition (B.11) is the existence of the long-run variance which is provided by Lemma $A.3$.

Verification of (B.12). This is the Lindeberg condition. We want to show that for $\psi(\varepsilon_{s+1})Z_{s,t} - E[\psi(\varepsilon_{s+1})Z_{s,t}] \equiv V_{s,t}$:

$$\lim_{T\to\infty} \sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} K\left(\frac{s-t}{Th}\right) E\left[\left|\frac{v^\top V_{s,t}}{\sqrt{Th}}\right|^2 1_{\{|\frac{v^\top V_{s,t}}{\sqrt{Th}}|>U\}}\right] = 0. \tag{B.15}$$

Note that for $R = pq/2(p+q) > 2$:

$$\begin{aligned}
&E\left[\left|\frac{v^\top V_{s,t}}{\sqrt{Th}}\right|^2 1_{\{|\frac{v^\top V_{s,t}}{\sqrt{Th}}|>U\}}\right] \\
&= E\left[\left|\frac{v^\top V_{s,t}}{\sqrt{Th}}\right|^2 \left(\frac{U}{U}\right)^{R-2} 1_{\{|\frac{v^\top V_{s,t}}{\sqrt{Th}}|>U\}}\right] \\
&< E\left[\frac{|v^\top V_{s,t}|^2}{Th}\left|\frac{v^\top V_{s,t}}{\sqrt{Th}}\right|^{R-2} U^{-(R-2)} 1_{\{|\frac{v^\top V_{s,t}}{\sqrt{Th}}|>U\}}\right] \\
&\leq E\left[\frac{|v^\top V_{s,t}|^R}{Th}(\sqrt{Th}U)^{-(R-2)}\right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} K\left(\frac{s-t}{Th}\right) E\left[\left|\frac{v^\top V_{s,t}}{\sqrt{Th}}\right|^2 1_{\{|\frac{v^\top V_{s,t}}{\sqrt{Th}}|>U\}}\right] \\
&< (\sqrt{Th}U)^{-(R-2)}\frac{1}{Th}\sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} K\left(\frac{s-t}{Th}\right) E[|v^\top V_{s,t}|^R] \\
&\leq (\sqrt{Th}U)^{-(R-2)}\frac{1}{Th}\sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} K\left(\frac{s-t}{Th}\right) \|\psi(\varepsilon_{s+1})\|_p^R \|Z_{s,t}\|_q^R + o(1) \\
&= O\left((Th)^{-(R-2)/2}\right) = o(1) \text{ as } Th\to\infty,
\end{aligned} \tag{B.16}$$

where p and q are defined in Assumption A.4. The application of Hölder's inequality in the penultimate inequality follows the logic in Lemma $C.3$. The $o(1)$ appears because $E[\psi(\varepsilon_{s+1})X_s] = E[E[\psi(\varepsilon_{s+1})|X_s]X_s]$ and $E[\psi(\varepsilon_{s+1})|X_s] = o(1)$ by condition A.2(ii). Thus, the condition is verified.

Verification of (B.13) *and* (B.13). We define, for any $u \in \mathbb{N}$, the indices $t - \lfloor Th\rfloor \leq s_1 < s_2 < ... < s_u < s_u + k \leq s_u + k' < t$. We focus on the time points prior to t because data after t are equivalent to points prior due to data reflection. We want to show that for a $R > 2$ and for any measurable function $g : \mathbb{R}^u \to \mathbb{R}$ with $\sup_{v\in\mathbb{R}^u} \|g(v)\| \leq 1$, the following:

$$\left|\frac{1}{Th}Cov\Big(\underbrace{g(v^\top V_{s_1,t}/\sqrt{Th}, v^\top V_{s_2,t}/\sqrt{Th}, ..., v^\top V_{s_u,t}/\sqrt{Th})v^\top V_{s_u,t}}_{\equiv A_{s_u,t}}, \underbrace{v^\top V_{s_u+k,t}}_{\equiv B_{s_u+k,t}}\Big)\right| = O\left(\frac{k^{-\varphi^*\frac{R-2}{R-1}}}{Th}\right) \tag{B.17}$$

$$\left| \frac{1}{Th} Cov \left(\underbrace{g(v^\top V_{s_1,t}/\sqrt{Th}, v^\top V_{s_2,t}/\sqrt{Th}, ..., v^\top V_{s_u,t}/\sqrt{Th})}_{\equiv g}, \underbrace{v^\top V_{s_u+k,t} v^\top V_{s_u+k',t}}_{\equiv D_{s_u+k,s_u+k'}} \right) \right| = O\left(\frac{k^{-\tilde{\varphi}}}{Th} \right). \quad \text{(B.18)}$$

To bind these covariances, we adopt a similar strategy as in our proof of Lemma $C.1$. First, define $\mathcal{F}_{-\infty}^{s_u} = \sigma(V_{s_u}, V_{s_u-1}, V_{s_u-2}, ...)$, the L_1-mixingale type coefficient $\gamma(\mathcal{F}_{-\infty}^{s_u}, B_{s_u+k,t}) = \|E(B_{s_u+k,t}|\mathcal{F}_{-\infty}^{s_u}) - E(B_{s_u+k,t})\|_1$, $Q_{|A_{s_u,t}|}$ to be the quantile of $|A_{s_u,t}|$, and $G_{|B_{s_u+k,t}|}$ to be the generalized inverse of $x \mapsto \int_0^x Q_{|B_{s_u+k,t}|}(u')du'$. We start with showing the first statement:

$$\begin{aligned}
\frac{1}{Th}|Cov(A_{s_u,t}, B_{s_u+k,t})| &\leq \frac{2}{Th} \int_0^{\gamma(\mathcal{F}_{-\infty}^{s_u}, B_{s_u+k,t})/2} Q_{|A_{s_u,t}|} \circ G_{|B_{s_u+k,t}|}(u')du' \\
&\leq \frac{2}{Th} \int_0^{\|B_{s_u+k,t}\|_1} 1_{\{u' < \gamma(\mathcal{F}_{-\infty}^{s_u}, B_{s_u+k,t})/2\}} Q_{|A_{s_u,t}|} \circ G_{|B_{s_u+k,t}|}(u')du' \\
&\leq \frac{1}{Th} \{\gamma(\mathcal{F}_{-\infty}^{s_u}, B_{s_u+k,t})\}^{\frac{R-2}{R-1}} \left(\int_0^{\|B_{s_u+k,t}\|_1} \left[Q_{|A_{s_u,t}|} \circ G_{|B_{s_u+k,t}|}(u') \right]^{R-1} du' \right)^{1/(R-1)} \\
&\leq \frac{C}{Th} \tau_k^{*\frac{R-2}{R-1}} \left[\left\{ \int_0^1 Q_{|A_{s_u,t}|}^R(y)dy \right\}^{\frac{R-1}{R}} \left\{ \int_0^1 Q_{|B_{s_u+k,t}|}^R(y)dy \right\}^{\frac{1}{R}} \right]^{1/(R-1)} \\
&= \frac{C}{Th} \tau_k^{*\frac{R-2}{R-1}} \|A_{s_u,t}\|_R \|B_{s_u+k,t}\|_R^{\frac{1}{R-1}},
\end{aligned}$$

where the second last equality follows from a change of variables, and Hölder's inequality twice (see proof of Lemma $C.1$ for further details). Note that because $sup_{v \in \mathbb{R}^u} \|g(v)\| \leq 1$:

$$\begin{aligned}
\|A_{s_u,t}\|_R &\leq \|v^\top \{\psi(\varepsilon_{s_u+1}) Z_{s_u,t} - E[\psi(\varepsilon_{s_u+1}) Z_{s_u,t}]\}\|_R \\
&\leq \|\psi(\varepsilon_{s_u+1}) Z_{s_u,t}\|_R + o(1) = O(1)
\end{aligned}$$

where $\|\psi(\varepsilon_{s_u+1}) Z_{s_u,t}\|_R = O(1)$ as demonstrated in Lemma $C.3$. The same argument is applicable to $\|B_{s_u+k,t}\|_R$. Using the assumption on the τ-mixing coefficient, $\tau_k^* = O(k^{-\varphi^*})$ and $\varphi^* > (R-1)/(R-2)$, we get

$$\frac{1}{Th}|Cov(A_{s_u,t}, B_{s_u+k,t})| = O\left(\frac{k^{-\varphi^* \frac{R-2}{R-1}}}{Th} \right).$$

The key requirement here is that the sequence $\{k^{-\varphi^* \frac{R-2}{R-1}}\}$ is summable (see proof of Lemma $A.3$ for details), which is guaranteed by the condition on φ^*. Hence, condition (2.3) of Theorem 2.1 in Neumann (2013) is satisfied. The verification of (B.18) (i.e. condition 2.4) is more convenient and we invoke the mixing conditions in assumption A.3(iii) instead:

$$\frac{1}{Th}|Cov(C_{s_u,t}, D_{s_u+k,s_u+k'})| \leq \frac{2}{Th} \int_0^{\gamma(\mathcal{F}_{-\infty}^{s_u}, D_{s_u+k,s_u+k'})/2} Q_{|g|} \circ G_{|D_{s_u+k,s_u+k'}|}(u')du'$$

$$\leq \frac{1}{Th}\gamma(\mathcal{F}_{-\infty}^{s_u}, D_{s_u+k,s_u+k'})$$
$$= \frac{1}{Th}\|E(v^\top V_{s_u+k,t}v^\top V_{s_u+k',t}|\mathcal{F}_{-\infty}^{s_u}) - E(v^\top V_{s_u+k,t}v^\top V_{s_u+k',t})\|_1$$
$$= O\left(\frac{(k'-k)^{-\tilde{\varphi}}}{Th}\right),$$

where we have used the fact that $\sup_x |g(x)| \leq 1$ and thus the quantile function is also upper bounded by 1. Given that $\tilde{\varphi} > 1$, the numerator is absolutely summable.

Therefore, 2.1 to 2.4 is verified, and the CLT can be applied. □

Proof of Lemma *A*.6

We analyze the component-wise Lipschitz continuity of g. By Lemma C.1 of Wang and He (2024), $\varepsilon_{s+1}^2\varphi_{\varepsilon_{s+1}^2}(\zeta^\top Z_{s,t})\nu_{b\|\zeta\|}(\zeta^\top Z_{s,t})$ is Lipschitz continuous in $\zeta^\top Z_{s,t}$ with Lipschitz constant 3. This is reduced to 1 with the scaling by 1/3 and the fact that $m_q \in (0,1)$.

Next, we analyze Lipschitz continuity of g with respect to ε_{s+1}^2. Specifically,

$$|g(\varepsilon_{s+1}^2, Z_{s,t}) - g(\varepsilon_{s+1}^{*2}, Z_{s,t})|$$
$$= \left|\frac{1}{3}m_q\left\{\varepsilon_{s+1}^2\varphi_{\varepsilon_{s+1}^2}(\zeta^\top Z_{s,t}) - \varepsilon_{s+1}^{*2}\varphi_{\varepsilon_{s+1}^{*2}}(\zeta^\top Z_{s,t})\right\}\nu_{b\|\zeta\|}(\zeta^\top Z_{s,t})\right|.$$

Note that if $b\|\zeta\| < |\zeta^\top Z_{t,s}|$, then $\nu_{b\|\zeta\|}(\zeta^\top Z_{s,t}) = 0$, and hence it is trivially Lipschitz. We therefore consider the case where $|\zeta^\top Z_{s,t}| \leq b\|\zeta\|$. Since both m_q and $\nu_{b\|\zeta\|}(\zeta^\top Z_{s,t})$ are ≤ 1, $|\frac{1}{3}m_q\{\varepsilon_{s+1}^2\varphi_{\varepsilon_{s+1}^2}(\zeta^\top Z_{s,t}) - \varepsilon_{s+1}^{*2}\varphi_{\varepsilon_{s+1}^{*2}}(\zeta^\top Z_{s,t})\}\nu_{b\|\zeta\|}(\zeta^\top Z_{s,t})| \leq \frac{1}{3}|\varepsilon_{s+1}^2\varphi_{\varepsilon_{s+1}^2}(\zeta^\top Z_{s,t}) - \varepsilon_{s+1}^{*2}\varphi_{\varepsilon_{s+1}^{*2}}(\zeta^\top Z_{s,t})| \equiv \frac{1}{3}|\Delta|$, and we just need to study this difference.

Here, we encounter several cases. If $|\zeta^\top Z_{s,t}| < \min\{\varepsilon_{s+1}^2, \varepsilon_{s+1}^{*2}\}$, then $\varphi_{\varepsilon_{s+1}^2}(\zeta^\top Z_{s,t}) = \varphi_{\varepsilon_{s+1}^{*2}}(\zeta^\top Z_{s,t}) = 0$ are 0, and Lipschitz holds trivially again. So we consider the converse of that: either $\varepsilon_{s+1}^2 \leq |\zeta^\top Z_{s,t}| \leq b\|\zeta\|$ or $\varepsilon_{s+1}^{*2} \leq |\zeta^\top Z_{s,t}| \leq b\|\zeta\|$ or both are true. This results in the following 4 sub-cases are:

Case (a): $\varphi_{\varepsilon_{s+1}^2}(\zeta^\top Z_{s,t}) = \varphi_{\varepsilon_{s+1}^{*2}}(\zeta^\top Z_{s,t}) = 1$. Then, $|\Delta| = |\varepsilon_{s+1}^2 - \varepsilon_{s+1}^{*2}|$, and Lipschizt holds.

Case (b): $\varphi_{\varepsilon_{s+1}^2}(\zeta^\top Z_{s,t}) \in (0,1)$ and $\varphi_{\varepsilon_{s+1}^{*2}}(\zeta^\top Z_{s,t}) = 1$. Then,

$$|\Delta| = \left|\varepsilon_{s+1}^2\left(-1 + \frac{|\zeta^\top Z_{s,t}|}{\varepsilon_{s+1}^2}\right) - \varepsilon_{s+1}^{*2}\right| = ||\zeta^\top Z_{s,t}| - \varepsilon_{s+1}^2 - \varepsilon_{s+1}^{*2}|,$$

and we ask if this is $\leq |\varepsilon_{s+1}^2 - \varepsilon_{s+1}^{*2}|$. For this case, we know that $\varepsilon_{s+1}^2 \leq |\zeta^\top Z_{s,t}| \leq 2\varepsilon_{s+1}^2$ and $\varepsilon_{s+1}^{*2} < |\zeta^\top Z_{s,t}|/2$, which implies $\varepsilon_{s+1}^{*2} < \varepsilon_{s+1}^2$, and so $|\varepsilon_{s+1}^2 - \varepsilon_{s+1}^{*2}| = \varepsilon_{s+1}^2 - \varepsilon_{s+1}^{*2}$. Now, suppose that $\Delta > 0$, we have

$$|\zeta^\top Z_{s,t}| - \varepsilon_{s+1}^2 - \varepsilon_{s+1}^{*2} \leq 2\varepsilon_{s+1}^2 - \varepsilon_{s+1}^2 - \varepsilon_{s+1}^{*2} = \varepsilon_{s+1}^2 - \varepsilon_{s+1}^{*2},$$

hence $\Delta \le \varepsilon_{s+1}^2 - \varepsilon_{s+1}^{*2}$. Next, suppose that $\Delta < 0$, since we have $2\varepsilon_{s+1}^{*2} < |\zeta^\top Z_{s,t}|$, we arrive at

$$-(|\zeta^\top Z_{s,t}| - \varepsilon_{s+1}^2 - \varepsilon_{s+1}^{*2}) = \varepsilon_{s+1}^2 + \varepsilon_{s+1}^{*2} - |\zeta^\top Z_{s,t}| < \varepsilon_{s+1}^2 + \varepsilon_{s+1}^{*2} - 2\varepsilon_{s+1}^{*2} = \varepsilon_{s+1}^2 - \varepsilon_{s+1}^{*2}.$$

Therefore, we conclude that $|\Delta| \le \varepsilon_{s+1}^2 - \varepsilon_{s+1}^{*2}$, and have shown that it is Lipschitz.

Case (c): $\varphi_{\varepsilon_{s+1}^2}(\zeta^\top Z_{s,t}) = 1$ and $\varphi_{\varepsilon_{s+1}^{*2}}(\zeta^\top Z_{s,t}) \in (0,1)$. This case is analogous to case (b), and an identical conclusion can be reached.

Case (d): $\varphi_{\varepsilon_{s+1}^2}(\zeta^\top Z_{s,t}) \in (0,1)$ and $\varphi_{\varepsilon_{s+1}^{*2}}(\zeta^\top Z_{s,t}) \in (0,1)$. Here,

$$|\Delta| = \left|\varepsilon_{s+1}^2\left(-1 + \frac{|\zeta^\top Z_{s,t}|}{\varepsilon_{s+1}^2}\right) - \varepsilon_{s+1}^{*2}\left(-1 + \frac{|\zeta^\top Z_{s,t}|}{\varepsilon_{s+1}^{*2}}\right)\right| = |\varepsilon_{s+1}^2 - \varepsilon_{s+1}^{*2}|,$$

and thus we arrive at the same conclusion. Hence, this shows that g is 1-Lipschitz continuous with respect to ε_{s+1}^2.

To conclude, $g(\varepsilon_{s+1}^2, Z_{s,t})$ is 1-Lipschitz in $(\varepsilon_{s+1}^2, Z_{s,t})$ with respect to the ℓ_1 norm. It is hence also τ-mixing since $(\varepsilon_{s+1}^2, X_s)$ is jointly τ-mixing and 1-Lipschitz maps preserve the mixing property. □

Proof of Lemma *A.7*

The result in (A.6) follows from the dominated convergence argument of equation (C.7) in Wang and He (2024).

To show the second part, first note that $1_{\{\zeta^\top Z_{s,t} > 2\varepsilon_{s+1}^2 > 0\}} \le \varphi_{\varepsilon_{s+1}^2}(\zeta^\top Z_{s,t})$ and $1_{\{|\zeta^\top Z_{s,t}| \le b\|\zeta\|/2\}} \le \nu_{b\|\zeta\|}(\zeta^\top Z_{s,t})$, and so

$$\begin{aligned} E[P_{t,T,b}(\zeta)] &\ge \frac{1}{3} m_q T^{-1} \sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} k_{s,t} E\Big[\varepsilon_{s+1}^2 1_{\{\zeta^\top Z_{s,t} > 2\varepsilon_{s+1}^2 > 0\}} 1_{\{|\zeta^\top Z_{s,t}| \le b\|\zeta\|/2\}}\Big] \\ &= \frac{1}{3} m_q T^{-1} \sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} k_{s,t} E\Big[E\Big(\varepsilon_{s+1}^2 1_{\{\zeta^\top Z_{s,t} > 2\varepsilon_{s+1}^2 > 0\}}\Big| Z_{s,t}\Big) 1_{\{|\zeta^\top Z_{s,t}| \le b\|\zeta\|/2\}}\Big]. \end{aligned} \tag{B.19}$$

Let $u_{s+1} = \varepsilon_{s+1}^2$, then by a change of variables, the conditional density of u_{s+1} given $Z_{s,t}$ can be written as, using the notation of assumption A.2:

$$\frac{f_{(s+1)/T,Z}(\sqrt{u}) + f_{(s+1)/T,Z}(-\sqrt{u})}{2\sqrt{u}} \ge \frac{2u_-}{2\sqrt{u}} \ge \frac{2u_-}{2\sqrt{u_0}}.$$

where by assumption, the conditional densities are greater than $u_- > 0$ uniformly in a neighborhood around 0 (i.e. $|u| \le u_0$). Hence,

$$E\Big(\varepsilon_{s+1}^2 1_{\{\zeta^\top Z_{s,t} > 2\varepsilon_{s+1}^2 > 0\}}\Big| Z_{s,t}\Big) = E\Big(u_{s+1} 1_{\{\zeta^\top Z_{s,t}/2 > u_{s+1} > 0\}}\Big| Z_{s,t}\Big)$$

$$= \int_0^{\zeta^\top Z_{s,t}/2} u \left[\frac{f_{(s+1)/T,Z}(\sqrt{u}) + f_{(s+1)/T,Z}(-\sqrt{u})}{2\sqrt{u}} \right] du$$

$$\geq \frac{u_-}{\sqrt{u_0}} \int_0^{\zeta^\top Z_{s,t}/2} u du = \frac{u_-}{2\sqrt{u_0}} \left(\frac{\zeta^\top Z_{s,t}}{2} \right)^2.$$

Therefore,

$$\begin{aligned}
(B.19) &\geq \frac{1}{4} \underbrace{\frac{u_-}{6\sqrt{u_0}} m_q}_{\equiv \bar{u}} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s\neq t}}^{t+\lfloor Th \rfloor} k_{s,t} E\left[(\zeta^\top Z_{s,t})^2 1_{\{|\zeta^\top Z_{s,t}| \leq b\|\zeta\|/2\}} \right] \\
&= \frac{\bar{u}}{4} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s\neq t}}^{t+\lfloor Th \rfloor} k_{s,t} E[(\zeta^\top Z_{s,t})^2] - \frac{\bar{u}}{4} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s\neq t}}^{t+\lfloor Th \rfloor} k_{s,t} E[(\zeta^\top Z_{s,t})^2 \underbrace{1_{\{|\zeta^\top Z_{s,t}| > b\|\zeta\|/2\}}}_{\leq 0.5 E[(\zeta^\top Z_{s,t})] \text{ by (A.6)}}] \\
&\geq \frac{\bar{u}}{8} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s\neq t}}^{t+\lfloor Th \rfloor} k_{s,t} E((\zeta^\top Z_{s,t})^2) \\
&= \frac{\bar{u}}{8} \frac{1}{Th} \sum_{\substack{s=t-\lfloor Th \rfloor \\ s\neq t}}^{t+\lfloor Th \rfloor} K\left(\frac{s-t}{Th} \right) \zeta^\top E(Z_{s,t} Z_{s,t}^\top) \zeta = \frac{\bar{u}}{8} \zeta^\top H(t/T) \zeta + o(1) \\
&\geq \frac{\bar{u}}{8} \rho_1 \|\zeta\|^2 + o(1)
\end{aligned}$$

where the last inequality uses assumption A.7 where $\lambda_{min}(H(t/T)) \geq \rho_1$. □

Proof of Lemma *A*.8

Since we have data reflection, for simplicity we focus on only the half of the data before time point t. Define $\xi_{s,\zeta} \equiv (\varepsilon_{s+1}^2, \zeta^\top Z_{s,t})^\top$ and write

$$\begin{aligned}
P_{t,T,b}(\zeta) &= \frac{1}{3} m_q T^{-1} \sum_{s=t-\lfloor Th \rfloor}^{t-1} k_{s,t} \varepsilon_{s+1}^2 \varphi_{\varepsilon_{s+1}^2}(\zeta^\top Z_{s,t}) \nu_{b\|\zeta\|}(\zeta^\top Z_{s,t}) \\
&= \frac{1}{Th} \sum_{s=t-\lfloor Th \rfloor}^{t-1} \frac{1}{3} m_q K\left(\frac{s-t}{Th} \right) \varepsilon_{s+1}^2 \varphi_{\varepsilon_{s+1}^2}(\zeta^\top Z_{s,t}) \nu_{b\|\zeta\|}(\zeta^\top Z_{s,t}) \\
&\equiv \frac{1}{Th} \sum_{s=t-\lfloor Th \rfloor}^{t-1} V_t(\xi_{s,\zeta}).
\end{aligned}$$

From Lemma *A*.6, we know that $V_t(\xi_{s,\zeta})$ is a τ-mixing process, so the goal is to first approximate it with independent blocks so that standard empirical process techniques can be applied. Define the l-th block of length L to be

$$G_{l,t,L} = V_t(\xi_{t-\lfloor Th \rfloor + (l-1)L,\zeta}) + \ldots + V_t(\xi_{t-\lfloor Th \rfloor + lL - 1,\zeta})$$

for $l = 1, 2, ..., \left\lfloor \frac{\lfloor Th \rfloor}{L} \right\rfloor$. The remainder block:

$$G_{\left\lfloor \frac{\lfloor Th \rfloor}{L} \right\rfloor +1,t,L} = \begin{cases} 0 & \text{if } \left\lfloor \frac{\lfloor Th \rfloor}{L} \right\rfloor \text{ is an integer,} \\ V_t(\xi_{t-\lfloor Th \rfloor + \left\lfloor \frac{\lfloor Th \rfloor}{L} \right\rfloor L,\zeta}) + \ldots + V_t(\xi_{t-1,\zeta}) & \text{otherwise.} \end{cases}$$

Let $\xi_s = (\varepsilon_{s+1}^2, Z_{s,t}^\top)^\top$ and define $U_{l,t,L} = (\xi_{t-\lfloor Th \rfloor+(l-1)L}^\top, ..., \xi_{t-\lfloor Th \rfloor+lL-1}^\top)^\top$ and $\{W_l : l = 1, ..., \left\lfloor \frac{\lfloor Th \rfloor}{L} \right\rfloor + 1\}$ to be i.i.d. uniform random variables in $(0, 1)$. Define the σ-algebras

$$\sigma(U_{1,t,L}, ..., U_{l-2,t,L}) = \sigma(\{\xi_s\}_{s \leq t-\lfloor Th \rfloor+(l-2)L-1}) \equiv \mathcal{F}_{\xi,l}, \quad \text{for } l = 3, ..., \left\lfloor \frac{\lfloor Th \rfloor}{L} \right\rfloor + 1.$$

By Lemma 5 of Dedecker and Prieur (2004) and similar to the construction in Appendix A of Pouzo (2024), for $l \geq 3$, there exists $U_{l,t,L}^* = (\xi_{t-\lfloor Th \rfloor+(l-1)L}^{*\top}, ..., \xi_{t-\lfloor Th \rfloor+lL-1}^{*\top})^\top$ that are distributed identically to $U_{l,t,L}$, and by extension $\{\xi_l^*\}$ has the same distribution as $\{\xi_l\}$, such that

1. $U_{l,t,L}^*$ is measurable with respect to $\mathcal{F}_{\xi,l} \vee \sigma(U_{l,t,L}) \vee \sigma(W_l)$;
2. $U_{l,t,L}^*$ is independent of $\{U_{m,t,L}\}_{m \leq l-2}$;
3. $\|U_{l,t,L} - U_{l,t,L}^*\|_1 = \tau(\mathcal{F}_{\xi,l}, U_{l,t,L})$.

For $l = 1, 2$, set $U_{l,t,L} = U_{l,t,L}^*$. Furthermore, note that by points (1) and (2), $\{U_{2l,t,L}^*\}_{2 \leq 2l \leq \left\lfloor \frac{\lfloor Th \rfloor}{L} \right\rfloor +1}$ and $\{U_{2l-1,t,L}^*\}_{1 \leq 2l-1 \leq \left\lfloor \frac{\lfloor Th \rfloor}{L} \right\rfloor +1}$ are independent sequences. Next,

$$\begin{aligned} |P_{t,T,b}(\zeta) - E[P_{t,T,b}(\zeta)]| &= \left| \frac{1}{Th} \sum_{s=t-\lfloor Th \rfloor}^{t-1} V_t(\xi_{s,\zeta}) - E[V_t(\xi_{s,\zeta})] \right| = \left| \frac{1}{Th} \sum_{l=1}^{\left\lfloor \frac{\lfloor Th \rfloor}{L} \right\rfloor +1} G_{l,t,L}(\zeta) - E[G_{l,t,L}(\zeta)] \right| \\ &\leq \left| \frac{1}{Th} \sum_{l=3}^{\left\lfloor \frac{\lfloor Th \rfloor}{L} \right\rfloor +1} G_{l,t,L}(\zeta) - G_{l,t,L}^*(\zeta) \right| + \left| \frac{1}{Th} \sum_{l=1}^{\left\lfloor \frac{\lfloor Th \rfloor}{L} \right\rfloor +1} \underbrace{G_{l,t,L}^*(\zeta) - E[G_{l,t,L}^*(\zeta)]}_{\equiv \overline{G}_{l,t,L}^*(\zeta)} \right| \\ &\leq \left| \frac{1}{Th} \sum_{l=3}^{\left\lfloor \frac{\lfloor Th \rfloor}{L} \right\rfloor +1} G_{l,t,L}(\zeta) - G_{l,t,L}^*(\zeta) \right| + \left| \frac{1}{Th} \sum_{l=1}^{(\left\lfloor \frac{\lfloor Th \rfloor}{L} \right\rfloor +1)/2} \overline{G}_{2l,t,L}^*(\zeta) \right| \\ &+ \left| \frac{1}{Th} \sum_{l=1}^{\left\lfloor \frac{\lfloor Th \rfloor}{L} \right\rfloor /2+1} \overline{G}_{2l-1,t,L}^*(\zeta) \right| \\ &\equiv A_{small}(\zeta) + A_1(\zeta) + A_2(\zeta), \end{aligned}$$

where we have used the notation $G^*_{l,t,L}$ to denote that the original variables $\{\xi_l\}$ have been replaced by the coupled variables, $\{\xi^*_l\}$, and $E[G_{l,t,L}(\zeta)] = E[G^*_{l,t,L}(\zeta)]$ in the second inequality due to the identical distribution. Furthermore, we have made explicit the dependence of $G_{l,t,L}$ on ζ.

Then,

$$P(Z_{t,T,J} \geq x) = P\left(\sup_{\zeta\in\Gamma_J} |P_{t,T,b}(\zeta) - E[P_{t,T,b}(\zeta)]| \geq x\right)$$
$$\leq P\left(\sup_{\zeta\in\Gamma_J} A_{small}(\zeta) \geq \frac{x}{3}\right) + P\left(\sup_{\zeta\in\Gamma_J} A_1(\zeta) \geq \frac{x}{3}\right) + P\left(\sup_{\zeta\in\Gamma_J} A_2(\zeta) \geq \frac{x}{3}\right).$$

We start with the first term:

$$P\left(\sup_{\zeta\in\Gamma_J} A_{small}(\zeta) \geq \frac{x}{3}\right) \leq P\left(\sup_{\zeta\in\Gamma_J} \frac{1}{Th} \sum_{l=3}^{\left\lfloor \frac{\lfloor Th\rfloor}{L}\right\rfloor+1} |G_{l,t,L}(\zeta) - G^*_{l,t,L}(\zeta)| \geq \frac{x}{3}\right)$$
$$\leq \frac{3}{x} E\left[\sup_{\zeta\in\Gamma_J} \frac{1}{Th} \sum_{l=3}^{\left\lfloor \frac{\lfloor Th\rfloor}{L}\right\rfloor+1} |G_{l,t,L}(\zeta) - G^*_{l,t,L}(\zeta)|\right]$$
$$= \frac{3}{x} E\left[\sup_{\zeta\in\Gamma_J} \frac{1}{Th} \sum_{l=3}^{\left\lfloor \frac{\lfloor Th\rfloor}{L}\right\rfloor+1} \left|\sum_{s=t-\lfloor Th\rfloor+(l-1)L}^{t-\lfloor Th\rfloor+lL-1} \{V_t(\xi_{s,\zeta}) - V_t(\xi^*_{s,\zeta})\}\right|\right].$$

Then,

$$|V_t(\xi_{s,\zeta}) - V_t(\xi^*_{s,\zeta})| = \underbrace{K\left(\frac{s-t}{Th}\right)}_{\leq 1} \left|\frac{m_q}{3}\varepsilon^2_{s+1}\varphi_{\varepsilon^2_{s+1}}(\zeta^\top Z_{s,t})\nu_{b\|\zeta\|}(\zeta^\top Z_{s,t}) - \frac{m_q}{3}\varepsilon^{2*}_{s+1}\varphi_{\varepsilon^{2*}_{s+1}}(\zeta^\top Z^*_{s,t})\nu_{b\|\zeta\|}(\zeta^\top Z^*_{s,t})\right|$$
$$\leq [|\varepsilon^2_{s+1} - \varepsilon^{2*}_{s+1}| + \underbrace{|\zeta^\top Z_{s,t} - \zeta^\top Z^*_{s,t}|}_{\leq \|\zeta\|\|Z_{s,t}-Z^*_{s,t}\|_{\ell_2} \leq 1\cdot\|Z_{s,t}-Z_{s,t}\|_{\ell_1}}] \text{ by Lemma } A.6$$
$$\leq \|\xi_s - \xi^*_s\|_{\ell_1}$$

where we have used the fact that $\|\zeta\| \leq 1$ and the property of the ℓ_p norms. Therefore,

$$P\left(\sup_{\zeta\in\Gamma_J} A_{small}(\zeta) \geq \frac{x}{3}\right) \leq \frac{3}{x}\frac{1}{Th} \sum_{l=3}^{\left\lfloor \frac{\lfloor Th\rfloor}{L}\right\rfloor+1} \sum_{s=t-\lfloor Th\rfloor+(l-1)L}^{t-\lfloor Th\rfloor+lL-1} \underbrace{E[\|\xi_s - \xi^*_s\|_{\ell_1}]}_{=\|\xi_s-\xi^*_s\|_1}$$
$$= \frac{3}{x}\frac{1}{Th} \sum_{l=3}^{\left\lfloor \frac{\lfloor Th\rfloor}{L}\right\rfloor+1} \|U_{l,t,L} - U^*_{l,t,L}\|_1$$
$$= \frac{3}{x}\frac{1}{Th} \sum_{l=3}^{\left\lfloor \frac{\lfloor Th\rfloor}{L}\right\rfloor+1} \tau(\mathcal{F}_{\xi,l}, (\xi_{t-\lfloor Th\rfloor+(l-1)L}, ..., \xi_{t-\lfloor Th\rfloor+lL-1}))$$

$$\leq \frac{3}{x} \underbrace{\frac{1}{L} \sup_{l+L+1 \leq l_1 < \ldots < l_L} \tau(\mathcal{F}_{\xi,l}, (\xi_{l_1}, \ldots, \xi_{l_L}))}_{\leq \tau_{L+1}}$$

$$= \frac{3}{x} C_\tau (L+1)^{-r_\tau},$$

where the second last inequality follows from the observation that $\mathcal{F}_{\xi,l}$ and $\xi_{t-\lfloor Th \rfloor+(l-1)L}$ are separated by $L+1$ timepoints, and by recalling the definition of τ_k in assumption A.3. For the final equality, we have $\tau_{L+1} = O((L+1)^{-r_\tau})$ where we use some positive constant C_τ. Finally set $x = 3C^* J\theta \sqrt{\frac{L \log(m_T)}{Th}}$, where C^* is a positive constant that is defined subsequently, then

$$P\left(\sup_{\zeta \in \Gamma_J} A_{small}(\zeta) \geq C^* J\theta \sqrt{\frac{L \log(m_T)}{Th}}\right) \leq \frac{1}{C^* J\theta} \sqrt{\frac{Th}{\log(m_T)}} C_\tau \frac{(L+1)^{-r_\tau}}{\sqrt{L}}. \tag{B.20}$$

Note that that block length should diverge to ∞ but slower than that of Th, so we let $L = (Th)^\alpha$ where $1 > \alpha > 1/r_\tau$. This implies that

$$\text{(B.20)} = O\left(\frac{1}{J}\sqrt{\frac{Th}{\log(m_t)}}(Th)^{-\alpha(r_\tau+1/2)}\right) = O\left(\frac{1}{J}\frac{(Th)^{-\frac{1}{2}-\frac{1}{2r_\tau}}}{\sqrt{\log(m_t)}}\right) = o(1),$$

since $r_\tau > 1$. Next, we study the second term:

$$\sup_{\zeta \in \Gamma_J} A_1(\zeta) = \sup_{\zeta \in \Gamma_J} \left| \frac{1}{Th} \sum_{l=1}^{(\lfloor \frac{\lfloor Th \rfloor}{L} \rfloor + 1)/2} \overline{G}^*_{2l,t,L}(\zeta) \right|$$

$$= \left| \sup_{\zeta \in \Gamma_J} \frac{1}{Th} \sum_{l=1}^{(\lfloor \frac{\lfloor Th \rfloor}{L} \rfloor + 1)/2} \left\{ \sum_{s=t-\lfloor Th \rfloor+(2l-1)L}^{t-\lfloor Th \rfloor+2lL-1} V_t(\xi^*_{s,\zeta}) - E\left[\sum_{s=t-\lfloor Th \rfloor+(2l-1)L}^{t-\lfloor Th \rfloor+2lL-1} V_t(\xi^*_{s,\zeta})\right] \right\} \right|.$$

Recall that $V_t(\xi^*_{s,\zeta}) = \frac{1}{3} m_q K(\frac{s-t}{Th}) \varepsilon^{*2}_{s+1} \varphi_{\varepsilon^{*2}_{s+1}}(\zeta^\top Z^*_{s,t}) \nu_{b\|\zeta\|}(\zeta^\top Z^*_{s,t})$ and note that

$$0 \leq V_t(\xi^*_{s,\zeta}) \leq \underbrace{\frac{1}{3} m_q}_{\equiv m^*} \underbrace{K(\frac{s-t}{Th})}_{\leq 1} \varepsilon^{*2}_{s+1} 1_{\{|\zeta^\top Z^*_{s,t}| > \varepsilon^{*2}_{s+1}\}} 1_{\{|\zeta^\top Z^*_{s,t}| < b\|\zeta\|\}} \leq m^* b\theta,$$

where $\|\zeta\| = \theta$ since $\zeta \in \Gamma_J$. If we were to replace one observation (say $\xi^{*'}_{s,\zeta}$ instead of $\xi^*_{s,\zeta}$ for a single timepoint s), we note that within that observation's block, the variables may all be dependent since we only have block-independent variables. Hence, the worst-case scenario would be that changing one variable at timepoint s induces a change in all the other variables of the same block. This implies that the value of $\sup_{\zeta \in \Gamma_J} A_1(\zeta)$ changes by at most $\frac{2m^* b\theta L}{Th} \equiv \overline{c}_l = \overline{c}$ where L comes from the length of the block that is being changed and since the length is the same for all blocks. Then, by the bounded difference inequality for independent variables (for e.g. proposition 2.15.3 in van der Vaart and Wellner, 2023), we get for any $\epsilon > 0$,

$$
\begin{aligned}
P\left(\sup_{\zeta\in\Gamma_J} A_1(\zeta) - E\left[\sup_{\zeta\in\Gamma_J} A_1(\zeta)\right] \geq \epsilon\right) &\leq \exp\left(-\frac{\epsilon^2}{\sum_{l=1}^{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2}\bar{c}^2}\right)\\
&= \exp\left(-\frac{\lfloor Th\rfloor^2\epsilon^2}{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2)4m^{*2}b^2\theta^2L^2}\right)\\
&\leq \exp\left(-\frac{Th^2\epsilon^2}{\frac{Th}{L}4m^{*2}b^2\theta^2L^2}\right)\\
&\leq \exp\left(-\frac{(Th/L)\epsilon^2}{4m^{*2}b^2\theta^2}\right). \qquad \text{(B.21)}
\end{aligned}
$$

Let $\{e_l; l = 1,..., \left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor\}$ to be Rademacher variables that are i.i.d. ± 1. Then,

$$
\begin{aligned}
E\left[\sup_{\zeta\in\Gamma_J} A_1(\zeta)\right] &\leq 2E\left[\sup_{\zeta\in\Gamma_J}\left|\frac{1}{Th}\sum_{l=1}^{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2} e_l\left(\sum_{s=t-\lfloor Th\rfloor+(2l-1)L}^{t-\lfloor Th\rfloor+2lL-1} V_t(\xi^*_{s,\zeta})\right)\right|\right]\\
&= 2E\left[E\left(\sup_{\zeta\in\Gamma_J}\left|\frac{1}{Th}\sum_{l=1}^{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2} e_l\left(\sum_{s=t-\lfloor Th\rfloor+(2l-1)L}^{t-\lfloor Th\rfloor+2lL-1} V_t(\xi^*_{s,\zeta})\right)\right|\middle|\{\xi^*_s\}\right)\right]\\
&\leq 4E\left[E\left(\sup_{\zeta\in\Gamma_J}\left|\frac{1}{Th}\sum_{l=1}^{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2} e_l\left(\sum_{s=t-\lfloor Th\rfloor+(2l-1)L}^{t-\lfloor Th\rfloor+2lL-1} \zeta^\top Z^*_{s,t}\right)\right|\middle|\{X^*_s\}\right)\right]\\
&= 4E\left[E\left(\sup_{\zeta\in\Gamma_J}\frac{1}{Th}\left|\zeta^\top\left(\sum_{l=1}^{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2}\sum_{s=t-\lfloor Th\rfloor+(2l-1)L}^{t-\lfloor Th\rfloor+2lL-1} Z^*_{s,t}e_l\right)\right|\middle|\{X^*_s\}\right)\right]\\
&\leq 4E\left[E\left(\sup_{\zeta\in\Gamma_J}\frac{1}{Th}\underbrace{\|\zeta\|_1}_{\leq J\theta}\left\|\sum_{l=1}^{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2}\sum_{s=t-\lfloor Th\rfloor+(2l-1)L}^{t-\lfloor Th\rfloor+2lL-1} Z^*_{s,t}e_l\right\|_{\ell_\infty}\middle|\{X^*_s\}\right)\right]\\
&\leq 4J\theta E\left[E\left(\frac{1}{Th}\left\|\sum_{l=1}^{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2}\sum_{s=t-\lfloor Th\rfloor+(2l-1)L}^{t-\lfloor Th\rfloor+2lL-1} Z^*_{s,t}e_l\right\|_{\ell_\infty}\middle|\{X^*_s\}\right)\right]
\end{aligned}
$$

where we have applied the symmetrization theorem (for e.g. Theorem 14.3 in Bühlmann and Van De Geer, 2011) in the first inequality, the contraction theorem (for e.g. Theorem 14.4 of the above-mentioned reference) in the third line, and finally $\|v\|_{\ell_\infty}$ for $v \in \mathbb{R}^k$ is defined to be $\max_{1\le i\le k} |v_i|$. Notice that,

$$
\begin{aligned}
&E\left(\frac{1}{Th}\left\|\sum_{l=1}^{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2}\sum_{s=t-\lfloor Th\rfloor+(2l-1)L}^{t-\lfloor Th\rfloor+2lL-1} Z_{s,t}^* e_l\right\|_{\ell_\infty}\Bigg|\{X_s^*\}\right)\\
&= E\left(\max_{1\le j\le 2m_T}\left|\sum_{l=1}^{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2}\underbrace{\frac{1}{Th}\sum_{s=t-\lfloor Th\rfloor+(2l-1)L}^{t-\lfloor Th\rfloor+2lL-1} Z_{s,t,j}^*}_{\equiv w_l(X_{t,j})} e_l\right|\Bigg|\{X_s^*\}\right),
\end{aligned}
$$

and that conditional on $\{X_s^*\}$, $H_{t,j} \equiv \sum_{l=1}^{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2} w_l(X_{t,j})e_l$ is a mean-zero sub-Gaussian random variable. Then by independence of the Rademacher variables and for $\lambda > 0$:

$$
\begin{aligned}
E[\exp(\lambda H_{t,j})] &= \prod_{l=1}^{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2} E[\exp(\lambda w_l(X_{t,j})e_l)] \le \prod_{l=1}^{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2} \cosh(\lambda w_l(X_{t,j}))\\
&\le \prod_{l=1}^{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2} \exp(\frac{\lambda^2 w_l(X_{t,j})^2}{2}) = \exp\left(\frac{\lambda^2}{2}\sum_{l=1}^{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2} w_l(X_{t,j})^2\right),
\end{aligned}
$$

where the first inequality is true for any random variable y with $Ey = 0$ and $|y| \le 1$ (see problem 2.15.10 of van der Vaart and Wellner, 2023), and the last inequality uses $\cosh(y) \le \exp(\frac{y^2}{2})$ for any $y \in \mathbb{R}$. Hence, by Lemma 17.5 of Van de Geer et al. (2016), we have

$$
\begin{aligned}
&E(\max_{1\le j\le m_T} |H_{t,j}||\{X_s^*\})\\
&\le \sqrt{2\log(4m_T)}\max_{1\le j\le 2m_T}\left(\sum_{l=1}^{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2} w_l(X_{t,j})^2\right)^{1/2}\\
&= \sqrt{2\log(4m_T)}\max_{1\le j\le 2m_T}\left(\frac{1}{(Th)^2}\sum_{l=1}^{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2}\left[\sum_{s=t-\lfloor Th\rfloor+(2l-1)L}^{t-\lfloor Th\rfloor+2lL-1} Z_{s,t,j}^*\right]^2\right)^{1/2}\\
&\le \sqrt{2\log(4m_T)/(Th)}\max_{1\le j\le 2m_T}\left(\frac{1}{(Th)}\sum_{l=1}^{(\left\lfloor\frac{\lfloor Th\rfloor}{L}\right\rfloor+1)/2} L\left(\sum_{s=t-\lfloor Th\rfloor+(2l-1)L}^{t-\lfloor Th\rfloor+2lL-1} Z_{s,t,j}^{*2}\right)\right)^{1/2}\\
&\le \sqrt{2c_X\log(4m_T)L/(Th)}
\end{aligned}
$$

where we have used Cauchy-Schwarz in the third line, and in the final inequality we used the fact that we are conditioning on the event $B_{X,t,T}$ as stated in the original lemma. Note that for sufficiently large

$m_T \geq 4$, $\log(4m_T) \leq 2\log(m_T)$, and hence, on $B_{X,t,T}$,

$$E[\sup_{\zeta\in\Gamma_J} A_1(\zeta)] \leq 4J\theta\sqrt{\frac{4c_X L\log(m_T)}{Th}} = \frac{1}{2}C^* J\theta\sqrt{\frac{L\log(m_T)}{Th}},$$

with $C^* = 16\sqrt{c_X}$. Substituting back into (B.21) and setting $\epsilon = \frac{1}{2}C^* J\theta\sqrt{\frac{L\log(m_T)}{Th}}$ yields

$$P\left(\sup_{\zeta\in\Gamma_J} A_1(\zeta) \geq C^* J\theta\sqrt{\frac{Llog(m_T)}{Th}}\right) = P\left(\sup_{\zeta\in\Gamma_J} A_1(\zeta) \geq C^* J\theta\sqrt{\frac{log(m_T)}{(Th)^{1-\alpha}}}\right) \leq \exp\left(-\frac{C^{*2}J^2}{16m^{*2}b^2}\log(m_T)\right),$$

where we have recalled that $L = (Th)^\alpha$ with $1 > \alpha > 1/r_\tau$.

Note that the proof for $A_2(\zeta)$ follows similarly. Hence, we conclude that

$$\begin{aligned}
&P\left(Z_{t,T,J} \geq C^* J\theta\sqrt{\frac{log(m_T)}{(Th)^{1-\alpha}}}\right) \\
&\leq 2P\left(\sup_{\zeta\in\Gamma_J} A_1(\zeta) \geq \frac{C^* J\theta}{3}\sqrt{\frac{log(m_T)}{(Th)^{1-\alpha}}}\right) + P\left(\sup_{\zeta\in\Gamma_J} A_{small}(\zeta) \geq \frac{C^* J\theta}{3}\sqrt{\frac{log(m_T)}{(Th)^{1-\alpha}}}\right) \\
&\leq 2\exp\left(-\frac{C^{*2}J^2}{144m^{*2}b^2}\log(m_T)\right) + \frac{C_Z}{J}\frac{(Th)^{1/2-\alpha(r_\tau+1/2)}}{\sqrt{\log(m_T)}},
\end{aligned}$$

for some constant $C_Z > 0$. □

Proof of Lemma *A.*9

Consider an arbitrary positive u, then

$$\begin{aligned}
P\left(\max_{1\leq j\leq 2m_T} |\Psi_{t,T,j}(\gamma_t^0)| > u\right) \leq &\sum_{j=1}^{2m_T} P\left(\left|T^{-1}\sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} k_{s,t}\psi(\varepsilon_{s+1})Z_{s,t,j}\right| > u\right) \\
&+ \sum_{j=1}^{2m_T} P\left(\left|T^{-1}\sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} k_{s,t}[\psi(\varepsilon_{s+1} + R_{s,t}(X_s)) - \psi(\varepsilon_{s+1})]Z_{s,t,j}\right| > u\right) \\
&\equiv I + II,
\end{aligned}$$

where $R_{s,t(X_s)} = \alpha_{0,s}^{0\top}X_s - \alpha_{0,t}^{0\top}X_s - \alpha_{1,t}^{0\top}X_s((s-t)/T)$. We begin with II. By Chebyshev's inequality and a similar argument to the proof of Proposition 1 along with Assumption A.2(iii) on the approximation of $E[\{\psi(\varepsilon_{s+1}+\epsilon) - \psi(\varepsilon_{s+1})\}^2|X_s]$, we can show that $II = O(\frac{1}{u^2}\frac{m_T s_T h^2}{Th})$. For I, we use a Fuk-Nagaev inequality (specifically, Theorem 3.1 in Babii et al. (2024)) for τ-mixing processes with only finite lower order moments (see Assumption A.3), to arrive at

$$I \leq 2c_1 m_T\frac{(Th)^{1-\kappa}}{u^\kappa} + 8m_T\exp\left(-\frac{c_2(Th)^2u^2}{B_T^2}\right)$$

where c_1, c_2 are positive constants, $\kappa = ((\varphi^* + 1)R - 1)/(\varphi^* + R - 1)$, $R = pq/2(p+q) + 1$, and $B_T^2 = \max_j \sum_{\substack{s=t-\lfloor Th \rfloor \\ s \neq t}}^{t+\lfloor Th \rfloor} \sum_{\substack{k=t-\lfloor Th \rfloor \\ k \neq t}}^{t+\lfloor Th \rfloor} Cov(Z_{s,t,j}\psi(\varepsilon_{s+1}), Z_{k,t,j}\psi(\varepsilon_{k+1}))$.

Therefore,

$$P\left(\max_{1 \leq j \leq 2m_T} \Psi_{t,T,j}(\gamma_t^0) > u\right) \leq 2c_1 m_T \frac{(Th)^{1-\kappa}}{u^\kappa} + 8m_T \exp\left(-\frac{c_2 (Th)^2 u^2}{B_T^2}\right) + \frac{c_3}{u^2} \frac{m_T s_T h^2}{Th},$$

for some positive c_3. Next, consider any $\delta \in (0,1)$ such that

$$2c_1 m_T \frac{(Th)^{1-\kappa}}{u^\kappa} \leq \frac{\delta}{3},$$

then we have for some positive C_1,

$$u \geq C_1 \left(\frac{m_T}{\delta (Th)^{\kappa-1}}\right)^{1/\kappa}.$$

Repeating the same steps for the second terms in the inequality above and we get

$$u \geq C_2 \sqrt{B_T^2 \frac{\log(24 m_T/\delta)}{(Th)^2}} = C_2 \sqrt{\frac{\log(24 m_T/\delta)}{(Th)}},$$

where we have used Lemma $C.2$ to get $B_T^2 = O(Th)$. For the final term,

$$u \geq C_3 \left(\frac{m_T}{\delta (Th)^{3/2}}\right)^{1/2},$$

where we have also used assumption A.8(i), $s_T h^2 \propto (Th)^{-1/2}$.

Hence, we conclude that there exists a positive C such that

$$P\left[\max_{1 \leq j \leq 2m_T} |\Psi_{t,T,j}(\gamma_t^0)| \leq C\left(\left(\frac{m_T}{\delta (Th)^{\kappa-1}}\right)^{1/\kappa} \vee \sqrt{\frac{\log(24 m_T/\delta)}{(Th)}} \vee \left(\frac{m_T}{\delta (Th)^{3/2}}\right)^{1/2}\right)\right] \geq 1 - \delta,$$

for every $\delta \in (0,1)$. □

Proof of Lemma $A.11$

For (i), we have

$$\min_{j \in \mathcal{S}} |\widehat{\gamma}_{t,j}^{\mathcal{S}}| \geq \min_{j \in \mathcal{S}} |\gamma_{t,j}^0| - \max_{j \in \mathcal{S}} |\widehat{\gamma}_{t,j}^{\mathcal{S}} - \gamma_{t,j}^0| > \min\{K_1, K_2\} c^* \sqrt{s_T}\lambda - c^* \sqrt{s_T}\lambda \geq a\sqrt{s_T}\lambda,$$

where we have used Lemma $A.10$ and the beta-min condition in assumption A.10(i). Therefore there exists $a^* > 0$ such that $\min_{j \in \mathcal{S}} |\widehat{\gamma}_{t,j}^{\mathcal{S}}| \geq (a + a^*)\sqrt{s_T}\lambda$.

For (ii), recall that for $j \in \mathcal{S}$, $\widehat{\gamma}_{t,j}^{\mathcal{S}}$ correspond to the elements in $\widetilde{\gamma}_t^{\mathcal{S}}$ which satisfies (A.13):

$$\Psi_{t,T}^{\mathcal{S}}(\widetilde{\gamma}_t^{\mathcal{S}}) + s_\lambda^{\mathcal{S}}(\widetilde{\gamma}_t^{\mathcal{S}}) - G'^{\mathcal{S}}(\widetilde{\gamma}_t^{\mathcal{S}}) = 0.$$

The j-th element of $G'^{\mathcal{S}}(\widetilde{\gamma}_t^{\mathcal{S}})$ is given by

$$G'_\lambda(h^i\alpha_{i,t,j}) = \begin{cases} 0, & \text{if } 0 \leq |h^i\alpha_{i,t,j}| < \lambda, \\ (h^i\alpha_{i,t,j} - \lambda sign(h^i\alpha_{i,t,j}))/(a-1), & \text{if } \lambda \leq |h^i\alpha_{i,t,j}| \leq a\lambda, \\ \lambda sign(h^i\alpha_{i,t,j}), & \text{if } |h^i\alpha_{i,t,j}| > a\lambda, \end{cases} \tag{B.22}$$

for either $i = 0, 1$ as long as $j \in \mathcal{S}$. Given the result in (i), since s_T is at least 1, we conclude that all the elements in $G'^{\mathcal{S}}(\widetilde{\gamma}_t^{\mathcal{S}})$ are given by $\lambda sign(h^i\alpha_{i,t,j})$ for $j \in \mathcal{S}$. Recall that this is equivalent to $s_\lambda^{\mathcal{S}}(\widetilde{\gamma}_t^{\mathcal{S}})$ i.e. $s_\lambda^{\mathcal{S}}(\widetilde{\gamma}_t^{\mathcal{S}}) = G'^{\mathcal{S}}(\widetilde{\gamma}_t^{\mathcal{S}})$ and hence $\Psi_{t,T}^{\mathcal{S}}(\widetilde{\gamma}_t^{\mathcal{S}}) = 0$.

Finally for (iii) we have,

$$\max_{j\in\mathcal{S}^c} |\Psi_{t,T,j}(\widehat{\gamma}_t^{\mathcal{S}})| \leq \max_{j\in\mathcal{S}^c} |\Psi_{t,T,j}(\gamma_t^0)| + \underbrace{\max_{j\in\mathcal{S}^c} |\Psi_{t,T,j}(\widehat{\gamma}_t^{\mathcal{S}}) - \Psi_{t,T,j}(\gamma_t^0)|}_{\equiv B}.$$

By Lemma $A.9$, with probability greater than $1-\delta$, $\max_{j\in\mathcal{S}^c} |\Psi_{t,T,j}(\gamma_{t,j}^0)| \leq C\lambda$. The second term is given by:

$$\max_{j\in\mathcal{S}^c} \left| T^{-1} \sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} k_{s,t} \left[\psi\left(y_{s+1} - \widehat{\alpha}_{0,t}^{\mathcal{S}\top} X_s - \widehat{\alpha}_{1,t}^{\mathcal{S}\top}\left(\frac{s-t}{T}\right) X_s \right) - \psi\left(y_{s+1} - \alpha_{0,t}^{0\top} X_s - \alpha_{1,t}^{0\top}\left(\frac{s-t}{T}\right) X_s \right) \right] Z_{s,t,j} \right|.$$

We can show that very similarly to the proof of Lemma $A.1$ (and specifically (B.6) and (B.7)), along with Lemma $A.10$ (to bind $|\widehat{\gamma}_t^{\mathcal{S}} - \gamma_t^0|$) and assumption A.10(ii), we have $P(B > u) = O\left(\frac{\lambda\sqrt{s_T}}{Thu^2}\right)$, for an arbitrary positive u. Hence, for a $\theta \in (0,1)$, we have $P\left(B \leq c(\frac{\lambda\sqrt{s_T}}{\theta Th})^{0.5}\right) \geq 1-\theta$. Note that $\sqrt{s_T}/(Th) = o(1)$ and so with a large enough sample, the second term would be dominated by the first term with high probability. □

Proof of Lemma $A.12$

We show this using the convex differencing result from Tao and An (1997). Recall the definition of $f(\gamma_t)$ in (A.7). The subdifferential $\partial f(\gamma_t)$ is given by

$$\partial f(\gamma_t) = \left\{ \vartheta_t = (\vartheta_{t,1}, ..., \vartheta_{t,2m_T})^\top \in \mathbb{R}^{2m_T} : \vartheta_j = \begin{cases} \Psi_{t,T,j}(\gamma_t) + \lambda_0 k_{t,j}, & \text{for } 1 \leq j \leq m_T, \\ \Psi_{t,T,j}(\gamma_t) + \lambda_1 k_{t,j}, & \text{for } m_T + 1 \leq j \leq 2m_T \end{cases} \right\},$$

where $k_{t,j} = sign(\gamma_{t,j})$ if $\gamma_{t,j} \neq 0$ and $k_{j,t} \in [-1, 1]$ otherwise. Recall that $\lambda_1, \lambda_2 \propto \lambda$ so we shall use λ instead moving forward.

Next, for our biased oracle solution $\widehat{\gamma}_t^{\mathcal{S}}$, define the set

$$\Upsilon = \left\{ \vartheta_t = (\vartheta_{t,1}, ..., \vartheta_{t,2m_T})^\top : \vartheta_j = \begin{cases} \lambda sign(\widehat{\gamma}_{t,j}^{\mathcal{S}}), & \text{for } j \in \mathcal{S} \\ \Psi_{t,T,j}(\widehat{\gamma}_t^{\mathcal{S}}) + \lambda k_{t,j}, & \text{for } j \in \mathcal{S}^c. \end{cases} \right\},$$

where $k_{j,t}$ ranges over $[-1,1]$. Lemma $A.11$ implies that $\Upsilon \subset \partial f(\widehat{\gamma}_t^{\mathcal{S}})$ with high probability.

Next, we define the following neighborhood around $\widehat{\gamma}_t^{\mathcal{S}}$ as such:

$$\mathcal{N} = \left\{\bar{\gamma}_t \in \mathbb{R}^{2m_T} : |\bar{\gamma}_{t,j} - \widehat{\gamma}_{t,j}^{\mathcal{S}}| < a^* \sqrt{s_T}\lambda,\ \forall j \in \mathcal{S}; |\bar{\gamma}_{t,j} - \widehat{\gamma}_{t,j}^{\mathcal{S}}| < \lambda,\ \forall j \in \mathcal{S}^c\right\}$$

where a^* and C are the same constants in Lemma $A.11$. Then pick any $\bar{\gamma}_t$ in the neighborhood. To prove the lemma, it is sufficient to show (by the results in Tao and An (1997)), that there exists $\vartheta_t^* = (\vartheta_{t,1}^*, ..., \vartheta_{t,2m_T}^*)^\top \in \Upsilon$ such that the following holds with high probability: $\vartheta_{t,j}^* = G'_\lambda(h^i\bar{\alpha}_{i,t,j})$ where $G'_\lambda(h^i\bar{\alpha}_{i,t,j})$ is the j-th element of $G'(\bar{\gamma}_t)$, and $i = 0$ for the first m_T elements and $i = 1$ otherwise. To satisfy this condition, we construct ϑ_t^* as follows:

(i) For $j \in \mathcal{S}$, set $\vartheta_{t,j}^* = \lambda sign(\widehat{\gamma}_{t,j}^{\mathcal{S}}) = \lambda sign(h^i\widehat{\alpha}_{i,t,j}^{\mathcal{S}})$. From (B.22), we can see that $G'_\lambda(h^i\bar{\alpha}_{i,t,j}) = \lambda sign(h^i\bar{\alpha}_{i,t,j})$ if $|h^i\bar{\alpha}_{i,t,j}| > a\lambda$. To establish this, note that with high probability, we have

$$\min_{j\in\mathcal{S}} |h^i\bar{\alpha}_{i,t,j}| \geq \min_{j\in\mathcal{S}} |h^i\widehat{\alpha}_{i,t,j}^{\mathcal{S}}| - \max_{j\in\mathcal{S}} |h^i\widehat{\alpha}_{i,t,j}^{\mathcal{S}} - h^i\bar{\alpha}_{i,t,j}| > (a + a^*)\sqrt{s_T}\lambda - a^*\lambda\sqrt{s_T} \geq a\sqrt{s_T}\lambda,$$

where recall that $h^i\widehat{\alpha}_{i,t,j}^{\mathcal{S}}$ is the j-th element of $\widehat{\gamma}_t^{\mathcal{S}}$ (see (A.14)), which is the center of $\mathcal{N}$. The second inequality is due to part (i) of Lemma $A.11$ and the construction of $\bar{\gamma}_t$. Next we verify that $sign(h^i\bar{\alpha}_{i,t,j}) = sign(h^i\widehat{\alpha}_{i,t,j}^{\mathcal{S}})$ holds with high probability for $j \in \mathcal{S}$:

a. When $h^i\alpha_{i,t,j}^0 > 0$, we have

$$h^i\widehat{\alpha}_{i,t,j}^{\mathcal{S}} = \underbrace{h^i\alpha_{i,t,j}^0}_{>\min\{K_1,K_2\}c^*\sqrt{s_T}\lambda} + \underbrace{(h^i\widehat{\alpha}_{i,t,j}^{\mathcal{S}} - h^i\alpha_{i,t,j}^0)}_{\geq -c^*\sqrt{s_T}\lambda} > 0,$$

$$h^i\bar{\alpha}_{i,t,j} = \underbrace{h^i\alpha_{i,t,j}^0}_{>\min\{K_1,K_2\}c^*\sqrt{s_T}\lambda} + \underbrace{(h^i\bar{\alpha}_{i,t,j} - h^i\alpha_{i,t,j}^0)}_{>-(a^*+c^*)\sqrt{s_T}\lambda} > 0,$$

since for the second equation, $|h^i\bar{\alpha}_{i,t,j} - h^i\alpha_{i,t,j}^0| \leq |h^i\bar{\alpha}_{i,t,j} - h^i\widehat{\alpha}_{i,t,j}^{\mathcal{S}}| + |h^i\widehat{\alpha}_{i,t,j}^{\mathcal{S}} - h^i\alpha_{i,t,j}^0| < (a^*+c^*)\sqrt{s_T}\lambda$ and the final inequality of being positive is guaranteed by choosing $\min\{K_1, K_2\}$ large enough.

b. When $h^i\alpha_{i,t,j}^0 < 0$, we get

$$h^i\widehat{\alpha}_{i,t,j}^{\mathcal{S}} = \underbrace{h^i\alpha_{i,t,j}^0}_{<-\min\{K_1,K_2\}c^*\sqrt{s_T}\lambda} + \underbrace{(h^i\widehat{\alpha}_{i,t,j}^{\mathcal{S}} - h^i\alpha_{i,t,j}^0)}_{\leq c^*\sqrt{s_T}\lambda} < 0,$$

$$h^i\bar{\alpha}_{i,t,j} = \underbrace{h^i\alpha_{i,t,j}^0}_{<-\min\{K_1,K_2\}c^*\sqrt{s_T}\lambda} + \underbrace{(h^i\bar{\alpha}_{i,t,j} - h^i\alpha_{i,t,j}^0)}_{<(a^*+c^*)\sqrt{s_T}\lambda} < 0,$$

where the second inequality is valid again with sufficiently large $\min\{K_1, K_2\}$. Therefore, we have $\vartheta_{t,j}^* = \lambda sign(h^i\widehat{\alpha}_{i,t,j}^{\mathcal{S}}) = \lambda sign(h^i\bar{\alpha}_{i,t,j}) = G'_\lambda(h^i\bar{\alpha}_{i,t,j})$ for $j \in \mathcal{S}$.

(ii) For $j \in \mathcal{S}^c$, by construction of the biased oracle estimator, $\widehat{\gamma}_{t,j}^{\mathcal{S}} = 0$ and $k_{t,j} \in [-1, 1]$. Note that

$$|\bar{\gamma}_{t,j}| \leq \underbrace{|\widehat{\gamma}_{t,j}^{\mathcal{S}}|}_{=0} + \underbrace{|\widehat{\gamma}_{t,j}^{\mathcal{S}} - \bar{\gamma}_{t,j}|}_{<\lambda},$$

and thus $G'_{\lambda}(h^i \bar{\alpha}_{i,t,j}) = 0$ from (B.22). For a sufficiently small constant $c < 1$, we have by Lemma $A.11$, $|\Psi_{t,T,j}(\widehat{\gamma}_t^{\mathcal{S}})| < c\lambda$ for all $j \in \mathcal{S}^c$. Then we can find $k_{t,j}^* \in (-1, 1)$ such that for $j \in \mathcal{S}^c, \vartheta_{t,j}^* = \Psi_{t,T,j}(\widehat{\gamma}_t^{\mathcal{S}}) + \lambda k_{t,j}^* = 0 = G'_{\lambda}(h^i \bar{\alpha}_{i,t,j})$ holds with high probability.

Hence, for both $j \in \mathcal{S}$ and $\mathcal{S}^c$, we have constructed a $\vartheta_t^* \in \Upsilon \subseteq \partial f(\widehat{\gamma}_t^{\mathcal{S}})$ that is equivalent to $G'(\bar{\gamma}_t)$ and thus the intersection is non-empty. By the result in Tao and An (1997), we therefore conclude that $\widehat{\gamma}_t^{\mathcal{S}}$ is a local minimizer of the original penalized objection function (5). □

Proof of Lemma $A.13$ We begin first with a modification of the RSC condition. Consider:

$$\begin{aligned}
[\Psi_{t,T}(\hat{\gamma}_t) - \Psi_{t,T}(\widehat{\gamma}_t^{\mathcal{S}})]^\top (\hat{\gamma}_t - \widehat{\gamma}_t^{\mathcal{S}}) &= [\Psi_{t,T}(\hat{\gamma}_t) - \Psi_{t,T}(\gamma_t^0)]^\top (\hat{\gamma}_t - \gamma_t^0) - [\Psi_{t,T}(\hat{\gamma}_t) - \Psi_{t,T}(\gamma_t^0)]^\top (\widehat{\gamma}_t^{\mathcal{S}} - \gamma_t^0) \\
&\quad - [\Psi_{t,T}(\widehat{\gamma}_t^{\mathcal{S}}) - \Psi_{t,T}(\gamma_t^0)]^\top (\hat{\gamma}_t - \gamma_t^0) + [\Psi_{t,T}(\widehat{\gamma}_t^{\mathcal{S}}) - \Psi_{t,T}(\gamma_t^0)]^\top (\widehat{\gamma}_t^{\mathcal{S}} - \gamma_t^0) \\
&\geq a_1 \|\hat{\zeta}_{1t}\|_2^2 - a_2 \sqrt{\frac{\log m_T}{(Th)^{1-\alpha}}} \|\hat{\zeta}_{1t}\|_1 + a_1 \|\hat{\zeta}_{2t}\|_2^2 - a_2 \sqrt{\frac{\log m_T}{(Th)^{1-\alpha}}} \|\hat{\zeta}_{2t}\|_1 \\
&\quad - \|\Psi_{t,T}(\hat{\gamma}_t) - \Psi_{t,T}(\gamma_t^0)\|_\infty \|\hat{\zeta}_{2t}\|_1 - \|\Psi_{t,T}(\widehat{\gamma}_t^{\mathcal{S}}) - \Psi_{t,T}(\gamma_t^0)\|_\infty \|\hat{\zeta}_{1t}\|_1, \qquad \text{(B.23)}
\end{aligned}$$

where the inequality holds with probability at least $1 - Q_T$, $\hat{\zeta}_{1t} = \hat{\gamma}_t - \gamma_t^0$ and $\hat{\zeta}_{2t} = \widehat{\gamma}_t^{\mathcal{S}} - \gamma_t^0$. Next, we condition on the events $\Sigma_{1T} \equiv \{\max_j |\Psi_{t,T,j}(\widehat{\gamma}_t^{\mathcal{S}}) - \Psi_{t,T,j}(\gamma_t^0)| \leq m_1 (\frac{m_T \sqrt{s_T} \lambda}{Th\sigma_1})^{0.5}\}$ and $\Sigma_{2T} \equiv \{\max_j |\Psi_{t,T,j}(\hat{\gamma}_t) - \Psi_{t,T,j}(\gamma_t^0)| \leq m_2 (\frac{m_T \sqrt{s_T} \lambda}{Th\sigma_2})^{0.5}\}$ for some positive constants m_1,m_2 and $\sigma_1, \sigma_2 \in (0, 1)$. We can show, in a similar fashion to (B.6) and (B.7) and the union bound, that $P(\Sigma_{1T}) \geq 1 - \sigma_1$ and $P(\Sigma_{2T}) \geq 1 - \sigma_2$. Therefore, with probability $1 - Q_T - \sigma_1 - \sigma_2$,

$$\text{(B.23)} \geq$$

$$a_1(\|\hat{\zeta}_{1t}\|_2^2 + \|\hat{\zeta}_{2t}\|_2^2) - \underbrace{\left[a_2 \sqrt{\frac{\log m_T}{(Th)^{1-\alpha}}} + m_1 \sqrt{\frac{m_T \sqrt{s_T} \lambda}{Th\sigma_1}}\right]}_{\equiv E_{1,T}} \|\hat{\zeta}_{1t}\|_1 - \underbrace{\left[a_2 \sqrt{\frac{\log m_T}{(Th)^{1-\alpha}}} + m_2 \sqrt{\frac{m_T \sqrt{s_T} \lambda}{Th\sigma_2}}\right]}_{\equiv E_{2,T}} \|\hat{\zeta}_{2t}\|_1.$$

By convexity[14] of $\frac{\mu}{2}\|\gamma_t\|_2^2 - g(\gamma_t)$ where recall that $g(\gamma_t)$ is from the decomposition in (A.7), we have:

$$[G'(\hat{\gamma}_t) - G'(\widehat{\gamma}_t^{\mathcal{S}})]^\top (\hat{\gamma}_t - \widehat{\gamma}_t^{\mathcal{S}}) \leq \mu \|\hat{\gamma}_t - \widehat{\gamma}_t^{\mathcal{S}}\|_2^2. \qquad \text{(B.24)}$$

[14]See proof of Lemma 3 in Loh and Wainwright (2017) and of Theorem 2 in Loh (2017).

Furthermore, by the first order optimality condition (see (A.9)) we have:

$$\Psi_{t,T}(\hat{\gamma}_t) + \lambda\hat{k}_t - G'(\hat{\gamma}_t) = 0 \implies [\Psi_{t,T}(\hat{\gamma}_t) + \lambda\hat{k}_t - G'(\hat{\gamma}_t)]^\top(\widehat{\gamma}_t^{\mathcal{S}} - \hat{\gamma}_t) = 0, \tag{B.25}$$

where $\hat{k}_t \in \partial\|\hat{\gamma}_t\|$. Combine (B.24),(B.25) along with the modified RSC condition and we have:

$$a_1(\|\hat{\zeta}_{1t}\|_2^2 + \|\hat{\zeta}_{2t}\|_2^2) - \mu\|\hat{\gamma}_t - \widehat{\gamma}_t^{\mathcal{S}}\|_2^2 - E_{1,T}\|\hat{\zeta}_{1t}\|_1 - E_{2,T}\|\hat{\zeta}_{2t}\|_1 \leq [G'(\widehat{\gamma}_t^{\mathcal{S}}) - \Psi_{t,T}(\widehat{\gamma}_t^{\mathcal{S}})]^\top(\hat{\gamma}_t - \widehat{\gamma}_t^{\mathcal{S}}) + \lambda\hat{k}_t(\widehat{\gamma}_t^{\mathcal{S}} - \hat{\gamma}_t).$$

By Lemma $A.12$, $\widehat{\gamma}_t^{\mathcal{S}}$ is a local minimizer of the original penalized loss function and therefore, $\Psi_{t,T}(\widehat{\gamma}_t^{\mathcal{S}}) + \lambda\widehat{k}_t^{\mathcal{S}} - G'(\widehat{\gamma}_t^{\mathcal{S}}) = 0$, where $\widehat{k}_t^{\mathcal{S}} \in \partial\|\widehat{\gamma}_t^{\mathcal{S}}\|_1$ and thus

$$\begin{aligned} a_1(\|\hat{\zeta}_{1t}\|_2^2 + \|\hat{\zeta}_{2t}\|_2^2) - \mu\|\hat{\gamma}_t - \widehat{\gamma}_t^{\mathcal{S}}\|_2^2 - E_{1,T}\|\hat{\zeta}_{1t}\|_1 - E_{2,T}\|\hat{\zeta}_{2t}\|_1 &\leq \lambda\widehat{k}_t^{\mathcal{S}\top}(\hat{\gamma}_t - \widehat{\gamma}_t^{\mathcal{S}}) + \lambda\hat{k}_t(\widehat{\gamma}_t^{\mathcal{S}} - \hat{\gamma}_t) \\ &= \lambda\widehat{k}_t^{\mathcal{S}\top}\hat{\gamma}_t - \lambda\|\widehat{\gamma}_t^{\mathcal{S}}\|_1 + \lambda\hat{k}_t\widehat{\gamma}_t^{\mathcal{S}} - \lambda\|\hat{\gamma}_t\|_1 \\ &\leq \lambda\widehat{k}_t^{\mathcal{S}\top}\hat{\gamma}_t - \lambda\|\hat{\gamma}_t\|_1. \end{aligned} \tag{B.26}$$

where we have used the property of subgradients in the equality, and the final inequality comes from the fact that $\lambda\hat{k}_t\widehat{\gamma}_t^{\mathcal{S}} \leq \lambda\|\hat{k}_t\|_\infty\|\widehat{\gamma}_t^{\mathcal{S}}\|_1 \leq \lambda\|\widehat{\gamma}_t^{\mathcal{S}}\|_1$. Following the proof of Theorem 1, we have $\|\hat{\zeta}_{1t}\|_1 \leq 4\sqrt{s_T}\|\hat{\zeta}_{1t}\|_2$ and similarly for $\hat{\zeta}_{2t}$. Furthermore, let $E_T = \max\{E_{1,T}, E_{2,T}\}$, then label the left hand side of (B.26) as (LHS) and we have:

$$a_1(\|\hat{\zeta}_{1t}\|_2^2 + \|\hat{\zeta}_{2t}\|_2^2) - \mu\|\hat{\gamma}_t - \widehat{\gamma}_t^{\mathcal{S}}\|_2^2 - 4E_T\sqrt{s_T}(\|\hat{\zeta}_{1t}\|_2 + \|\hat{\zeta}_{2t}\|_2) \leq (LHS).$$

By the Parallelogram law, $-\mu\|\hat{\gamma}_t - \widehat{\gamma}_t^{\mathcal{S}}\|_2^2 \geq -2\mu(\|\hat{\zeta}_{1t}\|_2^2 + \|\hat{\zeta}_{2t}\|_2^2)$, hence we get

$$\begin{aligned} &(a_1 - 2\mu)\|\hat{\zeta}_{1t}\|_2^2 + (a_1 - 2\mu)\|\hat{\zeta}_{2t}\|_2^2 - 4E_T\sqrt{s_T}(\|\hat{\zeta}_{1t}\|_2 + \|\hat{\zeta}_{2t}\|_2) \\ &= \{(a_1 - 2\mu)\|\hat{\zeta}_{1t}\|_2 - 4E_T\sqrt{s_T}\}\|\hat{\zeta}_{1t}\|_2 + \{(a_1 - 2\mu)\|\hat{\zeta}_{2t}\|_2 - 4E_T\sqrt{s_T}\}\|\hat{\zeta}_{2t}\|_2 \leq (LHS). \end{aligned}$$

Choose Th large enough such that $4E_T\sqrt{s_T} \leq a_1 - 2\mu$, then we get

$$0 \leq \lambda\widehat{k}_t^{\mathcal{S}\top}\hat{\gamma}_t - \lambda\|\hat{\gamma}_t\|_1.$$

Again, recall that $\lambda\widehat{k}_t^{\mathcal{S}\top}\hat{\gamma}_t \leq \lambda\|\hat{\gamma}_t\|_t$, hence this implies that $\widehat{k}_t^{\mathcal{S}\top}\hat{\gamma}_t = \|\hat{\gamma}_t\|_1$. We showed that $\max_{j\in\mathcal{S}^c}|\widehat{k}_t^{\mathcal{S}}| < 1$ in the proof of Lemma $A.12$, which together with the equality implies that the $supp(\hat{\gamma}_t) \subseteq \mathcal{S}$, where *supp* refers to the set of indices that correspond to non-zero coefficients.

Regarding uniqueness, we note that given the result above, all stationary points $\hat{\gamma}_t$ of (5) are supported in $\mathcal{S}$ and satisfies $\hat{\gamma}_t = (\widetilde{\gamma}_t^{\mathcal{S}\top}, 0_{\mathcal{S}^c}^\top)^\top$, where we recall that $\widetilde{\gamma}_t^{\mathcal{S}}$ are the solutions to the biased oracle problem. Lemma 1 of Loh and Wainwright (2017) shows that the biased oracle problem is strictly convex and therefore, $\widetilde{\gamma}_t^{\mathcal{S}}$ is unique. By extension, $\hat{\gamma}_t$ is also unique. □

Appendix C. Results on τ-mixing

Let $\{V_t\}_{t\in\mathbb{Z}}$ be a sequence of random variables. Define the L_1 mixingale-type coefficient: $\gamma_k = sup_t \gamma(\mathcal{F}^t_{-\infty}, V_{t+k})$, where $\mathcal{F}^t_{-\infty} = \sigma(V_t, V_{t-1}, ...)$ and

$$\gamma(\mathcal{F}^t_{-\infty}, V_{t+k}) = \|E(V_{t+k}|\mathcal{F}^t_{-\infty}) - E(V_{t+k})\|_1.$$

Note that we do not require $\{V_t\}_{t\in\mathbb{Z}}$ to be centered nor does the mixingale condition require stationarity since the coefficient is taken to be the supremum over t. Recall from Assumption A.3 that for a random variable $x \in \mathbb{R}$, $\|x\|_p$ is defined to be $E(|x|^p)^{1/p}$.

Lemma C.1 (Covariance inequality)**.** *Assume $R > 2$ exists such that $\|V_t\|_R < \infty$ for all t. Then,*

$$|Cov(V_t, V_{t+k})| \leq 2^{\frac{1}{R-1}} \gamma_k^{\frac{R-2}{R-1}} \|V_t\|_R^{\frac{R-1}{R}} \|V_{t+k}\|_R \leq 2^{\frac{1}{R-1}} \tau_k^{\frac{R-2}{R-1}} \|V_t\|_R^{\frac{R-1}{R}} \|V_{t+k}\|_R,$$

where τ_k is the τ-mixing coefficient.

Proof. Define $Q_{|V_t|}$ to be the quantile function of $|V_t|$, and $G_{|V_{t+k}|}$ to be the generalized inverse of $x \mapsto \int_0^x Q_{|V_{t+k}|}(u)du$. The proof strategy is similar to Lemma A.1.1 in Babii et al. (2024) but with explicit consideration of non-centered random variables with heterogeneous distributions. By Proposition 1 of Dedecker and Doukhan (2003), we have

$$|Cov(V_t, V_{t+k})| \leq 2 \int_0^{\gamma(\mathcal{F}^t_{-\infty}, V_{t+k})/2} Q_{|V_t|} \circ G_{|V_{t+k}|}(u)du.$$

Next, note that

$$\gamma(\mathcal{F}^t_{-\infty}, V_{t+k}) = \|E(V_{t+k}|\mathcal{F}^t_{-\infty}) - E(V_{t+k})\|_1 \leq \|E(V_{t+k}|\mathcal{F}^t_{-\infty})\|_1 + \|E(V_{t+k})\|_1,$$

where both $\|E(V_{t+k}|\mathcal{F}^t_{-\infty})\|_1 \leq \|V_{t+k}\|_1$ by Jensen's inequality and the law of iterated expectations, and $\|E(V_{t+k})\|_1 \leq E[|V_{t+k}|] = \|V_{t+k}\|_1$ likewise. Hence, $\gamma(\mathcal{F}^t_{-\infty}, V_{t+k})/2 \leq \|V_{t+k}\|_1$. So, we can rewrite the integral as:

$$2 \int_0^{\|V_{t+k}\|_1} 1_{\{u<\gamma(\mathcal{F}^t_{-\infty}, V_{t+k})/2\}} Q_{|V_t|} \circ G_{|V_{t+k}|}(u)du.$$

Note that $\frac{R-2}{R-1} + \frac{1}{R-1} = 1$, so by Hölder's inequality:

$$\begin{aligned} &2 \int_0^{\|V_{t+k}\|_1} 1_{\{u<\gamma(\mathcal{F}^t_{-\infty}, V_{t+k})/2\}} Q_{|V_t|} \circ G_{|V_{t+k}|}(u)du \\ &\leq 2 \Bigg(\underbrace{\int_0^{\gamma(\mathcal{F}^t_{-\infty}, V_{t+k})/2} 1du}_{=\gamma(\mathcal{F}^t_{-\infty}, V_{t+k})/2 \quad \leq \gamma_k/2} \Bigg)^{\frac{R-2}{R-1}} \left(\int_0^{\|V_{t+k}\|_1} \left[Q_{|V_t|} \circ G_{|V_{t+k}|}(u) \right]^{R-1} du \right)^{\frac{1}{R-1}}. \end{aligned}$$

Next, by a change of variables, let $u = \int_0^y Q_{|V_{t+k}|}(z)dz$ so that $du/dy = Q_{|V_{t+k}|}(y)$. Furthermore, for the upper limit of integral of $\|V_{t+k}\|_1$, $y = 1$. By definition, we have $G_{|V_{t+k}|}\left(\int_0^y Q_{|V_{t+k}|}(z)dz\right) = y$ so we can simplify the 2nd term:

$$\begin{aligned}\int_0^{\|V_{t+k}\|_1} \left[Q_{|V_t|} \circ G_{|V_{t+k}|}(u)\right]^{R-1} du &= \int_0^1 Q_{|V_t|}(y)^{R-1} Q_{|V_{t+k}|}(y)dy \\ &\leq \left(\int_0^1 Q_{|V_t|}(y)^R dy\right)^{(R-1)/R} \left(\int_0^1 Q_{|V_{t+k}|}(y)^R dy\right)^{1/R},\end{aligned}$$

where the last inequality is due to Hölder's inequality since $\frac{R-1}{R} + \frac{1}{R} = 1$. Finally we have by the properties of the quantile function, $\int_0^1 Q_{|V_t|}(y)^R dy = E|V_t^R|$, and similarly $\int_0^1 Q_{|V_{t+k}|}(y)^R dy = E|V_{t+k}^R|$.

The proof is complete by noticing that the L_1-mixingale coefficients satisfy $\gamma_k \leq \tau_k$ as given by the relations (2.2.13) and (2.2.18) in Dedecker et al. (2007). ☐

Lemma C.2. *Let $\{\bar{V}_s\}_{s\in\mathbb{Z}}$ be a sequence of centered random variables such that $\|\bar{V}_s\|_R < \infty$ for some $R > 2$. Furthermore, let the τ-mixing coefficient satisfy $\tau_k = O(k^{-\varphi})$ for $\varphi > \frac{R-1}{R-2}$. Then*

$$Var\left(\sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} \bar{V}_s\right) \leq \sum_{\substack{s=t-\lfloor Th\rfloor \\ s\neq t}}^{t+\lfloor Th\rfloor} \sum_{\substack{l=t-\lfloor Th\rfloor \\ l\neq t}}^{t+\lfloor Th\rfloor} |Cov(\bar{V}_s, \bar{V}_l)| = O(Th).$$

Proof. The proof is similar to Lemma A.1.2 of Babii et al. (2024) and involves counting terms in the sum, but here we use Lemma $C.1$ instead. ☐

For convenience, we consider a specific application of the covariance inequality to the product of individual forecasts with the score function of the forecast errors.

Lemma C.3. *Define $X_{ti}\psi(\varepsilon_{t+1}) \equiv V_{t,i}$. Under assumptions A.3(i)-(ii), for $k \geq 1$, and $i, j = 1, ..., d$ we can find an $R > 2$ such that*

$$|Cov(V_{t,i}, V_{t+k,j})| \leq 2^{\frac{1}{R-1}} \tau_k^{*\frac{R-2}{R-1}} \|V_{t,i}\|_R^{\frac{R-1}{R}} \|V_{t+k,j}\|_R < \infty,$$

where τ_k^* is defined in assumption A.3(ii).

Proof. The result is obtained if we have an $R > 2$ such that $\|V_{t,i}\|_R < \infty$. Consider $1/p + 1/q = (q+p)/(pq) \equiv 1/r$, where p and q are defined in assumption A.3(i), and $r = pq/(q+p)$. Immediately, from the condition that $p > 2q/(q-2)$, we have $pq/(q+p) = r > 2$. Hence, we can let $R = r > 2$ and so,

$$\|V_{t,i}\|_R = \|V_{t,i}\|_r \leq \|X_{ti}\|_q \|\psi(\varepsilon_{t+1})\|_p < \infty,$$

where we have applied Hölder's inequality and assumption A.3(i) to bind the moments. Therefore, we can directly apply Lemma $C.1$ along with the conditions on the τ-coefficients in assumption A.3(ii) to get our result. □

Lemma C.4. *Define* $X_{ti}X_{tj} \equiv V_{t,(i,j)}$. *Under assumptions A.3(i)-(ii) for* $k \geq 1$, *and indices* $i, j = 1, ..., d$, *we can set* $\tilde{R} \in (2, q/2]$ *such that*

$$|Cov(V_{t,(i,j)}, V_{t+k,(i,j)})| \leq 2^{\frac{1}{\tilde{R}-1}} \tau_k^{\frac{\tilde{R}-2}{\tilde{R}-1}} \|V_{t,(i,j)}\|_{\tilde{R}}^{\frac{\tilde{R}-1}{\tilde{R}}} \|V_{t+k,(i,j)}\|_{\tilde{R}} < \infty,$$

where τ_k *is defined in assumption A.3(ii) and* $q > 4$.

In order to apply Lemma $C.1$, we need to find $\tilde{R} > 2$ such that $\|X_{ti}X_{tj}\|_{\tilde{R}} < \infty$. Note that $\tilde{R}$ is arbitrary in Lemma $C.1$ and need not necessarily equate to R in assumption A.3(ii). By Hölder's inequality, we have $\|X_{ti}X_{tj}\|_{\tilde{R}} \leq \|X_{ti}\|_{2\tilde{R}}\|X_{tj}\|_{2\tilde{R}}$. We require both norms to be bounded and since we have at least 4 finite moments (i.e. $q > 4$), we need $2\tilde{R} \leq q$ which is $\tilde{R} \leq q/2$. Since $q/2 > 2$, we can set $\tilde{R} \in (2, q/2]$ and the result follows from Lemma $C.1$.

Appendix D. Specific loss function derivations

Assume throughout that $\{|u|\}$ is a sequence that goes to 0.

Appendix D.1. Mean absolute error loss.

For $\psi(e) = sign(e)$. Then,

$$E[sign(e+u)|x] = \int_{-u}^{\infty} f_{e|x}(z)dz - \int_{-\infty}^{-u} f_{e|x}(z)dz. \tag{D.1}$$

The first term:

$$\int_{-u}^{\infty} f_{e|x}(z)dz = \int_{0}^{\infty} f_{e|x}(z)dz + \underbrace{\int_{-u}^{0} f_{e|x}(z)dz}_{=f_{e|x}(0)u+o(u)}.$$

The second term in (D.1) is

$$\int_{-\infty}^{-u} f_{e|x}(z)dz = \int_{-\infty}^{0} f_{e|x}(z)dz - \underbrace{\int_{-u}^{0} f_{e|x}(z)dz}_{=f_{e|x}(0)u+o(u)}.$$

Putting these together:

$$E[sign(e+u)|x] = \int_{0}^{\infty} f_{e|x}(z)dz - \int_{-\infty}^{0} f_{e|x}(z)dz + 2f_{e|x}(0) + o(u).$$

By a symmetry restriction on the conditional density (see the discussion in lin-lin loss example below), we have $\int_0^\infty f_{e|x}(z)dz - \int_{-\infty}^0 f_{e|x}(z)dz = 0$.

Appendix D.2. Lin-lin loss

For $\psi_q(e) = q1_{e>0} + (q-1)1_{e<0}$. We have,

$$E[\psi_q(e+u)|x] = q\int_{-u}^{\infty} f_{e|x}(z)dz + (q-1)\int_{-\infty}^{-u} f_{e|x}(z)dz. \tag{D.2}$$

The derivation then mirrors that for the absolute error case, so we get:

$$\begin{aligned}
E[\psi_q(e+u)|x] &= q\left[\int_{0}^{\infty} f_{e|x}(z)dz + f_{e|x}(0)u\right] + (q-1)\left[\int_{-\infty}^{0} f_{e|x}(z)dz - f_{e|x}(0)u\right] + o(u) \\
&= \left[q\int_{0}^{\infty} f_{e|x}(z)dz + (q-1)\int_{-\infty}^{0} f_{e|x}(z)dz\right] + f_{e|x}(0)u + o(u) \\
&= \left[qP(e>0|x) + (q-1)P(e<0|x)\right] + f_{e|x}(0)u + o(u) \\
&= \left[q(1-P(e<0|x)) + (q-1)P(e<0|x)\right] + f_{e|x}(0)u + o(u)
\end{aligned}$$

$$= [q - P(e < 0|x)] + f_{e|x}(0)u + o(u).$$

As mentioned in the main text, $q - P(e < 0|x) = 0$ if we assume that $P(e < 0|x) = F_{e|x}(0) = q$ (see for e.g. Chen et al., 2019). We can adjust the intercept term to ensure that this assumption holds. This ties in with our symmetry discussion in the least absolute deviation case above. Over there, we have the median regression so the condition would be $F_{e|x}(0) = 0.5$, which yields symmetry. Nonetheless, with $P(e < 0|x) = q$, we have $E[\psi_q(e + u)|x] = f_{e|x}(0)u + o(u)$.

Next, we verify

$$\begin{aligned}
\Phi(x) = E[\psi_q(e)^2|x] &= E[q^2 1_{e>0} + (q-1)^2 1_{e<0}|x] \\
&= q^2 P(e > 0|x) + (q-1)^2 P(e < 0|x) \\
&= q^2[1 - P(e < 0|x)] + (q-1)^2 P(e < 0|x) \\
&= q^2 + \underbrace{P(e < 0|x)}_{=q} - 2q \underbrace{P(e < 0|x)}_{=q} \\
&= q - q^2 = q(1-q).
\end{aligned}$$

Appendix D.3. Asymmetric squared loss

For $\psi_q(e) = 2e(q1_{e>0} + (1-q)1_{e<0})$. We have

$$E[\psi_q(e+u)|x] = 2q \int_{-u}^{\infty} (z+u) f_{e|x}(z)dz + 2(1-q) \int_{-\infty}^{-u} (z+u) f_{e|x}(z)dz. \tag{D.3}$$

For the first term,

$$2q \int_{-u}^{\infty} (z+u) f_{e|x}(z)dz = 2q \Big\{ \underbrace{\int_{-u}^{\infty} z f_{e|x}(z)dz}_{\equiv A} + \underbrace{\int_{-u}^{\infty} u f_{e|x}(z)dz}_{\equiv B} \Big\}.$$

Then,

$$A = \int_0^{\infty} z f_{e|x}(z)dz + \int_{-u}^{0} z f_{e|x}(z)dz = \int_0^{\infty} z f_{e|x}(z)dz + f_{e|x}(o)\frac{u^2}{2} + o(u^2)$$

where the last equality follows from a Taylor expansion. Next,

$$B = u\left[\int_0^{\infty} f_{e|x}(z)dz + \int_{-u}^{0} f_{e|x}(z)dz\right] = u\left[\int_0^{\infty} f_{e|x}(z)dz + f_{e|x}(o)u + o(u)\right] = u\int_0^{\infty} f_{e|x}(z)dz + o(u^2).$$

Combining the results together, we get

$$2q \int_{-u}^{\infty} (z+u) f_{e|x}(z)dz = 2q \int_0^{\infty} z f_{e|x}(z)dz + 2qu \int_0^{\infty} f_{e|x}(z)dz + o(u^2).$$

The second term of (D.3) can be simplified analogously:

$$2(1-q)\int_{-\infty}^{-u}(z+u)f_{e|x}(z)dz = 2(1-q)\int_{-\infty}^{0} zf_{e|x}(z)dz + 2(1-q)u\int_{-\infty}^{0} f_{e|x}(z)dz + o(u^2).$$

Adding both terms up:

$$E[\psi_q(e+u)|x] = \underbrace{2q\int_{0}^{\infty} zf_{e|x}(z)dz + 2(1-q)\int_{-\infty}^{0} zf_{e|x}(z)dz}_{\equiv T_1}$$
$$+ \underbrace{2qu\int_{0}^{\infty} f_{e|x}(z)dz + 2(1-q)u\int_{-\infty}^{0} f_{e|x}(z)dz}_{\equiv T_2} + o(u^2).$$

For the first term,

$$T_1 = 2E\left[qe1_{\{e>0\}}\middle|x\right] + 2E\left[(1-q)e1_{\{e<0\}}\middle|x\right] = 2E\left[e\big(\underbrace{q1_{\{e>0\}} + (1-q)1_{\{e<0\}}}_{\equiv\omega_q(e)}\big)\middle|x\right].$$

Similar to the conditions for the absolute error loss and the lin-lin loss, an analogous condition imposed for expectile regression here is $E[e\omega_q(e)|x] = 0$ (see for e.g. condition 4 in Man et al., 2024), which implies that $T_1 = 0$.

Similarly we have,

$$T_2 = 2uE[q1_{\{e>0\}} + (1-q)1_{\{e<0\}}|x] = 2uE[\omega_q(e)|x].$$

So we conclude that

$$E[\psi_q(e+u)|x] = 2E[\omega_q(e)|x]u + o(u^2),$$

and thus $M_1(x) = 2E[\omega_q(e)|x]$. Finally,

$$\Phi(x) = E[\psi_q(e)^2|x] = E[4q^2e^21_{\{e>0\}} + 4(1-q)^2e^21_{\{e<0\}}|x] = 4E[e^2\{\underbrace{q^21_{\{e>0\}} + (1-q)^21_{\{e<0\}}}_{=\omega_q(e)^2}\}|x].$$

Appendix E. More empirical results

Table E.4: Comparison of forecast combinations with alternative loss functions for forecast evaluation.

	ASL(0.9)	ASL(0.8)	ASL(0.7)	ASL(0.6)	LL(0.9)	LL(0.8)	LL(0.7)	LL(0.6)
MF								
ASL (q = 0.9)	**0.689**	**0.680**	**0.670**	**0.659**	**0.864**	**0.863**	**0.862**	**0.861**
ASL (q = 0.75)	1.170	1.111	1.045	**0.971**	1.201	1.148	1.088	1.021
LL (q = 0.9)	**0.667**	**0.892**	1.144	1.427	**0.704**	**0.874**	1.064	1.277
LL (q = 0.75)	**0.891**	**0.957**	1.029	1.111	**0.881**	**0.954**	1.034	1.125
LS	2.152	2.023	1.880	1.718	1.701	1.610	1.508	1.393
peLasso	4.971	4.663	4.320	3.934	2.897	2.720	2.524	2.302
EQ	**0.643**	**0.661**	**0.682**	**0.705**	**0.826**	**0.852**	**0.882**	**0.915**
Hist. Avg.	1.162	1.124	1.082	1.034	1.044	1.039	1.033	1.026
AR(p)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
QF (0.6)								
ASL (q = 0.9)	**0.462**	**0.486**	**0.512**	**0.542**	**0.645**	**0.672**	**0.701**	**0.734**
ASL (q = 0.75)	1.111	1.057	**0.997**	**0.929**	1.159	1.112	1.059	1.000
LL (q = 0.9)	**0.661**	**0.915**	1.199	1.519	**0.625**	**0.817**	1.031	1.272
LL (q = 0.75)	2.658	2.650	2.641	2.631	**0.979**	1.055	1.140	1.235
LS	2.126	2.000	1.859	1.701	1.671	1.583	1.484	1.373
peLasso	5.279	4.952	4.587	4.176	2.998	2.814	2.609	2.378
EQ	**0.483**	**0.531**	**0.585**	**0.645**	**0.682**	**0.732**	**0.788**	**0.852**
Hist. Avg.	1.162	1.124	1.082	1.034	1.044	1.039	1.033	1.026
AR(p)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
QF (0.75)								
ASL (q = 0.9)	**0.560**	**0.566**	**0.573**	**0.581**	**0.729**	**0.747**	**0.766**	**0.788**
ASL (q = 0.75)	1.021	**0.973**	**0.918**	**0.857**	1.071	1.033	**0.992**	**0.945**
LL (q = 0.9)	**0.418**	**0.649**	**0.907**	1.198	**0.477**	**0.665**	**0.876**	1.113
LL (q = 0.75)	1.296	1.343	1.395	1.455	0.961	1.030	1.106	1.192
LS	2.087	1.963	1.824	1.667	1.666	1.577	1.477	1.365
peLasso	5.661	5.310	4.918	4.477	3.136	2.943	2.728	2.486
EQ	**0.298**	**0.403**	**0.520**	**0.652**	**0.462**	**0.566**	**0.682**	**0.813**
Hist. Avg.	1.162	1.124	1.082	1.034	1.044	1.039	1.033	1.026
AR(p)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Notes: See Table 3 for the list of abbreviations. Column headers indicate the loss function (and the associated asymmetry parameter) used for forecast evaluation and the values are averages.